Chemical Strategies for Bicyclic Peptide Formation

Philip Michael Morrison

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

The University of Leeds

Astbury Centre for Structural Molecular Biology

Faculty of Biological Sciences

September 2015

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others. Where work which has formed part of jointly-authored publications has been included the contribution of the candidate and the other authors to this work has been explicitly indicated below.

Chapter 2 includes work from "Chemical generation and modification of peptides containing multiple dehydroalanines" Chemical Communications, 2015, 51, 13470-13473, P. M. Morrison, P. J. Foley, S. L. Warriner and M. E. Webb. The candidate carried out all the work described in this paper with the exception of the synthesis of the set of stereochemically-varied PK15 sequence peptides which were made by P. J. Foley. The candidate undertook enzymatic assays of the cyclised products of these peptides with the assistance of P.J. Foley. The manuscript was prepared by the candidate in collaboration with S.L. Warriner and M.E. Webb.

Acknowledgements

Firstly I would like to thank my three supervisors for their advice, ideas and patience. Thanks to Mike for his day-to-day support, allowing me enough freedom to test my own ideas and even inflicting them on some poor Masters students. To Darren and Mike thank you for your useful discussions and tolerance of the endless synthetic chemistry monthly reports. Also thank you to Bruce Turnbull and in particular Stuart Warriner for helpful discussions. I would also like to acknowledge the Wellcome Trust for funding an enriching PhD scheme.

I owe a debt of gratitude to those I've shared labs with in the course of my research. All members of the Webb, Turnbull, Warriner, McPherson and Tomlinson groups have contributed to an enjoyable and proactive research environment. I doubtless would not have navigated the stormy seas of research without the guidance in the chemistry lab of Tom, Dan, Martin, Chadamas, Kristian and Clare, and in the biolab I particularly owe Heather, Jeff, James, Tom B and Darren for showing me the ropes. Ludwig, Matt, Zoe, Ivona, Arthur, Laura and Gemma have been valued colleagues, friends and bake-off adversaries. Thanks also to Christian and Anna in the phage lab for their assistance, protocols and explanations. To the aforementioned Masters students - Sam, Pat and Pete – thanks for your hard work and persistence, I learned as much supervising your research as carrying out my own. A special thank you is reserved for Kat and Diana – for four years we have shared the highs and levelled the lows via debates, jokes and drinks and I will miss finding answers at the bottom of a beer glass.

I also wish to thank my family for their support which has allowed me to follow such a privileged path full of opportunity. To Dad in particular thank you for showing me it can be fun to question things and that by finding out how something works you appreciate it so much more.

Finally thank you to Anna, for putting up with me, not asking about my work too much, and for saying yes.

Abstract

The chemical constraint of proteins has been demonstrated to improve the binding affinity of peptide ligands as well as improving properties such as cell penetrability and resistance to proteases. In this work two new chemical methodologies are presented for such constraint, widening the diversity of peptide structures available starting from a single linear peptide sequence. The combination of these chemical modifications with a library of peptide sequences access new classes of constrained peptides which bind protein surfaces. Such molecules could potentially be used in similar applications to antibodies: therapeutics, diagnostics, and molecular biology research tools.

A two-step strategy to produce stereochemically diverse bicyclic peptides has been developed where the first step involves the conversion of multiple cysteine residues to dehydroalanines. The application of methyl 2,5-dibromovalerate proved superior for this transformation when compared to other chemical modification methodologies. The second step of the strategy involved cyclisation via nucleophilic attack at the dehydroalanine residues by a small molecule core bearing multiple thiol functional groups – generating an alkylated L- or D- cysteine in the peptide sequence. The effect of scrambling the stereochemistry at each of these positions is exemplified via the use of a human plasma kallikrein inhibitor PK15 as a model system.

An alternative strategy to create diverse bicyclic peptides involved the synthesis of a new core molecule based upon Barbas reagent. The core molecule bearing three 2-mesyl-1,3,4-oxadiazole moieties cyclised peptides containing three cysteines efficiently - constraining peptide loops in distinctly different structures from other established core molecules. Attempts were made to create a phagemid-encoded phage display library of peptides to subsequently modify with these chemical methodologies.

A phage display library of Adhirons was screened for binders to the antifungal drug posaconazole. Adhirons are a non-antibody binding protein scaffold bearing variable peptide loops. The screening relied upon chemical derivatisation of the posaconazole structure to enable both screening and hit characterisation.

Table of Contents

Abbreviations

Abu	2-aminobutyric acid
AMC	Aminomethylcoumarin
ATIC	AICAR transformylase IMP cyclohydrolase
BME	β-Mercaptoethanol
BSTG	Bioscreening Technology Group (University of Leeds)
CD	Circular Dichroism
Da	Dalton
Dbu	Dehydrobutyrine
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCC	Dichloroacetyl chloride
Dha	Dehydroalanine
DIAD	Diisopropylazodicarboxylate
DiPCD	Diisopropylcarbodiimide
DIPEA	Diisopropylethylamine
DMAP	Dimethylaminopyridine
DMF	N,N-Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
DTT	Dithiothreitol
E. coli	Escherichia coli
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDT	Ethane-1,2-dithiol
ELISA	Enzyme-linked immunosorbent assay
EPO	Erythropoietin
ES-MS	Electrospray ionisation mass spectrometry
FA	Fluorescence anisotropy
FITC	Fluorescein isothiocyanate
HMDS	Bis(trimethylsilyl)amine
HPK	Human plasma kallikrein
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
LCMS	Liquid chromatography-mass spectrometry
LctM	Lacticin 481 synthetase
LLMOD	Large scale low-mode search
MALDI	Matrix assisted laser desorption/ionisation
mCPBA	m-Chloroperoxybenzoic acid
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
Ms	Mesyl
NBS	N-Bromosuccinimide

NHS	N-Hydroxysuccinimide
NMR	Nuclear magnetic resonance
NRP	Non-ribosomal peptide
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PPI	Protein-protein interaction
PTM	Post translational modification
r.m.s.d.h	Root mean square deviation of heavy atoms
RBS	Ribosome binding site
RNA	Ribonucleic acid
rt	Room temperature
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SICLOPPS	Split-intein circular ligation of peptides and proteins
SPSS	Solid phase peptide synthesis
TBAB	1,3,5-Tris(bromoacetamido)benzene
TBMB	1,3,5-Tris(bromomethyl)benzene
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TMMB	1,3,5-Tris(mercaptomethyl)benzene
TNF	Tumour necrosis factor
TPPO	Triphenylphosphine oxide
Tris	Tris(hydroxymethyl)methylamine
tRNA	Transfer ribonucleic acid
Trt	Trityl
Ts	Tosyl (toluenesulfonyl)
uPA	Urokinase plasminogen activator
WT	Wild type
Z	Carboxybenzyl

Chapter 1: An introduction to constrained peptides

This chapter introduces the potential value of constrained peptides as therapeutics, diagnostics and molecular tools. Examples from both nature and chemical synthesis are used to demonstrate the wide range of potential applications of constrained peptides. Research into generating new classes of constrained peptide including chemical modification methods and peptide library display platforms are described. Of particular interest are the bicyclic peptide ligand classes which have recently been shown to be high affinity inhibitors of protein-protein interactions. The use of dehydroalanine as a reaction handle in peptides is briefly summarised before the aims of this project - incorporating dehydroalanine chemistry and bicyclic peptide formation - are outlined.

1.1 The rise of biologics

1.1.1 Blockbuster biologics

Biologically-derived molecules (biologics) such as peptides and proteins play an ever-more-important role in healthcare - through both therapeutic and diagnostic applications. Of the top ten best-selling drugs of 2013, seven of these were biologics – in addition to insulin therapy there were five monoclonal antibodies and one TNF receptor-IgG Fc fragment fusion protein.[1]

Table 1: The top ten selling prescription drugs worldwide in 2013, showing the increased importance and value of biologic therapies.[1]

Rank	Brand	2013 Sales ($US m)	Type
1	Humira	10,659	Monoclonal antibody
2	Enbrel	8,776	Fusion protein
3	Remicade	8,386	Monoclonal antibody
4	Seretide/Advair	8,251	Small molecule
5	Lantus	7,592	Protein (Insulin)
6	Rituxan	7,503	Monoclonal antibody
7	Avastin	6,751	Monoclonal antibody
8	Herceptin	6,562	Monoclonal antibody
9	Crestor	5,622	Small molecule
10	Abilify	5,500	Small molecule

While the top sellers are currently dominated by antibodies and insulin analogues, there are many other examples of peptide (e.g. cyclosporine, bivalirudin, liraglutide) and protein drugs (e.g. Pulmozyme, EPO). Most drugs fit either into the small molecule (<500 Da) or large macromolecule (>5 kDa) categories; however the space between these has slowly been colonised as a result of the approval of peptide-based drugs.[2, 3]

Antibodies are fundamental for biological research, for example as purification aids, imaging reagents and assay reporters. However antibodies can often be costly to develop or purchase, and batch production can be inconsistent.[4] Molecules which can utilise the same recognition elements as antibodies, but can be produced more consistently, are more stable and are cheaper to use would be of great benefit to researchers. Such molecules would have a wide range of potential uses from diagnostics and therapeutics to lab reagents and purification aids.[5, 6]

There have been concerted research efforts into proteins which can present variable binding loops in a manner similar to antibodies but which are much smaller in size, leading to improved physical properties and simpler bacterial expression.[7] Adhirons are one such molecular class based on a cystatin scaffold.[8] The main advantage of using Adhirons over antibodies for diagnostic, therapeutic and research tool uses is their ease of production; Adhirons are easy to recombinantly express in *Escherichia coli*, require no post-translational modifications and their structure has a greater thermal stability compared to antibodies.[8] Developing antibodies is a more resource-intensive process – requiring more complicated expression and screening procedures and often the correct formation of disulfide bridges and post-translational modifications such as glycosylation.[9]

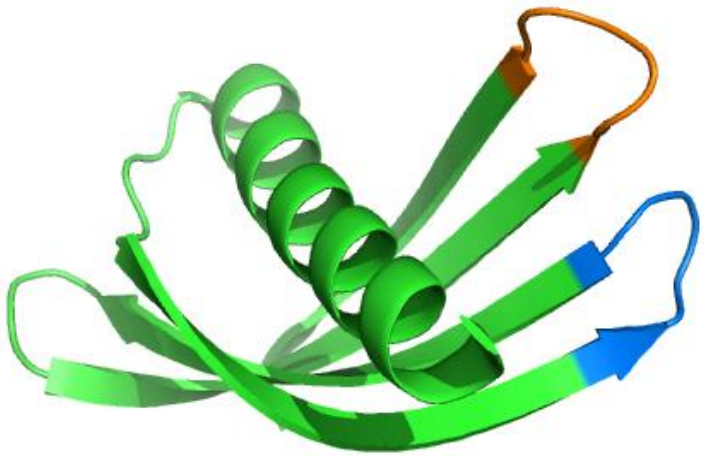


Figure 1: Adhiron – a non-antibody binding protein based upon a cystatin fold. Variable loops are coloured in orange and blue – these are each four residues in the crystal structure but can contain larger loops of up to ten residues in library design. (PDB: 4N6T)

Cystatins were used as the basis for the Adhiron scaffold because their favourable physical properties (in particular the high thermostability and low molecular weight) make them ideal for applications in which high-affinity binders are used, for example in biosensors or as lab reagents.[6] The cystatin sequence used in this work is a consensus sequence of 57 plant cystatins. A consensus sequence should theoretically give the most stable scaffold as evolutionary mutations which have been preserved in multiple variants are conserved while negative mutations specific to one variant are removed.

1.1.2 Peptides in therapeutics and diagnostics

Peptides occupy an underdeveloped region of synthetic chemical space, intermediate between small molecules and proteins.[10] However many peptidic structures of molecular weight 500-5000 Da are found throughout nature, playing a diverse array of roles including as toxins, neurotransmitters and defensive agents. Peptides are natural ligands in many

interactions, or alternatively a peptide can mimic the binding interface of a protein ligand; taking advantage of this, peptides have great potential as therapeutics, diagnostics and research tools.[11] Peptidic molecules are produced in nature via either ribosomal synthesis in the same manner as protein synthesis, or by non-ribosomal enzymatic methods. Both of these methods have been utilised in research in addition to the synthesis of peptides via solid-phase peptide synthesis (SPSS). The natural modularity of peptides lends itself to library screening - combined with the potential incorporation of non-canonical amino acids, peptides therefore provide an accessible and diverse source of potential for drug discovery.[12]

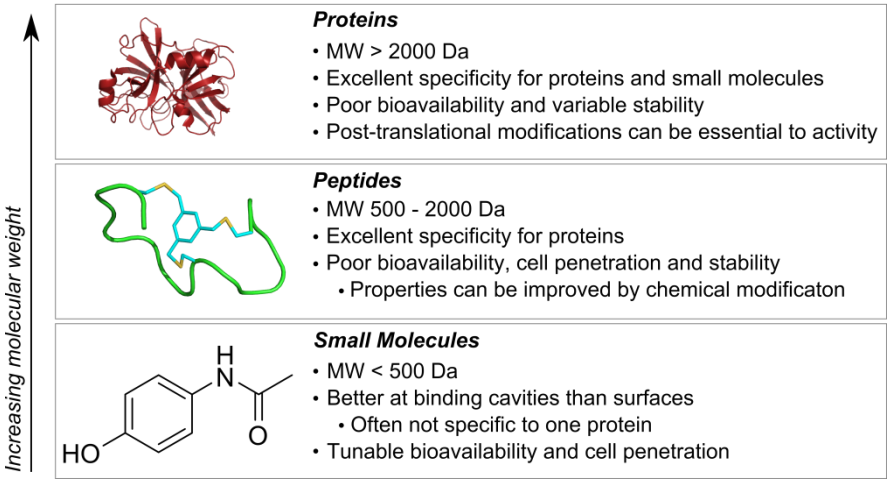


Figure 2: Peptides occupy an intermediate region of chemical space between proteins and small molecules. They share properties but offer inherent advantages in discovery and application compared to these bordering classes.

1.2 Constraining Peptides

While some peptides are active in their linear form, the constraint of these peptides into favourable binding conformations can improve their properties.[13] This has been demonstrated for peptide sequences which were selected as a constrained form[14], although the effect on proteins selected as linear binders is open to debate. It is proposed that constraint of the linear sequence reduces the entropic cost of binding to its target. In many cases constraint is essential for activity, while it also imparts increased resistance to proteases.[14-18] Peptides can be constrained via native methods such as disulfide formation and enzymatic modifications, or alternatively chemical modification can be used; some of the methods used to constrain peptides are described below.

1.2.1 Disulfide constraints

The simplest form of peptide constraint, the oxidative cross-linking of two cysteine residues restricts the peptide's movement. While this is a fundamental of protein structure it is also key to several smaller peptide structures. Conotoxins are potent neurotoxin peptides produced by venomous snails where a peptide is constrained by several disulfide linkages. Two classes exist: the ω- and α- families. ω-Conotoxins are 24-31 residue peptides with three disulfide bonds forming a four-looped structure. They have been shown to selectively block voltage gated calcium channels.[19, 20] α-Conotoxins are related 12-16 residue structures containing two disulfide bonds and therefore two loops of sequence. The α-conotoxin family is known to hit a target other than calcium channels, but it is not clear what this target is.[19] Due to their venomous origins the conotoxins have potential as pain relief therapeutics. For this reason several structures have been developed as potential therapeutics and synthetic libraries based upon conotoxin structures have been produced and screened.[21-23] In peptides containing multiple disulfide bonds there can be issues with switching of dynamic disulfides causing alternative structures to form. To alleviate this issue in their work with disulfide-rich peptide libraries, Heinis and co-workers synthesised an artificial amino acid **3**, bearing two thiols in its side chain.[24]

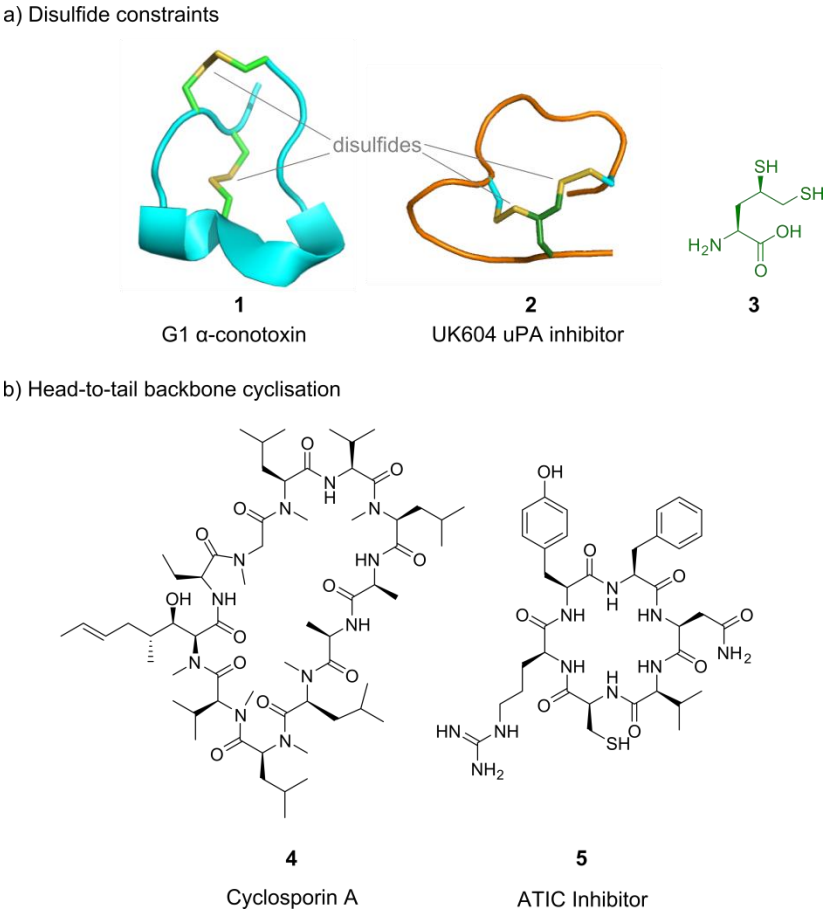


Figure 3: Peptides constrained by disulfide bonds and backbone cyclisations. a) Conotoxins such as G1 **1** are toxins found in venomous marine snails. They are short peptides constrained by disulfide bonds. Heinis and co-workers created synthetic peptides which were constrained by disulfide bonds in a similar manner – the incorporation of the artificial amino acid **3** avoided shuffling of disulfide bonds between cysteines at different positions – leading to a more consistent and stable structure of the urokinase plasminogen activator (uPA) inhibitor UK604 **2**.[24] b) Potent cyclic peptides with continuous backbone include cyclosporin A **4**, a non-ribosomal peptide, and the ATIC inhibitor **5** identified from a SICLOPPS library of small peptides.[25]

1.2.2 Head-to-tail backbone cyclisation

Continuous peptide backbone macrocycles are common features of non-ribosomal peptides such as cyclosporin A **4**. It is a less-common feature of ribosomal peptides but there are select known examples. There may be many more undiscovered but it is an onerous effort to identify these and to confirm that the cyclisation is via the backbone rather than cross-linkage involving a side-chain.[26] Θ-Defensins are of particular note – these peptides are formed from splicing of two ribosomally-expressed peptides. They are antimicrobials expressed in Old World monkeys and orangutans, their human homologues are not translated efficiently due to a premature stop codon mutation.[26] However the putative

historical sequence was used to generate retrocyclin which is not only an antimicrobial but also blocks HIV-1 infection.[27]

Cyclotides are a family of cyclic peptides produced in plants which contain a continuous backbone and are also further cross-linked by disulfide constraints, creating a knotted-like structure.[26, 28] These constraints improve peptide stability, improving resistance to both enzymatic digestion and thermal degradation.[29] They are probably the largest known family of constrained peptide, with an estimated 50000 members in plants.[30] Their native function appears to be in the plant's defences as they show insecticidal properties via inhibition of insect larvae development. However they have also been shown to have an array of other biological activities.[30] This has led to interest in applying these structures in pharmaceuticals and agrochemicals.

To generate cyclic peptide libraries *in vivo*, split-intein mediated cyclisation of peptides and proteins (SICLOPPS) was developed by Benkovic and co-workers.[25, 31] A DnaE split-intein was modified to swap the domains either side of a variable peptide sequence, causing cyclisation of the peptide upon excision. The bacterial expression of cyclic peptide libraries was combined with a host-cell survival genetic screening approach to identify disruptors of protein-protein interactions. From a library of 10^7 peptides an inhibitor of AICAR transformylase/IMP cyclohydrolase (ATIC) dimerisation was discovered ($K_i = 17 \pm 4$ μM).[25] These initial efforts have now been adopted by several groups, with some significant evolutions including the adaptation to fluorescent-reporter phenotypic screens[32], the incorporation of unnatural amino acids[33], and the use of SICLOPPS in both yeast[34] and human[35] cell lines.

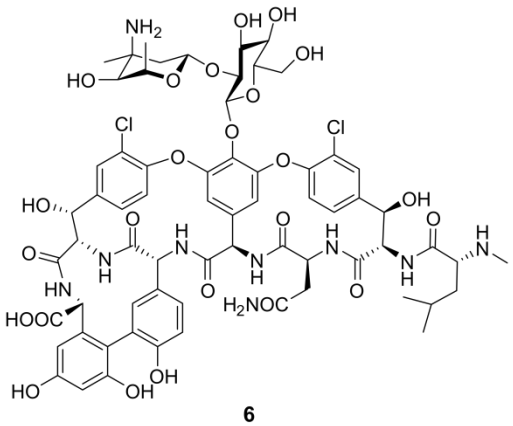
1.2.3 Cross-linked side chains

An incredibly diverse subset of constraints is the cross-linking of side chains. Again this is a common feature of NRPs – witness for example the tyrosine cross-linkage in vancomycin. Lantibiotics are a class of naturally occurring antibiotics produced by lactococcal bacteria which contain thioether linkages.[36] A two-step enzymatic process creates the thioethers: first an enzyme dehydrates serine or threonine to give dehydroalanine and dehydrobutyrine respectively. This is accomplished either via a phosphorylation (Class II enzymes) or an unusual glutamylation which requires the formation of an activated ester between the serine or threonine and the glutamic acid tRNA synthetase complex, which then undergoes elimination (Class I enzymes).[37, 38] A cyclase enzyme then catalyses the formation of thioether bonds in the structure with complete stereochemical control.[39-41] The source of this control is debated since some members of the cyclase family are promiscuous and act on a

range of peptide substrates from other lanthipeptide pathways.[42] The major products of these promiscuous cyclisations have the correct stereochemistry inserted, suggesting perhaps the peptide sequence is the defining factor in stereochemical control rather than the cyclase enzyme.[43]

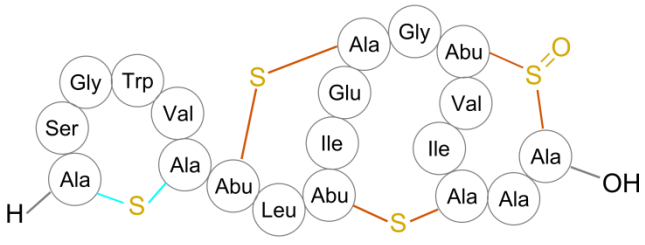
The enzymes involved in the lanthipeptide biosynthetic pathways are largely known, and some have been used *in vivo* on non-native sequences.[44] The specificity of the enzymes is determined by a leader sequence on the peptide which is removed post-modification. The cleavage can be difficult to replicate in *E. coli* production of lanthipeptides from other organisms – often a specific enzyme is required to cleave in the correct position. van der Donk and coworkers have avoided this by genetically incorporating a hydroxy acid at the cleavage site, allowing facile ester cleavage to remove the leader peptide.[45] It has been demonstrated that the leader sequence does not necessarily need to be attached to the substrate: a hybrid of the dehydratase LctM fused to the relevant leader sequence still dehydrated lacticin 481 analogues.[46] An *in vitro* application has allowed the incorporation of nonproteinogenic amino acids into lantibiotic-type structures.[47]

a) Cross-linked Peptides

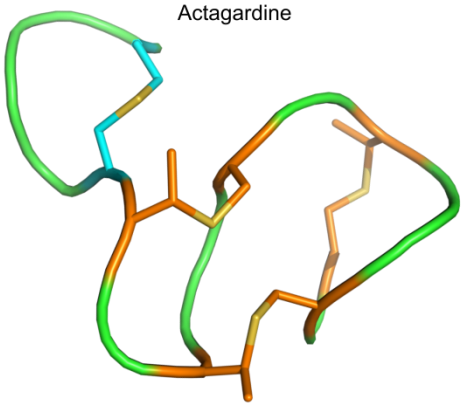


6
Vancomycin

b) Lanthipeptides



7
Actagardine



Actagardine (NMR structure)

Figure 4: Examples of cross-linked peptides. a) Vancomycin **6**, a non-ribosomal peptide product, has an unusual tyrosine-tyrosine linkage and another unusual triaryl cross-bridge. b) Actagardine **7** is a member of the lanthipeptide family, where thioether cross-linking is achieved by enzymatic control of cysteine attack on dehydroalanine or dehydrobutyrine respectively. (Abu = 2-aminobutyric acid) These multiple cross-links hold the peptide loops in defined binding conformations as shown in the NMR structure.

1.2.4 Synthetic cross-linking of side chains

Moving further away from native structures, the combination of peptide structures with organic prosthesis is now a common approach to improve the properties of peptide ligands.

The most frequent variant is the stapling of linear peptides to stabilise α-helical conformations; the many stapling methodologies including triazole formation[48], disulfide formation[49] and alkene metathesis[50-52] Metathesis of all-carbon side-chains has become the dominant methodology in the stapling field.[53-55] Most methodologies require the input of a pair of artificial amino acids to provide biorthogonal functionality or to sufficiently span the helical length. There is ongoing debate as to whether stapling improves peptide binding – although there is a reduced entropic cost of binding, often this is cancelled out by enthalpic penalties. Stapling has been observed to increase protease resistance and in some cases also improve cell penetration (although the precise rationale for the latter is unknown).[16, 17]

An alternative form of constraint is to cross-link across a cyclic peptide. Kritzer and co-workers took this approach to improve the protease resistance of cyclic peptide G1, an inhibitor of the Grb2 SH2 domain.[56] They replaced the disulfide bond in G1 with a covalent backbone and then stapled across two residues. These constraints are not dissimilar to the bicyclic peptides studied by Reymond and co-workers in their efforts to explore peptidic chemical space. Their structures were cross-linked by glutamic acid side chain amide coupling with amine side chains – leading to structures consisting entirely of amino acid subunits which were shown by NMR to have globular structure with rigid backbones.[10]

However these structures may not mimic discontinuous structures often found in protein binding interfaces – Liskamp and co-workers attempted to display 3 separate cyclic peptides on one central core via copper catalysed azide-alkyne cycloaddition.[57] They mimicked the binding loops of HIV gp120 in the discontinuous CD4-binding epitopes – showing that even incorporating just two of the three loops onto their central scaffold was sufficient for CD4 binding. The same group has also used the incorporation of artificial amino acids and subsequent intramolecular cycloadditions in attempts to mimic the peptidic CDE rings of vancomycin.[58]

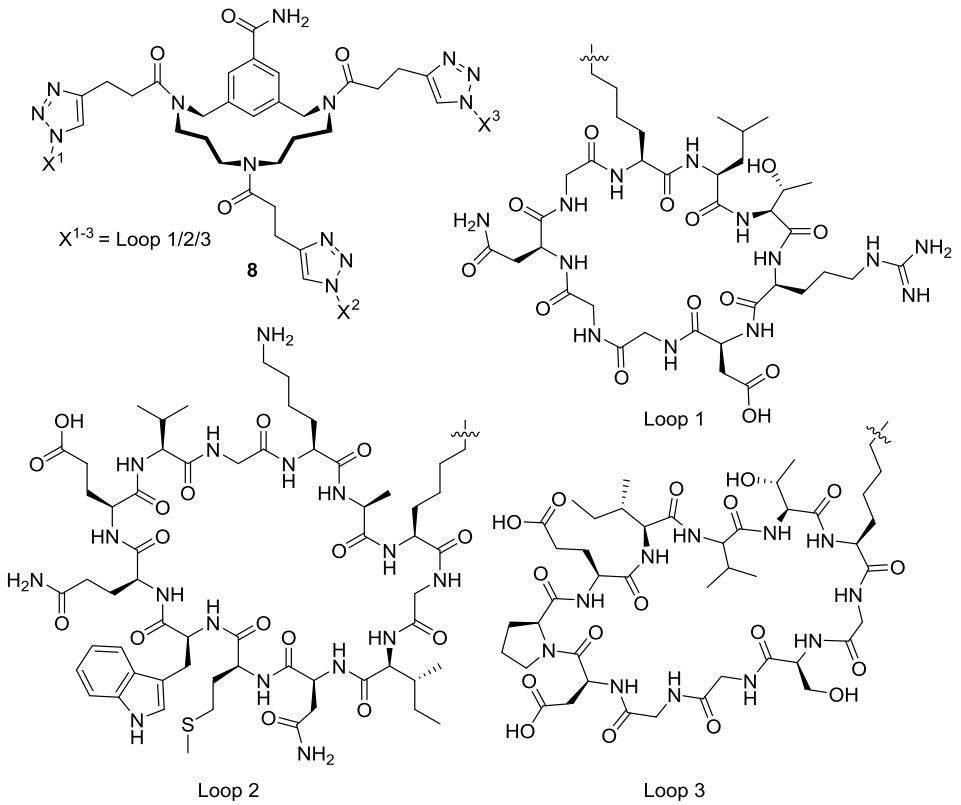
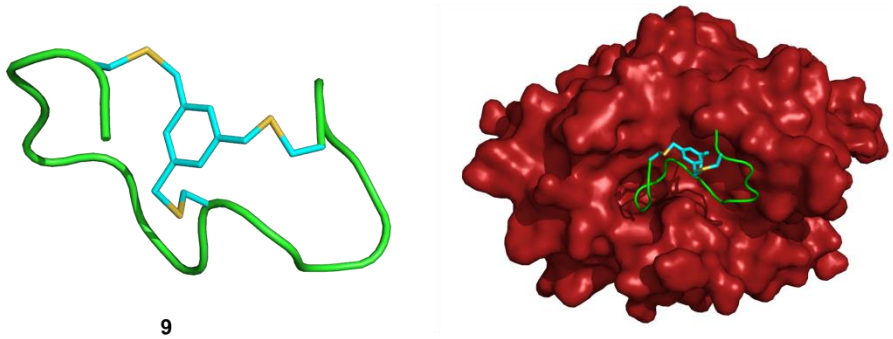


Figure 5: Liskamp and coworkers synthesised **8** in an effort to mimic the HIV gp120 binding loops. The most potent mimic competed with gp120 to bind CD4 (IC_{50} 16.9 μM).[57]

1.2.5 Bicyclic peptides

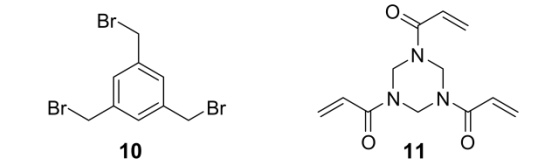
In a further extension of the cyclisations above, a new class of peptidic molecule is emerging consisting of two peptide loops constrained around a central core. These were pioneered as protein surface mimetics by Timmermann *et al.* with the development of CLIPS® technology – the reaction of cysteine residues with benzyl bromide linkers.[59] This was then applied to a phage-displayed peptide library by Winter and co-workers to identify PK15, a bicyclic peptide inhibitor of human plasma kallikrein (K_i = 2 nM). The same screening platform has been used to identify inhibitors of several other protease targets. UK18, an inhibitor of urokinase-type plasminogen activator (uPA), was co-crystallised with uPA, showing both peptide loops bound to the uPA with an interaction surface area of 701 $Å^2$ as shown in Figure 6a.[60]

a)

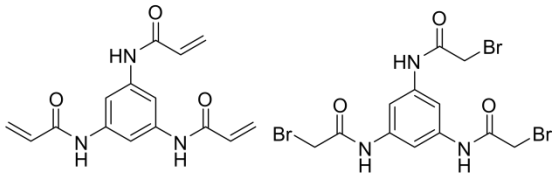


9

b)

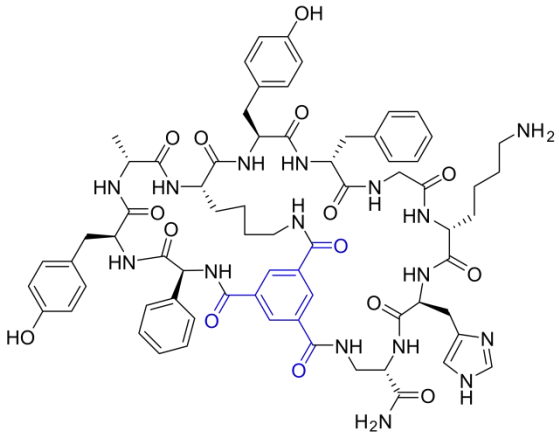


10 11



12 13

c)



14

Figure 6: Bicyclic peptides a) Bicyclic peptide inhibitor UK18: peptide sequence is constrained by a mesitylene core attached to three cysteine residues. The uPA inhibitor is shown bound to its target, with a large interaction area.[60] b) Different organic cores that have been reacted with cysteine-containing peptides by Heinis and co-workers. Different cores hold loops of the same peptide sequence in different orientation due to steric differences and in some cases hydrogen bonding between the core and the loops.[61] c) A bicyclic peptide inhibitor of TNFα identified by Pei and co-workers from a synthetic library containing both L- and D- amino acids including some non-canonical sidechains. A trimesic acid core (blue) is used to form loops via reaction with amine-containing side chains and the N-terminus.[62]

To demonstrate the diversity of structure available, Heinis and co-workers have shown that peptide loops of different sizes are tolerated and have produced phage-displayed libraries containing loops of varying size.[63] They have also created several new core molecules which react with cysteine residues in the same manner as the CLIPS® technology (Figure 6b). It was shown that different cores constraining identical peptide sequences create bicyclic peptides with varied inhibitory strength.[61] This suggests that not only does varying core size affect loop structure but that cores capable of hydrogen bonding to peptide backbones or side chains will further constrain the peptide into a unique binding conformation.[15] A similar bicyclic structure Anticachexin C1 **14**, was generated by Pei and co-workers through the reaction of trimesic acid with terminal amino acids and an internal lysine residue.[62] They screened a synthetic library for TNFα inhibitors – a target traditionally difficult for small molecule drug discovery – identifying a promising inhibitor (IC_{50} 3.1 ±0.3 μM). Their library included several artificial amino acids and D-amino acids.

An inherent problem with early iterations of these compounds was a low residency time *in vivo*. There are two contributing factors to this – firstly high clearance rates as the small structures are filtered from the blood via the kidneys. This particularly affects the bicyclic peptides which were extracellular protease inhibitors. Secondly the structures, although theoretically small enough to passively diffuse through a cell membrane, are not very cell permeable, thus they remain in circulation. To overcome the circulation issues, fusions of the linear peptide sequence were produced with either Fc antibody fragments[64], or with a protein tag which binds to serum albumin.[65] Cyclisation occurs as with the peptide alone to create a fusion of the inhibitor with a large protein partner. After linker optimisation a UK18-Fc fusion was shown to retain high potency (K_i = 76 nM), while terminal half-life had increased to 1.5 days in mice – a 100-fold improvement compared to similar non-conjugated bicyclic peptides.[64] When conjugated to the albumin-binding tag the bicyclic peptide is resistant to proteases and has 50-fold increase in terminal half-life.[65] In an alternative approach to increasing protease resistance, whilst also increasing activity, a uPA inhibitor had a glycine residue replaced by D-Ser.[66] This was guided by the crystal structure of the complex, showing the glycine residue had a large torsional twist, placing the side chain of the D-Ser mutant in a position where it would not sterically clash with the remainder of the peptide.

Suga and co-workers developed a combination of flexizyme technology to incorporate non-canonical amino acids and mRNA display (Section 1.3) to create genetically encoded macrocylic peptides.[67, 68] An *N*-chloroacetyl amino acid was incorporated via the flexizyme system, cyclisation then occurred spontaneously with an encoded C-terminal cysteine to form a thioether cross-linkage. This linkage is non-reducible and so genetically encoded

libraries can be screened to identify hits where the molecule is definitely macrocyclic, in contrast to the disulfide-bridged macrocycles. The fact that screening is of a precyclised state implies that the hits will be superior binders to the related linear pepitdes as demonstrated by Heinis *et al*. Screening targets for the mRNA-displayed macrocycles included VEGFR2, where inhibitors were identified with nanomolar IC_{50} values[69], and Hepatocyte growth receptor (also known as Met receptor) – in this case a binder was dimerised which in turn then brough two Met receptors together and initiated a signalling cascade.[70] The methodology has been combined with reaction with TBMB to create genetically encoded tricyclic peptides.[71] An investigation into the thioether cyclisation preferences when multiple cysteines are present in the encoded sequence showed that the first upstream cysteine is favoured unless it is in position two.[72] Thus the tricyclic linkages can be assumed based upon the first cysteine reacting with the chloroacetyl functionality leaving the other three cysteines to react with TBMB.

1.3 Library Display Methods

As mentioned in section 1.1.2 the modularity of peptides naturally lends itself to library screening. There are several methods of library display which have proved amenable to peptides, and in some cases to the chemical modification of displayed peptides. The key to successful display is the deconvolution of the original peptide sequence – usually this is through nucleic acid encoding of the peptide sequence. However in a synthetic peptide library the options are also extended to mass spectrometry-based deconvolution of fragmented peptide hits.

1.3.1 Phage Display

The use of phage to display a library of peptides or proteins is a well-established *in vitro* screening platform. Bacteriophage are viruses containing single stranded DNA which infect bacteria. Phage have two varieties depending on their effects on infected bacteria – lytic phage are released by cell lysis whereas other filamentous phage types are non-lytic as they are constantly secreted from infected cells. Typically non-lytic filamentous phage such as M13 or Fd are used to display a peptide, protein or antibody fragment library as a fusion with one of the phage coat proteins. The DNA within each phage encodes the library member displayed upon its surface. Binders to a target of interest are identified via an *in vitro* panning procedure, after several rounds of selection strong specific binders can be identified. Libraries can be generated displaying upwards of 10^9 different variants.

The most common phage proteins used to display peptides are the g3p and g8p proteins. As can be seen in Figure 7, g8p is the major coat protein, with each phage containing thousands of copies. g8p fusions therefore display many copies of a peptide per phage molecule, however there is a limit to the size of fusion which can be displayed – large fusions compromise the close packing of the phage coat. g3p is a tail protein with a role in phage infectivity and there are only five copies per phage molecule. Larger conjugates are tolerated however there can be issues with phage infectivity, particularly when polyvalent display is used. Both the packing and infectivity of g8p and g3p conjugates respectively can be overcome by the use of phagemid approaches described below. A further possibility is to use a truncated g3p variant containing the structural domain to allow association with the phage molecule but lacking the infective domain. This truncated variant requires wild type g3p to be provided in a helper phage vector to produce infective phage.

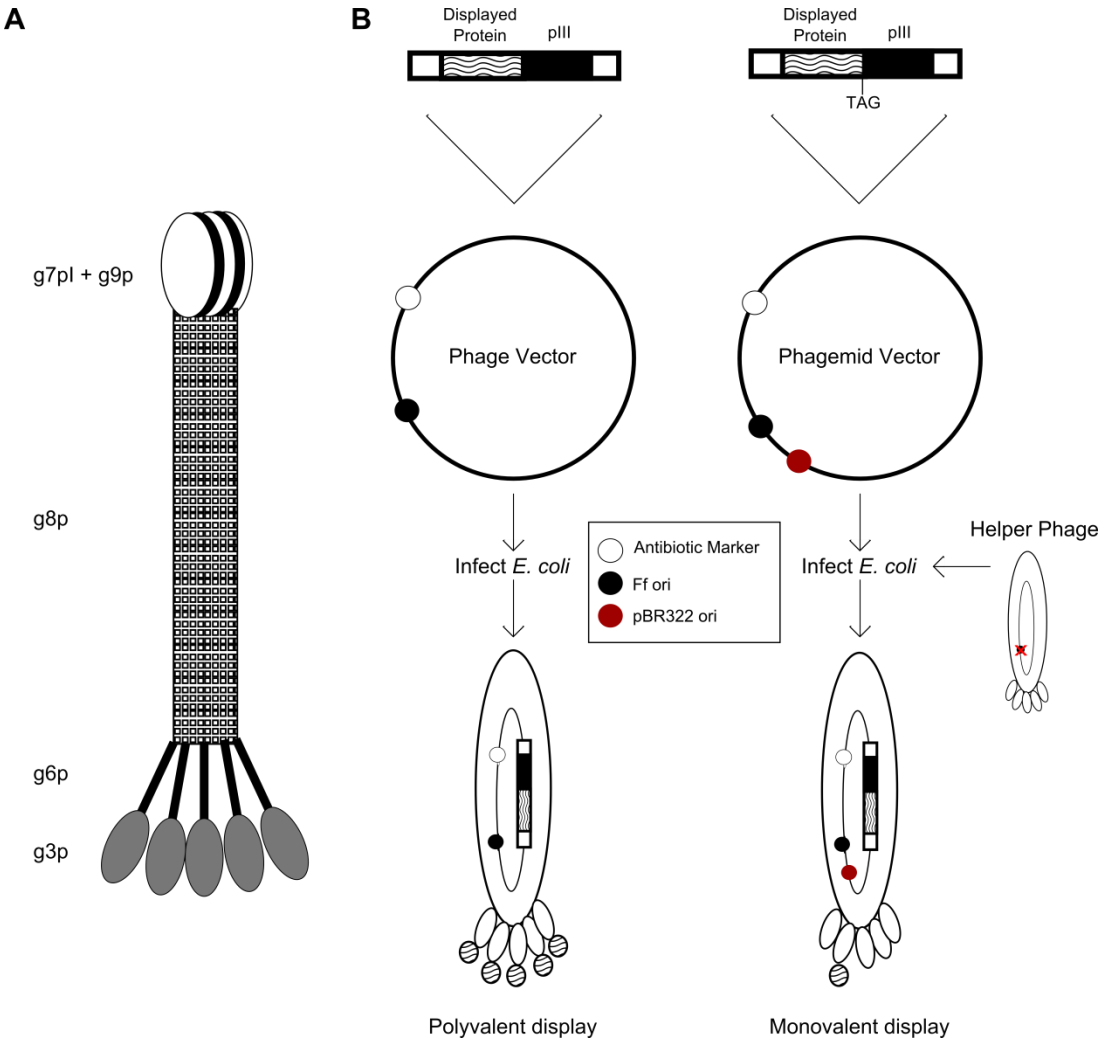


Figure 7: A) A schematic of M13 filamentous phage structure, with the most commonly used coat proteins g8p and g3p labelled. B) The two genetic approaches to generate phage displaying pIII conjugates and the resultant level of display on the phage dependent on the vector approach utilised.

There are two methods at the DNA level of displaying libraries on phage. The simplest approach is to edit the genome of the phage directly, thus all of the phage protein will carry a library conjugate. This gives high levels of display but may have negative effects on the phage structure or viability. To overcome this, a second approach has been developed termed the phagemid method. A phagemid vector is used which only encodes the fusion protein. To produce viable phage the *E. coli* need to be infected with a helper phage containing DNA encoding all of the M13 proteins, often including native forms of the protein used as a display fusion. The helper phage genome has a defunct Ff origin of replication, preventing it from being replicated and packaged in single stranded form. Therefore even though phage is produced from two distinct vectors, only the phagemid vector is packaged in the phage molecules produced. Due to the production of wild-type g3p

protein, the conjugate proteins are displayed effectively monovalently. In fact the majority of phage will only display wild-type coat protein, however these will not interfere with screening procedures.

In a typical screening round, phage is prepanned against blank wells to remove binders to the plate, reducing the occurrence of false positive hits. Prepanned phage is then incubated with a target of interest which is adhered to the plate. Phage which do not bind are then removed by washing stringently. Phage bound to the target are then eluted via either pH change to disrupt the binding interface, or reduction of a disulfide bond in the linker holding the target to the plate. Eluted phage are then used to infect to an *E. coli* culture, this culture is incubated for a short time before being plated onto LB agar and incubated overnight. Colonies will grow where the phage has infected and provided antibiotic resistance. A comparison between infection of phage panned against the target vs phage panned against a blank negative control gives indication of the hit rate and library enrichment. Bacteria can then be scraped and cultured, infecting with helper phage if necessary, to produce new phage for a new screening round. Alternatively individual colonies can be picked and grown individually to be used in phage ELISA to identify particularly strong or specific binders.

Phage-displayed libraries are usually composed of peptides or proteins composed of canonical amino acids. There are select examples of displayed libraries containing artificial amino acids using amber suppression[73-75] and methionine biosynthesis inhibition[76]. However amber suppression approaches can be limited by the translation efficiency of the amber codon – depending upon the position and number of codons being replaced truncation mutants could prove problematic. Each artificial amino acid also requires the development of its own tRNA-synthetase. The Met replacement requires amino acids which can use the Met synthetase, and may cause problems with Met being replaced in all of the phage proteins.

A more convenient approach to adding artificial functionality to phage-displayed peptide sequences is to use enzymatic or chemical modification to alter canonical amino acids. Enzymatic post translational modification (PTM) can take place in the host cell during phage production or *in vitro* after phage secretion. PTMs on phage-displayed sequences include biotinylation, glycosylation and ligation of molecules using Spf synthase or transglutamination. The main limitation of this approach with respect to library display is the requirement for recognition sites immediately around the site of modification. This limits the diversity available around the site, possibly reducing the value of the library approach. PTMs which have recognition signals distant from the modification sites may be of more interest.

Chemical modifications usually have less sequence-dependence, requiring only one or two specific residues. This makes them more appropriate for modification of a library of peptides. The toolkit of protein and peptide modification is ever-expanding as new methods are developed, however only a few have been used to modify phage-displayed libraries. While the incorporation of artificial amino acids extends this toolkit further it is possible to utilise the canonical amino acids to create diverse structures.

One of the early modifications was the use of chemical ligation to append a synthetic peptide containing artificial amino acids. In an approach termed biosynthetic phage display, the protease c inhibitor eglin was split in two.[77] Residues 8-40 were chemically synthesised including either kynurenine or norvaline at position 25. A library of residues 41-70 with randomisation at position 52 and 54 was created through either g3p or g8p fusions with an N-terminal cysteine. Native chemical ligation then fused the full-length sequence together which was subsequently tested for proper folding into the native eglin fold.

