DNA RECOGNITION AND ANTIMICROBIAL COMPOUNDS: EXPLORING THE VERSATILITY OF PEPTIDES

ABHISHEK IYER

Promoters: Prof. Dr. Annemieke Madder

Dr. Ishwar Singh

Dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

DOCTOR OF SCIENCE: CHEMISTRY / PHARMACY

2016

STATEMENT OF ORIGINALITY

“I, Abhishek Iyer, hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been published or accepted for the award of any other degree or diploma at University of Lincoln or any other educational institution, except where references have been made in the thesis. Any contribution made to the research by others, with whom I have worked at the University of Lincoln or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.”

Signed …….....

Date …….....

APPRECIATION AND ACKNOWLEDGEMENTS

After having written nearly three hundred pages as part of this PhD thesis, I am at a loss for words when it comes to writing this section. I am extremely grateful and would like to thank every single person I have met along the way for making this, by far, the best four years of my life. It is not possible for me to mention everyone by name, but I would like to mention the names of those people without whom this PhD would not have been possible.

First and foremost, I would like to thank my promoters Prof. Dr. Annemieke Madder and Dr. Ishwar Singh. I consider myself very fortunate for having the opportunity to work with two remarkable, very knowledgeable and unique individuals. I would like to thank Annemieke for her unwavering support and for always having faith in me even during those times when I lost confidence in myself. She has taught me, by example, that being a good human being is as important as being a good chemist and it is something I will remember for the rest of my life. I would like to thank Ishwar for his out-of-the-box thinking and ideas that he has shared with me. It has been wonderful discussing chemistry with Ishwar and I am grateful to him for always taking time out and answering patiently all the queries I have had in the last two years.

Among my colleagues and friends in Ghent I would first like to thank Yara Ruiz for the countless discussions we have had in the last four years – related and unrelated to chemistry. She has taught me a lot in Ghent as well as in Lincoln but also shown me that there is more to a PhD than just academics. I am very fortunate to have worked with Yara and am even more happy to have her as a very good friend. I am proud of my former master students – Lars, Dorien and Cecilia (for a short while) all of whom are doing PhDs. I would like to wish them all the best in their endeavours. Dorien, I would like to thank in particular for all the work that you have done for me for the articles as well as preparing and submitting countless mass samples on my behalf in the last two years. I would like to thank everyone in Ghent – Matthias, Lidia, Duchan, Vicky, Bram, Nathalie, Eri, Eva, Margarida, Smita, An, Kurt, Lieselot, Ellen, Willem, Maxime and Emily for the atmosphere in the lab. Last, but not the least, I would like to thank Jos for taking care of all the equipment and consumables and without whom the lab is simply incomplete. I would like to thank Prof. Bruno De Geest for his valuable advice and help for testing the peptides for cell uptake. I would also like to thank Maji for his help from time to time and Tim for the NMR measurements. A big thanks to Jan for running countless mass samples for four long years and for advice whenever I have needed it. Thanks to Tom and Veerle for all the administrative tasks.

I am fortunate to have been part of the PhosChemRec funded by a Marie Curie Initial Training Network. I would like thank the people from the network – Marta, Plamena, Alice, Zeyed, Martina, Charlotte, Emmanuel, Chris, Sylvain, Christina, Ute, Daniela, Luigi and Pierre-Yves for the time we spent in the network meetings – sharing chemistry as well as personal experiences.

From my friends in Lincoln, I would first like to thank Anish for the wonderful time we have shared in the past year and a half and all the light moments amidst tons of work and late night hours – which I will forever cherish. I would also like to thank Shreesha for helping me out initially, Paolo, Ranga and Louis for his witty remarks and everyone for maintaining a great atmosphere in the lab. Special thanks to Dr. Stephen Prior for some fantastic NMR work and Charlie for testing more than two hundred samples for biological activity. Without your contributions, this PhD would be incomplete. I would also

like to thank Dr. Timea Palmai-Pallag for testing the compounds for toxicity and Dr. Edward Taylor for advice regarding my project in Lincoln.

I would like to thank the members of the jury – Dr. Enrico Ferrari, Dr. Tobias Gruber, Dr. M. Frances Heaney and Prof. Dr. José Martins for taking out valuable time to read, correct and be a part of my jury.

I would like to thank my non-lab friends in Lincoln – Rashmi firstly for always being there for me and taking care of me. Mihir, Nivedita, Toral, Fawad, Khyati and Ramola we have spent some good time together and it has made my stay in Lincoln truly wonderful.

Last, but not the least, I would like to thank my family, particularly my mother for always wanting me to study hard and believing that I was capable of things even I did not think I could do at the time. My father for his patience, advice and values and my sister for being the joy of the house even when all of us were going through a tough time.

My PhD has been a long and tiring but amazing journey. I've learnt not just a lot of Chemistry but have grown as a person. It has broadened my perspective of life and given me a great platform and opportunity to contribute to science and help others less fortunate than me – something I will always keep in mind.

Abhishek Iyer

Lincoln, 15th August 2016

LIST OF ABBREVIATIONS

AA	amino acid
Aba	4-acetamidobenzoic acid
A. baumannii	*Acinetobacter baumannii*
Ac_2O	acetic anhydride
AcOH	acetic acid
Ahx	6-aminohexanoic acid
Alloc	allyloxycarbonyl (protecting group)
Boc	tert-butoxycarbonyl (protecting group)
b-HLH-ZIP	basic helix-loop-helix zipper
bZIP	basic region leucine zipper
CD	cyclodextrin
CuAAC	copper azide-alkyne cycloaddition
Dab	diaminobutyric acid
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
DIC	diisopropylcarbodiimide
DIPEA/DIEA	diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	*N*,*N*-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DPC	DNA-peptide/protein/polyamide cross-link
dsDNA	double stranded DNA
E. coli	*Escherichia coli*
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (hydrochloride salt)
EDT	ethanedithiol
EMSA	electrophoretic mobility shift assay
ESI	Electrospray ionization
EtOAc	ethyl acetate

EM	exact mass
Et_2O	Diethyl ether
Fmoc	9-fluorenylmethoxycarbonyl
Gaba	γ-aminobutylbenzoic acid
HATU	N-[(dimethylamino-1H-1,2,3-triazolo[4,5-b]pyridin-1-ylmethylene]-N-methylmethanaminium hexafluorophosphate N-oxide
HBTU	N-[(1H-benzotrioazole-1-yl)(dimethylamino) methylene]-N-methylmethanaminium hexafluorophosphate N-oxide
HCA	α-cyano-4-hydroxycinnaminic acid
hν	UV-light
HOAt	1-hydroxy-7-azabenztriazole
HOBt	1-hydroxybenzotriazole
HPLC	High-performance liquid chromatography
HRMS	high resolution MS
ICL	interstrand cross-link
K. pneumoniae	*Klebsiella pneumoniae*
LC	Liquid Chromatography
MALDI	matrix assisted laser desorption ionisation
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
MBC	Minimum Bactericidal Concentration
MeCN	acetonitrile
MeOH	methanol
MIC	Minimum Inhibitory Concentration
MoeA	Moenomycin A
mRNA	messenger RNA
MRSA	Methicillin resistant *Staphylococcus aureus*
MS	Mass Spectrometry
MW	molecular weight or microwave
mQ/ milliQ,	deionized H_2O
m/z	mass to charge ratio
NBS	*N*-Bromosuccinimide
NMP	*N*-Methyl-2-pyrrolidine

NMR	nuclear magnetic resonance
Oligo	oligonucleotide
Oxyma	Ethyl (hydroxyimino)cyanoacetate
PAGE	polyacrylamide gel electrophoresis
P. aeruginosa	*Pseudomonas aeruginosa*
Pbf	2,2,4,6,7-Pentamethyl-dihydrobenzofurane-5-sulfonyl
PBS	Phosphate-buffered saline
PDB	protein data bank
PEG	polyethyleneglycol
Pet ether	Petroleum ether
ppm	parts per million
PS	polystyrene
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
Rf	ratio to front
rt	room temperature
r.t.	retention time
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RP	reversed phase
S. aureus	*Staphylococcus aureus*
SPPS	Solid Phase Peptide Synthesis
ssDNA	single stranded DNA
TBE	TrisBoratEDTA (buffer) or tris(hydroxymethyl aminomethane, $B(OH)_3$, ethylenediamine tetraacetic acid
TCEP	tris(2-carboxyethyl)phosphine
TEA	triethylamine
TEAA	tetraethylammonium acetate
tBu	tert-Butyl
TF	transcription factor
TFA	trifluoroacetic acid
TIS	triisopropylsilane

Tris	Trizma hydrochloride buffer solution
tRNA	transfer RNA
Tris	tris(hydroxymethyl)aminomethane)
Trt	trityl or triphenylmethyl
UV	Ultraviolet
XL	cross-link

STRUCTURES OF AMINO ACIDS

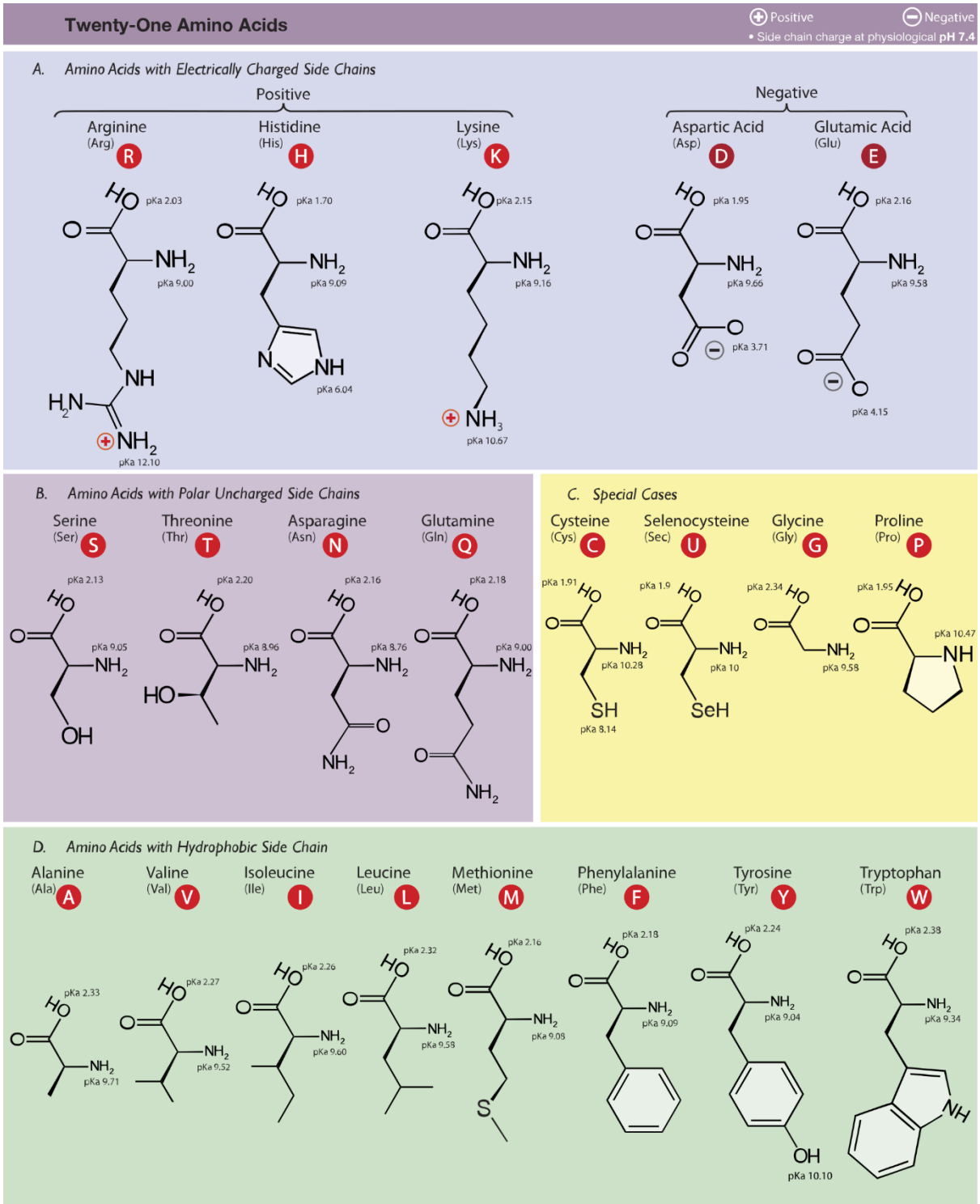


Image source: https://upload.wikimedia.org/wikipedia/commons/thumb/a/a9/Amino_Acids.svg/2000px-Amino_Acids.svg.pn

MESSAGE TO THE READER

This thesis contains the work done as part of my PhD over the last four years (Sept. 2012-Aug. 2016). It was a joint PhD, which allowed me to work two years in Ghent University in Belgium and the remaining two years in University of Lincoln in the UK.

Since there is a large amount of data in this thesis, for the purpose of convenience, I have divided the entire thesis and the experimental section, including the references and numbering of compounds chapter-wise. Research is a never-ending process and there is always a lot left to do. Therefore, I have done my level best to include as much detail and as many results as possible into this thesis while hoping to maintain a certain level of conciseness.

This thesis initially describes the quest for sequence selective DNA followed by the search for new antimicrobial compounds. The first four chapters are based on DNA recognition with new cancer therapeutics being the long term goal, while the remaining three focus on finding new drugs for bacteria. No thesis is ever perfect, but I have done my best to keep out errors, make the language lucid and explain the concepts in the simplest possible terms so as to be understood by a wide variety of audience.

Before you begin, just keep in mind that this thesis is a culmination of months of hard work, dedication and support not just by me but from my supervisors, colleagues, technicians and friends. Happy reading!

ABHISHEK IYER

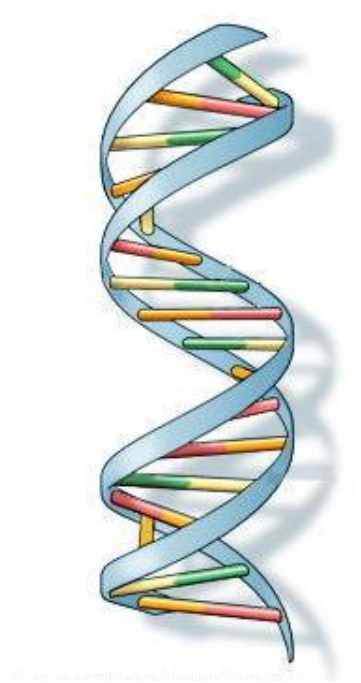
TABLE OF CONTENTS

II. CHAPTER-WISE EXPERIMENTAL SECTION

PART I:
DESCRIPTIVE SECTION

CHAPTER 1

AN INTRODUCTION TO DNA – DISCOVERY, STRUCTURE, FUNCTION AND IMPORTANCE OF dsDNA RECOGNITION[I]

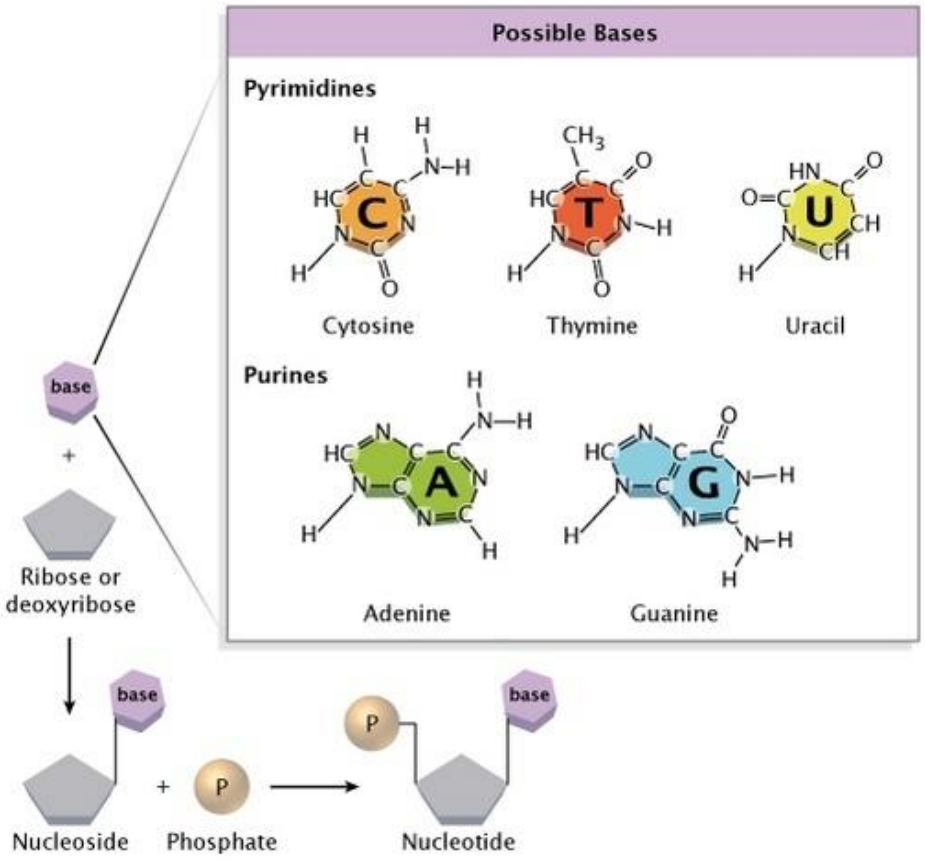


It is a misconception that American biologist James Watson and English physicist Francis Crick discovered DNA in the 1950s. DNA was in fact first identified several decades before Watson & Crick by Swiss chemist Friedrich Miescher[1, 2] *in the late 1860s. Of course at the time no one knew what it was or that it even existed. Miescher first identified what he called as "nuclein" present inside human white blood cells. (The name nuclein was later changed to nucleic acid and eventually to deoxyribonucleic acid or "DNA".) Decades later, other scientists – notably, Russian biochemist Phoebus Levene*[3] *and Austrian biochemist Erwin Chargaff*[4] *– carried out a series of systematic studies that revealed additional details about DNA. The most important among those details were identifying its primary chemical components and the ways in which they were connected with each other. Another noteworthy contribution was that of the English Chemist and crystallographer Rosalind Franklin*[5] *for her work on X-ray diffraction images of DNA. If this scientific foundation had not existed, Watson and Crick would probably not have reached their revolutionary conclusion in 1953: that the DNA molecule exists in the form of a three-dimensional double helix.*[6]

1.1 Basic constituents of DNA

Despite its complexity, it is quite amazing that DNA comprises only of four basic nucleotides (building blocks of DNA). A nucleotide, in turn, is made up of three sub units – a nitrogenous base, a sugar, and a phosphate group. The nitrogenous base can be either a pyrimidine or a purine. The three pyrimidines are: cytosine (C), thymine (T), and uracil (U). All three pyrimidines are made up of a single ring. The chemical structures of the two purines, adenine (A) and guanine (G), are also shown (Figure 1.1). Each purine is made up of a fused double ring. A nitrogenous base is attached to a ribose or deoxyribose sugar, forming a molecule called a nucleoside. When a phosphate moiety is attached to the other side of the sugar, the phosphate, sugar and base together form a nucleotide. Note that DNA does not contain Uracil (U). Thus, A, G, C and T form the four basic nucleotides which constitute DNA.

[I] **Note**: Throughout this chapter, wherever possible or available, the source has been mentioned for individual images. If not, the general rules regarding copyright are as follows: You may reproduce this material, without modifications, in print or electronic form for your personal, non-commercial purposes or for non-commercial use in an educational environment.



[II]**Figure 1.1**: The chemical structure of a nucleotide

1.2 Structure of dsDNA

The simplified illustration of dsDNA, which is shown in Figure 1.2, indicates certain important properties of dsDNA:

1. The first is that the two strands of the helix are arranged in an anti-parallel manner. This is evident from the grey ribbons, which represent the sugar-phosphate backbone, which have arrows that run in opposite directions.

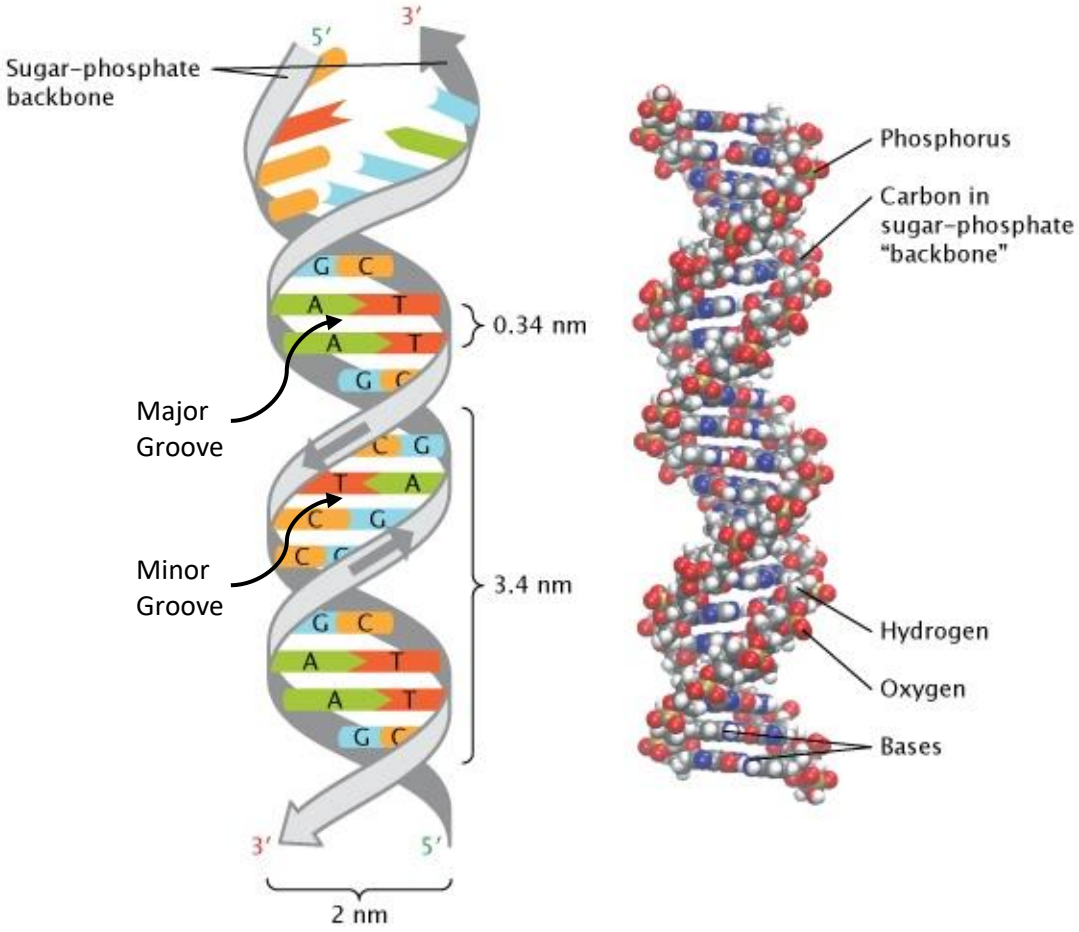
2. The second is that the upper end of one strand is labelled five prime (5'), and the lower end of the same strand is labelled three prime (3'). Also, the upper end of the opposite strand is labelled 3', and the lower end of the same strand is labelled 5'. As a result, the 5' end of one strand matches up with the 3' end of the other strand on each end of the double helix.

3. The third is that the two strands are held together by the pairing of complementary nucleotide bases on opposite DNA strands. The nucleotide bases are shown as differently coloured rectangles. These blocks fit together like pieces of a puzzle (meeting in the middle). The nucleotide guanine (G, shown in blue) binds with the nucleotide cytosine (C, shown in orange). The nucleotide adenine (A, shown in green) binds with the nucleotide thymine (T, shown in red). The base pairs between the two strands look like the rungs of a ladder. The distance between two base pairs, or "rungs", as proposed by Watson and Crick was 0.34 nanometres. The length of one turn of the double-helix was proposed to be 3.4 nanometres. The width of the DNA molecule was proposed to be two nanometres. Subsequent measurements over the years however revealed that this was not the case and that every turn

[II] Reproduced from Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* 1(1):100

was in fact closer to 3.5-3.6 Å (or 0.35-0.36 nanometres and not 0.34) and it was found that 10.4-10.5 base pairs (and not 10) constituted a helix turn (ref. 7).

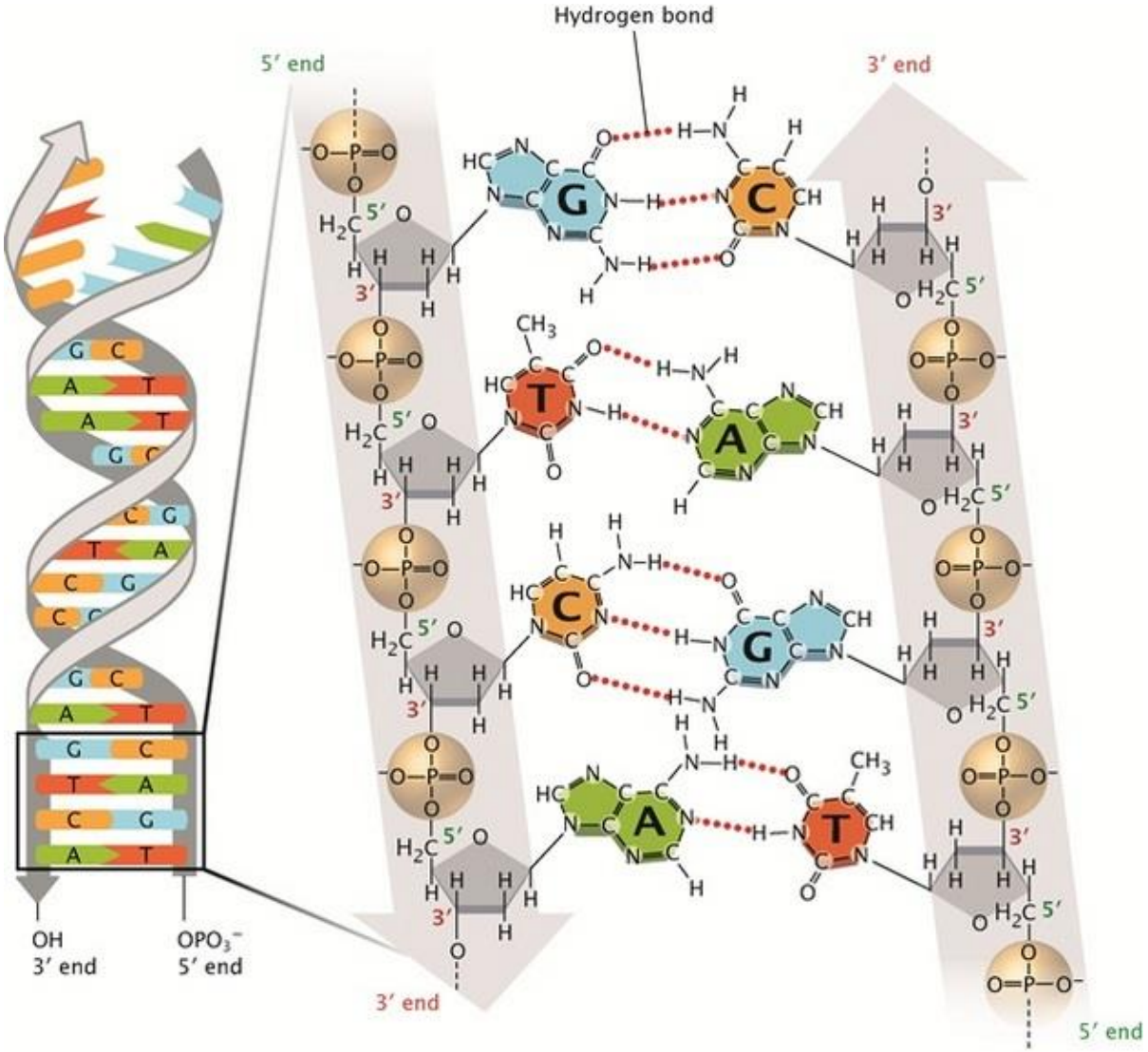
4. A space-filling molecular model is depicted on the right. The colour codes are: gold spheres for phosphorus atoms, grey spheres for carbon atoms, white spheres for hydrogen atoms, red spheres for oxygen atoms, and blue spheres for nitrogen atoms. The DNA molecule's sugar-phosphate backbone is made up of phosphorus, carbon, hydrogen and oxygen. The bases, which contain hydrogen, oxygen, nitrogen and carbon, connect the two sugar-phosphate backbone chains.



[III]**Figure 1.2**: The double-helical structure of DNA as proposed by Watson and Crick along with the formation of major and minor grooves.

5. Another key aspect of DNA assembly is the formation of major and minor grooves. This will be of great importance in the subsequent sections when we discuss DNA recognition. The major and minor grooves form as a result of the offset pairing between the two single strands of DNA as can be seen in Figure 1.2.

[III] Reproduced from Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* 1(1):100



[IV]**Figure 1.3**: Base pairing in DNA as correctly elucidated by Watson & Crick

Figure 1.3 shows the molecular structure of a DNA region illustrated in detail to show the pairing of the bases. The key points of the figure can be summarized as follows:

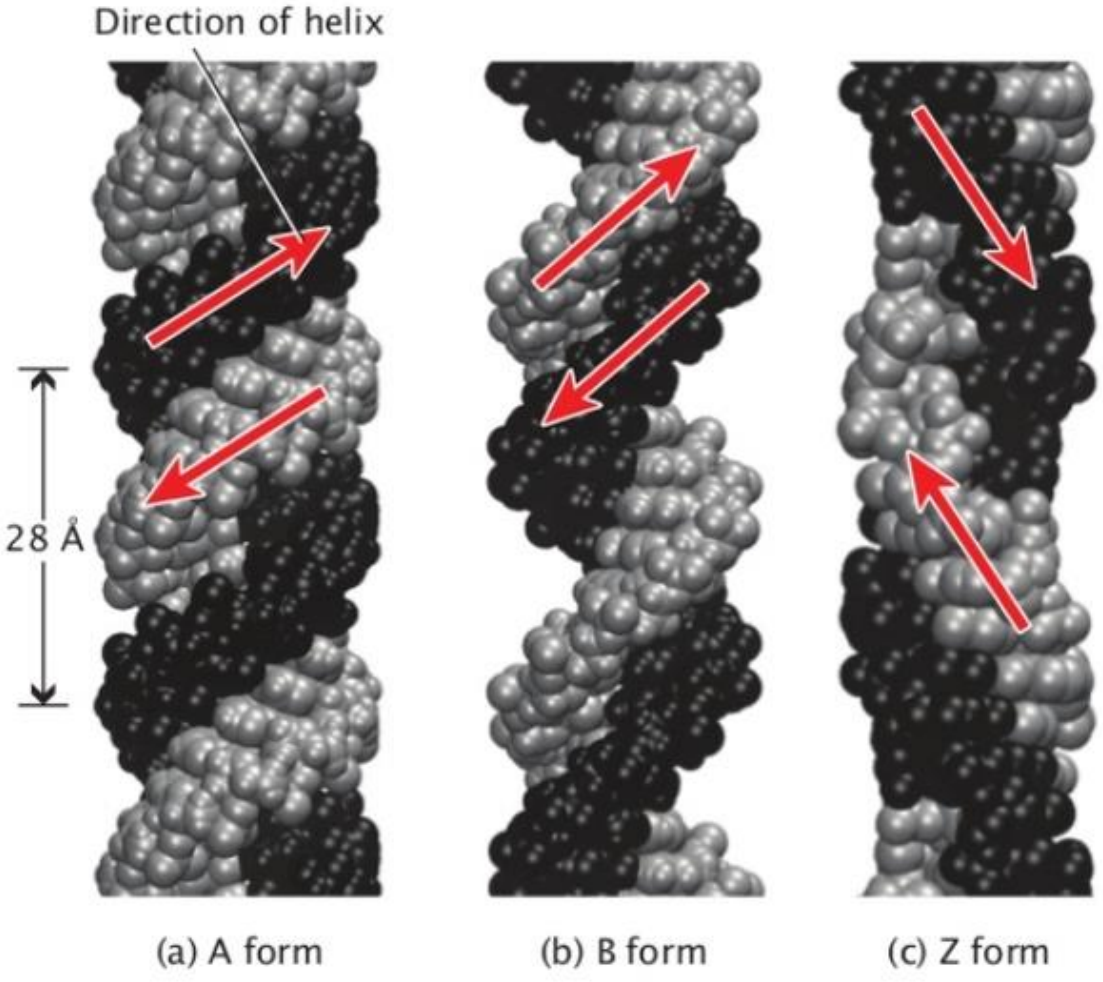
1. The simplified illustration on the left of Figure 1.2 depicts the basic structure of the DNA double helix. This structure has been explained in detail in the previous section. In this figure, the bottom of the DNA molecule has been expanded to elaborate upon the following: the strand on the left is labelled three prime hydroxyl (OH), and the strand at the right is labelled five prime phosphate (OPO_3^-). The bottom four base pairs are shown flattened instead of twisted, so that this region (marked with a black box) can be easily expanded to depict a more interesting, in-depth, close-up view.

2. The expanded section shows the individual atoms and bonds in the DNA molecule. Phosphate groups are depicted within light brown spheres. The bonds between the phosphate and oxygen atoms are also shown. The grey pentagons represent the sugars. The carbon atoms are unmarked to avoid over-crowding of the figure. From the figure, it is clear that the oxygen atoms and hydrogen atoms are attached to the carbon atoms. An oxygen atom from each phosphate moiety is connected by a solid black line to a carbon atom from the sugar moiety.

[IV] Reproduced from Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* 1(1):100

Covalent bonds between the sugars and phosphate groups are also represented by solid black lines. Each sugar molecule is individually attached to a single nitrogenous base. The nitrogenous bases from the two DNA strands form the central part of the molecule. The bases interact with each other and are thus linked via hydrogen bonds (shown by dotted, red lines).

3. Further introspection of the base pairing reveals that top-most the guanine base with two fused rings (**G**, shown in blue) is bound to a cytosine base with a single ring (**C**, shown in yellow) on the opposite strand. These two bases are held together by three hydrogen bonds. Below this base pair, a thymine base with a single ring (**T**, shown in red) is bound to an adenine base with two fused rings (**A**, shown in green) on the opposite strand. These two bases are held together by only two hydrogen bonds. Below this pair, a single-ringed cytosine base is bound to a double-ringed guanine base by three hydrogen bonds. In the final pair, an adenine base with two fused rings is bound to a single-ringed thymine by two hydrogen bonds.



[v]**Figure 1.4**: Three different conformations of the DNA double helix[8]

Although Watson & Crick's original model depicting the structure of dsDNA has remained more or less unchanged for decades, over time scientists have elaborated on the original model. One important discovery worth mentioning is the identification of three different conformations of the DNA double helix. In simpler terms, the precise geometries and dimensions of the double helix of DNA can vary.

[v] Reproduced from Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* 1(1):100

It should be noted here that most DNA double helices are right-handed. The simplest way to visualize this is to do as follows: hold your right hand out, with your thumb pointed up and your fingers curled around your thumb. For a right handed double helix, your thumb would represent the axis of the helix and your fingers would represent the sugar-phosphate backbone. Based on the directions of the helix, three different conformations of the DNA double helix exist: B-DNA, A-DNA and Z-DNA (Figure 1.5). Only one type of DNA, the Z-DNA, is left-handed. The most common conformation in most living cells (which is the one depicted in most diagrams of the double helix, the one proposed by Watson and Crick and the one which will be referred to throughout this thesis) is known as B-DNA. There are also two other less common conformations: A-DNA – a shorter and wider form of DNA found in dehydrated samples of DNA. A-DNA is rarely found under normal physiological circumstances. The other is Z-DNA, a left-handed conformation. Z-DNA is a transient form of DNA. Z-DNA only occasionally exists in response to certain types of biological activity. Z-DNA was first discovered in 1979, but it was considered unimportant resulting it being largely ignored until recently. Since then, however, scientists have discovered that certain proteins bind very strongly to Z-DNA. It is therefore a possibility that Z-DNA has a key biological role in protection against viral disease.[9]

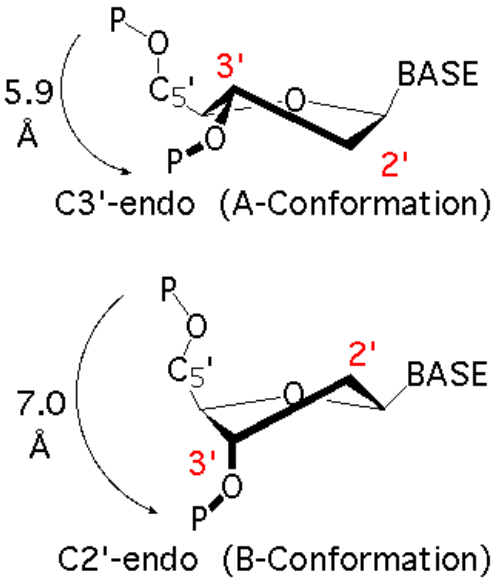
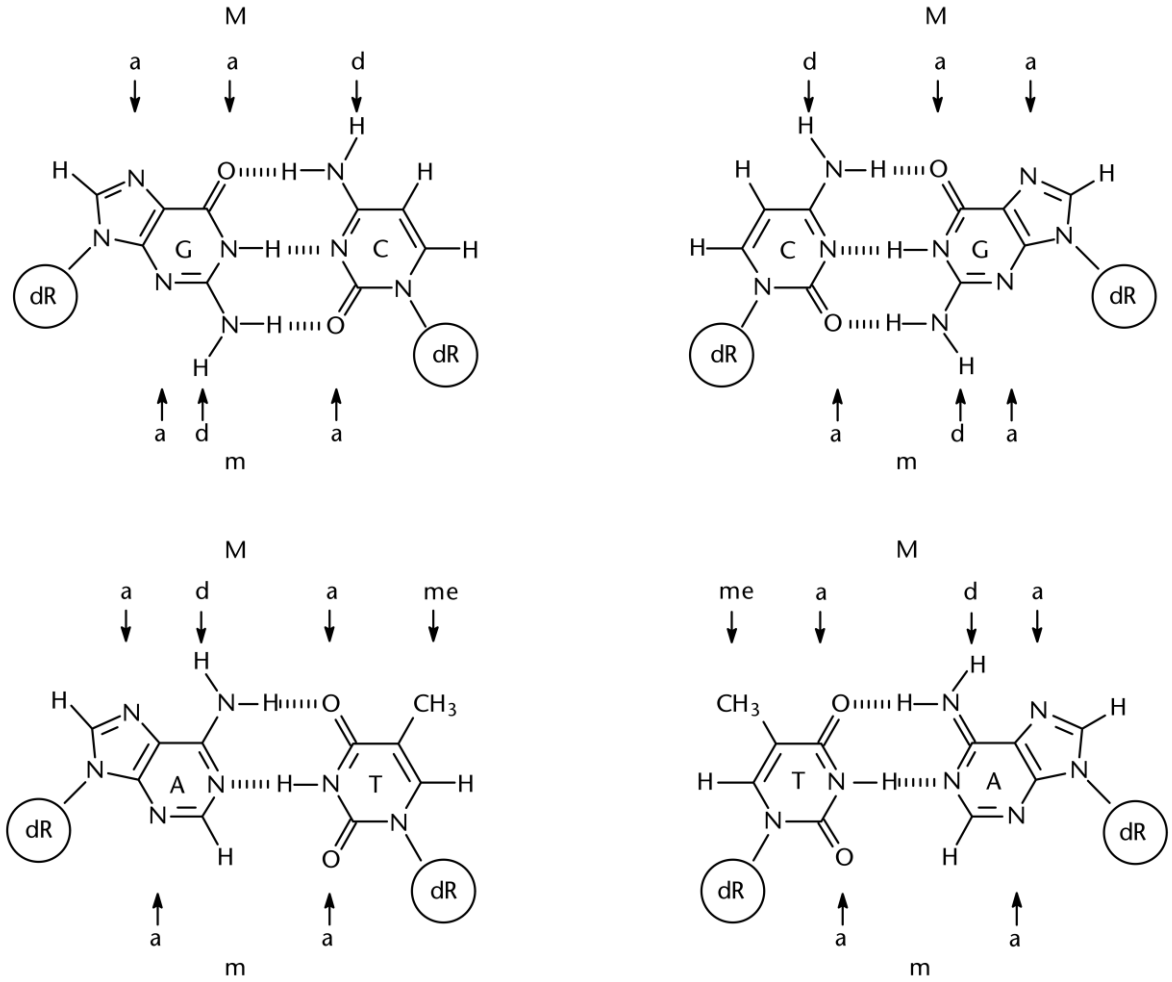


Figure 1.5: Figure showing the difference in conformation of the deoxyribose sugar ring between A-DNA and B-DNA.

The major difference between A-DNA and B-DNA[10] is in the conformation of the deoxyribose sugar ring. It is in the C2' endoconformation for B-DNA, whereas it is in the C3' endoconformation in A-DNA (Figure 1.5). For steric reasons, the 3' and 2' carbons present in the ribose ring will shift from the plane of the ring. As it can be seen, the C3'-endo A-form will hold the phosphate groups closer to the 3' OH whereas the C2'-endo B-form holds it further.

1.3 Structural basis of dsDNA recognition

Studying the three-dimensional structure of dsDNA (especially the B form) is critical to understanding dsDNA-protein interactions (which will be discussed extensively in the subsequent sections). Before this, however, an understanding on the available functional moieties within the major and minor grooves of dsDNA is necessary.



[VI]**Figure 1.6**: Points of recognition in the major (M) and minor (m) grooves of dsDNA for each of the four base pairs. a, electron acceptor; d, electron donor; me, methyl group. Hydrogen bonding in base pairs is indicated by dashed lines. dR in circles denotes the deoxyribose-phosphate backbone of dsDNA.

Sequence specificity in dsDNA can be achieved through a direct readout mechanism. This mainly involves various non-covalent interactions such as hydrogen bonding, electrostatic forces, salt bridges and van der Waals contacts (described in section 1.6) between the functional groups of the amino acids and/or back bone atoms of the protein and the functional groups available on the bases in the major and minor grooves of the DNA.[11] If there is no disruption in the normal Watson–Crick hydrogen bonding, each of the four types of base pair (AT, TA, GC, CG) presents a distinct pattern of hydrogen donor, hydrogen acceptor and hydrophobic groups available for bonding (Figure 1.6). A protein can thus easily distinguish between the four possibilities (through interaction in the major groove) by using two hydrogen bonds. In the minor groove, it is possible to distinguish GC/CG from

[VI] Protein-DNA Complexes: Specific, Mark A Strauch, Encyclopedia of Life Sciences, 2001 Nature Publishing Group / www.els.net

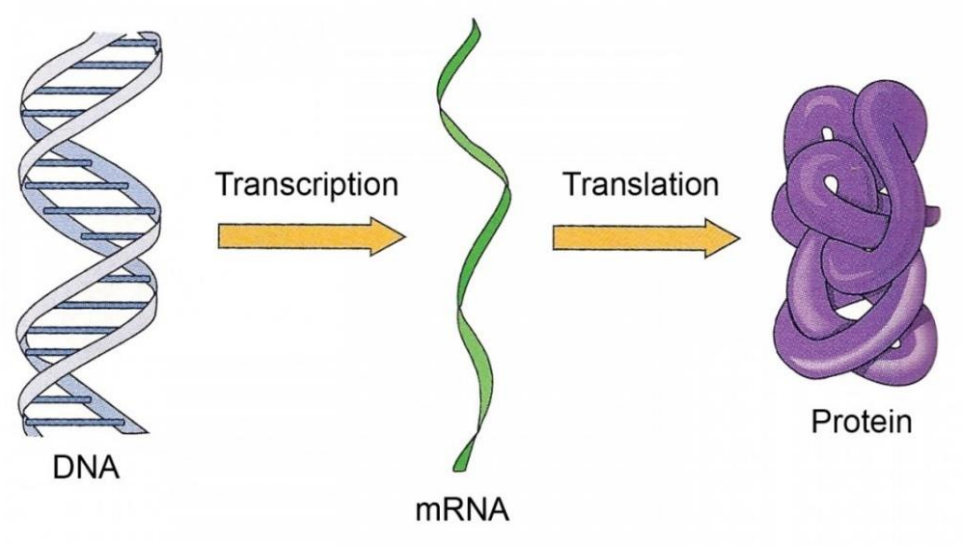
AT/TA through the presence of a hydrogen donor from the 2 amino group of guanine. However, in order to reduce the degeneracy further, discrimination of GC from CG or AT from TA in the minor groove must involve highly sensitive means exploiting the subtle different steric positioning and spacing between atoms. Alternatively, the slight difference in hydrogen bond energies to the O2 of pyrimidines versus the N3 of purines can also be used to distinguish GC from CG or AT from TA. Such selectivity is possible and has been described in literature.[12, 13] The sequence selectively of DNA-binding proteins which allows them to bind to sequences containing certain DNA base pairs over others arises as a result of their unique structural characteristics (described in section 1.5). In majority of the cases, study of different sites of a protein bound to DNA reveals a general basis for sequence selective recognition.

1.4 Functions of DNA

Our previous discussion allowed us to understand more about the structure of DNA. The more important question, however, is "what is the function of DNA"?[14] Unfortunately, no one knows for sure the full range of functions DNA can perform. Research is a continuous process and in terms of decoding the full range of functions performed by DNA, we have barely scratched the surface. That being said, certain things are known for sure. There is absolutely no doubt that the functions of DNA are vital for inheritance, for the coding of proteins and that DNA is the very genetic blueprint of life itself. Given the enormous number of functions DNA performs in the human body and the fact that it is invaluable in terms of growth and maintenance of life, it is not surprising that the discovery of DNA has led to countless developments in treating disease. In short, DNA holds the instructions for an organism's development, reproduction and ultimately, its survival. Some of the most essential known functions of DNA, namely the coding for proteins, replication and the role DNA plays in defining our genetic code have been discussed here in the simplest possible terms and some of them will be discussed in more depth and detail throughout the course of this thesis.

1.4.1 Coding for Proteins

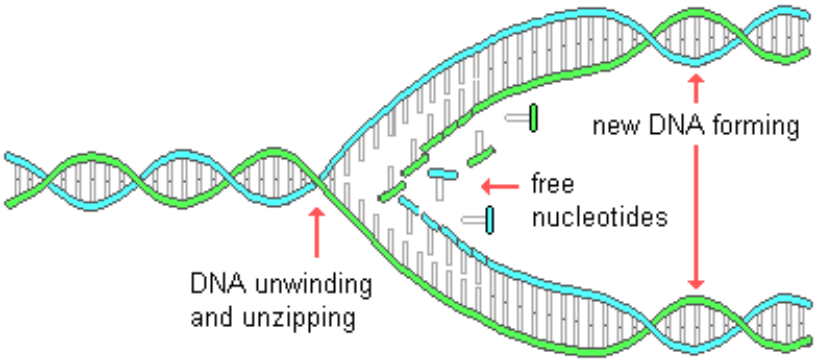
DNA holds the code for proteins, which are complex three dimensional molecules that are either involved in majority of the chemical reactions and/or which play critical roles and are vital for the structure, function and regulation of tissues in our body. Depending on the function they perform, proteins can be of the following types: antibodies, enzymes, messenger proteins. Special proteins may also provide structure and support for cells and transport cells or store small molecules throughout the body. Information in DNA is initially 'read' by certain proteins (called Transcription Factors, TFs) and it is then transcribed into messenger RNA (Ribonucleic acid). This process is known as transcription which is explained in more detail in section 1.5.2. Thereafter, the information held in this messenger molecule is translated (the process is known as translation) into a "language" that the body can understand. This language is nothing but that of amino acids, which are essentially the building blocks of proteins. There are twenty natural amino acids from which proteins are made. It is this specific language that dictates how the amino acids should produce a particular protein. Given that there are twenty different kinds of amino acids, it is but obvious that their ordering can produce a staggering number of varied proteins.



[VII]**Figure 1.7**: The flow of information from DNA to protein.

1.4.2 DNA Replication

DNA replication is a vital and virtually endless process occurring with a phenomenal speed and accuracy in the human body. DNA replication is a must for survival given the wide range of functions it is needed for – ranging from reproduction to maintenance and growth of cells, tissues and all body systems. To replicate itself, a DNA molecule essentially 'unwinds and unzips' (Figure 1.7), thus resulting in a series of bases without pairs along the backbone of the molecule. As explained previously, DNA has four bases – all part of a nucleotide that also consists of a sugar and phosphate. The base pairing in DNA is extremely specific and only 2 pairs can be formed out of the 6 possible pairs. This means that adenine will only pair with thymine and guanine will only pair with cytosine. As the free nucleotides start to assemble and pair up with the unpaired bases on the backbone of the DNA molecule, they build a new strand. This strand can complement or match up with the original sequence. The end result is the formation of a new strand which is a perfect match to the original one before the whole replication process began. Cells in the body replicate for purposes such as making new skin or blood cells. When mistakes occur, there are repair systems in place to remedy the mistake or alternatively, a cell has a marker for destruction. If a cell survives a mutation, there may still be benefits to an organism. In fact, this concept is essentially the basis for evolution.



[VIII]**Figure 1.8**: A simple schematic showing the DNA replication process

[VII] ©The McGraw-Hill Companies, Inc. Permission granted for reproduction

1.4.3 Genetic Code

One of the more complex functions of DNA is genetic coding, in the sense that it transfers genetic messages to all of the cells in your body. The basic physical and functional unit of heredity is called a gene which is made up of DNA. Without delving into too much detail, think about DNA in a reproductive sense: Consider the joining of an egg and sperm to create your first cell. This first cell already contains your complete genetic code that your body would use your entire life. Within that initial cell, half of your chromosomes which contain your DNA came from your father and half came from your mother.

To conclude, DNA clearly plays important roles in the human body and decoding the structure and function of DNA is one of the most significant discoveries of the twentieth century. Ongoing research into DNA and furthering our knowledge of DNA functions will definitely help us to learn even more about this all-important molecule.

1.5 Importance of dsDNA recognition

We have already discussed in detail the structure of dsDNA and established the importance of DNA in various functions in the human body. We have also briefly discussed the process of coding of proteins and DNA replication. In this section, we will take a closer look at why these processes are so important, what is cancer, what is transcription, the process of transcription and how Transcription Factors and cancer therapy relate to one another.

1.5.1 What is cancer?

Cancer can be defined as the out-of-control proliferation of a particular cell type. In all cases, this proliferation originates with an unwanted mutation[15]. This is subsequently followed by a large accumulation of defects in different classes of genes. There are two well-known types of primary genes that govern cell division and are responsible for cancer:

1. Protooncogenes (gain-of-function) are the first type of genes. These genes speed up the process of activation of the cell cycle.
2. The second are tumour suppressor genes (loss-of-function). These do the opposite, i.e. they slow down the growth of cells.

Till recently, there were 17 known signal transduction[16] (the process of transmitting molecular signals from the exterior of a cell to its interior) pathways, in addition to at least two stress-response pathways. All of these appear to be highly conserved in nematodes (roundworms), flies and all vertebrates. Ultimately, TFs participate at the ends of all 19 pathways. They cause the up- or down-regulation of specific genes. All primary and modifier genes which lead to cancer participate in one or another of these pathways. Countless exogenous and (autocrine and paracrine) endogenous signals bombard our cells each day. All these signals are channelled through these 19 pathways, leading to the cell's response to these signals. In simplistic terms, tumour progression represents a loss of normal cross-talk between cells. This may also involve breakdown in communication between classes of genes, DNA methylation abnormalities, genetic instability, and hypermutability. Cancer is thus an extremely complex disease: a multitude of defects in hundreds if not thousands of genes, over an extended period of time, leading to an invasive and lethal disease[17].

[VIII] Copyright © 1998-2014 by Dennis O'Neil. All rights reserved: http://anthro.palomar.edu/biobasis/bio_5.htm

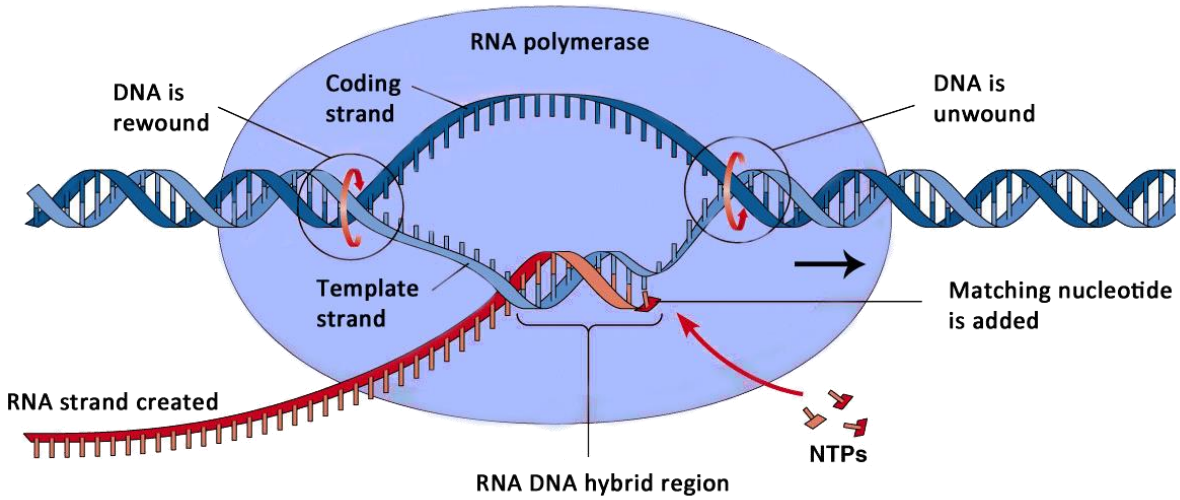
In the broadest of terms, two categories of cancer exist:

1. The inherited form occurs when an individual is born with a mutation in one of his primary cancer genes.
2. Oxidative DNA damage is the ultimate result of environmental agents. Additionally, endogenous compounds can also lead to reactive oxygenated metabolites and reactive oxygen species.

Not all damage is repaired and therefore goes on to become fixed mutations, resulting in defective gene by-products. With each increasing year of life, therefore, these mutations become more and more likely to cause crucial defects in primary and modifier cancer genes. This in turn leads to tumour initiation and progression in one or another specific cell type.

1.5.2 The process of transcription

The expression of genes is vital for functions of the body. The genes present in the nucleus of a cell need to be "read and decoded" for the biosynthesis of proteins or functional RNAs. This occurs through the processes of transcription (conversion of dsDNA into single stranded mRNA) and translation which is the subsequent production of proteins. RNA can be regarded as the chemical cousin of DNA. RNA is made up of A's, U's (no T's are present in RNA), G's and C's. RNA can be regarded as the set of molecules primarily involved with most of the interpretation and communication with primary DNA.



[IX]**Figure 1.9**: The DNA Transcription process in detail.

Gene expression begins when a chemical signal is originated inside or is received from outside the cell. This signal triggers a cascade of intracellular signalling and events. "Initiation" is the first step of the "transcription" phase of gene expression. Transcription Factors (TFs) perform intracellular signalling via binding to and regulating an enzyme called DNA polymerase. DNA polymerase itself attaches to a specific promoter region of DNA. The promoter region of DNA will be entirely dependent on the transcription factor. Once attached "elongation" begins, the RNA polymerase complex acts like a zipper and moves along the DNA sequence step-by-step. This unwinds the DNA, making a copy of the template strand of DNA. The template strand is defined as the strand from

[IX] Reproduced from https://limbiclab.com/ © 2012 Pearson Education, Inc.

which RNA is transcribed. Free-roaming A's, U's, G's and C's bind to the template strand creating a temporary RNA-DNA hybrid region, which ultimately creates the messenger RNA (mRNA) in the process. As the DNA polymerase moves along the DNA strand, it unwinds the DNA from the front while simultaneously stitching up the DNA at the back. Until a specific terminating region is reached, this process will continue. Finally, the mRNA will detach itself from the DNA, completing the "termination" phase.

1.5.3 Transcription factors in cancer therapy

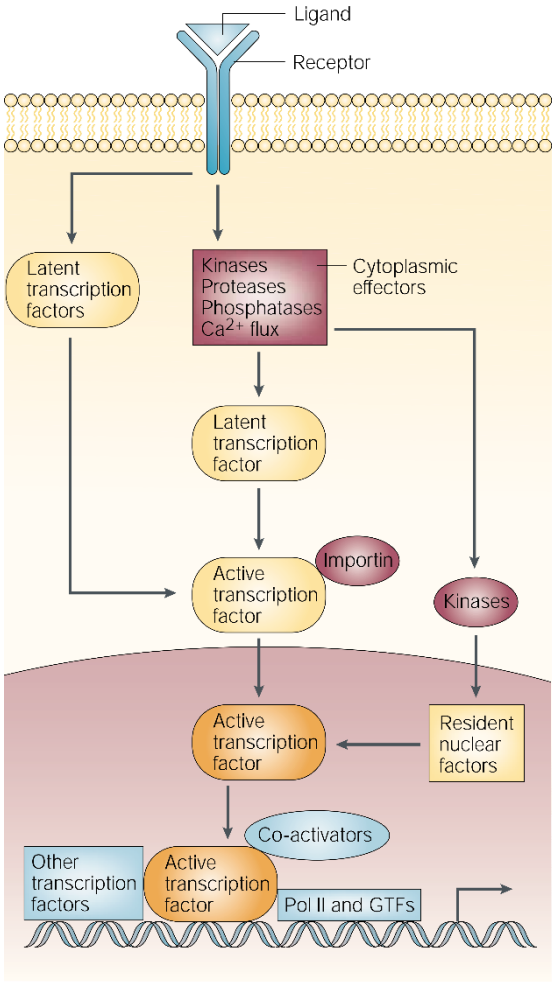
In human cancer, three main groups of transcription factors are known to be important. The steroid receptors, for example, were among the first to be recognized. Oestrogen receptors in breast cancer and androgen receptors in prostate cancer are two examples involving steroid receptors. Compounds such as amoxifen and bicalutamide which are anti-oestrogen and anti-androgen respectively have been in clinical use for many years. The ability to induce apoptosis, at least in lymphoid cells, have also resulted in the wide use of active glucocorticoids.[18, 19, 20]

The second group of transcription factors that were identified to have a role in cancer are resident nuclear proteins, which are activated by serine kinase cascades.[21] In 1987, JUN was shown to be an oncogene. Initially, the v-JUN of a retrovirus, then the normal cellular form, c-JUN. Soon thereafter, JUN was found to be a transcription factor. This was initially based on its similarity in a DNA-binding domain to the yeast transcription factor, GCN4[22 23] which was already identified at the time. (The GCN4 TF will be discussed in detail in the subsequent chapters). This was a highly significant discovery as it provided the first concrete example, besides steroid receptors, that directly linked cancer to transcriptional factors. Subsequent work showed that c-JUN was phosphorylated on two specific serines residues – serine 63 and 73. This was a result of the activation of signalling cascades, which made c-JUN very highly effective in stimulating transcription.[22, 24] c-JUN is just one among the few hundreds of nuclear proteins that are targets of serine kinase cascades. The entire set of nuclear proteins are initiated in the cytoplasm. It is now known that there are more than 500 serine kinases.[25] (Figure 1.10)

A third group of transcription factors which were only recently recognized to have oncogenic potential are latent cytoplasmic factors. The activation of these TFs is normally triggered by receptor–ligand interaction at the cell surface[21]. Direct activation of the latent cytoplasmic proteins can be achieved through tyrosine or serine kinases at the cell surface. Alternatively, they can also be activated by various different cytoplasmic biochemical events. These events may feature kinases (some regulated by Ca^{2+} flux) or specifically regulated proteolysis. Despite the pathways appearing different, they all have one thing in common which is a protein−protein interaction at the cell surface. This protein-protein interaction triggers cytoplasmic reactions, which result in delivery to the nucleus of a protein. The delivery of the protein in turn increases transcription via interaction with one or more of the many proteins affecting the initiation of transcription (Figure 1.10).

There is a high possibility that an oversupply or over-activity of one or more transcription factors from the above mentioned three classes will be required for the survival, unrestricted growth and metastatic behaviour of all human cancers.[22, 21, 26] Effective anticancer therapy can therefore be developed via a direct and promising route by the inhibition of excess transcription-factor activity. The therapeutics which have been developed to inhibit the steroid receptors form one such example.[19, 20] However, there is still a big knowledge deficit with regards to the other transcription-factor classes.

Further research into the prominent features of their activation and function, with special focus on their interactions with other nuclear proteins, should reveal a wider range of therapeutic targets.



[X]**Figure 1.10**: The figure shows the generalized signalling pathways summarizing key reactions in humans. Information for a few thousand cell-surface protein receptors and in general small, secreted proteins which are ligands for these receptors are contained in the human genome. The biochemistry of intracellular signalling has been very widely studied. Latent transcription factors (Figure 1.9, top) might be activated at the cell membrane by either tyrosine (STATs) or serine (SMADs) phosphorylation. The latent TFs then bind to the transport proteins called importins which convert them into an active TF. This active TF then enters the nucleus to participate in regulated gene transcription. Phosphorylation is one of the other, less direct pathways of activation for latent cytoplasmic factors. These events may be stimulated or inhibited and most often block or trigger proteases. The final result is importin-mediated delivery of an 'active transcription factor' (orange) to the nucleus. In a different set of pathways, the kinase enters the nucleus (Figure 1.9, right) and causes an activation of resident nuclear-transcription factors. The active transcription factors (in combination with DNA) can bind other transcription factors forming what can be called an "enhanceosome". This can attract co-activators which may include several dozen possible proteins. This ultimately leads to the final step of attracting the actual RNA synthesis machinery which comprises RNA polymerase II (Pol II) and general transcription factors (GTFs).

To summarize, it is clear that several complex mechanisms involving various TFs are involved in the onset of cancer. Therefore, despite years of research, there is still a lack of clear understanding and knowledge about the disease itself. This lack of knowledge has significantly hampered the development of gene therapy strategies in the past. Therefore, the study of TFs has gained a lot of traction in recent years, especially where other therapeutic strategies fail. The prevention of specific

[X] Reproduced from James E. Darnell, Jr www.nature.com/reviews/cancer, October 2002, Vol. 2, Pages 740-749

genes from being transcribed could ultimately result in a therapeutic platform with diverse biomedical applications.[27] The past few years have been no doubt fruitful with medicinal chemists & biologists making considerable progress targeting specific genes. In nature, certain proteins interact with DNA in a sequence-specific manner by means of their recognition domains, which consists of a short sequence of amino acids. Simplification of known TFs could ultimately lead to a peptide-based drug in cancer treatment. Moreover, due to the improvements in peptide manufacturing, peptide drugs can be produced on a large scale through synthetic methods and techniques. Therefore, the continuing study of Transcription Factors (TFs) in even greater detail for the modification of gene expression can be of considerable importance for the development of new anti-cancer therapies.

1.6 Interactions between DNA and proteins: A key factor regulating transcription

Section 1.2 described in detail the structure of dsDNA. In order to better understand the interactions between DNA and proteins, a closer look at the structural features of proteins along with their increasing levels of complexity is necessary.

1. **Primary structure**: The exact sequence of amino acids which constitute a protein and the type of covalent bonds (amide, ester, disulphide etc.) between various amino acids.

2. **Secondary structure**: The parts within a protein which are folded or coiled. Examples include α-helices and β-sheets. Most secondary structures are stabilized via hydrogen bonding or Van der Waals forces.

3. **Tertiary structure**: The final three-dimensional structure of a protein, formed as a result of a large number of non-covalent interactions between its constituent amino acids.

4. **Quaternary structure**: Large number of non-covalent interactions that bind multiple proteins or peptides into a single, larger protein. Example: Haemoglobin has quaternary structure due to association of two alpha globin and two beta globin polyproteins.

It has been established that certain Transcription Factors can specifically recognize and bind to short regulatory DNA sequences which control the gene. These interactions between DNA and proteins are extremely specific and of a very high affinity. Studying these interactions holds the key for synthesis of mimics of transcription factors which could be eventually used for cancer therapy.

The interactions between DNA and proteins is extremely complex. The lack of a general recognition code for the interaction between amino acid sequences within a protein and its corresponding specific DNA-binding site has been a major challenge in the study of the structure of TFs since this makes the prediction of its binding site virtually impossible. Although a general formulation of these interactions has not been achieved, some general rules can be followed, providing a starting point for the molecular basis of specific DNA recognition.

Due to the fact that secondary structures in case of proteins are extremely well defined, very specific interactions take place between TFs and dsDNA. In case of TFs, all these interactions occur in the major groove of the DNA, which is wide enough to accommodate the α-helix recognition unit. A detailed explanation of the four major forces between proteins and nucleic acids can be summarized as follows:

1. **Electrostatic forces** in the form of salt bridges: Long range, electrostatic interactions between groups of opposite charges are quite dominant due to the negatively charged DNA as a result of the phosphodiester backbone and the positively charged basic side chains of the amino acids of the protein, such as Lys, Arg and His. Electrostatic forces are unspecific and therefore only affect the free energy of association and not the specificity.
2. **Dipolar forces** in the form of hydrogen bonds: Short range interactions that contribute to a high degree to the specificity of the interactions that occur between DNA and proteins also occur in a substantial number. These interactions mainly occur between the polar side chains in the proteins and the sugars and bases in the DNA. Higher degrees of specificity can be achieved by bidentate hydrogen bonds between a base or base pair in the DNA with a single side chain present in the protein.
3. **Van der Waals forces** due to a hydrophobic effect: These are short range, non-directional forces which contribute to the free energy of association. Non-polar contacts also occur between the apolar amino acid side chains and the DNA base pairs.
4. **Water-mediated hydrogen bonds**: Water molecules are essential for indirect hydrogen bonding between the DNA and proteins. They also behave as space fillers in the DNA-protein complexes. Hydrogen bond donors and acceptors present especially at the edge of a DNA-protein complex are more influenced by water.

From strongest to weakest, the order of forces in an aqueous medium between DNA and proteins are generally: electrostatic forces > Water-mediated hydrogen bonds > Dipolar forces > Vander Waals forces. Electrostatic forces dominate due to the high number of negative charges present in DNA being strongly attracted to the positively charged proteins. Although an individual Hydrogen bond is not the strongest, the cumulative effect of a large number of hydrogen bonds makes this is dominating force in case of DNA-protein interactions. Dipolar forces are weaker than electrostatic forces and hydrogen bonds, but still stronger than Van der Waals forces.

1.7 Families of transcription factors: dsDNA recognition in the major groove

At present, there is great interest in the area involving human transcriptional regulation. However, very little is known about Transcription factors which regulate human transcription. Several questions come to mind. For example, how many TFs are present in the human genome? How are they expressed in various tissues? A list of 1,391 sequence-specific dsDNA-binding transcription factors, their functions, genomic organization and evolutionary conservation have been described in an excellent review article[28]. Despite this, a great deal remains unexplored. The study of TFs in detail is beyond the scope of this thesis; however, a short summary of the relevant families of TFs for this thesis is provided in this section[29].

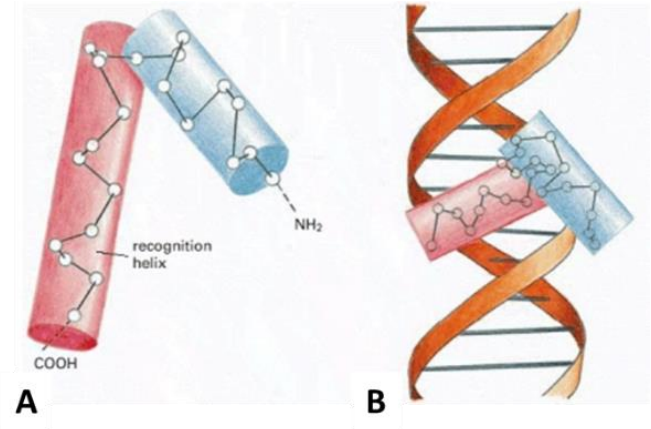
The protein structure of TFs has been divided into two broad domains for the purpose of easy understanding:

1. **The DNA binding domain**: This part of the protein is responsible for the selective interaction with the DNA control sequence for transcription.
2. **The DNA activation domain**: This part is responsible for the interaction with other TFs or with the RNA polymerase II.

There are several DNA binding domains found in TFs in eukaryotic cells: Helix-turn-helix, zinc finger and the basic domain.

1.7.1 The Helix-turn-helix (HTH) motif

The helix-turn-helix was the first DNA-binding protein motif to be discovered. It was first identified in bacterial proteins, but now this motif is known to be found in hundreds of DNA-binding proteins in both eukaryotes and prokaryotes. It has a very simple secondary structure consisting of two α helices connected together by a short slightly elongated chain of amino acids which constitute the "turn". (Figure 1.11). The angle between the two helices does not change, mainly due to interactions between the two helices. The helix which is more towards the C terminus is called the recognition helix. It gets its name from the fact that it can fit perfectly into the major groove of DNA. The amino acid side chains of the recognition helix differ from protein to protein, which enables it to recognize the specific DNA sequence to which the protein is required to bind.

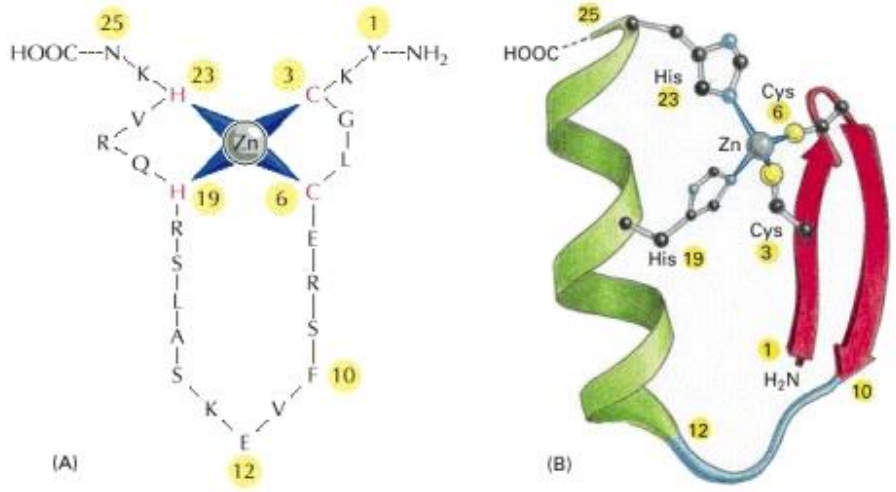


[XI]**Figure 1.11**: The helix-turn-helix motif is shown in (A). Each white circle denotes the α-carbon atom of an amino acid. The C-terminal α helix, i.e. the recognition region, is shown (red). Part (B) of the figure shows how this helix fits into the major groove of DNA, where it is in contact with the edges of the base pairs. It is clear from the figure that the N-terminal α-helix (blue) is present primarily to provide a structural component which helps to position the recognition helix.

1.7.2 Zinc-finger motifs

The Zinc-finger motifs are a second important group of DNA-binding motifs. They are known as Zinc-finger motifs because their structure contains one or more zinc ions. There are several types of zinc finger motifs, only one of which will be discussed here. Although we refer to all zinc-coordinated DNA-binding motifs as zinc fingers, this name refers only to the way they appear in schematic drawings ever since they were discovered (Figure 1.12). Structural studies over the years have shown that they can be divided into a number of distinct structural groups. Only the first one discovered is described here. The protein shown in Figure 1.12 belongs to the Cys-Cys-His-His family of Zinc finger proteins. It is known to activate the transcription of a eukaryotic ribosomal RNA gene. The structure is simple, consisting of two parts – an α helix and a β sheet which are held together by a zinc ion (Figure 1.12B). This type of zinc finger is often found in combination with additional zinc fingers which are arranged in succession so that the α helix of each monomeric unit can interact with the major groove of the DNA. The result is a near continuous stretch of α helices along the major groove of the DNA. Interaction at multiple points ensures that the DNA and protein bind very tightly to each other via a repeating basic structural unit (Fig 1.12A). During evolution, the number of zinc finger repeats can be adjusted thereby tuning the strength and specificity of the DNA-protein interaction. No other TF discussed here has this unique ability to form repeating chains.

[XI] Reproduced from DNA-Binding Motifs in Gene Regulatory Proteins Cover of Molecular Biology of the Cell . 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.



[XII]**Figure 1.12**: A schematic of the Cys-Cys-His-His family of zinc finger proteins. (A) Schematic drawing of a zinc finger showing the amino acid sequence from a frog protein belonging to this class. (B) Schematic showing the three-dimensional structure of this type of zinc finger comprising an antiparallel β sheet followed by an α helix. Amino acids 1 to 10 make up the β sheet and amino acids 12 to 24 make up the α helix. The four amino acids that bind to the zinc atom (Cys 3, Cys 6, His 19, and His 23) hold one end of the α helix tightly connecting it to one end of the β sheet.[30]

1.7.3 The Leucine Zipper motif

The leucine zipper motif is unique in the sense that, unlike other DNA binding motifs described previously, it binds to DNA via a dimer instead of a monomer. Usually, the portion of the protein responsible for the dimeric structure is distinct from the portion that is responsible for DNA recognition and binding. This motif, however, combines these two functions in a simple way. The leucine zipper motif, gets its name because of the way the two α helices provided by each monomer are joined together to form a short coiled-coil. The helices are held together due to strong hydrophobic interactions between the amino acid side chains particularly of leucines extending from one side of either helix. Almost immediately beyond the dimerization interface, the two α helices diverge, forming a Y-shaped structure thus allowing their side chains to interact with the major groove of DNA. The dimer grips the DNA double helix like a pair of tweezers (Figure 1.13).

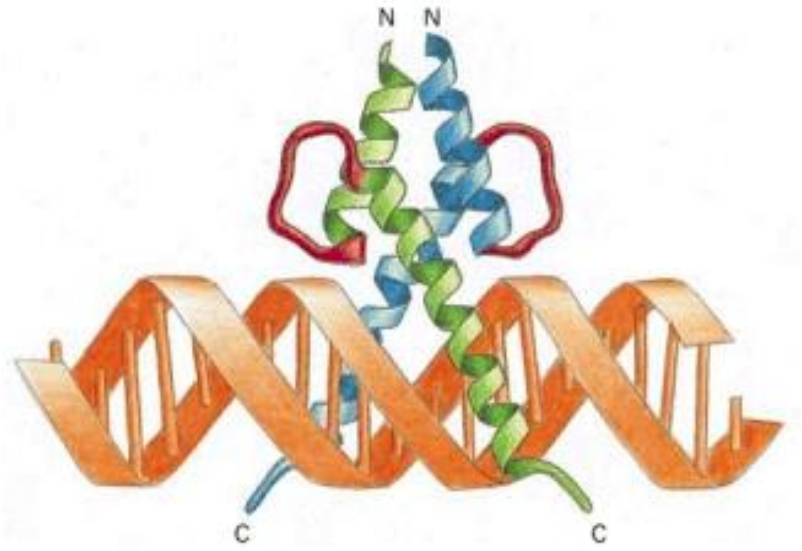
[XII] Reproduced from DNA-Binding Motifs in Gene Regulatory Proteins Cover of Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

[XIII]**Figure 1.13**: A leucine zipper dimer bound to DNA. It is clear from the figure that the bottom part of the two α-helical monomers are involved in DNA-binding whereas the top, α-helical, leucine zipper region allows for dimerization. Each of the two α helices binds to one-half of a symmetric DNA sequence. The structure shown is of the yeast GCN4 protein studied extensively in the subsequent chapters. The GCN4 TF regulates transcription in response to the availability of amino acids in the environment.[31]

1.7.4 Helix-loop-helix motif

The final important DNA-binding motif discussed here and related to the leucine zipper motif, is the helix-loop-helix (HLH) motif. It is to be noted that HLH should not be confused with the HTH discussed in section 1.7.1. An HLH motif consists of a short α helix connected to a second, longer α helix by a loop. The loop provides flexibility allowing one helix to fold back and squeeze against the other. As shown in Figure 1.14, this two-helix structure binds to DNA while at the same time binding to the HLH motif of a second HLH protein. Similar to the case of leucine zipper proteins, the second HLH protein can be the same resulting in a homodimer or different resulting in a heterodimer. In either case, the two α helices extending from the dimerization interface specifically interact with the DNA.

[XIII] Reproduced from DNA-Binding Motifs in Gene Regulatory Proteins Cover of Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.



[XIV]**Figure 1.14**: Schematic showing a helix-loop-helix dimer bound to DNA. The two monomeric units are held together in a four-helix bundle. Two α helices arise from each unit and are connected by a flexible loop of protein (*red*). The HLH is capable of recognizing a specific DNA sequence due to the two α helices that protrude out from the four-helix bundle.[32]

1.7 Recognition of dsDNA in the minor groove – the polyamides

In the previous sections, we have extensively studied the structure of natural dsDNA, the four Watson-Crick base pairs, the importance of DNA and how the information stored in DNA is manipulated in nature through major groove recognition via transcription factors. We now look upon briefly at an entirely different class of molecules, namely the minor groove binders or the polyamides capable of decoding the information stored in dsDNA. Minor groove recognition is also found in nature, although quite rarely. Two noteworthy examples are the natural products Distamycin A composed of three aromatic N-methylpyrrole rings (Figure 1.15A), capable of recognition of the A.T region in double stranded DNA. Distamycin, along with its close relative netropsin (Figure 1.15B), possess anti-viral, anti-tumor, and anti-bacterial properties. Unfortunately, these two molecules proved too toxic for clinical use. They have, however, been studied extensively to understand their base-specific minor-groove DNA-binding.

[XIV] Adapted from DNA-Binding Motifs in Gene Regulatory Proteins Cover of Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

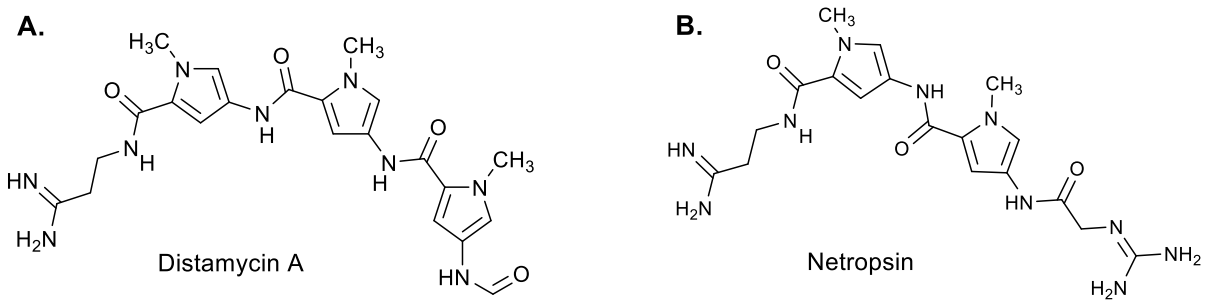


Figure 1.15: Structure of naturally occurring DNA minor groove binders A. Distamycin A and B. Netropsin

Both Distamycin and Netropsin have been shown to bind in the minor groove of A.T rich sequences of DNA in 1:1[33] and 2:1 complexes[34] (Figure 1.16). The charge distribution along DNA is sequence dependant. Uninterrupted A.T base pairs provide the greatest negative potential, whereas successive G.C base pairs give the most positive potential.[35] The differences in the potential in either groove between AT- and GC-rich sequences are due to the specific arrangement of polar groups at the base edges. The difference in potential can contribute significantly to sequence specific DNA recognition.[36] The majority of minor groove DNA binders carry positive charges, which helps explain their increased affinity for A.T rich sequences. Due to the backbone hydrogen atoms, the minor groove wall is also more hydrophobic than the major groove. This also contributes to the sequence selectivity of these binders to the minor groove. Further studies have shown that van der Waals forces are more dominant than initially anticipated.[37]

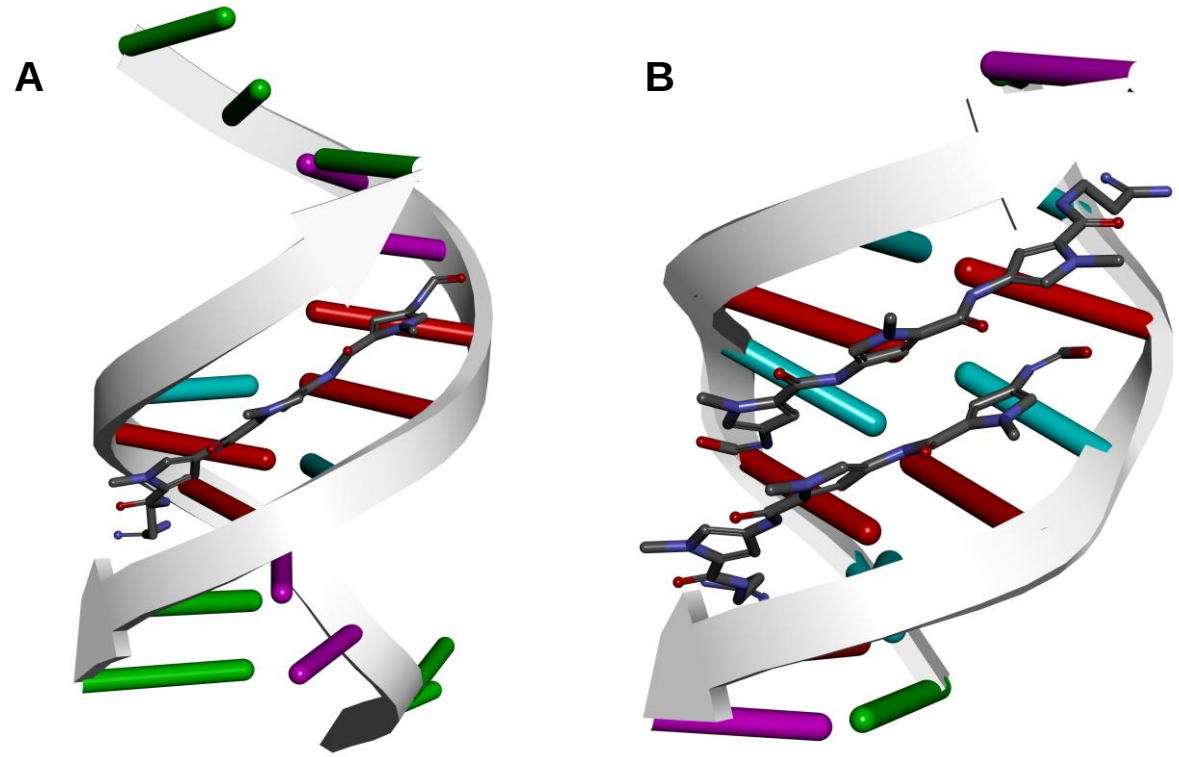
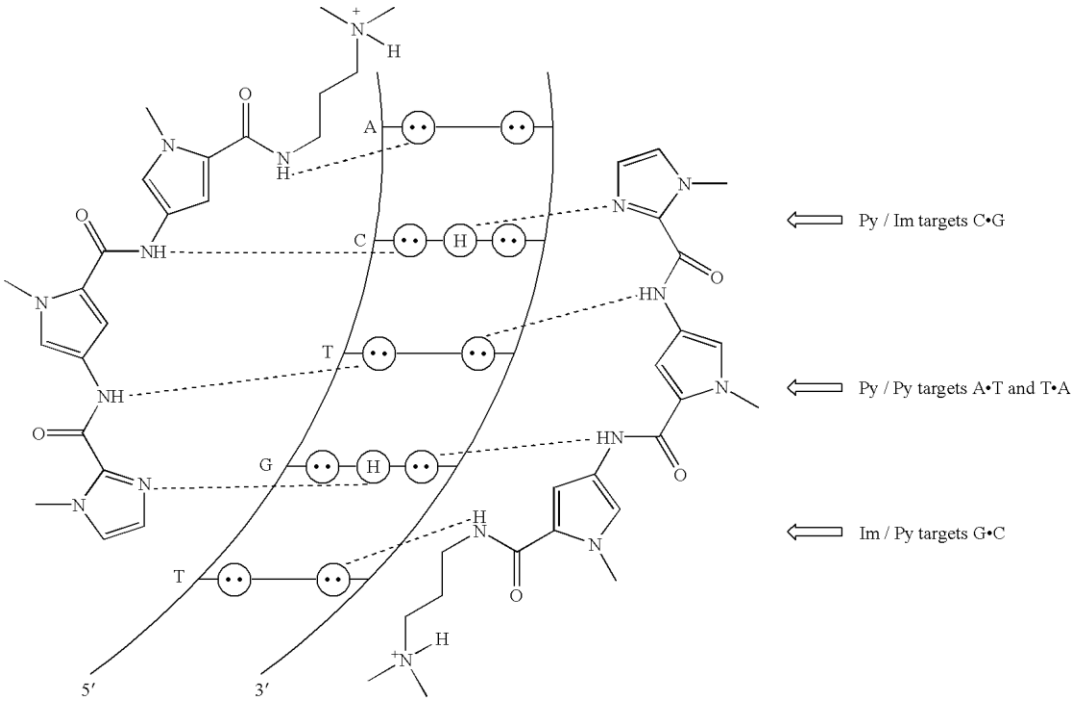


Figure 1.16: (A) Distamycin complexed with 5'-GGCCAATTGG-3' duplex DNA in a 1:1 binding mode. PBID: 1K2Z; DOI: 10.2210/pdb1k2z/pdb. **(B)** Distamycin complexed with 5'-GTATATAC-3' duplex DNA in a 2:1 binding mode. PBID: 378D; DOI: 10.2210/pdb378d/pdb. Images modified using Discovery Studio 4.5.

Many studies involving the NMR and crystal structures of DNA-protein complexes involve the major groove of B-type DNA. The wide, deep surface is rich with hydrogen bonding potential and, therefore, optimum for interactions.[35] In contrast to this, it is now widely accepted that minor-groove interactions generally result in a serious deformation of the DNA. It can be noticed from figure 1.16 that the binding of two Distamycin units has caused the widening of the minor groove and has subsequently distorted the DNA structure. In terms of TF binding, if a polyamide binds to the minor groove adjacent to the major groove to which the TF binds, there will be a detrimental effect on the binding affinity of the TF. This is, in fact, the keystone for cancer therapy using minor groove binders[27].

Inspired by the structure of Distamycin and Netropsin, a series of "synthetic ligands" were developed by Prof. Peter B. Dervan and, therefore, are also known as the Dervan peptides[38]. Chemically, they are N-methylimidazole (Im) and N-methylpyrrole (Py) rings connected with each other via amide bonds. Depending on the sequence of side-by-side pairings, they are capable of differentiating between the A.T, G.C and C.G base pairs present in dsDNA. A pairing between Im opposite Py targets a G.C base pair, whereas a Py opposite Im targets a C.G base pair. A Py-Py combination targets both T.A and A.T base pairs[39] [40] [41] (Figure 1.17).



[xv]**Figure 1.17**: A schematic representation of the polyamide base pairing rules in their simplest form

The specificity of the Py/Im units in case of the G, C base pairs arises from the formation of a hydrogen bond between the imidazole N3 and the exocyclic amino group of guanine[39, 40, 41]. The generality of the pairing rules has been extensively proven over the years by targeting a wide variety of DNA sequences and through several NMR structure studies[42].

[xv] Reproduced from Preparation and use of bifunctional molecules having DNA sequence binding specificity. http://www.google.com/patents/US6555692

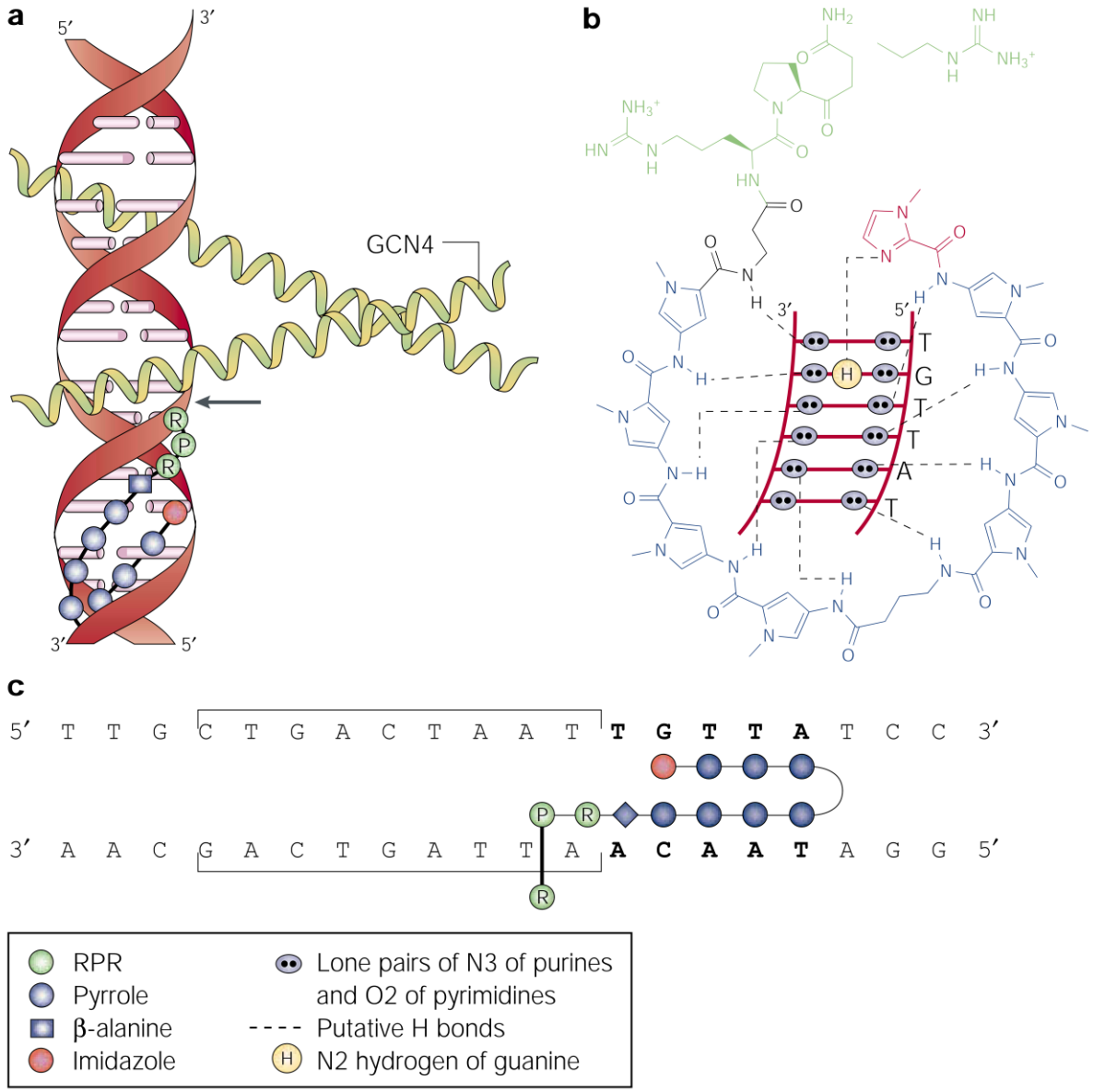
1.8 Blocking transcription factors using polyamides: an example

The introduction of hairpin polyamides capable of recognizing the C.G, A.T & T.A and G.C Watson-Crick base pairs (Figure 1.17) has allowed for the development of new molecular tools and reagents to block transcription factors which is an important strategy for cancer therapy. Figure 1.18 illustrates how the polyamides work through a typical example:

(a) An example of the blocking of GCN4 binding using a polyamide is shown in Figure 1.18a. As the α-helical GCN4 dimer (green-yellow) attempts binding to the adjacent major grooves, there is a steric hindrance between the tripeptide tail (shown in green) and the GCN4 TF preventing simultaneous binding. The point of steric crowding is shown by a black arrow.

(b) The hydrogen-bonding model of the eight-ring hairpin (hairpin denotes the turn) polyamide ImPyPyPy-γ-PyPyPyPy-β-RPR bound to the minor groove of 5'-TGTTAT-3'. γ represents the hairpin and β is short for β – alanine. The RPR tripeptide shown in green is attached to the C-terminus of the polyamide.

(c) The GCN4 binding site is shown in square brackets. The polyamide binding site is adjacent to the GCN4 binding site. The RPR tripeptide protruding into the adjoining major grove allows for blocking of the GCN4 transcription factor.



[XVI]**Figure 1.18**: Mechanism of action of a hairpin polyamide through a typical example.

1.10 Conclusion: DNA – the past, present & future

To conclude this chapter, it is obvious that the discovery of the dsDNA structure, as well as understanding the functions and recognition of dsDNA has played a pivotal role in cancer therapy. Continuing research into TFs and cancer therapy will eventually lead to new breakthroughs, as they have in the past. In the next three chapters of this Ph.D. thesis, we will apply the concepts discussed in this chapter for the synthesis of artificial mimics of TFs in an attempt to develop our own set of molecules capable of recognizing dsDNA.

[XVI] Reproduced from Hurley, L. H. DNA and its associated processes as Targets for Cancer Therapy. *Nat. Rev. Cancer* **2,** 188–200 (2002)

1.11 References:

1. Wolf, G. Friedrich Miescher: The man who discovered DNA. *Chem. Herit.* **21,** 37–41 (2003).
2. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122,** 565–581 (2008).
3. Levene, P. A. The structure of yeast nucleic acid. IV. Ammonia hydrolysis. *J. Biol. Chem.* **40,** 415–424 (1919).
4. Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **2,** 201–209 (1950).
5. Editorial. Due credit. *Nature* **496,** 270 (2013).
6. Watson, J. D., & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953).
7. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry 5th Ed.* (2008).
8. Pierce, B. *Genetics: A Conceptual Approach, 2nd Edition*. (2014).
9. Rich, A., &. Zhang, S. Z-DNA: The long road to biological function. *Nat. Rev. Genet.* **4,** 566–572 (2003).
10. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry, 5th Edition*. (2002).
11. Strauch, M. A. Protein – DNA Complexes : Specific. *Encycl. Life Sci. Nat. Publ. Gr.* 1–7 (2001).
12. Wong, J. M. & Bateman, E. TBP- DNA interactions in the minor between A : T and T : A base pairs groove discriminate. **22,** 1890–1896 (1994).
13. Kielkopf, C. L. *et al.* A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **282,** 111–115 (1998).
14. Alberts, B., Johnson, A. & Lewis, J. *Molecular Biology of the Cell, 4th Edition, New York. Garland Science*. (2002).
15. Blanpain, C. Tracing the cellular origin of cancer. *Nat. Cell Biol.* **15,** 126–34 (2013).
16. Kong, a-N. T. Signal transduction in cancer. *Perspect. Med.* **5,** a006098 (2015).
17. Nebert, D. W. Transcription factors : an overview. *Toxicology* **181–182,** 131–141 (2002).
18. Gibbs, J. B. Mechanism-based target identification and drug discovery in cancer. *Science (80-.).* **287,** 1969–1973 (2000).
19. Tilley, W. D., Clarke, C. L., Birrell, S. N. & Bruchovsky, N. Hormones and cancer: new insights, new challenges. *Trends Endocrinol. Metab* **12,** 186–188 (2001).
20. Barnes, P. J. Anti-inflammatory actions of glucocorticoids: molecular mechanisms. *Clin. Sci.* **94,** 557–572 (1998).
21. Brivanlou, A. H. & Darnell, J. E., J. Signal transduction and the control of gene expression. *Science (80-.).* **295,** 813–818 (2002).
22. Vogt, P. K. Jun, the oncoprotein. *Oncogene* **20,** 2365–2377 (2001).
23. Vogt, P. K., Bos, T. J. & Doolittle, R. F. Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxyl-terminal region of a protein coded for by the oncogene jun. *Proc. Natl Acad. Sci. USA* **97,** 3316–3319 (1987).
24. Derijard, B. et a. JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. *Cell* **76,** 1025–1037 (1994).
25. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411,** 355–365 (2001).
26. Denhardt, D. T. Oncogene-initiated aberrant signaling engenders the metastatic phenotype: synergistic transcription factor interactions are targets for cancer therapy. *Crit. Rev. Oncog* **7,** 261–291 (1996).
27. Hurley, L. H. DNA and its associated processes as Targets for Cancer Therapy. *Nat. Rev. Cancer* **2,** 188–200 (2002).

28. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10,** 252–263 (2009).
29. Yang, V. W. Issues and Opinions in Nutrition Eukaryotic Transcription Factors : Identification, Characterization. *J. Nutr.* 2045–2051 (1998).
30. Lee, M. S., Gippert, G. P., Soman, K. V, Case, D. A. & Wright, P. E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245,** 635–7 (1989).
31. Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted Helices : Crystal Structure of the Protein-DNA Complex. *Cell* **71,** 1223–1237 (1992).
32. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363,** 38–45 (1993).
33. Mitra, S. N., Wahl, M. C. & Sundaralingam, M. Structure of the side-by-side binding of distamycin to d(GTATATAC) 2. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 602–609 (1999).
34. Pelton, J. G. & Wemmer, D. E. Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 5723–5727 (1989).
35. Neidle, S. DNA minor-groove recognition by small molecules (up to 2000). *Nat. Prod. Rep.* **18,** 291–309 (2001).
36. Rohs, R. *et al.* Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79,** 233–69 (2010).
37. Sauers, R. R. An analysis of van der Waals attractive forces in DNA-minor groove binding. *Bioorganic Med. Chem. Lett.* **5,** 2573–2576 (1995).
38. Trauger, J. W., Baird, E. E. & Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentration. *Nature* **382,** 559–561 (1996).
39. Wade, W. S., Mrksich, M. & Dervan, P. B. Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114,** 8783–8794 (1992).
40. Mrksich, M. *et al.* Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci.* **89,** 7586–7590 (1992).
41. Wade, W. S., Mrksich, M. & Dervan, P. B. Binding affinities of synthetic peptides, pyridine-2-carboxamidonetropsin and 1-methylimidazole-2-carboxamidonetropsin, that form 2:1 complexes in the minor groove of double-helical DNA. *Biochemistry* **32,** 11385–11389 (1993).
42. Dervan, P. B. & Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **13,** 284–299 (2003).

CHAPTER 2

TOWARDS SEQUENCE SELECTIVE CROSS-LINKING TO dsDNA USING FURAN CONTAINING POLYAMIDES

In chapter 1, we have extensively discussed the major groove dsDNA binders (Transcription factors) and also briefly discussed the minor groove dsDNA binders (polyamides). In subsequent chapters, we will discuss more about the mimics of transcription factors developed in the OBCR group at UGent under the guidance of Prof. Dr. Annemieke Madder. In this chapter, we will discuss our initial efforts to develop the polyamides into probes for sequence selective cross-linking to dsDNA. The lab in UGent has considerable expertise in cross-linking via furan chemistry[1] *using combinations involving DNA*[2, 3]*, peptides*[4, 5]*, RNA*[6] *and PNA*[7]*. On the other hand, the research group at Lincoln University under Dr. Ishwar Singh has developed extensive expertise in the design and development of polyamide-based minor groove binders*[8, 9]*. A collaborative effort as a part of this joint PhD, is to create a series of minor groove binding DNA modifiers through design and synthesis of minor groove recognizing polyamides equipped with suitable furan moieties for DNA crosslinking.*

2.1 Introduction

Chapter 1 covered the interaction of proteins and DNA which has always been of great interest for a number of reasons. These interactions hold the key for decoding the mechanisms of essential functions such as DNA replication, transcription, and degradation. Developing drugs that can block processes like transcription are highly challenging due to very high selectivity and affinity that DNA binding proteins have for their target, even in a complex cellular environment.[10] In order to decode a protein-DNA interface, a common approach has been to covalently link proteins and DNA using cross-linking moieties, resulting in a covalently linked complex allowing analysis via crystal structures. For this purpose, several photoactivatable functionalities[11, 12, 13, 14] such as phenylazides, diazirines, and benzophenones have been used to form DNA−protein cross-links (DPCs). Unfortunately, there are several limitations to these systems. First, there is a major lack of selectivity, as the generated reactive species such as nitrenes in case of phenylazides, carbenes in case of diazirines and a bi-radical in case of benzophenone can target any amino acid. This makes the identification of the amino acid residues responsible for selective recognition at the DNA−protein binding interface almost impossible. Secondly, due to the occurrence of high-energy intermediates, the resulting final structure of the cross-linked molecule can be quite different from the actual structure, in the absence of the cross-link[5]. Thirdly, obtained DPC yields are, in general, rather low (10−20%)[15]. Finally, these approaches render the selective targeting of natural dsDNA virtually impossible, as they require the modification of one or more of the nucleotide base/bases. An excellent review has been published by Madder et. al. explaining in detail the methodology and applications of furan based cross-linking[1]. In this section, we will briefly discuss some of the aspects of the furan based chemistry developed at UGent relevant to this thesis.

2.2 An overview of the furan cross-linking technique developed in the OBCR group

Furan is a small heterocycle, normally stable under standard atmospheric conditions. However, if ingested, furan is known to undergo metabolic oxidation into butenedial by the enzyme cytochrome P450.[16] Butenedial is highly reactive towards nucleophiles such as amines, which are present in the DNA bases cytosine, adenine and guanine and also in amino acid side chains of lysine (amine) and

thiols present in proteins containing cysteines making it toxic for living tissue. Although, in general, a reaction is possible with A, C and G, we only observe this with A & C due to specific positioning of the furan moiety in the major groove. The relative inertness of furan before oxidation and its high reactivity after oxidation was one of the reasons the OBCR group at UGent under the guidance of Prof. Annemieke Madder ventured into the world of furan chemistry.

It all began more than ten years ago, when an article was published by the Madder group which describes the first efforts towards the development of a furan cross-linking strategy[17]. The article was focused mainly on DNA interstrand cross-linking. The strategy described in the article is to "*first synthesize a furan containing oligonucleotide strand followed by selective in situ oxidation of the furan moiety to a reactive enal species in the presence of a complementary DNA strand which gives rise to a fast and efficient interstrand cross-link"* (Figure 2.1).

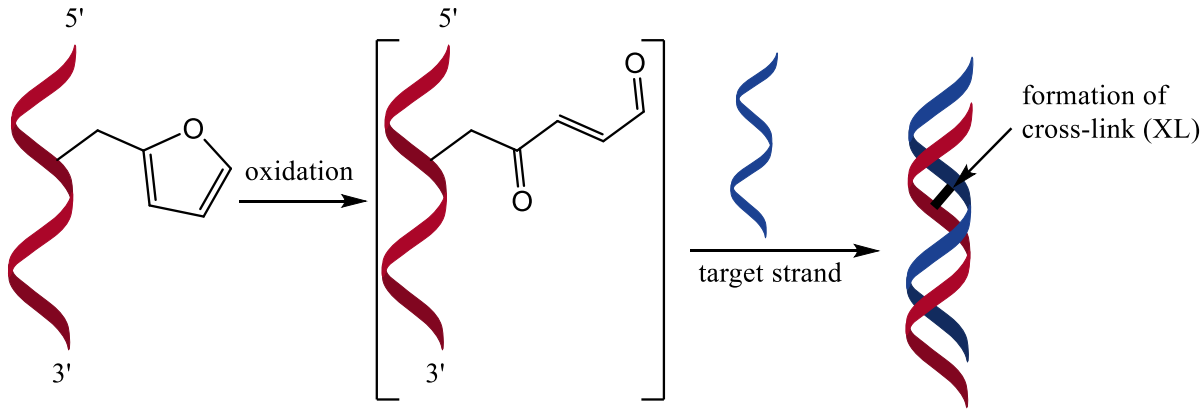
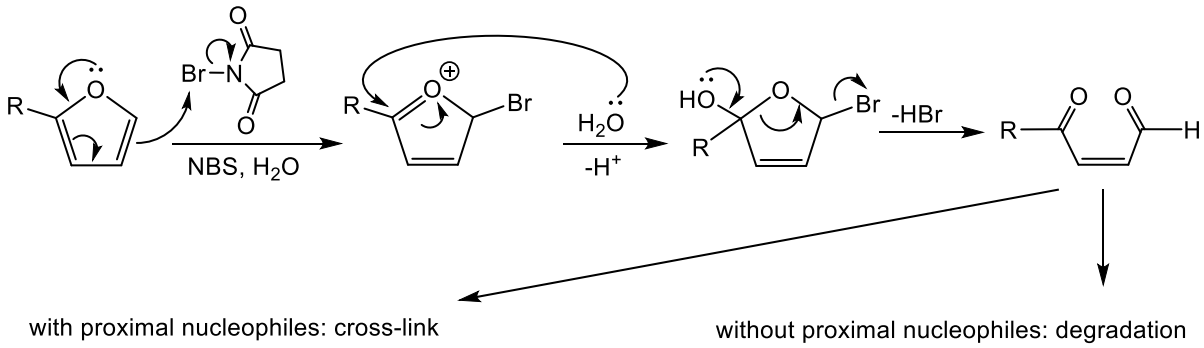
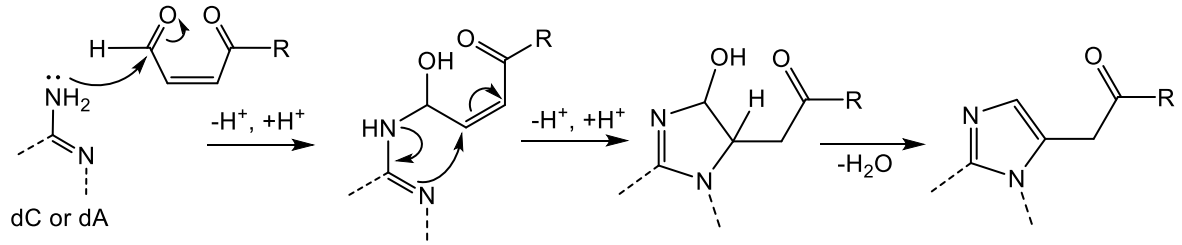


Figure 2.1: Schematic representation of the proposed cross-linking strategy as described in ref. [17]

Over the years, the furan cross-linking methodology has been developed considerably owing to the synthesis of various furan containing building blocks[18, 19, 20, 21] and via the improvement of techniques for selective furan oxidation[3]. Synthetically, oxidation of furan incorporated in DNA[19] and RNA[6] using NBS followed by cross-linking (Figure 2.2) has been the most widely used. Bromination of the furan ring can be an unwanted side-product, particularly when furan is conjugated to an aromatic system (Figure 2.3). The synthesis and incorporation of furan modified building blocks into DNA sequences are less relevant for this thesis and therefore will not be discussed. However, the mechanisms of furan oxidation established through this work are quite interesting and relevant and will be discussed as part of this chapter.



A. With DNA bases dC, dA or dG



B. With a lysine side-chain

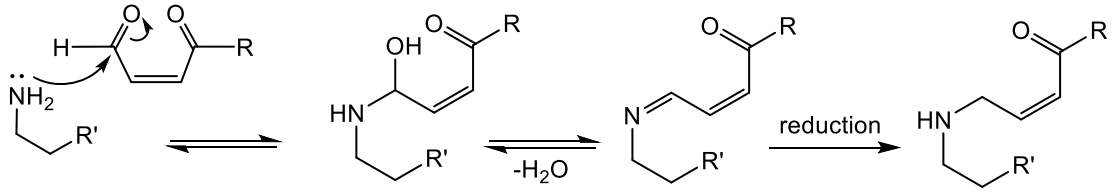


Figure 2.2: Mechanism of furan oxidation using N-Bromosuccinimide (NBS) resulting in the reactive aldehyde which undergoes cross-linking in the presence of proximal nucleophiles such as **A.** the amines of DNA bases dC, dA or dG and **B.** with a lysine side-chain. Adapted from ref. [1]

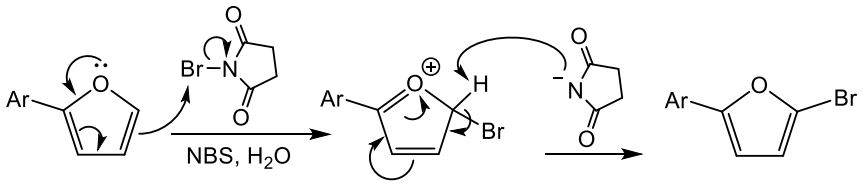


Figure 2.3: Formation of a brominated side reaction occurring in case of a substitution at the C2 position of furan. Adapted from ref. [1]

Recent developments with more biocompatible singlet oxygen strategy (Figure 2.4) through the use of photosensitizers such as methylene blue and Rose bengal and visible light[3] allows for an even wider application of furan based cross-linking. Using $^1O_2$ with organic solvents without a subsequent reduction step results in products containing an extra oxygen as shown in figure 2.4 (A) **a-c** [22]. In aqueous solution, however, furan oxidation using $^1O_2$ yields 4-oxo-enal (**d**) type products (as in the case of NBS/H_2O described previously).

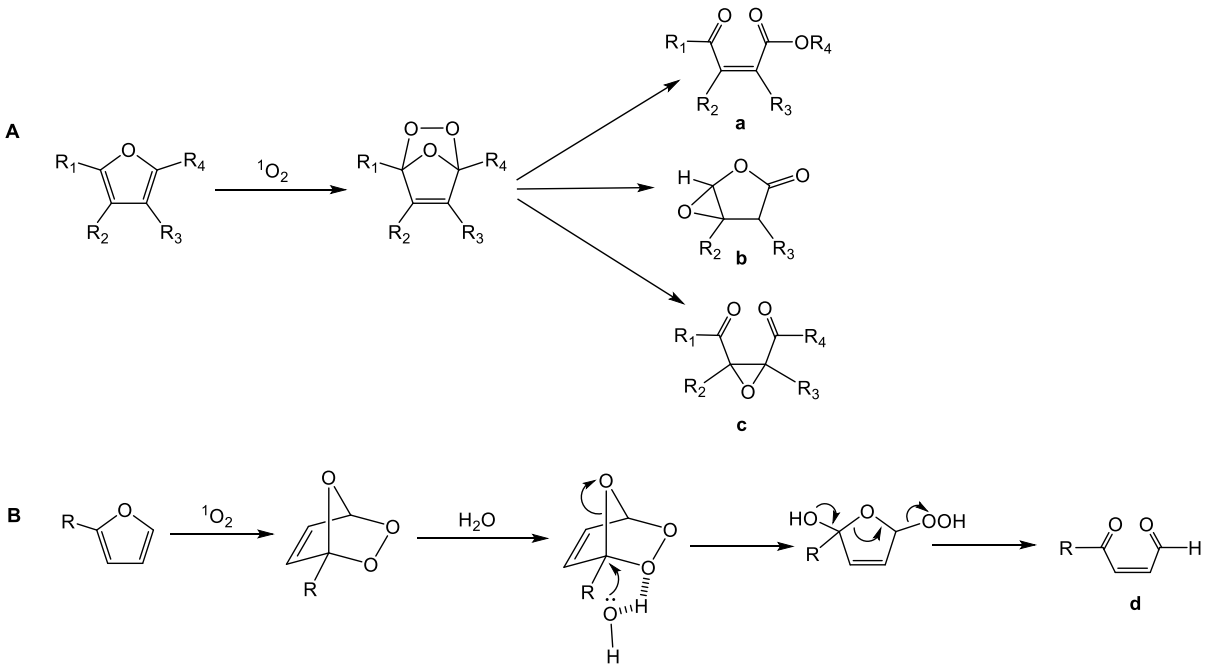


Figure 2.4: Overview of possible products resulting from singlet oxygen oxidation in (**A**) MeOH and (**B**) H_2O. Adapted from ref. [3]

It is well established that furan incorporation and subsequent oxidation in peptides is straightforward, as the furan ring is stable under the conditions for peptide synthesis. Moreover, various furan derivatives are commercially available or can be synthesized in a straightforward manner (Figure 2.5), allowing for the synthesis of different furan-containing polyamides. Considerable work has been done at UGent involving furan and nucleic acids and it has been proven that selective activation by oxidation to generate a reactive aldehyde is possible and can be carried out after duplex formation. This allows for minimization of formation of by products and enables site-selective cross-link formation[2, 3, 17, 23]. High yields and selectivity thus make the furan-oxidation methodology a general procedure to achieve DNA inter-strand cross-linking[24]. Further application of the furan-oxidation methodology has been demonstrated by Summerer, et. al. wherein genetically modified proteins containing a furan-amino acid have been cross-linked to RNA[25]. In this chapter, we attempt to expand the scope of the furan-oxidation cross-linking methodology even further to target natural, dsDNA in order to achieve a much broader application of furan-based probes. Unfortunately, furan cross-linking from any given molecule to dsDNA does not normally occur[1], as the nucleophiles present in dsDNA essential for a cross-link are involved in stable Watson-Crick base pairing which is inherent in the structure of natural dsDNA.

Figure 2.5: Furan building blocks for incorporation into peptides: I. 3-(furan-2-yl)propanoic acid. II. Fmoc-L-2-furylalanine.

2.3 Local disruption of dsDNA base-pairing

In earlier work at OBCR it was attempted to use furan-modified triplex-forming oligonucleotides (TFOs) to achieve cross-linking to dsDNA[26]. It was however shown in that study that the Watson-Crick base pairing within dsDNA, as opposed to ssDNA, significantly reduces the nucleophilic capacity of the exocyclic amine on the nucleobases to attack the aldehyde functionality generated upon furan oxidation. Therefore, in order to achieve a cross-link between polyamides and dsDNA, a local disruption between the base pairs must be externally introduced at the point of the cross-link. For this purpose, bulky amino acids such as racemic 2-([1,1'-biphenyl]-4-yl)-2-aminoacetic acid or L--α-cyclohexylglycine (Figure 2.6) can be used.[27] These amino acids are known to locally disrupt the DNA duplex, which should allow for a cross-link to be formed between an oxidized furan moiety and an unpaired base.

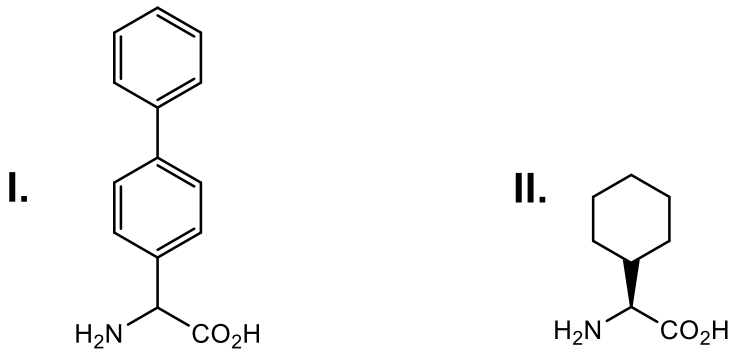


Figure 2.6: Amino acid building blocks for incorporation into polyamides: I. racemic 2-([1,1'-biphenyl]-4-yl)-2-aminoacetic acid. II. L-α-Cyclohexylglycine.

2.4 Using furan containing polyamides for cross-linking to dsDNA

As discussed in section 1.7, hairpin polyamides (or Dervan peptides) can distinguish all four base pairs combinations (GC, CG, AT and TA) in the minor groove of DNA.[28] Moreover, they bind only to dsDNA with very high affinities (their K_d is in the low nM range, even at room temperature)[29] and specificities[30] and can be designed to read DNA sequences up to 16 base pairs in length[31]. The protocols for the synthesis of hairpin polyamides have been greatly optimized over the years, making them synthetically accessible via standard solid phase synthesis.[9, 32] Mimics of transcriptions factors possessing these qualities, on the other hand, are complex and synthetically challenging.[33, 34] There are examples where polyamides have successfully been used to cross-link to dsDNA[35] and in continuation of our efforts towards simplified transcription factor models with high affinity for dsDNA we decided to combine the expertise areas of UGent and Lincoln in the development of furan modified polyamides as potential alternative cross-linking minor groove binders.

2.5 Design and synthesis of polyamides

A challenge to be overcome, if cross-linking to DNA using Dervan peptides is to be made feasible, was the incorporation followed by selective oxidation of the furan ring in presence of the Py/Im rings. The initial part of this project was to obtain conditions under which furan could be selectively oxidised in the presence of the aromatic Py/Im rings. Selective oxidation of furan in case of peptides containing aromatic Trp and Tyr residues has been achieved in the lab in UGent (unpublished results).

However, oxidation of furan in case of polyamides proved to be tricky (Table 2.1). The following polyamides were synthesised initially for testing furan oxidation:

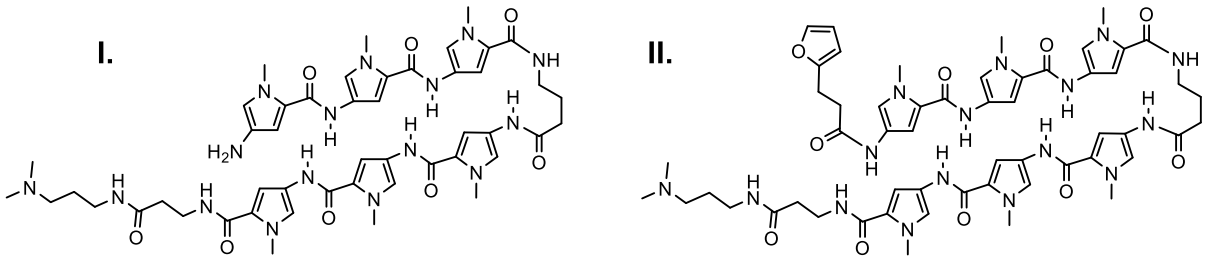


Figure 2.8: Polyamide **I** was designed as a control. Polyamide **II** was the furan containing polyamide.

For initial testing, only Py containing polyamides were chosen. Polyamide **I** was chosen as the control and does not contain any furan moiety. This allowed us to check for degradation of the Py rings when tested against the oxidation conditions described in Table 2.1. Polyamide **II** contains an N-terminal furan. An N-terminal furan was specifically chosen as it was less sterically hindered and, therefore, was expected to be more susceptible to oxidation.

2.6 Summary of results

Sr. No.	Reagent	Eq.	Time	Results (analysis using MALDI-MS)
1.	mCPBA	5	2h	No oxidation/degradation of polyamides
2.	mCPBA	50	6h	Trace oxidation of polyamide **II**
3.	NBS	2	1h	Oxidation/minor degradation of both polyamides
4.	Rose Bengal + $^1O_2$ (generated using white light)	0.02	30s	>80% degradation of both polyamides
5.	Rose Bengal + $^1O_2$ (generated using blue light)	0.02	30s	>80% degradation of both polyamides

Table 2.1: Summary of the best results obtained using different oxidizing agents, equivalents and times for furan oxidation on polyamide **I** & **II** (Figure 4.4)

2.7 Conclusion

Based on the experiments conducted thus far (refer experimental section for details) it can be concluded that use of $^1O_2$ resulted mainly in degradation of the polyamide. mCPBA on the other hand is too mild resulting in poor yields of the desired oxidized product. NBS gave the best results, however, optimization of the experimental conditions is required. At this stage of the project, in order to avoid orthogonality problems such as the sensitivity of the polyamide to oxidation/degradation while trying to selectively oxidize the furan, we decided to focus on an alternative non-polyamide based major groove binding peptide approach which we will discuss in chapters 3 & 4. Nonetheless, conditions suitable for the oxidation of peptides containing sensitive residues such as Tyr, Trp, His and Met are currently being developed at UGent on a related project. These conditions could be modified and used for selective furan oxidation in polyamides.

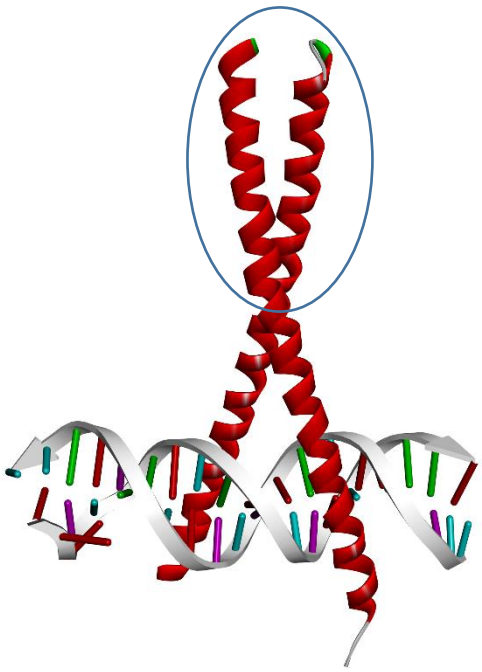
2.8 References:

1. Carrette, L. L. G., Gyssels, E., De Laet, N. & Madder, A. Furan oxidation based cross-linking: a new approach for the study and targeting of nucleic acid and protein interactions. *Chem. Commun.* **52,** 1539–1554 (2015).
2. De Beeck, M. O. & Madder, A. Unprecedented C-selective interstrand cross-linking through in situ oxidation of furan-modified oligodeoxynucleotides. *J. Am. Chem. Soc.* **133,** 796–807 (2011).
3. Beeck, M. Op De & Madder, A. Sequence Specific DNA Cross-Linking Triggered by Visible Light. *J. Am. Chem. Soc.* **134,** 10737–10740 (2012).
4. Hoogewijs, K., Buyst, D., Winne, J. M., Martins, J. C. & Madder, A. Exploiting furan's versatile reactivity in reversible and irreversible orthogonal peptide labeling. *Chem. Commun.* **49,** 2927–9 (2013).
5. Carrette, L. L. G., Morii, T. & Madder, A. Toxicity Inspired Cross-Linking for Probing DNA-Peptide Interactions. *Bioconjug. Chem.* **24,** 2008–14 (2013).
6. Carrette, L. L. G., Gyssels, E., Loncke, J. & Madder, a. A mildly inducible and selective cross-link methodology for RNA duplexes. *Org. Biomol. Chem.* **12,** 931–5 (2014).
7. Manicardi, A., Gyssels, E., Corradini, R. & Madder, A. Furan-PNA: a mildly inducible irreversible interstrand cross-linking system targeting single and double stranded DNA. *J. Am. Chem. Soc.* **52,** 6930–6933 (2016).
8. Singh, I. *et al.* Sequence-selective detection of double-stranded DNA sequences using pyrrole-imidazole polyamide microarrays. *Journal of the American Chemical Society* (2013).
9. Fallows, A. J., Singh, I., Dondi, R., Cullis, P. M. & Burley, G. A. Highly Efficient Synthesis of DNA-Binding Polyamides Using a Convergent Fragment-Based Approach. *Org. Lett.* **16,** 4654–4657 (2014).
10. Verzele, D., Carrette, L. L. G. & Madder, A. Peptide scalpels for site-specific dissection of the DNA-protein interface. *Drug Discov. Today Technol.* **7,** e115–e123 (2010).
11. Tate, J. J., Persinger, J. & Bartholomew, B. Survey of four different photoreactive moieties for DNA photoaffinity labeling of yeast RNA polymerase III transcription complexes. *Nucleic Acids Res.* **26,** 1421–1426 (1998).
12. Winnacker, M., Breeger, S., Strasser, R. & Carell, T. Novel diazirine-containing DNA photoaffinity probes for the investigation of DNA-protein-interactions. *Chembiochem* **10,** 109–18 (2009).
13. Shigdel, U. K., Zhang, J. & He, C. Diazirine-based DNA photo-cross-linking probes for the study of protein-DNA interactions. *Angew. Chemie - Int. Ed.* **47,** 90–93 (2008).
14. Zhu, G. & Lippard, S. J. Photoaffinity labeling reveals nuclear proteins that uniquely recognize cisplatin-DNA interstrand cross-links. *Biochemistry* **48,** 4916–4925 (2009).
15. G. Dolinnaya, N. Design and Synthesis of 2-Functionalised Oligonucleotides. Their Application for Covalent Trapping the Protein – DNA Complexes. *Curr. Org. Chem.* **13,** 1029–1049 (2009).
16. Peterson, L. A. Electrophilic intermediates produced by bioactivation of furan. *Drug Metab. Rev.* **38,** 615–26 (2006).
17. Halila, S., Velasco, T., Clercq, P. De & Madder, A. Fine-tuning furan toxicity: fast and quantitative DNA interchain cross-link formation upon selective oxidation of a furan containing oligonucleotide. *Chem. Commun.* 936–938 (2005). doi:10.1039/b415092a
18. Stevens, K. & Madder, A. Furan-modified oligonucleotides for fast, high-yielding and site-selective DNA inter-strand cross-linking with non-modified complements. *Nucleic Acids Res.* **37,** 1555–1565 (2009).

19. Beeck, M. Op De & Madder, A. Unprecedented C-Selective Interstrand Cross-Linking through in Situ Oxidation of Furan-Modified Oligodeoxynucleotides. *J. Am. Chem. Soc.* **133,** 796–807 (2011).
20. Stevens, K. *et al.* Furan-oxidation-triggered inducible DNA cross-linking: Acyclic versus cyclic furan-containing building blocks - On the benefit of restoring the cyclic sugar backbone. *Chem. - A Eur. J.* **17,** 6940–6953 (2011).
21. Jawalekar, A. M., Op de Beeck, M., van Delft, F. L. & Madder, A. Synthesis and incorporation of a furan-modified adenosine building block for DNA interstrand crosslinking. *Chem. Commun.* **47,** 2796–2798 (2011).
22. Wilkinson, F. & Brummer, J. G. Rate constants for the decay and reactions of the lowest electronically excited singlet state of molecular oxygen in solution. *J. Phys. Chem. Ref. Data* **10,** 809 (1981).
23. Stevens, K. & Madder, A. Furan-modified oligonucleotides for fast , high-yielding and site-selective DNA inter-strand cross-linking with non-modified complements y. *Nucleic Acids Res.* **37,** 1555–1565 (2009).
24. Carrette, L. L. G., Gyssels, E. & Madder, A. DNA interstrand cross-link formation using furan as a masked reactive aldehyde. *Curr. Protoc. Nucleic Acid Chem.* **54,** Unit 5.12. (2013).
25. Schmidt, M. J. & Summerer, D. Angewandte Red-Light-Controlled Protein – RNA Crosslinking with a Genetically Encoded Furan. *Angew. Chemie* **52,** 4690–4693 (2013).
26. Gyssels, E., Carrette, L. L. G., Vercruysse, E., Stevens, K. & Madder, A. Triplex Crosslinking through Furan Oxidation Requires Perturbation of the Structured Triple-Helix. *ChemBioChem* **16,** 651–658 (2015).
27. Singh, I., Hecker, W., Prasad, A. K., Parmar, V. S. & Seitz, O. Local disruption of DNA-base stacking by bulky base surrogates. *Chem. Commun.* **5,** 500–1 (2002).
28. Dervan, P. B. & Bürli, R. W. Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.* **3,** 688–693 (1999).
29. Trauger, J. W., Baird, E. E. & Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentration. *Nature* **382,** 559–561 (1996).
30. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391,** 468–71 (1998).
31. Trauger, J. W., Baird, E. E. & Dervan, P. B. Recognition of 16 Base Pairs in the Minor Groove of DNA by a Pyrrole-Imidazole Polyamide Dimer. *J. Am. Chem. Soc.* **120,** 3534–3535 (1998).
32. Su, W., Gray, S. J., Dondi, R. & Burley, G. a. Highly efficient synthesis of DNA-binding hairpin polyamides via the use of a new triphosgene coupling strategy. *Org. Lett.* **11,** 3910–3 (2009).
33. Mosquera, J., Jiménez-Balsa, A., Dodero, V. I., Vázquez, M. E. & Mascareñas, J. L. Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides. *Nat. Commun.* **4,** 1874 (2013).
34. Ruiz García, Y. *et al.* Sequence-selective DNA recognition and enhanced cellular up-take by peptide–steroid conjugates. *Chem. Commun.* **51,** 17552–17555 (2015).
35. Wang, Y. D. *et al.* DNA crosslinking and biological activity of a hairpin polyamide-chlorambucil conjugate. *Nucleic Acids Res.* **31,** 1208–1215 (2003).

CHAPTER 3

THE GCN4 TRANSCRIPTION FACTOR AND AN OVERVIEW OF THE SYNTHETIC MIMICS DEVELOPED IN THE OBCR GROUP AT UGENT



The GCN4 Transcription Factor (TF) is arguably one of the most studied transcription factors. Although this transcription factor is a yeast amino acid metabolism regulatory transcription factor[1] controlling the activation of several genes in response to amino acid starvation, it is of particular importance especially to synthetic chemists due to its well-defined dimerization domain (encircled in blue) and DNA recognition region. Its crystal structure bound to dsDNA was determined by Ellenberger et. al. in 1992[2]. This makes it one of the relatively easy TFs to study and mimic, which in turn allows chemists to make a diverse range of synthetic mimics of the GCN4 TF – a strategy which could eventually be used in cancer therapy by targeting TFs in humans[3].

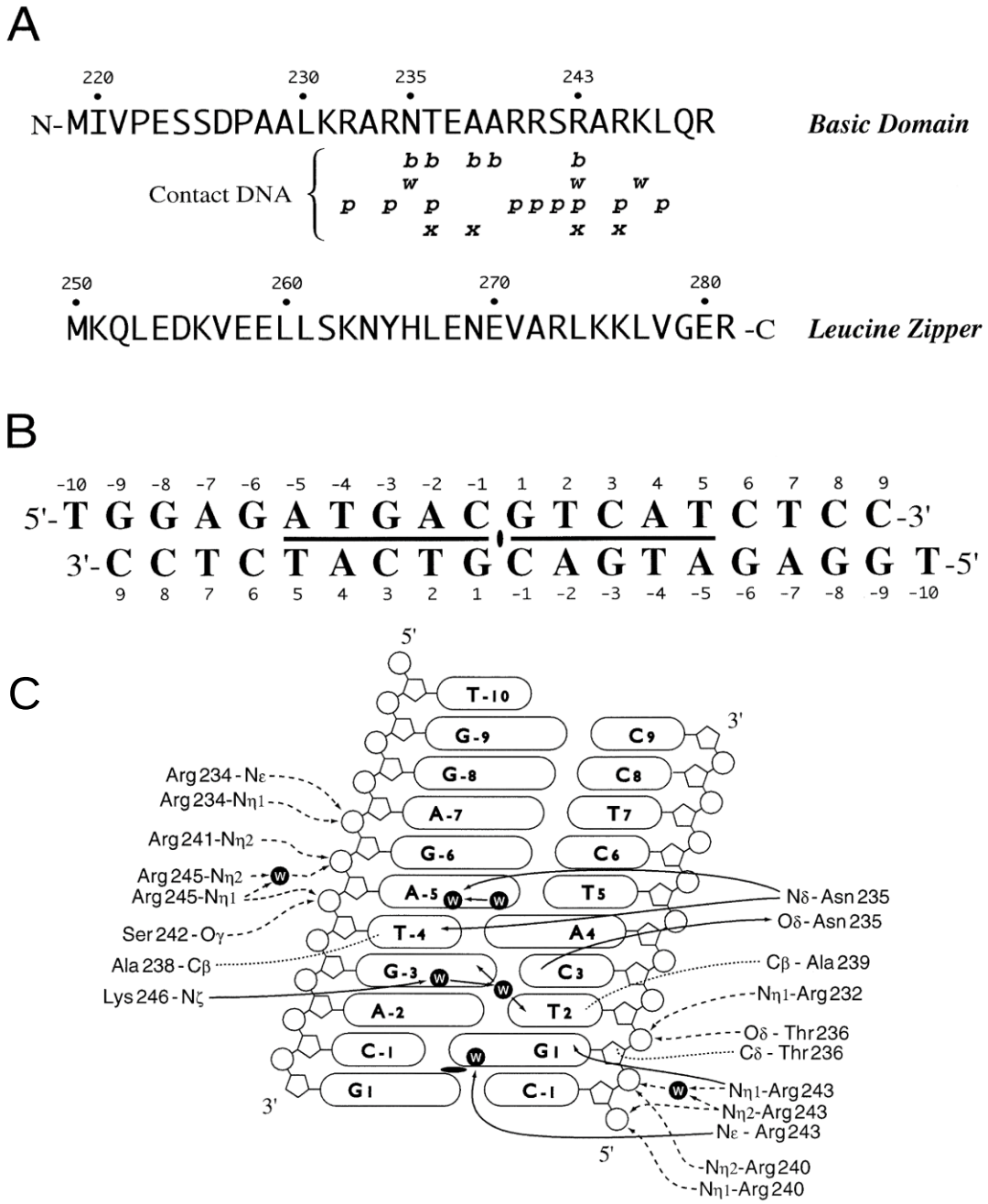
3.1 Introduction

In the previous chapter, we studied some of the families of TFs. In this chapter, we will study one model transcription factor in particular – the yeast TF GCN4 (General Control Nonderepressible). The GCN4 TF belongs to a family of TFs known as the basic region Leucine Zipper (bZIP). Over 30 known eukaryotic proteins contain the bZip DNA-binding motif. They are known as the bZIP because of the unique and well defined geometry of the bZIP region which firmly holds the two helices in place. The bZIP also gives these TFs a very high degree of specificity towards their respective DNA sequences by restricting not only the free movement of the helices with respect to each other, but also their rotation along a Z-axis. Genes containing the DNA sequence 5'-ATGACTCAT-3' are activated by the GCN4 TF. The GCN4 binds to its target DNA as a dimer which is critical for DNA binding. However, the TF is known to form stable dimers even in the absence of DNA[4].

3.2 Analysis of the primary & secondary structures of the GCN4 TF bound to its target DNA

The GCN4 TF is a relatively large protein comprising of 281 amino acids with a molecular weight slightly over 31kDa[5]. The full length sequence, although known, is of little importance. The amino acids forming the two important domains – the dimerization domain and the basic recognition region (Figure 3.1A) are, however, of much interest. The very first crystal structure of the GCN4 TF bound to DNA was solved with a resolution of 2.9 Å by Ellenberger in 1992[2]. Subsequently, Keller

improved it further, obtaining a resolution of 2.2 Å in 1995 for the same TF bound to DNA and containing the ATF/CREB recognition sequence. Crystallographic data is freely available at Protein Data Bank (pdb: 1YSA). A complete set of interactions is described in figure 3.1.



[I]Figure 3.1. The protein and DNA sequences of the GCN4-DNA complex crystallized and complete interaction summary.

A: GCN4-bZIP domain numbered according to the full-length sequence. Residues that make one or more DNA contacts are indicated: b, direct to base; w, via water to base; p, direct to phosphate; x, via water to phosphate.

B: Self-complementary, 18 bp palindromic ATF/CREB DNA with terminal thymidine. A molecular and crystallographic 2-fold axis passes through the C −1G1 base step. The specific recognition sequence is marked with a bar.

C: Protein-DNA interaction summary for one basic domain and DNA half-site. The location of the 2-fold axis is indicated. Direct and water (W) mediated interactions with bases are shown as continuous lines, interactions

[I] Reproduced from J. Mol. Biol. (1995) 254, 657-667

with the phosphate backbone as broken lines and hydrophobic interactions as dotted lines. Arrows indicate the direction of hydrogen bonds.

The secondary structure of the GCN4 TF is quite well defined and is a perfect example of a homodimer. The bZIP dimer is a pair of continuous α helices that form a parallel coiled coil over the C-termini of the recognition regions. The recognition region is around 30 residues and gradually diverges towards the N-termini while simultaneously passing through the major groove of the DNA-binding site. The interactions between the bases and phosphate groups present in DNA and the basic residues present in the protein constitute the main contacts of this region between the DNA and the protein.

