## *Total Synthesis of Bioactive Peptides and Whole Proteins*

LEAR, SAM

Department of Chemistry

# Total Synthesis of Bioactive Peptides and Whole Proteins

A Thesis Presented for the Degree of
Doctor of Philosophy

*Author:*
Sam Lear

*Supervisor:*
Dr Steven Cobb

August 2016

# Abstract

Total chemical synthesis is an essential tool for the validation of natural product structures and the discovery and elaboration of novel therapeutic scaffolds. Chapter 1 (Part I) surveys existing treatments for trypanosomatid neglected tropical diseases, and the synthesis of a novel class of antiparasitic cyclic depsipeptides is reported (Chapter 2) alongside a full NMR assignment and structure calculation using NMR-derived distance restraints. The synthetic peptides exhibit activity profiles in agreement with published results, and a series of ester-to-amide substitution analogues also synthesized show similar low micromolar potency. Chapter 3 describes the synthesis of lassomycin, a tuberculocidal lasso peptide reported to exhibit a unique unthreaded topology. The naturally occurring peptide was synthesized alongside C-terminally amidated and truncated analogues, but none were biologically active. Given clear differences observed in the two-dimensional NMR data for synthetic lassomycin, it is suggested that the reported natural product in fact exists in the threaded form. The chemical synthesis of whole proteins is addressed in Part II, and current progress in the field is reviewed (Chapter 4). Work towards the total chemical synthesis of acyl carrier protein using a two fragment approach is described in Chapter 5. The N-terminal fragment was synthesized using the sulfonamide linker, and while the C-terminal fragment presented difficulties due to extremely low solubility, solubilization using backbone protection was demonstrated. The development of a modified Dawson linker for the synthesis of peptide *N*-acylureas without overacylation is also described. Finally, Part III details the synthesis of tumor targeting peptides used for imaging and inhibition of cancer cell growth, and which cause tumor size reduction *in vivo*. A novel web utility used for the automated assignment of peptide mass spectra throughout this thesis is also presented.

# Contents

# Acknowledgements

Thanks go first and foremost to my supervisor Dr Steven Cobb for continued guidance and support throughout my time spent in research at Durham University. Thanks also go to Hannah Bolt, Dr Gabriela Eggimann and the other members of the Cobb Research Group. In addition I would like to thank Dr Jackie Mosely for acquiring and offering advice regarding interpretation of tandem MS and ion-mobility MS data, as well as Dr Dave Parker and Peter Stokes of the mass spectrometry service in the Department of Chemistry for carrying out the characterization of peptide samples. Thanks also go to Dr Alan Kenwright and Dr Juan Aguilar Malavia for invaluable advice regarding peptide NMR assignment and for the collection and processing of 2D NMR data.

Ambrose Crofton (undergraduate project student) carried out the initial synthesis of the chaiyaphumines and amide analogues and Sridévi Ramanoudjame (visiting Masters student) assisted with portions of the chaiyaphumine synthesis work. Parasitic and mammalian cytotoxicity screening was carried out by Dr Marcel Kaiser at the Swiss Tropical and Public Health Institute. Chaiyaphumine Marfey's and chromatographic analysis was carried out by Prof Helge Bode at Goethe Universität Frankfurt. Caitlin Hatton (summer studentship) carried out initial investigations into potential lassomycin synthetic approaches, while Dr Alex Hudson first suggested the use of side chain anchoring. Tuberculosis screening was performed by Dr Tulika Munshi in the research group of Dr Tim Bull at St. George's University of London. Charlie Tomlinson (undergraduate project student) assisted with the synthesis of the POLARIS Ala mutant described in Appendix C.

Special thanks go to Dr John Offer (NIMR, London) for generously allowing use of lab facilities and materials, and offering advice on the use of backbone protection and the sulfonamide safety-catch, as well as Dr Caroline Morris, Dr Richard Raz, Dr Abu Baker El Sayed and Dr George Papageorgiou of the Offer

group for invaluable advice and assistance. Thanks also go to Ian Edwards for assistance with ion-exchange purification, and for carrying out high-throughput crystallization screening, Dr Neil Colgin (Cambridge Research Biochemicals) for advice regarding MW-SPPS, and Prof Mohamed Marahiel (Philipps-Universität, Marburg) for helpful discussion on lasso peptides.

I am grateful to Dr Anna Mudge and her supervisor Prof Keith Lindsey (School of Biological and Biomedical Sciences) for collaboration on the POLARIS project, in particular Anna for assistance with synthesis and testing of PLS peptides, Dr Clarissa Sit (Harvard Medical School) for collaboration on the lassomycin work and Prof Ka-Leung Wong (Hong Kong Baptist University) for continued collaboration on various projects which have resulted in a number of the publications listed.

Finally, I would like to thank Dr Ehmke Pohl, a source of inexhaustible enthusiasm and guidance regarding peptide and protein crystallization, career advice, etc.

# Abbreviations

| | |
|---|---|
| Ac | Acetyl |
| Acm | Acetamidomethyl |
| ACP | Acyl carrier protein |
| Ahx | Aminohexanoic acid |
| All | Allyl |
| Alloc | Allyloxycarbonyl |
| $\alpha$-Syn | $\alpha$-Synuclein |
| AMP | Antimicrobial peptide |
| Ar | Aryl |
| AT | Acyl transferase |
| ATP | Adenosine triphosphate |
| Boc | *tert*-Butoxycarbonyl |
| cAMP | Cyclic adenosine monophosphate |
| Clp | Caseinolytic protease |
| CoA | Coenzyme A |
| COSY | Correlation spectroscopy |
| CTR1 | Copper transporter 1 |
| DARTS | Drug affinity responsive target stability |
| DBD | DNA-binding domain |
| DBU | 1,8-Diazabicyclo[5.4.0]undec-7-ene |
| Dbz | 3,4-Diaminobenzoic acid |
| DCHA | Dicyclohexylamine |
| DCM | Dichloromethane |
| DEBS | 6-Deoxyerythronolide B synthase |
| DH | Dehydratase |
| DIC | *N*,*N*′-Diisopropylcarbodiimide |
| DIPEA | *N*,*N*-Diisopropylethylamine |
| DMAP | 4-Dimethylaminopyridine |

| | |
|---|---|
| DMF | *N*,*N*-Dimethylformamide |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| Dpr | L-Diaminopropionic acid |
| DTT | Dithiothreitol |
| EBNA1 | Epstein-Barr virus nuclear antigen 1 |
| EBV | Epstein-Barr virus |
| EDCI | 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide |
| EIC | Extracted-ion chromatogram |
| EPO | Erythropoietin |
| ER | Endoplasmic reticulum |
| ER | Enoylreductase |
| ES | Electrospray |
| ESI | Electrospray ionization |
| ETR1 | Ethylene Receptor 1 |
| FAS | Fatty acid synthesis/synthase |
| Fmoc | Fluorenylmethyloxycarbonyl |
| FRET | Fluorescence resonance energy transfer |
| GalNAc | *N*-Acetylgalactosamine |
| GluNAc | *N*-Acetylglucosamine |
| HAT | Human African trypanosomiasis |
| HATU | 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate |
| HEPES | 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid |
| Hmb | 2-Hydroxy-4-methoxybenzyl |
| HMBC | Heteronuclear multiple-bond correlation spectroscopy |
| Hmnb | 2-Hydroxy-4-methoxy-5-nitrobenzyl |
| Hmsb | 2-Hydroxy-4-methoxy-5-methylsulfinyl benzyl |
| HOAt | 1-Hydroxy-7-azabenzotriazole |
| HOBt | 1-Hydroxybenzotriazole |
| HPLC | High-performance liquid chromatography |
| HRMS | High resolution mass spectrometry |
| HSD | Honest significant difference |
| HSQC | Heteronuclear single-quantum correlation |
| $IC_{50}$ | Half maximal inhibitory concentration |
| ICL | Isopeptide chemical ligation |
| IM | Ion-mobility |

| | |
|---|---|
| KCL | Kinetically controlled ligation |
| KR | Ketoreductase |
| KS | Ketosynthase |
| LC | Liquid chromatography |
| LDD | Loading didomain |
| LIFT | LIFT[I] |
| MALDI | Matrix-assisted laser desorption/ionization |
| MBS | 3-Maleimidobenzoyl *N*-hydroxysuccinimide ester |
| MeCN | Acetonitrile |
| MES | 2-(*N*-Morpholino)ethanesulfonic acid |
| MFD | Metal-free dethiylation |
| MIC | Minimum inhibitory concentration |
| MPAA | 4-Mercaptophenylacetic acid |
| MPD | 2-Methyl-2,4-pentanediol |
| $M_r$ | Relative molecular mass |
| MS | Mass spectrometry |
| MTT | 3-(4,5-Dimethyl-2-thiazolyl)-2,5-diphenyl-2*H*-tetrazolium bromide |
| MW | Microwave |
| $NAD^+$ | Nicotinamide adenine dinucleotide |
| $NADP^+$ | Nicotinamide adenine dinucleotide phosphate |
| NCL | Native chemical ligation |
| NHS | *N*-Hydroxysuccinimide |
| NMM | *N*-Methylmorpholine |
| NMP | *N*-Methyl-2-pyrrolidone |
| NMR | Nuclear magnetic resonance |
| NMT | *N*-Myristoyltransferase |
| NOE | Nuclear Overhauser effect |
| NOESY | Nuclear Overhauser effect spectroscopy |
| NTD | Neglected tropical disease |
| NTR | Nitroreductase |
| ODC | Ornithine decarboxylase |
| Pan | Pantetheine |
| Pbf | Pentamethyl-2,3-dihydrobenzofuran-5-sulfonyl |
| PDB | Protein Data Bank |
| PEG | Poly(ethylene glycol) |

[I] LIFT is not an acronym.

| | |
|---|---|
| PK | Polyketide |
| PKS | Polyketide synthase |
| PLS | POLARIS |
| PTFE | Polytetrafluoroethylene |
| PTM | Post-translational modification |
| PyBOP® | (Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate |
| QToF | Quadrupole time-of-flight |
| RNA | Ribonucleic acid |
| RNAi | RNA interference |
| RT | Room temperature |
| $R_t$ | Retention time |
| SDS | Sodium dodecyl sulfate |
| SEA | Bis(2-sulfanylethyl)amido |
| SeEA | Bis(2-selenylethyl)amido |
| SPPS | Solid-phase peptide synthesis |
| Su | Succinimidyl |
| TBAHS | Tetrabutylammonium hydrogen sulfate |
| TCEP | Tris(2-carboxyethyl)phosphine |
| TE | Thioesterase |
| TFA | Trifluoroacetic acid |
| THF | Tetrahydrofuran |
| Thz | Thiazolidine-4-carboxylic acid |
| TIPS | Triisopropylsilane |
| TLC | Thin layer chromatography |
| TMS | Trimethylsilyl |
| TOF | Time-of-flight |
| TQD | Tandem quadrupole detector |
| Tris | 2-Amino-2-(hydroxymethyl)-1,3-propanediol |
| Trt | Triphenylmethyl |
| Ts | Tosyl |
| Ub | Ubiquitin |
| VA-044 | 2,2′-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride |
| WHO | World Health Organization |

# Preface

The following thesis details a number of synthetic projects in peptide chemistry, covering a range of peptide target sizes and applications in biological research. Chapters 2 and 3 describe the total synthesis of a range of newly discovered peptide natural products possessing potent and selective antimicrobial activity. The usefulness of chemical synthesis in the elucidation of structure-activity relationships is demonstrated, in addition to its role in confirming (or disproving) published structural characterization for natural products. The compounds presented in Chapter 2 are active against the causative agents of a number of the so-called 'neglected tropical diseases', and a concise review of existing treatments is given in Chapter 1. The peptide described in Chapter 3 possesses a unique three-dimensional structure and is active against *Mycobacterium tuberculosis*.

Chapter 5 is concerned with the total chemical synthesis of a 77-amino acid protein involved in the biosynthesis of fatty acid natural products. A selection of illustrative examples of published chemical syntheses of whole proteins – representing an overview of the state of the art in the field – are reviewed in Chapter 4. Peptide thioesters are essential intermediates used in protein chemical synthesis, and a novel linker that can be used for the microwave-assisted synthesis of peptide thioester equivalents is also presented in Chapter 5.

The synthesis and fluorescent labelling of two tumor targeting peptides is described in Chapter 6. Their use in cancer cell imaging, and their *in vitro* and *in vivo* effectiveness against cancer cells and tumor xenografts respectively is also reported. Finally, Appendix C details the synthesis and investigation into sequence-activity relationships of a gene product found to be a regulator of root growth in *Arabidopsis*. The effects of the synthetic peptide on root length are documented. A number of publications arising as a result of the work described are also summarized.

# Publications and conferences

The following publications have arisen as a result of work carried out in the period described. The Pep-Calc.com manuscript is given in Appendix D.

- 'Total chemical synthesis of lassomycin and lassomycin-amide', Lear et al., *Org. Biomol. Chem.*, 2016, **14**, 4534.[1]
- 'Pep-Calc.com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment', Lear et al., *J. Comput. Aid. Mol. Des.*, 2016, **30**, 271.[2]
- 'The effects of morphology and linker length on the properties of peptide-lanthanide upconversion nanomaterials as G2 phase cell cycle inhibitors', Chan et al., *Eur. J. Inorg. Chem.*, 2015, 4539.[3]
- 'Directional Plk1 inhibition-driven cell cycle interruption using amphiphilic thin-coated peptide-lanthanide upconversion nanomaterials as *in vivo* tumor suppressors', Chan et al., *J. Mater. Chem. B*, 2015, **3**, 2624.[4]
- 'Monitoring and inhibition of Plk1: amphiphilic porphyrin conjugated Plk1 specific peptides for its imaging and anti-tumor function', Li et al., *Org. Biomol. Chem.*, 2014, **12**, 5876.[5]
- 'Targeting EBNA1: a therapeutic approach for EBV-related tumor with tailor responsive optical imaging', Jiang et al., *Nat. Commun.*, under review.[6]
- 'A peptide that regulates metallation of the *Arabidopsis* ethylene receptor', Mudge et al., *Nat. Plants*, under review.[7]
- 'Structural elaboration of the chaiyaphumine natural product scaffold reveals importance of ester bond and affords lactam analogues with

comparable antiparasitic activity', Lear et al., *ACS Infect. Dis.*, in preparation.[8]

The results described in Chapter 3 were presented by the author at the 2016 Belgian Peptide Group Meeting, Brussels, Belgium (poster presentation), the 2015 RSC Protein and Peptide Science Group (PPSG) Early Stage Researcher Meeting, Durham University (oral presentation), the 24th International American Peptide Symposium (APS) 2015, Orlando, Florida (poster presentation), the 2015 Mini-Symposium on Peptide and Protein Science, Durham University (oral presentation), the 2014 Dutch Peptide Symposium, Utrecht University, Netherlands (oral presentation), the 2014 SCI Young Chemists Symposium (oral presentation) and the 9th Chemistry and Biology of Peptides meeting 2014, NIMR, London (poster presentation). Those described in Chapter 2 were presented by the author at the British Society For Parasitology Autumn Symposium 2016, Durham University (poster presentation).

During the period of research covered the author also attended the RSC NorthEast Regional Organic Meeting 2016 (Newcastle University), the 5th International Meeting on Antimicrobial Peptides 2015 (Burlington House, London), the 6th Chemical Protein Synthesis Meeting 2015 (St. Augustine, Florida), the 2014 Peptides in Paris Symposium and Peptide & Protein Chemistry & Biology Training Workshop (University of Cergy-Pontoise, Paris, France), the 2013 and 2014 RSC PPSG Early Stage Researcher meetings (Durham University and Burlington House, respectively), the 8th Chemistry and Biology of Peptides meeting 2013 (University of Nottingham) and the PPI-Net Young Researchers Meeting 2013 (University of Leeds).

In addition, Pep-Calc.com[2] was presented by Hannah Bolt at the 9th Peptoid Summit, Lawrence Berkeley National Laboratory, Berkeley, California (oral presentation) and by the author at the 2013 and 2015 RSC PPSG Early Stage Researcher meetings. The Plk1 work[5] was presented by the author at the APS and the 2014 RSC PPSG meeting.

# Part I

# Total synthesis of antimicrobial peptide natural products

# Chapter 1

# Current and future therapies for neglected tropical diseases

The neglected tropical diseases are a group of 13 disabling chronic conditions that constitute the most common infections worldwide, affecting billions of people living below the poverty line.[9] The following chapter summarizes their occurrence, geographical distribution and current treatments, focusing specifically on those resulting from trypanosomatid infection: Chagas disease, human African trypanosomiasis (HAT) and leishmaniasis. The chapter concludes with an outlook towards future therapies and the elucidation of novel parasitic drug targets. The total synthesis and characterization of a series of compounds active against a number of NTDs and malaria is described in Chapter 2.

## 1.1 Occurrence and geographical distribution

Neglected tropical diseases (NTDs) affect some of the poorest communities in the world, up to 2.7 billion people who live on less than $2 a day.[9] It is estimated that NTDs cause over 500 000 deaths per year, and although this number is small compared to that for other tropical diseases such as malaria, their resulting contribution to severe disability and poverty means that they constitute a significant burden on healthcare services worldwide. The regions in which Chagas disease, HAT and leishmaniasis occur are summarized on the map shown in Figure 1.1. Other diseases which come under the NTD umbrella include three soil-transmitted helminth infections (ascariasis,

hookworm infection, and trichuriasis), lymphatic filariasis, onchocerciasis, dracunculiasis, schistosomiasis, Buruli ulcer, leprosy, and trachoma.[9]

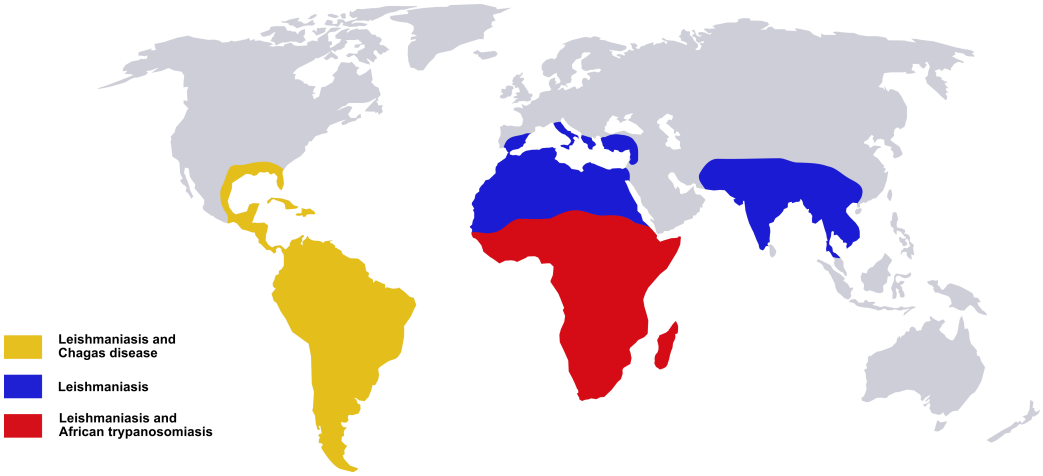



**Figure 1.1:** Geographic distribution of Chagas disease, human African trypanosomiasis (HAT) and leishmaniasis. Adapted from Cavalli et al.[10]

NTDs occur primarily in rural or poor urban areas of low-income countries in sub-Saharan Africa, Asia, and Latin America.[9] Because of this they receive comparatively little attention from the pharmaceutical industry, due to the fact that they offer little or no commercial incentive for drug development.[11] It is estimated that a central-nervous-system disorder or cancer drug has a 13-fold greater chance of being brought to market than one for a neglected disease.[12] Consequently for many NTDs treatment options are severely limited, and current front-line treatments often have undesirable toxic side effects or problems with antimicrobial resistance. A new generation of therapies is therefore needed to tackle the growing problem, which will be discussed in this chapter.

## 1.2 Symptoms and disease vectors

A description of the primary symptoms of Chagas disease, HAT and leishmaniasis – and their associated vectors – is shown in Figure 1.2. All three diseases result from infection by parasites of the order trypanosomatida.

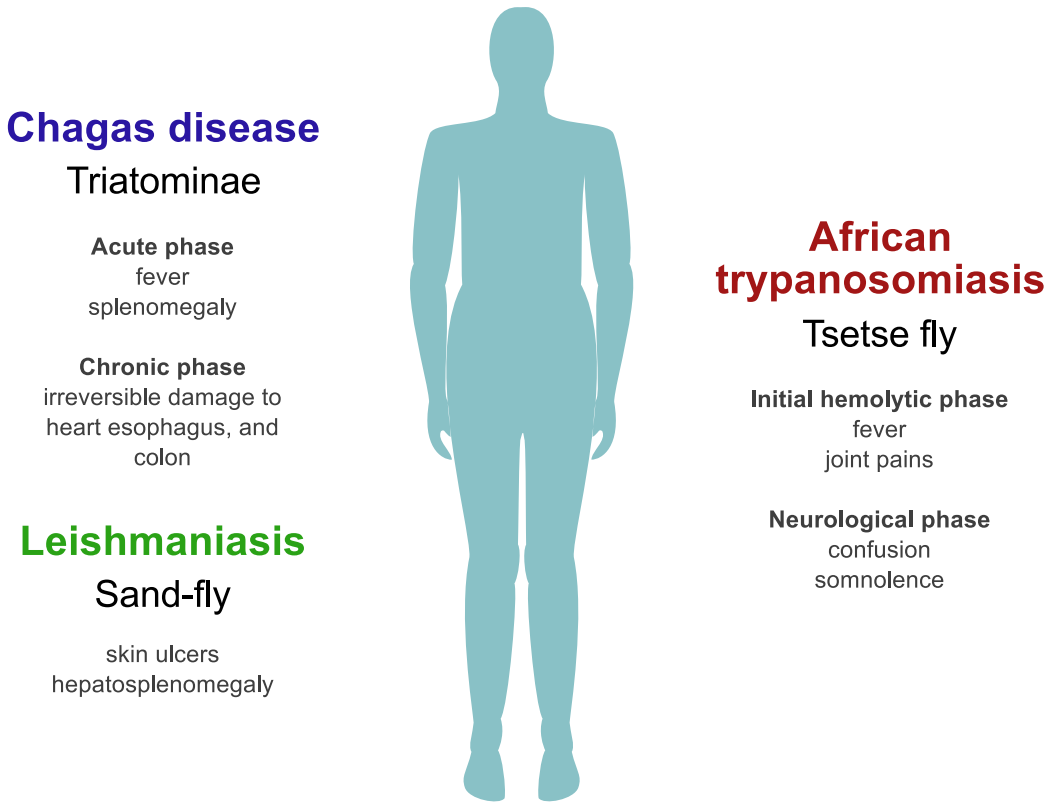



**Figure 1.2:** Summary of disease vectors and symptoms for the NTDs Chagas disease, African trypanosomiasis and leishmaniasis.[13]

HAT (or sleeping sickness) is caused by infection with either of the parasites *Trypanosoma brucei gambiense* or *Trypanosoma brucei rhodesiense*, the former a chronic form of the disease present in western and central Africa, and the latter an acute form found in eastern and southern Africa.[14] HAT is transmitted through the bite of an infected tsetse fly, and causes progressive neurological damage upon crossing of the blood-brain barrier by the parasite, leading to coma, severe organ failure, and eventually death if left untreated.[13,14]

Chagas disease is endemic to Latin America and is caused by the species *Trypanosoma cruzi*, carried by the triatomine (or 'kissing bug').[15] The parasite is transmitted when the insect defecates on the victim's skin whilst taking a blood meal. *T. cruzi* can enter the host through broken skin, such as cuts, abrasions or the insect bite, or it can penetrate the soft skin of the eyes and mouth. The initial acute infection often presents with only relatively mild symptoms, such as swelling of lymph nodes and tissues, conjunctivitis and skin lesions. The chronic infection can persist unnoticed for more than 30 years, and can lead to complications such as abnormal heart rhythm, heart failure, digestive problems and sudden cardiac death. Chagas disease affects 10–12 million people, and is responsible for over 15 000 deaths yearly.[15]

Leishmaniasis is most commonly caused by the parasites *Leishmania donovani* and *L. infantum*/*L. chagasi*, transmitted by the bite of an infected female phlebotomine sand fly.[16] The disease typically occurs in two forms: cutaneous, which causes skin ulcerations lasting months to years and often resulting in severe scarring, and visceral, which affects internal organs, especially the spleen, liver, and bone marrow. If left untreated, severe visceral leishmaniasis is often fatal. It is estimated that 2 million new cases of leishmaniasis occur each year.[16]

## 1.3 Existing treatments

Existing drugs for NTDs often suffer from severe toxic side effects. In addition, many are increasingly ineffective due to the problem of antimicrobial resistance. Figure 1.3 shows the current front-line treatments for HAT, Chagas disease and leishmaniasis.

Benznidazole (**2**) and nifurtimox (**1**), first used to treat Chagas disease in 1974 and 1970 respectively, are nitroheterocyclic prodrugs that undergo activation by nitroreductases (NTRs) within the parasite.[18] This results in

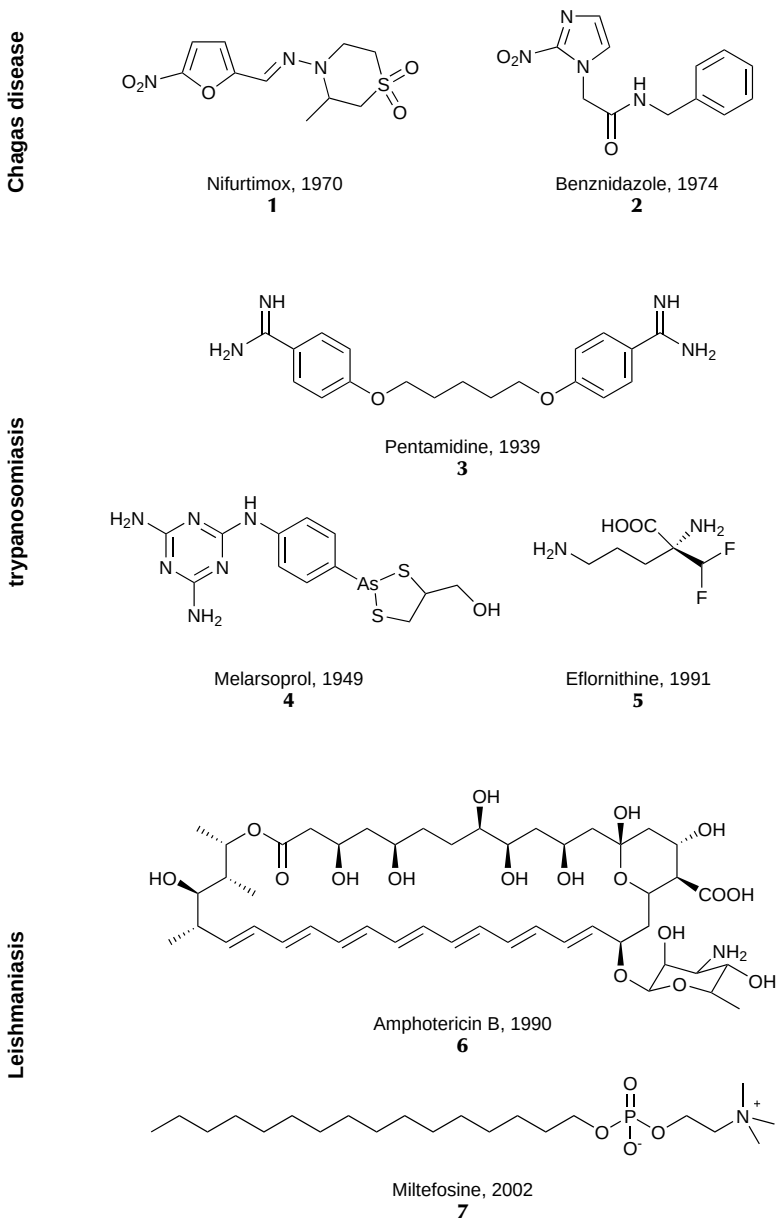
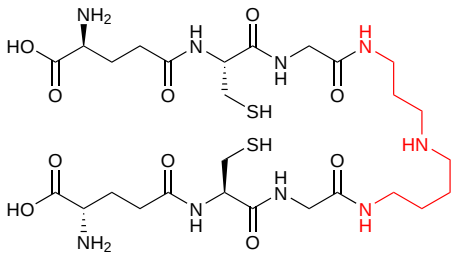



**Figure 1.3:** Structures of widely used or recently introduced therapies for Chagas disease, HAT and leishmaniasis. Current drugs carry risks of severe adverse effects, problems with antimicrobial resistance and/or a lack of efficacy in the later disease stages.[10,17]

the production of reactive moieties, potentially producing a cytotoxic effect through the promotion of DNA damage and/or induction of oxidative stress, although the nature and extent of these processes are not fully understood. Cases resistant to treatment with nitroheterocyclic therapies are commonly reported, and resistant parasites can be selected in the laboratory.[19] It is suggested that benznidazole resistance in *T. cruzi* arises due to a number of mechanisms, one of which involves mutations in the gene which encodes the nitroreductase TcNTR.[20]

Eflornithine (**5**), the front-line therapy against HAT (first used in 1991), is an ornithine analogue that targets parasitic ornithine decarboxylase (ODC).[21] Inhibition of ODC causes a reduction in polyamine biosynthesis and diminished production of trypanothione (**8**, Figure 1.4), the redox active metabolite occurring exclusively in trypanosomatids and essential for their survival.[22] Interestingly, nifurtimox/eflornithine combination therapy has been found to be effective against the Gambian form of HAT (*T. b. gambiense*), resulting in cure rates of ~97%, and has consequently been added to the WHO Essential Medicines List.[23] Eflornithine resistance resulting from the loss of a single amino acid transporter gene has been demonstrated.[21]



**8**

**Figure 1.4:** Trypanothione (**8**) consists of two glutathione molecules joined by a spermidine linker (shown in red). Trypanosomatids have evolved a unique thiol-redox system centred on **8**.[22]

Pentamidine (**3**, 1939) is the front-line treatment for early-stage *T. b. gambiense* infection, but cannot cross the blood-brain barrier, and so is ineffective once the parasite has entered the central nervous system.[24] Melarsoprol (**4**, 1949), the only available treatment in the late stage of infection

by the *T. b. rhodesiense* subspecies, acts by inhibiting parasite glycolysis.[24,25] However, **4** exhibits considerable toxicity, killing 5% of HAT patients receiving this treatment.[21] Cross-resistance to pentamidine and melarsoprol was first observed over 60 years ago, and has been linked to genetic changes in *T. brucei* transport proteins and aquaglyceroporins.[26]

Amphotericin B (**6**, 1990) is used for the treatment of visceral leishmaniasis, and while drug toxicity can be reduced by administration as a colloidal dispersion, this option is more expensive.[27,28] Miltefosine (**7**, 2002) is also used to treat visceral leishmaniasis, although its associated teratogenicity means that pregnancy must be avoided until at least two months after treatment.[29] The mechanism of action is not fully understood for either **6** or **7**.[29,30] While pore formation leading to depolarization of the membrane is assumed to be responsible for the cytotoxicity of amphotericin B, auto-oxidation and subsequent formation of free radicals may also play a part.[30] It is suspected that miltefosine and other alkyl-lysophospholipids can trigger programmed cell death (apoptosis), although the way that this occurs is not currently understood.[29]

Miltefosine is highly vulnerable to resistance due to its long half-life, and it is suggested that this arises due to suppression of oxidative stress-induced apoptosis.[31] Resistance to amphotericin B in *Leishmania donovani* may arise from alteration of parasitic membrane composition and fluidity, in addition to upregulation of the thiol metabolic pathway and overexpression of drug efflux machinery.[30]

## 1.4 Target validation and future therapies

There are very few validated molecular targets for parasitic diseases.[32] As can be seen from the previous section, the mechanism of action is not well understood for any of the drugs currently available for the treatment of trypanosomatid infections. The only exception is the ornithine decarboxylase inhibitor eflornithine (**5**), described previously.[32] Continued research into the validation of targets for existing antiparasitic drugs is therefore required, in addition to the discovery of new targets. Phenotypic screening can be used to ascertain the importance of a given molecular target for cell survival. This information can then be used to guide the selection of targets to use in target-based approaches, such as enzyme inhibition assays.

*N*-Myristoyltransferase (NMT) is an example of an enzyme found to be essential for survival in *T. brucei*, as elucidated through RNA interference (RNAi) studies.[33] NMT is responsible for myristoylation of the N-terminal glycine of a number of different peptides.[32] Brand et al. used high throughput screening followed by binding optimization to identify NMT inhibitors with nanomolar potency.[34,35] The lead molecule in the study (**9**, $IC_{50}$ = 2 nM) was found to cure rodent models of peripheral HAT infection. A crystal structure of **9** bound to NMT is shown in Figure 1.5.

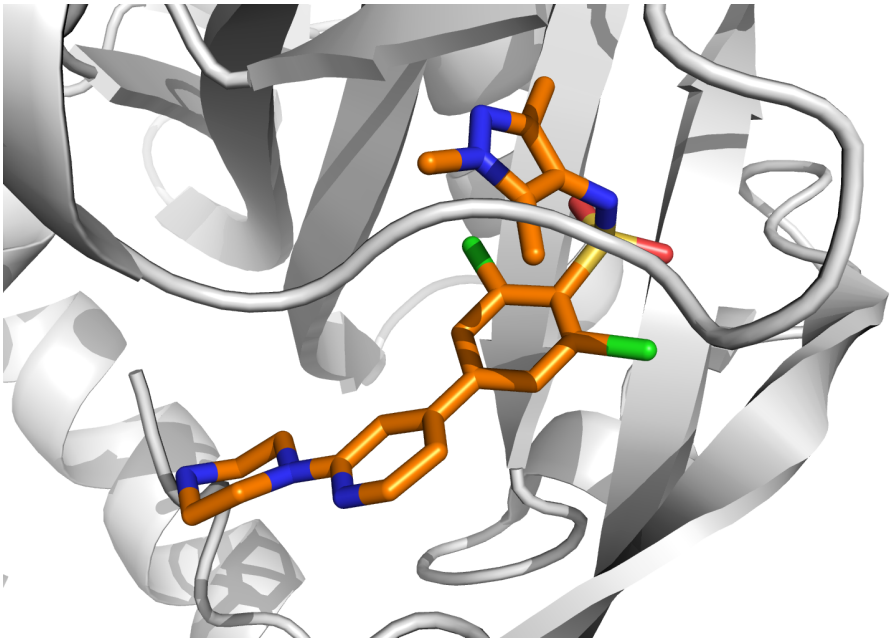


**Figure 1.5:** X-ray crystal structure of **9** bound to N-myristoyltransferase (NMT), PDB ID: 2WSA.[34,36]

Many other potential enzyme targets exist in trypanosomatida, such as inositol-3-phosphate synthase, glucose-6-phosphate dehydrogenase, Long chain fatty acyl-CoA synthetase, and the cysteine peptidases, histone deacetylases and topoisomerases.[11] The sterol 14-$\alpha$-demethylase inhibitors posaconazole and E1224 (a prodrug of ravuconazole) are active against Chagas disease and are currently in clinical trials.[32,37] *In vivo* efficacy was recently reported for a selective inhibitor of the kinetoplastid proteasome.[38] Other

possible targets include parasite surface glycoconjugates, and pathways such as apoptosis, cAMP signalling, glycolysis and lipid biosynthesis.[11]

Antimicrobial peptides (AMPs) have also been shown to be active against NTD-causing parasites, either through direct membrane or capsid interactions leading to membrane disruption, or by involvement in immunoregulatory processes such as angiogenesis, cell proliferation, cytokine regulation, leukocyte chemotaxis, degranulation of mast cells, stimulation of phagocytosis and even modulation of gene expression.[39] AMPs offer promise due to their broad spectrum of activity and limited probability of inducing pathogen resistance.

## 1.5 Conclusions and future work

It can be seen that sustained effort towards the discovery of improved drugs for NTDs is vitally important in the face of increasing resistance to current therapies. While this necessarily involves the improvement of available drugs and the elucidation of their mechanisms of action (many of which are poorly understood), a significant drive to elucidate new molecular targets in parasites is also necessary. The next chapter details the total synthesis of a number of newly discovered compounds found to be active against *T. brucei* and *T. cruzi*, in addition to *Plasmodium falciparum* – a causative agent of malaria. Investigations into the relationship between their molecular structure and activity are also described.

# Chapter 2

# Synthesis of the 'chaiyaphumine' series of antiparasitic natural products

Vector-borne protozoan infections constitute a major burden on healthcare services throughout the world. As discussed in Chapter 1, neglected tropical diseases represent a significant proportion of this burden. In addition, malaria – caused by protozoan parasites belonging to the genus *Plasmodium* – accounts for more than 1 million deaths per year worldwide.[40] Living organisms are a rich source of privileged molecular scaffolds for use in the search for bioactive candidates as leads in pharmaceutical development.[41,42] Numerous peptide natural products are produced by organisms such as fungi and bacteria, and these often offer promising leads in drug research.

The following chapter details the total synthesis of a newly discovered class of cyclic peptide natural products found to possess biological activity against the parasitic agents responsible for Chagas disease, human African trypanosomiasis and malaria. The synthesis of a series of more stable analogues that retain biological activity is also described.

## 2.1 Introduction

A suite of novel cyclic depsipeptides, termed 'chaiyaphumines', was recently isolated from *Xenorhabdus* sp. by Grundmann et al.[43] A number of members

were found to be active against *T. brucei*, *T. cruzi*, and *P. falciparum*, the causative agents of HAT, Chagas disease and malaria respectively (see Chapter 1). In order to confirm their structure and investigate the relationship between chaiyaphumine structure and antiparasitic activity, a total synthesis of the previously discovered compounds was proposed.

The general structures of the chaiyaphumines isolated are summarized in Figure 2.1. The sequence contains D-alanine and D-phenylalanine, in addition to an ester bond between the C-terminus and threonine side chain hydroxyl group. The N-terminus is acylated with a carbonyl fragment bearing an R-group which depends on the compound (shown).

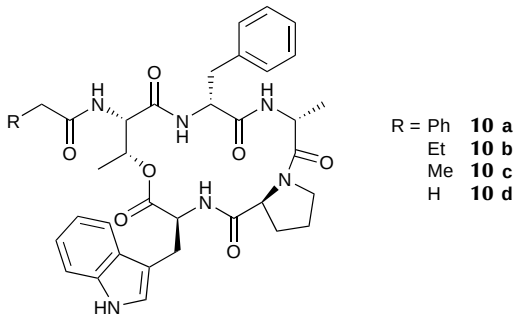
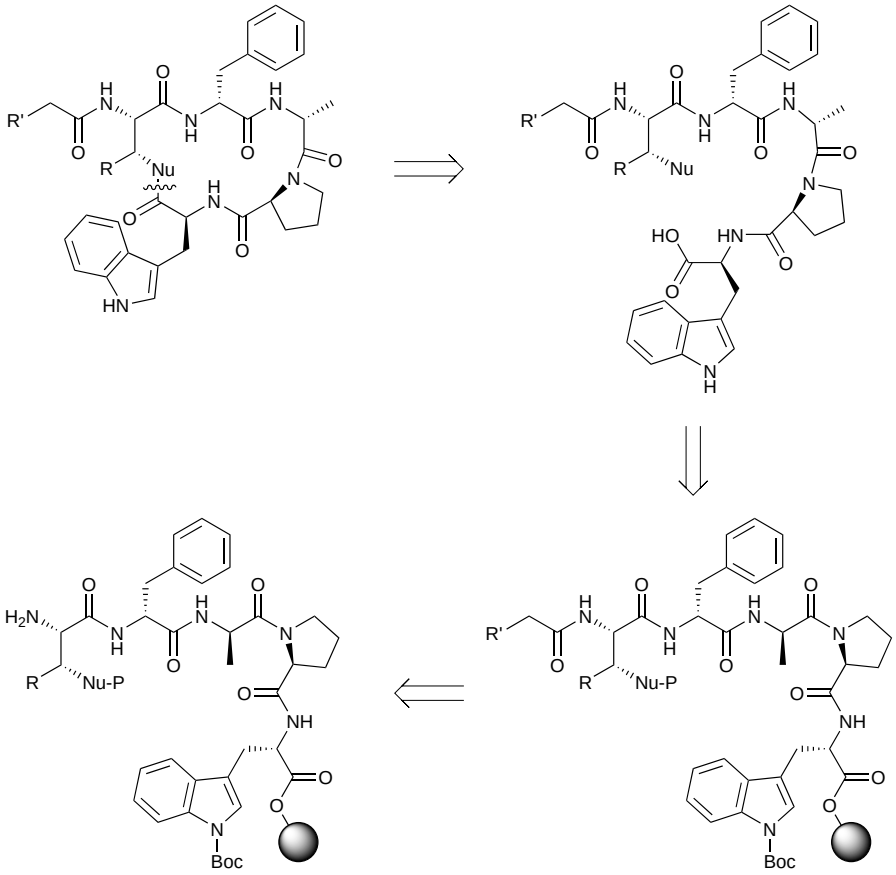



**Figure 2.1:** Isolated chaiyaphumine structures.[43]

Due to the absence of reactive side chain functionality, it was decided that a solution-phase cyclization approach would be the most straightforward for the total synthesis of the chaiyaphumines. This would involve formation of the ester bond in solution between the threonine side chain and C-terminus of the fully deprotected peptide. A similar approach was used in the total synthesis of the natural product lassomycin, described in Chapter 3. In order to elucidate the importance of the potentially labile ester bond, the synthesis of a set of analogues where the ester bond has been replaced by an amide linkage was also proposed and is described below.[44]

A retrosynthesis is given in Scheme 2.1. Routes to both the natural structures (**10a–d**) and amide-linked analogues (**11a–d**) are shown. The required sequence is synthesized via SPPS followed by installation of a range of N-terminal capping groups via acylation with the appropriate acid chloride or acetic anhydride. Following global deprotection and cleavage from the

resin, a final solution-phase cyclization step furnishes the target peptide. The synthesis and characterization of the chaiyaphumine ester and amide-linked analogue series is described in the following sections.

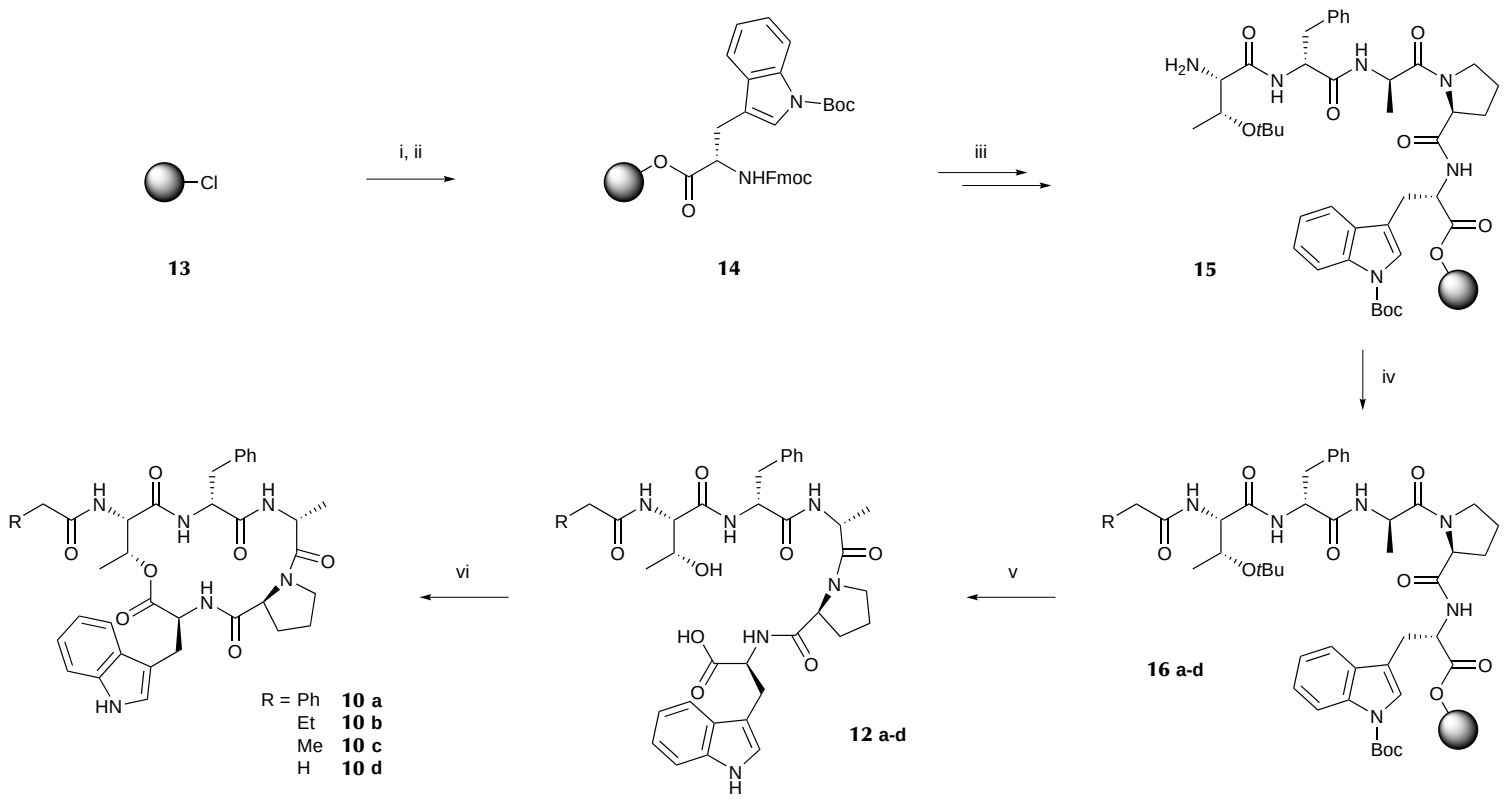


**Scheme 2.1:** Retrosynthetic route for chaiyaphumines and chaiyaphumine analogues (Nu = $NH_2$ or OH, R = H or $CH_3$ respectively, P = protecting group). R′ varies across the range of chaiyaphumines isolated (see text).

## 2.2 Synthesis of chaiyaphumine ester series

The synthetic route used for chaiyaphumines **10a–d** is shown in Scheme 2.2. Due to the presence of a proline residue close to the C-terminus, the sterically hindered 2-chlorotrityl linker was used to prevent diketopiperazine formation

early in the synthesis. The required sequence was synthesized via SPPS followed by installation of the appropriate N-terminal capping group via acylation with either 2-phenylacetyl, butyryl or propionyl chloride, or acetic anhydride. Cleavage from the resin and global deprotection was carried out using 95% TFA to give linear precursors **12a**–**d**, as confirmed by LCMS.




**Scheme 2.2:** Synthesis of chaiyaphumines **10a–d**. Reagents and conditions: (i) Fmoc-Trp(Boc)-OH (1.0 equiv) and DIPEA (5.0 equiv), DMF, RT, 2 × 1 h; (ii) MeOH, RT, 15 min; (iii) Fmoc SPPS (see Experimental section); (iv) $RCH_2COCl$ or $(CH_3CO)_2O$ for **10d** (5.0 equiv), DIPEA (10 equiv), DMF, RT, 2 × 1 h; (v) TFA (95%, water), TIPS, RT, 3–4 h; (vi) PyBOP® (3.0 equiv) and DIPEA (6.0 equiv) or EDCI/DMAP/HOBt for **10b** (see text), DMF, RT, 5 h (high dilution).

The cyclization step was carried out in solution at high dilution using a final peptide concentration of 0.003 M. The required total volume of DMF was calculated for each peptide, and a fifth of this was taken to dissolve the crude linear precursor. This solution was added dropwise over a period of 2 h to a stirred solution of PyBOP® (3.0 equiv) and DIPEA (6.0 equiv) in the remaining volume of DMF, and the reaction mixture was stirred for a further 3 h. Removal of the solvent under vacuum, followed by reversed-phase HPLC purification of the crude residue yielded chaiyaphumines **10a**–**d**. Peptides **10a** and **10c** were obtained in high purity by this method, as confirmed by analytical HPLC and ESI-MS (see below).

Peptide **10d** proved impossible to separate from the tripyrrolidinophosphine oxide by-product produced during amide bond formation using PyBOP® during the cyclization step ($m/z$ = 258.2), due to their coelution during HPLC purification. The cyclization was therefore repeated using the coupling reagent HATU, although this resulted in a second impurity ($m/z$ = 784.4) that also proved impossible to remove from the final product (see chromatogram for **10d** in Figure 2.2). The by-product was hypothesized to result from capping of the linear peptide by HATU (due to the initial large excess of the activator under the conditions used), thus preventing cyclization. While the synthesis of **10b** also initially produced a similarly impure product, subsequent optimization of the reaction and cleavage conditions, using syringe pump control with EDCI/DMAP/HOBt activation (as described in the Experimental section) significantly improved product quality.[45] Optimization of the synthesis of **10d** using alternative conditions for ester formation is ongoing. ESI-MS and analytical HPLC data for synthetic chaiyaphumines **10a**–**d** are shown in Figure 2.2.

## 2.3 Comparison with naturally isolated material

In order to compare the synthetic material with the natural product, a co-injection of **10a** with its naturally isolated counterpart was performed under the conditions described previously (Helge Bode, Goethe Universität Frankfurt).[43] A slight difference in retention time was observed for synthetic and natural chaiyaphumine (Figure 2.3), and in order to confirm the absolute configuration of the amino acids in both compounds, the advanced Marfey's method was used as described previously (chromatograms shown in Figure 2.4–

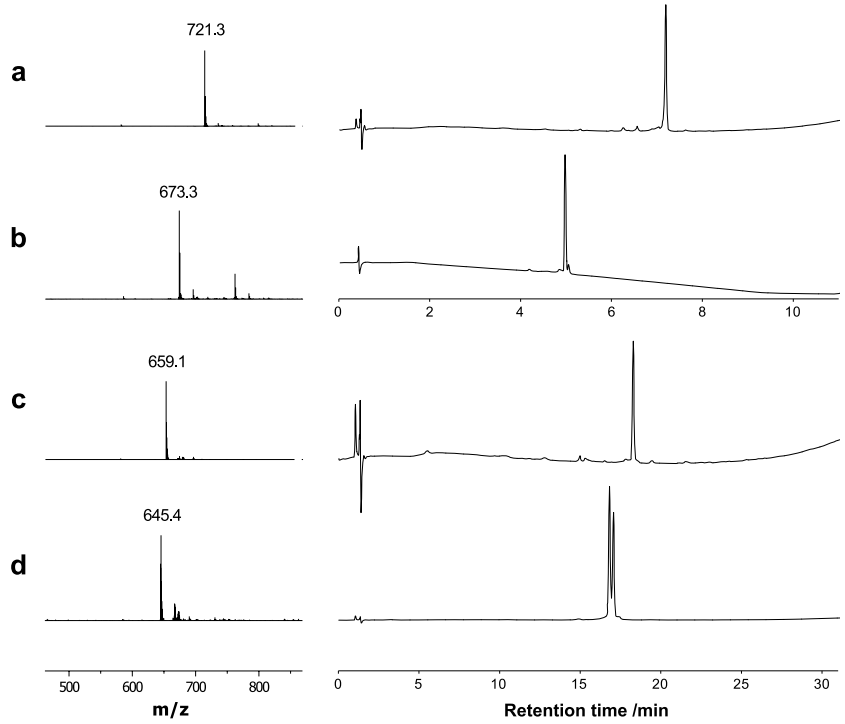



**Figure 2.2:** ESI-MS data (left) and HPLC traces (right) for synthetic chaiyaphumines **10a–d**. Labelled MS peaks are for $[M + H]^+$ ion. Absorbance recorded at 220 nm, gradient 0-100% B over 30 or 10 min as indicated for **10b** (A = 5:95:0.05 MeCN/$H_2O$/TFA; B = 95:5:0.03 MeCN/$H_2O$/TFA). The chromatogram for **10d** exhibits a double peak (see text).

Figure 2.8).[46–48] No difference in the amino acid configuration was observed however. It is therefore suggested that the differences in retention time and NMR data (below) might be the result of proline *cis*/*trans* isomerisation as observed by Zhou et al. for proline-containing xenoamicines, which naturally exhibit two distinct peaks for each mass.[49]

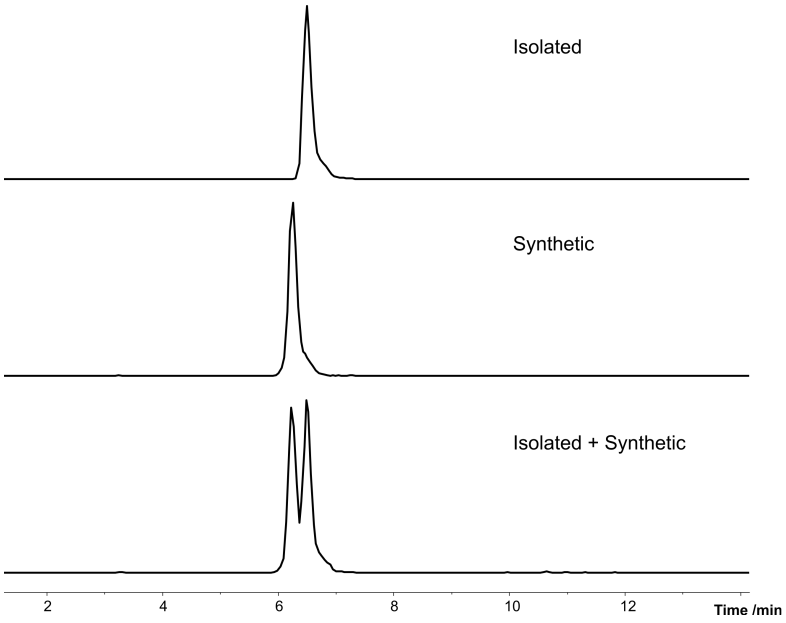



**Figure 2.3:** HPLC/MS of extracted ion chromatograms (EIC *m*/*z* 721.3 $[M + H]^+$) for analysis of natural (top) and synthetic (middle) chaiyaphumine **10a**, and co-injection of both (bottom). Conditions as previously described.[43]

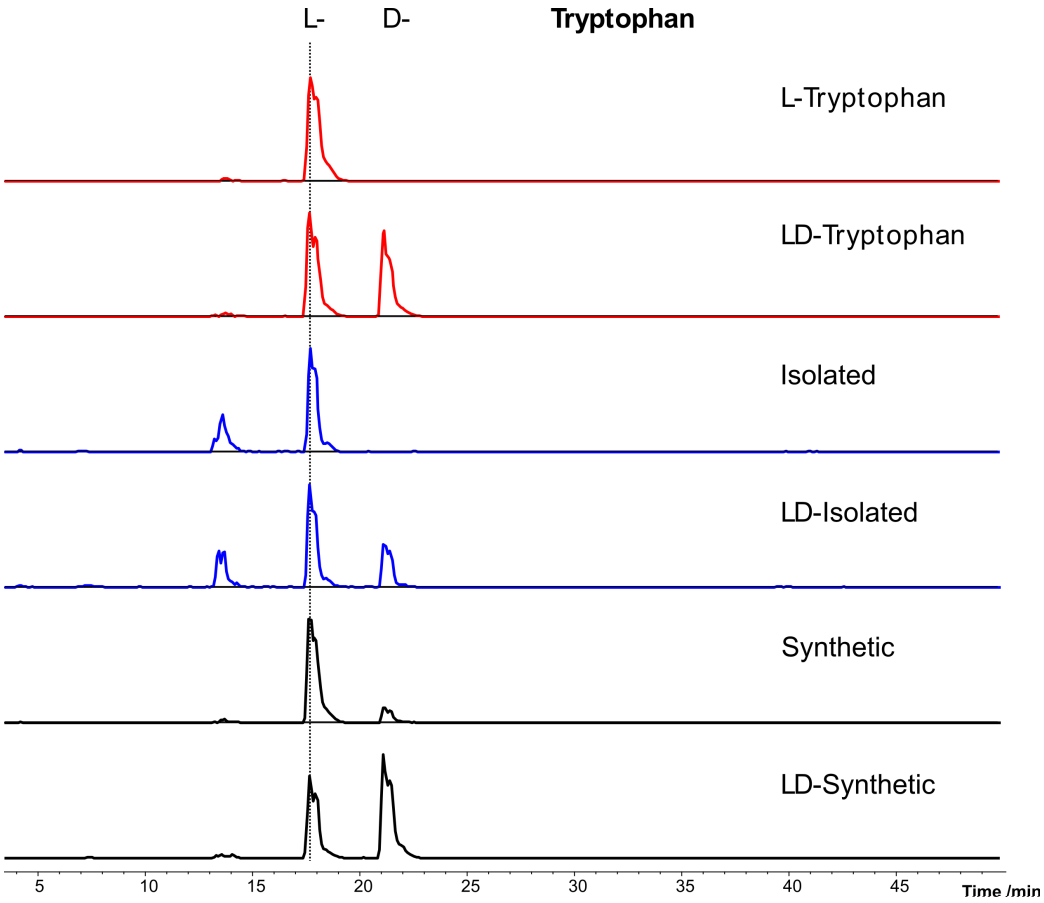



**Figure 2.4:** Advanced Marfey's analysis of synthetic **10a** and naturally isolated chaiyaphumine for the amino acid tryptophan (EIC *m/z* 499.0 $[M + H]^+$).

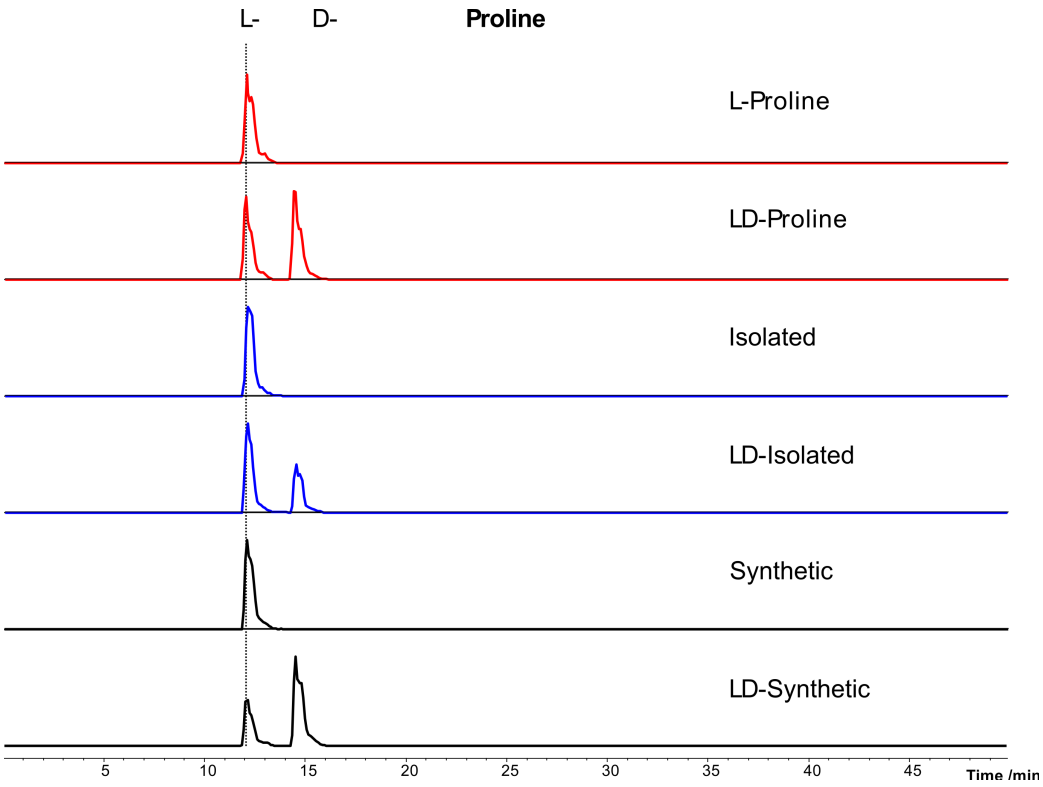



**Figure 2.5:** Advanced Marfey's analysis of synthetic **10a** and naturally isolated chaiyaphumine for the amino acid proline (EIC *m/z* 410.0 $[M + H]^+$).

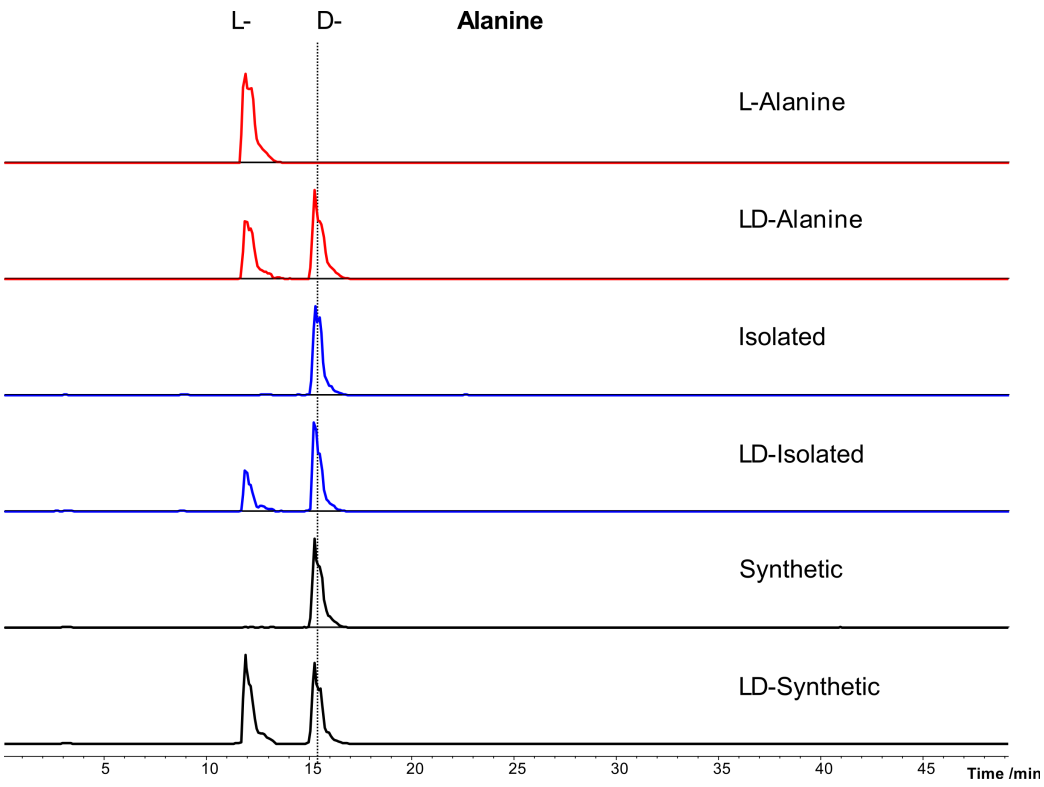



**Figure 2.6:** Advanced Marfey's analysis of synthetic **10a** and naturally isolated chaiyaphumine for the amino acid alanine (EIC $m/z$ 384.0 $[M + H]^+$).

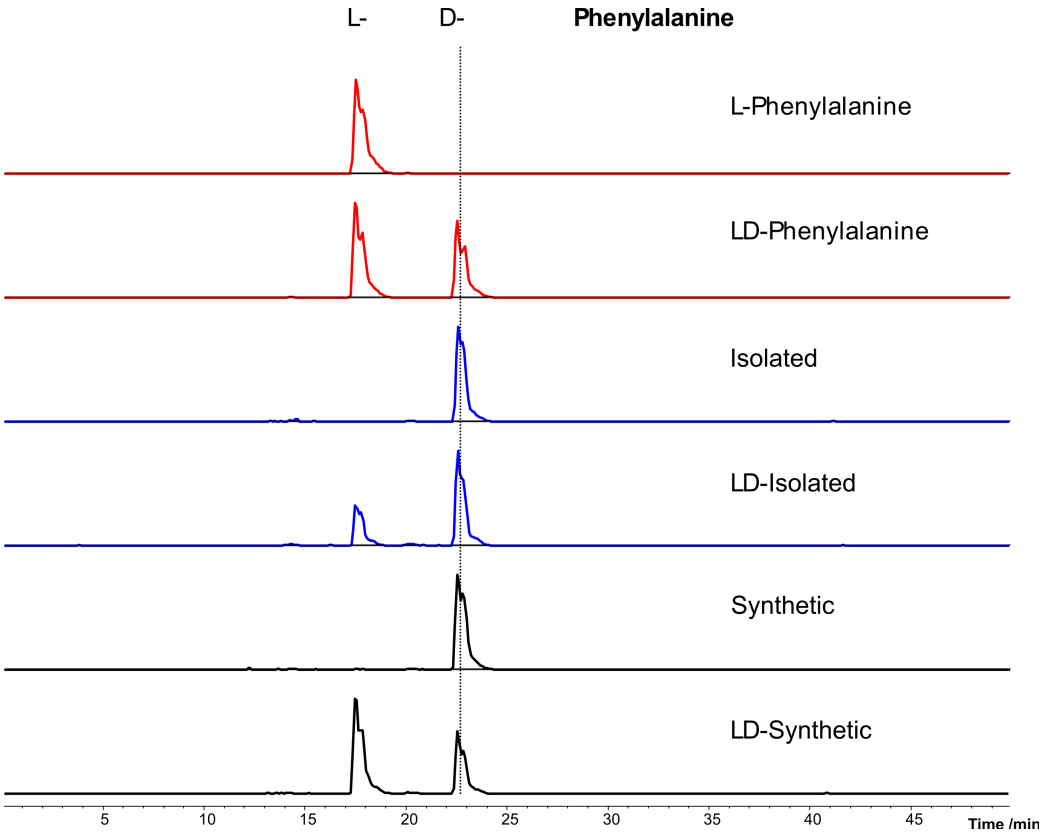



**Figure 2.7:** Advanced Marfey's analysis of synthetic **10a** and naturally isolated chaiyaphumine for the amino acid phenylalanine (EIC *m/z* 460.0 $[M + H]^+$).

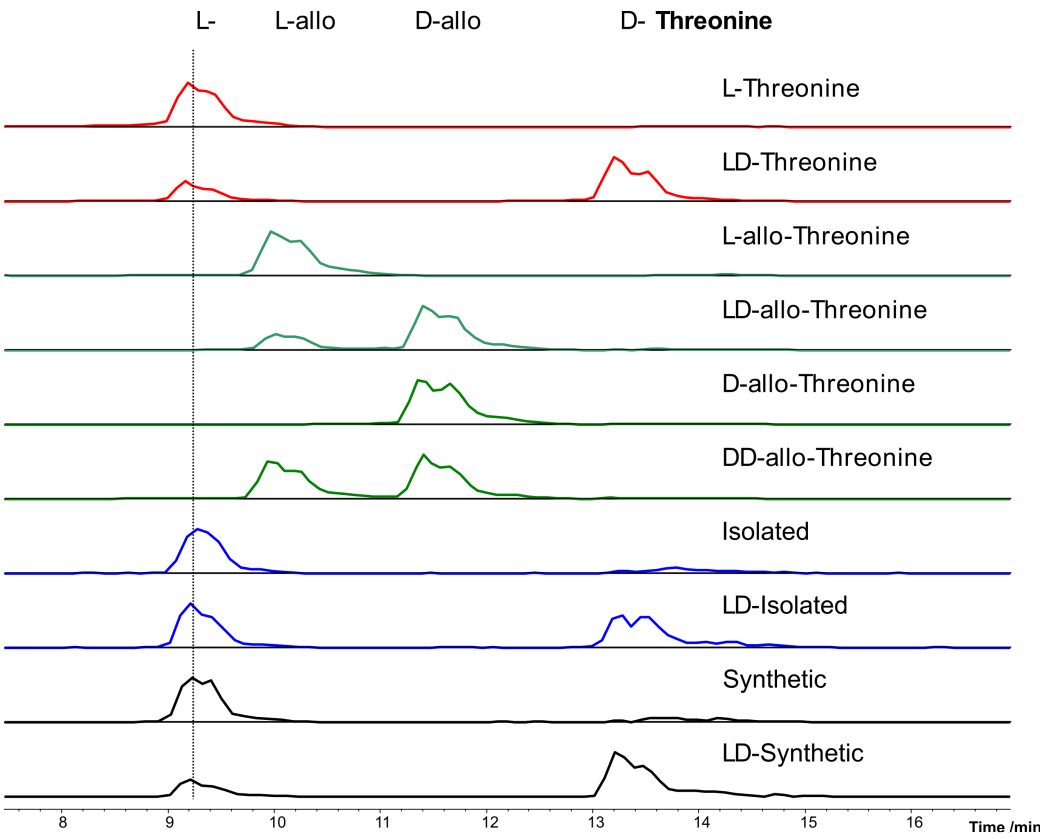
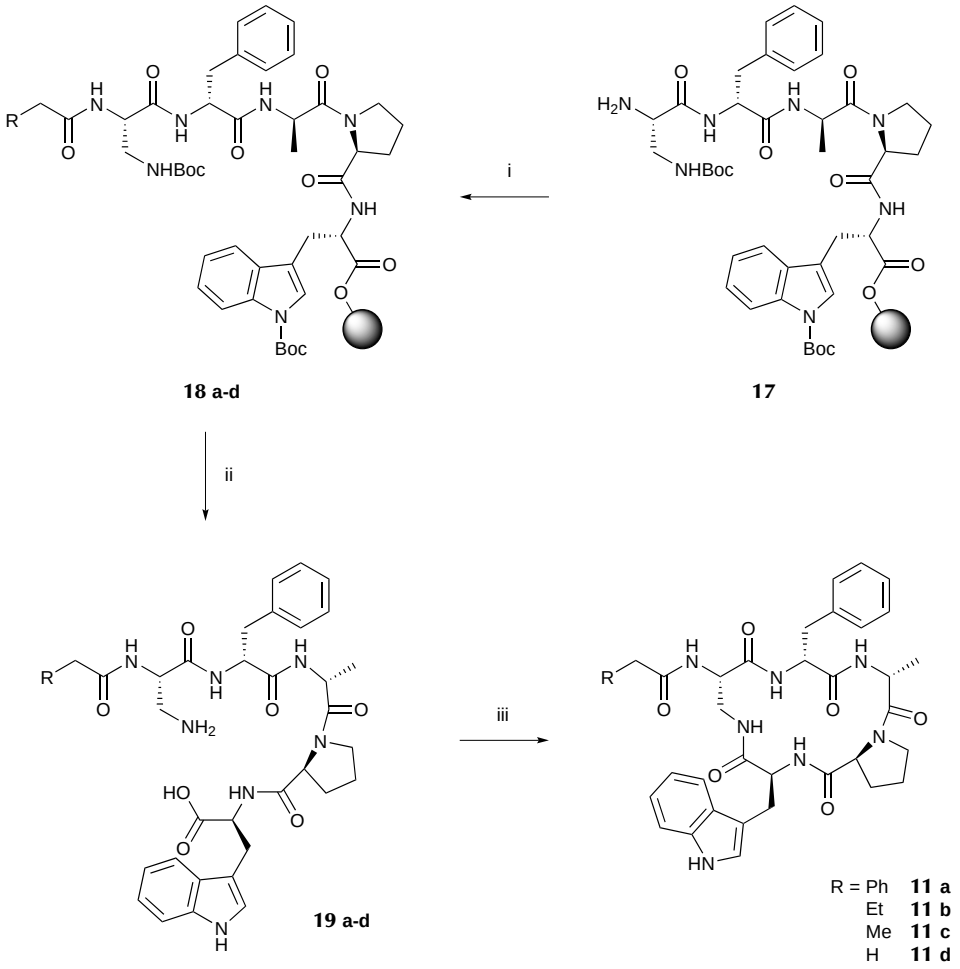



**Figure 2.8:** Advanced Marfey's analysis of synthetic **10a** and naturally isolated chaiyaphumine for the amino acid threonine (EIC $m/z$ 414.0 $[M + H]^+$).

## 2.4 Synthesis of amide-linked chaiyaphumine analogues

With a successful synthetic route to the chaiyaphumines in hand, it was decided that a series of more stable lactam analogues should be synthesized in order to evaluate their biological activity alongside the natural products.[44] A synthesis of the proposed derivatives, in which the ester bond has been replaced by an amide linkage that is less susceptible to hydrolysis, is outlined in Scheme 2.3. The synthetic route used previously for **10a–d** (Scheme 2.2) was repeated, instead incorporating the non-proteinogenic L-diaminopropionic acid (Dpr) in place of threonine. N-terminal acylation, cleavage and cyclization steps were then carried out as previously described, furnishing chaiyaphumine lactam analogues **11a–d**, confirmed by analytical HPLC and ESI-MS (Figure 2.9).

Some dimerization was observed during the synthesis of peptide **11b**, and cyclic dimeric by-product **20** (Figure 2.10) was observed in certain HPLC fractions. Figure 2.10 illustrates how singly-charged (monomeric) **11b** is impossible to distinguish from the doubly-charged (dimeric) **20** by *m*/*z* value alone (also the case for singly-charged **20** and singly-charged dimeric species formed from **11b** during ESI-MS analysis, shown). It was possible to confirm the charge state of the molecular ion peak in each case, however, by inspection of the isotopic envelope observed at high resolution to determine $\Delta_m$. A similar analysis was carried out on all peptides synthesized to ensure dimers were not present in the final products. A reduction in dimer formation was observed when additions of the linear precursor peptides were carried out under syringe pump control in order to reduce addition rates and increase reaction mixture dilution.[50]




**Scheme 2.3:** Synthesis of amide-linked chaiyaphumine analogues **11a–d**. SPPS and solution-phase cyclization were carried out as in the previous synthesis, instead incorporating L-diaminopropionic acid in place of threonine (resin loading and initial peptide elongation steps not shown). Reagents and conditions: (i) $RCH_2COCl$ or $(CH_3CO)_2O$ for **11d** (5.0 equiv), DIPEA (10 equiv), DMF, RT, 2 × 1 h; (ii) TFA (95%, water), TIPS, RT, 3–4 h; (iii) PyBOP® (3.0 equiv) and DIPEA (6.0 equiv), DMF, RT, 5 h (high dilution, see text).

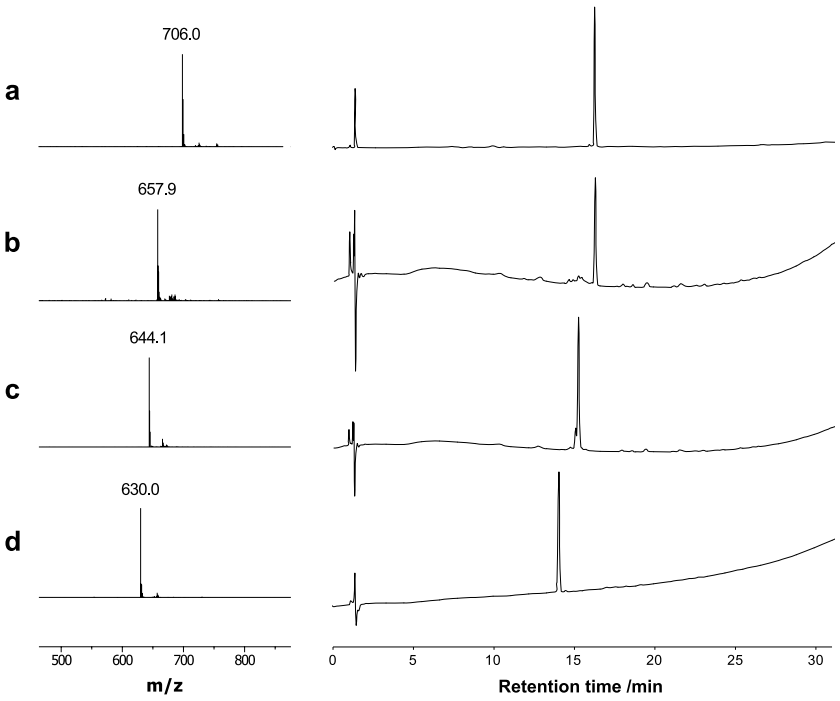



**Figure 2.9:** ESI-MS data (left) and HPLC traces (right) for amide-linked chaiyaphumine analogues **11a–d**. Labelled MS peaks are for $[M + H]^+$ ion. Absorbance recorded at 220 nm, gradient 0-100% B over 30 min (A = 5:95:0.05 MeCN/$H_2O$/TFA; B = 95:5:0.03 MeCN/$H_2O$/TFA).

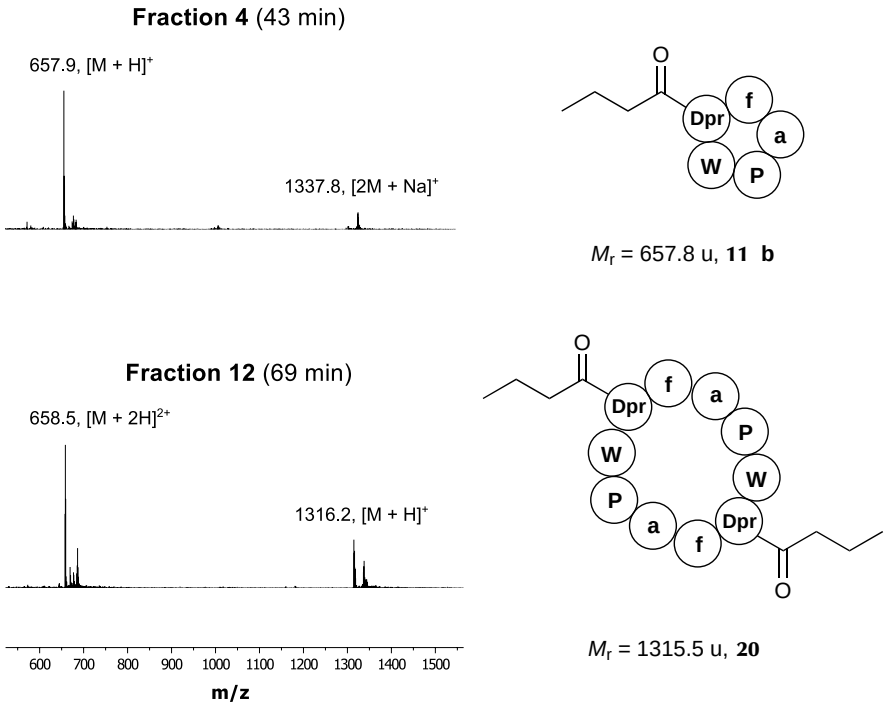



**Figure 2.10:** Low resolution ESI-MS analysis of fractions obtained during HPLC purification of crude peptide **11b**. Both the target peptide and covalent dimer **20** are observed, but cannot be distinguished by their *m*/*z* values alone at this resolution. Determination of the charge state of the molecular ion species present in each fraction enabled assignment to either the monomeric or dimeric peptide (see text).

## 2.5 Biological testing

Synthetic chaiyaphumines **10a** and **10c**, and analogues **11a–d** were screened for antiparasitic activity against *Trypanosoma brucei rhodesiense*, *Trypanosoma cruzi*, *Plasmodium falciparum* and *Leishmania donovani* (Marcel Kaiser, Swiss Tropical and Public Health Institute). Cytotoxic activity was determined against mammalian L6 cells. The bioactivity data are summarized in Table 2.1.

**Table 2.1:** Antiparasitic activity and cytotoxicity data for chaiyaphumines **10a** and **10c** and amide-linked analogues **11a–d**. Literature values for the naturally isolated peptides published previously are shown in brackets.[43]

| *Compound* | | $IC_{50}$ /$\mu M$ | | | | |
|---|---|---|---|---|---|---|
| | | *T. b. rhodesiense* | *T. cruzi* | *P. falciparum* | *L. donovani*[a] | L6 |
| **10** | **a** | 7.96 (5.11) | 9.01 (56.4) | 2.96 (0.61) | >10 | >10 (92.25) |
| | **c** | 60.9 (77.58) | 90.0 (98.94) | 21.6 (15.4) | >100 | >100 (>151) |
| **11** | **a** | 61.4 | 76.7 | 2.87 | >100 | >100 |
| | **b** | 7.00 | 8.67 | 4.7 | >10 | >10 |
| | **c** | 5.93 | 8.18 | >5 | >10 | >10 |
| | **d** | 63.1 | 79.5 | >50 | >100 | >100 |
| Ref[b] | | 0.01 (0.01) | 1.91 (1.73) | 0.004 (0.006) | 0.127 | 0.016 (0.02) |

[a] Axenic amastigotes.
[b] Melarsoprol (*T. b. rhodesiense*), benznidazole (*T. cruzi*), chloroquine (*P. falciparum*), miltefosine (*L. donovani*) or podophyllotoxin (mammalian L6 cells).

Synthetic peptides **10a** and **10c** had similar activity profiles to those of the naturally isolated versions published by Grundmann et al.[43] Propionyl derivative **10c** exhibited negligible inhibitory activity against *T. brucei* (60.9 μM) and *cruzi* (90.0 μM), and only weak activity against *P. falciparum* (21.6 μM), as previously described. Phenylacetyl chaiyaphumine **10a** displayed the inverse profile, with activity against *T. brucei* (7.96 μM), *cruzi* (9.01 μM) and *P. falciparum* (2.96 μM), in agreement with published results except for *T. cruzi*, against which the naturally isolated peptide showed negligible activity (56.4 μM).

Interestingly, the opposite appears to be true for amide-linked chaiyaphumine analogues **11a–d**. While alkyl derivatives **11b** and **11c** exhibit antiparasitic activity (5-9 μM), phenylacetyl derivative **11a** exhibits negligible activity against all parasites (>60 μM), with the exception of *P. falciparum* (2.87 μM). Reducing alkyl chain length, as in the case of **11d**, appears to have a deleterious effect on activity (>50 μM). This difference in activity profile observed for the amide variants may suggest that the lability of the ester bond present in **10a** is important for its biological activity, and that some form of ring-opening and linearization may occur in its mechanism of action. Furthermore, the fact that analogue **11a** appears to contradict this trend, exhibiting activity against *P. falciparum* when **10c** does not, may point to a different target for this compound in *P. falciparum* than in *Trypanosoma*.

The differing activity profile may also be a result of possible decreased cell permeability of the amide analogues. The replacement of an ester bond with a primary amide necessarily introduces a hydrogen bond donor into the peptide structure, which can negatively affect cell permeability.[51] The effect of this modification will of course depend on whether the peptides in question act on cell-surface or intracellular targets. The introduction of a methyl group onto the Dpr side chain/Trp C-terminal amide nitrogen of amides **11a–d** would remove hydrogen bonding potential, and would be useful for investigating the effect of hydrogen bonding ability on biological activity (see below).[52]

None of the synthetic peptides showed any appreciable inhibitory activity against *Leishmania donovani* at the concentrations tested. Significant cytotoxicity against mammalian L6 cells was not observed for any of the peptides tested.

## 2.6 NMR characterization

### 2.6.1 Spectral assignment

Structural characterization of **10a** was carried out using NMR spectroscopy. 2D $^{1}$H-$^{1}$H COSY, $^{1}$H-$^{13}$C HSQC, $^{1}$H-$^{13}$C HMBC and $^{1}$H-$^{1}$H NOESY NMR data were collected and are given in Appendix A. Distinctive resonances such as those belonging to the Trp indole NH (10.79 ppm), and the and Ala and Thr methyl groups were assigned first. The characteristic couplings between $CH_2$ protons in the Pro spin system, in addition to the benzyl $CH_2$ protons, were also identified from the COSY spectrum. Proton and carbon assignments within amino acid side chain spin systems were then carried out as far as possible by using COSY and HSQC data. HMBC data was then used to assist in the identification of remaining resonances, and to assign amino acid connectivity. A full assignment is given in Table 2.2, alongside the literature assignment of the natural isolate. Shift differences (Figure 2.11) were observed between the two compounds, but this is assumed to be due to putative proline *cis*/*trans* isomerisation (as discussed above).

### 2.6.2 Structure determination

Structural modelling using the NOESY NMR data was carried out using ARIA (Ambiguous Restraints for Iterative Assignment).[53] ARIA iteratively reassigns ambiguously assigned NOESY cross-peaks during the structure calculation, thus allowing automatic assignment of overlapping signals (such as phenyl group aromatic resonances). The software uses CNS (Crystallography & NMR System)[54,55] for structure refinement against NOE-derived distance restraints using a modified version of the PARALLHDG force field.[56] Typical simulation runs involve a combination of simulated annealing from a high initial temperature using fast torsion angle dynamics followed by conventional Cartesian dynamics structure refinement.[57] Peak reassignment occurs following each iteration.

Peak picking and assignment of the $^{1}$H-$^{1}$H NOESY spectrum was carried out using the CCPN (Collaborative Computing Project for NMR) software.[58] Initial distance restraints were then calculated using the software, and restraint and peak lists were exported to ARIA using the inbuilt interface. Due to the presence of D-amino acid residues, the N-terminal phenylacetyl capping

**Table 2.2:** NMR spectroscopic data for chaiyaphumine **10a**, 600 MHz ($^1$H) and 150 MHz ($^{13}$C) in $d_6$-DMSO.

| | *Synthetic peptide **10a*** | | *Peptide **10a** literature*[a] | |
|---|---|---|---|---|
| | $\delta$ $^{13}$C /ppm | $\delta$ $^1$H /ppm, mult (*J* /Hz) | $\delta$ $^{13}$C /ppm | $\delta$ $^1$H /ppm, mult (*J* /Hz) |
| C=O | 168.5 | | 169.4 | |
| NH | | 7.96, d (6.7) | | 8.06, d (10.1) |
| $\alpha$ | 55.0 | 4.35, t (6.8) | 54.9 | 4.57, m |
| $\beta$ | 69.0 | 5.08, quint (6.6) | 71.3 | 5.31, m |
| $\gamma$ | 15.0 | 0.99, d (6.5) | 15.5 | 0.66, d (6.5) |
| C=O | 170.8 | | 172.3 | |
| NH | | 8.96, d (7.4) | | 7.90, d (9.9) |
| $\alpha$ | 57.3 | 4.23, ddd (11.3, 7.4, 4.7) | 56.6 | 4.64, m |
| $\beta$1 | 35.9 | 3.00, dd (14.0, 4.7) | 35.7 | 3.20, m |
| $\beta$2 | | 2.89, dd (14.0, 11.3) | | 3.08, dd (14.2, 9.7) |
| Ar 1 | 137.6 | | 137.5 | |
| Ar 2 | 128.8 | 7.24-7.27, m | 128.4[b] | 7.26, m |
| Ar 3 | 128.2 or 128.1[c] | 7.21-7.29, m | 128.5[b] | 7.27, m |
| Ar 4 | 126.4[c] | 7.13-7.17, m | 126.6[d] | 7.18, m |
| C=O | 170.4 | | 172.7 | |
| NH | | 7.25-7.31, m | | 8.23, d (6.7) |
| $\alpha$ | 47.1 | 4.41, quint (6.6) | 47.4 | 4.54, m |
| $\beta$ | 17.7 | 1.20, d (6.6) | 15.9 | 1.36, d (7.0) |
| C=O | 170.3 | | 170.1 | |
| $\alpha$ | 61.4 | 3.97, dd (8.8, 5.4) | 60.8 | 4.27, dd (8.9, 2.3) |
| $\beta$1 | 28.4 | 1.92-1.99, m | 28.8 | 1.89, m |
| $\beta$2 | | 1.33, dq (12.2, 6.1) | | 1.46, m |
| $\gamma$1 | 24.4 | 1.79-1.87, m | 23.2 | 1.62, m |
| $\gamma$2 | | 1.79-1.87, m | | 1.03, m |
| $\delta$1 | 46.7 | 3.67-3.72, m | 46.7 | 3.73, m |
| $\delta$2 | | 3.48-3.53, m | | 3.46, m |
| C=O | 171.2 | | 170.5 | |
| NH | | 7.58, d (9.9) | | 6.9, d (8.1) |
| $\alpha$ | 55.4 | 4.58-4.63, m | 53.0 | 4.39, m |
| $\beta$1 | 27.9 | 2.94-2.97, m | 26.9 | 3.21, m |
| $\beta$2 | | 2.94-2.97, m | | 2.96, dd (14.5, 11.5) |
| Indole 1 | | 10.79, d (2.1) | | 10.81, d (1.9) |
| Indole 2 | 122.9 | 7.01, d (2.1) | 124.0 | 7.23, m |
| Indole 3 | 109.9 | | 109.9 | |
| Indole 3a | 127.2 | | 126.9 | |
| Indole 4 | 118.10 | 7.50 , d (8.0) | 118.0 | 7.54, d (7.9) |
| Indole 5 | 118.07 | 6.97, ddd (8.0, 7.0, 1.0) | 118.5 | 6.97, t (7.4) |
| Indole 6 | 120.9 | 7.05, ddd (8.0, 7.0, 1.2) | 121.0 | 7.03, t (7.7) |
| Indole 7 | 111.3 | 7.29-7.32, m | 111.4 | 7.31, d (8.1) |
| Indole 7a | 136.0 | | 136.2[e] | |
| C=O | 170.3 | | 171.3 | |
| $\alpha$(1/2) | 40.1 | 3.26, d (15.7) | 43.4 | 3.88, d (13.0) |
| | | 3.36, d (15.7) | | 3.49, d (13.0) |
| | 134.0 | | 136.1[e] | |
| | 126.7[c] | 7.19-7.27, m | 128.8 | 7.25, m |
| | 128.2 or 128.1[c] | 7.21-7.29, m | 129.1 | 7.37, d (7.1) |
| | 129.2 | 7.12-7.16, m | 136.5[b] | 7.18, m |

[a] Literature values for the naturally isolated compound as published previously.[43]
[b] One peak with an integral for two carbons.
[c,d,e] Assignments are interchangeable (second assignment not specified for d).

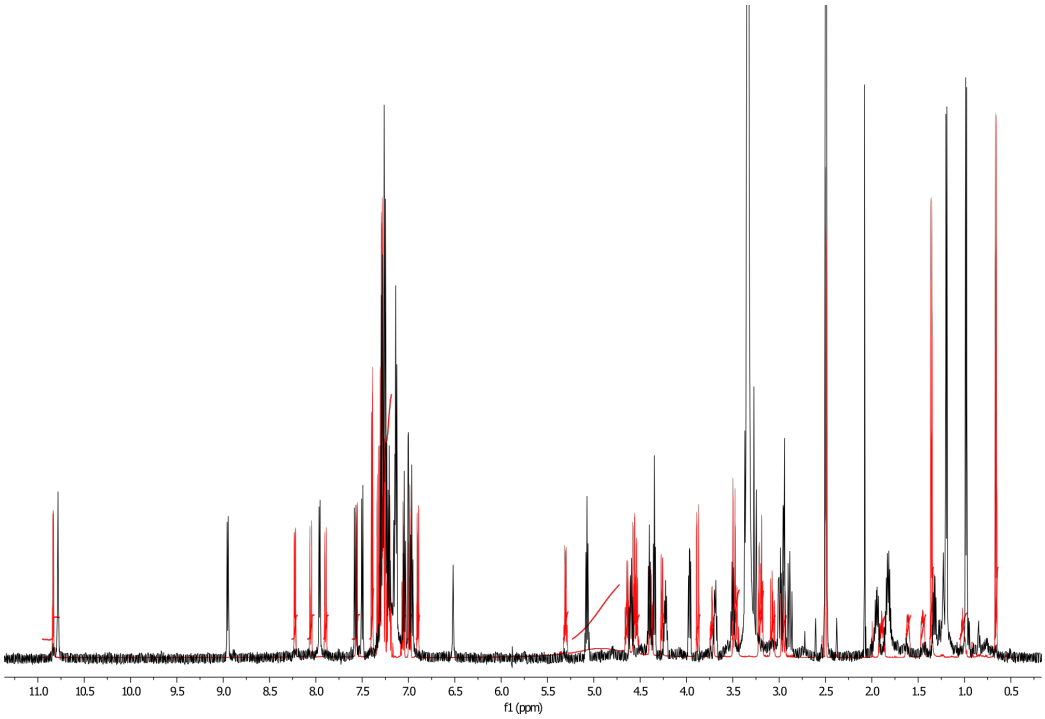



**Figure 2.11:** Overlay of $^{1}$H NMR spectra of **10a** (black) and natural isolate[43] (red). Chemical shift differences are observed.

group and a depsipeptide bond in the structure, the ARIA run parameters and scripts generated by default had to be heavily modified (as described in the Experimental section).

Torsion angle dynamics simulations attempted using the default parameters were found to cause instabilities due to the cyclization restraint and relatively small molecular size, which were not resolved upon decreasing high temperature values by more than half. Simulated annealing runs were therefore carried out using Cartesian dynamics only (with the default simulation total time period quadrupled), and were found to be stable when using the same high temperature value (10 000 K). An overlay of the final converging structural ensemble is given in Figure 2.12.

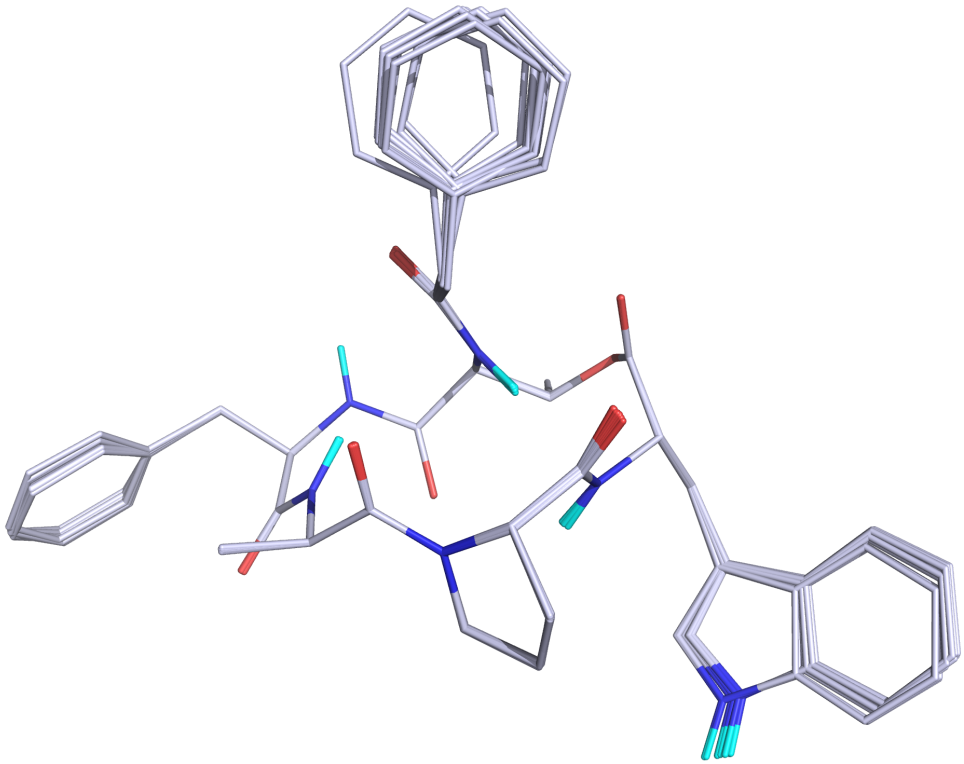


**Figure 2.12:** Overlay of 10 lowest energy conformers of **10a** following molecular dynamics energy minimization using NOE-derived distance restraints (see text).

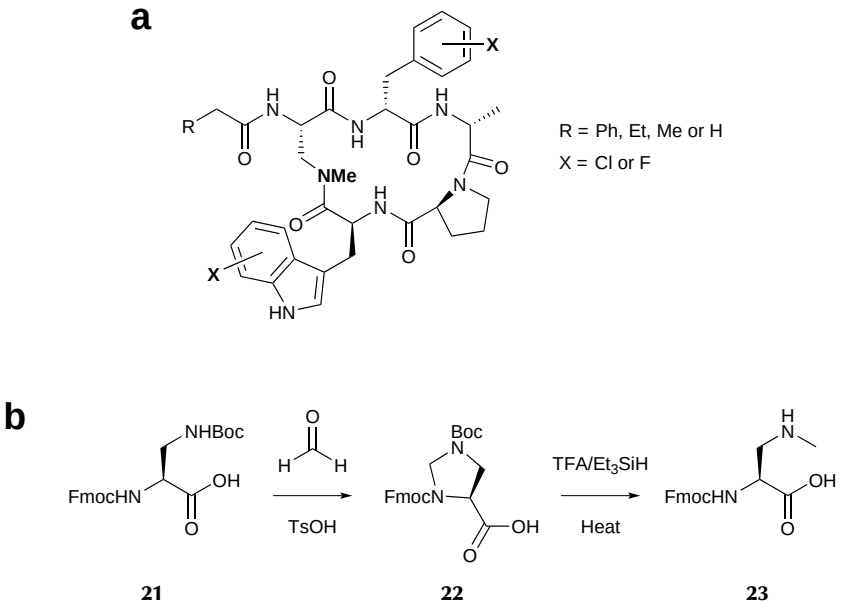
## 2.7 Conclusions and future work

A synthetic route to the chaiyaphumines, a newly discovered class of antiparasitic cyclic depsipeptide natural products, has been successfully developed. Correct stereochemistry was confirmed via Marfey's analysis, and the synthetic peptides exhibited antiparasitic bioactivity comparable to the naturally isolated peptides. In addition, a series of chaiyaphumine analogues were synthesized in which the ester bond was replaced by an amide linkage, and these exhibited a similar potency of antiparasitic action ($<3$ μM for **11a** against *P. falciparum*), although interestingly with respect to the N-terminal capping moiety the activity profile was reversed (see above).

Future work will involve assaying stability of the synthetic chaiyaphumines and analogues under different pH conditions and with regards to enzymatic degradation, to demonstrate the increased stability of the amide-linked analogues over the natural ester structures. Structure-activity studies will also be critical in further elucidating those structural features essential for activity. This will involve the synthesis of analogues bearing amino acid substitutions, such as the replacement of D- with L-amino residues to evaluate the effect of stereochemistry. In particular the role of the capping group functionality (previously demonstrated to be important) should be probed – for example further studies of chain length with regards to the phenyl-substituted version, to contrast that for the aliphatic capping moiety already investigated.

Of particular interest are those modifications likely to increase hydrophobicity and reduce hydrogen bonding ability, potentially resulting in increased cell permeability. Examples of this type, such as the halogenation of aromatic side chains or the introduction of the Dpr side chain methylation discussed in section 2.5, are given in Scheme 2.4a. These can be incorporated during Fmoc SPPS using the appropriate commercially available building blocks. In addition, a route to the methylated Dpr building block required for the latter modification is proposed in Scheme 2.4b. Although some difficulty is anticipated during cyclizations involving the relatively hindered methylated Dpr side chain, the already demonstrated propensity of the linear chaiyaphumines to readily undergo macrocyclizations at room temperature suggests that this will be feasible given the correct choice of activator and optimization of reaction conditions.

The elucidation of the mechanism of action of the chaiyaphumines is also necessary, and the identification of protein targets of this class of antiparasitics




**Scheme 2.4:** (a) Suggested modifications to the chaiyaphumine scaffold, potentially introducing enhanced cell permeability. *N*-Methylation of the isopeptide bond can be incorporated using the appropriate methylated Dpr building block (see below). Commercially available halogenated Phe and Trp building blocks facilitate introduction of chloro and fluoro substitutions on aromatic side chains. (b) A two-step route from Fmoc-Dpr(Boc)-OH (**21**) to methylated intermediate **23** recently published by Lindahl et al.[59] **23** can be Boc protected to give the methylated Dpr building block required above.

is vital. The work described provides access to much larger amounts of peptidic material than was previously available, thus paving the way for target validation studies using affinity chromatography-based techniques (in which the active peptide is immobilized on a stationary phase followed by washing with protein extracts to identify binders).[60] Drug affinity responsive target stability (DARTS) has also been used relatively recently to confirm or identify a number of small molecule-protein interactions, and selects proteins based on their induced stability with respect to proteolysis upon binding of the ligand of interest.[60] Finally, the most valuable advantage of obtaining natural products via a chemical route is the potential for chemical modification. The introduction of reactive electrophilic functionality into the chaiyaphumine scaffold would allow covalent attachment to catalytic residues in the active site of a potential target enzyme, and could be used in a target-based screen.[60]

# Chapter 3

# Synthesis of the lasso peptide lassomycin

As illustrated in Chapter 1, the pressing issue of antimicrobial resistance continues to drive the search for novel antiparasitic drug scaffolds. This is of course an equally important concern in *antibiotic* drug discovery, and emerging threats such as extensively drug-resistant tuberculosis require the use of innovative approaches in the search for potential leads.[61–63]

The following chapter describes the total synthesis of a peptide natural product found to be active against drug-resistant forms of *Mycobacterium tuberculosis*. The peptide is of particular interest due to the fact that it belongs to the class of 'lasso' peptides, which have been found to exhibit a unique three-dimensional fold. The use of total chemical synthesis in the elucidation of the structure of a natural product is demonstrated. An overview of lasso peptides is also given.

## 3.1 Introduction

### 3.1.1 Lasso peptides

Lasso peptides are a class of cyclic peptide natural products of bacterial origin.[64,65] Members of this class contain a post-translationally installed isopeptide bond between the N-terminus and an aspartate or glutamate side chain, resulting in a cyclic structure.[65,66] Some of the known lasso peptides

exhibit antibacterial activity and are secreted into the growth medium, where their putative role is to defend against competing bacterial strains within the same habitat.[67]

The compact cyclic structure of lasso peptides favours receptor interactions and confers resistance to unfolding or proteolytic degradation.[67,68] For this reason it is thought that the loop region could be used for the presentation of receptor binding epitopes – lasso peptides are therefore of interest in the development of peptide drug scaffolds.[65] In 2011 Knappe et al. demonstrated low cytotoxicity and high stability in physiological conditions of a lasso peptide integrin antagonist, where a tripeptide sequence in the loop region was substituted for the integrin recognition sequence 'RGD'.[69]

Only a handful of lasso peptide crystal structures have been published to date.[67,70–72] The first, published by Nar et al. in 2010, is shown in Figure 3.1. The peptide exhibits a conformation whereby the C-terminus is 'threaded' through the loop and held, either by steric interactions (in the case of the peptide shown) or disulfide bonds.[66] Attempts to chemically synthesize threaded lasso peptides have not been successful thus far,[67] although the synthesis of an interlocked peptide-containing rotaxane exhibiting a lasso topology has been described.[73]

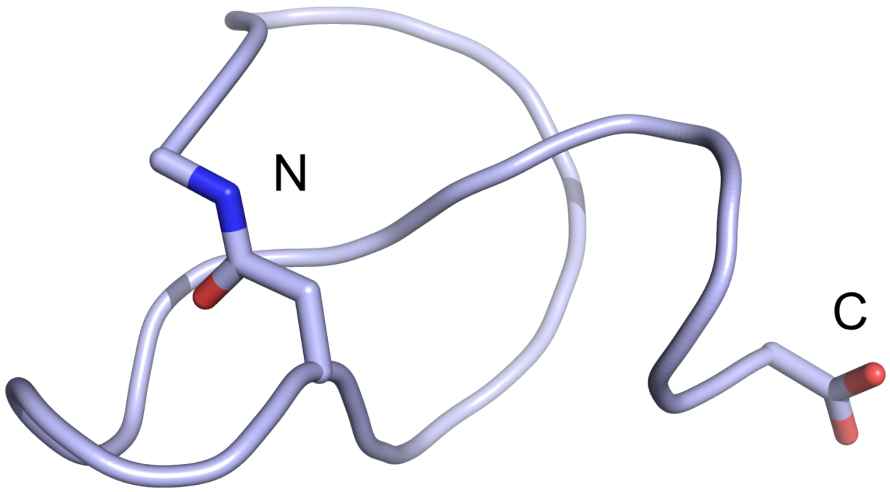



**Figure 3.1:** X-ray crystal structure of BI-32169 exhibiting the characteristic threaded loop conformation common to all lasso peptides discovered to date (except one, see text).[36,67]

### 3.1.2 Lassomycin

Lassomycin (**24**) is a newly discovered lasso peptide isolated from an extract of the Actinomycete *Lentzea kentuckiensis* sp. in a screen of a library of previously uncultured[I] soil bacteria.[74] Lassomycin was found to possess potent and specific bactericidal activity against mycobacteria, including drug-resistant forms of *Mycobacterium tuberculosis*. It was discovered that the peptide targeted the ClpC1 ATPase, an essential enzyme in mycobacteria, thus accounting for the specificity of **24**. ClpC1 ATPase is a key component of protein degradation together with the ClpP1P2 proteolytic complex.[75] The primary sequence and connectivity of **24**, as confirmed by preliminary MS and NMR analysis, is shown in Figure 3.2.

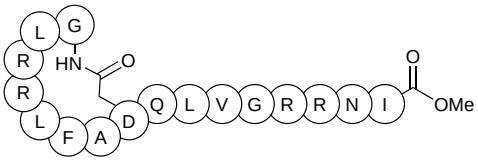



**Figure 3.2:** Primary sequence and connectivity of amino acids in lassomycin (**24**).[74] The peptide contains an isopeptide bond between the N-terminus and the carboxylate side chain of Asp-8, forming an 8-residue macrolactam ring. The C-terminus is methylated.

The reported NMR solution structure of lassomycin is shown in Figure 3.3. Remarkably, the peptide does not appear to adopt the characteristic threaded knot conformation reported for other homologous lasso peptides discovered to date.[65] In light of this discovery, and the potential of lasso peptides as novel drug scaffolds in general,[65,67,68] it was decided that a chemical synthesis of the natural product should be attempted to confirm its structure and investigate its biological activity. It was envisaged that a total synthesis would also provide enough material for crystallization and full structural characterization via X-ray diffraction methods.

[I] Bacteria that do not grow (or grow very poorly) *in vitro*, and represent an untapped source of secondary metabolites for drug discovery.[63,74]

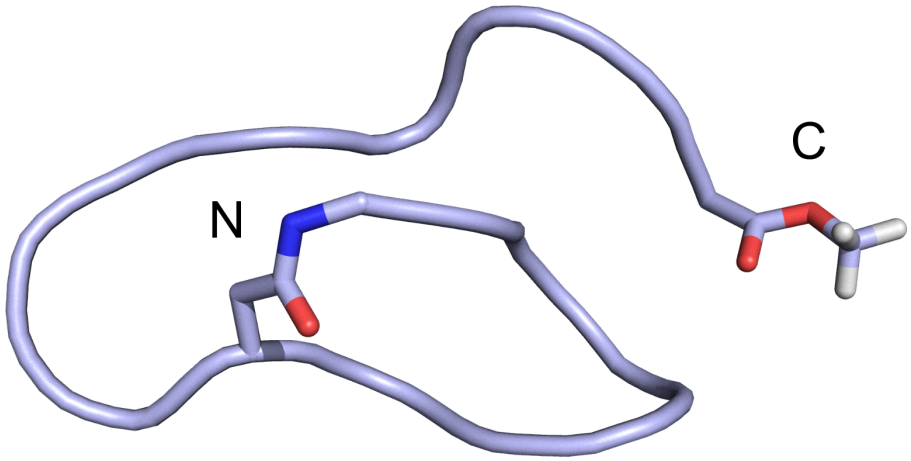


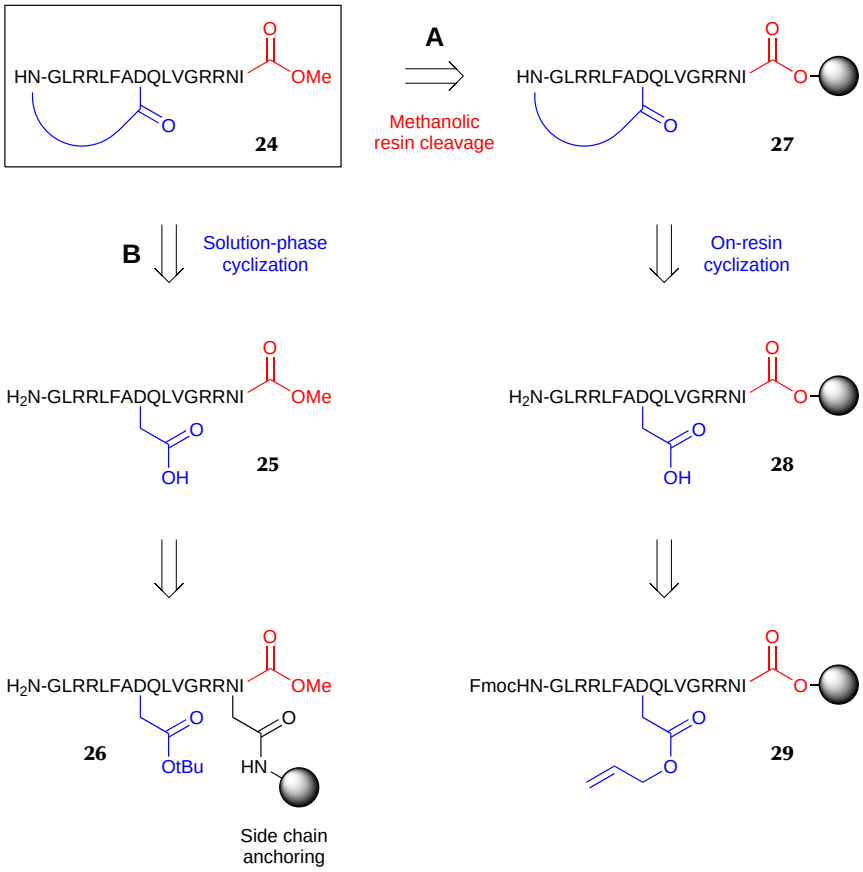

**Figure 3.3:** NMR solution structure of lassomycin (**24**).[36,74] The proposed structure of **24** is unique in that the C-terminus packs tightly against the N-terminal ring instead of passing through the macrolactam.

## 3.2 Proposed synthetic routes and initial attempts

Two key structural features that need to be addressed during a chemical synthesis of lassomycin (**24**) are the isopeptide bond between the N-terminus and the carboxylate side chain of Asp-8, and the methylated C-terminus (Figure 3.2). Two synthetic approaches were proposed and are outlined in the retrosynthesis given in Scheme 3.1.

Route **A** involves the synthesis of the linear sequence on 2-chlorotrityl chloride resin, followed by head-to-side-chain cyclization of the protected peptide on resin. The side chain of Asp-8 was protected using the orthogonal allyl ester protecting group. Treatment with methanolic anhydrous HCl was chosen as a convenient method of simultaneously cleaving the product peptide from the resin and generating the desired C-terminal methyl ester in the approach reported by Turner at al.[76]

Synthesis of the linear sequence was carried out successfully, and aqueous test cleavage of **29** (Scheme 3.1) using TFA generated the correct linear Fmoc/allyl ester intermediate bearing a C-terminal acid group, as confirmed by MALDI-TOF MS ($m/z$ = 2146.3). Obtaining C-terminal peptide methyl esters was found to be more problematic however, and the target peptide methyl




**Scheme 3.1:** Proposed synthetic routes to Lassomycin (**24**). On-resin cyclization approach **A** uses orthogonal protection in the form of an allyl ester protecting group, which was found to be difficult to remove (see text). A solution-phase cyclization strategy **B** was instead considered, which involves installation of the C-terminal methyl ester at the beginning of the synthesis as opposed to via a cleavage-esterification step at the end (as in **A**). (Amino acid side chain protecting groups are omitted for clarity.)

ester was not obtained using the previously published conditions (cleavage cocktail made up of a 3 M solution of acetyl chloride in methanol, to generate HCl *in situ*).[76]

Removal of the allyl ester protecting group also proved to be highly problematic, and no deprotected product was observed after 5 × 1 h treatments with 20 equivalents of phenylsilane in the presence of a $Pd(PPh_3)_4$ catalyst (0.35 equiv) under an inert atmosphere.[77] The conditions for various (unsuccessful) deprotection attempts are summarized in Table 3.1.

**Table 3.1:** Summary of conditions attempted for allyl ester deprotection.[77–79]

| *Pd(0) equiv* | *Other* | *Solvent* | *Duration* |
|---|---|---|---|
| 0.35 | $PhSiH_3$ (20 equiv) | DCM | 5 × 1 h |
| 1.0 | $PhSiH_3$ (20 equiv), $CH_3COOH$ (20 equiv) | DCM | 2 × 2 h |
| 2.0 | $CHCl_3$/$CH_3COOH$/NMM (37:2:1) | Chloroform | 2 × 2 h |

Allyl ester group removal was not observed despite the use of stoichiometric amounts of catalyst and silane, and trialling in a number of reaction solvents. The lack of success was assumed to arise from a difficulty in coordinating the allyl moiety to the catalyst, presumably due to steric hindrance around the metal centre. This effect will depend upon peptide sequence length and folding of the peptide, and as such is likely to be highly sequence dependent.

In light of the difficulties described, alternative orthogonal protecting groups were considered for the side chain of Asp-8, such as highly acid sensitive,[80–82] photolabile[83,84] or silicon-based protection.[85,86] It was decided that these were less than ideal, however, due to their sensitivity, requirement for specialized conditions for deprotection, or a lack of commercially available building blocks. Instead an alternative solution-phase cyclization approach (Route **B**) was proposed.

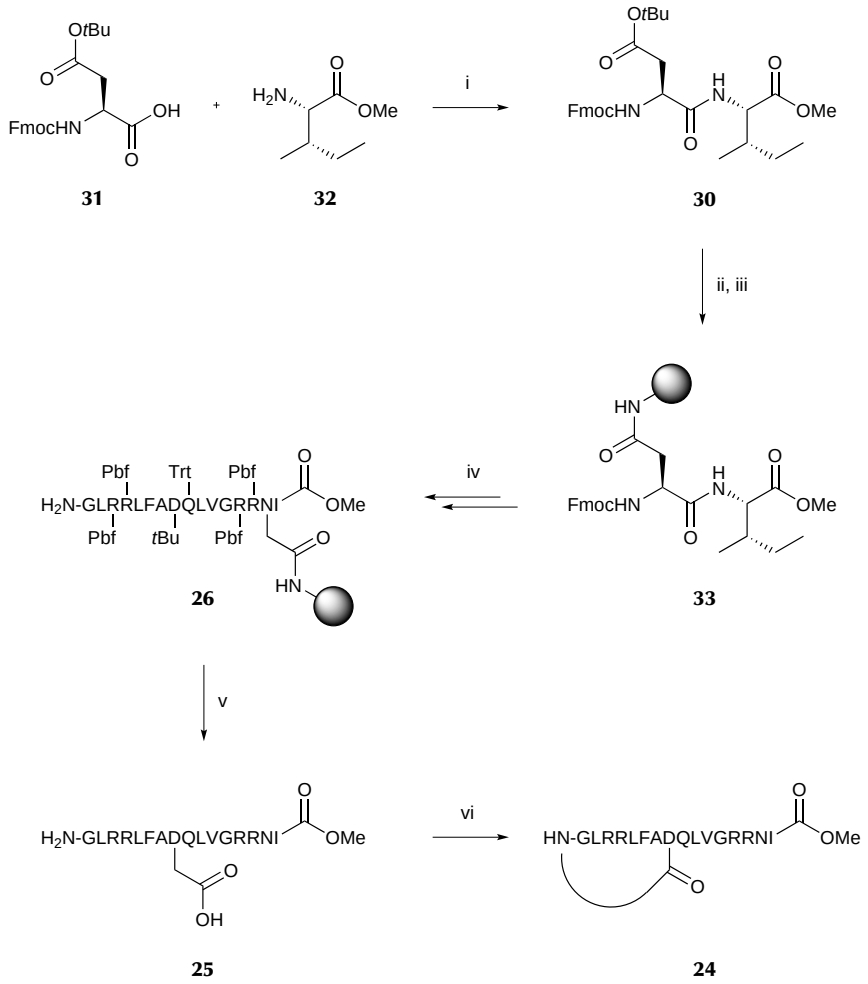
## 3.3 Solution-phase cyclization strategy

Route **B** is outlined in Scheme 3.1. This approach involves assembly of the side-chain-anchored linear peptide on resin, followed by cyclization of the product in solution after global deprotection and cleavage from the support. Due to the use of side chain anchoring, the C-terminal methyl ester functionality can be installed at the beginning of the synthesis.