A more common approach is to take advantage of the nucleophilicity of cysteine residues via alkylation. A number of functional groups will alkylate cysteine preferentially over other residues including haloacetamides[78] and maleimides[79]. An interesting example is the creation of light-responsive ligands through the addition of haloacetamide derivatives of azobenzene, developed independently by both Heinis and Derda.[80, 81] These cyclise peptide loops across two cysteine residues, with the ligand backbone changing from trans to cis under UV irradiation, altering the peptide loop structure. Through selective screening procedures, UV light can be used to either activate or deactivate peptide ligand binding conformation.

As mentioned in Section 1.2.5, Heinis and co-workers create bicyclic peptides through the reaction of tris(bromomethyl)benzene **10** (TBMB) and three displayed cysteine residues.[14] These reactions are not without pitfalls; M13 phage in particular has important cysteine residues in the g3p protein which when reacted with alkylating agents impair phage infectivity, for example Heinis and co-workers observed a 99% drop in infectivity after reaction with TBMB.[82] This was predicted to be particularly drastic due to the tris-reactive functionality of the alkylating agent causing cross linking of g3p proteins. To overcome this a cysteine-free g3p variant has been utilised, containing fifteen stabilising mutations.[83] Taki and co-workers found that the reaction of T7 phage gp10 protein did not cause the same problems when a displayed cysteine was reacted with tetramethylrhodamine-5-iodoacteamide.[78] Derda and co-workers used a two-step process to glycosylate peptides at the N-terminus.[84] An N-terminal serine or threonine was oxidised to the aldehyde and an oxime ligation attaches the mannose to the terminal residue specifically. Biotin was also attached in this manner and used to assess the yield of coupling.

One of the greatest challenges associated with chemical modification of phage is estimating the yield and rate of the reaction. Bacteriophage are too large to monitor reactions with traditional spectroscopic techniques such as NMR or mass spectrometry. In the case of g3p fusion modifications, each phage can carry five copies of the displayed peptide, and so estimating individual peptide conversion by phage-counting techniques assumes each peptide will have reacted identically. To overcome this problem Heinis and co-workers created a construct containing a trypsin digest site allowing release of displayed peptides and separation from phage by spin filtration and monitored by mass spectrometry.[85] While this is not quantitative it is an improvement on previous approximations.

1.3.2 mRNA and Ribosomal Display

In vitro display technologies which do not require cellular production stages have an advantage in that library diversity is not limited by the cellular transformation step. Two such display methods are ribosomal and mRNA display. Ribosomal display involves the translation of RNA libraries with ribosomes, however due to the lack of any stop codon both the RNA and produced protein are not released from the ribosome complex. After the separation of active mutants, the RNA can be released from the complex, amplified *in vitro* and be used for further rounds of selection.

mRNA display is similar in that mRNA is transcribed by ribosomes *in vitro* however in a preceeding step the RNA is ligated to a DNA section containing the ribosome inhibitor puromycin. The puromycin terminates the translation process and binds the synthesised protein to the RNA that encoded it. After removal of ribosomes the RNA-protein complexes can be purified and screened. RNA from hits can be amplified *in vitro* as before to generate an enriched library.

In early iterations of this display platform synthetic mRNA libraries were combined with cellular translation machinery obtained from cellular extracts. It is now common for the components to be either purified extensively or produced recombinantly. In addition the system can be supplemented with artificial components and enzymes such as tRNA synthetases to incorporate unnatural amino acids in the displayed peptide. It is estimated that libraries can contain up to 10^{13} members, but this is only limited by DNA synthesis and the number of positions being varied.[86]

Chemical modification of such libraries has fewer issues than with phage-displayed libraries. The use of artificial amino acids allows the application of biorthogonal chemistry, for example the installation of selenalysine allows the conversion to dehydroalanine with mild

oxidative conditions that do not affect the rest of the displayed peptide.[87] This can then be used in conjunction with cysteines in the peptide to form lanthipeptide linkages.[88]

1.3.3 Synthetic peptide libraries

Both of the above screening classes link the protein to the DNA sequence that encodes it – linking phenotype to genotype. There is an alternative approach where a peptide library can be synthesised and the structure determined by mass spectrometry fragmentation of the peptide hits. When cross-linking across a peptide the mass spectrometry fragmentation can be incomplete as a single amide bond cleavage will not lead to the loss of any amino acid subunits. To get around this Pei and co-workers used a carefully-controlled protecting group strategy to create a two-tiered resin where the outer deprotected layer was cyclised while the inner layer remained protected and linear. This allowed deconvolution of the sequence of TNFα inhibitor hit **14** (Figure 6) by fragmentation of the linear peptide.[62]

1.4 Chemical Modification of Peptides

1.4.1 Single-step modifications

As described previously, the chemical modification of peptides can improve their properties and widen the diversity of structures accessible.[15, 61, 89] The majority of chemical modifications of canonical amino acids in proteins and peptides rely on the nucleophilicity of side chains – predominantly cysteines and lysines. Labelling of lysine residues with molecules such as NHS esters is usually a high-yielding modification, however there can be drawbacks with undesired cross-reactivity leading to low selectivity and labelling of residues in binding interfaces leading to decreased efficacy of labelled molecules. Cysteines have been labelled by several classes of molecules including maleimides[90], allenamides[91] and iodoacetamides[92]. Since cysteines have a relatively low abundance in native peptide sequences these are often good choices for labelling – cysteine can often be introduced via mutation to then provide a single site for modification to occur. A caveat must be noted that introducing cysteine residues can disrupt the structure of proteins if disulfide bonds are formed or are shuffled by the introduction of a new cysteine.

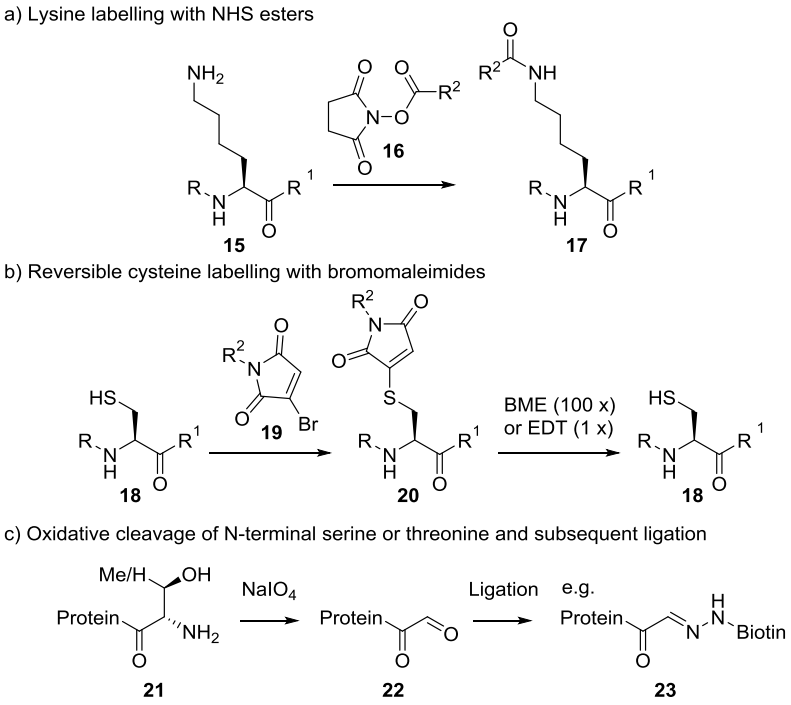


Figure 8: Common modifications to enable protein ligations a) Lysine labelling with NHS esters attaches cargoes via an amide linkage – this has low selectivity due to high lysine copy number. b) Cysteine labelling with bromomaleimide is reversible in the presence of high concentration of thiols or ethane dithiol. c) The two-step oxidation and ligation modification at protein N-terminus to selectively ligate a cargo to the protein.

Alternatively, the incorporation of non-canonical amino acids widens the scope of chemistry accessible. For example the incorporation of alkene-bearing alanine homologues has been used to create stapled peptides via metathesis as discussed in Section 1.2.3.[50, 52] There are also a wide range of dienes and dienophiles which can be incorporated and allow cycloadditions including the ubiquitous copper-catalysed Click chemistry. This bio-orthogonal chemistry is now being used in a wide variety of chemical biology applications including super-resolution microscopy[93], cell-surface glycan labelling[94] and lysine side chain decaging.[95]

1.4.2 Multi-step modifications

Multi-step modification strategies involve an initial modification of an amino acid to modulate its reactivity and then a subsequent step to complete the reaction. An example of this is the $NaIO_4$ oxidation of N-terminal serine and threonine residues to yield an aldehyde which can then react further with a wide range of nucleophiles – this oxidation step is often coupled with a reductive amination to label a protein at the N-terminus specifically (see Figure 8c).[84, 96] A second example is the oxidation and elimination of alkylated selenocysteine residues giving access to dehydroalanine.[97]

Both of the above strategies install electrophilic functionality into the protein at specific locations – electrophilic groups are rare in proteins, probably due to the likelihood of cross-reactivity with nucleophilic side-chains and molecules such as glutathione. Dehydroalanine has recently been utilised as an electrophilic handle to either cyclise a protein (in a manner equivalent to lanthipeptide linkages)[87] or to allow site-specific modification. These modifications have been to label with a dye[98], to introduce novel functionality into a specific position in a histone[99] or enzyme.[100, 101] Alternatively the dehydroalanine has been used as a chemical handle, placed in the linker region between ubiquitin dimers to investigate deubiquitinase selectivity and activity.[102] In many of these modifications, the addition of a nucleophile will give a mixture of isomers: for example the addition of a thiol will create a mixture of alkylated L- and D- cysteine derivatives. In many applications this is undesirable as the introduction of a D- amino acid can affect the structure and correct folding of a protein.

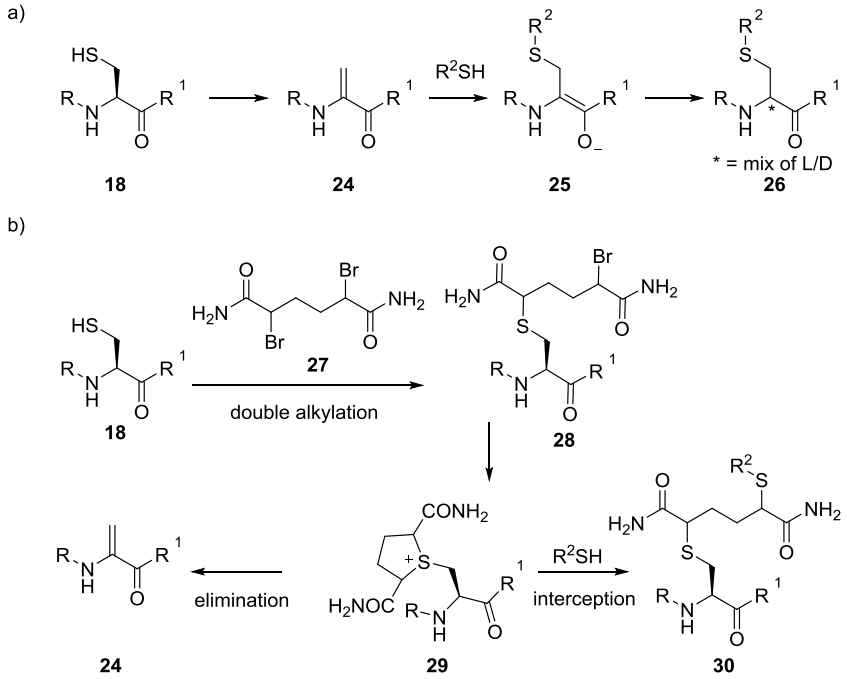


Figure 9: Dehydroalanine chemistry a) Cysteine can be converted into dehydroalanine which is subsequently reacted with thiols to generate alkylated cysteine variants with a mixture of L/D-stereochemistry. b) The double alkylation and elimination approach to convert cysteine to dehydroalanine using 2,5-dibromoadipamide pioneered by Davis and co-workers.[103] The cyclic sulfonium ion is usually eliminated to form dehydroalanine however due to its stability, the interception of the sulfonium with another thiol has been observed giving an elongated thiol alkylation.[104]

Dehydroalanine can be input by several chemical methods which were extensively reviewed by Chalker *et al.*[103] and are discussed in Section 2.1. Methods include the aforementioned oxidative elimination of alkylated selenocysteines, and elimination of activated serines (either via enzymatic or chemical methods). However the most prominent method currently used is a double alkylation to activate a cysteine residue for elimination. This method has the advantage of the mildest conditions, with reagents which are easy to prepare and handle. The double alkylation generates a sulfonium ion which can then be eliminated to yield dehydroalanine. Davis and co-workers have exemplified this approach to generate dehydroalanine-containing peptides and proteins in good yields.[103] A possible drawback of this method is the possibility of interception of the sulfonium ion by other thiol molecules – Nathani *et al.* took advantage of the long lifetime of the sulfonium to create extended thiol conjugates by adding a thiol to ring-open the sulfonium instead of eliminating.[104]

1.5 Research Outline

The preceding sections demonstrate the value of the chemical constraint of peptides to deliver structures with interesting biological activity. This body of work aimed to demonstrate the effective constraint of peptides via the reaction between a peptide containing multiple dehydroalanine residues and a small molecule core bearing multiple thiol functionalities. It was envisaged that the stereochemical scrambling which occurs upon thiol addition would generate libraries similar to those discussed in Section 1.2.5 but with increased stereochemical diversity in peptide loop structure due to the effect of D-amino acid incorporation. The incorporation of D-amino acids would also likely improve protease resistance.

To deliver this there was a requirement to investigate and optimise the following:

1. The generation of multiple dehydroalanine residues in a peptide sequence via the conversion of cysteine residues
2. The cyclisation of peptides bearing multiple dehydroalanine residues
3. The synthetic accessibility and tractability of multithiol cores

An ultimate aim of this project was to apply the two-step modification to phage-displayed peptide sequences. This would deliver access to a large library of peptides on a platform with easy amenability to screening against a diverse set of target molecules. The following aspects of this project are presented in this thesis:

1. The generation of multiple dehydroalanine residues in a peptide sequence using a new reagent better-suited to this application than previous double-alkylation elimination reagents (Chapter 2)
2. Exemplification of the effect of stereochemical scrambling on a bicyclic peptide structure using the model system PK15 kallikrein inhibitor (Chapter 2)
3. The attempted synthesis of novel cores containing trithiol functionality in addition to the development of a core which reacts directly with cysteine residues (Chapter 3)
4. An investigation into applying this two-step chemical modification to a peptide sequence displayed on phage (Chapter 4)

In addition an investigation into the applicability of non-antibody binding proteins to bind small molecules was enabled by chemical modification of the antifungal drug posaconazole. Initial biotinylation allowed screening of an Adhiron library for binders and hits were characterised via a second derivative of posaconazole which allowed fluorescence anisotropy biophysical analysis (Chapter 5).

Chapter 2: The chemical modification of peptides to incorporate multiple dehydroalanines and subsequent cyclisation to form bicyclic structures[i]

In this chapter the use of 2,5-dibromoadipamide to create multiple dehydroalanine residues is discussed. The associated problems then directed research into alternative reagents and the adoption of methyl 2,5-dibromovalerate as an improved reagent for the creation of multiple dehydroalanine residues. To exemplify this methodology a comparison with 2,5-dibromoadipamide was undertaken and the new reagent utilised in the conversion of a model system kallikrein inhibitor sequence PK15. The second step of the constraint strategy involved the cyclisation of a peptide containing multiple Dha with a small molecule core bearing multiple thiol functionality. This was exemplified through the synthesis of 1,3,5-tris(mercaptopmethyl)benzene and the subsequent cyclisation of the PK15-derived sequence to yield stereoisomers of the original PK15 bicyclic peptide which is formed through alkylation of the cysteine-containing peptide with 1,3,5-tris(bromomethyl)benzene. The mixture of stereoisomers was then tested in a human plasma kallikrein inhibition assay. Individual stereoisomers were resynthesised by solid phase peptide synthesis[ii] and cyclised with tris(bromomethyl)benzene to investigate the effect of stereochemical scrambling on inhibitor structure.

[i] The work presented in this chapter formed the basis of the publication "Chemical generation and modification of peptides containing multiple dehydroalanines", Morrison *et al.*, **Chem. Commun.**, 2015, **51**, 13470-13473

[ii] Individual stereoisomers were synthesised and cyclised by Patrick J. Foley (MChem student) and tested in a kallikrein inhibition assay by both PJF and PMM

2.1 Reported methods to generate dehydroalanine

As set out in Section 1.5, it was aimed to utilise a two-step chemical strategy to constrain peptides in which the first step involved the incorporation of multiple dehydroalanine (Dha) residues (Figure 10). In a second step a polythiol core was to be added to cyclise the peptide, scrambling the stereochemistry at the alpha carbons of the newly-formed cysteine derivatives.

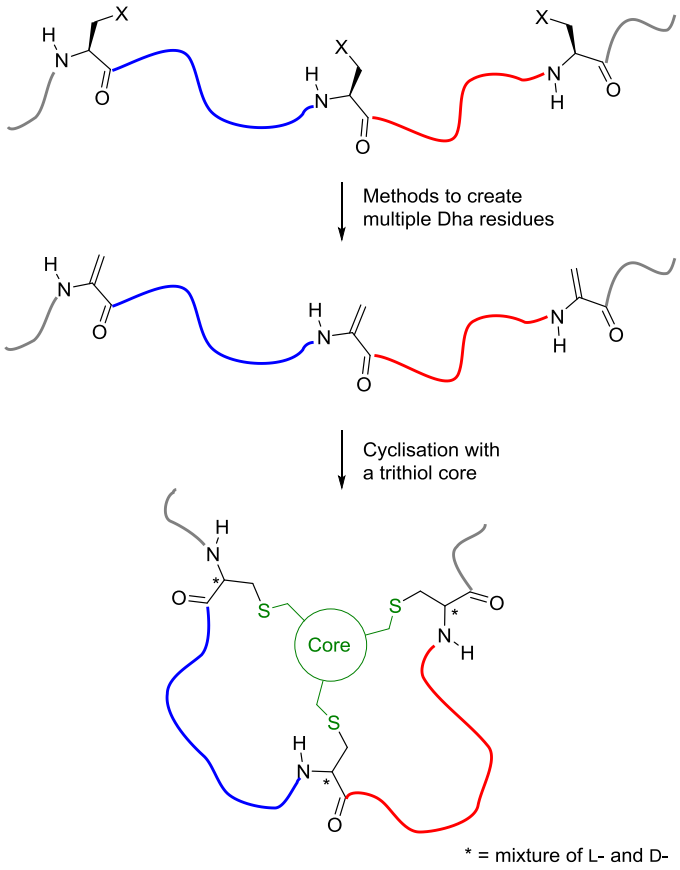


Figure 10: The two-step chemical strategy to constrain peptides. After generating a peptide containing three dehydroalanine versions it was aimed to cyclise the peptide via addition of a core bearing three thiol moieties (green). This would constrain the two linear sequences (blue and red) into constrained binding loops. The stereochemistry at the cysteine residues will be scrambled by the thiol addition – leading to a diverse subset of bicyclic peptides being generated from a single linear peptide.

Due to the highly reactive nature of dehydroalanine, the residue is best incorporated in a masked residue and subsequently converted to form Dha. Methods to generate Dha have been reported starting from serine, cysteine and selenocysteine derivatives. Synthetic methods which mimic the lantibiotic enzymatic dehydration of serine residues involve activation of the serine toward a subsequent elimination to form dehydroalanine. This activation typically involves tosylation[105, 106], or dichloroacetyl chloride ester formation.[107] These esters eliminate in the presence of sodium hydroxide, triethylamine or DBU bases,

often requiring elevated temperature to ensure complete reaction. In a similar vein, carbodiimide reagents such as EDC can be used to activate the serine and in the presence of catalytic copper this will eliminate[108], but it has also been shown that base elimination can be required to ensure complete reaction.[109] Lanthipeptide enzymatic elimination of serine residues (Section 1.2.3) has also been replicated in vitro by Van der Donk and co-workers.[110]

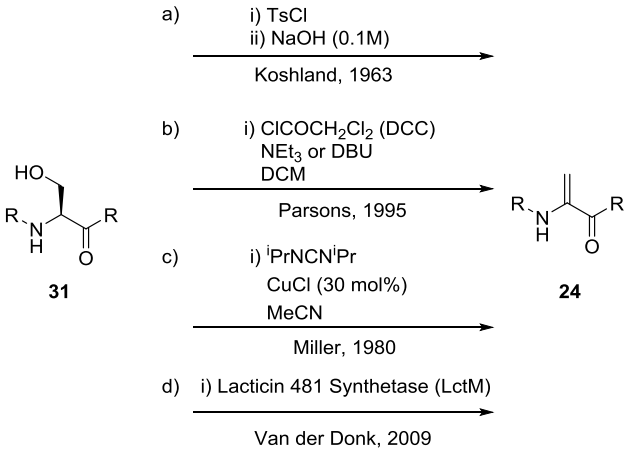


Figure 11: Methods of converting serine to dehydroalanine. Activation by a) tosylation and b) dichloroacetate ester formation both require base elimination. c) Carbodiimide activation with copper catalysed elimination first demonstrated by Miller with diisopropylcarbodiimide (DiPCD). d) Biosynthetic transformation using LctM, requiring adherence to a leader sequence peptide.

The elimination of the sulfur of cysteine residues has also been achieved via double S-alkylations[111] and also by the oxidation and elimination of S-methyl cysteine using periodate with subsequent elimination of the sulfinic acid.[112, 113] The oxidative elimination methodology has also been applied successfully to Se-alkylated selenocysteine amino acids. Selenoxide elimination can be achieved without the need for elevated temperatures.[87, 97, 114, 115] This method does however require the incorporation of artificial amino acids – methods such as amber suppression are efficient enough for high yielding incorporation at a single position, but incorporation at multiple positions is less efficient. mRNA display replacing lysine with selenalysine has been used to generate lantipeptide-like cyclic peptides (Figure 12b) however to date these examples contain only one Dha residue. Synthetic methods have produced multiple dehydroalanines but these are through purely synthetic efforts rather than methods which could be applied to peptides or proteins expressed using biosynthetic apparatus.[109, 113, 116]

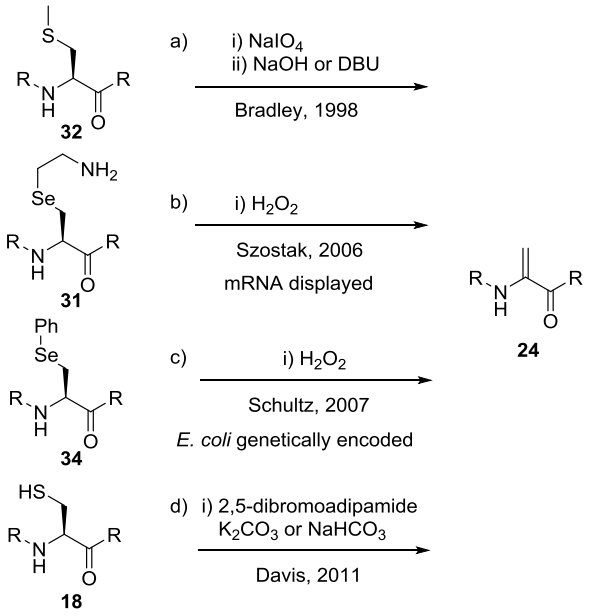


Figure 12: Alternative routes to dehydroalanine. a) The incorporation of S-methyl cysteine into a SPPS-built peptide allowed oxidation and subsequent elimination. b) mRNA display of a peptide containing selenalysine was followed by hydrogen peroxide oxidation to yield dehydroalanine. c) The genetic incorporation of phenylselenocysteine enabled dehydroalanine formation via Se oxidation. d) Cysteine can be converted to dehydroalanine by double-alkylation elimination with 2,5-dibromoadipamide.

The most accessible chemical approach to install dehydroalanine functionality is the double-alkylation and elimination methodology established by Davis and co-workers which converts cysteine residues to dehydroalanine via a cyclic sulfonium intermediate.[103] The reagent they developed for this transformation, 2,5-dibromoadipamide **27**, has since been used by many researchers to perform this transformation.[98, 100, 101] The methodology benefits from mild reaction conditions and the low occurrence of free cysteines in protein sequences. The double alkylation-elimination mechanism involves an initial alkylation at cysteine by **27** followed by a second alkylation to form a positively charged sulfonium species **29** which undergoes elimination to form dehydroalanine **24**. When using dibromoadipamide the cyclic sulfonium intermediate is long-lived and often requires heating at 37 °C to eliminate. This has been utilised by Nathani *et al.* to form elongated alkylated cysteines via the interception of the sulfonium with another thiol.[104]

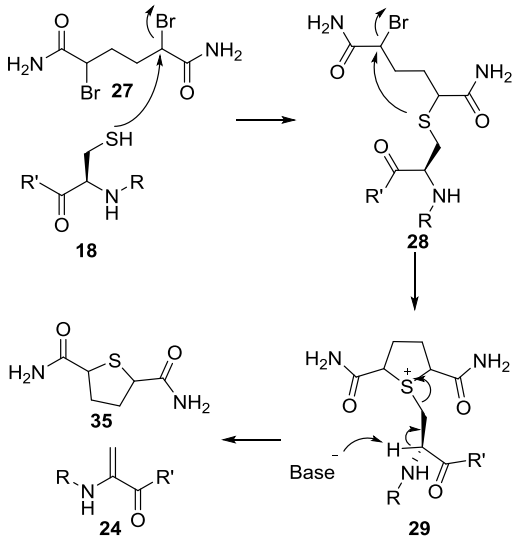


Figure 13: The double-alkylation elimination mechanism, utilising 2,5-dibromoadipamide **27** to convert cysteine to dehydroalanine.

2.2 Chemical generation of multiple dehydroalanine residues

2.2.1 Aims

An outline of the research objectives are summarised in Figure 14. It was aimed to generate multiple dehydroalanine residues in a peptide via the conversion of cysteine to dehydroalanine. Due to the literature precedent of 2,5-dibromoadipamide use this was the intended method. It was then aimed to synthesise the trithiol tris(mercaptomethyl)benzene TMMB **39** to create analogous structures to the bicyclic peptides formed by the cyclisation with TBMB **10**.

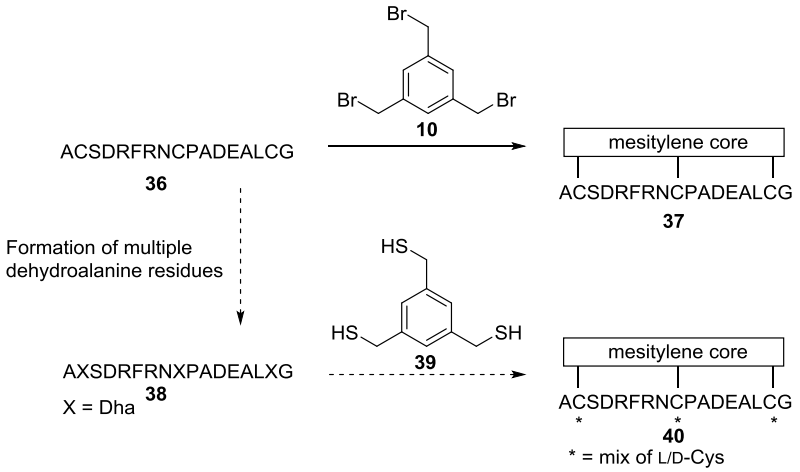


Figure 14: The proposed work undertaken in this project. A two-step strategy in which a peptide containing multiple dehydroalanines is initially formed and then reacted with a trithiol core to create peptide stereoisomers of the bicyclic peptides formed from direct reaction of cysteine-containing peptide with TBMB. PK15 sequence is used as an exemplar.

2.2.2 Application of dibromoadipamide for multiple dehydroalanine formation

The initial objective was to demonstrate the conversion of multiple cysteine residues to dehydroalanine using the double-alkylation elimination mechanism with 2,5-dibromoadipamide **27**. **27** was synthesised from adipic acid following literature procedure in 46% yield.[103] The product was isolated as a mixture of the *meso* and *d/l* diastereomers. Initial investigations confirmed the conversion of peptide H-Ala-Tyr-Cys-Asp-Gly-OH to H-Ala-Tyr-Dha-Asp-Gly-OH. Addition of alkyl thiols then created alkylated cysteine derivatives as expected. In contrast to literature reports[101] it was observed by LCMS that the addition of thiophenol to dehydroalanine was reversible.

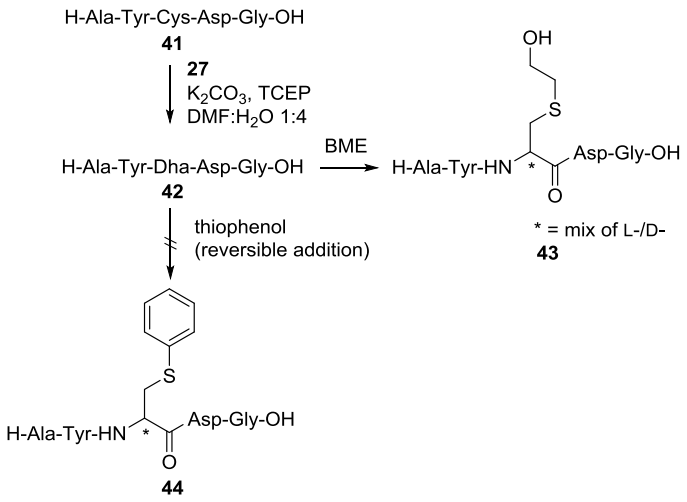


Figure 15: The conversion of cysteine to dehydroalanine using **27** and the results of subsequent thiol attack to form the mixture of alkylated cysteine derivatives **43**. Addition of thiophenol gave a small level of product **44** however this reaction appeared to be reversible.

The application of **27** to convert multiple cysteine residues to dehydroalanine residues proved more problematic – although the desired product was produced, there was parallel production of a cyclised by-product in which two cysteine residues had reacted with one equivalent of **27**. This problem was exacerbated with increasing numbers of cysteines in the peptide sequence. For example treatment of the peptide GCSDYVCGTHACA **45** with 30 equivalents of dibromoadipamide led to a 2 to 1 ratio of product **46**:stapled by-products **47** (Figure 16). This was assessed by extraction of the relevant ions for both the 1^+ and 2^+ charge states of the relevant molecules from the LCMS trace. Integration of the extracted ion chromatographs gave an approximation of the concentration of product and by-products. This assumes that the ionisations of the related peptide species are approximately comparable.

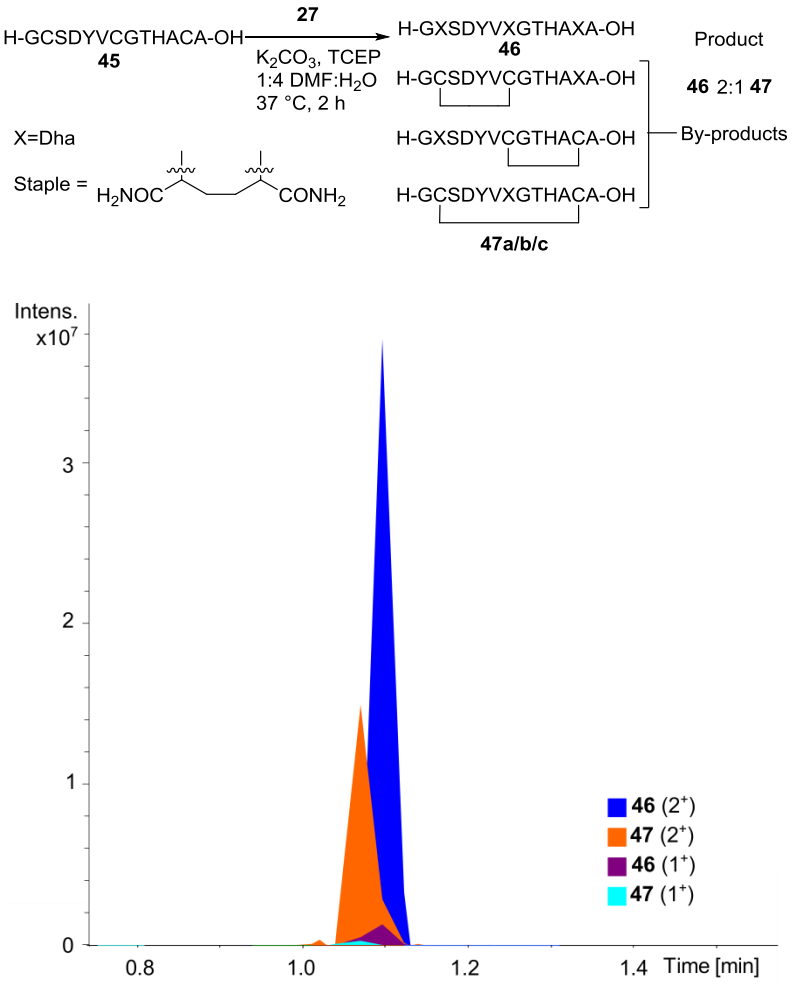


Figure 16: LCMS observation of the stapling issues observed when peptide **45** is reacted with **27**. Extracted ion chromatograms are depicted for ions corresponding to product **46** and by-products **47.** Despite use of excess **27** there was 32% formation of stapled by-products **47a/b/c** containing one dehydroalanine and two stapled cysteines.

The most likely mechanism of this by-product formation is the second bromide displacement being hijacked by the second cysteine (Figure 17). An alternative possible mechanism is that a neighbouring cysteine could intercept the sulfonium intermediate, especially given the long lifetime of this intermediate when using **27**. This would be similar to the thiol elongation observed by Nathani *et al.*[104]

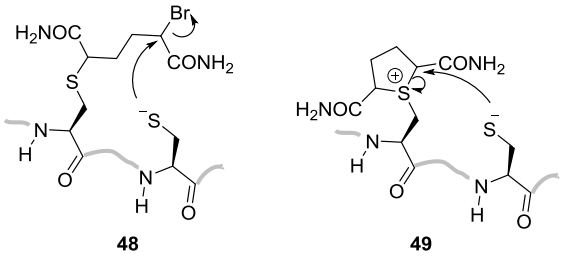


Figure 17: The two possible mechanisms of undesired cyclised peptide formation when using a double-alkylation elimination reagent (exemplified with 2,5-dibromoadipamide) to create multiple dehydroalanine residues.

2.2.3 Development of alternative reagents for dehydroalanine conversion

Due to the inconsistent and unfavourable stapled by-product formation it was decided to investigate alternative reagents for dehydroalanine production. The rationale was to decrease the likelihood of a second alkylation by a neighbouring cysteine and this was approached from two angles – the first of these was the use of steric restriction to prevent the intramolecular stapling.

The use of increased steric control with a symmetrical reagent was investigated through the synthesis of benzyl bromide derivatives. It was thought that the flexiblility of the adipamide may be allowing the second bromide to come into close proximity with the second cysteine residue. It was reasoned that the introduction of a rigid backbone may keep the second bromide in closer proximity to the initially alkylated cysteine. Chalker *et al.* showed the reagent 1,2-bis(bromomethyl)benzene was able to convert cysteine to dehydroalanine,[103] albeit in poor yield, thus this was used as a starting point. A combination of α-bromo-α-acyl benzylic positions would increase reactivity of the bromides and reduce the backbone conformational freedom. The amide and ester variants shown in Figure 18 were the initial targets to be synthesised.

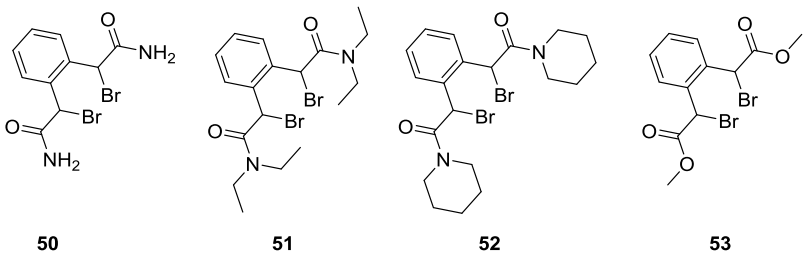


Figure 18: Targets of synthetic efforts to make double-alkylation elimination reagents selective for a single cysteine via steric constraint of the second bromide.

The synthetic route to these compounds were similar to the synthesis of **27** (Figure 19). The formation of the α,α'-dibromo-1,2-phenylenediacetoyl chloride intermediate **56** proceeded as

expected, however the reaction of this with amines proved more problematic. Reaction with ammonia and diethyl amine created multiple by-products, most likely due to attack on the very reactive bromides thus no single products were isolated. When piperidine was used the product **52** was initially formed and purified by column chromatography, however after being left under vacuum for 72 hours the product had degraded, possible degradation products are shown with the LCMS trace in Figure 20. In a final attempt amide derivatives were abandoned and the methyl ester **53** synthesised. Although it was possible to isolate and characterise **53**; upon addition to basic aqueous media, the solution immediately turned bright red and LCMS showed that **53** had degraded. Given this instability, it is probable that the benzylic positions are simply too reactive and will not be selective for cysteine residues thus this series of compounds was abandoned.

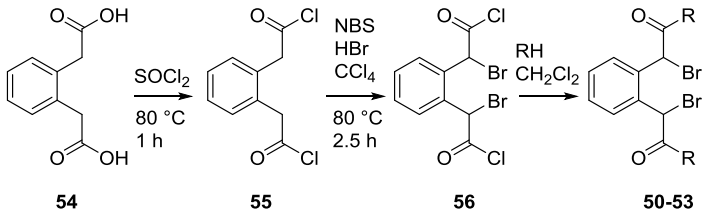


Figure 19: The one-pot synthetic route to the sterically hindered targets **50** to **53**.

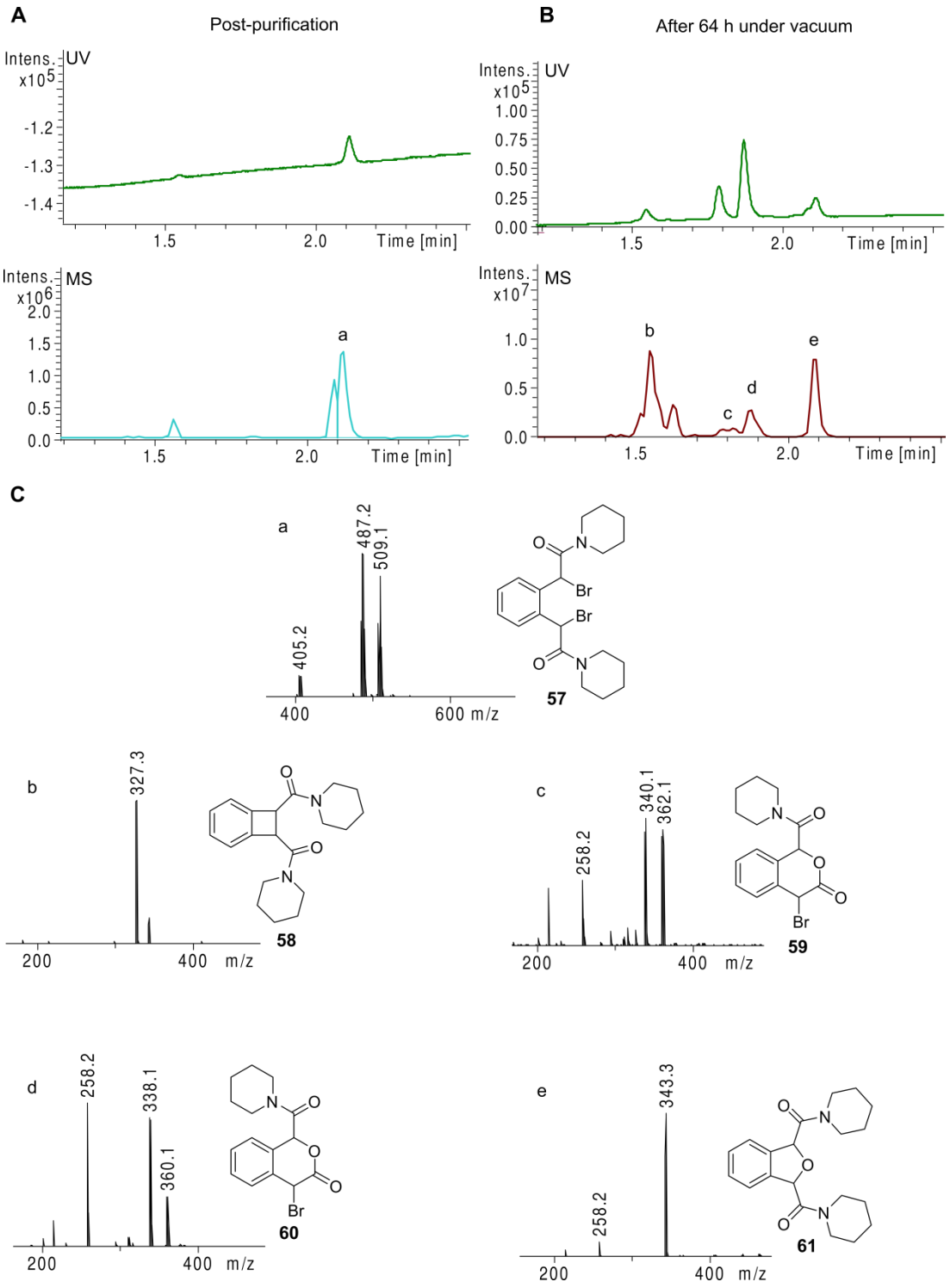


Figure 20: The breakdown of compound **52**. After initially appearing pure after column chromatography (A) after leaving under vacuum at room temperature for 64 hours the compound degraded to several by-products (B). The mass spec of these peaks gives indication of possible degradation products b-e which are shown beside the relevant MS peaks. The MS of c and d are very similar – suggesting possible diastereomers (MS signals labelled are different due to Br isotope pattern observed in both compounds). Br isotope patterns are in accordance with that expected for each compound.

2.2.4 The development of methyl 2,5-dibromovalerate for dehydroalanine generation

In an alternative approach, the reactivity of the second alkylation step could be fine-tuned to favour intermolecular alkylation with a second equivalent of reagent over intramolecular cyclisation with a single equivalent. The use of an unsymmetrical reagent containing one α-bromo carbonyl and a less reactive alkyl bromide was investigated through the synthesis of methyl 2,5-dibromovalerate **63**. It was predicted that due to the lower reactivity at the 5-position, neighbouring cysteine residues would favour alkylation with a further equivalent of reagent rather than form the cyclised by-product. Figure 21 demonstrates this strategy wherein we aimed to slow steps A and B_1 in order to favour step C.

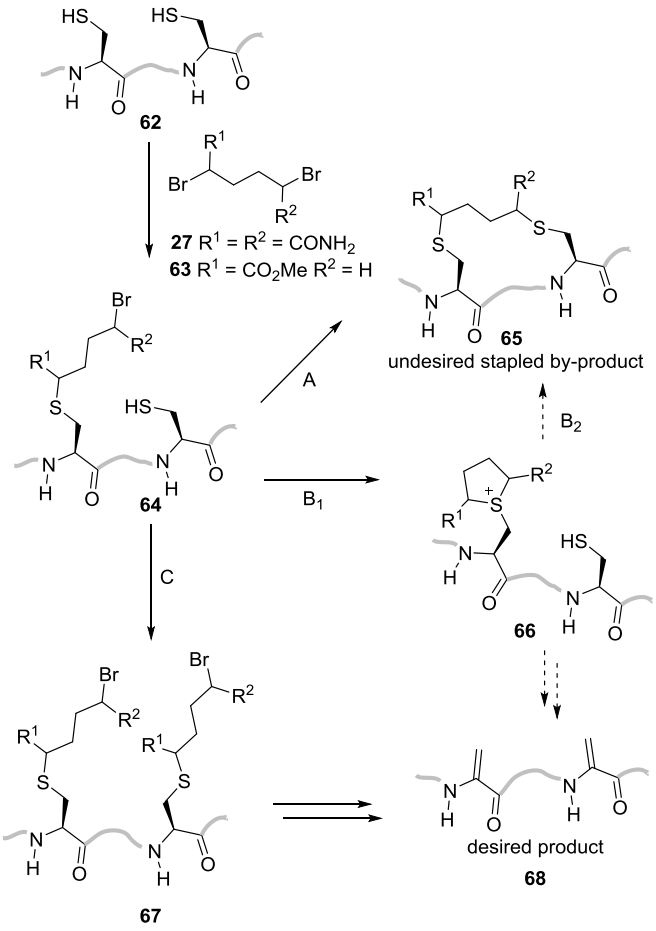


Figure 21: The rationale for the synthesis of methyl 2,5-dibromovalerate **63**. To prevent cyclisation of the peptide via routes A or B, it is necessary to slow these steps such that route C becomes favoured and the peptide reacts with a second equivalent of **63** rather than undergoing intramolecular cyclisation. The reduction of reactivity in the change from α-bromo carbonyl in 2,5-dibromoadipamide **27** to a simple alkyl bromide in **63** was designed to effect this bias.

Methyl 2,5-dibromovalerate **63** was deemed an appropriate target to validate the hypothesis that reducing reactivity at the second bromide position would moderate stapled peptide formation. An ester was chosen over the primary amide to avoid any possible lactam

formation through cyclisation at the 5 position. **63** was synthesised in good 77% yield from δ-valerolactone **69** adapting a literature preparation.[117] The bromination utilised substoichiometric PBr_3 although elementary red phosphorus has been utilised previously as this forms PBr_3 via reaction with Br_2.[118] This PBr_3 mediated reaction is similar to the Hell Volhard Zelinski bromination of carboxylic acids, forming a bromoacyl bromide which was then reacted with methanol to form methyl 2,5-dibromovalerate **63**. **63** proved to be very stable, even at room temperature and open to the air.

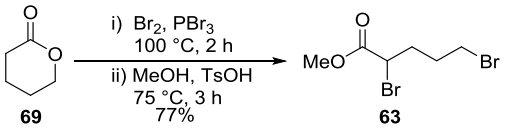


Figure 22: The synthesis of methyl 2,5-dibromovalerate **63** from δ-valerolactone **69** via PBr_3-mediated bromination.

Initial tests with peptide **70** proved that **63** was adept at converting cysteine to dehydroalanine through the same double alkylation-elimination mechanism as **27**. Interestingly the major intermediate observed was that of the monoalkylated cysteine **72**; in contrast to the use of **27** there was only extremely low observation of the sulfonium intermediate.

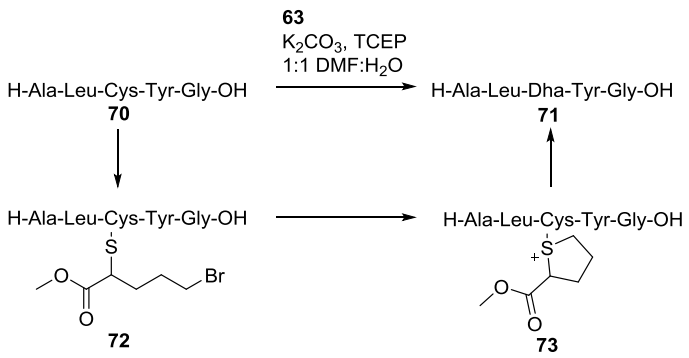


Figure 23: Conversion of peptide **70** to dehydroalanine-containing peptide **71** proved **63** was an alternative double-alkylation elimination reagent.