The leucine zipper region of the GCN4 TF is stabilized due to the hydrophobic and non-polar interactions within the monomer forming an extensive van der Waals contact surface. Intra and inter-helical salt bridges are also present. Although these are lower in number, they also contribute to the formation of the dimer.

The basic region of each monomer adopts the optimal conformation to bind to its respective half-site of the target DNA sequence. Although there are no sharp bends or kinks in either GCN4 monomer, the rigid coiled coil imposes a gentle bend of the α-helix towards the axis of the DNA. Consequently, it is able to grip the major groove like a pair of tweezers and interact specifically with the base pairs located there. The contacts of this region with the major groove are based on the positively charged polar residues namely the arginines and lysines, which establish hydrogen bonds with the oxygen atoms of the phosphate backbone and the base pairs. The N-terminal end of the basic region continues as a helix beyond the point of DNA contact and is the most flexible part of the protein.

3.3 Mimics of the GCN4 TF developed in the OBCR Group

As explained previously, the GCN4 TF is relatively easy to mimic due to its well-defined dimerization domain and basic recognition region. Synthetic chemists from several research groups have exploited this aspect and have designed and synthesized artificial mimics of the GCN4 TF[6]. The most noteworthy among them are the contributions from Mascarenhas[7, 8, 6], Morii[9, 10, 11], Schepartz[12, 13] and Kim[14]. From literature, it can be concluded that the minimum basic recognition domain can be reduced to the following 23 amino acids: DPAALKRARNTEAARRSRARKLQ.

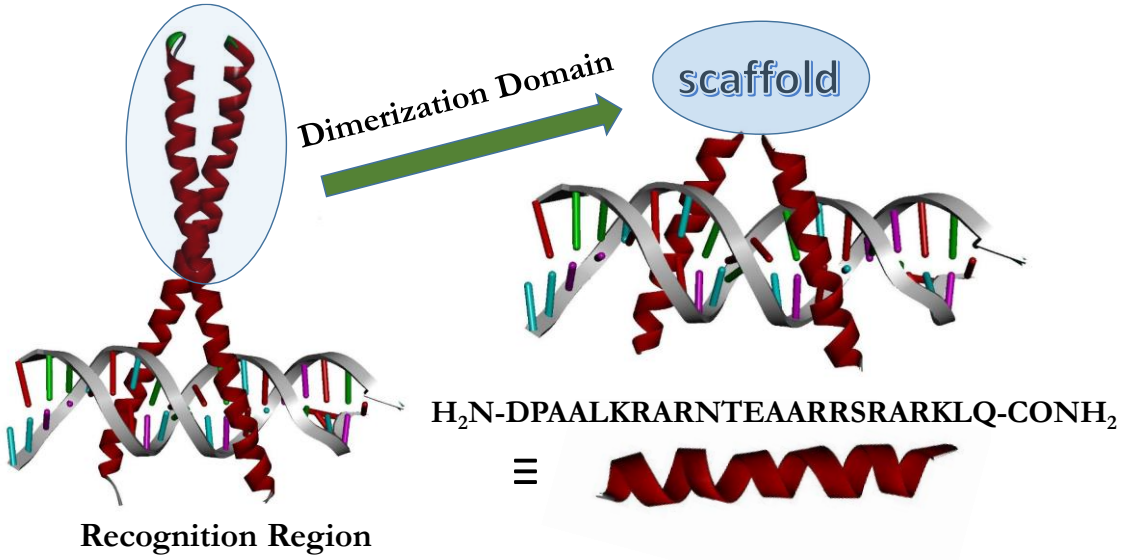


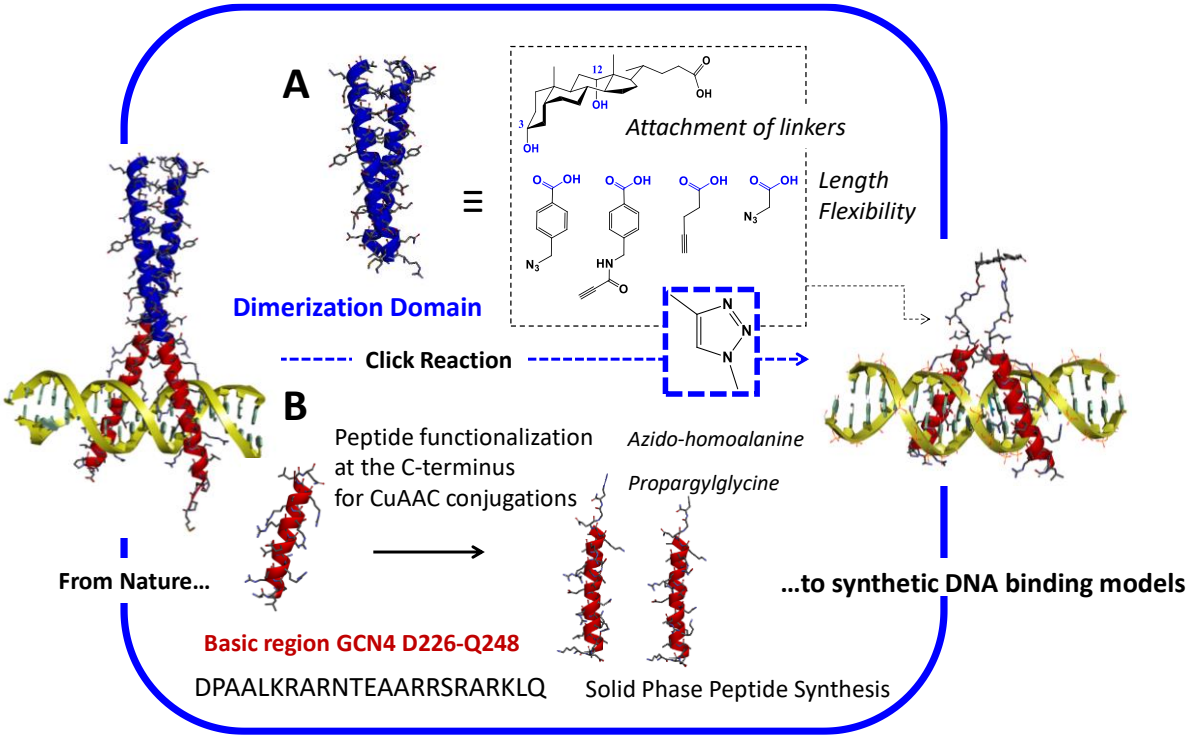
Figure 3.2: General schematic showing the most common strategy for the synthesis of artificial mimics of the GCN4 TF using the minimum basic recognition region and substituting the dimerization domain (encircled in blue) with a scaffold.

The Organic & Biomolecular Research Group (OBCR) at UGent has also been actively involved in phosphate recognition, especially focussing on the development of artificial DNA binders for the GCN4 TF. The initial work was started by Dieter Verzele[15] and was continued by Lieselot Carrette[16].

Yara Ruiz and myself further developed this work through the synthesis of new mimics of the GCN4 TF. The initial synthetic mimics were designed by replacing the dimerization domain of the GCN4 TF using rigid, modified deoxycholic acid steroid[17] and bi-functional cyclodextrin[18] scaffolds. Subsequently, as part of this PhD thesis, simplified and synthetically less-challenging monomeric mimics were designed and synthesized using stapled peptides[19]. All the constructs developed were also tested for cellular uptake as part of the master thesis of Dorien Van Lysebetten. The work done using stapled peptides has been described in detail in the next chapter. In this chapter, we will discuss, in brief, the constructs based on the dimerization of the GCN4 TF using functionalized steroid and cyclodextrin scaffolds.

3.4 DNA recognition and enhanced cellular uptake using peptide-steroid conjugates and cylcodextrin-peptide conjugates for DNA binding

This design, syntheses and purification of the final peptide-steroid conjugates was carried out by Yara Ruiz as part of her PhD thesis[17] along with myself and my former master thesis student Dorien at UGent who was also responsible for performing the biological studies. The overview of the synthesis is briefly described in this section. The complete work in the form of publications can be found in Annex I.



[II]Figure 3.3: Overview of the design and syntheses of artificial DNA binders through a convergent approach using a Copper Azide-Alkyne Cycloaddition (CuAAC) reaction involving: (A) Replacement of the dimerization

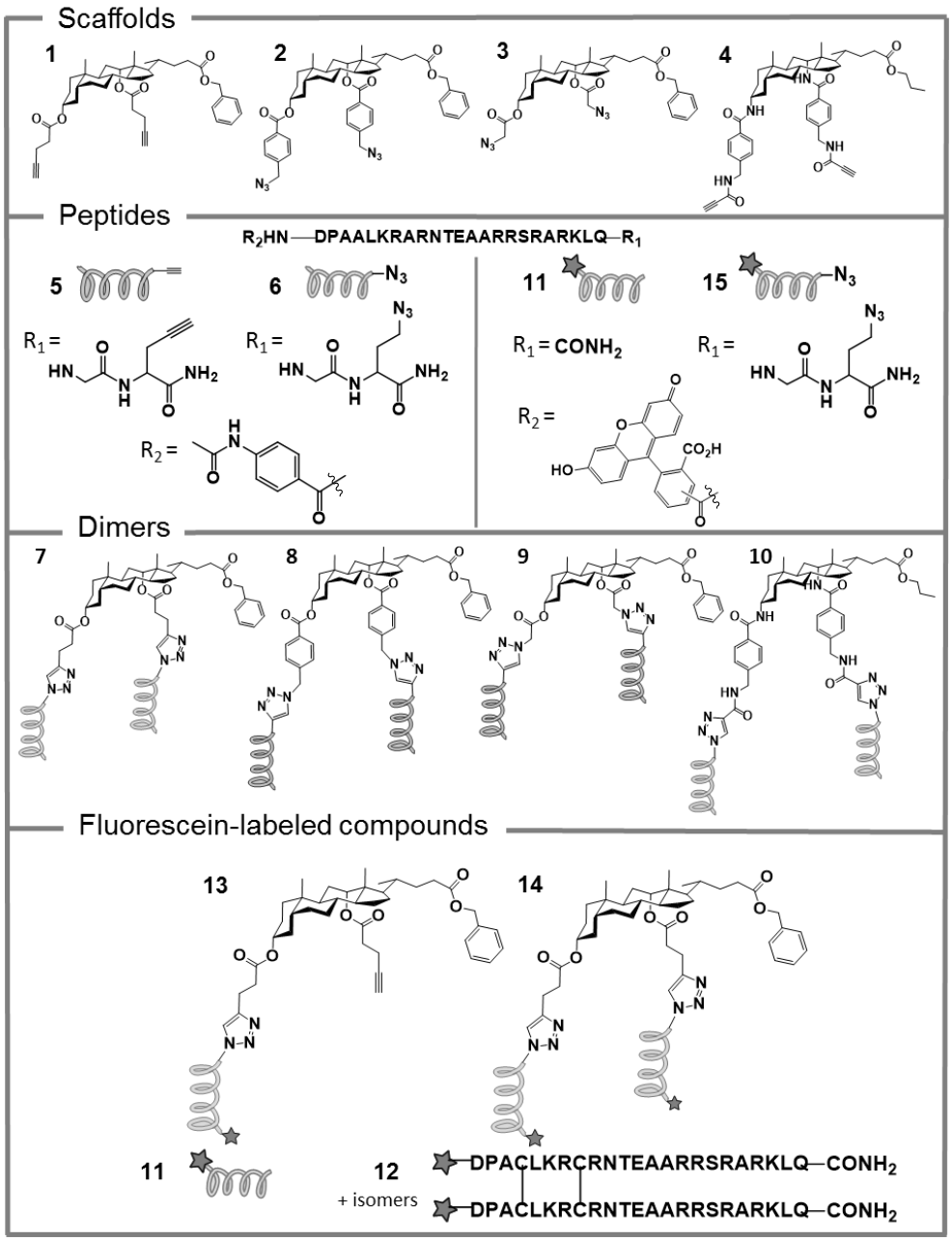
[II] Modified & reproduced from Ruiz García, Y. Iyer A. *et al.* Sequence-selective DNA recognition and enhanced cellular uptake by peptide–steroid conjugates. *Chem. Commun.* **51,** 17552–17555 (2015)

domain with a steroid scaffold equipped with a suitable linker and (B) The basic DNA recognition region peptide containing an azide or alkyne functionality at the C-terminus for conjugation to the modified scaffold.

Initial attempts to synthesize a mimic of the GCN4 TF using a linear and direct solid phase peptide synthesis approach were successful[16] but resulted in the final product with low yields and purity. Therefore, a new, convergent approach was designed as shown in figure 3.2. The dimerization domain was once again substituted with a rigid steroid scaffold possessing a suitable linker, but unlike the previous approach the scaffold and peptide were synthesized separately and conjugated via a CuAAC reaction.

Four different scaffolds with functionalised linkers capable of forming cycloaddition products (Figure 3.3: Scaffolds **1**-**4**) were designed and synthesized for peptide dimerization of the GCN4 basic region. Although the four scaffolds are deoxycholic acid derivatives, they possess linkers of varying lengths, flexibilities and functionalities (Fig. 3.3). In order to attach different linkers, the alcohol positions were modified using Steglich esterification which yielded the final scaffolds (**1**-**3**). The linkers chosen for the study were pentynoic acid, azido glycine, 4-azidomethyl-benzoic acid and (N-propynoylamino)-p-toluic acid (PATA). The PATA linker possess an active alkyne and has been used previously for preparation of peptide–oligonucleotide conjugates via CuAAC[20]. Due to the high reactivity of the PATA linker it was necessary to synthesize the diamino derivative of deoxycholic acid[21] as esterification resulted in byproducts.

For synthesis of the peptides, the unnatural amino acids propargyl glycine or azido homoalanine possessing an alkyne and azide functionality respectively were first loaded onto the resin to provide an attachment point for the peptides to the steroid scaffold. Peptides **5** and **6** (Figure 3.4) were synthesized using automated SPPS on Rink Amide Chemmatrix resin suitable for longer, difficult peptides. Four different transcription factor models (**7**-**10**) were synthesized using the peptides **5** & **6**. A DMSO/H_2O = 3:1 combination capable of dissolving both the hydrophobic scaffold and hydrophilic peptide was found to be optimal for the CuAAC reaction. After performing the reaction with different Cu catalysts such as $Cu(OAc)_2$, CuBr, CuI and $CuSO_4$, $Cu(CH_3CN)_4PF_6$ gave the best results. The presence of a large number of guanidine groups probably resulted in a high Cu complexation and therefore excess catalyst was needed to drive the reaction forward. The excess copper ions can be removed after completion of the reaction by using EDTA[20]. The desired product was surprisingly achieved through the use of excess scaffold and not excess peptide as expected. The reaction was complete after stirring the mixture for 3 hours at room temperature with the dipodal construct being the main product and without any major side products being formed despite the presence of unprotected amino acids side chains. The resulting final compounds **7**-**10** were purified using RP-HPLC.



[III]**Figure 3.4:** Complete list of peptides and peptide-steroid conjugates synthesized for DNA binding and cellular uptake.

To investigate the cellular uptake of a peptide-steroid conjugate, the best DNA binding construct **7** was suitably modified and synthesized with a fluorescein tag (Figure 3.4, **14**). The properties of this construct were compared the monomer **11**, the GCN4 peptide dimer **12** and the monomeric steroid conjugate **13**.

Additionally, GCN4 protein mimics were also synthesized by replacing the Leucine Zipper region with α, β and γ-cyclodextrins equipped with azide functionalities[18]. The peptides were conjugated via CuAAC reaction under the same conditions as those described before. The DNA binding studies, cell uptake studies and the published article can be found in Annex I.

[III] Reproduced from Ruiz García, Y. Iyer A. *et al.* Sequence-selective DNA recognition and enhanced cellular up-take by peptide–steroid conjugates. *Chem. Commun.* **51,** 17552–17555 (2015)

3.5 Conclusion:

In conclusion, we have described in detail the structure of the GCN4 TF and have subsequently designed and synthesized new artificial DNA binders which are mimics of the TF. This has been achieved using a convergent approach involving the GCN4 basic recognition region peptide and steroid as well as cyclodextrin scaffolds. The dissociation constant (K_D's) for the mimics of the TFs to their native DNA sequence are in the nM range. The synthetic route has been simplified as compared to previous models described in literature. It is the first time relatively long, cell permeable peptide-steroid conjugates have been synthesized and tested for cell uptake. The constructs show enhanced cell uptake even at 4°C and increased membrane interaction. Through this work we are a step closer to understanding the parameters which affect DNA recognition and cell uptake in case of the GCN4 TF.

3.6 References:

1. Pathak, D. & Sigler, P. B. Updating structure-function relationships in the bZip family of transcription factors. **116,** 116–123 (1992).
2. Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted Helices : Crystal Structure of the Protein-DNA Complex. *Cell* **71,** 1223–1237 (1992).
3. Darnell, J. E. Transcription Factors as Targets for Cancer Therapy. *Nature* **2,** 740–749 (2002).
4. Hopel, I. A. & Struhl, K. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6,** 2781–2784 (1987).
5. Saccharomyces Genome Database. at <http://www.yeastgenome.org/locus/S000000735/overview>
6. Pazos, E., Mosquera, J., Vazquez, M. E. & Mascarenas, J. L. DNA Recognition by Synthetic Constructs. *Chembiochem* **12,** 1958–1973 (2011).
7. Mosquera, J., Jiménez-Balsa, A., Dodero, V. I., Vázquez, M. E. & Mascareñas, J. L. Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides. *Nat. Commun.* **4,** 1874 (2013).
8. Caaman, A. M., Va, M. E., Castedo, L. & Mascaren, J. L. Design and Synthesis of a Peptide That Binds Specific DNA Sequences through Simultaneous Interaction in the Major and in the Minor Groove. 4723–4725 (2001).
9. Morii, T., Simomura, M. & Morimoto, S. Sequence-Specific DNA Binding by a Geometrically Constrained Peptide Dimer. *J. Am. Chem. Soc.* **32,** 1150–1151 (1993).
10. Morii, T., Tanaka, T., Sato, S., Hagihara, M. & Aizawa, Y. A General Strategy To Determine a Target DNA Sequence of a Short Peptide : Application to a D -Peptide. **124,** 2001–2002 (2002).
11. Shin-ichi Sato, Masaki Hagihara, Kenji Sugimoto, Takashi, M. Design of new DNA binding peptides Chemical Approaches Untangling Sequence-Specific DNA Binding. *Chem. - A Eur. J.* 5066–5071 (2002).
12. Cuenoud, B. & Schepartz, A. Design of a metallo-bZIP protein that discriminates between CRE and AP1 target sites: Selection against AP1. *Proc. Natl Acad. Sci. USA* **90,** 1154–1159 (1993).
13. Zondlo, N. J. & Schepartz, A. Highly Specific DNA Recognition by a Designed Miniature Protein. *J. Am. Chem. Soc.* **121,** 6938–6939 (1999).
14. Talanlan, R. V, Mcknight, C. J. & Kim, P. S. Sequence-Specific DNA Binding by a Short

Peptide Dimer. *Nature* **33,** 769–771 (1990).
15. Verzele, D. & Madder, A. Synthetic Progress in cMyc-Max Oncoprotein Miniaturization: Semi-On-Line Monitoring gives Solid-Phase Access to Hydrophobic b(-HLH-)ZIP Peptidosteroid Tweezers. *European J. Org. Chem.* 673–687 (2013).
16. Carrette, L. L. G., Morii, T. & Madder, A. Peptidosteroid Tweezers Revisited: DNA Binding Through an Optimised Design. *European J. Org. Chem.* **2014,** 2883–2891 (2014).
17. Ruiz García, Y. *et al.* Sequence-selective DNA recognition and enhanced cellular up-take by peptide–steroid conjugates. *Chem. Commun.* **51,** 17552–17555 (2015).
18. Ruiz García, Y. *et al.* Cyclodextrin–peptide conjugates for sequence specific DNA binding. *Org. Biomol. Chem.* **13,** 5273–5278 (2015).
19. Iyer, A. *et al.* Stapling monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake. *Org. Biomol. Chem.* **13,** 3856–3862 (2015).
20. Wenska, M. *et al.* An activated triple bond linker enables 'click' attachment of peptides to oligonucleotides on solid support. *Nucleic Acids Res.* **39,** 9047–9059 (2011).
21. Verzele, D. & Madder, A. Short Synthesis of Orthogonally Protected 3α,12α-Diamino-5β-cholan-24-oic Acid, a Dipodal Steroid Scaffold for Combinatorial Chemistry. *European J. Org. Chem.* **2007,** 1793–1797 (2007).

CHAPTER 4

SYNTHETIC MIMICS OF THE GCN4 TF USING STAPLED PEPTIDES FOR DNA RECOGNITION AND ENHANCED CELLULAR UPTAKE

Organic & Biomolecular Chemistry

COMMUNICATION

View Article Online
View Journal | View Issue

CrossMark click for updates

Cite this: *Org. Biomol. Chem.*, 2015, **13**, 3856

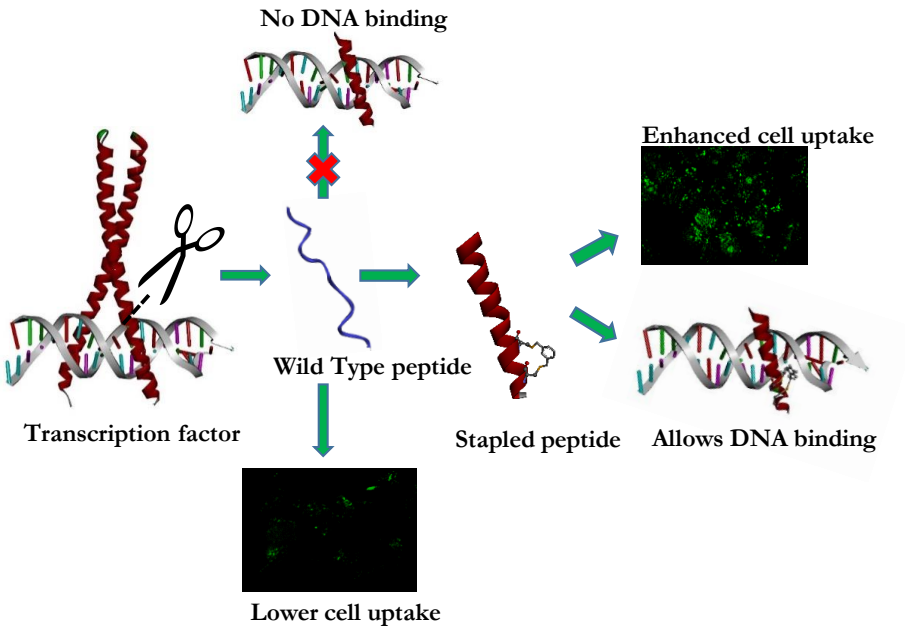
Received 24th December 2014,
Accepted 17th February 2015

DOI: 10.1039/c4ob02659d

www.rsc.org/obc

Stapling monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake†

Abhishek Iyer,[a] Dorien Van Lysebetten,[a] Yara Ruiz García,[a] Benoit Louage,[b] Bruno G. De Geest[b] and Annemieke Madder*[a]



DNA binding and concurrent enhancement of cellular uptake.

The basic DNA recognition region of the GCN4 protein comprising 23 amino acids has been modified to contain two optimally positioned cysteines which have been linked and stapled using cross-linkers of suitable lengths. This results in stapled peptides with a stabilized α-helical conformation which allows for

4.1 Introduction

As explained in Chapters 1 and 3, Transcription Factors (TFs) by definition are responsible for the decoding of genetic information from dsDNA to mRNA.[1] In the quest for sequence selective DNA recognition, TF proteins can be regarded as nature's models to design synthetic DNA binders. TFs are usually classified according to the fold of their DNA-binding domains and grouped into a small number of families, like the bZIP, bHLH, homeodomain, HTH, and zinc fingers which have already been studied in detail.[2] Most transcription factors are large proteins possessing complex secondary structures, rendering it difficult to design smaller synthetically accessible versions thereof retaining

the DNA binding capacity.[3] The GCN4 leucine zipper is, however, a relatively easy protein to mimic due to its well-defined dimerization domain and basic recognition region.

It has already been shown that dimerization of basic region peptides through non-peptide scaffolds can allow DNA binding.[4] However, construction of such smaller dimeric mimics of this TF, although feasible, has proven to be synthetically challenging.[5, 6] Previous attempts at more thorough structural minimization and reduction of complexity using the monomeric GCN4 peptide have shown that DNA binding is greatly reduced due to loss of secondary structure.[7] Indeed, since the basic region of the GCN4 transcription factor cannot adopt a helical fold in solution by itself nor bind to DNA[8] due to entropic reasons,[9] an external factor forcing the peptide into a helical conformation is needed. For this reason, unlike the existing synthetic bZip models in which DNA binding is induced via dimerization,[10, 5] we here aim to stabilize a single α helix via peptide stapling, reasoning that enhancing the helicity within the monomer should sufficiently stabilize the conformation to allow DNA binding.[7] Stapled peptides have been used extensively for improving helicity,[11] increasing cell-penetration,[12, 13] proteolytic stability and enhancing peptide-protein interactions (PPIs).[14] In all the above-mentioned cases, the benefits of stapled peptides have been demonstrated.[15] However, the use of stapled peptides in the miniaturization of zipper proteins has remained largely unexplored.

4.2 Stapled peptides as GCN4 TF mimics & selection of the stapling methodology

In this work, we have examined the DNA binding induced by stapling using the monomeric GCN4 transcription factor basic region as a model peptide, by comparing the i, i+4 and i, i+7 stapling methods and varying the positions of the staple along the helix. The cellular uptake of the constructs was also investigated using fluorescently labelled versions of the peptide.

Initial studies were dedicated to selecting the most suitable method for peptide stapling among the ones reported in literature.[12, 16, 17, 18, 19] The idea was to adopt a method as general as possible to be applicable to any potential DNA binding peptide by increasing its helicity. To ensure easy and cost-effective modification, the use of unnatural amino acids was avoided. The hydrocarbon stapling is generally considered as the best stapling methodology for improving helicity of peptides for subsequent enhancement of peptide-protein interactions. However, the labor intensive preparation of the required modified amino acid building blocks, coupling of the non-natural amino acids as well as peptide folding on resin is a cause for concern when stapling a peptide through this method.[20] Therefore, in the current study, we opted for cysteine cross-linking,[15] in view of the commercial availability and/or easy synthesis of cross-linking moieties, mild reactions conditions and easy scalability due to synthesis in solution.

4.3 Design and Synthesis of stapled peptides

A detailed analysis of the essential contacts for binding of the GNC4 protein to its cognate DNA sequence, as derived from the reported crystal structure, was described earlier by Ellenberger (pdb file: 1YSA).[21] Various dimeric peptides based on the D226-Q248 basic region of the GCN4 TF have been shown to retain their DNA binding properties.[22, 10]

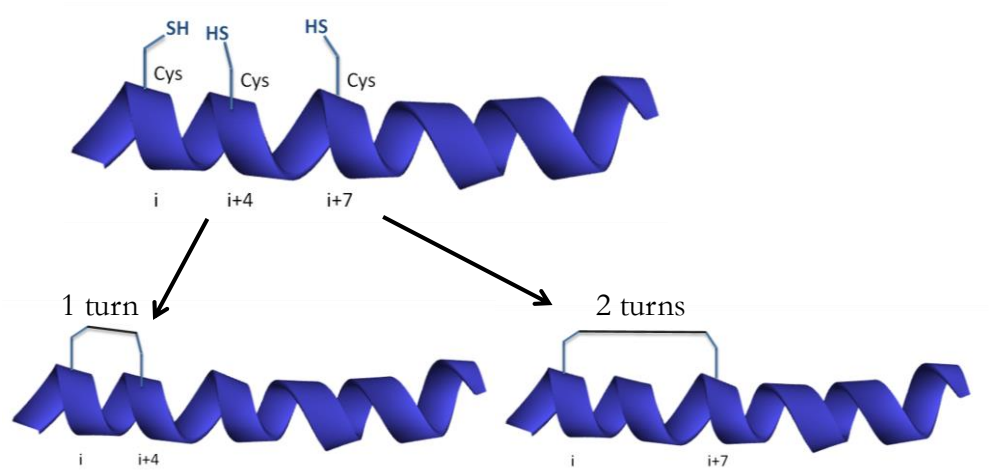


Figure 4.1: Schematic showing the stapling possibilities – the i, i+4 peptide stapling capable of stabilising a single helix turn and the i, i+7 peptide stapling capable of stabilising two helix turns.

Based on these studies, amino acids within this sequence, indicated as not involved in DNA contacts,[23] were identified and systematically replaced by Cys according to an i,i+4 or i,i+7 format (Figure 4.1). These positions were specifically chosen as they stabilise a peptide helix. Ideally, there are 3.6 amino acids per turn of a peptide helix. However, since it is not possible to stabilise the 3.6th amino acid we chose the nearest integer values, 4 (for one turn of the helix) and 7 (for two turns of the helix). Furthermore, molecular modelling aided visualisation based on the pdb file 1YSA[21] was used to ensure that the introduced linkers point away from the DNA and not towards it, thereby avoiding any steric repulsion which may arise due to peptide stapling. In this way, three different peptides, comprising the D226-Q248 sequence from the DNA binding basic region of GCN4, containing a double Cys substitution (at positions 237/244 for **1**, 229/233 for **2** and 233/237 for **3**) were synthesized on solid support, cleaved and subsequently treated with various linkers yielding a series of five stapled peptides as shown in Figure 4.2.

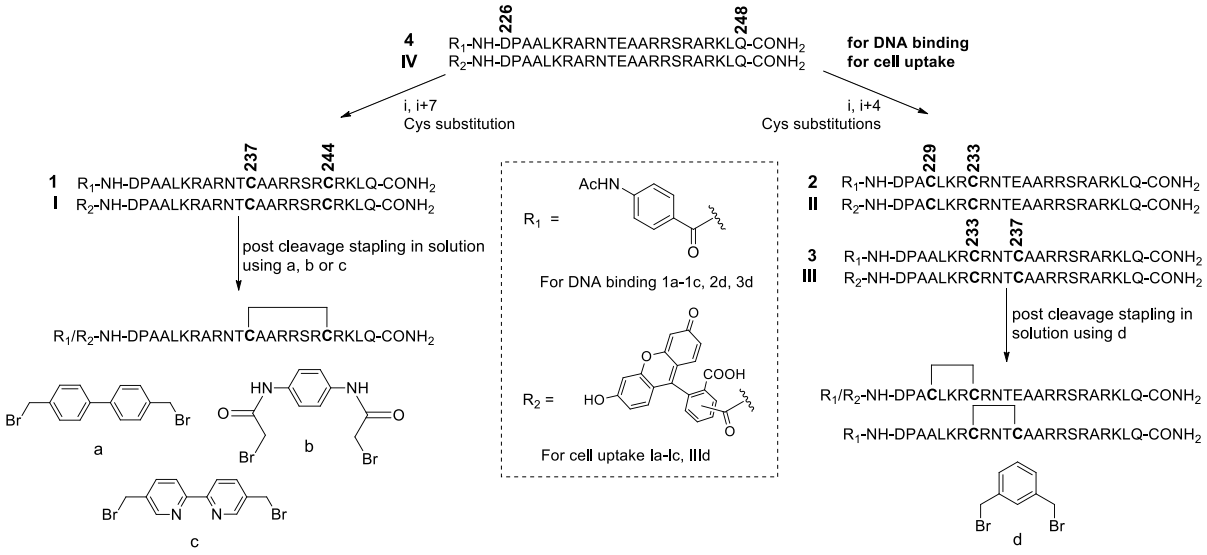


Figure 4.2: Synthetic stapled peptides 1a-c, 2d, 3d for DNA binding and unmodified basic region peptide 4. The peptides are N-terminally capped with a p-acetylamino benzoic acid (ABA) moiety to ensure UV-based detection and analysis in the case of DNA binding. For cell uptake studies, the ABA moiety was replaced by fluorescein.

The stapled peptides and the unmodified WT basic region peptide were successfully synthesized. The cross-linkers **a, c** & **d** are commercially available. The stapling moiety **b** has not been used thus far for peptide stapling and was designed and synthesized as a more polar alternative to the biphenyl and bipyridine cross-linkers.

4.4 DNA binding studies

Next, the DNA binding capacity of all peptides was evaluated through electrophoretic mobility shift assay (EMSA) titration of various peptide concentrations to the DNA sequence 5' – CGG ATG ACG TCA TTT TTT TTC – 3' (Figure 4.3) containing the cognate monomeric GCN4 binding site GTCAT. The WT peptide **4** does not bind DNA under the given EMSA conditions. For all the synthetic constructs **1a-c, 2d** & **3d,** we see enhanced DNA binding as compared to peptide **4**. Earlier, DNA binding with monomeric peptides was observed through the use of a grafting strategy, whereby the crucial contact residues of the GCN4 binding region are specifically positioned on an avian pancreatic polypeptide.[24] Apparently, our simple and straightforward stapling strategy also allows constraining the peptide into a suitable conformation for DNA binding. In the case of peptides, **1a-c** binding is only observed at higher concentrations of peptide as compared to **2d** and **3d**. The more hydrophobic nature of the biphenyl cross-linker makes peptide **1a** more susceptible to aggregation (followed by precipitation), as can be seen from the complete disappearance of the bands in the last lane. On the other hand, peptides **1b** and **1c** are less prone to aggregation, but the onset of binding occurs at higher concentrations than in the case of **1a**. Compared to peptides **1a-c**, the DNA binding pattern in gel **2d** is different. We propose that the appearance of two bands related to DNA-peptide complex formation is due to the presence of two binding sites in the CRE sequence (…GTCAT…).

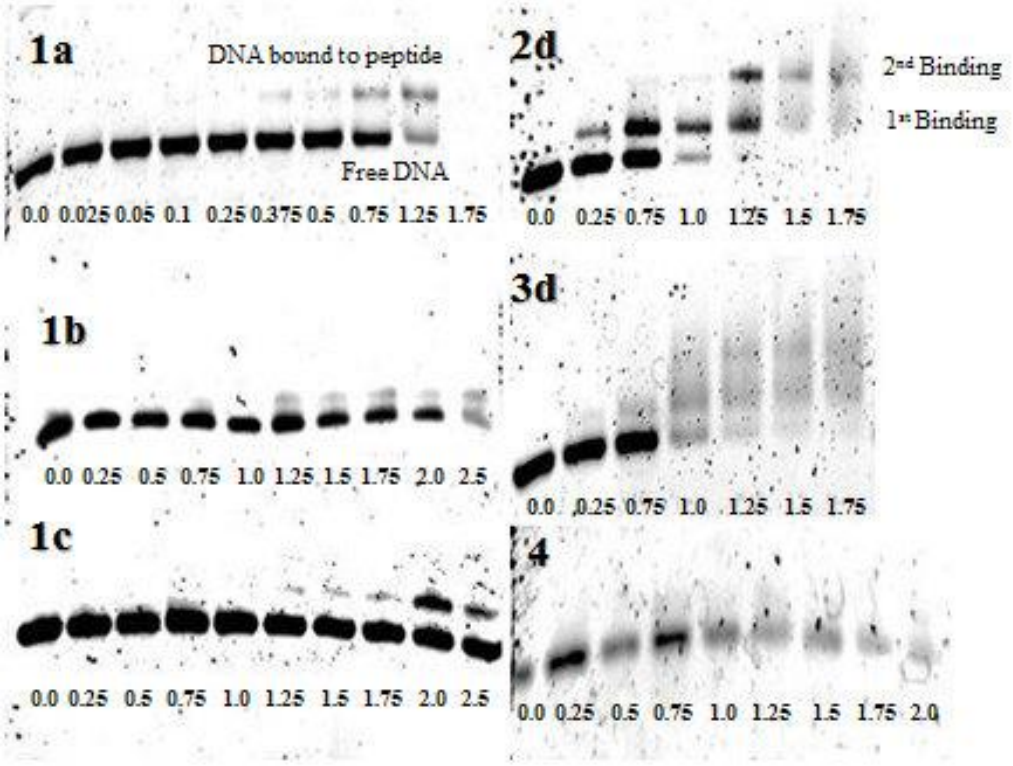


Figure 4.3: EMSA titrations for peptides 1a-c, 2d & 3d. Loading mixture comprises 5 µL from a mixture of 10 µL mQ, 4 µL sucrose, 2 µL loading buffer, 2 µL DNA, 2 µL peptide resulting in a total DNA concentration of 167 nM. The loading buffer consists of 20 µL Tris 1 M, pH = 7.6, 20 µL KCl 0.2 M, 20 µL $MgCl_2$ 0.1 M, 40 µL EDTA 0.025 M. Peptide concentrations from left to right (in µM) are indicated below each gel.

It can be noticed that after full occupation of one binding site by the peptide, remaining non-bound peptide can interact with the second binding site. The binding pattern of peptide **2d** is unique and only visible in the case of the two sequences that bind at a lower concentration range. It can further be noticed that peptide **3d** shows a similar binding pattern as peptide **2d** but suffers from non-specific interactions, which cause aggregation, bands getting blurred or disappearance of all bands.

From the experiments described in literature[7, 25] and from our own data, a general conclusion can be made regarding DNA binding stapled peptides. The major challenge in case of DNA binding peptides is that, increasing the helicity has to be complemented with a degree of flexibility in order to account for the conformational change which occurs when the peptide binds to DNA. Since this conformational change is more significant in the case of DNA binding peptides as compared to PPIs, a too-tight locking of the peptide into a helical conformation may result in the inability to achieve DNA binding. Indeed, our data show that peptide stapling by providing an N or C terminal helix stabilization, rather than centrally in the sequence, gives better results in terms of DNA binding, as observed from the binding pattern of peptides **2d** & **3d** versus **1a-c**. Though in the absence of DNA a low helical content is observed (see CD measurement results in SI), apparently the conformation can be adjusted into a structure whereby the contacts between the positively charged side chains, mainly involving the Lys and Arg residues, and the negatively charged backbone of the DNA can be maximized without a high entropic penalty. We believe peptide **2d** fits these criteria and hence is the best DNA binder from the constructs synthesized. Moreover, peptide **2d** is able to bind in a dimeric fashion without having been artificially dimerized.

4.5 Cell uptake & toxicity studies

Dorien Van Lysebetten, Benoit Louage and Prof. Bruno De Geest are thanked for their valuable contribution.

The previously related yeast derived GCN4 peptide (231-252) has been classified as a membrane permeable peptide.[26, 27] It varies slightly from the basic region peptide used in this article, which was specifically chosen for its DNA binding abilities in its dimeric form.[22] It has further been postulated that stabilization of the secondary structure via peptide stapling can enhance cell uptake[28]. Peptides with hydrocarbon staples in particular have shown considerable increase in uptake, as compared to their non-stapled counterparts[13,29]. Here, the fluorescently labelled versions of peptides **Ia-c** & **IId** were tested in a cellular environment using RAW 264.7 mouse macrophages. Confocal microscopy confirmed that cell uptake is achieved even at a low concentration of 0.25 µM for all peptides at 37°C using an incubation time of 3h, including native GCN4 sequence **IV**, as can be seen from fluorescence of fluorescein in figure 3A-E.

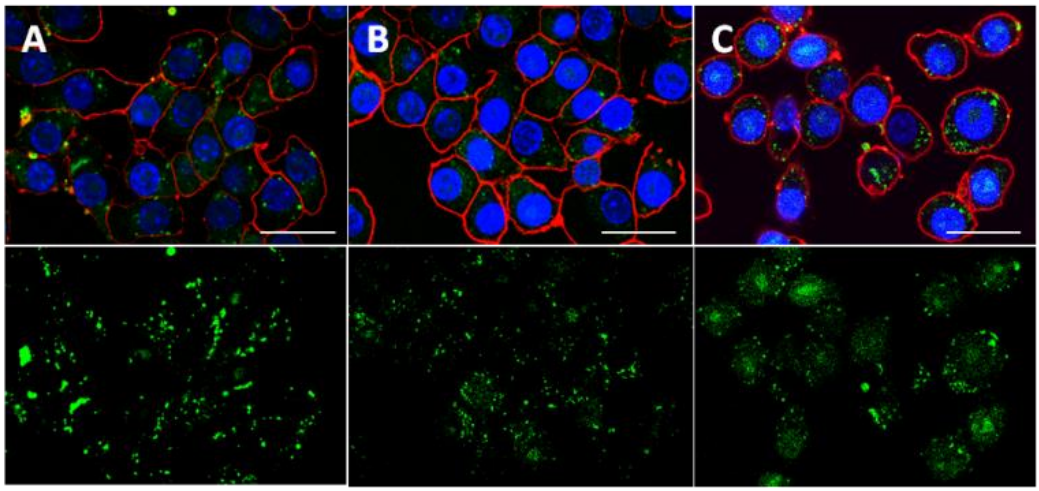


Fig. 4.4: Confocal microscopy images of the uptake of peptides (A) **Ib**, (B) **IId**, and (C) **IV** at 37 °C. The upper panel shows accumulated images of DNA in the nucleus (blue), cell membrane (red) and fluorescein (green). The lower panel shows the fluorescein image.

Quantification of the uptake by flow cytometry shows a concentration dependence as well as a temperature dependence on the mean fluorescence values measured (Fig. 4.5). At 37°C, a considerably larger uptake is observed at a 1 µM concentration of peptides. Furthermore, the mean fluorescence of the cells increased considerably for an i, i+7 staple, as can be seen from the results for **Ia-c**. Incubation at 4°C provides insight into the mode of uptake, as endocytic pathways are shut down at this temperature. Comparison of the data obtained at 4°C and 37°C with flow cytometry (Fig. 4.5) showed that there is significant uptake at 37 °C but almost no uptake at 4°C. This was further confirmed by confocal microscopy in a separate systematic study where we compared cell uptake for the non-stapled reference peptide **IV** and stapled peptide **Ib,** both at 4°C and 37°C (Fig. 5). Except for slight accumulation in and around the cell membrane, there was no uptake at 4°C. Thus, active uptake is the main internalization pathway for these peptides, based on the results obtained from flow cytometry and confocal microscopy.

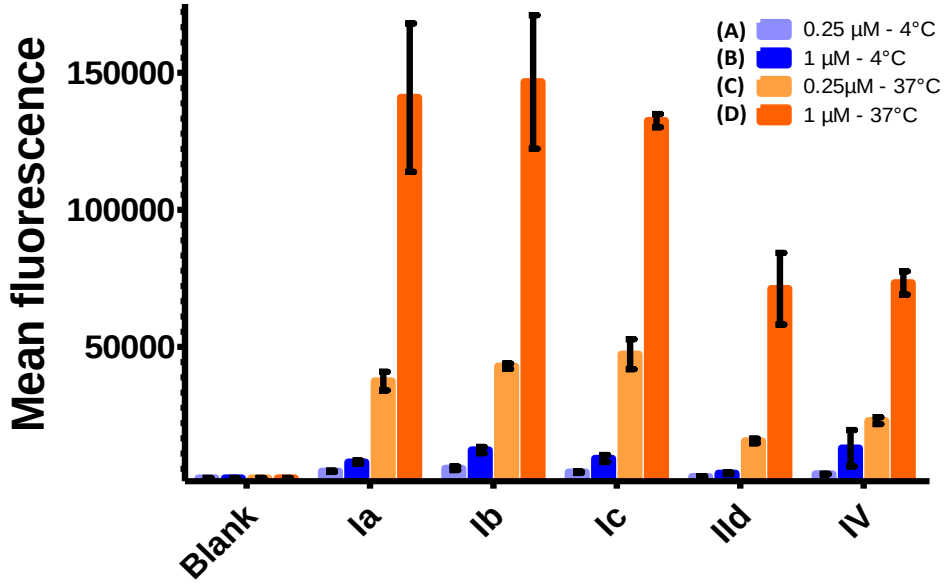


Fig. 4.5: Mean fluorescence of fluorescently labelled peptides by incubation at (A) 0.25 µM, 4°C, (B) 1 µM, 4°C, (C) 0.25 µM, 37°C, (D) 1 µM, 37°C.

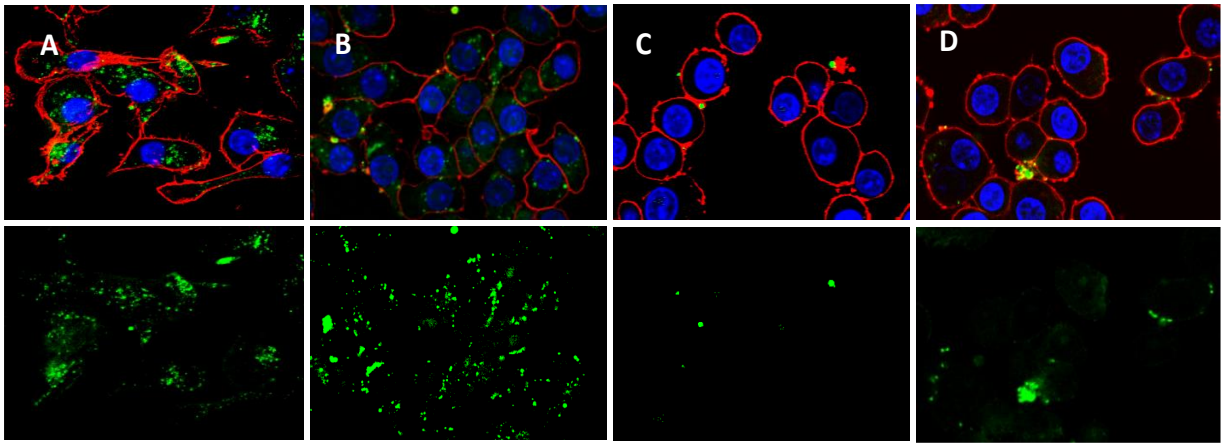


Fig. 4.6: Confocal microscopy images of the uptake of (A) peptide **IV** at 37°C, (B) peptide **Ib** at 37°C, (C) peptide **IV** at 4°C and (D) peptide **Ib** at 4°C at 0.25 µM. The upper panel shows accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). The lower panel shows the fluorescein image.

The cell uptake properties even at low concentrations for the current peptides are combined with low cytotoxicity values, as observed in an MTT-assay (Fig 4.7). From the MTT assay, we can conclude that for peptides **IV**, **Ia**, **Ib**, **Ic** and **IId** in general, more than 70% of the cells are viable at a concentration of 0.25 µM. Cell viability was evaluated according to the ISO10993-5 norms, which state that the compounds are cytotoxic if the assay points to a cell viability lower than 70%. Clearly, we see that the stapled peptides are not toxic at a concentration of 0.25 or 1 µM, except for **IId** which seems to be slightly toxic at a 1 µM concentration.

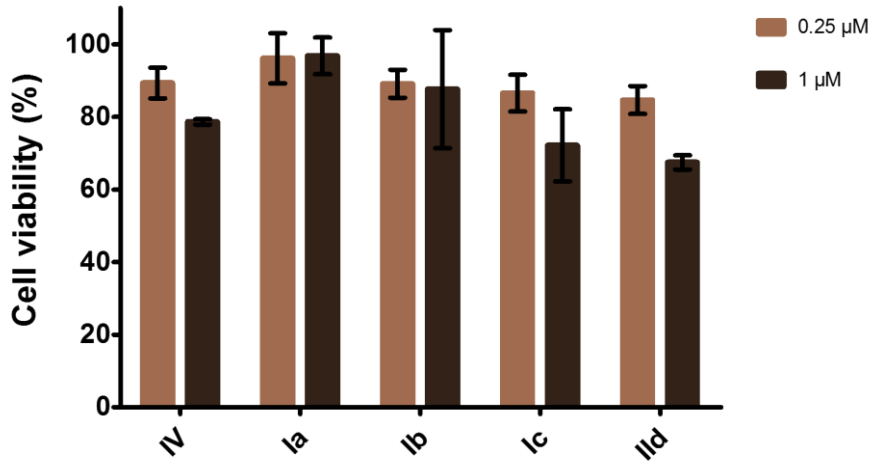


Fig. 4.7: MTT assay for fluorescently labelled peptides **IV**, **Ia**, **Ib**, **Ic** & **IId** measured at concentrations of 0.25 µM and 1 µM.

4.6 Peptide Stability

Dorien Van Lysebetten is thanked for performing the peptide stability tests.

Many hydrocarbon stapled peptides exhibit increased proteolytic resistance, proportional to the degree of α-helicity and number of staples introduced, resulting in enhanced therapeutic properties[30]. In case of cysteine cross-linking, no systematic comparison of peptide stability of stapled versus unstapled peptides has been reported. As a high number of Arg and Lys are present, the peptides will be very

susceptible to cleavage of trypsin. Therefore, a trypsin digest experiment was carried out using a trypsin to peptide ratio of 1/1000 wt% and samples were analysed after 0 min, 30 min, 1 h, 2 h and 24 h using RP-HPLC and MALDI-TOF. As expected, the control peptide **IV** was more susceptible to degradation than the stapled peptides, which can be seen in the higher number of different degradation products (fig. 4.8). The degradation products clearly reflect the influence of the staple on the cleavage process. Remarkably, while peptide **IV** was completely degraded after 24 h, between 20% (**Ic**) and 80 % (**Ib**) was left for the different stapled peptides (see experimental section for chapter 4 Fig. 51).

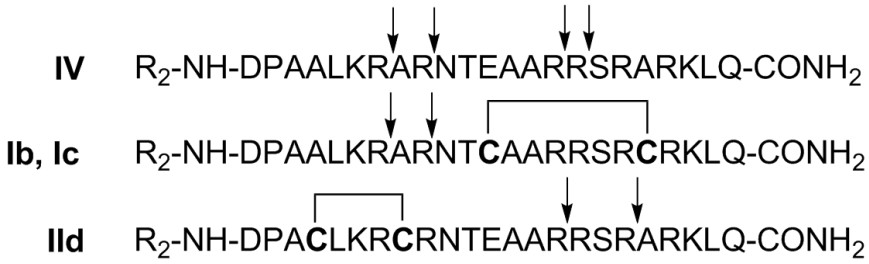


Fig. 4.8: Trypsin digestion of peptides showing different cleavage sites indicated by the arrows, observed with RP-HPLC analysis.

4.7 Conclusion

The elegant nature of the synthesis in combination with the observed DNA binding and cellular uptake properties render these constructs to be of considerable and specific interest among the mimics of the GCN4 transcription factor reported to date. Through this work and due to the nature of the stapled peptides, we believe that a general method is now at our disposal to allow DNA binding and enhance cellular uptake of a given DNA binding peptide, while avoiding tedious synthetic routes. It has been further shown that N-terminal helix stabilization, as in the case of **2d,** is more effective for enhancing DNA binding than stapling in the middle of the sequence (**1a-c**). For cell uptake, however, the more helical i, i+7 stapled peptides **Ia-c** have shown better uptake than **IV**. Although the two are not mutually exclusive, for future applications such as DNA binding *in cellulo*, a balance will have to be found between a peptide's DNA binding and cell penetration abilities.

The knowledge gained in the areas of peptide synthesis, transcription factors and in particular from the cell uptake studies carried out at UGent have helped tremendously to establish a platform for expanding the scope of peptides as a delivery system. Peptide based drug delivery into Gram negative bacteria will form the main discussion of this PhD thesis in the subsequent chapters.

4.8 References:

1. Latchman, D. S. *Eukaryotic Transcription Factors*. (Elsevier, 2011). doi:10.1075/z.172.00etra
2. Pazos, E., Mosquera, J., Vazquez, M. E. & Mascarenas, J. L. DNA Recognition by Synthetic Constructs. *Chembiochem* **12,** 1958–1973 (2011).
3. Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. An overview of the structures of protein-DNA complexes. *Genome Biol.* **1,** 1–10 (2000).
4. Carrette, L. L. G., Morii, T. & Madder, A. Peptidosteroid Tweezers Revisited: DNA Binding Through an Optimised Design. *European J. Org. Chem.* **2014,** 2883–2891 (2014).
5. Morii, T., Simomura, M. & Morimoto, S. Sequence-Specific DNA Binding by a Geometrically Constrained Peptide Dimer. *J. Am. Chem. Soc.* **32,** 1150–1151 (1993).
6. Vázquez, M Eugenio, A. M. & Mascareñas, J. L. A Light-Modulated Sequence-Specific. *Angew. Chem. Int. Ed. Engl.* **112,** 3234–3237 (2000).

7. Zhang, M. I. N., Wu, B., Zhao, H. & Taylor, J. W. The Effect of C-Terminal Helix Stabilization on Specific DNA Binding by Monomeric GCN4 Peptides. *J. Pept. Sci.* **136,** 125–136 (2002).
8. Park, C., Campbell, J. L. & Goddard, W. A. Can the Monomer of the Leucine Zipper Proteins Recognize the Dimer Binding Site without Dimerization ? *J. Am. Chem. Soc.* **118,** 4892–4896 (1996).
9. Henchey, L. K., Jochim, A. L. & Arora, P. S. Contemporary strategies for the stabilization of peptides in the a -helical conformation. *Curr. Opin. Chem. Biol.* **12,** 22–26 (2008).
10. Mosquera, J., Jiménez-Balsa, A., Dodero, V. I., Vázquez, M. E. & Mascareñas, J. L. Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides. *Nat. Commun.* **4,** 1874 (2013).
11. Estieu-gionnet, K. & Guichard, G. Stabilized helical peptides : overview of the technologies and therapeutic promises. *Drug Discov.* **6,** 937–963 (2011).
12. Madden, M. M. *et al.* Bioorganic & Medicinal Chemistry Letters Synthesis of cell-permeable stapled peptide dual inhibitors of the p53-Mdm2 / Mdmx interactions via photoinduced cycloaddition. *Bioorg. Med. Chem. Lett.* **21,** 1472–1475 (2011).
13. Inhibitor, H.- *et al.* A Cell-penetrating Helical Peptide as a Potential. *J. Mol. Biol.* **378,** 565–580 (2008).
14. Brown, C. J. *et al.* Stapled peptides with improved potency and specificity that activate p53. *ACS Chem. Biol.* **8,** 506–12 (2013).
15. Jo, H. *et al.* Development of α-helical calpain probes by mimicking a natural protein-protein interaction. *J. Am. Chem. Soc.* **134,** 17704–13 (2012).
16. Kim, Y.-W., Grossmann, T. N. & Verdine, G. L. Synthesis of all-hydrocarbon stapled α-helical peptides by ring-closing olefin metathesis. *Nat. Protoc.* **6,** 761–71 (2011).
17. Jacobsen, Ø., Maekawa, H., Ge, N., Görbitz, C. H. & Rongved, P. Stapling of a 3 10 -Helix with Click Chemistry. 1–73
18. Horne, W. S. ChemComm Promoting peptide a -helix formation with dynamic covalent oxime. *Chem. Comm.* **47,** 1–4 (2011).
19. Demizu, Y., Yamagata, N., Nagoya, S., Sato, Y. & Doi, M. Enantioselective epoxidation of α - β unsaturated ketones catalyzed by stapled helical L -Leu-based peptides. *Tetrahedron* **67,** 6155–6165 (2011).
20. Walensky, L. D. & Bird, G. H. Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *J. Med. Chem.* **57,** 6275–6288 (2014).
21. Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted Helices : Crystal Structure of the Protein-DNA Complex. *Cell* **71,** 1223–1237 (1992).
22. Jimønez-balsa, A., Pazos, E., Martínez-albardonedo, B., Mascarenas, J. L. & Vazquez, M. E. Temporary Electrostatic Impairment of DNA Recognition : Light- Driven DNA Binding of Peptide Dimers. *Angew. Chem. Int. Ed. Engl.* **51,** 8825–8829 (2012).
23. Keller, W., Ko, P. & Richmond, T. J. Crystal Structure of a bZIP / DNA Complex at 2 . 2 Å : Determinants of DNA Specific Recognition. *J. Mol. Biol.* **254,** 657–667 (1995).
24. Zondlo, N. J. & Schepartz, A. Highly Specific DNA Recognition by a Designed Miniature Protein. *J. Am. Chem. Soc.* **121,** 6938–6939 (1999).
25. Guerrero, L., Smart, O. S., Woolley, G. A. & Allemann, R. K. Photocontrol of DNA Binding Specificity of a Miniature Engrailed Homeodomain. *J. Am. Chem. Soc.* **127,** 15624–15629 (2005).
26. Futaki, S. *et al.* Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J. Biol. Chem.* **276,** 5836–40

(2001).
27. Langel, Ü. *Handbook of Cell-Penetrating Peptides*. (CRC Press, 2006).
28. Muppidi, A., Wang, Z., Li, X., Chen, J. & Lin, Q. Achieving cell penetration with distance-matching cysteine cross-linkers: a facile route to cell-permeable peptide dual inhibitors of Mdm2/Mdmx. *Chem. Comm.* **47,** 9396–8 (2011).
29. Nomura, W. *et al.* Cell-permeable stapled peptides based on HIV-1 integrase inhibitors derived from HIV-1 gene products. *ACS Chem. Biol.* **8,** 2235–44 (2013).
30. Walensky, L. D. & Bird, G. H. Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *J. Med. Chem.* **57,** 6275–6288 (2013).

CHAPTER 5

ANTIBIOTIC RESISTANCE: PAST, PRESENT & FUTURE

The problem of Antibiotic Resistance has plagued the 21st century with a host of acute problems. It has been estimated that by 2050 an additional 10 million people yearly will succumb to drug resistant infections.[1] *In 2013, the excess healthcare costs due to resistant infections in the US/EU totalled to a staggering $20/€ 1.6 billion!*[2] *Excess hospital days caused by resistant infections in the US and EU tallied to 8 and 2.5 million!*[3] *The list just never ends…and what's more... it is only going to get worse from here if something isn't done soon. In the previous chapters, we developed peptide based transcription factor models using cell permeable stapled peptides which could be further developed as anti-cancer drugs. In the subsequent chapters, we will use the experience gained in the area of cell penetrating peptides described in chapters 2 & 3 to establish a new peptide based delivery system for drugs capable of crossing the protective outer membrane of Gram negative bacteria which are mainly responsible for ushering of the antibiotic resistance. Before we venture into that aspect, however, we will first take a brief look at the reasons behind antibiotic resistance, the current drugs/strategies and what can be done to avoid the dreaded "post-antibiotic era".*

5.1 Introduction

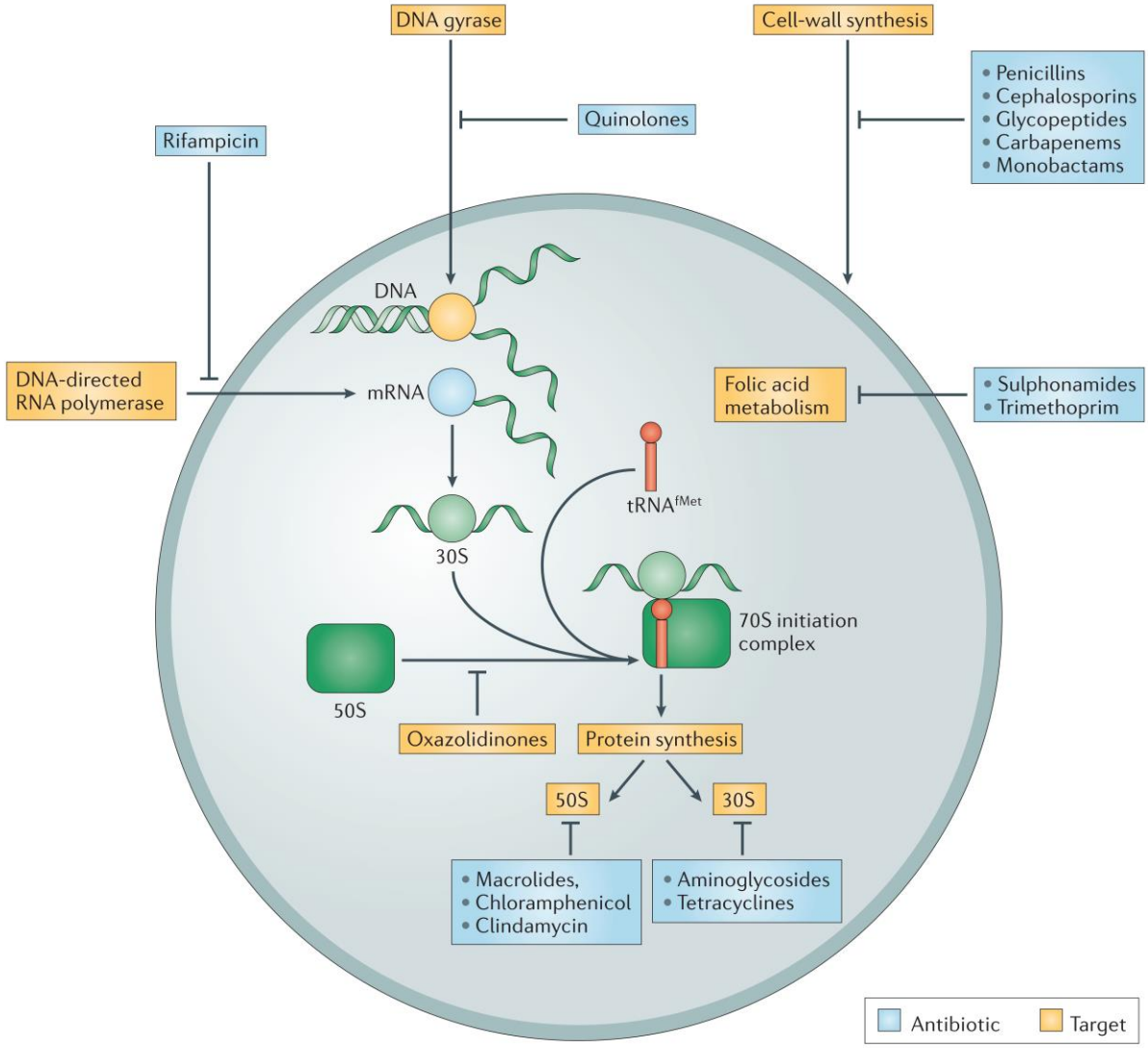
Truth be told, it is not entirely surprising that antibiotic resistance has emerged in the 21st century. Ever since the dawn of the antibiotic era, almost every scientist in the field knew that this day would eventually come. What they did not expect, however, was the alarming rate at which resistance would emerge. The past twenty years have seen a drastic rise in resistant strains of bacteria. The alarming rate at which new bacteria are emerging, coupled with a significant decline in new antibiotics, is a cause for major concern. Several questions arise from this, starting with what actually is antibiotic resistance? How does it occur? What can be done to prevent it? What does the future hold? In the subsequent chapters of this PhD, we will do our best to answer these questions providing details of the research that has been done in the University of Lincoln by me under the guidance of Dr. Ishwar Singh to develop new antibiotics.

5.1.1 Antibiotics & antibiotic resistance

An antibiotic can be defined as "Any substance that inhibits the growth and replication of a bacterium or kills it outright".[4] Antibiotic resistance can be defined as "The resistance of a microorganism to an antibiotic that was originally effective for treatment of infections caused by it".[5] It is to be noted here that antibiotics are different from antimicrobial compounds. In fact, antibiotics are a particular type of antimicrobial specially designed for the treatment of bacterial infections within (or on) the body. This makes them different from antiseptics (used to sterilize surfaces of living tissue) or disinfectants (which kill a wide range of organisms). Bacteria aren't the only microbes harmful to us. There's fungi and viruses too for which there are antifungals and antiviral compounds – but we will not discuss those here.

5.1.2 Targets of antibiotics

Research has shown that there are around 200 conserved essential proteins in bacteria. Unfortunately, the number of currently exploited targets is very small[6]. The most successful antibiotics target only a few of these pathways: cell wall synthesis[7] (β-lactams, such as penicillin or cephalosporins), DNA gyrase[8] (quinolones), DNA-directed RNA polymerase[9] (rifampicin) protein synthesis, i.e. the ribosome[10] consisting of 50S and 30S subunits, (examples include macrolides, chloramphenicol, clindamycin, aminoglycosides, tetracyclines and oxazolidinones) and enzymes[11] (sulphonamides and trimethoprim). A simple schematic of some of the known targets of antibiotics has been shown in Figure 5.1.

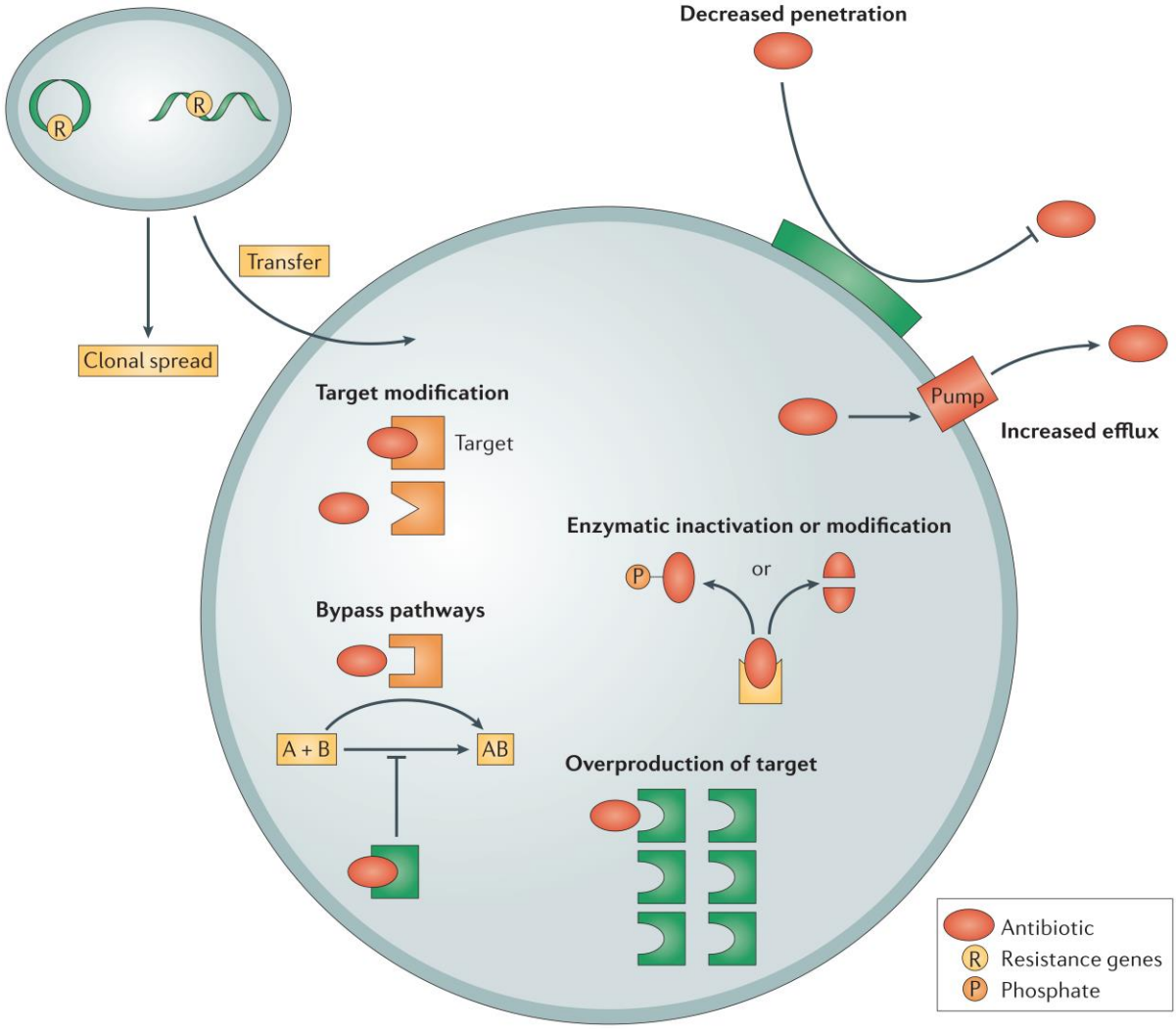


[I] [II]**Figure 5.1**: Schematic showing some of the known targets of antibiotics.

[I] Reproduced from Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12,** 371–87 (2013)
[II] Original article: Coates, A., Hu, Y., Bax, R. & Page, C. The future challenges facing the development of new antimicrobial drugs. *Nat. Rev. Drug Discov.* **1,** 895–910 (2002).

5.1.3 Mechanisms of antibiotic resistance and tolerance

The progress science and medicine makes is severely hampered by the rise and spread of antibiotic resistance. An abbreviation 'ESKAPE' has been assigned to multidrug resistant organisms which pose the most serious threat to human health. ESKAPE stands for *Enterococcus* spp., *Staphylococcus aureus*, *Klebsiella* spp., *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* spp.[12] (Note: spp stands for several species/subspecies) For *A. baumannii*, a resistant strain of Gram Negative Bacteria, there are reports saying "no available antibiotics"[13]. The general principle behind the design of an antibiotic is to attempt to shut down or disrupt essential cellular functions thereby considerably slowing down or killing entirely the growth of bacteria. Resistance mechanisms, on the other hand, operate on the strategy of preventing a drug from reaching its target. Majority of clinically relevant resistance mechanisms have been studied extensively and are therefore generally well understood[14, 15] (Figure 5.2). Examples include destruction of the antibiotic (by β-lactamases for example); target modification (Resistance to streptomycin is due to a mutation in the 30S ribosomal protein RpsL); as well as decreased penetration and/or efflux of the drug (efflux of linezolid by the AcrAB–TolC multidrug pump)[16, 17].



[III] [IV]**Figure 5.2**: Schematic showing some of the common ways in which antibiotic resistance has developed.

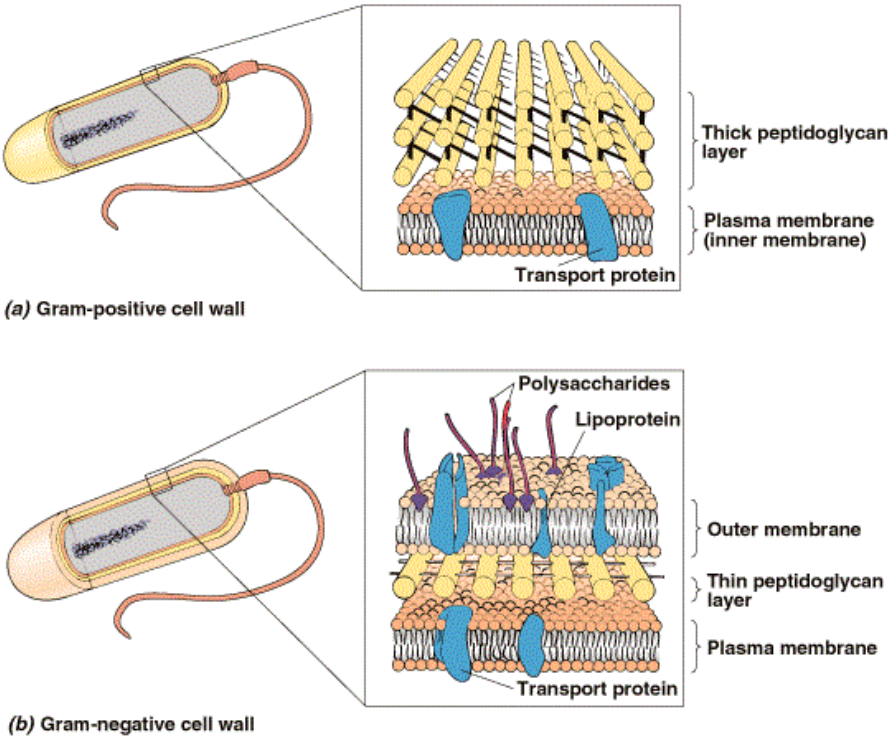
[III] Reproduced from Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12,** 371–87 (2013)
[IV] Original article: Coates, A., Hu, Y., Bax, R. & Page, C. The future challenges facing the development of new antimicrobial drugs. *Nat. Rev. Drug Discov.* **1,** 895–910 (2002).

Antibiotic resistance, therefore, is reasonably well-understood and manageable, but tolerance, however, is an entirely different challenge altogether. The main reason for the tolerance of pathogens to antibiotics is what is called a "persister" or a specialized survivor[18, 19]. Interestingly, persisters are not mutants, as one would expect. They are, in fact, phenotypic variants of cells which divide actively. They are produced unexpectedly in the population. Their relative abundance slowly rises and even at the late-exponential phase of growth does not rise to more than 1%[20]. It may not seem like much, but persisters are quite unlike the normal cells encountered. They are non-growing[21], dormant[22, 23] cells, which explains to a high degree their tolerance to bactericidal antibiotics since antibiotics need active targets for killing the cell[24]. From the aforementioned pathogens so far, all of them form persisters[19], but the mechanisms through which the persisters emerge is mostly unknown. In *Escherichia coli*, which can be regarded as the model organism for study, it has been shown that toxin–antitoxin modules are the key mechanism of persister formation.[23, 24, 25, 26] Unfortunately, it was also uncovered that the pathways of persister formation are highly redundant[27]. This redundancy makes target identification for drug discovery extremely challenging.

There is a close link between persisters and the emergence of drug tolerance. This was shown when an unexpectedly high number of persisters were selected in antimicrobial therapy in infections caused by *Pseudomonas aeruginosa*[28]. Overall, we can conclude that an important role is played by persisters in the development of antibiotic-resistant variants. Even if we do manage to kill them, persisters are killed rather slowly and more significantly resume growth when the antibiotic concentrations fall. This result is catastrophic because it results in a relapsing infection with a significant relative population size which will have a very high tendency to develop resistance[29]. Although the importance of persisters in drug resistant pathogens raises the bar considerably for drug discovery and is definitely something worth mentioning and considering, it is currently beyond the scope of this PhD thesis which will be more focussed on the discovery of new antibiotics capable of targeting the non-persisters.

5.2 Gram positive & Gram negative bacteria

It is common knowledge that there are two major types of bacteria: Gram positive and Gram negative. Their names arise from the Danish bacteriologist Hans Christian Gram who developed the Gram stain in 1884 comprising primarily of crystal violet and safranin. He noticed certain type of bacteria retain his stain and turn violet whereas others do not retain his stain and when counterstained, turn red. He therefore named the former Gram-positive bacteria and the latter Gram-negative bacteria. Little did he realize the significance of his discovery at that time. Today, of course, we know a lot more about Gram positive and Gram negative bacteria – the major difference being the presence of a robust outer membrane present in Gram negative bacteria but absent in Gram positive bacteria (Figure 5.3), making Gram negative bacteria particularly efficient in keeping out drugs. For the current thesis, we will be more focussed on development of antibiotics towards the more resistant Gram negative bacteria.



[V]**Figure 5.3**: Figure showing the difference between (a) Gram positive cell wall & (b) Gram negative cell wall.

5.3 Gram negative bacteria: Other modes of resistance

The current focus has been towards the development of drugs against Gram-positive bacteria such as *vancomycin resistant enterococci* **(**VRE) and methicillin-resistant *Staphylococcus aureus* (MRSA). The newly discovered antibiotic teixobactin (Figure 5.4) is an excellent example of an antibiotic active against MRSA[30]. We have already seen some of the modes of resistance in bacteria in general. However, Gram negative bacteria such as *Klebsiella pneumoniae* (*K. pneumoniae*), *Acinetobacter baumannii* (*A. baumannii*) and *Pseudomonas aeruginosa* (*P. aeruginosa*) in particular pose a more serious challenge to human health due to their natural ability to acquire resistance via multiple modes[31].

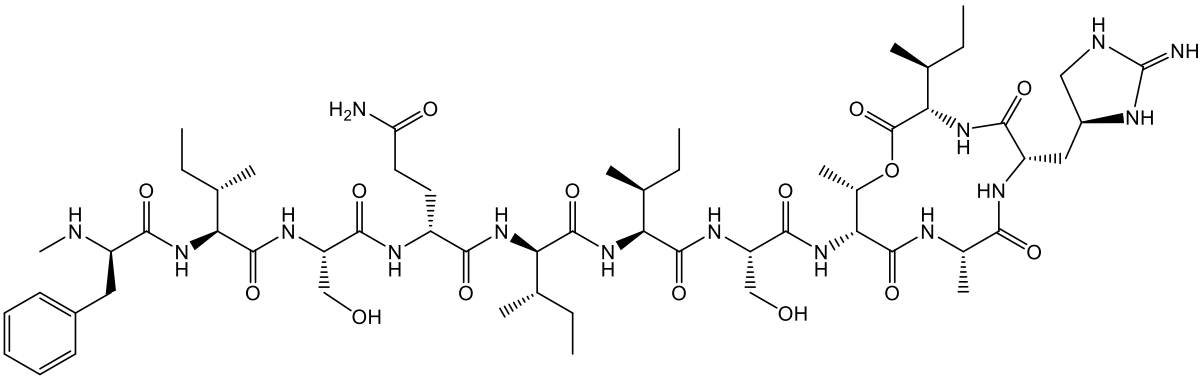


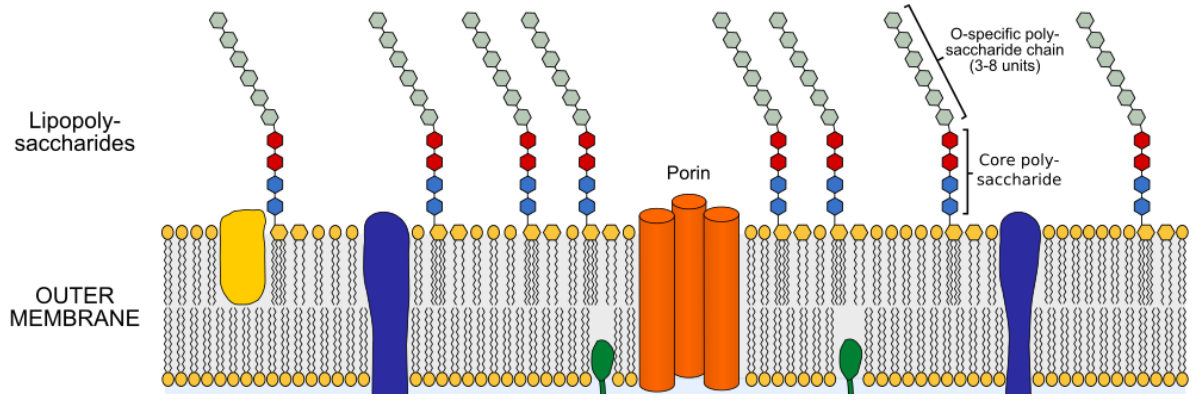
Figure 5.4: Structure of teixobactin.

[V] Reproduced from Solomon, Biology, 5th Edition, Figure 23.10

There has been virtually no new antibiotic for Gram negative bacteria in decades. Therefore, the polymyxins (described in section 5.4.2) discovered more than 50 years ago, have been revived and are being used as a drug of last resort against these bacteria[32] despite reports of their toxicity as well as increasing resistance[33]. Thus, the problem of antibacterial resistance is only getting worse – especially against the drug resistant Gram negative bacteria[34, 35]. Therefore, there is an urgent need to develop new antimicrobial compounds particularly against the resistant Gram negative bacteria. The outer membrane (OM) of Gram negative bacteria significantly limits the access of antibiotics to reach their targets, while no such barrier is present in Gram positive bacteria. One of the key challenge which is necessary to overcome to develop new antibiotics against Gram negative bacteria[36] is the lack of compounds which can penetrate through the outer membrane which acts as a barrier protecting these bacteria. The multidrug resistance (MDR) efflux pumps[37, 38] and the transfer of resistant genes in Gram negative bacteria via multiple modes such as Horizontal Gene Transfer (HGT)[39, 40] are the other challenges especially in resistant Gram negative bacteria. In what follows, we will briefly discuss the major problems associated with finding novel antibiotics against Gram negative bacteria.

5.3.1 Impermeable barrier

The major problem associated with antibiotic discovery against Gram negative bacteria is the presence of an impermeable outer membrane protecting it from a vast majority of compounds. The outer leaflet of the outer membrane bilayer is composed of an unusual lipid, lipopolysaccharide (LPS), compared to the usual glycerophospholipid found in most other biological membranes. Three main parts make up the LPS: a proximal hydrophobic lipid A region, a core oligosaccharide region and O-antigen (a repetitive glycan polymer)[41]. Essentially, all drugs are amphipathic due a dual requirement: Firstly, they need to be water soluble and secondly, they need to be able to cross the cytoplasmic membrane. Since the vast majority of clinically important antibiotics show some hydrophobicity it allows them to diffuse across the membrane especially through the lipid A region. However, since the LPS is an asymmetric bilayer, the bacterial outer membrane still serves as an efficient barrier considerably slowing down the penetration of lipophilic antibiotics.



[VI]**Figure 5.5**: Figure showing the outer membrane of Gram negative bacteria.

Since the outer membrane of Gram negative bacteria prevents the majority of molecules from entering, bacteria with this barrier have developed other methods to bring in nutrients from their surroundings. The outer membrane of Gram negative bacteria contains porins, a special class of

[VI] Modified from https://en.wikipedia.org/wiki/Bacterial_outer_membrane#/media/File:Gram_negative_cell_wall.svg

proteins, which produce non-specific aqueous diffusion channels across the membrane. Porin channels exclude antibiotics from crossing them by having a very small diameter (7 to 10 Å at the most) which considerably slows down or completely shuts down antibiotic influx. The porin channels are also lined with charged amino acid residues. This orients the water molecules in a fixed direction making the channel quite hydrophilic. These charged residues make the influx of lipophilic molecules difficult due to an energetically unfavourable orientation of the water molecules. Nutrients can diffuse into the cells through mycobacterial porin – a protein present in very small amounts and allowing only very slow diffusion of small molecules through its channel[42]. Antibiotics are severely restricted in entering the bacteria by the low permeability of the porin channels and the lipid matrix, which allows a high degree of resistance for these bacteria.

5.3.2 Multidrug resistance efflux pumps

Four general mechanisms responsible for antibiotic resistance have been identified by the World Health Organization (WHO). They are: resistance including target alteration, drug inactivation, decreased permeability and increased efflux. Among these, drug expulsion by the multidrug efflux pumps is an important mechanism of MDR. Efflux pumps have the ability to pump out a broad range of antibiotics due to their polysubstrate specificity[38]. Additionally, they can also enable the bacteria to acquire additional resistance mechanisms by lowering the intracellular antibiotic concentration as a result of their action, and subsequently promoting mutation accumulation. Over-expression of multidrug efflux pumps is a major challenge associated typically with drug resistance. More studies have revealed that efflux pumps are not just for the development of bacterial resistance but also play important physiological functions in bacteria[43]. This makes their expression very tightly regulated in response to various environmental and physiological signals. A thorough understanding of the mechanisms of drug expulsion and regulation and physiological functions of efflux pumps is necessary for the development of anti-resistance methodologies. It is to be noted here that the importance of efflux pumps to bacteria makes them a good target for antibiotics, and drugs which target these pumps have met with success[44].

5.3.3 Horizontal Gene Transfer

Mutation is not the only way in which bacteria evolve rapidly. Transfer of DNA can result in strains with mutations from more than one source making the new bacteria even more resistant to antibiotics. Evidence of Horizontal Gene Transfer (HGT) was documented as early as 1928 when researchers observed the transference of virulent determinants between pneumococci in infected mice. This phenomenon was later shown to be mediated by the uptake of the genetic material DNA in a process called transformation[45]. This transformation requires the release of naked DNA with subsequent uptake and recombination. DNA-repair processes and homologous recombination usually tend to limit this to DNA from similar bacteria. However, there is always the possibility that a gene moves onto a broad-host-range plasmid. In that case, it might be able to spread rapidly and without the need for recombination. Gram negative bacteria (among others), in particular, have the natural ability to acquire DNA[46] as part of their genetic trait, giving them access to not only nutrients or genetic information but also allowing them to acquire new modes of resistance.

5.4 Leading drugs against Gram negative bacteria

A vast majority of drugs against Gram negative bacteria are present on the market and therefore we will not discuss all of them in detail. The aim of this section is to get a concise overview of drugs against the resistant Gram negative bacteria as a background for the subsequent chapters. For a more detailed overview of the current drugs[47] and the evolution of drugs, several reviews are available[48, 33, 49]. Some of the classes of drugs against Gram negative bacteria relevant for this thesis (with examples) are described below.

5.4.1 β-lactams, resistance to β-lactams and their sub-classes

Despite their structural diversity, all β-lactams possess a common four-membered β-lactam ring (i.e. a four membered ring with an amide bond) which serves as the active pharmacophore for this class of antibiotics. β-lactams are a very old class of antibiotics – the first one discovered was benzylpenicillin (penicillin G, Figure 5.5) back in 1928[36]. The most approved class of antibiotics by the FDA are β-lactams. They also on top (in terms of numbers) of WHOs list of critically important antibiotics to human medicine. 28 β-lactams, including antibiotic/ β-lactamase inhibitor combinations, comprising three subclasses: penicillins, cephalosporins, and carbapenems are listed as critically important[50]. Their antibacterial activity is a result of their ability to act as suicide substrates for penicillin binding proteins (PBPs). They also inhibit cell wall biosynthesis, especially the maintenance of peptidoglycan. This leads to cell stress which ultimately results in cell lysis[51]. The currently used β-lactams have broad spectrum activity against most gram-positive and negative bacteria. This, coupled with low toxicity profiles makes them popular first-line antibiotics[16].

Resistance to β-lactams usually occurs via the hydrolysis of the β-lactam ring mediated by a wide range of β-lactamases with diverse classifications. We will briefly study the classification of these enzymes and which drugs they target. The new Ambler classification system has been updated in 2010 and three groups have been formed[52]. However, for simplicity, we will study them according to their classes:

1. **Class A**: Consisting of *Klebsiella Pneumoniae* Carbapenamase (KPCs) and most Extended Spectrum β-lactamases (ESBLs).
2. **Class B**: Metallo-β-lactamases (MBLs).
3. **Class C:** AmpC β-lactamases (AmpC).
4. **Class D**: OXA β-lactamases (OXA).

Many enzymes which can hydrolyze penicillins and cephalosporins, including some that can hydrolyze monobactams (β-lactams where the β-lactam ring is alone and not fused with other rings) and KPCs capable of hydrolyzing carbapenems, all belong to class A[53]. The ESBLs belonging to this class are mediated via plasmids which has aided in their diffusion to multiple species[54]. MBLs use divalent cations such as zinc as cofactors. Integrons[55] are genetic units possessing a unique ability to capture and incorporate gene cassettes (which contain a gene and recombination site) by site-specific recombination. Many are encoded in class 1 integrons – a major problem in antibacterial resistance[56]. Typically, they also affect gene cassettes responsible for coding aminoglycoside modifying enzymes (AMEs), found on transposons (a small piece of DNA capable of inserting itself into another part of the genome, thereby creating or reversing mutations), thereby facilitating their spread[57]. MBLs are capable of inactivating many β-lactams including carbapenems with currently no improved inhibitors for them. However, they have no activity against the monobactam aztreonam (Figure 5.9 I)[58].

AmpC β-lactamases are typically chromosomally encoded. Inactivation of many β-lactams including aztreonam occurs through AmpC and other class C β-lactamases. They show preferential activity against cephalosporins, but do not show any activity against carbapenems[53-59]. Many OXAs are encoded on integrons[60, 61, 62, 63]. Class D comprising only of OXAs are capable of hydrolysing cephalosporins and aztreonam . Some of them have carbapenemase activity as well[53-64]. Although they are not as potent as MBLs they are particularly problematic because they are the most commonly found β-lactamase in *Acinetobacter*[54]. *Streptococci*, in particular can possess altered PBPs, making the role of β-lactamases difficult[65]. Production of low affinity PBP2a in MRSA has caused resistance to Methicillin (**5**) and other β-lactam in greater than 90% of isolates[66]. A similar problem has been observed in *S. pneumoniae* due to expression of a variety of low affinity PBPs[16]. Resistance to β-lactams can also be caused by efflux by RND and ABC efflux pumps,[67] and outer membrane impermeability[68].

1. The penicillin subclass

The penicillin subclass of β-lactams form the oldest antibiotics developed. β-lactams have advanced mainly via semi-synthetic modifications. Amongst the penicillin subclass (Fig. 5.6) some early synthetic modifications were aimed at increasing stability to penicillinases through the attachment of bulky side chains as in the cases of methicillin (Figure 5.6 II) and oxacillin (Figure 5.6 III). A separate set of modifications were made to penicillin G (Figure 5.6 I) which is relatively narrow-spectrum to make it effective against gram-negatives. Examples include the aminopenicillins such as ampicillin (Figure 5.6 IV) and amoxicillin (Figure 5.6 V), and ureidopenicillins like piperacillin (Figure 5.6 VI). Despite most bacteria today being resistant to this class, many penicillins still remain important first-line antibiotics[16].

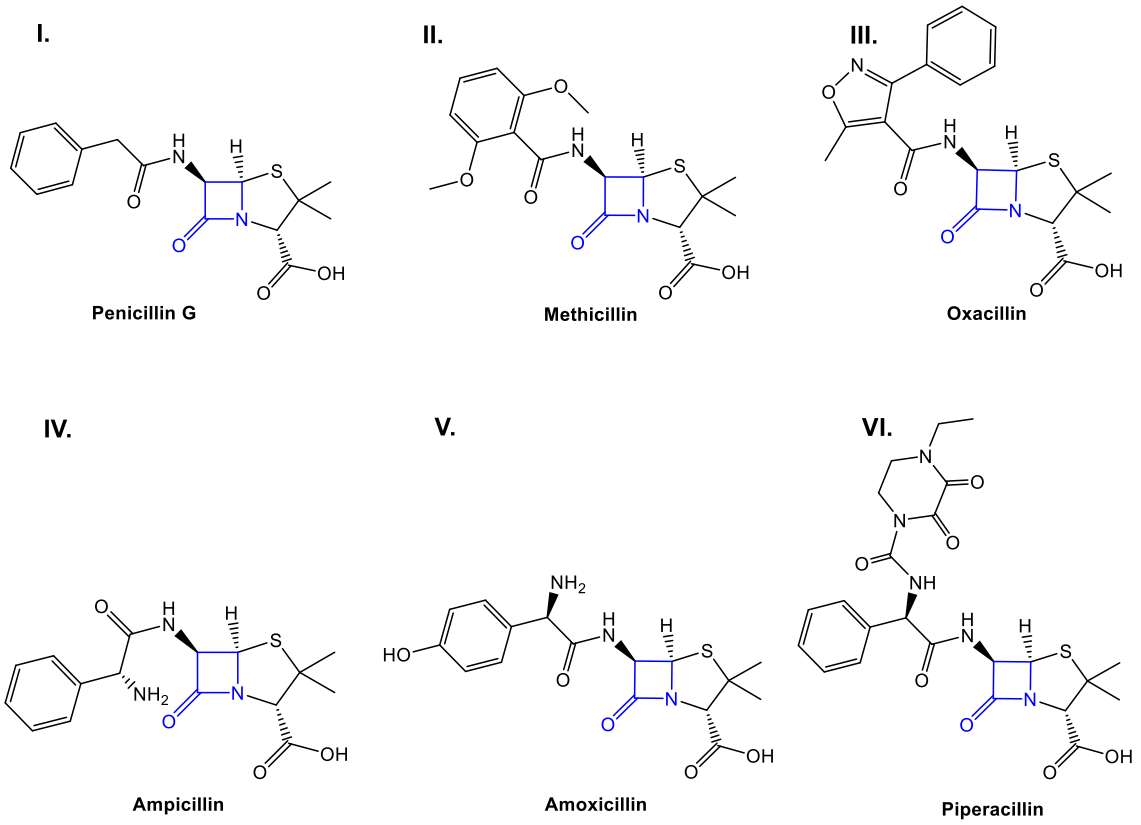


Figure 5.6: β-lactams of the Penicillin subclass. The β-lactam moiety is highlighted in blue.

2. The cephalosporins

The cephalosporins are a subclass of β-lactam antibiotics originally derived from the fungus *Acremonium*, known previously as "*Cephalosporium*". The first cephalosporin was cephalosporin C (Figure 5.7 I), discovered in 1948. Initial semi-synthetic cephalosporins were mainly developed to improve pharmacokinetics and increase spectrum of activity particularly against gram-negative bacteria through increased cellular penetration. The subsequent generations have become increasingly focused on combating β-lactam resistance[16]. Currently, the fifth generation of cephalosporins are in production. They possess excellent safety profiles and increased spectrum of activity, which have made them some of the most highly utilized first-line antibiotics. Ceftaroline (Figure 5.7 III), a fifth generation cephalosporin approved by the FDA in 2010, has shown increased activity against MRSA. Unfortunately, it is not potent against MDR gram-negatives by itself[69, 70]. However, a combination of ceftaroline with the β-lactamase inhibitor, tazobactam (Figure 5.10 II) was found to be quite effective against resistant strains, although its inactivation via certain β-lactamases could not be prevented[53]. Cubist's ceftolozane (Figure 5.7 II), which is in phase III trials, has shown activity complementary to Ceftaroline. It has low activity against MRSA, but is active against many MDR gram-negative bacteria including *E. coli* and *K. pneumoniae* strains. Most notably, it has extremely good activity against *P. aeruginosa* including strains with AmpC β-lactamases and upregulated efflux[71]. There is a strong likelihood of it being used in combination with tazobactam (Figure 5.9 II), which should broaden its range of activity[72].

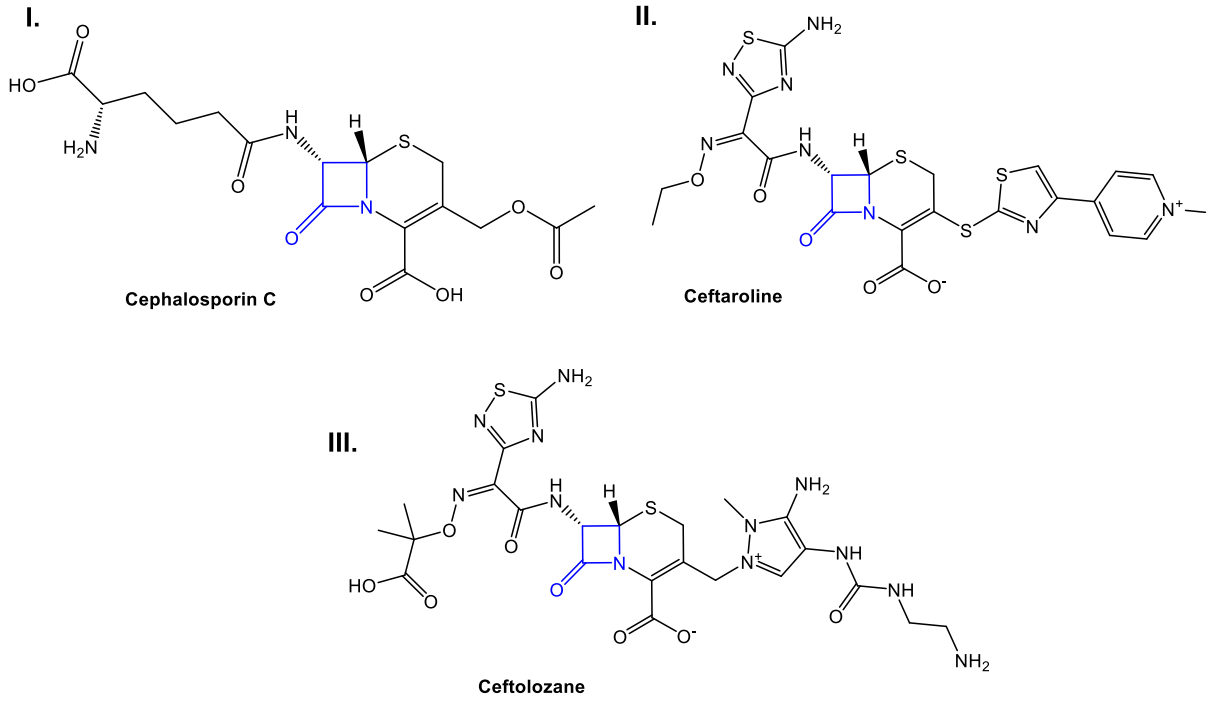


Figure 5.7: I-III. Structures of some of the cephalosporin subclass β-lactams. The β-lactam moiety is highlighted in blue.

3. The Carbapenems

The first carbapenem to be identified was Imipenem (Figure 5.8 I) in 1976. Their resistance to ESBLs makes carbapenems quite potent against many anaerobic and gram-negative bacteria compared to other β-lactams although more recently evolved carbapenmases have resulted in a decreased activity. *P. aeruginosa* has developed resistance to imipenem (Figure 5.8 II) and meropenem (Figure 5.8 II) through the loss of OprD porins and MexAB-OprM efflux upregulation[54]. Doripenem (Figure 5.8 III), was approved in Japan in 2005 and in the US in 2007. It is unaffected by certain KPCs and OXAs, but remains susceptible to all MBLs[73]. It shows similar activity as imipenem against most bacteria but it is superior to other carbapenems against *P. aeruginosa*. It, however, lacks activity against MRSA[12, 74]. Meropenem and biapenem (Figure 5.8 IV) (approved in Japan) have been used successfully against some imipenem resistant, MBL producing *P. aeruginosa*[75]. A carbapenem in phase II trials, Razupenem (Figure 5.6 V), has shown promising activity against *E. faecium* (gram positive) which is resistant to ampicillin[76].

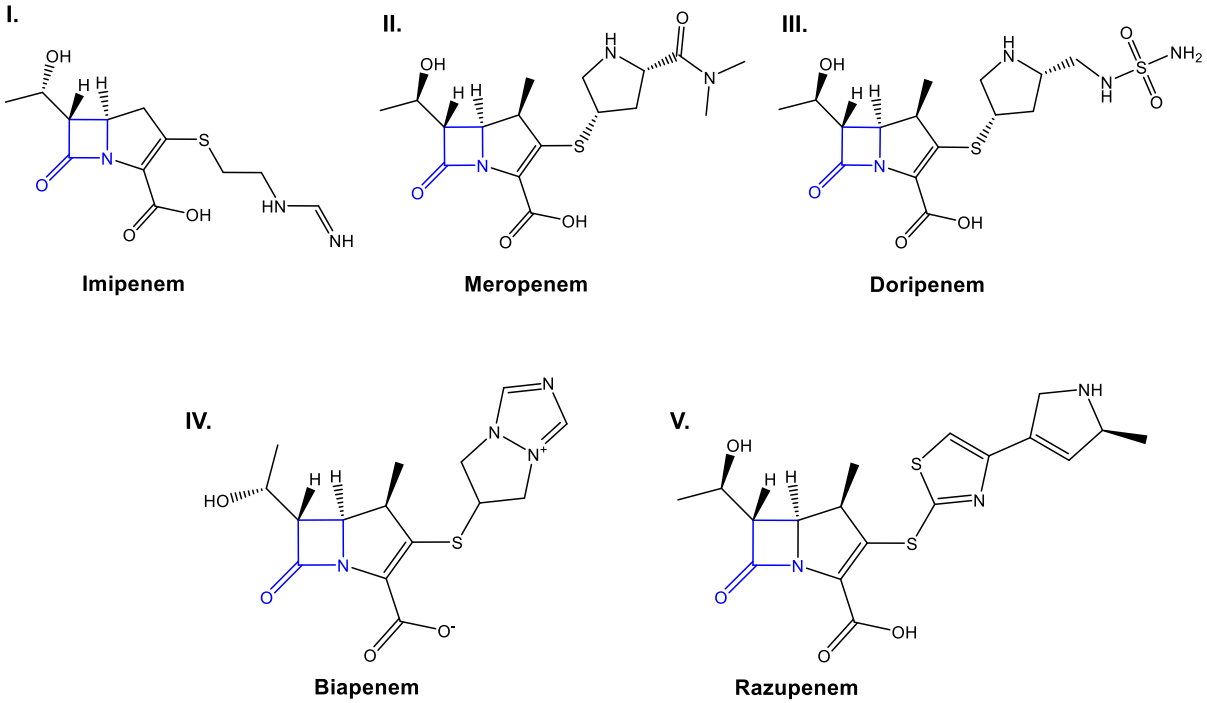


Figure 5.8: Structures of the Carbapenem subclass belonging to the β-lactam family. The β-lactam moiety is highlighted in blue.

4. The Monobactams

The first and only FDA-approved monobactam thus far, aztreonam (Figure 5.9 I), was identified in 1981. Though it is active only against gram-negative pathogens, it stands apart from all other antibiotics in its class because of its imperviousness to some of the most dreadful class B β-lactamases[16, 58]. Basilea's BAL30072 is still in phase I clinical trials (Figure 5.9 II), but has already proven to be a promising monobactam. It has shown excellent activity against MBL producing *P. aeruginosa* and *Acinetobacter* along with many KPC producing *Enterobacteriaceae*[77]. Likewise, it has also shown synergistic activity with meropenem (Figure 5.8 II) against *Acinetobacter*.

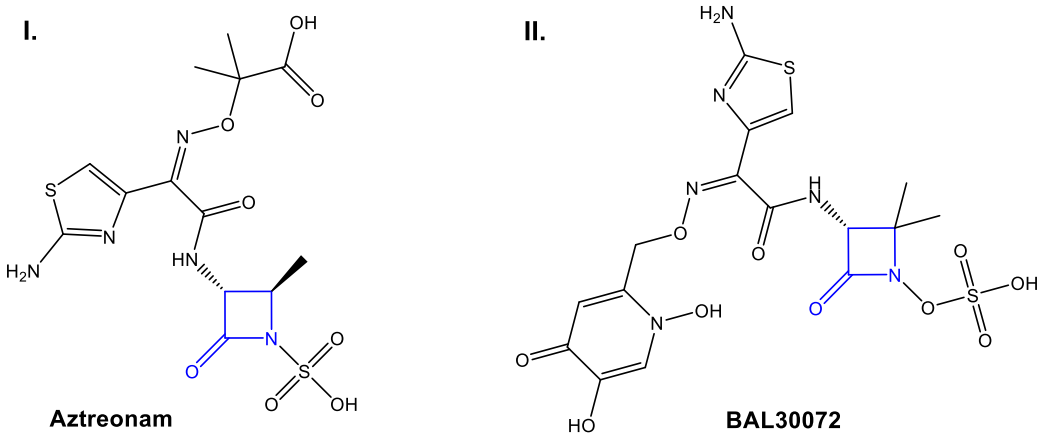


Figure 5.9: Structures of the monobactam subclass belonging to the β-lactam family. The β-lactam moiety is highlighted in blue.

5. The β-lactamases inhibitors

The first identified β-lactamase inhibitor was Clavulanic acid (Figure 5.10 I), which was discovered in 1976. A combination therapy of clavulanic acid (Figure 5.10 I) and amoxicillin (Figure 5.6 I), known as augmentin, is still amongst the most successful antibiotics on the market. Some *P. aeruginosa* infections, including those producing ESBLs, can be treated using piperacillin (Figure 5.6 VI) and tazobactam (Figure 5.10 II)[54]. Sulbactam (Figure 5.10 III) was initially used in combination with ampicillin in the past, with some success. Renewed interest in sulbactam was reported when it showed good activity in combination with meropenem (Figure 5.7 II) against a wide range of *A. baumannii* strains, although this combination is not yet clinically approved[78]. An interesting new β-lactamase inhibitor Avibactam (Figure 5.10 IV) is reported with broad spectrum activity against class A, C and D β-lactamases[79]. There are a number of β-lactam inhibitors, containing or not containing a β-lactam, that are either clinically approved or in clinical trials in combination with β-lactam antibiotics. Quite a few of these have activity against KPC, AmpC, and OXA β-lactamases[80]. However, there are very few with activity against MBLs and none that are currently clinically approved[81]. Some combinations of inhibitors, including ones that have siderophore (Greek for iron carriers; Siderophores are among the strongest known Fe^{3+} binders) activity, have shown some promise against MBLs[53]. Tricyclic competitive inhibitors of certain MBLs have also been isolated[82].

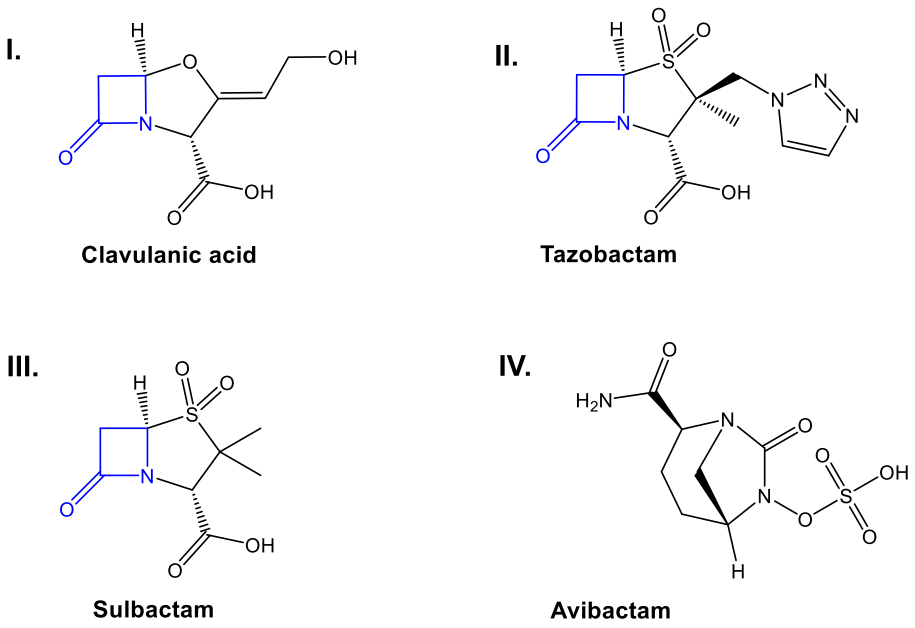


Figure 5.10: Structures of the β-lactamase inhibitors. The β-lactam moiety is highlighted in blue.

5.4.2 Polymyxins

The Polymyxins A-E are natural products of *Bascillus,* first discovered in 1947. Structurally, they are cyclic peptides with peptidyl side chains and a hydrophobic, saturated alkyl tail capping the N-terminus (Fig. 5.11). Colistin (polymyxin E, Figure 5.11 I) has been on the market since 1950. The polymyxins were discontinued due to significant nephrotoxicity and neurotoxicity[83] until recently, when interest was renewed due to a rapid increase of multidrug resistant bacteria. They are still, however, considered drugs of last resort. Recent studies have shown that the nephrotoxicity of colistin may have been overestimated possibly because of errors in dosing or inferior formulation[84, 85]. Polymyxins are polycationic due to the presence of as many as six diaminobutanoic acid residues, five of which have free amines (Figure 5.11 I). They operate by displacing stabilizing magnesium and calcium ions, allowing them to electrostatically interact with the outer layer of gram negative cell membranes containing anionic lipopolysaccharide (LPS). This disrupting interaction by polymyxins leads to an increase in bacterial cell membrane permeability, leakage, and rapid cell death[86]. Colistin has the added benefit of having potent anti-endotoxin activity[87].

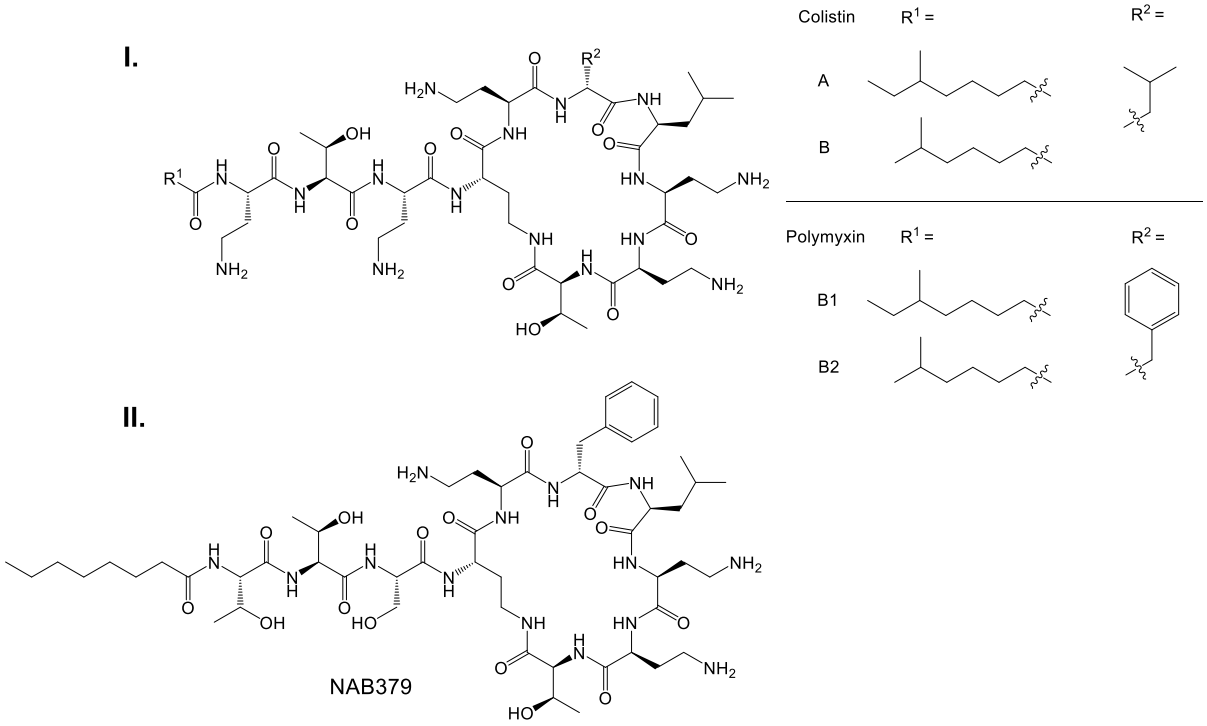


Figure 5.11: Figure I. showing the polymyxins B1 & B2 & collistin and Figure II. showing NAB739, the preclinical polymyxin B analogue.

The polymyxins are quite potent against most gram-negative bacteria (broad spectrum) although some reports say strains of *K. pneumoniae*, *P. aeruginosa*, *A. baumannii* and others have developed resistances[35]. Overall, however, resistance to polymyxins is quite uncommon, although its frequency varies significantly by bacterial species and by geographic region[88]. *E. coli* and *P. aeruginosa* have been known to express lower affinity modified LPS. This can also lead to resistance. *P. aeruginosa* can also upregulate membrane protein H1, thereby replacing divalent cations in the LPS and decreasing polymyxin affinity. *K. pneumoniae* is capable of increasing production of its capsule polysaccharide, which restricts polymyxin penetration to the LPS layer. Certain *B. polymyxa* are known to produce a degrading colistinase, but there are no reports which suggest these have spread to pathogenic bacteria strains[89].

Colistin is currently being used in the treatment of MDR gram-negative pathogens where other treatment options are unavailable. MDR resistant strains such as *Pseudomonas*, *Klebsiella*, and *Acinetobacter* including the NDM-1[90] producers can be targeted using Colistin. Recent studies have shown that reducing the overall positive charge in Polymyxin B analogues has maintained their antibacterial activity, while significantly improving in vitro toxicity profiles. One of these molecules, NAB739 (Figure 5.5II), is being actively developed in pre-clinical studies by Northern Antibiotics Ltd.[91, 92, 93, 94]

5.5 Conclusion: Learning from our mistakes to build a better future

A brief overview has been provided in this chapter covering the antibiotic problem in the 21st century, Gram positive & multidrug resistant Gram negative bacteria and the differences between them, mechanisms of antibiotics, modes of resistance and an overview of the drugs in the market or in development to cope against emerging resistance.

There is no denying that several mistakes have been made in the field of antibiotics research. After what was defined as the golden era of antibiotic research till the 1960s[95], there was a massive lull period for almost 20 years where nearly no new antibiotic was discovered. Significant complacency, ignorance coupled with irresponsible policies and practices, have time and again hampered antibiotic research. That being said, the situation today is far better than in the 1990s. The alarming rise of resistant bacteria has revived antibiotic research. The new 21st century has ushered in fresh perspectives. Advances in synthesis and technology have greatly helped in checking the further rise of resistant pathogens. The approval of new classes of antibiotics is definitely a step in the right direction. However, a lot of work remains to be done in the area of antibiotic research. The important thing is to keep steady on the path of science and constantly strive to avoid the dreaded "post-antibiotic" era.

5.6 References:

1. Review on Antimicrobial Resistance. *http://amr-review.org/* (2015). Available at: http://amr-review.org/.
2. Center for Disease Control and Prevention. World Heath Day: Media Fact Sheet. Available at: http://www.cdc.gov/media/releases/2011/f0407_antimicrobialresistance.pdf. (2013).
3. Center for Disease Control and Prevention. Antibiotic Resistance Threats in the United States, 2013. Available at: http://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf. (2013).
4. Antibiotic Resistance: a challenge for the 21st century. Available at: http://www.microbiologyonline.org.uk/media/transfer/doc/factfile_antibiotic_resistance.pdf. (2015).
5. World Heath Organization, fact sheet. Available at: http://www.who.int/mediacentre/factsheets/fs194/en/. (2015).
6. Mandell, G. L., Bennett, J. & Dolin, R. *Principles and Practice of Infectious Diseases 5th edn (Churchill Livingstone, Philadelphia)*. (2000).
7. Spratt, B. G. Biochemical and genetic approaches to the mechanism of action of penicillin. *Phil. Trans. R. Soc. Lond. B* **289,** 273–283 (1980).
8. Smith, J. T. The mode of action of 4-quinolones and possible mechanisms of resistance. *J. Antimicrob. Chemother.* **18,** 21–29 (1984).

9. Wehrli, W. Interaction of rifamycin with bacterial RNA polymerase. *Proc. Natl Acad. Sci. USA* **61,** 667–673 (1968).
10. Goldman, R. C., Fesik, S. W. & Doran, C. Role of protonated and neutral forms of macrolides in binding to ribosomes from Gram-positive and Gram-negative bacteria. (1990). *Antimicrob. Agents Chemother.* **34,** 426–431 (1990).
11. Bernstein, L. & Salter, A. Trimethoprim/Sulphamethozale in Bacterial Infections. *Churchill Livingstone, Edinburgh and London* **7–16,** (1973).
12. Boucher, H. W. *et al.* Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **48,** 1–12 (2009).
13. Higgins, P. G., Dammhayn, C., Hackel, M. & Seifert, H. Global spread of carbapenem-resistant Acinetobacter baumannii. *J. Antimicrob. Chemother.* **65,** 233–238 (2009).
14. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov.* **6,** 29–40 (2007).
15. Eliopoulos, G. M. *et al.* In vitro and in vivo activity of LY 146032, a new cyclic lipopeptide antibiotic. *Antimicrob. Agents Chemother.* **30,** 532–535 (1986).
16. Walsh, C. *Antibiotics: Actions, origins, resistance. DOI: 10.1128/9781555817886*. (2003). doi:10.1128/9781555817886
17. Alekshun, M. N. & Levy, S. B. Molecular Mechanisms of Antibacterial Multidrug Resistance. *Cell* **128,** 1037–1050 (2007).
18. Lewis, K. Persister cells, dormancy and infectious disease. *Nat. Rev. Microbiol.* **5,** 48–56 (2007).
19. Lewis, K. Persister cells. *Annu. Rev. Microbiol.* **64,** 357–72 (2010).
20. Keren, I., Kaldalu, N., Spoering, A., Wang, Y. & Lewis, K. Persister cells and tolerance to antimicrobials. *FEMS Microbiol. Lett.* **230,** 13–18 (2004).
21. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305,** 1622–1625 (2004).
22. Shah, D. *et al.* Persisters: a distinct physiological state of E. coli. *BMC Microbiol.* **6,** 53 (2006).
23. Dorr, T., Vulic, M. & Lewis, K. Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. *PLoS Biol.* **8,** 29–35 (2010).
24. Keren, I., Shah, D., Spoering, A., Kaldalu, N. & Lewis, K. Specialized Persister Cells and the Mechanism of Multidrug Tolerance in Escherichia coli. *J. Bacteriol.* **186,** 8172–8180 (2004).
25. Schumacher, M. a *et al.* Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. *Science* **323,** 396–401 (2009).
26. Maisonneuve, E., Shakespeare, L. J., Jørgensen, M. G. & Gerdes, K. Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13206–13211 (2011).
27. Hansen, S., Lewis, K. & Vulić, M. Role of global regulators and nucleotide metabolism in antibiotic tolerance in Escherichia coli. *Antimicrob. Agents Chemother.* **52,** 2718–2726 (2008).
28. Mulcahy, L. R., Burns, J. L., Lory, S. & Lewis, K. Emergence of Pseudomonas aeruginosa strains producing high levels of persister cells in patients with cystic fibrosis. *J. Bacteriol.* **192,** 6191–6199 (2010).
29. Levin, B. R. & Rozen, D. E. Non-inherited antibiotic resistance. *Nat. Rev. Microbiol.* **4,** 556–62 (2006).
30. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517,** 455–459 (2015).
31. Davies, J, & Davies, D. Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* **74,** 417–433 (2010).
32. Falagas, M. E., Kasiakou, S. K. & Saravolatz, L. D. Colistin: The Revival of Polymyxins for

the Management of Multidrug-Resistant Gram-Negative Bacterial Infections. *Clin. Infect. Dis.* **40,** 1333–1341 (2005).
33. Walsh, C. T. & Wencewicz, T. A. Prospects for new antibiotics: a molecule-centered perspective. *J. Antibiot. (Tokyo).* **67,** 7–22 (2014).
34. Liu, Y. Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *Lancet Infect. Dis.* **16,** 161–168 (2015).
35. Olaitan, A. O., Morand, S. & Rolain, J.-M. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. *Front. Microbiol.* **5,** 643 (2014).
36. Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12,** 371–87 (2013).
37. Mahamoud, A., Chevalier, J., Alibert-Franco, S., Kern, W. V. & Pagès, J. M. Antibiotic efflux pumps in Gram-negative bacteria: The inhibitor response strategy. *J. Antimicrob. Chemother.* **59,** 1223–1229 (2007).
38. Sun, J., Deng, Z. & Yan, A. Bacterial multidrug efflux pumps: Mechanisms, physiology and pharmacological exploitations. *Biochem. Biophys. Res. Commun.* **453,** 254–267 (2014).
39. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3,** 711–721 (2005).
40. Warnes, S. L., Highmore, C. J. & Keevil, C. W. Horizontal transfer of antibiotic resistance genes on abiotic touch surfaces: Implications for public health. *MBio* **3,** 1–10 (2012).
41. Schwechheimer, C. & Kuehn, M. J. Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nat. Rev. Microbiol.* **13,** 605–19 (2015).
42. Cohen, G. N. Microbial biochemistry. *Microb. Biochem. Second Ed.* 11–12 (2011). doi:10.1007/978-90-481-9437-7
43. Piddock, L. J. V. Multidrug-resistance efflux pumps - not just for resistance. *Nat. Rev. Microbiol.* **4,** 629–36 (2006).
44. Askoura, M., Mottawea, W., Abujamel, T. & Taher, I. Efflux pump inhibitors (EPIs) as new antimicrobial agents against Pseudomonas aeruginosa. *Libyan J. Med.* **6,** 1–8 (2011).
45. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond).* **27,** 113–159 (1928).
46. Lorenz, M. G. & Wackernagel, W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* **58,** 563–602 (1994).
47. World Heath Organization model lists of essential medicines. http://www.who.int/medicines/publications/essentialmedicines/en/. (2013).
48. Fair, R. J. & Tor, Y. Antibiotics and Bacterial Resistance in the 21st Century. *Perspect. Medicin. Chem.* **6,** 25–64 (2014).
49. Wikipedia. List of Antibiotics. Available at: https://en.wikipedia.org/wiki/List_of_antibiotics.
50. Collignon, P., Powers, J. H., Chiller, T. M., Aidara-Kane, A. & Aarestrup, F. M. World Health Organization ranking of antimicrobials according to their importance in human medicine: A critical step for developing risk management strategies for the use of antimicrobials in food production animals. *Clin. Infect. Dis.* **49,** 132–41 (2009).
51. Tomasz, A. The mechanism of the irreversible antimicrobial effects of penicillins: how the beta-lactam antibiotics kill and lyse bacteria. *Annu. Rev. Microbiol.* **33,** 113–37 (1979).
52. Bush, K. & Jacoby, G. A. Updated functional classification of β-lactamases. *Antimicrob. Agents Chemother.* **54,** 969–976 (2010).
53. Bassetti, M., Ginocchio, F., Mikulska, M., Taramasso, L. & Giacobbe, D. R. Will new antimicrobials overcome resistance among Gram-negatives? *Expert Rev. Anti. Infect. Ther.* **9,** 909–22 (2011).
54. Torres, J. A., Villegas, M. V. & Quinn, J. P. Current concepts in antibiotic-resistant gram-negative bacteria. *Expert Rev. Anti. Infect. Ther.* **5,** 833–43 (2007).

55. Mazel, D. Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* **4,** 608–20 (2006).
56. Gillings, M. *et al.* The evolution of class 1 integrons and the rise of antibiotic resistance. *J. Bacteriol.* **190,** 5095–5100 (2008).
57. Walsh, T. R., Toleman, M. A., Poirel, L. & Nordmann, P. Metallo-β-Lactamases : the Quiet before the Metallo-β-Lactamases : the Quiet before the Storm ? *Clin. Microbiol. Rev.* **18,** 306–325 (2005).
58. Nordmann, P. & Poirel, L. Emerging carbapenemases in Gram-negative aerobes. *Clin. Microbiol. Infect.* **8,** 321–331 (2002).
59. Bonomo, R. A. & Szabo, D. Mechanisms of multidrug resistance in Acinetobacter species and Pseudomonas aeruginosa. *Clin.Infect.Dis.* **43 Suppl 2,** S49–S56 (2006).
60. Navia, M. M., Ruiz, J. & Vila, J. Characterization of an integron carrying a new class D beta-lactamase (OXA-37) in Acinetobacter baumannii. *Microb. Drug Resist.* **8,** 261–5 (2002).
61. Kovach, M. E. *et al.* Cloning and nucleotide sequence analysis of a Brucella abortus gene encoding an 18 kDa immunoreactive protein. *Microb Pathog* **22,** 241–246 (1997).
62. Poirel, L. *et al.* Integron-located oxa-32 gene cassette encoding an extended-spectrum variant of oxa-2 β-lactamase from Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **46,** 566–569 (2002).
63. Poirel, L., Girlich, D., Naas, T. & Nordmann, P. OXA-28, an extended-spectrum variant of OXA-10 β-lactamase from Pseudomonas aeruginosa and its plasmid- and integron-located gene. *Antimicrob. Agents Chemother.* **45,** 447–453 (2001).
64. Naas, T. & Nordmann, P. OXA-type beta-lactamases. *Curr. Pharm. Des.* **5,** 865–79 (1999).
65. Hakenbeck, R., Grebe, T., Zähner, D. & Stock, J. B. β-lactam resistance in *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-binding proteins. *Mol. Microbiol.* **33,** 673–678 (1999).
66. Chambers, H. F. Methicillin resistance in staphylococci: molecular and biochemical basis and clinical implications. *Clin. Microbiol. Rev.* **10,** 781–791 (1997).
67. Mazzariol, A., Cornaglia, G. & Nikaido, H. Contributions of the AmpC beta-lactamase and the AcrAB multidrug efflux system in intrinsic resistance of Escherichia coli K-12 to beta-lactams. *Antimicrob. Agents Chemother.* **44,** 1387–1390 (2000).
68. Hancock, R. E. W. The bacterial outer membrane as a drug barrier. *Trends Microbiol.* **5,** 37–42 (1997).
69. Noel, G. J., Bush, K., Bagchi, P., Ianus, J. & Strauss, R. S. A Randomized, Double-Blind Trial Comparing Ceftobiprole Medocaril with Vancomycin plus Ceftazidime for the Treatment of Patients with Complicated Skin and Skin-Structure Infections. *Clin. Infect. Dis.* **46,** 647–655 (2008).
70. El Solh, A. Ceftobiprole: a new broad spectrum cephalosporin. *Expert Opin. Pharmacother.* **10,** 1675–86 (2009).
71. Takeda, S., Nakai, T., Wakai, Y., Ikeda, F. & Hatano, K. In vitro and in vivo activities of a new cephalosporin, FR264205, against Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **51,** 826–830 (2007).
72. Sader, H. S., Rhomberg, P. R., Farrell, D. J. & Jones, R. N. Antimicrobial activity of CXA-101, a novel cephalosporin tested in combination with tazobactam against Enterobacteriaceae, Pseudomonas aeruginosa, and Bacteroides fragilis strains having various resistance phenotypes. *Antimicrob. Agents Chemother.* **55,** 2390–2394 (2011).
73. Queenan, A. M., Shang, W., Flamm, R. & Bush, K. Hydrolysis and inhibition profiles of β-lactamases from molecular classes A to D with doripenem, imipenem, and meropenem. *Antimicrob. Agents Chemother.* **54,** 565–569 (2010).
74. Chastre, J. *et al.* Efficacy and safety of intravenous infusion of doripenem versus imipenem in

ventilator-associated pneumonia: a multicenter, randomized study. *Crit. Care Med.* **36,** 1089–96 (2008).

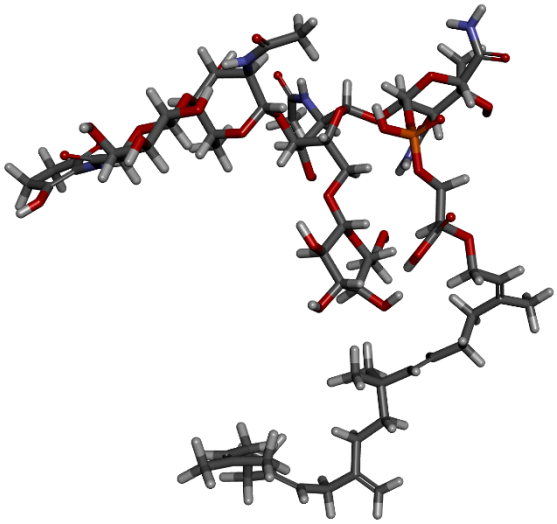
75. Fujimura, S., Nakano, Y., Sato, T., Shirahata, K. & Watanabe, A. Relationship between the usage of carbapenem antibiotics and the incidence of imipenem-resistant Pseudomonas aeruginosa. *J. Infect. Chemother.* **13,** 147–50 (2007).
76. Livermore, D. M., Mushtaq, S. & Warner, M. Activity of the anti-MRSA carbapenem razupenem (PTZ601) against Enterobacteriaceae with defined resistance mechanisms. *J. Antimicrob. Chemother.* **64,** 330–335 (2009).
77. Page, M. G. P., Dantier, C. & Desarbre, E. In vitro properties of BAL30072, a novel siderophore sulfactam with activity against multiresistant gram-negative bacilli. *Antimicrob. Agents Chemother.* **54,** 2291–2302 (2010).
78. Oleksiuk, L. M. *et al.* In vitro responses of Acinetobacter baumannii to two- and three-drug combinations following exposure to colistin and doripenem. *Antimicrob. Agents Chemother.* **58,** 1195–1199 (2014).
79. Ehmann, D. E. *et al.* Kinetics of avibactam inhibition against class A, C, and D β-lactamases. *J. Biol. Chem.* **288,** 27960–27971 (2013).
80. Butler, M. S., Blaskovich, M. A. & Cooper, M. A. Antibiotics in the clinical pipeline in 2013. *J. Antibiot. (Tokyo).* **66,** 571–91 (2013).
81. Bassetti, M., Righi, E. & Viscoli, C. Novel beta-lactam antibiotics and inhibitor combinations. *Expert Opin. Investig. Drugs* **17,** 285–96 (2008).
82. Payne, D. J. *et al.* Identification of a Series of Tricyclic Natural Products as Potent Broad-Spectrum Inhibitors of Metallo- β -Lactamases Identification of a Series of Tricyclic Natural Products as Potent Broad-Spectrum Inhibitors of Metallo- β -Lactamases. *Antimiccrobial Agents Chemother.* **46,** 1880–1886 (2002).
83. J, K.-W. *et al.* Adverse Effects of Sodium Colistimethate. Manifestations and Specific Reaction Rates During 317 Courses of Therapy. *Ann Intern Med* **72,** 857–868 (1970).
84. Ouderkirk, J. P., Nord, J. a, Turett, G. S., Ward, J. & Kislak, J. W. Polymyxin B Nephrotoxicity and Efficacy against Nosocomial Infections Caused by Multiresistant Gram-Negative Bacteria Polymyxin B Nephrotoxicity and Efficacy against Nosocomial Infections Caused by Multiresistant Gram-Negative Bacteria. **47,** 2659–2662 (2003).
85. Michalopoulos, A. S., Tsiodras, S., Rellos, K., Mentzelopoulos, S. & Falagas, M. E. Colistin treatment in patients with ICU-acquired infections caused by multiresistant Gram-negative bacteria: The renaissance of an old antibiotic. *Clin. Microbiol. Infect.* **11,** 115–121 (2005).
86. Schindler, M. & Osborn, M. Interaction of Divalent Cations and Polymyxin B with Lipopolysaccharide. *Biochemistry* **18,** 4425–30 (1979).
87. Şentürk, S. Evaluation of the anti-endotoxic effects of polymyxin-E (colistin) in dogs with naturally occurred endotoxic shock. *J. Vet. Pharmacol. Ther.* **28,** 57–63 (2005).
88. Falagas, M. E., Rafailidis, P. I. & Matthaiou, D. K. Resistance to polymyxins: Mechanisms, frequency and treatment options. *Drug Resist. Updat.* **13,** 132–138 (2010).
89. Grossman, T. H. *et al.* Target- and resistance-based mechanistic studies with TP-434, a novel fluorocycline antibiotic. *Antimicrob. Agents Chemother.* **56,** 2559–2564 (2012).
90. Kumarasamy, K. K. *et al.* Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: A molecular, biological, and epidemiological study. *Lancet Infect. Dis.* **10,** 597–602 (2010).
91. Vaara, M., Sader, H. S., Rhomberg, P. R., Jones, R. N. & Vaara, T. Antimicrobial activity of the novel polymyxin derivative nab739 tested against gram-negative pathogens. *J. Antimicrob. Chemother.* **68,** 636–639 (2013).
92. Mingeot-Leclercq, M. P., Tulkens, P. M., Denamur, S., Vaara, T. & Vaara, M. Novel

polymyxin derivatives are less cytotoxic than polymyxin B to renal proximal tubular cells. *Peptides* **35,** 248–252 (2012).

93. Ali, F. A. *et al.* Pharmacokinetics of novel antimicrobial cationic peptides NAB 7061 and NAB 739 in rats following intravenous administration. *J. Antimicrob. Chemother.* **64,** 1067–1070 (2009).
94. Vaara, M. *et al.* Novel polymyxin derivatives carrying only three positive charges are effective antibacterial agents. *Antimicrob. Agents Chemother.* **52,** 3229–3236 (2008).
95. Lewis, K. Antibiotics: Recover the lost art of drug discovery. *Nature* **485,** 439–440 (2012).

CHAPTER 6

DEVELOPING MOENOMYCIN A AS AN ANTIBIOTIC AGAINST GRAM NEGATIVE BACTERIA



Although it was discovered in 1965, the full potential of Moenomycin A (MoeA, 3D ball and stick model shown on the left[I]) has not been exploited till date. The recent surge of antimicrobial resistant bacteria has seen the revival of several underused molecules and MoeA is at the forefront. Revived interest in MoeA has allowed for a better understanding of the mechanisms of action of MoeA as well as problems and possible solutions towards developing MoeA as an antibiotic. Thus far, MoeA has only been used to target Gram positive bacteria. Herein, we present the first unique approach which enables MoeA to target the resistant Gram negative bacteria making it a potential new antibiotic against Gram negative bacteria. Subsequently, we have designed specific "delivery" peptides which bind to MoeA and enable it to cross the negatively-charged outer membrane of Gram negative bacteria. Peptides rich in Arg, Lys and/or Trp have been known for their cell permeability. This was shown in chapter 3 where we used stapled DNA binding peptides rich in Lys and Arg for enhancing cellular uptake. This is, however, the first instance where we have used short, cell permeable peptides to deliver MoeA. This has resulted in a 30 times improvement in antimicrobial activity of MoeA against the resistant Pseudomonas aeruginosa. Detailed NMR analysis gives valuable insight into the structure of the non-covalent complex between the lead delivery peptide and MoeA.
Note: Several people have contributed to this work. Their names are mentioned as and when their work is described.

6.1 Introduction

As we have seen in the previous chapter, the rise of antibiotic resistance poses a serious threat to human health. To keep up with the increasing threat of resistance, academia and industry are constantly researching for new and improved antibiotics, particularly against the resistant Gram negative bacteria. At the same time however, there is a renewed interest in exploring some of the existing validated targets using novel approaches. A notable example of a validated target is the bacterial cell wall biosynthesis pathway. The peptidoglycan biosynthetic machinery is well-established. It is now well-known that a number of these antimicrobial targets are present at or beyond the extra-cytoplasmic surface of the bacterial cell membrane. Moreover, they are also well-conserved across all bacterial species making them perfect targets for antibiotics[1, 2].

[I] Obtained from the RSCB protein data bank. The data is freely available at: http://ligand-expo.rcsb.org/pyapps/ldHandler.py?formid=cc-index-search&target=M0E&operation=ccid

In the previous chapter, we have reviewed the development of a large number of β-lactam based antibiotics. These antibiotics target the transpeptidase function of the penicillin binding proteins (PBPs)[3, 4]. Unfortunately, there has been relatively little development in the area of the essential transglycosylase (TG) function[5]. We will discuss the mechanism of the TG inhibitor MoeA in the subsequent sections. One of the reasons for the lack of research into the active site of the TG enzymes is probably because these exist at the membrane surface. This has been viewed as a difficult interface to target[6] particularly against Gram negative bacteria which possess robust outer membranes which protect the TG and prevent most molecules from accessing the inner membrane surface. Additionally, keeping the catalytic function of the enzymes in mind, the transglycosylase enzymes have long and extended active sites, which also make them difficult to target[7]. Despite these difficulties, nature has provided a perfect solution to this issue in the form of the moenomycin group of antimicrobials. The moenomycins appear to mimic the polymerised form of the substrate within the transglycosylase active site. Unfortunately, poor pharmacokinetics has disallowed the use of moenomycins in humans. Interestingly, the moenomycins (in the form of commercially available flavomycin) have been used for decades as cattle fodder with almost no reports of resistance indicating that the transglycosylase activity could be an excellent target for developing antibiotics[8].

6.2 Structure of MoeA

The moenomycins are a naturally occurring family of phosphoglycolipid antibiotics produced by *Streptomyces ghananensis*. MoeA (Figure 6.1) represents the most active compound among the moenomycins with potent antimicrobial activity[8, 9]. MoeA consists of a pentasaccharide of units B, C, D, E and F (green) with a chromophore (unit A, black). The unusual isoprenoid C25 lipid chain (shown in red) is connected to the F saccharide via a phosphoglycerate linker (blue). The C25 chain is essential for antimicrobial activity. A closer look reveals that the MoeA structure resembles that of the lipid IV (dimer of Lipid II, figure 6.1 B) product formed within the transglycosylase active site[10]. Moenomycin has amphiphilic properties due to the hydrophilic nature of the A ring and the carbohydrate units (green, rings B-F), the phosphate group of the 3-phosphoglyceric acid phosphorglycerate linker (blue) and the hydrophobic domain formed by the lipid chain (red).[11]

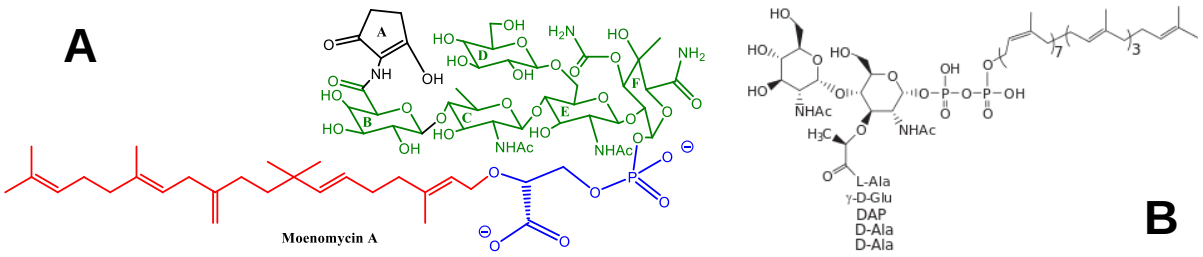


Figure 6.1: (A) Structure of MoeA and (B) Structure of Lipid II.

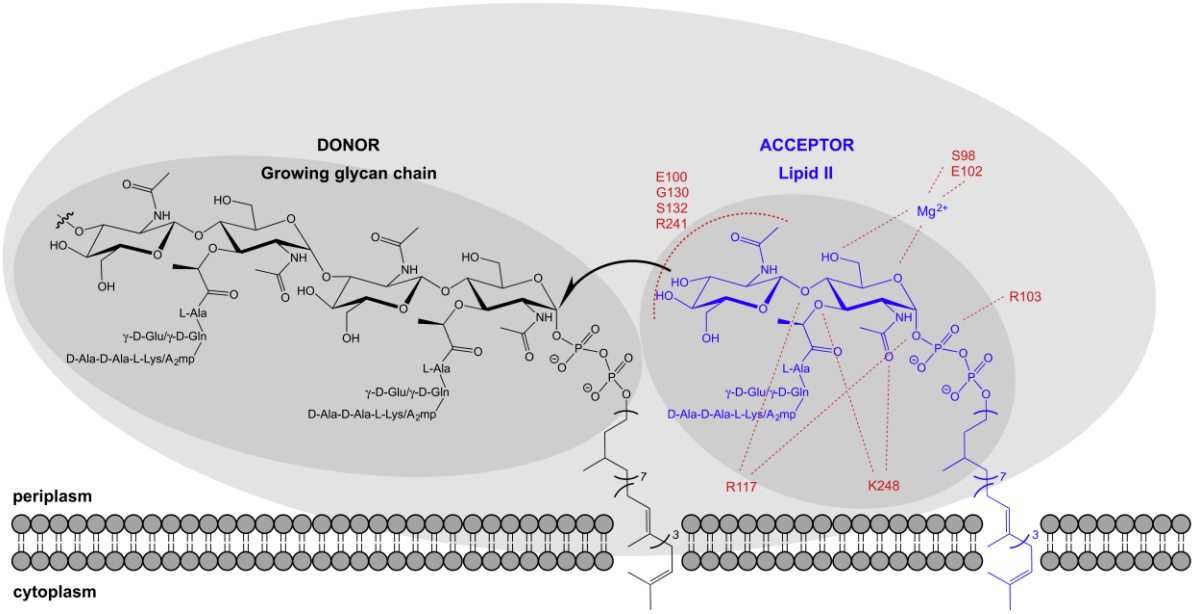
6.3 Biosynthetic pathway for the polymerisation of lipid II

The biosynthetic pathway which ultimately results in the formation of the peptidoglycan precursor lipid II and the general scheme for the polymerisation of lipid II into the peptidoglycan layer is well understood and documented. The short explanation is as follows[4]:

1. UDP-MurNAc-L-Ala-D-Glu-L-(Lys/meso-DAP)-D-Ala-D-Ala (Systematic name for UDP-MurNac: Uridine 5'Diphospho N-acetyl) or its L-lysine derivative UDP-MurNAcL-Ala-D-Glu-L-(Lys)-D-Ala-D-Ala (donor, part of the molecule shown in black in figure 6.2) is first produced in the cytoplasmic pathway.

2. It then links at the cytoplasmic membrane surface to an undecaprenyl (C55) carrier lipid prior to the addition of GlcNAc (a phosphate transferase enzyme), forming lipid II.[12]
3. This peptidoglycan precursor is then transferred to the outer surface of the cytoplasmic membrane where it is polymerised by monofunctional transglycosylases and class A bifunctional Penicillin Binding Proteins (PBPs) into long glycan chains[13] (Fig. 6.2).
4. The transpeptidase activity of Class A and B PBPs then produce inter-strand peptide cross-links from pentapeptides emanating from adjacent glycan chains.

The result is the formation of a polymer possessing high mechanical strength and rigidity sufficient to resist cytoplasmic osmotic stress. This polymer forms a scaffold for a number of extracellular structures and functions.



[II]**Figure 6.2**: Schematic diagram of the transglycoylase active site showing donor and acceptor sites. Residue numbers in the acceptor sites are in accordance with those determined for S. aureus monofunctional transglycosylase in relation to lipid II analogue as described by Huang et al.[14]

[II] Reproduced from Galley, N. F., O'Reilly, A. M. & Roper, D. I. Prospects for novel inhibitors of peptidoglycan transglycosylases. *Bioorg. Chem.* **55,** 16–26 (2014)

6.4 The inhibition of transglycosylase via MoeA: key interactions

There are two known natural molecules which bind to the transglycosylase domain of PBPs, MoeA and lipid II. Lipid II is the natural substrate while MoeA is the most potent inhibitor. The reason for this must lie in the structural differences between the two. A proper understanding of how these two molecules interact and bind to the transglycosylase domain is imperative if we wish to pursue the inhibition of transglycosylase using MoeA.

It is well known that MoeA binds with high affinity to the transglycosylase domain of several Class A PBPs. It is the most potent known inhibitor of the transglycosylase function of PBPs with MICs as low as 0.01–0.1 μg/mL[9]. The mode of inhibition by MoeA is via the direct (although reversible) binding to the active site of the transglycosylase domain preventing lipid II polymerisation. It first anchors itself to the cytoplasmic membrane via its lipid chain (which is shorter but similar to the C55 of lipid II), and then, owing to its sugar moieties, binds to the donor site of the transglycosylase.

Five motifs representing the GT51 fold-family[15] are part of the transglycosylase domain. The number of motifs may vary depending on the species but in general, these motifs are conserved among mono- and bi-functional PBPs. Six residues in the transglycosylase domain have been identified as important in the interaction with MoeA in *Aquifex aeolicus* PBP1A[16, 17] which are also conserved across other species. These 6 residues are known to bind to the F-ring and the phosphoglycerate portion of MoeA. As mentioned earlier, the transglycosylase active site is buried in the bacterial inner membrane. The lipid of MoeA is therefore required for interaction with the membrane, in order to grant access to the lipid II. It is believed that the C25 chain of MoeA interacts with the transmembrane (TM) segment of E. coli PBP1B[18], increasing the binding affinity of MoeA to the transglycosylase domain by 5-fold and highlighting the importance of the TM domain for activity. The C2 on the E ring, C3 on the F ring, the phosphoglycerate moiety and possibly the C10 region of the lipid tail are believed to be the other regions of MoeA that make essential contacts with the transglycosylase domain. The phosphoryl group and the carboxylate moiety interact with conserved active site residues of the transglycosylase domain[19]. The A ring is not essential for the interaction but provides higher binding affinity[8].

6.5 Problem, strategy and solution: Using MoeA against Gram negative bacteria

MoeA is highly active against Gram positive bacteria but largely inactive against Gram negative bacteria[8]. We are now aware that the activity is a result of the inhibition of the Glycosyltransferase enzyme (more specifically GT51) which is responsible for the peptidoglycan synthesis in the bacterial cell wall biosynthesis[4, 14]. So what makes GT51 such an attractive target? Peptidoglycan biosynthesis is mediated by Penicillin-Binding Proteins (PBPs) such as PBP1b. In *E.coli* this is a four domain protein comprising of the transglycosylase domain (GT51), transpeptidase domain (TP), transmembrane (TM) domain and UB2H domain[20]. The GT51 is an attractive target to develop new antibacterial drugs because the polysaccharide backbone (peptidoglycan) remains highly conserved in drug resistant bacteria. GT51 is essential for the survival of bacteria. It is ubiquitous to bacteria and thus confers high specificity with no mammalian couterpart[21]. Mutations in GT51 that would abolish its binding with MoeA are expected to be lethal to bacteria. This is because a number of amino acids involved in MoeA binding are also essential for catalysis of bacterial cell wall synthesis. Mutations that would decrease/abolish MoeA binding may also undermine GT51 activity. The occurrence of such mutations would not be advantageous for bacteria, thus slowing down the development of resistance. This is supported by a recent report, which described that a single point mutation in GT51 was lethal to bacteria and caused major cell division defects in *S aureus*[22].

Despite its high potency against gram positive bacteria (Minimum Inhibitory Concentration (MIC) ~1 µg/mL against *S. aureus*) and unique ability to disrupt the cell biosynthesis inhibiting GT51, interest in developing MoeA as an antibiotic was not pursued further because of its poor pharmacokinetic properties. In recent years, however, a phenomenal rise in drug resistant pathogens has once again revived interest in MoeA[8]. In this study, we have reasoned that MoeA is inactive against Gram negative bacteria because GT51 is inaccessible to MoeA due to the outer membrane barrier in Gram negative bacteria. To date, little research has been undertaken to enable MoeA to overcome this barrier in Gram negative bacteria. Polymyxin B (discussed in section 5.4.2) has been used to improve the delivery of MoeA in *E. coli* but the study was limited to only this organism[23]. Moreover, there have been no published studies with regards to MoeA delivery in resistant Gram negative bacteria. MoeA shows activity against *E. coli* BAS849 (which have damaged outer membranes[24]) as well as against stomach *H. pylori*[21] thereby providing strong evidence that the antibiotic potency of MoeA is limited by its ability to reach the GT51 target in case of the Gram negative bacteria. Additionally, it has been shown that the binding constant of MoeA to the full length PBPs isolated from 16 different species of bacteria (Gram negative and Gram positive) is in the same nM range in case of both bacteria[18]. The binding constants are indicative of the *in vitro* antimicrobial activity of MoeA. However, a similar trend is not observed for the *in vivo* model wherein MoeA is active against Gram positive bacteria and not against the Gram negative bacteria. This further suggests the inability of MoeA to reach its target GT51 in case of Gram negative bacteria. The main focus of the current work is to enable MoeA to reach its target GT51 in Gram negative bacteria by improving its delivery using a non-covalent approach. Through this approach we are exploiting the non-covalent interactions between MoeA and peptides, which result in the formation of non-covalent complexes capable of crossing the negatively charged barrier of Gram negative bacteria. Within this study, a non-covalent peptide based delivery system has been developed to enable the delivery of MoeA across the Gram negative outer membrane with the aim of improving localized MoeA concentrations in proximity to the GT51 target. The effect on activity MIC (minimal inhibitory concentration) was determined for 18 individual peptides in combination with MoeA. The structure for the peptide-MoeA complex for one of the leading compounds was determined by NMR spectroscopy.

6.6 Proposed Mechanism of delivery of MoeA in Gram negative bacteria

The surface of the outer membrane (OM) in Gram negative bacteria is overall *anionic* due to an excess of negatively-charged head-groups on the lipopolysaccharides (LPS) such as phosphates.[25] MoeA is negatively charged due to the presence of a phosphate and a carboxylate functionality, and as a result these negative charges of LPS and MoeA are expected to repel each other. For this reason, short cationic as well as amphipathic peptides have been selected in view of their known non-covalent interactions with negatively charged cell membranes (phosphates) and as they provide critical electrostatic, hydrogen bonding and hydrophobic interactions[26]. Interaction of cationic or amphipathic peptides with MoeA will mask the negative charges of MoeA thereby minimising repulsion and enabling the delivery of MoeA to the target GT51 in Gram negative bacteria. A series of short, cationic delivery peptides has been developed, which act as delivery agents. The peptide sequences are chosen based on their known ability to cross cell barriers and in certain instances deliver suitable cargoes such as proteins, drugs and fluorophores[27, 28]. They can be easily synthesized using automated Solid Phase Peptide Synthesis (SPPS) and commercially available, natural amino acids. Moreover, the syntheses do not require any special amino acid building blocks or complex steps and can therefore be done a highly cost-effective and efficient manner. MoeA is isolated in high purity from the commercially available Flavomycin using the procedure described in literature[10].

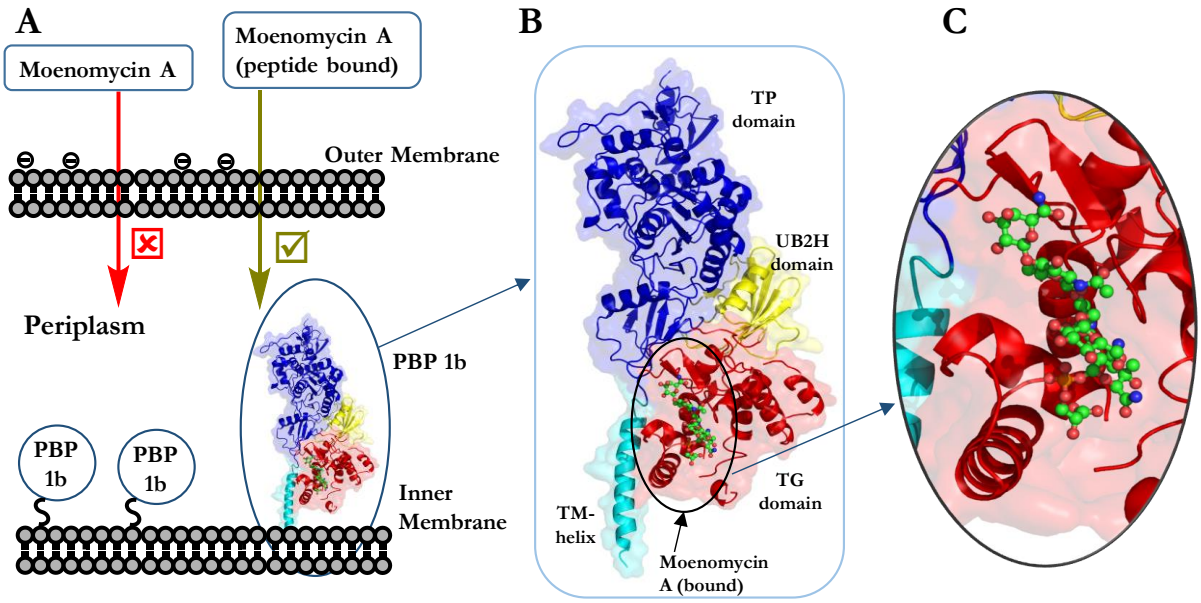


Figure 6.3: **A**. Scheme for the proposed mechanism of the delivery of MoeA using MoeA-peptide non-covalent complexes in case of Gram Negative Bacteria. **B.** The crystal structure of PBP1b is represented as a ribbon diagram. The TM, UB2H, TG, and TP domains are color coded in cyan, yellow, red, and blue respectively. All figures of 3D structural representations were made with PyMOL (www.pymol.org). **C.** Section showing Moenomycin A (bound) as a ball and stick model.

Simple mixing of these peptides with MoeA (in a 1:1 ratio in terms of moles) resulted in MoeA-peptide non-covalent complexes, the formation of which has been verified by NMR. An in depth NMR study described in subsequent sections further supports the hypothesis that MoeA-peptide complexes are able to cross the outer membrane barrier in Gram negative bacteria thereby enabling MoeA to reach its intended target – GT51 (Figure 6.3). This is the first strategy of its kind for MoeA and can be further expanded and developed into a screening platform for testing delivery capabilities of different peptides in Gram negative bacteria via a simple mixing approach.

6.7 Peptide Design and syntheses

Most antimicrobial peptides (AMPs) such as pexiganan (22 amino acids) are long and rich in positive charges. Short AMPs, such as MP196 (synthetic hexapeptides) are not active against Gram negative bacteria[29, 30]. For this study we have designed non-toxic short peptides (4-9 amino acids) as delivery systems for MoeA. Choosing longer peptides which have antimicrobial activity by themselves is not advisable in this case because that would result in increased toxicity[31].

In the current study, the residues of the delivery peptides have been tuned to obtain the most suitable peptide depending on the target bacteria. MoeA in itself is ineffective against *P. aeruginosia* and *K. pneumoniae* which are resistant to Ampicillin as well. However, it does show activity against *A. baumannii* (MIC ~4 µg/mL). In the other two cases, the most likely reason for inactivity is due to the like charge repulsion mainly between the phosphate and the carboxylate of MoeA with the phosphates of the bacterial outer membrane. This results in a non-insertion of the compound into the outer cell membrane. The hypothesis is that the cationic peptides rich in Arg[32] and Lys synthesized, known for their cell permeability, manage to mask the repulsion between MoeA and the outer cell membrane of Gram Negative Bacteria, enabling MoeA to reach its target. For the purpose of this study, the linear peptide sequences in Table 6.1 have been designed and synthesized as the most suited candidates for

the delivery of MoeA. The common denominator between all the peptides is the high percentage of cationic residues. These residues are vital for delivery of the negatively charged MoeA.

Sr. No.	Code	Peptide	# of Arg/Lys residues	Max. Theoretical Charge[a]
1	1	RRRRRRRRR	9/0	10+
2	2	KRRKRRKRR	6/3	10+
3	3	KKKKKKKKK	0/9	10+
4	4	RRRRRRRR	8/0	9+
5	5	RKKRRQRRR	6/2	9+
6	6	KKKKKKKK	0/8	9+
7	7	RRWWRRWRR	6/0	7+
8	8	KKKKKR	1/5	7+
9	9	RRRRR	5/0	6+
10	10	KKKKK	0/5	6+
11	11	RRRR	4/0	5+
12	12	KKKK	0/4	5+
13	13	RRR	3/0	4+
14	14	KKK	0/3	4+
15	15	RR	2/0	3+
16	16	KK	0/2	3+
17	17	H_2N-R-COOH	1/0	1+
18	18	H_2N-K-COOH	0/1	1+

Table 6.1: List of delivery peptides designed and synthesized (N-terminal free, C terminal amide except for entries 17 & 18). [a]Max Theoretical charge = No. of cationic residues + 1 due to the free N terminus except for entries 17 & 18.

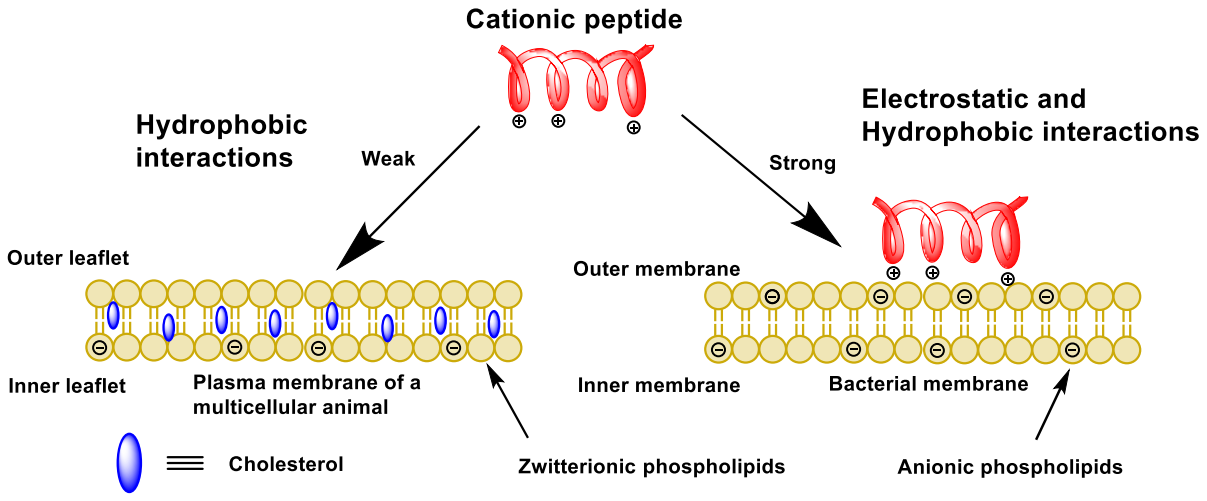


Figure 6.4: Schematic showing the membrane targeting of bacteria using cationic peptides and specificity arising due to differences in the outer membrane structure. Modified from ref. [33]

The common denominator between all these peptides is the high percentage of cationic residues. These residues are vital for the delivery of the negatively charged MoeA. An additional advantage of these cationic peptides lies in the fundamental difference in the structure and composition of the

membranes of microbes versus mammalian cells (Figure 6.4). Based on the current understanding, bacterial membranes in Gram negative bacteria are organized in such a way that the outer membrane is exposed to the environment. This membrane is heavily populated by lipids with negatively charged phospholipid head groups. In contrast, the outer leaflet of the membranes of plants and animals is composed principally of lipids with no net charge. As described by the Shai-Matsuzaki-Huang (SMH) model[34, 35, 36] most of the lipids with negatively charged head groups are segregated into the inner leaflet, facing the cytoplasm[33] (Figure 6.4). In case of the bacterial outer membrane the model proposes that the interaction of the peptide with the membrane is generally followed by displacement of lipids, alteration of membrane structure, and in certain cases entry of the peptide into the interior of the target cell. In case of plants and animals, the presence of cholesterol in the target membrane reduces the permeabilization capacity of cationic peptides in general, either due to stabilization of the lipid bilayer or due to interactions between cholesterol and the peptide.[34] We thus envisage obtaining selectivity towards bacterial membranes via this cationic peptide delivery approach.

The synthesis of hydrophobic protected peptides, particularly R8 which contains 8 Pbf protecting groups required special protocols to avoid deletion sequences. Therefore, optimized µwave assisted Fmoc Solid Phase Peptide Synthesis (SPPS) on a pure PEG based resin such as Rink Amide Chemmatrix Resin was used for the syntheses of the peptides. A hydrophilic resin backbone ensured that the sequence does not fold back on to the resin. Furthermore, DIC/Oxyma was used as the coupling mixture ensuring efficient coupling.[37] The peptides were subsequently cleaved using standard TFA conditions and purified using RP-HPLC (for detailed protocols see experimental section for chapter 6).

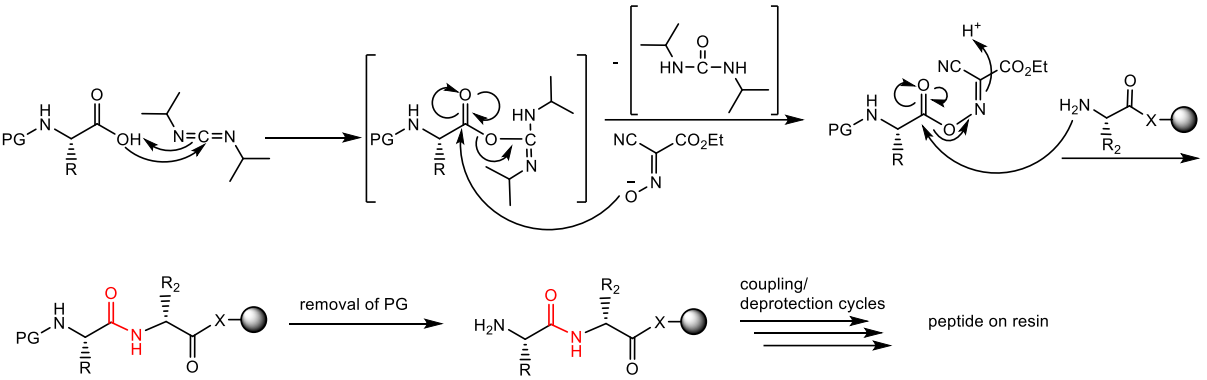


Figure 6.5: Amide bond coupling mechanism using DIC/Oxyma as a coupling mixture. PG stands for protecting group. X denotes the linker attached to the resin. The amide bond is marked in red.

6.8 Antibacterial activity: Minimum Inhibitory Concentration (MIC) values

MIC testing was performed by Charlotte S. Vincent who is thanked for her valuable contribution to this work. A summary of the MIC results are shown below:

Sr. No.	Code	Sample Name	*K. pneumoniae*	*A. baumannii*	*P. aeruginosa*
			MIC	MIC	MIC
1	19	1+Moe A	10.65	10.65	10.65
2	20	2+Moe A	5.48	5.48	5.48
3	21	3+Moe A	11.63	0.73	11.63
4	22	4+Moe A	11.24	11.24	11.24
5	23	5+Moe A	2.74	21.92	10.96
6	24	6+Moe A	12.20	0.38	12.20
7	25	7+Moe A	10.34	1.29	5.17
8	26	8+Moe A	13.36	6.68	13.36
9	27	9+Moe A	13.45	0.21	13.45
10	28	10+Moe A	28.58	0.45	14.29
11	29	11+Moe A	7.20	0.22	14.39
12	30	12+Moe A	15.16	0.24	30.32
13	31	13+Moe A	15.48	0.97	30.96
14	32	14+Moe A	32.27	1.01	32.27
15	33	15+Moe A	33.49	1.05	33.49
16	34	16+Moe A	34.50	1.08	69.01
17	35	17+Moe A	36.45	1.14	36.45
18	36	18+Moe A	37.04	1.16	74.08
19	37	Moe A	40.46	2.53	161.86
20	38	Ampicillin	GAW	-	GAW
21	39	Entries 1-18	GAW	GAW	GAW

Table 6.2: Summary of results when the delivery peptides are used with Moenomcyin in a ratio 1:1 (in terms of mM). MIC: Minimum Inhibitory concentration (in µM). GAW = Growth in all wells even at a concentration of 256 µg/mL. Strains used: *K. pneumoniae ATCC 700603, A. baumannii ATCC 19606, P. aeruginosa ATCC 27853.* Best MIC values have been marked in red.

In order to investigate the role of a peptide delivery system on MoeA activity three resistant Gram negative bacterial strains were selected namely, *K. pneumoniae*, *A. baumannii* and *P. aeruginosa*. In first instance, all the peptides were tested in the absence of MoeA for antibacterial activity and no activity was observed (Table 2, full table included in SI). Secondly, MoeA in the absence of a delivery peptide gave MICs of 40.46 µM, 2.53 µM and 161.86 µM for the three bacteria – *K. penumoniae, A. baumannii and P. aeruginosa*, respectively. The bacteria were found to be resistant to Ampicillin. Finally, a combination of 1:1 molar ratios of peptide to MoeA were tested. The main observations for each bacterium can be summarized as follows:

P. aeruginosa. Peptide **7** (Table 1) was found to be the best delivery peptide in case of *P. aeruginosa* closely followed by peptide **2** since their corresponding non-covalent complexes (formed by mixing the peptide with MoeA in ratio 1:1 in terms of moles), **2+MoeA** and **7+MoeA** (Table 6.2, entries 20 &

25) showed MIC values of 5.17 μM and 5.48 μM, respectively, which amounts to a 30 times improvement in antibacterial activity compared to MoeA alone (MIC 161.86 μM). We reasoned that the structure of polyarginine is possibly too rigid and also the fact that it binds too tightly to MoeA itself (polyarginines are particularly attracted to the phosphates as is observed in case of DNA binding peptides that all contain a significant percentage of arginines[38]) to be a suitable delivery peptide. Subsequently, insertion of Lys residues at positions 1, 3 & 6 of R9 was considered resulting in the sequence KRRKRRKRR of peptide **2**. It was indeed shown that peptide **2** shows better delivery capability for MoeA than **1**. Besides Arg and Lys, a very important residue for cell permeability is Tryptophan (Trp). Trp, has been known to interact with lipid membranes and especially within basic peptides it can trigger glycosaminoglycan-dependent endocytosis[39]. Certain Trp containing cationic peptides also have antimicrobial properties[40]. However, since Trp interacts strongly with lipids, it is a concern that it will interact strongly with the lipid tail of MoeA and render it useless. Therefore, the number of Trp residues is to be restricted in case of designing delivery peptides. We chose a Trp-mutant of the TAT peptide described in literature[40] (**7)** as the most suitable Trp containing delivery peptide for MoeA and from the MIC it is clear that **7** is indeed the lead peptide. Among the polyarginines, R9+MoeA (**19**) showed only a slight improvement in MIC as compared to R8+MoeA (**22**) with MoeA, but an overall 15 times improvement in antibacterial activity. The polylysines K9 (**3**) and K8 (**6**) also follow a similar efficiency profile with K9+MoeA (**21**) showing marginally improved activity over K8+MoeA (**24**). The truncated TAT sequence **5**+MoeA (**23**), has comparable activity to those of the polylysines and polyarginines. The complexes **8**+MoeA (**26**) and **9**+MoeA (**27**) show almost the same activity and so do **10**+MoeA (**28**) and **11**+MoeA (29). Interestingly, there is a difference of only 3.15 μM in terms of antibacterial activity for the complexes formed using R8 (**22**) and R4 (**29**). This is quite an unexpected result as R8 in particular is known to have a much higher cell penetration ability as compared to R4[41]. However, this makes the methodology quite valuable since longer Arg or Lys sequences are not required for the delivery of MoeA thereby considerably reducing the complexity of the system, meanwhile providing a simultaneous saving on material cost and synthesis time. Since the guanidium group of arginine is more basic (pKa ~12.5) than the side chain of Lysine (pKa ~10.5)[42], it was expected that R9 and R8 should have better permeabilization capacities and hence perform better delivery than K9 (**3**) and K8 (**6**)[43]. However, this was not the case as the MICs observed in case of **1**+MoeA & **4**+MoeA (**19** & **22**) as compared to **3**+MoeA & **6**+MoeA (**21** & **24**) are almost the same.

K. pneumoniae. Peptide **5**, truncated from the original TAT sequence (47-57) thereby resulting in the new sequence RKKRRQRRR was found to be the best delivery peptide in case of *K. pneumoniae* with its corresponding complex with MoeA (**5+MoeA**, **23**) giving an MIC of 2.74 μM, a 15 times improvement compared to MoeA alone. Two other impressive delivery peptides were found in case of *K. pneumoniae*, namely KRRKRRKRR (**2**) & R_4 (**11).** Their corresponding complexes with MoeA gave MICs of 5.48 μM (~7.3 times increase) and 7.2 μM (~5.5 times increase), respectively. Peptide **7** RRWWRRWRR in combination with MoeA (**25**) shows a ~4 times improvement in activity with an MIC of 10.34 μM closely followed by R9 **1**+MoeA (**19**) with an MIC of 10.65 μM.

A. baumannii. It is to be noted that MoeA in itself gives an impressive MIC of 2.53 μM in *A baumannii*. This value further improves with certain peptides. To summarize, we have seen a ~10 times improvement in MIC with values of 0.21 μM, 0.22 μM and 0.24 μM when the delivery peptides R5 (**9**), R4 (**11**) and K4 (**12**) were used in combination with MoeA (**27**, **29** and **30**), respectively. Surprisingly K8+MoeA (**24**) showed almost twice as much activity as K9+MoeA (**21**). The shorter peptide K5 (**10**) + MoeA (**28**) also showed an impressive 5.5 times improvement in antibacterial activity. Peptide sequences having 4 or lower than 4 positive charges show a comparable MIC value

with an overall 2 times improvement in MIC as can be seen from Table 2 entries **31**-3**6**. We thus have a large number of peptide-MoeA non-covalent complexes with excellent activity against *A baumannii* which is critical for antibiotic development since *A. baumannii* has developed resistance against almost all known classes of antibiotics.[44] The MIC trend when different peptides are used in combination with MoeA is thus quite interesting. Shorter sequences have shown an improvement in MIC as discussed above, whereas longer sequences such as R9 (**1**), KRRKRRKRR (**2**), R8 (**4**) and the truncated TAT sequence (**5**) have actually resulted in poorer MIC values. Rigidity and phosphate binding once again could be an explanation for this discrepancy in expected trend of MIC values particularly in the cases of R9 (**1**) and R8 (**4**).

Based on the MIC results summarized in table 6.2 we can conclude that at least 4 positively charged residues are required in the peptide sequence for effective delivery of MoeA. The higher number of charges are required for the efficient neutralization of the two negative charges of MoeA. We can also conclude that Trp is an important residue if activity is required in combination with a lower number of positive charges. A balance of Arg and Lys residues is required when targeting *P. aeruginosa* and *K. pneumoniae*, however, *A. baumannii* has higher affinity for Lys than Arg. Therefore, substitution of Arg with Lys permits the selective targeting of this bacterial species.

We have established that the inability of MoeA to cross the outer bacterial membrane and reach its target GT51 situated on the periphery of the inner membrane is the main reason for its inactivity. We have also shown that through the use of bacterial membrane permeable peptides, MoeA can be made effective against Gram negative bacteria. It is obvious that the higher the delivery capability of a peptide in a specific strain of bacteria, the better the delivery of MoeA. Increased antibiotic concentration results in a lower MIC. Although the MIC will also depend on the binding capacity of the peptides to MoeA, the MICs of peptide-MoeA complexes still provide a good indication of the delivery capability of peptides into bacterial membranes. In order to check for permeability into cells (as described in chapter 4) fluorescent labelling is almost always required. Currently, to the best of our knowledge, there are almost no general screening protocols capable of testing the permeability of peptides through gram negative bacterial membranes without labelling. Through the formation of peptide-MoeA complexes and measurement of MICs, we have essentially established protocols for the screening of peptides in terms of their delivery capacities without the requirement of any labelling.

6.9 Cytotoxicity Studies

Timea Palmai Pallag is thanked for her valuable contribution for testing the compounds for toxicity.

One of the key aspects in designing a delivery peptide is low toxicity. Therefore, we tested all the lead peptides without (refer experimental section) and in combination with MoeA (Figure 6.6) against the HeLa cell line and did not observe any significant toxicity (relative survival 70-90%) until a concentration of 100 µM. It should be noted that the Trp containing peptide **7** is more toxic compared to the other peptides at a 100 µM concentration.

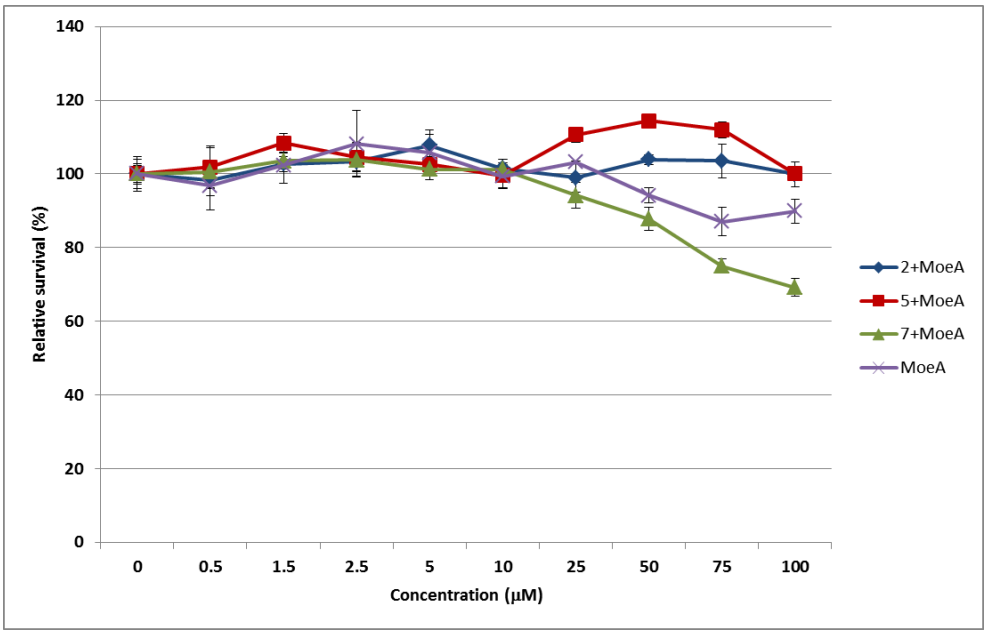


Figure 6.6. Toxicity results showing relative survival vs. Concentration (in µM).

6.10 NMR studies: Decoding the molecular structure of the MoeA-Peptide 7 non-covalent complex

*Dr. Stephen Prior is thanked for his invaluable contribution in performing the NMR experiments on the MoeA-peptide **7** non-covalent complex, structural calculations, figures, the structural determination of the complex and its description.*

In order to better explain the reasons behind the activity of the MoeA-peptide non-covalent complex, extensive NMR studies were performed, followed by detailed structural calculations. For this purpose, the MoeA-peptide **7** complex was chosen as it is one of the leading non-covalent complexes capable of targeting all three Gram negative bacteria.

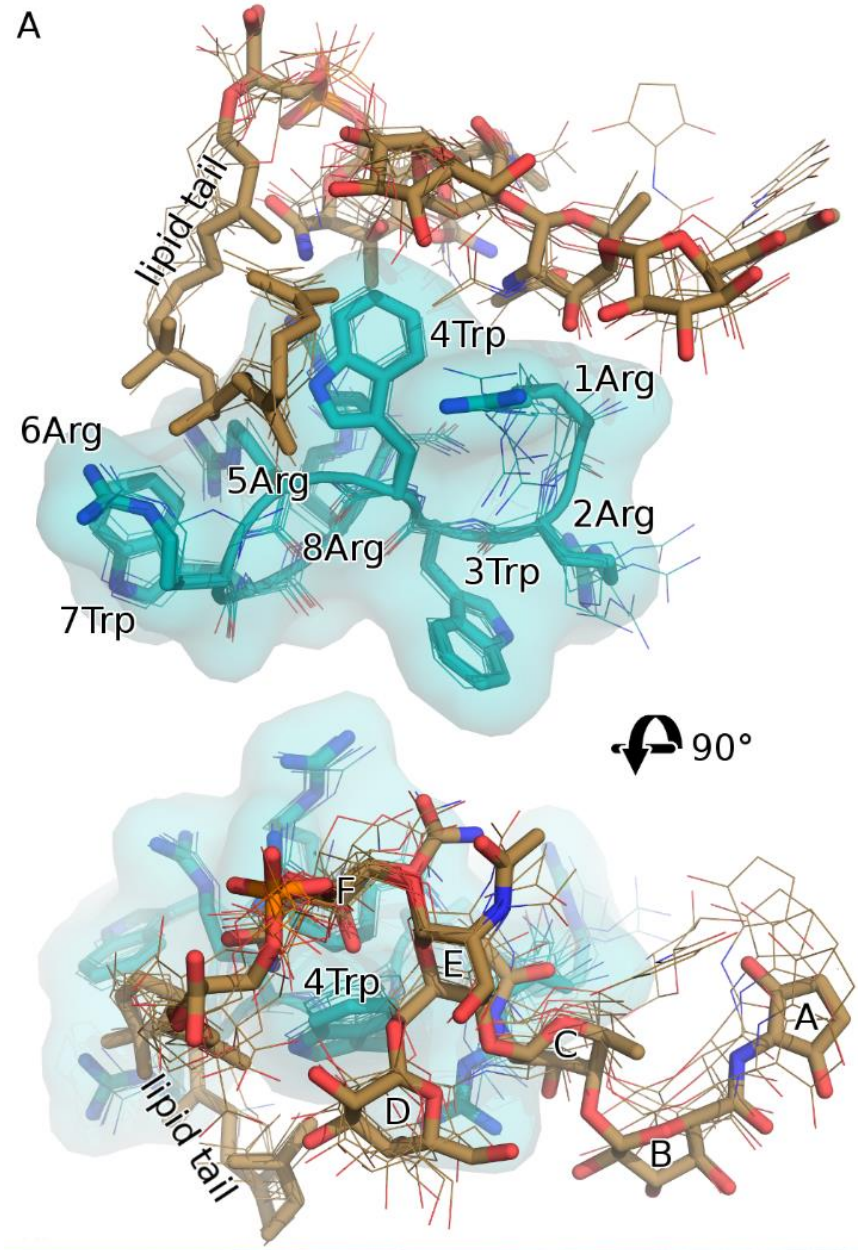
The structures of MoeA and peptide **7** were fully assigned at both 4 and 27°C (see experimental section for chapter 6, table 3 for chemical shifts). By following the linewidth of well-resolved protons throughout the titration of MoeA into peptide, the complex was shown to be in intermediate exchange, thus severely hampering spectral interpretation. Cooling the sample down to 4°C was found to move the complex into the fast to intermediate exchange regime. This sharpened the peaks considerably (for instance the NOE cross-peak between 4TrpHε1 and 4TrpHε3/Hζ2 sharpened from a half-height peak width of ~65 Hz at 27°C to ~20 Hz at 4°C) and therefore enabled the detection of significantly more

NOEs (Figure S36). Despite the relatively broad peaks ~600 NOEs were unambiguously assigned of which ~100 were intermolecular (Table 3), which was sufficient to allow a full structural characterisation of the complex (Figure 6.6A).

The resulting ensemble of the complex converged to a pairwise ordered-atom RMSD of 1.3 Å (experimental section for chapter 6, table 3 and table 4), with the MoeA component accounting for the majority of this variation, with an RMSD of 1.5 Å. The well-ordered atoms include the heavy atoms of the peptide; and sugars C, D, E and F, and the lipid tail of MoeA. Sugars A and B were found to be disordered. This is not surprising for a large flexible molecule involved in a complex of moderate affinity, and the structural flexibility of MoeA has been observed previously.[11] The peptide, by contrast, was extremely well resolved (RMSD of 0.9 Å). The complex is of sufficient resolution to identify the nature of the interaction. Central to this interaction is the indole ring of 4Trp which exhibits a number of strong NOEs to the lipid tail of MoeA (Figure S36). In the final structure this indole ring is observed to bury itself into a hydrophobic pocket formed by the lipid tail and sugars F, E and D of MoeA (Figure 4B). This buries a total of 92.9 ± 12.3 $Å^2$ at the interface. The interaction is further stabilised by hydrogen bonds between 1Arg and 9Arg of the peptide and sugars C and F (Figure 4 B), which provide the hydrophobic pocket with a slightly polar rim. Sugar A was not involved in any intermolecular interactions and therefore considerably more disordered in solution. Quite surprisingly the charged phosphate moiety of MoeA plays no part in the interaction.

It is an interesting observation that the MoeA-distal surface of the peptide is almost perfectly planar and projects a number of arginine and tryptophan side chains, thus putting these side chains in a perfect orientation to interact with a cell membrane: tryptophan and arginine are often involved in membrane recognition events[45], and could therefore facilitate trafficking of the MoeA across the bacterial membrane.

A schematic representation of the MoeA-peptide **7** complex has been displayed in figure 6.8 through which it is clear that the hydrophobic interaction between the side chain of the Trp residue and rings C, E and F and lipid tail of MoeA (figure 5) are the most dominant interactions of the complex. Hydrogen bonding between the N-terminal Arg and ring B of MoeA and the *C*-terminal Arg with the lipid tail and ring D further stabilize the MoeA-peptide non-covalent complex. A third hydrogen bond is observed between the C terminal amide and D ring of MoeA. Furthermore, the phosphate is a non-participant and is far away from the MoeA-peptide interface. This result provides a plausible explanation for the comparable activities of **11+MoeA**, **1+MoeA** and **4+MoeA**. Longer, polyarginine peptide sequences containing several basic residues show significantly higher phosphate binding affinity and could, in theory, also bind to the phosphate of MoeA. Since the phosphate is essential for activity, increasing the number of arginines does not ensure a better delivery peptide. It could also explain why Lys residues in combination with Arg show better activity (refer table 2, entry **1+MoeA** & **2+MoeA**) as, although polylysine is equally prone to unspecific interactions, Lys is less basic than Arg and consequently binds less tightly, thereby making these interactions reversible indicating a possibility of MoeA release once inside the periplasm. Trp in certain instances has been known to influence the binding of the peptide to lipids.[46] Moreover, given a polar protic medium like water, hydrophobic interactions have been known to have significant effect in Trp containing cationic peptides[47]. A similar result has been observed in this case making Trp containing peptides valuable for the delivery of MoeA.



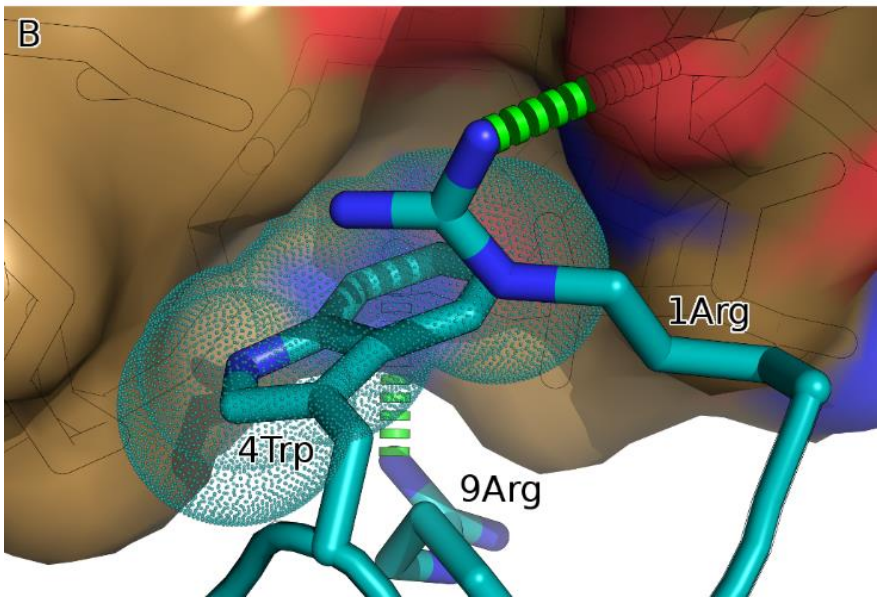


Figure 6.7: Solution structure of complex formed between peptide and Moenomycin A.

(A) NMR ensemble of 7 lowest energy conformers derived from 600 observed NOEs. Ensemble converged to an ordered atom pairwise RMSD of 1.33 Å. Peptide shown as *teal lines*; MoeA by *gold lines*. The conformer most structurally similar to the ensemble average is shown in *sticks*, and the peptide as *surface*. Sugars A and B were found to be disordered in solution. For clarity, hydrogen atoms have been omitted.

(B) The lipid tail and phosphate-proximal sugar rings of MoeA form a hydrophobic pocket with a relatively polar rim into which 4Trp of the peptide inserts, which is stabilised by hydrogen bonds donated by 1Arg and 9Arg to the polar rim. Peptide shown as *teal sticks*; MoeA by *gold surface*; hydrogen bonds as *green dashes*; and hydrophobic contacts (C-C distance < 3.9 Å) as *teal dots*.

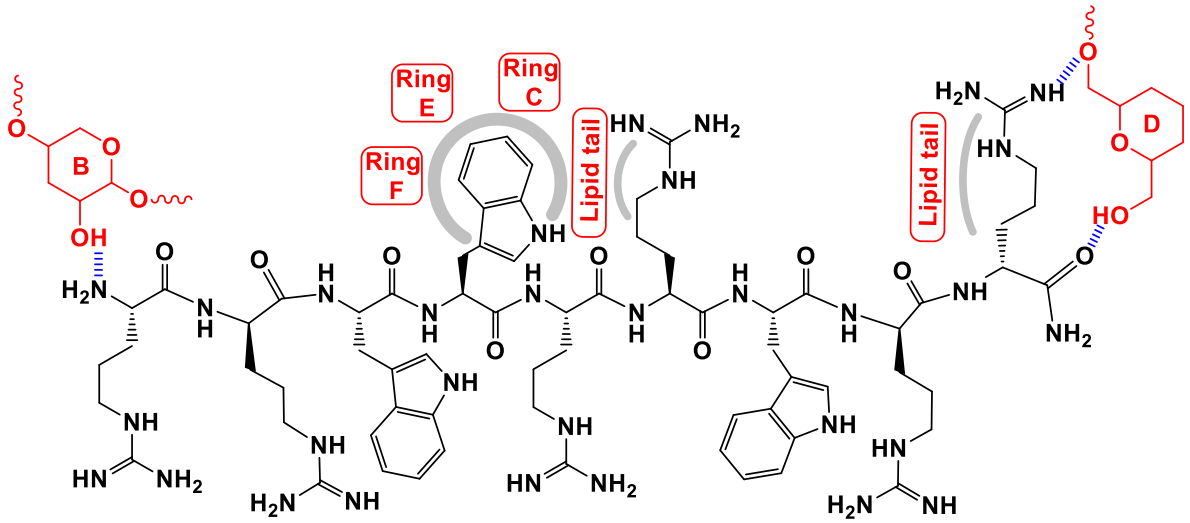


Figure 6.8. Schematic representation of observed contacts between peptide **7** and MoeA. Peptide **7** (from N to C, left to right) is shown as black lines; MoeA moeities are shown in red; hydrogen bonds are shown as blue stipples; areas of hydrophobic contact are shown as grey arcs.

A schematic showing the MoeA-peptide **7** complex has been shown in figure 6.8. From the schematic, one can clearly observe that the hydrophobic interaction between the side-chain of the Trp residue and the rings C, E and F and lipid tail of MoeA (figure 6.8) are the most dominant interactions of the complex. Hydrogen bonding between the N-terminal Arg and ring B of MoeA and the C-terminal Arg with the lipid tail and ring D further stabilize the MoeA-peptide non-covalent complex. A third hydrogen bonding is observed between the C terminal amide and D ring of MoeA. Furthermore, the phosphate is a non-participant and is far away from the MoeA-peptide interface. Since the phosphate is essential for the activity of MoeA, it could explain why having several basic residues, as in case of R_8 and R_9, does not improve the activity beyond a certain limit, as these peptides could, in theory, bind to the phosphate, thereby destroying the activity of MoeA. It could also explain why Lys residues in combination with Arg show better activity (refer table 2, entry **1+MoeA** & **2+MoeA**) as although polylysine is equally prone to unspecific interactions, Lys binds less tightly being a monodentate ligand as compared to Arg which is a bidentate ligand. Therefore, interactions involving lysine can be reversible, indicating a possibility of MoeA release once inside the periplasm. Trp, in certain instances, has been known to influence the binding of the peptide to lipids[46]. Moreover, given a polar, protic medium like water, hydrophobic interactions have been known to have significant effect in Trp containing cationic peptides.[47] A similar result has been observed in this case, making Trp containing peptides valuable for the delivery of MoeA.

A summary of the structural statistics of the MoeA-peptide 7 complex is mentioned here:

Pairwise RMSD of complex components	
Peptide heavy atom (Å)	0.91 ± 0.26
Moenomycin	
Core atoms (Å)	1.48 ± 0.23
Heavy atom[III] (Å)	1.81 ± 0.43
Complex	
Core atoms (Å)	1.33 ± 0.17
Heavy atom (Å)	1.62 ± 0.34
Experimental restraints	
Peptide NOEs	397
Moenomycin NOEs	112
Complex NOEs	
Intermolecular	98
Total	607
Restraint violations > 0.3 Å	3
RMSD from idealised covalent geometry	
Bond lengths (Å)	0.01 ± 0.00
Bond angles (°)	1.41 ± 0.05
Ramachandran analysis (determined by PROCHECK[48])	
Allowed regions (%)	93.9
Generously allowed regions (%)	6.1
Statistics of overall structural quality	
PROCHECK *G*-factor	-0.33 ± 0.05
ProSA[49] *Z*-score	0.81 ± 0.08
MolProbity[50] clash score	4.82 ± 0.09

Table 6.3. Structural statistics of peptide/MoeA complex ensemble

6.11 Conclusion

We have successfully identified poor delivery of MoeA as one of the key problems for its lack of activity against resistant Gram negative bacteria. Subsequently, we have designed a novel peptide-based, non-covalent delivery system specifically for MoeA, which enables for the first time in over five decades to target three resistant Gram Negative bacterial strains – *K. pneumoniae, A. baumannii and P. aeruginosa.* Furthermore, we have shown that although all three are Gram negative bacterial species with similar robust outer membranes, they behave quite differently towards the same set of peptide-MoeA non-covalent complexes. We have tuned the peptides in order to obtain up to 30 times improvement in MIC values. Until now, it has been believed that longer, synthetically more challenging polyarginine sequences are required for drug delivery[51]. We have shown, for the first time that MoeA delivery can be achieved effectively even through shorter sequences such as R4. By designing different delivery peptides (which are combined with MoeA via a simple mixing approach) we have demonstrated the selectivity of our approach to target a specific bacterial strain. Peptides **9 (**R5**), 11** (R4) & **12** (K4) promote the delivery of MoeA best in *A. baumannii*, whereas **5**

[III] Core atoms are well ordered moieties. These are: the heavy atoms of the peptide; and sugars C, D, E and F, and the lipid tail of MoeA. Sugars A and B were found to be disordered.

(RKKRRQRRR) is selective for MoeA delivery in *K. pneumoniae*. Peptide **2** (KRRKRRKRR) on the other hand can perform delivery in all 3 species of bacteria simultaneously. **7** (RRWWRRWRR)+MoeA shows excellent activity in *P. aeruginosa* and moderate to high activity in *A. baumannii* and *K. pneumoniae*. As toxicity is one of the key factors in case of delivery, we have established that all peptides as well as the MoeA-peptide non-covalent complexes they form are not toxic towards mammalian cell lines up to a tested concentration of 100 μM.

We have performed a detailed NMR analysis followed by extensive structural calculations revealing for the first time the molecular structure of a peptide-MoeA complex at an extremely impressive 1.3 Å resolution. We have proven that the two components interact in a 1:1 ratio further supporting the mechanism that the peptides bind to MoeA and then enable its delivery in Gram negative bacteria. We were thus able to deduce the key interactions between peptide **7** and MoeA. Surprisingly, we observe that the phosphate is not part of the interaction showing an example where although arginines and phosphates are present in the same solution, they need not necessarily interact with one another. Moreover, a simple mixing approach provides a platform to test the delivery capability of different peptides using MoeA without any conjugation.

Thus, we believe this work to be crucial for any further development of MoeA especially in the case of resistant Gram negative bacteria. In a time where new and novel antibiotics particularly against these resistant bacteria are critical, this work lays the foundation for the development of new antibiotics based on MoeA.

6.12 References:

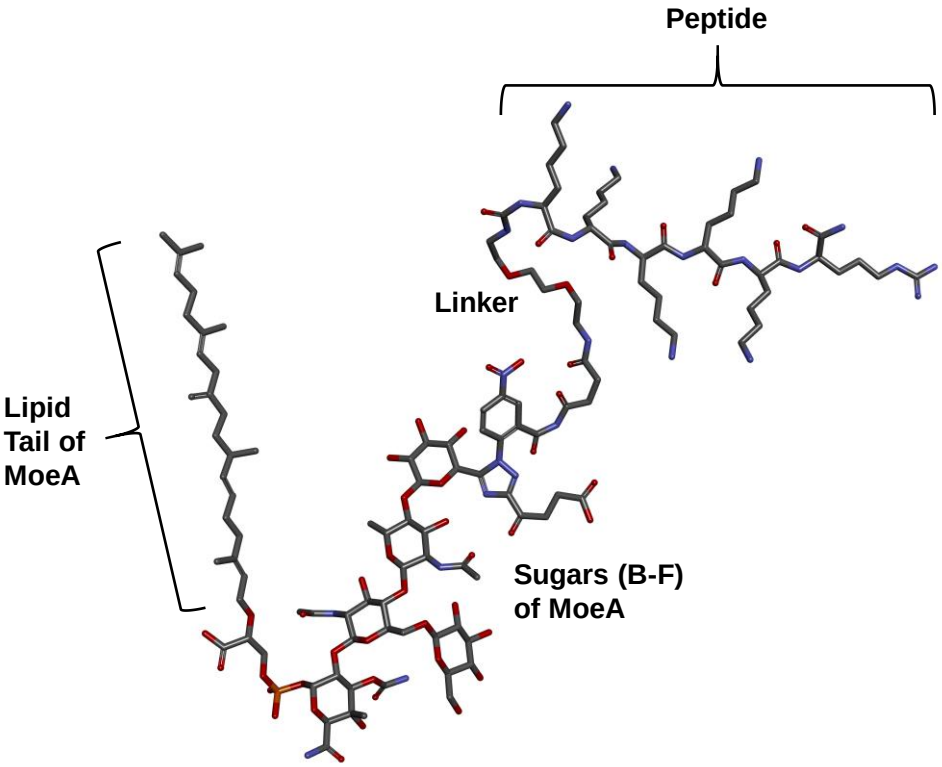
1. Bugg, T. D. H., Braddick, D., Dowson, C. G. & Roper, D. I. Bacterial cell wall assembly: Still an attractive antibacterial target. *Trends Biotechnol.* **29,** 167–173 (2011).
2. Silver, L. L. Challenges of Antibacterial Discovery. *Clin. Microbiol. Rev.* **24,** 71–109 (2011).
3. Page, M. P., Dougherty, T. J. & Pucci, M. J. *Beta-lactam antibiotics, in Antibiotic Discovery and Development, Springer, US, pp. 79-117*. (2012).
4. Galley, N. F., O'Reilly, A. M. & Roper, D. I. Prospects for novel inhibitors of peptidoglycan transglycosylases. *Bioorg. Chem.* **55,** 16–26 (2014).
5. Halliday, J., McKeveney, D., Muldoon, C., Rajaratnam, P. & Meutermans, W. Targeting the forgotten transglycosylases. *Biochem. Pharmacol.* **71,** 957–967 (2006).
6. Lovering, A. L., Gretes, M. & Strynadka, N. C. Structural details of the glycosyltransferase step of peptidoglycan assembly. *Curr. Opin. Struct. Biol.* **18,** 534–543 (2008).
7. Lovering, A. L., Safadi, S. S. & Strynadka, N. C. J. Structural perspective of peptidoglycan biosynthesis and assembly. *Annu. Rev. Biochem.* **81,** 451–78 (2012).
8. Ostash, B. & Walker, S. Moenomycin family antibiotics: chemical synthesis, biosynthesis, and biological activity. *Nat. Prod. Rep.* **27,** 1594–1617 (2010).
9. Welzel, P. Syntheses around the Transglycosylation Step in Peptidoglycan Biosynthesis. *Chem. Rev.* **105,** 4610–4660 (2005).
10. Adachi, M. *et al.* Degradation and reconstruction of moenomycin A and derivatives: Dissecting the function of the isoprenoid chain. *J. Am. Chem. Soc.* **128,** 14012–14013 (2006).
11. Kurz, M., Guba, W. & Vértesy, L. Three-dimensional structure of moenomycin A--a potent inhibitor of penicillin-binding protein 1b. *Eur. J. Biochem.* **252,** 500–7 (1998).
12. Barreteau, H. *et al.* Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32,** 168–207 (2008).
13. Mattei, P. J., Neves, D. & Dessen, A. Bridging cell wall biosynthesis and bacterial

morphogenesis. *Curr. Opin. Struct. Biol.* **20,** 749–755 (2010).
14. Huang, C.-Y. *et al.* Crystal structure of Staphylococcus aureus transglycosylase in complex with a lipid II analog and elucidation of peptidoglycan synthesis mechanism. *Proc. Natl Acad. Sci. USA* **109,** 6496–6501 (2012).
15. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77,** 521–55 (2008).
16. Yuan, Y. *et al.* Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *Pnas* **104,** 5348–5354 (2007).
17. Yuan, Y. *et al.* Structural analysis of the contacts anchoring moenomycin to peptidoglycan glycosyltransferases and implications for antibiotic design. *ACS Chem. Biol.* **3,** 429–36 (2008).
18. Rachel Cheng, T.-J. *et al.* Domain requirement of moenomycin binding to bifunctional transglycosylases and development of high-throughput discovery of antibiotics. *Pnas* **105,** 431–436 (2008).
19. Fuse, S. *et al.* Functional and structural analysis of a key region of the cell wall inhibitor moenomycin. *ACS Chem. Biol.* **5,** 701–11 (2010).
20. Sung, M. *et al.* Crystal structure of the membrane-bound bifunctional transglycosylase PBP1b from Escherichia coli. *Proc. Natl. Acad. Sci.* **106,** 8824–8829 (2009).
21. Tseng, Y. Y. *et al.* Development of bacterial transglycosylase inhibitors as new antibiotics: Moenomycin A treatment for drug-resistant Helicobacter pylori. *Bioorganic Med. Chem. Lett.* **24,** 2412–2414 (2014).
22. Rebets, Y. *et al.* Moenomycin resistance mutations in Staphylococcus aureus reduce peptidoglycan chain length and cause aberrant cell division. *ACS Chem. Biol.* **9,** 459–467 (2014).
23. El-Abadla, N. *et al.* Moenomycin A: The role of the methyl group in the moenuronamide unit and a general discussion of structure-activity relationships. *Tetrahedron* **55,** 699–722 (1999).
24. Baizman, E. R. *et al.* Antibacterial activity of synthetic analogues based on the disaccharide structure of moenomycin, an inhibitor of bacterial transglycosylase. *Microbiology* **146,** 3129–3140 (2000).
25. Beveridge, T. J. Sructure of fram-negative cell walls and their derived mebrane vesicles. *J. Bacteriol.* **181,** 4725–4733 (1999).
26. Zaro, J. L. & Shen, W. C. Cationic and amphipathic cell-penetrating peptides (CPPs): Their structures and in vivo studies in drug delivery. *Front. Chem. Sci. Eng.* **9,** 407–427 (2015).
27. Futaki, S. *et al.* Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J. Biol. Chem.* **276,** 5836–40 (2001).
28. Lättig-Tünnemann, G. *et al.* Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides. *Nat. Commun.* **2,** 453 (2011).
29. Wenzel, M. *et al.* Small cationic antimicrobial peptides delocalize peripheral membrane proteins. *Proc. Natl Acad. Sci. USA* **111,** E1409-18 (2014).
30. Ramesh, S., Govender, T., Kruger, H. G., de la Torre, B. G. & Albericio, F. Short AntiMicrobial Peptides (SAMPs) as a class of extraordinary promising therapeutic agents. *J. Pept. Sci.* **22,** 438–451 (2016).
31. Seo, M.-D., Won, H.-S., Kim, J.-H., Mishig-Ochir, T. & Lee, B.-J. Antimicrobial Peptides for Therapeutic Applications: A Review. *Molecules* **17,** 12276–12286 (2012).
32. Fretz, M. M. *et al.* Temperature-, concentration- and cholesterol-dependent translocation of L- and D-octa-arginine across the plasma and nuclear membrane of CD34+ leukaemia cells. *Biochem. J.* **403,** 335–342 (2007).

33. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415,** 389–395 (2002).
34. Matsuzaki, K. Why and how are peptide-lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. *Biochim. Biophys. Acta - Biomembr.* **1462,** 1–10 (1999).
35. Yang, L., Weiss, T. M., Lehrer, R. I. & Huang, H. W. Crystallization of antimicrobial pores in membranes: magainin and protegrin. *Biophys. J.* **79,** 2002–9 (2000).
36. Shai, Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim. Biophys. Acta - Biomembr.* **1462,** 55–70 (1999).
37. Subirós-Funosas, R., Khattab, S. N., Nieto-Rodríguez, L., El-Faham, A. & Albericio, F. Advances in acylation methodologies enabled by oxyma-based reagents. *Aldrichimica Acta* **46,** 21–44 (2013).
38. Milletti, F. Cell-penetrating peptides: classes, origin and current landscape. *Drug Discov. Today* **17,** 850–860 (2012).
39. Bechara, C. *et al.* Tryptophan within basic peptide sequences triggers glycosaminoglycan-dependent endocytosis. *FASEB J.* **27,** 738–749 (2013).
40. Bauke Albada, H. *et al.* Modulating the activity of short arginine-tryptophan containing antibacterial peptides with N-terminal metallocenoyl groups. *Beilstein J. Org. Chem.* **8,** 1753–1764 (2012).
41. Langel, Ü. *Handbook of Cell-Penetrating Peptides, Second Edition, page 478*. (2006).
42. Amino Acids Reference Chart Sigma-Aldrich. Available at: http://www.sigmaaldrich.com/life-science/metabolomics/learning-center/amino-acid-reference-chart.html.
43. Cutrona, K. J., Kaufman, B. A., Figueroa, D. M. & Elmore, D. E. Role of arginine and lysine in the antimicrobial mechanism of histone-derived antimicrobial peptides. *FEBS Lett.* (2015). doi:10.1016/j.febslet.2015.11.002
44. Fair, R. J. & Tor, Y. Antibiotics and Bacterial Resistance in the 21st Century. *Perspect. Medicin. Chem.* **6,** 25–64 (2014).
45. Rydberg, H. A. *et al.* Peptide-membrane interactions of arginine-tryptophan peptides probed using quartz crystal microbalance with dissipation monitoring. *Eur. Biophys. J.* **43,** 241–253 (2014).
46. Jin, Y. *et al.* Influence of tryptophan on lipid binding of linear amphipathic cationic antimicrobial peptides. *Biochemistry* **42,** 9395–9405 (2003).
47. Jin, Y. *et al.* Influence of Tryptophan on Lipid Binding of Linear Amphipathic Cationic Antimicrobial Peptides. *Biochemistry* **42,** 9395–9405 (2003).
48. Laskowski, R., Rullmann, J. A., MacArthur, M., Kaptein, R. & Thornton, J. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** 477–486 (1996).
49. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35,** 407–410 (2007).
50. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).
51. Rothbard, J. B. *et al.* Arginine-rich molecular transporters for drug delivery: Role of backbone spacing in cellular uptake. *J. Med. Chem.* **45,** 3612–3618 (2002).

CHAPTER 7

PEPTIDE-MOENOMYCIN A CONJUGATES AS MULTIFUNCTIONAL ANTIBIOTICS AGAINST GRAM NEGATIVE BACTERIA



In the previous chapter, we described a novel approach for the delivery of Moenomycin A (MoeA) in Gram negative bacteria and proposed a plausible explanation for the mechanism of action. In this chapter, we will take a look at some of the chemical modifications of MoeA described in literature, along with resistance to MoeA. We also describe an optimized synthetic route which allows, for the first time, covalent linkage of peptides to MoeA. A simple visualization of a peptide-MoeA covalent conjugate is shown on the left. The conjugate retains the structure of MoeA and the peptide is linked through MoeA via a suitable linker. Other molecules conjugated to MoeA have, in most instances, resulted in a decrease in antimicrobial activity against Gram positive bacteria. However, in this case, we have seen ~30 times improvement in antimicrobial activity of the peptide-MoeA conjugate against the resistant Pseudomonas aeruginosa.

7.1 Introduction

The results described in Chapter 6 show that MoeA can be used as an antibiotic against the resistant Gram negative bacteria. Other aspects need to be considered however if MoeA is to be developed into an antibiotic, such as resistance to MoeA particularly in case of Gram negative bacteria. As described in chapter 5, these bacteria have a tendency to develop resistance much more quickly than Gram positive pathogens. The other problem is binding of MoeA to plasma proteins which results in a low unbound fraction, decreased potency and poor pharmacokinetic properties of MoeA making it unsuitable for human use. In this chapter we will describe the covalent linkage of suitable peptides to MoeA resulting in a single molecule. Through this approach we envisage to further improve the potency of MoeA against Gram negative bacteria and reduce its plasma binding capability thereby increasing the scope for therapeutic applications.

7.2 Structure-activity relationships of MoeA

Through selective degradation of MoeA and the synthesis of di- and trisaccharide analogues, it is now believed that sugars C, E & F, the 3-phosphoglyceric acid and the lipid tail constitute the minimum pharmacophore of MoeA (Figure 7.2, highlighted in blue)[1, 2, 3]. The degradation of MoeA to chemical entities containing the carbohydrate units C, E and F can be performed. These molecules retain the transglycosylase inhibition and antibacterial activity[4, 5]. Degradation with removal of the C unit in addition to the A and B units to retain only the E and F carbohydrate units, still allows for transglycosylase inhibition but no antibacterial activity[2]. The C25 lipid chain is crucial for obtaining maximum anti-bacterial activity of MoeA. Unfortunately, it also results in an unusually long half-life in addition to poor bioavailability and making it incompatible as an antibiotic for humans[6]. The inhibitory ability of MoeA is directly proportional to the decrease in the length of the lipid chain[5], most likely as a result of a loss of ability to anchor itself into the cytoplasmic membrane. Modifying or truncating the lipid chain has been known to improve pharmacokinetic properties. However, the loss of activity needs to be compensated by maintaining essential polar active-site contacts or by the incorporation of other hydrophobic chemophores with suitable properties[7].

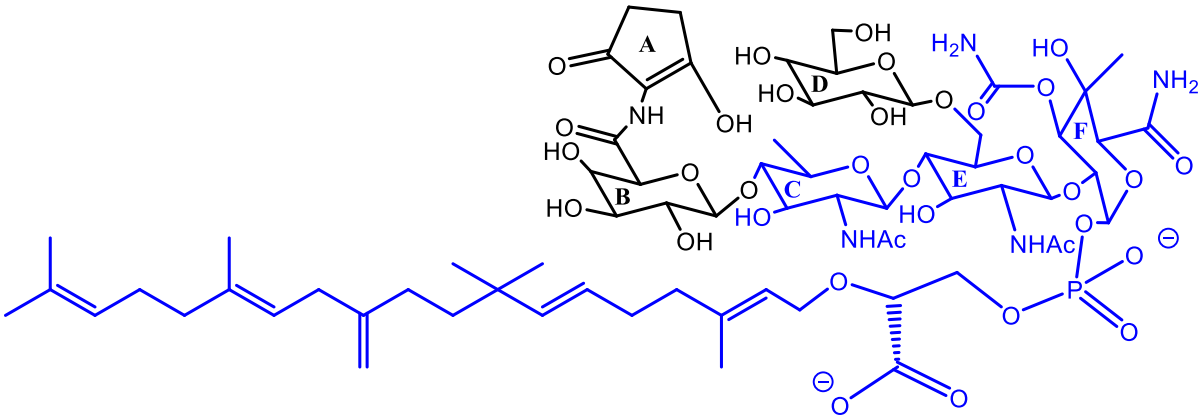


Figure 7.1: Structure of MoeA showing the minimal inhibitory pharmacophore (in blue) which can be used as a scaffold for the design of new potential inhibitors for bacterial transglycosylases. Modified from ref. 8

A break-up of the importance and roles of the individual units of MoeA can be summarized as follows[8]:

1. **A & B units**: These rings are not part of the minimal inhibitory pharmacophore of MoeA and, therefore, are ideal for chemical modification, and may be used as attachment points for other molecules.
2. **Sugar Units C, E & F**: These are the minimum units of the pentasaccharide required for antibacterial activity.
3. **Rings C and E**: These rings bind in the same way as a N-Acetylglucosamine-N-Acetylmuramic acid sugars (NAM-NAG disaccharide) of the lipid II substrate.
4. **Rings E and F:** The structure of ring E is similar to that of NAG whereas that of F is similar to NAM in Lipid II.
5. **D Ring:** Losing the D ring does not affect antibacterial activity.
6. **Phosphoric Acid diester group**: This group is similar in structure to the pyrophosphate in Lipid II and probably mimics it. The function of this group could be to direct the C25 lipid chain towards the membrane.
7. **C25 lipid chain**: The lipid chain part of MoeA is similar to the C35 chain in Lipid II and enables anchoring of MoeA.

8. Both the C25 chain and the phosphoglycerate are essential for the binding of MoeA to the transglycosylase. The phosphoryl group forms electrostatic interactions with conserved residues in the TG active site.

7.3 Important analogues of MoeA described in literature

The revival of MoeA ensured in-depth research into the molecule. From a synthetic point of view, the total synthesis of MoeA[9], synthesis of its analogues as well as degradation products[8] were published. Most of the analogues of MoeA described here are less potent than the parent molecule. We will not discuss those molecules which show little or no antimicrobial activity[10, 11, 12]. Some of the analogues possess antibacterial activity, particularly against a wide variety of Gram positive bacteria, and are also important from a synthetic point of view, and are therefore worth mentioning. One or two, however, are more potent than the parent molecule against certain species, particularly those with lipid tail modifications are of particular interest[7]. Depending on the type of modification, two types of analogues of MoeA are described herein: 1. degradation products of MoeA and 2. lipid tail modifications of MoeA. Some of the important studies, with brief descriptions of the analogues, are summarized herewith:

7.3.1 Type I: Degradation products of MoeA

1. The fact that the degradation products of MoeA exhibit inhibitory activity against transglycosylases led to a study in 1999 by Sofia et al. As part of this study, a combinatorial library consisting of 1300 analogues of the MoeA disaccharide core was synthesized.[13] These analogues were focussed on modifications at C2 of the E ring and C3 of the F ring of MoeA, which form the critical interaction points of MoeA with the transglycosylase domain. Modifications included aromatic groups attached to the E and F rings and a lipid tail of 12 rather than 25 carbons. Three compounds showed particular promise[14] (Figure 7.2, I-III. Original names: TS30153, TS30663 and TS30888 respectively). These compounds were significantly less potent than MoeA (<100 fold), but had IC50 values in the range of 10–15 μM.[15] Most noteworthy is that these analogues exhibited activity against Gram-positive strains, particularly *Enterococcus faecium*, which has tolerance to MoeA. This indicates simple degradation compounds of MoeA are capable of showing potency against clinically relevant pathogens.

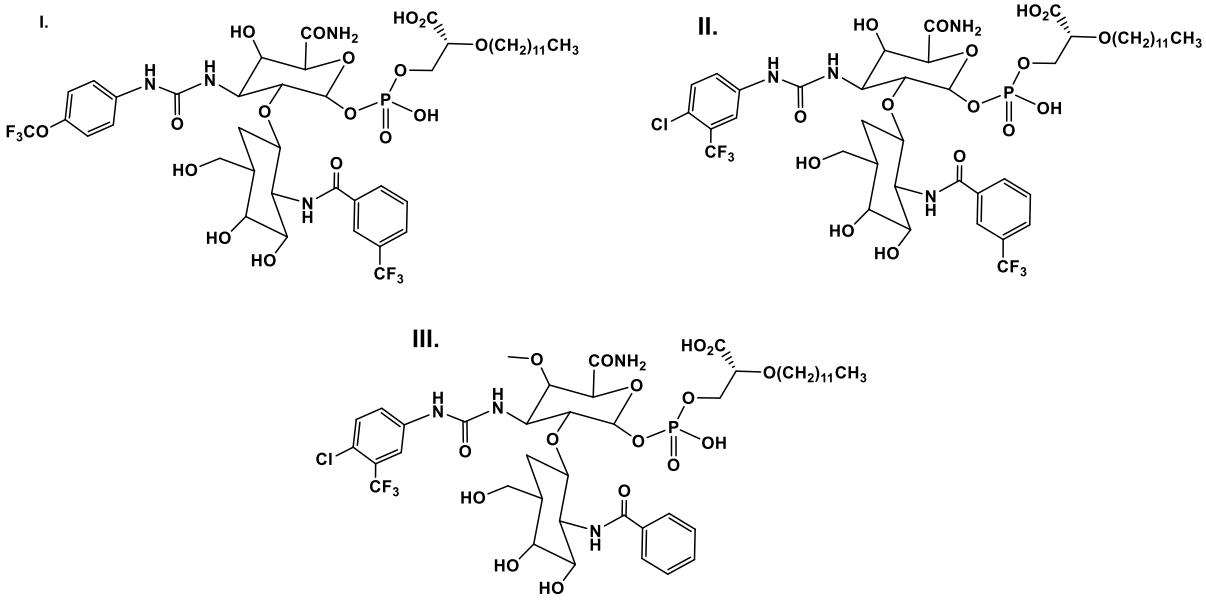


Figure 7.2: Analogues of MoeA as described in ref. 14.

2. In 2006, Halliday et al.[16] revisited the disaccharide scaffold described by Sofia et al. in 1999. His class of compounds also involved maintaining the important transglycosylase binding regions. MIC values of 1-4 µg/mL were obtained for the lead compounds against a broad range of Gram-positive organisms.[16] [13] One example compound from this class is ACL 19273 (Figure 7.3). This molecule showed direct binding and inhibition of the transglycosylase domain. The binding site, however, is unknown and could be either the acceptor or donor site of the enzyme. Inhibitors that bind to the transglycosylase acceptor site are of particular interest, as it is in complete contrast to where MoeA binds. This makes the study of these small disaccharides of particular interest, especially if they bind to the acceptor site. Determining the binding site of the inhibitors of transglycosylase is crucial in elucidating their mode of action.

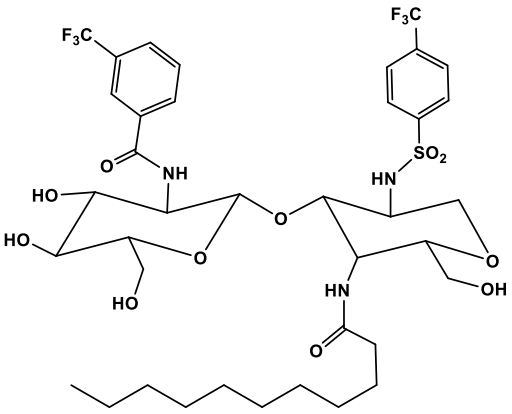


Figure 7.3: Structure of MoeA analogue ACL19273

7.3.2 Type II: Modification of the lipid tail of MoeA

1. He et al. in (2000) isolated a MoeA analogue, AC326-α (Figure 7.4)[17] which has a cyclic moenocinol chain, i.e. a diumycinol chain. The analogue has properties very similar to MoeA, except in case of certain Gram positive bacteria such as *Enterococcus faecium* where it is active (MIC 0.06-0.25 µg/mL and MoeA is not active).

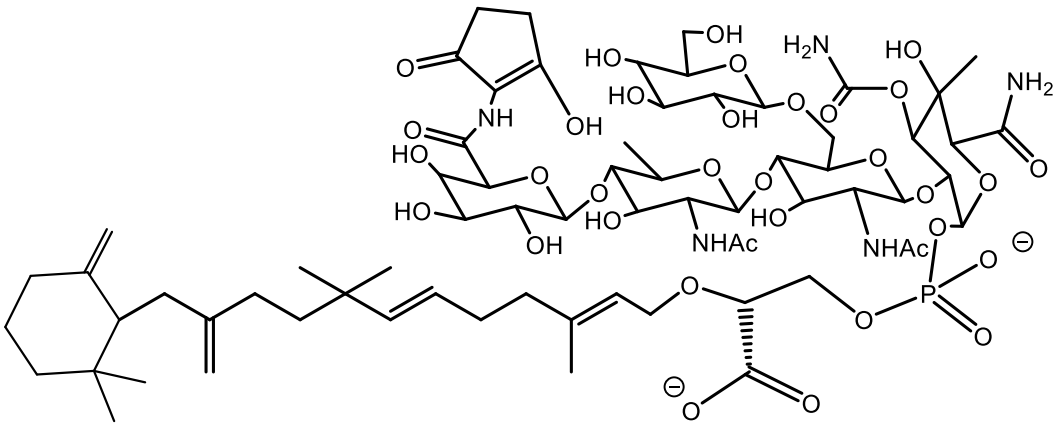


Figure 7.4: Structure of MoeA analogue AC326-α as described in ref. 17.

2. Several lipid modifications of MoeA were performed by Kahn, et. al. and are described in literature.[7] Only one example is shown herein, which is of particular interest. In this case, the C25 moenocinol chain has been replaced by a simple C20 fully saturated chain. The analogue (figure 7.5) shows a 4-fold lower activity than MoeA in *S. auerus* but a 4-fold and 2-fold improvement in activity is observed in case of *E. coli* MG1655 and MR698 respectively.

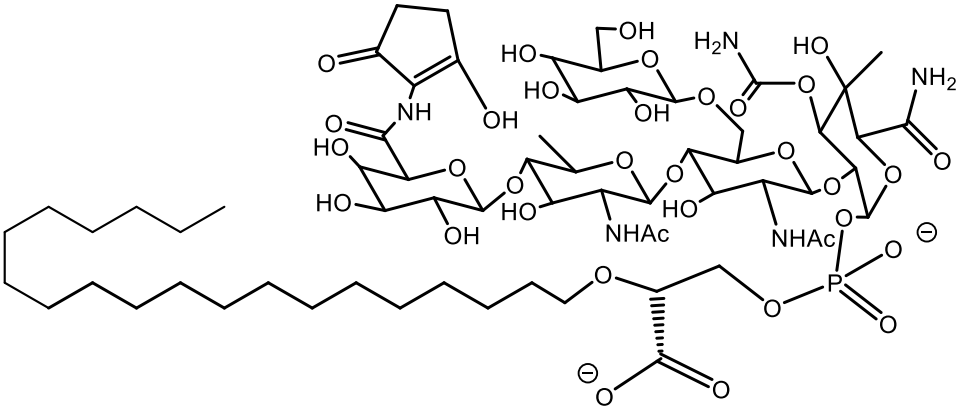


Figure 7.5: Structure of a MoeA analogue as described in ref. 7.

7.4 Design and syntheses of peptide-MoeA covalent conjugates

Our previous work involved the synthesis of MoeA-peptide non-covalent complexes. Although these complexes show up to 30-times improvement in antimicrobial activity, the non-covalent complexes have their limitations. The interactions between the lead peptide and MoeA in a non-covalent complex are strong enough to ensure that the MoeA-peptide complex is carried through the outer membrane in case of bacteria. However, the complex could break down if MoeA delivery is attempted in case of more complex multicellular organisms using the same approach. Therefore, there is a need to covalently link the delivery peptides to MoeA.

Three important parameters are considered in the development of the peptide-MoeA conjugates: 1. The synthesis strategy. 2. The choice of peptide. 3. The choice of linker linking the cationic peptide and negatively charged MoeA. To ensure that MoeA is carried through the outer bacterial membrane, cationic peptides were chosen for conjugation to MoeA. New delivery peptides as well as modified existing delivery peptides for MoeA delivery have also been reported through this work. The key residues for the delivery of MoeA have been identified as lysine and arginine.

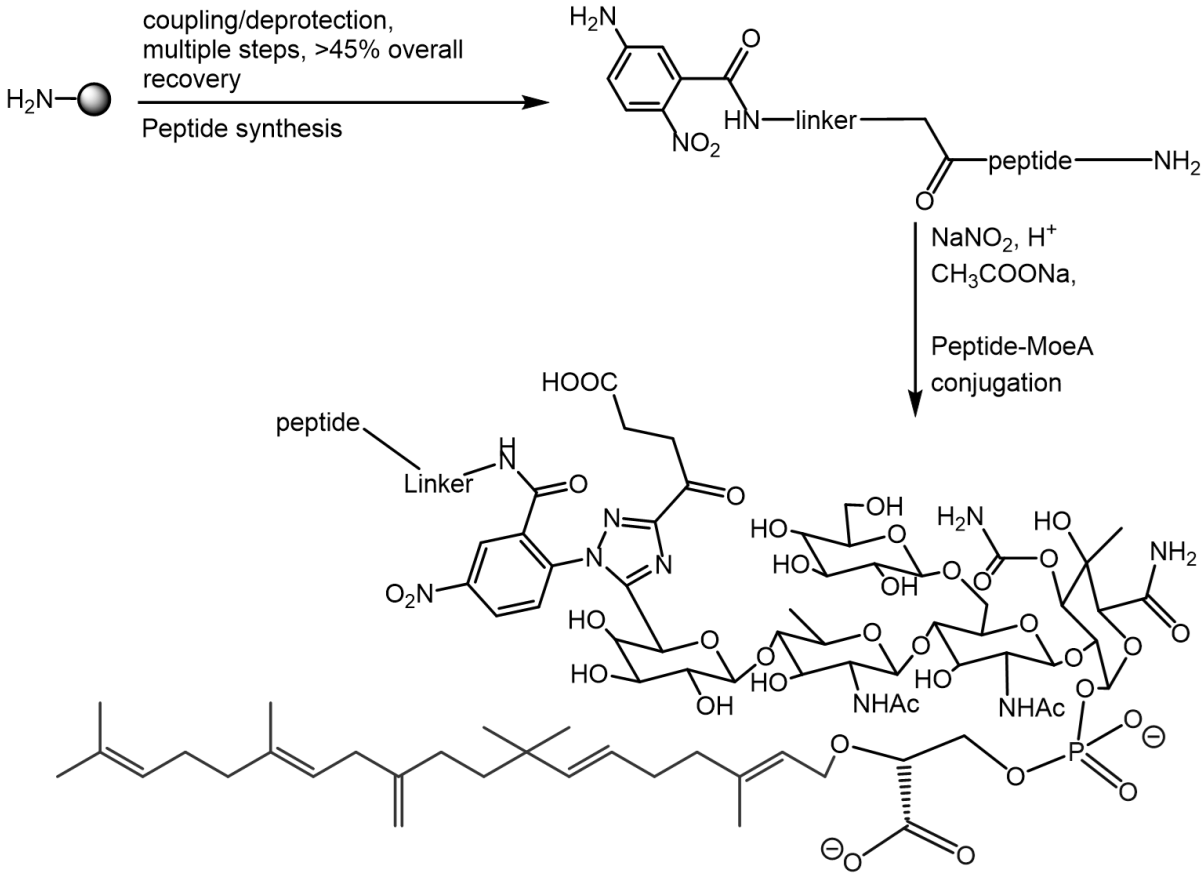


Figure 7.6: General strategy for the synthesis of a peptide-Moenomycin A conjugate via A ring conjugation

A general conjugation methodology (Figure 7.6) has been designed and optimized specially for peptide conjugation to MoeA based on the use of a commercially available 5-amino-2-nitrobenzoic acid (Figure 7.8, denoted by **X**) linker[12]. This linker can be directly coupled to the peptide on solid phase through an amide bond linkage using standard coupling condtions without the need for any protecting group due to the differences in pKa of the aromatic and aliphatic amines. MoeA has been known to react selectively with diazonium salts at the enolized β-diketone unit[18] (Figure 7.7 step 2A). The initial coupling product undergoes a Japp-Klingemann cleavage[19] forming an amidrazone (Figure 7.7 step 2B), which then cyclizes (Figure 7.7 step 2C) to form the final product which is a triazole. The initial diazonium salt formation (Figure 7.7 step 1) is very selective towards the aromatic amine, once again due to the differences in pKa between the amines present in the lysine and arginine side chains of the peptide and the amine present in the 5-amino-2-nitrobenzoic acid linker. Moreover, the reaction occurs at the A ring of MoeA, which has a minimal effect on the antimicrobial activity[5, 8], making this strategy perfect for peptide conjugation to MoeA.

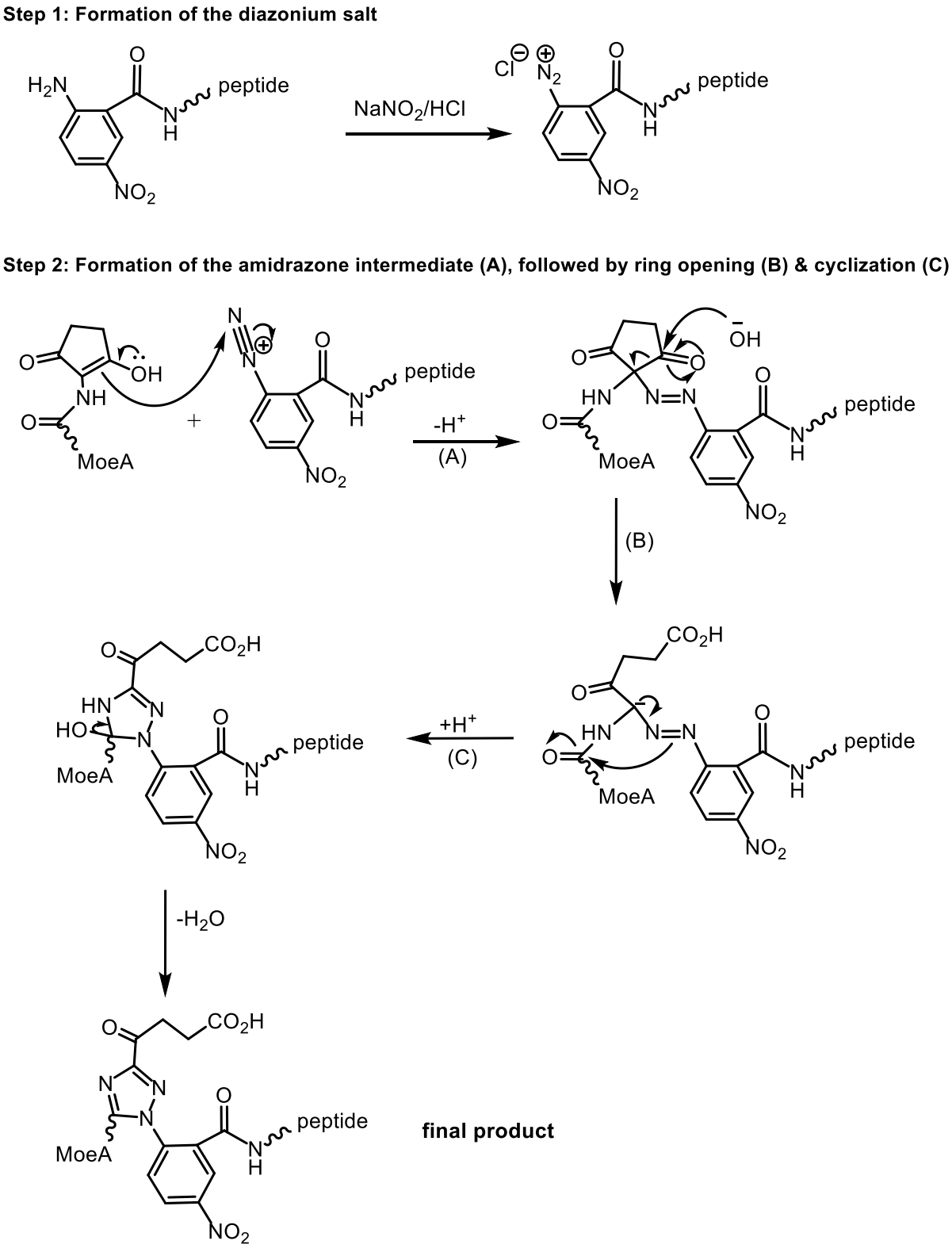


Figure 7.7: Proposed two-step mechanism of the formation of a peptide-MoeA conjugate.

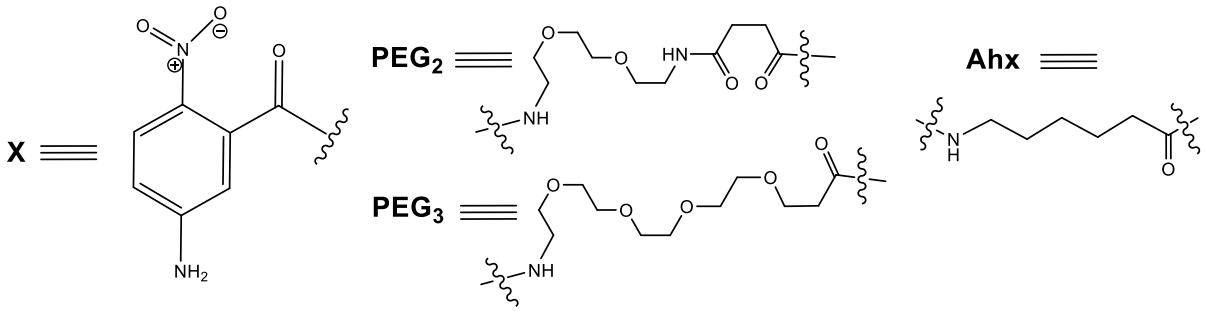


Figure 7.8: Structures of 5-amino-2-nitrobenzoic acid (**X**), PEG_2 linker (**PEG_2**), PEG_3 linker (**PEG_3**) and 6-Aminohexanoic acid linker (**Ahx**) used for the syntheses of peptides 1-22 (Table 7.1).

Another very important aspect in the synthesis and development of the peptide-MoeA conjugates was the choice of the linker. MoeA is a doubly negatively charged molecule in water, due to the presence of a phosphate and carboxylate group. On the other hand, the peptides are cationic due to the presence of multiple lysine and/or arginine groups. A linker is, therefore, essential; first to have flexibility for the overall peptide-MoeA conjugate and, more importantly, to prevent folding of the molecule onto itself due to intramolecular electrostatic interactions. Three linkers were chosen for this purpose: The commercially available 6-aminohexanoic acid (Ahx, 7 atoms) hydrophobic linker, which was coupled once or twice if a longer linker was needed, and a longer PEG_3 (16 atoms) hydrophilic linker. A PEG_2 (14 atoms) linker was chosen as the third linker and as a substitute for the PEG_3 linker in order to have the exact same length as two Ahx linkers. In order to have a peptide-MoeA conjugate which acted as a control, peptides without linkers were also synthesized. In these cases, the 5-amino-2-nitrobenzoic acid (X) linker was directly coupled to the peptide.

Based on our previous discussion and experience in the delivery of MoeA particularly against Gram Negative Bacteria, a select list of 22 peptides was chosen as most suitable for conjugation to MoeA (Table 7.1). These peptides are primarily polylysine and polyarginine sequences including combinations of Lys and Arg ranging from 4 to 10 amino acids. With the exceptions of peptides **2**, **8**, **11** and **15**, which were assigned as control peptides, all the peptides contain at least one of the three flexible linkers shown in Figure 7.5. The following list of peptides were synthesized for conjugation to MoeA:

Sr. No.	Code / Code for conjugate	Peptide sequence	Max. Theoretical Charge
1.	1 / 23	X-Ahx-RRRR	4+
2.	2 / 24	X-RRRRRRRR	8+
3.	3 / 25	X-Ahx-RRRRRRRR	8+
4.	4 / 26	X-$(Ahx)_2$-RRRRRRRR	8+
5.	5 / 27	X-Ahx-RRRRRRRRR	9+
6.	6 / 28	X-Ahx-RRRRRRRRRR	10+
7.	7 / 29	X-Ahx-KKRRRRRRRR	10+
8.	8 / 30	X-KRRKRRKRR	9+
9.	9 / 31	X-Ahx-KRRKRRKRR	9+
10.	10 / 32	X-$(Ahx)_2$-KRRKRRKRR	9+
11.	11 / 33	X-KKKKKR	6+
12.	12 / 34	X-Ahx-KKKKKR	6+
13.	13 / 35	X-$(Ahx)_2$-KKKKKR	6+
14.	14 / 36	X-Ahx-RRWWRRWRR	6+
15.	15 / 37	X-R	1+
16.	16 / 38	X-$(PEG)_3$-KKKKKR	6+
17.	17 / 39	X-$(PEG)_2$-KKKKKR	6+
18.	18 / 40	X-$(PEG)_3$-RRRRRRRR	8+
19.	19 / 41	X-$(PEG)_2$-RRRRRRRR	8+
20.	20 / 42	X-$(PEG)_3$-RRRR	4+
21.	21 / 43	X-$(PEG)_3$-RRRRR	5+
22.	22 / 44	X-$(PEG)_2$-RRRR	4+

Table 7.1: Complete list of peptides synthesized. All peptides have a C-terminal amide. Code for conjugate refers to the covalent conjugate between the corresponding peptide and MoeA.

7.5 Minimum Inhibitory Concentration (MIC) data

Charlotte S. Vincent is thanked for testing all the compounds for providing the MIC data.

Sr. No.	Code	*K. pnumoniae*	*A. baumannii*	*P. aeruginosa*
		MIC	MIC	MIC
1.	23	25.49	12.74	GAW
2.	24	42.35	2.65	42.35
3.	25	20.41	10.21	81.64
4.	26	2.46	19.70	9.85
5.	27	GAW	9.72	77.77
6.	28	GAW	18.56	74.25
7.	29	4.99	2.49	39.90
8.	30	GAW	1.29	41.36
9.	31	GAW	18.87	75.48
10.	32	77.09	4.82	38.55
11.	33	24.90	1.56	24.90
12.	34	23.85	0.37	23.85
13.	35	22.89	0.72	45.77
14.	36	18.93	GAW	18.93
15.	37	33.16	2.07	132.66
16.	38	11.36	45.43	11.36
17.	39	11.43	45.71	5.71
18.	40	78.30	4.89	78.30
19.	41	78.71	19.68	78.71
20.	42	GAW	2.86	GAW
21.	43	GAW	0.38	GAW
22.	44	97.40	3.04	GAW
23.	MoeA	40.46	2.53	161.83
24.	Peptides 1-22	GAW	GAW	GAW

Table 7.2: Complete list of MIC values obtained for the three Gram negative bacteria (in μM). GAW: Growth in all wells even at a concentration of 256 μg/mL. MIC values highlighted in red are of particular interest. Strains used: *K. pneumoniae ATCC 700603, A. baumannii ATCC 19606, P. aeruginosa ATCC 27853.*

The conjugates **23-44** were tested for their antimicrobial properties against three resistant Gram negative bacteria *K. pneumoniae*, *A. baumannii* and *P. aeruginosa*. Given the varied library of compounds synthesized, we obtained an intriguing set of MIC results. One very important aspect was that different peptide sequences showed selectivity against different bacteria and we can thus selectively target a single strain of bacteria by tuning the peptide sequence and the linker. Three lead peptides (**4**, **12** and **17**) and their corresponding conjugates (**26**, **34** and **39**) have been identified which are capable of selectively targeting *K. pneumoniae* (~16 times improvement in MIC), *A. baumannii* (~7 times improvement in MIC) and *P. aeruginosa* (~28 times improvement in MIC) respectively.

Apart from the lead compounds, there were several noteworthy improvements in the activity of the peptide-MoeA conjugates, as compared to MoeA, for all three bacteria. The results can be summarized as follows: Conjugate **43** shows ~7 times improvement in antibacterial activity followed

by **35** which shows ~3 times improvement in MIC against *A. baumannii*. Conjugate **29** shows ~8 times improvement in antibacterial activity against *K. pneumonaie*. Conjugates **26** shows ~16 times improvement against *P. aeruginosa* followed by **16** and **36** which show ~14 and ~8.5 times improvement in MIC against *P. aeruginosa*. A complete list of MIC data is shown in table 7.2.

With the exception of conjugate **43** which is active in *A. baumannii*, it can be concluded that peptides containing at least six positive charges are required for their corresponding conjugates to show noticeable improvement in MIC. This could be attributed to the fact (as in case of the MoeA-peptide non-covalent complexes) that the negative charges of MoeA must first be masked in order for better MoeA delivery. With the notable exception of conjugate **34**, we can also conclude that longer linkers such as $(Ahx)_2$ in case of **26**, $(PEG)_3$ and $(PEG)_2$ in case of **38** and **39** respectively result in better antibacterial activity. The main reason for this could be that a considerable degree of flexibility is required for the overall structure and that longer linkers prevent the folding of the peptide onto MoeA as proposed in section 7.4.

7.6 Synthesis of a Polymyxin-MoeA conjugate and its antimicrobial activity

Based on the discussions of Chapter 6, we know that the polymyxins are among the few molecules active against Gram negative bacteria, particularly against *P. aeruginosa*. Polymyxin B has been known to improve the delivery capability of MoeA against *E. coli*, but studies were limited to only this species of bacteria.[5] We, therefore, decided to pursue this further and hypothesized that a Polymyxin-MoeA conjugate should allow for better delivery of MoeA as compared to the linear peptides discussed in Section 7.3. Analogues of polymyxin have provided encouraging results in the past.[20] Moreover, the total synthesis of polymyxin B1 is known[21], allowing us to gain access to synthetic analogues of the molecule. We, therefore, developed a scheme for the modification of polymyxin B allowing us to covalently link it to MoeA.

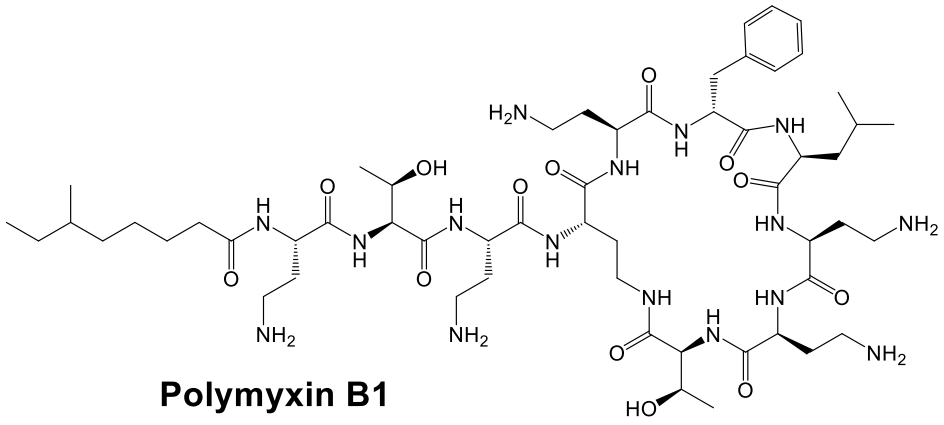


Figure 7.9: Structure of Polymyxin B1

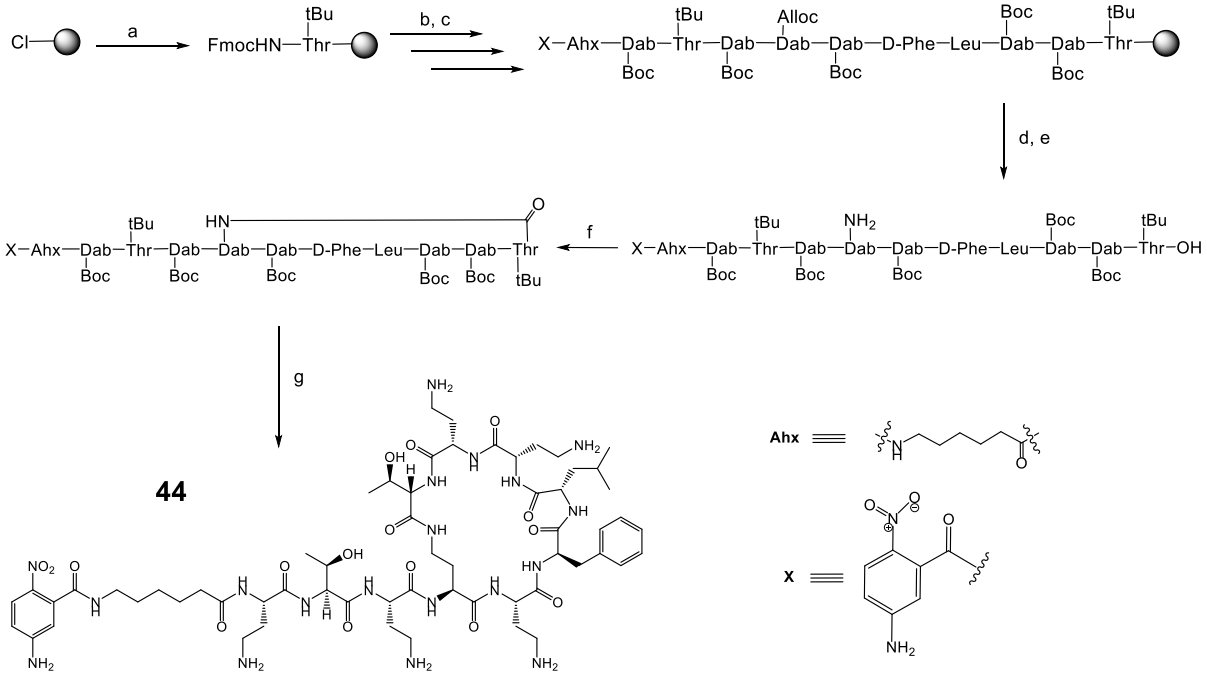


Figure 7.10: Scheme for the synthesis of polymyxin analogue **44** starting from 2-chlorotritylchloride resin, initial loading = 1.2 mmol/g. Reagents used are as follows: (a) 4 eq. Fmoc-Thr(tBu)-OH, 8 eq. DIPEA in DCM, 3 h, followed by capping with MeOH/DIPEA/DCM = 1:1:3, 30 min. loading = 0.68 mmol/g (b) Couplings were performed using 4 eq. Amino acid, 4 eq. DIC/Oxyma in DMF using microwave assisted SPPS. (c) Fmoc deprotection was performed using 20% piperidine in DMF by shaking for 3 min using a µwave, followed by 10 min at r.t. (d) Alloc removal was performed using 0.2 eq. $Pd(PPh_3)_4$ and 24 eq. $PhSiH_3$ in DCM (2 x 1h). (e) Partial cleavage was performed using 2% TFA/5% TIS in DCM, 2h. (f) Cyclization was performed using 1 eq. HATU/10 eq. DIPEA in DMF (peptide concentration: 15 mg/mL). (g) Full cleave was performed using 95% TFA/ 2.5% TIS/2.5% H_2O, 2h. (Refer to the experimental section for details.)

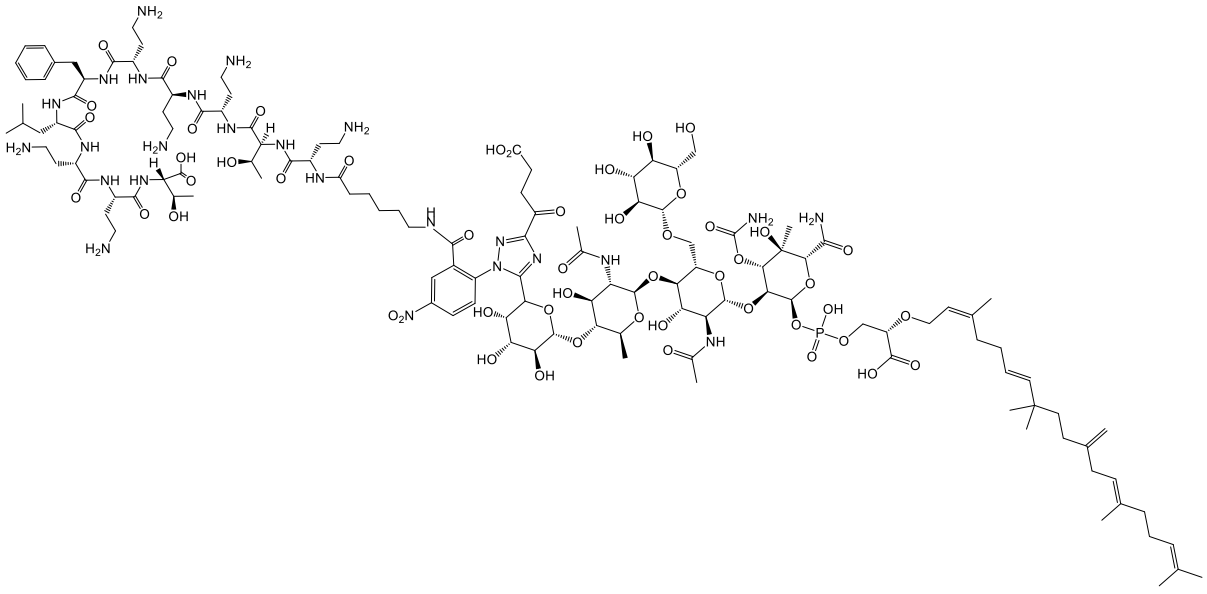


Figure 7.11: Structure of compound **45**: the covalent conjugate of **44** with MoeA.

44 was conjugated to MoeA and the conjugate was purified using the protocols described in the experimental section to obtain the corresponding conjugate **45**. **44** was tested against *P. aeruginosa*

with and without MoeA. The conjugate **45** was also tested against *P. aeruginosa*. These results were compared with those for commercially available polymyxin B.

Sr. No.	Code	*P. aeruginosa*
		MIC (μM)
1.	44	11.43
2.	44 + MoeA	1.37
3.	45	5.57
4.	Polymyxin B (commercial)	0.77

Table 7.3: MIC values obtained for the synthesized polymyxin analogue **44**, **44** + MoeA, **45** and commercially available Polymyxin B (in μM) when tested against *P. aeruginosa ATCC27853*.

The polymyxin analogue was found to be 16 times less effective, as compared to the commercially available polymyxin B. We reasoned that this was due to the attachment of the 5-amino-2-nitrobenzoic acid to the N-terminus of the lipid, resulting in a considerably more polar N-terminus as compared to that of polymyxin B. There was 9 times improvement in antibacterial activity when MoeA was mixed with **44,** suggesting that the polymyxin analogue allows for delivery of MoeA against Gram negative bacteria. Moreover, there was a 2 times improvement in antibacterial activity for compound **45** (as compared to **44**) suggesting that conjugation of the polymyxin analogue to MoeA improves the antibacterial activity of MoeA. Commercially available polymyxin has 5 positive charges, whereas compound **45** has only 2 net positive charges as 3 positive charges are neutralised by the phosphate and carboxylate of MoeA and the additional carboxylate generated due to the degradation of the A ring as a result of peptide conjugation to MoeA. The decrease in number cationic charges from 5+ to 2+ could be resulting in poor membrane permeability of the conjugate. We believe, synthesis of a polymyxin-MoeA conjugate containing additional cationic charges should help improve the permeability and MIC value.

7.7 Resistance to MoeA

The previous and current chapters of this thesis have established that there is great, unused potential for the development of MoeA as an antibiotic, particularly against the resistant Gram negative bacteria. However, development of an antibiotic, particularly against Gram negatives, is very challenging and time-consuming. Therefore, it is imperative to consider all aspects before venturing into this daunting task. One of the key aspects is development of resistance. We have established from the discussions in Chapter 5 that Gram negatives are particularly notorious for developing resistances against antibiotics via multiple mechanisms. Since MoeA has never been used against Gram negatives there are no reports pertaining to resistance regarding the same; however, there is an excellent review by Walker et. al. regarding resistance to MoeA[1] which we will discuss here.

Given that MoeA was first discovered in 1965, it is actually quite surprising that there are no reports which suggest significant animal[22] or human[23] resistance to MoeA. Several *Enterococcus faecium* strains are reported to be naturally resistant to MoeA. However, given varying experimental conditions and the fact that the very definition “resistant” has become a bit ambiguous makes the drawing of such conclusions unreliable. An important fact is that no significant natural cross-

resistance has been revealed between MoeA and other clinically useful classes of antibiotics. Plasmid-borne MoeA resistance has also not been reported.[22]

Resistances to cell-wall-active antibiotics, particularly to Vancomycin, are known. [24] [25] Since MoeA is one of them, it is but natural that the bacteria will be resistant to MoeA as well. [17] Other, rarer, obscure examples of resistance to MoeA can be found. For example[1], null mutations in bamB (formerly yfgL) suppress outer membrane defects in *E. coli* due to mutations in lptD (formerly imp). This suppression leads to increased resistance to vancomycin as well as MoeA.[26] Studies on Gram positive bacteria suggest that changes in membrane makeup or charge could result in resistance to MoeA. In *S. aureus*, reduced content of lysyl-phosphatidylglycerol has led to resistance to MoeA and Vancomycin.[27] The ruminal microbe Prevotella bryantii also suggests that increased availability of undecaprenylpyrophosphate may account for resistance to MoeA, as well as to vancomycin and bacitracin.[28]

The examples mentioned above relate to non-specific mechanisms of defence against MoeA. Currently, no dedicated MoeA resistance mechanisms have been described, although bacteria resistant to MoeA have been reported a long time ago.[29] Mutations in PGTs that would nullify its binding of MoeA have not yet been identified despite repeated efforts. [24] [25] [26] [27] Since quite a few amino acids involved in MoeA binding are also essential for catalysis, mutations that would decrease/abolish MoeA binding may also undermine PGT activity. If this is indeed true, such mutations would be unfavourable for the cell, thus slowing down the resistance development.[28] There is a recent report from Walker et. al. which proves MoeA resistance occurring in *S. Aureus* via a single point mutation in PGT (GT51) resulted in major cell defects and was lethal for the species.[30] Cleavage or modification of MoeA is, however, a viable resistance mechanism. Indeed, a *Bacillus* strain DSM4675 (isolated from contaminated moenomycin fermentation broth) was shown to cleave the phosphoglycosidic bond of MoeA.[31] No other moenomycin-modifying bacteria have been reported in literature thus far. Given that MoeA is excreted intact from animal body, degradation of MoeA does not occur rapidly, at least in the mammalian and avian microflora[1].

S. ghanaensis and *S. clavuligerus* gene clusters producing moenomycin do not carry resistance genes.[1] This raises the question: How do producers of MoeA avoid suicide? Analysis of a big collection of various moenomycin non-producers (such as *S. coelicolor*[32]), revealed their uniformly high resistance to MoeA[33]. *S. albus* J1074 is the only exception, but moenomycin-resistant clones of this strain occur frequently.[34] Resistance to MoeA appears to be a general trait of actinobacteria (Gram-positive bacteria with high guanine and cytosine content in their DNA), which may be a result of their PGT having low affinity to MoeA, or subtle differences in their cell wall organization compared to other bacteria.[1]

The fact that resistance to MoeA is quite uncommon is encouraging; however, little is known regarding the mechanisms of resistance to MoeA. Furthermore, there is no data regarding resistance of MoeA in case of Gram negatives. Therefore, we are currently conducting our own research into the development of resistance against MoeA, MoeA-peptide non-covalent complexes and MoeA-peptide covalent conjugates through a series of tests. At the time of writing this thesis, however, this data was not available.

7.8 Conclusion

The results of this chapter definitely prove that MoeA can be used to target resistant Gram negative bacteria. We have optimized the protocol for unprecedented MoeA-peptide conjugates. Conjugation of peptides has not only shown considerable improvement in antibacterial activity, but also allowed for specific targeting of individual bacteria. By varying the peptide sequences and linkers, the methodology can be developed into an effective screening technique to test the capability of peptides for delivery of MoeA. Moreover, the MoeA-peptide conjugates make use of cheap, commercially available and non-toxic starting materials, allowing for easy scaling up. Easy diversification allows us to synthesize multifunctional antibiotics synthesized using this approach. In a time where new antibiotics are the need of the hour, MoeA-peptide conjugates could prove to be a very useful class of antibiotics, particularly against resistant Gram negative bacteria and this work lays the foundation for the development for the same.

7.9 References:

1. Ostash, B. & Walker, S. Moenomycin family antibiotics: chemical synthesis, biosynthesis, and biological activity. *Nat. Prod. Rep.* **27,** 1594–1617 (2010).
2. Welzel, P. *et al.* Moenomycin a: Minimum structural requirements for biological activity. *Tetrahedron* **43,** 585–598 (1987).
3. Yang, G. *et al.* Studies on the synthesis of di- and trisaccharide analogues of moenomycin A. Modifications in unit E and in the lipid part. *Helv. Chim. Acta* **87,** 1807–1824 (2004).
4. Möller, U. *et al.* Moenomycin A - Structure-activity relations synthesis of the D-galacturonamide analogue of the smallest antibiotically active degradation product of moenomycin A. *Tetrahedron* **49,** 1635–1648 (1993).
5. El-Abadla, N. *et al.* Moenomycin A: The role of the methyl group in the moenuronamide unit and a general discussion of structure-activity relationships. *Tetrahedron* **55,** 699–722 (1999).
6. Fuse, S. *et al.* Functional and structural analysis of a key region of the cell wall inhibitor moenomycin. *ACS Chem. Biol.* **5,** 701–11 (2010).
7. Kahne, D. E. & Walker, S. Patent: Moenomycin Analogues, methods of Synthesis and uses thereof. US 2015/0119354 A1. 1–51 (2015).
8. Galley, N. F., O'Reilly, A. M. & Roper, D. I. Prospects for novel inhibitors of peptidoglycan transglycosylases. *Bioorg. Chem.* **55,** 16–26 (2014).
9. Taylor, J. G., Li, X., Oberthür, M., Zhu, W. & Kahne, D. E. The total synthesis of moenomycin A. *J. Am. Chem. Soc.* **128,** 15084–5 (2006).
10. Vogel, S. *et al.* Moenomycin analogues with long-chain amine lipid parts from reductive aminations. *Tetrahedron* **57,** 4147–4160 (2001).
11. Vogel, S., Stembera, K., Hennig, L., Findeisen, M. & Giesa, S. Moenomycin analogues with modified lipid side chains from indium-mediated Barbier-type reactions. *Tetrahedron* **57,** 4139–4146 (2001).
12. Müller, D., Giesa, S., Knoll, H. & Welzel, P. Synthesis of Fluorescent Derivatives of the Antibiotic Moenomycin A. *European J. Org. Chem.* 1149–1162 (2002).
13. Sofia, M. J. *et al.* Discovery of novel disaccharide antibacterial agents using a combinatorial library approach. *J. Med. Chem.* **42,** 3193–3198 (1999).
14. Baizman, E. R. *et al.* Antibacterial activity of synthetic analogues based on the disaccharide structure of moenomycin, an inhibitor of bacterial transglycosylase. *Microbiology* **146,** 3129–3140 (2000).
15. Goldman, R. & Gange, D. Inhibition of Transglycosylation Involved in Bacterial

Peptidoglycan Synthesis. *Curr. Med. Chem.* **7,** 801–820 (2000).
16. Halliday, J., McKeveney, D., Muldoon, C., Rajaratnam, P. & Meutermans, W. Targeting the forgotten transglycosylases. *Biochem. Pharmacol.* **71,** 957–967 (2006).
17. He, H., Shen, B., Korshalla, J., Siegel, M. M. & Carter, G. T. Isolation and Structural Elucidation of Moenomycin Group. **53,** 191–195 (2000).
18. Kempin, U. *et al.* Moenomycin A: New chemistry that allows to attach the antibiotic to reporter groups, solid supports, and proteins. *Tetrahedron* **53,** 17669–17690 (1997).
19. Phillips, R. R. No Title. *Org. React.* **10,** 143–178 (1959).
20. Rabanal, F. *et al.* A bioinspired peptide scaffold with high antibiotic activity and low in vivo toxicity. *Sci. Rep.* **5,** 10558 (2015).
21. Sharma, S. K. *et al.* Solid-phase total synthesis of polymyxin B1. *J. Pept. Res.* **53,** 501–506 (1999).
22. Butaye, P., Devriese, L. a & Haesebrouck, F. Antimicrobial Growth Promoters Used in Animal Feed : Effects of Less Well Known Antibiotics on Gram-Positive Bacteria Antimicrobial Growth Promoters Used in Animal Feed : Effects of Less Well Known Antibiotics on Gram-Positive Bacteria. *Society* **16,** 175–188 (2003).
23. Komatsuzawa, H., Suzuki, J., Sugai, M., Miyake, Y. & Suginaka, H. Effect of combination of oxacillin and non-beta-lactam antibiotics on methicillin-resistant Staphylococcus aureus. *J. Antimicrob. Chemother.* **33,** 1155–63 (1994).
24. Nishi, H. *et al.* Moenomycin-resistance is associated with vancomycin-intermediate susceptibility in Staphylococcus aureus. *Microbiol Immunol* **47,** 927–935 (2003).
25. Eggert, U. S. Genetic Basis for Activity Differences Between Vancomycin and Glycolipid Derivatives of Vancomycin. *Science (80-.).* **294,** 361–364 (2001).
26. Ruiz, N., Falcone, B., Kahne, D. & Silhavy, T. J. Chemical conditionality: A genetic strategy to probe organelle assembly. *Cell* **121,** 307–317 (2005).
27. Nishi, H., Komatsuzawa, H., Fujiwara, T., McCallum, N. & Sugai, M. Reduced content of lysyl-phosphatidylglycerol in the cytoplasmic membrane affects susceptibility to moenomycin, as well as vancomycin, gentamicin, and antimicrobial peptides, in Staphylococcus aureus. *Antimicrob. Agents Chemother.* **48,** 4800–4807 (2004).
28. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8,** 260–71 (2010).
29. Meyers, E. *et al.* Biological Characterization of Prasinomycin, a Phosphorus-containing Antibiotic. *Appl. Microbiol.* **16,** 603–608 (1968).
30. Rebets, Y. *et al.* Moenomycin resistance mutations in Staphylococcus aureus reduce peptidoglycan chain length and cause aberrant cell division. *ACS Chem. Biol.* **9,** 459–467 (2014).
31. Metten, K.-H. *et al.* The first enzymatic degradation products of the antibiotic moenomycin A. *Tetrahedron* **48,** 8401–8418 (1992).
32. Bentley, S. *et al.* Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417,** 141–147 (2002).
33. Tsypik, O. *et al.* Transcriptional regulators of GntR family in Streptomyces coelicolor A3(2): analysis in silico and in vivo of YtrA subfamily. *Folia Microbiol. (Praha).* **61,** 209–220 (2016).
34. Makitrynskyy, R. *et al.* Genetic factors that influence moenomycin production in streptomycetes. *J. Ind. Microbiol. Biotechnol.* **37,** 559–66 (2010).

CHAPTER 8

SUMMARY, CONCLUSION & PERSPECTIVES

8.1 Summary of work

This thesis contains the work done by me as part of a joint PhD between Ghent University, Belgium and University of Lincoln in the UK. The duration of this PhD was 4 years. The first 2 years of the PhD (Sep. 2012-Aug. 2014) were carried out at Ghent University in Belgium. Subsequently, I moved to the UK where I completed the remaining 2 years of my PhD (Sep. 2014-Aug. 2016) at the University of Lincoln. The joint PhD was designed so that I may use the skills gained in Ghent University to further my research in the University of Lincoln.

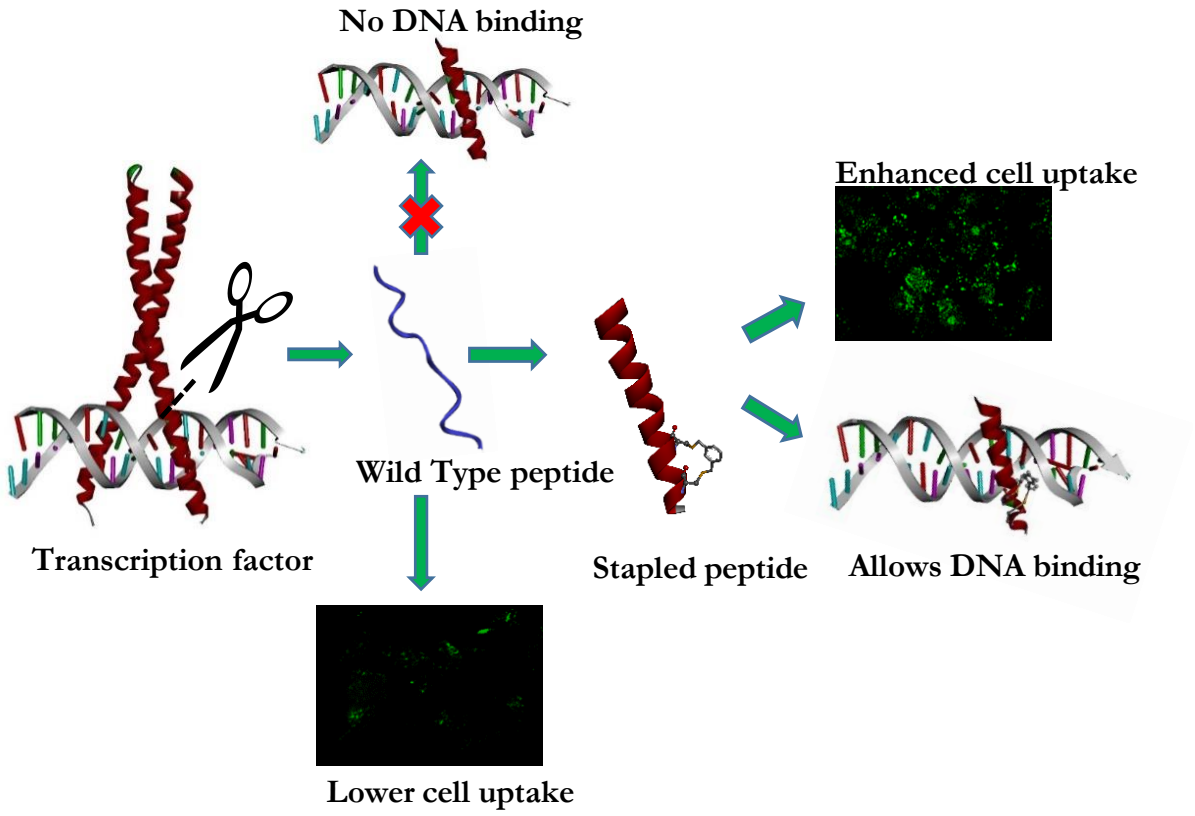


Figure 8.1: An overview of the work done involving stapled peptides for DNA recognition and enhanced cellular uptake at UGent.

My project at UGent involved the sequence specific recognition of dsDNA. An overview of the existing literature in this regard has been described in chapter 1. The aim was to look into and further develop existing stapling methods and subsequently apply them to enhance the DNA binding potential of the basic binding region of the GCN4 protein consisting of the 23 amino acids H_2N-DPAALK**R**A**R**NTEAA**RRS**RA**RK**LQ-OH. The amino acid residues depicted in bold represent the essential residues which interact specifically with DNA[1]. We have replaced and cross-linked two selectively positioned, non-DNA-contacting amino acids on the GCN4 peptide resulting in stapled peptides in an effort to achieve DNA binding and enhance cellular uptake. Classic families of transcription factors like the bZIP, bHLH, homeodomain, HTH, and zinc fingers have already been

studied and models have been developed[2]. However, unlike the existing synthetic bZip peptide models which induce α helicity via a dimeric structure, we here use a single α helix stabilized via peptide stapling. This work was successfully completed in Ghent University, Belgium and is described in chapters 3 & 4. Alternatively, the work has also been published through the following articles (see Annex I for full text):

> ***A. Iyer***, D. Van Lysebetten, Y. Ruiz García, B. Louage, B. G. De Geest and A. Madder, *Org. Biomol. Chem.*, 2015, **13**, 3856–3862.
>
> Y. Ruiz García, ***A. Iyer***, D. Van Lysebetten, Y. Vladimir Pabon, B. Louage, M. Honcharenko, B. G. De Geest, C. I. Edvard Smith, R. Strömberg and A. Madder, *Chem. Commun.*, 2015, **51**, 17552–17555
>
> Y. Ruiz García, J. Zelenka, Y. V. Pabon, ***A. Iyer***, M. Buděšínský, T. Kraus, C. I. Edvard Smith and A. Madder, *Org. Biomol. Chem.*, 2015, **13**, 5273–5278

During my time in Ghent I also successfully synthesized a cyclic-RGD peptide as part of an additional project. The peptide was conjugated to deoxynivalenol in order to develop new monoclonal antibodies. This work was subsequently published as (see Annex I for full text):

> M. Sanders, Y. Guo, ***A. Iyer***, Y. Ruiz García, A. Galvita, A. Heyerick, D. Deforce, M. D. P. Risseeuw, S. Van Calenbergh, M. Bracke, S. Eremin, A. Madder and S. De Saeger, *Food Addit. Contam. Part A. Chem. Anal. Control. Expo. Risk Assess.*, 2014, 1–9.

As I arrived at the University of Lincoln, I was already quite well trained in peptide synthesis, purification and analysis. The joint PhD gave me the opportunity to apply my skills in three different areas of research. The peptide skills – especially the syntheses and purification of arginine rich cationic peptides and use of orthogonal protection strategies for peptide synthesis which I learned at UGent helped me immensely to complete my PhD projects at the University of Lincoln. A brief description of the projects follows:

1. The problem of antibiotic resistance & the development of Moenomycin A as a drug candidate

The problem of antibiotic resistance poses a serious threat to human health[3]. There is a great demand for new antibiotics particularly against the resistant Gram negative bacteria. The various problems arising due to antibiotic resistance, brief description of the drugs available on the market along with the modes of resistance in bacteria to these drugs have been described in chapter 5. Subsequently, efforts have been made as part of this PhD under the guidance of Dr. Ishwar Singh at the University of Lincoln to develop our own set of antibiotics based on the natural product Moenomycin A.

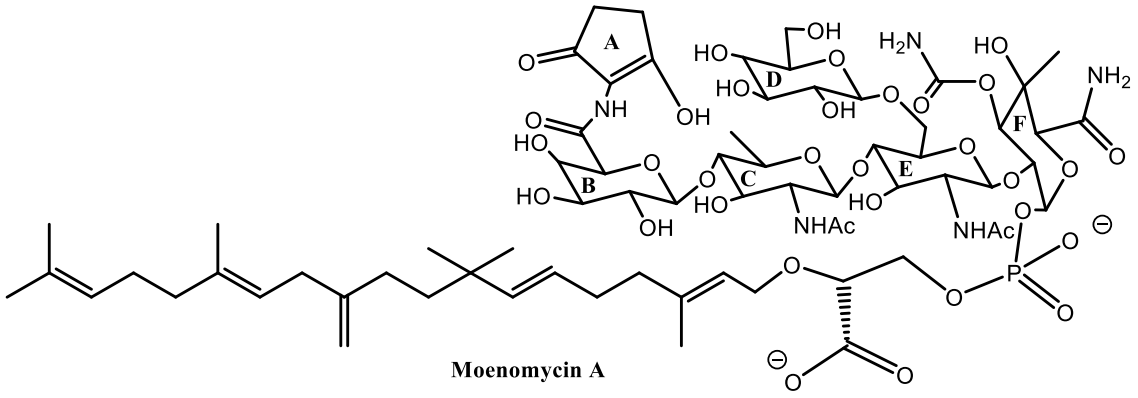


Figure 8.2: Structure of MoeA.

Research into Moenomycin A (MoeA) has not drawn the attention that was expected, although it was discovered more than 50 years ago. However, the recent surge of antimicrobial resistant bacteria has seen the revival of several underused molecules and MoeA is at the forefront. Thus far, MoeA has only been used to target Gram positive bacteria. Herein, we have developed the first unique approach which enables MoeA to target the resistant Gram negative bacteria. Subsequently, we have designed specific "delivery" peptides which bind to MoeA and enable it to cross the negatively charged outer membrane of Gram negative bacteria. Peptides rich in Arg, Lys and/or Trp have been known for their cell permeability. This was shown in chapter 3 where we used stapled DNA binding peptides rich in Lys and Arg to enhance cellular uptake. However, this is the first instance where we have used short, cell permeable peptides to deliver MoeA. This has resulted in a 30-times improvement in antimicrobial activity of MoeA against the highly resistant *Pseudomonas aeruginosa*. Detailed NMR analysis gives valuable insight into the structure of the non-covalent complex between the delivery peptide and MoeA. Furthermore, we have also covalently linked these peptides using suitable linkers to form peptide-MoeA covalent conjugates. The work has been described in detail in chapters 6 & 7.

2. The total syntheses & biological activities of teixobactin and its analogues

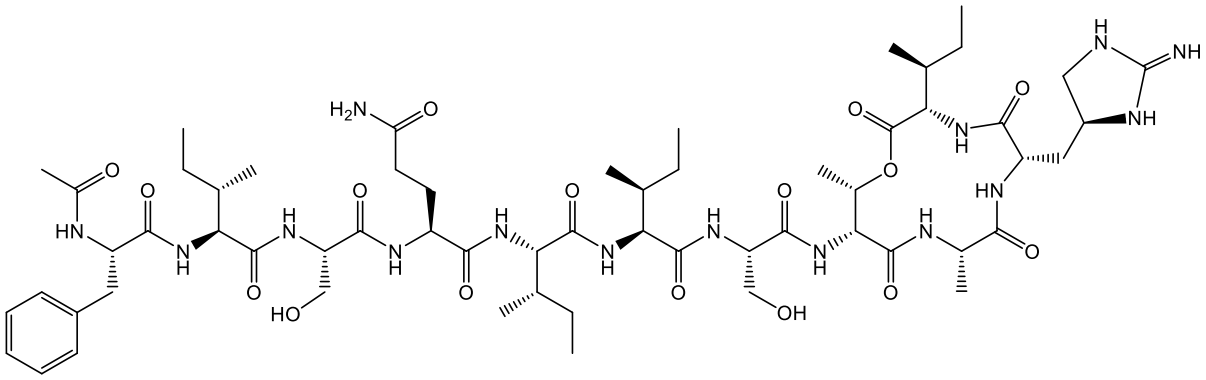


Figure 8.3: Structure of teixobactin.

The majority of work regarding the total syntheses & biological activities of recently discovered teixobactin[4] and its analogues has been done by my friend and colleague, Anish Parmar at the University of Lincoln and therefore has not been included as part of this thesis. Briefly, "*we have performed the total syntheses and biological activities of two teixobactin analogues. The approach is simple, efficient and has several advantages: it uses commercially available building blocks (except AllocHN-D-Thr-OH), has a single purification step and a good recovery (22%). We have also established that the D-amino acids are critical for the antimicrobial activity of these analogues*[5]*."* Along these lines, we have synthesized more than ten analogues of teixobactin (unpublished work). Part of this work has been published as (see Annex I for full text):

> A. Parmar, ***A. Iyer***, et. al. Efficient total syntheses and biological activities of two teixobactin analogues. Chem. Commun. 52, 6060–6063 (2016).

3. The syntheses of furan containing polyamides for dsDNA cross-linking

Complementary to my work on stapled peptides which were based on transcription factors which operate via major grove recognition of dsDNA (chapter 4), we also synthesized furan containing polyamides for dsDNA cross-linking in the minor groove. Although the syntheses of polyamides as well as the furan cross-linking strategy are well established, the following major challenges are to be overcome if cross-linking to dsDNA using this strategy is to be made feasible:

1. Selective oxidation of furan in presence of the Py/Im rings.
2. Local disruption of the DNA duplex in order to achieve cross-linking.

The initial part of this project was to obtain conditions under which furan could be selectively oxidised in the presence of the aromatic Py/Im rings. Through a series of experiments carried out (as described in chapter 2), we have been able to selectively oxidize an N-terminal furan contained in a polyamide, however, further optimization of conditions is required.

8.2 Conclusion

The work done over the past four years by me as part of my PhD thesis mainly revolves around the wide and versatile application of peptides – particularly cationic peptides which allow for DNA recognition and enhanced cell uptake via peptide stapling and the application of cationic peptides for the delivery of Moenomycin A, a compound with known antibacterial properties through bacterial membranes.

The initial part of this PhD involved the understanding of the dsDNA structure, transcription factors which recognize the major groove of DNA and polyamides which recognize the minor grove of DNA. We applied the concepts learned to develop our own set of molecules capable of recognizing dsDNA in the major and minor grooves. Based on the collective knowledge of Prof. Annemieke Madder at UGent and Dr. Ishwar Singh at University of Lincoln we designed and synthesized furan containing polyamides for minor groove cross-linking to dsDNA. Through a series of experiments, we concluded that it is possible to selectively oxidize furan in the presence of Pyrrole/Imidazole containing polyamides using N-Bromosuccinimide.

We also developed synthetic major groove DNA binders capable of mimicking the GCN4 Transcription Factor (TF) through the design and syntheses of five stapled peptides. Through this work we established a general method to allow for DNA binding and enhance cellular uptake of a given DNA binding peptide, while avoiding tedious synthetic routes. Furthermore, we showed that N-terminal helix stabilization is more effective for enhancing DNA binding than stapling in the middle of the sequence. For cell uptake, however, we observed that the more helical peptides are taken up better into cells. It can be concluded that although the two are not mutually exclusive, for future applications such as DNA binding *in cellulo*, a balance will have to be found between a peptide's DNA binding and cell penetration abilities.

As a result of the knowledge gained through our work on TFs involving cell permeable peptides, a new application of these peptides has been developed in the form of a peptide based delivery system for the antimicrobial compound Moenomycin A (MoeA). The system is based on our correct assessment that the delivery of MoeA is the key problem for its lack of activity against resistant Gram negative bacteria. The system is a novel peptide-based, non-covalent delivery system specifically for MoeA, which enables it to target three resistant Gram Negative bacterial strains – *K. pneumoniae, A.*

baumannii and P. aeruginosa. We have hypothesized and correctly proven that the peptides bind to MoeA and form non-covalent complexes and subsequently "carry" them across the outer bacterial membrane present in Gram negative species. Furthermore, we have shown that despite all of them being Gram negative bacteria, they behave quite differently towards the same set of peptide-MoeA non-covalent complexes. We have tuned the peptides in order to obtain up to 30 times improvement in MIC values. Until now, it has been believed that longer, synthetically more challenging polyarginine sequences are required for drug delivery[6]. We have shown, for the first time that MoeA delivery can be achieved as effectively even through shorter sequences such as R4. By designing different delivery peptides (which are combined with MoeA via a simple mixing approach) we have demonstrated the selectivity of our approach to target a specific bacterial strain. Peptides R5**,** R4 & K4 promote the delivery of MoeA best in *A. baumannii*, whereas the HIV-TAT 49-57 peptide RKKRRQRRR is selective for MoeA delivery in *K. pneumoniae*. The sequence KRRKRRKRR on the other hand can perform delivery in all three species of bacteria simultaneously. The Arg rich peptide containing Trp RRWWRRWRR in combination with MoeA shows excellent activity in *P. aeruginosa* and moderate to high activity in *A. baumannii* and *K. pneumoniae*. As toxicity is one of the key factors in case of delivery, we have established that all peptides as well as the MoeA-peptide non-covalent complexes they form are not toxic towards mammalian cell lines up to a tested concentration of 100 μM.

We have performed a detailed NMR analysis followed by extensive structural calculations revealing for the first time the molecular structure of a peptide-MoeA complex at an extremely impressive 1.3 Å resolution. We have proven that the two components interact in a 1:1 ratio further supporting the mechanism that the peptides bind to MoeA and then enable its delivery in Gram negative bacteria. We were thus able to deduce the key interactions between peptide RRWWRRWRR and MoeA. Surprisingly, we observe that the phosphate is not part of the interaction showing an example where although arginines and phosphates are present in the same solution, they need not necessarily interact with one another. Moreover, a simple mixing approach provides a platform to test the delivery capability of different peptides using MoeA without any conjugation. This work is crucial for any further development of MoeA especially in the case of resistant Gram negative bacteria. In a time where new and novel antibiotics particularly against these resistant bacteria are critical, this work lays the foundation for the development of new antibiotics based on MoeA.

Following the excellent results obtained through the formation of peptide-MoeA non-covalent complexes, we developed the protocol for unprecedented MoeA-peptide covalent conjugates. The results obtained through this work definitely prove that MoeA can be used to target resistant Gram negative bacteria. Conjugation of peptides to MoeA has not only shown considerable improvement in antibacterial activity, but also allowed for specific targeting of individual bacteria. By varying the peptide sequences and linkers, the methodology can be developed into an effective screening technique to test the capability of peptides for delivery of MoeA. Moreover, both the MoeA-peptide non-covalent complexes and covalent conjugates make use of cheap, commercially available and non-toxic starting materials, allowing for easy scaling up. Easy diversification allows us to synthesize multifunctional antibiotics synthesized using these approaches.

The work of this thesis has thus provided and expanded upon two major applications of peptides – stapled peptides for phosphate recognition and improvement cell permeability and the delivery of cargo through peptides into bacterial membranes. Although there is still considerable work to be done, I believe, this thesis lays the foundation for the development of new peptide based therapeutics.

8.3 Perspectives

Improvement in terms of binding affinity towards dsDNA in case of the synthetic major groove DNA binding stapled peptides has been proposed through the synthesis of a dimeric stapled peptide. Furthermore, since the GCN4 stapled peptides do not enter the nucleus, the conjugation of a Nuclear Localization Signal (or Sequence, in short NLS) to a GCN4 stapled peptide should help with nuclear uptake. This could allow for DNA binding in *cellulo* which is an important step towards the development of peptide based therapeutics.

In case of the polyamides, we have proposed to carry out the oxidation of the furan containing polyamides in the presence of dsDNA. As the Py/Im rings would be tightly bound to the dsDNA they would be less susceptible to oxidation as compared to the furan which would be unbound. Subsequently, we have also proposed to improve our design polyamides through the incorporation of amino acids capable of local disruption of DNA in addition to furan to enable cross-link formation.

Although we have had considerable success using linear peptides to inhibit the activity of MoeA, work still needs to be done to improve the MIC values if these compounds are to be converted into potential drug candidates. We have already observed improvement in MIC through the covalent conjugation of a modified polymyxin derivative to MoeA. Along those lines, we have proposed the syntheses of branched and cyclic peptides instead of the existing linear ones to be used as delivery peptides in a non-covalent manner. Cyclic peptides work better than linear peptides for cargo delivery for two main reasons:

1. Cyclization increases charge density thereby they are able to penetrate better into cell/bacterial membranes[7].
2. The backbone and the N/C termini of cyclic peptides is protected making them less susceptible to enzymatic degradation

If this is a success, they can then be modified to contain the 5-amino-2-nitrobenzoic acid linker which would allow for subsequent covalent conjugation to MoeA. Cyclic R6, R5, K6 and K5 are among the potential candidates for initial testing.

8.4 References

1. Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted Helices : Crystal Structure of the Protein-DNA Complex. *Cell* **71,** 1223–1237 (1992).
2. Pazos, E., Mosquera, J., Vazquez, M. E. & Mascarenas, J. L. DNA Recognition by Synthetic Constructs. *Chembiochem* **12,** 1958–1973 (2011).
3. Review on Antimicrobial Resistance. *2015* at <http://amr-review.org/>
4. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517,** 455–459 (2015).
5. Parmar, A. *et al.* Efficient total syntheses and biological activities of two teixobactin analogues. *Chem. Commun.* **52,** 6060–6063 (2016).
6. Rothbard, J. B. *et al.* Arginine-rich molecular transporters for drug delivery: Role of backbone spacing in cellular uptake. *J. Med. Chem.* **45,** 3612–3618 (2002).
7. Mika, J. T. *et al.* Structural basis for the enhanced activity of cyclic antimicrobial peptides: The case of BPC194. *Biochim. Biophys. Acta - Biomembr.* **1808,** 2197–2205 (2011).

PART II:

CHAPTER-WISE EXPERIMENTAL SECTION

EXPERIMENTAL SECTION FOR CHAPTER 2

I. Materials

The materials used for polyamide synthesis common to peptide synthesis are mentioned in Chapter 3. HATU, Boc-β-Ala-PAM resin were purchased from Iris Biotech GmbH. Boc-GABA-OH, HOAt solution 0.6M in DMF, 3-(Dimethylamino)-1-propylamine and BTC were purchased from Sigma Aldrich.

II. Manual synthesis protocol

1. General preparation of resin for deprotection and coupling

Note: If Boc-β-Ala-PAM resin of a loading higher than 0.26 mmol/g is purchased, the loading needs to be lowered to that value using HATU/Acetic acid/DIPEA in DMF before the first coupling.

For the first deprotection cycle, Boc-β-Alanine-PAM resin (100 mg; loading = 0.26 mmol/g, 0.026 mmol) was placed in a Teflon lined reaction vessel. The Boc protecting group was removed using three cycles of a TFA:Phenol:Water (TPW) mixture (92.5:5:2.5) (2 x 1 min, and 1 x 5 min) with shaking. The general deprotection procedure using TPW is as follows: 1 x 1 min, then 2 x 3 minute for pyrrole and 1 minute, 3 minutes then 20 minutes for imidazole.

For all couplings in general, the resin was then washed with DMF (3 x 1 min) and then washed with dry solvent (DMF for DCC/HOAt and HATU couplings, THF for BTC couplings) for 1 min before the vessel was drained. The activated monomer was added to the pre-swelled resin and the reaction mixed to form a slurry. The couplings were allowed to proceed for the required amount of time (dependant on sequence and coupling method), before being drained and washed with DMF (4 x 1 min)

2. General cleavage protocol and recovery of polyamides

The polyamides were cleaved from the resin by placing them into a sealed container with the appropriate cleavage amine and heating to 55 °C for 16 hours. The expired resin was removed by filtration through cotton wool and washed with a 1 mL of DCM. The crude material was precipitated by addition of cold diethyl ether and following centrifugation the supernatant was removed. The polyamide was washed twice more by dissolving in DCM (1 ml) and re-precipitating with diethyl ether to yield the crude material.

3. HPLC analysis and purification

Reported analytical HPLC retention times were acquired using 0.1 % TFA in water as buffer A and acetonitrile as buffer B with a flow rate of 1 mL/minute, starting at 5 % B and holding for 3 minutes before increasing to 95 % B over 20 minutes. Unless otherwise stated, polyamides were purified by semi-preparative HPLC with a gradient starting at 10 % B and holding for 3 minutes before increasing to 60 % B over 20 minutes and a constant flow rate of 4.5 mL/minute.

III. General activation protocols for polyamide synthesis

1. Activation with HATU

General procedure for 100 mg of Beta-Ala-Pam resin (loading = 0.26 mmol/g): Monomer (4 equivalents; 0.104 mmol) and HATU (3.6 equivalents; 0.3744 mmol; 36 mg) were dissolved in dry DMF (1 mL). Upon addition of DIEA (150 μL; 0.9 mmol) the solution was left to activate for 2 minutes, before being added to the resin.

2. Activation with DCC/HOAt

Before a DCC/HOAt coupling, the resin must be neutralized by washing it with 1 mL 10% DIPEA/DMF (for 100 mg resin) for 1x1 min.

General procedure for 100 mg of Beta-Ala-Pam resin (loading = 0.26 mmol/g): Monomer (4 equivalents; 0.104 mmol) was added to DCC (19.31 mg; 0.0936 mmol; 3.6 equivalents) and HOAt (13.45 mg; 0.0988 mmol; 3.8 equivalents) in dry DMF (1 mL) were mixed together in a vial and activated for 2 hours with constant shaking. The precipitated DCU is filtered off and the solution added to resin.

3. Activation with BTC

General procedure for 100 mg of Beta-Ala-Pam resin (loading = 0.26 mmol/g): To a solution of monomer (4 equivalents; 0.104 mmol) and BTC (9.6 mg; 0.0324 mmol; 1.2 equivalents) in dry THF (1 mL) was added 2,3,5-collidine (50 μL; 0.38 mmol) forming a white emulsion which was allowed to mix for 2 minutes. This was then added to resin with DIEA (150 μL; 0.9mmol).[1]

IV. Synthesis of polyamides

The Boc-Py-OH, and Boc-Im-OH monomers were synthesized using the protocols as described by Dervan et. al.[2]

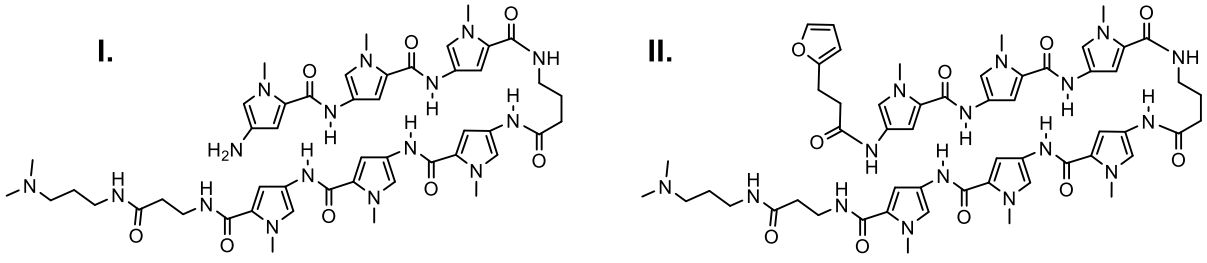


Figure 1: Polyamides synthesized for optimizing the conditions which would allow for selective oxidation of furan. I. is the control to check for degradation.

[1] W. Su, S. J. Gray, R. Dondi, and G. A. Burley, Organic letters, 2009, 11, 3910–3

[2] Baird, E. E. & Dervan, P. B. Solid Phase Synthesis of Polyamides Containing Imidazole and Pyrrole Amino Acids. *J. Am. Chem. Soc.* **118,** 6141–6146 (1996).

For both polyamides the coupling conditions were as follows: Couplings 1-5 (including turn) HATU/DIPEA. Coupling 6-7: BTC. In case of II. Coupling 8: HATU/DIPEA, 3 eq., 2h.

V. HPLC/MALDI-TOF Data

For MALDI-TOF, spectra were recorded up to 4 decimal places. However, information only up to 2 decimal places was considered accurate as a result of the calibration of the MALDI-TOF.

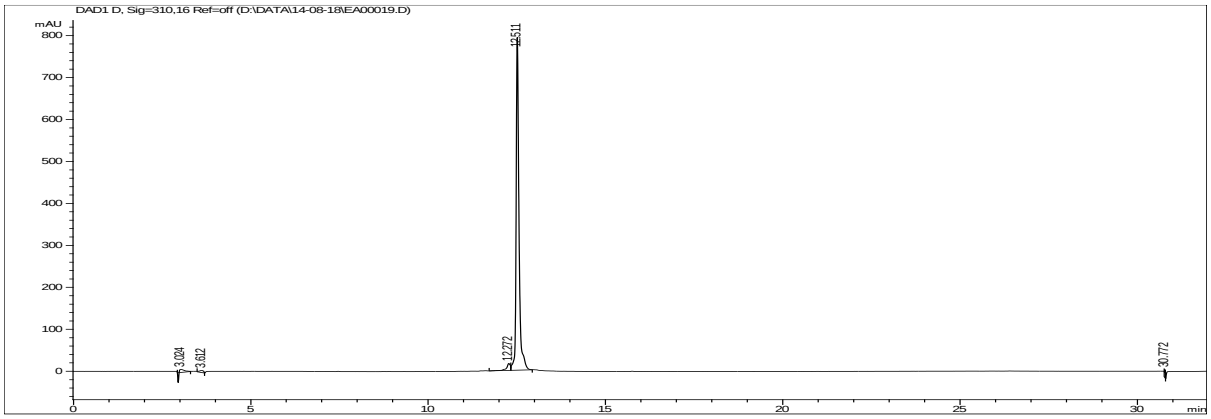


Figure 2: HPLC trace of HPLC purified polyamide **I**. r.t. = 12.511 min

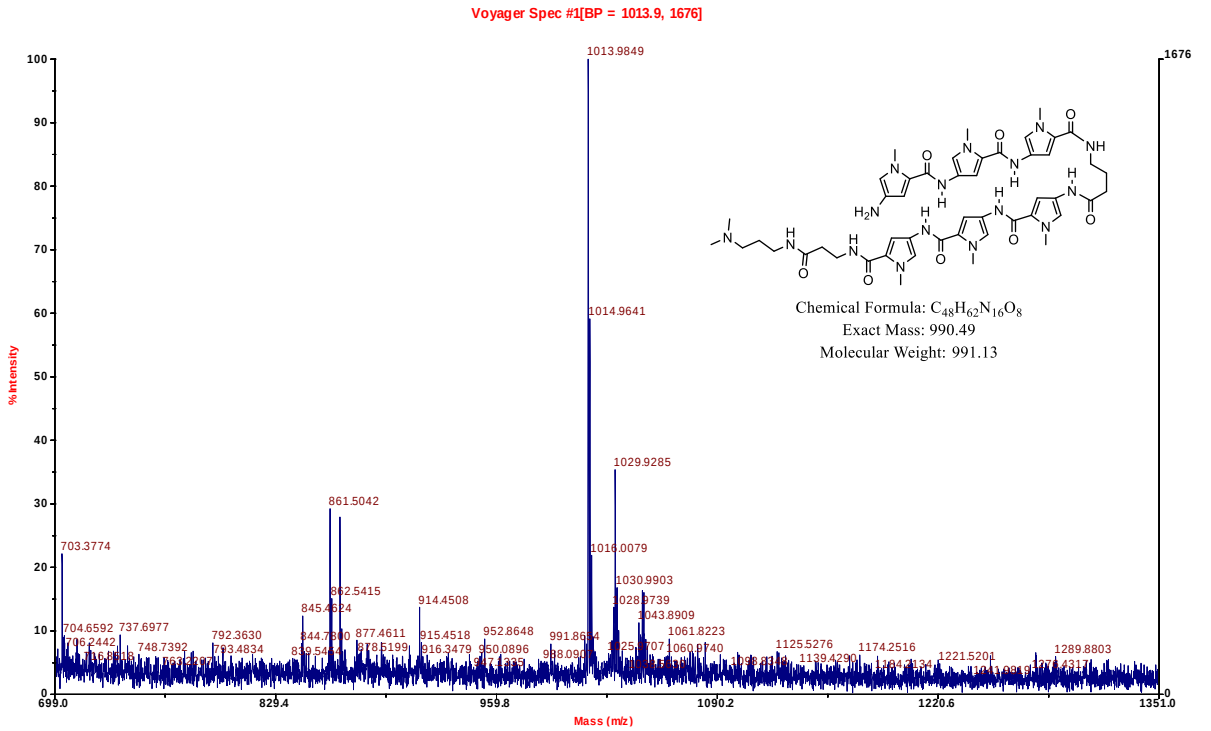


Figure 3: MALDI-TOF spectra for polyamide **I**. Exact mass calcd. 990.49, found M+ Na^+ = 1013.98, $M+K^+$ = 1029.92

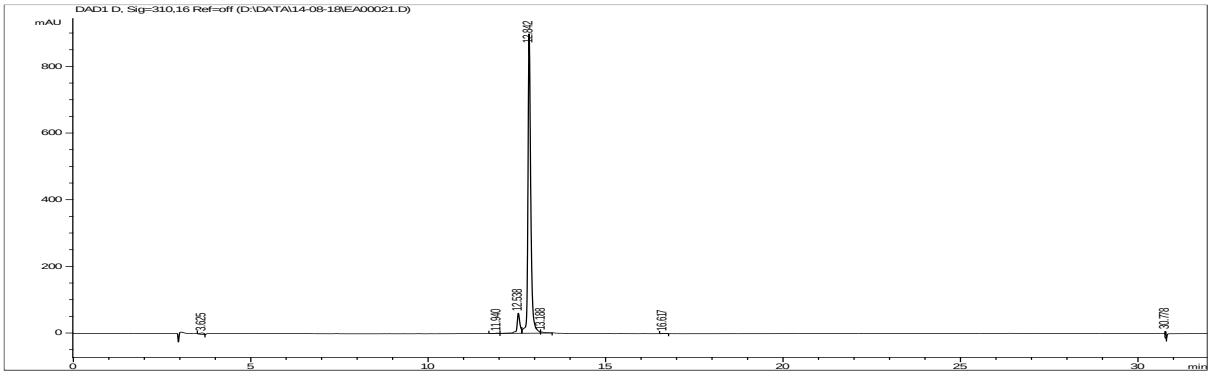


Figure 4: HPLC trace of HPLC purified polyamide **II**. r.t. = 12.842 min

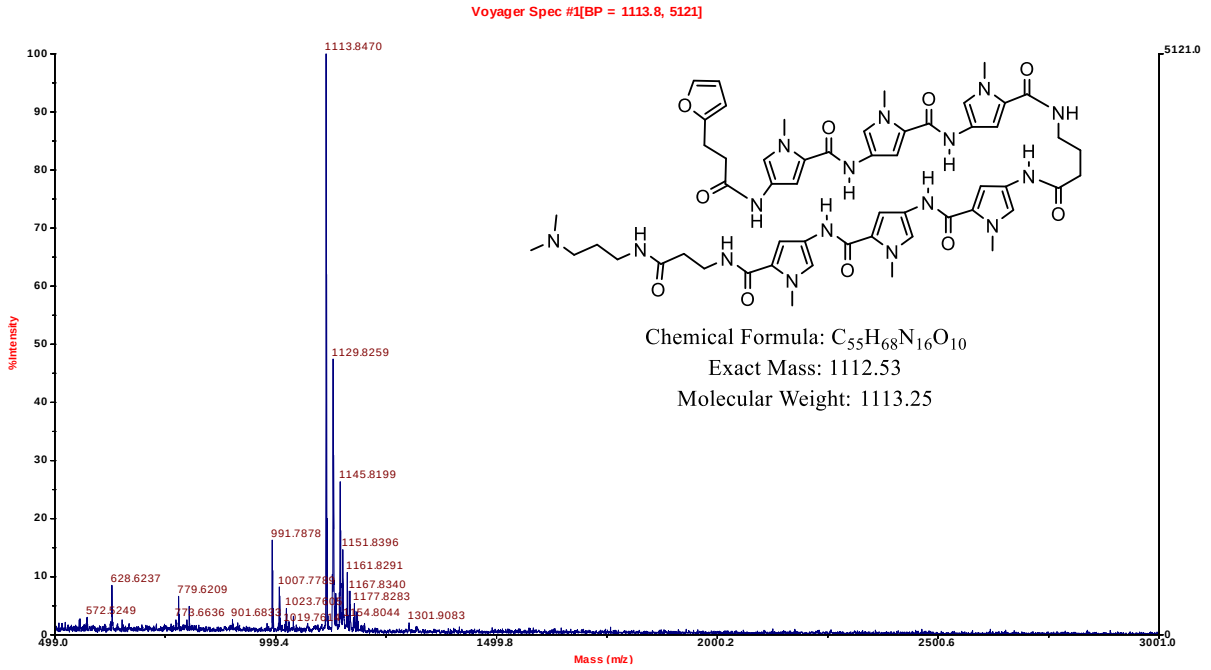


Figure 5: MALDI-TOF spectra for polyamide **II**. Exact mass calcd. 1112.53, found M+ H^+ = 1113.84.

VI. Results obtained for various oxidation conditions

1a. Polyamide concentration (polyamide **I**, no furan) = 250 µM, Rose Bengal concentration = 5 µM. Irradiation using white light (big lamp, all wavelengths), time = 2 min.

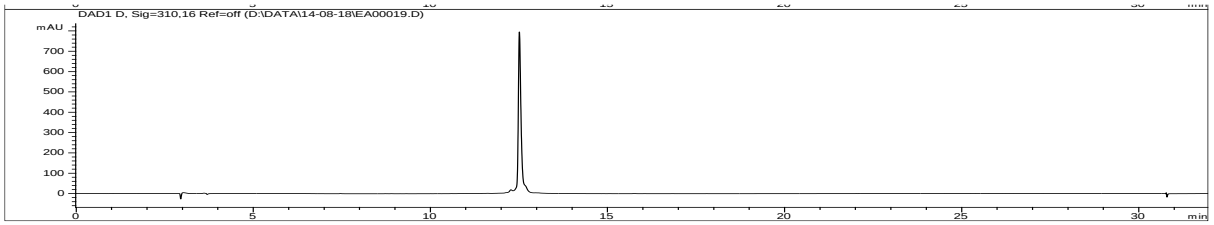


Figure 6: HPLC trace of polyamide **I** at t = 0 min.

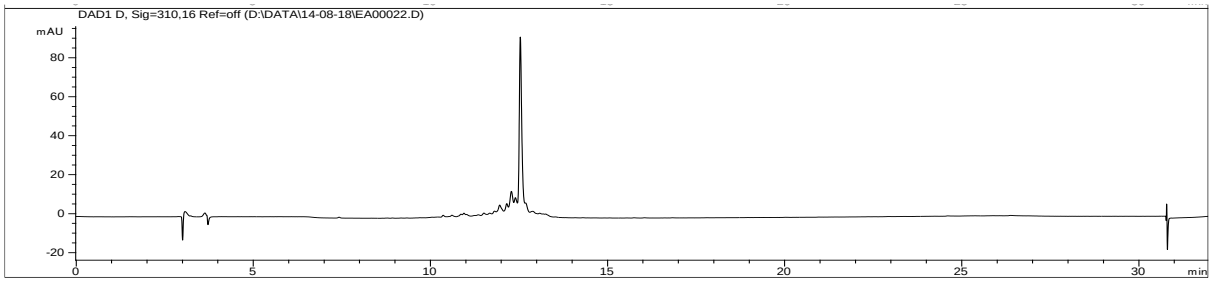


Figure 7: HPLC trace of polyamide **I** at t = 2 min, more than 80% of the polyamide has degraded (based on UV absorption)

1b. Polyamide concentration (polyamide **II**, contains furan) = 250 µM, Rose Bengal concentration = 5 µM. Irradiation using white light (big lamp, all wavelengths), time = 2 min.

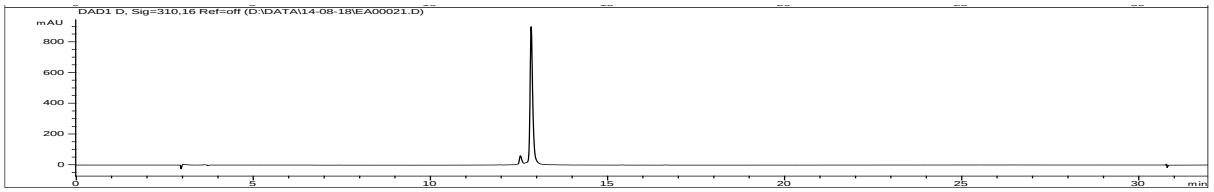


Figure 8: HPLC trace polyamide **II at** t = 0 min.

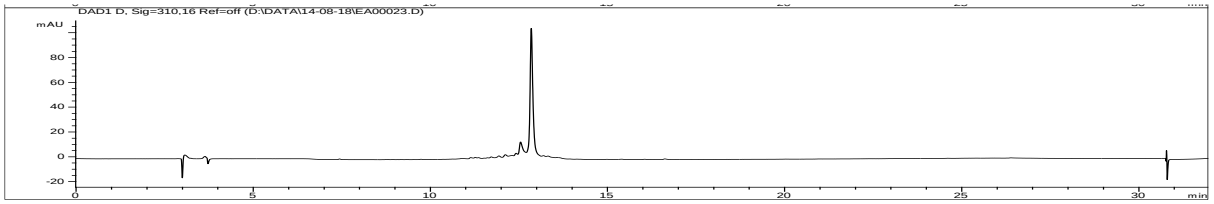


Figure 9: HPLC trace of polyamide **II** at t = 2 min, more than 80% of the polyamide has degraded (based on UV absorption)

2a. Polyamide **I** (without furan) concentration 10 µM, Rose Bengal conc. 0.5 µM, volume = 100 µL using green light filter (λ=495-570 nm). Analyzed after irradiating for 30s + 30s + 2 min

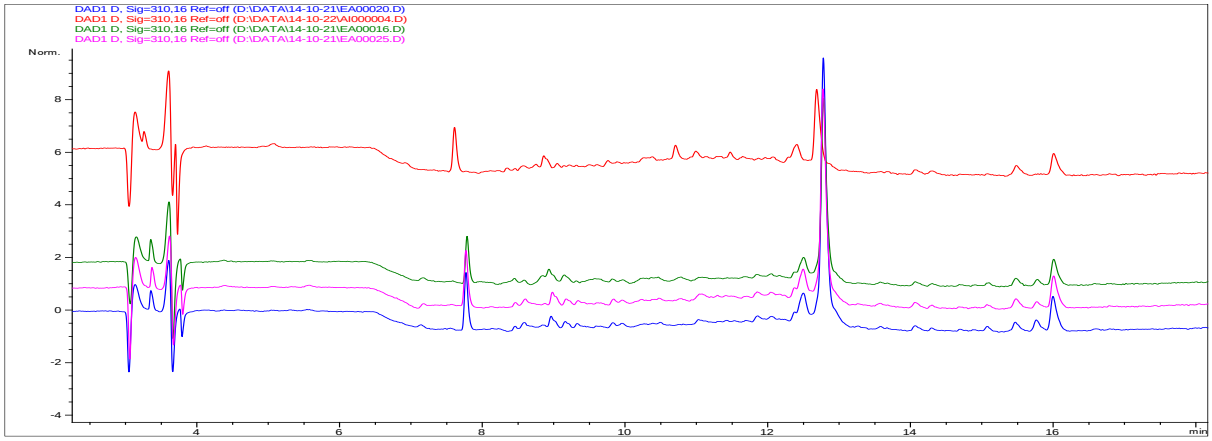


Figure 10: Overlay of HPLC traces for polyamide **I**. Blue = no light, Green = 30s irradiation, pink = 30s + 30s irradiation, Red = 30s + 30s + 2 min irradiation

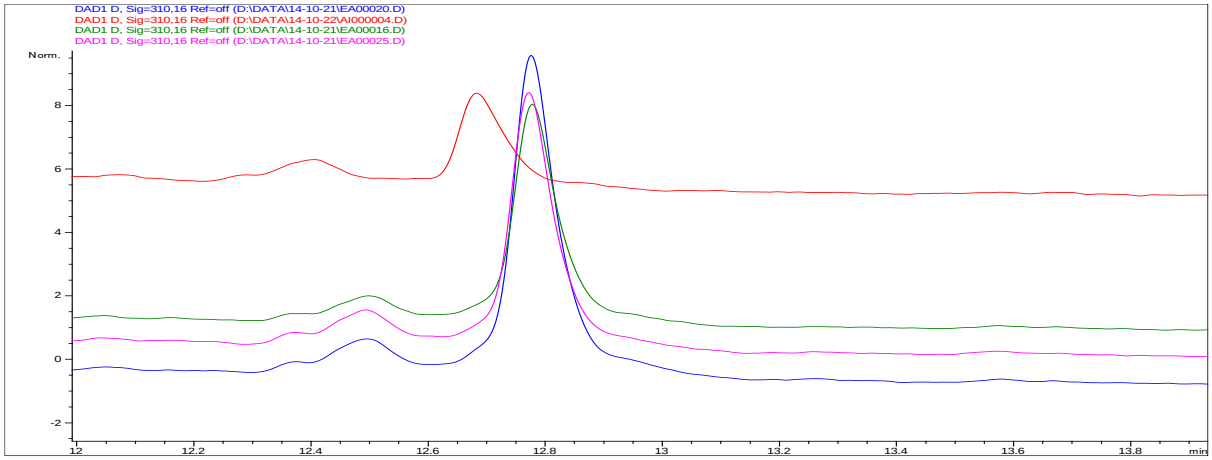


Figure 11: Overlay of HPLC traces (zoomed in) for polyamide **I**. Blue = no light, Green = 30s irradiation, pink = 30s + 30s irradiation, Red = 30s + 30s + 2 min irradiation

2b Polyamide **II** (with furan) concentration 10 µM, Rose Bengal conc. 0.5 µM, volume = 100 µL using green light filter (λ=495-570 nm). Analyzed after irradiating for 30s + 30s + 2 min

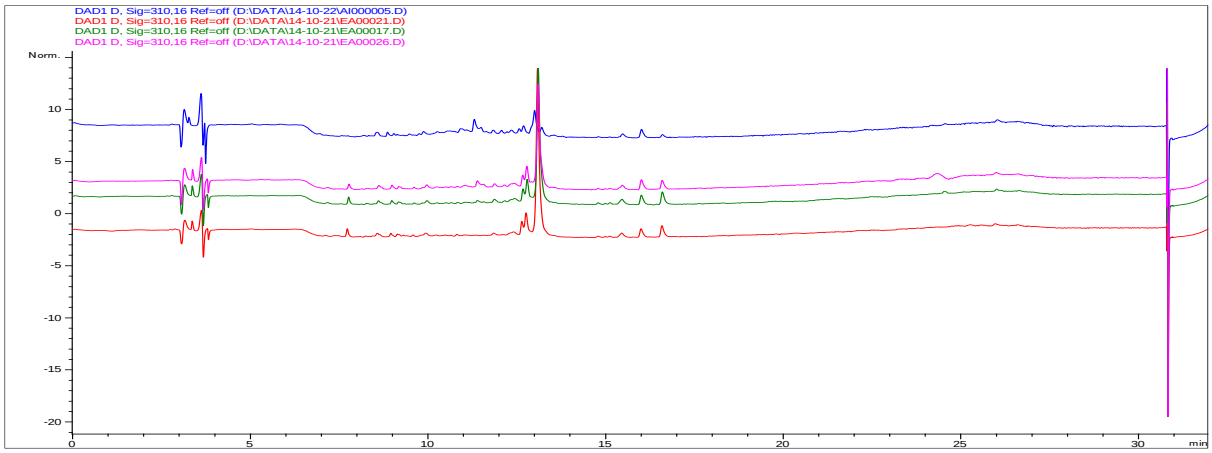


Figure 12: Overlay of HPLC traces for polyamide **II**. Red = no light, Green = 30s irradiation, pink = 30s + 30s irradiation, Blue = 30s + 30s + 2 min irradiation

- Concentration of polyamide: 10 µM, different concentrations of mCPBA, PBS buffer pH = 7.4

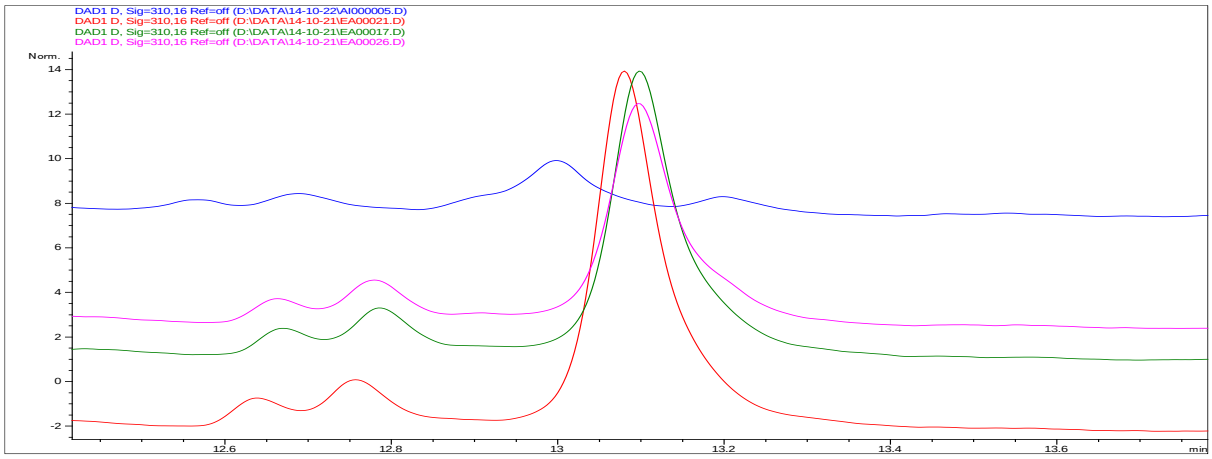


Figure 13: Overlay of HPLC traces for polyamide **II**. Red = no light, Green = 30s irradiation, pink = 30s + 30s irradiation, Blue = 30s + 30s + 2 min irradiation

3. Using mCPBA

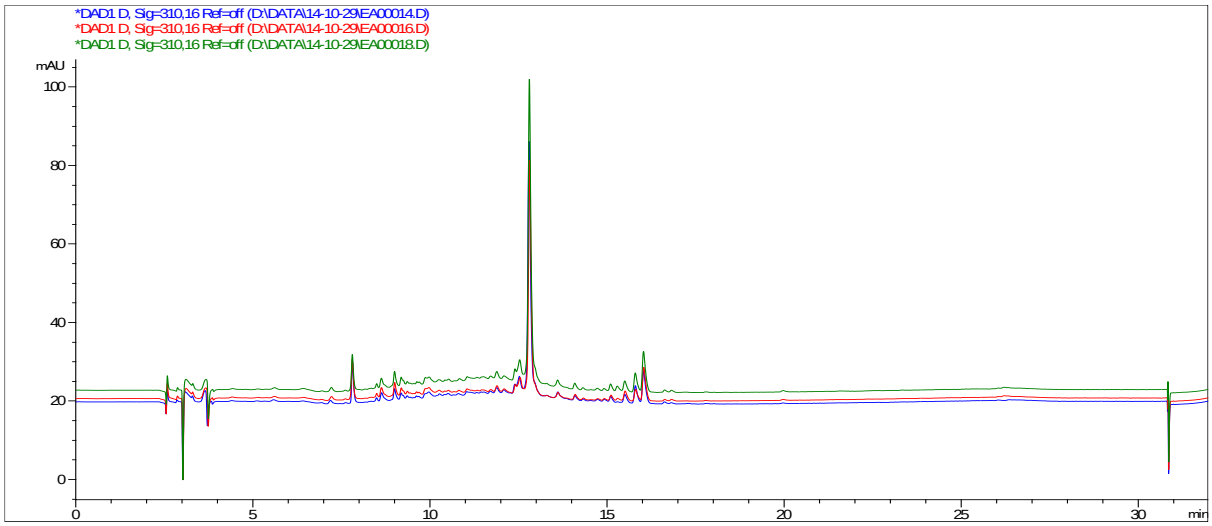


Figure 14: Overlay of HPLC traces for polyamide **I** using mCPBA. Blue = 1 eq. mCPBA. red = 2 eq. mCPBA , Green = 5 eq. mCPBA. time = 2 hrs. No change in peak was observed therefore no noticeable degradation.

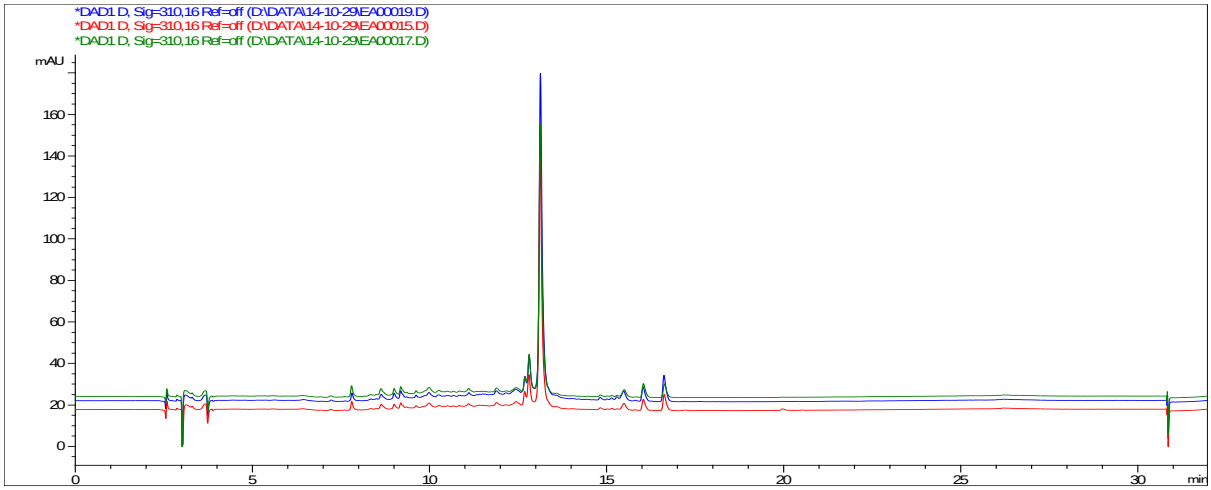


Figure 15: Overlay of HPLC traces for polyamide **II** using mCPBA. Blue = 1 eq. mCPBA. red = 2 eq. mCPBA , Green = 5 eq. mCPBA. time = 2 hrs. No change in peak was observed.

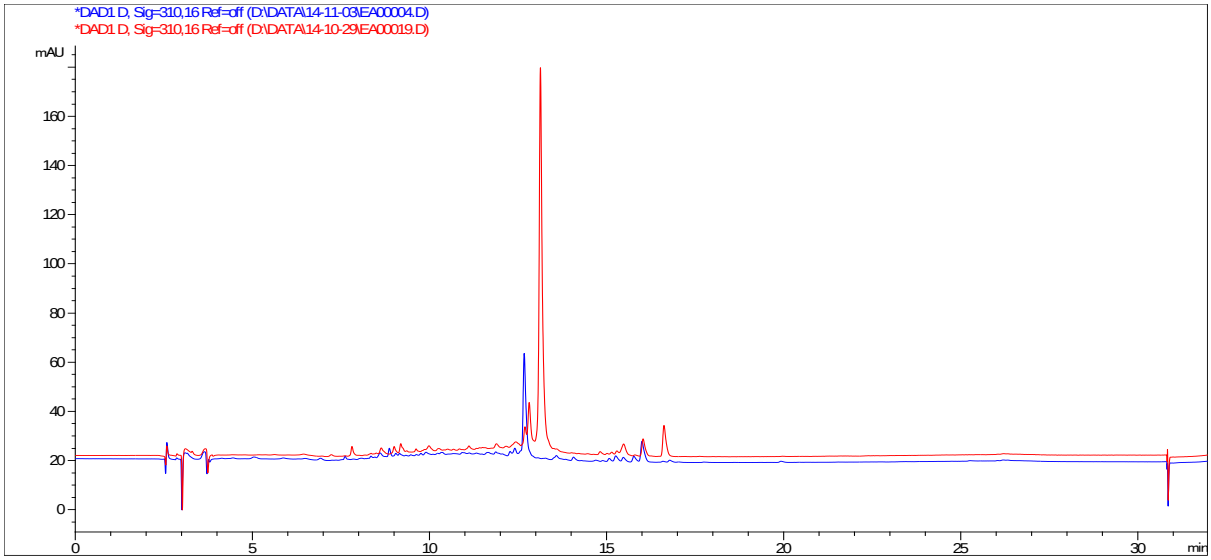


Figure 15: Overlay of HPLC traces for polyamide **II** using mCPBA. Red = 50 eq. mCPBA 6h; blue = 5 eq. mCPBA, 2 h. Different intensities due to different volumes of injection, but peak appears to have shifted.

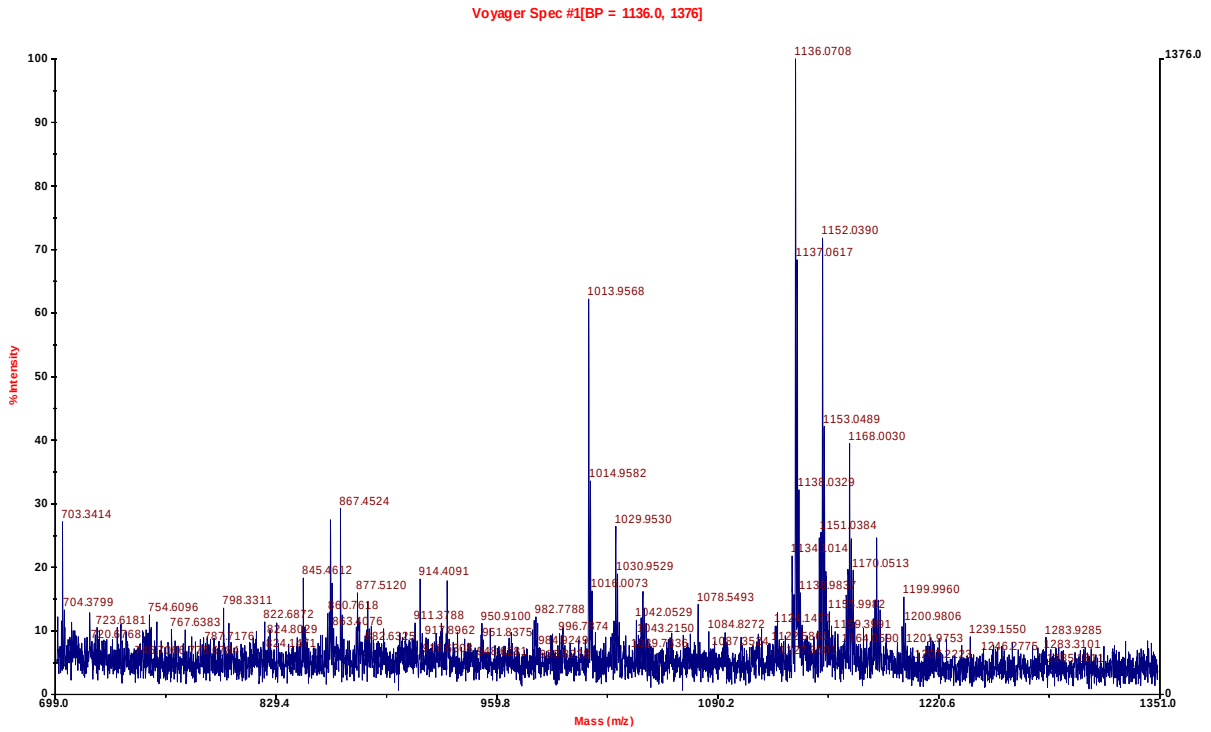


Figure 16: MALDI-TOF spectra for polyamide **II** after treating with 50 eq. mCPBA for 6h. Exact mass calcd. 1112.43, found M+ Na^+ = 1135.53, M + K^+ (or M + Na^+ + O?) = 1152.03, M + K^+ + O = 1167.53. Trace amounts of oxidation.

4. Using NBS

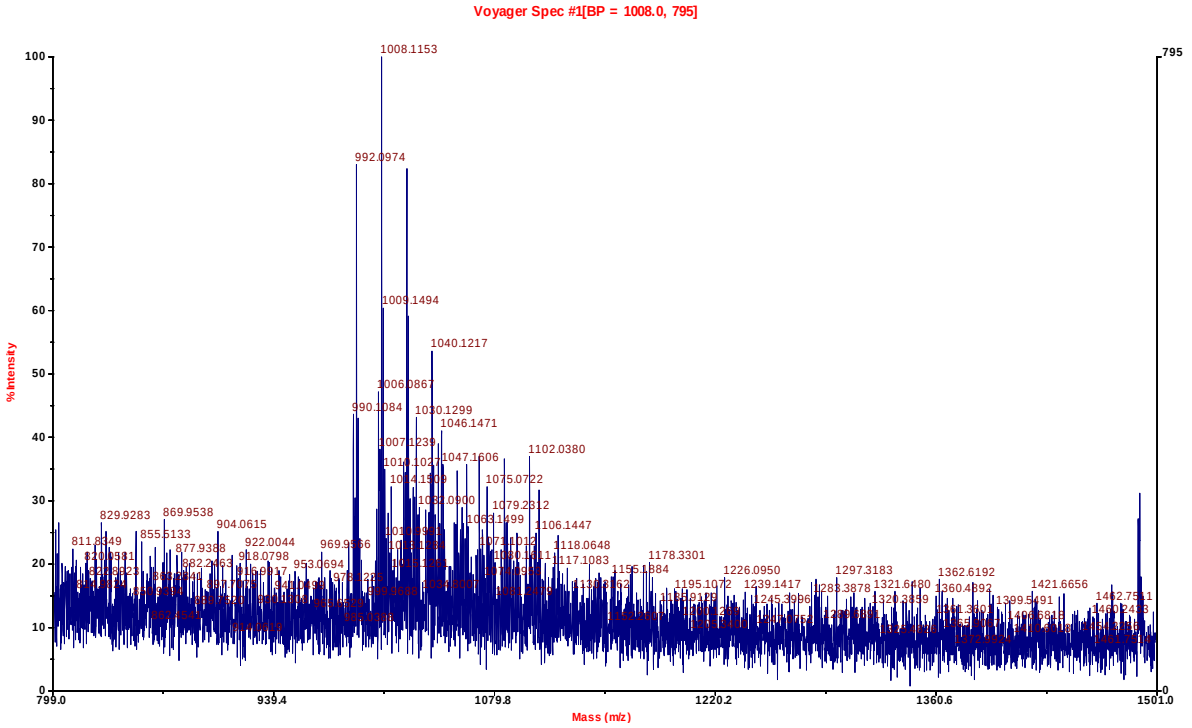


Figure 17: MALDI-TOF spectra for polyamide **I** after treating with 5 eq. NBS for 1h. Exact mass calcd. 990.49, found M+ H^+ = 992.69, found M+ Na^+ = 1013.49, M+K^+ (or M + Na^+ + O?) = 1029.49, M + H^+ + O = 1008.11. Degradation of polyamide (since no furan is present)

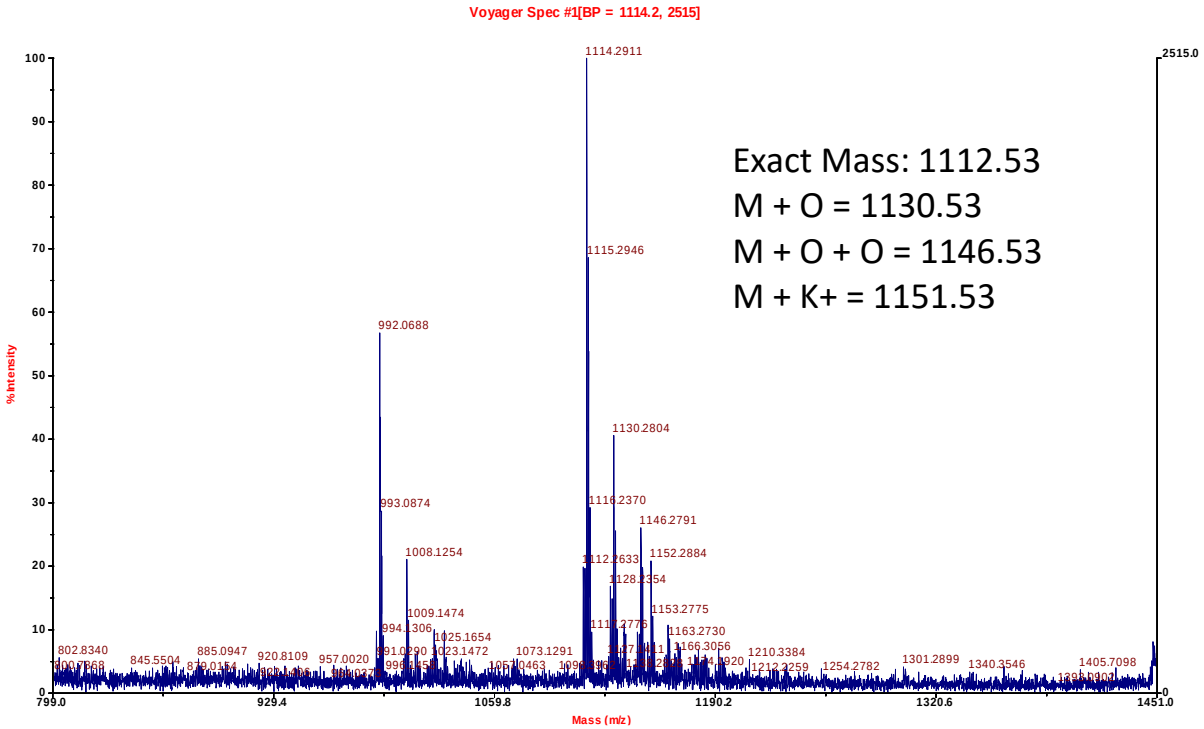


Figure 18: MALDI-TOF spectra for polyamide **II** after treating with 5 eq. NBS for 1h. Exact mass calcd. 1112.53, found M+ H^+ = 1114.29. M + O = 1130.28, M+K^+ (or M + Na^+ + O?) = 1151.53, M + O + O = 1146.27. Clear oxidation product formed along with over-oxidation/degradation.

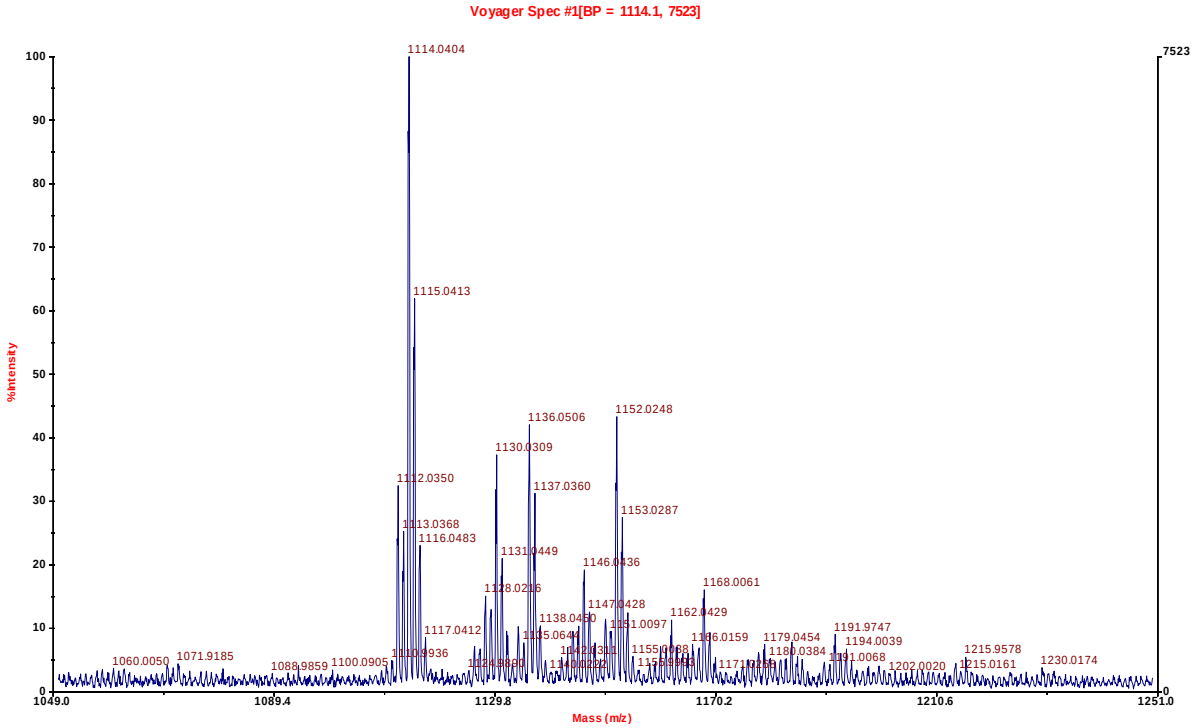


Figure 19: MALDI-TOF spectra for polyamide **II** after treating with 5 eq. NBS for 2h. Exact mass calcd. 1112.53, found M+ H^+ = 1114.04. M + O = 1130.28, M+K^+ (or M + Na^+ + O?) = 1152.02, M + O + O = 1146.27. Clear oxidation product formed along with over-oxidation/degradation.

EXPERIMENTAL SECTION FOR CHAPTER 4

I. Materials

All amino acids, Trifluoroacetic acid (TFA) and coupling reagent HBTU were purchased from Iris GmbH. L-amino acids were used throughout the syntheses. The protecting groups for the amino acids are ^{t}Bu for Asp, Glu, Ser, Thr, Boc for Lys, Pbf for Arg and Trt for Cys, Asn, Gln. 4-acetamidobenzoic acid (ABA), 4-4' Bis(bromomethyl)biphenyl, α,α'-Dibromo-m-xylene, p-Phenylenediamine, Ethyl Acetate, Acetonitrile, Methanol, Diethyl Ether DIPEA, supplied as extra dry, redistilled, 99.5 % pure and Triisopropylsilane were purchased from Sigma Aldrich. Bromoacteylbromide was purchased from Acros Organics. Dimmethylformamide (DMF) and N-Methylpyrolidine (NMP) peptide synthesis grade were purchased from Biosolve. HPLC grade quality Hexane and Chloroform were purchased from Fisher Scientific. Deuterated solvents D_2O (99.9% atom D) and DMSO-d6 (99.8% atom D) were obtained from Eurisotop. Water with the Milli-Q grade standard was obtained in-house either from a Millipore ROs 5 purification system or a Sartorius Arium 611 DI. Rink-Amide ChemMatrix (100-200 μm, manufacturer's loading: 0.52-0.54 mmol/g) was obtained from Biotage. NHS-Fluorescein was purchased from Thermo Scientific as a 5-6 isomer. All reagents were acquired from commercial sources and used without prior purification. All chemicals were used without further purification. All oligonucleotides used were commercially purchased from Eurogentec (HPLC purified using a RP-cartridge-Gold, 200 nm scale) and were used as such.

II. Methods

1. Peptide syntheses

Automated peptide syntheses were performed on a 24-reactor block SYRO Multiple Peptide Synthesizer equipped with a vortexing unit (Multisyntech, Witten, Germany). Peptides were synthesized by standard Fmoc/tBu strategy using HBTU/DIPEA couplings. All automated peptide synthesis were performed using Rink Amide Chemmatrix resin, loading = 0.54 mmol/g on a 100 mg scale using 10 eq. of amino acid, 10 eq. of HBTU and 20 eq. of DIPEA in DMF using single couplings of 1h. Fmoc deprotection was performed using 20% piperidine in DMF and shaking for 2, 5 and 15 min. For coupling fluorescein to the N-terminus 3 eq. of NHS fluorescein along with 3 eq. HOBt and 6 eq. of DIPEA was used and the reaction mixture was shaken overnight. The peptides were cleaved off using TFA/TIS/H_2O = 95/2.5/2.5 and precipitated using cold MTBE.

For peptide stapling, 10 mg of crude peptide was dissolved in 2-3 mL of NH_4HCO_3 solution (depending on the solubility) with 1.5 eq. of tris(2-carboxyethyl) phosphine (TCEP) and was stirred for 15 min at room temperature. Then 2.5 eq. of cross-linker dissolved in 1 mL of dry DMF was added drop-wise to the reaction mixture and was stirred for another 2-2.5h after which the crude reaction mixture was analyzed by MALDI-TOF to observe the progress. If starting material was detected, 1.5 eq. of cross-linker in 1 mL dry DMF was added and the reaction was continued for another 1.5-2h. MALDI-TOF samples were taken every 45 min. The solvents were removed and the peptides were obtained by precipitation (twice) by centrifugation at 4000 rpm at 0°C in cold MTBE (-20°C). The pellet was then redissolved in mQ H_2O and purified using RP-HPLC to give pure stapled peptides.

2. Analyses

LC-MS data were collected on an Agilent 1100 Series instrument with a Phenomenex Kinetex C18 100Å column (150 x 4.6 mm, 5 µm at 35 °C) connected to an ESMSD type VL mass detector with a flow rate of 1.5 ml/min was used with the following solvent systems: (A): 0.1% HCOOH in H_2O and (B) MeCN. The column was flushed with 100% A for 2 min, then a gradient from 0 to 100% B over 6 min was used, followed by 2 min of flushing with 100% B. RP-HPLC analyses and purification were performed on an Agilent 1100 Series instrument with a Phenomenex Luna C18 column (250 x 4.6 mm, 5 µm at 35 °C). A flow rate of 1 ml/min was used with the following solvent systems: (A): 0.1% TFA in H_2O and (B): MeCN. The column was flushed for 2 min with 100% A, then a gradient from 0 to 100% B over 15 min was used, followed by 5 min of flushing with 100% B. MALDI-TOF-MS spectra were acquired with a high performance nitrogen laser (337 nm), using the positive and reflector mode with delayed extraction. The matrix solution utilized was as follows: 4-5 mg α-cyano-4-hydroxycinnamic acid in 500µL MeCN, 490µL mQ, 10µL 1M ammoniumcitrate, 1µL TFA. CD spectra were measured using an Aviv 410 CD spectrophotometer equipped with a Peltier temperature control unit

III. Peptide Synthesis and HPLC/LC-MS/MALDI-TOF Data

1. Protocol for deprotection of peptides

Protected peptide-resin is treated with TFA/TIS/mQ = 95:2.5:2.5 (4 mL/100 mg). After 4 hours, the liquid is filtered of and the resin is washed 2 times with neat TFA. The filtrate is evaporated followed by precipitation (twice) by centrifugation at 4000 rpm at 0°C in cold MTBE (-20°C). The peptide was then redissolved in milliQ H_2O for RP-HPLC, LC-MS analysis, RP-HPLC purification or lyophilized for the next step.

2. Syntheses of peptides 1-7

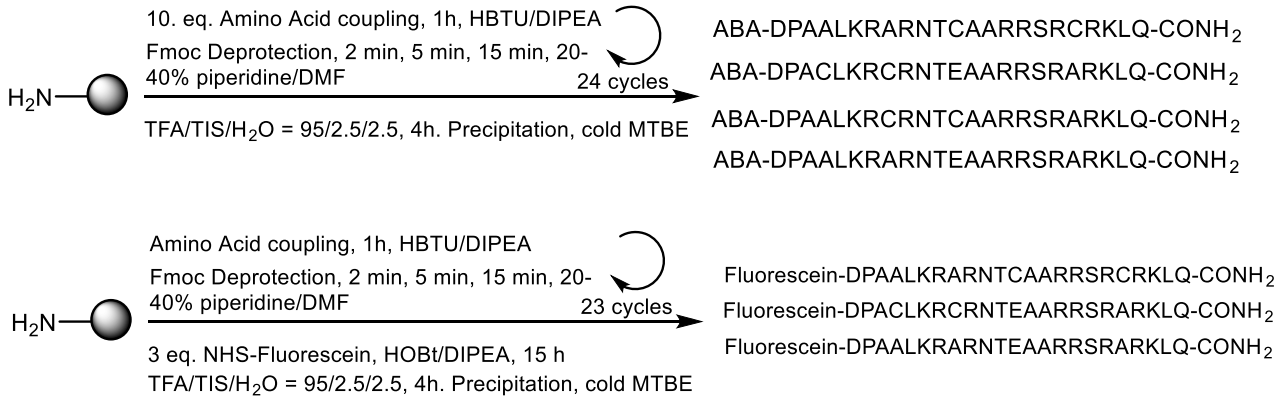


Figure 1: Syntheses of peptides **1-7**

The resin is preswollen in NMP for 10 min and then filtered off. Peptide synthesis is performed on an automated peptide synthesizer using the following protocols for Fmoc deprotection and coupling:
Fmoc deprotection: Fmoc deprotection is performed by adding a solution of 20% piperidine in NMP to the resin and shaking for 2 min, 5 min & 15 min durations. After each addition and shaking cycle, the resin is filtered off and washed with NMP (6 x 45 s).
Coupling: 10 equivalents of a 0.5 M solution of amino acid in NMP, 10 equivalents of a 0.5 M solution of HBTU in DMF and 10 equivalents of a 2.0 M solution of DIPEA in NMP are added to the resin. The

reaction mixture is shaken for 1 hour. The resin is filtered off and washed with NMP (9 x 2 min). The peptide is analyzed according to the above protocol for deprotection of peptides.
Capping: ABA was used for capping and was coupled using the protocol in *coupling* above.
Coupling of Fluorescein: 3 eq. of NHS-fluorescein predissolved in 0.5 mL DMF. 6 eq. of DIPEA and 3 eq. of HOBt predissolved in 0.5 mL DMF and are added successively to the resin. After shaking for 15 h, the reaction mixture was filtered off and the resin washed with 3x DMF, 3x MeOH, 3x DMF and 3x Et_2O.

No.	Sequence	Chemical Formula	Exact Mass	Obsd. Mass
1	ABA-DPAALKRARNTCAARRSRCRKLQ-$CONH_2$	$C_{115}H_{201}N_{47}O_{31}S_2$	2800.50	2801.88
2	ABA-DPACLKRCRNTEAARRSRARKLQ-$CONH_2$	$C_{117}H_{203}N_{47}O_{33}S_2$	2858.51	2859.47
3	ABA-DPAALKRCRNTCAARRSRARKLQ-$CONH_2$	$C_{115}H_{201}N_{47}O_{31}S_2$	2800.50	2801.60
4	ABA-DPAALKRARNTEAARRSRARKLQ-$CONH_2$	$C_{117}H_{203}N_{47}O_{33}$	2794.57	2795.28
I	Fluorescein-DPAALKRARNTCAARRSRCRKLQ-$CONH_2$	$C_{127}H_{204}N_{46}O_{35}S_2$	2997.50	2997.57
II	Fluorescein-DPACLKRCRNTEAARRSRARKLQ-$CONH_2$	$C_{129}H_{206}N_{46}O_{37}S_2$	3055.51	3056.61
IV	Fluorescein-DPAALKRARNTEAARRSRARKLQ-$CONH_2$	$C_{129}H_{206}N_{46}O_{37}$	2991.57	2992.54

Table 1: Calculated and observed masses of peptides

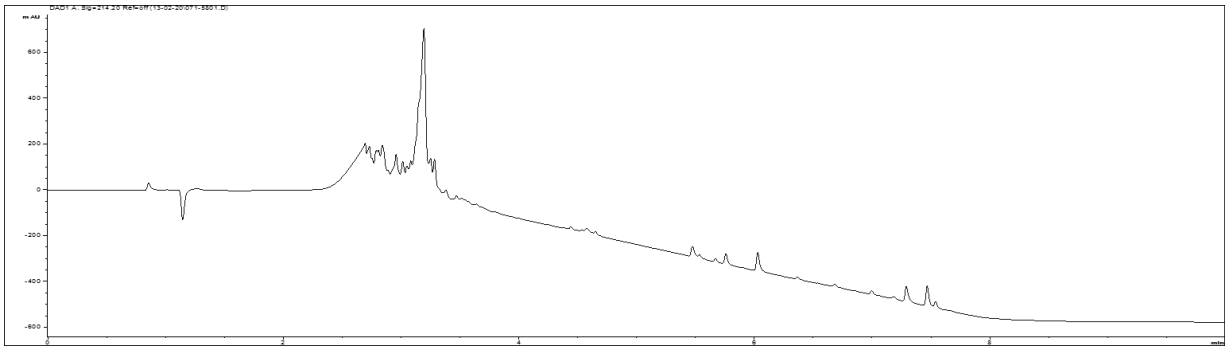


Figure 2: RP-HPLC trace from LC-MS of peptide **1** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

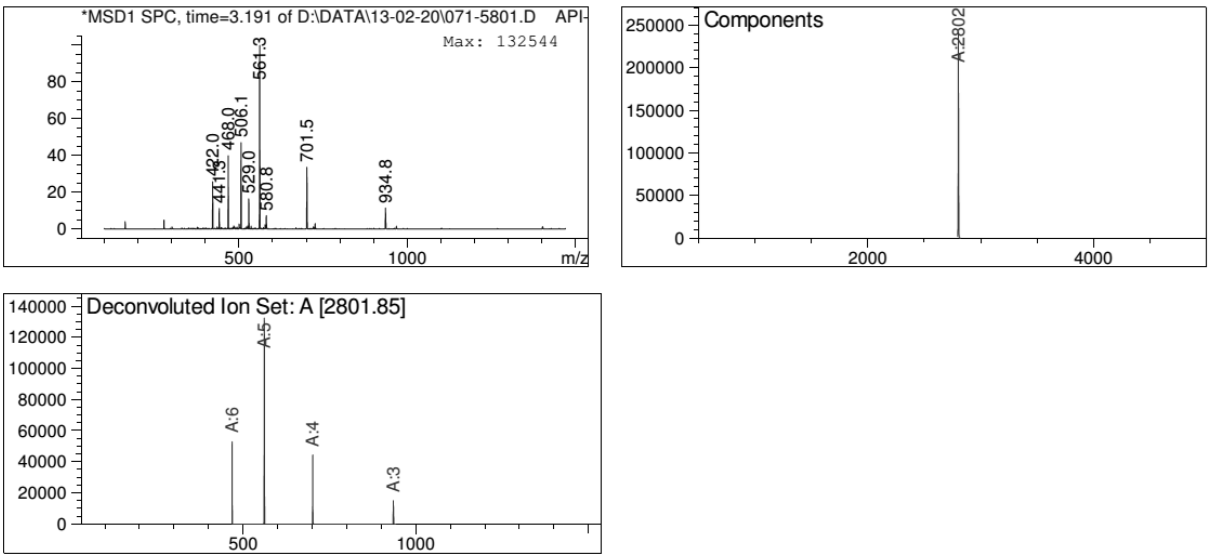


Figure 3: ESI-MS from LC-MS at t = 3.191 min. E.M. Calcd. for $C_{115}H_{201}N_{47}O_{31}S_2$ = 2800.50 and deconvoluted mass found 2801.85

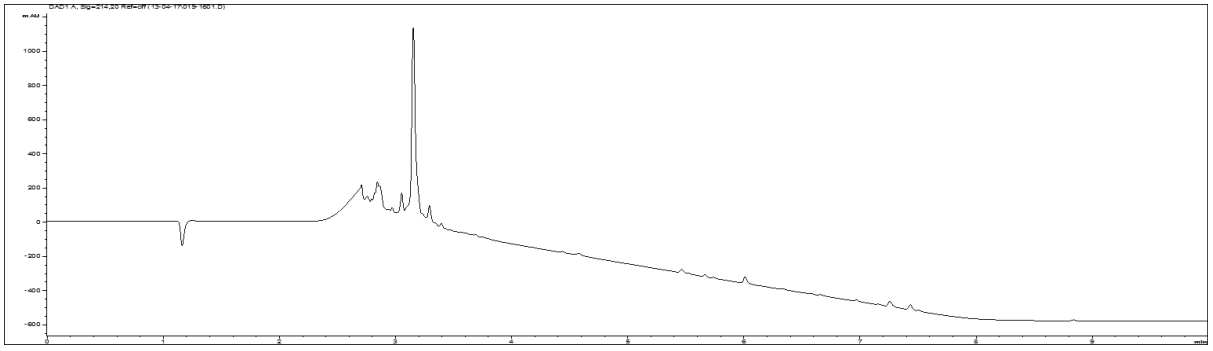


Figure 4: RP-HPLC trace from LC-MS of peptide **2** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

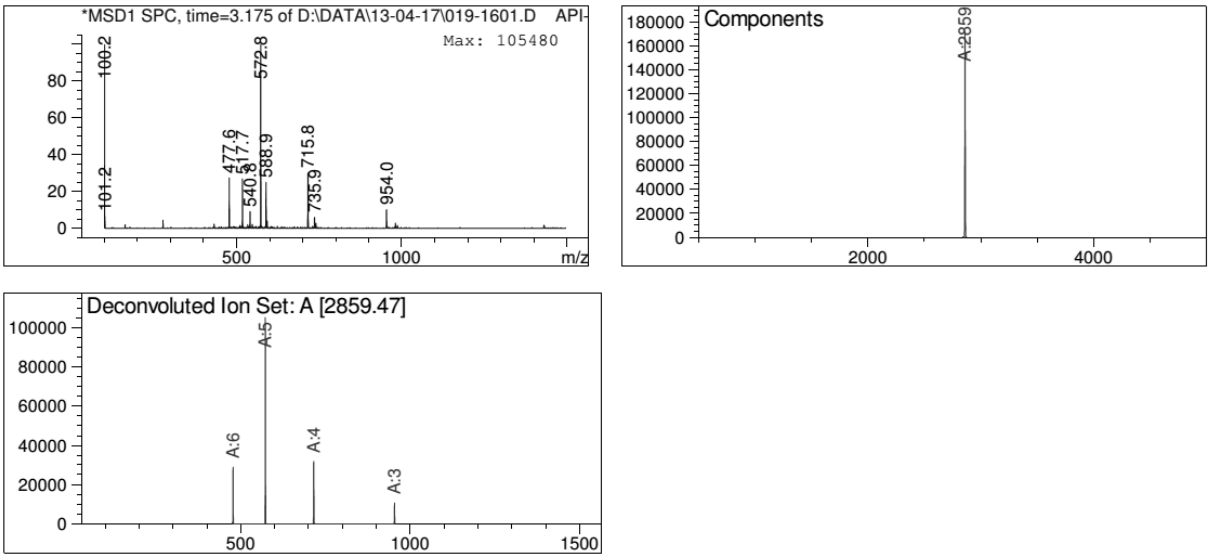


Figure 5: ESI-MS from LC-MS at t = 3.175 min. E.M. Calcd. for $C_{117}H_{203}N_{47}O_{33}S_2$ = 2858.51 and deconvoluted mass found 2859.47

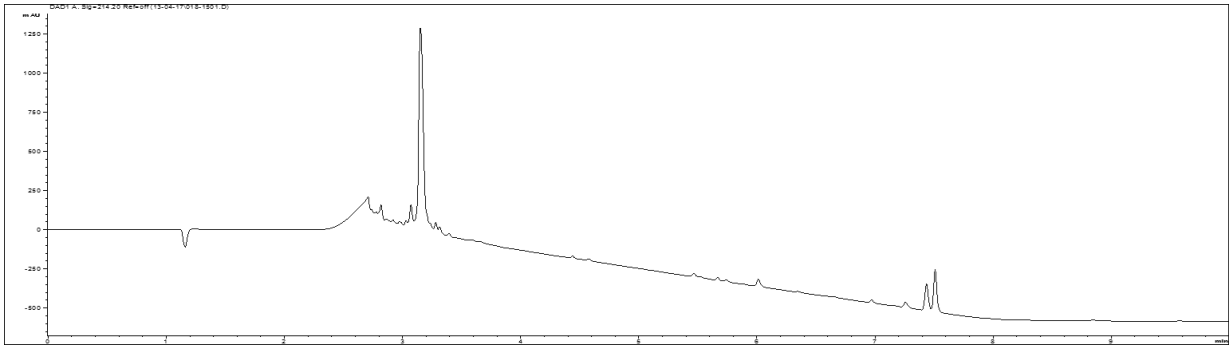


Figure 6: RP-HPLC trace from LC-MS of peptide **3** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

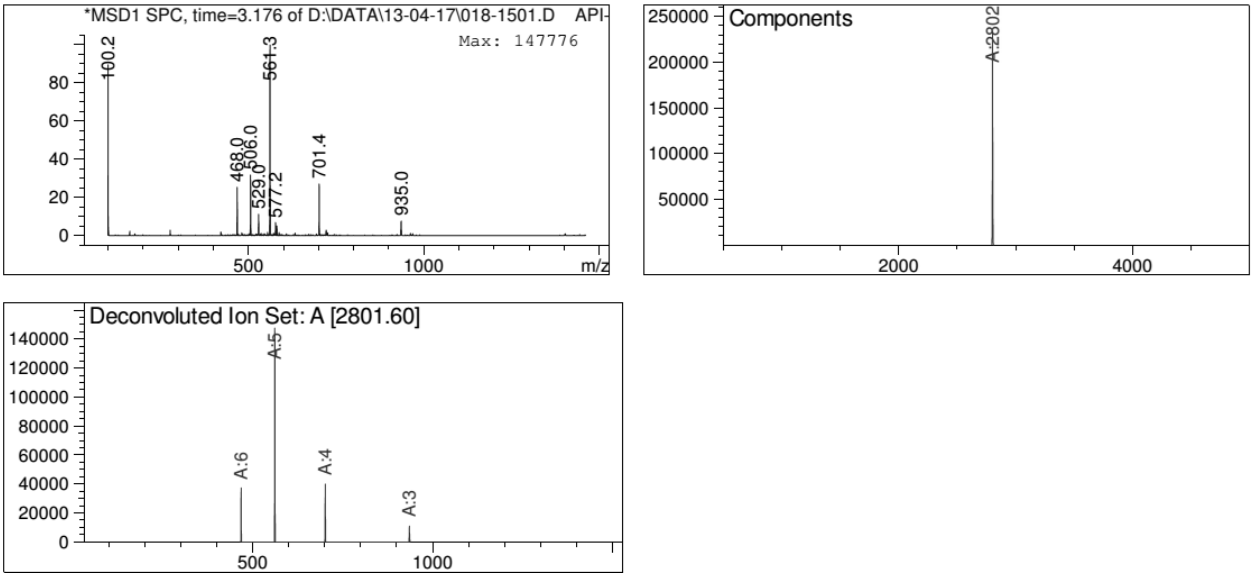


Figure 7: ESI-MS from LC-MS at t = 3.176 min. E.M. Calcd. for $C_{115}H_{201}N_{47}O_{31}S_2$ = 2800.50 and deconvoluted mass found 2801.60

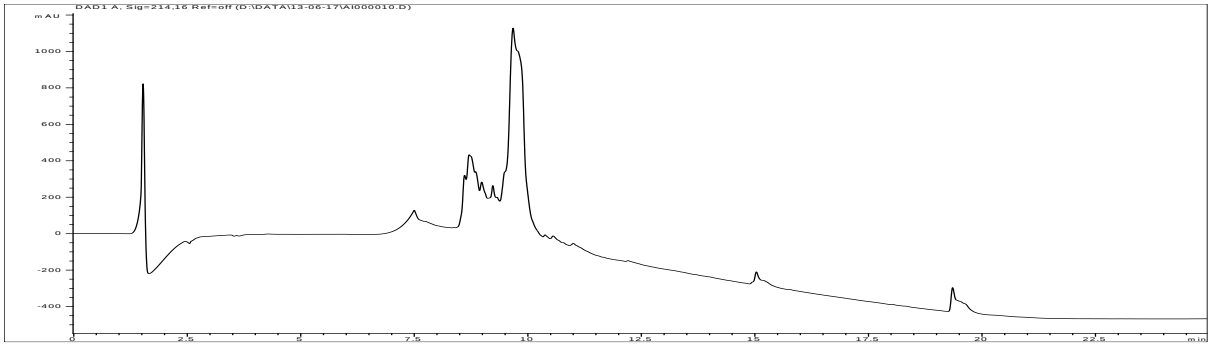


Figure 8: RP-HPLC trace of peptide **4** (0-100% ACN in 15 min on Luna C18(2) 100 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

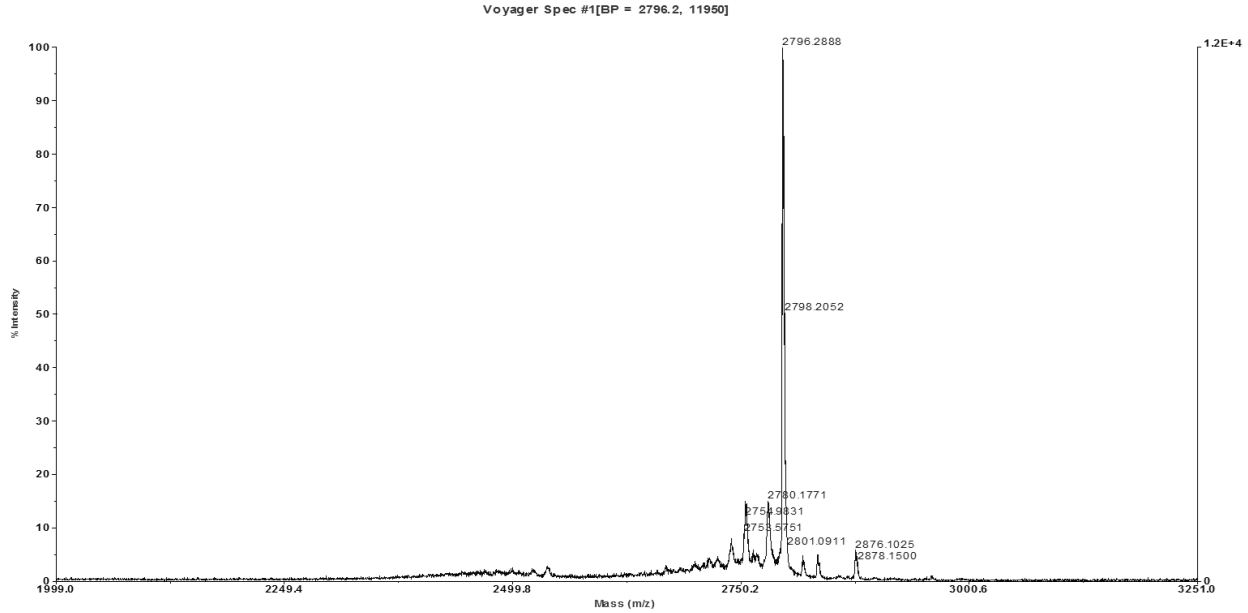


Figure 9: MALDI-MS spectrum of peak at t_R= 9.681. E.M. Calcd. for $C_{115}H_{201}N_{47}O_{31}S_2$ = 2794.57 and mass found $M+H^+$= 2796.28

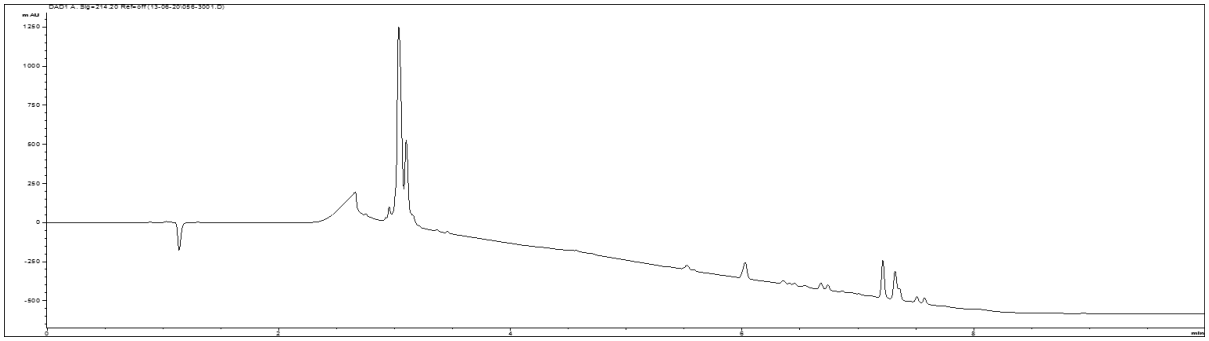


Figure 10: RP-HPLC trace from LC-MS of HPLC purified peptide **4** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

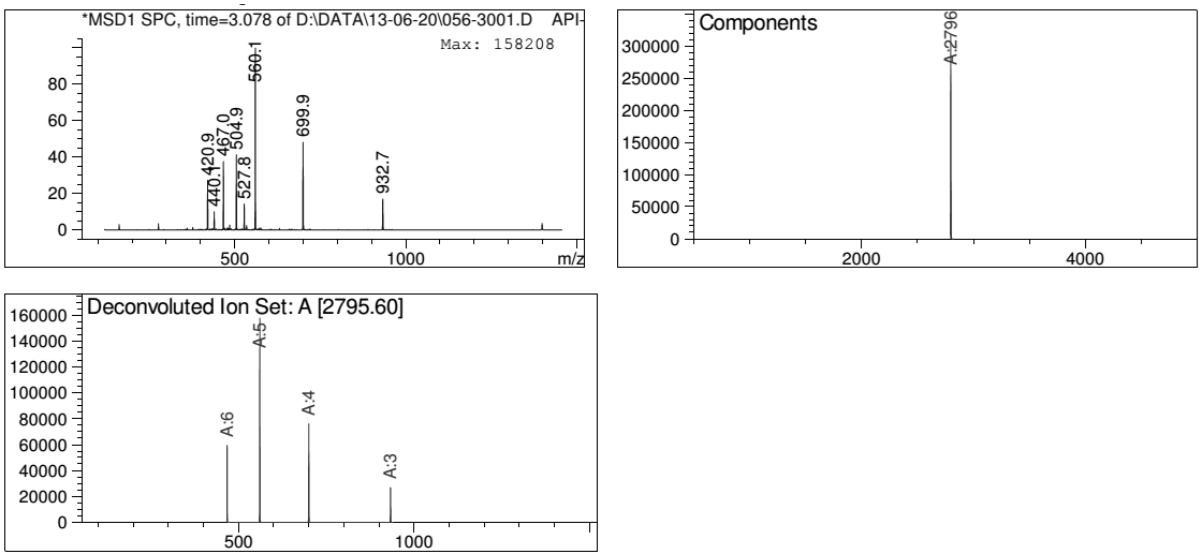


Figure 11: ESI-MS from LC-MS at t = 3.078 min. E.M. Calcd. for $C_{115}H_{201}N_{47}O_{31}S_2$ = 2794.57 and deconvoluted mass found 2795.60

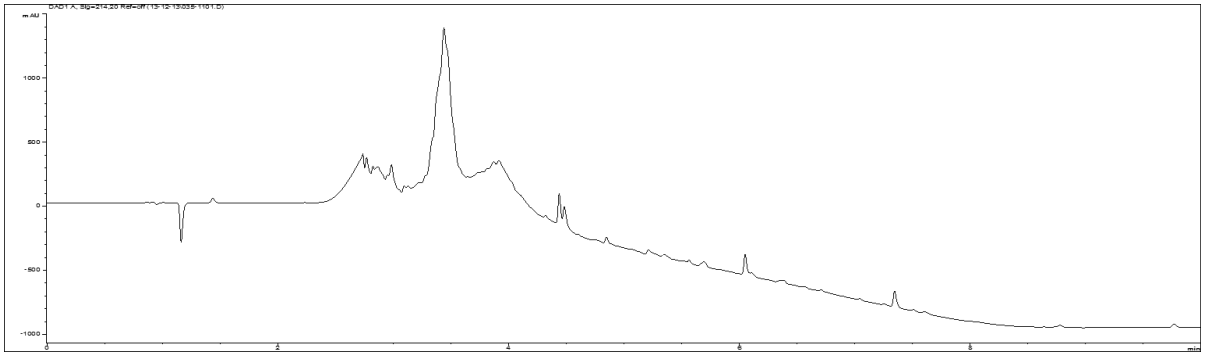


Figure 12: RP-HPLC trace from LC-MS of peptide **I** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

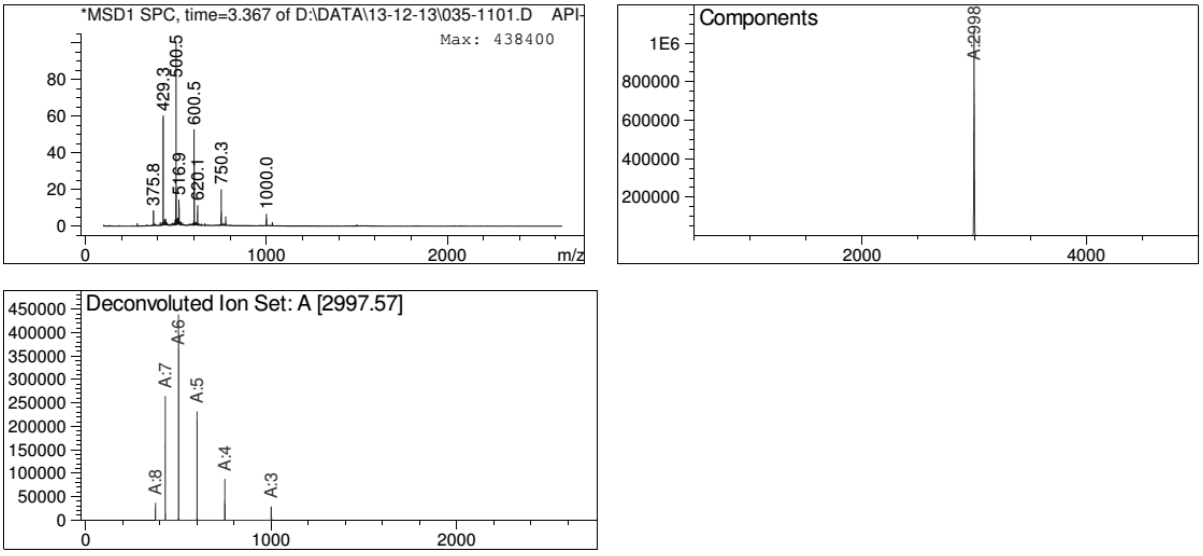


Figure 13: ESI-MS from LC-MS at t = 3.367 min. E.M. Calcd. for $C_{127}H_{204}N_{46}O_{35}S_2$ = 2997.50 and deconvoluted mass found 2997.57

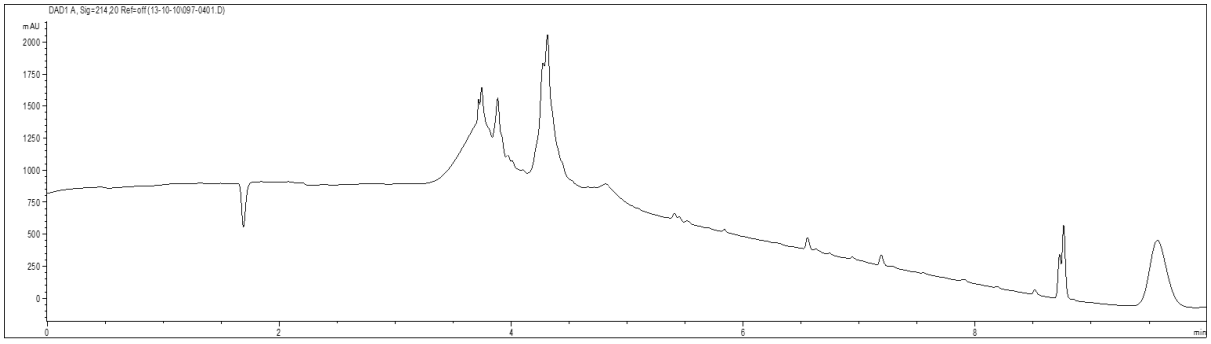


Figure 14: RP-HPLC trace from LC-MS of peptide **II** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

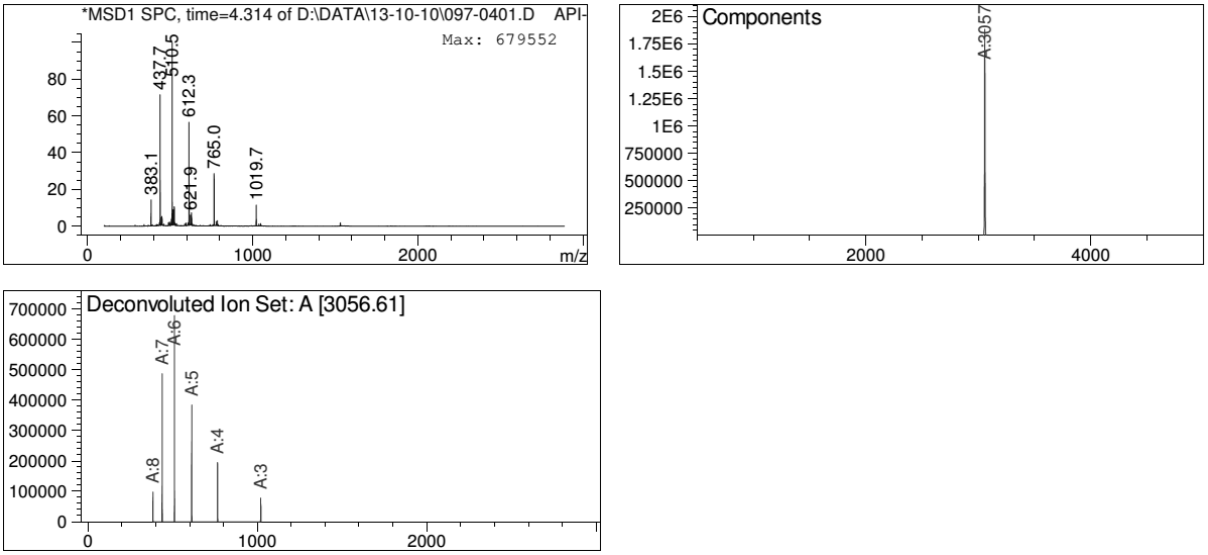


Figure 15: ESI-MS from LC-MS at t = 4.314 min. E.M. Calcd. for $C_{129}H_{206}N_{46}O_{37}S_2$ = 3055.51 and deconvoluted mass found 3056.61

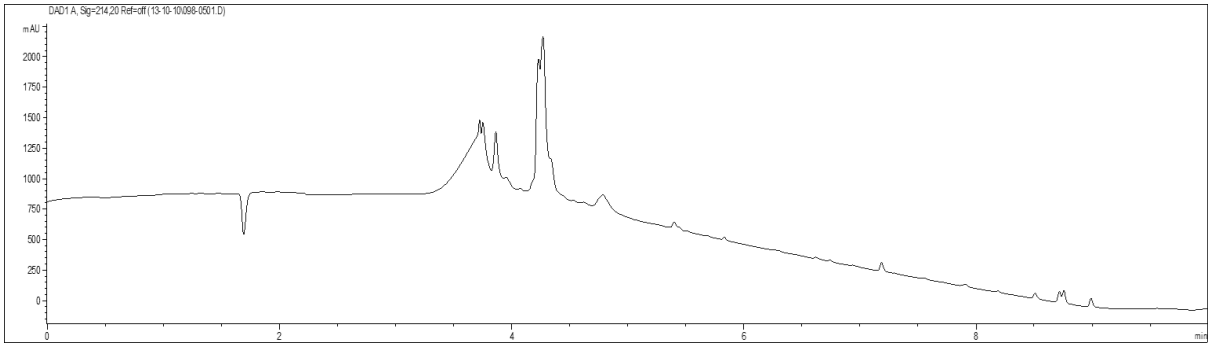


Figure 16: RP-HPLC trace of crude peptide **IV** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

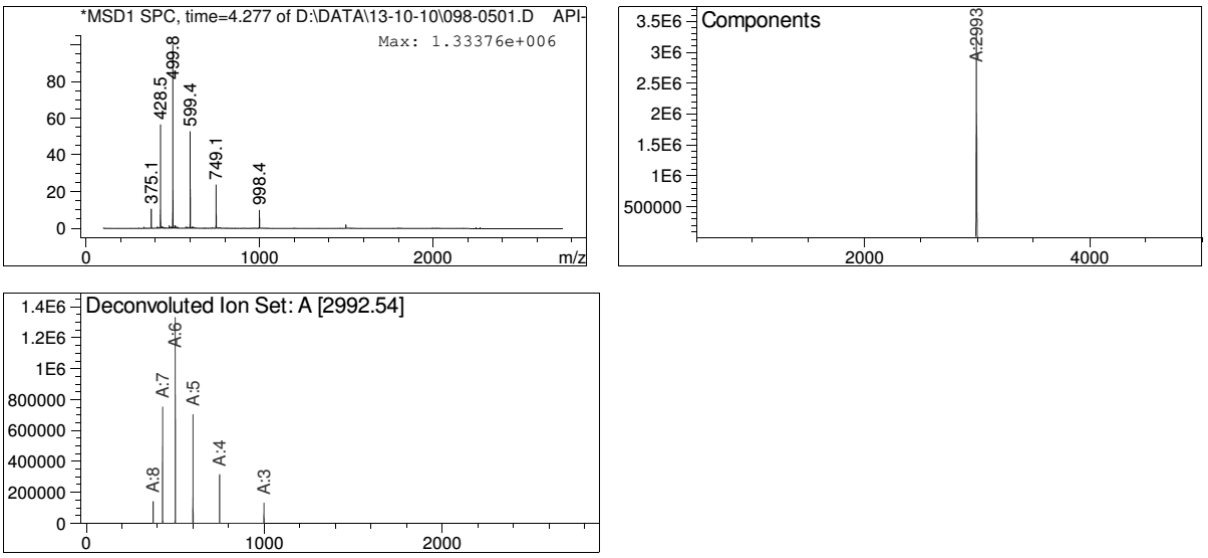


Figure 17: ESI-MS from LC-MS at t = 4.277 min for crude compound **7**. E.M calcd. for $C_{129}H_{206}N_{46}O_{37}$ = 2991.57 and deconvoluted mass found 2992.54.

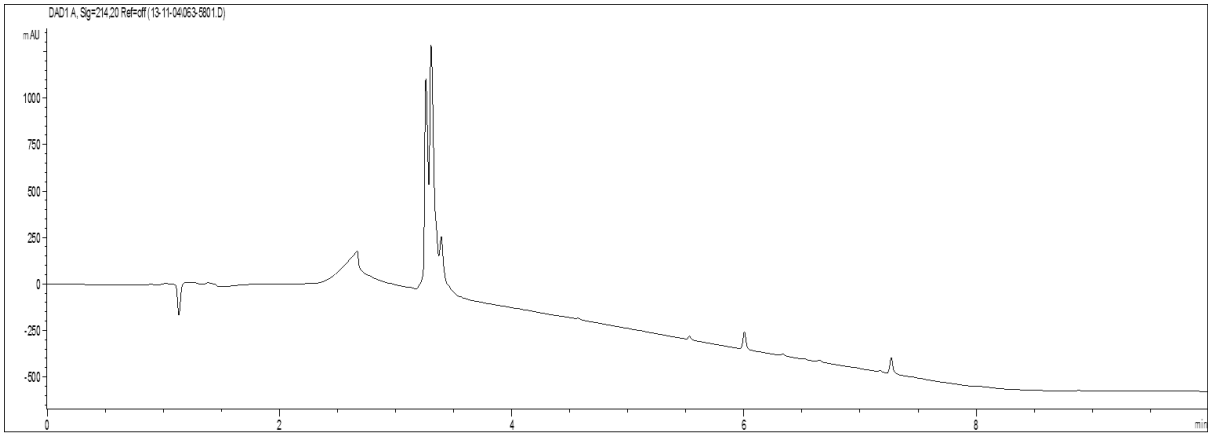


Figure 18: RP-HPLC trace of HPLC purified peptide **IV** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

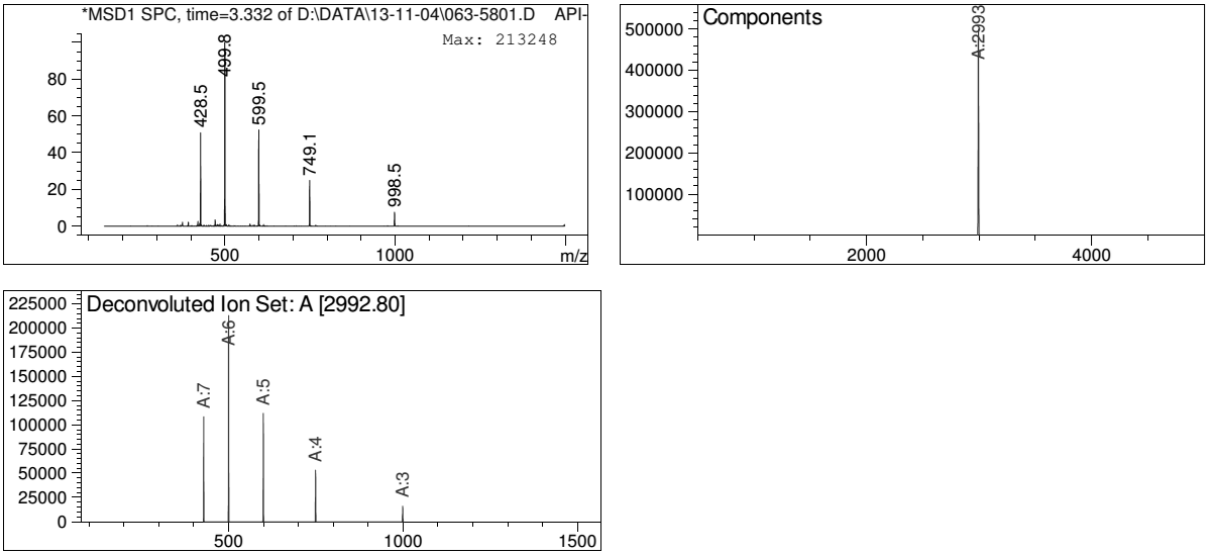
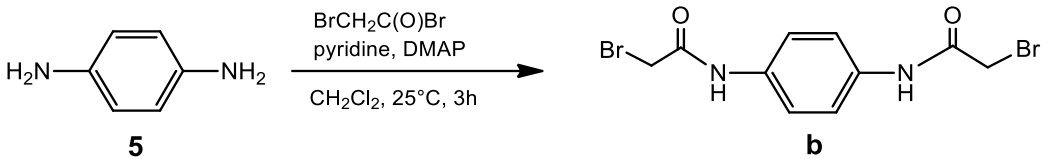


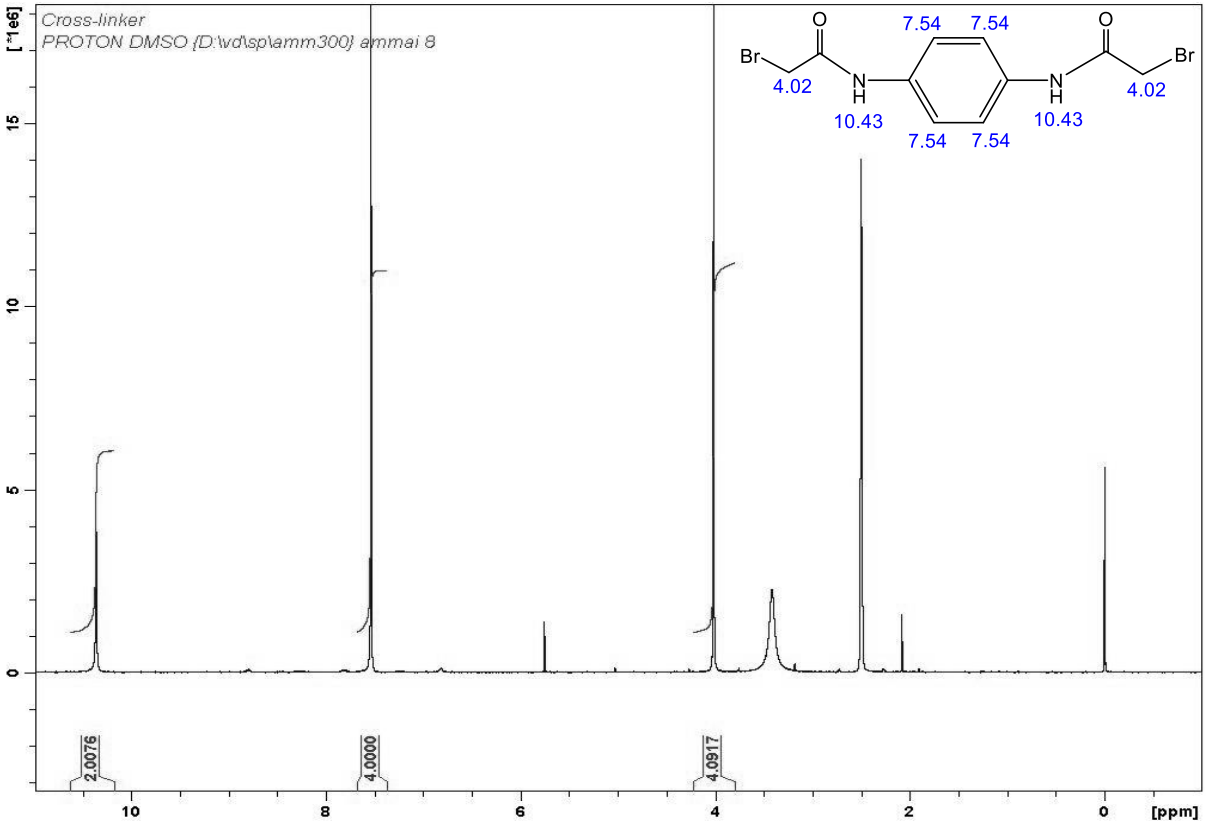
Figure 19: ESI-MS from LC-MS at t = 3.332 min. E.M calcd. for HPLC purified peptide **7** $C_{129}H_{206}N_{46}O_{37}$ = 2991.57 and deconvoluted mass found 2992.80.

IV. Synthesis of N,N'-(1,4-phenylene)bis(2-bromoacetamide)

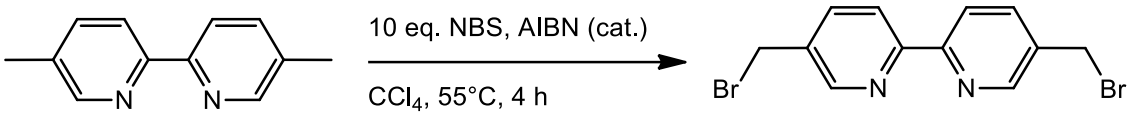


5 **b**

To a solution of p-Phenylenediamine **5** (54 mg, 0.5 mmol) in CH_2Cl_2 (5 mL), dry pyridine (96 µL, 1.2 mmol) and 4-(dimethylamino)pyridine (1.5 mg) was added bromoacetyl bromide (96 µL, 1.1 mmol) at 0°C, and the mixture was stirred at 25°C for 3h. The resulting precipitates were collected by filtration and washed with CH_2Cl_2 (3 x 5 mL) to give *N*,*N*'-(1,4-phenylene)bis(2-bromoacetamide) **b** (145 mg, 0.41 mmol, 82%) as a white amorphous solid and therefore a melting point was not determined for the compound. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 4.02 (s, 4H), δ 7.54 (s, 4H), δ 10.43 (s, 2H) ESI-MS: Calcd. for $C_{10}H_{10}Br_2N_2O_2$: 350.01, obsd. $M+H^+$= 350.8

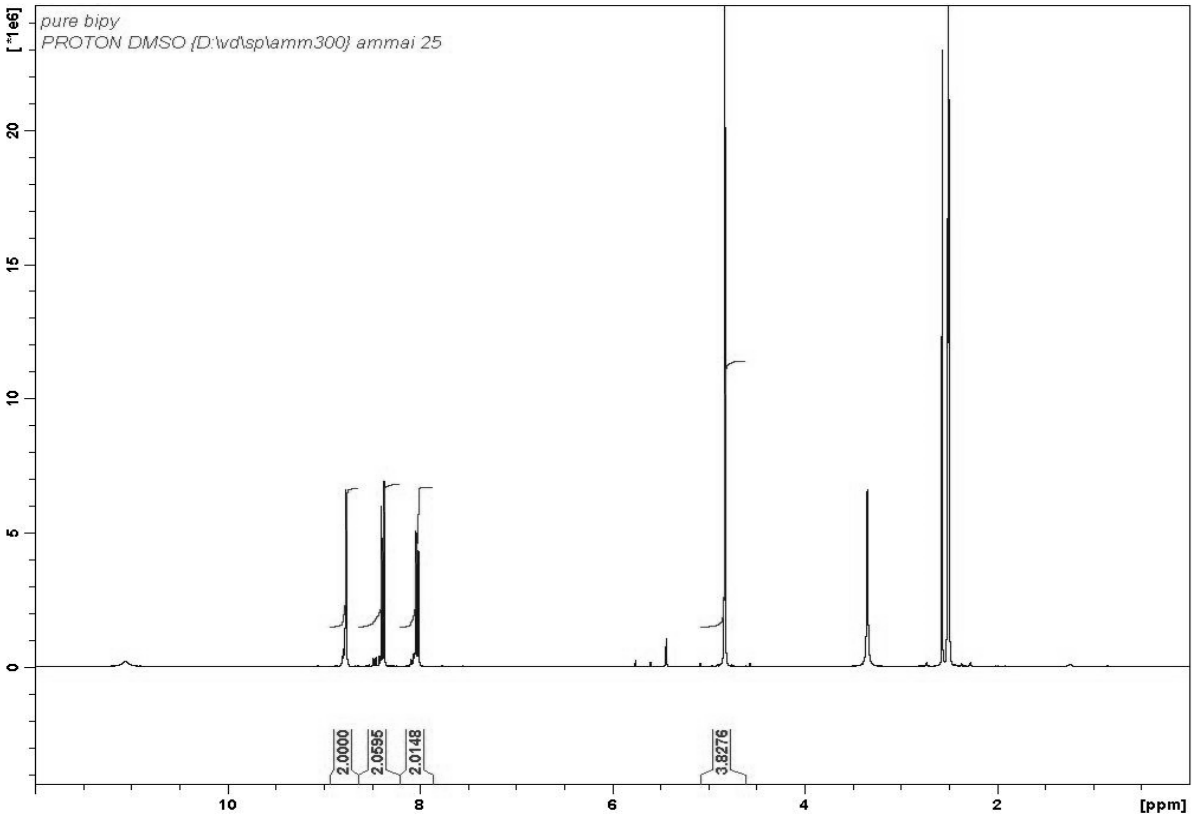


V. Synthesis of 5,5'-bis(bromomethyl)-2,2'-bipyridine



6 **c**

5, 5'-dimethyl-2,2' bipyridine **6** (2 mmol, 368 mg) was dissolved in 20 mL dry CCl_4 together with NBS (10 mmol, 1.78 g) and AIBN (20 mg) was stirred under reflux for 4 hrs at 55°C. The reaction mixture was then dried under vacuum and purified by silica gel column chromatography using EtOAc/Hexane (1:1) to obtain pure 5,5'-bis(bromomethyl)-2,2'-bipyridine **c**. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 8.85 (d, 2H), δ 8.43 (d, 2H), δ 8.10 (d, 2H), δ 4.88 (s, 4H), δ 2.63 (s, 1.5H) ESI-MS: Calcd. for $C_{12}H_{10}Br_2N_2$: 342.01, obsd. $M+H^+$= 342.9



VI. Stapled peptides

The following stapled peptides were synthesized:

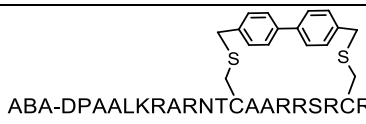
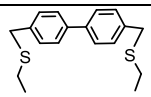
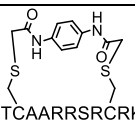
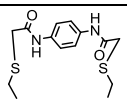
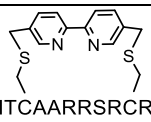
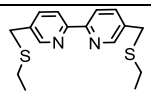
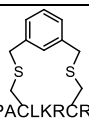
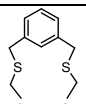
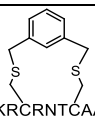
No.	Peptide	Chemical Formula	Exact Mass [M]	Observed
1a	 ABA-DPAALKRARNTCAARRSRCRKLQ-CONH$_2$	$C_{129}H_{211}N_{47}O_{31}S_2$	2978.58	M + H$^+$ = 2980.88
Ia	 Fluorescein-DPAALKRARNTCAARRSRCRKLQ-CONH$_2$	$C_{141}H_{214}N_{46}O_{35}S_2$	3175.58	M + H$^+$ = 3178.31
1b	 ABA-DPAALKRARNTCAARRSRCRKLQ-CONH$_2$	$C_{125}H_{209}N_{49}O_{33}S_2$	2988.56	M + H$^+$ = 2991.49
Ib	 Fluorescein-DPAALKRARNTCAARRSRCRKLQ-CONH$_2$	$C_{137}H_{212}N_{48}O_{37}S_2$	3185.56	3186.73
1c	 ABA-DPAALKRARNTCAARRSRCRKLQ-CONH$_2$	$C_{127}H_{209}N_{49}O_{31}S_2$	2980.57	2981.74
Ic	 Fluorescein-DPAALKRARNTCAARRSRCRKLQ-CONH$_2$	$C_{139}H_{212}N_{48}O_{35}S_2$	3177.57	3178.83
2d	 ABA-DPACLKRCRNTEAARRSRARKLQ-CONH$_2$	$C_{125}H_{209}N_{47}O_{33}S_2$	2960.56	M + H$^+$ = 2962.74
IId	 Fluorescein-DPACLKRCRNTEAARRSRARKLQ-CONH$_2$	$C_{137}H_{212}N_{46}O_{37}S_2$	3157.56	M + H$^+$ = 3161.14
3d	 ABA-DPAALKRCRNTCAARRSRARKLQ-CONH$_2$	$C_{123}H_{207}N_{47}O_{31}S_2$	2902.55	M + H$^+$ = 2904.71

Table 2: Calculated and observed masses of stapled peptides.

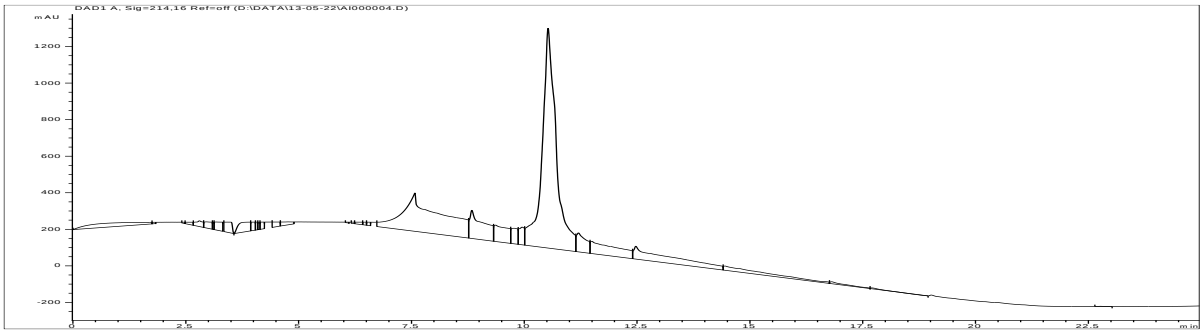


Figure 20: RP-HPLC trace of HPLC purified peptide **1a** with t_R= 10.526 min (0-100% ACN in 15 min on Luna C18(2) 100 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C).

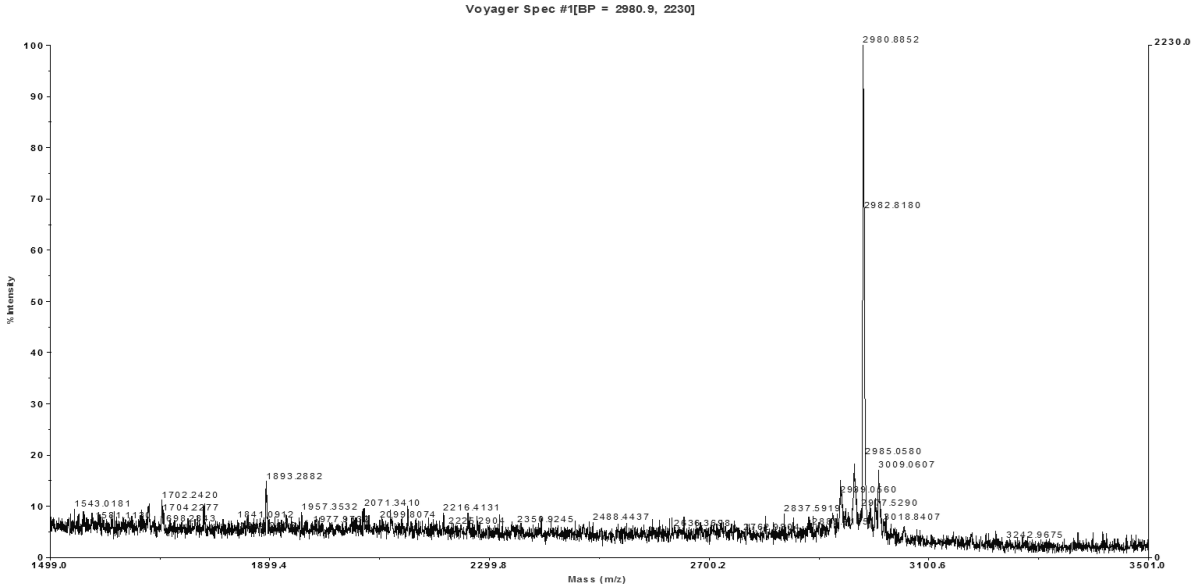


Figure 21: MALDI-MS spectrum of RP-HPLC purified peptide **1a**. E.M. Calcd. for $C_{129}H_{211}N_{47}O_{31}S_2$ = 2978.58 and mass found $M+H^+$= 2980.88

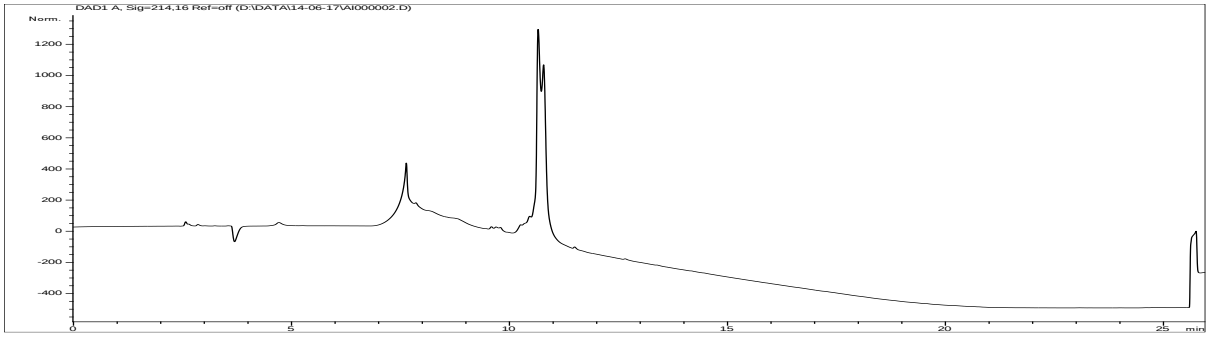


Figure 22: RP-HPLC trace of HPLC purified peptide **Ia** with t_R= 10.526 min (0-100% ACN in 15 min on Luna C18(2) 100 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

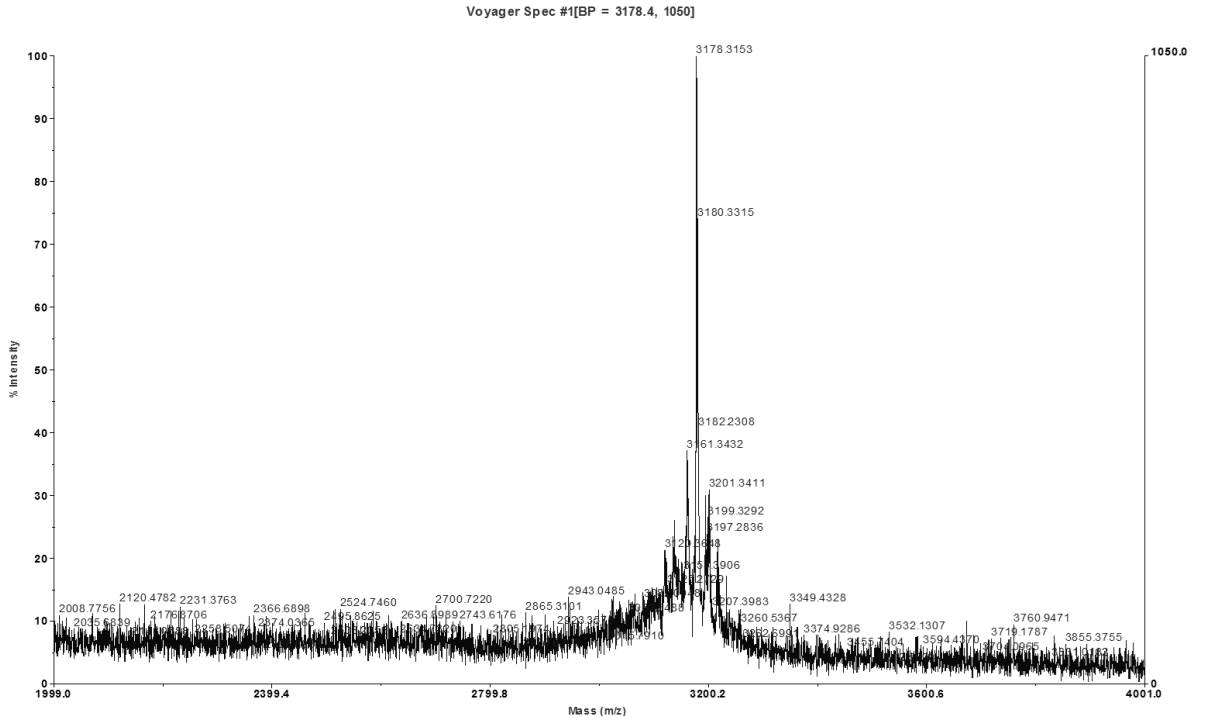


Figure 23: MALDI-MS spectrum of RP-HPLC purified peptide **Ia**. E.M. Calcd. for $C_{141}H_{214}N_{46}O_{35}S_2$ = 3175.58 and mass found $M+H^+$= 3178.31

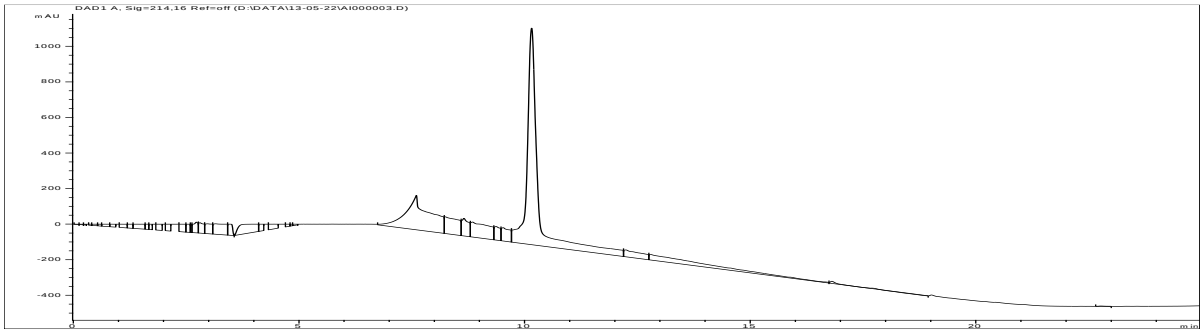


Figure 24: RP-HPLC trace of HPLC purified peptide **1b** with t_R= 10.155 min (0-100% ACN in 15 min on Luna C18(2) 100 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

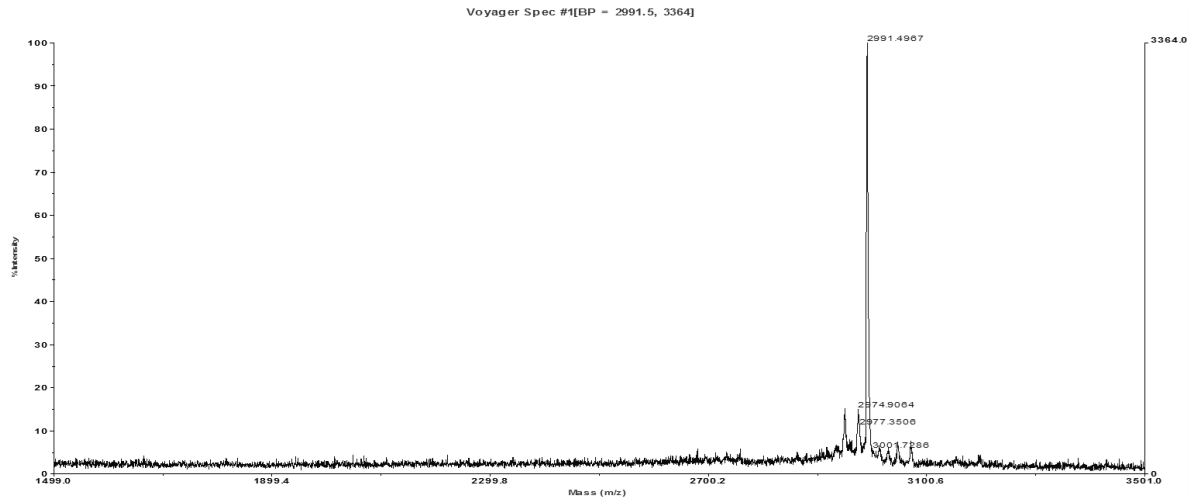


Figure 25: MALDI-MS spectrum of RP-HPLC purified peptide **1b**. E.M. Calcd. for $C_{125}H_{209}N_{49}O_{33}S_2$ = 2988.56 and mass found $M+H^+$= 2991.49

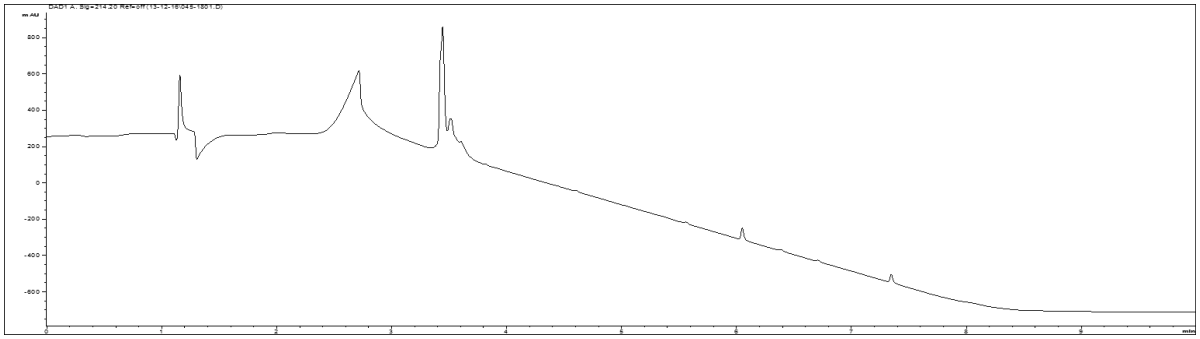


Figure 26: RP-HPLC trace of HPLC purified peptide **Ic** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

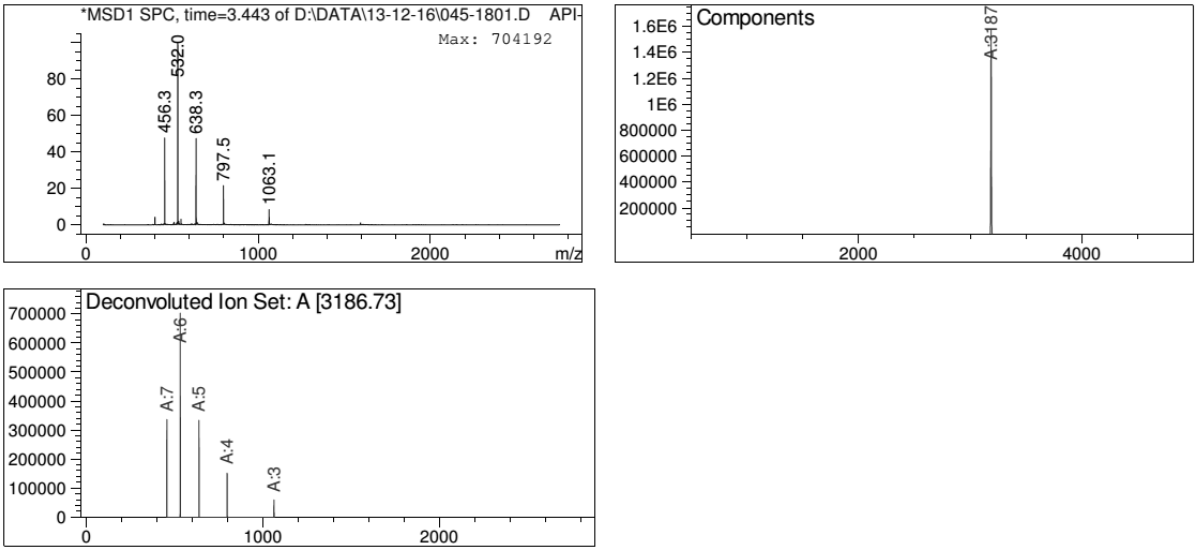


Figure 27: ESI-MS from LC-MS at t = 3.443 min. E.M calcd. for HPLC purified peptide **Ic** $C_{137}H_{212}N_{48}O_{37}S_2$ = 3185.56 and deconvoluted mass found 3186.73.

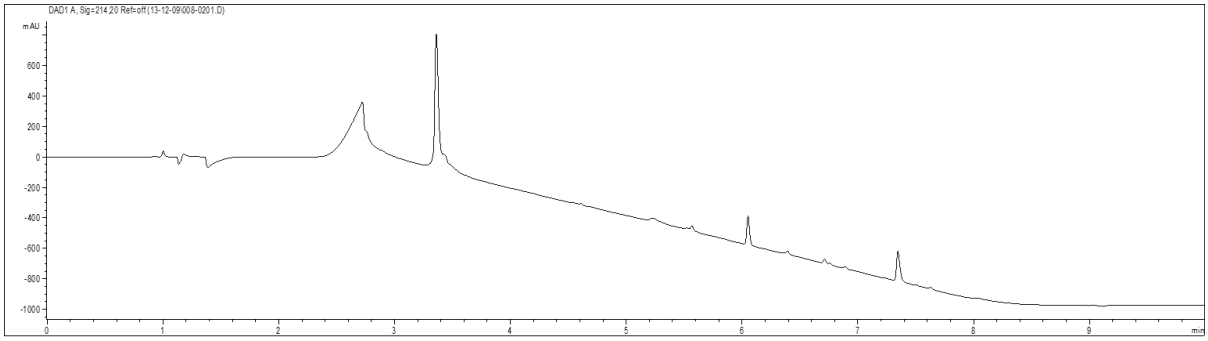


Figure 28: RP-HPLC trace of HPLC purified peptide **1c** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

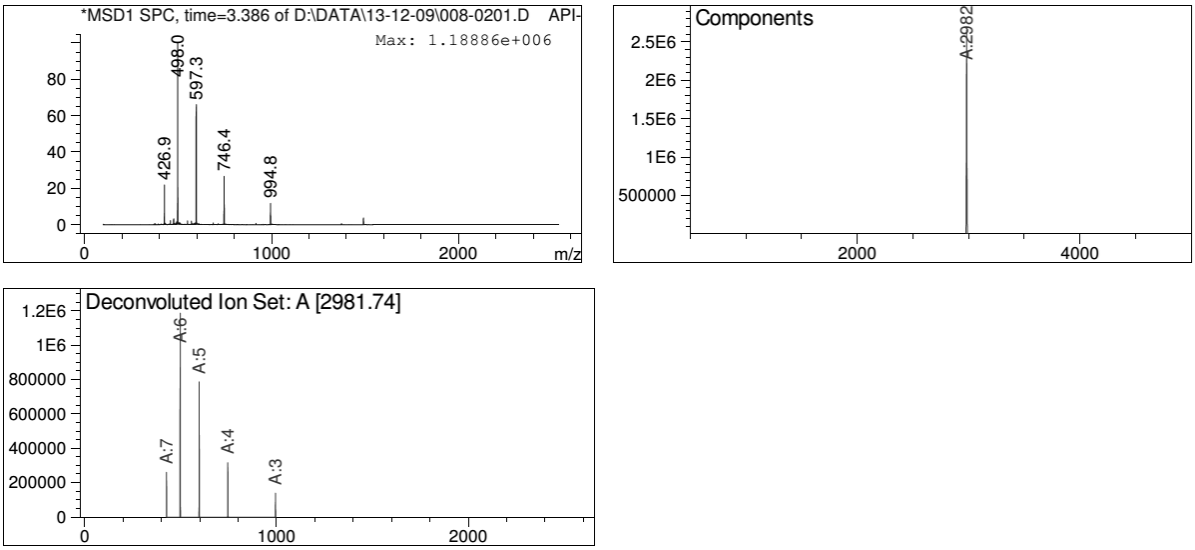


Figure 29: ESI-MS from LC-MS at t = 3.386 min. E.M calcd. for HPLC purified peptide **1c** $C_{127}H_{209}N_{49}O_{31}S_2$ = 2980.57 and deconvoluted mass found 2981.74.

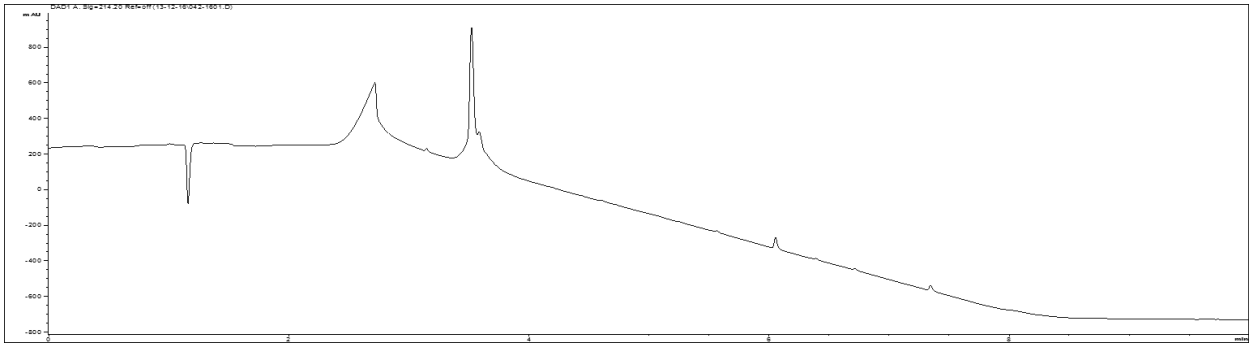


Figure 30: RP-HPLC trace of HPLC purified peptide **Ic** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

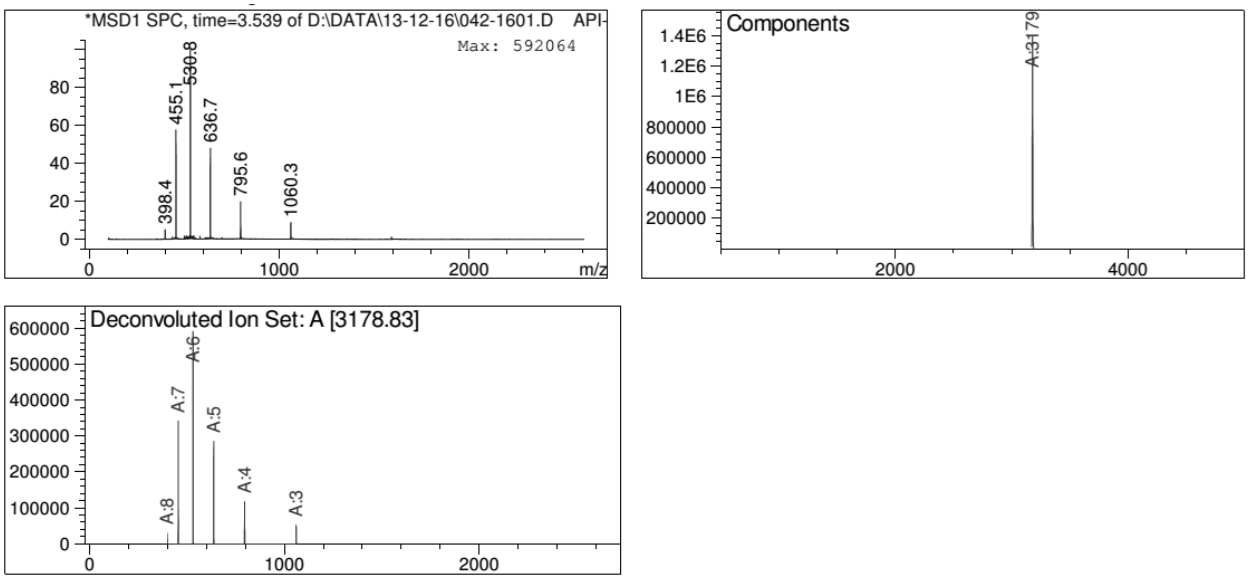


Figure 31: ESI-MS from LC-MS at t = 3.539 min. E.M calcd. for HPLC purified peptide **Ic** $C_{139}H_{212}N_{48}O_{35}S_2$ = 3177.57 and deconvoluted mass found 3178.83.

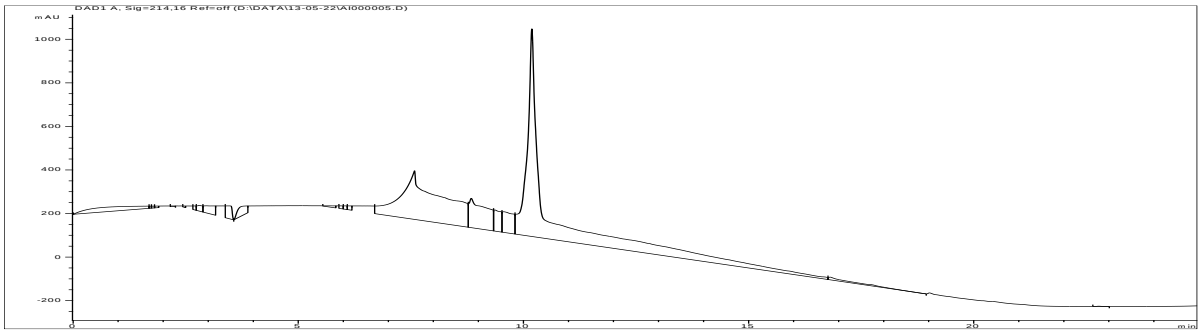


Figure 32: RP-HPLC trace of HPLC purified peptide **2d** with t_R= 10.185 min (0-100% ACN in 15 min on Luna C18(2) 100 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

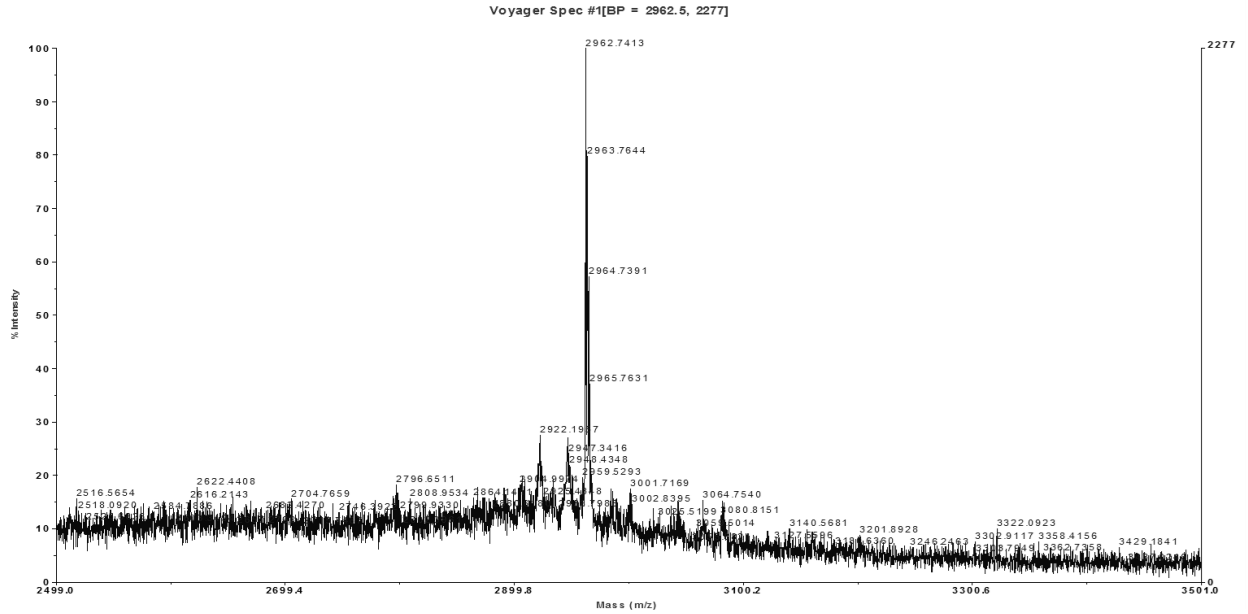


Figure 33: MALDI-MS spectrum of RP-HPLC purified peptide **2d**. E.M. Calcd. for $C_{125}H_{209}N_{47}O_{33}S_2$ = 2960.56 and mass found $M+H^+$= 2962.74

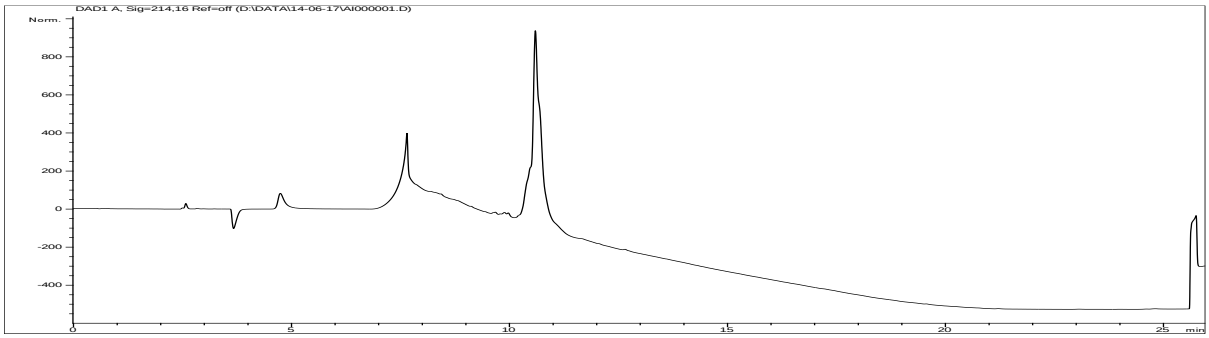


Figure 34: RP-HPLC trace of HPLC purified peptide **IId** with t_R= 10.492 min (0-100% ACN in 15 min on Luna C18(2) 100 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

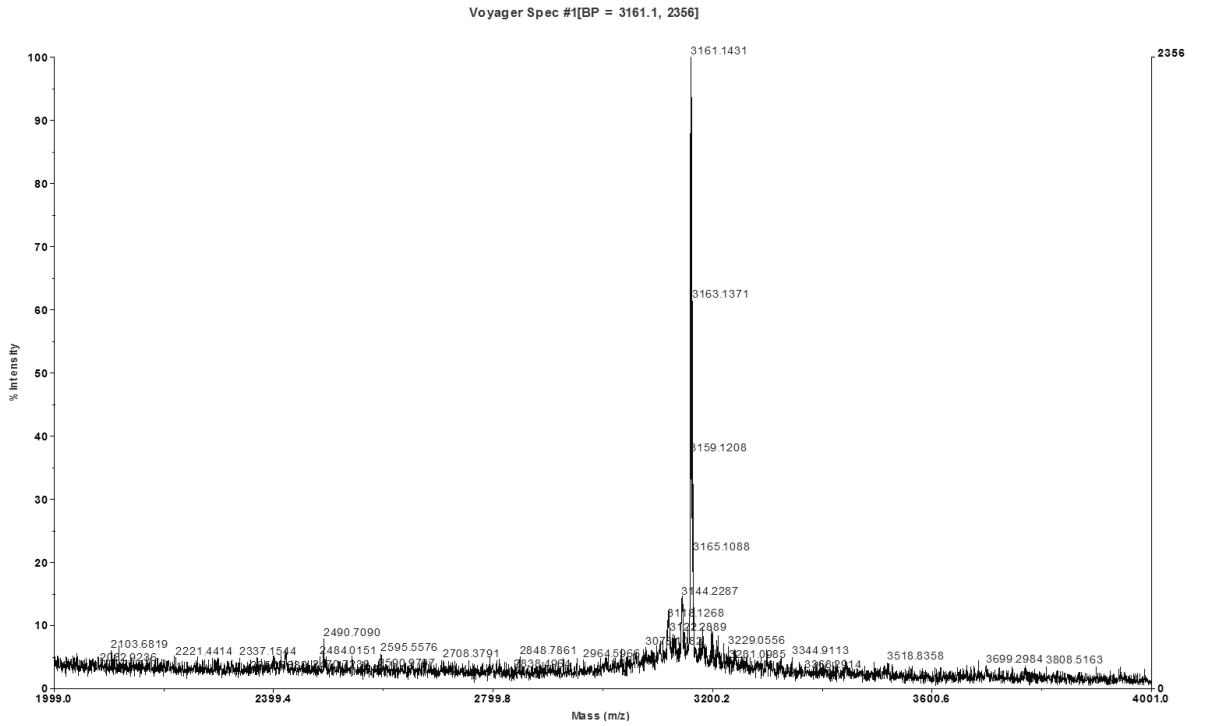


Figure 35: MALDI-MS spectrum of RP-HPLC purified peptide **IId**. E.M. Calcd. for $C_{137}H_{212}N_{46}O_{37}S_2$ = 3157.56 and mass found $M+H^+$= 3161.14

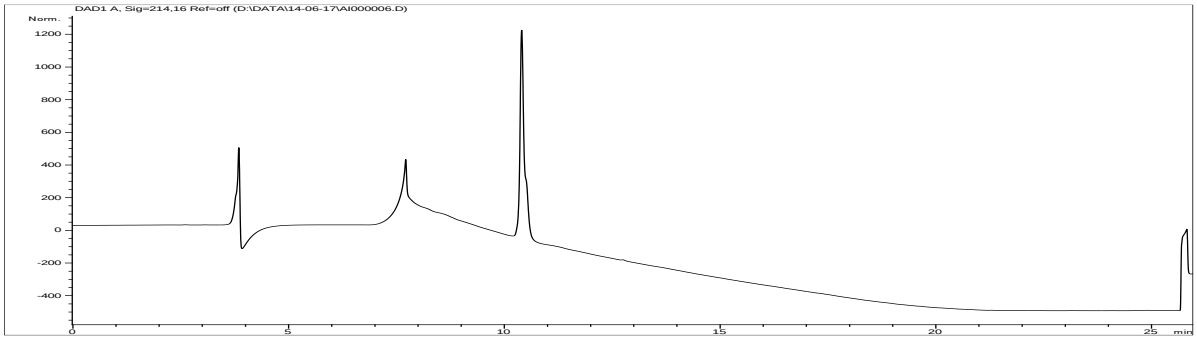


Figure 36: RP-HPLC trace of HPLC purified peptide **3d** with t_R = 10.397 min (0-100% ACN in 15 min on Luna C18(2) 100 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

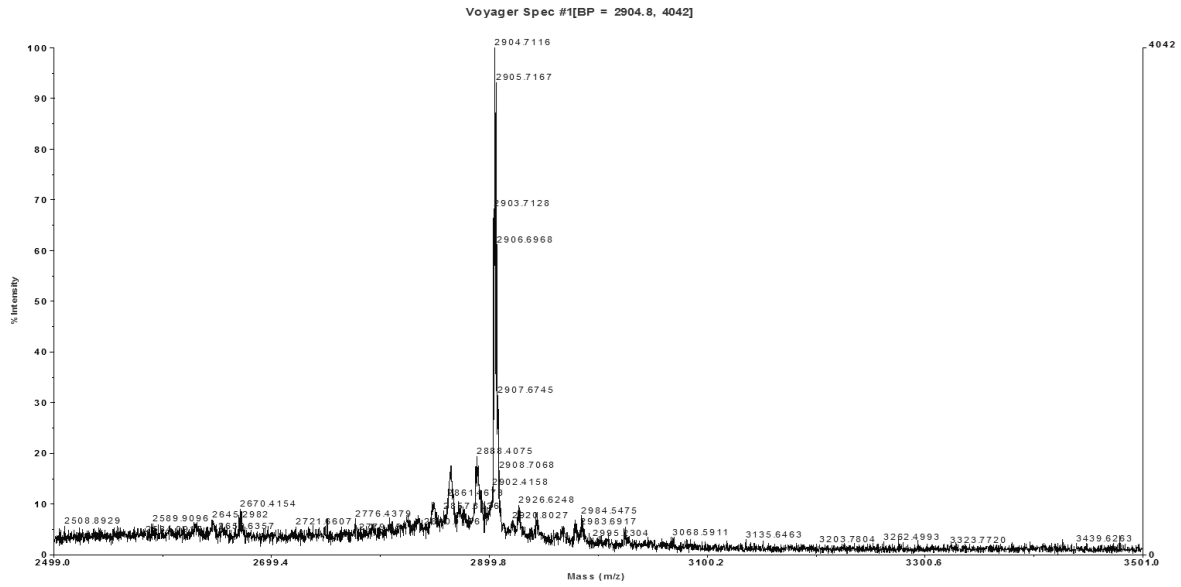


Figure 37: MALDI-MS spectrum of RP-HPLC purified peptide **3d**. E.M. Calcd. for $C_{123}H_{207}N_{47}O_{31}S_2$ = 2902.55 and mass found $M+H^+$= 2904.71

VII. Circular Dichroism studies

All HPLC purified peptide solutions were prepared at 5 μM in 50 mM phosphate buffer (pH 7.0) without TFE. CD studies were conducted at 25 °C on an Aviv 410 CD spectrophotometer equipped with a Peltier temperature control unit. Results (fig. 38) showed that the uncross-linked GCN4 sequence **4** shows almost no helical character whereas cross-linking with Biphenyl as in the case of **1a** shows most helical character. Surprisingly, cross-linking with Bipyridine does not improve the helicity significantly as can be seen from the curve of **1c**. Among the i, i+4 stapled peptides **2d** is more helical than **3d** which would explain its better DNA binding. In general the i, i+7 stapled peptides **1a** & **1b** show more helical character than the i, i+4 ones **2d** & **3d** which is expected given the two turn stabilization in case of the former.

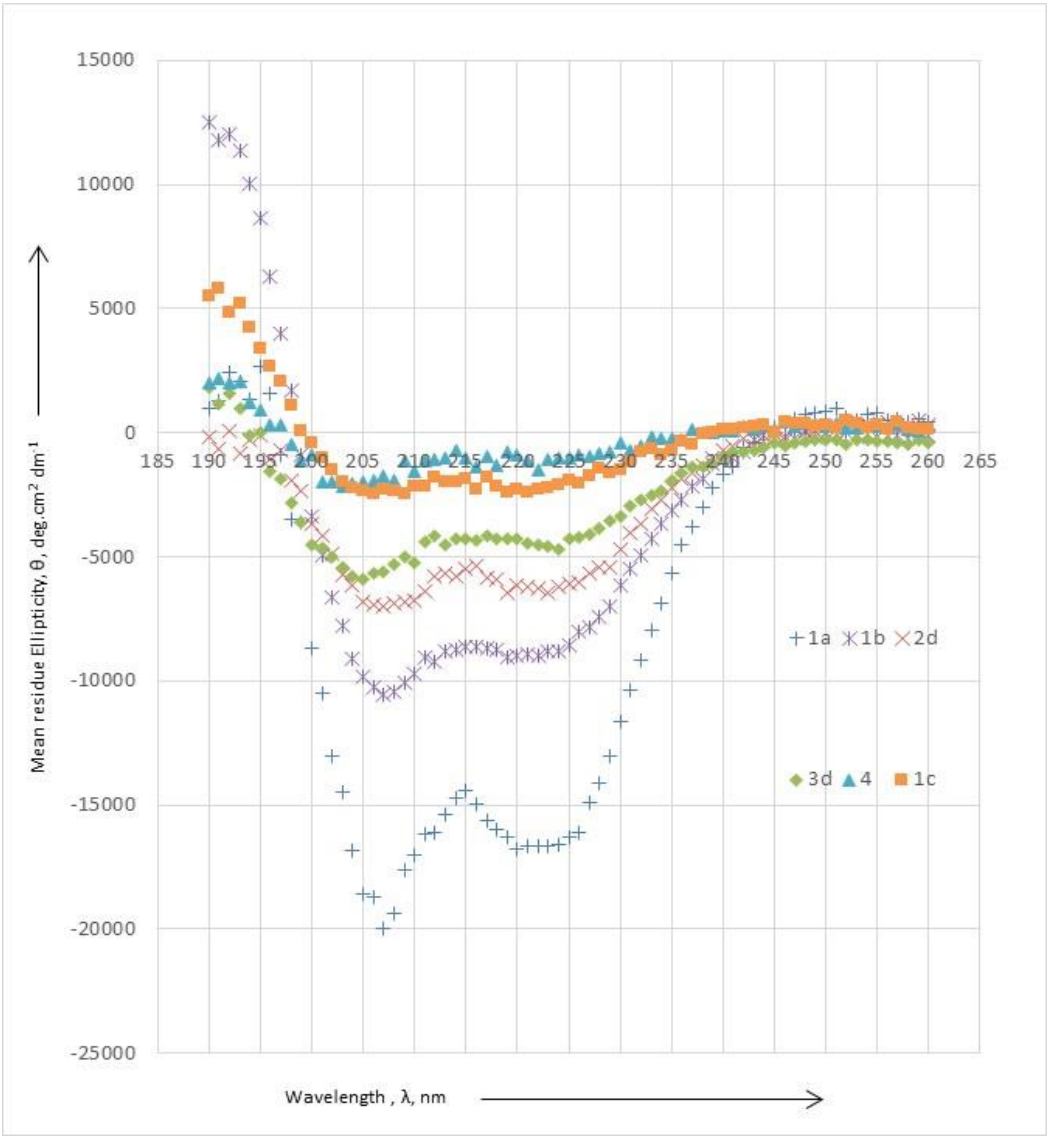


Figure 38: Plotted data obtained from Circular Dichroism studies showing Mean residue Ellipticity, θ, in deg.cm^2 dm^{-1} on the Y axis and Wavelength λ, in nm on the X axis for peptides **4**, **1a**, **1b**, **1c** and **2d**.

VIII. Electrophoretic Mobility Shift Assay (EMSA)

The following stock solutions were prepared (fresh each time, except for DNA and peptide):
DNA: 1.67 µM prepared from CRE (5' – CGG ATG ACG TCA TTT TTT TTC – 3') & CRE complement (5' – GAA AAA AAA TGA CGT CAT CCG – 3') and Random (5' – GCG CGA GAA GGA AAG AAA GCC GG – 3') & complement (5' – CCG GCT TTC TTT CCT TCT CGC GC – 3') DNA solutions (commercially purchased) by diluting with 20 µL 0.5 M Tris, pH = 8, 40 µL 2.5 M NaCl, 40 µL 0.025 M EDTA and then adding milliQ water such that the total volume is 1 mL. The DNA was annealed by heating in a Thermomixer from room temperature to 95°C and maintaining the temperature at 95°C in total time of 24 min. The machine was then turned off and the sample was allowed to cool down slowly.
Loading buffer: 20 µL Tris 1 M, pH = 7.6, 20 µL KCl 0.2 M, 20 µL $MgCl_2$ 0.1 M, 40 µL EDTA 0.025 M.
Sucrose: 30% sucrose in mQ (300 mg/mL)
Peptides: 10 µL stock solutions (10x) were prepared in MiliQ water (0, 1.67, 5.01, 6.68, 7.51, 8.35, 10.02, 11.69, 13.36, 16.7 µM)
Loading mixture: The loading mixture comprised of: 10 µL mQ, 4 µL sucrose, 2 µL loading buffer, 2 µL DNA, 2 µL peptide. The loading mixture was prepared only 1-2 hr prior to running of gels and kept on ice as soon as ready.
Preparation of Gels (for 2 Gels): In a clean falcon tube the following were added (in given order): 15.595 mL mQ, 0.4 mL TBE, 4.005 mL of 40% acrylamide solution, 200 µL APS (10% w/w in mQ). The solution was mixed by sonication to remove any air bubbles and cooled to 0°C (1 h under ice). 20 µL of TEMED was then added to the mixture and was again mixed properly before pouring it gently along parallel glass plates. The glass plates were tapped gently to ensure removal of all air bubbles and the markers were squeezed between the plates to ensure uniform width of each well. Sufficient time was given for polymerization (~1 h).
Gel Electrophoresis:A pre-run of the gels was performed prior to loading them. Care was taken to see that the gels were properly immersed in 0.2x TBE buffer (non-denaturing gel, without urea) and the loading wells were free from any air bubbles. Instrument settings: 150 V, 100 mA, 19 W for 30 mins at 4°C. The wells were washed after the pre-run. 5 µL of the loading mixture was then loaded onto the wells. Instrument settings: 150 V, 100 mA, 19 W for 45 mins at 4°C.
Staining of gels: After the run, the gels were removed from the glass and were stained using 100 mL of 0.2x TBE buffer + 10 µL Sybr Gold (commercially purchased stock solution 10,000X in DMSO). The gels were then washed twice with mQ and gently placed under a UV lamp (dark room) to observe the gel pattern.
Full gels and gels using random dsDNA are shown in figure 39-41. Figure 42 shows the gel of an additional gel to confirm that the black dot of peptide 3d at 2.5 µM (fig. 40, right) is an artifact. To investigate the double binding pattern of peptides 2d and 3d, additional tests with similar DNA containing only one binding site is shown in figure 43. Results show that scrambling the sequence is detrimental for the DNA binding capacity as binding is not observed with this mutated sequence 5'-CGG CGA TAG TCA TTT TTT TTC- 3'.

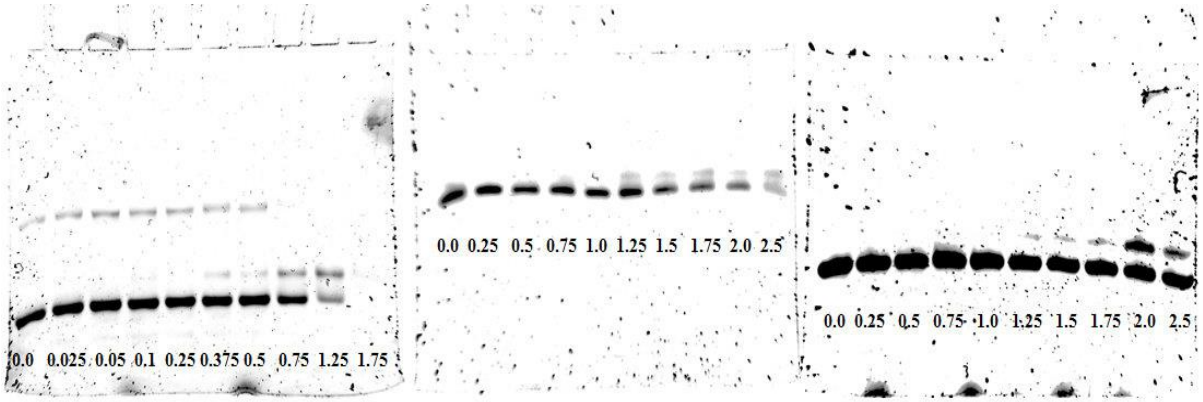


Figure 39: Full gels obtained from EMSA for peptides **1a**, **1b** and **1c** (from left to right) respectively using CRE DNA. DNA concentration = 167 nM for all gels. Peptide concentrations in µM (from left to right) are indicated below each gel.

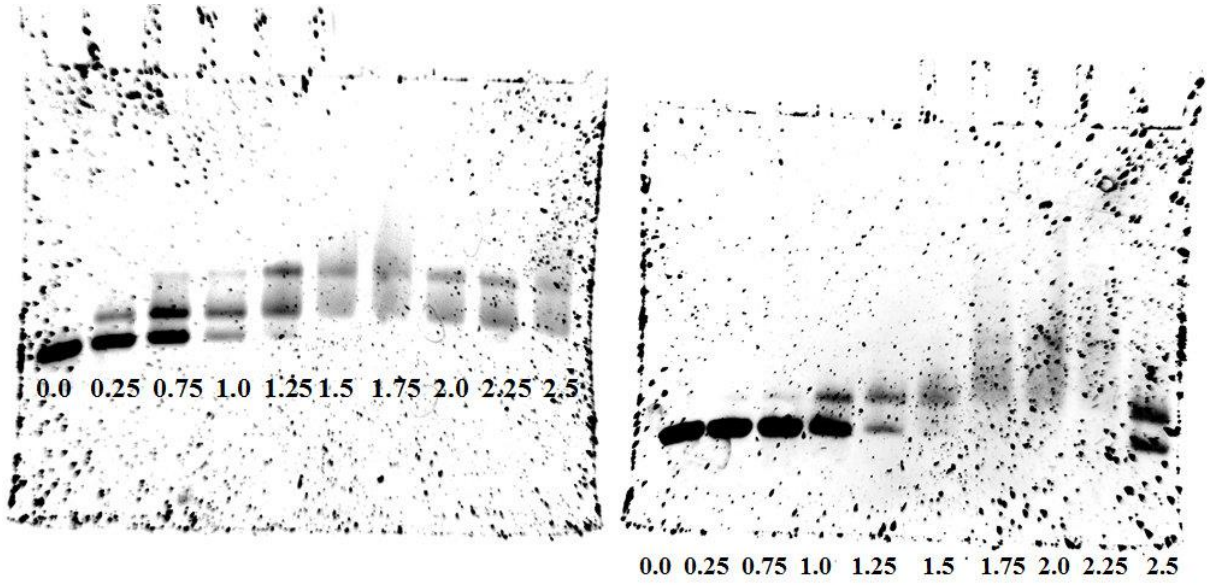


Figure 40: Full gels obtained from EMSA for peptides **2d** and **3d** (from left to right) respectively using CRE DNA. DNA concentration = 167 nM for all gels. Peptide concentrations in µM (from left to right) are indicated below each gel.

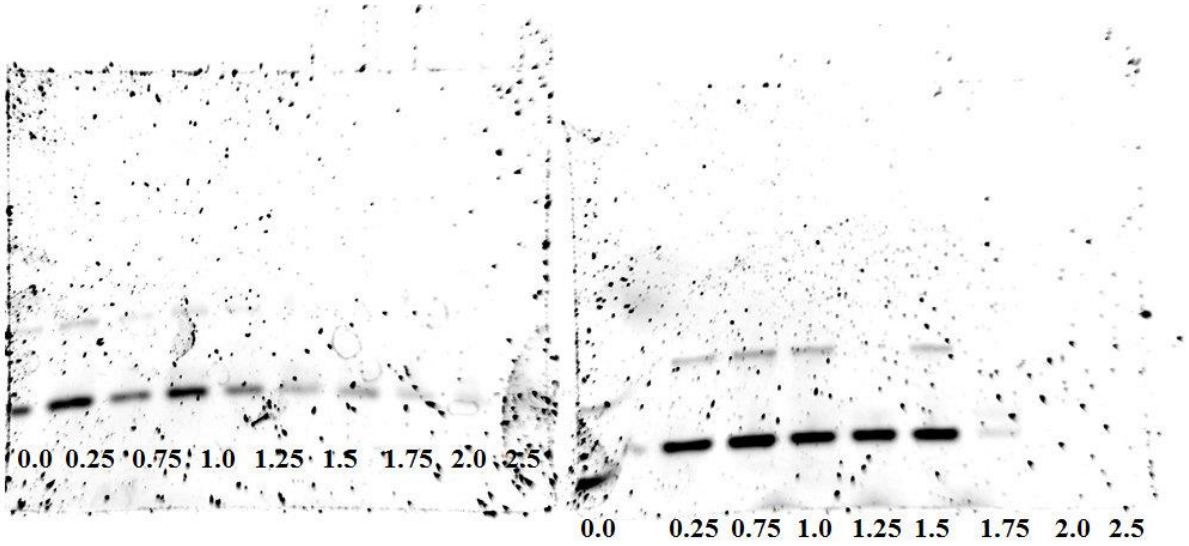


Figure 41: Full gels obtained from EMSA for peptides **4** using CRE DNA (left) and **2d** using random dsDNA (right). DNA concentration = 167 nM for both gels. Peptide concentrations in µM (from left to right) are indicated below each gel.

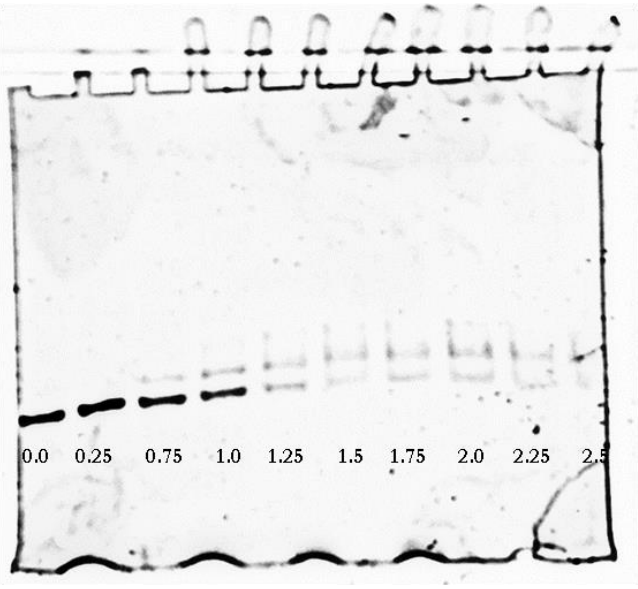


Figure 42: Full gel obtained from EMSA for peptides **3d** using CRE DNA (concentration = 167 nM). Peptide concentrations in µM (from left to right) are indicated below each gel. Gelred was used as staining agent instead of the previous used Sybr Gold. Therefore intensities are lower than in other gels.

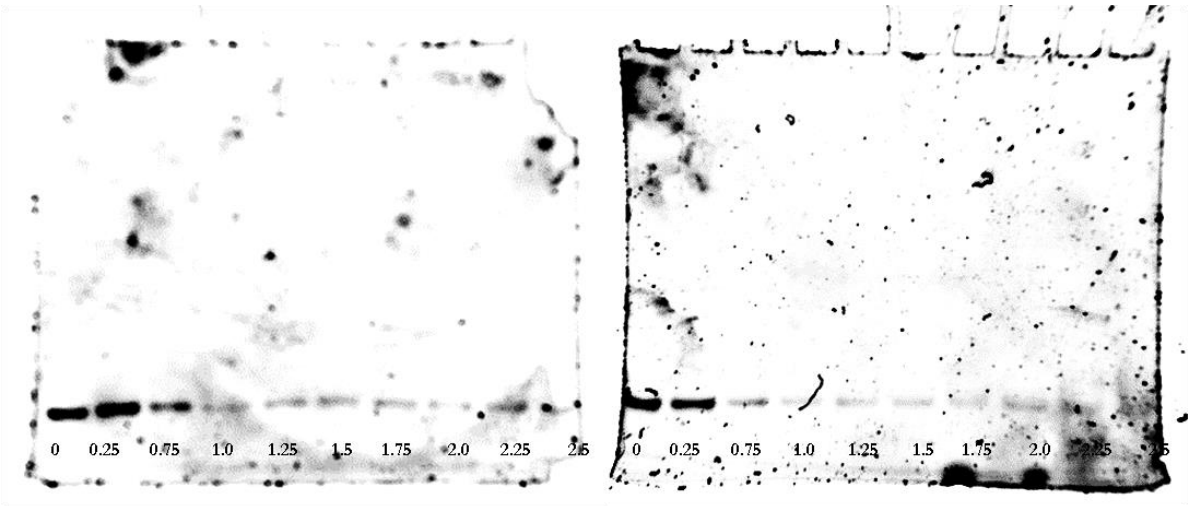


Figure 43: Full gels obtained from EMSA for peptides **2d** (left) and **3d** (right) using scrambled CRE. DNA concentration = 167 nM for both gels. Peptide concentrations in µM (from left to right) are indicated below each gel.

IX. DLS-measurement

In order to investigate the aggregation of DNA and peptide, leading to precipitation and a subsequent decrease in intensity of the band, the dynamic light scattering (DLS) is measured of peptides **1a** and **1c** in the same mixture as the loading mixture in the gel electrophoresis. DLS measurement was performed in fivefold on a Zetasizer Nano series, Malvern in a 100 µL cuvette. The results are shown in figure 45. On the left, the volume distribution displays the aggregation of peptides **1a** with DNA at higher concentration as a 'particle' with larger size is present. Also when looking to the count rate on the right of fig. 45, a higher amount of particles is present at high concentration which can be explained by aggregation. These results are conform the visual observation of the gels where peptide **1a** shows a decreased intensity of the band at higher concentration. Also a decreased intensity of the band of **1a** at higher concentration compared to peptide **1c** at 2 µM is observed.

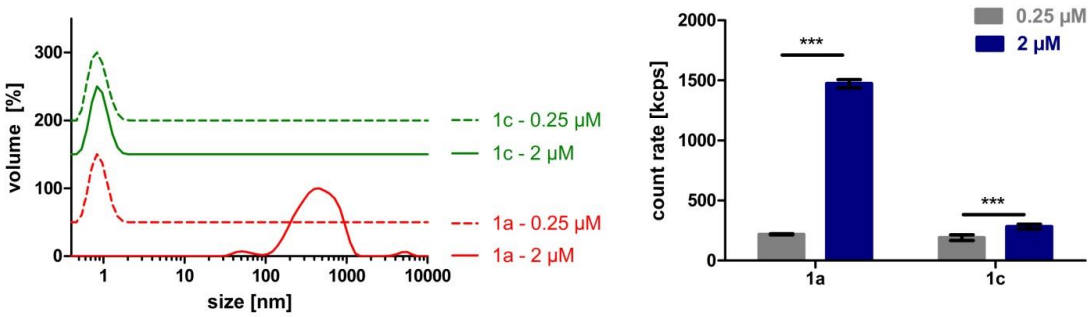


Figure 44: DLS measurement of peptides **1a** and **1c** at a peptide concentration of 0.25 µM and 2 µM and a DNA concentration of 167nM. On the left, the volume distribution is shown in function of the size of the measured particles. On the right, the count rate of each samples in shown.

X. Confocal Microscopy

RAW264.7 cells cultivated in DMEM (10^5/well, 300 µL) were plated in a confocal plates and incubated overnight at 37 °C and 5 % CO_2. Peptides were added in an overall concentration of 0.25 µM and incubated for 2 h at the same conditions or on ice. Cells were washed and fixated with 2 % of Paraformaldehyde for 30 min at 37°C. Cells were stained by 30 min incubation with a 0.2 % solution of CTB-AF647 and 0.2 % of Hoechst in PBS with 1 % BSA. This was done to visualize the cell nucleus (blue) and cell membrane (red) respectively. Cells were resuspended in PBS and measured with confocal microscope (Leica SP5 equiped with a 63x (1.4 NA) oil immersion objective) at 3 different wavelengths (405 nm, 488 nm and 643 nm). The peptides used were labeled with fluorescein which emits green light. To understand the uptake mechanism better, three samples are made in duplicates and incubated at 4 °C to inhibit active transport. Due to the difference in emission wavelength, the three dyes can be detected separately. Processing the data with Image J gave overlay images whereby the nucleus, cell membrane and fluorescein are shown in blue, red and green respectively.

XI. Flow cytometry

RAW264.7 cells cultivated in DMEM (10^5/well, 1 ml) were plated in a 24-well plates and incubated overnight at 37 °C and 5 % CO_2. Peptides were added in an overall concentration of 0.25 µM and 1 µM and incubated for 2 h at the same conditions or on ice. Cells were washed with PBS and detached with Na_4EDTA. Cells were re-suspended in PBS and added to the BD Accuri flow cytometer. Experiments were carried out in duplicate. The amount of peptide associated to the cell was measured by flow cytometry. Blank cells were measured and the signal received corresponded to auto-fluorescence. A threshold was chosen to separate cells emitting light due to auto-fluorescence from cells containing fluorescently labeled peptides. The area under the histograms beyond this threshold correspond to the number of cells which actually emit sufficient intensity of fluorescence to make sure fluorescently labeled peptides are present. This number is further referred to as the percent of peptide positive cells (fig. 45, right). The percent of peptide positive cells for two peptides may be 100 % yet one may better due to its higher mean fluorescence value. If a peptide is not taken up by all cells, the percentage of peptide positive cells is lower than 100 % proportional to the colored area under the curve. It is therefore important to both compare the mean fluorescence and the amount of peptide positive cells. The raw data, represented as histograms, can be found in fig. 46-50.

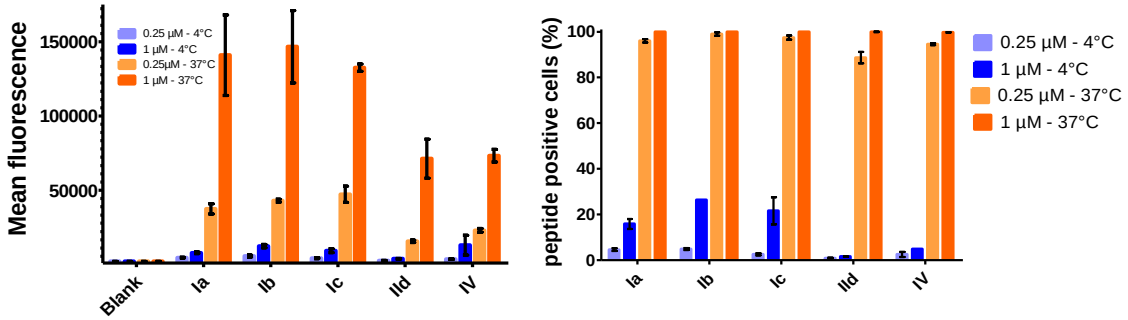
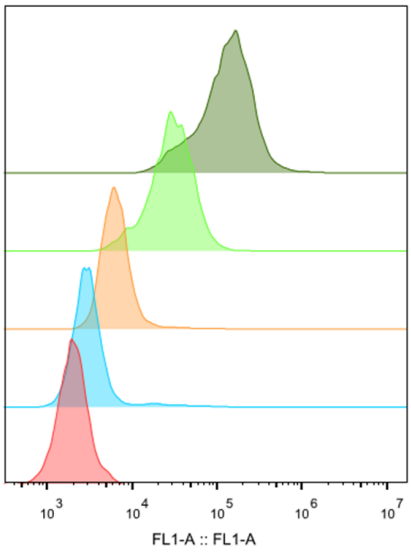
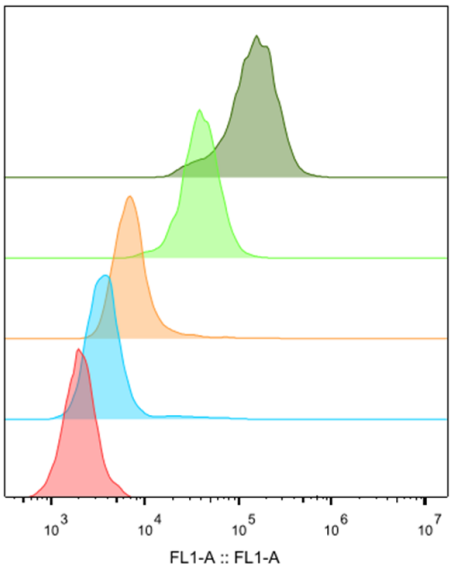


Figure 45: Mean fluorescence of fluorescently labelled peptides **Ia**, **Ib**, **Ic**, **IId** and **IV** (left) and peptide positive cells (right) using the following concentrations and incubation conditions: 0.25 µM at 4°C, 1 µM at 4°C, 0.25 µM at 37°C and 1 µM at 37°C.



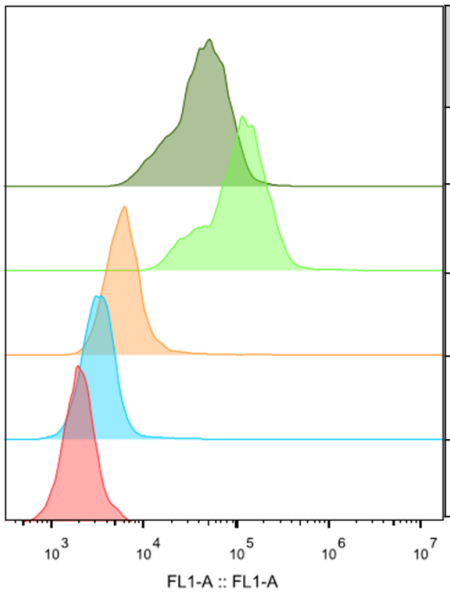
Conditions Peptide Ia	Count	Mean (FL1-A)	FL1-A+ (Freq. of parent)
37 °C 1 µM	5155	160044	100
37 °C 0.25 µM	5464	35170	95.5
4 °C 1 µM	4111	8361	17.4
4 °C 0.25 µM	3408	4786	5.05
Blank	23248	2207	0

Figure 46: Histograms responding to the uptake of peptide **Ia** under different conditions and compared with the blank measurement. 'Count' represents the cell count during each measurement. FL1-A+ represents the peptide positive cells.



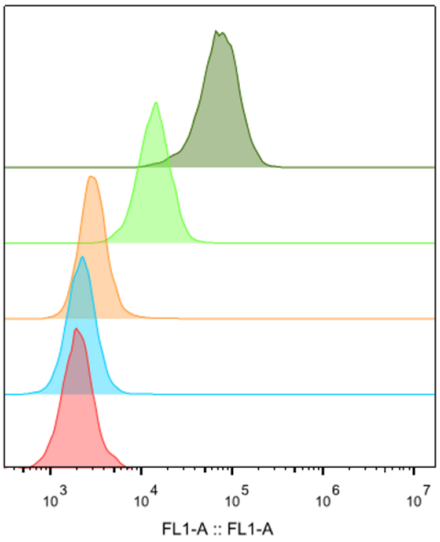
Conditions Peptide Ib	Count	Mean (FL1-A)	FL1-A+ (Freq. of parent)
37 °C 1 µM	6219	163849	100.0
37 °C 0.25 µM	7245	43630	99.5
4 °C 1 µM	4723	13203	26.4
4 °C 0.25 µM	5337	6165	5.17
Blank	23248	2207	0

Figure 47: Histograms responding to the uptake of peptide **Ib** under different conditions and compared with the blank measurement. 'Count' represents the cell count during each measurement. FL1-A+ represents the peptide positive cells.



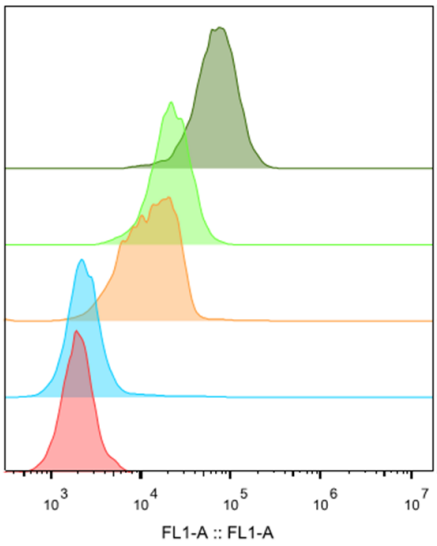
Conditions Peptide Ic	Count	Mean (FL1-A)	FL1-A+ (Freq. of parent)
37 °C 0.25 µM	5674	51197	98.1
37 °C 1 µM	4882	134245	100
4 °C 1 µM	3596	8395	17.4
4 °C 0.25 µM	3667	4140	2.26
Blank	23248	2207	0

Figure 48: Histograms responding to the uptake of peptide Ic under different conditions and compared with the blank measurement. 'Count' represents the cell count during each measurement. FL1-A+ represents the peptide positive cells.



Conditions Peptide IId	Count	Mean (FL1-A)	FL1-A+ (Freq. of parent)
37 °C 1 µM	10280	80470	100.0
37 °C 0.25 µM	8989	15141	86.9
4 °C 1 µM	6245	3738	1.55
4 °C 0.25 µM	6582	2604	0.77
Blank	23248	2207	0

Figure 49: Histograms responding to the uptake of peptide **IId** under different conditions and compared with the blank measurement. 'Count' represents the cell count during each measurement. FL1-A+ represents the peptide positive cells.



Conditions Peptide IV	Count	Mean (FL1-A)	FL1-A+ (Freq. of parent)
37 °C 1 µM	8194	76430	99.7
37 °C 0.25 µM	8793	23989	94.9
4 °C 1 µM	5124	17755	69.0
4 °C 0.25 µM	5437	3740	3.27
Blank	23248	2207	0

Figure 50: Histograms responding to the uptake of peptide **Ia** under different conditions and compared with the blank measurement. 'Count' represents the cell count during each measurement. FL1-A+ represents the peptide positive cells.

XII. MTT Assay

RAW264.7 cells cultivated in DMEM (10^4/well, 200 µL) were plated in a 96-well plates and incubated overnight at 37 °C and 5 % CO_2. The peptides and control compounds were added in an overall concentration of 0.25 µM and 1 µM and incubated for 24h under the same conditions. The MTT solution was then added to the aspirated wells and incubated for 3 h. After removal of the cell medium, purple formazan crystals were dissolved in DMSO. Then, UV-measurement at 570 nm was performed with a plate reader to check cell viability quantitatively. An observation with the naked eye already gave a good idea about the toxicity of the compounds due to the disappearance of color in the well. Comparison of the absorbance of the formazan solution of the sample to the absorbance of a positive (incubation with ultrapure water) and a negative control (incubation with DMSO) gave quantitative results of cell viability as in the equation:

$$Cell\ viability\ (\%) = \frac{A - A_{pos}}{A_{neg} - A_{pos}} * 100$$

XIII. Peptide stability

Peptides dissolved in a 50 mM NH_4HCO_3 buffer at a concentration of 0.5 mg/mL are incubated at 37°C with a trypsin solution (trypsin/ peptide: 1/1000 wt%) in AcOH 50 mM buffer. The pH was optimized to 7-9. Samples were taken after 0 min, 30 min, 1h, 2 h and 24 h and injected on RP-HPLC (Jupiter C4 300A, 0-100% CH_3CN in 15 min). Peaks were collected and analysed using MALDI-TOF. Peaks were integrated. The peak area at 0 min was used as control. The degradation in function of time shows that the stapled peptides are more stable than the control peptide **IV** (fig. 51). Figures 52-55 displays the degradation of the peptide in function of time.

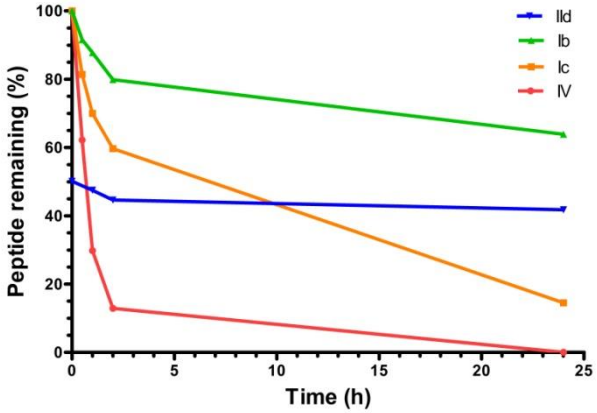
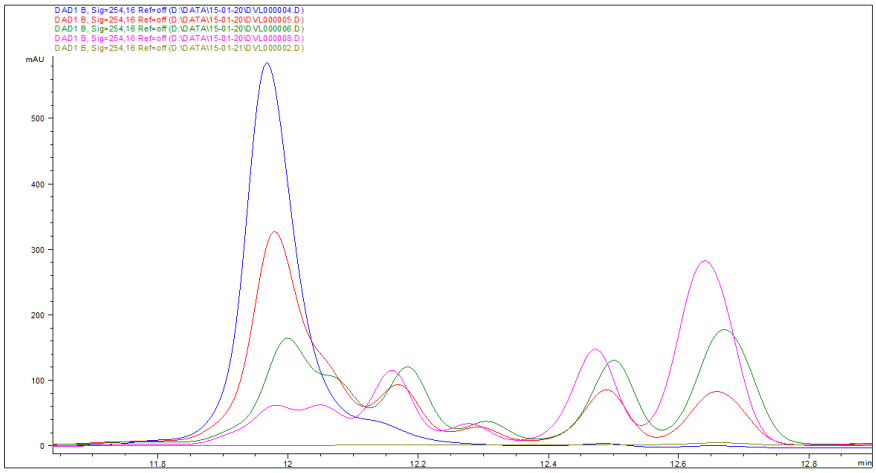
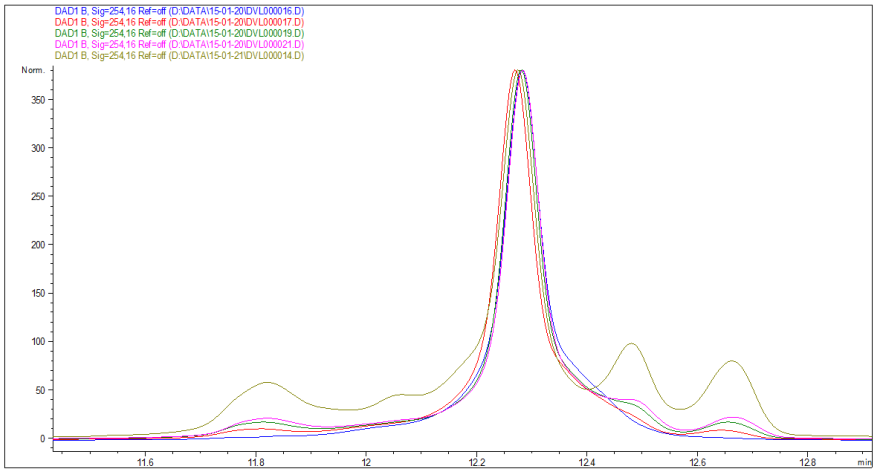


Figure 51: Chromatograms of RP-HPLC of samples of peptide **IV** taken after 0 min (blue), 30 min (red), 1 h (green), 2 h (pink),



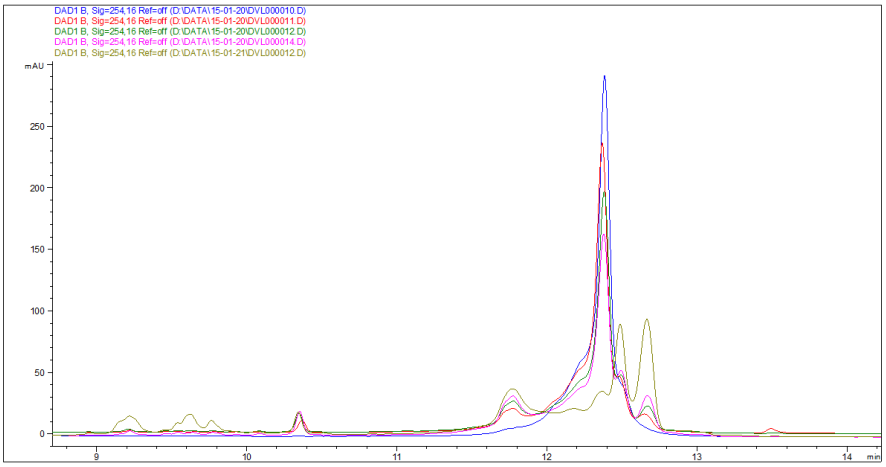
Peak (min)	MW (g/mol)	Peptide fragment
11.95	2995	Fluorescein-DPAALKRARNTEAARRSRARKLQ-NH_2
12.16	2154	Fluorescein-DPAALKRARNTEAARR-OH
12.29	1998	Fluorescein-DPAALKRARNTEAAR-OH
12.48	1355	Fluorescein-DPAALKRAR-OH
12.67	1129	Fluorescein-DPAALKR-OH

Figure 52: Chromatograms of RP-HPLC of samples of peptide **IV** taken after 0 min (blue), 30 min (red), 1 h (green), 2 h (pink).



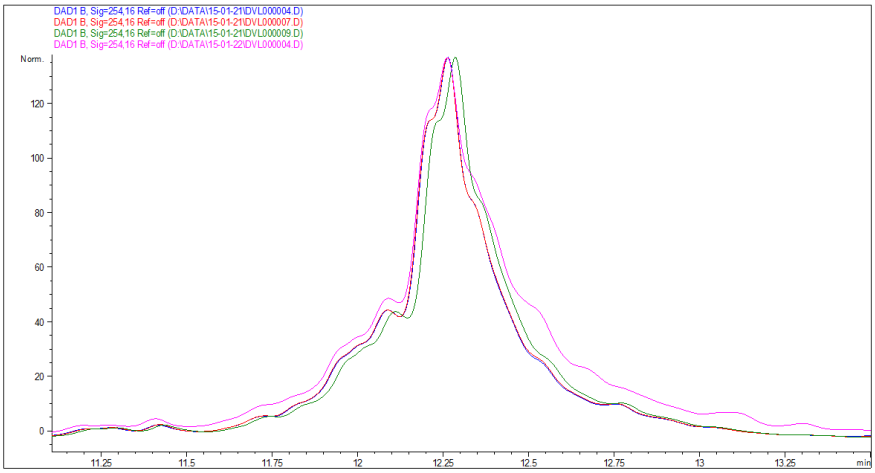
Peak (min)	MW (g/mol)	Peptide fragment
12.28	3189	Fluorescein-DPAALKRARNTC*AARRSRC*RKLQ-NH_2
12.47	1355	Fluorescein-DPAALKRAR-OH
12.67	1129	Fluorescein-DPAALKR-OH

Figure 53: Chromatograms of RP-HPLC of samples of peptide **Ib** taken after 0 min (blue), 30 min (red), 1 h (green), 2 h (pink).



Peak (min)	MW (g/mol)	Peptide fragment
12.38	3181	Fluorescein-DPAALKRARNTC*AARRSRC*RKLQ-NH_2
12.49	1357	Fluorescein-DPAALKRAR-OH
12.67	1129	Fluorescein-DPAALKR-OH

Figure 54: Chromatograms of RP-HPLC of samples of peptide **Ic** taken after 0 min (blue), 30 min (red), 1 h (green), 2 h (pink).



Peak (min)	MW (g/mol)	Peptide fragment
12.10	3162	Fluorescein-DPAC*LKRC*RNTEAARRSRARKLQ-NH_2
12.25	3162	Fluorescein-DPAC*LKRC*RNTEAARRSRARKLQ-NH_2
12.55	2566	Fluorescein-DPAC*LKRC*RNTEAARRSR-OH
12.75	2166	Fluorescein-DPAC*LKRC*NTEAAR-OH

Figure 55: Chromatograms of RP-HPLC of samples of peptide **IId** taken after 0 min (blue), 30 min (red), 1 h (green), 2 h (pink).

EXPERIMENTAL SECTION FOR CHAPTER 6

I. Materials

Amino acids and oxyma were purchased from Merck Millipore. Diisoproplycarbodiimide, Triisopropylsilane, Heptafluorobutyric acid and Trifluoroacetic acid (TFA) were purchased from Fluorochem, UK. L-amino acids were used throughout the syntheses. The protecting groups for the amino acids are ^{t}Bu for Glu, Boc for Pro, Tyr, Lys, Pbf for Arg and Trt for Gln. Diisopropylethylamine, supplied as extra dry, redistilled, 99.5 % pure, piperidine, NH_4HCO_3 and fully deprotected amino acids Lys and Arg were purchased from Sigma Aldrich. Dimmethylformamide (DMF) peptide synthesis grade was purchased from Rathburn chemicals. Diethyl ether, i-PrOH, MeOH (HPLC grade), 35% (S.G. 0.88) NH_4OH solution and Acetonitrile (HPLC grade) were purchased from Fisher Scientific. Water with the Milli-Q grade standard was obtained in-house from an ELGA Purelab Flex system. Rink-Amide ChemMatrix resin (100-200 μm, manufacturer's loading: 0.49 mmol/g) was obtained from Biotage. All chemicals were used without further purification.

II. Isolation of Moenomycin A from Flavomycin

Moenomycin was isolated according to the procedure described in literature[I].

III. General Procedure for peptide syntheses:

Peptide syntheses was performed using standard Fmoc Solid Phase Peptide Synthesis (SPPS) protocols on Rink Amide Chemmatrix Resin, loading = 0.49 mmol/g on a 0.1 mmol scale using a Biotage Initiator + Alstra fully automated microwave peptide synthesizer. All amino acid couplings were performed using 5 eq. Amino Acid with 5 eq. DIC/Oxyma in DMF as a coupling cocktail by irradiating at 70°C for 5 min. Fmoc deprotection was performed using 20% piperidine in DMF by irradiating at 70°C for 3 min, followed by shaking at r.t. for 10 min. 4 x 45s washes were performed after each coupling cycle and 3 x 30s washes were performed after each deprotection cycle.

Peptide cleavage was performed using TFA/TIS/H_2O = 95:2.5:2.5 (3 mL/100 mg resin). For sequences containing 1 or 2 Arg groups, the cleavage time was 2 h. For 3 or higher Arg containing peptides, cleavage time was 5h. Peptides were precipitated using cold Et_2O (-20°C) by adding approximately 5x volume of the TFA used for cleavage and centrifuging at 7000 rpm at 0°C.

IV. Analysis and Purification of peptides:

All peptides were analysed on a Thermo Scientific Dionex Ultimate 3000 RP-HPLC equipped with a Phenomenex Gemini NX C18 110 Å (150 x 4.6 mm) column using the following buffer systems: A: 1 mM Heptafluorobutyric acid in milliQ water. B: ACN using a flow rate of 1 ml/min unless specified otherwise. The column was flushed with 100% A for 5 min prior to an injection and was flushed for 5 min with 95% B and 5% A after the run was finished. All samples except **7** were dissolved in buffer A at a concentration of 1 mg/mL. Peptide **7** was dissolved in 0.1% HCOOH in mQ water.

Peptides were analysed using the following gradient: 100% A for 3 min. 0-95% B in 15 min. 95% B for 5 min. 100% A for 4 min. A: 1 mM Heptafluorobutyric acid in milliQ water. B: ACN

[I] Adachi, M.; Zhang, Y.; Leimkuhler, C.; Sun, B.; LaTour, J. V.; Kahne, D. E. *J. Am. Chem. Soc.* **2006**, *128*, 14012.

Peptides were purified using the same gradient and buffer system as mentioned above, on a Thermo Scientific Dionex Ultimate 3000 RP-HPLC with a flow rate of 4 mL/min using a Phenomenex Gemini NX C18 110 Å (150 x 10 mm) semi-prep column.

LC-MS/ESI-MS data were collected on an Agilent 1100 Series instrument with a Phenomenex Kinetex C18 100Å column (150 x 4.6 mm, 5 µm at 35 °C) connected to an ESMSD type VL mass detector with a flow rate of 1.5 ml/min was used with the following solvent systems: (A): 0.1% HCOOH in H_2O and (B) MeCN. The column was flushed with 100% A for 2 min, then a gradient from 0 to 100% B over 6 min was used, followed by 2 min of flushing with 100% B.

V. HPLC/LC-MS/ESI-MS Data

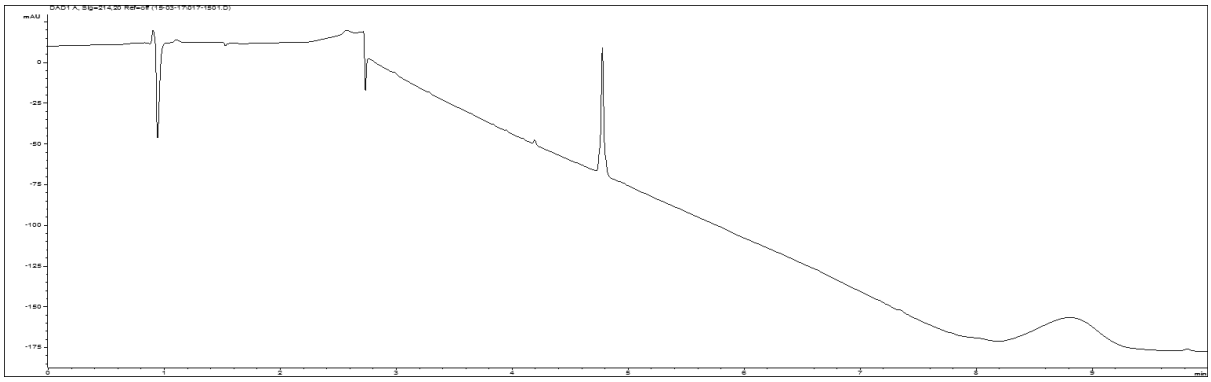


Fig. 1: HPLC trace of purified Moenomycin A using the following buffer system: A: 5 mM NH_4OAc, B: ACN. Gradient: 0-100% in 6 min (t_R = 4.779 min)

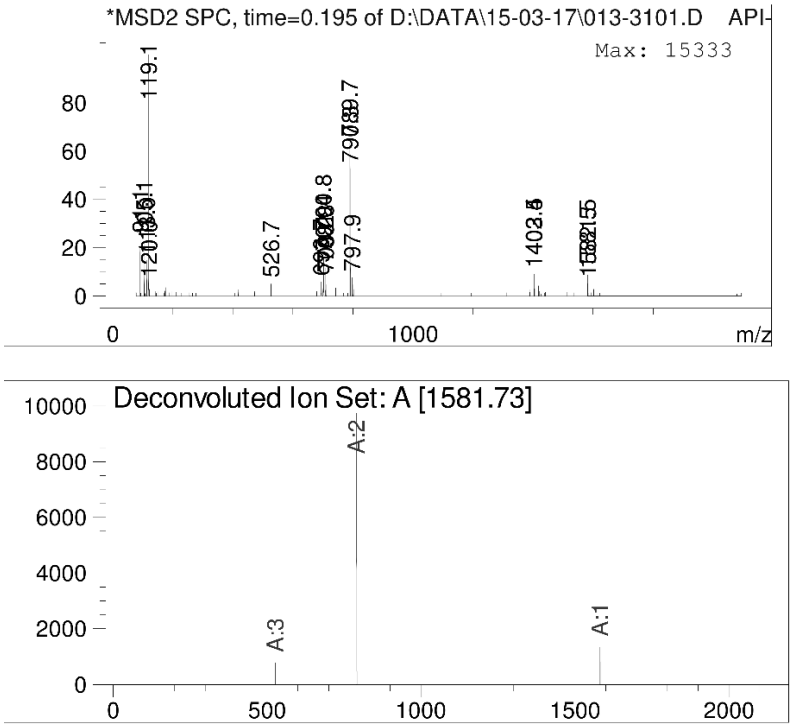


Fig. 2: ESI-MS spectra in negative mode of Moenomycin A with deconvolution. Exact Mass calcd. for $C_{69}H_{108}N_5O_{34}P$ = 1581.66, deconvoluted mass found 1581.73.

Sr. No.	Peptide	Chemical Formula	Exact Mass	Mass found
1	H_2N-RRRRRRRRR-$CONH_2$	$C_{54}H_{111}N_{37}O_9$	1421.94	1422.59
2	H_2N-KRRKRRKRR-$CONH_2$	$C_{54}H_{111}N_{31}O_9$	1337.92	1338.56
3	H_2N-KKKKKKKKK-$CONH_2$	$C_{54}H_{111}N_{19}O_9$	1169.88	1170.31
4	H_2N-RRRRRRRR-$CONH_2$	$C_{48}H_{99}N_{33}O_8$	1265.84	1266.44
5	H_2N-RKKRRQRRR-$CONH_2$	$C_{53}H_{107}N_{31}O_{10}$	1337.88	1338.56
6	H_2N-KKKKKKKK-$CONH_2$	$C_{48}H_{99}N_{17}O_8$	1041.79	1041.99
7	H_2N-RRWWRRWRR-$CONH_2$	$C_{69}H_{105}N_{31}O_9$	1511.87	1512.44
8	H_2N-KKKKKR-$CONH_2$	$C_{36}H_{75}N_{15}O_6$	813.60	813.88
9	H_2N-RRRRR-$CONH_2$	$C_{30}H_{63}N_{21}O_5$	797.53	798.5 [$M+H^+$]
10	H_2N-KKKKK-$CONH_2$	$C_{30}H_{63}N_{11}O_5$	657.50	658.5 [$M+H^+$]
11	H_2N-RRRR-$CONH_2$	$C_{24}H_{51}N_{17}O_4$	641.43	642.4 [$M+H^+$]
12	H_2N-KKKK-$CONH_2$	$C_{24}H_{51}N_9O_4$	529.41	530.4 [$M+H^+$]
13	H_2N-RRR-$CONH_2$	$C_{18}H_{39}N_{13}O_3$	485.33	486.3 [$M+H^+$]
14	H_2N-KKK-$CONH_2$	$C_{18}H_{39}N_7O_3$	401.31	402.3 [$M+H^+$]
15	H_2N-RR-$CONH_2$	$C_{12}H_{27}N_9O_2$	329.23	330.2 [$M+H^+$]
16	H_2N-KK-$CONH_2$	$C_{12}H_{27}N_5O_2$	273.22	274.3 [$M+H^+$]
17	H_2N-R-COOH	$C_6H_{14}N_4O_2$	174.11	-
18	H_2N-K-COOH	$C_6H_{14}N_2O_2$	146.11	-

Table 1: Complete list of peptides synthesized along with their respective chemical formula, exact mass and obtained deconvoluted mass.
Note: With the exception of peptide **7**, all the peptides elute in the first minute itself in case of standard HCOOH and TFA buffers. Therefore, we use a 1mM Heptafluorobutyric acid buffer to analyze these peptides. The broad peak in each spectrum at t ~ 7.8 min is a result of the Heptafluorobutyric acid buffer. The peak just after 25 min is a result of the re-equilibration of the column after the analysis is complete.

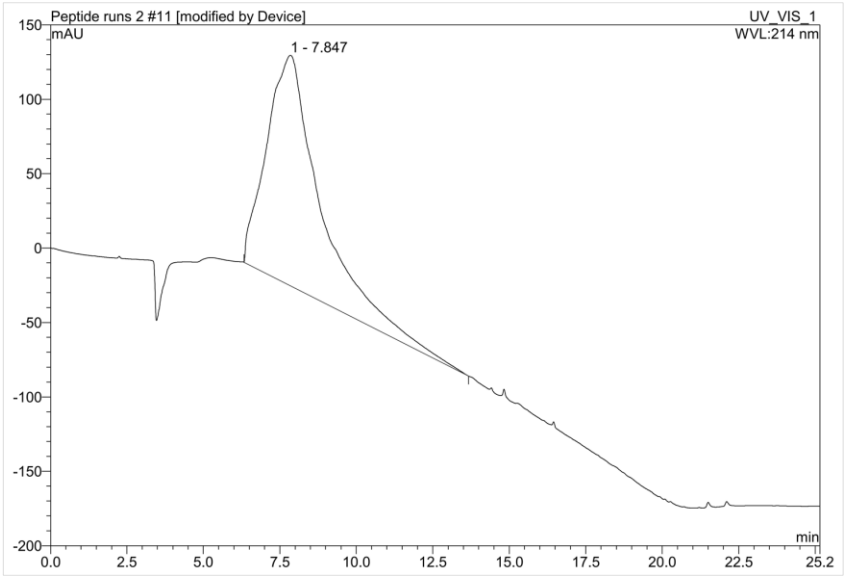


Fig. 3: RP-HPLC trace of mQ water using the gradient described in IV.

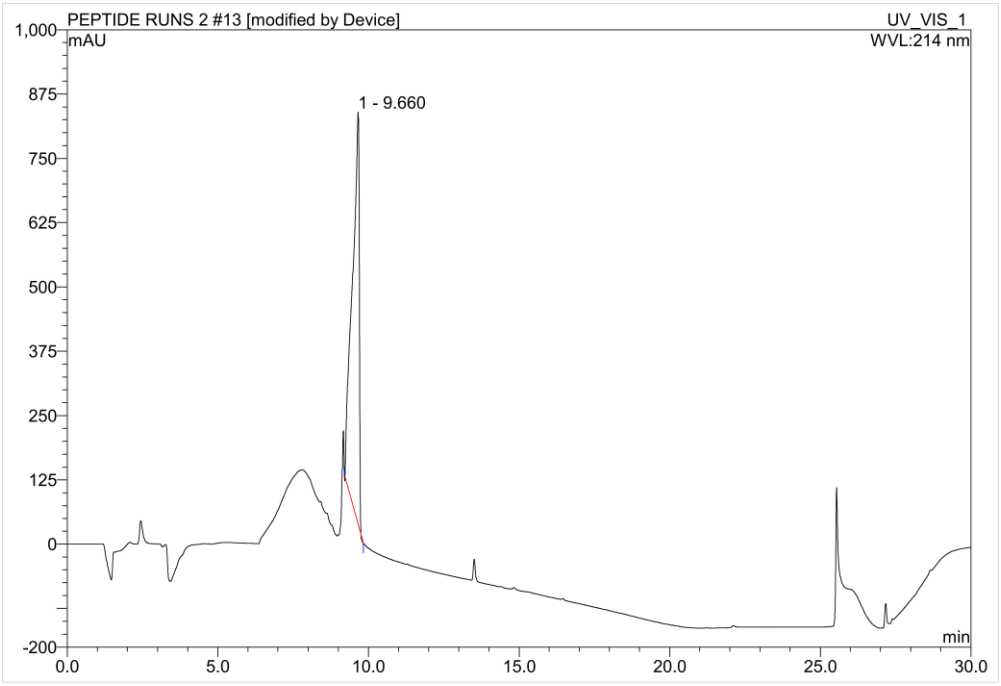


Fig. 4: RP-HPLC trace of HPLC purified peptide **1** showing product peak at r.t = 9.660 min using the gradient described in IV.

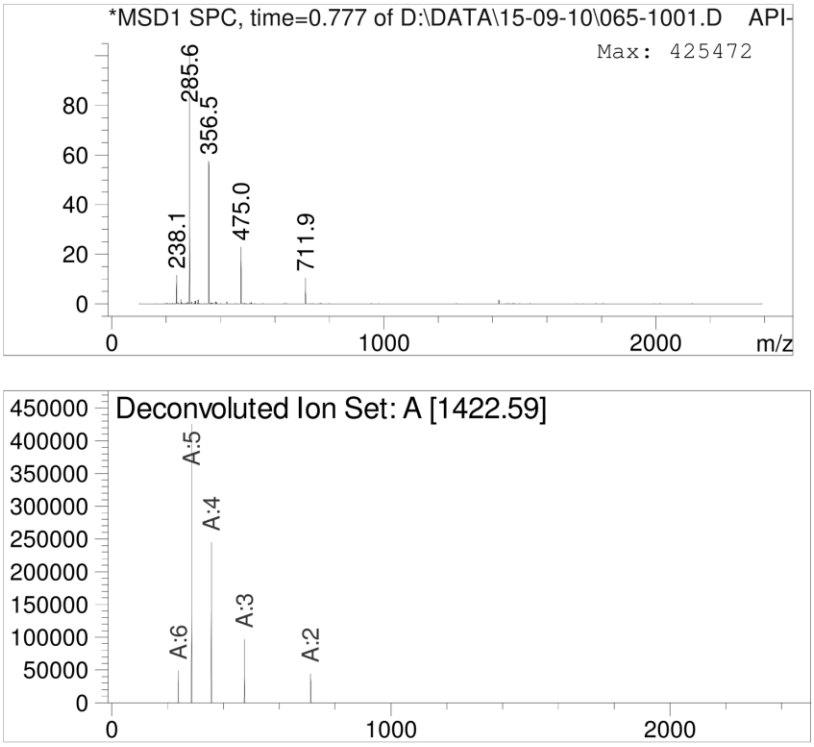


Fig. 5: ESI-MS spectra of RP-HPLC purified peptide **1** with deconvolution.

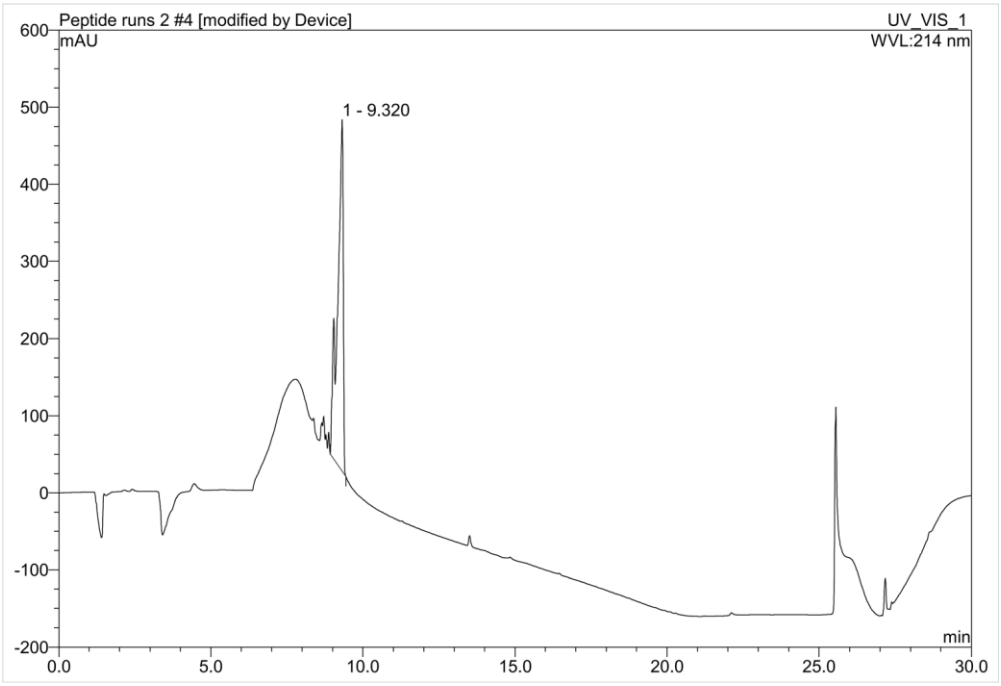


Fig. 6: RP-HPLC trace of HPLC purified peptide **2** showing product peak at r.t = 9.320 min using the gradient described in IV.

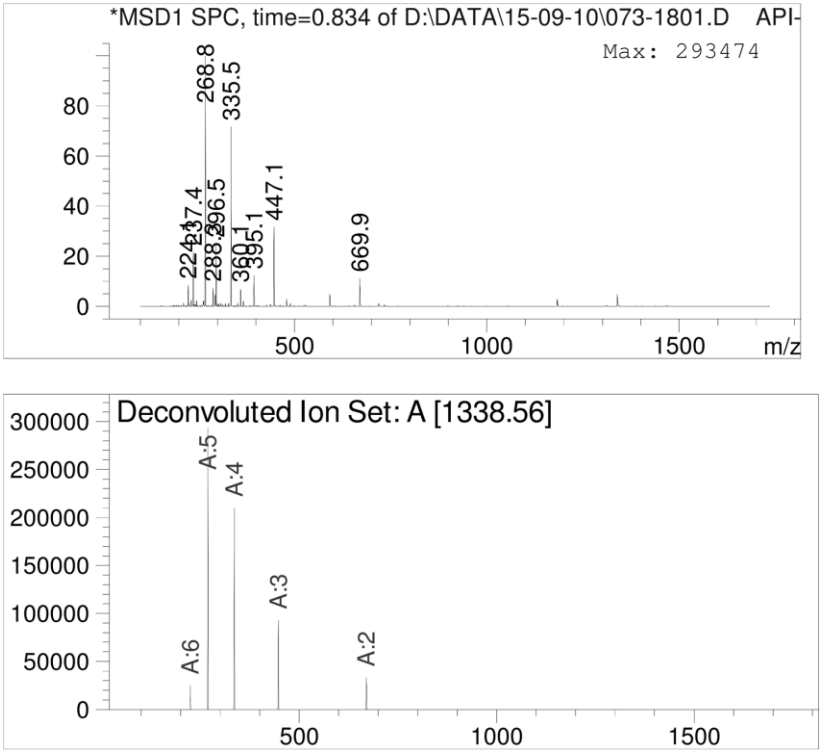


Fig. 7: ESI-MS spectra of RP-HPLC purified peptide **2** with deconvolution.

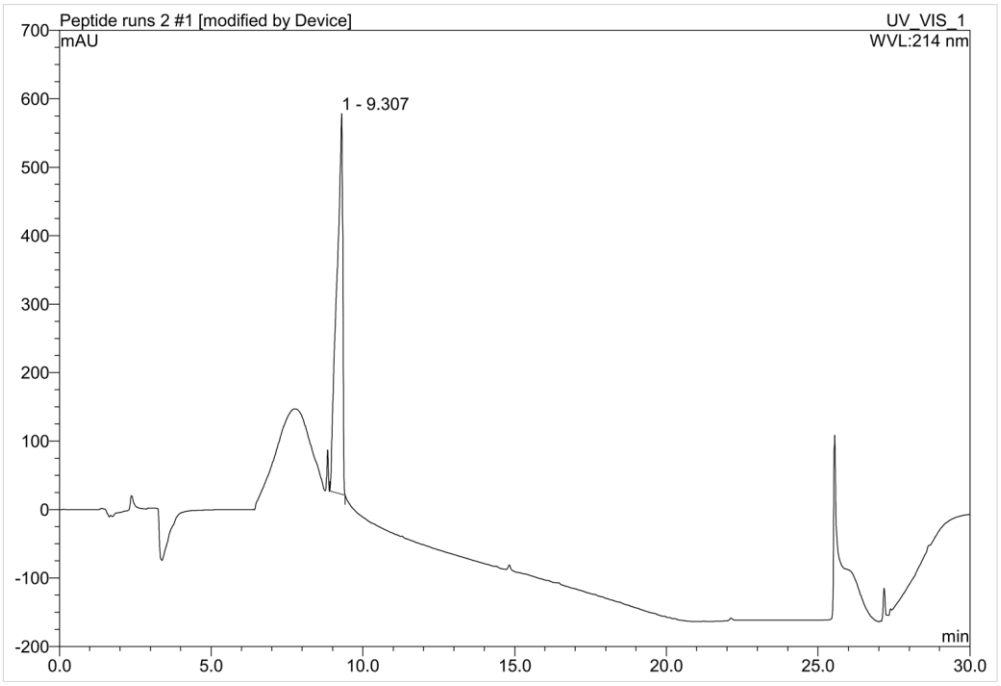


Fig. 8: RP-HPLC trace of HPLC purified peptide **3** showing product peak at r.t = 9.307 min using the gradient described in IV.

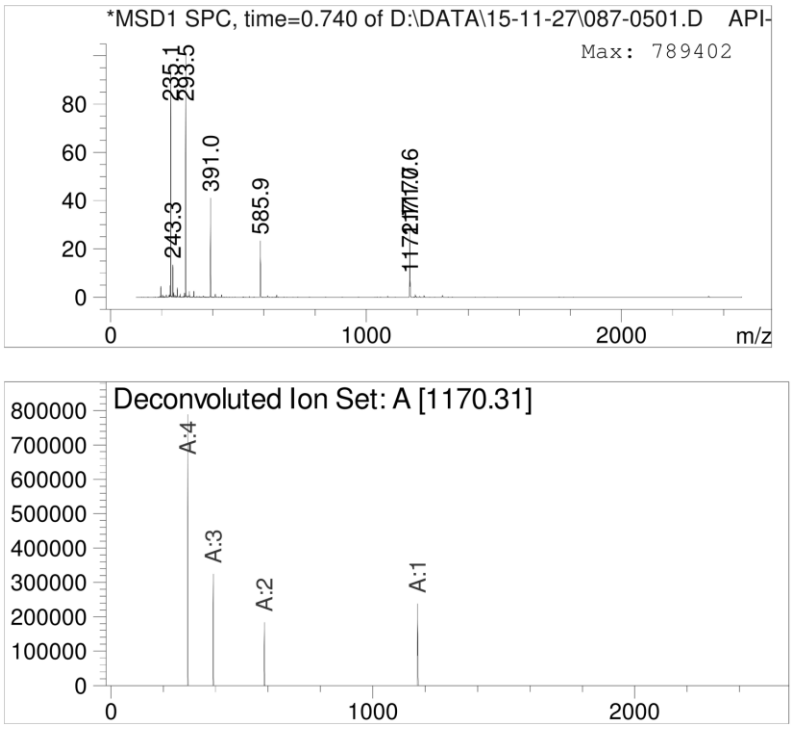


Fig. 9: ESI-MS spectra of RP-HPLC purified peptide **3** with deconvolution.

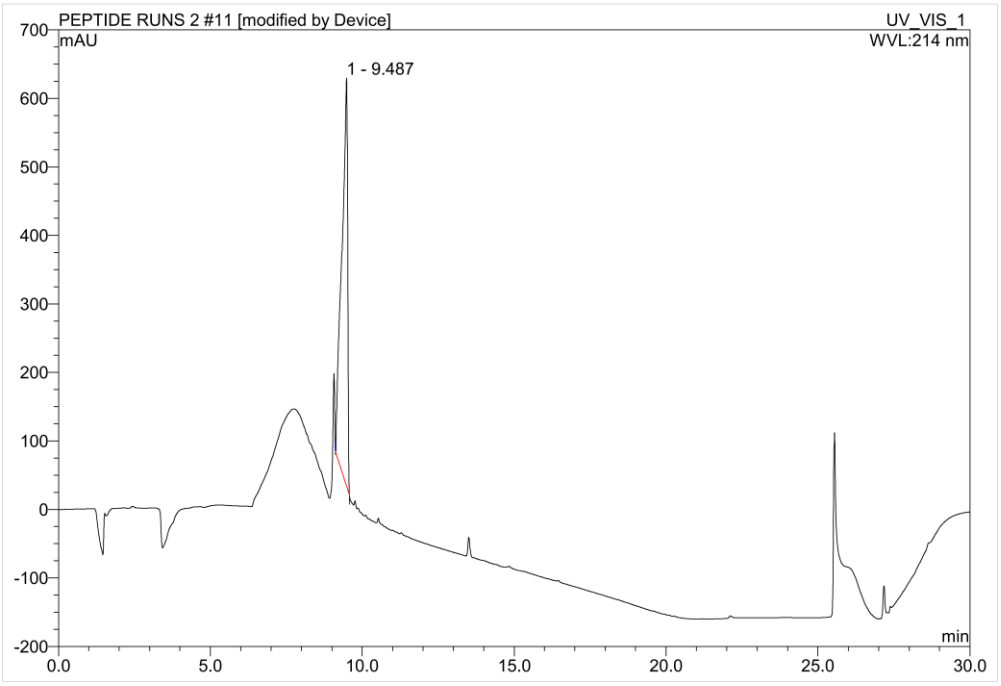


Fig. 10: RP-HPLC trace of HPLC purified peptide **4** showing product peak at r.t = 9.487 min using the gradient described in IV.

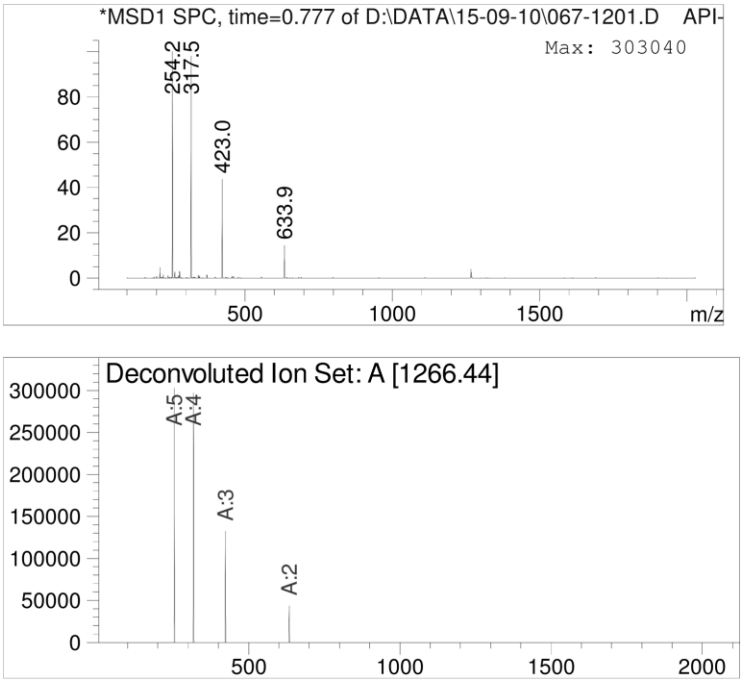


Fig. 11: ESI-MS spectra of RP-HPLC purified peptide **4** with deconvolution.

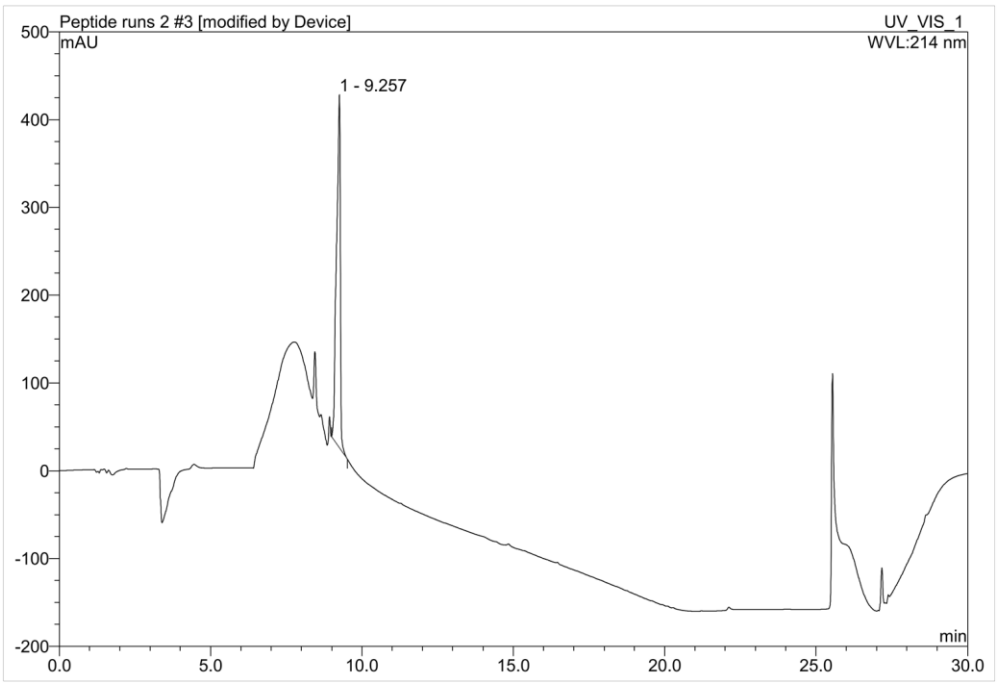


Fig. 12: RP-HPLC trace of HPLC purified peptide **5** showing product peak at r.t = 9.257 min using the gradient described in IV.

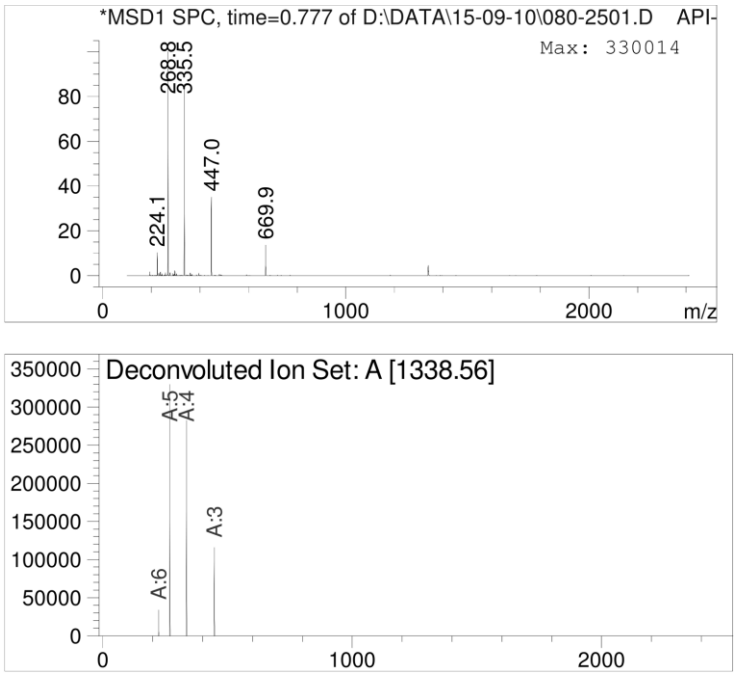


Fig. 13: ESI-MS spectra of RP-HPLC purified peptide **5** with deconvolution.

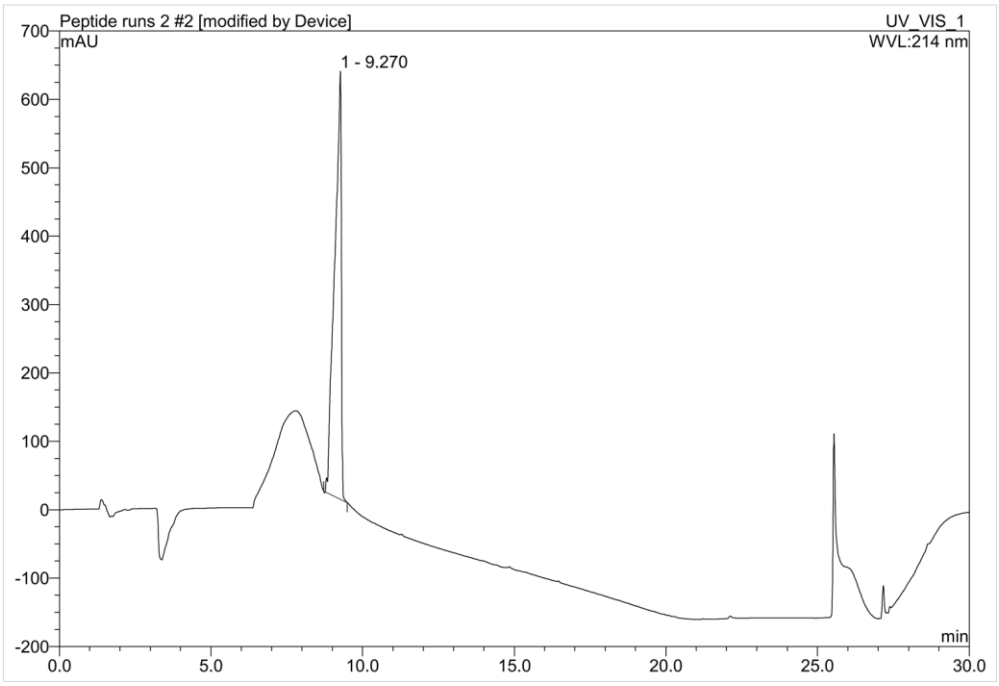


Fig. 14: RP-HPLC trace of HPLC purified peptide **6** showing product peak at r.t = 9.270 min using the gradient described in IV.

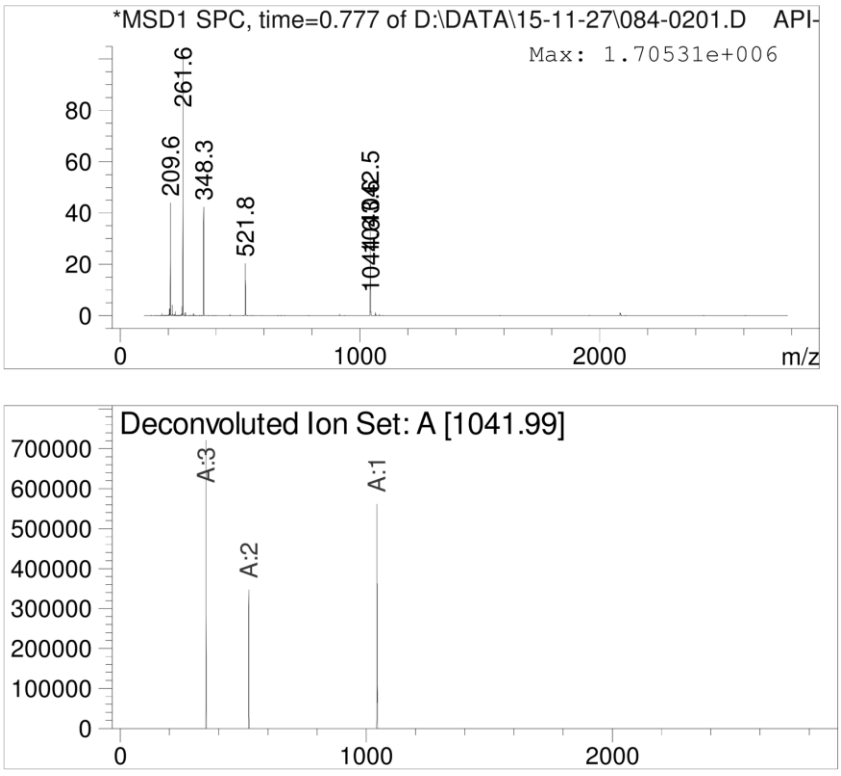


Fig. 15: ESI-MS spectra of RP-HPLC purified peptide **6** with deconvolution.

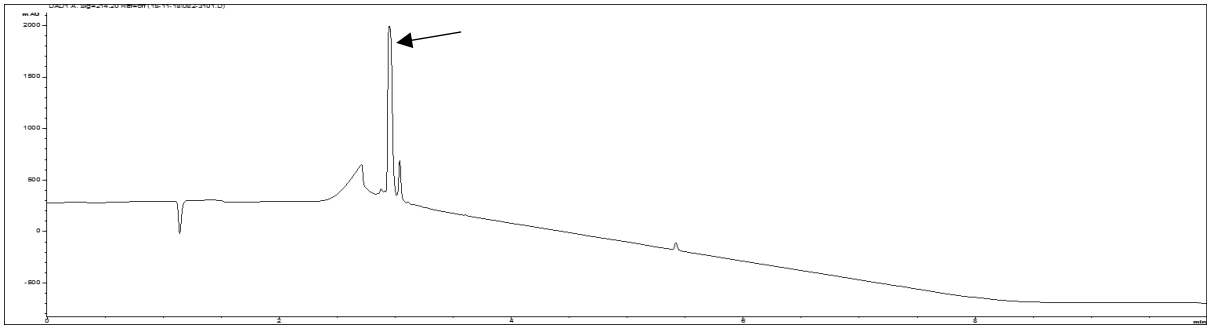


Fig. 16: HPLC trace of HPLC purified peptide **7** using the following gradient: 100% A for 2 min, then a gradient from 0 to 100% B over 6 min, followed by 2 min of flushing with 100% B. A: 0.1% HCOOH in mQ water. B: ACN

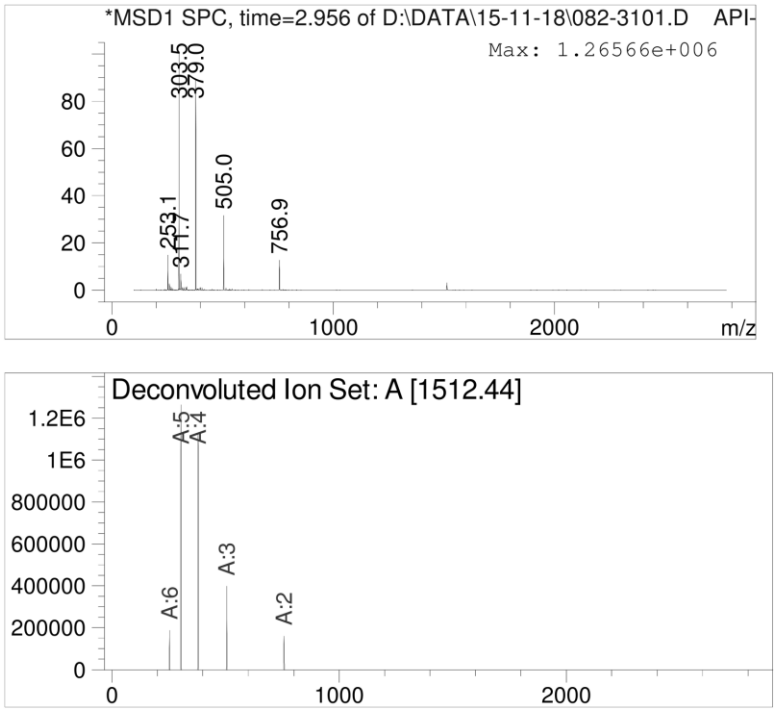


Fig. 17: ESI-MS spectra from LC-MS of RP-HPLC purified peptide **7** with deconvolution.

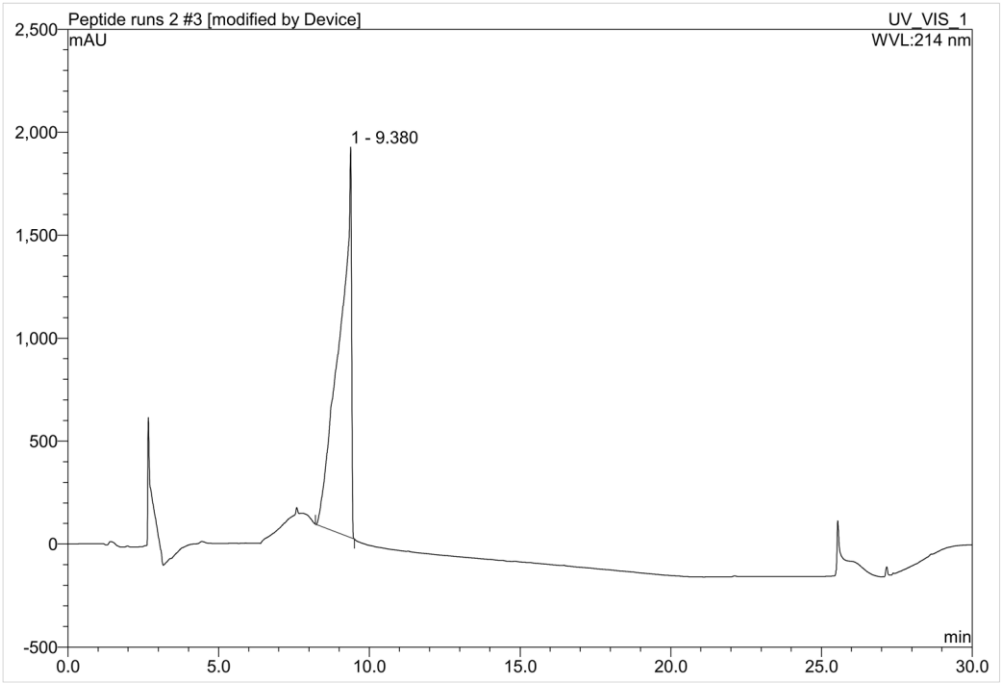


Fig. 18: RP-HPLC trace of HPLC purified peptide **8** showing product peak at r.t = 9.380 min using the gradient described in IV.

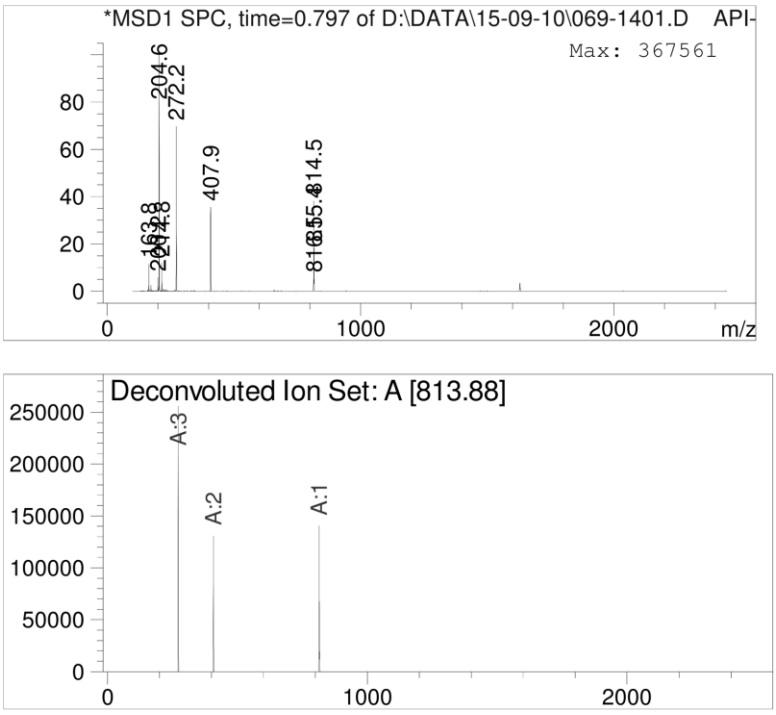


Fig. 19: ESI-MS spectra of RP-HPLC purified peptide **8** with deconvolution.

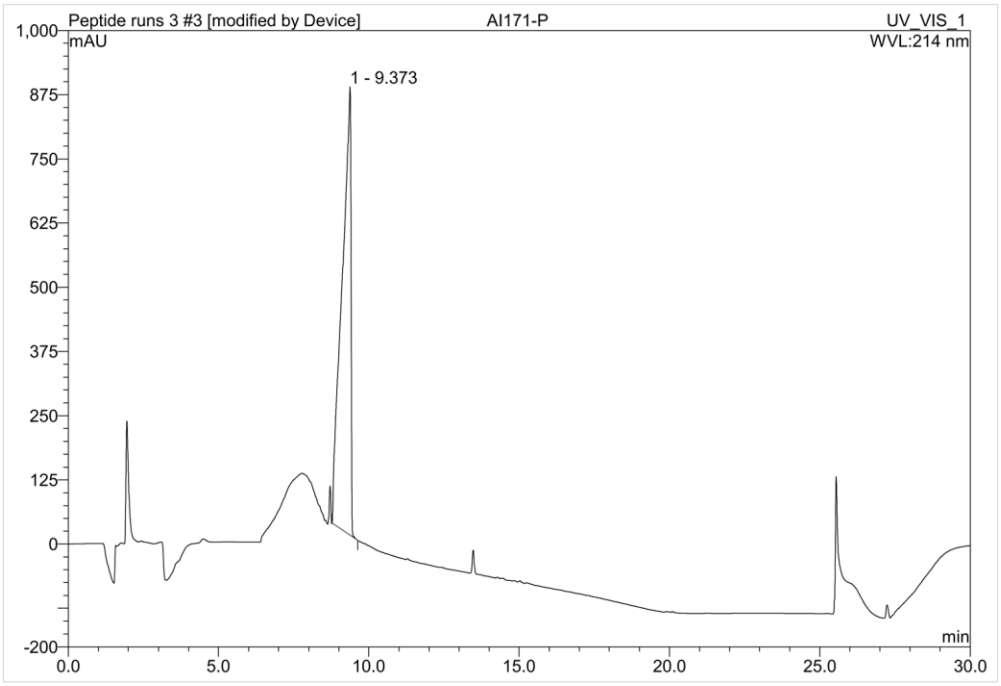


Fig. 20: RP-HPLC trace of HPLC purified peptide **9** showing product peak at r.t = 9.373 min using the gradient described in IV.

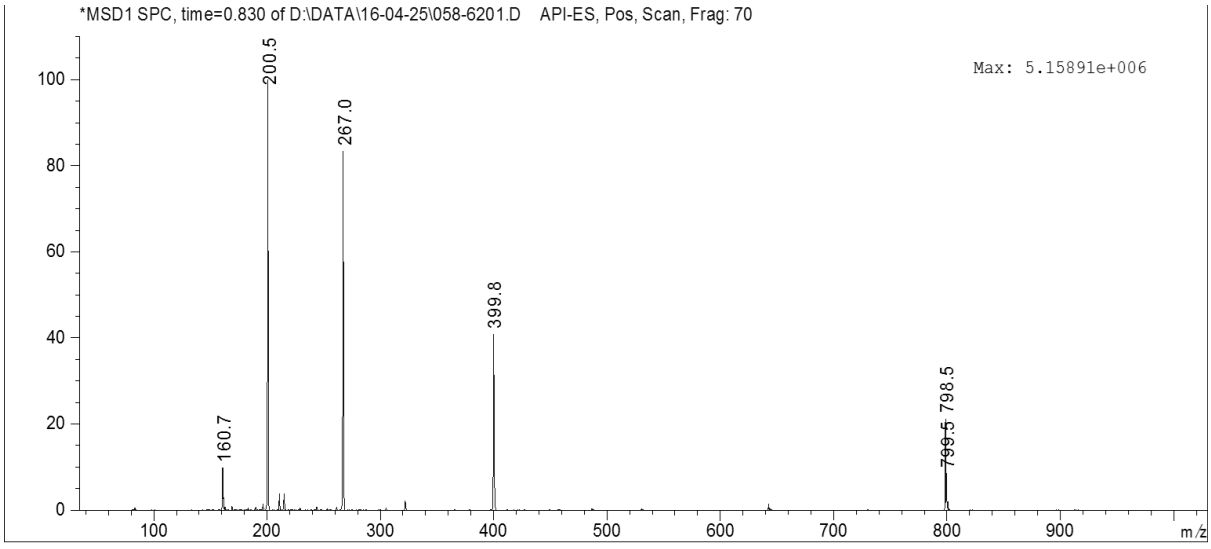


Fig. 21: ESI-MS spectra of RP-HPLC purified peptide **9**.

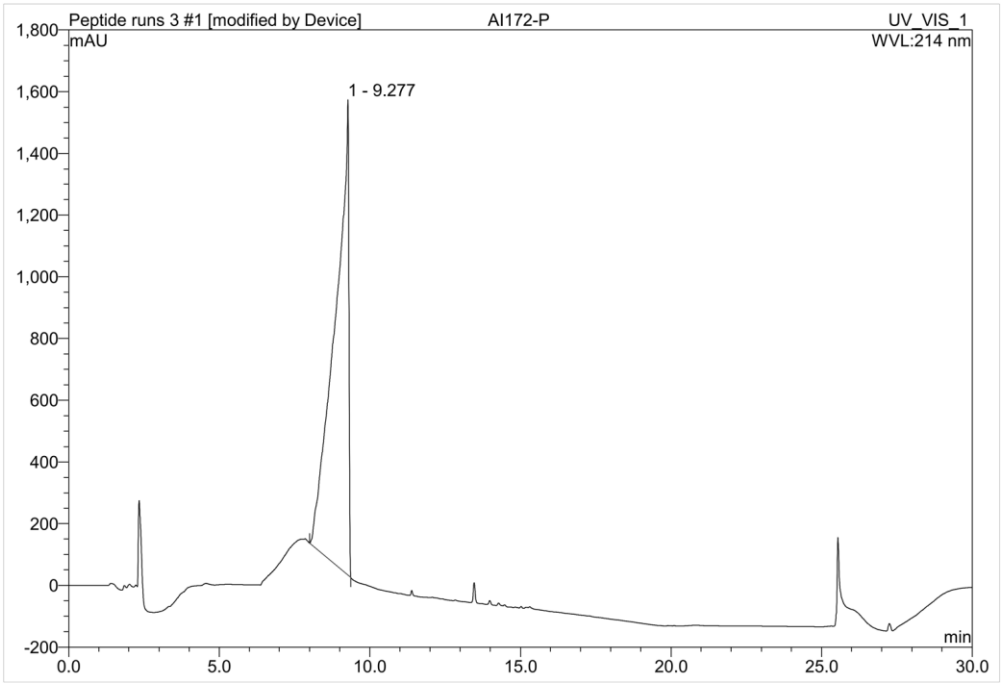


Fig. 22: RP-HPLC trace of HPLC purified peptide **10** showing product peak at r.t = 9.277 min using the gradient described in IV.

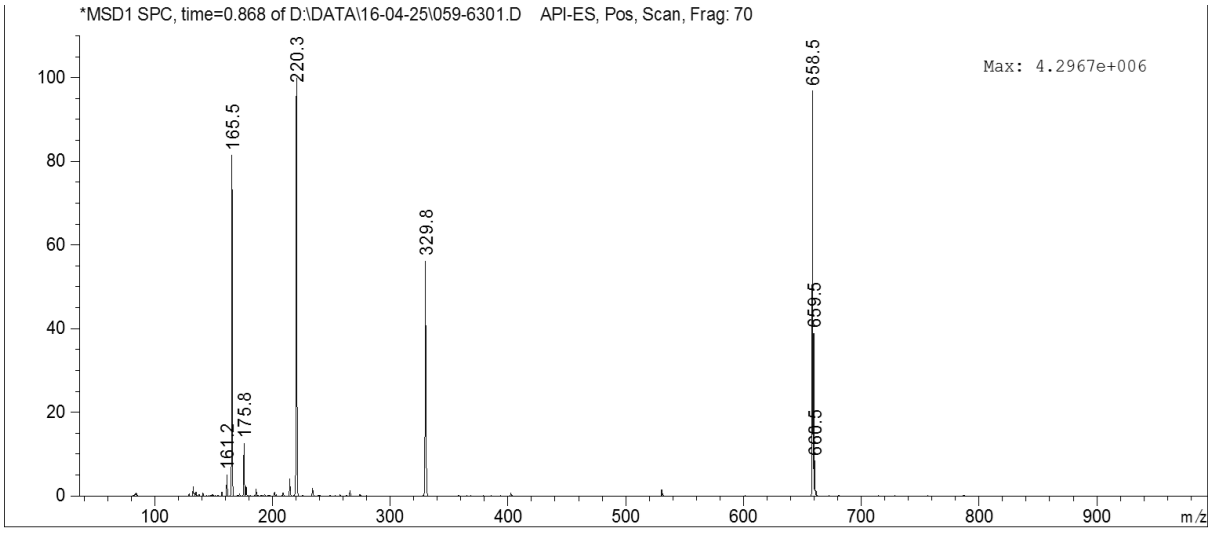


Fig. 23: ESI-MS spectra of RP-HPLC purified peptide **10**

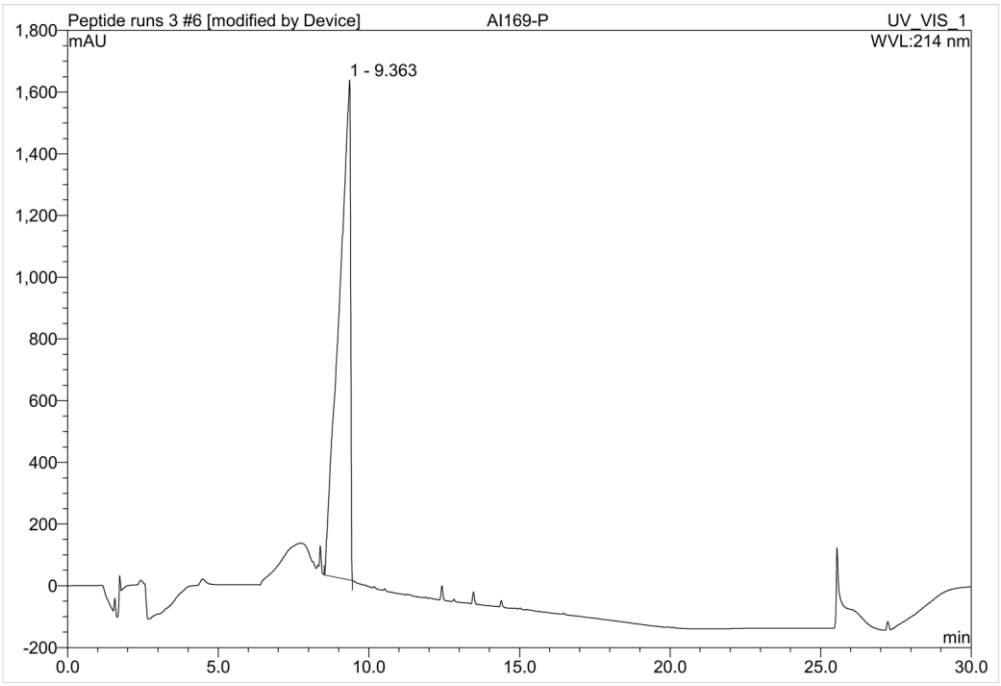


Fig. 24: RP-HPLC trace of HPLC purified peptide **11** showing product peak at r.t = 9.363 min using the gradient described in IV.

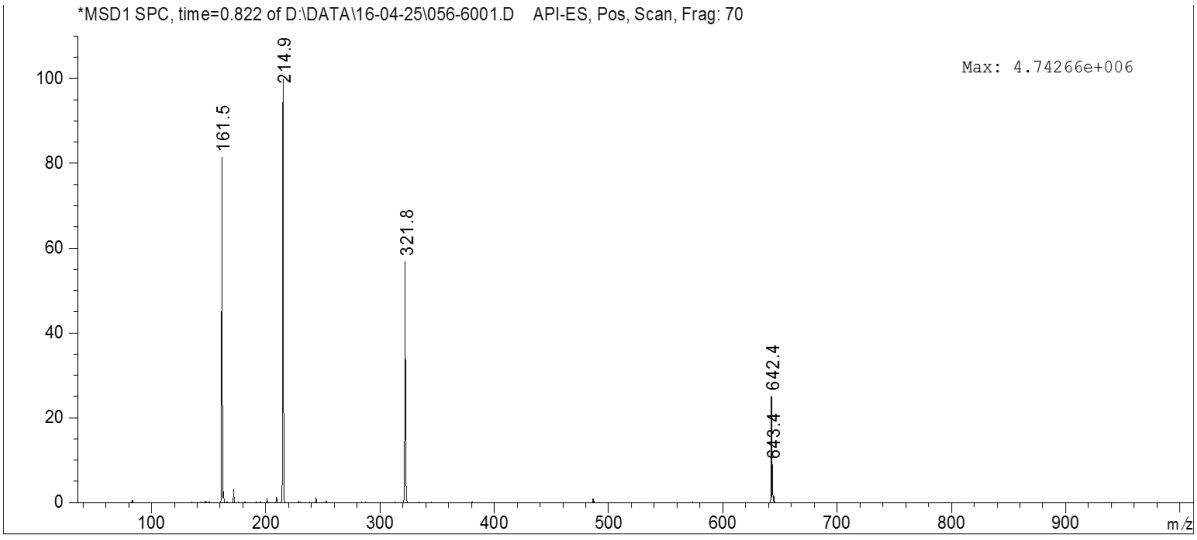


Fig. 25: ESI-MS spectra of RP-HPLC purified peptide **11**.

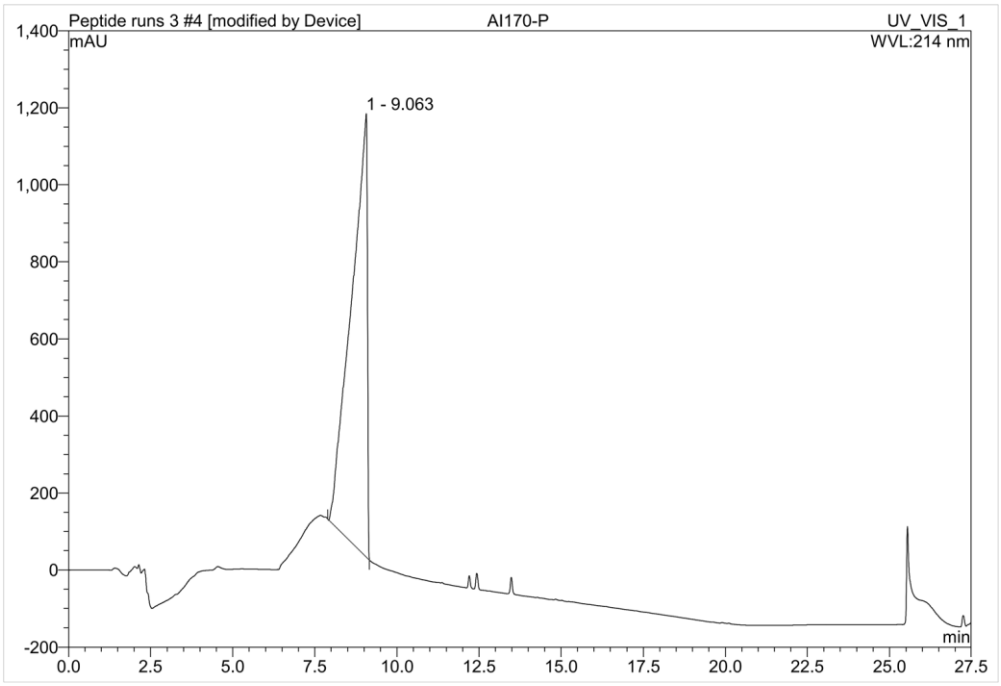


Fig. 26: RP-HPLC trace of HPLC purified peptide **12** showing product peak at r.t = 9.063 min using the gradient described in IV.

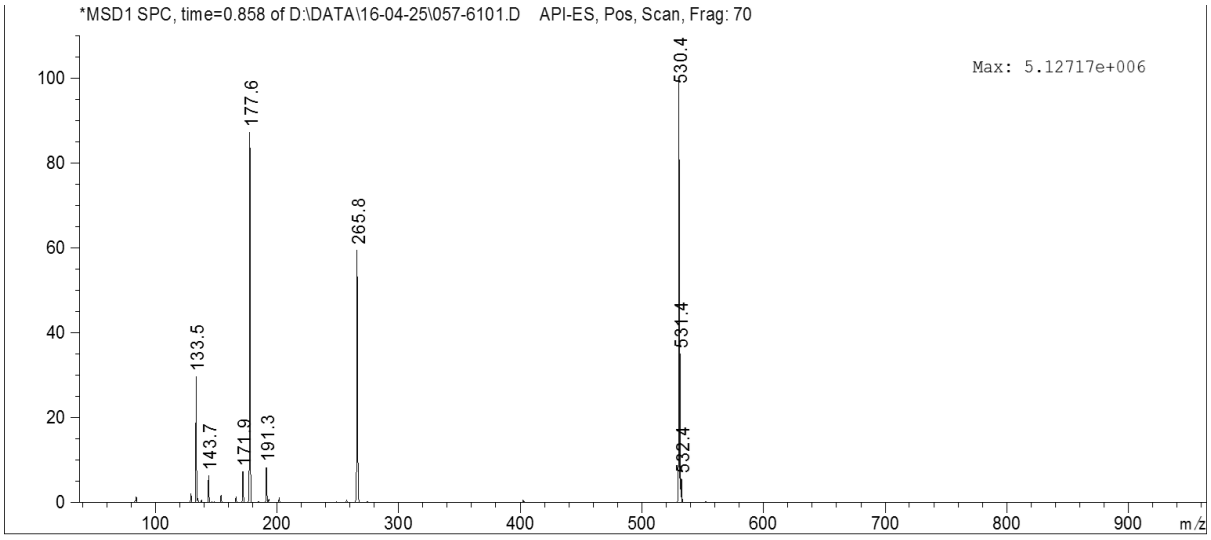


Fig. 27: ESI-MS spectra of RP-HPLC purified peptide **12**.

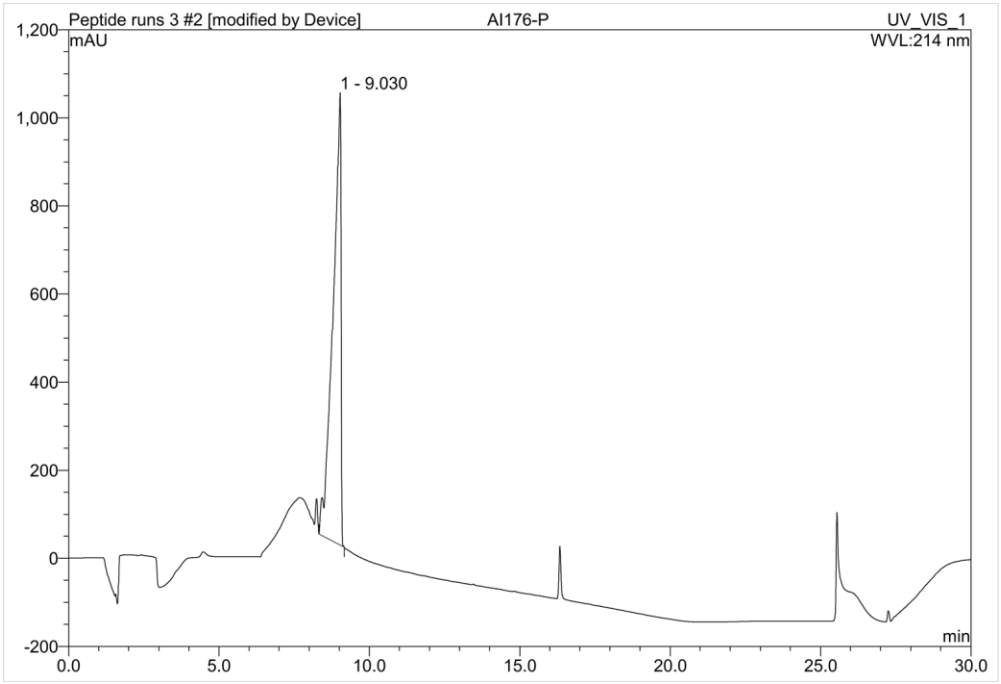


Fig. 28: RP-HPLC trace of HPLC purified peptide **13** showing product peak at r.t = 9.030 min using the gradient described in IV.

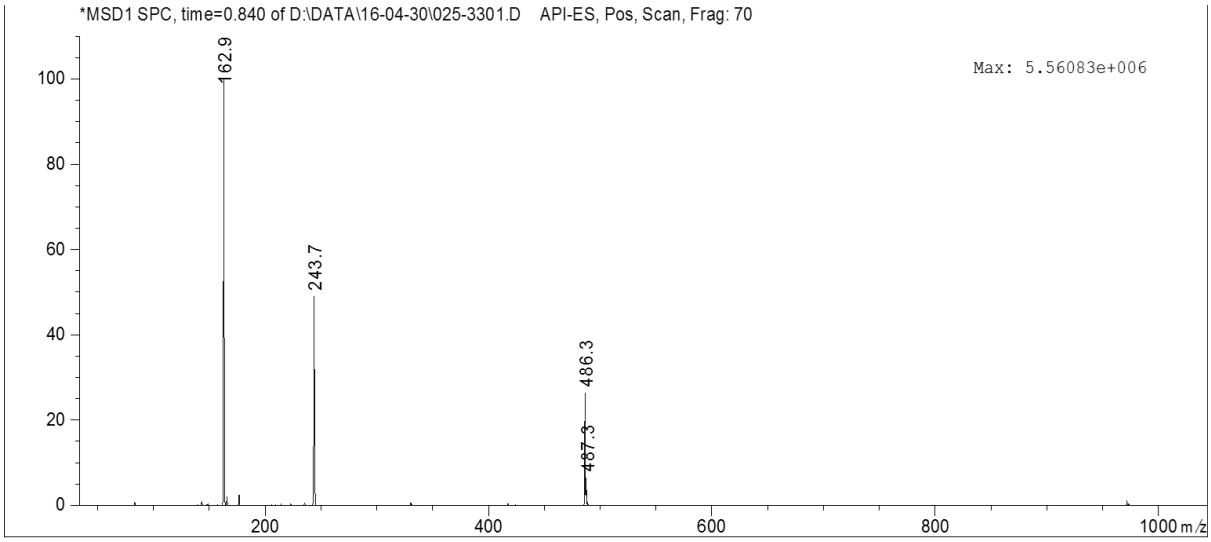


Fig. 29: ESI-MS spectra of RP-HPLC purified peptide **13**.

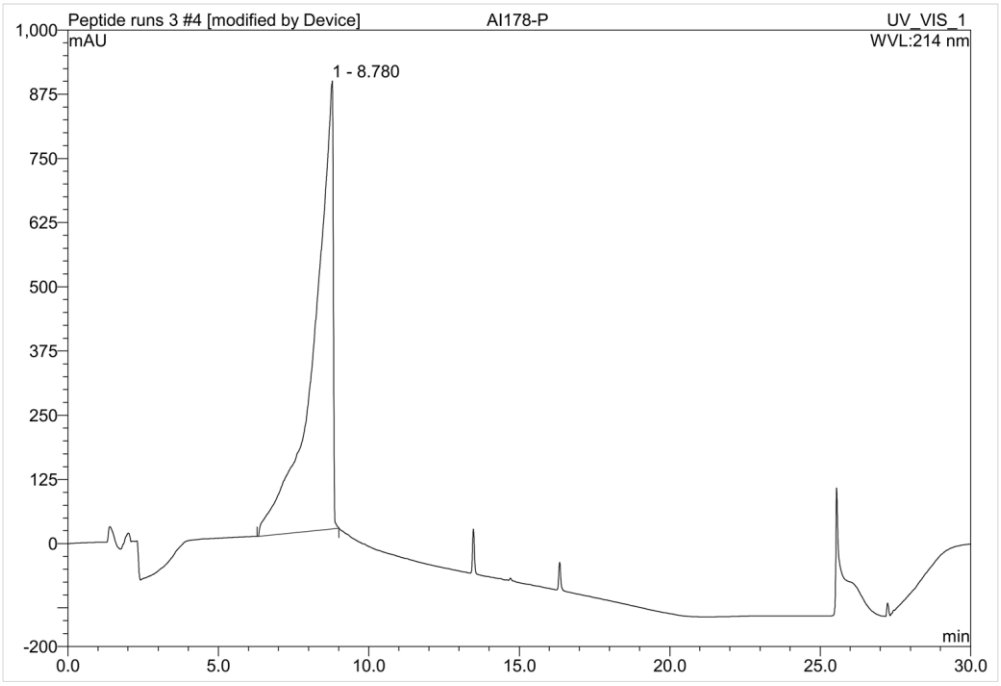


Fig. 30: RP-HPLC trace of HPLC purified peptide **14** showing product peak at r.t = 8.780 min using the gradient described in IV.

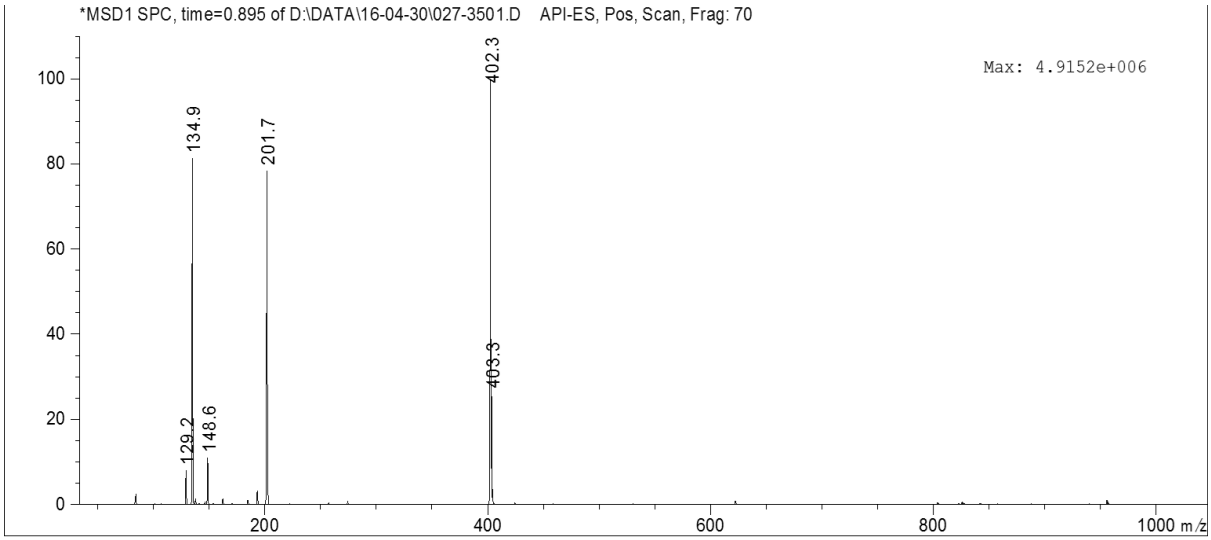


Fig. 31: ESI-MS spectra of RP-HPLC purified peptide **14**.

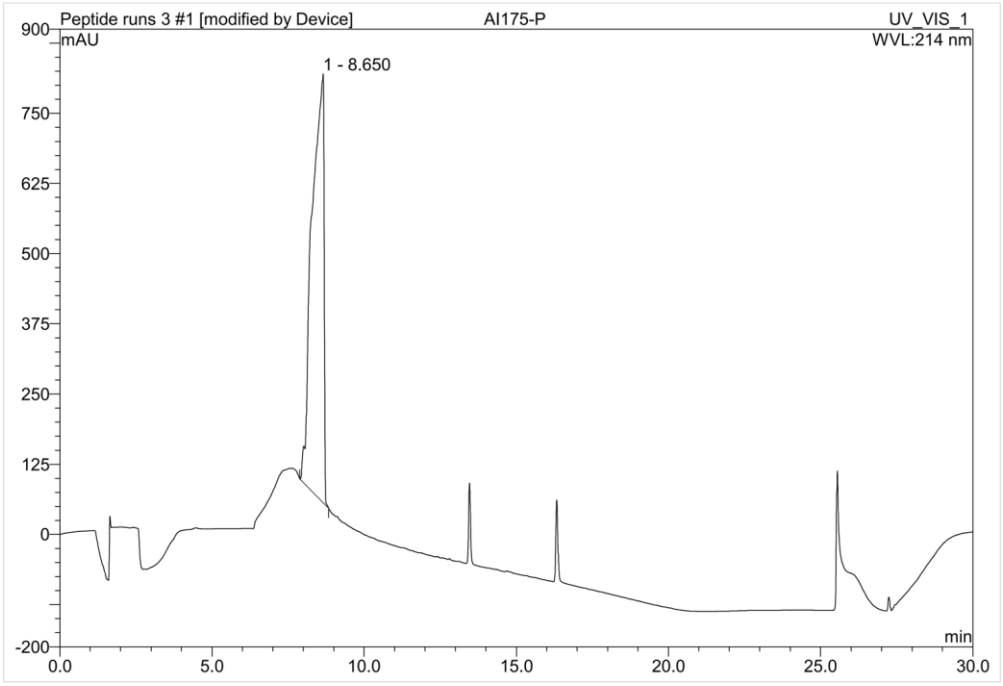


Fig. 32: RP-HPLC trace of HPLC purified peptide **15** showing product peak at r.t = 8.650 min using the gradient described in IV.

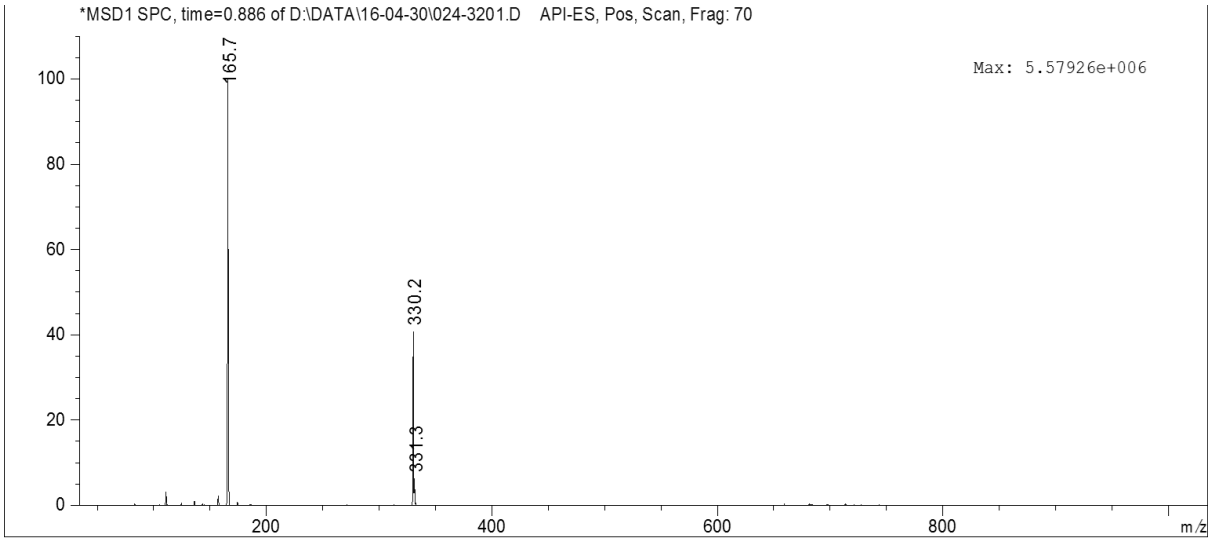


Fig. 33: ESI-MS spectra of RP-HPLC purified peptide **12**.

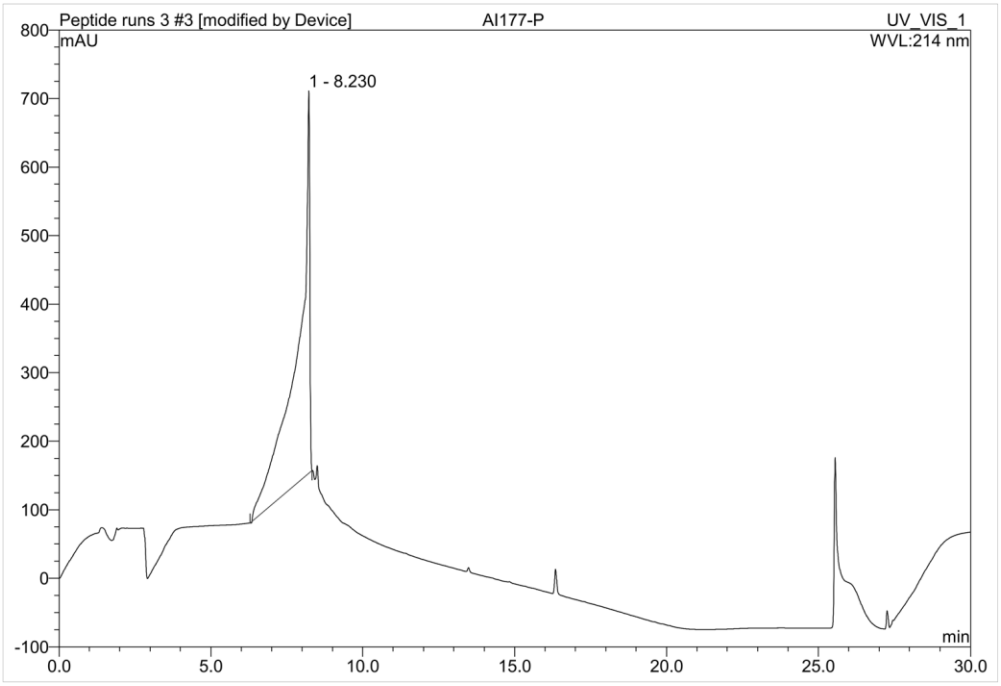


Fig. 34: RP-HPLC trace of HPLC purified peptide **16** showing product peak at r.t = 8.230 min using the gradient described in IV.

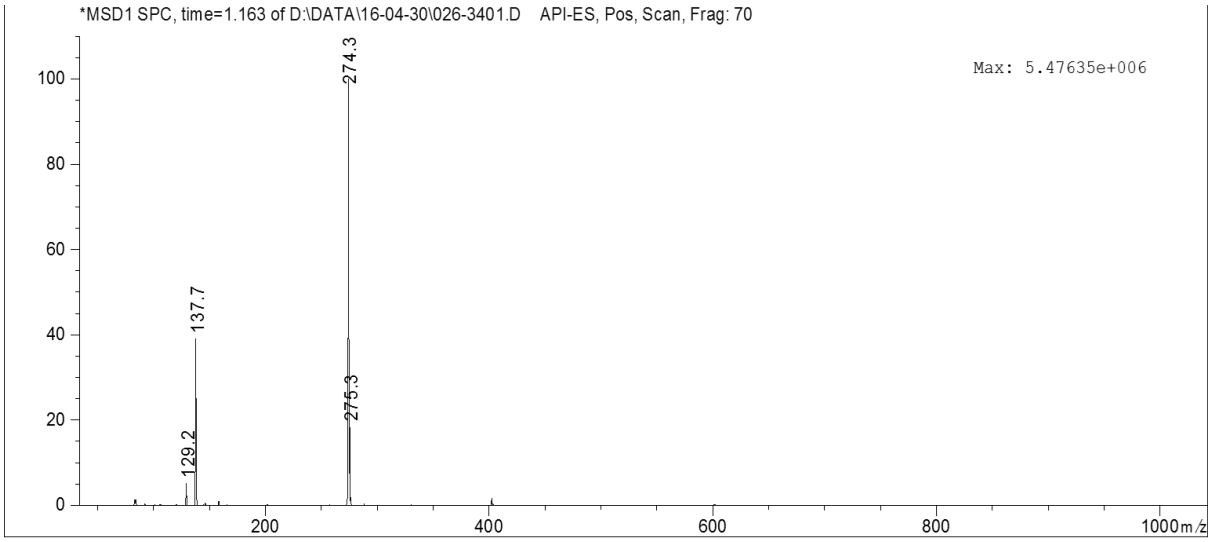


Fig. 35: ESI-MS spectra of RP-HPLC purified peptide **16**.

VI. MIC protocols

For MIC testing all peptides were dissolved in pure mQ water. Bacteria were grown on Mueller Hinton broth (oxoid) with no additional supplements unless otherwise stated. All incubations were at 37°C unless otherwise stated. When MoeA was tested in conjunction with the peptides, they were incubated together for 1 hour in a 1:1 ratio (in terms of mM, MoeA to peptide), before being used in the testing. All dilutions of MoeA and peptides were carried out using Mueller Hinton broth unless otherwise stated. 100 µl of autoclaved Mueller Hinton broth was added to wells 2-12 on a 96-well plate. 200 µl of the inhibitor was added to well one at a concentration of 512 µg/ml. 100µl of MoeA in well one was taken up and pipetted into well two. The mixture was the pipetted up and down three times before 100µl was taken up and pipetted into well three. This process was repeated up to well 11. Once MoeA has been added to well 11 100 µl was taken up and then discarded ensuring the well 12 had no inhibitor present. Each well was then inoculated with 100µl of bacteria that had been diluted to an OD_{600nm} of 0.1. This was repeated three times. The 96-well plates were then incubated for 24 hours. The results were measured using a plate reader (spectra max plus plate reader) where possible. The MIC was determined to be the lowest concentration at which there was no growth visible.

VII. Cytotoxicity assay by Formazan bioreduction

HeLa cells were seeded in a 96-well plate at 10^4 cells/cm^2 density in Dulbecco's Modified Eagles Medium (DMEM) supplemented with 10% serum. The cells were repeatedly rinsed with Hank's Balanced Salt Solution (HBSS) prior to be exposed to different peptide – MoeA complex or MoeA concentrations in the range of 0.5 – 100 µM in HBSS 24 hrs post-seeding. Following 6 hrs of exposure to the complex or MoeA, CellTiter 96 AQueous Nonradioactive Cell Proliferation Assay (Promega) was used according to the manufacturer's instructions[II]. Not ingested complex or MoeA was removed by repeated washings with fresh medium. 20 µL of the combined MTS/PMS solution was added to 100 µL fresh medium in each well and plates were incubated for 3 hrs at 37°C. Absorbance was measured at 490 nm on Tecan Infinite M200 PRO plate reader with i-control 1.10 software (Molecular Devices).

VIII. NMR Data

All NMR was performed on a Bruker Ascend 500 MHz equipped with a broadband probe. 600 µL 1mM solutions of MoeA and peptide **7** were made in 5 mM Ammonium acetate + 20 µL 0.1% acetic acid with 10% D_2O. ^1H-^1H TOCSY and ^1H-^1H NOESY were recorded on each sample at 4°C and 27°C with 2048 and 256 complex points in the direct and indirect dimensions, respectively. NOESY spectra were acquired with 200ms mixing time. MoeA was then titrated into the peptide at molar equivalents 0.1, 0.2, 04, 0.6, 0.8 and 1.0, with ^1H spectra recorded at each step. At a molar equivalence of 1:1 the NOESY and TOCSY experiments were repeated. Spectra were processed using Bruker TopSpin and analysed using CCPN Analysis[III]. The NOESY spectrum obtained from the complex was assigned iteratively,

[II] Cory, A. H.; *et al. Cancer communications* **1991**, *3*, 207

[III] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**:687-96

with structural calculations being performed using ARIA 2.3[IV]. The topology and parameters of MoeA were generated using PRODRG[V]. Distance restraints were calculated using NOE volume, and calibrated by optimising the NOE between anomeric sugar protons and other well-resolved sugar protons from the same sugar moiety of known distances. Rings B and F had suitable protons for NOE calibration. Calculated distances had an upper error bound of 20% applied. A lower bound of 1.72Å was used. Iterative NOE assignment was continued until all significant restraint violations had been eliminated. The generated ensemble was then subjected to a final round of refinement in explicit solvent and was visualised and analysed using PyMOL[VI]. Restraints have been deposited with the BMRB (accession no. xxxxx). Co-ordinates of the ensemble are available upon request.

[IV] Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**:381-382

[V] Schüttelkopf AW, van Aalten DMF (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D* **60**:1355–1363

[VI] The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC

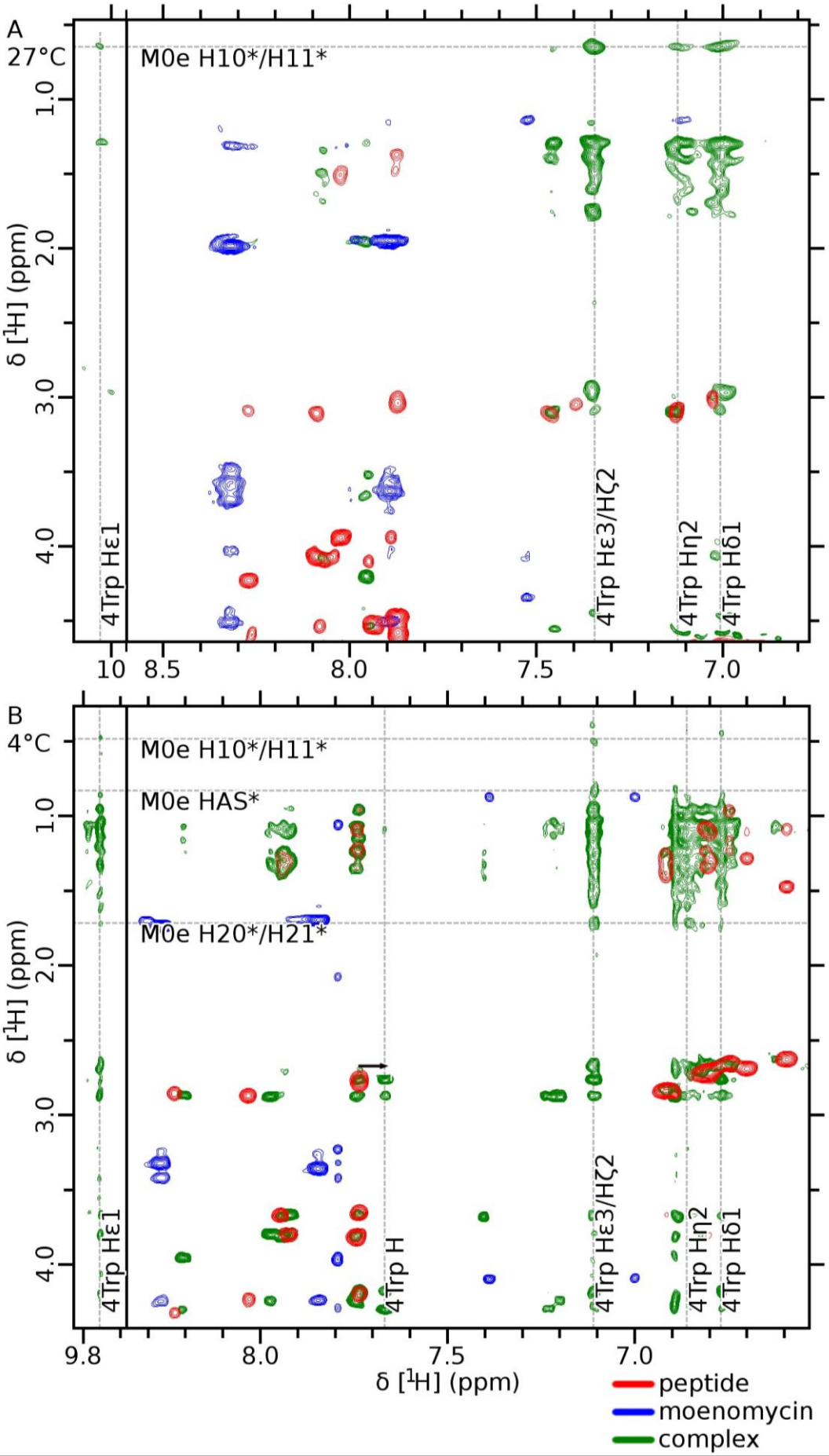

A
27°C
M0e H10*/H11*
δ [1H] (ppm)
1.0
2.0
3.0
4.0
4Trp Hε1
4Trp Hε3/Hζ2
4Trp Hη2
4Trp Hδ1
10
8.5
8.0
7.5
7.0
B
4°C
M0e H10*/H11*
M0e HAS*
M0e H20*/H21*
4Trp H
9.8
δ [1H] (ppm)
peptide
moenomycin
complex

Figure 36: Interaction between peptide **7** and MoeA revealed by NMR.
A. ^{1}H-^{1}H NOESY of amide and aromatic region of 1mM peptide (*red contours*), 1mM MoeA (*blue contours*) and 1:1 complex (*green contours*) recorded at 27°C, showing noticeable dispersion and clear evidence of intermolecular NOEs between 4Trp and the acyl chain of MoeA. Resolvable 4Trp protons and interacting MoeA protons are indicated by *dashed grey lines*. 600 µL 1mM solutions of MoeA and peptide **7** were made in 5 mM Ammonium acetate + 20 µL 0.1% acetic acid with 10% D_2O.
B. ^{1}H-^{1}H NOESY (MoeA and complex) and ^{1}H-1H TOCSY (peptide) recorded at 4°C. Sample and spectrum descriptions as before. Significantly more NOEs were visible at this temperature, allowing for a full structural characterisation of the complex. *Black arrow* indicates an example of the chemical shift perturbation experienced by the peptide upon interaction with MoeA.

Residue	Atom	Peptide		Moenomycin		Complex	
		4°C	37°C	4°C	37°C	4°C	37°C
1Arg	H						
1Arg	Hα	3.657	3.887			3.659	3.910
1Arg	Hβ	1.098	1.710			1.072	1.292
1Arg	Hγ	1.487	1.346			1.458	1.733 1.728
1Arg	Hδ	2.642	2.889			2.641	2.885
1Arg	Hε	6.605	6.821			6.625	6.698
1Arg	Hη1*					6.139	6.487
1Arg	Hη2*					6.519	6.696
2Arg	H	8.522				8.523	
2Arg	Hα	4.003	4.226			3.952	
2Arg	Hβ	1.137 1.294	1.515			1.137 1.238	
2Arg	Hγ	1.092 1.085	1.352			1.077	
2Arg	Hδ	2.705	2.945			2.677	
2Arg	Hε	6.712	6.908			6.745	
2Arg	Hη1*					6.127	
2Arg	Hη2*						
3Trp	H	8.234	8.272			8.206	
3Trp	Hα	4.324	4.573			4.295	4.541
3Trp	Hβ	2.887	3.098			2.859	3.095
3Trp	Hδ1	6.911	7.121			6.891	7.130
3Trp	Hε1	9.867	10.012			9.858	10.031
3Trp	Hε3	7.239	7.474			7.218	7.457
3Trp	Hη2	6.892	7.018			6.854	7.103
3Trp	Hζ2		7.361			7.109	7.353
3Trp	Hζ3	6.820	7.107			6.769	
4Trp	H	7.700	7.872			7.666	7.942
4Trp	Hα	4.204	4.475			4.180	4.453
4Trp	Hβ	2.785	3.017 3.036			2.841 2.762	2.956
4Trp	Hδ1	6.794	7.021			6.766	6.999
4Trp	Hε1	9.846	9.992			9.858	10.023
4Trp	Hε3	7.131	7.516			7.107	7.346

4Trp	Hη2	6.877	7.019			6.858	7.064
4Trp	Hζ2	7.132	7.516			7.107	7.345
4Trp	Hζ3		7.019			6.756	6.996
5Arg	H	7.757	7.874			7.736	7.947
5Arg	Hα	3.683	3.943			3.669	3.911
5Arg	Hβ	1.269 1.177	1.401 1.488			1.245 1.142	
5Arg	Hγ	0.977	1.224			0.954	1.162
5Arg	Hδ	2.678	2.907			2.664	2.910
5Arg	Hε	6.783	6.928			6.766	7.008
5Arg	Hη1*	6.143				6.136	
5Arg	Hη2*	6.384					
6Arg	H	7.948	8.020			7.934	8.067
6Arg	Hα	3.822	4.071			3.802	4.060
6Arg	Hβ	1.321 1.121	1.563			1.326 1.085	1.339
6Arg	Hγ	1.321	1.354			1.332	
6Arg	Hδ	2.732	2.954			2.698	2.954
6Arg	Hε	6.872	6.974			6.835	7.087
6Arg	Hη1*	6.145				6.113	
6Arg	Hη2*	6.550				6.514	
7Trp	H	8.002	8.090			7.960	
7Trp	Hα	4.265	4.531			4.244	4.455
7Trp	Hβ	2.895	3.105			2.871	3.098
7Trp	Hδ1	6.905	7.125			6.890	7.127
7Trp	Hε1	9.867	10.013			9.858	10.029
7Trp	Hε3	7.228	7.387			7.209	7.454
7Trp	Hη2	6.882	7.005			6.872	7.102
7Trp	Hζ2		7.359			7.109	7.353
7Trp	Hζ3	6.783	7.106			6.758	7.009
8Arg	H	7.768	7.937			7.739	8.071
8Arg	Hα	3.826	4.092			3.805	4.084
8Arg	Hβ	1.267 1.361	1.503 1.615			1.244 1.339	1.581 1.622
8Arg	Hγ	1.091	1.344			1.079 1.072	1.339 1.304
8Arg	Hδ	2.721	2.980			2.704	2.973
8Arg	Hε	6.862	6.998			6.859	7.102
8Arg	Hη1*	6.145				6.130	
8Arg	Hη2*	6.575					
9Arg	H	7.963	8.065			7.950	8.073
9Arg	Hα	3.699	3.992			3.673	3.990
9Arg	Hβ	1.276	1.666			1.250 1.100	1.502 1.295
9Arg	Hγ	1.367 1.435	1.527 1.553			1.354 1.403	1.623
9Arg	Hδ	2.848	3.077			2.848	3.076
9Arg	Hε	6.928	7.082			6.898	7.134

9Arg	Hη1*	6.162				6.133	
9Arg	Hη2*	6.615				6.148	
10Nh2	Hn1	6.896				6.886	7.192
10Nh2	Hn2	7.431				7.403	7.541
11M0e	Hdg			3.785		3.681	
11M0e	Hdk			4.254	4.502	3.800	
11M0e	Hax			5.431	5.677	5.434	5.755
11M0e	Haq			4.096	4.340	3.679	
11M0e	Hau			7.001 7.387	7.121 7.527	6.888 7.404	7.009 7.523
11M0e	Has*			0.881	1.140	0.858	1.170
11M0e	Hap					4.702	
11M0e	Hat			6.024 6.407		6.063	6.239
11M0e	Har			3.514	3.755	3.524	3.760
11M0e	H1			4.239	4.501	4.438	4.447
11M0e	H2			3.359	3.629	3.236	
11M0e	H3			3.265	3.550		
11M0e	Hn2			7.844	7.891	7.878	7.959
11M0e	Hah*			1.699	1.951	1.708	1.946
11M0e	H5			3.270		3.371	3.644
11M0e	H6			3.786	4.036	3.796	4.029
11M0e	Hbj			4.166	4.420	4.174	
11M0e	Hbn			3.146	3.379	3.279	3.375
11M0e	Hbo			3.596 3.605	3.839 3.848	3.749	
11M0e	Hbm			3.057	3.318	3.205	
11M0e	Hbl			3.199	3.458	3.220	
11M0e	Hbk			2.964	3.220		
11M0e	H4					3.225	
11M0e	Hbu			4.248	4.505	4.302	
11M0e	Hbv			3.418	3.666	3.578	3.650
11M0e	Hbw			3.315	3.573	3.461	3.520
11M0e	Hncc			8.268	8.322	8.205	8.262
11M0e	Hcb*			1.732	1.992	1.723	1.946
11M0e	Hby			3.304			
11M0e	Hbz*			1.056	1.312	1.071	1.289
11M0e	Hbx			3.226	3.472	3.225	3.422
11M0e	Hch			4.293	4.536	4.257	4.440
11M0e	Hci			3.256		3.461	3.512
11M0e	Hcj			3.421	3.658	3.379	3.655
11M0e	Hck			3.976		3.410	
11M0e	Hcl			3.974	4.222	3.928	4.197
11M0e	Hncs			7.795		7.740	7.950
11M0e	Hcw			2.072		2.250	2.304
11M0e	Hcx			2.072		2.250	2.304
11M0e	H31						

11M0e	H32			5.074	5.313	5.212	5.167
11M0e	H34*			1.417	1.651	1.408	1.643
11M0e	H35			1.838	2.083	1.959	
11M0e	H36			1.803	2.039	1.950	
11M0e	H7			4.981	5.216	5.103	
11M0e	H8			5.085	5.304	5.005	
11M0e	H10*			0.616	0.844	0.421	0.647
11M0e	H11*			0.616	0.844	0.486	0.647
11M0e	H12			1.172	1.447	1.057	1.385
11M0e	H13			1.288	1.528	1.534	1.574
11M0e	H15			4.430	4.643	4.568	
11M0e	H16			2.422	2.512	1.917 2.040	2.374
11M0e	H17			4.874	5.168	4.861	4.991
11M0e	H19*			1.594	1.737	1.324	1.752
11M0e	H20			1.739	1.960	1.723	
11M0e	H21			1.806		1.725	
11M0e	H22			5.079			
11M0e	H24*			1.051	1.236	1.236	1.296
11M0e	H25*			1.590	1.723	1.255	1.746

Table 3. Observed chemical shifts of peptide/MoeA complex constituents at 4°C and 37°C.

Sr. No.	Particulars	Value
1.	*Pairwise RMSD of complex components*	
	Peptide	
	Heavy atom (Å)	0.91 ± 0.26
	All Atoms (Å)	1.18 ± 0.32
	Moenomycin A	
	Core atoms[VII] (Å)	1.48 ± 0.23
	Heavy atom (Å)	1.81 ± 0.43
	All atoms (Å)	2.04 ± 0.39
2.	*Experimental restraints*	
	Numbers of experimental restraints	
	Peptide	
	Intraresidue	212
	Sequential	82
	Total	397
	Moenomycin A	
	Total	112
	Complex	
	Intermolecular	98
	Total	607
	Restraint violations	
	NOE violations > 0.5 Å	2
	NOE violations > 0.3 Å	1
	NOE violations > 0.1 Å	3
3.	*RMSD from idealized covalent geometry*	
	Bond lengths (Å)	0.01 ± 0.00
	Bond angles (°)	1.41 ± 0.05
	Ramachandran analysis (determined by PROCHECK[VIII])	
	Residues in allowed regions (%)	93.9
	Residues in generously allowed regions (%)	6.1

[VII] Core atoms are well ordered moieties. These are: the heavy atoms of the peptide; and sugars C, D, E and F and the lipid tail of MoeA. Sugars A and B were found to be disordered.

[VIII] Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**:477-486

Sr. No.	Particulars	Value
4.	*Statistics of overall structural quality*	
	PROCHECK G-factors	
	Dihedrals	-0.71 ± 0.06
	Covalent	0.35 ± 0.04
	Overall	-0.33 ± 0.05
	ProSA[IX] *Z*-score	0.81 ± 0.08
	MolProbity[X] clash score	4.82 ± 0.09

Table 4: Structural Statistics of peptide/MoeA complex ensemble

[IX] Wiederstein M, Sippl MJ (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucl Acids Res* **35**:W407-W410

[X] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D* **66**:12-21

EXPERIMENTAL SECTION FOR CHAPTER 7

I. Materials

All amino acids, Fmoc-Ɛ-Ahx-OH and oxyma pure were purchased from Merck Millipore. Fmoc-PEG_3-OH (16 atoms) was purchased from Novabiochem. Succinic anhydride was purchased from Alfa Aesar. 5-Amino-2-Nitrobenzoic acid, Diisoproplycarbodiimide, Trifluoroacetic acid (TFA) and Triisopropylsilane were purchased from Fluorochem, UK. L-amino acids were used throughout the syntheses. The protecting groups for the amino acids are ^{t}Bu for Glu, Boc for Pro, Tyr, Lys, Trp, Pbf for Arg and Trt for Gln. 2,2'-(Ethylenedioxy)bis(ethylamine), Diisopropylethylamine, supplied as extra dry, redistilled, 99.5 % pure, piperidine and NH_4HCO_3 were purchased from Sigma Aldrich. Dimmethylformamide (DMF) peptide synthesis grade was purchased from Rathburn chemicals. Diethyl ether, i-PrOH, MeOH (HPLC grade), Dichloromethane (Analytical grade), 35% (S.G. 0.88) NH_4OH solution and Acetonitrile (HPLC grade) were purchased from Fisher Scientific. Water with the Milli-Q grade standard was obtained in-house from an ELGA Purelab Flex system. Rink-Amide ChemMatrix resin (100-200 μm, manufacturer's loading: 0.49 mmol/g) was obtained from Biotage. All chemicals were used without further purification. Flavomycin, containing 12% by weight of Moenomycin A was purchased from Biovet J.S. Co, Sofia, Bulgaria.

II. Isolation of Moenomycin A:

Moenomycin A was isolated using the procedure described in literature[I].

III. General procedure for peptide syntheses:

Peptide syntheses was performed using standard Fmoc Solid Phase Peptide Synthesis (SPPS) protocols on Rink Amide Chemmatrix Resin, loading = 0.49 mmol/g on a 0.1 mmol scale using a Biotage Initiator + Alstra fully automated microwave peptide synthesizer. All amino acid couplings were performed using 5 eq. Amino Acid with 5 eq. DIC/Oxyma in DMF as a coupling cocktail by irradiating at 70°C for 5 min. Fmoc deprotection was performed using 20% piperidine in DMF by irradiating at 70°C for 3 min, followed by shaking at r.t. for 10 min. 4 x 45s washes were performed after each coupling cycle and 3 x 30s washes were performed after each deprotection cycle.

Fmoc-Ɛ-Ahx-OH, Fmoc-PEG_3-OH linker used in the peptides were coupled using 3 eq. of linker, 2.9 eq. of HATU and 6 eq. of DIPEA. 5-Amino-2-Nitrobenzoic acid was coupled using 2 x 1h couplings using 3 eq. of linker, 2.9 eq. of HATU, 3 eq. of DMAP and 6 eq. of DIPEA.

Peptide cleavage was performed using TFA/TIS/H_2O = 95:2.5:2.5 (3 mL/100 mg resin). For sequences containing 1 or 2 Arg groups, the cleavage time was 2 h. For 3 or higher Arg containing peptides, cleavage time was 5h. Peptides were precipitated using cold Et_2O (-20°C) by adding approximately 5x volume of the TFA used for cleavage and centrifuging at 7000 rpm at 0°C.

IV. General procedure for syntheses of peptide-MoeA covalent conjugates

IVa. For completely water soluble peptides (15 mg scale):

[I] Adachi, M.; Zhang, Y.; Leimkuhler, C.; Sun, B.; LaTour, J. V.; Kahne, D. E. *J. Am. Chem. Soc.* **2006**, *128*, 14012.

15 mg Moenomycin A together with 135 mg of NaOAc was weighed out in a clean 25 mL round bottom flask. 7 mL of milliQ water was added and the mixture was cooled down to 0°C. In an eppendorf tube 2 eq. of peptide was weighed out and dissolved in 800 µL 9% HCl and was cooled down to 0°C. 3 mg of $NaNO_2$ was dissolved in 300 µL milliQ water another Eppendorf and cooled to 0°C. The $NaNO_2$ was then added dropwise to the peptide whilst gently shaking the eppendorf. At this point, the yellow color of the peptide fades/disappears. The mixture was left standing for 5 min at 0°C and then added slowly to the 25 mL flask containing the Moenomycin. The yellow color of the peptide should now reappear. The reaction mixture was stirred for 4h and analyzed by RP-HPLC. If Moenomycin A was not completely consumed then the reaction mixture was left overnight.

IVb. For partially water soluble peptides to Moenomycin A (15 mg scale):

15 mg Moenomycin A together with 135 mg of NaOAc was weighed out in a clean 25 mL round bottom flask. 7 mL of milliQ water was added and the mixture was cooled down to 0°C. In a 10 mL falcon tube 2 eq. of peptide was weighed out and dissolved in 750 µL 9% HCl + 750 µL ACN and was cooled down to 0°C. 3 mg of $NaNO_2$ was dissolved in 300 µL milliQ water another Eppendorf and cooled to 0°C. The $NaNO_2$ was then added dropwise to the peptide whilst gently shaking the falcon tube. At this point, the yellow color of the peptide fades/disappears. The mixture was left standing for 5 min at 0°C and then added slowly to the 25 mL flask containing the Moenomycin. The yellow color of the peptide should now reappear. If the reaction mixture is cloudy add acetonitrile dropwise till the solution clears up. The reaction mixture was stirred for 4h and analysed by RP-HPLC. If Moenomycin A was not completely consumed then the reaction mixture was left overnight.

V. Analysis and purification of peptides/conjugates:

All peptides/conjugates were analysed on a Thermo Scientific Dionex Ultimate 3000 RP-HPLC equipped with a Phenomenex Gemini NX C18 110 Å (150 x 4.6 mm) column using the following buffer systems:

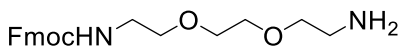
For peptides: Analysis of peptides was done using the conditions in chapter 6 using 1 mM Heptafluorobutyric acid buffer unless otherwise mentioned.

For conjugates: A: 0.1% HCOOH in milliQ water. B: ACN using a flow rate of 1 ml/min. The column was flushed with 100% A for 5 min prior to an injection and was flushed for 5 min with 95% B and 5% A after the run was finished. The following gradient was used: 100% A for 2 min. 0-95% B in 15 min. 95% B for 5 min. 100% A for 4 min.

Conjugates were purified using the same gradient as the analyses, on a Thermo Scientific Dionex Ultimate 3000 RP-HPLC with a flow rate of 5 mL/min using a Phenomenex Gemini NX C18 110 Å (150 x 10 mm) semi-prep column.

LC-MS data were collected on an Agilent 1100 Series instrument with a Phenomenex Kinetex C18 100Å column (150 x 4.6 mm, 5 µm at 35 °C) connected to an ESMSD type VL mass detector with a flow rate of 1.5 ml/min was used with the following solvent systems: (A): 0.1% HCOOH in H_2O and (B) MeCN. The column was flushed with 100% A for 2 min, then a gradient from 0 to 100% B over 6 min was used, followed by 2 min of flushing with 100% B.

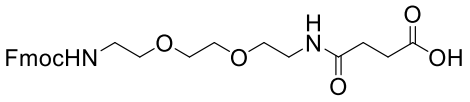
VI. Synthesis of a PEG_2 linker



A

To a cooled (0 °C) stirred solution of 2,2'-(ethane-1,2-diylbis(oxy))diethanamine (2.94 mL; 20.1 mmol) in DCM (200 mL) was added a solution of Fmoc-ONSu (1.35 g; 4 mmol) in DCM (2 mL) and the mixture stirred at 0 °C for 2 hours. The mixture was analysed by TLC using: pure DCM (to verify that all the Fmoc-ONSu was consumed) and 20% MeOH in DCM + 2% Et_3N (for identifying the product).

TLC was stained using ninhydrin. After reaction completion, the reaction mixture was washed successively with water (3 x 75 mL) before drying over Na_2SO_4. Removal of the solvent under vacuum afforded a brown oil **A**. Since the compound was found to be not stable under column chromatography conditions, no subsequent purification steps were performed.



B

To a stirred solution of crude (9H-fluoren-9-yl)methyl (2-(2-(2-aminoethoxy)ethoxy)ethyl)carbamate (**A**, 1.25 g; 3.36 mmol) in DCM (10 mL) was added succinic anhydride (0.772 g; 7.72 mmol). The reaction was stirred at room temperature for 3 hours. The mixture was analysed by TLC using: 20% MeOH in DCM + 0.5% AcOH. TLC was stained using PMA. Removal of the solvent under vacuum yielded a brown solid. The brown solid was purified using silica gel column chromatography using 10% MeOH in DCM + 0.5% AcOH to obtain the pure product **B** as colourless oil (881 mg; 44% overall yield from 2 steps).

^{1}H NMR ($CDCl_3$, 500MHz): δ = 7.73 (d, *J* = 7.3 Hz, 2 H), 7.59 (d, *J* = 7.3 Hz, 2 H), 7.36 (t, *J* = 7.5 Hz, 2 H), 7.23 - 7.32 (m, 2 H), 4.37 (d, *J* = 7.0 Hz, 2 H), 4.19 (br. s., 1 H), 3.44 - 3.60 (8 H), 3.21 - 3.44 (4 H), 2.54 - 2.64 (3 H), 2.42 - 2.52 ppm (2 H) (Figure 1)

^{13}C NMR ($CDCl_3$, 125 MHz) δ = 29.62, 31.77, 39.20, 40.81, 45.11, 47.18, 51.59, 66.54, 69.69, 70.08, 70.16, 119.93, 125.10, 127.05, 127.67, 141.21, 143.94, 156.69, 173.04, 173.50 (Figure 2)

Exact mass calcd. for $C_{25}H_{30}N_2O_7$ = 470.21, found M + H^+ = 471.1 (Figure 3)

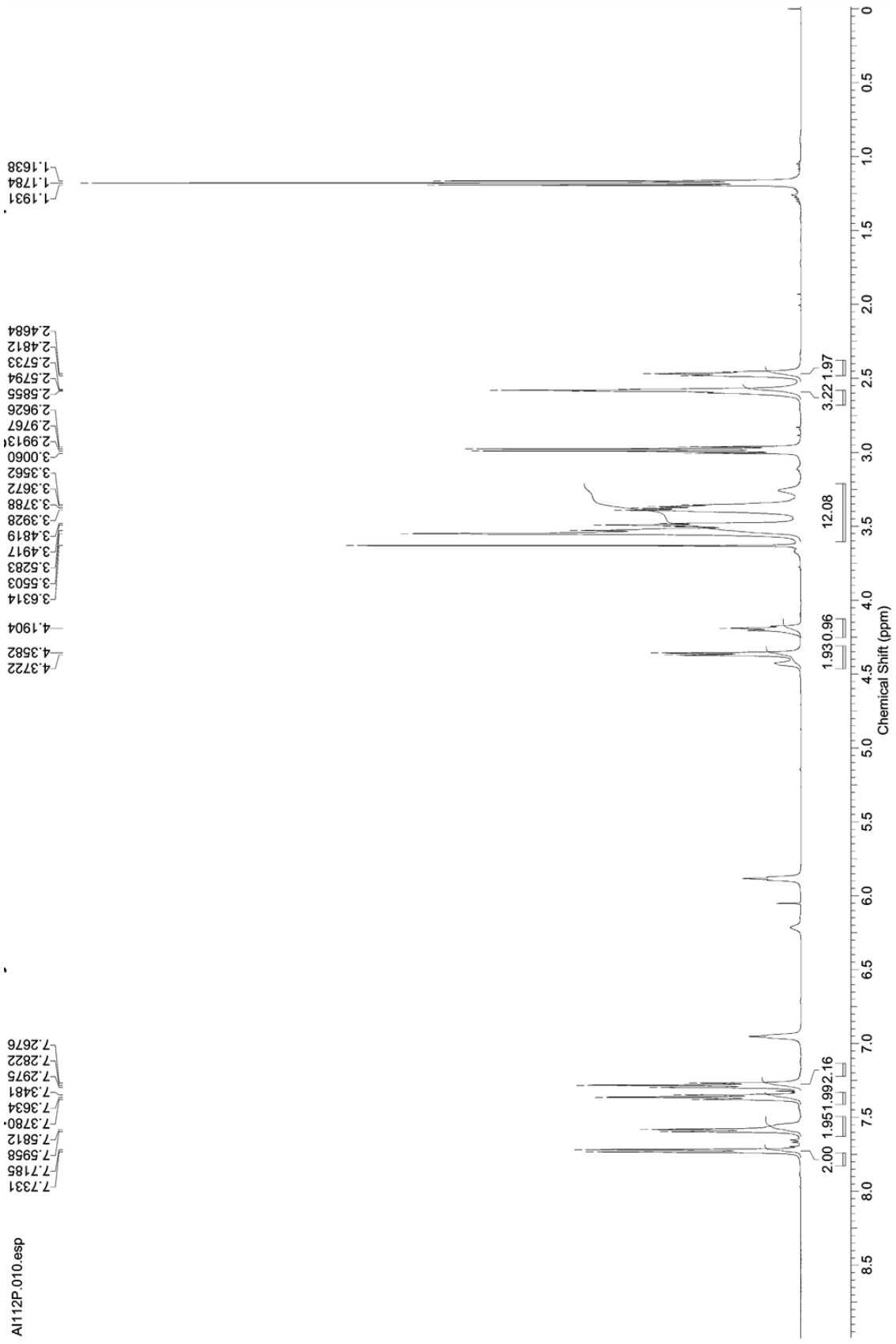


Figure 1: 1H NMR spectra of **B**.

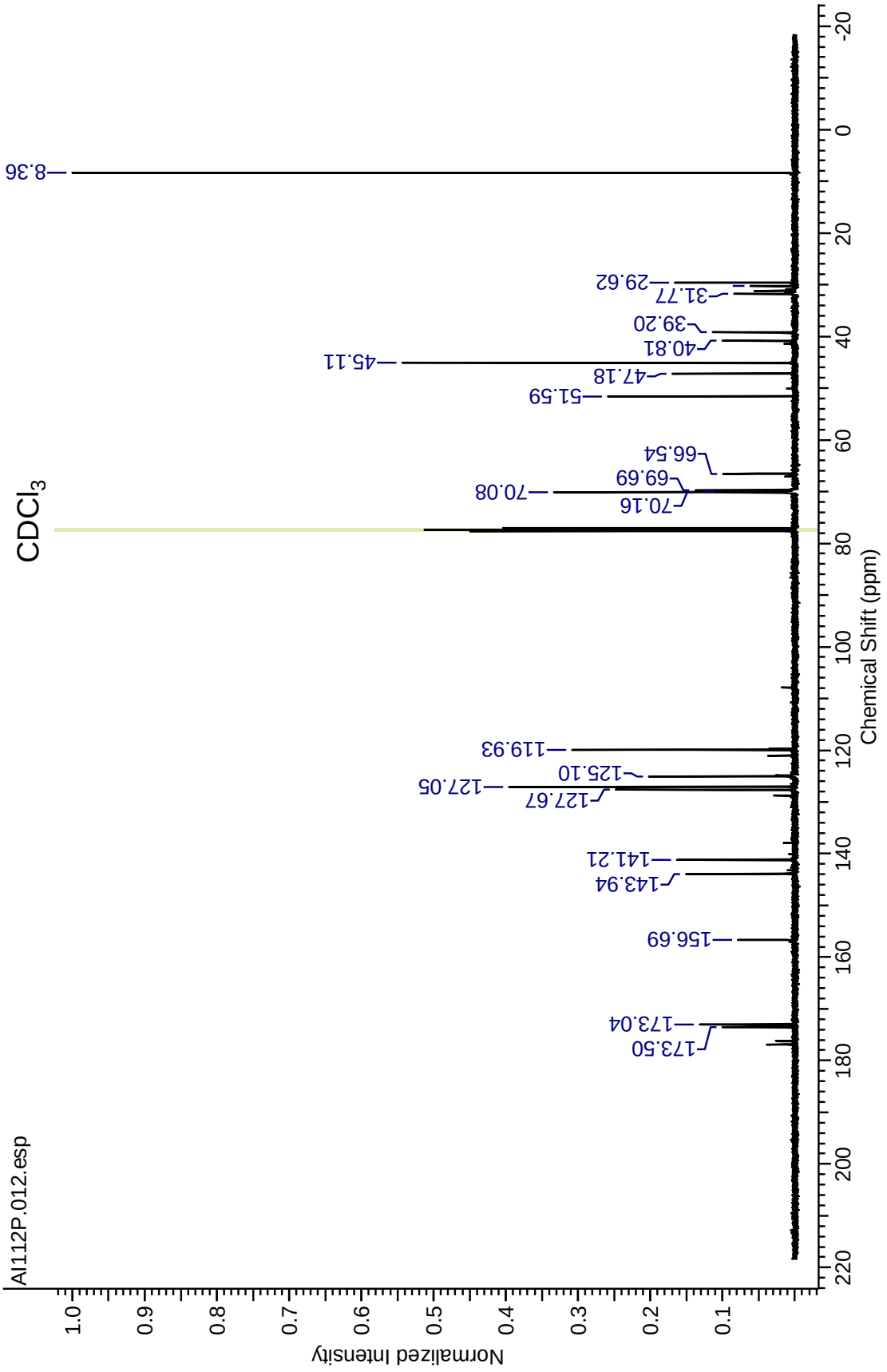


Figure 2: 13C NMR spectra of **B**.

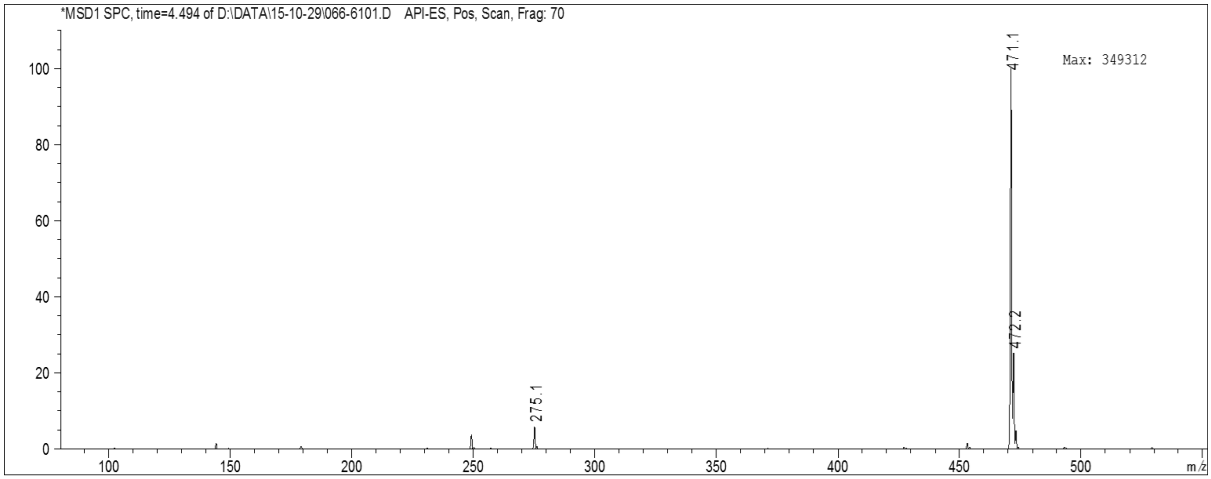


Figure 3: ESI-MS of compound **B**. Exact mass calcd. for $C_{25}H_{30}N_2O_7$ = 470.21, found M + H^+ = 471.1

VII. Complete list of peptides/conjugates synthesized

Complete list of peptides conjugated to Moenomycin A:

Sr. No.	Code	Peptide sequence	Chemical Formula	Exact Mass	Mass found
1.	1	X-Ahx-RRRR-$CONH_2$	$C_{37}H_{66}N_{20}O_8$	918.54	919.54
2.	2	X-RRRRRRRR-$CONH_2$	$C_{55}H_{103}N_{35}O_{11}$	1429.86	1430.47
3.	3	X-Ahx-RRRRRRRR-$CONH_2$	$C_{61}H_{114}N_{36}O_{12}$	1542.94	1543.53
4.	4	X-$(Ahx)_2$-RRRRRRRR-$CONH_2$	$C_{67}H_{125}N_{37}O_{13}$	1656.03	1656.38
5.	5	X-(Ahx)-RRRRRRRRR-$CONH_2$	$C_{67}H_{126}N_{40}O_{13}$	1699.04	1699.43
6.	6	X-(Ahx)-RRRRRRRRRR-$CONH_2$	$C_{73}H_{138}N_{44}O_{14}$	1855.14	1855.61
7.	7	X-Ahx-KKRRRRRRRR-$CONH_2$	$C_{73}H_{138}N_{40}O_{14}$	1799.13	1799.52
8.	8	X-KRRKRRKRR-$CONH_2$	$C_{61}H_{115}N_{33}O_{12}$	1501.94	1502.51
9.	9	X-Ahx-KRRKRRKRR-$CONH_2$	$C_{67}H_{126}N_{34}O_{13}$	1615.02	1615.40
10.	10	X-$(Ahx)_2$-KRRKRRKRR-$CONH_2$	$C_{73}H_{137}N_{35}O_{14}$	1728.11	1728.76
11.	11	X-KKKKKR-$CONH_2$	$C_{43}H_{79}N_{17}O_9$	977.62	978.50 $[M+H^+]$
12.	12	X-Ahx-KKKKKR-$CONH_2$	$C_{49}H_{90}N_{18}O_{10}$	1090.71	1090.6
13.	13	X-$(Ahx)_2$-KKKKKR-$CONH_2$	$C_{55}H_{101}N_{19}O_{11}$	1203.79	1204.00
14.	14.	X-Ahx-RRWWRRWRR-$CONH_2$	$C_{82}H_{120}N_{34}O_{13}$	1788.98	1789.47
15.	15	X-R-$CONH_2$	$C_{13}H_{19}N_7O_4$	337.15	338.1 $[M+ H^+]$
16.	16	X-$(PEG)_3$-KKKKKR-$CONH_2$	$C_{54}H_{100}N_{18}O_{14}$	1224.77	1226.5 $[M+ H^+]$
17.	17	X-$(PEG)_2$-KKKKKR-$CONH_2$	$C_{53}H_{97}N_{19}O_{13}$	1207.75	1208.08
18.	18	X-$(PEG)_3$-RRRRRRRR-$CONH_2$	$C_{66}H_{124}N_{36}O_{16}$	1677.00	1677.50
19.	19	X-$(PEG)_2$-RRRRRRRR-$CONH_2$	$C_{65}H_{121}N_{37}O_{15}$	1659.98	1660.43
20.	20	X-$(PEG)_3$-RRRR-$CONH_2$	$C_{42}H_{76}N_{20}O_{12}$	1052.59	
21.	21	X-$(PEG)_3$-RRRRR-$CONH_2$	$C_{48}H_{88}N_{24}O_{13}$	1208.70	1209.10
22.	22	X-$(PEG)_2$-RRRR-$CONH_2$	$C_{41}H_{73}N_{21}O_{11}$	1035.58	1035.89

Table 1: Complete list of peptides synthesized

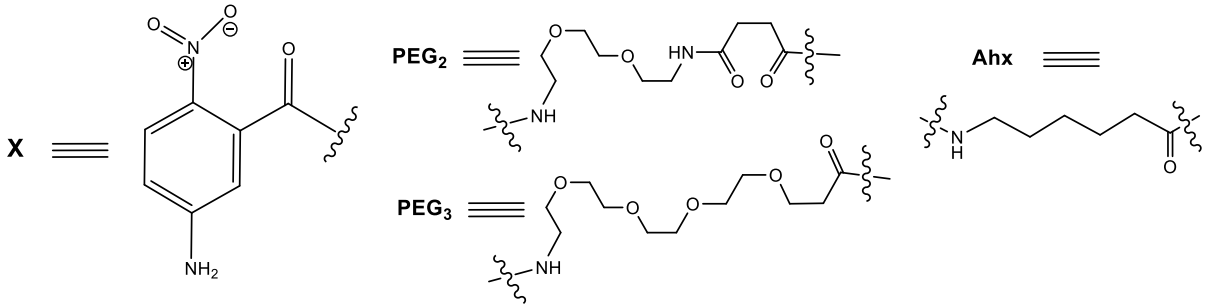
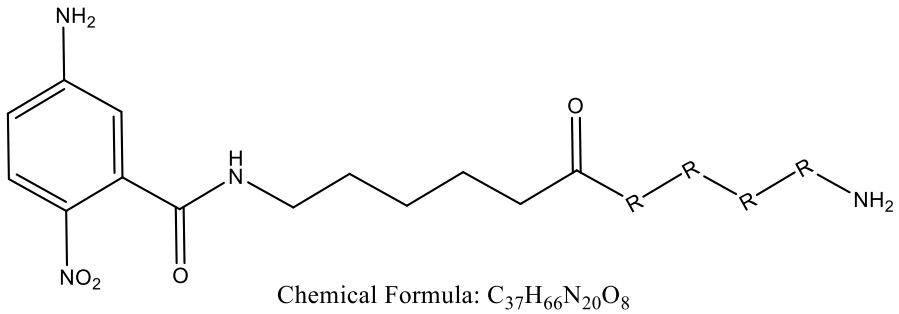


Figure 4: Structure of 5-amino-2-nitrobenzoic acid, PEG_2 linker, PEG_3 linker and 6-Aminohexanoic acid linker (Ahx).

Sr. No.	Code	Conjugate description	Chemical Formula	Exact Mass	Mass found
1.	23	Conjugate with peptide 1	$C_{106}H_{171}N_{26}O_{42}P$	2511.18	2512.2
2.	24	Conjugate with peptide 2	$C_{124}H_{208}N_{41}O_{45}P$	3022.50	3023.35
3.	25	Conjugate with peptide 3	$C_{130}H_{219}N_{42}O_{46}P$	3135.58	3137.1
4.	26	Conjugate with peptide 4	$C_{136}H_{230}N_{43}O_{47}P$	3248.67	3249.46
5.	27	Conjugate with peptide 5	$C_{136}H_{231}N_{46}O_{47}P$	3291.68	3292.46
6.	28	Conjugate with peptide 6	$C_{142}H_{243}N_{50}O_{48}P$	3447.78	3448.61
7.	29	Conjugate with peptide 7	$C_{136}H_{231}N_{40}O_{47}P$	3207.67	3208.34
8.	30	Conjugate with peptide 8	$C_{130}H_{220}N_{39}O_{46}P$	3094.58	3095.52
9.	31	Conjugate with peptide 9	$C_{142}H_{243}N_{46}O_{48}P$	3391.77	3392.52
10.	32	Conjugate with peptide 10	$C_{142}H_{242}N_{41}O_{48}P$	3320.75	3321.82
11.	33	Conjugate with peptide 11	$C_{112}H_{184}N_{23}O_{43}P$	2570.27	2571.00
12.	34	Conjugate with peptide 12	$C_{118}H_{195}N_{2}4O_{44}P$	2683.35	2683.99
13.	35	Conjugate with peptide 13	$C_{124}H_{206}N_{25}O_{45}P$	2796.43	2797.00
14.	36	Conjugate with peptide 14	$C_{151}H_{225}N_{40}O_{47}P$	3381.62	3383.55
15.	37	Conjugate with peptide 15	$C_{82}H_{124}N_{13}O_{38}P$	1929.79	1930.65
16.	38	Conjugate with peptide 16	$C1_{23}H_{205}N_{24}O_{48}P$	2817.41	2818.11
17.	39	Conjugate with peptide 17	$C_{122}H_{202}N_{25}O_{47}P$	2800.39	2801.07
18.	40	Conjugate with peptide 18	$C_{135}H_{229}N_{42}O_{50}P$	3269.64	3270.46
19.	41	Conjugate with peptide 19	$C_{134}H_{226}N_{43}O_{49}P$	3252.63	3253.47
20.	42	Conjugate with peptide 20	$C_{117}H_{193}N_{30}O_{47}P$	2801.34	2801.99
21.	43	Conjugate with peptide 21	$C_{111}H_{181}N_{26}O_{46}P$	2645.24	2645.85
22.	44	Conjugate with peptide 22	$C_{110}H_{178}N_{27}O_{45}P$	2628.22	2628.86

Table 2: Complete list of conjugates synthesized

VIII. HPLC/LC-MS/ESI-MS Data



Chemical Formula: $C_{37}H_{66}N_{20}O_8$
Exact Mass: 918.54
Molecular Weight: 919.07

Figure 5: Structure of peptide **1**

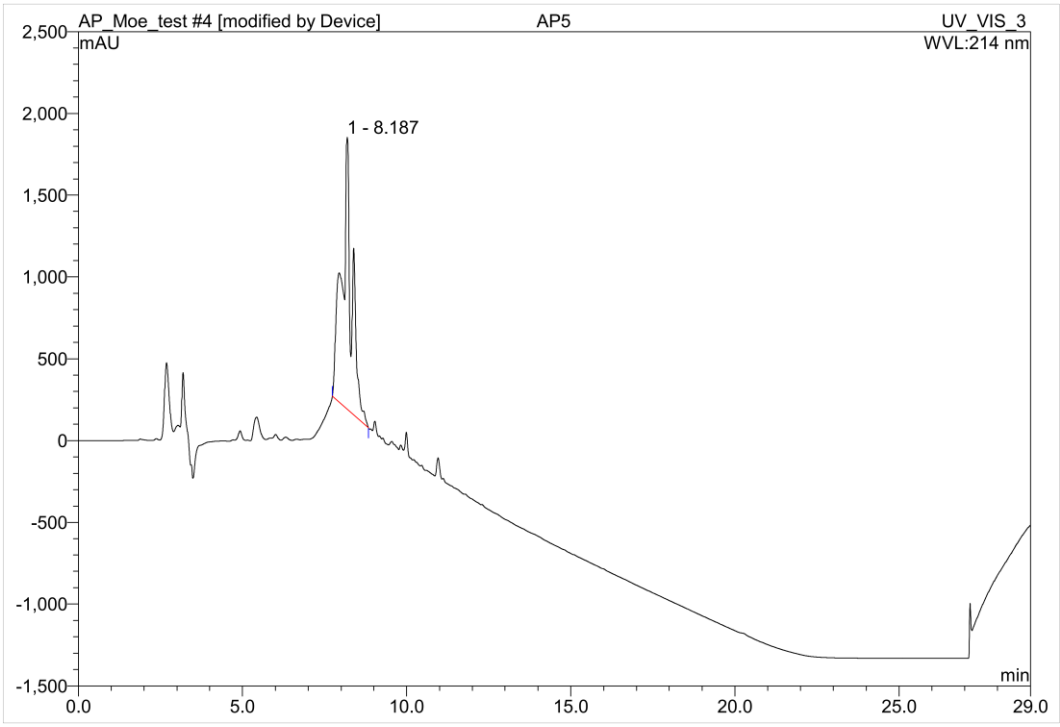


Figure 6: HPLC trace of crude peptide **1**, showing product peak at r.t. = 8.187 min. Gradient 0-95% ACN in 15 min.

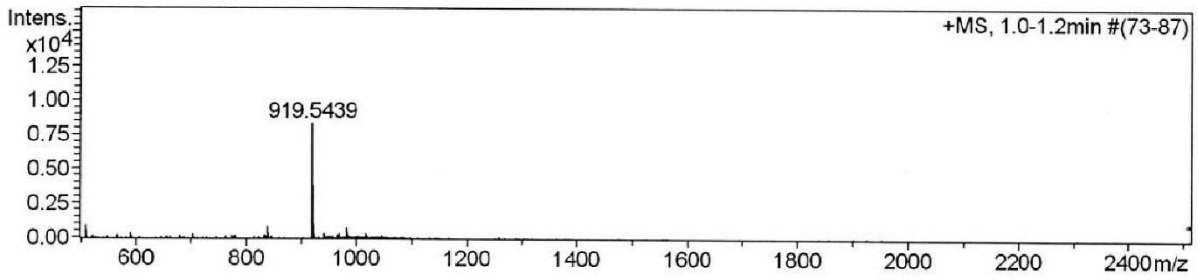
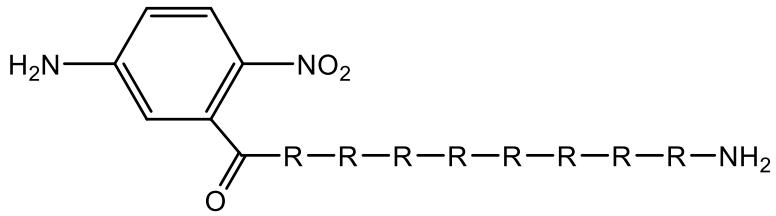


Figure 7: ESI-MS of peptide **1**. Exact mass calcd. for $C_{37}H_{66}N_{20}O_8$ = 918.54, found M + H^+ = 919.54.



Chemical Formula: $C_{55}H_{103}N_{35}O_{11}$
Exact Mass: 1429.86
Molecular Weight: 1430.66

Figure 8: Structure of peptide **2**

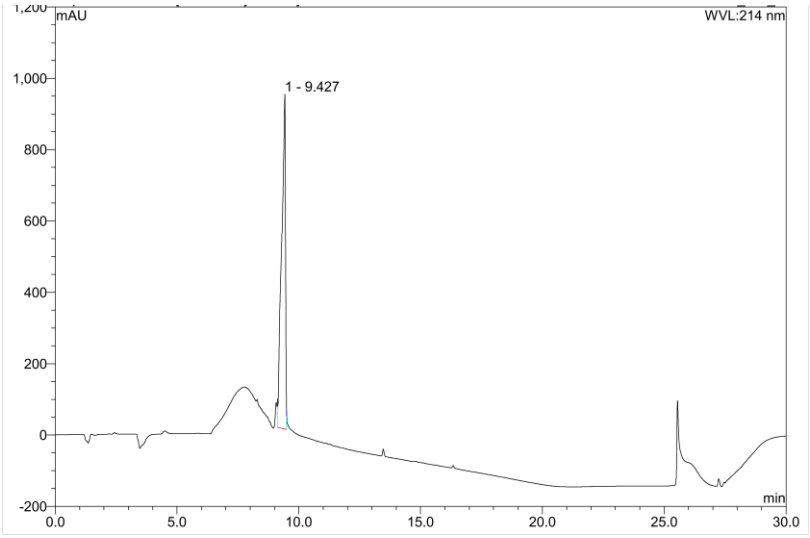


Figure 9: HPLC trace of crude peptide **2**, showing product peak at r.t. = 9.427 min. Gradient 0-95% ACN in 15 min.

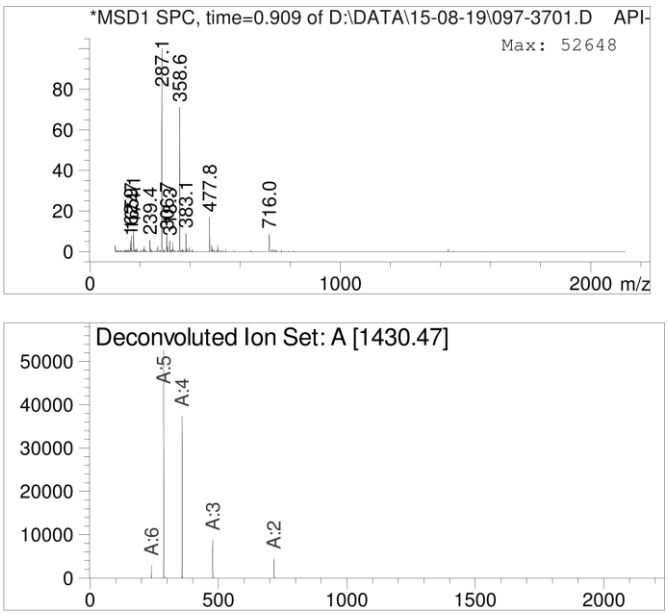
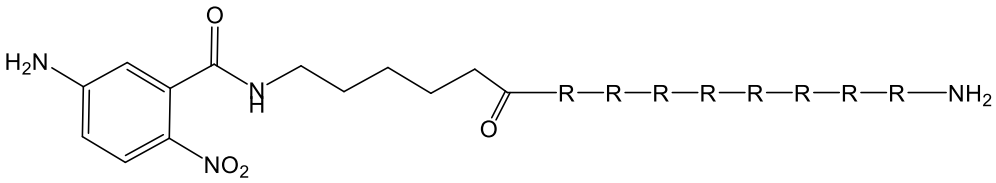


Figure 10: ESI-MS with deconvolution of peptide **2**. Exact mass calcd. for $C_{55}H_{103}N_{35}O_{11}$ = 1429.86, deconvoluted mass found = 1430.47.



Chemical Formula: $C_{61}H_{114}N_{36}O_{12}$
Exact Mass: 1542.94
Molecular Weight: 1543.82

Figure 11: Structure of peptide **3**

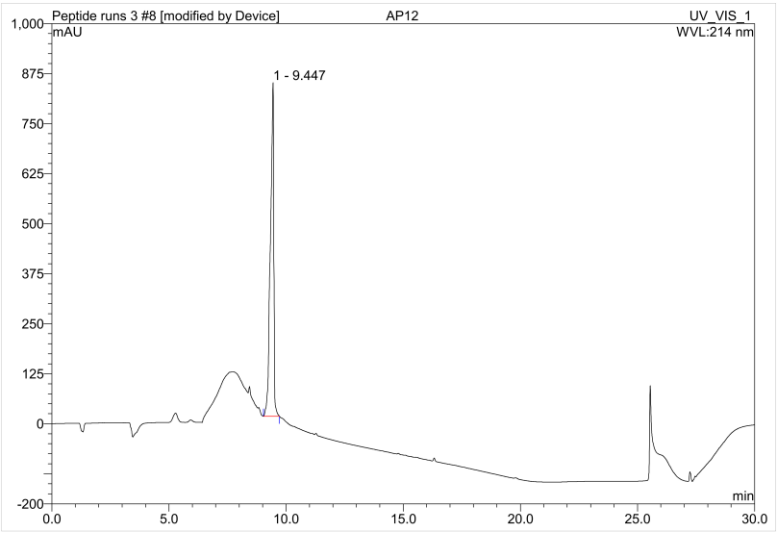


Figure 12: HPLC trace of crude peptide **3**, showing product peak at r.t. = 9.447 min. Gradient 0-95% ACN in 15 min.

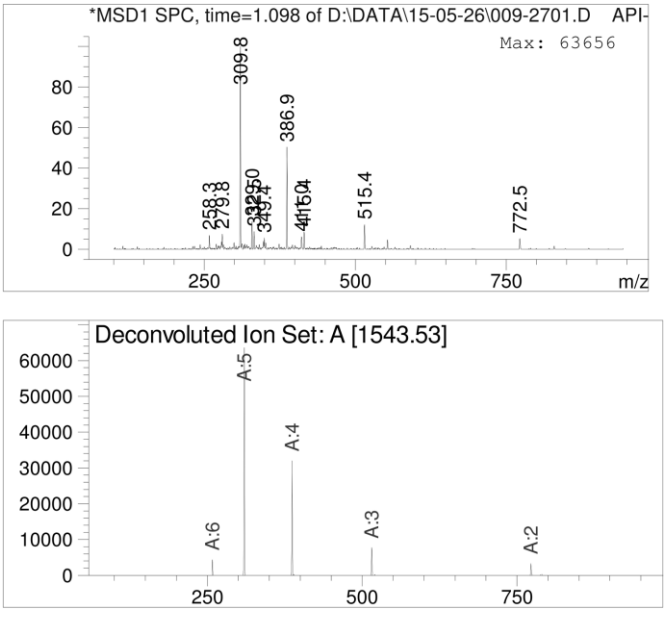
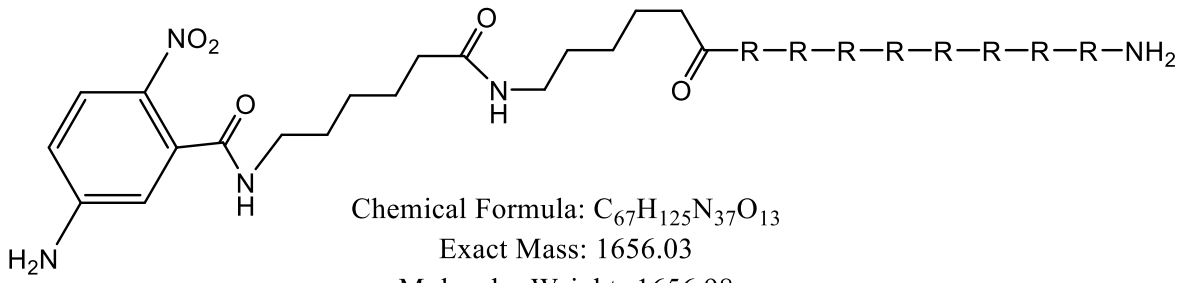


Figure 13: ESI-MS from LC-MS with deconvolution of peptide **3**. Exact mass calcd. for $C_{61}H_{114}N_{36}O_{12}$ = 1542.94, deconvoluted mass found = 1543.53.



Chemical Formula: $C_{67}H_{125}N_{37}O_{13}$
Exact Mass: 1656.03
Molecular Weight: 1656.98

Figure 14: Structure of peptide **4**

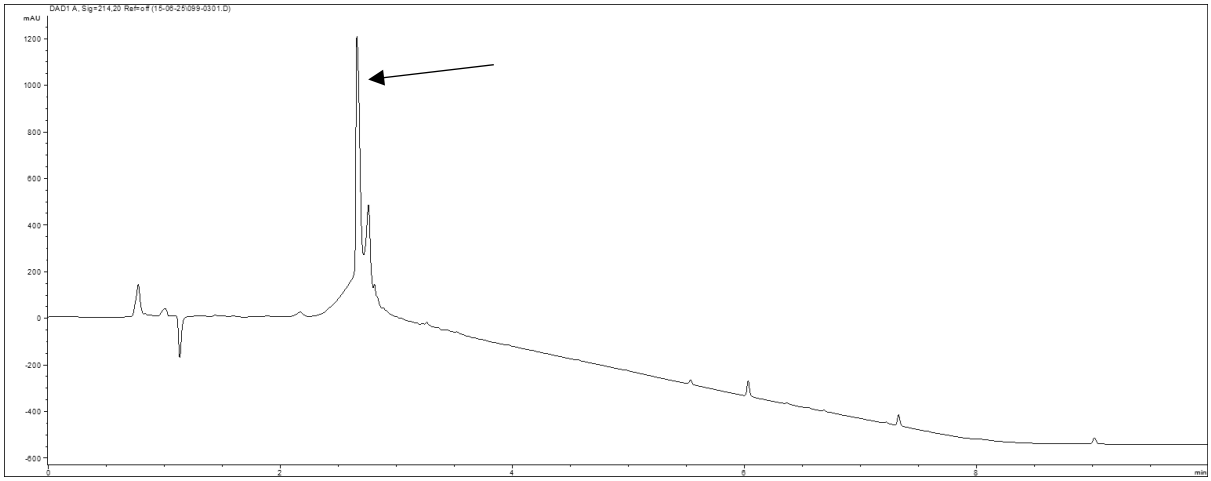


Figure 15: HPLC trace of crude peptide **4**, showing product peak at r.t. = 2.682 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.

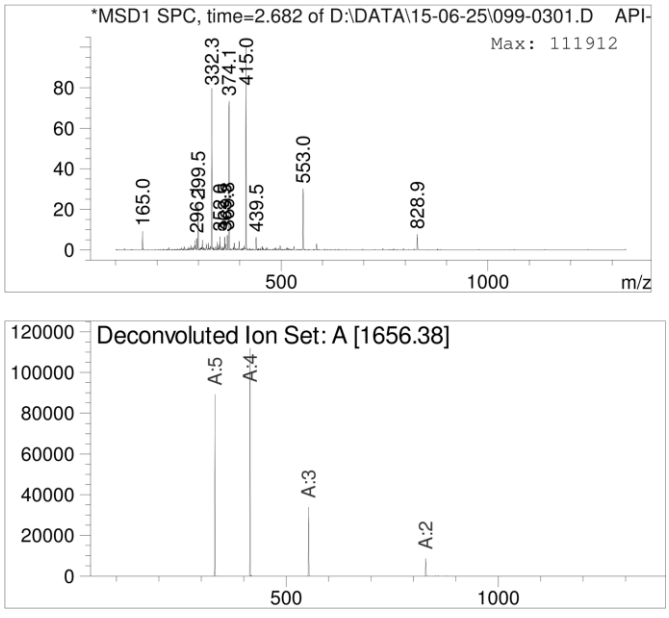
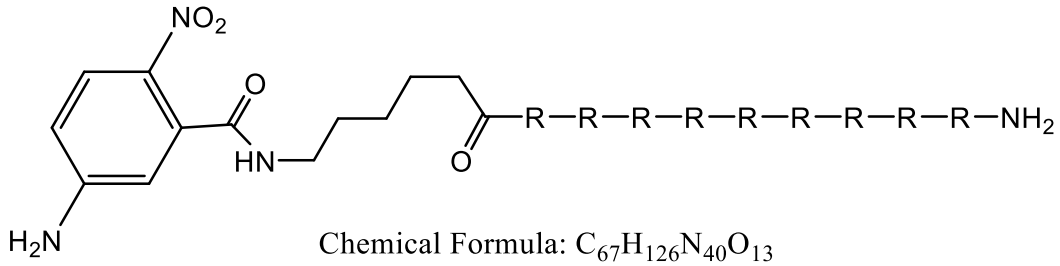


Figure 16: ESI-MS from LC-MS with deconvolution of peptide **4**. Exact mass calcd. for $C_{67}H_{125}N_{37}O_{13}$ = 1656.03, deconvoluted mass found = 1656.38.



Chemical Formula: $C_{67}H_{126}N_{40}O_{13}$
Exact Mass: 1699.04
Molecular Weight: 1700.01

Figure 17: Structure of peptide **5**

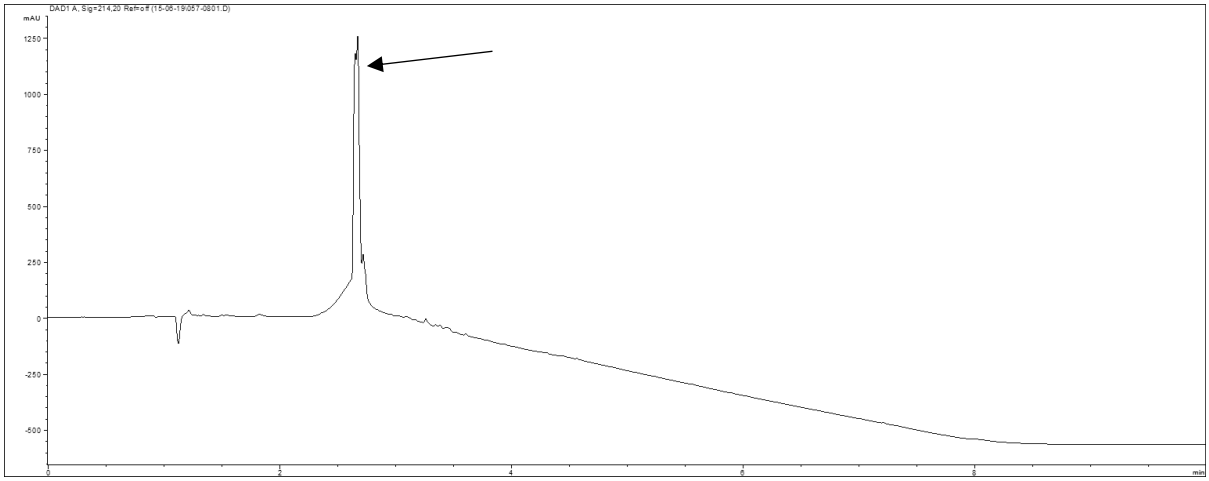


Figure 18: HPLC trace of crude peptide **5**, showing product peak at r.t. = 2.665 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.

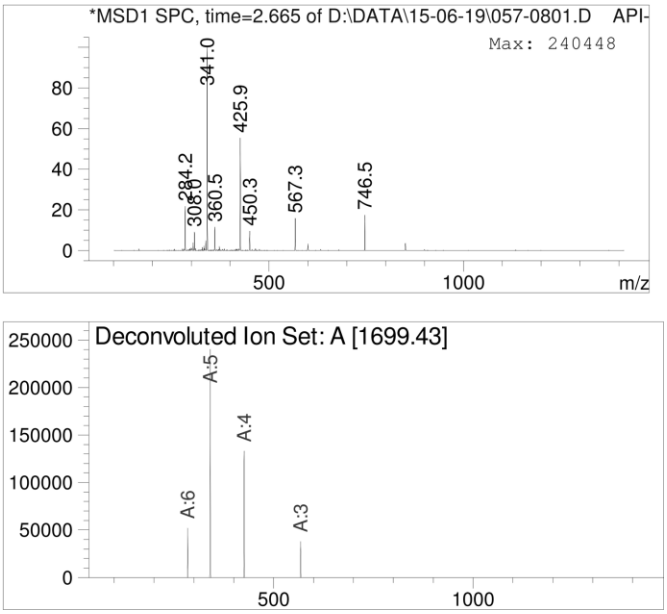
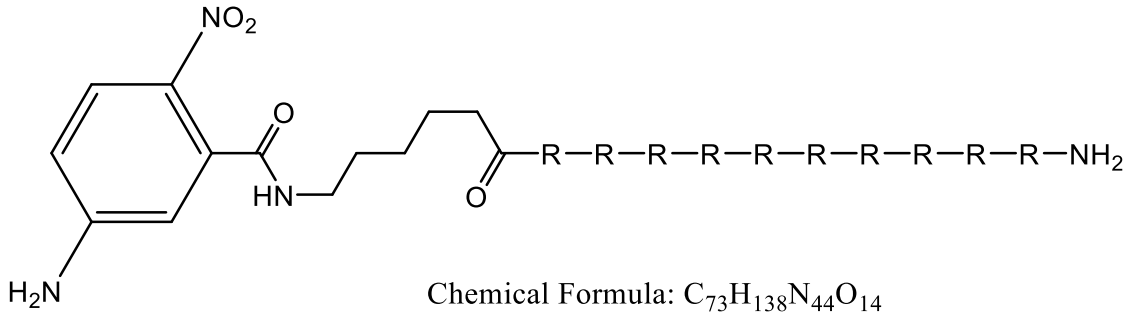


Figure 19: ESI-MS from LC-MS with deconvolution of peptide **5**. Exact mass calcd. for $C_{67}H_{126}N_{40}O_{13}$ = 1699.04, deconvoluted mass found = 1699.43.



Chemical Formula: $C_{73}H_{138}N_{44}O_{14}$
Exact Mass: 1855.14
Molecular Weight: 1856.16

Figure 20: Structure of peptide **6**

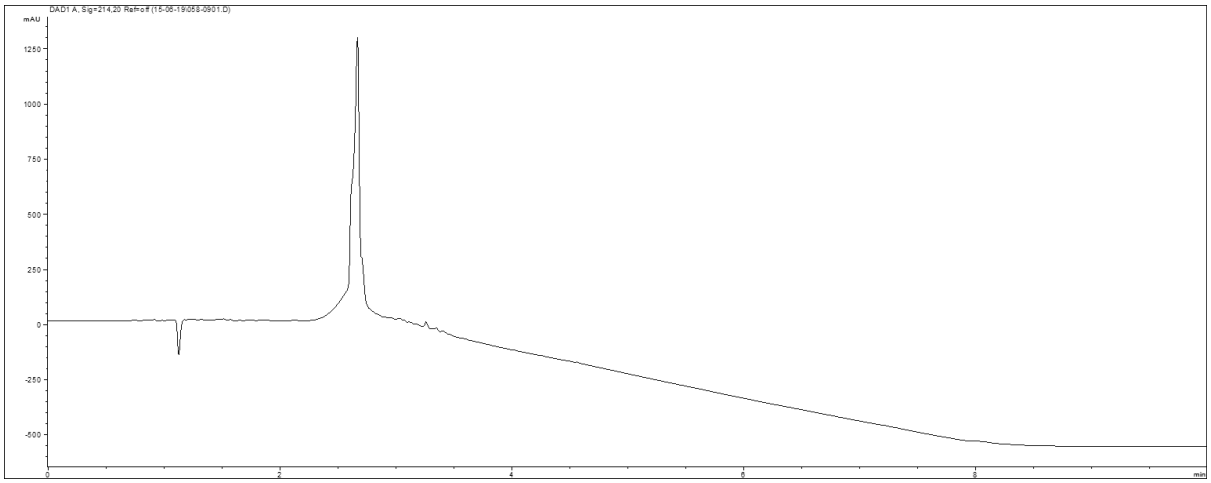


Figure 21: HPLC trace of crude peptide **6**, showing product peak at r.t. = 2.657 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.

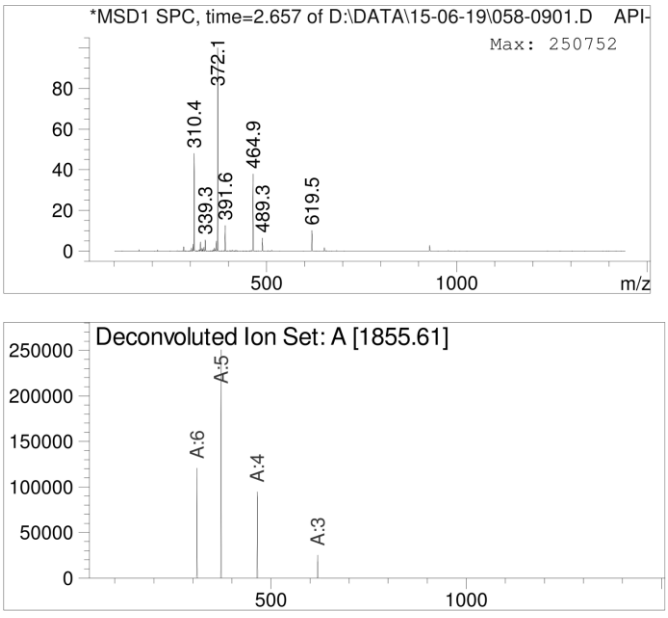
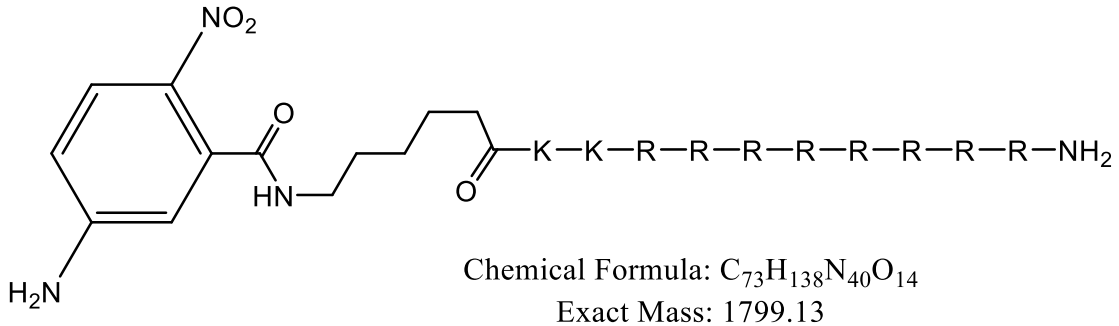


Figure 22: ESI-MS from LC-MS with deconvolution of peptide **6**. Exact mass calcd. for $C_{73}H_{138}N_{44}O_{14}$ = 1855.14, deconvoluted mass found = 1855.61.



Chemical Formula: $C_{73}H_{138}N_{40}O_{14}$
Exact Mass: 1799.13
Molecular Weight: 1800.14

Figure 23: Structure of peptide **7**

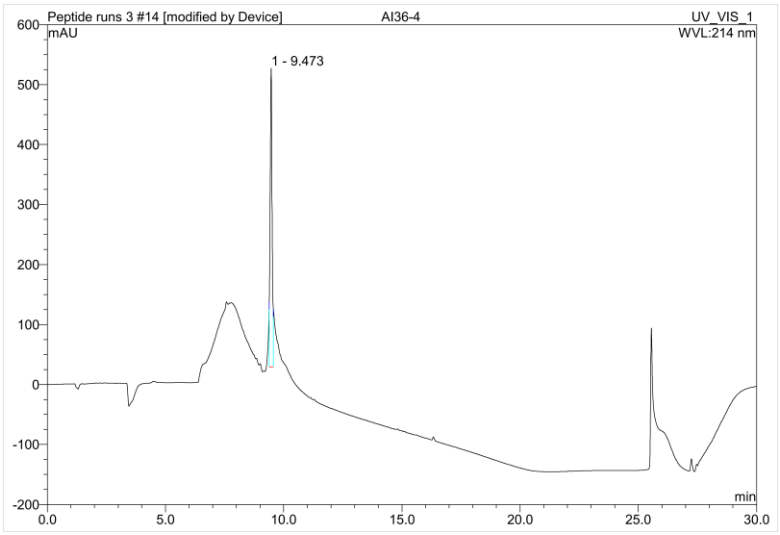


Figure 24: HPLC trace of crude peptide **7**, showing product peak at r.t. = 9.473 min. Gradient 0-95% ACN in 15 min.

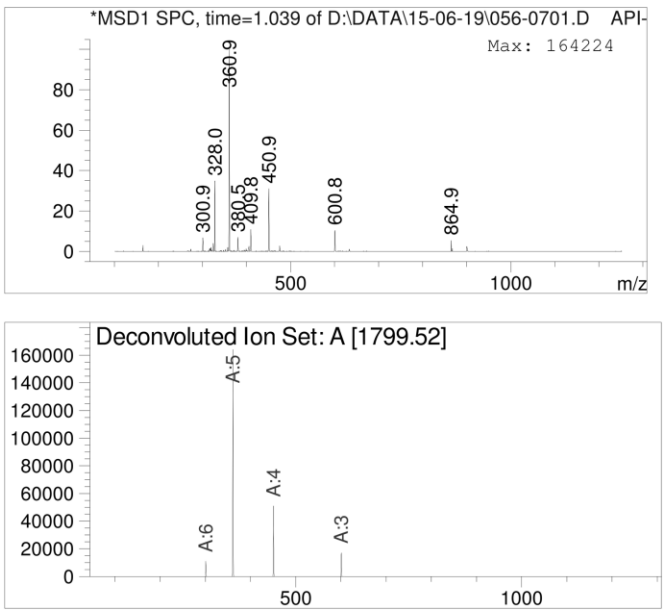
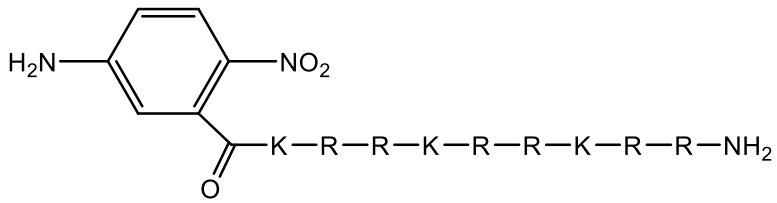


Figure 25: ESI-MS from LC-MS with deconvolution of peptide **7**. Exact mass calcd. for $C_{73}H_{138}N_{40}O_{14}$ = 1799.13, deconvoluted mass found = 1799.52.



Chemical Formula: $C_{61}H_{115}N_{33}O_{12}$
Exact Mass: 1501.94
Molecular Weight: 1502.81

Figure 26: Structure of peptide **8**

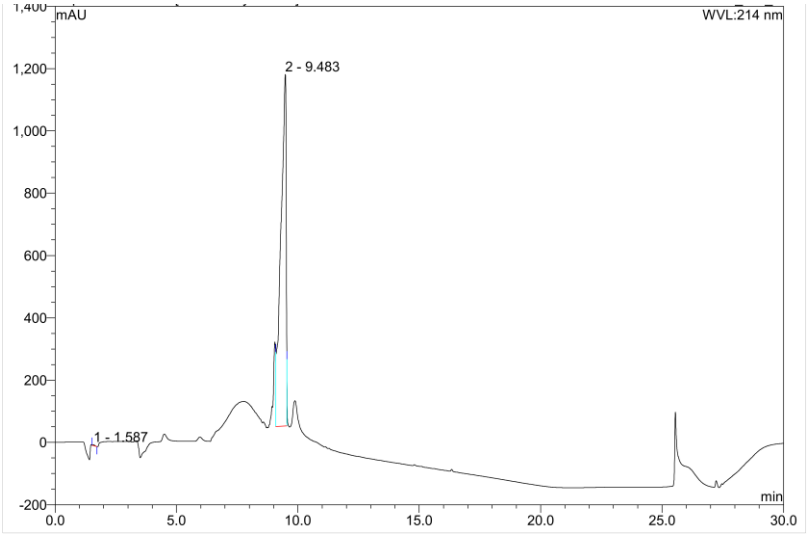


Figure 27: HPLC trace of crude peptide **8**, showing product peak at r.t. = 9.483 min. Gradient 0-95% ACN in 15 min.

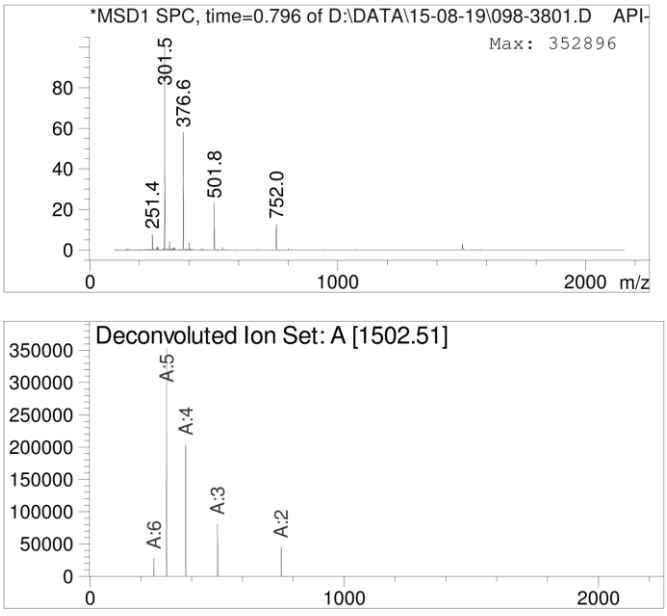
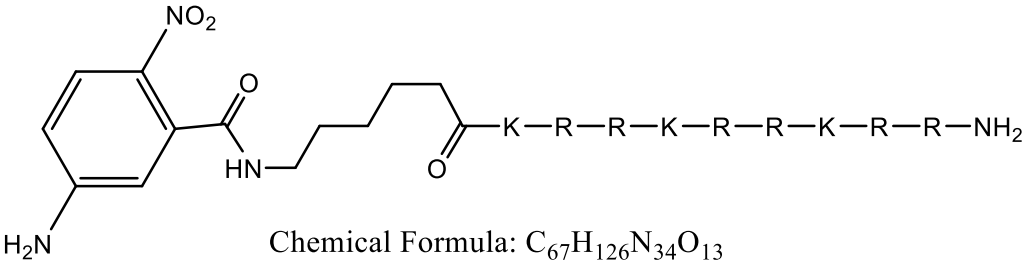


Figure 28: ESI-MS from LC-MS with deconvolution of peptide **8**. Exact mass calcd. for $C_{61}H_{115}N_{33}O_{12}$ = 1501.94, deconvoluted mass found = 1502.51.



Chemical Formula: $C_{67}H_{126}N_{34}O_{13}$
Exact Mass: 1615.02
Molecular Weight: 1615.97

Figure 29: Structure of peptide **9**

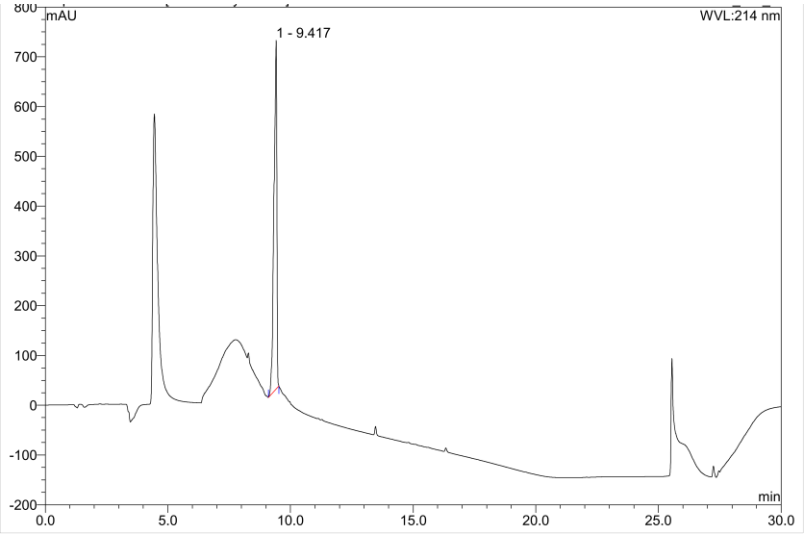


Figure 30: HPLC trace of crude peptide **9**, showing product peak at r.t. = 9.417 min. Gradient 0-95% ACN in 15 min.

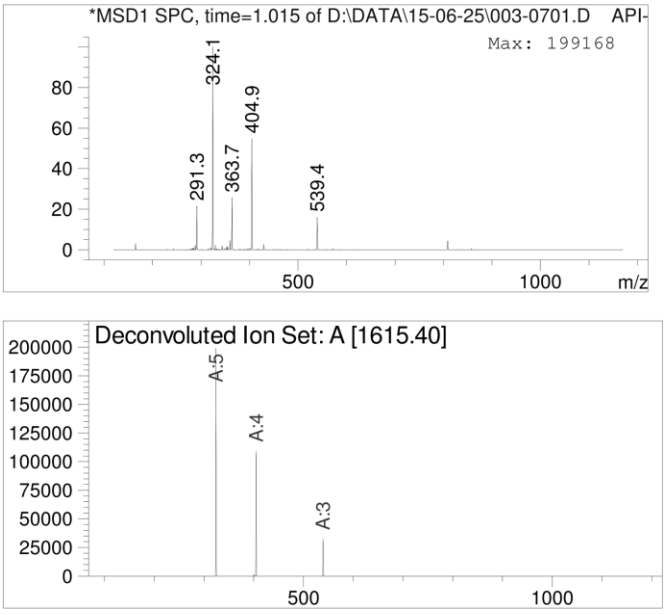
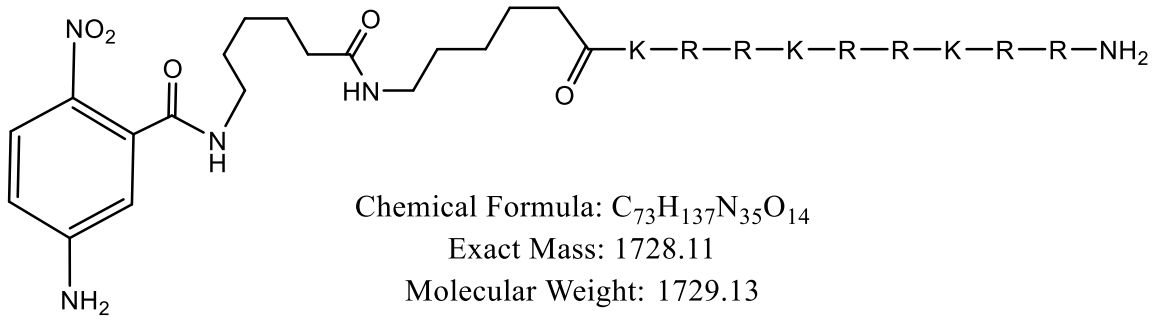


Figure 31: ESI-MS from LC-MS with deconvolution of peptide **9**. Exact mass calcd. for $C_{67}H_{126}N_{34}O_{13}$ = 1615.02, deconvoluted mass found = 1615.40.



Chemical Formula: $C_{73}H_{137}N_{35}O_{14}$
Exact Mass: 1728.11
Molecular Weight: 1729.13

Figure 32: Structure of peptide **10**

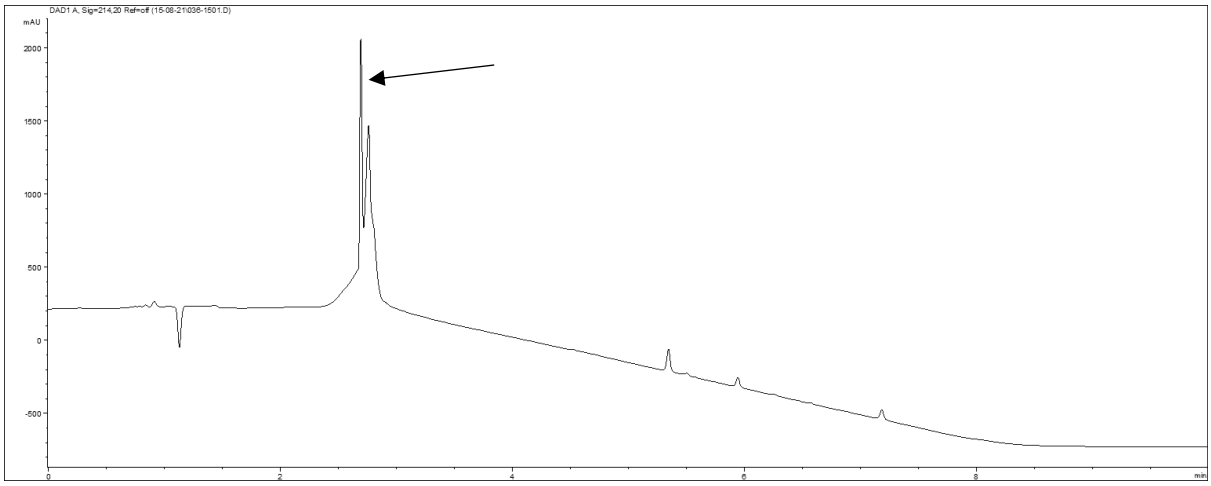
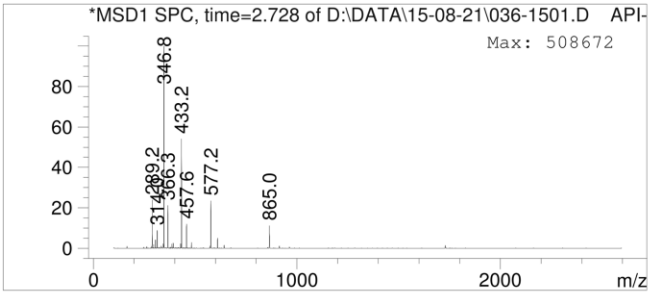


Figure 33: HPLC trace of crude peptide **10**, showing product peak at r.t. = 2.728 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.



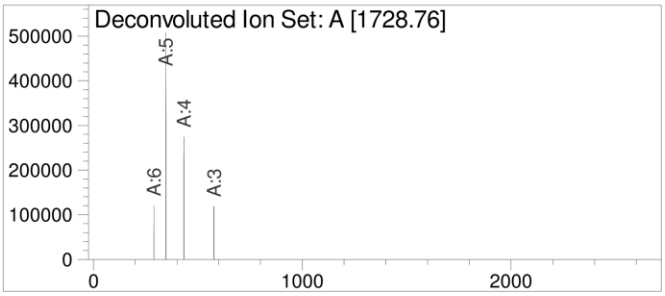
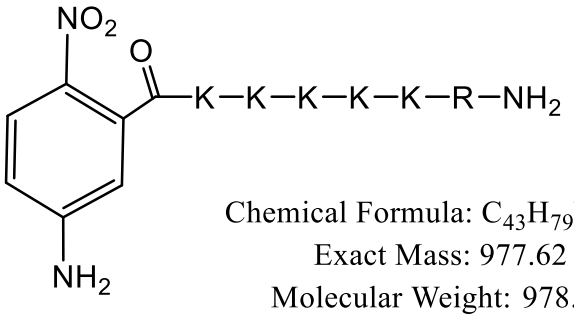


Figure 34: ESI-MS from LC-MS with deconvolution of peptide **10**. Exact mass calcd. for $C_{73}H_{137}N_{35}O_{14}$ = 1728.11, deconvoluted mass found = 1728.76.



Chemical Formula: $C_{43}H_{79}N_{17}O_9$
Exact Mass: 977.62
Molecular Weight: 978.22

Figure 35: Structure of peptide **11**

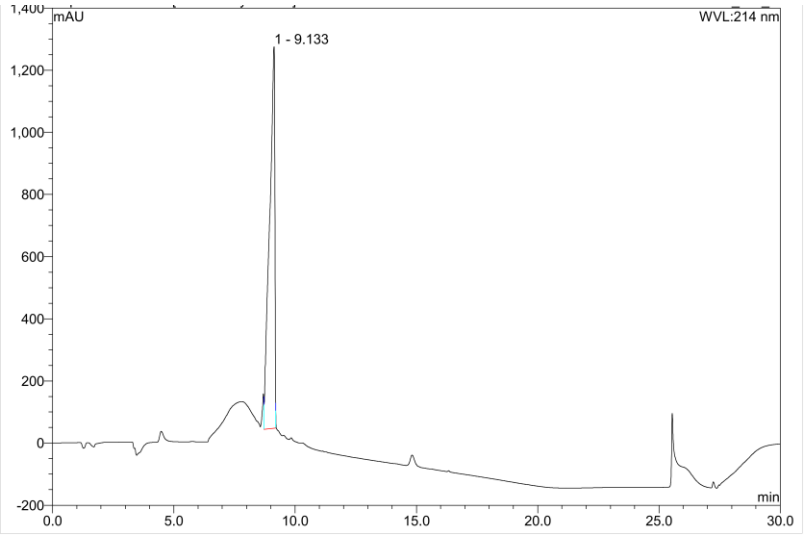


Figure 36: HPLC trace of crude peptide **11**, showing product peak at r.t. = 9.133 min. Gradient 0-95% ACN in 15 min.

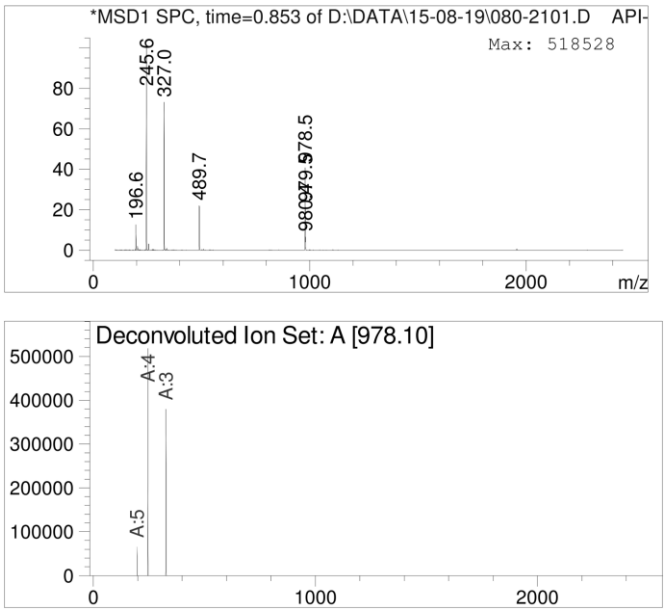
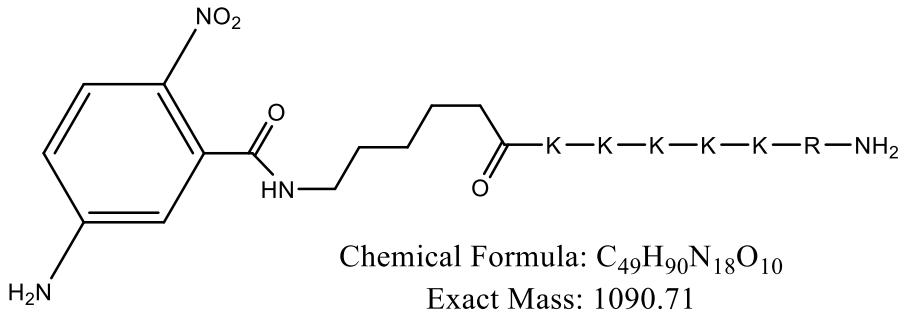


Figure 37: ESI-MS from LC-MS with deconvolution of peptide **11**. Exact mass calcd. for $C_{43}H_{79}N_{17}O_9$ = 977.62, deconvoluted mass found = 978.10.



Chemical Formula: $C_{49}H_{90}N_{18}O_{10}$
Exact Mass: 1090.71
Molecular Weight: 1091.38

Figure 38: Structure of peptide **12**

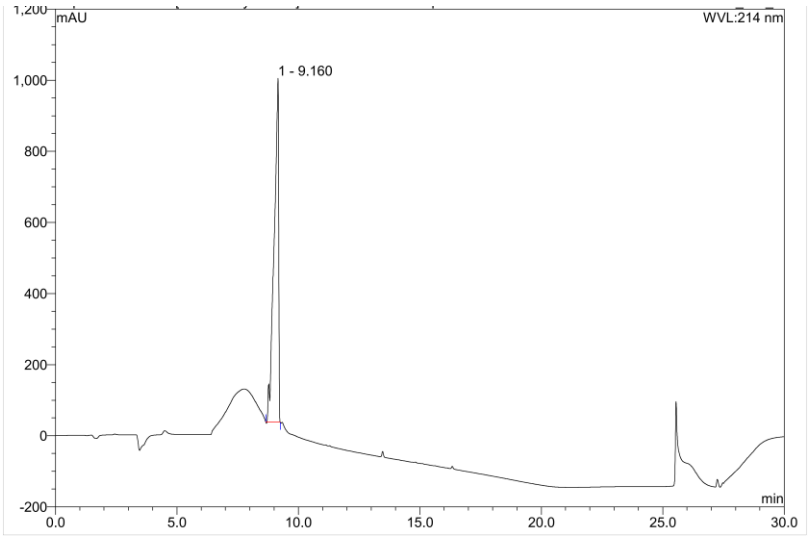


Figure 39: HPLC trace of crude peptide **12**, showing product peak at r.t. = 9.160 min. Gradient 0-100% ACN in 15 min.

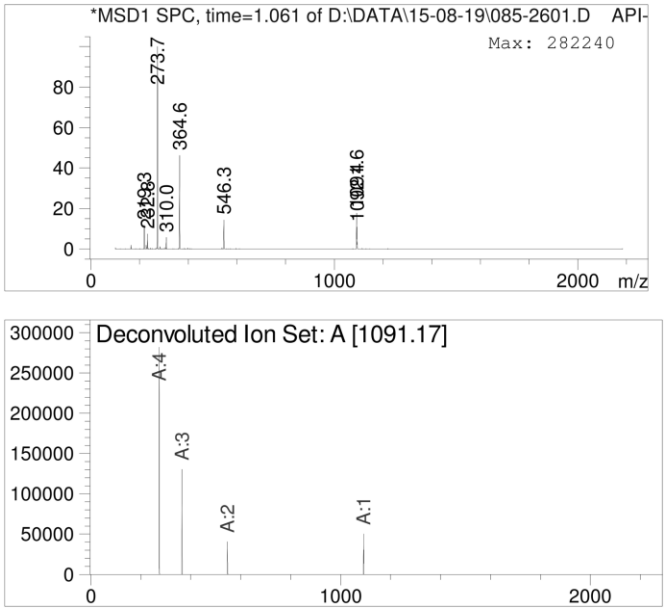
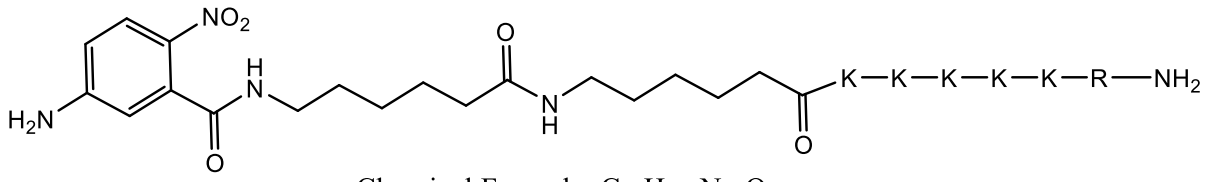


Figure 40: ESI-MS from LC-MS with deconvolution of peptide **12**. Exact mass calcd. for $C_{49}H_{90}N_{18}O_{10}$ = 1090.71, deconvoluted mass found = 1091.38.



Chemical Formula: $C_{55}H_{101}N_{19}O_{11}$
Exact Mass: 1203.79
Molecular Weight: 1204.54

Figure 41: Structure of peptide **13**

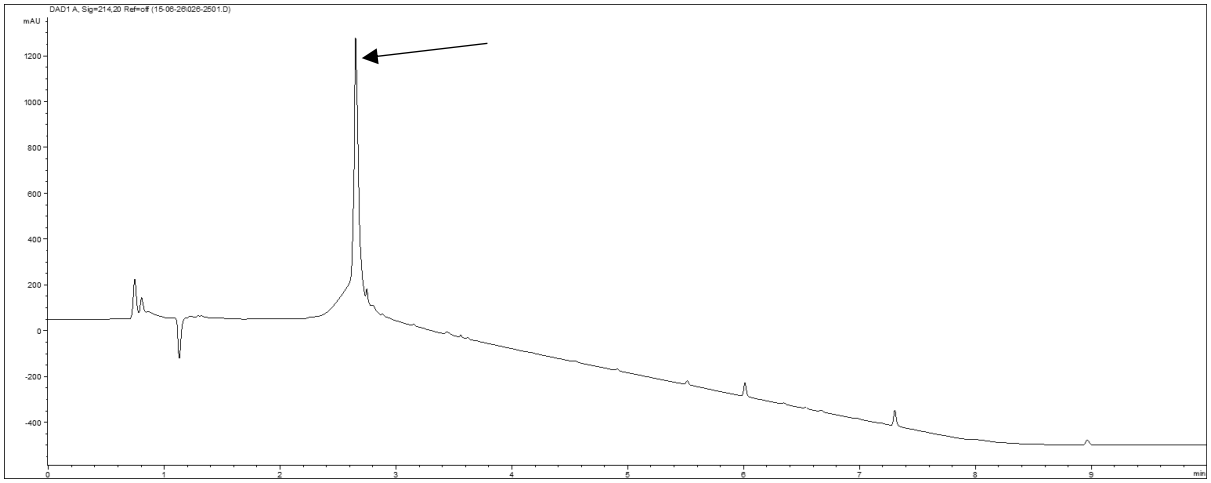
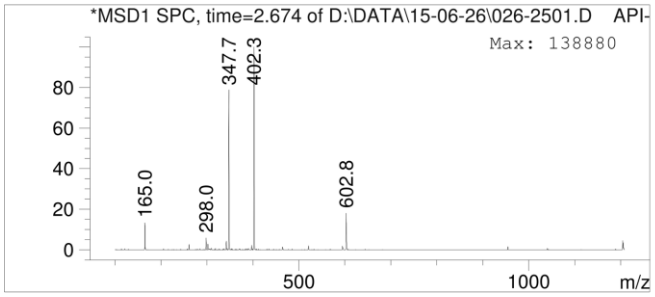


Figure 42: HPLC trace of crude peptide **13**, showing product peak at r.t. = 2.674 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.



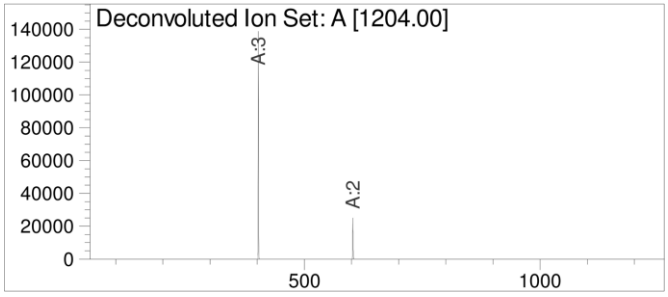
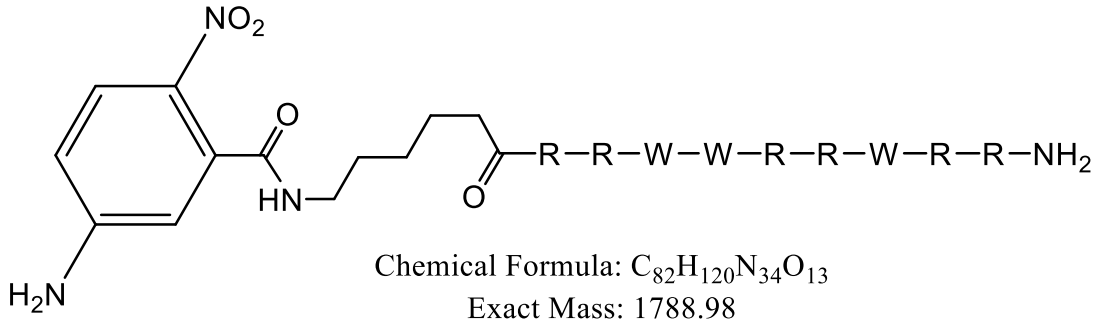


Figure 43: ESI-MS from LC-MS with deconvolution of peptide **13**. Exact mass calcd. for $C_{55}H_{101}N_{19}O_{11}$ = 1203.79, deconvoluted mass found = 1204.00.



Chemical Formula: $C_{82}H_{120}N_{34}O_{13}$
Exact Mass: 1788.98
Molecular Weight: 1790.05

Figure 44: Structure of peptide **14**

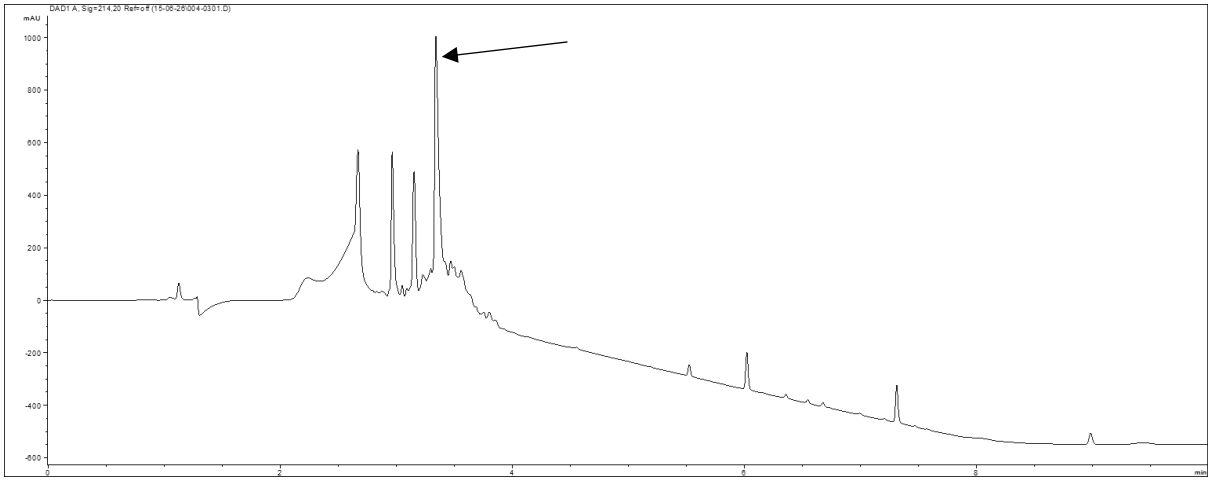


Figure 45: HPLC trace of crude peptide **14**, showing product peak at r.t. = 3.365 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.

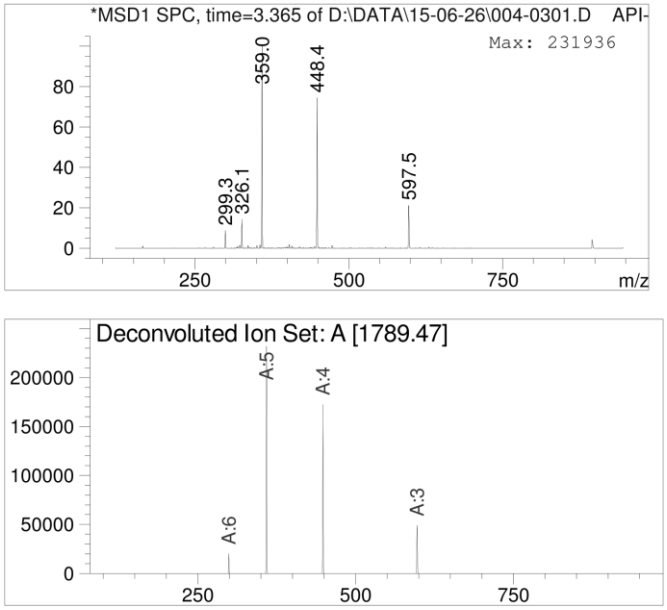
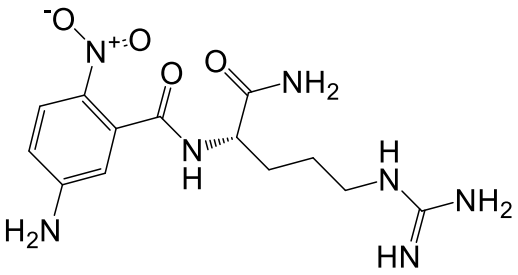


Figure 46: ESI-MS from LC-MS with deconvolution of peptide **14**. Exact mass calcd. for $C_{82}H_{120}N_{34}O_{13}$ = 1788.98, deconvoluted mass found = 1789.47.



Chemical Formula: $C_{13}H_{19}N_7O_4$
Exact Mass: 337.15
Molecular Weight: 337.34

Figure 47: Structure of compound **15**

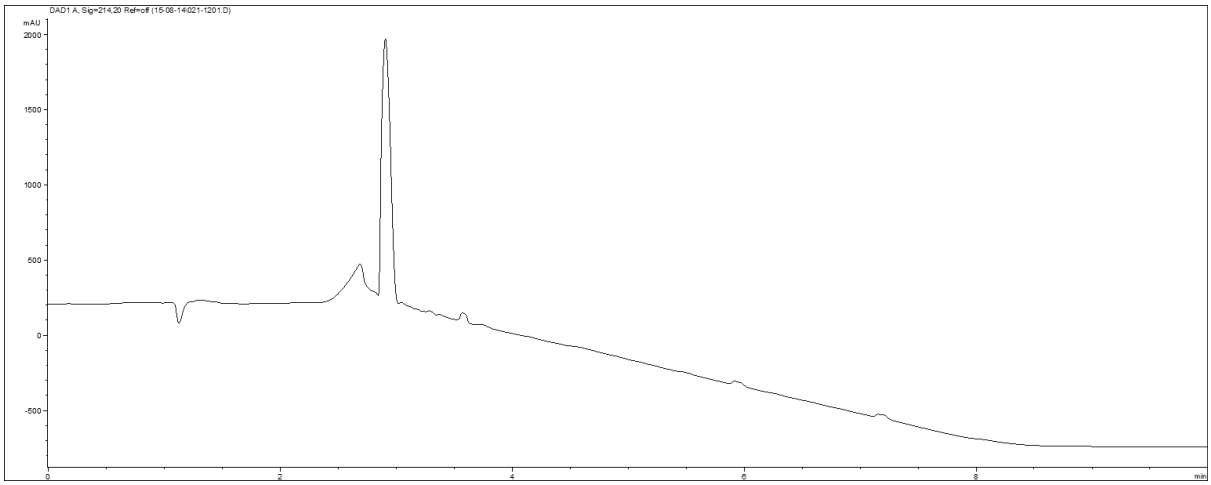


Figure 48: HPLC trace of crude compound **15**, showing product peak at r.t. = 2.918 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.

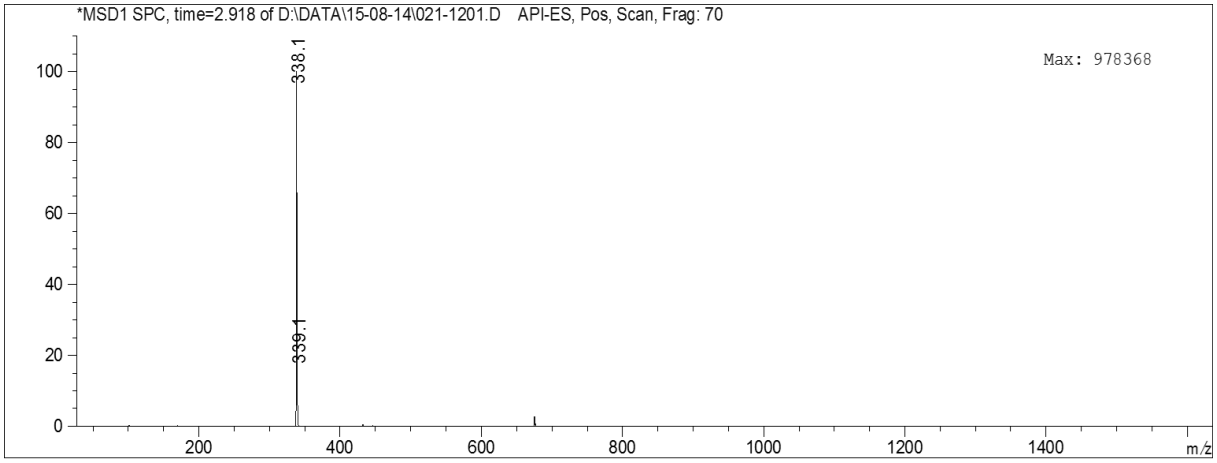


Figure 49: ESI-MS from LC-MS of compound **15**. Exact mass calcd. for $C_{13}H_{19}N_7O_4$ = 337.15, mass found $M + H^+$ = 338.10.

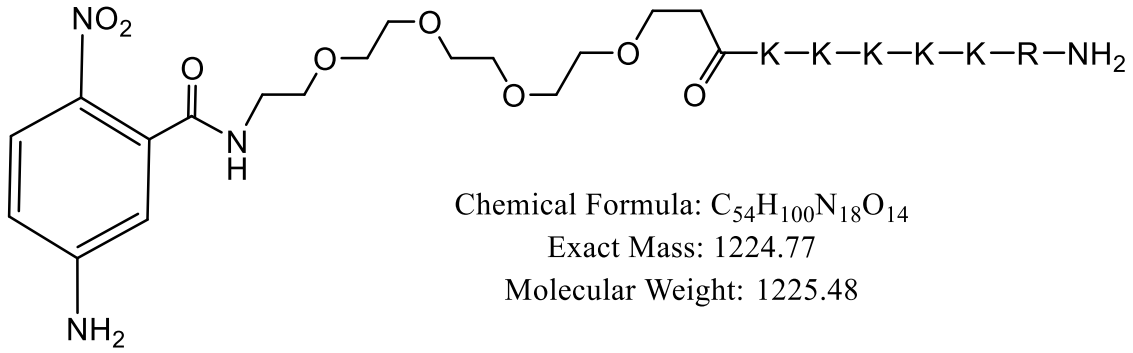


Figure 50: Structure of peptide **16**

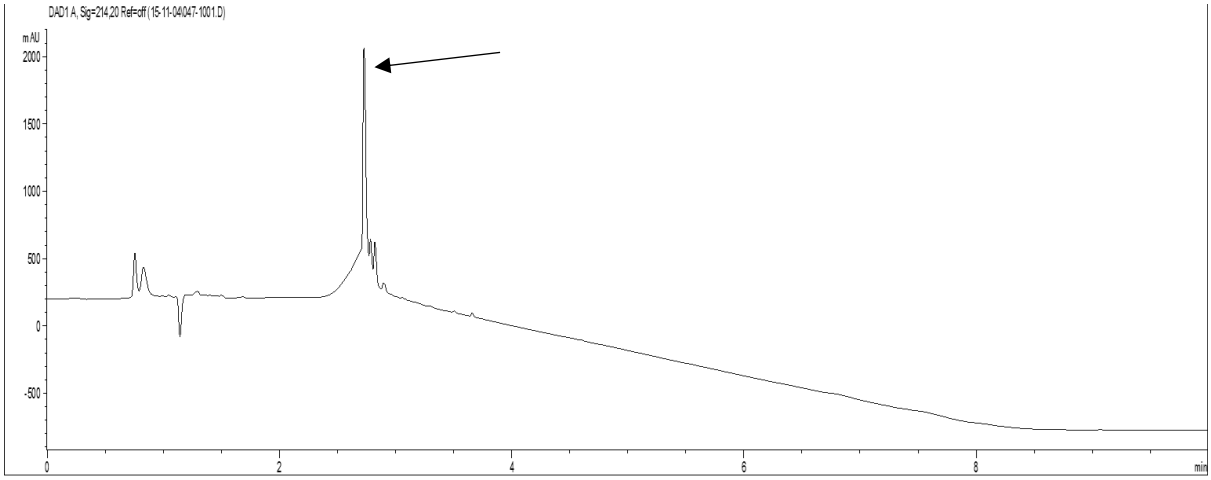


Figure 51: HPLC trace of crude peptide **16**, showing product peak at r.t. = 2.729 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.

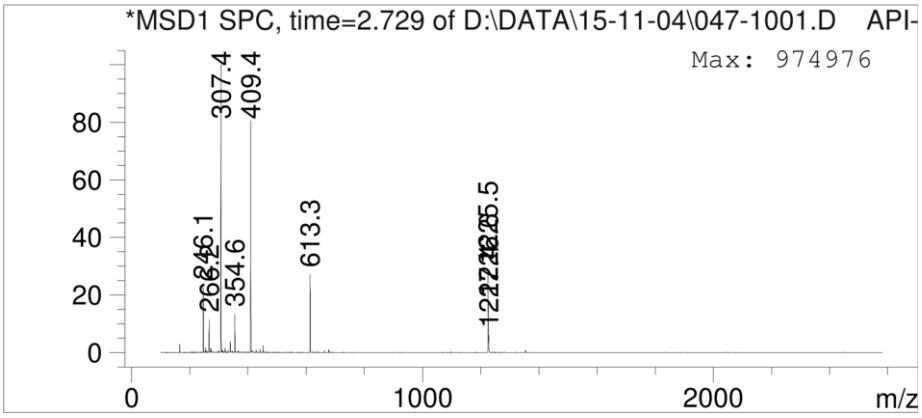
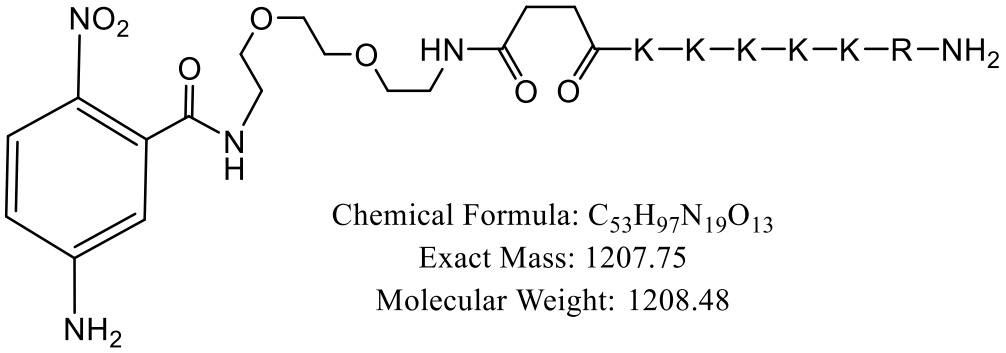


Figure 52: ESI-MS from LC-MS of peptide **16**. Exact mass calcd. for $C_{54}H_{100}N_{18}O_{14}$ = 1224.77, mass found M + H^+ = 1226.5, M/2 + H^+ = 613.3, M/3 + H^+ = 409.4, M/4 + H^+ = 307.4.



Chemical Formula: $C_{53}H_{97}N_{19}O_{13}$
Exact Mass: 1207.75
Molecular Weight: 1208.48

Figure 53: Structure of peptide **17**

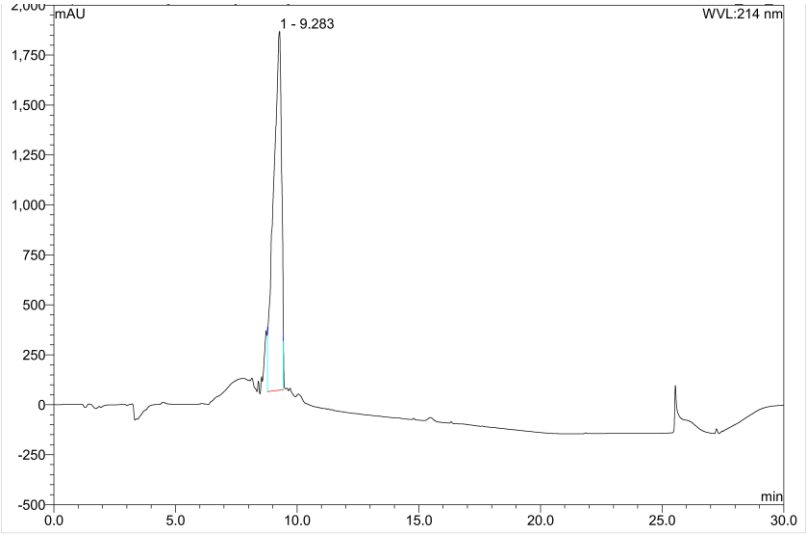


Figure 54: HPLC trace of crude peptide **17**, showing product peak at r.t. = 9.283 min. Gradient 0-95% ACN in 15 min.

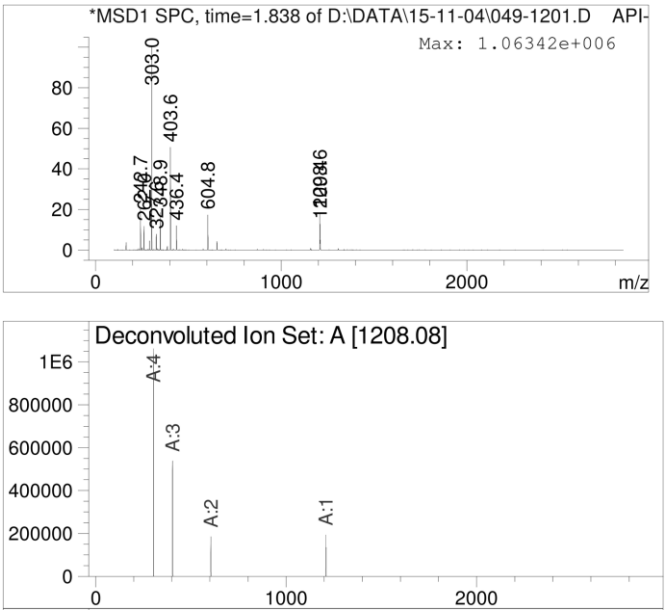
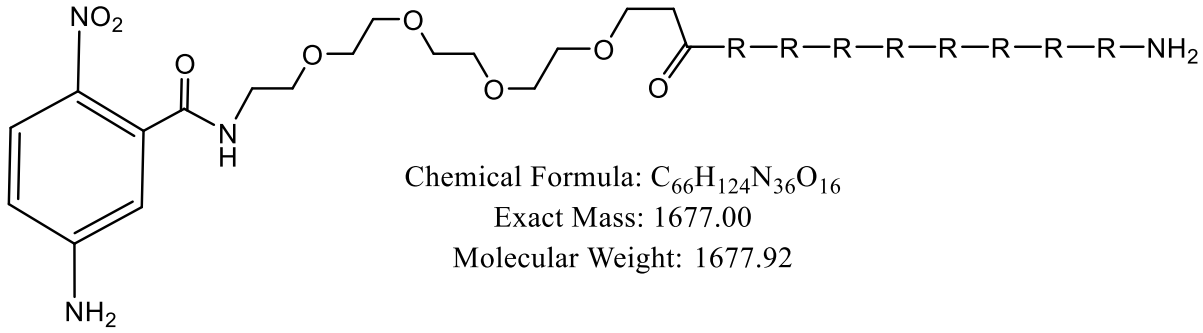


Figure 55: ESI-MS from LC-MS with deconvolution of peptide **17**. Exact mass calcd. for $C_{53}H_{97}N_{19}O_{13}$ = 1207.75, deconvoluted mass found = 1208.08.



Chemical Formula: $C_{66}H_{124}N_{36}O_{16}$
Exact Mass: 1677.00
Molecular Weight: 1677.92

Figure 56: Structure of peptide **18**

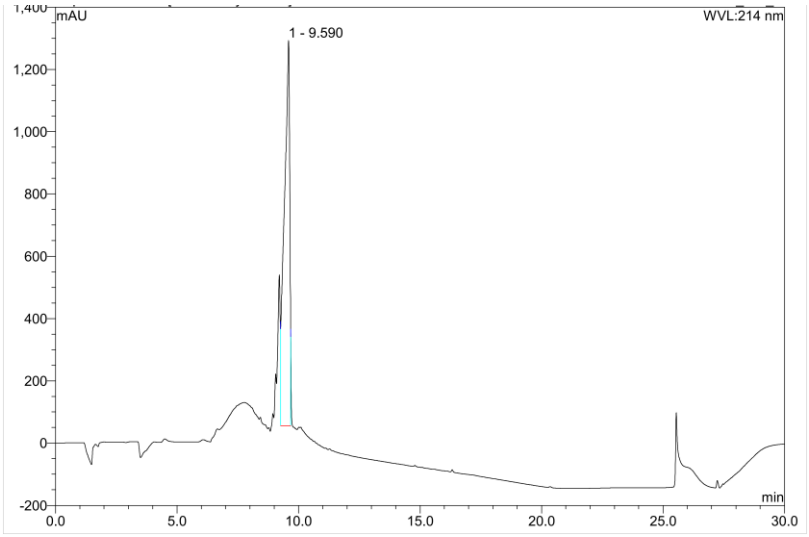
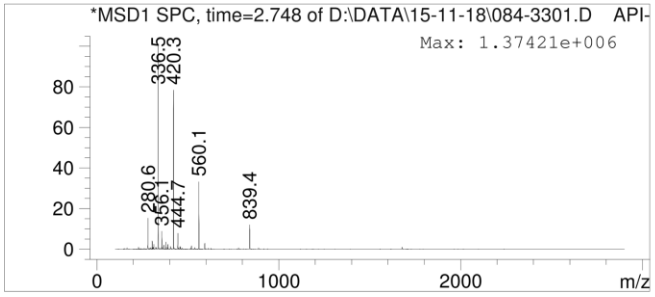


Figure 57: HPLC trace of crude peptide **18**, showing product peak at r.t. = 9.590 min. Gradient 0-95% ACN in 15 min.



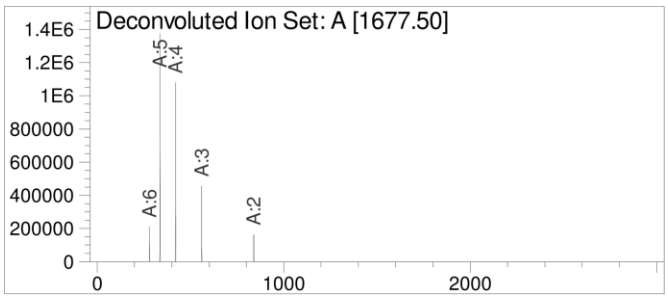
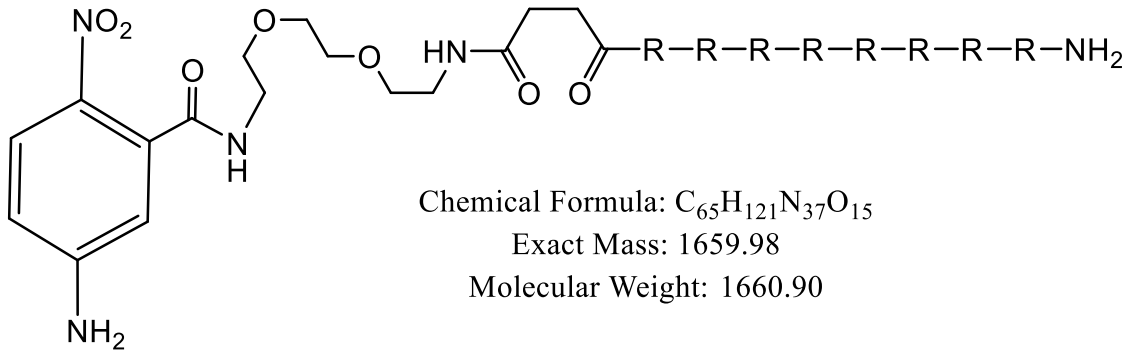


Figure 58: ESI-MS from LC-MS with deconvolution of peptide **18**. Exact mass calcd. for $C_{66}H_{124}N_{36}O_{16}$ = 1677.00, deconvoluted mass found = 1677.50.



Chemical Formula: $C_{65}H_{121}N_{37}O_{15}$
Exact Mass: 1659.98
Molecular Weight: 1660.90

Figure 59: Structure of peptide **19**

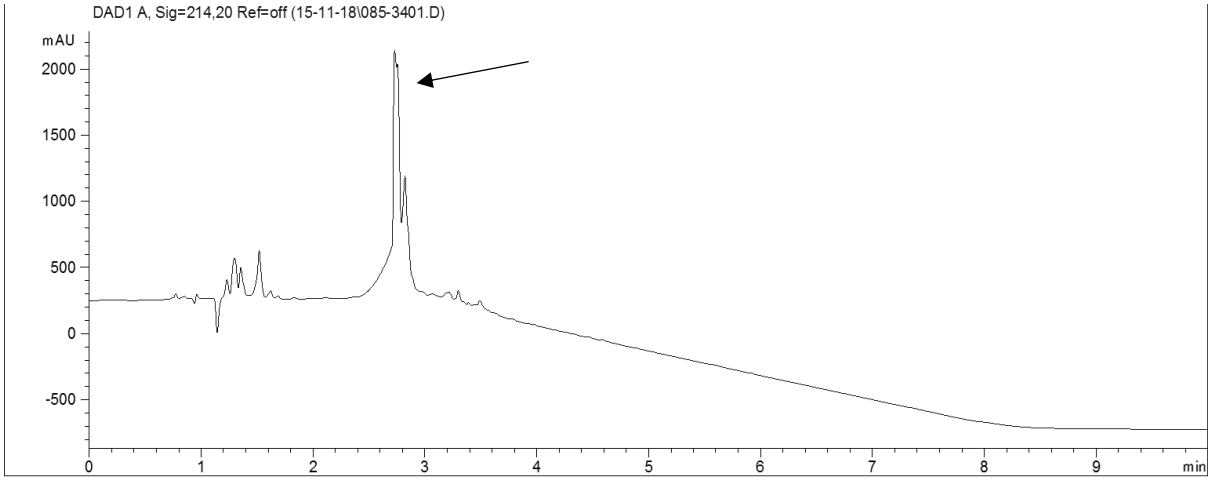
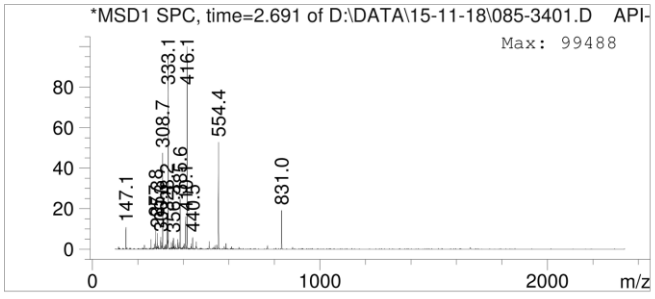


Figure 60: HPLC trace of crude peptide **19**, showing product peak at r.t. = 2.748 min. Gradient 0-100% ACN in 6 min using 0.1% HCOOH buffer.



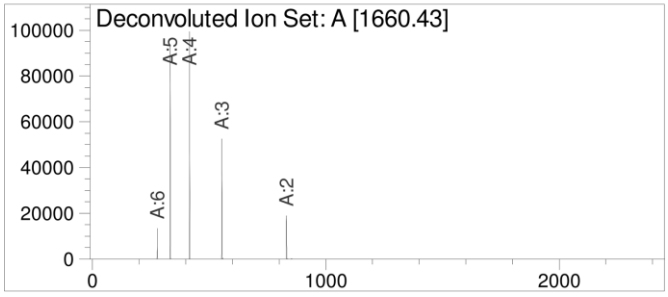
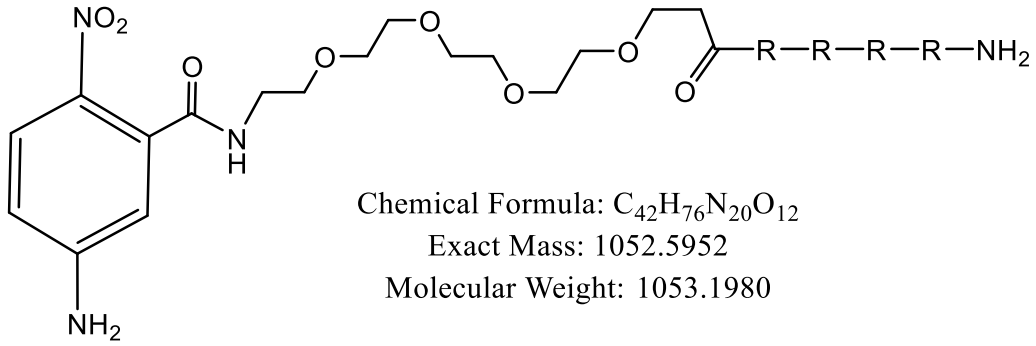


Figure 61: ESI-MS from LC-MS with deconvolution of peptide **19**. Exact mass calcd. for $C_{65}H_{121}N_{37}O_{15}$ = 1659.98, deconvoluted mass found = 1660.43.



Chemical Formula: $C_{42}H_{76}N_{20}O_{12}$
Exact Mass: 1052.5952
Molecular Weight: 1053.1980

Figure 62: Structure of peptide **20**

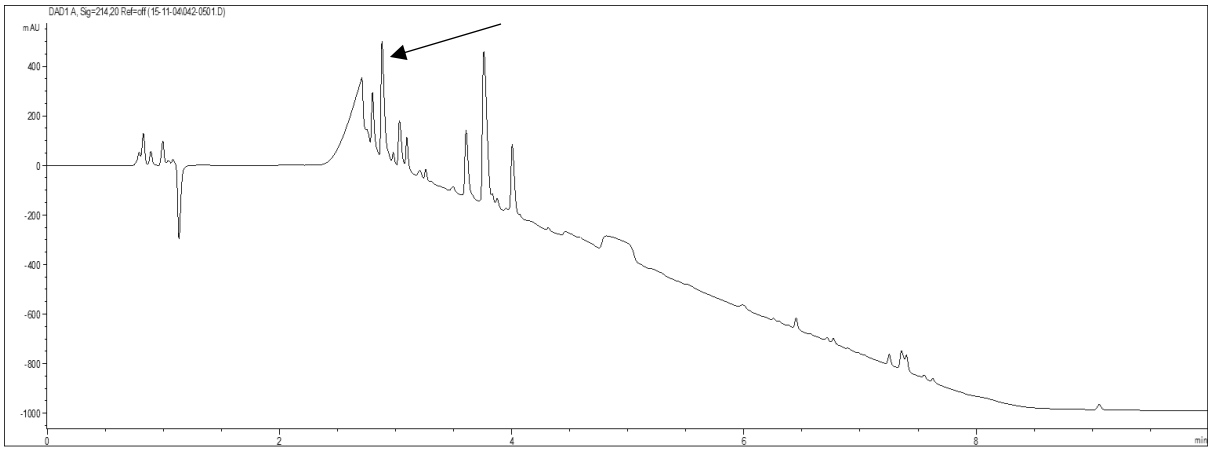


Figure 63: HPLC trace of crude peptide **20**, showing product peak at r.t. = 2.899 min. Gradient 0-100% ACN in 6 min. (Low crude purity due to an error in the peptide synthesizer – contains a mixture of peptide **20** & **21**)

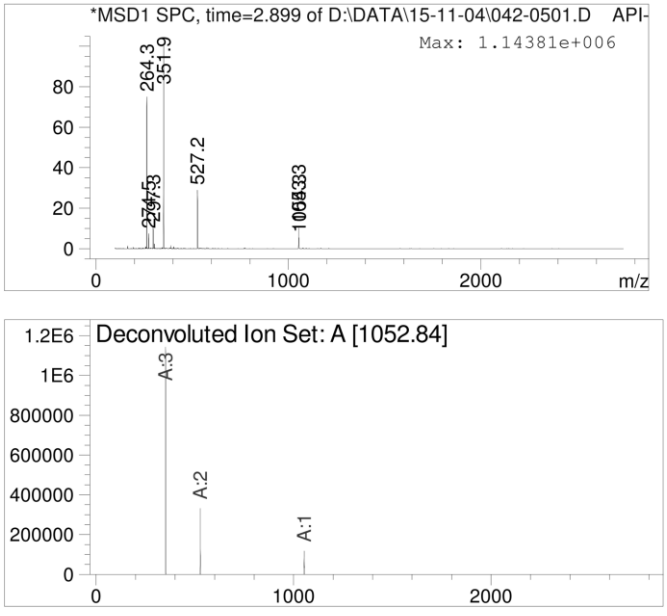
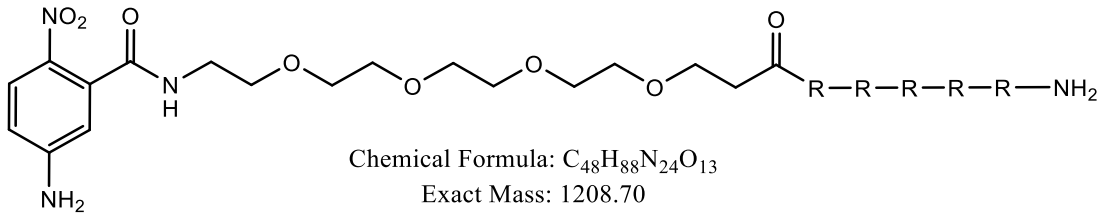


Figure 64: ESI-MS from LC-MS with deconvolution of peptide **20**. Exact mass calcd. for $C_{42}H_{76}N_{20}O_{12}$ = 1052.59, deconvoluted mass found = 1052.84.



Chemical Formula: $C_{48}H_{88}N_{24}O_{13}$
Exact Mass: 1208.70
Molecular Weight: 1209.39

Figure 65: Structure of peptide **21**

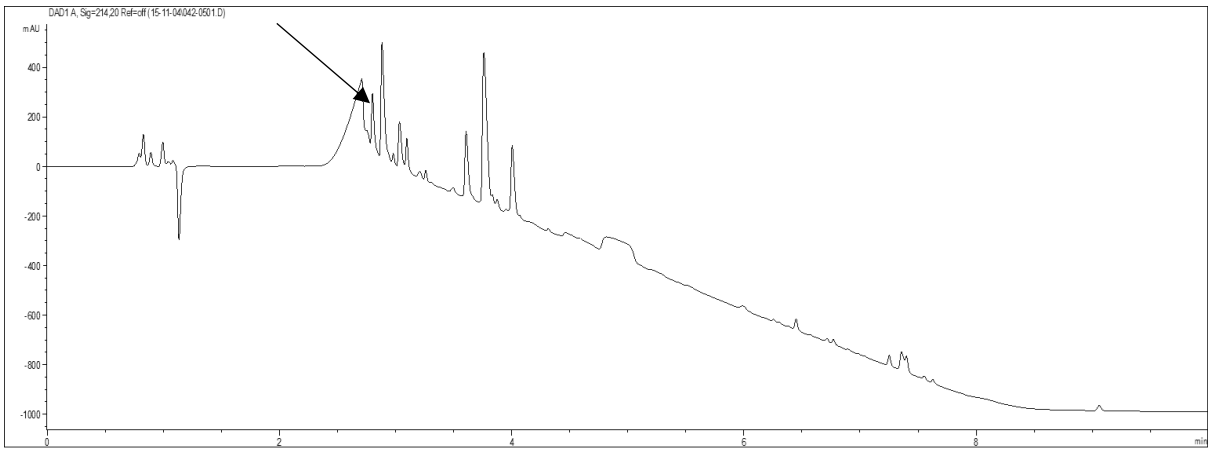
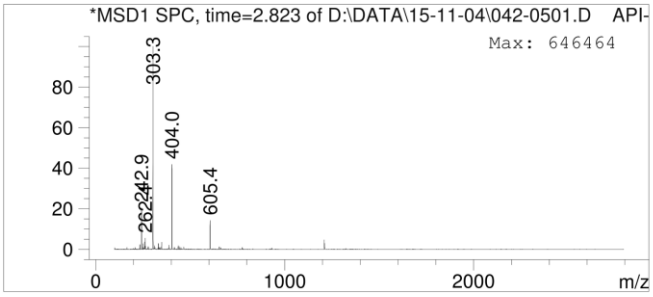


Figure 66: HPLC trace of crude peptide **21**, showing product peak at r.t. = 2.823 min. Gradient 0-100% ACN in 6 min. (Low crude purity due to an error in the peptide synthesizer – contains a mixture of peptides **20** & **21**)



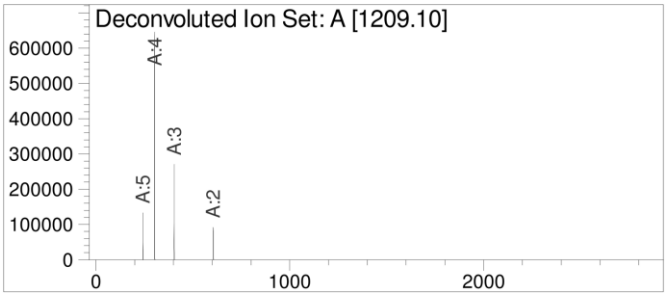
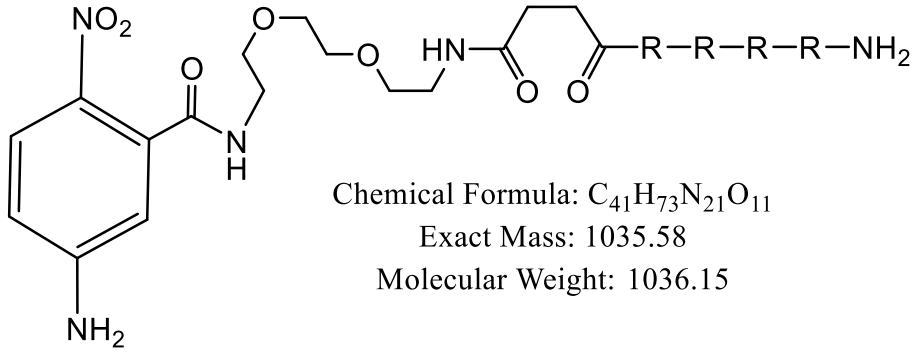


Figure 67: ESI-MS from LC-MS with deconvolution of peptide **21**. Exact mass calcd. for $C_{48}H_{88}N_{24}O_{13}$ = 1208.70, deconvoluted mass found = 1209.10.



Chemical Formula: $C_{41}H_{73}N_{21}O_{11}$
Exact Mass: 1035.58
Molecular Weight: 1036.15

Figure 68: Structure of peptide **22**

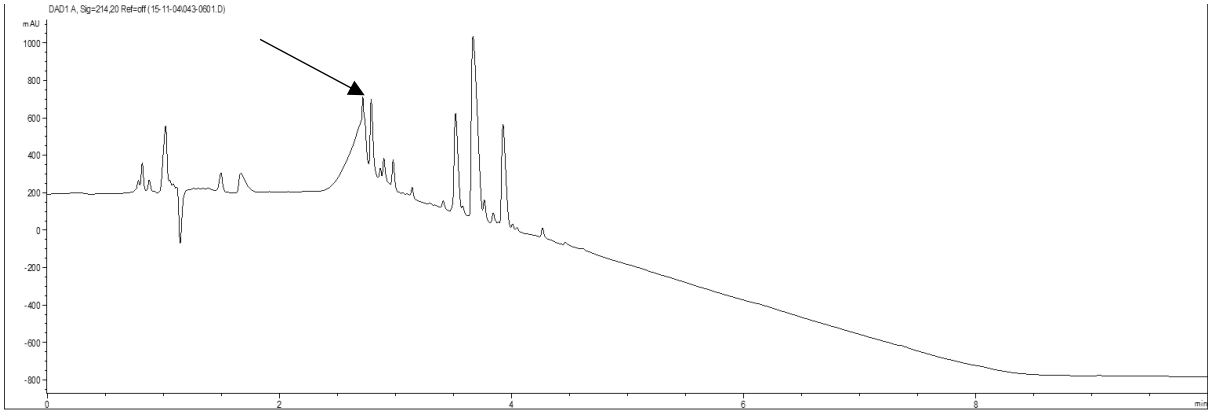
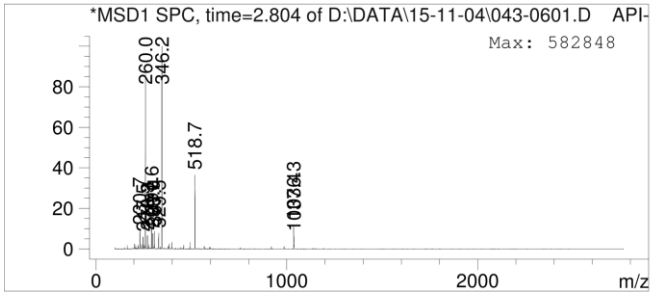


Figure 69: HPLC trace of crude peptide **22**, showing product peak at r.t. = 2.804 min. Gradient 0-100% ACN in 6 min. (Low crude purity due to an error in the peptide synthesizer)



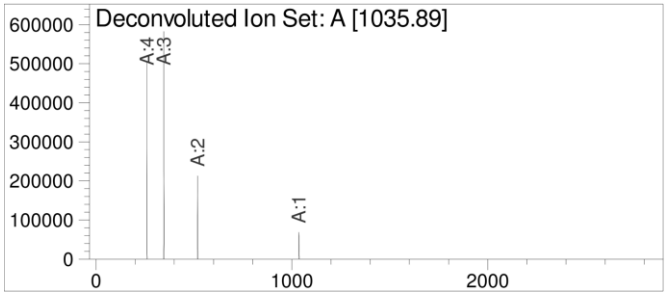


Figure 70: ESI-MS from LC-MS with deconvolution of peptide **22**. Exact mass calcd. for $C_{41}H_{73}N_{21}O_{11}$ = 1035.58, deconvoluted mass found = 1035.89.

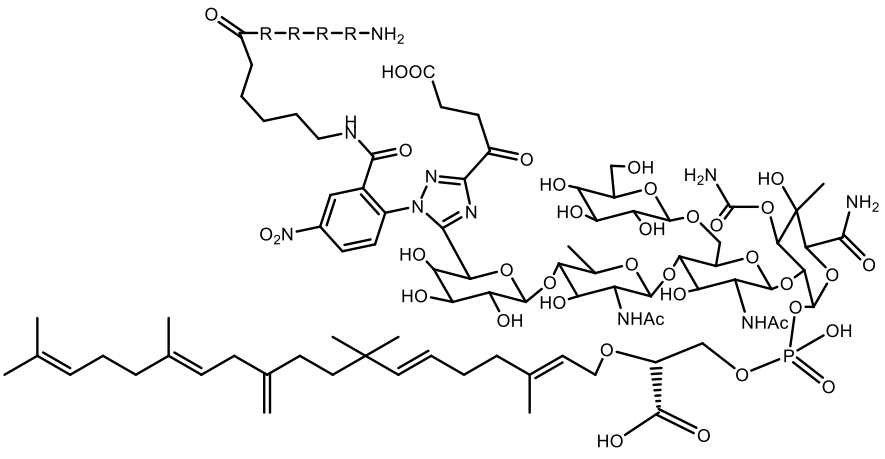


Figure 71: Structure of compound **23**

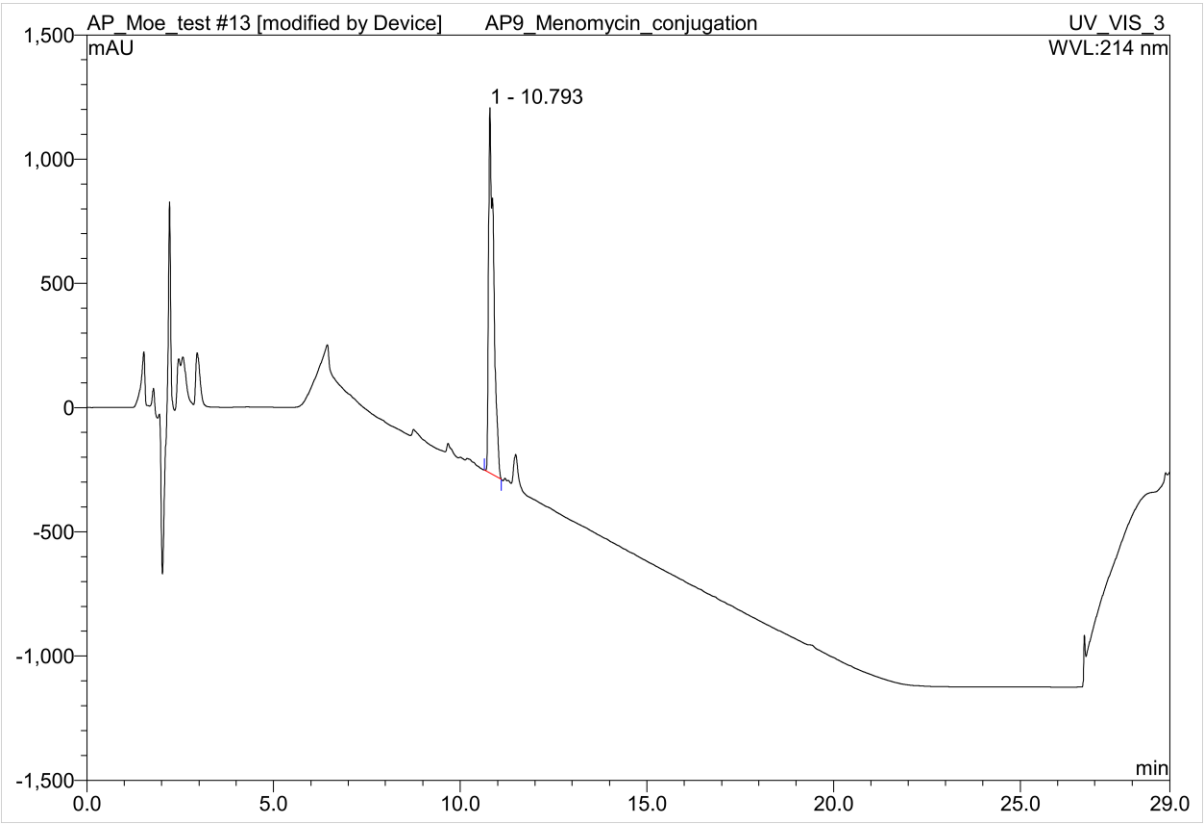


Figure 72: HPLC trace of RP-HPLC purified compound **23**, showing product peak at r.t = 10.793 min. Gradient 0-95% ACN in 15 min.

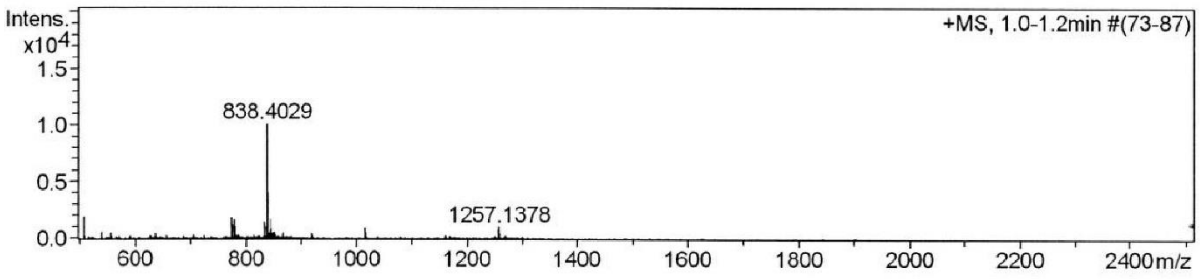


Figure 73: ESI-MS of RP-HPLC purified compound **23**. Exact mass calcd. for $C_{106}H_{171}N_{26}O_{42}P$ = 2511.18, mass found = M/2 + H^+ = 1257.1378, M/3 + H^+ = 838.4029.

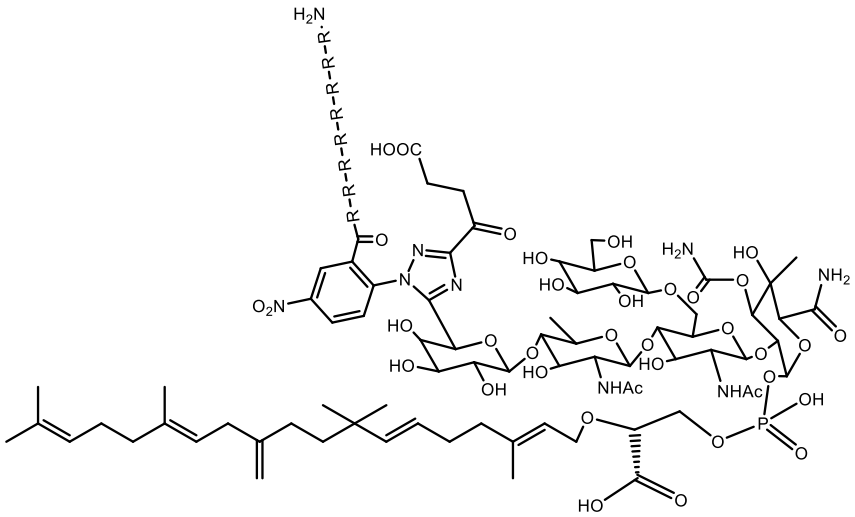


Figure 74: Structure of compound **24**

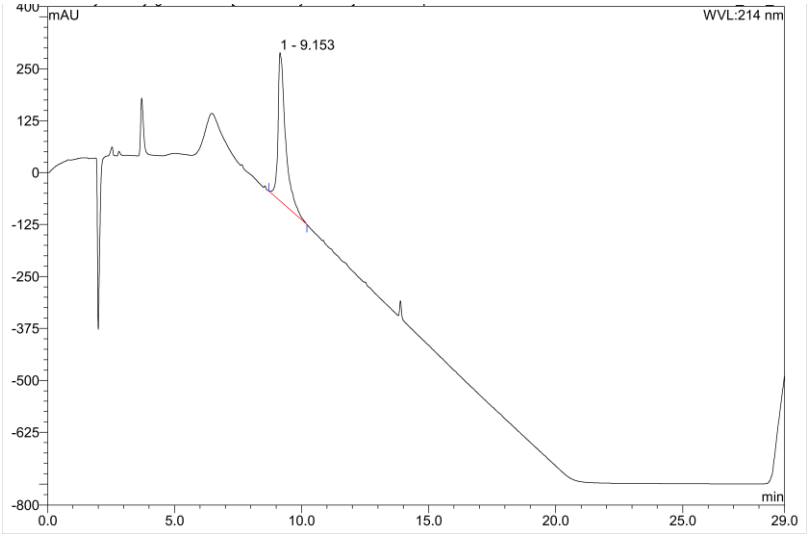


Figure 75: HPLC trace of RP-HPLC purified compound **24**, showing product peak at r.t. = 9.153 min. Gradient 0-95% ACN in 15 min.

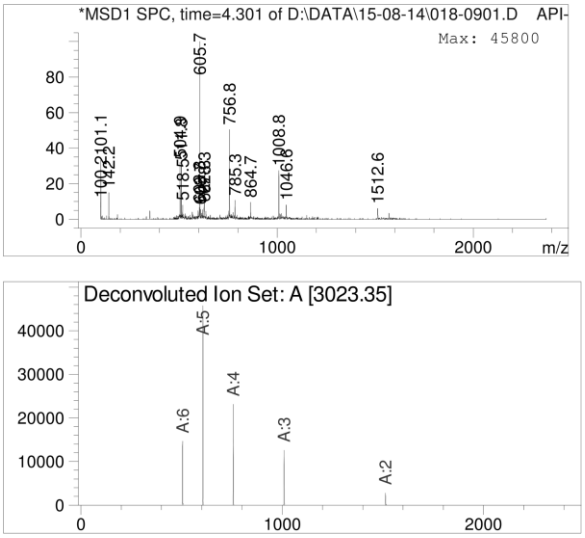


Figure 76: ESI-MS from LC-MS with deconvolution of compound **24**. Exact mass calcd. for $C_{124}H_{208}N_{41}O_{45}P$ = 3022.50, deconvoluted mass found = 3023.35.

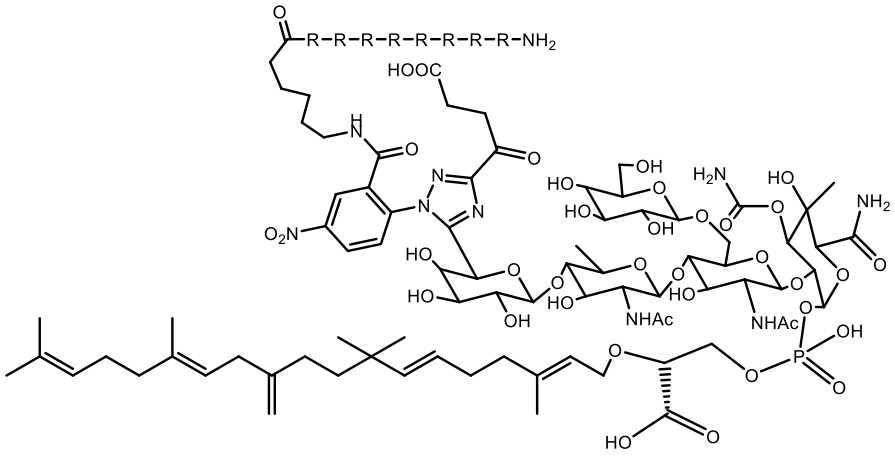


Figure 77: Structure of compound **25**

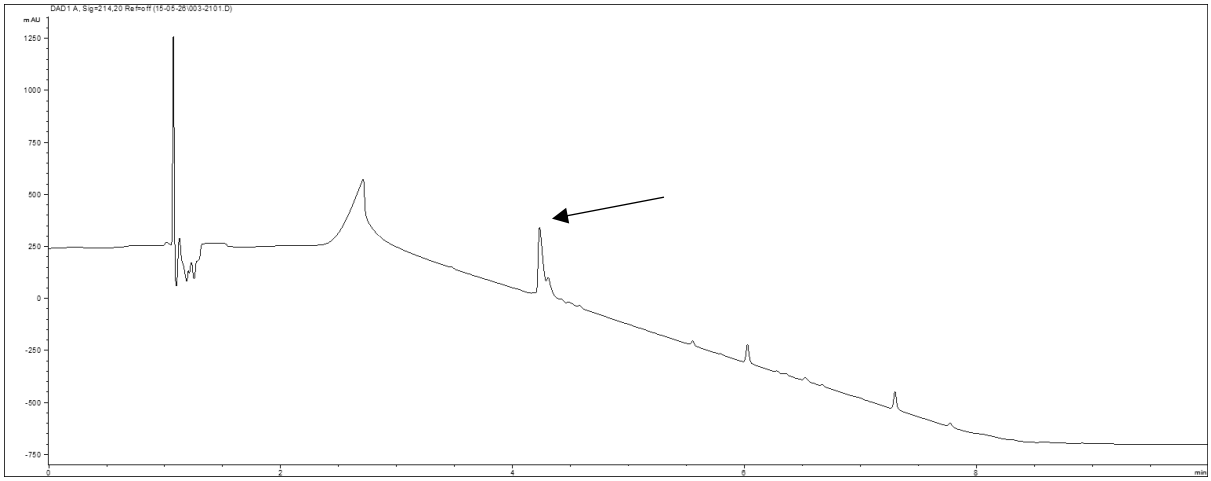


Figure 78: HPLC trace of RP-HPLC purified compound **25**, showing product peak at r.t. = 4.263 min. Gradient 0-100% ACN in 6 min.

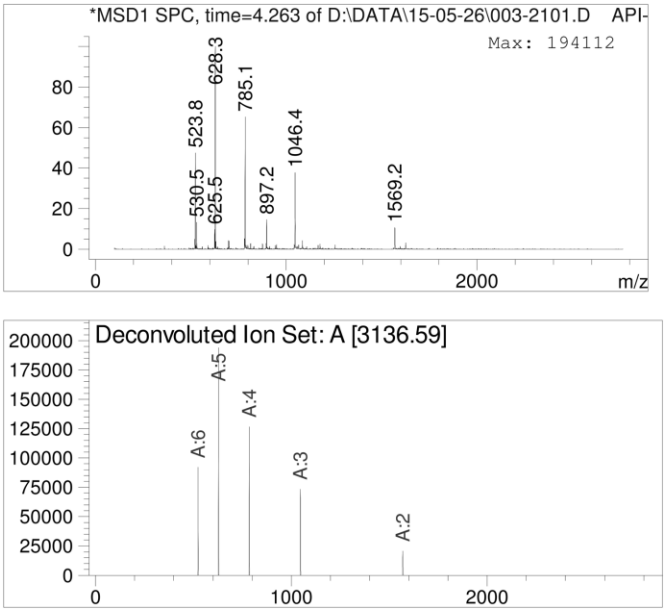


Figure 79: ESI-MS from LC-MS with deconvolution of compound **25**. Exact mass calcd. for $C_{130}H_{219}N_{42}O_{46}P$ = 3135.58, deconvoluted mass found = 3136.59.

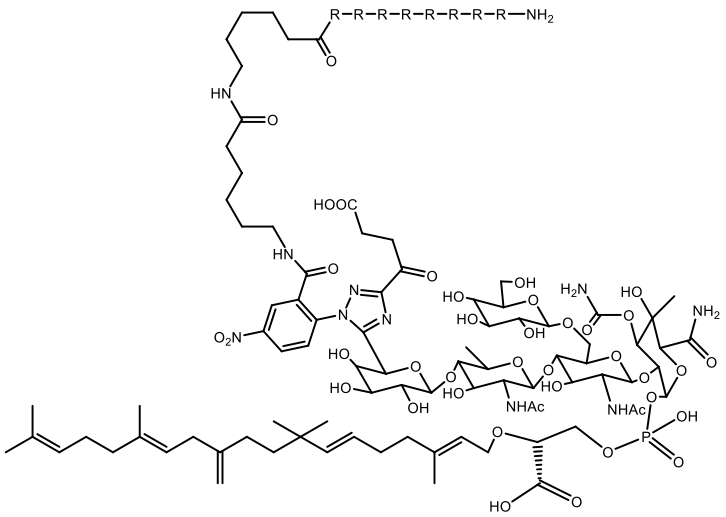


Figure 80: Structure of compound **26**

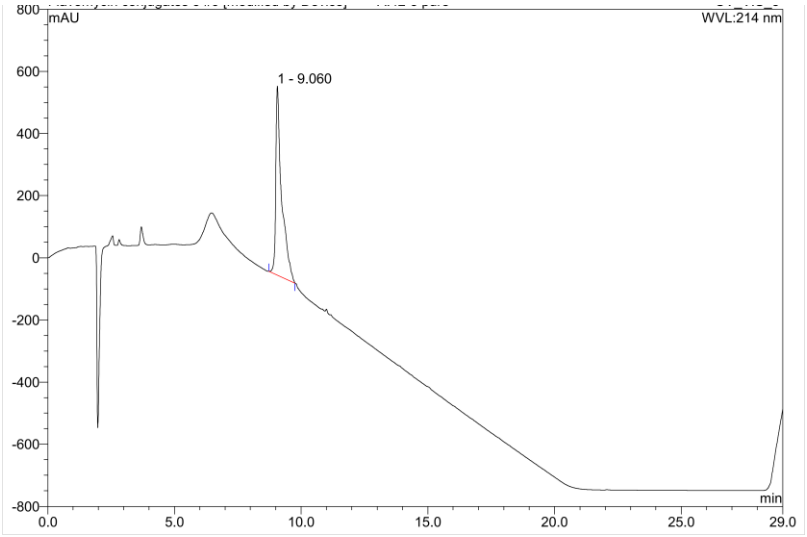


Figure 81: HPLC trace of RP-HPLC purified compound **26**, showing product peak at r.t. = 9.060 min. Gradient 0-95% ACN in 15 min.

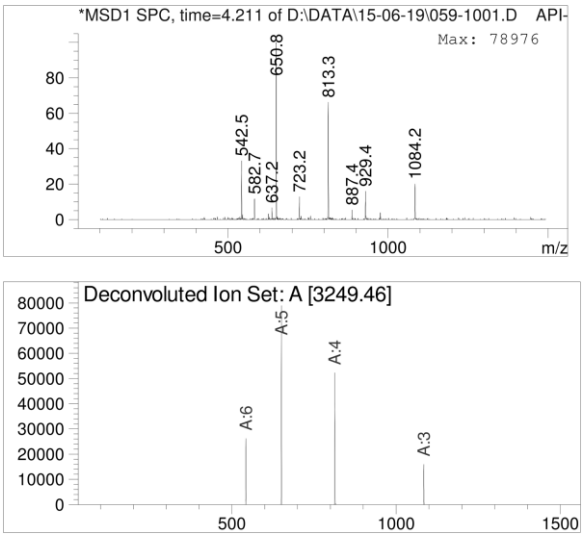
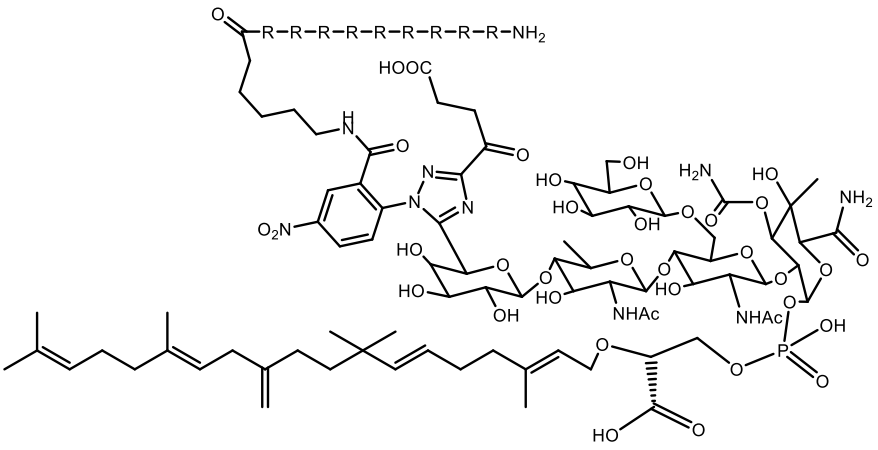


Figure 82: ESI-MS from LC-MS with deconvolution of compound **26**. Exact mass calcd. for $C_{136}H_{230}N_{43}O_{47}P$ = 3248.67, deconvoluted mass found = 3249.46.



Chemical Formula: $C_{136}H_{231}N_{46}O_{47}P$
Exact Mass: 3291.68
Molecular Weight: 3293.59

Figure 83: Structure of compound **27**

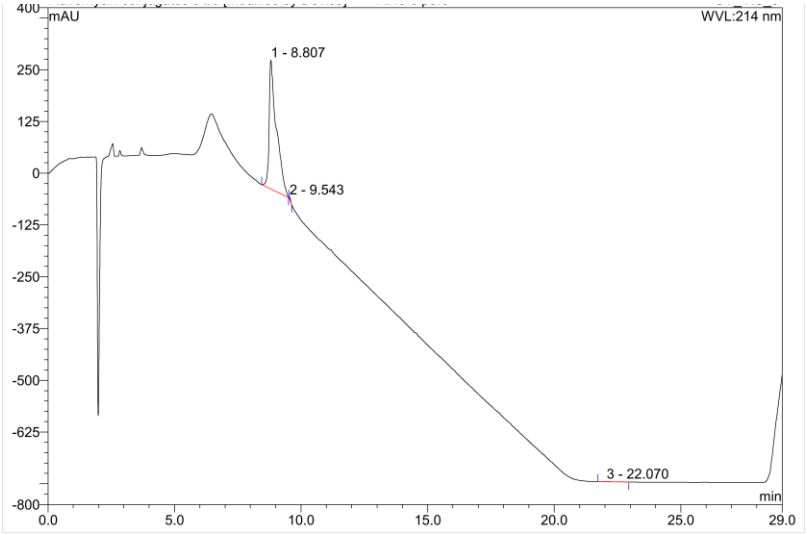


Figure 84: HPLC trace of RP-HPLC purified compound **27**, showing product peak at r.t. = 8.807 min. Gradient 0-95% ACN in 15 min.

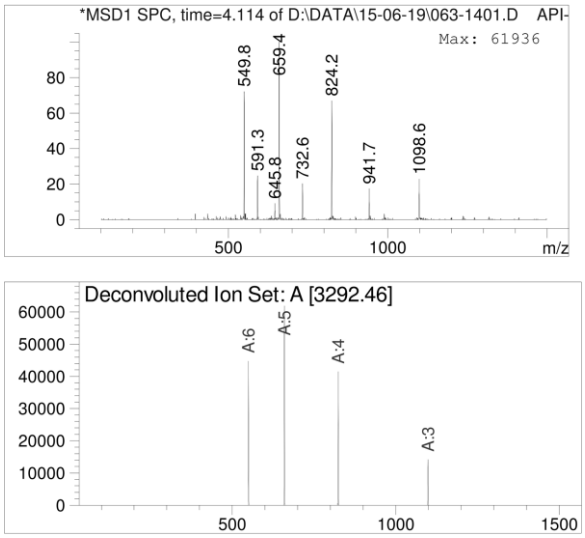
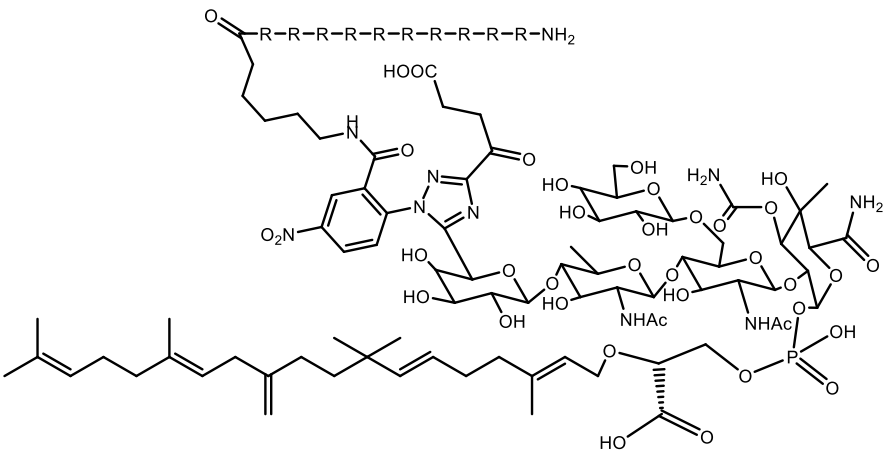


Figure 85: ESI-MS from LC-MS with deconvolution of compound **27**. Exact mass calcd. for $C_{136}H_{231}N_{46}O_{47}P$ = 3291.68, deconvoluted mass found = 3292.46.



Chemical Formula: $C_{142}H_{243}N_{50}O_{48}P$
Exact Mass: 3447.78
Molecular Weight: 3449.78

Figure 86: Structure of compound **28**

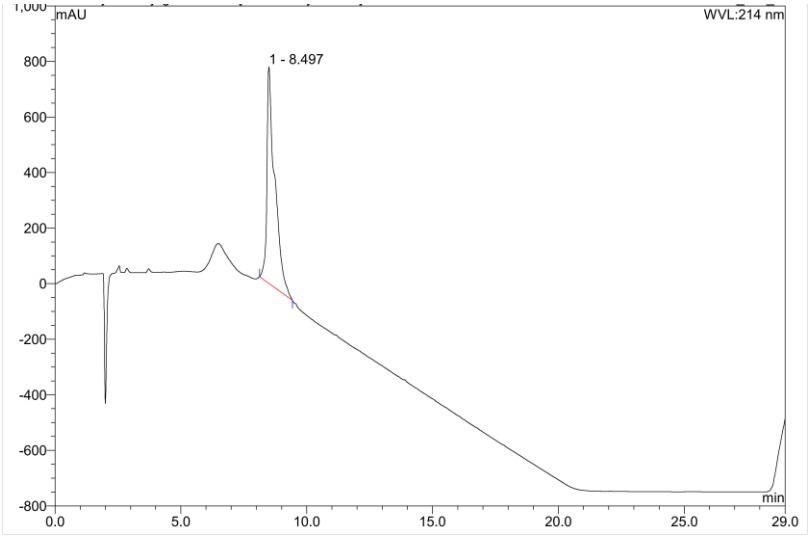


Figure 87: HPLC trace of RP-HPLC purified compound **28**, showing product peak at r.t. = 8.497 min. Gradient 0-95% ACN in 15 min.

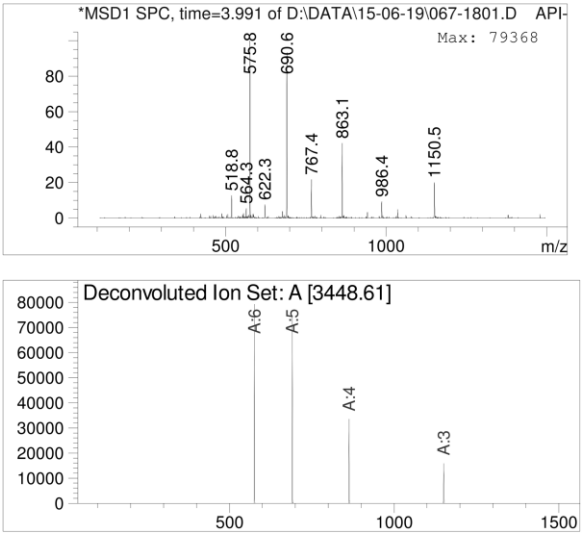


Figure 88: ESI-MS from LC-MS with deconvolution of compound **28**. Exact mass calcd. for $C_{142}H_{243}N_{50}O_{48}P$ = 3447.78, deconvoluted mass found = 3448.61.

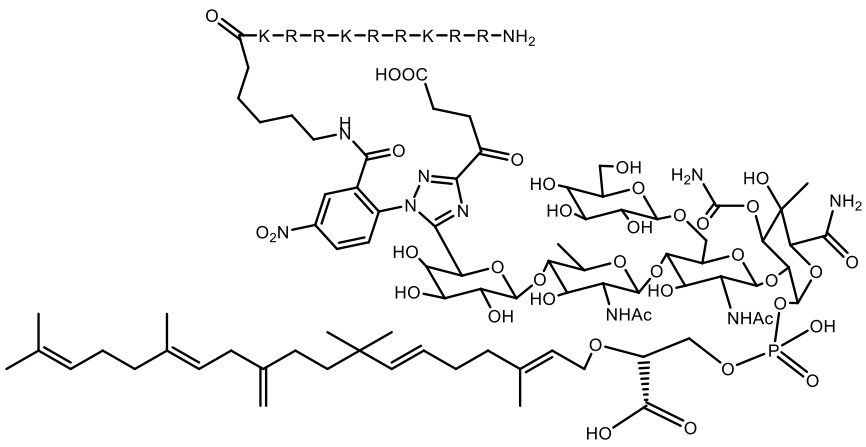


Figure 89: Structure of compound **29**

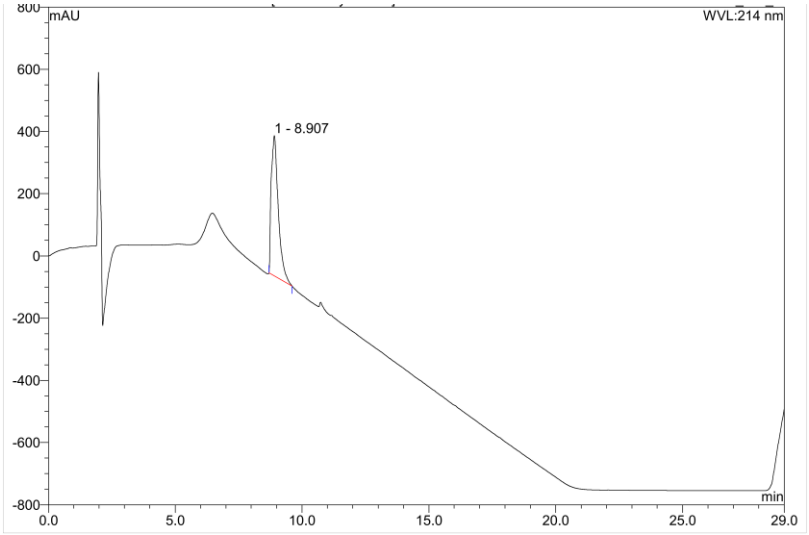


Figure 90: HPLC trace of RP-HPLC purified compound **29**, showing product peak at r.t. = 8.907 min. Gradient 0-95% ACN in 6 min.

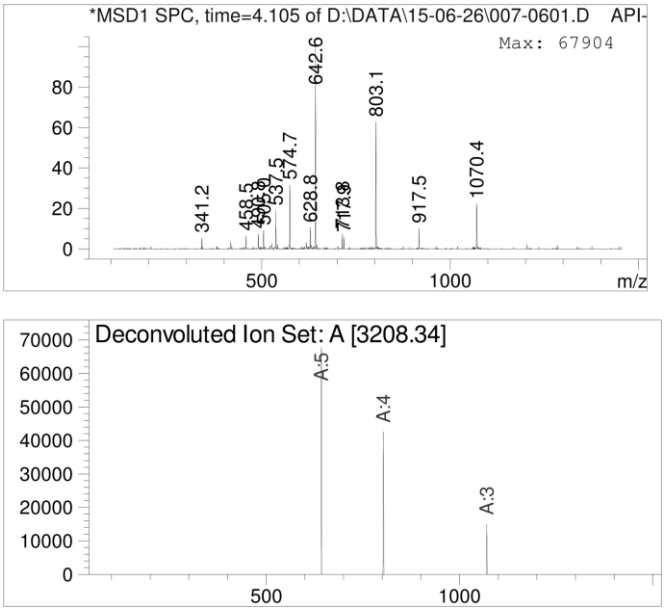
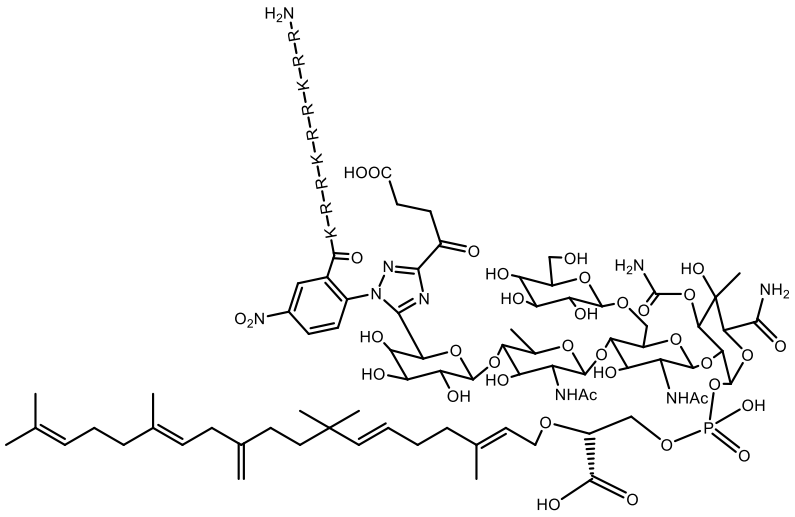


Figure 91: ESI-MS from LC-MS with deconvolution of compound **29**. Exact mass calcd. for $C_{136}H_{231}N_{40}O_{47}P$ = 3207.67, deconvoluted mass found = 3208.34.



Chemical Formula: $C_{130}H_{220}N_{39}O_{46}P$
Exact Mass: 3094.58
Molecular Weight: 3096.39

Figure 92: Structure of compound **30**

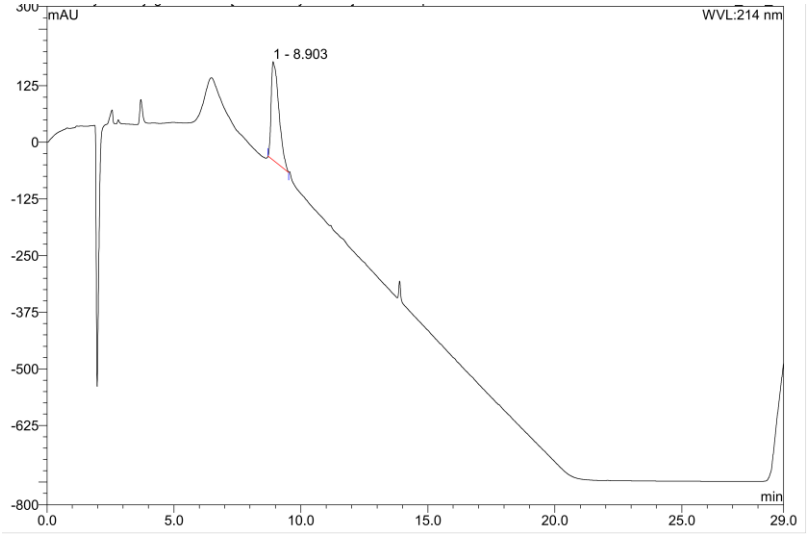


Figure 93: HPLC trace of RP-HPLC purified compound **30**, showing product peak at r.t. = 8.903 min. Gradient 0-95% ACN in 15 min.

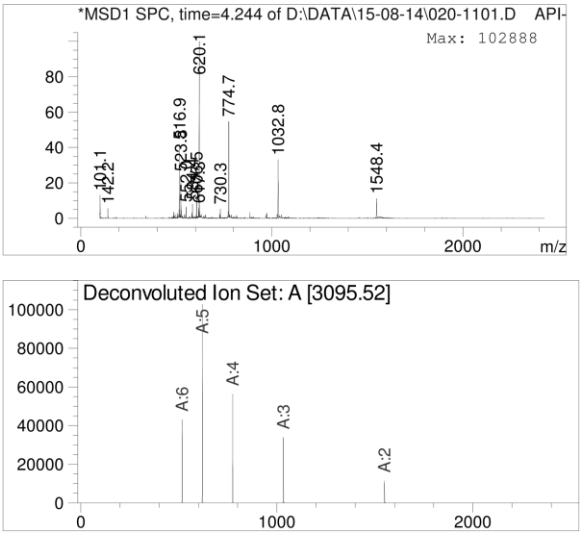
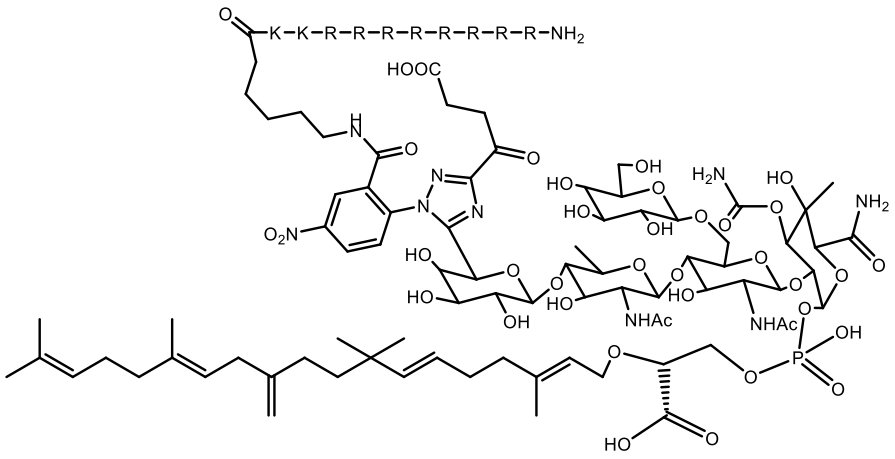


Figure 94: ESI-MS from LC-MS with deconvolution of compound **30**, **AI83-2**. Exact mass calcd. for $C_{130}H_{220}N_{39}O_{46}P$ = 3094.58, deconvoluted mass found = 3095.52.



Chemical Formula: $C_{142}H_{243}N_{46}O_{48}P$
Exact Mass: 3391.77
Molecular Weight: 3393.75

Figure 95: Structure of compound **31**

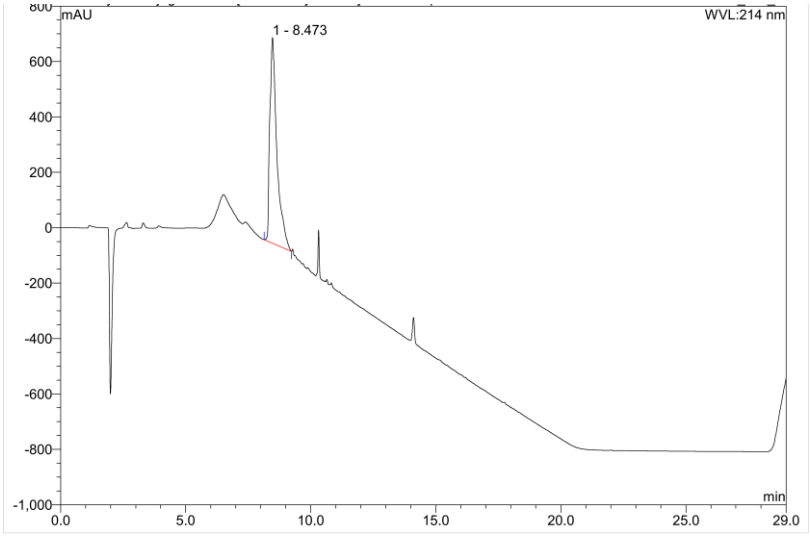


Figure 96: HPLC trace of RP-HPLC purified compound **31**, showing product peak at r.t. = 8.473 min. Gradient 0-95% ACN in 15 min.

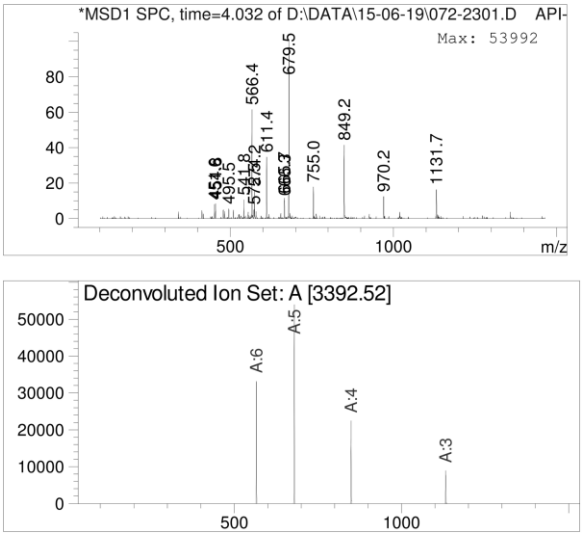


Figure 97: ESI-MS from LC-MS with deconvolution of compound **31**. Exact mass calcd. for $C_{142}H_{243}N_{46}O_{48}P$ = 3391.77, deconvoluted mass found = 3392.52.

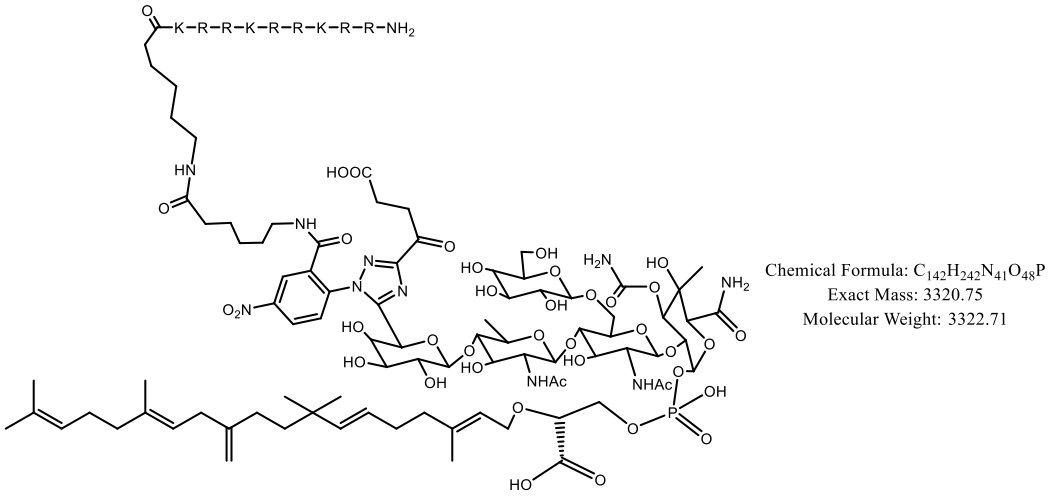


Figure 98: Structure of compound **32**

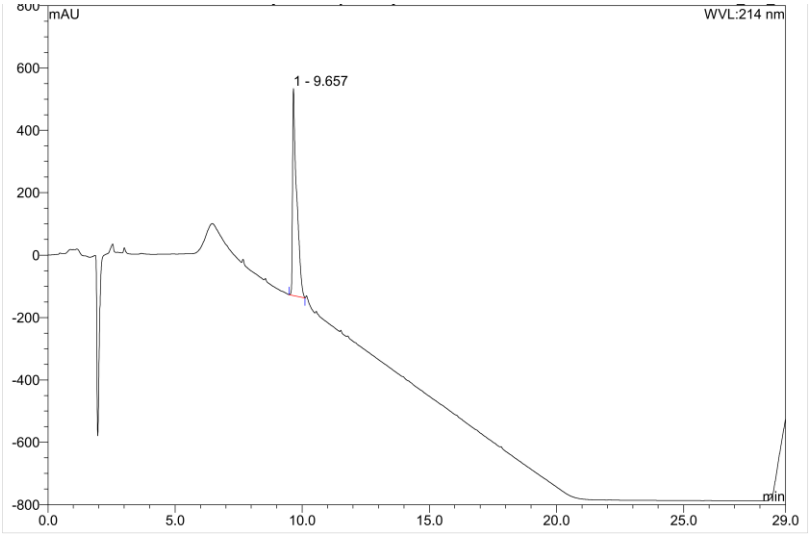


Figure 99: HPLC trace of RP-HPLC purified compound **32**, showing product peak at r.t. = 9.657 min. Gradient 0-95% ACN in 15 min.

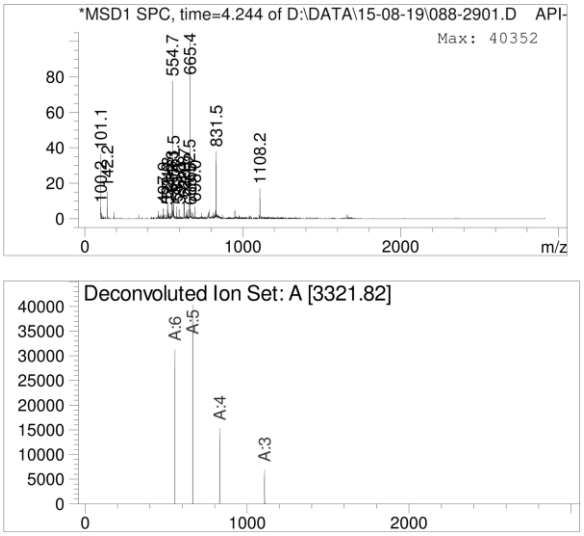


Figure 100: ESI-MS from LC-MS with deconvolution of compound **32**. Exact mass calcd. for $C_{142}H_{242}N_{41}O_{48}P$ = 3320.75, deconvoluted mass found = 3321.82.

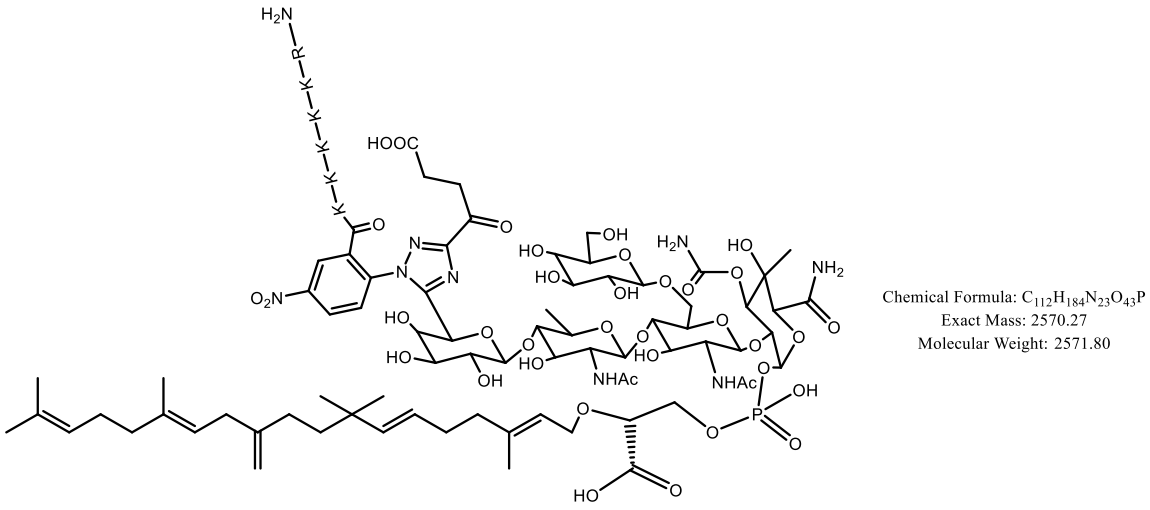


Figure 101: Structure of compound **33**

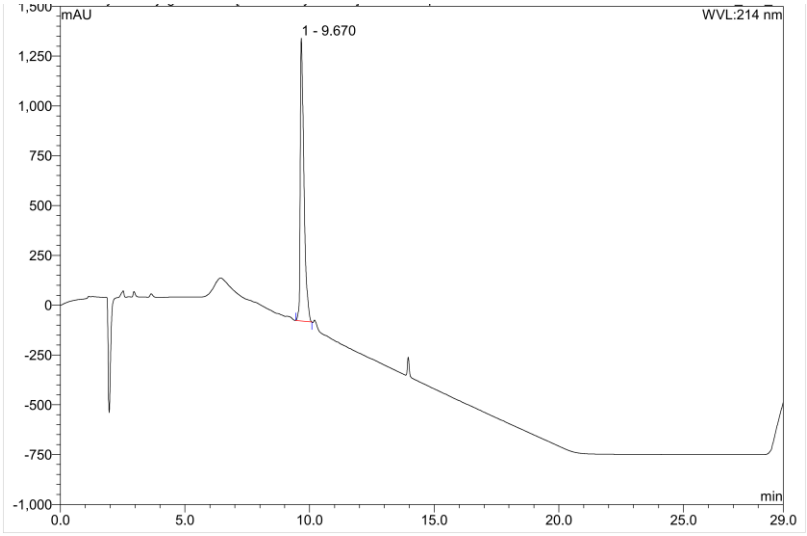


Figure 102: HPLC trace of RP-HPLC purified compound **33**, showing product peak at r.t. = 9.670 min. Gradient 0-95% ACN in 15 min.

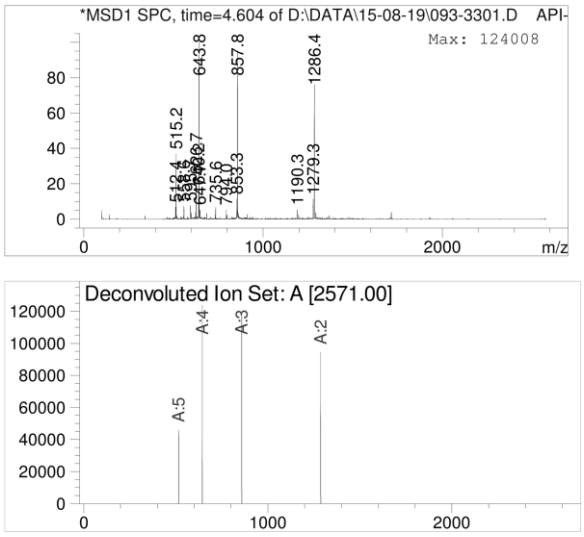
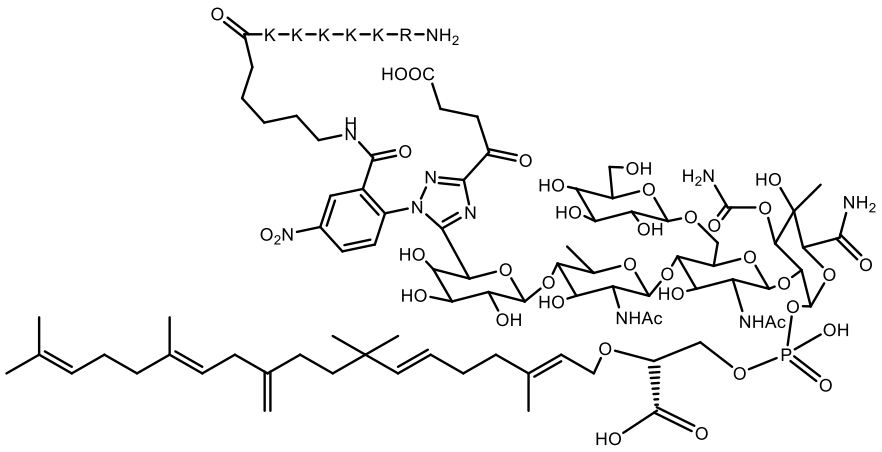


Figure 103: ESI-MS from LC-MS with deconvolution of compound **32**. Exact mass calcd. for $C_{112}H_{184}N_{23}O_{43}P$ = 2570.27, deconvoluted mass found = 2571.00.



Chemical Formula: $C_{118}H_{195}N_{24}O_{44}P$
Exact Mass: 2683.35
Molecular Weight: 2684.96

Figure 104: Structure of compound **34**

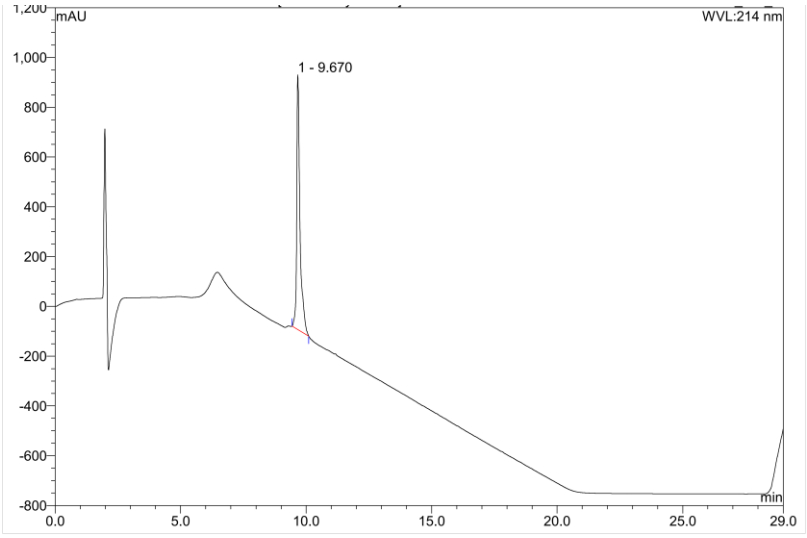


Figure 105: HPLC trace of RP-HPLC purified compound **34**, showing product peak at r.t. = 9.670 min. Gradient 0-95% ACN in 15 min.

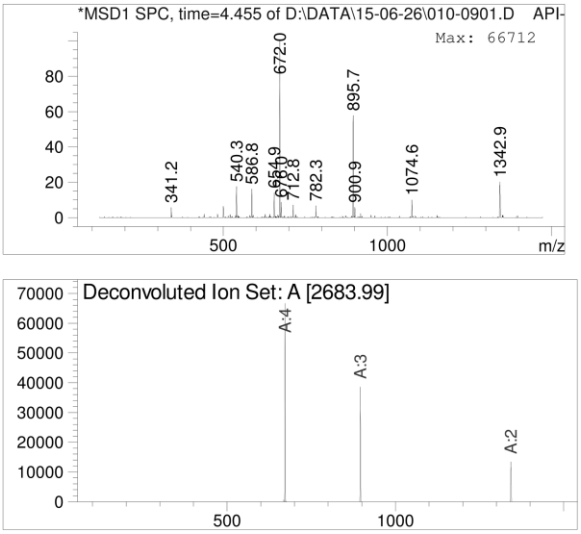


Figure 106: ESI-MS from LC-MS with deconvolution of compound **34**. Exact mass calcd. for $C_{118}H_{195}N_{24}O_{44}P$ = 2683.35, deconvoluted mass found = 2683.99.

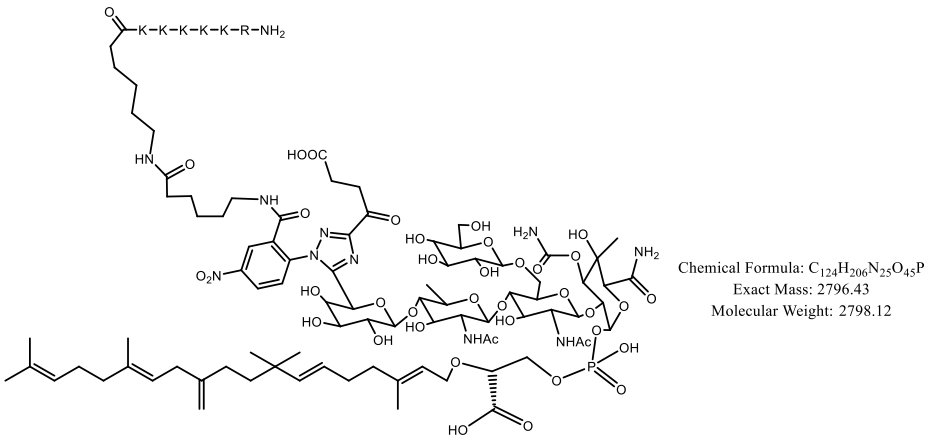


Figure 107: Structure of compound **35**

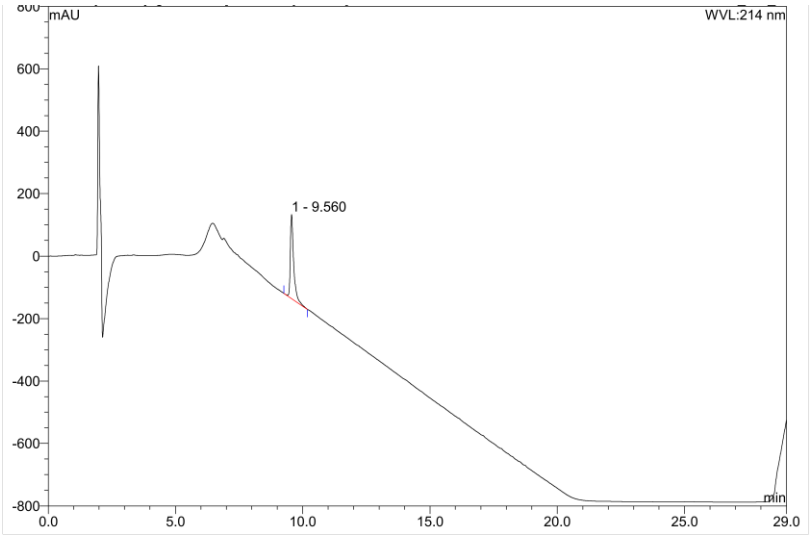


Figure 108: HPLC trace of RP-HPLC purified compound **35**, showing product peak at r.t. = 9.560 min. Gradient 0-95% ACN in 15 min.

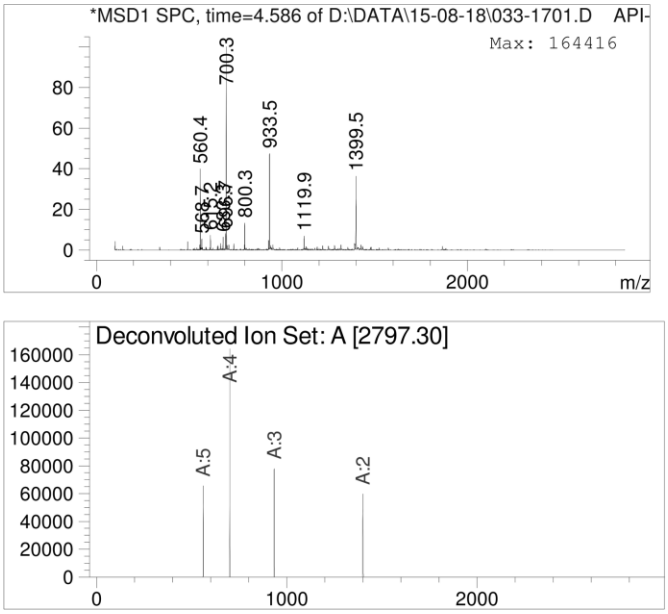
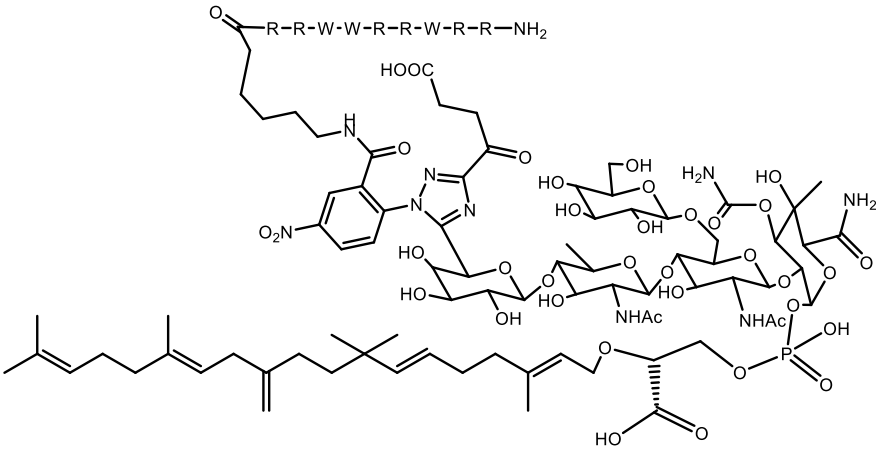


Figure 109: ESI-MS from LC-MS with deconvolution of compound **35**. Exact mass calcd. for $C_{124}H_{206}N_{25}O_{45}P$ = 2796.43, deconvoluted mass found = 2797.30.



Chemical Formula: $C_{151}H_{225}N_{40}O_{47}P$
Exact Mass: 3381.62
Molecular Weight: 3383.67

Figure 110: Structure of compound **36**

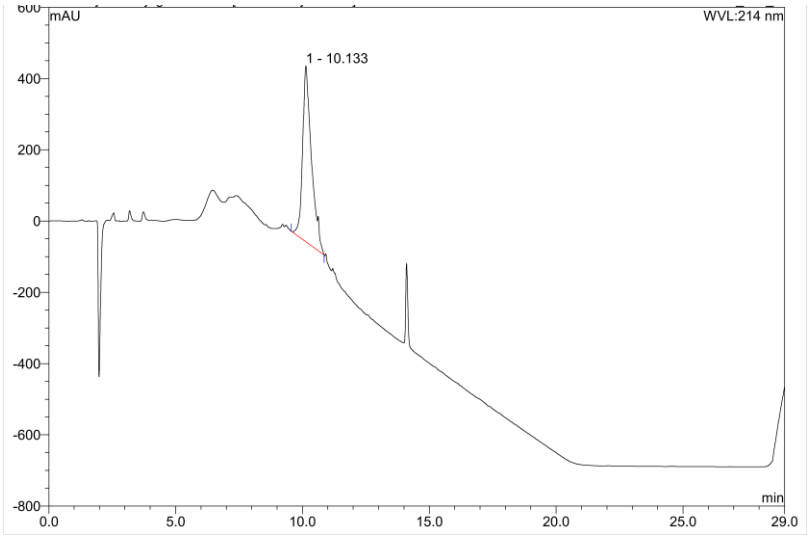


Figure 111: HPLC trace of RP-HPLC purified compound **36**, showing product peak at r.t. = 10.133 min. Gradient 0-95% ACN in 6 min.

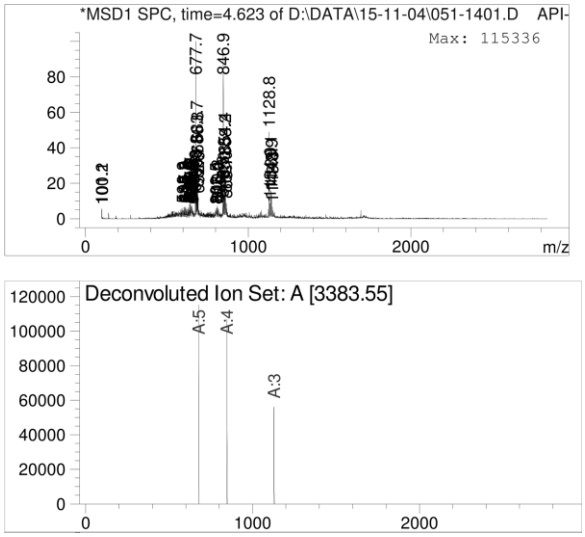
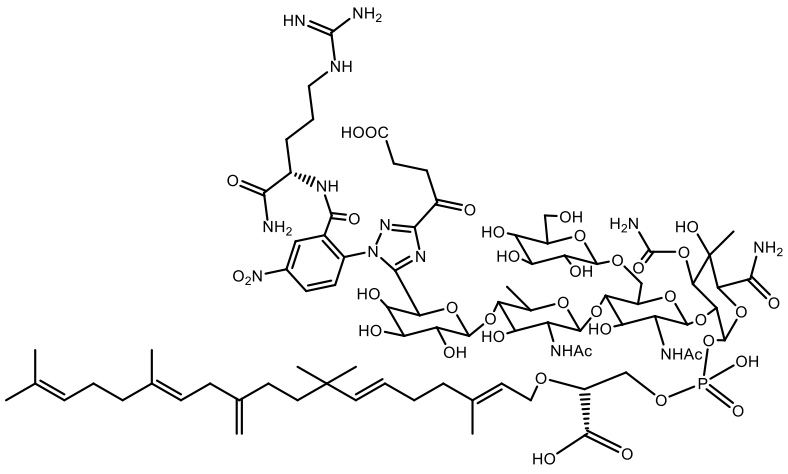


Figure 112: ESI-MS from LC-MS with deconvolution of compound **36**, **AI117-3**. Exact mass calcd. for $C_{151}H_{225}N_{40}O_{47}P$ = 3381.62, deconvoluted mass found = 3383.55.



Chemical Formula: $C_{82}H_{124}N_{13}O_{38}P$
Exact Mass: 1929.79
Molecular Weight: 1930.92

Figure 113: Structure of compound **37**

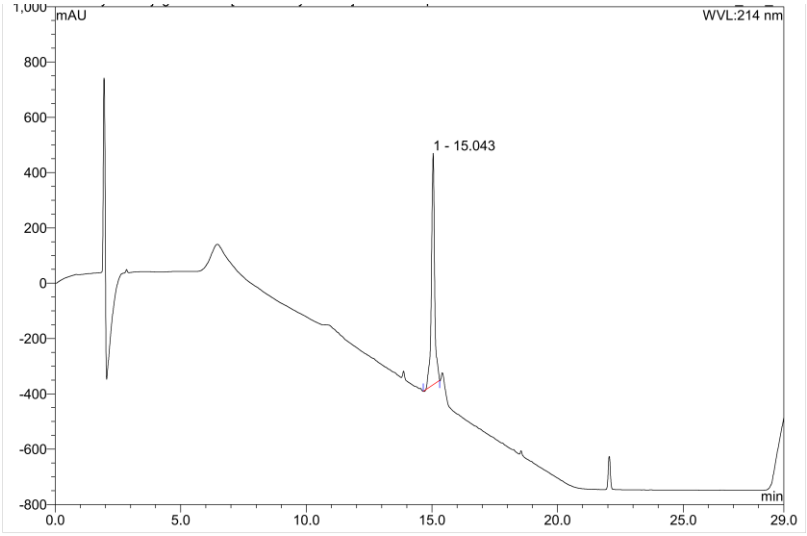


Figure 114: HPLC trace of RP-HPLC purified compound **37**, showing product peak at r.t. = 15.043 min. Gradient 0-95% ACN in 15 min

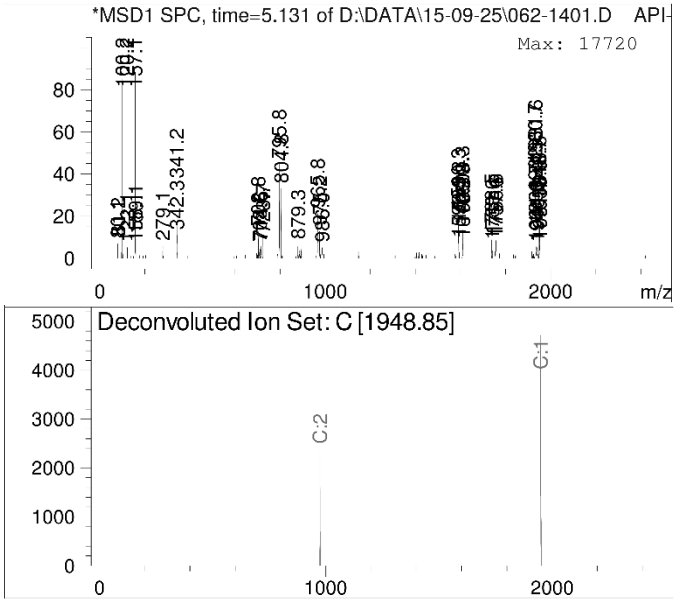


Figure 115: ESI-MS from LC-MS with deconvolution of compound **37**. Exact mass calcd. for $C_{82}H_{124}N_{13}O_{38}P$ = 1929.79, deconvoluted mass found = M + NH_4^+ = 1948.85

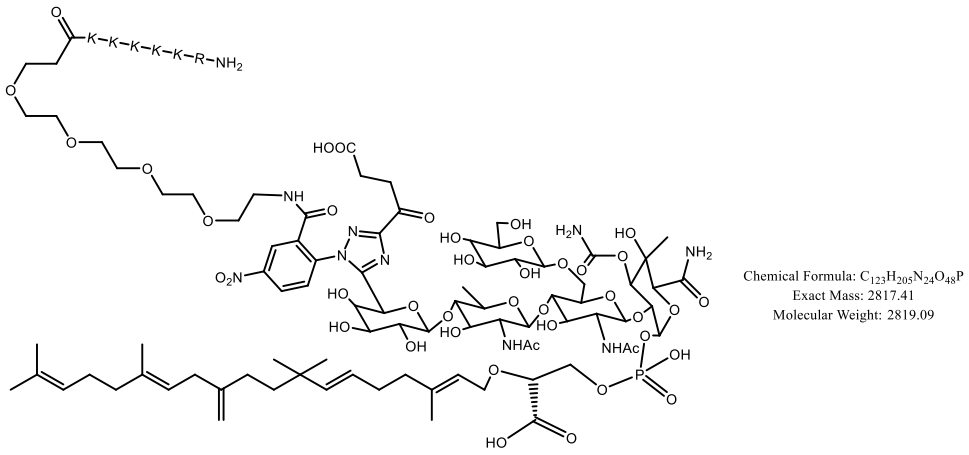


Figure 116: Structure of compound **38**

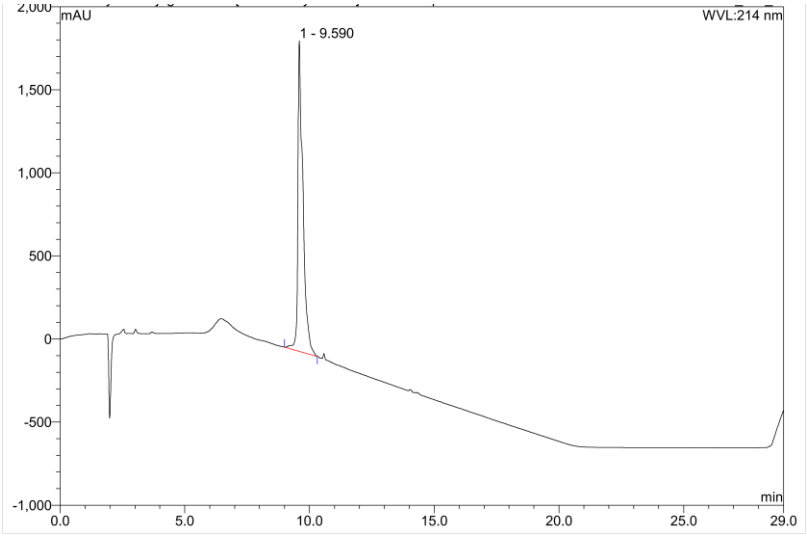


Figure 117: HPLC trace of RP-HPLC purified compound **38**, showing product peak at r.t. = 9.590. Gradient 0-95% ACN in 6 min.

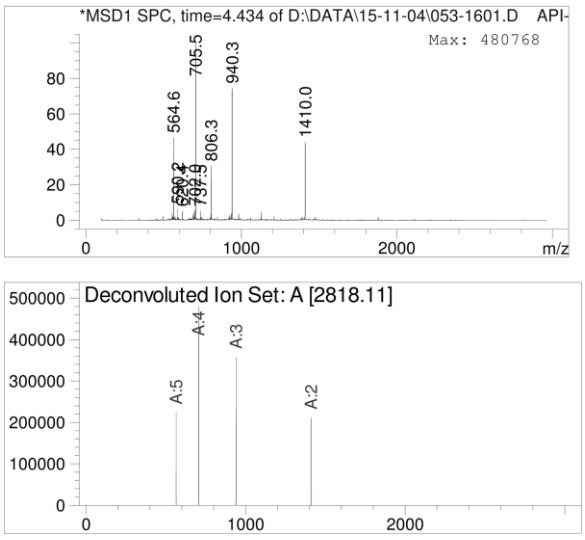
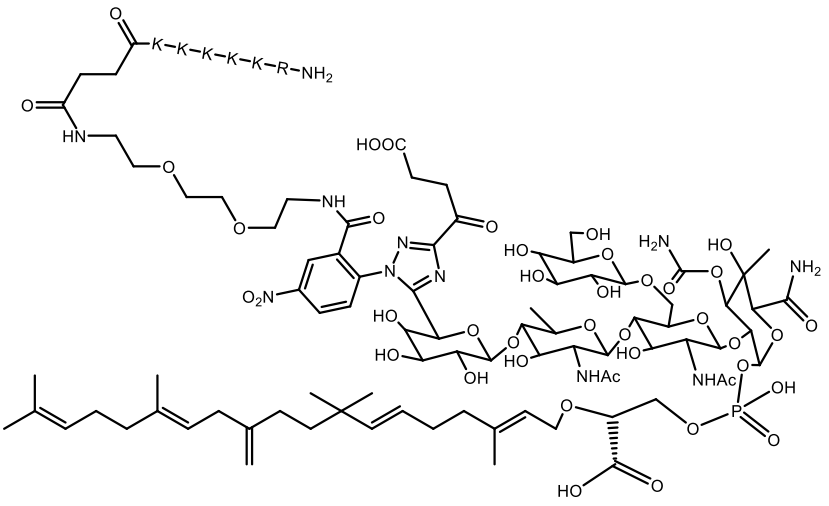


Figure 118: ESI-MS from LC-MS with deconvolution of compound **38**, **AI118-1**. Exact mass calcd. for $C_{123}H_{205}N_{24}O_{48}P$ = 2817.41, deconvoluted mass found = 2818.11.



Chemical Formula: $C_{122}H_{202}N_{25}O_{47}P$
Exact Mass: 2800.39
Molecular Weight: 2802.06

Figure 119: Structure of compound **39**

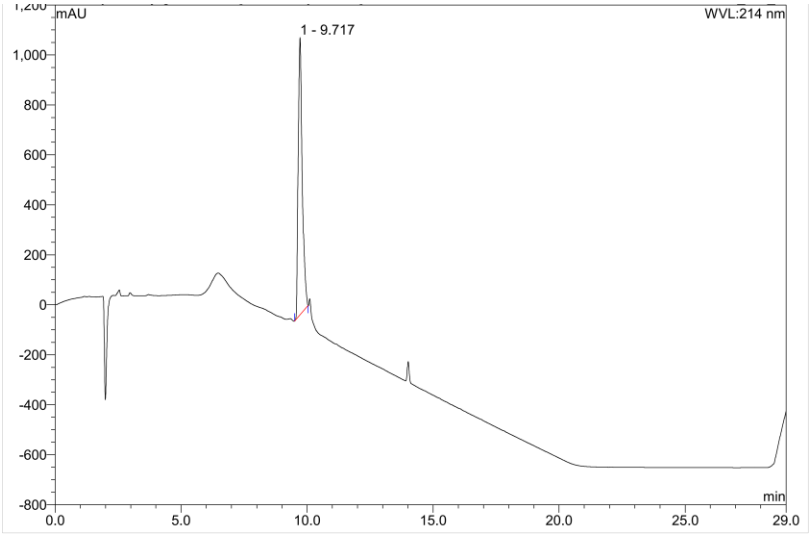


Figure 120: HPLC trace of RP-HPLC purified compound **39**, showing product peak at r.t. = 9.717 min. Gradient 0-95% ACN in 6 min.

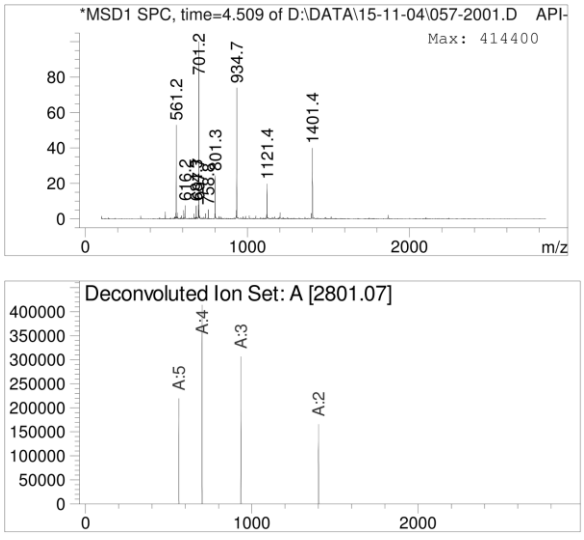


Figure 121: ESI-MS from LC-MS with deconvolution of compound **39**. Exact mass calcd. for $C_{122}H_{202}N_{25}O_{47}P$ = 2800.39, deconvoluted mass found = 2801.07.

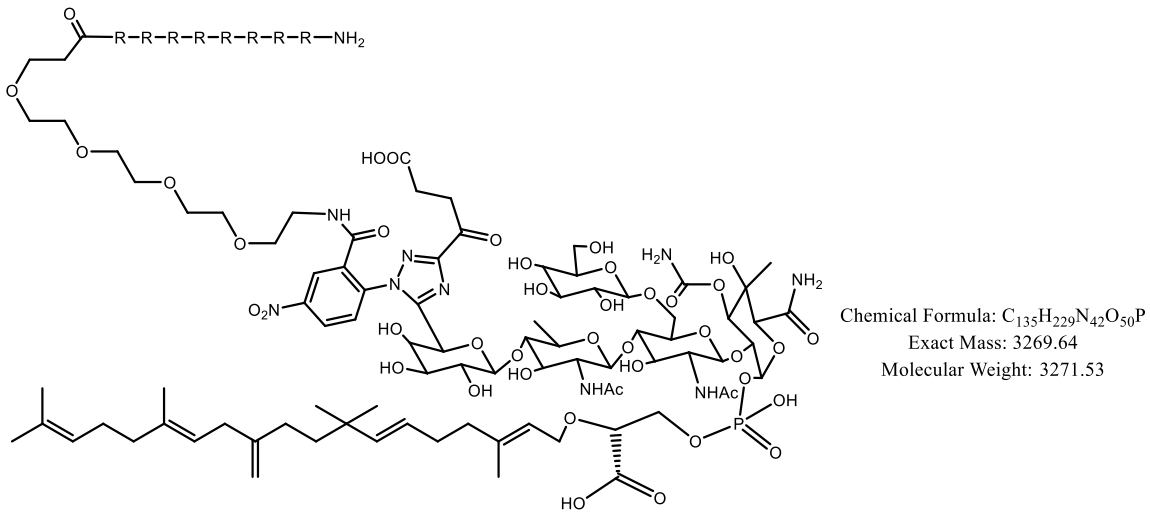


Figure 122: Structure of compound **40**

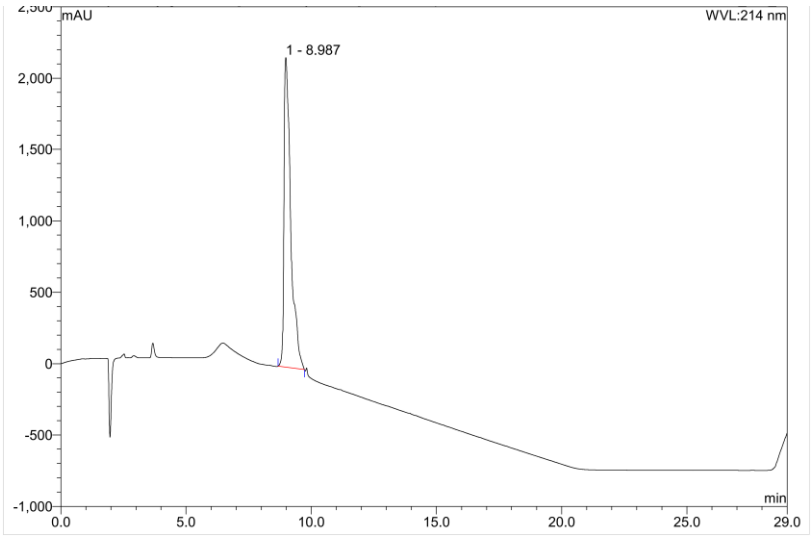


Figure 123: HPLC trace of RP-HPLC purified compound **40**, showing product peak at r.t. = 8.987 min. Gradient 0-95% ACN in 15 min.

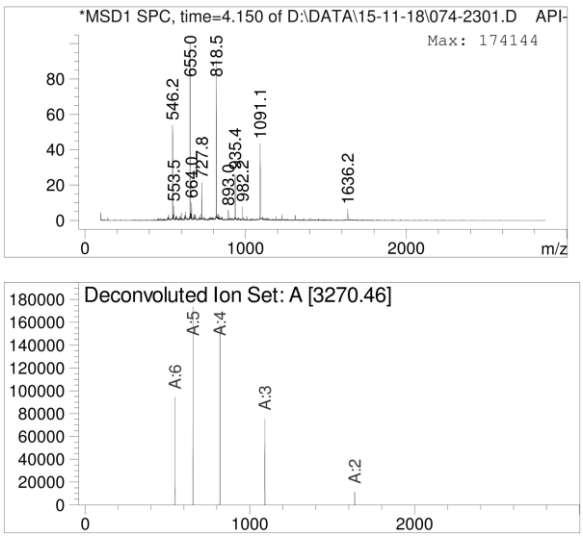
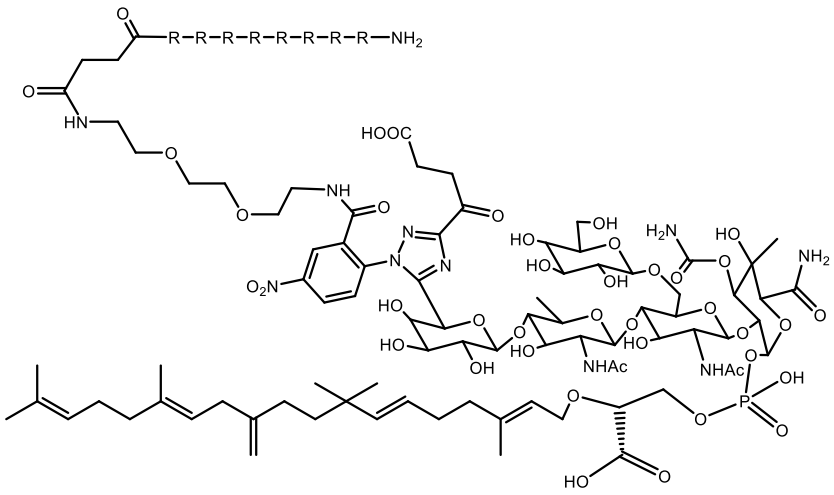


Figure 124: ESI-MS from LC-MS with deconvolution of compound **40**. Exact mass calcd. for $C_{135}H_{229}N_{42}O_{50}P$ = 3269.64, deconvoluted mass found = 3270.46.



Chemical Formula: $C_{134}H_{226}N_{43}O_{49}P$
Exact Mass: 3252.63
Molecular Weight: 3254.51

Figure 125: Structure of compound **41**

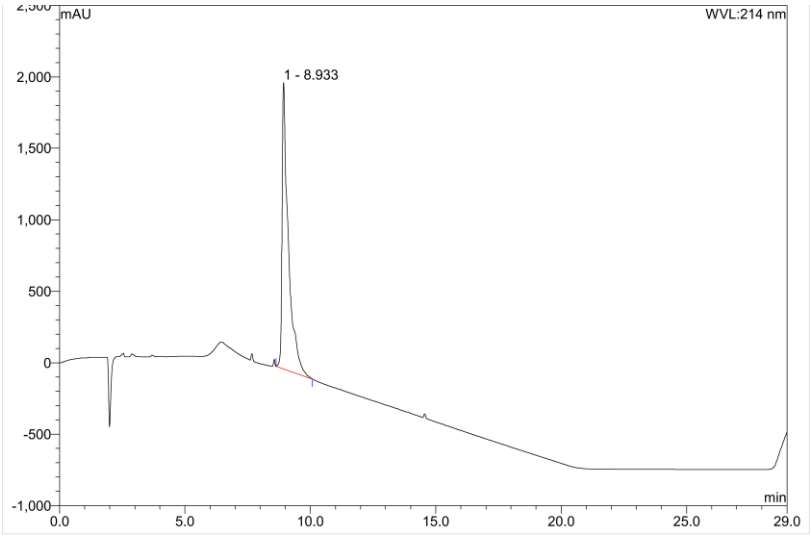


Figure 126: HPLC trace of RP-HPLC purified compound **41**, showing product peak at r.t. = 8.933 min. Gradient 0-95% ACN in 15 min.

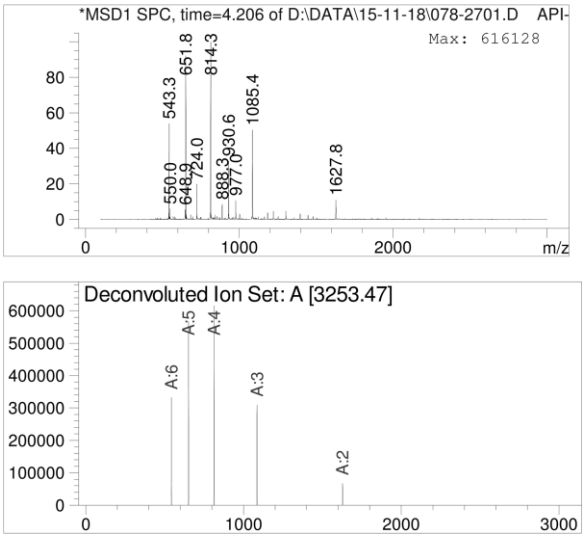


Figure 127: ESI-MS from LC-MS with deconvolution of compound **41**. Exact mass calcd. for $C_{134}H_{226}N_{43}O_{49}P$ = 3252.63, deconvoluted mass found = 3253.47.

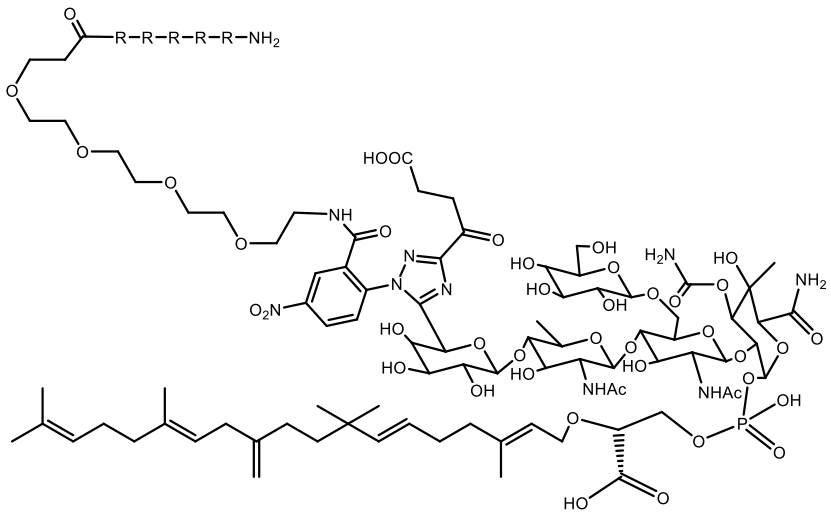


Figure 128: Structure of compound **42**

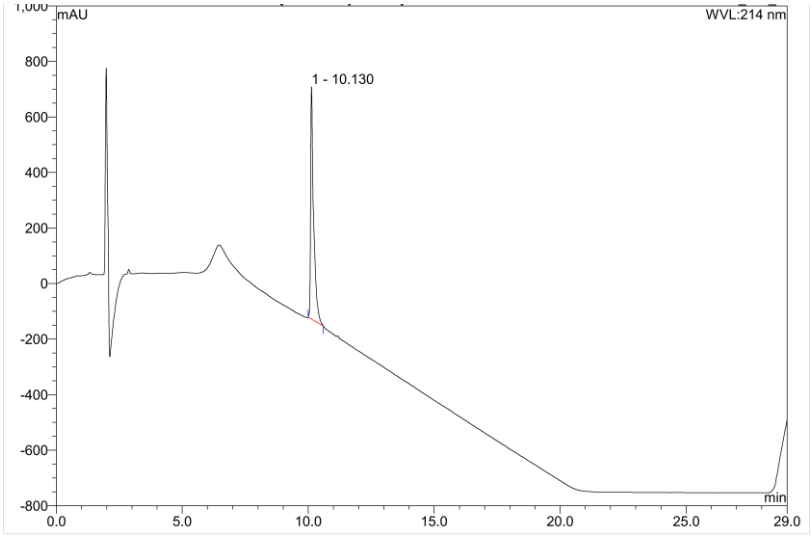


Figure 129: HPLC trace of RP-HPLC purified compound **42**, showing product peak at r.t. = 10.130 min. Gradient 0-95% ACN in 15 min.

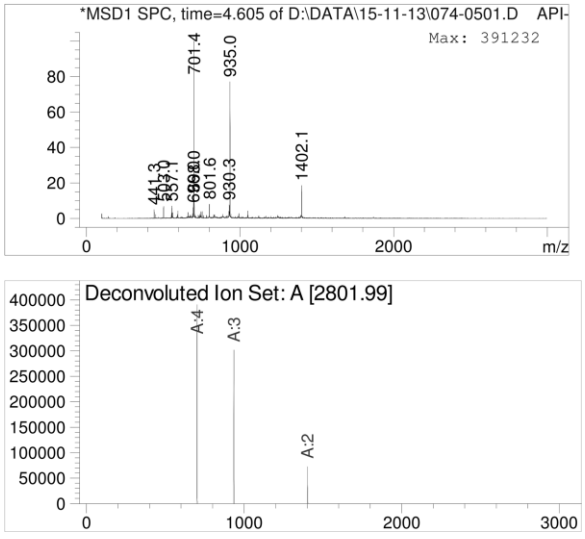


Figure 130: ESI-MS from LC-MS with deconvolution of compound **42**. Exact mass calcd. for $C_{117}H_{193}N_{30}O_{47}P$ = 2801.34, deconvoluted mass found = 2801.99.

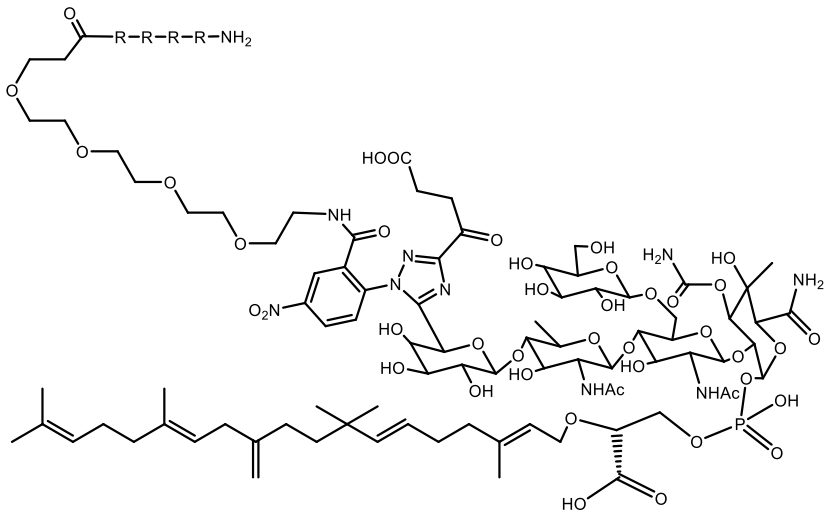


Figure 131: Structure of compound **43**

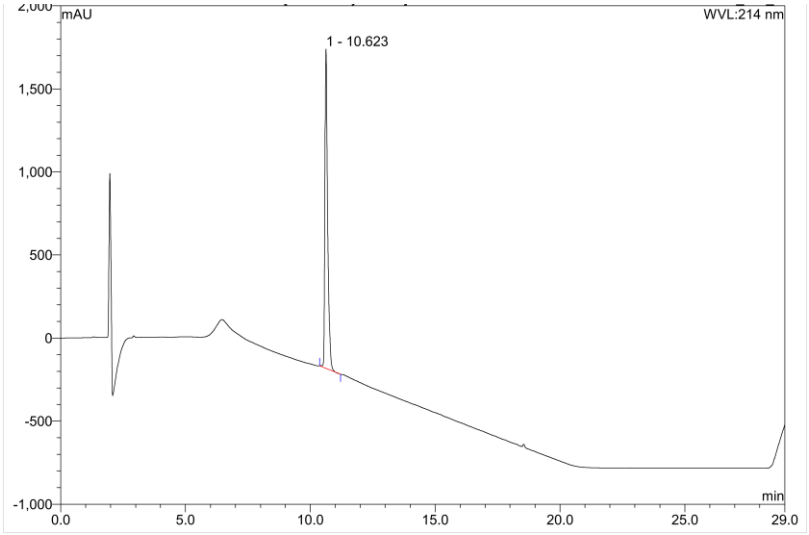


Figure 132: HPLC trace of RP-HPLC purified compound **43**, showing product peak at r.t. = 10.623 min. Gradient 0-95% ACN in 15 min.

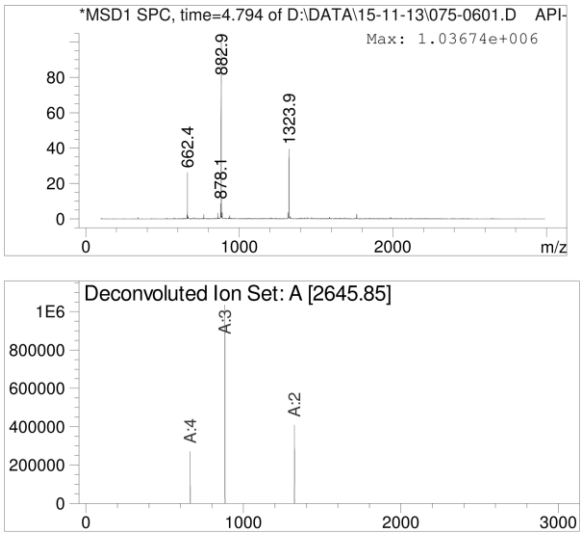
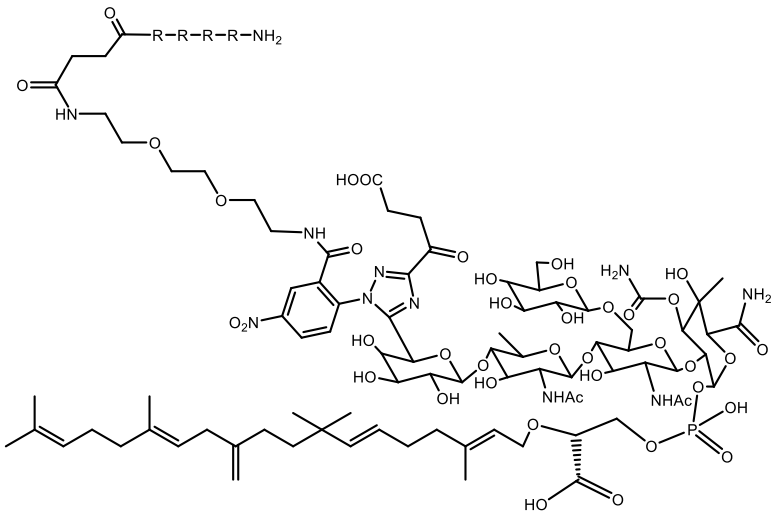


Figure 133: ESI-MS from LC-MS with deconvolution of compound **43**. Exact mass calcd. for $C_{111}H_{181}N_{26}O_{46}P$ = 2645.24, deconvoluted mass found = 2645.85.



Chemical Formula: $C_{110}H_{178}N_{27}O_{45}P$
Exact Mass: 2628.22
Molecular Weight: 2629.75

Figure 134: Structure of compound **44**

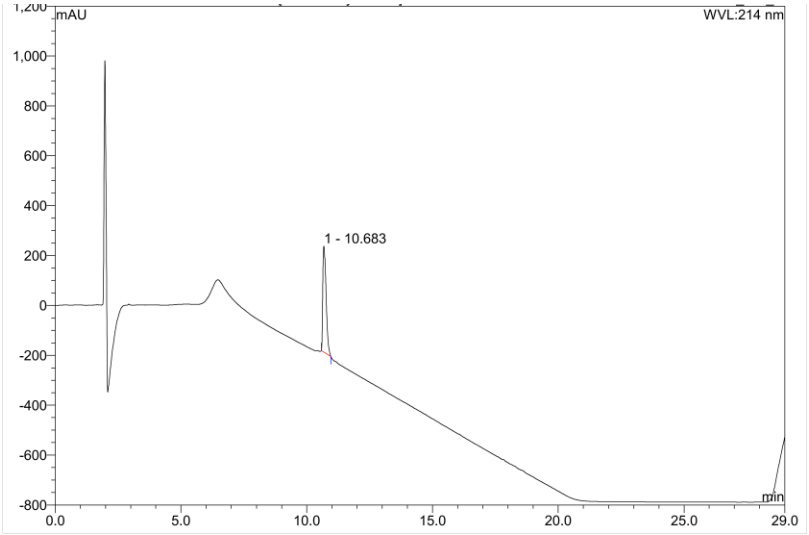


Figure 135: HPLC trace of RP-HPLC purified compound **44**, showing product peak at r.t. = 10.683 min. Gradient 0-95% ACN in 15 min.

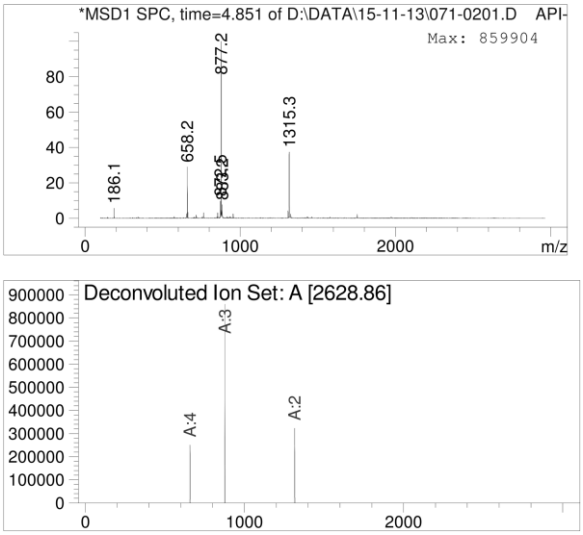


Figure 136: ESI-MS from LC-MS with deconvolution of compound **44**. Exact mass calcd. for $C_{110}H_{178}N_{27}O_{45}P$ = 2628.22, deconvoluted mass found = 2628.86.

IX. Synthesis of a polymyxin analogue & its conjugation to MoeA

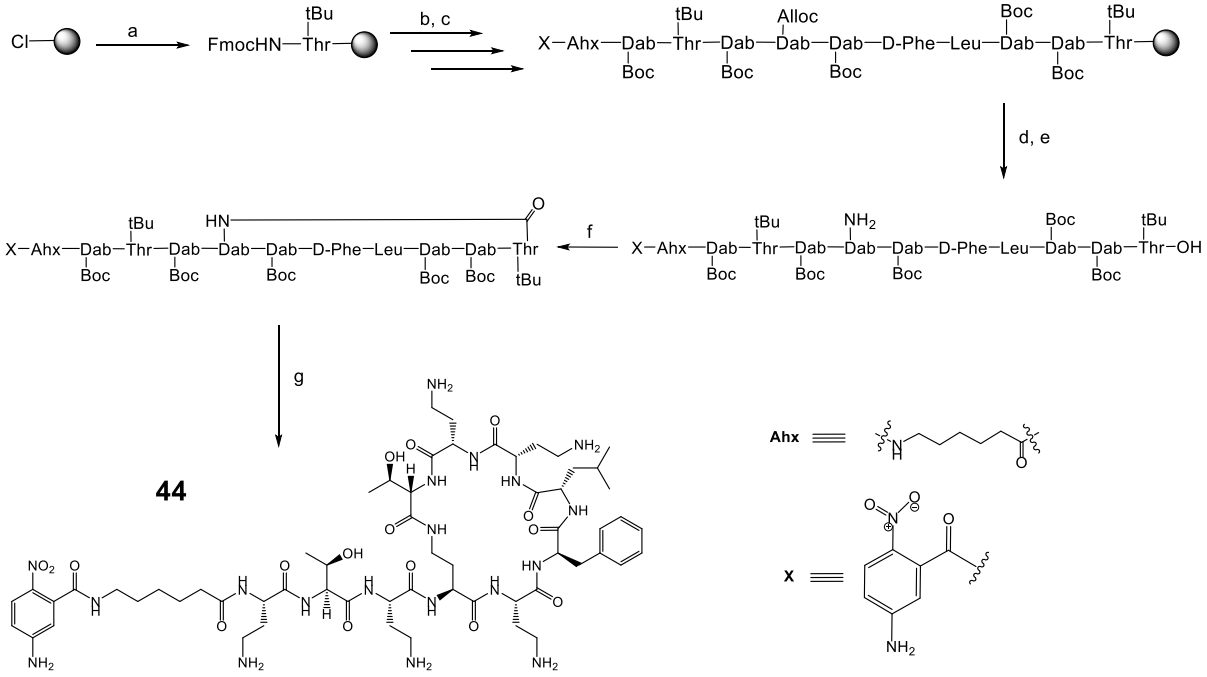


Figure 137: Scheme for the synthesis of polymyxin analogue **44.**

The resin used for the synthesis of **44** is 2-chlorotritylchloride resin, manufacturer's loading = 1.2 mmol/g, 100 mg scale. Steps/reagents are as follows: (a) 4 eq. Fmoc-Thr(tBu)-OH, 8 eq. DIPEA were dissolved in in 2 mL DCM and the mixture was added to the resin and shaken for 3 h. The resin was then washed 6 times using DCM. Unreacted resin was capped with MeOH/DIPEA/DCM = 1:1:3, 2 mL by shaking for 30 min. Loading was calculated as = 0.68 mmol/g (b) Couplings were performed using 4 eq. Amino acid, 4 eq. DIC/Oxyma in DMF using microwave assisted SPPS. (c) Fmoc deprotection was performed using 20% piperidine in DMF by shaking for 3 min using a µwave, followed by 10 min at r.t. (d) Alloc removal was performed using 0.2 eq. $Pd(PPh_3)_4$ and 24 eq. $PhSiH_3$ in DCM (2 x 1h). (e) Partial cleavage was performed using 2% TFA/5% TIS in DCM, 2h. (f) Cyclization was performed using 1 eq. HATU/10 eq. DIPEA in DMF (peptide concentration: 15 mg/mL). (g) Full cleave was performed using 95% TFA/ 2.5% TIS/2.5% H_2O, 2h. The filtrate was precipitated by adding cold diethyl ether (-20°C) and centrifuging twice with Et_2O at 7000 rpm at 0°C to obtain a yellow solid.

Crude **44** was conjugated to MoeA and the conjugate was purified using the protocols described in section IVa to obtain a polymyxin-MoeA conjugate **45**.

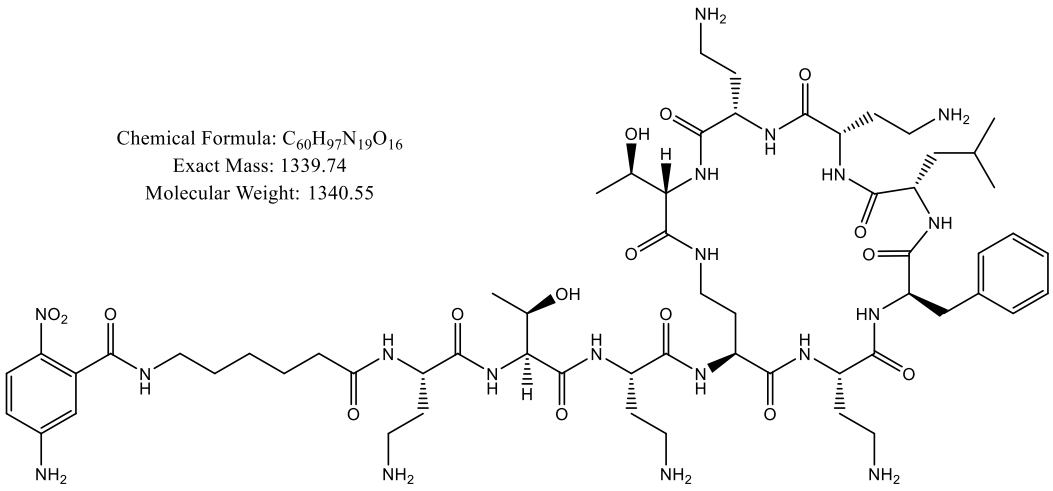


Figure 138: Structure of compound **44**

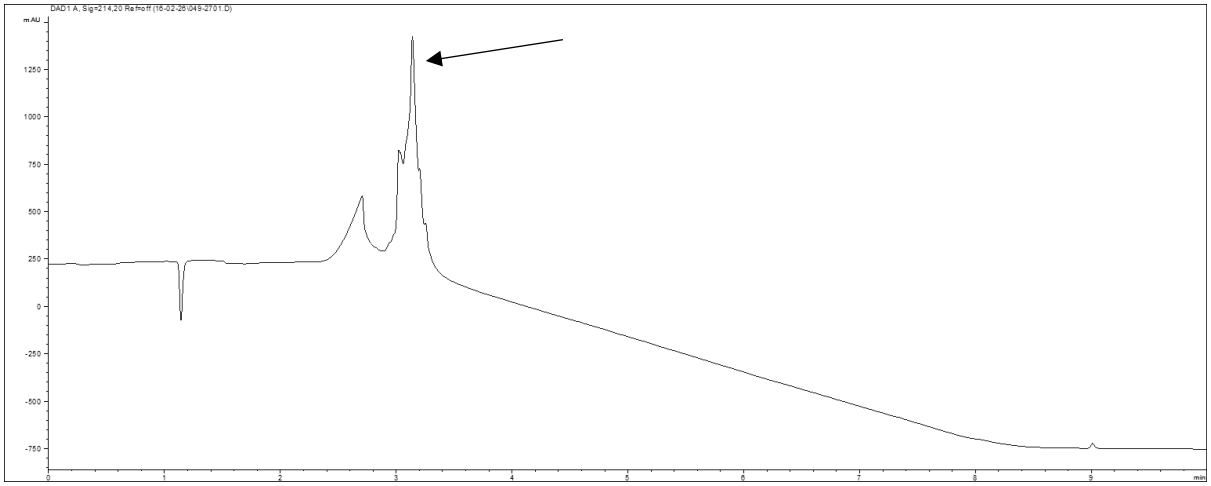


Figure 139: HPLC trace of RP-HPLC purified compound **44**, showing product peak at r.t. = 3.141 min. Gradient 0-100% ACN in 6 min.

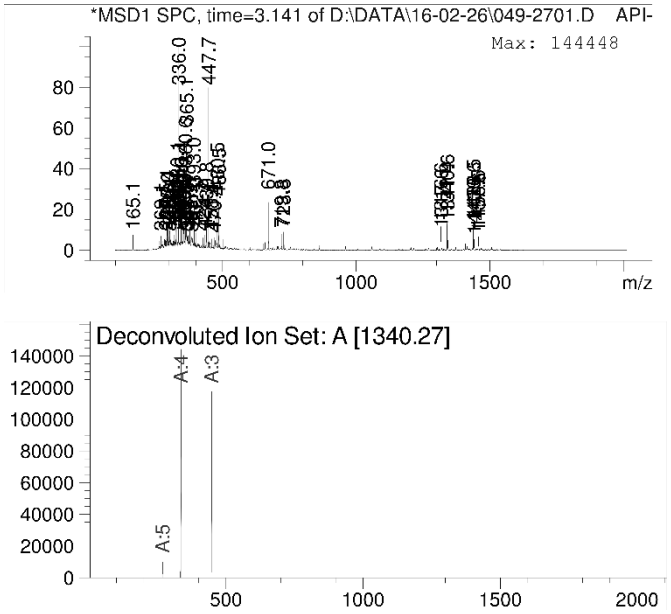


Figure 140: ESI-MS from LC-MS with deconvolution of compound **44**. Exact mass calcd. for $C_{60}H_{97}N_{19}O_{16}$ = 1339.74, deconvoluted mass found = 1340.27.

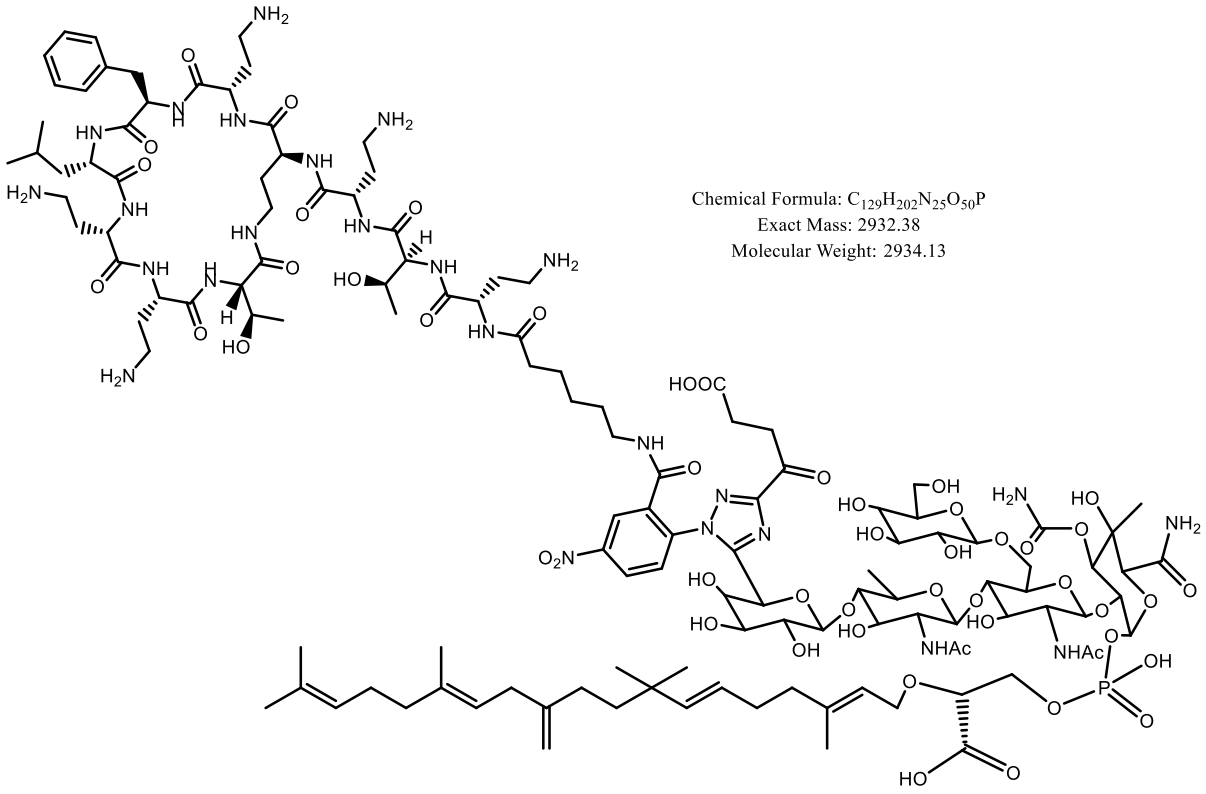


Figure 138: Structure of compound **45**

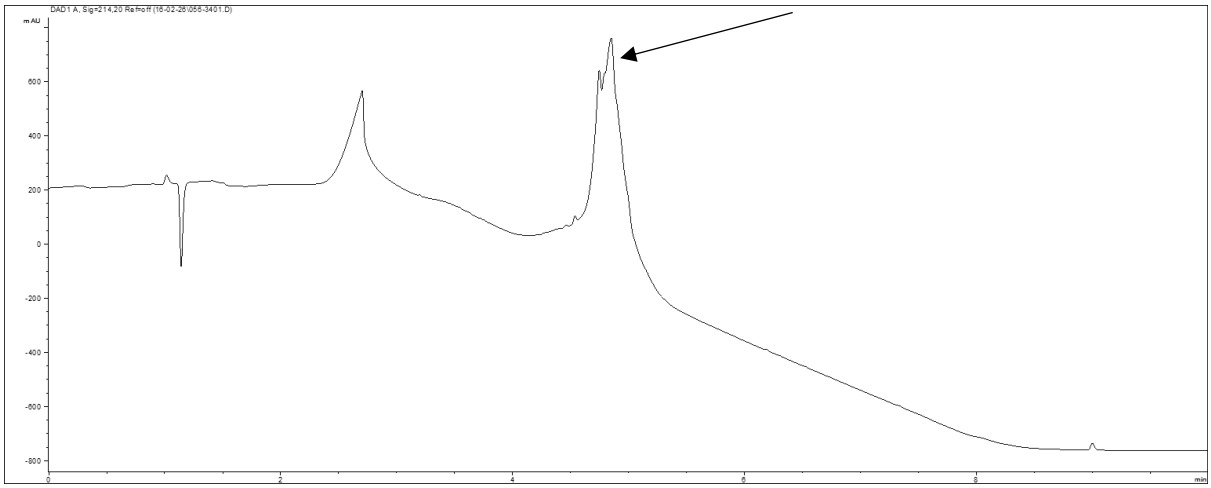


Figure 139: HPLC trace of RP-HPLC purified compound **45**, showing product peak at r.t. = 4.960 min. Gradient 0-100% ACN in 6 min.

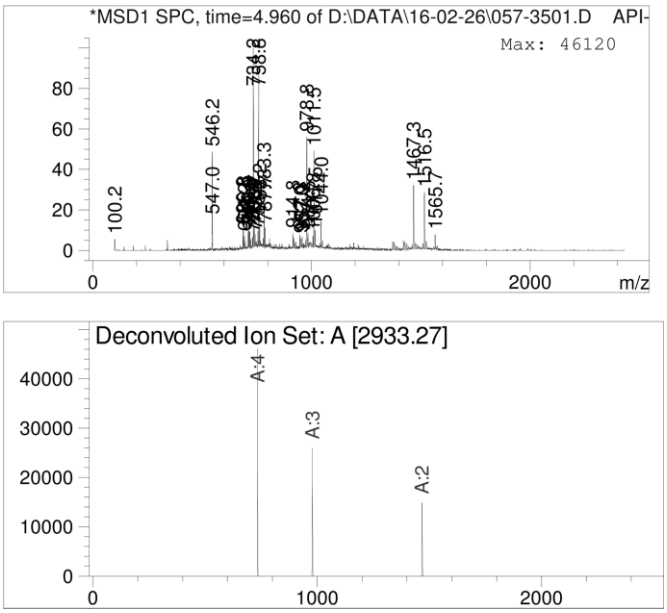


Figure 140: ESI-MS from LC-MS with deconvolution of compound **45**. Exact mass calcd. for $C_{60}H_{97}N_{19}O_{16}$ = 2933.27, deconvoluted mass found = 2933.27.

X. MIC protocols:

The protocol for the MICs are identical to those described in the experimental section of chapter 6.

ANNEX I:
NON-1ST AUTHOR PUBLICATIONS

Volume 52 | Number 36 | 4 May 2016 | Pages 6025–6198

ChemComm

Chemical Communications

www.rsc.org/chemcomm

ISSN 1359-7345

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

Edward J. Taylor, Ishwar Singh *et al.*

Efficient total syntheses and biological activities of two teixobactin analogues

175 YEARS

ChemComm

COMMUNICATION

Cite this: *Chem. Commun.*, 2016, 52, 6060

Received 13th December 2015, Accepted 9th March 2016

DOI: 10.1039/c5cc10249a

www.rsc.org/chemcomm

Efficient total syntheses and biological activities of two teixobactin analogues†

Anish Parmar,[a] Abhishek Iyer,[ab] Charlotte S. Vincent,[c] Dorien Van Lysebetten,[b] Stephen H. Prior,[d] Annemieke Madder,[b] Edward J. Taylor*[c] and Ishwar Singh*[a]

The discovery of the new antibiotic teixobactin has been timely in the race for unearthing novel antibiotics wherein the emergence of drug resistant bacteria poses a serious threat worldwide. Herein, we present the total syntheses and biological activities of two teixobactin analogues. This approach is simple, efficient and has several advantages: it uses commercially available building blocks (except AllocHN-D-Thr-OH), has a single purification step and a good recovery (22%). By using this approach we have synthesised two teixobactin analogues and established that the D-amino acids are critical for the antimicrobial activity of these analogues. With continuing high expectations from teixobactin, this work can be regarded as a stepping stone towards an in depth study of teixobactin, its analogues and the quest for synthesising similar molecules.

The decreased potency of drugs such as penicillin,[1] vancomycin[2] and oxacillin[3] due to their excessive use is a consequence of the emergence of drug resistant bacteria. It has been predicted that antimicrobial resistance (AMR) will have disastrous consequences and it is estimated that by 2050 an additional 10 million people yearly could succumb to drug resistant infections.[4] The recently published article in Nature[5] describing the discovery of teixobactin has provided a much needed breakthrough in the challenging field of antibiotic peptides. The bacteria against which it is effective have, thus far, not shown any detectable resistance to teixobactin. Moreover, given the multiple mechanisms of attack by teixobactin described by Süssmuth[6] resistance is unlikely in the near future. Unfortunately, although teixobactin provides some much needed answers, the problem is far from over. Organic chemists have worked round the clock to synthesise novel antibiotics and although progress has been steady, bacteria have time and again confounded even the best in this field.[7] This is where the multichannel device, iChip,[8] has made a considerable contribution enabling the identification of molecules like teixobactin, some of which could be active against drug resistant bacteria. However, the iChip device has its limitations. The probability of finding an antibiotic is extremely low (10^{-7} percent) and even with high-throughput screening, which takes considerable time and resources, there is no guarantee when will be the discovery of the next drug like teixobactin. Undeniably, breakthroughs *via* organic synthesis have to be made to keep the drug resistance problem under check. The time for this is indeed now and teixobactin is a good starting point. This study is aimed at doing so and provides a general approach through which, not only teixobactin, but also other analogues can be synthesised. The total synthesis of teixobactin will give access to analogues which nature cannot provide us with and therefore an organic synthesis approach to the bigger problem at hand should not be prematurely discarded. It is to be noted that during the preparation of this manuscript a report from the Albericio group appeared describing the synthesis of teixobactin analogue **1** obtained in a 6% yield. Moreover, the role of the D-amino acids have not been proven.[9] Our work presents an efficient syntheses of both the D and L versions of the analogues of teixobactin (22% yield for **1**) and their role in antimicrobial activity.

Since teixobactin has been fully characterised by NMR and LC-MS, the structural complexity of teixobactin can be described as moderate to difficult. It has also been described as an unusual depsipeptide due to the presence of the non-natural amino acids L-*allo*-enduracididine and *N*-methylphenylalanine along with four D-amino acids (Fig. 1). Despite not possessing a lipid tail, the commercial unavailability of the enduracididine amino acid makes the total synthesis of the molecule more time consuming than expected.[10] The commercially available natural amino acid arginine (a linear guanidine) is the closest structural match suitable for the replacement of the L-*allo*-enduracididine (a cyclic guanidine)

[a] *School of Pharmacy, JBL Building, University of Lincoln, Beevor St. Lincoln LN67DL, UK. E-mail: isingh@lincoln.ac.uk*

[b] *Organic and Biomimetic Chemistry Research Group, Department of Organic Chemistry, Ghent University, Krijgslaan 281 (S4), B-9000 Ghent, Belgium*

[c] *School of Life Sciences, JBL Building, University of Lincoln, Beevor St. Lincoln LN67DL, UK. E-mail: etaylor@lincoln.ac.uk*

[d] *School of Chemistry, JBL Building, University of Lincoln, Beevor St. Lincoln LN67DL, UK*

† Electronic supplementary information (ESI) available: Peptide synthesis, HPLC, LC-MS analysis, NMR spectra. See DOI: 10.1039/c5cc10249a

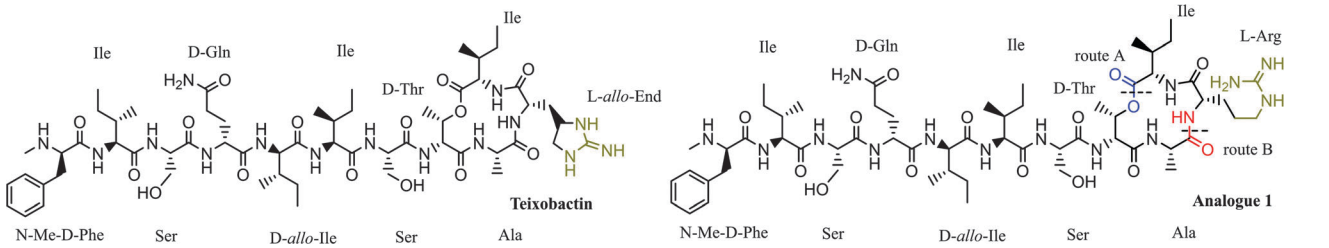


Fig. 1 (a) (left) Structure of teixobactin. (b) (right) Structure of the teixobactin analogue **1** showing the bonds to be cleaved for the synthesis routes A (in blue) and B (in red) and the structural differences (in green).

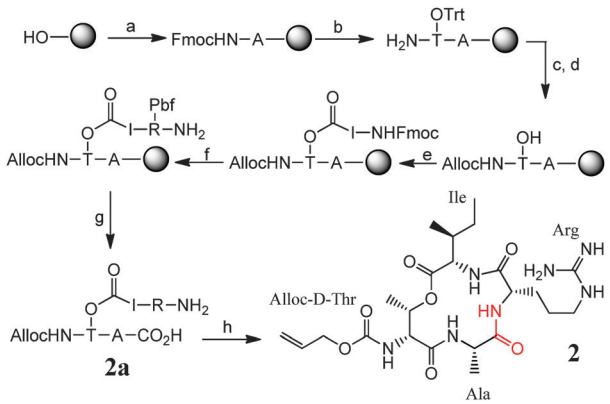


Fig. 2 Synthesis scheme for the core teixobactin fragment **2** starting from Wang resin: (a) 10 eq. Fmoc-Ala-OH, 10 eq. DIC, 1 eq. DMAP, 10% Ac2O/DIPEA in DMF followed by 20% piperidine in DMF. (b) 2.5 eq. Fmoc-D-Thr(Trt)-OH, 2.5 HATU/5 eq. DIPEA, 3 h DMF followed by 20% piperidine in DMF. (c) 4 eq. allyl chloroformate/8 eq. DIPEA in DCM, 1 h. (d) 1 : 5 : 96 TFA : TIS : DCM. 3 × 15 min. (e) 10 eq. Fmoc-Ile-OH, 10 eq. DIC, 10 mol% DMAP in DCM, 2 h followed by 10% Ac_2O/DIPEA in DMF, followed by 20% piperidine in DMF. (f) 4 eq. Fmoc-Arg(Pbf)-OH, 4 eq. HATU/8 eq. DIPEA in DMF, 1 h followed by 20% piperidine in DMF. (g) TFA : TIS : H_2O = 95 : 2.5 : 2.5, 1 h. (h) 1 eq. HATU/10 eq. DIPEA in DMF, 1 h, monitored on HPLC.

as can be seen in Fig. 1 (green). In this work, arginine was selected as a replacement of enduracididine for synthesis of teixobactin analogue **1**. A first approach towards tackling this molecule was to synthesise the complete peptide on solid phase and cyclise post-cleavage from the resin *via* an ester bond (ESI,† Fig. S1). This route has been used with success for the synthesis of the analogues of callipeltin B.[11]

Following this strategy, a fully protected linear peptide was synthesised and cleaved from the resin without cleaving the protecting groups. However, the synthesis failed at the esterification step. Several conditions have been tested for the esterification as described in Table S1 (ESI†). Unfortunately, none of them yielded the esterified product. This could be due to the steric bulk of protecting groups on the amino acids. This led to the conclusion that a direct and linear route is not the way to cyclisation. Therefore a new second synthetic route had been devised which involves cyclisation *via* amide bond formation (Fig. 1b). For the total synthesis of analogue **1**, an optimised pathway for the synthesis of the core ring structure (**2**) of teixobactin was required. Therefore, the initial efforts were focussed on obtaining this key fragment (Fig. 2).

For optimisation of the synthesis, we chose Wang resin. Fmoc-Ala-OH was loaded onto this resin *via* ester bond formation. The unreacted resin was capped using 10% acetic anhydride/diisopropylethylamine (Ac_2O/DIPEA) followed by (a) the attachment of Fmoc-D-Thr(Trt)-OH *via* amide bond formation and subsequent (b) Fmoc removal. The orthogonal Alloc protecting group was installed on the amine (c) followed by (d) trityl group removal by 1 : 5 : 96 TFA : TIS : DCM and proceeded with (e) the challenging esterification reaction between Ile and Thr. It is to be noted that excess Ile and base were required to drive the reaction to completion.[12] This was succeeded by (f) amide bond formation using Arg and subsequent (g) cleavage from the resin using 95 : 2.5 : 2.5 TFA : TIS : H_2O. The final step was (h) the amide bond formation between Arg and Ala which proceeds smoothly yielding the desired cyclised fragment. This approach was then slightly modified and used for the total synthesis of the teixobactin analogue **1** as described in Fig. 3.

The approach used in this work involves synthesising the peptide completely on solid phase and only a single purification was required after the final cleavage. The synthesis of building block AllocHN-D-Thr-OH (modified from ref. 13) was also reported for first time and which can be directly coupled to the resin without protecting the amino acid side chain. 2-Chlorotritylchloride resin was chosen instead of Wang resin as the peptide could be cleaved off from the resin without removing the side chain protecting groups of the attached amino acids.[14]

(a) The first amino acid loaded on the resin in this case is Fmoc-Ala-OH followed by (b) an amide bond coupling with Alloc-D-Thr-OH. (c) Fmoc-Ile-OH is then coupled at this stage *via* an ester bond to the free –OH side chain of threonine. Next, (d) arginine was coupled *via* an amide bond, the Fmoc protecting group is removed and (e) the N-terminus is protected *via* a trityl protecting group[15] (combining cleavage and deprotection in a single step) to facilitate the cleavage and cyclisation as described in reactions (h and i). (f) The Alloc group protecting the N-terminus of the threonine is then removed[16] and (g) the peptide chain is built *via* standard solid phase peptide synthesis (SPPS). Partial cleavage was performed using 2 : 5 : 93 TFA : TIS : DCM followed by cyclisation using 1-[bis(dimethylamino) methylene]-1*H*-1,2,3-triazolo[4,5-b]-pyridinium 3-oxid hexafluorophosphate (HATU) as a coupling reagent and DIPEA as a base in DMF for 1 h. The protecting groups are then cleaved off using 95 : 2.5 : 2.5 TFA : TIS : H_2O yielding the desired peptide (22% recovery, refer ESI,† S10). After successful synthesis of analogue **1**, the general applicability of

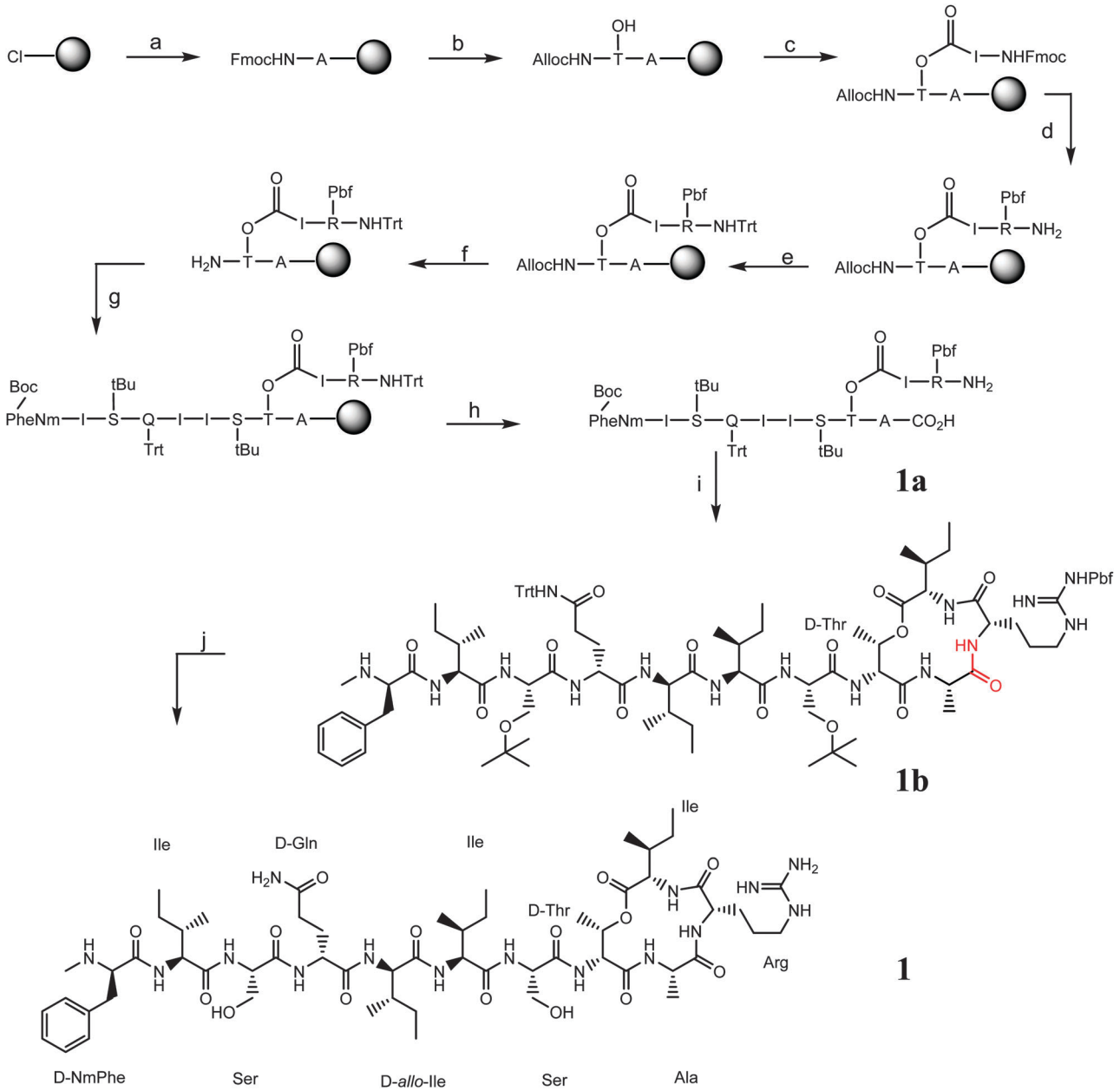


Fig. 3 Total synthesis of analogue **1** starting from 2-chlorotritylchloride resin: (a) 4 eq. Fmoc-Ala-OH/8 eq. DIPEA in DCM, 3 h. (b) 20% piperidine in DMF followed by 3 eq. AllocHN-D-Thr-OH, 3 eq. HATU/6 eq. DIPEA. (c) 10 eq. Fmoc-Ile-OH, 10 eq. DIC, 5 mol% DMAP in DCM, 2 h followed by capping with Ac_2O/DIPEA 10% in DMF, 20% piperidine in DMF (d) 4 eq. Fmoc-Arg(Pbf)-OH, 4 eq. HATU/8 eq. DIPEA in DMF, 1 h followed by 20% piperidine in DMF (e) 10 eq. Trt-Cl, 15% Et_3N in DCM, 1 h. (f) 0.2 eq. $Pd(PPh_3)_4$/24 eq. $PhSiH_3$ in DCM, 2 × 1 h. (g) Fmoc/Boc-AA(PG)-OH (AA = amino acid, PG = protecting group), HATU/DIPEA followed by 20% piperidine in DMF. (h) TFA : TIS : DCM = 2 : 5 : 93, 2 h. (i) 1 eq. HATU/10 eq. DIPEA in DMF, 1 h, monitored on HPLC. (j) TFA : TIS : H_2O = 95 : 2.5 : 2.5, 1 h.

this approach was tested for the synthesis of analogue **3** (Fig. 4). In analogue **3**, the three D-amino acids residues (Phe, Gln and Ile) were replaced by L-amino acid residues. The synthesis of analogue **3** also worked efficiently (17% recovery, refer ESI,† S10).

The detailed characterisation of **1** and **3** were performed using LC-MS and NMR. The NMR spectra of product **1** (refer ESI,† page S15) was shown to be identical as reported previously.[5,9] The NOEs of **1** were characteristic of a random coil, however, the NOEs of **3** suggested a considerable degree of structure.

The analogues **1** and **3** were evaluated for their antibacterial activity. MIC results showed a trend similar to teixobactin for analogue **1** against both Gram-positive and Gram-negative bacteria. Analogue **3** was not active against Gram-negative bacteria. Moreover, analogue **1** was 64 times more effective than analogue **3** (Table 1) against Gram-positive bacteria. This difference in antibacterial activity has established that the three D-amino acids residues of **1** are critical for the antibacterial activity.

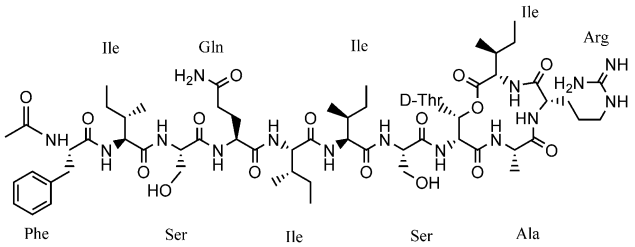


Fig. 4 Structure of analogue **3**.

Table 1 MIC (μg ml^{-1}) for **1** and **3** and Teixobactin

Entry	Organism	**1**	**3**	Teixobactin
1	*S. aureus* ATCC 25 923	2	128	0.25[a]
2	*E. coli* ATCC 25 922	64	GAW[b]	25

[a] MIC from ref. 5 was from a different strain of *S. aureus*. [b] Growth in all wells.

In conclusion, this work reports efficient total syntheses of two teixobactin analogues (22% yield of teixobactin analogue **1**). Analogue **1** is identical to teixobactin in all aspects with the exception of the L-*allo*-enduracididine amino acid. The methodology described here is not specific to only one molecule, but it can also be used as a general strategy for synthesis of other analogues of teixobactin. The role of three D-amino acids has also been established. The three D-amino acids present in the teixobactin analogue **1** but absent in analogue **3** are critical for antibacterial activity. This work also reports the synthesis of new AllocHN-D-Thr-OH building block and has incorporated it as such in the syntheses of teixobactin analogues. We believe that this work to be pivotal for the synthesis of teixobactin and its analogues and therefore will be helpful to address the current challenges of antimicrobial resistance.

The funding support from the Royal Society and Horizon 2020 is acknowledged. Anish Parmar, Abhishek Iyer and Charlotte Vincent would like to thank University of Lincoln for funding. Jan Goeman from Ghent University and Mick Lee from University of Leicester are thanked for the LC-MS/HRMS analysis. Dorien Van Lysebetten acknowledges the BOF for financial support.

Notes and references

1 F. W. Goldstein, L. D. M. Médicale, F. H. Saint-joseph and R. Losserand, *J. Antimicrob. Chemother.*, 1999, **44**, 141–144.
2 P. Courvalin, *Clin. Infect. Dis.*, 2006, **42**, 25–34.
3 J. H. Lee, *Appl. Environ. Microbiol.*, 2003, **69**, 6489–6494.
4 http://amr-review.org/, 2015.
5 L. L. Ling, T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, D. E. Hughes, S. Epstein, M. Jones, L. Lazarides, V. A. Steadman, D. R. Cohen, C. R. Felix, K. A. Fetterman, W. P. Millett, A. G. Nitti, A. M. Zullo, C. Chen and K. Lewis, *Nature*, 2015, **517**, 455–459.
6 F. von Nussbaum and R. D. Süssmuth, *Angew. Chem., Int. Ed.*, 2015, **54**, 6684–6686.
7 L. V. J. Blair, J. M. A. Webber, M. A. Baylay, A. J. Ogbolu and D. O. Piddock, *Nat. Rev. Microbiol.*, 2015, **42**, 42–51.
8 D. Nichols, N. Cahoon, E. M. Trakhtenberg, L. Pham, a. Mehta, a. Belanger, T. Kanigan, K. Lewis and S. S. Epstein, *Appl. Environ. Microbiol.*, 2010, **76**, 2445–2450.
9 Y. E. Jad, G. A. Acosta, T. Naicker, M. Ramtahal, A. El-Faham, T. Govender, H. G. Kruger, B. G. De La Torre and F. Albericio, *Org. Lett.*, 2015, **17**, 6182–6185.
10 W. Craig, J. Chen, D. Richardson, R. Thorpe and Y. Yuan, *Org. Lett.*, 2015, **17**, 4620–4623.
11 M. Kikuchi, Y. Watanabe, M. Tanaka, K. Akaji and H. Konno, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 4865–4868.
12 M. J. Martin, R. Rodriguez-Acebes, Y. Garcia-Ramos, V. Martinez, C. Murcia, I. Digon, I. Marco, M. Pelay-Gimeno, R. Fernández, F. Reyes, A. M. Francesch, S. Munt, J. Tulla-Puche, F. Albericio and C. Cuevas, *J. Am. Chem. Soc.*, 2014, **136**, 6754–6762.
13 M. Acid, L. Acid, N. Shibata, J. E. Baldwin, A. Jacobs and E. Wood, *Tetrahedron*, 1996, **52**, 12839–12852.
14 K. Barlos, D. Gatos, J. Kallitsis, G. Papaphotiu, P. Sotiriu, Y. Wenqing and W. Schäfer, *Tetrahedron Lett.*, 1989, **30**, 3943–3946.
15 C. Gros, C. Boulègue, N. Galeotti, G. Niel and P. Jouin, *Tetrahedron*, 2002, **58**, 2673–2680.
16 P. Grieco, P. M. Gitu and V. J. Hruby, *J. Pept. Res.*, 2001, **57**, 250–256.

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2015, **51**, 17552

Received 24th August 2015,
Accepted 12th October 2015

DOI: 10.1039/c5cc07097j

www.rsc.org/chemcomm

Sequence-selective DNA recognition and enhanced cellular up-take by peptide–steroid conjugates†

Yara Ruiz García,[a] Abhishek Iyer,[a] Dorien Van Lysebetten,[a] Y. Vladimir Pabon,[b] Benoit Louage,[c] Malgorzata Honcharenko,[d] Bruno G. De Geest,[c] C. I. Edvard Smith,[b] Roger Strömberg[d] and Annemieke Madder*[a]

Several GCN4 bZIP TF models have previously been designed and synthesized. However, the synthetic routes towards these constructs are typically tedious and difficult. We here describe the substitution of the Leucine zipper domain of the protein by a deoxycholic acid derivative appending the two GCN4 binding region peptides through an optimized double azide–alkyne cycloaddition click reaction. In addition to achieving sequence specific dsDNA binding, we have investigated the potential of these compounds to enter cells. Confocal microscopy and flow cytometry show the beneficial influence of the steroid on cell uptake. This unique synthetic model of the bZIP TF thus combines sequence specific dsDNA binding properties with enhanced cell-uptake. Given the unique properties of deoxycholic acid and the convergent nature of the synthesis, we believe this work represents a key achievement in the field of TF mimicry.

Gene expression at the transcriptional level is mainly regulated by proteins that bind DNA in a sequence-specific manner. These proteins, known as transcription factors (TFs), are responsible for controlling the transfer of genetic information from DNA to mRNA. More specifically, many oncoproteins are transcription factors responsible for cell-growth proliferation and tumor formation.[1] As a consequence of the specificity of these oncoproteins in the DNA sequence recognition during transcription, several approaches have been explored to develop inhibitors or modifiers of gene expression that can prevent specific genes from being transcribed.[2]

In addition, the lack of a general recognition code for the interaction between amino acid sequences within a protein and its specific DNA-binding site has promoted the study of the structure of TFs and their interaction with the DNA.[3] We here present our efforts towards mimicking the GCN4 bZIP TF (General Control Protein Leucine Zipper Transcription Factor) by a simplified synthetic construct and towards understanding how these protein mimics behave in a cellular context.[4] The GCN4 bZIP TF, which controls the activation of several genes in response to amino acid starvation in yeast, has been chosen as model system in order to allow comparison with previously published TF miniaturisation attempts. Many eukaryotic transcription factors feature homologous protein sequences forming a family generally referred to as the basic-region-leucine-zipper or bZIP motif of which the mammalian ATF/CREB family of transcription factors represent a large group. Our target has been to develop a new type of site-specific DNA binders which can recognize dsDNA by specific binding in the major groove and additionally show enhanced cellular uptake by exploiting the unique properties of the steroid moiety. We have been inspired by the bZIP (leucine zipper) and the b-HLH-ZIP (helix–loop–helix leucine zipper) motifs, in which the basic DNA recognition region binds to the major groove as a dimer, inserting two α-helices held in the correct position by a dimerization domain.[5] The main residues of the protein involved in the DNA recognition comprise the amino acids 226–248 of the basic region of the GCN4 protein. Previously, models of such transcription factors have been synthesized by different research groups employing a series of different dimerization domain substitutes. Pioneering work was carried out by the group of Kim,[6] developing an analogue in which the complete dimerization domain was substituted by a simple disulfide bond. Building on this successful idea of miniaturisation, Morii, Schepartz, Mascareñas and our group have enforced the proof-of-concept by using a variety of small dimerizing moieties.[7–12] Subsequently, we have shown that the attachment of the basic region peptides to a rigid scaffold, a derivative of deoxycholic acid in this case, also allows

[a] *Organic and Biomimetic Chemistry Research Group, Department of Organic and Macromolecular Chemistry, Ghent University, Krijgslaan 281 (S4), B-9000 Ghent, Belgium. E-mail: annemieke.madder@ugent.be*
[b] *Clinical Research Center, Department of Laboratory Medicine, Karolinska University Hospital Huddinge, SE-141 86, Stockholm, Sweden. E-mail: edvard.smith@ki.se*
[c] *Department of Pharmaceutics, Ghent University, Harelbekestraat 72, 9000 Ghent, Belgium. E-mail: br.degeest@ugent.be*
[d] *Department of Biosciences and Nutrition (BioNut), H2. Karolinska Institutet, Novum 141 83 Huddinge, Stockholm, Sweden. E-mail: Roger.Stromberg@ki.se*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5cc07097j

selective recognition of DNA. Indeed, our previous work on cMyc-Max b-HLH-ZIP and GCN4-bZIP proteins showed that this type of steroid-based constructs show potential for binding DNA.[10,13] The specific choice of the steroid scaffold was inspired by the fact that it is inexpensive, commercially available, versatile and easy to modify synthetically. In addition, its known ability to enhance proteolytic stability of attached peptides,[14–16] amphiphilicity,[17] the conformational properties ensuring correct positioning of the two appended chains[18] and the potential to increase cellular uptake and bioavailability[19] render it an excellent candidate for the attachment of the two DNA recognizing arms of the zipper motif of GCN4. The incorporation of a spacer between the peptides and the steroid scaffold was shown important in order to provide the final conjugate with enough flexibility to adopt an optimal conformation for specific interaction with the major groove of the target DNA sequence.[13] However, the concomitant further lengthening of the 23-mer peptide chains caused increased tendency of peptide aggregation necessitating the use of microwave assistance for successful synthesis.[13] In order to synthesize these peptidosteroids in a more effective and convenient way, we here explored a convergent approach involving the CuAAC reaction (copper-catalyzed alkyne–azide cycloaddition) to conjugate the recognition domains to a functionalized steroid scaffold. The geometric, steric and electronic properties of the 1,2,3-triazole resemble a *trans*-amide bond, while also affording resistance to enzymatic degradation,[20–23] hydrolysis and oxidation. In addition, successful replacement of two amino acids in α-helical peptides by a triazole unit has been shown to not significantly influence the secondary peptide structure.[20] To date the triazole linkage has scarcely been used for peptide–steroid conjugation and only to assemble short apolar tripeptides onto bile acid scaffolds,[24] despite the increasing interest of these type of conjugates in diverse applications such as HIV inhibitors[25,26] and immunogens for vaccine development.[27]

In the current work we have designed and synthesized four different scaffolds for peptide dimerization of the GCN4 basic region, which was made possible by optimizing conditions for the CuAAC mediated conjugation of the long, unprotected and functionalized zipper peptides (Fig. 1). A small series of deoxycholic acid derivatives was conceived, differing with respect to

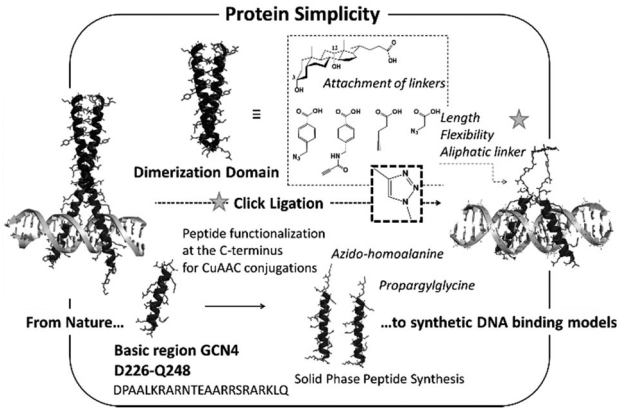


Fig. 1 Artificial DNA binder design.

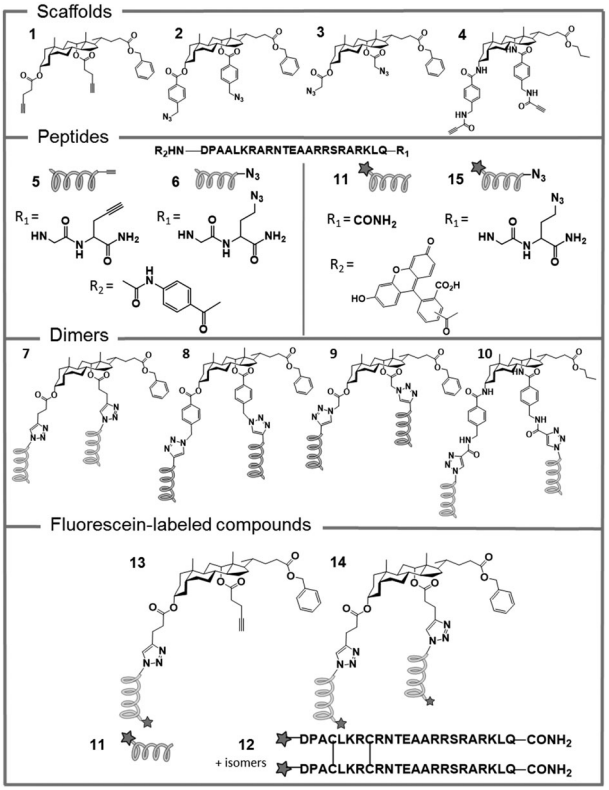


Fig. 2 Deoxycholic acid derivatives for the substitution of the dimerization domain with the corresponding peptides.

spacers between the peptide and the steroid skeleton, which have different lengths, rigidities and functionalities (Fig. 2). Hereto, commercially available deoxycholic acid was modified at the alcohol positions by attachment of different linkers. The linkers chosen for the study encompass pentynoic acid, azido glycine, 4-azidomethyl-benzoic acid and (*N*-propynoylamino)-*p*-toluic acid (PATA). The PATA linker has been specifically developed for bioconjugation purposes as an active alkyne for preparation of peptide–oligonucleotide conjugates *via* CuAAC.[28] Functionalization of the steroid nucleus was performed by Steglich esterification, affording final scaffolds (**1–3**). In case of PATA as linker (**4**), the diamino derivative of deoxycholic acid proved necessary for the coupling of the linker, as esterification gave rise to byproducts due to the high reactivity of the alkyne.

The GCN4 basic DNA binding region consists of 23 amino acids that specifically recognize the ATF/CREB binding site (5′-ATGA C/G TCAT-3′), which is the functional target of GCN4 *in vivo* and involved in inducing amino acid biosynthesis in yeast.[29]

In order to append the peptides to the bile acid scaffolds, peptides were modified at the C-terminus with unnatural amino acids bearing an alkyne or an azide. Peptides **5** and **6** were synthesized in an automated fashion using Fmoc/*t*Bu based solid phase peptide synthesis (SPPS). The alkyne and azide functionalized GCN4 basic region peptides, **5** and **6** were then attached to the central steroid core, affording four different transcription factor models. With this hydrophilic deprotected peptide and the hydrophobic scaffold, the DMSO/H_2O combination

was found to be optimal for the CuAAC reaction. As a catalyst, $Cu(CH_3CN)_4PF_6$ gave the best results. A high excess of catalyst was needed possibly due to complexation of copper with the nitrogen containing side chains of the peptide. However, reaction proceeds well with excess of copper ion complex and the copper ions can be readily removed after reaction by use of EDTA.[28] An excess of scaffold was also required for complete reaction of the peptide, the dipodal construct being favoured over the monopodal one under these conditions. The reaction was complete after 3 hours at room temperature and compatible with the presence of all unprotected amino acids in the sequence. Purification of final constructs after completion of the reaction was possible *via* RP-HPLC, affording compounds **7–10** in high purity for DNA binding studies. DNA binding affinity of mimics **7–10** was evaluated using an Electrophoretic Mobility Shift Assay (EMSA). The study was based on titration of a duplex DNA sequence containing the ATF/CREB recognition site with increasing concentrations of compounds **7–10** (Fig. 3).

No clear specific binding to DNA could be observed for mimics **8** and **10** (gel not shown). On the other hand, EMSA for compounds **7** and **9** in the absence of any competitor DNA shows that both compounds have a high affinity for the CRE DNA sequence (Fig. 3A). A closer visual inspection of the gels reveals that **7** has slightly higher DNA binding affinity as compared to **9**. Due to the more or less complete up-shifting of bands even at the lowest concentration, accurate determination of the dissociation constant (K_D) was not possible under these conditions. However, by a simple visual inspection we can conclude that almost all the DNA is completely bound at 50 nM in case of compound **7** and at 62.5 nM in case of **9**. Decreasing the peptide concentration to a value <50 nM results in large standard deviations in K_D calculations. Also, decreasing the ratio of DNA : peptide further without lowering the DNA concentration is not feasible due to problems in detection of the isotope. Therefore we conducted a second series of EMSA studies which involved a competitor DNA which allows for both checking of nonspecific interactions and calculation of the K_D. EMSA for compounds **7** and **9** in the presence of 500 nM competitor DNA reveal that nonspecific interactions do exist, however the specificity of the compounds towards the CRE sequence is still present to a large extent. This can be seen from the gradual gel shift in contrast to the previous gel (Fig. 3B). In support of this is that K_D calculations from competition experiments with 0.5–2 μM competitor DNA give values of 4–8 $\times$ 10 nm for **7** and 1–3 $\times$ 10^2 nm for **9** (see ESI†). We can also state here more conclusively that **7** is clearly not only the better DNA binding construct but also shows greater specificity as compared to **9**.

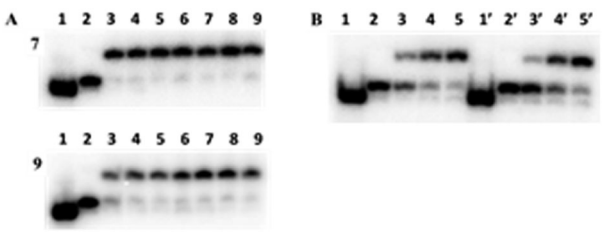


Fig. 3 (A) EMSA titration of the dipodal peptidosteroid conjugates **7** and **9** to the 5′-labeled ^{32}P-CRE sequence (5′-CGG ATG ACG TCA TTT TTT TTC-3′) and its complementary strand (5′-GAA AAA AAA TGA CGT CAT CCG-3′) at 5 nM: first lane in all the gels: pyrimidine strand (5′-labeled ^{32}P-CRE sequence). Lanes 2–9 contain peptide concentrations of 0, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.1125 and 0.125 μM for **7** and **9**. (B) EMSA titration of the dipodal peptidosteroid conjugates **7** and **9** to the 5′-labeled ^{32}P-CRE sequence (5-CGG ATG ACG TCA TTT TTT TTC-3′) at 5 nM in the presence of competitor DNA sequence (5′-AGCAGAGGGCGTGGGGGAAAAGAAAAAAG ATCCACCGGTCGCCAC-3′) at 500 nM: first lane in all the gels: pyrimidine strand. Lanes 2 and 2′ to 5 and 5′ contain peptide concentrations of 0, 0.05, 0.125 and 0.312 μM for **7** and **9** respectively.

A series of experiments were set up to investigate the cell up-take capacity of the various synthetic constructs. For this purpose the best DNA binding cholic acid based mimic **7** of the GCN4 protein was resynthesized incorporating a fluorescein tag to give **14**. The properties of this construct were compared to the labelled but otherwise unmodified GCN4 peptide dimer **12** and monomer **11** as well as the monomeric steroid conjugate **13**. The toxicity (MTT assay), quantification of uptake (flow cytometry) and intracellular localization of these constructs were studied on RAW264.7 mouse macrophages, which have been already used for cell penetration studies with a similar GCN4 peptide.[30] Compounds **12** and **14** showed a cell viability at 0,25 μM of 88% and 86% respectively (ESI†) and both are taken up at 37 °C but to a different degree. There is clearly an enhanced uptake when deoxycholic acid is used as a scaffold, as seen from the mean fluorescence values obtained by flow cytometry analysis, as they are more than four times higher for **14** than for **12** (Fig. 4II). The monopodal cholic acid derivative **13**, although non-DNA binding and more hydrophobic, also exhibits enhanced uptake. However, a decreased uptake is seen when the concentration is increased to 1 μM (ESI†). This could be attributed to the denaturant like properties of the cholic acid. The most interesting results were obtained when comparing the localisation of the DNA binding cholic acid dimer **14** at 4 °C and 37 °C (Fig. 4IA and B). At 4 °C, where endocytosis is blocked, only binding to the cell membrane was observed. This resulted in a high percentage of peptide positive cells (Fig. 4II left and middle). However, at 37 °C, whereby both passive and active uptake is possible, the uptake is considerably higher as evidenced by the higher mean cell fluorescence. This indicates that the deoxycholic acid coupled peptides are mainly internalized *via* active transport at 37 °C. The most likely ex-planation is that the peptides follow an endocytotic pathway.

In conclusion, we here have illustrated a strategy to conjugate relatively long, unprotected peptides to bile acid scaffolds in a convergent manner. From the four models of the GCN4 bZIP transcription factor presented, the one with the most flexible linker (7) proved to be the best synthetic DNA binder. This can be attributed to the fact that the linker allows the construct to grip the major groove of the DNA like a pair of tweezers which is less optimal in the case of the other linkers, which may be too long, too short, or too inflexible. The binding affinity of **7** is comparable with that of earlier described models of GCN4 TFs with K_Ds in the nM range, while the synthetic route is considerably less

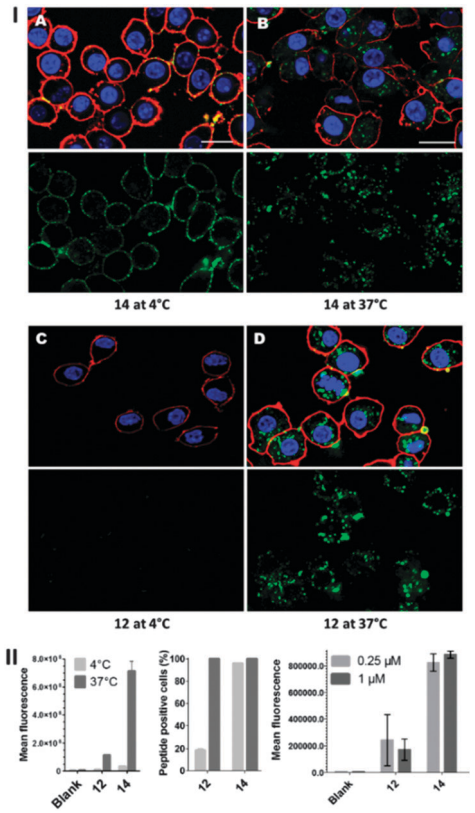


Fig. 4 (I) Confocal microscopy of RAW264.7 cells incubated with (A) **14** at 4 °C, (B) **14** at 37 °C, (C) **12** at 4 °C, and (D) **12** at 37 °C at 0.25 μM (green fluorescence signal). Cell nuclei were labeled with Hoechst (blue) and cell membranes with AlexaFluor647 conjugated cholera toxin subunit B (red). The lower panels only show the green fluorescence channel. (Scale bar = 20 μm). (II) Flow cytometry analysis of mean cell fluorescence (left) and peptide positive cells (middle) of the synthesized compounds **12** and **14** incubated with RAW264.7 cells at 4 °C and 37 °C at 0.25 μM. Mean fluorescence of compounds **12** and **14** at 0.25 μM and 1 μM.

complicated. In addition, compound 7 is more readily taken up by cells than non-steroid constructs as evidenced by the uptake in RAW264.7 mouse macrophages. This illustrates the particular properties of the peptide combined with the steroid nucleus, which allows uptake at low concentrations. Using four models of the GCN4 bZIP TF we were able to identify some of the parameters that affect dsDNA recognition in synthetic constructs. Additionally we also for the first time discuss and reveal the interesting cell-uptake properties of this type of peptidosteroid based TF mimics.

Tim Courtin is acknowledged for the determination of the concentration of the final compound by ERETIC-NMR and Jan Goeman for the LC-MS analysis. Yara Ruiz García and Abhishek Iyer are indebted to the Marie Curie Early Stage Research Training Fellowship of the European Community's Seventh Framework Programme under contract number PITN-GA-2010-238679. The FWO-Flanders, BOF (UGent) and the Swedish Research Council are also acknowledged for their financial support. Y. Vladimir Pabon is indebted to the Departamento Administrativo de Ciencia, Tecnología e Innovación COLCIENCIAS, grant 02007/24122010.

Notes and references

1 D. Hanahan and R. A. Weinberg, *Cell*, 2011, **144**, 646–674.
2 F. Fontaine, J. Overman and M. François, *Cell Regener.*, 2015, **4**, 2.
3 T. Suganuma, M. Kawabata, T. Ohshima and M.-A. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 13073–13078.
4 K. Arndt and G. R. Fink, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 8516–8520.
5 T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, 1992, **71**, 1223–1237.
6 R. V Talanian, C. J. McKnight and P. S. Kim, *Science*, 1990, **249**, 769–771.
7 A. M. Caamano, M. E. Vazquez, J. Martinez-Costas, L. Castedo and J. L. Mascareñas, *Angew. Chem., Int. Ed.*, 2000, **39**, 3104–3107.
8 M. Ueno, A. Murakami, K. Makino and T. Morii, *J. Am. Chem. Soc.*, 1993, **115**, 12575–12576.
9 J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vazquez and J. L. Mascareñas, *Nat. Commun.*, 2013, **4**, 1874–1878.
10 L. L. G. Carrette, T. Morii and A. Madder, *Eur. J. Org. Chem.*, 2014, 2883–2891.
11 B. Cuenoud and A. Schepartz, *Science*, 1993, **259**, 510–513.
12 Y. Ruiz García, J. Zelenka, V. Pabon, A. Iyer, M. Buděšínský, T. Kraus, C. I. E. Smith and A. Madder, *Org. Biomol. Chem.*, 2015, **13**, 5273–5278.
13 D. Verzele and A. Madder, *Eur. J. Org. Chem.*, 2013, 673–687.
14 W. Kramer, G. Wess, A. Enhsen, E. Falk, A. Hoffmann, G. Neckermann, G. Schubert and M. Urmann, *J. Controlled Release*, 1997, **46**, 17–30.
15 C. a. Bodé, T. Bechet, E. Prodhomme, K. Gheysen, P. Gregoir, J. C. Martins, C. P. Muller and A. Madder, *Org. Biomol. Chem.*, 2009, **7**, 3391–3399.
16 D. B. Salunke, B. G. Hazra and V. S. Pore, *Curr. Med. Chem.*, 2006, **13**, 813–847.
17 C. A. Bode, C. P. Muller and A. Madder, *J. Pept. Sci.*, 2007, **13**, 702–708.
18 A. Madder, L. Li, M. H. De, N. Farcy, H. D. Van, F. Fant, G. Vanhoenacker, P. Sandra, A. P. Davis and C. P. J. De, *J. Comb. Chem.*, 2002, **4**, 552–562.
19 C. A. Bode, T. Bechet, E. Prodhomme, K. Gheysen, P. Gregoir, J. C. Martins, C. P. Muller and A. Madder, *Org. Biomol. Chem.*, 2009, **7**, 3391–3399.
20 W. S. Horne, M. K. Yadav, C. D. Stout and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2004, **126**, 15366–15367.
21 H. C. Kolb and K. B. Sharpless, *Drug Discovery Today*, 2003, **8**, 1128–1137.
22 C. W. Tornoe, C. Christensen and M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057–3064.
23 I. E. Valverde, A. Bauman, C. A. Kluba, S. Vomstein, M. A. Walter and T. L. Mindt, *Angew. Chem., Int. Ed. Engl.*, 2013, **52**, 8957–8960.
24 R. Echemendía, O. Concepción, F. E. Morales, M. W. Paixão and D. G. Rivera, *Tetrahedron*, 2014, **70**, 3297.
25 A. Hollmann, P. M. Matos, M. T. Augusto, M. A. R. B. Castanho and N. C. Santos, *PLoS One*, 2013, **8**, e60302.
26 P. Ingallinella, E. Bianchi, N. A. Ladwa, Y.-J. Wang, R. Hrin, M. Veneziano, F. Bonelli, T. J. Ketas, J. P. Moore, M. D. Miller and A. Pessi, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5801–5806.
27 J. P. Tam, *J. Immunol. Methods*, 1996, **196**, 17–32.
28 M. Wenska, M. Alvira, P. Steunenberg, A. Stenberg, M. Murtola and R. Stroemberg, *Nucleic Acids Res.*, 2011, **39**, 9047–9059.
29 T. Hai and M. G. Hartman, *Gene*, 2001, **273**, 1–11.
30 S. Futaki, T. Suzuki, W. Ohashi, T. Yagami, S. Tanaka, K. Ueda and Y. Sugiura, *J. Biol. Chem.*, 2001, **276**, 5836–5840.

Organic & Biomolecular Chemistry

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2015, **13**, 5273

Cyclodextrin–peptide conjugates for sequence specific DNA binding†

Yara Ruiz García,[a] Jan Zelenka,[b] Y. Vladimir Pabon,[c] Abhishek Iyer,[a] Miloš Buděšínský,[b] Tomáš Kraus,[b] C. I. Edvard Smith[c] and Annemieke Madder*[a]

Received 27th March 2015,
Accepted 31st March 2015

DOI: 10.1039/c5ob00609k

www.rsc.org/obc

Synthetic models of bZIP transcription factors have been developed with the capability of specific DNA recognition. Our design is based on the CuAAC mediated conjugation of basic region Leucine Zipper peptides to different derivatives of α, β and γ-cyclodextrins equipped with azide functionalities. Thorough optimization of reaction conditions allowed convergent and simultaneous conjugation of two long unprotected cationic peptides to cyclodextrin-bis azide derivatives. The resulting constructs were shown to specifically recognize their cognate DNA sequence with nM affinities. In comparison with previously developed TF models, the derivatives described here combine the enhanced DNA binding capabilities with an easy and convergent synthetic route.

Introduction

The study of gene expression regulation is currently of particular interest due to the upcoming development of gene therapy strategies. Indeed, altered expression of particular genes can cause interferences with biological processes in the cell. Therefore, selective up- or down-regulation of specific gene transcription could ultimately result in a therapeutic platform with biomedical applications. In the last decade, biologists and medicinal chemists have been working together in order to target specific genes for this purpose.[1–3]

In this context, Transcription Factors (TFs) have been studied in great detail for the modification of gene expression and the study of DNA binding affinity in the cellular environment.[4–7] In particular, bZIP proteins have gained attention due to the simple arrangement of the protein structure consisting of a well-defined dimerization domain and a basic binding region.[8] These proteins are able to interact with DNA in a sequence-specific manner by means of inserting their recognition domains, consisting of defined sequences of amino acids, in the DNA major groove. Simplification of such proteins could ultimately result in peptide-based drugs for alternative disease treatment. Moreover, due to the improvements in peptide manufacturing, peptide drugs can now be produced in a straightforward way through synthetic methods and many techniques have been developed for improved peptide stability.

Therefore, the idea of transferring the protein properties to smaller systems whilst conserving the DNA recognition ability has received considerable attention. The bZIP leucine zipper TF binds double-stranded DNA as a dimer, presented as uninterrupted α-helices that grip the major groove and interact with the DNA through basic residues. The main residues involved in DNA recognition are the amino acids 226–248, located in the N-terminal basic region of the GCN4 protein.[8] The optimal position and spatial arrangement of the dimeric peptides are ensured by the leucine zipper domain. Simplified models of these proteins have been designed based on the substitution of the dimerization domain by a scaffold for the appendage of the basic region peptides in a correct geometry.[9] The first proof of this concept was delivered by Kim *et al.*[10] using a disulfide bond as a connector between the extended basic region of the peptides. Since then, several models have been developed trying to improve upon the pioneered one.[11–15]

Herein we report on cyclodextrin based TF models where the dimerization domain of the GCN4 protein is substituted by a scaffold which allows accurate regulation of the geometry and the mutual distance between the two peptide strands. This has been achieved by the use of α, β and γ-cyclodextrin derivatives in order to scan and evaluate the importance of the distance between the two anchoring points for the dimerizing peptides (Fig. 1). Due to the rigid structure of the cyclo-

[a]*Organic and Biomimetic Chemistry Research Group, Department of Organic and Macromolecular Chemistry, Ghent University, Krijgslaan 281 (S4), B-9000 Ghent, Belgium. E-mail: annemieke.madder@ugent.be; Fax: (+32) 9 264 4998*
[b]*Institute of Organic Chemistry and Biochemistry, v.v.i., Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 10 Praha 6, Czech Republic. E-mail: kraus@uochb.cas.cz; Fax: (+420) 220183560; Tel: (+420) 220183372*
[c]*Department of Laboratory Medicine, Clinical Research Center, Karolinska Institutet, 141 86 Huddinge, SE-141 86 Stockholm, Sweden. E-mail: edvard.smith@ki.se; Fax: (+46) 8 585 836 50; Tel: (+46) 8 585 838 00*

†Electronic supplementary information (ESI) available: Reaction conditions, RP-HPLC traces, ESI-MS, NMR spectra and EMSA studies. See DOI: 10.1039/c5ob00609k

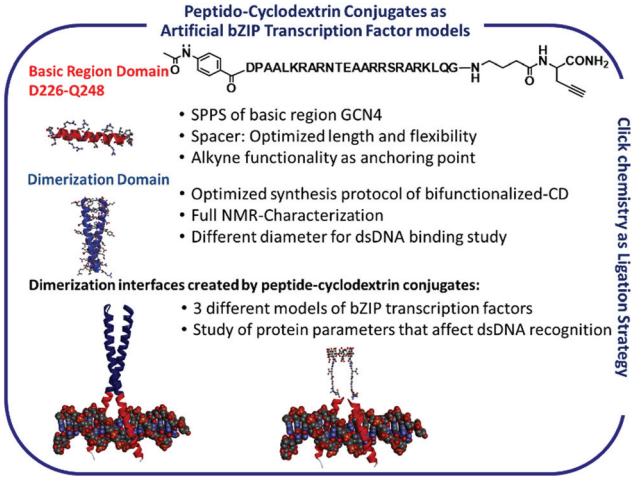


Fig. 1 Design of cyclodextrin based DNA binders.

dextrins, attachment of the basic region peptide of the GCN4 protein onto the opposite positions of the primary rim of the cyclodextrins indeed allows control of the dimerization distance. Based on the diameters of the primary rim of α, β and γ-cyclodextrins, which are 5.7, 7.8 and 9.5 Å respectively,[16] three different peptide–cyclodextrin conjugates have been synthesized, keeping in mind that the width of the major groove is 11.2 Å. For this purpose, we have synthesized $6^{I},6^{IV}$-diazido-α-cyclodextrin, $6^{I},6^{IV}$-diazido-β-cyclodextrin and $6^{I},6^{V}$-diazido-γ-cyclodextrin (Fig. 2), in order to conjugate peptides which are functionalized with an alkyne at the C-terminus *via* CuAAC.

The CuAAC is a broadly developed strategy for the bioconjugation of peptides to different building blocks due to its mildness, compatibility with a large variety of functional groups and the possibility of performing the reaction in a wide range of solvents and buffers. In view of the specific geometry, steric and electronic properties, a 1,2,3-triazole can be regarded as a *trans*-amide bond mimic.[17] Moreover, this linkage is stable under physiological conditions, thus representing a perfect heterocyclic moiety to replace unstable linkers.[18–20] In addition, successful replacement of dipeptide sequences in α-helical peptides by triazole units has been shown to only insignificantly influence the secondary peptide structure.[17]

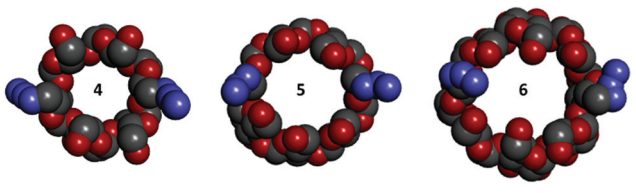


Fig. 2 Molecular visualization of the cyclodextrin derivatives synthesized as dimerization domains. From left to right: $6^{I},6^{IV}$-diazido-α-cyclodextrin **4**, $6^{I},6^{IV}$-diazido-β-cyclodextrin **5** and $6^{I},6^{V}$-diazido-γ-cyclodextrin **6**. Exact structures are described in the ESI.†

Peptide–cyclodextrin conjugates have been previously developed for a broad range of applications, such as drug release systems,[21,22] enantioselective ester hydrolysis catalysis,[23] and for the creation of new types of chemosensors,[24] enzyme mimics[25] and self-assembling materials.[26,27] Although CD conjugation to peptides has been intensively studied, we here for the first time used the cyclodextrin moiety as an artificial dimerizing unit for mimicking protein–DNA interactions.

Results and discussion

The required diazido cyclodextrin derivatives **4–6** were prepared in one step (Fig. 3) by reaction of the known $6^{I},6^{IV}$-dibromo-α-cyclodextrin[28a], $6^{I},6^{IV}$-dibromo-β-cyclodextrin[28b] and $6^{I},6^{V}$-dibromo-γ-cyclodextrin[29] with three-fold excess of sodium azide in DMF at 50 °C. Pure compounds **4–6** were isolated from the reaction mixtures by reversed-phase column chromatography with yields of 90%, 87% and 77% for the α, β and γ-cyclodextrin derivatives, respectively.

As we have previously illustrated, peptides attached directly to a dimerization unit are unable to adopt an adequate position for major groove binding.[30] Thus, a spacer is needed between the dimerization scaffold and the basic region GCN4 peptide to afford some flexibility required for binding. In addition to that, a C-terminal propargylglycine residue was introduced for conjugation of the deprotected peptide to the azide containing cyclodextrin scaffolds. The GCN4 basic region consists of the following 23 amino acids: DPAALKRARNTEAARRSRARKLQG which specifically recognize the ATF/CREB-binding site (5′-ATGA C/G TCAT-3′). The monomeric GCN4 sequence was synthesized using standard Fmoc SPPS on a Rink-amide ChemMatrix resin. After cleavage and deprotec-

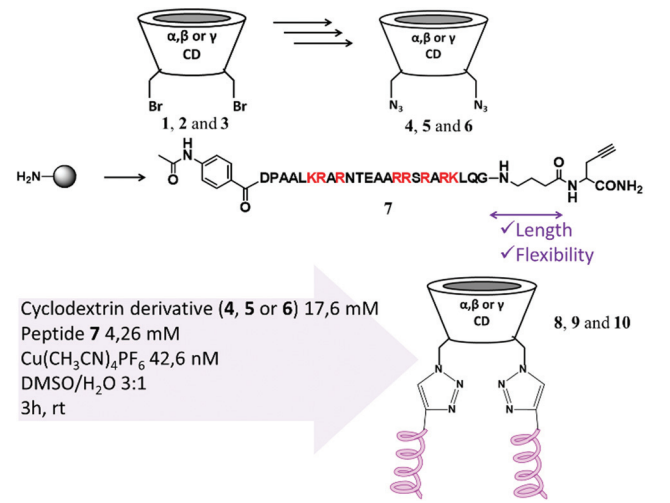


Fig. 3 Synthesis scheme for the conjugation of the different cyclodextrins **4**, **5** and **6** and peptide **7** to obtain the final compounds **8**, **9** and **10**. Residues that can chelate copper are marked in red.

tion, the peptide could be used for the next steps without intermediate purification.

The basic region GCN4 peptide functionalized with an alkyne at the C-terminus was then anchored to the opposite positions at the primary rim of α, β and γ-cyclodextrin derivatives (Fig. 3). In order to prevent aggregation of the peptide during the reaction, high polarity media were needed. In addition, polar solvents prevent copper species from aggregation. Therefore, we chose DMSO–H_2O as a solvent mixture. It is known that sodium ascorbate can cause covalent modifications of lysines and arginines, and can form reactive oxygen species.[31,32] Therefore, $Cu(CH_3CN)_4PF_6$ was chosen as the catalyst after carefully studying different reaction conditions. It is to be noted that a high excess of catalyst is needed which can probably be ascribed to the chelation of copper by the nitrogen containing amino acid side chains, which are intensely represented within this particular DNA binding peptide sequence. The reaction was stopped after 3 hours at room temperature under argon. The final purification of the reaction mixture was achieved *via* RP-HPLC and all conjugates were obtained in high purity (Fig. 4).

After successful synthesis of the constructs, the DNA binding capabilities of the systems were investigated. Quantitative binding analysis was performed by the Electrophoretic Mobility Shift Assay (EMSA) using double stranded ^{32}P-labeled CRE DNA (CRE: 5′-ATGACGTCAT-3′), which is the natural palindromic binding site of the GCN4 protein. This technique allows the sensitive determination of the dissociation constant (K_d) (Fig. 5).

The results show all conjugates to bind to the target DNA sequence as evident from the appearance of an up-shifted band. In contrast, no binding was observed when incubating the conjugates with random DNA sequences (see ESI†). Previous models developed by Morii and Mascareñas possess the same length of the basic region peptides (D226-Q248) and are therefore comparable to our systems. The group of Morii developed a system in which the dimerization unit consisted of a complex of cyclodextrin and adamantane. Both counterparts were anchored to a GCN4 basic region peptide and bind only when dimerized. EMSA studies were performed with <1 nM radiolabelled CRE, obtaining a K_d < 150 nM.[11] More recent models designed by the group of Mascareñas employed a diazobenzene moiety for dimerization, with a K_d < 5 nM using <1 nM radiolabelled CRE. They also reported the affinity of a disulfide-bridged GCN4 basic region, with a K_d of <150 nM approximately for a DNA concentration of 50 nM.[33] The latest model represents a dual mimic in which the dimerization arrangement and DNA selectivity are controlled by adjusting selected external parameters.[34] The basic region peptides were functionalized with a cysteine and a terpyridine moiety attached to the residues at the N- and C-termini of the peptide chain to achieve selective dimerization at both sites. Depending on the conditions, the disulfide-based or the metal–terpyridine complex-based dimers were preferred, and different DNA sequences could be targeted. By dimerizing the peptide in the presence of Ni^{2+}, the constructs were bound to the CRE sequence with a K_d of 299 ± 26 nM, using a DNA concentration of 100 nM, ~100 pM labeled with ^{32}P. For these last 3 models, it should be noted that the nonionic detergent (NP-40) and BSA were used during EMSA, known to decrease aggregation of peptides and proteins and therefore favor the interaction with DNA.[35,36] In our case, the obtained K_d values at 4 °C for compounds **8**, **9** and **10** were 50 ± 20, 30 ± 20 and 100 ± 60 nM respectively. A rough comparison thus indicates that our new constructs, apart from being synthesized in a straightforward and convergent manner, display comparable binding pro-

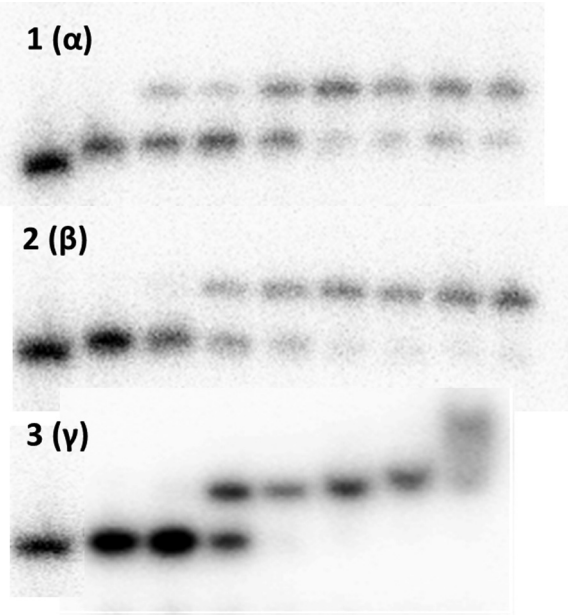


Fig. 5 EMSA titration at 4 °C of the dipodal peptidocyclodextrin conjugates to the 5′-labeled ^{32}P-CRE sequence (CRE: 5′-CGGATGACGTCATTTTTTTTC-3′ underlined portion indicates the binding region): concentrations: 5 nM dsDNA; first lane in all the gels: pyrimidine strand. Concentration of peptides in the lanes 2–9 for gel 1 and 2 (8 and 9): 0, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.1125 and 0.125 μM. Concentration of peptides in the lanes 2–8 for gel 3 (10): 0, 0.06, 0.125, 0.25, 0.5, 1, 2 μM. A Fuji FLA3000 phosphorimager was used for gel analysis and Multi Gauge V 3.0 software (Fujifilm) was used for quantification of the electrophoretic band intensities.

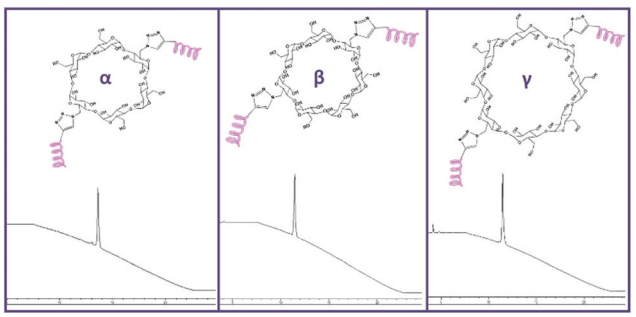


Fig. 4 Structure of the final compounds **8** (α), **9** (β) and **10** (γ) and RP-HPLC chromatograms of purified compounds (0–100% ACN for 15 min on C4, 300 Å).

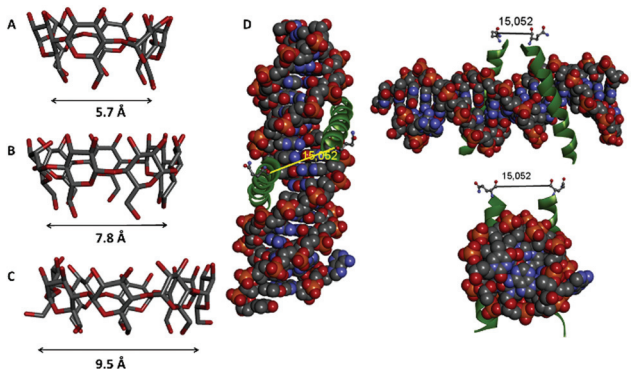


Fig. 6 3D structure of α, β and γ-cyclodextrins and the primary rim diameter distance (A, B and C respectively) and crystal structure of the GCN4 basic region peptides appended to the major groove of the DNA from different perspectives. The distance between C-termini is indicated in angstroms (PDB: 1YSA) (D).

perties to those described in the literature. The obtained values are further within the same order of magnitude as those calculated for the binding of bacterially expressed GCN4 and synthetic versions thereof. For instance, a dimer comprising of a 56mer GCN4 basic region (residues 226–281) also binds CRE in the nanomolar range ($K_d \sim 12$ nM).[37–39] No data for the dissociation constant of the natural protein have been reported so far.

In order to better explain the results obtained from the binding pattern on the gels and the determination of the K_d values, we performed molecular visualization of the dimerization interface of the bZIP GCN4 TF obtained by discarding the leucine zipper domain from the crystal structure of the natural protein (Fig. 6D). The distance between the C-termini of the basic regions is shown in Fig. 6 at different perspectives of the protein–DNA interface, and is found to be 15.052 Å. The K_d values showed that derivative **9** was the one with the best binding capability to the CRE sequence. Therefore, we consider that beta CD is the dimerizer which allows optimal anchoring of the peptides on the major groove of DNA. In the protein, the C-termini are placed at a distance of approximately 15 Å. In the case of our constructs, although the diameter of the primary rim of the cyclodextrins is known (Fig. 6, left), due to the flexibility of the linker, the exact distance between the C-termini of the peptides on **8**, **9** and **10** cannot be predicted accurately. However, we observed that increasing or decreasing the distance between the attachment points of the peptides by the use of different CD scaffolds results in a deviation of the optimum geometry of the system. This is reflected in terms of decreased binding affinity and lower K_ds. It should also be noted that these results are specific for the basic region GCN4 peptide sequence and the given spacers.

Conclusions

In conclusion, three peptide–cyclodextrin conjugates for sequence-selective DNA recognition were obtained. This was achieved by the use of α, β and γ-cyclodextrin diazido derivatives as scaffolds for the appendage of the peptides by CuAAC. We have successfully synthesized and fully characterized $6^I,6^{IV}$-diazido-α-cyclodextrin (**4**), $6^I,6^{IV}$-diazido-β-cyclodextrin (**5**) and $6^I,6^V$-diazido-γ-cyclodextrin (**6**). Though examples exist of CuAAC reactions with CD derivatives,[26,40] to the best of our knowledge long, deprotected peptides of this size have so far not been conjugated to cyclodextrins *via* CuAAC. We here presented the optimized conditions for the anchoring of such long peptides to cyclodextrin units *via* click chemistry.

Our results indicate the usefulness of an optimized dimerization configuration between both peptides in artificial TF models. Indeed, it was shown that the distance between the anchoring points has a notable influence on DNA binding. Successful models can be obtained by trying to approach the exact features of the protein at the interface between the dimerization and the basic region domains of the bZIP TF to achieve DNA binding comparable to that of the natural TF.

Experimental

Materials

All organic solvents were purchased from commercial suppliers and used without further purification or drying. DMF extra dry (with molecular sieves, water < 50 ppm) was acquired from ACROS Organics. DMF and NMP (peptide synthesis grade) were purchased from Biosolve. Ethyl acetate, acetonitrile, methanol, diethyl ether, DIPEA, supplied as extra dry, redistilled, 99.5% pure and triisopropylsilane were purchased from Sigma Aldrich. Water with the Milli-Q grade standard was obtained in-house from either a Millipore ROs 5 purification system or a Sartorius Arium 611 DI. Rink-Amide ChemMatrix (100–200 μm, manufacturer's loading: 0.52–0.54 mmol g^{-1}) from Biotage. All reagents were acquired from commercial sources and used without prior purification. Fmoc-Propargylglycine-OH (Fmoc-Pra-OH), tris-borate-EDTA buffer 10× pH 8.3 (TBE buffer), ammonium persulfate (APS), tetramethylethylenediamine (TEMED), PyBOP and HBTU coupling reagents were obtained from either Merck Novabiochem or IRIS Biotech GmbH, while HATU (purum ≥98.0%) was acquired from Fluka. TFA was obtained from IRIS Biotech GmbH. The Nα-Fmoc protected amino acids were purchased at Merck Novabiochem, IRIS Biotech GmbH and Fluka, or supplied by MultiSynTech GmbH. All chiral α-amino acids possessed the l configuration. Throughout this work, residues with standard acid-sensitive side-chain PGs were used: Asp(O*t*Bu) [D], Cys(Trt) [C], Glu(O*t*Bu) [E], His(Trt) [H], Lys(Boc) [K], Asn(Trt) [N], Gln(Trt) [Q], Arg(Pbf) [R], Ser(*t*Bu) [S], Thr(*t*Bu) [T]. All oligonucleotides used were commercially purchased from Eurogentec (HPLC purified using RP-cartridge-Gold, 200 nm scale) and were used as such.

Peptide syntheses

All automated peptide syntheses were synthesized on Rink Amide Chemmatrix resin with a loading of 0.54 mmol g^{-1}

using 10 eq. of amino acid, 10 eq. HBTU and 20 eq. DIPEA using Fmoc/*t*Bu SPPS. The coupling time was 1 h. Attachment of the first two residues glycine and GABA was performed manually. This was done to increase the space between the scaffold and the peptide. Residues D226-Q248 were then coupled *via* automated peptide synthesis. 20% piperidine/NMP was used for Fmoc deprotection. As the last residue and to increase UV-absorption and facilitate HPLC analysis, 4-acetamidobenzoic acid was coupled to the N-terminus. Cleavage and deprotection of the peptide was performed with a cocktail mixture containing TFA–TIS–H_2O (95/2.5/2.5) for 4 hours at room temperature. The peptide was obtained by precipitation with cold ether and lyophilization. No further purification was performed at this stage.

Peptide–CD conjugation *via* CuAAC

The selected deriviatized cyclodextrin scaffold **4**, **5** or **6** (4 eq.) was dissolved in dry DMSO in a round bottom flask. Peptide **7** (1 eq.) was dissolved in MilliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (10 eq.) was dissolved in dry DMSO and added to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC to obtain compound **8**. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.

Electrophoretic mobility shift assay (EMSA) for quantification of the dissociation constant

Preparation of ^{32}P-labeled double-stranded DNA target. Oligonucleotide CRE (5′-CGG ATG ACG TCA TTT TTT TTC-3′) was 5′-labeled using [γ-^{32}P] ATP and T4 polynucleotide kinase (Fermentas) according to the manufacturer's protocol and then purified using a QIAquick Nucleotide Removal Kit (Qiagen). The 5′-end labeled pyrimidine oligonucleotide was annealed with the unlabeled complementary strand. An amount of 5 nM dsDNA was prepared by diluting 20 μL of 0.5 M Tris, pH = 8, 40 μL of 2.5 M NaCl, 40 μL of 0.025 M EDTA and then adding MilliQ water such that the total volume is 1 mL. The DNA was annealed in a heat block by heating from 95 °C for 5 minutes and then slowly cooling to room temperature. Loading buffer: 20 μL Tris 1 M, pH = 7.6, 20 μL KCl 0.2 M, 20 μL $MgCl_2$ 0.1 M, 40 μL EDTA 0.025 M. Sucrose: 30% sucrose in mQ (300 mg mL^{-1}). Peptides: 10 μL stock solutions (10×) were prepared in MilliQ water. The first set of experiments were performed with 10× solutions of **8**, **9** and **10** at concentrations of 0, 0.6, 1.25, 2.5, 5, 10 and 20 μM. Loading mixture: The loading mixture was comprised of 10 μL mQ, 4 μL sucrose, 2 μL loading buffer, 2 μL DNA and 2 μL peptide. The loading mixture was prepared only 1 hour prior to running of gels and kept on ice as soon as it was ready.

Preparation of gels (for 1 gel). In a clean glass beaker the following were added (in the given order): 21.57 mL mQ, 0.6 mL 10× TBE, 7.5 mL of 40% acrylamide solution (29 : 1), and 0.3 mL APS (10% w/w in mQ), and 30 μL of TEMED was then added to the mixture and mixed properly before pouring it gently along parallel glass plates. The glass plates were tapped gently to ensure the removal of all air bubbles and the markers were squeezed between the plates to ensure uniform width of each well. Sufficient time was given for polymerization (40 minutes).

Electrophoresis. A pre-run of the gels was performed prior to loading them. Care was taken to see that the gels were properly immersed in 0.2× TBE buffer (non-denaturing gel, without urea) and the loading wells were free from any air bubbles. The wells were washed after the pre-run. Instrument settings: 150 V, 100 mA, 19 W for 30 minutes with a circulation water-cooling system. 5 μL of the loading mixture was then loaded onto the wells. Instrument settings: 150 V, 100 mA, 19 W for 45 minutes with a circulation water-cooling system.

Analysis of gels. The gels were frozen and analyzed by phosphor imaging using a Molecular Imager FX and the data were processed using Quantity One software (BioRad).

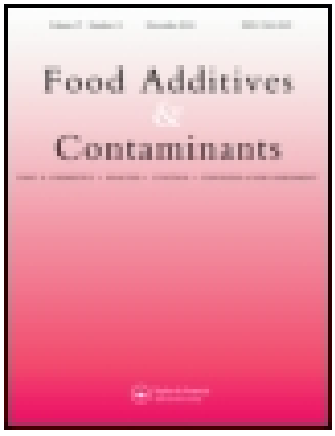
Acknowledgements

Tim Courtin is acknowledged for the determination of the concentration of the final compounds by ERECTIC-NMR. Jan Goeman is acknowledged for the LC-MS analysis. Yara Ruiz García and Abhishek Iyer are indebted to the Marie Curie Early Stage Research Training Fellowship of the European Community's Seventh Framework Programme under contract number PITN-GA-2010-238679. Support from the Ministry of Education, Youth and Sport of the Czech Republic to T. Kraus (LD12019) is greatly appreciated. Y. Vladimir Pabon is indebted to the Departamento Admistrativo de Ciencia, Tecnología e Innovación COLCIENCIAS, grant 02007/24122010. Ellen Gyssels and Nathalie de Laet are acknowledged for the determination of the DNA concentration using the Trinean Dropsense multichannel spectrophotometer.

Notes and references

1 S. Goverdhana, M. Puntel, W. Xiong, J. M. Zirger, C. Barcia, J. F. Curtin, E. B. Soffer, S. Mondkar, G. D. King, J. Hu, S. A. Sciascia, M. Candolfi, D. S. Greengold, P. R. Lowenstein and M. G. Castro, *Mol. Ther.*, 2005, **12**, 189–211.
2 A. Klug, *Annu. Rev. Biochem.*, 2010, **79**, 213–231.
3 X. Li, S. Chen, T. Sun, Y. Xu, Y. Chen, Y. Liu, R. Xiang and N. Li, *Clin. Lab.*, 2014, **60**, 909–918.
4 P. C. Ma, M. A. Rould, H. Weintraub and C. O. Pabo, *Cell*, 1994, **77**, 451–459.
5 J. N. Glover and S. C. Harrison, *Nature*, 1995, **373**, 257–261.
6 T. Shimizu, A. Toumoto, K. Ihara, M. Shimizu, Y. Kyogoku, N. Ogawa, Y. Oshima and T. Hakoshima, *EMBO J.*, 1997, **16**, 4689–4697.
7 J. E. Darnell Jr., *Nat. Rev. Cancer*, 2002, **2**, 740–749.
8 T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, 1992, **71**, 1223–1237.

9 E. Pazos, J. Mosquera, M. E. Vázquez and J. L. Mascareñas, *ChemBioChem*, 2011, **12**, 1958–1973.
10 R. V. Talanian, C. J. McKnight and P. S. Kim, *Science*, 1990, **249**, 769–771.
11 M. Ueno, A. Murakami, K. Makino and T. Morii, *J. Am. Chem. Soc.*, 1993, **115**, 12575–12576.
12 B. Cuenoud and A. Schepartz, *Science*, 1993, **259**, 510–513.
13 A. M. Caamano, M. E. Vazquez, J. Martinez-Costas, L. Castedo and J. L. Mascarenas, *Angew. Chem., Int. Ed.*, 2000, **39**, 3104–3107.
14 L. L. G. Carrette, T. Morii and A. Madder, *Eur. J. Org. Chem.*, 2014, 2883–2891.
15 E. Oheix and A. F. a. Peacock, *Chem. – Eur. J.*, 2014, **20**, 2829–2839.
16 M. E. Brewster and T. Loftsson, *Adv. Drug Delivery Rev.*, 2007, **59**, 645–666.
17 W. S. Horne, M. K. Yadav, C. D. Stout and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2004, **126**, 15366–15367.
18 H. C. Kolb and K. B. Sharpless, *Drug Discovery Today*, 2003, **8**, 1128–1137.
19 C. W. Tornoe, C. Christensen and M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057–3064.
20 I. E. Valverde, A. Bauman, C. A. Kluba, S. Vomstein, M. A. Walter and T. L. Mindt, *Angew. Chem., Int. Ed.*, 2013, **52**, 8957–8960.
21 N. Schaschke, I. Assfalg-Machleidt, W. Machleidt, T. Laßleben, C. P. Sommerhoff and L. Moroder, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 677–680.
22 C. Decroocq, A. Joosten, R. Sergent, T. Mena Barragán, C. Ortiz Mellet and P. Compain, *ChemBioChem*, 2013, **14**, 2038–2049.
23 H. Tsutsumi, H. Ikeda, H. Mihara and A. Ueno, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 723–726.
24 M. A. Hossain, K. Hamasaki, K. Takahashi, H. Mihara and A. Ueno, *J. Am. Chem. Soc.*, 2001, **123**, 7435–7436.
25 H. Tsutsumi, K. Hamasaki, H. Mihara and A. Ueno, *J. Chem. Soc., Perkin Trans. 2*, 2000, 1813–1818.
26 Y.-C. Lin, P.-I. Wang and S.-W. Kuo, *Soft Matter*, 2012, **8**, 9676.
27 J. N. Beck, A. Singh, A. R. Rothenberg, J. H. Elisseeff and A. J. Ewald, *Biomaterials*, 2013, **34**, 9486–9495.
28 (*a*) L. Kumprecht, M. Budesínský, J. Vondrásek, J. Vymetal, J. Cerný, I. Císarová, J. Brynda, V. Herzig, P. Koutník, J. Závada and T. J. Kraus, *Org. Chem.*, 2009, **74**, 1082; (*b*) A. Grishina, S. Stanchev, L. Kumprecht, M. Buděšínský, M. Pojarová, M. Dušek, M. Rumlová, I. Křížová, L. Rulíšek and T. Kraus, *Chem. – Eur. J.*, 2012, **18**, 12292–12304.
29 S. Volkov, L. Kumprecht, M. Buděšínský, M. Dušek and T. Kraus, *Org. Biomol. Chem.*, 2015, **13**, 2980–2985.
30 D. Verzele and A. Madder, *Eur. J. Org. Chem.*, 2013, 673–687.
31 V. Hong, S. I. Presolski, C. Ma and M. G. Finn, *Angew. Chem., Int. Ed. Engl.*, 2009, **48**(52), 9879–9883.
32 V. Hong, N. F. Steinmetz, M. Manchester and M. G. Finn, *Bioconjugate Chem.*, 2010, **21**, 1912–1916.
33 A. Jiménez-Balsa, E. Pazos, B. Martínez-Albardonedo, J. L. Mascareñas and M. E. Vázquez, *Angew. Chem., Int. Ed.*, 2012, **51**, 8825–8829.
34 J. Mosquera, A. Jiménez-Balsa, V. Dodero, M. E. Vázquez and J. L. Mascareñas, *Nat. Commun.*, 2013, **4**, 1874–1878.
35 L. M. Hellman and M. G. Fried, *Nat. Protoc.*, 2007, **2**, 1849–1861.
36 P. L. Molloy, *Cell Biology (Page 330)*, Elsevier, Oxford (UK), 3rd edn, Vol. 4, 2006.
37 J. J. Hollenbeck and M. G. Oakley, *Biochemistry*, 2000, **39**, 6380–6389.
38 G. H. Bird, A. R. Lajmi and J. A. Shin, *Biopolymers*, 2002, **65**, 10–20.
39 I.-S. Chan, A. V. Fedorova and J. a. Shin, *Biochemistry*, 2007, **46**, 1663–1671.
40 P.-A. Faugeras, B. Boëns, P.-H. Elchinger, F. Brouillette, D. Montplaisir, R. Zerrouki and R. Lucas, *Eur. J. Org. Chem.*, 2012, 4087–4105.



Food Additives & Contaminants: Part A
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tfac20

An immunogen synthesis strategy for the development of specific anti-deoxynivalenol monoclonal antibodies

Melanie Sanders[a], Yirong Guo[ab], Abhishek Iyer[c], Yara Ruiz García[c], Anastasia Galvita[a], Arne Heyerick[d], Dieter Deforce[d], Martijn D.P. Risseeuw[e], Serge Van Calenbergh[e], Marc Bracke[f], Sergei Eremin[g], Annemieke Madder[c] & Sarah De Saeger[a]

[a] Laboratory of Food Analysis, Ghent University, Ghent, Belgium

[b] Institute of Pesticide and Environmental Toxicology, Zhejiang University, Hangzhou, China

[c] Organic and Biomimetic Chemistry Research Group, Ghent University, Ghent, Belgium

[d] Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium

[e] Laboratory for Medicinal Chemistry, Ghent University, Ghent, Belgium

[f] Laboratory of Experimental Cancer Research, Ghent University, Ghent, Belgium

[g] Department of Chemical Enzymology, Moscow State University, Moscow, Russia

Published online: 18 Sep 2014.

Food Additives & Contaminants: Part A, 2014
http://dx.doi.org/10.1080/19440049.2014.955887

An immunogen synthesis strategy for the development of specific anti-deoxynivalenol monoclonal antibodies

Melanie Sanders[a]*, Yirong Guo[a,b], Abhishek Iyer[c], Yara Ruiz García[c], Anastasia Galvita[a], Arne Heyerick[d], Dieter Deforce[d], Martijn D.P. Risseeuw[e], Serge Van Calenbergh[e], Marc Bracke[f], Sergei Eremin[g], Annemieke Madder[c] and Sarah De Saeger[a]

[a]*Laboratory of Food Analysis, Ghent University, Ghent, Belgium;* [b]*Institute of Pesticide and Environmental Toxicology, Zhejiang University, Hangzhou, China;* [c]*Organic and Biomimetic Chemistry Research Group, Ghent University, Ghent, Belgium;* [d]*Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium;* [e]*Laboratory for Medicinal Chemistry, Ghent University, Ghent, Belgium;* [f]*Laboratory of Experimental Cancer Research, Ghent University, Ghent, Belgium;* [g]*Department of Chemical Enzymology, Moscow State University, Moscow, Russia*

(*Received 5 June 2014; accepted 13 August 2014*)

An immunogen synthesis strategy was designed to develop anti-deoxynivalenol (DON) monoclonal antibodies with low cross-reactivity against structurally similar trichothecenes. A total of eight different DON immunogens were synthesised, differing in the type and position of the linker on the DON molecule. After immunisation, antisera from mice immunised with different DON immunogens were checked for the presence of relevant antibodies. Then, both homologous and heterologous enzyme-linked immunosorbent assays (ELISAs) were performed for hybridoma screening. Finally, three monoclonal antibodies against DON and its analogues were generated. In addition, monoclonal antibody 13H1 could recognise DON and its analogues in the order of HT-2 toxin > 15-acetyldeoxynivalenol (15-ADON) > DON, with IC_{50} ranging from 1.14 to 2.13 μg ml^{-1}. Another monoclonal antibody 10H10 manifested relatively close sensitivities to DON, 3-acetyldeoxynivalenol (3-ADON) and 15-ADON, with IC_{50} values of 22, 15 and 34 ng ml^{-1}, respectively. Using an indirect ELISA format decreases the 10H10 sensitivity to 15-ADON with 92%. A third monoclonal antibody 2A9 showed to be very specific and sensitive to 3-ADON, with IC_{50} of 0.38 ng ml^{-1}. Using both 2A9 and 10H10 monoclonal antibodies allows determining sole DON contamination.

Keywords: deoxynivalenol; monoclonal antibodies; immunogens; ELISA; cross-reactivity

Introduction

Trichothecene mycotoxins are a group of naturally occurring secondary metabolites produced by *Fusarium* species, in particular *F. graminearum* and *F. culmorum*. Among the 150 related trichothecenes, deoxynivalenol (DON) (Figure 1) is of special importance as it is formed in the field prior to harvest and because its occurrence cannot be completely avoided by plant production-minimising strategies due to the high impact of weather conditions. Especially wheat, triticale and maize are vulnerable for *Fusarium* infection and subsequent DON production (Maragos et al. 2006; Döll & Dänicke 2011).

DON can cause disease in several animal species, especially in swine, and causes symptoms including reduced feed consumption, abdominal distress, increased salivation, malaise, diarrhoea, anorexia, leucopenia, haemorrhage, shock and death in extremely high DON doses (Maragos & McCormick 2000; Pestka 2007). Besides these effects, DON is also generally considered to be a potent inhibitor of protein and DNA synthesis and is known to be immunosuppressive (Krska et al. 2001; Danicke et al. 2006). Highly dividing cells, such as intestinal epithelial cells known to ensure a proper barrier function, are especially sensitive to trichothecenes. Exposure to DON can lead to a decrease in absorption of nutrients and a decrease in cell proliferation and consequently cell differentiation (Pinton et al. 2012).

Next to DON two acetylated derivatives are known to be coproduced, namely 3-acetyldeoxynivalenol (3-ADON) and 15-acetyldeoxynivalenol (15-ADON). JECFA stated that the acetylated DON levels were less than 10% of the total DON found in cereal grains, whereas De Boevre et al. (2012) found higher 3-ADON and 15-ADON co-contamination for several food and feed samples. In 2010 JECFA considered the toxicity of the acetylated derivatives equal to DON, but a recent study suggested that higher toxicity of 15-ADON should be taken into account. Another DON masked form detected in cereals and beers, namely DON-3-glucoside (DON-3-G), also shows a non-negligible contribution to the overall DON contamination (Berthiller et al. 2009; JECFA 2011; De Boevre et al. 2012; Pinton et al. 2012).

*Corresponding author. Email: melanie.sanders@ugent.be

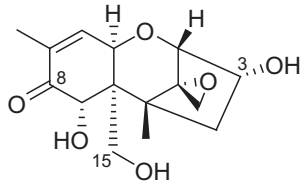


Figure 1. Chemical structure of deoxynivalenol.

Because of the economic importance of this toxin and from a food safety perspective, a variety of analytical techniques have been developed for the detection of DON and related trichothecene mycotoxins in food and feed. Most commonly used techniques include TLC, GC, HPLC and immunological approaches such as radioimmunoassay and ELISA. Recently, ELISA and ELISA-based procedures have gained acceptance as they offer the advantage of specificity, sensitivity, simplicity and rapidity, which are of importance for routine testing of mycotoxins (Maragos & McCormick 2000; Maragos et al. 2006, 2012). The specificity and sensitivity are dependent on the antibody used in the ELISA assay. Several antibodies for DON have been reported and all show high cross-reactivity against the acetylated derivatives 3-ADON and 15-ADON, which reduces their specific character. Sensitivity values of antibodies, measured by the IC_{50} value, range from approximately 20 to 3 ng ml^{-1} (Casale et al. 1988; Mills et al. 1990; Usleber et al. 1991; Sinha et al. 1995; Maragos & McCormick 2000).

The objective of this study is to improve the specificity of the ELISA assay by developing highly specific monoclonal anti-DON antibodies. The high affinity of the monoclonal antibody to 3-ADON or 15-ADON likely derives from the chemistry used to prepare the immunogen. The ester linkage of 3-ADON or 15-ADON may resemble the linkage of the DON-carrier protein immunogen. Therefore, it was decided to use different procedures to make linkers with varying length and chemical structure between DON and the carrier protein (Figure 2).

Materials and methods

Reagents

DON, 3-ADON and 15-ADON standards were obtained from Fermentek (Jerusalem, Israel). HT-2 toxin was purchased from Sigma-Aldrich (Bornem, Belgium).

Colorburst™ blue 3,3′,5,5′-tetramethylbenzidine (TMB) substrate solution containing hydrogen peroxide was supplied by Alerchek (Springvale, ME, USA). Rabbit anti-mouse immunoglobulins (anti-mouse IgG secondary antibody; protein concentration of 2.5 g l^{-1}) were purchased from DakoCytomation (Glostrup, Denmark). N, N′-carbonyldiimidazole, cyanuric chloride (CC), N,N′-diisopropylethylamine, glutaric anhydride, 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC), 1-butane boronic acid, sodium tetraborate, carboxymethoxylamine hemihydrochloride (CMO), γ-aminobutyric acid, sulfo-N-hydroxysuccinimide (NHS), tetrahydrofuran (THF), acetic acid, potassium carbonate, copper (II) sulfate, sodium ascorbate, 4,7-diphenyl-1,10-bathophenanthroline disulfonic acid disodium salt, arginine, glycine, aspartic acid, D-phenylalanine, lysine, trifluoroacetic acid (TFA), triisopropylsilane (TIS), methyl tertiary butyl ether (MTBE), polyethylene glycol (bovine serum albumin (BSA), ovalbumin (OVA), horseradish peroxidase (HRP), rabbit anti-mouse IgG secondary antibody labelled with horseradish peroxidase (Sec Ab-HRP), PBS (0.01 M, pH 7.4) tablet, carbonate-buffered saline (CBS, 0.05 M, pH 9.6) capsule, complete and incomplete Freund's adjuvants, Tween 20 and skim milk powder were obtained from Sigma-Aldrich (Bornem, Belgium). O-propargyl-hydroxylamine hydrochloride was purchased from Focus Synthesis (San Diego, CA, USA). Imidazole-1-sulfonylazide hydrochloride was kindly provided by the Laboratory of Medicinal Chemistry (Ghent, Belgium). The Oasis HLB cartridges were obtained from Waters (Zellik, Belgium). Chlorotrityl chloride resin was purchased from Merck (Darmstadt, Germany). Deionised water was purified by a Millipore Milli-Q system (Brussels, Belgium). Other chemicals and solvents were of analytical grade.

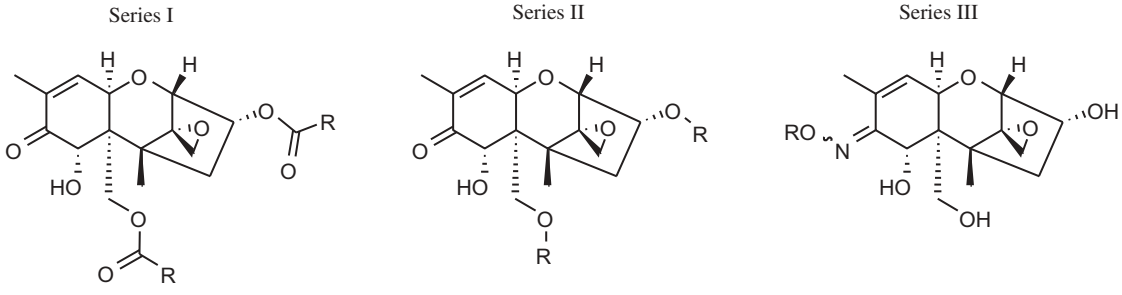


Figure 2. Three series of synthesised DON immunogens.

Nunc-Immuno™ F96-well microplates and Nunclon™ cell culture plates were from Nalge Nunc International (Roskilde, Denmark). Protein concentrators (9K MWCO, 20 ml) were purchased from Thermo Scientific (Rockford, USA).

Preparation of DON immunogens

DON is a hapten due to its small size and therefore cannot elicit an immune response. Consequently, for the synthesis of immunogens, DON was coupled to a protein via a linker to increase the molecular weight. In search for specific anti-DON antibodies, different synthetic strategies were followed. Firstly, DON was coupled to a carrier protein via the C3 and/or C15 position using a linker with a carboxyl function (Series I, Figure 2). To minimise the chance of cross-reacting monoclonal antibodies, a second DON series of immunogens was generated by coupling the carrier protein via a linker on C3 and/or C15 without a carboxyl function (Series II, Figure 2). To reduce even further the possibility of cross-reacting monoclonal antibodies the linkage to the carrier protein was introduced through reaction with the C8 carbonyl function of DON, delivering a last series of immunogens (Series III, Figure 2). The used carrier proteins were bovine serum albumin (BSA) for the immunogen and ovalbumin (OVA) or horseradish peroxidase (HRP) for the coating/competitive antigen in indirect or direct enzyme-linked immunosorbent assay (ELISA), respectively. The synthetic pathways underneath are described for BSA. The basic chemical structure of the three different series of immunogens is illustrated in Figure 2.

The first immunogen of the series I is DON-3,15-hemiglutaryl (HG)-BSA (Figure 3A) and was formed by the reaction of DON with glutaric anhydride in dry pyridine. Concisely, to a solution of 5 mg of DON in 500 µl of dry pyridine, 100 mg of glutaric anhydride was added and the solution was heated at 100°C for 8 h. After evaporation, the residue was dissolved in 7.5 ml of chloroform and washed three times with 5 ml of 0.1 M HCl. A purification step was performed using Oasis HLB® cartridges by a method derived from De Smet et al. (2010). Briefly, 5 µg of the crude DON-HG conjugate was dissolved in 1 ml of methanol–water (10:90, v/v) pH 2.3 and loaded onto a preconditioned column. After loading, the sorbent was dried for 15 s by applying vacuum. DON-HG was eluted by washing the column with 1 ml of methanol–water (55:45, v/v) followed by 1 ml of methanol. The sorbent was dried for 15 s by applying a vacuum. After purity determination by time-of-flight (TOF) MS, an aliquot of 1 mg of the synthesised DON-HG (1.92 µmol) was dissolved in 0.5 ml PBS using ultrasonication and transferred to 30 mg BSA (0.45 µmol) in 1 ml of PBS. Then 20 mg of EDC (0.10 µmol) was added to the above mixture and incubated for 5 h at RT with stirring (Mills et al. 1990; Usleber et al. 1991; Kohno et al. 2003; De Smet et al. 2010).

To obtain DON with one HG linker, namely DON-3-HG-BSA (Figure 3B), prior protection of the 7-OH and 15-OH groups as a cyclic boronate ester allowed one to use similar reaction conditions as for DON-3,15-HG-BSA (Casale et al. 1988). Additionally in the first series of immunogens, a common DON-BSA carbamate conjugate (Figure 3C and 3D) was synthesised by the N,N′-carbonyldiimidazole (CDI) coupling reaction, adopted from the published literature (Xiao et al. 1995; Maragos & McCormick 2000).

Within-series II, DON-CC-BSA (Figure 3E and 3F) was synthesised by first dissolving 1 mg of DON (3.374×10^{-3} mmol) in 480 µl of cold acetonitrile. A solution of 620 µg of cyanuric chloride (3.362×10^{-3} mmol) in 1.24 ml of acetonitrile was prepared and cooled until –20°C. The DON solution was added over 1 h to the vigorously stirred solution of cyanuric chloride. A solution of 870 µg N,N′-diisopropylethylamine in 260 µl of cold acetone was added to the solution of DON and cyanuric chloride and mixed for 5 h at 55–60°C followed by mixing for 16 h at RT. For coupling to BSA, a solution of 1830 µg of BSA (0.0270 µmol) in 1090 µl of 0.1 M sodium tetraborate (pH 9.2) was prepared. The solution was cooled to 4°C and 3.5 mg of DON, a 10-fold molar excess with respect to the available amino groups was added and the resulting mixture was stirred for 1 h at 4°C (Abuchowski et al. 1977; Abuknesha & Griffith 2005).

In an effort to reduce the cross-reactivity against 3-ADON and 15-ADON, we decided to use the C8 carbonyl function of DON to attach an appropriate linker via an oxime moiety (Series III). Hereto a carboxymethyloxime (CMO) strategy was applied (Figure 3G) (Burkin et al. 2000). To investigate the effect of the length of the linker between DON and the protein and with the aim of lowering the immune response against the linker itself, we also introduced a carboxypropyl imine (CPI) linker at the C8 position of DON (Figure 3H). For this purpose, 0.5 mg of DON (1.7 µmol) was reacted with 750 µg of CMO (7 µmol) in 500 µl of dry pyridine for the DON-CMO synthesis and with 1.6 mg of γ-aminobutyric acid (15.5 µmol) in 1 ml dimethylformamide (DMF) for the DON-CPI synthesis. After overnight reaction at RT, the reaction mixture was concentrated and the residue redissolved in 200 µl of water/DMSO (1:1.5, v/v). Next, 2 mg of sulfo-NHS and 4 mg of EDC were added together with 1.775 mg of BSA and the volume was adjusted to 1 ml by further dilution with water. The reactions were performed for 2 h at RT followed by washing and purification as described for previous immunogen synthesis.

Finally, click chemistry was adopted for linking C8 of DON without a carboxyl function. Following a modified procedure of Ikuina et al. (2003), DON was condensed to

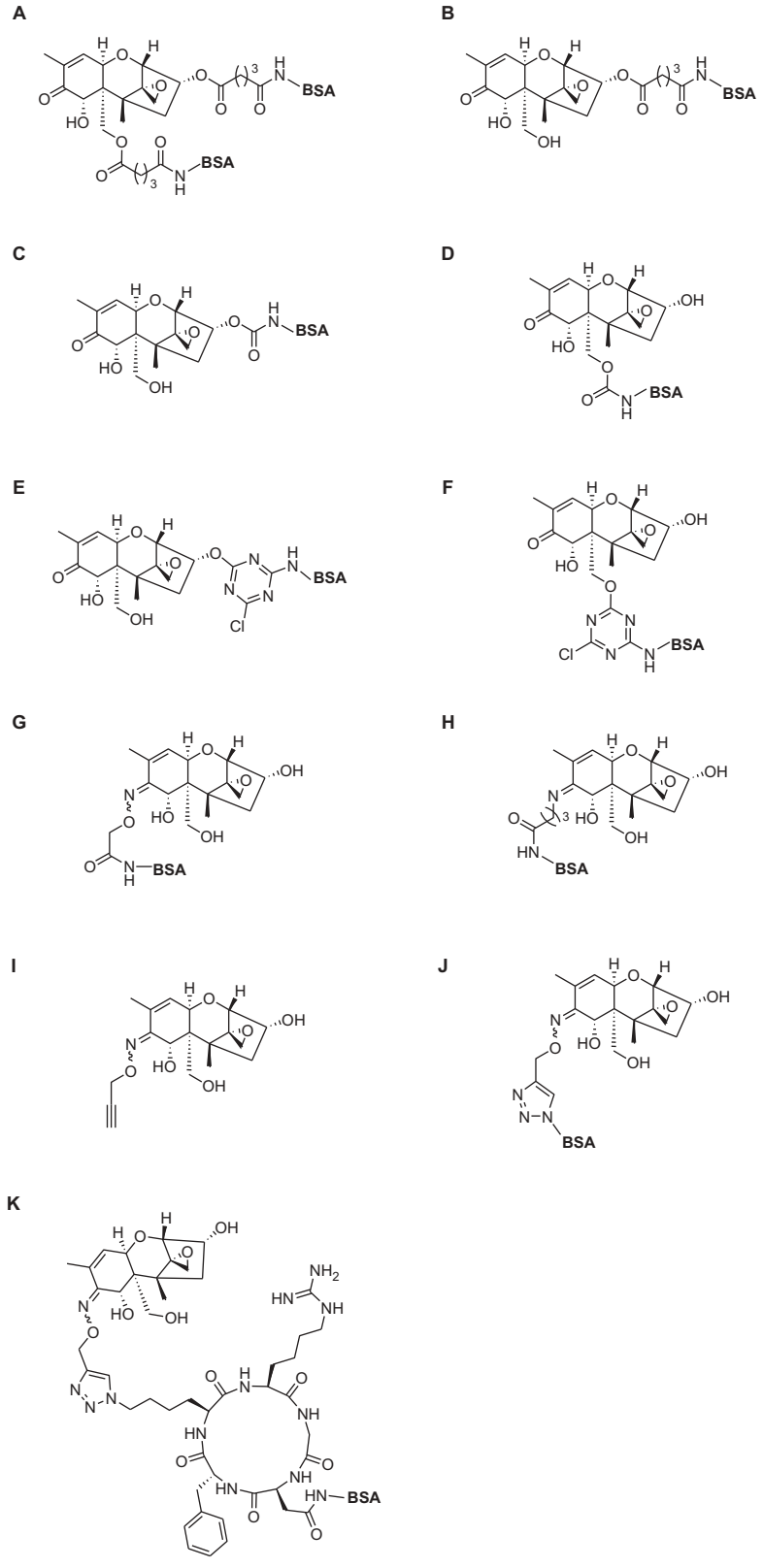


Figure 3. Overview of the deoxynivalenol immunogens (only the amino group of the protein is presented): A, DON-3,15-HG protein; B, DON-3-HG protein; C, DON-3 protein; D, DON-15 protein; E, DON-3-CC protein; F, DON-15-CC protein; G, DON-CMO protein; H, DON-CPI protein; I, DON-oxime; J, DON-azido protein; K, DON-cyclic peptide protein.

O-propargy-hydroxylamine hydrochloride (Figure 3I). Briefly, 2 mg of DON (6.76 µmol) was reacted with 1.81 mg of O-propargyl-hydroxylamine hydrochloride (25.52 µmol) in the presence of THF/acetic acid (1:1, v/v) at 40°C. After 2 h, an additional 170 µg of O-propargyl-hydroxylamine hydrochloride was added and the mixture was stirred for another 4.5 h to obtain DON-oxime. For the synthesis of azido-BSA, 10.686 mg of BSA (0.16 µmol) was dissolved in water together with 4 mg of potassium carbonate and 0.5 mg of copper (II) sulfate. Then, 1 mg of imidazole-1-sulfonylazide hydrochloride (4.79 µmol) was added and the reaction was agitated overnight. For the click reaction, 0.5 mg of DON-oxime (1.43 µmol) was transferred to an aqueous solution of azido-BSA (1.42 mg) containing 25 µg copper (II) sulfate (10 mM), 20 µg sodium ascorbate (10 mM) and 54 µg 4,7-diphenyl-1,10-bathophenanthroline disulfonic acid disodium salt (10 mM). The reaction went on for 14 h at RT (Figure 3J) (Ikuina et al. 2003; Van Dongen et al. 2009; Horak et al. 2010).

The click chemistry immunogen product was further expanded by the introduction of a N-azido cyclic peptide (630.3 g mol^{-1}) (Figure 3K). The cyclic peptide increases the distance between DON and BSA and renders the DON molecule more available for the immune system. Through solid-phase peptide synthesis using 2-chlorotrityl chloride resin and a Fmoc/*t*Bu protection scheme, the amino acids glycine, aspartic acid, D-phenylalanine, lysine and arginine were successively coupled to each other. After mild acid cleavage of the peptide from the resin, cyclisation of the peptide was performed followed by treatment with TFA–TIS–H_2O (95:2.5:2.5, v/v/v) to remove the protecting groups in solution. The peptide was then precipitated in MTBE–hexane (1:1, v/v) and redissolved in methanol for a diazo transfer. For the click chemistry, 1 mg of DON-oxime (2.84 µmol) was reacted with 1.78 mg of cyclic peptide (2.84 µmol) under the previously mentioned conditions. Under CDI reaction conditions, the obtained DON-cyclic peptide (2.84 µmol) was coupled to 3.9 mg of BSA in water–DMSO (1/5, v/v) for 3 h at RT (Dai et al. 2000; Dijkgraaf et al. 2007).

All DON immunogens were dialysed against 4 L PBS to remove low molecular weight substances and concentrated using Pierce Concentrator columns 20 ml/9K MWCO.

Characterisation of immunogens

The successful synthesis of the DON-3,15-HG-linker was confirmed by mass spectrometry LCT Premier XETM TOF (Waters, Milford, MA, USA) equipped with a standard electrospray ionisation and modular LockSpray TM interface in the positive and negative electrospray ionisation ($ESI^{+/-}$) mode. The purified DON-3,15-HG mixture was infused in acetonitrile–water (1:1, v/v) at 10 µl min^{-1}. The purity of the final product was assessed by HPLC and photo diode array (PDA) detection (190–400 nm) using a Phenomenex Luna 2.5 mm C_{18} (2)-HST column. A mobile phase consisting of eluents A (water, 0.1% formic acid) and B (acetonitrile, 0.1% formic acid) was used at a flow rate of 0.4 ml min^{-1}. A linear gradient of 10–100% solvent B was applied over 9 min. Other mass measurements of DON-linker syntheses were performed by direct injection of the reaction product into the Quattro Premier XE mass spectrometer (Waters, Milford, MA, USA) using the $ESI^{+/-}$ mode. Masslynx version 4.1 was used for data acquisition.

After coupling to the protein, the immunogen concentration was determined based on the amount of protein coupled by the use of the NanoDrop 2000c (Thermo Scientific, Rockford, USA) and the immunogen was characterised by the use of indirect competitive ELISA. All incubations except for the first coating step were carried out at 37°C. After each incubation, the plates were washed three times (300 µl/well) with PBST (PBS containing 0.05% Tween 20). High-binding Nunc-Immuno™ F96-well microplates were coated with 10 µg ml^{-1} of the synthesised conjugate in CBS (100 µl/well). After incubation at 4°C overnight, the plates were blocked with 2% skimmed milk in PBS (300 µl/well) for 30 min. Standard solutions of DON and PBS control were added (50 µl/well), followed by adding 50 µl/well of diluted DON monoclonal antibodies (clone 4 or 22) in PBS kindly provided by C. Maragos (USDA). After shaking and incubation for 1 h, 100 µl/well of Sec Ab-HRP was added and incubated for another 1 h. Then, 100 µl/well of TMB substrate solution was added. The reaction was stopped after 15 min with 2 M sulphuric acid (50 µl/well), and the absorbance at 450 nm was measured by a Bio-Rad 550 microplate reader (Richmond, CA, USA).

Immunisations

To obtain antibody-producing B-lymphocytes against DON, 6–8-week-old Balb/C female mice (ethical approval according to ethical commission for animals (ECD) 10/08) were subjected to an injection with the different DON immunogens emulsified with complete or incomplete Freund's adjuvant. For each group of mice, 100 µg of the DON immunogen was administered. Once the mice had reached a sufficient titre, cell fusion of mouse spleen cells and myeloma cells (NSO cells) was performed. Polyethylene glycol (PEG) 1500 was added as fusing reagent and hypoxanthine, aminopterine and thymidine (HAT) for the selection of the fused cells. The cells were distributed into 96-well culture plates and cultured in a humidified, 37°C, 5% CO_2 incubator. Culture supernatant was screened by indirect ELISA to determine the positive hybridomas producing antibody against DON. These hybridomas were further screened for the production of the target antibody and subcloned by limiting dilution.

Characterisation of monoclonal antibodies

Checkerboard assays, in which antibodies were titrated against various amounts of coating antigens, were conducted to select appropriate working concentrations for evaluation of assay sensitivities to DON. Standard competitive curves were obtained by plotting relative absorbance (ratio of absorbance measured at the standard concentration and zero concentration: $B/B_0 \times 100\%$) against the logarithm of analyte concentration. IC_{50} values (i.e., analyte concentrations at which the maximum absorbances were inhibited by 50%) were determined to assess the assay sensitivity. To evaluate the specificity or assay selectivity of the antibody, a set of DON analogues were utilised to perform cross-reactivity studies. The IC_{50} of each tested compound was based on its corresponding competitive curve. Cross-reactivity (CR) values were calculated according to the following equation: CR (%) = [IC_{50} (DON)/IC_{50} (analogue)] × 100%, where CR values were calculated using IC_{50} values with units of μg ml^{-1}.

Results and discussion

Characterisation of immunogens

When the desired synthesis products were obtained based on the exact mass measurements performed by direct injection into the Quattro Premier XE mass spectrometer or the LCT Premier XETM TOF mass spectrometer, further coupling to BSA and OVA or HRP was performed. The synthesised immunogens were characterised by competitive ELISA using reference DON monoclonal antibodies (clone 4 or 22, USDA). If the B_0 value was equal to or higher than 1 and decreasing B values were obtained when using increasing standard DON concentrations, the testing conjugates were confirmed and used for immunisation or further ELISA experiments.

Determination of antisera titres by using indirect ELISA

Antisera titres were determined by an indirect homologous ELISA using the immunising haptens coupled to OVA instead of BSA as coating antigen (Guo et al. 2014). By using the same format, the cross-reactivity against OVA was determined as well. The best antisera titres were obtained for the mice injected with DON-BSA and DON-CC-BSA. For DON-3,15-HG-BSA, high antisera titres were found, but the cross-reactivity against OVA was high. It was concluded that antibodies were probably formed against the protein instead of the target DON. This could be explained by the presence of two linkers on the DON molecule, which makes it less free for activation of the immune system of the mouse. For DON-CMO-BSA, DON-CPI-BSA, DON-azido-BSA and DON-cyclic peptide-BSA, a relatively lower titre, but no cross-reactivity against OVA was determined. When the ELISA response reached a plateau phase using a high serum dilution, the mice were sacrificed by cervical dislocation.

Hybridoma selection and subcloning

By screening and subcloning of the hybridomas, finally, a total of three different anti-DON monoclonal antibodies were obtained. Corresponding to the derived antisera titres, one DON-CC-BSA mouse produced the 13H1 monoclonal antibody, while the 10H10 and 2A9 monoclonal antibodies were produced by two different DON-BSA mice. Previous statement about the immunogenic importance of the C8 carbonyl function in the DON molecule was confirmed as no anti-DON monoclonal antibodies were derived using immunogens with a linker coupled to the C8 carbonyl function (Usleber et al. 1991).

Characterisation of monoclonal antibodies

The characterisation of monoclonal antibodies 13H1 and 10H10 by direct ELISA is described in detail by Guo et al. (2014). The 13H1 and 10H10 monoclonal antibody showed sensitivity values of 2.13 and 0.022 μg ml^{-1}, respectively. For the 13H1 monoclonal antibody the highest cross-reactivity was observed for 15-ADON (CR = 131%) and HT-2 toxin (CR = 187%). Monoclonal antibody 10H10 showed the highest cross-reactivity for 3-ADON (CR = 147%) and 15-ADON (CR = 65%). Characterisation of 10H10 was also repeated by indirect competitive ELISA using DON-OVA (4 μg ml^{-1}) and DON-CC-OVA (4 μg ml^{-1}) coating and DON, 3-ADON and 15-ADON for competition. This comparison between direct and indirect ELISA is illustrated in Table 1. The sensitivity of the antibody towards DON and 3-ADON measured by indirect ELISA is approximately half (DON-OVA coating) and 1/5 (DON-CC-OVA coating) of the value measured by direct ELISA. For 15-ADON, the sensitivity decreased at least ten times when using indirect ELISA. When looking to the CR_{molar} values for direct ELISA and indirect ELISA with DON-OVA coating, the same cross-reactivity is seen for 3-ADON, but the cross-reactivity towards 15-ADON lowered 6.7 times when using indirect ELISA. It can even be concluded that the 10H10 antibody shows only cross-reactivity to 3-ADON when using indirect ELISA with DON-OVA coating. When using an indirect ELISA format with DON-CC-OVA coating, the cross-reactivity against 3-ADON increased three times and the cross-reactivity against 15-ADON decreased four times compared to the direct ELISA format. So, the characterisation of the monoclonal antibody depends on the type of ELISA and the coating antigen used.

Characterisation of monoclonal antibody 2A9 was performed using direct ELISA and the standard curve using 3-ADON standard for competition is shown in Figure 4. The curve represents the relative absorbance

Table 1. Comparison between direct and indirect ELISA for 10H10 antibody characterisation.

			IC_{50} (μg ml^{-1})	CR (%)	IC_{50} molar (nmol ml^{-1})	CR molar (%)
Direct ELISA		DON	0.022	100	0.074	100
		3-ADON	0.015	147	0.044	167
		15-ADON	0.034	65	0.101	74
Indirect ELISA	DON-OVA	DON	0.040	100	0.135	100
		3-ADON	0.028	142	0.083	162
		15-ADON	0.421	9	1.250	11
	DON-CC-OVA	DON	0.113	100	0.383	100
		3-ADON	0.024	480	0.070	547
		15-ADON	0.637	18	1.885	20

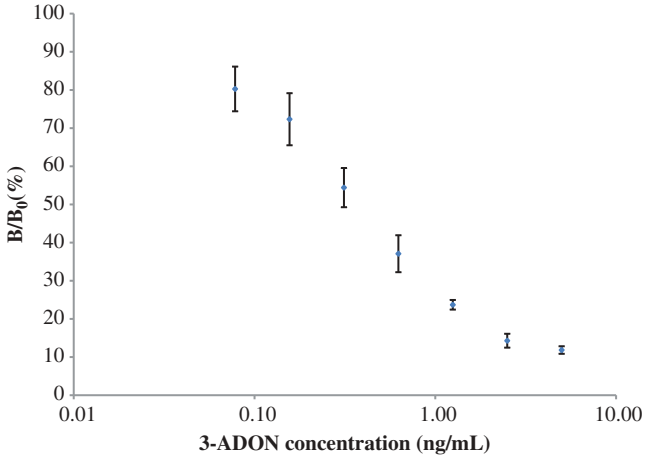


Figure 4. Standard curve of clone 2A9 for 3-ADON by direct competitive ELISA (n = 3).

values (B/B_0) of the ELISA experiment performed on 3 consecutive days. A very sensitive anti-3-ADON antibody was developed with an IC_{50} value of 0.38 ng ml^{-1}. Setting the monoclonal antibody activity for 3-ADON as 100%, the cross-reactivity values for DON, 15-ADON and DON-3-G (Table 2) were determined as 0.188%, 0.088% and 1.498% and can be considered as negligible. In comparison with other previously reported monoclonal antibodies for 3-ADON, the newly developed clone 2A9 is likely the most sensitive and specific antibody to 3-ADON (Casale et al. 1988; Maragos & McCormick 2000; Maragos et al. 2006; Baumgartner et al. 2010).

Conclusions

By using a three-series synthesis strategy for the development of specific monoclonal antibodies against DON, previous statements were confirmed and new conclusions could be made. The place of the linker on the DON molecule is of importance for the immunogenic response. When a linker is positioned on the C3 or C15 of DON for the synthesis of an immunogen, the produced antibody shows higher cross-reactivity against 3-ADON or 15-ADON, respectively. When coupling a linker to the carbonyl C8 of DON, no high anti-DON immune response was observed. This emphasises the immunogenic importance of the C8 carbonyl function. The size of the linker does not seem to influence the immunogenicity of the injected conjugate. The kind of linker has an influence on the electronic configuration of the DON molecule and therefore on the antigenic determinant of DON which is available for the immune system of the mouse.

Three different monoclonal antibodies were developed. One of them (2A9) can be called an anti-3-ADON monoclonal antibody, because of its very sensitive and specific characteristics towards 3-ADON. The second 10H10 monoclonal antibody is a broad specific antibody against DON and its acetylated derivatives 3-ADON and

Table 2. Comparison of cross-reactivity results between clone 2A9 and other DON-MAbs.[a]

	Clone 2A9 direct ELISA		Mab 1–6.2.6 (1) indirect ELISA		USDA Clone 22 (4) direct ELISA	
Compound	IC_{50} (ng ml^{-1})	CR (%)	IC_{50} (ng ml^{-1})	CR (%)	IC_{50} (ng ml^{-1})	CR (%)
3-ADON	0.38	100	1.70	100	2.88	100
DON	202.14	0.188	15.80	10.7	18.20	15.8
15-ADON	431.79	0.088	68.90	2.4	558.00	0.52
DON-3-G	25.36	1.498	n.a.	–	n.a.	–

Notes: [a]Herein, CR (%) = [IC_{50} (3-ADON)/IC_{50} (analogue)] × 100%.
For comparison, the results from reference (4) by Maragos & McCormick (2000) were competitive direct ELISA for clone 22; the results from reference (1) by Maragos et al. (2006) were competitive indirect ELISA for Mab 1–6.2.6.

15-ADON when using it in a direct ELISA format. By using an indirect ELISA format with DON-OVA or DON-CC-OVA coating, the selectivity of this antibody changes resulting in only cross-reactivity against 3-ADON. When combining the 2A9 and 10H10 antibodies in an indirect ELISA format, it is possible to measure the sole DON contamination.

Therefore, we conclude that it is possible to produce specific antibodies against trichothecene mycotoxins such as 3-ADON by synthesising the proper immunogen with the suitable linker and screening the hybridomas carefully. By changing the ELISA format and/or coating antigen, it is possible to influence the selectivity and cross-reactivity of the monoclonal antibody.

Acknowledgements

This work was executed thanks to the contribution of Ateknea Solutions (Hungary), ULUND (Sweden), HGFA (Hungary), SEEDYZ (Greece), CESFAC (Spain), ASEMAC (Spain), Synagra (Belgium), EASRET (Greece), Impuls Ltd (Poland), OSV Srl (Italy), Dunagabona Ltd (Hungary), Dimitriaki S.A. (Greece), ETIA (France), Bioforum S. A. (Greece), and EST Ltd (UK).

Funding

Research leading to the results received funding from the European Community's Seventh Framework Program (FP7/2007-2013) in the frame of the MYCOHUNT project [grant agreement number 243633].

Supplemental data

Supplemental data for this article can be accessed online: http://dx.doi.org/10.1080/19440049.2014.955887

References

Abuchowski A, Van Es T, Palczuk N, Davis F. 1977. Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. J Biol Chem. 252:3578–3581.

Abuknesha R, Griffith H. 2005. Generation of antiserum to Irgarol 1051 and development of a sensitive enzyme immunoassay using a new heterologous hapten derivative. Anal Bioanal Chem. 381:233–243.

Baumgartner S, Führer M, Krska R. 2010. Comparison of monoclonal antibody performance characteristics for the detection of two representatives of A- and B-trichothecenes: T-2 toxin and deoxynivalenol. World Mycotoxin J. 3:233–238.

Berthiller F, Dall'asta C, Corradini R, Marchelli R, Sulyok M, Krska R, Adam G, Schuhmacher R. 2009. Occurrence of deoxynivalenol and its 3-β-D-glucoside in wheat and maize. Food Addit Contam: Part A. 26:507–511.

Burkin A, Kononenko G, Soboleva N, Zotova E. 2000. Preparation of conjugated antigens based on zearalenone carboxymethyloxime and their use in enzyme immunoassay. Appl Biochem Microbiol. 36:282–288.

Casale WL, Pestka JJ, Hart LP. 1988. Enzyme-linked immunosorbent assay employing monoclonal antibody specific for deoxynivalenol (vomitoxin) and several analogs. J Agr Food Chem. 36:663–668.

Dai XD, Su Z, Liu JO. 2000. An improved synthesis of a selective $\alpha_v\beta_3$-integrin antagonist cyclo(-RGDfK-). Tetrahedron Lett. 41:6295–6298.

Danicke S, Goyarts T, Doll S, Grove N, Spolders M, Flachowsky G. 2006. Effects of the Fusarium toxin deoxynivalenol on tissue protein synthesis in pigs. Toxicol Lett. 165:297–311.

De Boevre M, Di Mavungu JD, Landschoot S, Audenaert K, Eeckhout M, Maene P, Haesaert G, De Saeger S. 2012. Natural occurrence of mycotoxins and their masked forms in food and feed products. World Mycotoxin J. 5:207–219.

De Smet D, Monbaliu S, Dubruel P, Van Peteghem C, Schacht E, De Saeger S. 2010. Synthesis and application of a T-2 toxin imprinted polymer. J Chromatogr A. 1217:2879–2886.

Dijkgraaf I, Rijnders AY, Soede A, Dechesne AC, van Esse GW, Brouwer AJ, Corstens FHM, Boerman OC, Rijkers DTS, Liskamp RMJ. 2007. Synthesis of DOTA-conjugated multivalent cyclic-RGD peptide dendrimers via 1,3-dipolar cycloaddition and their biological evaluation: implications for tumor targeting and tumor imaging purposes. Org Biomol Chem. 5:935–944.

Döll S, Dänicke S. 2011. The Fusarium toxins deoxynivalenol (DON) and zearalenone (ZON) in animal feeding. Prev Vet Med. 102:132–145.

Guo Y, Sanders M, Galvita A, Heyerick A, Deforce D, Bracke M, Eremin S, De Saeger S. 2014. Heterologous screening of hybridomas for the development of broad-specific monoclonal antibodies against deoxynivalenol and its analogues. World Mycotoxin J. 7:257–265.

Horak J, Hofer S, Lindner W. 2010. Optimization of a ligand immobilization and azide group endcapping concept via "Click-Chemistry" for the preparation of adsorbents for antibody purification. J Chromatogr B Anal Technologies Biomed Life Sci. 878:3382–3394.

Ikuina Y, Amishiro N, Miyata M, Narumi H, Ogawa H, Akiyama T, Shiotsu Y, Akinaga S, Murakata C. 2003. Synthesis and antitumor activity of novel O-carbamoylmethyloxime derivatives of radicicol. J Med Chem. 46:2534–2541.

[JECFA] Joint FAO/WHO Expert Committee on Food Additives. 2011. Safety evaluation of certain contaminants in food. Prepared by the seventy-second meeting of the Joint FAO/WHO Expert Committee on Food Additives. *WHO Food Additives Series 63*.

Kohno H, Yoshizawa T, Fukugi M, Miyoshi M, Sakamoto C, Hata N, Kawamura O. 2003. Production and characterization of monoclonal antibodies against 3,4,15-triacetylnivalenol and 3,15-diacetyldeoxynivalenol. Food Agric Immunol. 15:243–254.

Krska R, Baumgartner S, Josephs R. 2001. The state-of-the-art in the analysis of type-A and -B trichothecene mycotoxins in cereals. Fresenius J Anal Chem. 371:285–299.

Maragos C, Busman M, Sugita-Konishi Y. 2006. Production and characterization of a monoclonal antibody that cross-reacts with the mycotoxins nivalenol and 4-deoxynivalenol. J Food Addit Contam. 23:816–825.

Maragos CM, Li L, Chen D. 2012. Production and characterization of a single chain variable fragment (scFv) against the mycotoxin deoxynivalenol. Food Agric Immunol. 23:51–67.

Maragos CM, McCormick SP. 2000. Monoclonal antibodies for the mycotoxins deoxynivalenol and 3-acetyl-deoxynivalenol. Food Agric Immunol. 12:181–192.

Mills C, Alcock S, Lee H, Morgan M. 1990. An enzyme-linked immunosorbent assay for deoxynivalenol in wheat, utilizing novel hapten derivatization procedures. Food Agric Immunol. 2:109–118.

Pestka JJ. 2007. Deoxynivalenol: toxicity, mechanisms and animal health risks. Anim Feed Sci Technol. 137:283–298.

Pinton P, Tsybulskyy D, Lucioli J, Laffitte J, Callu P, Lyazhri F, Grosjean F, Bracarense AP, Kolf-Clauw M, Oswald IP. 2012. Toxicity of deoxynivalenol and its acetylated derivatives on the intestine: differential effects on morphology, barrier function, tight junction proteins, and mitogen-activated protein kinases. Toxicol Sci. 130:180–190.

Sinha R, Savard M, Laur R. 1995. Production of monoclonal antibodies for the specific detection of deoxynivalenol and 15-acetyldeoxynivalenol by ELISA. J Agr Food Chem. 43:1740–1744.

Usleber E, Maertlbauer E, Dietrich R, Terplan G. 1991. Direct enzyme-linked immunosorbent assays for the detection of the 8-ketotrichothecene mycotoxins deoxynivalenol, 3-acetyldeoxynivalenol, and 15-acetyldeoxynivalenol in buffer solutions. J Agr Food Chem. 39:2091–2095.

van Dongen SFM, Teeuwen RLM, Nallani M, van Berkel SS, Cornelissen JJLM, Nolte RJM, van Hest JCM. 2009. Single-step azide introduction in proteins via an aqueous diazo transfer. Bioconjug Chem. 20:20–23.

Xiao H, Clarke JR, Marquardt RR, Frohlich AA. 1995. Improved methods for conjugating selected mycotoxins to carrier proteins and dextran for immunoassays. J Agr Food Chem. 43:2092–2097.

An immunogen synthesis strategy for the development of specific anti-deoxynivalenol monoclonal antibodies

Melanie Sanders[a], Yirong Guo[a,b,] Abhishek Iyer[c], Yara Ruiz García[c], Anastasia Galvita[a], Arne Heyerick[d], Dieter Deforce[d], Martijn D.P. Risseeuw[e], Serge Van Calenbergh[e], Marc Bracke[f], Sergei Eremin[g], Annemieke Madder[c] and Sarah De Saeger[a]

[a]Laboratory of Food Analysis, Ghent University, Ghent, Belgium; [b]Institute of Pesticide and Environmental Toxicology, Zhejiang University, Hangzhou, China; [c]Organic and Biomimetic Chemistry Research Group, Ghent University, Ghent, Belgium; [d]Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium; [e]Laboratory for Medicinal Chemistry, Ghent University, Ghent, Belgium; [f]Laboratory of Experimental Cancer Research, Ghent University, Ghent, Belgium; [g]Department of Chemical Enzymology, Moscow State University, Moscow, Russia

Online Supplementary Material

The cyclic peptide was synthesized using manual Fmoc solid phase peptide synthesis (SPPS) according to the following procedure:

Coupling of first Amino acid:
Chlorotrityl chloride resin (100 mg, 0.16 mmol, loading = 1.6 mmol/g) was swollen in dichloromethane (dry, 10 mL/g) prior to the loading of the first amino acid. Fmoc-Gly-OH (0.190 g, 0.64 mmol) was dissolved in dry DCM with N,N'-diisopropylethylamine (DIPEA) (0.22 mL, 1.28 mmol) and the reaction was shaken for 3 h. The resin was then washed with 3 x DCM, 3 x DMF, 3 x DCM and was dried under vacuum for 2 h. The loading was determined via Fmoc determination and was calculated to be 0.9 mmol/g (82% yield). The resin was then capped with methanol/DIPEA/DCM = 2:1:17 by shaking for 2 h.

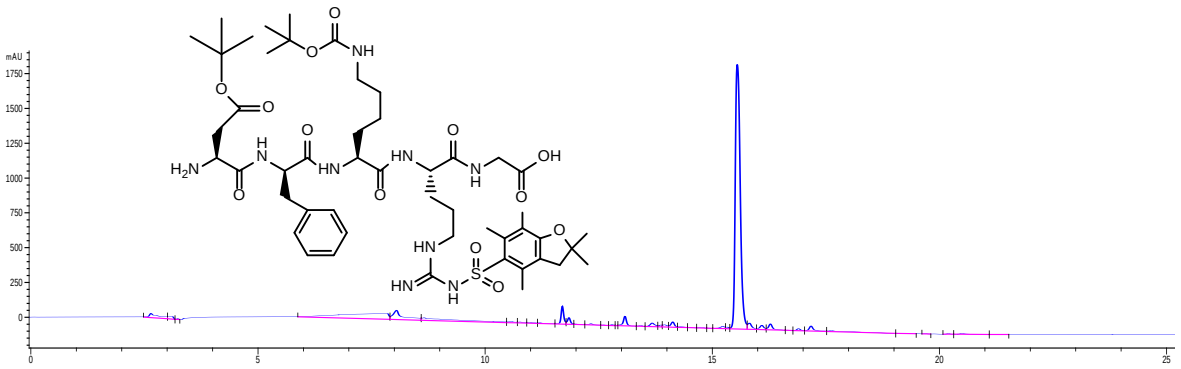
Fmoc deprotection for the first amino acid:
The resin was swollen in DMF prior to Fmoc deprotection. A solution of 20% piperidine in DMF was added to the resin and the mixture was shaken for 30 min. The solution was then washed with 3 x DMF, 3 x methanol, 3 x DMF. This protocol was repeated to ensure complete Fmoc removal.

Coupling for the remaining amino acids:
Fmoc protected glycine (38.9 mg, 0.131 mmol) along with benzotriazol-1-yl oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) (273.2 mg, 0.525 mmol) were dissolved in dry DMF. DIPEA (0.183 mL, 1.048 mmol) was added to it and the mixture was shaken for 30 s. The mixture was then added to resin pre-swollen in DMF (dry, 10 mL/g) and the reactor was shaken for 3 h. The solution was filtered followed by washing with 3 x DMF, 3 x methanol, 3 x DMF. Using this protocol, Fmoc-L-Arg(Pbf)-OH, Fmoc-L-Lys(Boc)-OH, Fmoc-D-Phe-OH and Fmoc-L-Asp-OH were coupled consecutively.

Fmoc deprotection for the remaining amino acids:
The resin was swollen in DMF prior to Fmoc deprotection. A solution of 20% piperidine in DMF was added to the resin and the mixture was shaken for 2 min, 5 min and 15 min. The solution was washed with 3 x DMF, 3 x methanol, 3 x DMF after each deprotection step.

Mild acid resin cleavage:
After assembly of the linear peptide, the resin pre-swollen in DCM (dry, 10 mL/g) was treated with trifluoroethanol (TFE)/acetic acid (AcOH)/DCM (1/1/3, v/v/v) for 2 h to ensure complete cleavage of the protected peptide from the resin. The solvents were evaporated under vacuum and the residual material was analysed using reversed phase (RP)-HPLC/MS (Phenomenex Luna C18 column (250 x 4.6 mm, 5 µm at 35 °C) connected to an ESI mass spectrometer. A mobile phase consisting of 5 mM ammonium acetate (A) and acetonitrile (B) was used at a flow rate of 1 mL/min. The gradient elution

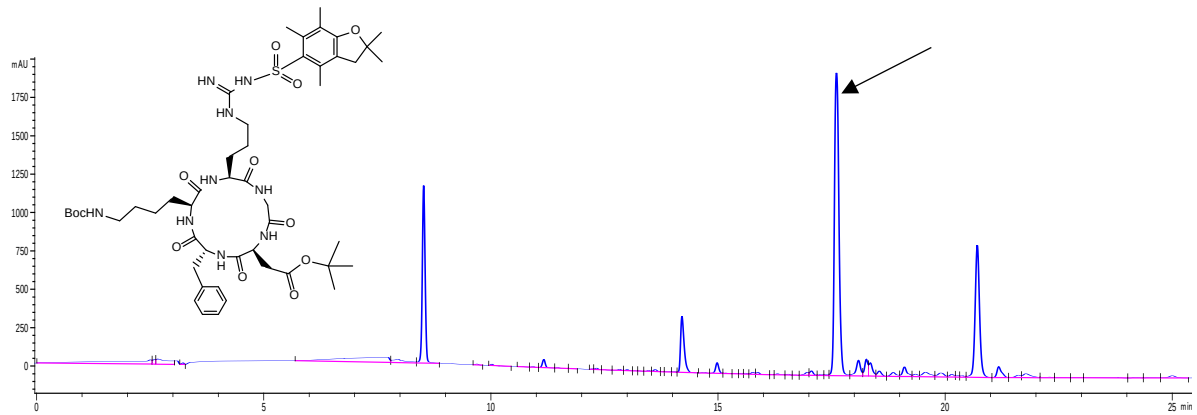
programme started with 100% mobile phase A for 2 min with a linear decrease to 0% mobile phase A in 15 min. An isocratic gradient of 100% mobile phase B was continued for 5 min.



HPLC trace of the protected linear peptide ($C_{49}H_{75}N_9O_{13}S$): eluting at 15.68 min with calculated *m/z*: 1029.52 and observed *m/z* (ESI^+): 1030.4

Peptide Cyclisation:

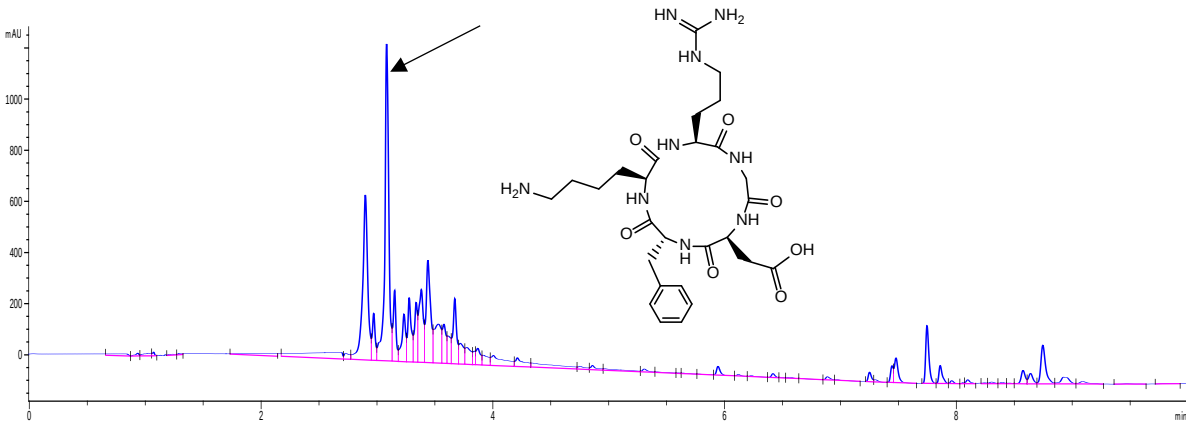
The crude linear peptide (77.5 mg, 0.077 mmol) was dissolved in 5 mL dry DCM, followed by the addition of 1-hydroxybenzotriazole (HOBt) (10.4 mg, 0.077 mmol), PyBOP (40.1 mg, 0.077 mmol) and DIPEA (30.5 mL, 0.175 mmol). The mixture was stirred for 24 h at room temperature. Only the cyclic peptide was observed as the main peptide product on HPLC/MS (HPLC/MS conditions same as above).



HPLC trace of the protected cyclic peptide ($C_{49}H_{73}N_9O_{12}S$): eluting at 17.66 min with calculated *m/z*: 1011.53 and observed *m/z* (ESI^+): 1012.4

Removal of protecting groups:

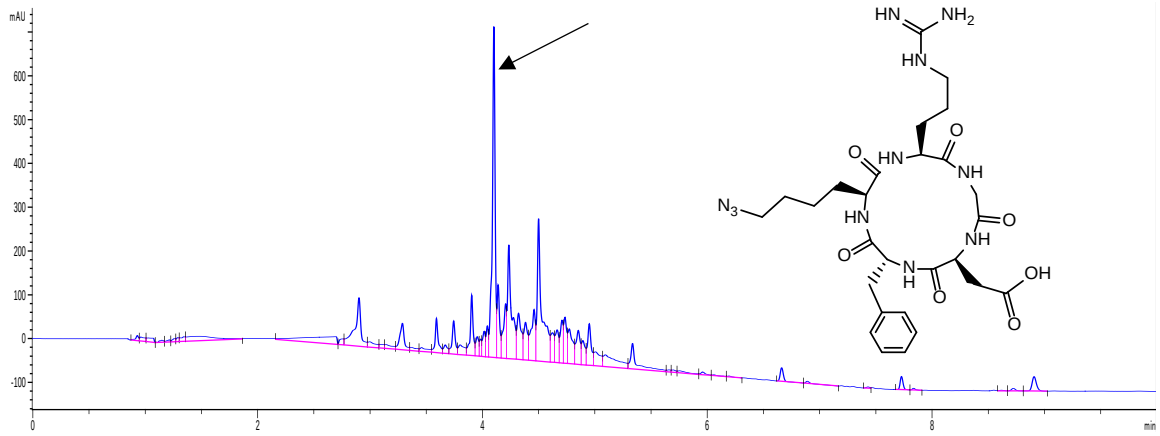
The reaction mixture was dissolved in 5 mL TFA/TIS/H_2O (95/2.5/2.5, v/v/v) to remove the protecting groups in solution. The peptide was then precipitated in MTBE/Hexane (1/1, v/v) and the reaction mixture was analysed using RP-HPLC/MS (Phenomenex Kinetex C18 100Å column (150 x 4.6 mm, 5 µm at 35 °C) connected to an ESI mass spectrometer. A mobile phase consisting of 5 mM ammonium acetate (A) and acetonitrile (B) was used at a flow rate of 1 mL/min. The gradient elution programme started with 100% mobile phase A for 1.5 min with a linear decrease to 0% mobile phase A in 6 min. An isocratic gradient of 100% mobile phase B was continued for 2.5 min.



HPLC trace of the deprotected cyclic peptide ($C_{27}H_{41}N_9O_7$)): eluting at 3.10 min with calculated *m/z*: 603.31 and observed *m/z* (ESI^+): 604.2

Diazo transfer:

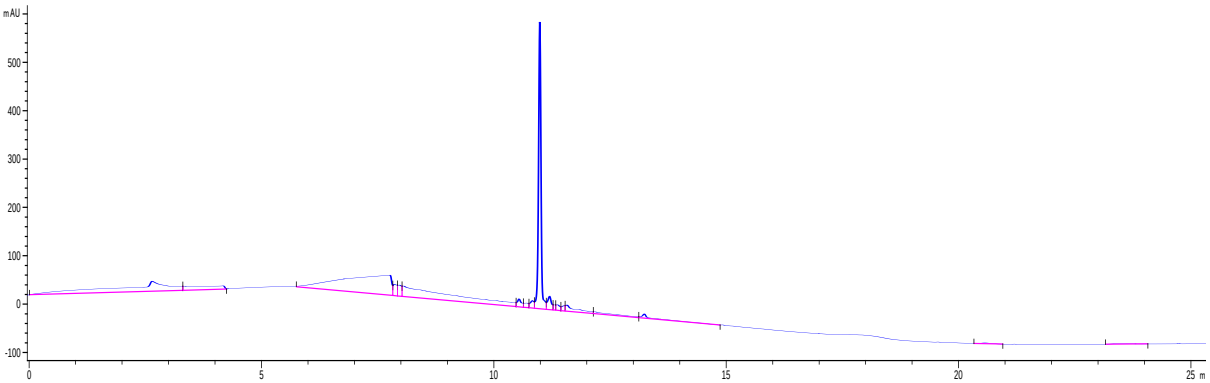
A diazo transfer was performed on 50.5 mg of the deprotected cyclic peptide using the imidazole-1-sulfonylazide hydrochloride reagent as described by Goddard-Borger and Stick (2007). The conversion was complete and no starting material was observed when analysed with HPLC/MS. The reaction mixture was redissolved in methanol and purified using RP-HPLC.



HPLC trace of the deprotected crude cyclic peptide after diazo transfer ($C_{27}H_{39}N_{11}O_7$): eluting at 4.10 min with calculated *m/z*: 629.30 and observed *m/z* (ESI^+): 630.2

HPLC purification of crude peptide:

The crude peptide was purified using RP-HPLC on an Agilent 1100 Series using a Phenomenex semiprep Luna C18 (250 x 10 mm) column. A mobile phase consisting of 0.1% TFA in H_2O (A) and acetonitrile (B) was used at a flow rate of 4.5 mL/min at 35°C. The gradient elution programme started with 100% mobile phase A for 2 min with a linear decrease to 0% mobile phase A in 15 min. An isocratic gradient of 100% mobile phase B was continued for 5 min.



HPLC trace of the final cyclic peptide after HPLC purification eluting at a retention time of 10.95 min.

References:

Dijkgraaf, I., Rijnders, A. Y., Soede, A., Dechesne, A. C., van Esse, G. W., Brouwer, A. J. et al. (2007). Synthesis of DOTA-conjugated multivalent cyclic-RGD peptide dendrimers via 1,3-dipolar cycloaddition and their biological evaluation: implications for tumor targeting and tumor imaging purposes. *Organic & Biomolecular Chemistry, 5,* 935-944.

Goddard-Borger, E. D. & Stick, R. V. (2007). An efficient, inexpensive, and shelf-stable diazotransfer reagent: Imidazole-1-sulfonyl azide hydrochloride. *Organic Letters, 9,* 3797-3800.

ANNEX II:
CURRICULUM VITAE

RESUME

PERSONAL DETAILS

Full Name: Abhishek Ramakrishna Iyer
Date of Birth: 25/07/1987
Nationality: Indian
Marital Status: Single
Current Address: 24 West Parade, Lincoln, LN11JT, United Kingdom
Mobile: +44(0)7477233783
E-mail Id: usariyer@gmail.com

ACADEMIC PROFILE

YEAR	INSTITUTE	DEGREE	CGPA / %	RANK
2014 –	University of Lincoln, United Kingdom	Doctor in Philosophy, Pharmacy & Pharmaceutical Sciences (as part of a joint PhD)	-	-
2012 – 2014	Ghent University, Belgium	Doctor in Philosophy, Chemistry (as part of a joint PhD)	-	-
2007 – 2012	Indian Institute of Technology, Kharagpur	Masters of Science in Industrial Chemistry	8.02/10	1st
2006	Swami Vivekanand Junior College	Maharashtra State Board (Class XII)	81%	–
2004	Udayachal High School	Maharashtra State Board (Class X)	90%	1st

ARTICLES PUBLISHED

- Total syntheses and biological evaluation of two teixobactin analogues. Anish Parmar, Abhishek Iyer, et. al. *Chem. Comm. 52, 6060-6063*. **DOI**: 10.1039/C5CC10249A. **2016**
- Stapling the monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake. Abhishek Iyer, et. al. *Org. Biol. Chem.* **2015**, 13, 3856-3862. **DOI**: 10.1039/C4OB02659D **2015**
- Sequence selective recognition and enhanced cellular uptake by peptide-steroid conjugates prepared through CuAAC. Yara Ruiz García, Abhishek Iyer, et. al. *Chem. Commun.,* **2015**, 51, 17552-17555. **DOI**:10.1039/C5CC07097J **2015**
- Cyclodextrin-peptide conjugates for sequence specific DNA binding. Yara Ruiz García, Jan Zelenka, Y. Vladimir Pabón Martínez, Abhishek Iyer, et. al. *Org. Biomol. Chem.*, 2015,**13**, 5273-5278 **DOI**: 10.1039/C5OB00609K **2015**
- An immunogen synthesis strategy for the development of specific anti-deoxynivalenol monoclonal antibodies. *Food additives and contaminants: Part A*. Melanie Sanders, Yirong Guo, Abhishek Iyer, et. al. **DOI**: 10.1080/19440049.2014.955887. **2014**

PhD PROJECTS

PhD at University of Lincoln, Lincoln, United Kingdom (as part of a joint PhD) **Sep. 2014 – present**

Development of antimicrobial compounds

Promoter: Dr. Ishwar Singh, School of Pharmacy, University of Lincoln, Lincoln, United Kingdom

Fellowship from University of Lincoln

- Successfully design and synthesized antimicrobial compounds for targeting Gram positive and Gram negative bacteria
- Synthesis of specialized building blocks for the synthesis of DNA cross-linking agents

PhD at Ghent University, Gent, Belgium (as part of a joint PhD) **Sep. 2012 – Aug. 2014**

Design and synthesis of stapled peptides for DNA recognition and enhanced cell penetration

Promoter: Dr. Annemieke Madder, Department of Organic and Macromolecular Chemistry, Ghent University, Belgium

PhosChemRec Network. Marie Curie Early Stage Research Training Fellowship

Successfully completed and published my work comprising 5 DNA binding stapled peptides synthesized via a cysteine cross-linking methodology and their concurrent cell uptake studies

OTHER PROJECTS / SUMMER INTERNSHIPS

Master Thesis, IIT Kharagpur, India **Aug. 2011 – June 2012**

To improve the yield of the Kinugasa reaction using Copper Iodide Nanoparticles

Guide: Prof. Amit Basak, Department of Chemistry, IIT Kharagpur

- Successfully synthesized Copper Iodide nanoparticles as catalyst for the Kinugasa reaction
- Improved the yield of the reaction, along with the time taken and the stereoselectivity of the products obtained.

Summer Internship at Ghent University, Belgium **May – July 2011**

To synthesize and purify a cyclic peptide and conjugate it with a mycotoxin to generate antibodies

Guide: Prof. Dr. Annemieke Madder, Department of Organic Chemistry, Ghent University, Belgium

- Successfully synthesized and purified the peptide N-azido cyclo (Arg-Gly-Asp-D-Phe-Lys) using Solid Phase Peptide Synthesis protocols.
- Compound submitted to the Department of Pharmacy at Ghent University for testing and to combine with a mycotoxin in an attempt to generate antibodies against the toxin

Short-term Industrial Project with Damodar Valley Corporation, India **Aug. – Sept. 2008**

Reduce the alkalinity and removal of impurities from raw river water

Guide: Prof. Manish Bhatacharjee, Department of Chemistry, IIT Kharagpur

- Analysis of river water sample to calculate exact pH value and other impurities in ppm
- Methods researched & suggested included use of acidic $AlCl_3$ (which was successfully regenerated) to reduce alkalinity and Purolite for Silica removal
- Analysis and calculations of flow rates and rate of regeneration based on theoretical models and columns of different

LAB EXPERIENCE

- Organic synthesis with specialization in peptide chemistry
- Cysteine cross-linking, formation of S-S bonds in peptides
- Synthesis of cyclic, branched and constrained peptides
- Peptide-drug conjugation
- Synthesis of Dervan Peptides
- Analytical techniques: NMR, HPLC, LC-MS, ESI-MS MALDI-TOF, Gel Shift Assay (EMSA), Circular Dichroism (CD)

CONFERENCES ATTENDED

DATE	NAME OF THE CONFERENCE	PARTICIPATION	LOCATION
17-18th Feb. 2016	Belgian Peptide Group Meeting	Oral Presentation	Brussels (BE)
27th May 2015	Dutch Peptide Symposium	Oral Presentation	Nijmegen (NL)
4th – 5th Dec. 2014	Sigma Aldrich Organic Synthesis Meeting	Oral presentation	Blankenberge (BE)
31st Aug – 5th Sep. 2014	European Peptide Symposium	Poster presentation	Sofia (Bulgaria)
14th Feb – 17th Feb. 2014	7th and final PhosChemRec Network Meeting	Oral presentation	Sheffield (UK)
29th Aug. – 2nd Sep. 2013	6th PhosChemRec Network Meeting	Oral presentation	Cambridge (UK)
5th – 6th Dec. 2013	Sigma Aldrich Organic Synthesis Meeting	-	Blankenberge (BE)
24th Apr. 2013	Dutch Peptide Symposium	-	Amsterdam (NL)
15th – 18th Jan. 2013	5th PhosChemRec Network Meeting	Oral presentation	Frankfurt (GE)