



UNIVERSITAT DE
BARCELONA

Development of synthetic strategies for lasso peptides with anticancer activity

Helena Martín Gómez



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – SenseObraDerivada 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – SinObraDerivada 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-NoDerivs 3.0. Spain License.**



UNIVERSITAT DE
BARCELONA



INSTITUTE
FOR RESEARCH
IN BIOMEDICINE

Programa de Química Orgànica

Tesi Doctoral

**DEVELOPMENT OF SYNTHETIC STRATEGIES
FOR LASSO PEPTIDES
WITH ANTICANCER ACTIVITY**

Helena Martín Gómez

Dirigida i revisada per:

Dr. Judit Tulla Puche
(Universitat de Barcelona)

Dr. Antoni Riera Escalé
(IRB Barcelona)

Departament de Química Orgànica

Facultat de Química

Universitat de Barcelona

Barcelona, 2018

*Mai t'adones del que ja has fet,
només pots veure el que et queda per fer.*

Marie Curie

Agraïments

En primer lloc vull agrair a la **Judit** i al **Fernando** per confiar en mi i haver-me donat la oportunitat de començar aquest gran repte, la tesi doctoral. Encara recordo la primera vegada que ens vam reunir al despatx i em vaig parlar dels lasso peptides... i quatre anys més tard, aquíestic, escrivint aquestes paraules. Agrair també al **Toni**, per haver-me acollit com una més al seu grup i ajudat a que aquesta tesi sigui possible.

Quatre anys es diuen ràpid, però donen per molt. Quan vaig començar no imaginava aquest moment. Hi ha hagut dies millors i dies pitjors, cada dia ha estat un aventura, però això és la tesi. Tota aquesta etapa no hagués estat el mateix sense els grans companys del laboratori que he tingut al costat. **Juan**, mis inicios no hubieran sido lo mismo sin ti. Tus consejos y ayudas han servido para aprender y llegar hasta aquí, gracias. **Ivan**, la persona con la que más cerca he estado estos últimos tiempos y casi compartiendo ordenador. Gran compañero de batallitas, cotilleos, bromas, charlas, pero sobretodo gran amigo. Siempre dispuesto a ayudar, a aportar tranquilidad y buen rollo al lab. Ibiz en el corazón quedará siempre pa los restos. **Nuria**, sort que vas arribar al lab, per fi una noia més! Vas aportat aire fresc i molt bon rotllo. Sempre atenta i disposada a donar-me un cop de mà. Amb les dues persones amb qui vaig iniciar aquesta etapa i ara l'acabarem plegats. Mitja tesi compartint lab amb cadascun de vosaltres. **Alex**, vaya vidorra! (no podía acabar sin volverlo a decir). **Judith**, per compartir moments d'estrès, d'enfadados i mooolts problemes amb l'UPLC. **Gerardo**, gran experto en el sintetizador! Gracias por esas increíbles purezas. ¿Has visto como me he portado bien este tiempo? Las birras y los frankfurts no han hecho descontrolarme. **Carlos**, **Omar**, les últimes incorporacions. Qui m'anava a dir que coincidiria amb vosaltres durant el doctorat. M'alegro d'haver compartit aquest últim any amb vosaltres. **Ioana**, for your annual visits, which brought fresh air and joy to the lab.

Recordar la gent que ja no hi és i amb la que vaig compartir els meus inicios. **Marta**, **Miriam**, la veu de l'experiència! Recordo els vostres consells sobre com "sobreviure" al doctorat i al lab. El Roger i el Martí estaran molt orgullosos de vosaltres! **Kamil**, el vecino de enfrente, que te fuistes a hacer las americas. Siempre dispuesto a hacer una birra y un Cangrejo. Gracias por la llama! **Laia**, pel teu non-stop i el teu gran somriure. **Anais**, recuerdo tu mal genio por intentar poner orden en

un laboratorio un tanto caótico. **Michael**, the crazy guy, aye! I still remember the first time that you entered to the lab, well that you asked me something... També als companys del Lab100 que van marxar més o menys quan jo arribava, **Janire, Adriana, Pau**.

Agrair a tots els companys del Lab300 amb els que vaig compartir els dos primers anys i que no han deixat mai de ser companys. Frau **Cris Fuster**, pel teu bon rotllo i sempre disposada a apuntar-te a tot i inclús a la immersió alemanya. El running team: **Pep**, el teacher de les cells. Gràcies per ensenyar-me el món de les cèl·lules, que deu ni do en quin merder em vaig ficar. **Macarena**, eres una gran científica y magnífica persona. Siempre se puede contar contigo y estás dispuesta a hechar una mano. Si la ciència no va bien, siempre nos quedará la costura i **Adam**, l'iniciador dels team buildings. Sempre amunt i avall, atent a tot, qualsevol cosa que es necessiti allà estàs tu, ets un crack! **Martí**, tu sempre a tope, ficat en mil coses a la vegada. Recta final, això ja ho tenim! **Monica, Cristina Diaz**, atentes i disposades a ajudar. **Salva**, a tu especialment, per la paciència, per estar al meu costat, recolzar-me i ajudar-me en tot moment. **Pol, Sonia**, els cafeteros oficials, pels moments que hem compartit plegats. Per tota la gent que en algun moment o altre han estat presents, **Julia, Montserrat, Txell, Toni T., Eduard, Silvia, Ester, Jan**.

Moltes gràcies a la **Marga**, per tota l'ajuda i consells en la part de RMN. Agrair també a la **Mireia**, l'ajuda i suport en els estudis de masses.

El meu pas pel IRB ha fet que creixés com a científica, però també com a persona. M'ha donat l'oportunitat de conèixer a molta gent i que ens convertíssim en una gran família. **Laura, Gemma**, mis pekeñajas! Que dir de vosaltres... me n'alegro moltíssim d'haver-vos conegut. La vostra paciència amb els meus bio-dubtes. Totes les aventures, riures i peripècies viscudes són extraordinàries, igual que vosaltres. Això ha estat només un inici! **Anna**, companya de pèptids i d'aventures. **Craig**, the oficial guiri, siempre con una sonrisa y una mano para ayudar. **Ricardo**, hablar contigo sin reir o hacer bromas es imposible! **Jürgen**, der Deutschlehrer. Wir begannen mit dem Unterricht, aber sie waren bald fertig. **Jordi**, pel teu carisma i les ganes d'organitzar coses. **Ernest**, els teus comentaris són mítics! El Santuari, un lloc espectacular, espero que puguem repetir la sortida. **Helena Roura**, molt bona cuinera i persona. La resta de URSA team: **Joan**, si la ciència no va enlloc, la pastisseria et portarà molt

lluny i **Amparo**, por nuestros encuentros en los microondas. Les companyes d'organització del retreat, **Celia, Teresa, Lada**, que no va ser feina fàcil.

Durant aquest temps he tingut l'oportunitat de fer una estada a Alemanya. I want to thank **Prof. Mohamed Marahiel** for giving me the opportunity to join his lab. I want to specially thank **Julian** very much during these months and for all your patience in teaching me. To all my lab colleagues, **Chris, Bastian, Zhu, Annika** and **Roman**. Thanks to **Antje**, for all the help and organization.

Natalia, Anna, les Trànsfugues! Companyes de màster i gran amigues. Les quedades per dinar, els cotis. Posar-nos al dia ha estat molt necessari! **Ari Grau**, la mestra, tu em vas ensenyar que era un pèptid ara fa sis anys. **Claudia**, vas aparèixer al lab d'un dia a un altre i vas acabar convertint-te en una gran amiga. Fins a Alemanya vam arribar! Anant més enrere, recordar als químics amb els quals vaig compartir una gran etapa. **Llucia**, la viatgera per excel·lència, des de que vas sortir de Menorca aquell setembre del 2008, no has parat! **Javi**, sempre tan enfeinat i nerviós, se que no tens remei! **Guillem**, encara recordo les dormides que et feies mentre analitzàvem espectres. **Marta**, haver anat de tardes va valer molt la pena! A les nenes: **Pili, Meri, Laura, Elena i Nuria**.

A las Palkue, **Jenni, Irene, Miriam, Andrea, Laura, Elia, Marta y Silvia**. Por ser como sois, porque medio siglo juntas ha hecho que crezcamos y aprendamos unas de las otras. ¡Gracias por estar ahí durante esta etapa y siempre!

Per últim, a aquelles persones que fa més temps que em coneixen, que sempre han estat quan les he necessitat, que m'han ajudat i recolzat en totes les situacions, que han fet que arribés on estic ara, als meus pares.

CONTENTS

ABBREVIATIONS	I
Proteinogenic amino acids	vii
Protecting groups	viii
Resins and linkers	viii
Coupling reagents and additives	ix
Synthetic peptides	x
INTRODUCTION AND OBJECTIVES	1
Natural products	3
Ribosomally synthesized and post-translationally modified peptides	6
Lasso peptides	7
Characteristics of lasso peptides	7
Biosynthesis of lasso peptides	11
Chemical synthesis	14
Chemoenzymatic approaches	17
Structural elucidation	19
Objectives	26
References	27
CHAPTER 1: DESIGN AND SYNTHESIS OF LASSO PEPTIDES	37
Structure elucidation of sungsanpin	39
Towards the synthesis of sungsanpin	40
General considerations	40
Protection scheme	40
Polymeric support	41
Coupling conditions	41
Ester bond conditions	42
Synthesis of the natural sungsanpin	43
First approach	43
Second approach	44
Synthesis of the bicyclic analogs via ester bond	46
Strategy A	46
First approach	49
Second approach	53

Strategy B.....	60
Strategy C.....	68
First approach	69
Second approach	70
Strategy D.....	72
First approach	73
Second approach	77
Third approach	81
Fourth approach.....	83
Synthesis of the branched cyclic analog.....	87
Synthesis of the bicyclic analogs via disulfide bond.....	88
Strategy 1: Cys3 - Cys13	88
Strategy 2: Cys3 - Cys15	91
Other synthetic bicyclic analogs	93
Bicyclic analogs side-chain-to-tail via amide bond	93
Discussion.....	95
References	96
CHAPTER 2: STRUCTURAL ELUCIDATION AND BIOLOGICAL EVALUATION ...	99
Chemical comparison.....	101
Stability assays	104
Crystallization	110
MS/MS analysis	111
IM-MS analysis.....	119
Conformational analysis	129
NMR spectroscopy and structural analysis	131
Synthetic sungsanpin	131
Spectral validation.....	131
Structure calculation	134
Bicyclic sungsanpin analog A	136
Hydrolysis of the ester bond	143
Spectral comparison with natural lasso peptides	145
Structure calculation	154
Biological evaluation	157

General: mechanism of action	157
Cytotoxicity assay	158
Migration and invasion assay	160
Discussion.....	167
References	169
CHAPTER 3: BIOSYNTHESIS OF LASSO PEPTIDES	173
Genome mining.....	175
Homologous expression	180
Heterologous expression	183
DNA isolation and cloning.....	183
Mutagenesis.....	186
Chemoenzymatic approach	189
Discussion.....	191
References	192
CONCLUSIONS.....	195
EXPERIMENTAL SECTION.....	199
MATERIALS AND METHODS.....	201
For peptide synthesis and characterization	203
Solvents and reagents	203
Analysis and purification.....	203
Solid-phase peptide synthesis (SPPS)	205
Basic treatment	212
NMR spectroscopy.....	212
Ion-mobility mass spectrometry (IM-MS)	214
MS/MS analysis	214
Crystallography	215
Computational experiments.....	215
Circular dichroism (CD)	216
Capillary electrophoresis (CE)	216
For the biosynthesis of lasso peptides	217
Strains and general methods	217
Analysis and purification.....	217

Primers	218
Homologous production	222
Cryostock of spores	223
DNA isolation	223
Gibson-assembly cloning	223
Slim mutagenesis	224
Heterologous expression	225
Chemoenzymatic approach.....	227
For the biological assays.....	228
Reagents	228
Enzymatic digestions and stability	228
Cell culture	229
Cytotoxicity assay	229
Transwell migration and invasion assay	230
Minimum inhibitory concentration (MIC)	230
Synergy study.....	231
References	232
PRODUCT OBTENTION AND CHARACTERIZATION	233
Amino acids synthesis	235
Peptide synthesis	246
ANNEX: MICROCIN J25, A BASIC-SENSITIVE LASSO PEPTIDE	271

ABBREVIATIONS

aa	amino acid
ACH	α -cyano-4-hydroxycinnamic acid
ACN	acetonitrile
AcOH	acid acetic
Alloc	allyloxycarbonyl
ATCC	american type culture collection
ATP	adenosine triphosphate
AtxE2	isopeptidase from astexin-2 and -3
BIPY	2,2'-bipyridine
Boc	<i>tert</i> -butoxycarbonyl
Bzl	benzyl
CCS	collision cross section
CD	circular dichroism
CE	capillary electrophoresis
CID	collision-induced dissociation
COSY	H-H correlation spectroscopy
CptB2	B2-gene of chaxapeptin
CuAAC	copper-catalyzed azide-alkyne cycloaddition
δ	chemical shift
Da	dalton
DCE	dichloroethane
DCM	dichloromethane
DEA	diethylamine
DIAD	diisopropyl azodicarboxylate
DIEA	<i>N,N</i> -diisopropylethylamine
DIPCDI	diisopropylcarbodiimide
DKP	diketopiperazine
DMAP	4-dimethylaminopyridine
DME	dimethoxyethane
DMEM	dulbecco's modified eagle medium
DMF	<i>N,N</i> -dimethylformamide
DMSO	dimethylsulfoxide
DTIMS	drift-time ion mobility spectrometry
ECD	electron capture dissociation
ECM	extracellular matrix
EDC	<i>N</i> -(3-dimethylaminopropyl)- <i>N'</i> -ethylcarbodiimide hydrochloride
eq	equivalent
ESI	electrospray ionization
EtOAc	ethyl acetate
EtOH	ethanol
FA	formic acid
FAIMS	field-asymmetric ion mobility spectrometry

Abbreviations

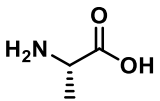
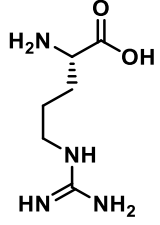
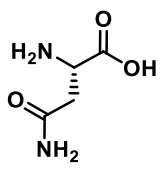
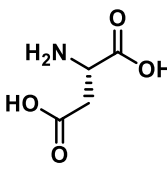
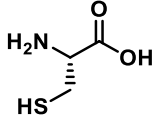
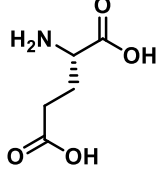
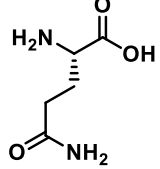
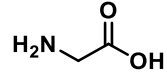
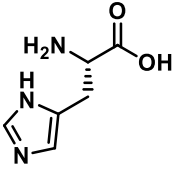
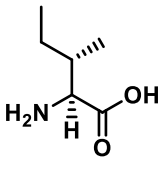
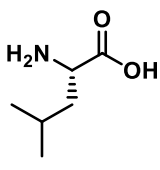
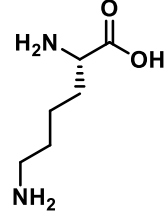
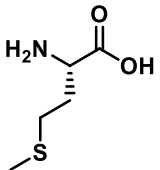
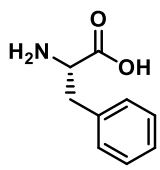
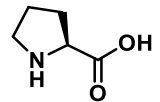
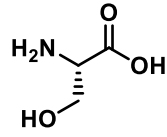
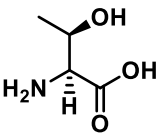
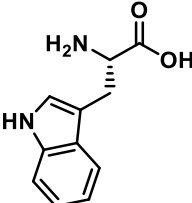
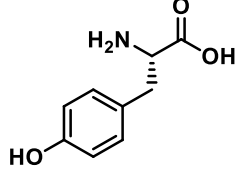
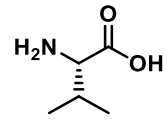
FBS	fetal bovine serum
FDA	US food and drugs administration
FIC	fractional inhibitory concentration
FIC _i	fractional inhibitory concentration index
Fm	9-fluorenylmethyl ester
Fmoc	9-fluorenylmethyloxycarbonyl
FP	forward primer
FWHM	full width at half maximum
GC	guanine-cytosine
GYM	glucose yeast media
h	hour
HATU	1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	hydrofluoric acid
HFIP	1,1,1,3,3,3-hexafluoro-2-propanol
HMBA	4-hydroxymethylbenzoic acid
HMBC	heteronuclear multiple bond correlation spectroscopy
HMPBA	3-(4-hydroxymethylphenoxy)propionic acid
HOAT	1-hydroxy-7-azabenzotriazole
HOSu	<i>N</i> -hydroxysuccinimide
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence spectroscopy
HTS	high-throughput screening
Hz	hertz
IM-MS	ion-mobility mass spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
IRMPD	infrared multiple photon dissociation
IsoP	isopeptidase
<i>J</i>	coupling constant
LB	lysogeny broth
LC-MS	liquid chromatography-mass spectrometry
<i>m/z</i>	mass charge relation
MALDI	matrix-assisted laser desorption/ionization
MccJ25	microcin J25
MEME	multiple EM for motif elicitation
MeOH	methanol
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MIC	minimum inhibitory concentration
min	minute
MMP	matrix metalloproteinase
MPD	4-pentenediol

MS	mass spectrometry
MS-MS	tandem mass spectrometry
NCI	US national cancer institute
NCL	native chemical ligation
NCS	<i>N</i> -chlorosuccinimide
NGS	next-generation DNA sequencing
NMP	<i>N</i> -methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NRPs	non-ribosomal peptides
NSCLC	non-small cell lung cancer
OD	optical density
Opfp	pentafluorophenyl ester
ORF	open reading frame
Osu	succinimide ester
PCR	polymerase chain reaction
PDA	photodiode array
PDB	Protein Data Bank
PEG	polyethyleneglycol
PfpOH	2,3,4,5,6-pentafluorophenol
2-Ph ⁱ Pr	2-phenyl isopropyl
pNZ	<i>p</i> -nitrobenzyloxycarbonyl
ppm	parts-per-million
PSI- BLAST	position-specific iterative basic local alignment search tool
pTsOH	<i>p</i> -toluenesulfonic acid
PyAOP	(7-azabenzotriazol-1-yloxy)trispyrrolidinophosphonium hexafluorophosphate
PyBOP	(benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
RBS	ribosomal binding site
RCM	ring-closing metathesis
REACH	registration, evaluation and authorization of chemicals
RiPPS	ribosomally synthesized and post-translationally modified peptides
RMSD	root mean square deviation
ROESY	rotating-frame Overhauser spectroscopy
RP	reverse primer
SLIM	site-directed ligase-independent mutagenesis
sp.	species
Spl-IsoP	isopeptidase from sphingopyxin I
SPPS	solid-phase peptide synthesis
tBu	<i>tert</i> -butyl
t _d	drift time

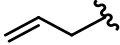
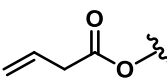
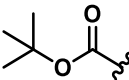

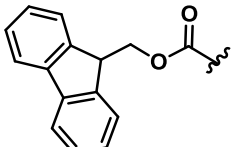
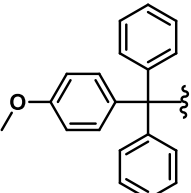
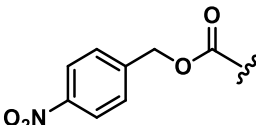
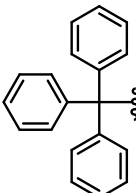
Abbreviations

TE	thioesterase
TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
THF	tetrahydrofuran
TIMP	tissue inhibitors of metalloproteinase
TIS	triisopropylsilane
Tmb	2,4,6-trimethoxybenzyl
TMSCl	chlorotrimethylsilane
TOCSY	total correlated spectroscopy
TOF	time-of-flight
T-P	terminator-promoter
t _R	retention time
TRIS	2-amino-2-(hydroxymethyl)propane-1,3-diol
Trt	trityl
TWIMS	travelling-wave ion mobility spectrometry
UPLC	ultra-high performance liquid chromatography
UV	ultraviolet
XTT	2,3-bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide

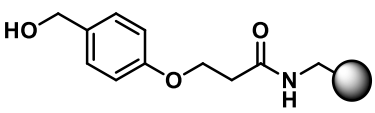
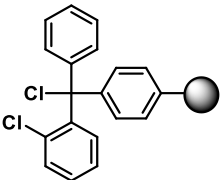
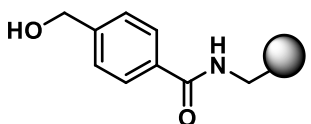
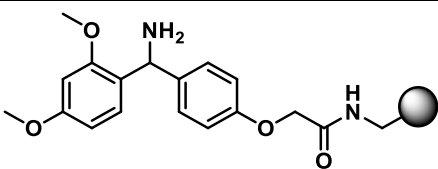
Proteinogenic amino acids

L-alanine, Ala, A	L-arginine, Arg, R	L-asparagine, Asn, N	L-aspartic acid, Asp, D
			
L-cysteine, Cys, C	L-glutamic acid, Glu, E	L-glutamine, Gln, Q	glycine, Gly, G
			
L-histidine, His, H	L-isoleucine, Ile, I	L-leucine, Leu, L	L-lysine, Lys, K
			
L-methionine, Met, M	L-phenylalanine, Phe, F	L-proline, Pro, P	L-serine, Ser, S
			
L-threonine, Thr, T	L-tryptophan, Trp, W	L-tyrosine, Tyr, Y	L-valine, Val, V
			

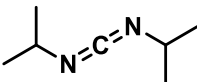
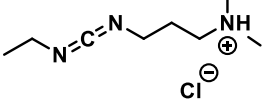
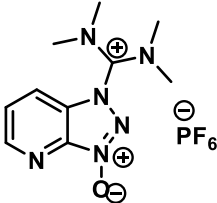
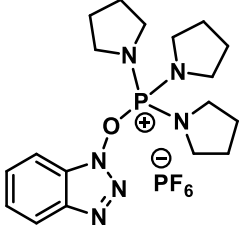
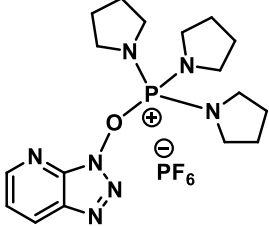
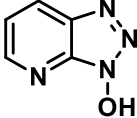
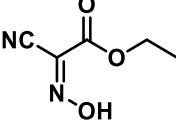
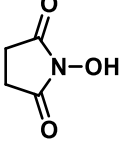
Protecting groups

allyl (OAll)	allyloxycarbonyl (Alloc)	<i>tert</i> -butoxycarbonyl (Boc)	<i>tert</i> -butyl (<i>t</i> Bu)
			
9-fluorenylmethoxycarbonyl (Fmoc)	4-methoxytrityl (Mmt)	<i>p</i> -nitrobenzyloxycarbonyl (pNZ)	trityl (Trt)
			

Resins and linkers

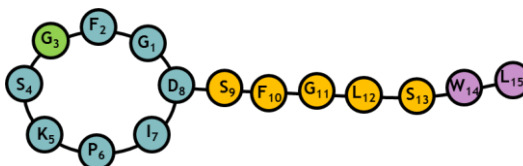
3-(4-hydroxymethylphenoxy)propionic acid (HMPBA) resin	2-chlorotrityl chloride (2-CTC) resin
	
4-hydroxymethylbenzoic acid (HMBA) resin	2-(4-(amino(2,4-dimethoxyphenyl)methyl)phenoxy)acetic acid (Rink amide) resin
	

Coupling reagents and additives

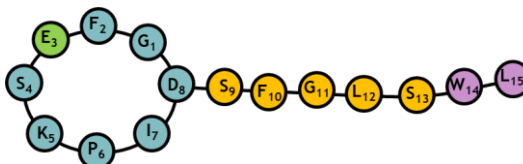
diisopropylcarbodiimide (DIPCDI)	1-ethyl-3-(3'-dimethylaminopropyl) carbodiimide (EDC·HCl)	1-[bis(dimethylamino)methylene]-1 <i>H</i> -1,2,3-triazolo[4,5- <i>b</i>]pyridinium 3-oxid hexafluorophosphate (HATU)	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP)
			
(7-azabenzotriazol-1-yloxy) trispyrrolidinophosphonium hexafluorophosphate (PyAOP)	1-hydroxy-7-azabenzotriazole (HOAt)	ethyl cyano(hydroxyimino)acetate (OxymaPure)	<i>N</i> -hydroxysuccinimide (HOSu)
			

Synthetic peptides

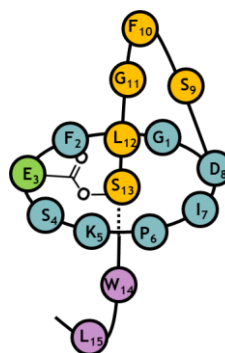
Synthetic sungsanpin



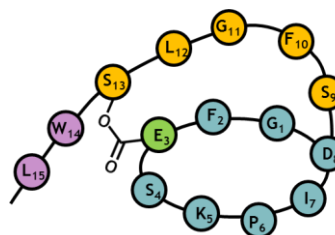
Branched cyclic analog



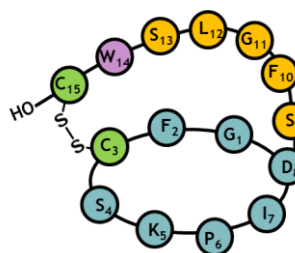
Bicyclic analog A



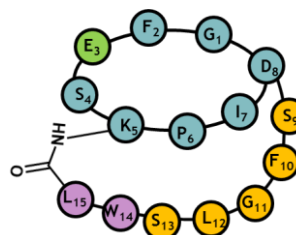
Bicyclic analog C and D



Bicyclic analog via disulfide bond
Cys3 - Cys15



Bicyclic analog side-chain-to-tail
via amide bond



INTRODUCTION AND OBJECTIVES

Natural products

For centuries, natural substances, particularly plants, have been used to manage and treat diseases, culminating in the discovery of the majority of modern pharmaceutical agents. For example, ancient Egyptians practiced medicine from as far back as 2900 BC.¹ Currently, it is estimated that over half the pharmaceuticals in clinical use are derived from natural products.² According to the US Food and Drugs Administration (FDA), in 2014 around 17% of the total number of approved drugs were isolated or derived from natural products and 25% were made by total synthesis, but the pharmacophore was from a natural product.³ Drugs based on natural products display a wide range of activities, including anticancer, antimicrobial, and antidiabetic action, and some of them are used to treat Alzheimer disease. Added to their diverse applications, natural products are more readily absorbed than synthetic drugs.⁴ Despite these advantages, many pharmaceutical companies have decreased their use in drug discovery screening due to the difficulties and complexities encountered with their isolation, chemistry, intellectual property rights and the inherent slowness of working with such compounds.⁵ The development of high-throughput screening (HTS) in the 1990s, together with combinatorial chemistry, allowed the generation of large libraries of compounds based on natural substances.⁶ This fact supposed an important improvement for the pharmaceutical industries. HTS allowed the rapid identification of a large number of active compounds that modulate a particular biomolecular pathway.

Plants are the oldest and most exploited source of natural products. In addition, these products have been found in fungi and terrestrial actinomycetes,⁷ the latter a particular type of Gram-positive bacteria found principally in soil. Terrestrial sources are the major focus of research efforts. Indeed, in the last two decades alone, more than 150 secondary metabolites from bacteria, fungi and plants have been identified.⁸ In particular, the genus *Streptomyces* is the major producer of bioactive compounds for the biotechnology industry.⁹ Published data reveal that 45% of all microbial bioactive secondary metabolites come from actinomycetes, and of these, 80% are produced by *Streptomyces* and show a wide range of bioactivities.¹⁰ Examples include antimicrobials (vancomycin and daptomycin), antifungals (amphotericin B), and anticancer agents (doxorubicin and epoxomicin)^{11,12} (Figure 1).

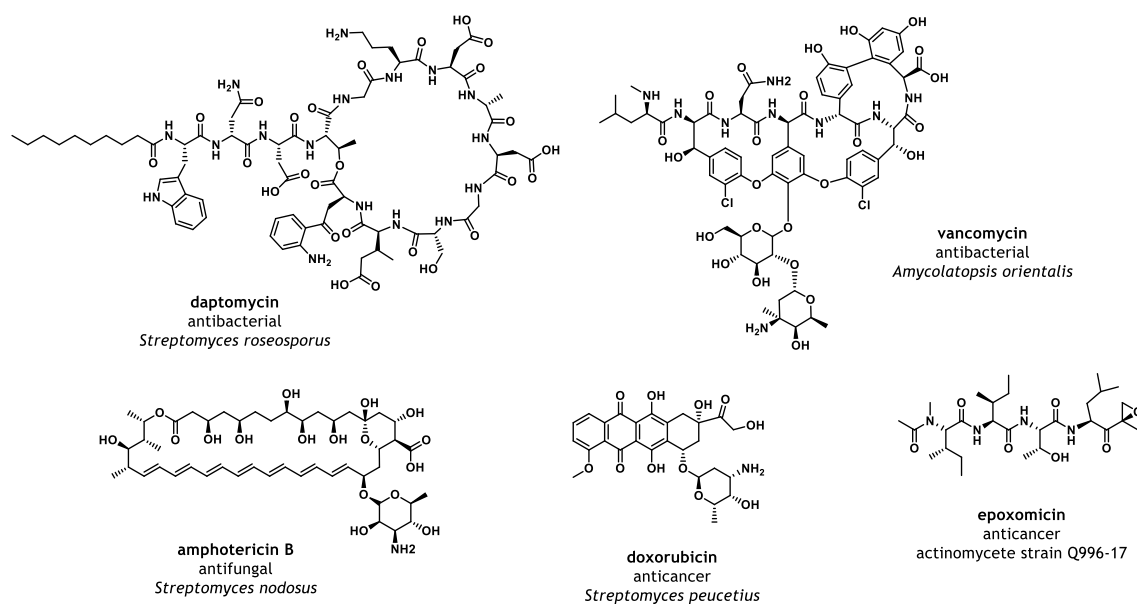


Figure 1: Examples of bioactive metabolites produced by Actinobacteria. (Adapted from Challis and coworkers)¹²

Regarding marine natural products, the first isolation of unusual nucleosides from a marine sponge near the coast of Florida was reported in 1951.^{13,14} Several uncharacterized compounds were found from marine sources and, compared to previously reported nucleosides from terrestrial sources, they displayed more potent biological activities. For example, Spongistatin (Figure 2), derived from tropical marine sponges, is the most active of all natural and synthetic compounds found by the US National Cancer Institute (NCI).¹⁵ However, drug discovery based on marine natural products is more challenging than that of terrestrial origin because these substances are located in places of difficult access and special equipment is required for their collection. Moreover, marine natural products are often isolated in low yields and the complexity of their structures makes chemical synthesis expensive and challenging. Despite these disadvantages, the development of new drugs from the marine environment is growing, especially those derived from marine actinomycetes, which produce novel bioactive metabolites.⁸ Taxonomically, diverse communities of actinomycetes have been detected in marine sediments, notably in deep-sea sediments.¹⁶ Chemical screening of novel *Streptomyces* isolates from unusual environments is therefore important for the discovery of new natural products.¹⁷

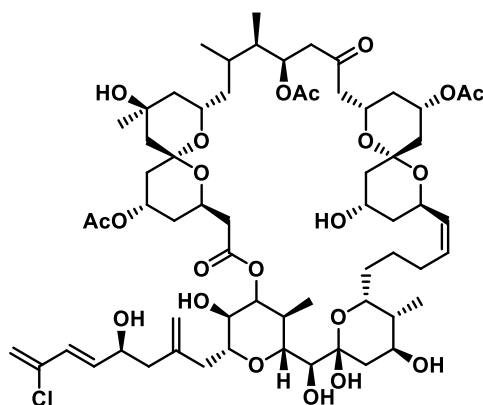


Figure 2: Structure of Spongistatin¹⁵

The first marine actinomycete, *Rhodococcus marinonascens*, was characterized in 1984¹⁸, but it was not until 2002, after the application of DNA sequence-based methods, that such actinomycetes were reported to differ from their terrestrial relatives.¹⁹ Before that, the existence of marine strains was questioned, hypothesizing that they derived from metabolically active spores in the sea.

The pioneering sequencing and analysis of the genomes of *Streptomyces coelicolor* A3(2)²⁰ and *Streptomyces avermitilis*²¹ revealed that actinomycetes have an unexpected abundance of natural biosynthetic gene clusters and thus, have the potential to produce a larger number of compounds than previously thought. The availability of these genomes allows comparison of genome sequences and secondary metabolite producers, thereby opening up new avenues in drug discovery.

Due to the exploitation of genomic sequencing, the field of natural products has been transformed during recent years. These genomes have been accumulated in public databases and *in silico* studies have been performed to discover new secondary metabolites.²² The availability of the genomes led to the genome mining approach for natural product discovery. Genome mining is a powerful technique that is used to discover the correlation between the primary genetic information and peptide function,²³ providing access to many novel biosynthetic pathways that would otherwise have remained undetected.¹² Moreover, it is useful for the analysis of novel and difficult microorganisms isolated from unexploited habitats such as oceans, deserts and surfaces of animals and plants.²⁴ Next-generation DNA sequencing (NGS) allows much faster DNA sequencing than the previous Sanger sequencing developed in 1977.²⁵

For several decades, the majority of secondary metabolites isolated from the marine ecosystem were terpenoids, alkaloids, polyketides and non-ribosomal peptides (NRPs), such as vancomycin or daptomycin (Figure 1). A main limitation of NRPs is that modifications in their biosynthetic clusters are highly complex as they depend on the ability of enzymes to recognize and modify substrates.²⁶ It was after the genome sequencing effort that another major class of secondary metabolites was recognized, namely ribosomally synthesized and post-translationally modified peptides (RiPPs).²⁷ This group includes lantibiotics,²⁸ cyclotides,²⁹ microcins³⁰ and lasso peptides.³¹

Ribosomally synthesized and post-translationally modified peptides

Ribosomal peptides can be formed only by proteinogenic amino acids, thus limiting their structural diversity. Nonetheless, post-translational modifications can be introduced, thereby leading to a wide range of products with features resembling those of non-ribosomal peptides.²⁷ A main advantage of ribosomal peptides compared to non-ribosomal peptides is that their sequences can be modified by simple mutation of a few codons.³²

In general, all RiPPs have similar structures and cluster organizations, as shown in Figure 3. Precursor peptides contain a core peptide, which is modified during the maturation step and usually flanked by sequences needed for enzymatic recognition. These modifications frequently include a proteolytic step, in which the precursor peptide is cleaved into a smaller active fragment. In addition, many of these precursor peptides contain sequences that can be recognized by enzymes. These sequences present a certain degree of variability. However, depending on enzyme selectivity, they continue to be substrates for specific post-translational modifications.³³

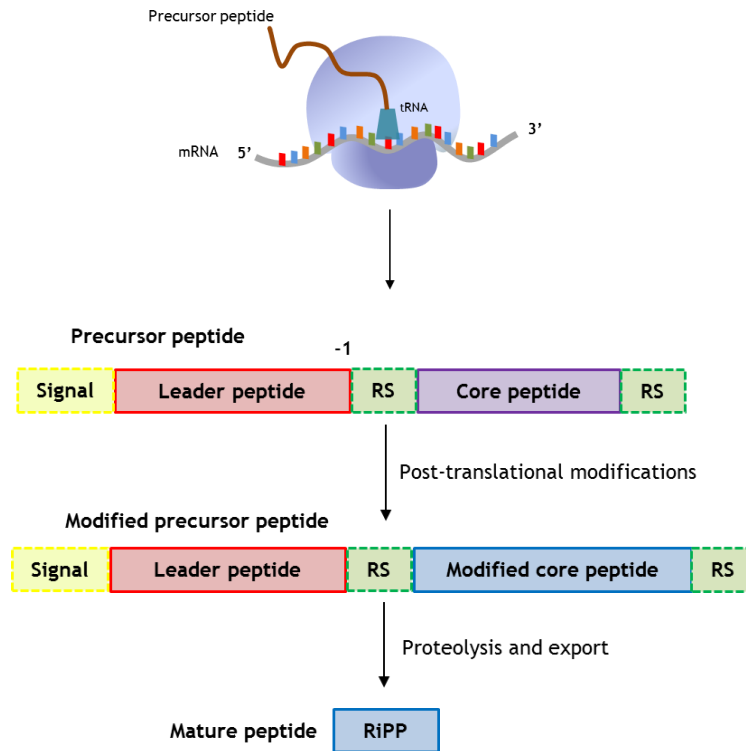


Figure 3: General synthetic pathway for RiPPs. RS = recognition sequence. (Adapted)^{33,34}

The biosynthesis of RiPPs is initiated by a precursor peptide, whose *N*-terminus contains a leader peptide that plays various roles during the biosynthesis of these molecules. For example, the leader peptide can act as a chaperone-like sequence that helps the precursor peptide to fold and stabilize. It can also serve as a recognition motif for post-translational modifications and as a protector to keep the precursor peptide inactive inside the host until proteolysis occurs.³⁵ Moreover, attached to the *C*-terminus of this leader sequence is a core peptide, which is transformed into the final product via proteolytic cleavage.

Lasso peptides

Characteristics of lasso peptides

Lasso peptides are a class of ribosomally synthesized and post-translationally modified natural products with a unique three-dimensional structure, in which the *C*-terminus threads through an *N*-terminal macrolactam ring in a right-handed conformation. These peptides consist of 15-26 proteinogenic amino acids and share an *N*-terminal 7- to 9- residue macrolactam ring where the *N*-terminal amino acid is

always glycine or cysteine and the amino acid that closes the ring is aspartic or glutamic acid. This rigid and folded structure confers lasso peptides with extraordinary stability against chemical, thermal, and proteolytic degradation.³¹ These features make them useful candidates as molecular scaffolds for drug design and epitope grafting.^{36,37} Considering this, it is possible to use a rational approach to further improve and optimize such a scaffold toward the generation of more potent and more selective bioactive compounds.

Generally, lasso peptides can be divided into three classes on the basis of the number of disulfide bridges present to maintain the lasso topology. Class I lasso peptides have two disulfide bonds, one involves the *N*-terminal cysteine and the second connects the ring to the tail. Class II lasso peptides, which are most frequently encountered, feature no disulfide bonds and the lasso topology is stabilized by steric interactions. Finally, class III, which is the newest and least abundant class, includes lasso peptides that have only one disulfide bond, which connects the ring to the tail³⁸ (Figure 4). Currently, a total of 43 lasso peptides have been described: 3 belong to class I, 39 to class II and 1 to class III.³¹ Regarding their origin, class I and III lasso peptides are extracted only from actinobacteria, while class II members derive mainly from proteobacteria.³⁹

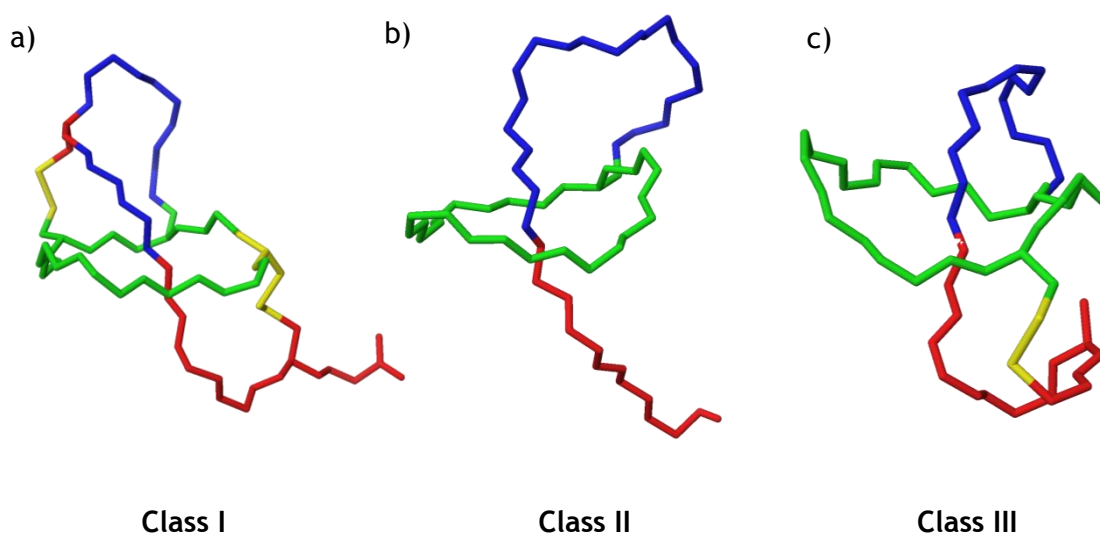


Figure 4: Backbones of lasso peptides of the three classes. The ring residues are shown in green, amino acids belonging to the loop in blue and the amino acids from the tail in red. Disulfide bridges are shown in yellow. a) Sviceucin (PDB code 2LS1), b) Caulosegnin (PDB code 2LX6) and c) BI-32169 (PDB code 3NJW).

The first lasso peptide, anantin,⁴⁰ was isolated from a strain of *Streptomyces coerulescens* in 1991. It was identified in a natural-product screening of antagonists of the atrial natriuretic factor (ANF) receptor. Two years later, RP-71955 was isolated as a new tricyclic peptide.^{41,42} Afterwards, the isolation and characterization of two lasso peptides with anti-HIV activity belonging to class I was published.⁴³ Around the same time, in 1992, the isolation of microcin J25 (MccJ25) was reported, incorrectly as a head-to-tail cyclic peptide.⁴⁴ However, it was not until 2003 that MccJ25 was demonstrated to be a lasso peptide in which the C-terminus was threaded through the ring.^{45,46,47} Regarding the biological activity of these peptides, a broad spectrum of pharmacological effects has been described. To date, they have been reported to act in the management of diabetes,⁴⁸ to inhibit HIV replication^{41,42} and to target cancer-involved receptors, such as the endothelin B receptor.⁴⁹ Moreover, some of these peptides show antimicrobial^{44,50,51} and receptor antagonist activity.^{52,53}

Prior to 2008, most of these lasso peptides were discovered by isolation from bacteria; however, capistruin, the first lasso peptide isolated by a genome mining approach, changed this scenario.⁵⁰ During the following years, several approaches to synthesize capistruin⁵⁴ and MccJ25^{55,56} were tested, but no similarity between the sequences of these compounds was found. This observation indicated that the lasso topology was responsible for the high stability and biological activity of these peptides.

In this thesis, we focus on the chemical synthesis of sungsanpin.⁵⁷ This lasso peptide was isolated from a *Streptomyces* sp. strain collected at a depth of 138 meters off the coast of Sungsanpo, in Korea, in 2012. Two years later, sungsanpin was isolated from the *Streptomyces* sp. 8LB7 collected from Atlantic Canadian marine sediment.⁵⁸ As mentioned before, lasso peptides are secondary metabolites and as such they can be produced in metabolic pathways that are not directly involved in the growth and development of the bacteria. This contrasts with primary metabolites, which are essential for the growth, development and reproduction of the microorganism.⁵⁹ Moreover, sungsanpin shows cytotoxicity in the human non-small cell lung cancer cell line A549 and was further demonstrated to inhibit the invasion of these cells. From a structural point of view, sungsanpin was the shortest lasso peptide and a unique member of lasso peptide family, until the isolation of chaxapeptin⁶⁰ from the *Streptomyces leeuwenhoekii* strain C58, found in the soil of

Atacama Desert, which was published in 2015. The complete characterization of these two lasso peptides was performed by NMR and the lasso structure was confirmed by performing NOE-based structural simulations, as well as by mass fragmentation and proteolytic assays.

From a taxonomic point of view, the sungsanpin-producing actinomycete *Streptomyces* strain SNJ013 is distantly related to *Streptomyces leeuwenhoekii* C58, and the two strains share 16S rRNA gene sequence similarity of 94%.⁶⁰ An analysis of their sequences reveals that the two lasso peptides are very similar, sharing the same number and character of residues (Figure 5). In this regard, they differ in only four residues, three of which are located at the C-terminal tail. In this position, NFF residues are replaced by SWL, thus maintaining hydrophobicity, and there is also a conservative change in the ring, namely Leu instead of Ile.

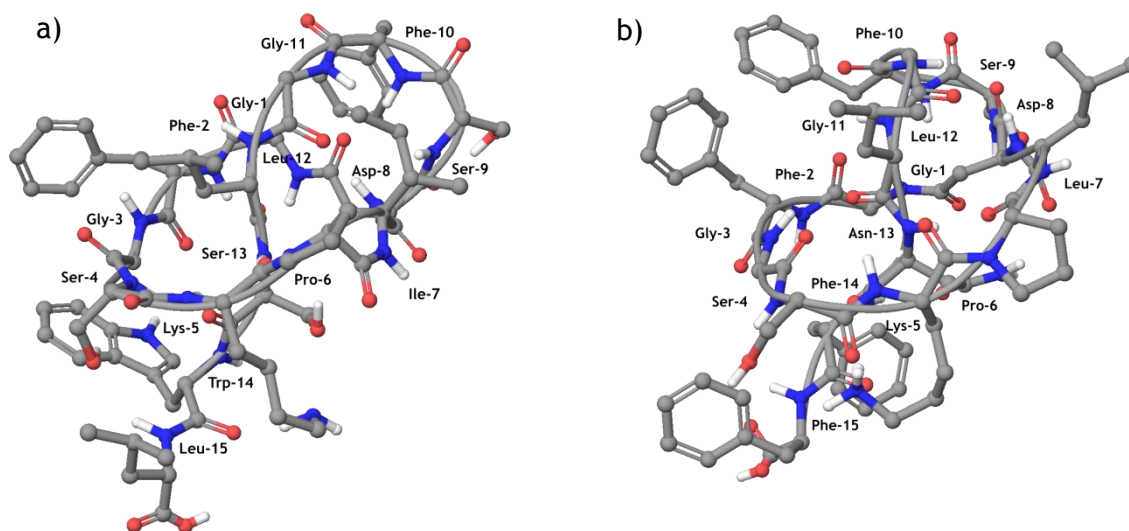


Figure 5: Structures of a) sungsanpin and b) chaxapeptin

Finally, chaxapeptin shows the same biological activity than sungsanpin, being inactive as an antimicrobial but exerting an inhibitory effect on the invasion of human lung cancer cells.

Biosynthesis of lasso peptides

Lasso peptide biosynthetic gene clusters consist of at least three genes, which encode the precursor peptide (A), an ATP-dependent cysteine protease (B) with homology to transglutaminase, and an ATP-dependent macrolactam synthetase (C) with homology to asparagine synthetase. Sometimes, a fourth gene encoding an ABC-transporter (D) can be found. This gene confers the lasso peptides with antimicrobial activity like for example in the case of microcin J25⁶¹ or capistruin.⁵⁰ However, it was identified by genome mining studies that lassomycin,⁶² lariatin^{63,64} or streptomycin⁶⁵ contain an additional open reading frame (ORF), called E-protein, which is an unknown protein, but essential for their biosynthesis (Figure 6). ORF is a section of a sequenced piece of DNA that contains a start codon (ATG normally) and stop codon (TAA, TAG, TGA). It has the potential to be transcribed into RNA and translated into protein.⁶⁶ Regarding this specific case, A, B, C, D and E are ORF because each of them are a DNA sequence with a start and stop codon and encode a specific protein.

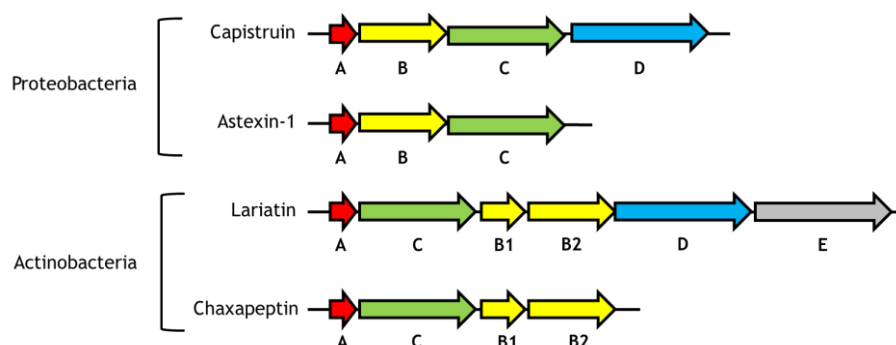


Figure 6: Schematic representation of the general organization of biosynthetic gene clusters for proteobacteria and actinobacteria (Adapted from Hegemann and coworkers).³¹

B-protein split is observed in different actinobacterial clusters,^{60,67} while in proteobacterial clusters B protein remains intact (Figure 6).^{68,69} B1 encodes the N-terminal domain while B2 encodes the C-terminal domain and it contains the catalytic triad characteristic of cysteine proteases. Both proteins have different functions, but they are complementary; B1 binds to the leader region of precursor peptide and then the substrate is delivered to B2 for processing and cleavage.⁷⁰ Whether the intact B protein evolved from the fusion of two genes, or the split B protein from division of one, is unknown.⁷¹ In any case, fusing the B1 and B2 proteins did not affect lasso peptide yield relative to the wild type system. On the other hand, the B protein of rubrivinodin system was split by introducing a stop codon.⁷² In

both cases, lasso peptide production was detected, revealing that both routes are feasible from an evolutionary standpoint and that the B protein activities could be assigned to distinct domains.

Actinobacteria are Gram-positive microorganisms and have high guanine-cytosine (GC) content in their DNA sequence, while proteobacteria are Gram-negative. Both type of bacteria differ mainly in the structure of their membrane. Gram-negative bacteria, but not Gram-positive, have an outer membrane formed by a lipopolysaccharide in the outer monolayer, which is held together by magnesium and calcium ions, that give stability to the membrane and it has anionic phospholipids, such as phosphatidylthanolamine and phosphatidylglycerol in the inner membrane.⁷³

Additionally, the main known lasso peptide biosynthetic gene clusters in proteobacteria have an ABC organization, while in actinobacteria an ACB formation is more prevalent (Figure 6), suggesting a common predecessor that diverged in parallel with the development of Gram-negative and Gram-positive cell walls.

The genetic sequence of MccJ25, the most studied lasso peptide, displays a gene cluster composed of four genes (*McjABCD*).⁶¹ In addition, it was revealed that during the maturation of lasso peptides the B-protein serves to cleave off the leader peptide from the precursor peptide using a source of ATP, thus providing suitable prefolding for the next step. In addition, *in vitro* studies of *McjB* and *McjC* pointed to a possible chaperone function for the B-protein.^{74,72} On the other hand, the C-protein catalyzes the ring formation reaction in a ATP-dependent manner, activating the carboxylic acid of the side chain. Finally, the mature lasso peptide is exported out of the cells by enzyme D (Figure 7).

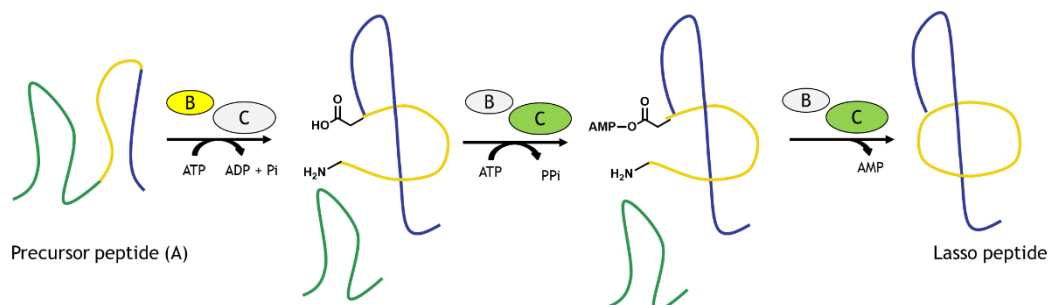


Figure 7: Proposed mechanism by which lasso peptides mature. (Adapted from Hegemann and coworkers)⁷⁵

As mentioned in the previous section, the first lasso peptide to be biosynthesized was capistruin, which has a gene cluster organization *CapABCD*.⁵⁴ In that study, Knappe and coworkers attempted to isolate the four key enzymes but failed due to the low solubility of the heterologously expressed enzymes. In this regard, the first *in vitro* assay to study the maturation enzymes of *MccJ25* was disappointing due to the low solubility and low expression levels,⁵⁵ thus hindering the development of a new strategy for lasso peptide synthesis. Five years later, Yan and coworkers modified the expression protocol to improve the production yields and were able to express and purify the enzymes *McjB* and *McjC*, together with some mutants, in order to identify the active site.⁷⁴ It was found that both enzymes are required for the lasso peptide maturation, generating a high affinity complex between B and C proteins. However, the enzymatic complex was not catalytically optimal under *in vitro* conditions and no lasso peptide was obtained; however, they synthesized two unthreaded topoisomers, called branched-cyclic peptides, for the first time and experimentally assigned the function of each enzyme.^{74,76} Branched cyclic peptides devoid of biological activity are susceptible to proteolytic degradation due to the mobile C-terminal tail (see Figure 8 for a schematic explanation).

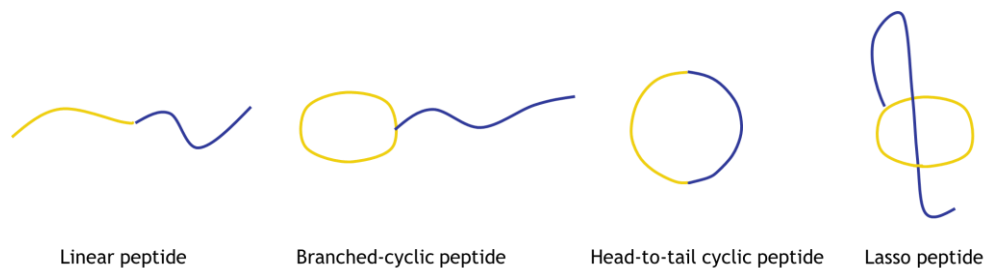


Figure 8: Schematic representation of different peptide structures.

As pointed out in the previous section, in 2008 there was a transition in the biosynthetic field, with the discovery of capistruin from *Burkholderia thailandensis* E264, the prediction of its genome sequence using genome mining, and the isolation of the desired molecule from culture supernatant. In addition, it was the first attempt to perform heterologous expression of this peptide in *E. coli*.⁵⁰ Genome mining for RiPPs is based on the homology search of the maturation enzymes that form the biosynthetic clusters of these natural products. For example, in the previous case of capistruin, enzymes *McjB* and *McjC* of *MccJ25* were used as the initial step toward finding the unknown biosynthetic cluster.^{54,56} Nowadays, more than half of the lasso peptides have been identified by genome mining approaches.⁷⁷

In general, lasso peptides are good targets for genome mining approaches mainly for two reasons. First, the ribosomal origin of the precursor peptides allows easy identification of the amino acid sequences of the mature lasso peptides and subsequent mass spectrometry characterization. The second reason is that B proteins share little homology to enzymes outside biosynthetic gene clusters of the lasso peptide and they constitute an ideal input for basic local alignment search tool (BLAST) homology searches to identify gene clusters.

Chemical synthesis

The diverse functionality of lasso peptides makes these molecules attractive candidates for drug development. In addition, given their high stability, these peptides are suitable scaffolds for epitope grafting approaches. However, to date, no lasso peptide has been synthesized using chemical protocols. This is explained by the highly challenging synthetic routes, especially due to the folding of peptides, which disturbs the necessary preorganization between the components to be assembled into an interlocked lasso molecule.

Regarding epitope grafting, cyclotides are the most well-known example of natural scaffolds used for grafting pharmaceutically relevant peptides and drugs.^{78,79,80} These molecules are head-to-tail macrocyclic peptides that comprise around 30 amino acids and three cysteine knots with a compact structure. Unlike lasso peptides, cyclotides are suitable candidates for solid-phase peptide synthesis (SPPS) via thioester-mediated strategy, native chemical ligation (NCL), in which the linear precursors are synthesized with a cysteine as *N*-terminal residue and a thioester at the *C*-terminus. When the last thioester bond is formed, there is a spontaneous *S,N*-acyl migration that gives the head-to-tail cyclization, followed by cysteine oxidation.^{81,82} Cyclotides show a wide range of biological activities and exceptional stability.⁸³ However, lasso peptides are emerging as promising candidates for epitope grafting, specifically MccJ25, which is the most suitable scaffold for epitope grafting because it has the longest loop, with 11 amino acid residues.^{36,37}

In addition, another type of RiPPs are lantipeptides, which have already been chemically synthesized.⁸⁴ These molecules are short peptides (19-38 residues) with several thioether links, and most of them possess antibacterial activity—these are known as lantibiotics.²⁸ The first total synthesis of a lantibiotic, nisin, was achieved

by Shiba and coworkers in 1988.⁸⁵ After the successful production of lactocin S⁸⁶, the lactocin 314711⁸⁷ and epilancin 15X⁸⁸ by SPSS highlighted the growing applicability of chemical approaches to produce complex lantibiotics and analogs.

The first total chemical synthesis of a lasso peptide, namely lassomycin, was reported by Cobb and coworkers.⁸⁹ Afterwards, another chemical approach to lassomycin using SPPS was published by Harris and coworkers.⁹⁰ Some years before, lassomycin was isolated and characterized as a lasso peptide by NMR and mass spectrometry,⁶² but after the synthetic attempts, researchers realized that the NMR spectra of the two synthetic lassomycins were identical and that they differed from that of the natural product. In contrast, by HPLC and LC-MS, synthetic lassomycins were found to be identical to the natural product. These controversial results suggest, that either lassomycin is not a proper lasso peptide, or it is a proper lasso peptide and its characterization is incorrect.

In this thesis we focus on the synthesis of class II lasso peptides. The tails of the peptides belonging to this class are threaded through the macrocyclic ring and they have bulky side chains amino acids, acting as molecular barriers that trap the tail in the ring. Some chemical approaches using SPPS have been used to synthesize branched cyclic MccJ25 and cyclic and bicyclic analogs using amide or disulfide bonds.^{91,92} Other approaches are based on cyclization via copper-catalyzed azide-alkyne cycloaddition (CuAAC),⁹³ where the peptide loop and the specific topology were observed to be critical for biological activity. Similar approaches with interlocked and constrained peptides have been reported recently, such as rotaxanes and catenanes.⁹⁴ These mechanically linked molecular rings are important for the tertiary structures of proteins and their function. For this reason, some of them can be found in natural systems, such as in DNA⁹⁵ and proteins, while others can be synthesized using metal templates via ring-closing metathesis (RCM)^{96,97} or via the common CuAAC click reaction.⁹⁸ In addition, NCL is another useful technique for the synthesis of peptide catenanes⁹⁹ and cyclotides, such as Kalata B1.^{81,100}

Currently, all research into new peptide drugs pursues two main common objectives: development of new compounds resistant to enzymatic degradation and the modulation of peptide topology, since the properties are highly related to the shape.¹⁰¹ In this regard, most synthetic strategies are based on the imitation of the rotaxane and catenane architecture, which is a self-entanglement.¹⁰² The most

common molecular machineries for interlocked molecules are crown ether macrocycles and ammonium moieties.^{101,103} Saito and coworkers reported the chemical synthesis of peptide-based [1]rotaxanes, reminiscent of natural MccJ25. In this example, two [2]rotaxanes with thermal stability and high proteolytic resistance were obtained.¹⁰⁴ They started from a 21-crown-7 ether and secondary dialkylammoniums¹⁰⁵ to provide the [2]rotaxane, which was later derivatized to form a lasso peptide by two amide-forming ligations, NCL¹⁰⁶ and the α -ketoacid-hydroxylamine (KAHA) ligation.¹⁰⁷ Other examples of interlocked compounds are quasi[1]rotaxane and quasi[1]catenane, which are based on template-directed backfolding macrocyclization to obtain the spiro geometry, where two rings are connected to the same tetrahedral core carbon atom.¹⁰⁸ This geometry is found in many natural products with a constrained 3D structure and a broad range of biological activities.¹⁰⁹ Some of these molecules have been synthesized chemically using several strategies.¹¹⁰ Moreover, lasso peptides are excellent starting materials for the generation of catenanes with much smaller ring sizes than the previously described protein catenanes.¹¹¹ In this example, [3]catenane and [4]catenane structures, which self-assemble from a cleaved lasso peptide building block, were obtained. These structures were achieved by the trypsin cleavage of the peptide loop of MccJ25, showing that the C-terminal tail remained locked in place within the 8-aa ring. The same behavior for lasso peptides was described in previous studies involving thermolysin cleavage.¹¹²

Interlocked structure	Chemical approaches				
Lasso peptides (lassomycin and MccJ25)	SPPS ^{89,90,91,92}		CuAAC ⁹³		
Rotaxanes and catenanes	Metal templates via RCM ^{96,97}		CuAAC ⁹⁸	NCL ⁹⁹	
MccJ25-based [1]rotaxanes	Crown ether macrocycles and ammonium moieties ¹⁰⁴				
Quasi[1]rotaxane and quasi[1]catenane	Template-directed backfolding macrocyclization base on spiro geometry ¹⁰⁸				
Pseudo[1]rotaxane	Radically promoted reactions ¹¹³				
Bicyclic peptides	SPPS ⁹¹	NCL ¹¹⁴	RCM ^{115,116}	Pd-catalyzed C-C bond formation ¹¹⁷	Michael additions ¹¹⁸

Table 1: Summary of chemical approaches used for the synthesis of interlocked structures resembling lasso peptides.

In recent years there has been an increase in the number of chemical approaches for the synthesis of lasso peptides (Table 1). A representative example is the case of radically promoted reactions using BIPY²⁺ via CuAAC and a bulky group at the end as a stopper, creating a pseudo[1]rotaxane.¹¹³ Those authors achieved self-interlacing of the tail through the ring with the reduction of the molecule, reminiscent of lasso peptides.

Bicyclic peptides are suitable analogs for the chemical synthesis of lasso peptides. They can be produced biologically via ribosomal synthesis¹¹⁹ or via chemical synthesis using various techniques, like NCL and intramolecular disulfide formation,¹¹⁴ RCM,^{115,116} palladium-catalyzed C-C bond formation¹¹⁷ and successive Michael additions.¹¹⁸ It has been demonstrated that the second cycle yields peptides with higher rigidity and stability against thermal and proteolytic degradation, which translates into extended *in vivo* lifetime, therefore increasing the biological effect.¹²⁰ In some cases, the second cycle enhances the tight binding to target, due to the smaller loss of conformational entropy on binding.¹¹⁹ The recognition domain, responsible for the high affinity, is in the loop, and is not the macrocycle, in contrast to what was previously thought.¹²¹ For this reason, a suitable chemical approach would be the synthesis of lasso peptide-like bicyclic peptides in which the loop sequence is tied with a covalent bond. In this regard, Soudy and coworkers synthesized various MccJ25 analogs using SPPS. One of these analogs was synthesized as a bicyclic molecule with a fold in the C-terminal tail, providing good results as an antimicrobial peptide; however, it was not resistant to proteolytic degradation.⁹¹

In conclusion, the ability to generate lasso peptides synthetically remains a challenging endeavor and it would open the door to the production of lasso peptide analogs with unnatural amino acids or other nonproteinogenic building blocks. From a therapeutic point of view, these small and constrained structures would represent a new paradigm in drug discovery.

Chemoenzymatic approaches

Chemical methodologies for peptide macrocyclizations are difficult, due to the large entropic barriers and the presence of regio- or stereo-selectivity problems.¹²² An alternative approach for the synthesis of compounds with pharmacological properties is the use of chemoenzymatic strategies that combine the strength of SPPS

with the powerful stereo- and regio-selectivity achieved by enzymes.¹²³ For example, several chemoenzymatic reactions have been carried out with NRPs, such as peptide cyclization. These are carried out by the last domain of an assembly line, which in most cases is a thioesterase (TE) domain, responsible for the release of the mature product.^{124,125} Isolated and purified TE domains are widely used for the cyclization of linear peptidyl substrates, thus facilitating and accelerating the production of libraries with new pharmacologically active compounds.¹²⁵ In addition, these domains can also catalyze the hydrolytic cleavage of NRPs, leading to the formation of linear peptides, such as vancomycin-type antibiotics.¹²⁶ Regarding enzymatic reactions, most of them are reversible and their reaction rates depend on several factors, such as pH, temperature and concentration of enzyme/substrate.¹²⁷⁻¹²⁹

Regarding lasso peptides, gene cluster organizations can differ depending on the type of bacteria (Figure 6). For example, proteobacteria normally contain an ACB-type gene cluster, but sometimes a fourth gene, corresponding to a peptidase, is present around this cluster (Figure 9). This is the case of astexin-1,^{130,131} xanthomonins III¹³² and sphingopyxin I,⁷² in which a isopeptidase (IsoP) is the fourth gene. This is the first example of an enzyme that handles RiPP catabolism.¹³³ IsoP is an enzyme that specifically hydrolyzes the isopeptide bond of lasso peptides to yield linear peptides.⁷¹

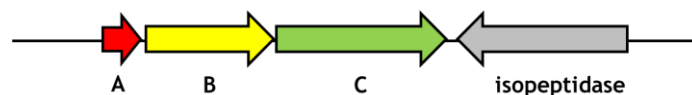


Figure 9: Schematic organization of an ABC-cluster with an adjacent isopeptidase

Maksimov and coworkers were the first to express a histidine-tagged isopeptidase from astexin-2 and -3 (AtxE2). AtxE2 was expressed in large amounts, and *in vitro* assays were performed with astexin-2/3. Mass spectrometry analysis revealed that the enzyme hydrolyzed the isopeptide bond. In addition, other lasso peptides, as well as unthreaded astexin-2 and -3 were tested, but in all cases the isopeptide bond remained intact.¹³³ This finding indicates that IsoP is highly selective; it recognizes lasso peptides only when these are produced by adjacent clusters.

To better understand IsoP specificity and to unveil its mechanism of action, other examples of this exciting family of enzymes, such as isopeptidase from sphingopyxin I (Spl-IsoP), were investigated (Figure 10). This 74.5 kDa protease consists of two

domains—an *N*-terminal β -propeller and a *C*-terminal α/β -hydrolase.¹³⁴ Despite its strict specificity, Spl-IsoP is highly tolerant to single amino acid substitutions in the peptide substrate, the loop being an exception as it plays an essential role in IsoP recognition.

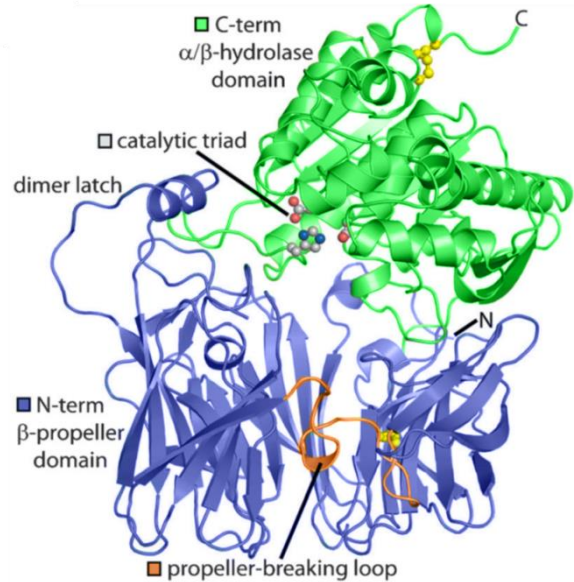


Figure 10: Front view of chain A of *SeMet*-derived *Spl-IsoP* crystal structure (PDB code 5JRK). The side chains of catalytic triad residues (S559, H671, D582) and disulfide bonds (yellow) are shown. (Adapted from Fage and coworkers)¹³⁴

Structural elucidation

The most common technique used for characterizing lasso peptides is NMR spectroscopy. To date, most three-dimensional structures of lasso peptides have been solved by NMR. It is a powerful tool for structural elucidation, but sometimes can be complex and can lead to confusion.¹³⁵ This may explain how the structures of MccJ25^{136,137} and BI-32169¹³⁸ were first incorrectly reported as head-to-tail cyclic and bicyclic peptides, respectively. For example, the correct structural characterization of MccJ25 required several publications.⁴⁵⁻⁴⁷

Furthermore, lassomycin,⁶² a highly selective and potent antibiotic against *Mycobacterium tuberculosis*, was characterized by NMR and was reported to be a lasso peptide but lacking the characteristic knot structure of the homologous lasso peptides. It shares only the isopeptide bond between the *N*-terminal Gly and the side-chain of Asp, forming an 8-residue macrolactam ring. This NMR characterization is ambiguous because a branched cyclic structure, as previously defined, is highly susceptible to proteolytic degradation and the structure obtained is unusual for

compounds with a mobile C-terminal tail. Another example in which NMR caused discrepancy was the first characterization of astexin-1.¹³⁹ In this case, the tail was threaded through the ring, but the amino acid plugs were incorrectly assigned. The amino acid plugs are normally key positions and they are essential for the stability of the lasso fold.^{68,140} Later, during a mutagenesis assay to study the thermal stability of astexin-1, where the supposed amino acids plugs were exchanged for Ala, the lasso production was observed not to be affected. However, when other positions were exchanged, such as Tyr14 and Phe15, no detectable lasso production was observed.¹³⁰ This observation was confirmed with a new NMR characterization, which indicated that these positions were the amino acid plugs, and not Glu17 and Arg19 as previously reported.

Regarding X-ray characterization, until now only two lasso peptide crystal structures, namely BI-32169¹⁴¹ and xanthomonin I,¹⁴² have been published. There is not a specific condition for lasso peptides crystallization. For example BI-32169 was crystallized under small-molecule conditions, while xanthomonin I was crystallized like a protein. Therefore, xanthomonin I was dissolved in water whereas the high hydrophobic BI-32169 was crystallized with 50% DMSO.

Other techniques, such as tandem mass spectrometry (MS/MS) in combination with thermal and enzymatic treatments, have been used to complement lasso peptide characterization. Several proteolytic assays with carboxypeptidase Y, pepsin, thermolysin, elastase, trypsin, chymotrypsin and proteinase K have shown that the vast majority of lasso peptides are highly stable against proteolysis by these enzymes.^{42,47,51,72,130,142,68,143} Interestingly, carboxypeptidase Y digestion can be used to differentiate a lasso and an unthreaded branched cyclic peptide. As the macrolactam ring shields the amino acids of the tail that are inside the ring against proteolytic digestion, a lasso peptide with a suitable amino acid plug at this position can be degraded only up to the amino acid plug. In contrast, the tail of a branched cyclic peptide is typically digested up to the macrolactam ring (Figure 11).¹⁴²

In tandem mass spectrometry, the molecular ions are dissociated, and their fragment ions are detected and can reveal the primary sequence of biomolecules. The fragmentation occurs at the same backbone linkage between each basic building block. From the mass differences for a series of cleavages at consecutive residues, the biomolecular sequence can be reconstructed.

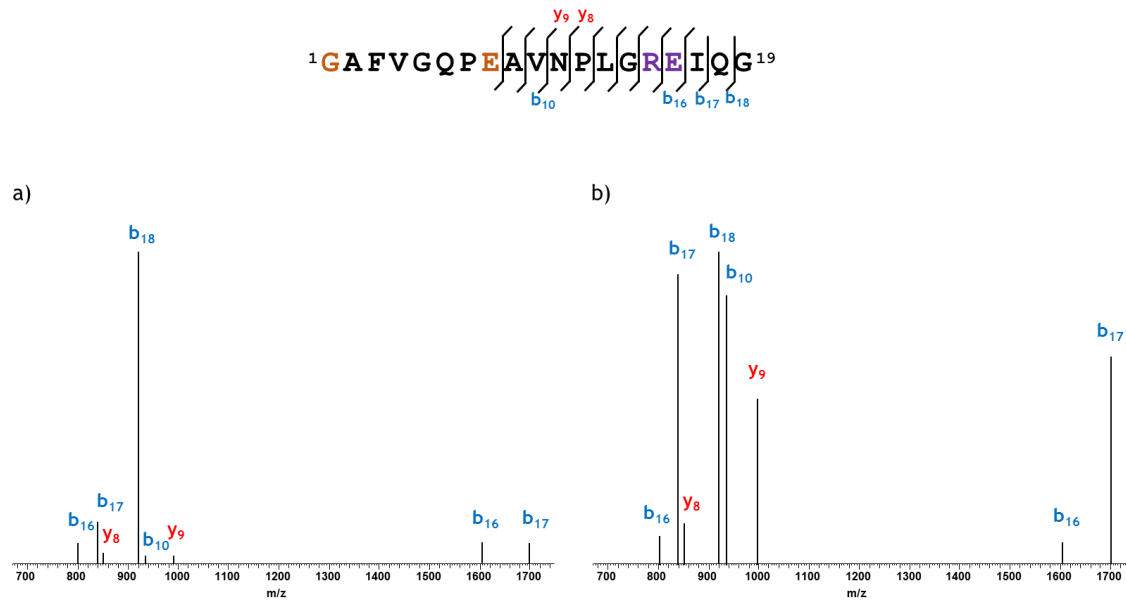


Figure 11: MS^2 spectra of the main fragmentation of caulosegnin I (a) and its branched cyclic topoisomer (b). The plug amino acids are presented in purple and the amino acids involved in the macrolactam ring formation are presented in brown. (Adapted from Hegemann and coworkers)⁶⁸

Regarding the MS^2 analysis, in lasso peptides the canonical fragmentation will occur only in the loop and tail regions of the peptide, providing preliminary information about the location of the ring. High intensities are usually associated with single bond breakages in the C-terminal tail region of the lasso peptide. In class II lasso peptides, the presence of two-product ions generated by double cleavage in the loop above the steric lock reveals the lassoed version of the peptide.¹⁴⁴ Moreover, comparison of the intensity of the fragmentations between lasso and branched cyclic peptides shows less intense fragmentation for the former as a result of the constrained structure (Figure 11). Concerning class I and III lasso peptides, their fragmentation pattern differs from that of peptides belonging to class II. Class I and III show weak fragmentation behavior and fragmentation around the ring, thereby hindering the distinction between a bicyclic and a lasso-structured peptide. This behavior could be a consequence of the fact that two or three bond breakages have to occur to observe fragmentation (Figure 12).^{52,67}

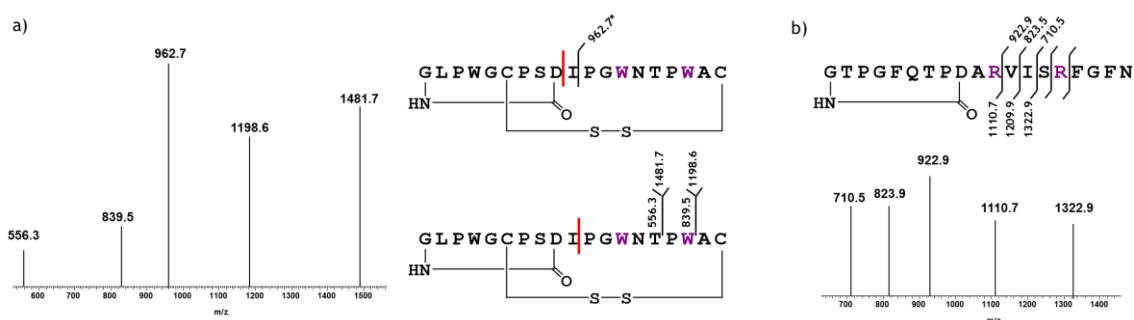


Figure 12: MS^2 spectra of doubly protonated precursor ion with the main fragmentations. a) BI-32169, class III lasso peptide. The fragments observed result from two peptide bond cleavages, indicated by the red and black bars, yielding a linear and a branched cyclic fragment. b) Capistruin, class II lasso peptide. The plug amino acids are presented in purple. (Adapted)^{52,145}

Other types of tandem mass spectrometry using distinct modes of activation are collision-induced dissociation (CID) and time-resolved electron capture dissociation (ECD).^{146,147} In ECD, precursor ions are irradiated with low-energy electrons and the fragmentation obtained is typically quite random, except for the fact that amine cleavage within Pro is rarely observed.¹⁴⁸ The major product normally is the charge-reduced species. On the other hand, CID shows a tendency for site-specific cleavage. A comparison of CID and ECD of proline-rich proteins (> 25% Pro) reported by Leymarie and coworkers demonstrated that both methodologies are useful for sequence analysis.¹⁴⁹

In CID and ECD experiments, the presence of disulfide bonds has been an impediment for MS/MS, limiting the available sequence information. Zubarev and coworkers reported that the most abundant fragment ion following ECD corresponds to S-S cleavage with retention of a hydrogen atom.¹⁵⁰ Mechanically interlocked molecules, such as catenanes, rotaxanes and lasso peptides constitute an attractive model to develop methods for characterizing gas-phase conformation.¹⁵¹ In this regard, different fragmentation patterns have been observed between the threaded and the unthreaded conformation and the ability to distinguish these two conformations depends on the amino acid sequence. Only in the case of MccJ25 were two-peptide product ions—a specific feature of lasso peptides—were detected under CID. However, these ions cannot always be detected, as already mentioned for the MS/MS analysis.⁵⁰ On the other hand, capistruin characterization by CID and ECD was more difficult and only the extent of H[•] transfer in the ECD spectra permitted differentiation between the lasso-folded peptide and its branched cyclic topoisomer.¹⁴⁴

Ion-mobility mass spectrometry (IM-MS) is a recent technique that allows rapid structural detection and dynamical features. Over the past 15 years, there has been a considerable increase in the use of this technique, as demonstrated by the growing number of scientific publications (Figure 13).

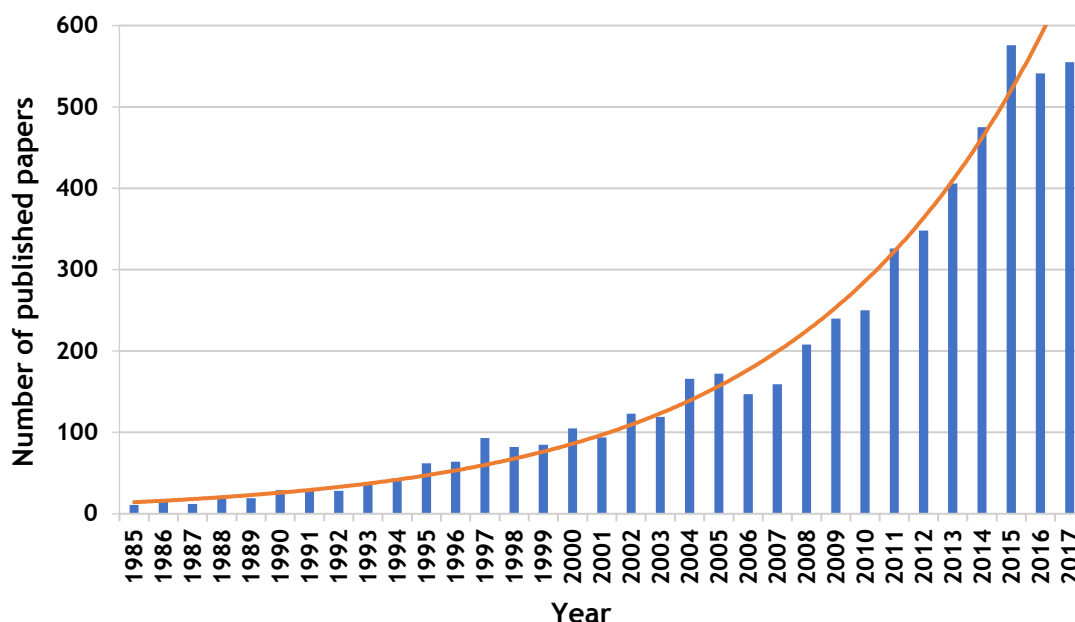


Figure 13: Number of published papers combining ion mobility and mass spectrometry. Data generated using Web of Science between 1985 and 2017, search terms “ion mobility” and “mass spectrometry”. The orange line indicates the increasing trend over the years.

IM-MS is a complementary approach to MS/MS experiments that provides information on the global shape of molecules,¹⁵² and has proven useful for the structural characterization of many molecules.¹⁵³⁻¹⁵⁶ There are three main types of IM-MS; drift-time ion mobility spectrometry (DTIMS), travelling-wave ion mobility spectrometry (TWIMS) and field-asymmetric ion mobility spectrometry (FAIMS).¹⁵⁷ DTIMS and TWIMS are based on the time mobility and can be used to determine information about cross-sectional area, while FAIMS operates as a mobility filter and typically finds application as an orthogonal separation technique to liquid chromatography, due to its high resolution power. DTIMS is the oldest and conceptually simplest form of ion mobility. Sometimes, the combined use of DTIMS/TWIMS enhance the structural analysis.

DTIMS is based on the measurement of the drift time of ion buffer gas, typically helium, under the influence of a weak electric field (Figure 14). The drift velocity (v_d) is proportional to the electric field (E) and the ion mobility (K), as shown in Equation 1.

$$v_d = KE \quad (1)$$

As collisions increase, there is a loss of energy, and consequently ions take longer time to cross the ion-mobility cell (the drift time). After that, ions are injected into the detector, accomplishing a separation on the basis of their mass to charge ratio and their collision cross section (CCS), Ω . The CCS is the time taken for an ion to drift through the tube, related to its rotationally averaged cross-sectional area. Different isomers of species can then be separated, the most compact travelling faster.¹⁵⁸ The interpretation of IM-MS spectra in terms of conformation then relies on comparison with the calculated CCS for candidate structures from simulations. This approach allows examination of peptide and protein conformations in the gas phase, thus facilitating differentiation between lasso and branched cyclic peptides. Several IM-MS studies on peptide confirm that charge location is the main driving force in the stabilization of gas-phase structures.¹⁵⁹

In a recent study, five distinct lasso peptides belonging to class II, together with their synthetic branched cyclic counterparts, were analyzed by IM-MS.¹⁶⁰ Differences between the two conformations were observed mainly at higher charge states. This high charge state is achieved through the addition of sulfolane as a supercharging reagent.¹⁶¹ Due to the lack of a mechanically interlocked structure in the branched cyclic peptide, higher CCS values and more conformations were found on these molecules than for lasso peptides.

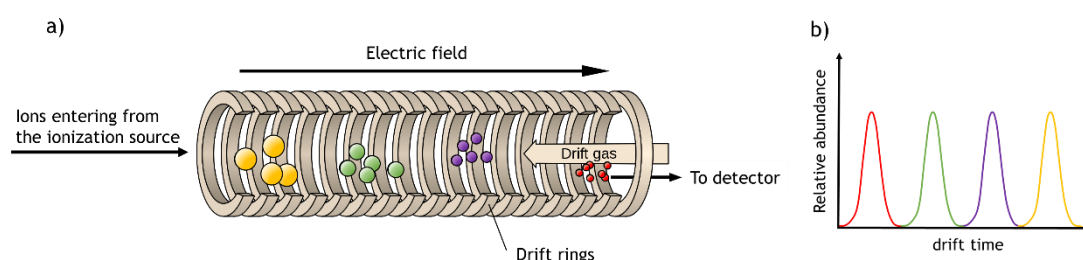


Figure 14: Separation of gaseous ions on the basis of ion mobility. a) a drift tube ion mobility spectrometer and b) schematic ion-mobility spectrum. (Adapted)¹⁶²

Finally, an alternative and complementary approach to IM-MS is infrared multiple photon dissociation (IRMPD), which provides structural information on mass-selected ions and has been used to resolve hydrogen bonding association to peptide structuration.¹⁶³ In IRMPD, precursor ions are subjected to infrared irradiation and slowly heated until they begin to dissociate. Amide bond cleavage is very extensive in

this technique, making it particularly useful for peptide sequence verification. For example, state charge $[M + 3H]^{3+}$ and $[M + 4H]^{4+}$ for MccJ25 show similar infrared spectra, which suggest that the gas phase conformation of the hydrogen bond interactions is not strongly affected by the charge state. However, for the branched cyclic MccJ25, different infrared spectra are observed. These evidences reveal that in the charge state $[M + 3H]^{3+}$, the conformation is stabilized by hydrogen bonds comparing to the $[M + 4H]^{4+}$, which displays more unfolding.¹⁶⁴ These results are in accordance with the IM-MS studies. Comparing both techniques, IM-MS provides information on the global shape of the molecule, whereas IRMPD informs on the presence of hydrogen bonding.^{165,166}

Objectives

The main focus of this PhD thesis is on the chemical synthesis of lasso peptides, specifically sungsanpin, a recently isolated lasso peptide with anticancer activity.

To date, no synthetic access to lasso peptides is available due to the difficulty in building and maintaining the threaded lasso structure. In order to accomplish this ambitious goal, the following specific objectives were established:

1. Design and synthesis of sungsanpin and bicyclic analogs with linkages able to maintain the threaded lasso structure.
2. Characterization and structural elucidation of the synthetic peptides to prove the presence of the lasso structure. For this objective, several structural techniques will be employed, such as mass spectrometry, chromatography, crystallization and NMR spectroscopy. The final 3D structure will be validated by computational tools.
3. *In vitro* evaluation of their anticancer activity, in order to establish the structure-activity relationship.

During the second year of my thesis I had the opportunity to join the group of Prof. Mohamed Marahiel in the Philipps-Universität Marburg, in Germany, to do a research stay. The main goal of this stay is the biosynthesis of sungsanpin and other lasso peptides. To achieve this goal, several specific objectives are established:

4. Perform genome mining to identify promising strains.
5. Production, isolation and purification of different lasso peptides homologously to compare their biological activity with the synthetic analogs.
6. Mutational analysis and heterologous expression of the identified gene clusters in *E.coli*.

References

- (1) Cheuka, P. M.; Mayoka, G.; Mutai, P.; Chibale, K. *Molecules* **2017**, *22* (1).
- (2) Clark, A. M. *Pharm. Res.* **1996**, *13* (8), 1133.
- (3) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016**, *79* (3), 629.
- (4) Ganesan, A. *Curr. Opin. Chem. Biol.* **2008**, *12* (3), 306.
- (5) Harvey, A. L. *Drug Discov. Today* **2008**, *13* (19-20), 894.
- (6) Fox, S.; Farr-Jones, S.; Yund, M. A. *J. Biomol. Screen.* **1999**, *4* (4), 183.
- (7) Kelecom, A. *An. Acad. Bras. Cienc.* **2002**, *74* (1), 151.
- (8) Salomon, C. E.; Magarvey, N. A.; Sherman, D. H. *Nat. Prod. Rep.* **2004**, *21* (1), 105.
- (9) Medema, M. H.; Breitling, R.; Takano, E. *Methods Enzymol.* **2011**, *497*, 485.
- (10) Bérdy, J. *J. Antibiot. (Tokyo)*. **2005**, *58* (1), 1.
- (11) Goodfellow, M.; Fiedler, H. P. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **2010**, *98* (2), 119.
- (12) Challis, G. L. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 219.
- (13) Bergmann, W.; Feeney, R. J. *J. Org. Chem.* **1951**, *16* (6), 981.
- (14) Bergmann, W.; Burke, D. C. *J. Org. Chem.* **1956**, *21* (2), 226.
- (15) Stonik, V. A. *Acta Naturae* **2009**, *1* (2), 15.
- (16) Pathom-aree, W.; Stach, J. E. M.; Ward, A. C.; Horikoshi, K.; Bull, A. T.; Goodfellow, M. *Extremophiles* **2006**, *10* (3), 181.
- (17) Lam, K. S. *Curr. Opin. Microbiol.* **2006**, *9* (3), 245.
- (18) Helmke, E.; Weyland, H. *Int. J. Syst. Bacteriol.* **1984**, *34* (69), 127.
- (19) Mincer, T. J.; Jensen, P. R.; Kauffman, C. A.; Fenical, W. **2002**, *68* (10), 5005.
- (20) Bentley, S.; Chater, K.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. a; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabinowitsch, E.; Rajandream, M.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. a. *Nature* **2002**, *417* (6885), 141.
- (21) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S. *Nat. Biotechnol.* **2003**, *21* (5), 526.
- (22) Rishton, G. M. *Am. J. Cardiol.* **2008**, *101* (10), S43.
- (23) Wilkinson, B.; Micklefield, J. *Nat. Chem. Biol.* **2007**, *3* (7), 379.

- (24) Gomez-Escribano, J. P.; Castro, J. F.; Razmilic, V.; Chandra, G.; Andrews, B.; Asenjo, J. A.; Bibb, M. J. *BMC Genomics* **2015**, *16* (1), 485.
- (25) Schuster, S. C. *Nat. Methods* **2007**, *5* (1), 16.
- (26) Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20* (3), 275.
- (27) McIntosh, J. a.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26* (4), 537.
- (28) Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61* (1), 477.
- (29) Daly, N. L.; Rosengren, K. J.; Craik, D. J. *Adv. Drug Deliv. Rev.* **2009**, *61* (11), 918.
- (30) Severinov, K.; Semenova, E.; Kazakov, A.; Kazakov, T.; Gelfand, M. S. *Mol. Microbiol.* **2007**, *65* (6), 1380.
- (31) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. *Acc. Chem. Res.* **2015**, *48* (7), 1909.
- (32) Widdick, D. A.; Dodd, H. M.; Barraille, P.; White, J.; Stein, T. H.; Chater, K. F.; Gasson, M. J.; Bibb, M. J. *Proc. Natl. Acad. Sci.* **2003**, *100* (7), 4316.
- (33) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. a; Bugni, T. S.; Bulaj, G.; Camarero, J. a; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. a; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, a J.; Liu, W.; Marahiel, M. a; Mitchell, D. a; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süßmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. a. *Nat. Prod. Rep.* **2013**, *30* (1), 108.
- (34) Yang, X.; Van Der Donk, W. A. *Chem. - A Eur. J.* **2013**, *19* (24), 7662.
- (35) Oman, T. J.; van der Donk, W. A. *Nat. Chem. Biol.* **2010**, *6* (1), 9.
- (36) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. *Angew. Chemie - Int. Ed.* **2011**, *50* (37), 8714.
- (37) Hegemann, J. D.; De Simone, M.; Zimmermann, M.; Knappe, T. A.; Xie, X.; Di Leva, F. S.; Marinelli, L.; Novellino, E.; Zahler, S.; Kessler, H.; Marahiel, M. A. *J. Med. Chem.* **2014**, *57* (13), 5829.
- (38) Maksimov, M. O.; Pan, S. J.; James Link, a. *Nat. Prod. Rep.* **2012**, *29* (9), 996.

-
- (39) Zhao, N.; Pan, Y.; Cheng, Z.; Liu, H. *Amino Acids* **2016**, *48* (6), 1347.
- (40) Weber, W.; Fischli, W.; Hochuli, E.; Kupfer, E.; Weibel, E. K. *J. Antibiot. (Tokyo)*. **1991**, *44*, 164.
- (41) Helynck, G.; Dubertret, C.; Mayaux, J. F.; Leboul, J. *J. Antibiot. (Tokyo)*. **1993**, *46* (11), 1756.
- (42) Fréchet, D.; Guitton, J. D.; Herman, F.; Faucher, D.; Helynck, G.; Monegier du Sorbier, B.; Ridoux, J. P.; James-Surcouf, E.; Vuilhorgne, M. *Biochemistry* **1994**, *33* (1), 42.
- (43) Tsunakawa, M.; Hu, S. L.; Hoshino, Y.; Detlefsen, D. J.; Hill, S. E.; Furumai, T.; White, R. J.; Nishio, M.; Kawano, K.; Yamamoto, S. *J. Antibiot. (Tokyo)*. **1995**, *48* (5), 433.
- (44) Salomón, R. a; Farías, R. N. *J. Bacteriol.* **1992**, *174* (22), 7428.
- (45) Wilson, K.-A. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A. *J. Am. Chem. Soc.* **2003**, *125* (41), 12475.
- (46) Bayro, M. J.; Mukhopadhyay, J.; Swapna, G. V. T.; Huang, J. Y.; Ma, L. C.; Sineva, E.; Dawson, P. E.; Montelione, G. T.; Ebright, R. H. *J. Am. Chem. Soc.* **2003**, *125* (41), 12382.
- (47) Rosengren, K. J.; Clark, R. J.; Daly, N. L.; G?ransson, U.; Jones, A.; Craik, D. J. *J. Am. Chem. Soc.* **2003**, *125* (41), 12464.
- (48) Shah, P.; Vella, A.; Basu, R.; Schwenk, W. F.; Rizza, R. A. *J Clin Endocrinol Metab.* **2000**, *85* (11), 4053.
- (49) Morishita, Y.; Chiba, S.; Tsukuda, E.; Tanaka, T.; Ogawa, T.; Yamasaki, M.; Yoshida, M.; Kawamoto, I.; Matsuda, Y. *J. Antibiot. (Tokyo)*. **1994**, *47* (3), 269.
- (50) Knappe, T. a.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. a. *J. Am. Chem. Soc.* **2008**, *130* (17), 11446.
- (51) Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Omura, S. *J. Am. Chem. Soc.* **2006**, *128* (23), 7486.
- (52) Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A. *FEBS Lett.* **2010**, *584* (4), 785.
- (53) Katahira, R.; Shibata, K.; Yamasaki, M.; Matsuda, Y.; Yoshida, M. *Bioorg. Med. Chem.* **1995**, *3* (9), 1273.
- (54) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. *Chem. Biol.* **2009**, *16* (12), 1290.
- (55) Clarke, D. J.; Campopiano, D. J. *Org. Biomol. Chem.* **2007**, *5* (16), 2564.