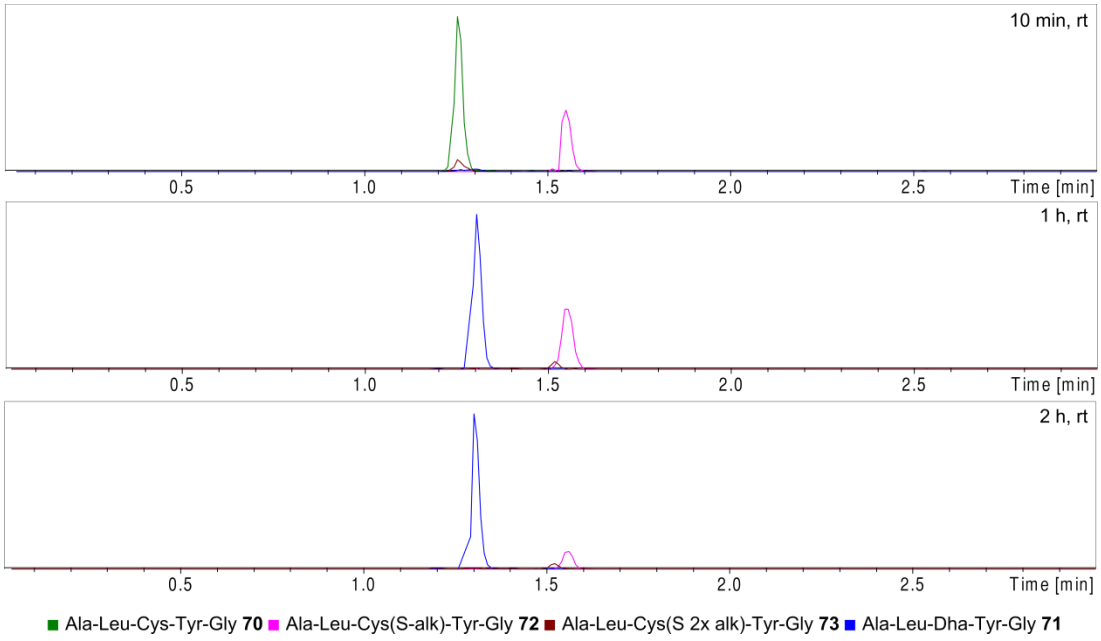


Figure 24: LCMS extracted ion chromatograms in the reaction of **70** with **63** to form **71**. Starting peptide **70** (green) is completely consumed and converted to the dehydroalanine-containing analogue **71** (blue) via the singly-alkylated cysteine intermediate **72** (pink). There is little evidence of any stable sulfonium intermediate **73** in contrast to the equivalent sulfonium when **27** is used to perform the transformation.

It was then verified that **63** was superior to **27** in the conversion of multiple cysteine residues to dehydroalanine. To compare the two reagents in more detail a number of solvent and base mixtures were examined utilising an LCMS assay (Figure 25 and Figure 26). The peptide **74** was converted to its dehydroalanine-containing analogue **75** using varying equivalents of either reagent **27** or **63**. The intensities of ions corresponding to cyclised by-products **76**/**77** and desired product **75** were extracted and the peak integrals compared to give an estimate of the product:by-product ratio.

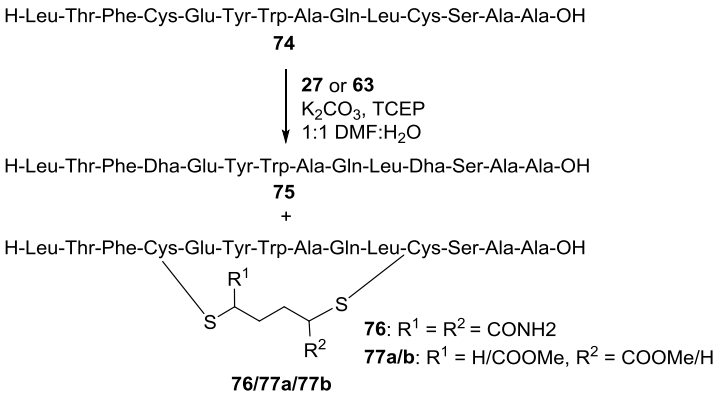


Figure 25: The reaction used to compare reagents dibromoadipamide **27** and methyl 2,5-dibromovalerate **63** in the conversion of multiple cysteine residues to dehydroalanine.

Conditions were selected based upon previous dehydroalanine forming conditions (K_2CO_3 and DMF), or upon conditions thought to be applicable to modification of residues displayed on phage (NH_4HCO_3 and acetonitrile – used by Heinis *et al.* in reaction with TBMB[14]). DMSO was also tested as it was thought the percentage of DMSO could be reduced to levels suitable for phage modification. **63** proved superior to **27** in all conditions tested. Although at low equivalents of **63** there was observation of the cyclised by-product in many solvent conditions, this was much lower than when using **27**. Of particular interest was that in 1:1 H_2O:DMSO there was no cyclisation observed using **63**, a feature not shared by **27**. Under all reaction conditions using **27** there remained a cyclised fraction of around 10% despite using large excess of dibromo reagent **27**. The equivalent residual by-product was not observed using **63**. A serendipitous advantage was also discovered in that at the concentrations used **63** proved much more soluble than **27**. In particular the solubility in acetonitrile proved promising as this solvent was used in the chemical modification of phage by Heinis *et al.*[14] The use of ammonium bicarbonate as the base appeared to produce less stapled product when **63** or **27** was used in DMF, however this trend was reversed when **27** was used in DMSO. Since the use of DMSO and K_2CO_3 with **63** led to extremely low levels of stapled by-product these conditions were used in subsequent experiments.

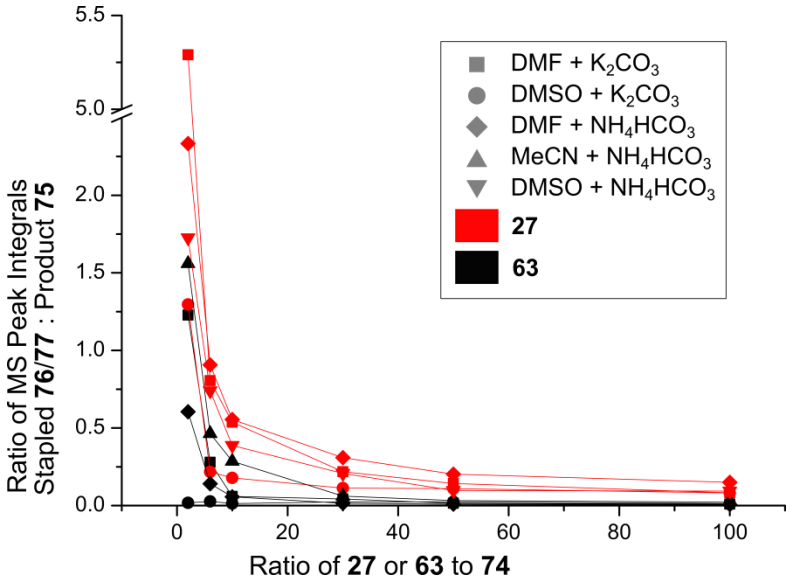


Figure 26: LCMS optimisation of dehydroalanine conversion in peptide **74** using **27** and **63. 63** proved to be superior to **27** in all comparable conditions and was additionally soluble in acetonitrile, allowing examination of the reaction by LCMS without filtration. Solvents were 1:1 mixture of H_2O and the organic solvent stated in the key. For each reaction condition, the ratio of stapled to dehydroalanine-containing peptide was determined after reaction for 3 hours at 37 °C; this corresponds to complete reaction of the peptide for all except the lowest ratios of **27** and **63** to peptide **74** in which small quantities of residual unmodified peptide were present.

2.3 Exemplification of the use of methyl 2,5-dibromovalerate to incorporate multiple Dha residues using the PK15 bicyclic peptide as a model system

2.3.1 One-pot synthesis of stereochemically-scrambled PK15 structures

In order to exemplify the use of **63** to incorporate multiple dehydroalanine residues in a peptide the human plasma kallikrein (HPK) inhibitor PK15 **37** developed by Heinis *et al*. was used as a model system.[14] HPK is a serine protease which forms part of the clotting cascade. HPK makes a good model system as there is a straight-forward *in vitro* assay and since it is found in the blood any *in vivo* assays are not complicated by cell penetration properties of the inhibitors being tested. The bicyclic peptide inhibitor PK15 was identified from a phage display library which had been cyclised using TBMB **10**. Incorporating the dehydroalanine formation strategy it was envisaged that eight stereoisomers of PK15 could be synthesised in one pot in two-steps. The first step would convert cysteine residues to dehydroalanine and subsequently a polythiol would cyclise the peptide via reaction with these dehydroalanine residues, scrambling the stereochemistry at these positions to create mixture **40**.

Starting from peptide **36**, **63** was used to convert the three cysteine residues to dehydroalanine forming the peptide **38** in 35% yield which was deemed good given the low purity of starting material (approx. 65%) and HPLC purification.

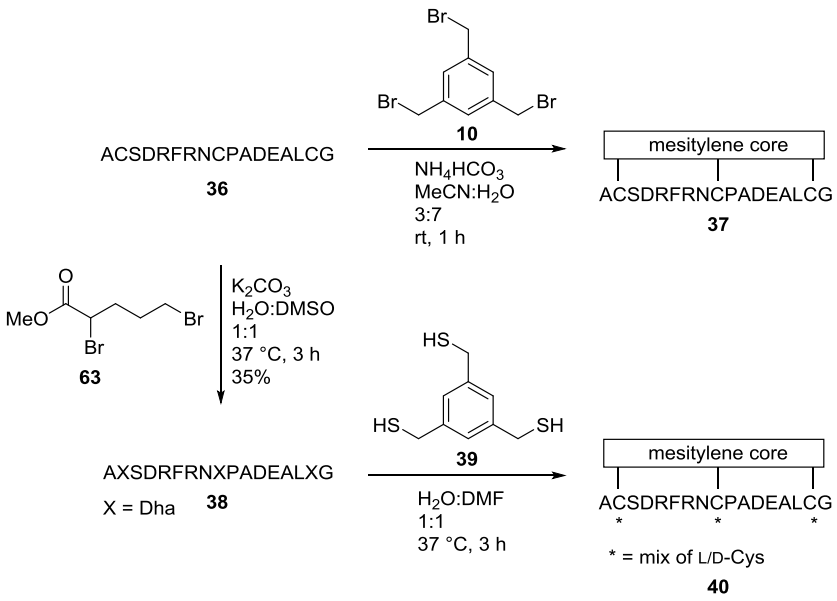


Figure 27: The PK15 kallikrein inhibitor model system used in this work. PK15 is synthesised by alkylation of three cysteine residues with TBMB **10**. The two step strategy would convert the cysteines to dehydroalanines and subsequent cyclisation with trithiol TMMB **39** would not only reproduce the original PK15 structure **37** but additionally form seven stereoisomers **40** with the stereochemistry at the cysteine residues being a mixture of L/D.

Cyclisation of peptide **38** was achieved using 1,3,5-tris(mercaptomethyl)benzene **39** to replicate the mesitylene core of PK15 **37**. **39** was synthesised from TBMB **10** via thiourea alkylation and hydrolysis (Figure 28).

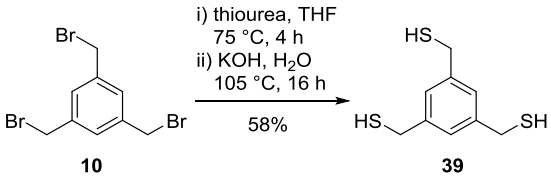


Figure 28: The synthesis of tris(mercaptomethyl)benzene from tris(bromomethyl)benzene via thiourea alkylation and hydrolysis.

The reaction of **39** with dehydroalanine-containing peptide **38** to form mixture **40** proved straight-forward. A caveat of the addition of a polythiol to dehydroalanine is that the product of single addition has the same mass as that of the products of multiple additions of the same core. Therefore although the initial addition of thiol was complete within 5 minutes, the reaction time was extended to ensure full cyclisation occurred. Purification of the bicyclic peptide ligands by mass-directed HPLC observed several closely-spaced peaks of mass corresponding to the desired peptides **40**, these were collected together.

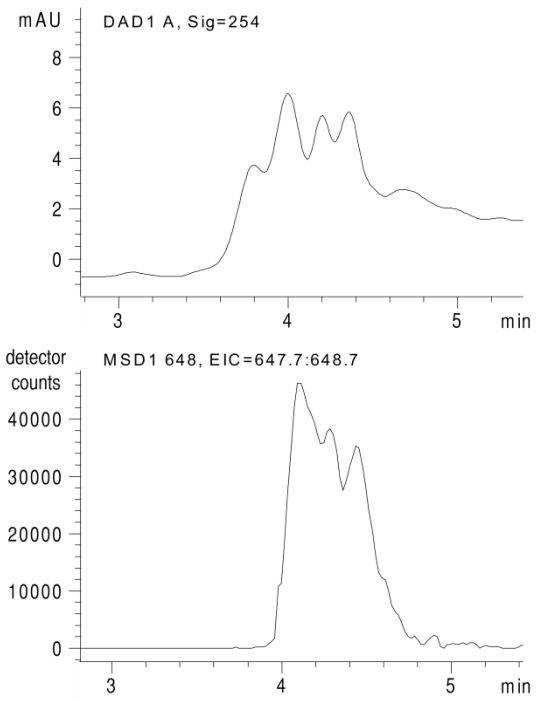


Figure 29: HPLC purification trace of PK15 stereoisomer mix **40**. The top trace shows the UV absorption trace at 254 nm while the bottom trace is an extracted ion trace for the mass corresponding to PK15 peptides **37**/**40**. Multiple peaks indicate the presence of multiple stereoisomers.

To probe for full cyclisation the purified product was tested with excesses of both N-methyl maleimide and β-mercaptoethanol as detailed in Table 2. Although an observed single addition of N-methyl maleimide was initially concerning, repeating this test using a sample of PK15 **37** independently synthesised through cyclisation of peptide **36** with TBMB **10** the bicyclic peptide **37** showed identical behaviour when reacted with N-methyl maleimide. Thus it can be assumed that the maleimide is reacting with the free N-terminal amino group rather than with any free cysteine. No addition of BME was observed, indicating full addition to the dehydroalanine residues.

Table 2: Chemical diagnostic tests to confirm full cyclisation of peptide **38** with **39**

Reactant	**Observed additions to**		**Conclusion**
	PK15 mix **40**	PK15 **37**	
N-methylmaleimide	1	1	N-terminal addition, no free thiol
β-mercaptoethanol	0	n/a	No free Dha

2.3.2 Testing of stereoisomers in kallikrein inhibition assay

The mixture of stereoisomers **40** was tested in a human plasma kallikrein inhibition assay (Figure 30). The bicyclic peptides were incubated with human plasma kallikrein for 30 minutes before addition of the substrate Z-Phe-Arg-AMC **78**. The release of fluorescent aminomethylcoumarin **80** was observed to estimate inhibition levels. The mixture of stereoisomers was tested as both crude reaction mixture (estimating the concentration of ligand by the initial starting material concentration) and as an HPLC-purified mixture.

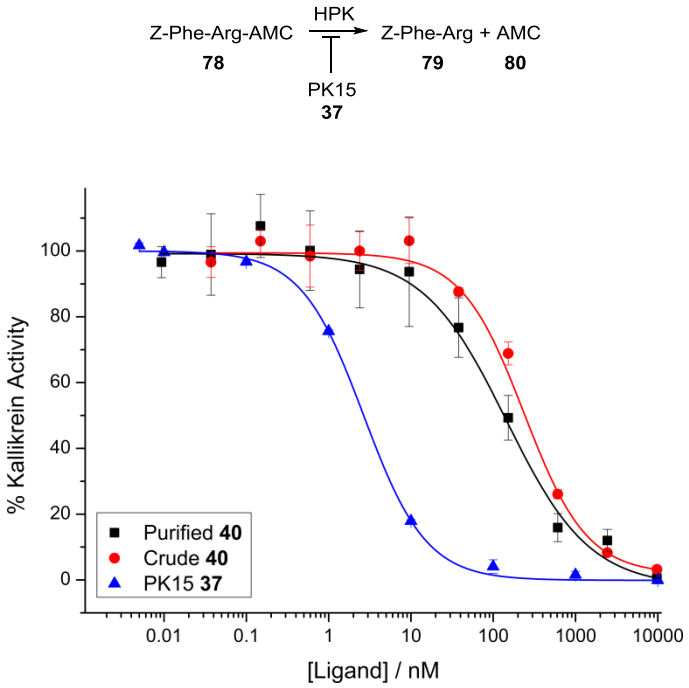


Figure 30: The inhibition of human plasma kallikrein. HPK cleaves substrate **78** to release free aminomethylcoumarin **80** (AMC) which can be observed by fluorescence (excitation 360 nm, emission 440 nm). The mixture of stereoisomers **40** (both as reaction crude and HPLC purified) was compared to that of the literature (all L-Cys) PK15 **37**.

The observed IC_{50} values for both crude (238 ± 21 nM) and purified (144 ± 31 nM) were encouraging although intriguing. The scrambling of the stereochemistry has clearly altered the bicyclic peptide loop structure leading to reduced inhibition of the HPK. However if a stoichiometric mixture of stereoisomers was produced it may be expected that the 2 nM activity of the LLL PK15 ligand[iii] should dominate and the IC_{50} of the mixture would be one eighth of this, i.e. 16 nM. Therefore the lower inhibition value suggests that a scalemic mixture was produced.

[iii] The nomenclature of the PK15 stereoisomers used in this work denotes the stereochemistry at each of the cysteine residues at positions 2, 9 and 16 in turn.

The nature of D- or L- cysteine formation upon addition of the thiol to dehydroalanine is determined by the reprotonation of the enolate intermediate **25 (**Figure 31). It is not surprising that this reprotonation can be biased thus producing non-stoichiometric mixtures of enantiomers. The local peptide sequence probably influences which products are more energetically favoured. To support this, it has been suggested that lanthipeptide stereochemistry is also encoded by sequence rather than by the cyclase enzyme.[43] This phenomenon is observed in lanthipeptide formation where cyclase substrates maintain their stereochemical bias of cysteine cross-linking when alternative cyclase enzymes are used.[43] The cyclisation with a trithiol may create bias in that there will likely be subtle differences in reactivity between the three dehydroalanine residues. The conformation of one thiol addition is then likely to affect the stereochemistry of subsequent additions – not through control of the thiol attack but via backbone strain and possibly steric hindrance of the reprotonation event.

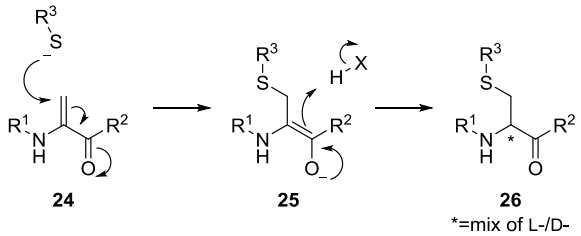


Figure 31: The attack on dehydroalanine by a thiol and subsequent reprotonation of the enolate intermediate. It is the second step which determines the stereochemistry at the alkylated cysteine derivative.

2.4 Further investigations into the effects of stereochemical scrambling on PK15

2.4.1 Synthesis and testing of individual PK15 stereoisomers

To further investigate the effect of stereochemical scrambling the eight individual stereoisomers were synthesised by solid phase peptide synthesis (SPPS).[iv] After the peptides were cleaved from the resin they were cyclised with TBMB using an established literature protocol.[14] The stereoisomers were then purified by HPLC and tested in the same HPK inhibition assay as the mixture (Section 2.3.2).

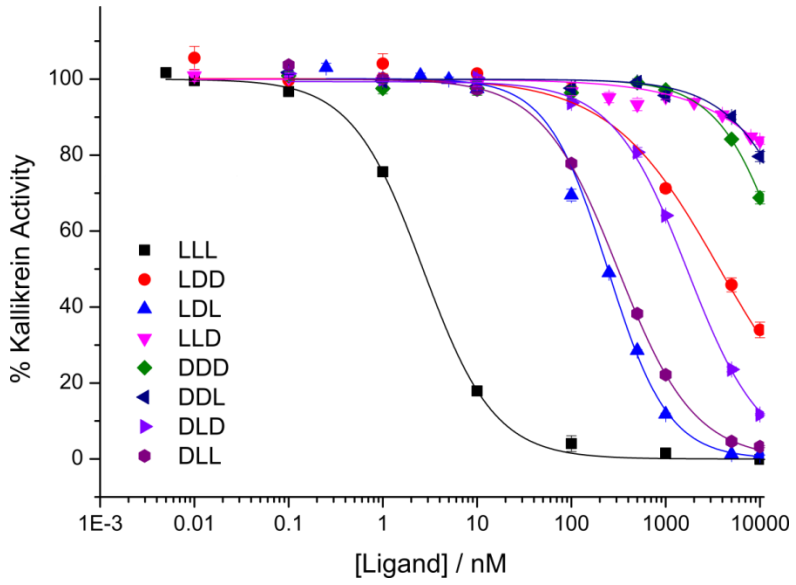


Figure 32: HPK inhibition by the 8 individual stereoisomers of PK15 where the stereochemistry at each of Cys2, Cys9 and Cys16 are either L or D.

Table 3: IC_{50} values of PK15 stereoisomers in kallikrein inhibition assay

	Isomer	IC_{50} /nM		Isomer	IC_{50} /nM
37	LLL[a]	2.7 ± 0.10	**84**	DDD	$> 1 \times 10^7$
81	LDD	3740 ± 480	**85**	DDL	$> 1 \times 10^7$
82	LDL	240 ± 14	**86**	DLD	1700 ± 50
83	LLD	$> 1 \times 10^7$	**87**	DLL	311 ± 15

[a] Naming refers to stereochemical configuration at Cys2, Cys9 and Cys16 in sequence

[iv] Synthesis and cyclisation carried out by P.J. Foley. Assay carried out by PMM with PJF's assistance.

This data validated the hypothesis that inverting the stereochemistry of the cysteine residues would significantly alter the structure of the peptide loops. The library of eight stereoisomers have inhibitory values spanning at least four orders of magnitude. Both LDL and DLL ligands remain potent inhibitors while LLD and the inhibitors with multiple stereochemical inversions had considerably reduced activity. It is not surprising that none of the stereoisomers are more active than the LLL PK15 standard – the mixture results suggested this, and the LLL ligand was already an extremely potent ligand identified through an extensive phage display screening process.

The stereochemical inversions give an insight into interesting structural-activity features of the PK15 molecule. It can be observed that single stereochemical inversions at the Cys2 and Cys9 positions, while reducing activity relative to the LLL ligand, still produce potent molecules. Of the three single inversion stereoisomers it was anticipated that the Cys9 inversion would have the largest effect due to being central in the peptide chain. However it was inversion of stereochemistry at Cys16 that almost destroyed the inhibition. To further compound this, a second stereochemical inversion to create DLD or LDD ligands rescues the activity relative to the LLD ligand. This suggests that the two loops may be interacting with each other as well as to the target kallikrein, or alternatively that the stereochemical inversions significantly alter loop structure throughout the molecule rather than just in the immediate surrounding area. Heinis *et al.* previously studied PK15 by NMR and found no interloop NOE signals indicating an absence of interaction between the loops.[14] However their data was for unbound ligand free in solution, there could be significant differences to the bound ligand conformation.

2.4.2 Modelling the PK15 stereoisomers

Further to the NMR work mentioned above, Heinis *et al.* attempted to model the structure of PK15 using the NOE constraints. This produced a model with two independent loops in an extended conformation. Recently the MacroModel (Schrodinger Inc.) suite has been shown to model macrocycles well, although large macrocycles are still a challenge for the program.[119] In the absence of NMR data for each isomer it was decided to use the MacroModel suite to investigate the effect of stereochemical inversion in the backbone. Given the large effects stereochemical inversions have on loop structure as observed via kallikrein inhibition above it was rationalised that modelling of the macrocyclic conformation may be insightful.

A 2D representation of each ligand was produced in ChemDraw and then converted to a random 3D structure using Pipeline Pilot.[v] This was used to seed the MacroModel suite's macrocycle sampling script, wherein five thousand rounds of simulated annealing were followed by five thousand rounds of energy minimisation with structures within 0.75 Å r.m.s.d.h. being discarded. The lowest twenty conformations were exported and visualised using Pymol. The results for the LLL and LLD ligands are displayed in Figure 33.

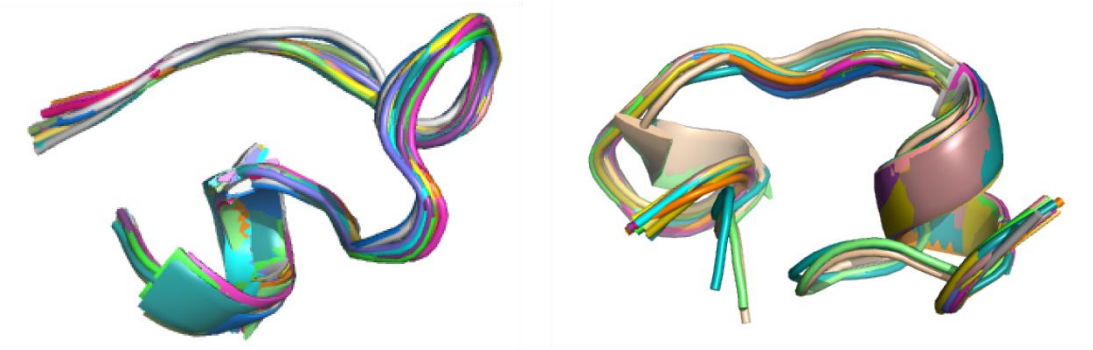
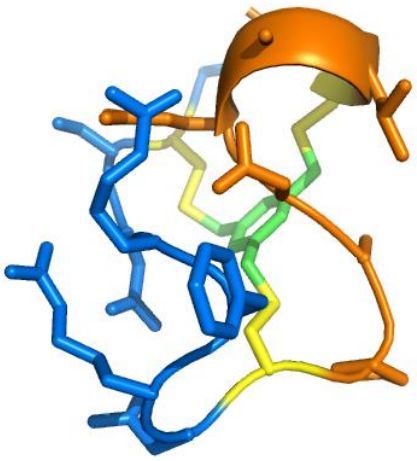


Figure 33: Comparison of LLL-PK15with LLD-PK15 molecular modelling results using the macrocycle sampling tools in Macromodel® The peptides are displayed as backbone ribbons with the mesitylene core and side chains omitted for clarity.

While it was encouraging that the modelling protocol was generating distinctly different peptide backbone structures it was of concern that the LLL structure would violate the NOE signals observed by Heinis *et al.*[14] They identified only one weak NOE signal across atoms more than two residues apart - HB Arg5: HN Cys2. Because of the lack of long- and medium-range NOE signals no single distinct structure could be proposed however the ensemble of structures which all fit with the gathered data all had an extended conformation with the loops not interacting with each other. The positions of the sidechains in the LLL ligand results are shown in Figure 34 – the close interactions between the loops would presumably have led to the observation of NOE signals.

[v] 3D structure generated by George Burslem



37

Figure 34: The lowest energy conformation of LLL-PK15 modelling results – the proximity of the sidechains is in clear violation of the NMR data obtained by Heinis and coworkers which showed only one weak NOE signal (HB Arg5: HN Cys2) between residues more than 2 positions apart. Loop 1 (blue) and 2 (orange) are clearly held closely together by the mesitylene core (green) in the modelling results which contrasts the NOE data.

To further investigate whether this modelling protocol is producing relevant structures an alternative bicyclic peptide ligand was used. UK18 is an inhibitor of urokinase plasminogen activator (uPA) which has been co-crystallised bound to its target (PDB 3QN7). The ligand was extracted from the pdb file to use as the starting point of the modelling protocol in parallel with a 3D structure generated at random from a 2D structure as above. The results of this control are shown in Figure 35 and 36. Using the pdb structure as the seed for the modelling procedure results in convergence towards a structure with similar structure in loop 1 but different loop 2 structure compared to the pdb structure (Table 4). However starting from the random 3D structure, the modelling converged on a different structure entirely. When aligned with the pdb structure the core is in an alternative conformation and there is very poor overlap with the loop structures.

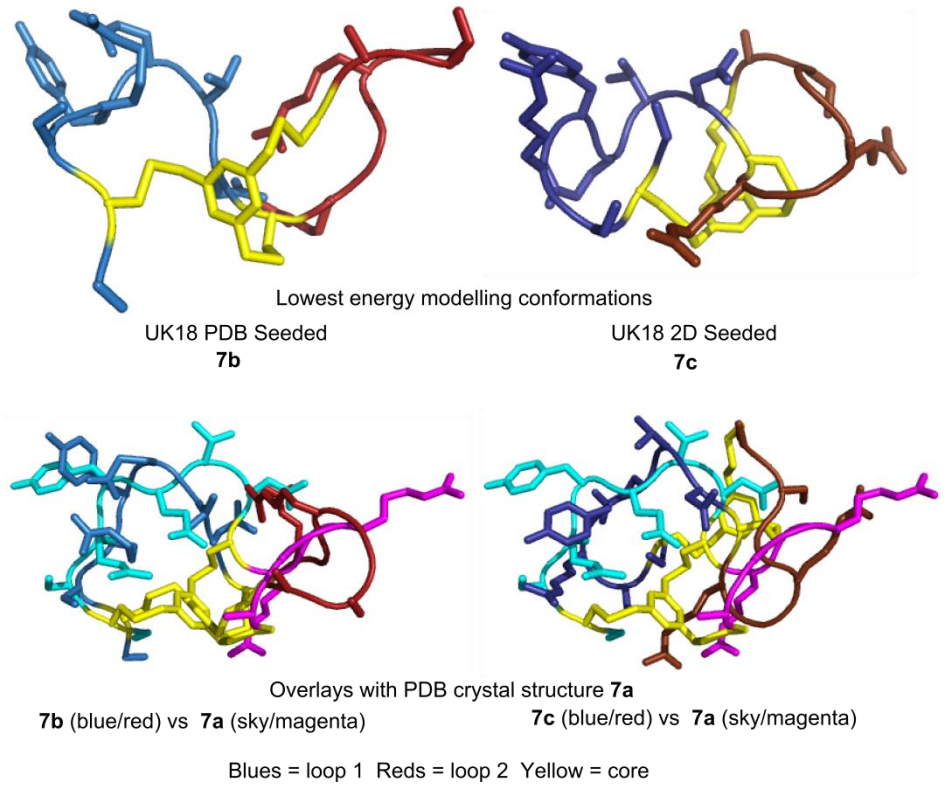


Figure 35: The lowest energy conformation of the UK18 structure determined in the macrocycle sampling protocol using both the PDB crystal structure and a random 3D structure generated from a 2D drawing as the initial seed. Not only are these different in both loops but it can be seen that when overlayed with the bound PDB ligand structure (sky/magenta) there is slight overlay in loop 1 (blues) but loop 2 overlays very poorly in both cases (reds/magenta). In the PDB seeded results the core aligns relatively well, however in the 2D seeded result the core is in an entirely different position.

Table 4: Calculated energies and r.m.s.d.h. values of modelling results vs the bound-ligand structure from the crystal structure

Structure	**Potential Energy (kJ/mol)[1]**	**r.m.s.d.h. from bound structure (Å)**		
		Total	**Loop 1 (res. 1-8)**	**Loop 2 (res. 10-16)**
PDB UK18 Structure **7a**	-2259.955	N/A	N/A	N/A
Model from PDB Seed **7b**	-3759.138	4.56	3.23	5.20
Model from 2D Seed **7c**	-3163.600	5.72	4.69	4.55

[1] Calculated using OPLS-2005 force field

These results with the UK18 peptide indicate that the modelling protocol is not converging on a global minimum. The calculated energies of the three compared structures are shown in

Table 4. This may suggest that the structure has a large degree of conformational flexibility and may reflect that the solution structure may be an ensemble of structures, and binding to a protein can stabilise one conformation in an induced fit manner. There is a large difference in energy between the bound structure and the modelling results – this is presumably offset by the binding interactions with the uPA target. Although the protocol discards conformers within 2 kcal/mol of existing hits, the backbone structure was conserved amongst the low energy hits. This suggests the sidechains have too much flexibility for this modelling protocol to work efficiently under the current restraints and a larger r.m.s.d. cut-off would need to be applied to identify alternative backbone constraints.

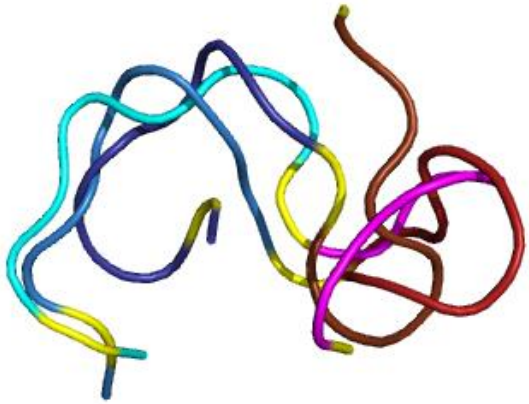


Figure 36 Overlay of the PDB bound UK18 structure (sky/magenta), the modelling result from the PDB UK18 seed (blue/red) and the modelling result from the 2D UK18 sequence seed (dark blue/brick red). Loop 1 (blues) tracks well for the PDB seed but due to the different position of the core in the 2D seed this loop does not overlay well. Loop 2 overlays poorly between all 3 structures.

2.5 Conclusions

The key step in this work involved the application of a new reagent methyl 2,5-dibromovalerate **63** to dehydroalanine conversion. This proved adept in the conversion of multiple cysteine residues to form multiple dehydroalanine residues. While this conversion was successfully incorporated into a two-step strategy to create diverse bicyclic peptides there are other applications where this conversion could prove useful.

Of primary interest is the application of this reagent to replace 2,5-dibromoadipamide in the formation of single dehydroalanine residues. **63** is more soluble than **27** and so has potential to be used in a wide range of protein modifications, using only low equivalent numbers of reagent. The reagent is stable at room temperature, although it was observed that a DMSO stock solution turned yellow over the course of about one month at 4 °C. Potential applications of dehydroalanine-containing peptides were covered in section 2.1. Preliminary

attempts at protein modification have successfully converted cysteine to dehydroalanine using **63**.[vi]

Applications requiring multiple dehydroalanine residues may be less frequent, however herein was presented an interesting use of the dehydroalanines to generate constrained peptides with diverse stereochemistry via reaction with a polythiol core. In the same manner that Heinis and co-workers have shown that different cores can then cyclise the same peptide sequence, different thiols may cyclise the dehydroalanine-containing peptides. The nature of the thiol may be vital to cyclisation though – attempts with alkyl thiols (dithiothreitol and trimethylolpropane tris(3-mercaptopropionate)) were inconclusive and the addition of thiophenol to dehydroalanine was observed to be reversible. It is interesting to hypothesise what might happen upon addition of 3,5-disulfhydrylthiophenol – in solution the ensemble of stereoisomers may equilibrate to form the most thermodynamically stable isomer. Alternatively temporary protection or labelling of a dehydroalanine residue with a thiophenol derivative could be achieved which is then removed by addition of a second thiol.

In the case of the stereochemically scrambled bicyclic peptides presented in this work, the deconvolution strategy which was applied to identify the individual activities of each stereoisomer could feasibly be applied to hit sequences from a library screen. The method of display, although intended as via phage display (see Chapter 4), could be through any of the peptide display platforms discussed in Chapter 1. Given that the crude reaction mixture gave good agreement with the purified mixture of stereoisomers this is also promising as cyclised products can immediately be used in assays without extensive purification to identify mixtures containing potent stereoisomers. This of course may be an assay-dependent feature and adequate controls would have to be established.

Perhaps the most valuable application of the new reagent would be in the production of lanthipeptides, both the final products and also intermediates to study the cyclase enzymes in detail. While preliminary attempts were made to carry out the transformation on solid-phase support (Section 6.2), these were complicated by inefficient deprotection of cysteine residues. This work merits further interest as the production of lantibiotic analogues by SPPS would allow the incorporation of non-canonical amino acids, or a focussed library approach.[120]

To further expand the applicability to lantibiotic synthesis, it would be of value to synthesise β-methylcysteine (See Section 6.2.2), a threonine mimic. This could in theory undergo double-alkylation elimination in the same manner as the cysteine residues to form

[vi] Work carried out by Rob Smith, Berry group, FBS Leeds

dehydrobutyrine. The stereoselective synthesis of this amino acid has been demonstrated starting from threonine.[121] Previous attempts to install Dhb have used solution phase chemistry to create peptide fragments containing Dhb which are then coupled by SPPS, or have created methyllanthionine building blocks.[122] The use of a chemical modification to insert Dhb would allow control over when a cyclisation reaction can occur.

Chapter 3: The synthesis of alternative cores for bicyclic peptide formation

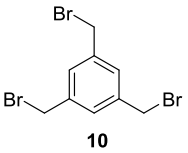
In this chapter, investigations into the synthesis and use of alternative classes of small molecule core for bicyclic peptide formation are presented; efforts toward both polythiol and alkylating agent cores are discussed. Of particular interest is the synthesis of a core based on the selective cysteine-labelling Barbas reagent - this new core forms different peptide structures from established cores.

3.1 Introduction

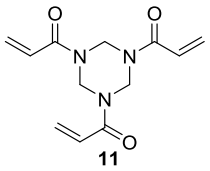
Bicyclic peptides consisting of a peptide attached around a small molecule core are potential modulators of protein-protein interactions (Section 1.2.5). Previous work by Heinis and co-workers has demonstrated that switching the small molecule core causes the same peptide sequence to adopt different loop structures (Figure 37a/b).[15, 61] This was first demonstrated using the kallikrein inhibitor PK15 sequence – switching the core used in the cyclisation of PK15 sequence **36** from **10** to 1,3,5-tris(bromoacetamido)benzene **13** (TBAB) altered its kallikrein inhibition.[61] They then observed that selecting against uPA from a phage library gave different results dependent upon which core was used for cyclisation.[15] Crystal structures of inhibitors demonstrated that the cyclisation with **13** creates a core wherein the amides are able to form intramolecular hydrogen bonds with the peptide, thus influencing the loop structure (Figure 37c).

The small molecule cores which have previously been used in the generation of bicyclic peptides are approximately similar in reactivity, size and topography (Figure 37a/b). In these cases the use of symmetrical cores avoids the production of constitutional isomers which simplifies any screening process and hit deconvolution. It was predicted that the use of a core based around a tetrahedral centre would generate an added layer of diversity through the production of enantiomeric structures where the fourth substituent of the tetrahedral centre is held on either face of the bicyclic peptide (Figure 38). This fourth substituent would also allow functionalisation of the bicyclic peptide structure without direct modification of the peptide, thus avoiding disruption of any binding interface as the fourth substituent should, in theory, be held away from the peptide structure. To combine these advantages it was aimed to either synthesise cores based upon pentaerythritol such as **89** (Figure 38b) or alternatively to investigate the use of cores based on the related tris(hydroxymethyl)methylamine base such as **90**. Both of these structures would allow further derivatisation to, for example, attach imaging reagents, adhere to surfaces or attach to another ligand.

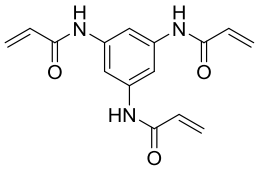
a) Winter, Heinis and co-workers



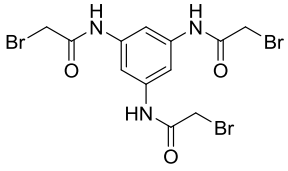
10



11

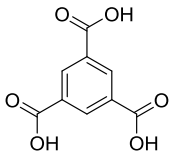


12



13

b) Pei and co-workers



88

c)

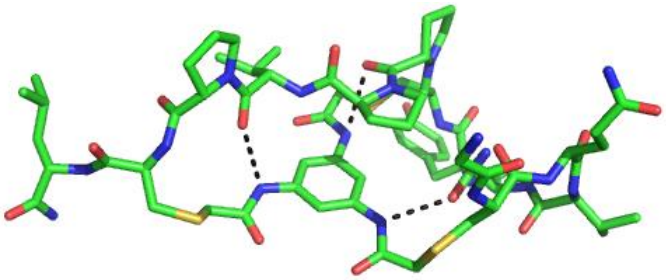


Figure 37: Alternative cores for bicyclic peptide formation. a) Cores used by Heinis and coworkers which imparted different loop structures when used to cyclise identical linear peptides. b) The trimesic acid core used by Pei and coworkers to identify **14**. c) The crystal structure of UK903 extracted from the bound crystal structure to target uPA (PDB: 4MNY). Three hydrogen bonds formed between the core and the peptide backbone are marked by black dashed lines between the heavy atoms involved.

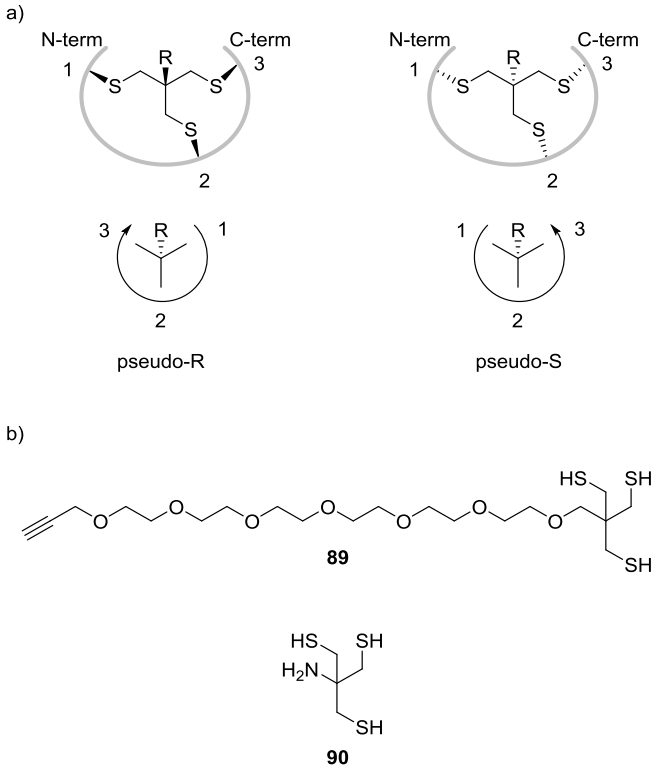


Figure 38: Tetrahedral core molecules a) The constraint of a peptide sequence by a tetrahedral core bearing three identical thiol arms generates two stereoisomers – here the residue number to prioritise the substituents and define the pseudo-stereochemistry at the tetrahedral centre. b) Two example target cores: alkyl trithiol core **89** bearing a tetrahedral centre derived from pentaerythritol **94** and a PEG chain to allow further conjugation and trithiol **90** derived from Tris base.

3.2 Synthesis of alkyl trithiol cores

Due to the success of cyclisation of dehydroalanine-containing peptides with TMMB **39** it was aimed to develop alternative thiol-based cores which could be used in the same role and test their cyclisation efficiency. Due to the observed reversible nature of thiophenol addition to Dha it was decided to focus synthetic efforts on alkyl thiols. It was aimed to incorporate a tetrahedral centre and a fourth substituent bearing a functionalised PEG tail for further conjugation.

3.2.1 Synthesis of the functionalised PEG tail

The initial objective was to synthesise the functionalised hexa-PEG **93** to be used to attach a fourth substituent to a tetrahedral core. A terminal alkyne was chosen to allow further derivatisation via copper-catalysed cycloaddition with an azide. The initial aim for the opposite terminus was to synthesise the alkyl bromide **93** for further alkylation of a tetrahedral core.

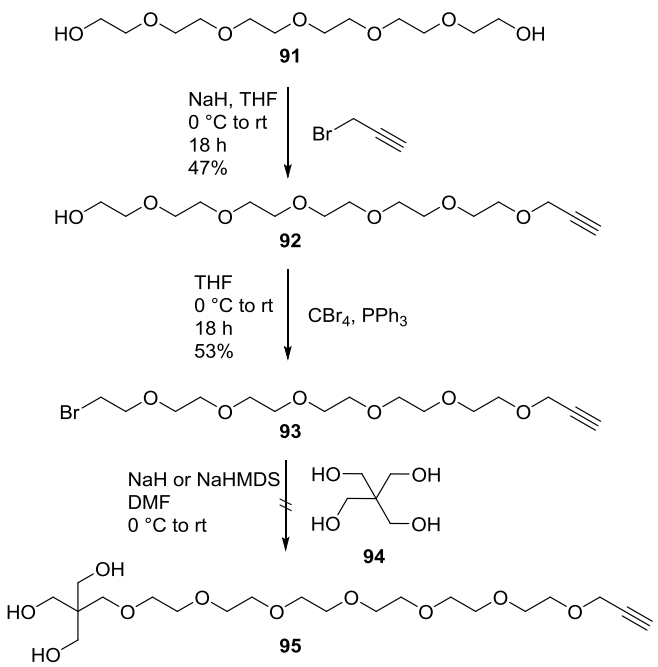


Figure 39: The synthesis of **93** from hexa-PEG **91** and attempted reaction to form **95**. An initial alkylation to add a propargyl group was followed by an Appel reaction to synthesise the alkyl bromide. The alkylation of pentaerythritol to deliver **95** proved problematic.

The synthesis of **93** was carried out in two steps from **91**. The initial alkylation produced a stoichiometric mixture of 1:2:1 double:single:zero alkylation products with the single alkylation product **92** being isolated in 47% yield. An Appel reaction[123] was then used to convert the alcohol **92** to the primary bromide **93** in 53% yield. With **93** in hand it was then aimed to insert a tetrahedral core bearing hydroxyl groups which could subsequently be

converted to thiols. Pentaerythritol **94** was chosen, although there were concerns about its low reactivity. Alkylation using pentaerythritol **94** was attempted but under conditions tested the bromide favoured elimination to form the alkene **96** (Figure 40). Pentaerythritol was a poor nucleophile – no reaction was observed when using mild bases triethylamine and imidazole, necessitating the use of the relatively strong bases. Some of the desired ether product **95** was observed by LCMS but upon addition of further base to drive reaction to completion the product also eliminated to form **96**.

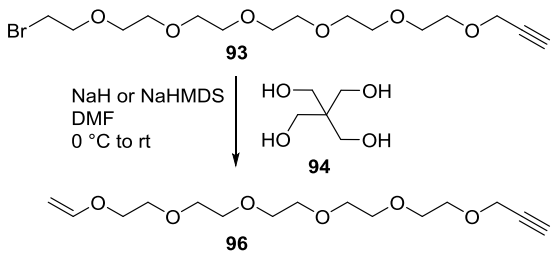


Figure 40: Elimination was observed in attempts to use **93** to alkylate pentaerythritol **94**. No reaction was observed using weaker bases.

