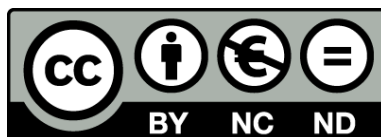




UNIVERSITAT DE
BARCELONA

Novel methodologies for the conjugation and cyclisation of polyamides

Omar Brun Cubero



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

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

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



UNIVERSITAT_{DE}
BARCELONA

Novel methodologies for the conjugation and cyclisation of polyamides

Omar Brun Cubero

Novel methodologies for the conjugation and cyclisation of polyamides

Manuscript submitted by:

Omar Brun Cubero

Directed and supervised by:

Prof. Anna Grandas Segarra

Universitat de Barcelona
Departament de Química Inorgànica i Orgànica
Secció de Química Orgànica

Departament de Química Inorgànica i Orgànica
Secció de Química Orgànica
Programa de Doctorat en Química Orgànica
Universitat de Barcelona, Juny 2016

Agraïments

M'agradaria començar aquesta tesi agraint-vos, Anna i Enrique l'oportunitat d'entrar al grup i de poder gaudir d'una beca per fer el doctorat. A tu, Anna, a més del teu tracte sempre proper, t'haig d'agrair que sempre hagis estat sempre accessible i disposada a rebre'm al despatx tant bon punt jo picava la porta, que no han estat pocs cops, i per tota la feina que has fet de seguiment i correcció d'aquest treball. Vull fer també extensiu el meu agraïment als altres membres del grup, Vicente, Núria i Jordi, per estar sempre disposats a ajudar. Also thanks to you, Röger, for letting me join your group for a few months. Irene i Laura de masses, i Francisco de RMN, moltes gràcies per la vostra amabilitat i ajuda. A tu Josep, que estàs en aquella nevera anomenada sala d'aparells, gràcies per la teva ajuda amb els HPLC/MS i pel teu bon humor.

A tota la gent que ha passat pel grup, Albert, Àlex, Javi, Ana, Natàlia, Jordi i molts d'altres, moltes gràcies per haver creat un bon ambient de treball i, de ben segur, haver-me ensenyat moltes coses (o recordar-me quants minuts dura el test de ninhidrina). Gràcies especials tant a l'Eli, amb qui vaig fer els primers passos en la química, molt abans de començar el doctorat, com al Xavi, que també ha estat un mentor per mi (i una font inesgotable de preguntes estúpides amb resposta estúpida). Merci à toi Clément pour l'aide avec les oligonucléotides et pour m'apprendre beaucoup de cochonneries, thanks to you Lewis for your hard work and dedication to the cyclics project, for the free English lessons and for your friendly character, a ti Cuqui por ser tan adorable (aunque discutamos por la música) y estar siempre ahí para echar una mano, i a tu Tomeu per la tranquil·litat que inspires i la teva ajuda. Fora del grup, però sempre per la uni, vull agrair-te a tu, Pau, els anys d'amistat, les cerveses, i haver estat amb mi en els moments difícils.

Fora de la uni m'agradaria agrair-vos a tots els i les de MacVoyage & The Real Family els bons moments, els riures i l'amistat. Gràcies també a aquells que heu estat amb mi des de fa també molt temps: Diego, Bertran, Dici i Edu.

Finalment, potser el més fort dels agraïments per vosaltres, Mama i Papa, que sempre m'heu fet costat, que m'heu posat totes les facilitats per arribar a poder acabar aquesta tesi, per la vostra estima incondicional i, en general, per tot el que heu fet i seguiu fent per mi.

Index

About the structure of this work

CHAPTER 1 Introduction and objectives

1.1 Biomolecules: biological role and use as drugs	3
1.2 Modification of biomolecules	4
1.2.1 Backbone modification.....	4
1.2.1.1 Oligonucleotide backbone modifications.....	6
1.2.1.2 Peptide backbone modifications	7
1.2.2 Conjugation	8
1.2.2.1 Synthesis of conjugates	8
1.2.2.2 Conjugates: desirable properties	11
1.2.3 Cyclisation	12
1.3 Survey of reactions employed for conjugation and cyclisation.....	13
1.3.1 Copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC).....	15
1.3.2 Strain-Promoted Azide-Alkyne Cycloaddition (SPAAC)	18
1.3.3 Staudinger ligation	19
1.3.4 Traceless Staudinger ligation.....	21
1.3.5 Thiol Michael-type reaction.....	22
1.3.6 Disulfide ligation	25
1.3.7 Native Chemical Ligation (NCL).....	27
1.3.8 Thiazolidine ligation	29
1.3.9 Oxime and hydrazone ligation	30
1.3.10 Diels-Alder (DA) ligation	31
1.3.11 Inverse Electron-demand Diels-Alder (IEDA)	33
1.4 Objectives	34
1.5 Abbreviations	35
1.6 Bibliography	35

CHAPTER 2

On-resin Diels-Alder reactions in water. Conjugates of diene-derivatised polyamides and diene-derivatised oligonucleotides

2.1 Conjugation of diene-derivatised polyamides	47
2.1.1 Background and objectives.....	47
2.1.2 Decomposition of dienes under the cleavage and deprotection conditions ...	48
2.1.3 Described alternatives to circumvent diene decomposition	50
2.1.4 Our proposal to solve the problem	52
2.1.5 The effect of water in Diels-Alder reactions.....	53
2.1.6 Selection of the solid support for the on-resin Diels-Alder reaction in water ..	55
2.1.7 Analysis of the reaction outcome	56
2.1.8 Selection of the diene	56
2.1.9 Assessment of the solvent influence on the on-resin Diels-Alder conjugation ...	59
2.1.10 Optimizing the reaction conditions	60
2.1.11 Broadening the scope of the reaction: conjugation of different polyamides	61
2.1.12 Use of protected maleimides to obtain conjugates with different linking sites....	64
2.1.13 Conjugation with molecules sparingly soluble or insoluble in water.....	65
2.1.14 Synthesis of double conjugates	69
2.2 Conjugation of diene-derivatised oligonucleotides.....	72
2.2.1 Background and objectives.....	72
2.2.2 Reaction parameters and analysis of the reaction outcome	74
2.2.3 Work with short ON models	75
2.2.4 Mixed ON sequences	79
2.2.5 MALDI-TOF MS monitoring of the conjugation reaction	80
2.2.6 Assessment of the fragmentation in the conjugation reaction conditions.....	85
2.3 Critical overview of the methodology developed.....	86
2.4 Abbreviations	87
2.5 Bibliography	88

CHAPTER 3

2,2-Disubstituted cyclopent-4-ene-1,3-diones: more than simple Michael acceptors

3.1 Background and objectives	93
3.2 Examining CPDs reactivity	97
3.2.1 Reaction of CPDs with peptides containing internal or C-terminal cysteines.....	
.....	98
3.2.2 Reaction of CPDs with peptides containing N-terminal cysteines.....	101
3.3 Structural determination of the M -20 Da adduct	105
3.4 Summary of the information available at this point	107
3.5 Screening of reaction conditions favouring the formation of the M -20 Da adduct	
.....	107
3.6 Attempts to gain insight into the oxidation process.....	109
3.7 Computational analyses.....	110
3.8 Reactions with other CPDs	111
3.9 Reactions between CPDs and other amino acids: cysteine and homocysteine ...	113
3.9.1 Reaction of CPDs with Cys.....	114
3.9.2 Reaction of CPDs with Hcy.....	118
3.10 Attachment of CPDs to peptide chains	119
3.11 Synthesis of bioconjugates exploiting the formation of an M -20 Da adduct and assessment of the stability of the generated products.....	120
3.12 Exploiting the selectivity of CPDs towards N-terminal cysteines.....	123
3.12.1 Selective derivatization of a peptide containing an N-terminal cysteine in a mixture of cysteine-containing peptides.....	123
3.12.2 Regioselective double labelling of a peptide containing N-terminal and internal cysteines	126
3.13 Estimation of the M -20 Da adduct molar extinction coefficient.....	128
3.14 Critical overview of the methodology	130
3.15 Abbreviations	131
3.16 Bibliography	131

CHAPTER 4

Cyclisation and derivatisation of peptides using 2,2-disubstituted cyclopent-4-ene-1,3-diones

4.1 First indication for the formation of a cyclic structure	137
4.2 First steps to set a procedure for the cyclisation reaction	139
4.3 Confirmation of the cyclic structure by NMR.....	145
4.4 Determination of the molar absorption coefficient.....	148
4.5 Search of alternatives to the use of TEMPO	154
4.6 Use of labelling CPDs in cyclisation reactions - 1.....	157
4.7 Effect of aromatic residues close to the <i>N</i> -terminal cysteine.....	159
4.7.1 Description of the experiments	159
4.7.2 Detection of a previously overlooked intermediate.....	162
4.7.3 Oxidation of the M -18 to M -20 Da adduct	163
4.7.4 Extent of by-product formation.....	164
4.7.5 Conclusion.....	165
4.8 Effect of lithium chloride on the reaction progress.....	165
4.9 Optimisation of the cyclisation procedure	167
4.10 Use of labelling CPDs in cyclisation reactions - 2. Epimerisation of the <i>N</i> -terminal cysteine	168
4.11 Confirmation of <i>N</i> -terminal cysteine epimerisation.....	171
4.12 Comparison of the stability of the D and L epimeric cyclic peptides	174
4.13 Use of labelling CPDs in cyclisation reactions - 3.....	175
4.14 Critical overview of the methodology developed.....	177
4.15 Abbreviations	177
4.16 Bibliography	178

Conclusions

Conclusions	181
-------------------	-----

Appendix 1

Attempts to identify a diene protecting group

A.1.1 Attempts performed	187
A.1.2 Abbreviations.....	194
A.1.3 Bibliography.....	195

Appendix 2

Experiments performed by Lewis Archibald

Experiments performed by Lewis Archibald	199
--	-----

Experimental part

Materials and Methods

E.M.M.1 Reagents and solvents	203
E.M.M.2 Chromatographic techniques	204
E.M.M.3 Spectroscopic techniques	204
E.M.M.4 Oligomer synthesis	206
E.M.M.5 Other instruments used.....	209
E.M.M.6 Abbreviations.....	209
E.M.M.7 Bibliography.....	210

Experimental part

Chapter 2

E.2 General methods	211
E.2 Analysis of the polyamide- and oligonucleotide-resins before diene introduction	213
E.2.1 Conjugation of diene-derivatised polyamides.....	214
E.2.1.2 Decomposition of dienes under the cleavage and deprotection conditions	214
E.2.1.8 Selection of the diene.....	215
E.2.1.9 Assessment of the solvent influence on the on-resin Diels-Alder conjugation	217

E.2.1.10	Optimizing the reaction conditions.....	218
E.2.1.11	Broadening the scope of the reaction: conjugation of different polyamides	219
E.2.1.12	Use of protected maleimides to obtain conjugates with different linking sites	220
E.2.1.13	Conjugation with molecules sparingly soluble or insoluble in water	222
E.2.1.14	Synthesis of double conjugates	228
E.2.2	Conjugation of diene-derivatised oligonucleotides	229
E.2.2.2	Reaction parameters and analysis of the reaction outcome.....	229
E.2.2.3	Work with short ON models.....	231
E.2.2.4	Mixed ON sequences	233
E.2.2.5	MALDI-TOF MS monitoring of the conjugation reaction.....	235
E.2.2.6	Assessment of the fragmentation in the conjugation reaction conditions	237
E.2	Characterisation of the products	240
E.2	Abbreviations.....	252
E.2	Bibliography.....	253

Experimental part

Chapter 3

E.3	General methods.....	255
E.3.1	Background and objectives.....	256
E.3.2	Examining CPDs reactivity.....	257
E.3.2.1	Reaction of CPDs with peptides containing internal or C-terminal cysteines	257
E.3.2.2	Reaction of CPDs with peptides containing N-terminal cysteines	260
E.3.3	Structural determination of the M -20 Da adduct.....	263
E.3.5	Screening of reaction conditions favouring the formation of the M -20 Da adduct	267
E.3.8	Reactions with other CPDs	268
E.3.9	Reactions between CPDs and other amino acids: cysteine and homocysteine	271
E.3.9.1	Reaction of CPDs with Cys	271
E.3.9.2	Reaction of CPDs with Hcy	274
E.3.10	Attachment of CPDs to peptide chains.....	275

E.3.11 Synthesis of bioconjugates exploiting the formation of an M -20 Da adduct and assessment of the stability of the generated products.....	275
E.3.12 Exploiting the selectivity of CPDs towards <i>N</i> -terminal cysteines	277
E.3.12.1 Selective derivatization of a peptide containing an <i>N</i> -terminal cysteine in a mixture of cysteine-containing peptides.....	277
E.3.12.2 Regioselective double labelling of a peptide containing <i>N</i> -terminal and internal cysteines.....	279
E.3.13 Estimation of the M -20 Da adduct molar absorption coefficient.....	280
E.3 Product characterisation	282
E.3 Abbreviations.....	292
E.3 Bibliography.....	292

Experimental part

Chapter 4

E.4 General methods.....	293
E.4.1 First indication for the formation of a cyclic structure.....	293
E.4.2 First steps to set a procedure for the cyclisation reaction.....	294
E.4.3 Confirmation of the cyclic structure by NMR	299
E.4.4 Determination of the molar absorption coefficient	303
E.4.5 Search of alternatives to the use of TEMPO	305
E.4.6 Use of labelling CPDs in cyclisation reactions – 1.....	308
E.4.7 Effect of aromatic residues close to the <i>N</i> -terminal cysteine	310
E.4.8 Effect of lithium chloride on the reaction progress.....	311
E.4.9 Optimisation of the cyclisation procedure.....	312
E.4.10 Use of labelling CPDs in cyclisation reactions - 2. Epimerisation of the <i>N</i> -terminal cysteine	313
E.4.11 Confirmation of <i>N</i> -terminal cysteine epimerisation	317
E.4.12 Comparison of the stability of the D and L epimeric cyclic peptides	318
E.4.13 Use of labelling CPDs in cyclisation reactions - 3	318
E.4 Characterisation of the products.....	320

Resum en català

Resum en català	341
-----------------------	-----

About the structure of this work

This work is divided into four parts.

The first one contains a first chapter with the introduction and objectives of this work, and three chapters explaining the development of three new conjugation or cyclisation methodologies. Each of them contains its own reference collection and abbreviations list. Conclusions reached during this work are also included in this part.

The second part is the experimental part and contains all the information necessary to reproduce the experiments performed as well as the product characterisation. Sections have been made so that experiments performed in a given section of the first part can be found in the experimental section with the same name. For clarity, numbering of the experimental sections is the same as in the first part, including the prefix "E." to each section. Those sections in which no experiments have been made are not included in the experimental part. Those experimental sections that do not possess their theoretical counterpart have been numbered only with the chapter number and the prefix "E.". Although, to help the reader, retention times of some products are listed in the corresponding experimental section, complete product characterisation can be found in an independent section of the same chapter.

The third part corresponds to the appendixes. The first one explains the efforts made to find an appropriate protecting group for a diene, while the second is a list of the experiments performed by Lewis Archibald.

Finally, the fourth part is a summary of this dissertation in catalan.