- (56) Duquesne, S.; Destoumieux-Garzón, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. *Chem. Biol.* **2007**, *14* (7), 793.
- (57) Um, S.; Kim, Y.-J. J.; Kwon, H. H. C.; Wen, H.; Kim, S.-H. H.; Kwon, H. H. C.; Park, S.; Shin, J.; Oh, D.-C. C. *J. Nat. Prod.* **2013**, *76* (5), 873.
- (58) Duncan, K. R.; Haltli, B.; Gill, K. A.; Correa, H.; Berru e, F.; Kerr, R. G. *J. Ind. Microbiol. Biotechnol.* **2015**, *42* (1), 57.
- (59) Wakefield, J.; Hassan, H. M.; Jaspars, M.; Ebel, R.; Rateb, M. E. *Front. Microbiol.* **2017**, *8* (JUL), 1.
- (60) Elsayed, S. S.; Trusch, F.; Deng, H.; Raab, A.; Prokes, I.; Busarakam, K.; Asenjo, J. A.; Andrews, B. A.; Van West, P.; Bull, A. T.; Goodfellow, M.; Yi, Y.; Ebel, R.; Jaspars, M.; Rateb, M. E. *J. Org. Chem.* **2015**, *80* (20), 10252.
- (61) Solbiati, J. O.; Ciaccio, M.; Far as, R. N.; Jos e, E.; Moreno, F.; Salom n, R. a; Solbiati, O.; Fari, R. N. *J. Bacteriol.* **1999**, *181* (8), 2659.
- (62) Gavrish, E.; Sit, C. S.; Cao, S.; Kandr or, O.; Spoering, A.; Peoples, A.; Ling, L.; Fetterman, A.; Hughes, D.; Bissell, A.; Torrey, H.; Akopian, T.; Mueller, A.; Epstein, S.; Goldberg, A.; Clardy, J.; Lewis, K. *Chem. Biol.* **2014**, *21* (4), 509.
- (63) Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. *Appl. Microbiol. Biotechnol.* **2012**, *95* (2), 451.
- (64) Cheung, W. L.; Chen, M. Y.; Maksimov, M. O.; Link, A. J. *ACS Cent. Sci.* **2016**, *2* (10), 702.
- (65) Metelev, M.; Tietz, J. I.; Melby, J. O.; Blair, P. M.; Zhu, L.; Livnat, I.; Severinov, K.; Mitchell, D. A. *Chem. Biol.* **2015**, *22* (2), 241.
- (66) Sugawara, E.; Nikaido, H. *Antimicrob. Agents Chemother.* **2014**, *58* (12), 7250.
- (67) Li, Y.; Ducasse, R.; Zirah, S.; Blond, A.; Goulard, C.; Lescop, E.; Giraud, C.; Hartke, A.; Guittet, E.; Pernodet, J. L.; Rebuffat, S. *ACS Chem. Biol.* **2015**, *10* (11), 2641.
- (68) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2013**, *135* (1), 210.
- (69) Pan, S. J.; Cheung, W. L.; Link, A. J. *Protein Expr. Purif.* **2010**, *71* (2), 200.
- (70) Zhu, S.; Fage, C. D.; Hegemann, J. D.; Mielcarek, A.; Yan, D.; Linne, U.; Marahiel, M. A. *Sci. Rep.* **2016**, *6* (1), 35604.
- (71) Maksimov, M. O.; Link, A. J. *J. Am. Chem. Soc.* **2013**, *135* (32), 12038.
- (72) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. *Biopolymers* **2013**, *100* (5), 527.
- (73) Zasloff, M. *Nature* **2002**, *415* (6870), 389.
- (74) Yan, K. P.; Li, Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.;

- Rebuffat, S. *ChemBioChem* **2012**, *13* (7), 1046.
- (75) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. *Biopolymers* **2013**, *100* (5), 527.
- (76) Pan, S. J.; Rajniak, J.; Cheung, W. L.; Link, A. J. *ChemBioChem* **2012**, *13* (3), 367.
- (77) Maksimov, M. O.; Link, A. J. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 333.
- (78) Craik, D. J.; Clark, R. J.; Daly, N. L. *Expert Opin. Investig. Drugs* **2007**, *16* (5), 595.
- (79) Henriques, S. T.; Craik, D. J. *Drug Discov. Today* **2010**, *15* (1-2), 57.
- (80) Eliassen, R.; Daly, N. L.; Wulff, B. S.; Andresen, T. L.; Conde-Frieboes, K. W.; Craik, D. J. *J. Biol. Chem.* **2012**, *287* (48), 40493.
- (81) Daly, N. L.; Love, S.; Alewood, P. F.; Craik, D. J. *Biochemistry* **1999**, *38* (32), 10606.
- (82) Gunasekera, S.; Daly, N. L.; Anderson, M. A.; Craik, D. J. *IUBMB Life* **2006**, *58* (9), 515.
- (83) Craik, D. J.; Clark, R. J.; Daly, N. L. *Expert Opin. Investig. Drugs* **2007**, *16* (5), 595.
- (84) Tabor, A. B. *Org. Biomol. Chem.* **2011**, *9* (22), 7606.
- (85) Fukase, K.; Kitazawa, M.; Sanu, A.; Fujita, H.; Horimoto, S.; Wakamita, T.; Shiba, T. *Tetrahedron Lett.* **1988**, *29* (7), 795.
- (86) Ross, A. C.; Liu, H.; Pattabiraman, V. R.; Vederas, J. C. *J. Am. Chem. Soc.* **2010**, *132* (2), 462.
- (87) Liu, W.; Chan, A. S. H.; Liu, H.; Cochrane, S. A.; Vederas, J. C. *J. Am. Chem. Soc.* **2011**, *133* (36), 14216.
- (88) Knerr, P. J.; Van Der Donk, W. A. *J. Am. Chem. Soc.* **2012**, *134* (18), 7648.
- (89) Lear, S.; Munshi, T.; Hudson, A. S.; Hatton, C.; Clardy, J.; Mosely, J. A.; Bull, T. J.; Sit, C. S.; Cobb, S. L. *Org. Biomol. Chem.* **2016**, 4534.
- (90) Harris, P. W. R.; Cook, G. M.; Leung, I. K. H.; Brimble, M. A. *Aust. J. Chem.* **2017**, *70* (2), 172.
- (91) Soudy, R.; Wang, L.; Kaur, K. *Bioorganic Med. Chem.* **2012**, *20* (5), 1794.
- (92) Hammami, R.; Bédard, F.; Gomaa, A.; Subirade, M.; Biron, E.; Fliss, I. *Amino Acids* **2015**, *47* (2), 417.
- (93) Ferguson, A. L.; Zhang, S.; Dikiy, I.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; James Link, A.; Link, A. J. *Biophys. J.* **2010**, *99* (9), 3056.
- (94) Gil-Ramírez, G.; Leigh, D. A.; Stephens, A. J. *Angew. Chemie - Int. Ed.* **2015**, *54* (21), 6110.

- (95) Hudson, B.; Vinograd, J. *Nature* **1967**, *216* (5116), 647.
- (96) Mohr, B.; Weck, M.; Sauvage, J.-P.; Grubbs, R. H. *Angew. Chem. Int. Ed. Engl.* **1997**, *36* (12), 1308.
- (97) Hogg, L.; Leigh, D. A.; Lusby, P. J.; Morelli, A.; Parsons, S.; Wong, J. K. Y. *Angew. Chemie - Int. Ed.* **2004**, *43* (10), 1218.
- (98) Hänni, K. D.; Leigh, D. A. *Chem. Soc. Rev.* **2010**, *39* (4), 1240.
- (99) Yan, L. Z.; Dawson, P. E. *Angew. Chemie Int. Ed.* **2001**, *40* (19), 3625.
- (100) Tam, J. P.; Lu, Y.-A.; Yang, J.-L.; Chiu, K.-W. *Proc. Natl. Acad. Sci.* **1999**, *96* (16), 8913.
- (101) Clavel, C.; Fournel-Marotte, K.; Coutrot, F. *Molecules* **2013**, *18* (9), 11553.
- (102) Xue, Z.; Mayer, M. F. *J. Am. Chem. Soc.* **2010**, *132* (10), 3274.
- (103) Coutrot, F.; Busseron, E. *Chem. - A Eur. J.* **2008**, *14* (16), 4784.
- (104) Saito, F.; Bode, J. W. *Chem. Sci.* **2017**, *8* (4), 2878.
- (105) Zhang, C.; Li, S.; Zhang, J.; Zhu, K.; Li, N.; Huang, F. *Org. Lett.* **2007**, *9* (26), 5553.
- (106) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science.* **1994**, *266* (5186), 776.
- (107) Bode, J. W.; Fox, R. M.; Baucom, K. D. *Angew. Chemie - Int. Ed.* **2006**, *45* (8), 1248.
- (108) S Steemers, L.; Wanner, M. J.; Lutz, M.; Hiemstra, H.; Van Maarseveen, J. H. *Nat. Commun.* **2017**, *8* (25), 15392.
- (109) Zheng, Y.; Tice, C. M.; Singh, S. B. *Bioorganic Med. Chem. Lett.* **2014**, *24* (16), 3673.
- (110) Smith, L. K.; Baxendale, I. R. *Org. Biomol. Chem.* **2015**, *13* (39), 9907.
- (111) Allen, C. D.; Link, A. J. *J. Am. Chem. Soc.* **2016**, *138* (43), 14214.
- (112) Rosengren, K. J.; Blond, A.; Afonso, C.; Tabet, J. C.; Rebuffat, S.; Craik, D. J. *Biochemistry* **2004**, *43* (16), 4696.
- (113) Wang, Y.; Sun, J.; Liu, Z.; Nassar, M. S.; Botros, Y. Y.; Stoddart, J. F. *Chem. Sci.* **2017**, *8* (4), 2562.
- (114) Sun, Y.; Lu, G.; Tam, J. P. *Org. Lett.* **2001**, *3* (11), 1681.
- (115) Ghalit, N.; Reichwein, J. F.; Hilbers, H. W.; Breukink, E.; Rijkers, D. T. S.; Liskamp, R. M. J. *ChemBioChem* **2007**, *8* (13), 1540.
- (116) Cromm, P. M.; Schaubach, S.; Spiegel, J.; Fürstner, A.; Grossmann, T. N.; Waldmann, H. *Nat. Commun.* **2016**, *7* (14), 11300.
- (117) Mendive-Tapia, L.; Preciado, S.; García, J.; Ramón, R.; Kielland, N.; Albericio, F.; Lavilla, R. *Nat. Commun.* **2015**, *6* (1), 7160.
- (118) Elduque, X.; Pedroso, E.; Grandas, A. *Org. Lett.* **2013**, *15* (8), 2038.

- (119) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat. Chem. Biol.* **2009**, *5* (7), 502.
- (120) Hacker, D. E.; Hoinka, J.; Iqbal, E. S.; Przytycka, T. M.; Hartman, M. C. T. *ACS Chem. Biol.* **2017**, *12* (3), 795.
- (121) Mathavan, I.; Zirah, S.; Mehmood, S.; Choudhury, H. G.; Goulard, C.; Li, Y.; Robinson, C. V.; Rebuffat, S.; Beis, K. *Nat. Chem. Biol.* **2014**, *10* (5), 340.
- (122) Cavelier-Frontin, F.; Achmad, S.; Verducci, J.; Jacquier, R.; Pèpe, G. *J. Mol. Struct. THEOCHEM* **1993**, *286* (1), 125.
- (123) Wang, L. X.; Amin, M. N. *Chem. Biol.* **2014**, *21* (1), 51.
- (124) Grünewald, J.; Kopp, F.; Mahlert, C.; Linne, U.; Sieber, S. A.; Marahiel, M. A. *Chem. Biol.* **2005**, *12* (8), 873.
- (125) Kopp, F.; Marahiel, M. A. *Nat. Prod. Rep.* **2007**, *24* (4), 735.
- (126) Recktenwald, J.; Shawky, R.; Puk, O.; Pfenning, F.; Keller, U.; Wohlleben, W.; Pelzer, S. *Microbiology* **2002**, *148* (4), 1105.
- (127) Monod, J. *Annu. Rev. Microbiol.* **1949**, *3* (1), 371.
- (128) Friedman, S.; Fraenkel, G. *Arch. Biochem. Biophys.* **1955**, *59* (2), 491.
- (129) Alton, L.; Institutes, M.; Bethesda, N. *J. Biol. Chem.* **1954**, *209* (9), 265.
- (130) Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A. *Chem. Biol.* **2013**, *20* (4), 558.
- (131) Allen, C. D.; Chen, M. Y.; Trick, A. Y.; Le, D. T.; Ferguson, A. L.; Link, A. J. *ACS Chem. Biol.* **2016**, *11* (11), 3043.
- (132) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Steuber, H.; Harms, K.; Xie, X.; Marahiel, M. A. *Angew. Chemie Int. Ed.* **2014**, *53* (8), 2230.
- (133) Maksimov, M. O.; Koos, J. D.; Zong, C.; Lisko, B.; James Link, A. *J. Biol. Chem.* **2015**, *290* (52), 30806.
- (134) Fage, C. D.; Hegemann, J. D.; Nebel, A. J.; Steinbach, R. M.; Zhu, S.; Linne, U.; Harms, K.; Bange, G.; Marahiel, M. A. *Angew. Chemie Int. Ed.* **2016**, *55* (41), 12717.
- (135) Xie, X.; Marahiel, M. A. *ChemBioChem* **2012**, *13* (5), 621.
- (136) Blond, A.; Péduzzi, J.; Goulard, C.; Chiuchiolo, M. J.; Barthélémy, M.; Prigent, Y.; Salomón, R. A.; Farías, R. N.; Moreno, F.; Rebuffat, S. *Eur. J. Biochem.* **1999**, *259* (3), 747.
- (137) Blond, A.; Cheminant, M.; Ségalas-Milazzo, I.; Péduzzi, J.; Barthélémy, M.; Goulard, C.; Salomón, R.; Moreno, F.; Farías, R.; Rebuffat, S. *Eur. J. Biochem.* **2001**, *268* (7), 2124.
- (138) Potterat, O.; Wagner, K.; Gemmecker, G.; Mack, J.; Puder, C.; Vettermann,

- R.; Streicher, R. *J. Nat. Prod.* **2004**, *67* (9), 1528.
- (139) Maksimov, M. O.; Pelczer, I.; Link, A. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (38), 15223.
- (140) Hegemann, J. D.; Fage, C. D.; Zhu, S.; Harms, K.; Di Leva, F. S.; Novellino, E.; Marinelli, L.; Marahiel, M. A.; Zahler, S.; Kessler, H.; Marahiel, M. A.; Akopian, T.; Mueller, A.; Epstein, S.; Goldberg, A.; Clardy, J.; Lewis, K. *Mol. BioSyst.* **2016**, *12* (4), 1106.
- (141) Nar, H.; Schmid, A.; Puder, C.; Potterat, O. *ChemMedChem* **2010**, *5* (10), 1689.
- (142) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Steuber, H.; Harms, K.; Xie, X.; Marahiel, M. A. *Angew. Chemie - Int. Ed.* **2014**, *53* (8), 2230.
- (143) Potterat, O.; Stephan, H.; Metzger, J. W.; Gnau, V.; Zähler, H.; Jung, G. *Liebigs Ann. der Chemie* **1994**, *1994* (7), 741.
- (144) Zirah, S.; Afonso, C.; Linne, U.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S.; Tabet, J. C. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (3), 467.
- (145) Knappe, T. A.; Linne, U.; Zirah, S. S.; Rebuffat, S.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2008**, *130* (17), 11446.
- (146) Cooper, H. J.; Håkansson, K.; Marshall, A. G. *Mass Spectrom. Rev.* **2005**, *24* (2), 201.
- (147) Pérot-Taillandier, M.; Zirah, S.; Rebuffat, S.; Linne, U.; Marahiel, M. A.; Cole, R. B.; Tabet, J. C.; Afonso, C. *Anal. Chem.* **2012**, *84* (11), 4957.
- (148) Cooper, H. J.; Hudgins, R. R.; Håkansson, K.; Marshall, A. G. *Int. J. Mass Spectrom.* **2003**, *228* (2-3), 723.
- (149) Leymarie, N.; Berg, E. A.; McComb, M. E.; O'Connor, P. B.; Grogan, J.; Oppenheim, F. G.; Costello, C. E. *Anal. Chem.* **2002**, *74* (16), 4124.
- (150) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121* (12), 2857.
- (151) Dietrich-Buchecker C., Colasson B.X., S. J. In *Topics in Current Chemistry. Templettes in Chemistry II*; Springer, Berlin, Heidelberg, 2005; pp 261-283.
- (152) Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32* (6), 577.
- (153) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995**, *117* (40), 10141.
- (154) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. *Nat. Protoc.* **2008**, *3* (7), 1139.
- (155) De Cecco, M.; Seo, E. S.; Clarke, D. J.; McCullough, B. J.; Taylor, K.; Macmillan, D.; Dorin, J. R.; Campopiano, D. J.; Barran, P. E. *J. Phys. Chem. B*

- 2010, 114 (6), 2312.
- (156) Canon, F.; Ballivian, R.; Chirot, F.; Antoine, R.; Sarni-Manchado, P.; Lemoine, J.; Dugourd, P. *J. Am. Chem. Soc.* **2011**, 133 (20), 7847.
- (157) Lanucara, F.; Holman, S. W.; Gray, C. J.; Evers, C. E. *Nat. Chem.* **2014**, 6 (4), 281.
- (158) Smith, D.; Knapman, T.; Campuzano, I.; Malham, R.; Berryman, J.; Radford, S.; Ashcroft, A. *Eur. J. Mass Spectrom.* **2009**, 15 (5), 113.
- (159) Jeanne, K.; Fouque, D.; Lavanant, H.; Zirah, S.; Lemoine, J.; Rebuffat, S.; Tabet, J.; Kulesza, A.; Afonso, C.; Dugourd, P.; Chirot, F. *Rapid Commun. Mass Spectrom.* **2015**, No. May, 1411.
- (160) Jeanne Dit Fouque, K.; Afonso, C.; Zirah, S.; Hegemann, J. D.; Zimmermann, M.; Marahiel, M. A.; Rebuffat, S.; Lavanant, H. *Anal. Chem.* **2015**, 87 (2), 1166.
- (161) Sterling, H. J.; Daly, M. P.; Feld, G. K.; Thoren, K. L.; Kintzer, A. F.; Krantz, B. A.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **2010**, 21 (10), 1762.
- (162) Dahl, J. Diagram of an ion mobility spectrometer
https://en.wikipedia.org/wiki/File:Ion_mobility_spectrometry_diagram.svg#filelinks (accessed Jan 17, 2018).
- (163) Hernandez, O.; Isenberg, S.; Steinmetz, V.; Glish, G. L.; Maitre, P. *J. Phys. Chem. A* **2015**, 119 (23), 6057.
- (164) Jeanne Dit Fouque, K.; Lavanant, H.; Zirah, S.; Steinmetz, V.; Rebuffat, S.; Maître, P.; Afonso, C. *J. Phys. Chem. A* **2016**, 120 (21), 3810.
- (165) Kupser, P.; Pagel, K.; Oomens, J.; Polfer, N.; Koks, B.; Meijer, G.; Von Helden, G. *J. Am. Chem. Soc.* **2010**, 132 (6), 2085.
- (166) Plowright, R. J.; Gloaguen, E.; Mons, M. *ChemPhysChem* **2011**, 12 (10), 1889.

**CHAPTER 1: DESIGN AND
SYNTHESIS OF LASSO
PEPTIDES**

Structure elucidation of sungsanpin

Sungsanpin (Figure 15) was isolated by Um and coworkers in 2012 from a *Streptomyces* sp. strain.¹ Chemical structure and amino acid sequence was determined by 1D NMR (¹H and ¹³C) and 2D NMR experiments (COSY, TOCSY, HMQC and HMBC) in pyridine-*d*₅. The 3D structure conformation was determined as well by HMBC and ROESY NMR data. The absolute configuration was established by acid hydrolysis. The NMR data were deposited in the Biological Magnetic Resonance Bank (accession number: BMRB 19154). Afterwards, sungsanpin was isolated again from another bacterial strain and it was characterized by 1D and 2D NMR experiments in DMSO-*d*₆.²

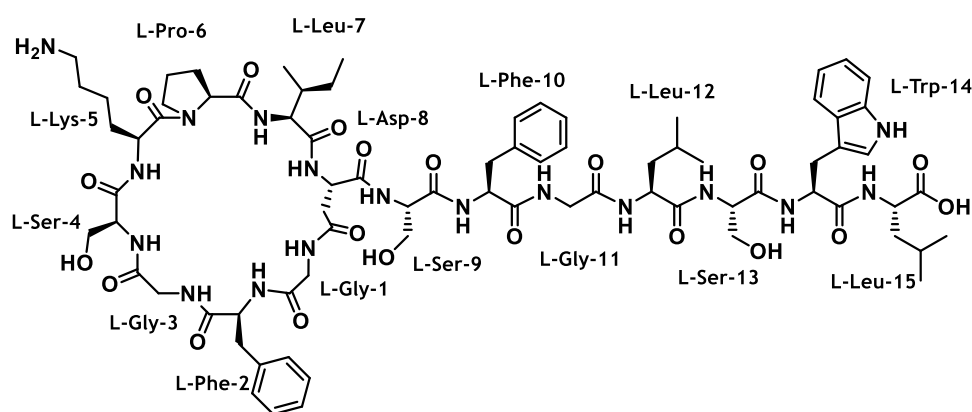


Figure 15: Chemical structure of sungsanpin.¹

On the basis of ROESY correlations, the 3D structure of sungsanpin was calculated. These results demonstrated that it holds a lasso structure (Figure 16). As it can be observed, all amino acids adopt a right-handed conformation. Ser13 was located in the middle of the ring, acting as the top plug whereas Trp14 acts as the bottom plug.

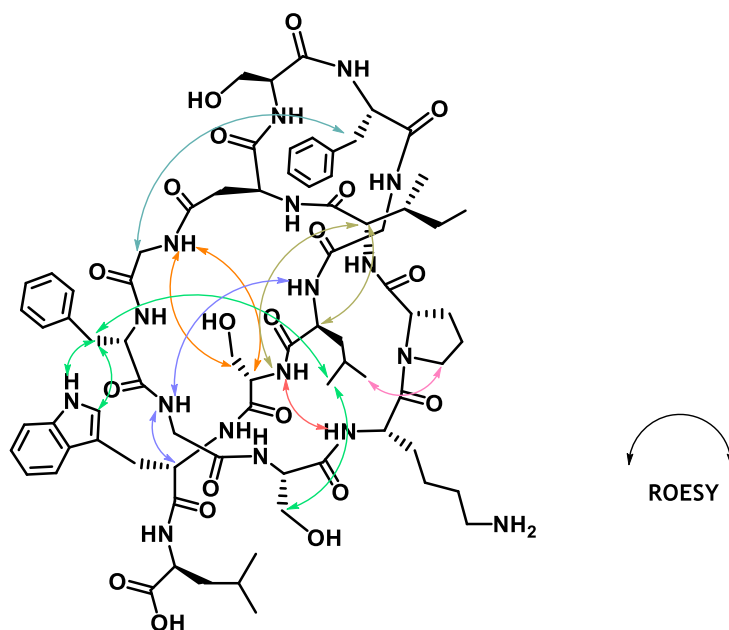


Figure 16: Key ROESY correlations of sungsanpin. Extracted from Um and coworkers.¹

Towards the synthesis of sungsanpin

General considerations

Sungsanpin has an 8-amino acids ring and a 7-amino acids tail that is threaded through the ring. The C-terminus remains in its free form, providing a negative charge, while the N-terminus is involved in the ring formation with the side chain of Asp8. All the amino acids are proteinogenic and commercially available. Added to this, all the linkages present in the compound are amide bonds, including the one involved in the macrolactamization. Regarding this, the difficulty of this synthesis resides in the 3D structure. No successful chemical synthesis approaches to lasso peptides have been reported so far.

Protection scheme

SPPS typically uses two protection strategies: Boc/Bzl and Fmoc/*t*Bu. The first strategy requires the use of harsh acid conditions, such as hydrofluoric acid (HF), in the final step for the removal of peptide chain from the solid support, while the second enables peptide synthesis in mild conditions. Since the desired peptide to be synthesized is relatively short and no aggregation was envisaged, Fmoc chemistry was

chosen for SPPS. Moreover, the stability of the ester bond of the analogs in front of strong acid was unknown.

Polymeric support

The permanent protecting group at the C-terminus was chosen among all the polymeric supports that provides the synthesis of peptide-acids. 2-chlorotriyl chloride resin (2-CTC)³ was initially chosen because it minimizes diketopiperazine (DKP) formation and furnishes the completely protected peptide after cleavage, enabling the key cyclization step not only on solid phase but also in solution. This resin is cleaved under extremely mild acid conditions (1% TFA in DCM, a mixture of AcOH and TFE in DCM, or HFIP in DCM), providing a partially orthogonal protection scheme. Furthermore, there are other resins that lead to peptide-acids, such as Wang-type resins. They are acid resistant and can be compatible with the temporary protecting trityl (Trt) or 2-phenyl isopropyl (2-PhⁱPr) groups. These resins are cleaved under 50 - 100% TFA conditions.

Coupling conditions

Activation of the carboxylic acid moiety of the amino acid is required to be able to react with the α -amino group of the growing peptide chain. Potential side reactions with carbodiimide-based reagents include the *O*-to-*N* rearrangement of the *O*-acylisourea intermediate by formation of the symmetrical anhydride, which can lead to epimerization. To avoid these side reactions different additives are used, such as OxymaPure⁴ or HOAt.⁶

Herein, DIPCDI-OxymaPure was used as the main coupling system for the amide formation. For the coupling of difficult buildings blocks or synthetic amino acids, HATU-HOAt-DIEA was chosen.⁵ However, HATU can lead to the formation of a guanidine side-product. To avoid this problem, preactivation of the reactive amino acid with the coupling reagent and DIEA was required, and also the use of short reaction times and a small excess of the aminium reagent. The coupling system PyBOP-HOAt-DIEA was used for the macrolactamization step. Phosphonium salts, such as PyBOP, have been described to be more convenient and efficient than uronium salts in slow couplings or when the carboxylic function is not in excess with respect to the amino group.⁷

Ester bond conditions

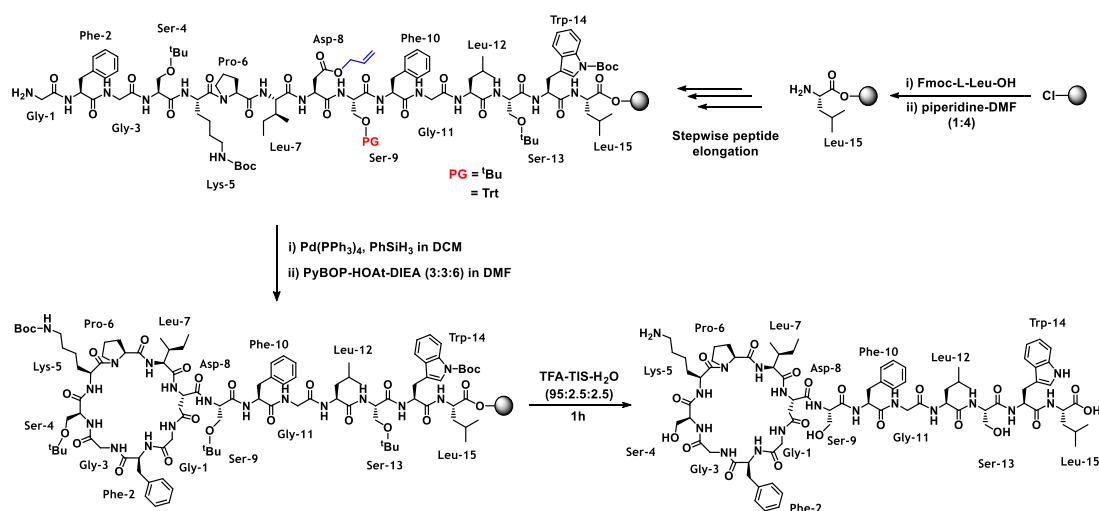
As mentioned in the previous section, phosphonium and aminium salts are highly efficient coupling reagents for amide bond formation. However, the active species formed are less appropriate for the formation of ester bonds. Thus, the carbodiimide method was selected for the activation of the carboxylic acids during ester bond formation. Since the Fmoc/*t*Bu was the strategy selected, DIPCDCI was preferred because of the high solubility of the derived urea. For the activation of carboxylic acid, carbodiimides can work through different activated species such as the *O*-acylisourea, in which equal equivalents of amino acid and carbodiimide are used, or the symmetrical anhydride, in which more equivalents of amino acid than carbodiimide are required. The enhanced reactivity of *O*-acylisoureas cause them to react more rapidly with available nucleophiles than the symmetrical anhydrides.⁸ For the ester bond formation, catalytic amounts of DMAP are added to significantly enhance the reaction rate. However, the racemization rate is also increased. This issue can be minimized or even eliminated by using less than a 10% of DMAP with respect to the amount of the protected amino acid used, and adding it onto the resin after the reactive species are formed.

Due to the low dielectric constant of DCM, the *O*-acylisourea formation occurs spontaneously. For this reason, DCM was selected as the first solvent option. However, if the activation is performed in a more polar solvent such as DMF, an immediate reaction does not take place, and a mixture of *O*-acylisourea, symmetrical anhydride, *N*-acylurea and urea can be formed.⁹ DCM has been extensively used in SPPS; however, it has some limitations due to the poor solubility of amino acids and coupling reagents in this solvent. Moreover, it can react with amines, such as piperidine during the Fmoc deprotection step.¹⁰ For these reasons, DCM is a good solvent for ester formation but sometimes a 10% of DMF is added to the mixture to ensure the complete solubilization of the amino acid and DMAP.

Synthesis of the natural sungsanpin

Considering all the above-mentioned characteristics, the strategy for the sungsanpin synthesis was designed. Two approaches were required to achieve the final product. The peptide chain was elongated in the $C \rightarrow N$ terminal direction starting from the C -terminus of Leu15. The main goal of this synthesis was to observe the spontaneous structural conformation that the peptide adopted and compare it with the already described natural product.

First approach



Scheme 1: First synthetic approach toward natural sungsanpin.

In the first approach (Scheme 1), the linear peptide was elongated on 2-chlorotrityl chloride resin (2-CTC). Allyl was used for the protection of Asp side-chain due to the need of an orthogonal protection during the cyclization step.

During the peptide chain elongation, an aspartimide was formed between Asp8-Ser9 when *tert*-butyl (*t*Bu) was used as the protecting group of Ser9. Thus, Trt, a more sterically hindered protecting group was used, but the aspartimide formation occurred in a similar proportion (Figure 17). Then, 0.1 M of OxymaPure was added into the piperidine solution to decrease the nucleophilicity of the amine during the Fmoc deprotection, but the side reaction was not eliminated.

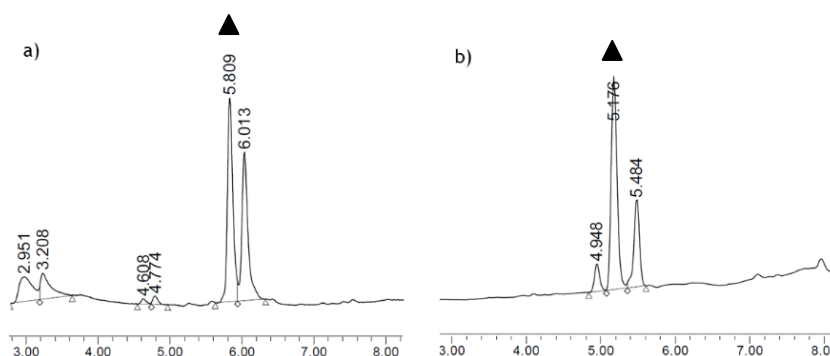
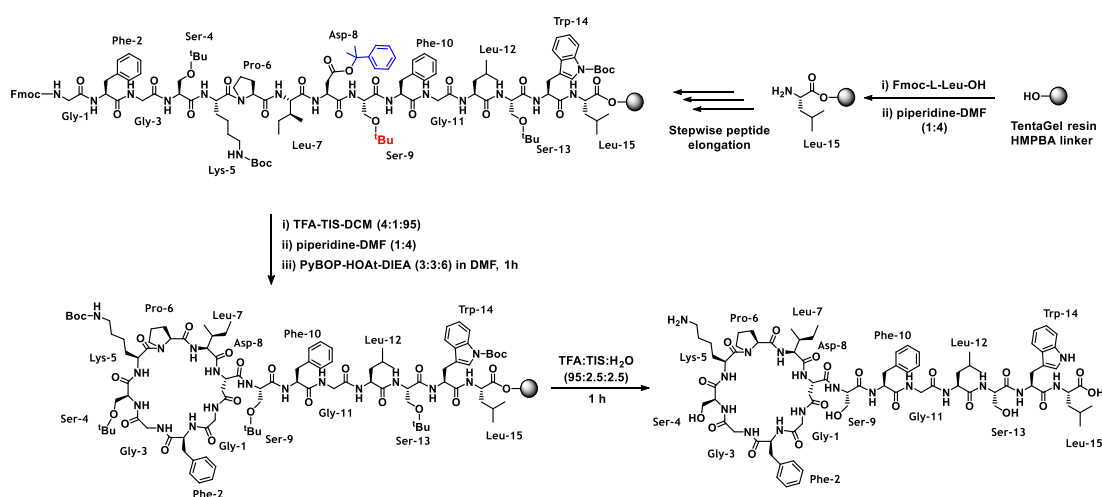


Figure 17: RP-HPLC chromatogram after the Pro6 coupling, using as a protecting group of Ser a) *t*Bu and b) Trt. Conditions: Sunfire™ C18 3.5 μ m (4.6 mm x 100 mm) reversed-phase analytical column; linear gradient from 30% to 100% of ACN over 8 min at 25 °C. \blacktriangle = aspartimide product.

Second approach



Scheme 2: Second synthetic approach toward natural sungsanpin.

In the second approach (Scheme 2), the same synthetic route than in the first approach was followed, but with 2-PhⁱPr ester instead of allyl ester as a protecting group for Asp8. The advantage of 2-PhⁱPr is that it can be removed with mild acid conditions and is compatible with Boc and *t*Bu. However, it is not compatible with 2-CTC resin. For this reason, a Wang-type linker, such as HMPBA on a TentaGel S-NH₂ support, was selected. The resin loading was lowered to ensure the correct cyclization and to avoid cyclodimerization.

The peptide chain was elongated from the C-terminus of Leu15 until the N-terminal Gly1. Then, the 2-PhⁱPr protecting group and the N-terminal Fmoc were

removed and the cyclization step was performed on solid phase. Finally, cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously.

Macrolactamization

This step was carried out after the complete elongation of the peptide chain while all the side-chains were protected. The 8-residue length favors the intramolecular cyclization. The cyclization was carried out with PyBOP-HOAt-DIEA and was monitored after 15, 30 min and 1 h to avoid long reaction times. The obtained crude was clean and without side products as analyzed by RP-HPLC (Figure 18).

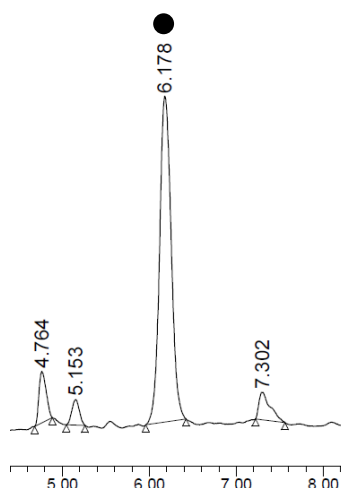


Figure 18: RP-HPLC chromatogram after the macrolactamization reaction. Conditions: Sunfire™ C18 3.5 μm (4.6 mm x 100 mm) reversed-phase analytical column; linear gradient from 25% to 50% of ACN over 8 min at 25 °C. ● = desired product.

Specific considerations

The HMPBA linker was attached to the resin with DIPCDI-OxymaPure and the first amino acid was then introduced through an ester bond via *O*-acylisourea activation with satisfactory loadings (0.25 mmol/g). Efficient couplings were obtained, and only Ile7 and Lys5 were re-coupled.

The synthesis of the linear peptide was optimized for the scale-up and for the automated synthesis. For the scale-up, double number of equivalents were used for the ester formation and the reaction time was reduced to 2 x 2 h, instead of overnight treatment. Normally, extending the time does not imply higher yield of the

ester product, whereas the chances of side reactions are increased. Thus, it is preferred to repeat the coupling with a second addition of reagents.

For the automated synthesis, the first three amino acids were coupled manually to avoid DKP formation. The cyclization and cleavage were also performed manually. The 2-Ph^tPr protecting group was compatible with the automated solid-phase conditions, since it is stable against high temperature and microwave conditions.

Synthesis of the bicyclic analogs via ester bond

The next step was the synthesis of bicyclic sungsanpin analogs, in which the lasso structure was forced by covalent bonds. The objective was to tie the peptide tail to the ring using an ester bond, which could later be removed easily while maintaining the final structure. To form the ester bond, one modification in the peptide sequence was required: the replacement of Gly³ by Glu³, such that an ester bond could be formed between the side-chain of Glu³ and the hydroxyl group of Ser¹³.

Four different strategies have been designed to produce different structural bicyclic analogs. The main goal during the design of each strategy was to modify the position of the tail with respect to the ring creating different 3D structures. Further analysis and comparison between them, will allow to find the most similar structure to lasso peptides.

Strategy A

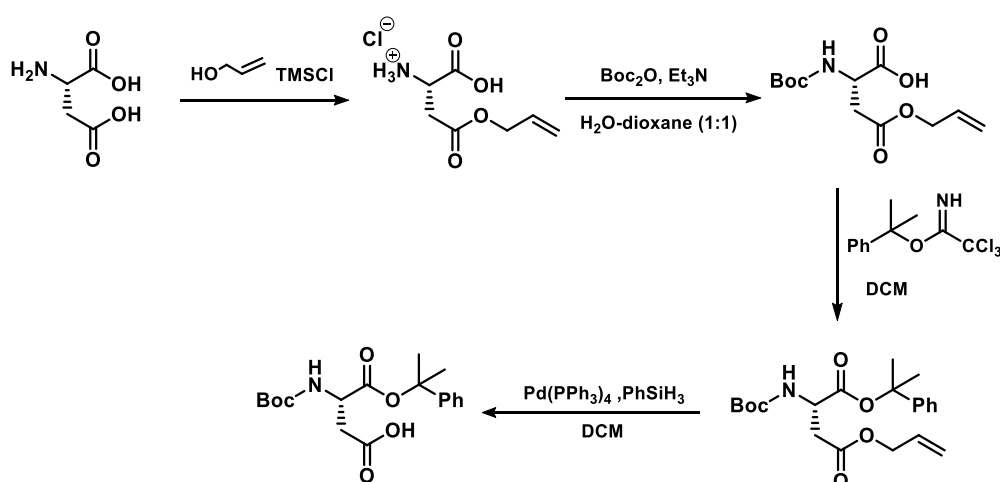
In the first strategy, the peptide chain was elongated in the $C \rightarrow N$ terminal direction starting from the C -terminus of Ile⁷. Regarding the synthesis of the natural sungsanpin (Scheme 2), the main problem was the steric hindrance of the tail generated by the C -terminal amino acids. The bulky side chains made the closure of the macrolactam ring difficult. For this reason, the main intention in this strategy was to perform first the bicycle, via the ester and the macrolactam bond and then, the coupling of the two last amino acids of the tail in order to avoid the steric hindrance of the C -terminus during the amide cyclization.

To carry out this strategy, two parallel peptide chains—linked by an ester bond—were elongated. One of these chains corresponded to the ring and the other chain to the loop. Secondly, the joining of the other extreme of the loop with the ring was performed, followed by a last cyclization step to close the ring. Finally, the peptide tail was coupled. Two approaches were designed to achieve the first bicyclic sungsanpin analog. In the first approach (Scheme 5), a basic design was set up and in the second (Scheme 7), the synthesis was improved with the implementation of some modifications. To carry out these synthesis, several synthetic amino acids were required. The synthesis described in Schemes 3 and 4 are common for both approaches.

Synthesis of Boc-L-Asp-O(2-PhⁱPr)

This amino acid was used in both approaches and was a key amino acid for the synthetic strategy. The main objective was the protection of the α -carboxylic acid with the 2-PhⁱPr ester and the orthogonal protection of the amino group.

On one hand, the corresponding imidate was prepared from the commercially available isopropyl alcohol and trichloroacetonitrile. On the other hand, the carboxylic acid side-chain of Asp was protected using chlorotrimethylsilane (TMSCl) and allyl alcohol, forming the ω -allyl ester in excellent yield. Then, the amino group was protected using di-*tert*-butyl dicarbonate (Boc₂O) (Scheme 3).

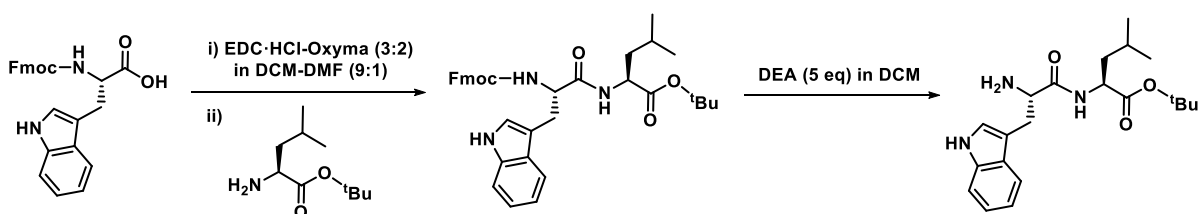


Scheme 3: Synthetic strategy to the Boc-L-Asp-O(2-PhⁱPr) residue.

Synthesis of the dipeptide H-Trp-Leu-OtBu

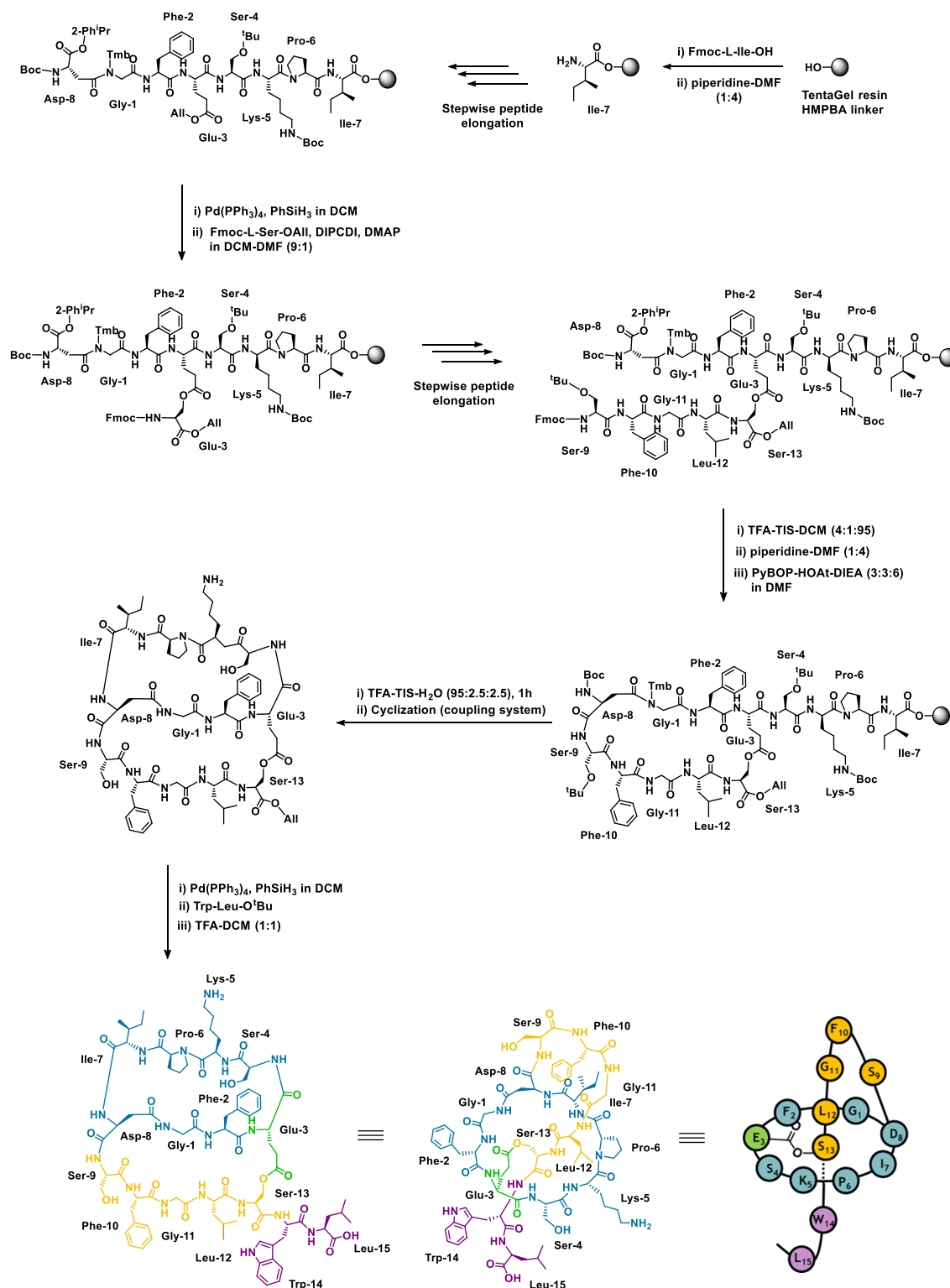
H-Trp-Leu-OtBu corresponds to residues 14 and 15 at the C-terminus of sungsanpin. The coupling of this dipeptide was a key reaction to ensure the total synthesis. It was synthesized with the objective to couple two amino acids in one step, thus avoiding two extra reaction steps.

This dipeptide was synthesized in two steps in solution (Scheme 4). Trp was activated with carbodiimide and OxymaPure and Leu was then added. Regarding the carbodiimide, EDC·HCl was used instead of DIPC DI because it is water soluble and easily removable in solution. For the Fmoc deprotection, diethylamine (DEA) was used because it allows the use of DCM and it is also easily eliminated in solution. Furthermore, *t*Bu was selected as the carboxylic acid protecting group of Leu15 because it was orthogonal and easily cleaved with 50% TFA in DCM.



Scheme 4: Synthetic strategy to the dipeptide H-Trp-Leu-OtBu

First approach



Scheme 5: First synthetic approach to the strategy A toward the bicyclic sunsanpin analog

In the first approach to the synthesis of the first bicyclic sunsanpin analog, HMPBA was selected as a Wang-type linker on TentaGel support with a low loading

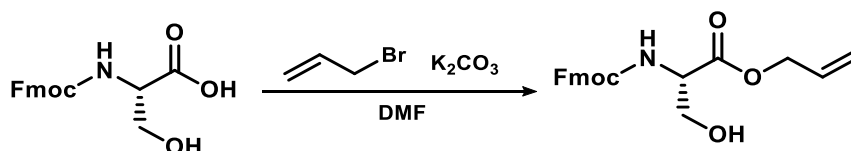
(0.29 mmol/g) (Scheme 5). In this approach commercially available amino acids have been mostly used. Only two synthetic amino acids were required (Scheme 3 and 4).

To avoid aspartimide formation at the Asp8-Gly1 sequence, 2,4,6-trimethoxybenzyl (Tmb) backbone protection was used for Gly1. Allyl was selected as protecting group of the α -carboxylic acid group of both Glu3 and Ser13. This was done to ensure orthogonal deprotections before the ester bond formation (in the case of Glu3) and before the dipeptide coupling (Ser13).

Asp8 was an important residue in the peptide sequence because it was the only residue involved in the ring and the loop. For this reason, a new protection strategy was used for this residue: Boc (instead of Fmoc) was used as protecting group of the amino group and 2-PhⁱPr was employed for the protection of the α -carboxylic acid, remaining the β -carboxylic acid free.

Synthesis of Fmoc-L-Ser-OAll

Herein, the carboxylic acid was protected as allyl ester through a nucleophilic substitution reaction. This protecting group provided full orthogonality with other protecting groups, such as Boc or Fmoc (Scheme 6).



Scheme 6: Synthetic strategy to the Fmoc-L-Ser-OAll residue.

Ester formation

The formation of the ester bond proceeded smoothly via the *O*-acylisourea activation. A high excess of DIPCDI and Fmoc-L-Ser-OAll was used to ensure its fast and efficient formation. The initial coupling mixture was at pH 9 and it was acidified to pH 7 with HCl in dioxane to avoid the formation of side-products due to the basic pH (Figure 19). Short reaction times and high excesses were needed to avoid the *N* → *O* transacylated product, which is a rapid intramolecular side reaction that normally occurs when the α -amino group is unprotected, or at basic pH, in which the amino group can be deprotonated.

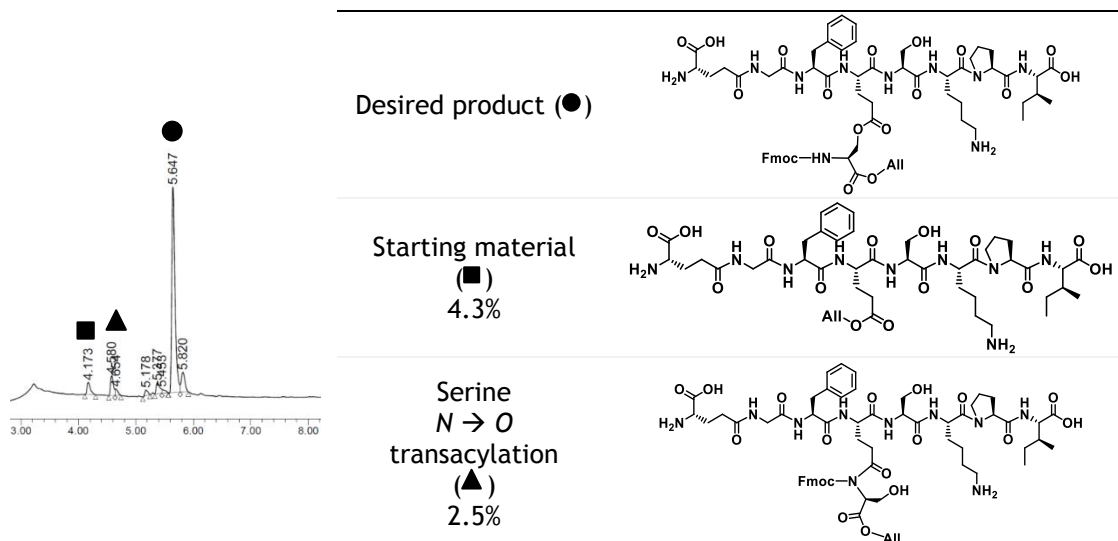


Figure 19: RP-HPLC chromatogram after the formation of the ester bond. Conditions: Xbridge™ C18 2.5 μ m (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 0% to 100% of ACN over 8 min at 25 $^{\circ}$ C.

Peptide chain elongation

After the formation of the ester bond, the second peptide chain was elongated using an Fmoc/*t*Bu strategy. Here, the main problems detected were the hydrolysis of the ester bond and the increase of serine transposition due to the use of piperidine during the Fmoc deprotection. Another side reaction that was detected was the DKP formation between Leu12 and Ser13, capping the peptide chain and decreasing the final yield (Figure 20). To avoid these side reactions, 0.1 M OxymaPure was added into the piperidine-DMF mixture to reduce the nucleophilicity of the amino group, but it was not an efficient solution. More experimentation was required to optimize this step and avoid or minimize these side reactions, before proceeding with the synthesis.

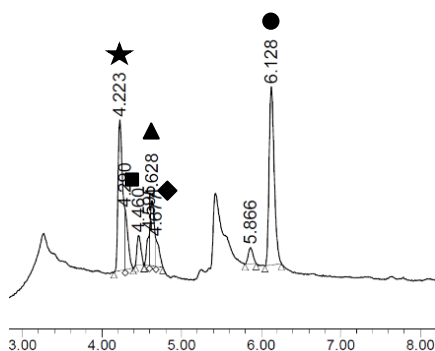


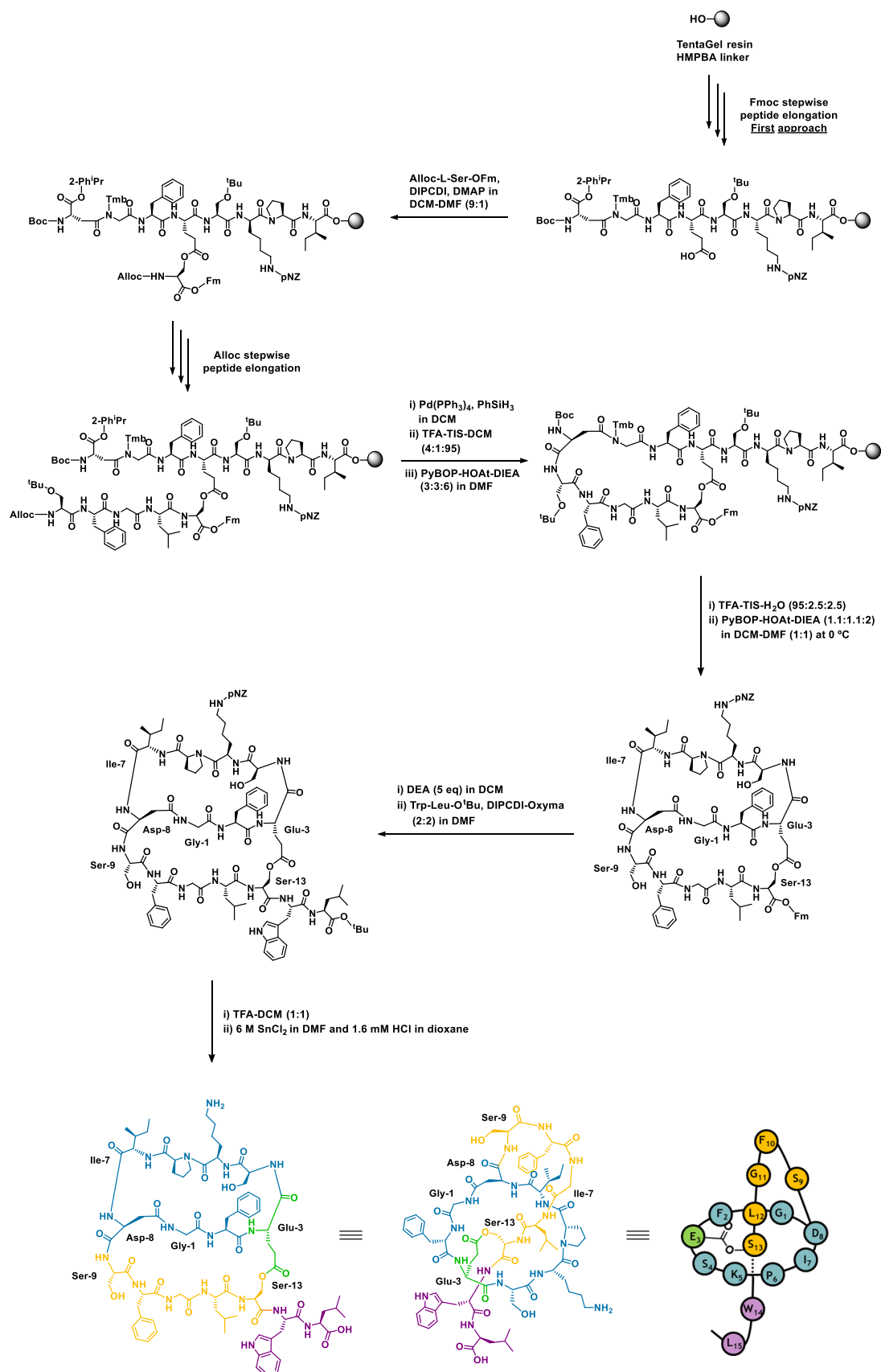
Figure 20: RP-HPLC chromatogram after the Phe10 coupling of the second chain. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 0% to 100% of ACN over 8 min at 25 °C. ★ = Ester bond hydrolysis, ■ = starting material (Glu3 with allyl), ▲ = serine N \rightarrow O transacylation, ◆ = DKP formation and ● = desired product.

Specific considerations

The HMPBA linker was attached to the resin with DIPCDI-OxymaPure and the first amino acid was then introduced through an ester bond via *O*-acylisourea activation with satisfactory loadings (0.28 mmol/g). Efficient couplings were obtained, and only Lys5 and Asp8 required to be re-coupled.

DIPCDI-OxymaPure was the coupling system used for all the amino acids, except for the coupling of Asp8. In this case, stronger coupling conditions, such as HATU-HOAt-DIEA (3:3:6), were required due to the presence of highly hindered protecting groups, including the *N*-substituted Gly1.

Second approach

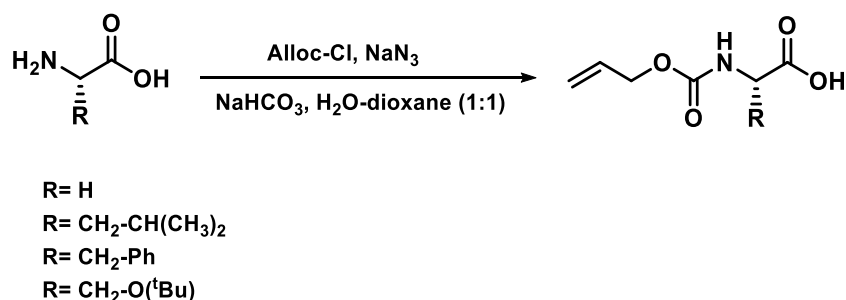


Scheme 7: Second synthetic approach to the strategy A toward the bicyclic sungsanpin analog

In this second approach to the strategy A, two main modifications were implemented with respect to the previous one to avoid the previous side reactions (Scheme 7). First, the second peptide chain was elongated via Alloc-chemistry instead of Fmoc to avoid the use of base during the deprotection step. This modification required the synthesis of Alloc-protected amino acids and the change of the protecting groups of Ser13. In this last case, 9-fluorenylmethyl ester (Fm) was used as a protecting group of the α -carboxylic acid rather than to protect the amino group, which this time was protected with Alloc. Secondly, the side-chain Boc protecting group of Lys5 was substituted by *p*-nitrobenzyloxycarbonyl (pNZ). This modification was performed because the free amino group from the Lys5 side chain could be a competitor during the second cyclization in solution, producing a non-desired cyclization.

Synthesis of Alloc-L-aa-OH

The amino group from Ser9, Phe10, Gly11 and Leu12 was protected with Alloc via “one pot” reaction using Alloc-azide as acylating agent (Scheme 8). The first step involves the *in situ* mixing of allyl chloroformate and sodium azide. By using this procedure, no temporary protection of the carboxyl function of the amino acid was needed.¹¹ All amino acids were obtained in excellent yields and no further purification was required. The advantage of this methodology was that azide byproducts could be easily removed.

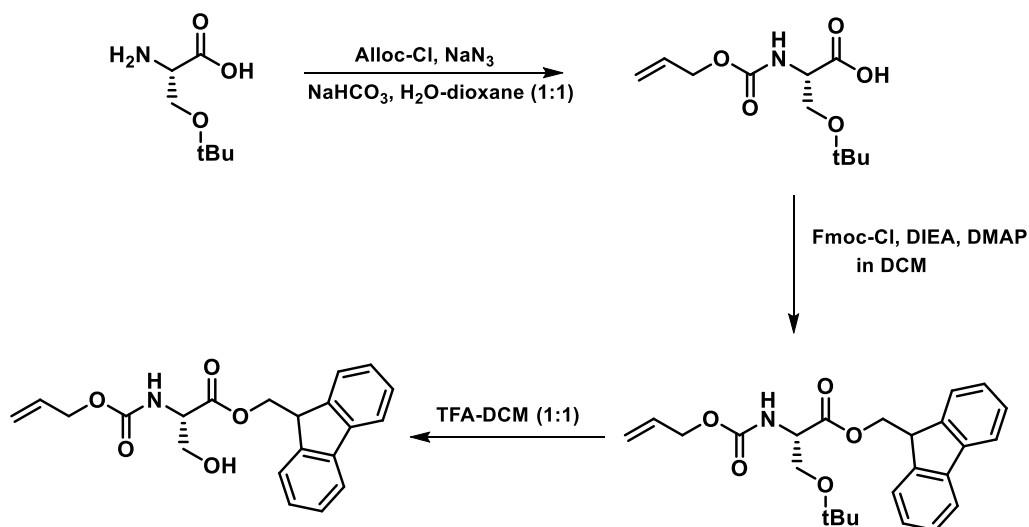


Scheme 8: Synthetic strategy to several Alloc-L-aa-OH residues.

Synthesis of Alloc-L-Ser-OFm

Ser13 is involved in ester formation via the hydroxyl group. For this reason, the other two reactive groups had to be orthogonally protected. First, the amino group was protected with Alloc using the methodology described in the above section.

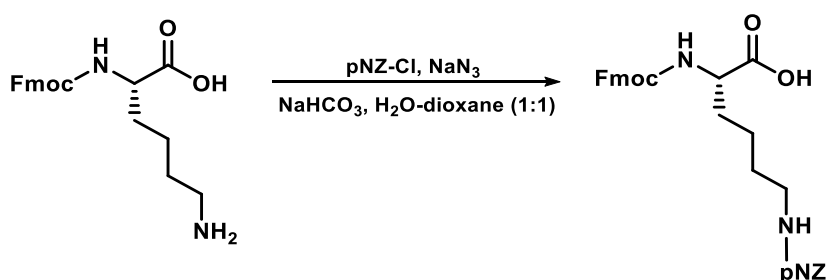
Then, the Fm ester was prepared by esterification using Fmoc-Cl, DIEA and DMAP. After the complete protection, the *t*Bu of the hydroxyl group was removed (Scheme 9).



Scheme 9: Synthetic strategy to the Alloc-L-Ser-OFm residue.

Synthesis of Fmoc-L-Lys(pNZ)-OH

Fmoc-L-Lys(pNZ)-OH was prepared from Fmoc-L-Lys-OH, where the pNZ protecting group was introduced using the previously described azide methodology in Scheme 8. 4-Nitrobenzyl chloroformate (pNZ-Cl) was mixed with sodium azide and Fmoc-L-Lys-OH was then added to the mixture, which led to the formation of the protected amino acid in excellent yields and without further purification (Scheme 10).



Scheme 10: Synthetic strategy to the Fmoc-L-Lys(pNZ)-OH residue.

Peptide chain elongation

The second peptide chain was elongated via Alloc-chemistry using the synthesized Alloc-aa-OH (Scheme 8). The elongation proceeded smoothly, and no side-reactions

were formed. Compared with the first approach, using Alloc-chemistry, the transacylation and the ester bond hydrolysis should be eliminated. Moreover, DKP formation was avoided using a bulkier protecting group of the Ser13 carboxylic acid, in this case an Fm ester instead of Allyl.

First macrolactamization on solid phase

The amide cyclization was performed using the conditions previously described in Scheme 2. The Alloc group of Ser9 was removed via palladium catalysis whereas the 2-PhⁱPr ester was cleaved under mild acidic conditions. The cyclization took place with PyBOP-HOAt-DIEA (3:3:6) at pH 8. The pH was adjusted to 7-8 to avoid the Fm elimination and the possible ester hydrolysis. The reaction proceeded satisfactorily.

Second macrolactamization in solution

This second cyclization took place in solution and different coupling systems, solvent mixtures and temperatures were tested. All the conditions are detailed in Table 2:

Assay	Coupling system	Concentration (mM)	pH	T (°C)	Time	Solvent
1	PyBOP-HOAt-DIEA (1.1:1.1:2)	0.1	7-8	25	1 h	DCM-DMF (9:1)
2	PyBOP-HOAt-DIEA (1.1:1.1:2)	0.1	7-8	0	30 min	DCM-DMF (9:1)
3	PyBOP-HOAt-DIEA (1.1:1.1:2)	0.1	6-7	0	30 min	DCM-DMF (9:1)
4	HOAt-DIPCDI (2:2)	0.1	--	0	6 h	DCM-DMF (9:1)
5	PyBOP-HOAt-DIEA (1.1:1.1:2)	0.1	7-8	0	3 h	DMF
6	PyBOP-HOAt-DIEA (2:2:4)	0.1	6-7	0	1 h	DMF
7	PyBOP-HOAt-DIEA (1.1:1.1:2)	0.1	6-7	0	20 min	DCM-DMF (1:1)
8	PyBOP-HOAt-DIEA (1.1:1.1:2)	0.2	6-7	0	20 min	DCM-DMF (1:1)
9	PyBOP-HOAt-DIEA (1.1:1.1:2)	0.4	6-7	0	20 min	DCM-DMF (1:1)
10	PyBOP-HOAt-DIEA (2:2:4)	0.8	6	0	2 h	DCM-DMF (1:1)

Table 2: Conditions for the cyclization in solution.

The cyclization was performed using high dilution conditions to avoid polymerization; 0.1 mM was selected as the starting concentration. The carboxylic acid was activated using the different coupling systems described in Table 1. DCM

was selected as the main solvent due to its ease of handling in solution, since DMF was not a good candidate due to its high boiling point and its miscibility with water.

The first assay was performed at 25 °C using the coupling system previously selected for the macrolactamization. However, three side reactions were detected: racemization, Fm deprotection and a highly hydrophobic side-product, which was uncharacterized. In assay 2, and to avoid these side reactions, the temperature and the pH were decreased to 0 °C and pH 6-7, respectively. Using these new conditions, two side reactions were eliminated, but the uncharacterized peak was still detected (Figure 21a).

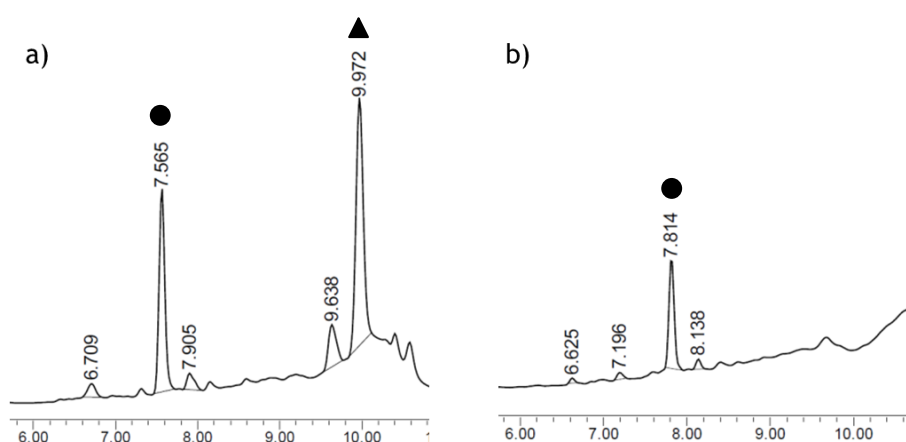


Figure 21: RP-HPLC chromatogram of the macrolactamization in solution using the conditions from: a) assay 2 or b) assay 7. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. ▲ = uncharacterized side-product and ● = desired product.

Afterwards, two parameters were separately modified: the coupling system and the solvent. Using a neutral coupling system (assay 3), the reaction did not proceed, even after long times. In the next assays, DMF was used as main solvent. In these cases, no side product appeared, but the conversion of the reaction was very low. Thus, the solvent mixture was optimized to DCM-DMF (1:1), yielding a good conversion without side products (Figure 21b). The reaction was scaled up to 0.8 mM without polymerization.

Coupling of the dipeptide

The neutral coupling system, DIPCDI-OxymaPure, was used to perform the dipeptide coupling. After 2 h the coupling was completed. The reaction worked smoothly.

Protecting groups removal

The last step consisted of the removal of the two last protecting groups: first, the *t*Bu group, using 50% TFA in a rapid and efficient reaction and then, the pNZ group. Regarding the last one, the most common method for the reduction of the nitro group to the corresponding amine is catalytic hydrogenation. However, in our case this methodology was discarded because the ester bond could be labile to hydrogenation. In Table 3, two other reducing methodologies are described:

Assay	Reducing agent	Solvent
1	4 to 16 eq Na ₂ S ₂ O ₄	ACN-EtOH-H ₂ O (4:1:1)
2	6 M SnCl ₂ with 1.6 mM HCl in dioxane	dry DMF

Table 3: Conditions for the removal of the pNZ.

First, sodium hydrosulphite was used due to the easy work-up. However, the reaction was very slow, even after the addition of a high excess of the nitro-reducing agent. After 11 h, the reaction was completed, but with several side products. Therefore, a stronger reducing agent, such as SnCl₂ with catalytic amounts of acid, was tested.¹² With these conditions a good yield was achieved in 5 h.

Specific considerations

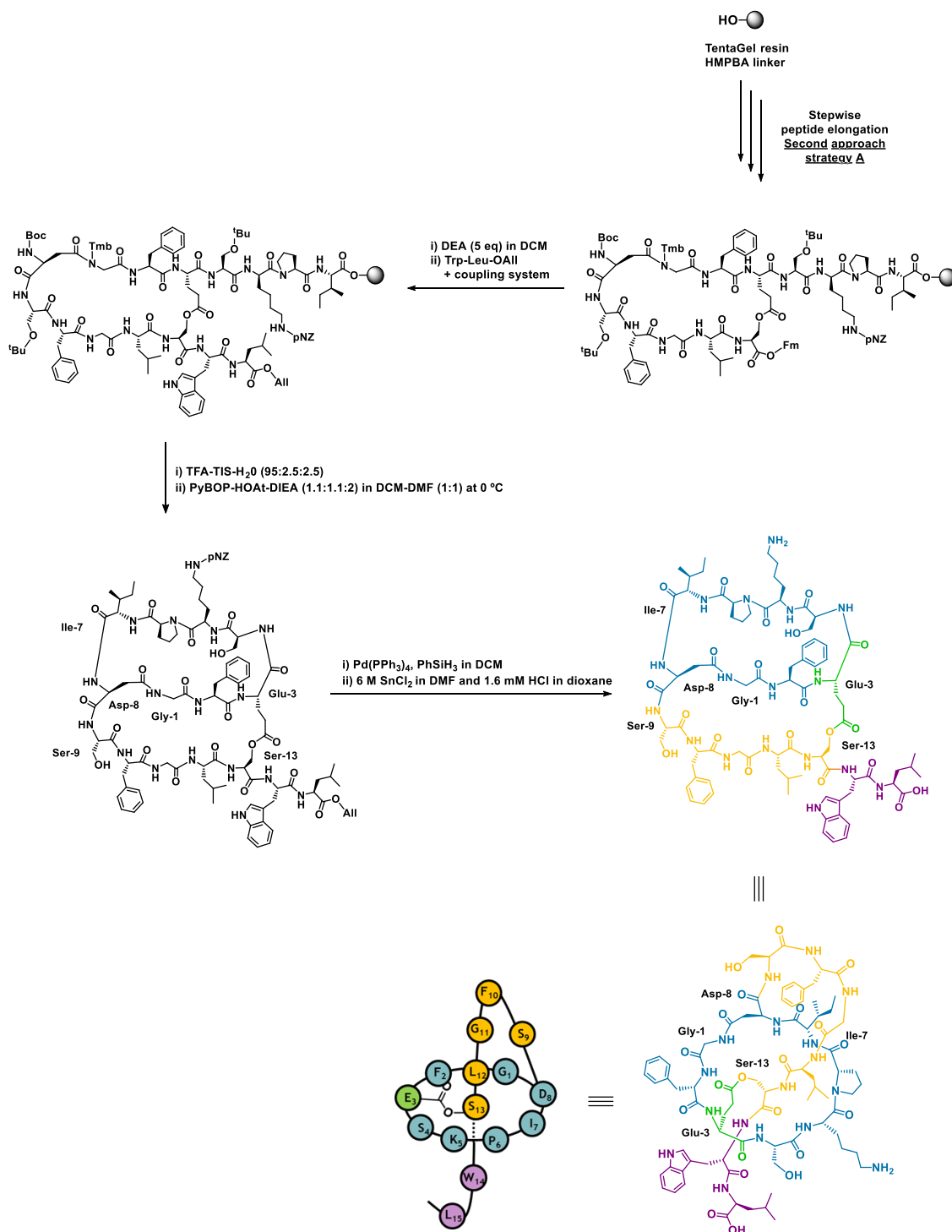
Once the peptide was cleaved from the resin, it was used without further purification. After the pNZ removal, the peptide was purified straightaway by means of a PoraPak Rxn cartridge, to remove the excess of DMF and salts, followed by a semi-preparative RP-HPLC purification.

This synthesis was scaled up. In this case, the same resin was used, but with a higher loading (0.54 mmol/g). The HMPBA linker and the first amino acid were attached to the resin using the same coupling system described before, yielding a

satisfactory loading (0.45 mmol/g). All reactions worked well in the conditions described for the small scale. The only exceptions were: the macrolactamization in solution, which was optimized to 0.8 mM instead of 0.1 mM, and the dipeptide coupling, in which more time was required.

Using this second strategy, we were able to obtain the first bicyclic sungsanpin analog through an ester bond linkage. The integrity of the ester bond was preserved by avoiding the use of harsh basic conditions, thus using Alloc-chemistry and SnCl₂ instead of hydrogenation conditions. Moreover, the C-terminal tail (Trp14-Leu15) was introduced in a straightforward way after the bicycle formation. The main advantage of this last coupling was that the steric hindrance of the ring was avoided, and the tail could be introduced through the middle of the ring. Further analysis will confirm the structure (Chapter 2).

Strategy B



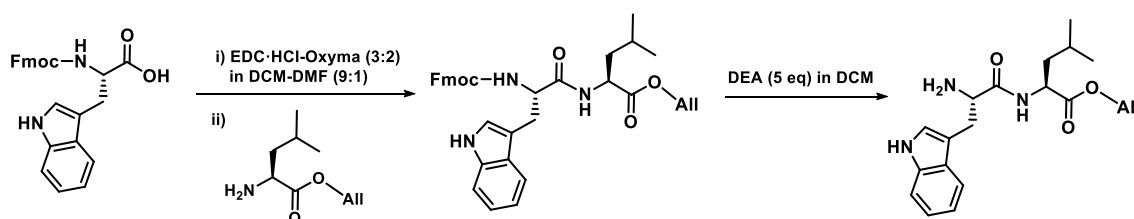
Scheme 11: Synthetic strategy B toward the bicyclic sungsanpin analog.

Compared with strategy A, the only difference in this strategy was the order followed to couple the peptide tail. Here, the peptide tail was coupled before the formation of the second cycle with the objective to determine if the ester bond linkage was able to fix the tail inside the ring during the cyclization (Scheme 11).

The first synthetic steps were the same as for the strategy A. However, instead of performing the cleavage after the first cyclization, this time the C-terminal dipeptide was coupled previously. Finally, the cleavage from the solid support was performed and the macrolactamization was carried out in solution. Therefore, another protecting group of the dipeptide had to be selected. To assure the orthogonality during the second cyclization, allyl was used instead of *t*Bu.

Synthesis of the dipeptide H-Trp-Leu-OAll

H-Trp-Leu-OAll was synthesized in solution in two steps, following the same methodology as in Scheme 4. The only difference was the protecting group of the Leu carboxylic acid. In this case, an allyl ester was selected due to its orthogonality during the cyclization in solution.



Scheme 12: Synthetic strategy to the H-Trp-Leu-OAll dipeptide.

Coupling of the dipeptide

Several conditions, detailed in Table 4, were tested to perform the dipeptide coupling on solid phase.

Assay	Active ester	Coupling system (eq)	Dipeptide (eq)	Pre-activation time (min)	pH	T (°C)	Time (h)	Solvent
1	--	DIPCDI-OxymaPure (3:3)	3	3	7	25	6	DMF
2	--	DIPCDI-OxymaPure (5:5)	5	3	7	25	2.5	DMF
3	--	DIPCDI-HOAt (5:5)	5	3	7	25	16	DMF
4	--	HATU-HOAt-DIEA (3:3:6)	3	3	8	25	3	DMF
5	--	HATU-HOAt-DIEA (3:3:6)	3	3	8	25	6	DMF
6	--	HATU-HOAt-DIEA (5:5:10)	5	3	8	25	5	DMF
7	--	HATU-HOAt-DIEA (10:10:20)	10	10	8	25	5	DMF
8	--	HATU-HOAt-DIEA (3:3:6)	3	3	8	40	5	DMF
9	--	HATU-HOAt-DIEA (5:5:10)	5 (HM15)	3	8	25	5	DMF
10	PfpOH-DIPCDI-DMAP (5:5:0.25) in dry DCM 1 h + 16 h	OxymaPure (5)	5	--	7	25	16	DMF
11	PfpOH-DIPCDI-DMAP (15:15:0.5) in dry DCM 1 h + 16 h	OxymaPure (10)	10	15	7	25	2 + 16	DMF
12	--	DIPCDI-OxymaPure (10:10)	5	15	7	25	16	DMF

13	--	HATU-HOAt-DIEA (10:10:20)	5	15	7	25	6	DMF
14	(*)	HATU-HOAt-DIEA (10:10:20)	10	5	8	25	5	DMF
15	--	DIPCDI-HOAt (10:10)	10	15	7	25	48	DMF
16	--	DIPCDI-HOAt (10:10)	10 (HM15)	15	7	25	5	DMF
17	--	DIPCDI-HOAt (10:10)	10	3	7	25	6	DMF
18	--	PyAOP-HOAt-DIEA (10:10:20)	10	3	8	25	6	DMF
19	--	PyAOP-HOAt-DIEA (10:10:20)	10	3	8	40	6	DMF
20	--	DIPCDI-HOAt (10:10)	5	5	7	25	5	ACN
21	HOSu-DIPCDI- DMAP (10:10:0.5) in dry DCM 2 h + 3 h	--	10	--	7	25	1.5	DMF
22	HOSu-DIPCDI- DMAP (10:10:0.5) in dry ACN 2 h + 48 h	--	10	--	7	25	3	DMF

Table 4: Conditions for the coupling of the dipeptide on solid phase. (*) 20% DEA in DCM was used for the Fm deprotection. In assays 11 to 20, a previous wash before dipeptide addition was performed. HM15: dipeptide used in the strategy A, with *t*Bu.

The neutral coupling system DIPCDI-OxymaPure was chosen to perform the coupling. However, no coupling was observed, and the number of equivalents was increased. In assay 3, OxymaPure was replaced by a stronger additive, like HOAt, but no desired product was detected even after an overnight reaction time.

Thus, for the next assays the coupling system was changed to HATU-HOAt-DIEA. The pH was kept around 8 to avoid the cleavage of the ester bond. The number of equivalents, the pre-activation time and the temperature were increased to facilitate the coupling, but no desired product was observed (assays 4 to 9). However, the peak of the starting material disappeared, forming a new peak with a mass of + 56 Da higher than the desired mass (Figure 22). The filtrate of the reaction mixture was collected and analyzed by LC-MS. A new main peak with a mass of + 100 Da appeared, corresponding to the guanidinylation of the dipeptide. For this reason, the dipeptide could not react.

In assay 9, another dipeptide was assayed in order to discard the possibility that the guanidinylation was dependent on the dipeptide nature (Scheme 4). However, the same peaks with the same mass (+ 100 Da) were observed.

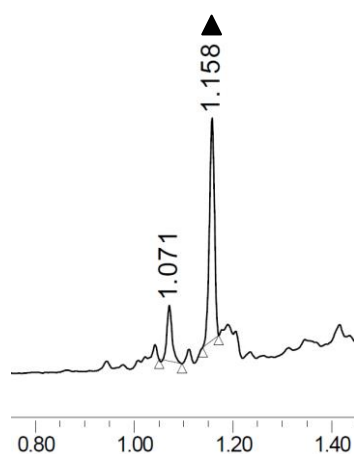


Figure 22: RP-UPLC chromatogram of assays 7, 8, 9 after the dipeptide coupling. Conditions: Acquity BEH C18 1.7 μm (2 mm x 50 mm) reversed-phase column; linear gradient from 30% to 100% of ACN over 2 min at 40 $^{\circ}\text{C}$. \blacktriangle = unknown side product (+ 56 Da).

We hypothesized that maybe the carboxylic acid was not easily accessible and the active ester was therefore not formed. For this reason, the side reactions would probably be faster than the desired one. Moreover, this would be in agreement with the evidence that when the equivalents of coupling system or dipeptide were increased, no improvement was observed.

In view of these problems, the activated ester of the carboxylic acid was changed. Pentafluorophenyl ester (OPfp) was tested in assays 10 and 11. The advantage of the OPfp was that it can be isolated after its formation due to its high stability. The OPfp was prepared on solid phase, by adding DIPCDI and 2,3,4,5,6-pentafluorophenol

(PfpOH) to the resin, followed by the addition of DMAP. The reaction was monitored by RP-UPLC, but no ester was formed, only starting material was detected together with a new peak with a mass of + 126 Da. This side product corresponded to the rearrangement of the *O*-acylisourea to an inactive *N*-acylurea (Figure 23).

From these assays, it could be extracted that the reaction rates of the ester formation and the dipeptide coupling were really slow compared with the side reactions, which were faster. For this reason, in the next conditions (assays 11 to 20) the resin was washed with DMF to remove the excess of coupling system after the pre-activation.

The previous coupling systems and conditions were tested again, but with the washes of the resin above described, and using also a high excess of equivalents. No conversion was observed for any condition, but the guanidinylated product was not detected.

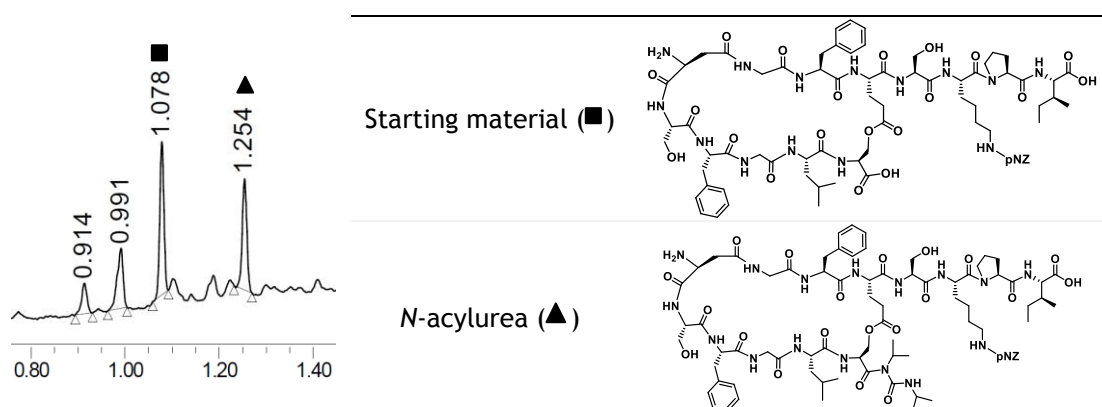


Figure 23: RP-UPLC chromatogram after the OPfp formation. ■ = starting material, ▲ = *N*-acylurea. Conditions: Acquity BEH C18 1.7 μ m (2 mm x 50 mm) reversed-phase column; linear gradient from 30% to 100% of ACN over 2 min at 40 °C.

The difference between assays 13 and 14 was based on the solution used for Fm deprotection. In assay 14, the Fm group was removed with DEA-DCM (1:4) instead of DMF in order to investigate whether the problem came from the deprotection step. However, it was found that DCM was not a good alternative since the deprotection was slow and more treatments were required.

Regarding assays 18 and 19 (Figure 24), we tested PyAOP,¹³ which is the most powerful phosphonium derivative and does not side react at the *N*-terminus of the peptide. Nevertheless, in these assays, the same side product than in assays 7, 8 and

9 was formed, although part of the starting material remained unreacted. Added to this, racemization in both peaks was observed, due to its high reactivity. As expected, the dipeptide was recovered intact after the coupling.

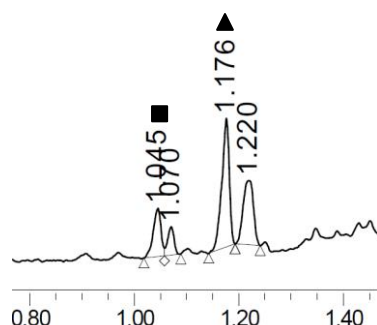


Figure 24: RP-UPLC chromatogram of assays 13 and 19 after the dipeptide coupling. Conditions: Acquity BEH C18 1.7 μm (2 mm x 50 mm) reversed-phase column; linear gradient from 30% to 100% of ACN over 2 min at 40 $^{\circ}\text{C}$. ■ = starting material, ▲ = unknown side-product (+ 56 Da). The double peaks indicate racemization.

Generally, DMF, *N*-methyl-2-pyrrolidone (NMP) and DCM are used as a solvent for SPPS, but they are toxic and cause environmental problems when are present in waste.¹⁴ For this reason, there is an urgent need to look for alternatives solvents, especially due to future restrictions caused by the implementation of the Registration, Evaluation and Authorization of Chemicals (REACH) program.^{10,15} As an alternative, acetonitrile (ACN) and tetrahydrofuran (THF) have been used in SPPS, but neither of them are considered green solvents.¹⁶ Regarding ACN, several years ago it was demonstrated that in combination with PEG-based resins, ACN was a good alternative to DMF for peptide coupling.¹⁷ Thus, ACN was selected in assay 20 as the solvent for performing the dipeptide coupling. Even though TentaGel was perfectly swollen in ACN and the dipeptide was soluble, no reaction was detected.

Going back to the hypothesis that the problem lied in the formation of the active ester, the succinimide ester (OSu) was tested in the last assays (Figure 25). OSu is compatible with solid phase and is also stable to the RP-HPLC conditions to be detected. The OSu ester was prepared on solid phase by adding a large excess of DIPCDI and *N*-hydroxysuccinimide (HOSu) to the starting material,¹⁸ followed by the addition of DMAP. However, the same side product than in 10 and 11 was obtained (Figure 23).

In assay 22, ACN was used as solvent for the ester formation and for the first time, the mass of the active ester was detected, but with very low intensity. Once the

ester bond was formed, the dipeptide was added to the resin, but no reaction took place.

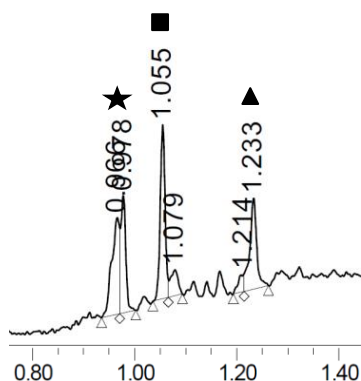


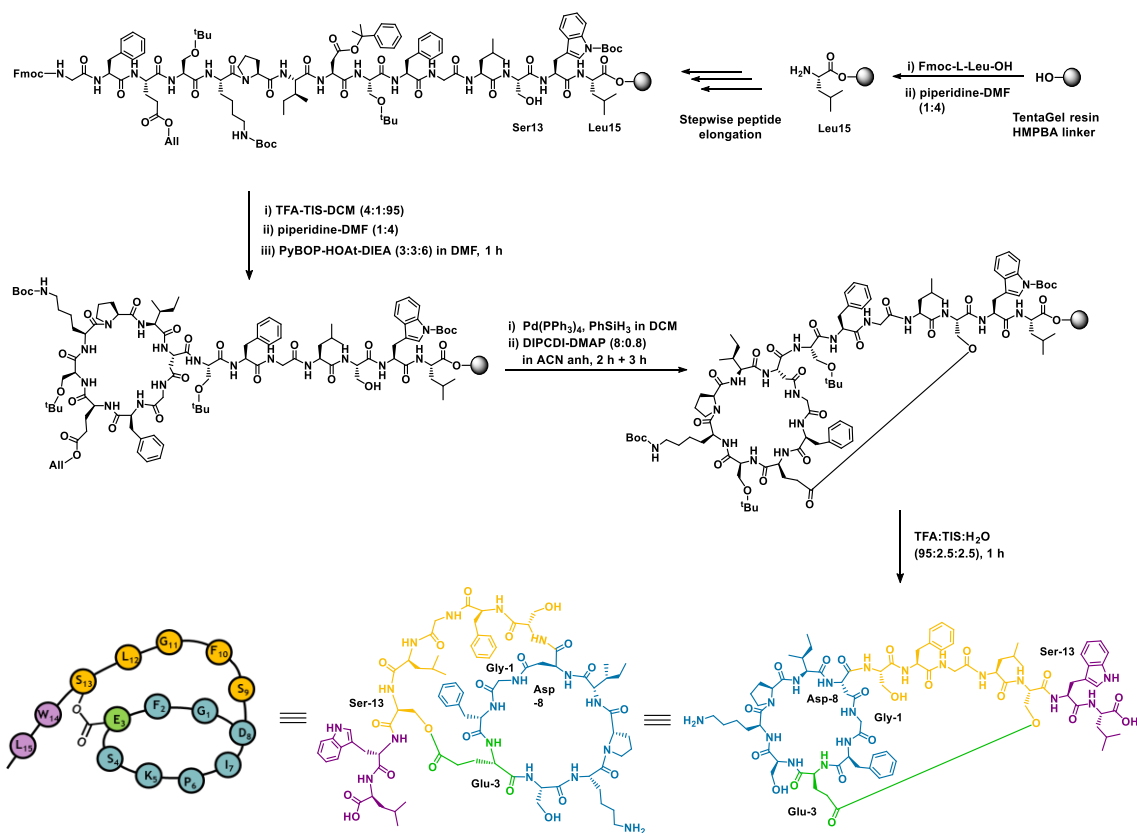
Figure 25: RP-UPLC chromatogram after the succinimide ester formation. Conditions: Acquity BEH C18 1.7 μm (2 mm x 50 mm) reversed-phase column; linear gradient from 30% to 100% of ACN over 2 min at 40 $^{\circ}\text{C}$. ■ = starting material, ▲ = O-acylurea rearrangement, ★ = unknown side reaction with m/z 1650 and 1764.6.

Specific considerations

Once the two peptide chains were linked through the ester bond, they were elongated, and after performing the cyclization on solid phase, the peptidyl-resin was split in two: one part was used to continue with strategy A and the other part was used for strategy B.

A large list of conditions was tested for the dipeptide coupling on solid phase, however none of them gave satisfactory results. Probably, the problem derived from the steric hindrance of the solid support and the resulting unapproachability to the free carboxylic acid. However, in strategy A, the same dipeptide was coupled in solution without any difficulty. In that case, the coupling took place once the bicycle was formed, which increased the accessibility and the correct location of the C-terminus. Here, with only one cyclization the C-terminal was not accessible. Therefore, this strategy was discarded and a new one was designed, which is described in the next section.

Strategy C



Scheme 13: Synthetic strategy C toward the bicyclic sungsanpin analog (second approach).

The objective of the third strategy was to synthesize the bicyclic analog but this time with the C-terminal tail outside the ring (Scheme 13). To achieve the desired goal, the complete peptide chain was elongated on solid-phase in the C → N terminal direction starting from the C-terminus of Leu15 until the N-terminal Gly1. The next step was the macrolactamization to close the ring, using the methodology described in Scheme 2 and also used in strategy A (Scheme 7). Finally, the ester bond formation between Glu3 and Ser13 was performed to ensure that the last two amino acids remained outside the ring.

Unlike the first two strategies, in which the main goal was to trap the tail inside the ring, the objective of this strategy was the opposite, so the product could serve as comparison with the previous one. To reach the final product, two approaches were set up.

First approach

A first approach was designed with an allyl ether as protecting group of the Ser13 hydroxyl (not shown in the synthetic strategy). The objective was the simultaneous deprotection of Glu3 and Ser13 side chains before ester bond formation. First, the synthesis of Fmoc-L-Ser(OAll)-OH was performed and then, several conditions for the allyl ether removal were tested. The conditions are detailed in Table 5:

Assay	Reagents (eq)	T (°C)	Time	Solvent
1	Pd(PPh ₃) ₄ -PhSiH ₃ (0.1:10)	25	4 x 15 min	DCM
2	Pd(PPh ₃) ₄ -PhSiH ₃ (0.2:20)	25	1 x 16 h	DCM
3	Pd(PPh ₃) ₄ -AcOH (0.07:1.1)	25	1 x 16 h	DME, N ₂
4	Pd(PPh ₃) ₄ -pTsOH (0.07:1.1)	25	1 x 16 h	DME, N ₂
5	Pd(PPh ₃) ₄ -PhSiH ₃ (0.1:10)	40	1 x 16 h	DME, N ₂
6	Pd(PPh ₃) ₄ -AcOH (0.2:10)	25	1 x 16 h	DME, N ₂
7	Pd(PPh ₃) ₄ -PhSiH ₃ (0.2:20)	60	1 x 16 h	DME, N ₂

Table 5: Conditions for the removal of the allyl ether on solid phase.^{19,20}

Palladium(0) catalysis was selected due to its high substrate tolerance and the fact that it can be employed for the mild and selective removal of allyl esters on solid phase. After several treatments, only the allyl ester from Glu3 was cleaved, while the allyl ether remained intact. Thus, the number of equivalents was duplicated, but the same result was obtained.

Allyl ethers need a Lewis acid to activate the ether oxygen, so acid acetic (AcOH) and *p*-toluenesulfonic acid (pTsOH) were tested in dimethoxyethane (DME), which was previously described to provide more cleavage efficiency.¹⁹ However, any of these conditions led to the free hydroxyl. Therefore, we decided to couple the Ser13 without protecting group at the hydroxyl functionality (Scheme 13).

Second approach

Peptide chain elongation

As mentioned above, Ser13 was introduced without protecting group at the hydroxyl functionality while allyl was used as a protecting group of the Glu3 side chain. Allyl was selected to ensure the orthogonality with the protecting groups in the macrolactamization step (Fmoc and 2-PhⁱPr). All the amino acids used in this strategy were commercially available.

During the peptide chain elongation, LC-MS analysis revealed the presence of two peaks: one corresponding to the desired product and the other corresponding to the elongation through the free hydroxyl (Figure 26). After consecutive treatments with a piperidine-DMF solution, this second peak disappeared. Therefore, this peak might be due to the basic hydrolysis of the ester bond. Additionally, a peak with a mass of + 44 Da appeared in all the chromatograms. This peak was not a side reaction, it corresponded to the Boc decarboxylation from Trp during the cleavage. This peak disappeared after some time providing a single peak.

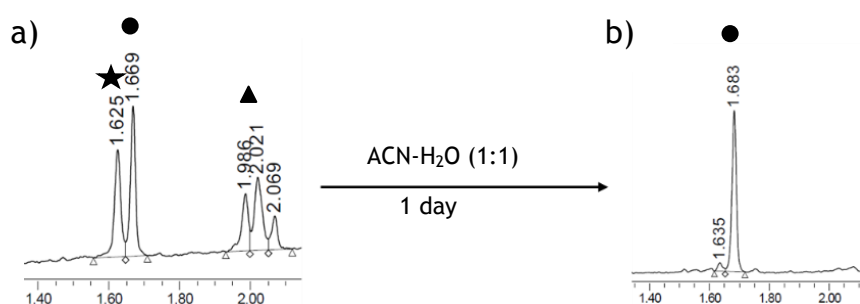


Figure 26: RP-UPLC chromatogram after Pro6 coupling. UPLC injection a) immediately after the cleavage and b) after 1 day in the UPLC vial. Conditions: Acquity BEH C18 1.7 μ m (2 mm x 50 mm) reversed-phase column; linear gradient from 30% to 100% of ACN over 2 min at 40 °C. ● = desired product, ▲ = products of ester formation, ★ = decarboxylation of Boc from Trp (+ 44 Da).

Macrolactamization

This step was performed using the same conditions than the macrolactamization described in Scheme 2.

Ester bond formation

The ester bond was formed on solid phase via intramolecular reaction. Hence, two parameters could only be modified: the carbodiimide and the solvent. All the conditions tested are detailed in Table 6:

Assay	Carbodiimide	Catalyst	T (°C)	Time (h)	Dry solvent
1	DIPCDI (4 eq)	DMAP (0.8 eq)	25	2 + 3	DCM
2	DIPCDI (4 eq)	DMAP (0.8 eq)	40	16	DCE
3	DIPCDI (4 eq)	DMAP (0.8 eq)	25	2 + 16	ACN
4	DIPCDI (8 eq)	DMAP (0.8 eq)	25	2 + 3	ACN

Table 6: Conditions for the formation of the ester bond on solid-phase.

In the first assay, using the standard conditions, the reaction did not work, and only starting material was detected. Therefore, the peptidyl-resin was heated up to 40 °C with dichloroethane (DCE), which has very similar properties to DCM, but it is less volatile. However, after the cleavage of a small aliquot of resin, no peptide was observed in the crude.

The complexity of this step was based on the high steric hindrance present in the resin due to the peptide chain elongation, meaning that a proper disposition of the two reactive groups was very important to ensure the ester linkage. Regarding this, a good alternative solvent was ACN.¹⁷ In assay 3, after the two consecutive treatments with DIPCDI-DMAP, 50% of starting material remained unreacted. Therefore, more equivalents and less time were the appropriate conditions for the ester formation (assay 4). With ACN, the ester bond formation worked smoothly.

Specific considerations

Efficient couplings were obtained in all cases, only Lys5 was re-coupled. Unprotected Ser13 was coupled with HATU-HOAt-DIEA to ensure a fast and efficient coupling.

In this strategy, the synthesis of a bicyclic sungsanpin analog with the unthreaded tail was achieved. To assure that the tail remained outside from the ring, all the synthesis was performed on solid phase. Moreover, the macrolactamization was

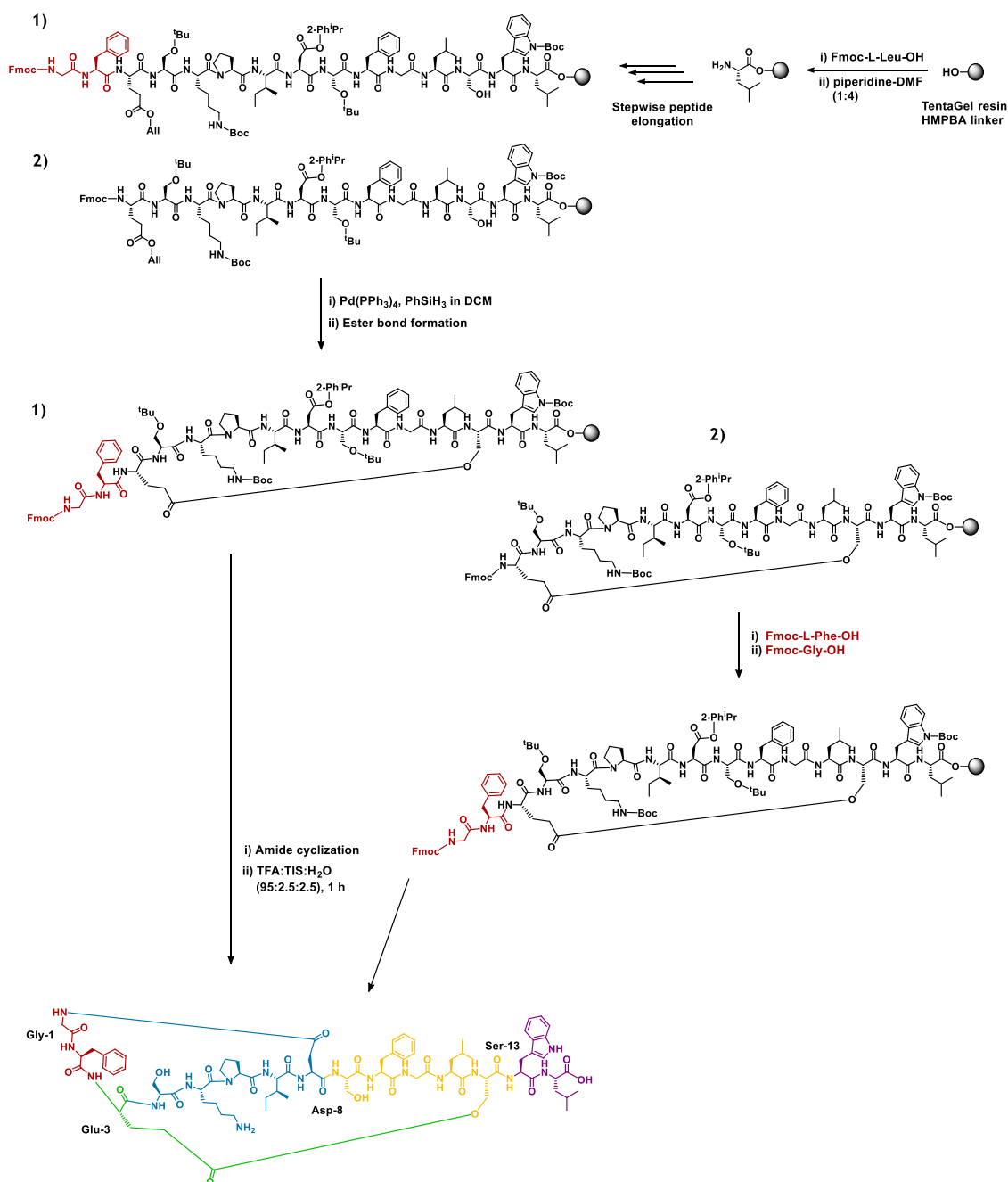
performed before the ester bond because once the ring amide was closed, the tail could not be trapped. After this result, a fourth strategy was set up in which the ester bond was formed before the amide cyclization.

Strategy D

This fourth strategy represented a mix of all the previous strategies. All the peptide chain elongation and further cyclizations were performed on solid phase. As previously mentioned, the objective of this strategy was the same as in strategy C (Scheme 13), but inverting the order of the key reactions: first, the formation of the ester bond and then, the macrolactamization.

In comparison with the strategy C, in which only two approaches were required, several synthetic approaches were needed here to reach the desired product. However, all the approaches had in common that they maintained the same order for the key reactions.

First approach



Scheme 14: First synthetic approach to the strategy D toward the bicyclic sungsanpin analog via two different routes (1 and 2).

Two different routes have been designed for the first approach to strategy D. In the first route, the peptide chain was elongated from the C-terminus of Leu15 until Gly1, whereas in the second route, the peptide chain was only elongated until Glu3 (Scheme 14). The two different lengths of the peptide chain were designed in order to evaluate its influence on the steric hindrance during the ester formation.

Ester bond formation

The ester bond was performed on solid phase via an intramolecular reaction. The conditions used for the ester formation in route 1 are detailed in Table 7. On the other hand, the conditions used in route 2 are detailed in Table 8.

Assay	Coupling agents	Catalyst	T (°C)	Time (h)	Dry solvent
1	DIPCDI (8 eq)	DMAP (0.8 eq)	25	2.5 + 3	ACN
2	DIPCDI (10 eq)	DMAP (0.5 eq)	25	2 + 3	DCM
3	DIPCDI (10 eq)	DMAP (0.5 eq)	45	48	DCM

Table 7: Conditions for the formation of the ester bond in route 1.

The initial conditions tested for the formation of the ester bond in this strategy were the same that achieved the desired results in strategy C (assay 1, Table 6). However, no desired product was this time detected, and only starting material was recovered. In the next two assays, the solvent and the number of equivalents were modified, but no conversion was either detected. This result might be attributed to the steric hindrance caused by the bulky *N*-terminal Fmoc group and by the long distance (11-amino acid residues) between the two reactive groups.

For this reason, in route 2, the peptide was only elongated until Glu3, in order to avoid the steric hindrance of the *N*-terminal amino acids.

Assay	Coupling agents (eq)	Active ester	Catalyst (eq)	pH	T (°C)	Time (h)	Dry solvent
1	DIPCDI (10)	--	DMAP (0.5)	--	25	2 x 3	DCM
2	DIPCDI (20)	--	DMAP (0.8)	--	45	4	DCM
3	DIPCDI (10)	--	DMAP (0.1)	--	0	3 + 48	DCM
4	DIPCDI (10)	--	DMAP (0.5)	--	0 → 25	2 x 3	DCM
5	DIPCDI (10) in dry DCM	(*)	DMAP (0.5)	--	25	2 x 3	DMF
6	DIPCDI (10) in dry DCM	(*)	DMAP (0.5)	--	45	3 + 6 + 16	DMF
7	DIPCDI- OxymaPure (5:5) in dry DMF	(*)	DMAP (0.5)	8	25	2 x 3	DMF
8	--	HOSu- DIPCDI-DMAP (10:10:0.5) in dry DCM 2 h + 3 h	DMAP (0.5)	8	25	1 + 16	DCM
9		PPh ₃ -DIAD (5:5)		--	25	3 + 16	Toluene

Table 8: Conditions for the formation of the ester bond in route 2. (*) The resin was washed after the addition of DIPCDI.

In the first four assays (Table 8), the same peak was observed with a mass of + 146 Da respect to the expected mass (Figure 27a). The reaction temperature was cooled to 0 °C to minimize or avoid this side reaction, but the desired reaction did not take place. It was not possible to identify this side product, and we hypothesized that it could be due to the excess of DIPCDI.

For this reason, in assays 5 and 6, after DIPCDI addition, the resin was washed with DMF and DMAP was then added. The reaction was monitored by LC-MS and the expected mass was detected. However, the peak could not be isolated due the high number of side products. When the same reaction was heated up, more side products appeared with no desired product being detected.

The next assays were based on the formation of an active ester from the carboxylic acid by adding either OxymaPure or HOSu. Regarding HOSu, the OSu ester was detected (Figure 27b) and DIEA was added to the resin to increase the pH and to facilitate the nucleophilic reaction. Nevertheless, after LC-MS analysis, none of the

detected peaks contained the expected mass. With addition of OxymaPure no improvement was achieved with respect to assays 5 and 6.

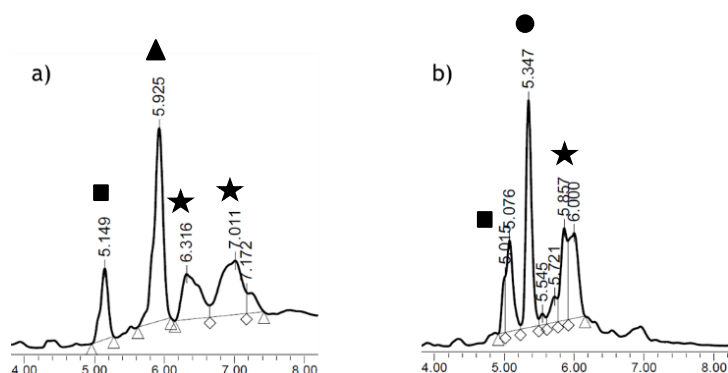


Figure 27: RP-HPLC chromatogram: a) after the formation of the ester of assays 1-4 and b) after the succinimide ester formation in assay 8. Conditions: Xbridge™ C18 2.5 μ m (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. ■ = starting material, ▲ = side product with $m/z + 146$ Da, ● = desired product (OSu ester), ★ = unknown side products.

The last condition tested (assay 9) was the Mitsunobu reaction,²¹ a well-known reaction for esterification. It was carried out with diisopropyl azodicarboxylate (DIAD) and triphenylphosphine (PPh₃) in toluene. After a total of 18 h, only the intact peak of the starting material was observed.

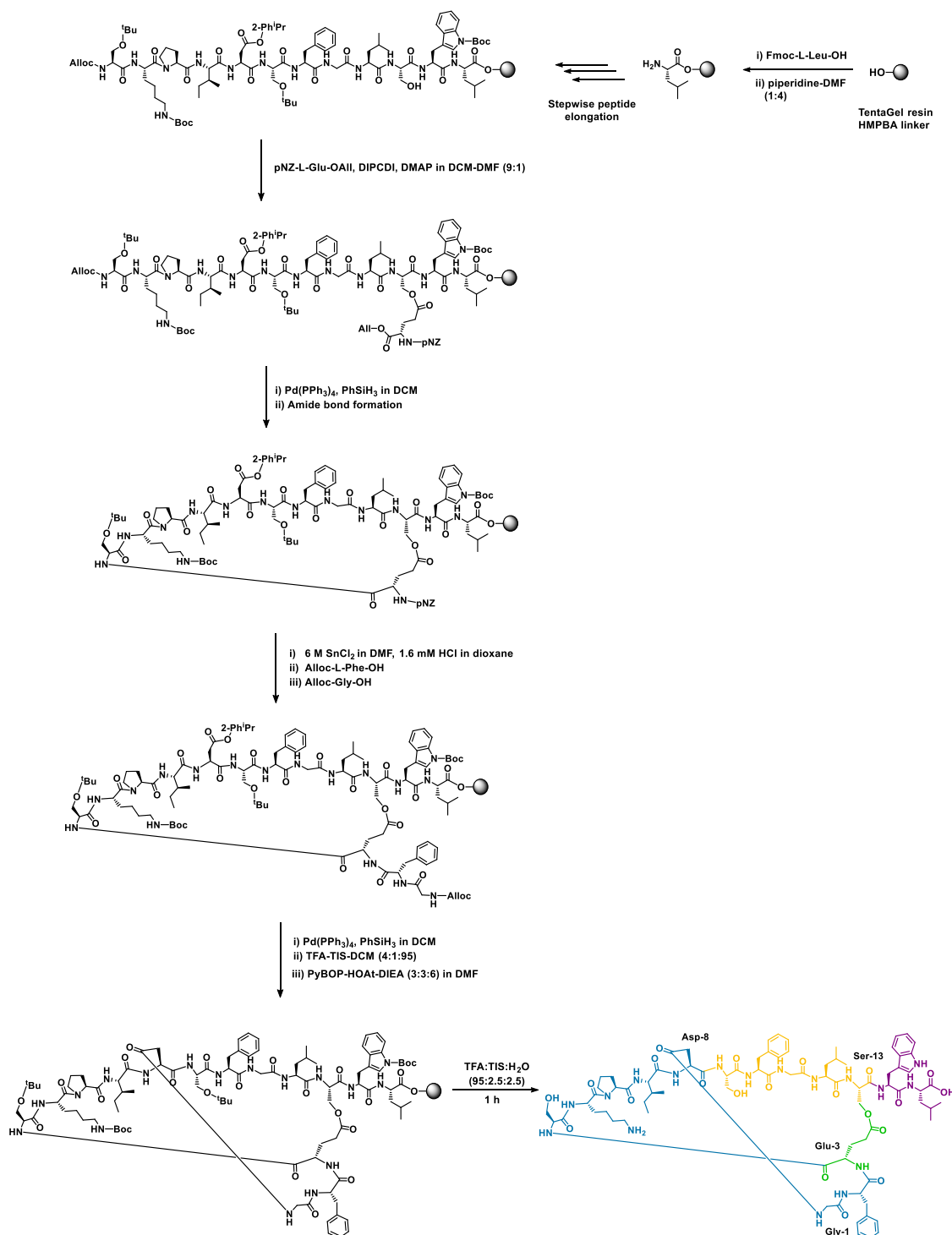
Specific considerations

The HMPBA linker was attached to the resin with DIPCDI-OxymaPure and then the first amino acid was introduced through an ester bond via *O*-acylisourea activation with satisfactory loadings (0.20 mmol/g). Both peptide chains were elongated having as reference the conditions and the methodology used in the second approach to the strategy C. Ser13 was introduced without protecting group on its side chain, like in strategy C. In this case, HATU-HOAt-DIEA were used to ensure a fast and efficient coupling.

It was observed that when the peptide chain was elongated, the formation of the ester bond was quite difficult, even making the peptide chain shorter. Thus, after the disappointing results, more experimentation was required. For this reason, a second approach was designed. The objective of the new approach was the same as before, the formation of the ester bond before the macrolactamization. The ester bond was formed directly through the side-chain of Glu3, after its addition to the

resin. Then, the amino acid was coupled to the peptide chain. The amide bond should be easier to form than the ester bond, making this step more feasible.

Second approach



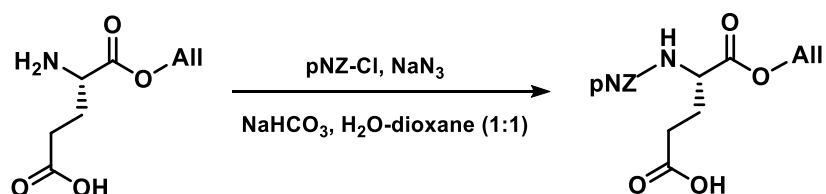
Scheme 15: Second synthetic approach to the strategy D toward the bicyclic sunsanpin analog.

A new synthetic design was set up, beginning with the ester bond formation through the introduction of Glu3 onto the resin, in order to avoid the steric hindrance of the peptide chain (Scheme 15). Thus, the peptide chain was elongated from the C-terminus of Leu15 until Ser4, and the ester bond formation was then accomplished through the side chain of Glu3. Afterwards, Glu3 was coupled to the rest of the peptide chain via formation of an amide bond through its α -carboxylic acid. Finally, the last two N-terminal amino acids were introduced.

The functionalities used for the protection of Glu3 were carefully selected to avoid the use of basic treatment for their removal: Allyl and pNZ became the best choice. Moreover, the last two amino acids were introduced in one step as an Alloc-dipeptide (Scheme 8), with the objective of preventing the hydrolysis of the ester bond.

Synthesis of pNZ-L-Glu-OAll

pNZ-L-Glu-OAll was prepared from the commercially available H-L-Glu-OAll using the previously described azide methodology in Scheme 8 for the pNZ protecting group introduction. First, the pNZ-Cl was mixed with sodium azide, and then, aa was added to the mixture, obtaining the protected aa in excellent yields.



Scheme 16: Synthetic strategy to the Fmoc-L-Lys(pNZ)-OH residue.

Ester formation

The formation of the ester bond worked smoothly using the conditions described in Table 9. A high excess of pNZ-L-Glu-OAll, previously activated with DIPCDI, and the subsequent addition of DMAP were used to obtain the desired product. The reaction was heated up to 40 °C to observe a complete and faster conversion than when carried out at room temperature (Figure 28).

Assay	Coupling agents	Catalyst	T (°C)	Time (h)	Dry solvent
1	DIPCDI (8 eq)	DMAP (0.5 eq)	25	2 + 3	DCM
2	DIPCDI (8 eq)	DMAP (0.5 eq)	40	3	DCM

Table 9: Conditions for the formation of the ester bond.

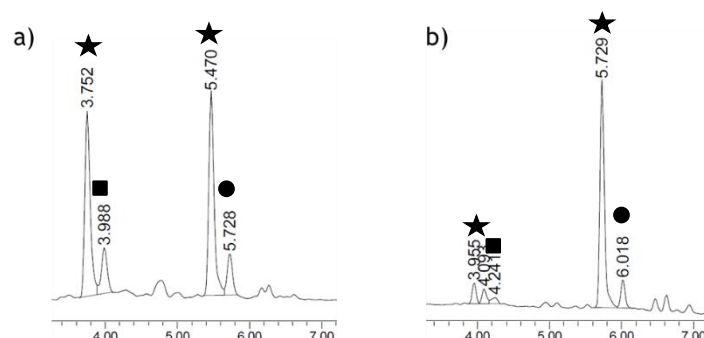


Figure 28: RP-HPLC chromatogram after the formation of the ester bond. a) Assay 1, only 50% of conversion was observed; and b) assay 2, with 90% of conversion. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. ★ = decarboxylation of Boc from Trp (+ 44 Da), ■ = starting material, ● = desired product.

The same peak as in Figure 26, with a mass of + 44 Da, was detected. This peak corresponds to a normal species observed after the cleavage of peptides containing Trp.

Amide bond formation

After the removal of allyl and Alloc groups, the amide bond between Glu3 and Ser4 was formed using two different coupling systems, one in basic and the other in neutral conditions (Table 10). In both cases, the same side product was obtained as a main peak with a mass of + 146 Da (Figure 29), which corresponded to the same unknown side product obtained in the first approach (Figure 27a).

Assay	Coupling system	pH	T (°C)	Time	Solvent
1	PyBOP-HOAt-DIEA (3:3:6)	7	25	30 min	DMF
2	DIPCDI-OxymaPure (3:3)	6 - 7	25	1 h	DMF

Table 10: Conditions for the formation of the amide bond between Glu3 and Ser4.

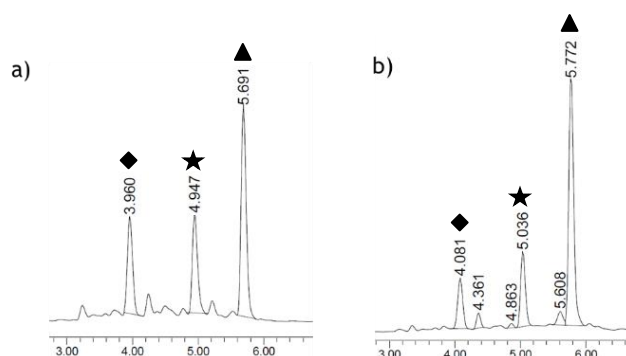


Figure 29: RP-HPLC chromatogram after the amide bond formation. a) With PyBOP-HOAt-DIEA at pH 7 and b) with DIPCDI-OxymaPure. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. \blacklozenge = ester bond hydrolysis, \star = unknown side product (+ 44 Da), \blacktriangle = unknown side product (+ 146 Da).

Peptide chain elongation

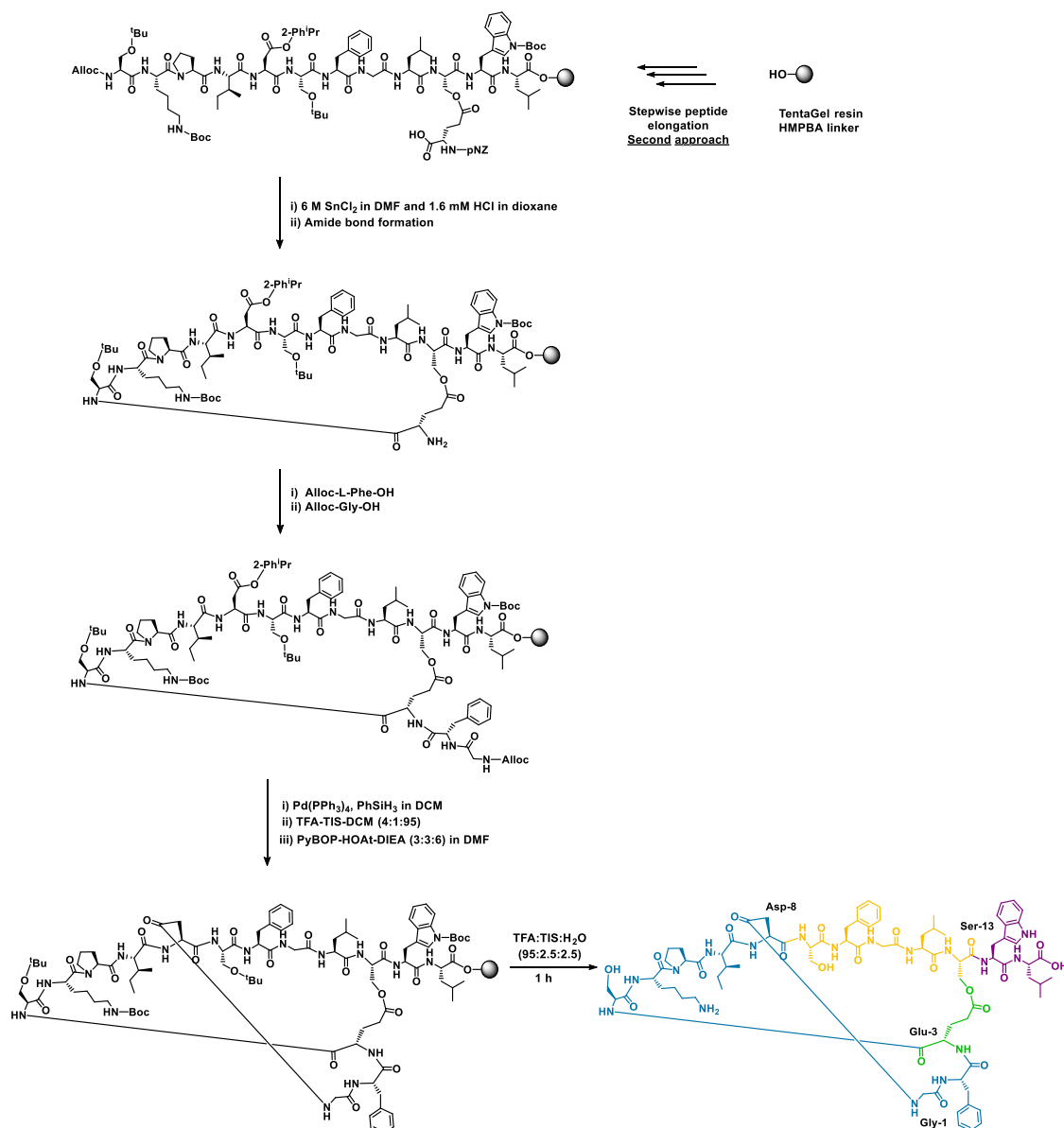
Despite the formation of the + 146 Da impurity, the synthesis was continued with the introduction of the last two amino acids. At this point, the two amino acids were coupled efficiently, but always with the gain of + 146 Da. We were unable to identify this side reaction, thus more experimentation would be required.

Specific considerations

The same conditions than in the first approach were used for the loading of the first amino acid and for the peptide couplings.

The first peptide chain elongation (until Glu3) was scaled up to have enough peptidyl-resin for the next approaches. Moreover, once the conditions for the ester bond formation had been optimized, the synthesis of the ester-containing peptide was scaled up. At this point no additional problems were detected, and the reaction worked smoothly. This peptidyl-resin was used in the next new approaches in which several modifications were attempted with the objective to optimize the amide bond formation and the final steps of the synthesis. However, formation of the same side product as in the first approach (see Figure 27) took place, which suggested that this impurity was independent of the synthetic step. Thus, in the next approaches, some modifications in the reaction order were performed in order to get the best synthetic route.

Third approach



Scheme 17: Third synthetic approach to the strategy D toward the bicyclic sungsanpin analog.

In this third approach to strategy D (Scheme 17), the order of the reactions after the formation of the ester bond was inverted with respect to the second approach. In consequence, the pNZ group from the amino functionality of Glu3 was removed before the formation of the amide bond. These modifications were based on the hypothesis that the pNZ could interfere during the formation of the amide bond, yielding the + 146 Da side product.

In principle, the free amino group from Glu3 should not be a problem during the formation of the amide bond, as it cannot compete with the *N*-terminal amino group of Ser4, since the cyclization over the same residue is not possible. In addition, the intermolecular reaction is not feasible on solid phase at low loading concentration.

Amide bond formation

In this step, the removal of the protecting groups (allyl, Alloc and pNZ) proceeded smoothly. Afterwards, the formation of the amide bond took place. The same coupling systems as in the second approach were tested (Table 10), but using more equivalents and extended reaction times. The detailed conditions are described in Table 11:

Assay	Coupling system	pH	T (°C)	Time (h)	Solvent
1	PyBOP-HOAt-DIEA (3:3:6)	7 - 8	25	2 x 1	DMF
2	DIPCDI-OxymaPure (5:5)	7	25	1.5	DMF

Table 11: Conditions for the formation of the amide bond between Glu3 and Ser4.

However, no desired product was obtained with none of these two conditions. The peak of the starting material was mainly observed (Figure 30), and the use of basic conditions (assay 1) resulted in more unknown side products, compared to assay 2.

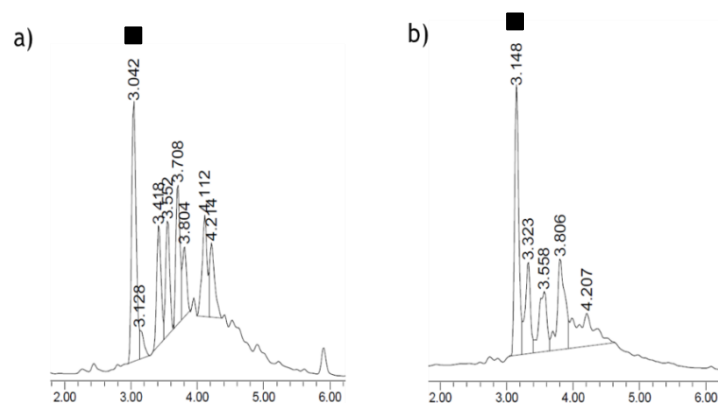
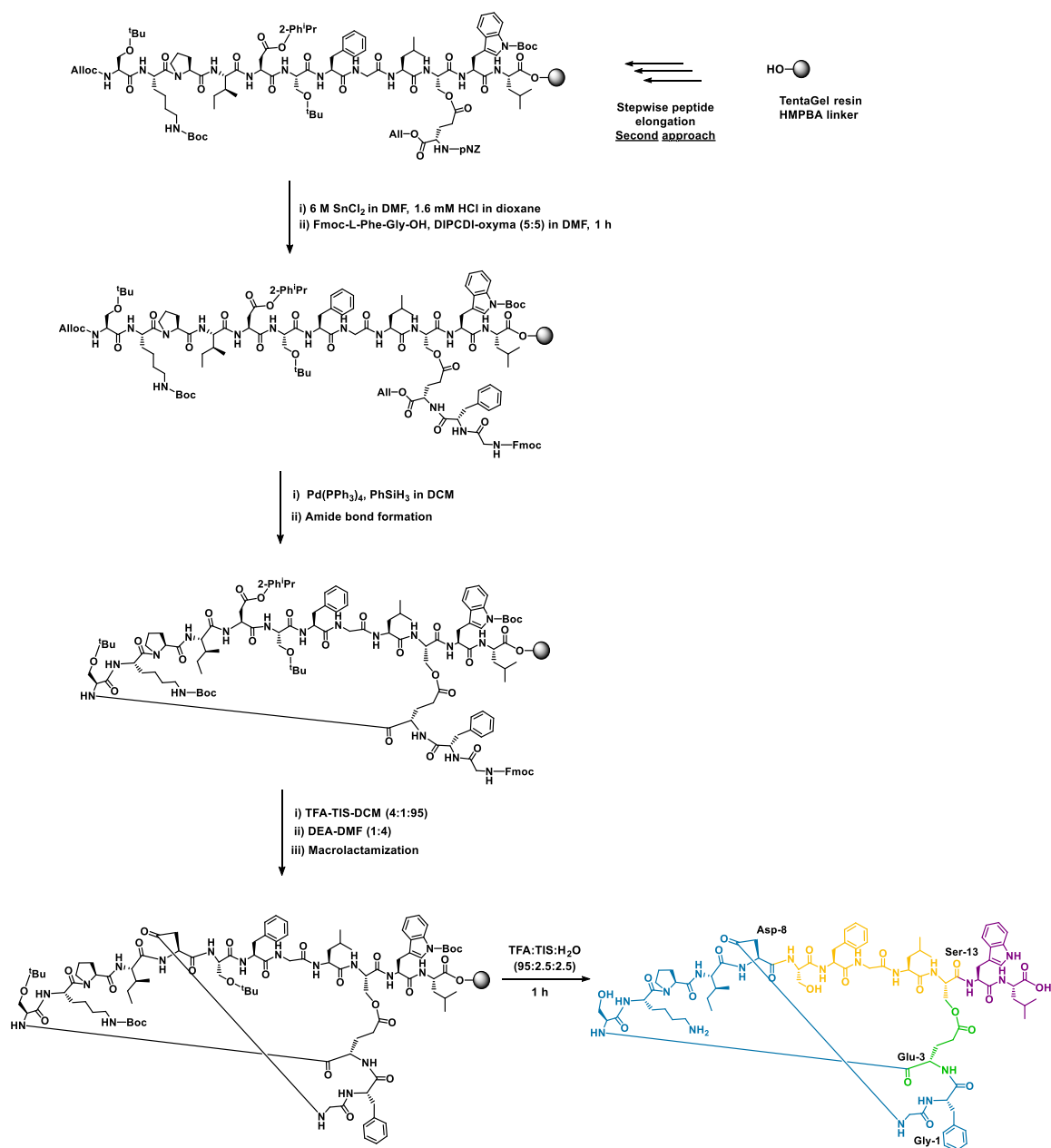


Figure 30: RP-HPLC chromatogram after the formation of the amide bond. a) With PyBOP-HOAt-DIEA at pH 7 - 8 (assay 1) and b) with DIPCDI-OxymaPure (assay 2). Conditions: Xbridge™ C18 2.5 μ m (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. ■ = starting material. The other peaks are unknown side-reactions with higher masses.

In conclusion, the removal of the pNZ before the formation of the amide bond did not result in any improvement. This fact meant that the steric hindrance factor was discarded. Satisfyingly, the + 146 Da side product did not appear, compared to the second approach (Figure 29). However, no reaction progress was detected, thus indicating that the problem arose at the amide bond formation step.

Fourth approach

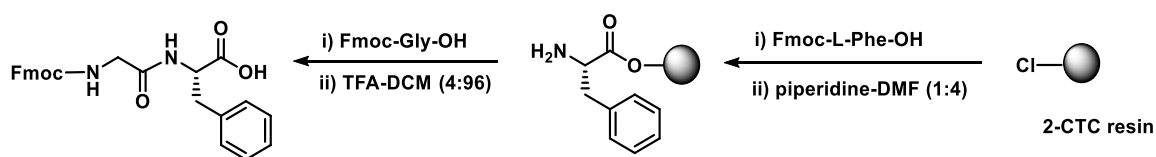


Scheme 18: Fourth synthetic approach to the strategy D toward the bicyclic sungsanpin analog.

The main objective of this approach was to make possible the formation of the amide bond. The last combination of reactions that remained to be tested was to couple the last two amino acids before the formation of the amide bond. Regarding this, two main modifications were implemented. The first was that Gly1 and Phe2 were directly coupled as an Fmoc-protected dipeptide, instead of as Alloc-protected dipeptide, due to the orthogonality requirement. The coupling of this dipeptide allowed the use of the Fmoc deprotection solution only once, in order to prevent the hydrolysis of the ester bond. The second introduced modification was that the amide bond formation was performed after the elongation of all the entire peptide chain (Scheme 18).

Synthesis of Fmoc-Phe-Gly-OH

The dipeptide was synthesized on solid phase (Scheme 19) providing an easier and faster access than in solution. Fmoc-L-Phe-OH was loaded onto the 2-CTC resin, followed by Fmoc deprotection and the introduction of Fmoc-Gly-OH. Finally, the dipeptide was cleaved from the resin under mild acidic conditions.



Scheme 19: Synthetic strategy to the Fmoc-Phe-Gly-OH dipeptide.

Coupling of the dipeptide

The neutral coupling system DIPCDI-OxymaPure was used to perform the dipeptide coupling. A re-coupling was required to ensure its complete introduction. The reaction worked smoothly.

Amide bond formation

Several conditions were tested to carry out the amide bond formation. The conditions are detailed in Table 12.

In neutral conditions (assay 1) the reaction was very slow; after one treatment the conversion was only 50%. After a second treatment, two side products were formed: one corresponding to a mass gain of 40 Da, and the other with a mass of + 143 Da, corresponding to the *N*-acylisourea formation with DIPCDI. Both mass gains were with respect to the starting material (Figure 31).

On the other hand, using basic conditions (assays 2 and 3) no desired product was obtained. Besides the starting material, the side product peak with a mass of + 40 Da was detected.

Assay	Coupling system	pH	T (°C)	Time	Solvent
1	DIPCDI-OxymaPure (5:5)	7	25	2 x 1 h	DMF
2	PyBOP-HOAt-DIEA (3:3:6)	7 - 8	25	2 x 30 min	DMF
3	PyBOP-HOAt-DIEA (3:3:6)	7 - 8	25	2 x 30 min	DCM-DMF (9:1)
4	DIPCDI-HOAt (5:5)	7	25	1 h	ACN

Table 12: Conditions for the formation of the amide bond between Glu3 and Ser4.

The ester-containing peptide, previously synthesized in the second approach (Figure 28), had been stored for a long time at 4 °C. Therefore, the resin was conditioned with a series of dry solvents before its use (DCM, Et₂O, DCM, DMF and DCM, 5 x 30 s each of them). Afterwards, conditions used in assay 1 were again tested and this time the desired product was detected (Figure 31c). The reaction was not totally clean, as some side products appeared—such as the hydrolysis of the ester bond—but in any case the expected product was formed.

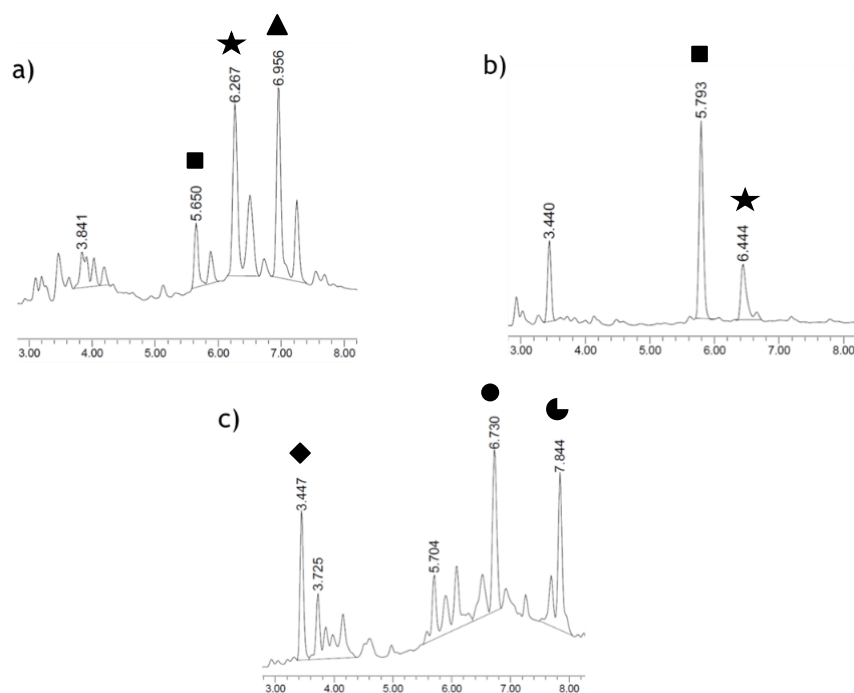


Figure 31: RP-HPLC chromatogram after the formation of the amide bond. a) Assay 1, after two treatments of coupling reagents; b) assay 2, after one treatment of coupling reagents; and c) after conditioning of the resin and using the conditions from assay 1. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 $^{\circ}\text{C}$. ■ = starting material, ★ = unknown side-product (+ 40 Da), ▲ = N-acylisourea (+ 143 Da), ◆ = ester bond hydrolysis, ● = desired product, ◐ = uncharacterized side-product.

Macrolactamization

To carry out the macrolactamization between the amino group of Gly1 and the side chain of Asp8, the same experimental procedure as in Scheme 2 was followed. Two modifications were implemented to avoid ester bond hydrolysis: first, the use of DEA in DMF (with 0.1 M OxymaPure) to remove the Fmoc group and second, the use of a neutral coupling system to perform the cyclization.

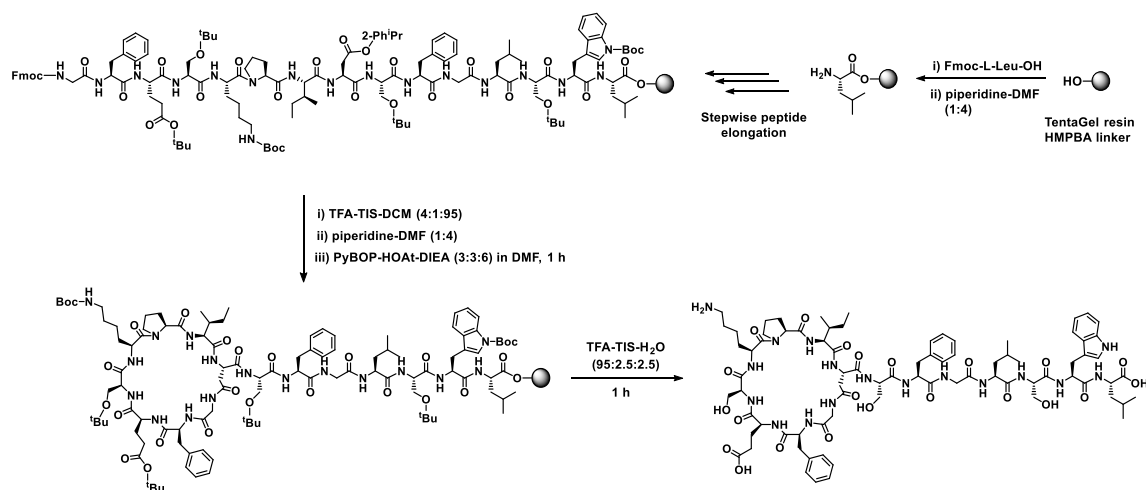
Specific considerations

It has been observed that after the removal of allyl and Alloc, better results were obtained when the synthesis was performed in small scale. For this reason, the last three steps were performed with small batches of resin. After the cleavage, the crudes were combined and purified together.

With this fourth approach we were able to obtain the third sungsanpin bicyclic analog via ester bond. By varying the order of the reactions, we could reach the total synthesis of the peptide. The introduction first of the *N*-terminal dipeptide and the suitable conditioning of the resin were the key steps for the successful synthesis.

Synthesis of the branched cyclic analog

So far, four different synthetic strategies were used for the synthesis of the bicyclic analogs, from which a total of three bicyclic analogs were obtained. Moreover, the natural sungsanpin was also synthesized on solid phase. In order to compare the bicycles with their branched cyclic analogs, it was necessary to perform the synthesis of the branched cyclic peptide with Glu at position 3 (Scheme 20).



Scheme 20: Synthetic strategy toward the branched cyclic analog.

Specific considerations

For the synthesis of this peptide, the same synthetic approach and experimental procedure than for the natural sungsanpin was followed (Scheme 2).

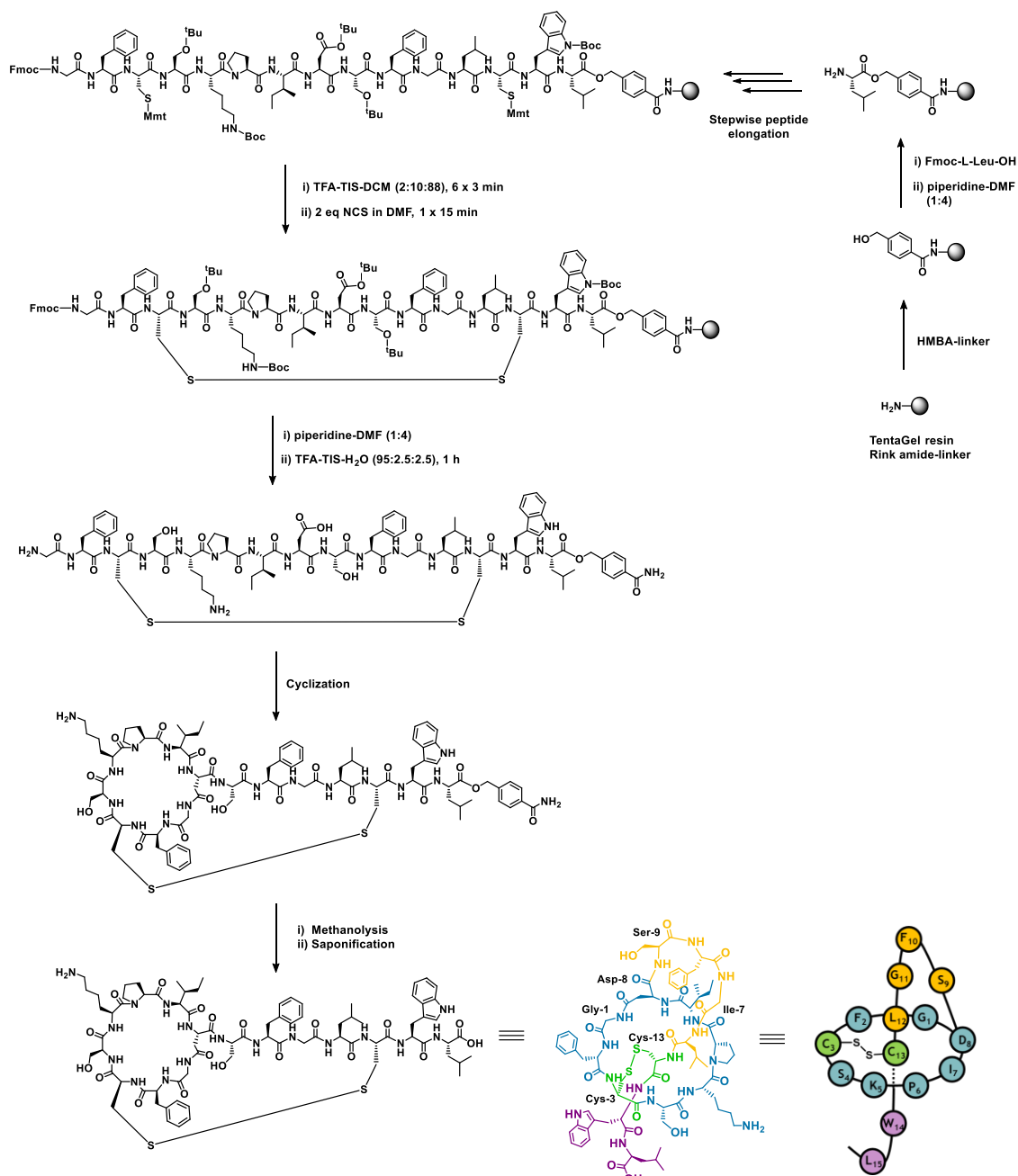
The peptide chain elongation was performed using automated synthesis, except for the introduction of the first three amino acids and the linker, which were coupled manually to avoid the DKP formation. Moreover, the removal of the *N*-terminal Fmoc group, the macrolactamization reaction and the cleavage steps were also performed manually. The synthesis worked smoothly and produced high yields of the branched cyclic peptide.

Synthesis of the bicyclic analogs via disulfide bond

Strategy 1: Cys3 - Cys13

In this section, a new strategy for the synthesis of bicyclic analogs was designed, in this case via disulfide bond. Based on the same principles used for the formation of the ester bond, the idea was that the disulfide bond could be reduced under mild conditions after the total synthesis, and provide that lasso structure. Thus, the objective was to position the disulfide bond in the same place as the previous ester bond. In consequence, two modifications in the peptide sequence were required: the replacement of Gly3 by Cys3 and Ser13 by Cys13 (Scheme 21).

However, a main difference was implemented in this strategy, compared with the previous four strategies used for the bicycle synthesis. The idea was to perform the peptide chain elongation and the disulfide bond on solid phase and afterwards the macrolactamization step in solution. The cyclization in solution could provide a more flexible structure to the cyclic peptide backbone, helping to trap the tail inside the ring if the suitable conformation was adopted. A main limitation was that the C-terminus should be protected with an orthogonal protecting group during the cleavage. For this, a small protecting group would be the best option to avoid steric hindrance during the cyclization step.



Scheme 21: Design of the synthetic strategy toward the bicyclic sungsanpin analog via disulfide bond between Cys3 - Cys13.

Specific considerations

The HMBA linker was chosen because it could provide an ester as C-terminal protecting group, which is resistant to the acidic conditions used for the cleavage. The first idea was to anchor directly the HMBA linker to the polymeric support, which allowed direct access to the methyl ester functionality via base-mediated methanolysis. However, methanolysis conditions, such as 0.2 M NaOMe, 0.1 M NaOH or 20% DIEA in MeOH-DMF (1:1), did not work. At high temperatures or even at room

temperature, racemization occurred and at 0 °C, the reaction did not progress. This lack of reactivity was due to the poor swelling of the polymeric support in MeOH, even with 50% in DMF. The PEG content of the TentaGel resin was not enough to bear this type of treatments.

This handicap meant that the re-introduction of an acid cleavable handle was needed to enable the release from the resin of the HMBA-peptide using standard TFA cocktails. Afterwards, the peptide-HMBA conjugate could undergo methanolysis in solution. The disadvantage of this second approach was the increase of the steric hindrance at the C-terminus, making more troublesome the introduction of the tail inside the ring.

The peptide chain elongation, the disulfide formation and the cleavage steps worked smoothly, but some problems occurred during the macrolactamization reaction in solution, where two products with the same expected mass were obtained (Figure 32). As explanation, one peak could belong to the desired cyclization, through Gly1-Asp8, while the other could be a cyclic side product through Lys5-Asp8.

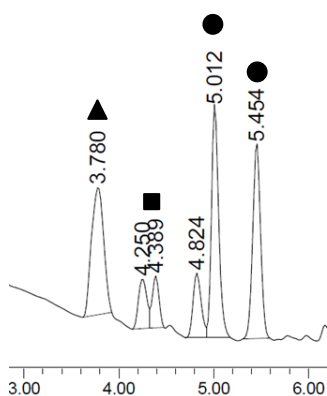


Figure 32: RP-HPLC chromatogram after the macrolactamization in solution. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. ▲ = coupling agents, ■ = starting material, ● = desired product.

Despite the fact that no successful results were obtained, this design was conceived as an alternative strategy to the synthesis of the bicyclic analogs. Much more experimentation would be required, like the protection of the Lys5 side chain and the optimization of the macrolactamization conditions. Added to this, the final saponification of the peptide-HMBA would also require an optimization.

Strategy 2: Cys3 - Cys15

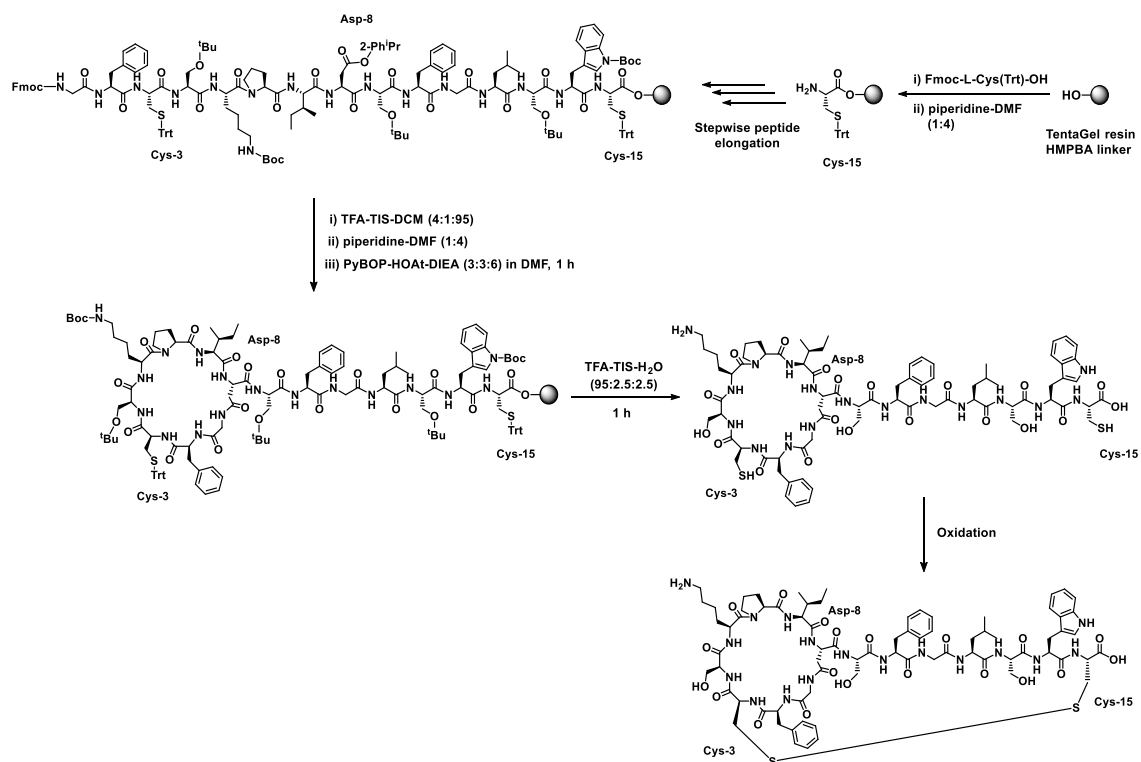
The lasso topology provides the high stability and activity to lasso peptides. For this reason, in the previous strategies the main aim was to tie the loop to the ring, trying to imitate the lasso topology. In strategy 1, a disulfide bond in the same position as the ester bond was tested. The idea with this modification was on one hand, the biological comparison between the ester-bond and disulfide-bond bicycles and, on the other hand, the structural comparison based on the different synthetic pathways.

It was previously reported that the recognition domain responsible for the high activity depends on the peptide loop.²² Regarding this, the idea of the analog presented below was to link the C-terminal tail with a Cys side chain, creating a bicyclic peptide, with no linear tail. Theoretically, this linear tail did not play an important role in the biological activity, being labile to degradation. For this reason, the linear tail was cyclized through a disulfide bond between Cys3 and Cys15.

Compared to strategy 1, here all the peptide chain was elongated on solid phase and only the oxidation was carried out in solution. The difference was that, in this case, the main objective was not to trap the tail inside the ring, as in the previous strategy. Here, the main goal was to cyclize the entire tail.

Comparing to the natural sungsanpin (Figure 15), two modifications in the peptide sequence were required to synthesize this analog: the replacement of Gly3 and Leu15 by cysteines to form a disulfide bond, maintaining the free C-terminus carboxylic acid.

The peptide chain was elongated in the $C \rightarrow N$ terminal direction starting from the C-terminus of Cys15 until the N-terminal Gly1. Trt was used as protecting group of the Cys thiol function, and was removed simultaneously during the cleavage together with the other protecting groups (Scheme 22).



Scheme 22: Synthetic strategy towards the bicyclic analog via disulfide bond between Cys3 and Cys15.

Macrolactamization

This step was performed after the complete elongation of the peptide chain using the same conditions as in the macrolactamization step (Scheme 2). The reaction worked smoothly.

Disulfide bond formation in solution

The disulfide was formed in solution under oxidizing conditions (5% DMSO-H₂O and a saturated solution of Na₂CO₃). The reaction was performed using high dilution conditions (0.6 mM) to favor the intramolecular cyclization.

After the disulfide bond formation, a small aliquot of resin was cleaved under acidic conditions and the crude was analyzed by LC-MS (Figure 33). Moreover, the Ellman test was performed to ensure complete disulfide bond formation.

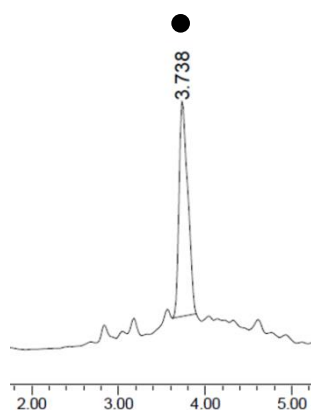


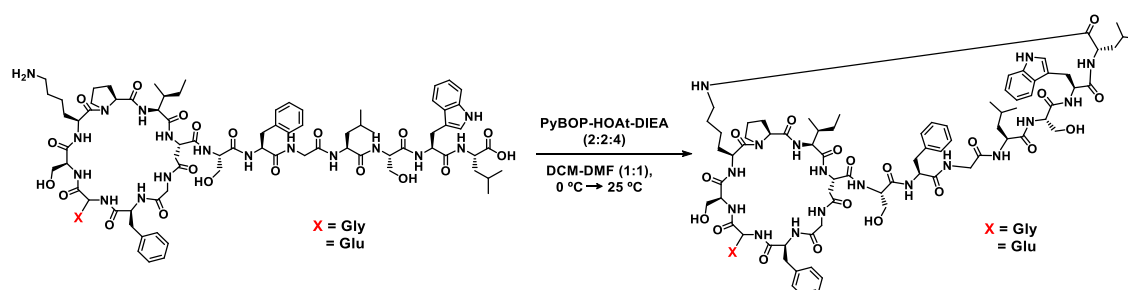
Figure 33: RP-HPLC chromatogram after the formation of the disulfide bond. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. ● = desired product.

Other synthetic bicyclic analogs

Bicyclic analogs side-chain-to-tail via amide bond

Continuing with the idea of the last strategy, in which the tail was linked to the ring, providing a bicycle with no linear part. Here, a new strategy was designed, in which the C-terminal carboxylic acid is linked with a side chain of one of the residues located in the ring.

The C-terminus was cyclized via an amide bond with the amino group of the Lys5 side-chain. With this modification, both peptide charges were eliminated, leading to a bicyclic peptide with no charge (Scheme 23).



Scheme 23: Synthetic strategy for the second cyclization of the synthetic sungsanpin and the branched cyclic analog.

The branched cyclic peptides obtained from Scheme 2 and 20 were subjected to a second cyclization. The standard coupling system used for the macrolactamization

produced satisfactory results (Table 13). The reaction was cooled to 0 °C to avoid racemization of the C-terminus. The cyclization with the synthetic sungsanpin was faster and cleaner than with the branched cyclic analog. For the latter, a wide peak with the expected mass and other unknown peaks were detected (Figure 34). The wide peak could be attributed to the side reaction between the side chains of Glu3 and Lys5.

The low product yield and the appearance of side products for the branched cyclic analog could be explained by the structure adopted in solution. In this sense, Glu3 and Lys5 may be involved in a electrostatic interaction, thus hindering the nucleophilic attack of the amine and allowing the formation of side products.

Assay	Coupling system	Concentration (mM)	pH	T (°C)	Time	Solvent
1	PyBOP-HOAt-DIEA (2:2:4)	0.4	7	0 → 25	1 h	DCM-DMF (1:1)

Table 13: Conditions for the macrolactamization.

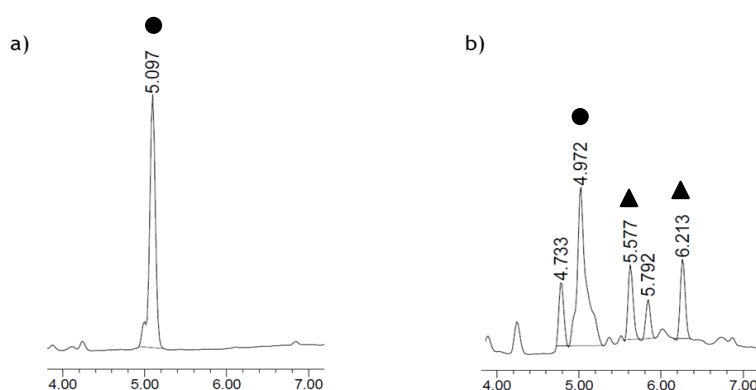


Figure 34: RP-HPLC chromatogram after the second cyclization a) synthetic sungsanpin b) branched cyclic analog. Conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 80% of ACN over 8 min at 25 °C. ● = desired product, ▲ = uncharacterized side products.

We were able to obtain and purify the bicyclic peptide from the synthetic sungsanpin. More experimentation was required for the one from the branched cyclic analog, in order to improve the reaction yields.

Discussion

With all the efforts dedicated to the different designs and synthetic approaches, we were able to synthesize the first sungsanpin bicyclic analogs via ester bond. From the four synthetic strategies that we designed, we were able to obtain three final products. In strategy A, a combination of solid-phase and in-solution reactions successfully led to the final product. On the other hand, the bicycles obtained from strategies C and D were entirely synthesized on solid phase. An additional strategy with a disulfide bond, instead of an ester bond, was set up, but the pure final product could not be achieved in this case. More experimentation will be required to optimize this last strategy.

Two alternative bicyclic peptides were synthesized, in one case keeping the sungsanpin sequence, and in the other introducing two modifications. In one case, the C-terminal carboxylic acid was linked to the ring through the side-chain of Lys5, leading to a side-chain-to-tail bicycle with no free tail and in the other case, the entire tail was cyclized via a disulfide bond, leading only the C-terminus free.

Finally, a simple and efficient solid phase strategy was envisaged for the synthesis of sungsanpin and the branched cyclic analog, which allowed the synthesis of both peptides in good yields. Their characterization will be analyzed in Chapter 2.

References

- (1) Um, S.; Kim, Y.-J. J.; Kwon, H. H. C.; Wen, H.; Kim, S.-H. H.; Kwon, H. H. C.; Park, S.; Shin, J.; Oh, D.-C. C. *J. Nat. Prod.* **2013**, *76* (5), 873.
- (2) Duncan, K. R.; Haltli, B.; Gill, K. A.; Correa, H.; Berru , F.; Kerr, R. G. *J. Ind. Microbiol. Biotechnol.* **2015**, *42* (1), 57.
- (3) Garc a-Mart n, F.; Bay -Puxan, N.; Cruz, L. J. J.; Bohling, J. C. C.; Albericio, F. *QSAR Comb. Sci.* **2007**, *26* (10), 1027.
- (4) Subir s-Funosas, R.; Prohens, R.; Barbas, R.; El-Faham, A.; Albericio, F. *Chem. Eur. J.* **2009**, *15* (37), 9394.
- (5) Carpino, L. a.; El-Faham, A.; Minor, C. a.; Albericio, F. *J. Chem. Soc. Chem. Commun.* **1994**, *0* (2), 201.
- (6) Carpino, L. A.; El-Faham, A.; Albericio, F. *Tetrahedron Lett.* **1994**, *35* (15), 2279.
- (7) Albericio, F.; Bofill, J. M.; El-Faham, A.; Kates, S. A. *J. Org. Chem.* **1998**, *63* (26), 9678.
- (8) Mikolajczyk, M.; Kielbasinski, P. *Tetrahedron* **1981**, *37* (2), 233.
- (9) Albericio, F.; Chinchilla, R.; Dodsworth, D. J.; N jera, C. *Org. Prep. Proced. Int.* **2001**, *33* (3), 203.
- (10) Jad, Y. E.; Acosta, G. A.; Khattab, S. N.; de la Torre, B. G.; Govender, T.; Kruger, H. G.; El-Faham, A.; Albericio, F. *Org. Biomol. Chem.* **2015**, *13* (8), 2393.
- (11) Cruz, L. J.; Beteta, N. G.; Ewenson, A.; Albericio, F. *Org. Process Res. Dev.* **2004**, *8* (6), 920.
- (12) Isidro-Llobet, A.; Guasch-Camell, J.;  lvarez, M.; Albericio, F. *European J. Org. Chem.* **2005**, *2005* (14), 3031.
- (13) Arai, K.; Miyajima, H.; Mushiroda, T.; Yamamoto, Y. *Chem. Pharm. Bull. (Tokyo)*. **1989**, *37* (12), 3229.
- (14) Lawrenson, S. B.; Arav, R.; North, M. *Green Chem.* **2017**, *19* (7), 1685.
- (15) MacMillan, D. S.; Murray, J.; Sneddon, H. F.; Jamieson, C.; Watson, A. J. B. *Green Chem.* **2013**, *15* (3), 596.
- (16) Prat, D.; Hayler, J.; Wells, A. *Green Chem.* **2014**, *16* (10), 4546.
- (17) Acosta, G. A.; Del Fresno, M.; Paradis-Bas, M.; Rigau-DeLlobet, M.; C t , S.; Royo, M.; Albericio, F. *J. Pept. Sci.* **2009**, *15* (10), 629.
- (18) Sheehan, J. C.; Hess, G. P. *J. Am. Chem. Soc.* **1955**, *77* (4), 1067.
- (19) Opatz, T.; Kunz, H. *Tetrahedron Lett.* **2000**, *41* (52), 10185.

- (20) Honda, M.; Morita, H.; Nagakura, I. *J. Org. Chem.* **1997**, *62* (25), 8932.
- (21) Mitsunobu, O.; Obata, T.; Mukaiyama, T. *J. Org. Chem.* **1965**, *30* (4), 1071.
- (22) Mathavan, I.; Zirah, S.; Mehmood, S.; Choudhury, H. G.; Goulard, C.; Li, Y.; Robinson, C. V; Rebuffat, S.; Beis, K. *Nat. Chem. Biol.* **2014**, *10* (5), 340.

**CHAPTER 2:
STRUCTURAL ELUCIDATION
AND BIOLOGICAL
EVALUATION**

The present chapter comprises the characterization and comparison of the synthetic peptides. They were compared chemically and by thermal and proteolytic assays. Afterwards, their cytotoxicity and inhibitory effects were evaluated in lung cancer human cells. Chaxapeptin, a natural lasso peptide, was used as a reference for the comparison studies. Its isolation will be further explained in Chapter 3.

For the structural elucidation, several techniques have been used, such as mass spectrometry (including mass fragmentation and ion mobility), crystallization and NMR spectroscopy. Then, structural calculations were carried out applying the NOE constraints obtained by NMR. In the end, the combination of all these techniques allowed the structural characterization of these peptides.

Chemical comparison

In order to compare the bicyclic products obtained by the different strategies and the branched cyclic analog, samples were dissolved in H₂O-ACN (1:1) and analyzed by RP-HPLC using a C18 analytical column. Moreover, their molecular weight was characterized by MALDI-TOF (Figure 35).

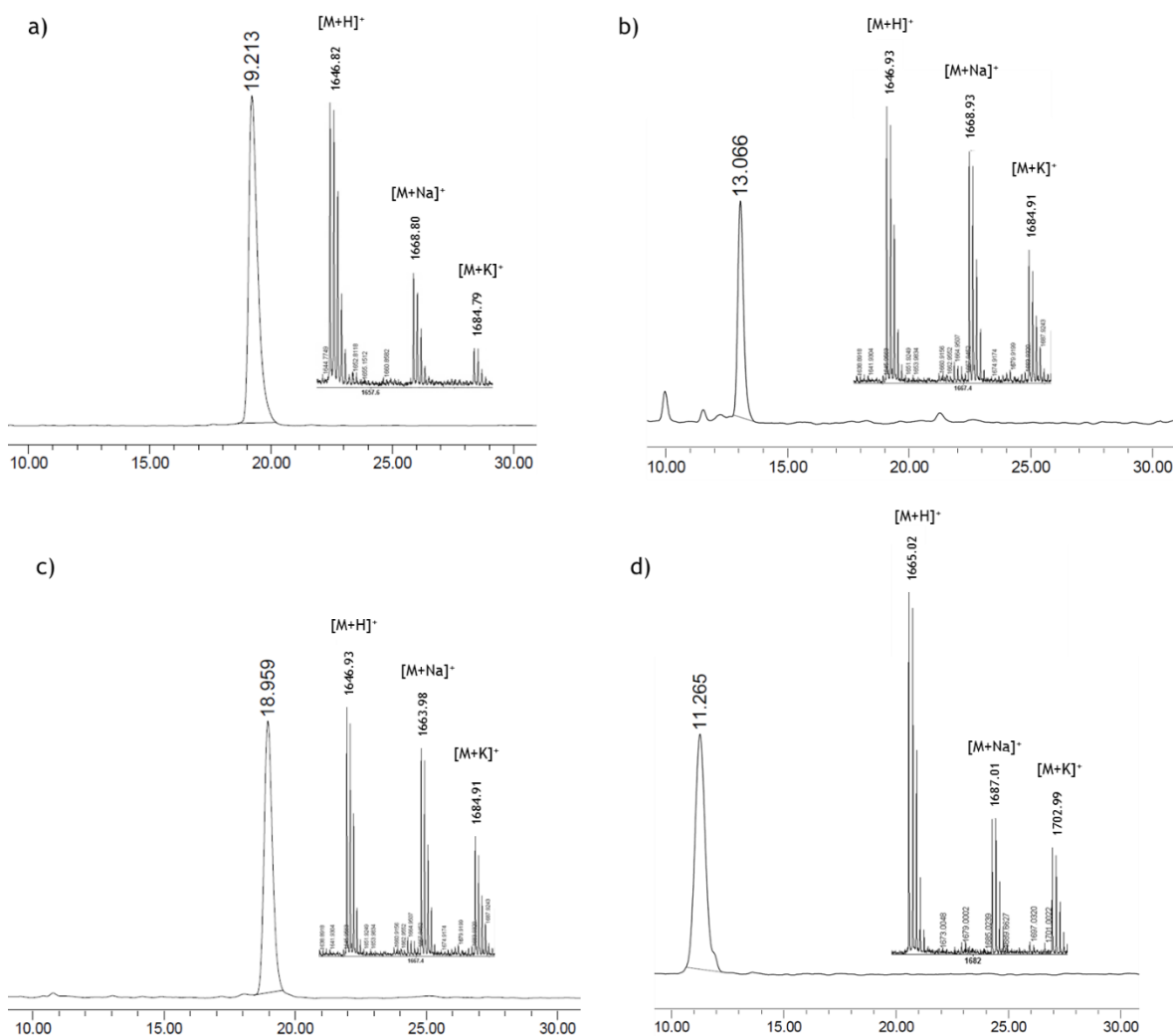


Figure 35: RP-HPLC chromatograms and MALDI-TOF spectra of the three bicyclic analogs obtained by the different strategies: a) strategy A, b) strategy C, c) strategy D, and d) branched cyclic analog. Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

In the analytical HPLC, peptides obtained by the strategies A and D co-eluted, showing a single peak (Figure 36), which indicated chemical equivalence between the two samples. On the other hand, it could be observed that the retention time (t_R) of the bicyclic peptide from strategy C was clearly different to the products obtained with the other strategies, its t_R being closer to the branched cyclic peptide than to the bicycles.

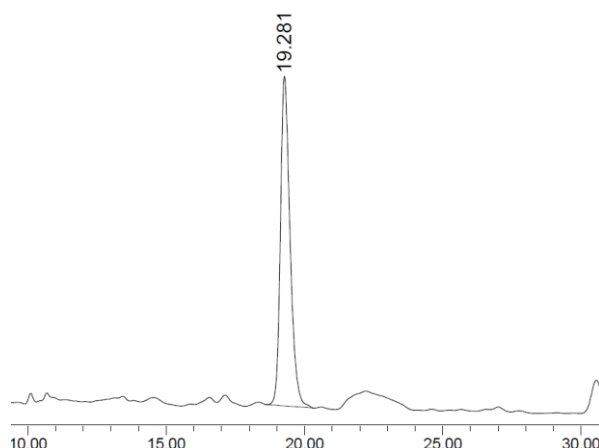


Figure 36: RP-HPLC chromatogram of the co-injection of peptides obtained by strategy A and D. Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

The single peak obtained in the co-injection experiment, showed that even using different strategies, the same product with the same t_{R} could be obtained. However, strategy C led to a different product, in which, using that synthetic strategy, it ensured that the tail could not be trapped inside the ring.

Subsequently, in order to assess the lasso formation, all bicyclic peptides were treated with a solution of 0.5 M NaOH in H_2O -ACN (1:1) to obtain the respective branched cyclic peptides (Figure 37) (see Materials and Methods for more details). The t_{R} of all peptides changed dramatically, resulting in wider peaks, in some cases, and t_{R} values closer to the branched cyclic analog. This significant change in the t_{R} meant that the peptides had lost their initial structure, resulting in a more unfolded peptide.

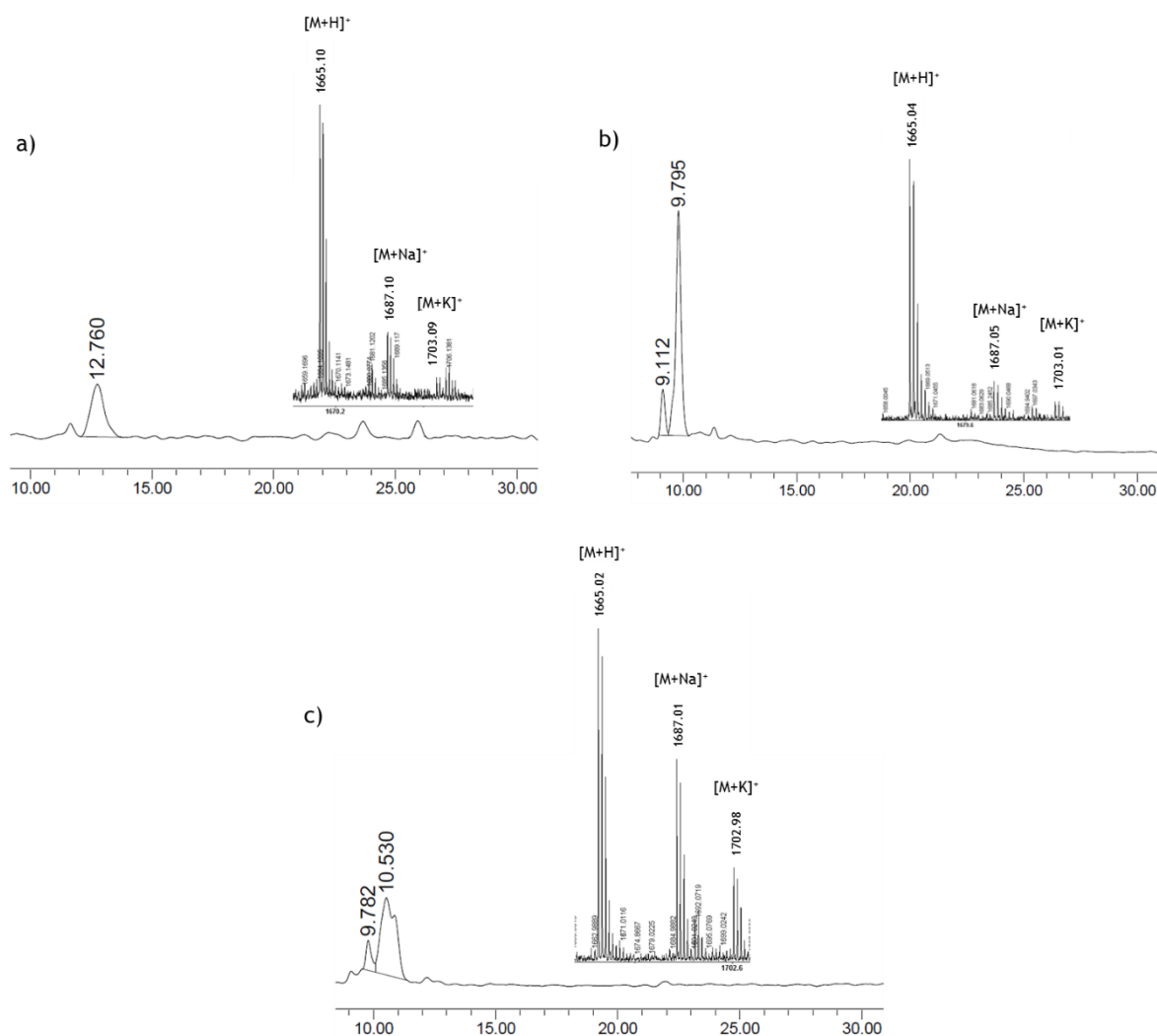


Figure 37: RP-HPLC chromatograms and MALDI-TOF spectra of the unthreaded bicyclic peptides obtained by: a) strategy A, b) strategy C and c) strategy D. Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

Stability assays

In order to study the biochemical stability of these molecules, all the obtained peptides, together with their respective unthreaded analogs, were assayed against carboxypeptidase Y (Figures 38 and 39). Moreover, their thermal stability at 95 $^{\circ}\text{C}$ was also tested (Figure 40). Carboxypeptidase Y is the most extensively used protease for lasso peptides evaluation. It was selected because it is relative unspecific and can remove any amino acid from the C-terminus. It can be used to evaluate the stability of all lasso peptides independently of their sequence.

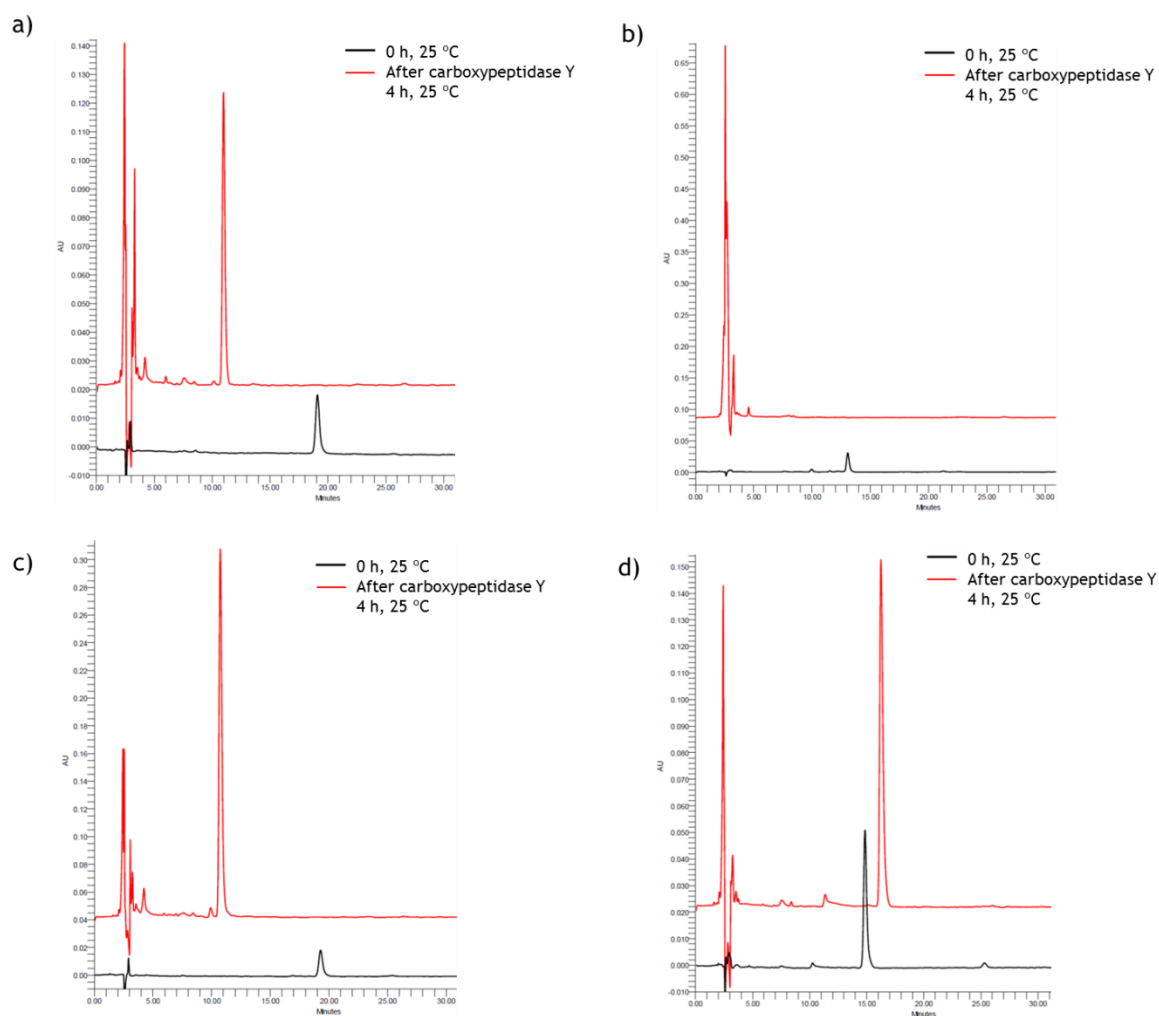


Figure 38: RP-HPLC chromatograms of the carboxypeptidase Y assay of bicyclic analogs obtained from: a) strategy A, b) strategy C, c) strategy D, and d) chaxapeptin. Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

The results obtained in the carboxypeptidase Y digestion assay allowed a clear distinction between the products obtained by the different strategies. Bicyclic peptides from strategies A and D showed the same digestion pattern, which corresponded to the loss of Leu15, whereas the peptide from strategy C was readily digested, showing a cleavage between Ser9 and Phe10 with ester bond hydrolysis. On the other hand, the natural lasso peptide, chaxapeptin, remained intact after the protease treatment, which is in agreement with the already published results from Elsayed and coworkers.¹

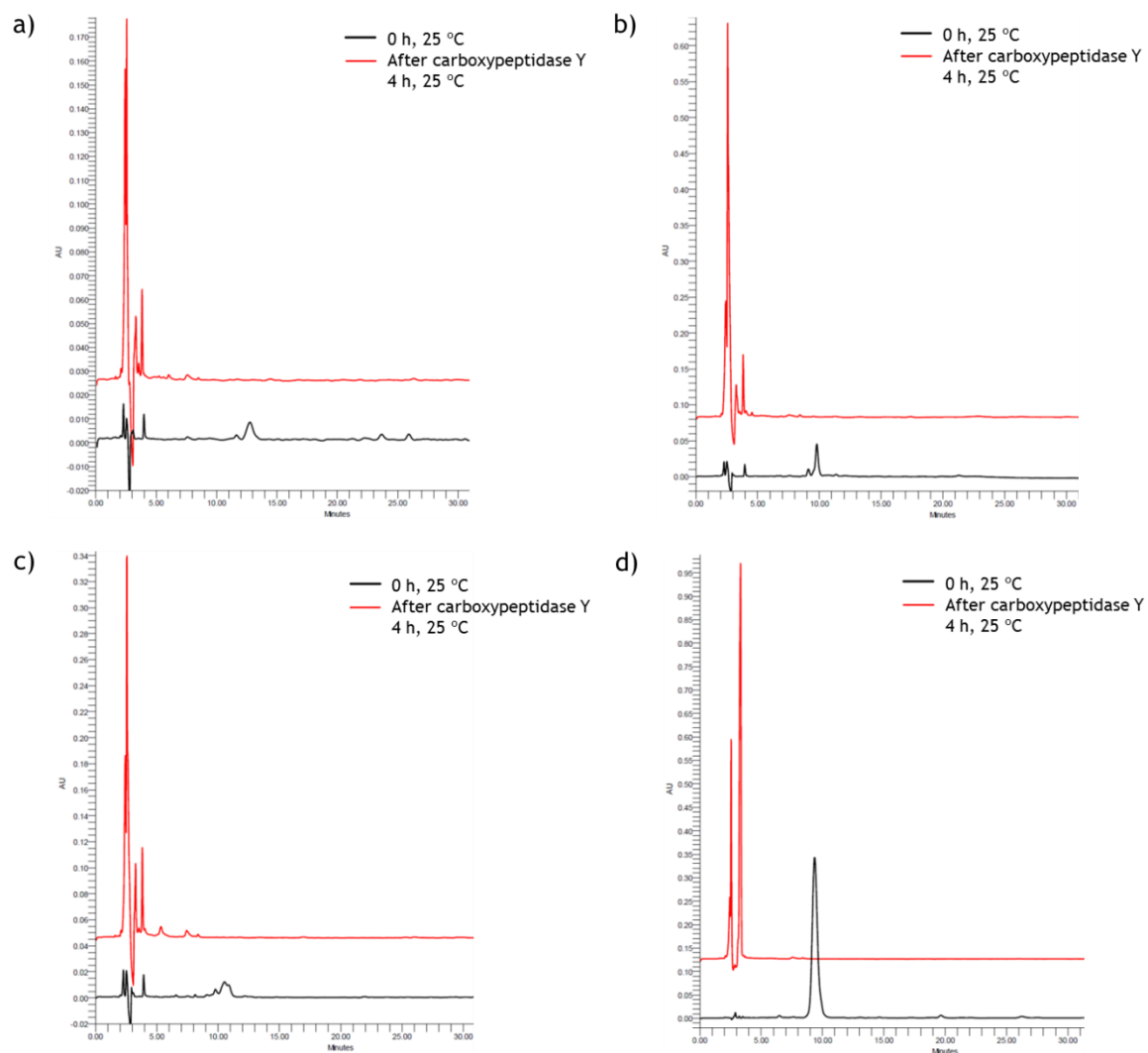


Figure 39: RP-HPLC chromatograms of the carboxypeptidase Y assay of the unthreaded bicyclic analogs obtained from: a) strategy A, b) strategy C, c) strategy D, and d) synthetic branched cyclic (Scheme 20). Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

As observed in Figure 39 the branched cyclic peptides obtained after the ester hydrolysis were completely degraded. Subsequent LC-MS analysis revealed the same digestion pattern for all of them, which corresponded to the amide cleavage between Ser9 and Phe10. It was the same behavior than for the branched cyclic analog.

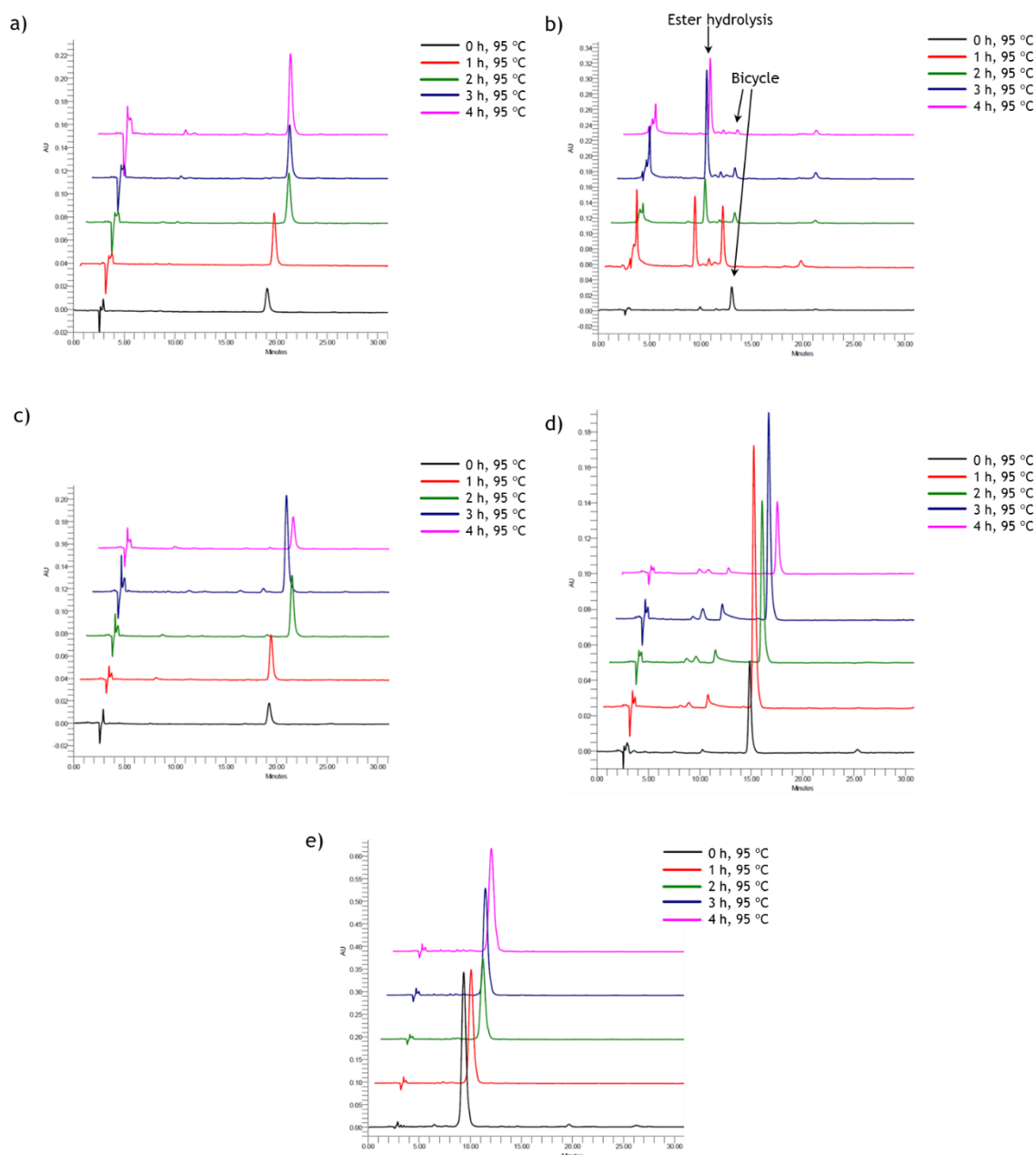


Figure 40: RP-HPLC chromatograms of the heat stability at 95 °C of bicyclic analogs obtained from: a) strategy A, b) strategy C, c) strategy D, d) chaxapeptin and e) synthetic branched cyclic. Conditions: Phenomenex Luna C18 5 μ m (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 °C.

Concerning the thermal stability of the peptides, the tendency observed was the same than for the proteolytic assay. Bicyclic peptides from strategies A and D were completely stable for 4 h at 95 °C, as well as chaxapeptin. Nevertheless, the bicyclic peptide from strategy C was not stable. After 1 h, 50% of the bicyclic peptide was hydrolyzed, yielding the corresponding branched cyclic peptide. The branched cyclic analog showed high stability after 4 h, reporting the same behavior as the unthreaded peptides (data not shown).

As reported in literature, high thermal stability is a general characteristic of lasso peptides, with some exceptions such as, astexin-1, caulosegnin I and III, caulonodins V, VI and VII, rhodanodin, rubrivinodin, sphingonodin II, sphingopyxin I and syanodin I.^{2,3,4,5} In some cases they showed a change in the chromatographic behavior, but in other cases no change was observed after prolonged exposure. In particular, for caulonodin VI, proteolytic digestion performed after the thermal treatment revealed sensitivity against carboxypeptidase Y, leading to the hypothesis that unfolding had happened but its t_R was very similar to the lasso peptide.⁴

Taking these facts as reference, bicyclic peptides from strategies A and D were subjected to thermal exposure at 95 °C followed by carboxypeptidase Y digestion, trying to detect some differences between both. Sphingopyxin I was used as a control because it is an example of lasso peptide labile to carboxypeptidase Y and heat sensitive (Figure 41). The results indicated that after 4 h at 95 °C, no difference in the t_R values and mass were observed. However, after proteolytic digestion, a clear difference between them was observed. For the bicyclic peptide from strategy A (Figure 41a), partial digestion was detected with loss of Leu15. This was the same hydrolysis pattern of this bicycle but without the previous thermal treatment (Figure 38a). On the other hand, the peptide obtained by strategy D was completely degraded (Figure 41 b), revealing that the bicycle was unfolded or the conformation changed after the thermal treatment. This result pointed to the fact that after the thermal treatment, the two bicyclic conformations were different, even though they showed the same t_R .

Furthermore, sphingopyxin I showed loss of the last two amino acids from its C-terminus (Gly20 and Gly21) after the thermal treatment, concomitant with an increase in the t_R (Figure 41c). After proteolytic digestion, cleavage of the amide bond between Tyr16 and Leu17 was detected, showing that the heat treatment produced the unfolding of the lasso structure, making these residues more accessible for proteolytic digestion.

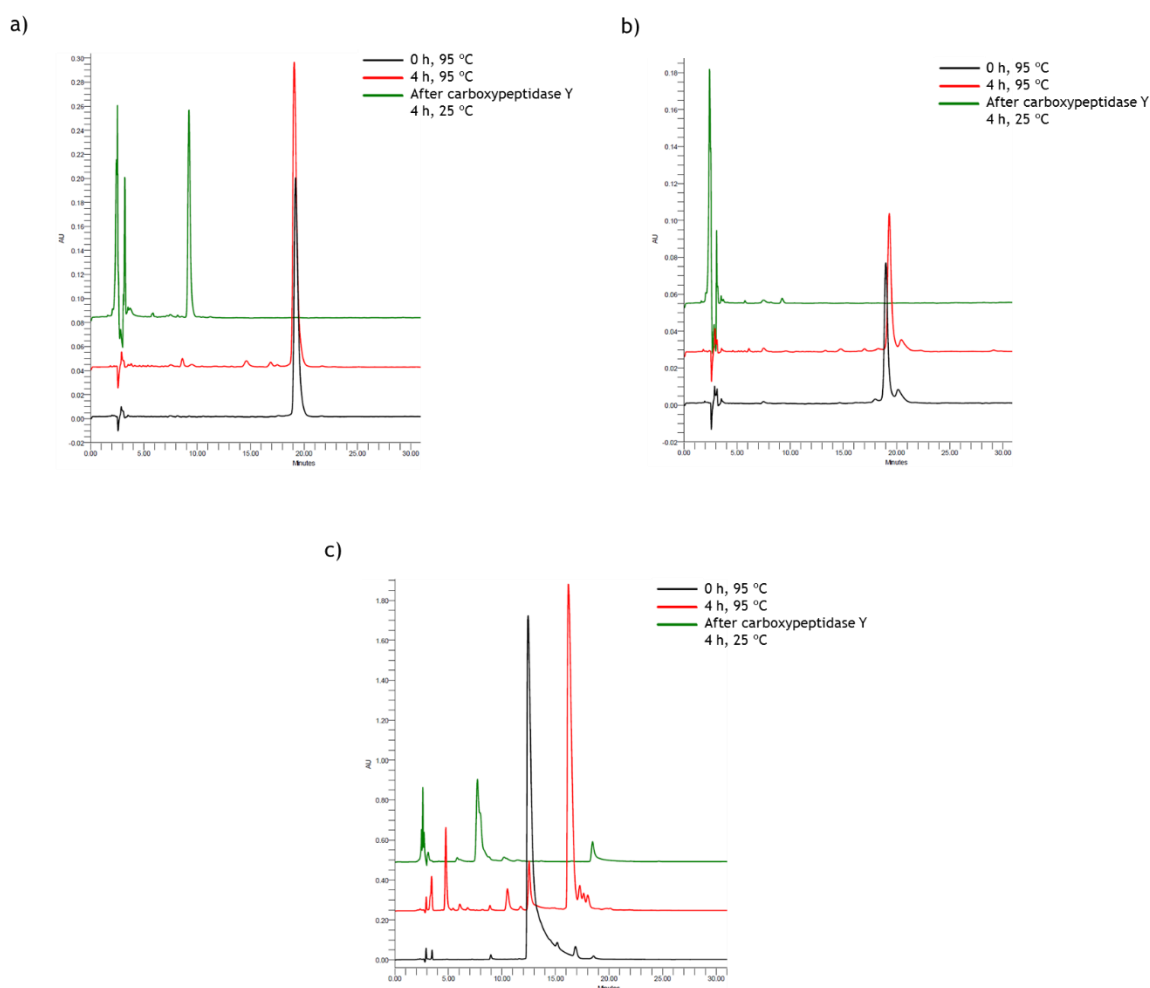


Figure 41: RP-HPLC chromatograms of the thermal treatment followed by the carboxypeptidase Y assay. a) bicyclic peptide from strategy A, b) bicyclic peptide from strategy D and c) sphingopyxin I. Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

Regarding the chemical comparison of the different analogs, the thermal and proteolytic stability studies showed that the peptide derived from strategy C was heat sensitive and proteolytically unstable. Therefore, it was not a good candidate for further development. On the other hand, the bicyclic peptide obtained by strategy A was the more promising candidate, showing a very similar behavior as chaxapeptin. It was thermally and proteolytically stable against carboxypeptidase Y, as only the last residue from the tail was cleaved, while maintaining the loop intact.

Concerning the bicyclic peptide derived from strategy D, it was previously thought that it exhibited the same stability behavior as the one obtained by strategy A. However, the last assays performed revealed that it was heat sensitive and consequently, susceptible to proteolytic degradation.

Finally, the treatment of the peptides with harsh basic conditions (0.5 M NaOH) resulted in a loss of the structure that had provided them with high stability. This result pointed to the fact that, in these peptides, the tail was located outside the ring, and therefore the cleavage of the ester bond, which maintained a tight structure, resulted in a more flexible and easily degraded molecule.

Crystallization

Crystallization studies were performed to go in depth into the characterization of the synthetic sungsanpin. The objective was to find the suitable condition that provides the crystal structure of the most stable conformation. In order to start this search, several 96-well-plate crystallization kits were used (see Materials and Methods for more details). Two screening packs were tested first. Under four of these conditions, some crystals appeared. In three of them, 2-methyl-2, 4-pentanediol (MPD) was present as precipitant. Subsequently, each condition was optimized in 24-well plates, by changing the pH and the percentage of precipitant. After this optimization, no improvement in the crystal growth was observed, indicating that they were not suitable conditions for such small peptides. For this reason, conditions suitable for small molecules were tested. However, with these new conditions, no improvement was achieved.

Afterwards, the crystallization of the racemic mixture was set up with the same buffer conditions assayed above. The use of racemic mixtures increases the probability to obtain a crystal because the number of chiral centers—and thus, the number of chiral space groups responsible of the crystallization—increases.⁶ With a highly concentrated sample (15.5 mg/mL), the racemic mixture precipitated after one day. The concentration was therefore decreased to 8.8 mg/mL and in three random conditions some crystals were detected. They were collected and flash-frozen in liquid nitrogen. However, subsequent analysis in the ALBA synchrotron revealed that they were salt crystals. Then, more conditions were tested, and a cube crystal was observed using a crystal grown at 20 °C from 0.1 M HEPES, pH 7.5, 20% w/v polyethylene glycol (PEG) 4000 and 10% w/v isopropanol. However, no diffraction was acquired, indicating that it was not a proper crystal. Due to the fact that a precipitate was often observed the next day of the seeding, the concentration of the racemic mixture was decreased to 4.4 mg/mL. However, under these last conditions, no crystals were observed.

One of the presumed problems that limited the crystallization success was the flexibility of this peptide, due to the 7-residue linear tail. For this reason, the bicyclic analogue from strategy A was tested using the two initial 96-well plate commercial packs. Due to the lack of sample and the complex synthesis, a lower peptide concentration was used in this attempt. However, despite having a more rigid and constrained peptide, no crystals were observed.

In summary, no crystals were obtained for the synthetic sungsanpin, neither with the pure sample nor with the racemic mixture. Regarding the bicyclic analog obtained by strategy A, no signal of crystal growth was observed. This fact meant that the constrained structure was probably not the hurdle. In fact, only two crystal structures of lasso peptides have been reported so far, from the total of 43 lasso peptides known. This low success rate highlights the limitations of this technique when applied to lasso peptides, largely due to the intrinsic properties of these molecules and the lack of reported specific crystallization conditions, altogether making it a challenging process.^{7,8}

MS/MS analysis

From the results obtained in the stability studies, we selected the bicyclic analog obtained by strategy A as a promising candidate for further studies. To allow for comparison, three other peptides were added to the study: the branched cyclic analog, the synthetic sungsanpin, and chaxapeptin. In order to find similarities or differences between the synthetic and the natural lasso peptide, all four peptides were subjected to an extensive analysis by mass spectrometry. Several features are already reported in the literature that will facilitate the comparison.

In the first study, the peptides were analyzed by MS/MS fragmentation. In this experiment, the fragmentation pattern of the branched cyclic analog and the synthetic sungsanpin was very similar (see Figures 42 and 43), showing only fragmentation around the tail, while the cycle remained intact. The closest fragmentation to the ring for both peptides took place between Phe10 and Gly11 (b_{10} , m/z 1090.52 and 1018.50, respectively). The most intense fragmentation was b_{12} (m/z 1260.62 and 1188.60, respectively), which corresponded to the fragmentation between Leu12 and Ser13, followed by Ser13-Trp14 fragmentation (b_{13} , m/z 1347.65 and 1275.63, respectively). The only difference observed between

these two peptides was based on the intensity of the y_3 fragment. The intensity of this fragment was lower for sungsanpin than for the branched cyclic analog; however, it was not a significant trend.

On the other hand, chaxapeptin showed a very similar fragmentation pattern to that of the two aforementioned peptides, where b_{12} was also the most intense fragmentation (m/z 1188.60) (Figure 44). This fragmentation corresponded to the loss of the last three residues. The next most intense fragmentation was b_{11} (m/z 1075.52), corresponding to the Gly11-Leu12 bond breakage, which is located in the peptide loop. This fragmentation pattern was not observed for any of the previously analyzed branched cyclic peptides, while for the chaxapeptin was very intense. Compared to the branched cyclic peptide, these two most intense fragmentations for chaxapeptin, adjacent to Leu12 (b_{11} and b_{12}), made this residue the major cleavage side. This significant difference indicated that the lasso structure played an important role in the fragmentation pattern.

The less intense fragmentations for chaxapeptin were b_{13} and b_{10} , which were less favored than for the branched cyclic peptides. The difference in the intensities could be due to the compact structure adopted for chaxapeptin, which made less accessible these positions. This behavior was in agreement with previous works, which reported that all class II lasso peptides showed fragmentations around the tail. However, these fragmentations were less intense than in branched cyclic peptides due to their constrained structures.^{5,9,10}

Considering the y fragmentations, the chaxapeptin spectrum was closer to the one obtained from the branched analog than for the synthetic sungsanpin, which showed a lower intensity in these fragmentations.

All these three peptides showed a high tendency of fragmentation between residues 12 and 13. This phenomenon might be due to the bulky side chains of the three last C-terminal residues, in which the steric hindrance complicated the fragmentations.

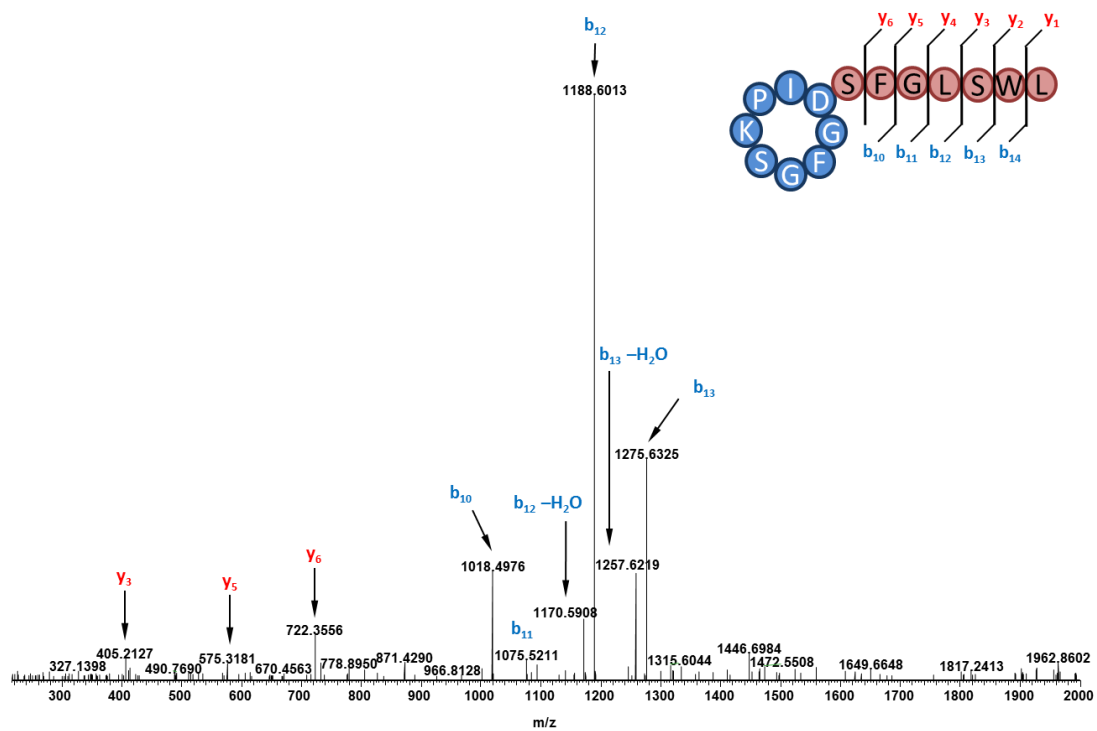


Figure 42: MS² fragmentation spectrum of the synthetic sungsanpin.

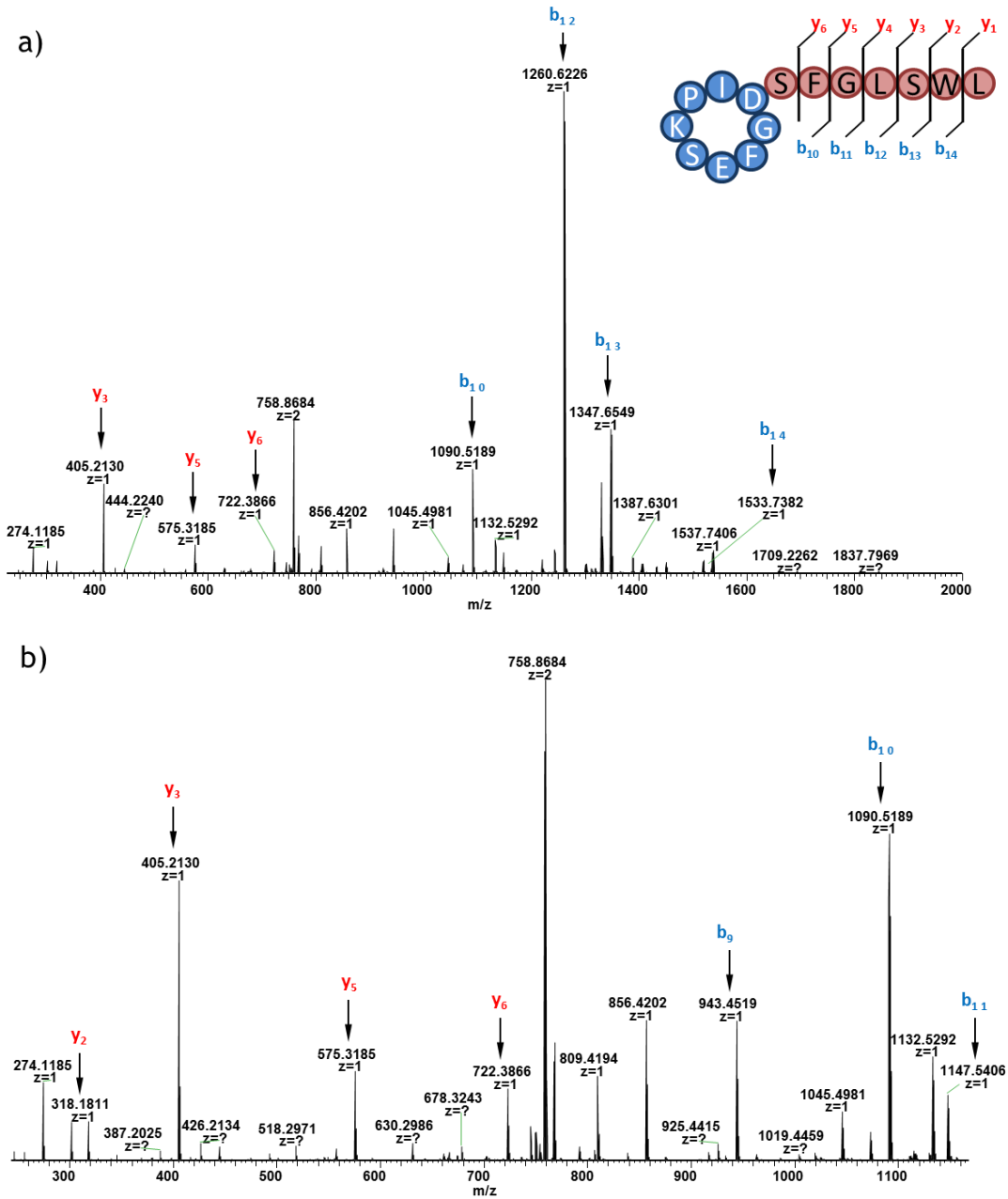


Figure 43: MS² fragmentation spectrum of the branched cyclic analog. a) complete spectrum m/z 250-2000 and b) zoom m/z 250-1150.

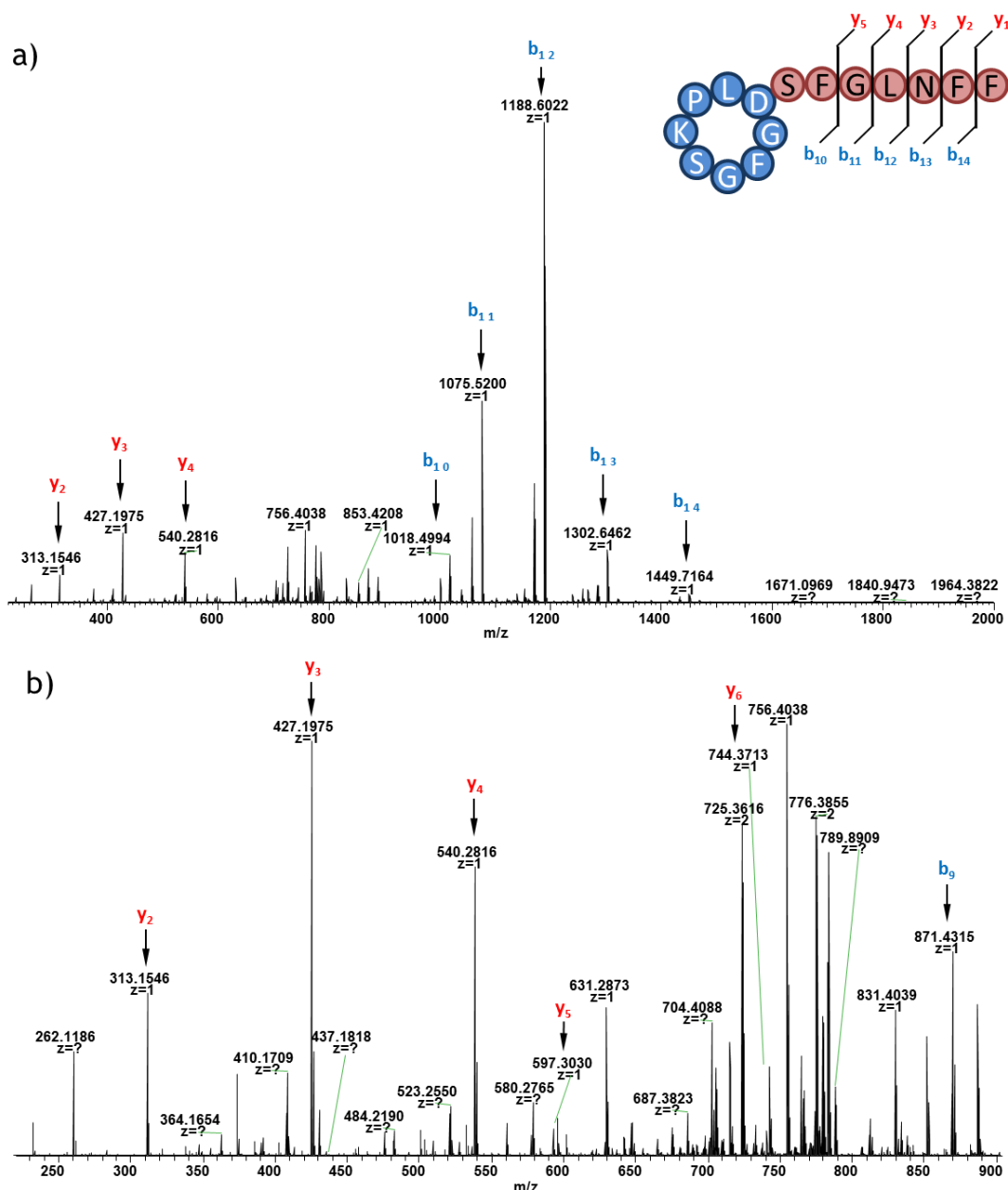


Figure 44: MS² fragmentation spectrum of the chaxapeptin. a) complete spectrum m/z 200-2000 and b) zoom m/z 220-990.

When the bicyclic analog A was analyzed, a totally different fragmentation spectrum was obtained, showing several fragments around the macrolactam ring (Figure 45). Regarding this observation, one of the most intense fragments was b_2+y_{12} (m/z 1518.72), which corresponded to a double fragmentation with the consequently loss of Lys5. The b_{13} (m/z 1329.64) fragment had the same intensity as b_2+y_{12} , being both of them the most intense fragments. This last fragment (b_{13}) was also present in the fragmentation pattern of the branched cyclic analog, the synthetic sungsanpin and the chaxapeptin, but with less intensity. In the case of the bicycle, b_{13} was more

intense than b_{12} due to the ester bond, which was formed between Ser13 and Glu3, making the fragmentation before Ser13 quite complex.

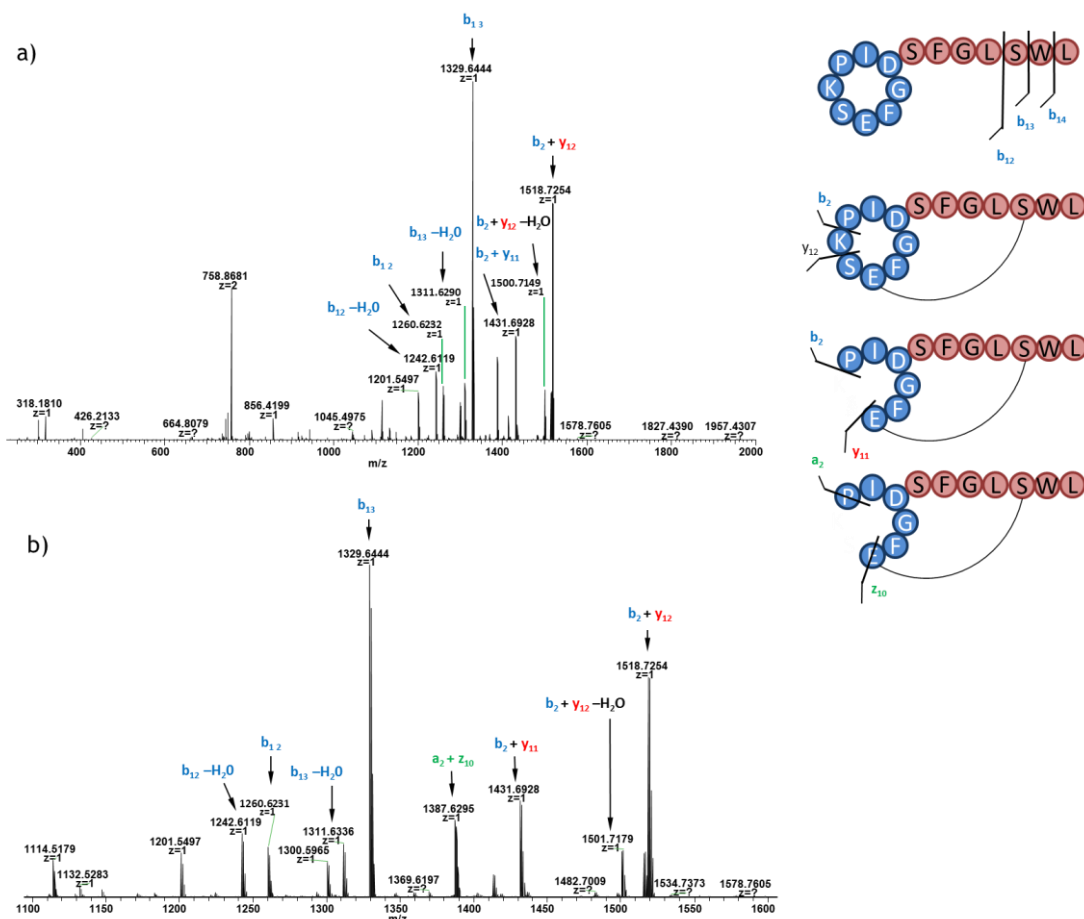


Figure 45: MS^2 fragmentation spectrum of the bicyclic analog obtained by strategy A. a) complete spectrum m/z 200-2000 and b) zoom m/z 1100-1600.

Two different sample sets of the bicyclic peptide were used to perform the mass fragmentation assays (Figure 46). Both samples resulted in the same spectrum except for the intensities of the main fragments. In one sample set, both b_{13} and b_2+y_{12} were more intense than in the other. An interesting observation was that all the ring fragmentations obtained, corresponded to sequences before Pro6. This suggested a constrained conformation of Pro6 in the bicycle, leading to the notion of Pro possibly being a turn-inducing element in the ring. In this regard, the *N*-terminal fragmentation of Pro is a typical feature for most peptides, one reason for this being the high proton affinity of the Pro ring.¹¹ However, other studies in which the substitution of Pro for higher proton-affinity residues resulted in a different selective cleavage, pointed to the existence of additional factors.¹²

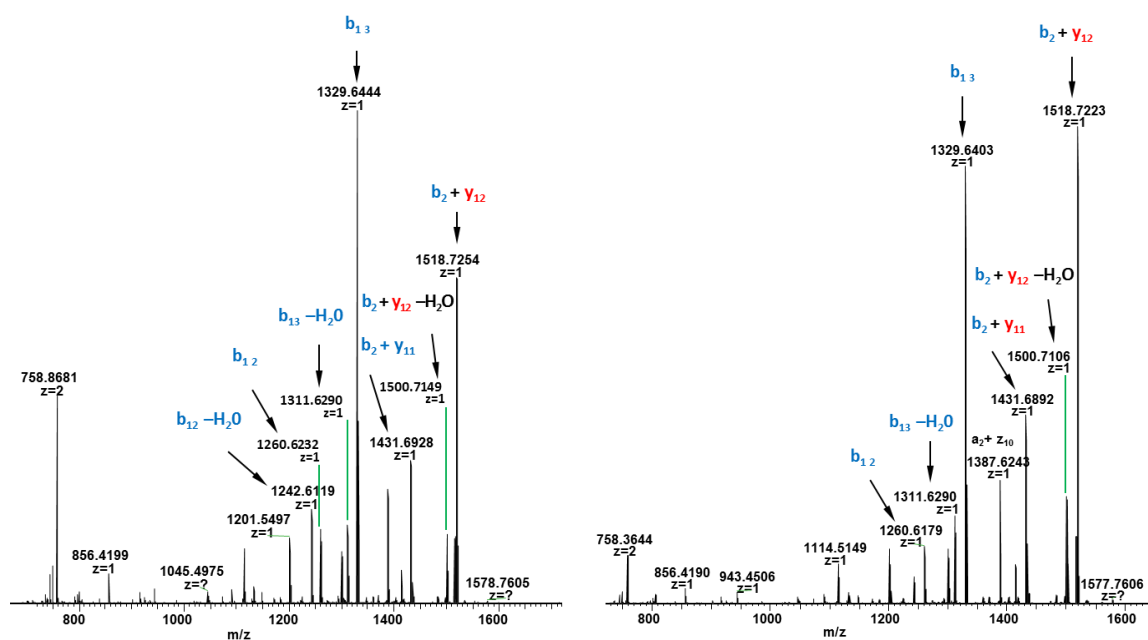


Figure 46: Comparison of MS^2 fragmentation spectra of the two different sample sets of the bicyclic analog obtained by strategy A. Zoom m/z 750-1650.

Despite being the *N*-terminal fragmentation of Pro a typical fragmentation, neither in the natural lasso peptide, chaxapeptin, nor in the branched cyclic peptides was observed. This fact indicated that for the bicycle, this fragmentation was more feasible than for other peptides, due to the constrained structure. In branched cyclic and lasso peptides, there were other fragmentations more accessible, like the linear tail. Concerning this behavior, this made us think that the bicyclic analog adopted a behavior more similar to class I and III lasso peptides than to class II.

The fragmentation pattern of class I and III lasso peptides is different from the one obtained from class II lasso peptides.² In this regard, class I and III share a weak fragmentation behavior and show fragmentation around the ring, which makes the distinction between a bicyclic and a lasso peptide difficult. This behavior could be a consequence of the fact that, in order to detect some fragments in bicycles, two or three bond breakages were required. Nevertheless, to characterize a lasso structure only with one cleavage was enough. This was the case for example of BI-32169,¹³ a class III lasso peptide, and svicucin, a class I lasso peptide.¹⁴ To detect fragments in BI-32169, two peptide bond breakages had to occur. In MS^2 studies of the reduced BI-32169, weaker fragments than expected were obtained from peptide bond cleavage at the *C*-terminal tail. Furthermore, svicucin mainly showed fragmentation in the ring and no degradation after carboxypeptidase Y digestion. This could be a consequence of the disulfide bonds. In order to exclude the effects of the disulfide

crosslink in peptide, the disulfide bonds were reduced in mild conditions and two peaks were detected. Afterwards, the carboxypeptidase Y assay showed that one peak remained intact, while the other was totally degraded. This fact revealed that the reduced svuceucin was a mixture of lasso and branched cyclic peptides. Subsequent MS^2 studies indicated that in both cases fragmentations in the tail were mainly detected, but in the case of the branched cyclic peptide they were more intense.

The bicyclic analog A showed a very similar behavior to BI-32169 and svuceucin, showing stability against carboxypeptidase Y and fragmentation in the macrocycle. Based on the MS^2 studies, we could not distinguish between a bicyclic and a lasso-structured peptide, as the weak fragmentation behavior could also be a consequence of the fact that more than one bond breakage was required rather than for the supposed lasso structure. Here, the ester bond of the bicycle was hydrolyzed (Figure 37a) and as observed previously, after the digestion with carboxypeptidase Y the hydrolyzed peptide was completely degraded (Figure 39a). This fact revealed that after the hydrolysis, the peptide lost its initial structure adopting a new one with more flexibility. This result indicated that the bicycle could not adopt a lasso structure and the fragmentation pattern was instead due to the bicycle structure.

To sum up the results obtained so far in this part, the synthetic sungsanpin and the branched cyclic analog showed very similar fragmentation spectra, which were different to chaxapeptin, thus revealing that they did not adopt a lasso structure. However, a common feature between these three peptides was that they only showed fragmentations in the tail, whereas the main fragmentations for the bicycle were in the macrocyclic ring. This fact demonstrated the totally different fragmentation pattern of the bicyclic analog A, which was more typical of a bicyclic peptide or a class I or III lasso peptide. Moreover, when the ester bond was hydrolyzed, the bicycle lost its proteolytic stability. Thus, all results pointed to the hypothesis that the bicyclic analog A did not adopt a lasso-like structure, as well as the results obtained from the stability assays.

IM-MS analysis

As mentioned in the introduction, IM-MS has become a useful and effective technique to differentiate lasso peptides from their branched cyclic topoisomers. In order to better characterize the bicyclic structure, ion mobility studies were performed. Herein, the four peptides previously reported were compared using this technique to find differential features between them.

In previous work by Dit Fouque and coworkers, five class II lasso peptides and their respective branched cyclic topoisomers were compared by IM-MS.¹⁵ From this study, three characteristic trends were established: 1) little change in collision cross section (CCS) with increasing charge state, 2) low intensity of their highly protonated species ($[M + 4H]^{4+}$), even when a supercharging agent, such as sulfolane, was added, and 3) narrow ion mobility peaks, which translates into a small number of conformations with different charge states.

From each of these trends, different parameters were measured, and the obtained values were used to compare the different conformations. The first parameter was the range of CCS covered by all protonated molecules ($\Delta\Omega/\Omega$), the second was the relative abundance of highly charged multiply protonated molecules (ζ) and the third parameter was related to the IM peak widths.

Considering these remarks, the discussion of our results was based on the comparison between the values of the four studied peptides and the previous features already described.

1) CCS values

The first trend described that narrow changes in CCS values at increasing charged states were characteristic for lasso peptides and that the differences between a lasso and a branched cyclic peptide were more pronounced at highly charge states.

Peptides	Charge state	no. of conformations	CCS (Å ²) (Ratio (%))
Chaxapeptin	[M + 2H] ²⁺	4	290 (10)/ 307 (53) /330 (33)/367 (4)
	[M + 3H] ³⁺	4	281 (18)/ 295 (52) /312 (23)/337 (8)
Bicyclic analog A	[M + 2H] ²⁺	4	306 (10)/ 319 (61) /340 (27)/373 (1)
	[M + 3H] ³⁺	4	296 (27)/ 317 (51) /337 (18)/350 (5)
Branched cyclic analog	[M + 2H] ²⁺	4	297 (24)/ 321 (43) /346 (29)/372 (4)
	[M + 3H] ³⁺	4	305 (16)/ 323 (52) /337 (23)/357 (9)
Synthetic sungsanpin	[M + 2H] ²⁺	3	346 (39)/328 (18)/ 312 (43)
	[M + 3H] ³⁺	3	388 (22)/ 352 (63) /310 (15)
	[M + 4H] ⁴⁺	2	391 (25)/ 426 (75)

Table 14: Number of conformations and CCS of the four studied peptides with 400 mM of sulfolane. Charge state $[M + H]^+$ is not shown because is a very unfolded structure with the highest drift time values and is not relevant for the comparison. Each CCS was calculated based on the drift times.¹⁶ When several conformations were evidenced, the major species is highlighted in bold.

From the results in Table 14, the following trend was observed: when the charge state was increased, the CCS remained relatively constant for all peptides, except for the synthetic sungsanpin, in which there was a substantial gain in the CCS value between $[M + 3H]^{3+}$ and $[M + 4H]^{4+}$. This increase and the higher values obtained in the high charge state were pointed to an unfolded structure.

In addition, the CCS values for $[M + 2H]^{2+}$ were found to be very similar, around 300 Å² for all the peptides. On the contrary, for $[M + 3H]^{3+}$ more differences could be observed. It is already reported that more differences between a lasso and a branched cyclic peptide can be observed at higher charge states, while for lower ones, the CCS values are similar due to the central role of charge location in the stabilization.¹⁵ The better differentiation in higher charge states could be justified by the presence of electrostatic interactions between charges, which tend to favor gas-phase conformations that minimize Coulombic repulsions. An example of this behavior was the study of the capistruiin conformations in the gas phase, in which a

similar behavior as here was obtained.¹⁷ In that paper, Dit Fouque and coworkers reported that they were only able to observe differences between capistrain and its branched cyclic peptide in the charge state $[M + 4H]^{4+}$. In that case, only CCS values from lasso peptides were constant, while the other increased.

The charge state $[M + 4H]^{4+}$ was only reached for the synthetic sungsanpin, whereas for the other peptides the maximum charge state detected was $[M + 3H]^{3+}$, even after the addition of sulfolane. Regarding the mass spectrum of the last peptides, a mix of peak fragmentations was detected around the m/z of the species $[M + 4H]^{4+}$, which made difficult the differentiation of the quadruply protonated species. Added to this, none of those peaks were major (data not shown).

In fact, the charge state that each peptide can achieve depends on different factors, such as the nature of the amino acids that define the peptide or the conformation. Therefore, the charge state distribution depends on the electrospray conditions and is usually thought to arise from the number of basic sites and their solvent availability during the ionization process. Consistently, it was reported in the literature, that the most abundant charge state of lasso peptides was generally the triply charged protonated molecules, in contrast with the unfolded branched cyclic forms for which the most abundant charge state is the quadruply.¹⁸ This observation was in agreement with our results, in which the charge state $[M + 3H]^{3+}$ was present for all the peptides and the $[M + 4H]^{4+}$ only for the synthetic sungsanpin. A closer behavior was expected between both branched cyclic peptides, but the branched cyclic analog with Glu3 showed a conformation more similar to the bicyclic analog A and chaxapeptin than to the synthetic sungsanpin. This fact could indicate that in the gas-phase the branched cyclic analog was stabilized through hydrogen bonds, resembling a bicycle, which agreed with the previous results of the MS^2 fragmentation.

From the ion mobility spectra and using relative CCS values, it could be affirmed that narrow changes in CCS values (Ω) were obtained when the relative range of CCS covered by all protonated molecules ($\Delta\Omega/\Omega$) was between 0 and 10.9%, and for the branched cyclic peptides when it was above 18%.¹⁸

Peptides	[M+2H] ²⁺	[M+3H] ³⁺	[M+4H] ⁴⁺	$\Delta\Omega$ (Å ²)	$\Delta\Omega/\Omega$ (%)
	CCS (Ω, Å ²)	CCS (Ω, Å ²)	CCS (Ω, Å ²)		
Chaxapeptin	307	295	n.d.	12	3.9
Bicyclic analog A	319	317	n.d.	2	0.6
Branched cyclic analog	321	323	n.d.	2	0.6
Synthetic sungsanpin	312	352	391	79	20.2

Table 15: CCS values (Ω) and $\Delta\Omega/\Omega$ from the ion mobility spectra of the main conformation. n.d.: no peak detected.

The wide range of CCS (20.2%) for the synthetic sungsanpin clearly demonstrated the flexibility and unfolding of this peptide (Table 15). The other three peptides presented a very low range of CCS ($\Delta\Omega$), even with the bicyclic and the branched cyclic analogs showing a lower value than chaxapeptin. These values were consistent with a rigid and constrained structure, due to the double cyclization in one case, and to the stabilization through hydrogen bonds in the other, producing a bicyclic structure in the gas-phase.

Considering only the $\Delta\Omega/\Omega$ values, the bicyclic and the branched cyclic analogs showed the same behavior as chaxapeptin. For this reason, their differentiation was not possible only with this trend. An example of this result resides in the study carried out by Dit Fouque and coworkers, in which two macrocyclic peptides, gramicidin S and cyclosporin A, were tested and compared with lasso peptides. Low $\Delta\Omega/\Omega$ values of 0.7% and 4.7% respectively, were obtained, due to the constrained structure of the macrocycles. However, they managed to differentiate them from a lasso structure by evaluating the third trend, which corresponds to the ion mobility peak widths. This value was clearly higher for the macrocycles than for the lassos, indicating that for example cyclosporin A had a bigger macrocycle and methyl groups on the peptide backbone, which might increase the flexibility of the molecule.¹⁸

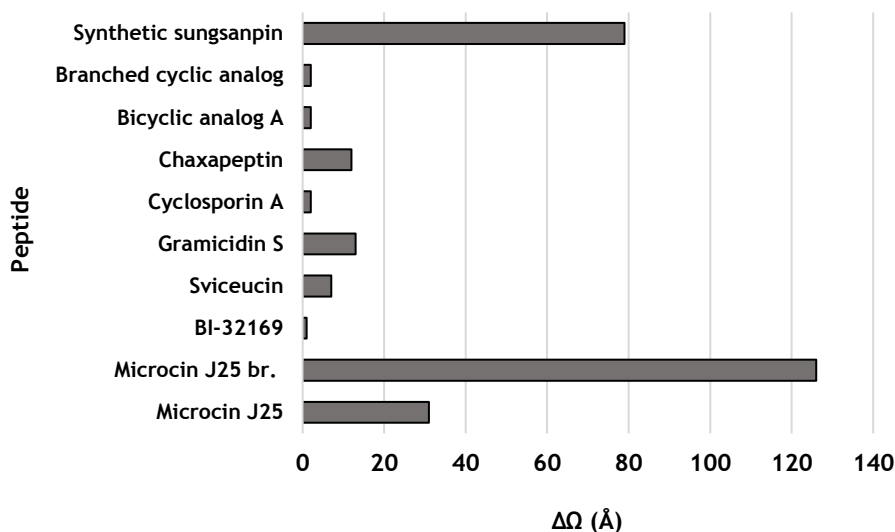


Figure 47: Range of CCS ($\Delta\Omega$) observed for all the multiply protonated ions $[M + nH]^{n+}$ of different peptides. br.: synthetic branched cyclic microcin J25.

The $\Delta\Omega$ value comparison for different peptides, including gramicidin S and cyclosporin A, is shown in Figure 47. The graphic shows a large range of CCS for the branched cyclic microcin J25 and for the synthetic sungsanpin, which was typical of flexible unfolded structures, as expected. On the other hand, the lowest values corresponded to the lasso peptide, BI-32169, followed by the bicyclic and the branched cyclic analogs and cyclosporin A, all with the same $\Delta\Omega$ value. These results indicated that the presence of second cycles or big head-to-tail macrocycles were more important than the lasso structure for the $\Delta\Omega$ value.

Concerning this first trend, we were able to confirm that the synthetic sungsanpin was clearly a branched cyclic peptide, but the other three peptides pointed out to be lasso-like.

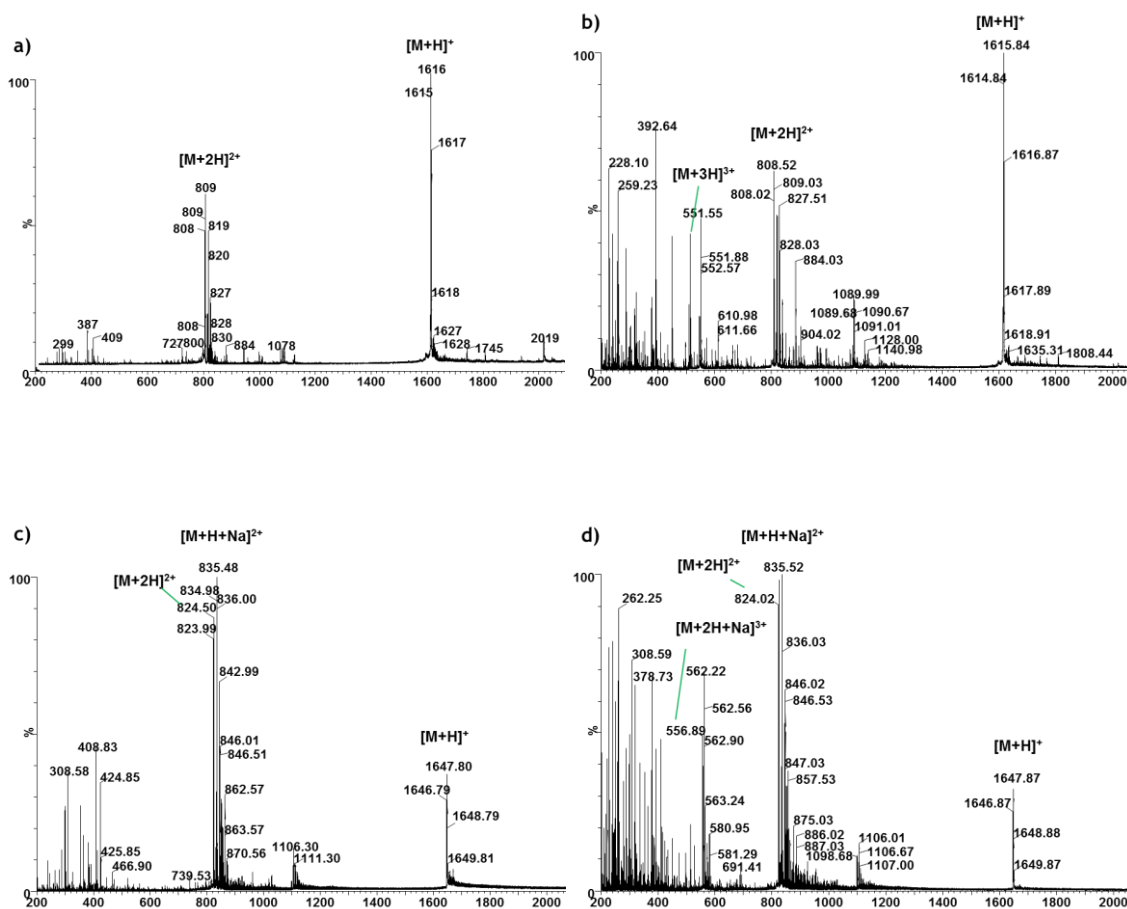
Peptides	$[M+2H]^{2+}$		$[M+3H]^{3+}$		$[M+4H]^{4+}$	
	no. of conformations	CCS (Å²)	no. of conformations	CCS (Å²)	no. of conformations	CCS (Å²)
Chaxapeptin	4	307	4	295	n.d.	n.d.
Bicyclic analog A	4	319	4	317	n.d.	n.d.
Branched cyclic analog	4	321	4	323	n.d.	n.d.
Synthetic sungsanpin	3	312	3	352	2	391

Table 16: Summary of the number of conformations and CCS of the main conformations for each studied peptide with 400 mM of sulfolane. n.d.: no peak detected.

2) Intensity of the highly protonated species

In the second distinctive trend of lasso peptides, the different mass spectra were compared with and without sulfolane (Figure 48). A low intensity peak corresponding to the highly protonated species, usually $[M + 4H]^{4+}$, has been detected in lasso peptides, even in the presence of the supercharging agent.¹⁸

Without sulfolane, only the doubly protonated ion was detected for all peptides. The intensity of this peak was lower than $[M + H]^+$ for all peptides, except for the synthetic sungsanpin. After the addition of sulfolane, the triply protonated ion was detected in a lower intensity compared to $[M + H]^+$ and $[M + 2H]^{2+}$. However, for the synthetic sungsanpin, the triply and the quadruply protonated ions were very intense, while the intensity of the doubly protonated ion decreased sharply.



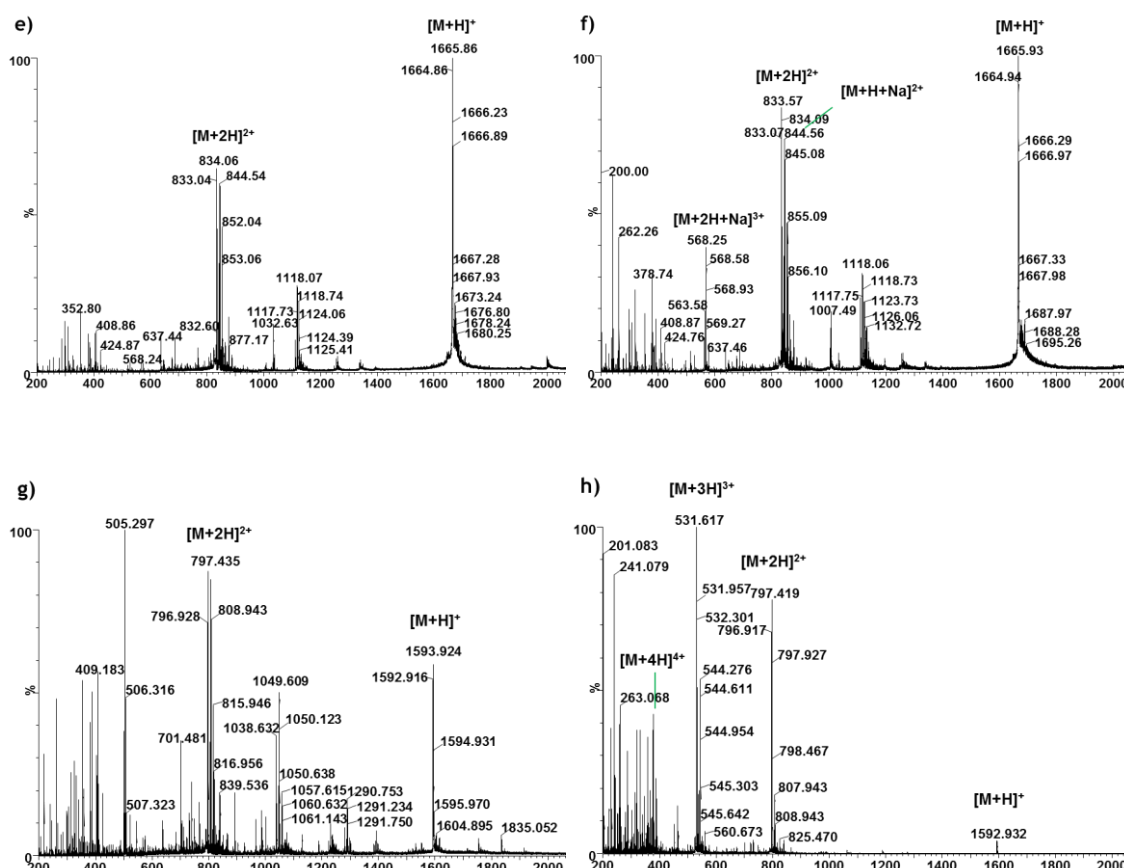


Figure 48: Mass spectra of chaxapeptin a) and b), bicyclic analog c) and d), branched cyclic analog e) and f) and synthetic sungsanpin g) and h). a), c), e) and g) without sulfolane and b), d), f) and h) with sulfolane.

Chaxapeptin, as expected for a lasso peptide, showed the lowest intensity of its highly protonated species. Moreover, the bicyclic analog A showed a slightly different spectrum after the addition of sulfolane compared to chaxapeptin, in which the intensity of the $[M + H]^+$ peak decreased, being the doubly protonated ion the most abundant. This fact was more readily detected for the synthetic sungsanpin, in which the peak of $[M + H]^+$ decreased considerably after the addition of sulfolane.

The second parameter extracted from the ion mobility spectra was the mean charge divided by the mass (ζ). This value measures the relative abundance of highly charged multiple protonated molecules independently from the mass of the peptide and it is represented by Equation 2.

$$\zeta = \frac{\sum n I_n}{m \sum I_n} \quad (2)$$

Where n and I_n are the charge and the relative intensity of the multiply protonated molecule $[M + nH]^{n+}$ and m is the mass of the neutral molecule in Da. Therefore, the resulting units for ζ are charge per Da, but they are expressed in $e \cdot kDa^{-1}$.

This value has been found to be higher for branched cyclic peptides than for lasso peptides, due to their unfolded structure and higher flexibility.^{15,18} Branched cyclic peptides can easily get high charge states with higher intensities. Thus, this ζ parameter provided a comparison of both the intensity and the charge state.

Peptides	$[M+2H]^{2+}$		$[M+3H]^{3+}$		$[M+4H]^{4+}$		Mass (Da)	ζ ($e \cdot kDa^{-1}$)
	t_d (ms)	Rel. Int. (%)	t_d (ms)	Rel. Int. (%)	t_d (ms)	Rel. Int. (%)		
Chaxapeptin	8.97	53	6.89	52	n.d.	n.d.	1614.8	1.55
Bicyclic analog A	9.19	61	7.19	51	n.d.	n.d.	1646.8	1.49
Branched cyclic analog	9.22	43	7.26	52	n.d.	n.d.	1663.8	1.53
Synthetic sungsanpin	8.89	46	7.53	63	7.11	75	1592.8	1.98

Table 17: Drift time (t_d), relative intensity and ζ from the ion mobility spectra of the main conformation for the four studied peptides. n.d.: no peak detected.

As shown in Table 17, the four peptides had similar ζ values ranging from 1.49 to 1.98 $e \cdot kDa^{-1}$. As expected, the synthetic sungsanpin had the highest value, which was in accordance with the results from the first trend, the relative range of CCS covered by all protonated molecules ($\Delta\Omega/\Omega$). The ζ values for the other three peptides indicated similar conformations.

To sum up, chaxapeptin, the bicycle and the branched cyclic analogs showed a low intensity of the highly protonated species, even after the addition of sulfolane, and similar ζ values, indicating more rigid and constrained structures than the synthetic sungsanpin. Added to this, the similar ζ values obtained for them indicated that the bicycle and the branched cyclic analog could adopt a lasso structure. The intensity-charge relation was still not enough to confirm or reject the presence of a lasso structure, thus requiring the combination with other parameters. On the other hand, the synthetic sungsanpin showed the opposite behavior: strong intensities of

the highly protonated species and high ζ value. Once again, this trend also demonstrated that the synthetic sungsanpin showed a branched cyclic structure.

3) Peak shape

The ion mobility peak width reflects the number of gas phase conformations. For more rigid structures, lower ion mobility peaks widths are expected. Usually, when the charge state increases, the number of conformations increases as well. This fact is more marked for branched cyclic peptides than for lasso peptides, for which small peaks widths are expected.

There are some factors that could affect the peak width in IM-MS, such as the time width for ion introduction in the ion mobility cell, the expansion of the ion pack due to normal molecular diffusion and the existence or conversion of several species of different mobilities within the drift region.¹⁹ To examine the last factor, full width at half maximum (FWHM) values were plotted as a function of the drift time divided by the square root of the charge number (Figure 49). A correlation with a determination coefficient of 0.98 was found after eliminating three values, corresponding to the synthetic sungsanpin, where the presence of multiple components was suspected. Multiple conformations and species were expected for this peptide, due to the flexibility of its C-terminal tail, and thus the same tendency than the other three peptides was not observed.

Peptides	FWHM (ms)		
	$[M+2H]^{2+}$	$[M+3H]^{3+}$	$[M+4H]^{4+}$
Chaxapeptin	0.48	0.29	n.d.
Bicyclic analog A	0.50	0.25	n.d.
Branched cyclic analog	0.49	0.28	n.d.
Synthetic sungsanpin	0.33	0.61	0.45

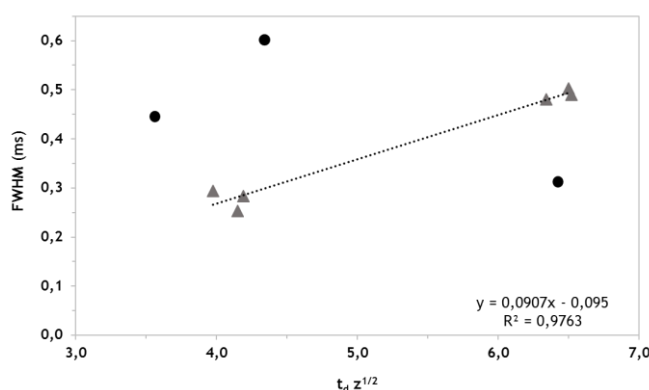


Figure 49: Left) FWHM values from the ion mobility spectra of the main conformation for the four studied peptides at different charge states. Right) Correlation between the FWHM and the drift time (Table 17) divided by the square root of the charge number. The circles are values far above the correlation line, corresponding to synthetic sungsanpin.

In general, the number of conformations obtained for all the peptides was around three (Table 16). The fourth conformation found for chaxapeptin, the bicycle and the branched cyclic analogs had low relative intensity, being less abundant. Focusing on the drift time profile of the doubly protonated ion of sungsanpin, a fourth peak around the drift time of 10 minutes could also be fitted (Figure 50). However, the drift time profile of the triply protonated ion was not feasible to fit another conformation. A common observation was the wider peak shape for the doubly protonated species, compared to the triply, which was due to the presence of more undefined and flexible conformations.

It was in the $[M + 3H]^{3+}$ charge state where more differences between them could be detected. The peak of the synthetic sungsanpin was the widest, compared to the other three peptides. This fact revealed that this peptide was quite flexible, featuring an unfolded structure. On the contrary, the chaxapeptin, the bicyclic and the branched cyclic analogs showed a similar peak profile. In other words, they showed lasso-like behavior.

The IM peak of the branched cyclic analog showed a more similar shape to chaxapeptin and to the bicyclic analog than to the synthetic sungsanpin, even though their peptide sequence differs only in one residue. This fact reflected once again the importance of the structure stabilization effect caused by the replacement of Gly3 by Glu3, producing a structure that either resembles a bicyclic or a lasso peptide.

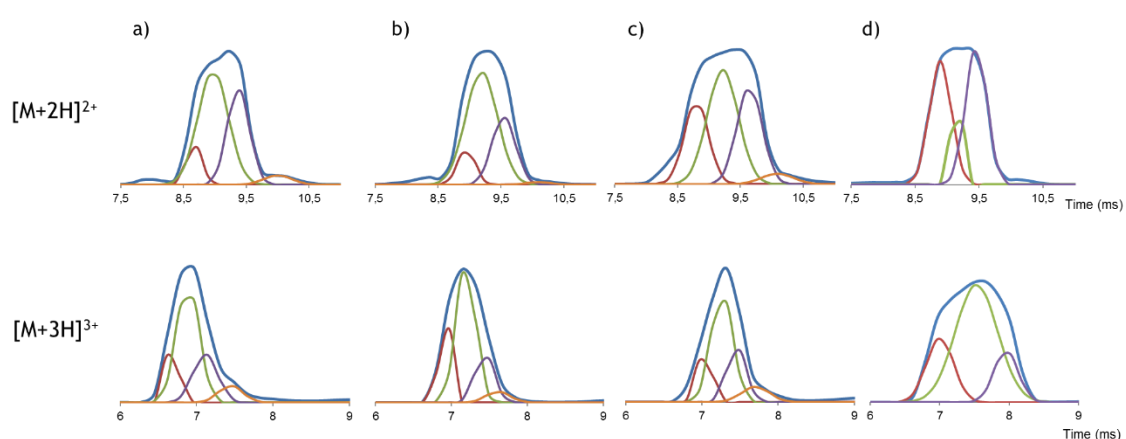


Figure 50: Drift time profiles (blue trace) and fitted peaks (red, green, purple and orange trace) of double protonated (top line) and triply protonated (bottom line) ion of a) chaxapeptin, b) bicyclic analog A, c) branched cyclic analog and d) synthetic sungsanpin.

The results obtained from the ion mobility peak shapes were in agreement with the previous trends. The synthetic sungsanpin showed the widest peak shape, whereas the other three peptides shared the same ion mobility profile. This fact revealed that the structure of the synthetic sungsanpin was more unfolded and flexible than the others, which displayed a more compact topology.

Conformational analysis

In order to better understand the different results obtained for the branched cyclic analog and the synthetic sungsanpin, an *in silico* conformational analysis was performed. The coordinates of the atoms were extracted from Protein Data Bank (PDB) 2N5C. Then, for sungsanpin, Leu7, Asn13, Phe14 and Phe15 were manually mutated to Ile, Ser, Trp and Leu, respectively. Additionally, Gly3 was mutated to Glu3 for the branched cyclic analog. Both peptides were cyclized through the side chain of Asp8 and the resulting structures were subjected to a minimization.

To identify the lowest-energy conformations, a conformational sampling algorithm was applied (with the help of Dr. Salvador Guardiola). In particular, the Macrocyclic Conformational Sampling tool, which is a part of the Schrodinger modeling suite, was used,²⁰ which includes repetitive cycles of heating/cooling and energy minimization (see Materials and Methods for more details). As a result, more conformations can be sampled faster, and a final lower-energy conformation can be usually accessed.

In the minimal energy structures, the root mean square deviation (RMSD), calculated for the peptide backbone (which is indicative of conformational similarity) was of 6.8 Å for the synthetic sungsanpin and the branched cyclic analog. This high RMSD value indicated considerable differences between both structures. As it can be observed in Figure 51, the synthetic sungsanpin adopted a more compact structure, due to its higher flexibility, whereas the branched cyclic analog showed a more rigid topology, like a bicycle. Added to this, the structure was stabilized by the electrostatic interaction between Glu3-Lys5. This reason may explain the IM results of the branched cyclic peptide, which were closer to the bicyclic peptide than to the synthetic sungsanpin.

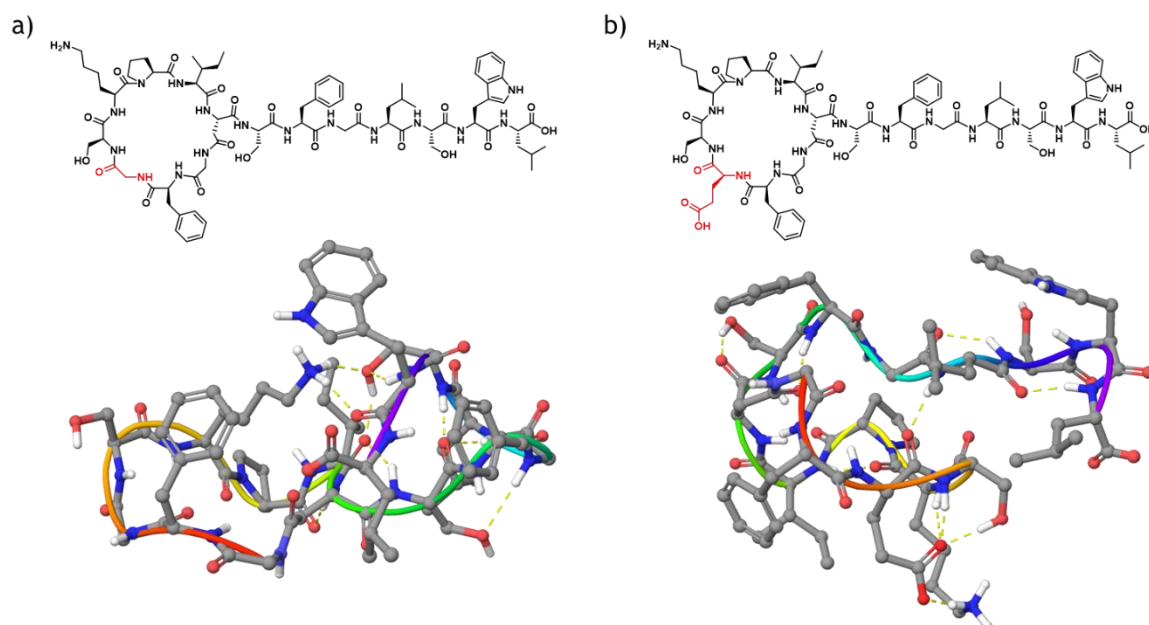


Figure 51: Lowest-energy conformations of a) synthetic sunspanin and b) branched cyclic analog. The different residue is highlighted in red. The yellow lines represent hydrogen bonds.

Considering the synthetic sunspanin, all the results for each trend pointed out to the same conclusion: it showed a branched cyclic structure. However, the branched cyclic analog had a totally different behavior, even though their sequences only differed in one residue. The *in silico* analysis showed that this difference could be explained by the different adopted conformations. The electrostatic interaction Glu3-Lys5 could be present in the gas-phase, having a stabilizing effect and as a result, a bicycle or lasso-like structure.

Regarding the IM-MS results and the analysis of the different parameters, it could be concluded that the bicycle and the branched cyclic analogs both adopted a lasso-like structure, as there was no differentiating feature between them.

NMR spectroscopy and structural analysis

Synthetic sungsanpin

Spectral validation

The structure of the synthetic sungsanpin was determined by performing exhaustive NMR analysis. For this purpose, a complete set of monodimensional and homonuclear heteronuclear bidimensional NMR experiments in pyridine-d₅ were recorded.

The NMR was performed in pyridine-d₅ in order to compare with the reported structure of the natural sungsanpin, and the same NMR parameters were used in the measurements. Spectral equivalence with the natural product was validated by comparison of monodimensional and bidimensional NMR data. The full assignment of ¹H and ¹³C signals is comprised in Table 18.

		Synthetic sungsanpin		Natural sungsanpin				Synthetic sungsanpin		Natural sungsanpin	
		δ_H	δ_C	δ_H	δ_C			δ_H	δ_C		
Gly1	NH	8.58		8.41		Ser9	NH	9.66		9.01	
	α H	3.71; 3.65	39.45	5.10; 3.80	45.2		α H	5.33	56.47	4.77	58.6
	CO		*		171.6		BH	4.47; 4.27	62.54	4.08; 3.85	62.0
Phe2	NH	10.11		11.01		Phe10	CO		*	171.0	
	α H	5.24	52.61	4.15	58.6		NH	9.91		9.16	
	BH	3.58; 3.30	38.34	3.92	35.0		α H	4.34	58.09	5.43	54.5
	H2-6	7.33	133.33	7.23	130.2		BH	3.83; 3.73	35.87	2.92; 2.69	39.0
	H3-5	7.25	138.70	7.25	129.1		H2-6	7.51	130.37	7.12	129.5
	H4	7.19	138.02	7.19	126.9		H3-5	7.32	129.81	7.15	128.8
	CO		123.23		171.5		H4	7.31	128.94	7.08	126.8
Gly3	NH	9.98		8.35		Gly11	CO		172.11	172.1	
	α H	4.54; 4.17	44.09	4.96; 4.19	46.0		NH	9.24		8.09	
	CO		171.00		169.5		α H	3.93; 4.64	43.93	5.01; 4.09	44.3
Ser4	NH	9.12		8.97		Leu12	CO		171.34	170.0	
	α H	5.27	56.27	5.36	57.4		NH	8.54		10.20	
	BH	4.51; 4.29	62.79	4.41; 4.36	64.5		α H	5.27	52.69	4.75	54.2
	CO		171.46		171.4		BH	2.20; 2.04	41.23	2.56; 1.55	39.3
Lys5	NH	9.03		9.70		γ H	1.24	28.31	1.75	15.3	
	α H	5.03	51.88	5.37	51.1	δ H	0.90	22.08	0.90	23.6	
	BH	2.24; 1.76	32.77	2.10; 1.87	31.9		0.83	22.08	0.86	20.1	
	γ H	1.91; 1.58	21.99	1.82; 1.62	26.6	CO		*	175.0		

	δ H	1.37; 1.21	29.83	1.74; 1.62	20.1	Ser13	NH	9.91		10.01	
	ϵ H	3.60; 2.67	38.48	3.14; 3.06	38.9		α H	5.50	56.87	5.37	55.2
	CO		*		173.2		BH	4.34; 4.49	63.54	4.42; 4.05	64.5
Pro6	α H	4.78	60.55	4.76	61.3		CO		*		174.5
	BH	2.06; 2.05	29.93	2.16; 2.09	30.4	Trp14	NH	9.26		8.30	
	γ H	1.75; 1.49	25.45	2.10; 1.78	26.7		α H	5.60	55.07	5.17	56.3
	δ H	3.81; 3.58	48.00	3.93; 1.62	48.4		BH	3.96; 3.68	28.67	4.09; 3.82	27.4
	CO		*		174.5		H2	7.76	123.91	8.35	126.1
Ile7	NH	9.45		8.04			H4	7.46	112.00	7.47	111.6
	α H	4.47	60.66	5.13	60.0		H5/6	7.84	119.28	7.20	121.8
	BH	1.97	35.46	2.24	43.4		H6/5	7.13	119.21	7.18	119.5
	γ H	1.61; 1.13	26.26	2.04; 1.56	25.8		H7	7.18	121.74	8.20	120.9
		0.87	15.48	1.04	16.0		Hind	11.63		11.75	
	δ H	0.58	10.81	0.99	12.9		CO		173.29		170.1
	CO		*			Leu15	NH	9.15		7.80	
Asp8	NH	9.31		9.51			α H	5.20	52.10	4.92	54.4
	α H	5.83	51.47	5.57	51.1		BH	1.97; 1.91	41.78	1.41; 1.03	43.9
	BH	3.19; 2.88	41.36	3.83; 3.69	37.5		γ H	*	*	1.50	25.2
	CO		*		171.3		δ H	0.95; 0.92	23.26	0.82; 0.73	23.2
							CO		*		179.7

Table 18: Chemical shifts of ^1H and ^{13}C (ppm) for the synthetic and natural sungsanpin in pyridine- d_5 at 298 K. (*) Missing chemical shifts.

Regarding the chemical shift values, no great difference between them was found, except for the amide values, in which more deviation was observed. The highest difference was observed for the residues Leu12 and Gly3. However, much closer chemical shifts were observed for the aromatic residues, which reveals that they are not affected by the different topology.

Comparison of the monodimensional ^1H NMR spectra (Figure 52) indicated that the synthetic sungsanpin spectrum had less resolution with wider peaks and several overlapped signals. These facts revealed that the synthetic sungsanpin showed more flexibility and consequently, a more disorganized structure than the natural sungsanpin. This evidence, in combination with the wider peaks, could indicate that multiple conformations were present.

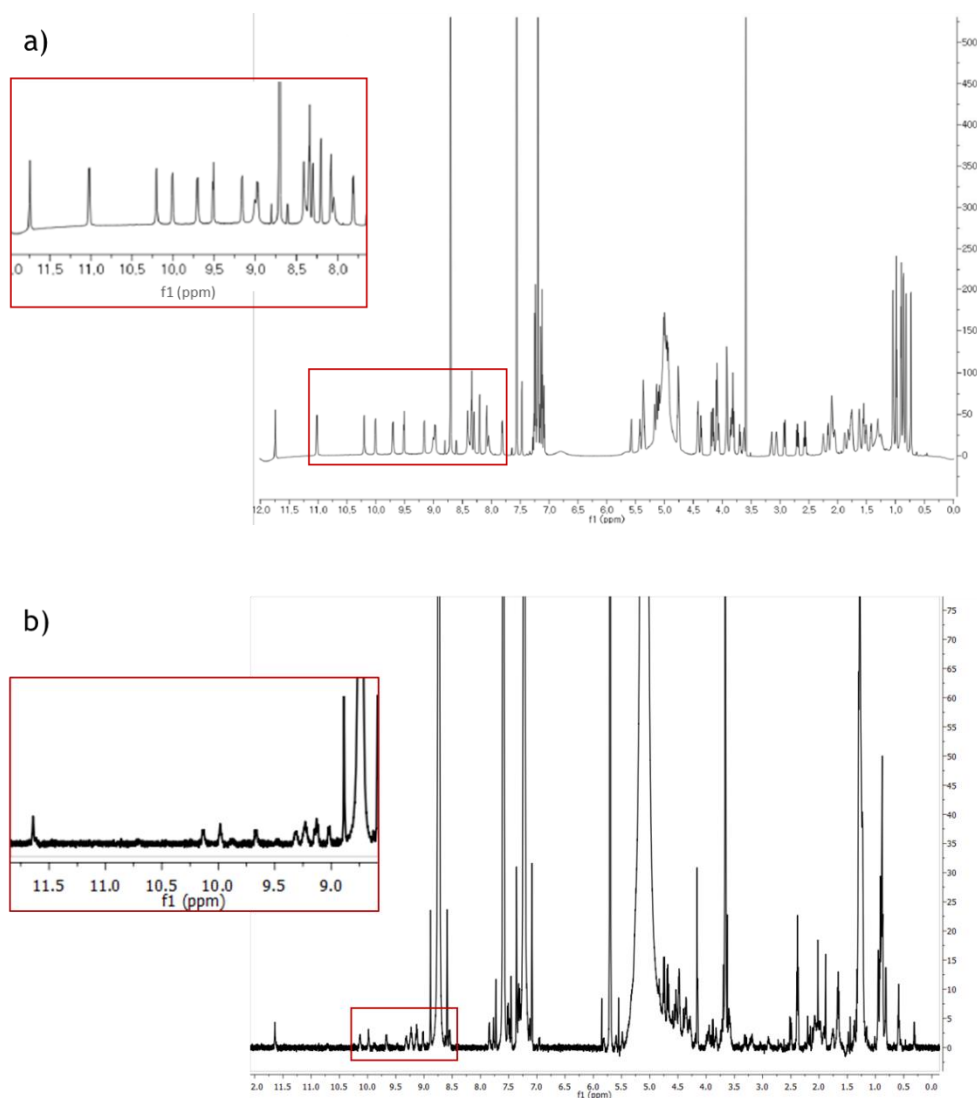


Figure 52: ¹H spectra of a) natural sungsanpin and b) synthetic sungsanpin in pyridine-*d*₅ at 298 K. In red, expanded amide region.

In peptides, the amide proton region provides useful information about the structure. Large ¹H signal dispersion being a typical feature of structured peptides. Thus, focusing in this region, some differences could be observed. For example, the natural product displayed large amide proton dispersion (7.80 - 11.01 ppm), whereas this region was less dispersed (8.54 - 10.11 ppm) for the synthetic sungsanpin, which revealed a less structured peptide. Moreover, in the aliphatic region of the methyl signals, there was a mixture of broad peaks—a typical feature of random coil structures—instead of narrow and organized peaks as expected in a lasso peptide.

Then, we proceeded with the identification of the NOE cross-peaks, in order to assess the formation of the lasso peptide. In these threaded structures, large number of long range NOEs are typically detected between the loop and the tail. In our case,

the high signal overlap, as well as the noise of solvent, made the NOEs assignment challenging. Most of the identified signals were sequential NOEs, like $\text{NH}_{(i)}\text{-H}\alpha_{(i-1)}$. One interesting NOE was found between the Lys5($\text{H}\alpha$) and Pro6($\text{H}\delta$), indicating that Pro adopted a *trans* conformation. This evidence was also confirmed by a method based on the ^{13}C chemical shift statistics (see below).²¹

Some long-range interactions were observed between Ser9(NH) and Leu12($\text{H}\beta$, $\text{H}\gamma$), indicating interaction around the loop. Furthermore, NOEs between both Gly11(NH) and Leu12(NH) with Asp8($\text{H}\alpha$, $\text{H}\beta$), the Leu12(NH) with Ser9($\text{H}\beta$) and the Asp8(NH) with Phe10($\text{H}\alpha$) were observed. All these NOE contacts indicated a strong interaction of Asp8 with the tail, providing evidence that the tail was located over the ring. Leu12 was also a residue highly involved in the structural interactions. Strong NOEs between the Trp14($\text{H}2$) and Asp8($\text{H}\alpha$) and Leu12($\text{H}\delta$) were detected, indicating proximity of them. NOEs around the residues of the ring were also observed, such as Lys5(NH) with Phe2($\text{H}\beta$) and Ile7($\text{H}\gamma$). However no long-range NOEs between the tail and the *N*-terminal macrolactam ring were detected, suggesting that no lasso structure was adopted.

There is a method that analyzes the chemical shift difference between the $^{13}\text{C}\beta$ and the $^{13}\text{C}\gamma$ of the same proline residue and allows the distinction between the *cis/trans* conformation. It is known that variations in these signals can be correlated with *cis* or *trans* conformations of the amide bond. In our case, the chemical shift differences between the $\text{C}\beta$ and $\text{C}\gamma$ atoms of the synthetic sungsanpin ($\Delta\text{C}_{\beta-\gamma} = 4.5$ ppm), indicated a *trans* conformation. This value was in agreement with the corresponding one in natural sungsanpin ($\Delta\text{C}_{\beta-\gamma} = 3.7$ ppm).

Structure calculation

In order to assess the 3D structure of the synthetic sungsanpin, conformational sampling using the Macrocyclic Conformational Sampling tool²⁰ was performed by applying the NMR-derived distance restraints (see Materials and Methods for more details), as the long-range NOE detected were not sufficient to elucidate the exact 3D structure. More restrains would be needed, but the low spectra resolution made their assignment difficult.

The lowest energy structure derived from the NMR-restraints, shown in Figure 53, was in agreement with our NMR predictions, in which the lasso structure was not adopted, but the loop was folded over the ring. The backbone of this structure was superposed with the lowest energy structure without restraints (Figure 51a), showing a low RMSD value of 4.2 Å. The similarity between both structures might be due to the few restraints obtained from the NOESY spectrum.

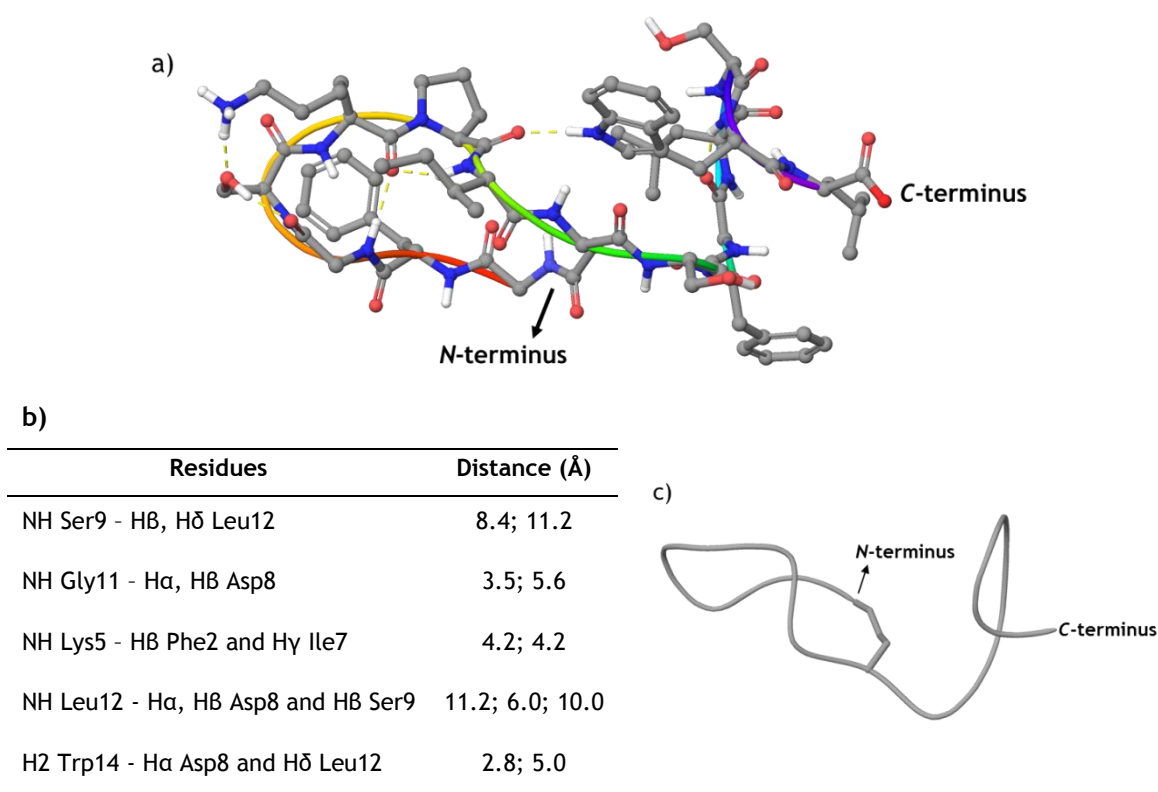


Figure 53: a) NMR-derived structure of the synthetic sungsanpin. Hydrogen bonds are shown in yellow. b) Long-range NOE restraints. c) Ribbon representation of the synthetic sungsanpin.

The general broadness of the spectral lines and the poor dispersion of amide signals led us to conclude that the synthetic sungsanpin did not show a lasso structure. Moreover, the few NOE cross-peaks detected and the lack of evidence for interactions between the C-terminal tail and the ring evidenced that the peptide did not adopt the lasso structure. Proteolytic assays with carboxypeptidase Y also revealed that the synthetic sungsanpin did not show a lasso structure, due to its rapid degradation. These results clearly point to a branched cyclic structure, which was also confirmed by the NMR-derived structure. The long-range NOEs between the ring and the loop evidenced that the linear part was folded over the ring, but with no threading.

Bicyclic sungsanpin analog A

Afterwards, the structure of the bicyclic sungsanpin analog was determined by a complete set of monodimensional and homonuclear heteronuclear bidimensional NMR experiments. DMSO- d_6 was chosen as solvent due to its high solubilizing capacity and its potential to stabilize conformations (in DMSO the side chains of the amino acids take on a preferred spatial orientation, fixing the peptide conformation).

The ^1H NMR was acquired at 298 K, showing wide peaks and low resolution. Added to this, two peaks of the NH indole of Trp14 were detected even though only one Trp residue was present in the peptide sequence. After an extensive NMR analysis, duplicity in residues Trp14, Pro6 and Leu15 was also detected. One of these conformations was more intense and abundant than the other. Xaa-Pro bond conformation was determined using a previously described method based on the ^{13}C chemical shift statistics.²¹ The chemical shift differences between the C_β and C_γ atoms in the most abundant Pro6 residue ($\Delta\text{C}_{\beta-\gamma} = 3.81$ ppm) and the minor Pro6 residue ($\Delta\text{C}_{\beta-\gamma} = 9.72$ ppm) indicated that, in both cases, the Xaa-Pro bond predominantly existed in the *trans* conformation. The exchange frequency between these two conformations was very slow, due to the high energy bandgap between them, resulting in double signal detection.

Due to the constrained and rigid structure of the bicycle, clear and organized NMR spectra were expected. Instead, the resulted spectra indicated no structural organization. The duplicity of the C-terminal residues (Trp14 and Leu15) suggested a high flexibility of this part, which is not constrained by the bicyclic structure. Regarding the *cis/trans* isomerization, the rate of interconversion is also solvent-dependent, being higher in nonpolar solvents.²² However, the *trans* configuration is more frequent for peptides, as the side chains suffer from a higher steric hindrance in the *cis* configuration.²³

The peak broadness observed could be related to the presence of, at least, two conformations. Thus, to explore the presence of these conformations, different ^1H NMR spectra were acquired at higher temperature (303 and 308 K) (Figure 54). Theoretically, the increase of temperature produces a faster conformational equilibrium, as long as the energy bandgap between them is small, thus producing a single peak in the spectrum.

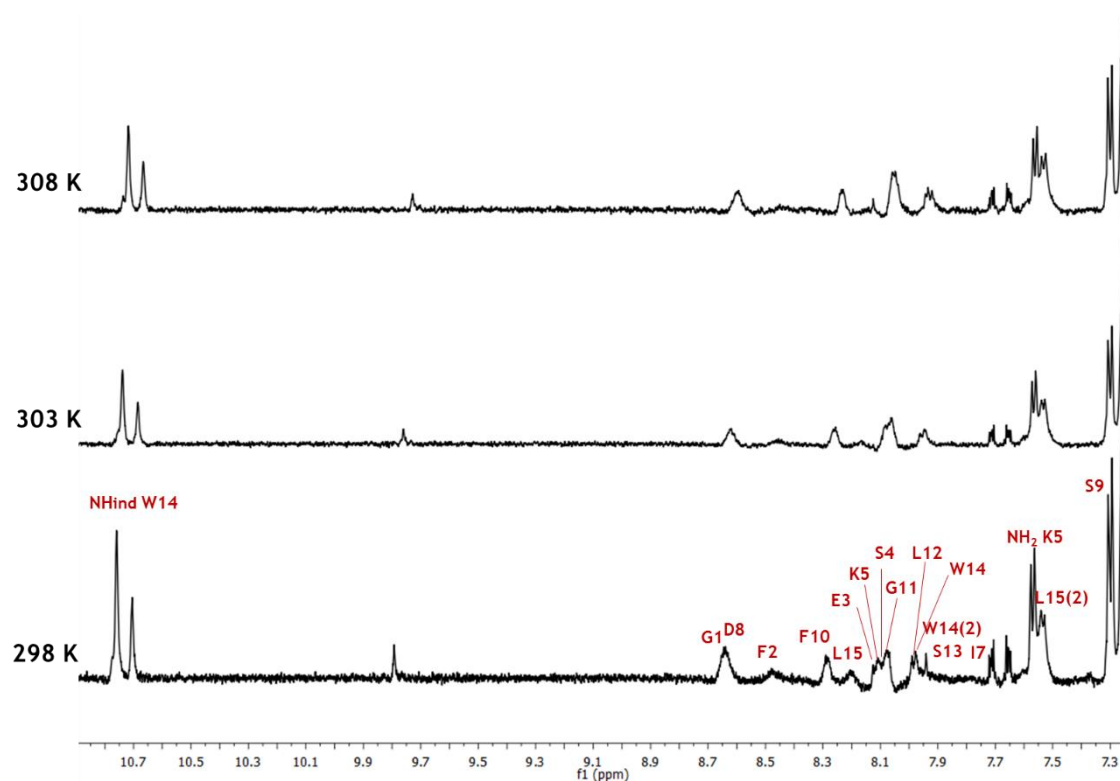


Figure 54: Expanded amide region of ^1H spectrum of the bicyclic sunsaanpin analog A in DMSO-d_6 at different temperatures. (2) Residues from the second conformation.

However, for this peptide the same spectra were obtained with the increase of temperature. The chemical shifts of all peaks remained almost constant, which could be explained by small temperature coefficients. The same peak shape was obtained at different temperatures, either indicating that the energy bandgap of these conformations was too high—and the equilibrium was still present, or that the temperature increase was not sufficient.

In the alpha proton region of the spectrum, a better resolution was obtained at 298 K than at higher temperatures (Figure 55). As a result, the 2D NMR experiments such as TOCSY, DQF-COSY, NOESY and ^1H - ^{13}C HSQC were acquired at 298 K. The complete ^1H and ^{13}C assignment in DMSO-d_6 at 298 K is detailed in the experimental section.

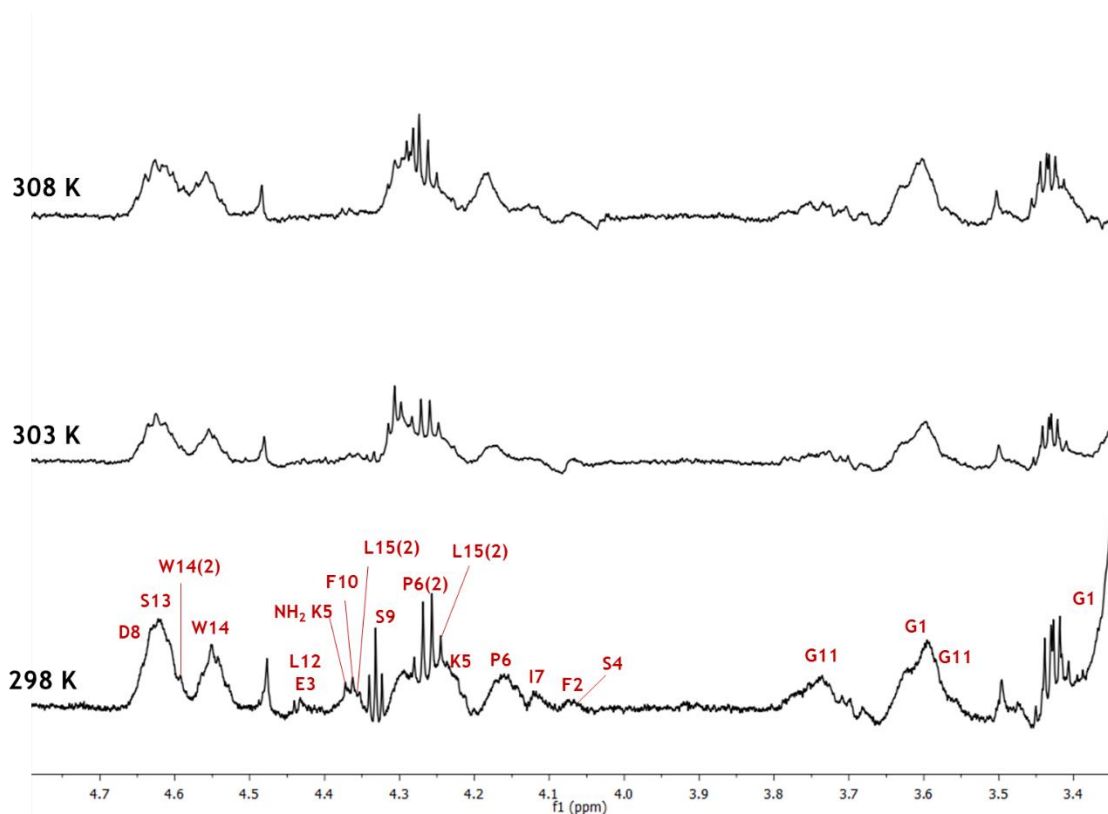


Figure 55: Expanded alpha region of ^1H spectrum of the bicyclic sungsanpin analog A in DMSO-d_6 at different temperatures. (2) Residues from the second conformation.

Due to the overlapping and low dispersion in some of the signals, a few long-range NOE correlations could be detected. To give an idea of the complex spectra, only between 8.11 and 8.08 ppm four residues were identified, thus making the NOEs assignment futile. The few NOEs that could be assigned (Figure 56) were not enough to predict the 3D structure.

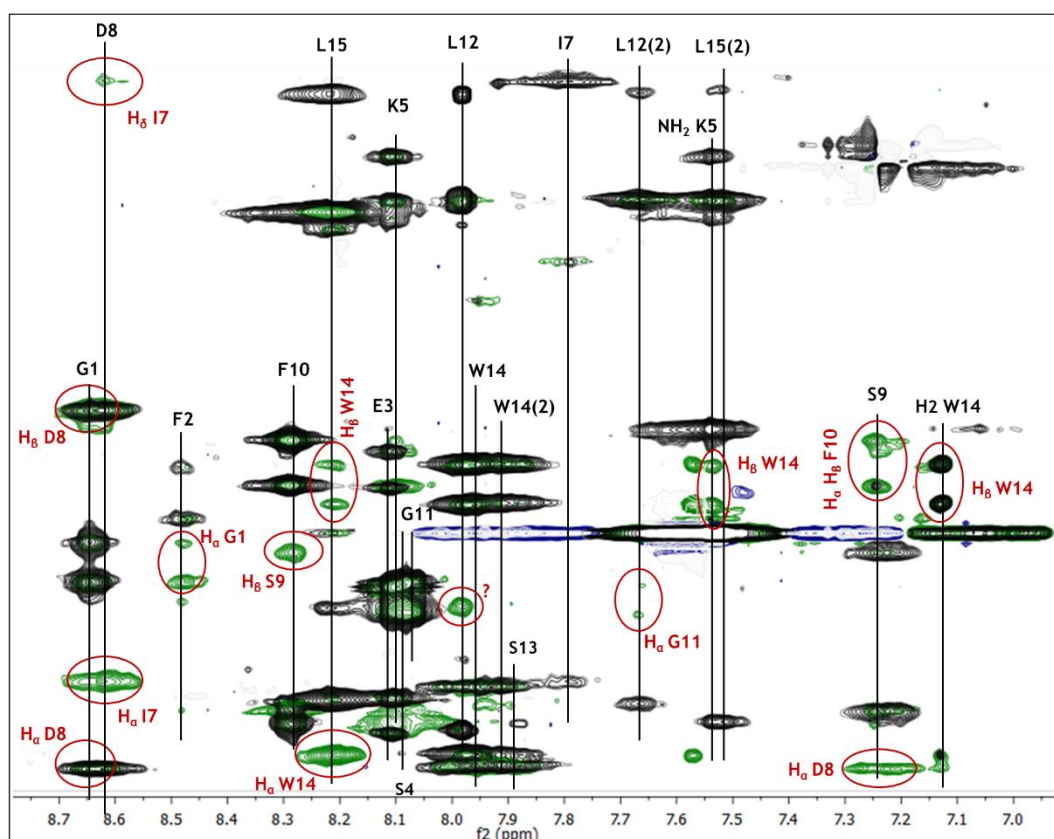


Figure 56: Superposition of expanded regions of the 2D TOCSY (black) and NOESY (green) spectra of the bicyclic sungsanpin analog A in DMSO- d_6 . In red, the NOE cross-peaks.

In order to obtain better spectra resolution, an alternative possibility would be to decrease the temperature to “freeze” the conformational equilibrium, producing a signal splitting. This was not a feasible option due to the high melting temperature of DMSO. Other solvents with lower melting temperature than DMSO- d_6 were tested, including methanol- d_3 , acetone- d_6 and acetonitrile- d_3 . Chloroform- d_3 was not assayed because a low peptide solubility was expected. Also, being a nonpolar solvent, more peptide conformations were expected in this solvent. The complete solubility of the bicycle was only achieved in acetonitrile- d_3 with H_2O . The ratio CD_3CN-H_2O was optimized to 1:5. Therefore, a new set of 1D and 2D NMR experiments were acquired in CD_3CN-H_2O (1:5) at 298, 288 and 278 K (Figure 57).

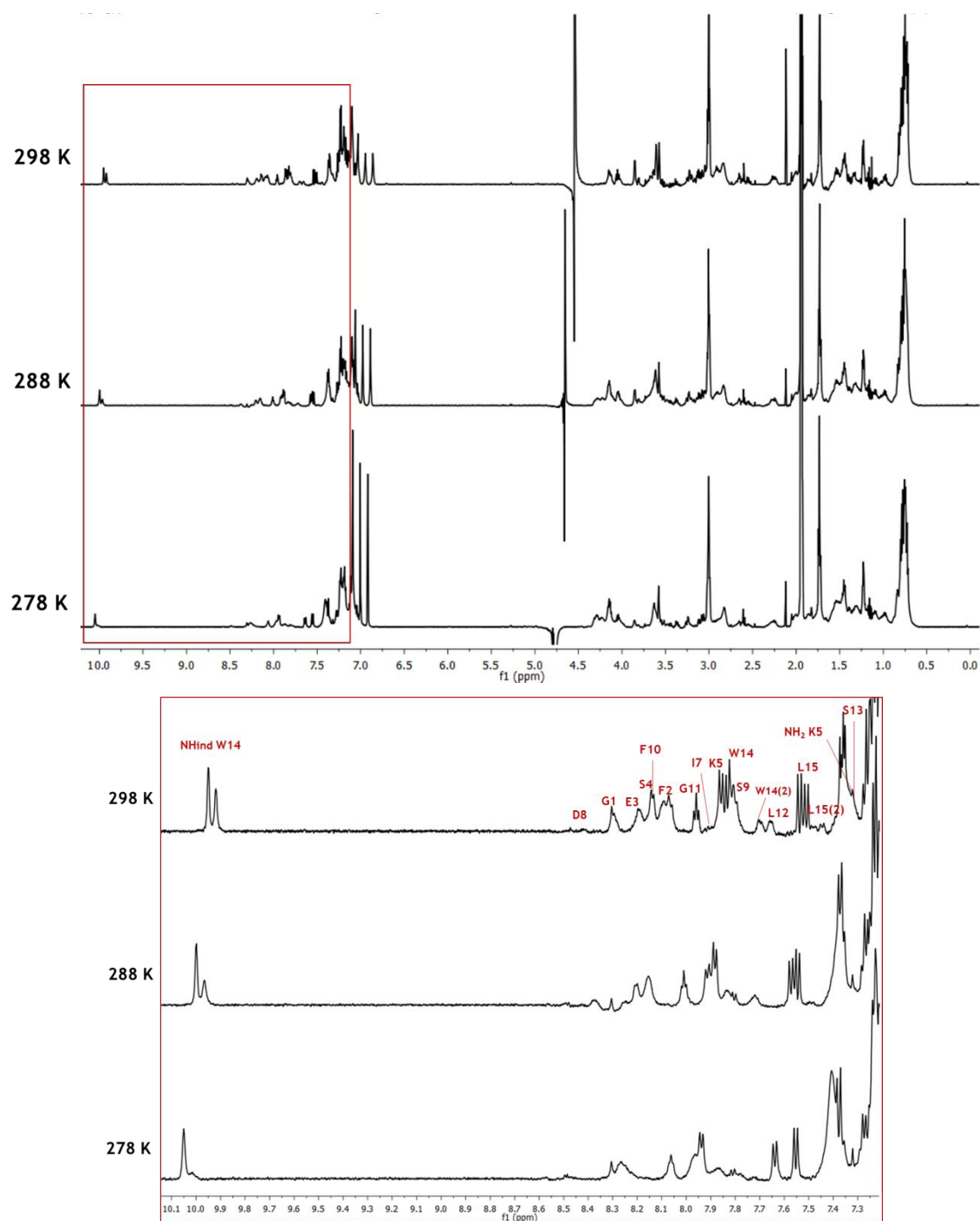


Figure 57: ^1H NMR spectra of the bicyclic sungsanpin analog A in $\text{CD}_3\text{CN}-\text{H}_2\text{O}$ (1:5) at different temperatures. Below) Expanded amide region.

Considerable differences in the chemical shift of the amide protons were observed when the temperature was decreased. For some proton signals, such as Glu3, Ser4 or Phe10, the peak completely disappeared as the temperature went down. Nevertheless, in the case of Leu12 and Leu15, two well-resolved peaks at different chemical shift appeared at 278 K. The change in the chemical shift evidenced that all the amide protons were solvent accessible, and they were not involved in any

hydrogen bond or hydrophobic interaction. By comparing it with the spectra in DMSO- d_6 at high temperatures (Figure 54), it can be noticed that the amide protons were less affected than in Figure 57. This could be a consequence of the solvent (ACN- H_2O has a weaker hydrogen-bond acceptor character than DMSO).

In general, the decrease in temperature caused a loss of spectral resolution with a widening of peaks. For example, the second peak of the NH indole became wider and decreased in intensity, but it did not disappear. However, the spectra obtained were different from expected, maybe due to the fact that the protons were still in a transition process and lower temperature would be required to reach the optimal conformation. Due to the cryotube limitation, the temperature could not be lowered below 278 K. The peaks in the aromatic region remained constant, whereas the amide and alpha protons region showed wide peaks and very low resolution. Compared to the spectrum at 298 K, low dispersion of the amide protons and loss of some signals is observed at 278 K (Figure 58).

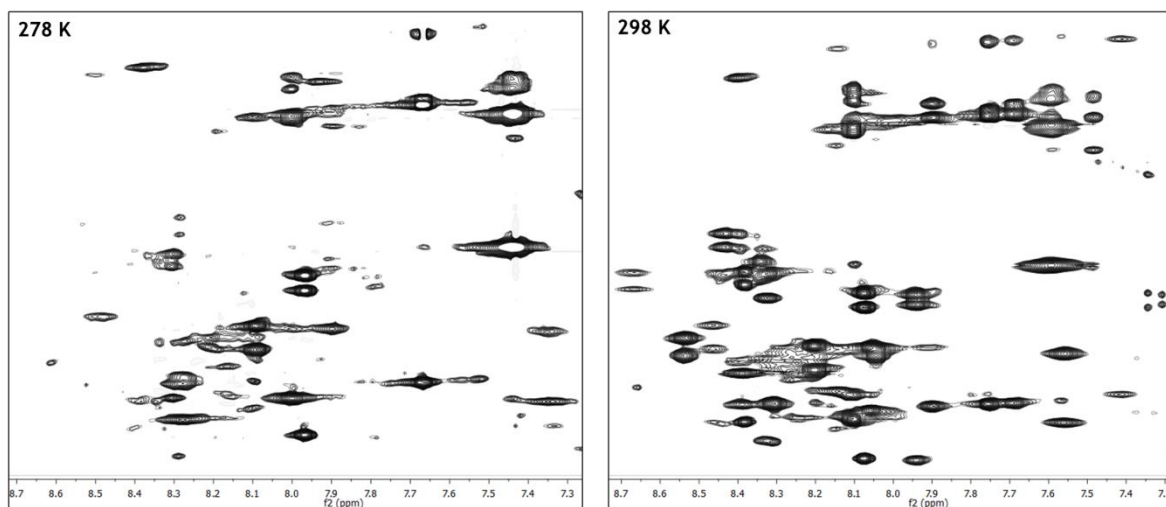


Figure 58: Expanded amide region of the 2D TOCSY spectra of the bicyclic analog A in CD_3CN-H_2O (1:5) at 278 and 298 K.

As we have seen, decreasing the temperature was not a suitable alternative to gain resolution and to detect long-range NOEs. However, with this solvent mixture (CD_3CN-H_2O 1:5), the proton dispersion at 298 K was better than in DMSO- d_6 . Thus, we decided to further proceed with the bicyclic NMR characterization. The complete 1H assignment in CD_3CN-H_2O (1:5) is detailed in the experimental section. Compared to DMSO, the amide dispersion was much better, and we were able to detect some long-range NOEs. The observation again of two peaks for the NH indole of Trp14, indicated that the two conformations were solvent independent (Figure 59).

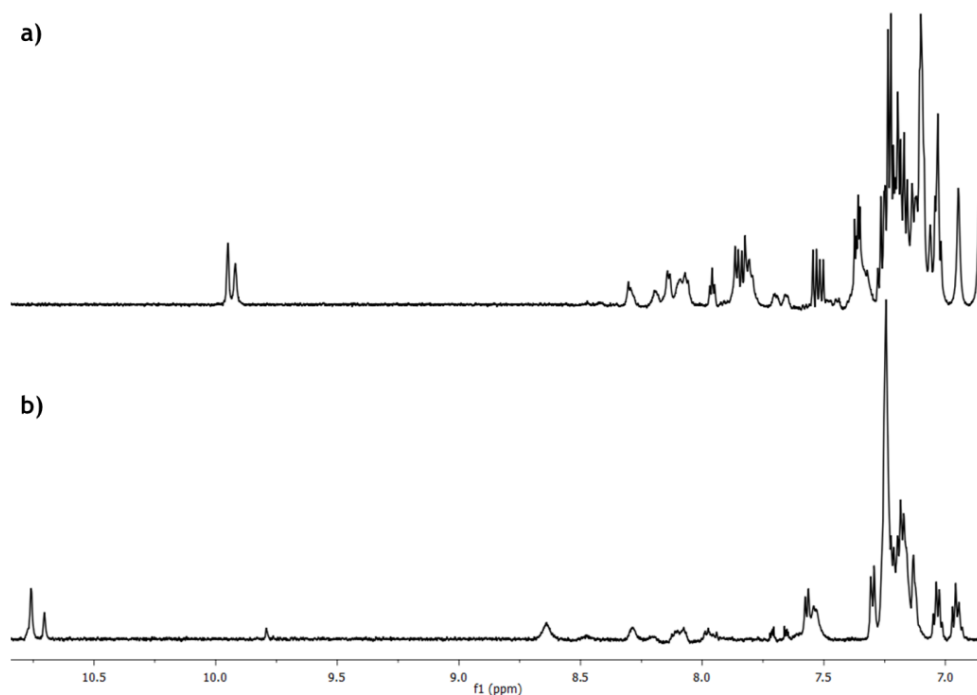


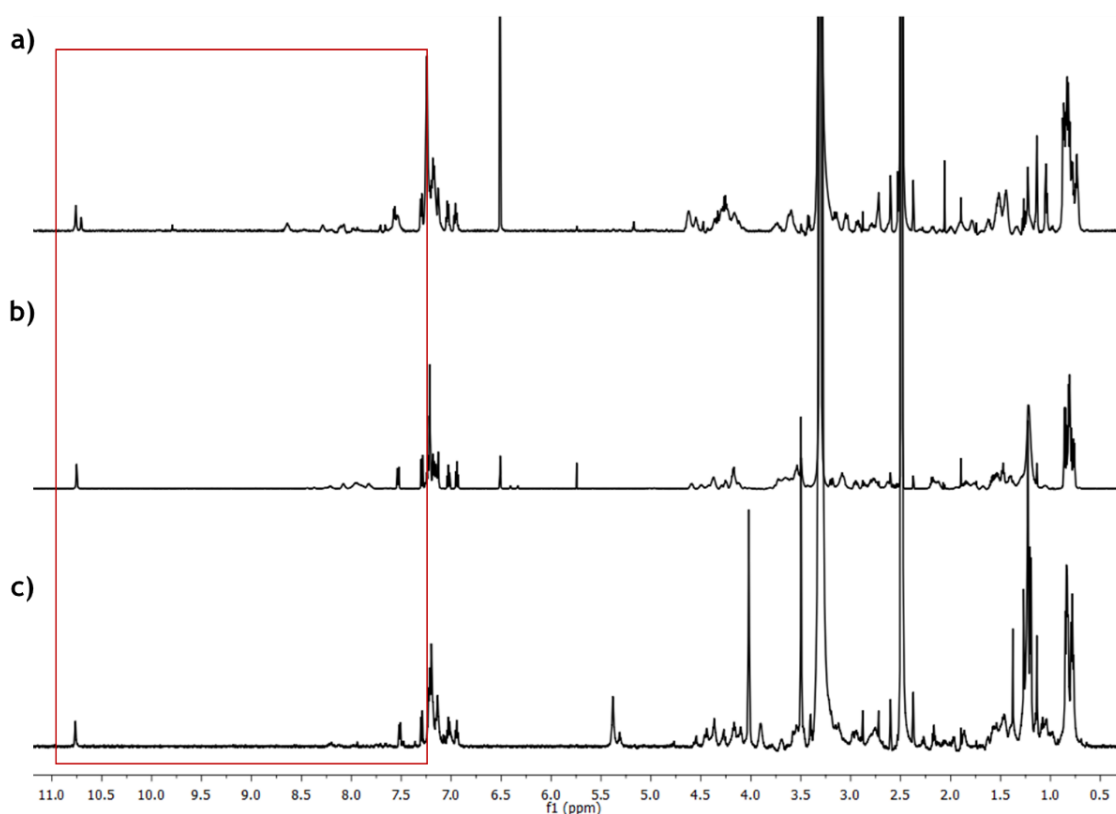
Figure 59: Expanded amide region of ^1H NMR spectra at 298 K in a) $\text{CD}_3\text{CN-H}_2\text{O}$ (1:5) and b) DMSO-d_6 .

Three interesting NOEs indicated interaction between the ring and the loop: the Phe2(NH) and Leu12(H β , H γ , H δ), the Lys5(NH $_2$) with Leu12(H δ) and Pro6(H α) with Leu12(H γ , H δ). As noticed, Leu12, which was located in the peptide loop, was involved in several interactions with the ring. This fact supported the hypothesis that the loop was closer to the ring. This hypothesis was also supported by the correlation between the Trp14(H2) and Ser9(H β), Phe2(H α , H β) and Leu12(H δ). Strong interactions inside the ring were observed in the NOE cross-peak between the NH of Gly1(NH) and Glu3(H β , H γ).

Regarding all these studies, it could be remarked that the spectra for the bicyclic analog A did not show a good resolution and the assignment of the ^1H and ^{13}C signals was very challenging, thus hampering the 3D structure prediction. Due to the low proton dispersion, NOE cross-peaks were difficult to detect. However, by changing the solvent to a 1:5 mixture of $\text{H}_2\text{O/ACN}$, a better dispersion was obtained and this allowed to measure some long-range NOEs. Contrary to what we expected, the rigid structure of the bicycle was not the important feature to provide an organized and structured spectrum. In fact, the lack of hydrophobic or electrostatic interactions that stabilize the structure was the main problem.

Hydrolysis of the ester bond

The main purpose of using an ester bond linkage when performing the bicycle synthesis, was to then hydrolyze it and obtain the lasso peptide. It is after the hydrolysis of the ester bond when the supposed lasso structure should keep intact. In lasso peptides, when the tail is trapped inside the ring it is impossible to unthread it, except in some cases when a large source of heat is applied.^{2,3} Bulky amino acids such as Phe, Trp, Leu or Tyr act as steric plugs to prevent slippage. Considering this, the pure unthreaded peptide derived from the strategy A (Figure 37a) was characterized by a set of 1D and 2D NMR experiments in DMSO-d₆ at 298 K and compared with the bicyclic peptide (Figure 60).



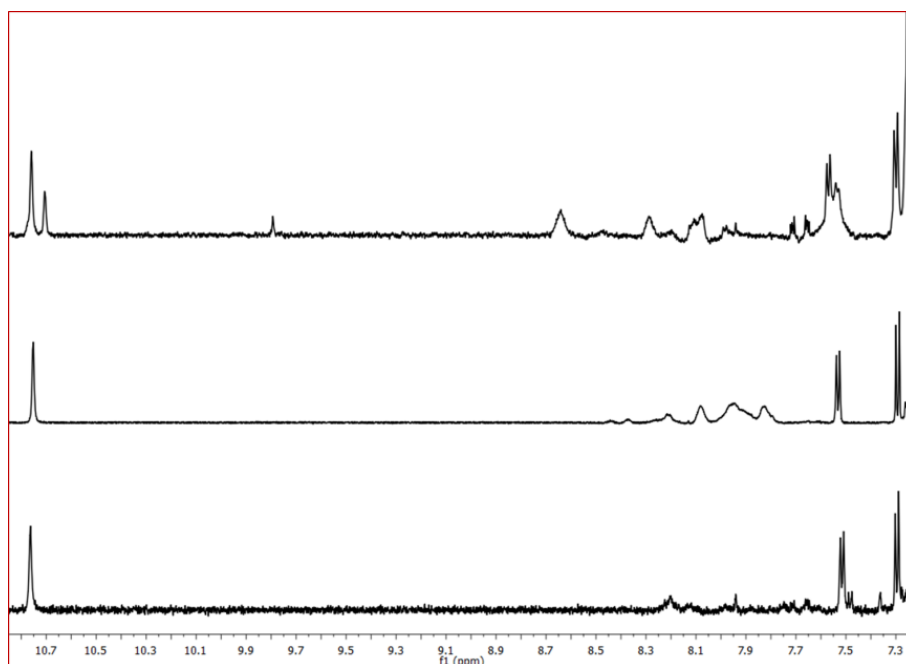


Figure 60: ^1H spectra of a) bicyclic sungsanpin analog A, b) synthetic branched cyclic analog and c) unthreaded peptide in DMSO-d_6 at 298 K. In red, expanded amide region.

The first difference observed in the ^1H spectra of the unthreaded peptide was the disappearance of the NH indole second peak. This fact suggested that the low conversion rate between conformations disappeared resulting in a fast exchange frequency and consequently, one signal detection. The amide region showed less dispersion and more signal overlapping. However, the aromatic and aliphatic region displayed more defined peaks.

On the other hand, the unthreaded peptide displayed a closer spectrum to the synthetic branched cyclic analog (Figure 60b). Nevertheless, 2D TOCSY spectra revealed more peaks in the alpha proton region for the unthreaded peptide than for the bicycle, indicating that more conformations—due to the increase of the flexibility—were present (Figure 61a). The bicyclic peptide showed a more organized spectrum, while the unthreaded peptide showed a large number of peaks, more than the number of alpha proton in the peptide sequence. Regarding the amide region, it was clearly visible the narrow signal dispersion and the overlap in the case of the unthreaded peptide (Figure 61b, c).

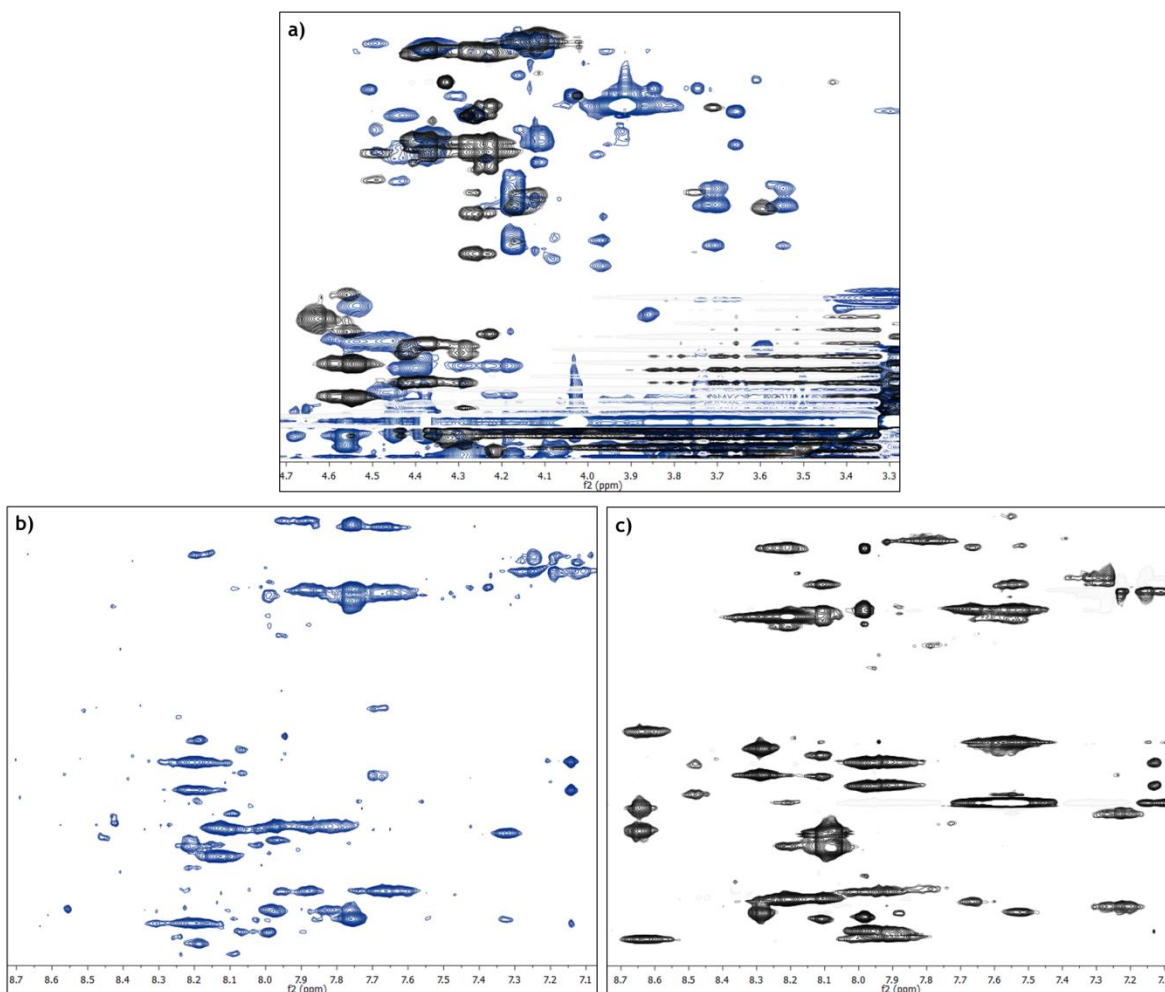


Figure 61: Expanded regions of the 2D TOCSY spectra in DMSO- d_6 at 298 K. a) Superposition of the alpha proton region. Unthreaded peptide in blue and bicyclic peptide in black. b) Amide region of the unthreaded peptide and c) amide region of the bicyclic analog A.

These evidences demonstrated that the ester bond was responsible for maintaining the rigidity in the structure. Once this linkage was not present, the peptide lost the compact topology and acquired a more flexible cyclic structure. This NMR characterization is in agreement with the stability studies, in which the unthreaded peptide was completely degraded after thermal and proteolytic assay (Figure 39).

Spectral comparison with natural lasso peptides

Lasso peptides have a defined structure, independently of their sequence. Based on the NMR studies of eight lasso peptides, several common features that are characteristic of the lasso structure were found.²⁴

- 1) Large ^1H signal dispersion in the amide proton region.
- 2) Large and irregular chemical shift deviation in αH compared to those of corresponding sequence in random coil state.
- 3) The ^1H signal of amide proton of Gly1/Cys1 is clearly visible and shows long-range NOE contacts to side-chain protons of Glu8/Asp9, which serves as evidence for a side-chain-to-backbone covalent linkage.
- 4) Rich in long-range NOEs between the C-terminal tail and the N-terminal lactam ring, which verifies tail- and-ring threading.
- 5) Slow H/D exchange for amide protons of those residues involved in the threading and the short β -sheets.
- 6) Low temperature coefficients of these amide protons, which were almost zero for the threading residues.
- 7) The structural folding is solvent independent, titration of mixed solvent does not change the chemical shifts of αH .
- 8) The corresponding synthetic sequences (linear, head-to-tail cyclic and branched cyclic) lack features 1-7, showing severe signal overlap in ^1H spectra.

The NMR results from the bicyclic analog A were compared with the features of lasso peptide structure based on the previous criteria. In the following section, each criterion is discussed in detail.

1) ^1H signal dispersion in the amide proton region

Generally, it is vital for successful and reliable structure elucidation that the backbone NH-signals show broad chemical shift dispersion instead of being clustered in a narrow region. This not only significantly facilitates the interpretation of the spectrum but is also a sign for rigid and defined structures, while clustering often corresponds to branched cyclic topologies.²

Regarding this previous remark, exhaustive analysis of the fingerprint region revealed poor signal dispersion compared to lasso peptides. This fact demonstrated that the structure was not well-organized despite the bicycle's presence. This lack of organization was observed both for the signal assignments and the NOE correlations, due to the signal overlap. In this regard, the chemical shift dispersion ($\Delta\delta$) of the bicyclic peptide was very low compared to reported lasso peptides (Figure 62). For

example, the lasso fold of BI-32169 is stabilized by one disulfide bond, forming a bicyclic structure. Even having this second covalent bond, the $\Delta\delta$ of the BI-32169 was higher than for the synthetic bicycle A, revealing that this second covalent bond was not the main responsible for providing the structural organization (Figure 63).

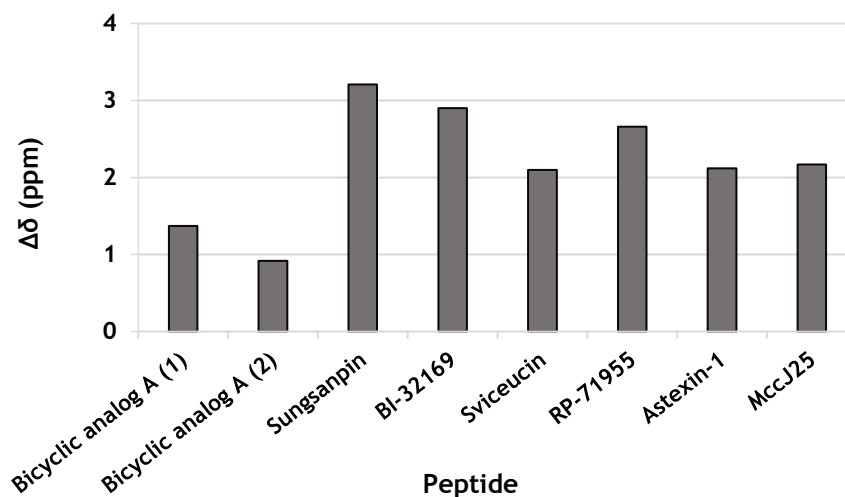


Figure 62: NH chemical shift dispersion ($\Delta\delta$) of the bicyclic analog A in (1) DMSO- d_6 , (2) CD_3CN-H_2O (1:5) and different lasso peptides from class I; sviveucin¹⁴ and RP 71955,²⁵ class II; sungsanpin,²⁶ astexin-1² and MccJ25²⁷ and class III; BI-32169.¹³

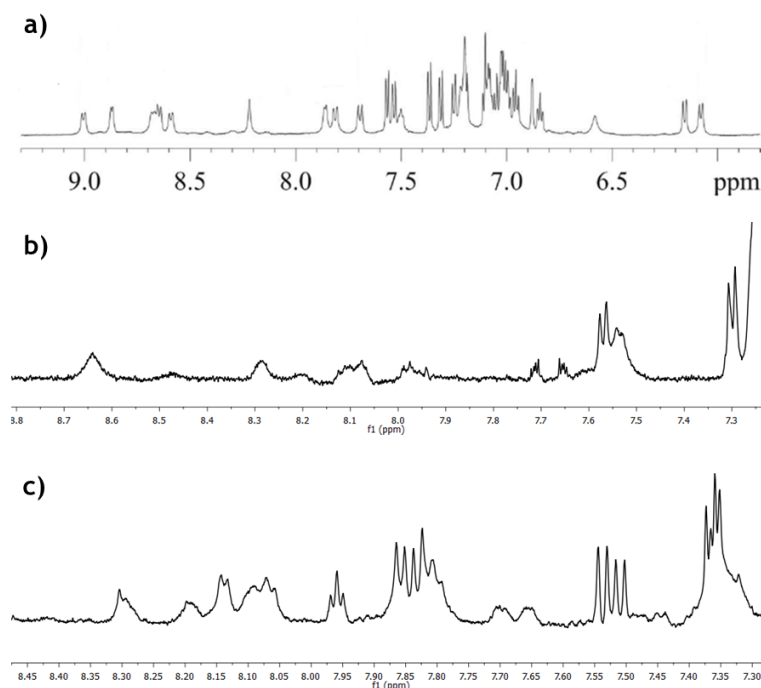


Figure 63: Expanded amide region from 1H spectra. a) BI-32169 in DMSO- d_6 at 313 K,¹³ b) bicyclic analog A in DMSO- d_6 at 298 K and c) bicyclic analog A in CD_3CN-H_2O (1:5) at 298 K.

2) Chemical shift deviation in α H compared to those of the corresponding random coil sequences

Large and irregular chemical shift deviations have been observed in lasso peptides. This feature indicates a defined and distant structure from random coil, which is in agreement with the first criterion. These values provide information about the secondary structure of the peptide. The tendency of most lasso peptides is to adopt short β -sheets between the threading moieties and the ring. This topology is also confirmed by the presence of long-range NOEs between these residues.

As it can be noticed in Figure 64, the chemical shift deviation values of both $H\alpha$ and $C\alpha$ did not show a clear tendency, as their magnitude was very small. Added to this, the observed values were different for those reported for α -helix (< -0.4 and > 2 ppm, for proton and carbon respectively) or β -sheet (> 0.4 and < 2 ppm, for proton and carbon respectively). It seemed that the negative values for α H and positive for α C for the Gly1-Ser4 segment could indicate an α -helical fold, but this trend did not extend to more consecutive residues. The α -helix shows a characteristic pattern of consecutive medium-to-strong NOE connectivities (Figure 65). Regarding this, we tried to identify NOE cross-peaks that allowed us to justify the presence of the α -helix. However, we were not able to detect this pattern. This evidence could indicate the presence of α -turns, which is very common in peptides and take place between amino acids 3-4 residues apart.

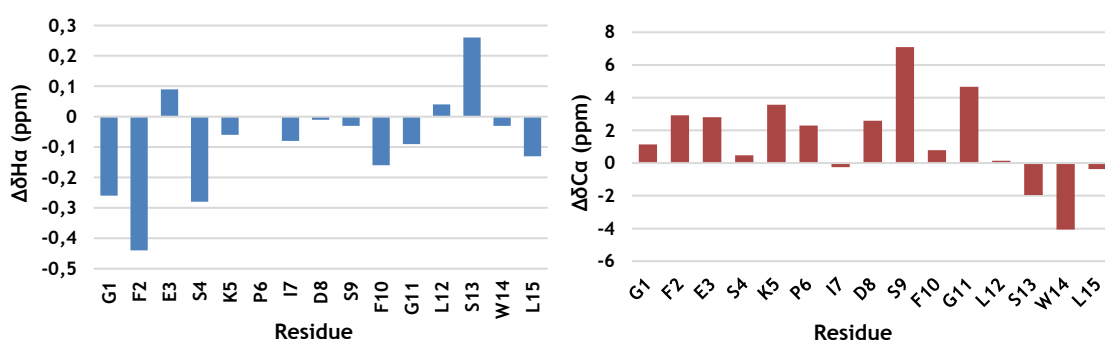


Figure 64: Histograms of chemical shift deviations from random coil, for $\Delta\delta H\alpha$ (blue) and $\Delta\delta C\alpha$ (red) in bicyclic analog A in $DMSO-d_6$. Random coil values were extracted from Wüthrich research.^{28,29}

Certainly, the NMR data analysis that we performed allowed us to discard the presence of β -sheet structure. This structure is more typical for proteins or longer

peptides. However, it is also a very common motif in lasso peptides (see below, criterion 5).

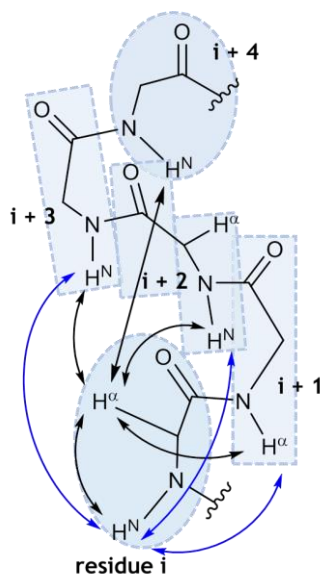


Figure 65: α -helix NOE pattern and summary of NOE distances.

3) NOE contact between the N-terminal residue and the side chain protons of the residue closing the ring

NMR characterization of lasso peptides shows a clear long-range NOE cross-peak between the NH of the N-terminal amino acid, always Gly1 or Cys1, and the side-chain protons of Glu or Asp, which are the amino acids that close the macrolactam ring. This fact evidences the presence of a backbone covalent linkage. It is important to remark that this criterion is very specific for lasso peptides because of their very well-defined structure that makes this NOE contact always visible. In other peptides this trend is not diagnostic, as its detection depends on the structure and the solvent.

In particular, for the bicyclic analog A, the NOE cross-peak between the NH of Gly1 and the side-chain of Asp8 could be observed in DMSO- d_6 , while in CD $_3$ CN-H $_2$ O (1:5) was not visible (see Figure 66). This observation does not imply that the bond did not exist, but simply that the NOE was not visible (the visibility depends on the peptide structure in solution). Instead, the presence of the macrolactam bond was previously identified by mass spectrometry analysis.

The same happened with the synthetic sungsanpin, in which this NOE cross-peak was missing. However, the mass fragmentation study clearly showed the presence of the covalent linkage Gly1-Asp8, showing only fragmentation around the linear part.

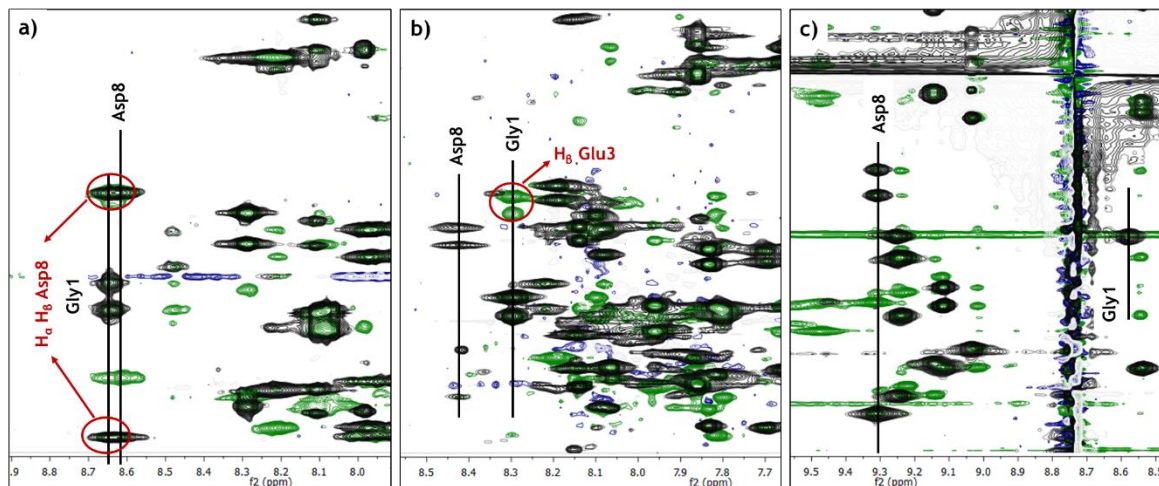


Figure 66: Superposition of expanded regions of the 2D TOCSY (black) and NOESY (green) spectra of a) bicyclic analog in DMSO- d_6 , b) bicyclic analog A in CD_3CN-H_2O (1:5) and c) synthetic sungsanpin in pyridine- d_5 . In red, the NOE cross-peaks of the NH of Gly1.

4) Long-range NOEs between the C-terminal tail and the N-terminal lactam ring

Due to the threaded tail, lasso peptides are rich in long-range NOEs between their C-terminal tail and the macrolactam ring. These contacts confirm the tail- and ring-threading and allow the 3D structural elucidation.

Regarding the case of the bicyclic analog A, most of the NOEs were sequential and few of them were located between the tail and the ring. To detect long-range NOEs, more spectrum definition and proton dispersion would be required. Therefore, it could not be guaranteed that the tail was threaded through the ring only with the detected NOE cross-peaks (detailed in Table 19).

NOE cross-peaks
NH Phe2 - H β , H γ , H δ Leu12
H α Pro6 - H γ , H δ Leu12
NH $_2$ Lys5 - H δ Leu12
H $_2$ Trp14 - H β Ser9; H α , H β Phe2; H γ Leu12

Table 19: Long-range NOEs from the bicyclic analog A in CD_3CN-H_2O (1:5) at 298 K.

5) Proton/deuterium (H/D) exchange for amide protons of those residues involved in the threading and the short β -sheets

For lasso peptides low H/D exchanges have been observed for the amide protons because most of them are involved in β -sheets. This is specially the case for the amino acids located within the amide ring, which are more hindered from the solvent. These small β -sheets normally comprise four residues, two consecutive residues located within the macrolactam ring and two residues from the ring (Figure 67).^{9,30,31}

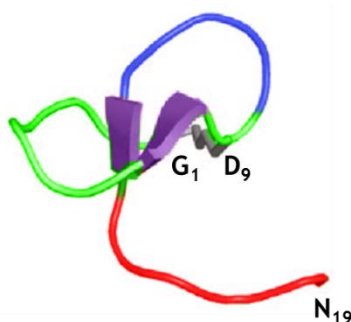


Figure 67: Ribbon representation of the 3D structure of capistruin (BMRB code: 20014). The macrocyclic ring (green), the loop (blue), the tail (in red) and the two fragments of antiparallel β -sheets (purple). (Adapted)³¹

We did not perform H/D exchange studies because taking the previous experiments and results into account, we thought that it did not add anything relevant to the comparison. Observing the low peak dispersion for this region of the bicyclic peptide, a fast H/D exchange would be expected.

6) Temperature coefficients of the amide protons with slow H/D exchange

This feature is directly related with criterion 5, in which low temperature coefficients are related with slow H/D exchange. These low values are associated with more hindered protons or protons which are involved in hydrogen bonds. For this reason, lasso peptides show a low temperature coefficient or almost zero for the threading residues.

Concerning the bicyclic analog A, when the temperature was increased, the chemical shifts almost remained constant, thus little changes were detected (Figure 68). The only residues that remained totally constant were the NH_2 of Lys5 and the Leu15 from the second conformation. More noticeable changes were detected in the

peak shapes. With the increase of temperature, the peaks became wider, yielding less resolution. We could not assure that this behavior could be related to a low temperature coefficient due to the lack of ^1H resolution.

On the other hand, when the temperature was decreased, significant changes on the chemical shifts were detected, indicating a higher temperature coefficient with DMSO-d_6 than with $\text{CD}_3\text{CN-H}_2\text{O}$ (1:5). This fact evidenced that the protons were solvent accessible and depending on the solvent properties (as hydrogen-bond acceptor) their behavior was different.

Another argument that could justify the low or even zero temperature coefficients was the poor solubility of the bicyclic analog A in water and methanol. This could indicate a compact structure with the amide protons buried inside the structure.

For example, in the ^1H spectrum of BI-32169, the lack of chemical shifts change at different temperatures is related with a low or even zero temperature coefficient (Figure 68a). In this case, the definition and the dispersion of the amide proton signals remained intact, and the best spectrum resolution was achieved at 313 K.

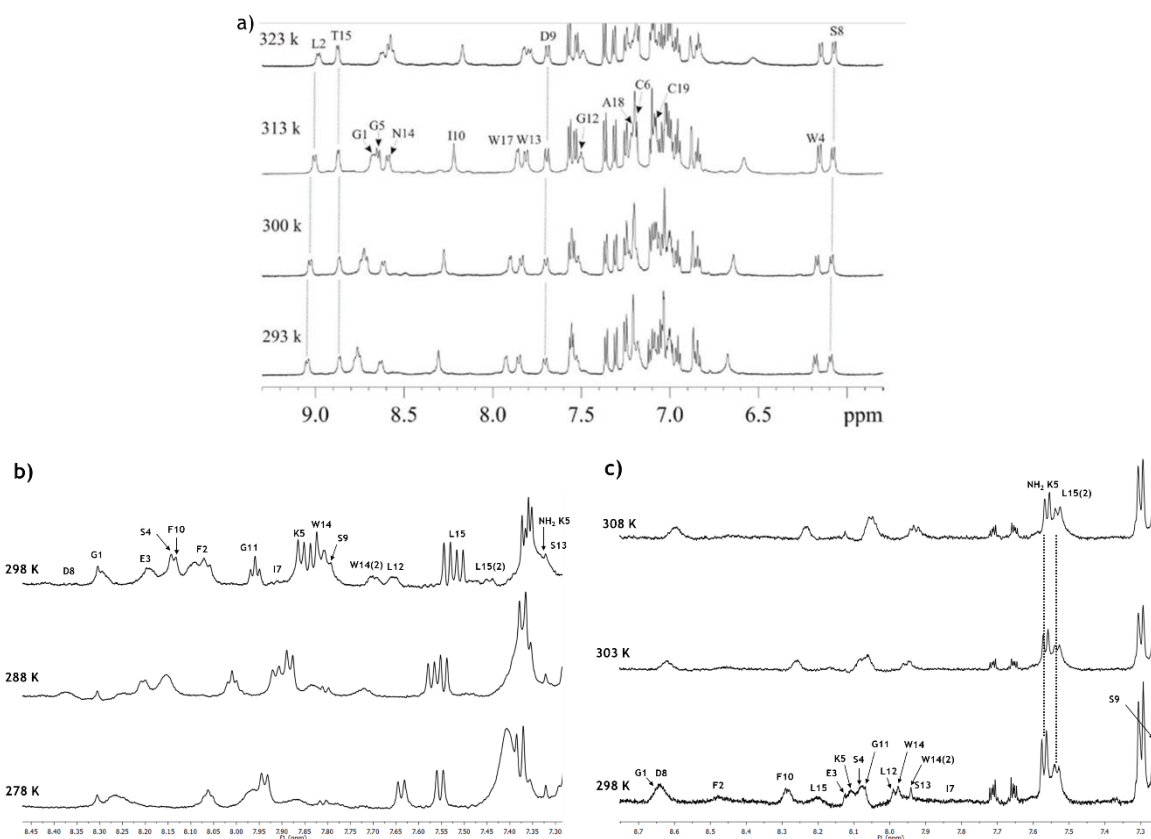


Figure 68: Expanded amide region of the ^1H spectra at different temperatures. a) BI-32169 in DMSO-d_6 , b) bicyclic analog A in $\text{CD}_3\text{CN-H}_2\text{O}$ (1:5) and c) bicyclic analog A in DMSO-d_6 .

7) How the titration of mixed solvent affects the α H chemical shifts.

This feature is also directly related with the previous two criteria, but here the focus is on the α H instead of the amide protons. Another consequence of the threaded arrangement and the compact topology is that the structure is not significantly affected by solvent. This fact was supported by ^1H NMR spectra that showed no significant change in the α H chemical shifts,^{9,10} which revealed that these chemical shifts were solvent independent.

In our study we did not perform titration assays of mixed solvent, however we performed the NMR analysis in DMSO- d_6 and using the solvent mixture CD₃CN-H₂O (1:5). In some of the peptide residues the difference was higher than in the others (Table 20), but none of them remained constant. Based on the previous results, in which the amide protons were solvent dependent, this was in fact the expected behavior. Moreover, this can be due to the different hydrogen-bond-acceptor strength of each solvent, which is a characteristic that depends on the polarity and Lewis acid basicity.³²

Residue	δH_α (ppm)		Residue	δH_α (ppm)	
	DMSO- d_6	CD ₃ CN-H ₂ O (1:5)		DMSO- d_6	CD ₃ CN-H ₂ O (1:5)
Gly1	3.60; 3.35	3.93; 3.76	Ser9	4.31	4.45
Phe2	4.12	4.73	Phe10	4.37	4.54
Glu3	4.43	4.57	Gly11	3.75; 3.59	4.05; 3.84
Ser4	4.06	5.10	Leu12	4.41	4.41
Lys5	4.23	4.52	Ser13	4.60	4.56
Pro6	4.17	3.86	Trp14	4.55	4.90
Ile7	4.17	4.61	Leu15	4.24	4.39
Asp8	4.62	4.64			

Table 20: α H chemical shift of the bicyclic analog A in two different solvents.

8) Overlap in ^1H spectra of the corresponding synthetic sequences

For our specific comparison, this last feature did not provide new information. As previously observed in Figure 60 a,b, the difference between the bicyclic and the branched cyclic peptides were not helpful for the structural elucidation. A similar ^1H

spectrum was observed for both peptides, with undefined peaks and several signal overlaps.

Structure calculation

In order to determine the 3D structure of the bicyclic sungsanpin analog A, a conformational sampling algorithm using the Macrocylic Conformational Sampling tool²⁰ was performed with NMR-derived distance restraints (see Materials and Methods for more details). Moreover, two bicyclic structures were used as the starting point. In one case, with the threaded tail and in the other, the tail was manually unthreaded from the ring (Figure 69).

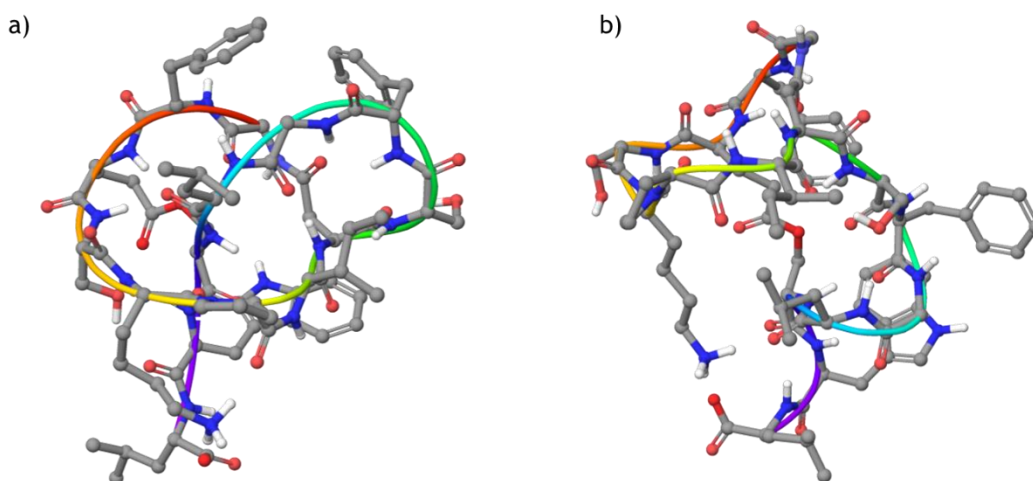
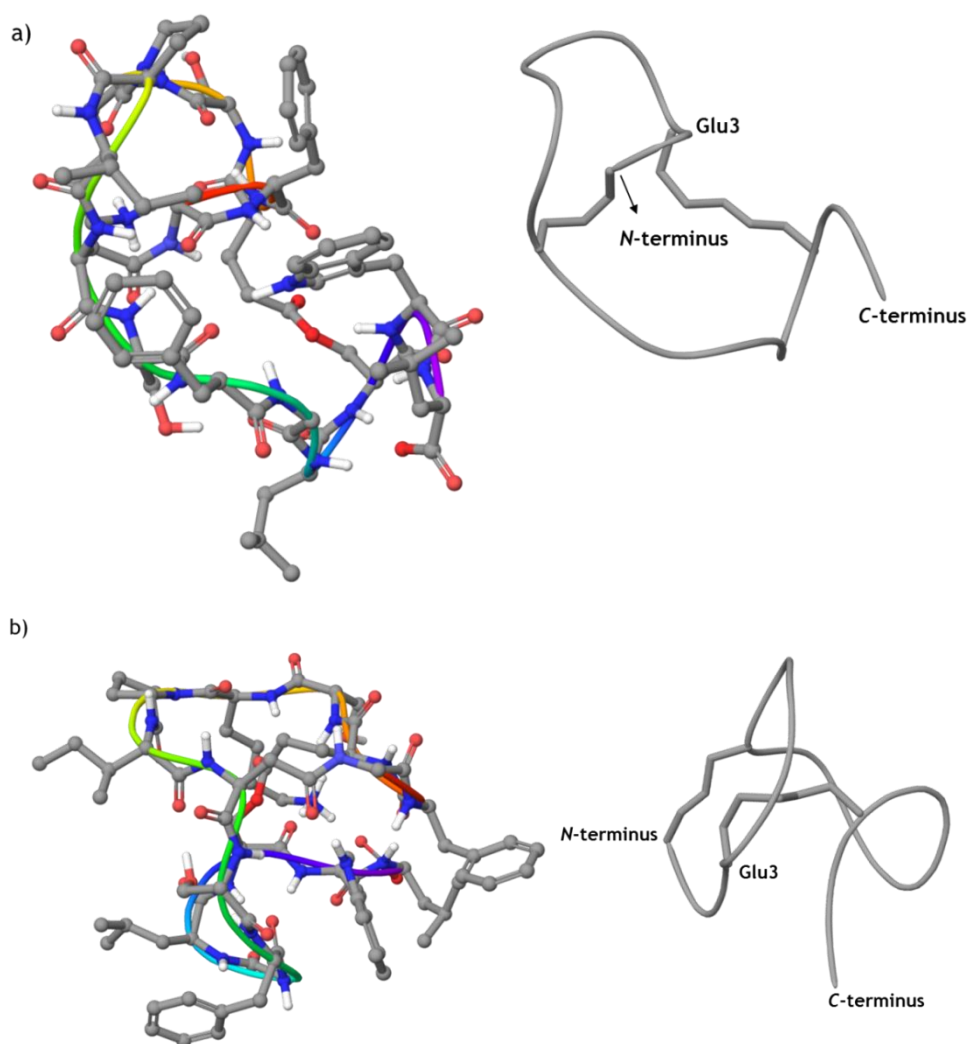


Figure 69: Starting structures for the conformational sampling of the bicyclic sungsanpin analog A with the threaded tail (a) and with the unthreaded tail (b). The coordinates were extracted from the crystal structure of chaxapeptin (PDB code: 2N5C).

The number of long-range NOE detected were not enough to elucidate the exact 3D structure. More long-range NOEs and dihedral angle restraints would be needed, but this was not possible due to the low spectra resolution. Nevertheless, the detected restraints were useful to guide the simulation of the bicyclic 3D structure. The lowest energy structures, shown in Figure 70, display similar structural features. In both structures the tail was unthreaded from the ring. The similarity between both structures was validated by the low backbone RMSD value of 4.2 Å.

The backbone of both minimal energy structures was superposed with their respective starting conformations. The initial and final structures were significantly different for the threaded bicycle (RMSD value of 6.3 Å), showing the incompatibility

of the NOE-derived restraints with the threaded structure. In fact, all the calculated minimal-energy conformations showed the unthreaded tail. For the bicyclic structures, on the other hand, an RMSD value of only 2.7 Å showed that the starting structure was closer to the modeled one with restraints. Moreover, both calculations led to the same bicyclic conformation, with the C-terminal tail outside the macrocyclic ring.



Residues	Distance (Å)
NH Gly1 - HB Glu3	1.5
NH Phe2 - HB, H γ , H δ Leu12	5.0; 7.0; 9.5
H α Pro6 - H γ , H δ Leu12	3.5; 6.0
NH Ser9 - H γ Glu3	7.0
NH ₂ Lys5 - H δ Leu12	12.0
H γ Trp14 - HB Ser9; H α , HB Phe2; H γ Leu12	5.0; 2.0; 2.9; 9.5

Figure 70: NMR-derived structures and ribbon representation of the bicyclic sungsanpin analog. a) The starting structure had the threaded tail and b) the starting structure had the unthreaded tail.

With all these NMR-based experiments and structural calculations, together with the results of the previous techniques, it could be confirmed that the bicycle had the C-terminal tail outside the ring. This evidence was also previously demonstrated with the hydrolysis assay of the ester bond, after which it was readily degraded by carboxypeptidase Y. In the NMR experiments, the spectra showed a very poor amide dispersion, high signal overlap—indicating high flexibility—and a very different topology compared to the bicyclic structure. On the other hand, the NMR of the bicyclic peptide showed less resolution and disorganization than the typical exhibited by a lasso peptide. The ester bond between the tail and the ring was responsible for providing proteolytic stability, but was not sufficient to stabilize the entire structure. The stabilization in natural lasso peptides was achieved by hydrogen bond contacts and hydrophobic and electrostatic interactions. Small β -sheets were formed between the ring and the threaded tail, which provided a high stability and structural organization. Finally, two structures with different topologies were used to perform the NOE-restrained structural modeling. With both conformations, the final structures were almost identical, with the tail clearly outside the ring.

Biological evaluation

General: mechanism of action

As previously presented in the introduction of this thesis, lasso peptides display a wide spectrum of biological activities. Sungsanpin, in particular, shows an inhibitory effect on the invasion of human non-small cell lung cancer (NSCLC), an effect that has been reported with the A549 cell line. It can inhibit the invasion of these tumor cells 25% and 53% versus the control, by exposure to 5 and 50 μM , respectively. NSCLC is the most common type of lung cancer and it is often not spread at the time of the diagnosis, allowing the surgical resection of the tumor.³³ According to the US Centers for Disease Control and Prevention, in 2014 lung cancer was the third most extended cancer.³⁴

Sungsanpin was initially evaluated as an antimicrobial agent and no activity was observed, as expected, due to the lack of D-gene in its gene cluster. Later, it was tested for immunosuppressive activity, showing no activity.²⁶ The cytotoxicity was also evaluated, showing no serious damage for 24 h at concentrations up to 50 μM . Chaxapeptin was also tested on the same invasion assay using the A549 cancer cell line, showing similar results than sungsanpin (28% at 5 μM and 56% at 50 μM). Hence, the difference in the tail of both lasso peptides did not affect the inhibition of cancer cell invasion.¹

Further attempts to explore the mechanism behind the inhibitory effect of sungsanpin revealed that it increased the levels of tissue inhibitors of metalloproteinases (TIMPs), while the levels of matrix metalloproteinases (MMPs) remained unaltered. MMPs are involved in the ECM degradation, playing a crucial role in tumor invasion.³⁵ Their expression is regulated by a variety of extracellular factors, such as growth factors and cytokines. An increase of MMPs expression level is associated with tumor invasion and metastasis.³⁶ TIMPs inhibit the activity of MMPs in extracellular space, binding to the highly conserved zinc binding site of active MMPs.³⁷ Therefore, an increase in the expression of TIMPs by either host or tumor cells results in reduced invasion and metastatic capacity of transformed cells.³⁸ Although TIMPs potentially inactivate MMPs *in vitro*, the regulatory mechanisms of MMP activity by TIMPs has not been yet fully uncovered.³⁹ For this reason, the mechanism of action of these lasso peptides is still unclear.

Cytotoxicity assay

To determine the cell viability, we used the XTT assay. It is based on the ability of living cells to reduce the tetrazolium XTT yellow salt to its orange formazan product, which absorbance is then measured. This value can provide an estimation of living cells compared to dead cells.

Preliminary studies to determine the best condition for the XTT assay

The optimal cell concentration to be seeded is dependent on the basic level of mitochondrial activity and the rate of proliferation. In order to establish this value, several concentrations of cells were seeded in plates, and the absorbance was measured for three days to determine the growth of the cell line. To prevent saturated UV reads, the maximum absorbance value was kept around 2. As shown in Figure 71, the best condition was with 4000 cells/well at 24 h because the growth with 5000 cells/well was too fast. On the other hand, with 2000 cells/well at 48 and 72 h the absorbance was almost the same, thus indicating a very low growth.

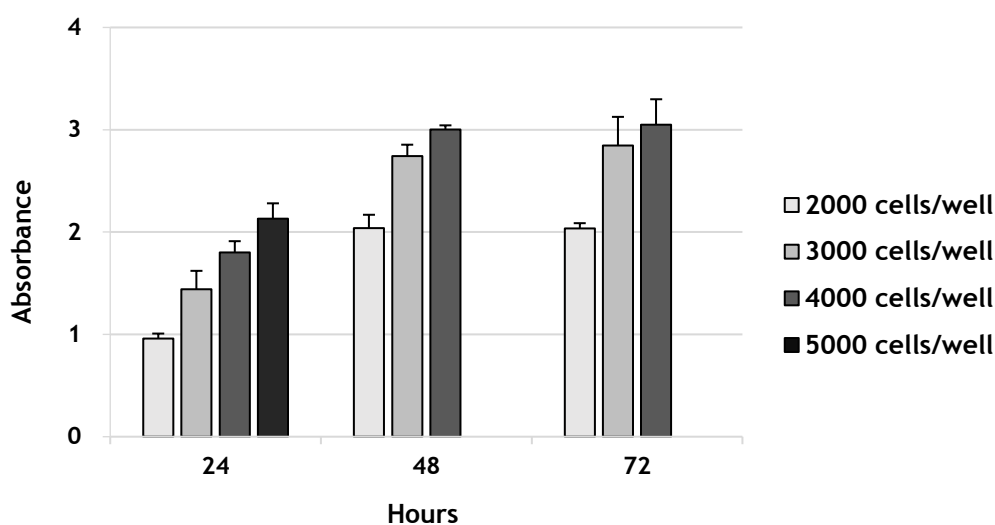


Figure 71: Correlation between the absorbance and the number of cells at different time for the cell line A549.

Cell viability assay

After having optimized the cell seeding concentration to 4000 cells/well, we performed the incubation experiment with different peptide concentrations (25, 50 and 100 μ M). Besides the four previously studied peptides (the synthetic sungsanpin,

the branched cyclic analog, the bicyclic analog A and the chaxapeptin), the cell viability was tested with several additional synthetic peptides (see Figure 72). Since the peptides do not absorb UV light at 475 nm (data not shown), the XTT reagent was added directly to the wells after the 24 h of incubation.

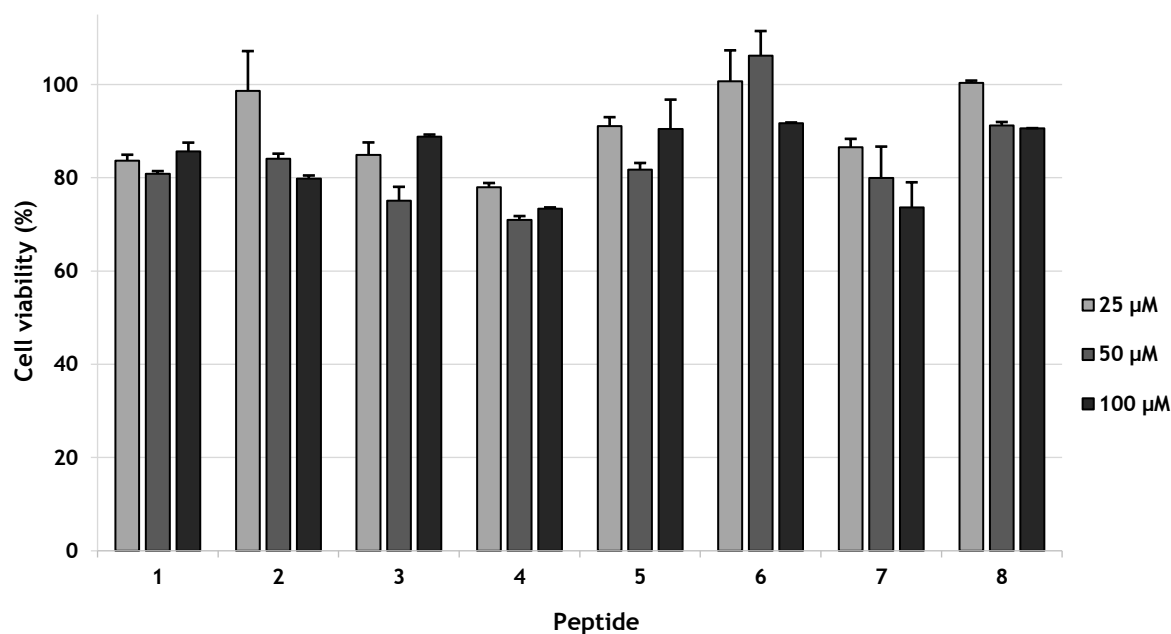


Figure 72: A549 cell viability after 24 h incubation at different concentration of peptides. 1) Chaxapeptin, 2) bicyclic analog A, 3) bicyclic analog C, 4) bicyclic analog D, 5) branched cyclic analog, 6) amide side-chain-to-tail bicyclic analog 7) disulfide Cys3 - Cys15 bicyclic analog and 8) synthetic sungsanpin.

The cytotoxicity of these peptides was compared with chaxapeptin (1), which was used as a control because it showed the same biological activity than sungsanpin. Under the same conditions, all peptides (including chaxapeptin) showed a cell viability around 80%, which can be considered nontoxic, even at the highest concentration (100 μM). However, there were small but reproducible differences between the different analogs tested. From all of them, the synthetic sungsanpin (8) and the branched cyclic analog (5) showed the lowest cytotoxicity.

The amide side-chain-to-tail bicyclic analog (6) had barely no activity in front of the A549 cells. On the other hand, higher cytotoxic activity (73% at 100 μM) was observed for the bicyclic analog (7), in which the tail is connected to the ring via a disulfide bond. From these results, it could be derived that the free C-terminus and Lys5 side chain were more important than the second cyclization involving the C-

terminal tail. Indeed, when these two reactive groups were linked through an amide bond (6), the cytotoxicity was non-existing.

Regarding the three bicyclic peptides having the ester bond (A, C and D), the cytotoxicity results showed opposing trends. The bicyclic analog D (4), showed the highest cytotoxicity (73% cell viability), even slightly higher than chaxapeptin. Otherwise, the bicyclic analog A (2), for which the highest cytotoxicity was expected—due to its high thermal and proteolytic stability—displayed lower activity than D, in the same order as chaxapeptin. Last, the bicyclic analog C (3) showed intermediate values between the other two bicycles.

From these results, no significant differences in the activity could be observed among peptides obtained from the three bicyclic strategies. Comparing these bicycles with the amide side-chain-to-tail bicyclic analog (6), more differences were observed. In this peptide, the second cycle covering all the peptide tail did not improve the activity. Consequently, in general terms, the activity of these peptides may derive from the characteristics of the loop and the spatial disposition rather than from the presence of a second cyclization.

Migration and invasion assay

General introduction

Tumor cells begin metastasis by invading the tissue surrounding the primary tumor. The initial step involves the detachment of epithelial cells from the extracellular matrix (ECM) and the disruption of the actin skeleton, leading to cell rounding.⁴⁰ These issues normally cause apoptotic processes.^{41,42} The ECM is a complex network composed of an array of multidomain macromolecules, such as collagen, proteoglycans, elastin, and fibronectin (Figure 73).⁴³

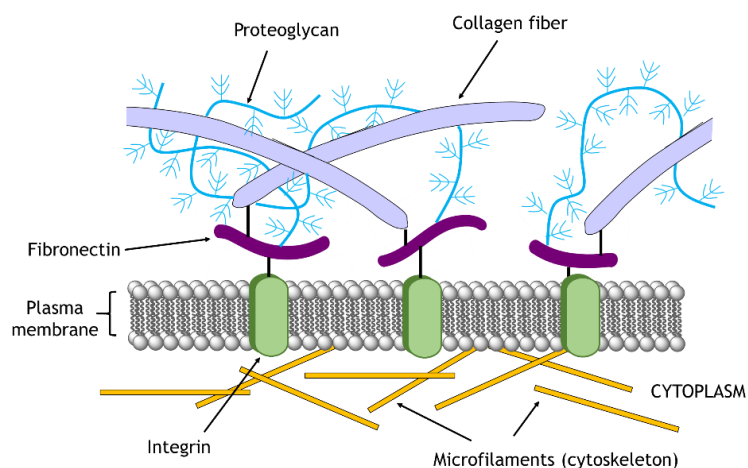


Figure 73: Schematic representation of the extracellular matrix (ECM). (Adapted)⁴⁴

Tumor cells use migration mechanisms to spread within the tissues. For this process to occur, cells must modify their shape and stiffness to interact with the surrounding tissue structures. Thus, the ECM provides the substrate, as well as a barrier towards the advancing of the cell body.^{45,46} Cell invasion refers to 3D migration of cells as they penetrate through the ECM. Both processes are related; the ability to migrate is a prerequisite for tumor cell invasion.

Transwell migration and invasion assay

Transwell chamber migration assay was carried out in order to determine the effect of our peptides on cell motility. This assay measures the number of cells that travel through a porous membrane, and provides information about the ability of cells to sense a particular chemo-attractant (a substance that induces movement of cells in the direction of its highest concentration) and migrate towards it through a physical barrier. In addition, this assay will be later used to investigate cell invasion by adding a layer of Matrigel® to mimic the process of ECM invasion, thus evaluating the ability of cancer cells to infiltrate into neighboring tissues.

In this assay, the size of the membrane pores is determined by the size of the cells to be analyzed. Considering that A549 cells have a diameter of 11-15 μM ,⁴⁷ a smaller pore diameter was selected (8 μm), which allowed an active migration and avoided the unspecific dropping of the cells. Furthermore, the incubation time for the motile cells to migrate to the other side of the membrane was also determined. Invasion is a more challenging process than migration because the cells have to disrupt the ECM in order to cross. Therefore, more incubation time is required (15 h for the migration

assay, versus 24 h for the invasion assay). Moreover, 50000 cells were found to be the optimal number to be seeded in each transwell. Thus, a cell solution in serum-free medium was added on the upper chamber, whereas rich medium (which contains growth factors that act as chemo-attractant) was added to the lower chamber (Figure 74).

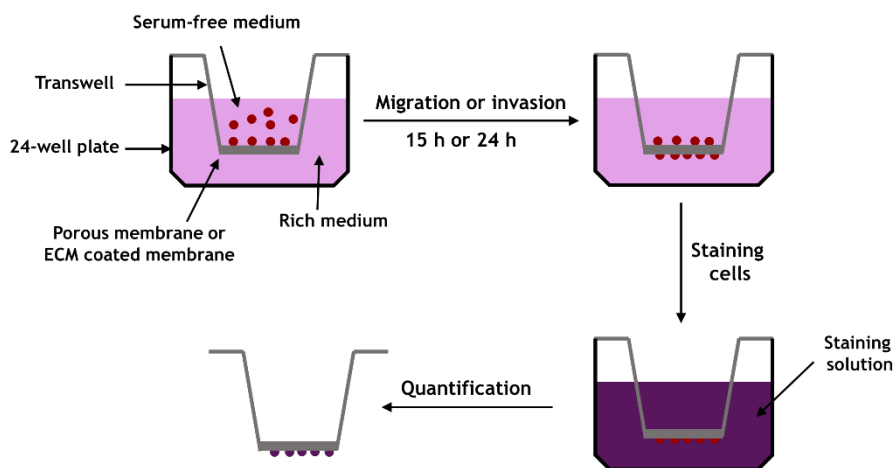


Figure 74: Schematic representation of the transwell migration and invasion assays.

Chaxapeptin was used as a control in both migration and invasion experiments. If the synthetic peptides are more active than chaxapeptin, less migration and invasion should be expected. The migration capacity of A549 is very large, even when a media depleted of chemoattractant is used in both chambers of the transwell. The cells were incubated with 100 μM of peptide for 24 h, conditions at which they were previously tested to be nontoxic.

In the migration assay (Figure 75), the cells incubated with the amide side-chain-to-tail bicyclic analog (6) and with the bicyclic analog A (2) showed more migration than chaxapeptin (1), even higher than cells incubated with the control medium, indicating that these peptides could stimulate the migration of A549 cells. For peptide 6, the high cell migration may be related with the absence of free charges (the amino group of Lys5 side chain and the free C-terminus). This behavior is consistent with the results of the cytotoxicity assay.

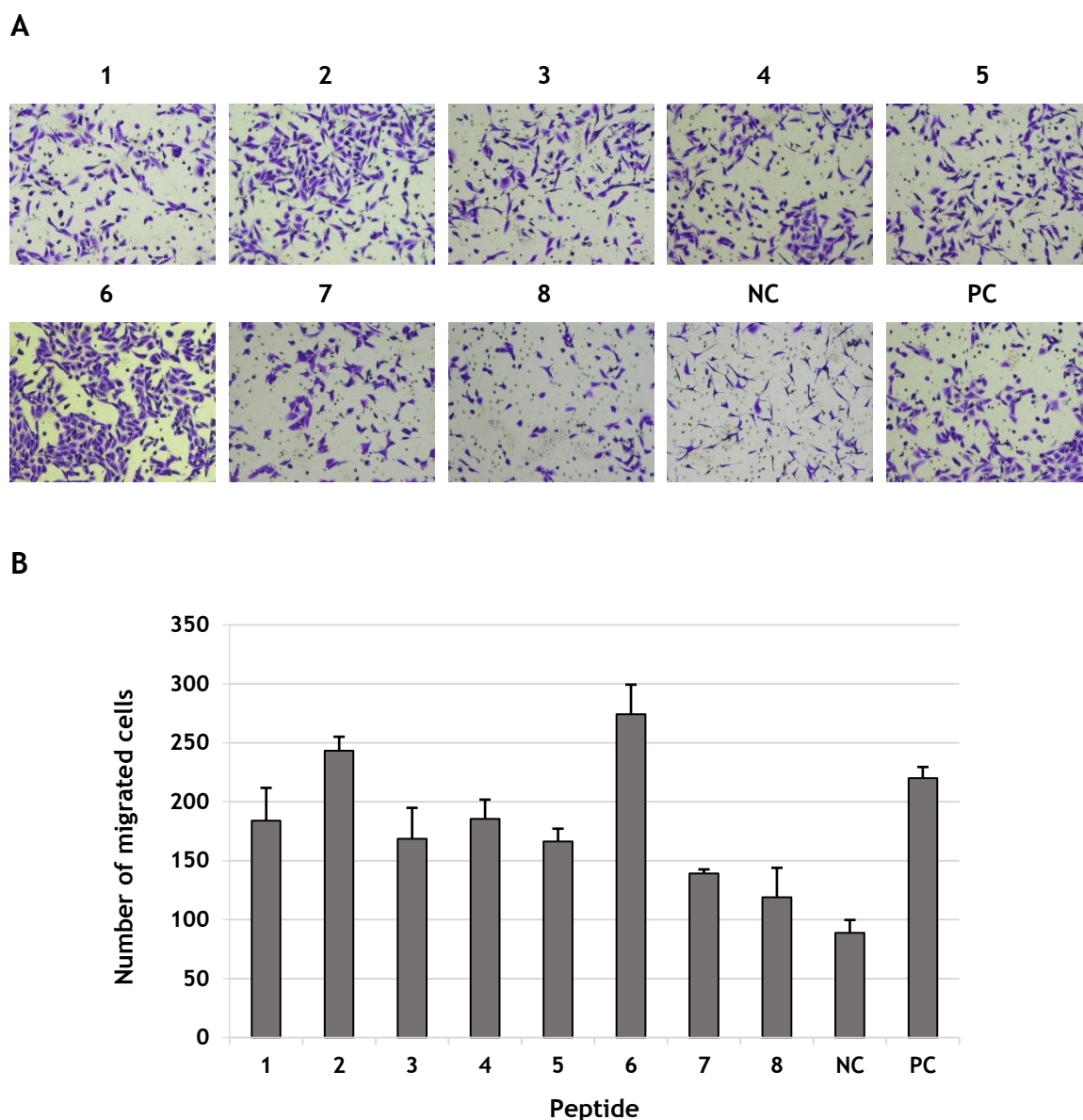


Figure 75: Migration assay after incubation of 100 μM of different peptides for 15 h. A) Pictures of A549 cells transwell migration, using an inverted microscope with a 10x objective. B) Quantification of A549 cell migration. 1) Chaxapeptin, 2) bicyclic analog A, 3) bicyclic analog C, 4) bicyclic analog D, 5) branched cyclic analog, 6) amide side-chain-to-tail bicyclic analog 7) disulfide Cys3-Cys15 bicyclic analog and 8) synthetic sungsanpin. NC and PC: negative and positive control, respectively.

The other peptides showed a similar or lower migration than chaxapeptin (1). For example, both the disulfide Cys3-Cys15 bicyclic analog (7) and the synthetic sungsanpin (8) produced a reduction of up to 50% of cell migration compared to the positive control. Compare to chaxapeptin (1), migration was reduced by 25% and 35%, respectively. On the contrary, the other two bicyclic analogs, C (3) and D (4) showed similar migration values than chaxapeptin, and the bicyclic analog A (2) showed more migration than the other bicycles. These results pointed out to the fact

that the bicycle was not the key feature for the activity. The different conformations adopted, together with the location of the C-terminal tail, were probably responsible of the different migration effects observed for the bicyclic analogs A (2) and C (3), D (4). Following this hypothesis, the migration effect of the branched cyclic analog (5) was also closer to those of chaxapeptin (1) and to the bicycles C (3) and D (4).

With regards to the invasion experiment (Figure 76), the bicyclic analog A (2), the synthetic sungsanpin (8) and the amide side-chain-to-tail bicyclic analog (6) produced high cellular invasion, peptide 6 in particular showing a stronger effect than the positive control, meaning that this peptide stimulated the invasion. On the other hand, the bicyclic analog C (3) was the most active, reducing the invasion up to 50% compared to the positive control and 30% respect to chaxapeptin (1). The second more active peptide was the bicyclic analog D (4) (60% of invasion compared to 82% of chaxapeptin), followed by the branched cyclic analog (5) and the disulfide Cys3-Cys15 bicyclic analog (7), both showing a 10% more reduction in invasion than chaxapeptin (1). In general, all these peptides showed a significative reduction of the tumor cellular invasion, the best result obtained was with peptide 3, with which invasion was reduced up to 50%.

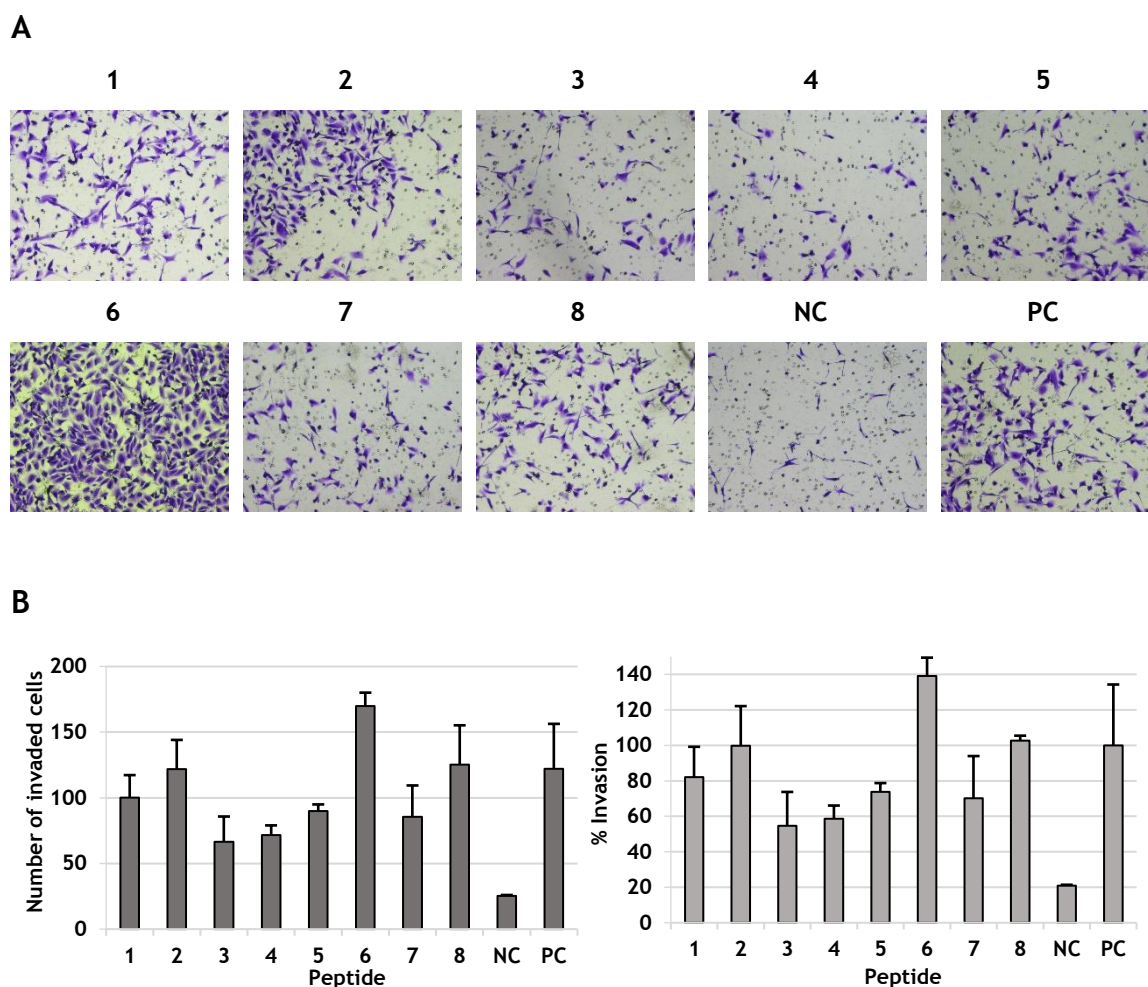


Figure 76: Invasion assay after incubation of 100 μ M of different peptides for 15 h. A) Pictures of A549 cells transwell invasion, using an inverted microscope with a 10x objective. B) Quantification of A549 cell invasion. Left- Number of invaded cells. Right- Number of invaded cells with respect to the number of cells of the positive control. 1) Chaxapeptin, 2) bicyclic analog A, 3) bicyclic analog C, 4) bicyclic analog D, 5) branched cyclic analog, 6) amide side-chain-to-tail bicyclic analog 7) disulfide Cys3-Cys15 bicyclic analog and 8) synthetic sungsanpin. NC and PC: negative and positive control, respectively.

The high ability of A549 tumor cells to migrate and invade was shown in the negative control (NC) of both assays, in which some cells were even detected on the lower chamber (Figure 76 A).

Comparing the results from the migration and invasion assays, both of them showed that the amide side-chain-to-tail bicyclic analog (6) stimulated the migration and consequently, the cell invasion, showing higher values than the positive control. We hypothesize that this effect could be related to the loss of the two sole charges of the peptide, the Lys5 and the C-terminus. The only difference between this peptide (6) and the synthetic sungsanpin (8) was the second cyclization through the amide bond between these two residues, which suggests that the interaction of these

two residues was highly important for activity. This was not the first time that this feature was observed. Previously, the high importance of the C-terminus in the RNAP inhibitory activity of MccJ25, has already been reported.⁴⁸ Vincent and coworkers introduced a methyl group into the C-terminal Gly, observing the absence of the RNAP inhibition. They suggested that the methyl group could either perturb the approximation to RNAP MccJ25-binding site⁴⁹ or cancel the possible electrostatic interaction between the unique carboxyl group of MccJ25 with a positive residue of RNAP. Additionally, it was also reported that the electrostatic interaction between the C-terminus and His5 (the only two charges of MccJ25) were involved in the stabilization of the threaded structure.⁹ These results could be extrapolated to our study, in which the absence of the free C-terminus causes an increase of the cell motility, and consequently low inhibitory effect against the migration and invasion. This fact was also confirmed by the comparison between peptides 6 and 7, both displayed a second cyclization between positions 5 and 15 and 3 and 15, respectively. However, peptide 7 did not show any charge, which had a significant effect on the inhibitory effect on the migration and invasion.

Regarding sungsanpin (8), the results obtained were more difficult to rationalize. It displayed a great activity against cell migration, reducing it by 50% and 35% compared to the positive control and the chaxapeptin (1), respectively. However, the invasion was not affected. The invasion results indicated that it stimulated the cells invasion. These results should be confirmed with a third replicate. Moreover, the bicyclic analog A (2) did not stimulate the migration, but it did not improve either the invasion activity of chaxapeptin (1). The lowest migration values, which were confirmed in the invasion assay, were obtained for the two bicyclic analogs C (3) and D (4), the branched cyclic analog (5) and the disulfide Cys3 - Cys15 bicyclic analog (7). With these four peptides, a significant decrease in cancer cell motility was observed.

Therefore, if the pending third replicate does confirm these results, the peptides that showed the highest biological activity in the migration and invasion assays were the bicyclic analogs from the strategies C (3) and D (4). In both cases the invasion was reduced by 30% with respect to chaxapeptin (1). This result indicates that the bicyclic structure with a specific disposition of this C-terminal tail, together with the side-chain of Lys5 and the free C-terminus, were important to reduce the motility of the A549 cells.

Discussion

Comparing the relation between the synthetic strategies, the structural analysis and the biological effect, some features could be summarized. The presence of the free charges and the electrostatic interactions between them, play an important role in the biological activity. For this reason, the branched cyclic analog (5), which has two negative and one positive charges showed a similar migration than chaxapeptin (1) and even less invasion. The same activity values were observed for the disulfide Cys3-Cys15 bicyclic analog (7), which displayed a bicycle between the side chains of two cysteines and the free C-terminus. Both showed free charges and the only difference was the presence or not of the second cyclization.

The lack of the charges was related with a considerable increase in cell migration and invasion. For example, in the case of the amide side-chain-to-tail bicyclic analog (6), in which the C-terminal charge was not present, the invasion was higher than the control. As mentioned before and in agreement with the already published results for MccJ25, the negative charge of the C-terminus was very important for the activity.

The ester bond from the bicyclic analogs was not enough to stabilize the structure and fix a defined conformation. It was the responsible for maintaining the rigidity in the structure and providing proteolytic stability, but only in the case of the bicyclic analog A. In the other two analogs with an ester bond, the C-terminal residues were more accessible to the carboxypeptidase Y, making them readily degradable. The opposite trend was observed by the migration and invasion assays, in which the analogs C (3) and D (4) were more active than the bicyclic analog A (2). In conclusion, the second cyclization did not provide the biological activity. This fact was also confirmed by the comparison between the amide side-chain-to-tail bicyclic analog (6) and the disulfide Cys3-Cys15 bicyclic analog (7), in which both displayed a bicycle between the same number of residues, but in one case through the C-terminal carboxylic acid and in the other, through a disulfide bond between two cysteines. Only in the latter, the C-terminal charge was present. This remark was in agreement with the previous mentioned, that the responsible of the biological activity were the hydrogen bond and hydrophobic and electrostatic interactions between the ring and the threaded tail.

The highest inhibitory effect on the invasion of A549 tumor cells was found for the bicyclic analogs from the strategies C (3) and D (4). Both peptides showed a reduction by 30% with respect to chaxapeptin (1). Moreover, the branched cyclic analog (5) and the disulfide tail-to-ring bicyclic analog showed less invasion than chaxapeptin (10% less). These results showed that these three peptides are more active than chaxapeptin and therefore, than sungsanpin.

References

- (1) Elsayed, S. S.; Trusch, F.; Deng, H.; Raab, A.; Prokes, I.; Busarakam, K.; Asenjo, J. A.; Andrews, B. A.; Van West, P.; Bull, A. T.; Goodfellow, M.; Yi, Y.; Ebel, R.; Jaspars, M.; Rateb, M. E. *J. Org. Chem.* **2015**, *80* (20), 10252.
- (2) Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A. *Chem. Biol.* **2013**, *20* (4), 558.
- (3) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. *Biopolymers* **2013**, *100* (5), 527.
- (4) Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A. *Chem. Sci.* **2014**, *5* (10), 4032.
- (5) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2013**, *135* (1), 210.
- (6) Yeates, T. O.; Kent, S. B. H. *Annu. Rev. Biophys.* **2012**, *41* (1), 41.
- (7) Nar, H.; Schmid, A.; Puder, C.; Potterat, O. *ChemMedChem* **2010**, *5* (10), 1689.
- (8) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Steuber, H.; Harms, K.; Xie, X.; Marahiel, M. A. *Angew. Chemie - Int. Ed.* **2014**, *53* (8), 2230.
- (9) Rosengren, K. J.; Clark, R. J.; Daly, N. L.; G??ransson, U.; Jones, A.; Craik, D. *J. Am. Chem. Soc.* **2003**, *125* (41), 12464.
- (10) Knappe, T. a.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. a. *J. Am. Chem. Soc.* **2008**, *130* (17), 11446.
- (11) Dong, N. P.; Zhang, L. X.; Liang, Y. Z. *Int. J. Mass Spectrom.* **2011**, *308* (1), 89.
- (12) Vaisar, T.; Urban, J. *J. Mass Spectrom.* **1996**, *31* (10), 1185.
- (13) Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A. *FEBS Lett.* **2010**, *584* (4), 785.
- (14) Li, Y.; Ducasse, R.; Zirah, S.; Blond, A.; Goulard, C.; Lescop, E.; Giraud, C.; Hartke, A.; Guittet, E.; Pernodet, J. L.; Rebuffat, S. *ACS Chem. Biol.* **2015**, *10* (11), 2641.
- (15) Jeanne Dit Fouque, K.; Afonso, C.; Zirah, S.; Hegemann, J. D.; Zimmermann, M.; Marahiel, M. A.; Rebuffat, S.; Lavanant, H. *Anal. Chem.* **2015**, *87* (2), 1166.
- (16) Smith, D.; Knapman, T.; Campuzano, I.; Malham, R.; Berryman, J.; Radford, S.; Ashcroft, A. *Eur. J. Mass Spectrom.* **2009**, *15* (5), 113.
- (17) Jeanne, K.; Fouque, D.; Lavanant, H.; Zirah, S.; Lemoine, J.; Rebuffat, S.;

- Tabet, J.; Kulesza, A.; Afonso, C.; Dugourd, P.; Chirot, F. *Rapid Commun. Mass Spectrom.* **2015**, No. May, 1411.
- (18) Fouque, K. J. D.; Lavanant, H.; Zirah, S.; Hegemann, J. D.; Zimmermann, M.; Marahiel, M. A.; Rebuffat, S.; Afonso, C. *J. Am. Soc. Mass Spectrom.* **2016**.
- (19) Watts, P.; Wilders, A. *Int. J. Mass Spectrom. Ion Process.* **1992**, *112* (2-3), 179.
- (20) Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. *J. Chem. Inf. Model.* **2014**, *54* (10), 2680.
- (21) Schubert, M.; Labudde, D.; Oschkinat, H.; Schmieder, P. *J. Biomol. NMR* **2002**, *24* (2), 149.
- (22) Kang, Y. K. *J. Phys. Chem. B* **2002**, *106* (8), 2074.
- (23) Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. *Biochemistry* **2002**, *41* (50), 14637.
- (24) Xie, X.; Marahiel, M. A. *ChemBioChem* **2012**, *13* (5), 621.
- (25) Fréchet, D.; Guitton, J. D.; Herman, F.; Faucher, D.; Helynck, G.; Monegier du Sorbier, B.; Ridoux, J. P.; James-Surcouf, E.; Vuilhorgne, M. *Biochemistry* **1994**, *33* (1), 42.
- (26) Um, S.; Kim, Y.-J. J.; Kwon, H. H. C.; Wen, H.; Kim, S.-H. H.; Kwon, H. H. C.; Park, S.; Shin, J.; Oh, D.-C. C. *J. Nat. Prod.* **2013**, *76* (5), 873.
- (27) Blond, A.; Cheminant, M.; Ségalas-Milazzo, I.; Péduzzi, J.; Barthélémy, M.; Goulard, C.; Salomón, R.; Moreno, F.; Farías, R.; Rebuffat, S. *Eur. J. Biochem.* **2001**, *268* (7), 2124.
- (28) Grathwohl, C.; Wüthrich, K. *J. Magn. Reson.* **1974**, *13* (2), 217.
- (29) Bundi, A.; Grathwohl, C.; Hochmann, J.; Keller, R. M.; Wagner, G.; Wüthrich, K. *J. Magn. Reson.* **1975**, *18* (1), 191.
- (30) Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Omura, S. *J. Am. Chem. Soc.* **2006**, *128* (23), 7486.
- (31) Knappe, T. A.; Linne, U.; Zirah, S. S.; Rebuffat, S.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2008**, *130* (17), 11446.
- (32) Tonan, K.; Ikawa, S. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **2003**, *59* (1), 111.
- (33) Tanley, M. D.; Swierzewski, J. Lung Cancer <http://www.healthcommunities.com/lung-cancer/lung-cancer-overview.shtml> (accessed Nov 21, 2017).
- (34) Centers for Disease Control and Prevention. 2014 Top Ten Cancers <https://nccd.cdc.gov/uscs/toptencancers.aspx> (accessed Nov 20, 2017).
- (35) Shapiro, S. D. *Curr. Opin. Cell Biol.* **1998**, *10* (5), 602.

-
- (36) Bolon, I.; Brambilla, E.; Vandenbunder, B.; Robert, C.; Lantuejoul, S.; Brambilla, C. *Lab. Invest.* **1996**, *75* (1), 1.
- (37) Brew, K.; Nagase, H. *Biochim. Biophys. Acta - Mol. Cell Res.* **2010**, *1803* (1), 55.
- (38) Krüger, A.; Sanchez-Sweatman, O. H.; Martin, D. C.; Fata, J. E.; Ho, A. T.; Orr, F. W.; Rüther, U.; Khokha, R. *Oncogene* **1998**, *16* (18), 2419.
- (39) Westermarck, J.; Kähäri, V. M. *FASEB J.* **1999**, *13* (8), 781.
- (40) Martin, S. S.; Vuori, K. *Biochim. Biophys. Acta - Mol. Cell Res.* **2004**, *1692* (2-3), 145.
- (41) Streuli, C. H.; Gilmore, A. P. *J Mammary Gland Biol Neoplasia* **1999**, *4* (2), 183.
- (42) Martin, S. S.; Leder, P. *Mol. Cell. Biol.* **2001**, *21* (19), 6529.
- (43) Yue, B. *J. Glaucoma* **2014**, *23* (8), S20.
- (44) Watts, P.; Wilders, A. *Int. J. Mass Spectrom. Ion Process.* **1992**, *112* (2-3), 179.
- (45) Friedl, P.; Bröcker, E. B. *Cell. Mol. Life Sci.* **2000**, *57* (1), 41.
- (46) Abercrombie, M.; Dunn, G. A.; Heath, J. P. *Soc. Gen. Physiol. Ser.* **1977**, *32*, 57.
- (47) Jiang, R. De; Shen, H.; Piao, Y. J. *Rom. J. Morphol. Embryol.* **2010**, *51* (4), 663.
- (48) Vincent, P. A.; Bellomio, A.; De Arcuri, B. F.; Farías, R. N.; Morero, R. D. *Biochem. Biophys. Res. Commun.* **2005**, *331* (2), 549.
- (49) Mukhopadhyay, J.; Sineva, E.; Knight, J.; Levy, R. M.; Ebright, R. H. *Mol. Cell* **2004**, *14* (6), 739.

CHAPTER 3:
BIOSYNTHESIS OF LASSO
PEPTIDES

The present chapter has been written with the work carried out during a research stay in the group of Prof. Mohamed Marahiel in the Philipps-Universität Marburg, in Germany, and under the supervision of Dr. Julian Hegeamann. The main goal of the stay was the biosynthesis of new sungsanpin-like lasso peptide and other lasso peptides. On the basis of the 16S rDNA sequence (GenBank accession number: KC871171) of the strain SNJ013, it was found that this strain was closely related to the *Streptomyces* sp. I08A-00426 (87.5% identity, 87.5% similarity, 12.1% gaps).¹

Genome mining

As mentioned in the introduction, lasso peptides are a class of ribosomally-synthesized and post-translationally modified natural products with diverse bioactivities and a unique three-dimensional structure in which the C-terminus threads through an N-terminal macrolactam ring.² Moreover, most of lasso peptides were discovered by isolation from bacteria, but this changed in 2008 with the report of the first lasso peptide isolated by a genome mining approach.³ Genome mining is a methodology that involved the search of new genomic information for the discovery of new pathways or metabolites. It provides access to a vast portion of natural product diversity that otherwise would remain unexploited.⁴ It consists in the *in silico* search of enzymes involved in the synthesis of secondary metabolites.

Regarding the previous remark, the first step was to perform B-protein centric genome mining to identify promising strains that would produce lasso peptides like sungsanpin. As mentioned in the introduction, lasso peptide biosynthetic gene clusters consist of at least three genes. These encode the precursor peptide (A), an ATP-dependent cysteine protease (B), and an ATP-dependent asparagine synthetase (C). Sometimes a fourth gene, called D-protein, encoding an ABC-transporter (D) can be found providing antimicrobial activity.

B-proteins are essential for lasso peptide maturation and their coding genes are commonly found in biosynthetic gene clusters.⁵ At some point B-protein was split into two separate ORFs: B1 and B2. In general, the gene cluster is formed by three proteins; B2 encodes about 150 amino acids in size, B1 encodes around 80 amino acids and C, which is the largest, around 500-600 amino acids. The length of the precursor peptide can differ, depending on the length of the leader and the core peptide. Concerning this fact, B2-protein was chosen as a centric genome because it

has been found exclusively in lasso peptide biosynthetic gene cluster and is larger and more specific than B1.

A new lasso peptide, called chaxapeptin,⁶ was isolated from the *Streptomyces* strain and its gene cluster was identified, taking Lariatins as a reference.⁷ For this purpose, the lasso peptide chaxapeptin was selected as a starting point for our genome mining because it shares 73.3% identity and 93.3% similarity with sungsanpin. Moreover, both contain 15 amino acids, 8-residue ring and 7-residue loop and tail. They only differ in four amino acids, three of them located in the tail and one in the ring. For this reason, B2-gene of chaxapeptin (CptB2) was used as a starting point to identify lasso-B2-homologs through a genome mining approach using a position-specific iterative basic local alignment search tool (PSI-BLAST).⁸ This tool is a sequence similarity search program that can be used to quickly search a sequence database for matches to a query sequence. Therefore, our B2 sequence was used to perform a protein BLAST and afterwards, manual checking was done to look for similar sequences and clusters.

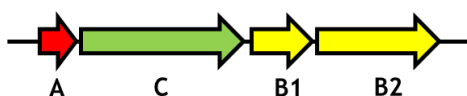


Figure 77: Schematic organization of the ACBB-cluster from chaxapeptin.

Regarding the gene cluster organization of chaxapeptin (Figure 77), at the surroundings clusters of the B2-gene, a smaller ORF encoding a B1-homolog must be found in front of it, followed by a bigger one encoding a larger protein, which should be a C-protein homolog. To validate the likely cluster, C-protein must be checked by BLAST looking for some distinctive hit to confirm that it is an asparagine synthetase. If all these features are present, then a lasso cluster has been identified. After this exhaustive search with our B2-gene, we were able to identify ten promising strains from a total of 100 (Table 21):

Strain	Precursor peptide	
	Leader region	Core peptide
<i>Streptomyces bingchenggensis</i>	MEMQARSDETVESYEPALVAVGEFSEDTL	GFGRHYEDVLGEQGWG
<i>Streptomyces lydicus</i>	MEEQNELSNVEPYAPPMLVEVGEFSEDTL	GFIGWGKDFGHYGG
<i>Streptomyces auranticus</i> JA 4570	MEEQTGIDERDTAASYEPALIAVGEFSEDTL	GFGRNRYNDVLGEQGW
<i>Streptomyces leeuwenhoekii</i> C34	MEPQMTELQPEAYEAPSLIEVGEFSEDTL	GFGRSKPLDSFGLNFF
<i>Streptomyces</i> sp. PBH53	MNENIEISNSEPYEPPTVVEVGEFNEDTL	GFIGWGKDFGHYSGGF
<i>Streptomyces</i> sp. WC-3641	MDEQTEFSTTEPYAPPVLEVGEFNEDTL	GFIGWGDDIFGHYSGDF
<i>Streptomyces</i> sp. S-118	VRESGDIEQAEVYEPVLEEVEGGFTEKTE	GYSGFSDGWGGFLHLG
<i>Streptomyces monomycini</i>	MQDVQERAEATEYEAPLAEAGRFTERTL	GFTGEHSDNWGGMHSTFW
<i>Streptomyces vitaminophilus</i>	MERETRGETAEVYEPVVLVGIGDFADETQ	GQWIGAEPDGSWGHHTPPS
<i>Streptomyces natalensis</i>	MEEPAVEPVTTSYEPVTEAGAFAEVTR	GGIGTFKEAGVGRFM

Table 21: Precursor peptide sequences.

10 precursor peptides were identified as promising sequences to produce lasso peptides. Looking carefully at the sequences, several residues in the leader region have been shown to have high importance in the lasso maturation process. For example, the case of Thr at the penultimate position of the leader region which is highly conserved.⁹ In addition, Gly-8 was found to be essential for the lasso production and has the same importance as Thr-2. Other conservative motifs are hydrophobic residues that can be found in positions -6 (Phe), -9 (Val/Ala), -11 (Val/Ile/Ala) and -12 (Val/Leu) and are highly important for the *in vivo* maturation. Moreover, it has been reported that small or medium residues at position -1, instead of large residues, are preferred for the maturation process, since this is the last position of the leader peptide and it can perturb the interaction with the protease containing the B-protein.¹⁰ Furthermore, there is a degree of conservation on the residue -1, which predominantly contains Leu on the residue -10, which contains an

ionic side chain (Glu in most of the sequences). Pro at position -14 and Tyr at the -17 are also conserved residues. An alignment of all these leader sequence was done with the MEME algorithm and it is shown in Figure 78.

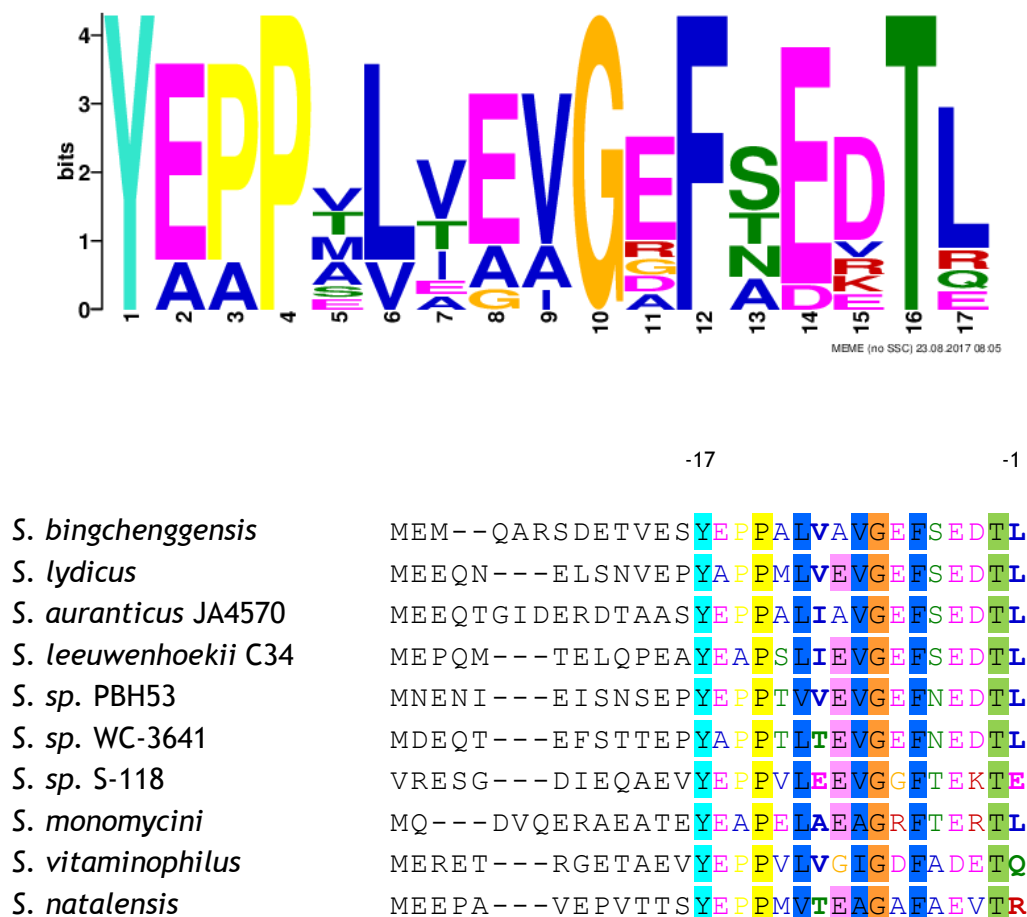


Figure 78: Leader peptide motif and sequence alignment using MEME.

From the 10 promising strains identified by genome mining to produce lasso peptides, we were able to reach only 7 from different commercial collections (see Materials and Methods for more details). Once we had these strains, the gene clusters of each strain were analyzed and compared with the chaxapeptin gene cluster (expressed from the *S. leeuwenhoekii* C34 and C58)^{11,12} using the EMBOSS Needle, a program that allows the alignment of two sequences¹³ (Table 22 and 23).

Strain	C-protein length (aa)	B1-protein length (aa)	B2-protein length (aa)
<i>S. leeuwenhoekii</i> C34	604	85	138
<i>S. lydicus</i>	572	83	142
<i>S. sp.</i> WC-3641	587	90	143
<i>S. sp.</i> S-118	544	81	139
<i>S. monomycini</i>	625	84	141
<i>S. vitaminophilus</i>	646	88	136
<i>S. natalensis</i>	611	74	143

Table 22: Gene cluster length of each strain

Strain	C-protein				B1-protein				B2-protein			
	I (%)	S (%)	G (%)	L (aa)	I (%)	S (%)	G (%)	L (aa)	I (%)	S (%)	G (%)	L (aa)
<i>S. lydicus</i>	51.7	63.7	8.8	615	43.5	62.4	2.4	85	65.5	83.2	2.8	142
<i>S. sp.</i> WC-3641	51.7	62.4	6.3	615	43.5	60.9	9.8	92	66.4	78.3	4.2	143
<i>S. sp.</i> S-118	31.5	46.8	14.8	620	34.5	52.9	9.2	87	51.7	58.5	12.2	147
<i>S. monomycini</i>	35.4	50.8	8.6	642	37.5	53.5	3.5	86	55.3	63.1	2.8	141
<i>S. vitaminophilus</i>	33.9	44.0	13.3	670	37.5	58	3.4	88	53.2	64.0	3.6	139
<i>S. natalensis</i>	35.7	47.6	15.3	658	27.6	46	17.2	87	53.1	65.7	4.2	143

Table 23: Comparison of each protein from chaxapeptin with the different strains. I: identity: exact same residues. S: similarity: same type of amino acids (e.g. Arg instead of Lys). G: gaps. L: length. The closer strains are presented in bold characters.

Regarding these results, gene clusters from the *S. lydicus* and *S. sp.* WC-3641 show the closest identities and similarities with chaxapeptin. This evidence can be related to the fact that their core peptides have higher identity with sungsanpin (31.2% and 29.4%, respectively) than the others, as it can be observed in the following alignment (Figure 79). Five residues are conserved, while the others are different. Conservative variations have been obtained, maintaining the hydrophobicity of the peptide tail. Residues with bulky side chains are used as a top and bottom plugs, preventing the tail from slipping out from the ring.

Sungsanpin	G	F	G	S	K	P	I	D	S	F	G	L	S	W	L	
<i>S. leeuwenhoekii</i> C34 (chaxapeptin)	G	F	G	S	K	P	L	D	S	F	G	L	N	F	F	
<i>S. lydicus</i>	G	F	I	G	W	G	K	D	I	F	G	H	Y	G	G	R
<i>S. sp.</i> WC-3641	G	F	I	G	W	G	D	I	F	G	H	Y	S	G	D	F

Figure 79: Alignment of core peptides of the closer strains. Underlined residues are the plugs of the lasso peptides.

In conclusion, after an exhaustive *in silico* search based on B2 protein-centric genome mining approach, we were able to reach 7 strains containing gene cluster-like lasso peptides. Two of them, *S. lydicus* and *S. sp.* WC-3641, showed the closest identities and similarities with chaxapeptin and therefore, to sungsanpin. The next step was the expression of these strains to test if lasso peptides are actually obtained.

Homologous expression

Homologous expression means that the bacterial strain is incubated directly in one or different growth medium during several days at a specific temperature to produce the secondary metabolites. This incubation could be performed in liquid medium or in agar-plates. Depending on the surfaces, different secondary metabolites could be expressed. As mentioned before, lasso peptides are secondary metabolites from bacteria production and referring to the initial definition, their production is strongly dependent on the growth media.¹⁴ Added to this, temperature, pH and inoculum size also play an important role in the *Streptomyces* growth. GYM is the standard medium used for *Streptomyces* cultivation. It contains yeast malt extract agar and it is indicated for the isolation of yeasts and molds.¹⁵ Concerning this previous remark, each strain was incubated in different growth medium (GYM, SM and BI) to evaluate how it affects to the lasso production. Moreover, the growth was performed in liquid culture and in agar plates to further extract the *Streptomyces* spores and check for lasso production. Morphological differentiation in *Streptomyces* involves the formation of a lawn of aerial hyphae on the colony surface that stands up into the air and differentiates into chains of spores.¹⁶ Incubation of *Streptomyces* in solid media produces aerial exospores with different pigments. These pigments are accumulated in the spores and each of them can contain a different gene cluster.¹⁷ Thus, the growth was performed in agar plates and in liquid culture at 30 °C for 7 days. In liquid culture, *Streptomyces* formed a white or pale brown mycelium, that looked like fluffy balls. Cultures were observed during several days and its appearance was

annotated to avoid undesired side products and contaminations (Table 24). Two of the strains (*S. vitaminophilus* and *S. natalensis*) were suspended in GYM media, but no growth was observed in these cases.

Strains	<i>S. lydicus</i>	<i>S. leeuwenhoekii</i> C34	<i>S. sp. WC-3641</i>	<i>S. monomycini</i>	<i>S. sp. S-118</i>
	Orange with transparent balls	Transparent with dark balls	Light yellow and transparent	Dark brown	Light brown-orange
Media	Dark brown and dense	Dark brown and opaque	Dark brown and transparent	Light brown-green	Light brown opaque
	White, light-yellow	Light brown and opaque	White-light brown	Red-brown	Pink-orange

Table 24: Color-appearance description of each liquid culture after 7 days.

After 7 days, cultures were harvested, and supernatant and pellet extracts were screened for the molecular masses by LC-MS. Both parts were analyzed, but in ACBB-gene cluster bacteria lasso peptide should be extracted from the pellet. Only in the cases where D gene (ABC-transporter) is present, lasso peptides are transported to the supernatant, like microcin J25,^{18,19} capistrin²⁰ or lariatin⁷. Presence of lasso peptide was not detected for *S. lydicus*, *S. sp. WC-3641* nor *S. sp. S-118* in any media. Only small amounts of lasso peptide in the pellet extract were detected for *S. monomycin* in GYM and SM medium. However, chaxapeptin was the only lasso peptide expressed in moderate yields. It was produced by the *S. leeuwenhoekii* C34 in the pellet extract in GYM media and some traces in SM and BI were detected. Afterwards, *S. leeuwenhoekii* C34 culture was inoculated in 10 L GYM media at 30 °C for 7 days and further purification was performed in a semi preparative HPLC, yielding low production. To improve production yield, heterologous expression was performed as it is shown in the next section.

On the other hand, the strain suspension was spread on the agar plates with GYM and SM media incubated at 37 °C. The plates were observed during several days to confirm that they had a homogeneous growth and the different pigments were compared (Figure 80). After 21 days, extraction from the plates was performed with methanol and the supernatant was analyzed by LC-MS. No amount of lasso peptide was detected for any strain.

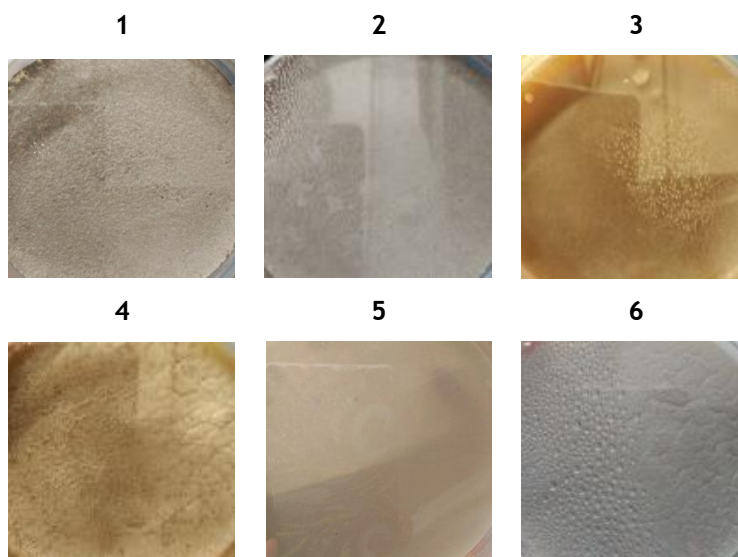


Figure 80: *Streptomyces* growth after 17 days in agar-plates with different media at 30°C. 1) *S.lyd_17d_GYM*, 2) *S.lyd_17d_SM*, 3) *S.mon_17d_GYM* (no growth), 4) *S.mon_17d_SM*, 5) *S.118_17d_GYM* AND 6) *S.118_17d_SM*. No growth for *S. WC-3641*.

In this section, the homologous production of the 7 strains previously identified by a genome mining approach containing the gene cluster like-lasso peptide, which was evaluated in liquid and solid cultures using different growth media. Only in the case of *S. leeuwenhoekii* with GYM medium, the lasso peptide was expressed. Moreover, in SM medium, small amounts of lasso peptide were detected, but comparing with GYM medium, the production was very low. Regarding these results, the homologous production of *S. leeuwenhoekii* was scaled up to 10 L, producing the lasso peptide in moderate yields. On the contrary, with the other strains no lasso peptide production was observed in any medium. Analyzing their gene clusters *in silico*, they were suitable candidates for lasso production, but during the expression process no metabolite with the expected mass was obtained. Concerning this, heterologous expression would be a good alternative to obtain lasso peptides.

Heterologous expression

The objective in this section was the heterologous expression of the previously identified gene clusters, which means the expression of a gene in either a different species, or cell type, than it originates. First, the DNA of each strain was extracted and then, some mutations were introduced in their gene clusters to create more similarity with the bacteria host. Finally, the modified gene clusters were subjected to fermentation in M9 minimal media under different growth conditions.

After the genome mining and the evaluation of lasso peptide production of each identified strains, the objective was the introduction of several modifications in their gene clusters. In this regard, more similarity and identity to sungsanpin could be created and even its total biosynthesis could be reached.

DNA isolation and cloning

Numerous lasso peptides, produced by proteobacteria, have been recently isolated using genome mining approaches and their gene cluster has been elucidated. Regarding these sequences, some similarities between them have been found,¹⁰ but only in Gram-positive microorganisms.⁵ Nowadays, there are more specific protocols for heterologous expression for proteobacteria clusters than for other phylums, such as actinobacteria, which are less feasible.²¹ Considering lasso peptide production in actinobacteria, the expression is more efficient in rich media, while proteobacteria have a preference for minimal media. Added to this, longer fermentation times are required for actinobacteria.²²

The most well-known host organism for lasso peptide production is *Escherichia coli*. However, it is not very well optimized for *Streptomyces* gene clusters.²³ For this reason, there are a large number of actinomycetes hosts for secondary metabolite production such as *Streptomyces coelicolor*,^{24,4} *Streptomyces lividans* or *Streptomyces albus* J1074,²⁵ but they are not optimized for lasso peptide production. They do not share the same regulatory signals for transcription and translation, and differences in codon usage may interfere in post-translational modifications. In *Streptomyces*, there are several promoter sequences that are not featured in *E. coli*.²⁶ A promoter is a DNA stretch that controls the rate at which the gene is

transcribed and the rate at which mRNA is degraded; summarizing, it controls the processing of enzyme expression.²⁷

In this work, new vectors containing the desired DNA isolated from our *Streptomyces* strains have been developed. Vectors are plasmids into which a foreign DNA fragment can be inserted for cloning purposes. As well, a new promoter was introduced in the DNA sequences to facilitate the recognition of the host organisms.

The first step was the DNA isolation of each strain, previously incubated in GYM media at 30 °C for 7 days. Once the DNA was extracted, it was amplified via PCR with the appropriate oligonucleotide primers (See Materials and Methods) as it is shown in Figure 81. Forward and reverse primers were designed for each strain with the same overhang, to perform later the Gibson-assembly cloning.²⁸ This type of cloning allows the joining of multiple DNA fragments in a single one without using restriction enzymes. Overhangs are DNA segments added to the primers that allow the overlap with the other DNA sequences during the cloning. They are complementary to the ends of linearized vector backbone, in this case to the *E. coli* plasmid (Figure 82). In addition, the length of the overlap sequence is dependent on the guanine-cytosine (GC) content of the sequence.²⁹ This has been pointed to the fact that the stop codon has a tendency towards adenine (A) and thymine (T) nucleotides, and, thus, the shorter the sequence, the higher the AT preference.³⁰

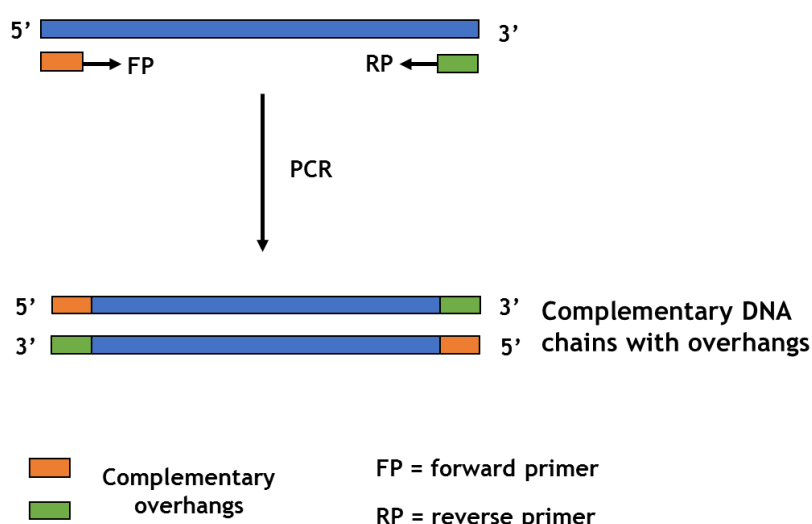


Figure 81: Schematic representation of overhang PCR

The first primers with overhang were designed using the standard parameters, which were; 18 - 25 base pairs, rich GC content (60 - 70%) and melting temperature 3-5 °C above the annealing temperature. Using these conditions, no DNA was

detected. Then, PCR conditions were optimized by changing DMSO concentrations (0, 5 and 10%), annealing temperatures (60, 65 and 72 °C) and buffers. However, no DNA amplification was detected. As a result, new long primers with higher melting temperature were designed. The main problem working with long primers is that their tendency to hybridize increases and makes them less specific. Added to this, the increase of the melting temperature implies a raise in the GC-content, which is not favorable for the amplification, due that it must be kept more or less constant. Despite these disadvantages, DNA of *S. monomycini* was amplified and scaled-up.

To discard the origin of the problem, new primers were designed without overhangs, making them shorter because maybe the problem would derive from the length of these overhangs, producing less specific primers. The problem of not using overhangs was that traditional cloning was required instead of Gibson-assembly cloning. The disadvantages of traditional cloning are that it is a longer process and requires the use of restrictions enzymes. For each DNA fragment it would be necessary to optimize the conditions to clone it into the vector. Following this methodology, no improvement in the DNA amplification was noticed. Thus, we turned back into the use of long primers with overhangs and new conditions were tested. Moreover, Q5 Polymerase, which is more specific for long and difficult amplifications was used instead of Phusion Polymerase. Finally, DNA amplification worked well for *S. leeuwenhoekii* C34 using long primers with overhangs, with 5% DMSO, HF buffer and DNA stock solution of template at 65 °C.

Afterwards, amplified DNA was cloned into pET41a expression vector via Gibson-assembly cloning using the protocol described in Materials and Methods. Ligated DNA was transformed into *E. coli* TOP10 cells (Figure 82). After overnight growth at 37 °C, some colonies were picked-up, followed by plasmid extraction, digestion and sequencing. Both plasmids were successfully created.

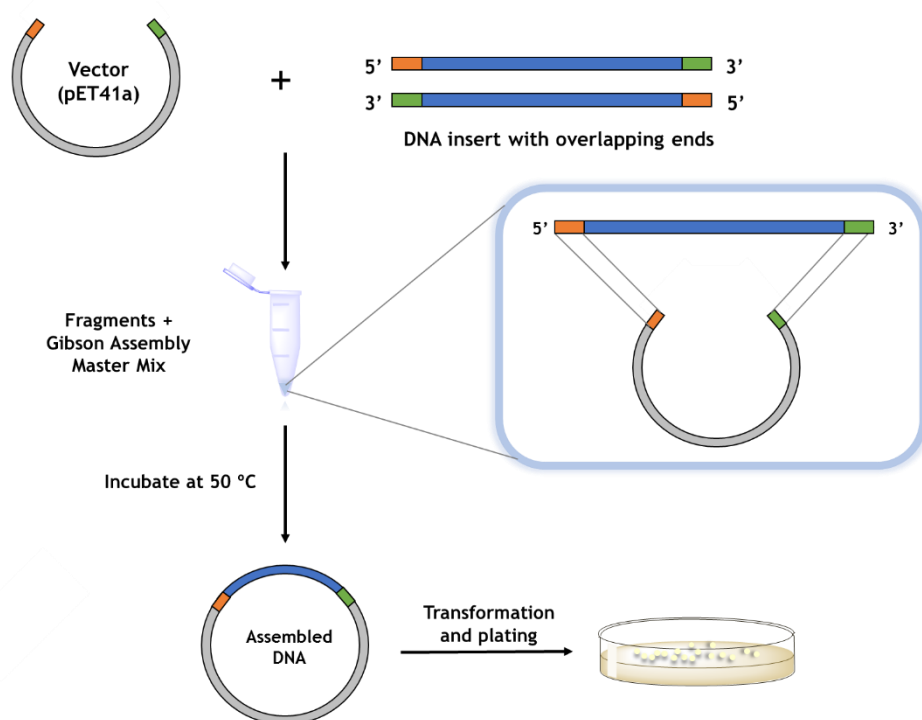


Figure 82: Schematic representation of Gibson Assembly cloning

Mutagenesis

Mutagenesis is a process by which the genetic information of an organism is changed, resulting in the onset of mutation. This process can occur spontaneously or after the exposure to mutagens. Mutations can also be induced experimentally by the use of genetic engineering techniques.³¹

Regarding the heterologous expression in *E. coli* of lasso peptides stemming from proteobacteria, it was shown that production could be highly improved through different means. On one hand, the production was optimized through exchange of the intergenic region between their respective precursor and processing enzyme genes with an *E. coli* optimized ribosomal binding site (RBS).³² Another way to improve lasso peptide yields was via the control of the precursor peptide through the use of strong inducible promoter, while the expression of the processing enzymes was controlled by the native promoter system.³³ Gene cluster of MccJ25 was the basis of all the mutations carried out in lasso peptides. It was isolated from *Escherichia coli* AY25³⁴ and it is the most studied lasso peptide because it can be produced easily by heterologous expression.^{35,19,36,37}

The following modifications were previously tested to be an efficient way to increase lasso peptide yield in astexin-1³⁸, capistrain³⁹ and caulosegnin.³² Three *in vitro* modifications based on MccJ25 enzymes were introduced into the plasmid, thus creating more similarity to the bacteria host. The first modification in the DNA sequence was the introduction of a terminator-promoter (T-P) in the intergenic region between A and C genes by Gibson-assembly (Figure 83). Additionally, the intergenic region between gene C-B1 and B1-B2 genes was modified with an optimized *E. coli* RBS resulting from enzyme *McjC* via site-directed ligase-independent mutagenesis (SLIM). SLIM is a methodology used to make specific and intentional changes to the DNA sequence of a gene.⁴⁰ The advantage of SLIM compared to cloning is that it does not require the use of ligases for plasmid circularization and avoids the formation of small transformable products.⁴¹ This methodology uses an inverse PCR amplification of the template by two tailed long primers and two short primers in a single reaction. The tailed primers contain the desired mutations on complementary overhangs. RBS could be introduced by SLIM because it is a small sequence (~15 bp), compared to the T-P (~200 bp), which had to be cloned. In this case, RBS sequence was used as an overhang and base primers were designed to bind to start codon (ATG) of the C-gene and the stop codon of A-gene (See Materials and Methods for primers sequences).

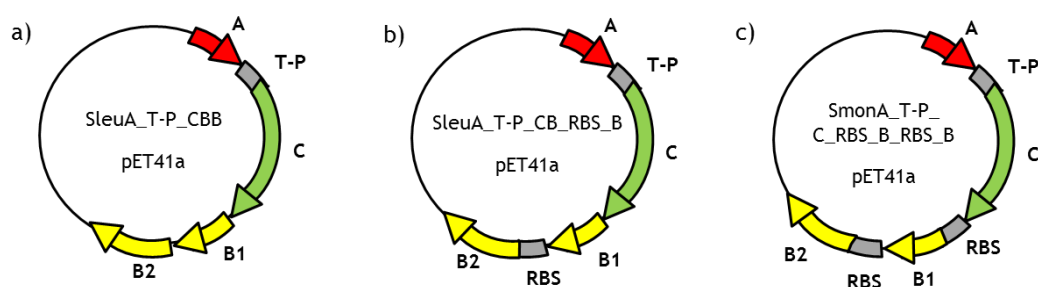


Figure 83: Schematic representation of gene cluster construction into pET41a expression vector. a) The intergenic region between A and D of chaxapeptin was replaced by a terminator-promoter sequence. b) The intergenic region between B1 and B2 of chaxapeptin was replaced by one ribosomal binding site (RBS) c) The intergenic regions of *S. monomykini* were replaced by a terminator-promoter and two RBS.

For the heterologous expression of our strains, plasmids bearing the correct sequences with T-P and RBS were transformed into *E. coli* BL21. M9 minimal medium was inoculated with LB-medium overnight culture of *E. coli* and incubated for 3 days at 20 °C and 1 day at 37 °C. Afterwards, they were harvested and the pellets extracts were analyzed by LC-MS. The expression of *S. monomykini* containing the T-P and the

two RBS (SmonA_T-P_C_RBS_B_RBS_B pET41a) did not produce any lasso. However, in the case of *S. leuvenhoekii* with the T-P and the second RBS (SleuA_T-P_CB_RBS_B pET41a), chaxapeptin was identified, but no isolated peak was detected based on UV intensity (Figure 83). This observation suggested that lasso production was very low. Nevertheless, we were able to produce for the first time chaxapeptin via heterologous expression.

Furthermore, heterologous expression of MccJ25,³³ capistruin²⁰ and sphingopyxin I⁵ was carried out. The plasmids that already contained the gene constructions were transformed into *E. coli* BL21 cells and subjected to fermentation in M9 minimal media. Then, they were harvested and the pellet and the supernatant were extracted, analyzed and purified by preparative HPLC. As mentioned before, the presence of D gene in MccJ25 and capistruin made the lasso peptide to be transported to the supernatant, while sphingopyxin I was extracted from the pellet. The sphingopyxin I heterologous production system was correctly optimized, producing very good yield. On the contrary, capistruin production system was not very well developed and the production yields were very low in both conditions. As expected, the best yield was found for MccJ25, as its original producing system is from *E. coli*.

In conclusion, only with the introduction of the T-P, lasso peptide production was not detected. So, more modifications had to be introduced to improve the production system. In the case of chaxapeptin, its production was improved with the introduction of the RBS between the intergenic region of B1 and B2. For capistruin, only with the introduction of the RBS between the intergenic region of A and B (capA_RBS_BCD pET41a) the production yield was improved compared to capABCD pET41a construct. Otherwise, the production yield of the sphingopyxin I construct containing the same modification, was vastly improved compared to the equivalent for capistruin. We were able to perform for the first time the heterologous expression for chaxapeptin production, but in low yields. The same occurred with capistruin, in which its plasmid was not well optimized and production yield was low, even incubating at two different conditions. These results indicated that the specific function of the amino acids in many lasso peptides was not clear, producing difficulties to select the appropriate position for further modifications.⁴² Future experiments should be performed on the optimization of plasmid constructions to improve production yields.

Chemoenzymatic approach

In this section we describe an initial approach for the chemoenzymatic synthesis for lasso peptides. It is already described in the introduction that isopeptidase (IsoP) is an enzyme that specifically hydrolyzes the isopeptide bond of lasso peptides to yield linear peptides, and it recognizes only lasso peptides and no unthreaded peptides.^{43,44} The question that we propose was: knowing that isopeptidase can only hydrolyze isopeptide bonds from lasso peptide, would it be possible to catalyze the isopeptide bond formation to provide the lasso peptide again?

To answer this question we focused on the isopeptidase from sphingopyxin I (Spi-IsoP), which was isolated and characterized by Fage and coworkers.⁴⁵ After the heterologous expression of sphingopyxin I, it was incubated with Spi-IsoP to linearize the lasso. The hydrolysis was fast and clean, total conversion was achieved in 30 minutes. Afterwards, the linear peptide was incubated with different amounts of enzyme with an initial peptide-enzyme ratio of 1:10 during different incubation times at 30 °C. In this case, the reaction was independent of the substrate concentration, compared to the direct reaction, in which the reaction rate was dependent on the sphingopyxin I concentration (Figure 84). Theoretically, enzyme activity is highest when substrate concentration is unlimited, for this reason it was expected that for short times the reaction was faster and more efficient. However, no conversion was observed at any time.

Changes in the amount of product formed over a specified period of time are dependent on the level of enzyme. Regarding this, enzyme amount was increased up to 50 times compared to the peptide concentration. After several hours, the enzyme precipitated, and no conversion was observed.

Further studies should be carried out by changing different parameters, like raising the temperature to increase the rate of an enzyme-catalyzed reaction, but bearing in mind that most enzymes are denatured around 40 °C. On the contrary, there are other enzymes that can resist up to 80 °C.⁴⁶

pH is also a variable factor. It has a considerable influence on the position of equilibrium. In this experiment, the reaction was performed at pH 7.5 because our first approach was to use the same conditions of the direct reaction, changing only

one parameter, which was the enzyme concentration. Moreover, this specific buffer was the one selected because it was used for the expression and purification of Spi-IsoP.

Additionally, as it was seen in the biosynthetic process for lasso peptides, a source of ATP was required for the lasso formation. Therefore, for the reverse process, an external source of energy would be necessary to overtake the energy gap.

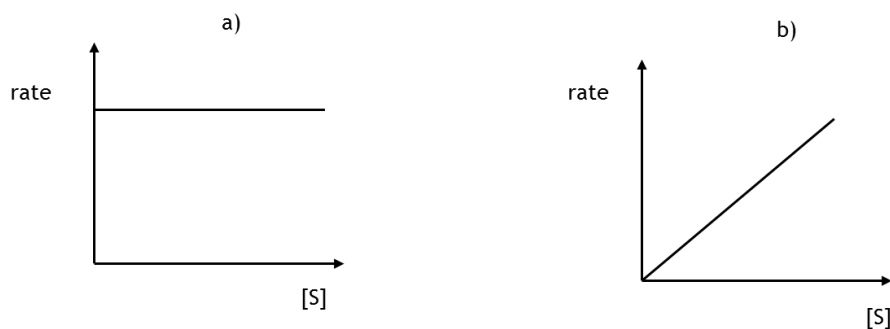


Figure 84: Reaction rate dependent versus substrate concentration. a) Order zero, b) First order.

Lasso formation is a challenging process and it is not clear that the reverse reaction is feasible because the final product is not entropically nor enthalpically favorable. It was observed that isopeptidase was not able to catalyze the isopeptide bond formation when changing enzyme concentration. So, changing only one parameter has been seen that was not feasible to achieve lasso formation. For this reason, high conditions to overtake the high energy gap are required. A good alternative to increase the reaction rate would be the combination of two parameters, such as, an increase of the temperature and the addition of an ATP source.

Discussion

Regarding the biosynthesis of new sungsanpin-like lasso peptide, from the genome mining we were able to identify ten promising strains, from which only seven were commercially available. Our search was based on the gene cluster of chaxapeptin, which was closer to sungsanpin. These strains by homologous expression were not able to produce any lasso peptide. Therefore, all efforts were based on the heterologous expression, in which we were able to produce for the first time chaxapeptin via heterologous expression. More modifications and experimentation will be required to optimize the plasmid construction to improve the production yields of chaxapeptin, in one case, and to produce new sungsanpin-like lasso peptides in the other.

References

- (1) Um, S.; Kim, Y.-J. J.; Kwon, H. H. C.; Wen, H.; Kim, S.-H. H.; Kwon, H. H. C.; Park, S.; Shin, J.; Oh, D.-C. C. *J. Nat. Prod.* **2013**, *76* (5), 873.
- (2) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. *Acc. Chem. Res.* **2015**, *48* (7), 1909.
- (3) Knappe, T. a.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. a. *J. Am. Chem. Soc.* **2008**, *130* (17), 11446.
- (4) Challis, G. L. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 219.
- (5) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. *Biopolymers* **2013**, *100* (5), 527.
- (6) Elsayed, S. S.; Trusch, F.; Deng, H.; Raab, A.; Prokes, I.; Busarakam, K.; Asenjo, J. A.; Andrews, B. A.; Van West, P.; Bull, A. T.; Goodfellow, M.; Yi, Y.; Ebel, R.; Jaspars, M.; Rateb, M. E. *J. Org. Chem.* **2015**, *80* (20), 10252.
- (7) Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. *Appl. Microbiol. Biotechnol.* **2012**, *95* (2), 451.
- (8) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25* (17), 3389.
- (9) Pan, S. J.; Rajniak, J.; Maksimov, M. O.; Link, a. J. *Chem. Commun.* **2012**, *48* (13), 1880.
- (10) Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A. *Chem. Sci.* **2014**, *5* (10), 4032.
- (11) Busarakam, K.; Bull, A. T.; Girard, G.; Labeda, D. P.; Van Wezel, G. P.; Goodfellow, M. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **2014**, *105* (5), 849.
- (12) Gomez-Escribano, J. P.; Castro, J. F.; Razmilic, V.; Chandra, G.; Andrews, B.; Asenjo, J. A.; Bibb, M. J. *BMC Genomics* **2015**, *16* (1), 485.
- (13) Needleman, S. B.; Wunsch, C. D. *J. Mol. Biol.* **1970**, *48* (3), 443.
- (14) Zhang, J.; Martin, C.; Shifflet, M. A.; Salmon, P.; Brix, T.; Greasham, R.; Buckland, B.; Chartrain, M. *Appl. Microbiol. Biotechnol.* **1996**, *44* (5), 568.
- (15) Pridham, T. G.; Hesseltine, C. W.; Benedict, R. G. *J. Biol. Chem. Can. J. Micro-biol* **1958**, *189* (2), 109.
- (16) Omura, S.; Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Takahashi, C.; Shinose, M.; Takahashi, Y.; Horikawa, H.; Nakazawa, H.; Osonoe, T.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M. *Proc. Natl. Acad. Sci.* **2001**, *98* (21), 12215.
- (17) Davis, N. K.; Chater, K. F. *Mol. Microbiol.* **1990**, *4* (10), 1679.

- (18) Clarke, D. J.; Campopiano, D. J. *Org. Biomol. Chem.* **2007**, *5* (16), 2564.
- (19) Duquesne, S.; Destoumieux-Garzón, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. *Chem. Biol.* **2007**, *14* (7), 793.
- (20) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. *Chem. Biol.* **2009**, *16* (12), 1290.
- (21) Zotchev, S. B. *J. Biotechnol.* **2012**, *158* (4), 168.
- (22) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. *Biopolymers* **2013**, *100* (5), 527.
- (23) Olano, C.; Lombó, F.; Méndez, C.; Salas, J. A. *Metab. Eng.* **2008**, *10* (5), 281.
- (24) Bentley, S.; Chater, K.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. a; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O’Neil, S.; Rabinowitsch, E.; Rajandream, M.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. a. *Nature* **2002**, *417* (6885), 141.
- (25) Galm, U.; Shen, B. *Expert Opin. Drug Discov.* **2006**, *1* (5), 409.
- (26) Lambertz, C.; Garvey, M.; Klinger, J.; Heesel, D.; Klose, H.; Fischer, R.; Commandeur, U. *Biotechnol. Biofuels* **2014**, *7* (1), 1.
- (27) Promoter | Learn Science at Scitable
<https://www.nature.com/scitable/definition/promoter-259> (accessed Nov 25, 2017).
- (28) Schiestl, R. H.; Petes, T. D. *Proc. Natl. Acad. Sci.* **1991**, *88* (17), 7585.
- (29) Parker, J. In *Encyclopedia of Genetics*; Elsevier, 2001; p 192.
- (30) Wuitschick, J. D.; Karrer, K. M. *J. Eukaryot. Microbiol.* **1999**, *46* (3), 239.
- (31) Crueger, A. In *Biotechnology Set*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2001; pp 4-45.
- (32) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2013**, *135* (1), 210.
- (33) Pan, S. J.; Cheung, W. L.; Link, A. J. *Protein Expr. Purif.* **2010**, *71* (2), 200.
- (34) Salomón, R. a; Farías, R. N. *J. Bacteriol.* **1992**, *174* (22), 7428.
- (35) Solbiati, J. O.; Ciaccio, M.; Farías, R. N.; José, E.; Moreno, F.; Salomón, R. a; Solbiati, O.; Fari, R. N. *J. Bacteriol.* **1999**, *181* (8), 2659.
- (36) Ducasse, R.; Yan, K. P.; Goulard, C.; Blond, A.; Li, Y.; Lescop, E.; Guittet, E.; Rebuffat, S.; Zirah, S. *ChemBioChem* **2012**, *13* (3), 371.

- (37) Yan, K. P.; Li, Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S. *ChemBioChem* **2012**, *13* (7), 1046.
- (38) Maksimov, M. O.; Link, A. J. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 333.
- (39) Maksimov, M. O.; Pan, S. J.; James Link, a. *Nat. Prod. Rep.* **2012**, *29* (9), 996.
- (40) DiStefano, J. K. *Disease Gene Identification*; DiStefano, J. K., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2011; Vol. 700.
- (41) Chiu, J. *Nucleic Acids Res.* **2004**, *32* (21), e174.
- (42) Zhao, N.; Pan, Y.; Cheng, Z.; Liu, H. *Amino Acids* **2016**, *48* (6), 1347.
- (43) Maksimov, M. O.; Link, A. J. *J. Am. Chem. Soc.* **2013**, *135* (32), 12038.
- (44) Maksimov, M. O.; Koos, J. D.; Zong, C.; Lisko, B.; James Link, A. *J. Biol. Chem.* **2015**, *290* (52), 30806.
- (45) Fage, C. D.; Hegemann, J. D.; Nebel, A. J.; Steinbach, R. M.; Zhu, S.; Linne, U.; Harms, K.; Bange, G.; Marahiel, M. A. *Angew. Chemie Int. Ed.* **2016**, *55* (41), 12717.
- (46) Daniel, R. M.; Dines, M.; Petach, H. H. *Biochem. J.* **1996**, *317* (1), 1.

CONCLUSIONS

Conclusions related to objective 1

- Sungsanpin and the branched cyclic analog (with Glu at position 3) were synthesized in good yields on solid phase.
- We were able to synthesize the first three sungsanpin bicyclic analogs mimicking the lasso structure via ester bond. A combination of solid-phase and in-solution reactions was used to accomplish the synthesis. A branched cyclic analog of them was synthesized in good yields by a simple and efficient solid phase strategy.
- Two alternative bicyclic peptides were synthesized, in one case keeping the sungsanpin sequence, and in the other introducing two residue modifications.

Conclusions related to objective 2

- The synthetic sungsanpin was characterized by a complete set of structural techniques that confirmed the presence of a branched cyclic structure. A preliminary 3D structure was elucidated using computational tools.
- The synthetic sungsanpin and the branched cyclic analog (with Glu at position 3) showed a totally different behavior by ion-mobility mass spectrometry (IM-MS), even though their sequences only differed in one residue. By *in silico* analysis it was demonstrated that this difference could be attributed to the different adopted conformations. The Glu3-Lys5 electrostatic interaction was still present in the gas-phase, thus stabilizing the lasso-like structure.
- From the three sungsanpin bicyclic analogs, only the one obtained by strategy A showed thermal and proteolytic stability. However, when the ester bond was hydrolyzed, the peptide lost the structure and the chemical properties. This fact revealed that the ester bond was not the responsible for the stabilization of the whole structure. This hypothesis was also confirmed by NMR.

- The bicyclic analog obtained by strategy A did not show a threaded tail, as shown by NMR characterization and NOE-restrained structural modeling. However, due to its high thermal and proteolytic stability, it could be a good candidate for epitope grafting or to undergo modifications in order to improve the activity.

Conclusions related to objective 3

- Electrostatic interactions—thus the presence of free charges in the peptide—are key for the inhibitory effect on the invasion of cancer cells. From the activity results of the two alternative bicycles, it was demonstrated that the amino group of the Lys5 side chain and the C-terminal carboxylic acid were more important for the biological activity than the second cyclization.
- The branched cyclic analog showed a reduced effect on the migration and invasion of cancer cells. This was due to the presence of three charges present in the structure, one more with respect to sungsanpin, produced by the free side chain of Glu3.
- The second cyclization did not provide the biological activity, as shown by the fact that from the three ester-bound bicyclic peptides, only the obtained by strategies C and D were active, while the bicyclic peptide A was not.

Regarding the last three objectives (4, 5 and 6) of the thesis, based on the biosynthesis of lasso peptides, two conclusions can be remarked:

- We were able for the first time to produce chaxapeptin via heterologous expression.
- Regarding the chemoenzymatic approach, changing the enzyme concentration was not enough to produce the lasso fold. An alternative process would be the combination of two parameters, such as temperature, pH, enzyme concentration and the addition of an ATP source.

EXPERIMENTAL SECTION

Materials and methods

For peptide synthesis and characterization

Solvents and reagents

Dichloromethane, dimethylformamide, methanol, diethyl ether were purchased from Scharlau; while acetone, acetonitrile and toluene were obtained from SDS. All the reagents employed in the present thesis have been purchased to suppliers Alfa Aesar, Bachem AG, Iris Biotech, Luxemburg Industries, Novabiochem, Panreac and Sigma-Aldrich.

Analysis and purification

Analytical HPLC

Analytical HPLC was performed on a Waters Alliance 2695 (Waters, Milford, MA) with an automatic injector and a photodiode array detector (Waters 2998 or Waters 996) and software Empower2. The columns used were a Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column and a Sunfire™ C18 3.5 μm (4.6 mm x 100 mm) reversed-phase analytical column run with linear gradients of ACN (0.036% TFA) into H₂O (0.045% TFA) over 8-min, and a Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column run with linear gradients over 40-min. UV detection was at 220 and 254 nm and the system was run at a flow rate 1 mL/min.

In the case of using a semi-analytical HPLC, the columns used were a Xbridge™ Prep BEH130 C18 5 μm (10 mm x 100 mm) reversed-phase analytical column and a Sunfire® C18 OBD™ Prep Column 5 μm (10 x 150 mm) reversed-phase column. Linear gradients of ACN (0.036% TFA) into H₂O (0.045% TFA) were run over 20-min. UV detection was at 220 and 254 nm and the system was run at a flow rate 3 mL/min.

UPLC

UPLC chromatograms were obtained on a Waters Acquity system (PDA e λ detector, sample manager FNT and Quaternary solvent manager) using an Acquity BEH C18 1.7 μm (2 mm x 50 mm) column. UV detection was at 220 and 254 nm, and linear

gradients of ACN (0.036% TFA) into H₂O (0.045% TFA) were run at a flow rate 0.6 mL/min over 2-min.

Semi-preparative HPLC

Semi-preparative RP-HPLC was carried on Waters instrument comprising a 2545 gradient module equipped with a sample manager module with automatic injector and fraction collector (Waters Alliance 2767), a Waters Alliance 2487 dual wavelength absorbance detector and a MassLynx v3.5 system controller. The column used was a XBridge® Prep C18 OBD™ 5 µm (19×100 mm) reversed-phase column. UV detection was at 220 and 254 nm and linear gradients of ACN (0.1% TFA) into H₂O (0.1% TFA) were run at a flow rate of 16 mL/min over 30 min.

HPLC-ESI MS

Mass spectra for some peptide and amino acids were performed on a Waters system Alliance 2695, equipped with a photodiode detector Waters 2998, an ESI-MS model Micromass ZQ and Masslynx v4.1 software (Waters). The column used was a Sunfire™ C18 3.5 µm (2.1 mm x 100 mm). UV detection was at 220 and 254 nm, and linear gradients of ACN (0.07% formic acid (FA)) into H₂O (0.1% FA) were run at a flow rate 1 mL/min over 8-min.

UPLC-ESI-MS

Mass spectra for some peptide and amino acids were performed on a Waters Acquity system (PDA eλ detector, sample manager FNT and Quaternary solvent manager) and an ESI-MS model SQ Detector2. The column used was an Acquity BEH C18 1.7 µm (2 mm x 50 mm). UV detection was at 220 and 254 nm, and linear gradients of ACN (0.07% FA) into H₂O (0.1% FA) were run at a flow rate 0.6 mL/min over 2-min.

MALDI-TOF

The determination of the molecular weight of some peptides was performed on a MALDI-TOF Applied Biosystem 4700 with a N₂ laser of 337 nm using α-cyano-4-hydroxycinnamic acid (ACH) matrix (10 mg/mL of ACH in ACN-H₂O (1:1, v/v) containing 0.1% TFA).

Sample preparation: 1 μ L of sample solution (0.5 - 2 mg/mL) mixed with 1 μ L matrix were seeded on the MALDI-TOF plate and air-dried.

CombiFlash® Rf TELEDYNE ISCO

The automatic purification of some peptides and amino acids was performed on a CombiFlash® Rf TELEDYNE ISCO by using prepacked Redisep® Rf Gold High Performance Column C18 50 g from Teledyne Technology Company. Liquid samples were charged directly into the column and eluted into the column with linear gradients of ACN (0.1% TFA) into H₂O (0.1% TFA) which were run at a flow rate of 40 mL/min. UV detection was at 220 and 254 nm.

Solid-phase peptide synthesis (SPPS)

General considerations

Solid-phase syntheses were performed manually in polypropylene syringes fitted with a polyethylene porous disc. All solvents and soluble reagents were removed by suction. Washings between deprotection, coupling and cyclization steps were carried out with DMF (5 x 30 s) and DCM (5 x 30 s) using 4 mL solvent/g resin for each wash. When not specified, all transformations and washes were performed at 25 °C. Short treatments were carried out with manual stirring while longer transformations took place in orbital shakers. All the peptides were synthesized on solid-phase using the Fmoc/*t*Bu protection strategy.

Automatic SPPS

Automatic SPPS was carried out in CEM Liberty Blue™ Microwave Peptide Synthesizer. Drain washings were performed with DMF (2 x 5 mL) with nitrogen gas agitation. Fmoc deprotection was carried out using 10% (w/v) piperazine and 0.1 M OxymaPure in a 9:1 mixture of NMP and EtOH. Coupling reactions were carried out with Fmoc-L-AA-OH (0.2 M), OxymaPure (1 M) and DIPCDI (0.5 M) in DMF. The mixtures were stirred for 3 min at 90 °C, except for cysteines, histidines and arginines, which were coupled at 50 °C for 10 min. After completion of the automatic synthesis, the peptidyl-resin was washed twice with 10 mL of DMF before being removed from the reactor.

Colorimetric tests

Kaiser test

The Kaiser or ninhydrin test is a colorimetric test that enables qualitative detection of free primary amines. It is commonly used in SPPS to monitor coupling and deprotection treatments.

The peptidyl-resin is washed with DCM and the solvent removed by suction. A small amount of resin beads is transferred to a glass tube and 6 drops of solution A and 2 drops of solution B are added. Then, the tube is incubated at 110 °C for 3 minutes. A dark blue or purple color (in the solution and/or in the resin beads) reveals the presence of free primary amines (positive test), whereas a yellow coloration ensures 99.5% coupling rate (negative test).

- Solution A: A solution of phenol (40 g) in absolute EtOH (10 mL) is prepared and heated until they are completely dissolved. After this, 65 mg of KCN solution in 100 mL of H₂O is prepared and mixed with 100 mL of freshly distilled pyridine. Both solutions are stirred with 4 g of Amberlite MB-3 resin for 45 minutes and finally filtered and mixed.
- Solution B: 2.5 g of ninhydrin is dissolved in 50 mL of absolute EtOH (50 mL). This solution must be kept in a flask protected from light.

Chloranil test

The chloranil test is a colorimetric test that enables qualitative detection of free secondary amines. It is commonly used in SPPS to monitor coupling and deprotection treatments onto proline or *N*-alkyl residues.

The peptidyl-resin is washed with DCM and the solvent removed by suction. A small amount of resin beads are transferred to a glass tube and 5 drops of saturated chloranil solution (2,3,5,6-tertavhloro-1,4-benzoquinone in toluene) and 20 drops of acetone are added. The resulting solution is shaken in a vortex for 5 min at room temperature. Blue to green beads mean free primary amines (positive test), whereas colorless or yellowish beads indicate the absence of amines (negative test).

De Clercq test¹

The De Clercq test, also known as the *p*-nitrophenyl ester test, is a colorimetric test used in SPPS that enables qualitative detection of free secondary and primary amines with a higher sensitivity than the chloranil test.

The peptidyl-resin is washed with DCM and the solvent removed by suction. A small amount of resin beads are transferred to a glass tube and 5 drops of the De Clercq reagent are added. Then, the tube is incubated at 70 °C for 10 minutes. The supernatant is removed with a Pasteur pipette and the resin beads washed with MeOH (x 3), DMF (x 3) and DCM (x 1). Red beads means free amines (positive test), whereas colorless beads indicates the absence of amines (negative test).

- De Clercq reagent preparation: To a solution of disperse red 1 (6.28 mg, 20 mmol) and Rh₂(OAc)₄ (150 mg, 0.34 mmol) in DCM-toluene (1:1, 200 mL), a solution of ethyl diazoacetate (8.4 mL, 80 mmol) in toluene (40 mL) was added at 40 °C over 1 h. The reaction mixture was stirred overnight at room temperature and purified by flash chromatography. The product obtained and KOH (4.0 g, 62.5 mmol) were dissolved in MeOH-toluene (4:3:1, 370 mL) and the resulting mixture was refluxed under N₂ for 1.5 h. The mixture was concentrated to 50 mL under reduced pressure, acidified with 25 mL of 10% HCl and further diluted with 120 mL of H₂O. The aqueous phase was extracted with DCM (x 4), the combined organic extracts were washed with MgSO₄. Finally, the purified product and *p*-nitrophenol (0.834 g, 6 mmol) were dissolved in pyridine-DCM (1:1.2, 220 mL), the resulting mixture was cooled down to -15 °C and a solution of POCl₃ in DCM (1 mL in 10 mL of DCM) slowly added over 1 h. The temperature was slowly raised to room temperature. After 4 h, the reaction mixture was poured into a mixture of DCM (250 mL) and ice cold H₂O (200 mL). The aqueous phase was extracted two times with DCM and the combined organic layers were washed with a sodium bicarbonate solution and with H₂O. The final *p*-nitrophenyl ester of disperse red 1 was dried over MgSO₄ and concentrated under *vacuo*. It is used as a 0.02 M solution in ACN.

Ellman test²

The Ellman test, also known as DTNB (5,5'-dithio(2-nitrobenzoic acid)) is a colorimetric test used in SPPS and in solution, that enables qualitative detection of thiol groups.

The peptidyl-resin is washed with DCM and the solvent removed by suction. A small amount of resin beads are transferred to a glass tube and 4-5 drops of the Ellman reagent are added. The resulting solution is shaken in a vortex for 3 min at room temperature. A yellow-orange solution means free thiols (positive test), whereas a colorless solution indicates the absence of thiols (negative test).

- Ellman reagent preparation: 5,5'-dithio(2-nitrobenzoic acid) dissolved in aqueous TRIS solution (1 M at pH 8).

Conditioning of the resin and incorporation of the first amino acid

Two different resins were used to synthesize all the peptides present in this thesis: 2-chlorotrityl chloride resin (2-CTC) and TentaGel® S- NH₂ resin with a substitution loading of 0.20 - 0.60 mmol/g, the latter with 3-(4-hydroxymethylphenoxy)propionic acid (HMPBA linker). Peptide cleavage from both resins provides the peptide with an acid function at the C-terminus. In both cases, prior to the first amino acid coupling, the resin was swollen in DCM for 3 min and subsequently washed with DMF (5 x 30 s).

Depending on the type of resin used, the following procedures for the coupling of the first amino acid and the capping? were followed:

	2-CTC	TentaGel
Coupling	1 eq Fmoc-AA-OH, 10 eq DIEA in DCM, 1 h	8 eq Fmoc-AA-OH, 8 eq DIPCDI, 0.5 eq DMAP in DCM-DMF (1:1), 2 h + 3 h
Capping	Addition of MeOH (0.8 mL/g resin) to the reaction mixture.	5 eq Ac ₂ O, 5 eq DIEA in DMF, 30 min.

Firstly, in the case of the TentaGel resin, the linker was attached to the resin through an amide bond using the coupling conditions: linker-DIPCDI-OxymaPure (3:3:3) in DMF for 1.5 h.

Elimination and quantification of the Fmoc group

The Fmoc group was removed before each coupling by treating the resin with piperidine-DMF (1:4, v/v). For the deprotection of primary amines, the resin was treated for 1 x 1 min and 1 x 3 min, while for secondary amines or more complex synthesis, a longer time was required, 1 x 1 min and 1 x 5 min or 1 x 1 min, 1 x 5 min, 1 x 10 min). After deprotection, the resin was washed with DMF (5 x 30 s) and DCM (5 x 30 s).

In order to measure the loading capacity of the resin, after coupling the first amino acid, piperidine washes were collected and measured by UV spectroscopy at 290 nm. The loading capacity was calculated using Equation 1:

$$X = \frac{A \cdot V}{\varepsilon \cdot m \cdot l} \quad \text{Equation 1}$$

where X is the loading of the resin, A is Fmoc absorbance at 290 nm, V is the dilution volume of solvent (in mL), ε is the molar extinction coefficient of Fmoc at 290 nm ($5800 \text{ M}^{-1} \text{ cm}^{-1}$), m is the mass of resin (in mg) and l is the optical path (typically, 1 cm).

Peptide chain elongation

The elongation of the peptide chain can be achieved by using several different coupling conditions. Method 1 was used as a default for all Fmoc-aa-OH couplings. All the couplings were performed at 25 °C. When the first coupling was not complete, a second coupling was performed. When the recoupling was unsuccessful or for the coupling of synthetic or more hindered amino acids, Method 2 was used.

Method 1:

Step	Treatment	Conditions
1	Washes	DCM (5 x 30 s)
2	Deprotection	Piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min)
3	Washes	DMF (5 x 30 s) and DCM (5 x 30 s)
4	Coupling	5 eq Fmoc-AA-OH, 5 eq DIPCDI, 5 eq OxymaPure in DMF, 30 min
5	Washes	DMF (5 x 30 s) and DCM (5 x 30 s)
6	Colorimetric test	Kaiser test, chloranil test or De Clercq, as needed

Method 2:

Step	Treatment	Conditions
1	Washes	DCM (5 x 30 s)
2	Deprotection	Piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min)
3	Washes	DMF (5 x 30 s) and DCM (5 x 30 s)
4	Coupling	3 eq Fmoc-AA-OH, 3 eq HATU, 3 eq DIEA in DMF, 1 h
5	Washes	DMF (5 x 30 s) and DCM (5 x 30 s)
6	Colorimetric test	Kaiser test, chloranil test or De Clercq, as needed

Allyl and Alloc group elimination

The allyl ester and the Alloc group were removed by treatment with a catalytic amount of Pd(PPh₃)₄ (0.1 eq) in the presence of the scavenger PhSiH₃ in DCM for 3 x 15 min. Between the treatments, the resin was washed with DCM (5 x 30 s) and then with DMF (5 x 30 s) and DCM (5 x 30 s). Additional washes with a 0.02 M solution of sodium diethyldithiocarbamate in DMF may be needed to remove all palladium traces.

pNZ removal and PoraPak purification

The pNZ group was removed by reduction with 6 M SnCl₂ in DMF containing 1.6 mM HCl/dioxane. It can be carried out in solid-phase or in solution.

In solution, the reaction mixture was then, directly purified with PoraPak Rxn RP cartridges with an elution gradient described below. The fractions of pure product were collected and lyophilized.

Step	Solvent	Time
Conditioning of the column	MeOH and H ₂ O	x 3
Product loading	Some H ₂ O could be added	
	H ₂ O	x 5
Elution	H ₂ O-ACN- (3:1)	x 4
	H ₂ O-ACN- (1:1)	x 4
	ACN	x 4

Otherwise, on solid-phase, the SnCl₂ solution was added to the resin for 2 x 45 min.

Cleavage of the peptide from the resin

- 2-Chlorotrytil chloride resin (2-CTC)

Cleavage from 2-CTC was achieved by acidolytic treatment in very mild TFA conditions to furnish the completely protected peptide, therefore enabling cyclizations in solution. After washing the resin with DCM (5 x 30 s), the resin could be treated with TFA-DCM (4:96) (5 x 1 min) and then washed with DCM to obtain the protected peptide. Otherwise, the resin could be treated with a high acid content, a TFA-TIS-H₂O (95:2.5:2.5) solution, to perform the cleavage and the side-chain deprotection in one step.

- Aminoethyl resin (TentaGel S-NH₂)

Cleavage from a Wang-type resin functionalized with the HMPBA linker requires harsh acidic conditions and furnishes the completely unprotected peptide. The cleavage cocktail is chosen according the sensitiveness of the residues present in the peptide sequence and their protecting groups. For all the synthesis performed in the present thesis using this resin, after washes with DCM (5 x 30 s), the polymeric support was treated with a TFA-TIS-H₂O (95:2.5:2.5) solution for 1 h.

Basic treatment

Bicyclic analog A

2.8 mg (1.7 μmol , 1 eq) of pure bicyclic analog A was dissolved in 1 mL of H₂O-ACN (1:1) and 80 μL of 0.5 M NaOH was added. The reaction mixture was stirred for 30 min at room temperature. The crude was purified by analytical RP-HPLC using Xbridge™ Prep BEH130 C18 5 μm (10 mm x 100 mm) reversed-phase analytical column and a linear gradient (30% to 55% over 25 min) of ACN with a flow rate of 2.0 mL/min to obtain 1.0 mg of pure unthreaded analog A (35%) with a purity > 99.9%.

Chaxapeptin, sphingopyxin I and MccJ25

0.1 mg of pure lasso peptides was dissolved in 30 μL H₂O-ACN (1:1) and 10 μL of 0.5 M NaOH was added. The reaction mixture was stirred at 40 °C and after 8, 16, 24 and 48 h an aliquot of sample was neutralized and analyzed by RP-HPLC and LC-MS. 16 h was found to be the optimum time. So, the reaction was neutralized with 5 μL of 0.5 M HCl and the crude was straightaway purified by analytical HPLC.

Scale-up of MccJ25

5.3 mg (2.5 μmol , 1 eq) of native MccJ25 was dissolved in 400 μL of H₂O-ACN (1:1). Then, 30 μL of NaOH 0.5 M (pH 12) was added and the mixture was stirred overnight at 40 °C. Afterwards, the reaction was neutralized by adding 10 μL of HCl 0.5 M and purified by analytical RP-HPLC using Sunfire® C18 OBD™ Prep Column 5 μm (10×150 mm) reversed-phase column and a linear gradient (30% to 35% over 15 min) of ACN with a flow rate of 2.0 mL/min to obtain 1.50 mg of pure hydrolyzed MccJ25 and 1.24 mg of pure SM MccJ25 with a purity > 99.9%.

NMR spectroscopy

¹H and ¹³C NMR spectra were recorded on a Varian MERCURY 400 (400 MHz for ¹H NMR, 100 MHz for ¹³C NMR) spectrometer for organic small molecules and on a Bruker 600 Avance II Ultrashield, provided with a cryoprobe TCI (600 MHz for ¹H NMR, 150 MHz for ¹³C NMR) spectrometer for peptidic samples. Chemical shifts (δ) are

expressed in parts per million (ppm) downfield tetramethylsilane (TMS). Coupling constants are expressed in Hertz (Hz).

Structural analysis of the synthetic sungsanpin

Pure peptide (1.75 mg) was dissolved in 550 μL of pyridine- d^5 . The spectrum was recorded at 25 $^\circ\text{C}$ using a Bruker 600 MHz NMR spectrometer equipped with a cryogenic probe. Residue assignments were obtained from 2D total correlated spectroscopy (TOCSY), while 2D nuclear Overhauser effect spectroscopy (NOESY) permitted sequence specific assignments. ^{13}C resonances were assigned from 2D ^{13}C - ^1H HSQC spectra. The TOCSY and NOESY mixing times were 70 and 400 ms, respectively.

Structural analysis of the bicyclic analog A

Pure peptide (0.77 mg) was dissolved in 550 μL of $\text{dms}\text{-d}^6$. The ^1H spectrum was recorded at 25, 30 and 35 $^\circ\text{C}$ using a Bruker 600 MHz NMR spectrometer equipped with a cryogenic probe. The 2D spectra were recorded at 25 $^\circ\text{C}$. Residue assignments were obtained from TOCSY. ^{13}C resonances were assigned from 2D ^{13}C - ^1H HSQC spectra. The TOCSY and NOESY mixing times were 70 and 400 ms, respectively.

Another batch of bicyclic peptide (0.83 mg) was dissolved in $\text{CD}_3\text{CN-H}_2\text{O}$ (1:5). First, the ^1H spectrum and TOCSY were recorded at 5 $^\circ\text{C}$ using a Bruker 600 MHz NMR spectrometer equipped with a cryogenic probe. Then, the temperature of ^1H was increased to 15 and 25 $^\circ\text{C}$. Final 2D spectra were recorded at 25 $^\circ\text{C}$. Residue assignments were obtained from TOCSY, while NOESY allowed sequence specific assignments. The TOCSY and NOESY mixing times were 70 and 400 ms, respectively.

Structural analysis of the unthreaded analog A

The unthreaded peptide (1.0 mg) was dissolved in 550 μL of $\text{dms}\text{-d}^6$. The spectrum was recorded at 25 $^\circ\text{C}$ using a Bruker 600 MHz NMR spectrometer equipped with a cryogenic probe. Residue assignments were obtained from TOCSY. The TOCSY mixing time was 70 ms.

Ion-mobility mass spectrometry (IM-MS)

IM-MS experiments were performed using a Synapt G1-HDMS mass spectrometer (Waters, Manchester, UK). 0.1 mg was reconstituted with 500 μ L of water (100 μ M). Sample were diluted 1/10 in H₂O-ACN (1:1) with 4% (FA) to obtain 10 μ M, 1/10 in H₂O-ACN (1:1) with 4% FA and 200 mM sulfolane or 1/10 in H₂O/ACN (1:1) with 4% FA and 200 mM or 400 mM sulfolane. Sample solutions were directly injected and were infused by automated chip-based nanoelectrospray using a Triversa Nanomate system (AdvionBioSciences, Ithaca, NY, USA) as the interface. The ionization was performed in positive mode using a spray voltage and a gas pressure of 1.75 kV and 0.5 psi, respectively. Cone voltage, extraction cone and source temperature were set to 40 V, 3 V and 20 °C, respectively. Trap and transfer collision energies were set to 6 V and 4 V, respectively. The pressure in the Trap and Transfer T-Wave regions were $5.99 \cdot 10^{-2}$ mbar of Ar and the pressure in the IMS T-Wave was 0.477 mbar of N₂. Trap gas and IMS gas flows were 8 and 25 mL/sec, respectively. The travelling wave used in the IMS T-Wave for mobility separation was operated at a velocity of 250 m/sec. The wave amplitude was a linear ramp from 1.0 to 28.0 V. The bias voltage for entering in the T-wave cell was 15 V. The instrument was calibrated over the m/z range 200-3000 Da using a solution of cesium iodide. MassLynx version 4.1 SCN 704 and Drift scope version 2.4 softwares were used for data processing. Ion mobility data analysis was performed with Driftscope software vs. 2.4 integrated in MassLynx software.

MS/MS analysis

MS² and MS³ analysis were performed on LTQ-FT Ultra mass spectrometer (Thermo Scientific). 0.1 mg was reconstituted with 500 μ L of water (100 μ M). 10 μ L (100 μ M) of sample were diluted 1/1 with ACN (1% FA) to obtain 20 μ L of 50 μ M. Sample solutions were directly injected and were infused by automated chip-based nanoelectrospray using a Nanomate system (AdvionBioSciences, Ithaca, NY, USA) as the interface. The ionization was performed in positive mode using a spray voltage and a gas pressure of 1.75 kV and 0.5 psi, respectively. Capillary temperature and voltage were set up to 200 °C and 44 V, respectively. The fragmentation technique used was CID (collision induced dissociation). The instrument was calibrated over the m/z range 200-2000 Da. Data was acquired with Xcalibur software, vs.2.0SR2 (ThermoScientific).

Crystallography

Lyophilized synthetic sungsanpin was dissolved in H₂O-ACN (1:1) at 11.4 mg/mL. On the other hand, the bicyclic peptide from strategy A was dissolved in H₂O-DMSO (1:1) at 1.4 mg/mL. Crystallization screening conditions were based on the commercially available Hampton index, such as crystal screen I and II, Index and SaltRX. Additionally, wizard I and II and Clear Strategy™ I and II were purchased from Molecular Dimensions and JBS Crystallization Freshman Kit-Junior from Jena Bioscience. Crystal screens I and II and wizard I and II were the two initial screening packs tested. Crystals were grown by sitting-drops vapor diffusion at 4 and 20 °C in 96-well plates using PHOENIX/RE from Art Robbins Instruments.

Racemic solutions of synthetic sungsanpin were prepared in H₂O with 35% DMSO at 15.5 mg/mL, 8.8 mg/mL and 4.4 mg/mL (15.5 mg/mL of L-peptide and 15.5 mg/mL of D-peptide). For the crystallization screening conditions, the same commercially available kits were used and the same procedure described above was followed.

In order to optimize the screening conditions, a set of different conditions was prepared in 24-well plates. Crystals were grown by sitting-drops vapor diffusion at 20 °C. The drops were generated by mixing manually 1 µL of peptide solution with 1 µL of reservoir solution and placed against 500 µL of reservoir solution. The crystals were transferred to cryoprotectant (reservoir solution plus 20% glycerol (v/v)) for a few seconds and flash-frozen in liquid nitrogen. Synchrotron X-ray data was collected at ALBA Synchrotron (Barcelona, Spain).

Computational experiments

Macrocylic Conformational Sampling³

The coordinates for the synthetic sungsanpin and the bicyclic analog A were extracted from the crystal structure of chaxapeptin (PDB code: 2N5C). Then, Leu7, Asn13, Phe14 and Phe15 were manually mutated to Ile, Ser, Trp and Leu, respectively, and the peptide was manually cyclized through the side chain of Asp8 and the ester bond through the side chains of Glu3 and Ser13. Additionally, the tail was manually unthreaded from the ring in the other bicyclic analog used as a starting structure. On the other hand, for the synthetic sungsanpin, the tail was unthreaded

and only was cyclized through the side chain of Asp8. Constraints from the NOESY cross-peaks were applied to each structure. The resulting structures were used as input for the Macrocycle Conformational Sampling module of Schrödinger, and conformations were generated using the OPLS-2005 force field and GBSA for electrostatic treatment. Conformations were kept when energies were below 10 kcal/mol, and redundant conformations (RMSD > 0.75 Å) were removed. For each peptide, 5000 simulation cycles were employed with 5000 LLMOD search steps.

Circular dichroism (CD)

CD spectra were recorded using a Jasco 810 UV-Vis spectropolarimeter, equipped with a CDF 426S/426L Peltier temperature controller. Peptide samples were dissolved in MeOH, and spectra were recorded at 200 μM. One sample was dissolved in H₂O with 5% TFE. The following parameters were used: sensitivity (standard, 100 mdeg), start (260 nm), end (190 nm), data pitch, (0.5 nm), scanning mode (continuous), scanning speed (200 nm/min), response (1 s), band width (1.0 nm), and accumulation (3). A blank spectrum of the buffer was subtracted from all recordings, and molar ellipticity values were calculated from experimental ellipticity using the Equation 2:

$$\theta = \frac{\theta_{exp} \cdot 10^6}{b \times C \times n} \quad \text{Equation 2}$$

where θ is the molar ellipticity in deg·cm²·dmol⁻¹, θ_{exp} is the measured ellipticity in mdeg, b is the optical path length in mm, C is the peptide concentration in μM and n is the number of residues in the peptide. After unit conversion, the spectra was smoothed using the Savitzky-Golay method (convolution width = 21) and taken to zero at the far-UV region ($\lambda = 260$ nm).

Capillary electrophoresis (CE)

CE was performed on a CE Agilent G1600AX in capillary cartridges (fused silica 50 μm i.d. x 49 cm length and 40.05 cm to detection). The capillary was equilibrated for 5 min with running buffer 50 mM sodium phosphate (pH 2.0) - ACN (9:1) and then, 200 μL of sample (230 μM) dissolved in the same buffer was injected at 25 °C, 5 s (50 mbar) pressure injection, separation for 40 min at 20 kV and detection at 214 nm.

For the biosynthesis of lasso peptides

Strains and general methods

Streptomyces lydicus (DSM 40461), *Streptomyces leeuwenhoekii* C34 (DSM 42122), *Streptomyces vitaminophilus* (DSM 41686) were purchased from the German Collection for Microorganisms and Cell Cultures (DSMZ). *Streptomyces sp.* WC-3641 (IMRU 3641) was purchased from the Waksman Institute of Microbiology, Rutgers, The State University of New Jersey, Piscataway, NJ, USA (IMRU). *Streptomyces sp.* S-118 (NRRL S-118) and *Streptomyces monomycini* (NRRL B-24309) were purchased from the Agricultural Research Service Culture Collection, National Center for Agricultural Utilization Research, US Department of Agriculture, Peoria, IL, USA (NRRL). *E. coli* TOP10, which was used for cloning, and *E. coli* BL21 (DE3), which was used for heterologous expression, were purchased from Invitrogen. Oligonucleotides were purchased from Sigma Aldrich. Carboxypeptidase Y was purchased from Alfa Aesar. Restriction enzymes, Phusion polymerase and Q5 Polymerase were purchased from New England Biolabs. DNA Sanger Sequencing was performed by GATC Biotech AG (Konstanz) to determine the identity of constructed plasmids and mutants.

Analysis and purification

HPLC-MSD

Analysis of pure lasso peptides was performed with a 1100 series MSD (Hewlett-Packard) coupled with a microbore 1100 HPLC system (Agilent). For sample separation, an EC 125/2 Nucleodur 100-3 C18 column at room temperature and a flow rate of 0.2 mL/min was used.

Preparative HPLC

Lasso peptides were purified by preparative HPLC using a microbore 1100 HPLC system (Agilent) with a VP 250/21 Nucleodur C18 Htec 5 μ m column (Macherey-Nagel) at room temperature and a flow rate of 18 mL/min. UV detection was at 215 nm and linear gradients of MeOH (0.05% FA) into H₂O (0.045% FA) were run at a flow rate of 18 mL/min over 30 min for the first-round purification. For the second-round

purification, linear gradients of ACN (0.1% TFA) into H₂O (0.1% TFA) were run at a flow rate of 18 mL/min over 30 min.

Mass Spectrometric Analysis

Mass spectrometric analysis of the extracts was done using a LTQ-FT ultra-instrument (Thermo Fisher Scientific) connected to a microbore 1100 HPLC system (Agilent). For detection, the UV absorption at 215 nm was recorded. Separation was achieved by using a CC 125/2 Nucleosil 300-8 C18 column (Macherey-Nagel) and applying a gradient of H₂O (0.1% TFA) and ACN (0.1% TFA) at 40 °C and a flow rate of 0.2 ml/min: holding 2% B for 2 min, followed by a linear increase from 2% to 30% B in 18 min, a subsequent increase from 30% to 90% B in 15 min and holding 95% B for an additional 2 min. Data was acquired with Xcalibur software (ThermoScientific).

Primers

First design of the primer of all strains. Gibson-assembly overhang is underlined:

Name	Tm(°C)	Sequence (5'-3')
FP_S118_ACBB_Gibson	69.0	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG</u> CGC GAA AGT GGC GAC ATC GAA CAG GCC
RP_S118_ACBB_Gibson	70.4	<u>GAG CCA CCA CCG GTA CCC AGA TCT TCA</u> CGG CGA CGC GTG GGG CGG GAC GAC
FP_SWC3641_ACBB_Gibson	69.5	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG</u> GAC GAG CAG ACC GAG TTC TCG ACC ACC GAG
RP_SWC3641_ACBB_Gibson	70.7	<u>GAG CCA CCA CCG GTA CCC AGA TCT TCA</u> CCG GTT CCG TTC GTC CGG CAC CGG CGG
FP_Smon_ACBB_Gibson	72.7	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG</u> CAG GAC GTA CAG GAA CGT GCC GAG GCG ACC GAG
RP_Smon_ACBB_Gibson	71.2	<u>GAG CCA CCA CCG GTA CCC AGA TCT CTA</u> TGC CAC CGG CTC TCC CGG CGG CAC G
FP_Slyd_ACBB_Gibson	70.8	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG</u> GAA GAG CAG AAC GAG CTC TCG AAC GTA GAG CCC

RP_Slyd_ACBB_Gibson	71.8	<u>GAG CCA CCA CCG GTA CCC AGA TCT CTA</u> TCC CTC GCC GCT CCG CAG CGG AGG TAT G
FP_Sleu_ACBB_Gibson	70.4	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG</u> GAA CCC CAG ATG ACT GAG CTT CAG CCG GAG GC
RP_Sleu_ACBB_Gibson	69.3	<u>GAG CCA CCA CCG GTA CCC AGA TCT TCA</u> GGA AAC CGG AGG AAC GGT CAG CAA AGG CCG
FP_pET41a_BglII_Gibson	65.1	AGA TCT GGG TAC CGG TGG TGG CTC CGG
RP_pET41a_NdeI_Gibson	62.4	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG

Design of long primers. Gibson-assembly overhang is underlined:

Name	Tm(°C)	Sequence (5'-3')
FP80_SNRRLS-118_ACBB	80.2	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG CGC</u> GAA AGT GGC GAC ATC GAA CAG GCC GAG GTG TAC GAG CCG CCG GTG CTG
RP80_SNRRLS-118_ACBB	80.4	<u>GAG CCA CCA CCG GTA CCC AGA TCT TCA</u> CGG CGA CGC GTG GGG CGG GAC GAC GAG GAG TGG CCG GTA GGC GGC
FP80_SWC-3641_ACBB	80.2	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG GAC</u> GAG CAG ACC GAG TTC TCG ACC ACC GAG CCG TAC GCG CCC CCG ACG CTG
RP80_SWC-3641_ACBB	77.5	<u>GAG CCA CCA CCG GTA CCC AGA TCT TCA</u> CCG GTT CCG TTC GTC CGG CAC CGG CGG GAT GGA CAT CAG TGT CTT GAA GTG CCC
FP80_Smon_ACBB	80.2	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG CAG</u> GAC GTA CAG GAA CGT GCC GAG GCG ACC GAG TAC GAG GCG CCG GAG CTG
RP80_Smon_ACBB	79.7	<u>GAG CCA CCA CCG GTA CCC AGA TCT CTA</u> TGC CAC CGG CTC TCC CGG CGG CAC GCT CAG CAA CGG CTG GTA GAG
FP80_Slyd_ACBB	80.5	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG GAA</u> GAG CAG AAC GAG CTC TCG AAC GTA GAG CCC TAC GCG CCC CCG ATG CTG GTC GAG

RP80_Slyd_ACBB	80.0	<u>GAG CCA CCA CCG GTA CCC AGA TCT CTA</u> TCC CTC GCC GCT CCG CAG CGG AGG TAT GGA GAT CAG CGG CTT GAA GTG GCC
FP80_Sleu_ACBB	80.0	<u>TAA CTT TAA GAA GGA GAT ATA CAT ATG GAA</u> CCC CAG ATG ACT GAG CTT CAG CCG GAG GCC TAC GAG GCG CCG TCC CTC ATC GAG
RP80_Sleu_ACBB	77.8	<u>GAG CCA CCA CCG GTA CCC AGA TCT TCA</u> GGA AAC CGG AGG AAC GGT CAG CAA AGG CCG ATA GTG ACC GTC CGG GTA CGG CTC G

Long primers without overhang:

Name	Tm(°C)	Sequence (5'-3')
FP80_SNRRLS-118_ACBB_noOH	80.2	ATG CGC GAA AGT GGC GAC ATC GAA CAG GCC GAG GTG TAC GAG CCG CCG GTG CTG
RP80_SNRRLS-118_ACBB_noOH	80.4	TCA CGG CGA CGC GTG GGG CGG GAC GAC GAG GAG TGG CCG GTA GGC GGC
FP80_SWC-3641_ACBB_noOH	80.2	ATG GAC GAG CAG ACC GAG TTC TCG ACC ACC GAG CCG TAC GCG CCC CCG ACG CTG
RP80_SWC-3641_ACBB_noOH	77.5	TCA CCG GTT CCG TTC GTC CGG CAC CGG CGG GAT GGA CAT CAG TGT CTT GAA GTG CCC
FP80_Slyd_ACBB_noOH	80.5	ATG GAA GAG CAG AAC GAG CTC TCG AAC GTA GAG CCC TAC GCG CCC CCG ATG CTG GTC GAG
RP80_Slyd_ACBB_noOH	80.0	CTA TCC CTC GCC GCT CCG CAG CGG AGG TAT GGA GAT CAG CGG CTT GAA GTG GCC
FP80_Sleu_ACBB_noOH	80.0	ATG GAA CCC CAG ATG ACT GAG CTT CAG CCG GAG GCC TAC GAG GCG CCG TCC CTC ATC GAG
RP80_Sleu_ACBB_noOH	77.8	TCA GGA AAC CGG AGG AAC GGT CAG CAA AGG CCG ATA GTG ACC GTC CGG GTA CGG CTC G

Primers for the modifications of Terminator Promoter. Optimal RBS region is underlined. GTG start codon was mutated to ATG (in black):

Name	Tm(°C)	Sequence (5'-3')
FP_TermProm_Gibson	64.5	GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CC
RP_TermProm_Gibson	64.0	<u>ATGTGACCTCCTTA</u> AGT TAT GCC TTT ACA CTC AAT TGG AAT AAT AAA ACT TCT GAA TTT ATC
FP_Smon_TermProm_Gibson	70.0	<u>CATAACTTAAGGAGGTCACAT</u> ATG GAT TTC GTC GTC TTC CCC GAC CAT CCC GCG ACC
RP_Smon_TermProm_Gibson	70.6	<u>TCAGCTAATTAAGCTGGTACC</u> TCA CCA GAA GGT GCT GTG CAT CCC GCC CCA GTT GTC
FP_Sleu_TermProm_Gibson	70.4	<u>CATAACTTAAGGAGGTCACAT</u> ATG AGC ACC GGG TTT CTT ATC CTG CCG GAC TCT GTC G
RP_Sleu_TermProm_Gibson	71.1	<u>TCAGCTAATTAAGCTGGTACC</u> CTA AAA GAA GTT CAG GCC GAA GCT GTC CAG CGG CTT G

Primers for SLIM:

Name	Sequence (5'-3')
FP_SleuC_RBS_B1	CGC CTG CAC CCT GAT ATC GTC ATC GCG GAA ACC G
RP_SleuC_RBS_B1	GGG TGC AGG CGC ATC TTG TCT CTT TGG GGT GGT GC
FPTail_SleuC_RBS_B1	TAA AGA GGA GAA ATT AAC CAT GCG CCT GCA CCC TGA TAT CGT CAT CGC GGA AAC CG
RPTail_SleuC_RBS_B1	CAT GGT TAA TTT CTC CTC TTT AGG GTG CAG GCG CAT CTT GTC TCT TTG GGG TGG TGC
FP_SmonC_RBS_B1	TTC GCC CTC GCC CCC CAC GTC AGC AGC GTG
RP_SmonC_RBS_B1	GGC TCC TTC CGC GGC GGG AAC GGT GGG GCG
FPTail_SmonC_RBS_B1	TAA AGA GGA GAA ATT AAC CAT GTT CGC CCT CGC CCC CCA CGT CAG CAG CGT G
RPTail_SmonC_RBS_B1	CAT GGT TAA TTT CTC CTC TTT AGG CTC CTT CCG CGG CGG GAA CGG TGG GGC G

FP_SleuB1_RBS_B2	AGT ACC CCC GCC GCC CTG GTC AAG CGC AGA AAA C
RP_SleuB1_RBS_B2	CGA GGT GCT CTC CAC TAT CTG GGC CGC GTG TAG TTG
FPTail_SleuB1_RBS_B2	TAA AGA GGA GAA ATT AAC CAT GAG TAC CCC CGC CGC CCT GGT CAA GCG CAG AAA AC
RPTail_SleuB1_RBS_B2	CAT GGT TAA TTT CTC CTC TTT ACG AGG TGC TCT CCA CTA TCT GGG CCG CGT GTA GTT G
FP_SmonB1_RBS_B2	AGC ATC CCG GTC GCC CTG GGC GTG CCC GCG
RP_SmonB1_RBS_B2	CGA GGT CAC CAG CTT CGC CGA CCT CAG CGA CTC CAG
FPTail_SmonB1_RBS_B2	TAA AGA GGA GAA ATT AAC CAT GAG CAT CCC GGT CGC CCT GGG CGT GCC CGC G
RPTail_SmonB1_RBS_B2	CAT GGT TAA TTT CTC CTC TTT ACG AGG TCA CCA GCT TCG CCG ACC TCA GCG ACT CCAG

Homologous production

Liquid culture

Bacterial strain was dissolved in 3 mL of GYM medium. 500 μ L of culture was inoculated in 100 mL GYM-medium and incubated for 7 days at 30 °C. Then, 5 mL was inoculated in 500 mL of GYM medium (10 g/L malt extract, 4 g/L glucose, 4 g/L yeast extract, 20 g/L agar at pH 7.0), SM medium (20 g/L soy flour, 20 g/L mannitol at pH 7.4) and BI medium (20 g/L glucose, 4 g/L caseinpepton, 0.5 g/L yeast extract, 2.5 g/L NaCl, 3 g/L CaCO₃ at pH 7.2) for 7 days at 30 °C. After that time, cultures were harvested by centrifugation (2 x 20 min, 7000 rpm). Cell pellets were extracted with MeOH, shaken vigorously and incubated overnight at 4 °C. The extract was centrifuged for 20 min at 7000 rpm and the supernatant was collected. Culture supernatants were applied to solid phase extractions using 10 mL of XAD-16. After 1 h shaking at 100 rpm, the supernatant was removed by filtration and the resin was washed with water and extracted with 50 mL of MeOH. Solvent from supernatant and pellet extract was removed under *vacuo*. Dry pellet and supernatant extracts were resuspended in 900 μ L of MeOH-H₂O (1:1, v/v) and analyzed by LC-MS.

Solid culture

20 μ L of spores cryostock were inoculated in 300 mL GYM medium and incubated overnight at 30 °C. 2 mL of culture was spread into GYM and SM agar plates and they were incubated for 20 days at 37°C. The agar was introduced in falcons and extracted with 50 mL of MeOH, shaken vigorously and kept overnight at 4 °C with shaking. Falcons were centrifuged at 3500 rpm for 30 min. The supernatant was collected and concentrated under *vacuo*. 900 μ L of MeOH-H₂O (1:1, v/v) were used to resuspend the sample which was analyzed with LC-MS.

Cryostock of spores

500 μ L of culture was inoculated in 100 mL GYM-medium and incubated for 7 days at 30 °C. Then, 2 mL were spread into GYM-agar plates and incubated at 30 °C for 10 days. 3 mL of 10% glycerol were added to each plate and the surface was scrubed slowly with sterile cotton buds. The procedure was repeated and the supernatant was filtered through glass-wool. The resultant suspension was centrifuged 20 min at 4500 rpm. The supernatant was discarded and the pellet resuspended with 1 mL of 20% glycerol. The spore stock was transferred to 2 mL screw cap tube, frozen and stored at -80 °C.

DNA isolation

Bacterial strain was dissolved in 3 mL of GYM medium. 500 μ L of culture was inoculated in 100 mL GYM-medium and incubated for 7 days at 30 °C. Culture was centrifuged at 3000 rpm for 20 min. The supernatant was discarded and the pellet was kept on ice. DNA was extracted using a commercially available FastDNA SPIN kit. DNA concentration was measured by reading the absorbance at 260 and 280 nm using a NanoDrop. DNA concentration for *S. monomycini*: 143 ng/ μ L and 42.1 ng/ μ L for *S. leeuwenhoekii*.

Gibson-assembly cloning

For cloning, genomic DNA was amplified by PCR using appropriate oligonucleotide primers (see above), Phusion polymerase, GC and HF-buffer and 0, 5, or 10% DMSO.

Annealing temperature was set to 65 °C and extension was performed for 4.5 minutes at 72 °C. DNA was purified by electrophoresis using an agarose gel. After purification, the resulting amplicon was cloned into pET41a vector using Gibson-Assembly Master Mix.

$$m(\text{vector}) \cdot \frac{I}{V} \cdot X_{si} = m(\text{insert}) \quad \text{Equation 3}$$

where m is the size of the vector or insert (in kb), I is the size of the insert (in kb), V is the size of the vector (in kb) and I/V is the relation in size (in bp) and X_{si} is the ratio $I/V = 1.5$.

Following Equation 3, 2.5 μL of insert were mixed with 2.5 μL of vector and 5 μL of Gibson Assembly Master Mix and incubated for 1 h at 50 °C. Dialysis was performed with 10 μL for 1 h. 20 μL of ligated DNA were transformed into *E. coli* TOP10 cells (40 μL) using electroporation. 360 μL of LB-medium was added to the cells and incubated for 1 h at 37 °C. 30 μL of cells were spread into LB agar plates containing 50 $\mu\text{g}/\text{mL}$ of kanamycin and incubated overnight at 37 °C. After incubation, some colonies were picked up and inoculated in 5 mL of LB medium containing 50 $\mu\text{g}/\text{mL}$ of kanamycin. After overnight incubation at 37 °C, plasmids were isolated using GenElute™ Plasmid Miniprep Kit and digested with the restriction enzymes HindIII in for *S. leeuwenhoekii* and with KpnI for *S. monomycini* during 2h at 37 °C. Each plasmid was analyzed by agarose gel and DNA sequencing.

Slim mutagenesis

The gene cluster containing the terminator-promoter was amplified by two PCR using appropriate oligonucleotide primers (see above), Phusion polymerase, GC-buffer and 5, 10% DMSO. Annealing temperature was set to 60 and 65 °C and extension was performed for 4.5 min at 72 °C. The resulting amplicons were digested with DpnI for 2 h at 37 °C and 20 min at 80 °C to remove the original template. Hybridization was carried out mixing 10 μL of each digested PCR, 10 μL of H-buffer (750 mM NaCl, 125 mM Tris, 100 mM Na₂EDTA at pH 8.0) and 20 μL of H₂O at 99 °C for 3 min. Then, DNA was dialyzed and transformed into *E. coli* TOP10 cells (40 μL) using electroporation. 360 μL of LB medium was added to the cells and incubated for 1 h at 37 °C. 30 μL of cells was spread into LB agar plates containing 50 $\mu\text{g}/\text{mL}$ of kanamycin and incubated overnight at 37 °C. After incubation, some colonies were

picked up and inoculated in 5 mL of LB medium containing 50 µg/mL of kanamycin. After overnight incubation at 37 °C, plasmids were isolated using GenElute™ Plasmid Miniprep Kit and digested with the restriction enzymes HindIII for *S. leeuwenhoekii* and KpnI for *S. monomycini* during 2h at 37 °C. Each plasmid was analyzed by agarose gel and DNA sequencing.

Heterologous expression

For the heterologous expression, plasmids bearing the correct sequences (MccJ25 pTuc262, CapA_RBS_BCD pET41a, sala1_RBS_BC pET41a, SmonA_T-P_C_RBS_B_RBS_B pET41a and SleuA_T-P_CB_RBS_B pET41a) were transformed into *E. coli* BL21 by electroporation and incubated 1 h at 37 °C. 30 µL of cells was spread into LB agar plates containing 50 µg/mL of kanamycin and incubated overnight at 37 °C. One colony was inoculated in LB medium supplemented with 50 µg/mL kanamycin and grown overnight at 37 °C. 5 mL of culture was inoculated in 500 mL of M9 media (17.1 g/L Na₂HPO₄ × 12 H₂O, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 1.0 g/L NH₄Cl, 1.0 mL/L MgSO₄ 2 M, 0.2 mL/L CaCl₂ 0.5 M at pH 7.0) and 10 mL/L glucose solution (40% w/v), 0.2 mL/L vitamin mix and antibiotic (amphenicol 17 µg/mL or kanamycin 50 µg/mL). After inoculation, cells were grown at 37 °C until an OD₆₀₀ of ~0.4 and then slowly cooled to 20 °C during 1 h upon reaching an OD₆₀₀ of ~0.6. The expression was induced by addition of IPTG to a final concentration of 0.05 mM, except for MccJ25 expression, in which the isopropyl β-D-1-thiogalactopyranoside (IPTG) addition was not required. Capistruin, chaxapeptin and the one of *S. monomycini* were incubated in two batches, one for 1 day at 37 °C and other for 3 days at 20 °C. For Mcc25 expression, cells were grown for 3 days at 37 °C, while for sphingopyxin I cells were grown for 3 days at 20 °C. Once OD₆₀₀ reach ~1.5, cells were harvested by centrifugation (2 x 20 min at 7000 rpm). The pellet was discarded for microcin J25 and capistruin and the supernatant was treated with 10 mL of XAD-16. After 1 h shaking at 100 rpm, the supernatant was removed by filtration and the resin was washed with water and extracted with 50 mL of MeOH.

In the case of sphingopyxin I, chaxapeptin and *S. monomycini*, the supernatant was discarded, and the pellet was resuspended in 30 mL MeOH, shaken vigorously and kept overnight at 4 °C. The extract was centrifuged for 20 min at 7000 rpm and the supernatant was collected. In both cases, the supernatant was removed under *vacuo*. Dry pellet and supernatant extracts were resuspended in MeOH-H₂O (1:1, v/v)

and purified by preparative HPLC. For all lasso peptides two round purifications were needed. For the first purification MeOH (0.045% FA) and H₂O (0.05% FA) were used as solvents whereas for the second, ACN (0.1% TFA) and H₂O (0.1% TFA) were used. For MccJ25 purification, a linear gradient from 40% to 50% in 5 min, followed by 50% to 95% in 30 min was used. For the second purification, a linear gradient from 25% to 40% in 30 min was used. For the rest of the peptides, a linear gradient from 20% to 35% in 5 min, followed by 30% to 95% in 30 min was used. For the second purification, a linear gradient from 10% to 60% in 30 min was used. All lasso peptides obtained were characterized by HPLC-MSD.

Microcin J25: analytical HPLC (linear gradient from 25:40 (0.036% TFA in ACN/0.045% TFA in H₂O) in 40 min; $t_R = 20.318$, 99.9% purity) and LC-MS (ESI⁺) m/z calculated for C₁₀₁H₁₃₉N₂₃O₂₇ [M+H]⁺ 2106.02, found 2107.5. Heterologous production yield: 3 days at 37 °C, 15.6 mg/L.

Chaxapeptin: analytical HPLC (linear gradient from 35:40 (0.036% TFA in ACN/0.045% TFA in H₂O) in 40 min; $t_R = 14.848$, 99.9% purity) and LC-MS (ESI⁺) m/z calculated for C₇₉H₁₀₇N₁₇O₂₀ [M+H]⁺ 1613.79, found 1615.17. Homologous production yield: 0.711 mg/L.

Capistruin: analytical HPLC (linear gradient from 20:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 3.365$, 99.9% purity) and LC-MS (ESI⁺) m/z calculated for C₉₃H₁₃₇N₂₇O₂₆ [M+H]⁺ 2048.02, found 2049.42. Heterologous production yield: 3 days at 20 °C, 0.27 mg/L and 1 day at 37 °C, 0.57 mg/L.

Sphingopyxin I: analytical HPLC (linear gradient from 25:40 (0.036% TFA in ACN/0.045% TFA in H₂O) in 40 min; $t_R = 12.445$, 99.9% purity) and LC-MS (ESI⁺) m/z calculated for C₉₈H₁₄₀N₂₄O₃₃ [M+H]⁺ 2181.00, found 2182.38. Heterologous production yield: 3 days at 20 °C, 2.1 mg/L.

M9 Vitamin Mix composition:

Component	Amount
biotin	0.2 g
choline chloride	1.0 g
disodium adenosine 5'-triphosphate	0.3 g
folic acid	1.0 g
myo-inositol	2.0 g
nicotinamide	1.0 g
panthothenic acid	1.0 g
pyridoxal hydrochloride	1.0 g
riboflavin	0.1 g
thiamine	1.0 g
H ₂ O	300 mL

10 M NaOH were added to the resulting suspension until complete dissolution. The solution was then sterile filtered and stored at 4 °C

Chemoenzymatic approach

Lasso peptide unthreading

5 mg of pure sphingopyxin I was dissolved in 1 mL of H₂O and 80 µg of isopeptidase in 80 µL of buffer (10 mM HEPES + 150 mM NaCl at pH 7.5). Different aliquots of peptide-enzyme were mixed in a ratio 5:1, respectively. The mixture was incubated at 30 °C and after 30 min the peptide was linearized. Linear sphingopyxin was purified in a semi preparative HPLC using 0.1% TFA in ACN/0.1% TFA in H₂O as solvents and a linear gradient from 10% to 60% in 30 min.

Reverse reaction

0.41 µL of peptide solution (420 µg/400 µL H₂O) were mixed with 118 µL of isopeptidase solution (1.37 µg/µL in buffer) in 1:10 (w/w) and incubated at 30 °C. The reaction was followed by LC-MS by taking aliquots every 30 minutes. The enzyme

amount was increased every 30 min, from 1:10, 1:30 to 1:50 and incubated overnight.

For the biological assays

Reagents

DMEM-Dulbecco's Modified Eagle Medium and cell Proliferation Kit II were obtained from Sigma-Aldrich and fetal bovine serum (FBS) was obtained from Gibco™. A549 cell line was obtained from the American Type Culture Collection (ATCC). Pepsin and thermolysin were purchased from Sigma Aldrich. Carboxypeptidase Y was purchased from Alfa Aesar. Migration and Matrigel¹¹ invasion chambers were obtained from Corning®. Collistin sulphate was purchased from Zhejiang Shenghua Biok Biology Co. LTD, China. Mueller-Hinton Broth (MHB) was obtained from Scharlau.

Enzymatic digestions and stability

Carboxypeptidase Y Digestion

Purified peptide (0.5 µg) was dissolved in 25 µL of MES buffer (50 mM MES, 1 mM CaCl₂ at pH 6.75). 10 µL of carboxypeptidase Y solution in MES buffer (1 mg/mL) was added to the peptide solution and incubated for 4 h at 25 °C. The reaction was quenched by the addition of 75 µL of ACN, filtered and analyzed with LC-MS.

Pepsin Digestion

The purified peptide (20 µg) was dissolved in 150 µL HCl 0.01 M (pH 2). 40 µL of pepsin solution in HCl 0.01 M (1 mg/mL) was added to the peptide solution and incubated at 37 °C for 2-20 h. The reaction was quenched by the addition of HCl 4 M until pH 2, filtered and analyzed with LC-MS.

Thermolysin Digestion

Condition 1

The purified peptide (5-10 µg) was dissolved in 50 µL of buffer (50 mM Tris, 0.5 mM CaCl₂ at pH 8). 15 µL of thermolysin solution in buffer (1 mg/mL) was added to the peptide solution and incubated for 30 min and 1 h at 70 °C. The reaction was quenched by the addition of NaOH until pH 8-8.5, filtered and analyzed via LC-MS.

Condition 2: 0.1 M NH₄CO₃, 1 mM CaCl₂ at pH 8 for 30 min and 1h at 60 °C.

Condition 3: 10 mM Tris-HCl at pH 8 for 2 h at 25 °C.

Thermal stability

The purified peptide (10 µg) was dissolved in 100 µL of H₂O-ACN (1:1, v/v) and incubated at 95 °C for 1, 2, 3 and 4 h. Samples were cooled to 4 °C and analyzed via LC-MS.

Cell culture

A549 cells were incubated in DMEM medium supplemented with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin and 100 mg/mL streptomycin in a humidified atmosphere of 5% CO₂ at 37 °C.

For sub-culturing, cells were detached from culture flasks by incubation with 1 mL 1 x trypsin in PBS 1 min (at 37 °C, 5% CO₂). Trypsinated cells were centrifuged for 5 min at 1000 rpm at room temperature and gently resuspended in 10 mL of fresh medium preheated at 37 °C. Cellular density was determined in a Neubauer counting plate and the appropriate volume of cells was placed into a fresh culture flask or a well-plate.

Cytotoxicity assay

A549 cells were seeded in a 96-well Corning microplate (4 x 10³ cells/well). After 24 h of incubation at 37 °C, cells were treated with the peptides dissolved in DMEM containing 0.5% DMSO (concentrations of 25, 50 and 100 µM, well volume 100 µL) for 24 h. Afterwards, 50 µL of activated-XTT solution (0.1 mL of the XTT reagent in 5.0 mL of the electron coupling reagent) was added to each well and incubated for 4 h at 37 °C. The absorbance at 475 nm was measured in ELx800 BioTEK UV-Vis plate

reader. Absorbance values were compared with the positive control (cells treated with DMSO) and with the negative control (untreated cells). Each experiment was performed in triplicate.

Transwell migration and invasion assay

A549 cells were treated with 100 μM of peptides for 24 h in DMEM medium supplemented with 0.2% FBS. Afterwards, 5×10^4 cells (500 μL) were added to each 24-well transwell, previously hydrated for at least 2 h at 37 °C and 5% CO_2 . 700 μL of DMEM supplemented with 10% FBS was added into the bottom of the 24-well plates. The migration assay was carried out for 15 h and the invasion for 24 h. After this time, the transwells were washed with PBS and the cells were fixed with 4% *p*-formaldehyde for 15 min at room temperature. Then, the transwells were rinsed in H_2O and the cells into the top of the chamber were rubbed with a cotton swab. Finally, the cells were stained with 0.2% crystal violet for 15 min, followed by the gently rinsed with H_2O . The cells from the bottom of the membrane were counted using an inverted Nikon TE200 microscope equipped with an Olympus DP72 camera. The number of cells was compared with the positive control (untreated cells with DMEM containing 10% FBS into the bottom of the well plate) and with the negative control (untreated cells with DMEM containing 0.2% FBS into the bottom of the well plate).

Minimum inhibitory concentration (MIC)

MIC values were determined by broth microdilution method and interpreted according to Clinical & Laboratory Standards Institute (CLSI) and European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines.¹²

Briefly, the different strains were grown overnight at 37 °C with shaking at 200 rpm in MHB. After doing a refresh of 2 h with the same conditions, the bacterial cultures were adjusted to $\text{OD}_{625\text{nm}}$ of 0.08-0.1 and diluted 1:100 in fresh MHB medium. 5 μL of each diluted suspension was added to a 96-well plates previously filled with MHB and serially diluted peptides. The plates were incubated at 37 °C for 24 h, after which the MIC was determined macroscopically, based on the visually turbidity of the wells. All experiments were performed in triplicate.

Synergy study

A checkerboard test was used to determine the fractional inhibitory concentrations (FICs) of colistin in combination with the peptide. Each well in a 96-well plate was inoculated with 100 μ L of a bacterial inoculum of 1×10^5 CFU/mL, and the plates were incubated at 37 °C for 24 h. The FIC was calculated after identifying the first well in each row without growth (MIC), according to Equation 4.

$$FIC_A = \frac{\text{MIC drug A in combination}}{\text{MIC drug A}} \quad FIC_B = \frac{\text{MIC drug B in combination}}{\text{MIC drug B}} \quad \text{Equation 4}$$

The FIC index (FIC_i) values were calculated by adding the FIC A (colistin) to the FIC B (peptide). FIC_i values were interpreted as follows: FIC_i < 0.5, synergistic; FIC_i \geq 0.5 and < 4, no interaction; FIC_i > 4, antagonistic.¹³ This assay was performed in triplicate.

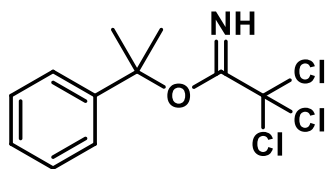
References

- (1) Madder, A.; Farcy, N.; Hosten, N. G. C.; De Muynck, H.; De Clercq, P. J.; Barry, J.; Davis, A. P. *European J. Org. Chem.* **1999**, 1999 (11), 2787.
- (2) Ellman, G. L. *Arch. Biochem. Biophys.* **1959**, 82 (1), 70.
- (3) Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. *J. Chem. Inf. Model.* **2014**, 54 (10), 2680.
- (4) Wessel, H.-P.; Iversen, T.; Bundle, D. R. *J. Chem. Soc. Perkin Trans. 1* **1985**, 0 (24772), 2247.
- (5) Belshaw, P. J.; Mzengeza, S.; Lajoie, G. a. *Synth. Commun.* **1990**, 20 (20), 3157.
- (6) Yue, C.; Thierry, J.; Potier, P. *Tetrahedron Lett.* **1993**, 34 (2), 323.
- (7) Galaud, F.; Lubell, W. D. *Biopolym. - Pept. Sci. Sect.* **2005**, 80 (5), 665.
- (8) Isidro-Llobet, A.; Guasch-Camell, J.; Álvarez, M.; Albericio, F. *European J. Org. Chem.* **2005**, 2005 (14), 3031.
- (9) Cruz, L. J.; Beteta, N. G.; Ewenson, A.; Albericio, F. *Org. Process Res. Dev.* **2004**, 8 (6), 920.
- (10) Mérette, S. A. M.; Burd, A. P.; Deadman, J. J. *Tetrahedron Lett.* **1999**, 40 (4), 753.
- (11) Kleinman, H. K.; Martin, G. R. *Semin. Cancer Biol.* **2005**, 15 (5), 378.
- (12) Committee, T. E.; Testing, A. S.; Changes, N.; Pseudomonas, E. European Committee on Antimicrobial Susceptibility Testing Breakpoint tables for interpretation of MICs and zone diameters European Committee on Antimicrobial Susceptibility Testing Breakpoint tables for interpretation of MICs and zone diameters. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.pdf (accessed Jan 12, 2018).
- (13) Odds, F. C. *J. Antimicrob. Chemother.* **2003**, 52 (1), 1.

Product obtention and characterization

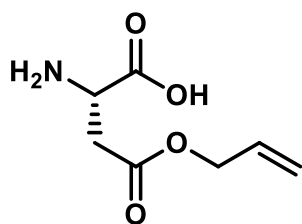
Amino acids synthesis

2-phenylisopropyl 2,2,2-trichloroacetimidate (1)

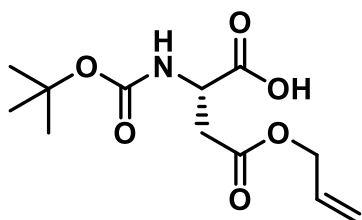


Sodium hydride (10.54 g, 263.6 mmol, 2 eq) was suspended in dry THF (350 mL) in a round bottomed flask, previously purged with N₂, and under inert atmosphere, following the procedure described by Wessel *et al.*⁴ A solution of 2-phenylisopropanol (17.95 g, 131.8 mmol, 1 eq) was added dropwise to the stirring suspension. After 20 min, the solution was cooled to 0 °C and trichloroacetonitrile (26.4 mL, 263.6 mmol, 2 eq) was added dropwise for 20 min and the reaction mixture was allowed to reach room temperature over 3 h. The resulting mixture was cooled at 0 °C, and MeOH was added and concentrated under *vacuo*. Hexane was added to the brown thick oil obtained, followed by vigorous shaking, filtration and concentration under *vacuo*. 35.2 g as a brown-yellow oil was obtained of compound 1 (91%), which was used without further purification. ¹H NMR (400 MHz, CDCl₃) δ 1.90 (s, 6H), 7.23-7.51 (m, 5H), 8.21 (s, 1H); ¹³C (400 MHz, CDCl₃) δ 28.25 (q, CH₃), 72.68 (s, C), 85.18 (s, C), 124.47 (d, CH), 127.31 (d, CH), 128.41 (d, CH), 145.21 (s, C), 159.56 (s, C). LC-MS (ESI⁺) *m/z* calculated for C₁₁H₁₂Cl₃NO [M+H]⁺ 279.00, found 280.10.

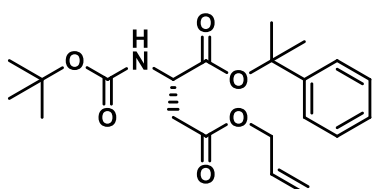
H-L-Asp(OAll)-OH (2)



To a stirred suspension of L-aspartic acid (7.63 g, 57.3 mmol, 1 eq) in dry allyl alcohol (224.1 mL) under N₂ atmosphere, TMSCl (20 mL, 143.2 mmol, 2.5 eq) was added dropwise according to the procedure of Belshaw *et al.*⁵ The resulting mixture was stirred at room temperature for 18 h. The solution was cooled at 0 °C and cold Et₂O was added to obtain a precipitate, which was collected by filtration to give 10.1 g (84%) of a white solid. It was used in the next reaction without further purification. ¹H NMR (400 MHz, DMSO- *d*₆) δ 3.01 (dd, *J* = 5.7, 3.8 Hz, 2H), 4.19 (t, *J* = 5.6 Hz, 1H), 4.60 (dt, *J* = 5.0, 1.5 Hz, 2H), 5.23 (dq, *J* = 10.5, 1.4 Hz, 1H), 5.33 (dq, *J* = 17.2, 1.7 Hz, 1H), 5.91 (ddt, *J* = 17.3, 10.7, 5.5 Hz, 1H); ¹³C (400 MHz, DMSO- *d*₆) δ 34.09 (t, CH₂), 48.41 (d, CH), 65.20 (t, CH₂), 118.09 (t, CH₂), 132.20 (d, CH), 169.00 (s, CO), 169.60 (s, CO). LC-MS (ESI⁺) *m/z* calculated for C₇H₁₁NO₄ [M+H]⁺ 173.07, found 174.3.

Boc-L-Asp(OAll)-OH (3)

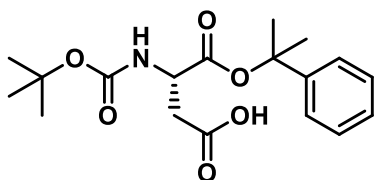
To a solution of **2** (9.24 g, 44.18 mmol, 1 eq) in 200 mL of H₂O/1,4-dioxane (1:1, v/v), Boc₂O (11.6 g, 53.02 mmol, 1.2 eq) in 4 mL of solvent was added dropwise. Then, Et₃N (11.7 mL, 83.9 mmol, 1.9 eq) was added dropwise to the stirred solution. After 16 h, the solution was extracted with hexane (x 3) and the aqueous phase was cooled to 0 °C and acidified to pH 2-3 by slowly addition of HCl 1 M. The acidic aqueous layer was extracted with EtOAc (x 3) and the combined organic layers were washed twice with brine, dried over MgSO₄, filtered and concentrated under *vacuo* to give 11.2 g of crude **3** (93%) (as a pale-yellow oil, which was used in the next reaction without further purification. ¹H NMR (400 MHz, CDCl₃) δ 1.44 (s, 9H), 2.87 (dd, *J* = 17.2, 4.8 Hz, 1H), 3.05 (dd, *J* = 17.3, 4.7 Hz, 1H), 4.60 (dq, *J* = 5.8 Hz, 3H), 5.24 (dq, *J* = 10.4, 1.3 Hz, 1H), 5.31 (dq, *J* = 17.2, 1.5, 1H), 5.56 (d, *J* = 8.6 Hz, 1H), 5.84-5.94 (m, 1H); ¹³C (400 MHz, CDCl₃) δ 28.42 (q, CH₃), 36.66 (t, CH₂), 49.93 (d, CH), 65.95 (t, CH₂), 80.64 (s, C), 118.89 (t, CH₂), 131.72 (d, CH), 155.74 (s, CO), 170.98 (s, CO), 175.61 (s, CO). LC-MS (ESI⁺) *m/z* calculated for C₁₂H₁₉NO₆ [M+H]⁺ 273.12, found 274.31.

Boc-L-Asp(OAll)-O(2-PhⁱPr) (4)

To a solution of **3** (12.84 g, 47.01 mmol, 1 eq) in dry DCM (160 mL), compound **1** (27.55 g, 94.03 mmol, 2 eq) dissolved in 90 mL of dry DCM was added slowly.^{6,7} The reaction mixture was stirred for 3 h at room temperature. Then, the mixture was extracted with 0.5% aqueous citric acid (x 3). The combined organic layers were washed with saturated NaHCO₃ (x 3), dried over MgSO₄, filtered and concentrated under *vacuo*. The residue obtained was purified by reversed phase C18 flash column chromatography (H₂O/ACN) to give 7.4 g (40%) of a white powder. ¹H NMR (400 MHz, CDCl₃) δ 1.44 (s, 9H), 1.77 (s, 3H), 1.79 (s, 3H), 2.83 (dd, *J* = 16.8, 5.0 Hz, 1H), 3.03 (dd, *J* = 16.8, 4.6 Hz, 1H), 4.51-4.61 (m, 3H), 5.24 (dq, *J* = 10.4, 1.3 Hz, 1H), 5.31 (dq, *J* = 17.2, 1.5, 1H), 5.46 (d, *J* = 8.6 Hz, 1H), 5.89 (ddt, *J* = 17.2, 10.5, 5.8 Hz, 1H), 7.22-7.37 (m, 5H); ¹³C (400 MHz, CDCl₃) δ 28.46 (q, CH₃), 28.53 (q, CH₃), 36.98 (t, CH₂), 50.72 (d, CH), 65.73 (t, CH₂), 80.09 (s, C), 83.70 (s, C), 118.78 (t, CH₂), 124.44 (d, CH), 127.35 (d, CH), 128.44 (d, CH),

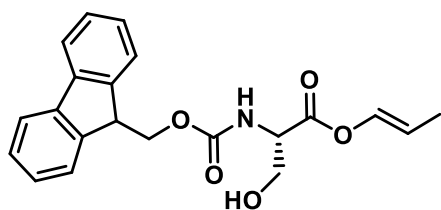
131.89 (d, CH), 145.25 (s, C), 155.69 (s, CO), 169.61 (s, CO), 170.81 (s, CO). LC-MS (ESI⁺) m/z calculated for C₂₁H₂₉NO₆ [M+H]⁺ 391.20, found 392.17.

Boc-L-Asp-O(2-PhⁱPr) (5)



Compound 4 (3.249 g, 8.3 mmol, 1 eq) in dry DCM (140 mL) was dissolved in a round bottomed flask, previously purged with N₂, and under inert atmosphere, pyrrolidine (1.4 mL, 16.6 mmol, 2 eq) was added dropwise. Then, Pd(PPh₃)₄ (960 mg, 0.83 mmol, 0.1 eq) and PPh₃ (43.7 mg, 1.7 mmol, 0.2 eq) were added. The resulting mixture was stirred for 30 min. After further adding DCM, the mixture was extracted with 0.5% aqueous citric acid (x 3). The combined organic layers were washed twice with brine, dried over MgSO₄, filtered and concentrated under *vacuo*. The residue was purified by flash column chromatography (DCM-EtOAc = 100:0 to 0:100) to give 1.74 g (60%) of a white-yellow solid. ¹H NMR (400 MHz, CDCl₃) δ 1.45 (s, 9H), 1.77 (s, 3H), 1.79 (s, 3H), 2.87 (dd, *J* = 17.2, 4.9 Hz, 1H), 3.08 (dd, *J* = 18.0, 3.4 Hz, 1H), 4.55 (t, *J* = 5.0 Hz, 1H), 5.45 (d, *J* = 8.4 Hz, 1H), 7.23-7.36 (m, 5H); ¹³C (400 MHz, CDCl₃) δ 28.46 (q, CH₃), 28.53 (q, CH₃), 36.76 (t, CH₂), 50.51 (d, CH), 80.33 (s, C), 83.87 (s, C), 124.46 (s, CH), 127.43 (s, CH), 128.45 (s, CH), 145.07 (s, C), 155.65 (s, CO), 169.9 (s, CO), 176.15 (s, CO). LC-MS (ESI⁺) m/z calculated for C₁₈H₂₅NO₆ [M+H]⁺ 351.17, found 352.09.

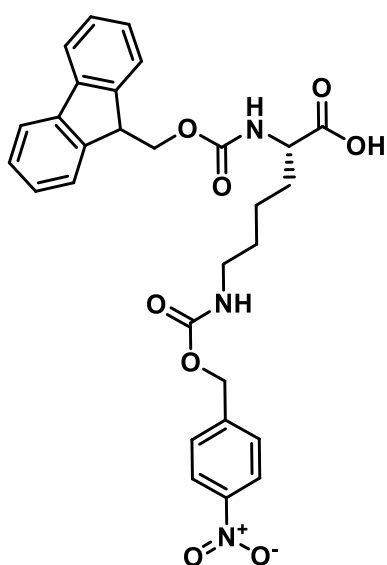
Fmoc-L-Ser-OAll (6)



Fmoc-L-Ser-OH (1.006 g, 2.91 mmol, 1 eq) was dissolved in dry DMF (5 mL) and K₂CO₃ (402.7 mg, 2.91 mmol, 1 eq) was added. The reaction mixture was stirred for 15 min and then, allyl bromide (1.261 mL, 14.55 mmol, 5 eq) was added dropwise. The resulting white mixture was stirred for 3 h at room temperature. The mixture was dissolved in 75 mL of 0.5% aqueous citric acid and extracted with EtOAc (x 3). The combined organic layers were washed with saturated NaHCO₃ (x 3), dried over MgSO₄, filtered and concentrated under *vacuo*. 980 mg (92%) of a pale-yellow product were obtained. It was used in the next reaction without further purification. ¹H NMR (400 MHz, CDCl₃) δ 3.95 (d, *J* = 11.1 Hz, 1H), 4.04 (d, *J* = 10.9 Hz, 1H), 4.23 (t, *J* = 6.9 Hz, 1H), 4.38 - 4.52 (m, 3H), 4.70 (d, *J* = 5.7 Hz, 2H), 5.27 (dt, *J* = 10.4, 1.2 Hz, 1H), 5.35 (dd, *J* = 17.1, 1.6 Hz, 1H), 5.70 (d, *J* = 7.6 Hz, 1H), 5.92 (ddt, *J* =

16.4, 10.9, 5.7 Hz, 1H), 7.32 (tt, $J = 7.5, 1.2$ Hz, 2H), 7.41 (dddt, $J = 8.2, 7.5, 1.3, 0.7$ Hz, 2H), 7.61 (d, $J = 7.5$ Hz, 2H), 7.75 - 7.79 (m, 2H). ^{13}C NMR (400 MHz, CDCl_3) δ 47.29 (t, CH_2), 56.24 (d, CH), 6.22 (d, CH), 66.52 (t, CH_2), 67.35 (t, CH_2), 119.18 (d, CH), 120.14 (d, CH), 125.20 (d, CH), 127.23 (d, CH), 127.89 (d, CH), 131.44 (d, CH), 141.48 (s, C), 143.68 (s, C). HPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H_2O) in 8 min; $t_{\text{R}} = 6.637$, 92% purity) and LC-MS (ESI⁺) m/z calculated for $\text{C}_{21}\text{H}_{21}\text{NO}_5$ $[\text{M}+\text{H}]^+$ 367.14, found 368.09.

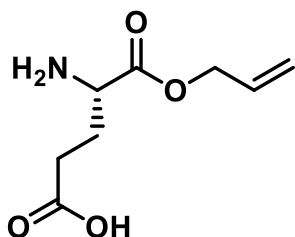
Fmoc-L-Lys(pNZ)-OH (7)



p-Nitrobenzyl chloroformate (2.41 g, 11.2 mmol, 1 eq) was dissolved in 1,4-dioxane (20 mL) and a solution of sodium azide (874.4 mg, 13.5 mmol, 1.2 eq) in H_2O (20 mL) was added following the protocol described by Isidro-Llobet et al.⁸ The resulting emulsion was stirred for 1h. Fmoc-L-Lys-OH (4.13 g, 11.2 mmol, 1 eq) was dissolved in 1,4-dioxane/2% Na_2CO_3 (1:1, v/v) (80 mL), and then added slowly. The resulting white suspension was stirred for 16 h, keeping the pH between 9 and 10 by adding 10% aqueous Na_2CO_3 . The solution was extracted with Et_2O (x 3) and the aqueous phase was acidified to pH 2-3

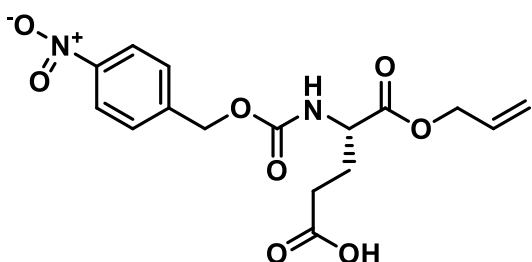
by the slow addition of HCl 1 M. The acidic suspension was extracted with EtOAc (x3). The combined organic layers were dried over MgSO_4 , filtered and concentrated under *vacuo* to yield 5.01 g of oil (68%) which solidified as a white solid, which was used without further purification. ^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 1.37 (dt, $J = 30.1, 7.6$ Hz, 4H), 1.54-1.72 (m, 2H), 3.00 (q, $J = 6.6$ Hz, 2H), 3.90 (td, $J = 8.7, 4.5$ Hz, 1H), 4.19-4.24 (m, 1H), 4.28 (d, $J = 8.0$, 2H), 5.15 (s, 2H), 7.32 (dd, $J = 7.9, 6.8$ Hz, 2H), 7.41 (t, $J = 7.5$ Hz, 2H), 7.59 (dd, $J = 8.3, 6.2$ Hz, 2H), 7.72 (d, $J = 7.5$ Hz, 2H), 7.89 (d, $J = 7.6$ Hz, 2H), 8.21-8.25 (m, 2H); ^{13}C (400 MHz, $\text{DMSO}-d_6$) δ 23.35 (t, CH_2), 29.35 (t, CH_2), 30.89 (t, CH_2), 47.04 (t, CH_2), 54.25 (d, CH), 64.37 (d, CH), 66.02 (t, CH_2), 66.79 (t, CH_2), 120.53 (d, CH), 123.98 (d, CH), 125.70 (d, CH), 127.48 (d, CH), 128.06 (d, CH), 128.50 (d, CH), 141.14 (s, C), 144.25 (s, C), 145.76 (s, C), 147.33 (s, C), 156.19 (s, CO), 156.58 (s, CO), 174.40 (s, CO). LC-MS (ESI⁺) m/z calculated for $\text{C}_{29}\text{H}_{29}\text{N}_3\text{O}_8$ $[\text{M}+\text{H}]^+$ 547.20, found 548.6.

H-L-Glu-OAll (8)



To a solution of Fmoc-L-Glu-OAll (3.23 g, 7.88 mmol, 1 eq) in DCM (25 mL), DEA (4.1 mL, 39.4 mmol, 5 eq) was added and the resulting solution was stirred for 3 h. Then, the solvent was removed *in vacuo*, followed by consecutive co-evaporations with DCM. The residue obtained was purified by reversed phase C18 flash column chromatography (H₂O/ACN) to give 1.14 g (98%) of a pale-yellow solid. ¹H NMR (400 MHz, CDCl₃) δ 2.07 (dt, *J* = 15.6, 7.5 Hz, 1H), 2.19 (dd, *J* = 9.7, 5.8 Hz, 1H), 2.55 (t, *J* = 6.2 Hz, 2H), 3.95 (dd, *J* = 8.3, 3.6 Hz, 1H), 4.67 (ddd, *J* = 5.8, 2.8, 1.4 Hz, 2H), 5.29 (dd, *J* = 10.5, 1.3 Hz, 1H), 5.34 (dd, *J* = 17.2, 1.4 Hz, 1H), 5.95 - 5.84 (m, 1H). LC-MS (ESI⁺) *m/z* calculated for C₈H₁₃NO₄ [M+H]⁺ 187.08, found 188.19.

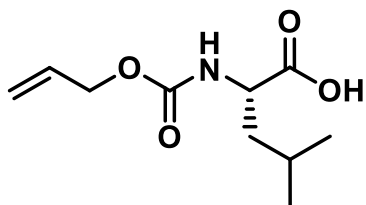
pNZ-L-Glu-OAll (9)



p-Nitrobenzyl chloroformate (1.85 g, 8.60 mmol, 1.1 eq) was dissolved in 1,4-dioxane (10 mL) and a solution of sodium azide (610.1 mg, 9.4 mmol, 1.2 eq) in H₂O (10 mL) was added following the protocol described by Isidro-Llobet et al.⁸ The resulting emulsion was stirred for 1 h. Compound **8** (1.46 g, 7.82 mmol, 1 eq) was dissolved in 1,4-dioxane/2% Na₂CO₃ (1:1, v/v) (80 mL), and then added slowly. The resulting white suspension was stirred for 5 h, keeping the pH between 7 and 8 by adding 10% aqueous Na₂CO₃. The solution was extracted with Et₂O (x 3) and the aqueous phase was acidified to pH 2-3 by the slow addition of HCl 1 M. The acidic suspension was extracted with EtOAc (x 3). The combined organic layers were dried over MgSO₄, filtered and concentrated under *vacuo* to yield 2.27 g of oil (79%) which solidified as a yellow solid, which was used without further purification. ¹H NMR (400 MHz, DMSO-*d*₆) δ 1.86 - 1.75 (m, 1H), 1.93-2.05 (m, 1H), 2.36 - 2.30 (m, 2H), 4.13 (ddd, *J* = 9.6, 7.8, 5.1 Hz, 1H), 4.59 (dq, *J* = 5.4, 1.4 Hz, 2H), 5.22 - 5.16 (m, 3H), 5.30 (dt, *J* = 17.3, 1.7 Hz, 1H), 5.89 (ddt, *J* = 17.2, 10.6, 5.3 Hz, 1H), 7.63 - 7.58 (m, 2H), 7.94 (d, *J* = 7.8 Hz, 1H), 8.26 - 8.21 (m, 2H). ¹³C NMR (400 MHz, DMSO- *d*₆) δ 25.95 (t, CH₂), 29.89 (t, CH₂), 53.22 (d, CH), 64.34 (t, CH₂), 64.90 (t, CH₂), 117.73 (t, CH₂), 123.53 (d, CH), 128.12 (d, CH), 132.36 (d, CH), 144.89 (s, C), 146.96 (s, C),

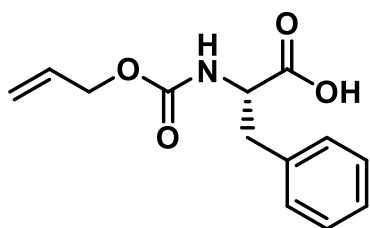
155.85 (s, CO), 171.63 (s, CO), 173.62 (s, CO). LC-MS (ESI⁺) *m/z* calculated for C₁₆H₁₈N₂O₈ [M+H]⁺ 366.11, found 367.3.

Alloc-L-Leu-OH (10)



Allyl chloroformate (747.7 μ L, 7.03 mmol, 1 eq) was dissolved in 1,4-dioxane (5 mL) and a solution of sodium azide (625 mg, 9.6 mmol, 1.2 eq) in H₂O (4 mL) was added following the azide methodology.⁹ The resulting emulsion was stirred for 1 h. H-L-Leu-OH (1.03 g, 7.82 mmol, 1 eq) was dissolved in 1,4-dioxane/1% Na₂CO₃ (1:1, v/v) (60 mL), and then added slowly. The resulting white suspension was stirred for 16 h, keeping the pH between 9 and 10 by adding 10% aqueous Na₂CO₃. The solution was extracted with Et₂O (x3) and the aqueous phase was acidified to pH 2-3 by the slow addition of HCl 1 M. The acidic suspension was extracted with EtOAc (x 3). The combined organic layers were dried over MgSO₄, filtered and concentrated under *vacuo* to yield 1.55 g of colorless oil (92%), which was used without further purification. ¹H NMR (400 MHz, CDCl₃) δ 0.95 (s, 3H), 0.97 (s, 3H), 1.57 (ddd, *J* = 13.5, 9.5, 5.3 Hz, 1H), 1.65-1.80 (m, 2H), 4.39 (td, *J* = 9.2, 6.4 Hz, 1H), 4.58 (d, *J* = 5.7, 2H), 5.10 (s, *J* = 8.6 Hz, 1H), 5.22 (dt, *J* = 10.5, 1.3 Hz, 1H), 5.31 (m, 1H), 5.91 (ddt, *J* = 16.4, 10.8, 5.6 Hz, 1H). LC-MS (ESI⁺) *m/z* calculated for C₁₀H₁₇NO₄ [M+H]⁺ 215.12, found 216.26.

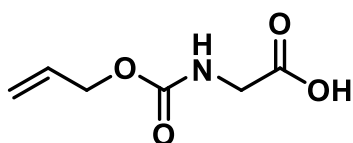
Alloc-L-Phe-OH (11)



Allyl chloroformate (721 μ L, 6.78 mmol, 1 eq) was dissolved in 1,4-dioxane (5 mL) and a solution of sodium azide (602.6 mg, 9.3 mmol, 1.2 eq) in H₂O (4 mL) was added following the azide methodology.⁹ The resulting emulsion was stirred for 1 h. H-L-Phe-OH (1.24 g, 7.53 mmol, 1 eq) was dissolved in 1,4-dioxane/1% Na₂CO₃ (1:1, v/v) (60 mL), and then added slowly. The resulting white suspension was stirred for 16 h, keeping the pH between 9 and 10 by adding 10% aqueous Na₂CO₃. The solution was extracted with Et₂O (x 3) and the aqueous phase was acidified to pH 2-3 by the slow addition of HCl 1 M. The acidic suspension was extracted with EtOAc (x 3). The combined organic layer was dried over MgSO₄, filtered and concentrated under *vacuo* to yield 1.1 g of a pale yellow oil (59%), which was used without further purification. ¹H NMR (400 MHz,

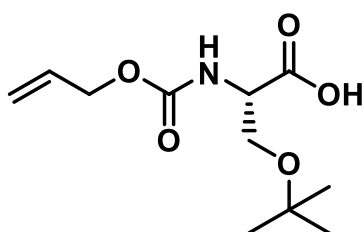
CDCl₃) δ 3.12 (dd, *J* = 14.1, 6.3 Hz, 1H), 3.21 (dd, *J* = 14.0, 5.5 Hz, 1H), 4.56 (d, *J* = 5.7, 2H), 4.68 (q, *J* = 6.6 Hz, 1H), 5.14 (d, *J* = 8.0 Hz, 1H), 5.21 (dd, *J* = 10.5, 1.4 Hz, 1H), 5.28 (d, *J* = 16.9 Hz, 1H), 5.89 (ddt, *J* = 16.3, 10.8, 5.6 Hz, 1H), 7.16-7.20 (m, 2H), 7.24-7.34 (m, 3H). LC-MS (ESI⁺) *m/z* calculated for C₁₃H₁₅NO₄ [M+H]⁺ 249.10, found 250.40.

Alloc-Gly-OH (12)



Allyl chloroformate (1.28 mL, 12.04 mmol, 1 eq) was dissolved in 1,4-dioxane (5 mL) and a solution of sodium azide (1.07 g, 16.5 mmol, 1.2 eq) in H₂O (4 mL) was added following the azide methodology.⁹ The resulting emulsion was stirred for 1 h. H-Gly-OH (1.00 g, 13.4 mmol, 1 eq) was dissolved in 1,4-dioxane/1% Na₂CO₃ (1:1, v/v) (60 mL), and then added slowly. The resulting white suspension was stirred for 16 h, keeping the pH between 9 and 10 by adding 10% aqueous Na₂CO₃. The solution was extracted with Et₂O (x 3) and the aqueous phase was acidified to pH 2-3 by the slow addition of HCl 1 M. The acidic suspension was extracted with EtOAc (x 3). The combined organic layers were dried over MgSO₄, filtered and concentrated under *vacuo* to yield 1.32 g of a pale-yellow oil (63%), which was used without further purification. ¹H NMR (400 MHz, CDCl₃) δ 4.02 (d, *J* = 5.8 Hz, 2H), 4.60 (d, *J* = 5.6 Hz, 2H), 5.23 (d, *J* = 10.4 Hz, 1H), 5.28-5.35 (m, 1H), 5.91 (ddt, *J* = 17.2, 10.4, 5.6 Hz, 1H). LC-MS (ESI⁺) *m/z* calculated for C₆H₉NO₄ [M+H]⁺ 159.05, found 160.10.

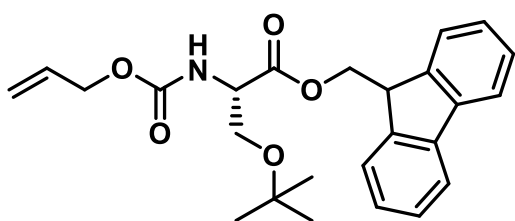
Alloc-L-Ser(*t*Bu)-OH (13)



Allyl chloroformate (5.3 mL, 49.7 mmol, 1 eq) was dissolved in 1,4-dioxane (5 mL) and a solution of sodium azide (3.88 g, 59.6 mmol, 1.2 eq) in H₂O (4 mL) was added following the azide methodology.⁹ The resulting emulsion was stirred for 1 h. H-L-Ser(*t*Bu)-OH (8.00 g, 49.7 mmol, 1 eq) was dissolved in 1,4-dioxane/1% Na₂CO₃ (1:1, v/v) (60 mL), and then added slowly. The resulting white suspension was stirred for 16 h, keeping the pH between 9 and 10 by adding 10% aqueous Na₂CO₃. The solution was extracted with Et₂O (x 3) and the aqueous phase was acidified to pH 2-3 by the slow addition of HCl 1 M. The acidic suspension was extracted with EtOAc (x 3). The combined organic

layers were dried over MgSO_4 , filtered and concentrated under *vacuo* to yield 11.1 g of colorless oil (91%), which was used without further purification. ^1H NMR (400 MHz, CDCl_3) δ 1.19 (s, 9H), 3.57 (dd, $J = 8.9, 5.0, 1\text{H}$), 3.90 (dd, $J = 9.0, 3.4, 1\text{H}$), 4.45 (dd, $J = 11.4, 4.0\text{ Hz}, 1\text{H}$), 4.60 (dt, $J = 5.7, 1.4\text{ Hz}, 2\text{H}$), 5.23 (dq, $J = 10.4, 1.3\text{ Hz}, 1\text{H}$), 5.28-5.36 (m, 1H), 5.59 (d, $J = 8.1\text{ Hz}, 1\text{H}$), 5.87-5.98 (m, 1H). LC-MS (ESI⁺) m/z calculated for $\text{C}_{11}\text{H}_{19}\text{NO}_5$ $[\text{M}+\text{H}]^+$ 245.13, found 246.30.

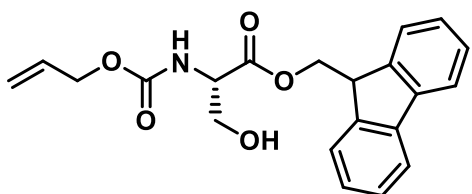
Alloc-L-Ser(*t*Bu)-OFm (14)



Compound **14** was prepared as described by Mérette *et al.*¹⁰ A solution of **13** (11.1 g, 45.12 mmol, 1 eq) and 9-fluorenylmethyl chloroformate (11.7 g, 45.12 mmol, 1 eq) in

dry DCM (130 mL) was cooled to 0 °C, followed by addition of DIEA (9.44mL, 54.14 mmol, 1.2 eq). After 5 min, DMAP (661.5 mg, 5.4 mmol, 0.12 eq) was added and the reaction mixture was allowed to warm to room temperature over 2 h. The resulting mixture was extracted with 2% aqueous citric acid (x 3) and 0.5 M aqueous NaHCO_3 (x 3). The combined organic layers were dried over MgSO_4 , filtered and concentrated under *vacuo* to give 17.6 g of crude (92%). The residue was purified by flash column chromatography with DCM to give 11.4 g of pure product (69%) as a colorless oil. ^1H NMR (400 MHz, CDCl_3) δ 1.12 (s, 9H), 3.60 (dd, $J = 9.0, 3.3, 1\text{H}$), 3.85 (dd, $J = 9.0, 2.9, 1\text{H}$), 4.24 (t, $J = 7.1\text{ Hz}, 1\text{H}$), 4.42 (dd, $J = 10.7, 7.1\text{ Hz}, 2\text{H}$), 4.45-4.53 (m, 2H), 4.61 (dt, $J = 5.7, 1.4, 2\text{H}$), 5.20-5.25 (m, 1H), 5.34 (dd, $J = 17.2, 1.7, 1\text{H}$), 5.61 (d, $J = 8.9\text{ Hz}, 1\text{H}$), 5.94 (ddt, $J = 16.3, 10.8, 5.7\text{ Hz}, 1\text{H}$), 7.31 (tt, $J = 7.5, 1.1\text{ Hz}, 2\text{H}$), 7.41 (tq, $J = 7.6, 1.1\text{ Hz}, 2\text{H}$), 7.61 (ddq, $J = 7.5, 4.6, 0.9\text{ Hz}, 2\text{H}$), 7.77 (d, $J = 7.5\text{ Hz}, 2\text{H}$). LC-MS (ESI⁺) m/z calculated for $\text{C}_{25}\text{H}_{29}\text{NO}_5$ $[\text{M}+\text{H}]^+$ 423.20, found 424.60.

Alloc-L-Ser-OFm (15)



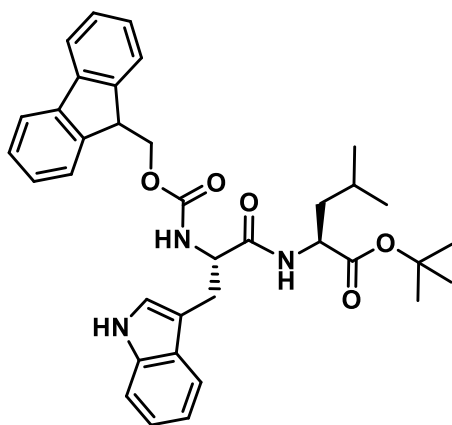
Compound **14** (11.4 g, 27 mmol, 1 eq) was dissolved in TFA-DCM (1:1, v/v) (200 mL) at room temperature. After 1 h, the resulting mixture was concentrated under *vacuo*. The

crude was dissolved in H_2O -ACN (1:1, v/v) and lyophilized to give 10 g (99%) of a white solid, which was used without further purification. ^1H NMR (400 MHz, CDCl_3) δ , 3.75 (d, $J = 3.5, 2\text{H}$), 4.24 (t, $J = 5.9\text{ Hz}, 1\text{H}$), 4.36-4.41 (m, 1H), 4.53 (dd, $J = 10.9,$

5.9, 1H), 4.58 (dt, $J = 5.6, 1.4$, 2H), 4.77 (dd, $J = 10.9, 6.0$ 1H), 5.20-5.25 (m, 1H), 5.32 (d, $J = 16.5$ Hz, 1H), 5.63 (d, $J = 7.9$ Hz, 1H), 5.90 (td, $J = 11.0$, 1H), 7.33 (qd, $J = 7.5, 1.2$, 2H), 7.42 (ttt, $J = 7.5, 1.3, 0.6$, 2H), 7.58 (dt, $J = 7.5, 1.5$, 2H), 7.78 (dt, $J = 7.6, 0.9$, 2H). ^{13}C (400 MHz, CDCl_3) δ 47.10 (s, C), 56.04 (d, CH), 63.48 (t, CH_2), 66.43 (t, CH_2), 67.04 (t, CH_2), 110.16 (t, CH_2), 120.27 (d, CH), 124.88 (d, CH), 127.35 (d, CH), 127.51 (d, CH), 128.17 (t, CH), 141.60 (s, C), 143.48 (s, C). LC-MS (ESI⁺) m/z calculated for $\text{C}_{12}\text{H}_{21}\text{NO}_5$ $[\text{M}+\text{H}]^+$ 367.14, found 368.10.

DIPEPTIDE SYNTHESIS

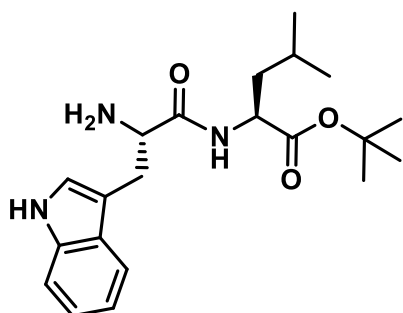
Fmoc-Trp-Leu-OtBu (16)



Fmoc-L-Trp-OH (3.95 g, 9.26 mmol, 1 eq) was dissolved in DCM-DMF (9:1, v/v) (150 ml), and OxymaPure (2.63 g, 18.5 mmol, 2 eq) and EDC·HCl (5.32 g, 27.8 mmol, 3 eq) were added to the stirred solution. After 5 minutes, a solution of H-L-Leu-OtBu (2.07 g, 9.26 mmol, 1 eq) in DCM-DMF (9:1, v/v) (50 mL) was added. The resulting mixture was stirred for 1 h. Then, additional DCM was added to the reaction, which

was extracted with 0.5 M aqueous NaHCO_3 (x 3). The combined organic layers were washed twice with brine, dried over MgSO_4 , filtered and concentrated under *vacuo*. 5.4 g (98%) of product were obtained as a brown-red oil. The crude was used in the next reaction without further purification. RP_HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H_2O) in 8 min; $t_R = 7.520$ min) and LC-MS (ESI⁺) m/z calculated for $\text{C}_{36}\text{H}_{41}\text{N}_3\text{O}_5$ $[\text{M}+\text{H}]^+$ 596.30, found 596.83.

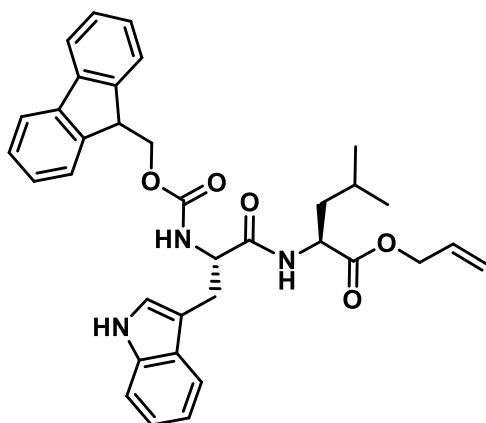
H-Trp-Leu-OtBu (17)



To a solution of **16** (5.4 g, 10.1 mmol, 1 eq) in DCM (15 mL), DEA (5.2 mL, 50.4 mmol, 5 eq) was added and the reaction mixture was stirred for 1 h. Then, the solvent was removed under *vacuo*, followed by consecutive coevaporations with DCM. The residue obtained was purified by reversed-phase C18 flash

column chromatography (H₂O/ACN) to give 1.7 g (45%) of a pale yellow solid. ¹H NMR (400 MHz, CDCl₃) δ 0.72 (d, *J* = 5.6 Hz, 1H), 0.81 (d, *J* = 5.8 Hz, 1H), 1.41-1.54 (m, 12H), 3.20 (dd, *J* = 15.1, 7.6 Hz, 1H), 3.36 (dd, *J* = 15.1, 5.4 Hz, 1H), 4.25 (t, *J* = 6.7 Hz, 2H), 7.13 - 7.02 (m, 3H), 7.23-7.27 (m, 1H), 7.32 (d, *J* = 8.1 Hz, 1H), 7.51 (d, *J* = 8.0 Hz, 1H), 8.69 (s, 1H). ¹³C NMR (400 MHz, CDCl₃) δ 21.78 (q, CH₃), 22.53 (d, CH), 24.76 (t, CH₂), 28.06 (q, CH₃), 31.27 (t, CH₂), 41.05 (d, CH), 51.69 (d, CH), 82.69 (s, C), 106.88 (s, CH), 111.89 (d, CH), 118.36 (s, CH), 119.86 (d, CH), 122.45 (d, CH), 125.32 (d, CH), 126.87 (s, C), 136.49 (s, C). HPLC (linear gradient from 20:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; *t_R* = 5.789 min, 99.9% purity). LC-MS (ESI⁺) *m/z* calculated for C₈H₁₃NO₄ [M+H]⁺ 373.24, found 374.70.

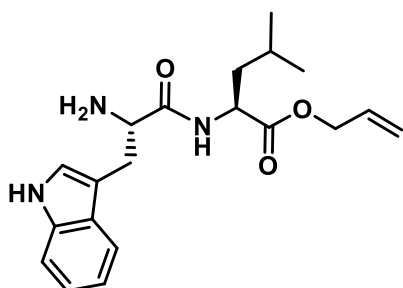
Fmoc-Trp-Leu-OAll (18)



H-L-Leu-OAll (3.12 g, 9.08 mmol, 1 eq) was dissolved in DCM-DMF (9:1, v/v) (150 ml), and OxymaPure (2.58 g, 18.2 mmol, 2 eq) and EDC·HCl (5.25 g, 27.3 mmol, 3 eq) were added to the stirred solution. After 5 min, a solution of Fmoc-L-Trp-OH (3.87 g, 9.08 mmol, 1 eq) in DCM-DMF (9:1, v/v) (50 mL) was added. The resulting mixture was stirred for 1 h. Then, additional DCM was added to the reaction,

which was extracted with 0.5 M aqueous NaHCO₃ (x 3). The combined organic layers were washed twice with brine, dried over MgSO₄, filtered and concentrated under *vacuo*. 5.0 g (95%) of product were obtained as a brown oil. The crude was used in the next reaction without further purification. LC-MS (ESI⁺) *m/z* calculated for C₃₅H₃₇N₃O₅ [M+H]⁺ 579.27, found 580.40.

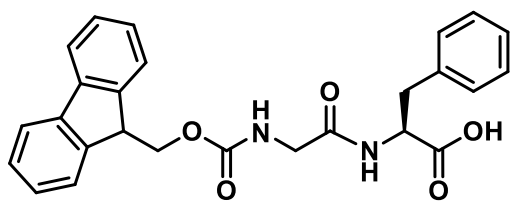
H-Trp-Leu-OAll (19)



To a solution of **18** (5 g, 8.63 mmol, 1 eq) in DCM (15 mL), DEA (4.5 mL, 43.2 mmol, 5 eq) was added, and the reaction mixture was stirred for 1 h. Then, the solvent was removed under *vacuo*, followed by consecutives co-evaporations with DCM. The residue obtained was purified by reversed phase C18 flash

column chromatography (H₂O/ACN) to give 839.0 mg (27%) of a brown oil. ¹H NMR (400 MHz, CDCl₃) δ 0.78 (d, *J* = 6.0 Hz, 3H), 0.82 (d, *J* = 5.9 Hz, 3H), 1.43-1.58 (m, 3H), 3.22 (dd, *J* = 15.0, 7.6 Hz, 1H), 3.36 (dd, *J* = 15.0, 6.0 Hz, 1H), 4.33 (t, *J* = 7.0 Hz, 1H), 4.41 (q, *J* = 7.4, 6.8 Hz, 1H), 4.55 (dd, *J* = 5.8, 1.4 Hz, 2H), 5.25 (dq, *J* = 10.5, 1.2 Hz, 1H), 5.31 (dq, *J* = 17.2, 1.5 Hz, 1H), 5.86 (ddt, *J* = 17.2, 10.4, 5.8 Hz, 1H), 7.05 (td, *J* = 7.5, 7.0, 1.1 Hz, 1H), 7.09 - 7.15 (m, 2H), 7.22 (d, *J* = 7.4 Hz, 1H), 7.33 (d, *J* = 8.1 Hz, 1H), 7.53 (d, *J* = 7.9 Hz, 1H), 8.53 (s, 1H); ¹³C (400 MHz, CDCl₃) δ 21.60 (q), 22.54 (q), 24.73 (t), 40.76 (t), 51.83 (d), 54.07 (d), 66.38 (t), 107.09 (s), 118.32 (d), 119.33 (d), 120.12 (d), 122.70 (d), 126.75 (d), 131.43 (d), 136.50 (t), 140.00 (s), 172.04 (s). HPLC (linear gradient from 20:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; *t_R* = 5.059 min, 99.9% purity). LC-MS (ESI⁺) *m/z* calculated for C₂₀H₂₇N₃O₃ [M+H]⁺ 357.21, found [M+Na]⁺ 379.15.

Fmoc-Gly-Phe-OH (20)

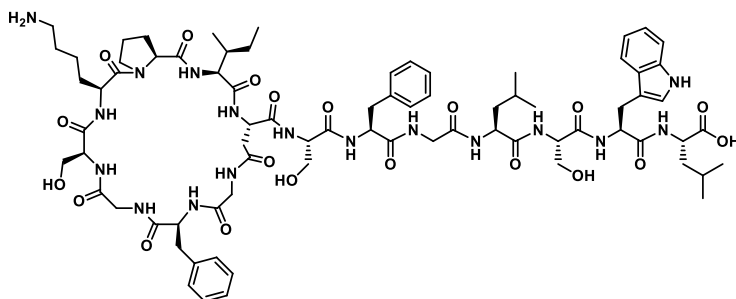


2-Chlorotrityl chloride resin (844.9 mg, 1.4 mmol/g) was placed in a 10-mL polypropylene syringe fitted with a polyethylene filter disc. The resin was then washed with DMF (5 × 1 min) and DCM (5 × 1 min). A solution of Fmoc-L-Phe-OH (458.3 mg, 1.18 mmol) and DIEA (687.6 μL, 5.5 mmol) in DCM was added. After 5 min, more DIEA (1.38 mL, 11.1 mmol) was added and the mixture was stirred for 1 h at room temperature. The reaction was quenched by the addition of MeOH (676 μL) and the mixture was stirred for 10 min to cap any unreacted sites. After filtration, the peptide resin was washed with DCM (5 × 1 min), DMF (5 × 1 min), piperidine-DMF (1:4, 1 × 1 min, 1 × 3 min), and, again, DMF (5 × 1 min) and DCM (5 × 1 min). Fmoc-Gly-OH (1.76 g, 5.9 mmol) was introduced with OxymaPure (840.5 mg, 5.9 mmol) and DIPCDI (915.8 μL, 5.9 mmol) in DMF, and the mixture was stirred for 30 min. After resin filtration and washes, cleavage was affected by treatment with a TFA-DCM (4:96) solution (5 × 1 min) and then washes with DCM. The resulting solution was concentrated under *vacuo* and lyophilized to give 165 mg of white solid (32%), which was used without further purification. ¹H NMR (400 MHz, DMSO-*d*₆) δ 2.88 (dd, *J* = 13.7, 8.7 Hz, 1H), 3.04 (dd, *J* = 13.8, 5.0 Hz, 1H), 3.60 (ddd, *J* = 32.7, 16.9, 6.2 Hz, 3H), 4.28 - 4.18 (m, 3H), 4.44 (td, *J* = 8.4, 5.1 Hz, 1H), 7.27 - 7.16 (m, 6H), 7.32 (t, *J* = 7.1 Hz, 2H), 7.41 (dd, *J* = 7.5, 6.7 Hz, 2H), 7.50 (t, *J* = 6.1 Hz, 1H), 7.71 (d, *J* = 7.5 Hz, 2H), 7.89 (d, *J* = 7.5 Hz, 2H), 8.11 (d, *J* = 7.9 Hz, 1H), 12.82 - 12.70 (m, 1H). ¹³C

NMR (400 MHz, DMSO- d_6) δ 36.81 (t, CH₂), 42.98 (t, CH₂), 46.60 (d, CH), 53.41 (d, CH), 65.72 (t, CH₂), 120.09 (d, CH), 125.25 (d, CH), 126.44 (d, CH), 127.07 (d, CH), 127.61 (d, CH), 128.17 (d, CH), 129.11 (d, CH), 137.37 (s, C), 140.70 (s, C), 143.83 (s, C), 156.40 (s, CO), 168.97 (s, CO), 172.77 (s, CO). HPLC (linear gradient from 0:100 (0.036% TFA in ACN)/0.045% TFA in H₂O) in 8 min; t_R = 6.359 min, 99.9% purity) and LC-MS (ESI⁺) m/z calculated for C₂₆H₂₄N₂O₅ [M+H]⁺ 444.17, found 445.36.

Peptide synthesis

Synthetic sungsanpin



TentaGel S-NH₂ resin (864.6 mg, 0.29 mmol/g) was placed in a 20 mL polypropylene syringe fitted with two polyethylene filter discs. The resin was then washed

with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s); the HMPBA linker and the first amino acid (Fmoc-L-Leu-OH) were incorporated and the excess of reactive positions terminated following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.27 mmol/g.

Peptide elongation was achieved by automatic synthesis following the methodology described in the section 'Automated SPPS', except for the three first amino acids which were coupled manually using Method 1. Fmoc removal was carried out by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min).

The macrolactamization step was carried out manually on solid phase. The 2-PhiPr ester was removed by treatment of the resin with TFA-TIS-DCM (5:1:94) (5 mL, 4 x 10 min). After the four treatments, the resin was washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Then, Fmoc was removed by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min, 1 x 5 min, 1 x 30 min). Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 25:50 (0.036% TFA in ACN)/0.045% TFA in H₂O) in 8 min;

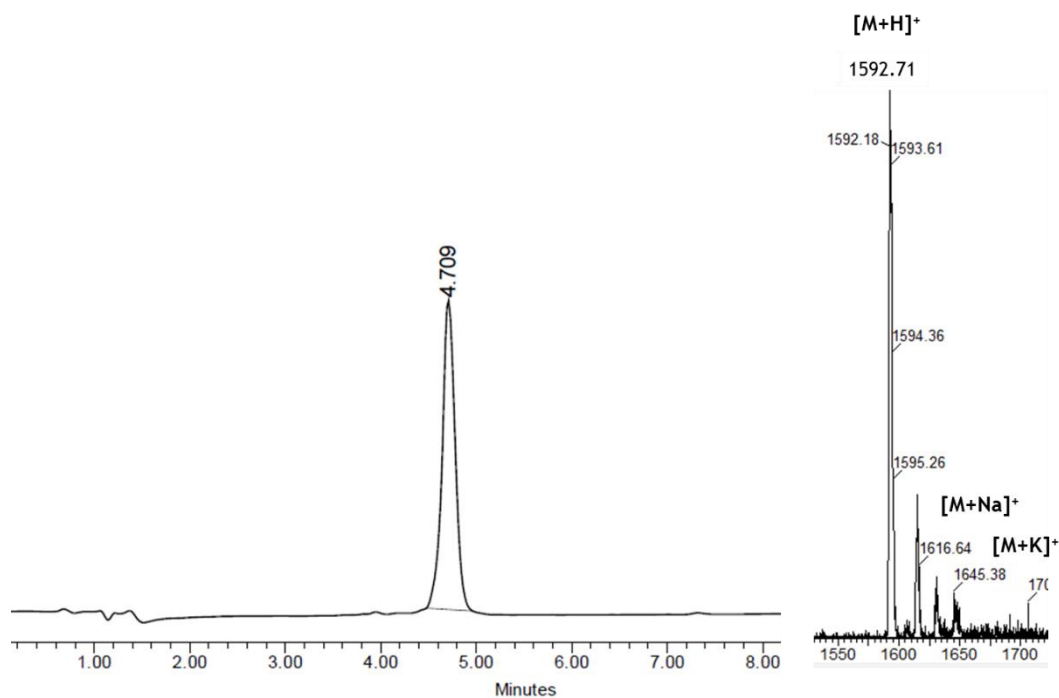
$t_R = 4.621$ min) and LC-MS (ESI⁺) m/z calculated for $C_{77}H_{111}N_{17}O_{21}$ $[M+H]^+$ 1609.81, found 1610.89. After the deprotection, HOAt (95.3 mg, 0.70 mmol, 3 eq) was dissolved in DMF (0.3 mL) and DIEA (0.24 mL, 1.4 mmol, 6 eq) was added. The reaction mixture was then added to the resin followed by the addition of solid PyBOP (364.44 mg, 0.70 mmol, 3 eq). The pH was adjusted by the addition of more DIEA until pH 8. It was allowed to react for 1 h at 25 °C. The solvents were removed and the resin washed with DMF (5 x 30 s) and DCM (5 x 30 s). The cyclization step was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 25:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 6.180$ min) and LC-MS (ESI⁺) m/z calculated for $C_{77}H_{109}N_{17}O_{20}$ $[M+H]^+$ 1591.80, found 1592.87.

The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (10 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated to 1-2 mL and precipitated onto ice-cold Et₂O. Finally, it was lyophilized in H₂O-ACN (1:1) to give 122 mg of crude (33%). RP-HPLC (linear gradient from 25:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 6.178$, 82% purity). LC-MS (ESI⁺) m/z calculated for $C_{77}H_{109}N_{17}O_{20}$ $[M+H]^+$ 1591.80, found 1592.94.

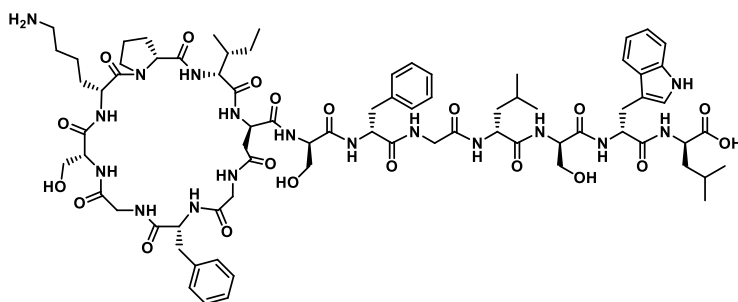
The crude was purified by semi-preparative RP-HPLC using a XBridge® Prep C18 OBD™ 5 μm (19 × 100 mm) column and a linear gradient (0% to 30% over 2 min and 30% to 50% over 20 min) of ACN with a flow rate of 16.0 mL/min to give 32.2 mg of synthetic sungsanpin (9%) with a purity > 99.9%.

Characterization

- RP-HPLC conditions: Sunfire™ C18 3.5 μm (4.6 mm x 100 mm) reversed-phase analytical column; linear gradient from 30% to 60% of ACN over 8 min at 25 °C.
- UPLC-ESI⁺ MS: $C_{77}H_{109}N_{17}O_{20}$ $[M+H]^+$ 1591.80



D-synthetic sungsanpin



TentaGel S-NH₂ resin (865.1 mg, 0.29 mmol/g) was placed in a 20 mL-polypropylene syringe fitted with two polyethylene filter discs.

The resin was then washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s); the HMPBA linker and the first amino acid (Fmoc-D-Leu-OH) were incorporated and the excess of reactive positions terminated following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.26 mmol/g.

Peptide elongation was achieved by automatic synthesis following the methodology described in the section 'Automated SPPS', except the three first amino acids which were coupled manually using the method 1. Fmoc removal was carried out by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min).

The macrolactamization step was carried out manually on solid phase. The 2-PhiPr ester was removed by the treatment of the resin with TFA-TIS-DCM (5:1:94) (5 mL, 4

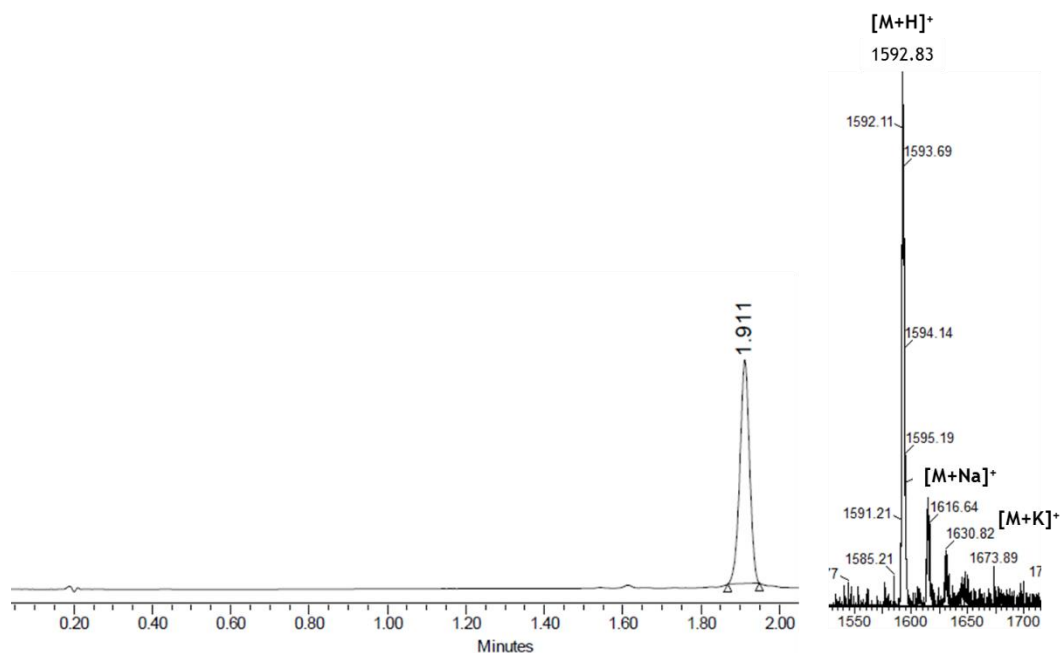
x 10 min). After the four treatments, the resin was washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Then, Fmoc was removed by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min, 1 x 5 min, 1 x 30 min). Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h), followed by analysis of the crude by RP-HPLC (linear gradient from 25:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.315 min) and LC-MS (ESI⁺) *m/z* calculated for C₇₇H₁₁₁N₁₇O₂₁ [M+H]⁺ 1609.81, found 1610.70. After the deprotection, HOAt (91.8 mg, 0.67 mmol, 3 eq) was dissolved in DMF (0.3 mL) and DIEA (0.23 mL, 1.34 mmol, 6 eq) was added. The reaction mixture was then added to the resin followed by the addition of solid PyBOP (351.1 mg, 0.67 mmol, 3 eq). The pH was adjusted by the addition of more DIEA until reaching pH 8 - 9. The cyclization was allowed to react for 1 h at 25 °C. Then, the solvents were removed, and the resin washed with DMF (5 x 30 s) and DCM (5 x 30 s). The cyclization step was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 25:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 6.059 min) and LC-MS (ESI⁺) *m/z* calculated for C₇₇H₁₀₉N₁₇O₂₀ [M+H]⁺ 1591.80, found 1593.10.

The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (10 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated to 1-2 mL and precipitated onto ice-cold Et₂O. It was lyophilized in H₂O-ACN (1:1) to give 90.2 mg of crude (25%). RP-HPLC (linear gradient from 25:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 6.103 min, 85% purity). LC-MS (ESI⁺) *m/z* calculated for C₇₇H₁₀₉N₁₇O₂₀ [M+H]⁺ 1591.80, found 1593.09.

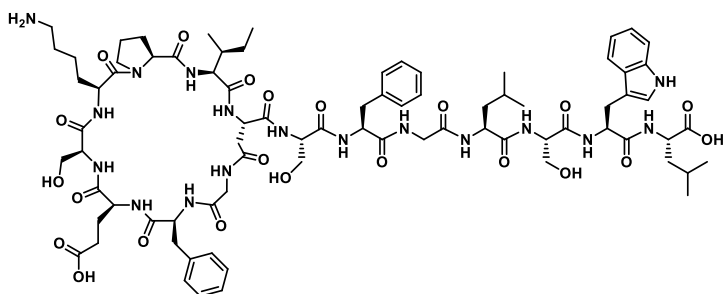
The crude was purified by semi-preparative RP-HPLC using a XBridge® Prep C18 OBD™ 5 μm (19×100 mm) column and a linear gradient (0% to 30% over 2 min and 30% to 50% over 20 min) of ACN with a flow rate of 16.0 mL/min to give 12.5 mg of synthetic sungsanpin (4%) with a purity > 99.9%.

Characterization

- RP-UPLC conditions: Acquity BEH C18 1.7 μm (2 mm x 50 mm) reversed-phase column; linear gradient from 25% to 50% of ACN over 2 min at 40 °C.
- UPLC-ESI⁺ MS: C₇₇H₁₀₉N₁₇O₂₀ [M+H]⁺ 1591.80



Branched cyclic analog



TentaGel S-NH₂ resin (591.4 mg, 0.54 mmol/g) was placed in a 20 mL-polypropylene syringe fitted with two polyethylene filter discs. The resin was then

washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s); the HMPBA linker and the first amino acid (Fmoc-L-Leu-OH) were incorporated and the excess of reactive positions terminated following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.54 mmol/g.

Peptide elongation was achieved by automatic synthesis following the methodology described in the section 'Automated SPPS', except for the three first amino acids which were coupled manually using Method 1. Fmoc removal was carried out treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min).

The macrolactamization step was carried out manually on solid-phase. The 2-PhiPr ester was removed by the treatment of the resin with TFA-TIS-DCM (4:1:95) (5 mL, 4 x 10 min). After the four treatments, the resin was washed with DCM (5 x 30 s), DMF

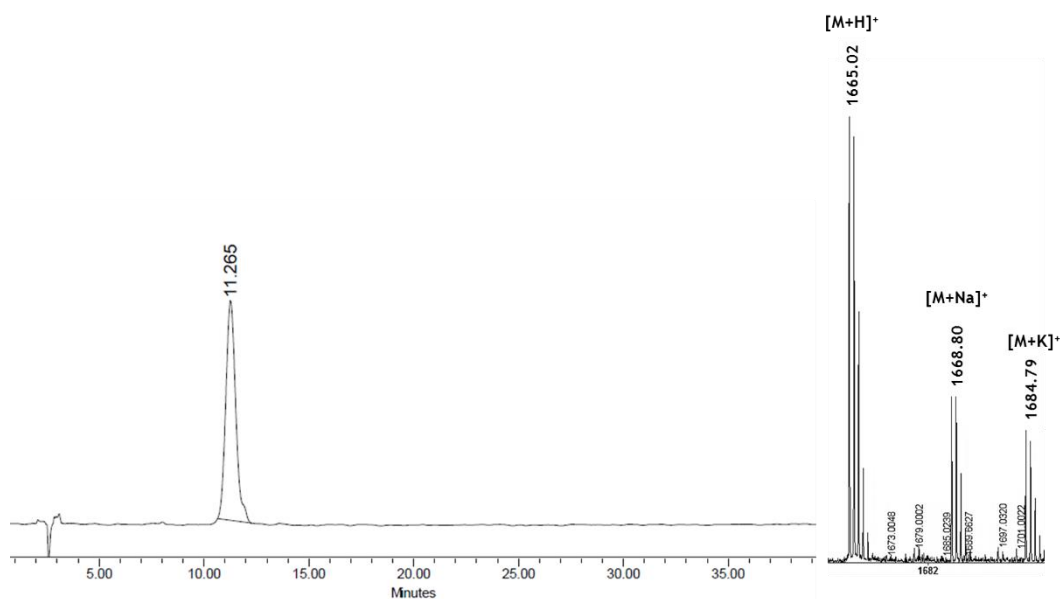
(5 x 30 s) and DCM (5 x 30 s). Then, Fmoc was removed by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min, 2 x 5 min). Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 3.372 min) and LC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₅N₁₇O₂₃ [M+H]⁺ 1681.84, found 1682.8. After the deprotection, HOAt (130.4 mg, 0.96 mmol, 3 eq) was dissolved in DMF (0.3 mL) and DIEA (0.33 mL, 1.92 mmol, 6 eq) was added. The reaction mixture was then added to the resin followed by the addition of solid PyBOP (498.6 mg, 0.96 mmol, 3 eq). The pH was adjusted by the addition of more DIEA until reaching pH 8. The cyclization was allowed to react for 1 h at 25 °C. Then, the solvents were removed and the resin washed with DMF (5 x 30 s) and DCM (5 x 30 s). The cyclization step was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 35:40 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.550 min) and LC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₃N₁₇O₂₂ [M+H]⁺ 1663.82, found 1664.6.

The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (10 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated to 1 -2 mL and precipitated onto ice-cold Et₂O. It was lyophilized in H₂O-ACN (1:1) to give 91.7 mg of crude (19%). RP-HPLC (linear gradient from 35:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.280 min, 82% purity). LC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₃N₁₇O₂₂ [M+H]⁺ 1663.82, found 1664.84.

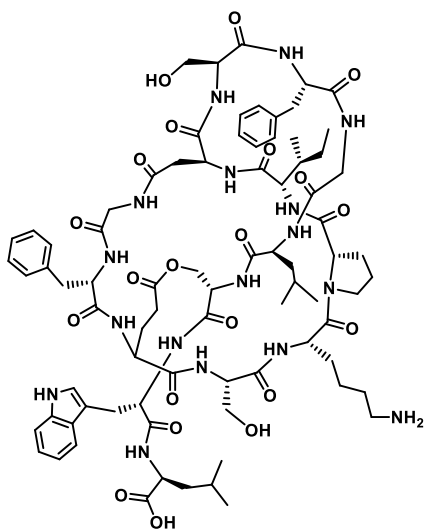
The crude was purified by semi-preparative RP-HPLC using a XBridge® Prep C18 OBD™ 5 μm (19 × 100 mm) column and a linear gradient (0% to 35% over 2 min and 35% to 40% over 30 min) of ACN with a flow rate of 16.0 mL/min to give 21.4 mg of branched cyclic analog (4%) with a purity > 99.9%.

Characterization

- RP-HPLC conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 °C.
- MALDI-TOF: C₈₀H₁₁₃N₁₇O₂₂ [M+H]⁺ 1663.82



Strategy A



TentaGel S-NH₂ resin (4.061 g, 0.54 mmol/g) was placed in a 60 mL-polypropylene syringe fitted with two polyethylene filter discs. The resin was then washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s); the HMPBA linker and the first amino acid (Fmoc-L-Ile-OH) were incorporated and the excess of reactive positions terminated following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.45 mmol/g.

Peptide elongation (Asp⁸-Gly¹-Phe²-Glu³-Ser⁴-Lys⁵-Pro⁶-Ile⁷) was achieved by means of different coupling conditions, being Method 1 the most commonly used. Thus, the incorporation of Boc-L-Asp-O(2-PhiPr)-OH (1.93 g, 5.48 mmol, 3 eq) was accomplished using Method 2. In all cases, Kaiser's test showed a quantitative coupling after 30 min of treatment. Fmoc removal was carried out by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min), except after Fmoc-L-Pro-OH where more time was required (1 x 1 min, 1 x 5 min). Fmoc-L-Lys(pNZ)-OH (3.1 g, 5.48 mmol, 3 eq) was re-coupled by default and chloranil's test was used to check the completion reaction. Allyl ester was removed from Glu following the protocol

described in the section 'Allyl and Alloc deprotection' and then, Fmoc-L-Ser-OFm (6.7 g, 17.8 mmol, 10 eq) dissolved in 8 mL of dry DCM was incorporated to the resin to perform the ester bond with DIPC DI (2.8 mL, 17.8 mmol, 10 eq) and DMAP (111.6 mg, 0.91 mmol, 0.5 eq) dissolved in dry DMF (0.2 mL). The treatment was repeated. The ester formation was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-UPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H₂O) in 2 min; $t_R = 1.444$ min) and LC-MS (ESI⁺) m/z calculated for C₆₉H₈₅N₁₁O₂₂ [M+H]⁺ 1419.59, found 1420.75.

The second peptide chain elongation (Ser⁹-Phe¹⁰-Gly¹¹-Leu¹²) with Alloc-aa was achieved by using coupling conditions from Method 1. The Alloc group was removed following the protocol described in the section 'Allyl and Alloc deprotection'. Thus, the peptide was treated with Pd(PPh₃)₄ (211.2 mg, 0.18 mmol, 0.1 eq) and PhSiH₃ (2.26 mL, 18.25 mmol, 10 eq) dissolved in 2 mL of DCM. After 15 min, the solvents were removed and the resin washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). The treatment was repeated three times. Deprotection completion was monitored by Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-UPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H₂O) in 2 min; $t_R = 1.489$ min) and LC-MS (ESI⁺) m/z calculated for C₈₉H₁₁₃N₁₅O₂₇ [M+H]⁺ 1823.79, found 1824.70.

The first macrolactamization step was carried out on solid phase. First, the Alloc group was removed following the protocol described in the section 'Allyl and Alloc deprotection' and the same procedure than the described above. For the 2-PhiPr ester removal, the resin was treated with TFA-TIS-DCM (5:1:94) (10 mL, 4 x 10 min). After the fourth treatment, the resin was washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-UPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H₂O) in 2 min; $t_R = 1.269$ min) and LC-MS (ESI⁺) m/z calculated for C₈₅H₁₀₉N₁₅O₂₅ [M+H]⁺ 1739.77, found 1741.27. After the deprotection, HOAt (746.2 mg, 5.25 mmol, 3 eq) was dissolved in DMF (0.3 mL) and DIEA (1.91 mL, 10.5 mmol, 6 eq) was added. The reaction mixture was then added to the resin followed by the addition of solid PyBOP (2.86 g, 5.25 mmol, 3 eq). The pH was adjusted by the addition of more DIEA until reaching pH 6. The cyclization was allowed to react for 2 h at 25 °C. The solvents were removed and the resin washed

with DMF (5 x 30 s) and DCM (5 x 30 s). The cyclization step was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-UPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H₂O) in 2 min; t_R = 1.462 min) and LC-MS (ESI⁺) m/z calculated for C₈₅H₁₀₇N₁₅O₂₄ [M+H]⁺ 1721.76, found 1723.11.

The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (10 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated to 1 -2 mL and precipitated onto ice-cold Et₂O. It was lyophilized in H₂O-ACN (1:1) to give 2.86 g of cyclic peptide (91%). RP-UPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H₂O) in 2 min; t_R = 1.462 min, 75% purity). LC-MS (ESI⁺) m/z calculated for C₈₅H₁₀₇N₁₅O₂₄ [M+H]⁺ 1721.8, found 1723.1. The crude was used without further purification.

1.58 g (0.92 mmol, 1 eq) of cyclic peptide was dissolved in 700 mL of DCM-DMF (1:1) at 0 °C in a round-bottom flask. HOAt (260.8 mg, 1.8 mmol, 2 eq) was dissolved in DCM-DMF (1:1) (0.3 mL) and DIEA (0.64 mL, 3.67 mmol, 4 eq) was added. The mixture was then added to the flask followed by the addition of solid PyBOP (955 mg, 1.8 mmol, 2 eq). The pH was adjusted with DIEA until pH 6. The reaction mixture was stirred for 5 min at 0 °C and then it was allowed to reach room temperature for 2 h. The solvents were removed under *vacuo* and then, more DCM was added. The organic layer was extracted with 0.5% acid citric (x 2), dried over MgSO₄, filtered and concentrated under *vacuo* to give an oily crude.

The obtained crude was dissolved in 40 mL of DCM, and DEA (1.4 mL, 13.5 mmol, 10 eq) was added. The mixture was stirred for 1 h at 25 °C. The solvent was removed under *vacuo* and the peptide was precipitated by adding ice-cold Et₂O. It was lyophilized in H₂O-ACN (1:1) to give 802.5 mg (57%) of bicyclic peptide without the Fm ester. RP-HPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.675 min, 61% purity). LC-MS (ESI⁺) m/z calculated for C₇₁H₉₅N₁₅O₂₃ [M+H]⁺ 1525.7, found 1526.8. It was used in the next step without further purification.

The bicyclic peptide without the Fm ester was dissolved in 10 mL of DMF and it was pre-activated with DIPCDI (155.7 μL, 1.05 mmol, 2 eq) and OxymaPure (149.5

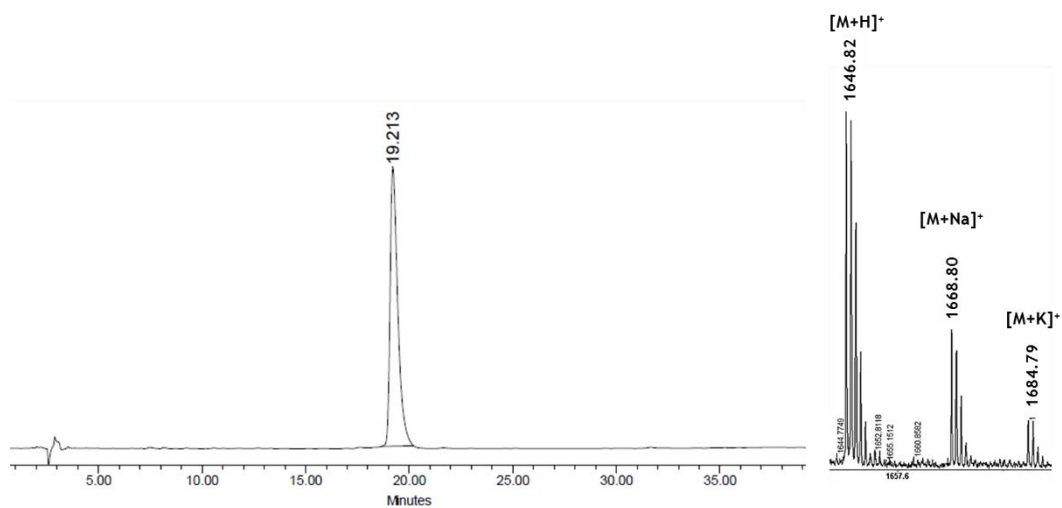
mg, 1.05 mmol, 2 eq) in a round-bottom flask for 5 min at 25 °C. The dipeptide H-Trp-Leu-O^tBu (196.3 mg, 0.53 mmol, 1 eq) was dissolved in 5 mL of DMF and added to the reaction mixture. After 3 h, DCM was added and it was extracted with 0.5% acid citric (x 2), dried over MgSO₄, filtered and concentrated under *vacuo*. No more work-up was required because in the next step, the crude was dissolved in 8 mL of TFA-DCM (1:1) and the mixture was stirred for 30 min. Afterwards, the solvents were removed under *vacuo* and the peptide was precipitated by adding ice-cold Et₂O. It was lyophilized in H₂O-ACN (1:1) to give 365.3 mg of impure bicyclic peptide with the dipeptide tail (21%). RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 6.299 min, 35% purity). LC-MS (ESI⁺) m/z calculated for C₈₈H₁₁₆N₁₈O₂₅ [M+H]⁺ 1824.84, found 1825.96.

Then, the pNZ was removed following the procedure described in the section 'pNZ removal'. The crude was dissolved in 7 mL of DMF and 500 μL of a 6 M solution of SnCl₂ and 1.6 mM of HCl/dioxane was added. The resulting mixture was stirred for 6 h at room temperature with further purification with a PoraPak™ Rxn RP 20 cc cartridge. It was lyophilized in H₂O-ACN (1:1) to give 145.4 mg of final product (10%). RP-HPLC (linear gradient from 30:100 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.016 min, 45% purity). LC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81, found 1647.05.

The final product was re-purified by semi-analytical RP-HPLC using a Xbridge™ Prep BEH130 C18 5 μm (10 mm x 100 mm) column and a linear gradient (5% to 30% over 5 min and 30% to 55% over 20 min) of ACN with a flow rate of 3.0 mL/min. Different peaks with the expected mass were detected, the main peak at t_R = 17.426 min being obtained with a purity of 78%. Then, it was re-purified again using a Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) column and a linear gradient (35% to 40% over 40 min) of ACN with a flow rate of 1.0 mL/min (t_R = 18.853 min) to give 6.7 mg (0.4%) of bicyclic analogue A with a purity of > 99.9%.

HPLC characterization

- RP-HPLC conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 °C.
- MALDI-TOF: C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81



NMR characterization of the bicyclic sungsanpin analog A

The ^1H and ^{13}C NMR (600 MHz, DMSO-d_6) chemical shifts of the bicyclic peptide are detailed in Table 1. The ^1H NMR (600 MHz, $\text{CD}_3\text{CN-H}_2\text{O}$ (1:5)) chemical shifts of the bicyclic peptide are detailed in Table 2.

First conformation				Second conformation		First conformation				Second conformation			
		δ_{H}	δ_{C}	δ_{H}	δ_{C}			δ_{H}	δ_{C}	δ_{H}	δ_{C}		
Gly1	NH	8.64				Ser9	NH	7.27					
	αH	3.58; 3.35	43.25				αH	4.31	53.94				
Phe2	NH	8.48				Phe10	βH	3.42; 3.60	61.90				
	αH	4.12	56.52				NH	8.29					
	βH	3.23; 2.93	42.16				αH	4.37	54.38				
	H2-6	7.17	128.81				βH	3.05; 2.77	36.40				
	H3-5	7.65	131.35				H2-6	7.18	126.14				
Glu3	H4	7.25	126.21			H3-5	7.72	128.44					
	NH	8.11				H4	7.25	128.00					
	αH	4.43	54.21			Gly11	NH	8.08					
	βH	2.84	36.10				αH	3.75; 3.59	46.77				
Ser4	γH	3.06	45.64			Leu12	NH	7.98		NH	7.66		
	NH	8.10					αH	4.41	50.65	αH	4.27	*	
	αH	4.06	55.28				βH	1.58	31.82	βH	1.44	40.47	
Lys5	βH	3.64; 3.75	61.97			γH	1.45	30.80	γH	1.54	23.99		
	NH	8.10				δH	0.87	24.11	δH	0.85	25.12		
	αH	4.23	55.37				0.84	24.18		0.81	24.18		
	βH	1.55	31.94			Ser13	NH	7.92					
	γH	1.20	21.52				αH	4.60	52.85				
	δH	1.45	26.74				βH	4.18	63.32				
Pro6	ϵH	2.71	38.48			Trp14	NH	7.97		NH	7.92		
	NH_2	7.55					αH	4.55	49.03	αH	4.59	49.02	
	αH	4.17	61.39	αH	4.17		60.31	βH	3.15; 2.92	27.64	βH	3.15; 2.92	*
	βH	2.11	28.63	βH	2.11		31.59	H2	7.13	123.30	H2	7.13	122.07
	γH	1.88	24.82	γH	1.88		21.87	H4	7.58	118.22	H4	7.53	118.22
Ile7	δH	3.58	42.24	δH	3.58	41.90	H5/6	7.05	120.67	H5/6	7.03	118.01	
	NH	7.80					H6/5	6.98	118.01	H6/5	6.95	*	
	αH	4.14	55.95				H7	7.32	111.15	H7	7.30	*	
	βH	1.80	34.64				Hind	10.76		Hind	10.71		
	γH	1.34; 1.01	24.48			Leu15	NH	8.21		NH	7.52		
	0.78	14.78			αH		4.24	50.14	αH	4.36	50.59		
δH	0.73	10.01			βH		1.51	39.87	βH	1.54	*		
Asp8	NH	8.62					γH	1.61	24.19	γH	1.45	*	
	αH	4.62	51.82				δH	0.82	21.40	δH	0.84	*	
	βH	2.61; 39.46					0.87	22.83		*	*		

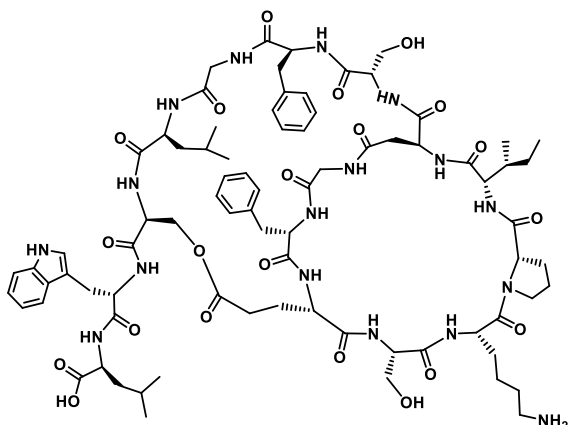
2.70

Table 1: Chemical shifts of ^1H and ^{13}C (ppm) for the bicyclic sungsanpin analog A in DMSO-d_6 at 298 K. (*) Missing chemical shifts.

	First conformation		Second conformation			First conformation		Second conformation	
		δ_{H}		δ_{H}			δ_{H}		δ_{H}
Gly1	NH	8.30			Ser9	NH	8.06		
	αH	3.69; 3.53				αH	4.45		
Phe2	NH	8.08			Phe10	βH	3.93; 3.88		
	αH	4.49				NH	8.38		
	βH	3.14; 2.92				αH	4.54		
	H2-6				βH	3.26; 3.15			
	H3-5				H2-6				
	H4				H3-5				
Glu3	NH	8.20				H4			
	αH	4.33			Gly11	NH	7.96		
βH	2.67; 2.54			αH		3.82; 3.60			
	γH	2.93			Leu12	NH	7.66		
Ser4	NH	8.15				αH	4.17		
	αH	4.86				βH	1.32		
	βH	4.15; 3.86			γH	1.46			
Lys5	NH	7.86			δH	0.76			
	αH	4.28				*			
	βH	1.30; 1.19			Ser13	NH	7.32		
	γH	1.54				αH	4.32		
	δH	*				βH	3.66		
		ϵH	2.83			Trp14	NH	7.83	NH
	NH_2	7.35		αH	4.65		αH	4.67	
Pro6	αH	3.61	αH	3.48	βH	3.24; 3.10	βH	3.19; 3.11	
	βH	2.16	βH	2.26; 2.01	H2	7.11	H2	7.07	
	γH	1.88	γH	1.85; 1.76	H4	7.36	H4	7.36	
	δH	3.40	δH	3.39	H5/6	7.48	H5/6	7.48	
Ile7	NH	7.90			H6/5	7.48	H6/5	7.48	
	αH	4.37			H7	7.55	H7	7.54	
	βH	1.72			Hind	9.95	Hind	9.92	
	γH	1.57			Leu15	NH	7.51	NH	7.45
		*				αH	4.16	αH	4.13
δH	0.80			βH		1.44	βH	1.43	
Asp8	NH	8.43			γH	1.30	γH	1.33	
	αH	4.40			δH	0.74	δH	0.72	
	βH	2.91; 3.07				*		*	

Table 2: Chemical shifts of ^1H (ppm) for the bicyclic sungsanpin analog A in $\text{CD}_3\text{CN-H}_2\text{O}$ (1:5) at 298 K. (*) Missing chemical shifts.

Strategy C



TentaGel S-NH₂ resin (1.897 g, 0.26 mmol/g) was placed in a 20 mL-polypropylene syringe fitted with two polyethylene filter discs. The resin was then washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s); the HMPBA linker and the first amino acid (Fmoc-L-Leu-OH) were incorporated and the excess of reactive positions terminated

following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.25 mmol/g.

Peptide chain elongation (Gly1-Phe2-Glu3-Ser4-Lys5-Pro6-Ile7-Asp8-Ser9-Phe10-Gly11-Leu12-Ser13-Trp14-Leu15) was achieved by means of different coupling conditions, being Method 1 the most commonly used. Thus, the incorporation of the following Fmoc-L-Leu-OH at position 12 (522.8 mg, 1.48 mmol, 3 eq) was accomplished using Method 2. In all cases, Kaiser's test showed a quantitative coupling after 30 min of treatment. Fmoc removal was carried out by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min), except after Fmoc-L-Pro-OH where more time was required (1 x 1 min, 1 x 5 min). Fmoc-L-Lys(Boc)-OH (1.16 mg, 2.47 mmol, 5 eq) was re-coupled by default and chloranil's test was used to check the completeness of the reaction.

The macrolactamization step was carried out on solid phase. The 2-PhiPr ester was removed by the treatment of the resin with TFA-TIS-DCM (5:1:94) (5 mL, 4 x 10 min). After the four treatments, the resin was washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Then, Fmoc was removed by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min, 1 x 5 min, 1 x 30 min). Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 3.858 min) and UPLC-MS (ESI⁺) m/z calculated for C₈₃H₁₁₉N₁₇O₂₃ [M+H]⁺ 1721.87, found 1722.89. After the deprotection, HOAt (201.4 mg, 1.48 mmol, 3 eq) was dissolved in DMF (0.3 mL) and DIEA (0.52 mL, 2.96 mmol, 6 eq) was added. The

reaction mixture was then added to the resin followed by the addition of solid PyBOP (769.9 mg, 1.48 mmol, 3 eq). The pH was adjusted by the addition of more DIEA until pH 7 - 8. The cyclization was allowed to react for 1.5 h at 25 °C. Then, the solvents were removed, and the resin washed with DMF (5 x 30 s) and DCM (5 x 30 s). The cyclization step was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.593 min) and UPLC-MS (ESI⁺) m/z calculated for C₈₃H₁₁₇N₁₇O₂₂ [M+H]⁺ 1703.86, found 1705.01.

Next, to remove the allyl group the conditions described in the section 'Allyl and Alloc deprotection' were used. Thus, the peptide was treated with Pd(PPh₃)₄ (56.9 mg, 0.05 mmol, 0.1 eq) and PhSiH₃ (0.61 mL, 4.93 mmol, 10 eq) dissolved in 1 mL of DCM. After 15 min, the solvents were removed and the resin washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). The treatment was repeated three times. Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 3.909 min) and UPLC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₃N₁₇O₂₂ [M+H]⁺ 1663.82, found 1665.41..

The formation of the ester bond was carried out on solid phase. The resin was washed with ACN (5 x 30 s) and then, DIPCDI (0.62 mL, 3.95 mmol, 8 eq) was added to the resin together with 5 mL of ACN. The resin was stirred for 3 min and then, DMAP (48.2 mg, 0.39 mmol, 0.8 eq) dissolved in dry DMF (0.2 mL) was added, and the mixture was allowed to react 2 h. Then, the solvents were removed, and the resin was washed with ACN (5 x 30 s), DMF (5 x 30 s), DCM (5 x 30 s), ACN (5 x 30 s) and the treatment was repeated. The ester bond formation was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.033 min) and UPLC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81, found 1646.66.

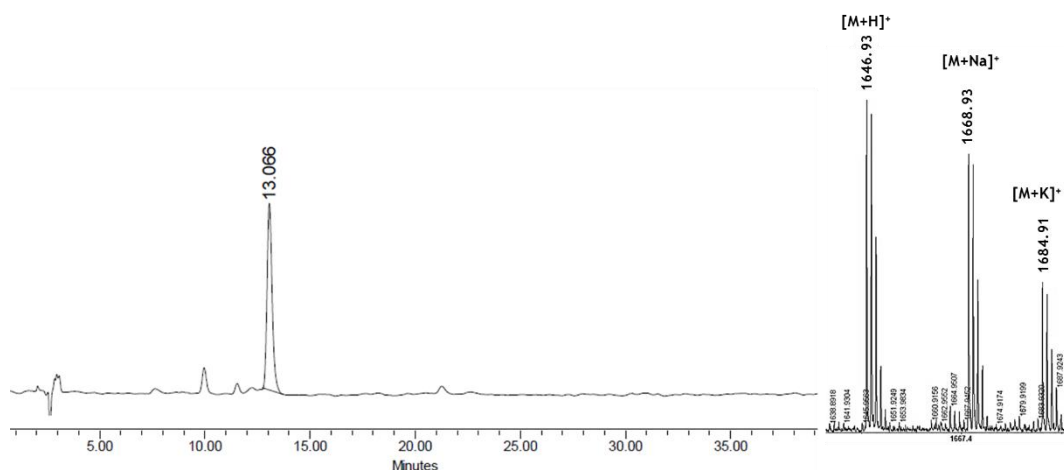
The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (10 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated to 1 -2 mL and precipitated onto ice-cold Et₂O. It was lyophilized in

H₂O-ACN (1:1) to give 103 mg of impure final product (13%). RP-HPLC (linear gradient from 35:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 4.071$ min, 25% purity). LC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81, found 1646.66.

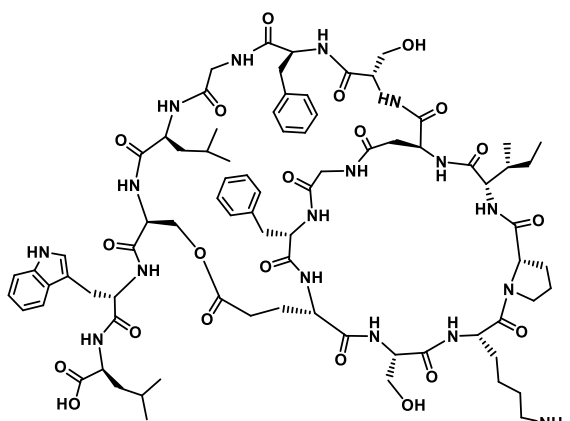
The crude was purified by semi-preparative RP-HPLC using a XBridge® Prep C18 OBD™ 5 μm (19×100 mm) column and a linear gradient (0% to 35% over 2 min and 35% to 50% over 30 min) of ACN with a flow rate of 16.0 mL/min to give 4.7 mg of bicyclic analogue C (0.6%) with a purity of 93%.

Characterization

- RP-HPLC conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 °C.
- MALDI-TOF: C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81



Strategy D



TentaGel S-NH₂ resin (480 mg, 0.26 mmol/g) was placed in a 20 mL-polypropylene syringe fitted with two polyethylene filter discs. The resin was then washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s); the HMPBA linker and the first amino acid (Fmoc-L-Leu-OH) were incorporated and the excess of reactive positions terminated

following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.26 mmol/g.

Peptide chain elongation (Ser4-Lys5-Pro6-Ile7-Asp8-Ser9-Phe10-Gly11-Leu12-Ser13-Trp14-Leu15) was achieved by means of different coupling conditions, being Method 1 the most commonly used. Thus, the incorporation of the following Fmoc-L-Leu-OH in position 12 (132.3 mg, 0.37 mmol, 3 eq) was accomplished using Method 2. In all cases, Kaiser's test showed a quantitative coupling after 30 min of treatment. Fmoc removal was carried out by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min), except after Fmoc-L-Pro-OH where more time was required (1 x 1 min, 1 x 5 min). Fmoc-L-Lys(Boc)-OH (292.4 mg, 0.62 mmol, 5 eq) was re-coupled by default and chloranil's test was used to check the completion of the reaction. The peptide chain elongation was performed until the incorporation of Alloc-L-Ser(^tBu)-OH and then, the ester bond formation was carried out.

After washings with DCM (5 x 30 s), the resin was dried under vacuum for 10 min and transferred to a PYREX® culture tube provided with a magnetic stirrer. pNZ-L-Glu-OAll (365.5 mg, 1.0 mmol, 8 eq) was dissolved in dry DCM (2 mL) and DIPCDI (0.15 mL, 1.0 mmol, 8 eq) was added. After stirring for 3 min, the mixture was incorporated onto the resin. DMAP (7.62 mg, 0.06 mmol, 0.5 eq) dissolved in dry DMF (0.2 mL) was finally added to the tube. The mixture was reacted for 3 h at 40 °C with smooth stirring. The reaction was transferred back to the syringe, the solvents removed by filtration, and the resin washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Ester bond formation was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 5.651$ min) and UPLC-MS (ESI⁺) m/z calculated for C₈₄H₁₁₆N₁₆O₂₇ [M+H]⁺ 1780.80, found 1781.70.

The pNZ group was then removed following the procedure described in the section 'pNZ removal'. A solution of 6 M SnCl₂ in DMF and 1.6 mM of HCl/dioxane was added to the resin (2 x 45 min). After each treatment, the resin was washed DMF (5 x 30 s), DCM (5 x 30 s), and DMF (5 x 30 s).

The last two moieties were incorporated to the growing peptide chain as a single unit. Fmoc-Gly-Phe-OH (>99.9% purity, 277.18 mg, 0.62 mmol, 5 eq), DIPCDI

(96.6 μL , 0.62 mmol, 5 eq) and OxymaPure (88.7 mg, 0.62 mmol, 5 eq) were dissolved in DMF (4 mL) and pre-activated for 3 min. Then, the mixture was added to the resin and it was allowed to react for 45 min, followed by a re-coupling for further 30 min. The solvents were removed, and the resin was washed with DMF (5 x 30 s) and DCM (5 x 30 s). Dipeptide coupling completion was monitored by the Kaiser test and cleavage of an aliquot of resin with TFA-TIS- H_2O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H_2O) in 8 min; t_{R} = 6.897 min) and UPLC-MS (ESI⁺) m/z calculated for $\text{C}_{102}\text{H}_{132}\text{N}_{16}\text{O}_{28}$ $[\text{M}+\text{H}]^+$ 2027.96, found 2029.70.

Next, the allyl and Alloc group were removed following the conditions described in the section 'Allyl and Alloc deprotection'. Thus, the peptidyl-resin was treated with $\text{Pd}(\text{PPh}_3)_4$ (14.4 mg, 0.01 mmol, 0.1 eq) and PhSiH_3 (0.15 mL, 1.25 mmol, 10 eq) dissolved in 1 mL of DCM. After 15 min, the solvents were removed and the resin washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). The treatment was repeated three times. Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS- H_2O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H_2O) in 8 min; t_{R} = 5.583 min) and UPLC-MS (ESI⁺) m/z calculated for $\text{C}_{95}\text{H}_{124}\text{N}_{16}\text{O}_{26}$ $[\text{M}+\text{H}]^+$ 1903.91, found 1904.80. Subsequently, DIPCDI (96.6 μL , 0.62 mmol, 5 eq) and OxymaPure (88.7 mg, 0.62 mmol, 5 eq) were dissolved in 3 mL of DMF and the mixture was added to the resin. The reaction mixture was allowed to react for 1 h at 25 °C. The amide formation was monitored by cleavage of an aliquot of resin with TFA-TIS- H_2O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H_2O) in 8 min; t_{R} = 6.730 min) and UPLC-MS (ESI⁺) m/z calculated for $\text{C}_{95}\text{H}_{122}\text{N}_{16}\text{O}_{25}$ $[\text{M}+\text{H}]^+$ 1886.88, found 1887.50, as the Kaiser's test was not reliable at this point of the synthesis.

The macrolactamization step between *N*-terminal Gly1 and the side chain of Asp8 was carried out on solid phase. The 2-PhiPr ester was removed by the treatment of the resin with TFA-TIS-DCM (3:1:96) (4 mL, 4 x 10 min). After the four treatments, the resin was washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Then, Fmoc was removed by treating the resin with DEA-DMF (1:4) with 0.1 M OxymaPure (1 x 30 s, 1 x 1 min, 1 x 5 min). Deprotection completion was monitored by cleavage of an aliquot of resin with TFA-TIS- H_2O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis

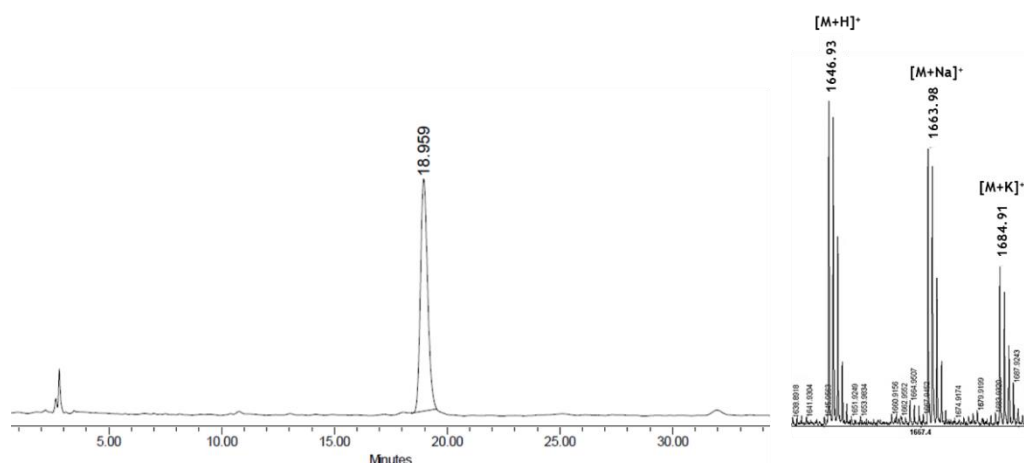
of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 4.052$ min) and UPLC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₂N₁₆O₂₃ [M+H]⁺ 1663.81, found 1665.10. After the deprotection, DIPCDI (96.6 μ L, 0.62 mmol, 5 eq) and OxymaPure (88.7 mg, 0.62 mmol, 5 eq) were dissolved in 1.5 mL of DMF and added to the resin. The reaction mixture was allowed to react for 1 h at 25 °C. The cyclization step was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 4.853$ min) and UPLC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81, found 1647.10.

The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (5 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated and lyophilized in H₂O-ACN (1:1) to give 3.2 mg of impure final product (2%). RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 5.077$, 43% purity). LC-MS (ESI⁺) m/z calculated for C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81, found 1646.66.

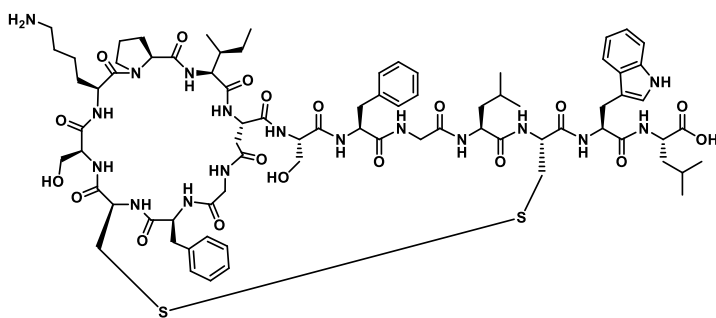
The crude was purified by semi-analytical RP-HPLC using a Sunfire® C18 OBD™ Prep Column 5 μ m (10 \times 150 mm) column and a linear gradient (35% to 40% over 35 min) of ACN with a flow rate of 2.0 mL/min to 1.16 mg of bicyclic analog D (0.7%) with a purity of 90%.

Characterization

- RP-HPLC conditions: Phenomenex Luna C18 5 μ m (4.6 mm \times 250 mm) reversed-phase analytical column; linear gradient from 35% to 40% of ACN over 40 min at 25 °C.
- MALDI-TOF: C₈₀H₁₁₁N₁₇O₂₁ [M+H]⁺ 1645.81



Bicyclic analog via disulfide bond Cys3 - Cys13



TentaGel S-NH₂ resin (780.3 mg, 0.26 mmol/g) was placed in a 20 mL - polypropylene syringe fitted with two polyethylene filter discs. The resin was then washed with DCM (5 x 30 s),

DMF (5 x 30 s) and DCM (5 x 30 s); the Fmoc-Rink amide linker, the HMBA linker and the first amino acid (Fmoc-L-Leu-OH) were introduced and the excess of reactive positions terminated following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.20 mmol/g.

Peptide chain elongation (from Gly1 to Leu15) was achieved by means of different coupling conditions, being Method 1 the most commonly used. In all cases, Kaiser's test showed a quantitative coupling after 30 min of treatment. Fmoc removal was carried out treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min), except after Fmoc-L-Pro-OH where more time was required (1 x 1 min, 1 x 5 min). Fmoc-L-Lys(Boc)-OH (292.4 mg, 0.62 mmol, 5 eq) was re-coupled by default and chloranil's test was used to check the completion of the reaction.

Next, the Mmt group was removed by the treatment of the resin with TFA-TIS-DCM (2:10:88) (8 mL, 6 x 3 min). After the sixth treatment, the resin was washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Then, *N*-chlorosuccinimide (NCS)

(54.1 mg, 0.41 mmol, 2 eq) dissolved in 5 mL of DMF was added. After 15 min, the solvents were removed and the resin washed with DMF (5 x 30 s) and DCM (5 x 30 s). Oxidation completion was monitored by the Ellman test.

Fmoc was removed by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min, 1 x 5 min). Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.483 min) and UPLC-MS (ESI⁺) m/z calculated for C₈₆H₁₁₈N₁₈O₂₁S₂ [M+H]⁺ 1802.82, found 1803.65.

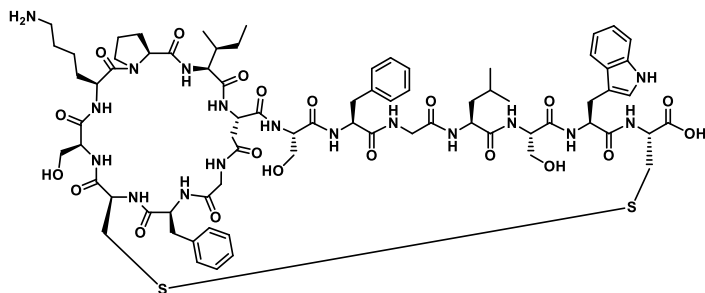
The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (10 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated and lyophilized in H₂O-ACN (1:1) to give 85 mg (30%) of cyclic peptide through disulfide bond. RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.495 min, 33% purity). LC-MS (ESI⁺) m/z calculated for C₈₆H₁₁₈N₁₈O₂₁S₂ [M+H]⁺ 1802.82, found 1803.50. Subsequently, the crude was purified with a PoraPak™ Rxn RP 20 cc cartridge using an ACN gradient (0% to 50%) to elute the product. 21.7 mg of purified product (8%) was obtained. RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.377 min, 68% purity).

The macrolactamization step was carried out in solution. The purified crude was dissolved in 25 mL of DCM-DMF (1:1) at 0 °C in a round-bottom flask. HOAt (4.1 mg, 0.03 mmol, 2 eq) was dissolved in DCM-DMF (1:1) (0.3 mL) and DIEA (10.5 μL, 0.06 mmol, 4 eq) was added. The mixture was then added to the flask followed by the addition of solid PyBOP (15.6 mg, 0.03 mmol, 2 eq). The pH was adjusted with DIEA until pH 6. The reaction mixture was stirred at 0 °C for 2 h. The solvents were removed under *vacuo*. The cyclization was monitored by analysis of the crude by RP-HPLC and HPLC-MS. Two peaks with the desired mass appeared (see Figure 32, Chapter 1).

The synthesis was continued and in the same reaction flask, 10 mL of H₂O-ACN (1:1) was added, and the mixture was cooled down to 0 °C, and a saturated solution of Na₂CO₃ was added until pH 12 was attained. After 30 min, the reaction was

allowed to warm to room temperature for 10 h. A mix of products and a messy chromatogram were obtained, making impossible the further purification.

Bicyclic analog via disulfide bond Cys3 - Cys15



TentaGel S-NH₂ resin (668.7 mg, 0.54 mmol/g) was placed in a 20 mL-polypropylene syringe fitted with two polyethylene filter discs. The resin was then washed with DCM (5 x 30 s), DMF (5 x 30 s)

and DCM (5 x 30 s); the HMPBA linker and the first amino acid (Fmoc-L-Cys(Trt)-OH) were incorporated, and the excess of reactive positions terminated following the procedure described in the section 'conditioning of the resin'. The loading, as calculated by UV absorbance at 290 nm, was 0.54 mmol/g.

Peptide chain elongation (from Gly1 to Leu15) was achieved by automatic synthesis following the methodology described in the section 'Automated SPPS', except for the three first amino acids which were coupled manually using Method 1. Fmoc removal was carried out by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min).

The macrolactamization step was carried out manually on solid phase. The 2-PhiPr ester was removed by treatment of the resin with TFA-TIS-DCM (4:1:95) (10 mL, 4 x 10 min). After the four treatments, the resin was washed with DCM (5 x 30 s), DMF (5 x 30 s) and DCM (5 x 30 s). Then, Fmoc was removed by treating the resin with piperidine-DMF (1:4) (1 x 1 min, 1 x 3 min, 1 x 5 min, 1 x 30 min). Deprotection completion was monitored by the Kaiser test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; *t_R* = 2.987 min) and UPLC-MS (ESI⁺) *m/z* calculated for C₇₆H₁₀₈N₁₆O₂₁S₂ [M+H]⁺ 1644.73, found 1646.50. After the deprotection, HOAt (147.4 mg, 1.08 mmol, 3 eq) was dissolved in DMF (0.3 mL) and DIEA (0.38 mL, 2.17 mmol, 6 eq) was added. The reaction mixture was then added to the resin followed by the addition of solid PyBOP (563.74 mg, 1.08 mmol, 3 eq). The pH was adjusted by the addition of more DIEA

until pH 7 - 8 was attained. The cyclization was allowed to react for 1 h at 25 °C. The solvents were removed and the resin washed with DMF (5 x 30 s) and DCM (5 x 30 s). The cyclization step was monitored by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 3.850 min) and UPLC-MS (ESI⁺) m/z calculated for C₇₆H₁₀₆N₁₆O₂₀S₂ [M+H]⁺ 1626.72, found 1628.60.

The cleavage of the peptide and the elimination of side-chain protecting groups were accomplished simultaneously by treating the resin with TFA-TIS-H₂O (95:2.5:2.5) (10 mL for 1 h). The filtrate was collected in a round-bottom flask, concentrated to 1 - 2 mL and precipitated onto ice-cold Et₂O. It was lyophilized in H₂O-ACN (1:1) to give 240 mg of impure cyclic peptide (41%). RP-HPLC (linear gradient from 30:80 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 3.956 min, 47% purity). LC-MS (ESI⁺) m/z calculated for C₇₆H₁₀₆N₁₆O₂₀S₂ [M+H]⁺ 1626.72, found 1627.7.

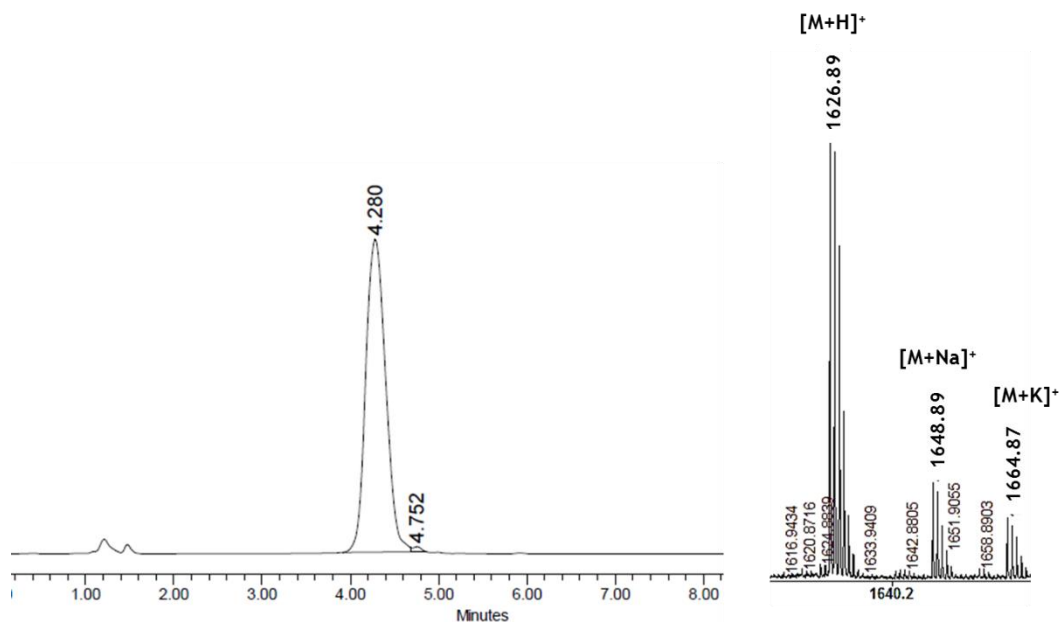
The cyclic peptide was dissolved in 300 mL of H₂O with 5% DMSO and a saturated solution of Na₂CO₃ was added until reaching basic pH. The reaction mixture was allowed to react overnight at 25 °C. The disulfide formation was monitored by Ellman's test and by cleavage of an aliquot of resin with TFA-TIS-H₂O (95:2.5:2.5) (0.5 mL for 1 h) followed by analysis of the crude by RP-HPLC (linear gradient from 30:50 (0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; t_R = 4.846 min) and UPLC-MS (ESI⁺) m/z calculated for C₇₆H₁₀₄N₁₆O₂₀S₂ [M+H]⁺ 1624.72, found 1626.60. It was lyophilized in H₂O-ACN (1:1) to give 195 mg of oxidized product (33%).

The oxidized product was purified by semi-preparative RP-HPLC using a XBridge® Prep C18 OBD™ 5 μm (19 × 100 mm) column and a linear gradient (0% to 30% over 2 min and 30% to 50% over 30 min) of ACN with a flow rate of 16.0 mL/min to give 3.0 mg of bicyclic analog (0.5%) with a disulfide bond between Cys3 and Cys15 with a purity of 72%.

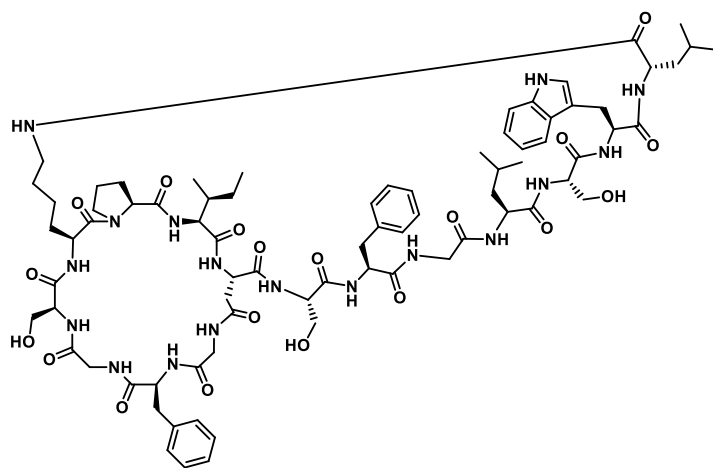
Characterization

- RP-HPLC conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 60% of ACN over 8 min at 25 °C.

- MALDI-TOF: $C_{76}H_{104}N_{16}O_{20}S_2$ $[M+H]^+$ 1624.72



Bicyclic analog side-chain-to-tail via amide bond

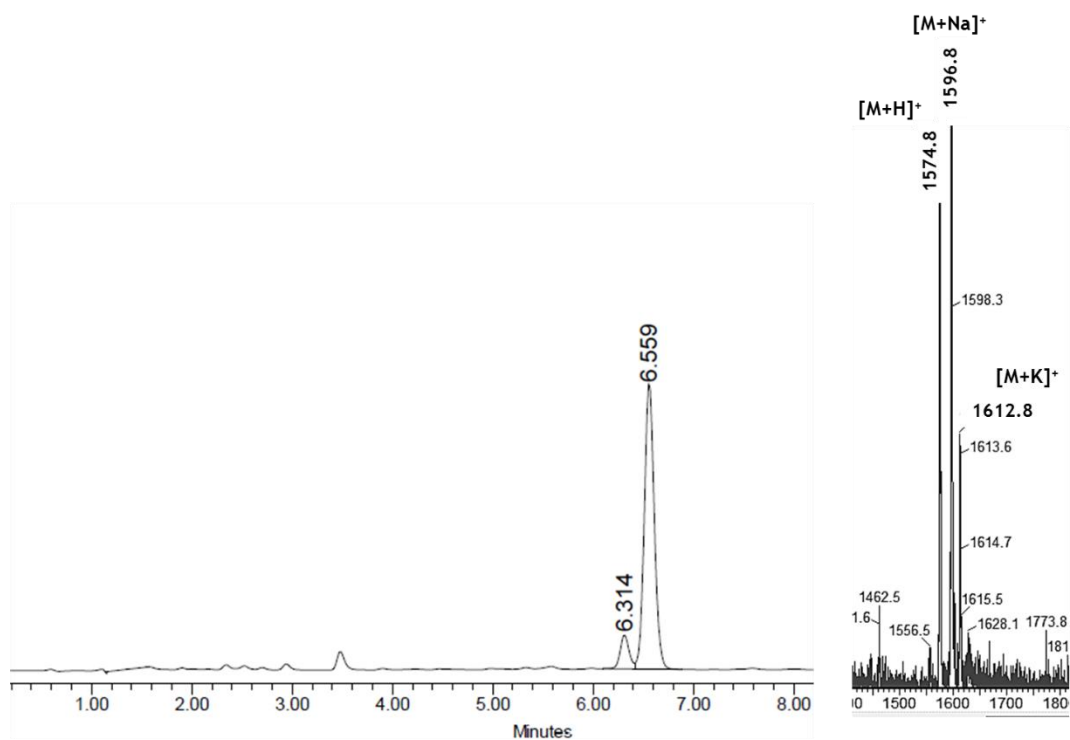


5.5 mg (3.5 μmol , 1 eq) of synthetic sungsanpin was dissolved in 6 mL of DCM-DMF (1:1) at 0 °C. HOAt (0.94 mg, 6.9 μmol , 2 eq) was dissolved in DMF (0.2 mL) and DIEA (2.4 μL , 13.8 μmol , 4 eq) was added. The reaction mixture was then added to the peptide solution followed by the addition of solid PyBOP (3.6 mg, 6.9 μmol , 2 eq). The pH was adjusted by the addition of more DIEA until pH 7 was attained. The reaction was stirred 5 min at 0 °C and then, it was allowed to warm to room temperature for 1 h. The solvents were removed under *vacuo* with the addition of more DCM. The organic layer was extracted with 0.5% acid citric solution (x 2), dried over MgSO_4 , filtered and concentrated under *vacuo*. Some product moved to the aqueous phase, so it was purified with a PoraPak™ Rxn RP 6 cc cartridge using an ACN gradient (0% to 50%). 3.1 mg of bicyclic product (56%) were obtained. RP-HPLC (linear gradient from 35:50

(0.036% TFA in ACN/0.045% TFA in H₂O) in 8 min; $t_R = 6.559$ min, 96% purity). LC-MS (ESI⁺) m/z calculated for C₇₇H₁₀₇N₁₇O₁₉ [M+H]⁺ 1573.80, found 1574.91.

Characterization

- RP-HPLC conditions: Xbridge™ C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 35% to 50% of ACN over 8 min at 25 °C.
- UPLC-ESI⁺ MS: C₇₇H₁₀₇N₁₇O₁₉ [M+H]⁺ 1573.80



**ANNEX: MICROCIN J25, A
BASIC-SENSITIVE LASSO
PEPTIDE**

Introduction

This chapter has been considered as an independent part of the main body of the thesis, as it was a parallel study performed mainly with antimicrobial—instead of anticancer—lasso peptides, which are the main goal of the present thesis. This investigation has been carried out in collaboration with Prof. Miquel Viñas from the Medical School-IDIBELL of the University of Barcelona.

This study has been performed after the results obtained after the basic treatment of the bicyclic analog A (Chapter 2). The ester bond was hydrolyzed, and the supposed lasso structure produced by the bicycle was not maintained after the hydrolysis. As a result, the obtained structure was chemically labile; in addition, the NMR experiments showed a clear branched cyclic structure. The loss of the initial structure suggested that the ester bond was the responsible of maintaining the bicyclic structure and providing chemical stability. Regarding this, we wanted to evaluate if the same result could be achieved by applying the same treatment to a natural lasso peptide. Thus, the basic treatment was applied to the chaxapeptin, sphingopyxin I and MccJ25. For more in-depth studies and further characterization, MccJ25 was selected because more amount of peptide was required (MccJ25 can be easily expressed in large amounts).

Overview about Microcin J25

MccJ25 is a 21-residue peptide with an 8-residue macrolactam ring, formed between the *N*-terminal Gly1 and the Glu8 side-chain, and a 13-residue *C*-terminal tail, which is threaded through the ring (Figure 1). It is classified as a class II lasso peptide, in which the residues Phe19 and Tyr20 are the steric locks of the lasso peptide. The threaded-lasso structure is stabilized by two short double-stranded antiparallel β -sheets. The first sheet comprises residues 6-7 and 18-19 and it is formed between part of the ring and the threaded *C*-terminal tail. The second sheet, which involves residues 10-11 and 15-16, is associated with a β -turn involving residues 11-14, and forms a hairpin-like structure.¹ MccJ25 shows the largest loop among all lasso peptides, with 11 residues, which has been used for epitope grafting.^{2,3} It was isolated from the *E. coli* AY25 strain in 1992 and showed antimicrobial activity against a range of Gram-negative bacteria.⁴ In particular, it

exhibits remarkable antibiotic activity towards *Salmonella Newport* and several *E.coli* strains, with minimum inhibitory concentrations typically in the submicromolar range.

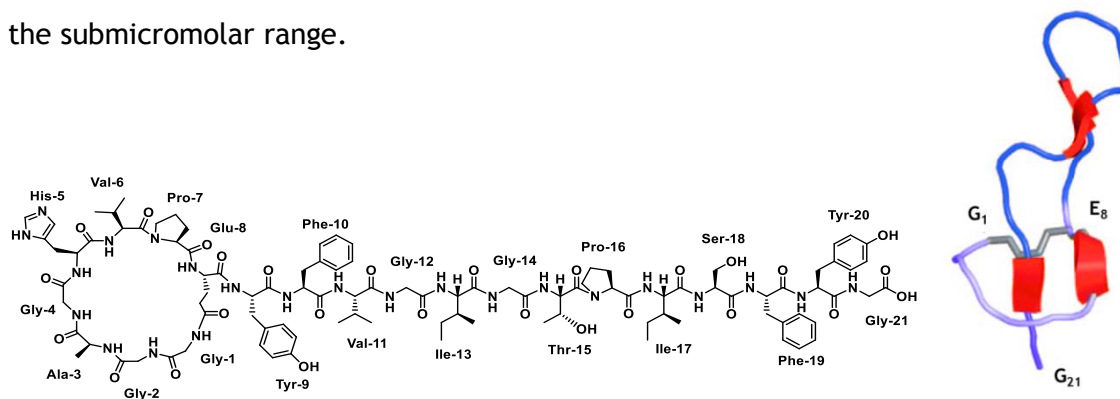


Figure 1: Chemical structure and ribbon representation of the 3D structure of MccJ25 (PDB code: 1Q71). The four fragments of antiparallel β -sheets are shown in red. (Adapted)⁵

MccJ25 gene cluster was the first biosynthetic gene cluster identified and it is formed by four genes: *Mcj-A*, *-B*, *-C* and *-D*. The last one corresponds to the ABC-transporter and provides the lasso peptide with the antimicrobial activity. This cluster (*McjABCD*) was found on a plasmid and was proven to be sufficient for MccJ25 production by heterologous expression.⁶ Regarding this, MccJ25 is the most extensively studied lasso peptide and several variants of it has been generated by mutagenesis.⁷

MccJ25 shows high thermal and proteolytic stability against several proteases such as chymotrypsin, trypsin and pepsin.⁸ Moreover, it is stable to highly denaturing conditions such as 8 M urea and temperatures above 100 °C.⁹ However, it is sensitive to thermolysin, which cleaves the peptide chain between Phe10 and Val11, resulting in a single product with a mass of + 18 Da. This product consists of two peptide chains that remain associated due to the topological trapping of the tail inside the ring.¹⁰

Mechanism of action

Although most antimicrobial peptides are characterized by a large number of positive charges and are generally considered to interact with bacterial membrane structures, MccJ25 contains only a single positive charge and has been shown to

inhibit bacterial transcription by interacting with the B' subunit of the *E. coli* RNA polymerase (RNAP), which is the target for its antibiotic action.⁹

MccJ25 also disrupts the *Salmonella newport* inner membrane, inhibiting several essential processes for cell viability such as oxygen consumption;¹¹ whereas in many *E. coli* strains, the peptide inhibits the RNA transcription⁹ without affecting the oxygen consumption. The uptake of MccJ25 by *E. coli* is dependent on the outer membrane receptor FhuA.¹² FhuA is a multifunctional protein from the outer membrane of *E. coli* which binds and transports ferrichrome-iron across the outer membrane.¹³ It also acts as a receptor for colicin M, bacteriophages T1, T5, phi80 and UC-1 and the presently studied MccJ25.¹⁴ For example, treatment of MccJ25 with thermolysin abolishes its binding to FhuA and results in no antibacterial activity, although it is still able to inhibit *E. coli* RNAP activity and *S. newport* respiration *in vitro*.¹⁵ This suggests that the intact ring and tail parts interact with the target, whereas the B-hairpin loop (hydrolyzed after the thermolysin treatment) are involved in the reception and translocation by FhuA. Additional supporting evidence that the Val11 to Pro16 residues are important for MccJ25 uptake emerged from an extensive mutational scanning analysis.⁷

In another experiment, in which a methyl group was introduced on the C-terminal Gly, it was found that the C-terminal region was critical for the RNAP inhibitory activity of MccJ25, but not for the capability to inhibit oxygen consumption nor for the peptide transportation.¹⁶ This finding suggested that the methyl ester group could either perturb its approximation to the RNAP MccJ25-binding site,¹⁷ or cancel a possible electrostatic interaction between the unique carboxyl group of MccJ25 with a positive residue of RNAP. Regardless of the mechanism, an electrostatic interaction between the C-terminus and His5 was detected, which was also involved in stabilizing the threaded structure.¹ Furthermore, mutation studies showed that Ile13, located on the loop of the lasso, was essential for the uptake of MccJ25 into cells.¹⁸

Interestingly, a second mode of action is reported for MccJ25, and this plays an important role in its activity against highly sensitive *Salmonella* and *E. coli* species overexpressing FhuA. This activity has been validated to be completely independent of RNAP inhibition. In these cases, MccJ25 appears to interact with the membrane and thereby depolarizes it and decreases oxygen consumption, although the fundamental mechanism is not yet understood.¹⁹

Chemical part

Basic treatment

The expression and purification of the lasso peptides chaxapeptin, sphingopyxin I and MccJ25 were performed as described in the ‘heterologous expression’ section of Chapter 3 (see Materials and Methods for more details). As mentioned in the introduction of this investigation, harsh basic conditions (0.5 M NaOH) were applied to the synthetic bicyclic analog A and the ester bond was hydrolyzed. Here, the same conditions were applied to three natural lasso peptides in order to evaluate if a change in their structure could also be possible. However, using the same conditions as for the ester bond hydrolysis of the bicyclic analog A, no reaction progress was detected, even after overnight exposure. When the temperature was increased to 40 °C, the appearance of new peaks was detected. As shown in Figure 2, the three lasso peptides were affected by the basic treatment after 16 h. At this point, the ratio between the starting material and the new product was 1:1 and even with the addition of more NaOH or the increase of the reaction time, no reaction progress was observed.

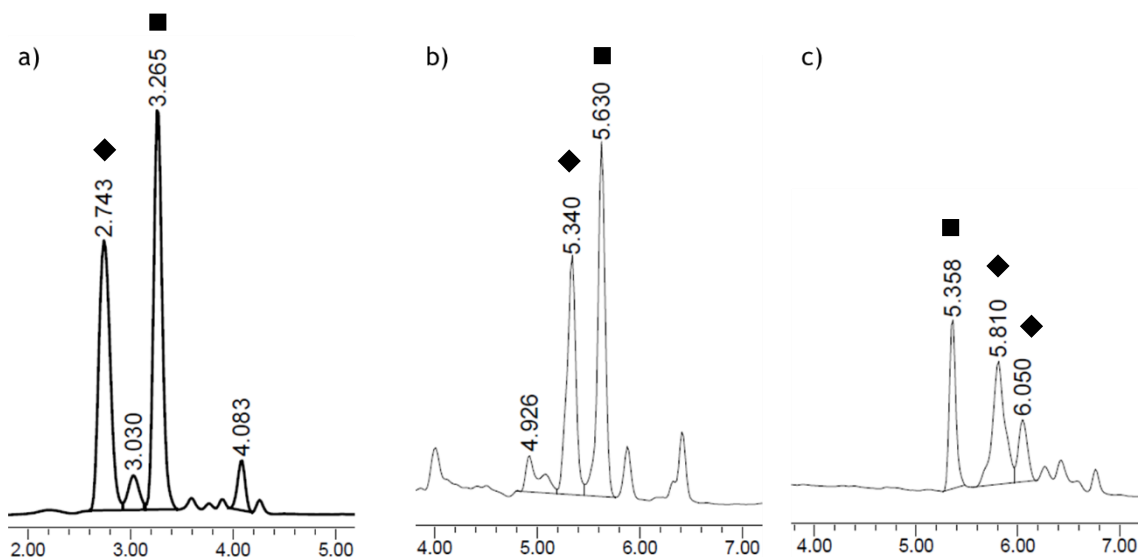


Figure 2: RP-HPLC chromatogram after the basic treatment. a) MccJ25, b) chaxapeptin and c) sphingopyxin I. Conditions: XBridge C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 60% of ACN over 8 min at 25 °C (a and b) and 25% to 50% of ACN over 8 min at 25 °C (c). ■ = starting material (SM), ◆ = new product (hydrolyzed MccJ25).

The crude of MccJ25 was the only that could be purified because more amount of sample was initially used. From the other two peptides there was not enough sample to isolate the peaks and perform further analysis and studies. In order to facilitate the names of the peptides, we called “hydrolyzed MccJ25” to the new peak and “SM MccJ25” to the starting material that remained unreacted.

As shown by this experiment, lasso peptides are not totally chemically stable. After a long treatment with harsh basic conditions, lasso peptides were converted into another product. The main objective in this investigation was the characterization of the new product obtained from the basic hydrolysis and the comparison to the native lasso peptide.

Chemical comparison

The two isolated peaks from MccJ25, together with the native MccJ25, were dissolved in H₂O-ACN (1:1) and analyzed by RP-HPLC using a C18 analytical column. Moreover, their molecular weight was characterized by ESI⁺ (Figure 3). In RP-HPLC, both peptides co-eluted and their different chemical equivalence was confirmed by the different retention time (t_R), even though they had the same molecular mass (2107.8 Da). The SM MccJ25 showed the same t_R as the native MccJ25, indicating that after the basic treatment the peptide that remained unreacted preserved the initial conformation. Regarding this chemical equivalence, the following studies will be based on the comparison of the hydrolyzed MccJ25 with the native MccJ25.

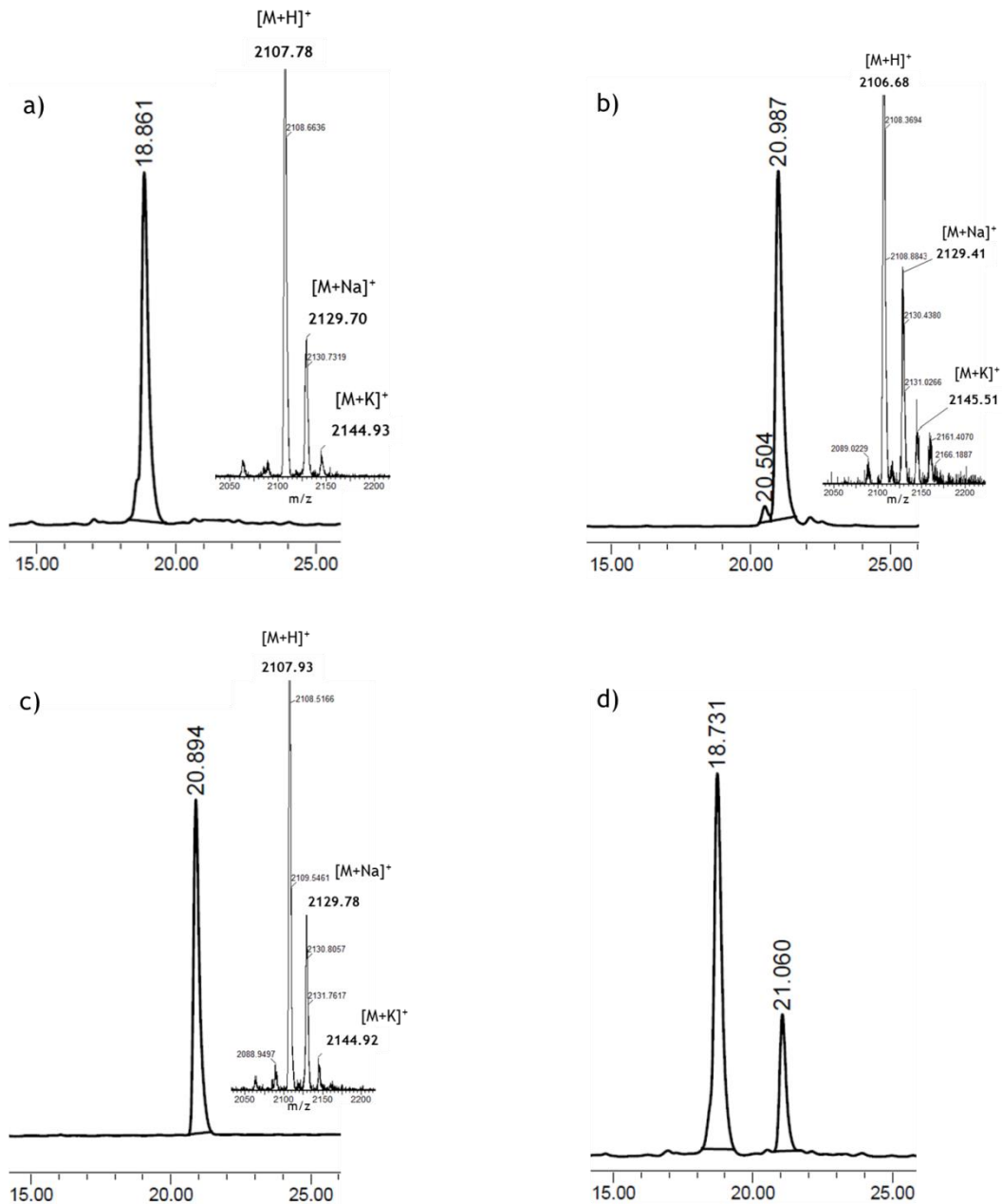


Figure 3: RP-HPLC chromatograms and ESI⁺ spectra of the: a) hydrolyzed MccJ25, b) SM MccJ25, c) native MccJ25 and d) co-injection of the hydrolyzed and SM MccJ25. Conditions: Phenomenex Luna C18 5 μ m (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 25% to 40% of ACN over 40 min at 60 °C.

Stability assays

In order to compare the two products, the thermal and proteolytic stability against carboxypeptidase Y was tested (Figure 4 and 5).

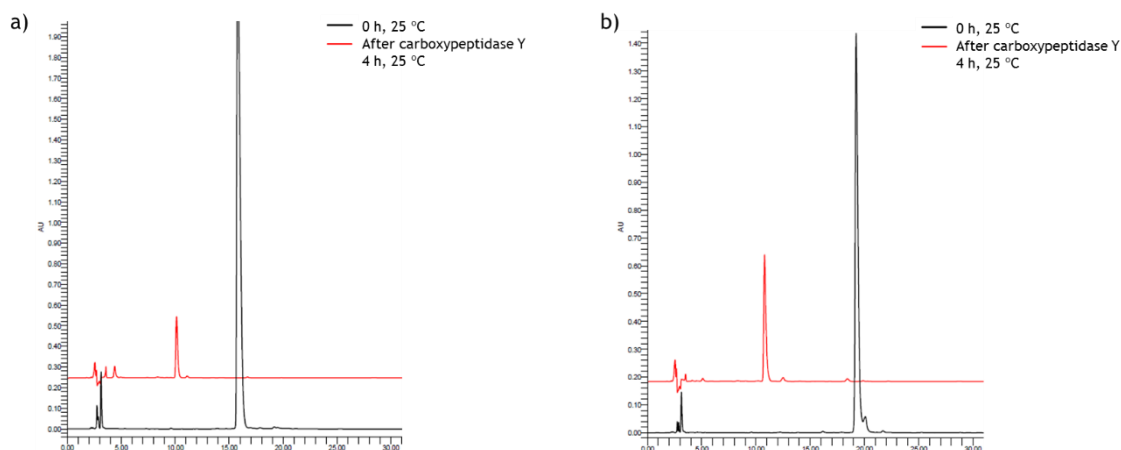


Figure 4: RP-HPLC chromatograms of the carboxypeptidase Y assay of the a) hydrolyzed MccJ25 and b) native MccJ25. Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 25% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

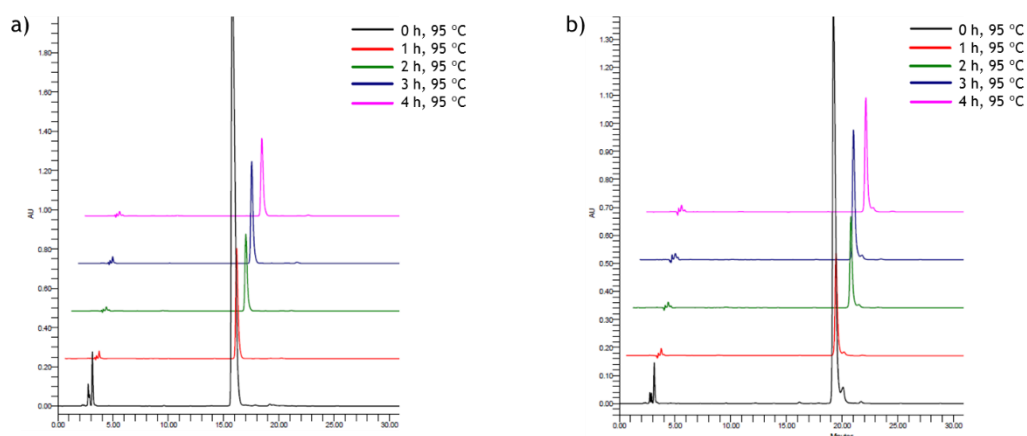


Figure 5: RP-HPLC chromatograms of the heat stability at 95 $^{\circ}\text{C}$ of the a) hydrolyzed MccJ25, b) native MccJ25. Conditions: Phenomenex Luna C18 5 μm (4.6 mm x 250 mm) reversed-phase analytical column; linear gradient from 25% to 40% of ACN over 40 min at 25 $^{\circ}\text{C}$.

Regarding the carboxypeptidase Y assay, the same cleavage pattern was detected for both peptides, corresponding to the loss of 81 Da. This loss corresponded to a double cleavage in the loop between the amide bond of Phe10 - Val11 and Val11-Gly12, resulting in the loss of Val11. The treated peptide kept the linear tail still threaded through the ring.²⁰ Despite showing the same cleavage

pattern, the t_R of the both peptides was still different: $t_R = 10.790$ min for the native MccJ25 and 10.123 min for the hydrolyzed MccJ25. Both peptides were also thermally stable for 4 h at 95 °C. The carboxypeptidase Y treatment was also performed to the resulting peptides after heating for 4 h at 95 °C, and the same cleavage pattern as in Figure 4 was observed.

Both products were subjected to proteolytic hydrolysis with strong proteases, such as pepsin and thermolysin. These two enzymes were selected because they can tolerate harsh conditions, like acidic pH and high temperatures, respectively. As expected from the pepsin treatment, no hydrolysis was observed even after several hours at pH 2 at 37 °C. For this reason, the temperature was increased to 50 °C, as pepsin is stable at this temperature. However, no cleavage was detected indicating that both peptides were totally stable (Figure 6).⁸

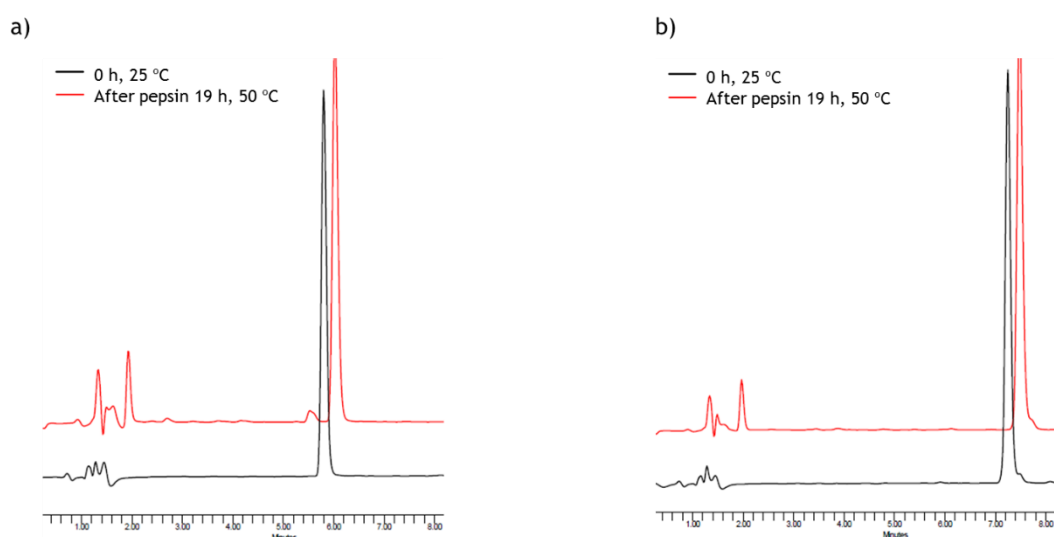


Figure 6: RP-HPLC chromatograms of the pepsin assay of the a) hydrolyzed MccJ25, b) native MccJ25. Conditions: XBridge C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 25% to 35% of ACN over 8 min at 25 °C.

On the other hand, treatment with thermolysin has been reported in several articles because it is the only enzyme able to cleave MccJ25 and maintain the threaded structure. Thermolysin hydrolyzes the Phe10-Val11 amide bond giving rise to two fragments, Gly1 - Phe10 and Val11-Gly21, with a gain of mass of 18 Da.¹⁰ The bulky aromatic side chains of Phe19 and Tyr20, positioned on each side of the macrolactam ring, play a crucial role in the stability of this peptide. The cleaved structure retains some antimicrobial activity, although the efficacy is considerably reduced when compared to the native MccJ25.²¹ This region comprises a β -sheet

between the same residues 10-11 and is directly involved in the β -hairpin. Thus, the activity was reduced due to the fact that thermolysin affects the secondary structure.

In order to study this effect, the native and the hydrolyzed MccJ25 were subjected to thermolysin cleavage using different buffers and temperatures (see Materials and Methods for more details). For both peptides, the same cleavage pattern was obtained corresponding to the loss of 137 Da (m/z 1969.9), independently of the conditions used (Figure 8). Analyzing the MccJ25 structure, the loss of 137 Da could be attributed to two types of cleavage: the double cleavage in the ring between Gly4-His5 and His5-Val6 amide bonds, with the concomitant loss of His5, or to the double cleavage between Phe10-Val11 and Gly12-Ile13, with the loss of Val11-Gly12 and the subsequent gain of 18 Da. The last cleavage leads to a structure with two fragments stabilized by the steric hindrance of Phe19 and Tyr20 and the presence of noncovalent interactions.

In a study about the elucidation of MccJ25 structure, Wilson and coworkers described the cleavage that we observed, but only when they used large amounts of thermolysin.²² They described three major degradation products (Figure 7). However, using low quantities of thermolysin they were only able to observe the cleavage of the amide bond between Phe10 and Val11, with a mass gain of + 18 Da. Considering this, we repeated the thermolysin assay again with less amount of enzyme, and the same cleavage as in Figure 8 was obtained. In our case, the thermolysin cleavage was independent of the amount of enzyme used.

Fragments	m/z
G ¹ GAGHVPEYF ¹⁰ + S ¹⁸ FYG ²¹	14887.7
G ¹ GAGHVPEYF ¹⁰	1015.5
G ¹ GAGHVPEYF ¹⁰ + I ¹³ GTPISFYG ²¹	1968.9

Figure 7: Degradation products obtained with large amounts of thermolysin.²²

An interesting observation after the treatment was that the t_R of both peptides was the same (Figure 8d). This result indicated that the cleavage led to the same structure, suggesting that the variable part was eliminated during the hydrolysis. This observation suggests that Val11-Gly12 were the residues affected during the basic

treatment. The only possibility would be that the chiral center of Val11 had suffered racemization. This fact would be in agreement with the evidence that the residues Val11 to Pro7 were important for the antimicrobial activity of MccJ15, because this region showed high interaction with RNAP of the bacteria.⁷

Thus, concerning all this information, the previous cleavage proposed around His5 could be discarded. In addition, there is no reported literature about the presence of this cleavage site. However, an alteration of this residue, which is the only positive charge of the peptide, would greatly affect the interaction with the outer membrane receptor FhuA and with RNAP. This hypothesis will be discussed in detail in further experiments.

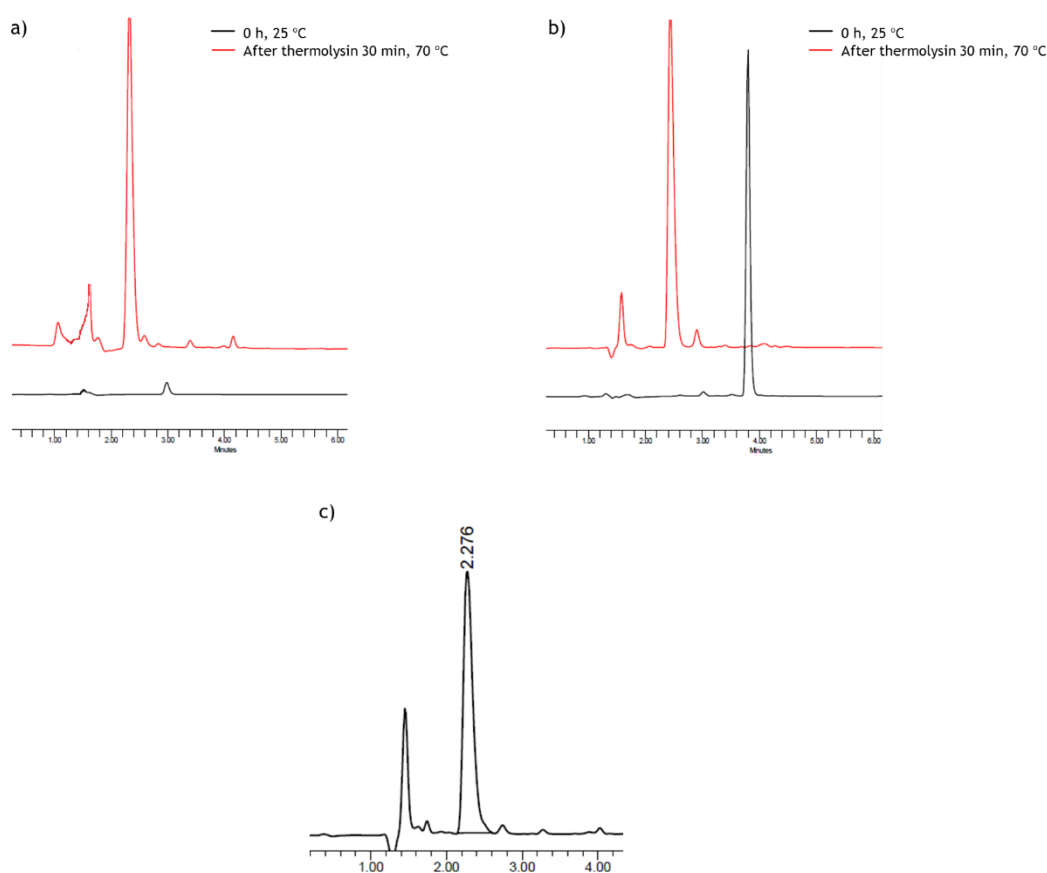


Figure 8: RP-HPLC chromatograms after 30 min of thermolysin treatment at 70 °C in 50 mM TRIS pH 8, 0.5 mM CaCl₂ buffer. a) hydrolyzed MccJ25, b) native MccJ25 and c) co-injection of the three peptides. Conditions: XBridge C18 2.5 μm (4.6 mm x 75 mm) reversed-phase analytical column; linear gradient from 30% to 60% of ACN over 8 min at 25 °C.

The thermal and proteolytic studies did not allow a proper characterization of the hydrolyzed MccJ25, as it showed exactly the same behavior as the native MccJ25. Therefore, the results pointed out that the hydrolyzed MccJ25 showed a lasso

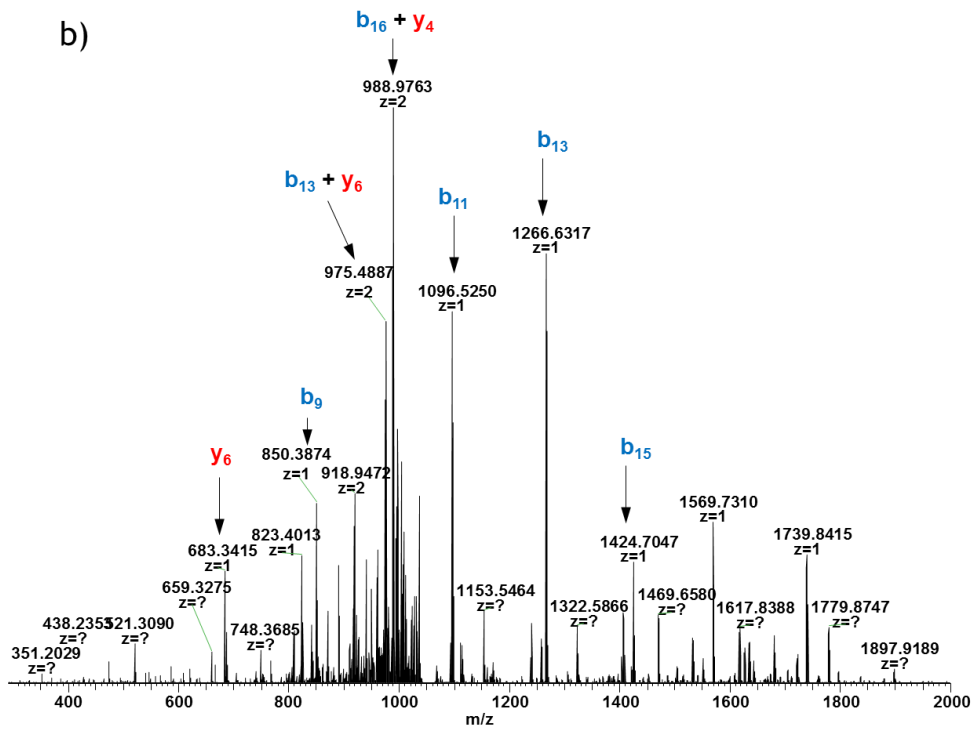
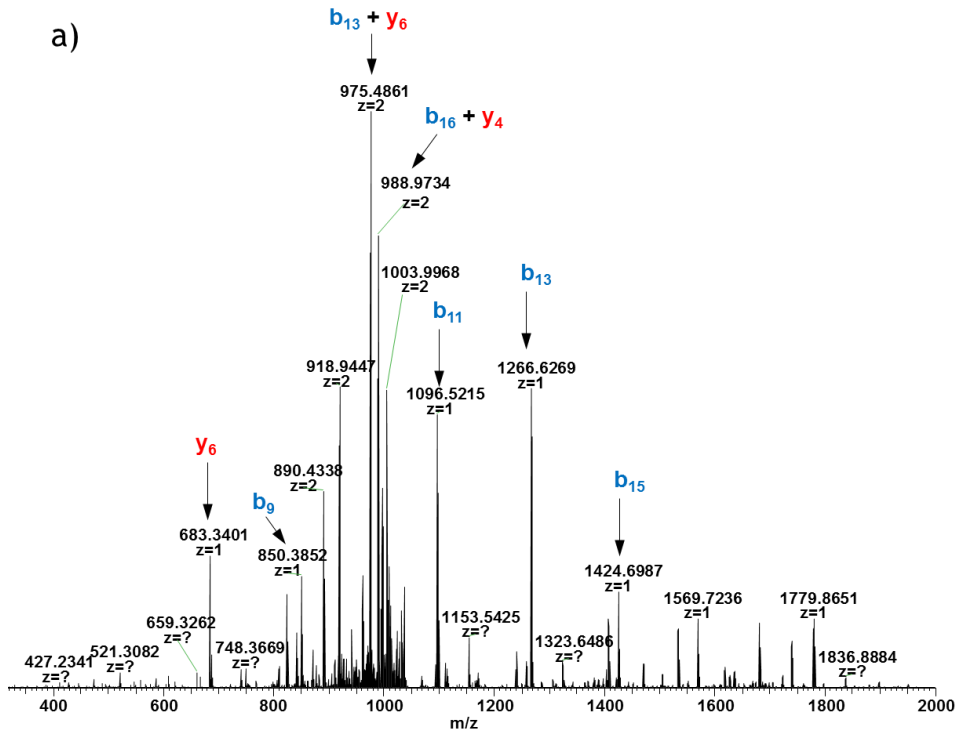
structure, due to the proteolytic stability. However, this hypothesis should be confirmed in further experiments.

The most interesting feature was that the same t_R was achieved after the thermolysin treatment, therefore suggesting that maybe they acquired the same topology. However, more experimentation was required to confirm these results (see below).

MS/MS analysis

In order to go a step further into the characterization of the hydrolyzed MccJ25, MS/MS studies were performed. This technique also allowed the comparison of the MS² fragmentation spectrum with the native MccJ25 with the SM MccJ25.

Based on the previous thermal and proteolytic stability studies, two hypotheses were considered. First, that due to the basic treatment, the tail could have been moved up or down to the ring, leading to a different topology, as it has been observed for astexins-2 and -3 after thermal treatment.²³ This topology would translate into a different fragmentation pattern. The second hypothesis was that the tail would be unthreaded from the ring, leading to a branched cyclic MccJ25. However, this last hypothesis was not clear because the previous proteolytic studies showed that the hydrolyzed product was highly stable. In contrast, the proteolytic stability of a branched cyclic analog is very low, being readily degradable.



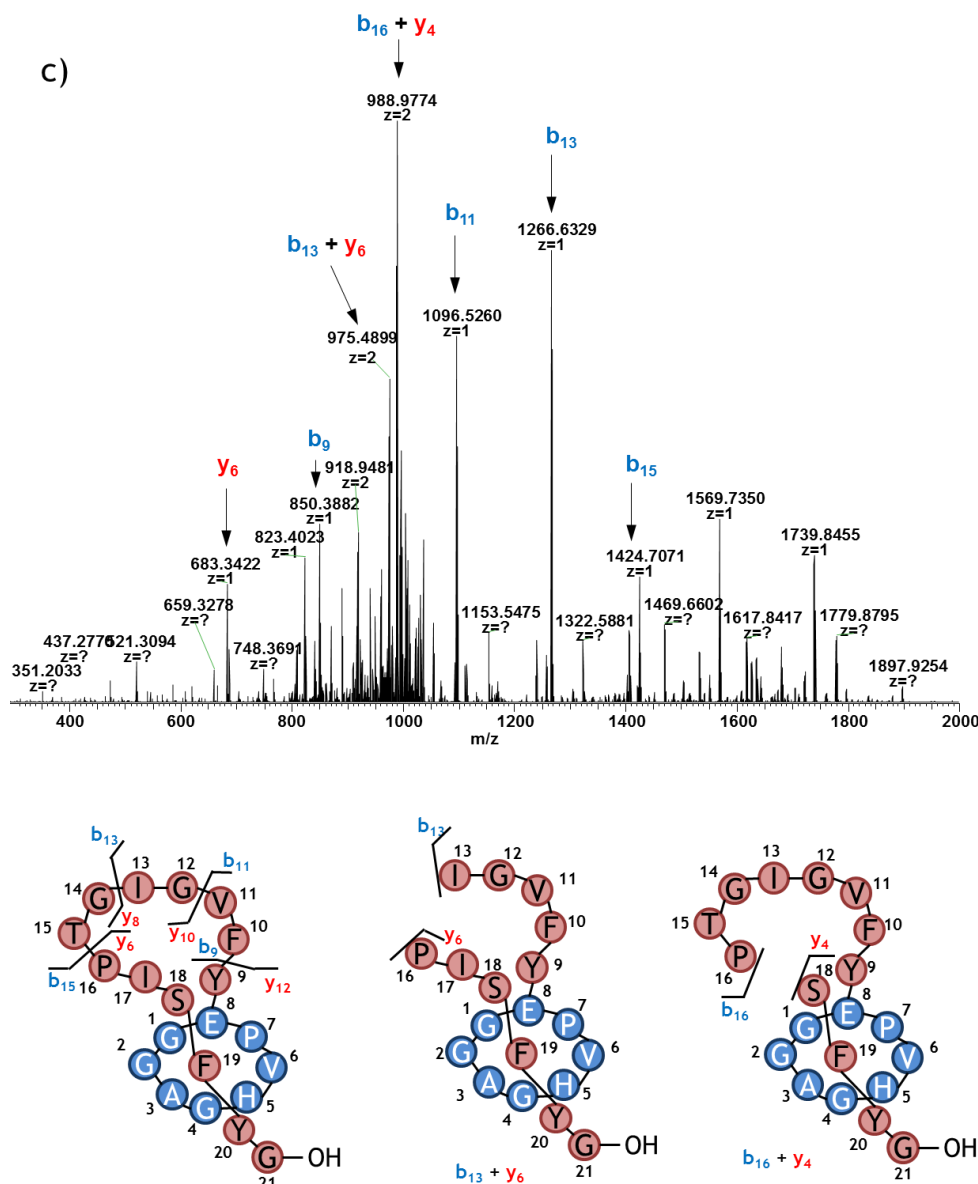


Figure 9: MS² fragmentation spectrum of the: a) hydrolyzed MccJ25, b) SM MccJ25 and c) native MccJ25. Below) Schematic representation of the fragmentations derived from the MS² analysis.

The MS² fragmentation spectra revealed that the main fragmentations took place around Pro16 (Figure 9). The same fragmentations were obtained for all the peptides, but with different intensities (Table 1). The SM and the native MccJ25 showed a very similar spectrum, being $b_{16} + y_4$ the most abundant fragmentation. On the other hand, the spectrum of the hydrolyzed MccJ25 differed from the previous two, being $b_{13} + y_6$ the most intense fragmentation, although $b_{16} + y_4$ was also present. These fragmentations were characteristic of a lasso structure because they contained both N-terminal and C-terminal parts of the molecule. The steric hindrance between the ring and the bulky side chains of a putatively threaded C-terminal part of the

peptide most likely explain this remarkable fragmentation pattern. Regarding the native MccJ25, the fragmentations obtained here are in agreement with those already described in the literature.^{1,22}

Peptide	Most intense fragmentation	Description
Hydrolyzed MccJ25	$b_{13} + y_6$	Fragmentation around N-terminus Pro16. Loss of Gly14 and Thr15
SM MccJ25	$b_{16} + y_4$	Fragmentation around C-terminus Pro16. Loss of Ile17
Native MccJ25	$b_{16} + y_4$	Fragmentation around C-terminus of Pro16. Loss of Ile17

Table 1: Summary of the main MS² fragmentations.

Due to the fact that the MS² spectrum of the SM and native spectra were identical, only the native was further used in the comparison. MS³ of the main fragments of the hydrolyzed and native MccJ25 were acquired. The MS³ spectra led to the same fragmentation pattern, the only difference being that the MS³ fragmentations of the $b_{13}+y_6$ fragment were more intense for the native MccJ25 than for the hydrolyzed product (Figure 10). There was not any remarkable fragmentation that allowed the distinction of the hydrolyzed product from the native MccJ25, only the different intensities. However, the intensity could vary depending on the sample batch. For this reason, the difference in the abundance was not a relevant trend to consider.

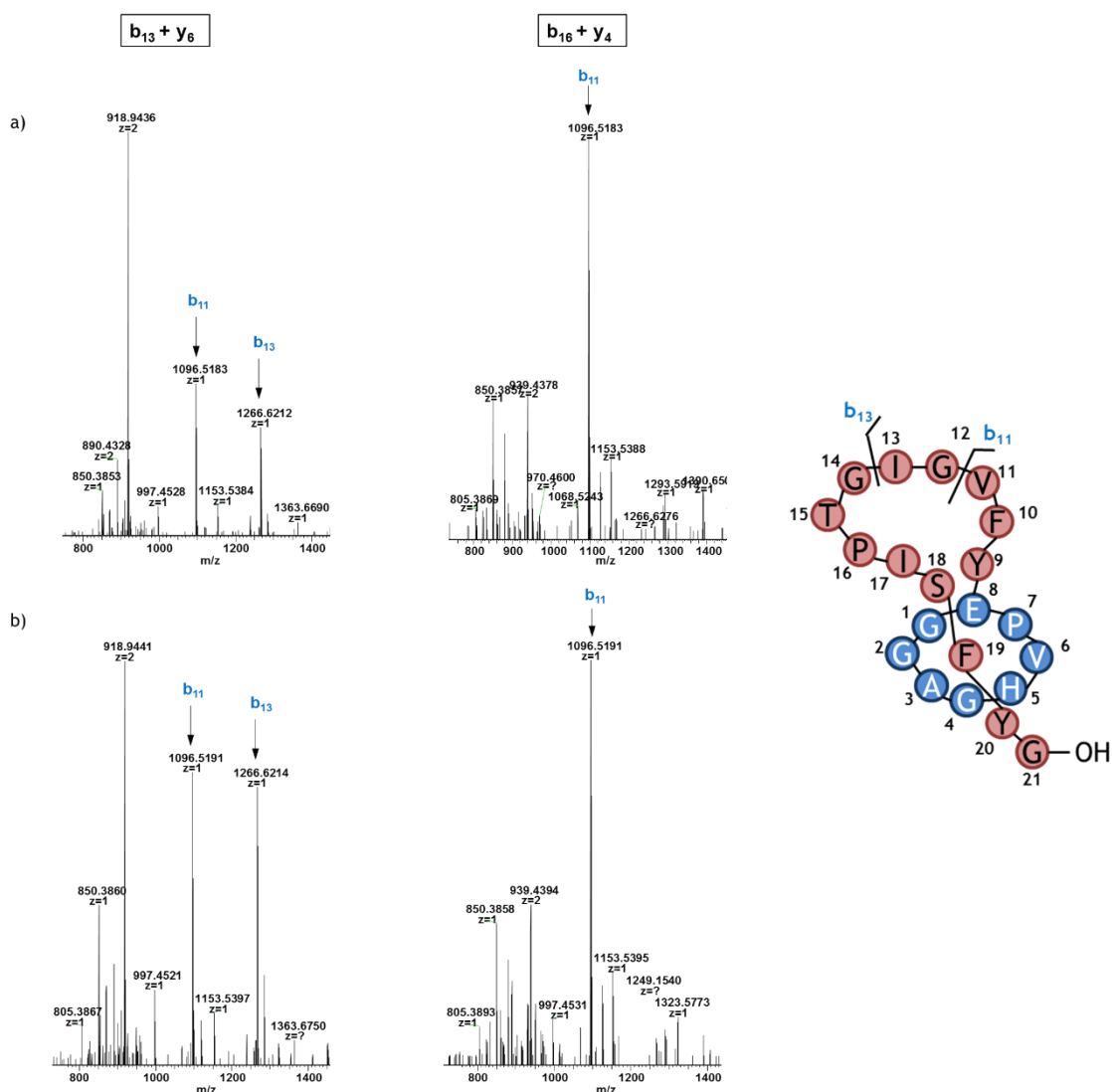


Figure 10: MS^3 spectrum of the $b_{13} + y_6$ and $b_{16} + y_4$ fragments of the: a) hydrolyzed MccJ25 and b) native MccJ25. Right) Schematic representation of the fragmentations derived from the MS^3 analysis.

Considering the MS^2 analysis, it did not provide any new information that allowed to differentiate and characterize the hydrolyzed MccJ25. From these results, we could confirm that the hydrolyzed MccJ25 showed a lasso structure and its structure was very similar or even the same as the native MccJ25, except for the different t_R that they displayed. Concerning the first hypothesis proposed, if the loop had been shortened because the tail has been moved up to the ring, the fragmentation $b_{16} + y_4$ would not be possible due to the steric hindrance caused by the proximity to the ring. On the other hand, if the loop was longer, Gly21 would act as the bottom plug and since it is not a bulky residue, after thermal treatment the peptide would unthread.

IM-MS analysis

The next technique used for the characterization of the hydrolyzed MccJ25 was IM-MS, an extensively used technique for lasso peptides characterization. In this regard, IM-MS should confirm whether the hydrolyzed MccJ25 showed or not the lasso structure. Previous IM-MS experiments demonstrated several characteristic features that allowed the distinction of a lasso peptide from other topologies.²⁴

Peptides	Charge state	no. of conformations	CCS (Å ²) (Ratio (%))
Hydrolyzed MccJ25	[M + 2H] ²⁺	4	315 (11)/370 (40)/ 396 (47) /442
	[M + 3H] ³⁺	3	361 (10)/ 382 (85) /435 (5)
	[M + 4H] ⁴⁺	2	375 (77) /409 (23)
SM MccJ25	[M + 2H] ²⁺	4	315 (14)/360 (14)/ 379 (67) /437 (4)
	[M + 3H] ³⁺	3	363 (14)/ 379 (75) /427 (11)
	[M + 4H] ⁴⁺	2	380 (77) /417 (23)
Native MccJ25 ²⁴	[M + 2H] ²⁺	3	373 (63) /383 (27)/393 (10)
	[M + 3H] ³⁺	1	383 (100)
	[M + 4H] ⁴⁺	2	404 (88) /442 (12)

Table 2: Number of conformations and CCS with 400 mM of sulfolane. Charge state [M + H]⁺ is not shown because is a very unfolded structure with the highest drift time values and is not relevant for the comparison. Each CCS was calculated based on the drift times.²⁵ When several conformations were evidenced, the major species is highlighted in bold. CCS values of native MccJ25 are obtained from previous work from Dit Fouque and coworkers.²⁴

In the three peptides, small changes in CCS values were observed upon increasing in the charge state (Table 2). This feature was demonstrated with the low range of CCS values ($\Delta\Omega$) and the concomitant low $\Delta\Omega/\Omega$ parameter that indicates the relative range of CCS covered by all protonated molecules (Table 3). Small changes in CCS values were characteristic for lasso peptides because as the charge state increases there is more unfolding and consequently, lower drift time (t_d) values. However, due to the organized and constrained structure of lasso peptides, this increase in the CCS was very low compared to branched cyclic or lineal peptides. Following this theory, the t_d of the [M+4H]⁴⁺ charge state was the lowest for all the peptides. The $\Delta\Omega/\Omega$ value of the hydrolyzed MccJ25 was within the accepted range for lasso peptides (0

to 10.9%).²⁶ So, it could be extracted that this peptide clearly showed a lasso structure.

The different t_d value between both the hydrolyzed and SM MccJ25, compared with the native MccJ25, were due to the different parameters used during the acquisition (Table 3). The t_d values cannot be compared between independent experiments because they are highly related to the equipment, the methodology and the parameters used. Only the CCS values can be compared because they are calculated based on the t_d using a calibration curve with tryptic peptides from human hemoglobin and BSA (of known CCS values).

Peptides	[M+2H] ²⁺		[M+3H] ³⁺		[M+4H] ⁴⁺		$\Delta\Omega$ (Å ²)	$\Delta\Omega/\Omega$ (%)
	CCS (Ω , Å ²)	t_d (ms)	CCS (Ω , Å ²)	t_d (ms)	CCS (Ω , Å ²)	t_d (ms)		
Hydrolyzed MccJ25	396	10.51	382	7.96	375	6.62	21	5.3
SM MccJ25	379	10.24	379	7.93	380	6.67	1	0.3
Native Mccj25	373	8.66	383	4.67	404	3.21	31	7.7

Table 3: CCS values (Ω) and drift time (t_d) of the main conformation from the ion mobility spectra. Range of CCS ($\Delta\Omega$) and $\Delta\Omega/\Omega$ parameter. CCS values of native MccJ25 are obtained from previous work from Dit Fouque and coworkers.²⁴

The number of conformations for the hydrolyzed and the SM MccJ25 was the same, however it was higher than for the reported native MccJ25 (Table 2). Regarding this difference, several questions related with the methodology and the parameters used during the acquisition were suggested: could the high number of conformations be associated with a too high voltage in the sampling cone, which would result in the detection of unthreaded structures? Or would the use of a linear Wave Height ramp (1-28 V), instead of a fixed the Wave Height, modify the number of conformations?

To answer to these questions, we performed several IM-MS spectra with SM MccJ25, which displayed the same characteristics than the native MccJ25, by modifying the sampling cone and the Wave Height values. First, sampling cone values of 15, 20, 30 and 40 V were tested and the same number of conformations for all charge states were detected. Therefore, using high sampling cone values, peptides were not being unthreaded. On the other hand, fixing the Wave Height values to 4, 5, 6, 7 and 8 V, for the charge states [M + 2H]²⁺ and [M + 3H]³⁺ resulted in the same

the number of conformations. However, for $[M + 4H]^{4+}$ two conformations were detected with Wave Height 4 to 7 V, but only one with 8 V as Wave Height, due to the fact that the two conformations appeared at the same t_d and it was not possible to separate them (data not shown).

In conclusion, the parameters and the data processing developed here were better optimized than those previously reported. As a result, we were able to fit more peaks in the experimental Gaussian curve, allowing a more accurate result and an improved detection of conformations (Table 2).

The number of multiple conformations was higher for lower charge states, which was expected due to the larger number of possible available protonation sites and the coexistence of several protomers.²⁴ The presence of more conformations at low charge states was related to width ion mobility peaks. When the charge state increases, narrow ion mobility peaks, associated with the existence of few conformations, are expected for lasso peptides (Figure 11).

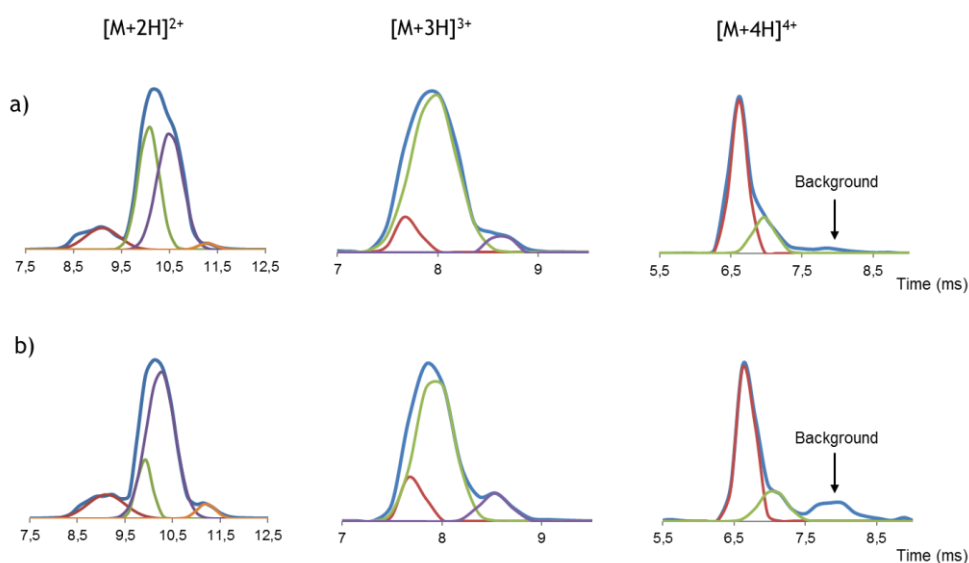


Figure 11: Drift time profiles (blue trace) and fitted peaks (red, green, purple and orange trace) of double, triply and quadruply protonated ion of a) hydrolyzed MccJ25 and b) SM MccJ25. Background is the residual drift time profile, in which no more peaks could be fitted.

By IM-MS, we could confirm once again that the hydrolyzed MccJ25 showed a lasso structure because it showed the same features as SM MccJ25, as well as the reported native MccJ25. If the lasso structure had been moved up or down to the ring, it would not be detected by IM-MS because the CCS values were very close one

to each other. To characterize the secondary structure of the hydrolyzed MccJ25, other techniques would be required.

Circular dichroism (CD) spectroscopy

In order to evaluate the secondary structure of the hydrolyzed MccJ25 and compare it with the native and SM MccJ25, CD was carried out. CD provides a rapid and accurate fingerprint of the peptide folding in solution. In the experiment, MeOH was used as solvent because the CD of native MccJ25 was already reported in this solvent.^{27,28} The spectrum of the native MccJ25 was also acquired in H₂O with 5% of 2,2,2-trifluoroethanol (TFE) to favor the folding of the peptide. Usually, the addition of TFE accentuates the secondary structure of peptides and proteins.

The CD spectra (Figure 12) of the three peptides displayed a clear minimum at 200 nm, which is characteristic of unstructured peptides, indicating a predominant random coil state. The native MccJ25 in MeOH showed a maximum at 220 nm and a smooth positive shoulder at 210 nm. These bands have been previously attributed to the Phe L_a and Tyr L_a transitions.²⁹ However, for the native MccJ25 in H₂O-TFE, the maximum was located at 222 nm and was more intense than in MeOH. Added to this, the shoulder at 210 nm was not marked. On the contrary, the SM MccJ25 presented the same CD spectrum as the native MccJ25, but in this case, the shape of the positive band was different. Both positive bands at 210 and 220 nm could be detected, but they were not as clear as for the native MccJ25.

The positive band at 220 nm for the hydrolyzed MccJ25 was displaced to 225 nm and the one at 210 nm was more defined and clear than for the other peptides. As mentioned before, these positive bands were attributed to the Phe and Tyr transitions, suggesting that the L_a transition of these residues had more contribution in the hydrolyzed MccJ25.

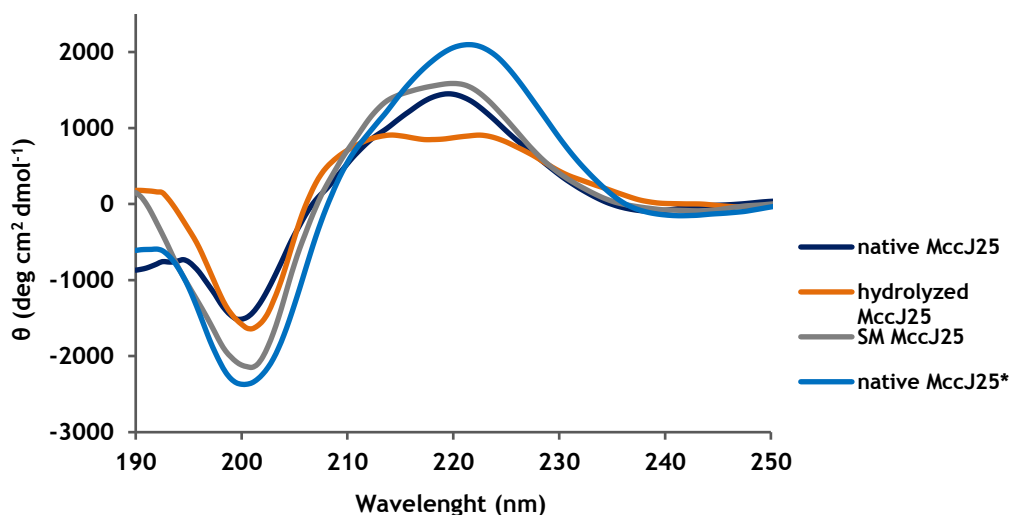


Figure 12: CD spectra of the hydrolyzed, SM and native MccJ25 in MeOH at 25 °C. *in H₂O with 5% TFE.

CD did not add new information about the characterization of the hydrolyzed MccJ25. From the CD spectra some differences in the positive bands could be observed, but the clear minimum at 200 nm indicated a random coil structure. These changes in the positive bands could suggest a different topology.

Capillary electrophoresis (CE)

A previously presented hypothesis (see Stability studies, pag. 281) was the alteration of His5. Deprotonation of His5 during the basic treatment could produce a modification of the peptide charge and a subsequent alteration in the biological activity and the 3D structure. In order to evaluate this hypothesis, the three peptides were analyzed by CE, a technique that allows the separation of charged species based on their charge and size, due to their different migration rate in an electric field.

An acidic buffer was selected to assure that the peptides were totally protonated. Moreover, 10% (v/v) ACN was added to the buffer to enhance the solubility, a percentage that has been shown to not affect to the CE.³⁰

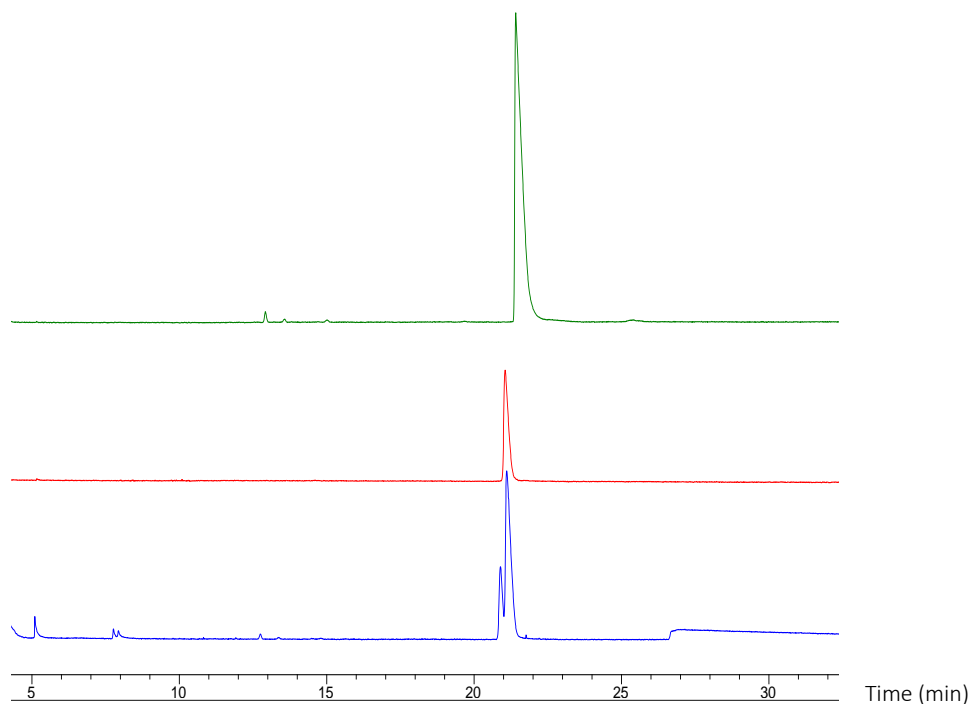


Figure 13: Overlap of the electropherograms of the hydrolyzed MccJ25 (red), native MccJ25 (green) and the co-injection of both (blue) dissolved in 10 % (v/v) ACN and 50 mM sodium phosphate pH 2.0. Conditions: CE capillary (fused silica 50 μm i.d. x 49 cm length) with 20 kV of voltage at 25 $^{\circ}\text{C}$ over 40 min.

As shown in Figure 13, both peptides had very similar t_R , but when a co-elution experiment was done two peaks appeared, indicating that they did not have the same charge. The first peak corresponded to the hydrolyzed MccJ25 ($t_R = 20.889$ min), which revealed a higher content of negative charges compared to the native MccJ25 ($t_R = 21.104$ min).

The presence of more negative charges could be associated with deprotonation of His5 and subsequent structure modification. The electrostatic interaction between His5 and the C-terminus of the peptide, which participated in the structure stabilization, may have disappeared. Otherwise, a new interaction could have been formed during the basic treatment due to His5 deprotonation, giving rise to a new stable structure. Possibly, a strong hydrogen bond interaction between His5 - Ser18 could be formed due to their proximity. In some proteins the effect of the pH on His has been already observed. Changes in the pH produced different conformations with the most stable being achieved at high pH.³¹

This hypothesis would be in agreement with the previous thermolysin cleavage, in which the hydrolysis produced a loss of 137 Da, which could be related to His5. If the

new conformation was due to a new hydrogen bond interaction with the deprotonated His5, its loss could lead to the same conformation for both the hydrolyzed and the native MccJ25. On the other hand, the other cleavage between Phe10-Val11 and Gly12-Ile13 would not be in agreement with the CE results.

The CD spectra did not provide new information about differences in the secondary structure of the hydrolyzed MccJ25. All data pointed out to the fact that the hydrolyzed MccJ25 showed the same secondary structure as the native MccJ25. The same positive and negative bands were observed, but the maximums were more defined for the hydrolyzed MccJ25 than for the native, suggesting some alteration.

The hypothesis of the movement of the lasso structure up or down to the ring only could be evaluated by proteolytic assays, MS/MS analysis or NMR. By proteolytic and thermal assays, the same behavior was observed. Thus, independently of their initial conformation, the cleavages took place between the same residues. This result meant that the enzyme accessibility was the same for the hydrolyzed and the native MccJ25. The same occurred in the MS² analysis, in which the same fragmentation pattern was observed for both. If the lasso structure had moved, a different fragmentation should be detected because the loop would be larger or shorter and this difference would allow a different fragmentation.

The NMR analysis has not been performed yet, but it is scheduled for the near future. This study will determine the structure of the hydrolyzed MccJ25 and the comparison to the native MccJ25 will allow the identification of the modified residue or residues.

Biological part

After the exhaustive chemical characterization of the hydrolyzed MccJ25, its antimicrobial activity was evaluated and compared to the SM and the native MccJ25. This was done by standard microbiological methodologies as indicated in the experimental section. The experiments of this section have been performed by Marta Jorba, a PhD student from the group of Prof. Miquel Viñas.

Minimum inhibitory concentration (MIC)

The results showed similar MIC values for the SM MccJ25 and the native MccJ25, in almost all the tested strains (Table 4). These two peptides displayed activity against four Gram-negative strains: two *E. coli* and two *S. enterica* strains.

Strains	MIC ($\mu\text{g/mL}$)		
	SM MccJ25	Hydrolyzed MccJ25	Native MccJ25
<i>Escherichia coli</i> MDR 39255	0.5-1	>256	0.5
<i>Escherichia coli</i> MDR 208691	>128	>128	>128
<i>Escherichia coli</i> MDR 246415	0.5-1	>128	0.0625
<i>Escherichia coli</i> MDR 239910	>128	>128	>128
<i>Salmonella enterica</i> ATCC 14028	>128	>128	>128
<i>Salmonella enterica</i> ATCC 13076	<0.0625	128	<0.0625
<i>Salmonella enterica</i> ATCC 49214	0.8	>128	n.d
<i>Salmonella typhimurium</i> SY5015	>128	>128	>128

Table 4: Minimum Inhibitory Concentration (MIC) of SM MccJ25, hydrolyzed MccJ25 and native MccJ25 against eight Gram-negative strains. The sensitive strains to SM and native MccJ25 are highlighted in bold. n.d.: no determined.

The hydrolyzed MccJ25 showed the lowest activity against the eight strains, compared to the SM MccJ25 and the native MccJ25. Its MIC values were in almost all cases higher than 128 $\mu\text{g/mL}$. The diminished activity of the hydrolyzed MccJ25 may be caused by changes in its conformation as an undesired result of the basic treatment. Thus, we tried to find out if the loss of activity was due to the inability of the hydrolyzed MccJ25 to penetrate the target bacteria or if, alternatively, the loss of activity was due to changes in its ability to inhibit the RNAP.

Synergy study

Colistin, is a peptide of bacterial origin able to disrupt the outer membrane structure and subsequently, the stability of bacteria. In other words, low concentrations of colistin seriously disorganize the outer membrane and, therefore, chemicals can easily penetrate into Gram-negative bacteria. Therefore, when the resistance mechanism is due to restrictions in the entry of the antibiotic, colistin and antimicrobials have a synergistic mode of action. In this study, the synergy of colistin and the bacteriocins was explored.

A checkerboard assay using hydrolyzed MccJ25 and colistin, in combination, was performed against the *E. coli* MDR 39255 strain and the *Salmonella* ATCC 13076 strain, which are sensitive to SM and native but not to the hydrolyzed MccJ25, as shown in the MIC assay (Table 4).

To quantitatively determine the interaction between the hydrolyzed MccJ25 and colistin, the Fractional Inhibitory Concentration (FIC) values were calculated. These indexes were calculated according to the following formula: FIC of drug A (FIC A) = (MIC of drug A in combination)/(MIC of A); FIC of drug B (FIC B) = (MIC of drug B in combination)/(MIC of B), see Material and Methods for more details. The FIC index (FIC_i) is calculated by adding FIC A and FIC B.

The FIC_i values of the bacterial strains tested with combinations of the hydrolyzed MccJ25 and colistin are shown in Table 5. According to the international standard, FIC_i values below 0.5 should be interpreted as synergistic.³² Thus, these experiments pointed to the existence of a synergistic effect of the two antimicrobials tested, colistin and the bacteriocin, hydrolyzed MccJ25.

Strains	FIC _i
<i>E. coli</i> MDR 39255	0.40
<i>S. enterica</i> ATCC 13076	0.41

Table 5: Fractional Inhibitory Concentration index (FIC_i) values of the hydrolyzed MccJ25 in combination with colistin.

These results clearly show that the hydrolyzed MccJ25 retains the ability to inhibit the growth of these two strains, although probably it is not able to enter to the

bacteria through the outer membrane because is unable to bind to the receptor FhuA. However, in combination with colistin, the hydrolyzed MccJ25 can penetrate, showing antibacterial activity because colistin interacts with the bacterial outer membrane changing its permeability.

Discussion

The microbiological results showed that the hydrolyzed MccJ25 has lost the antimicrobial activity because it was unable to penetrate the outer membrane. However, when it was used in combination with colistin, which facilitated the entrance of the hydrolyzed MccJ25, it showed antimicrobial activity. These results were in agreement with the chemical characterization, in which a change in the topology was observed, but the lasso structure was still present. As previously discussed, a movement of the tail up or down to the ring was discarded because a different fragmentation pattern with respect to the native MccJ25 would be observed. Thus, the only possibility that could justify the loss of the activity would be a change in the His5 or Val11-Gly12 region. The NMR results will confirm or discard these hypotheses.

Despite the microbiological results are relatively preliminary, we can speculate that the combination of bacteriocins with “door openers”, such as colistin or other antimicrobial peptides, able to facilitate the antimicrobial permeation, may be a promising strategy to fight multidrug resistant infections, whose treatment is becoming more and more unsuccessful, and has thus emerged as a central challenge of current microbiological research.³³

Conclusions

- The lasso structure of the hydrolyzed MccJ25 was confirmed by three independent properties: a high proteolytic stability and identical MS² fragmentation pattern and ion-mobility (IM) spectra as the native MccJ25.
- The different conformation was not related to a movement of the tail up or down the ring. An alteration of the His5 residue is supported by the results obtained in capillary electrophoresis (CE) experiments. However, the reported literature pointed to a modification in the Val11-Gly12 region, which was previously demonstrated to be important for the interaction with the bacteria.⁷
- The hydrolyzed MccJ25 lost its antimicrobial activity because it is unable to cross the bacterial outer membrane.

References

- (1) Rosengren, K. J.; Clark, R. J.; Daly, N. L.; Göransson, U.; Jones, A.; Craik, D. *J. J. Am. Chem. Soc.* **2003**, *125* (41), 12464.
- (2) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. *Angew. Chemie - Int. Ed.* **2011**, *50* (37), 8714.
- (3) Hegemann, J. D.; De Simone, M.; Zimmermann, M.; Knappe, T. A.; Xie, X.; Di Leva, F. S.; Marinelli, L.; Novellino, E.; Zahler, S.; Kessler, H.; Marahiel, M. A. *J. Med. Chem.* **2014**, *57* (13), 5829.
- (4) Salomón, R. a; Farías, R. N. *J. Bacteriol.* **1992**, *174* (22), 7428.
- (5) Knappe, T. a.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. a. *J. Am. Chem. Soc.* **2008**, *130* (17), 11446.
- (6) Solbiati, J. O.; Ciaccio, M.; Farías, R. N.; José, E.; Moreno, F.; Salomón, R. a; Solbiati, O.; Fari, R. N. *J. Bacteriol.* **1999**, *181* (8), 2659.
- (7) Pavlova, O.; Mukhopadhyay, J.; Sineva, E.; Ebright, R. H.; Severinov, K. *J. Biol. Chem.* **2008**, *283* (37), 25589.
- (8) Blond, A.; Péduzzi, J.; Goulard, C.; Chiuchiolo, M. J.; Barthélémy, M.; Prigent, Y.; Salomón, R. A.; Farías, R. N.; Moreno, F.; Rebuffat, S. *Eur. J. Biochem.* **1999**, *259* (3), 747.
- (9) Delgado, M. A.; Rintoul, M. R.; Farias, R. N.; Salomon, R. A. *J. Bacteriol.* **2001**, *183* (15), 4543.
- (10) Rosengren, K. J.; Blond, A.; Afonso, C.; Tabet, J. C.; Rebuffat, S.; Craik, D. J. *Biochemistry* **2004**, *43* (16), 4696.
- (11) Rintoul, M. R.; De Arcuri, B. F.; Salomón, R. A.; Farías, R. N.; Morero, R. D. *FEMS Microbiol. Lett.* **2001**, *204* (2), 265.
- (12) Salomon, R. A.; Farias, R. N. *J. Bacteriol.* **1993**, *175* (23), 7741.
- (13) Bonhivers, M.; Plançon, L.; Ghazi, A.; Boulanger, P.; Le Maire, M.; Lambert, O.; Rigaud, J. L.; Letellier, L. *Biochimie* **1998**, *80* (5-6), 363.
- (14) Destoumieux-Garzón, D.; Duquesne, S.; Peduzzi, J.; Goulard, C.; Desmadril, M.; Letellier, L.; Rebuffat, S.; Boulanger, P. *Biochem. J.* **2005**, *389* (3), 869.
- (15) Bellomio, A.; Vincent, P. A.; De Arcuri, B. F.; Salomón, R. A.; Morero, R. D.; Farías, R. N. *Biochem. Biophys. Res. Commun.* **2004**, *325* (4), 1454.
- (16) Vincent, P. A.; Bellomio, A.; De Arcuri, B. F.; Farías, R. N.; Morero, R. D. *Biochem. Biophys. Res. Commun.* **2005**, *331* (2), 549.
- (17) Mukhopadhyay, J.; Sineva, E.; Knight, J.; Levy, R. M.; Ebright, R. H. *Mol. Cell*

- 2004, 14 (6), 739.
- (18) Socias, S. B.; Severinov, K.; Salomon, R. A. *FEMS Microbiol. Lett.* **2009**, 301 (1), 124.
- (19) Bellomio, A.; Vincent, P. A.; De Arcuri, B. F.; Farias, R. N.; Morero, R. D. *J. Bacteriol.* **2007**, 189 (11), 4180.
- (20) Ducasse, R.; Yan, K. P.; Goulard, C.; Blond, A.; Li, Y.; Lescop, E.; Guittet, E.; Rebuffat, S.; Zirah, S. *ChemBioChem* **2012**, 13 (3), 371.
- (21) Blond, A.; Cheminant, M.; Destoumieux-Garzón, D.; Ségalas-Milazzo, I.; Peduzzi, J.; Goulard, C.; Rebuffat, S. *Eur. J. Biochem.* **2002**, 269 (24), 6212.
- (22) Wilson, K.-A. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A. *J. Am. Chem. Soc.* **2003**, 125 (41), 12475.
- (23) Allen, C. D.; Chen, M. Y.; Trick, A. Y.; Le, D. T.; Ferguson, A. L.; Link, A. J. *ACS Chem. Biol.* **2016**, 11 (11), 3043.
- (24) Jeanne Dit Fouque, K.; Afonso, C.; Zirah, S.; Hegemann, J. D.; Zimmermann, M.; Marahiel, M. A.; Rebuffat, S.; Lavanant, H. *Anal. Chem.* **2015**, 87 (2), 1166.
- (25) Smith, D.; Knapman, T.; Campuzano, I.; Malham, R.; Berryman, J.; Radford, S.; Ashcroft, A. *Eur. J. Mass Spectrom.* **2009**, 15 (5), 113.
- (26) Fouque, K. J. D.; Lavanant, H.; Zirah, S.; Hegemann, J. D.; Zimmermann, M.; Marahiel, M. A.; Rebuffat, S.; Afonso, C. *J. Am. Soc. Mass Spectrom.* **2017**, 28 (2), 315.
- (27) Blond, A.; Cheminant, M.; Ségalas-Milazzo, I.; Péduzzi, J.; Barthélémy, M.; Goulard, C.; Salomón, R.; Moreno, F.; Farías, R.; Rebuffat, S. *Eur. J. Biochem.* **2001**, 268 (7), 2124.
- (28) Soudy, R.; Wang, L.; Kaur, K. *Bioorganic Med. Chem.* **2012**, 20 (5), 1794.
- (29) Woody, R. W. *Biopolymers* **1978**, 17 (6), 1451.
- (30) Kulsing, C.; Boysen, R. I.; Hearn, M. T. W. *Anal. Chem.* **2016**, 88 (24), 12255.
- (31) Valéry, C.; Deville-Foillard, S.; Lefebvre, C.; Taberner, N.; Legrand, P.; Meneau, F.; Meriadec, C.; Delvaux, C.; Bizien, T.; Kasotakis, E.; Lopez-Iglesias, C.; Gall, A.; Bressanelli, S.; Le Du, M.-H.; Paternostre, M.; Artzner, F. *Nat. Commun.* **2015**, 6, 7771.
- (32) Rudilla, H.; Fusté, E.; Cajal, Y.; Rabanal, F.; Vinuesa, T.; Viñas, M. *Molecules* **2016**, 21 (9), 1.
- (33) Sierra, J. M.; Fusté, E.; Rabanal, F.; Vinuesa, T.; Viñas, M. *Expert Opin. Biol. Ther.* **2017**, 17 (6), 663.