It was therefore decided to replace the ether linkage with a more stable amide linkage (Figure 41). Jones oxidation of **92** delivered the carboxylic acid **97**. An EDC-activated amide coupling was carried out with tris(hydroxymethyl)methylamine (Tris) to provide the trishydroxy compound **98** (Figure 41). EDC-activated amide formation was initially attempted in aqueous conditions to solubilise the Tris. However the amide bond formation was low yielding due to competing hydrolysis of the EDC-activated acid. The polar product required continuous extraction to purify. Addition of DMAP to the reaction failed to improve the conversion and additionally DMAP was co-purified in the continuous extraction process. An attempt was made to replace the EDC coupling with a two-step conversion via acid chloride synthesis with thionyl chloride followed by reaction with Tris in DMF. Although the first step proceeded quantitatively by NMR, the second step proved more problematic and only a 19% yield was achieved following continuous extraction.

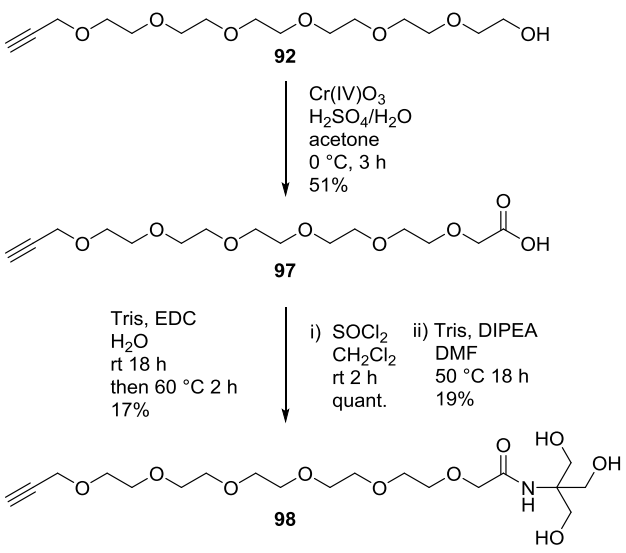


Figure 41: The synthesis of **98** from alcohol **92**. Jones oxidation of **92** provided acid **97** which was either directly coupled with Tris via an EDC-activated amide formation, or was converted to the acid chloride before direct reaction with Tris.

3.2.2 Synthesis of trithiol subunits

With **98** in hand the next step in the proposed synthesis was the conversion of hydroxyl groups to thiols. An alternative route to the pentaerythritol-derived core **89** was targeted via the route shown in Figure 42. Tribromide **99** was synthesised via bromination of pentaerythritol in 31% yield. The tribromide **99** was shown to be unreactive toward the PK15 peptide sequence **36** under the cyclisation conditions used for the reaction of PK15 **36** with TBMB **10**.[vii] It was aimed to convert **99** to trithiol core **100** via thiourea alkylation and hydrolysis. Although an initial addition of thiourea was observed by MS, subsequent additions were not observed and intermediates or product **100** could not be detected by NMR.

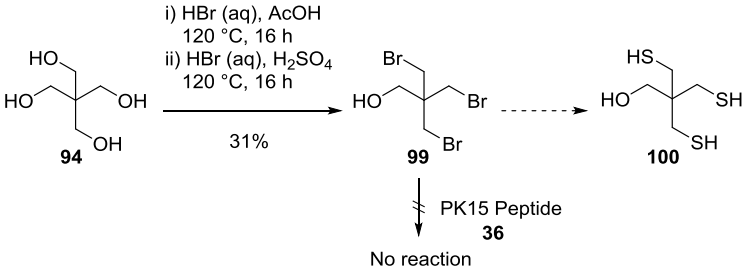


Figure 42: Attempted synthesis of **100** via bromination and thiourea alkylation. The bromide **99** was inert to peptide cyclisation, and alkylation with thiourea also proved problematic.

It was decided to attempt a Mitsunobu reaction with thioacetic acid to install the thiol functionality (Figure 43). **102** was used as a model system to develop conditions for this conversion. The alkylation of pentaerythritol was again hampered by its poor solubility and nucleophilicity. Reaction with one equivalent of benzyl bromide led to multiple additions due to enhanced solubility of the alkylated products compared to pentaerythritol. The polyalkylated by-products were removed by ether extraction and a subsequent extraction of the aqueous layers with dichloromethane then gave straight-forward purification of **101**. Though low-yielding this reaction gave sufficient material to proceed in preference to investigating protecting group strategies.

[vii] Attempted cyclisation with **36** carried out by Patrick Foley

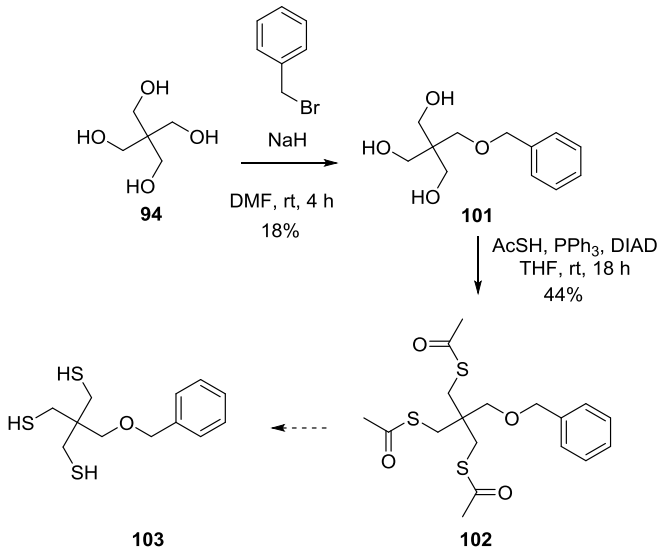


Figure 43: The attempted synthesis of **103** via monobenzylation of pentaerythritol followed by a Mitsunobu reaction to convert the hydroxyls of **101** to thioacetates. Deacetylation of **102** was thought to be achievable via either hydrolysis or reduction however attempts were unsuccessful.

A Mitsunobu reaction with thioacetic acid was used to install the masked thiol functionality, there remained trace triphenylphosphineoxide (TPPO) impurity present.[124] In contrast to the literature reports,[124] deprotection of the acetate groups of **102** was not without complication. Lithium aluminium hydride did not appear to deprotect the thioacetates as intended. Due to further problems with thioacetate deprotection, as discussed below, this route was abandoned and an alternative thiol synthesis sought.

Synthetic efforts also focussed on Tris-based cores and attentions returned to the use of thioacetate to install thiol functionality. The Mitsunobu reaction with thioacetic acid was attempted using **98** as a substrate however this failed – the amide appeared to intercept the activated alcohol, forming the oxazoline structure **105**.

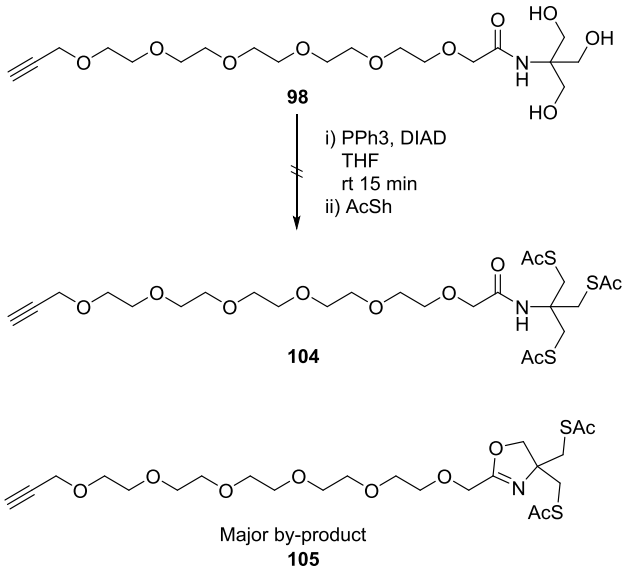


Figure 44: The unsuccessful attempted Mitsunobu reaction to install thioacetate functionalities and form **104** starting from **98** was disrupted by formation of the oxazoline **105**.

An alternative strategy would be to generate a protected tris(mercaptomethyl)methylamine prior to amide coupling to the PEG tail. It was rationalised that using Tris directly as a substrate in the Mitsunobu reaction with thioacetic acid would be unwise due to acetyl transfer from thioesters to the free amine. This would obstruct attempts at further functionalization at this position. It was thought that the use of tris(hydroxymethyl)nitromethane may circumvent this if the thiols could be deprotected prior to nitro reduction. The Mitsunobu reaction proceeded in 54% yield however the removal of the acetate groups again proved problematic – a white polymeric substance formed which solubilised upon treatment with DTT. Presumably a disulfide polymer was formed which still formed oligomeric species upon reduction. Due to the difficulties in obtaining the tris(mercaptomethyl)methane-based cores **104** and **108** work on this motif was abandoned.

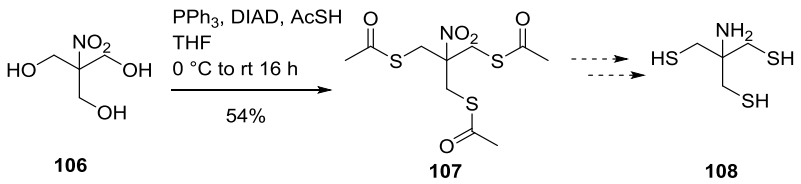


Figure 45: The synthesis of **107** via a Mitsunobu reaction with thioacetic acid. Deprotection attempts were unsuccessful – leading to suspected polymerisation.

3.3 Synthesis of alternative alkylating core molecules

It was also aimed to test the effects of varying the core topography on peptide structure via the synthesis of cores **109** and **110** (Figure 46). Both cores were intended to react directly with cysteines in peptides. **109** was previously shown to react with nucleophiles to form trisubstituted tertiary amines.[125] The 1,3,4-oxadiazole core was inspired by work by Barbas and co-workers. They demonstrated that 2-mesyl-5-phenyl-1,3,4-oxadiazole selectively labels cysteine residues as an alternative to maleimide chemistry.[126] The extended aromatic ring structure should provide a more rigid and planar core than the previously-used cores. Heinis and co-workers have made cores slightly larger than TBMB **10** (Figure 37) but these have flexible arms and so the distance between cysteine residues can vary for example in a range of 3.5 to 13.1 Å using TBAB **13**.[15]

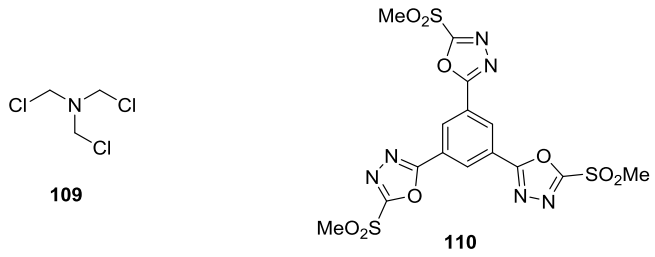


Figure 46: Target core molecules to investigate the effect of small and large steric constraints on peptide structure

3.3.1 Tris(chloromethyl)amine-derived cores

The synthesis of tris(chloromethyl)amine **109** was first reported in 1967 via the reaction of trisodium 1,1′,1″-nitrilotrismethylsulfonate and PCl_5.[127] This was improved upon by Fluck and Meiser with the reaction of hexamethylenetetramine **111** with PCl_5 (Figure 47).[128] They then demonstrated the reaction of **109** with a range of nucleophiles, including alkoxides, amines and Grignard reagents. All classes of nucleophile tested underwent triple addition to the trimethylamine core with the exception of the addition of β-mercaptoethanol which was observed to undergo only two additions, with the product appearing to have molecular formula $N(CH_2SCH_2CH_3)_2CH_2Cl$.[125] **109** has subsequently been shown to react with sodium azide and ammonium 5-nitrotetrazolate to form explosive materials by Klapötke and co-workers.[129] It was decided to use the small trimethylamine core as the basis of potential new cores for bicyclic peptides. The close resemblance of **109** to carcinogenic nitrogen mustard gases was noted – tris(chloroethyl)amine is the only nitrogen mustard still used in chemical warfare. However due to the inability of **109** to form a reactive aziridinium intermediate it was hoped that this structure would be less hazardous.

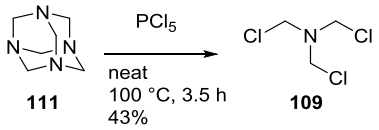


Figure 47: The synthesis of **109** from **111** following literature procedure[128]

Tris(chloromethyl)amine **109** was synthesised following the literature procedure from hexamethylenetetramine **111** (Figure 47).[128] Due to technical difficulties with sublimation apparatus product **109** was instead purified by recrystallization from dichloromethane in 43% yield. An initial reaction of **109** with peptide ACAHTGCVYDSCG did not lead to any observed cyclisation in either water or acetonitrile. This was not unexpected as **109** appears extremely reactive with water – when exposed to air an acute smell of HCl was observed.

Using **109** as a starting point, it was targeted to synthesise the trithiol **113**. It was aimed to synthesise this via the displacement of the chlorides with a protected thiol and subsequent deprotection. In contradiction to Fluck and Meiser's observation of βME addition noted above, the addition of *tert*-butyl thiol gave the product of triple-addition **112** in 56% yield (Figure 48). Attempts to deprotect the tBu groups were unsuccessful. It may be possible to synthesise **113** directly from **109** with sodium sulfide, however this was not attempted.

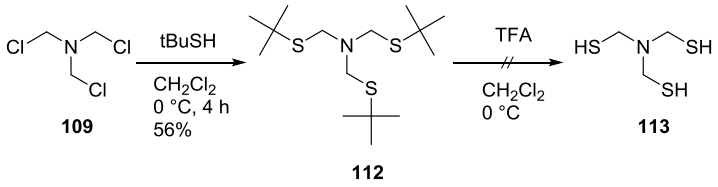


Figure 48: The synthesis of **112** was achieved from **109** however deprotection attempts were unsuccessful.

3.3.2 Cores based on Barbas reagent - synthesis of oxadiazole core 110

As described in Section 3.3 it was aimed to synthesise the core **110**, containing multiple 2-mesyl-1,3,4-oxadiazole moieties. It was thought this extended aromatic system would have a rigid extended structure when compared to the cores in Figure 37.

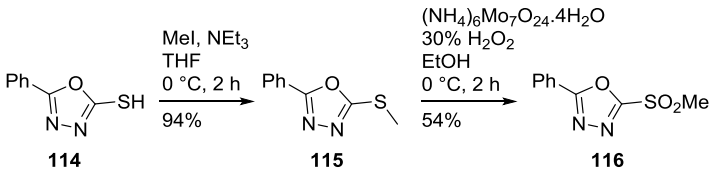


Figure 49: The synthesis of 2-mesyl-5-phenyl-1,3,4-oxadizole **116** by Toda *et al.*[126]

The synthesis of 2-mesyl-5-phenyl-1,3,4-oxadiazole **116** by Barbas and co-workers[126] is shown in Figure 49 (molecules derived from this have since been classified as Barbas reagents). These conditions were adapted to the proposed synthesis of core **110** displayed in Figure 50.

This synthetic route adapted to the synthesis of multifunctional core **110** (Figure 50) proved inefficient. While the initial hydrazide formation proved straight-forward, the cyclisation reaction with carbondisulfide proved low-yielding. The methylation was also disappointingly low-yielding – it was observed that hydrazide by-products were formed.

The final oxidation step also proved to be problematic – it was observed that over-oxidation of the mesylate species **120** to form oxadiazal-2-one by-products occurred. This was presumably not a major problem for the synthesis of **116**, where the lower levels of oxidant required would lead to less by-product formation. However when six oxidations were attempted simultaneously, the reaction profile was more complicated – the over-oxidation to form a 1,3,4-oxadiazol-2-one appears to be at least as favoured as the sulfur oxidations. The extended conjugation in **120** relative to **115** could also mean the carbon is even more electropositive.

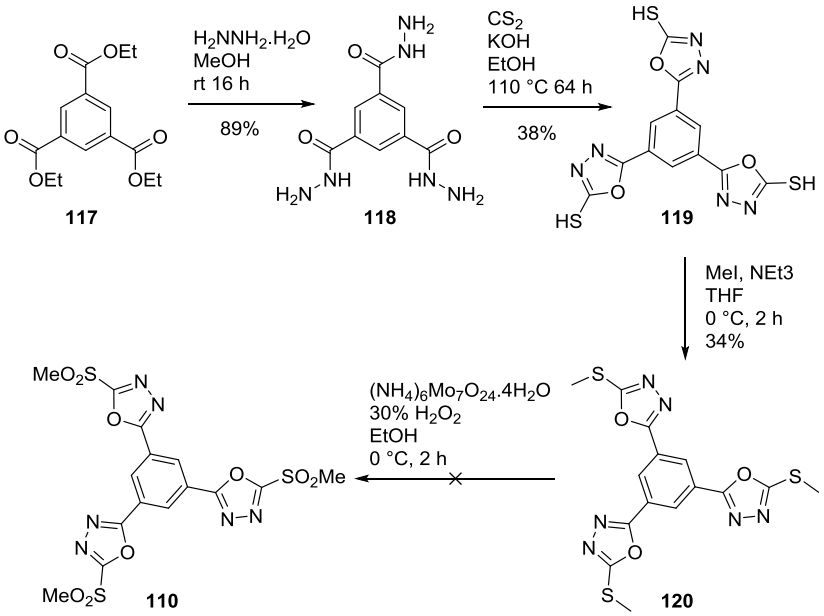


Figure 50: Initial synthetic efforts to core **110** replicating methods used in the synthesis of **116**.

The yield of the methylation step to form **120** was improved by a switch of solvent to DMSO. The starting material **119** proved more soluble in DMSO than THF, and fortuitously the product **120** precipitated from the reaction mixture. This serendipitous development avoided the need for further work-up or purification steps. This change improved the yield, however for a simple alkylation it was still below expectations.

Several attempts were made to improve the oxidation step (Table 5) using either **120** or 5-phenyl-2-thiomethyl-1,3,4-oxadiazole **115** as substrates. Only minimal improvements were achieved via the use of peroxide based approaches. The greatest improvement was achieved using a biphasic reaction involving ruthenium catalysis and sodium periodate. This forms ruthenium tetroxide *in situ*, which although usually unstable, is stable in carbon tetrachloride. This oxidant proved superior as, similar to oxidation with chromium trioxide (Section 3.2.1), the target of oxidation must interact with the metal centre. The lone pairs of the sulfide are capable of this co-ordination whereas the carbon is unable to co-ordinate. The reaction proceeded quickly and cleanly and after HPLC purification a greatly improved 65% yield was obtained. It should be noted that purification by silica chromatography led to the degradation of **110**.

Table 5: Synthetic methods attempted in the optimisation of the oxidation of **120** to **110**

Entry	Conditions	Observations	Yield of
	Substrate 115		**116**
1	Oxone, 2:1 MeOH:H_2O, rt	Over-oxidation observed	69%
2	MMPP.6H_2O, MeCN, rt	Complete over-oxidation	-
3	$NaIO_4$, $RuCl_3$, 1:1:2 CCl_4:MeCN:H_2O, rt	Quantitative conversion	100%
	Substrate 120		**120**
4	$(NH_4)_6Mo_7O_{24}.4H_2O$, 30% H_2O_2, EtOH, 60 °C	Over-oxidation	-
5	5% $KMnO_4$ (aq), AcOH, 10 °C	No reaction (insoluble)	-
6	$(NH_4)_6Mo_7O_{24}.4H_2O$, 30% H_2O_2, [bmim]PF_6, 60 °C	Poor solubility, Over-oxidation of soluble material	-
7	mCPBA, 3:4 DMF:CH_2Cl_2, 40 °C	Over-oxidation observed	8%[a]
8	$NaIO_4$, $RuCl_3$, 1:1:2 CCl_4:MeCN:H_2O, rt	No over-oxidation observed.	65%[a]

[a] NB do not column on silica – product degrades

The optimisation of the alkylation and oxidation steps greatly improved the synthesis of **110 (**Figure 50). Combining the original synthetic route to **119** with the use of DMSO as solvent in the methylation reaction and the use of the ruthenium catalyst in the oxidation (Table 5 entry 8) gives an 11% yield over the four steps to synthesise **110**. While there is still room for further optimisation of these conditions this route delivered 81 mg of final product – more than sufficient to investigate bicyclic peptide formation with **110**.

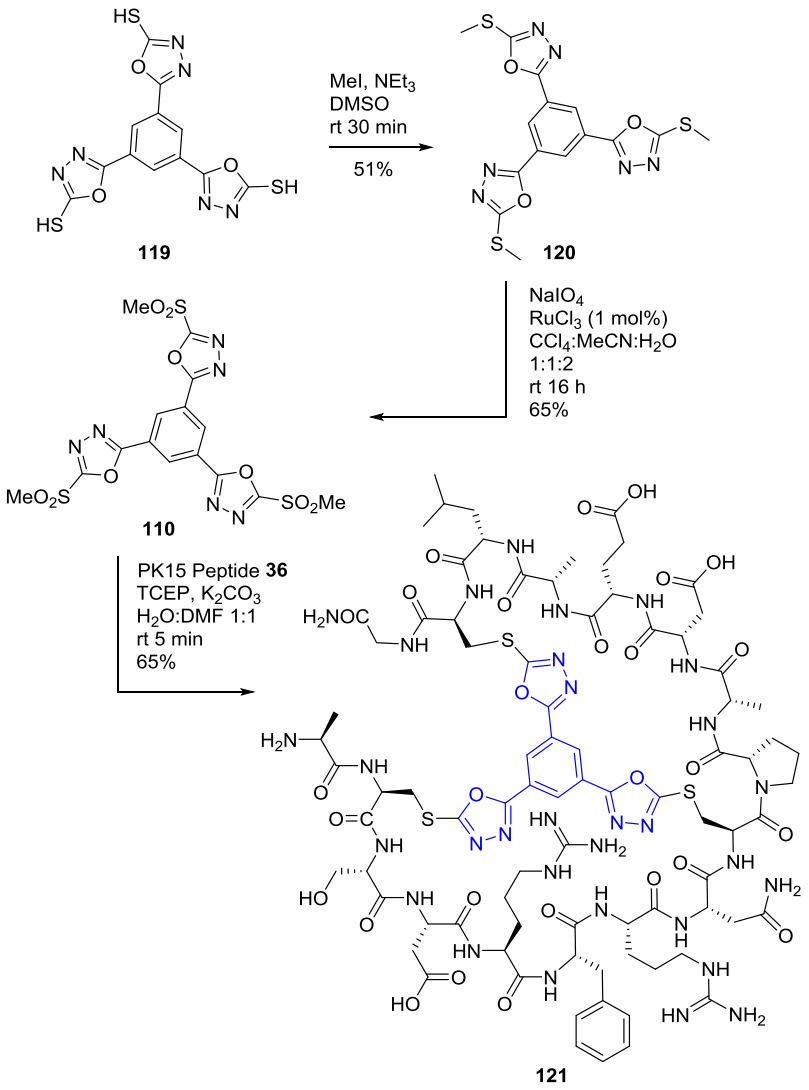


Figure 51: Optimised synthesis of **110** from intermediate **119** and subsequent use in the cyclisation of PK15 sequence peptide **36** to generate bicyclic peptide **121**.

3.4.3 Testing and modelling bicyclic peptide PK15-oxa 121

To demonstrate the potential of the core **110** to cyclise cysteine-containing peptides it was reacted with the PK15 kallikrein inhibitor sequence **36** (Figure 51). The reaction appeared complete within five minutes by LCMS and after HPLC purification **121** was obtained in 65% yield. This compound was tested in the same kallikrein inhibition assay used in Section 2.3.2 and shown to be inactive within the concentration range tested (Figure 52); as an approximation a logistic fit with the baseline activity set at 0% and the power set as 1 gives an approximate IC_{50} of 360 ± 62 µM. This is 10^5 times lower than the PK15 sequence when cyclised with the mesitylene core and 100-fold worse than the TBAB core Heinis and co-workers tested. This suggests that the peptide structure is very different between the two

cores; in this case the loops presumably have less backbone flexibility due to the rigid core. It would be of interest to obtain an NMR or crystal structure of **121** and compare it to the modelled structure shown in Figure 53. A single linear peptide library would give rise to very different bicyclic peptide libraries if cyclised in parallel with TBMB **10** and the Barbas-based core **121**.

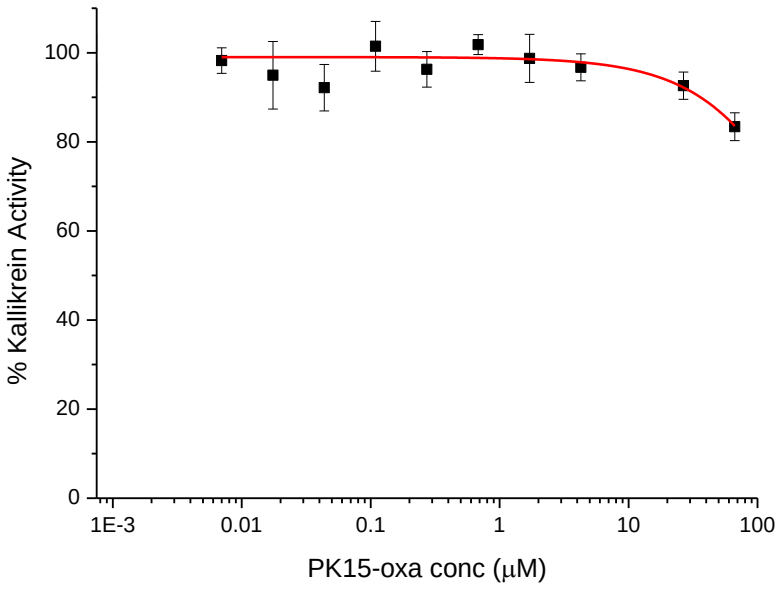
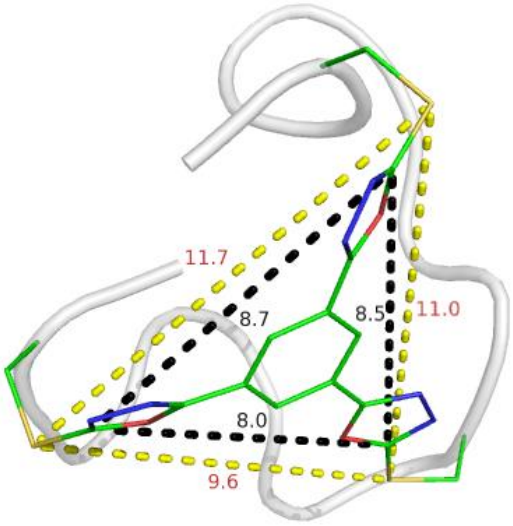


Figure 52: The results of testing compound **121** for inhibition of HPK via the cleavage of Z-Phe-Arg-AMC **78**. **121** proved almost totally inactive in the concentration range tested – at least 5 orders of magnitude worse than the same peptide sequence cyclised by TBMB **10** (PK15 **37)**, and 2 orders of magnitude worse than the same sequence cyclised by TBAB **13**.

The lack of activity indicates the peptide is adopting a significantly different structure and is held unable to form the correct binding pose. The same macrocyclic modelling protocol was followed as for the stereoisomers in Section 2.4.2 (Figure 53). The results for the PK15 sequence attached around the oxadiazole core determined a relatively rigid core with the sulfur atoms of the cysteine residues spaced between 9.6 and 11.7 Å apart – with the variance due to slight twisting of the aryl rings and the relative orientation of the oxadiazole rings (Figure 53). This result supports our hypothesis that the core forms a less flexible scaffold than the molecules used previously (Figure 1) and leads to a peptide of significantly different structure.



121

Figure 53: The structure of the PK15 sequence attached around a 1,3,5-tri-1,3,4-oxadiazolylbenzene core modelled using the same macrocyclic sampling protocol as in section 2.4.2. The core is relatively rigid which holds the sulfur atoms of three cysteines approximately planar with the core.

3.4.4 A bis-Barbas stapling reagent

In preliminary efforts to synthesise **110,** the hydrazide formation step to form **118** produced a high yield of bis-hydrazide material **123** rather than tris-hydrazide **118** or even a stoichiometric mixture of hydrazides despite the use of ten equivalents of hydrazine hydrate. In contrast the use of fifteen equivalents gave clean production of **118**. This is difficult to rationalise other than to assume that the concentration of hydrazine was below that stated. This material was used to test the subsequent reaction steps in the synthetic route to **110** and often the products were taken on directly as this was only intended as preliminary method development. The product of the oxadiazole formation step **124** was taken on into the methylation step without further purification. Post-cyclisation the methylation step was straight-forward, in contrast to the problems with the synthesis of **120**. The oxidation step to form **126** was also more straight-forward than in the synthesis of **110** – the use of the molybdenum catalyst and hydrogen peroxide gave **126** in 23% yield. Although **126** was further reacted with a peptide (see below), it degraded prior to complete characterisation.

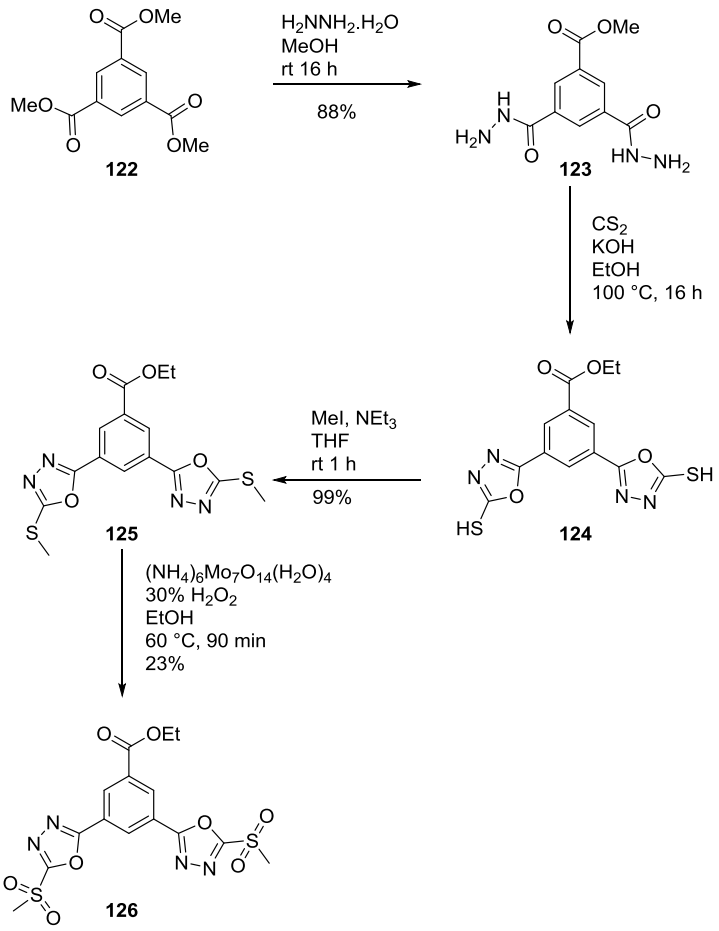


Figure 54: The synthesis of bis-Barbas reagent **126**

It was subsequently realised that **126** may make a useful reagent for the cyclisation or stapling of peptides. As well as incorporating the 2-mesyl-1,3,4-oxadiazole moieties the third substituent acts as a point for further functionalization. To prove the appropriate reactivity of **126**, it was reacted with peptide LTFCEYWAQLCSAA **74** wherein the cysteines have been introduced as potential cyclisation handles at positions 4 and 11. This peptide sequence was chosen as it closely resembles that of an MDM2 inhibitor **127** developed by Aileron Therapeutics (Figure 55).[130]

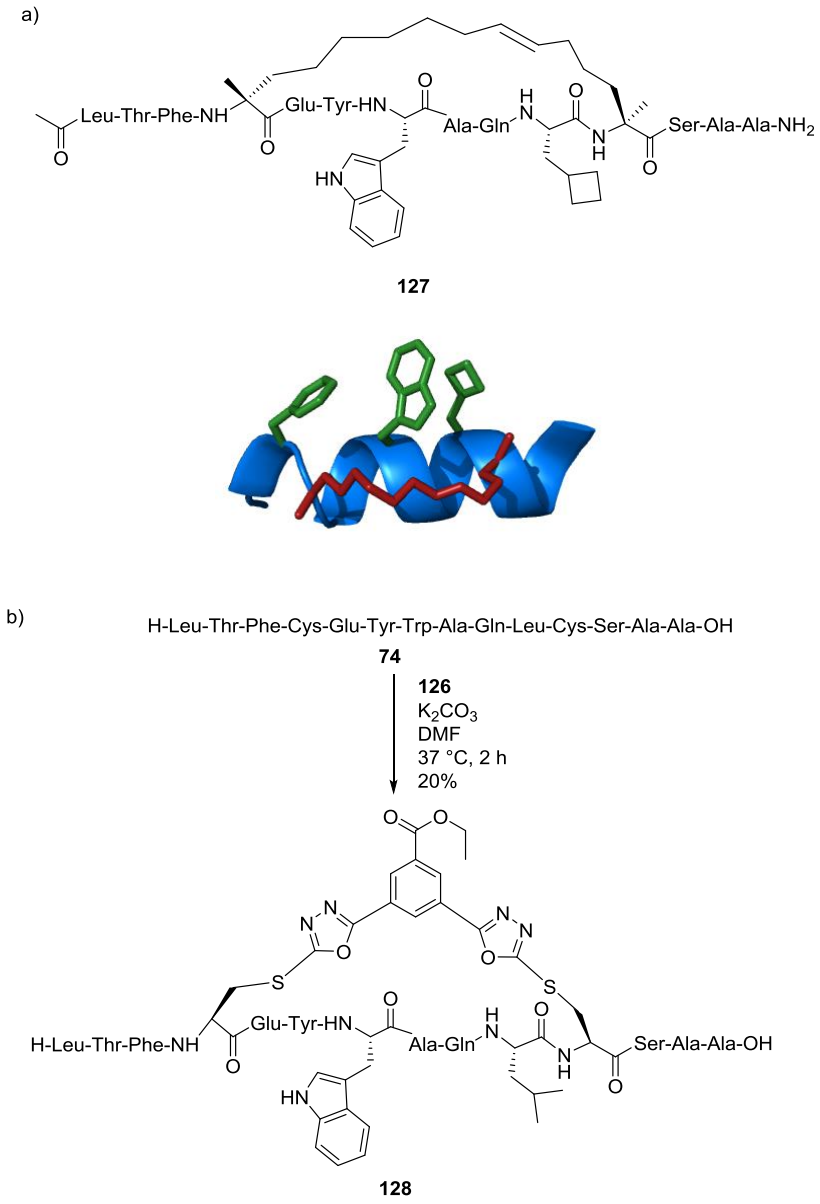


Figure 55: Application of the bis-Barbas reagent **126** as a staple. a) the MDM2 inhibitor ATSP-7041 developed by Aileron Therapeutics contains a hydrocarbon staple which presents the three residues in a position to mimic the P53 helix which binds to MDM2.[130] b) Cyclisation of the similar peptide **74** with reagent **126** in an effort to stabilise a helical conformation.

The synthesis of **128** proceeded cleanly with the 20% yield attributable to losses on the HPLC purification system. **128** was compared to the parent peptide **74** by circular dichroism (CD) in 1:1 water:acetonitrile for evidence of alpha helicity. Unfortunately **128** appears to have even less helical character than the parent peptide **74**. The stapling has instead induced an alternative cyclised form. There are two likely causes of this – firstly the stereochemistry of the cysteine at position 4 is the opposite of the amino acid used in the MDM2 inhibitor. Secondly the choice of solvent used in the CD was not optimised – the peptide is likely to adopt different conformations dependent upon the solvent used. It is likely that a switch to a

mixture of water:trifluoroethanol might enhance helix formation. Spring and co-workers have previously used a similar class of i, i+7 staple formed via an azide-alkyne cycloaddition as demonstrated in stapled peptide **129**.[131] The stapling using **126** would create a constraint of similar size and geometry if homocysteine was used in place of cysteine in the peptide sequence.

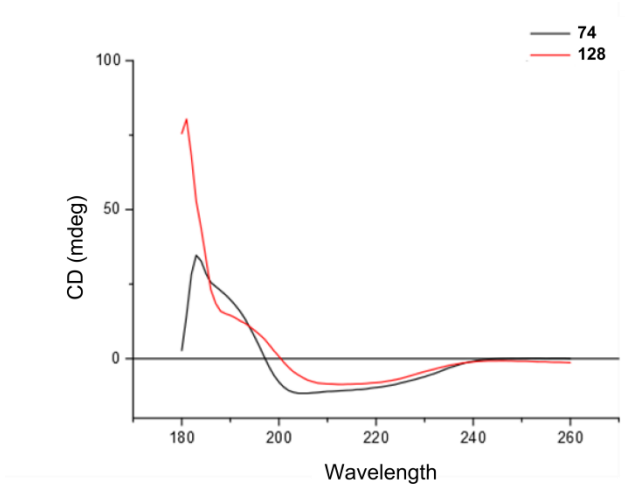
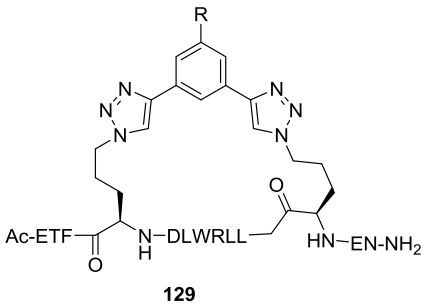


Figure 56: The CD spectra of the peptide **74** (black) and its cyclised variant **128** (red) in 1:1 water:acetonitrile. **74** has minor alpha helical character which is not observable in cyclised peptide **128**.

In summary the 2-mesyl-5-phenyl-1,3,4-oxadiazole is a synthetically tractable alternative to halides for the selective alkylation of cysteine residues to create constrained peptides. The rigid cores have potential to restrain peptides in bioactive structures which may be useful as modulators of PPIs.



129

Figure 57: A similar triaryl staple developed by Spring and co-workers.[131]

Chapter 4: Phage display of bicyclic peptides via g3p phagemid vector

In this chapter, work toward the development of a phage display construct for the display of a peptide containing multiple cysteine residues is presented. Two approaches are presented – display of a fusion with either cysteine-free g3p protein or a g3p truncation mutant. The challenges of efficient display and chemical modification of phage are discussed.

4.1 Introduction

Following the development of both the dehydroalanine chemistry (Chapter 2) and the oxadiazole-based core (Section 3.3.2) it was aimed to test these methods on peptides displayed using a library display platform. Previous work with bicyclic peptide structures has used either phage display[14, 82] or solid-phase resin approaches.[62] Phage display methodology was selected due to existing expertise and facilities in the McPherson and Tomlinson groups. Phage display of the linear peptides and subsequent cyclisation has previously been achieved by Heinis and co-workers using a single phage vector approach (Section 1.2.5) to display a peptide sequence containing three cysteine residues as a g3p conjugate. g3p is the 'tail' protein of a phage, present in five copies per phage unit, with an important role in binding to *E. coli* pili to initiate infection of the bacterium. g3p consists of three domains – two N-terminal domains with similar beta barrel structure and a C-terminal domain which attaches to the rest of the phage. Heinis and co-workers reported treatment with TBMB caused problems with phage infectivity – proposed to be due to cross-linking of WT g3p proteins via native cysteines. These problems were alleviated by use of a cysteine-free g3p mutant. This cysteine-free g3p (Figure 58) was developed by Kather *et al.* in a protein folding mutagenesis study;[83] six native cysteines in the N1 and N2 domains have been removed and the protein structure stabilised by a further nine mutations. The mutated g3p has reduced infectivity compared to the wild-type g3p, however this was still greater than the infectivity observed when WT g3p had been reacted with TBMB.[82]

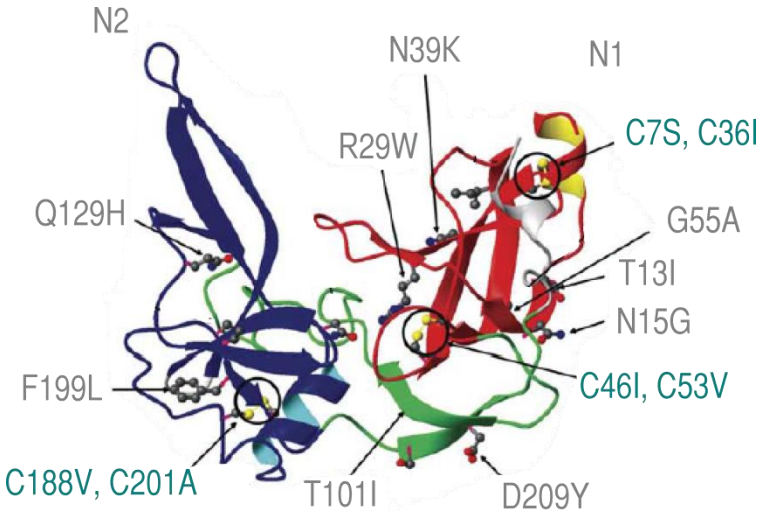


Figure 58: The mutations introduced into g3p by Kather, Bippes and Schmid to replace three disulfide bonds in the N1 and N2 domains of g3p.

In contrast to the single vector approach employed by Heinis, the Adhiron library developed at Leeds (Section 1.1 and Chapter 5) is displayed using a phagemid approach. In this approach one vector – pBSTG1 – encodes the Adhiron as a fusion with a truncated g3p

variant consisting of only the C-terminal domain. A second vector encodes the remainder of the phage proteins including a WT g3p to ensure infectivity. This second phage vector is delivered to the *E. coli* via use of a secondary helper phage. This approach avoids the use of a whole-phage vector which could cause contamination issue for other researchers. In addition the in-house standard phagemid protocols are well optimised and have been used for a variety of displayed libraries. It was aimed to adapt this approach to display a peptide containing two variable loops spaced by three cysteine residues which can be modified chemically.

4.2 First iteration phage studies

The target pBSTG1 vector is depicted in Figure 59. The leader sequence is responsible for export of the fusion protein to the periplasm where it can fold correctly. Several leader sequences were tested as different sequences have been shown to be better for different fusion protein exports.

The mutated cysteine-free g3p, hereafter termed g3p*, was used as although we did not anticipate the same problems with cross-linking when using methyl 2,5-dibromovalerate **63** to install dehydroalanine residues, it is possible that the oxadiazole reagent would cross-link the g3p proteins. The PK15 sequence **36** was used as the model system peptide to test display and the success of chemical modifications as it had previously been reacted successfully with both methyl 2,5-dibromovalerate **63** and the oxadiazole-based core **110**. In addition it was aimed to use the TBMB **10** cyclisation chemistry to verify the level of display and infectivity.

With respect to the secondary phage used to supply the second vector, two options were considered. The standard helper phage was the first option – there were concerns that the WT g3p would be cross-linked and impair infection but this might be overcome by combination with the use of the g3p* fusion. Additionally the WT g3p might be incorporated into phage in preference to the g3p* fusion leading to low display levels of the conjugate. A second option was the use of commercially available Hyperphage® - a helper phage system which lacks a g3p variant in the genome it delivers. This hyperphage system would be compatible with all the regular helper phage protocols but should produce phage which display multiple copies of the g3p* fusion. The major drawback of Hyperphage is it is an expensive commercial option (approximately £360 per litre of culture).

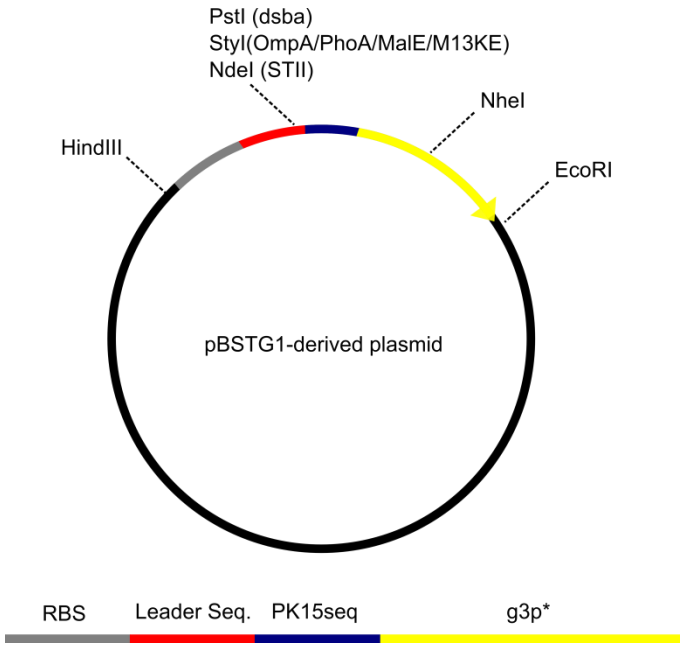


Figure 59: The target vector to express a conjugate between the PK15 peptide sequence and g3p* . Important features include the ribosome binding sequence, a leader sequence for periplasmic export containing a restriction site which along with the NheI site in the g3p* sequence could be used to create a library of peptide sequences in place of the PK15 sequence.

The g3p* gene was supplied in the pUC57 vector (Genscript), codon optimised for *E. coli* with the dsba leader sequence and HindIII/EcoRI restriction sites. A two-step PCR procedure was undertaken to change the leader sequences to yield a library of six leader sequences detailed in Table 6. Silent mutations were introduced into the leader sequence DNA to input a restriction site which could be used in any future library generation. A second glycine was inserted at the N-terminus of the PK15 sequence to ensure appropriate peptide cleavage as guided by the SignalP prediction algorithms.

Table 6: The periplasmic export leader sequences used in this work

Leader sequence	**Amino acid sequence**
dsba	MKKIWLALAGLVLAFSASA
OmpA	MKKTAIAIAVALAGFATVAQA
MalE	MKIKTGARILALSALTTMMFSASALA
STII	MKKNIAFLLASMFVFSIATNAYA
PhoA	MKQSTIALALLPLLFTPVTKA
M13KE	MKKLLFAIPLVVPFYSHS

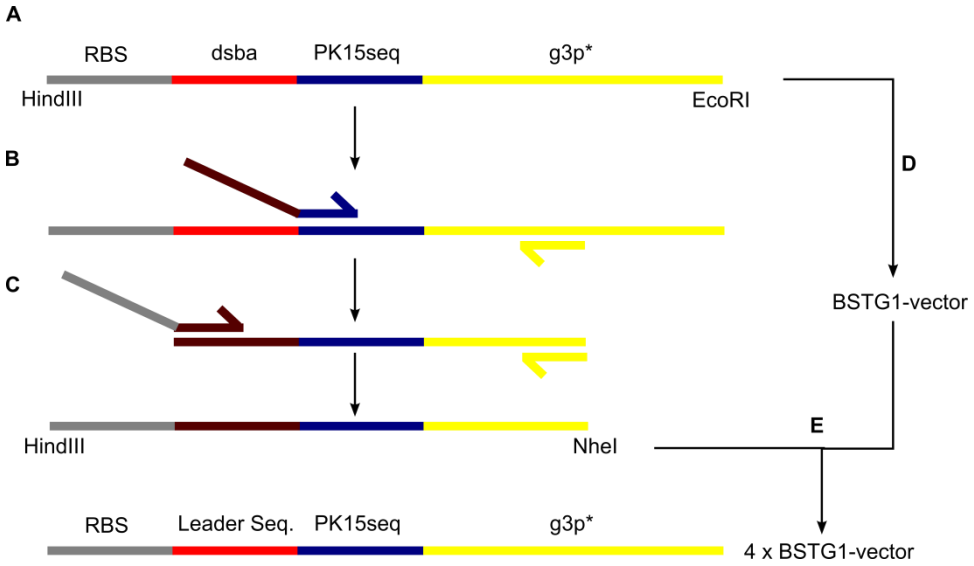


Figure 60: The PCR and cloning strategy to create the library of BSTG1 vectors containing the PK15-g3p* fusion protein and a range of leader sequences. Starting from a gene in pUC57 (A), a two-step PCR process installed the leader sequences (B) and reinstalled the ribosome binding site and HindIII restriction site (C). The original gene was subcloned into BSTG1 plasmid (D) and the resultant plasmid used as the recipient plasmid for the PCR products (E). This gave a set of four leader sequences to test.

An intial PCR step inserted the leader sequences ahead of the PK15 sequence (Figure 60B). The reverse primer used covered the NheI site which was incorporated in the g3p* sequence to allow cloning of an intermediate fragment – this site could then be used in any future library generation. A second PCR step then extended each gene to reinsert the ribosome binding site and the HindIII restriction site (Figure 60C). Optimisation of annealing temperature led to successful first steps for all five sequences. The original dsba construct was cloned into the pBSTG1 vector using HindIII and EcoRI (Figure 60D). The resultant pBSTG1-g3p* vector was then used to clone in the three new leader sequence fragments produced from the PCR strategy using HindIII and NheI (Figure 60E). Results of the PCR reactions are shown in Figure 61. The first PCR step was successful for all five leader sequences however the second PCR step was only successful for MalE, PhoA and M13KE. Attempts were made to resolve the problematic second PCR step for OmpA and STII leader sequences. This included using a wide range of annealing temperatures, supplementing the reaction with 1% glycerol in an attempt to disrupt any secondary structure of the long primers used, and lowering the magnesium concentration. However none of these attempts were successful. It was decided to press on with the set of four leader sequences to test including the original dsba vector.

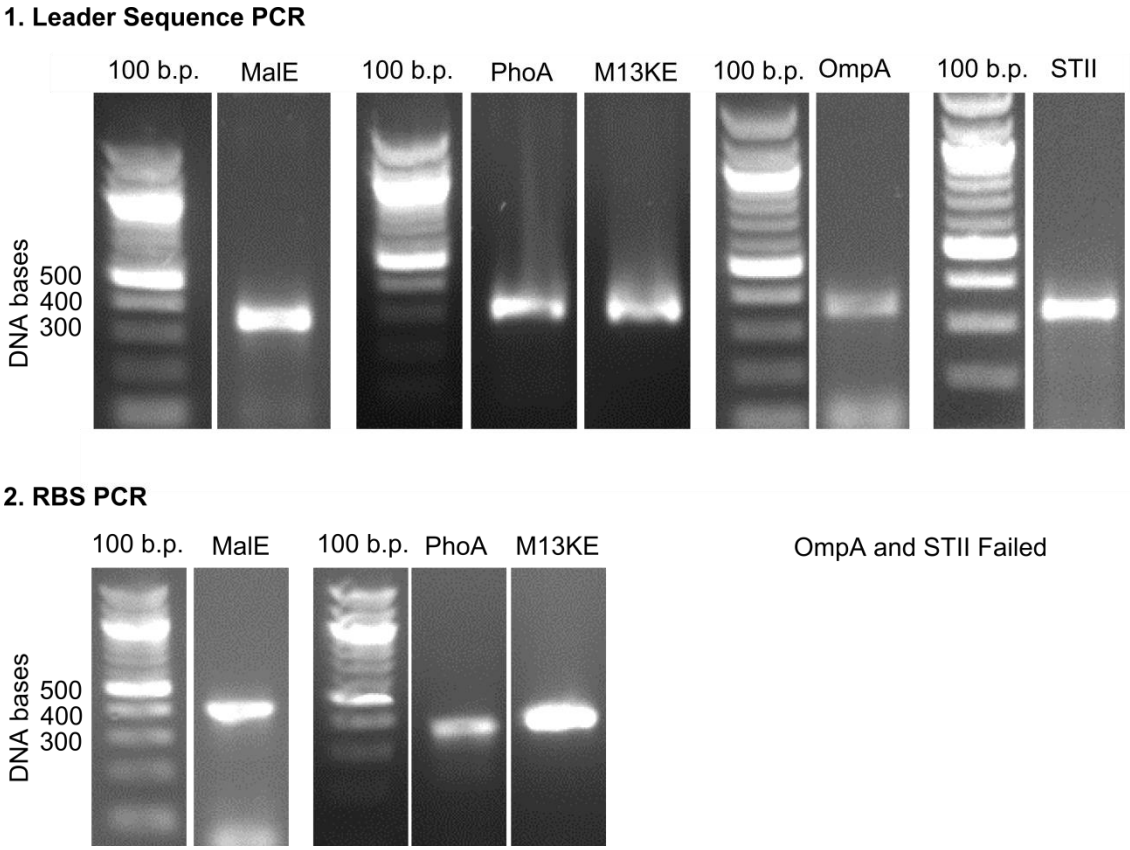


Figure 61: The results of the two-step PCR process to vary the leader sequence on the PK15-g3p*. The first step used primers which coded for the new leader sequences (MalE, PhoA, M13KE, OmpA and STII) and overlapped with the PK15 sequence, giving fragments around 350 base pairs in length. These were then extended in a second step with primers encoding the ribosome binding site and HindIII restriction site.

The four PK15-g3p* genes were cloned into the pBSTG1 vector using HindIII and NheI-HF. Sequences were confirmed by Sanger sequencing before the vectors were transformed into *E. coli* ER2738 cells. Cultures of each leader sequence variant were grown with IPTG induction and infected with either helper phage or hyperphage to produce phage overnight. Upon centrifugation to isolate phage it was observed that some of the hyperphage cultures had white cell debris in the pellet. After precipitation, centrifugation, resuspension and filtering, phage concentration was measured using UV absorption at 269 and 320 nm (Table 7). It appeared that the total phage produced by each combination of leader sequence and secondary phage method was approximately the same. The infectivity of the phage batches was tested via trial infections of ER cell cultures. 10^{10} phage produced only around 1 to 3

million infected bacteria when using the phage produced using helper phage. The phage derived from hyperphage had 100-fold less infectivity.

Table 7: Phage titres of different phage cultures

Secondary Phage	Leader Sequence	Phage Titre (25 mL culture)
Helper	Dsba	1.92×10^{13}
	MalE	4.62×10^{13}
	PhoA	3.20×10^{13}
	M13KE	8.60×10^{13}
Hyper	Dsba	2.61×10^{13}
	MalE	2.18×10^{13}
	PhoA	7.36×10^{13}
	M13KE	4.70×10^{13}

To test whether the peptide was indeed being displayed, it was necessary to cyclise the displayed peptide using TBMB to create the functional PK15 kallikrein inhibitor. To ensure sufficient levels of functional phage, one litre cultures of the dsba and PhoA variants were grown, with addition of helper phage (hyper phage was not an economically viable long-term option for litre-scale production). These cultures gave phage titres of an estimated 10^{15} phage per mL.

The litre cultures were modified with TBMB **10** following literature protocol.[82] Following precipitation, centrifugation and resuspension the phage titre was measured as around 10^{15} phage per mL. However in an ELISA assay to test binding of the phage to human plasma kallikrein there appeared to be no binding (Figure 62). Given that there was enough phage present to bind to the background of the plate with a strong signal, it was disappointing that the phage did not appear to bind to kallikrein at all. There were two possible reasons for this – either the peptide was not displayed at all, or the chemical modification failed. Attempts to examine the phage by SDS-PAGE were dominated by the g8p protein. The g8p protein is present in around 10^3-fold excess relative to the g3p. Due to difficulties in observing or isolating the other phage proteins from the SDS-PAGE gel this was abandoned.

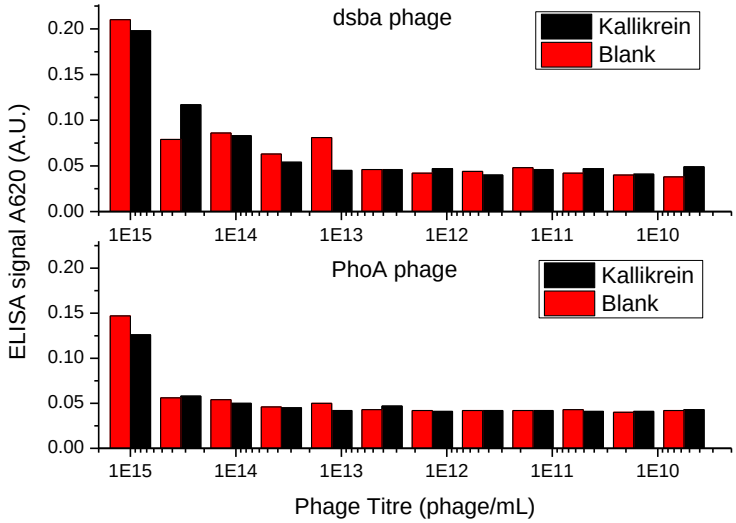


Figure 62: Phage ELISA results for binding of phage which should display the PK15 peptide modified by TBMB show no binding to the kallikrein target of PK15 with any signal being equal to that of the phage binding to the background.

4.3 Second Iteration Phage Studies

The failure of the first generation construct could not be attributed directly to either poor expression or failure of the chemical modification, therefore both of these factors were taken into account in designing a new construct. The new construct contained both an internal myc-tag for use in immunohistochemistry and a tobacco etch virus (TEV) cleavage site to allow release of any expressed peptide from the g3p conjugate. This TEV site was included with the aim of observing chemical modifications by mass spectrometry which is not possible on intact phage due to their large size. The peptide was to be displayed as a conjugate with a truncated variant of g3p consisting only of the C-terminal domain which is used in the Adhiron library in the McPherson and Tomlinson groups; helper phage would then supply infective WT g3p as before. Given the lack of difference in expression levels observed with different leader sequences in the first iteration, the dsba leader sequence was retained.

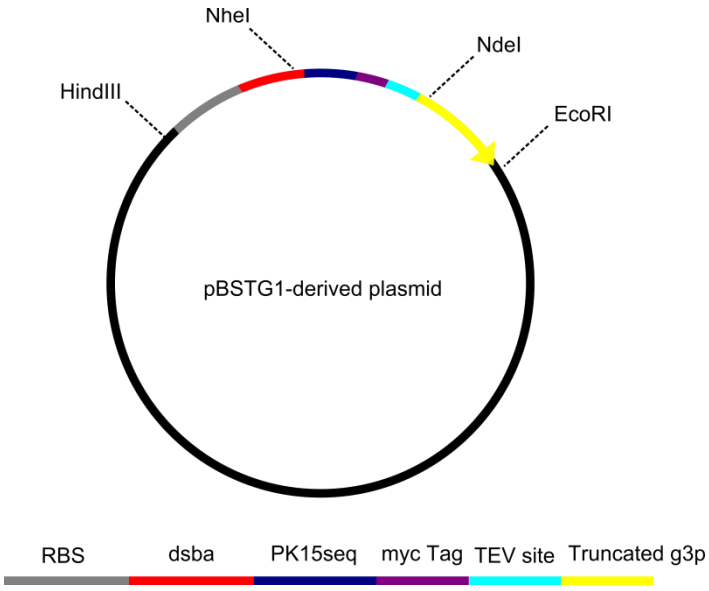


Figure 63: The second targeted vector containing the same PK15 sequence as previously but with additions of a myc-sequence for antibody detection, a TEV site for cleavage and MS studies and a truncated WT g3p variant consisting of the C-terminal domain.

The second generation construct (Figure 63) was supplied in puc57 (Genscript) and cloned into a pBSTG1 plasmid containing truncated g3p using NheI and NdeI. Successful ligation was confirmed by Sanger sequencing. Phage were produced using the standard BSTG protocol from ER cells which had been transformed with the second generation pBSTG1 construct. Helper phage were used to provide the remainder of the phage genome. Phage were precipitated, centrifuged and resuspended – it was found that the phage concentration

was low – in the order of 10^{11} phage per 8 mL culture. This is 10^{2}-10^{4} times lower than Adhiron-displaying phage using the same protocol.

The low levels of phage produced led to difficulty with detection of the displayed peptide by mass spec. The phage were concentrated in a centrifugal spin concentrator and incubated with TEV protease. Any cleavage products were then isolated via centrifugal concentration and a sample examined using high resolution ES-MS. However no signal corresponding to PK15 peptide was observed.

The phage was tested for presence of the myc epitope via an ELISA with anti-myc antibody as the target. No binding to the target was observed – suggesting either the peptide is not being expressed, or the truncated g3p is not being incorporated into the phage secreted from the *E. coli* cells. No binding to the background was observed either – not surprising due to the low phage titre used. It is possible that a Western blot with the anti-myc antibody would confirm the presence of the myc tag. However it was thought that due to inexperience with the technique a negative result would not be reliable, and if sufficient display was achieved the anti-myc phage ELISA should have given signal. Due to these disappointing results and impending time constraints further work on the phage display of the PK15 sequence was halted.

4.4 Discussion

It was clearly disappointing that attempts to display the PK15 peptide using phagemid display approaches failed. With the first generation construct it was difficult to assess whether the phage produced were displaying the PK15-g3p* fusion protein. Although phage appeared to be produced using all leader sequences, there was low infectivity with helper phage as well as with hyper phage. The method of assessing phage concentration was based only upon UV absorption of both protein and DNA – there is no indication whether phage is fully constructed or viable.

Chemical modification with TBMB was carried out following literature procedure but the phage did not appear to bind to the kallikrein target – with the first generation construct it was impossible to determine whether this was due to the lack of displayed peptide or due to failure of the chemical modification. The second generation construct also expressed poorly with low levels of phage produced. However the failure to bind to the anti-myc antibody suggests the level of displayed conjugate is either extremely low or non-existent. When using helper phage, phage will be produced which do not contain the display conjugate but which will be included in any concentrations assessed via UV absorbance – again this measure of phage concentration is at best a rough estimate.

The poor or non-existent expression was unexpected due to the successful expression of PK15-g3p* by Heinis and co-workers. They used a leader sequence very similar to M13KE, only differing in the last 3 amino acids. This leader sequence should have been included in the set of sequences chosen. It was considered that the expression may be failing due to inadvertent disulfide formation. There is a pair of cysteine residues in the C-terminal domain which may be disrupted by the PK15's free cysteines, this may particularly be of concern with the second construct used. It may be of use to express the PK15 sequence with the cysteines mutated to alanine, to test this hypothesis. The cell strain used for phage display was not the same in both pieces of work; in this work ER2738 cells were used whereas Heinis and co-workers use TG1. Therefore it is possible but unlikely that a lack of chaperones or transport machinery was responsible for the poor expression.

It is possible that the peptide is being successfully displayed but the myc binding was inhibited by the construct. The myc tag was an internal sequence however this should still be identified by the anti-myc antibody. It may be possible that a Western blot would identify the presence of the myc tag and have a lower detection limit, however if the peptide was expressing to sufficient levels for display purposes it would be expected that the ELISA protocol would be sufficient to detect the myc sequence. The TEV cleavage site was intended to be used to act as a secondary check, however the TEV-cleaved peptide could not

be detected by MS. Due to the low concentration of phage produced this is not unexpected; the concentration of peptide assuming full expression and cleavage was low nanomolar - below the recommended detection limit of the mass spectrometry setup although other researchers have reported signals at this concentration. Previous work by Chen *et al.* used a MALDI mass spectrometry approach[85] whereas here it was attempted to use an ES-MS setup including a column separation step.

It is difficult within the context of these results to assess whether the conjugate was being expressed at all. If a phagemid vector approach is retained there are a few potential options to further investigate this. First there may be an issue with the leader sequence used, although this has been used for the truncated g3p previously. Similar to the work with the first construct, the leader sequence could be varied to investigate phage production. Alternatively a reporter system other than the myc tag could be used. For example the display of fluorescent proteins including mCherry and green fluorescent protein has previously been achieved using filamentous phage[132] – the expression of a variable peptide conjugated to these could be assessed via phage solution fluorescence.

A whole-phage vector could be used as an alternative phage production system, as demonstrated by Winter, Heinis and co-workers. However if aiming to use the dehydroalanine chemistry then it is likely that a return to WT g3p or the truncated g3p should be feasible. It would be of interest to treat phage displaying these g3p variants with methyl 2,5-dibromovalerate and examine the effect on infectivity.

In conclusion the standard protocols for the BSTG1 plasmid phage production were co-opted from the expression of Adhiron-displaying phage to the display of a peptide and ultimately this may not have been the most appropriate approach. A great deal more work needs to be undertaken to investigate and optimise expression of the PK15 peptide using a phagemid approach.

Chapter 5: Screening of a non-antibody binding protein library against a small molecule target

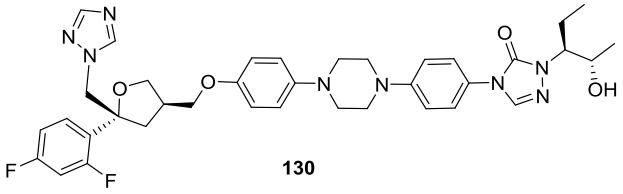
In this chapter the synthesis of a posaconazole biotin conjugate is presented alongside efforts to incorporate a disulfide linker. Biophysical analysis of the interaction between posaconazole and Adhirons identified in a screening process[viii] is presented. This involved creating further conjugates of posaconazole with both a PEG tail in attempts to increase solubility and then fluorescein to enable a fluorescence anisotropy assay to accurately characterise the interaction.

[viii] Screening carried out by Anna Tang and Darren Tomlinson, BSTG (FBS, University of Leeds)

5.1 Introduction to Adhiron diagnostics for small molecule detection

Antibodies most commonly bind to protein or peptide surfaces and can be tailored to have extremely high selectivity for small differences in molecular structure such as phosphorylation of individual amino acids.[133, 134] Binding of antibodies to small molecules is also feasible and can be used in diagnostics and perhaps even in therapeutic roles binding to signalling molecules or cofactors for example. Antibody screening is not without practical and ethical hurdles – often either requiring animal hosts to challenge with an antigen or *in vitro* display of antibody fragments via phage or yeast display.[133] In addition the detection of small molecules using an animal challenge requires the use of haptens – conjugates between the molecule and an inert protein such as albumin. Non-antibody binding proteins which are cheaper and more reliable to produce could potentially replace antibodies in many roles. Adhirons (Section 1.1) - a non-antibody protein binder family which contain two variable nine-residue loops - could be used in exactly the same manner and may be more suited to incorporation into diagnostics due to their increased stability.

It was aimed to screen the Adhiron library for binders to the small molecule antifungal drug posaconazole **130** (Figure 64). Posaconazole is variably absorbed by patients depending on many factors including diet, dosage routine and gut pH.[135] It is therefore an important diagnostics target, especially in the accurate and selective measurement of drug levels in infants. The viability of therapies has been shown to be reliant on the serum drug levels. Current testing regimes involve either microbiological assays or HPLC analysis. However the microbiological assays are often unsuitable in examining drugs used in a combination therapy – as is often the case for antifungal therapies. HPLC analysis requires specialised equipment and knowledge – leading to wait times for results if the equipment is available at all. A biosensor device which is specific for posaconazole would allow bedside monitoring specific for posaconazole, overcoming these associated problems. It was therefore aimed to screen the Adhiron library for binders to the small molecule antifungal drug posaconazole **130** (Figure 64). Posaconazole contains a secondary alcohol at one end of the molecule which was also a convenient synthetic handle for modification.



130

Figure 64: Posaconazole, the antifungal drug used as a target for Adhiron binders in this work

5.2 Synthesis of a posaconazole-biotin conjugate

To fix the target to the plate surface in an *in* vitro screening process biotinylation is often the method of choice. In order to screen for Adhiron binders to posaconazole it was necessary to create a posaconazole-biotin conjugate – due to the lack of amine or thiol groups in **129** common biotinylation reagents were not appropriate. These reagents are designed for protein biotinylation via NHS ester reaction with lysine residues or maleimide conjugation to thiols. A desirable feature of the linker between posaconazole and biotin would be a cleavable moiety such as a disulfide. This would allow elution of bound phage from the screening platform via disulfide reduction. Direct labelling of a small molecule with biotin may leave the molecule too close to the surface of the plate for the protein of the phage display library to bind efficiently. An extra linker is incorporated between the biotin and the small molecule to prevent this. It was also of interest to incorporate a linker containing a disulfide bond as this aids with the elution of the phage from the plate using a reducing agent such as DTT.

It was therefore aimed to synthesise the biotin linker **131** (Figure 65), containing a disulfide bond for cleavage and an acid functionality to react with the hydroxyl of posaconazole. The commercially-available linker **132** was purchased as a contingency. This can still be used in phage display screening assays using either pH change elution or a competition elution method rather than a disulfide reduction.

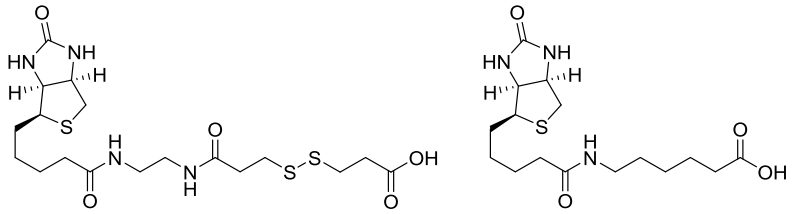


Figure 65: The two biotin linkers discussed in this section. Linker 45 is a novel target which would allow cleavage from a streptavidin plate using DTT. Linker 46 is commercially available but is reliant upon competition elution in the screening process.

The synthesis of **131** is shown in Figure 66. The first amide coupling had literature precedent but was found to be low yielding.[136] A switch to DMF as a solvent rather than the reported methanol/acetonitrile mixture improved the isolated yields, and heating the reaction increased this further, far beyond that achieved following the method reported in the literature. After Boc group deprotection using TFA, a second amide coupling was attempted. This appeared to give a small level of product by LCMS which also appeared to run on a tlc plate. However upon attempting to purify by silica column chromatography the product could not be isolated. When the column was washed with methanol/CH_2Cl_2/acetic acid

(10:88:2) 3,3′-dithiodipropionic acid was eluted. This suggests that the biotin conjugate is sticking to the column and the use of acetic acid initiated disulfide exchange but the biotin remained on the column. Since the previously synthesised Boc-protected biotin derivative **135** had been successfully purified by silica column chromatography it was suspected that the incorporation of the thiopropionic acid **137** was responsible for the retention on the column. Due to this it was anticipated that the biotinylated drugs with this linker may also be irreversibly retained on silica. Therefore focus switched to the simpler, commercially available biotin derivative **132**.

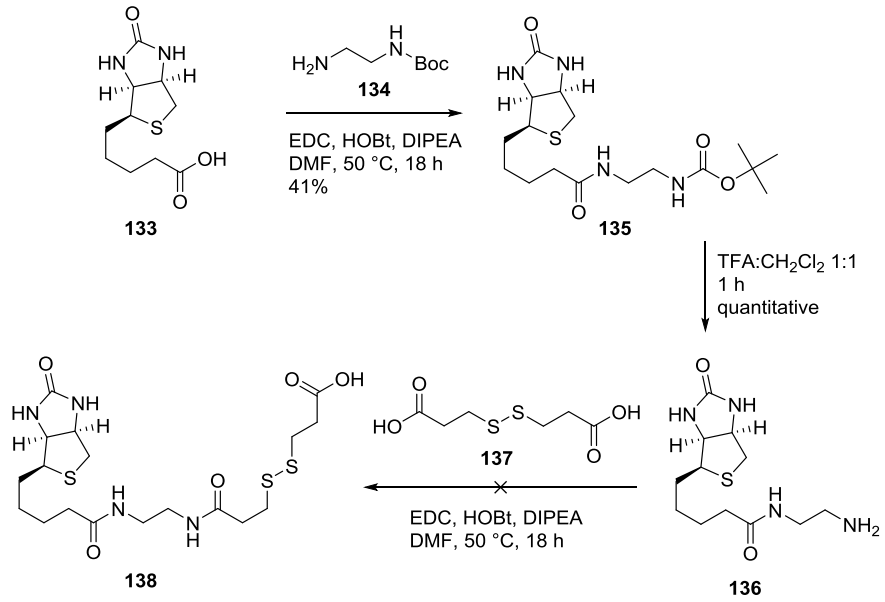


Figure 66: The attempted synthesis of biotin linker **138**. An amide coupling between biotin **133** and Boc protected ethylenediamine **134** yielded **135** which was then deprotected to yield **136**. However attempts to purify **138** failed and led to disulfide exchange regenerating 3,3'-dithiodipropionic acid **137**.

Posaconazole was extracted from the pharmaceutical Noxafil oral suspension, with all other suspension components remaining in the aqueous phase. **130** was successfully coupled to the biotin linker **132** using EDC-activated ester formation (Figure 67). The conjugate **139** was obtained in 18% yield (limiting reagent being the biotin derivative). This biotinylated molecule was purified by silica column chromatography unlike the previously discussed biotin linker **138**.

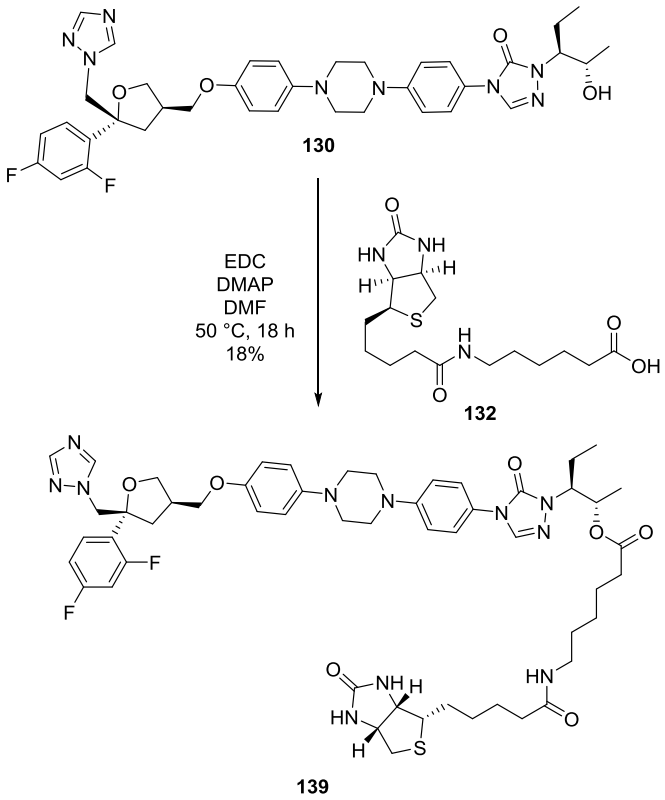


Figure 67: The biotinylation of posaconazole. EDC ester coupling was used to generate drug-biotin conjugate **139** in 18% yield.

5.3 Screening results from the Adhiron screen against posaconazole

The biotinylated posaconazole **139** was used in a phage display screen in the BioScreening Technology Group (BSTG) at the University of Leeds.[ix] Binders to biotinylated posaconazole **139** were selected from a phage-displayed library, using three iterative selection rounds to enrich the best binders. This process identified several binders from the Adhiron library of non-antibody binding proteins.

48 colonies expressing individual phage conjugates were cultured and phage isolated from the media. A binding ELISA of this phage expressing the Adhiron-g3p conjugates bound to biotinylated posaconazole is shown in Figure 68 which demonstrated that the binders were binding to the posaconazole rather than the plate background. Phage were incubated with the target before washing off unbound phage and treatment with an anti-phage antibody conjugated to horseradish peroxidase. HRP oxidation of 3,3′,5,5′-tetramethylbenzidine is then quantified to estimate the levels of bound phage. 16 of the hits were sent for sequencing.

[ix] Screening carried out by Darren Tomlinson and Anna Tang

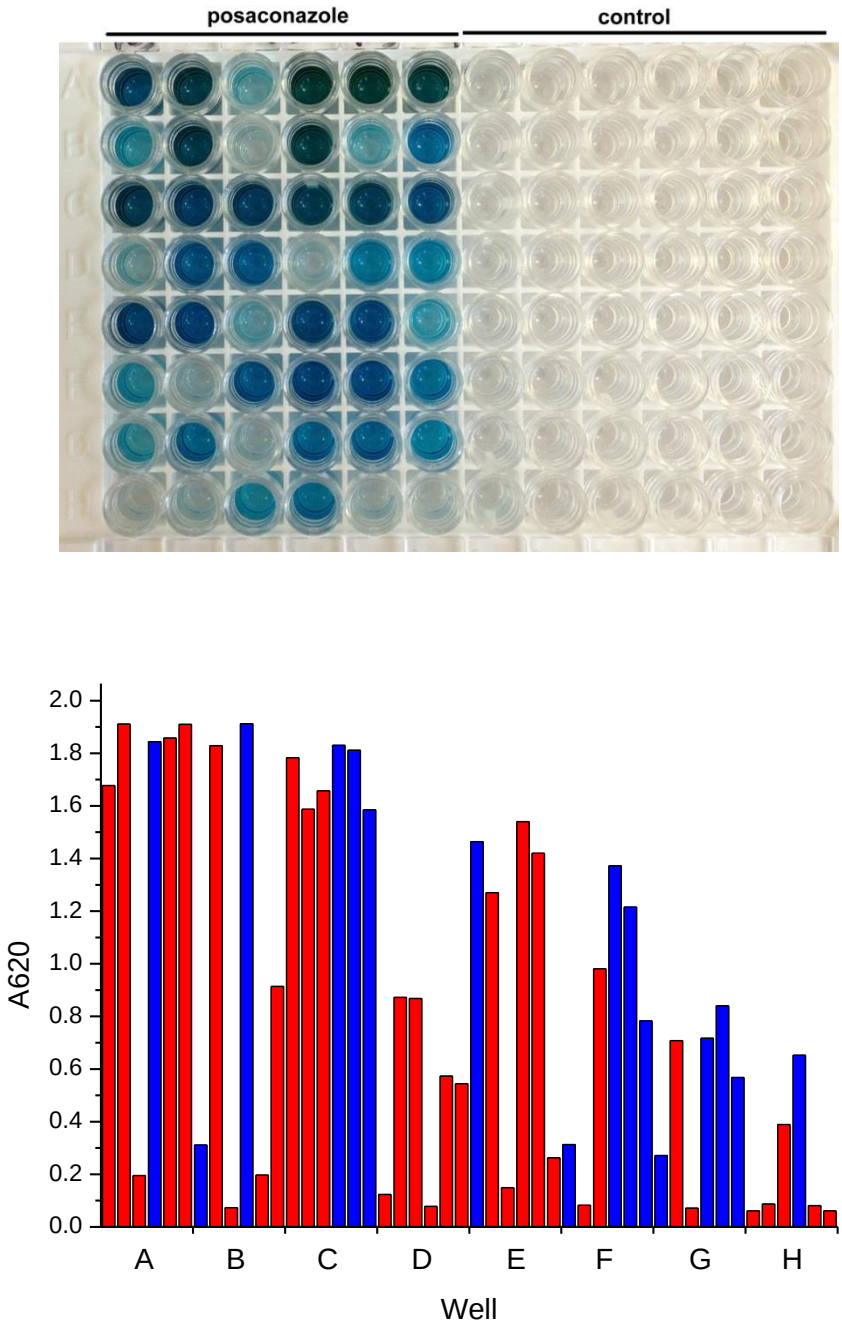


Figure 68: The results of a phage ELISA binding test against biotinylated posaconazole. Wells on the left half have been precoated with biotinylated posaconazole, the right half are blank negative controls. Phage displaying the same protein sequence are added to the same relative well in left and right halves. An anti-phage antibody and secondary antibody conjugated with horseradish peroxidase gives readout of the binding affinity of the phage-displayed protein. The more intense the blue colour, the stronger the binding event between protein and posaconazole is. Several different strength binders have been identified which were to be further investigated for binding affinity quantification and physical properties. Quantification of the ELISA signal is shown in the bar chart (blue bars represent hits which were sent for sequencing)

The sequencing results are shown in Table 8. A4 was the predominant hit – the same sequence was returned from 8 other clones out of the 16 tested. F5 also shares a large sequence similarity with A4 in both loops – suggesting they would bind in very similar orientations. Outside of these two sequences, no further consensus sequence can be seen but it is evident that both loops contain a high proportion of aromatic and hydrophobic residues. This is not surprising as the posaconazole has several aromatic rings and very few groups capable of hydrogen bonding. It was interesting that two of the hits have truncated second loops – this is not uncommon in Adhiron screens and is an artefact of library production. It does suggest that in some cases the Adhiron may rely on only one loop to bind the posaconazole. Loop two in particular is very proline rich – perhaps these influence the loop structure in an important manner.

Table 8: The variable loop sequences of the three possible posaconazole binders identified from a phage-displayed Adhiron library.

Well	N	Loop 1	Loop 2
A4	9	YWVVQWGMD	WAREWPSFH
B1	1	MAQWMQYYS	Truncated (TNRG ΔKE)
E1	1	HRWDPYNWW	YEWLGGPIM
F1	1	HTLVLEDRW	WPPFSFGYL
F5	1	VRVVQWTMD	SWREWPSFM
G1	1	QMQYQSYLS	Truncated (QHQT ΔKE)
H4	2	WQQHNFFIY	WYQVTPPYW

N = Occurrences in sequencing results (Total N=16)

In most screening procedures successful binders typically have their interactions characterised biophysically by techniques including ITC, SPR and bio-layer interferometry. Therefore it was aimed to characterise the posaconazole binders' interactions. Three of the Adhiron sequences were subcloned (A4/E1/H4) and expressed by the BSTG. Initial characterisation of the Adhirons' interactions with posaconazole was not possible due to the drug's poor solubility. While use of DMSO increases posaconazole solubility, the Adhirons do not tolerate large percentages of DMSO – typically 5 to 10% maximum. In addition the small size of posaconazole precludes its use in SPR or biolayer inferometry without further derivatisation to facilitate adhesion to the chip surface. Therefore it was decided to investigate the synthesis of further posaconazole conjugates which would enable biophysical characterisation via either improvement in solubility or production of a fluorescent reporter molecule.

5.4 Synthesis of further posaconazole conjugates for biophysical characterisation of the interaction between Adhiron hits and posaconazole

5.4.1 Synthesis of a posaconazole-PEG conjugate for ITC studies

Given that it was likely the Adhirons were binding at the opposite end of the posaconazole from the site of biotinylation, it was decided to use the same point for further derivatisation. It was first aimed to create a conjugate with the addition of a PEG chain to increase aqueous solubility and facilitate ITC characterisation of the interaction with the Adhiron. The PEG acid **140** was used as the azide functionality would serve as a convenient handle for any further modification. This subunit was attached to posaconazole via the same EDC ester formation as in synthesis of **139** above.

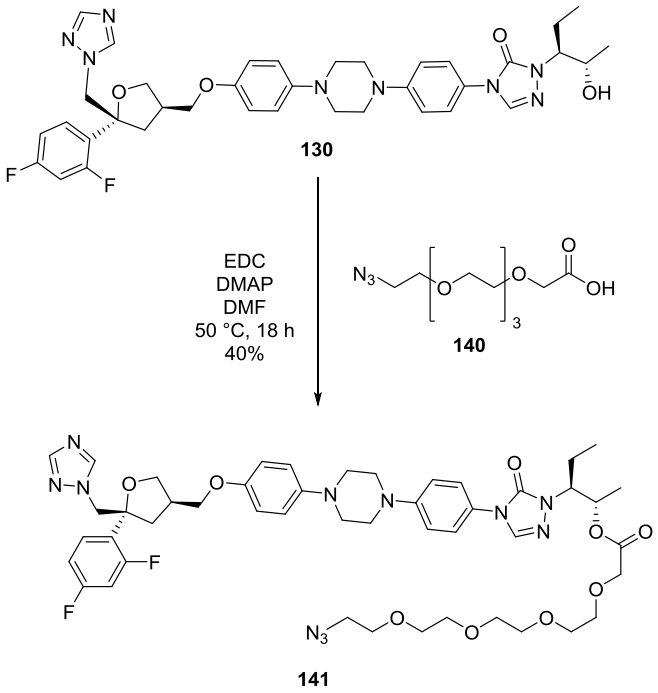


Figure 69: The synthesis of PEGylated posaconazole **141** via EDC ester formation between posaconazole **130** and the azido-PEG derivative **140**.

The posaconazole-PEG conjugate **141** was still not entirely soluble at high µM concentration in 5% DMSO. This prevented ITC analysis due to domination of the signal in the instrument by the solvation energy upon dilution, although as will become apparent in section 5.4.2 it is also possible that the ITC was carried out at concentrations too low to detect the interaction.

5.4.2 Synthesis of a posaconazole-fluorescein conjugate for fluorescence anisotropy studies

Due to the failure of ITC experiments it was decided to use fluorescence anisotropy (FA) as an alternative assay. It was aimed to synthesise a conjugate of posaconazole attached to a fluorophore to allow analysis of the binding via FA. FA was deemed an appropriate reporter for this interaction as the fluorescently labelled posaconazole probe could be used at low concentration (100 nM), thus avoiding the solubility problems. The difference in size between drug and Adhiron should facilitate an observable change in the anisotropy upon binding.

The inclusion of the azide moiety on the PEG chain proved advantageous as it allowed further derivatisation. In order to attach fluorescein to the posaconazole it was aimed to use a copper catalysed azide-alkyne cyclisation. This necessitated the synthesis of **143** via reaction of fluorescein isothiocyanate (FITC) **142** with propargyl amine. Initially this was attempted in the presence of triethylamine and although some product was formed it appeared that the thiourea may be cleaving to release fluorescein. Repeating the reaction in DMSO and without any additional base, the reaction proceeded in just 5 minutes. Purification by HPLC gave **143** in a 30% yield giving enough material to perform the cycloaddition reaction.

The copper-catalysed reaction between **143** and **141** formed a triazole as expected, presumably in the 1,4-substitution common to the Cu(I) catalysed Huisgen cycloaddition. In an unexpected development the thiourea was also converted to a urea, presumably due to the presence of ascorbic acid – this conversion is well precedented in acidic conditions,[137] although not with ascorbic acid.

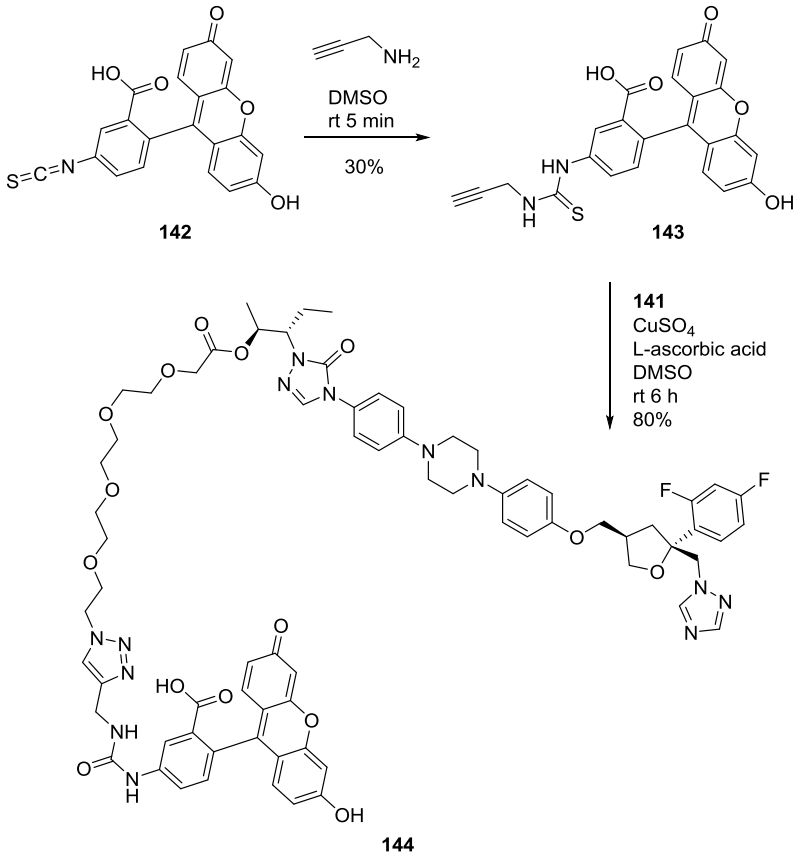


Figure 70: The synthesis of a posaconazole-fluorescein conjugate **144** via a copper-catalysed cycloaddition reaction between **141** and the fluorescein-derived alkyne **143**.

144 was used in a fluorescence anisotropy assay with the three subcloned Adhiron binders. Binder A4C gave the highest quality data (Figure 71), and binds to **144** with a K_d of 158 ± 44 µM. Binders E1C and H4C proved weaker binders and were not observed to bind within the concentration range tested. Binder E1C appeared to precipitate under concentration in a centrifugal spin concentrator – the addition of BME to the buffer did not seem to alleviate this. Although the concentration of the soluble fraction was measured and used in the assay it is likely this was not a homogenous protein sample. The binding of A4C was weaker than the typical affinities of Adhiron binders for protein targets which have been measured as low nanomolar affinity.

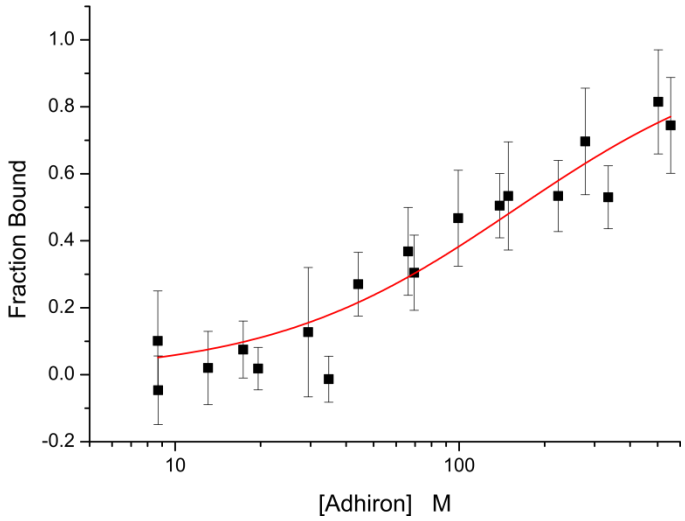


Figure 71: Binding of Adhiron A4C to posaconazole derivative **144** – as observed by fluorescence anisotropy measurements.

5.5 Discussion

The direct labelling of posaconazole via ester formation enabled screening to identify a binder from a non-antibody binding library of Adhirons. This direct labelling avoided the need to display the molecule on the surface of a protein, as is often required for antibody screening. The use of **132** should be transferable to the labelling of other small molecules bearing either amine or hydroxyl functionality.

Further synthetic derivitisation of posaconazole **130** then allowed characterisation of the binding between posaconazole and the Adhiron binder A4C. The labelling with a fluorophore and subsequent use of fluorescence anisotropy avoided allowed confirmation that the Adhiron was binding posaconazole regardless of which linker was attached to the alcohol. It is possible that the biotinylated posaconazole **139** could have enabled this characterisation via SPR – the addition of **139** to a streptavidin-coated surface would then allow observation of the binding of the adhiron to the chip. The observed anisotropy was relatively small ranging from 0.05 to 0.08 compared to the maximum obtainable with bound fluorescein (0.37).[138] The small change in anisotropy observed is probably due to the flexibility of the PEG linker and the large distance between where the Adhiron is assumed to bind and where the fluorescein is attached.

It was encouraging that A4C was the hit most-enriched in the phage library and also the best binder of those tested – in both the phage ELISA and the FA assay. The binding affinity was disappointingly low – Adhirons have previously been found to bind to protein targets with nanomolar affinity.[8] Closer examination of the posaconazole structure shows it has no

hydrogen bond donors and only a few acceptors. The majority of interactions will be hydrophobic, including interactions with the many aromatic rings. It is therefore no surprise that Adhiron A4C has hydrophobic loops. The hydropathicity of the loop regions as determined using protparam are shown in Table 9. Loop 1 of Adhiron A4C is particularly hydrophobic which may explain the better binding affinity to posaconazole.

Table 9: Hydropathicity scores as calculated using ProtParam.[139]

Adhiron	Loop 1	Score	Loop 2	Score
A4C	YWVVQWGMD	-0.022	WAREWPSFH	-1.2
E1C	HRWDPYNWW	-2.25	YEWLGGPIM	0.23
H4C	WQQHNFFIY	-0.64	WYQVTPPYW	-0.844

Whether this binding affinity is sufficient for biosensor use would require further investigation. An Adhiron which binds to an anti-myc antibody with nanomolar affinity (K_d = 360 ± 10 nM) was incorporated into an electrochemical biosensor and measured levels of target in the pM range, saturating above 330 pM.[140] The recommended treatment levels of posaconazole in adults is 1 µM serum concentration. Therefore for a similar set-up to the proof-of-principle electrochemical biosensor above, the weaker binding of A4C may be advantageous, although the proof-of-principle device did not measure serum samples. If a higher affinity Adhiron is required a focussed library could be produced based on the loop regions in A4C. Due to the low affinity of hits, it is probable that a focussed approach may yield better results than repeating the phage display process from the completely randomised library.

Chapter 6: Conclusions and Future Work

6.1 Conclusions

The synthetic methods developed in this body of work broaden the diversity of accessible bicyclic peptide structures. Starting from one linear amino acid sequence, two chemical methodologies were developed which lead to bicyclic peptides with significantly different structures.

It was first demonstrated that methyl 2,5-dibromovalerate **63** can cleanly convert multiple cysteine residues to dehydroalanine (Chapter 2). This avoids the formation of cyclic by products present when other double-alkylation elimination reagents are used and also requires only native amino acids and mild reaction conditions to perform the transformation. The dehydroalanine-forming step was incorporated into a two-step strategy to access a diverse subset of bicyclic peptide structures, exemplified via the one-pot synthesis of a library of stereoisomers **40** of kallikrein inhibitor PK15. Although none of the individual stereoisomers were more potent than the LLL-PK15 **37**, the spread of IC_{50} values over at least four orders of magnitude was validation that altering the stereochemistry of these peptides had a significant effect on their structure.

The development of a new core based on Barbas reagent (Chapter 3) should give access to bicyclic peptides which are distinct in structure when compared to the same sequence cyclised with established cores. It was aimed to widen the range of bicyclic peptide scaffolds which can be accessed by dehydroalanine cyclisation with a polythiol. The synthesis of polythiols proved problematic but the intermediates **102**, **107** and **112** produced require only deprotection. Further work into the development of these deprotection steps and preventing polymerisation of cores could provide alkyl polythiols which cyclise the dehydroalanine-containing peptides.

The power of phage display libraries was demonstrated via the Adhiron screening for binders to posaconazole (Chapter 5). Synthetic modification of posaconazole enabled both its use in *in vitro* screening and the biophysical characterisation of the interaction between posaconazole and the best binder Adhiron A4C. Attempts to set up a phage display platform for library display were ultimately unsuccessful (Chapter 4) however if these library display problems can be overcome then bicyclic peptides could be used in a variety of potential applications. The following sections present suggested future directions for both the chemical methods developed and the use of bicyclic peptides in general.

6.2 Expanding the scope of dehydroalanine formation with methyl 2,5-dibromovalerate

6.2.1 Modifying expressed peptides and proteins

In Chapter 2 an example use of peptides containing multiple dehydroalanines was presented in which a trithiol core cyclised the three dehydroalanine positions, scrambling the stereochemistry at each alpha carbon and creating a diverse library of peptide structures. The same cyclisation chemistry could be used on a library scale, with hits from a screening procedure being deconvoluted by individual synthesis of peptides containing D- or L-cysteine and cyclisation with TBMB as was carried out in Section 2.4.1. It has been demonstrated that combining a binding loop with a poly-arginine loop can improve cell penetration.[141] It is possible that combining a known binding epitope with a polyarginine loop could create potent and cell penetrant compounds. However since the introduction of the polyarginine loop may lead to altered structure in the epitope and lower affinity it might be of use to generate the D- and L- stereosiomers of the peptide structure to identify an improved inhibitor. The dehydroalanine chemistry could be used on a small focussed library using the epitope as a starting point. Hits can then be deconvoluted by individual SPPS and retesting.

One area where the use of methyl 2,5-dibromovalerate **63** can be applied is in the chemical modification of proteins. Dibromoadipamide **27** has become the standard for dehydroalanine conversion from cysteine but **63** is perhaps more suitable to this transformation due to its increased solubility. Preliminary work into using **63** to perform this transformation is taking place in the Berry research group at Leeds with positive results so far. The valerate compound produces a cleaner reaction profile than the adipamide compound, is more soluble and requires fewer equivalents of reagent to perform the transformation.[x]

6.2.2 Using methyl 2,5-dibromovalerate to create Dha and Dbu in lanthipeptides

However perhaps the most interesting and useful area where **63** could be applied to insert multiple dehydroalanine residues is in lanthipeptide research (Section 1.2.3). The solid-phase synthesis of entire peptides or distinct rings of the lanthipeptides allows understanding of their properties and investigation of their structure and binding to targets. The substrates of known cyclases could be synthesised and then cyclised *in vitro*, this would be particularly useful for the study of such cyclases when the dehydratase is unknown or not available for *in vitro* studies. For example a photocaged cysteine could be combined with the free cysteine

[x] Robert Smith, personal communication

which is converted to Dha. Uncaging would then allow temporal control of the lanthionine formation.

Preliminary attempts to synthesise Dha on resin via the double alkylation-elimination of cysteine are described below. In addition it would be of interest to synthesise the thiol amino acid **149** – a threonine analogue. This should in theory undergo the same double-alkylation elimination upon reaction with **63** to give access to dehydrobutyrine. This can then be crosslinked with cysteine to form methyllanthionine.[121] The amino acid **149** was synthesised by Gellman and co-workers[142] via the aziridine **148** which had been synthesised previously by Vederas and co-workers.[143] Allyl deprotection of **149** to form **150** would access a subunit which could be directly used in an Fmoc synthesis strategy, alternatively experiments on the amino acid conversion could first be facilitated by deprotection of the trityl protecting group of **149**.

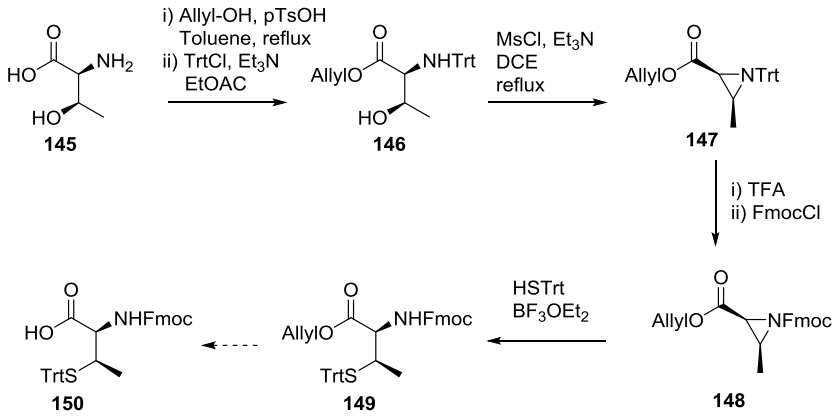


Figure 72: The synthesis of protected thiol amino acid **150** could be achieved via allyl deprotection of **149** which has previously been synthesised via ring opening of the aziridine **148**.[142, 143]

6.2.3 Preliminary attempts to synthesise Dha on resin

With a new reagent for dehydroalanine in hand it was decided to test the scope of the reagent through the conversion of a cysteine in a peptide attached to solid phase resin. A preliminary investigation was undertaken with the solid phase synthesis of the peptide **151** on chlorotrityl resin. The use of differential protecting groups was designed to allow selective deprotection of the tBS protecting group from one cysteine while leaving the other protected to give **152**. Samples of the resin were isolated and cleaved with either TFA or hexafluoroisopropanol to monitor the progress of reactions by LCMS.

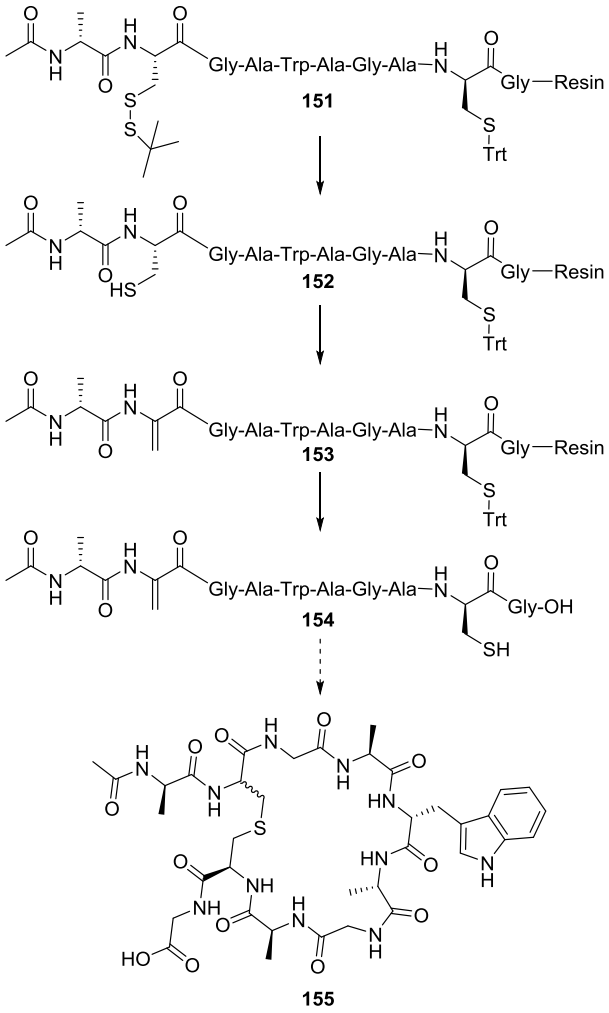


Figure 73: The attempted synthesis of cyclic peptide **155** via lanthionine formation on-resin. Attempts were hindered by inefficient deprotection of the tBS protecting group.

Synthesis of the peptide **151** proceeded successfully, although the coupling of the Cys(tBuS) amino acid proved less efficient than the other Fmoc amino acids as the peptide Ac-AGAWAGASG-OH was observed as a major by-product. Deprotection of the tBS protecting group was attempted with a variety of reagents with limited success (Table 10) and full deprotection to yield **152** could not be achieved. As a proof of principle **63** was applied to the resin and conversion to dehydroalanine in **153** was observed by MS. It was difficult to estimate the success of conversion due to both the presence of the truncated peptide impurity and the extensive attempts at deprotection leading to several by-products and traces of reagents being present.

Table 10: Conditions which were used in the attempted deprotection of the tBS group on peptide **151**.

Reagents	**Conditions**	**Observations**
βME (100 eq), DMF	rt, 18 h	Low levels of deprotection
PBu_3 (1.1 eq), DMF:H_2O 99:1	rt, 1 h	No further deprotection
PBu_3 (10 eq), DMF:H_2O:DCM 2:1:2	rt, 18 h	No further deprotection
DTT (10 eq), NEt_3, DMF:H_2O 3:1	rt, 16 h	Low levels of deprotection
DTT, K_2CO_3, DMF:H_2O 1:1	37 °C, 2 h	Overswollen resin/insoluble K_2CO_3

Due to time constraints this work was not repeated. In a future iteration it would be interesting to replace the tBS protecting group with a methoxytrityl group which should selectively deprotect using hexafluoroisopropanol. Resynthesis of the peptide on a non-trityl resin such as Rink amide would then allow reaction monitoring with TFA cleavage.

6.3 Further research with the core molecules

6.3.1 Utilising the Barbas-type cores

The trifunctional core **110** based on Barbas reagent has been demonstrated to cyclise a cysteine-containing peptide efficiently. This core could be applied directly to the phage display approach developed by Heinis and co-workers and would produce peptides of different backbone structure to those currently accessible. The problems with oxidation of this core may actually prove useful – it may be possible that a peptide wherein cysteines are attached to the 1,3,4-oxadiazole could be treated with oxidant, oxidising cysteines and eliminating them to form the oxadiazolone and free sulfinic acids. The core could also be used as a peptide stapling reagent – the third 2-mesyl-1,3,4-oxadiazole can be used as a functional handle for dyes or conjugation to other molecules.

6.3.2 Alternative thiol-based cores

There are other cores which could potentially be synthesised which could cyclise the dehydroalanine-containing peptides appropriately. For example although the aim of the tetrahedral cores was to attach a cargo, cores such as **156** and **157** might also be of interest. Recently the groups of Derda and Heinis independently developed light-responsive peptide ligands wherein an azobenzene derivative was used to cyclise a peptide.[80, 81] This allowed photoswitching to moderate the binding of the peptide. Similar cores could be created with more than two attachment points such as **158**.

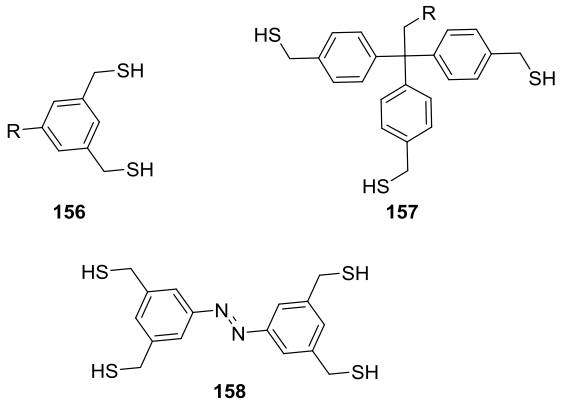


Figure 74: Potential cores based on TMMB which might expand the applications of constrained peptides.

6.4 Expanding the roles of bicyclic peptides

Bicyclic peptides have currently been used individually for therapeutics – they are in theory easily adaptable to uses in diagnostics or as molecular tools such as imaging reagents due to their specificity of binding. However there are many possible applications outside of these uses. The use of chemical cyclisation allows an on-switch event for binding. For example the expression of a bicyclic peptide sequence on the terminus of its own binding partner may then allow controllable protein polymerisation via addition of the cyclisation reagent. An interesting avenue may be the use of bicylic peptides to form interfaces in heterogenous supramolecular protein structures for example in artificial capsids or nanotubes. This would allow tunable structure formation and avoid the need for complex interface engineering. It may also be interesting to create conjugates of different bicyclic peptides linked by a flexible linker - a double bicyclic peptide could be used as a polymerisation reagent acting as the glue between two protein interfaces.

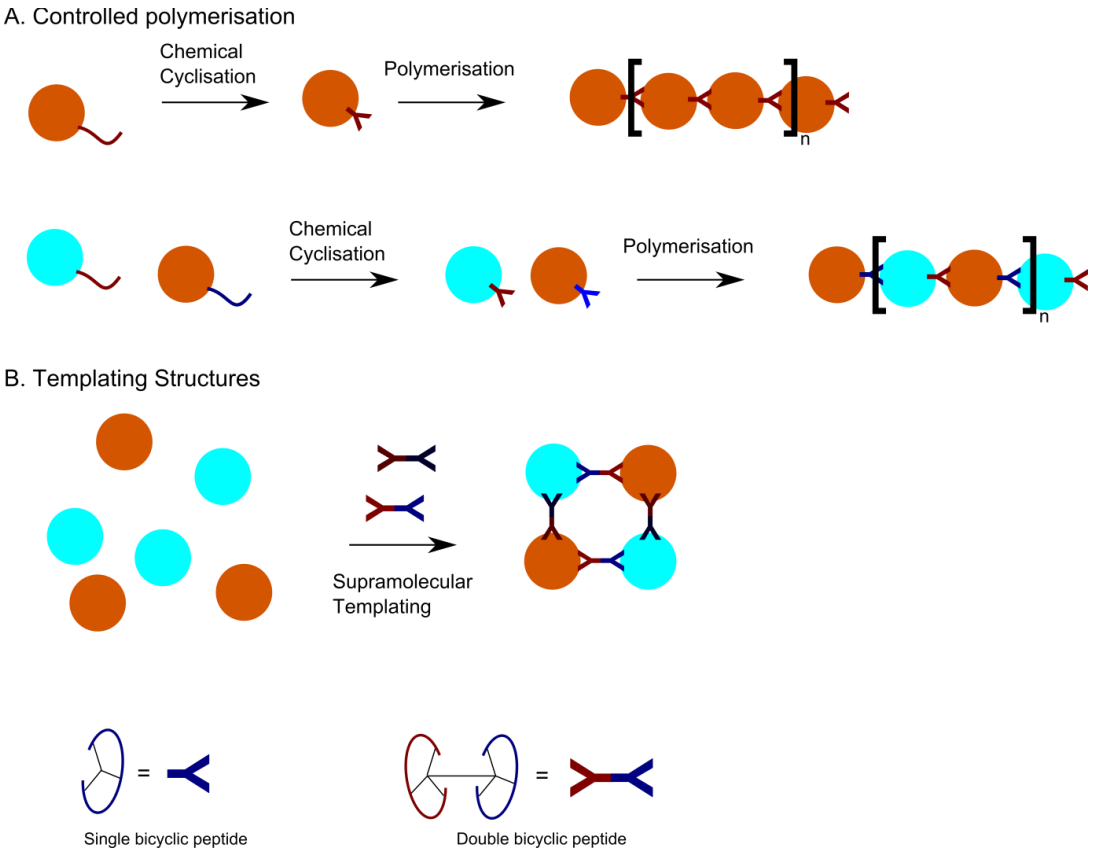


Figure 75: Potential supramolecular structures which could use bicyclic peptides as a controllable on-switch. A. Expression of the peptide sequence as a terminal extension would then allow the cyclisation reaction to control the polymerisation event. B. The combination of multiple binding sites with specific bicyclic peptides acting as interface junctions could template supramolecular structures.

6.5 Effective display of peptide libraries for chemical modification

The key step which was unsuccessful in this work was the creation of a library display platform. Successful expression of the g3p-peptide conjugate may deliver an effective display platform such as that used by Heinis and co-workers. It is optimistic to assume that the chemical modification to form dehydroalanine and cyclise with tris(mercaptomethyl)benzene would not alter the phage viability but this cannot be ruled out and merits further investigation. Likewise the use of a g8p fusion does not guarantee that chemical modification does not affect the WT g3p protein.

It is plausible that this two-step chemical modification strategy is more suited to a less complex display platform. Phage contain many substituent parts and the display process relies on successful reinfection of *E. coli*. If it can be demonstrated that the chemical modification does not modify nucleic acid bases it is likely that a shift to mRNA display would be advantageous. mRNA display allows purification of expressed components and their RNA partners and post-screening only an *in vitro* amplification is required to generate the new library. In addition the translation machinery is tunable to incorporate artificial amino acids more easily than when phage is used.

Alternatively a solid-phase synthesis based approach could be used to create a synthetic library in the same manner as Pei and co-workers used to identify **14**.[62] Although in theory diversity is limited by the synthetic resources – specifically the handling of the resin (Pei and co-workers estimated of a theoretical library of 10^{13} they synthesised around 10^{7} sequences). With advances in synthetic technologies, microscale synthesis and screening methods the library size may be increased. In addition second and third iteration libraries can be built which improve upon initial hits in a focussed screening approach.

Chapter 7: Experimental Details

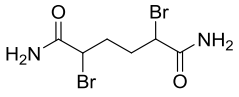
7.1 Reagents and Equipment

Unless otherwise stated, all reagents were purchased from Sigma Aldrich, Alfa Aesar, or Fisher Scientific and were used without further purification. Human plasma kallikrein was purchased from Molecular Innovations (HPKA-MIN), Z-Phe-Arg-AMC was purchased from Cambridge Bioscience.

NMR data were collected using a Bruker Avance 500, Bruker DRX500, or Bruker DPX300 and analysed using MestReNova software. IR spectra were recorded using a Bruker Platinum-ATR. High resolution mass spectrometry was carried out on a Bruker maXis Impact or on a Bruker Daltonics micrOTOF using electrospray ionisation. Isotopic distributions in routine mass spectra were as expected. Liquid Chromatography Mass Spectrometry (LCMS) was performed on an Agilent 1200 series LC system comprising a Bruker HCT Ultra ion trap mass spectrometer and a Phenomenex Luna C18 50 × 2 mm column (5 µm particle size) using a gradient of 5–90% MeCN over 1.8 min. ES-MS data were obtained using a Bruker MaXis Impact Mass Spectrometer. HPLC analysis was performed on an Agilent 1290 infinity LC using an Ascentis® Express Peptide ES-C18 column (10 cm × 2. 1 mm, 2.7 µM particle size). Mixtures of solvents, such as those used in column chromatography, are *v/v* and all column chromatography was carried out using Geduran Si 60 silica gel. Unless otherwise stated all non-aqueous reactions were performed under an atmosphere of nitrogen.

7.2 Chemical synthesis

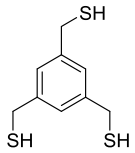
2,5-dibromoadipamide (27) [103]



2,5-dibromoadipamide was synthesised following a literature procedure.[103] Adipic acid (12.5 g, 85.6 mmol) was dissolved in $SOCl_2$ (37.5 mL), heated to 80 °C under reflux open to the air for 90 min then cooled to rt. Carbon tetrachloride (50 mL) was added followed by

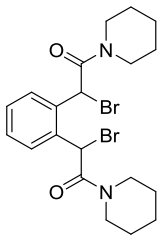
NBS (34 g, 191 mmol) (NBS freshly recrystallized from H_2O and dried). With vigorous stirring HBr (aq) (48 %, 5 drops) were added and the reaction heated to 80 °C. The reaction turned from red to black over the course of 2.5 h at which point it was cooled to rt and then to 0 °C. The precipitated solid was removed by filtration and the flask washed with ether (50 mL). The filtrate was concentrated *in vacuo* to give the crude acid chloride as a dark red viscous liquid. Ammonium hydroxide solution (25%, 140 mL) was cooled to 0 °C and the crude acid chloride was added dropwise, turning the solution blue. The reaction was stirred for 1 h at 0 °C before the blue solid was isolated by filtration. The solid was suspended in MeOH–H_2O (1:1, 100 mL) and heated to 60 °C for 30 min. Filtration and washing with methanol (100 mL) gave the product **27** (11.36 g, 44%), a mixture of diastereomers, as a grey solid. Characterisation was in line with the literature.[103] δ_H (500 MHz, DMSO-d_6): 7.70 (2H, s, NH_2), 7.32 (2H, s, NH_2), 4.37-4.32 (2H, m, BrCH), 2.10-1.79 (4H, m, CH_2-CH_2); *m/z* (ES): Found M+Na^+ 322.900256, expected for [$C_6H_{10}O_2N_2{}^{79}Br_2{}^+$ Na] 322.900122.

1,3,5-tris(mercaptomethyl)benzene (39)



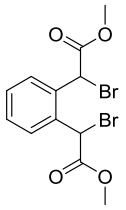
1,3,5-tris(bromomethyl)benzene (205 mg, 0.575 mmol) and thiourea (230 mg, 3 mmol) were combined and THF (5 mL) added before stirring at 75 °C for 4 h. After cooling to rt, potassium hydroxide solution (0.68 M, 5 mL) was added and the reaction stirred at 105 °C for 17 h. After cooling to rt the reaction mixture was concentrated under reduced pressure before addition of ethyl acetate (20 mL). The mixture was extracted with sodium hydroxide solution (0.1 M, 2 × 20 mL), aqueous layers were combined and acidified (pH 1) with hydrochloric acid (1M) before extraction with ethyl acetate (2 × 30 mL). Organic fractions were combined, dried ($MgSO_4$), filtered and concentrated under reduced pressure. The product was further purified by column chromatography (1:1 ethyl acetate:hexane) to yield **39** (72 mg, 0.33 mmol, 58%) as a colourless oil. δ_H (300 MHz, DMSO-d_6): 7.18 (3H, s, benzene CH), 3.70 (6H, d, *J* 7.0, benzyl CH_2), 2.84 (3H, t, *J* 7.0, SH); δ_C (75 MHz, d_6-DMSO): 142.2 (benzene C_q), 126.8 (benzene CH), 28.0 (benzyl CH_2); ν_{max}/cm^{-1} (film) 2549 (SH); *m/z* (ES negative): Found [M−H] 215.002092, expected for [$C_9H_{11}S_3$] 215.002837.

1,1'-[1,2-phenylenebis(2-bromo-1-oxo-2,1-ethanediyl)]bispiperidine (52)



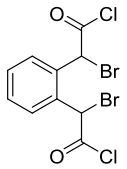
A solution of the acid chloride intermediate **56** (approx. 5.15 mmol) in CH_2Cl_2 (20 mL) and CCl_4 (3 mL) was cooled to 0 °C before the dropwise addition of piperidine (3.2 mL, 31.4 mmol) over 15 min. After stirring for 45 min total the reaction was concentrated under reduced pressure, dry-loaded using Celite and purified by silica column chromatography (ethyl acetate:hexane gradient, 1:4 to 1:1) to yield **52**. Upon drying under vacuum the product degraded.

Dimethyl α^1,α^2-dibromo-1,2-phenylenediacetate (53)



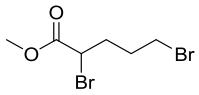
A solution of the acid chloride intermediate **56** (approx. 5.15 mmol) in CH_2Cl_2 (15 mL) and CCl_4 (3 mL) was cooled to 0 °C before the dropwise addition of methanol (5 mL) in CH_2Cl_2 (20 mL) over 15 min. After stirring for 30 min total the reaction was concentrated under reduced pressure, dry-loaded using Celite and purified by silica column chromatography (ethyl acetate:hexane gradient, 1:9 to 1:3) to yield **53** (1.858 g, 4.89 mmol, 95% over 3 steps) as a 1:1 mixture of diastereomers. δ_H (300 MHz, $CDCl_3$): 7.66 (2H, m, C4-H/H'), 7.41 (1H, d, *J* 5.8, C3-H'), 7.40 (1H, d, *J* 5.8 C3-H), 5.80 (2H, s, CHBr), 3.83 (6H, s, methyl); δ_C (75 MHz, $CDCl_3$): 168.2 (2 × CO), 134.2 (C1/2), 130.7 (C4/5), 129.9 (C3/6), 53.8 (Me), 53.7 (Me), 43.4 (CBr), 43.1 (CBr);

α^1,α^2-dibromo-1,2-phenylenediacetyl chloride (56)



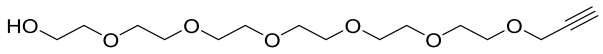
1,2-phenylenediacetic acid (1.00 g, 5.15 mmol) and thionyl chloride (2.30 mL, 31.15 mmol) were combined and heated to 80 °C for 1 h. After cooling to rt, carbon tetrachloride (3 mL), *N*-bromosuccinimide (2.8 g, 15.7 mmol, recrystallized from H_2O) and HBr (3 drops) were added and the reaction heated to 80 °C for 2 hours. After cooling to rt, a further portion of *N*-bromosuccinimide (500 mg, 2.8 mmol) was added and the reaction heated to 80 °C for 30 min. The reaction mixture was cooled to 0 °C and stirred for 30 min to ensure complete precipitation of succinimide by-products. The reaction mixture was filtered and washed with CH_2Cl_2 (15 mL) to yield a solution of the intermediate **56**. The reactions were monitored by adding samples to methanol and observing the methyl esters by LCMS.

Methyl 2,5-dibromovalerate (63)



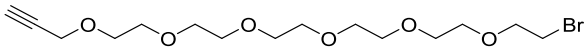
δ-Valerolactone (2.59 g, 25.9 mmol), bromine (1.99 mL, 38.9 mmol) and phosphorus tribromide (47 µL, 0.5 mmol) were combined and the reaction heated at 110 °C for 2.5 h. The reaction was then cooled to 0 °C before the addition of methanol (5.23 mL, 130 mmol) and *p*-toluenesulfonic acid monohydrate (47 mg, 0.25 mmol). The reaction was heated to 80 °C for 3 h before cooling and removal of excess methanol under reduced pressure. CH_2Cl_2 (50 mL) was added and the lower organic layer obtained, washed sequentially with water (10 mL), NaOH (10%, 10 mL), and water (10 mL). Organics were dried ($MgSO_4$) and concentrated under reduced pressure. The product was further purified by column chromatography (1:4 EtOAc:hexane) to yield **63** (5.457 g, 77%) as a colourless oil. δ_H (300 MHz, $CDCl_3$): 4.27 (1H, dd, *J* 8.0, 6.0, C2-H), 3.80 (3H, s, methyl), 3.43 (2H, t, *J* 6.3, C5-H_2), 2.39-1.88 (4H, m, C3-H_2, C4-H_2); δ_C (75 MHz, $CDCl_3$): 169.9 (C1), 53.0 (MeO), 44.5 (C2), 33.3, 32.0 (C5), 30.1; ν_{max}/cm-1 (film) 1736 (C=O); The product proved difficult to obtain M.S. data for – either by accurate mass ESI-MS or the LCMS system.

3,6,9,12,15,18-hexaoxahenicosa-20-ynol (92)



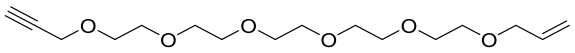
Hexaethyleneglycol (25.0 g, 88.5 mmol) was added to anhydrous THF (400 mL) and the solution cooled to 0 °C. Sodium hydride (60% in mineral oil, 3.57 g, 89.3 mmol) was added in 10 portions over 1 h. Propargyl bromide (80% in toluene, 10.0 mL, 111.7 mmol) was added dropwise over 45 min. The reaction was allowed to warm to rt and was stirred for 18 h. The reaction was quenched with brine (50 mL) and solvent removed *in vacuo*. The reaction mixture was extracted with CH_2Cl_2 (3 × 100 mL), combined organic fractions were washed with brine (50 mL), dried ($MgSO_4$) and concentrated *in vacuo* to yield a yellow oil. The crude product was purified by two-stage column chromatography (1. MeOH–CH_2Cl_2, 2:98; 2. MeOH–CH_2Cl_2, 5:95) to give propargyl-PEG **92** (13.28 g, 47%) as a yellow oil. δ_H (300 MHz, $CDCl_3$): 4.21 (2H, d, *J* 2.4, C19-H_2), 3.58-3.75 (22H, m, C1-C17 H_2), 2.70 (1H, t, *J* 6.3, OH), 2.44 (1H, t, *J* 2.4, C21-H); δ_C (75 MHz, $CDCl_3$): 79.7 (C20), 74.5 (C21), 72.6 (PEG CH_2), 70.7-70.2 (multiple PEG CH_2), 69.1 (PEG CH_2), 61.7 (C1), 58.4 (C19); ν_{max}/cm^{-1} (film): 3354 (alkyne C-H), 3241 (OH), 2112 (alkyne C≡C), 1087 (C-O); *m/z* (ES): Found M+NH_4^+ 338.217975, expected for [$C_{15}H_{28}O_7$ + NH_4] 338.217329.

21-bromo-4,7,10,13,16,19-hexaoxahenicosyne (93)



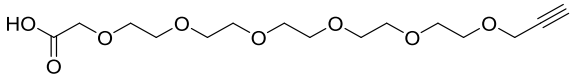
92 (200 mg, 0.53 mmol) was dissolved in CH_2Cl_2 (2 mL) and cooled to 0 °C. Carbon tetrabromide (444 mg, 1.33 mmol) was added followed by a solution of triphenylphosphine (348 mg, 1.33 mmol) in CH_2Cl_2 (2 mL). The reaction was stirred at rt for 18 h. The solution was concentrated *in vacuo* and the residue triturated in ice-cold pentane (2 × 20 mL). The soluble fractions were combined, concentrated *in vacuo* and the crude purified by column chromatography (MeOH–CH_2Cl_2 gradient, 0:1 to 5:95). The bromide **93** was isolated as a yellow oil (68% pure w/w, 158 mg, 53% yield). δ_H (500 MHz, $CDCl_3$): 4.21 (2H, d, *J* 2.5, propargyl-CH_2), 3.81 (2H, t, *J* 6.5, C20-H_2), 3.72-3.65 (20H, m, 10 x PEG CH_2), 3.48 (2H, t, J 6.5, C21-H_2); δ_C (75 MHz, $CDCl_3$): 79.7 (propargyl C), 74.5 (propargyl CH), 71.2 (PEG CH_2), 70.8-70.5 (PEG CH_2), 70.4 (PEG CH_2), 69.1 (CH_2-OH), 58.4 (propargyl CH_2), 30.3 (C21-Br); ν_{max}/cm^{-1} (CCl_4 solution): 3270 (alkyne C-H), 2130 (alkyne C≡C), 1090 (C-O); *m/z* (ES): Found M+NH_4^+ 400.133562, expected for [$C_{15}H_{27}O_6{}^{79}Br$ + NH_4] 400.132926.

Attempted alkylation of pentaerythritol with 93 to form 95, forming instead by-product 4,7,10,13,16,19-hexaoxahenicos-1-en-21-yne 96



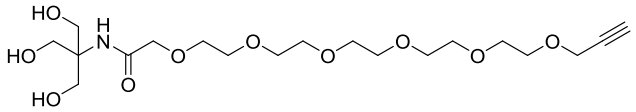
Pentaerythritol (28.6 mg, 0.21 mmol) was dissolved in DMF (1 mL) and NaH (10.4 mg, 0.26 mmol) added and the reaction stirred for 10 minutes. **93** (100 mg, 0.26 mmol) was dissolved in DMF (1 mL) and added dropwise. No sign of **95** was observed by LCMS however the major peak corresponded to 96 (m/z 302 and 325 – H^+ and Na^+ adducts respectively). Reaction was stirred for 18 h before the addition of water (20 mL). The reaction mixture was extracted with CH_2Cl_2 (2 × 40 mL) and organics combined and concentrated under reduced pressure. COSY NMR of the crude extracted material indicated the presence of alkene protons supporting formation of **96**.

3,6,9,12,15,18-hexaoxahenicosa-20-ynoic acid (97)



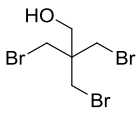
Chromium(VI)oxide (0.625 g, 6.25 mmol) was added to conc. H_2SO_4 (1.2 mL) in a glass vial and the mixture transferred to a flask containing water (4 mL). Reaction was cooled to 0 °C and a solution of **92** (1.0 g, 3.125 mmol) in acetone (10 mL) was added dropwise. The reaction was stirred open to air for 3 h at 0 °C. The reaction mixture was concentrated *in vacuo* and extracted with CH_2Cl_2 (3 × 100 mL). Organic fractions were combined and washed with H_2O (30 mL). Organics were concentrated to 100 mL before extraction with NaOH (2 × 50 mL). Hydroxide layers were combined and washed with CH_2Cl_2 (10 mL), acidified with HCl (aq) (3M, to pH 1) and extracted with CH_2Cl_2 (5 × 50 mL). Organic fractions were combined, dried ($MgSO_4$) and concentrated under reduced pressure to give the acid **97** (373 mg, 36%) as a yellow oil. δ_H (300 MHz, $CDCl_3$): 4.20 (2H, d, *J* 2.4, C19-H_2), 4.16 (2H, s, C2-H_2), 3.72-3.65 (20H, m, C4-C17 H_2), 2.44 (1H, t, *J* 2.4, C21-H); δ_C (75 MHz, $CDCl_3$): 172.3 (C1), 79.6 (C20), 74.6 (C21), 71.1 (PEG CH_2), 70.6-70.3 (multiple PEG CH_2), 69.1 (C2), 69.0 (PEG CH_2), 58.3 (C19); ν_{max}/cm^{-1} (film): 3466 (OH), 3247 (alkyne C-H), 2116 (alkyne C≡C), 1087 (C-O); *m/z* (ES): Found M+NH_4^+ 352.197665, expected for [$C_{15}H_{26}O_8$ + NH_4] 352.196593.

***N*-[tris(hydroxymethyl)methyl]-3,6,9,12,15,18-hexaoxahenicosa-20-ynamide (98)**



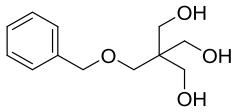
97 (200 mg, 0.6 mmol) was added to a flask followed by water (5 mL), Tris base (218 mg, 1.8 mmol) and EDC.HCl (230 mg, 1.2 mmol). The reaction was stirred open to air for 16 h. DMAP (218 mg, 1.8 mmol) was added and the pH adjusted to pH 8 using HCl (aq) (1M). Further EDC.HCl (230 mg, 1.2 mmol) was added and the reaction was stirred for 1 h at rt followed by heating at 60 °C for 2 h. After cooling, the reaction was extracted with CH_2Cl_2 (3 × 50 mL). The aqueous fraction was diluted with sat. $NaHCO_3$ (aq) (30 mL) and continuously extracted over 16 h using CH_2Cl_2 as the organic phase (200 mL). The organic phase was concentrated *in vacuo* and the crude product purified by column chromatography (MeOH–CH_2Cl_2, 1:9). The alcohol **98** (45 mg, 17%) was obtained as a colourless gum. δ_H (500 MHz, $CDCl_3$): 7.65 (1H, bs, NH), 4.54 (2H, t, *J* 6.3, OH), 4.21 (2H, d, *J* 2.4, C19-H_2), 4.02 (2H, s, C2-H_2), 3.74-3.63 (26H, m, CH_2OH + C4-C17 CH_2), 2.47 (1H, t, *J* 2.4, C21-H); δ_C (125 MHz, $CDCl_3$): 171.3 (C1), 79.6 (C20), 74.7, 70.7-70.1, 69.0, 63.1, 61.7, 58.4; ν_{max}/cm^{-1} (CCl_4 solution): 3410 (OH) 1670 (amide C=O), 1550 (N-H bend), 1100 (C-O); *m/z* (ES): Found $M+H^+$ 438.233891, expected for [$C_{19}H_{35}O_{10}N$ + H] 438.233373.

2,2-bis(bromomethyl)-3-bromopropanol (99) [144]



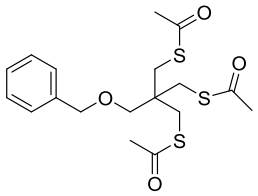
Pentaerythritol (12.8 g, 94 mmol) was dissolved in acetic acid (10 mL) and HBr (aq) (48%, 50 mL) added. The reaction was heated to 120 °C for 16 h. After cooling, further HBr (aq) (48%, 50 mL) and sulfuric acid (conc., 23 mL) was added and the reaction heated to 120 °C for 16 h. The reaction was cooled and extracted with CH_2Cl_2 (3 × 100 mL). The product was isolated by column chromatography (EtOAc:hexane gradient 1:9 to 2:8) as an off-white solid (9.35 g, 31%). 2,2-bis(bromomethyl)-3-bromopropanyl acetate impurity (approx. 27%) was observed by NMR. Characterisation was in accordance with the literature.[144] δ_H (300 MHz, $CDCl_3$): 3.75 (2H, s, CH_2OH), 3.54 (6H, s, CH_2Br); δ_C (75 MHz, $CDCl_3$): 62.3 (CH_2OH), 44.1 (propyl C2), 34.4 (CH_2Br).

1-benzoxy-2,2,2-tris(hydroxymethyl)ethane (101)



Pentaerythritol (1.0 g, 7.34 mmol) was added to anhydrous DMF (5 mL) and cooled to 0 °C. Sodium hydride (60% in mineral oil, 320 mg, 8.0 mmol) was added and the reaction stirred for 30 min. Benzyl bromide (960 μL, 8.1 mmol) was added and the reaction was stirred for 4 h. at rt before being kept at –20 °C overnight. The reaction was quenched with brine (15 mL) and extracted with CH_2Cl_2 (3 × 50 mL). The combined organic fractions were washed with brine (20 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude was purified by column chromatography (MeOH–CH_2Cl_2, 1:9) to yield alcohol **101** (291 mg, 18%) as an off-white gum. δ_H (300 MHz, D_3COD): 7.37-7.23 (5H, m, phenyl H), 4.52 (2H, s, benzyl CH_2), 3.64 (6H, s, CH_2OH), 3.52 (2H, s, ethane C1-H_2); δ_C (75 MHz, D_3COD): 140.0 (Ph C1), 129.4 (Ph), 128.6 (Ph), 128.6 (Ph), 74.6 (Benzyl CH_2), 71.1 (ethane C1), 63.4 (hydroxymethyl CH_2), 46.9 (ethane C2); ν_{max}/cm^{-1} (film): 3260 (OH), 1479 (Ar C=C), 1403 (Ar C=C), 1015 (C-O); *m/z* (ES): Found $M+Na^+$ 249.109717, expected for [$C_{12}H_{18}O_4$ + Na] 249.109730.

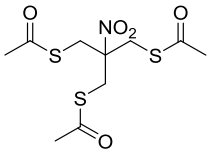
2,2,2-tris(acetylsulfanylmethyl)-1-benzoxyethane (102)



Triphenylphosphine (1.04 g, 3.98 mmol) was added to anhydrous THF (10 mL) followed by DIAD (805 mg, 3.98 mmol) to give a yellow solution. A solution of **101** (200 mg, 0.88 mmol) and thioacetic acid (284 μL, 3.98 mmol) in anhydrous THF (10 mL) was added dropwise, giving a green solution. The reaction was stirred overnight at rt before quenching with brine (30mL) and extracting with CH_2Cl_2 (2 × 75 mL). Combined organic fractions were washed with brine (20 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude mixture was purified by column chromatography (EtOAc–hexane gradient, 0:1 to 1:4) to give a mixture of triphenylphosphine and **102**. Trituration in pentane (2 × 30 mL) yielded product **102** (145 mg, 44%) as a yellow oil. δ_H (300 MHz, $CDCl_3$): 7.39-7.25 (5H, m, phenyl H), 4.45 (2H, s, benzyl CH_2), 3.29 (2H, s, ethane C1-H_2), 3.09 (6H, s, CH_2S), 2.33 (9H, s, SCH_3); δ_C (75 MHz, $CDCl_3$): 194.9 (acetyl CO), 128.3 (Ph), 127.6 (Ph), 127.6 (Ph), 73.4 (benzyl CH_2), 73.0 (ethane C1), 33.5 (CH_2S), 30.5 (acetyl CH_3) (quaternary carbons too

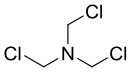
weak to assign); ν_{max}/cm^{-1} (solid): 1689 (C=O), 1093 (C-O); *m/z* (ES): Found $M+NH_4^+$ 418.118998, expected for $[C_{18}H_{24}O_4S_3 + NH_4]$ 418.117498.

Tris(acetylthiomethyl)nitromethane (107)



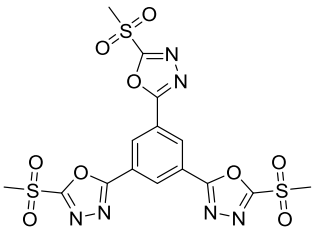
Tris(hydroxymethyl)nitromethane (500 mg, 3.31 mmol) and triphenylphosphine (3.9 g, 14.9 mmol) were combined followed by addition of THF (10 mL) and the reaction cooled to 0 °C. DIAD (3.1 mL, 14.9 mmol) and thioacetic acid (1.06 mL, 14.9 mmol) were added upon which point the reaction turned yellow-orange and cloudy. The reaction was allowed to warm to rt and stirred for 16 h – the reaction turned orange and clear. Saturated $NaHCO_3$ (aq) (20 mL) was added and the reaction mixture extracted with CH_2Cl_2 (2 × 50 mL). Organics were combined and washed with saturated NH_4Cl (aq) (25 mL), dried ($MgSO_4$), filtered and concentrated under reduced pressure. Purification by two stage column chromatography (1. EtOAc–hexane gradient 3:17 to 3:10, 2. EtOAc–hexane gradient 1:19 to 3:17) yielded **107** as a yellow solid (585 mg, 54%) δ_H (300 MHz, $CDCl_3$): 3.60 (6H, s, CH_2), 2.36 (9H, s, SAc CH_3); δ_C (75 MHz, $CDCl_3$): 193.0 (CO), 92.6 (NO_2C), 33.3 (CH_2), 30.3 (SAc CH_3); m/z (ES): Found $M+NH_4^+$ 343.045814, expected for $[C_{10}H_{15}NO_5S_3 + NH_4]$ 343.045061.

Tris(chloromethyl)amine (109)[125]



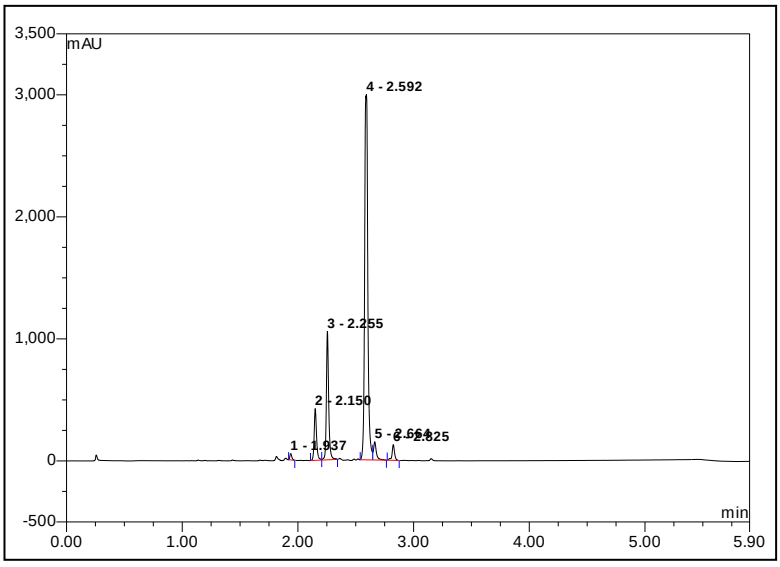
Hexamethylene tetramine (2.8 g, 20 mmol) was crushed using a mortar and pestle before addition of PCl_5 (12.56 g, 60 mmol). A drying tube ($CaCl_2$) was fitted and the reaction heated to 100 °C for 3.5 h whereupon the reaction mixure turned to a red/brown liquid. The product **109** was crystallised from CH_2Cl_2 at rt as light brown crystals (1.4 g, 43%) Characterisation was in accordance with the literature.[125] δ_H (300 MHz, $CDCl_3$): 5.14 (6H, s, NCH_2Cl); δ_C (75 MHz, $CDCl_3$): 65.4 (NCH_2Cl).

1,3,5-tris(5-methylsulfonyl-1,3,4-oxadiazol-2-yl)benzene (110)

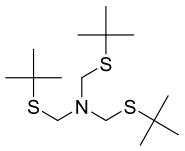


120 (100 mg, 0.24 mmol), sodium periodate (459 mg, 2.14 mmol), and ruthenium chloride trihydrate (0.6 mg, 2.4 µmol) were combined. Carbon tetrachloride (2 mL), acetonitrile (2 mL) and water (4 mL) were added and the reaction stirred open to air for 16 h. CH_2Cl_2 (20 mL) was added and the precipitate collected by filtration. The reaction mixture had brine (10 mL) added and further precipitate was collected. The precipitate was washed in acetonitrile (3 × 10 mL) and product obtained by removal of acetonitrile under reduced pressure to yield **110** as a white solid (81 mg, 65%). δ_H (300 MHz, DMSO-*d6*): 8.93 (3H, s, benzene CH), 3.79 (9H, s, SO_2CH_3); δ_C (75 MHz, DMSO-*d6*): 163.9 (oxadiazole C5), 162.7 (oxadiazole C2), 129.2 (benzene C2), 125.2 (benzene C1), 43.0 (SO_2CH_3); ν_{max}/cm^{-1} 1341 (S=O), 1151 (S=O) *m/z* (ES): Found [2M+NH_4^+] 1049.998070, expected for [$C_{30}H_{28}N_{13}O_{18}S_6$] 1049.99940.

HPLC of **110**

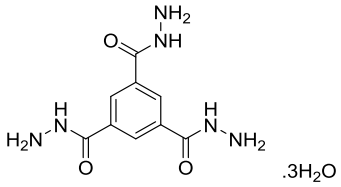


Tris(*tert*-butylthiomethyl)amine (112)



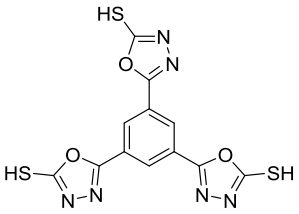
109 (150 mg, 0.924 mmol) was dissolved in CH_2Cl_2 (2 mL) and cooled to 0 ° C before addition of *tert*-butyl thiol (360 µL, 3.22 mmol). The reaction was stirred for one hour before addition of H_2O (5 mL) and extraction with CH_2Cl_2 (2 × 10 mL). Organics were combined, dried ($MgSO_4$) and filtered before solvent was removed in vacuo to yield **112** (167 mg, 56%). δ_H (300 MHz, $CDCl_3$): 4.10 (6H, s, NCH_2S), 1.35 (27H, s, tBu CH_3); δ_C (75 MHz, $CDCl_3$): 52.6 (NCH_2S), 42.6 (tBu C_q), 31.5 (tBu CH_3); *m/z* (ES): Found $M+Na^+$ 346.167575, expected for [$C_{15}H_{33}NS_3$ + Na] 346.166734.

1,3,5-benzenetricarboxylic acid trihydrazide trihydrate (118)[145]



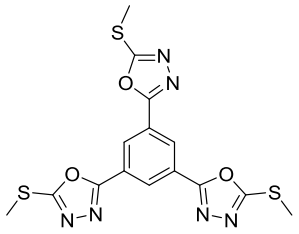
To a flask containing triethyl 1,3,5-benzenetricarboxylate (22.15 g, 75.3 mmol) was added methanol (400 mL) and hydrazine hydrate (64%, 86 mL, 1.13 mol) and the mixture stirred for 18 h at rt The precipitate was filtered, washed with methanol (2 x 100 mL) and dried under vacuum. The product was ground to a fine powder and further dried under vacuum to yield the product **118** as a white solid (20.5 g, 89%) Spectral characterisation was in accordance with literature. δ_H (300 MHz, DMSO-*d6*): 9.83 (3H, bs, NH), 8.32 (3H, s, CH), 4.55 (6H, bs, NH_2); δ_C (75 MHz, DMSO-*d6*): 164.9 (CO), 133.8, 128.0; ν_{max}/cm^{-1} 3281, 3187, 3053 (NH_2 and H_2O), 1621 (CO), 1502; *m/z* (ES): Found [$2M+H^+$] 505.200763, expected for [$C_{18}H_{25}N_{12}O_6$] 505.201453.

1,3,5-tris(5-mercapto-1,3,4-oxadiazol-2-yl)benzene (119)



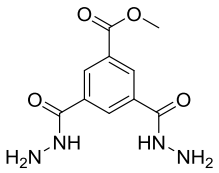
118 (15.0 g, 49.0 mmol) and potassium hydroxide (15.0 g, 267.9 mmol) were combined and ethanol (300 mL) added. Carbon disulphide (16.9 mL, 267.9 mmol) was added in 5 mL portions followed by stirring at 110 °C for 64 h. After cooling to rt, the reaction mixture was concentrated under reduced pressure. Water (200 mL) was added and the solution filtered to remove unreacted starting material (2.6 g, 17%). The filtrate was acidified to pH 3 with HCl (aq) (1M) upon which a white slurry formed. The solid was obtained by filtration and dried under vacuum to yield **119** as a yellow solid (6.97 g, 38%). An unknown impurity was observable by NMR. δ_H (300 MHz, d6-DMSO): 8.25 (3H, s, benzene CH); δ_C (75 MHz, d6-DMSO): 177.8 (oxadiazole C2), 158.4 (oxadiazole C5), 133.3 (benzene C1), 125.0 (benzene C2); ν_{max}/cm^{-1} 2737 (SH); *m/z* (ES negative): Found [M−H] 376.960039, expected for $[C_{12}H_5N_6O_3S_3]$ 376.959075.

1,3,5-tris(5-methylthio-1,3,4-oxadiazol-2-yl)benzene (120)



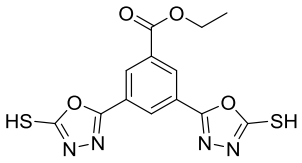
119 (470 mg, 1.24 mmol) and DMSO (10 mL) were combined and heated to 50 °C for 20 min to dissolve **119** before cooling to rt. Triethylamine (605 µL, 4.34 mmol) and methyl iodide (255 µL, 4.1 mmol) were added and the reaction stirred for 30 min. The product was obtained by precipitation at 0 °C. The solid was obtained by filtration and triturated in CH_2Cl_2, impurities were removed by filtration and the CH_2Cl_2 removed under reduced pressure to yield **120** as a yellow solid (267 mg, 51%). δ_H (300 MHz, $CDCl_3$): 8.74 (3H, s, benzene C2-H), 2.83 (9H, s, SCH_3); δ_C (75 MHz, $CDCl_3$): 166.5 (oxadiazole C2), 163.8 (oxadiazole C5), 126.6 (benzene C2), 125.9 (benzene C4), 14.7 (SCH_3); ν_{max}/cm^{-1} 1462, 1180; *m/z* (ES): Found $[2M+H^+]$ 841.033187, expected for $[C_{30}H_{25}N_{12}O_6S_6]$ 841.033879.

Methyl 3,5-bis(hydrazinocarbonyl)benzoate (123)



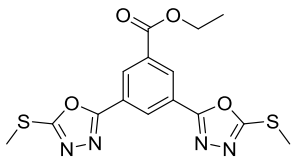
To trimethyl 1,3,5-benzenetricarboxylate (1.01 g, 4.00 mmol) was added methanol (50 mL) and hydrazine hydrate (66%, 2.9 mL, 40 mmol) and the reaction stirred for 16 h. The precipitate was collected by filtration, upon which more precipitation from the filtrate occurred – this batch was also collected by filtration. Batches were combined and dried to give **123** (883 mg, 83%) This material was taken into the next step of the synthesis without further purification. δ_H (300 MHz, d6-DMSO): 10.03 (2H, bs, NH), 8.49 (3H, s, benzene C2-H and C4-H), 4.58 (4H, bs, NH_2), 3.92 (3H, s, methyl).

Ethyl 3,5-bis(2-mercapto-1,3,4-oxadiazol-5-yl)benzoate (124)



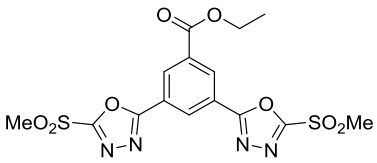
123 (900 mg, 3.6 mmol) and potassium hydroxide (900 mg, 16.1 mmol) were combined and EtOH (100 mL) added. Carbon disulfide (966 µL, 16.1 mmol) was added dropwise before the mixture was heated to 110 °C for 22 h. The reaction was cooled and solvent removed *in vacuo* before addition of H_2O (100 mL) and acidification (pH 1) with HCl (1 M). The white precipitate was obtained by filtration, redissolved in ethyl acetate (50 mL), dried ($MgSO_4$), filtered and solvent removed *in vacuo* to yield a mixture of **124** and an unidentified impurity (770 mg total). Purification of a 400 mg sample of this mixture by mass-directed HPLC yielded **124** (90 mg, 7%). δ_H (300 MHz, d6-DMSO): 8.47 (2H, d, J 1.5, benzene C2-H), 8.40 (1H, t, J 1.5, benzene C4-H), 4.43 (2H, q, J 7.1, ethyl CH_2), 1.39 (3H, t, J 7.1, ethyl CH_3); δ_C (75 MHz, d6-DMSO): 163.7 (COO), 158.7 (oxadiazole C2), 158.7 (oxadiazole C5), 132.1 (benzene C1), 128.7 (benzene C2), 126.8 (benzene C4), 124.5 (benzene C3), 61.9 (ethyl C1), 14.0 (ethyl C2); *m/z* (ES): Found $M+H^+$ 351.021929, expected for [$C_{13}H_{10}N_4O_4S_2$ + H] 351.021623.

Ethyl 3,5-bis(5-methylthio-1,3,4-oxadiazol-2-yl)benzoate (125)



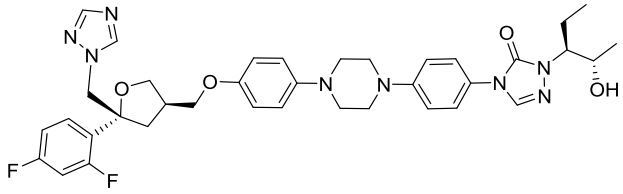
124 (90 mg, 0.26 mmol) was dissolved in THF (1 mL) and methyl iodide (35 µL, 0.57 mmol) added followed by triethylamine (79 µL, 0.57 mmol). The reaction was stirred for 2 h before quenching with H_2O (2 mL) and extraction with ethyl acetate (2 × 10 mL). Organics were combined, dried ($MgSO_4$), filtered and solvent removed under reduced pressure to give product **125** (99 mg, 99%). δ_H (300 MHz, DMSO-*d6*): 8.62 (1H, t, J 1.3, benzene CH), 8.58 (2H, d, J 1.3, benzene CH), 4.44 (2H, q, J 7.1, ethyl CH_2), 2.82 (6H, s, methyl CH_3), 1.40 (3H, t, J 7.1, ethyl CH_3); δ_C (75 MHz, DMSO-*d6*): 14.1 (ethyl CH_3), 14.4 (methyl CH_3), 61.9 (ethyl CH_2), 125.0 (benzene C3/5), 127.4 (benzene C4), 128.9 (benzene C2/6), 132.2 (benzene C1), 163.5 (oxadiazole C2), 163.8 (benzoic COO), 165.8 (oxadiazole C5).

Ethyl 3,5-bis(2-mesyl-1,3,4-oxadiazol-5-yl)benzoate (126)



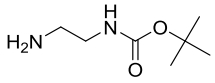
To **125** (56 mg, 0.15 mmol) was added $(NH_4)_6Mo_7O_{24}(H_2O)_4$ (198 mg, 0.16 mmol) followed by ethanol (10 mL). The reaction was cooled to 0 °C before addition of hydrogen peroxide (30%, 245 µL, 2.4 mmol). The reaction was then heated to 60 °C for 90 min. The reaction was cooled followed by addition of sodium thiosulfate solution (10%, 10 mL). The mixture was extracted with ethyl acetate (2 × 30 mL), organics were combined, dried ($MgSO_4$) and solvent removed in vacuo. Purification by mass directed HPLC gave **126** (10 mg, 23%) as a white solid. **126** degraded to the related oxadiazol-2-one before full characterisation was obtained. δ_H (300 MHz, DMSO-*d6*): 8.90 (1H, t, *J* 1.6, benzene C4), 8.80 (2H, d, *J* 1.6, benzene C2), 4.48 (1H, q, *J* 7.1, Et CH_2), 1.42 (2H, t, *J* 7.1, Et CH_3); *m/z* (ES): Found $M+NH_4^+$ 460.059448, expected for $[C_{15}H_{14}N_4O_8S_2 + NH_4]$ 460.059131.

Extraction of Noxafil solution yielding posaconazole (130)



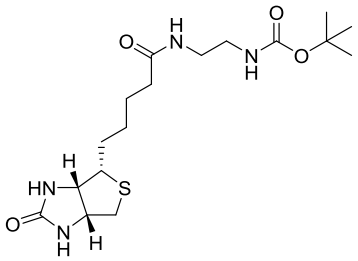
Noxafil® suspension (40 mg/mL, 5 mL) was diluted with water (50 mL) and extracted with EtOAc (3 × 50 mL). Organics were combined and washed with brine (20 mL) followed by sat. $NaHCO_3$ (aq) (30 mL). Solvent was removed under reduced pressure to yield **130** (130 mg, 65%).

***tert*-butyl *N*-(2-aminoethyl)carbamate (133)**



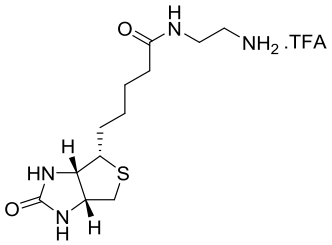
Ethylene diamine (6.0 mL, 90 mmol) was added to a flask containing anhydrous CH_2Cl_2 (50 mL). A solution of Boc-anhydride (3.27 g, 15 mmol) in CH_2Cl_2 (53 mL) was added dropwise with vigorous stirring over 3 h. The reaction was stirred for 16 h at rt before addition of Na_2CO_3 (aq) (20%, 40 mL) and extraction with CH_2Cl_2 (3 × 50 mL). The combined organic fractions were dried ($MgSO_4$) and concentrated *in vacuo* to give product **133** (1.76 g, 73%) as a white solid. δ_H (500 MHz, $CDCl_3$): 4.85 (1H, bs, NH), 3.20-3.15 (2H, bm, C2-H_2), 2.80 (2H, t, *J* 5.8, C1-H_2), 1.45 (9H, s, Boc CH_3), 1.22 (2H, bs, NH_2); δ_C (125 MHz, $CDCl_3$): 156.2 (Boc NC(O)O), 79.1 (Boc $OC(Me)_3$), 43.4 (CH_2), 41.9 (CH_2), 28.4 (Boc CH_3); ν_{max}/cm^{-1} (solid): 3357 (NH), 1683 (C=O); *m/z* (ES): Found $M+H^+$ 161.128271, expected for [$C_7H_{16}N_2O_2$ + H] 161.128454.

_N_-[(2-_tert_-butoxycarbonylamino)ethyl]-5-[(3a_S_,4_S_,6a_R_)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (134)



Biotin (2 g, 8.19 mmol), EDC.HCl (1.88 g, 9.82 mmol), **133** (1.31 g, 8.19 mmol), HOBt (1.33 g, 9.82 mmol) and DIPEA (3.56 mL, 20.5 mmol) were added to anhydrous DMF (30 mL). The reaction was heated to 55 °C for 16 h. Brine (30 mL) was added and the reaction mixture extracted with CH_2Cl_2 (3 × 50 mL). The combined organic fractions were dried ($MgSO_4$) and concentrated *in vacuo*. The product was purified by two-stage column chromatography (1. MeOH–CH_2Cl_2, 1:9; 2. MeOH–CH_2Cl_2, 1:9). Product elution was monitored by NMR during the second column procedure. The amide **134** (1.29 g, 41%) was isolated as a white solid. δ_H (300 MHz, D_3COD): 4.51 (1H, dd, *J* 7.9, 4.8, Biotin C6a-H), 4.32 (1H, dd, *J* 7.9, 4.5, Biotin C3a-H), 3.29-3.11 (5H, m, Biotin C4-H, 2 × CH_2), 2.95 (1H, dd, *J* 12.8, 5.0, Biotin C6-H), 2.72 (1H, d, *J* 12.8, Biotin C6-H'), 2.22 (2H, t, *J* 7.1, CH_2), 1.82-1.54 (4H, m, 2 × CH_2), 1.52-1.41 (11H, m, CH_2, Boc CH3); δ_C (75 MHz, D_3COD): 176.4 (amide C, Boc C(O)O), 166.1 (Biotin C2), 80.2 (Boc $OC(Me)_3$), 63.4 (Biotin C3a), 61.6 (Biotin C6a), 57.0 (Biotin C4), 41.0 (Biotin C6), 40.6 (CH_2), 36.9 (CH_2), 29.8 (CH_2), 29.5 (CH_2), 28.8 (Boc CH_3), 26.9 (CH_2); *m/z* (ES): Found $M+Na^+$ 409.1868, expected for [$C_{17}H_{30}O_4N_4S$ + Na] 409.1880.

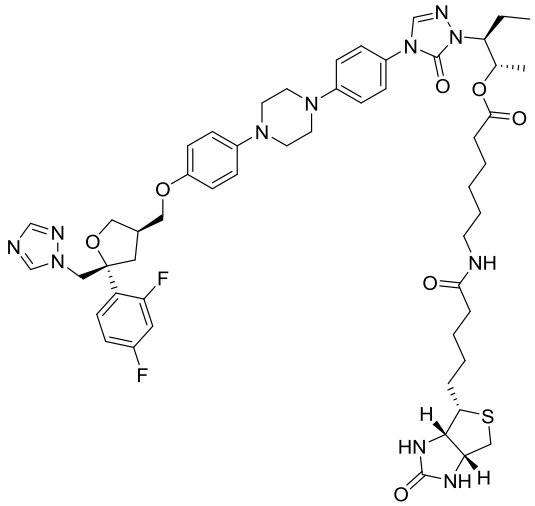
_N_-[(2-amino)ethyl]-5-[(3a_S_,4_S_,6a_R_)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (135)



49 (840 mg, 2.17 mmol) was added to CH_2Cl_2 (10 mL) and TFA (3 mL). The reaction was stirred for 90 min. Concentration *in vacuo* yielded **135** as a TFA salt (1.28 g, >100%). δ_H (500 MHz, D_3COD): 4.53 (1H, dd, *J* 7.9, 4.8, Biotin C6a-H), 4.33 (1H, dd, *J* 7.9, 4.8, Biotin

C3a-H), 3.47 (2H, dt, *J* 6.0, 5.3, ethyl C1-H_2), 3.27-3.21 (1H, m, Biotin C4-H), 3.07 (2H, t, *J* 6.0, ethyl C2-H_2), 2.96 (1H, dd, *J* 12.8, 4.9, Biotin C6-H), 2.74 (1H, d, *J* 12.8, Biotin C6-H'), 2.29 (2H, t, *J* 7.6, Biotin C4'-H_2), 1.82-1.61 (4H, m, Biotin CH_2), 1.52-1.45 (2H, m, Biotin CH_2); δ_C (75 MHz, D_3COD): 177.4 (amide C), 166.1 (Biotin C2), 63.4 (Biotin C3a), 61.7 (Biotin C6a), 57.0 (Biotin C4), 41.0 (Biotin C6), 41.0 (CH_2), 38.2 (CH_2), 36.5 (CH_2), 29.8 (CH_2), 29.5 (CH_2), 26.5 (CH_2); ν_{max}/cm^{-1} (solid): 3000-3500 (NH), 1671 (C=O), 1648 (C=O); *m/z* (ES): Found $M+Na^+$ 309.1344, expected for [$C_{12}H_{22}N_4O_2S$ + Na] 309.1356.

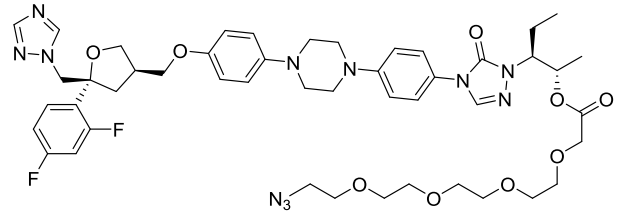
(2*S*,3*S*)-3-{4-[4-(4-{4-[(3*R*, 5*R*)-5-(2,4-difluoro-phenyl)-5-([1,2,4]triazol-1-ylmethyl)tetrahydrofuran-3-ylmethoxy]phenyl}piperazin-1-yl)phenyl]-5-oxo-4,5-dihydro-[1,2,4]triazol-1-yl}pent-2-yl 6-{5-[(3a*S*,4*S*,6a*R*)-2-oxo-hexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanoylamino}hexanoate (137)



N-biotinyl-6-aminohexanoic acid **131** (40 mg, 0.112 mmol), posaconazole **130** (93 mg, 0.134 mg), EDC.HCl (21 mg, 0.112 mmol) and DMAP (16 mg, 0.134 mmol) were added to anhydrous DMF (1 mL). The reaction was heated to 50 °C for 18 h before quenching with brine (10 mL) and extracting with CH_2Cl_2. Combined organic fractions were dried ($MgSO_4$) and concentrated *in vacuo*. The crude was purified by two-stage column chromatography (1. MeOH–DCM, 1:9; 2. MeOH–DCM gradient, 0:1 to 2:98). The biotinylated drug **137** (21 mg, 18%) was obtained as a white powder. δ_H (500 MHz, $CDCl_3$): 8.12 (1H, s, oxotriazole C3-H), 7.80 (1H, s, triazole H), 7.69 (1H, s, triazole H), 7.43 (2H, d, *J* 8.5, N-Ar-N C2-H), 7.40-7.35 (1H, m, difluorophenyl C-H), 7.03 (2H, d, *J* 8.5, N-Ar-N C3-H), 6.93 (2H, d, *J* 9.0, N-Ar-O C2-H), 6.89-6.80 (2H, m, difluorophenyl C-H), 6.77 (2H, d, *J* 9.0, N-Ar-O C3-H), 6.11 (1H, s, amide NH), 6.04 (1H, s, biotin N3-H), 5.28-5.17 (2H, m, biotin N1-H, pentyl C2-H), 4.65 (1H, d, *J* 14.2, triazolylmethyl CH_2), 4.55-4.45 (2H, m, triazolylmethyl CH_2, biotin C6a-H), 4.32-4.27 (1H, m, biotin C3a-H), 4.24-4.15 (1H, m, pentyl C3-H), 4.12 (1H, t, *J* 8.0, tetrahydrofuran C2-H), 3.78 (1H, dd, *J* 8.0, 6.8, tetrahydrofuran C2-H), 3.72-3.67 (1H, m, tetrahydrofuranylmethoxy CH_2), 3.67-3.58 (3H, m, tetrahydrofuranylmethoxy CH_2), 3.40-3.33 (4H, m, piperazine C2-H_2), 3.25-3.20 (4H, m, piperazine C3-H_2), 3.20-3.07 (3H, m, hexanoic acid C6-H_2, biotin C4-H), 2.89 (1H, dd, *J* 12.7, 5.0, biotin C6-H), 2.72 (1H, dd, *J* 12.7, biotin C6-H'), 2.66-2.51 (2H, m, tetrahydrofuran C3-H, tetrahydrofuran C4-H), 2.24 (2H, t, *J* 7.3, hexanoic acid C2-H_2), 2.17 (2H, t, *J* 7.3, biotin C4'-H_2), 2.08 (1H, dd, *J* 12.5, 8.4, tetrahydrofuran C4-H'), 1.97-1.85

(1H, m, pentyl C4-H), 1.84-1.59 (3H, m, pentyl C4'-H, biotin C1'-H_2), 1.58-1.48 (2H, m, hexanoic acid C3-H_2), 1.47-1.38 (4H, m, hexanoic acid C5-H_2, biotin C3'-H_2), 1.35-1.21 (7H, m, pentyl C1-H_3, hexanoic acid C4-H_2, biotin C2'-H_2), 0.89 (3H, t, *J* 7.1, pentyl C5-H_3), δ_C (75 MHz, $CDCl_3$): 173.1, 172.8, 163.5 (biotin C2), 153.1 (triazole), 151.1, 150.6, 134.4 (triazole), 123.4 (N-Ar-N C2), 118.6 (N-Ar-O C2), 116.6 (N-Ar-N C3), 115.2 (N-Ar-O C3), 84.1, 71.0, 70.8, 70.6, 69.0, 61.8, 60.4, 60.2, 56.0, 55.5, 50.7, 49.2, 40.5, 39.1, 38.9, 35.9, 34.3, 29.0, 28.1, 28.0, 26.3, 25.6, 24.5, 22.4, 17.4, 10.5; δ_F (282 MHz, $CDCl_3$): -109.0 (1F, d, *J* 7.5), -110.5 (1F, d, *J* 7.5); *m/z* (ES): Found $M+H^+$ 1040.499815, expected for $[C_{53}H_{67}F_2N_{11}O_7S + H]$ 1040.498647.

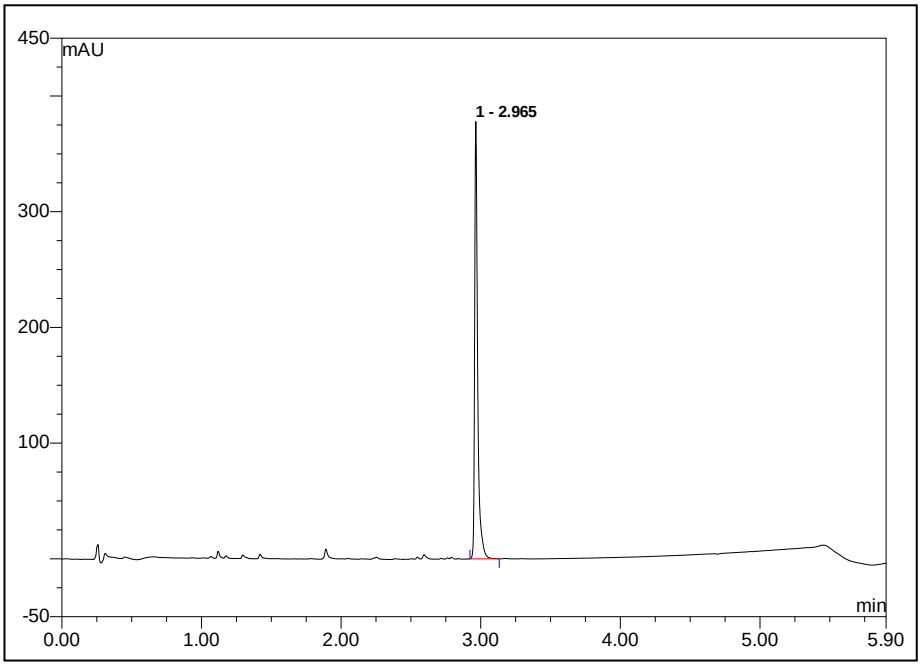
(2*S*,3*S*)-3-{4-[4-(4-{4-[(3*R*, 5*R*)-5-(2,4-difluoro-phenyl)-5-([1,2,4]triazol-1-ylmethyl)tetrahydrofuran-3-ylmethoxy]phenyl}piperazin-1-yl)phenyl]-5-oxo-4,5-dihydro-[1,2,4]triazol-1-yl}pent-2-yl 14-Azido-3,6,9,12-tetraoxatetradecanoate (140)



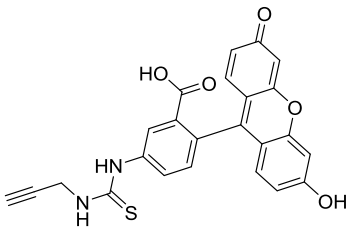
Posaconazole (200 mg, 0.285 mmol), 14-Azido-3,6,9,12-tetraoxatetradecanoic acid (0.5M in TBME, 856 µL, 0.428 mmol), EDC (82 mg, 0.428 mmol), DMAP (52 mg, 0.428 mmol) and DMF (3 mL) were combined and heated to 50 °C for 16 h. After cooling the reaction was purified by mass-directed HPLC. Product-containing fractions were combined and acetonitrile removed under reduced pressure. Water was then removed by lyophilisation to yield **140** (109.6 mg, 0.114 mmol, 40%) as an orange solid. δ_H (300 MHz, $CDCl_3$): 8.11 (1H, s, oxotriazole C3-H), 7.80 (1H, s, triazole H), 7.63 (1H, s, triazole H), 7.44-7.36 (3H, m, N-Ar-N C2-H, difluorophenyl C-H), 7.02 (2H, d, *J* 9.0, N-Ar-N C3-H), 6.93 (2H, d, *J* 8.9, N-Ar-O C2-H), 6.88-6.80 (2H, m, difluorophenyl C-H), 6.78 (2H, d, *J* 9.0, N-Ar-O C3-H), 5.34-5.28 (1H, m, pentyl C2-H), 4.65 (1H, d, *J* 14.4, triazolylmethyl CH_2), 4.52 (1H, d, *J* 14.4, triazolylmethyl CH_2), 4.23-4.18 (1H, m, pentyl C3-H), 4.12 (1H, t, *J* 8.0, tetrahydrofuran C2-H), 3.78 (1H, dd, *J* 8.7, 6.8, tetrahydrofuran C2-H), 3.73-3.60 (16H, m, tetrahydrofuranylmethoxy CH_2, 7 x PEG CH_2), 3.39-3.34 (6H, m, piperazine C2-H_2, PEG C14-H_2), 3.25-3.21 (4H, m, piperazine C3-H_2), 2.64-2.51 (2H, m, tetrahydrofuran C3-H, tetrahydrofuran C4-H), 2.08 (1H, dd, *J* 12.9, 8.4, tetrahydrofuran C4-H'), 1.97-1.88 (1H, m, pentyl C4-H), 1.83-1.74 (1H, m, pentyl C4'-H), 1.33 (3H, m, pentyl C1-H3), 0.89 (3H, t, *J* 7.1, pentyl C5-H3); ν_{max}/cm^{-1} (solid): 2102 (N_3),1710, 1690 (C=O), 1509; *m/z* (ES): Found

$M+H^+$ 960.456668, expected for [$C_{47}H_{59}F_2N_{11}O_9$ + H] 960.453805. HPLC retention time: 2.965 min.

HPLC trace of **140**

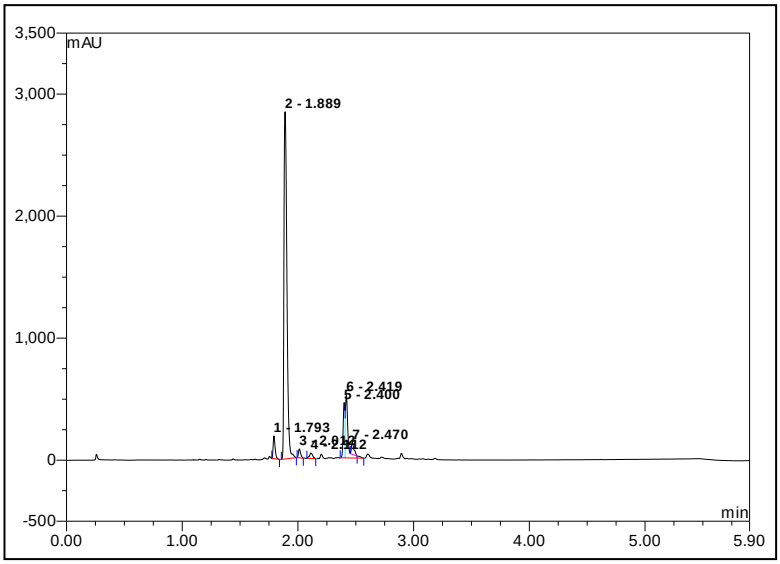


N-fluoresceinyl-N'-2-propyn-1-ylthiourea (142) [146]

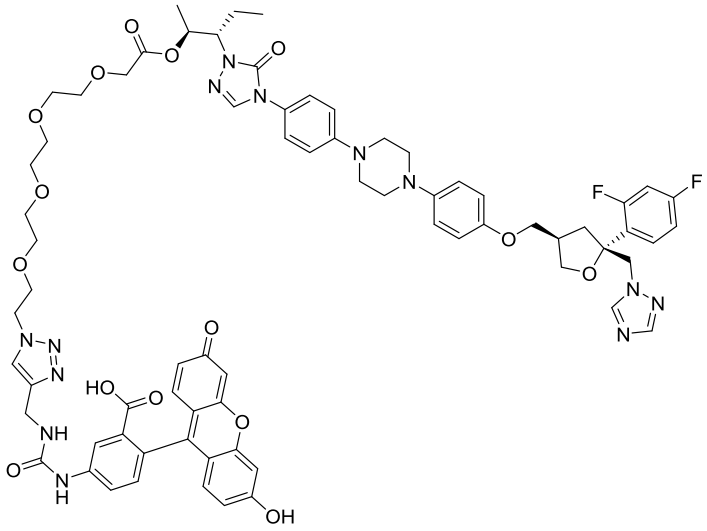


Fluorescein isothiocyanate (50 mg, 0.128 mmol) was dissolved in DMSO (1 mL) and propargyl amine (12.3 µL, 0.192 mmol) added, with an instant colour change to red. After ten minutes the reaction mixture was purified by mass directed HPLC. Product-containing fractions were combined and acetonitrile removed under reduced pressure. Water was then removed by lyophilisation to yield **142** (17 mg, 0.038 mmol, 30%) as an orange solid. Characterisation was in accordance with literature.[146] *m/z* (ES): Found M+Na$^+$ 467.067886, expected for [$C_{24}H_{16}N_2O_5S$ + Na] 467.067213. HPLC retention time 1.889 (73%)

HPLC of **142**

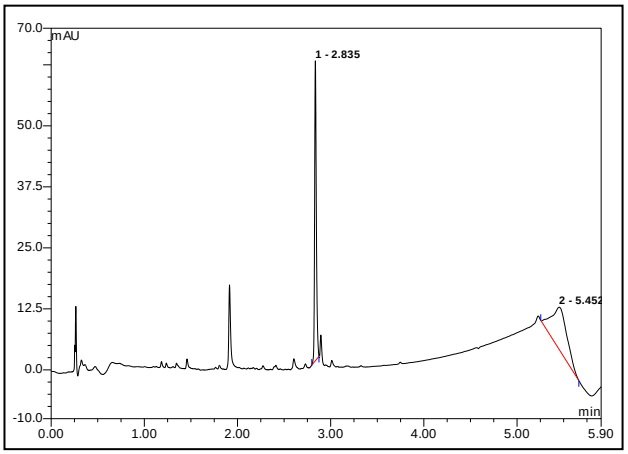


(2*S*,3*S*)-3-{4-[4-(4-{4-[(3*R*, 5*R*)-5-(2,4-difluoro-phenyl)-5-([1,2,4]triazol-1-ylmethyl)tetrahydrofuran-3-ylmethoxy]phenyl}piperazin-1-yl)phenyl]-5-oxo-4,5-dihydro-[1,2,4]triazol-1-yl}pent-2-yl 14-(4-[[(fluoresceinylamino)carbonyl]aminomethyl]-[1,2,3]triazinyl-3,6,9,12-tetraoxatetradecanoate (143)



130 (5 mg, 0.005 mmol), **142** (4.45 mg, 0.010 mmol), copper (II) sulfate (2.5 mg, 0.010 mmol), sodium ascorbate (5 mg, 0.025 mmol) and DMSO (500 µL) were combined and stirred for 6 h at rt. Half of the reaction mixture was then purified by mass directed HPLC. Product-containing fractions were combined and acetonitrile removed under reduced pressure. Water was then removed by lyophilisation to yield **143** (2.8 mg, 0.002 mmol, 80%) as an orange solid. *m/z* (ES): Found [M+H$^+$] 1388.554642, expected for [$C_{71}H_{76}F_2N_{13}O_{15}$] 1388.55464. HPLC retention time 2.835 min

HPLC of **143**



7.3 Peptides

7.3.1 General Procedures

General Procedure 1: Solid Phase Peptide Synthesis

Peptides were synthesised using Rink amide or pre-loaded glycine chlorotrityl resin depending on C-terminal functionality. During all incubations the tubes were placed on a rotator. Couplings were performed using Fmoc-protected amino acids (5 eq.), HCTU (5 eq.) and DIPEA (5 eq.) with 1 h incubation time at rt The resin was washed with DMF (3 × 3 mL, 2 min incubation), with 10% piperidine in DMF (5 × 3 mL, 2 min incubation), DMF (5 × 3 mL, 2 min incubation). After the final coupling reaction the resin was washed with DMF (3 × 3 mL, 2 min incubation), DCM (3 × 3 mL, 2 min incubation) and MeOH (3 × 3 mL, 2 min incubation). The peptide was cleaved and deprotected using a cleavage cocktail (TFA:TIS:EDT:H_2O, 95:1:2.5:2.5, total volume = 3 mL) with incubation for 1 h. The cleavage mixture was eluted into ice-cold ether (40 mL) and the precipitate collected by centrifugation (4000 r.p.m., 5 min). The precipitate was resuspended and washed with ice-cold ether (5 × 15 mL). After drying overnight the peptide was dissolved in water and lyophilised to yield final product.

General Procedure 2: Conversion of cysteine to dehydroalanine in peptides using 27

The peptide was dissolved in water (0.4 mL) and TCEP (0.5 eq.) added. A solution of 2,5-dibromoadipamide **27** (10 eq.) in DMF (0.1 mL) was added before the entire solution was transferred to a vial containing K_2CO_3 (5 eq.). The vial was incubated in a 37 °C shaking incubator at 140 rpm. Reaction was monitored by LCMS.

7.3.2 Linear Peptides

H-ACSDRFRNCPADEALCG-NH_2 (36)

Synthesised using general procedure 1 *m/z* (ES): Found $M+2H^+$ 913.890594, expected for $[C_{72}H_{117}N_{25}O_{25}S_3]$ 913.890180.

H-AXSDRFRNXPADEALXG-NH_2 (X=Dha) (38)

Crude peptide H-ACSDRFRNCPADEALCG-NH_2 **36** (10 mg, 5.5 µmol) was incubated with TCEP (0.7 mg) in H_2O (4.4 mL) for 45 min before addition of methyl 2,5-dibromovalerate (90 mg, 328 µmol) in DMSO (4.4 mL) and K_2CO_3 (113 mg, 821 µmol). The reaction was incubated at 37 °C for 3 h before purification by mass-directed HPLC (5-50% MeCN:H_2O gradient + 0.1% formic acid). Product-containing fractions were combined, concentrated under reduced pressure before removal of water by lyophilisation to yield pure product **38** as a white flocculent solid (3.5 mg, 37%). *m/z* (ES): Found $M+2H^+$ 862.908689, expected for $[C_{72}H_{111}O_{25}N_{25}]$ 862.908599.

H-ACSDRFRNCPADEALCG-NH_2 (8 stereoisomers) (individual isomers of 36/precursors of 37+81-87)

Synthesised using general procedure 1 by Patrick Foley and used crude.

H-AYCDG-OH (41)

Synthesised using general procedure 1 and used without further purification. *m/z* (ES): Found $M+H^+$ 528.175738, expected for $[C_{21}H_{29}O_9N_5S + H]$ 528.175875. HPLC: 1.56 min (5-95% Ascentis Peptide Gradient).

H-AYXDG-OH (X=Dha) (42)

General procedure 2 from **41**, observed by LCMS – not isolated.

H-GCSDYVCGTHACA-OH (45)

Synthesised using general procedure 1 and used crude. Observed $M+2H^+$ by LCMS 643.7

H-GXSDYVXGTHAXA-OH (46)

General procedure 2 from **45**, observed by LCMS – not isolated

H-ALCYG-OH (70)

Synthesised using general procedure 1 and used without further purification.

H-LTFCEYWAQLCSAA-OH (74)

Purchased from Thermo Scientific (>80% purity)

Ac-AC(tBS)GAW(Boc)AGAC(Trt)G-OH (150)

Synthesised using general procedure 1 omitting cleavage steps (0.055 mmol scale) on pre-loaded Gly chlorotrityl resin. Acetylation was carried out through addition of acetic anhydride (26 µL, 0.275 mmol) and DIPEA (102 µL, 0.275 mmol) and incubation on the rotator for 1 h. Acetylation was followed by washes with DMF (3 × 2 mL), DCM (3 × 2 mL), and MeOH (3 × 2 mL) as in procedure 1.

Deprotection of tBS protecting group on peptide 150 to form 151

To a sample of peptide **150** (approx. 0.055 mmol) was added β-mercaptoethanol (38.8 µL, 0.55 mmol) and DMF (2 mL) and the tube incubated on a rotator for 18 h with samples taken after 1 h, 4 h and 18 h. Samples for LCMS were taken by drainage of the liquid phase, washing the resin with DMF (2 × 2 mL) and adding a small amount of resin to a sample of 100 µL TFA or hexafluoroisopropanol. Further deprotection attempts were made on the same peptide sample as follows: 1. PBu_3 (15 µL, 0.061 mmol), H_2O (20 µL) and DMF (2 mL), rt 1 h. 2. PBu_3 (150 µL, 0.61 mmol), H_2O (150 µL), DCM (300 µL) and DMF (300 µL), rt 18 h. 3. DTT (85 mg, 0.55 mmol), NEt_3 (200 µL), DMF (750 µL), H_2O (250 µL), rt 16 h. 4. DTT (850 mg, 5.5 mmol) K_2CO_3 (760 mg, 5.5 mmol), DMF (500 µL) and H_2O (500 µL), 37 °C, 2 h. Each deprotection attempt had samples taken for LCMS as above.

7.3.3 Cyclised Peptides

Peptide Cyclisation with tris(bromomethyl)benzene (TBMB) to generate cyclised peptides 37/81-87

Cyclisation reaction was carried out by Patrick Foley using the protocol in Heinis et al.[14] Crude peptides (1.83 mg, 1 µmol) were dissolved in 70% (v/v) ammonium bicarbonate (20 mmol, 0.3 mL, pH 8) 30% acetonitrile (0.3 mL) with TBMB (0.36 mg, 1 µmol) added and agitated for an hour. The cyclised peptides were then purified using mass directed HPLC (5-95% MeCN + 0.1 % formic acid, 8 min) to give: PK15 (0.58 mg, 0.29 µmol, 29%); LLD

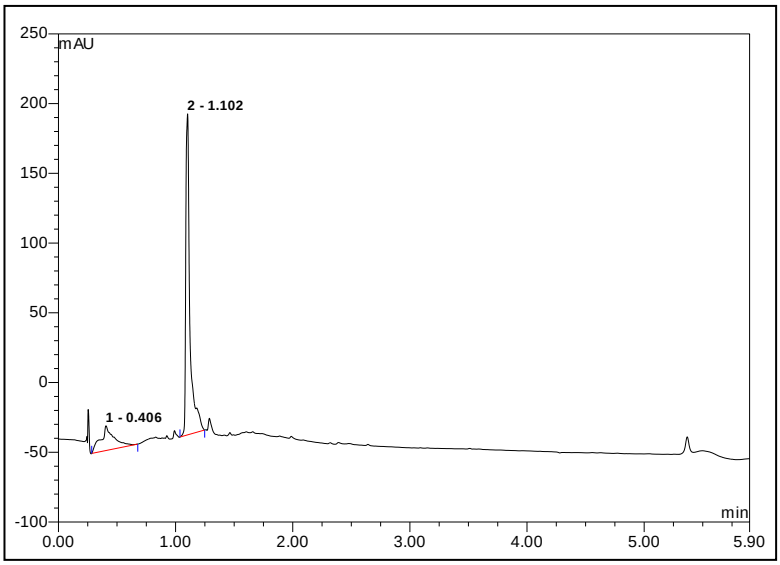
(0.19 mg, 0.10 µmol, 10%); LDL (0.32 mg, 0.16 µmol, 16%); LDD (0.46 mg, 0.24 µmol, 24%); DDD (0.50 g, 0.257 µmol, 26%); DLL (0.50 g, 0.257 µmol, 26%); DLD (0.3 mg, 0.155 µmol, 16%); DDL (0.5 mg, 0.257 µmol, 26%).

Peptide Cyclisation with tris(mercaptomethyl)benzene (TMMB) to make peptide mixture 40

Peptide **38** (2 mg, 1.15 µmol) was combined with 1,3,5-tris(mercaptomethyl)benzene **39** (0.3 mg, 1.39 µmol) and K_2CO_3 (1.6 mg, 11.59 µmol) in 1:1 H_2O:DMF (1 mL). The reaction was incubated at 37 °C for 3 h before purification by mass-directed HPLC. Product-containing fractions were combined, concentrated under reduced pressure before removal of water by lyophilisation to yield pure product **40** as a white solid (1.1 mg, 43%) A quarter of the reaction mixture was kept crude for testing in the kallikrein inhibition assay. *m/z* (ES): Found $M+2H^+$ 970.910013, expected for $[C_{81}H_{123}O_{25}N_{25}S_3]$ 970.913655.

PK15 cyclisation with 110 (Synthesis of 121)

Peptide **36** (2.0 mg, 1.1 µmol) was combined with TCEP (0.12 mg, 0.5 µmol) and K_2CO_3 (0.45 mg, 3.3 µmol) in water (500 µL) before addition of a solution of **110** (0.56 mg, 1.1 µmol) in DMF (500 µL). Complete reaction was observed after 5 minutes. HPLC purification yielded **121** (1.4 mg, 67%) as a white solid.



Peptide 74 cyclised with Barbas core 126 (Synthesis of 128)

Peptide **74** (1 mg, 0.6 µmol) was incubated with **126** (0.4 mg, 0.9 µmol) and K_2CO_3 (0.1 mg, 0.7 µmol) in DMF (0.5 mL) for 2 h at 37 °C. The reaction was purified by mass directed HPLC to yield **128** (0.2 mg, 20%)

7.4 Assays

7.4.1 LCMS assay to optimise conditions of dehydroalanine formation using 27 and 63

Peptide (1.2 mg, 747 nmol) was pretreated with TCEP (0.09 mg, 373.5 nmol, 0.5 eq.) in water (300 µL), after half an hour at 37 °C this solution was used as the peptide stock solution. In a 96-well plate (polypropylene, Greiner Bio-one 651201), reactions were set up with the following conditions: peptide **6** 0.623 mM, base (K_2CO_3 or NH_4HCO_3, 62.3 mM), dibromo compound (**3** or **4**, ranging from 1.2 mM to 62.3 mM), 1:1 water:organic (DMF/DMSO/MeCN) total volume 20 µL. After incubation at 37 °C for 3 hours, the reactions were examined using LCMS with blank runs as spacers between samples. Ions corresponding to the product **7** and stapled by-products **8**/**9** were extracted using Bruker ESI Compass data analysis software and the extracted ion peaks integrated. The ratio of these MS peak integrals of **7** to **8**/**9** was plotted to estimate levels of both product and stapled by-product.

7.4.2 Kallikrein Inhibition Assay

The inhibitory activity (IC_{50}) of each bicyclic peptide was determined by measuring the residual activity of HPK upon incubation for 30 min at room temperature with concentrations of bicyclic peptide ranging from 10 µM to 0.05 nM. The activity of HPK (2×10^{-4} µM) was measured using fluorogenic substrate Z-Phe-Arg-AMC (0.1 mM) in Tris (10 mM, pH 7.4) supplemented with NaCl (150 mM), $MgCl_2$ (10 mM), $CaCl_2$ (1 mM), BSA (0.1%, *w*/*v*), Triton X-100 (0.01%, *v*/*v*) and DMSO (5%, *v*/*v*). Excitation at 355 nm, emission 460 nm. Measurements were taken every 30 seconds. A linear fit of the data points from 1.5 min to 30 min gave the rate of substrate cleavage. Standard errors were calculated from the rates across triplicate wells. Rates were plotted against inhibitor concentration and a logistic fit used to determine the IC_{50} using equation [1]:

$$[1] \qquad y = A2 + (A1 - A2)/(1 + (x/x0)\text{^}p)$$

7.4.3 Fluorescence anisotropy assay

Adhirons were expressed by Thomas Taylor in the BSTG at Leeds. Adhirons were concentrated in Amicon 3K spin filters at 4000 r.p.m. Appropriate dilution series (A4C: 556 µM to 0.5 µM) were set up in triplicate before addition of probe **143** (2 µM DMSO stock, 5 µL) Total volume 100 µL, 5% DMSO, 100 nM probe. Fluorescence Anisotropy measurements were taken in 96-well plate format using an Envision 2103 Multilabel Reader.

G-factor was calculated by the average of the readings on a blank probe sample where G=S/P

Anisotropy (r) of individual wells was then calculated using equation [2]

$$r = \frac{(S - G*P)}{(S + 2G*P)} \quad [2]$$

The total fluorescence (S+2G*P) was plotted vs Adhiron concentration to confirm there was no observable quenching.

The anisotropy values for the triplicate wells were averaged and standard errors calculated. These values were used to produce a plot of anisotropy vs Adhiron concentration. This plot was fitted with a logistic method (fixed power of 1) and this used to convert the anisotropy values to fraction bound. A hyperbolic fit of fraction bound vs Adhiron concentration gave the K_d value.

7.5 Macrocyclic Sampling Modelling Protocol

2D representations of PK15 stereoisomers were created using ChemDraw with appropriate stereochemistry at all stereocentres. These were then converted to a random 3D structure using PipelinePilot. This 3D structure was loaded into MacroModel (Schrodinger Inc.), hydrogens were automatically added via an initial energy minimisation and the macrocycle sampling script used with the following settings:

Force field	OPLS_2005
Energy window for saving structures	10 kcal/mol
Eliminate conformers within RMSD	0.75 Å
Simulation cycles	5000
LLMOD search steps	5000

The modelling procedure undertook 0.5 ps simulation at 1000 K followed by 0.5 ps at 300 K and then energy minimisation using Macromodel's large scale low-mode search (LLMOD). The twenty lowest energy conformation molecular structures were exported and the pdb files edited using Notepad++ to allow visualisation in pymol. Structures were overlayed in pymol using the align command.

The UK18 pdb structure was extracted from the pdb file 3QN7 and loaded into MacroModel and the same modelling protocol as above undertaken.

7.6 Molecular Biology

7.6.1 Buffers, Media and Reagents

Table 11: Media, buffers and reagents used in this work

	Recipe/Supplier
Media	**(per litre where applicable)**
LB	25 g LB broth mix
LB-agar	15 g agar, 25 g LB
2TY	10 g yeast extract, 16 g tryptone, 5 g NaCl
Buffers	
TE Buffer	10 mM Tris, 1mM EDTA, pH 8.0
PEG-NaCl solution	20% (w/v) PEG 8000, 2.5 M NaCl
PBS-T	137 mM NaCl, 10 mM phosphate, 2.7 mM KCl, 0.1% Tween-20, pH 7.4
TBMB Reaction Buffer	20 mM NH_4HCO_3, 5 mM EDTA, pH 8.0
TCEP Buffer	TBMB Reaction buffer + 1 mM TCEP
10× blocking buffer	Sigma Cat. No. B6429
2× blocking buffer	10× Buffer diluted in PBS-T
Reagents	
TMB	SeramunBlau® fast TMB/substrate solution) (Seramun, Cat. No. S-001-TMB)
Anti-Fd-Bacteriophage-HRP	(Seramun Diagnostica GmbH, Cat. No. A-020-1-HRP)

7.6.2 General procedures

General Procedure 1: Transformation of *E. coli* with plasmid DNA

Cells and plasmid were mixed together in 1.5 mL Eppendorf tubes and incubated on ice for 5 minutes. Samples were then immersed in a 42 °C water bath for 45 seconds before immediate cooling on ice for 5 minutes. LB (0.5 mL) was added to each sample and the samples incubated at 37 °C for 45-60 minutes. A 20 µL sample was plated on LB-agar containing the relevant antibiotic. The remaining transformation sample was centrifuged at 2000 *g* for 2 minutes and the cell pellet resuspended in 100 µL before plating on LB-agar. Spreading was achieved via the use of sterilised glass shot.

7.6.3 1st Generation Construct

The construct (dsba)PK15-g3p* (pUC57) was purchased from genscript. XL10 *E. Coli* (15 µL) were mixed with plasmid (1.5 µL) before transformation via general procedure 1. Colonies were minicultured in LB-Amp (5 mL) overnight at 37 °C. DNA was obtained via miniprep following the standard instructions for the Qiagen miniprep kit.

PCR reactions

PCR to install different leader sequences

PCR reactions were set up using primer indicated in Table 12 (0.5 µL), reverse primer RP-PK15-g3p*(NheI) (0.5 µL), Pwo DNA polymerase (0.2 µL), Pfu buffer (5 µL), DNTPs mix (0.4 µL), template (dsba)PK15-g3p*(puc57) (0.2 µL), DMSO (2.5 µL) and water (40.7 µL). PCR reactions were run under the conditions listed in Table 12. Reactions were examined and purified via gel electrophoresis using agarose gel (1.6%), bands excised by scalpel using ethidium bromide and UV illumination and DNA purified following the standard procedure in the Qiagen Gel extraction kit.

Table 12: PCR conditions for the first PCR step in first generation plasmid construction

Product	Primer	Conditions	Anneal Temp (°C)
OmpA-PK15-g3p*	OmpA-PK15	b	54.4
MalE-PK15-g3p*	MalE-PK15	a	53.6
STII-PK15-g3p*	STII-PK15	b	60.3
PhoA-PK15-g3p*	PhoA-PK15	a	53.6
M13KE-PK15-g3p*	M13KE-PK15	a	53.6

[a] 5% DMSO, 30 cycles: 30 s denaturation (95 °C), 60 s anneal, 210 s extension (72 °C)
[b] 5% DMSO, 35 cycles: 30 s denaturation (95 °C), 60 s anneal,120 s extension (72 °C)

PCR to install RBS site

PCR reactions were set up using primer indicated in Table 13 (0.5 µL), reverse primer RP-PK15-g3p*(NheI) (0.5 µL), Pwo DNA polymerase (0.2 µL), Pfu buffer (5 µL), DNTPs mix (0.4 µL), template (product of PCRs above – listed in Table 13) (1 µL), DMSO (2.5 µL) and water (39.9 µL). PCR reactions were run under the conditions listed in Table 13. Reactions were examined via gel electrophoresis using agarose gel (1.6 %) and the remainder of the reaction purified via PCR clean-up: following the standard procedure in the Qiagen Gel extraction kit omitting the gel-running steps. Concentrations were measured using Nanodrop UV absorption readings at 260 nm.

Table 13: PCR conditions for the 2nd PCR step to reinsert a ribosome binding site

Product	Primer	Template	Conditions
R-MalE-PK15-g3p*	RBS-MalE	MalE-PK15-g3p*	a
R-PhoA-PK15-g3p*	RBS-PhoA	PhoA-PK15-g3p*	a
R-M13KE-PK15-g3p*	RBS-M13KE	M13KE-PK15-g3p*	a

[a] 5% DMSO, 35 cycles: 30 s denaturation (95 °C), 60 s anneal (53.4 °C),120 s extension (72 °C)

Cloning into pBSTG1

pBSTG1 [HindIII and EcoRI digested]

pBSTG1 vector (15 µL, 1.5 µg) was digested using restriction enzymes EcoRI HF (1 µL) and HindIII HF (1 µL) in NEB restriction digest buffer 4 (5 µL) and water (28 µL). The digestion mixture was incubated at 37 °C for 2.5 hours.

(dbsa)PK15-g3p*[HindIII and EcoRI digested]

(dbsa)PK15-g3p*(pUC57) vector (5.35 µL, 1.5 µg) was digested using restriction enzymes EcoRI HF (1 µL) and HindIII HF (1 µL) in NEB restriction digest buffer 4 (5 µL) and water (37.65 µL). The digestion mixture was incubated at 37 °C for 2.5 hours.

(dsba)PK15-g3p*(pBSTG1)

Both pBSTG1(HindIII+EcoRI) and (dsba)PK15-g3p*(pUC57)(HindIII+EcoRI) were purified by gel electrophoresis. Ligation of the two components was set up (insert 17.5 µL, vector 4 µL, T4 10× buffer 2.5 µL, T4 ligase 1 µL) at rt for 10 min. Ligation mix (2.5 µL) was used to transform XL10 *E. coli* following general procedure 1. Isolated vector (dsba)PK15-g3p*(pBSTG1) was confirmed by Sanger sequencing (GATC Biotech).

g3p*(pBSTG1) [HindIII and NheI digested]

(dbsa)PK15-g3p*(pBSTG1) vector (4.28 µL, 1.5 µg) was digested using restriction enzymes NheI HF (1 µL) and HindIII HF (1 µL) in NEB restriction digest CutSmart buffer (5 µL) and water (38.21 µL). The digestion mixture was incubated at 37 °C for 1 hour. Phosphatase and $ZnCl_2$ were then added followed by incubation at 37 °C for 45 min.

(MalE/PhoA/M13KE)PK15-g3p*(pBSTG1)

PCR products R-MalE-PK15-g3p* (0.5 µg) was digested with NheI HF (1 µL) and HindIII HF (1 µL) in SmartCut buffer (5 µL) and water (26.3 µL) at 37 °C for 1 hour. The same conditions were used with R-PhoA-PK15-g3p* and R-M13KE-PK15-g3p*. Digested products were purified via PCR clean-up. Ligation of digested genes (13.3 ng) and g3p*(pBSTG1) (20 ng) was undertaken using T4 ligase (1 µL) in T4 buffer (10×, 2 µL) and water (total reaction volume 20 µL). Reaction proceeded at rt for 10 min. Ligation mix (1 µL) was used to transform XL10 *E. coli* (10 µL) following general procedure 1. Isolated vectors (MalE/PhoA/M13KE)PK15-g3p*(pBSTG1) were confirmed by Sanger sequencing (GATC).

Phage Production

ER *E. coli* cells were transformed with one of the leader sequence variants (dsba/MalE/PhoA/M13KE)PK15-g3p*(pBSTG1) following general procedure 1 and colonies grown on LB-Agar-Amp. Minicultures were grown in 2TY-Amp (5 mL) overnight. Minicultures were used to seed 2TY-Amp cultures (50 mL) at OD_{600} = 0.1. Cultures were incubated at 37 °C and 170 rpm until OD_{600} = 0.5 at which point IPTG (0.5 M, 100 µL) was added. Cultures were incubated at 37 °C and 170 rpm for 90 minutes. Each culture was then split in two and either Hyperphage (0.47 mL, MOI=20) or Helperphage (1 ul, MOI = 15) were added. Cultures were incubated at 37 °C and 90 rpm for 1 hour before addition of Kanamycin (24 µL). Cultures were then incubated at 25 °C and 170 rpm overnight.

Phage cultures had bacteria removed by centrifugation (3000 rpm, 10 min). Supernatant (25 mL) was decanted into a PEG solution (25% PEG, 2.5 M NaCl, 6.25 mL) and stored at 4 °C for 2 hours. Phage pellet was obtained by centrifugation (4600 rpm, 10 min). The pellet was resuspended in TE buffer (4 mL) and passed through a 0.45 µM filter.

Concentration of phage was assessed using formula [2]

[2] $$\frac{(A_{269}-A_{320})\times 6\times 10^{16}}{6400}$$

Absorptions were measured on a 1 in 10 dilution.

Test of phage infectivity

E. coli ER2738 were cultured to OD_{600} = 0.5 and split into batches (9 × 3 mL). Batches were infected with phage in an array of 4 leader sequences × 2 secondary phage types (MOI=100). Cultures were incubated at 37 °C and 90 rpm for 30 minutes. A dilution series (1 µL, 1 in 100 and 1 in 10000) of each culture was plated on LB-Agar-Amp and incubated at 37 °C overnight.

Modification of phage with TBMB

1 litre 2TY cultures of ER2738 cells transformed with either dsba-PK15-g3p*(pBSTG1) or PhoA-PK15-g3p*(pBSTG1) were grown at 37 °C and 190 rpm. At OD_{600} = 0.7 M13 Helper phage (40 µL) was added. Cultures were incubated at 37 °C and 90 rpm for 30 min before addition of Kan (1 mL). Cultures were then incubated at 25 °C and 170 rpm overnight. Bacteria were pelleted by centrifugation 4000 *g* for 15 min. Supernatant was decanted into PEG-NaCl solution (125 mL) and stored at 4 °C for 1 hour followed by ice incubation for 30 min. Precipitated phage was pelleted by centrifugation at 4000 *g* for 30 min. The pellet was resuspended in 20 mL of TCEP buffer (Table 11) and incubated at 37 °C for 1 h. The phage solution was concentrated to 1 mL and washed with reaction buffer (2 × 10 mL) before addition to reaction buffer (32 mL) and followed by addition of TBMB solution (50 µM, 8 mL). Reaction was incubated at 30 °C for 1 hour. PEG-NaCl solution (10 mL) was added and phage precipitated overnight at 4 °C. Phage pellet was obtained by centrifugation at 4000 *g* for 30 min. Pellet was resuspended in TE buffer (dsba 1 mL, PhoA 2 mL) and cell debris removed by centrifugation (10000 *g*, 10 min). Concentration was estimated using equation 1.

Kallikrein-binding phage ELISA

2× Blocking buffer was added to streptavidin coated 96-well plate and incubated at 37 °C overnight. The plate was washed 1× in PBS-T before addition of biotinylated human plasma kallikrein (1 in 500 in 2× blocking buffer, 50 µL) to each well before incubation at rt for 1 h. To control wells 50 µL of 2× blocking buffer was added in place of kallikrein. The plate was washed 1× in PBS-T before addition of phage stock (approx. 10^{15} phage/mL, 60 µL) to the 1st column, and TE buffer (40 µL) to columns 2-12. A 3-fold dilution series of phage was created by 20 µL transfer and mixing from column 1 toward 12. 10× Blocking buffer (10 µL) was added to each well and the plate incubated at rt for 1 h. The plate was washed 1× in PBS-T before addition of anti-phage HRP (1 in 1000 in 2× blocking buffer, 50 µL) to each well and incubation at rt for 1 h. The plate was washed 10× in PBS-T before addition of 50

µL TMB solution. The developed blue colour was measured by the A620 using a plate-reader.

7.6.4 2nd Generation Phage

Dsba-PK15-g3p(C-term truncation)

The second generation dsba-PK15-g3p(C) gene was digested from puc57 into pBSTG1 using NheI and NdeI. pUC57 vector (27.4 µL, 10 µg) was combined with NheI-HF (1 µL), NdeI (1 µL), SmartCut buffer (5 µL) and water (15.6 µL) and incubated at 37 °C for 1 hour. The digestion was monitored and purified by gel electrophoresis (1% agarose). Gel band containing digested insert was excised and purified following the Qiagen gel extraction protocol.

Recipient pBSTG1 plasmid was digested with NheI and NdeI. pBSTG1 plasmid (2 µg, 20 µL) was incubated with NheI-HF (1 µL) and NdeI (1 µL), Smart Cut buffer (5 µL), and water (16 µL) for 45 min at 37 °C. Antarctic phosphatase (2 µL) and $ZnCl_2$ (5 µL) were added and the reaction incubated at 37 °C for 45 min. Reaction was monitored and purified by gel electrophoresis (1% agarose). Gel bands were excised and DNA obtained via Qiagen gel extraction protocol.

Digested pBSTG1 vector (20 ng, 0.64 µL) and PK15 gene (15 ng, 0.55 µL) were combined with T4 10× buffer (1 µL), water (7.3 µL) and T4 DNA Ligase (0.5 µL). The reaction was incubated for 10 min at rt. 1.5 µL of ligation mix was used to transform XL10 *E. coli* following general procedure 1. Isolated vector (dsba)PK15-g3p(Ctrunc)(pBSTG1) was confirmed by Sanger sequencing (GATC).

Phage Production

ER2738 cells were transformed with (dsba)PK15-g3p(Ctrunc)(pBSTG1) plasmid following general procedure 1 with bacteria plated on LB-Amp agar. Minicultures (2TY, 5 mL) were grown overnight at 37 °C and used to seed 8 mL cultures (2TY, OD_{600} = 0.2) which were incubated at 37 °C and 230 rpm for 2 hours. M13K07 helper phage (0.32 µL) was added to each culture and the cultures incubated at 37 °C, 90 rpm for 30 min. Kanamycin (25 mg/mL, 16 µL) was added and the cultures incubated at 25 °C and 170 rpm overnight. Phage-infected cultures were centrifuged at 3500 *g* for 10 min before decanting the supernatant into PEG-NaCl solution (2 mL) – phage were precipitated at 4 °C overnight and pelleted by centrifugation at 4800 *g* for 30 min. The pellet was resuspended in TE buffer (320 µL) and centrifuged at 16000 *g* for 30 min to remove cell debris.

Anti-myc ELISA

Anti-myc antibody (Abcam AB9106) (50 µL, 1 in 100, 1 in 1000 and 1 in 10000 dilution in TE buffer) was added to Nunc-Immuno™ MaxiSorp™ strips (Thermo Scientific, Cat. No. 469949) and incubated overnight at 4 °C. Nunc strips were washed 3× in PBS-T before addition of 10× blocking buffer (250 µL) and incubation at 37 °C for 3.5 hours. Strips were washed 1× in PBS-T before addition of phage (2-fold dilution series starting from 10^{12} phage per mL, 50 µL per well) and incubation at rt. for 1 hour on a shaking platform. Strips were washed 1× in PBS-T before addition of anti-phage HRP (50 µL, 1 in 1000) and incubation at rt. for 1 hour on shaking platform. Strips were washed 10× with PBS-T before addition of 50 µL TMB solution. No signal was observed.

TEV cleavage

Phage samples were spin-concentrated in an Amicon 30 kDa centrifugal spin filter (4000 *g*) to 50 µL and incubated with TEV-protease (5 mol%) for 2 hours at rt. The reaction was spin-filtered and the flow-through analysed by high resolution MS but no signal was observed.

Appendix 1 – DNA Sequences

Gene Sequences

(dsba)PK15-g3p* (Genscript gene ordered)

AAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAAAAGATTTGG
TTGGCTCTGGCTGGTCTGGTTCTGGCGTTTTCTGCGTCTGCAGGTGGCTGCTCTG
ACCGTTTTCGCAACTGCCCGGCCGATGAAGCATTATGTGGCGGGTCTGGAGCGG
CAGAAACAGTTGAATCCTCACTCGCCAAACCGCATATCGAAGGTAGCTTTACGA
ACGTATGGAAAGATGACAAAACCTTAGACTGGTACGCCAACTATGAAGGCATTC
TGTGGAAGGCCACCGGCGTGGTTGTCATCACGGGTGACGAGACGCAAGTTTATG
CAACGTGGGTCCCTATCGGGCTAGCGATTCCCGAGAACGAAGGTGGCGGCAGC
GAAGGTGGTGGGAGTGAAGGCGGGGGTAGCGAAGGTGGGGGCACGAAACCGC
CGGAATATGGTGATACTCCTATCCCGGGCTATATTTACATTAACCCGCTGGACG
GCACGTACCCCCCGGGTACCGAACAGAACCCGGCAAATCCGAATCCATCCTTAG
AAGAAAGCCATCCATTAAACACGTTTATGTTCCAGAATAATCGTTTCCGCAATC
GTCAGGGCGCACTCACCGTCTATACGGGTACTGTGACACAGGGCACTGACCCGG
TAAAAACCTATTATCAGTATACCCCGGTTTCTTCTAAAGCGATGTATGACGCCTA
TTGGAATGGTAAATTCCGTGACGTCGCATTCCACTCTGGCTTCAACGAAGACCC
GCTGGTGGCGGAGTATCAAGGCCAATCGTCCTACCTGCCGCAGCCGCCCGTCAA
TGCGGGTGGTGGGAGCGGTGGTGGGTCCGGTGGAGGGTCTGAGGGGGGGGGCA
GCGAAGGGGGTGGAAGTGAGGGGGGCGGAAGCGAAGGGGGTGGCTCTGGTGG
GGGCAGCGGGTCCGGGGACTTCGACTACGAAAAAATGGCGAATGCGAATAAAG
GTGCAATGACCGAAAATGCAGATGAAAACGCACTGCAATCTGACGCCAAAGGC
AAGCTGGACTCAGTTGCTACGGATTACGGCGCCGCCATCGATGGGTTTATCGGC
GATGTGTCTGGCCTCGCGAATGGTAACGGAGCAACTGGCGATTTTGCTGGCTCG
AACTCCCAGATGGCGCAAGTGGGAGATGGTGATAACTCTCCTTTGATGAATAAC
TTCCGGCAATATCTGCCCTCCCTGCCTCAGTCAGTTGAATGTCGCCCGTACGTGT
TCGGTGCCGGTAAACCGTATGAATTTTCCATTGATTGCGACAAAATCAATCTGTT
CCGGGGCGTATTCGCTTTCCTGTTATATGTGGCTACCTTCATGTATGTTTTTTCCA
CGTTTGCGAACATCCTGCGCAACAAGGAGTCATAATAAGAATTC

2nd generation construct (NheI)-PK15-g3p(trunc)(NdeI)

GCTAGCGGTTGTTCGGATCGTTTCCGTAATTGTCCTGCGGATGAAGCTCTGTGTG
GAGGTAGTGGCGAACAGAAACTGATTTCTGAAGAAGACCTGGAAAATCTTTATT
TTCAAGGTGCGGCAGGTAACGGTGGTGGTAACGGTAACTCTGGTGGTGGTTCTG
GTACCGGCGGCTCCGGTTCCGGTGATTTTGATTATGAAAAAATGGCAAACGCTA

ATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACGCT
AAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGATGGTTTCA
TTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGG
CTCTAATTCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAAT
AATTTCCGTCAATATTTACCTTCTTTGCCTCAGTCGGTTGAATGTCGCCCTTATGT
CTTTGGCGCTGGTAAACCATATG

Primers used in PCR reactions

Leader Sequence Oligos

STII-PK15

GAAAAAAAATATCGCCTTTCTCCTCGCGTCAATGTTCGTATTCTCAATTGCTACG
AACGCATATGCGGGTGGCTGCTCTGACC

OmpA-PK15

GAAGAAAACCGCAATTGCAATCGCTGTCGCCTTGGCCGGGTTCGCAACCGTTGC
ACAAGCGGGTGGCTGCTCTGACC

PhoA-PK15

GAAGCAGAGCACAATCGCCCTTGCTTTGCTTCCCCTGCTGTTTACCCCAGTTACC
AAGGCGGGTGGCTGCTCTGACC

MalE-PK15

GAAAATCAAAACCGGTGCCCGTATTCTGGCCCTGAGTGCACTCACAACGATGAT
GTTTAGTGCTAGCGCCTTGGCGGGTGGCTGCTCTGACC

M13KE-PK15

GAAAAAACTGTTGTTCGCCATTCCCTTGGTGGTCCCTTTTTATAGCCACAGCGGT
GGCTGCTCTGACC

Overlap Region: GGTGGCTGCTCTGACC (Same for all 5) Tm 55.9

RBS Oligos

RBS-PelB

GCAATGAAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAGTAT
CTCCTGCCTACTG

Overlap Region: GAAGTATCTCCTGCCTACTG Tm 58.4

RBS-StII

GCAATGAAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAAAAA
AATATCGCCTTTCTCC

Overlap Region: GAAAAAAAATATCGCCTTTCTCC Tm 57.6

RBS-OmpA

GCAATGAAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAGAAA
ACCGCAATTGCAATC

Overlap Region: GAAGAAAACCGCAATTGCAATC Tm 58.4

RBS-PhoA

GCAATGAAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAGCAG
AGCACAATCGCC

Overlap Region: GAAGCAGAGCACAATCGCC Tm 59.5

RBS-MalE

GCAATGAAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAAATC
AAAACCGGTGCCC

Overlap Region: GAAAATCAAAACCGGTGCCC Tm 58.4

RBS-M13KE

GCAATGAAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAAAAA
CTGTTGTTCGCCATTC

Overlap Region: GAAAAAACTGTTGTTCGCCATTC Tm 59.2

Reverse Primer

RP-PK15-g3p*(NheI) Reverse Primer NheI site in PK15-g3p*

CGG GAA TCG CTA GCC CG Tm 59.8

References

1. EvaluatePharma, *World Preview 2014, Outlook to 2020*, 2014.
2. D. J. Craik, D. P. Fairlie, S. Liras and D. Price, *Chem. Biol. Drug. Des.*, 2013, **81**, 136-147.
3. T. Uhlig, T. Kyprianou, F. G. Martinelli, C. A. Oppici, D. Heiligers, D. Hills, X. R. Calvo and P. Verhaert, *EuPA Open Proteom.*, 2014, **4**, 58-69.
4. M. Baker, *Nature*, 2015, **521**, 274-276.
5. S. Johnson, D. Evans, S. Laurenson, D. Paul, A. G. Davies, P. Ko Ferrigno and C. Walti, *Anal. Chem.*, 2008, **80**, 978-983.
6. L. K. J. Stadler, T. Hoffmann, D. C. Tomlinson, Q. Song, T. Lee, M. Busby, Y. Nyathi, E. Gendra, C. Tiede, K. Flanagan, S. J. Cockell, A. Wipat, C. Harwood, S. D. Wagner, M. A. Knowles, J. J. Davis, N. Keegan and P. Ko Ferrigno, *Protein Eng. Des. Sel.*, 2011, **24**, 751-763.
7. A. Skerra, *Curr. Opin. Biotechnol.*, 2007, **18**, 295-304.
8. C. Tiede, A. A. Tang, S. E. Deacon, U. Mandal, J. E. Nettleship, R. L. Owen, S. E. George, D. J. Harrison, R. J. Owens, D. C. Tomlinson and M. J. McPherson, *Protein Eng. Des. Sel.*, 2014, **27**, 145-155.
9. A. Beck, T. Wurch, C. Bailly and N. Corvaia, *Nat. Rev. Immunol.*, 2010, **10**, 345-352.
10. M. Bartoloni, R. U. Kadam, J. Schwartz, J. Furrer, T. Darbre and J. L. Reymond, *Chem. Commun.*, 2011, **47**, 12634-12636.
11. L. Nevola and E. Giralt, *Chem. Commun.*, 2015, **51**, 3302-3315.
12. C. Falciani, L. Lozzi, A. Pini and L. Bracci, *Chem. Biol.*, **12**, 417-426.
13. T. A. Hill, N. E. Shepherd, F. Diness and D. P. Fairlie, *Angew. Chem. Int. Ed.*, 2014, **53**, 13020-13041.
14. C. Heinis, T. Rutherford, S. Freund and G. Winter, *Nat. Chem. Biol.*, 2009, **5**, 502-507.
15. S. Chen, D. Bertoldo, A. Angelini, F. Pojer and C. Heinis, *Angew. Chem. Int. Ed.*, 2014, **53**, 1602-1606.
16. Q. Chu, R. E. Moellering, G. J. Hilinski, Y.-W. Kim, T. N. Grossmann, J. T. H. Yeh and G. L. Verdine, *MedChemComm*, 2015, **6**, 111-119.
17. G. L. Verdine and G. J. Hilinski, in *Methods Enzymol.*, eds. K. D. Wittrup and L. V. Gregory, Academic Press, 2012, vol. Volume 503, pp. 3-33.
18. L. D. Walensky and G. H. Bird, *J. Med. Chem.*, 2014, **57**, 6275-6288.
19. D. J. Adams and G. Berecki, *Acta Biochim. Biophys.*, 2013, **1828**, 1619-1628.
20. D. J. Adams, B. Callaghan and G. Berecki, *Br. J. Pharmacol.*, 2012, **166**, 486-500.
21. C. J. Armishaw, J. Banerjee, M. L. Ganno, K. J. Reilley, S. O. Eans, E. Mizrachi, R. Gyanda, M. R. Hoot, R. A. Houghten and J. P. McLaughlin, *ACS Comb. Sci.*, 2013, **15**, 153-161.
22. Y. Mochizuki and N. Nemoto, *Methods Mol. Biol.*, 2012, **805**, 237-250.
23. S. Chen, I. Rentero Rebollo, S. A. Buth, J. Morales-Sanfrutos, J. Touati, P. G. Leiman and C. Heinis, *J. Am. Chem. Soc.*, 2013, **135**, 6562-6569.
24. S. Chen, R. Gopalakrishnan, T. Schaer, F. Marger, R. Hovius, D. Bertrand, F. Pojer and C. Heinis, *Nat. Chem.*, 2014, **6**, 1009-1016.
25. A. Tavassoli and S. J. Benkovic, *Angew. Chem. Int. Ed.*, 2005, **44**, 2760-2763.
26. L. Cascales and D. J. Craik, *Org. Biomol. Chem.*, 2010, **8**, 5035-5047.
27. A. M. Cole, T. Hong, L. M. Boo, T. Nguyen, C. Zhao, G. Bristol, J. A. Zack, A. J. Waring, O. O. Yang and R. I. Lehrer, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 1813-1818.
28. D. J. Craik, N. L. Daly, T. Bond and C. Waine, *J. Mol. Biol.*, 1999, **294**, 1327-1336.
29. M. Trabi and D. J. Craik, *Plant Cell*, 2004, **16**, 2204-2216.

30. C. W. Gruber, A. G. Elliott, D. C. Ireland, P. G. Delprete, S. Dessein, U. Goransson, M. Trabi, C. K. Wang, A. B. Kinghorn, E. Robbrecht and D. J. Craik, *Plant Cell*, 2008, **20**, 2471-2483.
31. A. R. Horswill and S. J. Benkovic, *Cell Cycle*, 2005, **4**, 552-555.
32. L. Cheng, T. A. Naumann, A. R. Horswill, S.-J. Hong, B. J. Venters, J. W. Tomsho, S. J. Benkovic and K. C. Keiler, *Protein Sci.*, 2007, **16**, 1535-1542.
33. T. S. Young, D. D. Young, I. Ahmad, J. M. Louis, S. J. Benkovic and P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11052-11056.
34. J. A. Kritzer, S. Hamamichi, J. M. McCaffery, S. Santagata, T. A. Naumann, K. A. Caldwell, G. A. Caldwell and S. Lindquist, *Nat. Chem. Biol.*, 2009, **5**, 655-663.
35. T. M. Kinsella, C. T. Ohashi, A. G. Harder, G. C. Yam, W. Li, B. Peelle, E. S. Pali, M. K. Bennett, S. M. Molineaux, D. A. Anderson, E. S. Masuda and D. G. Payan, *J. Biol. Chem.*, 2002, **277**, 37512-37518.
36. C. Chatterjee, M. Paul, L. Xie and W. A. van der Donk, *Chem. Rev.*, 2005, **105**, 633-684.
37. N. Garg, L. M. Salazar-Ocampo and W. A. van der Donk, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7258-7263.
38. M. A. Ortega, Y. Hao, Q. Zhang, M. C. Walker, W. A. van der Donk and S. K. Nair, *Nature*, 2015, **517**, 509-512.
39. P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, *Nat. Prod. Rep.*, 2013, **30**, 108-160.
40. P. J. Knerr and W. A. van der Donk, *Annu. Rev. Biochem.*, 2012, **81**, 479-505.
41. J. Lubelski, R. Khusainov and O. P. Kuipers, *J. Biol. Chem.*, 2009, **284**, 25962-25972.
42. Q. Zhang, Y. Yu, J. E. Velasquez and W. A. van der Donk, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 18361-18366.
43. W. Tang, G. Jimenez-Oses, K. N. Houk and W. A. van der Donk, *Nat. Chem.*, 2015, **7**, 57-64.
44. J. A. Majchrzykiewicz, J. Lubelski, G. N. Moll, A. Kuipers, J. J. Bijlsma, O. P. Kuipers and R. Rink, *Antimicrob. Agents Ch.*, 2010, **54**, 1498-1505.
45. N. A. Bindman, S. C. Bobeica, W. R. Liu and W. A. van der Donk, *J. Am. Chem. Soc.*, 2015, **137**, 6975-6978.
46. T. J. Oman, P. J. Knerr, N. A. Bindman, J. E. Velasquez and W. A. van der Donk, *J. Am. Chem. Soc.*, 2012, **134**, 6952-6955.
47. M. R. Levengood, P. J. Knerr, T. J. Oman and W. A. van der Donk, *J. Am. Chem. Soc.*, 2009, **131**, 12024-12025.
48. S. A. Kawamoto, A. Coleska, X. Ran, H. Yi, C. Y. Yang and S. Wang, *J. Med. Chem.*, 2012, **55**, 1137-1146.
49. D. Y. Jackson, D. S. King, J. Chmielewski, S. Singh and P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 9391-9392.
50. H. E. Blackwell and R. H. Grubbs, *Angew. Chem. Int. Ed.*, 1998, **37**, 3281-3284.
51. C. E. Schafmeister, J. Po and G. L. Verdine, *J. Am. Chem. Soc.*, 2000, **122**, 5891-5892.
52. Y.-W. Kim, T. N. Grossmann and G. L. Verdine, *Nat. Protoc.*, 2011, **6**, 761-771.

53. M. L. Stewart, E. Fire, A. E. Keating and L. D. Walensky, *Nat. Chem. Biol.*, 2010, **6**, 595-601.
54. D. J. Yeo, S. L. Warriner and A. J. Wilson, *Chem. Commun.*, 2013, **49**, 9131-9133.
55. G. H. Bird, E. Gavathiotis, J. L. LaBelle, S. G. Katz and L. D. Walensky, *ACS Chem. Biol.*, 2014, **9**, 831-837.
56. J. S. Quartararo, P. Wu and J. A. Kritzer, *ChemBioChem*, 2012, **13**, 1490-1496.
57. G. E. Mulder, J. A. Kruijtzer and R. M. Liskamp, *Chem. Commun.*, 2012, **48**, 10007-10009.
58. J. Zhang, J. Kemmink, D. T. Rijkers and R. M. Liskamp, *Chem. Commun.*, 2013, **49**, 4498-4500.
59. P. Timmerman, J. Beld, W. C. Puijk and R. H. Meloen, *ChemBioChem*, 2005, **6**, 821-824.
60. A. Angelini, L. Cendron, S. Chen, J. Touati, G. Winter, G. Zanotti and C. Heinis, *ACS Chem. Biol.*, 2012, **7**, 817-821.
61. S. Chen, J. Morales-Sanfrutos, A. Angelini, B. Cutting and C. Heinis, *ChemBioChem*, 2012, **13**, 1032-1038.
62. W. Lian, P. Upadhyaya, C. A. Rhodes, Y. Liu and D. Pei, *J. Am. Chem. Soc.*, 2013, **135**, 11990-11995.
63. V. Baeriswyl, H. Rapley, L. Pollaro, C. Stace, D. Teufel, E. Walker, S. Chen, G. Winter, J. Tite and C. Heinis, *ChemMedChem*, 2012, **7**, 1173-1176.
64. A. Angelini, P. Diderich, J. Morales-Sanfrutos, S. Thurnheer, D. Hacker, L. Menin and C. Heinis, *Bioconjugate Chem.*, 2012, **23**, 1856-1863.
65. A. Angelini, J. Morales-Sanfrutos, P. Diderich, S. Chen and C. Heinis, *J. Med. Chem.*, 2012, **55**, 10187-10197.
66. S. Chen, D. Gfeller, S. A. Buth, O. Michielin, P. G. Leiman and C. Heinis, *ChemBioChem*, 2013, **14**, 1316-1322.
67. T. Morioka, N. D. Loik, C. J. Hipolito, Y. Goto and H. Suga, *Curr. Opin. Chem. Biol.*, 2015, **26**, 34-41.
68. P. C. Reid, Y. Goto, T. Katoh and H. Suga, in *Ribosome Display and Related Technologies: Methods and Protocols*, eds. A. J. Douthwaite and H. R. Jackson, Springer New York, New York, NY, 2012, pp. 335-348.
69. T. Kawakami, T. Ishizawa, T. Fujino, P. C. Reid, H. Suga and H. Murakami, *ACS Chem. Biol.*, 2013, **8**, 1205-1214.
70. K. Ito, K. Sakai, Y. Suzuki, N. Ozawa, T. Hatta, T. Natsume, K. Matsumoto and H. Suga, *Nat Commun*, 2015, **6**.
71. N. K. Bashiruddin, M. Nagano and H. Suga, *Bioorg. Chem.*, 2015, **61**, 45-50.
72. K. Iwasaki, Y. Goto, T. Katoh and H. Suga, *Org. Biomol. Chem.*, 2012, **10**, 5783-5786.
73. M. Kang, K. Light, H. W. Ai, W. Shen, C. H. Kim, P. R. Chen, H. S. Lee, E. I. Solomon and P. G. Schultz, *ChemBioChem*, 2014, **15**, 822-825.
74. F. Tian, M. L. Tsao and P. G. Schultz, *J. Am. Chem. Soc.*, 2004, **126**, 15962-15963.
75. C. C. Liu, A. V. Mack, E. M. Brustad, J. H. Mills, D. Groff, V. V. Smider and P. G. Schultz, *J. Am. Chem. Soc.*, 2009, **131**, 9616-9617.
76. R. K. V. Lim, N. Li, C. P. Ramil and Q. Lin, *ACS Chem. Biol.*, 2014, **9**, 2139-2148.
77. M. A. Dwyer, W. Lu, J. J. Dwyer and A. A. Kossiakoff, *Chem. Biol.*, 2000, **7**, 263-274.
78. K. Fukunaga, T. Hatanaka, Y. Ito and M. Taki, *Mol. Biosyst.*, 2013, **9**, 2988-2991.
79. B. Santoso, S. Lam, B. W. Murray and G. Chen, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 5680-5683.
80. M. R. Jafari, L. Deng, P. I. Kitov, S. Ng, W. L. Matochko, K. F. Tjhung, A. Zeberoff, A. Elias, J. S. Klassen and R. Derda, *ACS Chem. Biol.*, 2014, **9**, 443-450.
81. S. Bellotto, S. Chen, I. Rentero Rebollo, H. A. Wegner and C. Heinis, *J. Am. Chem. Soc.*, 2014, **136**, 5880-5883.
82. I. Rentero Rebollo and C. Heinis, *Methods*, 2013, **60**, 46-54.
83. I. Kather, C. A. Bippes and F. X. Schmid, *J. Mol. Biol.*, 2005, **354**, 666-678.

84. S. Ng, M. R. Jafari, W. L. Matochko and R. Derda, *ACS Chem. Biol.*, 2012, **7**, 1482-1487.
85. S. Chen, J. Touati and C. Heinis, *Chem. Commun.*, 2014, **50**, 5267-5269.
86. A. D. Keefe and J. W. Szostak, *Nature*, 2001, **410**, 715-718.
87. F. P. Seebeck and J. W. Szostak, *J. Am. Chem. Soc.*, 2006, **128**, 7150-7151.
88. F. T. Hofmann, J. W. Szostak and F. P. Seebeck, *J. Am. Chem. Soc.*, 2012, **134**, 8038-8041.
89. J. E. Bock, J. Gavenonis and J. A. Kritzer, *ACS Chem. Biol.*, 2013, **8**, 488-499.
90. L. M. Tedaldi, M. E. Smith, R. I. Nathani and J. R. Baker, *Chem. Commun.*, 2009, 6583-6585.
91. A. Abbas, B. Xing and T.-P. Loh, *Angew. Chem. Int. Ed.*, 2014, **53**, 7491-7494.
92. A. Wawrzynow, J. H. Collins and C. Coan, *Biochemistry*, 1993, **32**, 10803-10811.
93. A. Raulf, C. K. Spahn, P. J. M. Zessin, K. Finan, S. Bernhardt, A. Heckel and M. Heilemann, *RSC Adv.*, 2014, **4**, 30462-30466.
94. K. A. Andersen, M. R. Aronoff, N. A. McGrath and R. T. Raines, *J. Am. Chem. Soc.*, 2015, **137**, 2412-2415.
95. J. Li, S. Jia and P. R. Chen, *Nat. Chem. Biol.*, 2014, **10**, 1003-1005.
96. K. F. Geoghegan and J. G. Stroh, *Bioconjug Chem*, 1992, **3**, 138-146.
97. N. M. Okeley, Y. Zhu and W. A. van Der Donk, *Org. Lett.*, 2000, **2**, 3603-3606.
98. E. Branigan, C. Pliotas, G. Hagelueken and J. H. Naismith, *Nat. Protoc.*, 2013, **8**, 2090-2097.
99. J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield and B. G. Davis, *Angew. Chem. Int. Ed.*, 2012, **51**, 1835-1839.
100. N. Timms, C. L. Windle, A. Polyakova, J. R. Ault, C. H. Trinh, A. R. Pearson, A. Nelson and A. Berry, *ChemBioChem*, 2013, **14**, 474-481.
101. F. Rowan, M. Richards, M. Widya, R. Bayliss and J. Blagg, *PLoS One*, 2014, **9**, e103935.
102. N. Haj-Yahya, H. P. Hemantha, R. Meledin, S. Bondalapati, M. Seenaiah and A. Brik, *Org. Lett.*, 2014, **16**, 540-543.
103. J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield and B. G. Davis, *Chem. Sci.*, 2011, **2**, 1666.
104. R. Nathani, P. Moody, M. E. Smith, R. J. Fitzmaurice and S. Caddick, *ChemBioChem*, 2012, **13**, 1283-1285.
105. D. H. Strumeyer, W. N. White and D. E. Koshland, Jr., *Proc. Natl. Acad. Sci. U. S. A.*, 1963, **50**, 931-935.
106. H. Weiner, W. N. White, D. G. Hoare and D. E. Koshland, *J. Am. Chem. Soc.*, 1966, **88**, 3851-3859.
107. K. Goodall and A. F. Parsons, *Tetrahedron Lett.*, 1995, **36**, 3259-3260.
108. M. J. Miller, *J. Org. Chem.*, 1980, **45**, 3131-3132.
109. J. C. Slootweg, E. F. van Herwerden, M. F. M. van Doremalen, E. Breukink, R. M. J. Liskamp and D. T. S. Rijkers, *Org. Biomol. Chem.*, 2015, **13**, 5997-6009.
110. Y. O. You, M. R. Levengood, L. A. Ihnken, A. K. Knowlton and W. A. van der Donk, *ACS Chem. Biol.*, 2009, **4**, 379-385.
111. T. J. Holmes, Jr. and R. G. Lawton, *J. Am. Chem. Soc.*, 1977, **99**, 1984-1986.
112. S. A. Burrage, T. Raynham and M. Bradley, *Tetrahedron Lett.*, 1998, **39**, 2831-2834.
113. S. Burrage, T. Raynham, G. Williams, J. W. Essex, C. Allen, M. Cardno, V. Swali and M. Bradley, *Chem.-Eur. J.*, 2000, **6**, 1455-1466.
114. K. Hashimoto, M. Sakai, T. Okuno and H. Shirahama, *Chem. Commun.*, 1996, 1139-1140.
115. J. Wang, S. M. Schiller and P. G. Schultz, *Angew. Chem.*, 2007, **119**, 6973-6975.
116. Y. Zhu, M. D. Gieselman, H. Zhou, O. Averin and W. A. van der Donk, *Org. Biomol. Chem.*, 2003, **1**, 3304-3315.
117. WO2005028449A1, 2005.

118. A. Kamal, K. V. Ramana, A. V. Ramana and A. H. Babu, *Tetrahedron: Asymmetry*, 2003, **14**, 2587-2594.
119. K. S. Watts, P. Dalal, A. J. Tebben, D. L. Cheney and J. C. Shelley, *J. Chem. Inf. Model.*, 2014, **54**, 2680-2696.
120. A. B. Tabor, *Bioorg. Chem.*, 2014, **55**, 39-50.
121. R. S. Narayan and M. S. Van Nieuwenhze, *Org. Lett.*, 2005, **7**, 2655-2658.
122. P. J. Knerr and W. A. van der Donk, *J. Am. Chem. Soc.*, 2012, **134**, 7648-7651.
123. R. Appel, *Angew. Chem. Int. Ed.*, 1975, **14**, 801-811.
124. P. Jutzi and K. Wojczykowski, *Synlett*, 2006, **2006**, 0039-0040.
125. E. Fluck and P. Meiser, *Chem. Ber.*, 1973, **106**, 69-77.
126. N. Toda, S. Asano and C. F. Barbas, *Angew. Chem. Int. Ed.*, 2013, **52**, 12592-12596.
127. T. Moeller and A. H. Westlake, *J. Inorg. Nucl. Chem.*, 1967, **29**, 957-960.
128. E. Fluck and P. Meiser, *Angew. Chem. Int. Ed.*, 1971, **10**, 653-653.
129. T. M. Klapötke, B. Krumm, M. Scherr, F. X. Steemann, K. Banert and Y.-H. Joo, *Chem. Eur. J.*, 2009, **15**, 11341-11345.
130. Y. S. Chang, B. Graves, V. Guerlavais, C. Tovar, K. Packman, K. H. To, K. A. Olson, K. Kesavan, P. Gangurde, A. Mukherjee, T. Baker, K. Darlak, C. Elkin, Z. Filipovic, F. Z. Qureshi, H. Cai, P. Berry, E. Feyfant, X. E. Shi, J. Horstick, D. A. Annis, A. M. Manning, N. Fotouhi, H. Nash, L. T. Vassilev and T. K. Sawyer, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E3445-3454.
131. Y. H. Lau, P. de Andrade, S.-T. Quah, M. Rossmann, L. Laraia, N. Sköld, T. J. Sum, P. J. E. Rowling, T. L. Joseph, C. Verma, M. Hyvönen, L. S. Itzhaki, A. R. Venkitaraman, C. J. Brown, D. P. Lane and D. R. Spring, *Chem. Sci.*, 2014, **5**, 1804.
132. J. Speck, K. M. Arndt and K. M. Müller, *Protein Eng. Des. Sel.*, 2011, **24**, 473-484.
133. A. R. M. Bradbury, S. Sidhu, S. Dubel and J. McCafferty, *Nat Biotech*, 2011, **29**, 245-254.
134. J. M. Kee, R. C. Oslund, D. H. Perlman and T. W. Muir, *Nat. Chem. Biol.*, 2013, **9**, 416-421.
135. D. Andes, A. Pascual and O. Marchetti, *Antimicrob. Agents Chemother.*, 2009, **53**, 24-34.
136. A. Eisenfuhr, P. S. Arora, G. Sengle, L. R. Takaoka, J. S. Nowick and M. Famulok, *Bioorg. Med. Chem.*, 2003, **11**, 235-249.
137. J. W. Lown and S. M. S. Chauhan, *J. Chem. Soc., Chem. Commun.*, 1981, 675-676.
138. L. B. Johansson, *J. Chem. Soc., Faraday Trans.*, 1990, **86**, 2103.
139. E. Gasteiger, C. Hoogland, A. Gattiker, S. e. Duvaud, M. Wilkins, R. Appel and A. Bairoch, in *The Proteomics Protocols Handbook*, ed. J. Walker, Humana Press, 2005, ch. 52, pp. 571-607.
140. M. Raina, R. Sharma, S. E. Deacon, C. Tiede, D. Tomlinson, A. G. Davies, M. J. McPherson and C. Walti, *Analyst*, 2015, **140**, 803-810.
141. W. Lian, B. Jiang, Z. Qian and D. Pei, *J. Am. Chem. Soc.*, 2014, **136**, 9830-9833.
142. S. J. Maynard, A. M. Almeida, Y. Yoshimi and S. H. Gellman, *J. Am. Chem. Soc.*, 2014, **136**, 16683-16688.
143. H. Liu, V. R. Pattabiraman and J. C. Vederas, *Org. Lett.*, 2007, **9**, 4211-4214.
144. B. G. Davis, K. Khumtaveeporn, R. R. Bott and J. B. Jones, *Biorg. Med. Chem.*, 1999, **7**, 2303-2311.
145. M. C. Davis, *Synth. Commun.*, 2007, **37**, 1457-1462.
146. M. Ikeda, J. Minari, N. Shimada, M. Numata, K. Sakurai and S. Shinkai, *Org. Biomol. Chem.*, 2007, **5**, 2219-2224.