Chapter 1. Introduction and objectives

1.1 Biomolecules: biological role and use as drugs

Biomolecules are those molecules in living organisms ranging from large, complex structures such as proteins weighing hundreds of kDa to low molecular weight metabolites. Every single biomolecule has a role in the living organism to which it belongs: DNA stores information, RNA mediates in the protein synthesis, proteins realise most of the functions in the organism, coenzymes are involved in the transference of electrons or functional groups, etc.

Not only they make life possible, but also many of them have been proved to be related with disease development. They can either play a key role, such as proteins in Alzheimer's, Huntington's and Parkinson's diseases,^{1,2} or just be related to higher risks of suffering illnesses, such as homocysteine with strokes.³ Discovering and understanding how biomolecules behave in our organism and all the mechanisms in which they are involved is, therefore, of utmost importance.

Biomolecules, especially proteins, peptides and oligonucleotides, can also be used as drugs for the treatment of various diseases. The FDA has approved around 180 proteins for their clinical use as drugs or diagnostic agents,⁴ and although not many oligonucleotides have been approved, some of them have reached phase III clinical trials.⁵ Peptides have also found their spot in the medicinal field: cyclosporines act as

immunosuppressants,⁶ somatostatin analogues such as octreotide have been used for the treatment of neuroendocrine tumours,⁷ polymyxins have successfully been used as antibiotics,⁸ etc.

Although biomolecules are intrinsically excellent drugs, they normally need to be modified to improve or modulate their pharmacokinetic and pharmacodynamic properties. For this purpose, three main strategies have been extensively employed: backbone modification, conjugation and cyclisation (which although it actually is a special type of backbone modification, due to its importance will be treated apart). The three strategies will be discussed for oligonucleotides, peptides and proteins, although backbone modification will be only briefly explained as it is not related to the main object of this work. All molecules resulting from these strategies will be treated as artificially generated molecules, but this does not mean that they compulsory need human intervention to exist: cyclic peptides⁹ and conjugates¹⁰ can be found in nature, for example.

1.2 Modification of biomolecules

1.2.1 Backbone modification

As mentioned above, backbone modification is one of the strategies employed to tune biomolecule properties. Modifications are performed at the backbone in order not to alter the recognition motifs usually found in the side-chains of amino acids or the nitrogenous bases of oligonucleotides. Backbone modifications in oligonucleotides and peptides have provided good results in improving the stability towards enzymatic degradation, in the pre-organization of complex structures and also in the activity.^{11,12}

Some of the most important modifications of peptide and oligonucleotides backbones are listed in **Table 1.1**. PNAs Will be the most extensively discussed modification, as they have been used in this work. More exhaustive discussions for the other modifications can be found in appropriate reviews.^{13,14} Cyclisation, which is another backbone modification usually employed, will be discussed in section **1.2.3**.

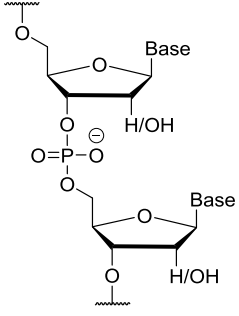
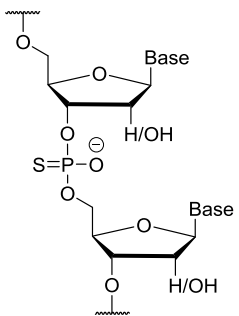
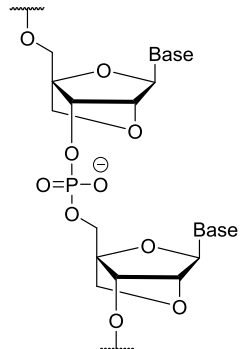
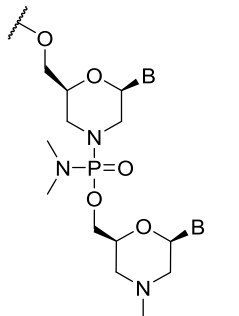
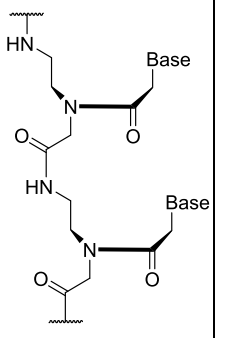
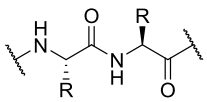
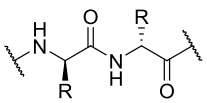
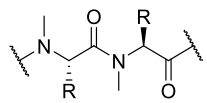
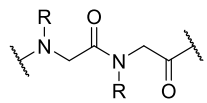
			
Natural oligonucleotide backbone			
			
Phosphorothioate Oligonucleotides (PSOs)	Locked Nucleic Acids (LNAs)	Phosphorodiamidate morpholino oligonucleotides (PMOs)	Peptide Nucleic Acids (PNAs)
			
Natural peptide backbone			
			
Peptides incorporating D- amino acids	<i>N</i> -methylated peptides	Peptoids	

Table 1.1. Main backbone modifications of oligonucleotides (top part) and peptides (bottom part).

1.2.1.1 Oligonucleotide backbone modifications

Phosphorothioate oligonucleotides (PSOs) replace one of the non-linking oxygen atoms in the phosphate diester with a sulfur atom, distributing the negative charge unsymmetrically and generating two diastereomers. This change results in an increased stability to nucleases and base-catalysed hydrolysis while the solubility in water is maintained.¹⁵ Furthermore, improved cell uptake when compared to the wild-type oligonucleotides has been demonstrated.¹⁶ From a synthetic point of view, a great step-forward on their synthesis occurred when Beaucage and co-workers implemented the use of 3*H*-1,2-benzodithiol-3-one 1,1-dioxide, which allowed for the use of standard phosphoramidites in the synthesis of PSOs.¹⁷

Locked Nucleic Acids (LNAs) were designed and synthesised by the Wengel group in the late 90s and have proved to be excellent modifications of natural oligonucleotides. The introduction of a 2'-O-4'-C-methylene bridge generates an outstanding resistance against phosphodiesterases without losing water solubility.¹⁸ Modified oligonucleotides incorporating LNA monomers exhibit thermal stabilities increased by *ca.* 3 °C and 6-8 °C when paired to complementary DNA and RNA sequences, respectively.¹⁵ Furthermore, LNA monomers can be incorporated to oligonucleotide chains by means of the standard phosphoramidite chemistry.¹⁹

Phosphorodiamidate morpholino oligonucleotides (PMOs) are one of the most important modifications performed on oligonucleotides, and some of them are currently in clinical trials.¹³ In these entities, the sugar ring is replaced by a six-membered morpholino ring, and the phosphodiester linkage is changed to a phosphorodiamidate. They have shown improved nuclease stability in comparison with their unmodified partners, and good water solubility, especially when compared to other morpholino modifications.²⁰

Peptide nucleic acids (PNAs) were first introduced by Nielsen and Berg in 1992, and their usefulness is today beyond question; more than 300 articles were published dealing with PNAs only in 2014. Based on a (2-aminoethyl)glycine skeleton, the nucleobases are appended from the glycine N atom with a methylenecarbonyl linker. This skeleton is homomorphous with the deoxyribose backbone found in nucleic acids and was computationally designed to be able to mimic a B-DNA duplex. Thanks to the lack of backbone negative charges, when paired to complementary chains PNAs show improved thermal stabilities and melting temperatures almost unaffected by ionic strength.²¹ From the degradation point of view, they have shown good stability in human serum and in eukaryotic cytoplasmic extract.²² Total resistance to isolated proteinases and peptidases at high enzyme concentrations was also found,²²

altogether accounting for improved stability properties in comparison with both natural oligonucleotides and peptides.

On the negative side, their lack of charges also involves low solubility in water, but this problem can easily be solved by attaching lysine residues to the PNA chain. Even worse, studies regarding cellular uptake showed low phospholipid membrane permeability,²³ making it almost impossible to distribute unmodified PNAs through cell membranes. Certain methods have been used for *in vitro* delivery of PNAs into cells (microinjection, use of cells with wall/membrane defects, electroporation), but their use is impossible in complex living systems.²⁴ To ensure cell permeability in said systems, conjugation (see section 1.2.2) of PNAs to other entities has proved to be an excellent method. Among the entities destined to be linked to PNAs one can find lipophilic moieties, peptides and cell-specific receptor ligands.²⁴ From a synthetic point of view, the synthesis of PNAs is accomplished as for peptides. It started with the use of the Boc/Cbz protecting groups scheme, but soon many groups began to propose different Fmoc-based protecting schemes: the Fmoc/Cbz by the Noble group,²⁵ the Fmoc/acyl by the Kovács group,²⁶ the Fmoc/Mmt by the Uhlmann group,²⁷ etc. Nowadays, the most common scheme is the one based in the Fmoc and Bhoc protecting groups,²⁸ and a proof of it is the amount of commercial sources for Fmoc/Bhoc protected PNA monomers. Another alternative that has received some attention is the Fmoc/bis-Boc protecting scheme²⁹, although PNA monomers derivatised in such way do not seem to be commercially available.

It is also worth mentioning that some groups have modified the PNA backbone, mainly with the objective of enhancing water solubility, hybridization properties and cellular uptake. For instance, the Ly group has synthesised and evaluated α -guanidine derivatised,³⁰ γ -guanidine derivatised³¹ and γ -diethylenglycol derivatised PNAs.³²

1.2.1.2 Peptide backbone modifications

Peptides incorporating D-amino acids do not present any changes in the connectivity of the different atoms, but just a change in the configuration of certain stereocentres. This change triggeres a huge improvement in the activity of certain peptides.³³ From a synthetic point of view, these entities are synthesised by means of normal solid-phase peptide synthesis (SPPS), with D-aminoacids being commercially available.

N-Methylation of amide bonds improved the pharmacological properties of several peptides and is an increasingly employed strategy to gain oral bioavailability.³⁴ The stability towards enzymes and membrane permeability are also greater than in the wild-type peptides.³⁵ As to their synthesis, they can be obtained either by employing N-methylated amino acids, which are commercially available, or by a methylation reaction

after the incorporation of every amino acid residue.³⁵ The acylation reaction is more problematic because the increased bulkiness of *N*-methylated *N*-terminal amino acids entails longer reaction times or the need for special activating agents, which might result in higher racemisation degrees.³⁵

Peptoids are characterised by the lack of a chiral centre in the polyamide chain, as functional groups are appended to the nitrogen atom instead of to the α -carbon. Resistance towards proteolytic enzymes³⁶ and membrane permeability³⁷ are enhanced compared to normal peptides.

Many synthesis approaches have been used for their synthesis, and some of them are briefly discussed in this review by Zuckermann.³⁸

1.2.2 Conjugation

1.2.2.1 Synthesis of conjugates

Conjugates or bioconjugates are the entities generated upon linking a biomolecule to another moiety to obtain a product that either combines the features of both species or in which the properties of one component are ameliorated.

As stated before, conjugation is one of the main techniques used to modulate biomolecule properties, therefore allowing some of those with initially poor pharmacokinetic or pharmacodynamic properties to be used as drugs.

Conjugation can be used to gain specificity and reduce side-effects, as for example when some already active drugs are linked to specific recognition motifs, such as antibodies. Since 2000, three different anti-cancer antibody-drug conjugates (ADCs) have been approved by the FDA for clinical use: Mylotarg, Adcetris and Kadcyła, although the first one had to be recently removed from the market due to toxicity reasons.³⁹ Despite the warning offered by Mylotarg, evidences that ADCs are extremely useful tools for the treatment of several diseases are conclusive, and different groups continue conducting active research on anti-cancer ADCs.⁴⁰

Conjugation can also be used to make biomolecules more resistant to degradation. It has been shown since long ago⁴¹ that conjugation of proteins to poly(ethylene glycol) (PEG) moieties enhances their stability and bioavailability, and several FDA-approved PEGylated proteins are in clinical use.⁴² Not only PEG has proved to be a valuable polymer for conjugation with proteins, but many other organic polymers are currently being investigated as promising candidates for properties modulation.⁴³

Cellular uptake can also be dramatically improved by means of conjugation. The best example may be the use of cell-penetrating peptides (CPPs). These entities have been extensively used as vectors to deliver a large variety of molecules into cells. Active

drugs such as taxol⁴⁴ or doxorubicin⁴⁵ have been linked to CPPs to improve cell permeability. Imaging systems have also experienced a boost since they have been used in combination with cell penetrating peptides.^{46,47} Oligonucleotides and analogs have also been linked to these entities to improve pharmacokinetic properties. As said in section 1.2, PNAs possess low membrane permeability, so many of them have been derivatised in this way to ensure cell uptake.^{48,49}

Conjugation can also serve as a platform for the synthesis of prodrugs as long as the linker between the active drug and the other moiety is designed to be cleavable. One of the most common strategies is the use of hydrolysable⁴⁴ or reducible linkers, such as oximes and disulfides. As an example, well-known anticancer agents have been linked to a variety of biomolecules to generate prodrugs.^{44,50}

Finally, the modification of biomolecules with reporter groups, such as fluorophores or radiolabels, can be used for imaging or localization purposes. The molecules to be imaged or localised can be labelled beforehand and then administrated to the living organism of study,⁵¹ or, on the contrary, they can be labelled *in situ*.⁵² This last strategy needs the use of biorthogonal reactions that do not produce toxicity.

The different strategies used for the synthesis of bioconjugates in which both moieties are oligomers can be split into two main categories: convergent synthesis or fragment conjugation and divergent or stepwise synthesis. The former consists in the union of the two conjugate moieties once they have been independently synthesised, while the latter implies the construction of the second moiety, step by step, directly onto the first. Fragment conjugation can be performed with both blocks in solution, the so-called Liquid-Phase Fragment Conjugation (LPFC), or with one of the blocks still anchored to the solid support, the Solid-Phase Fragment Conjugation (SPFC) (see **Figure 1.1**). As for the protection strategy, LPFC is nowadays performed with both blocks unprotected and using click-type reactions (see section 1.3),⁵³ but some examples exist in the early literature in which one or both oligomers are protected.⁵⁴ SPFC Can be performed linking both a protected or unprotected oligomer to the one anchored to the solid support.⁵⁵

On the other hand, the stepwise synthesis is only performed on a solid matrix. Protection of both the synthesised oligomer and the building blocks of the one in construction is always needed, although it has been observed that partial deprotection of oligonucleotides may be useful when synthesizing peptide-oligonucleotide conjugates.⁵⁶ Normally a linker must be introduced to allow the incorporation of the first monomer of the second oligomer, although sometimes one of the functional groups present, if unprotected, can be used as appendage site.

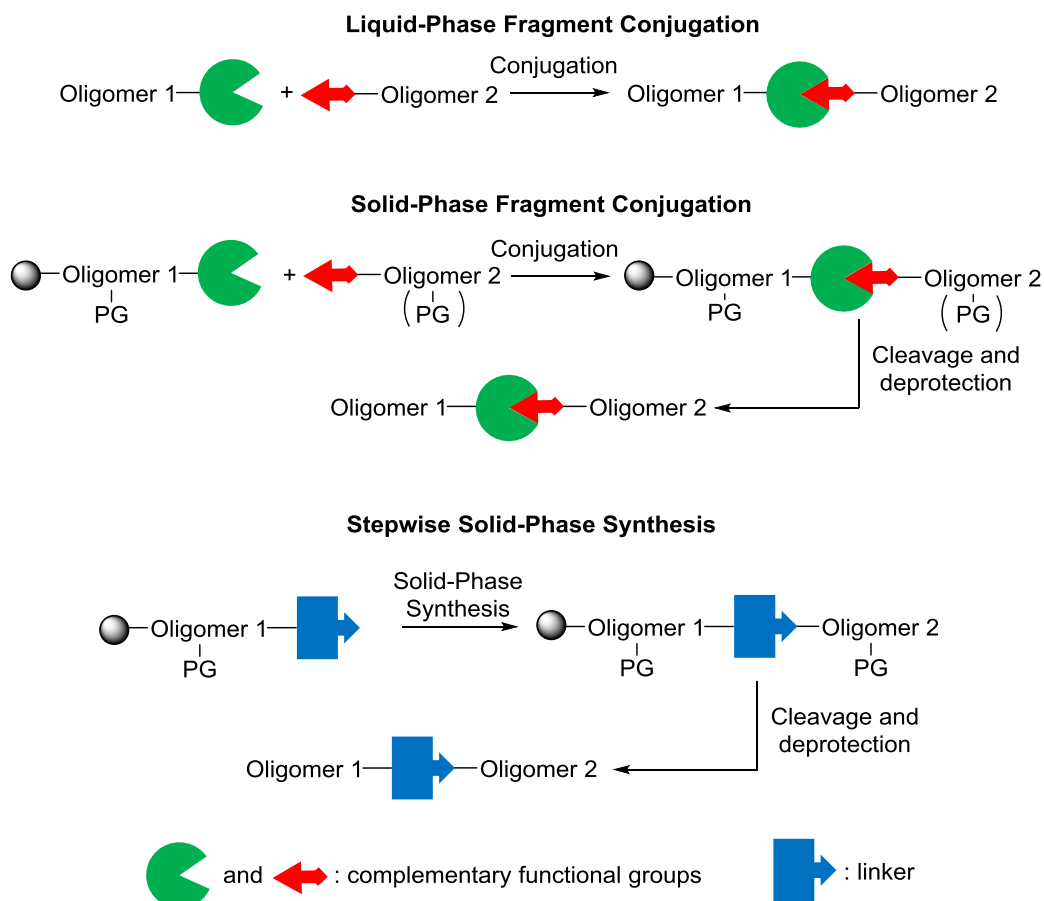


Figure 1.1. Different strategies for the synthesis of conjugates. PG = Protecting Group.

Analysis and purification of such large molecules is not easy, and separation may not be effective if the impurities have small differences in molecular weights. This may lead to impossible-to-purify crudes or erroneously considered pure products. Furthermore, most of the methods available for purification (HPLC, electrophoresis, dialysis...) tend to cause severe drops in the isolation yields, in contrast with classical methods such as recrystallisation or precipitation. Analysis methods are more restricted, since techniques such as NMR are typically precluded due to the low amounts of product normally obtained (nano- to few micromols) and expertise in the analysis of highly complex spectra is required. Chemo- and regioselectivity become major issues when dealing with molecules containing several reactive groups that can be repeated dozens of times. A classical approach to this problem would be the use of protecting groups, but this is not an option when working with molecules from natural sources.

1.2.2.2 Conjugates: desirable properties

In order to obtain the best-performing bioconjugates possible, several properties are desirable and some aspects must be controlled, especially if they are intended to be used as drugs.

The growing field of ADCs has provided very useful information about the parameters affecting the pharmacokinetic and therapeutic properties of conjugates, from both *in vivo* and *in vitro* experiments. Thus, it will be used as an example, but all comments and reflections apply to any other type of conjugate.

It has been shown that relatively unstable linkers such as hydrazones lead to the premature loss of payload, which might result in systemic toxicity and lower amounts of the active ADC distributed to the target tissues.⁵⁷ It is, consequently, of great interest to use chemical reactions that furnish stable enough linkages between the antibody and the drug. However, in some cases a cleavable linker can be desired or needed for the drug to accomplish its function. In this case, fragmentation of the ADC to release the active drug must happen only when the target has been reached, and therefore must be accomplished by means of specific mechanisms existing only in those areas.

Generation of different regioisomers should be avoided, as it has been shown that there is a clear impact of the linking position to the activity of the ADC. It is not only due to the fact that the immediate environment modulates the stability of certain linkers, like in the case of thiosuccinimides,⁵⁸ but because the linking point, *per se*, influences the physical and pharmacokinetic properties of ADCs.⁵⁹

It has also been noted that the drug-to-antibody ratio (DAR) affects the therapeutic index and pharmacokinetic properties of several ADCs,⁶⁰ thus highlighting the need to find methodologies to generate homogenous DARs in ADCs.

The medical, biological and chemical communities have grown a major concern towards regio- and stereospecificity. Not only are they desirable from a chemical point of view, to obtain homogenous crudes easier to purify, for example, but have proved exceptionally relevant in some cases, such as in the unfortunate incident with thalidomide.⁶¹ However, this is very difficult to achieve and some molecules that are not stereoisomerically pure, such as PSOs, can be used without problems.

Chemoselectivity does not seem a major problem with the actual toolbox of reactions, but the fact that the functional groups used for conjugation are repeated several times in antibodies complicates regiochemical and DAR control when synthesizing ADCs. An obvious solution seems to use reactive functional groups that are not naturally present, which, if orthogonal, should avoid the above-mentioned problems. Unluckily, this means that bioengineering methods must be developed to introduce these groups into each useful antibody (or any other oligomer from natural sources).

A purely chemical methodology to take advantage of the naturally occurring functional groups can only be envisaged if these groups are present in the molecule only once, such as specific amino acid sequences or specific *N*- or *C*-terminal motifs. Only in the case of synthetic molecules, when protecting groups can be introduced, selectivity can be achieved even in the case of repeated functional groups.

1.2.3 Cyclisation

Cyclisation of biomolecules is another important backbone modification, normally employed to increase stability towards enzyme-promoted degradation or to preorganise structures to improve their affinity with target molecules.

Cyclic oligonucleotides have been explored in the past few years and have shown promising results regarding stability, recognition properties and diagnostic and therapeutic applications. Cyclic di-nucleotides possess a huge biological potential, and many methods have been reported for their synthesis. The importance of the cyclic oligonucleotides is, therefore, beyond question, but as they have not been used in this work, more extensive discussions should be found elsewhere.^{62,63}

Cyclic peptides can be found in nature playing very different roles: from human hormones, as oxytocin and somatostatin, to natural venoms such as amanitines and phalloidines (found in *Amanita Phalloides*). They are endowed with improved stability towards enzyme degradation, especially against amino- and carboxypeptidases, and have a lower conformational flexibility.

They have been used for the regulation of protein-protein interactions, such as preventing the amyloid fibrils formation,⁶⁴ a process known to be involved in Alzheimer's, Huntington's and Parkinson's diseases. Further, their rigidity makes them useful to mimic structures with a high level of organization where linear peptides fail, such as epitopes and other protein regions. This concept has been applied to mimic β -strands, turns and helices.⁶⁵

Their role as antibiotics has been broadly explored, as antibacterial⁶⁶ and antifungal⁶⁷ agents. Examples of the former are polymyxins, which disrupt the structure of bacterial cell membranes in Gram-negative bacteria. Caspofungin, which is in the latter category, also bases its mechanism of action in disrupting the cell wall of fungi.

Furthermore, cyclic peptides have also found their spot in the antitumoural fight. Octreotide is used in the treatment of neuroendocrine tumours⁷ and the so-called RGD peptides, in other words, those peptides containing an RGD motif, have been cyclised and modified to enhance integrin recognition, a receptor overexpressed in tumour cells.⁶⁸

1.3 Survey of reactions employed for conjugation and cyclisation

A large amount of transformations has been employed for the synthesis of conjugates and for the cyclisation of peptides and oligonucleotides (with natural or modified backbones). From a chemical point of view, although certain parameters such as concentration must be adjusted, conjugation and cyclisation are similar, as they require complementary functional groups to selectively react with each other.

Many of these processes have been performed using enzymes that recognise specific amino acid residues, as the acyl transfer reaction carried out by transglutaminases,⁵⁹ but they will not be discussed in this work. Besides from enzymatic reactions, dozens of organic reactions have been used to link two different entities. Only the so-called “click” reactions will be commented here, not only because they are the most often used type of reaction nowadays, but also because they possess the most desirable properties for a chemical transformation. All of them will be explained without distinction for conjugation and cyclisation.

Sharpless introduced the concept of click reaction in the year 2001,⁶⁹ and defined the requisites that such transformations should possess:

- give very high yields
- be performed in simple reaction conditions
- be modular
- generate inoffensive by products removable by non-chromatographic techniques
- use of no solvent or benign solvents (such as water)
- purification (if required) must be performed by non-chromatographic techniques

This general concept soon spread to the bioconjugate chemistry, and although it is continuously applied, it must be stated that it is not fully fulfilled, as chromatographic separations are usually needed, especially when working with biomolecules. Some of the most important click reactions are listed in **Table 1.2**.

Entry and Name	Functional group 1	Functional group 2	Product
1, CuAAC	R_1-N_3		
2, SPAAC			
3, Staudinger ligation			
4, Traceless Staudinger ligation			
5, Thiol Michael-type	R_1-SH		
6, Disulfide ligation		$HS-R_2$	
7, NCL			
8, Thiazolidine ligation			
9, Oxime and hydrazone ligation			
10, DA reaction			
11, IEDA reaction			

Table 1.2. Summary of the main methods used for conjugation and cyclisation reactions. R_1 and R_2 belong to the same molecule in the case of cyclisations and to

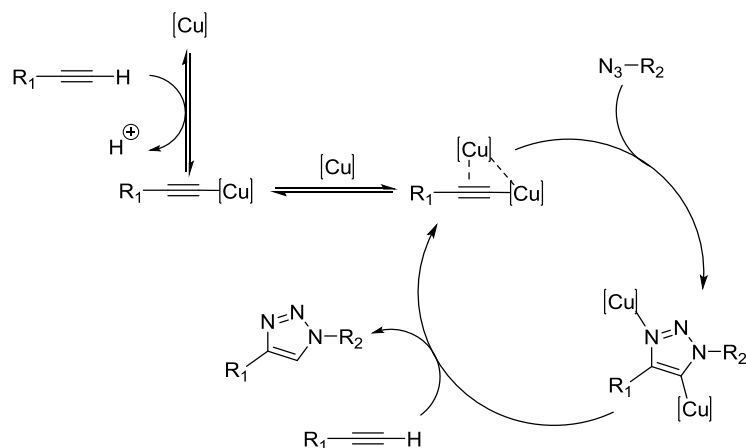
different molecules in the case of conjugations. Abbreviations: CuAAC: Copper (I)-catalysed azide-alkyne cycloaddition, SPAAC: Strain-Promoted Azide-Alkyne Cycloaddition, NCL: Native Chemical Ligation, DA: Diels-Alder, IEDA: Inverse Electron-demand Diels-Alder.

Apart from the reactions employed in this work, that is, the Diels-Alder and the Michael-type reaction, the CuAAC reaction will be also thoroughly discussed, according to its importance nowadays.

1.3.1 Copper (I)-catalysed azide-alkyne cycloaddition (CuAAC)

The CuAAC (**Table 1.2**, entry 1) reaction was first described, independently, by Sharpless and Meldal in 2002.^{70,71} It represents the catalysed version of the Huisgen cycloaddition, discovered in the mid-1900s, and copper not only increases the reaction rate but confers regioselectivity to the transformation. It has been very often used, and it has been sometimes considered the “click reaction”.

The reaction mechanism has been a controversial issue, and since the first one was proposed by Sharpless and co-workers⁷⁰ many attempts to establish a definitive reaction pathway have been made. The most recent studies in the Fokin⁷² and Bertrand⁷³ labs point to a mechanism involving copper dinuclear complexes. The first step (see **Scheme 1.1**) involves formation of a copper(I) acetylide and subsequent coordination of a second copper atom via the acetylide π -orbitals. Both groups have found, by different methods, that this two copper atoms are interchangeable and do not have independent roles. This complex would then react with the azide to generate a bis(copper)triazole, which upon an acidolysis step performed by the reacting alkyne, would generate the final product and regenerate the active dinuclear copper acetylide. The reaction pathway that leads to the formation of the triazole ring from the σ, π -complex is not clear, but computational studies performed with a σ -complex indicated that it is improbable that it is a concerted mechanism.⁷⁴



Scheme 1.1. Mechanism for the CuAAC reaction postulated by the Bertrand group.

The reaction exhibits an enormous scope regarding both alkynes and azides, can be performed in a wide variety of solvents, including water, and is tolerant to a wide range of pH and temperatures.⁷⁴

The drawbacks of this reaction are the oxidation of Cu(I) to unreactive Cu(II) and the toxicity of Cu(I). The latter leads to the formation of Reactive Oxygen Species (ROS) that ultimately can trigger cell death⁵² if the conjugation reaction is performed *in vivo* or *in vitro*, as in certain labelling studies.

Two main strategies have been developed to enhance the reactivity of the Cu(I) species, protect Cu(I) from oxidation and reduce the toxic effects *in vitro* and *in vivo*.

The first one is based in the use of Cu(I) ligands. The most successful family of ligands is that based on polytriazoles, which were found to accelerate the reaction rate and stabilise Cu(I) towards oxidation (**Figure 1.2**). TBTA Was the first ligand of this family to be employed, and showed enhanced catalytic activity and stabilizing properties.⁷⁵ Since then, the research on new polytriazoles has been focused mainly in the reduction of toxicity. TBTA Proved to be troublesome in this aspect, as it is not very soluble in water and inhibits the replication of cells.⁷⁶ THPTA Was shown to be more soluble in water and act a sacrificial reductant, intercepting the formed ROS and thus reducing Cu toxicity.⁷⁷ Soon after, a 6:1 BTTES/Cu formulation showed improved reaction rates and was described as totally non-toxic. Cells exposed to this catalyst system showed no difference in long-term viability when compared to the untreated ones.⁵² BTTAA, a modification of BTTES, showed increased reaction rates than those obtained with the former when working with zebrafish embryos, and did not cause defects in long-term development studies, demonstrating the biocompatibility of this catalytic system.⁷⁸

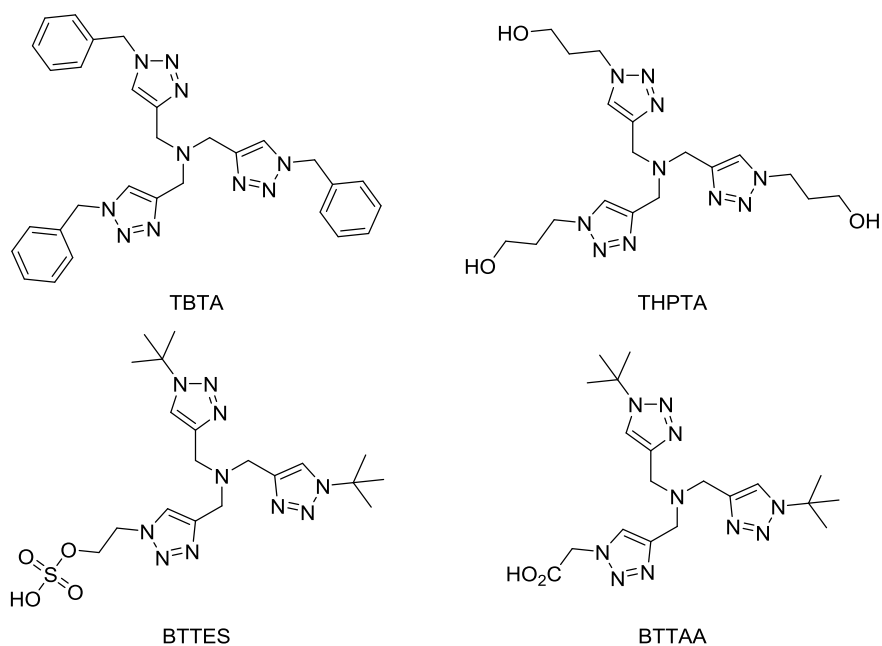


Figure 1.2. Polytriazole ligands for the CuAAC.

The second strategy that has been employed to solve the Cu(I)-related problems is the use of copper-chelating azides (**Figure 1.3**). These molecules raise the effective concentration of copper ions next to the reactive groups and therefore allow for a decrease in the total amount of copper necessary to achieve a reasonable reaction rate. The use of picolyl azides in conjunction with the BTAA ligand, described by Ting, proved to be extremely useful: the reaction is 1.6 times faster than when using the ligand alone, and as non-toxic as the SPAAC reaction (**Table 1.2**, entry 2).⁷⁹ A modification of this picolyl azide by Wu and co-workers afforded a new molecule that reacted faster than the former and showed no apparent toxicity.⁸⁰ Inspired by these works, the group of Taran designed and synthesised a large variety of copper-chelating azides based in bi-, tri- and tetradentate *N*-heterocyclic ligands. Among them, a bis(triazole)azide showed the ability to undergo ultra-fast click reactions, reaching yields of *ca.* 60 % in just 40 seconds, avoiding the use of any other ligands, and needing really low concentrations of copper.⁸¹ Despite its advantages, it must be taken into account that in this strategy the chelating group remains in the final “click” adduct, a possible drawback especially in the cases where this chelating group is bulky, such in the bis(triazole)azides developed by Taran.

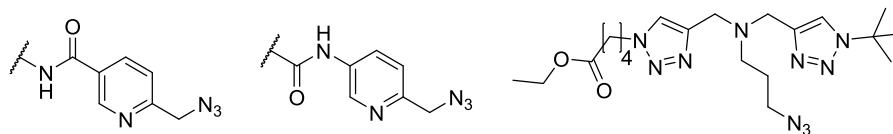


Figure 1.3. Picolyl azides developed by Ting (left) and Wu (centre), and the bis(triazole)azide developed by Taran.

While these results are very positive and seem to solve all the problems related to the copper oxidation and toxicity, it must be stated that most of the studies cited here have investigated toxicity exposing cells or embryos to the catalytic systems for really short periods of time, normally below 5 minutes (some exceptions reached 20 minutes). It is true that with this brief incubations the conversion achieved is normally enough for the desired purposes, but normally longer times are needed to reach complete derivatization. More prolonged exposures to the catalytic system tend to produce toxic side-effects,⁷⁷ which seems to preclude the use of this reaction in systems where Cu(I) quenching is not possible, such as in pretargetting strategies.

The CuAAC reaction has been used for the conjugation of peptides to a large variety of molecules, such as macrocyclic copper-chelators for radiopharmaceuticals,⁸² ¹⁸F radiolabels,⁸³ and polymers,⁸⁴ among others. Oligonucleotides have also been conjugated to multiple structures taking advantage of the CuAAC reaction. For example, peptide-oligonucleotide conjugates were synthesised using a SPFC strategy,⁵⁵ anandamide-siRNA conjugates showing great silencing and transfection properties could be obtained by means of this click reaction,⁸⁵ and heteroglyco 5'-oligonucleotide conjugates were produced using a mixed SPFC-LPFC strategy.⁸⁶

A wide variety of cell-surface imaging systems, mostly relying on glycans and glycoproteins, have also been developed using the CuAAC reaction, and, in fact, this application seems to have been the benchmark for the development and comparison of new ligands and copper-chelating azides.^{77,52,78,80} Other applications of this reaction include Activity-Based Protein Profiling⁸⁷ and studies on the aspirin-dependent protein modification in living cells, in which aspirin analogues bearing alkynes⁸⁸ were used.

1.3.2 Strain-Promoted Azide-Alkyne Cycloaddition (SPAAC)

The Strain-Promoted Azide-Alkyne Cycloaddition (**Table 1.2**, entry 2) was first described by Bertozzi's group in 2004 as an alternative to the CuAAC reaction. It implies the cycloaddition of an azide and a strained alkyne, typically cyclooctyne. It normally takes place readily, with high yields and at room temperature.⁸⁹ Mild reaction

conditions lead to rate constants up to $1 \text{ M}^{-1} \text{ s}^{-1}$. It does not require any catalyst, thus avoiding the use of the toxic Cu(I) ion needed for the CuAAC reaction. This might broaden the scope of this reaction compared to its Cu (I)-catalysed counterpart, especially for *in vivo* and *in vitro* labelling applications. SPAAC is a useful transformation because both reacting groups are not present in biomolecules, thus providing a good starting point towards chemoselectivity. Several cyclooctynes have been developed to enhance chemoselectivity and reaction kinetics, providing this reaction with a wide toolbox of reagents.^{90, 91, 92}

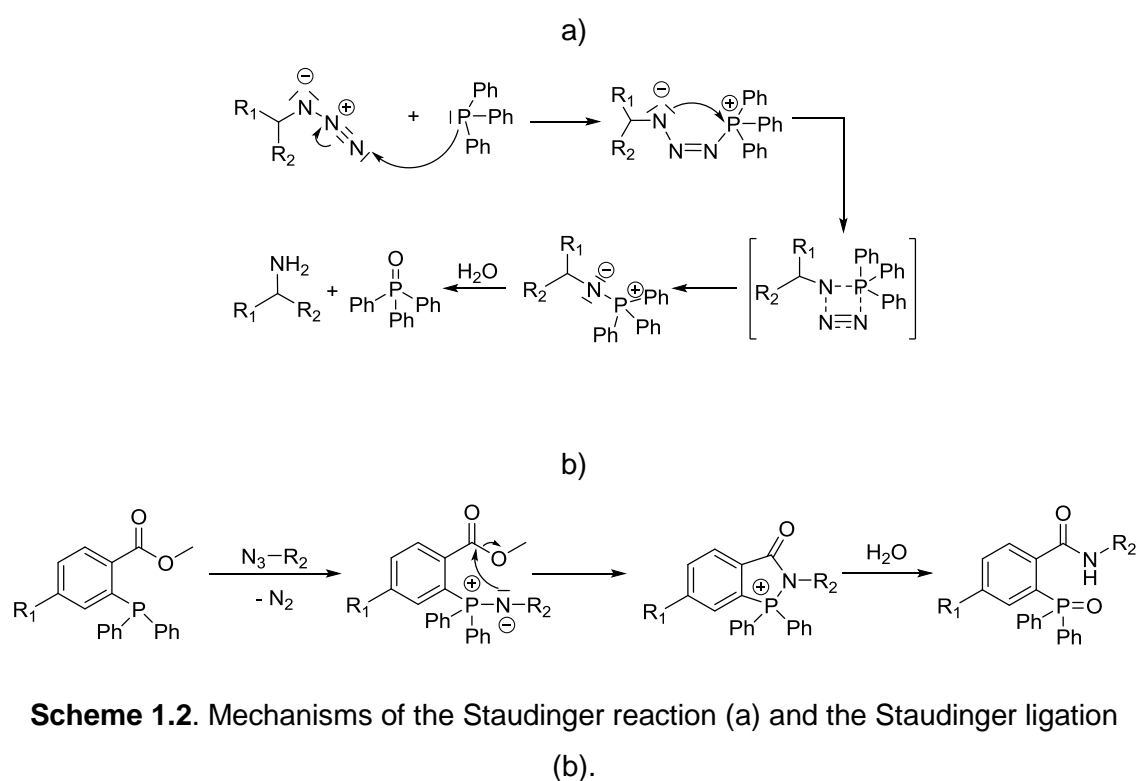
The SPAAC has been used for a wide variety of purposes since its discovery. From the very beginning both proteins and living cells were modified employing this reaction.⁸⁹ Cell-surface labelling has been performed by reacting biotin-derivatised cyclooctynes with glycoproteins metabolically modified with azides. After treatment with an avidin-fluorescein dye, the fluorescence intensity of the cell-membrane could be recorded and measured.⁹⁰ Surface functionalization has also been performed taking advantage of the SPAAC reaction,⁹¹ and quantum dots could be efficiently decorated with biomolecules avoiding the copper-related alteration of luminescent properties.⁹³ Peptide conjugates incorporating radiolabels such as ^{18}F and ^{64}Cu have been synthesised taking advantage of this transformation and successfully used for Positron Emission Tomography (PET) imaging *in vivo*.^{94, 95} Oligonucleotides have also been derivatised with fluorophores and linked to surfaces using this reaction.⁹⁶

Apart from all its advantages and potency, the SPAAC reaction also presents certain problems. First of all, the reaction is not regioselective, and renders two isomeric products that might show different behaviours.⁹⁷ In the second place, although the reaction is fast and very chemoselective, it has been found that cyclooctynes do react, although slowly, with thiols through polar and radical (thiol-yne) additions.⁹⁷ Actually, these have been found to be responsible for most of the azide-independent reactions of cyclooctynes, although some non thiol-related reactions that lead to the labelling of proteins lacking an azide moiety also exist.⁹⁸ This is not a problem for synthetic purposes, but may suppose a drawback for *in vivo* applications, especially taking into account the high concentration of some endogenous thiols (glutathione and cysteine, basically) in the human body. Finally, some cyclooctyne derivatives possess low solubility in water,⁹⁰ a problem expected to be more important for small derivatives rather than for proteins or peptides containing these strained alkynes.

1.3.3 Staudinger ligation

The Staudinger ligation (**Table 1.2**, entry 3) is a methodology originally developed by Saxon and Bertozzi⁹⁹ that is inspired by the Staudinger reaction, developed in 1919. In

the Staudinger reaction, a phosphine reacts with an azide to generate, after N_2 extrusion, an iminophosphorane (see **Scheme 1.2a**). This iminophosphorane can react with a wide variety of electrophiles, and also with water to generate a phosphine oxide and an amine. The Staudinger ligation was designed so that an electrophilic trap as an ester, placed close to the phosphorus atom, reacts with the nucleophilic nitrogen of the iminophosphorane prior to hydrolysis. This generates an amidophosphonium salt (see **Scheme 1.2b**) that hydrolyses to generate an amide bond and a phosphine oxide. The reaction is chemoselective and biorthogonal, as none of the reacting groups is naturally present in biomolecules, and proceeds well in water and biological media without generating too many by-products.



It has been used for the labelling of glycans *in vitro*,⁹⁹ for the study of protein modification *in vitro* by fatty acids¹⁰⁰ and by carbohydrates¹⁰¹ and for the modification of both proteins¹⁰² and oligonucleotides.¹⁰³

Despite its good performance in the above-mentioned cases, the Staudinger ligation presents certain problems. The first is its low rate, with second-order constants typically around $0.002 \text{ M}^{-1} \text{ S}^{-1}$. This implies that large amounts of one of the reagents must be used to achieve proper conversions, which can be a problem in imaging purposes due to excessive background signal.¹⁰⁴ Another consequence of this slow reactivity is that

phosphines tend to be oxidised in physiological medium, which might be a drawback for *in vitro* and *in vivo* applications.¹⁰⁴

Finally, the phosphine, whose role is just to aid in the creation of the amide bond, remains present in the final adduct, a problem that its traceless counterpart (**Table 1.2**, entry 4) solves.

1.3.4 Traceless Staudinger ligation

The traceless Staudinger ligation (**Table 1.2**, entry 4) was reported, at the same time, by Bertozzi and by Raines in year 2000.^{105,106} The reaction relies on the intramolecular attack of an iminophosphorane, generated upon Staudinger reaction (see **Scheme 1.2a**), to a thioester carbonyl and subsequent hydrolysis of the amidophosphonium salt (see **Scheme 1.3**). As its name implies, this reaction does not leave any atom from the phosphine in the final adduct, in contrast with the previously described Staudinger ligation (see **Scheme 1.2b**).



Scheme 1.3. Mechanism of the traceless Staudinger ligation.

Each group used different phosphines for the ligation, although both chose aromatic phosphines as starting point, due to the improved stability towards oxidation by atmospheric oxygen. Nowadays the most common method implies the use of the thioester depicted in **Table 1.2**, which was developed by Raines one year after his first publication on the traceless Staudinger ligation.¹⁰⁷ It has proved to react as fast as the one described by Bertozzi and co-workers, but its chemoselectivity makes it more suitable when working with molecules containing more functionalities than the reacting azide.¹⁰⁸

The traceless Staudinger ligation was initially proposed by Raines as a method for the chemical synthesis of proteins avoiding the need of cysteines present in the native chemical ligation. As an example of this application, the 124 amino acid protein ribonuclease A was synthesised using several orthogonal chemical ligation methods, including the traceless Staudinger ligation.¹⁰⁹ Furthermore it has also been applied to the cyclisation of peptides¹¹⁰ and the synthesis of microarrays for the immobilization of proteins.¹¹¹ From a synthetic point of view, the stereoselective synthesis of glycosyl

amides¹¹² and of ¹⁸F probes for PET imaging¹¹³ have been achieved. No *in vivo* or *in vitro* applications seem to have been published to date.

Even though this reaction has been employed for very different purposes, the traceless Staudinger ligation still presents some problems. It has been described that glycine is needed either in the *N*-terminal or the *C*-terminal position of the junction of the peptidic fragments to obtain good yields. When this is not the case, reaction yields drop to below 50 %, ¹⁰⁸ restricting the applicability of the reaction. Although it has been described that electron-donating groups in the aromatic rings of the phosphine help to increase these yields, an organic, apolar solvent is needed,¹¹⁴ precluding the use of this strategy with some biomolecules, and, of course, its use in *in vivo* and *in vitro* assays. To solve this problem several groups have focused on the synthesis and evaluation of new phosphines capable of performing the reaction in water, although with reported yields below 80 %.^{115,116}

It is also one of the slowest click reactions, with second-order rate constants around $8 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ in the best experiments, which involves reaction times longer than 5 h to reach good yields in the best-performing cases.¹⁰⁸

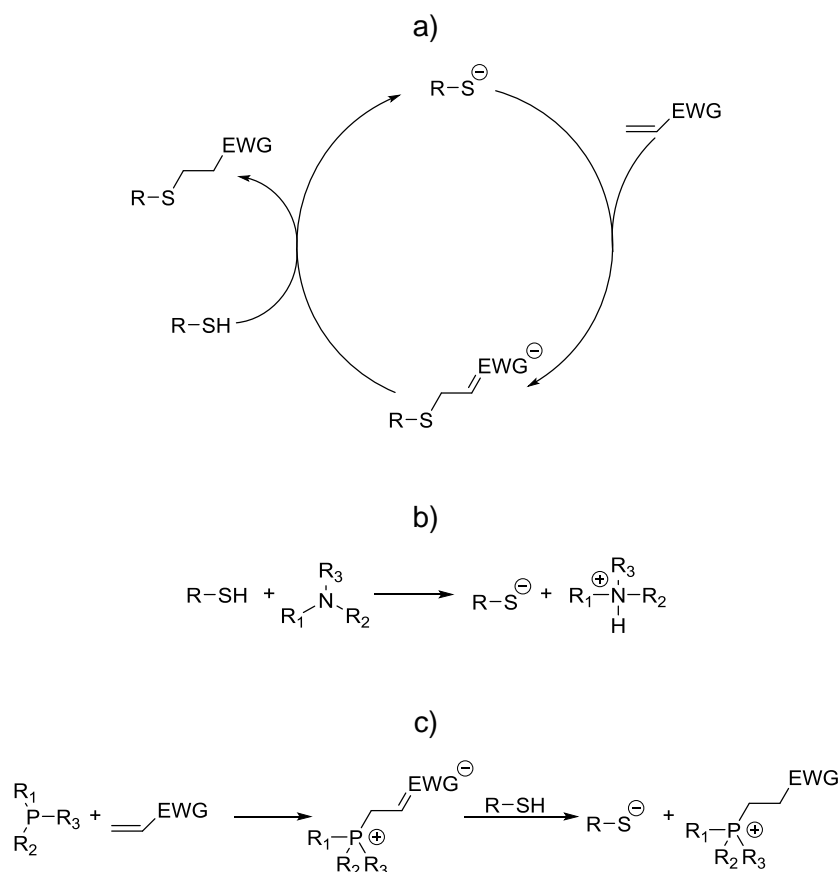
1.3.5 Thiol Michael-type reaction

Reactions involving the highly nucleophilic thiols have been known since many years ago and include the thiol-halide reaction, the thiol-ene and thiol-yne radical reactions, disulfide formation, the thiol-epoxy reactions, and many others. Among all, the Michael-type reaction (**Table 1.2**, entry 5) has recently been the focus of extensive studies and has been used for a great variety of purposes. It consists in the conjugate addition of a thiol to the β -carbon of an α,β -unsaturated carbonyl (or equivalent), rendering a thioether linkage.

The reaction relies on the inherent electron density of the sulfur atom, which enables thiols to react under mild (catalysed or not) reaction conditions with a broad scope of substrates. This reaction can be performed in a wide variety of solvents, ranging from apolar, organic solvents such as benzene to water, and gives high yields and selectivities.

From a mechanistic point of view, the reaction proceeds as follows: small traces of thiolate are created in the medium by different mechanisms, and it is this species that attacks the activated, electron-poor double bond to generate an enolate (or equivalent, see **Scheme 1.4a**). This strong base will now subtract a proton from a thiol, rendering the final product and regenerating the active thiolate.¹¹⁷ Two main paths lead to the formation of the active thiolate, the base-promoted and the nucleophile-promoted mechanisms (see **Scheme 1.4b** and **Scheme 1.4c**). Both of them ultimately generate

the conjugated base of the thiol through an acid-base reaction, although each of them does it in a different way. In the first mechanism, an organic base, normally an amine, is the species that subtracts the thiol proton, while in the latter a nucleophile, typically a phosphine, reacts with the vinyl group to generate a stabilised carbanion, the true basic species that will subtract the thiol proton.



Scheme 1.4. Michael-type reaction between a thiol and an activated alkene (a), base-initiated mechanism (b) and nucleophile-initiated mechanism (c).

However, several studies have recently shown that reality is not as simple as that. Bowman and co-workers published in 2010 a paper suggesting that the catalytic action of amines should not be exclusively ascribed to their basicity.¹¹⁸ A comparison of three different alkylamines with a very small difference in pK_a showed reaction rates that correlated to their ability to act as nucleophiles. Other comparisons evidenced that the nucleophilicity of the amine had a greater effect on the reaction rate than its pK_a value. Altogether, evidences collected by the Bowman's group suggest that a purely acid-base mechanism is rarely operating when using amines as catalysts, and mixed or nucleophilic ones must be taken into account. A recent computational study by Northrop and co-workers¹¹⁹ showed that only in very polar solvents the base-catalysed

mechanism can account for the overall reactivity. In other solvents, a mixture of this and other mechanisms, such as ion-pair formation (in this mechanism the initiator species is the thiolate/ammonium salt ion-pair) (see **Scheme 1.5**) and nucleophilic initiation, must be taken into account to explain the observed reactivity. Polar solvents such as DMF do not only facilitate the base-catalysed generation of free thiolates, but these species can spontaneously form without the need of an external base or nucleophile, therefore avoiding the need for catalysts.^{117, 119}



Scheme 1.5. Ion-pair formation.

The overall reaction rate in the thiol Michael-type reaction depends on the solvent, the initiator (base or nucleophile), the double bond and the thiol. Thiol reactivity matches with the ease of forming a thiolate anion, and not the nucleophilicity of this species.¹¹⁹. This is directly related to the thiol pK_a , as shown by the fact that the more acidic ethyl 2-mercaptopropionate reacts faster than methyl 3-mercaptopropionate despite the fact that the latter is less sterically hindered.¹¹⁸ The fact that any thiolate is a strong nucleophile that will react readily with the conjugated double bond, whereas the deprotonation of the starting thiol is not so fast, explains this observation.

As to the double bonds, it has been noticed that the reactivity order for typical vinyl compounds is as follows: maleimides > vinyl sulfones > acrylates > acrylamides > methacrylates; which correlates with the electrophilicity of the different β -carbons.

Regarding the initiator, it has been previously said that the factors governing the reaction rate are not totally clear when using bases, as many different initiation mechanisms can occur, depending on the solvent. A clear trend is not easy to establish, and important differences can be found from one article to another (compare, for example, references 118 and 119). On the other hand, the trend for different phosphines, which are known to act only as nucleophiles,¹¹⁸ has been established on the basis of their nucleophilicity: $\text{PAI}_3 > \text{PAI}_2\text{Ar} > \text{PAIAr}_2 > \text{PAR}_3$ (AI= alkyl, Ar = aryl). Although trialkylphosphines induce faster reaction rates they tend to be pyrophoric, which makes the use of PAI_2Ar phosphines more convenient.

Finally, polar solvents have demonstrated to give faster reaction rates than non-polar ones, as they favour the formation of free thiolates, thus by-passing the slower ion-pair mechanism.¹¹⁹

The thiol Michael-type reaction has been broadly used in the polymer field, although no examples will be described here as it goes beyond the scope of this work. Regarding the biological field, only a few practical examples of this reaction not involving maleimides as electrophiles can be found. One of the most interesting is the primer extension of vinyl sulphonamide- and acrylamide-derivatised nucleotides to introduce activated double bonds in oligonucleotides, which enabled the synthesis of their peptide conjugates.¹²⁰ Regarding the thiol-maleimide chemistry, many applications have been described in the literature. Attachment of cysteine moieties to nucleobases afforded thiol-modified oligonucleotides that could be conjugated to maleimide-containing peptides.¹²¹ Long peptides containing cysteine have been conjugated to maleimide-derivatised carbohydrates and carbohydrate dendrimers,¹²² and c(RGD) peptide dendrimers have been synthesised and conjugated to ⁶⁸Ga labels by means of the thiol-maleimide reaction.¹²³ Several surfaces have been functionalised using the thiol-maleimide chemistry, to render, for example, an antibacterial surface.¹²⁴

Not many years ago, our group described the protection of maleimides using 2,5-dimethylfuran, which enabled the on-resin derivatization of oligonucleotides with maleimides for the first time. By means of this protection strategy, several oligonucleotides could be readily derivatised with masked maleimides for their posterior deprotection and reaction with different thiols, such as long peptides and cholesterol.¹²⁵ Phosphorothioate oligonucleotides were also derivatised with protected-maleimides and further conjugated to biotin and peptides, with no side-reactions with the phosphorothioate linkage.¹²⁶ Simultaneous introduction of a protected maleimide and a thiol into oligonucleotide chains enabled the synthesis of cyclic oligonucleotides.¹²⁷

Introducing maleimides anywhere in the peptide chain enabled the labelling of peptides at different positions, and the combination of a protected and an unprotected maleimide in the same oligomer was used for the synthesis of cyclic and bicyclic peptides and to prepare doubly derivatised polyamides.¹²⁸ Further, the combination of differently-protected thiol groups enabled a regiospecific tandem cyclisation-conjugation procedure with maleimide-containing oligonucleotides.¹²⁹

Despite being quick, clean and selective, the thiol-maleimide reaction also suffers from some problems that will be described in chapter 3.

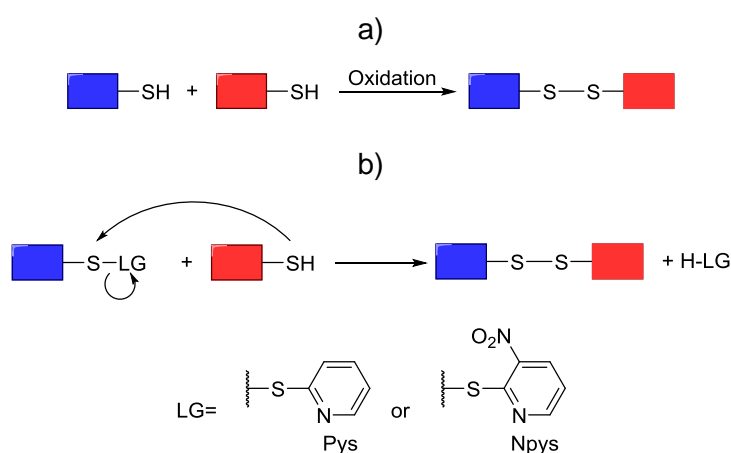
1.3.6 Disulfide ligation

Disulfide ligation (**Table 1.2**, entry 6) is basically used when cleavable linkers must be obtained, and relies on the variable stability of the disulfide bond in different environments. While disulfide bonds can be synthesised, handled and stored in non-reducing solutions, reducing environments, as those found inside cells,¹³⁰ trigger the

reduction of this linkage. Based on this different behaviour, reversible linkages can be synthesised and used in cases where they must be cleaved in a controlled fashion, as for example in the synthesis of ADCs.

Both the formation and cleavage of the disulfide bond take place in mild aqueous conditions, with kinetics that strongly depend in the disulfide environment. The reaction of sulfhydryl groups is advantageous because it uses functional groups naturally present in biomolecules such as peptides and proteins.

The synthesis of disulfide conjugates is normally accomplished by two different strategies: the direct oxidation of two thiols or by preliminary activation of one of the thiols with a good leaving group, normally the pyridylsulfenyl (Pys) or 3-nitropyridylsulfenyl (Npys) groups (see **Scheme 1.6**). Direct oxidation of thiols normally yields the symmetric disulfide together with the desired heterodimer, so preliminary activation is normally recommended.¹³¹



Scheme 1.6. Synthesis of disulfide conjugates by a) direct oxidation and b) thiol-disulfide exchange with preliminary activation.

The local environment of the disulfide bond modulates its stability, basically through steric hindrance or by electrostatic attractions/repulsions. Kellogg *et al.* demonstrated that steric hindrance around the disulfide bond could tune its stability both in buffered solutions and *in vivo*, therefore regulating the pharmacokinetic properties and activity of different ADCs.¹³² On the other hand, Gauthier and co-workers have shown that the electrostatic microenvironment of a disulfide bond can modulate its cleavage rate over several orders of magnitude.^{133,134} These last studies postulate that the reducing agent, glutathione or cysteine in their case, is attracted or repelled by the electrostatic environment of the disulfide bond, thus accelerating or inhibiting its cleavage, respectively. It must be stated that, although glutathione can act as a reducing agent, it

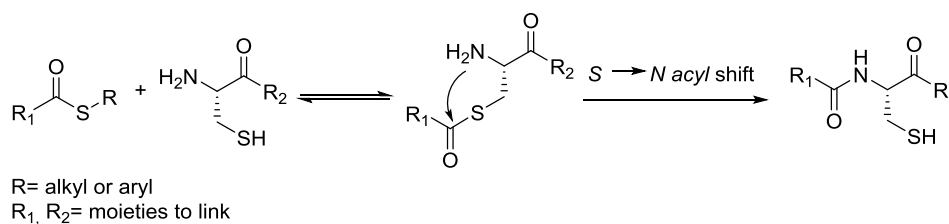
is not the only found in the intracellular milieu, and therefore this study does not cover all the possible reduction paths.¹³⁰

Disulfide bond formation has been applied to the synthesis of different types of conjugates. Potent ADCs have been obtained by linking the huC242 antibody to maytansinoid moieties,¹³² and peptide-oligonucleotide conjugates of various kinds have been synthesised through thiol-thiol coupling.^{135,136} Disulfide linkage has also been used for the synthesis of peptide-nanoparticle¹³⁷ and peptide-polymer conjugates.¹³⁸

On the negative side, the need for solvent-accessible thiols for the conjugation process may be sometimes not reached or exceeded, causing lack of reaction or lack of positional selectivity, respectively. These problems are especially important when dealing with large peptides and proteins, where more than one accessible cysteine residue may be found or complex folding patterns can hinder the sulfhydryl groups. Control over disulfide bond stability can be troublesome, as it is strongly influenced by the position of sterically-hindered amino acids around the disulfide bond and cannot be easily predicted.¹³² Regarding the electrostatic regulation, a specific peptide sequence in the conjugate is commonly needed, precluding the adjustment of the disulfide electrostatic environment by modulating the peptide primary structure. So far, there seems to be no general rules to achieve an appropriate stability balance that avoids the premature loss of payload and achieving the desired conjugate activity.

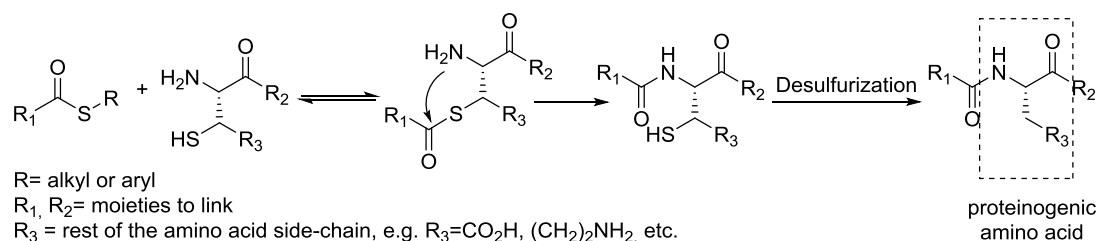
1.3.7 Native Chemical Ligation (NCL)

Native Chemical Ligation (**Table 1.2**, entry 7) is the name given to the reaction of an *N*-terminal cysteine residue with a thioester to render an amide bond. Developed in 1994 by Kent and co-workers,¹³⁹ this reaction permits the coupling of two completely unprotected peptide fragments yielding a natural amide bond. The reaction is selective for *N*-terminal cysteines (or other 2-aminothiols) even in the presence of other thiols, and proceeds in mild conditions. Furthermore, it yields a native amide bond, which means that traceless ligations can be performed, expanding the possibilities for the chemical synthesis of long peptides. Mechanistically, the reaction consists in a reversible transthioesterification followed by an *S*→*N* acyl transfer through a five-membered transition state, rendering the final amide bond with a cysteine residue at the ligation site (see **Scheme 1.7**).¹⁴⁰ To maintain a balance between thioester reactivity and stability, the commonly employed strategy is the synthesis of alkylthioesters, which are more stable than their aryl counterparts, and the use of arylthiophenols as catalysts.¹⁴¹ These additives ensure the *in situ* generation of a more reactive arylthioester that will increase the ligation rate.



Scheme 1.7. Native Chemical Ligation mechanism.

Of course the reaction was, at the beginning, limited to the ligation or conjugation of peptide fragments that contained an *N*-terminal cysteine. Nowadays, though, many amino acids can be introduced at the ligation site taking advantage of desulfurization methodologies. These allow for the use of unnatural, thiol-containing amino acids that will, after removal of the thiol group, render a natural amino acid and amide bond (see **Scheme 1.8**). Using such strategy, precursors of amino acids with reactive groups on the side chain have been used for ligation, for example precursors of aspartic acid,¹⁴² arginine,¹⁴³ tryptophan¹⁴⁴ and lysine.¹⁴⁵



Scheme 1.8. The ligation + desulfurization strategy.

As for the introduction of the thioester functionality, many different strategies have been developed, in particular to allow for the use of Fmoc SPPS. However, this goes beyond the scope of this work and other sources might be consulted.¹⁴⁶

NCL Has been used to obtain long peptides or proteins, often incorporating post-translational modifications, by chemical synthesis,^{139,147,148} for the synthesis of peptide conjugates with DNA¹⁴⁹ and RNA,¹⁵⁰ and for the modification of silicon nanowires surfaces with peptides,¹⁵¹ among others. Also, the protection of cysteine residues using thiazolidine formation enabled for sequential NCL processes, as the one described by Kent and co-workers that afforded the chemical synthesis of Human insulin-like growth factor 1 and some analogues.¹⁵²

Whilst widely used, NCL still presents some problems. The methodology needs cysteine residues in the protein or peptide chain, a relatively infrequent amino acid in natural peptides and proteins. Although, as mentioned, the ligation + desulfurization

strategy allows thiol-modified amino acids to be involved in amide formation, these amino acid surrogates are not always commercially available. Furthermore, a desulfurization step must be performed after ligation, and typical yields for these procedures are not very high (60-80 %). Another factor that might be troublesome is the steric influence of the amino acid bearing the thioester moiety. A study performed by Hackeng *et al.* showed that some amino acids enabled ligation in less than 4 hours, but others did not reach good conversions even after 48 hours.¹⁵³ This was partially solved by the introduction of the oxo-ester ligation by Danishefsky and co-workers.¹⁵⁴ This method, which uses the more active *p*-nitrophenol esters to achieve better reaction rates, improved the reaction times for bulky amino acids, but unfortunately yields only ranged from moderate to good (50-80%).

1.3.8 Thiazolidine ligation

The reaction of an aldehyde with a 1,2-aminothiol to render a thiazolidine (**Table 1.2**, entry 8) has been known for many years, but it was Tam and co-workers who first introduced it for the synthesis of biomolecule analogs, using cysteine as the natural source for 1,2-aminothiols.¹⁵⁵ The reaction occurs readily (reaction times below 10 minutes have been described) and orthogonally in acidic conditions, as demonstrated in some of the earliest studies.¹⁵⁶ Noteworthy, thiazolidine formation is orthogonal to *N*-acylated cysteines, as thiol groups react with aldehydes to yield reversible hemithioacetals. This implies that thiazolidine ligation can be envisaged as a platform for the selective derivatization of *N*-terminal cysteines in peptides that also contain internal cysteines.

The stability of thiazolidines can be modulated modifying substitution at the 2 position. 2-Alkyl thiazolidines tend to be more stable than 2-aryl thiazolidines, and these are more stable than those 2,2-doubly-substituted.¹⁵⁷

The introduction of the two necessary moieties is quite straightforward, especially in the case of peptides. The *N*-terminal cysteine residue can be introduced by means of SPPS, and the aldehyde moiety can be installed by mild oxidation of *N*-terminal Ser and Thr residues using periodate¹⁵⁶ or by incorporation of its acetalic precursor during SPPS.¹⁵⁵ For the introduction of the carbonyl moiety in oligonucleotides, the common strategy is the use of a phosphoramidite bearing a masked aldehyde (see below).

This reaction has been used for the modification of both peptides and oligonucleotides. Peptide dendrimers have been synthesised employing the ligation between an *N*-terminal cysteine and an aldehyde.¹⁵⁵ This reaction proved to be more useful than oxime and hydrazine ligations, as it was faster and yielded more stable products. Cyclic peptides could also be obtained using this strategy, with no apparent size limitation for

the cyclisation reaction.¹⁵⁸ More complicated peptide structures such as antibodies were also modified using thiazolidine ligation. By these means, tumour-targeting ADCs were synthesised, and their activity tested.¹⁵⁹ In this case, the thiazolidine linker was designed to be hydrolysable, so the drug attached to the antibody could be slowly released to perform its action.

Oligonucleotides have also been used in thiazolidine ligation methodologies. The group of Gait and Oretskaya synthesised peptide-oligonucleotide conjugates taking advantage of a phosphoramidite bearing a protected 1,2-diol in the 2' position of uracil. This was introduced at the internal positions of an oligonucleotide chain and oxidised in mild conditions to render the necessary aldehyde for conjugation.¹⁶⁰ Dumy and co-workers used a different phosphoramidite to link dystamycin-based peptides to the 5'-end of different oligonucleotide chains.¹⁶¹

Finally, and taking advantage of the reversibility of the thiazolidine adducts mentioned above, these have also been used for the protection of *N*-terminal cysteines in sequential NCL processes.¹⁵²

The reaction, though, is far from perfection. It cannot be performed *in vitro* or *in vivo*, as the acidic conditions needed for its orthogonality are not possible in living systems. Although normally described to be fast, thiazolidine formation may need really long reaction times, that can reach four days, to achieve good yields.¹⁵⁹ This, therefore, implies that reducing agents have to be added to prevent oxidation of the cysteine thiol. As many others, the reaction generates two isomers, which can be separated only when low molecular weight adducts are generated.¹⁵⁶

1.3.9 Oxime and hydrazone ligation

The reaction of carbonyl compounds, especially aldehydes, with hydroxylamines and hydrazines (**Table 1.2**, entry 9) has been extensively used for bioconjugation purposes. The reaction, as in the case of thiazolidine formation (**Table 1.2**, entry 8), is normally performed in water, at acidic pH (4.5-5.5) and under mild conditions. It is selective due to the α -effect, which increases the nitrogen nucleophilicity and allows them to react in such conditions. This effect also contributes to the major stability of oximes and hydrazones towards hydrolysis when compared to their imine counterparts, especially in the case of oximes, which proved to be the most stable linkages.¹⁶² Still, both oxime and hydrazone formation can be used for the preparation of pH-sensitive biomaterials, as their stability can be easily modulated by the carbonyl group used for the conjugation.¹⁶³

Peptide conjugates have been synthesised taking advantage of oxime ligation: daunorubicin-linked peptides have been synthesised for cancer therapy,¹⁶⁴ and sugar

conjugated peptides have been used as tumour-related antigens.¹⁶⁵ On the other hand, hydrazone ligation has been employed for the synthesis of antibody-conjugated quantum dots.¹⁶⁶

Oxime formation has also been investigated as another possible methodology for the synthesis of cyclic peptides,¹⁶⁷ as well as in conjunction with the CuAAC reaction for the regioselective ligation of cyclic RGD peptides onto cyclopeptide scaffolds.¹⁶⁸

Oligonucleotide-peptide conjugates have been generated through oxime and hydrazone ligations, as in a paper by Zatsepin *et al.* where the two methodologies and the thiazolidine formation reaction are used.¹⁶⁰

Despite their usefulness, these two reactions also have some limitations. The first, as it happens with the thiazolidine ligation, is that they cannot be performed *in vivo* or *in vitro* due to pH-related limitations: at neutral pH the reactions are not selective and need catalysis to proceed at reasonable rates.¹⁶⁹ Although these condensation reactions are described to be typically high-yielding, there are several examples in which yields are moderate, especially when working with complex systems.¹⁷⁰ Finally, a fast comparison of the three carbonyl-involving condensation reactions reviewed in this work shows that both oxime and hydrazone ligations perform worse than thiazolidine formation, with reaction times that can double those of the latter reaction and lower product stabilities.¹⁵⁵

1.3.10 Diels-Alder (DA) ligation

The DA reaction (**Table 1.2**, entry 10) was discovered by Diels and Alder in 1928 and has been, since then, one of the most often used reaction in organic chemistry. It is a [4+2] cycloaddition reaction in which a 1,3-diene reacts with an alkene to yield a cyclohexene adduct. The driving force of the reaction is the generation of new σ -bonds, which are more stable than π -bonds. The classical reaction is normally performed with electron-poor alkenes (or alkynes) and electron-rich dienes, although alternative versions such as the inverse electron-demand DA (IEDA) reaction exist (see **Table 1.2**, entry 11).

The reaction is very selective and almost completely orthogonal, as the reacting groups are not normally present in biomolecules. The only incompatibility that can be found is the presence of thiols, which, depending on the nature of the electron-withdrawing substituent, can react with the electron-poor alkene in a Michael-type reaction (see **Table 1.2**, entry 5). Furthermore, the reaction does not need catalysis, although several catalysts can be used to accelerate this transformation, especially on those cases that require harsh conditions. Apart from the extremely important influence of the solvent in the reaction rate, which will be discussed in chapter 2, catalysis by iminium

ion with α -nucleophiles¹⁷¹ and by Lewis acid activation¹⁷² have been described. Another benefit of the DA reaction is that both the kinetics of the reaction¹⁷³ and the stability of the generated adducts¹⁷⁴ can be tuned. That means that, by just modulating the substituents present in the reaction partners, acceleration of certain DA reactions is possible and that, in some cases, reversible adducts can be turned to irreversible, or vice versa. Generally, though, DA adducts tend to be very stable and need high temperatures to revert to the starting materials through a retro-Diels-Alder reaction (rDA).

Thanks to its advantages, the DA reaction has been used for multiple purposes in the field of bioconjugation. Oligonucleotide conjugation has classically been performed in water by reacting diene-modified oligonucleotides with maleimide-containing compounds. A first approach was first performed by Hill *et al.* when they conjugated 5'-diene-containing oligonucleotides to small organic molecules such as fluorophores and biotin.¹⁷⁵ Not long after, our group published methodology to react the same kind of oligonucleotides with maleimide-containing peptides to render peptide-oligonucleotide conjugates.⁵³ Other research groups have expanded this strategy, using diene-modified nucleosides that allowed for the derivatization of oligonucleotides at internal positions.¹⁷⁶

Another application of the DA reaction related to oligonucleotides is the one developed in our group some years ago, which consisted in a DA-rDA reaction sequence to allow for the introduction of maleimides into oligonucleotide chains using standard solid-phase methodologies (see section 1.3.5). Other biomolecules, such as carbohydrates,¹⁷⁷ have also been conjugated taking advantage of this reaction. Surface immobilization of different biomolecules has been performed using the DA reaction. Specifically, the DA cycloaddition was used to functionalise a surface with alkyne moieties that were afterwards used for the derivatization with specific recognition motifs using the CuAAC reaction. Maleimide-modified peptides have also been used for DA ligation with other molecules apart from oligonucleotides, although not so many examples can be found. In one of them, Tischer *et al.* have used them for a light-induced DA reaction with modified biosurfaces.¹⁷⁸ Introduction of dienes to peptide chains for their DA reactions will be discussed in chapter 2.

On the negative side, the DA reaction can furnish different isomers. If the diene or the dienophile are not symmetrical, different regioisomers might be formed, although normally some of them are preferred over the others (see **Figure 1.4a**). This problem is normally solved by working with symmetrical molecules, such as *N*-substituted maleimides. Apart from regioisomers, several stereoisomers can be generated depending on the relative orientation of the reagents when approximating each other.

As seen in **Figure 1.4b** two diastereomeric sets of enantiomers can be generated. The *endo* approximation is the most kinetically favourable, although the *exo* product is normally also observed. Facial selectivity is only achieved when the faces of the diene or the dienophile are asymmetrical.

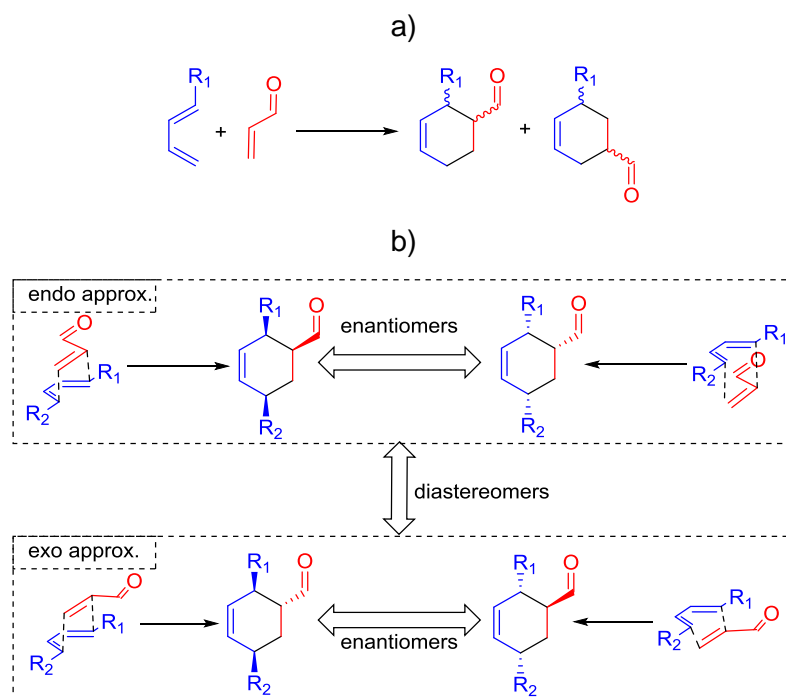


Figure 1.4. Regioisomers (a) and stereoisomers (b) generated in a Diels-Alder reaction.

1.3.11 Inverse Electron-demand Diels-Alder (IEDA)

The Inverse Electron-demand Diels-Alder ligation (**Table 1.2**, entry 11) involves the reaction of an electron-poor, conjugated diene with an electron-rich or strained double bond. The archetypical transformation is performed with *trans*-cyclooctene as the dienophile and 1,2,4,5-tetrazines as dienes, and proceeds with extrusion of N_2 to yield the final cycloadduct in extremely short reaction times ($k = 2000 \text{ M}^{-1} \text{ s}^{-1}$ at room temperature, although some constants above $10^6 \text{ M}^{-1} \text{ s}^{-1}$ have been measured).¹⁷⁹

Several articles describe the use of this reaction for the synthesis of conjugates. Ternary conjugates have been synthesised using tandem S_NAr -IEDA reactions on 3,6-dichloro-1,2,4,5-tetrazines,¹⁸⁰ and the group of Kolmar showed that oxime ligation allows the site-specific introduction of tetrazines and strained alkenes for peptide conjugation.¹⁸¹ The reaction is so selective, fast and clean that has also been used for several *in vivo* or *in vitro* imaging purposes. Tumour pretargeting and imaging was

accomplished using a *trans*-cyclooctene/tetrazine reaction,¹⁸² and PET ¹⁸F labels have been successfully developed using similar approaches.⁵¹

Another advantage of the IEDA reaction, besides its bioorthogonality and, most of the times, fast reaction rates, is its orthogonality to the CuAAC reaction (**Table 1.2**, entry 1). Taking advantage of this fact, double dye-tagged oligonucleotides have been synthesised.¹⁸³

Despite the fact that it is nowadays one of the most commonly used reactions and its great applicability, several issues still remain fully or partially unsolved. It has been described that, although *trans*-cyclooctene reacts readily with tetrazines, it can also react with thiols and isomerise to the unreactive *cis*-cyclooctene,¹⁸⁴ a fact that imposes the need to search for other dienes. Apart from *trans*-cyclooctene, other strained alkenes have been used for the IEDA, as for example norbornenes¹⁸⁵ and modified cyclopropenes,¹⁸⁴ although these react hundreds of times more slowly (rate constants ranging from 10 to 0.1 M⁻¹ s⁻¹). Terminal alkenes, which have also been used as alternatives to *trans*-cyclooctene and which are expected to be more stable than strained alkenes, are the ones that react more slowly, with rate constants ranging from 0.1 to 0.001 M⁻¹s⁻¹. On the other hand, 1,2,4,5-tetrazines have dominated the IEDA landscape, although they also present several problems. The most reactive tetrazines towards dienophiles tend to be the least stable, and their solubility is, in general, quite unpredictable.¹⁸⁶ Finally, it must be stated that tetrazines are normally quite bulky, although some exceptions can be found in the literature.¹⁸⁰ This has triggered the search for other alternatives, and it seems that the use of 1,2,4-triazines might solve some of these problems and rise as a generally applicable alternative to 1,2,4,5-tetrazines.¹⁸⁷

Lastly, as both derivatised *trans*-cyclooctene and tetrazines are unsymmetrical, the concomitant cycloadditions can form different regioisomers, a fact that does not seem to have been investigated thoroughly. In the case of terminal alkenes, several regio- and stereoisomers may also be formed.¹⁸⁸

1.4 Objectives

The objectives of this work were:

- 1- To find an alternative allowing for the use of the Diels-Alder reaction for the synthesis of conjugates using diene-derivatised polyamides.
- 2- To explore the use of 2,2-disubstituted cyclopent-4-ene-1,3-diones as maleimide analogs in Michael-type reactions for the conjugation and cyclisation of peptides.

1.5 Abbreviations

ADC: Antibody-Drug Conjugate
 CPP: Cell-Penetrating Peptide
 CuAAC: Copper (I)-catalysed azide-alkyne cycloaddition
 DA: Diels-Alder
 DAR: Drug-to-Antibody Ratio
 FDA: Food and Drug Administration
 HPLC: High Performance Liquid Chromatography
 IEDA: Inverse Electron-demand Diels-Alder
 LNA: Locked Nucleic Acids
 LPFC: Liquid-Phase Fragment Conjugation
 NCL: Native Chemical Ligation
 NMR: Nuclear Magnetic Resonance
 Npys: 3-Nitropyridylsulfenyl
 PEG: Polyethylenglycol
 PMO: Phosphorodiamidate morpholino oligonucleotides
 PNA: Peptide Nucleic Acid
 PSO: Phosphorothioate Oligonucleotides
 Pys: Pyridylsulfenyl
 SPAAC: Strain-Promoted Azide-Alkyne Cycloaddition
 SPFC: Solid-Phase Fragment Conjugation
 SPPS: Solid-Phase Peptide Synthesis

1.6 Bibliography

- (1) Iqbal, K.; del C. Alonso, A.; Chen, S.; Chohan, M. O.; El-Akkad, E.; Gong, C.-X.; Khatoun, S.; Li, B.; Liu, F.; Rahman, A.; Tanimukai, H.; Grundke-Iqbal, I. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **2005**, 1739 (2-3), 198.
- (2) Rubinsztein, D. C. *Nature* **2006**, 443 (7113), 780.
- (3) Wang, W.; Rusin, O.; Xu, X.; Kim, K. K.; Escobedo, J. O.; Fakayode, S. O.; Fletcher, K. A.; Lowry, M.; Schowalter, C. M.; Lawrence, C. M.; Fronczek, F. R.; Warner, I. M.; Strongin, R. M. *J. Am. Chem. Soc.* **2005**, 127 (45), 15949.
- (4) DrugBank: Biotech Drugs
<http://www.drugbank.ca/drugs?approved=1&c=name&d=up&type=biotech> (accessed Oct 22, 2015).
- (5) Tse, M. T. *Nat. Rev. Drug Discov.* **2013**, 12 (3), 179.

- (6) Hebert, M. F. *Adv. Drug Deliv. Rev.* **1997**, 27 (2), 201.
- (7) Wolin, E. M. *Gastrointest. Cancer Res. GCR* **2012**, 5 (5), 161.
- (8) Falagas, M. E.; Kasiakou, S. K.; Saravolatz, L. D. *Clin. Infect. Dis.* **2005**, 40 (9), 1333.
- (9) Moore, R. E. *J. Ind. Microbiol.* **1996**, 16 (2), 134.
- (10) Moreau, R. A.; Whitaker, B. D.; Hicks, K. B. *Prog. Lipid Res.* **2002**, 41 (6), 457.
- (11) Micklefield, J. *Curr. Med. Chem.* **2001**, 8 (10), 1157.
- (12) Dragulescu-Andrasi, A.; Rapireddy, S.; Frezza, B. M.; Gayathri, C.; Gil, R. R.; Ly, D. H. *J. Am. Chem. Soc.* **2006**, 128 (31), 10258.
- (13) Sharma, V. K.; Sharma, R. K.; Singh, S. K. *Med Chem Commun* **2014**, 5 (10), 1454.
- (14) Kazmaier, U.; Deska, J. *Curr. Org. Chem.* **2008**, 12 (5), 355.
- (15) Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, 90 (4), 543.
- (16) Gao, W. Y.; Storm, C.; Egan, W.; Cheng, Y. C. *Mol. Pharmacol.* **1993**, 43 (1), 45.
- (17) Iyer, R. P.; Egan, W.; Regan, J. B.; Beaucage, S. L. *J. Am. Chem. Soc.* **1990**, 112 (3), 1253.
- (18) Singh, S.; Koshkin, A.; others. *Chem. Commun.* **1998**, No. 4, 455.
- (19) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. *Tetrahedron* **1998**, 54, 3607.
- (20) Summerton, J.; Weller, D. *Antisense Nucleic Acid Drug Dev.* **1997**, 7 (3), 187.
- (21) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, 114 (5), 1895.
- (22) Hyrup, B.; Nielsen, P. E. *Bioorg. Med. Chem. Lett.* **1996**, 4, 5.
- (23) Wittung, P.; Kajanus, J.; Edwards, K.; Nielsen, P.; Nordén, B.; Malmström, B. G. *FEBS Lett.* **1995**, 365, 27.
- (24) Koppelhus, U.; Nielsen, P. E. *Adv. Drug Deliv. Rev.* **2003**, 55 (2), 267.
- (25) Thomson, S. A.; Josey, J. A.; Cadilla, R.; Gaul, M. D.; Hassman, C. F.; Luzzio, M. J.; Pipe, A. J.; Reed, K. L.; Ricca, D. J.; Wiethe, R. W.; Noble, S. A. *Tetrahedron* **1995**, 51, 6179.
- (26) Kovács, G.; Timár, Z.; Kupihár, Z.; Kele, Z.; Kovács, L. *J. Chem. Soc. [Perkin 1]* **2002**, No. 10, 1266.
- (27) G. Breipohl; J. Knolle; D. Langer; G. O'Malley; E. Uhlmann. *Bioorg. Med. Chem. Lett.* **1996**, 6, 665.
- (28) Casale, R.; Jensen, I.; Egholm, M. *Pept. Nucleic Acids* **1999**, 39.
- (29) Wojciechowski, F.; Hudson, R. H. E. *J. Org. Chem.* **2008**, 73 (10), 3807.

- (30) Zhou, P.; Wang, M.; Du, L.; Fisher, G. W.; Waggoner, A.; Ly, D. H. *J. Am. Chem. Soc.* **2003**, *125* (23), 6878.
- (31) Sahu, B.; Chenna, V.; Lathrop, K. L.; Thomas, S. M.; Zon, G.; Livak, K. J.; Ly, D. H. *J. Org. Chem.* **2009**, *74* (4), 1509.
- (32) Sahu, B.; Sacui, I.; Rapireddy, S.; Zanotti, K. J.; Bahal, R.; Armitage, B. A.; Ly, D. H. *J. Org. Chem.* **2011**, *76* (14), 5614.
- (33) Dawson, D. W.; Volpert, O. V.; Pearce, S. F. A.; Schneider, A. J.; Silverstein, R. L.; Henkin, J.; Bouck, N. P. *Mol. Pharmacol.* **1999**, *55* (2), 332.
- (34) Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. *Acc. Chem. Res.* **2008**, *41* (10), 1331.
- (35) Fernández-Llamazares, A. I.; Spengler, J.; Albericio, F. *Pept. Sci.* **2015**, *104* (5), 435.
- (36) Miller, S. M.; Simon, R. J.; Ng, S.; Zuckermann, R. N.; Kerr, J. M.; Moos, W. H. *Drug Dev. Res.* **1995**, *35* (1), 20.
- (37) Kwon, Y.-U.; Kodadek, T. *J. Am. Chem. Soc.* **2007**, *129* (6), 1508.
- (38) Zuckermann, R. N. *Pept. Sci.* **2011**, *96* (5), 545.
- (39) Ornes, S. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (34), 13695.
- (40) Saber, H.; Leighton, J. K. *Regul. Toxicol. Pharmacol.* **2015**, *71* (3), 444.
- (41) Jevševar, S.; Kunstelj, M.; Porekar, V. G. *Biotechnol. J.* **2010**, *5* (1), 113.
- (42) Alconcel, S. N. S.; Baas, A. S.; Maynard, H. D. *Polym. Chem.* **2011**, *2* (7), 1442.
- (43) Pelegri-O'Day, E. M.; Lin, E.-W.; Maynard, H. D. *J. Am. Chem. Soc.* **2014**, *136* (41), 14323.
- (44) Kirschberg, T. A.; VanDeusen, C. L.; Rothbard, J. B.; Yang, M.; Wender, P. A. *Org. Lett.* **2003**, *5* (19), 3459.
- (45) Liang, J. F.; Yang, V. C. *Bioorg. Med. Chem. Lett.* **2005**, *15* (22), 5071.
- (46) Santra, S.; Yang, H.; Dutta, D.; Stanley, J. T.; Holloway, P. H.; Tan, W.; Moudgil, B. M.; Mericle, R. A. *Chem. Commun.* **2004**, No. 24, 2810.
- (47) Santra, S.; Yang, H.; Stanley, J. T.; Holloway, P. H.; Moudgil, B. M.; Walter, G.; Mericle, R. A. *Chem. Commun.* **2005**, No. 25, 3144.
- (48) Good, L.; Awasthi, S. K.; Dryselius, R.; Larsson, O.; Nielsen, P. E. *Nat. Biotechnol.* **2001**, *19* (4), 360.
- (49) Lewis, M. R.; Jia, F.; Gallazzi, F.; Wang, Y.; Zhang, J.; Shenoy, N.; Lever, S. Z.; Hannink, M. *Bioconjugate Chem.* **2002**, *13* (6), 1176.
- (50) Vrudhula, V. M.; MacMaster, J. F.; Li, Z.; Kerr, D. E.; Senter, P. D. *Bioorg. Med. Chem. Lett.* **2002**, *12* (24), 3591.
- (51) Selvaraj, R.; Giglio, B.; Liu, S.; Wang, H.; Wang, M.; Yuan, H.; Chintala, S. R.; Yap, L.-P.; Conti, P. S.; Fox, J. M.; Li, Z. *Bioconjugate Chem.* **2015**, *26* (3), 435.

- (52) Soriano del Amo, D.; Wang, W.; Jiang, H.; Besanceney, C.; Yan, A. C.; Levy, M.; Liu, Y.; Marlow, F. L.; Wu, P. *J. Am. Chem. Soc.* **2010**, *132* (47), 16893.
- (53) Marchán, V.; Ortega, S.; Pulido, D.; Pedroso, E.; Grandas, A. *Nucleic Acids Res.* **2006**, *34* (3), e24.
- (54) Kuył-Yeheskiely, E.; Tromp, C. M.; Lefeber, A. W. M.; van der Marel, G. A.; van Boom, J. H. *Tetrahedron* **1988**, *44* (20), 6515.
- (55) Wenska, M.; Alvira, M.; Steunenberg, P.; Stenberg, A.; Murtola, M.; Stromberg, R. *Nucleic Acids Res.* **2011**, *39* (20), 9047.
- (56) Zaramella, S.; Yeheskiely, E.; Strömberg, R. *J. Am. Chem. Soc.* **2004**, *126* (43), 14029.
- (57) Han, T. H.; Zhao, B. *Drug Metab. Dispos.* **2014**, *42* (11), 1914.
- (58) Shen, B.-Q.; Xu, K.; Liu, L.; Raab, H.; Bhakta, S.; Kenrick, M.; Parsons-Reponte, K. L.; Tien, J.; Yu, S.-F.; Mai, E.; Li, D.; Tibbitts, J.; Baudys, J.; Saad, O. M.; Scales, S. J.; McDonald, P. J.; Hass, P. E.; Eigenbrot, C.; Nguyen, T.; Solis, W. A.; Fujii, R. N.; Flagella, K. M.; Patel, D.; Spencer, S. D.; Khawli, L. A.; Ebens, A.; Wong, W. L.; Vandlen, R.; Kaur, S.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Junutula, J. R. *Nat. Biotechnol.* **2012**, *30* (2), 184.
- (59) Strop, P.; Liu, S.-H.; Dorywalska, M.; Delaria, K.; Dushin, R. G.; Tran, T.-T.; Ho, W.-H.; Farias, S.; Casas, M. G.; Abdiche, Y.; Zhou, D.; Chandrasekaran, R.; Samain, C.; Loo, C.; Rossi, A.; Rickert, M.; Krimm, S.; Wong, T.; Chin, S. M.; Yu, J.; Dilley, J.; Chaparro-Riggers, J.; Filzen, G. F.; O'Donnell, C. J.; Wang, F.; Myers, J. S.; Pons, J.; Shelton, D. L.; Rajpal, A. *Chem. Biol.* **2013**, *20* (2), 161.
- (60) Hamblett, K. J.; Senter, P. D.; Chace, D. F.; Sun, M. M. C.; Lenox, J.; Cerveny, C. G.; Kissler, K. M.; Bernhardt, S. X.; Kopcha, A. K.; Zabinski, R. F.; Meyer, D. L.; Francisco, J. A. *Clin. Cancer Res.* **2004**, *10* (20), 7063.
- (61) Melchert, M.; List, A. *Int. J. Biochem. Cell Biol.* **2007**, *39* (7-8), 1489.
- (62) Gaglione, M.; Di Fabio, G.; Messere, A. *Curr. Org. Chem.* **2012**, *16* (11), 1371.
- (63) Clivio, P.; Coantic-Castex, S.; Guillaume, D. *Chem. Rev.* **2013**, *113* (10), 7354.
- (64) Luo, J.; Abrahams, J. P. *Chem. – Eur. J.* **2014**, *20* (9), 2410.
- (65) Hill, T. A.; Shepherd, N. E.; Diness, F.; Fairlie, D. P. *Angew. Chem. Int. Ed.* **2014**, *53* (48), 13020.
- (66) Mogi, T.; Kita, K. *Cell. Mol. Life Sci.* **2009**, *66* (23), 3821.
- (67) Guo, J.; Hu, H.; Zhao, Q.; Wang, T.; Zou, Y.; Yu, S.; Wu, Q.; Guo, Z. *ChemMedChem* **2012**, *7* (8), 1496.
- (68) Pfaff, M.; Tangemann, K.; Müller, B.; Gurrath, M.; Müller, G.; Kessler, H.; Timpl, R.; Engel, J. *J. Biol. Chem.* **1994**, *269* (32), 20233.

- (69) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004.
- (70) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41* (14), 2596.
- (71) Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67* (9), 3057.
- (72) Worrell, B. T.; Malik, J. A.; Fokin, V. V. *Science* **2013**, *340* (6131), 457.
- (73) Jin, L.; Tolentino, D. R.; Melaimi, M.; Bertrand, G. *Sci. Adv.* **2015**, *1* (5), e1500304.
- (74) Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V. *J. Am. Chem. Soc.* **2005**, *127* (1), 210.
- (75) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. *Org. Lett.* **2004**, *6* (17), 2853.
- (76) Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125* (37), 11164.
- (77) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. *Bioconjugate Chem.* **2010**, *21* (10), 1912.
- (78) Besanceney-Webler, C.; Jiang, H.; Zheng, T.; Feng, L.; Soriano del Amo, D.; Wang, W.; Klivansky, L. M.; Marlow, F. L.; Liu, Y.; Wu, P. *Angew. Chem. Int. Ed.* **2011**, *50* (35), 8051.
- (79) Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y. *Angew. Chem. Int. Ed.* **2012**, *51* (24), 5852.
- (80) Jiang, H.; Zheng, T.; Lopez-Aguilar, A.; Feng, L.; Kopp, F.; Marlow, F. L.; Wu, P. *Bioconjugate Chem.* **2014**, *25* (4), 698.
- (81) Bevilacqua, V.; King, M.; Chaumontet, M.; Nothisen, M.; Gabillet, S.; Buisson, D.; Puente, C.; Wagner, A.; Taran, F. *Angew. Chem. Int. Ed.* **2014**, *53* (23), 5872.
- (82) Cai, Z.; Li, B. T. Y.; Wong, E. H.; Weisman, G. R.; Anderson, C. J. *Dalton Trans* **2015**, *44* (9), 3945.
- (83) Roberts, M. P.; Pham, T. Q.; Doan, J.; Jiang, C. D.; Hambley, T. W.; Greguric, I.; Fraser, B. H. *J. Label. Compd. Radiopharm.* **2015**, *58* (13-14), 473.
- (84) Poon, C. K.; Chapman, R.; Jolliffe, K. A.; Perrier, S. *Polym. Chem.* **2012**, *3* (7), 1820.
- (85) Willibald, J.; Harder, J.; Sparrer, K.; Conzelmann, K.-K.; Carell, T. *J. Am. Chem. Soc.* **2012**, *134* (30), 12330.
- (86) Meyer, A.; Noël, M.; Vasseur, J.-J.; Morvan, F. *Eur. J. Org. Chem.* **2015**, *2015* (13), 2921.
- (87) Speers, A. E.; Adam, G. C.; Cravatt, B. F. *J. Am. Chem. Soc.* **2003**, *125* (16), 4686.

- (88) Bateman, L. A.; Zaro, B. W.; Miller, S. M.; Pratt, M. R. *J. Am. Chem. Soc.* **2013**, *135* (39), 14568.
- (89) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126* (46), 15046.
- (90) Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G.-J. *Angew. Chem. Int. Ed.* **2008**, *47* (12), 2253.
- (91) Kuzmin, A.; Poloukhine, A.; Wolfert, M. A.; Popik, V. V. *Bioconjugate Chem.* **2010**, *21* (11), 2076.
- (92) Friscourt, F.; Fahrni, C. J.; Boons, G.-J. *Chem. - Eur. J.* **2015**, *21* (40), 13996.
- (93) Bernardin, A.; Cazet, A.; Guyon, L.; Delannoy, P.; Vinet, F.; Bonnaffé, D.; Texier, I. *Bioconjugate Chem.* **2010**, *21* (4), 583.
- (94) Sachin, K.; Jadhav, V. H.; Kim, E.-M.; Kim, H. L.; Lee, S. B.; Jeong, H.-J.; Lim, S. T.; Sohn, M.-H.; Kim, D. W. *Bioconjugate Chem.* **2012**, *23* (8), 1680.
- (95) Chen, K.; Wang, X.; Lin, W.-Y.; Shen, C. K.-F.; Yap, L.-P.; Hughes, L. D.; Conti, P. S. *ACS Med. Chem. Lett.* **2012**, *3* (12), 1019.
- (96) Marks, I. S.; Kang, J. S.; Jones, B. T.; Landmark, K. J.; Cleland, A. J.; Taton, T. A. *Bioconjugate Chem.* **2011**, *22* (7), 1259.
- (97) Lutz, J.-F. *Angew. Chem. Int. Ed.* **2008**, *47* (12), 2182.
- (98) van Geel, R.; Pruijn, G. J. M.; van Delft, F. L.; Boelens, W. C. *Bioconjugate Chem.* **2012**, *23* (3), 392.
- (99) Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287* (5460), 2007.
- (100) Hang, H. C.; Geutjes, E.-J.; Grotenbreg, G.; Pollington, A. M.; Bijlmakers, M. J.; Ploegh, H. L. *J. Am. Chem. Soc.* **2007**, *129* (10), 2744.
- (101) Vocadlo, D. J.; Hang, H. C.; Kim, E.-J.; Hanover, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci.* **2003**, *100* (16), 9116.
- (102) Tsao, M.-L.; Tian, F.; Schultz, P. G. *ChemBioChem* **2005**, *6* (12), 2147.
- (103) Weisbrod, S. H.; Marx, A. *Chem. Commun.* **2007**, No. 18, 1828.
- (104) Sletten, E. M.; Bertozzi, C. R. *Acc. Chem. Res.* **2011**, *44* (9), 666.
- (105) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. *Org. Lett.* **2000**, *2* (14), 2141.
- (106) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2000**, *2* (13), 1939.
- (107) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2001**, *3* (1), 9.
- (108) Soellner, M. B.; Nilsson, B. L.; Raines, R. T. *J. Am. Chem. Soc.* **2006**, *128* (27), 8820.
- (109) Nilsson, B. L.; Hondal, R. J.; Soellner, M. B.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125* (18), 5268.
- (110) Kleineweischede, R.; Hackenberger, C. P. R. *Angew. Chem. Int. Ed.* **2008**, *47* (32), 5984.

- (111) Soellner, M. B.; Dickson, K. A.; Nilsson, B. L.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125* (39), 11790.
- (112) Bianchi, A.; Bernardi, A. *J. Org. Chem.* **2006**, *71* (12), 4565.
- (113) Pretze, M.; Wuest, F.; Peppel, T.; Köckerling, M.; Mamat, C. *Tetrahedron Lett.* **2010**, *51* (49), 6410.
- (114) Soellner, M. B.; Tam, A.; Raines, R. T. *J. Org. Chem.* **2006**, *71* (26), 9824.
- (115) Weisbrod, S.; Marx, A. *Synlett* **2010**, *2010* (05), 787.
- (116) Tam, A.; Raines, R. T. *Bioorg. Med. Chem.* **2009**, *17* (3), 1055.
- (117) Nair, D. P.; Podgórski, M.; Chatani, S.; Gong, T.; Xi, W.; Fenoli, C. R.; Bowman, C. N. *Chem. Mater.* **2014**, *26* (1), 724.
- (118) Chan, J. W.; Hoyle, C. E.; Lowe, A. B.; Bowman, M. *Macromolecules* **2010**, *43* (15), 6381.
- (119) Northrop, B. H.; Frayne, S. H.; Choudhary, U. *Polym Chem* **2015**, *6* (18), 3415.
- (120) Dadová, J.; Orság, P.; Pohl, R.; Brázdová, M.; Fojta, M.; Hocek, M. *Angew. Chem. Int. Ed.* **2013**, *52* (40), 10515.
- (121) Eberhard, H.; Diezmann, F.; Seitz, O. *Angew. Chem. Int. Ed.* **2011**, *50* (18), 4146.
- (122) Ni, J.; Singh, S.; Wang, L.-X. *Bioconjugate Chem.* **2003**, *14* (1), 232.
- (123) Wängler, C.; Maschauer, S.; Prante, O.; Schäfer, M.; Schirmacher, R.; Bartenstein, P.; Eisenhut, M.; Wängler, B. *ChemBioChem* **2010**, *11* (15), 2168.
- (124) Gao, G.; Yu, K.; Kindrachuk, J.; Brooks, D. E.; Hancock, R. E. W.; Kizhakkedathu, J. N. *Biomacromolecules* **2011**, *12* (10), 3715.
- (125) Sánchez, A.; Pedroso, E.; Grandas, A. *Org. Lett.* **2011**, *13* (16), 4364.
- (126) Sánchez, A.; Pedroso, E.; Grandas, A. *Bioconjugate Chem.* **2012**, *23* (2), 300.
- (127) Sánchez, A.; Pedroso, E.; Grandas, A. *Chem Commun* **2013**, *49* (3), 309.
- (128) Elduque, X.; Pedroso, E.; Grandas, A. *Org. Lett.* **2013**, *15* (8), 2038.
- (129) Elduque, X.; Pedroso, E.; Grandas, A. *J. Org. Chem.* **2014**, *79* (7), 2843.
- (130) Saito, G.; Swanson, J. A.; Lee, K.-D. *Adv. Drug Deliv. Rev.* **2003**, *55* (2), 199.
- (131) Lu, K.; Duan, Q.-P.; Ma, L.; Zhao, D.-X. *Bioconjugate Chem.* **2010**, *21* (2), 187.
- (132) Kellogg, B. A.; Garrett, L.; Kovtun, Y.; Lai, K. C.; Leece, B.; Miller, M.; Payne, G.; Steeves, R.; Whiteman, K. R.; Widdison, W.; Xie, H.; Singh, R.; Chari, R. V. J.; Lambert, J. M.; Lutz, R. J. *Bioconjugate Chem.* **2011**, *22* (4), 717.
- (133) Wu, C.; Wang, S.; Brülisauer, L.; Leroux, J.-C.; Gauthier, M. A. *Biomacromolecules* **2013**, *14* (7), 2383.
- (134) Wu, C.; Belenda, C.; Leroux, J.-C.; Gauthier, M. A. *Chem. – Eur. J.* **2011**, *17* (36), 10064.

- (135) Antopolsky, M.; Azhayeva, E.; Tengvall, U.; Auriola, S.; Jääskeläinen, I.; Rönkkö, S.; Honkakoski, P.; Urtti, A.; Lönnberg, H.; Azhayev, A. *Bioconjugate Chem.* **1999**, *10* (4), 598.
- (136) Turner, J. J.; Arzumanov, A. A.; Gait, M. J. *Nucleic Acids Res.* **2005**, *33* (1), 27.
- (137) Hirosue, S.; Kourtis, I. C.; van der Vlies, A. J.; Hubbell, J. A.; Swartz, M. A. *Vaccine* **2010**, *28* (50), 7897.
- (138) Duvall, C. L.; Convertine, A. J.; Benoit, D. S. W.; Hoffman, A. S.; Stayton, P. S. *Mol. Pharm.* **2010**, *7* (2), 468.
- (139) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266* (5186), 776.
- (140) Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chem. Int. Ed.* **2008**, *47* (52), 10030.
- (141) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128* (20), 6640.
- (142) Thompson, R. E.; Chan, B.; Radom, L.; Jolliffe, K. A.; Payne, R. J. *Angew. Chem. Int. Ed.* **2013**, *52* (37), 9723.
- (143) Malins, L. R.; Cergol, K. M.; Payne, R. J. *ChemBioChem* **2013**, *14* (5), 559.
- (144) Malins, L. R.; Cergol, K. M.; Payne, R. J. *Chem. Sci.* **2013**, *5* (1), 260.
- (145) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. *J. Am. Chem. Soc.* **2009**, *131* (38), 13592.
- (146) Mende, F.; Seitz, O. *Angew. Chem. Int. Ed.* **2011**, *50* (6), 1232.
- (147) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *J. Am. Chem. Soc.* **1997**, *119* (19), 4325.
- (148) Dang, B.; Kubota, T.; Mandal, K.; Bezanilla, F.; Kent, S. B. H. *J. Am. Chem. Soc.* **2013**, *135* (32), 11911.
- (149) Diezmann, F.; Eberhard, H.; Seitz, O. *Pept. Sci.* **2010**, *94* (4), 397.
- (150) Geiermann, A.-S.; Polacek, N.; Micura, R. *J. Am. Chem. Soc.* **2011**, *133* (47), 19068.
- (151) Dendane, N.; Melnyk, O.; Xu, T.; Grandidier, B.; Boukherroub, R.; Stiévenard, D.; Coffinier, Y. *Langmuir* **2012**, *28* (37), 13336.
- (152) Sohma, Y.; Pentelute, B. L.; Whittaker, J.; Hua, Q.; Whittaker, L. J.; Weiss, M. A.; Kent, S. B. H. *Angew. Chem. Int. Ed.* **2008**, *47* (6), 1102.
- (153) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci.* **1999**, *96* (18), 10068.
- (154) Wan, Q.; Chen, J.; Yuan, Y.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2008**, *130* (47), 15814.
- (155) Shao, J.; Tam, J. P. *J. Am. Chem. Soc.* **1995**, *117* (14), 3893.
- (156) Zhang, L.; Tam, J. P. *Anal. Biochem.* **1996**, *233* (1), 87.

- (157) Butvin, P.; Al-Ja'afreh, J.; Světlík, J.; Havranek, E. *Chem. Pap.* **1999**, *53* (5), 315.
- (158) Botti, P.; Pallin, T. D.; Tam, J. P. *J. Am. Chem. Soc.* **1996**, *118* (42), 10018.
- (159) Casi, G.; Huguenin-Dezot, N.; Zuberbühler, K.; Scheuermann, J.; Neri, D. *J. