Synthesis and Characterisation of Peptide-Oligonucleotide Conjugates Derived from β-sheet Forming or Collagen Mimetic Peptides

Akiko Sato

A thesis submitted for the degree of Doctor of Philosophy

School of Physical Sciences

University of Kent

July 2021

Abstract

Peptide-oligonucleotide conjugates (POCs) are hybrid compounds consisting of peptides covalently linked to nucleic acids such as DNA or RNA. Integrating the unique characteristics of oligonucleotides with peptides through conjugation has the potential to create novel hybrids that surpass the properties within the individual biomolecules. Despite the value that POCs could bring to biomaterial and biomedical sciences, finding a suitable conjugation strategy remains a critical challenge.

This project explored two conjugation strategies (thiol-maleimide and amide-coupling reactions) for generating a variety of POCs comprising β-sheet-forming or collagen-mimetic peptides and DNA oligonucleotides. Different purification strategies were conducted following POC synthesis to isolate the desired conjugates. The self-assembly of POCs was investigated using AFM following hybridisation and annealing experiments to explore their unique structural properties.

Thiol-maleimide chemistry was confirmed to be an effective strategy for conjugating β-sheet-forming and collagen-mimetic peptides with DNA oligonucleotides. Hybridisation of conjugate with DNA resulted in reduced fibril formation compared with the original peptide sequence. Conjugate-to-conjugate hybridisation showed large rod-like structures that were not seen in the original peptide sequence, whereas fibril formation was not observed. Annealing of conjugates led to the formation of rod-like structures or dot-like structures that were not seen in the original peptide sequences.

This research demonstrated that hybridisation and annealing caused POCs to acquire structural characteristics that were not observed in the individual non-conjugated biomolecules, highlighting the potential of peptide-oligonucleotide conjugation for developing novel biomaterials with emergent properties.

Acknowledgements

Firstly, I would like to express my special appreciation to my supervisors, Dr Christopher J. Serpell and Dr Michael R. Reithofer for giving me the opportunity to undertake this project and for their guidance throughout the course of this research. I would also like to thank the Leverhulme Trust for its generous financial support that allowed me to dedicate myself to this research.

I am deeply grateful to my fellow colleague from the Serpell group, Emerald R. Taylor, for synthesising maleimide-functionalised DNA and for her generous support during this project. My gratitude extends to all the members in the research group for their insightful suggestions and kind help over the past years.

I am thankful to Kevin Howland, from the School of Biosciences, for helping me conduct LC-MS measurements on countless occasions. I am also grateful to Dr Andrew Morrell, from the School of Physical Sciences, for giving me access to the LC-MS instrument, and to Dr Wei-Feng Xue from the School of Biosciences, for giving me access to the AFM instrument. Without their support, it would not have been possible to conduct this research.

Lastly, I would like to thank my parents and friends for their encouragement and support over the course of my studies. In the last instance, I would like to express my gratitude to Juan, who provided me with invaluable support during these tough years.

Table of Contents

List of Abbreviations and Acronyms

Abbreviation	Meaning
δ	NMR chemical shift
λ	Wavelength
µL	Microlitre
µM	Micromolar
A	Adenine
Ac	Acetyl
ACN	Acetonitrile
Ac_2O	Acetic anhydride
AFM	Atomic force microscopy
Ala (A)	Alanine
Ar	Aryl
Arg (R)	Arginine
Asn (N)	Asparagine
^{t}Bu	*tert*-butyl
C	Cytosine
°C	Degree Celsius
CD	Circular dichroism
$CDCl_3$	Chloroform
CTC	2-Chlorotrityl Chloride
Cys (C)	Cysteine
d	Doublet
Da	Dalton
DMAP	4-dimethylamino pyridine
DCM	Dichloromethane
dd	doublet of doublets
DIPEA	N,N-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
DTT	Dithiothreitol

EDC.HCl	N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray
Et_2O	Diethyl ether
Equiv	Equivalent
Fmoc	Fluorenylmethyloxycarbonyl
G	Guanine
g	Grams
Gly (G)	Glycine
HCl	Hydrochloric acid
HFIP	Hexafluoroisopropanol
HIC	Hydrophobic interaction chromatography
HOBt	1-hydroxybenzotriazole hydrate
HPLC	High performance liquid chromatography
Hyp (O)	Hydroxyproline
Hys (H)	Histidine
Hz	Hertz
Ile (I)	Isoleucine
J	NMR coupling constant
KCN	Potassium cyanide
Leu (L)	Leucine
Lys (K)	Lysine
LC-MS	Liquid chromatography mass spectrometry
M	Multiplet
mA	Milliampere
mal	Maleimide
MALDI	Matrix-assisted laser desorption/ionization
MBHA	Methylbenzhydryl amine
Me	Methyl
Met (M)	Methionine
MeOH	Methanol
mg	Milligram
MHz	Megahertz
mL	Millilitre

mM	Millimolar
mol	Moles
MS	Mass spectrometry
MW	Molecular weight
m/z	Mass to charge ratio
NaCl	Sodium chloride
nm	Nanometre
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
pH	$-\log_{10}[H]^{+}$
Ph	Phenyl
Phe (F)	Phenylalanine
POC	Peptide-oligonucleotide conjugate
ppm	Parts per million
Pro (P)	Proline
PS	Polystyrene
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
q	Quartet
RNA	Ribonucleic acid
RT	Room temperature
S	Serine
s	Singlet
SPPS	Solid-phase peptide synthesis
T	Thymine
t	Triplet
TAC	Trichloroacetic acid
TAMg	Tris-acetate magnesium
TBE	Tris-borate-EDTA
TBTU	2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethylaminium tetrafluoroborate
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
TLC	Thin layer chromatography

Tyr (Y)	Tyrosine
UV-Vis	Ultraviolet visible spectroscopy
V	Volt
v	Volume
Val (V)	Valine

1 Introduction

1.1 Supramolecular Chemistry

1.1.1 Overview

Supramolecular chemistry refers to chemistry beyond the molecule, building on the concept of molecular recognition with the aim of developing highly complex and functional chemical systems through non-covalent interactions of molecular components.[1] This field dates back to the 1960s when a number of researchers across various scientific disciplines, including Charles J. Pedersen, Donald J. Cram and Jean-Marie Lehn, began to study the chemistry of molecules beyond the covalent bond.[2] In contrast to conventional chemistry focusing on covalent bonds, supramolecular chemistry opens new possibilities by allowing researchers to further explore molecular behaviour through the weaker, yet more dynamic non-covalent intermolecular forces.[3]

The origin of supramolecular chemistry stems from the discovery of crown ethers by Charles Pedersen in 1967, embodying the concept of molecular recognition using synthetic molecules.[4] Crown ethers are cyclic polyethers that possess selective inclusion properties (Figure 1.1).

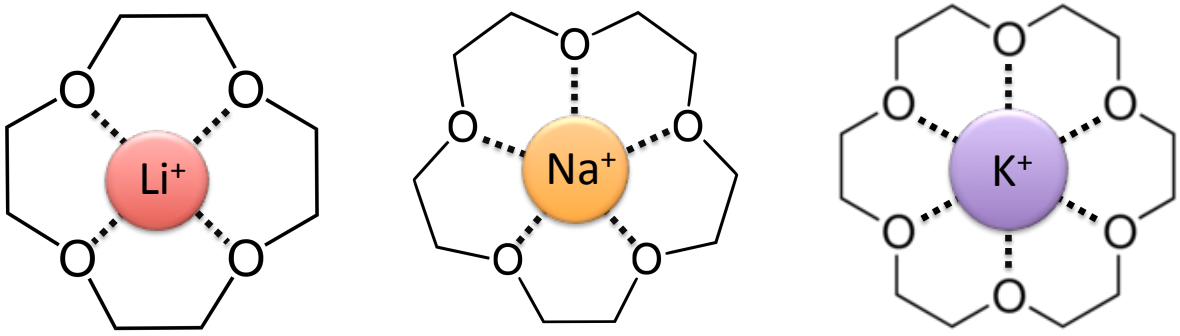


Figure 1.1. Examples of crown ethers from left to right: [12] crown-4 a host for Li^+, [15] crown-5 a host for Na^+ and [18] crown-6 a host for K^+. The driving force of the inclusion comes from electrostatic interactions between the metal cation and the oxygen atoms (anion).

Due to their cyclic structures, crown ethers selectively bind to metal ions based on their cavity size, acting as functional molecules through molecular recognition. Following Pedersen's discovery, Donald Cram advanced the concept of molecular recognition to a wide range of molecular systems, leading to the birth of host-guest chemistry.[5,6] This field diverged from conventional chemistry by focusing on non-covalent interactions between host and guest molecules. Host-guest chemistry highlighted the relations between molecular structure and functions, bringing new insights into the design and synthesis of functional molecular systems by exploiting non-covalent interactions. The term supramolecular chemistry was eventually proposed by Jean-Marie Lehn to refer to a multidisciplinary field that would act as the bridging point across host-guest chemistry, molecular self-assembly, and conventional scientific fields including chemistry, physics and biology.[1]

Non-covalent interactions act as driving forces for the formation of higher-order structures, being important for understanding biological processes and designing nanomaterials.[7] Several types of non-covalent interactions exist, including hydrogen bonds, π-π stacking, electrostatic interactions, hydrophobic interactions, coordination interactions and van der Waals forces.

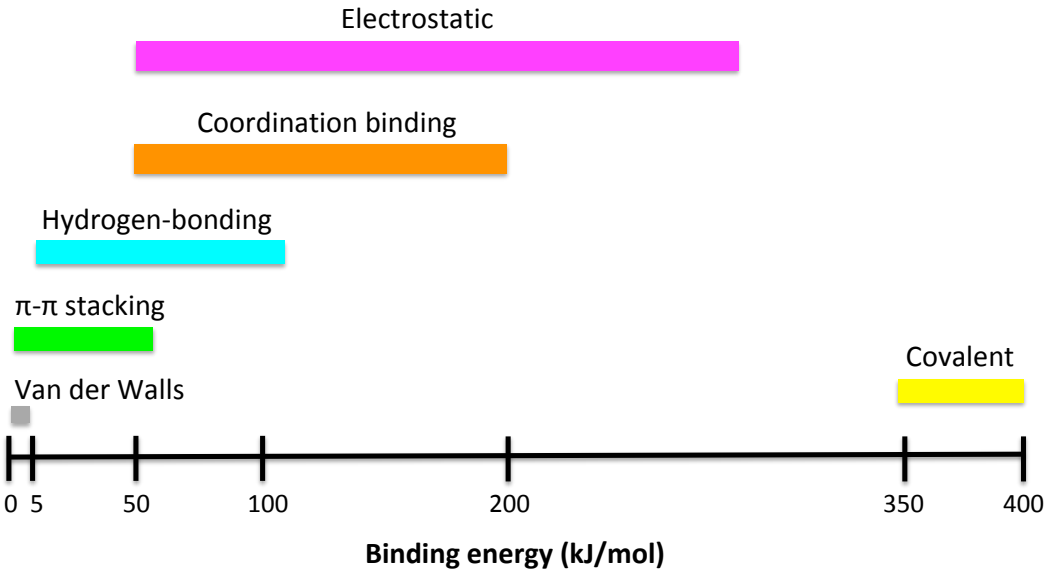


Figure 1.2. Strength of common non-covalent interactions.[8]

While non-covalent interactions may be weak on their own (Figure 1.2), their combination can act synergistically to form stable associations between molecules that lead to the generation of well-defined structures.[9] This process, whereby individual molecules spontaneously form into higher-order structures through non-covalent interactions is referred to as self-assembly (Figure 1.3).[10]

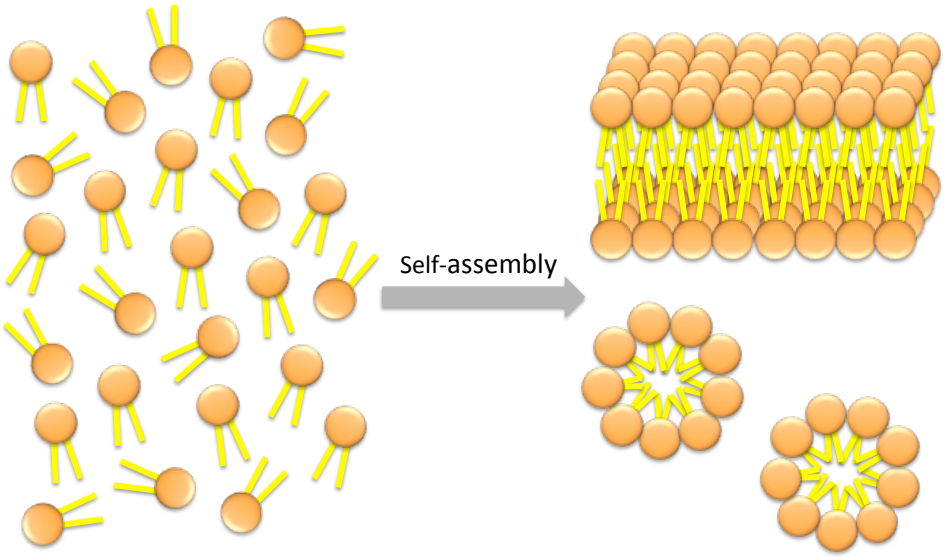


Figure 1.3. Schematic illustration showing the process of self-assembly in which individual molecules spontaneously form into higher-order structures through non-covalent interactions.

Supramolecular chemistry highlights the notion of emergence, whereby self-assembled structures acquire greater properties than those present in the sum of their individual molecular components.[11] Through emergence, the resultant self-assembled structures can reach new levels of complexity with novel characteristics that were previously absent in the individual components. The self-assembly of molecules plays a fundamental role in generating organised structures across all scales of matter, ranging from simple molecules to highly complex living organisms.[12] Molecular self-assembly serves the foundation for the generation of most biological nanostructures, including phospholipid bilayers of cells, double-helical structures of DNA, as well as higher-ordered protein structures.[13] Hence, molecular self-assembly is a key driver for the origination and maintenance of life.

By utilising the concept of self-assembly, highly sophisticated and functional materials in synthetic systems can be developed due to the intrinsic ability of molecules to self-organise into well-defined architectures.[14] While the molecules are developing into more complex structures, non-covalent bonds provide a degree of flexibility that facilitates reversible self-assembly. Unveiling the precise mechanisms of molecular self-assembly driven by non-covalent interactions has the potential to accurately predict the resultant structures, opening new possibilities for developing life-like materials with self-replicating,[15] self-catalysing[16] or regenerative properties[17]. Nevertheless, precisely predicting the final structure remains difficult given that the self-assembly processes are governed by the random movements of molecules. Consequently, improving our understanding on how self-assembling systems can generate a high level of complexity similar to that seen in biological systems is a key goal for supramolecular chemistry.[18]

1.1.2 Static and Dynamic Self-assembly

Self-assembly can be classified as either static or dynamic.[19] While static self-assembly has been widely studied in the field of supramolecular chemistry, dynamic self-assembly has not been explored to the same extent.[20] Static self-assembly occurs at either global or local equilibrium following the law of thermodynamics (Figure 1.4 a), leading to structures that are thermodynamically stable.[19] Examples of structures derived from static self-assembly include micelles, lipid bilayers, DNA double helices and amyloid fibrils, all of which can be produced in a laboratory environment. Given that complex molecular structures can be generated through static self-assembly by exploiting multiple non-covalent interactions, this type of self-assembly has been utilised to design functional synthetic nanomaterials.[21] While this approach has led to major advances in nanotechnology, the structures generated are always thermodynamically stable and thus cannot mimic the dynamic properties seen in living systems.

In contrast to static self-assembly, dynamic self-assembly occurs outside thermodynamic equilibrium by continuously consuming energy from the external environment (Figure 1.4 b), thereby making the resultant structure thermodynamically metastable.[22] This type of

self-assembly goes against the law of increasing entropy that only applies to closed systems, and instead, it occurs in open systems to generate molecular structures using an incoming external energy supply.[23] Such structures obtained under non-equilibria are known as dissipative structures, showing distinctive characteristics compared with those of thermodynamic constructs generated via static self-assembly.[24] The molecular constituents of dissipative structures constantly undergo changes while preserving a defined overall architecture.

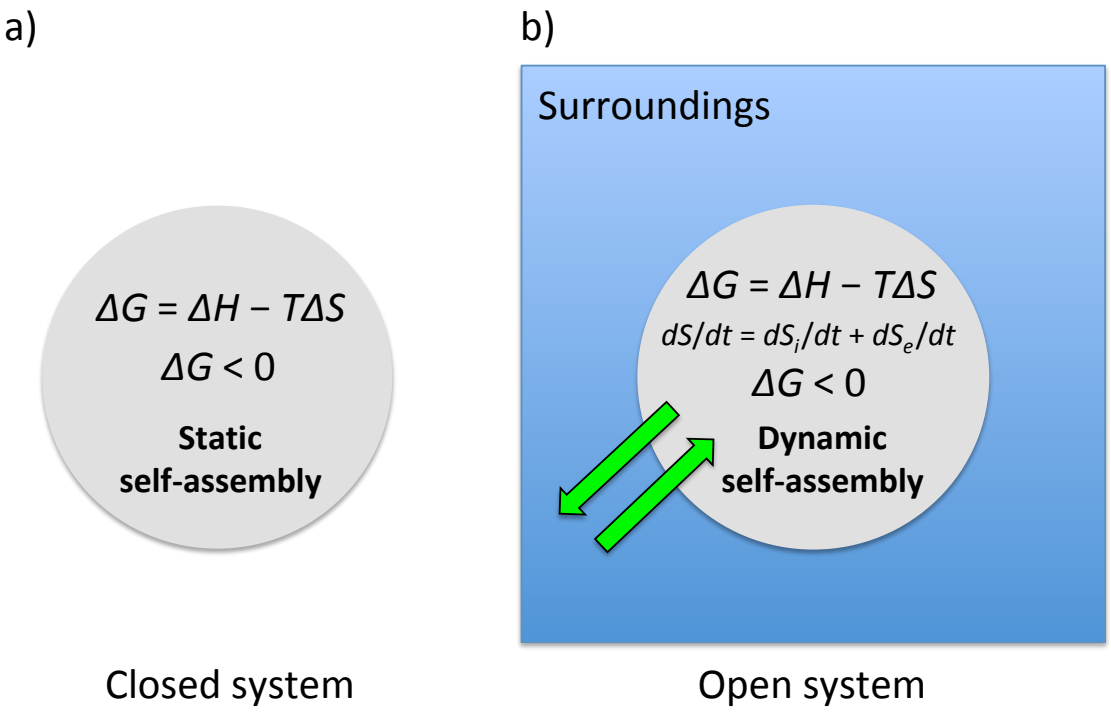


Figure 1.4. Schematic illustration showing **a)** static self-assembly that occurs in a closed system and **b)** dynamic self-assembly that occurs in an open system. **a)** Static self-assembly follows the equation $\Delta G = \Delta H - T\Delta S$ at constant temperature and pressure (where G is Gibbs free energy, H is enthalpy, T is temperature and S is entropy), and molecular self-assembly occurs spontaneously when $\Delta G < 0$. **b)** In dynamic self-assembly, entropy is exchanged with the surroundings, where the variation of entropy (dS) is the sum of the amount of internal entropy created and the amount of entropy exchanged with the surroundings (total entropy variation rate = $dS/dt = dS_i/dt + dS_e/dt$).[25] Dynamic self-assembly generates orders in the system by releasing entropy to the surroundings.

Dissipative structures undergo dynamic interactions with their surroundings by consuming and dissipating energy (Figure 1.5). Through this continuous energy exchange, the resulting structures may be imparted with greater and more sophisticated functionalities. This phenomenon can be explained by the theory of dissipative structures developed by Ilya Prigogine, outlining that thermodynamically open systems may lead to the formation of highly ordered structures with extraordinary properties.[26] Based on this theory, living systems can be seen as the ultimate forms of dissipative structures, given that life is maintained in a state of dynamic equilibrium where the simultaneous degradation and synthesis of components is made possible by relying on an external energy supply. Bringing the concept of dissipative structures to the field of supramolecular chemistry may help to unveil the origin of order and complexity seen in living systems.

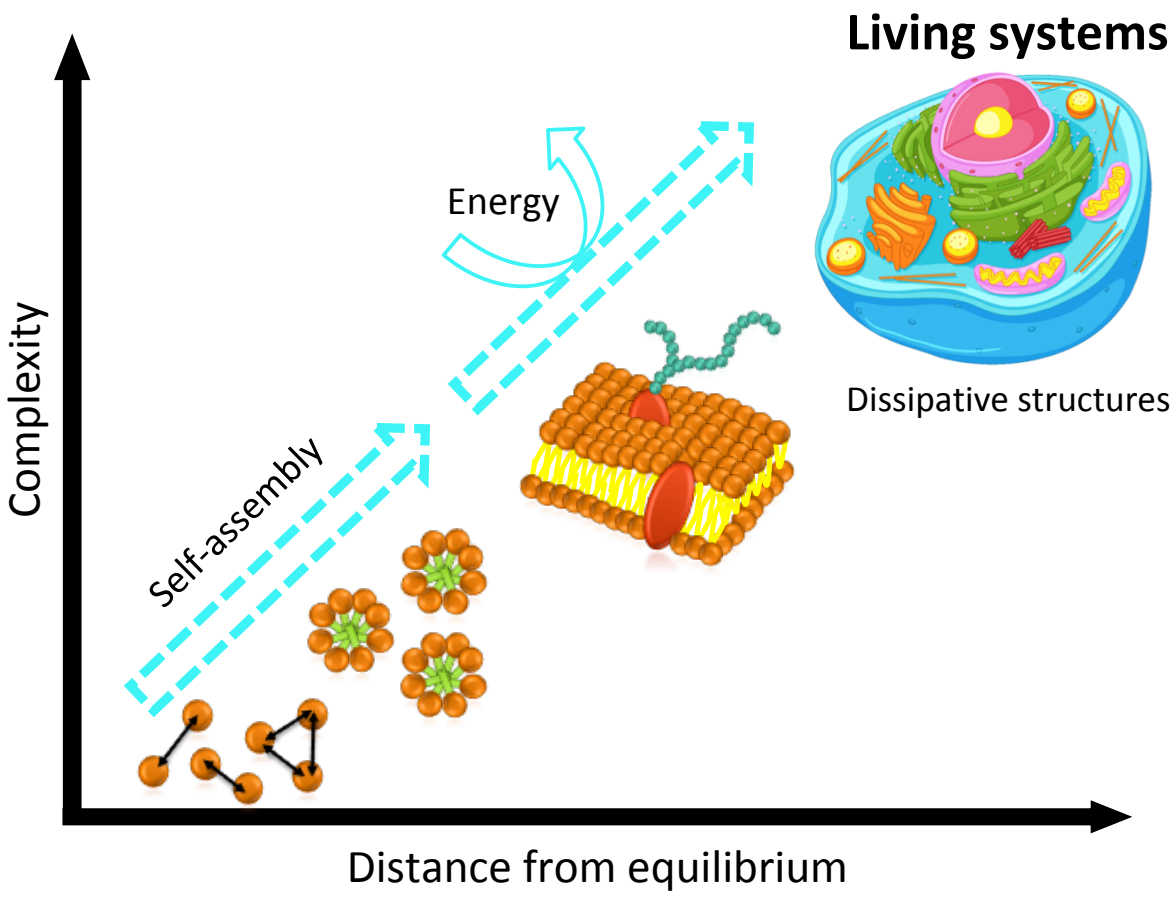


Figure 1.5. Dissipative structures are highly complex and distant from equilibrium, relying on an energy supply from the surroundings.

Although dynamic self-assembly can be seen as a means to create life-like, reconfigurable, and intelligent materials, it remains elusive how to impart such self-assembling processes in synthetic systems.[27] In order to approach these challenges, the chemical sciences are witnessing an increasing interest in the study of dynamic interactions resulting from numerous molecular components, rather than specific interactions in single components. Towards achieving a higher level of complexity, supramolecular chemistry has shifted its focus on understanding how novel and emergent properties arise from multi-component molecular systems, leading to the birth of systems chemistry.[28]

1.1.3 Systems Chemistry

The origin of systems chemistry as a scientific field dates back to the start of the 21st century. This field strives to comprehend the complex, dynamic and emergent properties of biological systems using synthetic chemical frameworks.[29] In contrast to conventional approaches that mainly focus on individual molecules, systems chemistry studies the multi-component interactions present within complex chemical systems (Figure 1.6).[30]

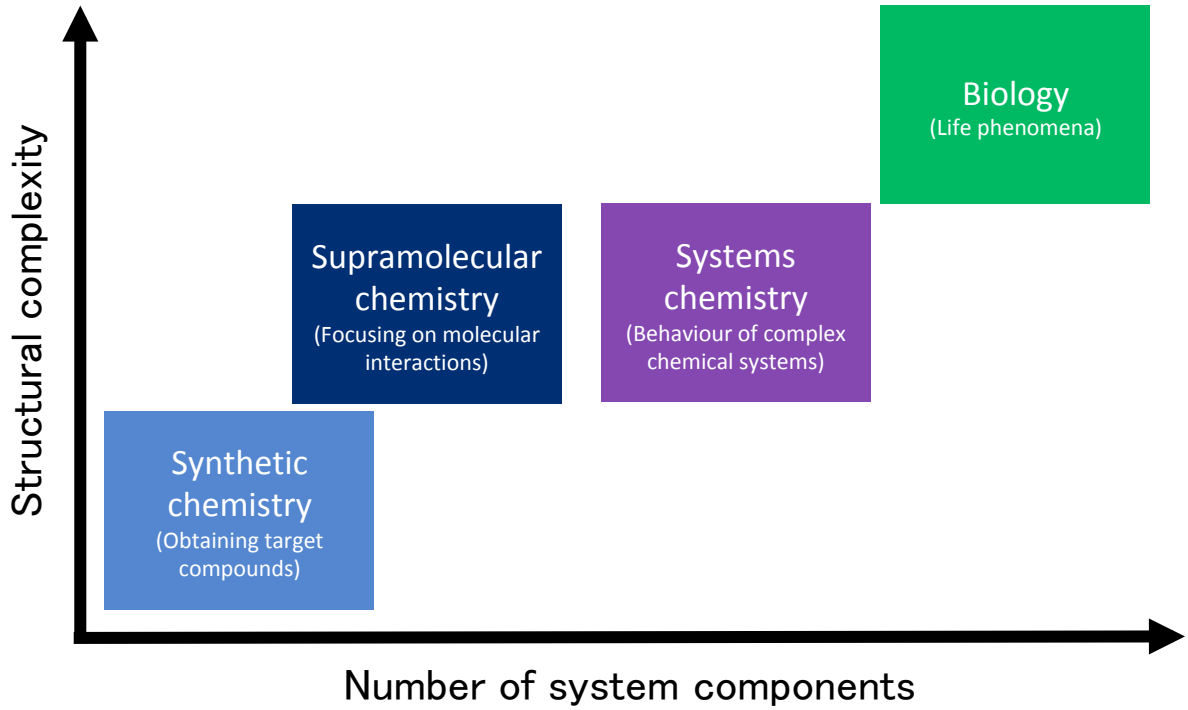


Figure 1.6. An overview of where systems chemistry, supramolecular chemistry, synthetic chemistry and biology sit among each other in terms of structural complexity and the number of system components.

By building on our current understanding of how different molecular components self-assemble to generate complexity and order, systems chemistry may facilitate the construction of highly biomimetic structures that more closely align with the characteristics seen in living organisms.

Systems chemistry addresses the void that exists between chemistry and biology by studying self-replicating molecules and the networks they establish.[31] Unlike conventional chemistry, which studies non-replicating systems, or biology which studies highly complex self-replicating systems, this emerging field focuses on how complexity in biological systems arises using simpler replicating systems in a synthetic environment.[32]

The fundamental principle in systems chemistry is the potential to generate emergent phenomena using simple synthetic molecules. Emergence refers to properties that are greater than the sum of the characteristics in the individual components within the system.[33] Obtaining emergent properties in synthetic systems is highly desirable as it can help to bridge the existing differences between chemistry and biology, but also facilitate the design and development of materials with enhanced characteristics.

Systems chemistry has tested various approaches towards the imitation of biological systems including dynamic self-assembly, chemically fuelled molecular motion, compartmentalised chemical networks and the autocatalytic self-replication of biological macromolecules.[34] These efforts have extended to the design of biomolecule-based dynamic chemical networks by exploiting the properties of nucleic acids, peptides, lipids and other biomolecules. By utilising a combination of synthetic and biological components, systems chemistry aims to generate materials with new system-level properties.[35]

1.1.4 Achieving Emergence through Hierarchical Self-assembly

Hierarchical self-assembly refers to the multi-step formation of complex and organised structures through various combinations and interactions of non-covalent bonds (Figure 1.7).[36] There are countless examples of hierarchical self-assembled structures in nature across all scales. Biological components such as proteins, polysaccharides, lipids and

oligonucleotides can direct structural formation of hierarchically organised complexes that generate life.[37] The hierarchal systems seen in biological organisms are dynamic and reconfigurable, being regulated by their individual constituents to great levels of precision. By learning from nature as a source of inspiration, the concept of hierarchical self-assembly can be utilised to design complex supramolecular architectures with enhanced levels of function.[38]

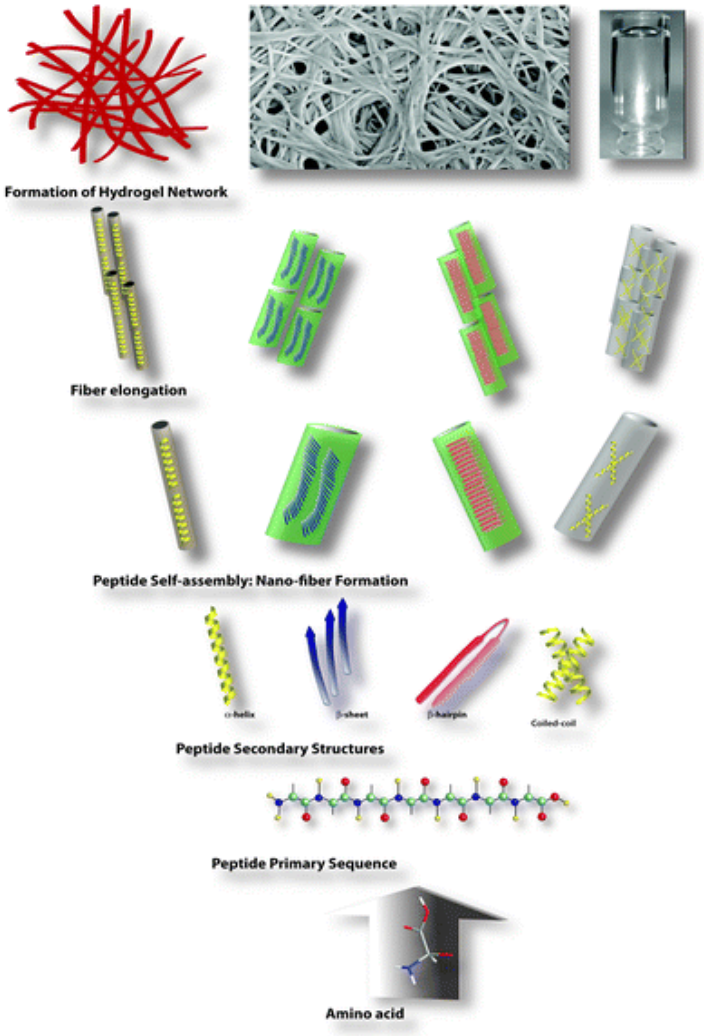


Figure 1.7. Simplified schematic illustrations showing the hierarchical self-assembly process leading to the formation of complex hydrogel networks from simple amino acid molecules. Reproduced with permission from ref 39.

Hierarchical self-assembly requires well-defined molecular building blocks that can continuously self-assemble through additional non-covalent interactions, eventually leading

to the formation of highly-organised structures.[40] The complexity seen in nature is often achieved through the non-covalent arrangement of simple components into nanoscale fibres that subsequently align, twist or bundle to form more sophisticated biological complexes.[41] A famous example is collagen formation, which involves three polypeptide chains folding into triple-strand helices that further self-assemble into fibres (Figure 1.8).[42] The hierarchical structure of collagen allows it to play a pivotal role in animals, providing mechanical stability, elasticity and strength to connective tissues. Understanding how simple molecules arrange in nature to form such complex and highly functional structures may pave the way forward for designing a new class of biomimetic materials such as smart contractile materials, molecular sensors and nano-actuators with optimised responses.[43]

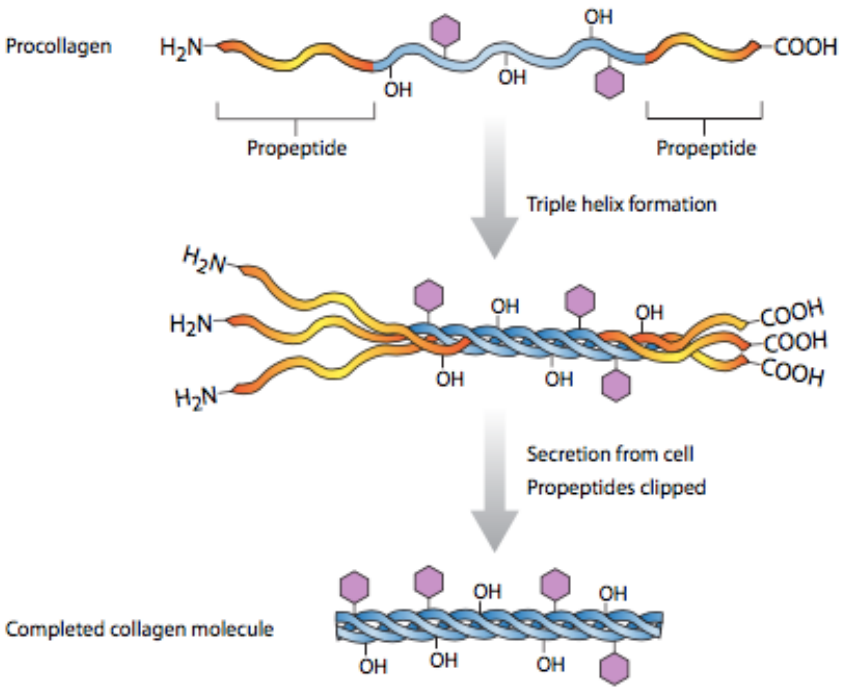


Figure 1.8. The formation of collagen involves the folding of three polypeptide chains into triple-strand helices. Reproduced with permission from ref 44.

The formation of complex structures through hierarchical self-assembly relies on a multi-step process, whereby each new level formed within the hierarchy exhibits greater complexity than previous steps.[45] Each level in the hierarchy displays a unique array of non-covalent interactions that partly determine the outline of subsequent layers.[46]

Ultimately, the interplay of the various building blocks and their non-covalent interactions enables the formation of a well-defined final construct capable of displaying emergent properties that were absent in previous levels within the hierarchy.[47] Consequently, to understand how emergent characteristics arise in biological systems, it is essential to dissect and characterise the different interactions that lead to the formation of the final product.

In addition to the complexity demonstrated by biological structures, an even more outstanding feature is their dynamic nature that allows them to undergo change in response to both external and internal stimuli.[43] For example, organelles constantly undergo unique structural changes in response to chemical and mechanical signals that enable them to perform highly specific functions within cells. Being able to replicate the dynamic properties seen in living systems has great potential for designing artificial systems with enhanced functionalities.

Designing complex hierarchical constructs is currently achievable; but obtaining a level of control similar to that seen in living systems remains a crucial challenge. DNA-mediated assembly has emerged as one of the approaches to attain dynamic and reversible control over synthetic hierarchical materials.[48] However, most of the research in DNA-mediated assembly has been confined to inorganic nanoparticles and pure nucleic acid materials.[49] It has only been over the last decade that the first examples of DNA-functionalised supramolecular polymers were reported.[50,51] Bioconjugation is a promising avenue for achieving complexity and emergent properties as those seen in nature by utilising the unique building blocks present in biological systems.

1.2 Bioconjugation

1.2.1 Overview

Bioconjugation has emerged as a valuable tool to integrate the unique characteristics of biomolecules within a unified system (Figure 1.9).[52] The goal of this field is to generate hybrid compounds that not only integrate the qualities of different biomolecules, but also those that surpass them.[53] Covalently linking separate biomolecules may induce structural changes in the resulting system that could potentially bestow it with novel functions that were previously absent. Bioconjugation is a unique approach that merges the fields of biology and chemistry by exploring how the manipulation of key biological constituents may lead to emergent properties.

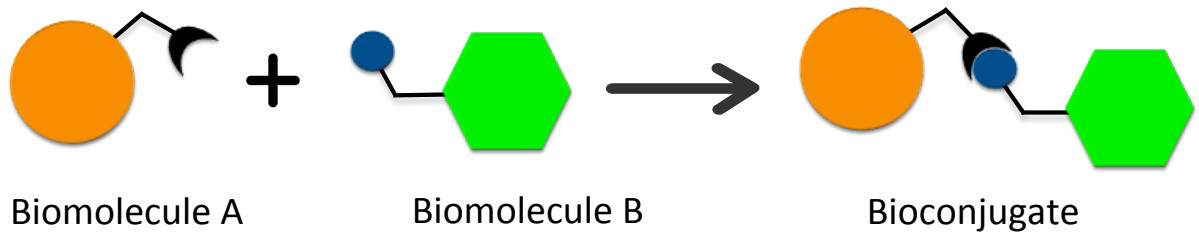


Figure 1.9. Schematic illustration of bioconjugation, showing the process in which two different biomolecules are conjugated through covalent bonding.

Living organisms highlight how chemical systems give rise to diverse properties and enormous complexity through the associations of different components. For instance, cellular activities are maintained through the synergies of multiple biological molecules including proteins, nucleic acids and lipids. The multi-component self-assembly of biological molecules generates coordinated non-covalent interactions, providing adaptability, dynamics and responsiveness.[54] While living systems generate emergent properties from the interaction of their biological components, achieving this in synthetic systems remains a great challenge. Mixing different biomolecules *in vitro* does not result in multi-component self-assembly, but rather leads to segregation from each other and self-assembly of individual molecules. Bioconjugation strives to attain a new level of complexity through the manipulation of biomolecules that closely mimics what is seen in nature.[55]

Bioconjugation explores how emergent properties can be achieved through the design of multi-component materials using biomolecules. The generation of bioconjugates can integrate the advantageous characteristics of the individual components, as well as introducing new characteristics that were previously absent. Examples of emergent properties seen in nature include the enhanced toughness of silk resulting from the interactions between fibroin and sericin, the dynamic structures generated by actin and myosin in muscle tissue, or the precise motion in microtubules resulting from the interaction between kinesin and tubulin.[56] Achieving multi-component self-assembly through bioconjugation has enormous potential for generating next-generation biomaterials with enhanced properties that could transform different scientific disciplines, including medicine and material engineering.[57,58]

1.2.2 Current Progress and Challenges

A wide variety of bioconjugate complexes have already been produced, ranging from highly biomimetic materials to those that cannot be obtained through biological synthesis. As a result, bioconjugation has made it possible to develop various hybrid materials with different applications, including antibody-drug conjugates for therapeutic applications,[59] hybrid conjugate systems capable of interacting with cells,[60] and bioconjugates incorporating semiconducting polymer dots for cell imaging.[61] Most of the successes in bioconjugation have relied on merging the useful characteristics of different biomolecules into a final hybrid compound, but it remains challenging to introduce desired emergent properties to the resultant system.[62] This field aims to impart synthesised bioconjugates with characteristics resembling those in living systems, such as bioselectivity, biocompatibility and biostability.[17]

A challenge with bioconjugation is that synthetic bioconjugates may not retain the desirable features of selected biomolecules previously present in their native environment.[63] This is because the covalent bonding of the individual components can affect their properties, meaning that they may not be preserved in the formed hybrid.[64] Designing bioconjugates of interest that retain the useful features of their constituents requires extensive knowledge of

the structural, chemical and biological characteristics of each manipulated biomolecule. In addition, maximising the potential activity and function of the final bioconjugate necessitates a careful assessment of the functional groups present on the biomolecule, as well as the chemical reagents employed during conjugation reactions.

1.2.3 Approaches to Bioconjugation

In order to design bioconjugates with specific properties, a wide variety of bioconjugation techniques have been developed. In the beginning, techniques focused on functionalising proteins with small molecules, eventually extending to the conjugation of multiple biomolecules. The methods that covalently link biomolecules can be grouped into three main categories, depending on whether the conjugation reactions are chemical, enzyme-mediated or photo-induced. Although these techniques can be considered as chemical in nature due to the chemical linkages, they can be more specifically distinguished based on the presence of biological or physical stimuli during reactions.[52]

Chemical conjugation is the most widely used strategy for generating bioconjugates given that the reactions involve relatively simple techniques and equipment, with examples including amide-coupling, maleimide-thiol addition and cycloaddition.[65] Meanwhile, enzyme-mediated conjugation is generally only used for the modification of oligosaccharides, whereas the use of photo-mediated approaches is limited by the risk that ultraviolet light poses to oligonucleotides and cells during reactions.[52] Consequently, the use of chemical conjugation is deemed preferable when using oligonucleotides given the versatility of this approach compared to the limitations of the two other methods.

Achieving successful bioconjugation is difficult due to the chemical incompatibility that exists between different biomolecules.[66] Moreover, there is no established universal standardised conjugation method that guarantees success, meaning that suitability of possible techniques needs to be selected on an individual basis, and will depend on the type of biomolecules used.[67] Key elements that need to be considered for bioconjugation include rate of reaction, site-selectivity, orthogonality (where chemical reactions occur between specific functional groups without affecting other functional groups present on biological

components) and reaction partner accessibility.[66] Achieving site selectivity is particularly challenging owing to the limited diversity of chemical functional groups in native biomolecules. This makes it difficult to react biomolecules at the desired position and to obtain the target conjugate. To overcome this issue, functional groups are commonly modified to ensure that biomolecules will react at a predetermined position. This modification step is fundamental for conjugation to succeed because it not only enables biomolecules to react selectively, but also helps to avoid non-specific functional groups generating unwanted hybrid compounds.

The first thing to be considered in bioconjugation is the type of biomolecule that will be conjugated and the desired outcome in the final bioconjugate.[68] This is because there are various types of biomolecules with highly distinctive properties. Additionally, the intrinsic properties of individual biomolecules may negatively affect each other, ruling out the desired features in the final conjugate. Therefore, it is important to carefully select biomolecules that can leverage their unique characteristics following covalent linkage to achieve greater properties in the final product than in the individual constructs.

1.3 Peptide-Oligonucleotide Conjugation

1.3.1 Overview

Peptides and oligonucleotides are highly attractive biomolecules due to their unique chemical properties and the diverse functions they mediate in living systems. Oligonucleotides are the constituents of the universal code of life present in all living organisms, including DNA and RNA. The structure of oligonucleotides is dictated by Watson-Crick base paring, which provides a highly predictable scaffold for the design of complex nanostructures.[69] Peptides are also interesting biomolecules as they form the constituents of the numerous enzymes and proteins that shape life. In contrast to oligonucleotides, peptides are able to generate an enormous diversity of proteins through their various self-assembly motifs.[70]

The chemical diversity and unique biological roles of peptides and oligonucleotides within

living systems make the conjugation of these biomolecules an exciting prospect. Given that peptides and oligonucleotides coexist together in living systems to make life possible (e.g. nucleic acids, chromatin, ribosomes and transcription factors), conjugating these biomolecules may guide the design of synthetic systems with similar capabilities as those seen in nature.[71] In addition, the development of peptide-oligonucleotide conjugates (POCs) may allow us to understand the behaviour of these biomolecules further, potentially shedding clues of how life may have originated.

1.3.2 Peptides

Peptides are short chains of amino acids linked through amide bonds between the carboxyl group (-COOH) of one amino acid and the amino group (-NH_2) of another one (Figure 1.10).[72] Both the carboxyl and amino groups are conserved among all amino acids, whereas the variable side chain provides variety, which differentiates amino acids from one another. Generally, peptides are referred to as molecules comprising 2-50 amino acid resides, whereas those with more than 50 amino acids are known as proteins.[72]

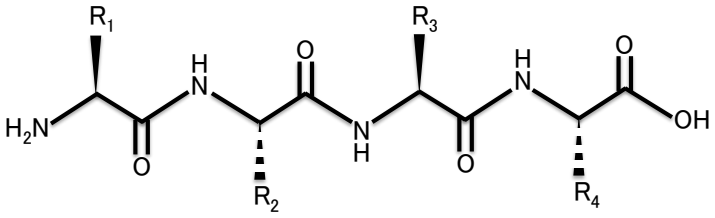


Figure 1.10. Chemical structure of an amino acid (R represents a side chain).

Peptides are attractive building blocks due to their intrinsic ability to self-assemble into a diverse range of structures depending on their amino acid sequences.[73] The distinctive side chain of amino acids is responsible for imparting the final peptide or protein with unique physicochemical and biochemical properties.[74] The different combination of amino acids induces changes in their higher-order structures, resulting in enormous variety in the resultant peptide. Consequently, by alternating amino acid sequences, it is possible to generate numerous peptides possessing highly distinctive structures and functions.

The potential of peptides to self-assemble emanates from the occurrence of various non-covalent interactions, including electrostatic interactions, hydrogen bonding, π-π stacking and hydrophobic effects. These forces are pivotal in facilitating the interactions between peptides in solution that ultimately lead to the formation of nanostructures.[75] Additionally, different amino acid types rely on unique non-covalent interactions when they self-assemble into peptides. For example, aromatic amino acids and non-polar amino acids self-assemble through hydrophobic interactions and/or π-π stacking, while the self-assembly of polar amino acids occurs predominantly through hydrogen bonding or electrostatic interactions.[73]

Peptide chains spontaneously self-assemble into secondary structures through the formation of hydrogen bonds between carbonyl and amino groups in the peptide backbone.[76] Two of the most commonly known secondary structures are α-helices and β-sheets.[77] In α-helices, the carbonyl group of one amino acid forms hydrogen bonds with the amino group of an amino acid four positions down the chain, leading to a coiled configuration in the resulting peptide. In this type of configuration, the side chains are projected outwards and are consequently free to interact with their surrounding environment. Certain amino acids such as alanine, glutamic acid, leucine and histidine favour the formation of α-helices when they self-assemble.[78]

In contrast, β-sheets consist of peptide chains connected by interstrand hydrogen bonding that form extended amphipathic sheets in which hydrophilic and hydrophobic side-chains face in opposite directions. β-sheets are further subdivided into parallel and anti-parallel forms, depending on the direction between the strands.[77] Parallel β-sheets are characterised by two peptide strands running consistently in the same direction from the N- to C-terminus (Figure 1.11 a).

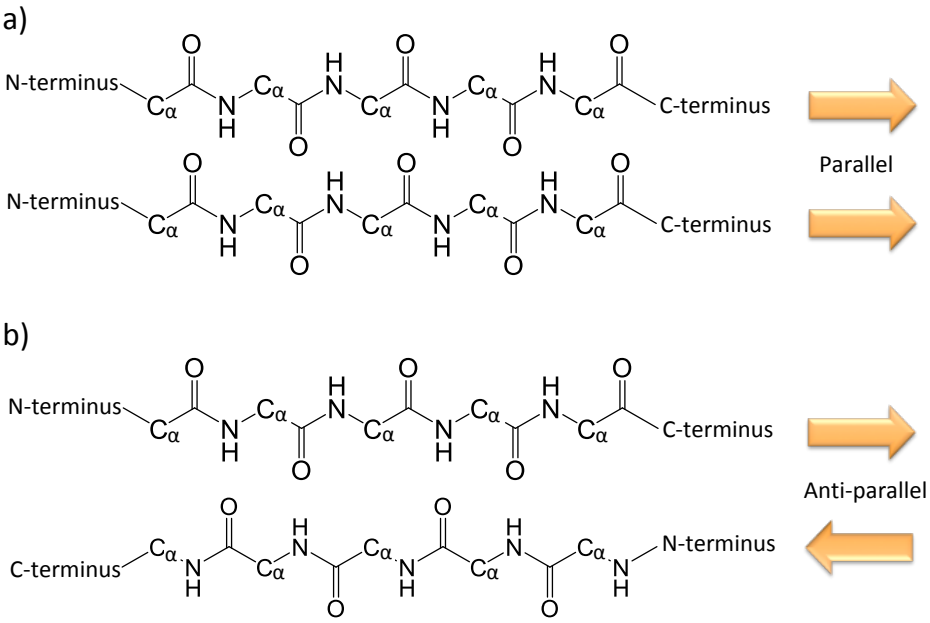


Figure 1.11. a) Parallel and **b)** anti-parallel β-sheet peptide secondary structure.

Whereas, anti-parallel β-sheets consist of strands running in opposite directions, meaning that the direction of one strand runs from the N- to C-terminus, whilst the next one follows the reverse order (Figure 1.11 b). In β-sheets, the amino acid side chains always protrude above or below the plane of the strand.[79]

The formation of β-sheets in an aqueous environment is principally mediated by the formation of hydrogen bonds of the amide groups along the peptide backbones and ionic interactions between the side chains of positively and negatively charged residues.[80] In addition, hydrophobic interactions that result from amino acids displaced by water molecules, and water-mediated hydrogen bond formation also play an important role in driving peptide β-sheet formation.[81]

In addition to intermolecular forces, environmental factors such as changes in temperature and pH can affect the morphology of the resultant peptide structure.[75] Manipulating these environmental factors can affect the interacting forces within peptides, which can affect the self-assembly process by inducing structural organisation changes. By understanding how environmental factors induce specific chemical and physical changes, it may be possible to engineer novel peptide-based materials with high precision.

Peptides have the advantage of being relatively short and easily obtainable via solid-phase synthesis, which enables their incorporation into other nanomaterials.[82] Moreover, peptides possess multiple structural motifs and self-assembly modes that generate a rich variety of mechanical and chemical properties.[83] Despite their useful characteristics, it remains challenging to precisely predict and control the resulting structures of peptides due to their rich self-assembly behaviour.[84] To overcome this limitation and achieve greater control over the resulting structure, peptides can be conjugated with molecules such as oligonucleotides that possess highly predictable modes of self-assembly.

1.3.3 Oligonucleotides

Oligonucleotides are short-chain polymers consisting of 8-50 nucleotides.[85] Nucleotides are the basic units of DNA and RNA, being made up of three different components: a nitrogenous base, a 5-carbon sugar and a phosphate group (Figure 1.12).[86] The four different nitrogenous bases in DNA are adenine (A), guanine (G) cytosine (C) and thymine (T), whereas uracil (U) is found in RNA in place of thymine.[86] Nucleotides not only constitute the genetic material in living organisms, but they also form part of other molecules that make life possible, including adenosine triphosphate (ATP) and coenzymes such as NAD and NADP.[87]

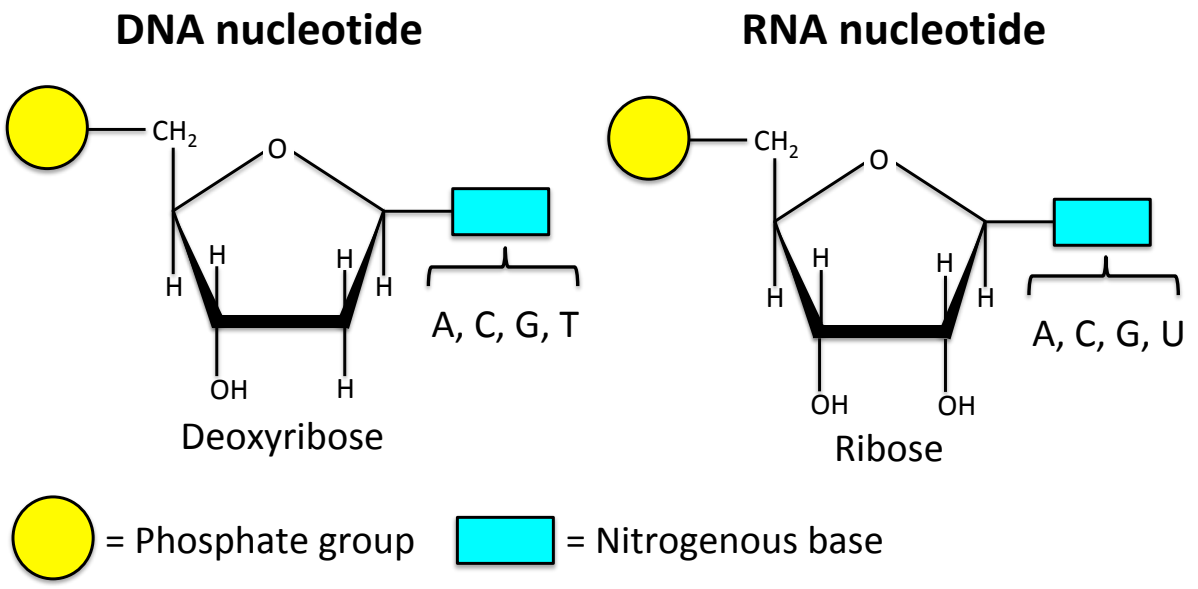


Figure 1.12. Chemical structure of a DNA and an RNA nucleotide.

Nitrogenous bases are either categorised into purines (i.e. adenine and guanine) or pyrimidines (i.e. cytosine, thymine and uracil) based on the number of rings they have. Purines are larger than pyrimidines and consist of double-ringed structures, whereas pyrimidines possess a single ring.[86] Purines form base pairs with pyrimidines in a highly specific manner through hydrogen bonding; with adenine always paring up with thymine or uracil, and guanine being specific to cytosine.[86] Guanine links with cytosine through three hydrogen bonds, making this linkage stronger than that between adenine and thymine/uracil, which forms two hydrogen bonds.[86]

The intermediate part of nucleotides consists of a 5-carbon sugar, which remains the same across the different nucleotides.[88] However, the 5-carbon sugar in DNA is deoxyribose, whereas in RNA the sugar is ribose, and hence the differences in the way these genetic molecules are named.[88] Both sugars are almost identical with the exception that deoxyribose lacks a hydroxyl group (-OH) at its 2' carbon.[88] The last part of nucleotides is the phosphate group, which can form phosphodiester bonds between sugar molecules of adjacent nucleotides.[88] Phosphodiester bonds are covalent linkages that occur between the 3' carbon of one sugar and the 5' of another, being central to life on earth as they form the backbone of the strands that carry the genetic material.[88]

While their structure is simple, nucleotides can form complex structures through the hydrogen bonding that occurs between complementary base pairs. A well-known example of the complexity mediated by nucleotide base pairing is the double-helical structure of DNA that comprises two nucleotide chains twisted around each other.[88] The arrangement of nucleotides is directed by specific hydrogen bonding patterns, known as Watson-Crick base pairing. This highly defined bonding occurring between bases allows for a precise manipulation of oligonucleotides.

The process whereby two single strands of DNA are joined together to form double-stranded DNA (dsDNA) is referred to as DNA hybridisation, being commonly utilised as a tool to create desired nanostructures.[89] DNA hybridisation is a reversible process since

it relies on the formation of hydrogen bonds between complementary base pairs. Heating can be used to induce the breakage of hydrogen bonds in dsDNA to subsequently release the individual strands in a process known as DNA denaturation. Conversely, the hydrogen bonds can be reformed when the DNA is allowed to cool down, reverting back to dsDNA.

Owing to the relatively simple and predictable nature of DNA hybridisation, this process has been widely recognised as a powerful means for the construction of nanostructures. DNA hybridisation techniques have been deployed for the production of an ever-increasing variety of nanostructures, ranging from simple 1D arrays to highly complex and anisotropic 2D and 3D structures including DNA origami.[69] DNA technology has been exploited for bottom-up DNA self-assembly, paving the way for programming hierarchical architectures. However, research based on DNA technology has been primarily confined to inorganic nanoparticles and pure nucleic acid materials.[49] The exploration of DNA-mediated assembly combined with different organic materials has only emerged over the last decade, yet it holds enormous potential in developing hierarchal complexes with an additional level of control.[48]

The attractive aspect of oligonucleotides for bioconjugation is their highly predictable Watson-Crick base paring, which provide highly programmable scaffold and bioinstructive properties for designing complex nanostructures, sequence-specific ligand immobilisation and reversible crosslinking materials through strand-displacement reactions.[69] Particularly, the programmability of DNA hybridisation has received increased interest as an important component to enhance molecular control in hybrid molecular systems.[60] The relatively simple structure of oligonucleotides imparts them with high programmability but this also limits their chemical and functional diversity. To overcome this drawback and build on their advantageous characteristics, peptide-oligonucleotide conjugation can be utilised as a means to merge the structural programmability of oligonucleotides with the chemical and structural diversity generated by peptides.

1.3.4 Potential Applications of Peptide-Oligonucleotide Conjugates (POCs)

The novel and growing field of POCs embeds enormous potential for developing highly biomimetic nanomaterials, including those materials that cannot be obtained through biological synthesis.[53] Current research is focusing on how the conjugation of peptides with oligonucleotides in a laboratory environment may lead to emergent properties not seen in the individual biomolecules; with the long-term aim being the development of biomaterials with therapeutic (e.g. drug delivery[90] and tissue engineering[52]) as well as functional applications (e.g. hybrid self-assembling systems[91], controlling peptide self-assembly by DNA and DNA-templated peptide ligation[92]).

However, the generation of functional biomaterials from oligonucleotides or from peptides has shortcomings. For example, oligonucleotides lack chemical and functional diversity compared to peptides due to specificity of the complementary base paring rules. On the other hand, the rich self-assembly behaviour of peptides complicates the rational design and control of their resultant structures. Current efforts in bioconjugation are focusing on integrating the programmability of oligonucleotides with the self-assembling properties of peptides to derive conjugates that possess the advantageous characteristics of both biomolecules.

The first example of self-assembling POC was reported in 2012 by the Vebert-Nardin group.[93] They generated a conjugate consisting of a 12-nt CT-rich oligonucleotide and a diphenylalanine (FF) peptide that self-assembled into a spherical structure comprising a POC bilayer surrounding a hollow core (Figure 1.13). Fluorescence microscopy revealed that the spherical structure had the capability of encapsulating and releasing a dye in response to pH changes. According to the authors, the release of the dye was attributed to protonation of the cytosine-rich oligonucleotides and disruption of the membrane. Although further assessment on its biocompatibility is required, their POC structure may potentially act as a delivery cargo capable of transporting small molecules following cellular endocytosis.

Furthermore, the authors investigated the self-assembling mechanism of the spherical

structure and hypothesised that hydrogen-bonding (but not base-paring) played a crucial role between the oligonucleotides. To examine the effect of hydrogen bonding, they co-incubated the conjugates and dipeptide FF fibres with urea and compared the effects in their morphology. While there was no significant change in dipeptide fibres, the conjugates underwent a drastic shift to an elongated morphology transitioning between the spheres and fibres. This work highlighted that the conjugated oligonucleotides displayed self-assembling capacity beyond the characteristic Watson-Crick base pairing. Importantly, this shows that bioconjugation may lead to the appearance of novel properties that would not normally be seen in the individual biomolecules.

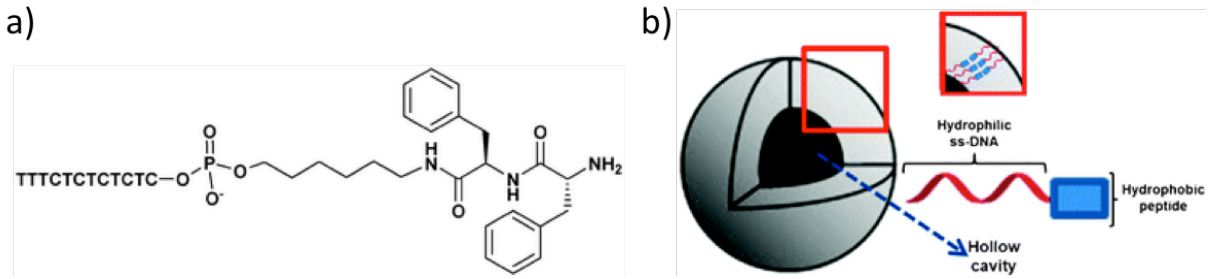


Figure 1.13. a) Structure of POC generated by the Vebert-Nardin group and **b)** a self-assembled spherical structure with a hollow core consisting of a POC bilayer. Adapted and reproduced with permission from ref 93.

The self-assembly of POCs can potentially provide synergistic properties of peptides and oligonucleotides under certain conditions, which could ultimately lead to improved properties in the resulting conjugate. However, simply mixing POCs in solution results in the formation of heterogeneous particles with little to no added value.[94] The current objective in this field is to develop an understanding as to how multiple self-assembly modes can be precisely controlled within a single system in order to design POCs with desired properties.

A recent study by Kye and Lim showed that it was possible to form well-defined homogenous nanostructures using a dual pathway approach, but only when the orthogonal self-assembly modes of POCs (i.e. β-sheet formation and Watson-Crick base paring) were controlled (Figure 1.14).[95] The first pathway involved the formation of β-sheets, followed by DNA hybridisation and subsequent cooling of the solution to anneal the duplex. The second

approach involved DNA hybridisation below the critical aggregation concentration of the peptide, followed by concentrating the solution to induce peptide assembly.

This study revealed that both pathways were able to generate identical toroidal POC nanostructures with homogeneous distributions. In contrast, heterogeneous POC nanostructures were formed when the solution was mixed without controlling their self-assembly modes. This research highlighted that it was possible to produce well-defined hybrid nanostructures by precisely controlling the orthogonal self-assembly of POCs.

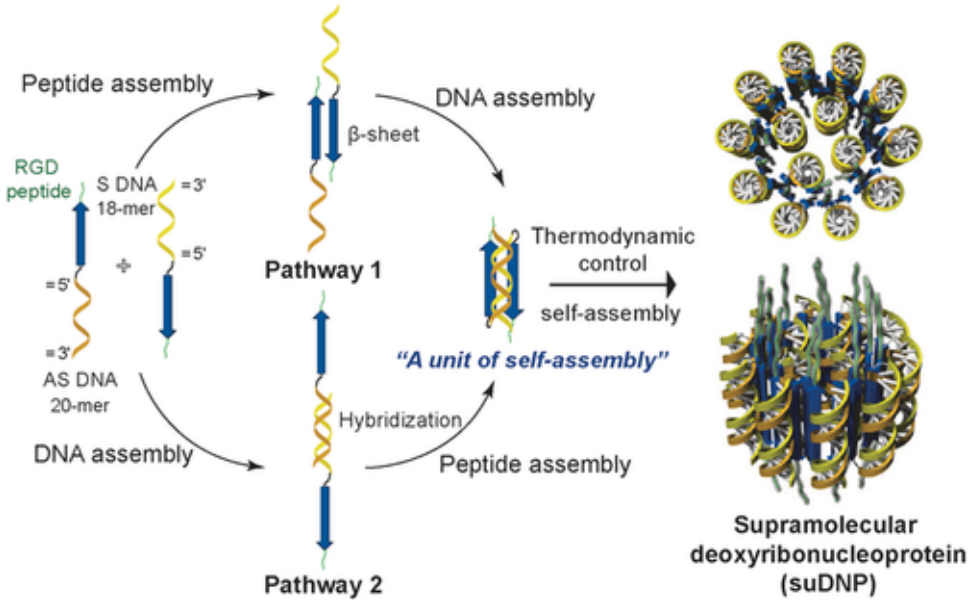


Figure 1.14. Pathways for the dual-mode self-assembly of complementary POCs into toroidal structures. Adapted and reproduced with permission from ref 95.

For the first time in 2018, the Luijten and Stupp groups demonstrated reversible control in cross-linking peptide hydrogels using DNA (Figure 1.15).[96] To obtain these hydrogels, they prepared nanofibres through the co-assembly of alkylated peptides with a similar monomer containing a covalently linked oligonucleotide terminal segment. These nanofibres were subsequently mixed with fibres bearing complementary DNA sequences, which resulted in cross-linking hydrogels. Notably, the resultant POC hydrogels contained thick micrometre-sized bundles enriched in the DNA-modified molecules surrounded by DNA-poor individual fibres, rather than those containing fibres randomly cross-linked.

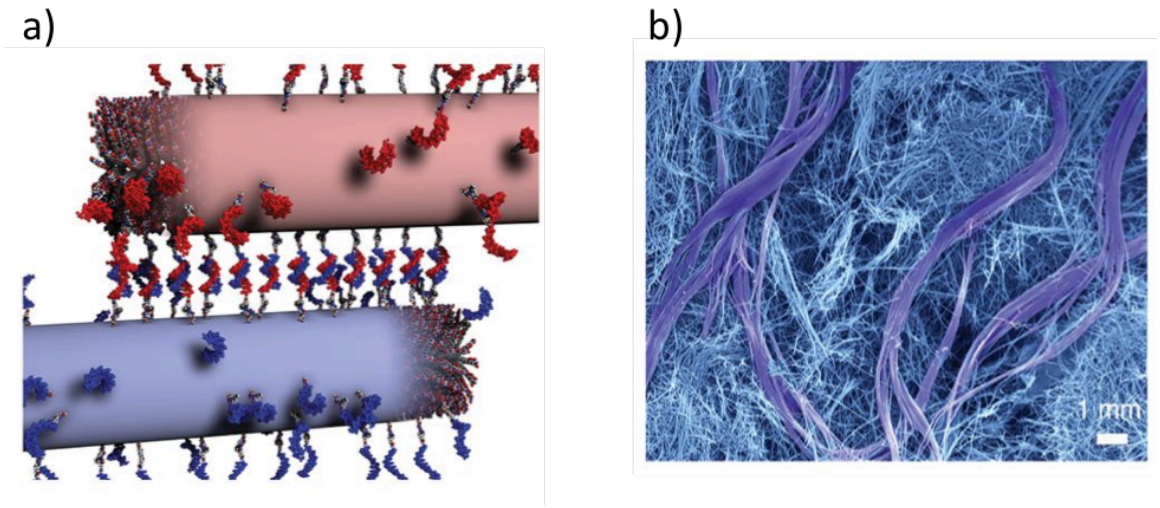


Figure 1.15. a) Illustration of peptide amphiphilic fibres cross-linked by DNA hybridisation and **b)** SEM image of the hydrogel formed upon DNA cross-linking. Adapted and reproduced with permission from ref 96.

According to the authors, the bundle formation could have been due to the redistribution of DNA-modified molecules between fibres, followed by hybridisation with complementary molecules. Additionally, the DNA-rich bundles were reversible over multiple cycles through strand displacement and by cycling the temperature. The formation of the hydrogels demonstrated hierarchical organisation involving multiple levels of molecular interactions: from peptide-filament formation via hydrogen bonding and hydrophobic interactions, followed by intertwining of fibres through DNA hybridisation.

In this study, the reversible control of such hierarchical self-assembly was comparable to that seen in proteins that shape up the extracellular matrix (ECM), such as collagen.[97] In addition, the hydrogels displayed similar structural properties as proteins found in living organisms, in which the constant remodeling of bundled fibres controls their organisation and stiffness. This work has potential implications for manipulating and precisely guiding cell behaviour, which may lead to advancements in biotechnology.

Bioconjugation also holds great potential in overcoming some of the key limitations encountered with the use of therapeutic oligonucleotides in gene therapy, which could lead to further advancements in this medical field. Antisense oligonucleotide therapy involves

the use of highly specific oligonucleotide sequences that block and induce the destruction of altered sections of mRNA, ultimately preventing the production of faulty proteins implicated in human disease.[98]

Antisense therapy can also be used to direct oligonucleotides against aberrant sections of immature mRNA known as introns, to promote the production of fully functional proteins in individuals who would otherwise produce non-functional proteins.[99] The rise of therapeutic oligonucleotides is regarded as a major advancement in the field of pharmacology due to their enormous potential at treating certain gene-related conditions, neurological disorders, cancers and infectious diseases with a high degree of specificity.[100-102]

Despite their potential in gene therapy, the chemical properties of therapeutic oligonucleotides can limit their full potential. Their hydrophilicity, multiple anion charges and ready interaction with degrading enzymes cause oligonucleotides to suffer from poor stability and rapid degradation.[103] Additionally, these chemical properties also interfere with the cellular uptake of oligonucleotides and lead to poor target specificity.[104] The combination of these factors negatively affects the pharmacokinetic profile of oligonucleotides and limits their capacity of successfully binding to their cellular targets. To overcome some of these barriers, bioconjugation has been proposed as a solution to integrate the advantageous properties of peptides into the final oligonucleotide hybrid. This can lead to improved cellular permeability and enhanced resistance to enzymatic degradation of oligonucleotides, ultimately improving their capacity to achieve long lasting target-specific effects.[105]

There are some research examples showing that oligonucleotides are efficiently taken up by cells when combined with suitable peptides either in the form of non-covalent complexes or covalent conjugates.[106] Different types of peptides have been used to enhance the cellular uptake of oligonucleotides, including cell-penetrating peptides (CPPs), arginine rich MTS peptides, and Tat and histidine-rich peptides.[104] However, cell toxicity associated with some of these peptides is a key limitation that needs to be overcome.[106]

1.3.5 Conjugation Strategies for POCs

While POCs have great potential in many fields of science, their synthesis can be challenging given that peptides and oligonucleotides are often structurally and chemically incompatible with each other.[107] Consequently, a great deal of research has taken place over the last decades to develop conjugation strategies that can facilitate the synthesis of POCs.[107] Currently, conjugation strategies can be grouped into two main categories: either involving on-resin synthesis of the entire conjugate, or those involving the synthesis and purification of each component separately prior to conducting conjugation.[53]

For on-resin synthesis, peptides and oligonucleotides are synthesised on solid support using specific protected monomers.[109] This strategy constructs the POC in a linear approach without having to synthesise and then merge peptides and oligonucleotides separately. However, a key limitation with on-resin synthesis is that the reaction conditions for one biomolecule are often incompatible with the other, which may structurally damage the final POC.[53] For example, the cleavage of the peptide fraction of the final POC when using Fmoc-peptide synthesis requires strong acidic conditions, which may damage the oligonucleotide segment.[110]

To avoid these issues, it is more common to use an alternative strategy in which peptides and oligonucleotides are synthesised separately on solid phase followed by linking them directly, or through a bifunctional linker (Figure 1.16).[53] After coupling, the conjugate can be purified using common methods such as preparative HPLC, anion exchange, elution from a denaturing polyacrylamide gel or ethanol precipitation.[53]

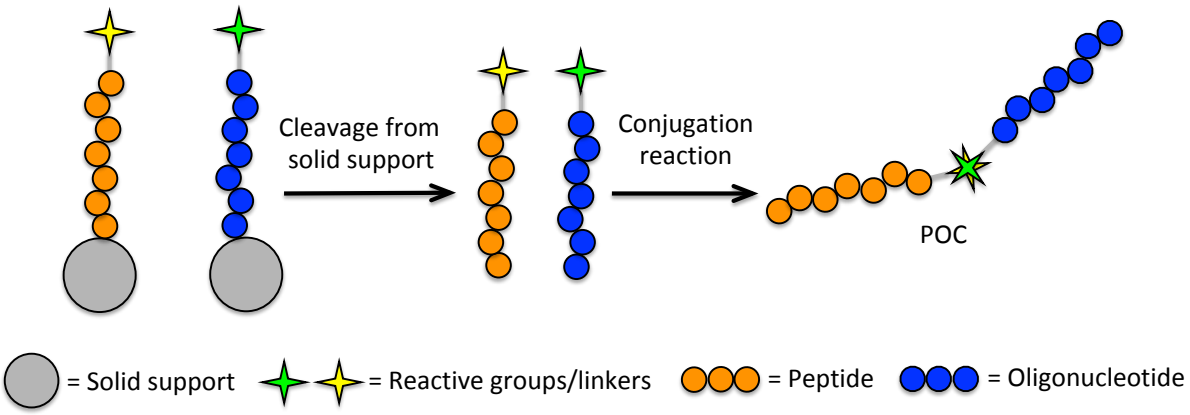


Figure 1.16. Peptide-oligonucleotide conjugation following synthesis of individual components and their subsequent cleavage from a solid support.

Given that oligonucleotides lack suitable nucleophiles or electrophiles that can react with peptides, functionalisation is required prior to conjugation. The common reactive groups introduced to oligonucleotides include amines, thiols, azides/alkynes, alkoyamines or hydrazines and aldehydes. For peptides, some reactive groups contained in natural amino acids can be used for conjugation: amines are available in lysine, carboxylic acids in glutamic acid and aspartic acid or thiols in cysteine. The N-terminus or C-terminus of the peptide can also be used for conjugating with oligonucleotides using an activated linker.[111]

1.4 Project Aims

The principal aim of this project is to generate a variety of POCs by conjugating different peptides with DNA oligonucleotides, with the secondary aim exploring how the different self-assembly properties of the synthesised POCs affected their resultant structures.

The first objective of this research focuses on synthesising the peptides that will later be used for conjugation experiments, as well as assessing their predicted self-assembling behaviour. These peptides include three β-sheet forming peptides (**KLVFFA**, **HYFNIF** and **RVFNIM**) and a collagen mimetic peptide **$(POG)_6$**. The second objective involves finding a suitable conjugation strategy for synthesising different POCs. Lastly, the project explores the self-assembly of the POCs, specifically investigating how the final structures were affected by the interplay of different interactions.

1.5 References

1. J. -M. Lehn, *Chem. Soc. Rev.*, 2007, **36**, 151–160.
2. J. -M. Lehn, Angew. *Chem., Int. Ed. Engl.*, 1988, **27**, 89.
3. J. -M. Lehn, *Science*, 2002, **295**, 2400–2402.
4. C. J. Pedersen, *Angew. Chem. Int. Ed.*, 1988, **27,** 1021–1027.
5. D. J. Cram, *Angew. Chem. Int. Ed.*, 1988, **27**, 1009–1020.
6. G. Yu and X. Chen, *Theranostics.*, 2019, **9**, 3041–3074.
7. D. Lombardo, M. A. Kiselev, S. Magazù and P. Calandra, *Advances in Condensed Matter Physics*, 2015, **2015**, 1–22.
8. M. S. Ekiz, G. Cinar, M. A. Khalily and M. O Guler, *Nanotechnology*, 2016, **27**, 1–37.
9. D. N. Reinhoudt and M. Crego-Calma, *Science*, 2002, **295**, 2403–2407.
10. G. M. Whitesides, J. P. Mathias, C. T. Seto, *Science,* 1991, **254**, 1312–1319.
11. J. H. V. Esch, *Nature*, 2010, **466**, 193–194.
12. M. Eigen, *Naturwissenschaften*, 1971, **58**, 465–523.
13. S. Yadav, A. K. Sharma and P. Kumar, *Frontiers in Bioengineering and Biotechnology*, 2020, **8**, 1–24.
14. L. Brunsveld, B. J. B. Folmer, E. W. Meijer and R. P. Sijbesma, *Chem. Rev.*, 2001, **101**, 4071–4098.
15. J. M. A Carnall, C. A. Waudby, A. M. Belenguer, M. C. A. Stuart, J. J. P. Peyralans, S. Otto, *Science*, 2010, **327**, 1502–1506.

16. S. N. Semenov, L. J. Kraft, A. Ainla, *et al.*, *Nature*, 2016, **537**, 656–660.
17. O. J. G. M. Goor, S. I. S. Hendrikse, P. Y. W. Dankers and E. W. Meijer, *Chem. Soc. Rev.*, 2017, **46**, 6621–6637.
18. J. -M. Lehn, in *Supramolecular Chemistry: Concepts and Perspectives*, VCH, Weinheim, Cambridge, 1995, vol. 1, ch. 10, pp. 199–206.
19. G. M. Whitesides and B. Grzybowski, *Science*, 2002, **295**, 2418–2421.
20. K. Subramani and W. Ahmed, in *Emerging Nanotechnologies in Dentistry*, Elsevier, Amsterdam, 2nd edn, 2018, vol. 1, ch. 12, pp. 231–249.
21. S. Zhang, *Interface Focus*, 2017, **7**, 1–18.
22. M. Fialkowski, K. J. M. Bishop, R. Klajn, S. K. Smoukov, C. J. Campbell, and B. A. Grzybowski, *J. Phys. Chem. B*, 2006, **110**, 2482–2496.
23. W. Wang, J. Giltinan, S. Zakharchenko and M. Sitti, *Sci. Adv.*, 2017, **3**, 1–9.
24. A. Goldbeter, *Phil. Trans. R. Soc. A.*, 2018, **376**, 1–25.
25. A. Barbacci, V. Magnenet and M. Lahaye, *Thermodynamical journey in plant biology. Front. Plant Sci.*, 2015, **6**, 1–8.
26. D. kondepudi, T. Petrosky and J. A. Pojman, *Chaos*, 2017, **27**, 104501–104505.
27. B. A. Grzybowski, K. Fitzner, J. Paczesn and S. Granick, *Chem. Soc. Rev.*, 2017, **46**, 5647–5678.
28. E. Mattia and S. Otto, *Nature Nanotechnology*, 2015, **10**, 111–119.
29. G. Ashkenasy, T. M. Hermans, S. Otto and A. F. Taylor, *Chem. Soc. Rev.*, 2017, **46**, 2543–2554.
30. R. F. Ludlow and S. Otto, *Chem. Soc. Rev.*, 2008, **37**, 101–108.
31. S Otto, S. Kubik, S. I. Pascu, P. T. Corbett and Z. Rodriguez-Docampo, *Beilstein-Institut*, 2007, 207–219.
32. H. Duim and S. Otto, *Beilstein J Org Chem.*, 2017, **13**, 1189–1203.
33. P. L. Luisi, *Foundations of Chemistry*, 2002, **4**, 183–200.
34. P. Strazewski, *Life*, 2019, **9**, 1–7.
35. Y. Bai, A. Chotera, O. Taran, C. Liang, G. Ashkenasy and D. G. Lynn, *Chem. Soc. Rev.*, 2018, **47**, 5444–5456.
36. I. S. Choi, N. Bowden and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 1999, **38**, 3078–3081.
37. S. Datta, M. L. Saha and P. J. Stang, *Acc. Chem. Res.*, 2018, **51**, 2047–2063.
38. T. Lin, J. Yan, L. L. Ong, J. Robaszewski, H. D. Lu, Y. Mi, P. Yin and B. Wei, *Nano Letters*, 2018, **18**, 4791–4795.

39. A. Dasgupta, J. H. Mondal and D. Das, *RSC Adv.*, 2013, **3**, 9117–9149.
40. M. Piota, B. Abécassisb, D. Brourid, C. Troufflarda A. Prousta and G. Izzeta, *PNAS*, 2018, **115**, 8895–8900.
41. R. Ajdary, B. L. Tardy, B. D. Mattos, L. Bai and O. J. Rojas, *Adv. Mater.*, 2020, **2001085**, 1–31.
42. J. A. A. W. Elemans, A. E. Rowan and R. J. M. Nolte, *J. Mater. Chem.*, 2003, **13**, 2661–2670.
43. P. Egan, R. Sinko, P. R. LeDuc and S. Keten, *Nature Communications*, 2015, **6**, 1–12.
44. Collagen, https://bio.libretexts.org/@go/page/16171, (accessed Jun 29, 2021).
45. L. Wang, J. Lin, X. Zhang, *Polymer*, 2013, **54**, 3427–3442.
46. L. Cademartiri, K. J. M. Bishop, P. W. Snyder and G. A. Ozin, *Phil. Trans. R. Soc. A.*, 2012, **370**, 2824–2847.
47. R. W. Korn, *Biol Philos*, 2005, **20**, 137–151.
48. S. P. W. Wijnands, E. W. Meijer and M. Merkx, *Bioconjugate Chem.*, 2019, **30**, 1905−1914.
49. M. L. Daly, Y. Gao, and R. Freeman, *Bioconjugate Chem.*, 2019, **30**, 1864–1869.
50. X. Li, Y. Kuang, J. Shi, Y. Gao, H. -C. Lin and B. Xu, *J. Am. Chem. Soc.*, 2011, **133**, 17513–17518.
51. Y. Vyborna, M. Vybornyi and Robert Häner, *J. Am. Chem. Soc.*, 2015, **137**, 14051–14054.
52. C. D. Spicer, E. T. Pashuck, and M. M. Stevens, *Chem. Rev.*, 2018, **118**, 7702–7743.
53. T. MacCulloch, A. Buchberger and N. Stephanopoulos, *Org. Biomol. Chem.*, 2019, **17**, 1668–1682.
54. L. Li, R. Sun and R. Zheng, *Materials and Design*, 2021, **197**, 1–23.
55. J. Kalia and R. T. Raines, *Curr Org Chem.*, 2010, **14**, 138–147.
56. B. O. Okesola and A. Mata, *Chem. Soc. Rev.*, 2018, **47**, 3721–3736.
57. F. Li and R. I. Mahato, *Mol Pharm.*, 2017, **14**, 1321–1324.
58. K. Werengowska-Ciećwierz, M. Wiśniewski, A. P. Terzyk and S. Furmaniak, *Advances in Condensed Matter Physics*, 2015, **2015**, 1–27.
59. L. Juen, C. B. Baltus, C. Gély, *et al.*, *Bioconjugate Chemistry*, 2021, **32**, 595–606.
60. R. Freeman, N. Stephanopoulos, Z. Álvarez, *et al.*, *Nat Commun*, 2017, **8**, 1–11.
61. C. Wu, T. Schneider, M. Zeigler, *et al.*, *Journal of the American Chemical Society*, 2010, **132**, 15410–15417.
62. C. J. Pickens, S. N. Johnson, M. M. Pressnall, M. A. Leon and C. J. Berkland, *Bioconjugate Chemistry*, 2018, **29**, 686–701.

63. E. Mahon, A. Salvati, F. B. Bombelli, I. Lynch and K. A. Dawson, *Journal of Controlled Release*, 2012, **161**, 164–174.
64. B. A. Grzybowski and W. T. S. Huck, *Nature Nanotechnology*, 2016, **11**, 585–592.
65. J. M. J. M. Ravasco, H. Faustino, A. Trindade and P. M. P. Gois, *Chem. Eur. J.*, 2019, **25**, 43–59.
66. C. D. Spicer, and B. G. Davis, *Nat. Commun.*, 2014, **5**, 1−14.
67. S. A. Fisher, A. E. G. Baker, M. S. Shoichet, *J. Am. Chem. Soc.*, 2017, **139**, 7416−7427.
68. C. Chen, D. Y. W. Ng and T. Weil, *Progress in Polymer Science*, 2020, **105**, 1–40.
69. F. Hong, F. Zhang, Y. Liu and H. Yan, *Chem. Rev.*, 2017, **117**, 12584–12640.
70. N. Stephanopoulos, J. H. Ortony and S. I. Stupp, Acta Mater., 2013, **61**, 912–930.
71. B. A. Juodka, *Nucleosides Nucleotides Nucleic Acids*, 1984, **3**, 445–483.
72. IUPAC-IUB Comm. on Biochem. Nomenclature, *Biochem. J.*, 1984, **219**, 345–373.
73. D. Mandal, A. N. Shirazi and K. Parangb, *Org Biomol Chem.*, 2014, **14**, 3544–3561.
74. J. Lee, M. Ju, O. H. Cho, Y. Kim and K. T. Nam, *Adv. Sci.*, 2019, **6**, 1–15.
75. B. B. Gerbelli, et al., *Macromol. Chem. Phys*., 2019, **220**, 1–22.
76. R. Mehta, J. Lund and B. A. Parviz, in *Nanofabrication: Fundamentals and Applications*, ed. A. A. Tseng, World Scientific, Singapore, 1st edn, 2008, vol. 1, ch. 4, pp. 127–158.
77. M. M. Gromiha, in *Protein Bioinformatics From Sequence to Function*, Elsevier, Amsterdam, 1st edn, 2010, vol. 1, ch. 1, pp. 1–27.
78. L. Skipper, in *Encyclopedia of Analytical Science*, ed. P. Worsfold, A. Townshend and C. Poole, Elsevier, Amsterdam, 2nd edn, 2005, vol. 1, pp. 344–352.
79. D. P. Clark, N. J. Pazdernik and M. R. McGehee, in *Molecular Biology*, Elsevier, Amsterdam, 3rd edn, 2019, vol. 1, ch. 14, pp. 445–483.
80. H. -X. Zhou and X. Pang, *Chem Rev.*, 2018, **118**, 1691–1741.
81. D. Eisenberg, *Proc. Natl. Acad. Sci. USA.*, 2003, **100**, 11207–11210.
82. N. Stephanopoulos, *Bioconjugate Chem.*, 2019, **30**, 1915–1922.
83. A. Buchberger, C. R. Simmons, N. E. Fahmi, R. Freeman and N. Stephanopoulos, *J. Am. Chem. Soc.*, 2020, **142**, 1406–1416.
84. S. Lou, X. Wang, Z. Yu, and L. Shi, *Adv. Sci.*, 2019, **6**, 1–24.
85. R. Lakhia, A. Mishra and V. Patel, in *Methods in Cell Biology*, ed. T. Weimbs, Elsevier, Amsterdam, 1st edn, 2019, vol. 154, ch. 7, pp. 109–120.
86. H. Lodish, A. Berk, S. L. Zipursky, *et al.*, in *Molecular Cell Biology*, W. H. Freeman, New York, 4th edn, 2000, ch. 4, pp. 101–145.
87. N. Pollak, C. Dölle, and M. Ziegler, *Biochem J.*, 2007, **402**, 205–218.

88. B. Alberts, A. Johnson, J Lewis J, *et al.*, in *Molecular Biology of the Cell*, Garland Science, New York, 4th edn, 2002, ch. 4.
89. C. K. McLaughlin, G. D. Hamblin and H. F. Sleiman, *Chem. Soc. Rev.*, 2011, **40**, 5647–5656.
90. H. Lee, *Nat. Nanotechnol*., 2012, **7**, 389−393.
91. A. Chotera, H. Sadihov, R. Cohen-Luria, P. A. Monnard and G. Ashkenasy, *Chem. Eur. J.*, 2018, **24**, 10128–10135.
92. B. A. R. Williams, C. W. Diehnelt, P. Belcher, M. Greving, N. W. Woodbury, S. A. Johnston and J. C. Chaput, *J. Am. Chem. Soc.*, 2009, **131**, 17233–17241.
93. N. Gour, D. Kedracki, I. Safir, K. X. Ngo and C. Vebert-Nardin, *Chem. Commun.*, 2012, **48**, 5440–5442.
94. Y. Miura, *J. Mater. Chem. B*, 2020, **8**, 2010–2019.
95. M. Kye and Y. B. Lim, *Angew. Chem., Int. Ed.*, 2016, **55**, 12003–12007.
96. R. Freeman et al., *Science*, 2018, **362**, 808–813.
97. C. Frantz, K. M. Stewart and V. M. Weaver, *J Cell Sci.*, 2010, **123**, 4195–4200.
98. C. Rinaldi and M. J. A. Wood, *Nature Reviews Neurology*, 2018, **14**, 9–21.
99. R. Kole, A. R. Krainer and S. Altman, *Nat Rev Drug Discov.*, 2012, **11**, 125–140.
100. D. D. Fusco, V. Dinallo, I. Marafini, M. M. Figliuzzi, B. Romano and G. Monteleone, *Front. Pharmacol.*, 2019, **10**, 1–7.
101. R. L. Juliano, *Nucleic Acids Res.*, 2016, **44**, 6518–6548.
102. R. Martier and P. Konstantinova, *Front. Neurosci.*, 2020, **14**, 1–26.
103. J. Winkler, *Ther Deliv.*, 2013, **4**, 791–809.
104. N. Venkatesan and B. H. Kim, *Chem. Rev.*, 2006**, 106**, 3712−3761.
105. E. Valeur, *Angew. Chem. Int. Ed.*, 2017, **56**, 10294−10323.
106. M. Gait, *J. Cell. Mol. Life Sci.*, 2003, **60**, 844–853.
107. O. Koniev and A. Wagner, *Chem. Soc. Rev.,* 2015, **44**, 5495–5551.
108. K. Lu, Q. -P. Duan, L. Ma, and D. -X. Zhao. *Bioconjugate Chem.*, 2010, **21**, 187–202.
109. M. Wenska, M. Alvira, P. Steunenberg, A. Stenberg, M. Murtola and R. Stromberg, *Nucleic Acids Research*, 2011, **39**, 9047–9059.
110. T. Suzuki, S. Ohsumi and K. Makino, *Nucleic Acids Res*., 1994, **25**, 4997–5003.
111. K. Gogoi, M. V. Mane, S. S. Kunte, and V. A. Kumar, *Nucleic Acids Research*, 2007, **35**, 1−7.

2 Peptide Synthesis

2.1 Chapter Overview

This chapter describes the synthesis and characterisation of β-sheet forming peptides (**KLVFFA**, **HYFNIF** and **RVFNIM**)[1,2] and a collagen mimetic peptide **$(POG)_6$**[3] that would be subsequently used for conjugation reactions with DNA oligonucleotides (**DNA1** and **DNA1'**). **KLVFFA**, **HYFNIF** and **RVFNIM** were chosen because these peptide sequences are known to form amyloid-like fibrils,[1,2] while **$(POG)_6$** was selected given that proline-hydroxyproline-glycine repeating **(POG)** units have been shown to display collagen triple-helix characteristic.[3] Peptide sequences displaying these unique self-assembling structures were relevant for the purpose of studying their self-assembly properties before and after being conjugated with DNA. The desired peptide sequences were generated with additional modifications to make the linkage with DNA possible.

The first part of this chapter focuses on the use of Fmoc solid-phase synthesis to obtain the desired peptide sequences, highlighting the three different types of resins used, with the characterisation of each peptide by NMR and LC-MS being discussed. Lastly, the chapter describes the self-assembling behaviour of each peptide, showing the results of circular dichroism (CD) and atomic force microscopy (AFM), as well as comparing these findings to those reported in the literature.

2.2 Synthetic Strategies

Specific synthetic strategies were required to produce suitable peptides that would react with DNA oligonucleotides using three different conjugation approaches: thiol-maleimide reactions;[4] amide coupling;[5] and strain-promoted alkyne-azide cycloadditions (SPAAC).[6]

For thiol-maleimide conjugation reactions with maleimide-functionalised DNA, cysteine-containing peptides were synthesised using Rink amide resin or Wang resin. For amide coupling with amine-functionalised DNA, side-chain protected peptides with a free C-terminus were synthesised using Chlorotrityl Chloride (CTC) resin. For conducting strain-promoted alkyne-azide cycloadditions (SPAAC) with alkyne-modified DNA,

azide-modified peptides were synthesised using Rink amide resin.

The peptide sequence **CGSG** was attached to the target peptides to act as a flexible spacer,[7] prior to undergoing conjugation with the reactive DNA linker. This spacer provides flexibility with low hindrance, reducing the likelihood of DNA and peptide interfering with each other upon self-assembly.

2.3 Materials and Methods

Fmoc-protected amino acids used for peptide synthesis were purchased from Fluorochem or Sigma-Aldrich. DCM, DMF, MeOH and Et_2O were purchased from Fisher Scientific. Acetic anhydride and 1-butanol were purchased from Acros Organics. Piperidine was purchased from Alfa Aesar. 2-Chlorotrityl Chloride resin, Rink Amide MBHA resin, EDC hydrochloride, DMAP, TBTU, PyBOP, HOBt, DIPEA, TFA and TIPS were purchased from Fluorochem. 3-Azidoproploanic acid was purchased from BroadPharm. KCN was purchased from Amersham.

2.4 Peptide Synthesis and Characterisation

All the peptides reported in this chapter were generated through standard Fmoc solid-phase peptide synthesis (SPPS) procedures (Figure 2.1). In this strategy, the desired peptides were synthesised stepwise by being immobilised to an insoluble resin support using Fmoc-protected amino acids. The Fmoc group at the N-terminus of amino acids was removed using 20% piperidine in DMF, followed by coupling with the C-terminus of another amino acid. Each coupling reaction was monitored by the Kaiser test[8] to ensure the disappearance of the primary amine before proceeding to the next coupling cycle. Coupling reactions were repeated when the Kaiser test showed a positive result.

The resin was subsequently washed with DMF and DCM to remove the by-products generated during each reaction. Following DCM washing, MeOH was used to shrink the resin before the peptide was cleaved from the resin by employing acid treatment (the conditions are defined in the peptide synthesis section below). All the reaction steps were carried out manually at R.T. (20-25°C) using a 25 mL peptide synthesis vessel consisting of a

sintered glass funnel and a 3-way stopcock for easy filtration and washing.

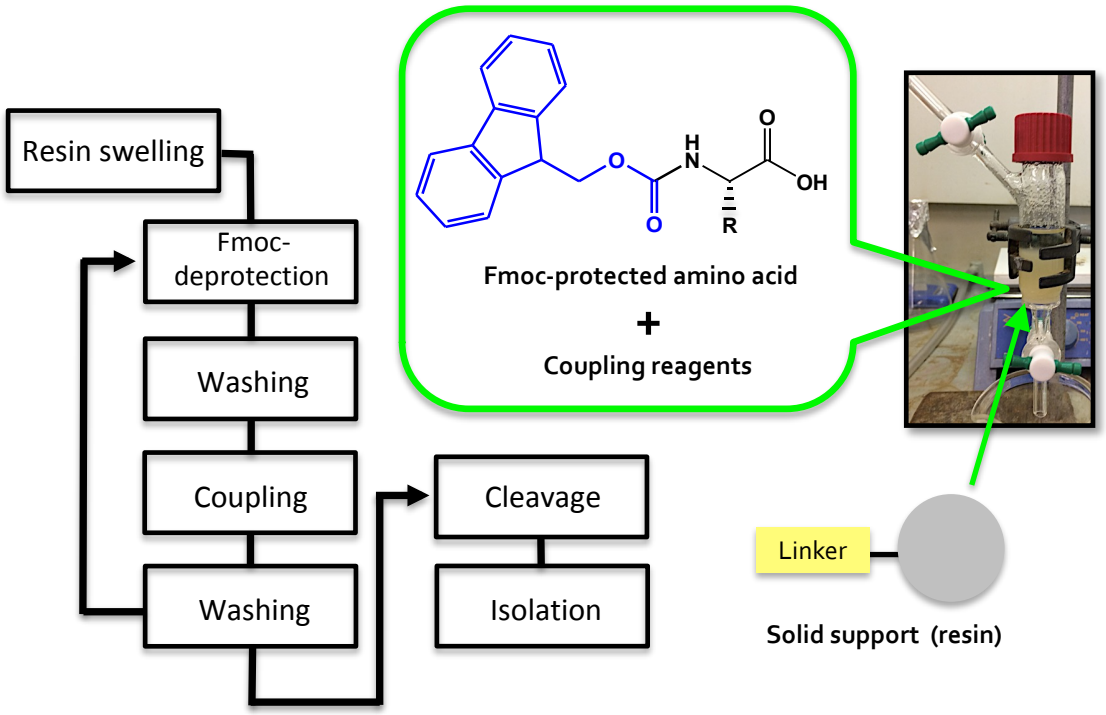


Figure 2.1. Schematic representation of Fmoc solid-phase peptide synthesis (SPPS) and an image showing the peptide synthesis vessel employed.

Kaiser Test

The Kaiser test detects the presence of primary amines and is commonly used in solid-phase peptide synthesis to assess whether coupling reactions are successful.[8] The Kaiser test is based on the reaction of ninhydrin with primary amines, which produces an intense blue colour known as Ruhemann's purple (Figure 2.2). This colour is observed when unreacted N-terminal amines are present following a coupling reaction, indicating a positive result. The solutions used in the Kaiser test include pyridine, phenol and potassium cyanide (KCN), which facilitate the formation of Ruhemann's purple.[9]

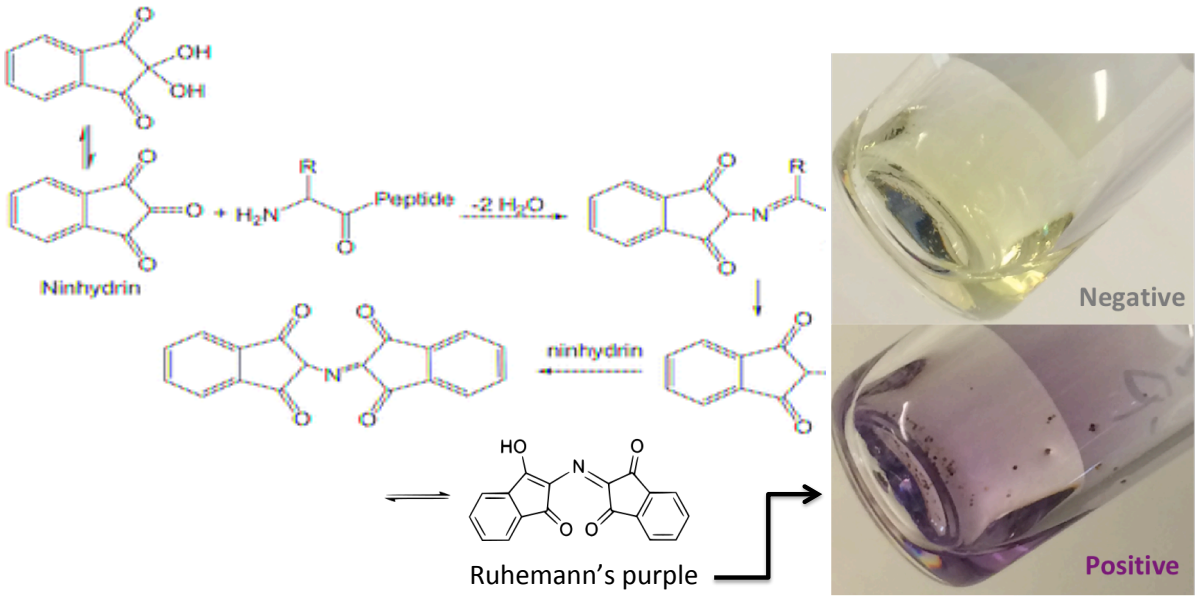


Figure 2.2. Ninhydrin reaction mechanism[10] and images of the Kaiser test showing a negative and a positive result. Reaction mechanism of peptides with ninhydrin (Kaiser test) was adapted and reproduced with permission from ref 10.

The Kaiser test solutions were prepared following the protocols as illustrated in Table 1.[8] The Kaiser test was carried by adding 2-3 drops of Reagent A, B and C to a glass vial containing a few resin beads. A reference sample was prepared by adding 2-3 drops of Reagent A, B and C to an empty glass vial. Both vials were heated at 110 °C for 5 min, prior to comparing their colour.

Table 1. Chemicals used for each Kaiser test solution

Reagent A	33 mg of KCN was dissolved in 25 mL of deionized water. 1.0 mL of this solution was diluted with 49 mL of pyridine
Reagent B	1.0 g of ninhydrin was dissolved in 20 mL of 1-butanol
Reagent C	20 g of phenol was dissolved in 20 mL of 1-butanol

Chloranil Test

The Chloranil test is used to detect the presence of secondary amines. The solution used for this test was prepared by dissolving 1 g of p-chloranil in 49 mL of DMF. To carry out the test, 2-3 drops of this solution was added to a glass vial containing a few resin beads. A reference sample was prepared by adding 2-3 drops of this solution to an empty glass vial. Both vials were left to stand at R.T. for 5 min to compare their colour. Presence of blue/purple colour

indicated a positive result.[11]

Synthesis of Peptides with a free C-terminus using Wang Resin

Wang resin (4-alkoxybenzyl alcohol resin) is used in solid phase peptide synthesis for generating peptides with C-terminal carboxylic acids.[12] Peptides were synthesised on solid support using Wang resin (1.0 g, loading capacity: 1 mmol/g) through standard Fmoc procedures (Figure 2.3). The resin was swollen in DCM for 30 min, and then filtered. 2 equivalents of the first amino acid, 2 equivalents of HOBt, 2 equivalents of EDC and 0.1 equivalent of DMAP were dissolved in a minimum amount of DMF, which was then added to the resin. The reaction mixture was left to stand overnight with occasional agitation.

The resin was washed with DMF (5 mL × 3), DCM (5 mL × 3) and DMF (5 mL × 3), followed by capping any unreacted sites on the resin with 2 equivalents of Ac_2O and 2 equivalents of DIPEA in DMF for 2 h. The resin was washed with DMF (5 mL × 5), prior to conducting Fmoc deprotection by suspending the resin with 20% piperidine in DMF for 20 min. The resin was washed with DMF (5 mL × 5), prior to conducting the coupling by reacting 3 equivalents of an amino acid, 3 equivalents of HOBt and 3 equivalents of TBTU, and 6 equivalents of DIPEA in DMF for 40-60 min. The resin was subsequently washed with DMF (5 mL × 5).

The Fmoc deprotection and coupling steps were repeated until the desired sequence was obtained on the resin. The N-terminus of the peptide was acetylated by suspending the resin with 1 mL of Ac_2O and 1 mL of DIPEA per gram of resin in DMF for 2 h. Following washing with DMF (5 mL × 3), DCM (5 mL × 3) and MeOH (5 mL × 3), the resin was transferred to a glass vial and dried under vacuum overnight. The peptide was cleaved by suspending the dry resin with a cold cleavage solution consisting of TFA/H_2O/TIPS (94:3:3 v/v, 10 mL per gram of resin) for 90 min. The resin was filtered and washed with TFA (3 × 0.5 mL). The solution was then collected in a round bottom flask and TFA was removed under a stream of N_2. The peptide was precipitated with cold Et_2O, which was then collected by filtration and dried under vacuum.

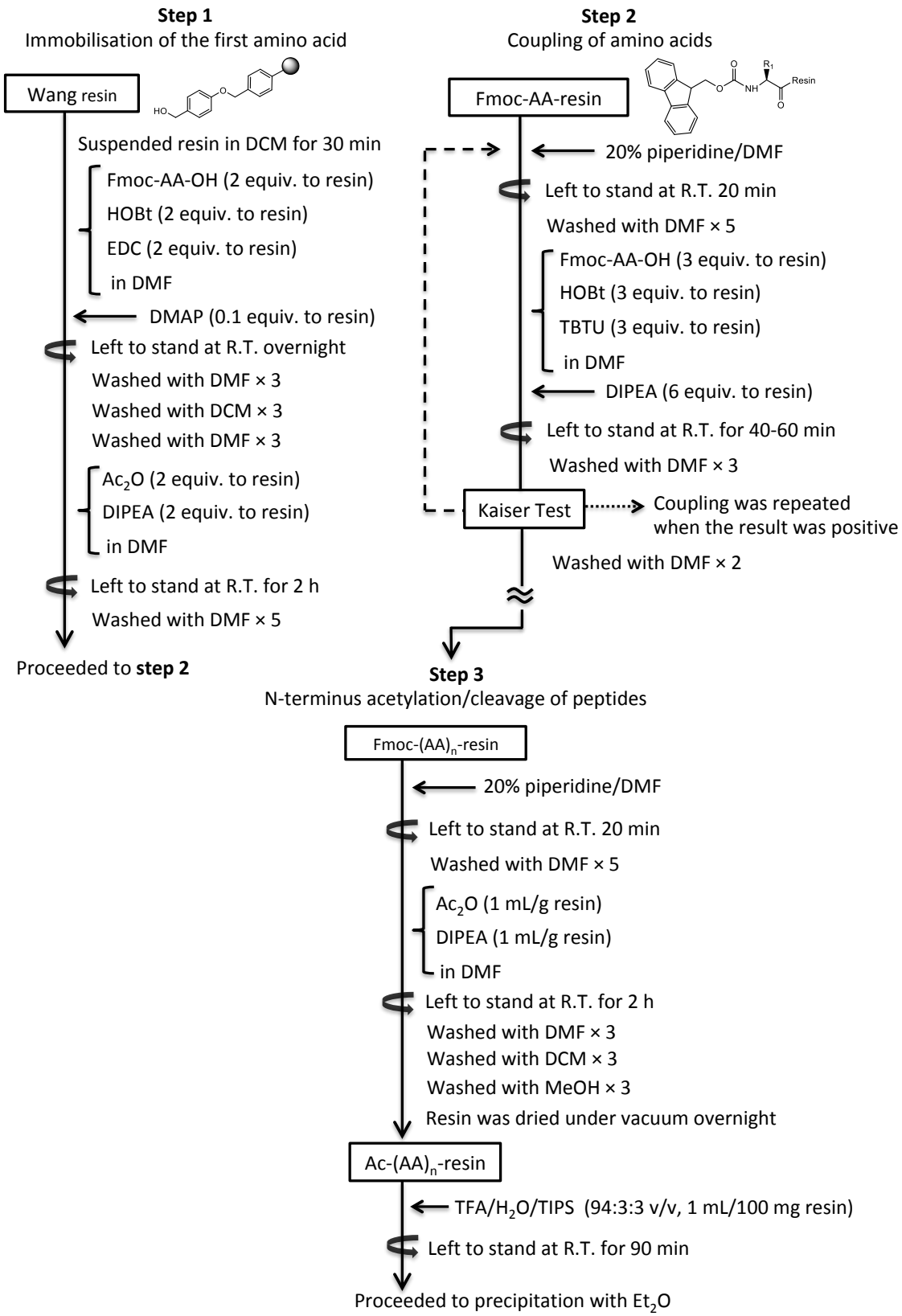


Figure 2.3. Immobilisation, coupling and cleaving steps of peptide synthesis using Wang resin. Coupling steps were repeated where appropriate.

Synthesis of Peptides with an Amidated C-terminus using Rink MBHA Amide Resin

Rink MBHA Amide resin is used in solid phase peptide synthesis for generating peptides with an amidated C-terminus. Peptides were synthesised on a solid phase Rink Amide MBHA resin (1.0 g, loading capacity: 0.55 mmol/g) through standard Fmoc procedures (Figure 2.4). The resin was swollen in DCM for 30 min and then filtered, followed by Fmoc deprotection of the resin by suspending with 20% piperidine in DMF for 20 min.

The resin was then washed with DMF (5 mL × 5), prior to coupling by reacting 3 equivalents of the first amino acid, 3 equivalents of HOBt and 3 equivalents of TBTU or PyBOP, and 6 equivalents of DIPEA in DMF. The reaction mixture was left to stand overnight with occasional agitation. The resin was washed with DMF (5 mL × 5), followed by capping any unreacted sites on the resin with 2 equivalents of Ac_2O and 2 equivalents of DIPEA in DMF for 2 h. The resin was subsequently washed with DMF (5 mL × 5), prior to conducting Fmoc deprotection by suspending the resin with 20% piperidine in DMF for 20 min. The resin was then washed with DMF (5 mL × 5), prior to conducting the coupling by reacting 3 equivalents of an amino acid, 3 equivalents of HOBt and 3 equivalents of TBTU or PyBOP, and 6 equivalents of DIPEA in DMF for 40-90 min. The resin was subsequently washed with DMF (5 mL × 5).

The Fmoc deprotection and coupling steps were repeated until the desired sequence was obtained on the resin. The N-terminus of the peptide was acetylated by suspending the resin with 1 mL of Ac_2O and 1mL of DIPEA per gram of resin in DMF for 2 h. Following washing with DMF (5 mL × 3), DCM (5 mL × 3) and MeOH (5 mL × 3), the resin was transferred to a glass vial and dried under vacuum overnight. The peptide was cleaved by suspending the dry resin with a cold cleavage solution consisting of TFA/H_2O/TIPS (94:3:3 v/v, 10 mL per gram of resin) for 90 min. The resin was filtered and washed with TFA (3 × 0.5 mL). The solution was then collected in a round bottom flask and TFA was removed under a stream of N_2. The peptide was precipitated with cold Et_2O, which was then collected by filtration and dried under vacuum.

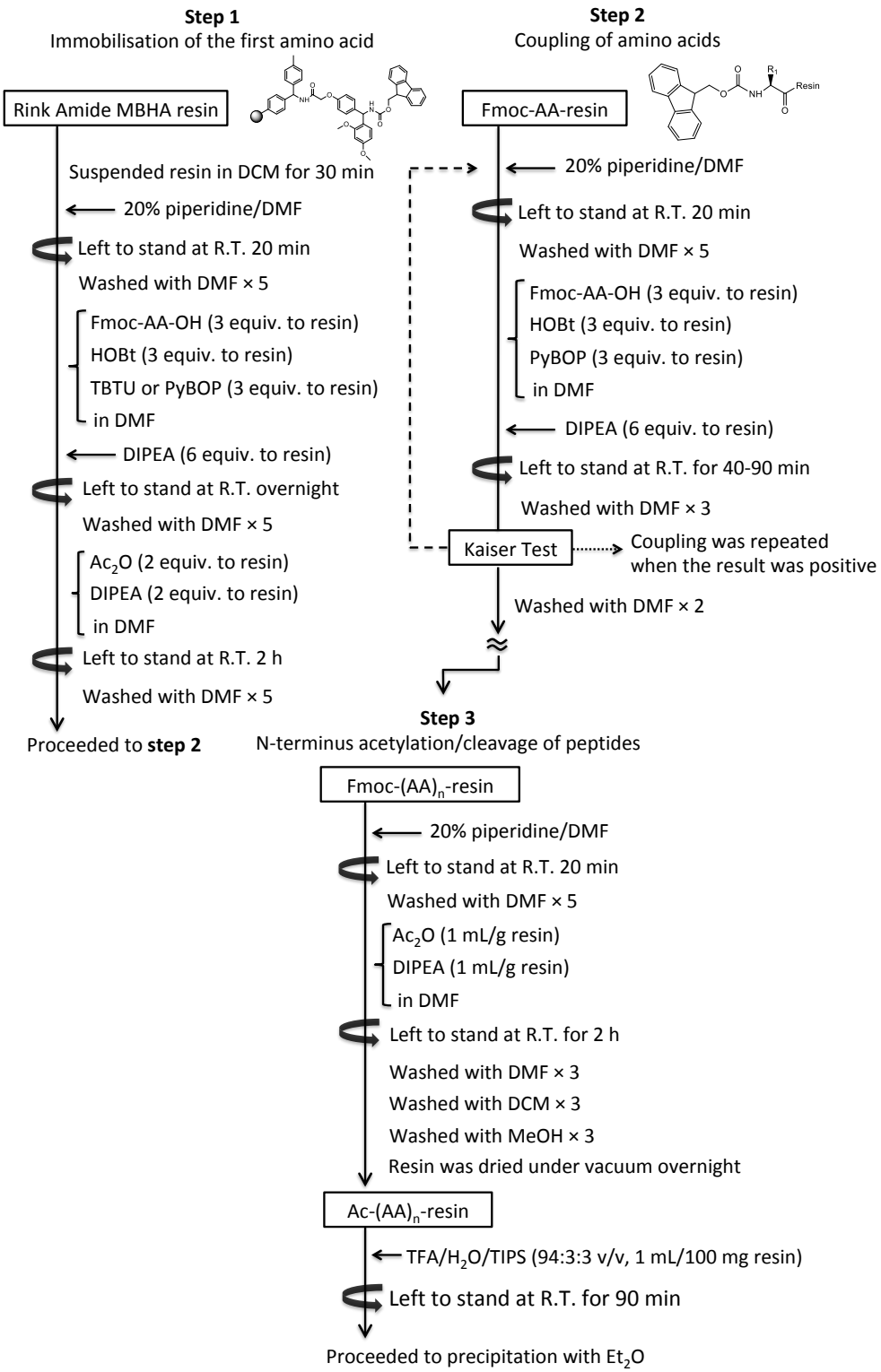


Figure 2.4. Immobilisation, coupling and cleaving steps of peptide synthesis using Rink Amide MBHA resin. Coupling steps were repeated where appropriate.

Synthesis of Side-chain Protected Peptide using 2-Chlorotrityl Chloride Resin

2-Chlorotrityl chloride resin is used in solid phase peptide synthesis for generating peptides with a carboxylic acid C-terminus and fully protected side-chains.[13] Peptide synthesis was performed on a solid support using 2-chlorotrityl chloride resin (0.5 g, 1.9 mmol/g loading) through standard Fmoc procedures (Figure 2.5). The resin was swollen in dry DCM for 30 min, and then filtered. 2 equivalents of the first amino acid dissolved in DMF and 2 equivalents of DIPEA were added to the resin. Following the agitation of the reaction mixture for 5 min, 3 equivalents of DIPEA were added to the resin, which was then left to stand for 2 hours with occasional agitation.

The resin was washed with DMF (5 mL × 5), followed by capping any unreacted sites on the resin with 0.4 mL of MeOH for 15 min. The resin was subsequently washed with DMF (5 mL × 5), prior to conducting Fmoc deprotection by suspending the resin with 20% piperidine in DMF for 20 min. The resin was then washed with DMF (5 mL × 5), prior to conducting the coupling by reacting 2 equivalents of an amino acid, 2 equivalents of HOBt and 2 equivalents of PyBOP, and 4 equivalents of DIPEA in DMF for 90 min. The resin was subsequently washed with DMF (5 mL × 5).

The Fmoc deprotection and coupling steps were repeated until the desired sequence was obtained on the resin. The N-terminus of the peptide was acetylated by suspending the resin with 1 mL of Ac_2O and 1mL of DIPEA per gram of resin in DMF for 2 h. Following washing with DMF (5 mL × 5), the peptide was cleaved by suspending in 5 mL of 20% HFIP in DCM for 90 min. The resin was filtered and collected in a round bottom flask. The solution was evaporated under a stream of N_2. The peptide was precipitated with cold Et_2O, which was then collected by filtration and dried under vacuum.

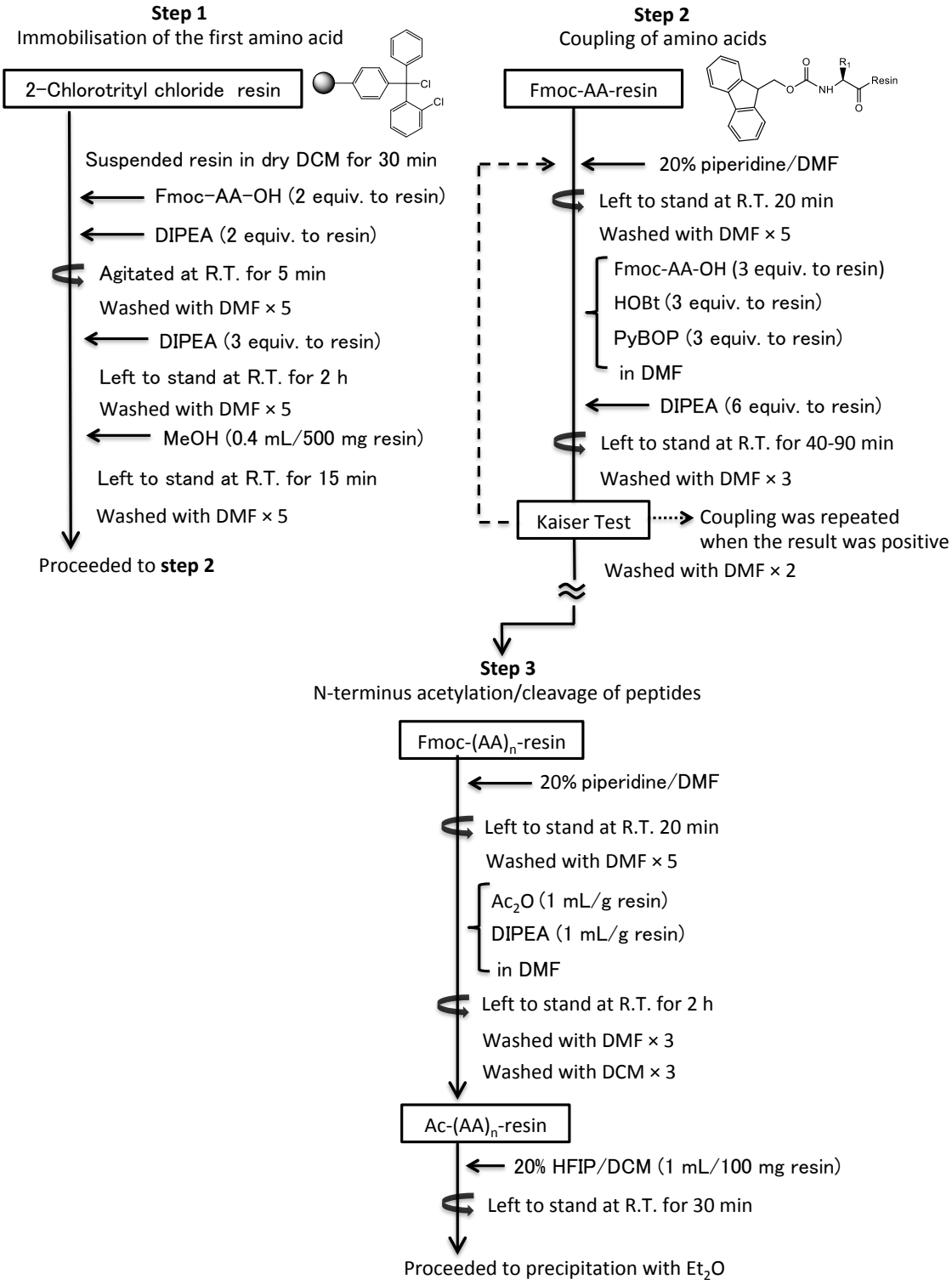


Figure 2.5. Immobilisation, coupling and cleaving steps of peptide synthesis using CTC resin. Coupling steps were repeated where appropriate.

Azide Modification of HYFNIF

3-azidopropionic acid was introduced at the N-terminus of **HYFNIF** following the synthesis by rink amide MBHA resin. 250 mg of Fmoc-HYFNIF-rink amide MBHA resin was deprotected by suspending in 20% piperidin in DMF for 20 min. The resin was then washed with DMF (5 mL × 5), prior to conducting the coupling by reacting 3-azidopropionic acid (47 µL, 0.5 mmol), HOBt (77 mg, 0.5 mmol), PyBOP (260 mg, 0.5 mmol) and DIPEA (175 µL, 1.0 mmol) in DMF for 2.5 h. The resin was subsequently washed with DMF (5 mL × 3), DCM (5 mL × 3) and MeOH (5 mL × 3). The resin was transferred to a glass vial and dried under vacuum overnight. The peptide was cleaved by suspending the dry resin with 5 mL cold cleavage solution consisting of TFA/H_2O/TIPS (94:3:3 v/v) for 95 min. The resin was filtered and washed with TFA (3 × 0.5 mL). The solution was then collected in a round bottom flask and TFA was removed under a stream of N_2. The peptide was precipitated with cold Et_2O, which was then collected by filtration and dried under vacuum.

Azide Modification of $(POG)_6$

3-azidopropionic acid was introduced at the N-terminus of **$(POG)_6$** following the synthesis by rink amide MBHA resin. 150 mg of Fmoc-$(POG)_6$-rink amide MBHA resin was deprotected by suspending in 20% piperidine in DMF for 20 min. The resin was then washed with DMF (5 mL × 5), prior to conducting the coupling by reacting 3-azidopropionic acid (28 µL, 0.3 mmol), HOBt (46 mg, 0.3 mmol), PyBOP (156 mg, 0.3 mmol) and DIPEA (105 µL, 0.6 mmol) in DMF for 2.5 h. The resin was subsequently washed with DMF (5 mL × 3), DCM (5 mL × 3) and MeOH (5 mL × 3). The resin was transferred to a glass vial and dried under vacuum overnight. The peptide was cleaved by suspending the dry resin with 4 mL cold cleavage solution consisting of TFA/H_2O/TIPS (94:3:3 v/v) for 90 min. The resin was filtered and washed with TFA (3 × 0.5 mL). The solution was then collected in a round bottom flask and TFA was removed under a stream of N_2. The peptide was precipitated with cold Et_2O, which was then collected by filtration and dried under vacuum.

^{1}H NMR

Proton nuclear magnetic resonance (^{1}H NMR) spectra were recoded with a Bruker AV2 400

MHz spectrometer in DMSO-d_6. Chemical shifts were reported in delta (δ) units, parts per million (ppm) downfield from trimethylsilane (TMS). Coupling constants were reported in Hertz (Hz). NMR data were analysed and integrated with ACD labs. Peptides were characterised by ^{1}H NMR based on theoretical estimates of hydrogen atom chemical shifts, or by making comparisons with ^{1}H NMR data collected for the other synthesised peptides.

LC-MS

Mass spectroscopy data of the peptides were collected on either a Thermo MSQ Plus LC-MS system or a Bruker micrOTOF-Q II mass spectrometer. Samples were separated by reverse-phase HPLC on a Phenomenex Aeris Widepore column (XB-C18, 3.6 µm, 2.1 mm x 150 mm) running on an Agilent 1100 HPLC system at a flow rate of 0.2 mL/min using a water, acetonitrile and 0.05% TFA gradient: mobile phase (A), 0.05% TFA in water; mobile phase (B), 80% acetonitrile, 0.045% TFA in water; 5% (B) for 5 min, then a linear gradient 5-100% (B) for 25 min, then held at 100% (B) for 5 min before returning to initial conditions. The eluent was monitored at 214 nm and then directed into the electrospray source at 4.5 kV. The mass spectra were recorded from 500-3000 m/z and analysed with Bruker's Compass Data Analysis software.

CD

Circular dichroism (CD) refers to the unequal absorption of left-handed and right-handed circular polarised light that is seen in optically active (chiral) molecules.[14] CD spectroscopy is a well-established method for characterising the structural properties of biological molecules based on their unique CD spectra.[15] CD spectroscopy was performed to gain insights into the secondary structure of the peptides synthesised.

CD spectra were obtained on a J-715 (Jasco) spectrophotometer using a quartz cuvette with a 1 mm path length and a volume of 350 µL. CD spectra were recorded at 20°C in the scan range of 190 nm to 260 nm, with a bandwidth of 1.0 nm and a scanning speed of 20 nm/min. Each CD spectrum was obtained by taking the average of 5 scans, prior to subtracting the background signal. Peptide samples were prepared in MilliQ water at 1 mg/mL, before

diluting at lower concentrations prior to conducting CD analysis.

AFM

Atomic force microscopy (AFM) measurements were carried out using a Bruker Multi-Mode microscope with a Quadrexed Nanoscope III controller using Bruker ScanAsyst-Air silicon tip on nitride lever, by scanning 512 × 512 pixels covering surfaces between 1 and 20 μm^2. AFM images obtained were processed by Nanoscope analysis software (Version 1.5, Bruker). The sample concentration used for AFM analysis (0.2 mg/mL) was chosen based on the CD intensity previously measured to ensure that self-assembled peptides could be visualised by AFM. To prepare samples, 20 µL of peptide solution in water was deposited into freshly cleaved mica for 10 min. The excess liquid was wiped off with the edge of a filter paper. Subsequently, the mica surface was dried under a stream of N_2. For annealing experiments, the samples were incubated in a thermocycler from 95 °C to 4 °C at a cooling rate of 1 °C per minute.

2.5. Results and Discussion

The peptides listed in Table 2.2 were synthesised and characterised by NMR and LC-MS. In addition to the four main peptide sequences (**KLVFFA**, **HYFNIF**, **RVFNIM** and **$(POG)_6$**), additional sequences were synthesised to gain practice with this methodology or as required for conjugation with DNA. Different crude yields were obtained among the synthesised peptides, with values ranging from 23-79%.

Table 2.2. List of synthesised peptides showing their predicted secondary structures and crude yields.

Peptide	Secondary structure	Crude yield, %
Ac-AFFVLK-COOH	Antiparallel β-sheet	N/A*
Ac-VIYKI-COOH	Parallel β-sheet	N/A*
Ac-KLVFFA-$CONH_2$	Antiparallel β-sheet	77
Ac-CGSGKLVFFA-$CONH_2$	Antiparallel β-sheet	75
Ac-HYFNIF-$CONH_2$	Parallel β-sheet	54
Ac-CHYFNIF-$CONH_2$	Parallel β-sheet	52
Ac-CGSGHYFNIF-$CONH_2$	Parallel β-sheet	49
Ac-H(Trt)-Y(tBu)-F-N(Trt)-F-GS(Trt)-G-COOH	Parallel β-sheet	23
N_3-$(CH_2)_2$-HYFNIF-$CONH_2$	Parallel β-sheet	52
Ac-RVFNIM-$CONH_2$	Parallel β-sheet	62
Ac-CGSGRVFNIM-$CONH_2$	Parallel β-sheet	49
Ac-$(POG)_3$-$CONH_2$	Collagen triple helix	77
Ac-$(POG)_6$-$CONH_2$	Collagen triple helix	76
Ac-C$(POG)_6$-$CONH_2$	Collagen triple helix	33
Ac-CGSG$(POG)_6$-$CONH_2$	Collagen triple helix	75
N_3-$(CH_2)_2$-$(POG)_6$-$CONH_2$	Collagen triple helix	79

*NMR and LC-MS could not be conducted owing to poor solubility of **Ac-AFFVLK-$COOH_2$** and **Ac-VIYKI-COOH** in aqueous or organic solvents, meaning that it was not possible to confirm whether the peptide was obtained.

Although the peptides with a free C-terminus (**Ac-AFFVLK-COOH** and **Ac-VIYKI-COOH**) were synthesised using Wang resin, their poor solubility in water and organic solvents made it

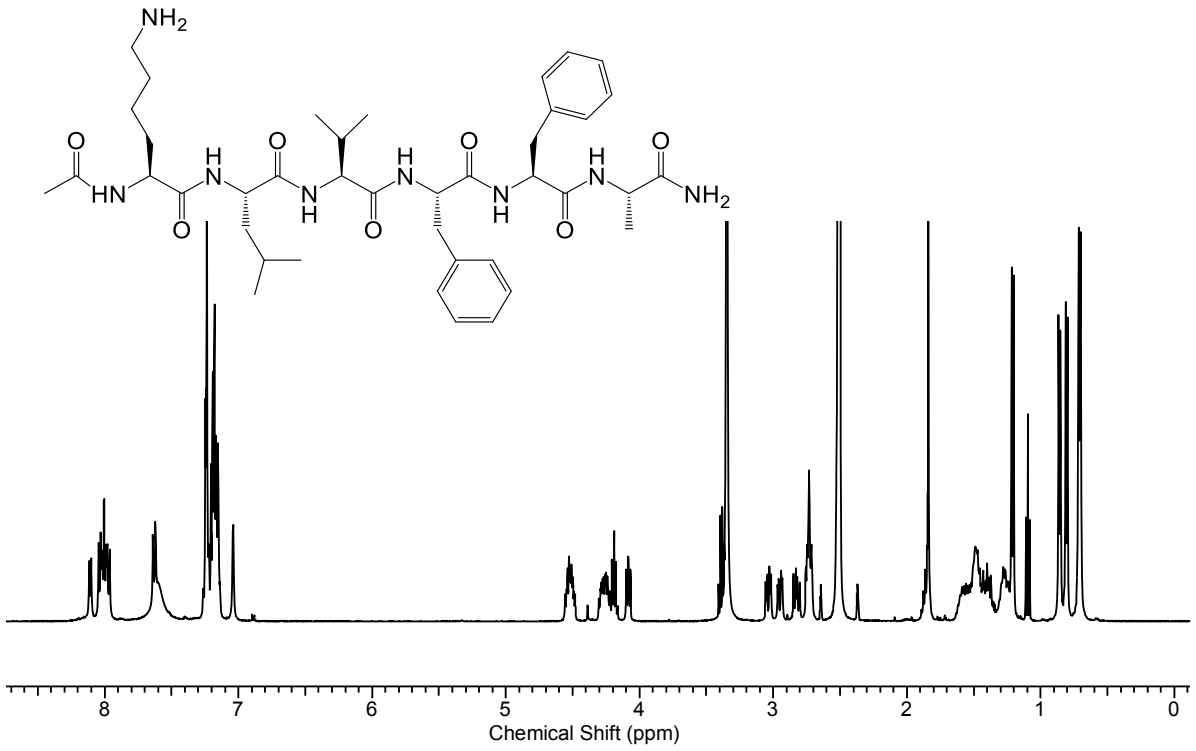
difficult to conduct further experiments including NMR and LC-MS measurements. Given that **Ac-AFFVLK-COOH** and **Ac-VIYKI-COOH** possess a net charge of zero at a neutral pH where the molecular interactions are enhanced, this potentially contributed to their poor solubility.[16]

To overcome this solubility issue, peptides with both termini capped (acetylated N-terminus and amidated C-terminus) were synthesised using Rink Amide MBHA resin. All the peptides obtained using this resin dissolved in water, allowing for further experiments including NMR, LC-MS, CD and AFM analysis.

Characterisation of Peptides by ^{1}H NMR and LC-MS

Ac-KLVFFA-CONH$_2$

The ^{1}H NMR showed all the protons peaks of **Ac-KLVFFA-CONH$_2$**, with the proton integration being similar to theoretical values (Figure 2.5). LC-MS confirmed the mass of the desired peptide (Figure 2.6).



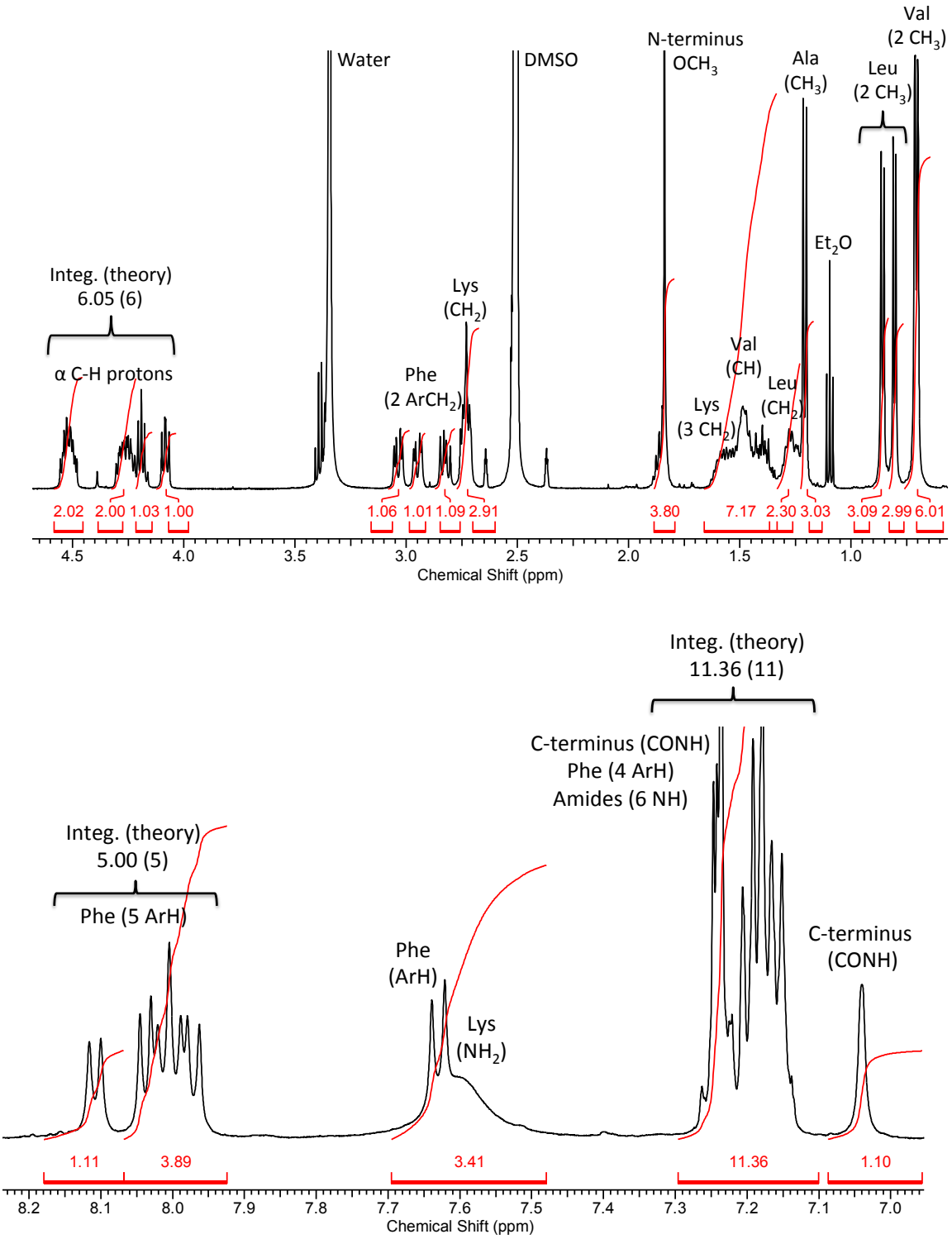


Figure 2.5. ^{1}H NMR of **Ac-KLVFFA-CONH$_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.71 (d, *J* = 5.4 Hz, CH_3 [Val], 6H), 0.76 (d, *J* = 5.2 Hz, CH_3 [Leu], 3H), 0.86 (d, *J* = 5.2 Hz, CH_3 [Leu], 6H), 1.21 (d, *J* = 5.6 Hz, CH_3 [Ala], 3H), 1.23-1.32 (m, CH_2 [Ile], 2H), 1.34-1.64 (m, CH [Val] and CH_2 [Lys], 7H), 1.84 (s, OCH_3 [N-terminus], 3H),

2.69-2.77 (m, CH_2 [Lys] and ArCH [Phe], 3H), 2.83 (dd, *J* = 9.9 and 6.2 Hz, ArCH [Phe], 1H), 2.95 (dd, *J* = 12.4 and 3.8 Hz, ArCH [Phe], 1H), 3.04 (dd, *J* = 10.2 and 5.6 Hz, ArCH [Phe], 1H), 4.08 (t, *J* = 6.4 Hz, αC-H, 1H), 4.15-4.22 (m, αC-H, 1H), 4.22-4.32 (m, αC-H, 2H), 4.47-4.57 (m, αC-H, 2H), 7.04 (s, CONH [C-terminus], 1H), 7.13-7.27 (m, CONH [C-terminus], ArH [Phe] and NH, 11H), 7.50-7.70 (m, NH_2[Lys] and ArH [Phe], 3H), 7.93-8.07 (m, ArH [Phe], 4H), 8.11 (d, *J* = 6.4 Hz, ArH [Phe], 1H).

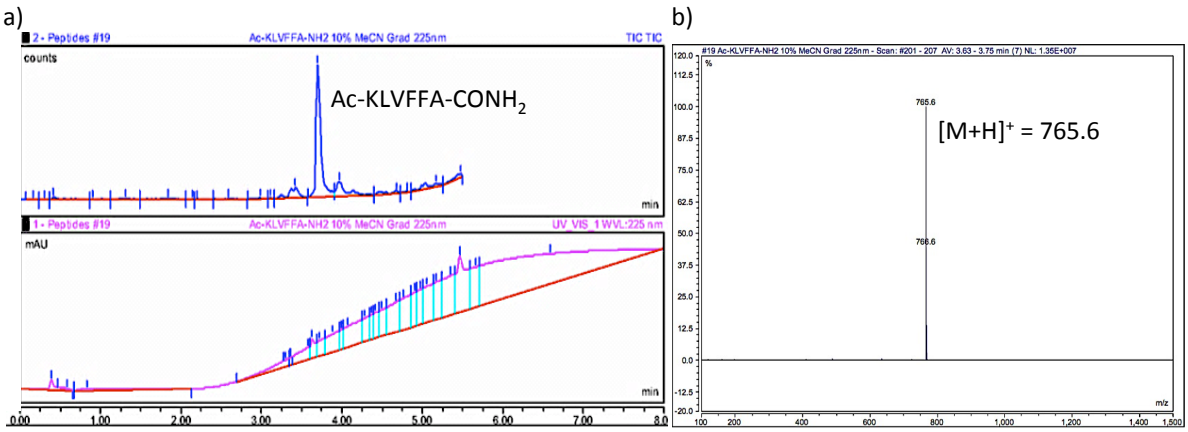
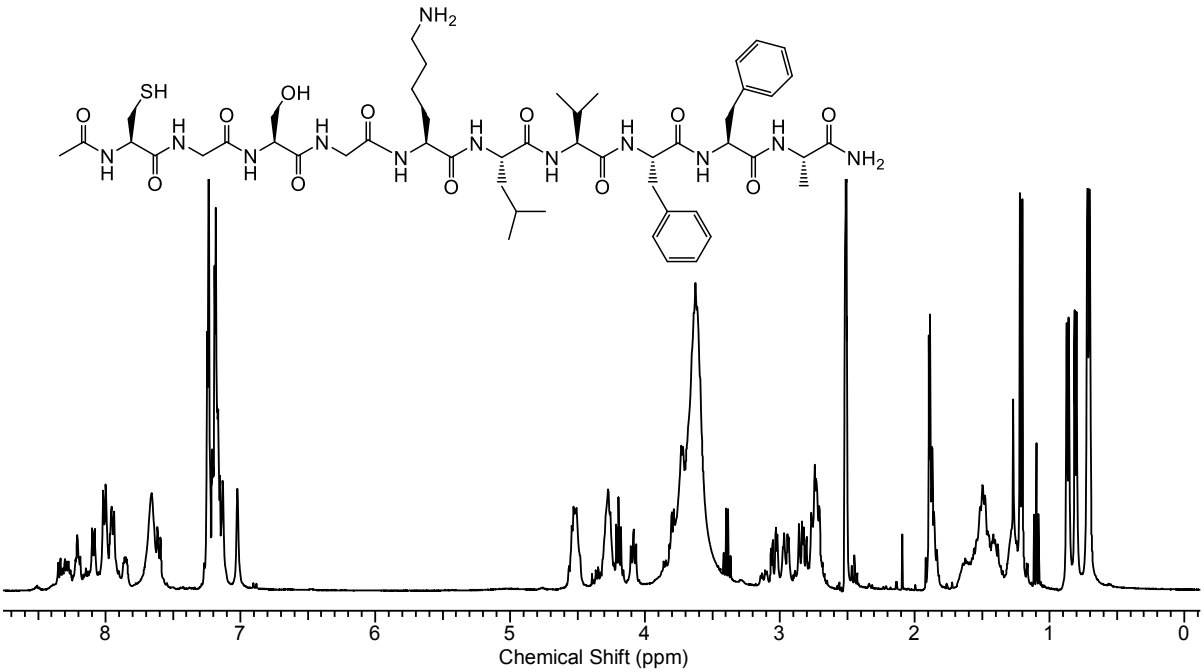
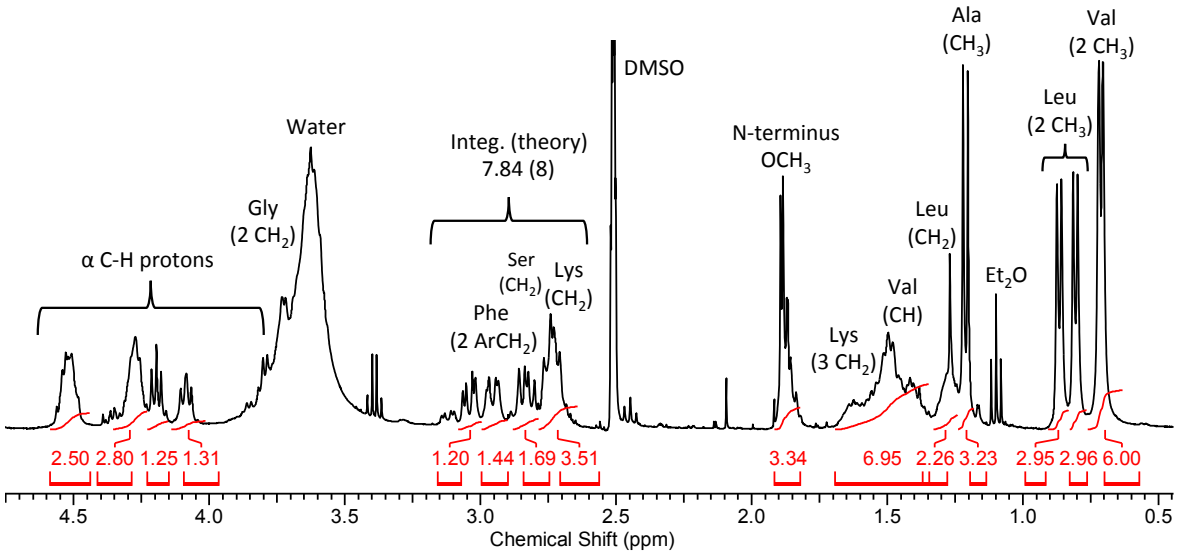


Figure 2.6. a) Total ion current (TIC) chromatography of **Ac-KLVFFA-CONH$_2$** (monitoring absorbance at 225 nm) showing that the peptide peak at 3.7 min. **b)** Mass spectrum of **Ac-KLVFFA-CONH$_2$**. LC-MS (m/z) calc. for: $[M+H]^+$ 765.5, found 765.6.

Ac-CGSGKLVFFA-CONH$_2$

The ^{1}H NMR of **Ac-CGSGKLVFFA-CONH$_2$** showed most of its protons peaks, with the proton integration being similar to theoretical values (Figure 2.7). The appearance of broad peaks could have been due to this peptide undergoing extensive self-assembly, given that it did not readily dissolve in the solvent.[17] LC-MS confirmed the mass of the desired peptide (Figure 2.8).





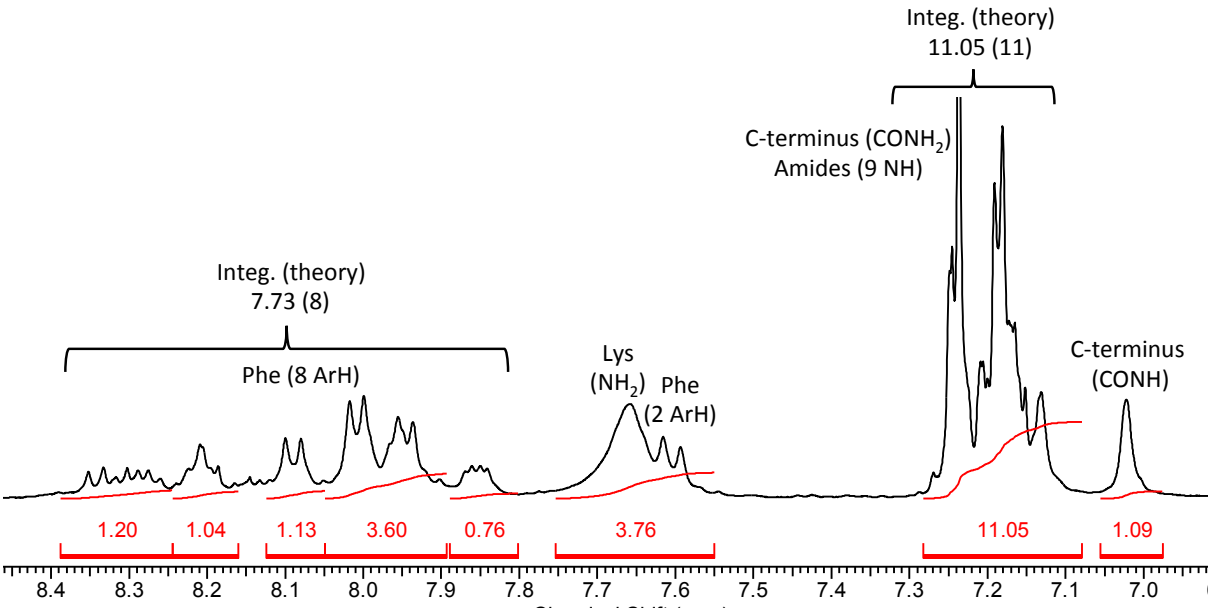


Figure 2.7. ^{1}H NMR of **Ac-CGSGKLVFFA-NH$_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.71 (d, *J* = 6.6 Hz, CH_3 [Val], 6H), 0.81 (d, *J* = 6.5 Hz, CH_3 [Leu], 3H), 0.87 (d, *J* = 6.6 Hz, CH_3 [Leu], 6H), 1.21 (d, *J* = 7.1 Hz, CH_3 [Ala], 3H), 1.24-1.33 (m, CH_2 [Ile], 2H), 1.34-1.69 (m, CH [Val] and CH_2 [Lys], 6H), 1.89 (s, OCH_3 [N-terminus], 3H), 2.67-2.78 (m, CH_2 [Lys], 2H), 2.78-2.88 (m, ArCH [Phe], 1H), 2.95 (dd, *J* = 14.5 and 4.0 Hz, ArCH [Phe], 1H), 3.04 (dd, *J* = 14.0 and 4.9 Hz, ArCH [Phe], 1H), 4.09 (t, *J* = 8.0 Hz, αC-H, 1H), 4.15-4.23 (m, αC-H, 1H), 4.24-4.34 (m, αC-H, 3H), 4.46-4.58 (m, αC-H, 3H), 7.02 (s, CONH [C-terminus], 1H), 7.11-7.28 (m, CONH [C-terminus], ArH [Phe] and NH, 11H), 7.56-7.75 (m, NH_2[Lys] and ArH [Phe], 3H), 7.91-8.04 (m, ArH [Phe], 4H), 8.09 (d, *J* = 8.0 Hz, ArH [Phe], 1H).

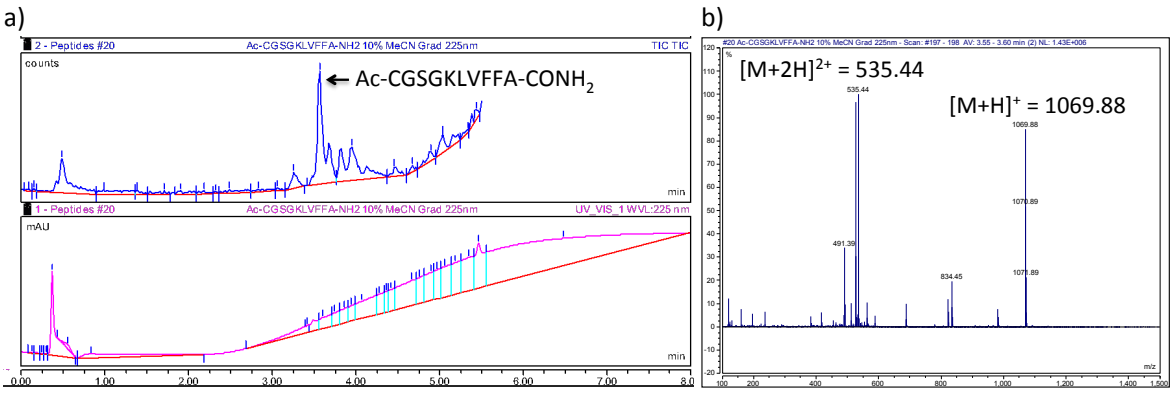
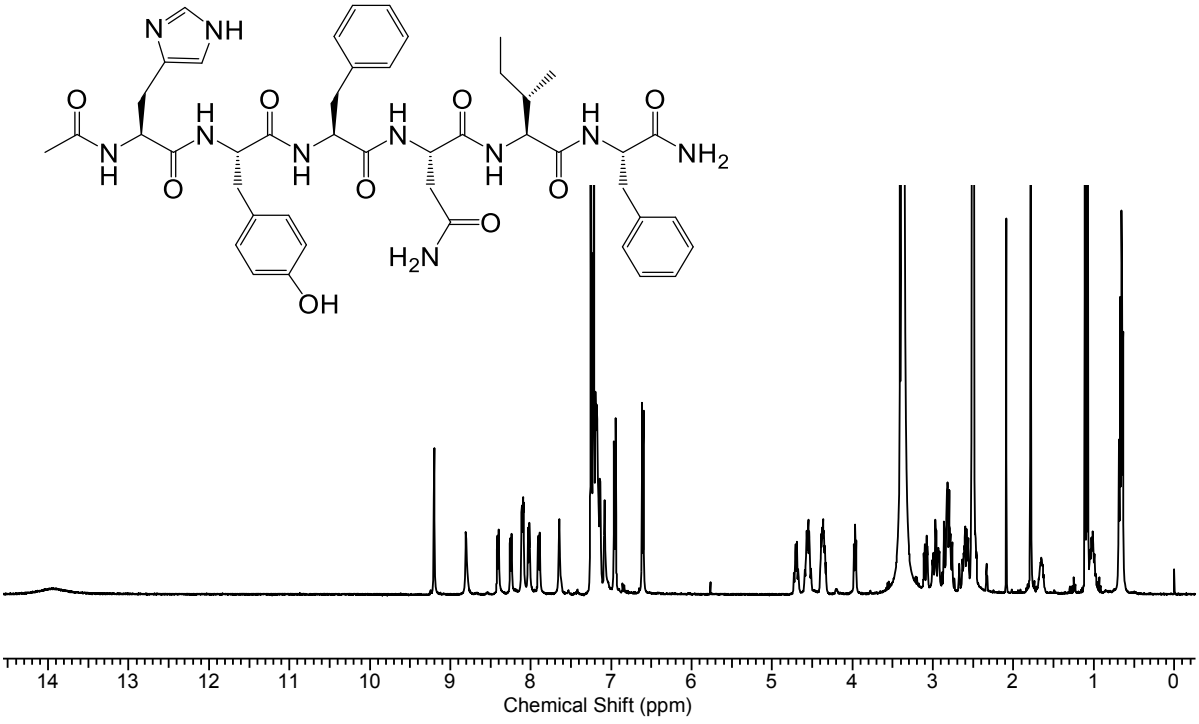
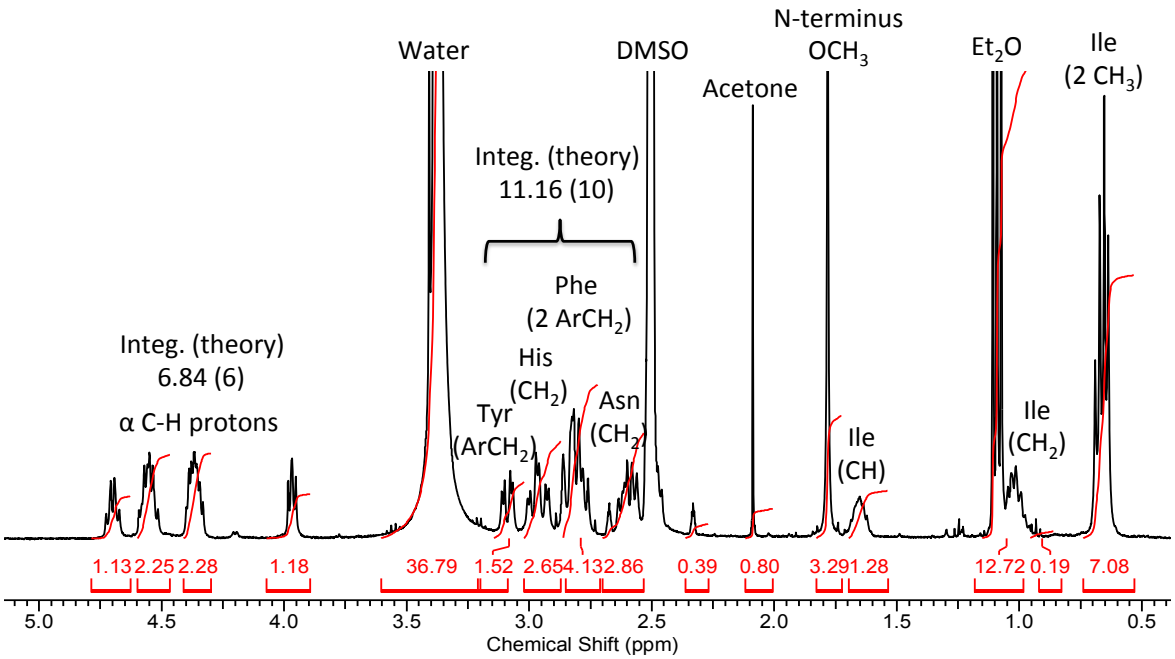


Figure 2.8. a) Total ion current (TIC) chromatography of **Ac-CGSGKLVFFA-CONH$_2$** (monitoring absorbance at 225 nm) showing that the peptide peak at 3.6 min. **b)** Mass spectrum of **Ac-CGSGKLVFFA-CONH$_2$**. LC-MS (m/z) calc. for: $[M+H]^+$ 1069.50, found 1069.88.

Ac-HYFNIF-CONH$_2$

The ^{1}H NMR showed all the protons peaks of **Ac-HYFNIF-CONH$_2$**, with the proton integration being similar to theoretical values (Figure 2.9). LC-MS confirmed the mass of the desired peptide (Figure 2.10).





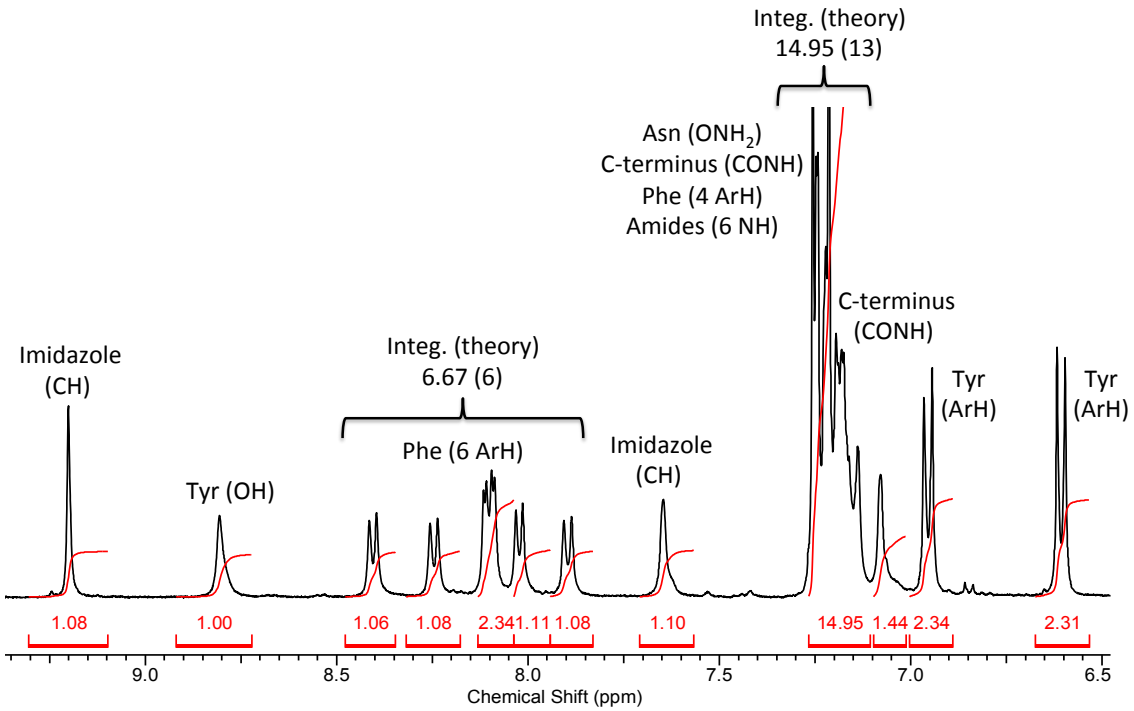


Figure 2.9. ^{1}H NMR of **Ac-HYFNIF-CONH$_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.63-0.70 (m, CH_3 [Ile], 6H), 0.97-1.06 (m, CH_2 [Ile], 2H), 1.61-1.71 (m, CH, [Ile], 1H), 1.82 (s, OCH_3 [N-terminus], 3H), 2.55-2.70 (m, OCH_2 [Asn], 2H), 2.72-2.86 (m, $ArCH_2$ [Phe], 4H), 2.97 (dd, *J* = 9.2 Hz, CH_2 [His], 2H), 3.09 (dd, *J* = 4.4 and 10.0 Hz, $ArCH_2$ [Tyr], 2H), 3.97 (t, *J* = 6.4 Hz, αC-H, 1H), 4.32-4.41 (m, αC-H, 2H), 4.50-4.60 (m, αC-H, 2H), 4.66-4.74 (m, αC-H, 1H), 6.61 (d, *J* = 8.4 Hz, ArH [Tyr], 2H), 6.96 (d, *J* = 8.4 Hz, ArH [Tyr], 2H), 7.08 (s, CONH [C-terminus], 1H), 7.14-7.27 (m, ONH_2 [Asn], CONH [C-terminus], ArH [Phe] and NH, 13 H), 7.65 (s, imidazole, 1H), 7.90 (d, *J* = 8.0 Hz, ArH [Phe], 1H), 8.02 (d, *J* = 7.1 Hz, ArH [Phe], 1H), 8.10 (dd, *J* = 4.5 and 8.7 Hz, ArH [Phe], 2H), 8.25 (d, *J* = 8.1 Hz, ArH [Phe], 1H), 8.41 (d, *J* = 7.6 Hz, ArH [Phe], 1H), 8.81 (s, OH [Tyr], 1H), 9.20 (s, imidzole, 1H), 13.94 (s, NH [His], 1H).

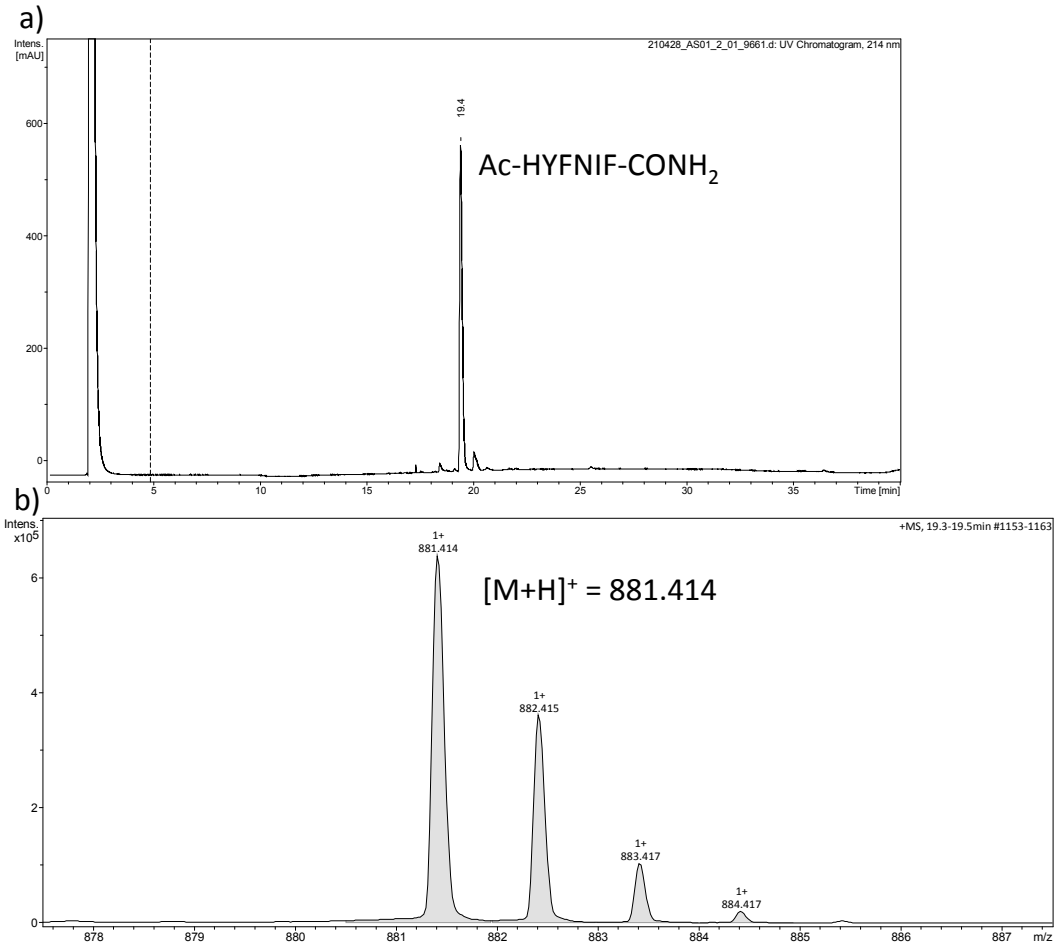
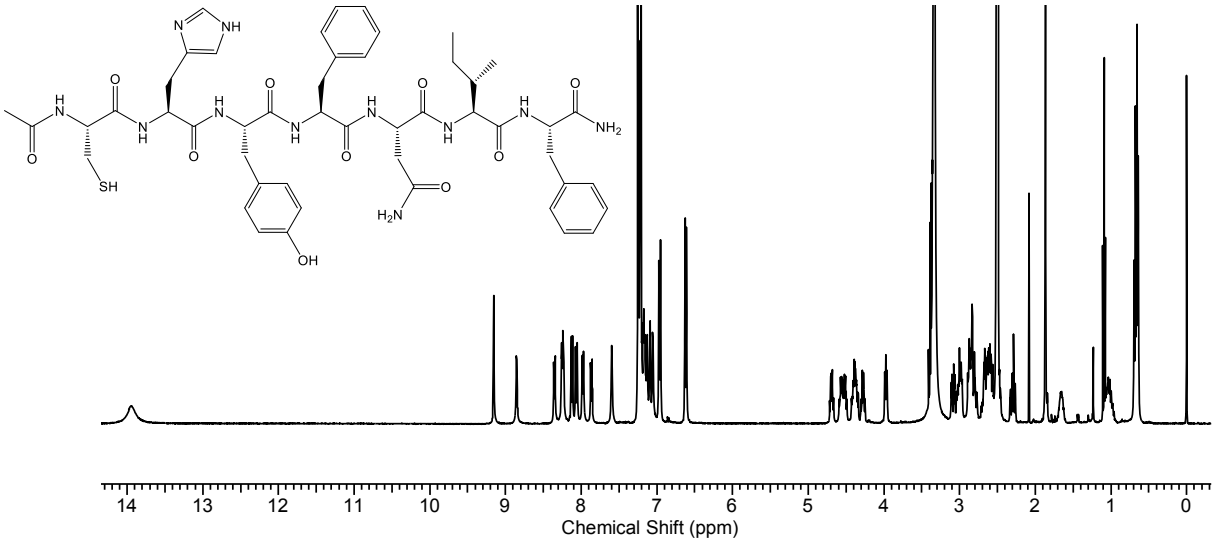
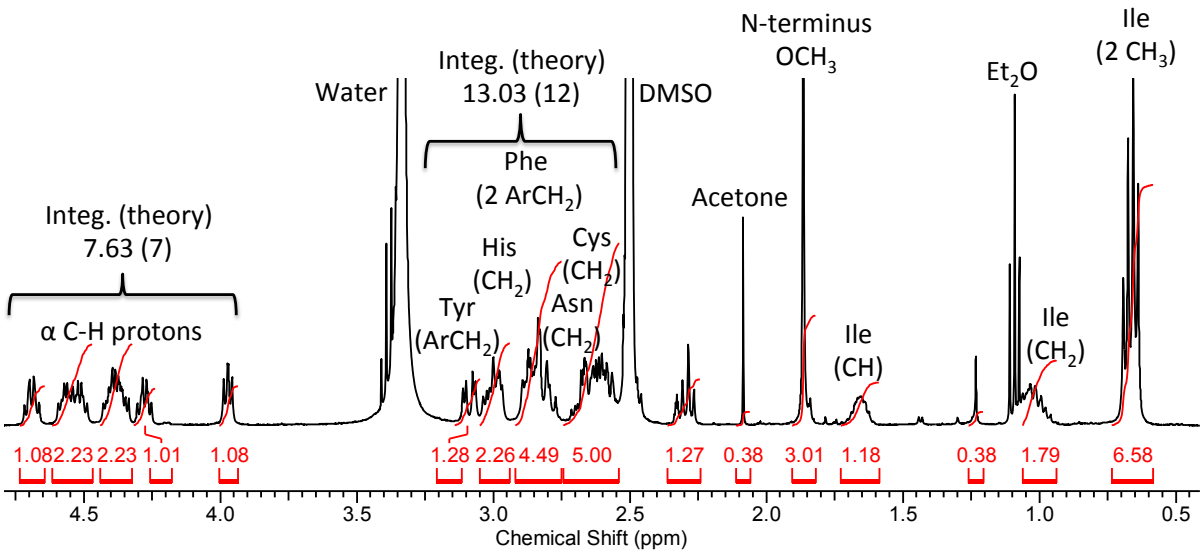


Figure 2.10. a) Analytical HPLC chromatograph of **Ac-HYFNIF-CONH$_2$** (monitoring absorbance at 214 nm) showing that the peptide peak at 19.4 min. **b)** Expanded trace of desired peptide peak. LC-MS (m/z): calc. for $[M+H]^+$ 881.4, found 881.414.

Ac-CHYFNIF-CONH$_2$

The ^{1}H NMR of **Ac-CHYFNIF-CONH$_2$** showed most of its protons peaks, with the proton integration being similar to theoretical values (Figure 2.11). LC-MS confirmed the mass of the desired peptide (Figure 2.12).





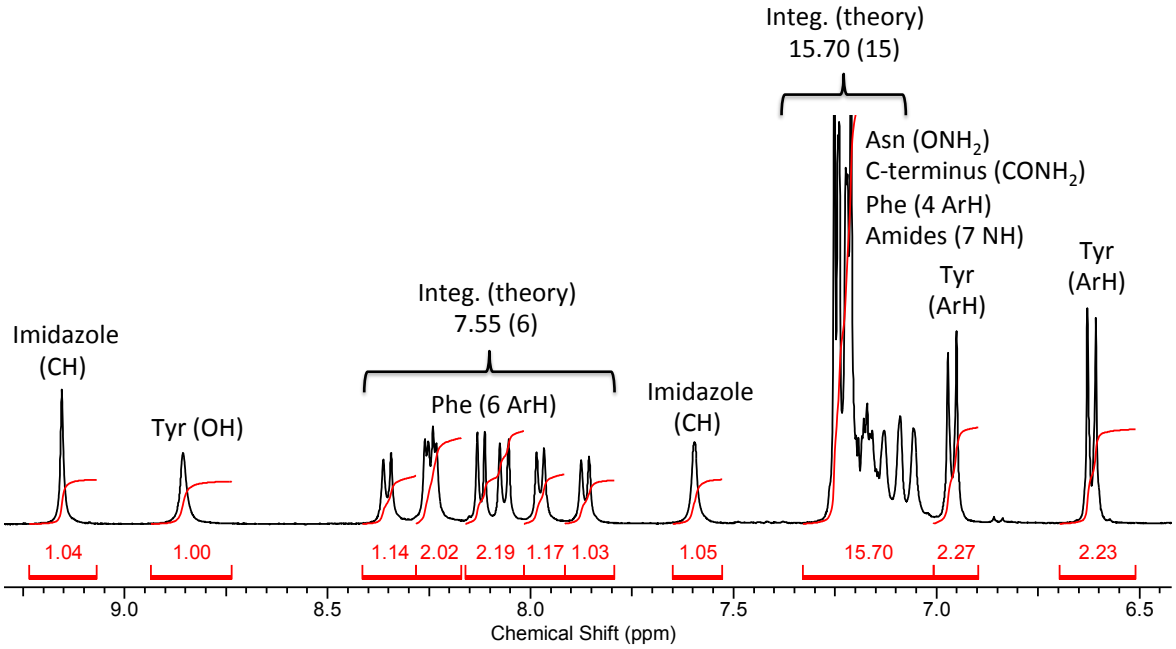


Figure 2.11. ^{1}H NMR of **Ac-CHYFNIF-CONH$_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.65 (d, *J* = 6.9 Hz, CH_3 [Ile], 3H), 0.68 (t, *J* = 7.2 Hz, CH_3 [Ile], 3H), 0.94-1.97 (m, CH_2 [Ile], 2H), 1.23 (s, SH [Cys], 1H), 1.59-1.73 (m, CH [Ile], 1H), 1.86 (s, OCH_3 [N-terminus], 3H), 2.55-2.72 (m, OCH_2 [Asn] and CH_2 [Cys], 4H), 2.75-2.92 (m, $ArCH_2$ [Phe], 4H), 2.94-3.05 (m, CH_2 [His] and ArCH [Tyr], 3H), 3.09 (dd, *J* = 5.4 and 13.6 Hz, ArCH [Tyr], 1H), 3.97 (t, *J* = 6.3 Hz, αC-H, 1H), 4.24-4.31 (m, αC-H, 1H), 4.33-4.44 (m, αC-H, 2H), 4.47-4.61 (m, αC-H, 2H), 4.65-4.73 (m, αC-H, 1H), 6.62 (d, *J* = 8.5 Hz, ArH [Tyr], 2H), 6.69 (d, *J* = 6.7 Hz, ArH [Tyr], 2H), 7.06 (s, ONH_2 [Asn], 2H), 7.09 (s, $CONH_2$ [C-terminus], 2H), 7.14−7.28 (m, ArH [Phe] and NH, 11 H), 7.60 (s, imidazole, 1H), 7.87 (d, *J* = 7.8 Hz, ArH [Phe], 1H), 7.98 (d, *J* = 7.1 Hz, ArH [Phe], 1H), 8.06 (d, *J* = 8.5 Hz, ArH [Phe], 1H), 8.12 (d, *J* = 7.5 Hz, ArH [Phe], 1H), 8.25 (dd, *J* = 8.3 and 4.5 Hz, ArH [Phe], 1H), 8.35 (d, *J* = 7.6 Hz, ArH [Phe], 1H), 8.85 (s, OH [Tyr], 1H), 9.15 (s, imidzole, 1H), 13.94 (s, NH [His], 1H).

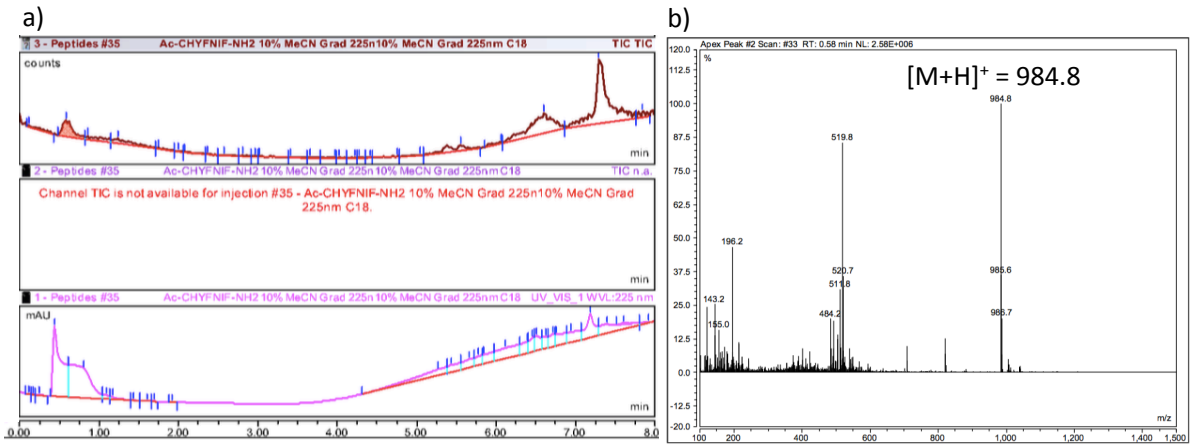
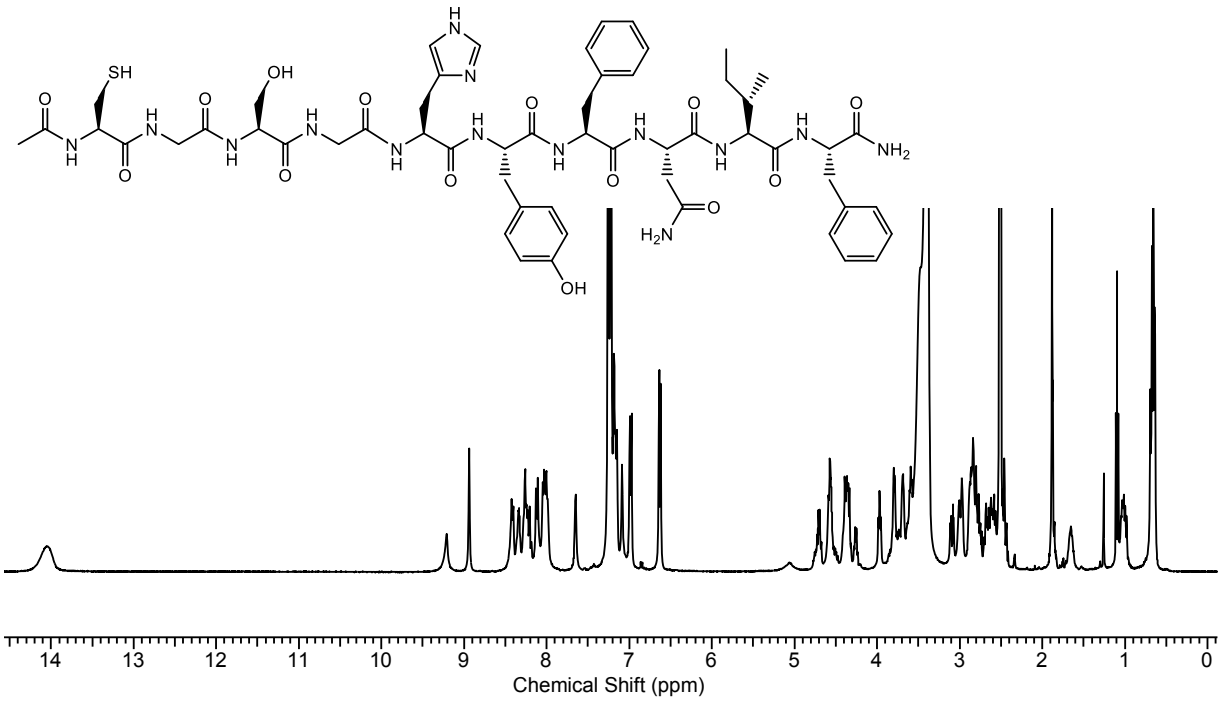


Figure 2.12. a) Total ion current (TIC) chromatography of **Ac-CHYFNIF-CONH$_2$** (monitoring absorbance at 225 nm) showing that the peptide peak at 0.58 min. **b)** Mass spectrum of **Ac-CHYFNIF-CONH$_2$**. LC-MS (m/z): calc. for $[M+H]^+$ 984.4, found 984.8.

Ac-CGSGHYFNIF-CONH$_2$

The ^{1}H NMR of **Ac-CGSGHYFNIF-CONH$_2$** showed most of its protons peaks, with the proton integration being similar to theoretical values (Figure 2.13). LC-MS confirmed the mass of the desired peptide (Figure 2.14).



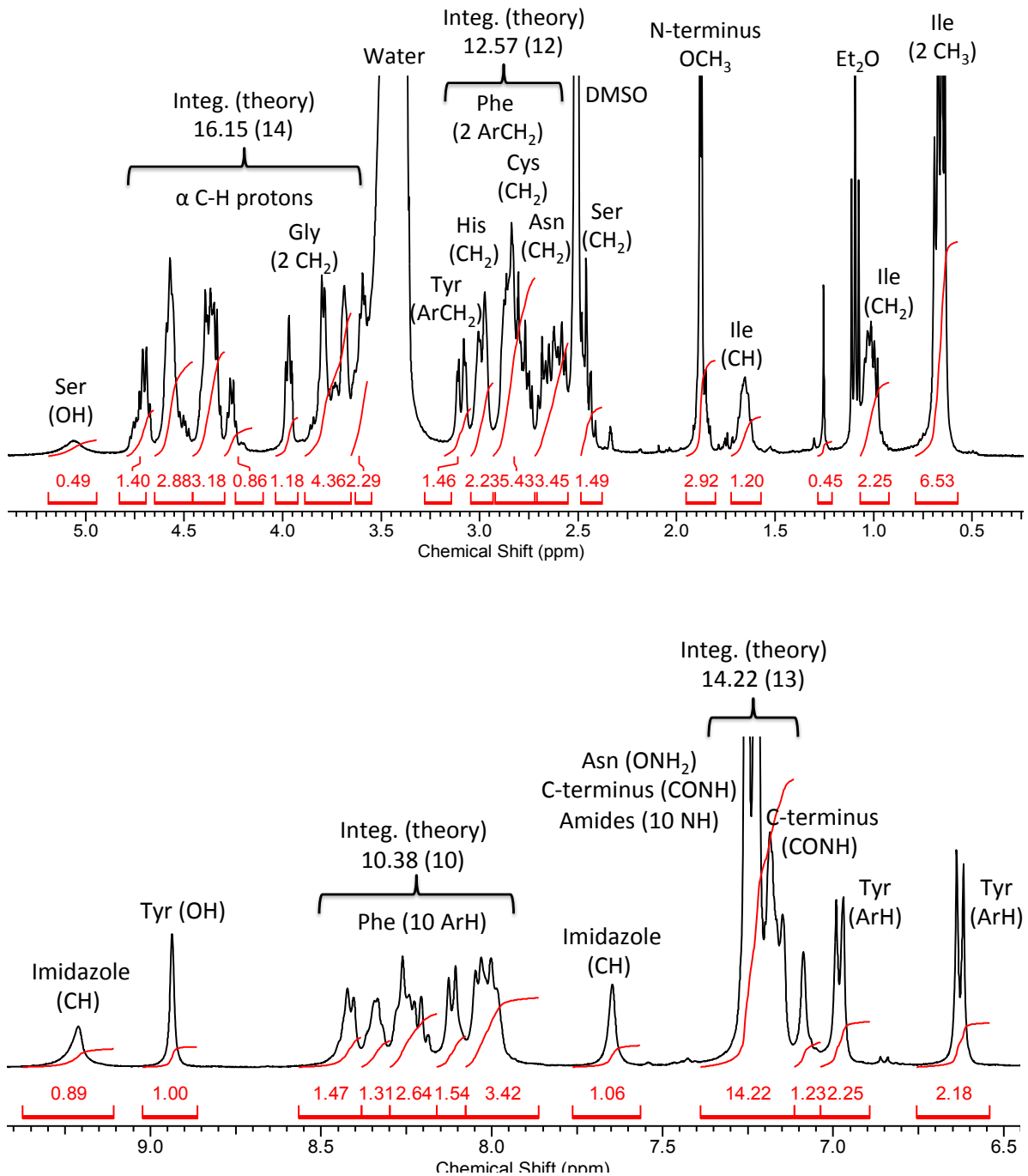


Figure 2.13. ^{1}H NMR of **Ac-CGSGHYFNIF-CONH$_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.61-0.72 (m, CH_3 [Ile], 6H), 0.95-1.06 (m, CH_2 [Ile], 2H), 1.60-1.71 (m, CH [Ile], 1H), 1.88 (s, OCH_3 [N-terminus], 3H), 2.40-2.48 (m, CH_2 [Ser], 2H), 2.55-2.64 (m, OCH_2 [Asn], 2H), 2.64-2.71 (m, CH_2 [Cys], 2H), 2.72-2.91 (m, $ArCH_2$ [Phe], 4H), 2.99 (d, *J* = 11.2 Hz, CH_2 [His], 2H), 3.09 (dd, *J* = 3.9 and 14.2 Hz, $ArCH_2$ [Tyr], 2H), 3.55-3.65 (m, αC-H, 2H), 3.65-3.84 (m, CH_2 [Gly], 4H), 3.97 (t, *J* = 6.4 Hz, αC-H, 1H), 4.22-4.29 (m, αC-H, 1H), 4.30-4.45 (m, αC-H, 3H), 4.51-4.65 (m, αC-H, 2H), 4.66-4.78 (m, αC-H, 1H), 5.06 (s, OH

[Ser], 2H), 6.63 (d, *J* = 8.4 Hz, ArH [Tyr], 2H), 6.98 (d, *J* = 8.4 Hz, ArH [Tyr], 2H), 7.09 (s, CONH [C-terminus], 1H), 7.14-7.29 (m, ONH_2 [Asn], CONH [C-terminus], NH, 13H), 7.65 (s, imidazole, 1H), 7.95-8.07 (m, ArH [Phe], 2H), 8.12 (d, *J* = 8.4 Hz, ArH [Phe], 1H), 8.17-8.30 (m, ArH [Phe], 1H), 8.31-8.38 (m, ArH [Phe], 1H), 8.41 (d, *J* = 7.6 Hz, ArH [Phe], 1H), 8.94 (s, OH [Tyr], 1H), 9.21 (s, imidzole, 1H), 14.05 (s, NH [His], 1H).

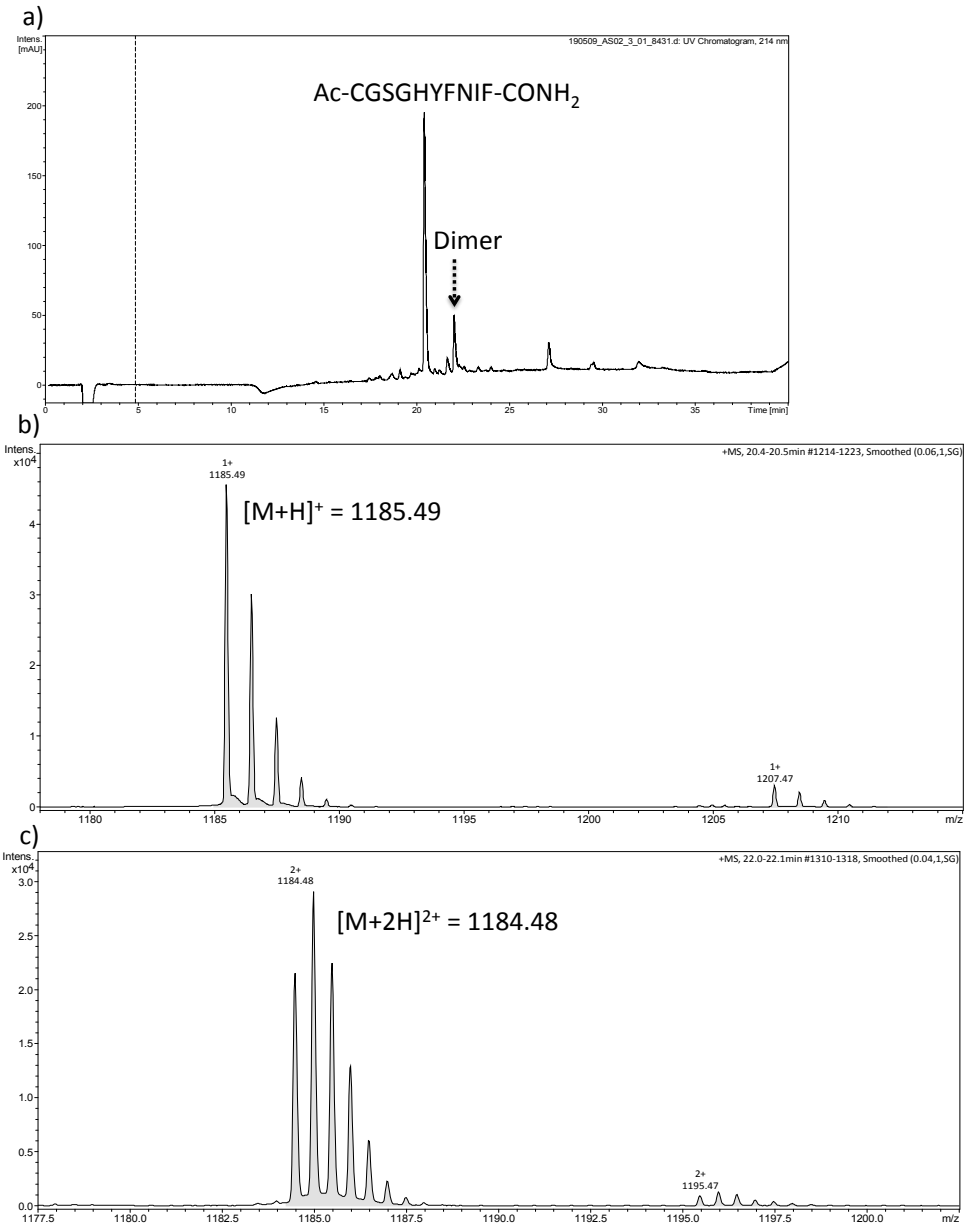
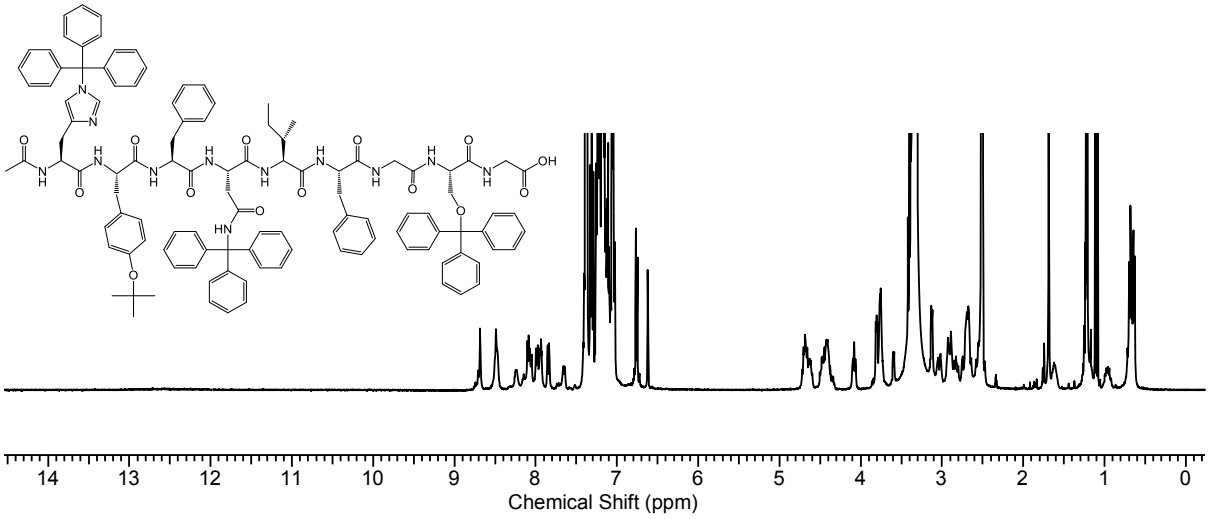


Figure 2.14. a) Analytical HPLC chromatograph of **Ac-CGSGHYFNIF-CONH$_2$** (monitoring absorbance at 214 nm) showing that the peptide peak at 20.5 min and the dimer peak at 22 min. Expanded trace of **b)** desired peptide peak and **c)** dimer peak. LC-MS (m/z) calc. for: $[M+H]^+$ 1185.49, found 1185.48.

Ac-H(Trt)-Y(tBu)-F-N(Trt)-F-GS(Trt)-G-COOH

The ^{1}H NMR showed most of the proton peaks of the desired peptide, with the proton integration being similar to theoretical values (Figure 2.15). The disappearance of the hydroxyl proton signal of serine at ~5 ppm and the amine proton signal of imidazole at ~14 ppm were confirmed, suggesting that those peptide side-chains remained protected following the cleavage of peptide from the resin. However, additional peaks were also observed at 1.75, 3.60, 7.65 and 8.24 ppm, indicating the presence of by-products.

LC-MS confirmed the mass of the desired peptide, in addition to showing some impurity peaks (Figure 2.16 a). The loss of a trityl group from the peptide was suggested given that there was a mass difference of 242 between the two peptide peaks detected at 34.0 min (Figure 2.16 b).



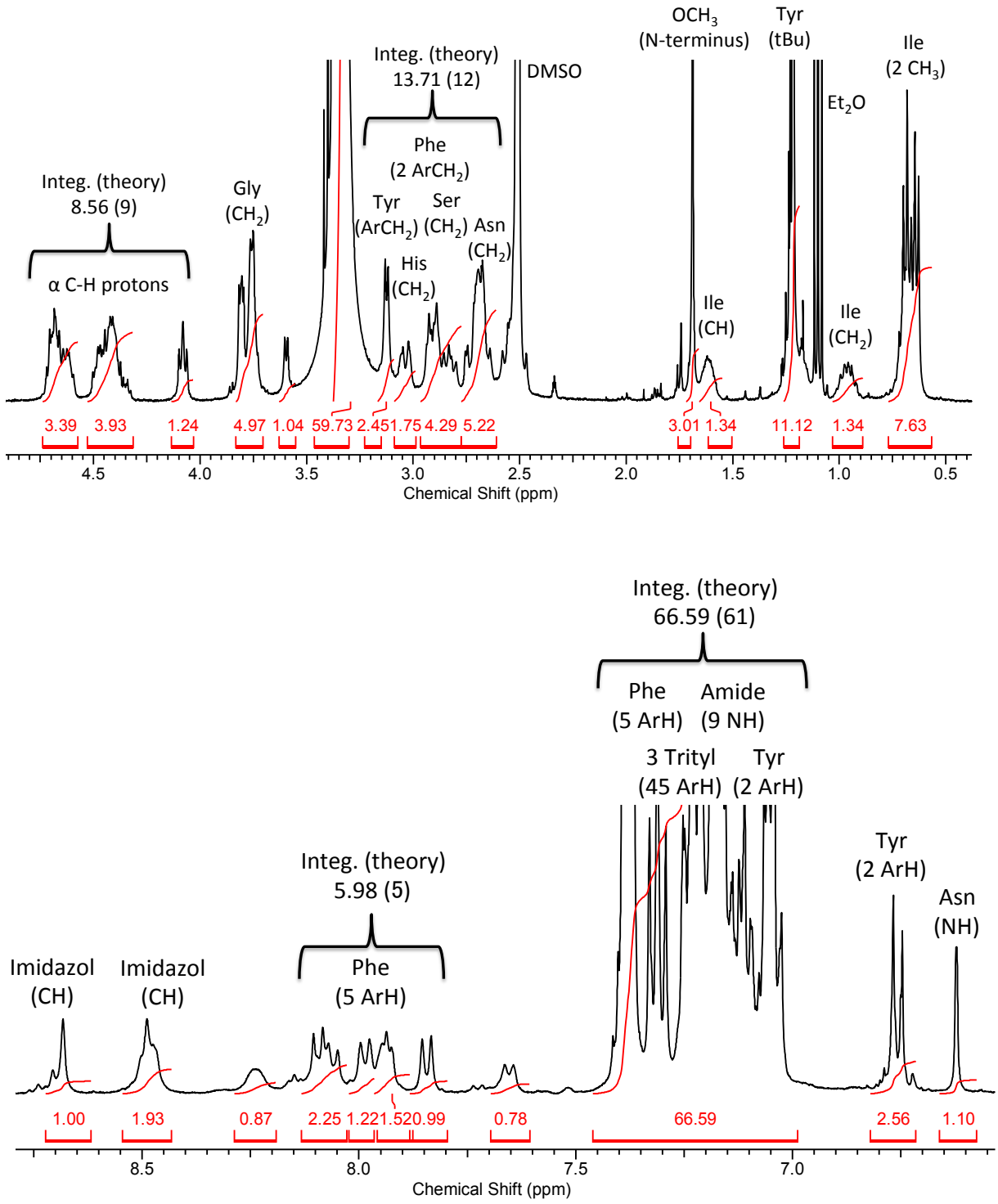


Figure 2.15. ^{1}H NMR of **Ac-H(Trt)-Y(tBu)-F-N(Trt)-F-GS(Trt)-G-COOH**.

^{1}H NMR (400 MHz, DMSO) δ 0.64 (d, *J* = 6.8, CH_3 [Ile], 3H), 0.68 (t, *J* = 7.4, CH_2 [Ile], 3H), 1.22 (s, tert-butyl [Tyr], 9H), 1.55-1.66 (m, CH [Ile], 1H), 1.69 (s, OCH_3 [N-terminus], 3H), 2.62-2.77 (m, OCH_2 [Asn] and CH_2 [Ser], 4H), 2.78-2.97 (m, $ArCH_2$ [Phe], 2H), 3.04 (d, *J* = 8.0 Hz, CH_2 [His], 2H), 3.13 (d, *J* = 5.9 Hz, $ArCH_2$ [Tyr], 2H), 3.71-3.83 (m, CH_2 [Gly], 4H), 4.08 (t, *J*

= 7.0 Hz, αC-H, 1H), 4.32-4.52 (m, αC-H, 4H), 4.58-4.74 (m, αC-H, 4H), 6.62 (s, ONH [Asn], 1H), 6.76 (d, *J* = 8.4 Hz, ArH [Tyr], 2H), 7.01-7.42 (m, ArH [Tyr], NH, Trityl and ArH [Phe], 61H), 7.84 (d, *J* = 8.2 Hz, ArH [Phe], 1H), 7.90-7.96 (m, ArH [Phe], 1H), 7.96-8.01 (m, d, *J* = 8.2 Hz, ArH [Phe], 1H), 8.03-8.13 (m, ArH [Phe], 2H), 8.49 (s, imidzole, 1H), 8.68 (s, imidzole, 1H).

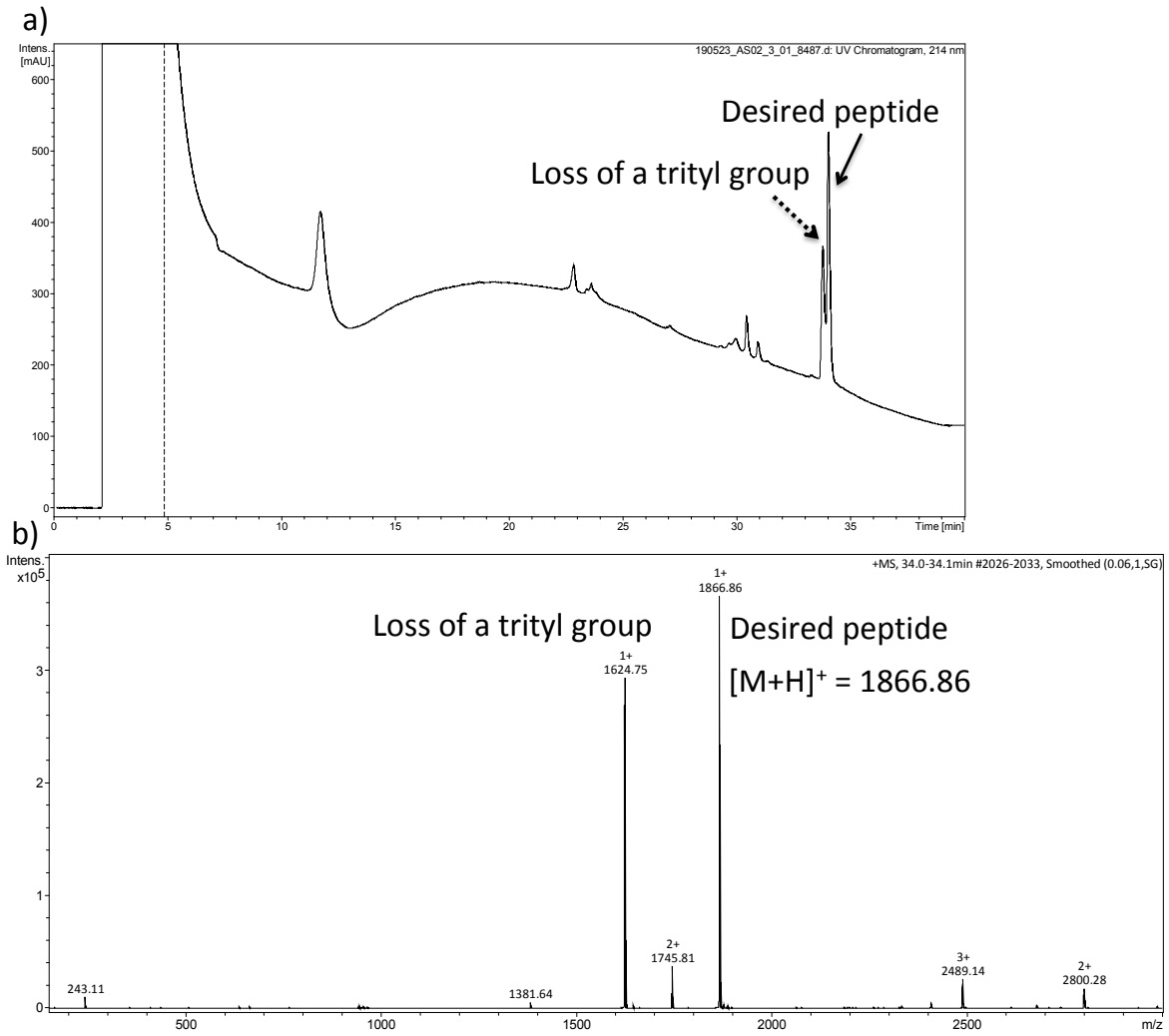
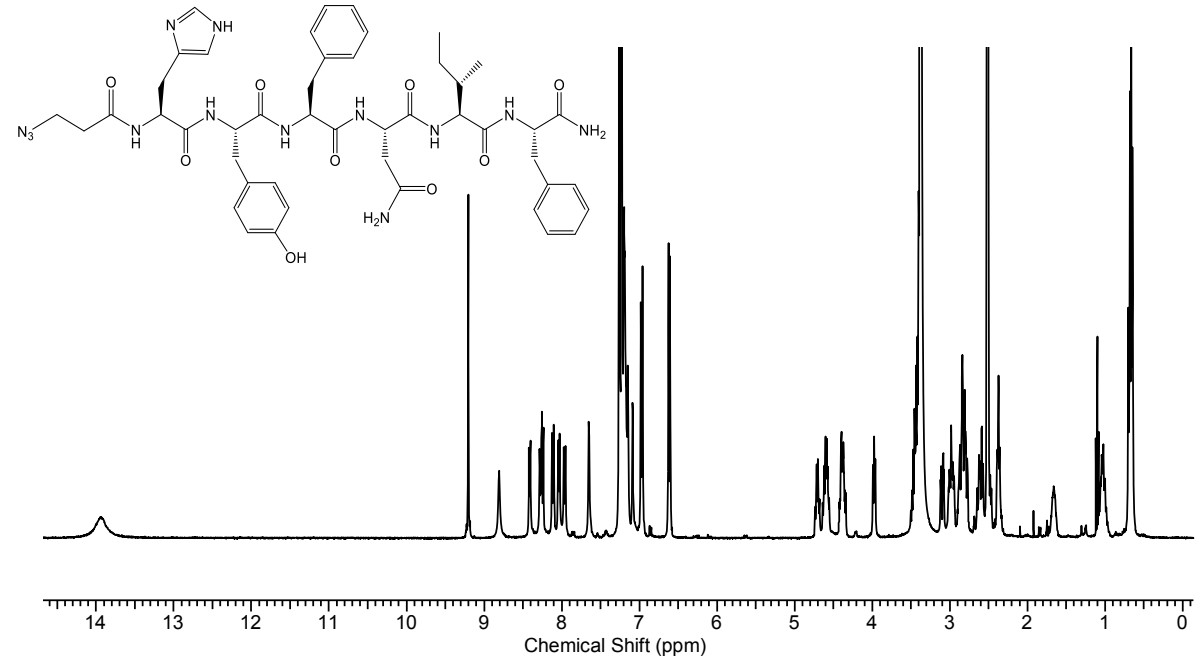
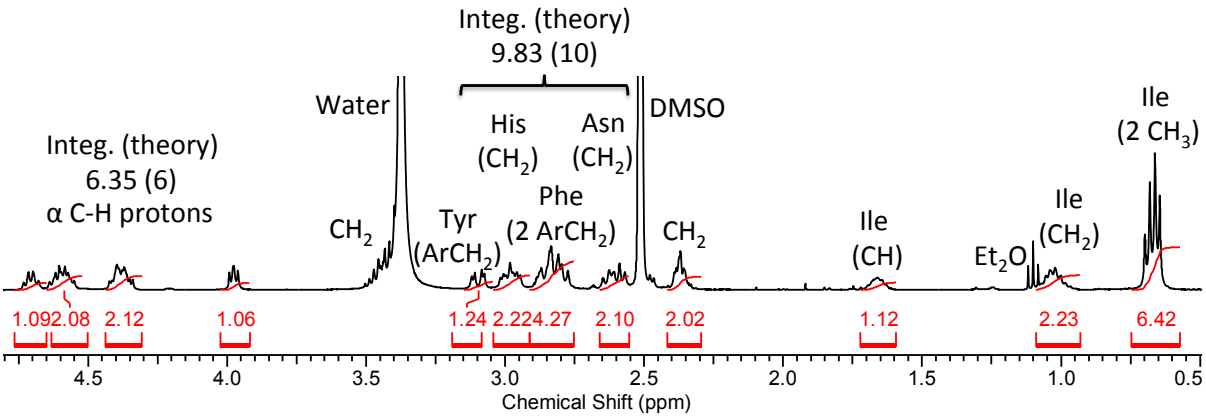


Figure 2.16. a) Analytical HPLC chromatograph of **Ac-H(Trt)-Y(tBu)-F-N(Trt)-F-GS(Trt)-G-COOH** (monitoring absorbance at 214 nm) showing that the peptide peak at 34.0 min. **b)** Expanded trace of desired peptide peak. LC-MS (m/z) calc. for: $[M+H]^+$ 1866.66, found 1866.86.

N_3-$(CH_2)_2$-HYFNIF-$CONH_2$

The ^{1}H NMR showed all the protons peaks of **N_3-$(CH_2)_2$-HYFNIF-$CONH_2$**, with the proton integration being similar to theoretical values (Figure 2.17). LC-MS confirmed the mass of the desired peptide (Figure 2.18).





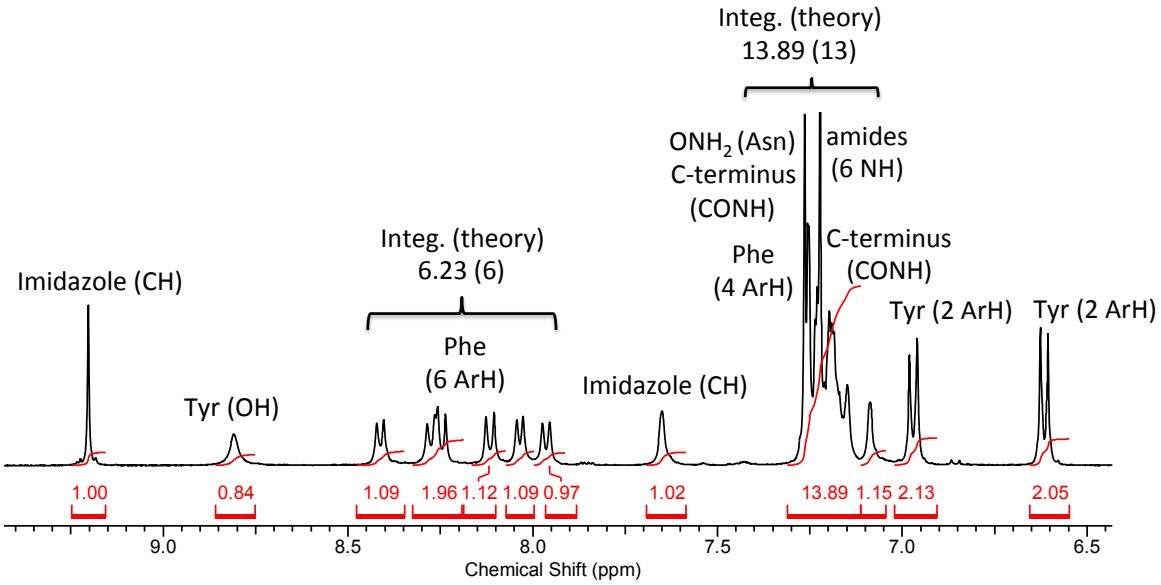


Figure 2.17. ^{1}H NMR of **N_3-$(CH_2)_2$-HYFNIF-$CONH_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.63-0.71 (m, CH_3 [Ile], 6H), 0.95-1.08 (m, CH_2 [Ile], 2H), 1.61-1.71 (m, [Ile], 1H), 2.32-2.41 (t, N_3CH_2, 3H), 2.55-2.66 (m, OCH_2 [Asn], 2H), 2.76-2.90 (m, $ArCH_2$ [Phe], 4H), 2.93-3.03 (m, CH_2 [His], 2H), 3.10 (dd, *J* = 4.4 and 6.4 Hz, $ArCH_2$ [Tyr], 2H), 3.41−3.50 (m, CH_2, 2H), 3.98 (t, *J* = 6.4 Hz, αC-H, 1H), 4.33-4.44 (m, αC-H, 2H), 4.53-4.65 (m, αC-H, 2H), 4.67-4.75 (m, αC-H, 1H), 6.62 (d, *J* = 8.4 Hz, ArH [Tyr], 2H), 6.97 (d, *J* = 8.4 Hz, ArH [Tyr], 2H), 7.09 (s, CONH [C-terminus], 1H), 7.15-7.26 (m, ONH_2 [Asn], CONH [C-terminus], ArH [Phe] and NH, 13H), 7.65 (s, imidazole, 1H), 7.96 (d, *J* = 8.0 Hz, ArH [Phe], 1H), 8.04 (d, *J* = 7.0 Hz, ArH [Phe], 1H), 8.12 (d, *J* = 8.5 Hz, ArH [Phe], 1H), 8.25 (d, *J* = 8.7 Hz, ArH [Phe], 1H), 8.28 (d, *J* = 8.4 Hz, ArH [Phe], 1H), 8.41 (d, *J* = 6.5 Hz, ArH [Phe], 1H), 8.81 (s, OH [Tyr], 1H), 9.20 (s, imidzole, 1H), 13.94 (s, NH [His], 1H).

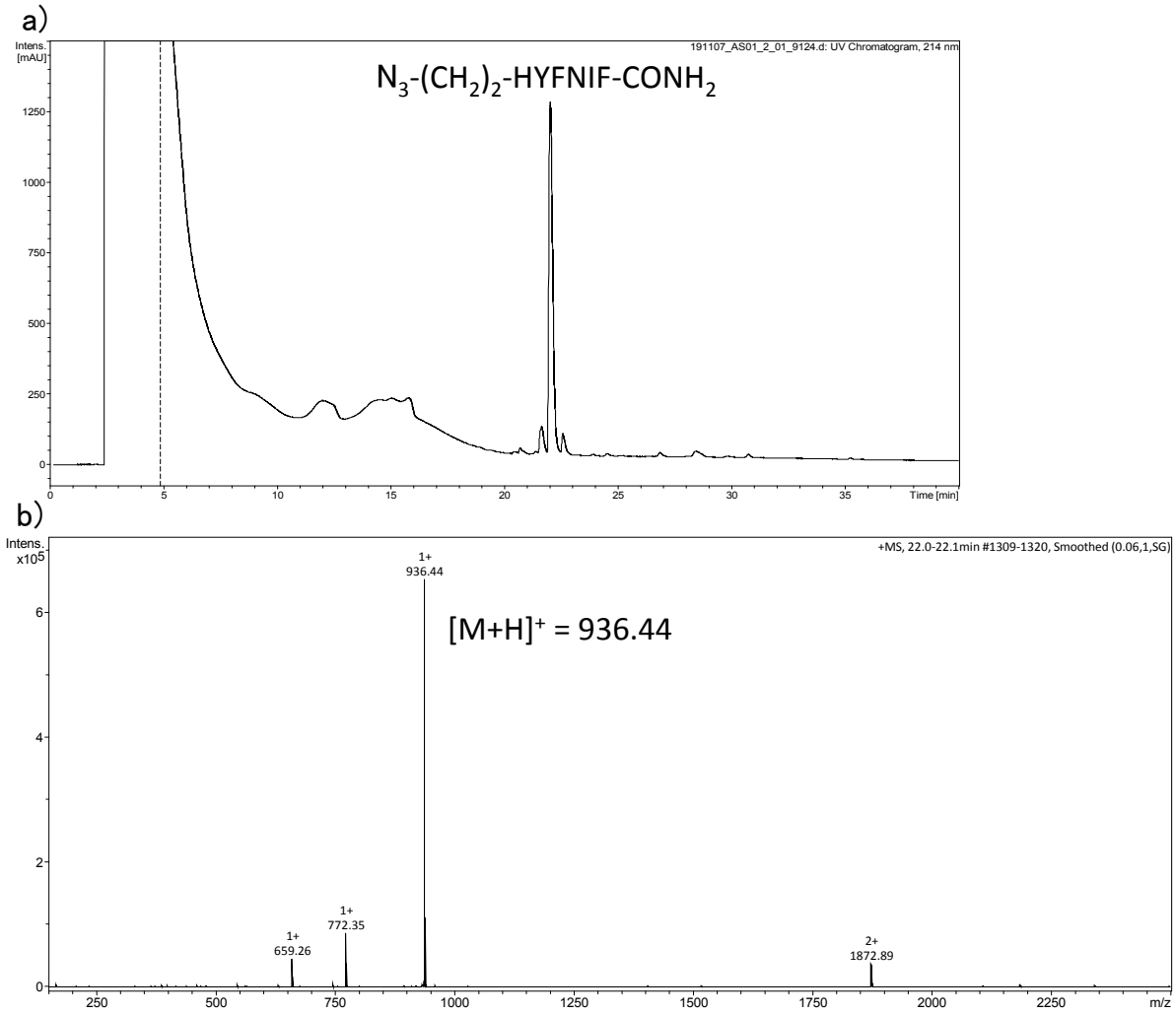
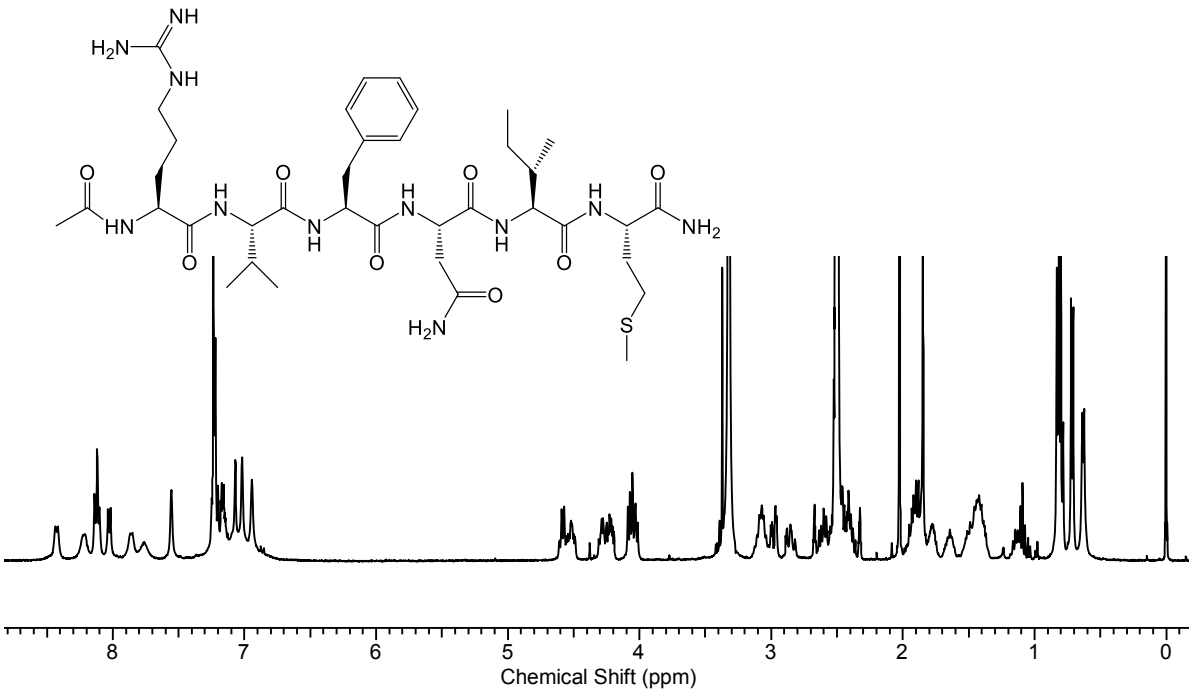
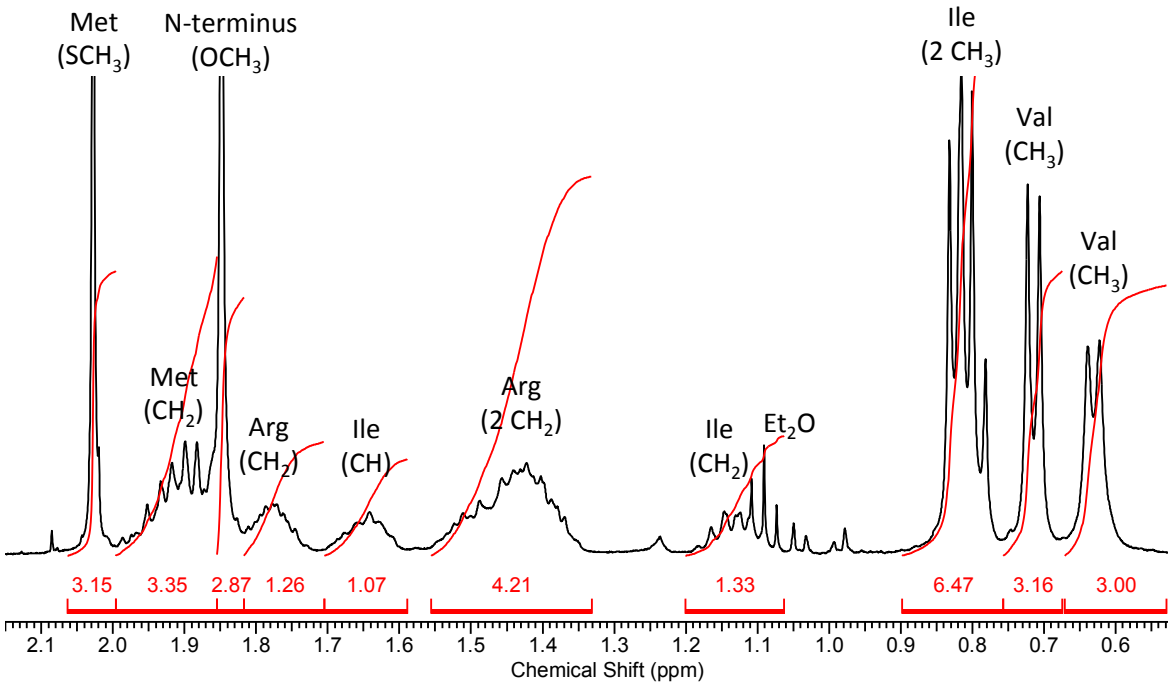


Figure 2.18. a) Analytical HPLC chromatograph of **N_3-$(CH_2)_2$-HYFNIF-$CONH_2$** (monitoring absorbance at 214 nm) showing that the peptide eluted at 22 min. **b)** Expanded trace of the peptide peak. LC-MS (m/z) calc. for: $[M+H]^+$ 936.44, found 936.44.

Ac-RVFNIM-CONH$_2$

The ^{1}H NMR showed all the protons peaks of **Ac-RVFNIM-CONH$_2$**, with the proton integration being similar to theoretical values (Figure 2.19). LC-MS confirmed the mass of the desired peptide, as well as showing an additional peak that indicated the presence of a by-product (Figure 2.20).





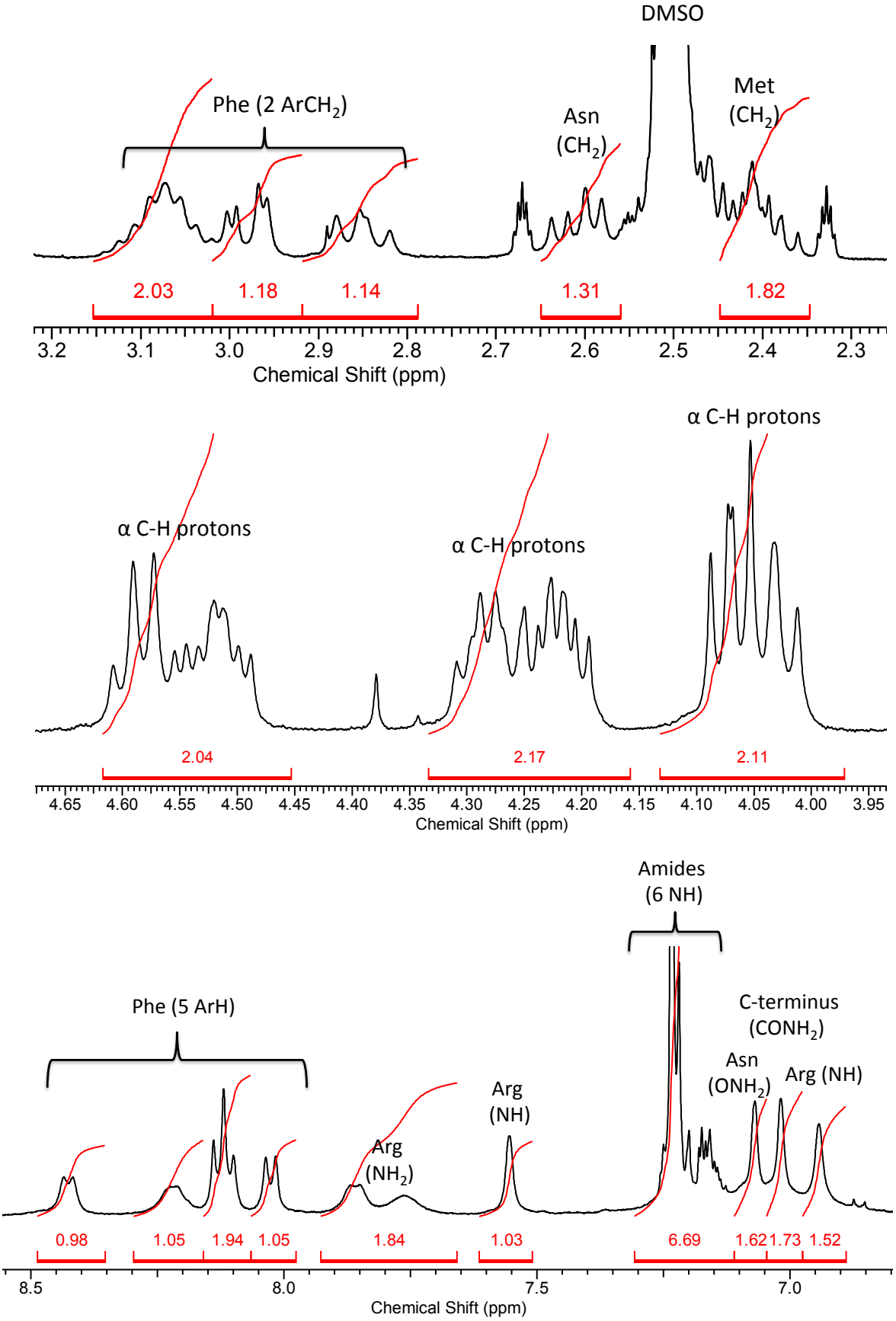


Figure 2.19. ^{1}H NMR of **Ac-RVFNIM-CONH$_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.63 (d, *J* = 6.8 Hz, CH_3 [Val], 3H), 0.71 (d, *J* = 6.8 Hz, CH_3 [Val], 3H), 0.77-0.85 (m, CH_3 [Ile], 6H), 1.10-1.18 (m, CH_2 [Ile], 2H), 1.34-1.56 (m, CH_2 [Arg], 4H), 1.60-1.70 (m, CH [Ile], 1H), 1.72-1.82 (m, CH_2 [Arg], 2H), 1.85 (s, OCH_3 [N-terminus], 3H), 1.87-1.98 (m, CH_2 [Met], 2H), 2.02 (s, SCH_3 [Met], 3H), 2.35-2.45 (m, CH_2 [Met], 2H), 2.57-2.65 (m, OCH_2 [Asn], 2H), 2.80-2.90 (m, $ArCH_2$ [Phe], 1H), 2.98 (dd, $ArCH_2$ [Phe], 1H), 3.03-3.14 (m, $ArCH_2$ [Phe], 2H), 4.00-4.10 (m, αC-H, 2H), 4.18-4.32 (m, αC-H, 2H), 4.48-4.62 (m, αC-H, 2H), 6.94 (s, NH [Arg], 1H), 7.02 (s, $CONH_2$ [C-terminus], 2H), 7.07 (s, ONH_2 [Asn], 2H), 7.13-7.26 (m, NH, 6H), 7.55-7.92 (m, NH_2 [Arg], 2H), 8.03 (d, *J* = 8.0 Hz, ArH [Phe], 1H), 8.12 (t, *J* = 8.0 Hz, ArH [Phe], 2H), 8.17-8.28 (m, ArH [Phe], 1H), 8.43 (d, *J* = 8.0 Hz, ArH [Phe], 1H).

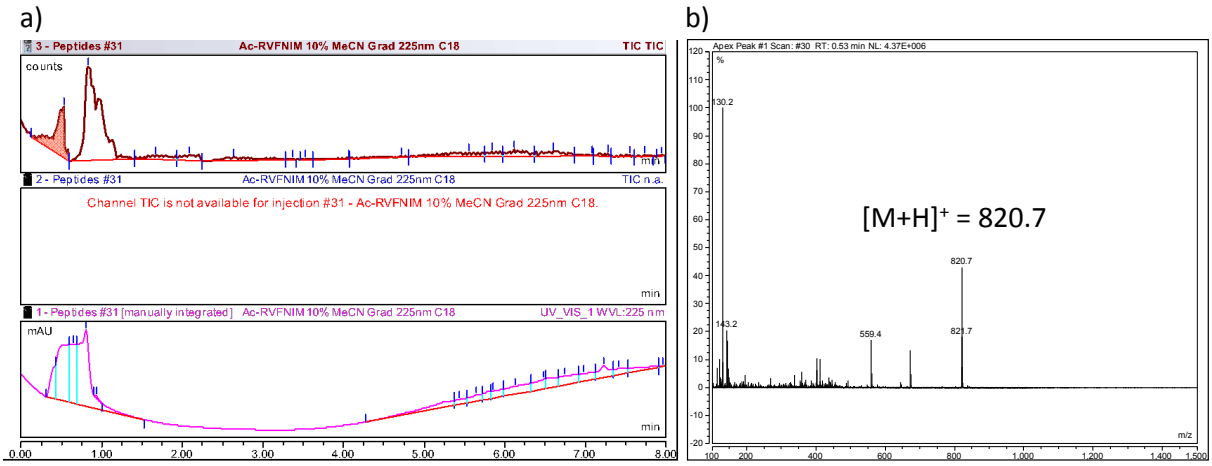
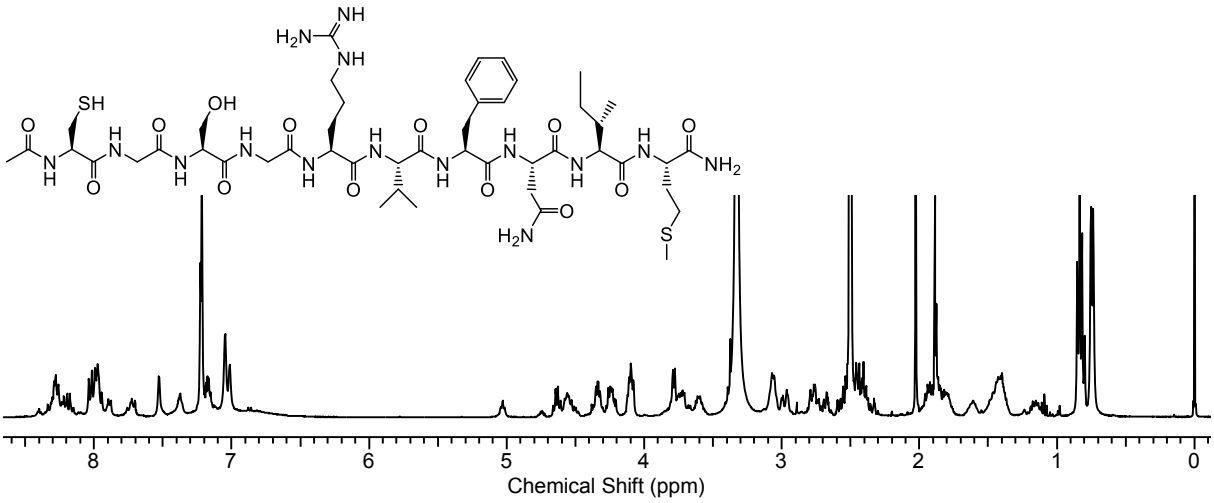
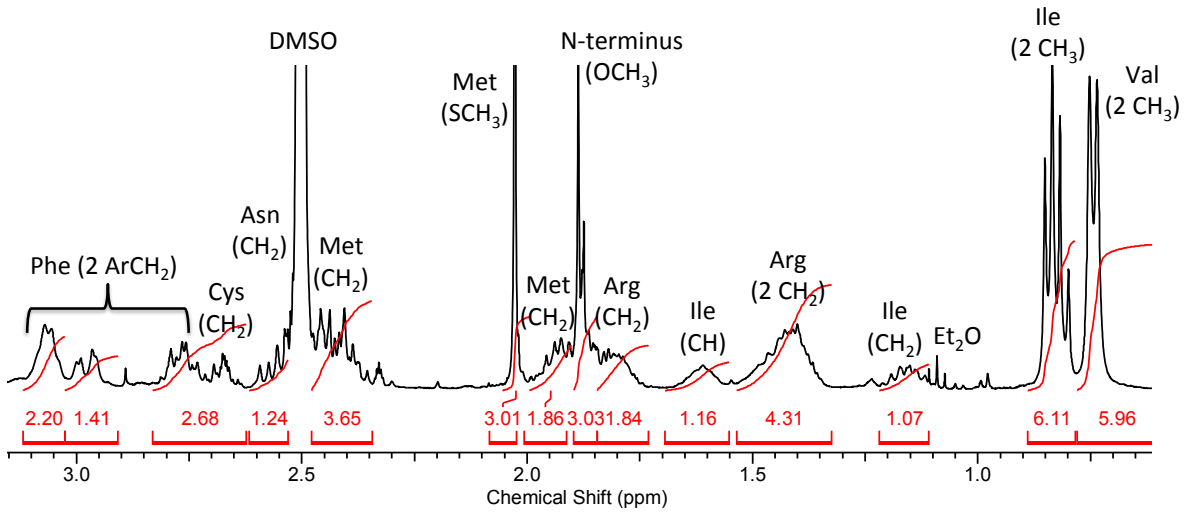


Figure 2.20. a) Total ion current (TIC) chromatography of **Ac-RVFNIM-CONH$_2$** (monitoring absorbance at 225 nm) showing that the peptide peak at 0.53 min. **b)** Mass spectrum of **Ac-RVFNIM-CONH$_2$**. LC-MS (m/z): calc. for $[M+H]^+$ 820.4, found 820.7.

Ac-CGSGRVFNIM-CONH$_2$

The ^{1}H NMR of **Ac-CGSGRVFNIM-CONH$_2$** showed most of the protons peaks, with the proton integration being similar to theoretical values except amide protons (Figure 2.21). LC-MS confirmed the mass of the desired peptide, as well as showing additional peaks that indicated the presence of by-products (Figure 2.22).





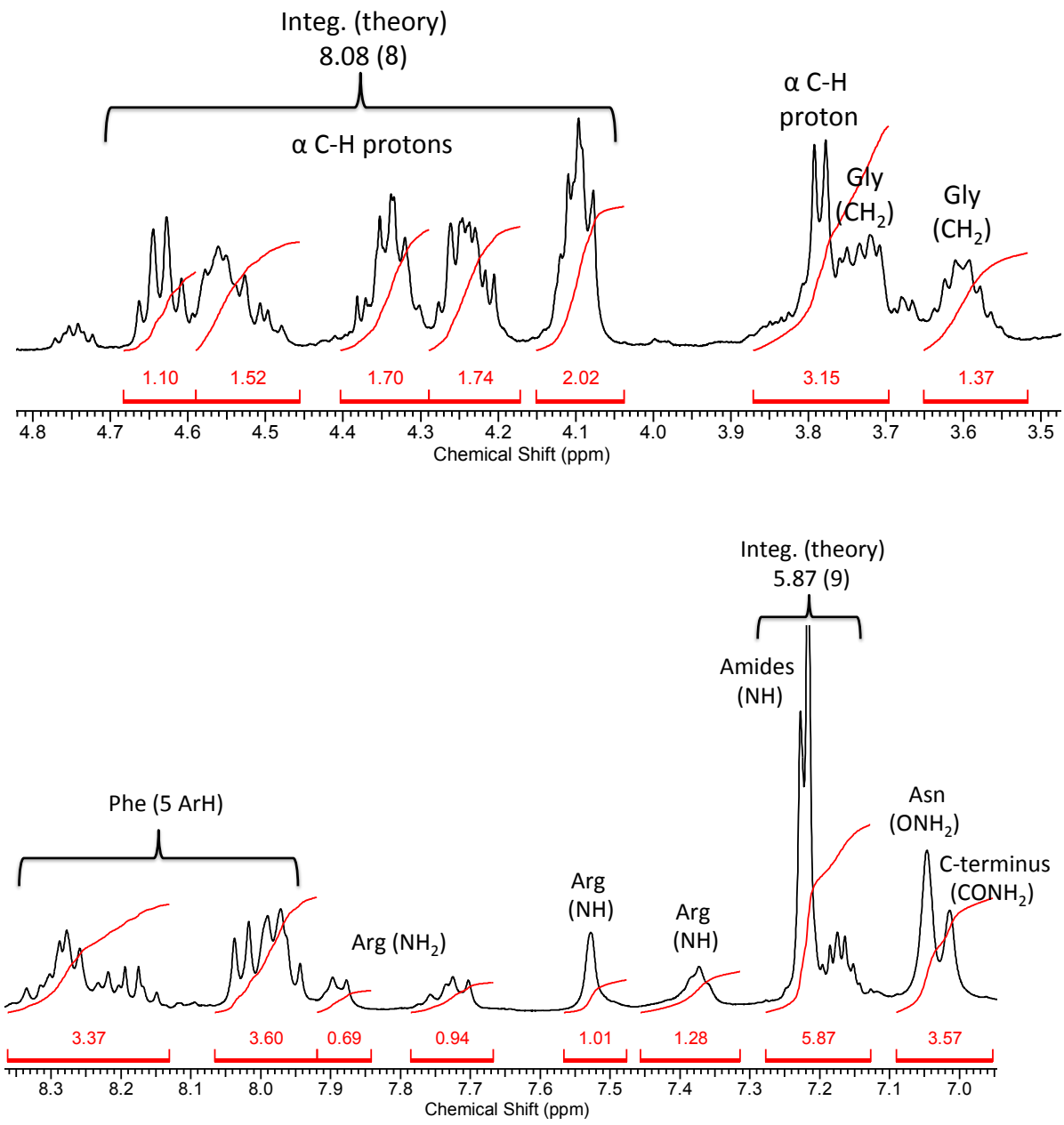


Figure 2.21. ^{1}H NMR of **Ac-CGSGRVFNIM-CONH$_2$**.

^{1}H NMR (400 MHz, DMSO) δ 0.74 (d, *J* = 6.7 Hz, CH_3 [Val], 6H), 0.81 (d, *J* = 7.4 Hz, CH_3 [Ile], 3H), 0.83 (t, *J* = 6.8 Hz, CH_3 [Ile], 3H), 1.10-1.21 (m, CH_2 [Ile], 2H), 1.34-1.54 (m, CH_2 [Arg], 4H), 1.56-1.68 (m, CH [Ile], 1H), 1.72-1.84 (m, CH_2 [Arg], 2H), 1.89 (s, OCH_3 [N-terminus], 3H), 1.90-1.97 (m, CH_2 [Met], 2H), 2.03 (s, SCH_3 [Met], 3H), 2.35-2.47 (m, CH_2 [Met], 2H), 2.52-2.61 (m, CH_2 [Asn], 2H), 2.63-2.70 (m, CH_2 [Cys], 2H), 2.71-2.83 (m, $ArCH_2$ [Phe], 1H), 2.98 (dd, *J* = 14.0, 4.0, $ArCH_2$ [Phe], 1H), 3.06 (d, *J* = 6.1, $ArCH_2$ [Phe], 2H), 3.52-3.65 (m, CH_2 [Gly], 2H), 3.69-3.76 (m, CH_2 [Gly], 2H), 3.77-3.81 (d, *J* = 5.9, αC-H, 1H), 4.06-4.13 (m, αC-H,

2H), 4.19-4.28 (m, αC-H, 2H), 4.46-4.60 (m, αC-H, 2H), 4.60-4.68 (m , αC-H, 2H), 7.01 (s, $CONH_2$ [C-terminus], 2H), 7.05 (s, ONH_2 [Asn], 2H), 7.13-7.24 (m, NH, 9H), 7.34−7.43 (m, NH [Arg], 1H), 7.53 (s, NH [Arg], 1H), 7.69-7.78 (m, NH [Arg], 1H), 7.93-8.06 (m, ArH [Phe], 2H), 8.13−8.36 (m, ArH [Phe], 3H).

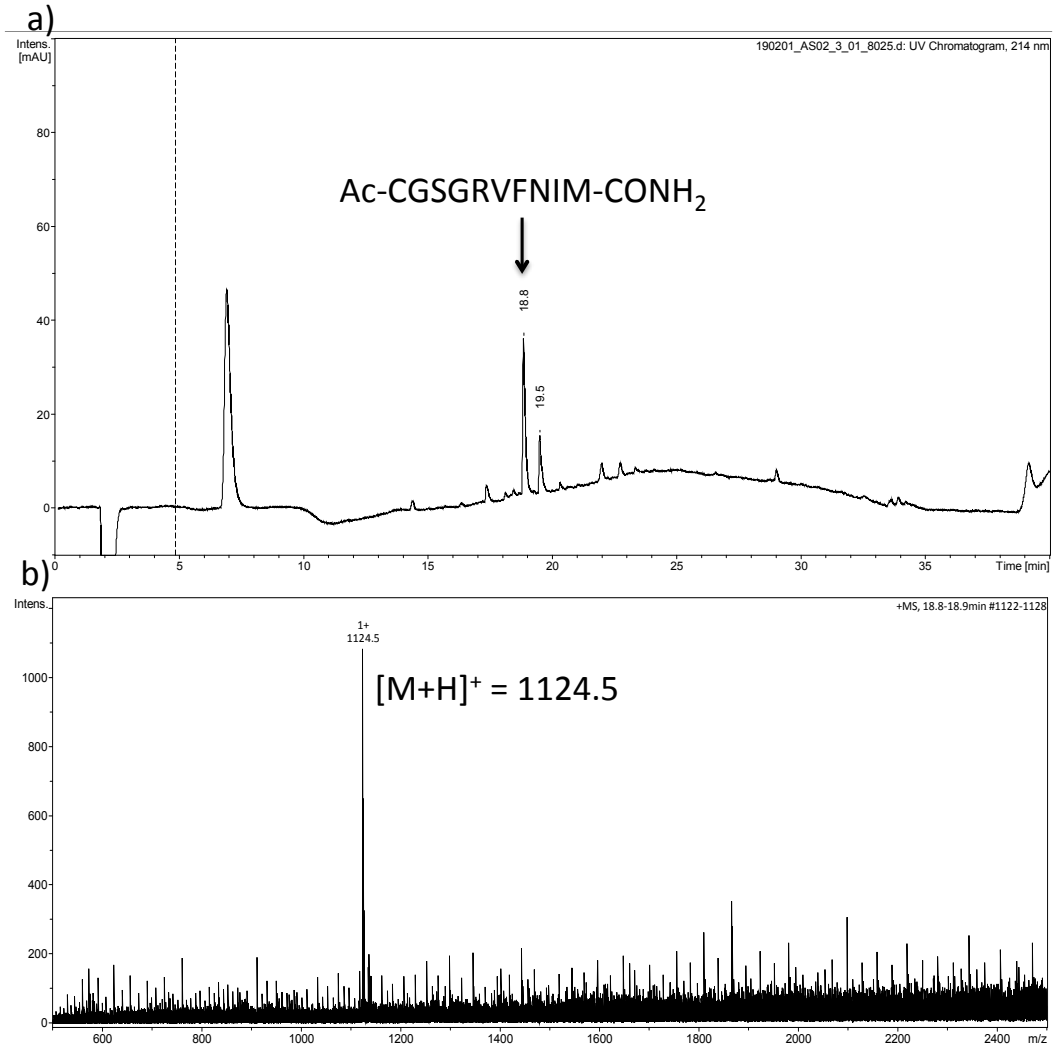


Figure 2.22. a) Analytical HPLC chromatograph of **Ac-CGSGRVFNIM-CONH$_2$** (monitoring absorbance at 214 nm) showing that the peptide peak at 18.8 min. Expanded trace of **b)** desired peptide peak. LC-MS (m/z) calc. for: $[M+H]^+$ 1124.34, found 1124.5.

Ac-(POG)$_3$-CONH$_2$

The ^{1}H NMR of **Ac-(POG)$_3$-CONH$_2$** showed a number of broad peaks that were likely attributed to peptide aggregation, making it difficult to assign the corresponding proton peaks of the peptide (Figure 2.23). The LC-MS measurement of **Ac-(POG)$_3$-CONH$_2$** showed a strong peak of the desired peptide, as well as small impurity peaks on the HPLC chromatogram, indicating that the peptide was obtained with a high degree of purity (Figure 2.24).

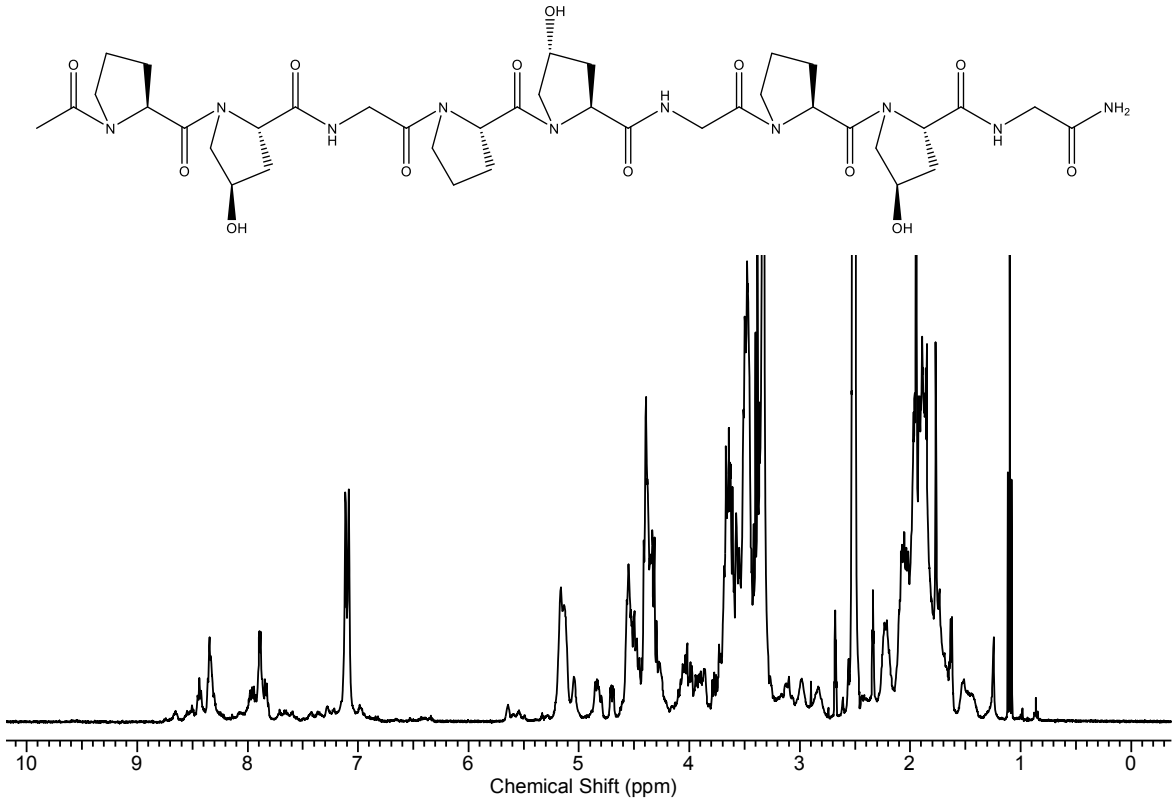


Figure 2.23. ^{1}H NMR of **Ac-(POG)$_3$-CONH$_2$**.

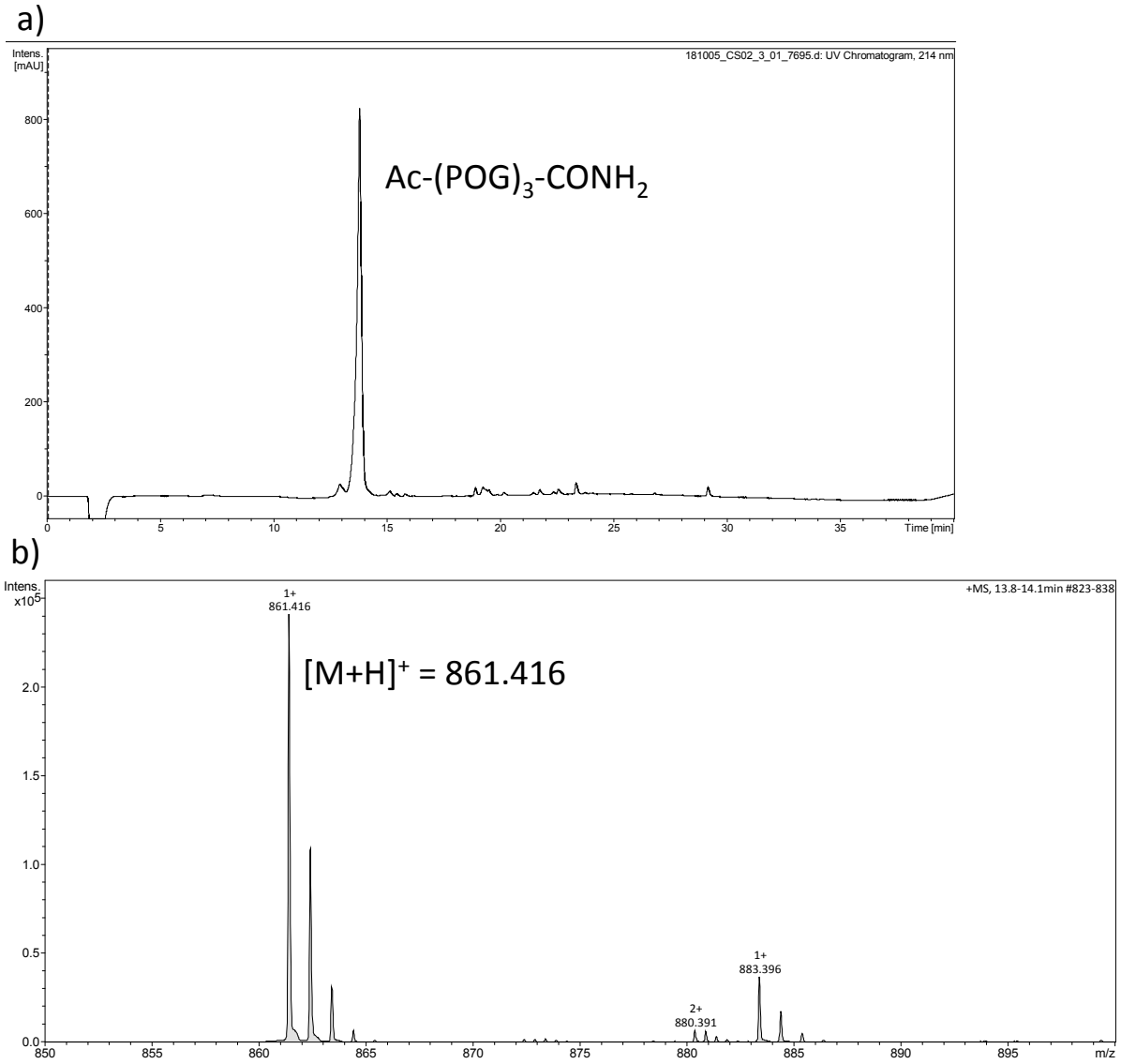


Figure 2.24. Analytical HPLC chromatograph of **Ac-(POG)$_3$-CONH$_2$** (monitoring absorbance at 214 nm) showing that the peptide peak at 13.8 min. Expanded trace of **b)** desired peptide peak. LC-MS (m/z) calc. for: $[M+H]^+$ 861.4, found 861.426.

Owing to the difficulties encountered when assigning the proton peaks of **Ac-(POG)$_3$-CONH$_2$** due to peptide aggregation, the remaining collagen mimetic peptides (**Ac-(POG)$_6$-CONH$_2$**, **Ac-C(POG)$_6$-CONH$_2$** and **Ac-CGSG(POG)$_6$-CONH$_2$**) were only assessed by LC-MS.

Ac-(POG)$_6$-CONH$_2$

The LC-MS measurement of **Ac-(POG)$_6$-CONH$_2$** revealed a strong peak corresponding to the desired peptide on the HPLC chromatogram, as well as showing small impurity peaks, indicating that the peptide was obtained with a high degree of purity (Figure 2.25).

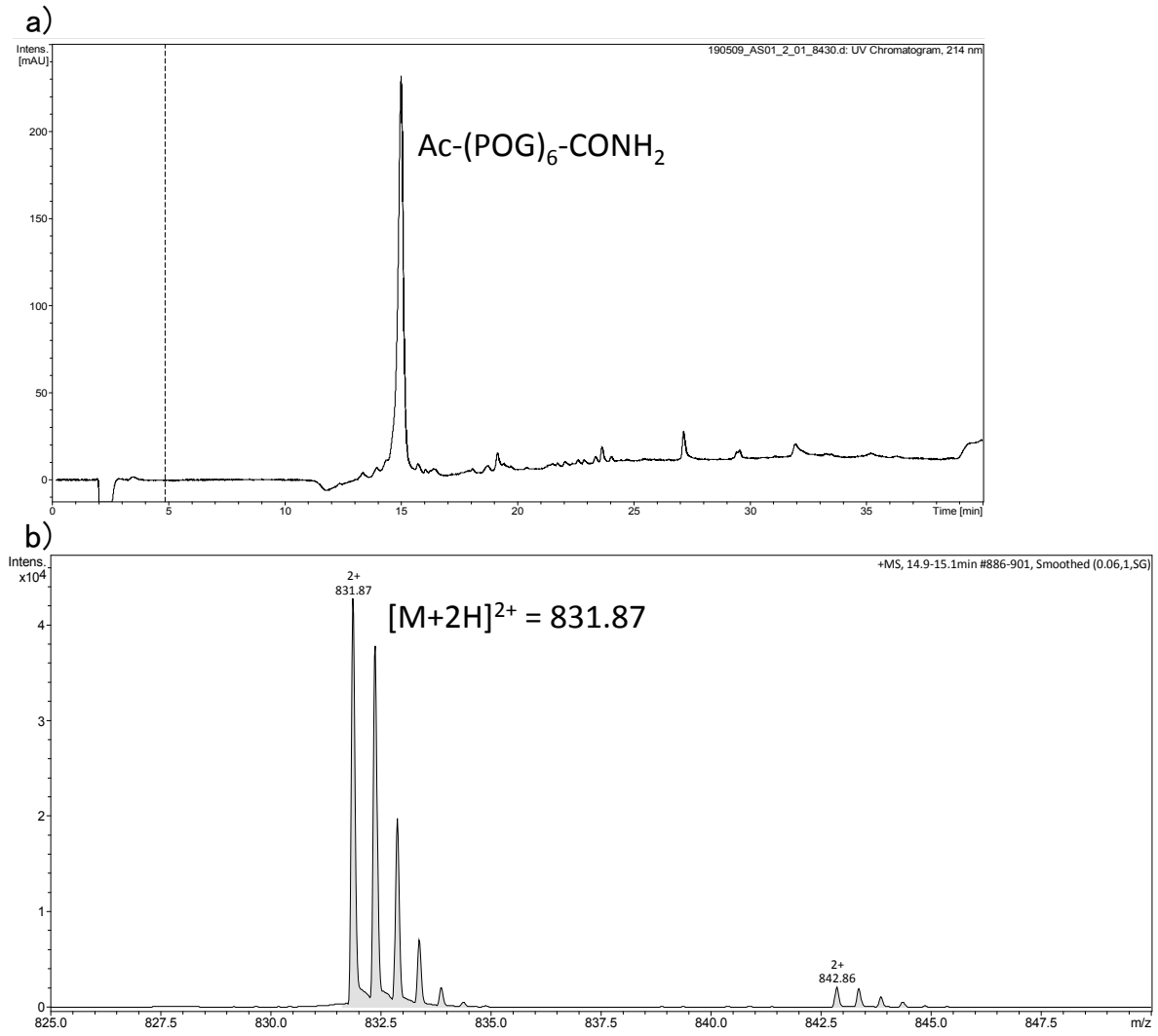


Figure 2.25. a) Analytical HPLC chromatograph of **Ac-(POG)$_6$-CONH$_2$** (monitoring absorbance at 214 nm) showing that the peptide eluted at 22 min. **b)** Expanded trace of the peptide peak. LC-MS (m/z) calc. for: $[M+2H]^{2+}$ 831.3, found 831.87.

Ac-C(POG)$_6$-CONH$_2$

The LC-MS measurement of **Ac-C(POG)$_6$-CONH$_2$** showed a peak corresponding to the desired peptide as a dimer on the HPLC chromatogram, as well as showing several impurity peaks (Figure 2.26).

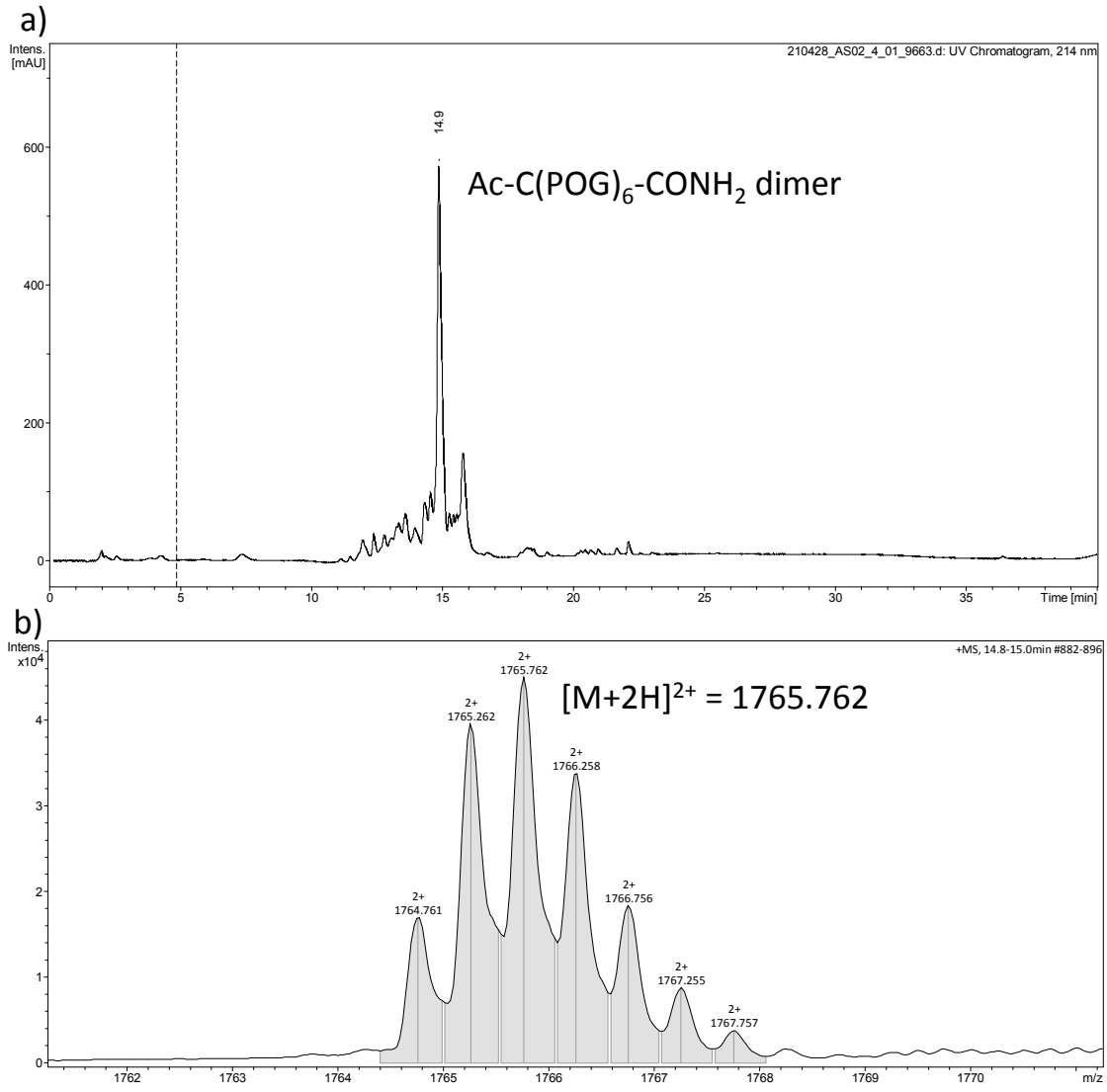
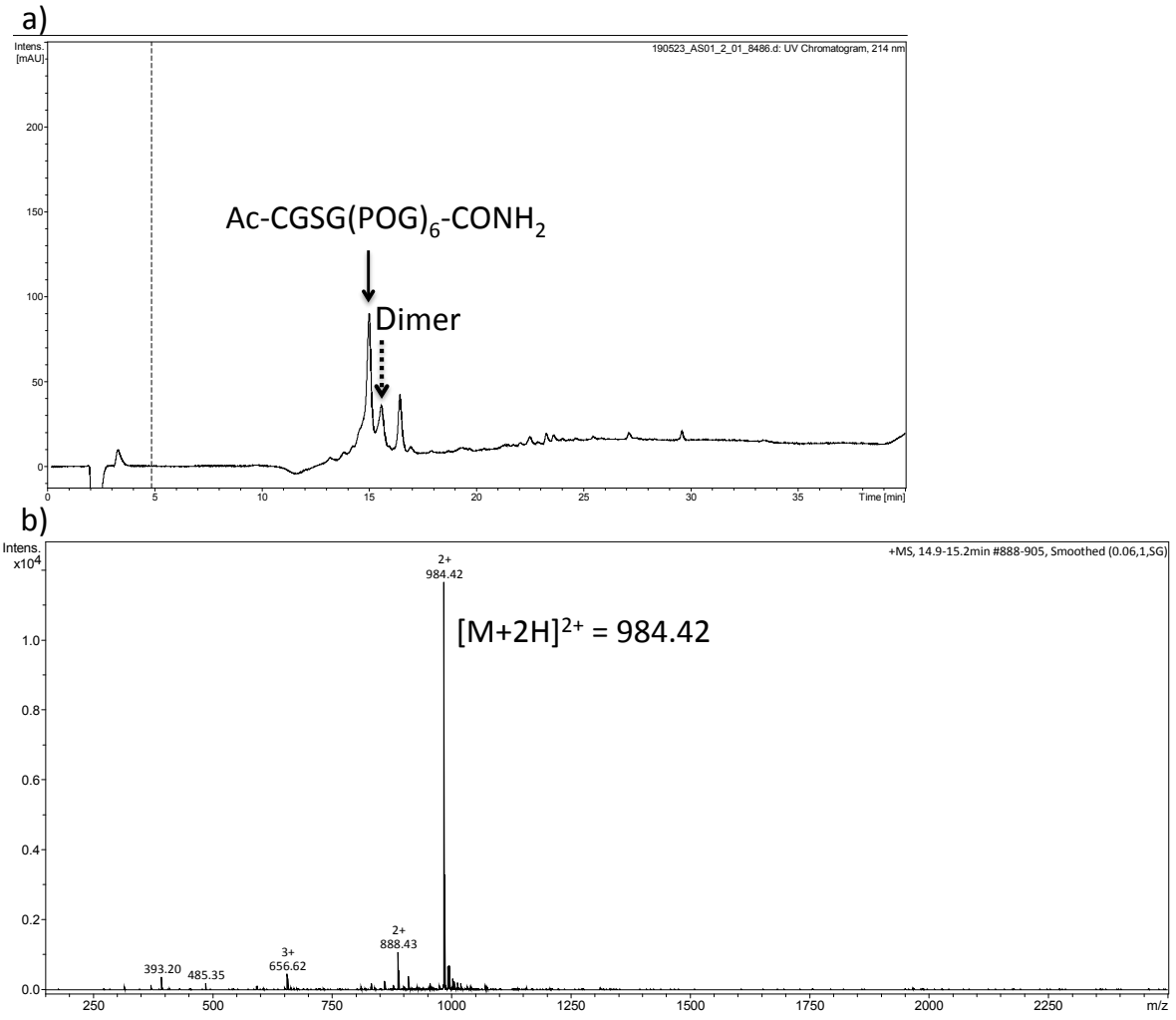


Figure 2.26. a) Analytical HPLC chromatograph of **Ac-C(POG)$_6$-CONH$_2$** (monitoring absorbance at 214 nm) showing that the peptide eluted as a dimer at 14.8 min. **b)** Expanded trace of the peptide peak. LC-MS (m/z) calc. for: $[M+2H]^{2+}$ 1765.9, found 1765.762.

Ac-CGSG(POG)$_6$-CONH$_2$

The LC-MS measurement of **Ac-CGSG(POG)$_6$-CONH$_2$** revealed a peak corresponding to the desired peptide on the HPLC chromatogram, as well as two additional peaks (Figure 2.27 a and b). The second peak at 15.5 min appeared to correspond to the formation of dimer (Figure 2.27 c), whereas the third peak at 16.3 min corresponded to a compound that was larger in mass by 56 compared with the desired peptide (Figure 2.27 d). The third peak appeared to have resulted due to the presence of a t-butyl group, suggesting that this protecting group was not completely removed from the desired peptide.



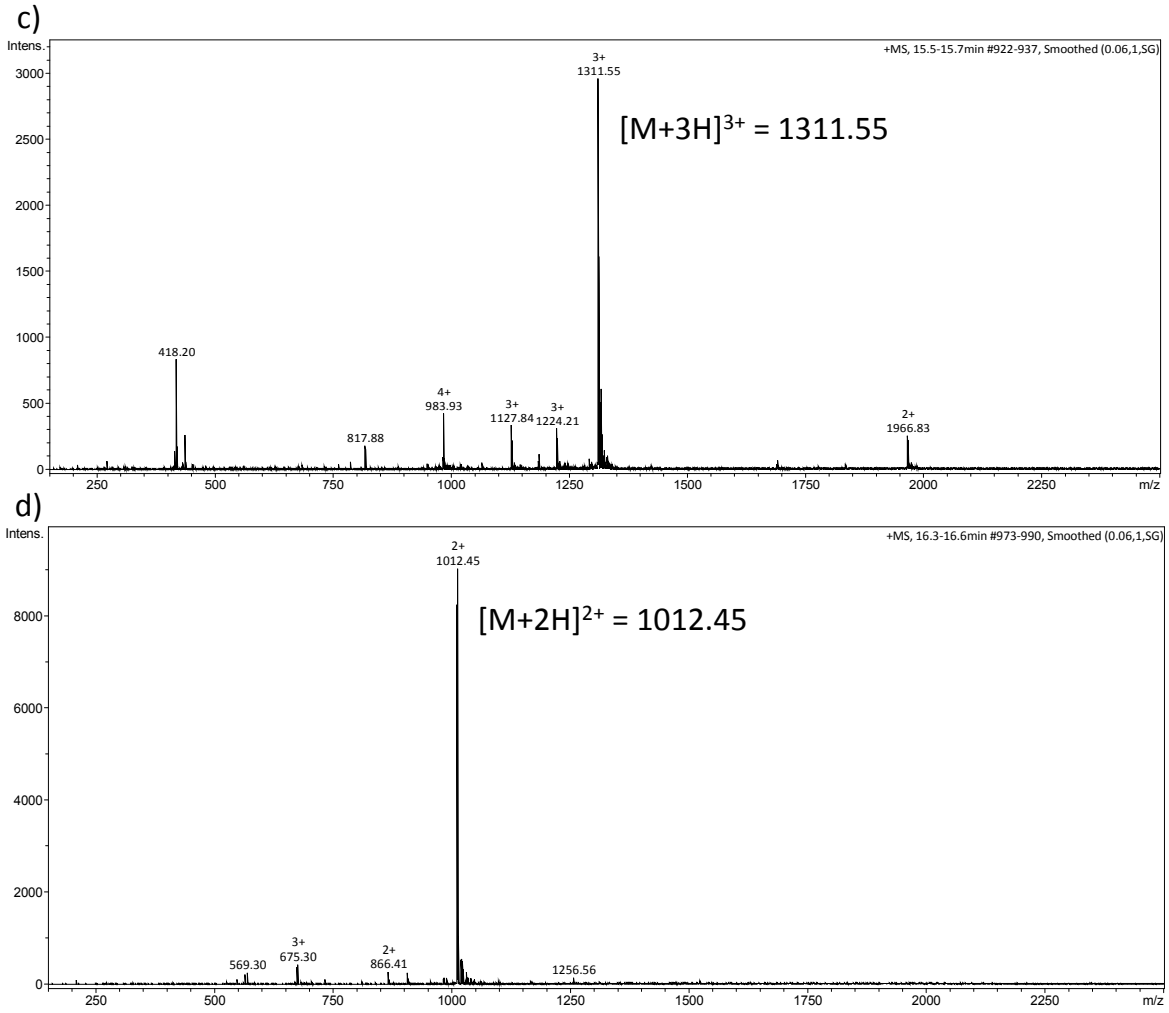


Figure 2.27. a) Analytical HPLC chromatograph of **Ac-CGSG(POG)$_6$-CONH$_2$** (monitoring absorbance at 214 nm) showing that the peptide eluted at 14.9 min. **b)** Expanded trace of the peptide peak. LC-MS (m/z) calc. for: $[M+2H]^{2+}$ 984.6, found 984.42. **c)** Expanded trace of the second peak at 15.5 min, suggesting the formation of dimer. **d)** Expanded trace of the third peak at 16.3 min, corresponding to a compound with a mass difference of +56 compared with the desired peptide.

N_3-$(CH_2)_2$-$(POG)_6$-$CONH_2$

The LC-MS measurement of **N_3-$(CH_2)_2$-$(POG)_6$-$CONH_2$** revealed a peak corresponding to the desired peptide on the HPLC chromatogram (Figure 2.28 a and c). In addition, it showed a peak corresponding to an unidentified compound that was smaller in mass by 11.97 compared with the desired peptide (Figure 2.28 b).

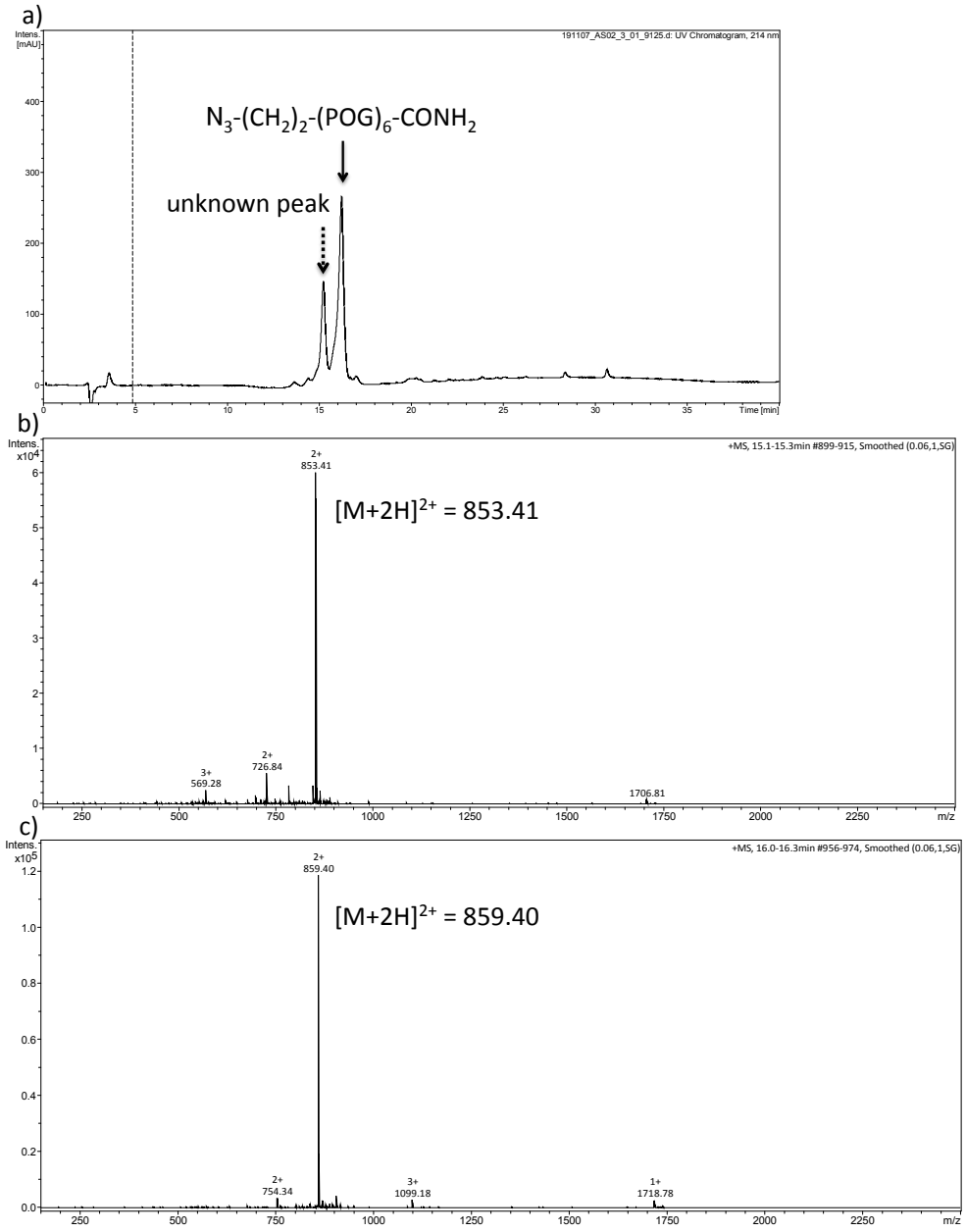


Figure 2.28. a) Analytical HPLC chromatograph of **N_3-$(CH_2)_2$-$(POG)_6$-$CONH_2$** (monitoring absorbance at 214 nm) showing that the desired peptide eluted at 16 min. **b)** Expanded trace of unknown peak at 15.1 min. **c)** Expanded trace of the desired peptide peak at 16.0 min. LC-MS (m/z) calc. for: $[M+2H]^{2+}$ 859.5, found 859.40.

Analysis of Self-assembling Behaviour of Peptides

The self-assembling behaviour of the following β-sheet forming peptides; **Ac-KLVFFA-$CONH_2$**, **Ac-HYFNIF-$CONH_2$**, **N_3-$(CH_2)_2$-HYFNIF-$CONH_2$** and **Ac-RVFNIM-$CONH_2$**, and collagen mimetic peptides **Ac-$(POG)_3$-$CONH_2$** and **Ac-$(POG)_6$-$CONH_2$** was investigated by CD and AFM to assess their self-assembling behaviour. These peptides will be referred to as **KLVFFA**, **HYFNIF**, **N_3-HYFNIF**, **RVFNIM**, **$(POG)_3$** and **$(POG)_6$**, respectively. The obtained results were compared with previous literature findings (CD and/or AFM) of the same or similar peptide sequences to confirm the predicted self-assembling properties of the synthesised peptides.

While CD spectroscopy provides insights into the structural properties of peptides, it is not possible to fully depict their secondary structure owing to the wide structural diversity of peptides. Even small changes within peptides such as variations in length and degree of twist can lead to large changes in CD spectra, thereby limiting the capacity of CD spectroscopy to accurately predict the secondary structure of peptides[15]. For example, while a classic β-sheet spectrum is expected to show a positive peak at around 195 nm and a negative peak at around 216 nm, a group of researchers reported that Waltz peptides that are known to form β-sheets did not follow these CD patterns, owing to their anisotropic and high aromatic content nature[1].

AFM provides morphological information on samples in response to tip-surface interactions between the cantilever and the sample surface[18]. However, a key limitation is that the measured diameter of structures (e.g. peptides) may not reflect their exact size due to due to the tip-broadening effect[19].

Self-assembling Behaviour of KLVFFA

The CD spectra of **KLVFFA** showed negative peaks at 200 nm and shoulder signals at 230 nm at concentrations $\geqq$ 0.05 mg/mL (Figure 2.29). These profiles were similar to those reported in the literature for KLVFFAE (with both N and C termini capped), which showed negative CD peaks at 215 nm, after incubation for 20 h in 40% acetonitrile/water with 0.1% TFA at pH 2.[20] Although direct comparisons cannot be made as the experimental conditions

were different, **KLVFFA** showed an expected self-assembly behaviour that is suggestive of β-sheet formation.[21]

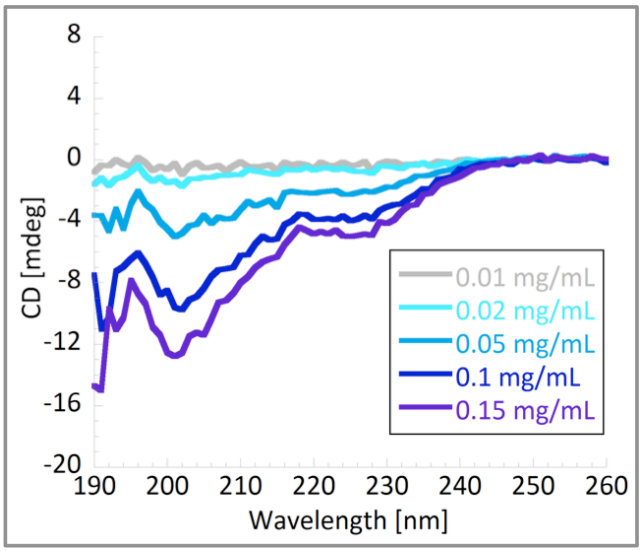
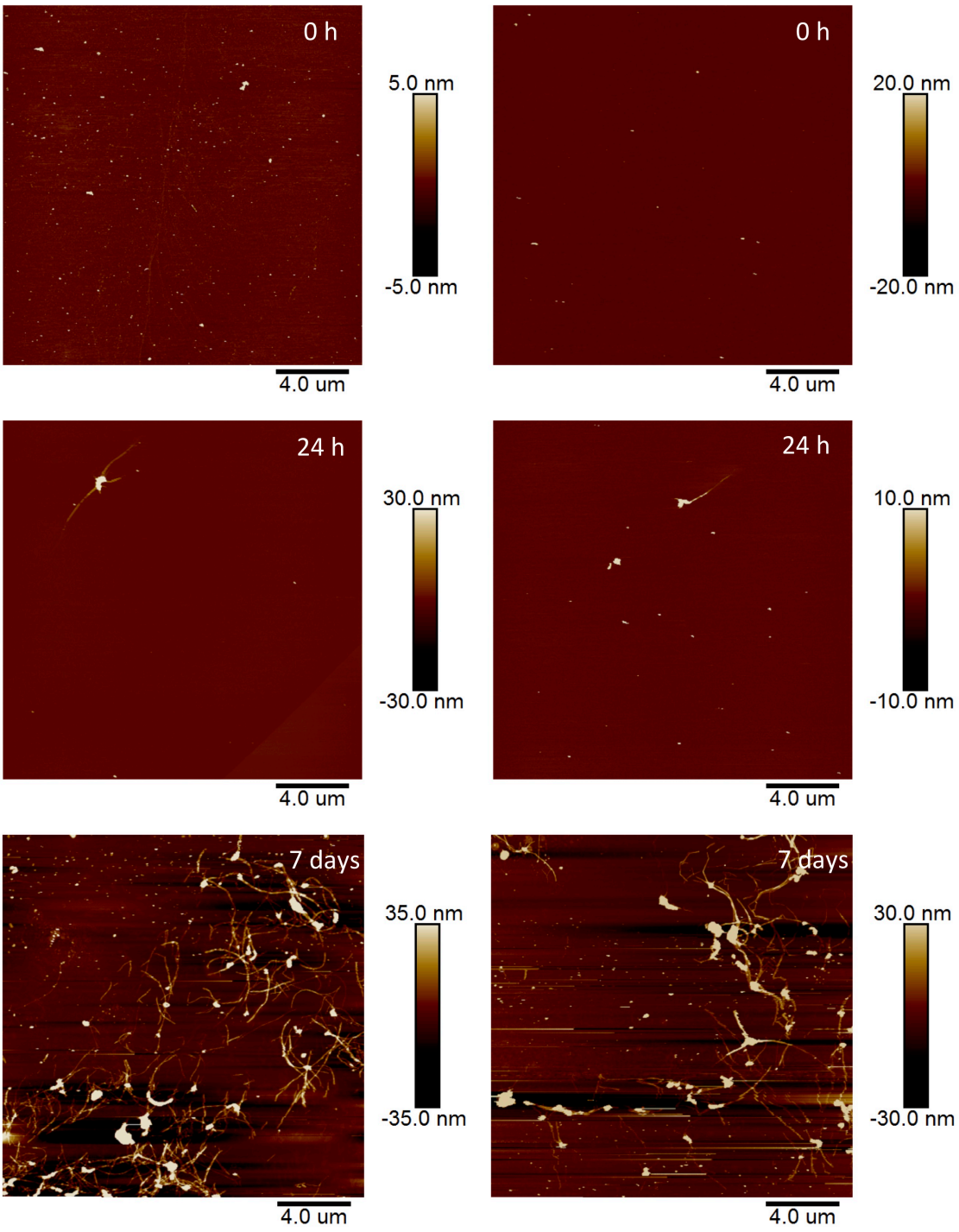


Figure 2.29. CD spectra of **KLVFFA** in water at concentrations of 0.01, 0.02, 0.05, 0.1 and 0.15 mg/mL.

The AFM study of KLVFFA at R.T. revealed that the peptide underwent morphological changes during a 28-day period (Figure 2.30). At 0 h, only dot-like or thin fibre-like structures were observed, but these further elongated at 24 h, suggesting a structural transition into fibres. Subsequently, twisted fibres were observed at 7 days, which further elongated and bundled over the next few weeks. In contrast, when **KLVFFA** was annealed (heating to 95°C, then gently cooling down to 4°C at a cooling rate of 1 °C per minute), this resulted in sheet-like structures not previously observed at R.T. (Figure 2.31), highlighting the polymorphic nature of this peptide in response to different self-assembling conditions.[22]


0 h
5.0 nm
-5.0 nm
4.0 um
0 h
20.0 nm
-20.0 nm
4.0 um
24 h
30.0 nm
-30.0 nm
4.0 um
24 h
10.0 nm
-10.0 nm
4.0 um
7 days
35.0 nm
-35.0 nm
4.0 um
7 days
30.0 nm
-30.0 nm
4.0 um

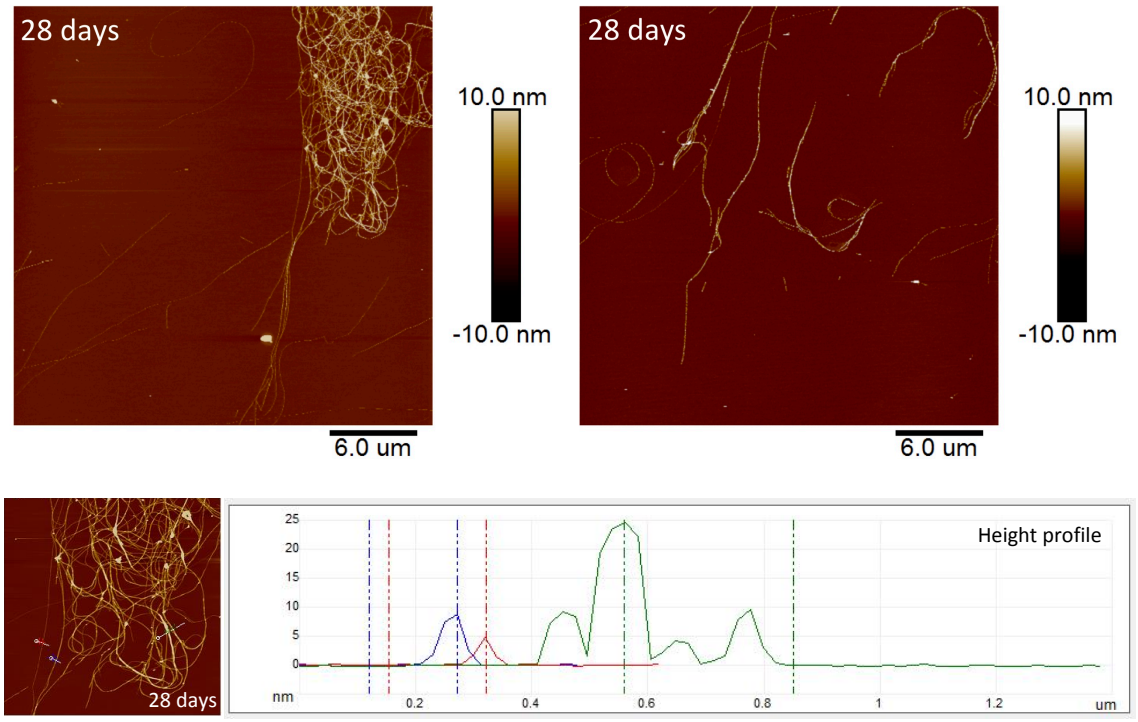


Figure 2.30. AFM images of **KLVFFA** at a concentration of 0.2 mg/mL in water at 0h, 24 h, 7 days and 28 days of aging.

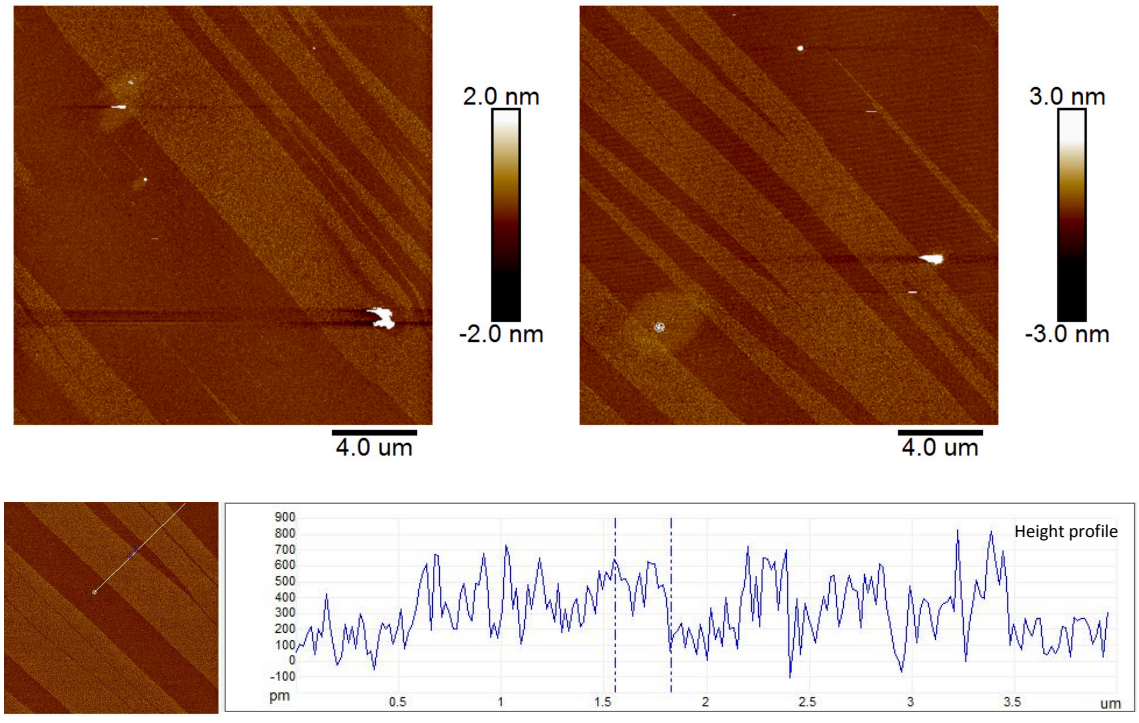


Figure 2.31. AFM images of annealed **KLVFFA** at a concentration of 0.2 mg/mL in water following annealing.

Self-assembling Behaviour of HYFNIF

The CD spectra of **HYFNIF** showed broad negative peaks at around 210 nm at concentrations ≧0.05 mg/mL and positive signals between 190-195 nm at concentrations ≧0.1 mg/mL, indicative of β-sheet structure formation (Figure 2.32).[1] The CD profiles obtained for HYFNIF were consistent with those previously reported for **HYFNIF**, thereby confirming the expected self-assembling behaviour of this peptide.[1]

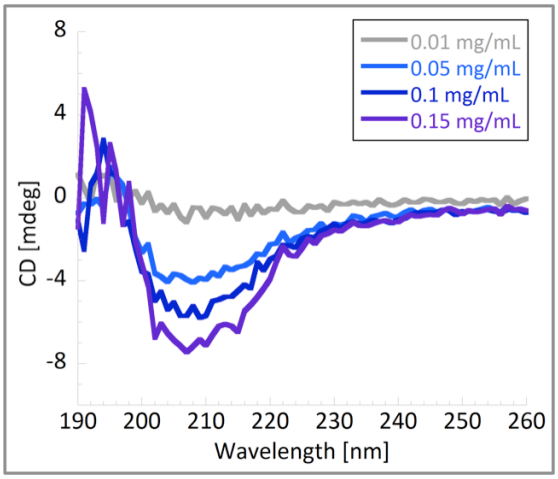


Figure 2.32. CD spectra of **HYFNIF** in water at concentrations of 0.01, 0.05, 0.1 and 0.15 mg/mL.

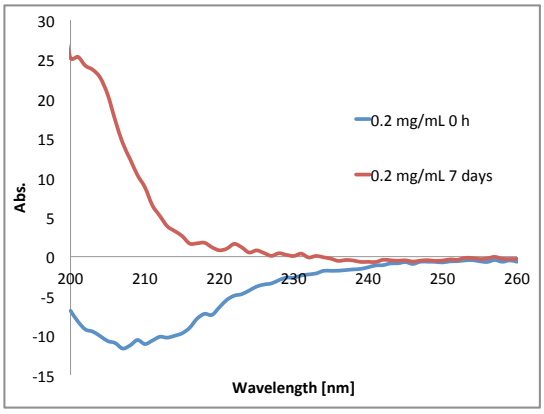
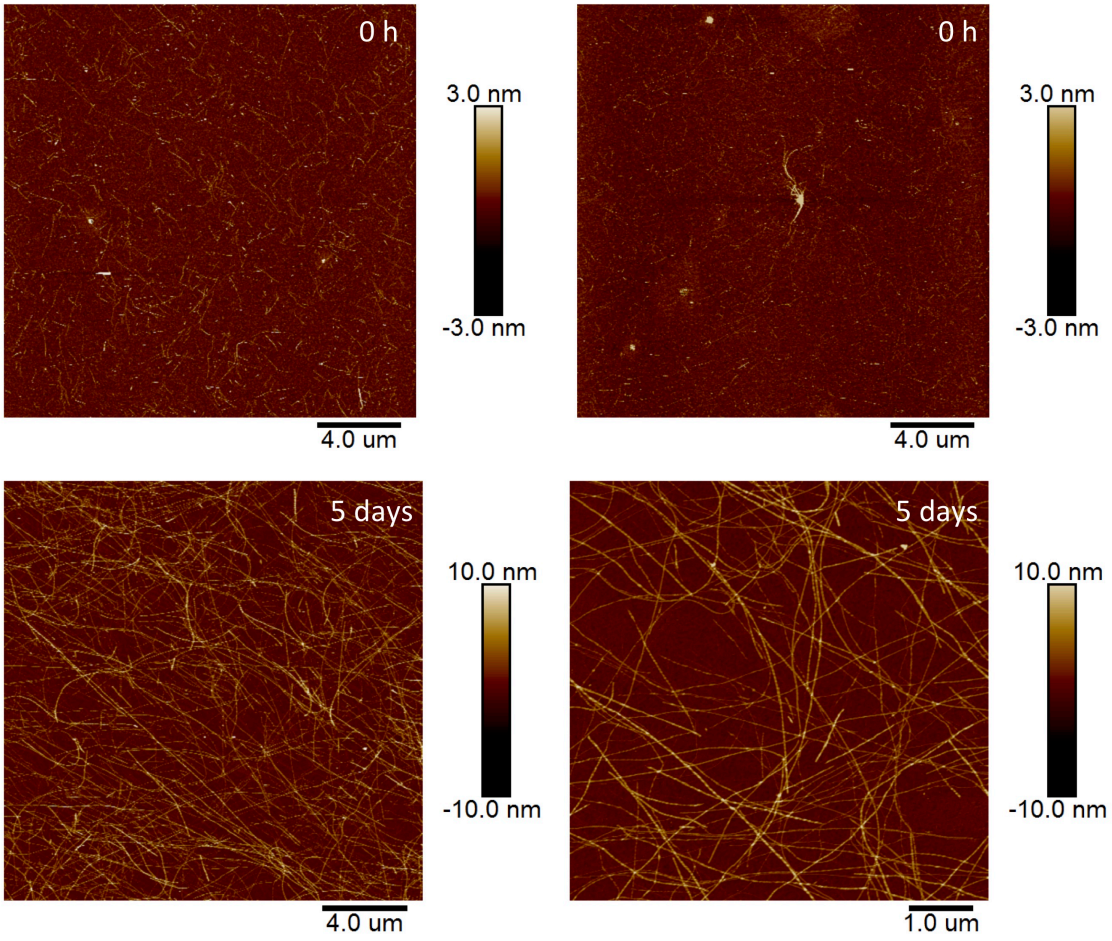


Figure 2.33. CD spectra of **HYFNIF** in water at a concentration of 0.2 mg/mL at 0 h and 7 days of aging.

In addition, the CD spectra of **HYFNIF** changed considerably following incubation at R.T. during 7 days. At 0 h, a negative peak at around 210 nm was observed, whereas at 7 days this disappeared and instead a positive signal at 200 nm was observed (Figure 2.33). This positive signal suggested fibril alignment, which was previously reported for HYFNIF fibrils in

the literature.[23] Moreover, AFM analysis showed the presence of thin fibres at 0 h, which eventually developed into long twisted fibres during a 5-day period (Figure 2.34), being consistent with the CD spectra changes observed for **HYFNIF** over time. AFM observations of **N_3-$(CH_2)_2$-HYFNIF** were in line with those seen for **HYFNIF**, showing that the azide modification had limited to no influence on the self-assembly of this peptide (Figure 2.35 and Figure 2.36). In contrast, spherical aggregates and disrupted fibres were observed when **HYFNIF** was annealed, implying that the fibres are susceptible to heat (Figure 2.37).



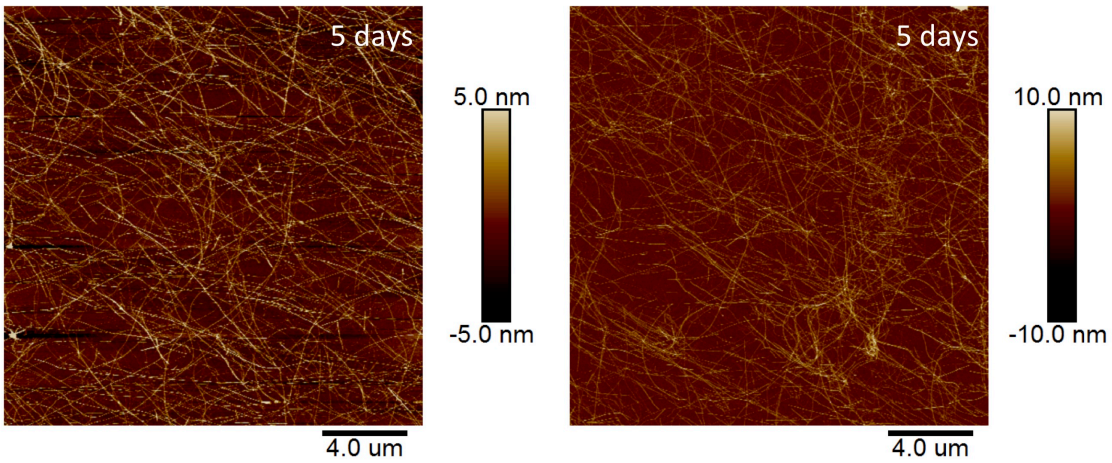
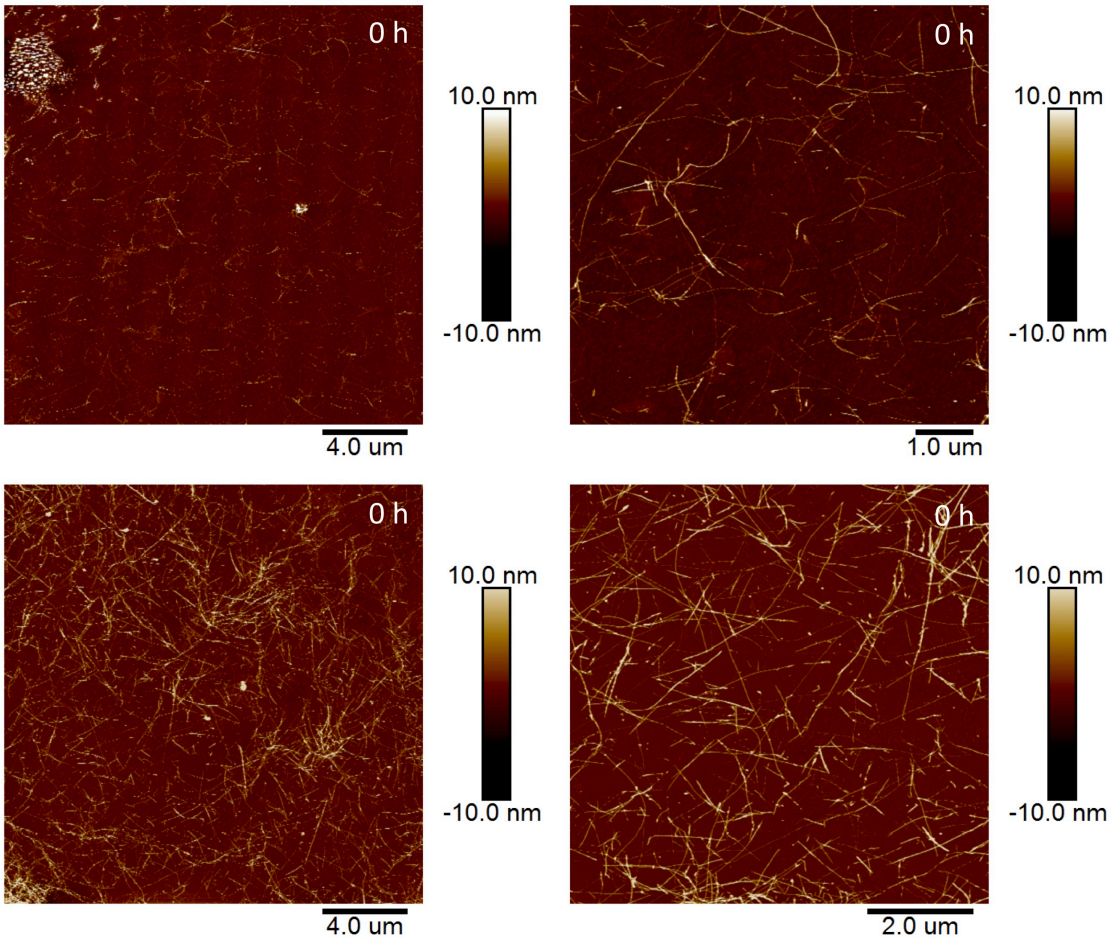


Figure 2.34. AFM images of **HYFNIF** at a concentration of 0.2 mg/mL in water at 0 h and 5 days of aging.



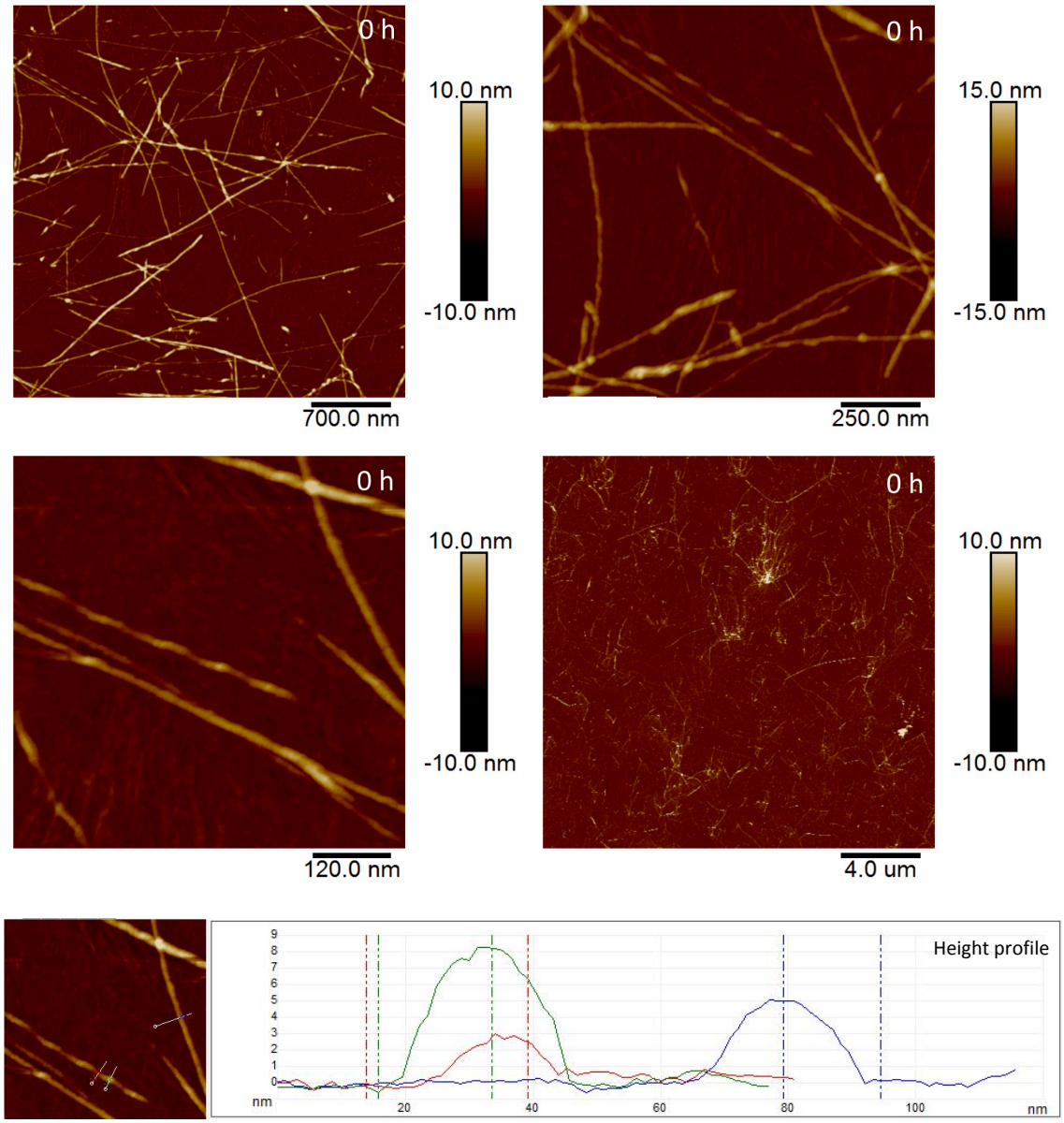
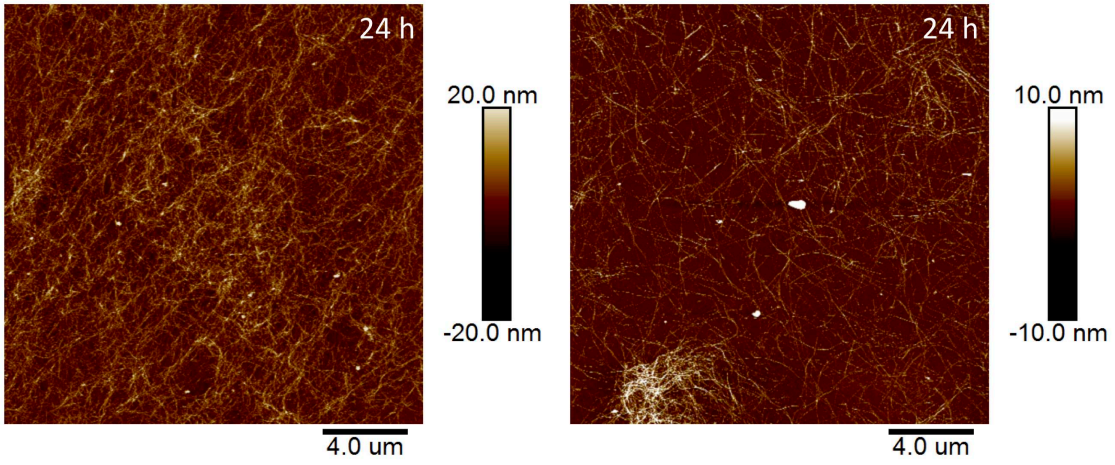


Figure 2.35. AFM images of **HYFNIF-N_3** at a concentration of 0.2 mg/mL in water.



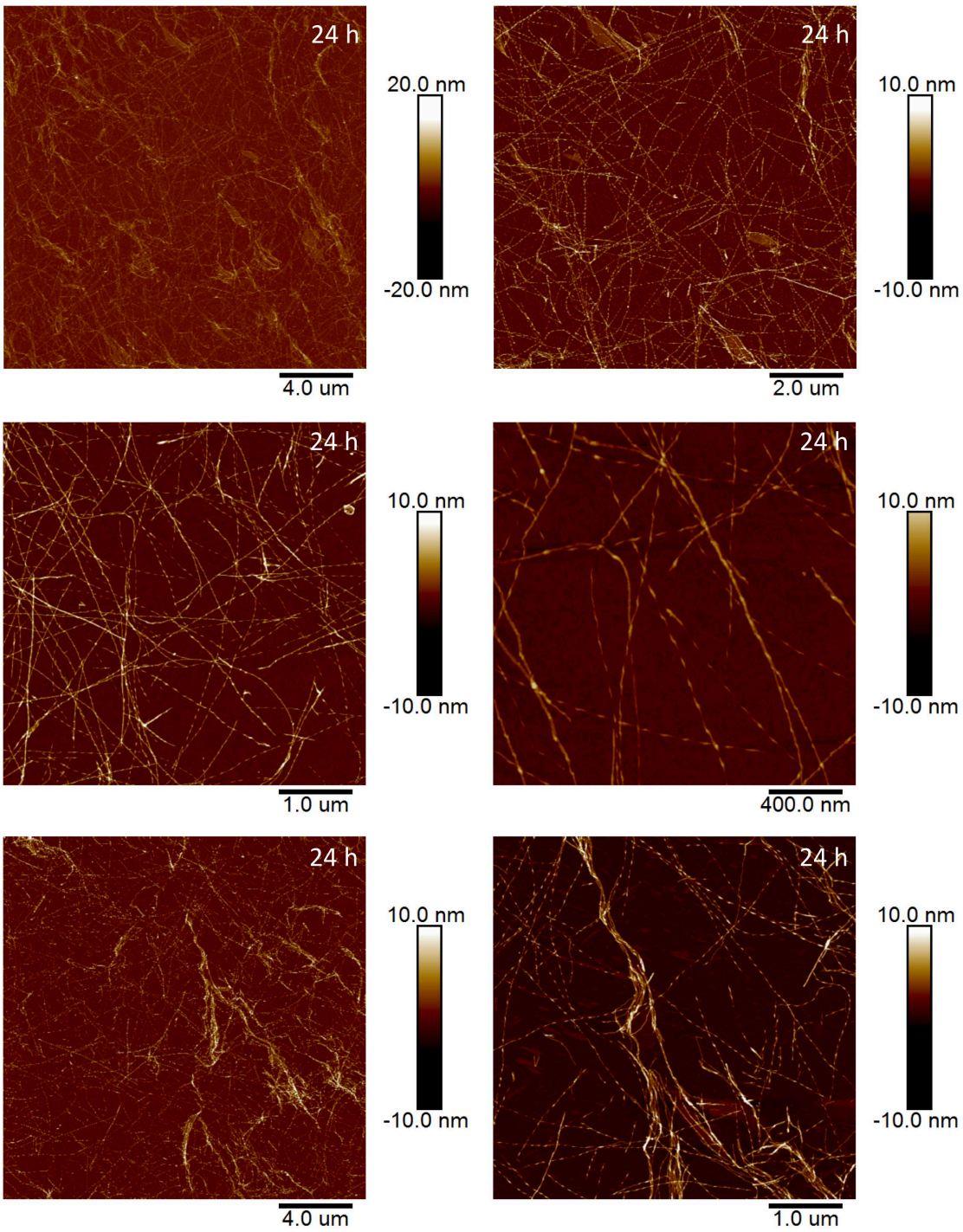


Figure 2.36. AFM images of **HYFNIF-N_3** at a concentration of 0.2 mg/mL in water at 24 h.

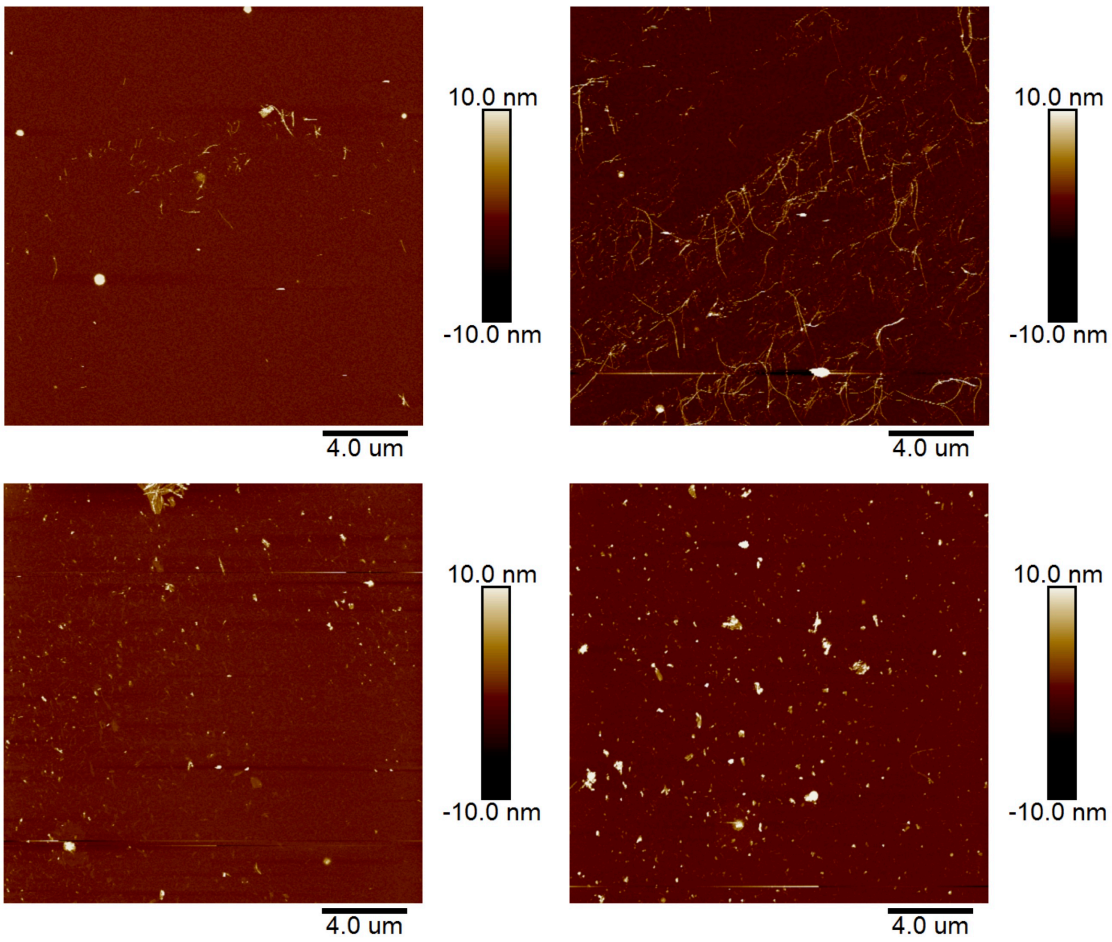


Figure 2.37. AFM images of annealed **HYFNIF** at a concentration of 0.2 mg/mL in water.

Self-assembling Behaviour of RVFNIM

The CD spectra of **RVFNIM** showed negative signals at around 202 nm and broad signals between 220-230 nm at concentrations ≧0.05 mg/mL, as well as positive signals at 195 nm (Figure 2.38). This result is consistent with the CD spectra previously reported for **RVFNIM** in the literature, thereby confirming the expected self-assembling behaviour of this peptide.

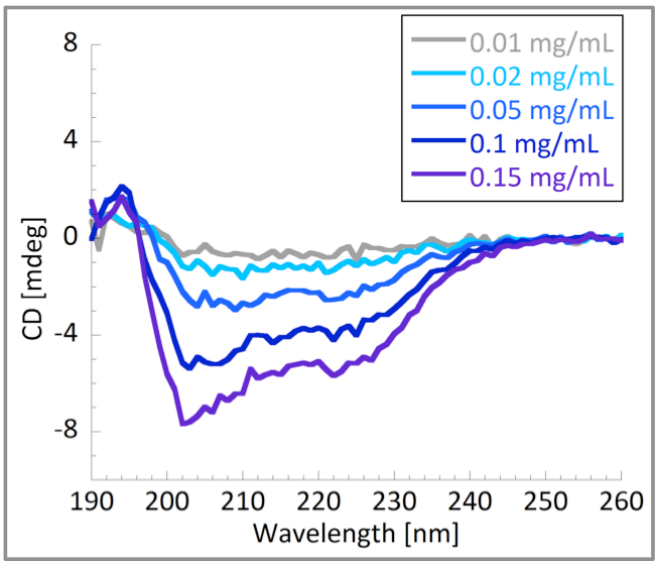


Figure 2.38. CD spectra of **RVFNIM** in water at concentrations of 0.01, 0.02, 0.05, 0.1 and 0.15 mg/mL.

The AFM study of **RVFNIM** showed long (10-20 µm) thin fibres with a height of 1-3 nm at 0 h, which developed into thicker and twisted fibres with a height of ~6 nm and with a width of ~20 nm following incubation at R.T. over 10 days (Figure 2.39 and 2.40). In contrast, annealed **RVFNIM** resulted in shorter and more rigid fibres (10-30 nm height and ~50 nm width) compared with those seen in non-annealed conditions (Figure 2.41). Given that it took up to 10 days at R.T. until bundled fibres were observed, the annealing experiment may suggest that applying heat to this peptide could have facilitated the formation of bundled fibres. Furthermore, the incubation of the annealed sample at R.T. for 10 days led to long straight fibres consisting of smaller bundled fibres, which were possibly generated due to elongation of the rigid fibres post annealing. These results suggested that varying the temperature at an early stage of the self-assembly process had the potential to affect the resultant peptide structure.[24]

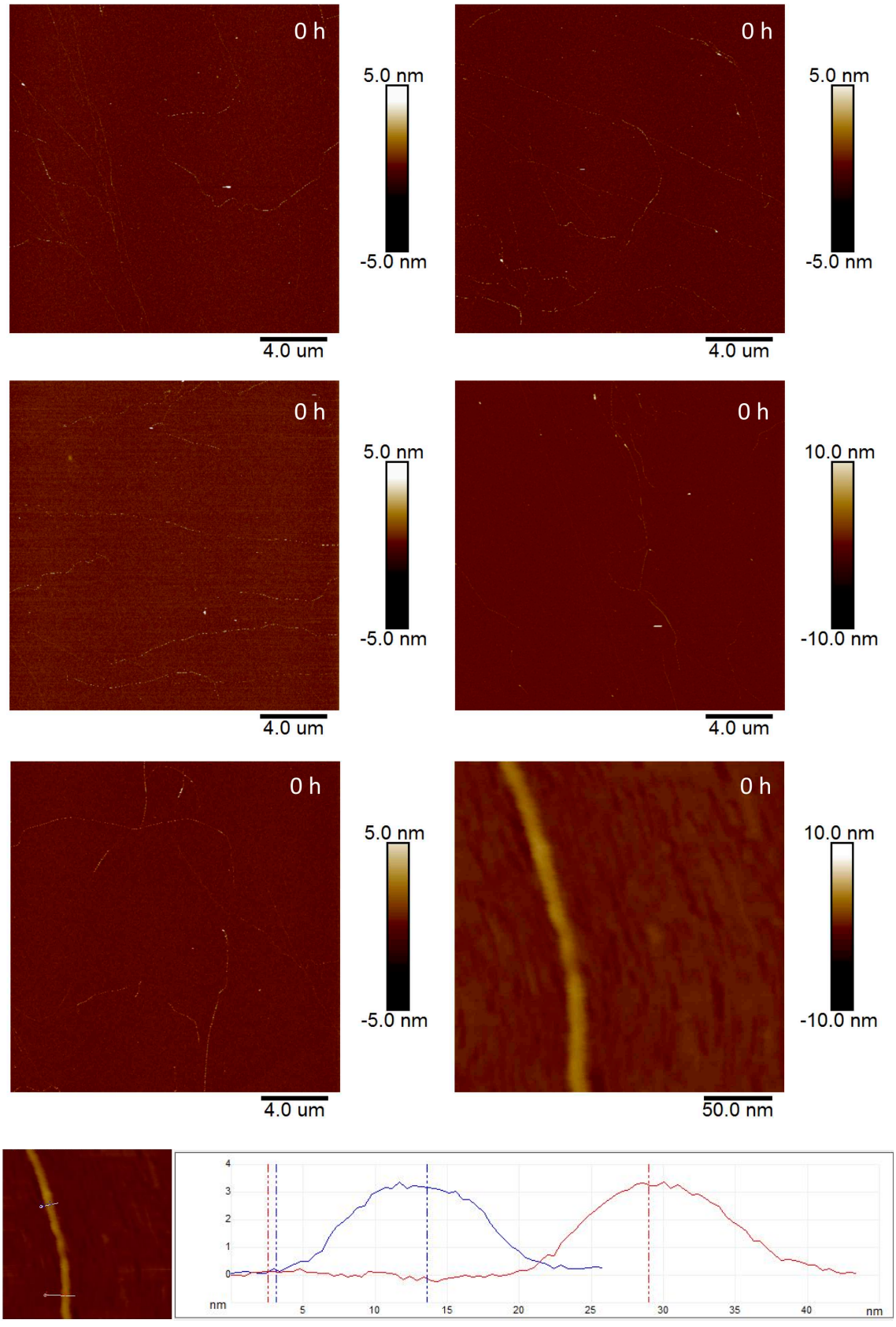


Figure 2.39. AFM images of **RVFNIM** at a concentration of 0.2 mg/mL in water at 0 h.

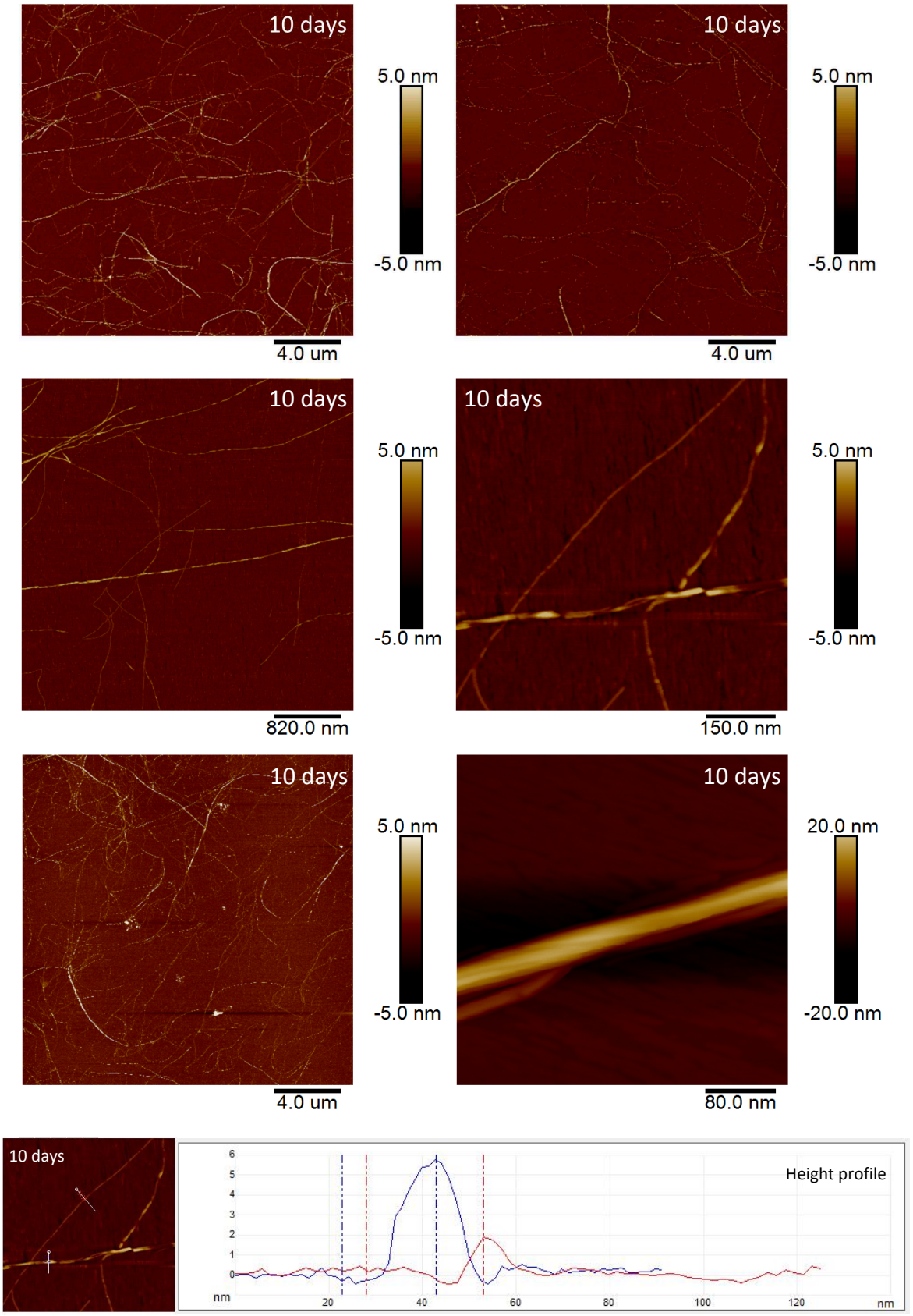


Figure 2.40. AFM images of **RVFNIM** at a concentration of 0.2 mg/mL in water at 10 days.

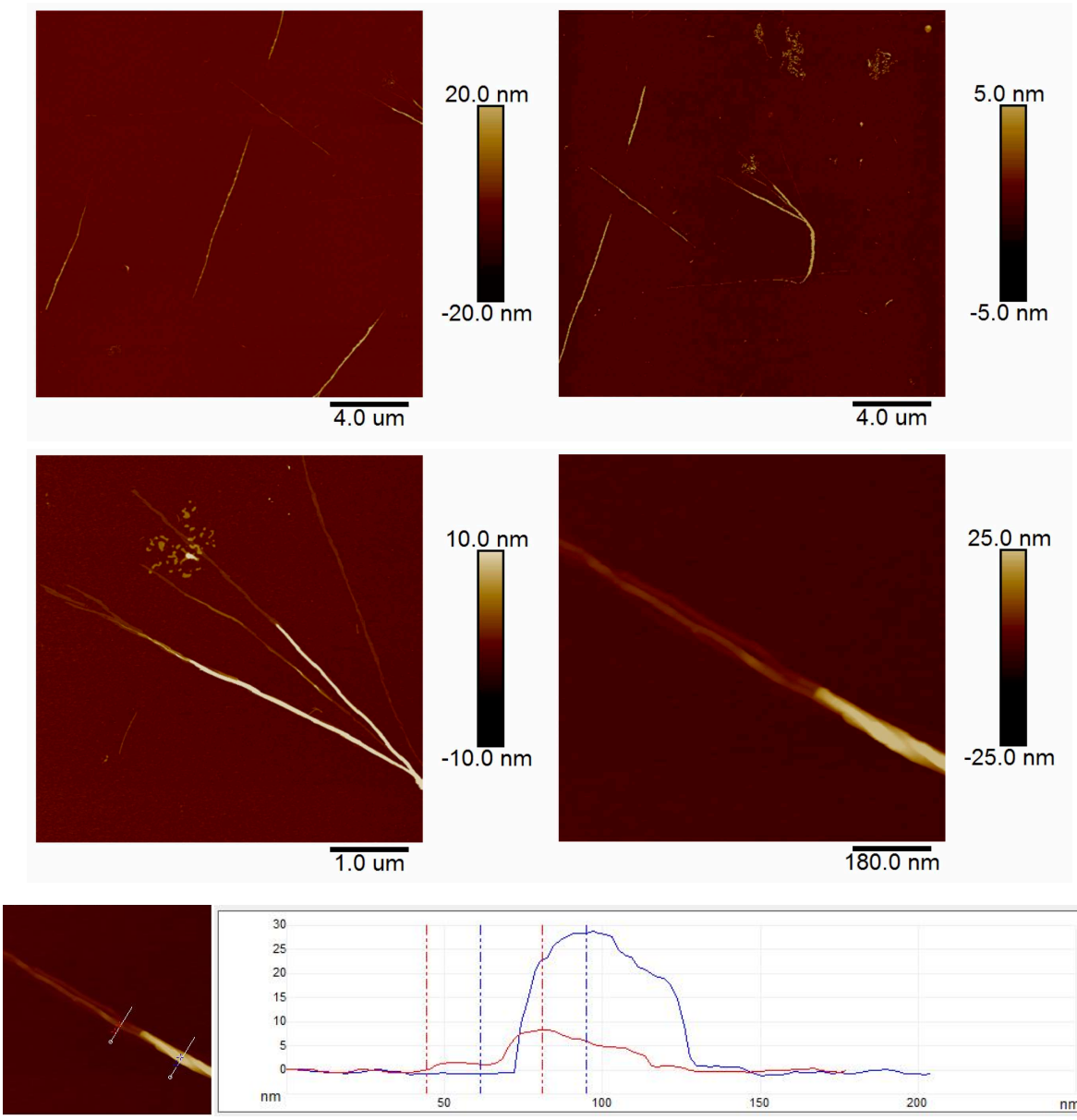


Figure 2.41. AFM images of annealed **RVFNIM** at a concentration of 0.2 mg/mL in water.

Self-assembling Behaviour of $(POG)_3$

The CD spectra of **$(POG)_3$** at 0 h showed a negative absorption at 200 nm and a positive absorption at 225 nm, suggestive of triple helix formation (Figure 2.42).[3] This result is consistent to the CD spectra reported for longer collagen mimetic peptides in the literature,[3] highlighting that it is also possible for shorter peptides to display CD spectra characteristic of collagen triple helices.

The CD spectra of **$(POG)_3$** at 5 days remained relatively unchanged with those observed at 0 h, implying there were little to no structural changes over this time period (Figure 2.42). There were only minor increases of negative signals at 200 nm detected for the peptide samples at low concentrations (0.05, 0.1 and 0.15 mg/mL). This may have been because at lower concentrations the peptide required slightly more time to reach its final self-assembled structure than when higher concentrations were used.

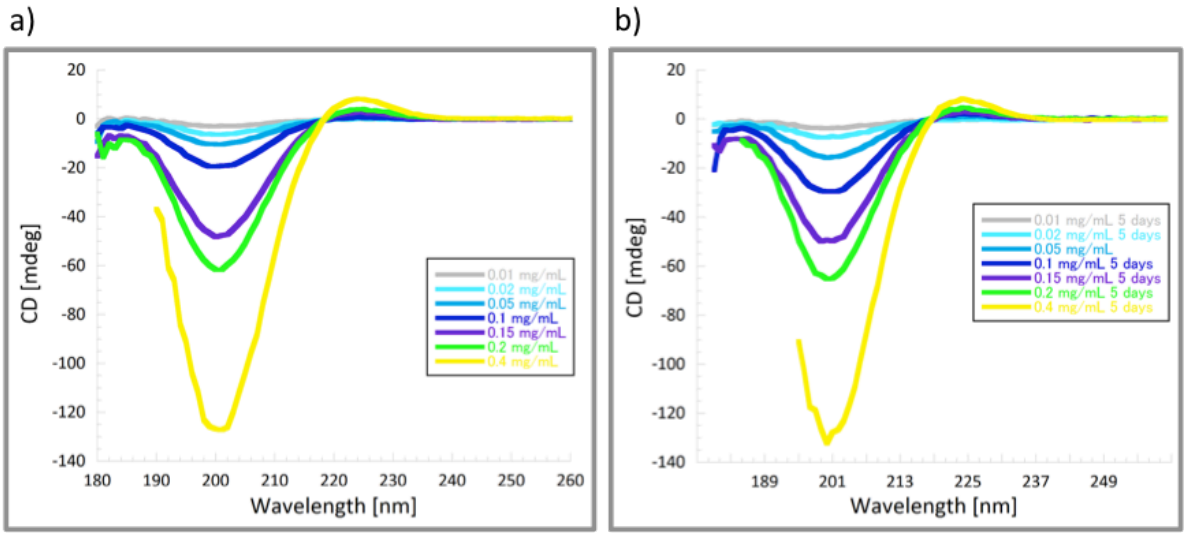


Figure 2.42. CD spectra of **a) $(POG)_3$** at 0 h and **b) $(POG)_3$** at 5 days in water. The CD spectra of **$(POG)_3$** at 0.4 mg/mL were not measured below the wavelength of 190 nm due to signal noise.

The AFM study of **$(POG)_3$** at 0.2 mg/mL and 0.4 mg/mL showed few dot-like structures at 0 h, as opposed to the expected structure of collagen fibrils predicted by the CD results obtained (Figure 2.43). Fibrils were not observed following 5 days of peptide incubation at 0.2 mg/mL,

even when the concentration was increased to 10 mg/mL, implying that **$(POG)_3$** might not have the ability to develop into fibrils.

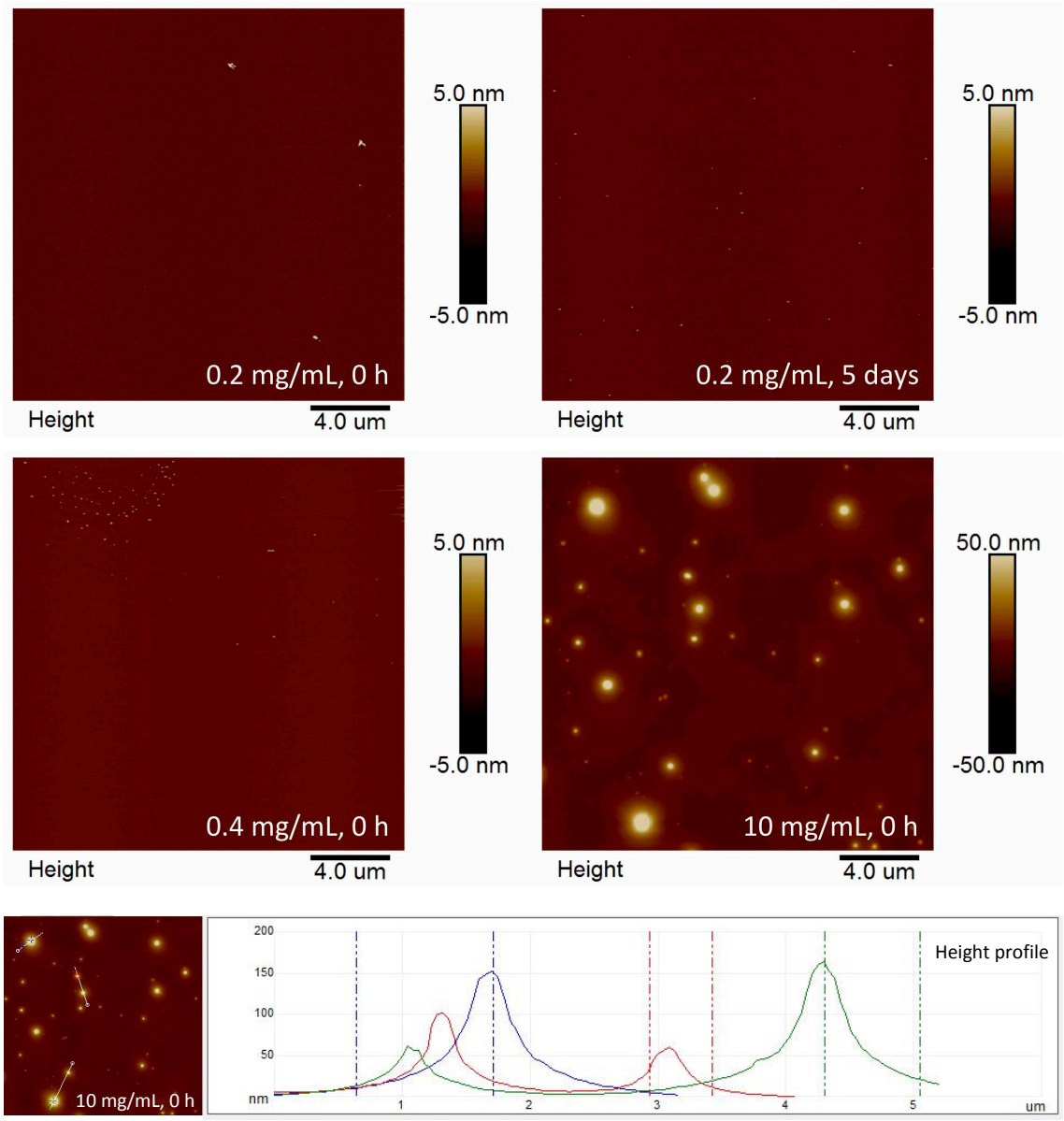


Figure 2.43. AFM images of **$(POG)_3$**.

Self-assembling behaviour of $(POG)_6$

The CD spectra of **$(POG)_6$** at 0 h was highly similar to that of **$(POG)_3$**, having also shown a negative absorption at 200 nm and a positive absorption at 225 nm, suggestive of triple helix formation (Figure 2.44).[3]

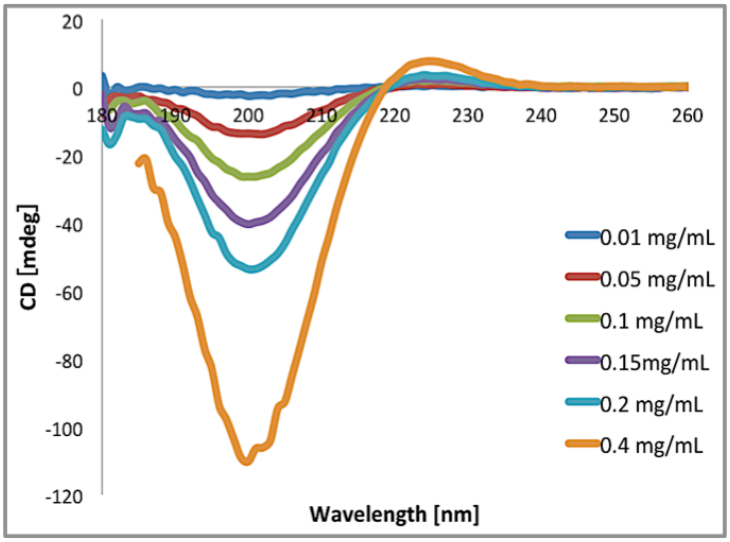


Figure 2.44. CD spectra of **$(POG)_6$** in water at concentrations of 0.01, 0.02, 0.05, 0.1 and 0.15 mg/mL.

The AFM study of **$(POG)_6$** at 0 h only showed dot-like structures at a concentration of 1 mg/mL, whereas spherical structures and few ring-like structures were observed at a concentration of 10 mg/mL (Figure 2.45). In contrast, when **$(POG)_6$** was annealed at a concentration of 10 mg/mL, a majority of ring-like structures were observed with a lesser number of spherical structures being present. In addition, a number of tape-like structures were observed to stem from the ring-like structures (Figure 2.46). It may have been possible that these tape-like structures were generated from stretching of the ring-like structures given that their heights were similar (5-6 nm for both structures). Overall, it appeared that the use of higher temperatures, as demonstrated in the annealing experiment, potentially accelerated the self-assembling process of **$(POG)_6$**, given that more ring-like structures were observed.

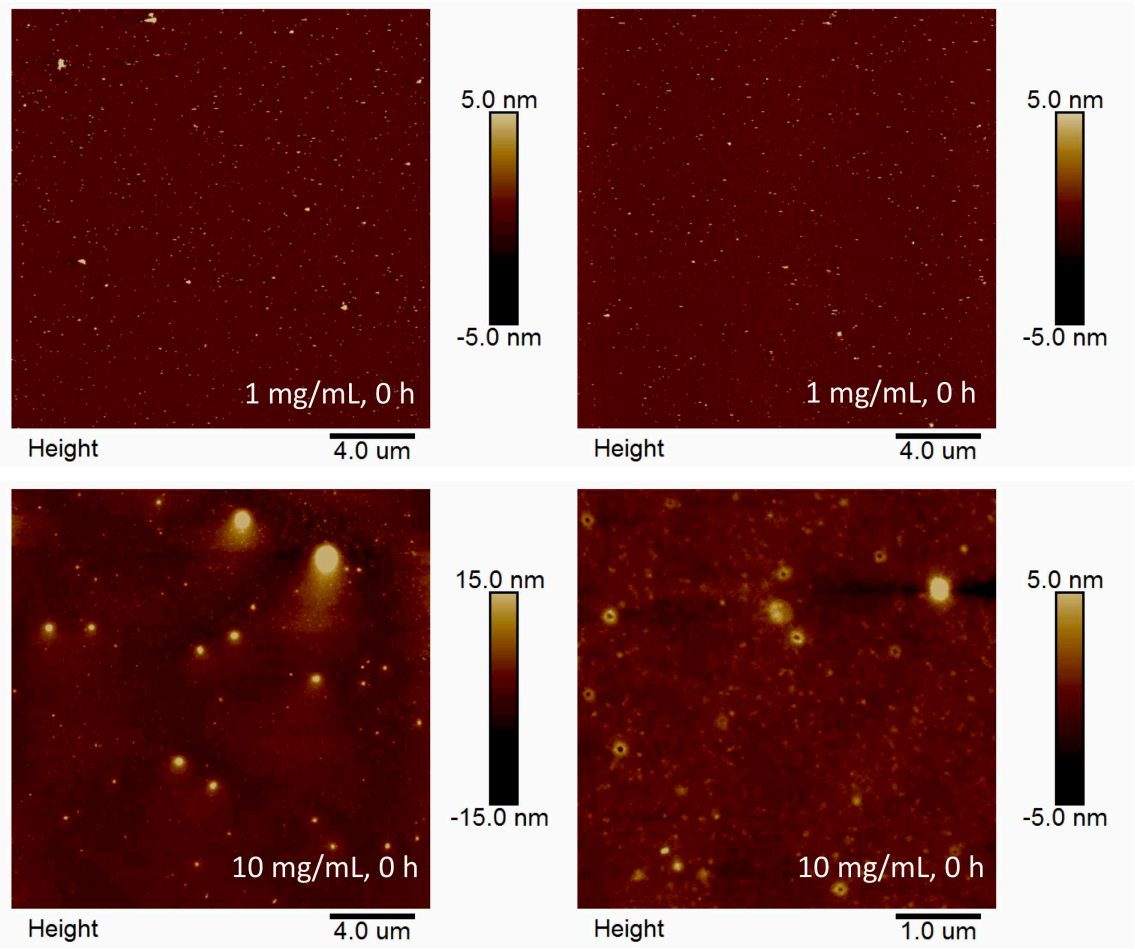
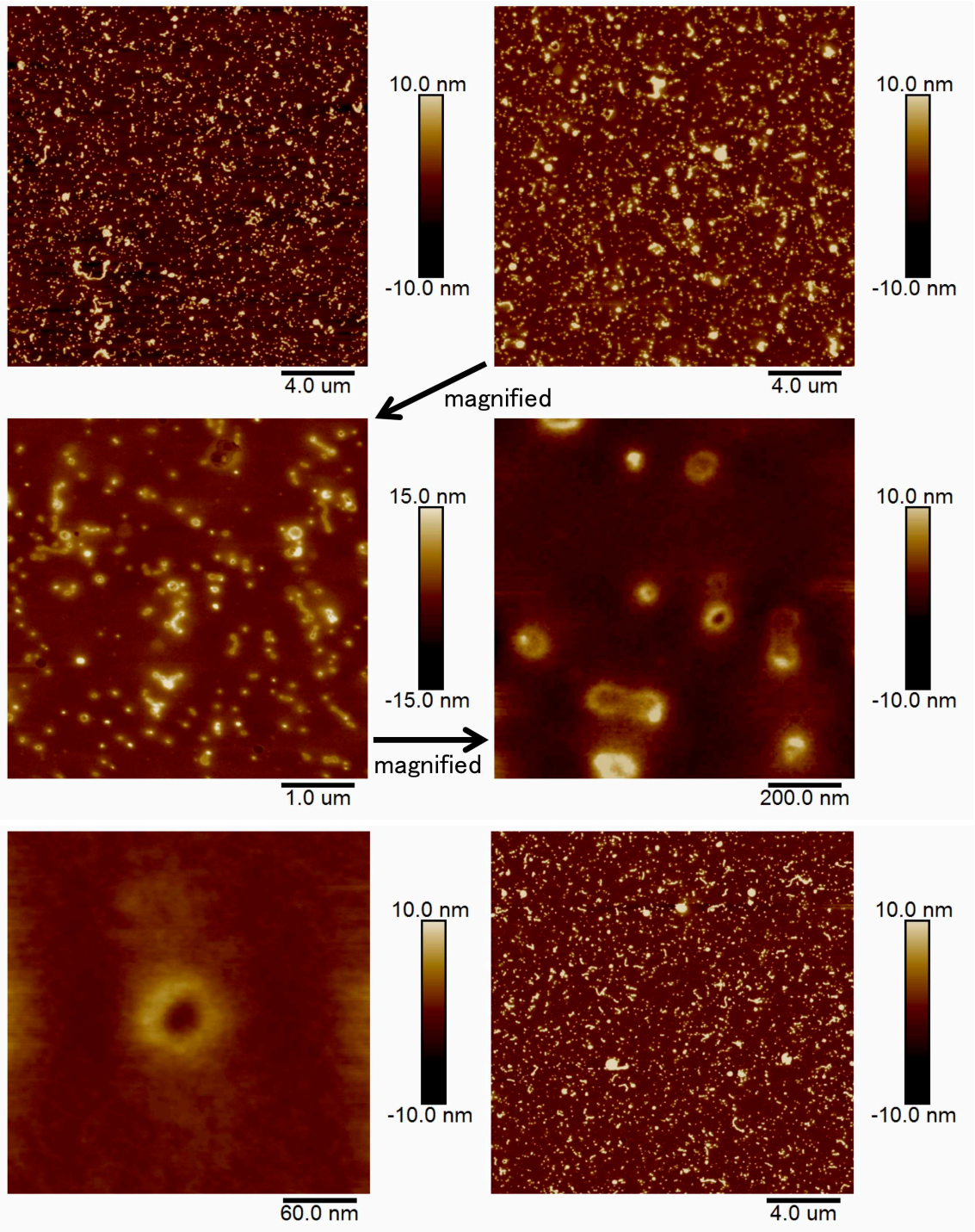
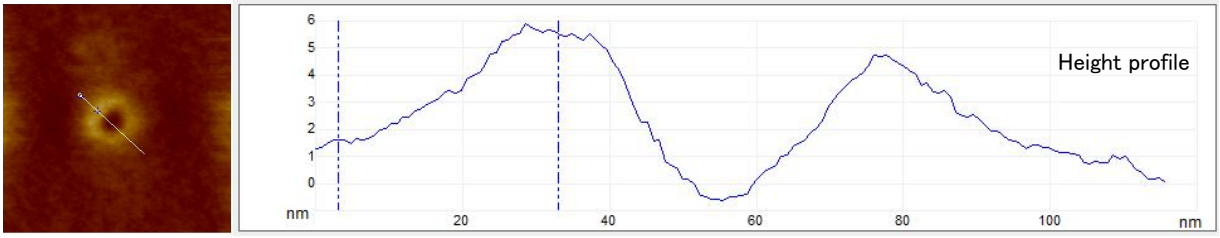
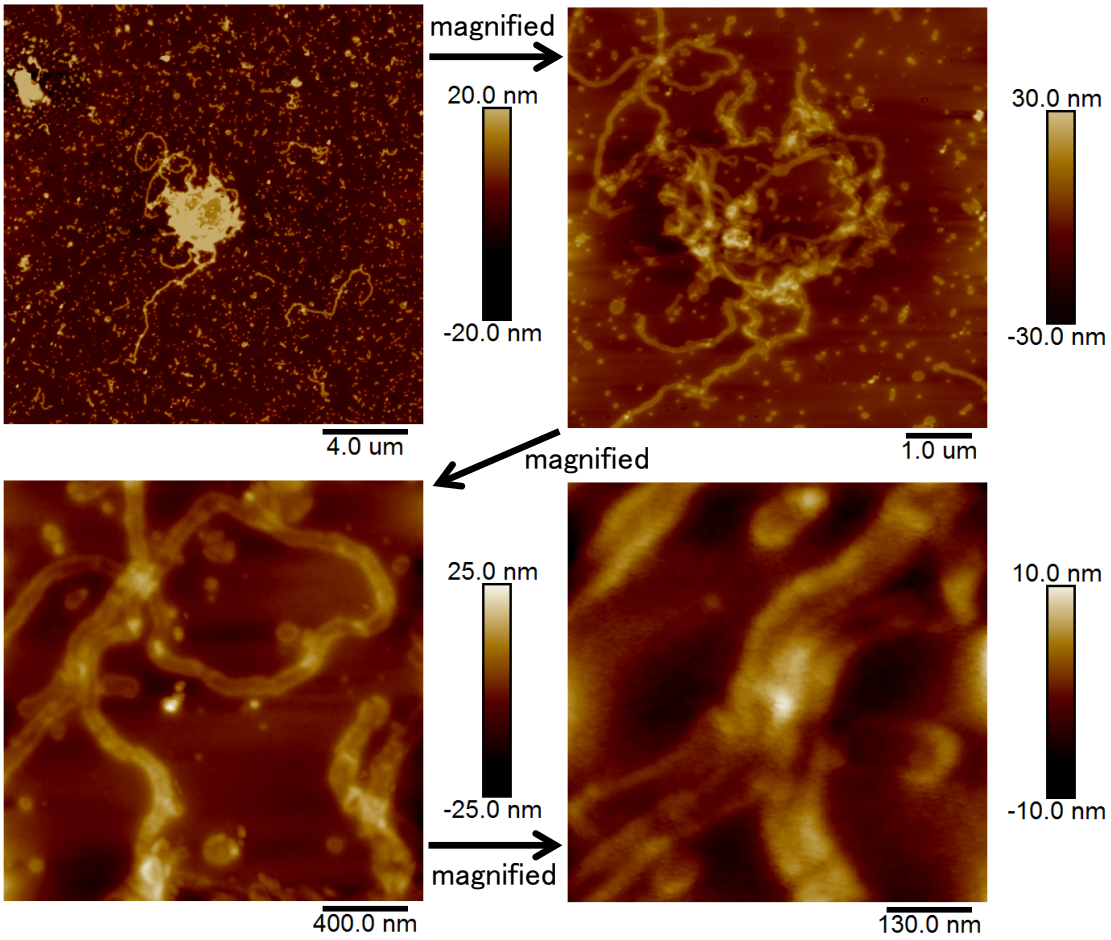
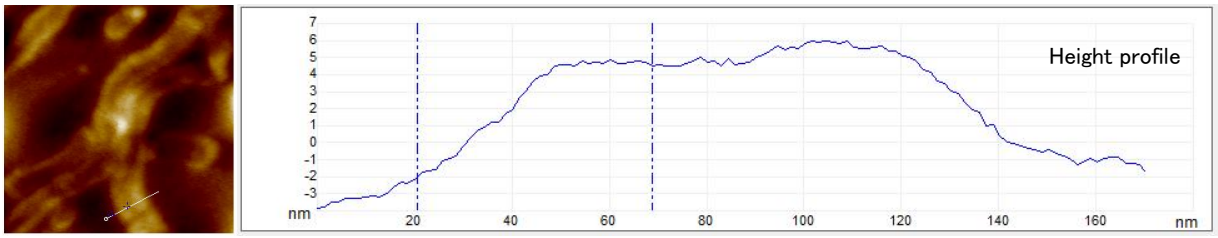


Figure 2.45. AFM images of **$(POG)_6$** at 0 h at a concentration of 1 mg/mL and 10 mg/mL.


10.0 nm
-10.0 nm
4.0 um
10.0 nm
-10.0 nm
4.0 um
magnified
15.0 nm
-15.0 nm
1.0 um
magnified
10.0 nm
-10.0 nm
200.0 nm
10.0 nm
-10.0 nm
60.0 nm
10.0 nm
-10.0 nm
4.0 um


Height profile
nm
nm


magnified
20.0 nm
-20.0 nm
30.0 nm
-30.0 nm
4.0 um
1.0 um
magnified
25.0 nm
-25.0 nm
10.0 nm
-10.0 nm
magnified
400.0 nm
130.0 nm


Height profile
nm
7
6
5
4
3
2
1
0
-1
-2
-3
20
40
60
80
100
120
140
160
nm

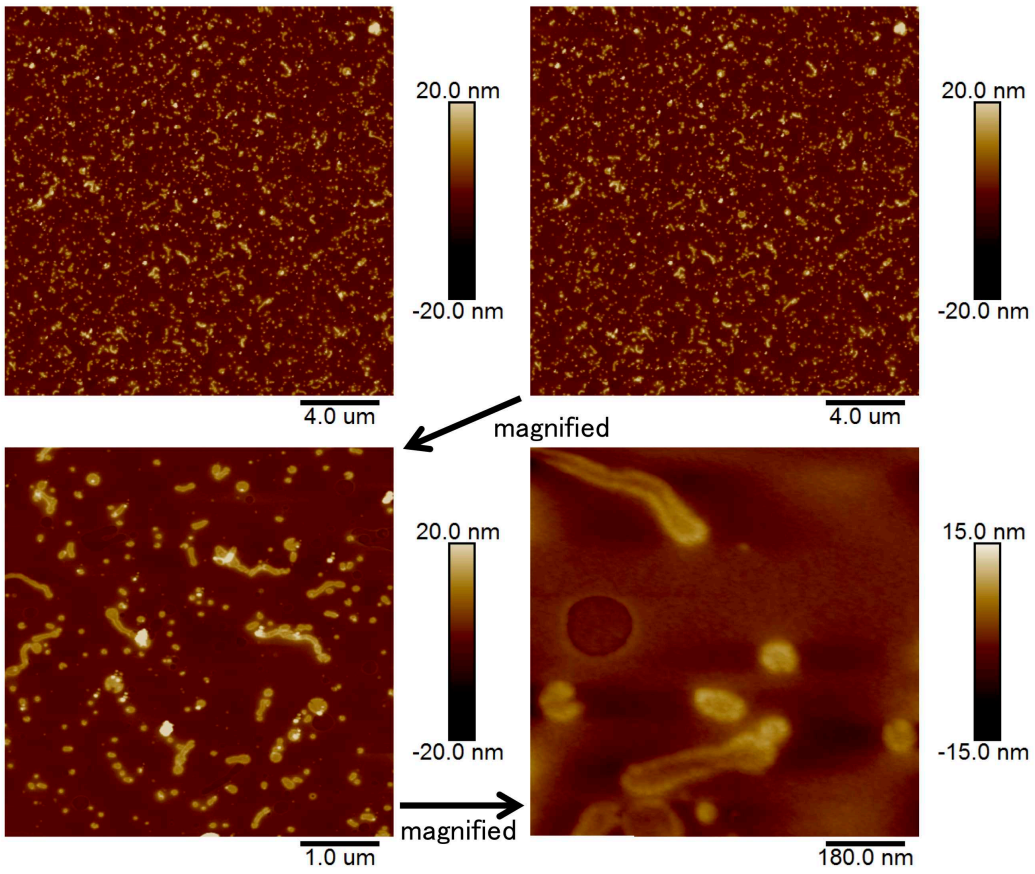


Figure 2.46. AFM images of annealed **$(POG)_6$** at a concentration of 10 mg/mL in water.

2.6 Summary

Three types of β-sheet forming peptides (**KLVFFA**, **HYFNIF** and **RVFNIM**) and collagen mimetic peptides **$(POG)_3$** and **$(POG)_6$** were successfully obtained based on NMR and/or LC-MS analysis. In addition, terminal modification and attachment of additional sequences were successfully conducted, yielding multiple modified versions of the main peptide sequences as listed in Table 2. Different crude yields were obtained among the synthesised peptides, with values ranging from 23-79%.

Overall, these β-sheet forming peptides (**KLVFFA**, **HYFNIF** and **RVFNIM**) showed their expected self-assembling behaviour based on the results of CD and AFM analysis. The AFM studies confirmed the fibre formation of these peptides following incubation in water at R.T. While the collagen mimetic peptide **$(POG)_3$** and **$(POG)_6$** may possess a characteristic of collagen triple helices based on the CD results, fibril or fibre formation from these peptides were not confirmed based on the limited results of AFM analysis.

2.7 References

1. K. L. Morris, A. Rodger, M. R. Hicks, M. Debulpaep, J. Schymkowitz, F. Rousseau and L. C. Serpell, *Biochem. J.*, 2013, **450**, 275–283.
2. J.-P. Colletier, A. Laganowsky, M. Landau, M. Zhao, A. B. Soriaga, L. Goldschmidt, D. Flot, D. Cascio, M. R. Sawaya and D. Eisenberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 16938–16943.
3. L. E. R. O'Leary, J. A. Fallas, E. L. Bakota, M. K. Kang and J. D. Hartgerink, *Nature Chemistry*, 2011, **3**, 821–828.
4. B. H. Northrop, S. H. Fraynea and U. Choudhary, *Polym. Chem.*, 2015, **6**, 3415–3430.
5. T. R. Wilks and R. K. O'Reilly, *Scientific Reports*, 2016, **6**, 1–11.
6. E. Kim and H. Koo, *Chem. Sci.*, 2019, **10**, 7835–7851.
7. Y. Xue, M. L. O'Mara, P. P. T. Surawski, M. Trau and A. E. Mark, *Langmuir*, 2011, **27**, 296–303.
8. D. A. Wellings and E. Atherton, in *Methods in Enzymology*, ed. G. B. Fields, Academic Press, San Diego, 1997, vol. 289, ch. 1, p. 54.
9. W. Troll and R. K. Cannan, *Journal of Biological Chemistry*, 1953, **200**, 803–811.
10. J. M. Palomo, *RSC Adv.*, 2014, **4**, 32658–32672.
11. T. Vojkovsky, *Pept Res.*, 1995, **8**, 236–237.
12. S. Wang, *J. Am. Chem. Soc.*, 1973, **95**, 1328–1333.
13. C. L. Tsai, S. Y. Wu, H. K. Hsu, S.B. Huang, C. H. Lin, Y. T. Chan and S. K. Wang, *Nanoscale*, 2021, **13**, 4592–4601.
14. N. J. Greenfield, *Nat Protoc.*, 2006, **1**, 2876–2890.
15. A. J. Miles and B. A. Wallace, *Chem. Soc. Rev.*, 2016, **45**, 4859–4872.
16. K. L. Shaw, G. R. Grimsley, G. I. Yakovlev, A. A. Makarov and C. N. Pace, *Protein Sci.*, 2001, **10**, 1206–1215.
17. P. Macek, R. Kerfah, E. B. Erba, E. Crublet, C. Moriscot, G. Schoehn, C. Amero and J. Boisbouvier, *Sci. Adv.*, 2017, **3**, 1–9.
18. S. S. Ray, in *Clay-Containing Polymer Nanocomposite*, Elsevier, Amsterdam, 1st edn, 2013, vol.1, ch.3, 29-66.
19. R. Wang and D. Kurouski, *Phys. Chem. C.*, 2018, 122, **42**, 24334–24340.
20. K. Lu, J. Jacob, P. Thiyagarajan, V. P. Conticello and D. G. Lynn, *J. Am. Chem. Soc.*, 2003,

125, 6391–6393.

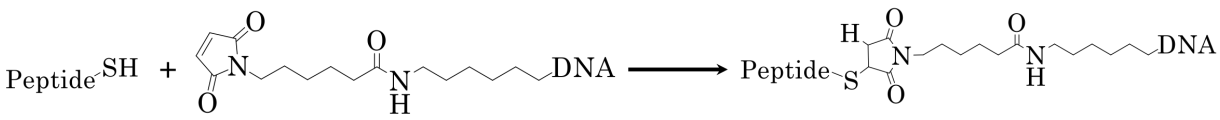
21. J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed, and R. Tycko, *Biochemistry*, 2000, **39**, 13748–13759.
22. B. Dai, *et al*., *PNAS*, 2015, **112**, 2996–3001.
23. S. I. Sampani, Y. K. Al-Hilaly, S. Malik, L. C. Serpell and G. E. Kostakis, *Dalton Trans.*, 2019, **48**, 15371–15375.
24. M. A. Kiselev, S. Magazù and P. Calandra, *Advances in Condensed Matter Physics*, 2015, **2015**, 1–22.

3 Peptide-Oligonucleotide Conjugation through Thiol-maleimide Chemistry

3.1 Chapter Overview

This chapter reports the synthetic and purification methods employed for the production of peptide-oligonucleotide conjugates (POCs) through thiol-maleimide chemistry. The first part of this chapter focuses on thiol-maleimide conjugation reactions and their outcomes, highlighting the rationale for modifying conditions at each step of the process in order to overcome the challenges that were faced to increase the conjugate yield. Thereafter, the chapter discusses the results obtained from different purification strategies that were used to isolate the desired conjugates. The purification strategies tested involved denaturing polyacrylamide gel electrophoresis (PAGE), reversed-phase high-performance liquid chromatography (RP-HPLC), anion-exchange chromatography and hydrophilic interaction chromatography (HIC).

Thiol-maleimide conjugation is used to link biomolecules together through the reaction between the cysteine group of one biomolecule with the maleimide group of another. The cysteine-containing peptides previously discussed in Chapter 2 were conjugated with two complementary DNA oligonucleotides (**DNA1** or **DNA1'**) that contained a maleimide linker, which had been synthesised by Emerald Taylor in the Serpell Group using thiol-maleimide chemistry (Scheme 1). The oligonucleotide sequence of **DNA1** was CTG TAT GGT CAA CTG TTTTT, and that of **DNA1'** was CAG TTG ACC ATA CAG TTTTT. Both DNA strands were purchased from Integrated DNA Technologies. For clarity, each thiol-maleimide conjugation reaction described in this chapter will be referred to as *TCR* (*thiol-maleimide conjugation reaction*), followed by a number (e.g. *TCR-1*). A summary of all reaction results, including those not discussed in this chapter can be found in Table 3.2.



Scheme 1. Thiol-maleimide conjugation showing a single cysteine-containing peptide reacting with maleimide-functionalised DNA, resulting in POC formation.

3.2 Methods

Thiol-maleimide conjugation reactions were carried using two maleimide-functionalised DNA oligonucleotides with complementary sequences (**DNA1-mal** and **DNA1'-mal**) and several peptides containing a single N-terminal cysteine residue (**CGSGHYFNIF**, **CHYFNIF**, **CGSGKLVFFA**, and **CGSG(POG)$_6$**) with an N-terminal acetylation and a C-terminal amidation. Reaction conditions, including temperature, solvents, reaction time, reaction volume, DNA oligonucleotide and peptide concentration were modified until the desired conjugates were obtained. The methods for each individual thiol-maleimide conjugation reaction are further described in the results section.

Reaction mixtures were analysed by denaturing PAGE to assess the presence of the desired conjugates. PAGE studies were carried out using either 10% or 20% denaturing polyacrylamide gels. 10% PAGE gels were run in 1× TBE buffer at a constant current of 15 mA and a set voltage of 300 V for 20 min. 20% PAGE gels were run in 1× TBE buffer at a constant current of 15 mA and a set voltage of 300 V for 60 min. Gels were stained in Stains-All stain prepared in 1× TBE buffer for a minimum of 20 min. Small amounts of HFIP were used in reactions and PAGE studies to dissolve peptides (**CGSGHYFNIF** and **CGSGKLVFFA**) with poor solubility in aqueous solution. For *TCR-39*, 12% native PAGE was run in 1× TAMg buffer at a constant current of 15 mA and a set voltage of 300 V for 90 min. Band intensity of TCR-32 and TCR-33 was quantified using ImageJ and the band intensity ratio between conjugate and unreacted DNA bands was calculated.

Anion Exchange Chromatography

Anion exchange chromatography is a type of ion exchange chromatography that separates molecules based on their net surface charge.[1] This type of chromatography uses a positively charged ion exchange resin that binds molecules with net negative surface charges. Relying on electrostatic interactions, this technique allows for separation of a wide range of molecules, including oligonucleotides, peptides, proteins and possibly POCs. DNA is negatively charged, whereas our peptides are neutral or positively charged in the buffer used. Assuming that DNA and POCs have different affinities to the resin, the removal of

unreacted DNA and excess peptide was attempted by subsequent washing with buffer and a linear gradient of NaCl solution.

TCR-30

50 µL of Q-Sepharose resin (containing 20% ethanol) was equilibrated by washing with pH 8 phosphate buffer (100 µL ×4). 25 µL of *TCR-30* in pH 6.8 phosphate buffer was added to the resin and incubated for 10 min to facilitate conjugate binding to the resin. Following incubation, the resin was centrifuged and the supernatant was collected as fraction 1. Thereafter, the resin was washed with pH 8 phosphate buffer (100 µL × 2) to remove the peptide that would have less affinity to the resin than the conjugate. After each washing cycle, the resin was centrifuged and the supernatant was collected as fractions 2 and 3, respectively. The resin was washed further with 1M NaCl (100 µL × 2), followed by 2M NaCl washing (100 µL × 2) in an attempt to remove the unreacted DNA-mal and the desired conjugate from the resin at different points. The UV-Vis of each collected fraction was measured by a NanoDrop spectrometer to determine the respective DNA absorbance. Fractions were subsequently analysed by 10% denaturing PAGE to evaluate whether successful separation between unreacted DNA-mal and the desired conjugate had occurred.

TCR-42

50 µL of Q-Sepharose resin (containing 20% ethanol) was equilibrated by washing with pH 8 phosphate buffer (100 µL ×4). *TCR-42* precipitate (suspended in 50 µL of pH 9 borate buffer) was added to the resin and incubated for 10 min, to allow for the conjugate to bind to the resin. After the incubation, the resin was centrifuged and the supernatant was collected as fraction 1.

The resin was washed with pH 8 phosphate buffer (50 µL × 3) to remove the peptide that had less affinity to the resin than the conjugate. After each round of washing, the resin was centrifuged and the supernatant was collected as fractions 2-4, respectively. Subsequently, the removal of the desired conjugate from the resin was attempted by washing with a linear NaCl gradient (0.1-2M). After each round of washing, the resin was centrifuged and the

supernatant was collected as fractions 5-9, respectively. Each fraction was quantified based on A260 using a NanoDrop and fractions that showed DNA absorption were analysed by 10% denaturing PAGE.

Denaturing Urea PAGE

Denaturing urea PAGE employs 6-8 M urea, which denatures secondary oligonucleotide structures, and is used for their separation in a polyacrylamide gel matrix.[2] PAGE is then used to separate oligonucleotides based on their length to allow for differentiation of full-length products from shorter by-products.[3]

A 20% denaturing PAGE gel of 20 cm × 22 cm × 1.5 mm was produced by mixing 50 mL of 20% polyacrylamide solution containing 8M urea with 50 µL of TEMED and 130 µL of 40% APS solution. The gel was poured between two glass plates and then a 1.5 mm comb was inserted. After the gel was polymerised, the comb was removed and the well was washed with 1× TBE buffer. Prior to loading the sample, the gel was pre-run at constant current of 30 mA and a set voltage of 300 V for 1 h in 1× TBE buffer (pH 8) in order to equilibrate and preheat the gel.[4] 260 µL of 224 µM *TCR-34* was mixed with 8M urea (1:1, v/v) and heated at 90 °C for 15 min. The sample was loaded onto the gel, which was then run at constant current of 30 mA and a set voltage of 250 V for 30 min, followed by a set voltage of 300 V for 2h in 1× TBE buffer.

After the completion of electrophoresis, the gel was removed from the glass plates and wrapped in cling film, which was then placed on to a silica TLC plate. The gel was illuminated with a UV lamp (254 nm) to visualise the bands that displayed shadows on the TLC plate. The top band was excised and crushed in a 15 mL falcon tube. 8 mL of autoclaved water was added to the gel pieces, which were then frozen in liquid nitrogen, followed by incubation at 55 °C for 17 h. After the solution was centrifuged for 5 min, the supernatant was collected and the precipitate was washed with 2 mL of autoclaved water 3 times. The collected supernatants were dried by a SpeedVac vacuum concentrator at 60°C for 3 h. The dried samples were re-suspended in a total of 1 mL of autoclaved water, which was then desalted

using Zetadex-25 (10 mL). The eluent from the Zetadex column was collected as Fraction 1 (F1). The column was washed with 1 mL of autoclaved water, and the eluent was collected as F2. The washing was repeated 14 times and each eluent was labeled as F3-F16, respectively. Each fraction was quantified based on A260 using a NanoDrop spectrometer. F3-F6 was analysed by 20% analytical PAGE to check for the presence of the desired conjugate.

RP-HPLC Purification

RP-HPLC purification separates molecules based on their hydrophobicity using a non-polar stationary phase, while aqueous buffers and organic solvents are used in the mobile phase to elute the analytes.[3] RP-HPLC purification was carried out on an Agilent 1100 system using a C18 column at a flow rate of 1 mL/min, with solvent systems: 0.1% TFA/water (solvent A) and 80% ACN containing 0.09% TFA (solvent B). A linear gradient (10-100%) of solvent B was used over the course of 50 min to elute the sample (Table 3.1). 500 µL of sample was used per RP-HPLC run. The oven temperature of the column and the detection wavelength were set at 40°C and at 260 nm, respectively.

Table 3.1. RP-HPLC solvent gradient

Time (min)	Solvent A (%)	Solvent B (%)
0	90	10
5	90	10
10	80	20
40	40	60
50	0	100
51	90	10
70	90	10

TCR-37

TCR-37 was prepared for RP-HPLC purification by reacting **DNA1'-mal** (396 µM) with **CGSGKLVFFA** (7.9 mM) in unbuffered water at R.T. overnight (total reaction volume: 520 µL). LC-MS and 20% PAGE were conducted to determine the presence of the desired conjugate.

TCR-37 was diluted 5 times prior to being injected to the RP-HPLC system. UV-Vis measurements of the fractions were conducted to assess for the presence of DNA.

TCR-39

TCR-39 was prepared for RP-HPLC purification by reacting **DNA1'-mal** (472 μM) with **CGSGHYFNIF** (7.1 mM) in unbuffered water at R.T. overnight (total reaction volume: 400 μL). *TCR-39* was diluted 4 times prior to being injected to the RP-HPLC system. UV-Vis measurements of the fractions were conducted to assess for the presence of DNA. These fractions were individually analysed by LC-MS to determine whether the desired conjugate was present. In addition, **DNA1'-mal** and **CGSGHYFNIF** were run by RP-HPLC, acting as controls to compare the position of their peaks with the conjugate peaks.

TCR-46

TCR-46 was prepared by incubating 20 μL of *TCR-45* (described in Table 3.2) in 180 μL of pH 9 borate buffer at R.T. overnight. This reaction mixture was used in the experiment that tested the stability of DNA under RP-HPLC purification conditions. To mimic the RP-HPLC gradient used for purification, 0.1% TFA/water (solvent A) and 80% ACN/water containing 0.09% TFA (solvent B) were prepared. 1 μL of **DNA1** (100 μM) was diluted 10-fold in a PCR tube with either Milli-Q water, or with solvent A to solvent B in the following ratios: 20:80, 40:60 and 60:40, respectively. In addition, 5 μL of *TCR-46* (32 μM) was diluted 6-fold in a PCR tube with either Milli-Q water, or with solvent A to solvent B in the following ratios: 20:80, 40:60 and 60:40, respectively. The samples were then incubated at 40 °C for 1h in a thermocycler, followed by UV-Vis measurements and 10% denaturing PAGE analysis. **DNA1'-mal** run by RP-HPLC was also analysed by UV-Vis spectroscopy and 10% denaturing PAGE, which were compared with the results of *TCR-46*.

Purification by HIC Chromatography

HIC is a technique commonly used for purifying proteins in their native state.[5] Proteins are separated based on differences in their surface hydrophobicity through the interaction between the proteins and the hydrophobic surface of a HIC resin. The retention time is

determined by adsorption-desorption equilibrium in the presence of salts. A high salt concentration increases the hydrophobic interaction between the proteins and the HIC medium whereas lowering the salt concentration reduces this interaction. With this principle, when the salt concentration is reduced, the protein with the lowest degree of hydrophobicity elutes first, while the protein with the highest hydrophobicity elutes last.

PureCube Butyl Agarose resin containing 20% ethanol was equilibrated by washing with 2M Na_2SO_4 four times. The conjugate sample was suspended in the resin for 20 min, which was then centrifuged at 8,860 rpm at R.T. for 1 min. The supernatant was collected as a fraction. The resin was then washed twice with 2M Na_2SO_4 (100 µL) to remove unreacted DNA. The salt washing step (50 µL) was repeated by gradually lowering Na_2SO_4 concentrations (1.5 M - 0 M). After each washing cycle, the resin was centrifuged and the supernatant was collected in a microtube. Each fraction was quantified based on A260 using a NanoDrop spectroscopy. Fractions which showed DNA absorption were analysed by 10% denaturing PAGE to check for the presence of the desired conjugate.

For *TCR-51*, 100 µL of resin was used to purify 80 µL of supernatant fraction 2. For *TCR-52*, 60 µL of resin was used to purify 50 µL of supernatant fraction 3. For *TCR-53*, 60 µL of resin was used to purify 50 µL of supernatant fraction 3. For *TCR-54*, 150 µL of resin was used to purify 50 µL of supernatant fraction 2.

3.3 Results and Discussion

3.3.1 Reactions in TBE Buffer (*TCR-1*)

Peptide **CGSGHYFNIF** (100 µM) was reacted with **DNA1-mal** (10 µM) or with **DNA1'-mal** (10 µM) in TBE buffer at pH 8 (total reaction volume: 850 µL). Reactions were either conducted at R.T. or 37°C and allowed to sit overnight. The 20% PAGE study did not show the appearance of any new bands, only showing presence of unreacted DNA (Figure 3.1). Given this result, it may have been possible that the amount of peptide used in *TCR-1* was insufficient to have reacted with DNA. Another possible explanation could have been that the slightly basic pH prevented the maleimide group of DNA from reacting with the thiol

group of the peptide, given that the optimal pH of this reaction has been previously reported to be between 6.5-7.5.[6,7]

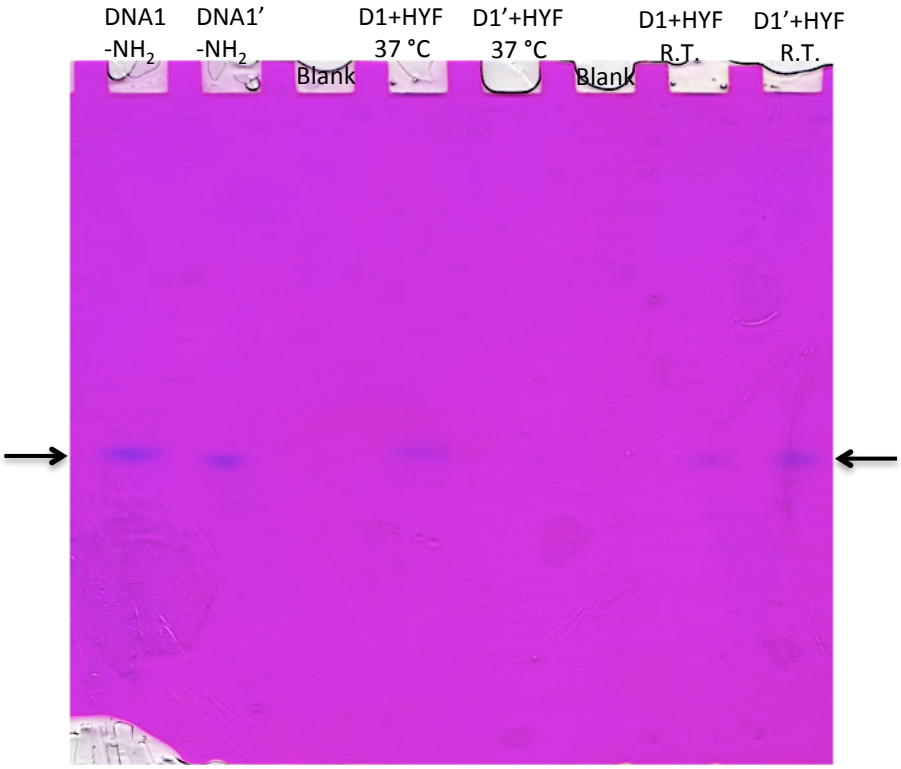


Figure 3.1. 20% denaturing PAGE of *TCR-1* showing unreacted DNA bands. The band positions are indicated with arrows.

3.3.2 Reactions under Lower pH (*TCR-2*)

To improve the reactivity of the maleimide group of DNA with the thiol group of the peptide, reactions were carried out at lower pH conditions compared with *TCR-1* using unbuffered water or pH 7.4 phosphate-buffered saline (PBS) buffer. HFIP was used to help dissolve the peptide in solution (water or PBS:HFIP = 19:1, v/v). Moreover, the concentration of peptide **CGSGHYFNIF** (4.2 mM) used in *TCR-2* was substantially higher than in *TCR-1*, which was reacted with **DNA1-mal** (10 µM) or with **DNA1'-mal** (10 µM) in water/HFIP or PBS buffer/HFIP at 37°C, and allowed to sit overnight (total reaction volume: 200 µL).

The 20% denaturing PAGE study showed two new bands above the band of unreacted **DNA-mal** in all tested reaction mixtures (Figure 3.2), suggesting that the peptide and DNA partially reacted, as well as implying the formation of a by-product. It may be possible that

the peptide concentration used in *TCR-2* was excessive, meaning that the peptide reacted with DNA, as well with impurities that might have been present in the reaction mixture, resulting in the formation of these two additional bands. However, the reason why these new bands appeared on the gel was not investigated given that the priority of the research was to check the reproducibility of conjugation reactions and to obtain the desired conjugates.

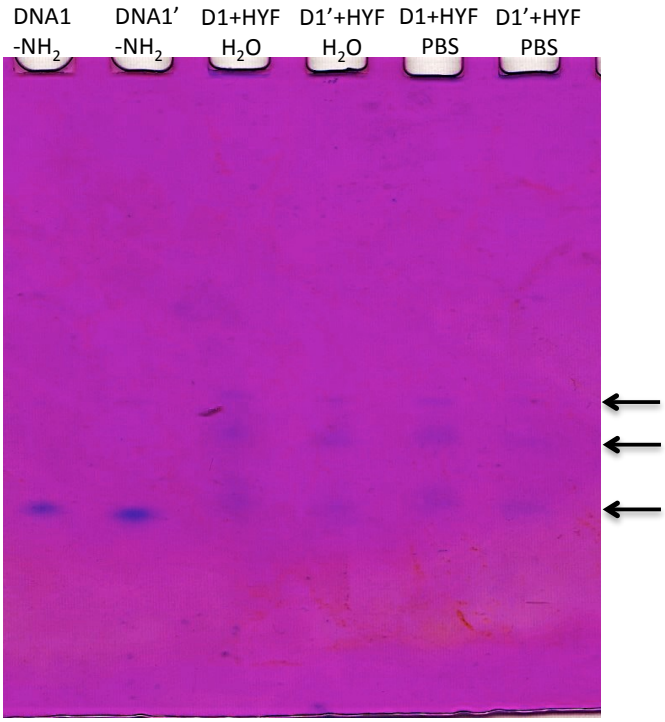


Figure 3.2. 20% denaturing PAGE of *TCR-2* showing two new bands above the bands of unreacted DNA. The band positions are indicated with arrows.

3.3.3 Determining Appropriate Peptide Concentration (*TCR-3* to *TCR-6*)

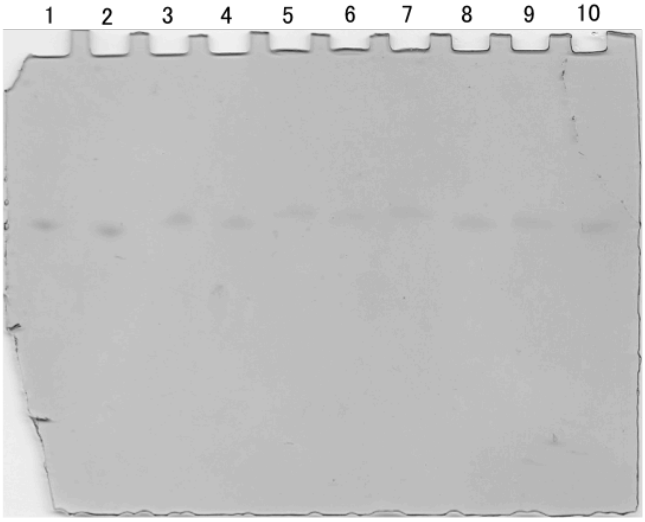
The following conjugation reactions were aimed at determining the optimal concentration of peptide for reacting with DNA. In *TCR-3*, four different concentrations of **CGSGHYFNIF** (10, 20, 30 and 40 equivalents of peptide) were reacted with 10 µM **DNA1-mal** or **DNA1'-mal**. In *TCR-4*, *TCR-5* and *TCR-6*, higher concentrations of **CGSGHYFNIF** (100, 200, 300 and 400 equivalents of peptide) were reacted with 10 µM **DNA1-mal** or **DNA1'-mal**.

The reactions were carried out at R.T. overnight in pH 7.4 PBS/HFIP (4:1, v/v) for *TCR-3*, *TCR-4* and *TCR-5*, and in unbuffered water for *TCR-6* (**CGSGHYFNIF** showed better solubility

in water compared with PBS buffer, which allowed for conjugation reactions without the use of HFIP. Thus, from *TCR-6* onwards, conjugation reactions were carried out in water). The reaction mixtures were individually analysed using 20% denaturing PAGE. For reactions where large amounts of peptide precipitate were obtained, centrifugation was conducted for separating the supernatant (SN) from the precipitate (PPT) prior to PAGE analysis.

The formation of the desired conjugates was not confirmed by 20% denaturing PAGE in any of the reactions conducted. *TCR-3* only showed bands of unreacted **DNA-mal**, suggesting that 40 equivalents of **CGSGHYFNIF** were not enough to react with **DNA-mal** (Figure 3.3). Meanwhile, *TCR-4* (Figure 3.4) and *TCR-5* did not show bands on the gels, suggesting that **DNA-mal** disappeared before it was analysed by PAGE. Given that excess amounts of **CGSGHYFNIF** (100, 200, 300 and 400 equivalents to 10 μM **DNA-mal**) were used in these reactions, **DNA-mal** might have been trapped within the peptide precipitate that had formed in the reaction mixture, resulting in the sample not being loaded to the gels.

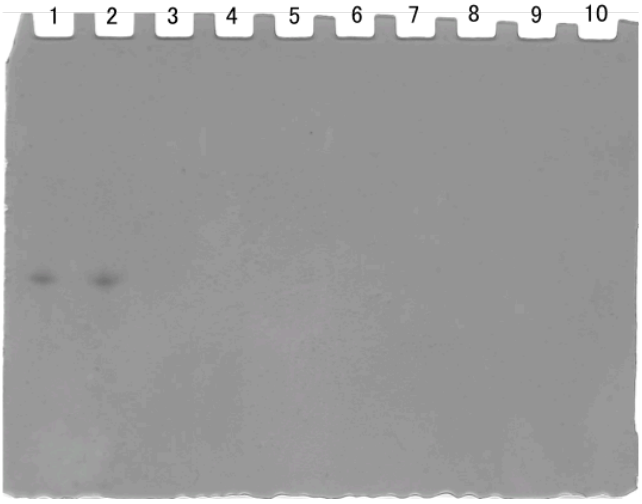
In *TCR-6* precipitate, stained wells were observed, confirming the formation of aggregates that did not migrate to the gel (Figure 3.5). Given that no bands were observed in the reaction mixtures of *TCR-4*, *TCR-5* and in *TCR-6* supernatant, it may have been possible that the high concentration of **CGSGHYFNIF** led to precipitation of the conjugate. In addition, the aggregates in the reaction mixtures may have in turn prevented **DNA-mal** from reacting with the peptide.



Well No.	Sample
1	10 μM DNA1-NH_2 control
2	10 μM DNA1'-NH_2 control
3	DNA1-mal + HYF (40 equiv.)
4	DNA1'-mal + HYF (40 equiv.)
5	DNA1-mal + HYF (30 equiv.)
6	DNA1'-mal + HYF (30 equiv.)
7	DNA1-mal + HYF (20 equiv.)
8	DNA1'-mal + HYF (20 equiv.)
9	DNA1-mal + HYF (10 equiv.)
10	DNA1'-mal + HYF (10 equiv.)

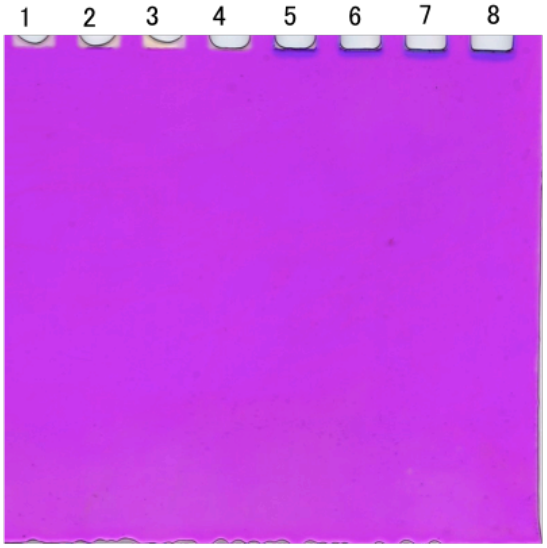
Figure 3.3. 20% denaturing PAGE of *TCR-3* showing unreacted DNA bands. HYF stands for

CGSGHYFNIF.



Well No.	Sample
1	10 μM DNA1-NH_2 control
2	10 μM DNA1'-NH_2 control
3	DNA1-mal + HYF (400 equiv.)
4	DNA1'-mal + HYF (400 equiv.)
5	DNA1-mal + HYF (300 equiv.)
6	DNA1'-mal + HYF (300 equiv.)
7	DNA1-mal + HYF (200 equiv.)
8	DNA1'-mal + HYF (200 equiv.)
9	DNA1-mal + HYF (100 equiv.)
10	DNA1'-mal + HYF (100 equiv.)

Figure 3.4. 20% denaturing PAGE of *TCR-4* only showed DNA control bands. HYF stands for CGSGHYFNIF.



Well No.	Sample
1	SN: DNA1-mal + HYF (400 equiv.)
2	SN: DNA1-mal + HYF (200 equiv.)
3	SN: DNA1'-mal + HYF (400 equiv.)
4	SN: DNA1'-mal + HYF (200 equiv.)
5	PPT: DNA1-mal + HYF (400 equiv.)
6	PPT: DNA1-mal + HYF (200 equiv.)
7	PPT: DNA1'-mal + HYF (400 equiv.)
8	PPT: DNA1'-mal + HYF (200 equiv.)

Figure 3.5. 20% denaturing PAGE of *TCR-6* did not show any bands. Wells 5-8 were stained in blue. HYF stands for CGSGHYFNIF.

3.3.4 Reactions with Different Peptides (*TCR-8* and *TCR-9*)

In *TCR-8* and *TCR-9*, three different peptides (**$C(POG)_6$**, **CHYFNIF** and **CGSGHYFNIF**) were reacted with **DNA1'-mal** using 10, 50, 100 and 200 equivalents to 10 μM DNA. New faint bands above unreacted DNA appeared for peptides **CHYFNIF** (at concentrations ≧100 equivalents of peptide to DNA) and **CGSGHYFNIF** (at concentrations ≧50 equivalents of peptide to DNA) when conjugation was conducted in water, but not in phosphate or PBS

buffer (Figure 3.6). The presence of a new band above DNA control suggested that **DNA1'-mal** partially reacted with these two peptides in water. This might be explained by the fact that **CHYFNIF** and **CGSGHYFNIF** were more soluble in water than phosphate or PBS buffer, allowing for better conversion to the conjugate. Meanwhile, the PAGE study of **C(POG)$_6$** only showed unreacted DNA bands, indicating that the conjugate was not formed.

Well No.	Sample (reaction conditions)
1	10 μM DNA1'-NH_2 control
2	DNA1'-mal + 200 equiv. peptide in PBS buffer
3	DNA1'-mal + 200 equiv. peptide in phosphate buffer
4	DNA1'-mal + 10 equiv. peptide in water
5	DNA1'-mal + 50 equiv. peptide in water
6	DNA1'-mal + 100 equiv. peptide in water
7	DNA1'-mal + 200 equiv. peptide in water

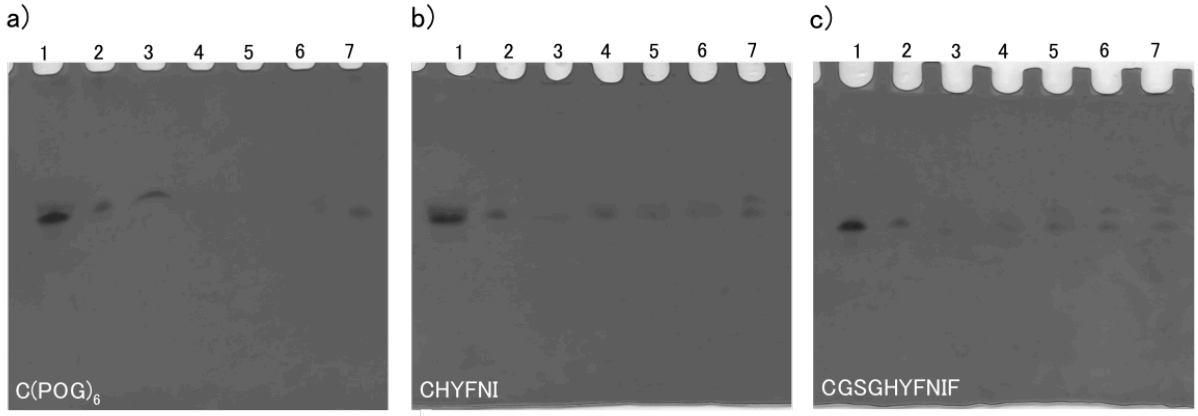


Figure 3.6. 20% denaturing PAGE of *TCR-9* showing the results of conjugation reactions with three different peptides: **a)** C(POG)$_6$, **b)** CHYFNIF and **c)** CGSGHYFNIF.

3.3.5 Reactions to Improve the Conjugate Conversion Yield (*TCR-10* to *TCR-29*)

The objective of the following conjugation reactions focused on improving the conversion yield of DNA to conjugate. To undertake this, different reaction conditions were explored, including changes in temperature, pH and solvent, as well as the use of a reducing agent (TCEP). Despite this, PAGE studies at best only showed presence of unreacted DNA, with some studies showing no bands, highlighting that there was no improvement of the conversion yield. Consequently, amide coupling was tested as an alternative conjugation

strategy (see Chapter 4) for several months, until discovering the reason why thiol-maleimide conjugation reactions had not been successful.

3.3.6 Rationale for Restarting Thiol-maleimide Conjugation

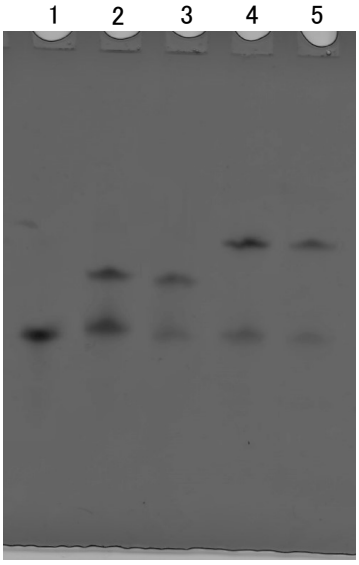
While amide-coupling experiments were being conducted, Emerald Taylor in the Serpell Group successfully identified the key causes as to why the conjugates were not obtained from the latest thiol-maleimide conjugation experiments. Firstly, it was discovered that the maleimide linker had not been present in the DNA samples, explaining why the peptides did not react with DNA. In addition, given that the DNA concentration was found to be substantially lower than what was initially predicted, it explained why some PAGE studies did not show any bands. Following the identification of these issues and once the DNA was successfully resynthesised with a maleimide linker, the conjugation reactions were undertaken again.

3.3.7 Conjugation Reactions with DNA Containing a Maleimide Linker (*TCR-30*)

Based on previous conjugation reactions, it was suggested that DNA-mal would react with peptide in a 1:20 ratio in aqueous solution at around pH 7. Following this, peptides **CGSGHYFNIF** (2.5 mM) and **CGSG(POG)$_6$** (2.5 mM) were reacted with **DNA1-mal** (50 μM) or with **DNA1'-mal** (50 μM) in pH 6.8 phosphate buffer at 37°C and allowed to sit overnight (total reaction volume: 200 μL). The PAGE study showed a new band above the DNA control band, suggesting conjugate formation (Figure 3.7), which was further confirmed by LC-MS (Figure 3.8, 3.9, 3.10 and 3.11). The conjugate band for **DNA-CGSG(POG)$_6$** was observed at a higher position than that of **DNA-CGSGHYFNIF** since the molecular mass of **DNA-CGSG(POG)$_6$** was larger. However, the presence of unreacted DNA bands indicated that the DNA to conjugate conversion was incomplete.

Given that some DNA oligonucleotides were shown to have ring-opened maleimide groups by LC-MS, this may have explained why DNA did not fully react with the peptide. Since the hydrolytic stability of maleimide groups decreases at higher temperatures, having conducted the conjugation reaction at 37 °C could have contributed to the hydrolysis of the

maleimide group.[8] Consequently, lowering the temperature of the conjugation reactions was hypothesised to result in an improved conjugate conversion yield.



Well No.	Sample
1	10 µM DNA1'-mal control
2	DNA1-mal + CGSGHHYF
3	DNA1'-mal + CGSGHHYF
4	DNA1-mal + $CGSG(POG)_6$
5	DNA1'-mal + $CGSG(POG)_6$

Figure 3.7. 20% denaturing PAGE of *TCR-30,* showing unreacted DNA and conjugate bands.

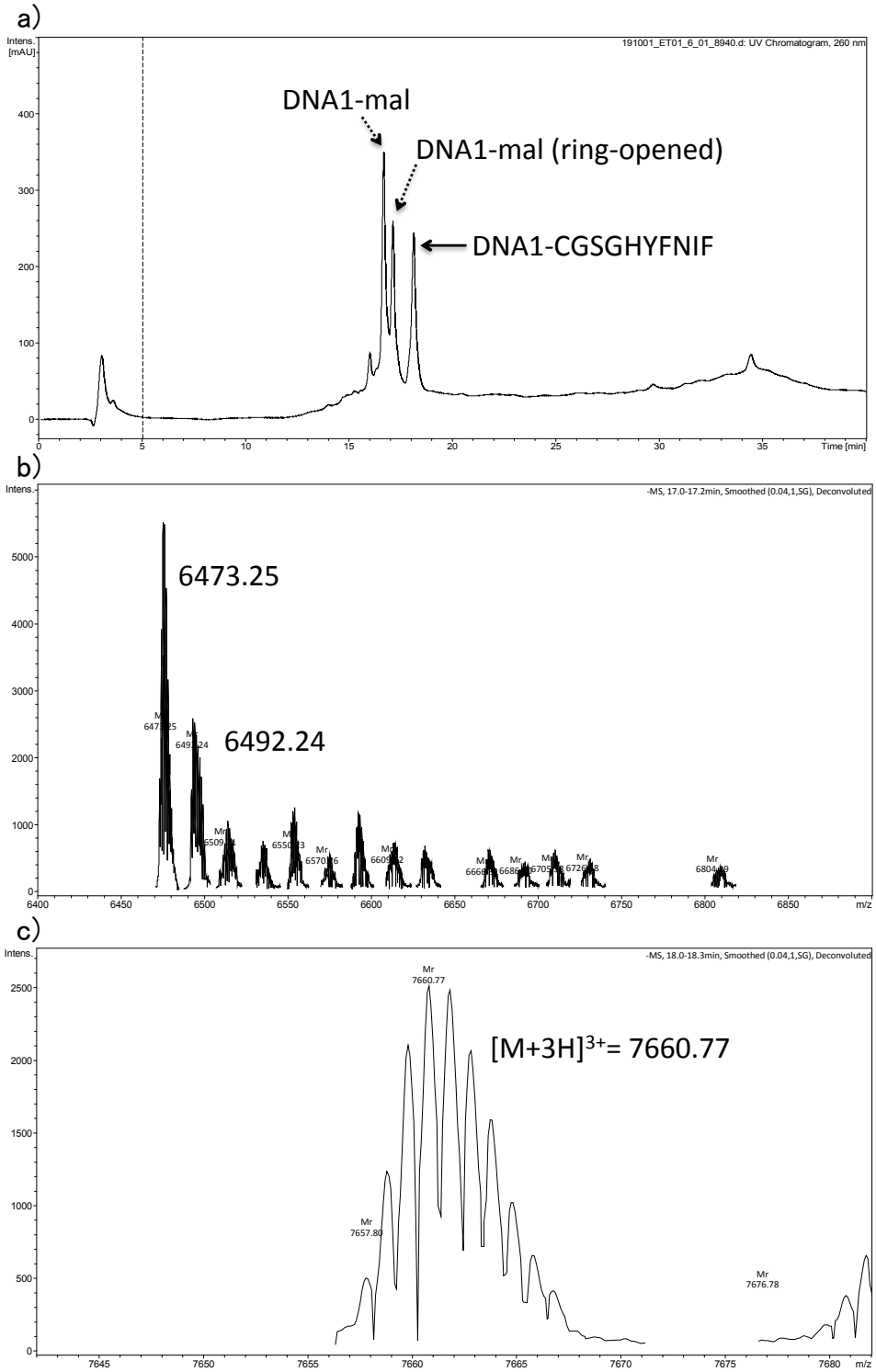


Figure 3.8. a) Analytical HPLC chromatograph of **DNA1-CGSGHYFNIF**, *TCR-30* (monitoring DNA absorbance at 260 nm). Unreacted **DNA1-mal** eluted at 17 min, and **DNA1-CGSGHYFNIF** eluted at 18 min. **b)** Expanded trace of DNA-mal (6473.25) and ring-opened form of DNA-mal (6492.24). **c)** Expanded trace of the **DNA1-CGSGHYFNIF** peak. LC-MS (m/z) calc. for: $[M+3H]^{3+}$ 7660.68, found 7660.77.

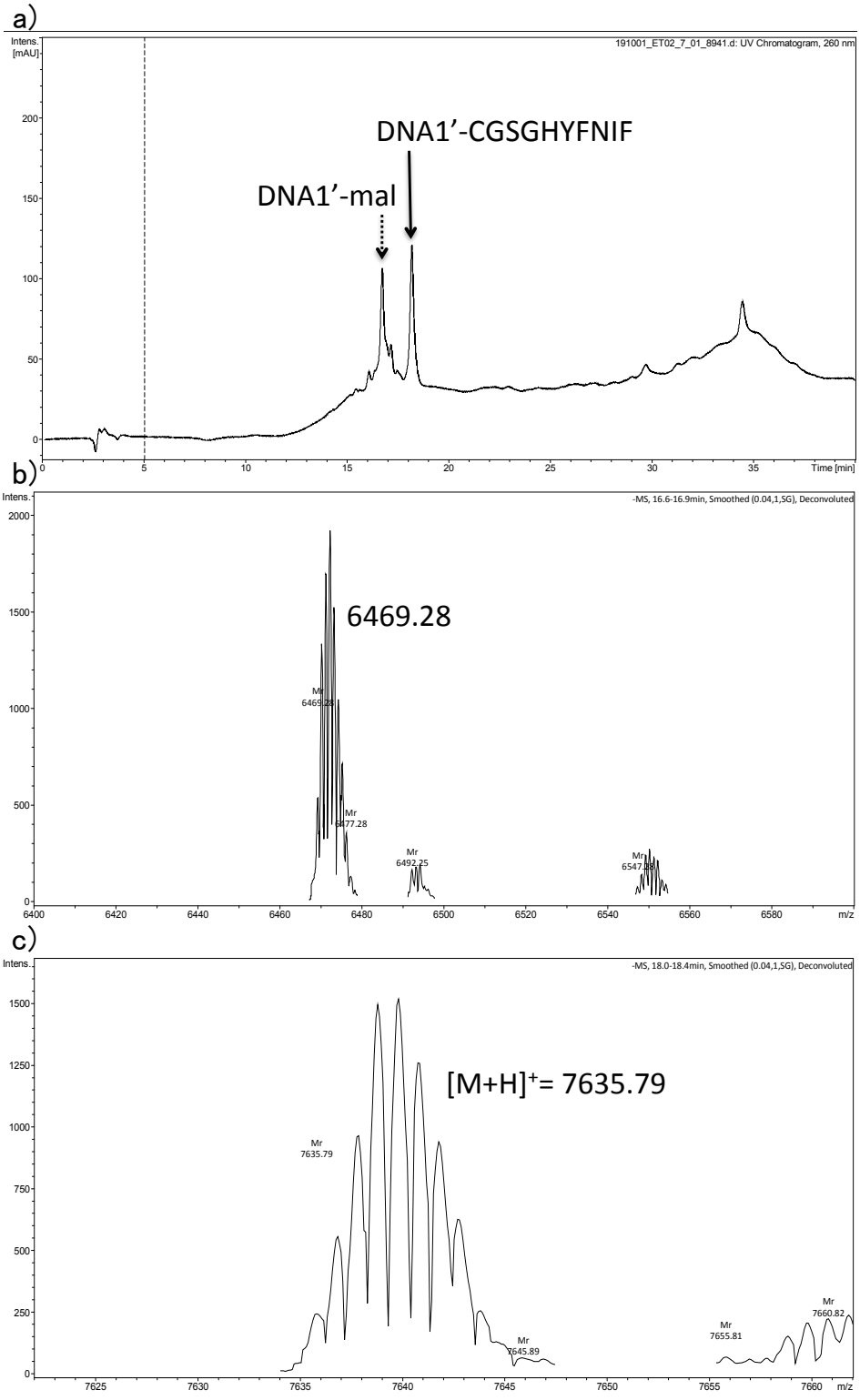


Figure 3.9. a) Analytical HPLC chromatograph of **DNA1'-CGSGHYFNIF**, *TCR-30* (monitoring DNA absorbance at 260 nm). Unreacted **DNA1'-mal** eluted at 16.6 min, and **DNA1'-CGSGHYFNIF** eluted at 18 min. **b)** Expanded trace of ring-opened form of **DNA-mal** (6469.28). **c)** Expanded trace of the **DNA1'-CGSGHYFNIF** peak. LC-MS (m/z) calc. for: $[M+H]^+$ 7636.69, found 7635.79.

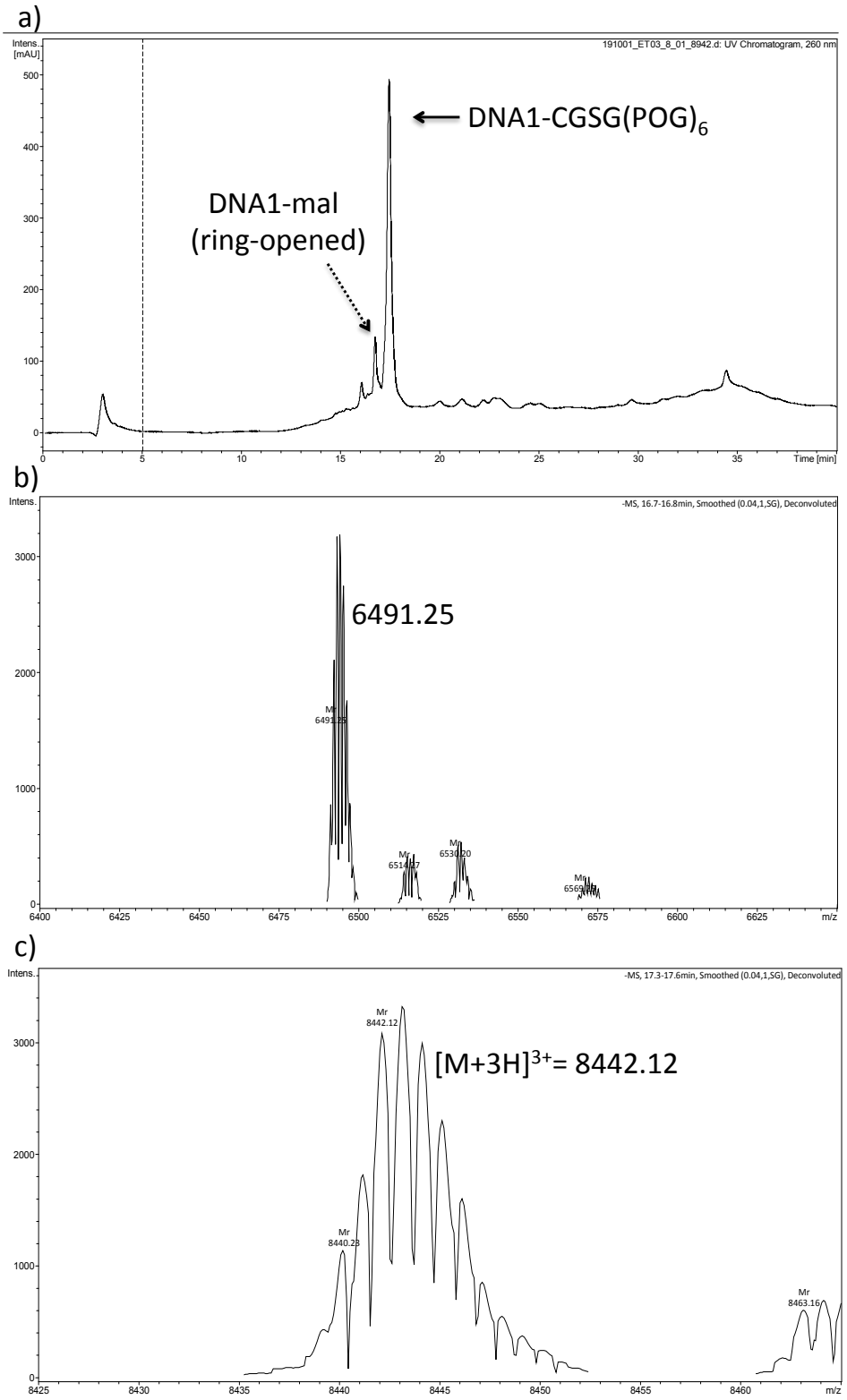


Figure 3.10. a) Analytical HPLC chromatograph of **DNA1-CGSG(POG)$_6$**, *TCR-30* (monitoring DNA absorbance at 260 nm). Ring-opened form of DNA1-mal eluted at 16.7 min, and **DNA1-CGSG(POG)$_6$** eluted at 17.3 min. **b)** Expanded trace of ring-opened form of **DNA-mal** (6491.25). **c)** Expanded trace of the **DNA1-CGSG(POG)$_6$** peak. LC-MS (m/z) calc. for: $[M+3H]^{3+}$ 8442.02, found 8442.12.

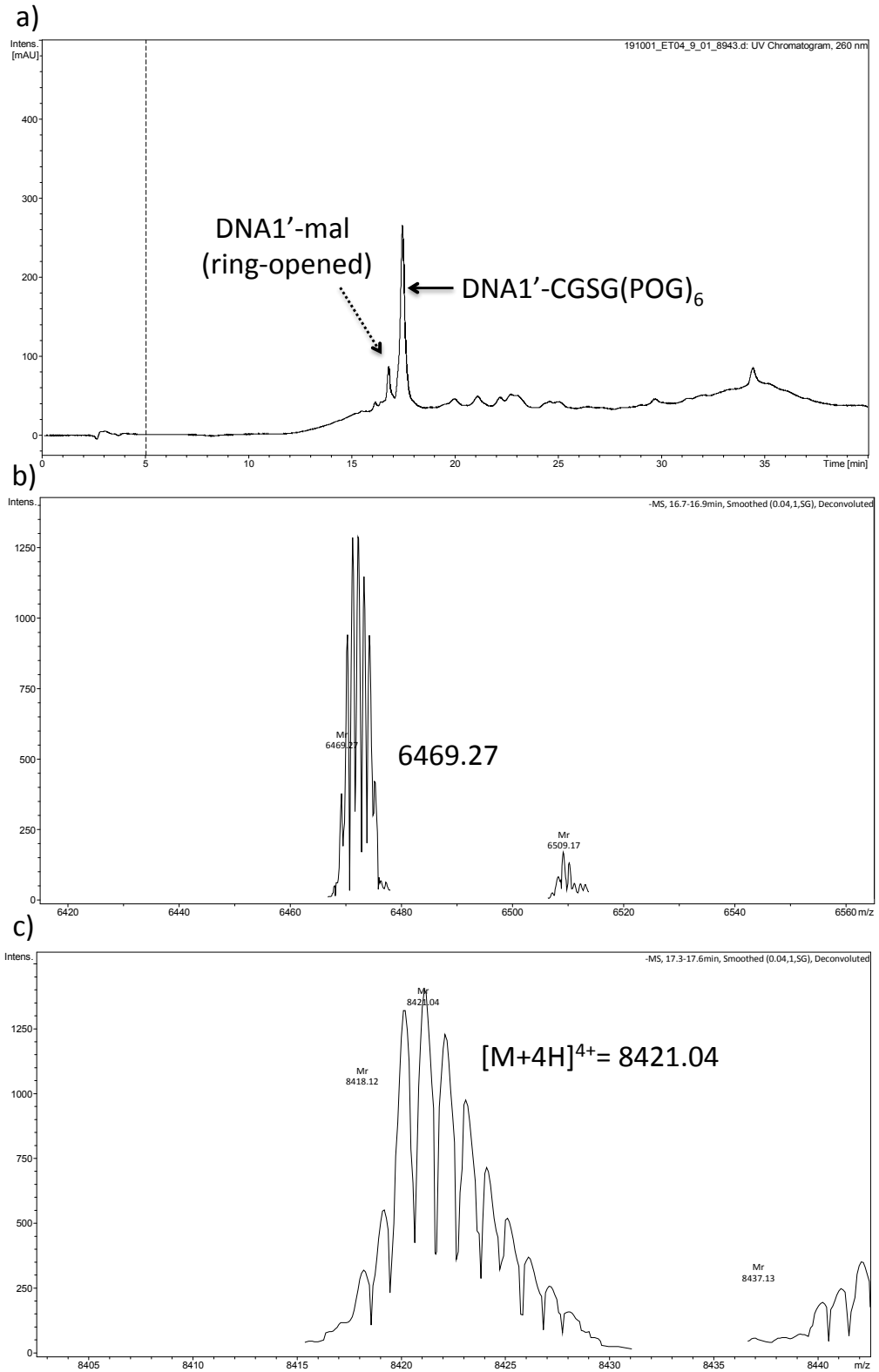
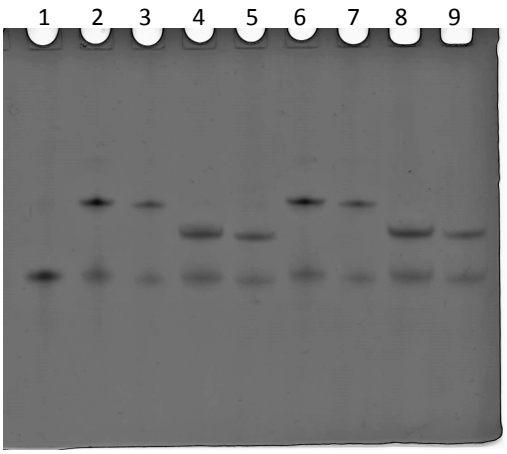


Figure 3.11. a) Analytical HPLC chromatograph of **DNA1'-CGSG(POG)$_6$**, *TCR-30* (monitoring DNA absorbance at 260 nm). Ring-opened form of **DNA1-mal** eluted at 16.7 min, and **DNA1'-CGSG(POG)$_6$** eluted at 17.3 min. **b)** Expanded trace of ring-opened form of **DNA1'-mal** (6469.27). **c)** Expanded trace of the **DNA1'-CGSG(POG)$_6$** peak. LC-MS (m/z) calc. for: $[M+4H]^{4+}$ 8421.04, found 8421.04.

3.3.8 Modulating Reaction Temperatures (*TCR-32*)

TCR-32 was similar to *TCR-30* with the exception that the reactions were carried out at room temperature to minimise the risk of DNA maleimide group hydrolysis. 20 µL from each reaction were taken out and quenched using 2 µL of 100 mM idoacetamide after 2 hours to check if hydrolysis had occurred by this point. The PAGE study suggested that the conjugate conversion yield was improved in *TCR-32* given that the intensity of the conjugate bands was stronger compared with the bands of unreacted DNA (Figure 3.12). This result highlighted that lowering the reaction temperature potentially reduced the hydrolysis of maleimide groups, resulting in a higher DNA to conjugate conversion yield.

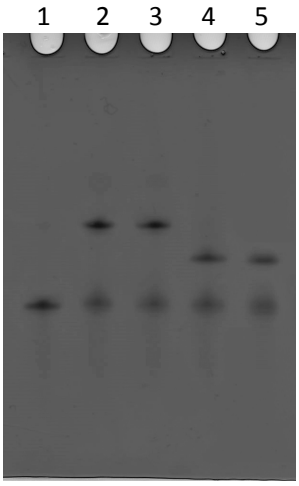


Well No.	Sample	Band intensity ratio (conjugate:unreacted DNA)
1	10 µM DNA1'-mal control	N/A
2	DNA1 + POG (2h)	1.9:1.0 (66:34)
3	DNA1' + POG (2h)	1.6:1.0 (62:38)
4	DNA1 + HYF (2h)	1.0:1.0 (51:49)
5	DNA1' + HYF (2h)	1.4:1.0 (59:41)
6	DNA1 + POG (O/N)	1.7:1.0 (63:37)
7	DNA1' + POG (O/N)	1.6:1.0 (61:39)
8	DNA1 + HYF (O/N)	1.6:1.0 (61:39)
9	DNA1' + HYF (O/N)	1.3:1.0 (56:44)

Figure 3.12. 20% denaturing PAGE of *TCR-32*. POG stands for **CGSG(POG)$_6$**, and HYF for **CGSGHYFNIF**.

3.3.9 Reactions to Reduce the Risk of Maleimide Group Hydrolysis (*TCR-33*)

To further minimise maleimide group hydrolysis, the reactions were carried out at a lower temperature (4°C) in the fridge. The PAGE study of *TCR-33* showed stronger conjugate bands than unreacted DNA bands (Figure 3.13). However, the unreacted DNA bands observed showed a greater intensity compared with those observed for *TCR-32* (Figure 3.12), potentially suggesting a smaller conversion yield of DNA to the conjugate in *TCR-33*. It is possible that lowering the temperature to 4°C slowed down the reaction of peptide with DNA, leading to a smaller formation of conjugate.



Well No.	Sample	Band intensity ratio (conjugate : unreacted DNA)
1	10 µM DNA1'-mal control	N/A
2	DNA1 + POG (4°C, 2h)	1.4:1.0 (58:42)
3	DNA1 + POG (4°C, O/N)	1.6:1.0 (61:39)
4	DNA1 + HYF (4°C, 2h)	1.0:1.0 (51:49)
5	DNA1 + HYF (4°C, O/N)	1.0:1.0 (51:49)

Figure 3.13. 20% denaturing PAGE of *TCR-33*. POG stands for **CGSG(POG)$_6$**, and HYF for **CGSGHYFNIF**.

3.3.10 Purification by Anion Exchange Chromatography

Although altering reaction conditions improved the DNA to conjugate conversion yield, full conversion was not achieved. Subsequently, in an attempt to obtain the pure conjugates, different purification strategies were tested, with anion exchange purification of *TCR-30* (**DNA1-CGSGHYFNIF**) being tried in first place.

UV-Vis measurements revealed that unreacted DNA and the conjugate had adhered to the resin given that the supernatant fraction (F1) showed a reduced absorption at 260 nm compared with the reaction mixture fraction (Figure 3.14 a). Although DNA absorption at 260 nm was observed after washing in 1M NaCl for a second time, the PAGE study revealed that the separation of unreacted DNA and the conjugates was unsuccessful, given that DNA bands in two fractions were observed whereas the remaining fractions showed no conjugate or DNA bands (Figure 3.14 b).

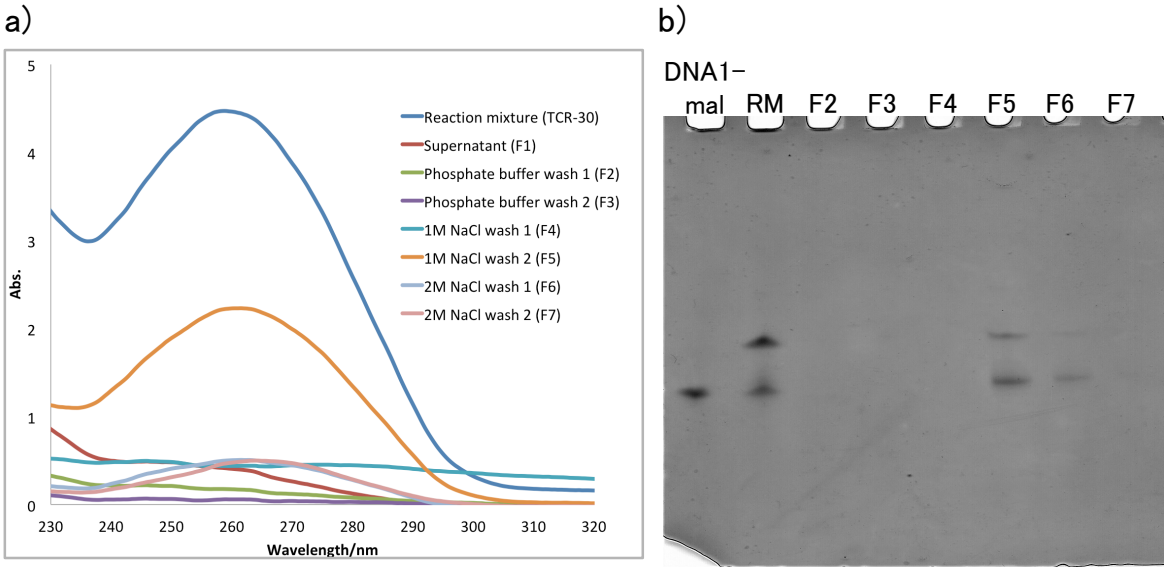


Figure 3.14. a) UV-Vis spectra of fractions 1-7 following the first purification attempt by anion exchange of *TCR-30* **b)** 20% denaturing PAGE of fractions 2-7, 10 µM of **DNA1-mal** control, and *TCR-30* prior to purification (labeled as RM).

As the separation of **DNA1-CGSGHYFNIF** (*TCR-30*) from the reaction mixture was unsuccessful, the second purification attempt by anion exchange was conducted. This time the washing of the resin was carried out using a slower gradient of salt concentration (0.1 -1.0 M) to allow for better separation of the conjugate from unreacted DNA than in the previous purification attempt. UV-Vis measurements of fractions 7 to 9 showed the presence of DNA absorption peaks (Figure 3.15). The PAGE study showed both unreacted DNA and faint conjugate bands in fractions 7, 8 and 9. Although the separation appeared to have been improved compared with the previous purification attempt, the pure conjugate was not obtained. The incomplete separation between unreacted DNA and the conjugate may be due to the possibility that both these biomolecules had similar binding affinities to the resin. Given that only limited improvements were obtained in this second attempt, an alternative purification strategy was explored instead.

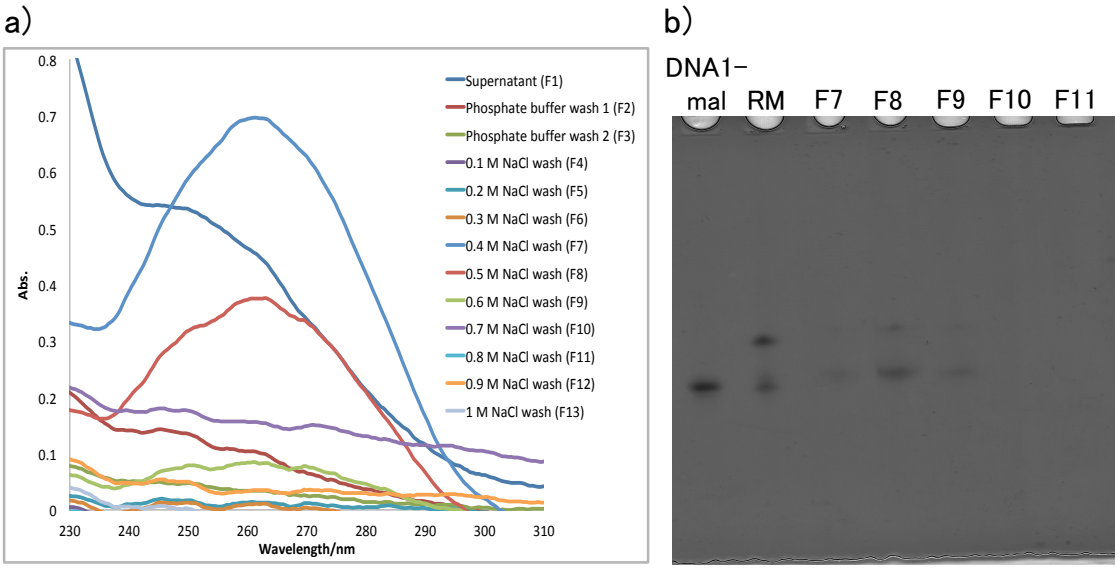


Figure 3.15. a) UV-Vis spectra of fractions 1-13 following the second attempt of anion exchange purification of *TCR-30* **b)** 20% denaturing PAGE of fractions 7-11, 10 µM of **DNA1-mal** control, and *TCR-30* before purification (labeled as RM).

3.3.11 Purification by Denaturing Urea PAGE

In an attempt to effectively remove unreacted DNA from the conjugation reaction mixture, denaturing urea PAGE purification was tested using a new batch of reaction mixture (*TCR-34*). *TCR-34* was prepared for PAGE purification by reacting **DNA1-mal** (224 µM) with **CGSGHYFNIF** (4.5 mM) in HFIP/phosphate buffer (1:3, v/v) at R.T. overnight (total reaction volume: 850 µL), followed by 20% denaturing PAGE to confirm the presence of the desired conjugate. HFIP was used to help to dissolve the peptide.

Figure 3.16. 20% denaturing PAGE of *TCR-34* and 10 µM **DNA1-mal** control.

Prior to proceeding with PAGE purification, a preliminary PAGE study of *TCR-34* showed the presence of unreacted DNA and a conjugate band (Figure 3.16). The purification PAGE unexpectedly showed three bands (Figure 3.17), of which the top one was assumed to contain the conjugate. Whereas, the bottom band on the gel may have resulted due to tailing from the middle band, given that a large amount of reaction mixture was loaded to the purification gel (520 µL of 112 µM *TCR-34*).

To assess whether the desired conjugate was successfully extracted, the top band was analysed by UV-Vis spectroscopy and 20% denaturing PAGE following PAGE purification. UV-Vis measurements showed a broad absorption between 260-280 nm in fractions 4-8, indicating the presence of conjugate sample extracted from the purification gel. In addition, fractions 6-8 showed a large absorbance between 220-240 nm, which suggested presence of urea.[9]

The subsequent 20% denaturing PAGE of fractions 3-6 showed unreacted DNA and conjugate bands (Figure 3.18). It is possible that the unsuccessful separation between unreacted DNA and the desired conjugate may have been due to overloading the sample to the gel, which led to incomplete separation between them.

Figure 3.17. A gel illuminated by UV light following the first PAGE purification attempt of *TCR-34*.

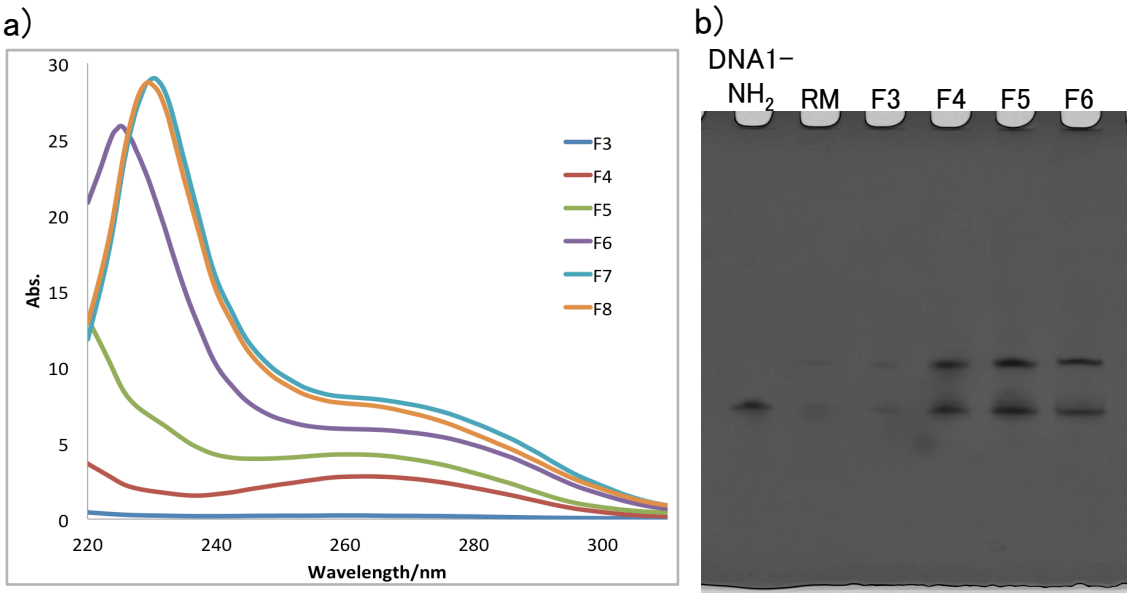


Figure 3.18. a) UV-Vis spectra of fractions 3-8 following PAGE purification and desalting **b)** 20% denaturing PAGE of 20 µM **DNA1-NH_2** control, *TCR-34* prior to PAGE purification and fractions 3-6.

To further improve the separation of the conjugate from unreacted DNA, a second attempt at PAGE purification was carried out using a smaller amount of reaction sample. Fractions 4-8 obtained following the first attempt of PAGE purification and desalting were concentrated to 300 µL of 80 µM solution. Using this concentrated sample, PAGE purification was performed again.

Figure 3.19. A gel illuminated by UV light following the second PAGE purification attempt of *TCR-34*.

The gel obtained following the second PAGE purification showed two bands (Figure 3.19), being consistent with the preliminary PAGE study of *TCR-34* (Figure 3.16). To assess whether the desired conjugate was successfully extracted, the top band was analysed by UV-Vis spectroscopy and 20% denaturing PAGE following PAGE purification.

UV-Vis measurements showed a large absorption between 230-240 nm and a broad absorption between 260-290 nm in fractions 7-10 corresponding to the presence of urea, which overlapped with the absorption of the conjugate sample (Figure 3.20 a). The subsequent 20% denaturing PAGE of fractions 4-7 showed the presence of both **DNA1-mal** and conjugate bands (Figure 3.20 b), meaning that the isolation of the conjugate from unreacted DNA was unsuccessful.

The cause of extraction failure may have been attributed to the incomplete separation between **DNA1-mal** and the conjugate on the purification gel. However, it may have been possible that **DNA1-mal** was regenerated from the conjugate following PAGE purification, given that thiol-maleimide reactions have been reported to be reversible under certain conditions. The band extraction steps that involved the incubation of the conjugate sample at 60°C in basic conditions for several hours could have potentially resulted in a reversible thiol-maleimide reaction[10].

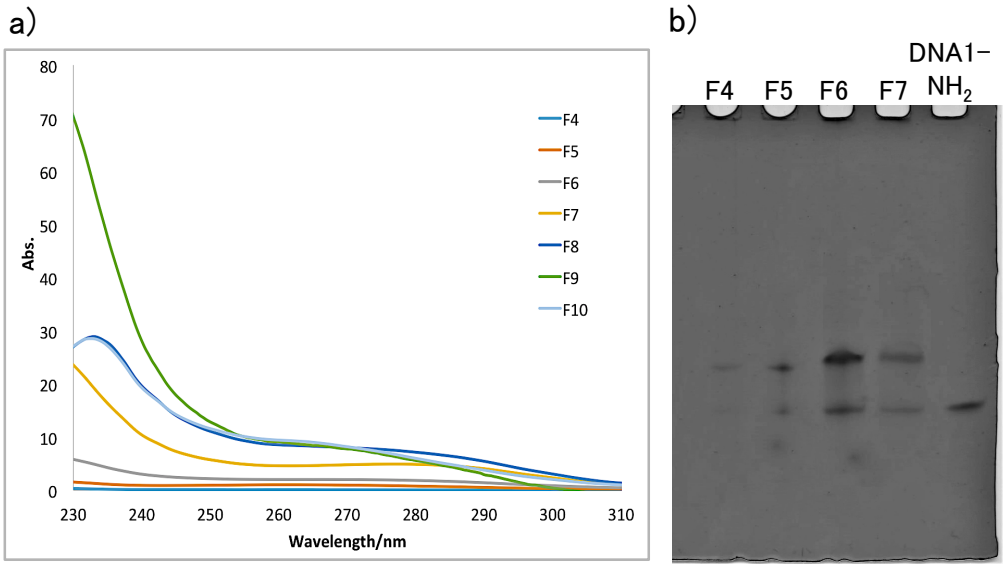
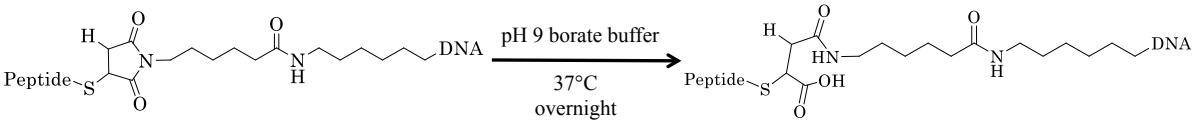


Figure 3.20. a) UV-Vis spectra of fractions 4-10 following PAGE purification and desalting **b)** 20% denaturing PAGE of 20 µM **DNA1-NH_2** control and fractions 4-7.

3.3.12 Succinimide-thioether Ring-opening Reaction of the Conjugate (*TCR-36*)

To reduce the risk of thiol-maleimide reactions reversing, hydrolysis of the succinimide-thioether ring was attempted to increase the stability of the thiol-maleimide linkage (Scheme 3.2).[10] To do this, 10 µL of *TCR-34* reaction mixture was mixed with 90 µL of 0.1 M borate buffer at pH 9 in a PCR tube, which was then heated at 37°C in a thermocycler overnight.



Scheme 3.2. Hydrolysis reaction of the succinimide-thioether ring.

To assess whether the hydrolysis reaction had occurred, *TCR-36* was analysed by LC-MS. The HPLC trace at 260 nm showed a strong peak of the original conjugate (Figure 3.21 a), indicating that the succinimide-thioether ring did not undergo hydrolysis. Conversely, the HPLC trace suggested that the reaction conditions employed led to an improved conversion of DNA to the desired conjugate. The denaturing 20% PAGE study of *TCR-36* only showed a faint band of unreacted DNA, which corresponded to a small peak at 15 min on the HPLC chromatogram (Figure 3.21 c). While no conjugate bands were seen on the gel, the confirmation of the conjugate peak previously obtained by HPLC suggested that the conjugate precipitated, implying that the conjugate was not loaded to the gel.

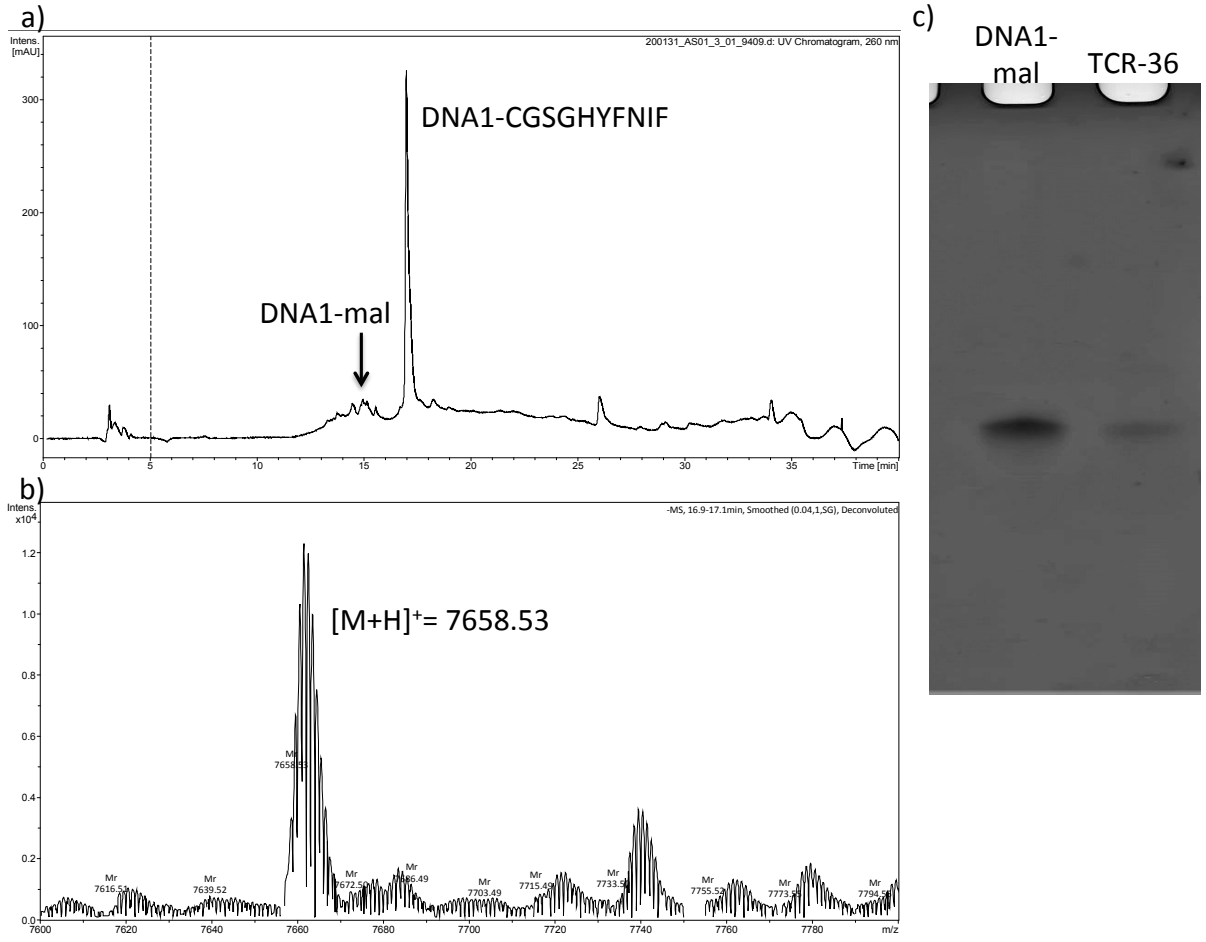


Figure 3.21. a) Analytical HPLC chromatograph of *TCR-36* (monitoring DNA absorbance at 260 nm). Unreacted **DNA1-mal** eluted at 15 min, and **DNA1-CGSGHYFNIF** eluted at 17 min. **b)** Expanded trace of the **DNA1-CGSGHYFNIF** peak. LC-MS (m/z) calc. for: $[M+H]^+$ 7658.68, found 7658.53. **c)** Denaturing 20% PAGE showing **DNA1-mal** (control) and *TCR-36*.

3.3.13 Conjugate Extraction by Borate Buffer Washing of *TCR-42*

For *TCR-42*, the conditions used in *TCR-36* were repeated using the conjugate solution obtained in *TCR-34*, for the purpose of obtaining more conjugate sample. The supernatant and the precipitate that had excess peptide were separated by centrifugation, prior to analysing both samples individually using 20% denaturing PAGE (Figure 3.22 b).

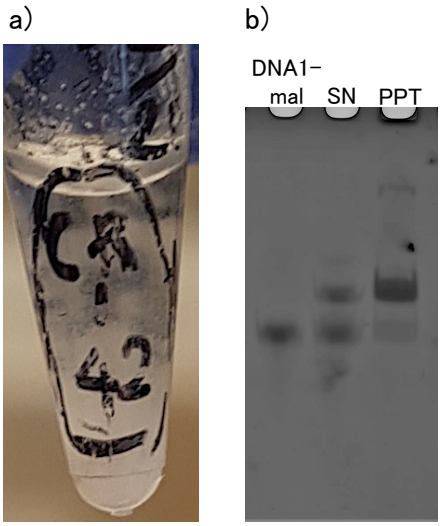


Figure 3.22. a) *TCR-42* reaction mixture. **b)** 20% denaturing PAGE showing **DNA1-mal** (control), *TCR-42* supernatant (SN) and *TCR-42* precipitate (PPT). Sample preparation for PAGE: (SN) 5 µL of supernatant was mixed with 7.5 µL of MilliQ water and 7.5 µL of 8M urea in a PCR tube. (PPT) 2.5 µL of solution containing the precipitate was taken to a PCR tube, which was then mixed with 15 µL of HFIP and 15 µL of 8M urea. 10 µL of each solution was loaded to the gel.

The PAGE analysis of the supernatant showed both unreacted DNA and conjugate bands of similar intensity. In contrast, the PAGE analysis of the precipitate revealed a substantially fainter unreacted DNA band, as well as a more prominent conjugate band compared with the supernatant. This PAGE study appeared to demonstrate that most of the DNA converted to the conjugate in the precipitate sample.

To isolate the conjugate from unreacted DNA, the precipitate was washed with additional pH 9 borate buffer (50 µL ×2). After each round of additional washing, the sample was centrifuged to separate the supernatant from the precipitate. To assess whether the additional removal of unreacted DNA was successful, the supernatants obtained were subsequently analysed by 10% PAGE (Figure 3.23).

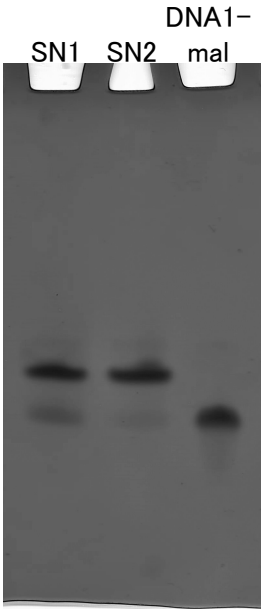


Figure 3.23. 10% denaturing PAGE showing the supernatant obtained following one round of washing (SN1) and two rounds of washing (SN2) of the precipitate with borate buffer, and **DNA1-mal** control.

The PAGE analysis revealed that the unreacted DNA band became fainter following two rounds of washing the precipitate. However, the precipitate was not washed for more than two rounds to minimise the loss of conjugate that was observed in both supernatant samples (SN1 and SN2). Instead, anion exchange purification chromatography was carried out to extract the desired conjugate from the precipitate, without the risk of losing the conjugate.

3.3.14 Anion exchange purification of *TCR-42*

UV-Vis measurements showed that DNA absorption at 260 nm was observed for fractions 8-13, suggesting that unreacted DNA and conjugate were gradually released from the resin (Figure 3.24 a). The two fractions (F9 and F10) that showed the most prominent UV spectra were analysed by 10% denaturing PAGE to assess whether the desired conjugate was separated from unreacted DNA. The UV spectra of F11-F15 suggested that these fractions also contained the desired conjugate. However, these fractions were not analysed by PAGE because the sample concentration was not high enough to visualise bands on the gel.

The PAGE analysis of fractions 9 and 10 showed prominent conjugate bands, whereas very faint unreacted DNA bands were observed, suggesting that the reaction mixture was almost composed of the desired conjugate (Figure 3.24 b). It is possible that washing the precipitate with borate buffer prior to conducting anion exchange purification played a critical role in removing unreacted DNA from the sample, given that this washing step was not conducted prior to anion exchange purification *TCR-30*, where removal of unreacted DNA was unsuccessful.

Despite the acceptable outcomes obtained with anion exchange at removing unreacted DNA, this strategy was discontinued due to the significant amount of product loss caused by the multiple washing steps of this purification strategy.

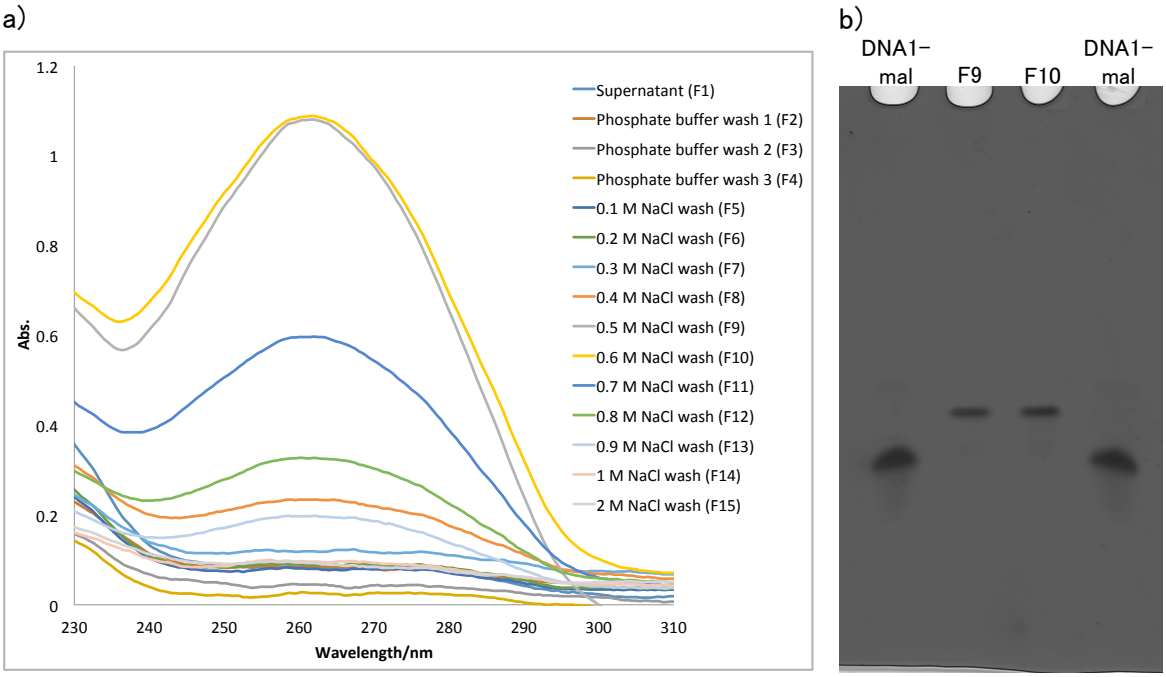


Figure 3.24. a) UV-Vis spectra of fractions 1-15 following anion exchange purification of *TCR-42* SN2 **b)** 10% denaturing PAGE of 10 µM **DNA1-mal** as a control, fraction 9 and fraction 10.

3.3.15 Purification by RP-HPLC (*TCR-37*)

RP-HPLC of *TCR-37* (**DNA1'-mal + CGSGHYFNIF**) and *TCR-39* (**DNA1'-mal + CGSGKLFFA**) was conducted as an alternative purification strategy, with the aim of minimising the loss of the conjugate samples. The PR-HPLC elution fractions obtained between 11-12 min were labeled as F13-F16, respectively (Figure 3.25).

The LC-MS of *TCR-37* prior to RP-HPLC purification showed multiple peaks, including those corresponding to the desired conjugate (**DNA1'-CGSGKLVFFA**), as well as showing impurities in the samples (Figure 3.25). The PAGE analysis of this sample showed unreacted DNA and conjugate bands (Figure 3.26).

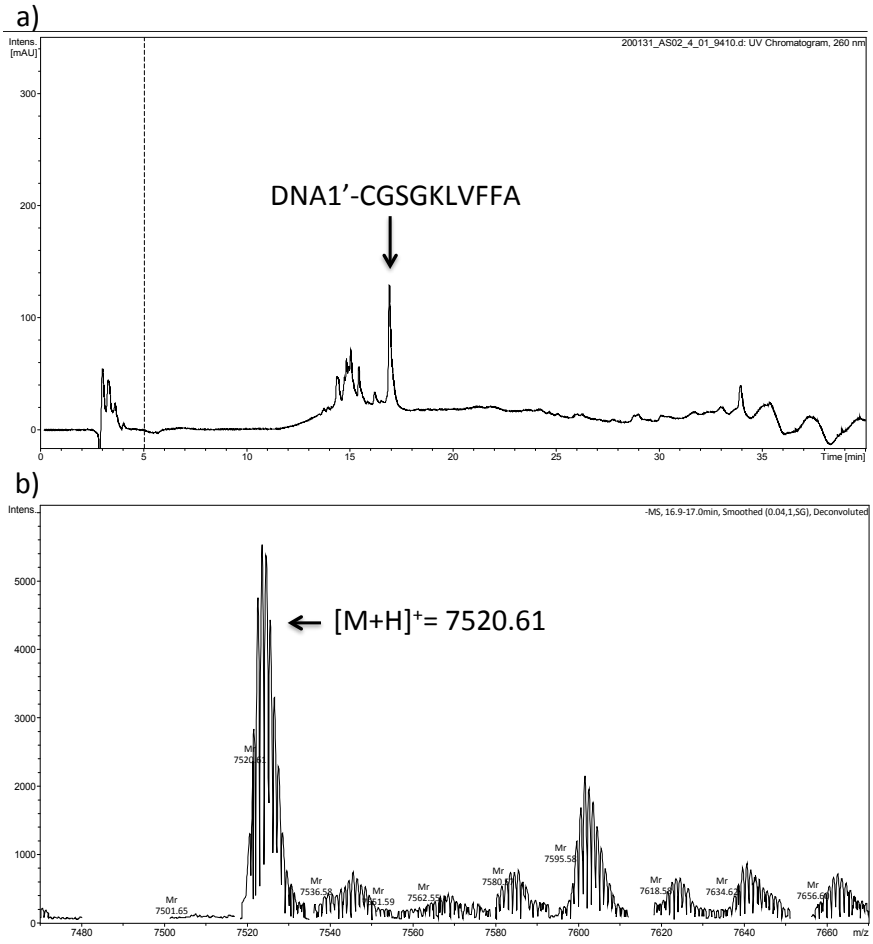


Figure 3.25. a) LC-MS of *TCR-37* prior to PR-HPLC purification showing the presence of **DNA1'-CGSGKLVFFA**. **a)** Analytical HPLC chromatograph of *TCR-37* (monitoring DNA absorbance at 260 nm) showing that **DNA1'-CGSGKLVFFA** eluted at 17 min. **b)** Expanded trace of the **DNA1'-CGSGKLVFFA** peak. LC-MS (m/z) calc. for: $[M+H]^+$ 7520.46, found 7520.61.

The UV-Vis measurements of F13, F14 and F15 showed characteristic DNA absorption peaks at 260 nm, suggesting conjugate formation (Figure 3.26 b). Meanwhile, the DNA absorption peak of F16 was weak, possibly due to its low concentration. F15 showed the highest DNA absorption and was subsequently analysed by 20% PAGE, but no bands were observed on the gel (Figure 3.26 c). This may have been due to the degradation of DNA into smaller oligonucleotides during RP-HPLC purification, which would not have been detectable in the gel.

Another possibility was that additional impurities, which might have formed following RP-HPLC potentially interfered with the gel stain, preventing bands from appearing. This

hypothesis was based on the fact that while the post-purification sample showed no bands, a conjugate band was observed prior to conducting RP-HPLC purification (Figure 3.26 c).

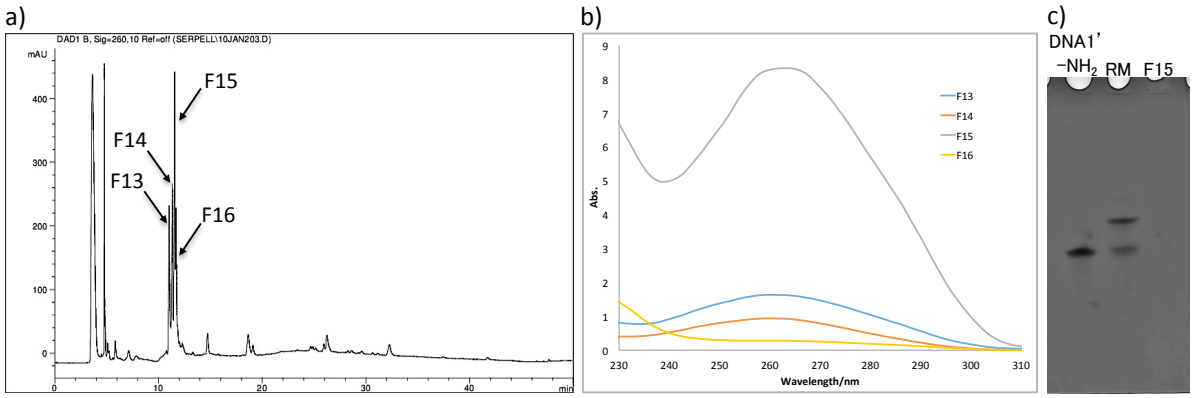


Figure 3.26. a) RP-HPLC chromatograph of *TCR-37* (monitoring DNA absorbance at 260 nm). **b)** UV-Vis of F13-F16 obtained following RP-HPLC purification of *TCR-37*. **c)** 20% denaturing PAGE of 20 µM **DNA1'-NH$_2$**, *TCR-37* prior to RP-HPLC purification and F15.

3.3.16 Purification by RP-HPLC (*TCR-39*)

The PR-HPLC elution fractions obtained between 12-16 min were labeled as F13, F18, F20 and F21 (as indicated in Figure 3.27). UV-Vis of F13, F18, F20 and F21 showed a characteristic DNA absorption peak at 260 nm, suggesting presence of conjugates. However, the subsequent LC-MS analysis did not show the desired conjugate peak (**DNA1'-CGSGHYFNIF)** in any of the fractions despite the large oligonucleotide signal that was observed on the RP-HPLC trace of F18, F20 and F21 (Figure 3.30-32). Moreover, F18 and F20 did not show peaks above 2000 m/z, indicating absence of oligonucleotides in the fractions, which could have been indicative of conjugate degradation. Another possibility is that sample aggregation may have occurred, failing to be analysed by LC-MS and thereby showing no oligonucleotide peaks.

Although F21 showed multiple peaks ranging between 4000-7000 m/z, none of them corresponded to the mass of unreacted **DNA1'-mal** (MW: 6450.18) or **DNA1'-CGSGHYFNIF** (MW: 7520.46), suggesting that the conjugate was not present (Figure 3.32). In addition, multiple peaks observed in the mass range of 6000 m/z might have suggested that oligonucleotide-derived impurities were formed.

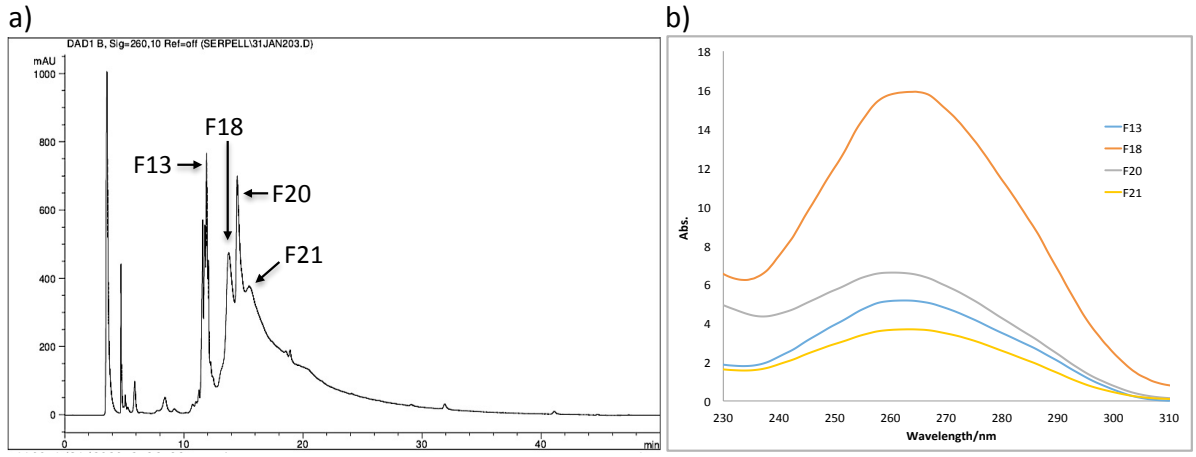


Figure 3.27. a) RP-HPLC chromatograph of *TCR-39* (monitoring DNA absorbance at 260 nm). **b)** UV-Vis of F13, F18, F20 and F21 obtained following RP-HPLC purification of *TCR-39* (**DNA1'-CGSGHYFNIF**).

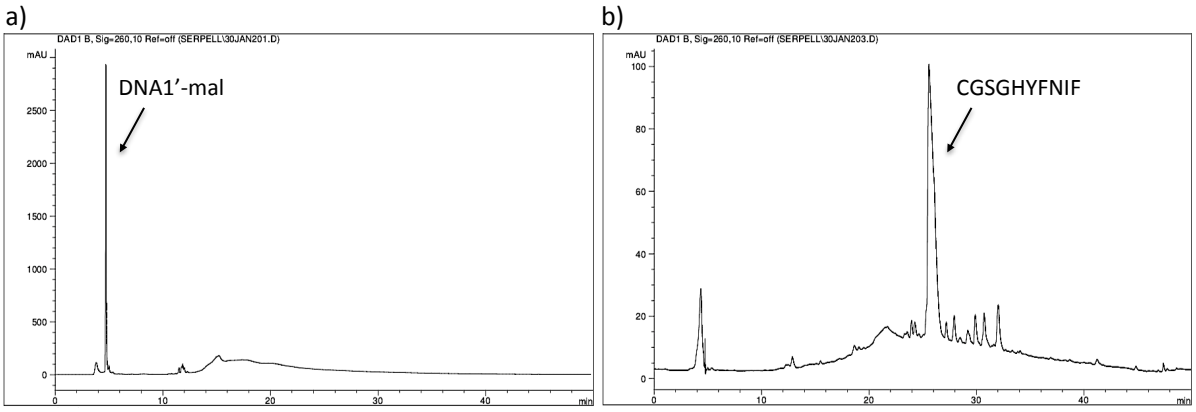


Figure 3.28. RP-HPLC chromatograph of **a) DNA1'-mal** and **b) CGSGHYFNIF** controls (monitoring DNA absorbance at 260 nm). **DNA1-mal'** eluted at 4.5 min and **CGSGHYFNIF** eluted at 26 min.

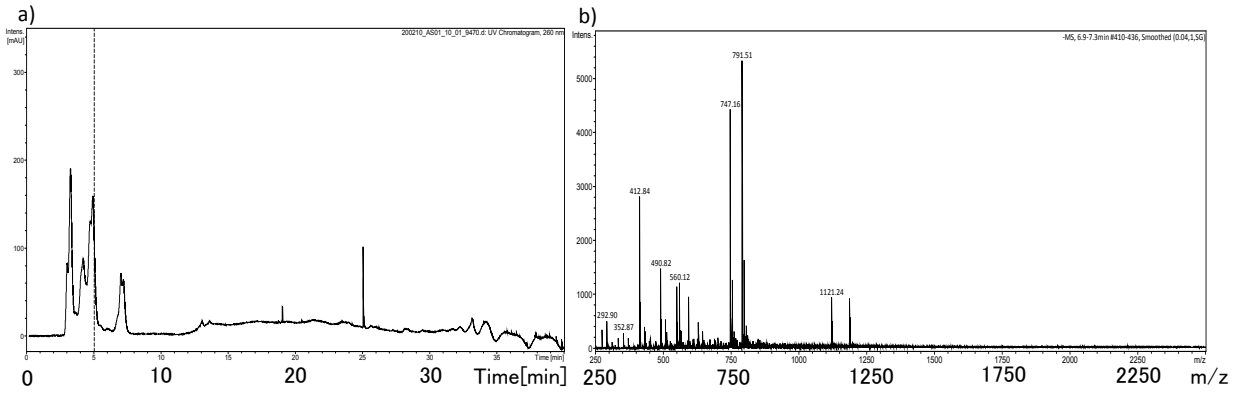


Figure 3.29. a) Analytical RP-HPLC chromatograph (monitoring DNA absorbance at 260 nm) of F13 obtained following RP-HPLC purification of *TCR-39* (**DNA1'-CGSGHYFNIF**). **b)** Expanded trace of the peak between 6.9-7.3 min.

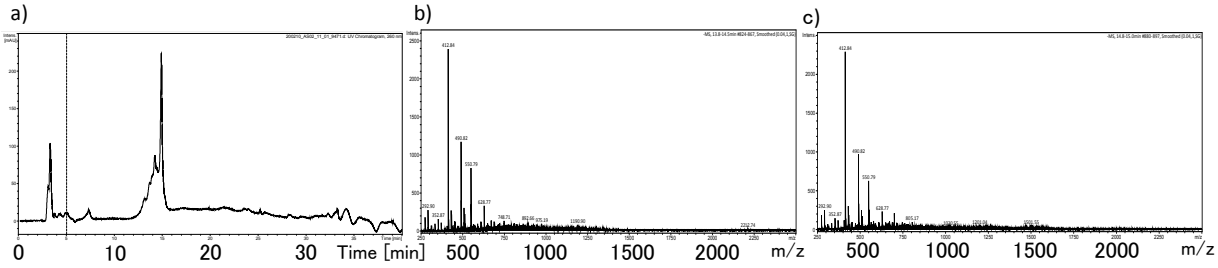


Figure 3.30. a) Analytical RP-HPLC chromatograph (monitoring DNA absorbance at 260 nm) of F18 obtained following RP-HPLC purification of *TCR-39* (**DNA1'-CGSGHYFNIF**). Expanded trace of the peak between **b)** 13.8-14.5 min and **c)** 14.8-15.0 min, respectively.

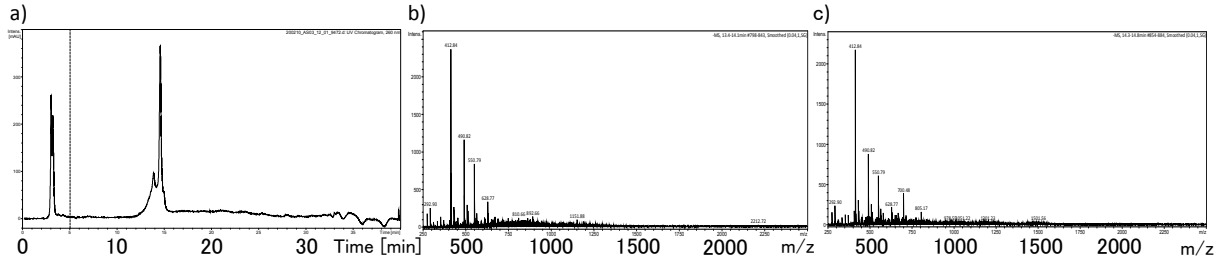


Figure 3.31. a) Analytical RP-HPLC chromatograph (monitoring DNA absorbance at 260 nm) of F20 obtained following RP-HPLC purification of *TCR-39* (**DNA1'-CGSGHYFNIF**). Expanded trace of the peak between **b)** 13.4-14.1 min and **c)** 14.3-14.8 min, respectively.

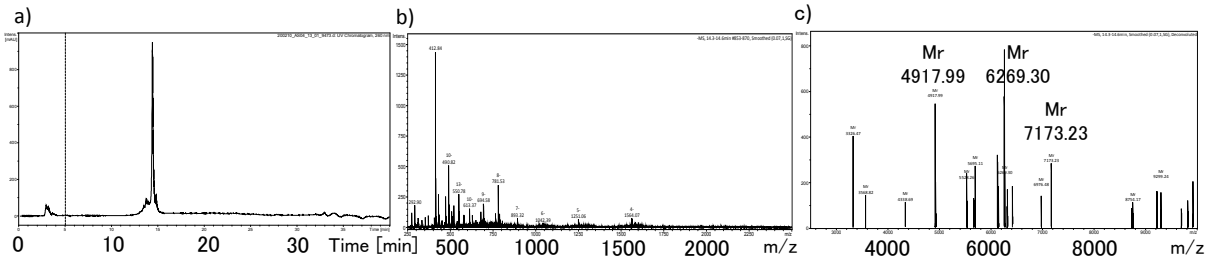


Figure 3.32. a) Analytical RP-HPLC chromatograph (monitoring DNA absorbance at 260 nm) of F21 obtained following RP-HPLC purification of *TCR-39* (**DNA1'-CGSGHYFNIF**). **b-c)** Expanded trace of the peak between 14.3-14.6 min.

3.3.17 Native PAGE study of *TCR-39* F12

To test the hypothesis that aggregate formation had occurred, 12% native PAGE was conducted using F12 (13 µM) obtained following RP-HPLC of *TCR-39* (**DNA1'-mal** + **CGSGHYFNIF**). In addition, F12 mixed with **DNA1** in a 1:1 ratio was also analysed by 12%

native PAGE to investigate if the sample could be visualised on the gel due to hybridisation.

F12 showed a characteristic DNA absorption peak with a similar spectral shape as **DNA1** (Figure 3.33), but it did not show any bands on 12% native PAGE, being similar to the result seen with the denaturing PAGE study of *TCR-37* F15. In addition, only a single band of **DNA1** control was confirmed after being mixed with F12, suggesting that hybridisation did not occur. Since aggregate formation was not confirmed on the gel, this further supported the possibility that the absence of bands could have been caused as a result of conjugate degradation.

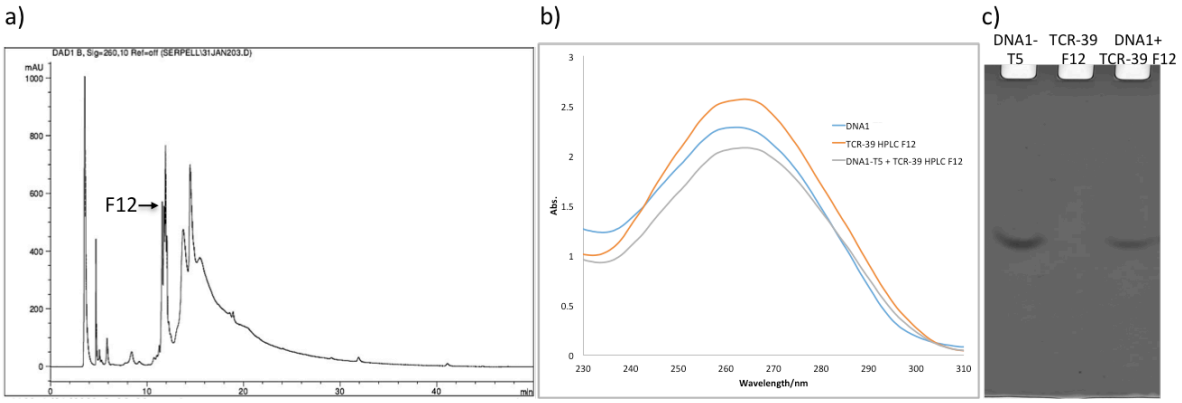


Figure 3.33. a) RP-HPLC chromatograph of *TCR-39* (monitoring DNA absorbance at 260 nm), highlighting F12. **b)** UV-Vis and **c)** 12% native PAGE of **DNA1**, F12 and solution made by mixing **DNA1** and F12.

The conjugate samples that were run using RP-HPLC were not visualised by either denaturing or native PAGE, despite that a characteristic DNA absorption peak at 260 nm was confirmed by UV-Vis spectroscopy. This might have implied that either the conditions used for RP-HPLC purification interfered with the DNA staining on the PAGE, or that they caused degradation of DNA into smaller oligonucleotide fragments that would not have been detectable by PAGE.

3.3.18 Stability of DNA under RP-HPLC Purification Conditions

To investigate this issue, **DNA1** and *TCR-46* reaction mixture containing **DNA1'-mal** and **DNA1'-CGSGHYFNIF** were treated with similar conditions used for RP-HPLC purification, but they were not run by RP-HPLC. The UV-Vis of **DNA1** showed characteristic DNA absorption

peaks, with slight variations in their absorbance, but with comparable spectra shapes (Figure 3.34 a). Despite applying the same dilution across the samples, the small differences in absorbance seen might have been caused by uneven mixing prior to conducting the UV-Vis measurements. Nevertheless, the subsequent 10% denaturing PAGE confirmed a single DNA band for every **DNA** analysed (Figure 3.34 b).

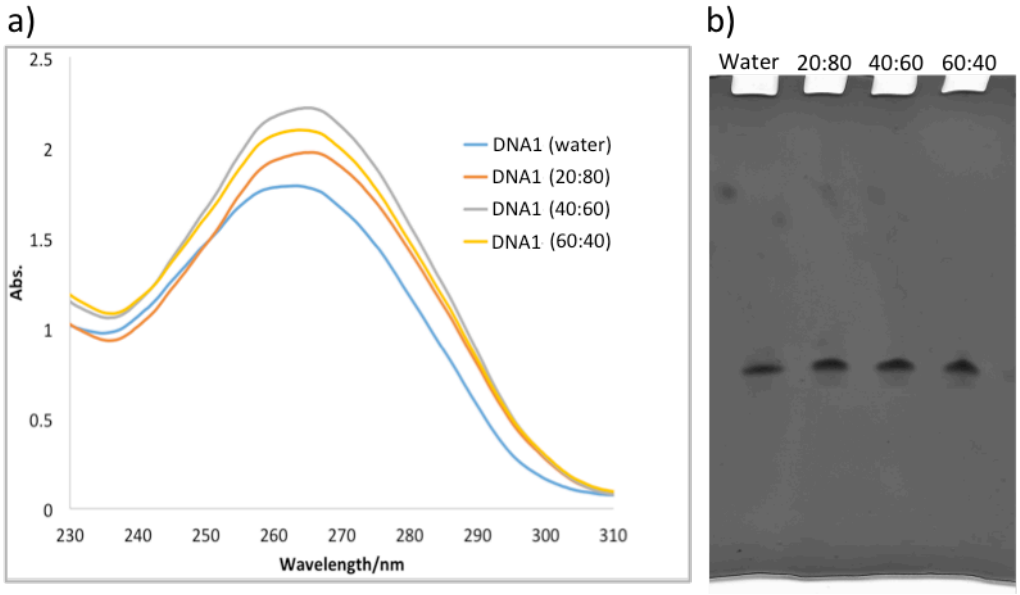


Figure 3.34. a) UV-Vis and **b)** 10% denaturing PAGE of **DNA1** incubated in water, solvent A: solvent B = 20:80, 40:60 and 60:40.

The *TCR-46* samples also showed characteristic DNA absorption peaks, and the appearance of both conjugate and unreacted DNA was confirmed by 10% denaturing PAGE (Figure 3.35 b). In contrast, unreacted DNA run by RP-HPLC did not show bands on the gel (Figure 3.35 b), despite the fact that a DNA absorption peak had been observed by UV-Vis (Figure 3.35 a).

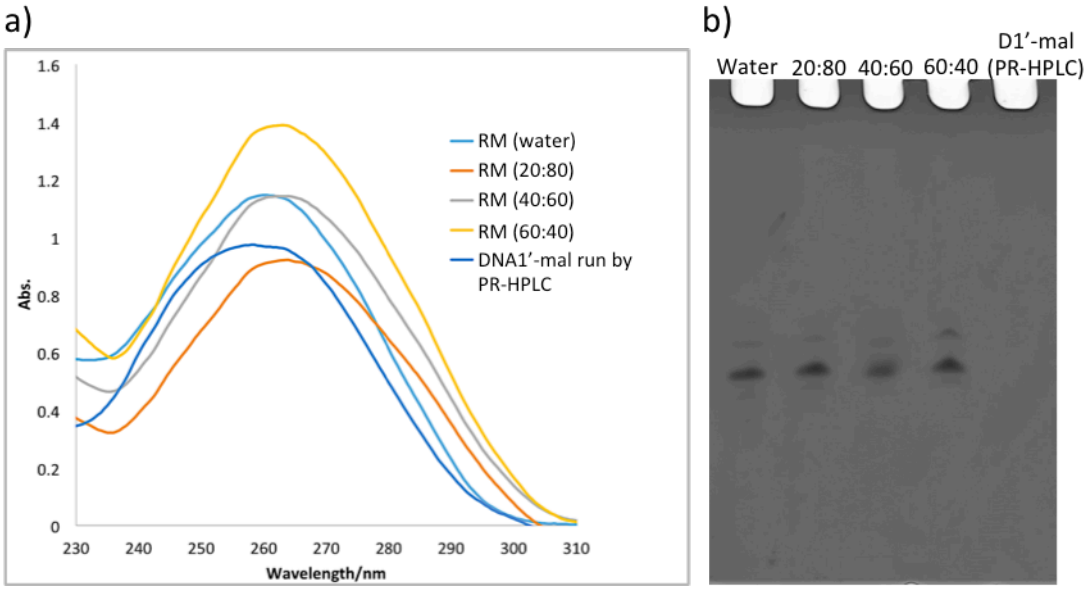


Figure 3.35. a) UV-Vis and **b)** 10% denaturing PAGE of *TCR-46* incubated in water, solvent A: solvent B = 20:80, 40:60 and 60:40, in addition to **DNA1'-mal** that was run by RP-HPLC.

These results suggested that mimicking RP-HPLC conditions in a thermocycler without using the RP-HPLC instrument did not affect DNA staining on the gel, or cause DNA degradation. Although these findings suggested that the disappearance of bands may have been attributed to running the samples on the RP-HPLC instrument, the exact cause as to why this happened could not be figured out. Given that it was not possible to confirm the presence of desired conjugates following RP-HPLC purification, hydrophobic interaction chromatography (HIC) was tested as an alternative purification method.

3.3.19 Purification by HIC Chromatography

Hydrophobic interaction chromatography (HIC) of **DNA1'-CGSGHYFNIF** (*TCR-51*), **DNA1-CGSGHYFNIF** (*TCR-52*), **DNA1'-CGSGKLVFFA** (*TCR-53*) and **DNA1-CGSGKLVFFA** (*TCR-54*) were carried out to isolate the desired conjugate from the reaction mixture.

TCR-51

TCR-51 was prepared for HIC purification by reacting **DNA1'-mal** (319 µM) with **CGSGHYFNIF** (6.38 mM) in unbuffered water at R.T. overnight (total reaction volume: 120 µL). The reaction mixture was centrifuged to separate the supernatant (SN1) and the

precipitate. The precipitate was washed with 80 µL of pH 9 borate buffer and then centrifuged to separate the supernatant (SN2) from the precipitate. This step was repeated four times, with the supernatants obtained following each wash cycle being labeled as SN3, SN4, SN5 and SN6, respectively. The UV-Vis spectra of SN1-SN6 obtained were compared with that of **DNA1'-mal** used for *TCR-51* (Figure 3.36 a). 10% denaturing PAGE of SN2-SN6 was conducted to confirm the presence of the conjugate (**DNA1'-CGSGHYFNIF**), as well as to assess for any remaining unreacted DNA following each washing cycle. SN2-SN6 were diluted to ~5 µM before being analysed by PAGE.

The UV-Vis measurements revealed that the absorption of **DNA1'-mal** substantially decreased after it had reacted with **CGSGHYFNIF**, showing a significantly reduced DNA absorption in SN1. This is presumably because the majority of the conjugate precipitated with unreacted peptide. Notably, SN2 showed an increased DNA absorption, indicating that washing with borate buffer played a role in partially dissolving unreacted DNA and the conjugate in the supernatant. The UV-Vis spectra of SN3-SN6 showed a steady decrease in DNA absorption, indicating that unreacted DNA and the conjugate were gradually released to the supernatant from the precipitate over the repeated washing with borate buffer. 10% PAGE also confirmed the presence of both unreacted DNA and the conjugate. However, the band intensity of unreacted DNA became fainter as the borate buffer washing proceeded, whereas the conjugate band showed a similar intensity throughout SN2-SN6. This result demonstrated that unreacted DNA dissolved more readily in borate buffer compared with the conjugate, which was slowly released in the supernatant after each washing cycle.

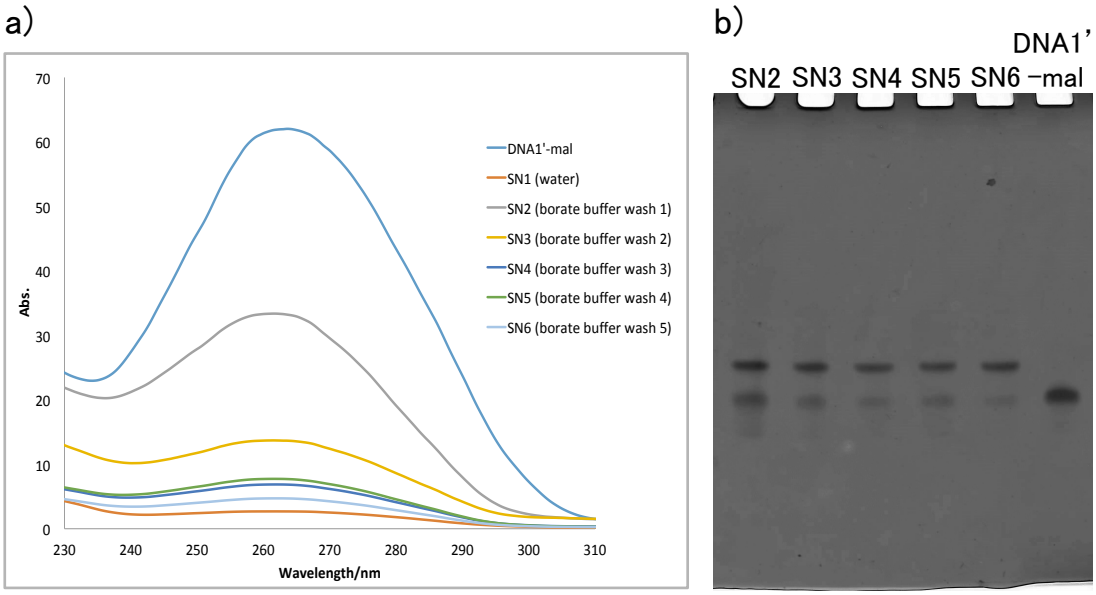


Figure 3.36. a) UV-Vis spectra of **DNA1'-mal** used in *TCR-51* and SN1-SN6 obtained following each round of *TCR-51* precipitate washing. **b)** 10% denaturing PAGE of 10 µM **DNA1'-mal** control and SN2-SN6.

***TCR-51* SN2**

In an attempt to isolate the desired conjugate from the reaction mixture, HIC was used to attempt the complete removal of remaining peptide and unreacted DNA from the supernatant. UV-Vis measurements conducted following HIC purification confirmed DNA absorption at 260 nm of varying intensity in all the fractions (F1-F15), indicating that unreacted DNA and the conjugate (**DNA1'-CGSGHYFNIF**) were gradually released from the resin over the course of Na_2SO_4 washing (Figure 3.37 a). The highest DNA absorption was observed in F1 (Abs.=9.3), whereas F10-F15 (2<Abs.<3) showed a lower absorption, with F2-F9 showing the lowest absorption of the analysed fractions (Abs.<1.0).

The subsequent 10% denaturing PAGE revealed that most of the unreacted DNA was released from the resin, given that F1 showed a strong band of unreacted DNA and a faint conjugate band (Figure 3.37 b). F10-F12 showed a clear conjugate band and a faint band of unreacted DNA. Meanwhile, F13-F15 only showed single conjugate bands, indicating that **DNA1'-CGSGHYFNIF** was successfully separated from unreacted DNA. Across the PAGE results, Na_2SO_4 potentially affected the migration of the samples given that the band

position upshifted at higher Na_2SO_4 concentrations, whereas it downshifted with lower salt concentrations.

The LC-MS result of F15 further confirmed the presence of **DNA1'-CGSGHYFNIF**, with a single peak at 19 min in the RP-HPLC chromatograph (Figure 3.38 a). Although a strong absorbance was confirmed on the chromatograph, the intensity in the mass peak of the conjugate was low (Figure 3.38 b). This was potentially due to the formation of metal adduct that could have suppressed the ionisation of the conjugate. The gel result and subsequent LC-MS analysis of F15 confirmed that the desired conjugate was successfully isolated following HIC purification.

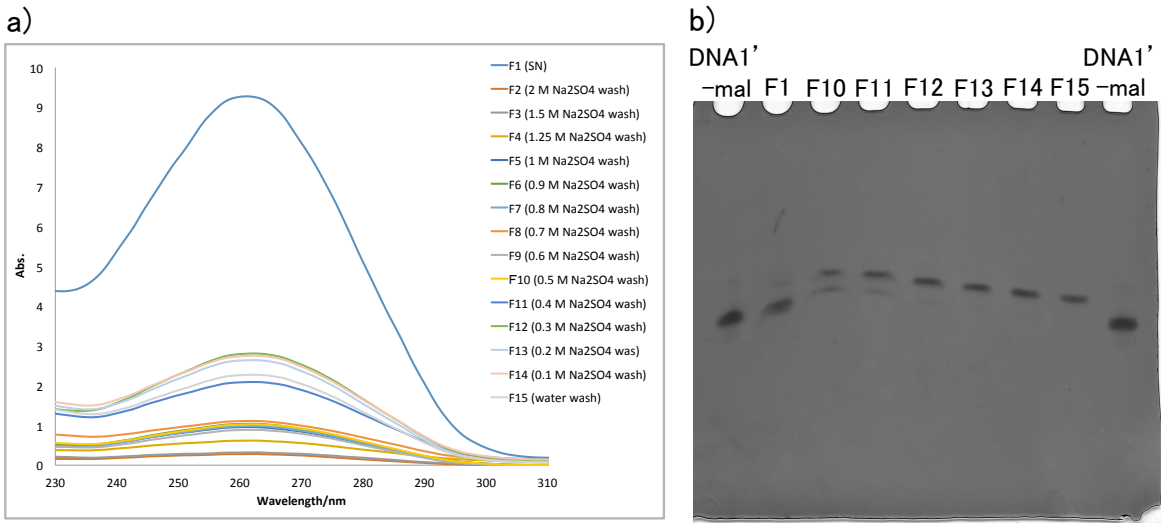


Figure 3.37. a) UV-Vis spectra of **F1-F15** obtained following HIC purification of *TCR-51* SN2. **b)** 10% denaturing PAGE of 10 µM **DNA1'-mal** control, F1 and F10-F15.

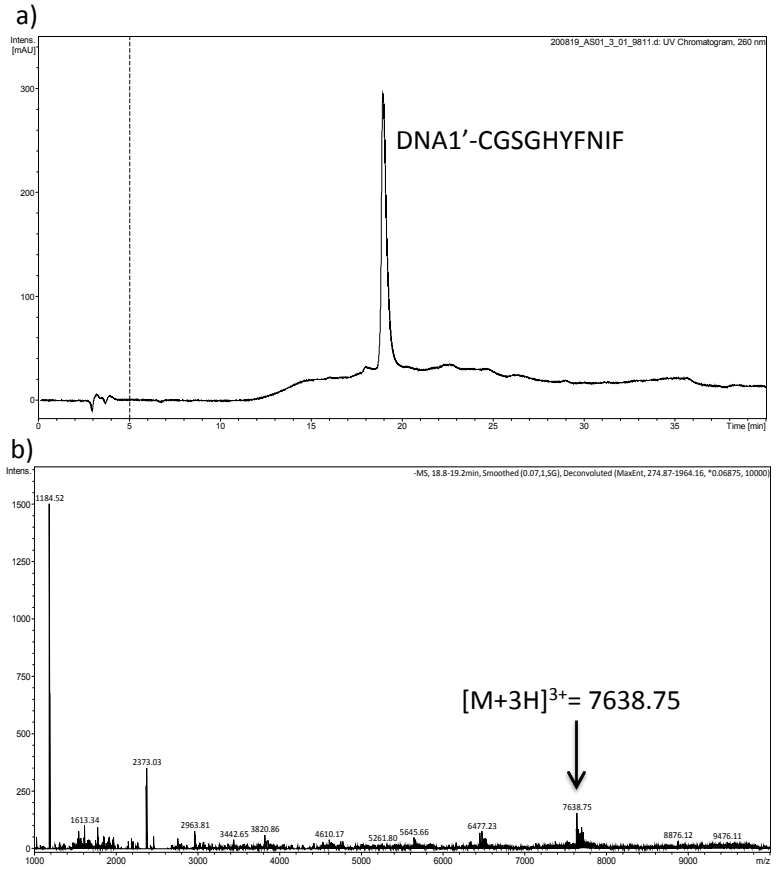


Figure 3.38. a) Analytical HPLC chromatograph of *TCR-51* F15 (monitoring DNA absorbance at 260 nm). **DNA1'-CGSGHYFNIF** eluted at 19 min. **b)** Expanded trace of the **DNA1'-CGSGHYFNIF** peak. LC-MS (m/z) calc. for: $[M+3H]^{3+}$ 7638.69, found 7638.75.

TCR-52

TCR-52 was prepared for HIC purification by reacting **DNA1-mal** (217 µM) with **CGSGHYFNIF** (4.34 mM) in unbuffered water at R.T. overnight (total reaction volume: 100 µL). The reaction mixture was centrifuged to separate the supernatant (SN1) from the precipitate. The precipitate was washed with 70 µL of unbuffered water, with the supernatant obtained labeled as SN2. The precipitate washing was repeated four times using 50 µL of pH 9 borate buffer, with the supernatants obtained following each washing cycle being labeled as SN3, SN4, SN5 and SN6, respectively.

The UV-Vis spectra of SN1-SN6 obtained were compared with that of **DNA1-mal** used in *TCR-51* (Figure 3.39 a). The absorption of **DNA1-mal** substantially decreased after it had reacted with **CGSGHYFNIF**, consistent to the UV-Vis of *TCR-51*. Meanwhile, the DNA

absorption remained very low in the supernatant (SN2) obtained after the precipitate was washed with water, suggesting that the conjugate (**DNA1-CGSGHYFNIF**) was still present within the precipitate. In contrast, the DNA absorption was higher in SN3 after the precipitate was washed with borate buffer. The DNA absorption showed a steady decline for SN4-6, compared with that of SN3, suggesting that the majority of the conjugate was released from the precipitate into SN3.

***TCR-52* SN3**

HIC purification of SN3 was conducted to isolate the desired conjugate. UV-Vis measurements conducted following HIC purification confirmed DNA absorption of varying intensity in all the fractions (F1-F15), indicating that unreacted DNA and the conjugate were gradually released from the resin over the course of Na_2SO_4 washing (Figure 3.40 a). 10% denaturing PAGE analysis showed that most of the unreacted DNA was released from the resin, given that F1 showed a strong band of unreacted DNA and a faint conjugate band (Figure 3.40 b). F10 and F11 showed clear conjugate bands and faint bands of unreacted DNA. Meanwhile, F12-F15 showed strong conjugate bands and almost undistinguishable unreacted DNA bands, suggesting that almost all of the unreacted DNA was removed, with the sample mostly consisting of **DNA1-CGSGHYFNIF**. In addition, the LC-MS result of F15 further confirmed the presence of **DNA1-CGSGHYFNIF** (Figure 3.41 b), with a single peak at 19 min in the HPLC chromatograph (Figure 3.41 a).

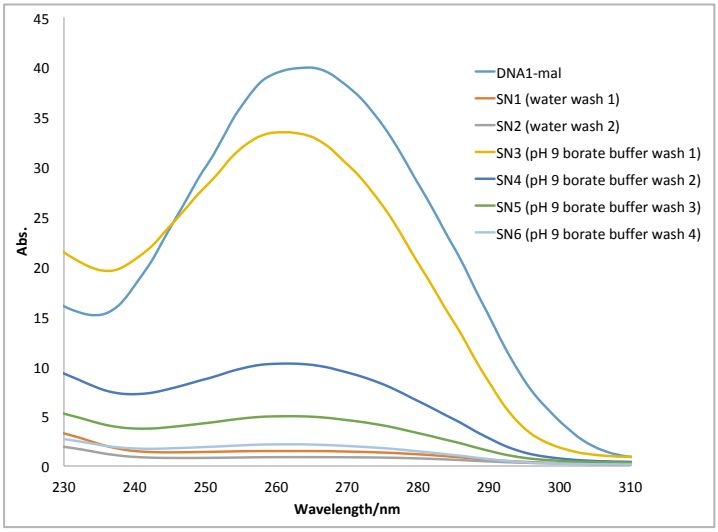


Figure 3.39. a) UV-Vis spectra of **DNA1-mal** used in *TCR-52* and SN1-SN6 obtained following each round of *TCR-52* precipitate washing.

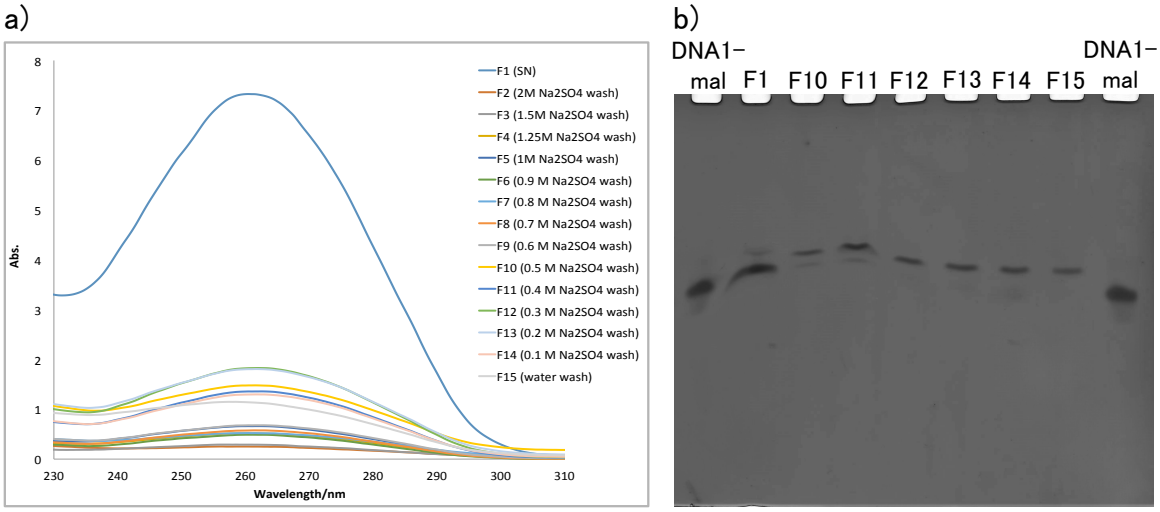


Figure 3.40. a) UV-Vis spectra of F1-F15 obtained following HIC purification of *TCR-52* SN3. **b)** 10% denaturing PAGE of 10 µM **DNA1-mal** control, F1 and F10-F15. F1 showed a clear band of unreacted **DNA1-mal**; F10-F11 showed a conjugate band and a faint band of unreacted **DNA1-mal**. F12-F15 showed a single band of **DNA1-CGSGHYFNIF**.

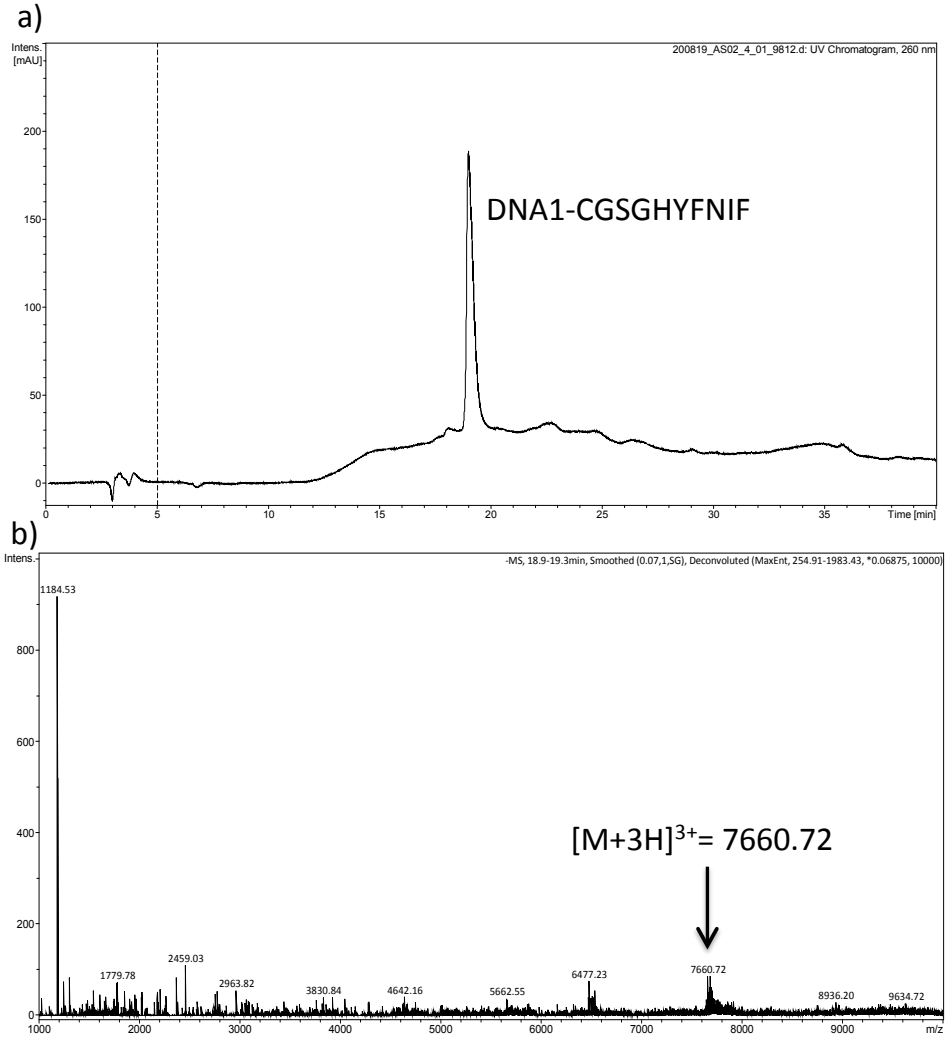


Figure 3.41. a) HPLC chromatograph of *TCR-52* F15 (monitoring DNA absorbance at 260 nm). **DNA1-CGSGHYFNIF** eluted at 19 min. **b)** Expanded trace of the **DNA1-CGSGHYFNIF** peak. LC-MS (m/z) calc. for: $[M+3H]^{3+}$ 7660.68, found 7660.72.

TCR-53

TCR-53 was prepared for HIC purification by reacting **DNA1'-mal** (215 µM) with **CGSGKLVFFA** (4.3 mM) in unbuffered water at R.T. overnight (total reaction volume: 150 µL). The reaction mixture was centrifuged to separate the supernatant (SN1) and the precipitate. The precipitate was washed with 50 µL of unbuffered water and then centrifuged to separate the supernatant (SN2) from the precipitate. The precipitate was then washed with 50 µL of pH 9 borate buffer, followed by 30 µL of pH 9 borate buffer six times, with the supernatants obtained following each wash cycle being labeled as SN3-SN9, respectively. The LC-MS result of SN7 showed a strong conjugate (**DNA1'-CGSGKLVFFA**) peak on the

chromatograph, as well as smaller peaks corresponding to impurities (Figure 3.43), indicating that washing the precipitate with borate buffer helped to remove a large amount of unreacted DNA from the supernatant. 10% PAGE of SN9 also confirmed that most of the unreacted DNA was removed by repeating the borate buffer washing step, showing a prominent conjugate band and a very faint band of unreacted DNA (Figure 3.45).

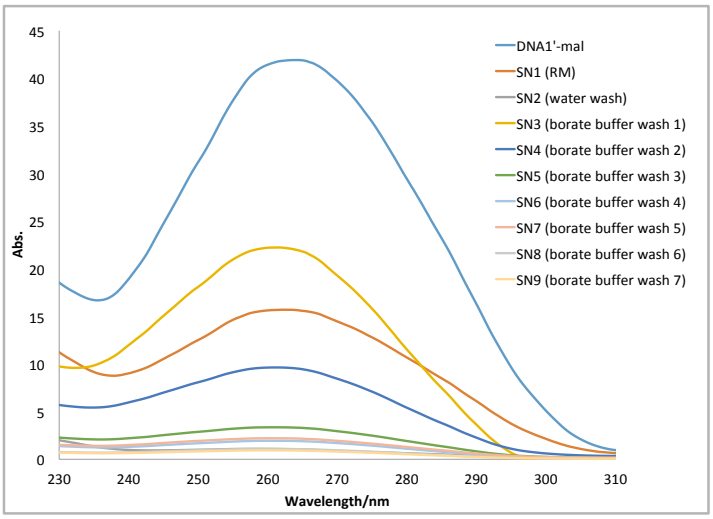


Figure 3.42. a) UV-Vis spectra of **DNA1'-mal** used in *TCR-53*, SN1 (supernatant of *TCR-53*) SN2 obtained following *TCR-53* precipitate washing with water, SN3-SN9 obtained following each round of *TCR-53* precipitate washing with borate buffer.

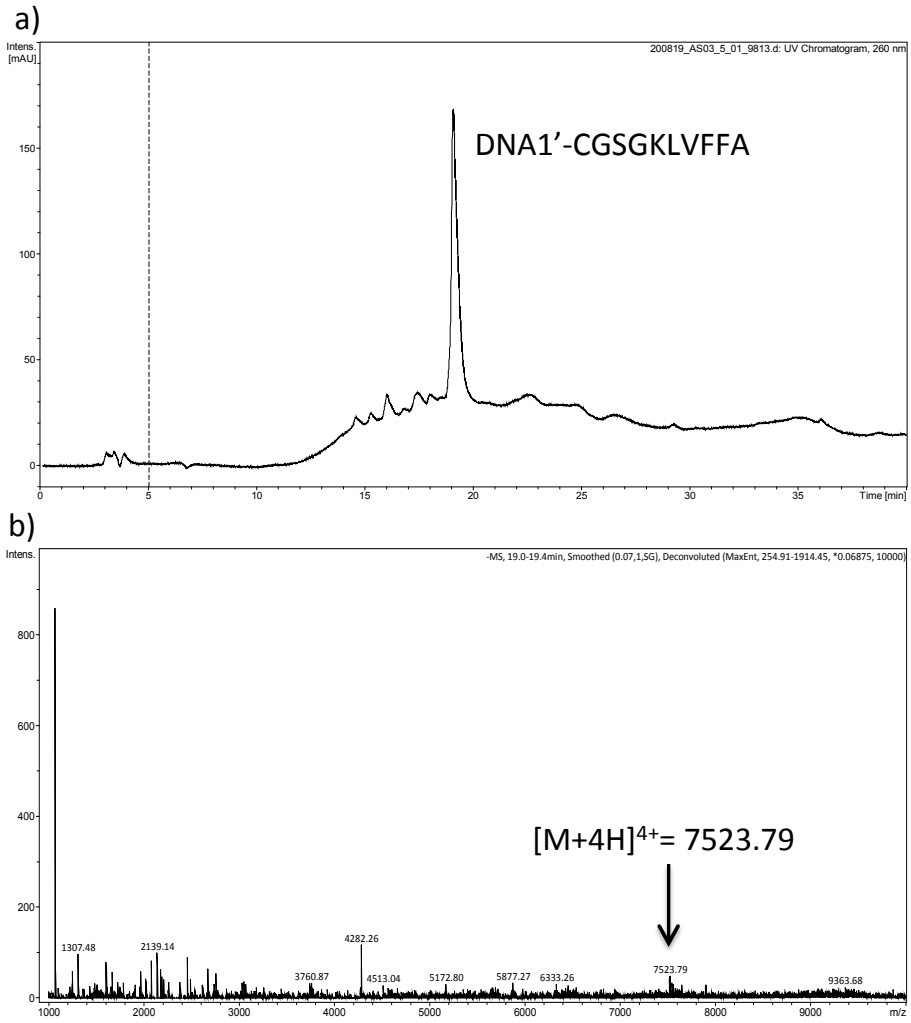


Figure 3.43. a) Analytical HPLC chromatograph of *TCR-53 SN7* (monitoring DNA absorbance at 260 nm). **DNA1'-CGSGKLVFFA** eluted at 19 min. **b)** Expanded trace of the **DNA1'-CGSGKLVFFA** peak. LC-MS (m/z) calc. for: $[M+4H]^{4+}$ 7523.46, found 7523.79.

***TCR-53* SN3**

HIC purification of SN3 was conducted to isolate the desired conjugate. UV-Vis measurements conducted following HIC purification confirmed DNA absorption at 260 nm in all the fractions (Figure 3.44). However, the remaining fractions (F2-F14) showed a very low intensity of DNA absorption. The 10% denaturing PAGE of F1 showed a prominent conjugate band and unreacted DNA band, suggesting that a large portion of conjugate did not adhere to the resin (Figure 3.45). This happened presumably because the amount of resin used was not sufficient for the amount of conjugate sample that was loaded. However, repeated washing of the resin with Na_2SO_4 led to the gradual removal of unreacted DNA (fractions

9-13), given that strong conjugate bands where observed, with the bands of unreacted DNA being almost negligible. Complete removal of unreacted DNA appeared to have been achieved in F14, given that only a single conjugate band was seen on the gel (Figure 3.45).

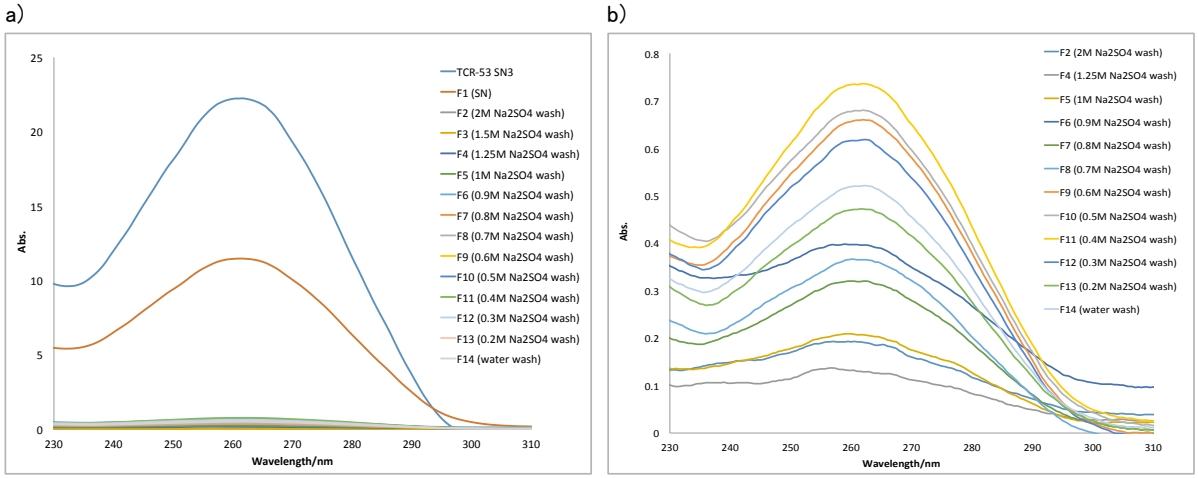


Figure 3.44. a) UV-Vis spectra of F1-F15 obtained following HIC purification of *TCR-53* SN3. **b)** UV-Vis spectra of F2-F14.

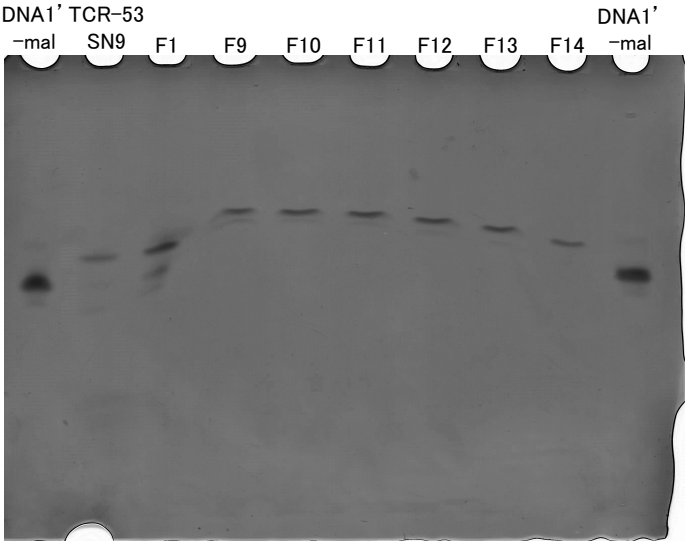


Figure 3.45. 10% denaturing PAGE of 10 µM **DNA1'-mal** control, *TCR-53 SN9*, F1 and F9-F14.

TCR-54

TCR-54 was prepared for HIC purification by reacting **DNA1-mal** (263 µM) with **CGSGKLVFFA** (5.33 mM) in unbuffered water at R.T. overnight (total reaction volume: 100 µL). The reaction mixture was centrifuged to separate the supernatant (SN1) from the precipitate. The precipitate was washed with 50 µL of pH 9 borate buffer three times, with the

supernatants obtained following each wash cycle being labeled as SN2, SN3 and SN5, respectively. The precipitate was subsequently washed with a smaller amount of pH 9 borate buffer (30 µL ×1, 25 µL ×1 and 20 µL ×2), labeled as SN6, SN7, SN8 and SN9, respectively. SN2 was used for HIC purification given that it showed the highest DNA absorption amongst all of the obtained supernatants.

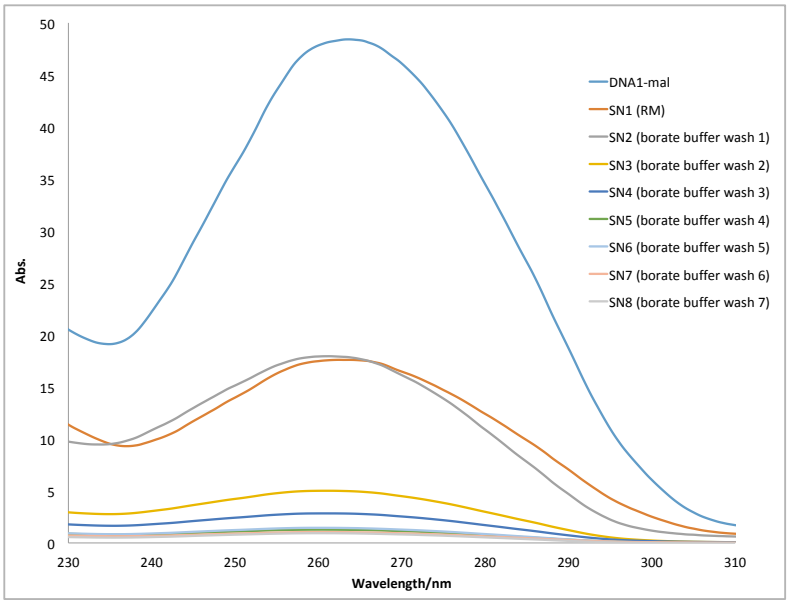


Figure 3.46. UV-Vis spectra of **DNA1-mal** used in *TCR-54*, SN1 (supernatant of *TCR-54*), SN2-SN9 obtained following each round of *TCR-43* precipitate washing with borate buffer.

TCR-54 SN2

UV-Vis measurements following HIC purification of the fractions confirmed DNA absorption at 260 nm of varying intensity (Figure 3.47 a). The 10% denaturing PAGE showed prominent conjugate bands and very faint bands of unreacted DNA in F1, F9-F14 (Figure 3.47 b). The gel study of F15 only showed a single conjugate band, suggesting that **DNA1-CGSGKLVFFA** was successfully isolated from unreacted DNA.

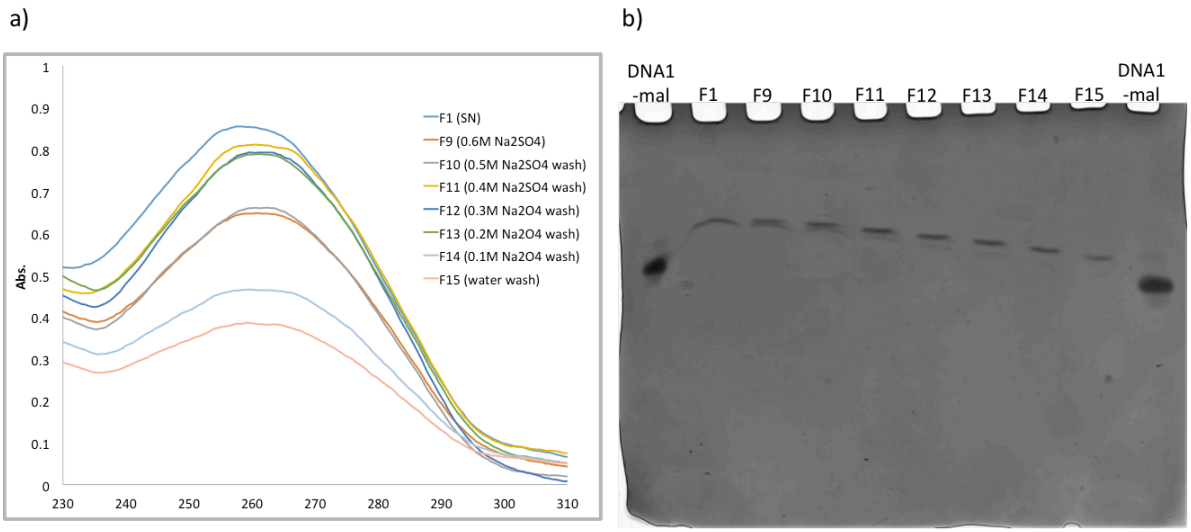


Figure 3.47. a) UV-Vis spectra of F1, F9-F14 obtained following HIC purification of *TCR-54* SN2 **b)** 10% denaturing PAGE of 10 µM **DNA1-mal** control, F1 and F9-F15 obtained following HIC purification of *TCR-54* SN2.

3.4 Summary

Using thiol-maleimide chemistry, the formation of desired conjugates including **DNA-CGSGHYFNIF**, **DNA-CGSG(POG)$_6$** and **DNA-CGSGKLVFFA** was confirmed by 10% denaturing PAGE and LC-MS. However, full conversion from **DNA-mal** to conjugate was not achieved due to the presence of ring-opened forms of the DNA maleimide groups before reacting with peptides.

To isolate the conjugate from the reaction mixture, four different purification strategies were tested. PAGE purification failed to separate the conjugate from unreacted DNA, since two bands were observed on the gel study. Meanwhile, PR-HPLC purification did not show any bands on the gel study, suggesting that DNA degradation occurred.

Anion exchange failed to separate the conjugate from unreacted DNA when no prior precipitate washing step was conducted, showing two bands on the gel study. In contrast, when the precipitate in the reaction mixture was washed with borate buffer prior to conducting anion exchange, most of the unreacted DNA was removed given that the gel study showed a prominent conjugate band and a very faint unreacted DNA band.

HIC purification proved successful in isolating the desired conjugates (**DNA-CGSGHYFNIF** and **DNA-CGSGKLVFFA)** from their reaction mixtures following washing of the precipitates with borate buffer, given that only single conjugate bands were observed in the gel study. LC-MS further confirmed a single peak of **DNA-CGSGHYFNIF**.

Table 3.2. Summary of Thiol-maleimide conjugation reactions

Reaction No.	Reactants	Conditions	Results
TCR-1	10 µM of **DNA1-mal**; 10 µM of **DNA1'-mal**; 100 µM of **CGSGHYFNIF**	Overnight reaction at R.T. or 37°C in pH 8 TBE buffer (850 µL)	20% denaturing PAGE showed only DNA-mal bands, showing no formation of conjugates
TCR-2	10 µM of **DNA1-mal**; 10 µM of **DNA1'-mal**; 4.2 mM of **CGSGHYFNIF**	Overnight reaction at 37°C in pH 7.4 PBS buffer/HFIP or water/HFIP (200 µL)	20% denaturing PAGE showed two new bands in addition to DNA-mal band, suggesting that DNA and peptide had reacted
TCR-3	10 µM of **DNA1-mal**; 10 µM of **DNA1'-mal**; 0.1 mM, 0.2 mM, 0.3 mM or 0.4 mM of **CGSGHYFNIF**	Overnight reaction at 37°C in pH 7.4 PBS buffer/HFIP (170 µL)	20% denaturing PAGE showed only DNA-mal bands, showing no formation of conjugates
TCR-4	10 µM of **DNA1-mal**; 10 µM of **DNA1'-mal**; 1 mM, 2 mM, 3 mM or 4 mM of **CGSGHYFNIF**	Overnight reaction at 37°C in pH 7.4 PBS buffer/HFIP (200 µL)	Absence of bands on 20% denaturing PAGE
TCR-5	10 µM of **DNA1-mal**; 10 µM of **DNA1'-mal**; 1 mM, 2 mM, 3 mM or 4 mM of **CGSGHYFNIF**	Overnight reaction at 37°C in pH 7.4 PBS buffer/HFIP (200 µL)	Absence of bands on 20% denaturing PAGE
TCR-6	10 µM of **DNA1-mal**; 10 µM of **DNA1'-mal**; 2mM or 4 mM of **CGSGHYFNIF**	Overnight reaction at 37°C in water (400 µL)	20% denaturing PAGE showed no bands in SN, whereas stained wells in blue were observed in PPT, suggesting

			aggregate formation in reaction mixture
TCR-7	10 μM of **DNA1-mal**; 10 μM of **DNA1'-mal**; 2 mM of **CGSGHYFNIF**	Overnight reaction at 37°C in pH 7.4 PBS buffer/8M urea (800 μL)	10% denaturing PAGE showed only DNA-mal bands. Stained wells were observed in PPT
TCR-8	10 μM of **DNA1'-mal**; 0.1, 0.5, 1 or 2 mM of **CGSGHYFNIF**, **CHYFNIF** and **C(POG)$_6$**	Overnight reaction at 37°C in water, pH 7.4 PBS buffer or pH 7.2 phosphate buffer (80 μL)	10% denaturing PAGE showed conjugate and DNA-mal bands when reactions were carried out in water with ≧50 equivalents of CGSGHYFNIF or ≧100 equivalents of CHYFNIF
TCR-9	10 μM of **DNA1'-mal**; 0.1, 0.5, 1 or 2 mM of **CGSGHYFNIF**, **CHYFNIF** and **C(POG)$_6$**	Overnight reaction at 37°C in pH 7.4 PBS buffer/8M urea (80 μL). Repeat experiment of *TCR-8*	10% denaturing PAGE showed conjugate and DNA-mal bands when reactions were carried out in water with ≧50 equivalents of CGSGHYFNIF or ≧100 equivalents of CHYFNIF
TCR-10 to TCR-29	DNA1 or DNA1' (10 or 20 μM) were reacted with a variety of peptide (**CGSGHYFNIF**, **C(POG)$_6$** and **CGSG(POG)$_6$**) at different concentrations	Overnight reaction at variable temperatures (30, 37, 42, 60 and 90 °C), pH 6.6-7.4 and solvents (water, phosphate buffer or PBS buffer)	Absence of DNA or conjugate bands on 10% or 20% PAGE. The absence of bands was later found out to have been caused by the absence of maleimide group from DNA
TCR-30	50 μM of **DNA1-mal**; 50 μM of **DNA1'-mal**; 2.5 mM of **CGSGHYFNIF** and **CGSG(POG)$_6$**	Overnight reaction at 37°C in pH 6.8 phosphate buffer (200 μL)	20% denaturing PAGE and LC-MS confirmed the presence of desired conjugates
TCR-31	50 μM of **DNA1-mal**; 50 μM of **DNA1'-mal**; 2.5 mM of **CGSGHYFNIF**	Overnight reaction at 50°C in pH 6.8 phosphate buffer	20% denaturing PAGE confirmed the presence of desired conjugates as well as

	CGSGKLVFFA or **CGSG(POG)$_6$**	(200 µL)	unreacted DNA-mal, except for DNA1'-mal + CGSGKLVFFA
TCR-32	50 µM of **DNA1-mal**; 50 µM of **DNA1'-mal**; 2.5 mM of **CGSGHYFNIF** and **CGSG(POG)$_6$**	Overnight reaction at R.T. in pH 6.8 phosphate buffer (200 µL)	20% denaturing PAGE confirmed the presence of desired conjugates as well as unreacted DNA-mal
TCR-33	50 µM of **DNA1-mal**; 50 µM of **DNA1'-mal**; 2.5 mM of **CGSGHYFNIF** and **CGSG(POG)$_6$**	2h reaction at 4°C or R.T. in water. HFIP was added for reaction with CGSGHYNIF (100 µL)	20% denaturing PAGE confirmed the presence of desired conjugates as well as unreacted DNA-mal
TCR-34	449 µM of **DNA1-mal**; 9.0 mM of **CGSGHYFNIF**	Overnight reaction at 37°C in pH 6.8 PBS buffer/HFIP, 1:4 v/v, (200 µL). Preparation of reaction mixture for PAGE purification	20% denaturing PAGE showed unreacted DNA and conjugate bands. PAGE purification was conducted, but the separation of the conjugate from unreacted DNA was unsuccessful
TCR-35	290 µM of **DNA1-mal**; 5.8 mM of **CGSGHYFNIF**	Overnight reaction at R.T. in water (580 µL). Preparation of reaction mixture for RP-HPLC	RP-HPLC was conducted, but no evidence of successful separation was obtained
TCR-36	*TCR-34* reaction mixture containing **DNA1-CGSGHYFNIF** and **DNA-mal**	Overnight incubation of 10 µL *TCR-34* at 37°C in pH 9 borate buffer (1:9, v/v)	LC-MS showed a strong peak of t **DNA1-CGSGHYFNIF** conjugate. RP-HPLC trace suggested an improved conversion of DNA to conjugate
TCR-37	396 µM of **DNA1'-mal**; 7.95 mM of	Overnight reaction at R.T. in water	20% denaturing PAGE and LC-MS confirmed

	CGSGKLVFFA	(520 µL). Preparation of reaction mixture for RP-HPLC	DNA1'-mal-CGSGHYFNIF in the reaction mixture. RP-HPLC did not work, indicating sample degradation during the purification
TCR-38	396 µM of **DNA1'-mal**; 6.23 mM of **CGSGKLVFFA**	Overnight reaction at R.T. in water (240 µL). Preparation of reaction mixture for RP-HPLC	Repeat of *TCR-37* using the same reactants. RP-HPLC purification was conducted, but no evidence of successful separation was obtained
TCR-39	472 µM of **DNA1'-mal**; 7.1 mM of **CGSGKLVFFA**	Overnight reaction at R.T. in water (240 µL). Preparation of reaction mixture for RP-HPLC	RP-HPLC purification was conducted, but no evidence of successful separation was obtained
TCR-40	472 µM of **DNA1'-mal**; 7.95 mM of **CGSGKLVFFA**	Overnight reaction at R.T. in water /HFIP, 7:1 v/v, (240 µL). Preparation of reaction mixture for RP-HPLC	RP-HPLC purification was conducted, but no evidence of successful separation was obtained
TCR-41	TCR-34 reaction mixture containing **DNA1-mal-CGSGHYFNIF** and **DNA-mal**	Overnight incubation of 15 µL *TCR-34* at 37°C in pH 9 borate buffer (1:1, v/v)	20% denaturing PAGE showed DNA-mal and conjugate bands of similar intensity
TCR-42	*TCR-34* reaction mixture containing **DNA1-mal-CGSGHYFNIF** and **DNA1-mal**	Overnight incubation of 10 µL *TCR-34* at 37°C in pH 9 borate buffer (1:9, v/v), followed	20% denaturing PAGE showed DNA-mal and conjugate bands of similar intensity in SN, whereas a substantially faint DNA band and a more

		by anion exchange purification	prominent conjugate band in PPT. Anion exchange purification was successful, showing a single band of conjugate on 10% PAGE
TCR-43	194 μM of **DNA1-mal**; 3.75 mM of **CGSGHYFNIF**	Overnight reaction at R.T. in water (180 μL). Sample preparation to conduct *TCR-44*	10% denaturing PAGE showed DNA-mal and conjugate bands of similar intensity
TCR-44	*TCR-43* reaction mixture containing **DNA1-mal-CGSGHYFNIF** and **DNA1-mal**	Overnight incubation of 10 μL TCR-43 at R.T. in pH 9 borate buffer (1:9, v/v)	20% denaturing PAGE showed DNA-mal and conjugate bands of similar intensity in SN, whereas a substantially faint DNA band and a more prominent conjugate band in PPT
TCR-45	453 μM of **DNA1'-mal**; 9.37 mM of **CGSGHYFNIF**	Overnight reaction at R.T. in water (90 μL), followed by borate buffer washing. Sample preparation to conduct *TCR-46*	10% denaturing PAGE showed that appearance of unreacted DNA band became substantially fainter after four cycles of borate buffer washing, showing an essentially single band of conjugate
TCR-46	*TCR-45* reaction mixture containing **DNA1'-CGSGHYFNIF** and **DNA1'-mal**	Overnight incubation of *TCR-45* (20 μL) at R.T. in pH 9 borate buffer (1:9, v/v).	10% denaturing PAGE showed DNA-mal and conjugate bands
TCR-47	272 μM of **DNA1'-mal**;	Overnight reaction	10% denaturing PAGE showed

	5.43 mM of **CGSGHYFNIF**	at R.T. in water (90 µL)	DNA-mal and conjugate bands
TCR-48	229 µM of **DNA1'-mal**; 5.18 mM of **CGSGHYFNIF** Supernatant of *TCR-45*	3.5h reaction at R.T. in water (90 µL), followed by incubation in pH 9 borate buffer at R.T. o/n	10% denaturing PAGE showed DNA-mal and conjugate bands
TCR-49	219 µM of **DNA1'-mal**; 4.39 mM of **CGSGHYFNIF**	Overnight reaction at R.T. in water (98 µL). Preparation of reaction mixture for HIC	10% denaturing PAGE showed a conjugate band and a more prominent conjugate band
TCR-50	143 µM of **DNA1'-mal**; 2.9 mM of **CGSGHYFNIF**	Overnight reaction at R.T. in water (100 µL). Attempt to extract the conjugate by borate buffer washing	10% denaturing PAGE showed a conjugate band, along with the disappearance of unreacted DNA following five cycles of borate buffer washing
TCR-51	319 µM of **DNA1'-mal**; 6.4 mM of **CGSGHYFNIF**	Overnight reaction at R.T. in water (120 µL). Preparation of reaction mixture for HIC	The desired conjugate was successfully purified by HIC, being confirmed by 10% denaturing PAGE and LC-MS
TCR-52	217 µM of **DNA1-mal**; 4.3 mM of **CGSGHYFNIF**	Overnight reaction at R.T. in water (100 µL). Preparation of reaction mixture for HIC	The desired conjugate was successfully purified by HIC, being confirmed by 10% denaturing PAGE and LC-MS
TCR-53	215 µM of **DNA1'-mal**; 4.3 mM of	Overnight reaction at R.T. in water	The desired conjugate was successfully purified by HIC,

	CGSGKLVFFA	(150 µL). Preparation of reaction mixture for HIC	being confirmed by 10% denaturing PAGE
TCR-54	263 µM of **DNA1-mal**; 5.3 mM of **CGSGKLVFFA**	Overnight reaction at R.T. in water (100 µL). Preparation of reaction mixture for HIC	The desired conjugate was successfully purified by HIC, being confirmed by 10% denaturing PAGE

3.5 References

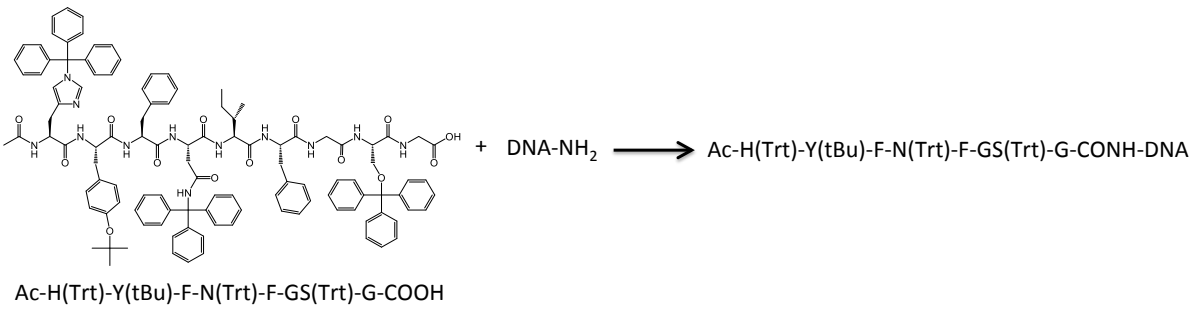
1. C. Shen, in *Diagnostic Molecular Biology*, Elsevier, Amsterdam, 2019, ch. 8, pp. 187–214.
2. H. Summer, R. Grämer, and P. Dröge, *J Vis Exp.*, 2009, **32**, 1485.
3. S. Lopez-Gomollon and F. E. Nicolas, in *Methods in Enzymology*, Elsevier, Amsterdam, 2013, vol. 529, ch. 6, pp. 65–83.
4. A. Petrov, T. Wu, E. V. Puglisi, and J. D. Puglisi, in *Laboratory Methods in Enzymology: RNA*, ed. J. Lorsch, Elsevier, Amsterdam, 1st edn, 2013, vol. 530, ch. 17, pp. 315–330.
5. J. T. McCue, *Methods Enzymol.*, 2009, **463**, 405–414.
6. T. A. Young, et al, in *Methods in Enzymology*, Elsevier, Amsterdam, 2018, vol. 610, ch. 8, pp. 167–190.
7. G. Hermanson, in *Bioconjugate Techniques*, 3rd edn, 2013, ch, 15, pp. 589–740.
8. S. Kirchhof et al., *J. Mater. Chem. B*, 2015, **3**, 449–457.
9. Y. Q. Wang, S. S. Wang, J. Zhu, L. Wang, B. H. Jiang and W. J. Zhao, *Journal of Food and Drug Analysis*, 2016, **24**, 399–405.
10. L. N. Tumey, et al., *Bioconjugate Chem.*, 2014, **25**, 1871–1880.

4 Peptide-Oligonuceotide Conjugation Using Amide Coupling

4.1 Chapter Overview

Due to the difficulties encountered in obtaining the desired conjugates using thiol-maleimide chemistry, amide coupling was tested as an alternative conjugation strategy. To conduct amide-coupling conjugation, synthesis of a side-chain protected peptide with a free C-terminus (see Chapter 2) was required in order to react it with **DNA-NH_2** (Scheme 1).

This section covers the results of amide-coupling conjugation reactions, including Kaiser tests, UV-Vis spectra and PAGE studies. Each amide-coupling conjugation reaction described in this chapter will be referred to as *ACR* (*amide-coupling reaction*), followed by a number (e.g. *ACR-1*). For reference, the side-chain protected peptide (**Ac-H(Trt)-Y(tBu)-F-N(Trt)-F-GS(Trt)-G-COOH**) used for amide coupling will be abbreviated as **Ac-HYFNIFGSG-COOH** across this chapter.



Scheme 1. Amide-coupling conjugation showing a side-chain protected peptide with a free C-terminus (**Ac-H(Trt)-Y(tBu)-F-N(Trt)-F-GS(Trt)-G-COOH**) reacting with **DNA-NH_2**, resulting in the formation of the POC: **Ac-H(Trt)-Y(tBu)-F-N(Trt)-F-GS(Trt)-G-CONH-DNA**.

4.2 Methods

Amide-coupling reactions were carried out using appropriate coupling reagents in DMF at R.T. (20-25°C) or 37°C overnight. Following reactions, peptide side-chain deprotection was carried out with an acid (TFA, TCA or HCl). The acid was removed under a stream of N_2 after the reaction had occurred, and the pellet obtained was suspended in 200 µL of autoclaved water, which was then desalted using Zetadex-25 (10 mL). The eluent from the Zetadex column was collected as Fraction 1 (F1). Subsequently, the column was washed with 1 mL of

autoclaved water, and the eluent was collected as F2. This washing procedure was repeated until all the sample was removed from the column. UV-Vis measurements were conducted to estimate the concentration of DNA in the fractions. Each fraction was quantified based on A260 (DNA absorption at 260 nm) using a NanoDrop spectrophotometer, which was then concentrated for a subsequent PAGE study to assess for the presence of the desired conjugate.

PAGE studies were carried out using either 10% or 20% denaturing polyacrylamide gels. 10% PAGE gels were run in 1× TBE buffer at a constant current of 15 mA and at a set voltage of 300 V for 20 min. 20% PAGE gels were run in 1× TBE buffer at a constant current of 15 mA and at a set voltage of 300 V for 60 min.

The Kaiser test was employed to determine the presence of primary amines, which would confirm whether **DNA-NH_2** had reacted with the peptide. The test is based on the ninhydrin reaction with primary amines, which produces a characteristic blue colour. In the absence of **DNA-NH_2**, the ninhydrin reaction does not occur and the solution remains yellow in colour. To conduct this test, two drops of reaction mixture and two drops of **DNA-NH_2** solution were added into two separate glass vials using a Pasteur pipette. Subsequently, two drops of Reagent A, B and C were added to each vial, respectively. The vials were heated at 110 °C for 5 min to assess for colour changes. The full details of the Kaiser test protocol were previously described in Chapter 2.

LC-MS was used to identify the products obtained based on their molecular mass. For LC-MS measurements, 50 µL of 10 µM solution was prepared in a mass spec vial.

AFM analysis was conducted from *ACR-10* onwards to assess for structural changes of the reaction-mixture components. The desalted fraction that showed the highest DNA absorption by UV-Vis were analysed by AFM. In *ACR-12*, the sample was compared against different controls, which included DNA control (**DNA-NH_2**), peptide control (**Ac-HYFNIFGSG-COOH**) and a control consisting of unreacted DNA and peptide in solution.
To prepare samples for AFM analysis, 20 µL of each sample was deposited onto freshly

cleaved mica for 10 min, with the excess liquid being wiped off with the edge of a filter paper. Subsequently, the mica surface was dried under a stream of N_2. The full methodology of AFM analysis was previously described in Chapter 2.

4.3 Results and Discussion

Exploration of Suitable Reaction Conditions (*ACR-1* to *ACR-3*)

In *ACR-1*, **DNA1-NH_2** (100 µM) in water (9.3 µL) was reacted with 1.1 equivalents of **Ac-HYFNIFGSG-COOH** using 2.2 equivalents of TBTU, HOBt and DIPEA in DMF at 37°C overnight (total reaction volume: 71 µL). Subsequently, the reaction mixture was desalted and the collected fractions 1-5 were analysed by 20% denaturing PAGE (Figure 4.1). The gel study only showed unreacted DNA bands, suggesting that the DNA did not react with the peptide. This was possibly due to the presence of water in the reaction mixture, which may have prevented amide bond formation or reacted preferentially with the activated ester, given the relative concentrations.

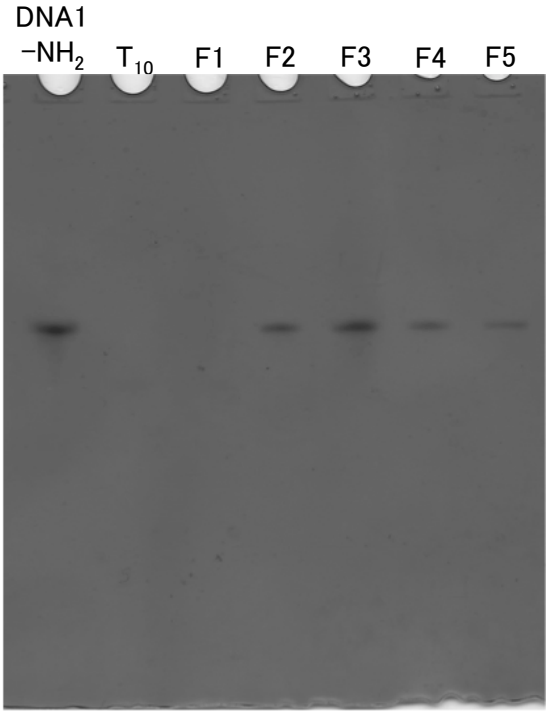


Figure 4.1. Denaturing 20% PAGE of 10 µM **DNA1-NH_2** control, **T_{10} DNA** and fractions 1-5 obtained in *ACR-1* following desalting. **T_{10} DNA** was run on the gel on behalf of a different student, but this was not part of the current research project.

Following the result of *ACR-1*, the subsequent reaction (*ACR-2*) was conducted in DMF using freeze-dried **DNA1-NH_2** (20 µM), which was reacted with **Ac-HYFNIFGSG-COOH**, PyBOP, HOBt and DIPEA in a 2-fold molar excess at 37°C overnight (total reaction volume: 161 µL). Subsequently, peptide side-chain deprotection was carried out by treating the pellet with a cleavage solution consisting of TFA/H_2O/TIPS (94:3:3 v/v) for 80 min, followed by desalting. The collected fractions 1-8 were then analysed by 10% denaturing PAGE, which only showed unreacted DNA bands (Figure 4.2).

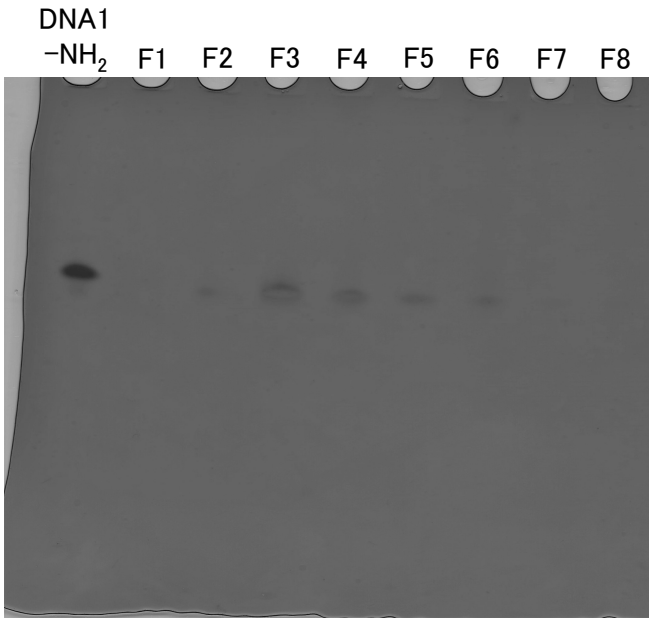


Figure 4.2. Denaturing 20% PAGE of 10 µM **DNA1-NH_2** control and fractions 1-8 obtained in *ACR-2* following desalting.

The absence of conjugation bands in previous reactions may have occurred due to the low concentration of reaction mixture that was used. Consequently, to overcome this, *ACR-3* was conducted in more concentrated conditions using 100 µM of **DNA1-NH_2**, which was then reacted with **Ac-HYFNIFGSG-COOH**, PyBOP, HOBt and DIPEA in a 2-fold molar excess at 37°C overnight (total reaction volume: 71 µL). After the reaction, peptide side-chain deprotection was carried out by treating the pellet with TFA/H_2O/TIPS (94:3:3 v/v) for 60 min, followed by desalting.

UV-Vis of fractions 4 and 5 showed prominent absorption peaks at around 270 nm, being

slightly different compared with the characteristic DNA absorption at 260 nm (Figure 4.3). This may have suggested that DNA reacted with the peptide. In addition, fraction 5 showed a slightly red-shifted spectrum with a broad absorption range between 300-330 nm, possibly due to the presence of coupling reagents.

To assess whether DNA had reacted in fraction 4 as indicated by the UV-Vis result, the Kaiser test was conducted to check for the presence of primary amines, with **T10 DNA** and **DNA-NH_2** being used as controls (Figure 4.4). The Kaiser test result was negative for fraction 4 and for **T10 DNA**, confirming the absence of primary amines, and thereby suggesting that DNA potentially reacted. The Kaiser test result for **DNA-NH_2** was positive, which was expected due to the presence of primary amines.

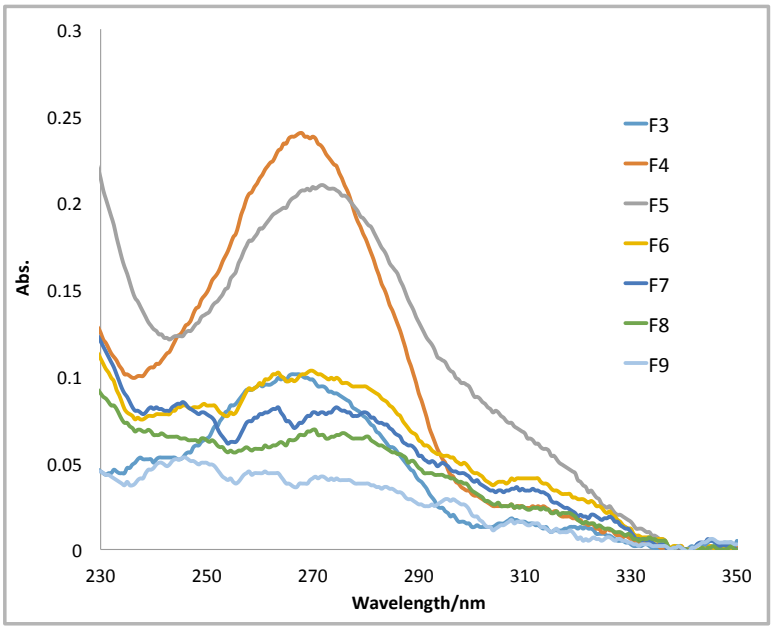


Figure 4.3. UV-Vis spectra of fractions 3-9 in *ACR-3* following desalting.

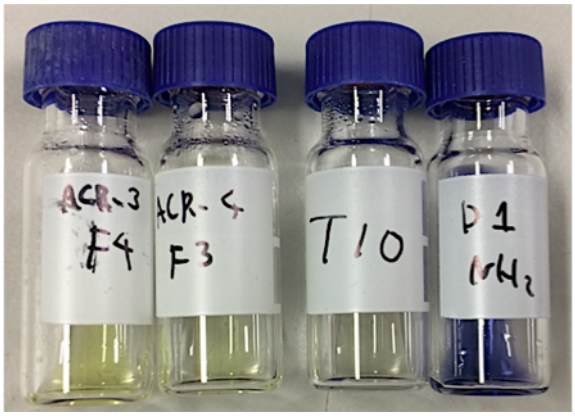


Figure 4.4. The results of the Kaiser test from left to right are as follows: fraction 4 of *ACR-3*, fraction 3 of *ACR-4*, **T10 DNA** and **DNA-NH_2**. Only **DNA-NH_2** showed a positive result, as indicated by the blue colour.

10% denaturing PAGE of fractions 1-5 appeared to show bands slightly below the **DNA-NH_2** control (Figure 4.5 a). To confirm that these bands were different from the unreacted DNA bands, **DNA-NH_2** control was added to all the fractions. The subsequent PAGE study confirmed that reacted DNA bands were present just below the **DNA-NH_2** control bands, suggesting that the DNA reacted (Figure 4.5 b). Given that these new bands were observed below the DNA control, it indicated that the resulting products had a smaller mass than **DNA-NH_2** or the desired conjugate. It is possible that the use of strong acidic conditions by employing TFA for peptide side-chain deprotection led to DNA degradation, which would have resulted in a smaller DNA sequence than the original **DNA-NH_2** control.[1]

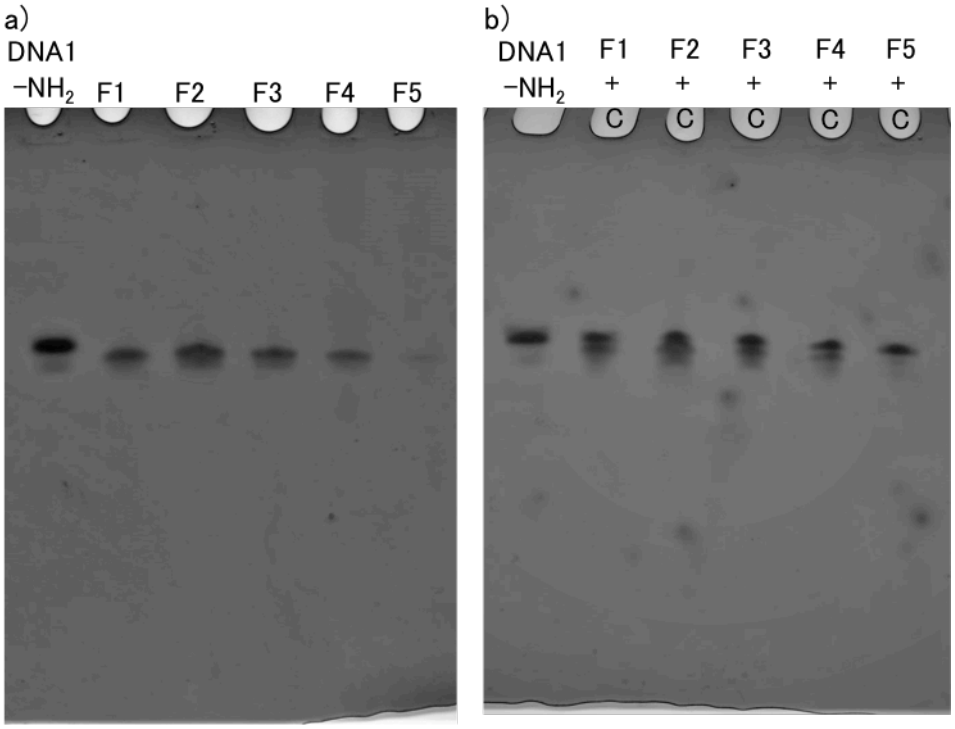


Figure 4.5. Denaturing 10% PAGE of **DNA1-NH_2** control and **a)** fractions 1-5 obtained in *ACR-3* following desalting, and **b)** fractions 1-5 mixed with **DNA1-NH_2** control.

Exploration of the Gel Findings (*ACR-6* and *ACR-8*)

Additional conjugation reactions (A*CR-6* and *ACR-8*) employing the same conditions as in ACR-3 were carried out to obtain further clarity on the PAGE studies previously described. For ACR-6, fractions 4 and 5 were submitted for LC-MS analysis. The LC-MS showed no peaks above 2000 m/z (data not shown), implying that the conjugate was not present in the samples since this value was smaller than its predicted mass of 7366 m/z. However, it may also have been possible that the samples were not ionised in the LC-MS system, failing to detect DNA or conjugate peaks.

To understand why PAGE or LC-MS did not show evidence of conjugate formation, it was subsequently investigated whether the conjugate may have been obtained prior to conducting the peptide side-chain deprotection step. In *ACR-8*, the reaction mixture was analysed by the Kaiser test and 10% denaturing PAGE without peptide side-chain deprotection. The Kaiser test result was negative, confirming the absence of primary amines and potentially suggesting that DNA reacted (Figure 4.6). The 10% PAGE study of ACR-8

showed a single band, whereas the gel study of *ACR-8* mixed with **DNA1-NH$_2$** control showed a band slightly above the DNA control band, which might have implied that a conjugate was formed (Figure 4.7).

Figure 4.6. Kaiser test results for **DNA1-NH$_2$** control (left) and *ACR-8* (right). The result was negative in ACR-8, confirming the absence of primary amines.

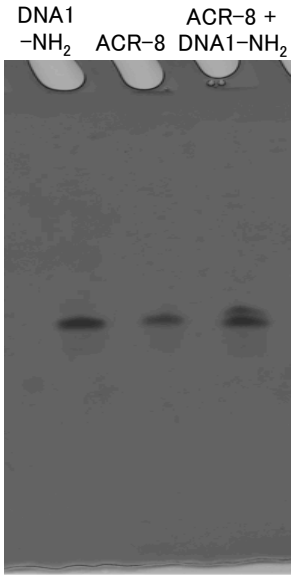


Figure 4.7. 10% denaturing PAGE of **DNA1-NH$_2$** control, *ACR-8* and *ACR-8* mixed with **DNA1-NH$_2$** control.

Deprotection Strategy using TAC (*ACR-9*)

Given that TFA might have caused DNA degradation in previous reaction mixtures based on PAGE data, an alternative acid was tested for peptide side-chain deprotection. In *ACR-9*, 100 µM of **DNA1-NH$_2$** was reacted with **Ac-HYFNIFGSG-COOH**, PyBOP, HOBt and DIPEA in a

2-fold molar excess at 37°C overnight (total reaction volume: 75 µL). After the reaction, peptide side-chain deprotection was carried out by treating the pellet with 5% TAC in DCM/ TIPS (94:6 v/v) for 90 min, followed by desalting. Collected fractions 3 and 4 were analysed by 10% denaturing PAGE (Figure 4.8).

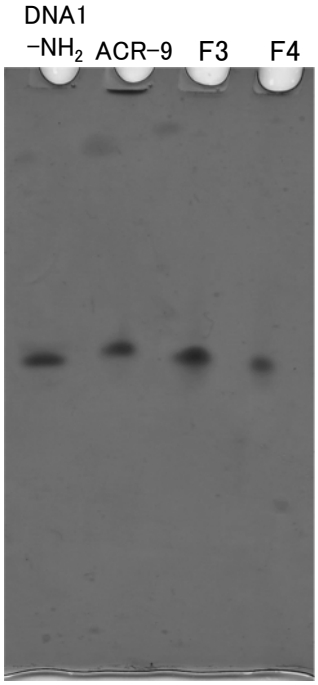


Figure 4.8. 10% denaturing PAGE of **DNA-NH_2** control, *ACR-9* prior to desalting and F3 and F4 of *ACR-9* obtained following desalting.

The PAGE study suggested that TAC did not cause DNA degradation, given that the bands for fractions 3 and 4 were on the same level as the **DNA-NH_2** control band. The lack of significant band shifts compared with the DNA control suggests that the conjugate was not formed. While the band corresponding to *ACR-9* prior to desalting was slightly above the DNA control, this was likely attributed to the possibility that DMF increased the density and viscosity of the sample. The presence of a stained well in the *ACR-9* lane may have been caused by precipitation of peptide, giving that there was an excessive amount of peptide in the reaction mixture.

Deprotection Strategy using HCl (*ACR-10*)

ACR-10 (**Ac-HYFNIFGSG-COOH** + **DNA1-NH_2**) was carried out using the same reaction

conditions as in *ACR-9*, with the exception that the deprotection step was carried out using HCl instead of TAC. The UV-Vis of *ACR-10* showed that fraction 3 had the highest DNA absorption, and was subsequently analysed by MALDI to determine the presence of the desired conjugate (Figure 4.9). However, the MALDI analysis did not show DNA or conjugate peaks (data not shown), suggesting that the desired conjugate was not present in the fraction. Given that the gel results of previous reaction mixtures had consistently shown DNA bands, the absence of a DNA peak in the MALDI analysis of *ACR-10* could suggest that the sample was not ionised, resulting in no DNA peaks.

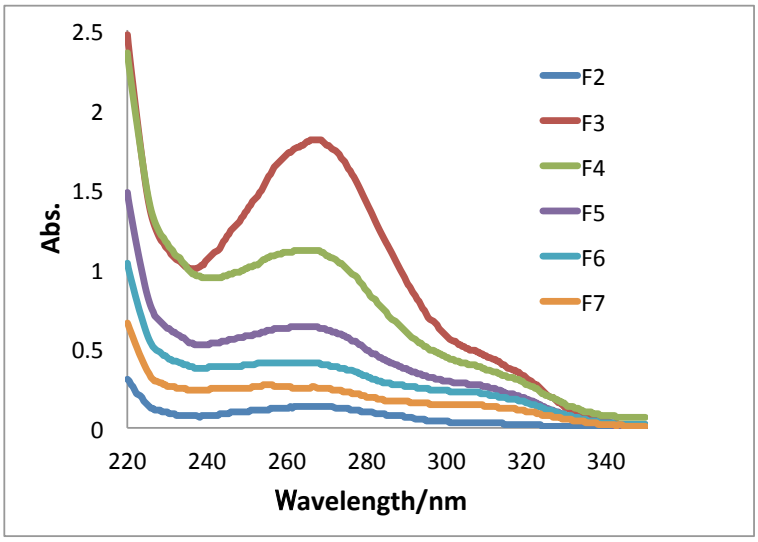


Figure 4.9. UV-Vis spectra of fractions 2-7 in *ACR-10* following desalting.

AFM Analysis of *ACR-10*

Given that the previous analytical techniques did not provide conclusive evidence of conjugation formation, AFM was conducted to assess for any structural differences between the reaction mixture and the DNA and peptide controls. It was hypothesised that the potential conjugate should display a different structure to either the non-conjugated reaction mixture, or DNA and peptide on their own. To investigate morphological changes of the sample over time, fraction 3 (9 µM) was incubated at R.T. and analysed at 0 h, 24 h and 7 days (Figure 4.10-12).

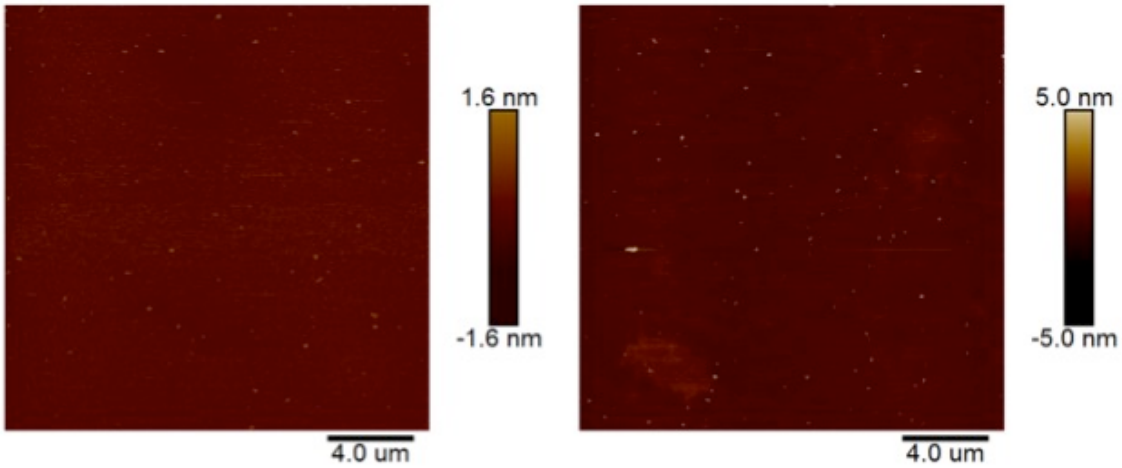


Figure 4.10. AFM images of *ACR-10* at 0 h.

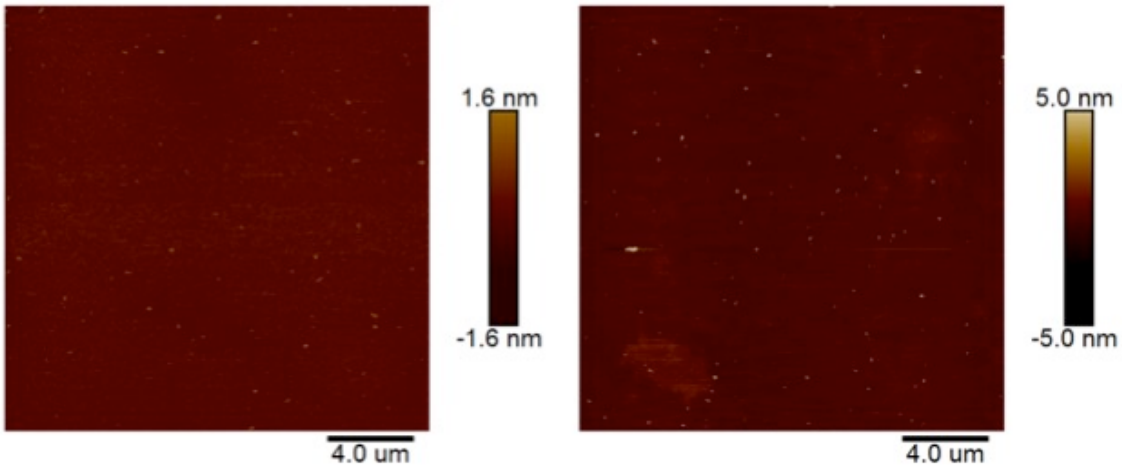
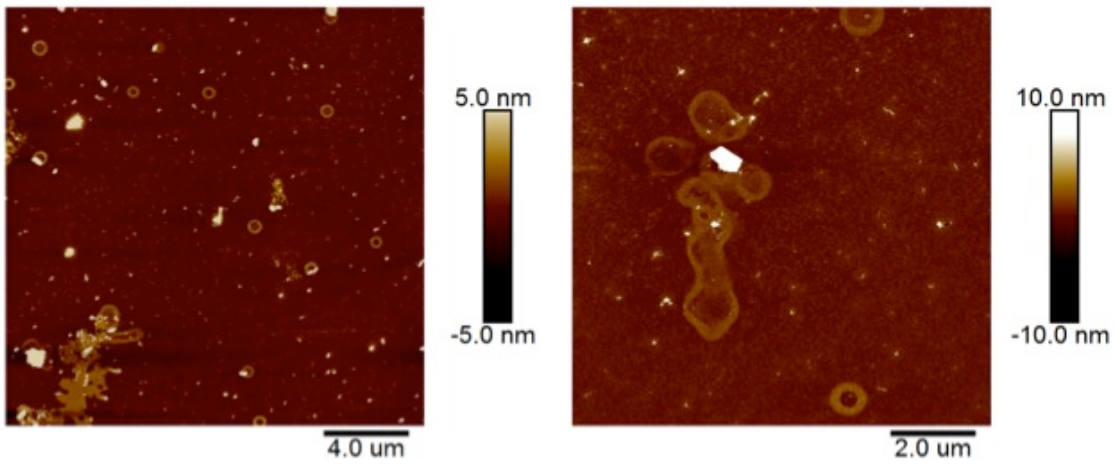


Figure 4.11. AFM images of *ACR-10* at 24 h.



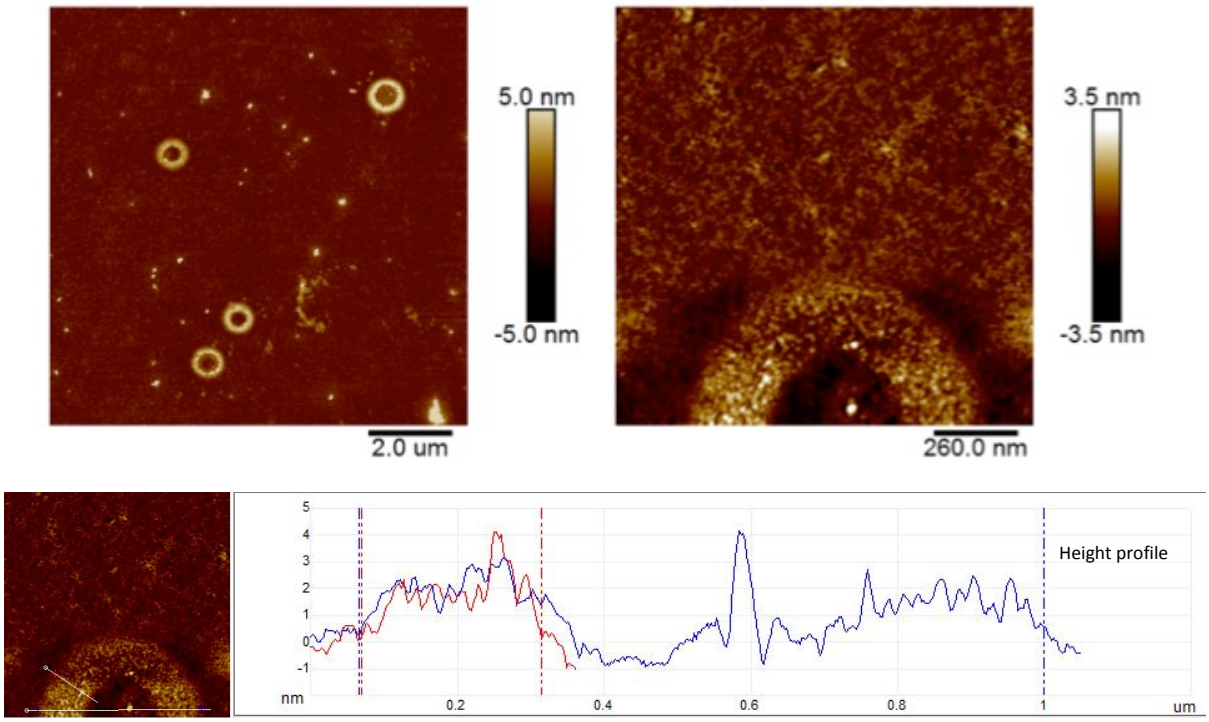


Figure 4.12. AFM images of *ACR-10* at 7 days and corresponding height profile.

At 0 h and 24 h, the AFM analysis of fraction 3 showed dot-like structures (Figure 4.10 and 4.11), whereas ring-like structures were observed at 7 days of aging. Some of these ring-like structures appeared to be flattened or collapsed, which could have indicated that vesicles were previously formed in the solution.[2] Since ring-like structures were not seen during the first 24 h, the potential formation of vesicles may have occurred after this time frame and before the final analysis conducted at 7 days. In contrast, the AFM analysis of unreacted DNA and peptide control only showed dots and fibres, suggesting that DNA and peptide had reacted, leading to the formation of new structures. Furthermore, the potential formation of vesicles in the AFM images of *ACR-10* suggested that amphiphilic molecules might have been present in the solution (Figure 4.12), which could have consisted of the DNA and peptide used in the reaction.

Complementary DNA (**DNA1'-NH_2**) was added to fraction 3 of *ACR-10* in a 1:1 ratio and left to stand at R.T. for 7 days to evaluate potential molecular interactions in solution. The AFM analysis (Figure 4.13) showed ring-like structures with thinner edges (~100 nm) compared to those observed in the sample without **DNA1'-NH_2** at 7 days of aging (~200 nm). In addition,

it appeared that there was fragmentation within the ring-like structures, with the rings showing little signs of uniformity, suggesting that the addition of DNA might have interfered with their formation. However, it was also possible that the addition of complementary DNA did not contribute to the morphological changes of the ring-like structures between the two samples, given that their diameter was relatively similar (0.5-1 µm for both samples).

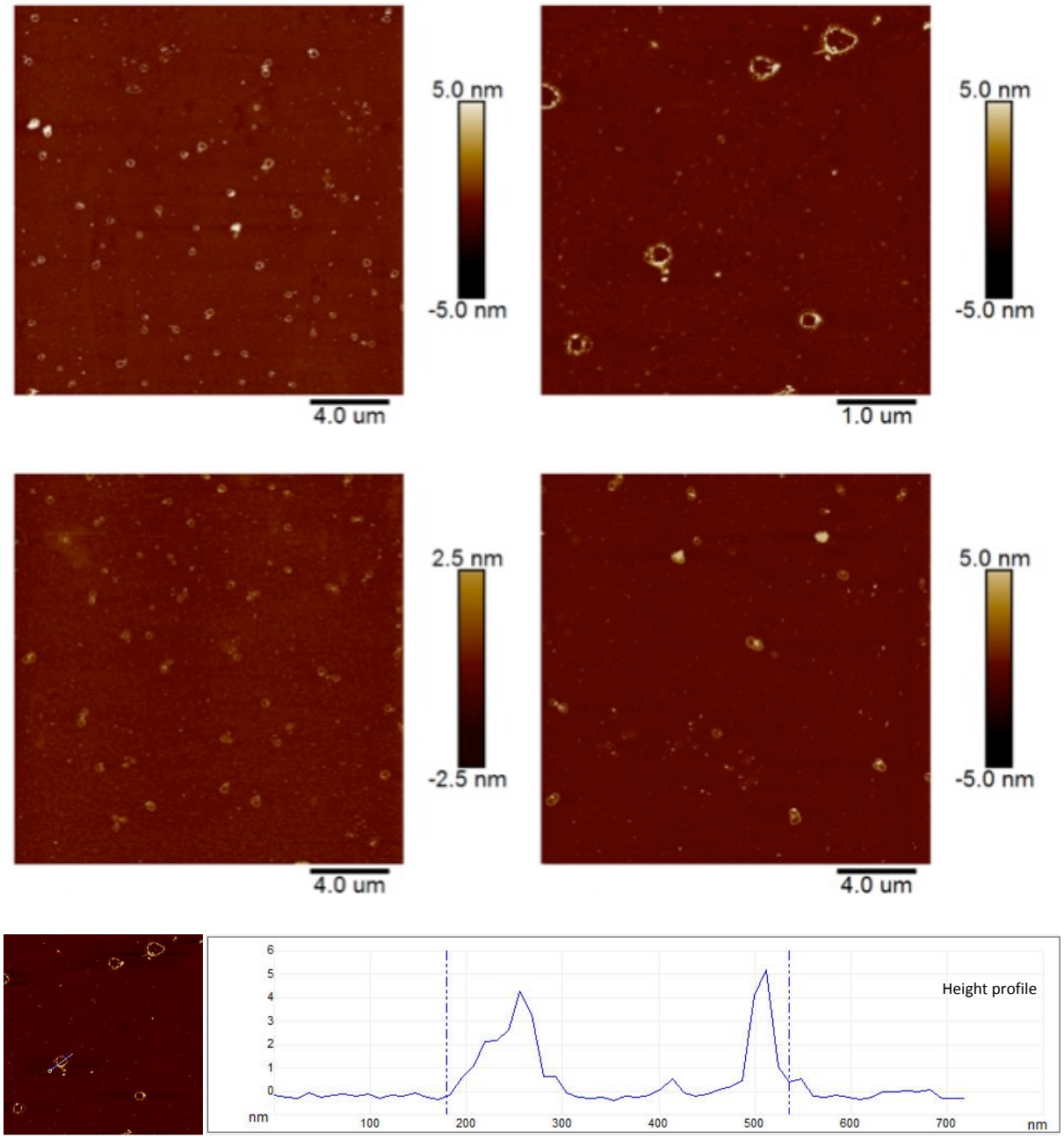


Figure 4.13. AFM images of *ACR-10* mixed with **DNA1'-NH_2** at 7 days and corresponding height profile.

Despite the possibility that a conjugate may have formed in *ACR-10*, there were limitations that need to be considered. Firstly, the presence of the desired conjugate in fraction 3 could not be confirmed as neither LC-MS or PAGE were conducted. Although MALDI analysis was carried out, the formation of amorphous crystals on the MALDI target possibly prevented ionisation, which resulted in the absence of oligonucleotide peaks (data not shown). Moreover, given that the aging experiments of *ACR-10* were not repeated, it is difficult to extrapolate the results obtained.

AFM Analysis of *ACR-12*

In *ACR-12,* **DNA1'-NH_2** was used for conjugation instead of **DNA1-NH_2**. In addition, the AFM sample was prepared with TAMg buffer to allow for efficient hybridisation upon the addition of complementary DNA. Fraction 3 of *ACR-12* was freeze-dried and suspended in TAMg buffer to make 10 μM solution, which was then mixed with complementary DNA (**DNA1-NH_2**) in a 1:1 ratio. The solution was left to stand at R.T. for 7 days to evaluate potential molecular interactions in solution. The AFM images obtained for *ACR-12* were compared with DNA control (**DNA-NH_2**), peptide control (**Ac-HYFNIFGSG-COOH**) and a control consisting of unreacted DNA and peptide in solution.

To analyse the self-assembly of **Ac-HYFNIFGSG-COOH** (peptide control) by AFM, the side-chain deprotection was carried out with a cleavage solution consisting of TFA/TIPS (97:3), followed by Et_2O precipitation. To prepare AFM samples, 10 μM and 1 mM peptide solutions were prepared in water, respectively, which were then left to stand at R.T. for 7 days. Following the incubation, two peptide control samples were individually prepared on mica for AFM analysis.

At a concentration of 10 μM, the AFM analysis of the peptide control showed small dot-like structures and thin fibril-like structures with a height of ~2 nm (Figure 4.14). In contrast, thin fibre-like structures with a height of 0.5 nm were observed at a concentration of 1 mM (Figure 4.15). These results revealed that the peptide control only showed a small numbers of fibres at a concentration $\leqq$10 μM, whereas it formed thin fibres at a concentration of 1

mM.

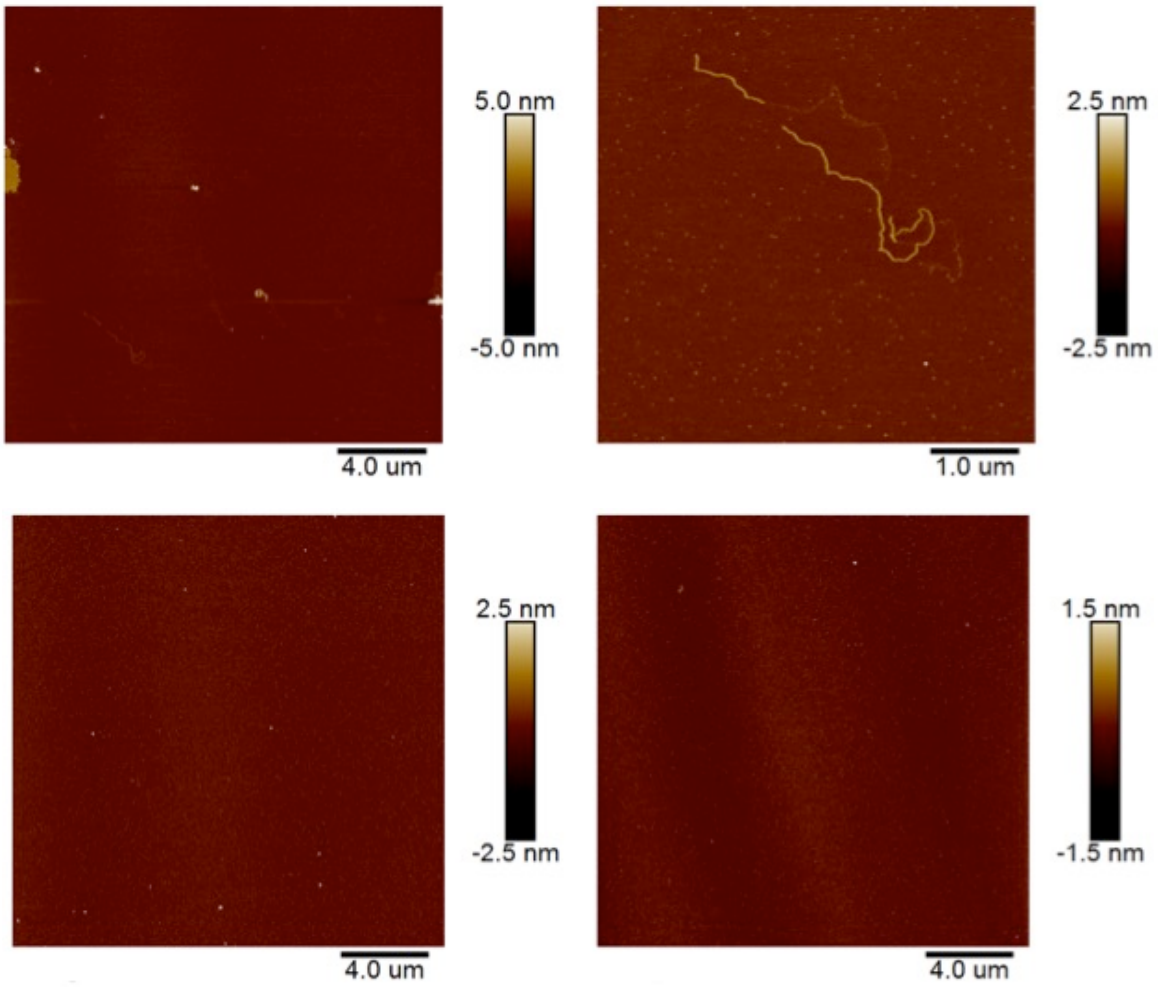
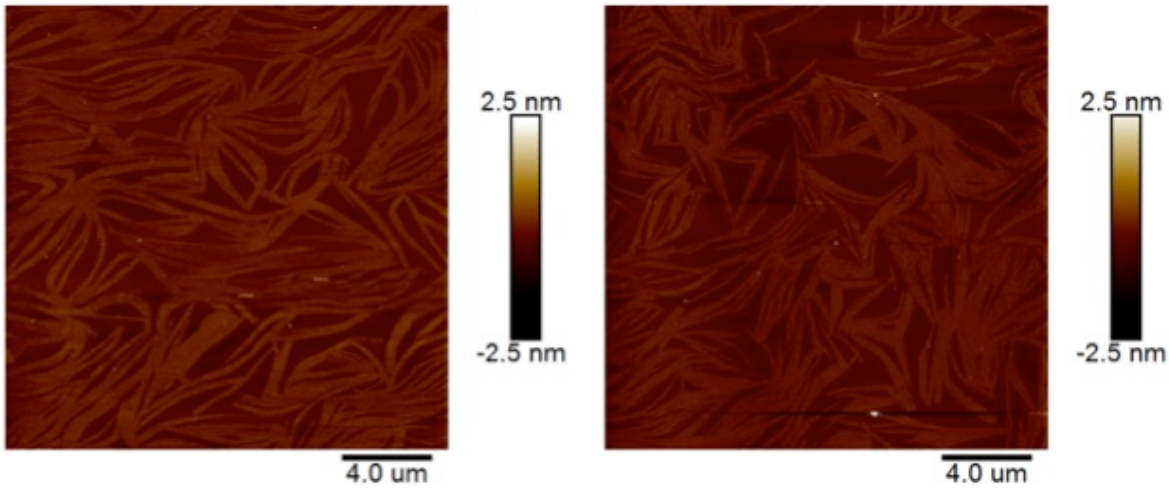


Figure 4.14. AFM images of peptide control (**Ac-HYFNIFGSG-COOH**, side-chain deprotected) in water at a concentration of 10 µM at 7 days, showing small dot-like structures and thin fibril-like structures.



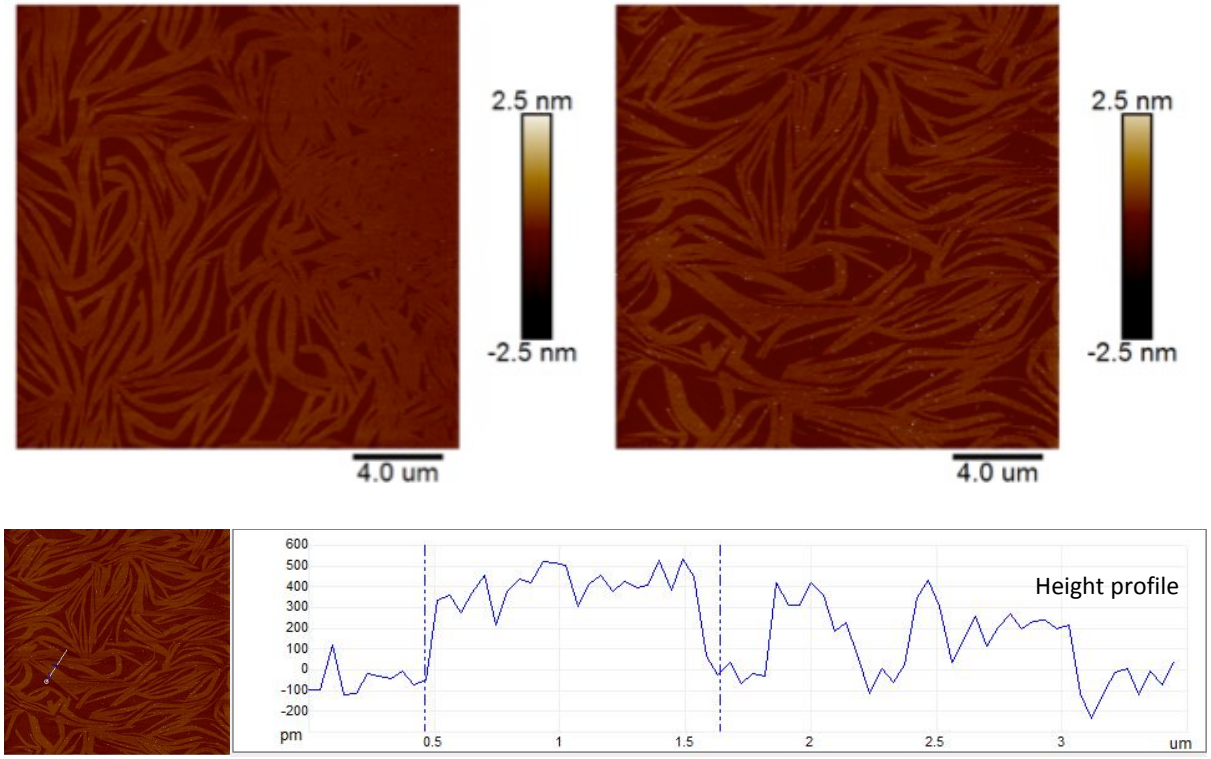


Figure 4.15. AFM images of peptide control (**Ac-HYFNIFGSG-COOH**, side-chain deprotected) in water at a concentration of 1 mM, showing thin fibre-like structures with a height of 0.5 nm.

To analyse the self-assembly of the DNA control (**DNA1-NH_2** mixed with **DNA1'-NH_2**) by AFM, 10 µM of **DNA1-NH_2** and 10 µM of **DNA1'-NH_2** were mixed in water in a 1:1 (v/v) ratio and left to stand at R.T. for 24 h. In addition, another DNA control sample was prepared in TAMg buffer and left to stand at R.T. for 7 days. Following incubation, two DNA control samples were individually prepared on mica for AFM. The analysis of both DNA control samples showed undefined structures (potentially artefacts), highlighting that the DNA components did not self-assemble into large structures on their own.

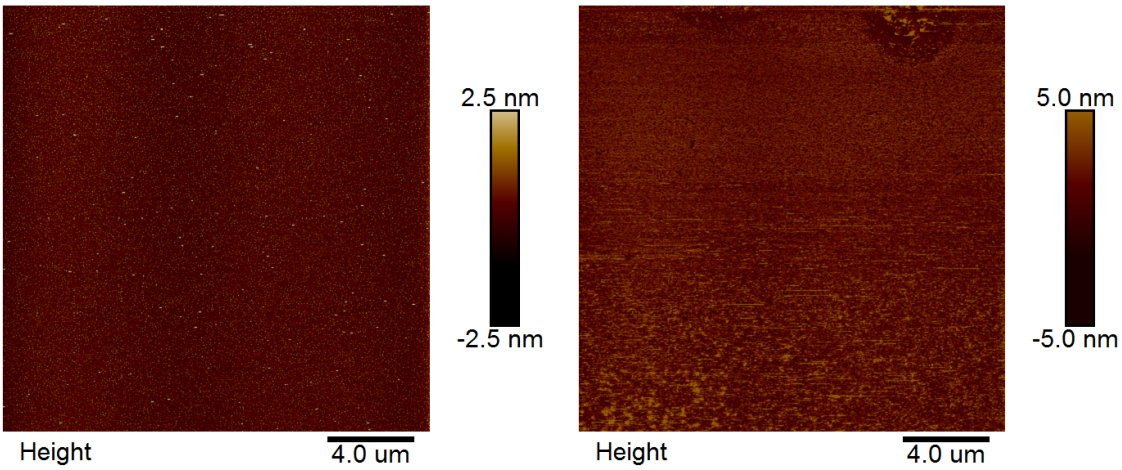


Figure 4.16. AFM images of DNA control (10 µM **DNA1-NH_2** mixed with 10 µM **DNA1'-NH_2**) in water at 24 h (left) and TAMg at 7 days (right).

To analyse the self-assembly of unreacted DNA and peptide control, 10 µM of **Ac-HYFNIFGSG-COOH** (side-chain deprotected) and 10 µM **DNA1'-NH_2** were mixed in water in a 1:1 (v/v) ratio without coupling reagents, which was then left to stand at R.T. for 7 days. Following incubation, the solution was prepared on mica for AFM. The analysis of this control sample showed small dot-like structures and fibril like structures, which were presumably formed as a result of the peptide self-assembly (Figure 4.17). However, there appeared to be no appreciable structural differences in this control sample compared with the peptide or the DNA controls previously described.

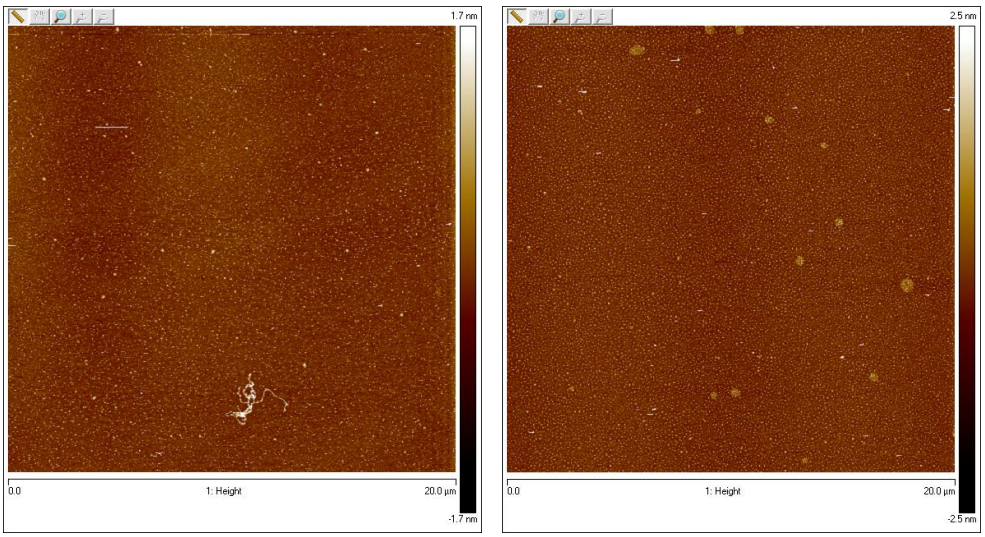
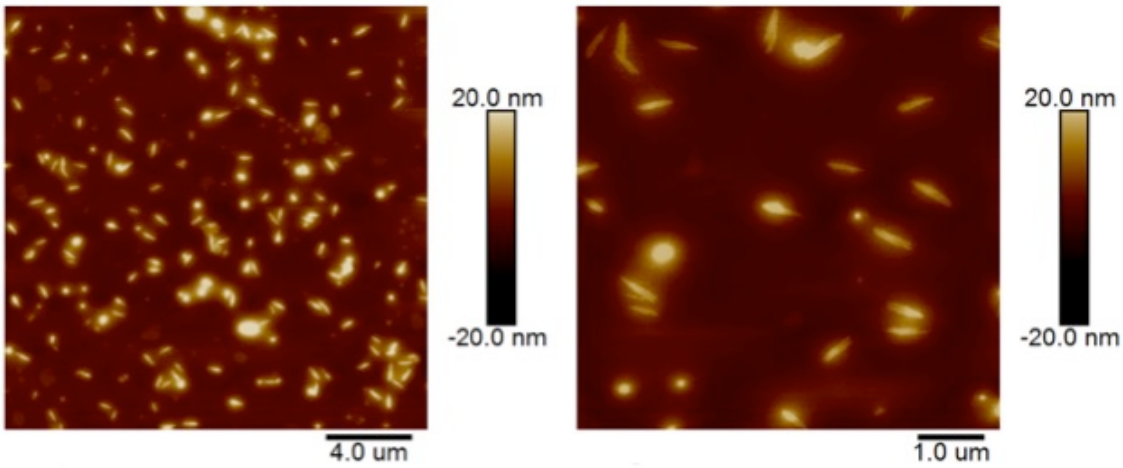


Figure 4.17. AFM images of unreacted DNA and peptide control. These two AFM images were not processed in a publication format because the original files were corrupted.

The AFM analysis of *ACR-12* fraction 3 revealed the formation of fibre-like structures with a height of 10-15 nm, which appeared to possess flat morphologies consisting of individual fibres (Figure 4.18). The fibre-like structures observed in *ACR-12* appeared to be thicker in morphology compared with the sheet-like structures formed by the peptide control (**Ac-HYFNIFGSG-COOH**), with the respective height being ~0.5 nm (Figure 4.15). Given these differences, it was possible that the structures observed in *ACR-12* were not only attributed to the self-assembly of peptide, suggesting that the DNA component was also involved. Moreover, the fibre-like structures observed in *ACR-12* were absent in the unreacted DNA and peptide control, potentially implying that the conjugation of peptide with DNA was necessary for these structures to form.

However, these findings need to be interpreted with caution given that the self-assembly of the peptide control and unreacted DNA plus peptide control were investigated using water as a solvent; whereas in *ACR-12* TAMg buffer was used, which might have partially influenced the structural differences observed.



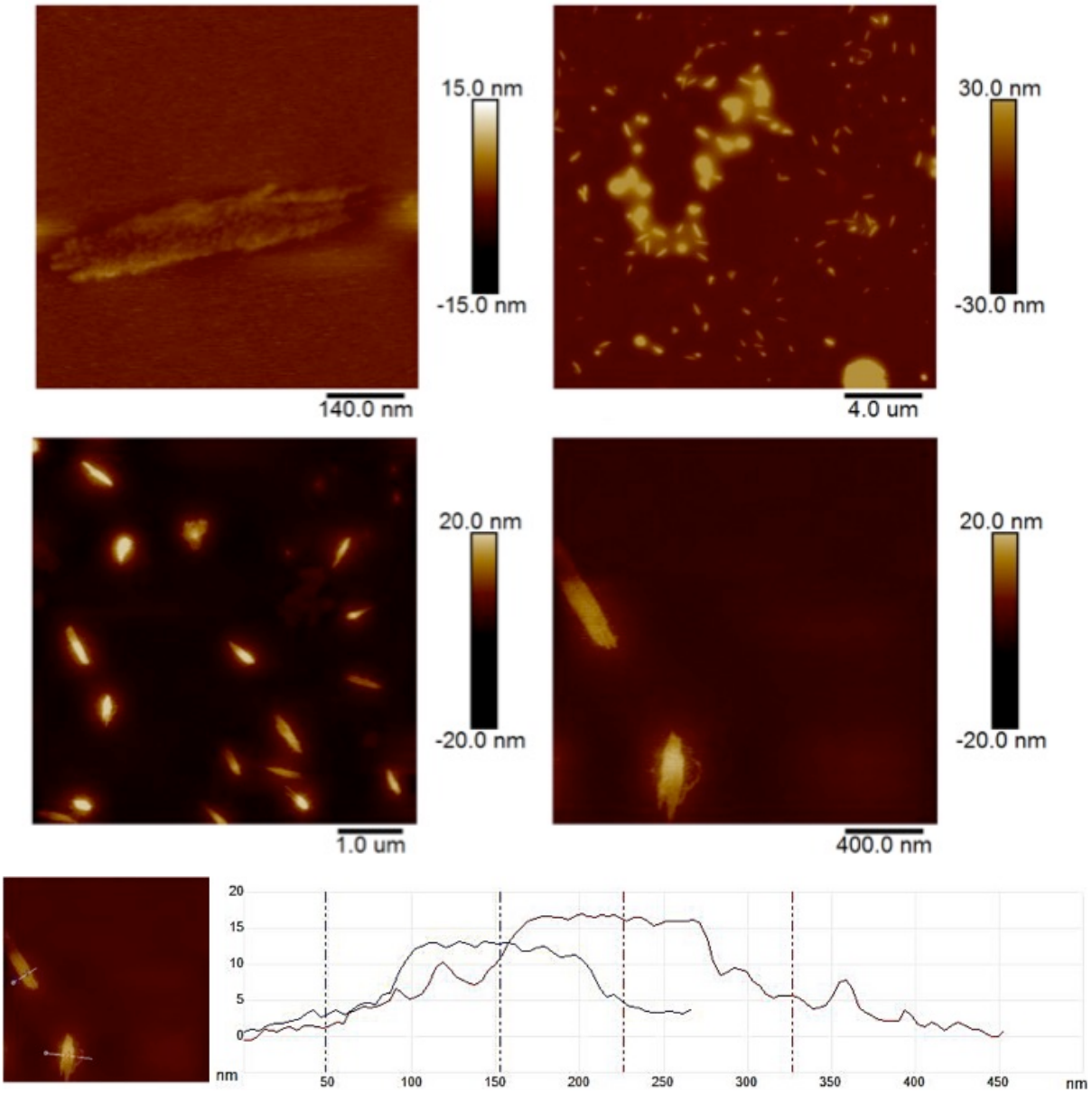


Figure 4.18. AFM images of *ACR-12* mixed with **DNA1-NH_2** at 7 days and corresponding height profile.

AFM Analysis of *ACR-14*

In ACR-14, **Ac-HYFNIFGSG-COOH** was reacted with either **DNA1-NH_2** or **DNA1'-NH_2**. For each reaction, fraction 3 was used for AFM analysis following freeze-drying and incubation in TAMg buffer at R.T. for 6 days. The AFM images of both samples showed spherical structures with a height of 20-40 nm (Figure 4.19 and 4.20), with no apparent morphological differences between the structures being observed. Given the similar structures that were seen between both samples, it appeared that the DNA sequence (**DNA1-NH_2** or **DNA1'-NH_2**)

had little to no influence on the resultant morphology. The large variation in the size of the spherical structures possibly indicated random aggregate formation.

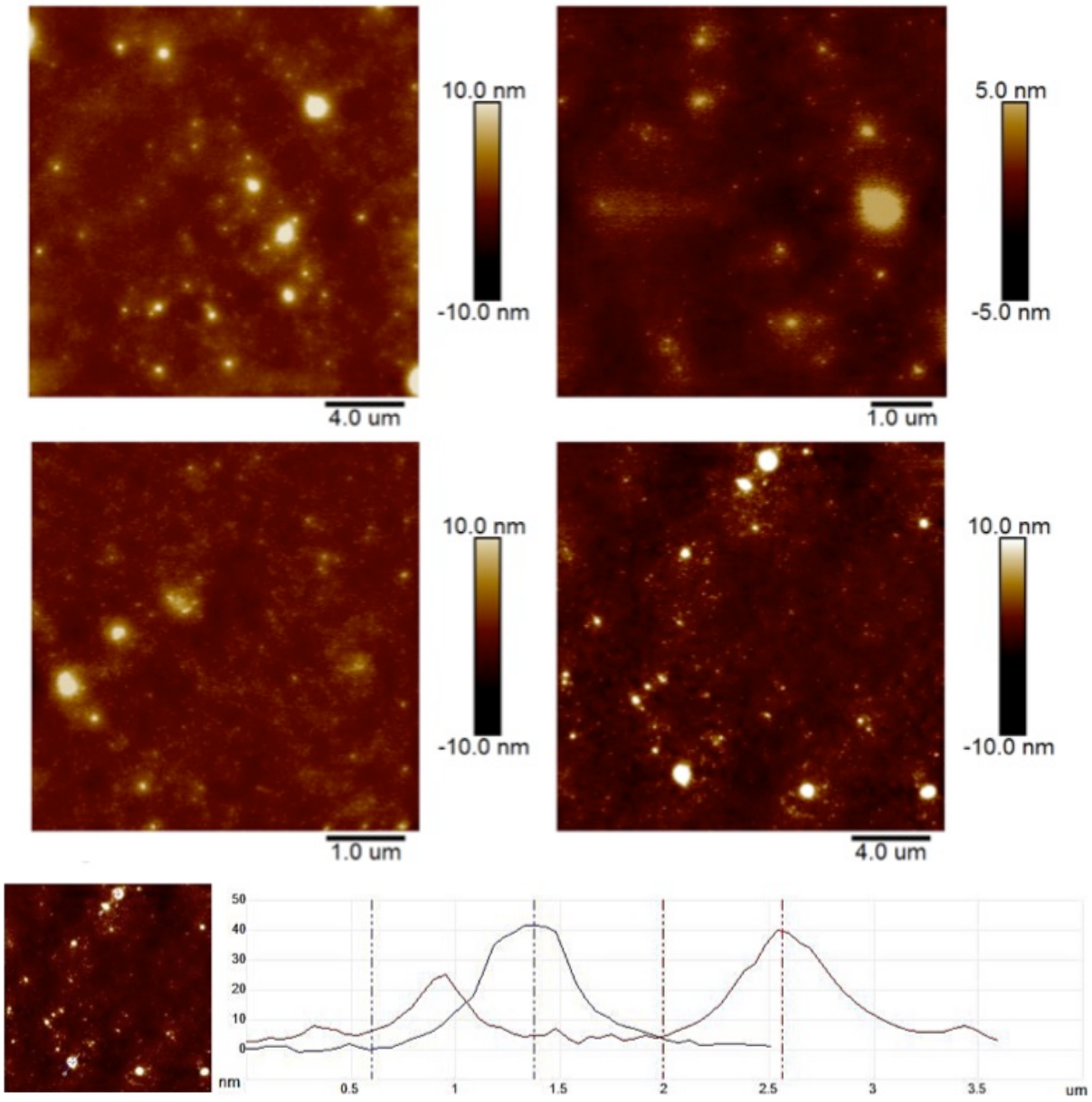


Figure 4.19. AFM images of *ACR-14* (**Ac-HYFNIFGSG-COOH** + **DNA1-NH_2**) at 6 days and corresponding height profile.

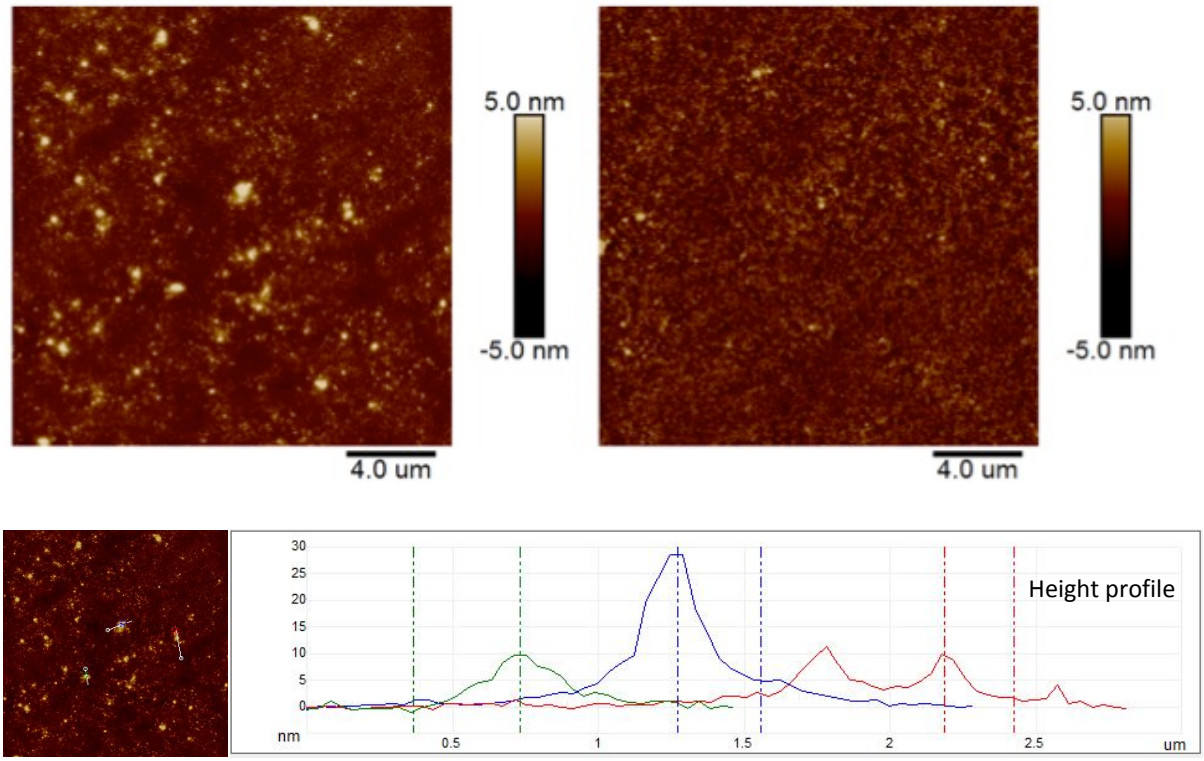


Figure 4.20. AFM images of *ACR-14* (**Ac-HYFNIFGSG-COOH** + **DNA1'-NH_2**) at 6 days.

4.4 Summary

Amide coupling was conducted as an alternative strategy to thiol-malemeide conjugation, but the reactions did not provide conclusive evidence that the conjugate was obtained. The PAGE studies only showed minimal band shifts, suggesting that the desired conjugate was not formed. Although the Kaiser tests were negative, suggesting that **DNA-NH_2** may have reacted, the PAGE results did not support this finding. This contradictory outcome indicated that the Kaiser tests were false negatives. While the potential cause of the false negative result could not be identified, it may have been attributed to DNA degradation following TFA treatment.

AFM showed the presence of unique structures in the reaction mixture samples that were not present in the different analysed control samples (DNA control; peptide control; unreacted DNA plus peptide control), suggesting the potential formation of a new compound, but it was not possible to confirm if this was a conjugate.

Overall, given the lack of evidence that the desired conjugate was obtained using amide

coupling, this strategy was discontinued. Thereafter, thiol-maleimide chemistry was resumed in an effort to obtain the desired conjugates.

4.5 References

1. C. Portela, J. L., Mascareñas, F. Albericio, S. Mazzini, C. Caminal, R. Ramos, S.M. Ocampo and R. Eritja, *Nucleosides Nucleotides Nucleic Acids*, 2007, **26**, 963–967.
2. N. Gour, D. Kedracki, I. Safir, K. X. Ngo and C. Vebert-Nardin, *Chem. Commun.*, 2012, **48**, 5440–5442.

5 Analysis of Peptide-Oligonucleotide Conjugate Self-assembly

5.1 Chapter Overview

This chapter explores the results of the self-assembling studies of POCs (**DNA1-CGSGHYFNIF**, **DNA1'-CGSGHYFNIF**, **DNA1-CGSGKLVFFA** and **DNA1'-CGSGKLVFFA**) that were purified by HIC following thiol-maleimide conjugation (see Chapter 3). The first part of the chapter focuses on conjugate-to-conjugate and conjugate-to-DNA hybridisation studies, highlighting the key structural features observed by AFM, as well as structural differences compared with the original peptide sequence. The latter part of the chapter discusses the AFM results of conjugates following annealing experiments, featuring the formation of unique structures, as well as outlining structural differences compared with the original peptide sequence.

5.2 Methods

Hybridisation

DNA1'-CGSGHYFNIF was hybridised with **DNA1-CGSGHYFNIF** or with two complementary DNA strands which present either one or two binding tracts for **DNA1'** (**DNA1** or **$(DNA1)_2$**). To prepare the samples for conjugate-to-conjugate hybridisation, 15 µL of 10 µM **DNA1'-CGSGHYFNIF** (HIC purified, F13) and 15 µL of 10 µM **DNA1-CGSGHYFNIF** (HIC purified, F13) were mixed in a PCR tube. To prepare the samples for conjugate-to-DNA hybridisation, **DNA1'-CGSGHYFNIF** was mixed with **DNA1** or the double-length complementary DNA strand **$(DNA1)_2$**. Samples were prepared in a PCR tube by mixing 5 µL of 14 µM **DNA1'-CGSGHYFNIF** (HIC purified, F14) with **DNA1** in a 1:1 ratio, or with **$(DNA1)_2$** in a 2:1 ratio. All samples were subsequently incubated in a thermocycler at 55 °C for 10 min, 35 °C for 15 min, 25°C for 15 min and 4 °C for 20 min.

Annealing

DNA1-CGSGKLVFFA (HIC purified F15, 2 µM), **DNA1'-CGSGKLVFFA** (HIC purified F15, 5 µM) and **DNA1'-CGSGHYFNIF** (HIC purified F15, 12 µM) were used for annealing. 5 µL of each sample was incubated in a thermocycler from 95 °C to 4 °C at a cooling rate of 1 °C per

minute.

12% Native PAGE

12% native PAGE was carried out to assess whether hybridisation had occurred based on band shifts compared with the control samples analysed (**DNA1**, **DNA1-CGSGHYFNIF**, **DNA1'-CGSGHYFNIF** and **$(DNA1)_2$**). Samples were mixed with glycerol in a 1:1 ratio (v/v) in a PCR tube prior to native PAGE. The gel was run in 1× TAMg buffer at a constant current of 15 mA and at a set voltage of 300 V for 90 min. Gels were stained using Stains-All prepared in 1× TAMg buffer for a minimum of 20 min. DNA ladder (GeneRuler Ultra Low Range DNA Ladder, ready-to-use, purchased from Thermo Scientific) was used to obtain a relative measure of oligonucleotide length.

AFM

AFM measurements were conducted for hybridised or annealed conjugate samples using the same methods described in Chapter 2. To prepare samples, solution (5 µL for annealed samples or 10 µL for hybridised samples and for peptide comparators) was deposited onto freshly cleaved mica for 20 min. The excess liquid was wicked off with the edge of a filter paper. Subsequently, the mica surface was washed with 50 µL of MilliQ water four times, which was then dried under a stream of N_2. To make comparisons with the hybridised samples, **CGSGHYFNIF** was analysed at a concentration of 50 µM and 5 µM in water. To make comparisons with the annealed samples, **CGSGKLVFFA** was analysed at a concentration of 50 µM in water. In addition, 0.2 M Na_2SO_4 solution was analysed by AFM to determine whether the conjugate samples had any remaining salt residues from HIC purification. The samples analysed by AFM are listed in Table 5.1.

Table 5.1. Samples analysed by AFM.

Samples	Conditions	Figure No.
CGSGHYFNIF (peptide comparator)	Prepared in water at a final conc. of 50 and 5 µM at 0 h	5.2
Hybridised **DNA1'-CGSGHYFNIF** and **$(DNA1)_2$**	Prepared in 0.2 M Na_2SO_4, at a final conc. of 7 µM at 0 h	5.3
Hybridised **DNA1'-CGSGHYFNIF** and **DNA1**	Prepared in 0.2 M Na_2SO_4, at a final conc. of 7 µM at 0 h	5.4
Hybridised **DNA1-CGSGHYFNIF** and **DNA1'-CGSGHYFNIF**	Prepared in 0.2 M Na_2SO_4, at a final conc. of 5 µM at 0 h	5.5 5.6
Na_2SO_4 (salt control)	Prepared at 0.2 M	5.7
CGSGKLVVFA (peptide comparator)	Prepared in water at a final conc. of 50 µM at 0 h	5.8
Annealed **DNA1-CGSGKLVFFA**	Prepared in water at a final conc. of 2 µM. Annealed (95→4°C, 1 h)	5.9
Annealed **DNA1'-CGSGKLVFFA**	Prepared in water at a final conc. of 5 µM. Annealed (95→4°C, 1 h)	5.10
Annealed **DNA1'-CGSGHYFNIF**	Prepared in water at a final conc. of 12 µM. Annealed (95→4°C, 1 h)	5.11

5.3 Results and Discussion

5.3.1 Confirmation of Conjugate Hybridisation

Native PAGE confirmed that **DNA1'-CGSGHYFNIF** successfully hybridised with **DNA1** given that the band position was higher compared with its respective DNA and conjugate controls (**DNA1** and **DNA1-CGSGHYFNIF**). Likewise, PAGE confirmed hybridisation of **DNA1'-CGSGHYFNIF** with **$(DNA1)_2$** given that the band appeared higher compared with its respective DNA and conjugate controls (Figure 5.1 a). The hybridisation of **DNA1-CGSGHYFNIF** with **DNA1'-CGSGHYFNIF** was also confirmed by native PAGE, showing a band above than the individual conjugate control samples (Figure 5.1 b). In addition, the DNA ladder showed that the different samples on the gels corresponded to their predicted molecular weight.

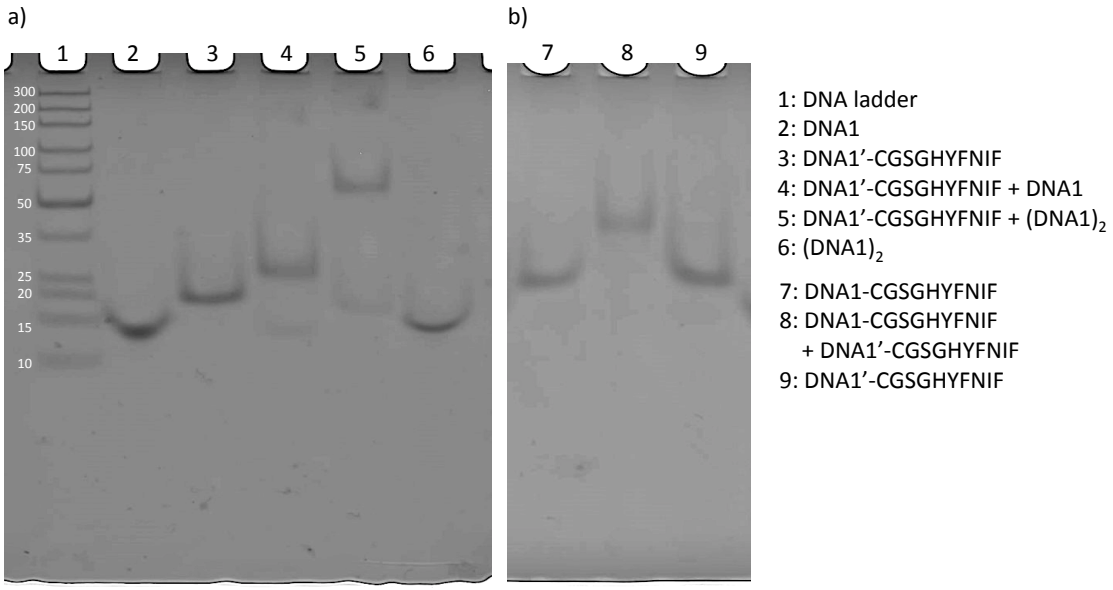


Figure 5.1. a) 12% native PAGE following hybridisation of **DNA1'-CGSGHYFNIF** with **DNA1** or with **$(DNA1)_2$**. The conjugate (**DNA1'-CGSGHYFNIF**) and complementary DNA strands (**DNA1** and **$(DNA1)_2$**) were run as controls. The DNA ladder is notated by number of DNA bases. **b)** 12% native PAGE following hybridisation of **DNA1-CGSGHYFNIF** with **DNA1'-CGSGHYFNIF**, along with both conjugate controls (**DNA1-CGSGHYFNIF** and **DNA1'-CGSGHYFNIF**).

5.3.2 Structural Characteristics of CGSGHYFNIF

CGSGHYFNIF was analysed by AFM to act as a comparator against the hybridised conjugates, given that they contained this peptide sequence. The AFM analysis of **CGSGHYFNIF** showed

fibril-like structures at 50 µM. Likewise, the formation of fibril-like structures was also confirmed when **CGSGHYFNIF** was prepared at a lower concentration of 5 µM. The self-assembled structures observed for **CGSGHYFNIF** were in line with those observed for **HYFNIF** (see Chapter 2), indicating that the self-assembly behaviour of this peptide was not altered by the presence of the spacer sequence **CGSG**.

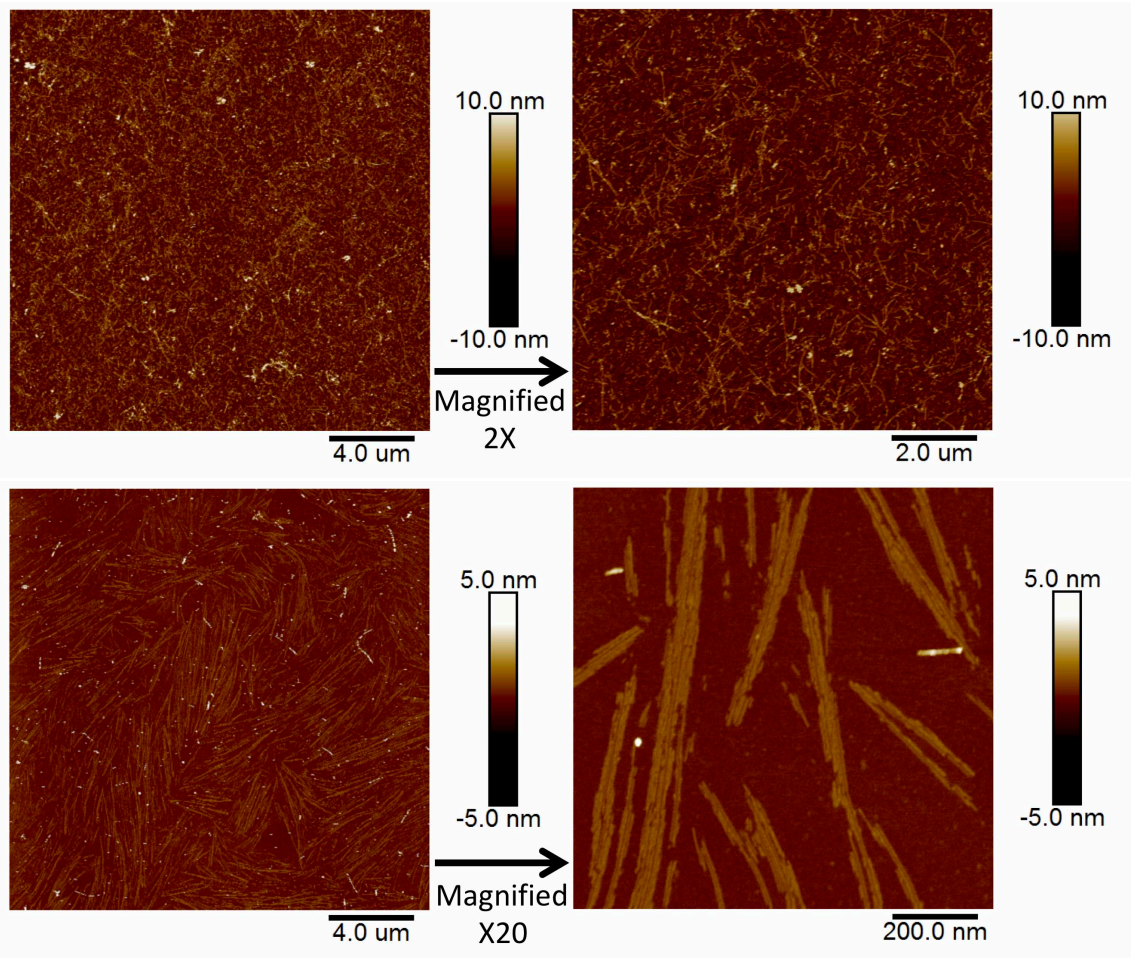


Figure 5.2. AFM images of **CGSGHYFNIF** at a concentration of 50 µM (above) and 5 µM (below) at 0 h.

5.3.3 AFM Analysis of Hybridised Conjugate with DNA

The AFM analysis of **DNA1'-CGSGHYFNIF** hybridised with **$(DNA1)_2$** at 7 µM showed small dot-like structures of 2-4 nm in height as well as fibril-like structures of 3-4 nm in height (Figure 5.3).

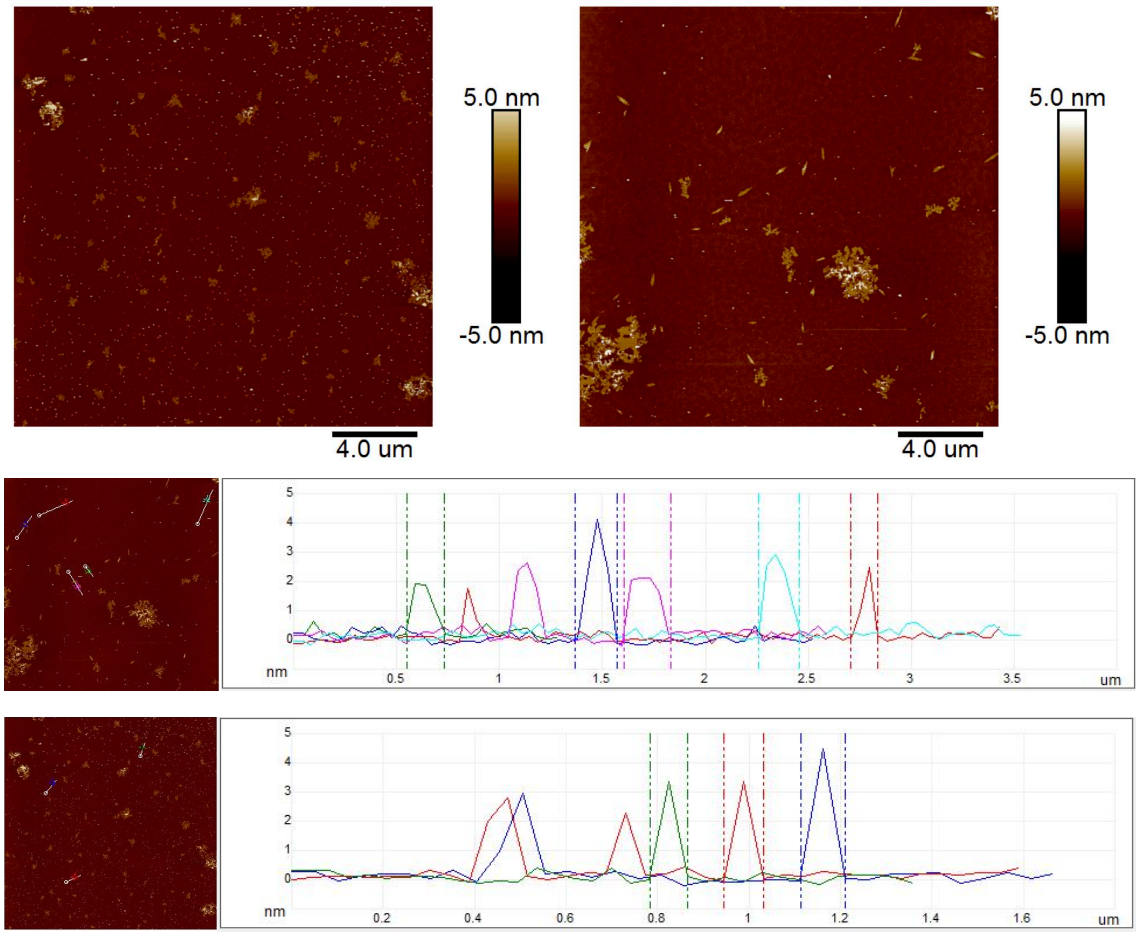


Figure 5.3. AFM images of **DNA1'-CGSGHYFNIF** hybridised with **$(DNA1)_2$** at a final concentration of 7 µM in 0.2 M Na_2SO_4 solution and corresponding height profiles.

The AFM images of **DNA1'-CGSGHYFNIF** hybridised with **$(DNA1)_2$** showed smaller fibril-like structures (300-500 nm in length) compared with the fibrils (0.5-4 µm in length) observed for **CGSGHYFNIF** (Figure 5.2). In addition, hybridisation of the conjugate with **$(DNA1)_2$** would have increased the hydrophilicity of the molecular system compared with the unhybridised conjugate, possibly resulting in chemical incompatibility with the hydrophobic peptide backbone. Consequently, this may have prevented cooperative self-assembly between each component, thereby suppressing fibril formation.[1,2]

In addition, hybridisation of the conjugate with **$(DNA1)_2$** inhibited the formation of hydrogen bonding between DNA molecules among the conjugates, as these bonds instead formed between the DNA of the conjugates and **$(DNA1)_2$**. The reduced number of hydrogen bonding across the DNA component of the conjugates and the other alternative interactions (peptide hydrophobic interactions, π-π stacking of phenylalanine residues, hydrogen bonding between the peptide backbone, and electrostatic interaction between the lysine residue and the phosphate groups of DNA) may not have been sufficient to drive further self-assembly at the low concentration used.

In contrast, no fibril-like structures were observed in the AFM analysis of **DNA1'-CGSGHYFNIF** hybridised with **DNA1** (Figure 5.4). However, the possibility that fibril-like structures were formed cannot be excluded given that the observation for this sample came from a single AFM image. A small number of undefined sheet-like structures were observed when **DNA1'-CGSGHYFNIF** was hybridised with **$(DNA1)_2$** or with **DNA1**, but these were likely to be artefacts given that these structures were observed across several AFM images of different samples.

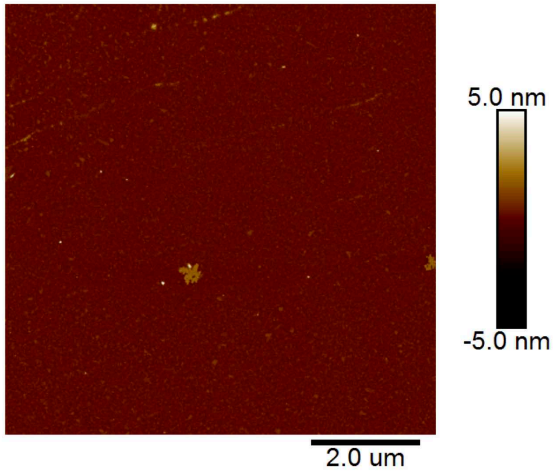
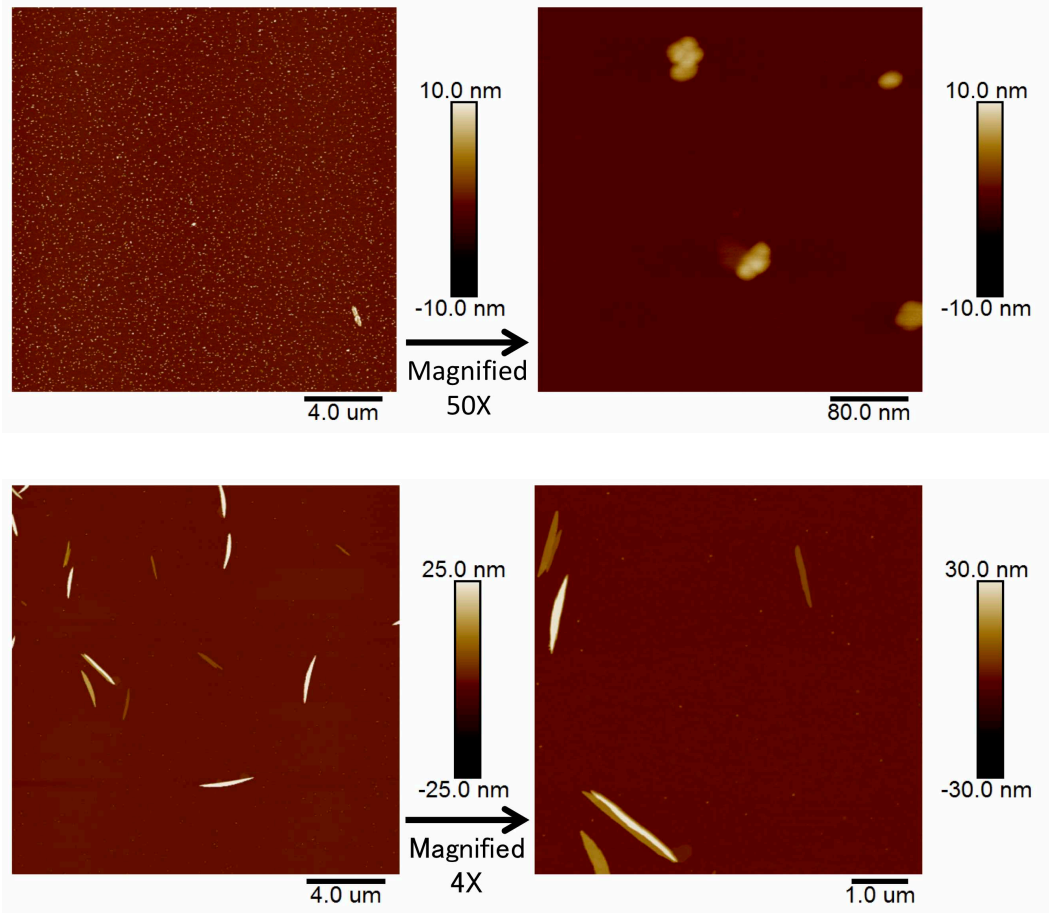


Figure 5.4. AFM image of **DNA1'-CGSGHYFNIF** hybridised with **DNA1** at a final concentration of 7 µM in 0.2 M Na_2SO_4 solution.

5.3.4 AFM Analysis of Hybridised Conjugate with Conjugate

The AFM analysis of **DNA1-CGSGHYFNIF** hybridised with **DNA1'-CGSGHYFNIF** showed two different types of structures coexisting together. These included small dot-like structures with a diameter of 20-40 nm and a height of 5-8 nm and rod-like structures with a width of ~250 nm and a height of 5-40 nm (Figure 5.5). While the shape of the dot-like structures was variable in morphology, the rod-like structures were well defined and appeared to have a flat morphology (Figure 5.6).



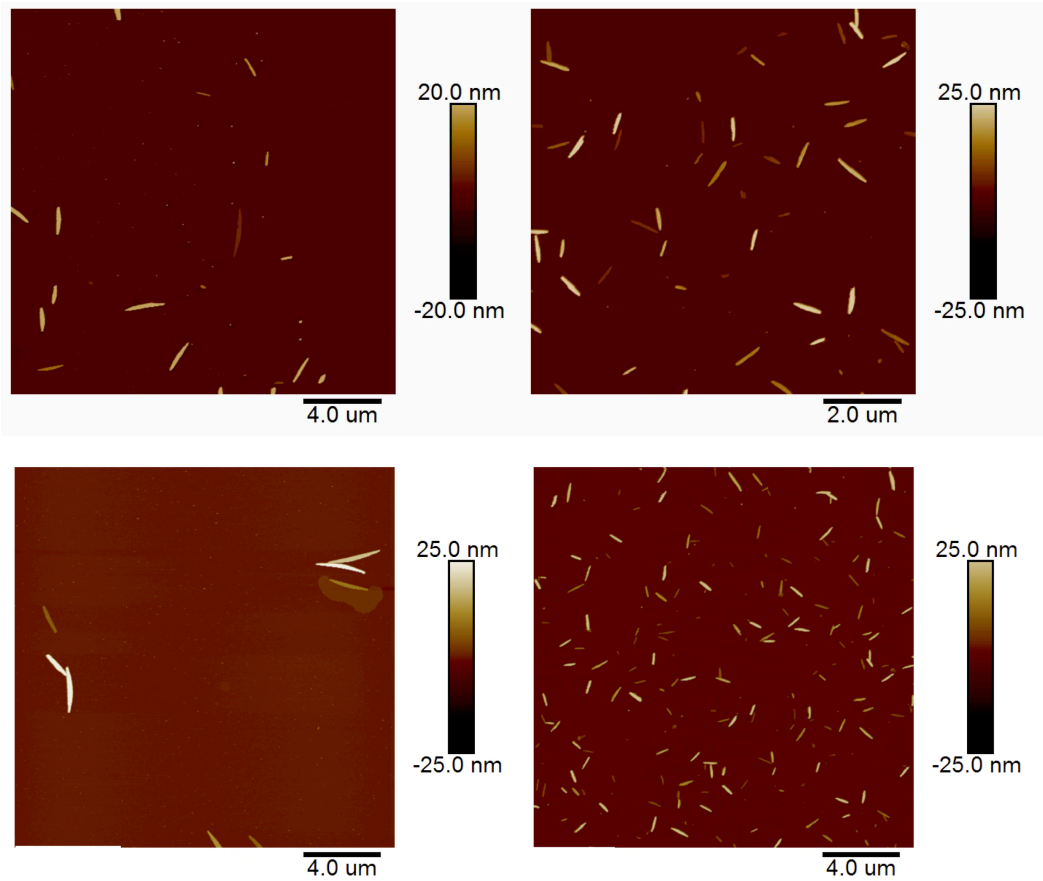


Figure 5.5. AFM images of **DNA1-CGSGHYFNIF** hybridised with **DNA1'-CGSGHYFNIF** at a concentration of 5 µM in 0.2 M Na_2SO_4 solution. The images were taken at different points on mica.

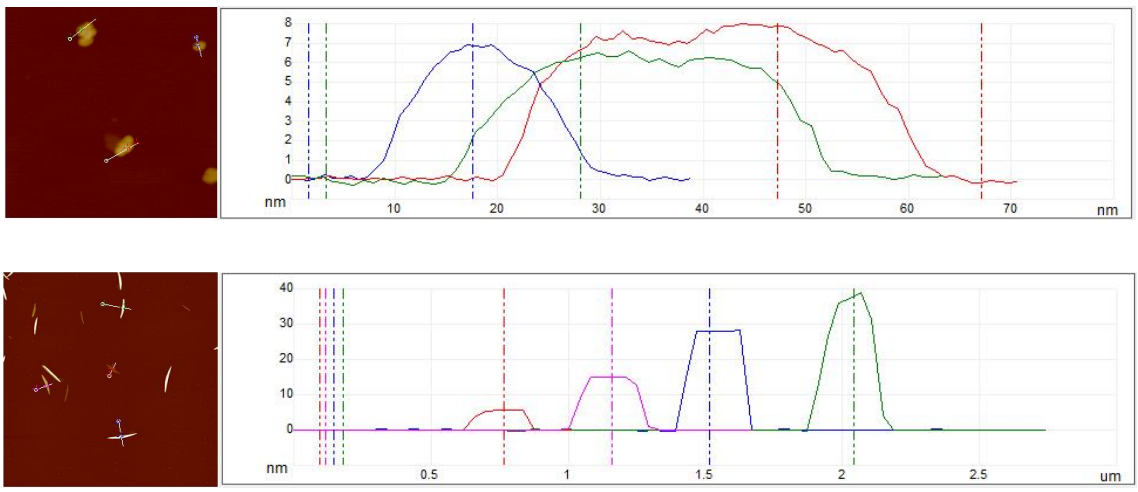
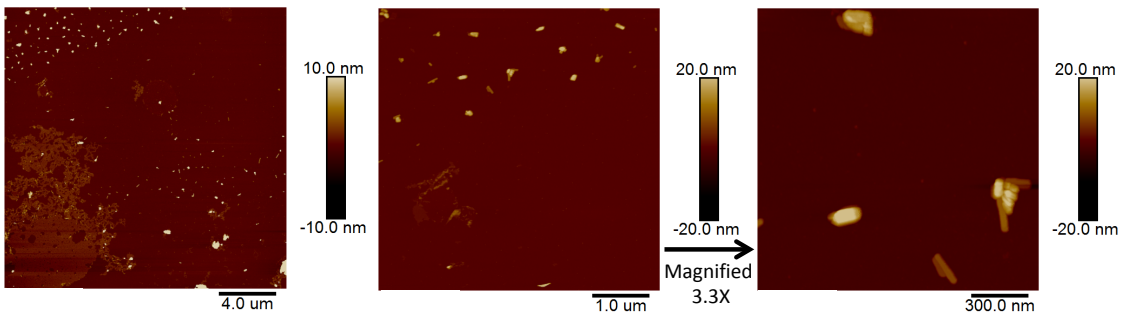


Figure 5.6. AFM height profiles of **DNA1-CGSGHYFNIF** hybridised with **DNA1'-CGSGHYFNIF** and corresponding height profiles.

In addition, hybridised **DNA1'-CGSGHYFNIF** with **DNA1-CGSGHYFNIF** did not show any fibril-like structures unlike **DNA1'-CGSGHYFNIF** hybridised with **$(DNA1)_2$**, which was surprising given that the peptide component (**CGSGHYFNIF**) showed a high propensity of fibril formation. It is possible that the additional peptide component (two moieties of **CGSGHYFNIF**) in the hybridised conjugate-to-conjugate resulted in the structural mismatch between DNA and peptide. In turn, this may have prevented the self-assembly of these biomolecules in a cooperative manner, disturbing fibril formation mediated by the peptide self-assembly.[3]

The flat morphology of rod-like structures may indicate a regular alignment of the hybridised conjugate molecules, in which DNA duplexes are tightly packed.[4] Furthermore, some of the rod-like structures appeared to be curved, which may be caused by repulsive forces that were generated upon further self-assembly into larger structures (i.e. coulombic repulsion between negatively charged phosphate backbones) and/or the structural mismatch between DNA and peptide.[4]

However, there is a possibility that the rod-like structures were not part of the hybridised sample and instead were formed from the residual salt (Na_2SO_4) that was used during HIC purification, given that their morphology was flat in appearance (Figure 5.7). A number of aggregates were observed even after washing of the mica surface with water (50 µL × 2) following deposition of Na_2SO_4 solution. This suggested that Na_2SO_4 may have been observed as a residual salt in the conjugate samples.



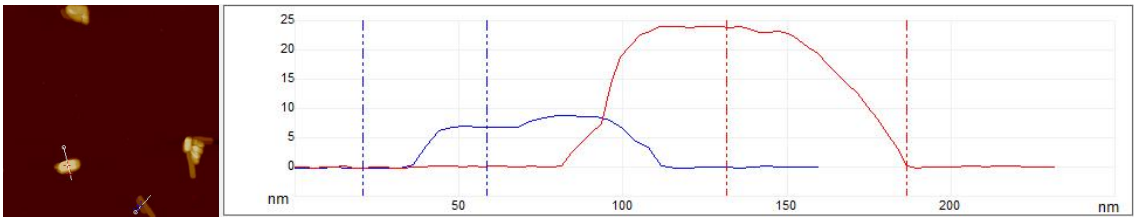
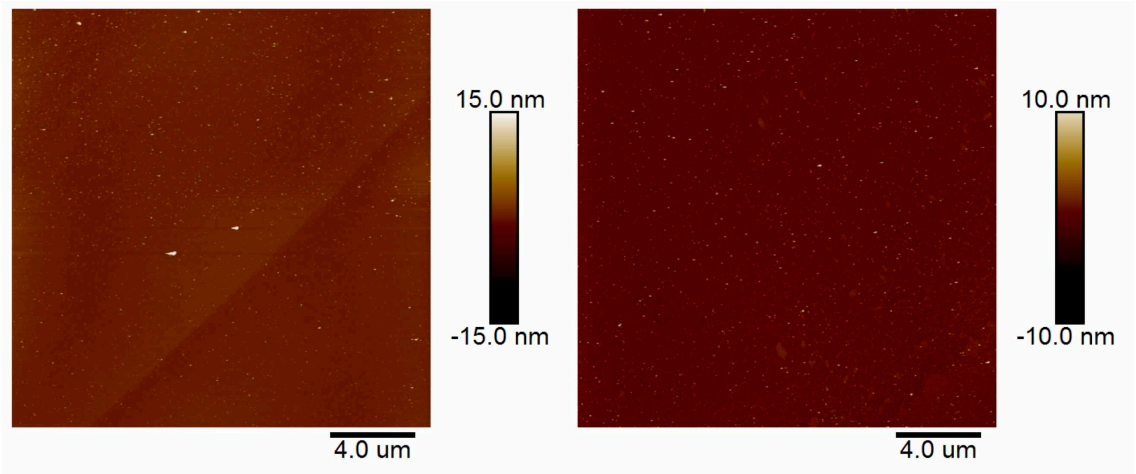


Figure 5.7. AFM images of 0.2 M Na_2SO_4 (salt control) and corresponding height profile, showing a number of aggregates with a flat morphology (5-25 nm in height).

The dot-like structures observed for the hybridised conjugate-to-conjugate appeared to be similar in size and height to those observed for annealed **CGSGHYFNIF** (Figure 5.11). This may suggest that the dot-like structures were attributed to the self-assembly of unhybridised conjugate. However, given that native PAGE showed a single band of the hybridised conjugate, only a small amount of unhybridised conjugate was present in the sample.

5.3.5 Structural Characteristics of CGSGKLVFFA

CGSGKLVFFA was analysed by AFM to act as a comparator against the annealed conjugates, given that they contained this peptide sequence. The AFM analysis of **CGSGKLVFFA** (peptide comparator) showed dot-like structures (Figure 5.8). The self-assembled structures observed for **CGSGKLVFFA** were in line with those observed for **KLVFFA** (see Chapter 2), indicating that the self-assembly behaviour of this peptide was not altered by the presence of the spacer sequence **CGSG**.



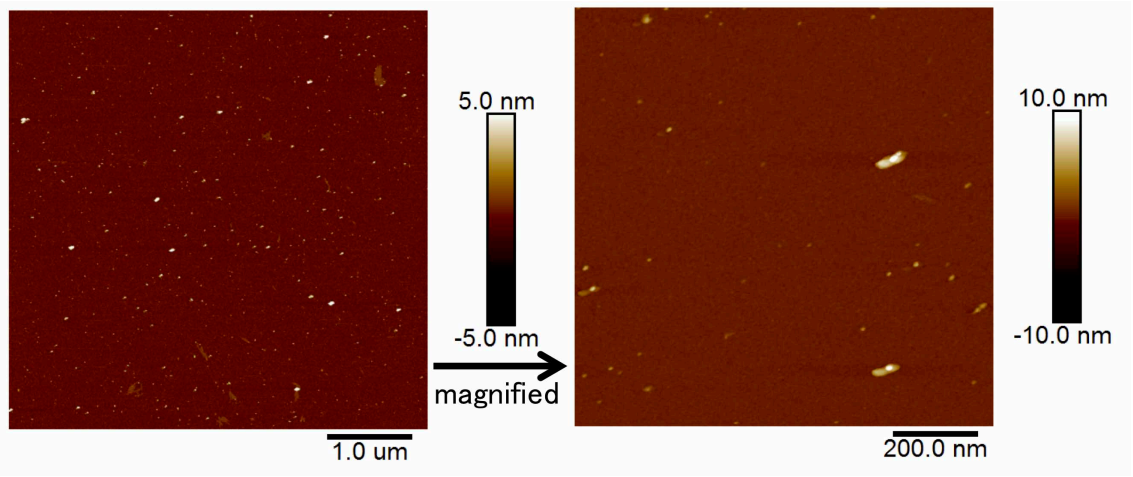
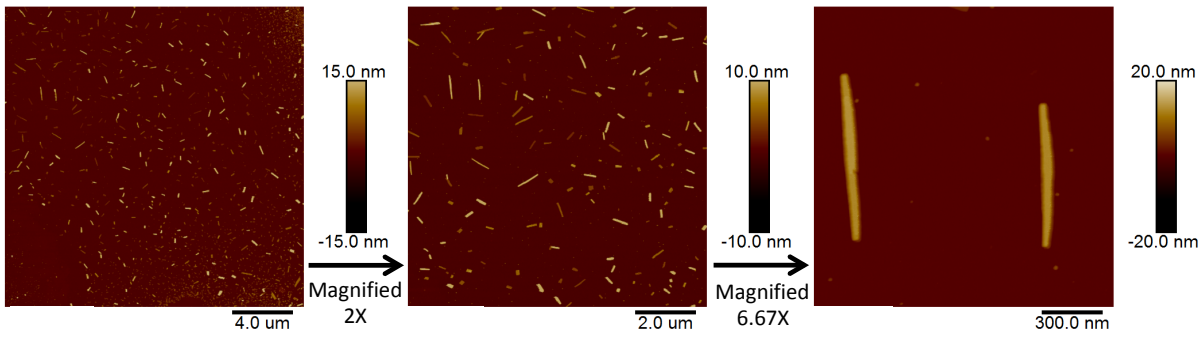


Figure 5.8. AFM images of **CGSGKLVFFA** at a concentration of 50 µM at 0 h.

5.3.6 AFM Analysis of Annealed Conjugates

The AFM analysis of annealed **DNA1-CGSGKLVFFA** and annealed **DNA1'-CGSGKLVFFA** showed rod-like structures with a width of ~100 nm a height of 5-20 nm, which appeared to be flat in morphology (Figure 5.9 and Figure 5.10). Given that both annealed conjugates showed homogeneous structures, it appeared that the choice of DNA (**DNA1** or **DNA1'**) had little to no impact in the resultant structures. The flat surface observed by AFM may be indicative that the structures were generated as a result of highly-ordered packing of the conjugate molecules, in which hydrophilic DNA and hydrophobic peptide blocks were segregated.[5] Given that the length of the conjugate used was approximately 10 nm, the rod-like structures (10-20 nm in height) observed could have consisted of one to two layers of conjugates stacked together.



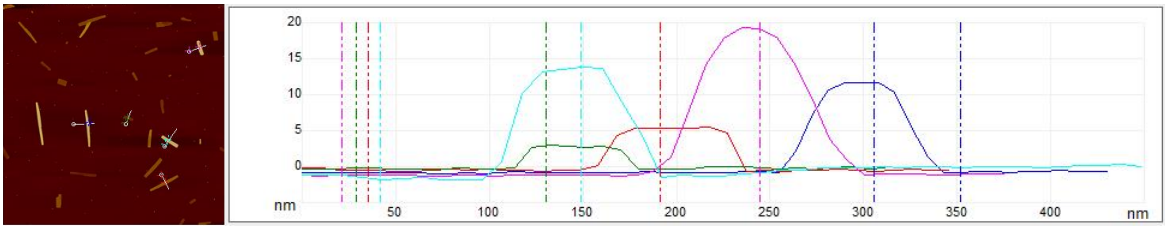


Figure 5.9. AFM images of annealed **DNA1-CGSGKLVFFA** at a concentration of 2 µM in water and corresponding height profile.

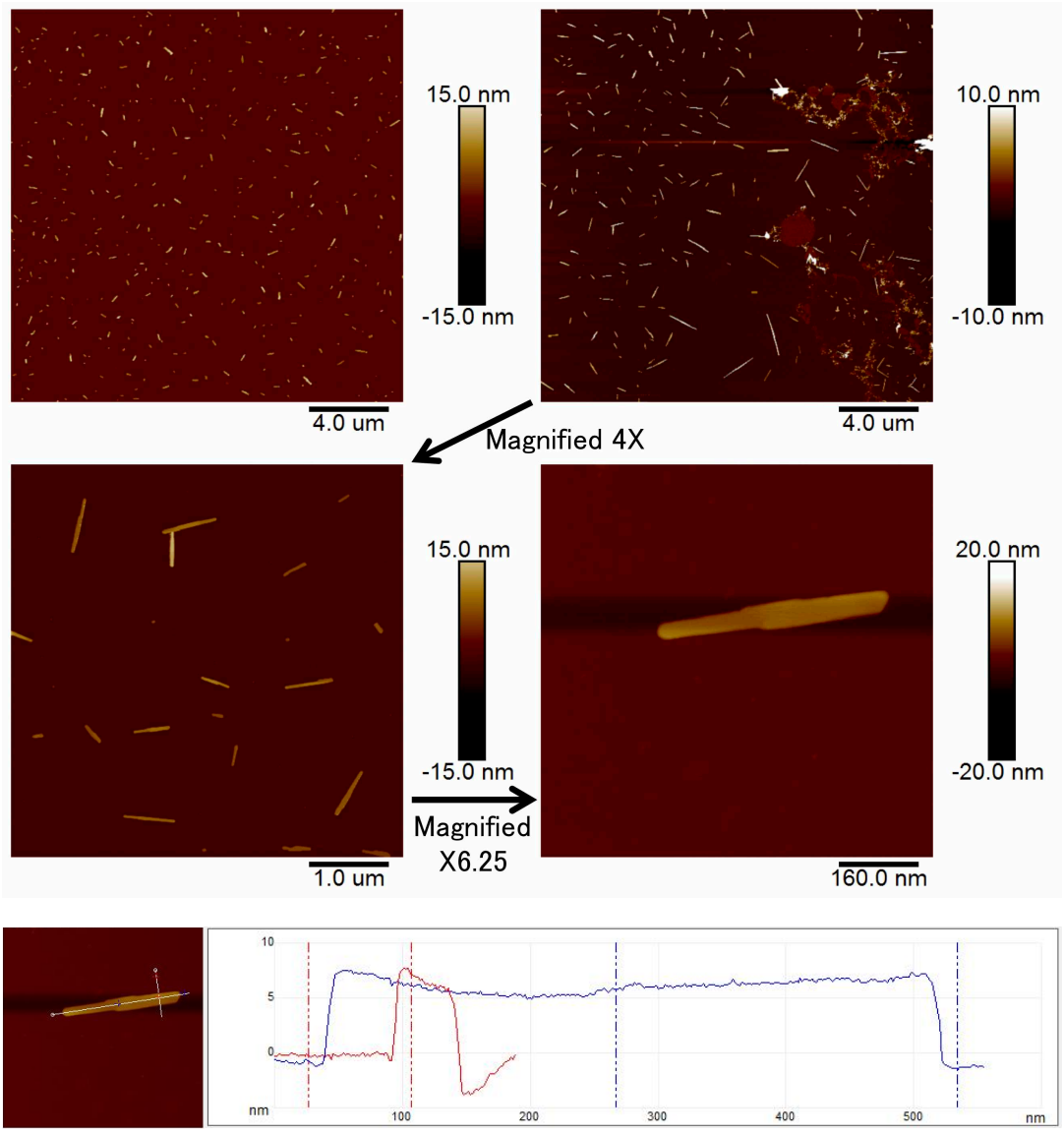
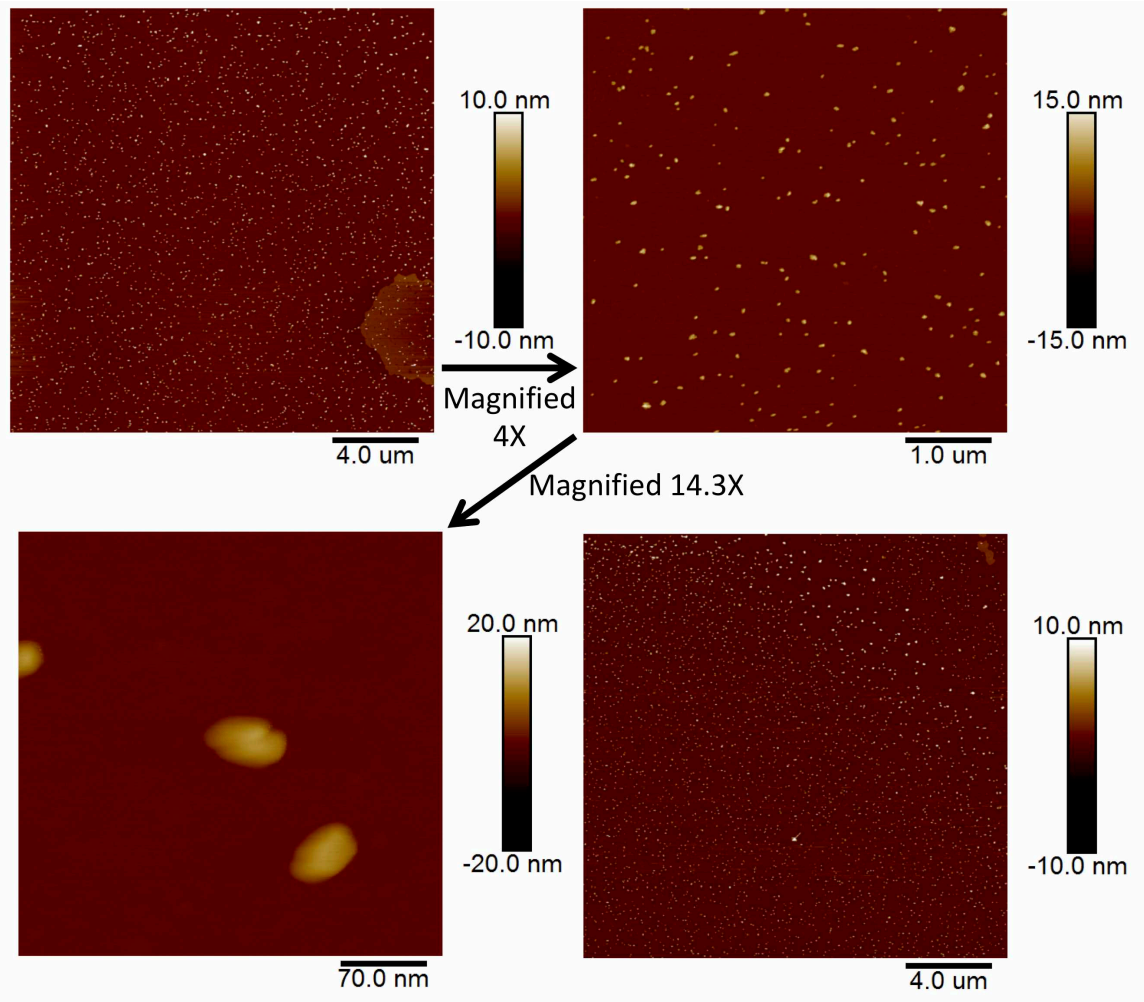


Figure 5.10. AFM images of annealed **DNA1'-CGSGKLVFFA** at a concentration of 5 µM in water and corresponding height profile.

However, given that the concentration of conjugates used was very low (2 µM for **DNA1-CGSGKLVFFA** and 5 µM for **DNA1'-CGSGKLVFFA**) and that structures with a flat morphology were observed, these structures may have been derived from the residual salt (Na_2SO_4), rather than from the annealed samples. In addition, given that the AFM analysis Na_2SO_4 control showed structures with a similar flat morphology (Figure 5.7) to those seen for the annealed conjugates, this hypothesis may be possible.

In contrast, the AFM analysis of annealed **DNA1'-CGSGHYFNIF** showed dot-like structures with a height of ~10 nm (Figure 5.11), while no rod-like structures were observed. Given that **DNA1'-CGSGHYFNIF** is amphiphilic and that the rod-like structures were approximately twice the predicted length of the conjugate, it may have been suggested potential vesicle formation occurred. However, to fully confirm this, further analysis in solution is required.



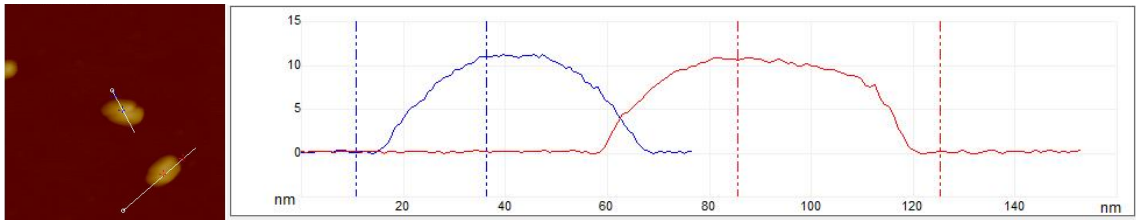


Figure 5.11. AFM images of annealed **DNA1'-CGSGHYFNIF** at a concentration of 12 µM in water and corresponding height profile.

Overall, annealing experiments for both **DNA-CGSGKLVFFA** and **DNA-CGSGHYFNIF** resulted in the formation of structures with uniform morphology, even at low concentrations. This may suggest that annealing facilitated the self-assembly of conjugates, yielding thermodynamic products. These findings are consistent with a previous study by the Freeman group which also reported that annealing accelerated the self-assembling process of peptide (diphenylalanine, FF)-DNA conjugates prepared at 50 µM, yielding twisted fibres.[6] Although further investigation is required, annealing of **DNA-CGSGKLVFFA** and **DNA-CGSGHYFNIF** at a higher concentration may result in the formation of hierarchical structures (e.g. twisted/bundled fibres) given that the peptide component of these conjugates have fibre-forming characteristics.

5.4 Summary

The hybridisation of **DNA1'-CGSGHYFNIF** with **DNA1**, **$(DNA1)_2$** or **DNA1-CGSGHYFNIF** was confirmed by native PAGE. Hybridisation of **DNA1'-CGSGHYFNIF** with complementary DNA strands resulted in reduced fibril formation compared with the unconjugated peptide. The presence of an additional DNA moiety may have caused the hybridised conjugate to become more hydrophilic, interfering with the self-assembly of the peptide and its subsequent fibril formation.

Hybridisation of **DNA1'-CGSGHYFNIF** with **DNA1-CGSGHYFNIF** resulted in the formation of small dot-like structures and rod-like structures with a flat morphology, whereas fibril formation was not observed. The absence of fibril-like structures in this hybridised sample may have been due to the structural mismatch between DNA and peptide, inhibiting these

biomolecules from self-assembling in a cooperative manner. The flat morphology may indicate a regular alignment of the hybridised conjugate molecules, in which DNA duplexes are tightly packed.

Annealed **DNA1-CGSGKLVFFA** and annealed **DNA1'-CGSGKLVFFA** showed rod-like structures with a flat morphology, which were not observed in the unconjugated peptide comparator. The flat surface may be indicative of layer formation generated from highly ordered arrangement of conjugate molecules. Meanwhile, annealed **DNA1'-CGSGHYFNIF** showed dot-like structures, which may suggest potential formation of vesicles based on the amphiphilic nature of the conjugate and the height of the structure.

While DNA does not self-assemble into large structures on its own, it demonstrated the capacity to self-assemble into unique structures following conjugation with β-sheet forming peptides (**CGSGHYFNIF** and **CGSGKLVFFA**) and subsequent hybridisation or annealing. Furthermore, DNA appeared to exert an inhibitory role on peptide self-assembly, reducing or even preventing the formation of fibrils in the analysed conjugates. Overall, this highlights that peptide-oligonucleotide conjugation leads to the formation of unique structures that would not be obtainable through the self-assembly of the individual biomolecules.

5.5. References

1 M. Kye and Y. Lim, *Angew. Chem. Int. Ed.*, 2016, **55**, 12003–12007.

2 N. Gour, J. N. Abraham, M. Chami, A. Castillo, S. Vermac and C. Vebert-Nardin, *Chem.Commu*n., 2014, **50**, 6863–6865.

3 I. W. Hamley, *Biomacromolecules*, 2014, **15**, 1543–1559.

4 S. Rothenbühler, I. Iacovache, S. M. Langenegger, B. Zuber and Robert Häner, *Nanoscale*, 2020, **12**, 21118–21123.

5 K. Mukai, M. Hara, S. Nagano and T. Seki, *Angew. Chem. Int. Ed.*, 2016, **55**, 14028–14032.

6 M. L. Daly, Y. Gao, and R. Freeman, *Bioconjugate Chem.*, 2019, **30**, 1864–1869.

6 Conclusion

The first objective of this project was met, with different β-sheet forming peptides (**KLVFFA**, **HYFNIF** and **RVFNIM**) and collagen mimetic peptides including **$(POG)_3$** and **$(POG)_6$** being successfully synthesised, as well as yielding multiple modified versions of the main peptide sequences. Overall, the CD profiles of **KLVFFA**, **HYFNIF** and **RVFNIM** revealed their expected β-sheet formation, with the AFM results showing fibre formation. While the CD profiles of **$(POG)_3$** and **$(POG)_6$** revealed their expected collagen triple helix formation, the AFM results did not confirm fibril or fibre formation. A spacer sequence (**CGSG**) was successfully attached to all peptides prior to conducting conjugation with oligonucleotides.

Chosen peptides (**CGSGKLVFFA**, **CGSGHYFNIF** and **CGSG$(POG)_6$**) were successfully conjugated with DNA oligonucleotides (**DNA1-mal** or **DNA1'-mal**) using thiol-maleimide chemistry. However, full conversion from **DNA-mal** to conjugate was not achieved due to the presence of some ring-opened forms of the DNA maleimide groups prior to reacting with peptides. Instead, the desired conjugates (**DNA-CGSGHYFNIF** and **DNA-CGSGKLVFFA**) were successfully isolated from the reaction mixture using HIC purification. Amide coupling was tested as an alternative strategy, but successful conjugation could not be confirmed with this method.

Hybridisation of conjugates with DNA led to reduced fibril formation compared with the original peptide. Whereas, conjugate-to-conjugate hybridisation led to the formation of large rod-like structures that were absent in the original peptide. In contrast, the characteristic fibril structures of the original peptide vanished in the hybridised conjugates. Lastly, the annealing experiments showed that conjugates formed rod-like structures or dot-like structures that were not seen in the unconjugated peptides.

Overall, thiol-maleimide was shown to be an effective strategy for conjugating peptides with oligonucleotides, with hybridisation and annealing experiments demonstrating that POCs acquired novel structural characteristics that were absent in the original unconjugated biomolecules.