A dipeptide of aspartic acid and isoleucine methyl ester (**30**) was synthesized in solution and the *t*-butyl group removed from the aspartate side chain under standard acidic conditions. The dipeptide was then attached via the deprotected side chain to Rink amide resin (0.77 mmol/g) in order to give the required Asn-Ile-OMe C-terminal sequence upon TFA cleavage (Scheme 3.2). The required sequence was then elongated via automated SPPS at room temperature using PyBOP®/DIPEA activation and 2 × 1 h couplings (3 × 1 h for Arg). Fmoc deprotection of the penultimate residue (Leu-2) was found to be inefficient, and an additional two further deprotection steps were incorporated into the synthesis at this point. The fully deprotected linear peptide (**25**) was obtained after cleavage from the resin using 95% TFA, as confirmed by ESI-MS ($m/z$ = 1898.4, $[M + H]^+$).

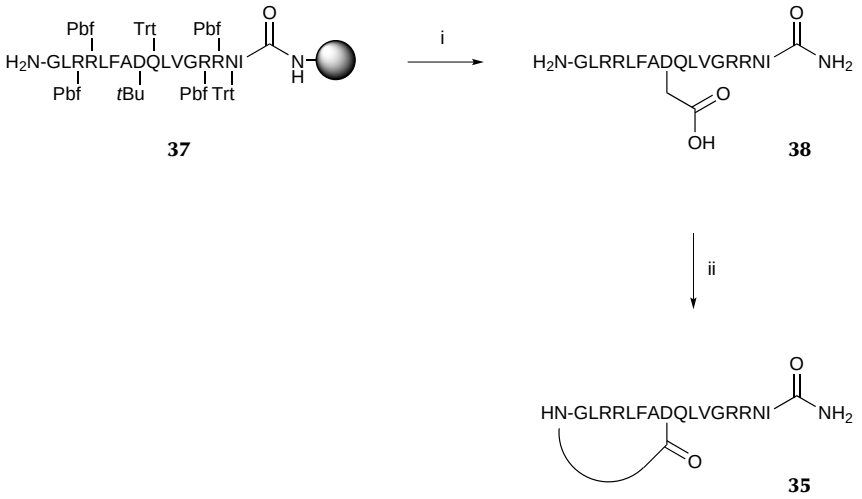
Cyclization of linear precursor **25** was then attempted. While peptide macrocyclizations are often best performed under high dilution to minimize unwanted oligomer formation,[87] the pseudo-high dilution conditions described by Malesevic et al. were instead utilized as a more convenient alternative in order to minimize solvent usage.[50] 11.3 mL each of solutions of **25** (0.01 M) and HATU (0.03 M) in DMF were added to a reaction flask containing DIPEA and HOAt over a period of ~20 h at a rate of 0.01 mL/min (syringe pump controlled). Slow addition was used in order to keep the effective reaction concentration low, while requiring only a relatively small final solvent volume (34 mL). Removal of the solvent under vacuum and HPLC purification of the crude reaction product yielded lassomycin (**24**) in high purity, as confirmed by analytical HPLC and MALDI-TOF MS (Figure 3.4).

The C-terminal amide analogue (**35**) of lassomycin was also synthesized (Scheme 3.3). Synthesis of the linear peptide was carried out directly on Rink amide resin via Fmoc SPPS (without side chain anchoring). Cleavage from the resin, solution-phase cyclization and HPLC purification to yield **35** were carried out as previously described for the synthesis of **24**. Purity and correct molecular weight were confirmed by analytical HPLC and MALDI-TOF MS




**Scheme 3.2:** Synthesis of lassomycin. The dipeptide Fmoc-Asp(OH)-Ile-OMe (**34**, not shown) was preformed in solution and anchored to Rink amide resin via the Asp side chain. The required sequence was then elongated via SPPS, before global deprotection and cleavage from the resin. Cyclization of the linear peptide in solution yielded the cyclic target. Reagents and conditions: (i) PyBOP® (1.0 equiv) and NMM (3.0 equiv), DCM, RT, 18 h; $SiO_2$, 98/2% hexane/EtOAc → 50/50% hexane/EtOAc; (ii) 40% TFA, DCM, RT, 30 min; (iii) Rink amide resin (0.79 mmol/g), **34** (3.1 equiv), PyBOP® (3.1 equiv) and DIPEA (6.2 equiv), DMF, RT, 2 × 1 h; (iv) Fmoc SPPS (see Experimental section); (v) TFA (95%, water), TIPS, RT, 4 h; (vi) HATU (3.0 equiv) and DIPEA (6.0 equiv), DMF, RT, 20 h (pseudo-high dilution, see text).

(Figure 3.4). Lassomycin(1-9) (**36**), a truncated analogue, was synthesized by the same route in order to assess the effect of loss of the tail region on biological activity (characterization shown).




**Scheme 3.3:** Synthesis of lassomycin-amide (**35**). The peptide sequence was synthesized via SPPS on Rink amide resin (0.79 mmol/g), before global deprotection and cleavage from the resin. Cyclization of the linear peptide in solution yielded the cyclic target. Reagents and conditions: (i) TFA (95%, water), TIPS, RT, 4 h; (ii) HATU (3.0 equiv) and DIPEA (6.0 equiv), DMF, RT, 20 h (pseudo-high dilution, see text).

## 3.4 MS characterization and biological testing

MALDI TOF/TOF MS/MS using LIFT[88] technology was used to provide sequence information for the peptides synthesized. This technique enables the detection of product ions resulting from fragmentation in the field free region of the mass spectrometer, after acceleration from the ion source and as a consequence of elevated laser power. Conservation of momentum means the product ions will maintain the velocity of their corresponding precursor ion. In the case of LIFT, the product ions subsequently experience an additional voltage potential that 'lifts' them to an energy that enables them to pass through the rest of the mass spectrometer and be detected. Backbone amide

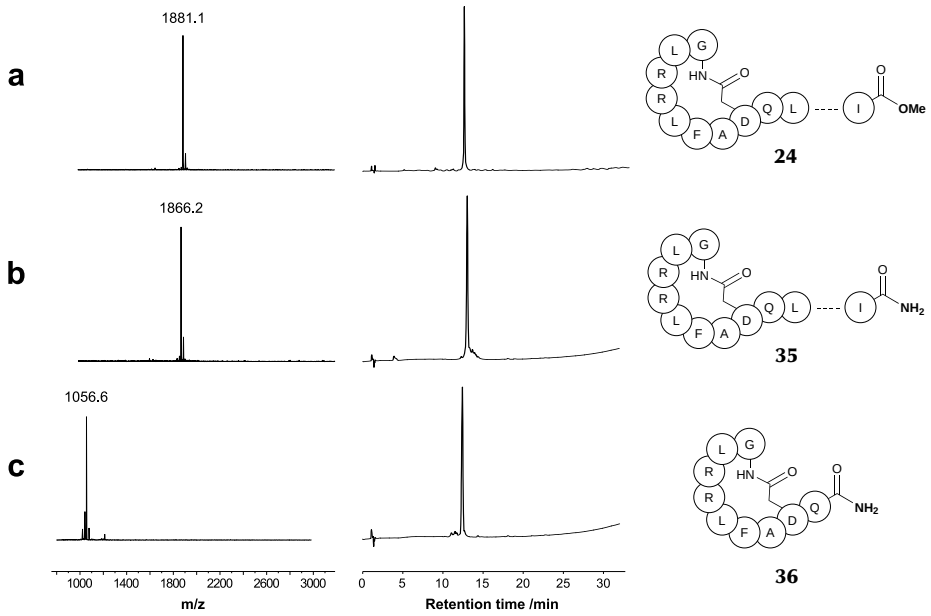



**Figure 3.4:** MALDI-TOF mass spectra (left) and HPLC traces (centre) for (a) lassomycin (**24**), (b) lassomycin-amide (**35**) and (c) lassomycin(1-9) (**36**). Labelled MS peaks are for $[M + H]^+$ ion. Absorbance recorded at 220 nm, gradient 0-100% B over 30 min (A = 5:95:0.05 MeCN/$H_2O$/TFA; B = 95:5:0.03 MeCN/$H_2O$/TFA).

bond cleavages are the most commonly used for peptide sequencing and result in series of product ions extending from either the N-terminus ('**b**' ions, retaining the carbonyl part of the amide bond) or C-terminus ('**y**' ions, retaining the nitrogen atom).[89]

Product ions observed for lassomycin (in addition to linear and amide analogues) are shown in Figure 3.5. While a complete set of product ions were observed for the uncyclized peptide, the cyclic peptides only exhibited fragmentation up to Asp-8 (side chain cyclized onto N-terminus). This is in agreement with the results from MS/MS sequencing of the natural product by Gavrish et al.,[74] and is unsurprising given that for a cyclic peptide *two* bond cleavages must occur in order to observe **b** or **y** ion series in the ring region.[II]


**Figure 3.5:** Summary of product ions observed via MALDI LIFT-TOF/TOF sequencing of (a) linear (uncyclized) lassomycin (**25**), (b) lassomycin (**24**) and (c) lassomycin-amide (**35**). Ions resulting from amide (CO-NH) bond cleavage are shown, with the charge retained on either the N-terminal ('**b**' ion) or C-terminal ('**y**' ion) fragment (charged fragment is observed).[89] Asterisked ions are observed only as M-17 (loss of ammonia).

[II]A single cleavage in the ring does not divide the peptide into two fragments as they are still attached at the branching point, hence the resulting 'fragment' has the same mass as the fully intact peptide.

Ion-mobility (IM) mass spectrometry was also carried out on the synthetic peptide. This technique involves separation of ions based on their differential mobility through a buffer gas,[90] and has been used previously to distinguish threaded and unthreaded topoisomers in lasso peptides.[91] The mobiliogram (plot of *m*/*z* against drift time in Bins) is shown in Figure 3.6. A number of peaks with distinct collisional cross sections were resolved belonging to either lassomycin or a trace amount of the C-terminal acid hydrolysis product (**39**). Only a single peak was observed for each species suggesting that only one conformation is present, as opposed to distinct populations of threaded and unthreaded peptide (or that the collisional cross section for threaded/unthreaded ions is identical).

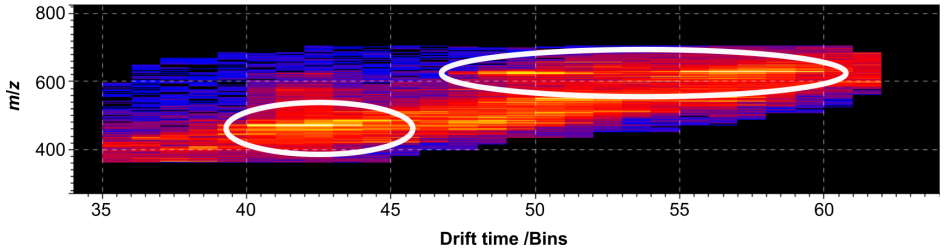



**Figure 3.6:** Mobiliogram for lassomycin (LC ES IM MS, full scan). The collisional cross section for the lassomycin (methyl ester, **24**) and trace hydrolysis product (C-terminal acid, **39**) ions in 4+ charge state are virtually identical (left-hand ellipse), whereas the 3+ charge state ions for the two peptides have different drift times and can be separated (right-hand ellipse). The peak at 50 Bins belongs to **39**, and at 58 Bins to **24**.

Peptides **24**, **35** and **36** were tested against *Mycobacterium tuberculosis* H37Rv in minimum inhibitory concentration (MIC) assays utilizing both solid-agar based and liquid-media based methods (Tim Bull, St. George's University of London). None of the peptides showed inhibitory activity when tested, up to concentrations of 100 μg/mL. These results were surprising, in particular for peptide **24**, given the previously reported biological activity profile for the natural product.[74] Due to the unlikely nature of spontaneous threading during the synthesis of **24** (and **35**), it is logical to assume that the peptides reported are unthreaded. The biological data presented therefore suggests that the unthreaded structure previously reported may not be correct.

## 3.5 Investigation of effect of temperature on peptide conformation

The structural discrepancy from the previous data may possibly be due to unthreading of lassomycin between biological testing and NMR data collection in the published report.[74] Several examples have been documented of lasso peptides unthreading upon heat treatment. Caulosegnin I unthreads completely after 4 hours at 95 °C, while caulosegnin III unthreads and decomposes at elevated temperatures.[92] Similarly, astexin-1 unthreads when heated to 50 °C for 4 hours[68] and astexin-2 unthreads and decomposes after heating at 95 °C for 2 hours.[93] In contrast, caulosegnin II and astexin-3 show minimal to no unthreading when heated at 95 °C for prolonged periods of time.[92,93] NMR data collection was carried out previously at 40 °C however, at which temperature unthreading seems unlikely.

In order to further elucidate the effect of temperature on the conformational state of lassomycin, structural characterization of synthetic peptide **24** was carried out using two-dimensional NMR. A comparison of $^{1}$H-$^{13}$C HSQC spectra for **24** at two different temperatures is shown in Figure 3.7. The NMR data gathered is also compared with naturally isolated labelled [$^{13}$C, $^{15}$N]lassomycin.[74] No noticeable changes in chemical shift values are observed between the spectrum of **24** at 25 °C and at 40 °C. Furthermore, clear differences in the $^{1}$H-$^{13}$C HSQC spectrum can be seen between the synthetic (**24**) and naturally isolated peptide at 40 °C (Figure 3.7). These results strongly suggest that the conformation of **24** does not change appreciably upon heating, and they highlight that there are significant differences between the conformation of synthetically prepared and naturally isolated lassomycin.

In light of this evidence, it is proposed that synthetic lassomycin (**24**) either exists in the unconstrained unthreaded form, or as a threaded lasso that does not unthread upon heating to at least 40 °C (the latter implying that unthreading would also be impossible for the natural isolate, should it exist in a threaded conformation). Given the clear conformational differences observed between synthetic and naturally isolated peptides at the point of NMR data collection, it seems likely that one is threaded while the other is not. As an identical situation has already been established at the point of biological testing, it is proposed that i) neither peptide has undergone unthreading (or threading) upon heating during NMR data collection, or ii)

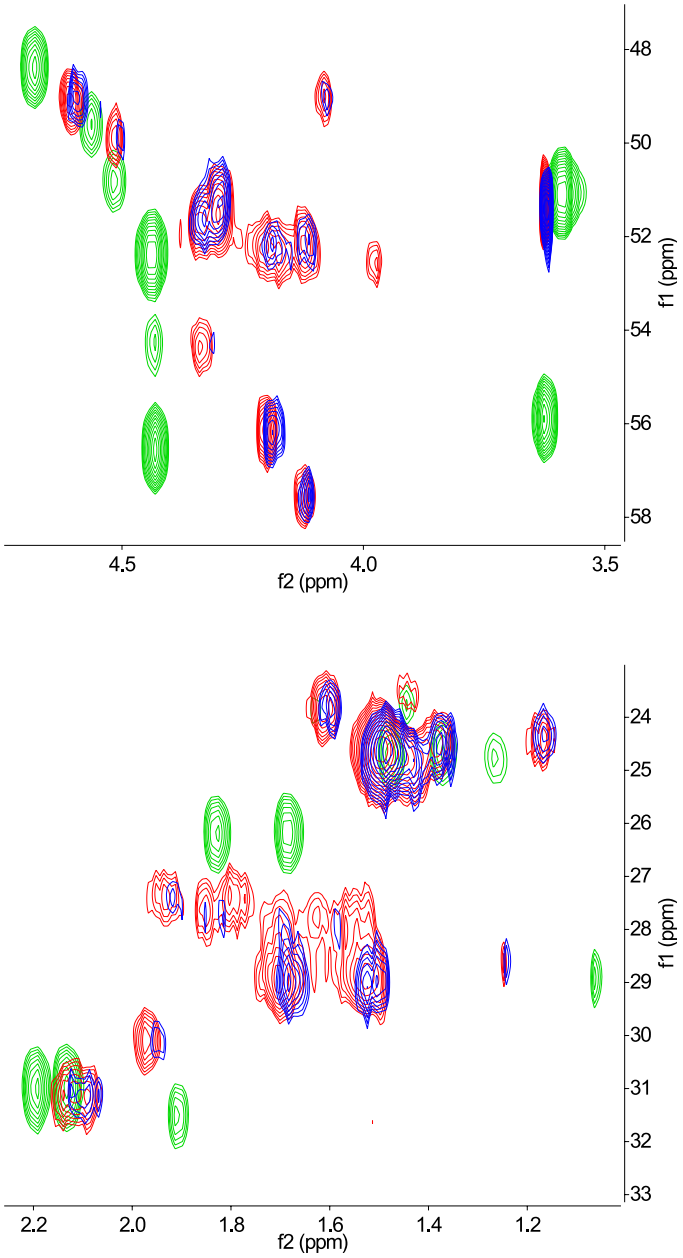



**Figure 3.7:** Overlay of expanded regions of $^{1}$H-$^{13}$C HSQC (heteronuclear single-quantum correlation) NMR spectra of naturally isolated [$^{13}$C, $^{15}$N]lassomycin at 40 °C (green), and synthetic lassomycin (**24**) at 25 °C (blue) and 40 °C (red). (Full spectra given in Appendix B.)

one has undergone threading from an initial unthreaded state whilst the other has undergone unthreading. For case ii) the peptide undergoing unthreading cannot be the natural isolate, as this would imply that the synthetic peptide (**24**) has undergone threading and is threaded at the point of NMR data collection, thus paradoxically implying (given the NMR study of temperature dependence already discussed) that a threaded lassomycin cannot unthread. Neither scenario therefore supports the hypothesis that the naturally isolated peptide has undergone unthreading upon heating to 40 °C. Furthermore, given the general implausibility of scenario ii), taken together with the previously discussed likelihood that synthetic peptide **24** is in fact unthreaded, it is proposed that the previously published naturally isolated peptide exists as a threaded lasso that does not unthread upon heating to 40 °C.

## 3.6 Crystallization screening

Crystallization of synthetic lassomycin (**24**) was also attempted via high-throughput screening of 10 mg/mL peptide solution using a standard protein crystallization buffer set on an Innovadyne™ Screenmaker 96+8 Xtal™ nanolitre scale screening robot. No crystals were obtained for any of the conditions tested, and so a novel sparse matrix screen developed specifically for peptide crystallization[94] was investigated.

Vapour diffusion conditions in sitting-drop format were set up manually using 1 μL of 40 mg/mL peptide solution per well, in addition to either 1 or 2 μL of the buffer solution (1:1 and 2:1). An additional 1:1 tray was also set up to run at 4 °C. The complete buffer set prepared is shown in Table 3.2. Each of the 24 solutions were used for 1:1, 2:1 and 1:1 (4 °C) conditions (72 conditions total).

**Table 3.2:** Buffer set prepared for peptide sparse matrix crystallization screen.[94]

| | *Precipitant 1* | *Salt/precipitant 2* | *Additive* | *Buffer (0.1 M)* | *pH* |
|---|---|---|---|---|---|
| 1 | 2.0 M NaCl | - | - | NaOAc | 4.5 |
| 2 | 20% 2-PrOH | - | 0.01 M $CaCl_2$ | ″ | ″ |
| 3 | 10% (w/v) PEG 3000 | 0.1 M $ZnCl_2$ | - | ″ | ″ |
| 4 | 30% (w/v) PEG 4000 | - | - | ″ | ″ |
| 5 | 0.8 M $(NH_4)_2SO_4$ | - | - | Na citrate | 5.0 |
| 6 | 30% (v/v) MPD | - | - | ″ | ″ |
| 7 | 30% (v/v) *t*-BuOH | - | - | ″ | 5.5 |
| 8 | 20% (w/v) PEG 4000 | 20% (v/v) 2-PrOH | - | ″ | ″ |
| 9 | 20% (w/v) PEG 3350 | - | - | ″ | ″ |
| 10 | 20% (w/v) PEG 8000 | 0.2 M $ZnCl_2$ | - | MES | 6.0 |
| 11 | 30% (v/v) Jeffamine® M-600 | - | - | ″ | ″ |
| 12 | 30% (v/v) MPD | 0.2 M $Mg(OAc)_2$ | - | ″ | 6.5 |
| 13 | 20% (w/v) PEG 3000 | 0.2 M $MgCl_2$ | - | ″ | ″ |
| 14 | 0.1 M $K_2HPO_4$ | - | - | HEPES | 7.0 |
| 15 | 1.0 M NaOAc | 0.1 M $ZnSO_4$ | - | ″ | ″ |
| 16 | 20% 2-PrOH | 0.1 M NaCl | 20% (v/v) glycerol | ″ | 7.5 |
| 17 | 25% (w/v) PEG 3350 | 10% (v/v) 2-PrOH | - | ″ | ″ |
| 18 | 30% (v/v) Jeffamine® M-600 | - | - | ″ | ″ |
| 19 | 45% (v/v) MPD | 0.2 M $NH_4OAc$ | - | ″ | ″ |
| 20 | 20% (v/v) MPD | - | 0.05 M NaCl | Tris | 8.0 |
| 21 | 10% (w/v) PEG 8000 | - | - | ″ | ″ |
| 22 | 25% (w/v) PEG 3350 | 0.2 M $NH_4OAc$ | - | ″ | 8.5 |
| 23 | 10% 2-PrOH | - | 0.05 M NaCl | ″ | ″ |
| 24 | 70% (v/v) MPD | - | - | ″ | ″ |

Abbreviations: HEPES - 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid, MES - 2-(*N*-morpholino)ethanesulfonic acid, MPD - 2-methyl-2,4-pentanediol, PEG - poly(ethylene glycol), Tris - 2-amino-2-(hydroxymethyl)-1,3-propanediol.

Peptide aggregates were seen in a number of wells, but no crystalline material was observed after 1 week. The trays will be checked at monthly intervals, and alternative protein crystallization screens will be considered. Truncated peptide **36** or analogues from alanine scanning may also potentially be more amenable to crystallization.

## 3.7 Conclusions and future work

A practical synthetic route has been developed for the synthesis of an unthreaded version of the lasso peptide lassomycin (**24**). Lassomycin(1-9) (**36**), in addition to lassomycin-amide (**35**), a C-terminal amide derivative of **24**, were also prepared. A lack of biological activity for the synthetic peptides against *Mycobacterium tuberculosis* suggests the structure of the synthetic peptide does not match that of the previously published compound.[74] Given that the correct sequence was confirmed by MS analysis, it is assumed that the peptides differ only in their threading state. Furthermore, IM data suggested that the synthetic peptide exists only in a single conformation, and not in a mixture of threaded and unthreaded states.

Due to threading of the peptide during chemical synthesis being unlikely, it is suggested that the previously published natural product in fact exists in a threaded conformation, and that the previously published unthreaded NMR structure is incorrect. In order to rule out the possibility of unthreading of the natural product during the NMR data collection previously reported, NMR characterization was carried out on the synthetic peptide. The results reported suggest that for lassomycin there is no interconversion between threaded and unthreaded states upon heating to 40 °C. Furthermore, clear differences between the NMR data presented and the previously published data support the hypothesis that synthetic and naturally isolated lassomycin adopt different conformations.

Together these results strongly suggest that the previously published, naturally isolated lassomycin peptide exists in the threaded lasso conformation observed for all other lasso peptides discovered to date. Crystallization assays are ongoing, and an X-ray crystal structure will provide definitive evidence for the lack of threading in the synthetic peptide.

Threaded lasso peptides remain out of the reach of chemical synthesis techniques. Given the conclusion that lassomycin is likely to be expressed

as a threaded lasso, the lack of cysteine residues in its sequence suggests that stabilization of the threaded topoisomer may be due to a 'steric lock' formed from the bulky arginine side chains present, as opposed to disulfide bonding.[65] It is possible therefore that threaded conformations may be accessed via replacement of Arg side chains with less sterically hindered residues such as Ala or Gly, and that equilibrium populations of threaded and unthreaded forms could be isolated through covalent 'trapping' via cyclization or disulfide bond formation. A potential strategy could involve introduction of Acm-protected Cys residues during SPPS, which would be deprotected following purification and solution-phase cyclization. Threading dynamics could then feasibly be observed in aqueous solution via disulfide trapping of threaded and unthreaded topoisomers, which may be subsequently separated and characterized by ion-mobility MS (Figure 3.8). Such an approach would facilitate elucidation of the effect of peptide sequence and structural modifications upon the threading, stability and biological activity of lasso peptides.

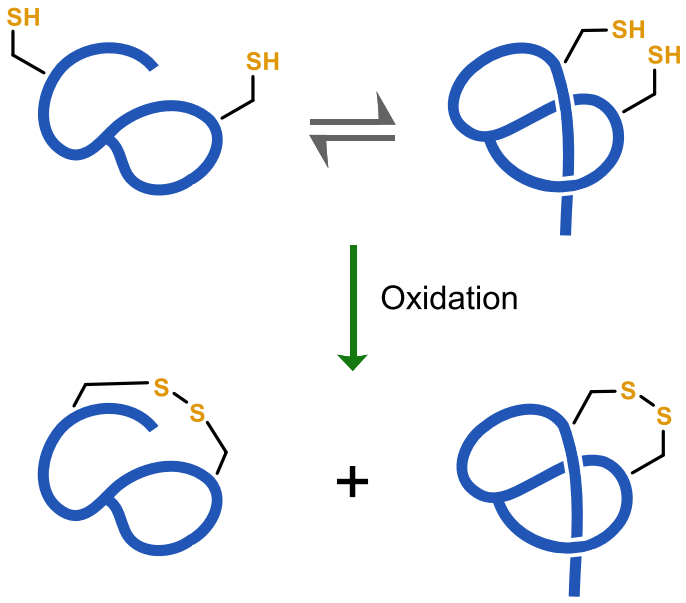



**Figure 3.8:** Suggested strategy for the isolation of chemically synthesized threaded lasso peptides. Equilibrium populations of threaded and unthreaded conformations could be 'trapped' via disulfide formation and subsequently purified and characterized.

# Part II

# Total chemical synthesis of whole proteins

# Chapter 4

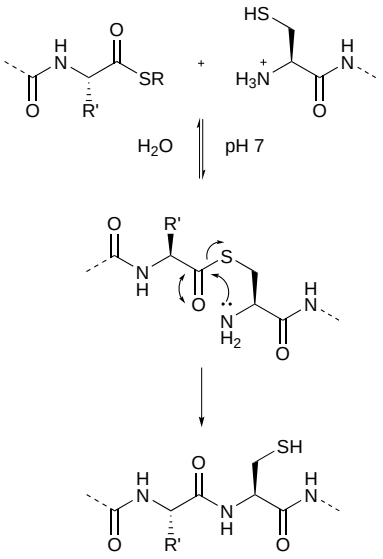
# Current progress in whole protein chemical synthesis

The applicability of chemical synthesis to the investigation of the structure and function of small bioactive peptide natural products has been demonstrated thus far in Chapters 2 and 3. Chemical synthesis can also be invaluable in elucidating the role of much larger biomolecules, such as proteins. The following chapter introduces the reaction central to the field of whole protein chemical synthesis – native chemical ligation (NCL) – which enables the application of peptide synthesis techniques to targets which which would otherwise be inaccessible due to their size. A selection of illustrative examples are also provided, representing the state of the art in the field.

## 4.1 Introduction

Peptide chemistry has changed immeasurably since Emil Fischer's synthesis of glycyl-glycine in 1901.[95] Bruce Merrifield's pioneering work using Boc chemistry and a solid support enabled peptide synthesis to become a routine technique,[96] and stepwise Fmoc SPPS has since been used to produce peptides[97] and even whole proteins[98–100] of increasing length. Nevertheless, the purification of longer synthetic peptides quickly becomes problematic due to the increasing proportion of deletion by-products resulting from incomplete couplings.[101,102] Instead the target peptide can be synthesized in smaller sections that are subsequently ligated in solution, thus circumventing issues

associated with the SPPS of longer targets. Native chemical ligation (NCL) is a well-established technique involving the reaction between a peptide bearing an N-terminal cysteine and a second bearing a C-terminal thioester group (Scheme 4.1). The reaction can be carried out on unprotected peptides in aqueous buffer, and the technique has enabled the total chemical synthesis of countless whole proteins inaccessible by SPPS alone.[103–109] The various methods available that can be used to produce the peptide thioesters required for NCL will be discussed at the end of this chapter.




**Scheme 4.1:** General mechanism for native chemical ligation (NCL).[104]

An area of particular interest in the field is the chemical synthesis of proteins bearing post-translational modifications (PTMs).[110] While PTMs often play a crucial role in the biological processes associated with a particular protein's function, obtaining homogeneously modified protein bearing one or more specific PTMs at one or more specific sites is often impossible by conventional means. A post-translationally modified protein isolated from a biological source may comprise multiple forms, each bearing different PTMs at a number of possible sites, hence rendering the purification of a given form in order to study the effect of a specific PTM extremely difficult. For example, despite the area of therapeutic glycoproteins being the fastest

growing field in the pharmaceutical industry,[111] glycosylated proteins cannot be recombinantly expressed routinely in bacterial systems such as *E. coli*,[112,113] and the variability of glycosylation sites for recombinant protein can render the successful purification of a desired glycoform impossible.[111,114,115]

By contrast, a host of covalent modifications can be incorporated site-specifically into a target protein during a total chemical synthesis, including labels such as fluorophores[116] or NMR probes,[117] or PTMs such as phosphorylations[118] or glycosylations.[119] Indeed any natural or unnatural amino acid can be inserted at the desired point during Fmoc SPPS, given an appropriately protected building block is available or can be chemically synthesized.[120] An understanding of the relationship between PTM occurrence and function is crucial for a given protein of interest, and total chemical synthesis affords unparalleled control over sequence and covalent modification, allowing access to site-specifically modified synthetic proteins.

A selection of recent examples of whole protein total chemical synthesis is provided in the following sections, representing the state of the art in the field. The focus of the chapter will be those targets bearing important PTMs and for which obtaining site-specifically modified protein has proved extremely difficult or impossible by conventional means.

## 4.2 Obtaining erythropoietin as a single glycoform

The glycoprotein hormone erythropoietin (EPO) is secreted by the kidneys during decreased oxygen circulation and is responsible for triggering erythropoiesis – the production of red blood cells (erythrocytes).[121] Despite the primary sequence of EPO and its glycosylation sites being highly conserved, the oligosaccharides linked at asparagines 24, 38, and 83 are not and exhibit considerable structural variation, resulting in a family of inseparable EPO glycoforms.[122] While a number of semisynthetic homogeneous EPO glycoforms have been reported,[123] including those by Flitsch (Macmillan et al., 2001)[124] and Kajihara (Hirano et al., 2009),[125] the first total chemical synthesis of a synthetic EPO analogue was published by Kent in 2003 (Kochendoerfer et al.).[126] While this analogue was unglycosylated, PEG-derived branched polymer side chains were introduced at two of EPO's four native glycosylation sites, resulting in enhanced biological activity and lifetime *in vivo*. A second fully synthetic EPO analogue was reported by Kent in 2012 (Liu et al.), where

Asn residues at three of the native glycosylation sites were replaced with Lys, retaining biological activity despite the lack of glycosylation.[127]

Kajihara (Murakami et al.) reported a fully synthetic glycosylated EPO analogue in 2012, bearing a complex oligosaccharide at Asn-83.[128] However, the first total chemical synthesis of EPO bearing oligosaccharides at all four native glycosylation sites (**40**) was completed by Danishefsky in 2012 (Wang et al.).[121] The approach used is summarized in Figure 4.1a. The protein was assembled in four fragments (I-IV) using using NCL. Alanines 79 and 125 were replaced with Cys residues allowing sequential ligations to be carried out at those points, with subsequent conversion back to Ala using metal-free dethiylation (MFD). MFD is carried out using the reducing agent tris(2-carboxyethyl)phosphine (TCEP) in the presence of the radical initiator 2,2′-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044).[129,130] Acm protecting groups were used for those Cys residue side chains that were to remain in the final protein.

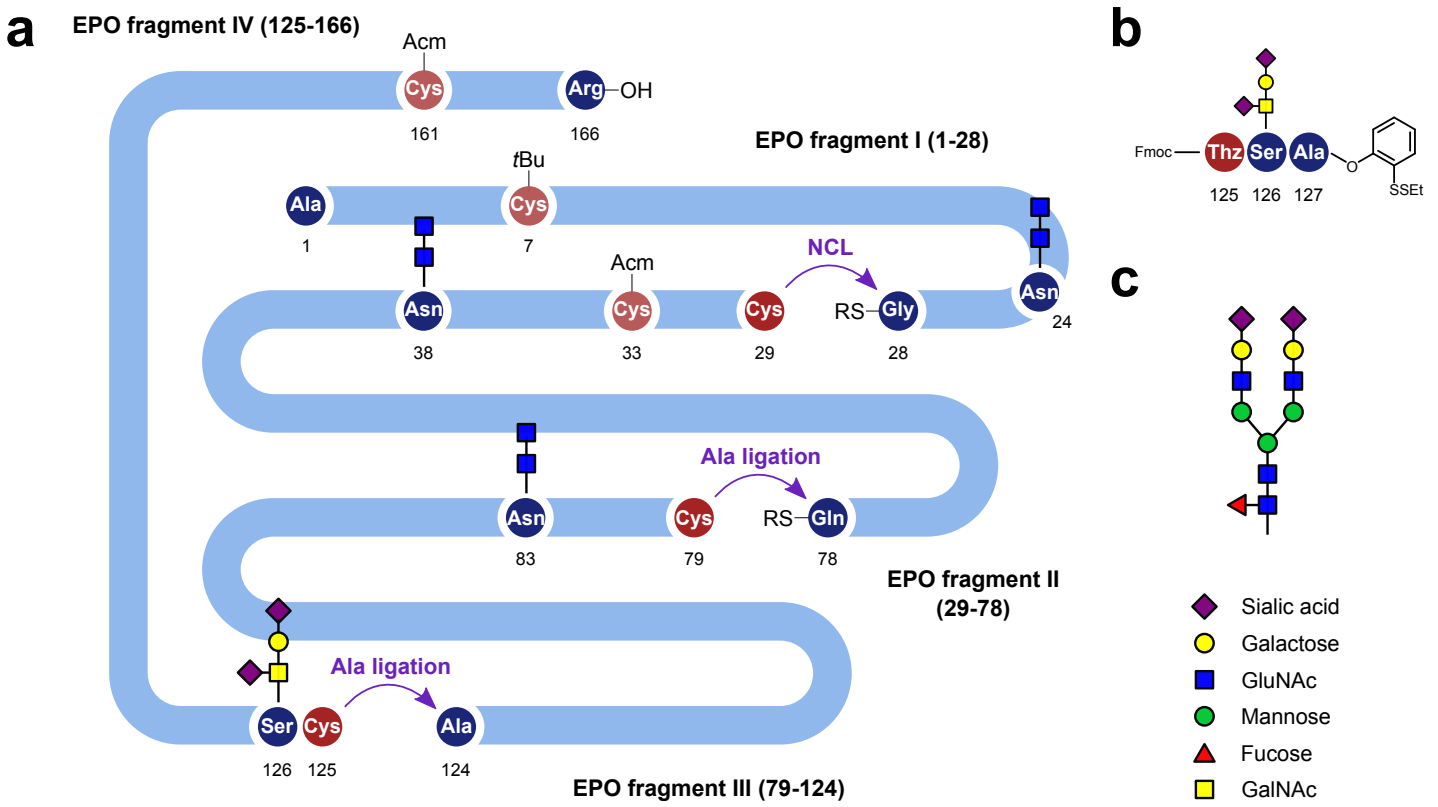



**Figure 4.1:** (a) Chemical synthesis of a single glycoform of erythropoietin (EPO, **40**). (b) Tripeptide incorporating fully deprotected glycosyl-amino acid cassette, which was ligated onto the N-terminus of fragment IV (see text). The Thz residue was subsequently unmasked to give the required Cys residue. (c) Example oligosaccharide structure occurring at asparagine sites 24, 38 and 83 in the naturally isolated protein[121] (glycan symbol nomenclature shown[131,132]). Adapted from Wang et al.[121]

The oligosaccharide required at position Ser-126 was first incorporated as a fully deprotected glycosyl-amino acid cassette into a tripeptide bearing the masked C-terminal thioester equivalent previously reported by Chen et al.,[133] while the required N-terminal cysteine was masked as a thiazolidine-4-carboxylic acid (Thz) residue (Figure 4.1b). This tripeptide was then introduced onto the N-terminus of fragment IV via NCL, with subsequent deprotection of the Fmoc group and treatment with methoxyamine to unmask the thiazolidine group. Simplified disaccharide units were coupled at the required Asn positions using the previously reported one-flask aspartylation/deprotection protocol, which was carried out on the Asp-containing thioester fragments.[134] An example of the type of carbohydrate structure that would appear at these sites in the naturally isolated protein is shown in Figure 4.1c. An explanation of the standard glycan symbol nomenclature is also given.[131,132]

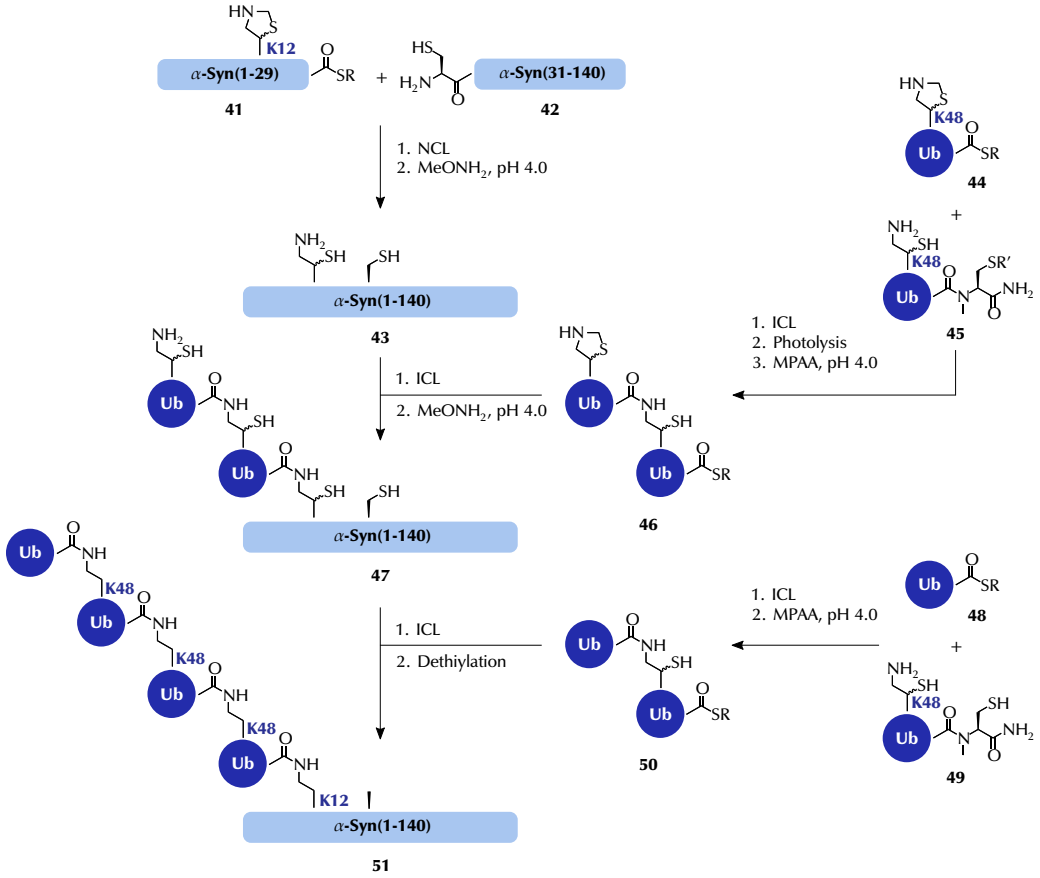
Biological activity of **40** and unglycosylated EPO was evaluated by observing erythroid colony formation in a hematopoietic stem and progenitor cell differentiation assay. Glycosylated **40** exhibited greater activity than the unglycosylated protein, potentially due to the poor stability of the latter. Furthermore, the unglycosylated form 'did not give rise to a supportive mass spectrum'.[121] These results demonstrate the importance of glycosylation in maintaining glycoprotein stability, and thus presumably biological activity also.

## 4.3 Elucidating the role of polyubiquitination

Ubiquitination is a PTM most well known as a signal for proteasomal degradation, but is also involved in many and varied cellular processes and has been implicated in the pathogenesis of several neurodegenerative diseases.[135,136] The modification involves either the attachment of a single molecule of the 76 amino acid protein ubiquitin (Ub) to a protein target (monoubiquitination), or a chain of Ub molecules (polyubiquitination). The latter is greatly complicated by the fact that linkage of consecutive Ub molecules in the chain can occur at any of the seven Lys residues present in the Ub protein.[137] Therefore rigorous study of the effect of ubiquitination on the function of a given protein requires access to homogeneously ubiquitinated analogues of the protein in question.

$\alpha$-Synuclein ($\alpha$-Syn) is a 140-residue intrinsically disordered protein that plays a central role in Parkinson's disease and several other neurodegenerative

disorders.[136] Ubiquitination is hypothesized to be involved in regulating the formation of fibrillar $\alpha$-Syn aggregates, which occurs in diseased brain tissue. In 2013 Haj-Yahya et al. reported the semisynthesis of site-specifically mono-, di-, and tetraubiquitinated forms of $\alpha$-Syn.[136] A summary of the approach is given in Scheme 4.2.




**Scheme 4.2:** Semisynthesis of di- and tetraubiquitinated $\alpha$-Syn from recombinant fragment [A30C]$\alpha$-Syn(30-140) (**42**), synthetic fragment $\alpha$-Syn(1-29) thioester incorporating a thiazolidine-protected $\delta$-mercaptolysine residue (**41**), and two synthetic diubiquitin thioester building blocks (**46** and **50**). R groups: R = $SCH_2CH_2COOH$, $SC_6H_4CH_2COOH$ or another unspecified thioester; R′ = 2-nitrobenzyl. Adapted from Haj-Yahya et al.[136]

Recombinantly expressed fragment [A30C]$\alpha$-Syn(30-140) (**42**) was first ligated with synthetic thioester **41**, incorporating the thiazolidine-protected form of $\delta$-mercaptolysine at the Lys-12 position. Deprotection of the thiazolidine group (using methoxyamine) yielded the full $\alpha$-Syn sequence (**43**). Ubiquitin monomers were produced by total chemical synthesis, which enabled the necessary site-specific attachment at defined Lys residues within the Ub sequence. *N*-Methylcysteine-mediated latent thioester formation[138] was proposed to produce the necessary Ub-thioesters for attachment to $\alpha$-Syn, although attempts to produce the entire tetraubiquitin molecule were found to be problematic due to difficulties in the final thioester formation step.

Instead two separate diubiquitin fragments were synthesized (**46** and **50**) and ligated sequentially with $\alpha$-Syn. Monoubiquitin thioesters **44** and **48** underwent isopeptide chemical ligation (ICL) with *N*-methylcysteine 'latent' thioesters **45** and **49**, respectively, to form diubiquitin units linked at the side chain of Lys-48. Diubiquitin thioesters **46** and **50** were obtained following an N→S acyl shift in the presence of 4-mercaptophenylacetic acid (MPAA) (photolysis of the 2-nitrobenzyl protecting group was required first for **45**).

Sequential ligation of **46** (followed by thiazolidine deprotection) and **50** to the $\delta$-mercaptolysine side chain of **43** was then carried out. A final MFD step was used to convert Cys to Ala in the $\alpha$-Syn sequence, and all $\delta$-mercaptolysine residues in the polyubiquitin section to Lys, to give site-specifically tetraubiquitinated $\alpha$-synuclein.

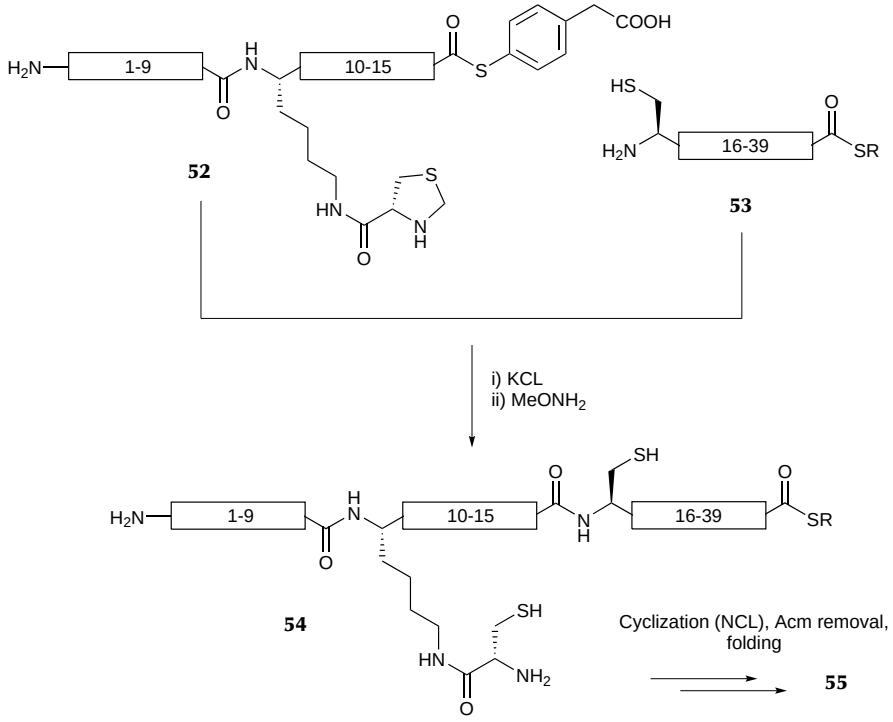
With a route to semisynthetic $\alpha$-Syn allowing full control over the extent and position of ubiquitination in hand, Haj-Yahya et al. were able to characterize the biological properties of mono-, di-, and tetraubiquitinated forms of the protein. *In vitro* fibrillization studies were carried out at low concentration (5 μM) to mimic the estimated concentration in neurons. While wild type $\alpha$-Syn does not form aggregates at low concentration, the tetraubiquitinated protein was found to form SDS-resistant high-molecular weight species at 5 μM *in vitro*. Transmission electron microscopy was used to confirm that the wild type protein only formed mature amyloid fibrils after 8 days, whereas tetraubiquitinated $\alpha$-Syn formed electron-dense aggregates at 84 h, not possessing an amyloid-like structure. Haj-Yahya et al. suggest therefore that tetraubiquitination at this position results in the formation of nonfibrillar aggregates but blocks $\alpha$-Syn fibrillization *in vitro*.

It was also demonstrated that phosphorylation at Tyr-125 of $\alpha$-Syn destabilized both di- and tetraubiquitinated forms, causing a tendency to precipitate (unlike the wild type), suggesting the possibility for crosstalk between these two PTMs. In addition di- and tetraubiquitinated $\alpha$-Syn were shown to undergo efficient proteasomal degradation, whereas the single ubiquitin moiety is efficiently cleaved by deubiquitinases for the monoubiquitinated form (thus preventing degradation). It was therefore suggested by Haj-Yahya et al. that polyubiquitination serves as a signal for $\alpha$-Syn degradation by the proteasome. The clearly demonstrated role of ubiquitination in the regulation of $\alpha$-Syn stability, aggregation, phosphorylation and clearance therefore suggests that the function of $\alpha$-Syn in disease may be under the control of complex regulatory mechanisms involving crosstalk among different PTMs.[136]

## 4.4 Exploring novel protein chain topologies

The small, disulfide-rich protein crambin was previously chemically synthesized by Bang et al. in 2004 using NCL.[108,109] Subsequent synthetic work by the Kent group, in which the carboxylate of the C-terminal Asn-46 residue was replaced with an amide, demonstrated the importance of the salt bridge between Arg-10 and the C-terminus for correct folding and disulfide bond formation.[139] It was proposed that replacing the salt bridge by a covalent linkage would result in a protein bearing an unusual topology lacking a C-terminus.[140] Furthermore the structure would be of particular interest due to the threading of the linear N-terminal section through the covalent macrolactam ring, resulting in a topology reminiscent of the lasso peptide fold (see Chapter 3). A chemical synthesis was carried out incorporating Lys at position 10 in place of Arg, with an amide bond installed between the Lys side chain and the C-terminus. The approach reported is summarized in Scheme 4.3.

Fragments **52** and **53** were used in a 'kinetically controlled ligation' (KCL) whereby the N-terminal Cys of **53** reacts more rapidly with C-terminal peptide aryl thioester **52** than the alkyl thioester C-terminus of **53**. KCL reactions must necessarily be carried out in the *absence* of exogenous thiol, as the presence of a catalyst such as MPAA would cause interconversion to the more reactive aryl thioester. The more rapid reaction of the Ala thioester (at position 15)




**Scheme 4.3:** Synthesis of topological analogue of crambin (**55**). Intermediate **54** was assembled from fragments **52** and **53** via a kinetically controlled ligation (KCL, see text), followed by conversion of Thz to Cys. Subsequent cyclization, Acm protecting group removal and folding gave topological analogue **55** (shown in Figure 4.2). Fragments were synthesized using *in situ* neutralization Boc SPPS. R = $CH_2CH_2CO(Arg)_4$–Ala–COOH. Adapted from Mandal et al.[140]

than the hindered Thr-39 thioester was also exploited in this case. The Thz residue was converted to Cys in the same pot following the ligation reaction, using methoxyamine at pH 4.0. The reaction mixture was diluted and the pH adjusted to 6.8 before carrying out the intramolecular NCL reaction to form the cyclic product. Acm protecting groups were used for side chains of Cys residues not at ligation sites (not shown), and these were removed before HPLC purification, folding and oxidation in a redox buffer to yield **55**.

An X-ray crystal structure of the synthetic protein is shown in Figure 4.2. While crystallization trials with **55** over a period of several months were not initially successful, microcrystals were obtained from a quasi-racemate[l] of **55** and synthetic D-crambin after 4 days in 40-50% of the conditions tested. Topologue **55** was found to share a similar fold with native crambin, with correct disulfide bond formation. It can be seen however that introduction of the amide bond between the C-terminus and Lys-10 results in a folded structure where the N-terminus is threaded through a macrolactam ring and is held in place by two disulfide bonds (as shown). The structure presented by Mandal et al. therefore represents a previously undiscovered protein topology made accessible by chemical synthesis.

## 4.5 Methods used for the synthesis of peptide thioesters

Despite the applicability of native chemical ligation to whole protein chemical synthesis (as illustrated by the examples given), a major drawback is the difficulty in producing peptide thioesters using Fmoc SPPS. While multiple strategies exist,[142] no single method has emerged as the dominant technique. Direct Fmoc-based solid-phase synthesis of peptide C-terminal thioesters can be achieved using a thioester linkage to the solid support, although this linkage is susceptible to cleavage by piperidine during the Fmoc deprotection step.[142] While this can be avoided using less nucleophilic bases for Fmoc deprotection (such as hexamethyleneimine or DBU[143,144]), linker aminolysis still occurs during early synthesis cycles.[142]

A number of auxiliary-mediated approaches exist for the synthesis of thioesters via intramolecular N→S or O→S acyl shift in solution following

[l]For further reading on racemic protein crystallography see the review by Yeates and Kent (2012).[141]

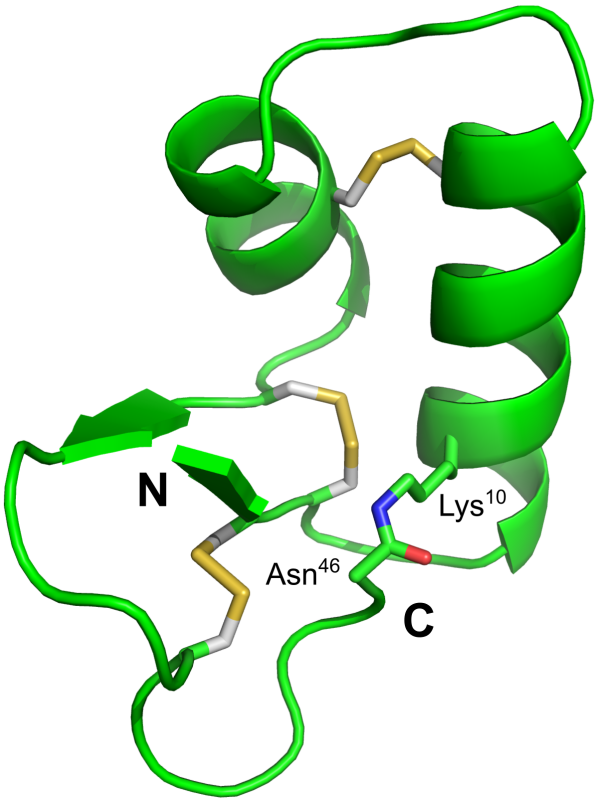



**Figure 4.2:** Crystal structure of chemically synthesized topological analogue of crambin (**55**). The salt bridge present in the wild type between Arg-10 and the C-terminal carboxylate group is instead replaced by a covalent linkage. Disulfide bridges are shown in yellow. PDB ID: 3UE7.[36,140]

cleavage from the resin.[142] Various auxiliaries have been reported for this purpose, including modified thiol-containing amino acids and thiol-functionalized aromatic backbone 'protecting' groups. Indeed, in some cases so-called 'reverse' NCL can occur for peptides bearing internal Cys residues, which can rearrange to the corresponding thioester fragment upon treatment with MPAA.[145] Burlina et al. reported that methylation of cysteine at the alpha carbon allowed its use as a thioester surrogate when sited at the C-terminus of a peptide.[146]

The most frequently utilized methods for thioester generation involve synthesis on the solid phase using specially adapted linkers, such as the 'sulfonamide safety-catch'.[147–149] Safety-catch linkers are designed so as not to be susceptible to attack unless activated by chemical modification, thus remaining untouched during peptide synthesis cycles. The sulfonamide safety catch and another safety-catch linker – the Dawson Dbz linker[150] – are shown in Scheme 4.4. More recently introduced strategies for the synthesis of thioesters or thioester surrogates are also described below.

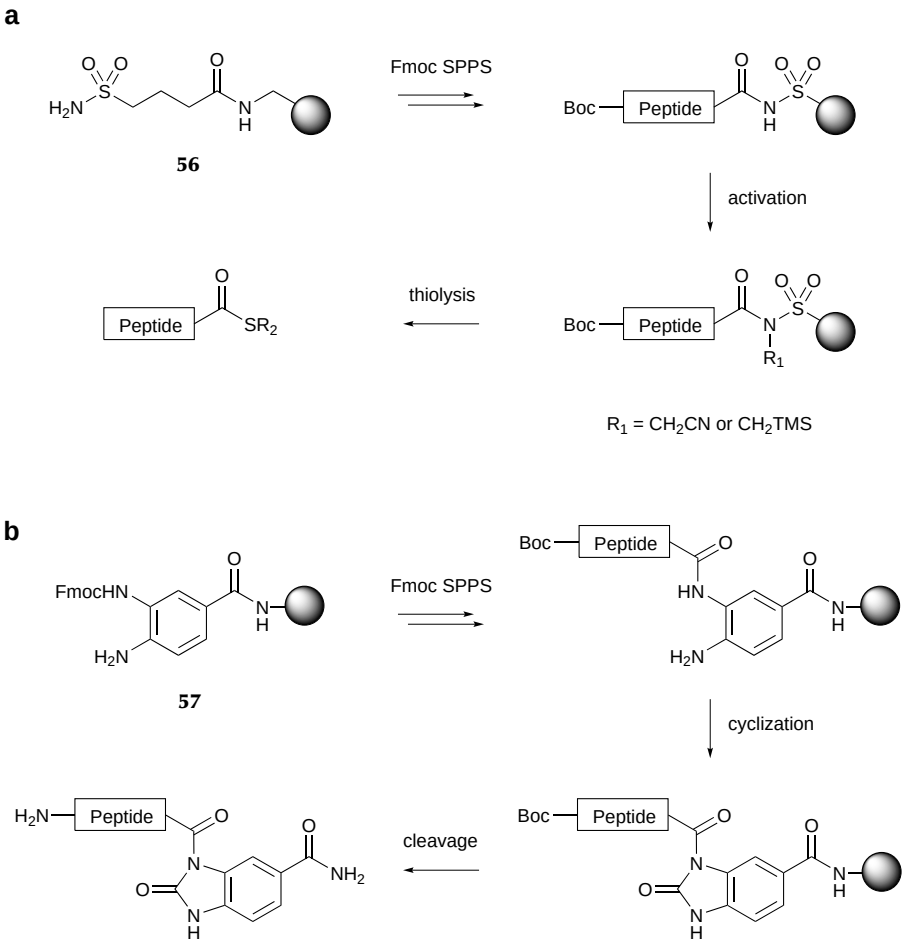
### 4.5.1 The sulfonamide safety-catch

The sulfonamide safety-catch is a linker that can be used for Fmoc SPPS. After completion of the synthesis the linker nitrogen is alkylated (using either iodoacetonitrile or trimethylsilyldiazomethane), and the desired peptide thioester is released via attack with an appropriate thiol (Scheme 4.4). A drawback of this method is that the final thiolysis step can be very low yielding (as demonstrated by previous work within the group[115]), although the use of 2 M LiBr/THF as the solvent during this step was found to give a significant yield increase by Quaderer et al.[151]

The sulfonamide linker can also be attached to the support via an acid-labile linkage (such as Rink amide) to facilitate synthesis monitoring by small-scale TFA test cleavages. This approach also allows direct cleavage of the *N*-alkylsulfonamide (without thiolysis) which can be used in ligations directly (see Chapter 5).[152]

### 4.5.2 The Dawson Dbz linker

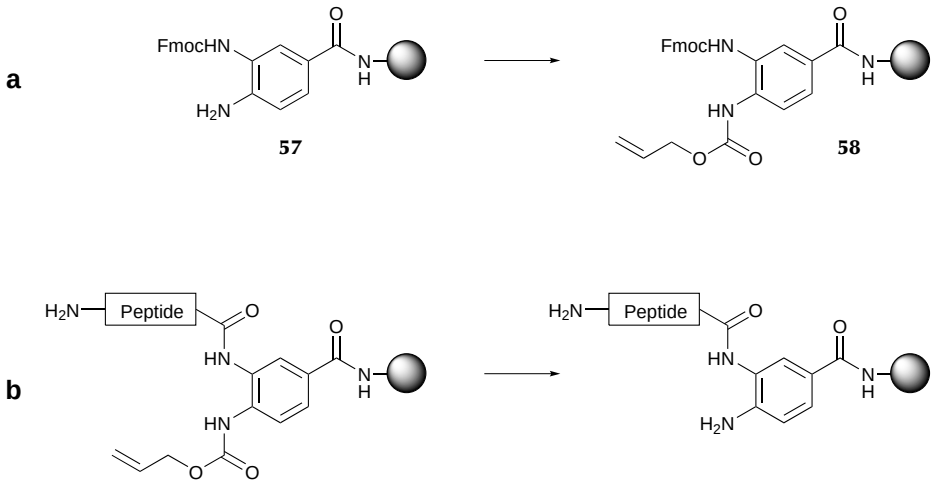
After Fmoc SPPS the Dawson Dbz linker is cyclized to the reactive *N*-acylurea species using *p*-nitrophenylchloroformate, and the peptide is released by




**Scheme 4.4:** SPPS linkers used for the synthesis of peptide thioesters or *N*-acylureas.[142] (a) The sulfonamide safety-catch method involves synthesizing the peptide on sulfamylbutyryl resin (**56**), alkylating the linker (using $ICH_2CN$ or TMS-$CHN_2$) and releasing the peptide thioester from the activated linker via attack with the desired thiol.[142] (b) The Dawson Dbz linker (**57**) is used to produce peptide *N*-acylureas that can either be used directly in ligations or converted to thioesters in solution. The products are released after SPPS by treatment with *p*-nitrophenylchloroformate and cleavage of the Rink linker using TFA (see text).[150]

treatment with TFA (Scheme 4.4). The resulting peptide *N*-acylurea can be used directly in ligations or converted to the thioester in solution via treatment with thiol.[150] While steric and/or electronic factors normally prevent reaction at the second linker amine, coupling can nevertheless occur – particularly with sequences rich in sterically unhindered amino acids such as glycine – resulting in so-called 'overacylation' and the formation of undesired branched by-products.[150]

Previous work within the group (S. Lear, MSc thesis 2013)[115] showed that MW heating promoted overacylation of the Dbz linker, even for sequences without glycine and containing relatively hindered amino acids. To circumvent this problem an Alloc-protected Dbz linker[77] was used (Scheme 4.5), which was found to be effective. The use of a Di-Fmoc-Dbz linker for MW-SPPS was recently published by Gunasekera et al.[153]



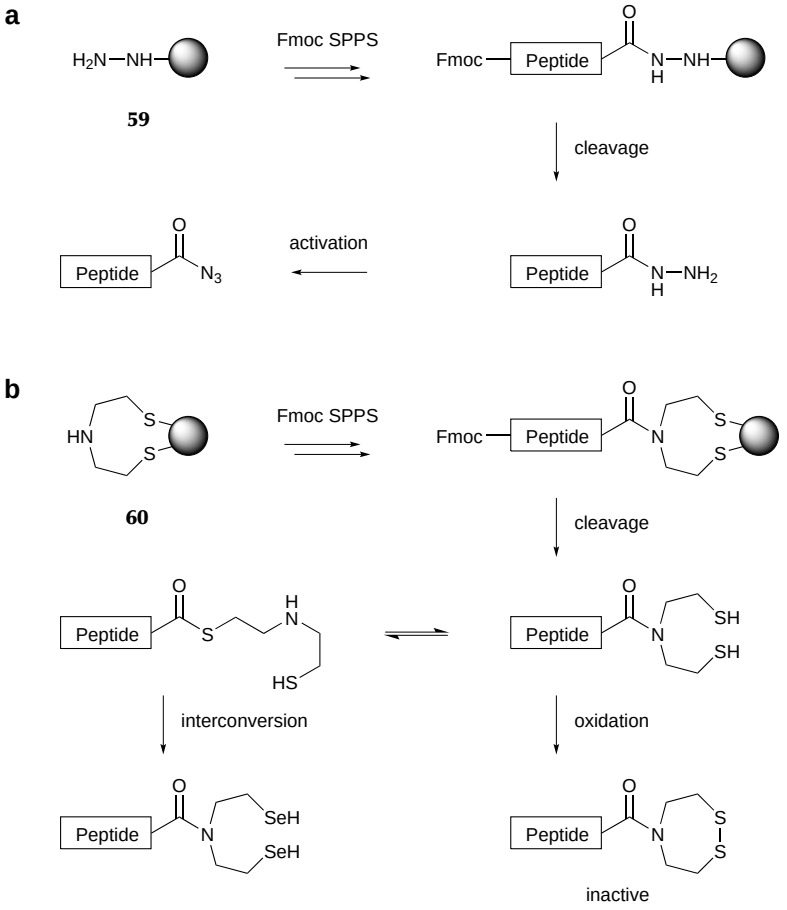

**Scheme 4.5:** Reaction onto the second Dbz linker amine during SPPS (overacylation) can be prevented using an Alloc-protected linker (**58**).[77] (a) Alloc protection is accomplished by treatment with allyl chloroformate. (b) Following SPPS, the Alloc group is removed using Pd(0) (see Chapter 3).

### 4.5.3 Peptide hydrazides

More recently, peptide hydrazides have been found to act as suitable thioester surrogates for NCL.[154] Peptide hydrazides can be synthesized directly on a solid support using the linker shown in Scheme 4.6. Cleavage from the resin and global deprotection using TFA occurs as normal to release the peptide. The activation step involves $NaNO_2$ oxidation to the azide, with subsequent *in situ* conversion to the thioester using MPAA in a one-pot fashion.[155] As peptide hydrazides are unreactive until unmasked in this manner, they can be utilized for convergent assembly of large peptides and proteins using multiple fragment approaches.[156] The generation of peptide hydrazides via direct hydrazinolysis of peptides attached to TentaGel® Wang resin has also been reported.[157]

### 4.5.4 SEA and SeEA peptides

Another recently reported highly versatile approach is the use of bis(2-sulfanylethyl)amido (SEA) peptides as thioester surrogates.[158] SEA peptides can be synthesized on resin using the linker shown in Scheme 4.6 and released using standard TFA cleavage protocols. The C-terminal SEA moiety can undergo an N→S acyl transfer to generate the thioester in water at neutral or mildly acidic pH, which can be used directly in NCL reactions.[159] The selenium bis(2-*selenyl*ethyl)amido (SeEA) analogue (also shown) is significantly more reactive and therefore adds an additional layer of orthogonality, enabling one-pot three fragment assembly under kinetic control (see section 4.4).[160] The corresponding cyclic disulfide (shown) and diselenide forms of these functional groups are unreactive towards Cys peptides, and can therefore be used as masked intermediates in ligation reactions.[162] Furthermore, the diselenide analogue is less easily reduced – the disulfide can be reduced using dithiothreitol (DTT) whereas the diselenide cannot and requires TCEP – which adds yet a further orthogonal aspect the the approach.[161]




**Scheme 4.6:** Recently developed approaches for the synthesis of peptide thioester surrogates. (a) Peptide hydrazides can be synthesized directly on resin and the resulting hydrazides can be used in ligations following activation via $NaNO_2$ oxidation to the azide.[155] (b) SEA peptides can also be produced by Fmoc SPPS and can be used directly in ligation reactions or can undergo thioester exchange.[120,158,159] The corresponding selenium analogue (SeEA) is more reactive and can be used for KCL.[160] The greater redox stability of SeEA over SEA introduces an additional layer of control (see text).[161]

## 4.6 Conclusions and future work

It can be seen that total chemical synthesis is an important tool that can be used to answer important biological questions, not only for small bioactive molecules but for whole proteins also. Furthermore the importance of whole protein total chemical synthesis in addressing challenges that cannot be solved by other means has been illustrated. The native chemical ligation reaction has proven essential to the field, enabling the total chemical synthesis of numerous targets of interest with those bearing precise chemical modifications (such as PTMs) at defined sites being of particular significance. The next chapter describes work towards the total chemical synthesis of acyl carrier protein employing many of the techniques discussed here. The development of a novel linker which can be used to synthesize peptide *N*-acylureas without overacylation is also described.

# Chapter 5

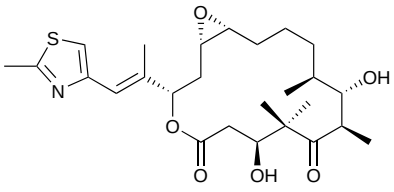
# Total chemical synthesis of *E. coli* acyl carrier protein

Substantial amounts (>50 mg) of a protein of interest are often required in order to characterize its role in a biological system fully. While recombinant expression is conventionally used for this purpose,[163] total *chemical* synthesis can provide a target protein bearing site-specific incorporations of any of an almost endless array of unnatural amino acids, typically in >100 mg quantities. The following chapter details progress made in the total chemical synthesis of acyl carrier protein (ACP), an important component of the cellular machinery responsible for the biosynthesis of fatty acids.[164] Chemical synthesis will provide a route to modified ACPs, which will be used to elucidate the role of protein-protein interactions in fatty acid biosynthesis, while offering insights into the functioning of modular synthase systems in general.

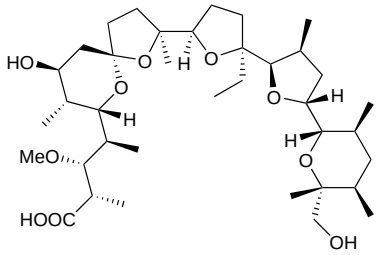
## 5.1 Introduction

Multienzyme megasynthetases are large polypeptides responsible for the biosynthesis of polyketides, nonribosomal peptides and fatty acids in living organisms.[165] The sequential action of independently folded catalytic domains within each multienzyme results in a range of complex and highly diverse secondary metabolites, examples of which are shown in Figure 5.1. A schematic detailing the path of a polyketide through such a multienzyme 'assembly line' is shown in Scheme 5.1. Catalytic domains of various types

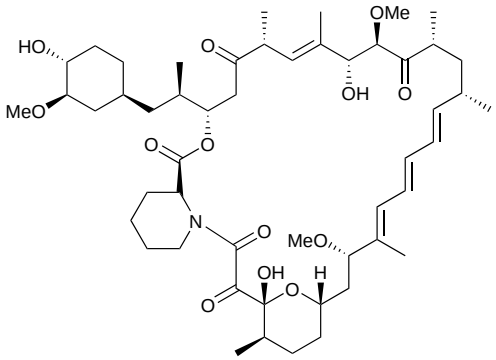
are arranged into *modules*, each acting in turn to elongate the nascent polyketide by a two-carbon unit bearing specific functionality. This colinearity between the sequence of enzymatic transformations and positioning of catalytic domains is such that features of the product structure can in some cases be deduced from inspection of the gene sequence.[165,166]



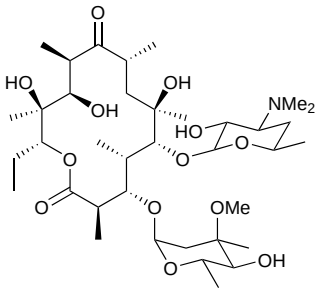
Epothilone A (anticancer)
**61**



Monensin A (antibiotic)
**62**



Rapamycin (immunosuppressant)
**63**

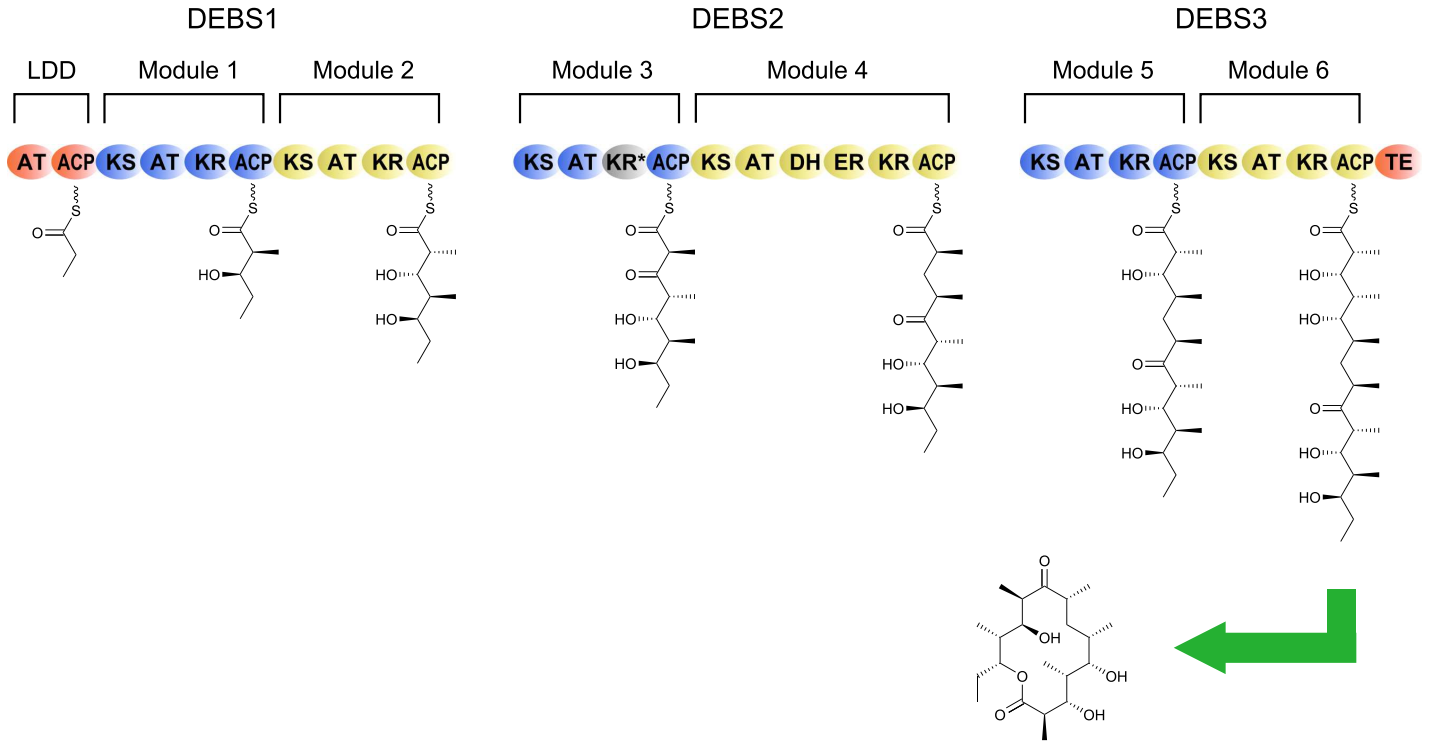


Erythromycin A (antibiotic)
**64**

**Figure 5.1:** Examples of polyketide (PK) natural product structures.[165]

The modularity of megasynthetases makes engineering of chimeric systems an attractive option for the biosynthesis of a diverse range of secondary metabolites. While the combining of polyketide synthase (PKS) domains from different systems has been demonstrated, the resulting PKS turnover rate often suffers due to the perturbation of protein-protein interactions.[167,168] Furthermore, ACP-ketosynthase (KS) domain interactions are weak but specific, ensuring other ACPs coexisting in the same cell do not transfer their

biosynthetic intermediates to undesignated KS partners.[169] This specificity is an issue where hybrid PKSs are concerned, and a detailed understanding of protein-protein interactions within megasynthetases is important for effective engineering of these systems.

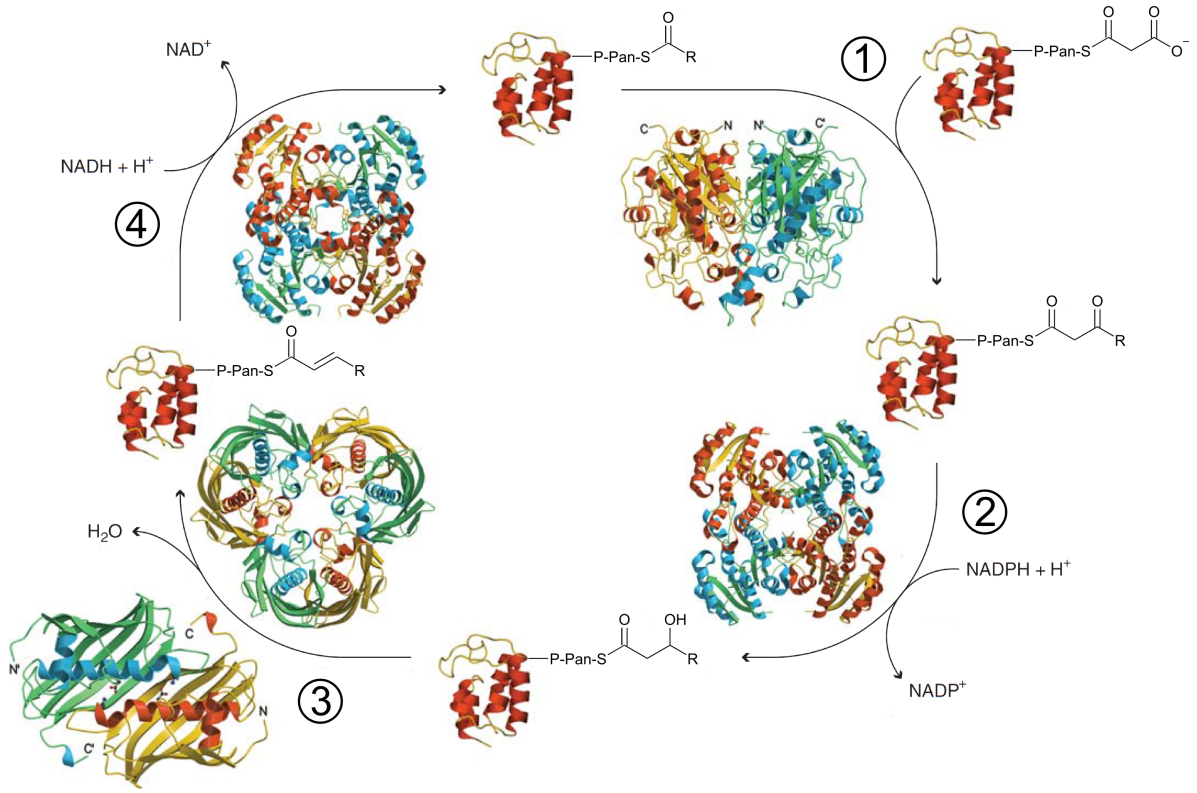



**Scheme 5.1:** 6-Deoxyerythronolide B synthase (DEBS). Chain elongation occurs due to the action of ketosynthase (KS) and acyl transferase (AT) domains, while the final oxidation level of the $\beta$-carbon is controlled by the combination of ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains present in a given module. The nascent polyketide is bound to an acyl carrier protein (ACP) domain at each stage before cyclization and release by the thioesterase (TE) domain at the C-terminus of the polyketide synthase. LDD - loading didomain; KR* - inactive ketoreductase domain. Adapted from Khosla et al.[168]

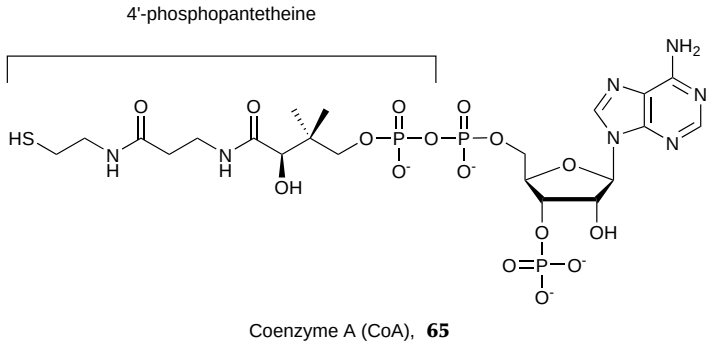
### 5.1.1 Fatty acid biosynthesis

Fatty acid biosynthesis is a vital process in the synthesis of structural lipids for bacterial membranes.[170] Medium-chain (4–12 carbon) fatty acids are valuable precursors to industrial chemicals and biofuels,[171] and therefore reengineering of the fatty acid synthesis (FAS) pathway in order to produce 'tailored' fatty acid metabolites is a desirable goal.[172] While *E. coli* FAS has been used to produce fatty acids, alcohols, esters and alkanes,[171] the significance of biomolecular interactions within megasynthetase systems – of which fatty acid synthase (FAS) is an example – means a complete understanding of these interactions is required. In addition, bacterial FAS is thought to be an attractive target for new antimicrobials.[170]

Scheme 5.2 shows a so-called 'type II' FAS, a dissociated system occurring in bacteria, plants and parasites, in which each protein is encoded by a separate gene.[173] ('Type I' FAS occurs in mammals and consists of a single polypeptide.[173]) The fatty acid product is generated through successive iterations of the FAS cycle, each resulting in elongation of the carbon chain by two atoms. *Acyl carrier protein* (ACP) acts to shuttle fatty acyl intermediates between enzymes in the system, as well as introducing malonate (from malonyl-CoA,[173] not shown) which is condensed with the growing fatty acid acyl-ACP at step ① during each cycle. ACP also bears a 4′-phosphopantetheine prosthetic group, which is transferred from CoA (**65**) to Ser-36 of apo-ACP by ACP synthase (not shown).[173,174] 4′-Phosphopantetheine and related CoA are shown in Figure 5.2.




**Scheme 5.2:** Each enzyme of the Type II fatty acid synthesis (FAS II) pathway is encoded by a separate gene and exists as a soluble protein – the elongation cycle of *E. coli* FAS II is shown.[173] Each cycle is initiated by condensation of the growing fatty acid acyl-ACP with malonyl-ACP, catalysed by $\beta$-ketoacyl-ACP synthase I or II, ①. The product is then reduced by the NADPH-dependent $\beta$-ketoacyl-ACP reductase, ②, and subsequently dehydrated by either $\beta$-hydroxydecanoyl-ACP dehydratase or $\beta$-hydroxyacyl-ACP dehydratase, ③. The final reduction is carried out by the NADH-dependent enoyl-ACP reductase I, ④. P-Pan-SH = 4′-phosphopantetheine (Figure 5.2).




Coenzyme A (CoA), **65**

**Figure 5.2:** 4′-Phosphopantetheine (which is covalently attached to Ser-36 of ACP) forms part of the structure of (and is transferred from) coenzyme A (CoA, **65**).[175]

Enzyme-ACP interactions are of crucial importance in FAS, in addition to the closely related sequestration of the nascent fatty acid chain by ACP and the role of this process in substrate recognition.[176] It is hoped that the chemical synthesis of ACP will allow the generation of tools with which to investigate enzyme-ACP and ACP-substrate interactions in order to gain a broader understanding of FAS, as well as related modular synthase systems.

### 5.1.2 Acyl carrier protein (ACP)

An NMR solution structure of ACP with a bound butyryl acyl intermediate is shown in Figure 5.3. Binding of the substrate alters the conformation of ACP,[174] and it is thought that in this way the nature of the bound substrate may affect the stability of the ACP-enzyme complex, hence conferring substrate specificity.[176]

Zornetzer et al. found that attaching either decanoate (C10:0-ACP) or stearate (C18:0-ACP) to the pantetheine prosthetic group of spinach ACP caused the protein to adopt a single conformer in contrast to the unacylated protein, which interconverts between two major conformers in solution.[176] A macromolecular docking study also suggested that the fatty acid was positioned down the centre of the helix bundle, and it was hypothesized that for longer substrate chain lengths increased solvent exposure may induce

**Figure 5.3:** Ribbon representation of butyryl-ACP (butanoyl group bound to the thiol of the 4′-phosphopantetheine prosthetic group, see text). NMR solution structure, PDB ID: 2K94.[36,174]

greater flexibility in the phosphopantetheine loop, hence facilitating protein-protein interactions with the enzyme.

The flexibility of ACP results in crystallization of the protein being challenging, and the importance of dynamic protein-protein interactions for its function means that protein NMR is often a more appropriate technique for structural characterization of ACP.[164,174,176] $^{19}$F NMR is a powerful technique that can be used for investigating interactions in fluorine-labelled proteins,[117] and total chemical synthesis of ACP will allow the site-specific incorporation of fluorinated amino acids to produce analogues for use in structural $^{19}$F NMR studies.

Perhaps most importantly, chemical synthesis will allow facile incorporation of a variety of amide and ester substrate-bound ACP analogues (Figure 5.4). These can be used to study enzyme-ACP and ACP-substrate interactions and are more resistant to hydrolysis than naturally occurring acyl-ACPs, where the substrate is attached via a *thioester* linkage to the 4′-phosphopantetheine prosthetic group.[177,178]

### 5.1.3 Previous total chemical synthesis

*E. coli* ACP has previously been chemically synthesized via automated Boc SPPS (Hancock et al., 1972).[179–181] The protein was synthesized in a stepwise manner as a single peptide (no ligations used). Cleavage from the resin and deprotection was carried out using HBr/TFA (with catalytic hydrogenation used to remove remaining protecting groups) and the crude protein was purified via gel filtration and ion-exchange chromatography. The resulting apo-ACP was obtained in 42% yield, and the 4′-phosphopantetheine prosthetic group was introduced enzymatically using ACP synthase.

An Fmoc SPPS strategy is preferred however, as repeated exposure of the peptide to TFA during Boc SPPS can result in degradation of the product and is not likely to be compatible with modifications such as ‘non-hydrolysable’ prosthetic group analogues (which can be installed site-specifically during Fmoc SPPS). In addition, the use of NCL allows shorter fragments bearing site-specific incorporations to be synthesized quickly, before attaching the corresponding ligation partner (which can be synthesized at the same time) to rapidly generate multiple analogues in parallel (cf. stepwise approach of the previous synthesis).

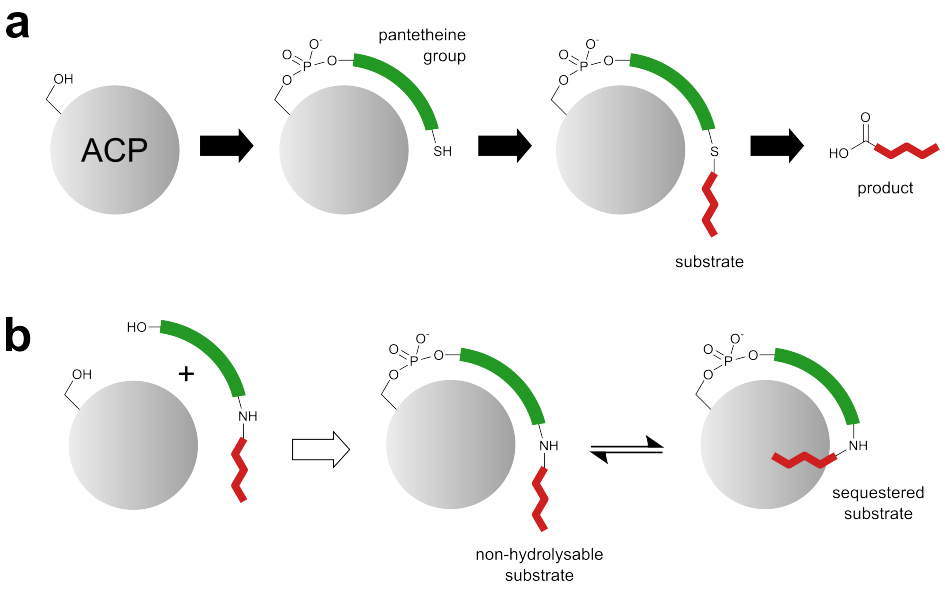
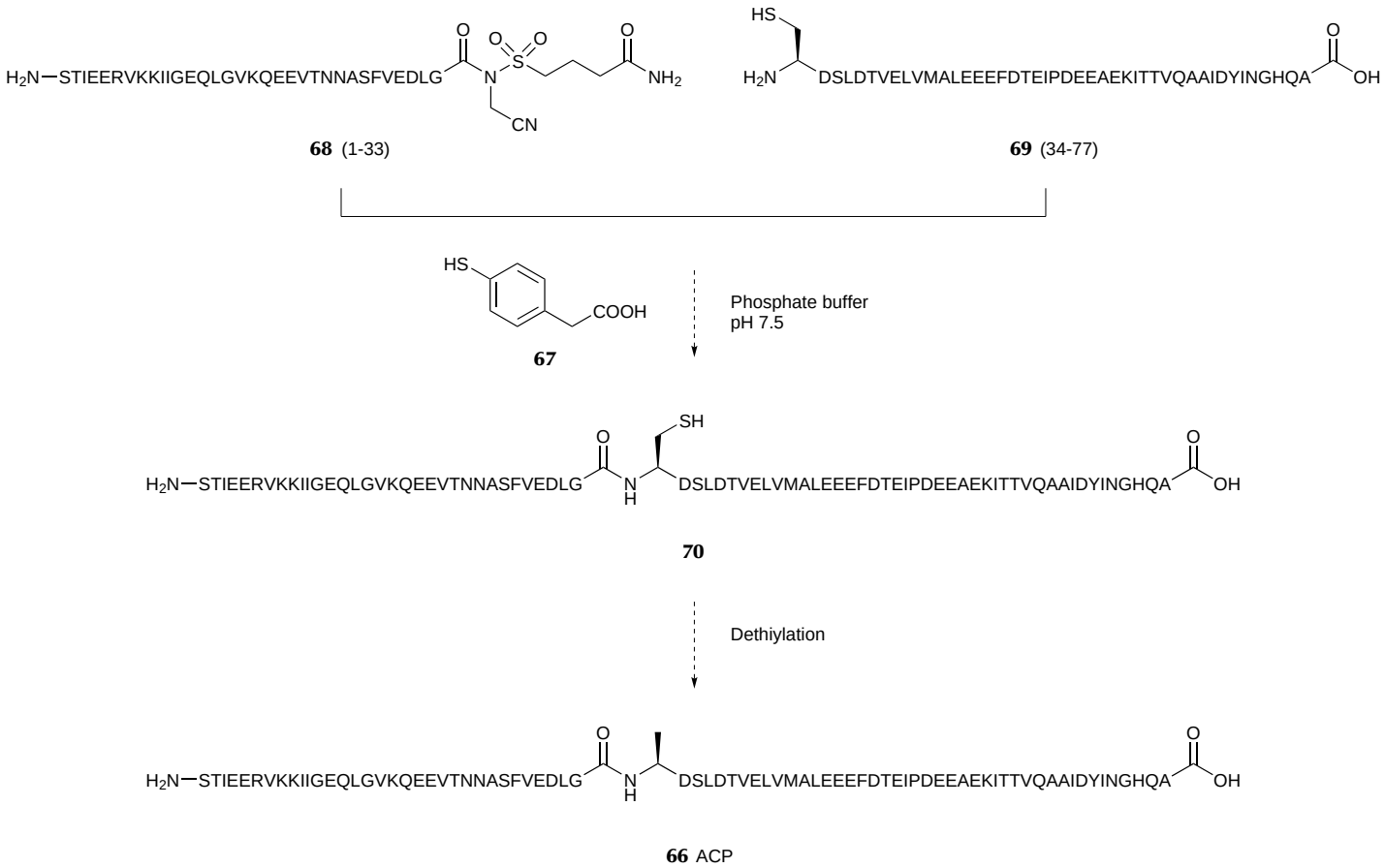



**Figure 5.4:** (a) During FAS the 4′-phosphopantetheine prosthetic group is installed (for full structure see Figure 5.2), and the product is subsequently elongated and released.[177] (b) Substrate sequestration by ACP can be investigated using non-hydrolysable amide substrate analogues, which have previously been enzymatically installed (unfilled arrow).[177] Chemical synthesis will allow access to a broader range of substrate analogues of this type, as modifications such as these can be incorporated relatively easily during Fmoc SPPS.

## 5.2 Synthetic strategy

The proposed strategy for the synthesis of *E. coli* fatty acid apo-ACP (**66**) using Fmoc SPPS is given in Scheme 5.3. The whole protein is formed via native chemical ligation (NCL) of two smaller fragments, one bearing an N-terminal cysteine and the other with an *N*-alkylsulfonamide instead of a carboxyl group at its C-terminus (see Chapter 4). The ligation reaction is carried out in aqueous solution at neutral pH, with the addition of 4-mercaptophenylacetic acid (MPAA, **67**) as a catalyst.[182] Peptide *N*-alkylsulfonamide **68** and acid **69** are synthesized via Fmoc SPPS and fully deprotected before NCL (see below). The resulting product, peptide **70**, contains an undesired Cys residue (shown) which is removed via metal-free dethiylation (MFD, see Chapter 4) to give ACP (**66**).




**Scheme 5.3:** Strategy for total chemical synthesis of *E. coli* fatty acid apo-ACP (**66**) via ligation of *N*-alkylsulfonamide **68** and peptide acid **69**. Fragment **69** contains a Cys residue in place of Ala (which occurs in the final protein), enabling the use of native chemical ligation (NCL) in the presence of 4-mercaptophenylacetic acid (MPAA, **67**). The Cys thiol is removed in the final step to generate the desired Ala residue (see text).

## 5.3 Synthesis of C-terminal section

### 5.3.1 Initial attempts

Automated Fmoc SPPS of [A34C]ACP(34-77) (**69**) was carried out using a CEM Liberty1 peptide synthesizer. Polystyrene Wang resin[l] was used in order to obtain the target peptide with an acid at the C-terminus, as desired. Preloaded resin was used to circumvent problems typically associated with loading, such as racemization.[183] A low substitution (0.33 mmole/g) was chosen in order to minimize steric hindrance and potential aggregation, which can affect synthesis quality for SPPS of long peptides. Double couplings were carried out at room temperature (1 h per reaction) using 5 equivalents of Fmoc amino acid. Deprotection of the Fmoc group was effected via treatment with 20% piperidine twice (5 + 10 min).

Synthesis efficiency was monitored at various points by liberating a sample of the product peptide from a test aliquot of resin. The cleavage cocktail used consisted of 95% trifluoroacetic acid (TFA) and triisopropylsilane (TIPS) as a scavenger, unless otherwise stated. After washing with ether to remove cleaved protecting groups, the resulting peptide mixture was analysed using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. Mass spectra of test cleavages at various points during the synthesis of **69** are shown in Figure 5.5.

Due to the fact that fragment ACP(65-74) (underlined in Figure 5.5) is traditionally used as a 'difficult' sequence to test SPPS protocols, it was anticipated that the coupling of those residues close to the C-terminus would be problematic, particularly Val-65.[185–188] No problems were encountered in this instance however, and the use of dimethyl sulfoxide (DMSO) as the solvent during the coupling of the valine in question (in order to disrupt aggregation[187,188]) did not appear to have any effect on reaction efficiency.

The synthesis appears to proceed without problems until Asp-35 (43-mer, spectrum **E**), where subsequent coupling and Fmoc deprotection of the N-terminal cysteine results in a mass spectrum (**F**) containing multiple peaks which are attributed to side products resulting from amino acid deletions. As mass spectrometry is not ideally suited to quantifying the relative amounts of peptide species present in a crude product mixture, chromatographic

[l]The Wang linker is the most commonly used linker for the solid-phase synthesis of peptide acids.[183,184]

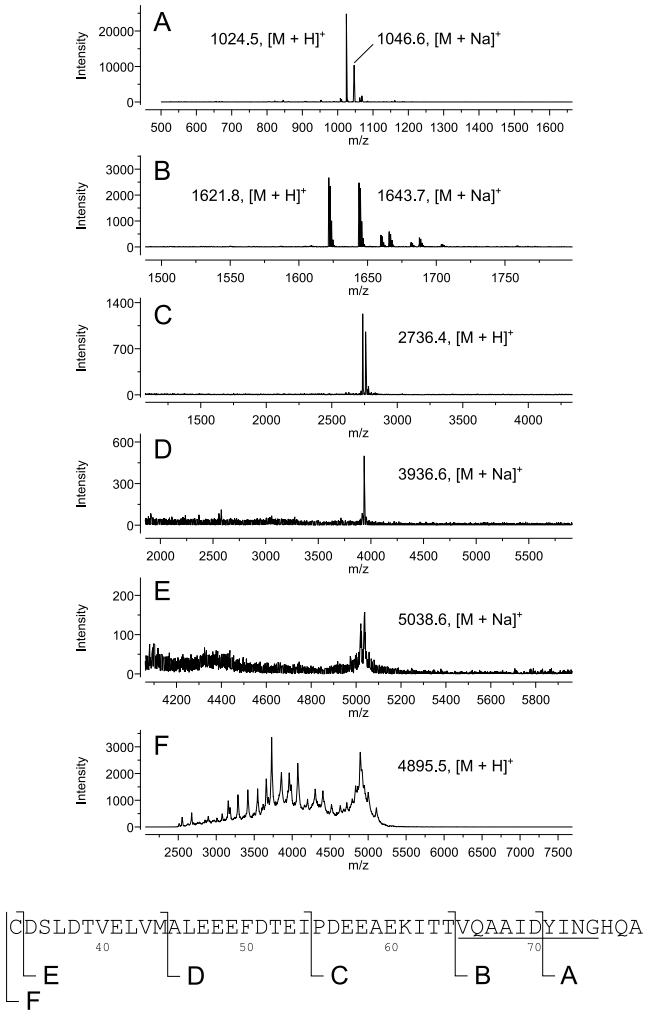



**Figure 5.5:** MALDI-TOF mass spectra from analysis of products of test cleavages performed at various points during SPPS of [A34C]ACP(34-77) (**69**). It can be seen that that the synthesis is largely unproblematic, although coupling of the final Cys residue and deprotection of the N-terminal Fmoc protecting group results in spectrum **F**, which contains multiple peaks belonging to putative deletion by-products resulting from sequential amino acid omissions (other spectra show Fmoc-protected peptide). 'Difficult' sequence ACP(65-74) is underlined (see text).

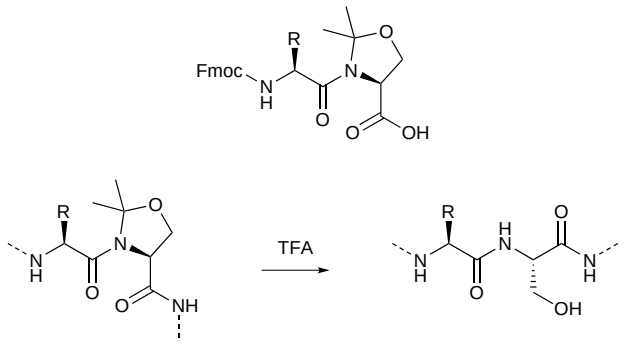
purification was attempted in order to separate the desired target **69** ($m/z$ = 4895.5) from any deletion peptides present in the crude material. Reversed-phase HPLC purification of **69** was hindered however by the acidic peptide's extremely poor solubility in conventional low pH solvent systems. Attempted solubilization of the sample via the addition of DMF, DMSO or guanidinium chloride was unsuccessful, and while basification using DIPEA was sufficient to dissolve the sample before injection, elution of the product was not observed presumably due to precipitation occurring upon introduction of the peptide into the acidic HPLC solvent flow. Due to the large number of charged residues present in the sequence (Asp/Glu), purification using ion-exchange chromatography was attempted, but was also unsuccessful.

### 5.3.2 Addressing fragment solubility

As illustrated, attempts to synthesize **69** resulted in poor quality crude product, and by far the most significant difficulty proved to be the extremely poor solubility of the crude peptide and the resulting difficulties associated with HPLC purification. In light of these difficulties, the use of pseudoproline dipeptides during the synthesis of **69** was proposed in order to disrupt aggregation and thus improve synthesis quality (Scheme 5.4). Pseudoprolines revert to the parent dipeptide upon cleavage of the peptide from the resin however, therefore this strategy does not address the issue of the poor solubility of the fully deprotected peptide during purification and handling.

The incorporation of *backbone* protection is another strategy that can be used to disrupt peptide secondary structure formation and hence improve synthesis quality during SPPS.[190–194] Furthermore, backbone protecting groups such as 2-hydroxy-4-methoxybenzyl (Hmb) can be rendered impervious to TFA-mediated removal via acetylation, and retained after cleavage from the resin in order to solubilize the product peptide during subsequent purification steps.[146,195] The steps involved in an Hmb protection strategy are illustrated in Scheme 5.5.

The automated synthesis of fragment **69** was repeated as previously carried out. Pseudoprolines were incorporated at the positions shown in Figure 5.6, and Val-43 was incorporated as the Hmb-protected building block (as this was shown to be close to the point where product quality deteriorated in previous syntheses) in a minimum volume of DMF. The leucine residue following Val-43 was double coupled using the symmetrical anhydride (each overnight,




**Scheme 5.4:** Pseudoproline dipeptides (of the form XS or XT, where X is another amino acid) can be incorporated via Fmoc SPPS and undergo acidolysis during the TFA cleavage step to generate the uncyclized dipeptide.[189] The Fmoc-protected serine dipeptide building block is shown (side chain of X represented by R group).

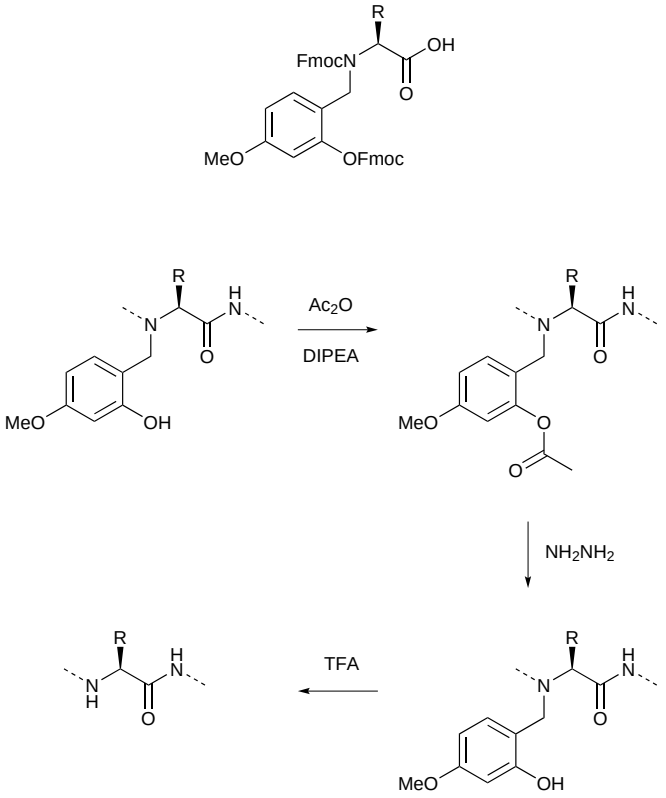
RT) due to the difficulty in acylating the phenolic Hmb hydroxyl group. This reaction was carried out in DCM, to favour the required O→N acyl migration.[‖] Microwave assistance (with DIC/HOBt activation) was also used in an attempt to increase the efficiency of the coupling and deprotection reactions.[197]

CDSLDTVEL**V**MALEEEF**DT**EIPDEEAEK**IT**TVQAAIDYINGHQA
40 50 60 70

**Figure 5.6:** Sequence of ACP C-terminal fragment [A34C]ACP(34-77) (**69**) showing the sites of incorporation of pseudoproline dipeptides and Hmb-protected building block (at Val-43), bold underlined.

Boc-Cys(Trt)-OH was coupled as the N-terminal residue to prevent reaction of the N-terminus during acetylation of the Hmb hydroxyl. The peptide-resin was first treated with piperidine to free any Hmb hydroxyl groups that may have become acylated during the synthesis, before acetylation with acetic anhydride/DIPEA.[146] DIPEA is required as Hmb acetylation was shown by Quibell et al. to be unsuccessful using acetic anhydride alone.[195] Cleavage

[‖]Acylation of the exposed Hmb hydroxyl occurs preferentially over coupling onto the hindered secondary amine at the N-terminus of the peptide. The aryl ester intermediate then undergoes an intramolecular acyl migration to generate the required amide bond.[196]




**Scheme 5.5:** Hmb backbone protection can be introduced into a peptide via incorporation of the appropriate Fmoc-protected building block.[193] Acetylation of the Hmb hydroxyl prevents loss of the protecting group during TFA-mediated cleavage of the peptide from the resin. Hmb protection can be retained to enhance solubility during subsequent purification/handling, and removed when required via deacetylation in dilute hydrazine solution, followed by TFA treatment.[146]

of the backbone-protected peptide from the resin was then carried out using standard TFA cleavage conditions.

The solubility of the crude material was found to be vastly improved when compared to that of the crude product from the previous synthesis (where Hmb protection was not utilized). The Hmb-protected peptide dissolved readily in a 1:1 MeCN/water solvent blend (containing 0.1% TFA), and injection onto a preparative HPLC column resulted in a strong absorbance signal at 220 nm that was significantly higher than the baseline level. Unfortunately however, the peptidic material observed appeared to elute as a single broad peak assumed to be comprised of an ensemble of deletion peptides, as MALDI-TOF MS analysis of multiple fractions taken from within the peak failed to identify the target peptide.

It was concluded that while incorporation of backbone protection during the synthesis did not appear to markedly improve product quality, the assumed retention of the acetylated Hmb group appeared to drastically improve solubility of the peptide, thus rendering HPLC purification possible. This demonstrates that backbone protection is a hugely valuable strategy when attempting the synthesis and purification of poorly soluble peptides. Future work will involve resynthesis of the peptide with incorporation of Hmb protection at multiple sites, with the aim of disrupting *on-resin* aggregation (in addition to solubilization of the cleaved product) enabling synthesis of the full target sequence. Various alternative backbone protecting groups have recently been reported by Abdel-Aal et al. and may also prove useful, such as 2-hydroxy-4-methoxy-5-methylsulfinyl benzyl (Hmsb)[196] and 2-hydroxy-4-methoxy-5-nitrobenzyl (Hmnb).[198] Both Hmsb and Hmnb are rendered acid labile via reduction, and can be incorporated via fully automated MW-SPPS protocols.

## 5.4 Synthesis of N-terminal section

A synthesis of ACP(1-33) *N*-alkylsulfonamide (**68**, Scheme 5.3) was also attempted via automated Fmoc SPPS using the sulfonamide linker (Chapter 4). Aspartic acid and asparagine residues were triple coupled to remove deletion sequences present in test cleavage samples from a preliminary synthesis (not

shown). Boc-Ser(O*t*Bu)-OH[III] was coupled as the final amino acid to protect the N-terminus during the subsequent linker alkylation (Scheme 5.6). The mass spectrum for the test cleavage of **71** (unalkylated) is shown in Figure 5.7.

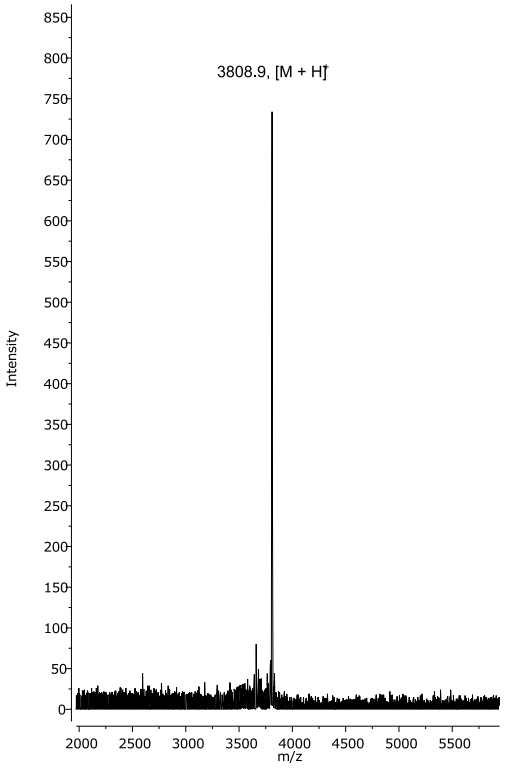
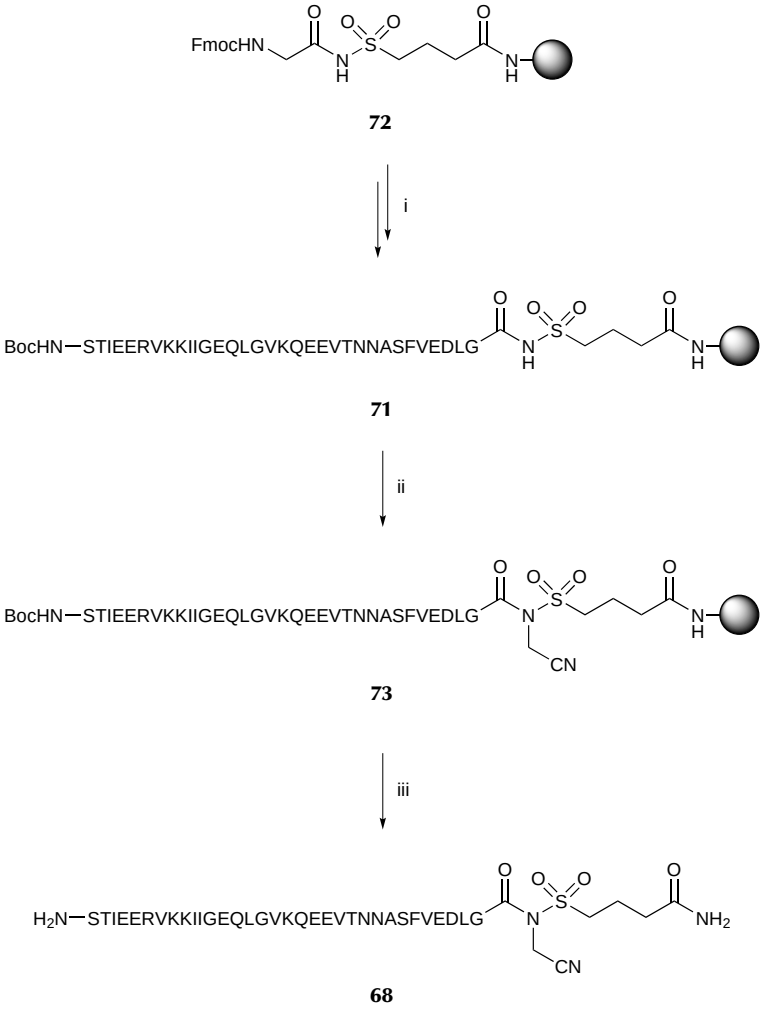



**Figure 5.7:** MALDI-TOF mass spectrum from test cleavage of unalkylated ACP(1-33) *N*-acylsulfonamide (**71**, Scheme 5.6).

Alkylation of the linker was effected by treatment of **71** with iodoacetonitrile and DIPEA in DMF for 24 h. A mass spectrum of the resulting test cleavage mixture is shown in Figure 5.8. Alkylated peptide is present alongside the corresponding peptide acid, possibly indicating that attack on the more activated C-terminal carbonyl has occurred resulting in loss of the sulfonamide linker (Scheme 5.7).

*Cyano*methylation[199] with iodoacetonitrile was used initially instead of the more widely performed methylation[200] due to the extreme toxicity[201] of the

[III]Commercially available as the DCHA salt, which is liberated with 10% $H_3PO_4$ before use.




**Scheme 5.6:** Synthesis of ACP(1-33) *N*-alkylsulfonamide (**68**) using sulfamylbutyryl Rink amide resin. Side chain protecting groups are not shown. Reagents and conditions: (i) Fmoc SPPS (see Experimental section); (ii) $ICH_2CN$ (10 equiv), DIPEA (10 equiv), DMF, RT, 24 h; (iii) TFA (95%, water), TIPS, RT, 4 h.

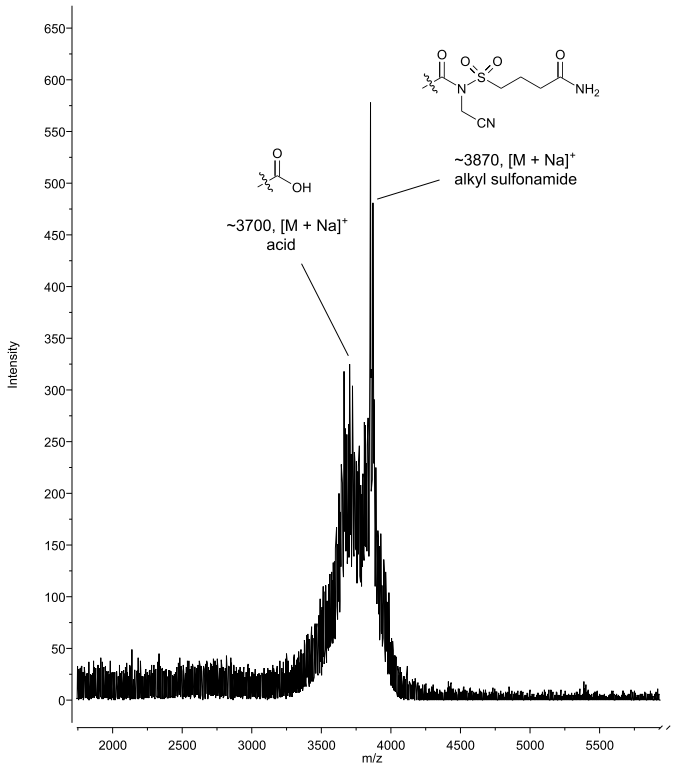
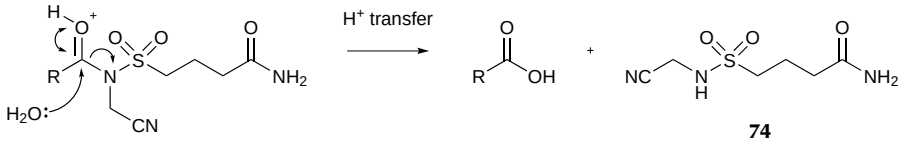



**Figure 5.8:** MALDI-TOF mass spectrum from test cleavage carried out after attempted on-resin alkylation of **71** (Scheme 5.6). While peptide bearing the alkylated linker appears to be present, the corresponding peptide acid is also observed, suggesting partial hydrolysis has occurred (resulting in elimination of the sulfonamide linker).




**Scheme 5.7:** Suggested mechanism for hydrolysis of peptide *N*-cyanomethylsulfonamides in acidic conditions (0.1% TFA solution).

methylating agent required (trimethylsilyldiazomethane). It is thought that peptide *N*-cyanomethylsulfonamides are more susceptible to hydrolysis than their methylated counterparts however, due to the presence of the electron withdrawing nitrile moiety.[199] In light of this, the linker alkylation step was repeated using trimethylsilyldiazomethane in order to generate the more stable *N-methyl*sulfonamide peptide (**75**, Figure 5.9) instead. This was carried out using facilities available at the MRC National Institute for Medical Research, London (Offer group).


**68**

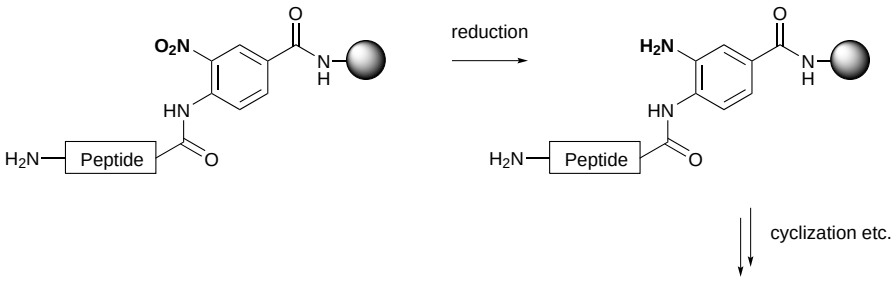

**75**

**Figure 5.9:** Comparison of *N*-alkylsulfonamides produced using the sulfonamide safety-catch linker. Cyanomethylation using iodoacetonitrile/DIPEA produces *N*-cyanomethylsulfonamides, such as fragment **68**. *Methylation* using TMS-diazomethane should instead generate more stable *N-methyl*sulfonamide **75**.

Unfortunately the required methylated peptide was not observed, and the resulting MALDI-TOF MS product peak could not be assigned (potentially due to side reactions occurring during the methylation reaction). Incorporation of a pseudoproline into the N-terminal fragment has also been carried out, in the hope that this will improve the solubility of the resin-bound peptide during treatment with the methylating agent (used as the commercially available hexane solution). Given the challenges encountered it was decided that other strategies for the synthesis of peptide thioester equivalents should be investigated.

## 5.5 A novel nitro-modified Dawson linker

The Dawson Dbz linker can be used as an alternative method for synthesizing C-terminal peptide thioester equivalents on the solid phase. Undesirable overacylation can occur when using this approach however, and while the Alloc protecting group strategy discussed in Chapter 4 (Scheme 4.5) addresses this problem, it necessarily adds an extra protection step to the synthesis. It was suggested that replacing one Dbz amine with a nitro group would also stop overacylation, but without the requirement of an initial protection reaction (Scheme 5.8). At the completion of the synthesis, reduction of the nitro group would generate the required amine, allowing cyclization and release of the product as illustrated in Chapter 4. An exploration of this approach is detailed in the following sections.

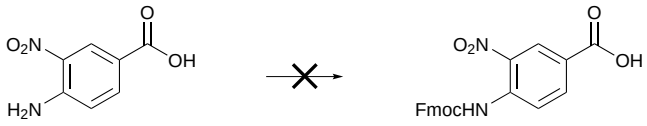



**Scheme 5.8:** Replacing an amine group on the Dbz linker with a nitro group would stop overacylation during SPPS. An amine could then be generated via on-resin reduction of the nitro group (shown in **bold**), allowing cyclization and release of the product to be carried out as normal (see Scheme 4.4, Chapter 4).

### 5.5.1 Linker loading and MW-SPPS

Initially, attempts were made to Fmoc protect commercially available 4-amino-3-nitrobenzoic acid (**76**) as illustrated in Scheme 5.9. These were unsuccessful however, and no Fmoc-protected product was observed by LCMS after $>7$ h. It was suggested that the deactivating effect of the nitro group was causing the observed difficulties in acylation of the aniline nitrogen of **76**, and consequently Fmoc protection of the linker before loading onto the solid support was

deemed to be unnecessary. Furthermore, undesirable multiple addition of the linker would involve coupling of the apparently unreactive and relatively sterically hindered amine onto a similarly hindered carboxylic acid group, and was predicted to be unlikely. Indeed, the standard Dbz linker is loaded routinely bearing Fmoc protection on only one of its amine groups, as steric and electronic factors are expected to prevent reaction at the second.[150]

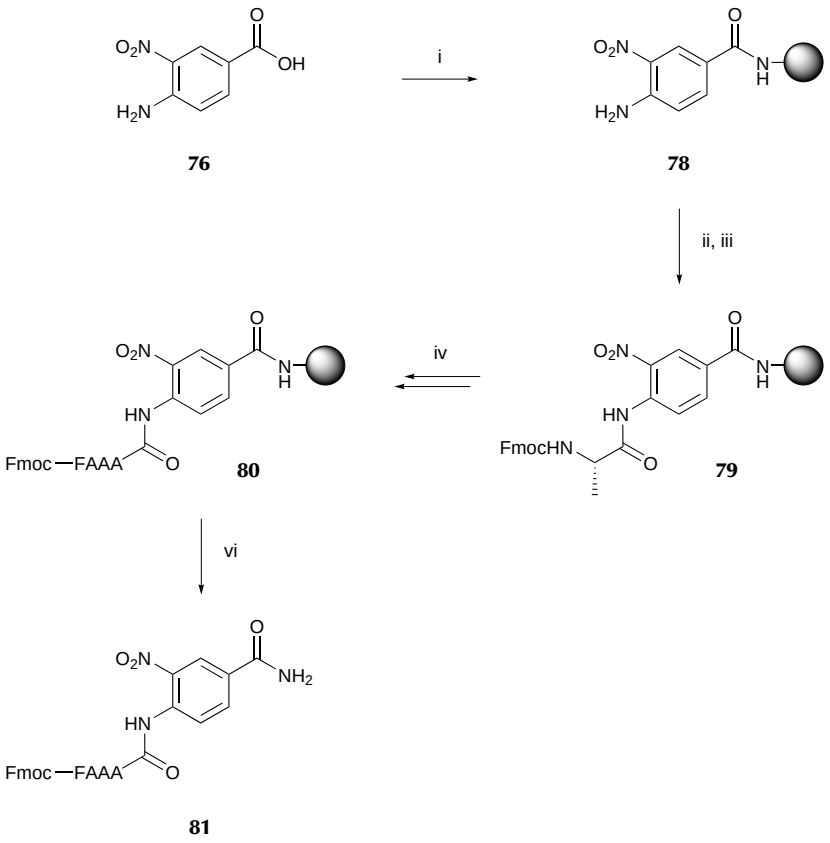



**76** **77**

**Scheme 5.9:** Attempted Fmoc protection of 4-amino-3-nitrobenzoic acid (**76**). No Fmoc-protected product was observed whatsoever after >7 h (by LCMS). Reagents and conditions: Fmoc-OSu (1.0 equiv), 1:1 $NaHCO_{3(aq)}$ (0.1 M)/MeCN, RT, >7 h.[77,150]

Linker loading and SPPS are shown in Scheme 5.10. Unprotected **76** was loaded directly onto Rink amide polystyrene resin (0.33 mmole/g) using a single 1 h coupling (5.0 equiv) and HATU activation. Forcing conditions were necessary as the loading reaction was found to be relatively inefficient (see below), presumably due to the electron withdrawing effect of the nitro group. After capping with acetic anhydride (to acetylate remaining Rink amide resin amine groups) the test peptide FAAA was built onto the linker using automated MW-SPPS (double couplings). While capping was anticipated to potentially also acetylate linker aniline groups, it was necessary to ensure all unreacted resin amine groups were removed prior to MW-SPPS to prevent C-terminal peptide amides being released upon cleavage.

A MALDI-TOF mass spectrum of products from the test cleavage is shown in Figure 5.10. While target peptide **81** is present ($m/z$ = 786.3), a larger peak at $m/z$ = 622.2 could be assigned to either a double alanine deletion or **81** missing the desired aryl nitro C-terminus (peptide amide only).




**Scheme 5.10:** Loading of and MW-SPPS using novel nitro-Dbz linker. Reagents and conditions: (i) Rink amide, **76** (5.0 equiv), HATU (4.8 equiv) and DIPEA (10 equiv), DMF, RT, 1 × 1 h; (ii) 20% acetic anhydride, DMF, RT, 30 min; (iii) Fmoc-Ala-OH (5.0 equiv), PyBOP® (10 equiv) and DIPEA (20 equiv), DMF, 75 °C, 2 × 10 min, MW; (iv) Fmoc MW-SPPS (see Experimental section); (vi) TFA (95%, water), TIPS, RT, 4 h.

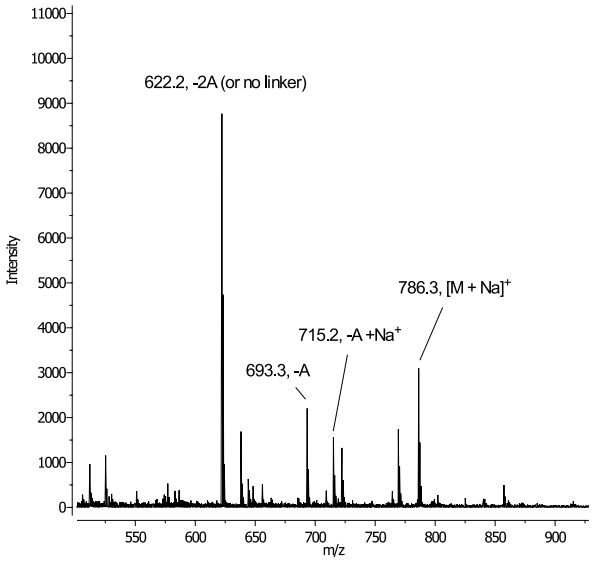


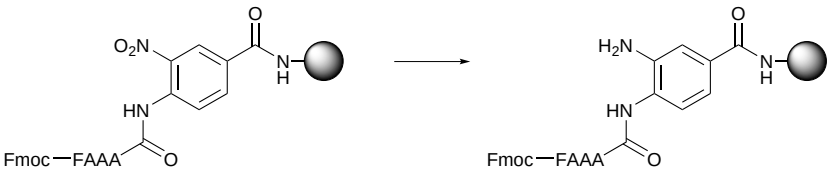

**Figure 5.10:** MALDI-TOF mass spectrum from test cleavage of peptide **80** (Scheme 5.10). While target **81** and deletions (bearing the nitro linker) are present, the largest peak can be assigned to **81** missing the linker (peptide amide only).

## 5.5.2 Linker nitro group reduction

On-resin reduction of the nitro group was attempted using the mild reducing agent sodium hydrosulfite ($Na_2S_2O_4$) under the phase-transfer conditions developed by Kaplánek et al. for reduction of resin-bound nitroarenes (Scheme 5.11).[202] Although $Na_2S_2O_4$ has been demonstrated to effectively reduce compounds of this type,[203] the reagent suffers from poor solubility in organic solvents and needs to be used in aqueous solution, which is incompatible with conventional polystyrene solid supports. While attempts at reductions using blends of water and water-miscible solvents such as NMP, DMF and THF have met with some success,[203] it was thought that a biphasic system of DCM and water with an added phase-transfer catalyst would be more suitable to ensure swelling of the resin and simultaneous solvation of the reducing agent.[202]

The resin was agitated with $Na_2S_2O_4$ and $K_2CO_3$ in a 1:1 mixture of water and DCM to which the phase-transfer catalyst tetrabutylammonium hydrogen sulfate (TBAHS) had been added. After 9 h the yellow colouration of the support had disappeared, and a test cleavage was performed (mass spectra shown in Figure 5.11). While reduction of the nitro species present appears

to be have been partially successful, it was tentatively suggested that only a small proportion of the aryl amine (Dbz-functionalized) peptide is present after reaction under the conditions used. Repeating the reduction in water/DMF with microwave heating at 40 °C/10 W[IV] was not successful. No deprotection of the Fmoc group was observed in the presence of $K_2CO_3$.




**80** **82**

**Scheme 5.11:** On-resin reduction of the linker nitro group under phase-transfer conditions. Reagents and conditions: $Na_2S_2O_4$ (20 equiv), $K_2CO_3$ (28 equiv), TBAHS (2.0 equiv), DCM/$H_2O$ (1:1), RT, 9 h.

[IV] $Na_2S_2O_4$ disproportionates above 50 °C.[203]

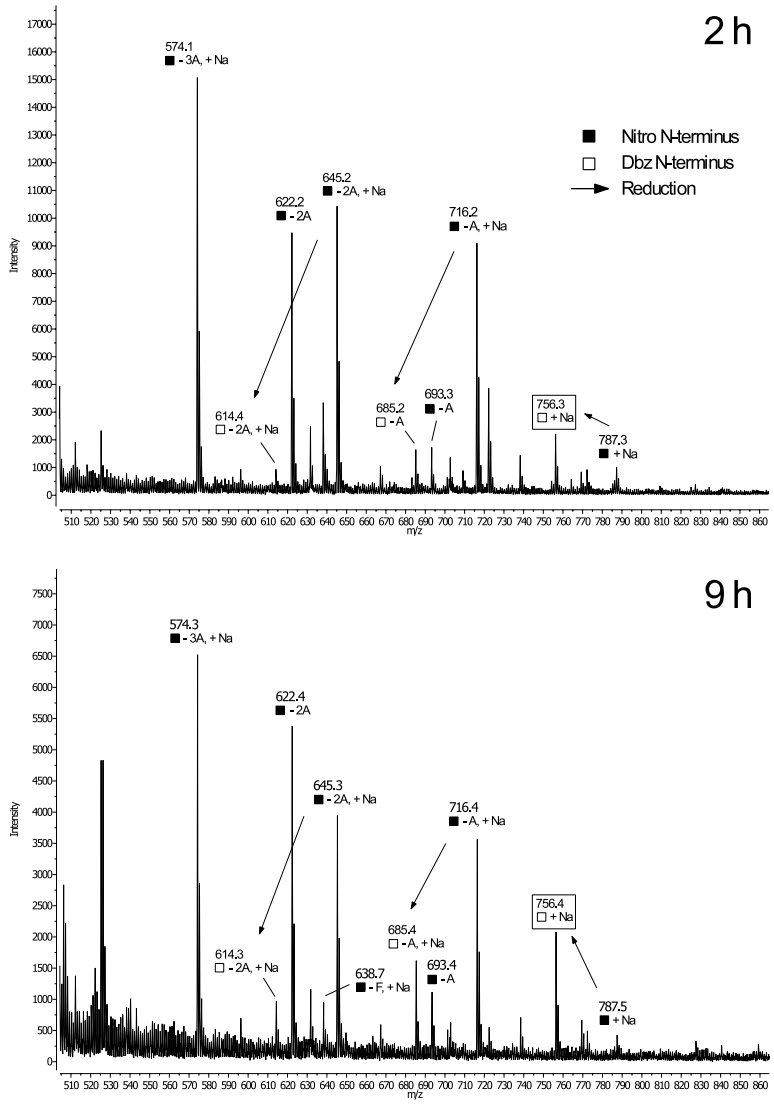



**Figure 5.11:** MALDI-TOF mass spectra from test cleavages of linker nitro group reduction after 2 h (top spectrum) and 9 h (bottom spectrum). Peaks corresponding to peptides bearing aryl nitro C-termini are labelled with filled boxes (■), while those belonging to *reduced* (Dbz) peptides are marked with unfilled boxes (□). The target peptide peak is boxed. Reduction of the linker nitro group appears to be partially successful, with an apparent increase in reduced species with prolonged reaction time. Automated peak assignment was carried out using Peptide Calculator (see Chapter 6).[2]

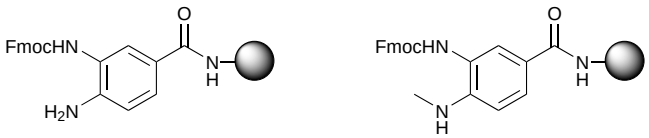
## 5.6 Conclusions and future work

Significant progress has been made towards the total chemical synthesis of *E. coli* ACP. The complete N-terminal fragment has been synthesized as the *N*-alkylsulfonamide by SPPS, and while cyanomethylation was found to produce potentially unstable peptide intermediates, attempts to produce the more stable methylated sulfonamide analogues are ongoing. Synthesis of the C-terminal fragment could not be completed due to solubility issues, although these were circumvented by introduction of Hmb backbone protection, resulting in a soluble crude product that was observed by HPLC. Further syntheses will be trialled involving the incorporation of the Hmb group at other positions in the sequence, with the aim of increasing the quality of the crude product. Native chemical ligation of the two sections will be carried out in neutral aqueous buffer in the presence of a thiol catalyst, and native apo-ACP will be obtained via metal-free dethiylation in the final step.

The application of a novel nitro-functionalized linker based on Dawson's Dbz linker has also been demonstrated successfully, which should allow overacylation-free MW-SPPS. This serves as an alternative approach for the synthesis of peptide thioester equivalents on resin, such as the N-terminal fragment required for the total chemical synthesis of ACP. Improvement of the linker reduction step is necessary, and different reduction conditions should be investigated. For example, a hydrophilic poly-$\varepsilon$-lysine support such as SpheriTide® may be more suited to reactions using water-soluble reagents in aqueous solvent.[204,205] An alternative approach for the reduction of aromatic nitro groups on resin using chromium chloride ($CrCl_2$)[206] and was also recently demonstrated by Abdel-Aal et al. to be significantly more effective than the $Na_2S_2O_4$ reduction under phase-transfer conditions attempted here.[198]

However, during the period of work described Blanco-Canosa et al. reported an improvement to their existing Dbz approach in the form of a 'second generation' linker bearing a methylation on the alternate aniline nitrogen (Figure 5.12), which was demonstrated to effectively suppress overacylation. It is also interesting to note the use of a glycine spacer in this study to facilitate coupling of the sterically hindered aromatic linker to the resin, an approach which may improve linker loading for the nitro-functionalized linker described. The 'MeDbz' linker represents an elegant solution to the problem of overacylation, and while there are certainly benefits to the method discovered here, such as rendering undesirable overacylation

completely impossible, given the apparent poor linker loading and reduction efficiency it is suggested that use of the MeDbz linker for the synthesis of the N-terminal ACP fragment may be preferable.




**57** **83**

**Figure 5.12:** Comparison of Dbz linker (**57**) and second-generation 'MeDbz' linker (**83**) recently reported by Blanco-Canosa et al.[207]

Once complete, the total chemical synthesis will enable the creation of site-specifically modified ACP analogues which will be used to investigate the role of ACP-subunit and ACP-substrate interactions in fatty acid synthetases, and the identification of specific residues involved at those interfaces. The incorporation of fluorescent probe residues[116,208] will allow site-specific labelling of the protein, offering the potential to use fluorescence resonance energy transfer (FRET) experiments to investigate domain arrangement and interaction distances between ACP and labelled substrate or catalytic subunit residues. The specificity afforded will circumvent selectivity issues due to off-site tagging encountered during previous attempts to measure intersubunit distances in fatty acid synthetases using FRET.[209,210]

Labelling using fluorinated residues will enable structural characterization via fluorine-19 NMR studies.[117] The incorporation of non-hydrolysable $^{13}C$-labelled fatty acid substrate analogues at Ser-36 will be possible during Fmoc SPPS (see Figure 5.4). By inserting a fluorinated phenylalanine building block at residue Phe-50, heteronuclear nuclear Overhauser effect (NOE) experiments between $^{13}C$ and $^{19}F$ labels will be used to probe protein structure and dynamics in solution. This will help to elucidate the crucial role of substrate sequestration within the hydrophobic binding cleft of ACP,[176] shedding light on how this causes conformational changes within the protein, and the extent to which these modulate ACP-subunit interactions and confer specificity for designated catalytic domains.

The introduction of residues bearing selectively reactive groups will also allow site-specific cross-linking between ACP and catalytic subunits. Azide-tagging, for example, would enable the use of Azide-Alkyne Cycloaddition chemistry, whereby the tagged ACP could be covalently 'clicked' onto a catalytic domain bearing a genetically encoded alkyne moiety.[211] The ability to site-specifically incorporate each tag means that the cross-linking could be carried out with high spatial precision, with no risk of off-site reactivity. High-resolution structural characterization of cross-linked ACP-enzyme complexes will facilitate elucidation of synthetase protein-protein interactions via crystallographic and NMR techniques, and cryo-electron microscopy.

Total chemical synthesis will enable the generation of diverse libraries of selectively modified ACPs, thus paving the way for the production of novel tools with which the function of modular synthetases can be interrogated. The toolkit of chemically modified ACPs will be invaluable in shedding light on the mechanisms behind the function of FAS, PKS and multienzyme megasynthetases in general.

# Part III

# Synthesis of dual function cancer targeting peptides

# Chapter 6

# Synthesis of tumor targeting peptides for cancer cell imaging and tumor growth inhibition

The inhibition of a protein target shown to be essential for tumor growth is a desirable goal for researchers seeking to understand the cellular processes underlying tumorigenesis. The conjugation of emissive probe moieties to peptides incorporating inhibitory sequences and/or those which target cellular compartments offers the potential for selective imaging of these processes alongside tumor growth inhibition. The following chapter describes the synthesis of a number of peptides designed to target a specific protein shown to be essential for tumor growth. Their incorporation into luminescent probes and use in cancer cell imaging and tumor growth inhibition is also described.

## 6.1 Introduction

Epstein-Barr virus (EBV) is a member of the herpesvirus family that infects over 90% of humans worldwide.[212] EBV has been discovered in tissues from patients with numerous cancer types including T-cell lymphomas, Hodgkin's disease, non-Hodgkin's lymphoma, Burkitt's lymphoma and nasopharyngeal carcinoma. In the case of the latter, tumorigenesis has been attributed to the latent EBV infection.[213] Epstein-Barr nuclear antigen (EBNA1) is a critical component of the EBV lifecycle and has been found to directly

contribute to the tumorigenic cell phenotype. Given that homodimerization is required for EBNA1 to function, the development of agents that inhibit the EBNA1 dimerization process represents a significant opportunity for potential therapeutic targeting of EBV-associated cancers.[6]

While a number of EBNA1-specific protein-protein interaction inhibitors have been reported, most cannot be easily imaged *in vitro* and suffer from low bioavailability and poor water solubility.[6] To overcome these disadvantages a novel hybrid system was proposed consisting of a water-soluble EBNA1 targeting peptide conjugated to a small organic fluorophore. The approach is illustrated in Figure 6.1, and offers the potential for simultaneous imaging and EBV-associated tumor growth inhibition *in vitro*. The design, synthesis and biological evaluation of the proposed system is described in the following sections.

## 6.2 Design and synthesis of EBNA1 targeting peptides

As EBNA1 is primarily located in the nucleus of EBV-infected cells, the targeting peptide required both EBNA1 and nuclear targeting regions to be incorporated into its sequence. Binding free energies of potential peptide conjugates were evaluated *in silico* following molecular dynamics simulation of each conjugate in complex with the dimerization interface of the EBNA1 DNA-binding domain (DBD).[6] The targeting sequences shown in Figure 6.2 were proposed, incorporating an EBNA1 targeting pentapeptide taken from chain B of the EBNA1 DBD and a polyarginine nuclear targeting sequence. Flexible linking regions were also included between the two targeting sequences and at the N-terminus (for attachment to the fluorophore).

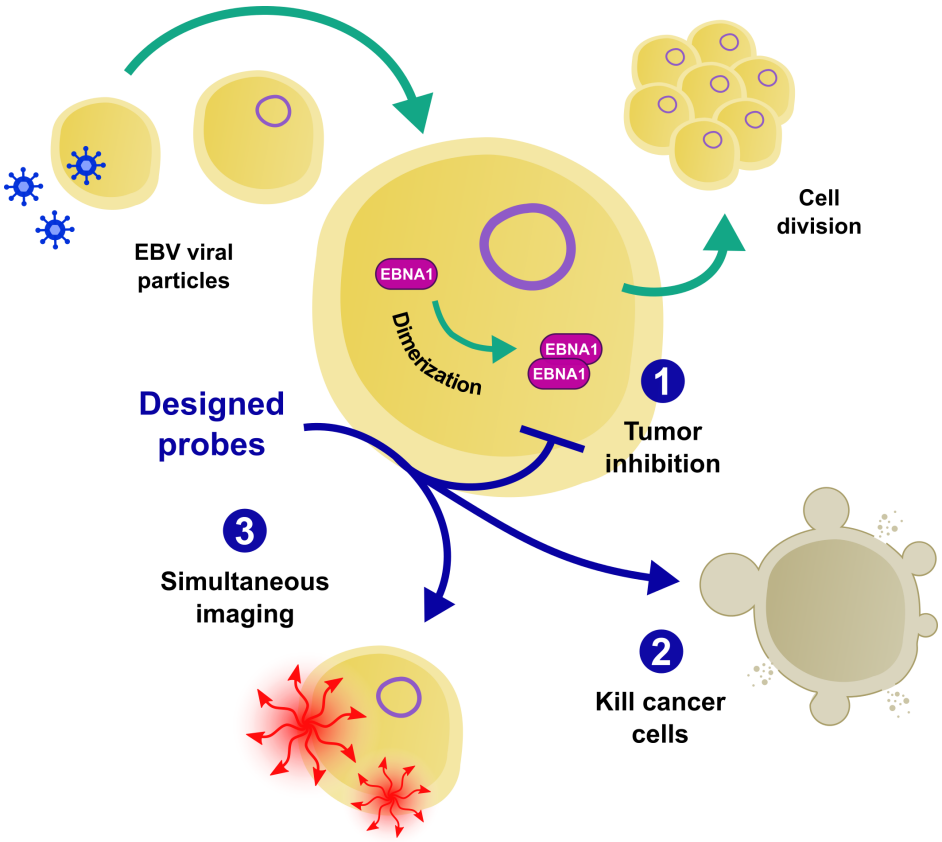
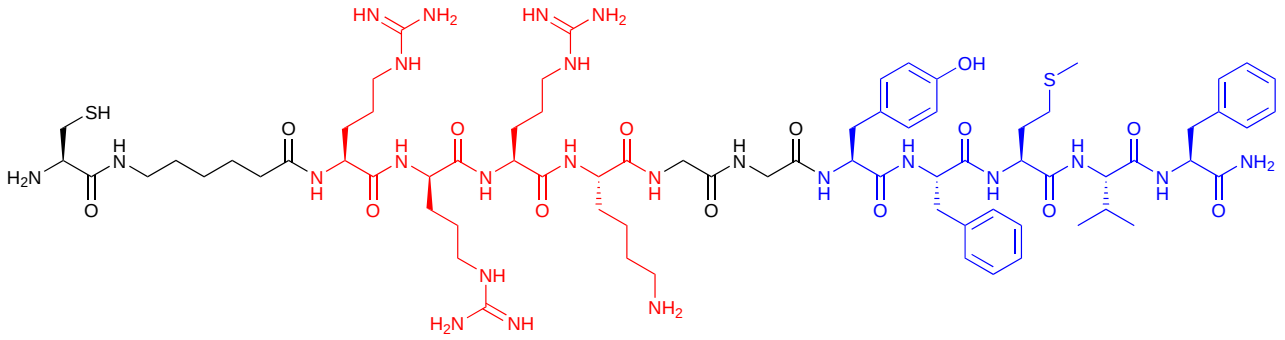
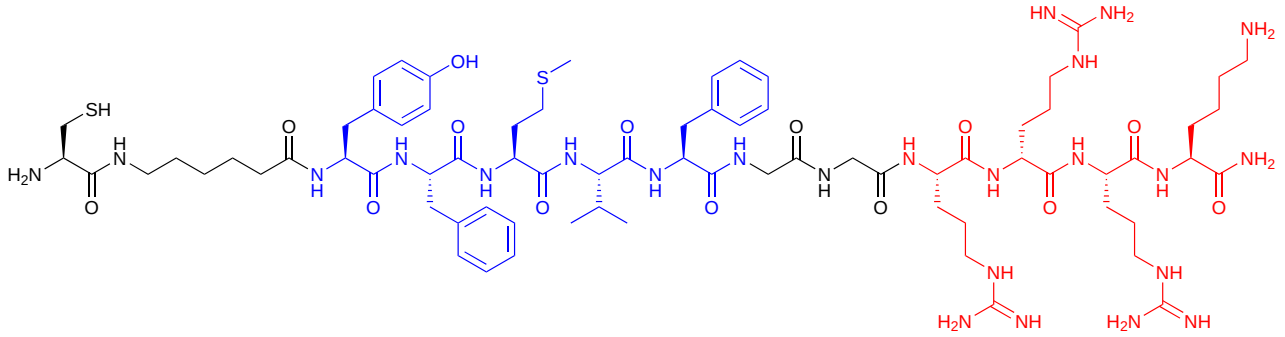



**Figure 6.1:** Targeting EBV-associated tumors using a dual-function probe. An emissive inhibitor of EBNA1 dimerization can be used for tumor growth inhibition and cancer cell killing (① and ②), with simultaneous imaging, ③. Adapted from Jiang et al.[6]



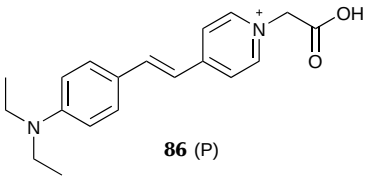
C-Ahx-RrRK-GG-YFMVF-$NH_2$ (**84**)



C-Ahx-YFMVF-GG-RrRK-$NH_2$ (**85**)

**Figure 6.2:** Peptide sequences synthesized incorporating EBNA1 targeting motif YFMVF (shown in blue) and nuclear localization sequence RrRK (red). A diglycine region between the two targeting sequences and an aminohexanoic acid (Ahx) residue at the N-terminus act as flexible linkers. Lowercase lettering indicates D-amino residues.

Automated synthesis of peptides **84** and **85** was carried out via Fmoc MW-SPPS using a CEM Liberty1 peptide synthesizer. Rink amide resin (0.82 mmol/g) was used in order to install the required C-terminal amide group. The amino acid building block was added manually to the reaction vessel at the required point in the coupling cycle for D-Arg and Ahx residues. Double couplings and extended Fmoc deprotection reaction times (using MW heating) were utilized for Arg residues due to the difficulty in coupling multiple consecutive Pbf-protected arginine building blocks. Double couplings were also found to be necessary for Cys residues, and an extended cycle was required for Ahx. The presence of each target peptide was confirmed by MALDI-TOF MS after test cleavage ($m/z$ = 1632.0 and 1631.7, for **84** and **85** respectively). Resin-bound peptides were sent to the research group of Gary Ka-Leung Wong (Hong Kong Baptist University) for on-resin coupling of the fluorophore shown in Figure 6.3 to the N-terminus, cleavage from the resin, HPLC purification and biological evaluation.[6]




**86** (P)

**Figure 6.3:** Structure of small organic fluorophore coupled to N-terminus of peptides **84** and **85** (to give P-**84** and P-**85**).[6]

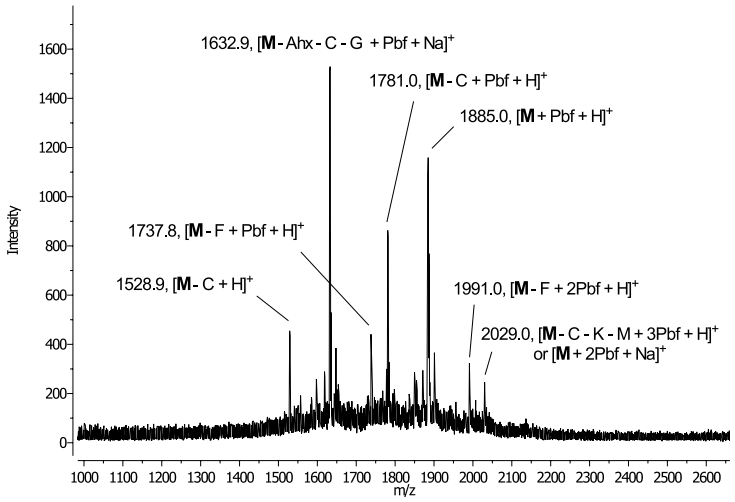
## 6.3 Automated assignment of peptide mass spectra

Mass spectrometry is the principle technique used in peptide chemistry for initial confirmation that a target molecule is present following a synthesis (as evidenced by the work presented thus far). Where inefficient couplings occur during syntheses of so-called 'difficult' sequences (such as those encountered in Chapter 5) multiple amino acid deletion peptides are often present, which appear as by-product peaks in the mass spectrum and can greatly complicate

analysis. Furthermore, most of the synthetic peptides described in this and the preceding chapters contain unnatural or non-canonical amino acids, or chemically modified termini, which cause calculation of peptide molecular weight to be non-trivial. These complications, in combination with the fact that a given peptide may also appear in the mass spectrum in the form of a metal and/or protecting group adduct, often result in the assignment of peptide mass spectra being a challenging and time-consuming process.

To address this problem the web utility Peptide Calculator was developed (Lear and Cobb, 2016), and has been used for the assignment of mass spectra shown throughout this report.[2] An example assignment automatically generated by Peptide Calculator is shown in Figure 6.4 for a MALDI-TOF mass spectrum from a synthesis of peptide **85** (Figure 6.2). Many by-product peaks are present in the spectrum and upon initial inspection the molecular ion for the target peptide is not found. However, through reference to the assignment table given it is apparent that the target peptide is present as the Pbf adduct ($m/z = 1885.0$), indicating that an extended TFA deprotection time is required. During sequence input multiple non-canonical residues can be incorporated such as Ahx (Figure 6.4), Dpr (Chapter 2), and Alloc- and allyl ester-protected residues (Chapter 3), in addition to user-specified N- and C-terminal modifications, including N-terminal acylation (Chapter 2), C-terminal alkylation (Chapter 3), and the plethora of thioesters and thioester equivalents described previously in Chapters 4 and 5.

Various other features are also available alongside molecular weight calculation and MS assignment, including calculation of isoelectric point and molar extinction coefficient, and ChemDraw structure file export. A plot of $\beta$-strand propensity is also generated for sequences where an indication of potential synthesis difficulty may be required. The MS assignment features described are also available for peptoids through the sister utility Peptoid Calculator, and both utilities are available online at http://www.pep-calc.com. More information is available in the article included in Appendix D. Peptide Calculator has greatly simplified the assignment of all mass spectral data reported, in particular the characterization of complex mixtures of peptides exhibiting a variety of unusual C-terminal functionality described in Chapter 5 (Figure 5.11).




| *Peak* (m/z) | *Deletion*(s) | *Adduct*(s) | *Loss*(es) | *Calculated mass* |
|---|---|---|---|---|
| 1528.9 | C | - | - | 1528.87 |
| 1632.9 | Ahx, C, G | $Na^+$, Pbf | - | 1633.75 |
| 1737.8 | F | Pbf | - | 1737.81 |
| 1781.0 | C | Pbf | - | 1781.87 |
| 1885.0 | - | Pbf | - | 1884.88 |
| 1991.0 | F | 2Pbf | - | 1990.81 |
| 2029.0 | C, K, M | 3Pbf | - | 2028.73 |
| ″ | M | $Na^+$, 2Pbf | - | 2028.82 |

**Figure 6.4:** MALDI-TOF mass spectrum from test cleavage of peptide **85** (sequence shown in Figure 6.2) annotated with peak assignment automatically generated by Peptide Calculator. The assignment includes deletions, unremoved protecting groups and metal ion adducts, with support for non-canonical residues (such as Ahx). Custom formulae can also be specified in the case of N- or C-terminal modification.

## 6.4 Investigation of tumor imaging and growth inhibition

The binding interaction of fluorescent conjugates P-**84** and P-**85** with EBNA1 was first characterized experimentally via luminescence titration experiments. A significant responsive emission enhancement was observed for P-**85** upon addition of EBNA1, corresponding to an apparent binding constant log $K_a$ of 6.82. An enhancement was also observed for P-**84** (log $K_a$ = 5.50). A chemical cross-linking assay was performed to evaluate the ability of the peptide conjugates to disrupt EBNA1 dimer formation in the presence of 3-maleimidobenzoyl *N*-hydroxysuccinimide ester (MBS), an amine-to-thiol crosslinker containing NHS ester and maleimide reactive groups. The dimerization efficiency was greatly decreased upon mutation of the YFMVF motif in the binding site, and furthermore also significantly for the wild type in the presence of conjugates P-**84** and P-**85**.

Fluorescence microscopy was used to image C666-1 cells (a native nasopharyngeal carcinoma cell line with latent EBV infection) treated with P-**84** and P-**85**, and co-stained with a nuclear dye (Figure 6.5). Subcellular localization of both peptides is observed in the nucleus, with P-**85** exhibiting the best cellular uptake. Nuclear localization was also observed for P-**85** in EBV-negative cancer cell lines (CNE-2 and HeLa).

The colourimetric MTT assay was used to assess cytotoxicity of P-**84** and P-**85** towards cancer cells *in vitro*. Both conjugates were found to exhibit selective cytotoxicity against the EBV-positive C666-1 cell line (P-**85** $IC_{50}$ = 15 μM), with $IC_{50}$ values greater than 0.5 mM for EBV-negative MRC-5 and HeLa cell lines. *In vivo* effectiveness was also investigated via intratumoral injections of P-**85** or unconjugated peptide **85** in mice carrying C666-1 xenografts (Figure 6.6). After 21 days average tumor volumes reduced by 65.5% when treated with the peptide alone ($p < 0.005$) and 92.3% when treated with the conjugate ($p < 0.001$), compared to the control. After sacrifice and tumor excision, average tumor weights were found to have decreased by 86.6% ($p < 0.05$) and 92.8% ($p < 0.005$) respctively, compared to the control.

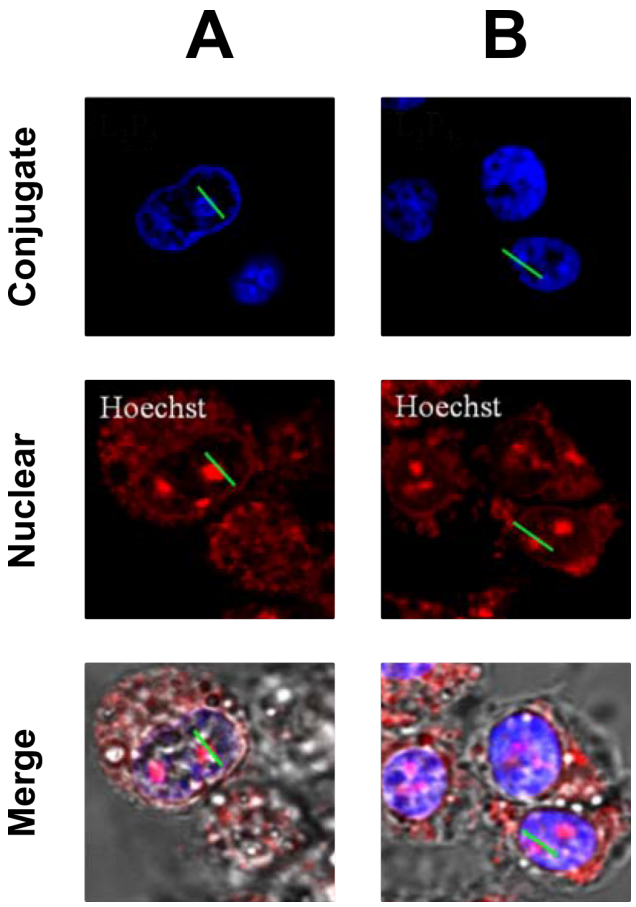



**Figure 6.5:** *In vitro* imaging of C666-1 cells treated with P-**84** (**A**) or P-**85** (**B**) for 6 h (10 μM), then co-stained with the nuclear dye Hoechst 33342 for 1 h (1 nM). Emission intensity profiles were plotted along the green line indicated, and were found to match between the conjugates (found in the nucleus) and the nuclear stain (under the same excitation).[6]

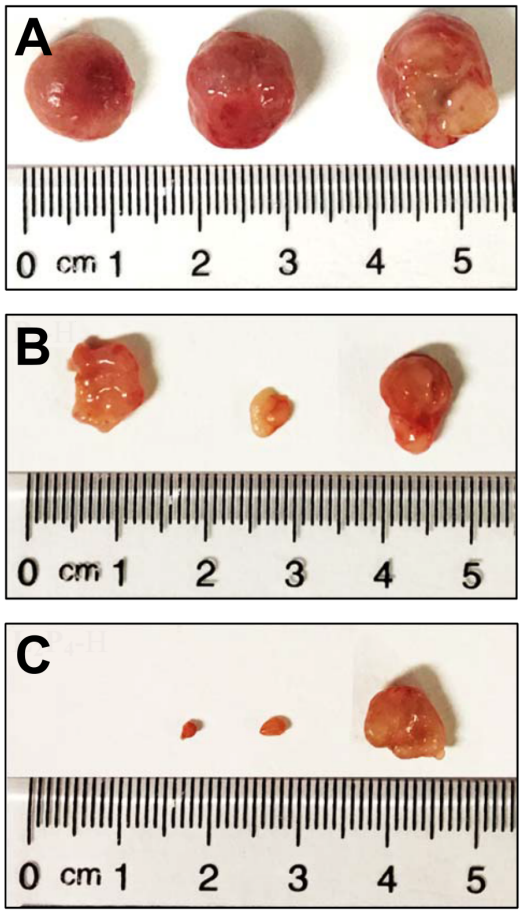



**Figure 6.6:** C666-1 mouse xenografts after 21 days of twice-weekly intratumoral injections of DMSO control (**A**), or 4 μg of unconjugated **85** (**B**) or P-**85** (**C**).[6]

## 6.5 Conclusions and future work

Two peptides have been successfully synthesized incorporating both EBNA1 and nuclear targeting sequences. Couplings in the potentially difficult polyarginine region were accomplished successfully through optimization of the coupling and deprotection conditions. The peptides were conjugated to a fluorescent probe and shown to bind to EBNA1 *in vitro* and disrupt homodimerization, a process known to be crucial to its function as part of its role in EBV-associated tumor growth. Tumor cell imaging was also demonstrated using the conjugates, which were found to undergo effective nuclear localization in a number of cancer cell lines. In addition to their imaging potential, selective *in vitro* cytotoxicity against EBV-positive cancer cells was also demonstrated, as well as *in vivo* tumor xenograft growth reduction for conjugate P-**85**.

The results reported will allow detailed investigation into the processes underlying tumorigenesis for EBV-associated malignancies. The conjugates described will be used to unravel the nature of EBNA1 tethering to host cell chromosomes, a crucial but as yet poorly understood process.[6] The insights gained will pave the way in the development of diagnostic tools and therapeutics against EBV-associated cancers.

# Part IV

# Experimental details and references

# Chapter 7

# Experimental

## 7.1 Peptide synthesis

All reagents were purchased from Sigma-Aldrich unless otherwise specified. Peptide synthesis grade DMF was purchased from AGTC Bioproducts (Hessle, UK) or Fisher Scientific, and amino acid derivatives were purchased from CEM, Novabiochem (Merck) or AGTC. DIC and HOBt were purchased from Fluorochem (Hadfield, UK), and PyBOP® from Apollo Scientific (Stockport, UK). All resins were purchased from Novabiochem.

Fmoc SPPS procedures are detailed in the following sections. Fritted polypropylene reaction vessels were used for all manual reactions and resin swelling, with DMF as the reaction solvent. For peptide couplings DIC or PyBOP® was used as the activator unless otherwise indicated. Amino acid side chain functionality was protected as follows: Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(O*t*Bu)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Dpr(Boc)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(O*t*Bu)-OH, Fmoc-His(Trt)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Ser(O*t*Bu)-OH, Fmoc-Thr(O*t*Bu)-OH, Fmoc-Trp(Boc)-OH and Fmoc-Tyr(O*t*Bu)-OH. Fmoc deprotections were carried out using a 20% (v/v) solution of piperidine in DMF. Pre-swelling of the resin was carried out in DMF for a minimum of 1 h (overnight preferred), followed by washing with DMF. The resin was shrunk in diethyl ether to remove DMF in preparation for TFA cleavage. Peptide-resin was stored at −18 °C. Peptide molecular weight calculations and mass assignments were carried out using Pep-Calc.com.[2]

### 7.1.1 Automated Fmoc SPPS

Automated SPPS was carried out on a CEM Liberty1 single-channel microwave peptide synthesizer equipped with a Discover microwave unit. All reactions were carried out at 0.1 mmol scale using the 30 mL PTFE reaction vessel, with agitation by bubbling nitrogen. Couplings were carried out using Fmoc-protected amino acid (5.0 equiv), PyBOP® (5.0 equiv) and DIPEA (10 equiv, 2 M solution in NMP) unless otherwise stated. For double and triple couplings the reaction vessel was drained after each cycle and fresh reagents were added. For DIC couplings, a 0.8 M solution of DIC in DMSO was used in the activator base position, and a 0.5 M solution of HOBt in DMF was used in the activator position.

#### Microwave-assisted

Microwave couplings were carried out for 10 min at 75 °C and 25 W power unless otherwise stated. Cys and His residues were coupled at low temperature: 10 min at room temperature followed by 10 min at 50 °C (25 W). Arg residues were double coupled, with the first coupling carried out for 45 min at RT followed by 5 min at 75 °C (25 W), and the second using the standard microwave conditions given. Removal of the Fmoc group was carried out at room temperature using two successive treatments with piperidine solution (5 + 10 min). For DIC couplings a room temperature 'preactivation' period of 30 min, 1 h or 2 h was used.

#### Room temperature

Reactions were carried out at room temperature using 2 × 1 h couplings (3 × 1 h for Arg residues). The Fmoc group was removed by two successive treatments with piperidine solution (5 + 10 min). For lassomycin, two additional deprotections were carried out for the penultimate residue (Leu-2).

### 7.1.2 Manual Fmoc SPPS

Manual SPPS was carried out in fritted polypropylene reaction vessels. Couplings and deprotections were carried out using the conditions for automated room temperature SPPS already described (see section 7.1.1). Efficient mixing during the synthesis was achieved using a shaker (without

nitrogen agitation). Where indicated, capping of the peptide N-terminus was carried out via acetylation using a 20% solution of acetic anhydride in DMF for 30 min at RT.

For the chaiyaphumines, syntheses were carried out at 0.2 or 0.4 mmol scale using 2 equivalents of the amino acid, with PyBOP® (2.0 equiv) as the activator in the presence of DIPEA (4.0 equiv) (apart from Pro residues, which used 5:5:10). N-terminal acylation was carried out according to procedure 7.1.5. The enantiomeric derivatives Fmoc-D-Ala-OH and Fmoc-D-Phe-OH were used for the corresponding residues.

### 7.1.3 Cleavage of peptides from acid-labile resins

Peptide-resin was treated with 2.85 mL TFA and 0.15 mL deionized water, with 0.15 mL TIPS as a scavenger (unless otherwise stated) for 3–4 h at room temperature (volumes were reduced by two thirds for test cleavages). Longer cleavage times were used for peptides containing Pbf or benzyl protecting groups. The resin was then removed by filtration and the filtrate concentrated *in vacuo* before precipitation using ether and decanting of the liquid (followed by subsequent ether washes). For peptides bearing only protecting groups releasing volatile products upon cleavage (such as Boc), the ether was simply re-evaporated instead of decanting. The resulting solid peptide was dissolved in deionized water and lyophilized. For hydrophobic sequences a small amount of acetonitrile was added ($< 50\%$ v/v) in the event of poor solubility (water and MeCN containing 0.1% TFA or triethylamine were also used to improve solubility for particularly insoluble peptides).

### 7.1.4 Loading of 2-chlorotrityl chloride resin

2-Chlorotrityl chloride resin was pre-swelled for 1 h in DMF followed by addition of Fmoc-Trp(Boc)-OH (1.0 equiv) and DIPEA (5.0 equiv) in DMF ($2 \times 1$ h). After washing with 5 portions of DMF, the linker was capped via treatment with MeOH (15 min). Swelling was repeated in DMF for at least 1 h before continuation of the synthesis.

### 7.1.5 Acylation with N-terminal capping group

For the chaiyaphumines, acylation of the N-terminal residue was carried out via treatment of peptide-resin with 5.0 equivalents of the acyl chloride specified (or acetic anhydride) and DIPEA (10 equiv) in DMF (2 × 1 h). The resin was washed with 5 portions of DMF between couplings.

### 7.1.6 Standard solution-phase peptide cyclization

Cyclizations were carried out as follows unless otherwise stated. Fully deprotected linear peptide was dissolved in dry DMF and added dropwise over a period of 2 h to a further 4 volume parts of a stirred room temperature DMF solution containing PyBOP® (3.0 equiv) and DIPEA (6.0 equiv), with a final peptide concentration of 0.003 M. The reaction was stirred for a further 3 h before removal of the solvent using a rotary evaporator with a water bath temperature of ~80 °C. The resulting crude oil was kept at −20 °C until needed for purification.

### 7.1.7 Pseudo-high dilution solution-phase cyclization

Pseudo-high dilution peptide cyclization conditions were used as described by Malesevic et al.[50] Fully deprotected peptide (113 μmol) and HATU (339 μmol) were dissolved separately in 11.3 mL each of dry DMF. Both solutions were added simultaneously using two syringe pumps at a rate of 0.01 mL/min each to a stirred solution of HATU (11.3 μmol) and DIPEA (678 μmol) in dry DMF (11.3 mL) over a period of ~20 h. After the addition was complete, the reaction was stirred for a further 30 min and the solvent was removed under vacuum. The resulting crude oil was kept at −20 °C until needed for purification.

### 7.1.8 Preparation of anhydrous methanolic HCl solution

Anhydrous HCl solution for methanolic resin cleavages was prepared by dropwise addition of acetyl chloride to a stirred solution of methanol which was cooled in an ice bath (final concentration 3 M). The solution was left to stir for 1 h before use.[76]

### 7.1.9 Allyl ester protecting group removal

Peptide-resin was swollen in DCM for 1 h and sparged with argon. $PhSiH_3$ and $Pd(PPh_3)_4$ were added in a small amount of DCM and the reaction was agitated under argon for 2 h at room temperature. The resin was then washed with DCM and DMF. The procedure was repeated as needed.[77]

### 7.1.10 Coupling of Hmb building block

Fmoc-(FmocHmb)Val-OH (3.0 equiv) was coupled manually overnight at room temperature (DIC/HOBt, 3.0 equiv each) in a minimum volume of DMF. The Fmoc group was removed using the conditions described. The symmetrical anhydride[214] of Fmoc-Leu-OH was preformed via addition of DIC (5.0 equiv) to a solution of Fmoc-protected leucine (10 equiv) in dry DCM (3 mL) that had been sparged with argon gas. The mixture was stirred at 0 °C in a flask fitted with a calcium chloride drying tube. After 20 min the anhydride solution was added directly to the resin, which had been washed with DCM to remove residual DMF. After 24 h the resin was drained and the coupling repeated.

### 7.1.11 Acetylation of Hmb group

Peptide-resin was first treated with piperidine solution followed by washing with 5 portions of DMF. The resin was then agitated with a solution of acetic anhydride (10 equiv) and DIPEA (5.0 equiv) (1:1 volume) in a minimum volume of DMF at RT.[146] After 30 min the solvent was drained and the resin washed with 5 further portions of DMF.

### 7.1.12 Liberation of Boc-amino acid from DCHA salt

Boc-Ser(O*t*Bu)-OH·DCHA was suspended in ice-cold ethyl acetate (30 mL/g) and 10% phosphoric acid was added until all the solid had dissolved. The aqueous layer was separated and the pH checked using pH paper (2–3). The organic layer was washed with 2 volume parts phosphoric acid (10%) and $3 \times 2$ volume parts water. The pH of the final aqueous wash should be 4. Complete conversion of the starting material was confirmed by TLC (1:1 hexane/ethyl acetate). The organic layer was dried ($MgSO_4$) and concentrated *in vacuo* to afford Boc-Ser(O*t*Bu)-OH as a colourless oil.[215]

### 7.1.13 Alkylation of sulfonamide linker

Peptide-resin was cyanomethylated via treatment with iodoacetonitrile (10 equiv) and DIPEA (10 equiv) in DMF (1 mL) for 24 h with agitation in a foil-wrapped reaction vessel.[216,217] The solvent was removed and the resin washed with DMF (3 × 5 mL), THF (3 × 5 mL) and DMF again. Alternatively, methylation was carried out by treatment of peptide-resin with 2.0 M trimethylsilyldiazomethane in hexane for 24 h. The resin was drained and washed with DCM and DMF. TMS-diazomethane residues were destroyed using base.

### 7.1.14 On-resin reduction of nitro linker

Peptide-resin was swollen for 1 h in DMF. The solvent was exchanged for DCM and a solution of $Na_2S_2O_4$ (20 equiv), $K_2CO_3$ (28 equiv) and TBAHS (2.0 equiv) in a 1:1 mixture of DCM/$H_2O$ was added to the resin (2 mL total).[202] After agitating for 9 h at room temperature, the solvent was removed and the resin washed (3 × $H_2O$/3 × DCM/3 × DMF).

## 7.2 Purification and characterization

### 7.2.1 Preparative high-performance liquid chromatography (HPLC)

Samples were dissolved in deionized water (with MeCN for insoluble peptides), centrifuged and injected in ~30 mg amounts (0.5–1 mL) per run onto a Speck and Burke Analytical C18 Column (5.0 μm, 10.0 × 250 mm) attached to a PerkinElmer Series 200 LC Pump and 785A UV/Vis Detector. A linear gradient of 0-50% B (solvent A = 5:95:0.1 MeCN/$H_2O$/TFA; B = 95:5:0.1 MeCN/$H_2O$/TFA) over 60 min, followed by 50-100% B over 15 min, with a flow rate of 2 mL/min was used for separations unless otherwise stated. Absorbance data were collected at 220 nm. For chaiyaphumines the gradient was started at 10% B and absorbance recorded at 280 nm. For lassomycin (and analogues) a gradient of 10-36% B over 40 min, followed by 36-100% B over 20 min was used. Selected fractions were lyophilized and a mass assigned using MALDI-TOF MS or ESI-MS, and peak fractions of interest were pooled and lyophilized.

### 7.2.2 Analytical chromatography (HPLC)

10 μL of sample solution was injected using a PerkinElmer Series 200 Autosampler onto a Waters XBridge BEH C18 Column (3.5 μm, 4.6 mm × 100 mm) attached to a PerkinElmer Series 200 LC Pump and Series 200 UV/Vis Detector. The column temperature was maintained at 40 °C using a PerkinElmer Series 200 Peltier Column Oven. A linear gradient of 0-100% B (solvent A = 5:95:0.05 MeCN/$H_2O$/TFA; B = 95:5:0.03 MeCN/$H_2O$/TFA) over 30 min with a flow rate of 1 mL/min was used for all analyses. Solvents were degassed using an in-line vacuum degasser. Absorbance data were collected at 220 nm and processed using the TotalChrom software (version 6.3.1).

### 7.2.3 Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)

MALDI-TOF and LIFT-TOF/TOF spectra were acquired on an Autoflex II ToF/ToF mass spectrometer (Bruker Daltonik GmBH) equipped with a 337 nm nitrogen laser. Peptides were dissolved in 1:1 deionized water/MeCN

(solvents containing 0.1% TFA or triethylamine were used for particularly insoluble analytes). Sample solution (1 mg/mL) was mixed with matrix solution ($\alpha$-cyano-4-hydroxy-cinnamic acid, ~50 mg/mL) in a ratio of 1:9, and 1 μL of the resulting solution spotted onto a metal target and placed into the MALDI ion source. Reflectron mode was used for molecules with $m/z < 4000$. MS data was processed using FlexAnalysis 2.0 (Bruker Daltonik GmBH) or MestReNova 10.0 (Mestrelab Research S.L.). MALDI MS/MS was performed using LIFT technology, which enables detection of product ions that result from elevating the laser power.

### 7.2.4 Analytical liquid chromatography mass spectrometry (LCMS)

Analytical LCMS was carried out using an Acquity UPLC system (Waters Ltd, UK) equipped with a photodiode array detector. Samples were injected onto an Acquity UPLC BEH C18 column (1.7 μm, 2.1 × 50 mm) and a gradient of 5–95% B (solvent A = $H_2O$, 0.1% formic acid; B = MeCN) was run over 3.8 min with a flow rate of 0.6 mL/min. The flow of solvent from the UPLC system was introduced into the electrospray ion source of an Aquity TQD or QToF Premier mass spectrometer, and positive ions were measured.

### 7.2.5 Ion-mobility (IM) mass spectrometry

Ion-mobility mass spectrometry was carried out using a Synapt G2S mass spectrometer equipped with an Acquity UPLC (Waters Ltd, UK). 1 μL of sample was injected onto a BEH C18 column (1.7 μm, 2.1 × 100 mm) and a gradient of 5-95% B (solvent A = $H_2O$, 0.1% formic acid; B = MeCN, 0.1% formic acid) was run over 6.0 min with a flow rate of 0.4 mL/min. A photodiode array detector provided absorbance data from 200 nm to 500 nm. The solvent flow was introduced into the electrospray ion source, and full scan MS carried out for positive ions. IM separation was performed with nitrogen as the gas. MS and photodiode array absorbance data was processed using MassLynx 4.1 (Waters Ltd, UK). IM results are represented as a plot of $m/z$ against drift time in Bins (mobiliogram).

### 7.2.6 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectra were recorded on a Bruker Avance 400 MHz or Varian VNMRS 700 MHz. Multiplicities: s = singlet, d = doublet, dd = doublet of doublets, ddd = doublet of doublet of doublets, dq = doublet of quartets, t = triplet, quint = quintet, m = multiplet. Chemical shifts are reported in $\delta$ units and are referenced to residual solvent peaks: $CHCl_3$ ($^1$H 7.26 ppm, $^{13}$C 77.0 ppm) and DMSO ($^1$H 2.50 ppm, $^{13}$C 39.5 ppm). Chaiyaphumine NMR spectra were recorded on a Varian VNMRS 600 spectrometer operating at a proton frequency of 599.7 MHz and a $^{13}$C-carbon frequency of 150.8 MHz. The instrument was equipped with a 5 mm OneNMR probe head. The experiments were carried out with compounds dissolved in 500 μL $d_6$-DMSO at 298 K. For structure elucidation and complete assignment of proton and carbon resonances $^1$H-$^1$H COSY, $^1$H-$^{13}$C multiplicity edited-HSQC, $^1$H-$^{13}$C HMBC and $^1$H-$^1$H NOESY (mixing time 500 ms) spectra were acquired.

## 7.3 Biological testing

Parasitic and mammalian cytotoxicity screening was carried out by Marcel Kaiser (Swiss Tropical and Public Health Institute). Testing against *Mycobacterium tuberculosis* was performed by Tim Bull (St. George's University of London).

### 7.3.1 Activity against *Trypanosoma brucei rhodesiense*

This stock was isolated in 1982 from a human patient in Tanzania and after several mouse passages cloned and adapted to axenic culture conditions.[218] Minimum Essential Medium (50 μL) supplemented with 25 mM HEPES, 1 g/L additional glucose, 1% MEM non-essential amino acids (100x), 0.2 mM 2-mercaptoethanol, 1 mM Na-pyruvate and 15% heat inactivated horse serum was added to each well of a 96-well microtiter plate. Serial drug dilutions of eleven 3-fold dilution steps covering a range from 100 to 0.002 μg/mL were prepared. $4 \times 10^3$ bloodstream forms of *T. b. rhodesiense* STIB 900 in 50 μL was added to each well and the plate incubated at 37 °C under a 5% $CO_2$ atmosphere for 70 h. 10 μL Alamar Blue (resazurin, 12.5 mg in 100 mL double-distilled water) was then added to each well and incubation continued for a further 2–4 h.[219] The plates were read with a Spectramax Gemini XS microplate fluorometer (Molecular Devices Cooperation, Sunnyvale, CA, USA) using an excitation wavelength of 536 nm and an emission wavelength of 588 nm. The $IC_{50}$ values were calculated by linear regression[220] from the sigmoidal dose inhibition curves using SoftmaxPro software (Molecular Devices Cooperation, Sunnyvale, CA, USA). Melarsoprol (Arsobal Sanofi-Aventis, received from WHO) is used as control.

### 7.3.2 Activity against *T. cruzi*

Rat skeletal myoblasts (L6 cells) were seeded in 96-well microtitre plates at 2000 cells/well in 100 μL Roswell Park Memorial Institute (RPMI) 1640 medium with 10% fetal bovine serum and 2 mM L-glutamine. After 24 h the medium was removed and replaced by 100 μL per well containing 5000 trypomastigote forms of *T. cruzi* Tulahuen strain C2C4 containing the $\beta$-galactosidase (Lac Z) gene.[221] After 48 h the medium was removed from the wells and replaced by 100 μL fresh medium with or without a serial

drug dilution of eleven 3-fold dilution steps covering a range from 100 to 0.002 μg/mL. After 96 h of incubation the plates were inspected under an inverted microscope to assure growth of the controls and sterility. Then the substrate chlorophenol red-β-D-galactopyranoside (CPRG)/Nonidet (50 μL) was added to all wells. A color reaction developed within 2–6 h and could be read photometrically at 540 nm. Data were analyzed with the graphic programme Softmax Pro (Molecular Devices), which calculated $IC_{50}$ values by linear regression[220] from the sigmoidal dose inhibition curves. Benznidazole is used as control ($IC_{50}$ 0.5 ± 0.2 μg/mL).

### 7.3.3 Activity against *L. donovani* axenic amastigotes

Amastigotes of *L. donovani* strain MHOM/ET/67/L82 were grown in axenic culture at 37 °C in SM medium[222] at pH 5.4 supplemented with 10% heat-inactivated fetal bovine serum under an atmosphere of 5% $CO_2$ in air. 100 μL of culture medium with $10^5$ amastigotes from axenic culture with or without a serial drug dilution were seeded in 96-well microtitre plates. Serial drug dilutions of eleven 3-fold dilution steps covering a range from 90 to 0.002 μg/mL were prepared. After 70 h of incubation the plates were inspected under an inverted microscope to assure growth of the controls and sterile conditions. 10 μL of Alamar Blue (12.5 mg resazurin dissolved in 100 mL distilled water)[223] were then added to each well and the plates incubated for another 2 h. Then the plates were read with a Spectramax Gemini XS microplate fluorometer (Molecular Devices Cooperation, Sunnyvale, CA, USA) using an excitation wavelength of 536 nm and an emission wavelength of 588 nm. Data were analyzed using the software Softmax Pro (Molecular Devices Cooperation, Sunnyvale, CA, USA). Decrease of fluorescence (= inhibition) was expressed as percentage of the fluorescence of control cultures and plotted against the drug concentrations. From the sigmoidal inhibition curves the $IC_{50}$ values were calculated.

### 7.3.4 Activity against *P. falciparum*

*In vitro* activity against erythrocytic stages of *P. falciparum* was determined using a $^3$H-hypoxanthine incorporation assay,[224,225] using the drug sensitive NF54 strain (Schipol Airport, The Netherlands[226]) or the chloroquine and pyrimethamine resistant K1 strain that originates from Thailand[227] and the

standard drug chloroquine (Sigma C6628). Compounds were dissolved in DMSO at 10 mg/mL and added to parasite cultures incubated in RPMI 1640 medium without hypoxanthine, supplemented with HEPES (5.94 g/L), $NaHCO_3$ (2.1 g/L), neomycin (100 U/mL), Albumax$^R$ (5 g/L) and washed human red cells $A^+$ at 2.5% haematocrit (0.3% parasitaemia). Serial drug dilutions of eleven 3-fold dilution steps covering a range from 100 to 0.002 μg/mL were prepared. The 96-well plates were incubated in a humidified atmosphere at 37 °C; 4% $CO_2$, 3% $O_2$, 93% $N_2$. After 48 h 50 μL of $^3$H-hypoxanthine (=0.5 μCi) was added to each well of the plate. The plates were incubated for a further 24 h under the same conditions. The plates were then harvested with a Betaplate$^{TM}$ cell harvester (Wallac, Zurich, Switzerland), and the red blood cells transferred onto a glass fibre filter then washed with distilled water. The dried filters were inserted into a plastic vial with 10 mL of scintillation fluid, and counted in a Betaplate$^{TM}$ liquid scintillation counter (Wallac, Zurich, Switzerland). $IC_{50}$ values were calculated from sigmoidal inhibition curves by linear regression[220] using Microsoft Excel. Chloroquine and artemisinin are used as control.

### 7.3.5 *In vitro* cytotoxicity with L6 cells

Assays were performed in 96-well microtiter plates, each well containing 100 μL of RPMI 1640 medium supplemented with 1% L-glutamine (200 mM) and 10% fetal bovine serum, and 4000 L6 cells (a primary cell line derived from rat skeletal myoblasts[228,229]). Serial drug dilutions of eleven 3-fold dilution steps covering a range from 100 to 0.002 μg/mL were prepared. After 70 h of incubation the plates were inspected under an inverted microscope to assure growth of the controls and sterile conditions. 10 μL of Alamar Blue was then added to each well and the plates incubated for another 2 h. Then the plates were read with a Spectramax Gemini XS microplate fluorometer (Molecular Devices Cooperation, Sunnyvale, CA, USA) using an excitation wavelength of 536 nm and an emission wavelength of 588 nm. The $IC_{50}$ values were calculated by linear regression[220] from the sigmoidal dose inhibition curves using SoftmaxPro software (Molecular Devices Cooperation, Sunnyvale, CA, USA). Podophyllotoxin (Sigma P4405) is used as control.

### 7.3.6 Activity against *Mycobacterium tuberculosis*

The MICs of the compounds were determined using standard liquid-broth and solid-agar based methods. Briefly, serial dilutions of each compound were made between 100 mg/L and 0.78 mg/L with MB7H9/oleic albumin dextrose catalase (OADC) (BD Biosciences) media in 96-well U-bottom plates. *Mycobacterium tuberculosis* H37Rv was grown to log phase ($OD_{600}$ = 1.0) and $10^4$ cells were added to all wells. Solid-agar based method was carried out as previously described[230] using a 48-well plate format. Compounds were tested in triplicate by adding serial dilutions into separate wells overlaid with MB7H11/OADC (BD Biosciences) agar based media. *Mycobacterium tuberculosis* culture (5 μL containing $10^4$ cells) was then inoculated into the centre of each well on top of the solidified media. Plates were sealed and incubated at 37 °C in 5% $CO_2$ for 10 days. Each plate also contained standardised serial dilution controls of first line tuberculosis drugs isoniazid (INH [Sigma, UK]; concentration range 8, 4, 2, 1, 0.5, 0.25, 0.125, 0.0625 μg/mL) and rifampicin (RIF [Sigma, UK]; concentration range 16, 8, 4, 2, 1, 0.5, 0.25, 0.125 μg/mL), in addition to negative (media only) controls. The MICs were determined as the first dilution to show complete growth inhibition as determined by a total lack of colony formation within the test period. It was not possible to obtain an authentic sample of naturally isolated lassomycin to compare in this assay because the *Lentzea kentuckyensis* subspecies from which lassomycin was originally prepared is the property of NovoBiotic Pharmaceuticals and was unavailable.

## 7.4 NMR structure calculation

$^{1}$H-$^{1}$H NOESY NMR data were processed separately using vnmrJ (version 4.2) and imported into CcpNmr Analysis[58] (version 2.4) for manual peak picking and assignment to a molecular system. Custom residue definitions were created using CcpNmr ChemBuild (version 1.0). Initial distance restraints were calculated from assigned peaks using the Analysis software and split into ambiguous and unambiguous restraint lists. Complete peak lists including unassigned peaks were also generated. The inbuilt graphical interface was used to export peak and restraint lists to ARIA[53] (version 2.3), and set up the run directory tree.

Structure refinement was carried out by ARIA using CNS[54,55] (version 1.2 at patch level 1) and a modified version of the PARALLHDG force field[56] (version 5.3). New force field topology definitions were added manually for D-amino acid residues, the phenylacetyl capping group and the ester linkage in `topallhdg5.3.pro`. Parameters for new bond length, angle, dihedral and improper torsion angle types were added to `parallhdg5.3.pro`. Bond lengths and angles for similar cyclic depsipeptides were taken from the literature[231–233] and dihedrals were generated using SwissParam.[234] Phenylacetyl group parameters were taken from HIC-Up.[235] Greater flexibility was introduced into the force field by dividing all force constants for bond angles and impropers by 10, as described by Mareuil et al.[236] Topology and parameter files were modified in both the run and ARIA distribution directories.

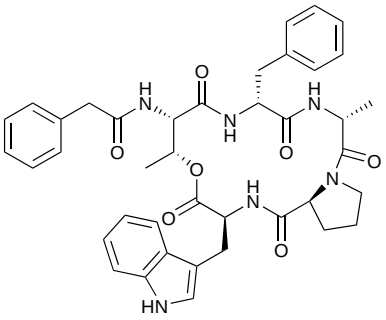
The CNS `generate.inp` script was modified to include the sequence (declared manually within the chain statement, incorporating nonstandard residues) and topology patch for the side-chain-to-tail ester bond. The linkage definitions in `topallhdg5.3.pep` were also modified to prevent terminal hydrogens being added to the capping residue, as an N-terminal amine group is not present in this case.

Simulated annealing was carried out using Cartesian dynamics only (torsion angle dynamics disabled) on an ensemble of 20 structures over 8 iterations each (without final refinement in explicit solvent). The Cartesian high temperature value was increased to 10 000 K (default value for torsion angle annealing) and the default simulation total time period was quadrupled (step length kept the same). The 10 lowest energy structures were chosen for the final ensemble.

# 7.5 Chaiyaphumines

## 7.5.1 Natural ester series

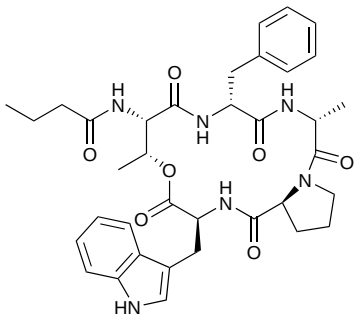
### Chaiyaphumine 10a



**10 a**

The linear sequence was synthesized as described in section 7.1.2 and capped with 2-phenylacetyl chloride (procedure 7.1.5), before undergoing resin cleavage and deprotection as described in procedure 7.1.3. Solution-phase cyclization of 224 mg of the linear precursor according to procedure 7.1.6, followed by HPLC purification (procedure 7.2.1) yielded **10a** (16 mg, 7%), HPLC $R_t = 20.3$ min (procedure 7.2.2); HRMS *m/z* (ESI) 721.3364, consistent with empirical formula $C_{40}H_{45}N_6O_7$ with an accuracy of 1.9 ppm (accepted as $[M + H]^+$).

### Chaiyaphumine 10b

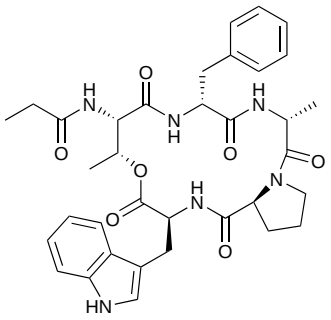


**10 b**

The linear sequence was synthesized as described in section 7.1.2 and capped with butyryl chloride (procedure 7.1.5). Resin cleavage was carried out using a 1% TFA solution in DCM (5 × 30 s).[237] The resulting combined filtrates were concentrated *in vacuo* and lyophilized. Global deprotection was carried out on the crude solid as described in procedure 7.1.3. A modification of solution-phase cyclization procedure 7.1.6 was carried out whereby the linear precursor was added under syringe pump control at a rate of 0.35 mL/h using EDCI as the activator (1.3 equiv) in the presence of DMAP (1.3 equiv) with HOBt (1.3 equiv).[50,238] The procedure was carried out on 66 mg of the linear precursor, followed by HPLC purification (procedure 7.2.1), to yield **10b** (5 mg, 8%), HPLC $R_t = 5.0$ min (procedure 7.2.2, 10 min gradient); HRMS *m/z* (ESI)

673.3353, consistent with empirical formula $C_{36}H_{45}N_6O_7$ with an accuracy of 0.4 ppm (accepted as $[M + H]^+$).

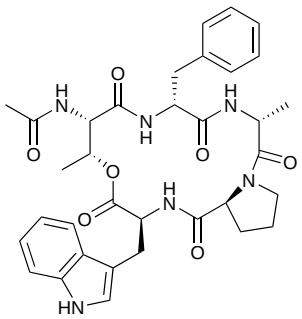
**Chaiyaphumine 10c**



**10 c**

The linear sequence was synthesized as described in section 7.1.2 and capped with propionyl chloride (procedure 7.1.5), before undergoing resin cleavage and deprotection as described in procedure 7.1.3. Solution-phase cyclization of 27 mg of the linear precursor according to procedure 7.1.6, followed by HPLC purification (procedure 7.2.1) yielded **10c** (2 mg, 8%), HPLC $R_t$ = 18.3 min (procedure 7.2.2); HRMS *m/z* (ESI) 659.3174, consistent with empirical formula $C_{35}H_{43}N_6O_7$ with an accuracy of 2.9 ppm (accepted as $[M + H]^+$).

**Chaiyaphumine 10d**

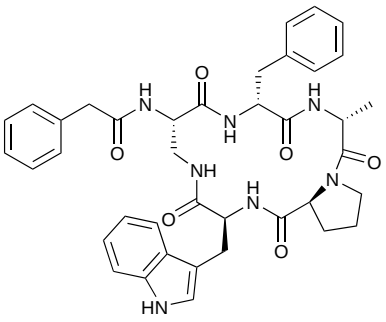


**10 d**

The linear sequence was synthesized as described in section 7.1.2 and capped with acetic anhydride (procedure 7.1.5), before undergoing resin cleavage and deprotection as described in procedure 7.1.3. Solution-phase cyclization of 10 mg of the linear precursor according to procedure 7.1.6 using HATU as the activator, followed by HPLC purification (procedure 7.2.1) yielded **10d** (2 mg, 21%), HPLC $R_t$ = 16.8 min (procedure 7.2.2); HRMS *m/z* (ESI) 645.3044, consistent with empirical formula $C_{34}H_{41}N_6O_7$ with an accuracy of 1.1 ppm (accepted as $[M + H]^+$).

## 7.5.2 Amide-linked analogues

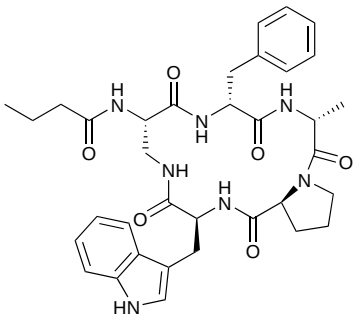
### Chaiyaphumine 11a



**11 a**

The linear sequence was synthesized as described in section 7.1.2 and capped with 2-phenylacetyl chloride (procedure 7.1.5), before undergoing resin cleavage and deprotection as described in procedure 7.1.3. Solution-phase cyclization of 18 mg of the linear precursor according to procedure 7.1.6, followed by HPLC purification (procedure 7.2.1) yielded **11a** (4 mg, 23%), HPLC $R_t$ = 16.3 min (procedure 7.2.2); HRMS *m/z* (ESI) 706.3355, consistent with empirical formula $C_{39}H_{44}N_7O_6$ with an accuracy of 0.3 ppm (accepted as $[M + H]^+$).

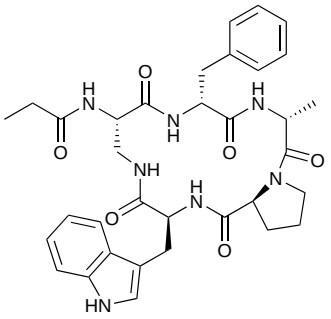
### Chaiyaphumine 11b



**11 b**

The linear sequence was synthesized as described in section 7.1.2 and capped with butyryl chloride (procedure 7.1.5), before undergoing resin cleavage and deprotection as described in procedure 7.1.3. Solution-phase cyclization of 12 mg of the linear precursor according to procedure 7.1.6, followed by HPLC purification (procedure 7.2.1) yielded **11b** (1 mg, 9%), HPLC $R_t$ = 16.3 min (procedure 7.2.2); HRMS *m/z* (ESI) 658.3358, consistent with empirical formula $C_{35}H_{44}N_7O_6$ with an accuracy of 0.8 ppm (accepted as $[M + H]^+$).

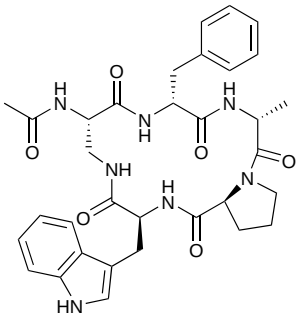
**Chaiyaphumine 11c**



**11 c**

The linear sequence was synthesized as described in section 7.1.2 and capped with propionyl chloride (procedure 7.1.5), before undergoing resin cleavage and deprotection as described in procedure 7.1.3. Solution-phase cyclization of 15 mg of the linear precursor according to procedure 7.1.6, followed by HPLC purification (procedure 7.2.1) yielded **11c** (1 mg, 7%), HPLC $R_t$ = 15.3 min (procedure 7.2.2); HRMS *m/z* (ESI) 644.3176, consistent with empirical formula $C_{34}H_{42}N_7O_6$ with an accuracy of 3.3 ppm (accepted as $[M + H]^+$).

**Chaiyaphumine 11d**

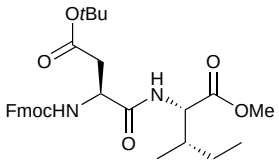


**11 d**

The linear sequence was synthesized as described in section 7.1.2 and capped with acetic anhydride (procedure 7.1.5), before undergoing resin cleavage and deprotection as described in procedure 7.1.3. Solution-phase cyclization of 23 mg of the linear precursor according to procedure 7.1.6, followed by HPLC purification (procedure 7.2.1) yielded **11d** (2 mg, 9%), HPLC $R_t$ = 14.1 min (procedure 7.2.2); HRMS *m/z* (ESI) 630.3034, consistent with empirical formula $C_{33}H_{40}N_7O_6$ with an accuracy of 1.0 ppm (accepted as $[M + H]^+$).

## 7.6 Lassomycin and lassomycin-amide

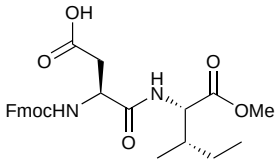
### Fmoc-Asp(O*t*Bu)-Ile-OMe (30)




**30**

Fmoc-Asp(O*t*Bu)-OH (1.14 g, 2.8 mmol), H-Ile-OMe (0.40 g, 2.8 mmol) and PyBOP® (1.44 g, 2.8 mmol) were suspended in DCM (30 mL) and stirred at room temperature. *N*-Methylmorpholine (0.91 mL, 8.3 mmol) was added and the reaction was stirred at room temperature for 18 h. The crude reaction mixture was evaporated under reduced pressure and purified via column chromatography ($SiO_2$; 98/2% hexane/EtOAc → 50/50% hexane/EtOAc) to yield Fmoc-Asp(O*t*Bu)-Ile-OMe (**30**) as a white powder (0.93 g, 63%), $\delta_H$ (700 MHz, $CDCl_3$) 0.88–0.93 (3H, m, Ile-δ$CH_3$), 0.88–0.93 (3H, m, Ile-γ$CH_3$), 1.15–1.23 (1H, m, Ile-γ$CH_2$), 1.38–1.44 (1H, m, Ile-γ$CH_2$), 1.46 (9H, s, O*t*Bu), 1.89–1.95 (1H, m, Ile-βCH), 2.64 (1H, dd, *J* 17.3 and 7.3 Hz, Asp-β$CH_2$), 2.93 (1H, dd, *J* 17.3 and 4.1 Hz, Asp-β$CH_2$), 3.71 (3H, s, $CO_2CH_3$), 4.23 (1H, t, *J* 7.1 Hz, Fmoc-CH), 4.41 (2H, d, *J* 7.1 Hz, Fmoc-$CH_2$), 4.53 (1H, dd, *J* 8.6 and 4.8 Hz, Ile-αCH), 4.58–4.63 (1H, m, Asp-αCH), 6.08 (1H, d, *J* 8.2 Hz Asp-NH), 7.16 (1H, d, *J* 8.6 Hz, Ile-NH), 7.31 (2H, t, *J* 7.5 Hz, FmocArH), 7.40 (2H, t, *J* 7.5 Hz, FmocArH), 7.59 (2H, d, *J* 7.5 Hz, FmocArH), 7.76 (2H, d, *J* 7.5 Hz, FmocArH); $\delta_C$ (176 MHz, $CDCl_3$) 11.6, 15.6, 25.1, 28.1, 37.6, 47.2, 50.9, 52.1, 56.9, 67.4, 82.0, 120.1, 125.1, 127.2, 127.8, 141.4, 143.7, 143.9, 156.0, 170.6, 171.6, 171.9; HRMS *m/z* (ESI) 539.2756, consistent with empirical formula $C_{30}H_{39}N_2O_7$ with an accuracy of 0.2 ppm (accepted as $[M + H]^+$).

### Fmoc-Asp(OH)-Ile-OMe (34)




**34**

Fmoc-Asp(O*t*Bu)-Ile-OMe (**30**, 0.33 g, 0.6 mmol) was treated with 20 mL of TFA solution (40% in DCM) for 30 min. Complete removal of the *t*-butyl group was confirmed using TLC. The reaction mixture was evaporated under reduced pressure to give Fmoc-Asp(OH)-Ile-OMe (**34**) as a white powder (quantitative yield, 0.29 g) which was used without further purification, $\delta_H$ (700 MHz, $CDCl_3$) 0.86–0.91 (3H, m, Ile-δ$CH_3$), 0.86–0.91 (3H, m, Ile-γ$CH_3$), 1.11–1.20 (1H, m, Ile-γ$CH_2$), 1.35–1.44 (1H, m,

Ile-$\gamma CH_2$), 1.85–1.96 (1H, m, Ile-$\beta$CH), 2.78 (1H, dd, *J* 17.4 and 6.4 Hz, Asp-$\beta CH_2$), 3.03 (1H, d, *J* 17.4 Hz, Asp-$\beta CH_2$), 3.72 (3H, s, $CO_2CH_3$), 4.22 (1H, t, *J* 6.6 Hz, Fmoc-CH), 4.43 (2H, d, *J* 6.6 Hz, Fmoc-$CH_2$), 4.53 (1H, dd, *J* 8.4 and 4.7 Hz, Ile-$\alpha$CH), 4.60–4.69 (1H, m, Asp-$\alpha$CH), 5.97 (1H, d, *J* 8.0 Hz, Asp-NH), 7.09 (1H, d, *J* 8.4 Hz, Ile-NH), 7.31 (2H, t, *J* 7.5 Hz, FmocArH), 7.40 (2H, t, *J* 7.5 Hz, FmocArH), 7.57 (2H, d, *J* 7.5 Hz, FmocArH), 7.76 (2H, d, *J* 7.5 Hz, FmocArH); HRMS *m/z* (ESI) 483.2115, consistent with empirical formula $C_{26}H_{31}N_2O_7$ with an accuracy of 3.3 ppm (accepted as $[M + H]^+$).

**Lassomycin (24)**

The dipeptide Fmoc-Asp(OH)-Ile-OMe (**34**) was coupled manually to Rink amide AM resin (0.77 mmol/g substitution, 0.10 mmol scale) using two 1 h couplings at RT. The sequence GLRRLFADQLVGRR was then built onto the resin-supported dipeptide via automated RT SPPS (as described in section 7.1.1). Resin cleavage (procedure 7.1.3) and pseudo-high dilution cyclization (procedure 7.1.7) were carried out as described, and the crude cyclization product was dissolved in 2.5 mL deionized water/acetonitrile, filtered and centrifuged. Purification by high-performance liquid chromatography (procedure 7.2.1) afforded lassomycin (26 mg, 14%), HPLC $R_t$ = 12.6 min (procedure 7.2.2); MALDI-TOF mass: found *m/z* 1881.1 $[M + H]^+$, calcd 1881.2 $[M + H]^+$; HRMS *m/z* (ESI) 940.55785, consistent with empirical formula $C_{83}H_{144}N_{30}O_{20}$ with an accuracy of 0.3 ppm (accepted as $[M + 2H]^{2+}$).

**Lassomycin-amide (35)**

Linear peptide GLRRLFADQLVGRRNI-amide was synthesized via automated SPPS at room temperature (section 7.1.1) on Rink amide AM resin (0.79 mmol/g substitution, 0.10 mmol scale). Resin cleavage (procedure 7.1.3), cyclization (procedure 7.1.7) and purification by high-performance liquid chromatography (procedure 7.2.1) were carried out as detailed above to afford lassomycin-amide (4 mg, 8%), HPLC $R_t$ = 13.0 min (procedure 7.2.2); MALDI-TOF mass: found *m/z* 1866.2 $[M + H]^+$, calcd 1866.2 $[M + H]^+$; HRMS *m/z* (ESI) 933.06000, consistent with empirical formula $C_{82}H_{143}N_{31}O_{19}$ with an accuracy of 1.8 ppm (accepted as $[M + 2H]^{2+}$).

**Lassomycin(1-9) (36)**

Linear peptide GLRRLFADQ-amide was synthesized via automated MW-SPPS (section 7.1.1) on Rink amide AM resin (0.79 mmol/g substitution, 0.10 mmol scale), using DIC as the activator. Resin cleavage (procedure 7.1.3), cyclization (procedure 7.1.6) and purification by high-performance liquid chromatography (procedure 7.2.1) afforded lassomycin(1-9) (31 mg, 29%), HPLC $R_t$ = 12.4 min (procedure 7.2.2); MALDI-TOF mass: found *m/z* 1056.6 $[M + H]^+$, calcd 1057.2 $[M + H]^+$; HRMS *m/z* (ESI) 1056.6082, consistent with empirical formula $C_{47}H_{78}N_{17}O_{11}$ with an accuracy of 1.4 ppm (accepted as $[M + H]^+$).

## 7.7 ACP fragments

**[A34C]ACP(34-77) (69)**

Sequence CDSLDTVELVMALEEEFDTEIPDEEAEKITTVQAAIDYINGHQA synthesized via automated SPPS at room temperature (section 7.1.1) on Fmoc-Ala-Wang resin LL (0.33 mmol/g substitution, 0.10 mmol scale). Double couplings were used for all residues. For the final 11 couplings the reaction volume was doubled (equivalents also doubled to maintain concentration). Resin cleavage and deprotection using TFA was carried out according to procedure 7.1.3 to afford fragment **69** (101 mg from half resin, 41%), MALDI-TOF mass: found *m/z* 4895.5 $[M + H]^+$, calcd 4895.3 $[M + H]^+$.

**ACP(1-33) *N*-acylsulfonamide (unalkylated) (71, cleaved)**

Synthesis of sequence STIEERVKKIIGEQLGVKQEEVTNNASFVEDLG carried out via automated SPPS at room temperature (section 7.1.1) on Fmoc-Gly-4-Sulfamylbutyryl Rink amide AM resin (0.26 mmol/g substitution, 0.10 mmol scale). Double couplings were used for all residues except Asn and Asp, which were triple coupled. For the final 13 couplings the reaction volume was doubled (equivalents also doubled to maintain concentration). Boc-Ser(O*t*Bu)-OH was liberated from the commercially available DCHA salt according to procedure 7.1.12 and coupled manually as the N-terminal amino acid under the conditions used for previous couplings. Resin test cleavage and deprotection using TFA was carried out according to procedure 7.1.3 to afford ACP(1-33) *N*-acylsulfonamide (unalkylated), MALDI-TOF mass: found *m/z* 3808.9 $[M + H]^+$, calcd 3807.9 $[M + H]^+$.

## 7.8 Nitro-functionalized linker

**Fmoc-FAAA-[76]-amide (81)**

Linker **76** (5.0 equiv) was coupled manually (section 7.1.2) to Rink amide AM resin (0.79 mmol/g substitution, 0.10 mmol scale) for 1 h using HATU (4.8 equiv) with DIPEA (10 equiv). Unreacted amine groups were capped according to the procedure given in section 7.1.2, and the Fmoc group removed using the conditions given previously. The sequence FAAA was then built on-resin via automated MW-SPPS (section 7.1.1) using double couplings. Resin cleavage and deprotection using TFA was carried out according to procedure 7.1.3 to afford **81**, MALDI-TOF mass: found *m/z* 786.3 $[M + Na]^+$, calcd 786.3 $[M + Na]^+$.

**Fmoc-FAAA-diaminobenzamide (82, cleaved)**

Resin-bound **81** (**80**) was reduced using $Na_2S_2O_4$ according to procedure 7.1.14. Resin cleavage and deprotection using TFA was carried out according to procedure 7.1.3 to afford Fmoc-FAAA-diaminobenzamide, MALDI-TOF mass: found *m/z* 756.4 $[M + Na]^+$, calcd 756.3 $[M + Na]^+$.

## 7.9 EBNA1 targeting peptides

### C-Ahx-RrRKGGYFMVF-amide (84)

Synthesized via automated MW-SPPS (section 7.1.1) on Rink amide AM resin (0.82 mmol/g substitution, 0.10 mmol scale), using DIC as the activator. A 1 h preactivation period was used for all residues except Ahx, which required 2 h. Double couplings were used for Arg and Cys residues, with extended Fmoc deprotection times for Arg (3 min MW + 20 min). Resin-bound peptide was sent directly to the collaborator for fluorophore conjugation and biological characterization. MALDI-TOF mass: found *m/z* 1632.0 $[M + H]^+$, calcd 1631.9 $[M + H]^+$.

### C-Ahx-YFMVFGGRrRK-amide (85)

Synthesized via automated MW-SPPS (section 7.1.1) on Rink amide AM resin (0.82 mmol/g substitution, 0.10 mmol scale), using DIC as the activator. A 1 h preactivation period was used for all residues except Ahx, which required 2 h. Double couplings were used for Arg and Cys residues, with extended Fmoc deprotection times for Arg (3 min MW + 20 min). Resin-bound peptide was sent directly to the collaborator for fluorophore conjugation and biological characterization. MALDI-TOF mass: found *m/z* 1631.7 $[M + H]^+$, calcd 1631.9 $[M + H]^+$.

# References


1. S. Lear, T. Munshi, A. S. Hudson, C. Hatton, J. Clardy, J. A. Mosely, T. J. Bull, C. S. Sit and S. L. Cobb, *Org. Biomol. Chem.*, 2016, **14**, 4534.
2. S. Lear and S. L. Cobb, *J. Comput. Aid. Mol. Des.*, 2016, **30**, 271.
3. C. -F. Chan, C. Xie, M. -K. Tsang, S. Lear, L. Dai, Y. Zhou, J. Cicho, M. Karbowiak, D. Hreniak, R. Lan, S. L. Cobb, M. H. -W. Lam, J. Hao and K. -L. Wong, *Eur. J. Inorg. Chem.*, 2015, 4539.
4. C. F. Chan, R. Lan, M. -K. Tsang, D. Zhou, S. Lear, W. -L. Chan, S. L. Cobb, W. -K. Wong, J. Hao, W. -T. Wong and K. -L. Wong, *J. Mater. Chem. B*, 2015, **3**, 2624.
5. H. Li, C. -F. Chan, W. -L. Chan, S. Lear, S. L. Cobb, N. -K. Mak, T. C. Lau, R. Lan, W. -K. Wong and K. -L. Wong, *Org. Biomol. Chem.*, 2014, **12**, 5876.
6. Jiang et al., *Nat. Commun.*, under review.
7. Mudge et al., *Nat. Plants*, under review.
8. S. Lear, S. M. Ramanoudjame, A. N. Crofton, A. M. Kenwright, M. Kaiser, H. B. Bode and S. L. Cobb, *ACS Infect. Dis.*, in preparation.
9. P. J. Hotez, D. H. Molyneux, A. Fenwick, J. Kumaresan, S. E. Sachs, J. D. Sachs and L. Savioli, *N. Engl. J. Med.*, 2007, **357**, 1018.
10. A. Cavalli and M. L. Bolognesi, *J. Med. Chem.*, 2009, **52**, 7339.
11. H. K. Majumder, W. de Souza and K. P. Chang, *Mol. Biol. Int.*, 2011, ID 185413.

12. P. Trouiller, P. Olliaro, E. Torreele, J. Orbinski, R. Laing and N. Ford, *Lancet*, 2002, **359**, 2188.

13. M. Reddy, S. S. Gill, S. R. Kalkar, W. Wu, P. J. Anderson and P. A. Rochon, *JAMA*, 2007, **298**, 1911.

14. J. R. Franco, P. P. Simarro, A. Diarra and J. G. Jannin, *Clin. Epidemiol.*, 2014, **6**, 257.

15. J. Clayton, *Nature*, 2010, **465**, S4.

16. A. Clem, *J. Global. Infect. Dis.*, 2010, **2**, 124.

17. R. Pink, A. Hudson, M. -A. Mouriès and M. Bendig, *Nat. Rev. Drug. Discov.*, 2005, **4**, 727.

18. S. R. Wilkinson, M. C. Taylor, D. Horn, J. M. Kelly and I. Cheeseman, *P. Natl. Acad. Sci. USA*, 2008, **105**, 5022.

19. A. M. Mejia, B. S. Hall, M. C. Taylor, A. Gómez-Palacio, S. R. Wilkinson, O. Triana-Chávez and J. M. Kelly, *J. Infect. Dis.*, 2012, **206**, 220.

20. M. C. O. Campos, L. L. Leon, M. C. Taylor and J. M. Kelly, *Mol. Biochem. Parasit.*, 2014, **193**, 17.

21. I. M. Vincent, D. Creek, D. G. Watson, M. A. Kamleh, D. J. Woods, P. E. Wong, R. J. S. Burchmore and M. P. Barrett, *PLoS Pathog.*, 2010, **6**, e1001204.

22. B. Manta, M. Comini, A. Medeiros, M. Hugo, M. Trujillo and R. Radi, *Biochim. Biophys. Acta*, 2013, **1830**, 3199.

23. S. Wyllie, B. J. Foth, A. Kelner, A. Y. Sokolova, M. Berriman and A. H. Fairlamb, *J. Antimicrob. Chemother.*, 2016, **71**, 625.

24. M. Balasegaram, S. Harris, F. Checchi, S. Ghorashian, C. Hamel and U. Karunakara, *B. World Health Organ.*, 2006, **84**, 783.

25. I. Kuepfer, C. Schmid, M. Allan, A. Edielu, E. P. Haary, A. Kakembo, S. Kibona, J. Blum and C. Burri, *PLoS Negl. Trop. Dis.*, 2012, **6**, e1695.

26. N. Baker, H. P. de Koning, P. Mäser and D. Horn, *Trends Parasitol.*, 2013, **29**, 110.

27. D. Mondal, J. Alvar, M. G. Hasnain, M. S. Hossain, D. Ghosh, M. M. Huda, S. G. Nabi, S. Sundar, G. Matlashewski and B. Arana, *Lancet Glob. Health*, 2014, **2**, e51.

28. J. Berman and R. Dietze, *Chemotherapy*, 1999, **45 (Suppl. 1)**, 54.

29. S. Sundar and P. L. Olliaro, *Ther. Clin. Risk Manage.*, 2007, **3**, 733.

30. B. Purkait, A. Kumar, N. Nandi, A. H. Sardar, S. Das, S. Kumar, K. Pandey, V. Ravidas, M. Kumar, T. De, D. Singh and P. Das, *Antimicrob. Agents Ch.*, 2012, **56**, 1031.

31. J. Mishra and S. Singh, *Exp. Parasitol.*, 2013, **135**, 397.

32. I. H. Gilbert, *J. Med. Chem.*, 2013, **56**, 7719.

33. H. P. Price, M. L. S. Güther, M. A. J. Ferguson and D. F. Smith, *Mol. Biochem. Parasit.*, 2010, **169**, 55.

34. S. Brand, L. A. T. Cleghorn, S. P. McElroy, D. A. Robinson, V. C. Smith, I. Hallyburton, J. R. Harrison, N. R. Norcross, D. Spinks, T. Bayliss, S. Norval, L. Stojanovski, L. S. Torrie, J. A. Frearson, R. Brenk, A. H. Fairlamb, M. A. J. Ferguson, K. D. Read, P. G. Wyatt and I. H. Gilbert, *J. Med. Chem.*, 2012, **55**, 140.

35. D. Spinks, V. Smith, S. Thompson, D. A. Robinson, T. Luksch, A. Smith, L. S. Torrie, S. McElroy, L. Stojanovski, S. Norval, I. T. Collie, I. Hallyburton, B. Rao, S. Brand, R. Brenk, J. A. Frearson, K. D. Read, P. G. Wyatt and I. H. Gilbert, *ChemMedChem*, 2015, **10**, 1821.

36. The PyMOL Molecular Graphics System, Version 1.2r1, Schrödinger, LLC.

37. P. J. Hotez, B. Pecoul, S. Rijal, C. Boehme, S. Aksoy, M. Malecela, R. Tapia-Conyer and J. C. Reeder, *PLoS Negl. Trop. Dis.*, 2016, **10**, e0003895.

38. S. Khare, A. S. Nagle, A. Biggart, Y. H. Lai, F. Liang, L. C. Davis, S. W. Barnes, C. J. N. Mathison, E. Myburgh, M. -Y. Gao, J. R. Gillespie, X. Liu, J. L. Tan, M. Stinson, I. C. Rivera, J. Ballard, V. Yeh, T. Groessl, G. Federe, H. X. Y. Koh, J. D. Venable, B. Bursulaya, M. Shapiro, P. K. Mishra, G.

Spraggon, A. Brock, J. C. Mottram, F. S. Buckner, S. P. S. Rao, B. G. Wen, J. R. Walker, T. Tuntland, V. Molteni, R. J. Glynne and F. Supek, *Nature*, 2016, DOI: 10.1038/nature19339.

39. A. Lewies, J. F. Wentzel, G. Jacobs and L. H. Du Plessis, *Molecules*, 2015, **20**, 15392.

40. B. M. Greenwood, K. Bojang, C. J. M. Whitty and G. A. T. Targett, *Lancet*, 2005, **365**, 1487.

41. D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2012, **75**, 311.

42. J. W. -H. Li and J. C. Vederas, *Science*, 2009, **325**, 161.

43. F. Grundmann, M. Kaiser, M. Schiell, A. Batzer, M. Kurz, A. Thanwisai, N. Chantratita and H. B. Bode, *J. Nat. Prod.*, 2014, **77**, 779.

44. N. Bionda, R. M. Fleeman, L. N. Shaw and P. Cudic, *ChemMedChem*, 2013, **8**, 1394.

45. A. Parenty, X. Moreau and J. -M. Campagne, *Chem. Rev.*, 2006, **106**, 911.

46. Q. Zhou, A. Dowling, H. Heide, J. Wöhnert, U. Brandt, J. Baum, R. ffrench-Constant and H. B. Bode, *J. Nat. Prod.*, 2012, **75**, 1717.

47. K. Fujii, T. Shimoya, Y. Ikai, H. Oka and K. Harada, *Tetrahedron Lett.*, 1998, **39**, 2579.

48. K. Fujii, Y. Ikai, T. Mayumi, H. Oka, M. Suzuki and K. Harada, *Anal. Chem.*, 1997, **69**, 3346.

49. Q. Zhou, F. Grundmann, M. Kaiser, M. Schiell, S. Gaudriault, A. Batzer, M. Kurz and H. B. Bode, *Chem. Eur. J.*, 2013, **19**, 16772.

50. M. Malesevic, U. Strijowski, D. Bächle and N. Sewald, *J. Biotechnol.*, 2004, **112**, 73.

51. K. Josephson, A. Ricardo and J. W. Szostak, *Drug Discov. Today*, 2014, **19**, 388.

52. J. Chatterjee, F. Rechenmacher and H. Kessler, *Angew. Chem. Int. Ed.*, 2013, **52**, 254.

53. W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, and M. Nilges, *Bioinformatics*, 2007, **23**, 381.

54. A. T. Brunger, *Nat. Protoc.*, 2007, **2**, 2728.

55. A. T. Brunger, P. D. Adams, G. M. Clore, P. Gros, R. W. Grosse-Kunstleve, J. -S. Jiang, J. Kuszewski, N. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren, *Acta Cryst.*, 1998, **D54**, 905.

56. J. P. Linge and M. Nilges, *J. Biomol. NMR*, 1999, **13**, 51.

57. E. G. Stein, L. M. Rice and A. T. Brünger, *J. Magn. Reson.*, 1997, **124**, 154.

58. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins Struct. Funct. Bioinf.*, 2005, **59**, 687.

59. F. Lindahl, H. N. Hoang, D. P. Fairlie and M. A. Cooper, *Chem. Commun.*, 2015, **51**, 4496.

60. B. Lomenick, R. W. Olsen and J. Huang, *ACS Chem. Biol.*, 2011, **6**, 34.

61. M. A. Fischbach and C. T. Walsh, *Science*, 2009, **325**, 1089.

62. S. E. Dorman and R. E. Chaisson, *Nat. Med.*, 2007, **13**, 295.

63. T. Kaeberlein, K. Lewis and S. S. Epstein, *Science*, 2002, **296**, 1127.

64. J. D. Hegemann, M. Zimmermann, X. Xie and M. A. Marahiel, *Acc. Chem. Res.*, 2015, **48**, 1909.

65. M. O. Maksimov, S. J. Pan and A. J. Link, *Nat. Prod. Rep.*, 2012, **29**, 996.

66. T. A. Knappe, U. Linne, S. Zirah, S. Rebuffat, X. Xie and M. A. Marahiel, *J. Am. Chem. Soc.*, 2008, **130**, 11446.

67. H. Nar, A. Schmid, C. Puder and O. Potterat, *ChemMedChem*, 2010, **5**, 1689.

68. M. Zimmermann, J. D. Hegemann, X. Xie and M. A. Marahiel, *Chem. Biol.*, 2013, **20**, 558.

69. T. A. Knappe, F. Manzenrieder, C. Mas-Moruno, U. Linne, F. Sasse, H. Kessler, X. Xie and M. A. Marahiel, *Angew. Chem. Int. Ed.*, 2011, **50**, 8714.

70. J. D. Hegemann, C. D. Fage, S. Zhu, K. Harms, F. S. Di Leva, E. Novellino, L. Marinelli and M. A. Marahiel, *Mol. BioSyst.*, 2016, **12**, 1106.

71. J. D. Hegemann, M. Zimmermann, S. Zhu, H. Steuber, K. Harms, X. Xie and M. A. Marahiel, *Angew. Chem. Int. Ed.*, 2014, **53**, 2230.

72. I. Mathavan, S. Zirah, S. Mehmood, H. G. Choudhury, C. Goulard, Y. Li, C. V. Robinson, S. Rebuffat and K. Beis, *Nat. Chem. Biol.*, 2014, **10**, 340.

73. C. Clavel, K. Fournel-Marotte and F. Coutrot, *Molecules*, 2013, **18**, 11553.

74. E. Gavrish, C. S. Sit, S. Cao, O. Kandror, A. Spoering, A. Peoples, L. Ling, A. Fetterman, D. Hughes, A. Bissell, H. Torrey, T. Akopian, A. Mueller, S. Epstein, A. Goldberg, J. Clardy and K. Lewis, *Chem. Biol.*, 2014, **21**, 509.

75. T. Akopian, O. Kandror, R. M. Raju, M. Unnikrishnan, E. J. Rubin and A. L. Goldberg, *EMBO J.*, 2012, **31**, 1529.

76. R. A. Turner, R. J. Weber and R. S. Lokey, *Org. Lett.*, 2010, **12**, 1852.

77. S. K. Mahto, C. J. Howard, J. C. Shimko and J. J. Ottesen, *ChemBioChem*, 2011, **12**, 2488.

78. T. Walther, S. Renner, H. Waldmann and H. -D. Arndt, *ChemBioChem*, 2009, **10**, 1153.

79. S. Friedrich-Bochnitschek, H. Waldmann and H. Kunz, *J. Org. Chem.*, 1989, **54**, 751.

80. S. M. Ocampo, F. Albericio, I. Fernández, M. Vilaseca and R. Eritja, *Org. Lett.*, 2005, **7**, 4349.

81. M. Mergler, F. Dick, B. Sax, P. Weiler and T. Vorherr, *J. Peptide Sci.*, 2003, **9**, 36.

82. C. Yue, J. Thierry and P. Potier, *Tetrahedron Lett.*, 1993, **34**, 323.

83. A. G. Russell, M. -E. Ragoussi, R. Ramalho, C. W. Wharton, D. Carteau, D. M. Bassani and J. S. Snaith, *J. Org. Chem.*, 2010, **75**, 4648.

84. R. S. Givens, J. F. W. Weber, P. G. Conrad, G. Orosz, S. L. Donahue and S. A. Thayer, *J. Am. Chem. Soc.*, 2000, **122**, 2687.

85. B. Mothia, A. N. Appleyard, S. Wadman and A. B. Tabor, *Org. Lett.*, 2011, **13**, 4216.

86. C. K. Marlowe, *Bioorg. Med. Chem. Lett.*, 1993, **3**, 437.

87. C. J. White and A. K. Yudin, *Nature Chem.*, 2011, **3**, 509.

88. D. Suckau, A. Resemann, M. Schuerenberg, P. Hufnagel, J. Franzen and A. Holle, *Anal. Bioanal. Chem.*, 2003, **376**, 952.

89. P. Roepstorff and J. Fohlman, *Biol. Mass Spectrom.*, 1984, **11**, 601.

90. F. Lanucara, S. W. Holman, C. J. Gray and C. E. Eyers, *Nat. Chem.*, 2014, **6**, 281.

91. K. Jeanne Dit Fouque, C. Afonso, S. Zirah, J. D. Hegemann, M. Zimmermann, M. A. Marahiel, S. Rebuffat and H. Lavanant, *Anal. Chem.*, 2015, **87**, 1166.

92. J. D. Hegemann, M. Zimmermann, X. Xie and M. A. Marahiel, *J. Am. Chem. Soc.*, 2013, **135**, 210.

93. M. O. Maksimov and A. J. Link, *J. Am. Chem. Soc.*, 2013, **135**, 12038.

94. E. Midforth and E. Pohl, unpublished work.

95. M. Goodman, W. Cai and N. D. Smith, *J. Peptide Sci.*, 2003, **9**, 594.

96. G. R. Marshall, *J. Peptide Sci.*, 2003, **9**, 534.

97. M. A. Roggero, B. Filippi, P. Church, S. L. Hoffman, U. Blum-Tirouvanziam, J. A. Lopez, F. Esposito, H. Matile, C. D. Reymond, N. Fasel and G. Corradin, *Mol. Immunol.*, 1995, **32**, 1301.

98. F. El Oualid, R. Merkx, R. Ekkebus, D. S. Hameed, J. J. Smit, A. de Jong, H. Hilkmann, T. K. Sixma and H. Ovaa, *Angew. Chem. Int. Ed.*, 2010, **49**, 10149.

99. S. Frutos, J. Tulla-Puche, F. Albericio and E. Giralt, *Int. J. Pept. Res. Ther.*, 2007, **13**, 221.

100. F. García-Martín, P. White, R. Steinauer, S. Côté, J. Tulla-Puche and F. Albericio, *Biopolymers*, 2006, **84**, 566.

101. B. Bacsa, K. Horváti, S. Bõsze, F. Andrea and C. O. Kappe, *J. Org. Chem.*, 2008, **73**, 7532.

102. F. García-Martín, M. Quintanar-Audelo, Y. García-Ramos, L. J. Cruz, C. Gravel, R. Furic, S. Côté, J. Tulla-Puche and F. Albericio, *J. Comb. Chem.*, 2006, **8**, 213.

103. B. L. Nilsson, M. B. Soellner and R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 91.

104. S. B. H. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338.

105. T. L. R. Grygiel, A. Teplyakov, G. Obmolova, N. Stowell, R. Holland, J. F. Nemeth, S. C. Pomerantz, M. Kruszynski and G. L. Gilliland, *Biopolymers*, 2010, **94**, 350.

106. V. Y. Torbeev and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2007, **46**, 1667.

107. D. Bang, B. L. Pentelute and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2006, **45**, 3985.

108. D. Bang and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2004, **43**, 2534.

109. D. Bang, N. Chopra and S. B. H. Kent, *J. Am. Chem. Soc.*, 2004, **126**, 1377.

110. S. Bondalapati, M. Jbara and A. Brik, *Nat. Chem.*, 2016, **8**, 407.

111. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, in *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, New York, 2nd edn., 2009.

112. H. Nothaft and C. M. Szymanski, *J. Biol. Chem.*, 2013, **288**, 6912.

113. F. Schwarz, W. Huang, C. Li, B. L. Schulz, C. Lizak, A. Palumbo, S. Numao, D. Neri, M. Aebi and L. -X. Wang, *Nat. Chem. Biol.*, 2010, **6**, 264.

114. D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131.

115. S. Lear, MSc(Res) thesis, Durham University, 2013.

116. M. Sainlos and B. Imperiali, *Nat. Protoc.*, 2007, **2**, 3201.

117. S. L. Cobb and C. D. Murphy, *J. Fluorine Chem.*, 2009, **130**, 132.

118. P. W. R. Harris, G. M. Williams, P. Shepherd and M. A. Brimble, *Int. J. Pept. Res. Ther.*, 2008, **14**, 387.

119. S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem. Int. Ed.*, 2005, **44**, 1650.

120. R. Behrendt, P. White and J. Offer, *J. Pept. Sci.*, 2016, **22**, 4.

121. P. Wang, S. Dong, J. A. Brailsford, K. Iyer, S. D. Townsend, Q. Zhang, R. C. Hendrickson, J. Shieh, M. A. S. Moore and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, 2012, **51**, 11576.

122. R. S. Rush, P. L. Derby, D. M. Smith, C. Merry, G. Rogers, M. F. Rohde and V. Katta, *Anal. Chem.*, 1995, **67**, 1442.

123. M. Chilver-Vaughan, dissertation, Durham University, 2016.

124. D. Macmillan, R. M. Bill, K. A. Sage, D. Fern and S. L. Flitsch, *Chem. Biol.*, 2001, **8**, 133.

125. K. Hirano, D. Macmillan, K. Tezuka, T. Tsuji and Y. Kajihara, *Angew. Chem. Int. Ed.*, 2009, **48**, 9557.

126. G. G. Kochendoerfer, S. Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent and J. A. Bradburne, *Science*, 2003, **299**, 884.

127. S. Liu, B. L. Pentelute and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2012, **51**, 993.

128. M. Murakami, R. Okamoto, M. Izumi and Y. Kajihara, *Angew. Chem. Int. Ed.*, 2012, **51**, 3567.

129. S. Shang, Z. Tan, S. Dong and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2011, **133**, 10784.

130. Q. Wan and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, 2007, **46**, 9248.

131. A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J. J. Prestegard, R. L. Schnaar, H. H. Freeze, J. D. Marth, C. R. Bertozzi, M. E. Etzler, M. Frank, J. F. G. Vliegenthart, T. Lütteke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K. F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi and S. Kornfeld, *Glycobiology*, 2015, **25**, 1323.

132. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, J. D. Marth, C. R. Bertozzi, G. W. Hart and M. E. Etzler, *Proteomics*, 2009, **9**, 5398.

133. G. Chen, J. D. Warren, J. Chen, B. Wu, Q. Wan and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2006, **128**, 7460.

134. P. Wang, B. Aussedat, Y. Vohra and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, 2012, **51**, 11571.

135. L. Spasser and A. Brik, *Angew. Chem. Int. Ed.*, 2012, **51**, 6840.

136. M. Haj-Yahya, B. Fauvet, Y. Herman-Bachinsky, M. Hejjaoui, S. N. Bavikar, S. V. Karthikeyan, A. Ciechanover, H. A. Lashuel and A. Brik, *P. Natl. Acad. Sci. USA*, 2013, **110**, 17726.

137. C. M. Pickart and D. Fushman, *Curr. Opin. Chem. Biol.*, 2004, **8**, 610.

138. L. A. Erlich, K. S. Ajish Kumar, M. Haj-Yahya, P. E. Dawson and A. Brik, *Org. Biomol. Chem.*, 2010, **8**, 2392.

139. D. Bang, V. Tereshko, A. A. Kossiakoff and S. B. H. Kent, *Mol. BioSyst.*, 2009, **5**, 750.

140. K. Mandal, B. L. Pentelute, D. Bang, Z. P. Gates, V. Y. Torbeev and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2012, **51**, 1481.

141. T. O. Yeates and S. B. H. Kent, *Annu. Rev. Biophys.*, 2012, **41**, 41.

142. F. Mende and O. Seitz, *Angew. Chem. Int. Ed.*, 2011, **50**, 1232.

143. X. Bu, G. Xie, C. W. Law and Z. Guo, *Tetrahedron Lett.*, 2002, **43**, 2419.

144. X. Li, T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, 1998, **39**, 8669.

145. J. Kang, J. P. Richardson and D. Macmillan, *Chem. Commun.*, 2009, 407.

146. F. Burlina, G. Papageorgiou, C. Morris, P. D. White and J. Offer, *Chem. Sci.*, 2014, **5**, 766.

147. B. J. Backes and J. A. Ellman, *J. Org. Chem.*, 1999, **64**, 2322.

148. R. Ingenito, E. Bianchi, D. Fattori and A. Pessi, *J. Am. Chem. Soc.*, 1999, **121**, 11369.

149. P. Heidler and A. Link, *Bioorg. Med. Chem.*, 2005, **13**, 585.

150. J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem. Int. Ed.*, 2008, **47**, 6851.

151. R. Quaderer and D. Hilvert, *Org. Lett.*, 2001, **3**, 3181.

152. F. Burlina, C. Morris, R. Behrendt, P. White and J. Offer, *Chem. Commun.*, 2012, **48**, 2579.

153. S. Gunasekera, T. L. Aboye, W. A. Madian, H. R. El-Seedi and U. Göransson, *Int. J. Pept. Res. Ther.*, 2013, **19**, 43.

154. G. -M. Fang, Y. -M. Li, F. Shen, Y. -C. Huang, J. -B. Li, Y. Lin, H. -K. Cui and L. Liu, *Angew. Chem. Int. Ed.*, 2011, **50**, 7645.

155. J. -S. Zheng, S. Tang, Y. -K. Qi, Z. -P. Wang and L. Liu, *Nat. Protoc.*, 2013, **8**, 2483.

156. G. -M. Fang, J. -X. Wang and L. Liu, *Angew. Chem. Int. Ed.*, 2012, **51**, 10347.

157. C. Bello, F. Kikul and C. F. W. Becker, *J. Pept. Sci.*, 2015, **21**, 201.

158. W. Hou, X. Zhang, F. Li and C. -F. Liu, *Org. Lett.*, 2011, **13**, 386.

159. N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain and O. Melnyk, *Org. Lett.*, 2010, **12**, 5238.

160. L. Raibaut, M. Cargoët, N. Ollivier, Y. M. Chang, H. Drobecq, E. Boll, R. Desmet, J. -C. M. Monbaliu and O. Melnyk, *Chem. Sci.*, 2016, **7**, 2657.

161. L. Raibaut, H. Drobecq and O. Melnyk, *Org. Lett.*, 2015, **17**, 3636.

162. N. Ollivier, J. Vicogne, A. Vallin, H. Drobecq, R. Desmet, O. El Mahdi, B. Leclercq, G. Goormachtigh, V. Fafeur and O. Melnyk, *Angew. Chem. Int. Ed.*, 2012, **51**, 209.

163. T. S. Young and P. G. Schultz, *J. Biol. Chem.*, 2010, **285**, 11039.

164. A. C. Mercer and M. D. Burkart, *Nat. Prod. Rep.*, 2007, **24**, 750.

165. K. J. Weissman and R. Müller, *ChemBioChem*, 2008, **9**, 826.

166. S. Lautru, R. J. Deeth, L. M. Bailey and G. L. Challis, *Nat. Chem. Biol.*, 2005, **1**, 265.

167. M. Hans, A. Hornung, A. Dziarnowski, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2003, **125**, 5366.

168. C. Khosla, S. Kapur and D. E. Cane, *Curr. Opin. Chem. Biol.*, 2009, **13**, 135.

169. C. Khosla, *J. Org. Chem.*, 2009, **74**, 6416.

170. J. E. Cronan, *Annu. Rev. Microbiol.*, 2003, **57**, 203.

171. J. P. Torella, T. J. Ford, S. N. Kim, A. M. Chen, J. C. Way and P. A. Silver, *P. Natl. Acad. Sci. USA*, 2013, **110**, 11290.

172. P. Handke, S. A. Lynch and R. T. Gill, *Metab. Eng.*, 2011, **13**, 28.

173. S. W. White, J. Zheng, Y. -M. Zhang and C. O. Rock, *Annu. Rev. Biochem.*, 2005, **74**, 791.

174. B. -N. Wu, Y. -M. Zhang, C. O. Rock and J. J. Zheng, *Protein Sci.*, 2009, **18**, 240.

175. R. Leonardi, Y. -M. Zhang, C. O. Rock and S. Jackowski, *Prog. Lipid Res.*, 2005, **44**, 125.

176. G. A. Zornetzer, B. G. Fox and J. L. Markley, *Biochemistry*, 2006, **45**, 5217.

177. R. W. Haushalter, F. V. Filipp, K. -s. Ko, R. Yu, S. J. Opella and M. D. Burkart, *ACS Chem. Biol.*, 2011, **6**, 413.

178. A. S. Worthington and M. D. Burkart, *Org. Biomol. Chem.*, 2006, **4**, 44.

179. W. S. Hancock, G. R. Marshall and P. R. Vagelos, *J. Biol. Chem.*, 1973, **248**, 2424.

180. W. S. Hancock, D. J. Prescott, G. R. Marshall and P. R. Vagelos, *J. Biol. Chem.*, 1972, **247**, 6224.

181. P. R. Vagelos, W. S. Hancock, D. J. Prescott, W. L. Nulty, J. Weintraub and G. R. Marshall, *J. Am. Chem. Soc.*, 1971, **93**, 1799.

182. E. C. B. Johnson and S. B. H. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640.

183. M. Amblard, J. -A. Fehrentz, J. Martinez and G. Subra, *Mol. Biotechnol.*, 2006, **33**, 239.

184. F. Guillier, D. Orain and M. Bradley, *Chem. Rev.*, 2000, **100**, 2091.

185. C. A. Chantell, M. A. Onaiyekan and M. Menakuru, *J. Pept. Sci.*, 2012, **18**, 88.

186. I. M. Krishna Kumar, V. N. Rajasekharan Pillai and B. Mahew, *J. Pept. Sci.*, 2002, **8**, 183.

187. C. Hyde, T. Johnson and R. C. Sheppard, *J. Chem. Soc., Chem. Commun.*, 1992, 1573.

188. C. A. Hood, G. Fuentes, H. Patel, K. Page, M. Menakuru and J. H. Park, *J. Pept. Sci.*, 2008, **14**, 97.

189. P. White, J. W. Keyte, K. Bailey and G. Bloomberg, *J. Peptide Sci.*, 2004, **10**, 18.

190. M. Quibell, W. G. Turnell and T. Johnson, *J. Chem. Soc., Perkin Trans. 1*, 1995, 2019.

191. M. Quibell, W. G. Turnell and T. Johnson, *J. Org. Chem.*, 1994, **59**, 1745.

192. W. R. Sampson, H. Patsiouras and N. J. Ede, *J. Peptide Sci.*, 1999, **5**, 403.

193. E. Nicolás, M. Pujades, J. Bacardit, E. Giralt and F. Albericio, *Tetrahedron Lett.*, 1997, **38**, 2317.

194. N. J. Ede, K. H. Ang, I. W. James and A. M. Bray, *Tetrahedron Lett.*, 1996, **37**, 9097.

195. M. Quibell, W. G. Turnell and T. Johnson, *Tetrahedron Lett.*, 1994, **34**, 2237.

196. A. -B. M. Abdel-Aal, G. Papageorgiou, M. Quibell and J. Offer, *Chem. Commun.*, 2014, **50**, 8316.

197. J. M. Collins and N. E. Leadbeater, *Org. Biomol. Chem.*, 2007, **5**, 1141.

198. A. -B. M. Abdel-Aal, G. Papageorgiou, R. Raz, M. Quibell, F. Burlina and J. Offer, *J. Pept. Sci.*, 2016, **22**, 360.

199. B. J. Backes, A. A. Virgilio and J. A. Ellman, *J. Am. Chem. Soc.*, 1996, **118**, 3055.

200. G. W. Kenner, J. R. McDermott and R. C. Sheppard, *J. Chem. Soc. D*, 1971, 636.

201. N. G. Murphy, S. M. Varney, J. M. Tallon, J. R. Thompson and P. D. Blanc, *Clin. Toxicol.*, 2009, **47**, 702.

202. R. Kaplánek and V. Krchňák, *Tetrahedron Lett.*, 2013, **54**, 2600.

203. R. A. Scheuerman and D. Tumelty, *Tetrahedron Lett.*, 2000, **41**, 6531.

204. *Novabiochem Letters*, 2012, 2, **12**, EMD Millipore.

205. *Novabiochem Innovations*, 2012, 1, **12**, EMD Millipore.

206. A. Hari and B. L. Miller, *Tetrahedron Lett.*, 1999, **40**, 245.

207. J. B. Blanco-Canosa, B. Nardone, F. Albericio and P. E. Dawson, *J. Am. Chem. Soc.*, 2015, **137**, 7197.

208. E. K. Kainmüller and W. Bannwarth, *Helv. Chim. Acta.*, 2006, **89**, 3056.

209. J. L. Meier and M. D. Burkart, *Chem. Soc. Rev.*, 2009, **38**, 2012.

210. R. J. Foster, A. J. Poulose, R. F. Bonsall and P. E. Kolattukudy, *J. Biol. Chem.*, 1985, **260**, 2826.

211. T. Plass, S. Milles, C. Koehler, C. Schultz and E. A. Lemke, *Angew. Chem. Int. Ed.*, 2011, **50**, 3878.

212. J. I. Cohen, *N. Engl. J. Med.*, 2000, **343**, 481.

213. L. S. Young and C. W. Dawson, *Chin. J. Cancer*, 2014, **33**, 581.

214. R. Hancock, M. Schaap, H. Pfister and G. Wells, *Org. Biomol. Chem.*, 2013, **11**, 3553.

215. Bachem Technical Note: Conversion of a DCHA salt to the Free Acid, http://www.bachem.com/service-support/technical-library/technical-notes/, (accessed May 2013).

216. J. -w. Park and K. -H. Lee, *Bull. Korean Chem. Soc.*, 2009, **30**, 2475.

217. Q. Li, M. Moutiez, J. -B. Charbonnier, K. Vaudry, A. Ménez, E. Quéméneur and C. Dugave, *J. Med. Chem.*, 2000, **43**, 1770.

218. T. Baltz, D. Baltz, C. Giroud and J. Crockett, *EMBO J.*, 1985, **4**, 1273.

219. B. Räz, M. Iten, Y. Grether-Bühler, R. Kaminsky and R. Brun, *Acta Trop.*, 1997, **68**, 139.

220. W. Huber and J. C. Koella, *Acta Trop.*, 1993, **55**, 257.

221. F. S. Buckner, C. L. Verlinde, A. C. La Flamme and W. C. Van Voorhis, *Antimicrob. Agents Chemother.*, 1996, **40**, 2592.

222. I. Cunningham, *J. Protozool.*, 1977, **24**, 325.

223. J. Mikus and D. Steverding, *Parasitol. Int.*, 2000, **48**, 265.

224. R. E. Desjardins, C. J. Canfield, J. D. Haynes and J. D. Chulay, *Antimicrob. Agents Chemother.*, 1979, **16**, 710.

225. H. Matile and J. R. Pink, L. Plasmodium falciparum Malaria Parasite Cultures and Their Use in Immunology. In: I. Lefkovits and B. Pernis (eds.) *Immunological Methods*. San Diego, CA: Academic Press; 1990. p. 221–234.

226. T. Ponnudurai, A. D. Leeuwenberg and J. H. Meuwissen, *Trop. Geogr. Med.*, 1981, **33**, 50.

227. S. Thaithong, G. H. Beale and M. Chutmongkonkul, *Trans. R. Soc. Trop. Med. Hyg.*, 1983, **77**, 228.

228. B. Page, M. Page and C. Noel, *Int. J. Oncol.*, 1993, **3**, 473.

229. S. Ansar Ahmed, R. M. Gogal and J. E. Walsh, *J. Immunol. Methods.*, 1994, **170**, 211.

230. D. Evangelopoulos and S. Bhakta, *Methods Mol. Biol.*, 2010, **642**, 193.

231. A. Stamm, D. Bernhard and M. Gerhards, *Phys. Chem. Chem. Phys.*, 2016, **18**, 15327.

232. U. M. Reinscheid, J. Farjon, M. Radzom, P. Haberz, A. Zeeck, M. Blackledge and C. Griesinger, *ChemBioChem*, 2006, **7**, 287.

233. J. N. Tabudravu, L. A. Morris, B. F. Milne and M. Jaspars, *Org. Biomol. Chem.*, 2005, **3**, 745.

234. V. Zoete, M. A. Cuendet, A. Grosdidier and O. Michielin, *J. Comput. Chem.*, 2011, **32**, 2359.

235. G. J. Kleywegt, *Acta Cryst.*, 2007, **D63**, 94.

236. F. Mareuil, T. E. Malliavin, M. Nilges and B. Bardiaux, *J. Biomol. NMR*, 2015, **62**, 425.

237. F. García-Martín, N. Bayó-Puxan, L. J. Cruz, J. C. Bohling and F. Albericio, *QSAR Comb. Sci.*, 2007, **26**, 1027.

238. Y. Zhang, R. Boyer, X. Sun, J. Paschal and S. -H. Chen, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 775.

239. J. S. I. Ingram and J. R. Porter, *Nat. Plants*, 2015, **1**, 15173.

240. S. Smith and I. De Smet, *Phil. Trans. R. Soc. B*, 2012, **367**, 1441.

241. Y. Uga, K. Sugimoto, S. Ogawa, J. Rane, M. Ishitani, N. Hara, Y. Kitomi, Y. Inukai, K. Ono, N. Kanno, H. Inoue, H. Takehisa, R. Motoyama, Y. Nagamura, J. Wu, T. Matsumoto, T. Takai, K. Okuno and M. Yano, *Nat. Genet.*, 2013, **45**, 1097.

242. A. Karaba, S. Dixit, R. Greco, A. Aharoni, K. R. Trijatmiko, N. Marsch-Martinez, A. Krishnan, K. N. Nataraja, M. Udayakumar and A. Pereira, *P. Natl. Acad. Sci. USA*, 2007, **104**, 15270.

243. K. Růžička, K. Ljung, S. Vanneste, R. Podhorská, T. Beeckman, J. Friml and E. Benková, *Plant Cell*, 2007, **19**, 2197.

244. C. Ju and C. Chang, *AoB PLANTS*, 2012, pls031.

245. S. N. Shakeel, Z. Gao, M. Amir, Y. -F. Chen, M. I. Rai, N. Ul Haq and G. E. Schaller, *J. Biol. Chem.*, 2015, **290**, 12415.

246. B. M. Binder, F. I. Rodríguez and A. B. Bleecker, *J. Biol. Chem.*, 2010, **285**, 37263.

247. A. J. Mudge, PhD thesis, Durham University, 2015.

248. R. F. Lacey and B. M. Binder, *J. Inorg. Biochem.*, 2014, **133**, 58.

249. P. M. Chilley, S. A. Casson, P. Tarkowski, N. Hawkins, K. L. -C. Wang, P. J. Hussey, M. Beale, J. R. Ecker, G. K. Sandberg and K. Lindsey, *Plant Cell*, 2006, **18**, 3058.

# Part V

# Appendices

# Appendix A

# Chaiyaphumine NMR spectra

The following NMR characterization is included for chaiyaphumine **10a** (Chapter 2)



NOE-derived distance restraints for the final molecular dynamics energy minimization run are also included (Table A.1, page 180).

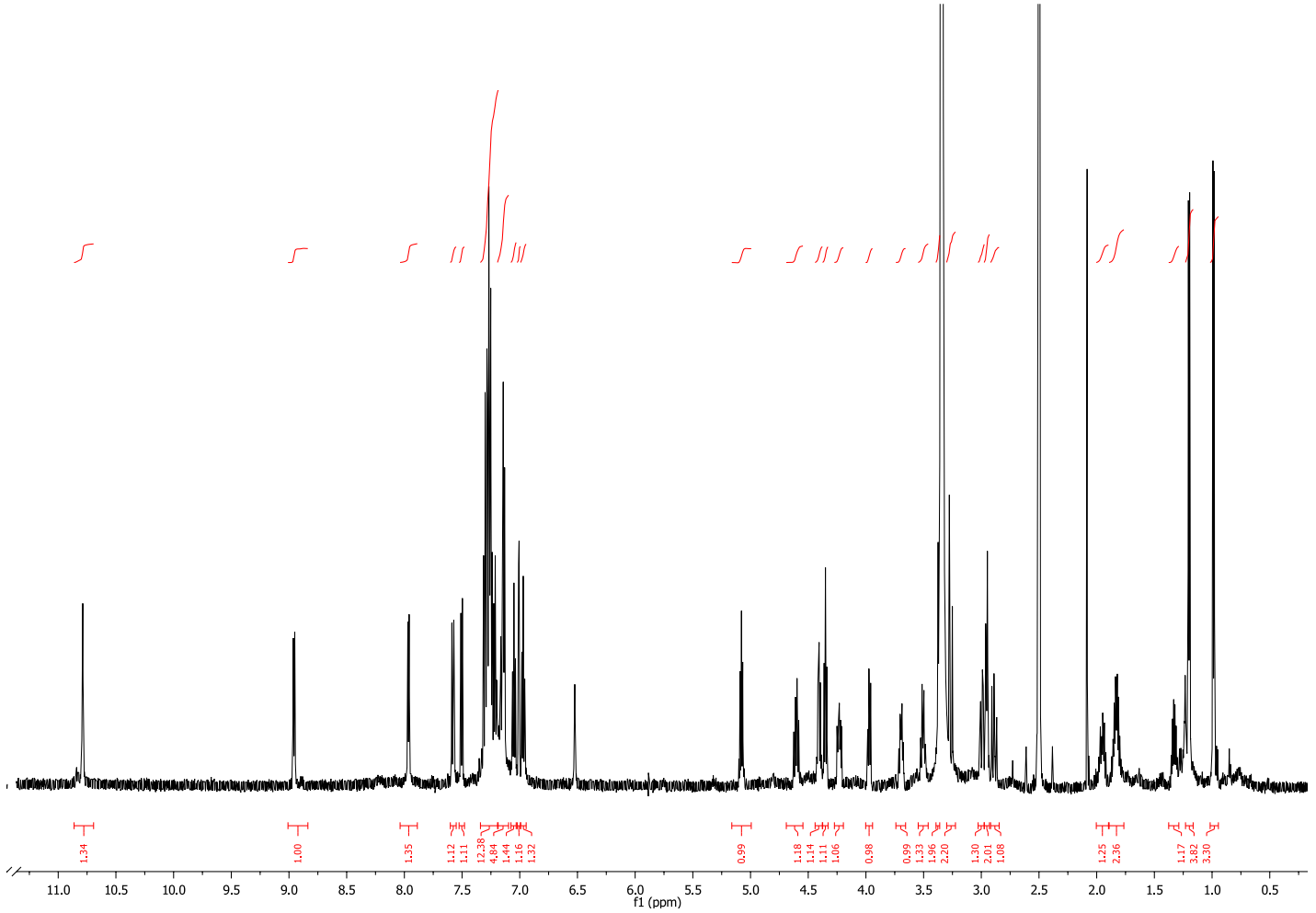



**Figure A.1:** $^{1}$H NMR spectrum of chaiyaphumine **10a**, $d_6$-DMSO, 298 K, 600 MHz.

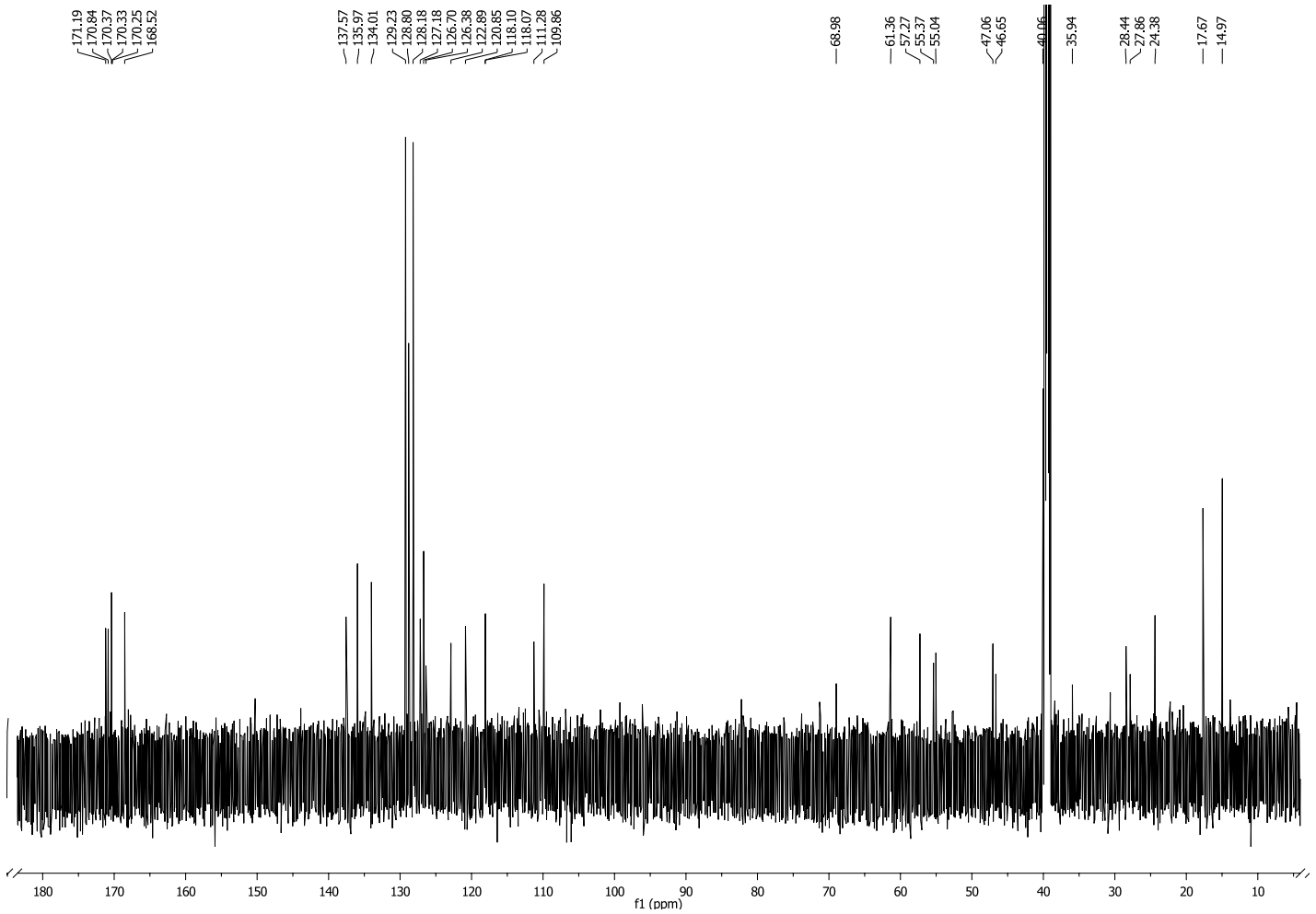



**Figure A.2:** $^{13}$C NMR spectrum of chaiyaphumine **10a**, $d_6$-DMSO, 298 K, 150 MHz.

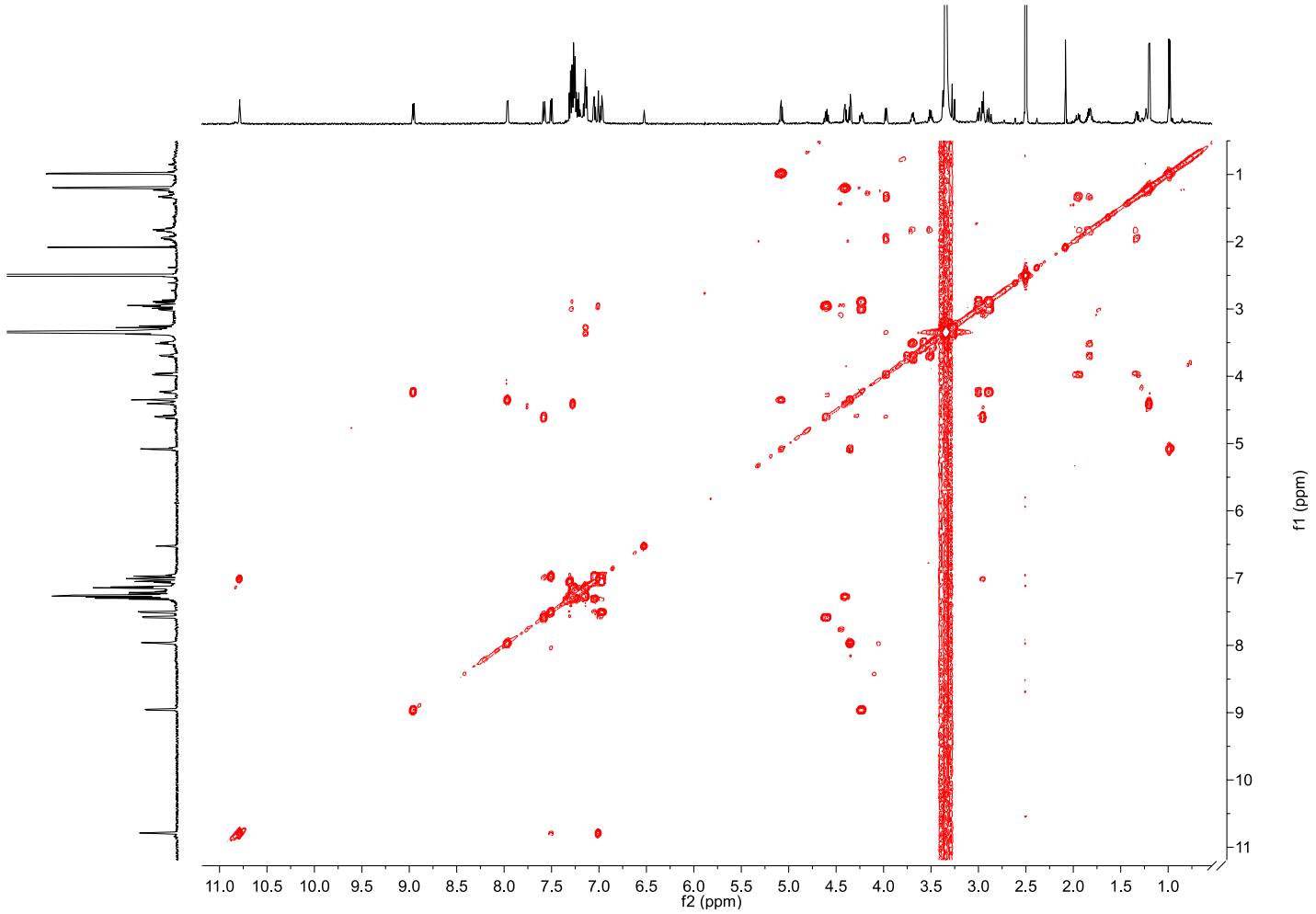



**Figure A.3:** $^{1}$H-$^{1}$H COSY NMR spectrum of chaiyaphumine **10a**, $d_6$-DMSO, 298 K, 600 MHz.

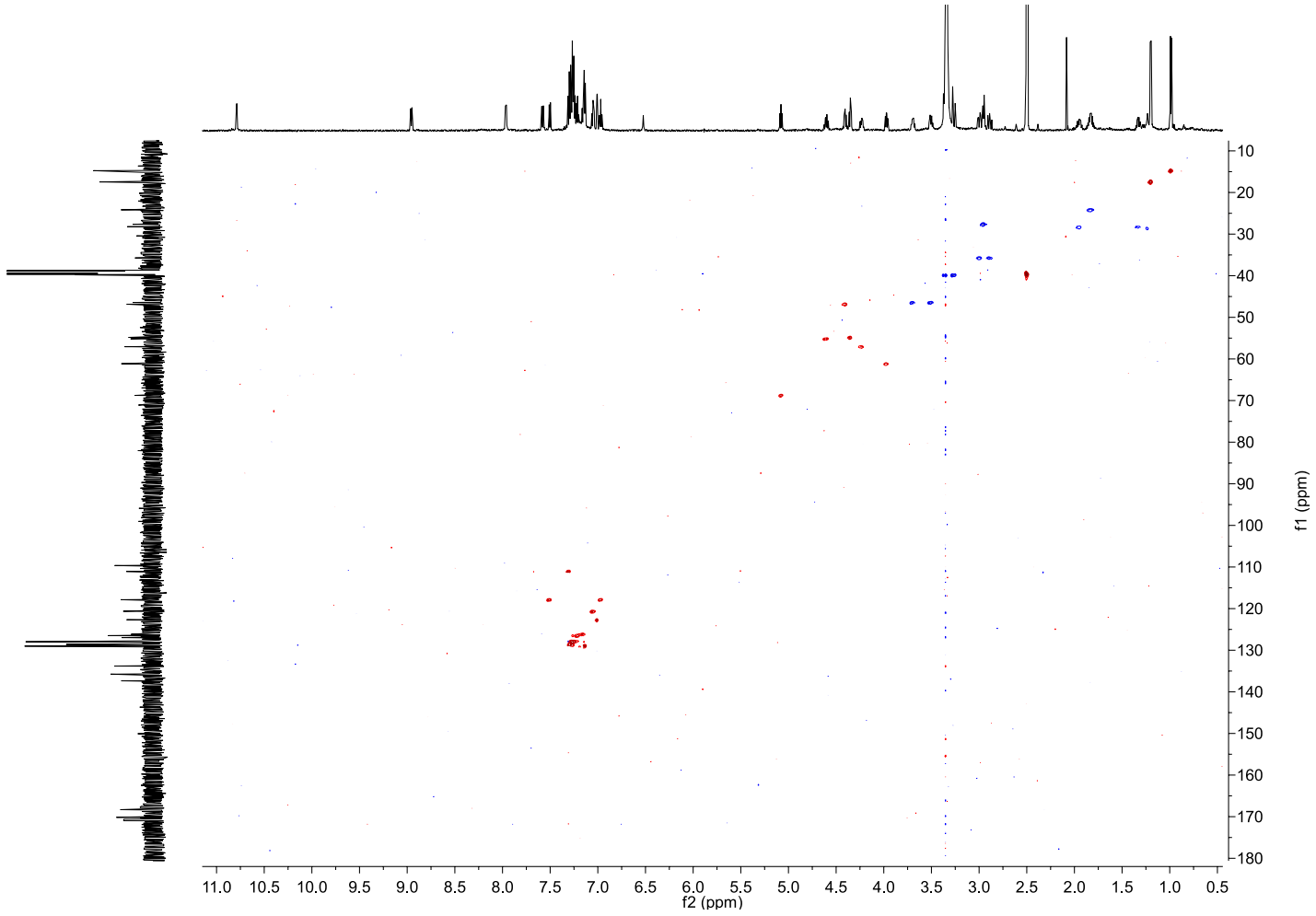



**Figure A.4:** $^{1}H$-$^{13}C$ HSQC NMR spectrum of chaiyaphumine **10a**, $d_6$-DMSO, 298 K, 600 MHz.

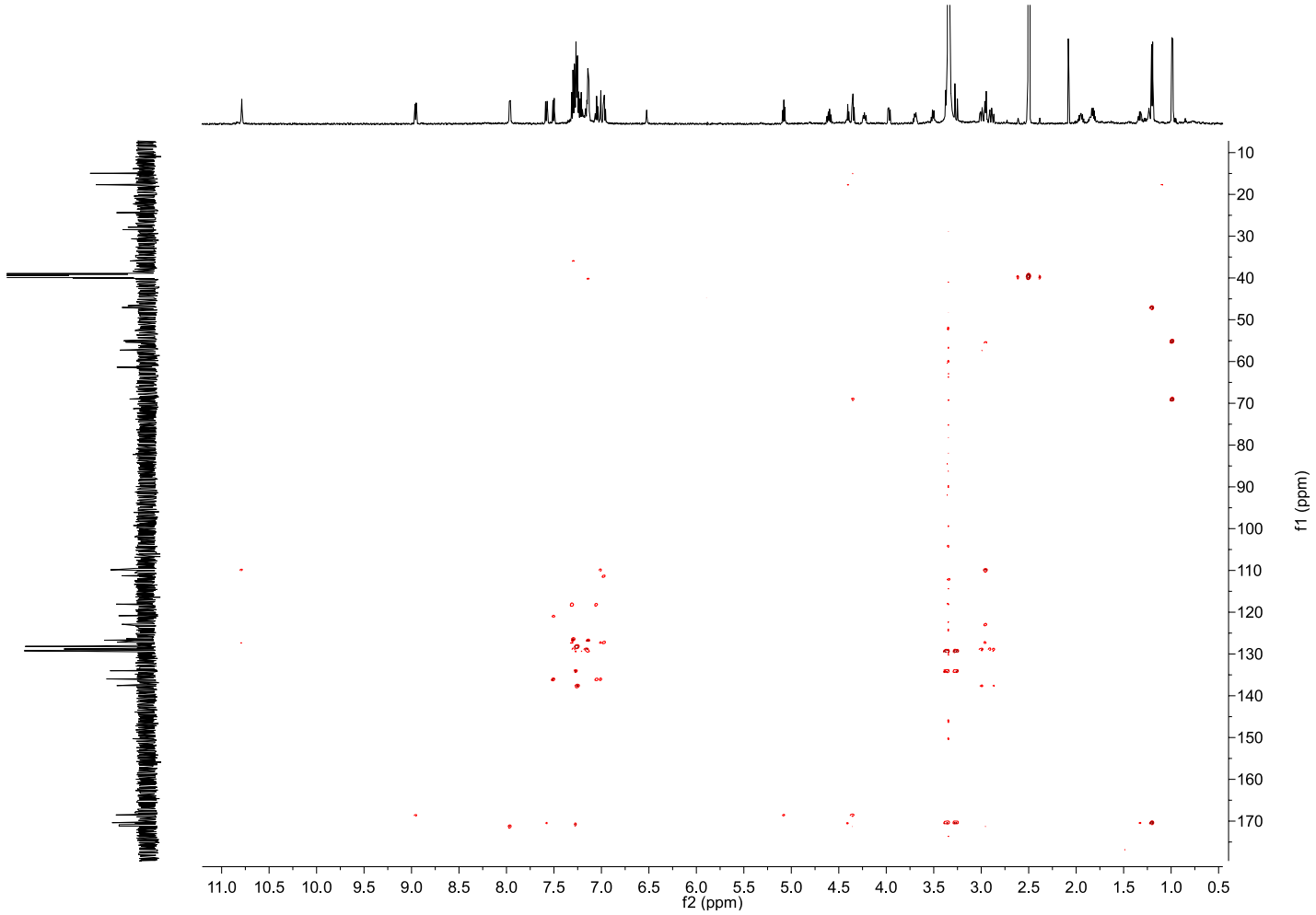



**Figure A.5:** $^{1}H$-$^{13}C$ HMBC NMR spectrum of chaiyaphumine **10a**, $d_6$-DMSO, 298 K, 600 MHz.

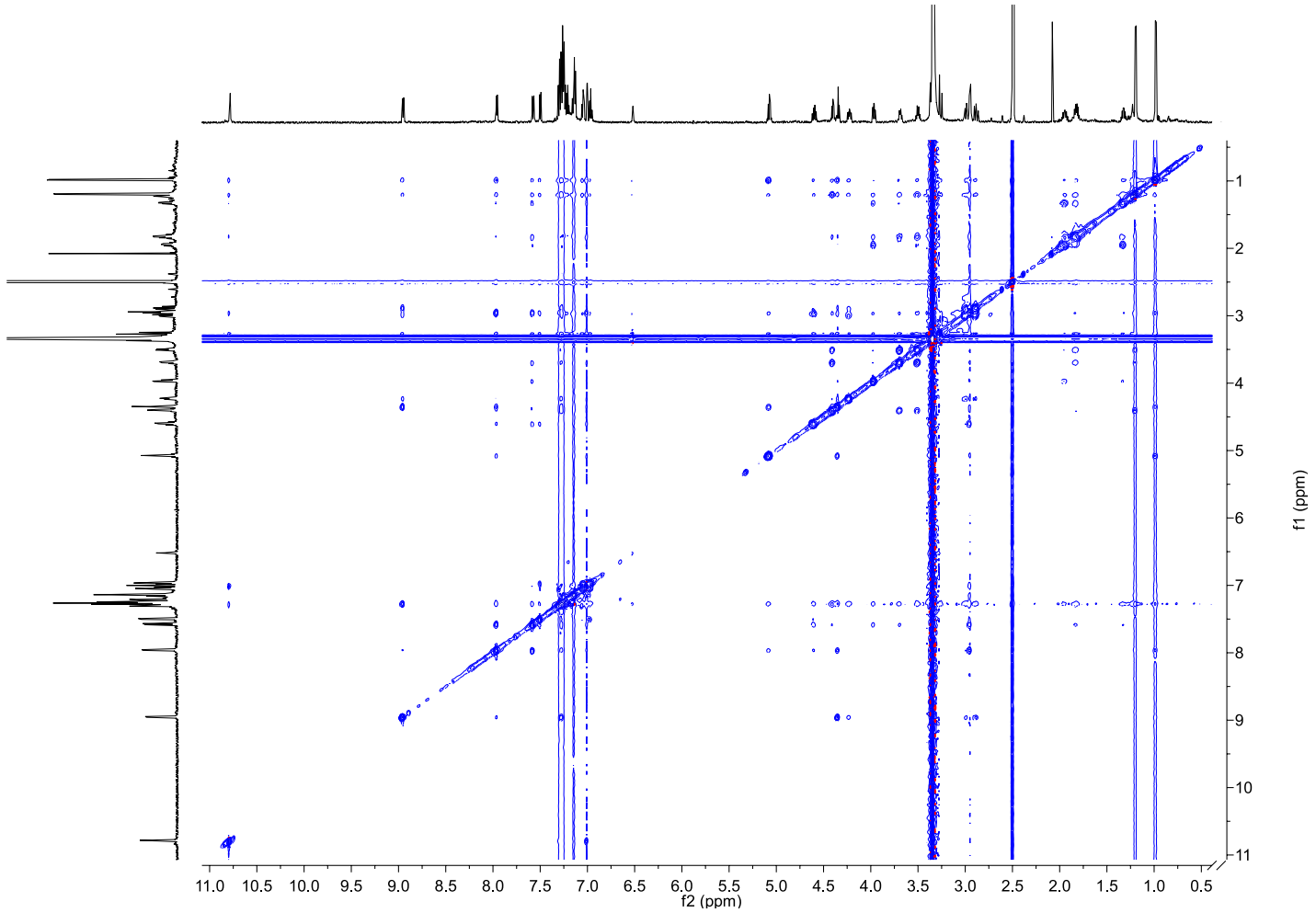



**Figure A.6:** $^{1}$H-$^{1}$H NOESY NMR spectrum of chaiyaphumine **10a**, $d_6$-DMSO, 298 K, 600 MHz.

**Table A.1:** NOE-derived distance restraint list following final molecular dynamics energy minimization run.

| *Atom 1* | *Atom 2* | *Lower* | *Upper* | *Av dist* | *Viol* |
|---|---|---|---|---|---|
| 1-Thr α | 1-Thr NH | 1.7 | 3.2 | 2.90 | 0 |
| 1-Thr α | 2-Phe NH | 1.5 | 2.4 | 2.24 | 0 |
| 1-Thr α | 1-Thr γ | 1.7 | 3.2 | 2.20 | 0 |
| 1-Thr α | 2-Phe β | 1.9 | 4.6 | 4.41 | 0 |
| 1-Thr α | 1-Thr NH | 1.7 | 3.1 | 2.90 | 0 |
| 1-Thr α | 2-Phe NH | 1.4 | 2.3 | 2.24 | 0 |
| 1-Thr α | 1-Thr γ | 1.7 | 3.4 | 2.20 | 0 |
| 1-Thr α | 1-Thr γ | 1.7 | 3.2 | 2.20 | 0 |
| 1-Thr α | 2-Phe β | 1.9 | 4.7 | 4.41 | 0 |
| 1-Thr α | 1-Thr NH | 1.7 | 3.1 | 2.90 | 0 |
| 1-Thr α | 2-Phe NH | 1.4 | 2.3 | 2.24 | 0 |
| 1-Thr β | 1-Thr NH | 1.8 | 3.9 | 2.68 | 0 |
| 1-Thr β | 1-Thr γ | 1.5 | 2.6 | 2.09 | 0 |
| 1-Thr β | 1-Thr α | 1.6 | 2.9 | 2.96 | 0.07 |
| 1-Thr β | 1-Thr NH | 1.8 | 3.5 | 2.68 | 0 |
| 1-Thr β | 2-Phe NH | 1.9 | 4.2 | 4.25 | 0.03 |
| 1-Thr β | 1-Thr γ | 1.6 | 2.9 | 2.09 | 0 |
| 1-Thr β | 1-Thr γ | 1.6 | 2.7 | 2.09 | 0 |
| 1-Thr β | 1-Thr α | 1.6 | 2.9 | 2.96 | 0.11 |
| 1-Thr β | 1-Thr NH | 1.8 | 3.6 | 2.68 | 0 |
| 1-Thr β | 2-Phe NH | 1.9 | 4.3 | 4.25 | 0 |
| 1-Thr γ | 2-Phe NH | 1.9 | 4.3 | 3.20 | 0 |
| 1-Thr γ | 5-Trp δ | 1.9 | 4.5 | 4.53 | 0.06 |
| 1-Thr γ | 2-Phe NH | 1.9 | 4.3 | 3.20 | 0 |
| 1-Thr γ | 1-Thr NH | 1.8 | 3.7 | 3.74 | 0.05 |
| 1-Thr γ | 1-Thr NH | 1.8 | 3.5 | 3.74 | 0.21 |
| 1-Thr γ | 5-Trp δ | 1.9 | 4.6 | 4.53 | 0 |
| 1-Thr γ | 2-Phe NH | 1.9 | 4.4 | 3.20 | 0 |
| 1-Thr γ | 1-Thr γ | 1.1 | 1.5 | 1.31 | 0 |
| 1-Thr NH | 5-Trp NH | 1.5 | 2.4 | 2.42 | 0.03 |
| 1-Thr NH | 5-Trp NH | 1.5 | 2.5 | 2.42 | 0 |
| 1-Thr NH | 5-Trp NH | 1.5 | 2.5 | 2.42 | 0 |
| 1-Thr NH | 5-Trp NH | 1.5 | 2.4 | 2.42 | 0 |
| 1-Thr NH | 3-D-Ala α | 1.9 | 4.7 | 4.69 | 0 |
| 2-Phe α | 2-Phe NH | 1.8 | 3.7 | 2.78 | 0 |
| 2-Phe α | 2-Phe β | 1.7 | 3.0 | 2.31 | 0 |
| 2-Phe α | 2-Phe β | 1.8 | 3.4 | 2.98 | 0 |
| 2-Phe α | 2-Phe β | 1.9 | 4.0 | 2.31 | 0 |
| 2-Phe α | 2-Phe β | 1.8 | 3.6 | 2.31 | 0 |
| 2-Phe α | 2-Phe NH | 1.7 | 3.4 | 2.78 | 0 |
| 2-Phe α | 2-Phe β | 1.9 | 4.6 | 2.98 | 0 |
| 2-Phe α | 2-Phe β | 1.9 | 4.7 | 2.98 | 0 |
| 2-Phe α | 2-Phe β | 1.7 | 3.1 | 2.31 | 0 |
| 2-Phe α | 2-Phe β | 1.8 | 3.5 | 2.98 | 0 |
| 2-Phe α | 2-Phe NH | 1.8 | 3.4 | 2.78 | 0 |
| 2-Phe β | 2-Phe NH | 1.8 | 3.6 | 2.82 | 0 |
| 2-Phe β | 2-Phe NH | 1.6 | 2.9 | 2.19 | 0 |
| 2-Phe β | 2-Phe NH | 1.9 | 4.1 | 2.82 | 0 |
| 2-Phe β | 2-Phe NH | 1.8 | 3.7 | 2.82 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| 2-Phe β | 2-Phe NH | 1.7 | 3.3 | 2.19 | 0 |
| 2-Phe β | 2-Phe NH | 1.7 | 3.1 | 2.19 | 0 |
| 2-Phe β | 2-Phe NH | 1.8 | 3.6 | 2.19 | 0 |
| 2-Phe β | 1-Thr γ | 1.9 | 4.6 | 4.70 | 0.07 |
| 2-Phe β | 2-Phe NH | 1.8 | 3.6 | 2.82 | 0 |
| 2-Phe β | 2-Phe NH | 1.6 | 3.0 | 2.19 | 0 |
| 2-Phe β | 1-Thr γ | 1.9 | 4.7 | 4.70 | 0.01 |
| 2-Phe β | 1-Thr γ | 2.0 | 4.8 | 4.96 | 0.21 |
| 2-Phe NH | 1-Thr NH | 1.9 | 4.3 | 4.05 | 0 |
| 2-Phe NH | 1-Thr NH | 1.9 | 4.0 | 4.05 | 0.06 |
| 2-Phe NH | 1-Thr NH | 1.9 | 4.1 | 4.05 | 0 |
| 3-D-Ala α | 5-Trp NH | 1.9 | 4.5 | 4.12 | 0 |
| 3-D-Ala α | 5-Trp NH | 1.9 | 4.6 | 4.12 | 0 |
| 3-D-Ala β | 3-D-Ala α | 1.6 | 2.7 | 2.12 | 0 |
| 3-D-Ala β | 4-Pro δ | 1.7 | 3.2 | 3.21 | 0.03 |
| 3-D-Ala β | 3-D-Ala α | 1.6 | 3.0 | 2.12 | 0 |
| 3-D-Ala β | 4-Pro δ | 1.7 | 3.2 | 3.21 | 0.02 |
| 3-D-Ala β | 2-Phe β | 1.9 | 4.6 | 4.04 | 0 |
| 3-D-Ala β | 4-Pro δ | 1.9 | 4.7 | 4.06 | 0 |
| 3-D-Ala β | 3-D-Ala α | 1.6 | 2.8 | 2.12 | 0 |
| 3-D-Ala β | 4-Pro δ | 1.9 | 4.0 | 4.06 | 0.11 |
| 3-D-Ala β | 4-Pro δ | 1.7 | 3.2 | 3.21 | 0 |
| 3-D-Ala β | 2-Phe β | 1.9 | 4.7 | 4.04 | 0 |
| 3-D-Ala β | 3-D-Ala β | 1.1 | 1.5 | 1.31 | 0 |
| 4-Pro α | 5-Trp NH | 1.9 | 4.1 | 3.58 | 0 |
| 4-Pro α | 5-Trp NH | 1.9 | 4.0 | 3.58 | 0 |
| 4-Pro α | 4-Pro β | 1.7 | 3.2 | 2.64 | 0 |
| 4-Pro α | 4-Pro β | 1.6 | 2.7 | 2.31 | 0 |
| 4-Pro α | 5-Trp NH | 1.8 | 3.5 | 3.58 | 0.05 |
| 4-Pro α | 4-Pro β | 1.8 | 3.9 | 2.64 | 0 |
| 4-Pro α | 4-Pro β | 1.7 | 3.2 | 2.31 | 0 |
| 4-Pro α | 4-Pro β | 1.7 | 3.3 | 2.64 | 0 |
| 4-Pro α | 4-Pro β | 1.6 | 2.7 | 2.31 | 0 |
| 4-Pro α | 5-Trp NH | 1.8 | 3.6 | 3.58 | 0 |
| 4-Pro β | 5-Trp NH | 1.9 | 4.6 | 4.14 | 0 |
| 4-Pro β | 5-Trp δ | 2.0 | 4.9 | 4.66 | 0 |
| 4-Pro β | 5-Trp NH | 1.8 | 3.9 | 3.41 | 0 |
| 4-Pro β | 5-Trp NH | 1.8 | 3.9 | 3.41 | 0 |
| 4-Pro β | 5-Trp NH | 1.8 | 3.8 | 3.41 | 0 |
| 4-Pro β | 4-Pro β | 1.5 | 2.6 | 1.77 | 0 |
| 4-Pro β | 4-Pro β | 1.5 | 2.6 | 1.77 | 0 |
| 4-Pro β | 5-Trp NH | 1.9 | 4.7 | 4.14 | 0 |
| 4-Pro β | 5-Trp δ | 2.0 | 5.0 | 4.66 | 0 |
| 4-Pro β | 5-Trp NH | 1.8 | 3.8 | 3.41 | 0 |
| 4-Pro β | 4-Pro β | 1.5 | 2.6 | 1.77 | 0 |
| 4-Pro β | 3-D-Ala β | 1.9 | 4.6 | 4.60 | 0 |
| 4-Pro δ | 5-Trp NH | 1.8 | 3.7 | 2.06 | 0 |
| 4-Pro δ | 5-Trp NH | 1.8 | 3.7 | 2.06 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.4 | 4.04 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.0 | 3.94 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.5 | 3.89 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.6 | 3.22 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.6 | 3.22 | 0 |
| 4-Pro δ | 3-D-Ala α | 1.6 | 2.8 | 1.95 | 0 |
| 4-Pro δ | 4-Pro δ | 1.4 | 2.2 | 1.77 | 0 |
| 4-Pro δ | 3-D-Ala α | 1.5 | 2.5 | 2.52 | 0.01 |

| | | | | | |
|---|---|---|---|---|---|
| 4-Pro δ | 4-Pro δ | 1.4 | 2.3 | 1.77 | 0 |
| 4-Pro δ | 3-D-Ala α | 1.7 | 3.0 | 1.95 | 0 |
| 4-Pro δ | 5-Trp NH | 1.8 | 3.8 | 2.06 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.6 | 3.22 | 0 |
| 4-Pro δ | 5-Trp NH | 1.8 | 3.7 | 2.06 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.5 | 4.04 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.1 | 3.94 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.6 | 3.89 | 0 |
| 4-Pro δ | 3-D-Ala α | 1.6 | 2.9 | 1.95 | 0 |
| 4-Pro δ | 4-Pro δ | 1.4 | 2.3 | 1.77 | 0 |
| 4-Pro δ | 3-D-Ala α | 1.5 | 2.6 | 2.52 | 0 |
| 4-Pro δ | 4-Pro β | 1.9 | 4.6 | 3.22 | 0 |
| 5-Trp α | 1-Thr NH | 1.9 | 4.4 | 3.33 | 0 |
| 5-Trp α | 5-Trp NH | 1.8 | 3.5 | 2.89 | 0 |
| 5-Trp α | 5-Trp NH | 1.8 | 3.7 | 2.89 | 0 |
| 5-Trp α | 5-Trp δ | 1.9 | 4.3 | 4.36 | 0.06 |
| 5-Trp α | 1-Thr NH | 1.9 | 4.2 | 3.33 | 0 |
| 5-Trp α | 1-Thr NH | 1.9 | 4.1 | 3.33 | 0 |
| 5-Trp α | 5-Trp NH | 1.7 | 3.3 | 2.89 | 0 |
| 5-Trp α | 5-Trp NH | 1.7 | 3.4 | 2.89 | 0 |
| 5-Trp α | 5-Trp δ | 1.9 | 4.4 | 4.36 | 0.01 |
| 5-Trp α | 1-Thr NH | 1.9 | 4.2 | 3.33 | 0 |
| 5-Trp α | 5-Trp NH | 1.7 | 3.4 | 2.89 | 0 |
| 5-Trp δ | 5-Trp NH | 1.9 | 4.2 | 2.74 | 0 |
| 5-Trp δ | 5-Trp NH | 1.9 | 4.2 | 2.74 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp α | 1.9 | 4.2 | 2.83 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp Indole 5,7 | 1.7 | 3.3 | 2.44 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp Indole 5,7 | 1.7 | 3.4 | 2.44 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp α | 1.9 | 4.7 | 2.83 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp δ | 1.5 | 2.5 | 2.56 | 0.08 |
| 5-Trp Indole 1 (ArNH) | 5-Trp δ | 1.6 | 2.8 | 2.56 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp α | 1.8 | 3.8 | 2.83 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp Indole 5,7 | 1.6 | 3.0 | 2.44 | 0 |
| 5-Trp Indole 1 (ArNH) | 5-Trp δ | 1.5 | 2.5 | 2.56 | 0.04 |

Average restraint violation: 0.01

# Appendix B

# Lassomycin NMR and MALDI LIFT spectra

The following NMR and MS characterization is included for lassomycin (Chapter 3)

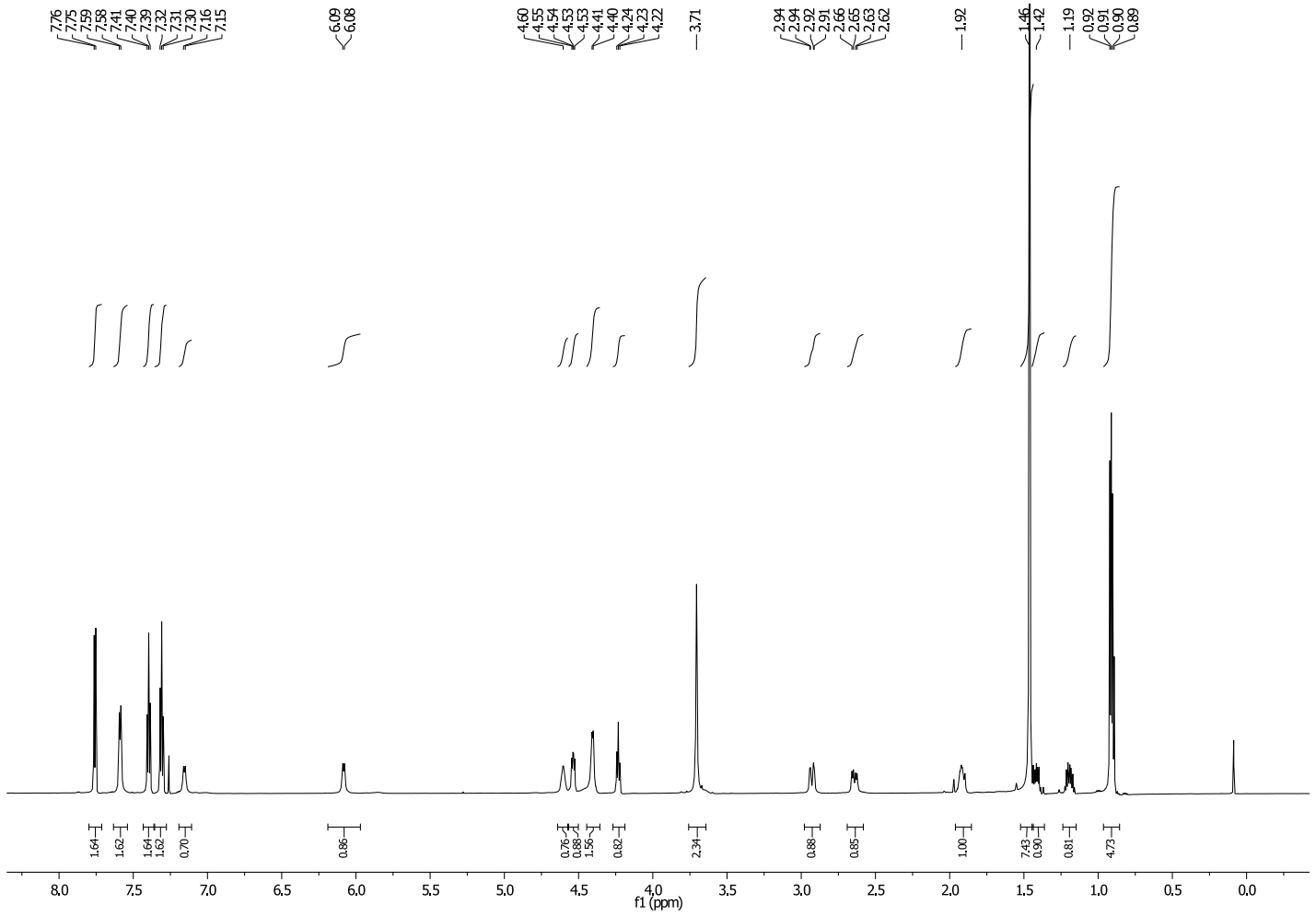



**Figure B.1:** $^{1}$H NMR spectrum of Fmoc-Asp(O*t*Bu)-Ile-OMe (**30**) recorded at 700 MHz ($CDCl_3$).

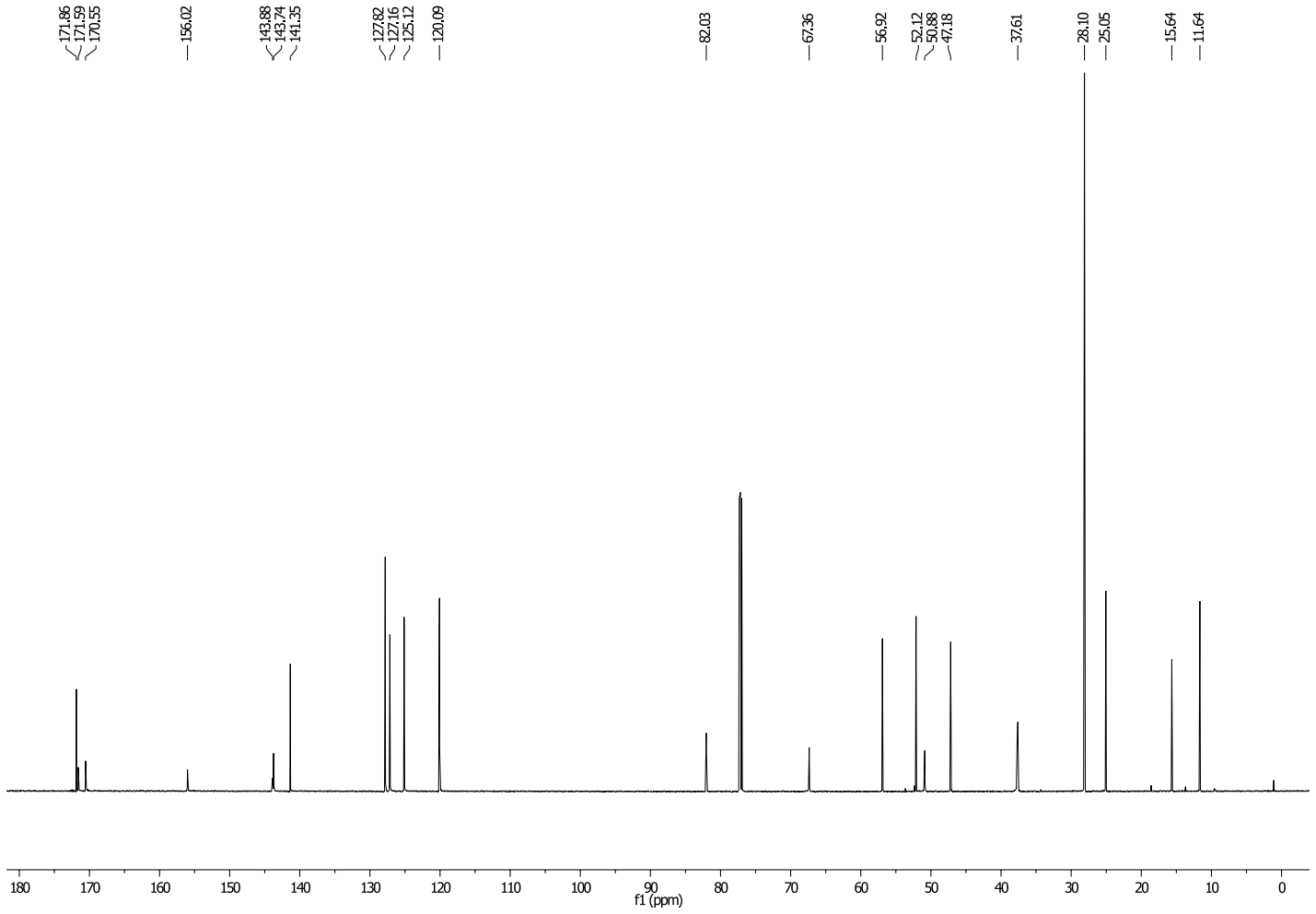



**Figure B.2:** $^{13}$C NMR spectrum of Fmoc-Asp(O*t*Bu)-Ile-OMe (**30**) recorded at 176 MHz ($CDCl_3$).

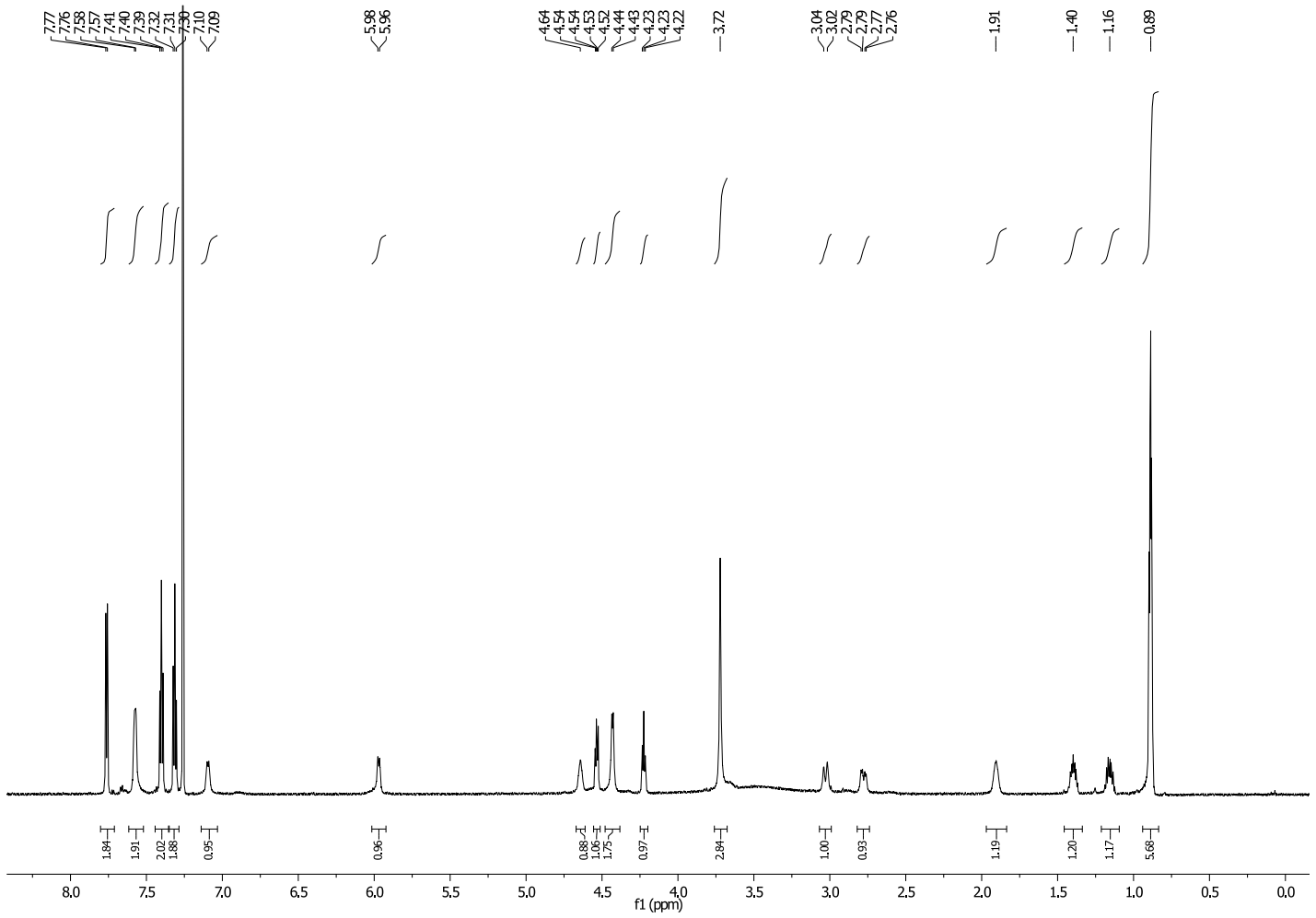



**Figure B.3:** $^{1}$H NMR spectrum of Fmoc-Asp(OH)-Ile-OMe (**34**) recorded at 700 MHz ($CDCl_3$).

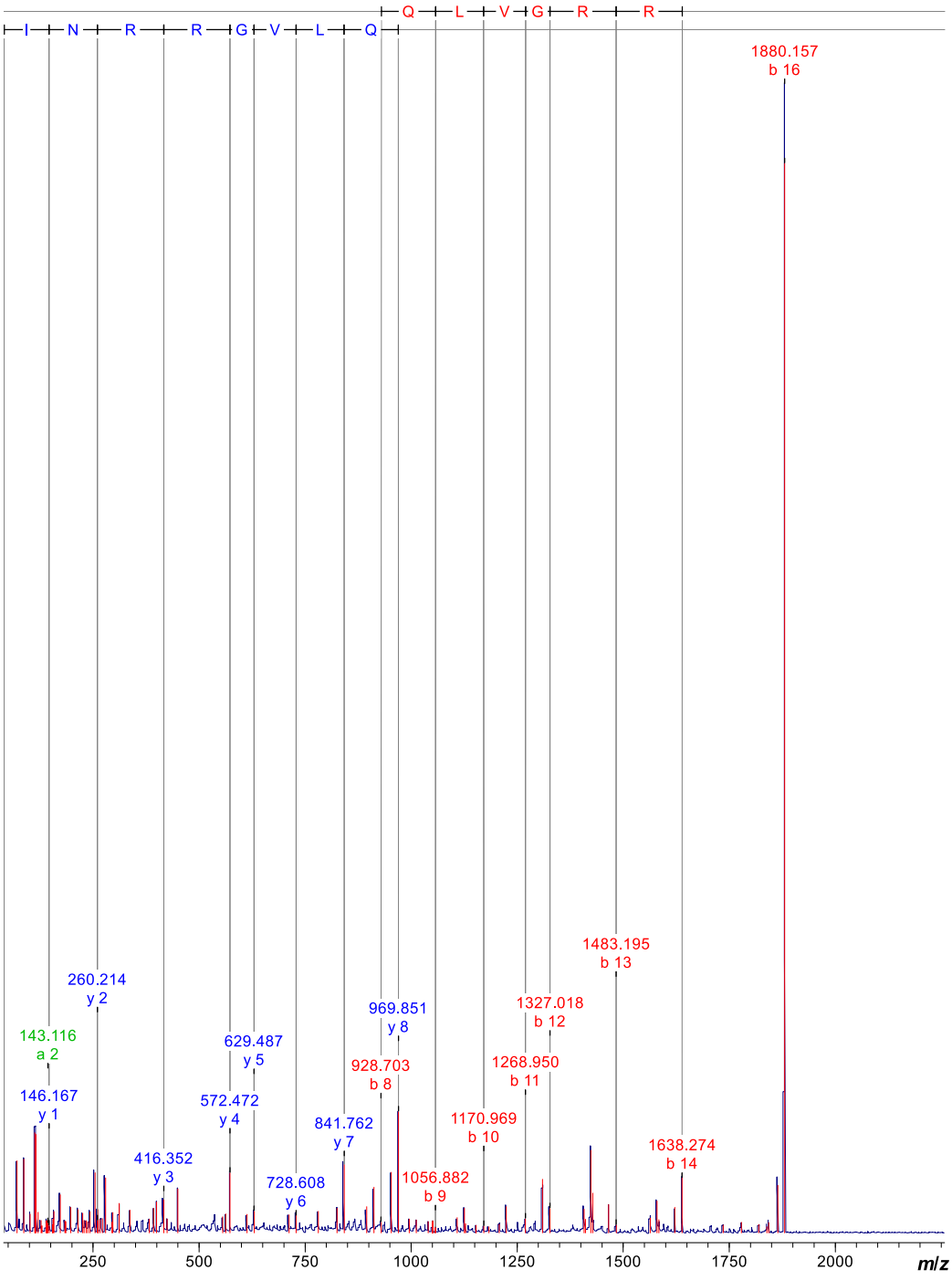



**Figure B.4:** MALDI LIFT-TOF/TOF mass spectrum showing fragmentation of lassomycin (**24**) with ions labelled.

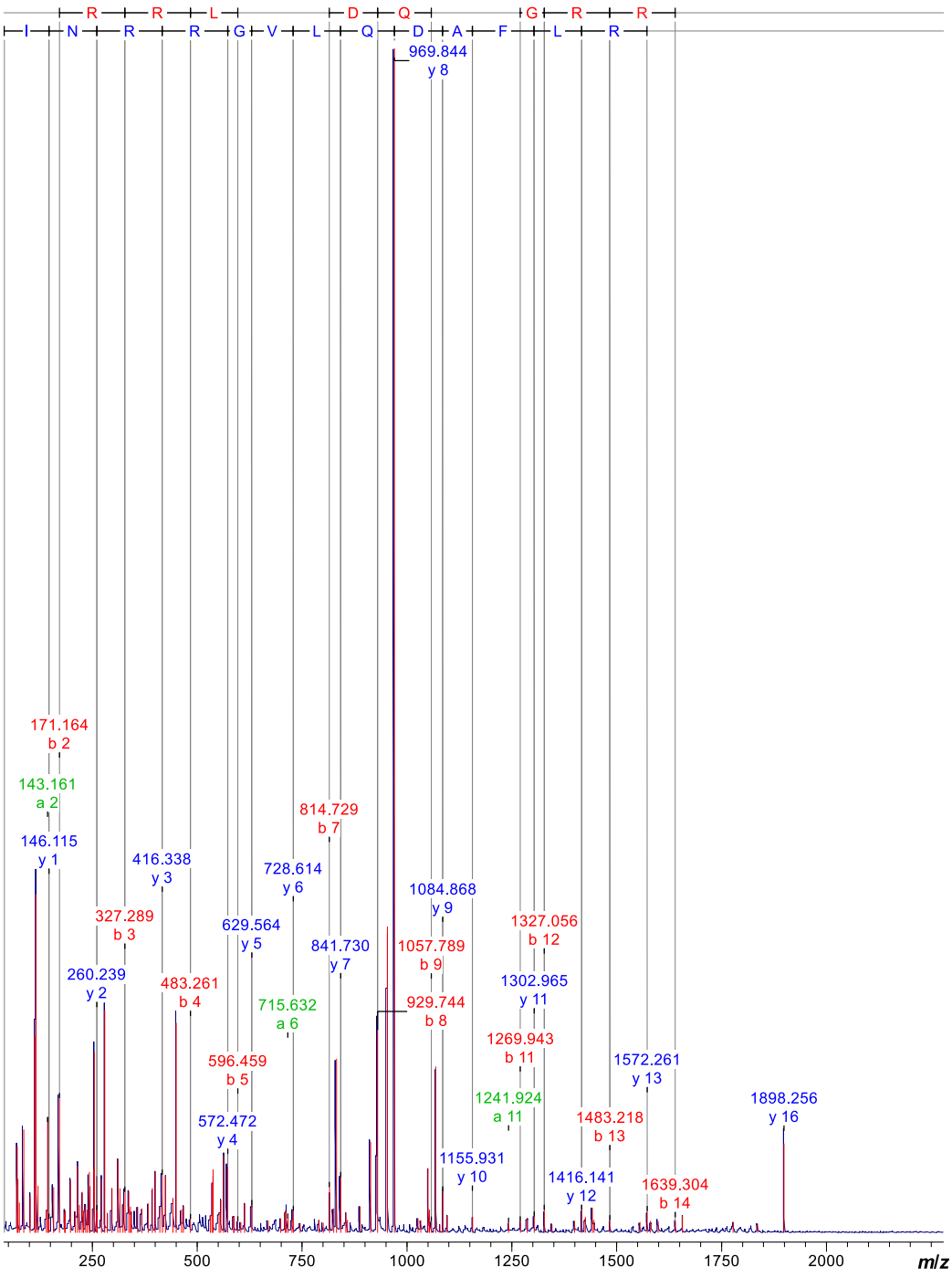



**Figure B.5:** MALDI LIFT-TOF/TOF mass spectrum showing fragmentation of linear (uncyclized) lassomycin (**25**) with ions labelled.

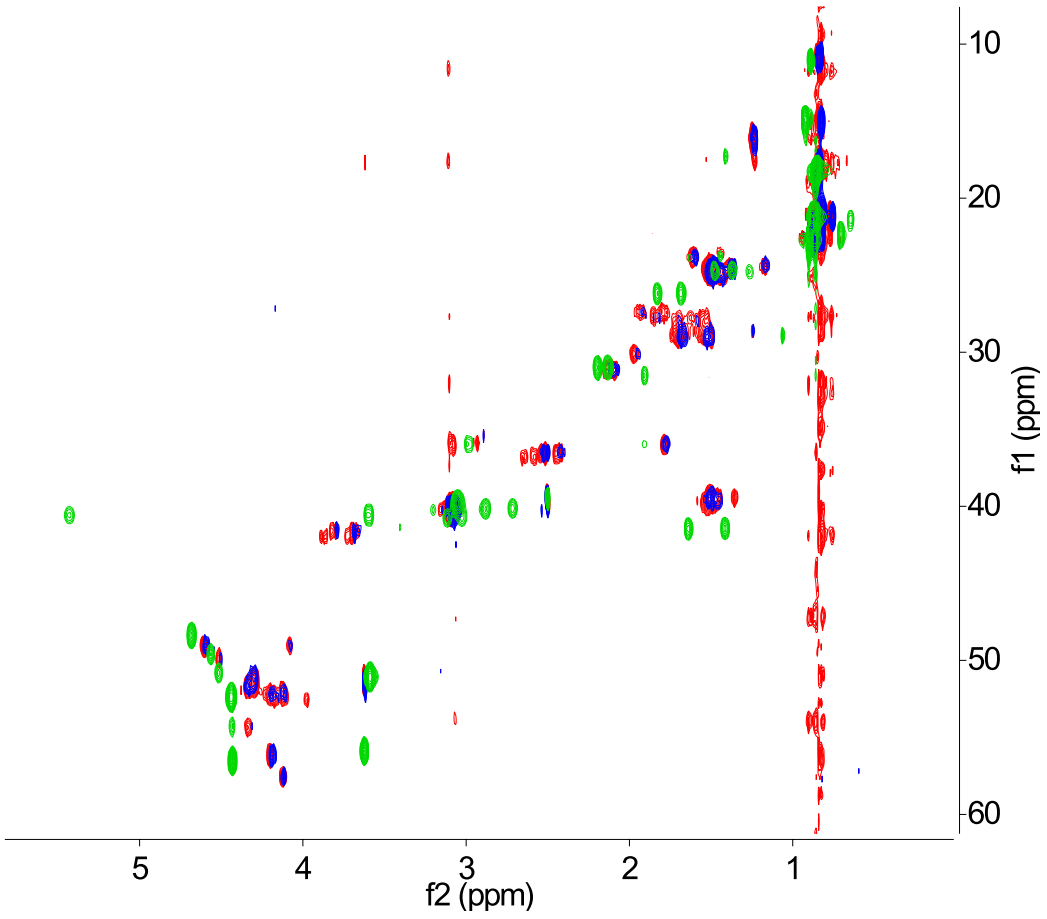



**Figure B.6:** Overlay of $^{1}$H-$^{13}$C HSQC (heteronuclear single-quantum correlation) NMR spectra of naturally isolated [$^{13}$C, $^{15}$N]lassomycin at 40 °C (green), and synthetic lassomycin (**24**) at 25 °C (blue) and 40 °C (red).

# Appendix C

# Synthesis of the POLARIS peptide: A regulator of ethylene signalling in *Arabidopsis*

Synthetic peptides can be useful in elucidating the role of a given gene product in a biological system in order to complement genetic studies. The following chapter describes the synthesis of a peptide found to regulate ethylene signalling in the plant *Arabidopsis*, and the effect of the synthetic peptide on root length in both wild type and mutant seedlings.

## C.1 Introduction

Food security is a pressing issue facing societies worldwide.[239] With the global population projected to rise to over 9 billion by 2050, it is vitally important to ensure sufficient food crops can be grown in increasingly drought-affected or nutrient-stressed environments.[240] Genetic engineering offers a potential solution to this problem, as it can be used to create plant lines which produce deeper root systems or make more efficient use of available water.[241,242] A detailed genetic understanding of plant physiology is therefore crucial, and much of the current research effort is focused on elucidating the processes governing root growth and the mechanisms by which root shape and structure are regulated.

Root growth is controlled by a number of plant hormones including auxin and ethylene. Although the molecular mechanisms behind this interplay are

not well understood, ethylene is thought to regulate growth by achieving local activation of the auxin signalling pathway.[243] A summary of ethylene signalling and the receptors involved is given in the following section.

## C.1.1 Ethylene signalling

Ethylene is perceived by a family of receptor complexes which are associated with or lie within the endoplasmic reticulum (ER) membrane.[244] As a gaseous hormone, ethylene diffuses into and out of cells and so can be detected at the ER, which is not a typical site for receptor-ligand binding. A five-member family of ethylene receptors is responsible for ethylene sensing in the model organism *Arabidopsis thaliana*, composed of ETR1, ERS1, ETR2, ERS2, and EIN4.[245] The C-terminal half of ETR1 has been shown to be important as a docking site for the kinase CTR1, a downstream negative regulator of the ethylene pathway.[245]

A three-state model for ETR1 receptor signalling, including the involvement of CTR1, is shown in Figure C.1. As indicated, the presence of copper is important – the copper transporter RAN1 has been demonstrated to be essential for ethylene-binding activity, and is thought to play a role in the delivery of copper to ethylene receptors.[246]

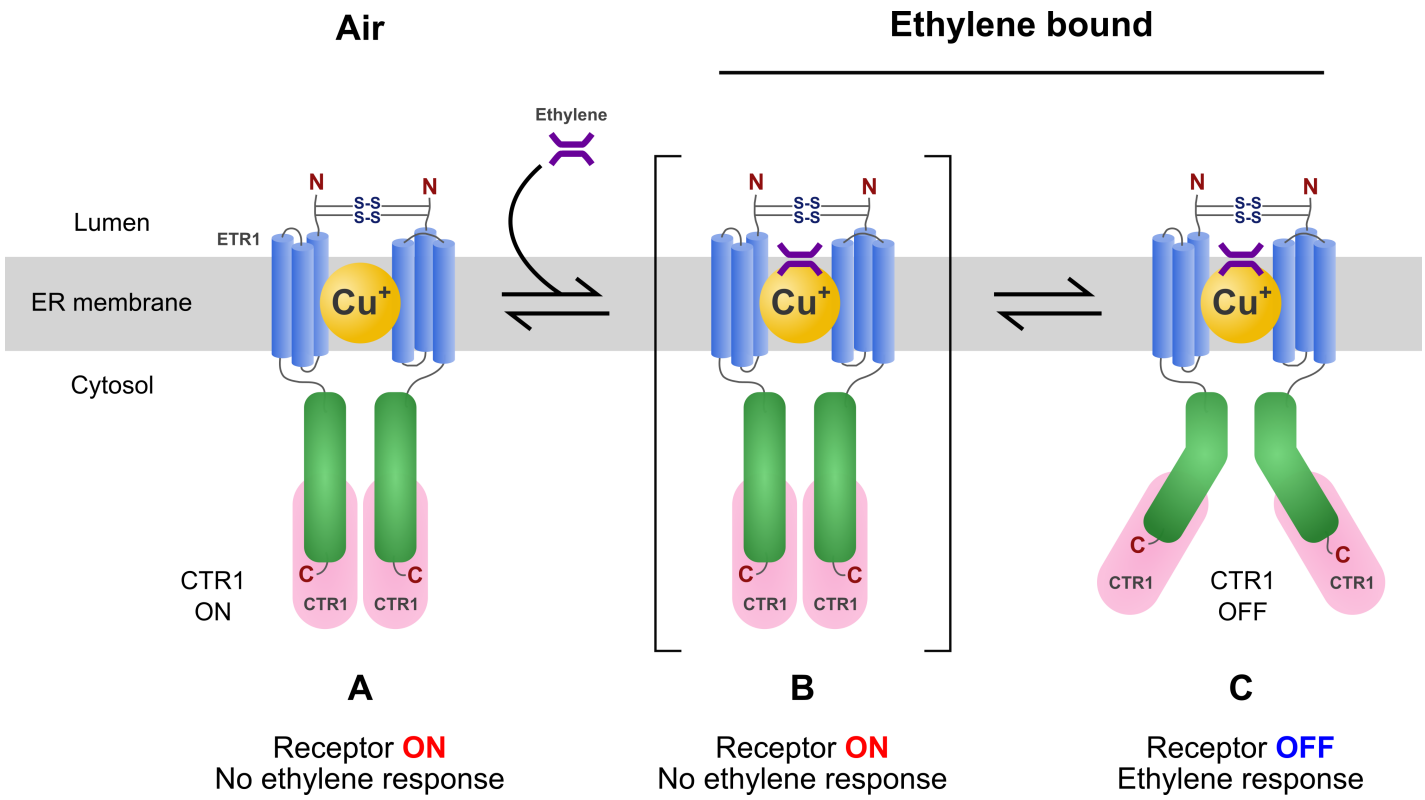



**Figure C.1:** A three-state model for ETR1 receptor signalling. The ethylene binding domain of ETR1 is shown as a disulfide-stabilized homodimer anchored in the ER membrane by three $\alpha$-helical domains on each monomer. Copper(I) and ethylene are shown binding to regions demonstrated to be important in mutation studies. Upon ethylene binding the receptor adopts an intermediate conformation in which it is still transmitting a signal. The third equilibrium state, where the receptor is turned off, leads to diminished activity of CTR1. CTR1 negatively regulates the ethylene response, thus in the third state inhibition of downstream signalling is released. Adapted from Mudge and Lacey et al.[247,248]

### C.1.2 The *Arabidopsis* POLARIS (PLS) peptide

The *POLARIS*[l] (*PLS*) gene from *Arabidopsis thaliana* is required for correct root growth (cell division and elongation) and root architecture (lateral root formation).[249] *PLS* encodes a peptide of 36 amino acids in length, and *pls* mutants exhibit a phenotype with reduced root length. Furthermore, *PLS* is required for correct control of several ethylene-mediated responses, including growth in the dark, microtubule dynamics and auxin transport and homeostasis.

A model proposed by Chilley et al. for the interaction of PLS with the ethylene and auxin signalling pathways, and its function in root growth, is shown in Figure C.2. According to the model, treatment with ethylene results in a decrease in the concentration of auxin in the root tip which in turn leads to a reduction in *PLS* transcription, allowing ethylene to alter root architecture (for example leading to the production of thicker roots and fewer lateral roots).[249] Auxin signalling from the shoot would act antagonistically by increasing *PLS* transcription, thus PLS is suggested to act as a modulator of the auxin-ethylene interaction. This interaction is thought to affect root cell division and shape – and therefore root growth – through effects on the cytoskeleton.

Due to the importance of PLS in the mediation of crosstalk between auxin and ethylene signalling, and by extension for the regulation of root growth and architecture, it was decided that direct action of the PLS peptide upon *Arabidopsis* root growth should be investigated. For this reason a chemical synthesis of PLS was proposed, in order that sufficient quantities of PLS could be produced for use in root growth assays. Crucially, access to the synthetic peptide also allows the facile generation of mutant and truncated PLS analogues, with which regions of the sequence or specific residues essential for activity can be identified. Synthesis and testing of PLS and PLS analogues is described in the following sections.

## C.2 Synthesis of PLS, truncations and mutants

The full length PLS sequence (**87a**, Figure C.3) was synthesized via automated MW-SPPS on polystyrene Wang resin preloaded with the C-terminal amino acid histidine. Truncated sequences (also shown) were synthesized in order to

[l]Nomenclature used in this chapter as follows: POLARIS (PLS) refers to the peptide product of the *POLARIS* (*PLS*) gene, *polaris* (*pls*) refers to the mutant (short-root phenotype).

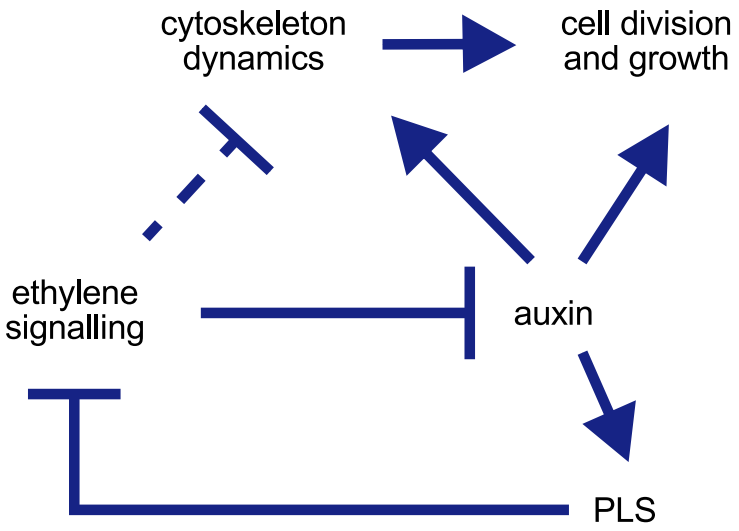



**Figure C.2:** Model for PLS interaction with ethylene and auxin signalling, and its function in root growth. Auxin plays a role in cytoskeleton dynamics, cell division, and cell expansion, while ethylene suppresses these effects. Auxin is a positive regulator of *PLS* expression, and PLS suppresses the effects of ethylene. The relationship between the inhibitory effects of ethylene on cytoskeleton dynamics (dotted line) and auxin transport are unclear mechanistically. Adapted from Chilley et al.[249]

investigate which sections of the sequence are essential for biological function. Truncations **87d** and **87e** were synthesized using Rink amide resin to give an amide instead of an acid group at the C-terminus. A C-terminal amide was assumed to be more similar to the internal amide bond present at the corresponding point in the parent sequence, in contrast to an carboxylic acid, which would carry a negative charge. In a similar manner, truncation peptides **87b** and **87c** were N-terminally acetylated via on-resin treatment with acetic anhydride in order to remove the charge on the N-terminal amine group. Ser and Ala mutants (peptides **87f** and **87g**) were also synthesized to investigate the effect of removing cysteine residues (and any associated potential for disulfide bonding or coordination to metal centres) from the peptide (see below). Correct masses for the PLS peptides were confirmed by MALDI-TOF MS.

| | | |
|---|---|---|
| PLS | **a** | MKPRLCFNFRRRSISPCYISISYLLVAKLFKLFKIH |
| PLS(23-36) | **b** | YLLVAKLFKLFKIH |
| PLS(13-36) | **c** | SISPCYISISYLLVAKLFKLFKIH |
| PLS(1-22) | **d** | MKPRLCFNFRRRSISPCYISIS |
| PLS(1-9) | **e** | MKPRLCFNF |
| [C6S; C17S]PLS | **f** | MKPRL**S**FNFRRRSISP**S**YISISYLLVAKLFKLFKIH |
| [C6A; C17A]PLS | **g** | MKPRL**A**FNFRRRSISP**A**YISISYLLVAKLFKLFKIH |

**Figure C.3:** The sequence of the PLS peptide (**87a**), in addition to those of the truncations **87b**–**e** and mutant **87f** (in which the Cys residues have been replaced by Ser) that were synthesized. Peptides **87b** and **87c** are acetylated at the N-terminus, and **87d** and **87e** are C-terminally amidated.

## C.3 Effect of synthetic peptides on root growth

The synthetic peptides were tested on both wild type (C24) *Arabidopsis* seedlings, and the loss-of-function *pls* mutant (in which the PLS peptide is absent) to investigate their effects on root growth (Anna Mudge, Durham).[247]

Seedlings were grown for 10 days in liquid growth medium supplemented with full length PLS (**87a**), truncation peptides **87b–e**, or no peptide. Mean primary root length after 10 days is plotted in Figure C.4.

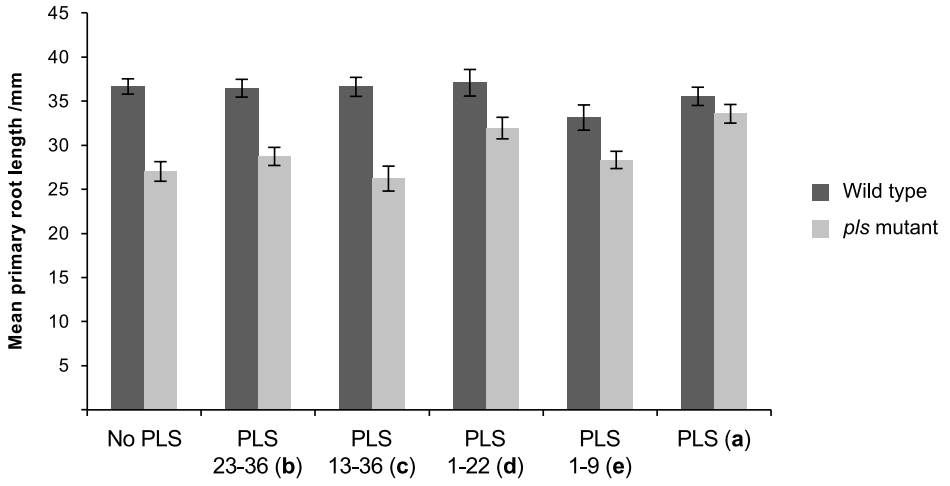



**Figure C.4:** Mean primary root length of wild type and *pls* mutant seedlings treated with peptides **87a–e** (50 nM final concentration, 0.01% DMSO v/v) at 10 days after germination. Error bars show ±1 standard error.

The primary root length of the wild type does not change significantly with the addition of full length PLS (**87a**) or any of the truncated peptides (**87b–e**).[II] In contrast, there is a significant effect on the *pls* mutant upon addition of PLS peptides,[III] with **87a** and PLS(1-22) (**87d**) significantly increasing the primary root length (Tukey's HSD test).[247]

Both full length **87a** and PLS(1-22) (**87d**) were shown to rescue root length in the *pls* mutant, whereas PLS(1-9) (**87e**) and PLS(13-36) (**87c**) were not. As almost all of the sequence of **87d** is overlapped by both **87c** and **87e** (see Figure C.3), it was suggested that the presence of both Cys residues may be important for activity, a characteristic shared by all peptides found to rescue root growth.

To investigate the role of the Cys residues, the effect of Ser mutant **87f** on root growth was examined. (Ala mutant **87g** was not available at the time the

[II] Analysis of variance $F(5, 121) = 1.6$, $p = 0.16$.[247]

[III] Analysis of variance $F(5, 112) = 6.13$, $p = 4.65 \times 10^{-5}$.[247]

assay was carried out and will be tested in the future.) Mean primary root length after 10 days is plotted in Figure C.5.

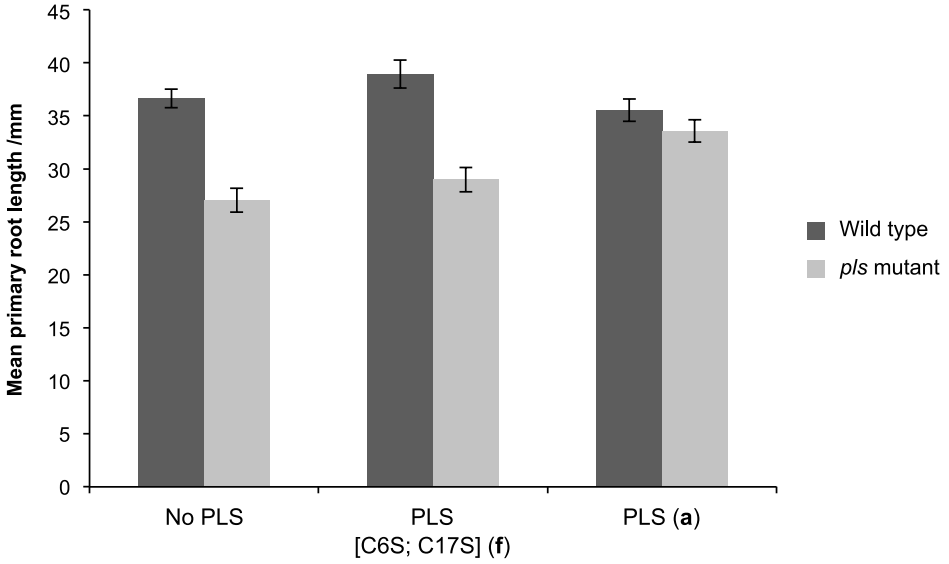



**Figure C.5:** Mean primary root length of wild type and *pls* mutant seedlings treated with full length PLS (**87a**) and mutant **87f** (100 nM final concentration, 0.01% DMSO v/v), at 10 days after germination. Error bars show ±1 standard error.

The PLS peptide lacking the two Cys residues (**87f**) did not exhibit the same activity in *Arabidopsis* roots as the unmodified peptide (**87a**). While the mean primary root length did not change significantly for the wild type when treated with either **87a** or **87f**,[IV] the *pls* seedlings were significantly affected.[V] However, only the unmodified full sequence (**87a**) significantly affected the mean *pls* root length (Tukey's HSD test).[247] The presence of both Cys residues is therefore required for PLS function.

## C.4 Conclusions and future work

PLS, a peptide found to be important in the regulation of root growth and architecture in *Arabidopsis thaliana*, was chemically synthesized, and sufficient

[IV] Analysis of variance $F(2, 58) = 2.42$, $p \geqslant 0.05$.[247]
[V] Analysis of variance $F(2, 51) = 9.48$, $p = 3.15 \times 10^{-4}$.[247]

quantities were obtained to evaluate its effect on *Arabidopsis* seedlings in a root growth assay. The synthetic route to full length PLS (**87a**) also allowed access to a number of truncation peptides (**87b**–**e**), which were used to look for regions in the PLS sequence important for activity. Both **87a** and the truncation **87d** were shown to rescue root growth in the *pls* mutant, and due to a lack of activity for all other peptides, it was suggested that the two cysteine residues present only in the active sequences were responsible for their ability to rescue root growth.

Mutants **87f** and **87g**, in which the cysteines were replaced by Ser or Ala respectively, were synthesized. Peptide **87f** was found to lack any ability to rescue root growth, supporting the hypothesis that the Cys residues are essential. (The activity of **87g** will be evaluated in future assay work.) These results suggest that the presence of both Cys residues in the sequence of PLS is necessary for its function, which may be related to cysteine's ability to coordinate metal ions. Considering the importance of copper binding in ethylene receptor signalling (discussed previously), is it suggested that PLS may potentially modulate the availability of copper at the ethylene receptor, and further work is underway in the Lindsey group to investigate this.[247]

PLS may exert its effect on ethylene signalling by decreasing the availability of copper at the ETR1 receptor in order to prevent ethylene binding, thus maintaining the receptor in its 'on' state and negatively regulating the downstream ethylene response (see Figure C.1).[247] Future work will involve investigation of the interaction between copper(I) and PLS using circular dichroism spectroscopy and crystallization studies.

# Appendix D

# Pep-Calc.com *J. Comput. Aid. Mol. Des.* article

The following article describing the web utility presented in Chapter 6 was published in the Journal of Computer-Aided Molecular Design.

J Comput Aided Mol Des
DOI 10.1007/s10822-016-9902-7

# Pep-Calc.com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment

**Sam Lear**[1] · **Steven L. Cobb**[1]

Received: 11 December 2015 / Accepted: 12 February 2016


**Abstract** The ability to calculate molecular properties such as molecular weights, isoelectric points, and extinction coefficients is vital for scientists using and/or synthesizing peptides and peptoids for research. A suite of two web utilities: Peptide Calculator and Peptoid Calculator, available free at http://www.pep-calc.com, are presented. Both tools allow the calculation of peptide/peptoid chemical formulae and molecular weight, ChemDraw structure file export and automatic assignment of mass spectral peaks to deletion sequences and metal/protecting group adducts. Peptide Calculator also provides a calculated isoelectric point, molar extinction coefficient, graphical peptide charge summary and $\beta$-strand contiguity profile (for aggregation-prone sequences), indicating potential regions of synthesis difficulty. In addition to the unique automatic spectral assignment features offered across both utilities, Peptoid Calculator represents a first-of-a-kind resource for researchers in the field of peptoid science. With a constantly expanding database of over 120 amino acids, non-natural peptide building blocks and peptoid building blocks, it is anticipated that Pep-Calc.com will act as a valuable asset to those working on the synthesis and/or application of peptides and peptoids in the biophysical and life sciences fields.

**Keywords** Peptide · Peptoid · Calculated properties · Automatic mass assignment · ChemDraw export · Calculator

✉ Sam Lear
contact@pep-calc.com

[1] Department of Chemistry, Durham University, South Road, Durham DH1 3LE, UK

## Introduction

Convenient and rapid access to calculated molecular properties is essential for researchers using and/or synthesizing peptides and peptidomimetics for biophysical or life sciences applications. Furthermore, the process of assigning peptide byproducts in mass spectra resulting from residue deletions or incomplete protecting group removal during a synthesis can be a laborious and time consuming process, and access to freely available automatic assignment tools is necessary to improve workflow and increase research efficiency. While a plethora of peptide and protein property calculation tools are accessible online, very few offer mass spectral peak assignment functionality, and for those that do this is often extremely limited.

While the ExPASy portal [1] acts as the most comprehensive protein property calculation resource for molecular biology, other more specific tools exist, such as ChemCalc [2], PredictProtein [3], IMSPeptider [4], POTAMOS [5], Top Pred [6], CheckMyMetal [7], AFAL [8] and a host of other peptide property calculation utilities [9–16]. Few of these are designed specifically with the synthetic peptide chemist in mind however, and furthermore, to the best of our knowledge, no freely available web services exist for the calculation of *peptoid* molecular properties or assignment of peptoid synthesis mass spectra.

We present a pair of web tools: Peptide Calculator and Peptoid Calculator, for chemical formula and molecular weight calculation of peptides and peptoids. In addition, both sites offer automatic assignment of mass spectral peaks to deletion sequences, metal ion adducts and protected byproducts, as well as the option to download structures in ChemDraw format for the sequences entered. Peptide Calculator can also give calculated values for isoelectric point and molar extinction coefficient (at

Published online: 24 February 2016

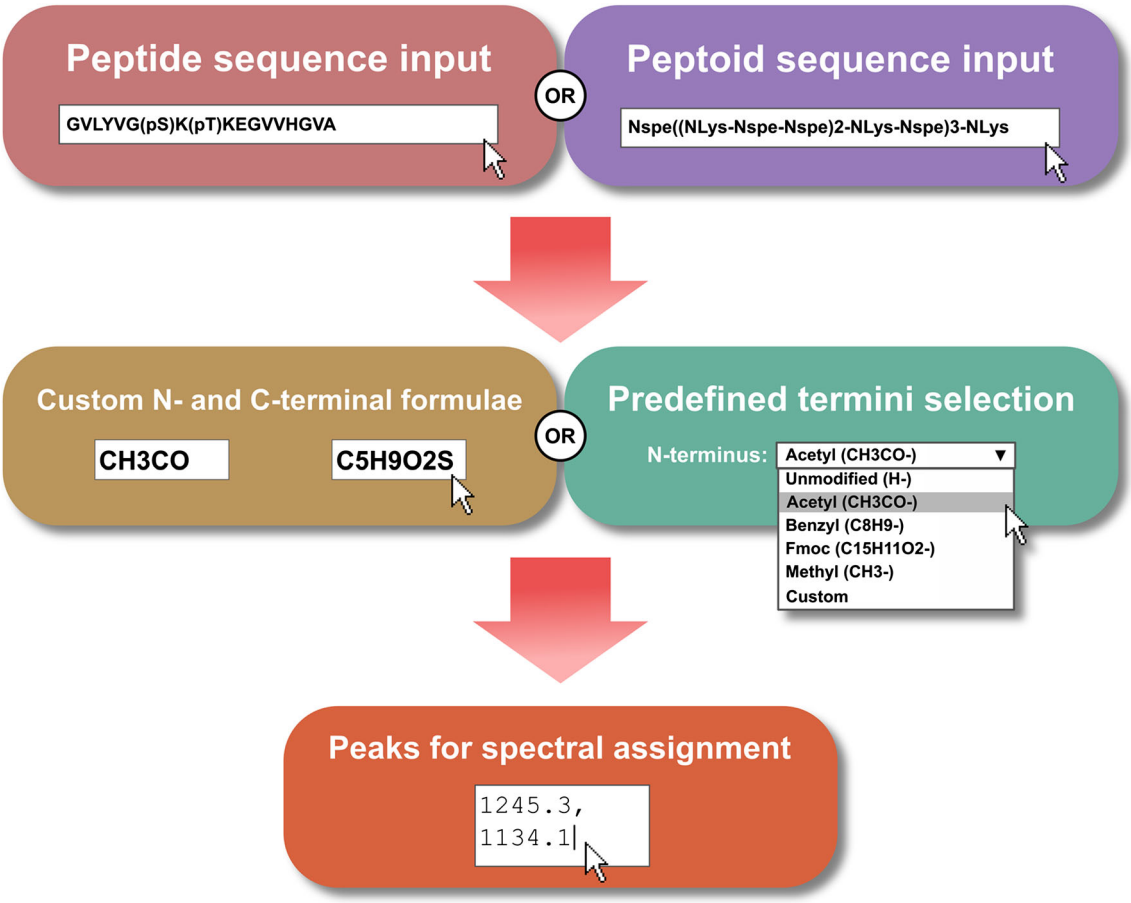



**Fig. 1** Summary of input options available for Peptide Calculator and Peptoid Calculator. Sequences can be specified using a large variety of residue types, and Peptoid Calculator also accepts input strings containing repeating sequence motifs indicated by *nested parentheses*. Termini formulae can be selected from available options and are also fully customizable. Optionally, *m/z* values can be specified for automatic peak assignment

280 nm), as well as a plot of calculated $\beta$-strand propensity for the sequence. Both utilities are available at http://www.pep-calc.com.

## Features summary

### Sequence input

Peptide and peptoid sequences up to 150 residues in length can be entered, containing any combination of amino acids or peptoid building blocks present in the database. For peptides, the input string may include any of the standard single-letter amino acid codes in addition to a number of 'nonstandard' residues (such as phosphoserine, *pS*), which must appear in parentheses within the string. An equivalent set of single-letter codes does not exist for peptoid building blocks, therefore Peptoid Calculator instead accepts a string of residue codes separated by dashes, without the requirement for multiple-letter codes to be enclosed in brackets. As peptoids can often consist of repeating motifs, Peptoid Calculator additionally allows parentheses to be used to indicate repeat sequences within the input string. Peptide and peptoid sequence input options are summarized in Fig. 1.

Both utilities also offer the option of specifying formulae for the N- and C-termini of the input sequence. These can be entered as a custom molecular formula string, or selected from lists of predefined formulae (Fig. 1). A full list of available residue types (showing residue code, molecular formula and molecular structure) and predefined termini available on Peptide Calculator and Peptoid Calculator is given on each site's Help page.

A final (optional) input field can be used to specify *m/z* values belonging to singly-charged species in mass spectra, for automatic assignment to peptide or peptoid deletion sequences and/or adducts (described below).

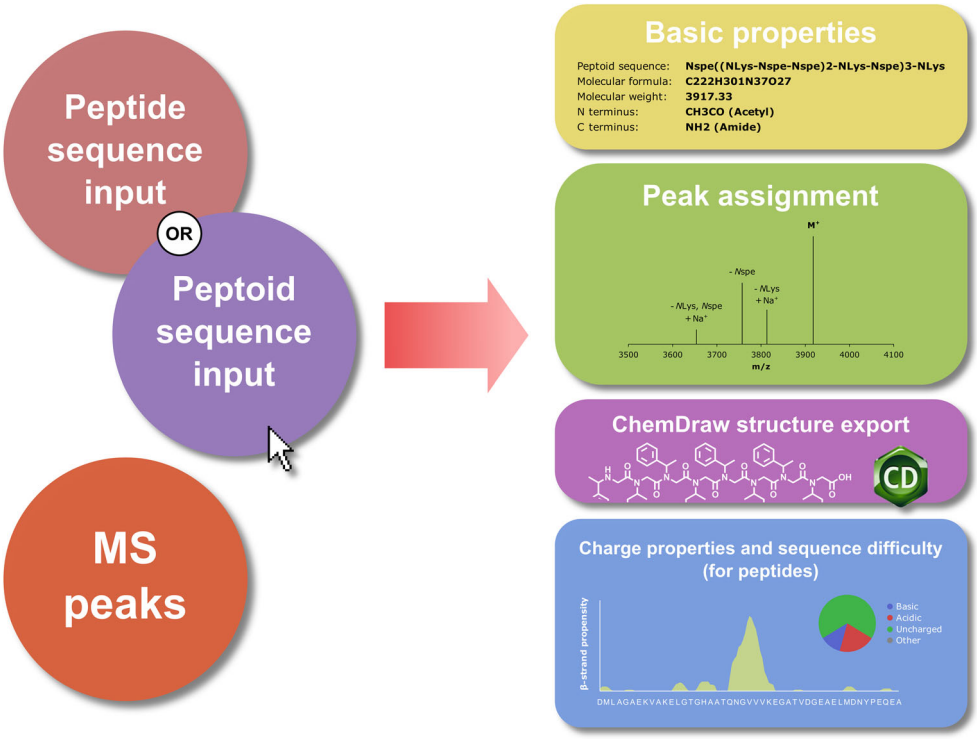



**Fig. 2** Both Peptide Calculator and Peptoid Calculator will output a number of basic calculated properties, in addition to a peak assignment and ChemDraw structure file for the sequence. A number of additional parameters are also provided for peptides, including estimated isoelectric point and molar extinction coefficient, as well as a graphical residue charge summary and $\beta$-strand contiguity profile

### Calculated parameters

Both utilities will provide a molecular formula and calculated molecular weight for peptide/peptoid sequences entered, in addition to an automatically generated ChemDraw structure in .cdxml format (Fig. 2). A spectral assignment for the peptide/peptoid will also be given if *m*/*z* values were provided as part of the input.

An example of an automatic peak assignment is illustrated in Fig. 3 (assignment output shown in Table 1). A number of peaks are present in the spectrum and have been assigned to either deletion sequences (where one or more residues are missing from the target sequence), sequences with unremoved protecting groups, metal adducts or a combination of two or more of the conditions described. Peptide Calculator and Peptoid Calculator will attempt to assign any *m*/*z* values provided to either the target sequence or a formula containing single or multiple residue deletions, metals, unremoved protecting groups or any combination thereof.

A number of calculated parameters specific to peptides are also available. Peptide Calculator will provide estimated values for sequence isoelectric point and molar extinction coefficient (at 280 nm), as well as a pie chart summarizing proportions of acidic, basic and uncharged residues in the sequence (Fig. 2). For sequences that are 10 residues or longer in length, a $\beta$-strand contiguity profile is calculated (Fig. 2). This provides an ab initio prediction of the location of $\beta$-strand forming regions within the sequence, and hence may offer an indication of aggregation-prone sequences, or those which are likely to present difficulties during synthesis.

## Methods

Peptide Calculator and Peptoid Calculator make use of a database each containing either amino acids or peptoid building blocks defined by residue codes (single- or multiple-letter) and accompanying molecular formulae. Molecular weights are calculated by reference to a table of atomic masses (most abundant isotope). Methods used to generate other calculated parameters are described below. All Pep-Calc.com functionality is scripted using an extensible framework written in the Python programming language, and the site is accessed using an HTML web interface. Residue formulae can be added to either database upon request.

### Isoelectric point and molar extinction coefficient calculation

Theoretical peptide isoelectric points are calculated using the bisection method described by Kozlowski [17–19]. The net charge of the peptide can be found using the Henderson–Hasselbalch equation, taking into account

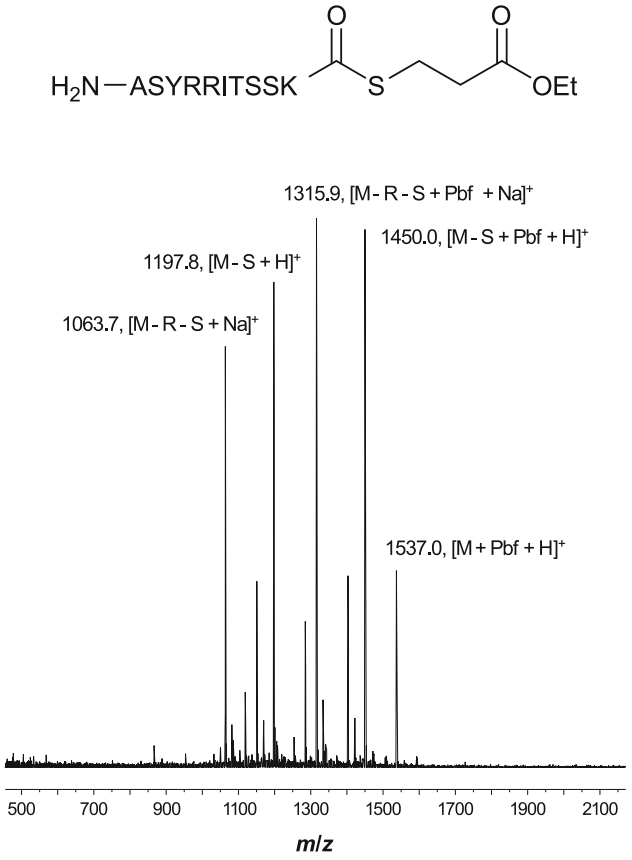



**Fig. 3** Example spectrum automatically assigned by Peptide Calculator (assignment is also available for Peptoid Calculator). A number of single- and multiple-residue deletions have been identified, in combination with sodiation and/or unremoved 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) protecting groups. The ethyl 3-mercaptopropionate thioester is available as a predefined C-terminus and can be selected during sequence input

**Table 1** Example peak assignment output (sequence and mass spectrum shown in Fig. 3). For each peak specified by the user, a set of all possible combinations of residue deletions and/or adducts agreeing with that *m/z* value are provided, in addition to a calculated mass for each suggestion

| Peak (*m/z*) | Deletion(s) | Adduct(s) | Calculated mass |
|---|---|---|---|
| 1063.7 | R, S | $Na^+$ | 1063.52 |
| 1197.8 | S | – | 1197.64 |
| 1315.9 | R, S | $Na^+$, Pbf | 1316.52 |
| 1450.0 | S | Pbf | 1450.64 |
| 1537.0 | – | Pbf | 1537.67 |

contributions from negatively and positively charged groups (first and second terms in Eq. (1) respectively, where $K_a$ is the acid dissociation constant of the amino acid).

$$charge = \sum \frac{-1}{1 + 10^{pK_a - pH}} + \sum \frac{1}{1 + 10^{pH - pK_a}} \quad (1)$$

As the isoelectric point (pI) represents the pH at which the net charge of the peptide equals zero, finding the root of this equation (in this case numerically, using the bisection method) gives the pI (or pH at zero charge).

Peptide Calculator takes into account side chain charge contributions from Arg, Asp, Cys, Glu, His, Lys and Tyr residues, in addition to the N-terminal amine and C-terminal carboxyl groups (only if the terminus types are set to 'Unmodified' and 'Acid' respectively). Other residue side chains are not taken into account for pI estimation, and are designated 'Other' in the charge summary pie chart.

Molar extinction coefficients are estimated using Eq. (2), described by Pace et al. [20]. The formula takes into account numbers of Trp and Tyr residues in the peptide ($n_{Trp}$ and $n_{Tyr}$ respectively), in addition to the number of *cystine* residues ($n_{cystine}$) formed via disulfide bond formation between pairs of cysteine side chains (reduced cysteine residues do not contribute significantly to the absorbance above 275 nm [20]).

$$\varepsilon_{280}\ \mathrm{M^{-1}cm^{-1}} = 5500 n_{Trp} + 1490 n_{Tyr} + 125 n_{cystine} \quad (2)$$

Peptide Calculator outputs two values for $\varepsilon_{280}$, calculating the theoretical molar extinction coefficient based on either formation of the maximum number of disulfide bonds possible ($n_{cystine}$ equal to the number of cysteine residue *pairs*), or complete reduction resulting in the absence of disulfides ($n_{cystine} = 0$).

## Automatic mass spectral peak assignment

User-entered *m/z* values are assigned through the process summarized in the flowchart given in Fig. 4. Pep-Calc first compiles lists of possible single-amino-acid deletions and single modifications (metal adducts and unremoved protecting groups), including null entries for no deletion or no modification. A complete set of combinations of these lists is then generated, and the molecular weight of the peptide/peptoid sequence incorporating each combination of single deletion and/or single modification calculated. Each input peak is then compared against the list of molecular weights, and a peak is assigned to a particular peptide if it falls within ±1.0 u of the calculated molecular weight of the peptide.

In the event that all the input peaks are not assigned on the first pass, Pep-Calc calculates the molecular weights for all peptide/deletion/modification combinations incorporating single *or double* deletions and single *or double* modifications, and checks remaining peaks against these (omitting already assigned peaks). This process is repeated until all peaks are assigned, or until up to the maximum allowed number of deletions/modifications have been checked. To prevent excessive computation times the

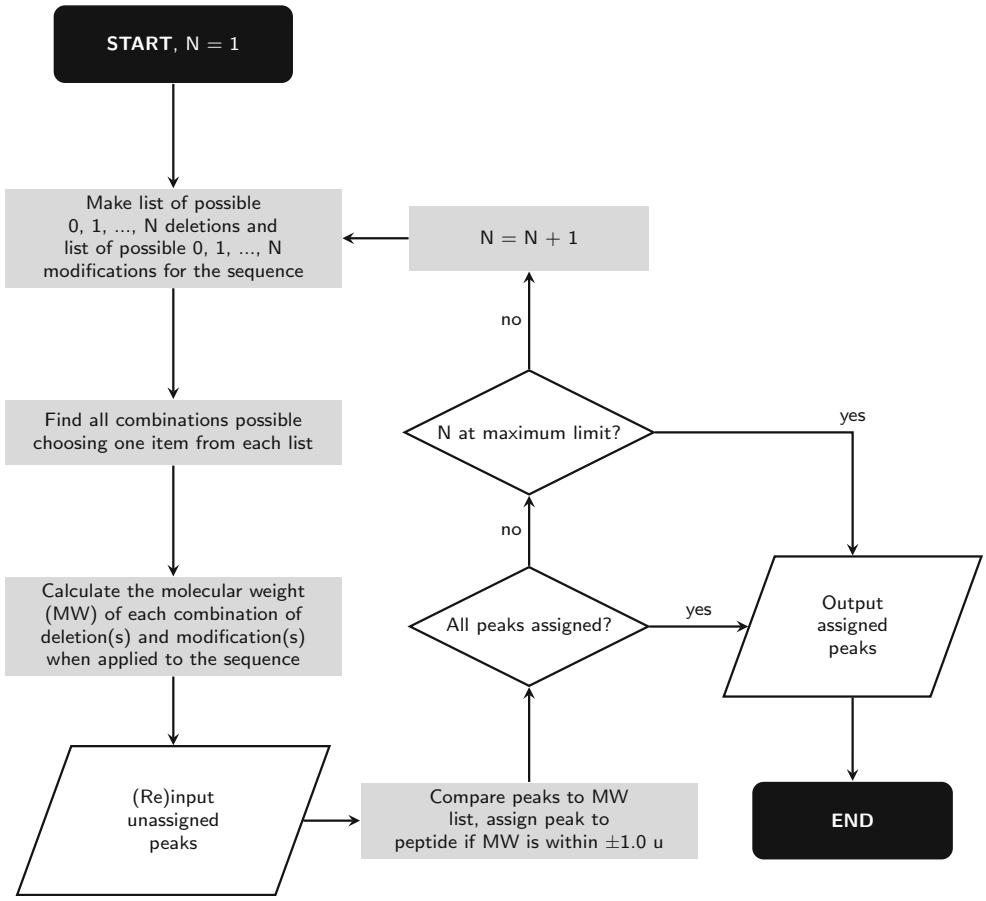



**Fig. 4** Flowchart summarizing the mass spectral peak assignment algorithm used by Peptide Calculator and Peptoid Calculator. Residues missing from the expected full sequence are termed 'deletions' and any other atom or group that causes a change in the molecular weight of the sequence (including metal adducts and unremoved protecting groups) is termed a 'modification'. Which deletions and modifications are allowed depends on the residues present in the sequence (unremoved Pbf protecting groups, for example, are only permitted for Arg residues). Only sequences bearing a single deletion and/or a single modification are considered on the first iteration (N = 1), increasing to two of each on the second etc. The maximum allowed value for N depends on the length of the input sequence, and is set at 5 iterations for sequences up to and including 30 residues in length, 4 up to 60 residues and 3 up to the maximum 150 residues

maximum number of deletions/modification depends on the sequence length, and is set at 5 iterations for sequences up to 30 residues in length, 4 for 60-mers and 3 up to the maximum 150 residue sequence input.

## Calculation of sequence β-strand propensity

β-Strand contiguity profiles for peptides greater than 9 residues in length are calculated using an implementation of the simple algorithm for sliding averages (SALSA) described by Zibaee et al. [21]. A window of size 4 residues is scanned across the input sequence and each fragment within the window scored using Eq. (3), where $P_\alpha$, $P_\beta$ and $P_t$ are the Chou–Fasman secondary structure probability parameters (for α-helix, β-strand and reverse turn preference, respectively) [22]. This process is repeated for all window sizes up to 20 residues or the sequence length (whichever is reached first), and all fragments with scores lower than 1.2 are discarded.

$$fragment\ score = \frac{\sum P_\beta}{\frac{1}{2}(\sum P_\alpha + \sum P_t)} \quad (3)$$

β-Strand propensity values are then calculated for each residue in the sequence by summing the scores of all remaining windows which contain the residue. These final values are then plotted to produce a β-strand contiguity profile for the peptide. Chou–Fasman parameters are only available for the 20 canonical amino acids and hence only these are taken into account when calculating β-strand propensity values.

It should be noted that β-strand propensity alone may not be indicative of aggregation likelihood or sequence difficulty. In addition, ab initio secondary structure prediction methods based on probability parameters alone can

in some cases give false predictions or fail to predict regions of a given secondary structure. SALSA was chosen with speed in mind, and for this reason the calculated profile is intended to serve only as a guide.

## Conclusions

Peptide Calculator and Peptoid Calculator form a set of full featured, freely available web utilities for peptide and peptoid molecular property calculation and mass spectral peak assignment. Modern peptide research demands tools that can handle residue types beyond the canonical amino acids (such as phosphorylated peptide building blocks [23–25]), and with unique spectral assignment capabilities and an expanding amino acid database Peptide Calculator offers a service beyond that of current freely available web utilities. Furthermore, similar services for peptoid research are non-existent, and Peptoid Calculator represents a first-of-a-kind resource for researchers in the field of peptoid science. The tools described have found broad application in our lab, and are used frequently in peptide and peptoid research activities [26–28]. It is anticipated that PepCalc.com (http://www.pep-calc.com) will act as a valuable asset to those synthesizing and/or using peptides or peptoids as part of their research in the biophysical and life sciences fields.

**Acknowledgments** The authors wish to thank Hannah Bolt for the compilation of peptoid building block data, and Ehmke Pohl for assistance during preparation of the manuscript.

**Funding** Financial support was provided by the Engineering and Physical Sciences Research Council (EPSRC).



## References

1. ExPASy: SIB Bioinformatics Resource Portal. http://www.expasy.org. Accessed 03 Dec 2015
2. Patiny L, Borel A (2013) ChemCalc: a building block for tomorrow's chemical infrastructure. J Chem Inf Model 53:1223–1228
3. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B (2014) PredictProtein—an open resource for online prediction of protein structural and functional features. Nucl Acids Res 42:W337–W343
4. de Carvalho RV, Lopez-Ferrer D, Guimarães KS, Lins RD (2013) IMSPeptider: a computational peptide collision cross-section area calculator based on a novel molecular dynamics simulation protocol. J Comput Chem 34:1707–1718
5. Vlachopanos A, Soupsana E, Politou AS, Papamokos GV (2014) POTAMOS mass spectrometry calculator: computer aided mass spectrometry to the post-translational modifications of proteins. A focus on histones. Comput Biol Med 55:36–41
6. Claros MG, von Heijne G (1994) TopPred II: An improved software for membrane protein structure predictions. Comput Appl Biosci 10:685–686
7. Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, Minor W (2014) Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. Nat Protoc 9:156–170
8. Arenas-Salinas M, Ortega-Salazar S, Gonzales-Nilo F, Pohl E, Holmes DS, Quatrini R (2014) AFAL: a web service for profiling amino acids surrounding ligands in proteins. J Comput Aided Mol Des 28:1069–1076
9. Peptide Mass Calculator. http://immweb.vet.uu.nl/P&P_fac/pepcalc.htm. Accessed 04 Dec 2015
10. Peptide Property Calculator. http://www.basic.northwestern.edu/biotools/proteincalc.html. Accessed 04 Dec 2015
11. Fragment Ion Calculator. http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html. Accessed 04 Dec 2015
12. N2.cz Peptide Calculator. http://pept.n2.cz/. Accessed 04 Dec 2015
13. Oligonucleotide- and Peptide calculations. http://www.chemie.hu-berlin.de/seitz/oligo-tools_e.htm. Accessed 04 Dec 2015
14. Sheffield ChemPuter. http://winter.group.shef.ac.uk/chemputer/. Accessed 04 Dec 2015
15. PepCalc.com—Innovagen peptide property calculator. http://pepcalc.com/. Accessed 22 Jan 2016
16. GenScript Peptide Property Calculator. https://www.genscript.com/ssl-bin/site2/peptide_calculation.cgi/. Accessed 22 Jan 2016
17. Kozlowski LP (2015) Calculation of protein isoelectric point. http://isoelectric.ovh.org. Accessed 23 Sept 2015
18. Cameselle JC, Ribeiro JM, Sillero A (1986) Derivation and use of a formula to calculate the net charge of acid-base compounds. Its application to amino acids, proteins and nucleotides. Biochem Educ 14:131–136
19. Sillero A, Maldonado A (2006) Isoelectric point determination of proteins and other macromolecules: oscillating method. Comput Biol Med 36:157–166
20. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4:2411–2423
21. Zibaee S, Makin OS, Goedert M, Serpell LC (2007) A simple algorithm locates $\beta$-strands in the amyloid fibril core of $\alpha$-synuclein, A$\beta$, and tau using the amino acid sequence alone. Protein Sci 16:906–918
22. Chou PY, Fasman GD (1974) Conformational parameters for amino acids in helical, $\beta$-sheet, and random coil regions calculated from proteins. Biochemistry 13:211–222
23. Chan C-F, Lan R, Tsang M-K, Zhou D, Lear S, Chan W-L, Cobb SL, Wong W-K, Hao J, Wong W-T, Wong K-L (2015) Directional Plk1 inhibition-driven cell cycle interruption using amphiphilic thin-coated peptide-lanthanide upconversion nanomaterials as in vivo tumor suppressors. J Mater Chem B 3:2624–2634
24. Chan C-F, Xie C, Tsang M-K, Lear S, Dai L, Zhou Y, Cicho J, Karbowiak M, Hreniak D, Lan R, Cobb SL, Lam MH-W, Hao J, Wong K-L (2015) The effects of morphology and linker length on the properties of peptide-lanthanide upconversion nanomaterials as G2 phase cell cycle inhibitors. Eur J Inorg Chem 2015:4539–4545. doi:10.1002/ejic.201500321

25. Li H, Chan C-F, Chan W-L, Lear S, Cobb SL, Mak N-K, Lau TC, Lan R, Wong W-K, Wong K-L (2014) Monitoring and inhibition of Plk1: amphiphilic porphyrin conjugated Plk1 specific peptides for its imaging and anti-tumor function. Org Biomol Chem 12:5876–5882
26. Bolt HL, Cobb SL (2016) A practical method for the synthesis of peptoids containing both lysine-type and arginine-type monomers. Org Biomol Chem 14:1211–1215
27. Eggimann GA, Bolt HL, Denny PW, Cobb SL (2015) Investigating the anti-leishmanial effects of linear peptoids. ChemMedChem 10:233–237
28. Eggimann GA, Sweeney K, Bolt HL, Rozatian N, Cobb SL, Denny PW (2015) The role of phosphoglycans in the susceptibility of *Leishmania mexicana* to the temporin family of antimicrobial peptides. Molecules 20:2775–2785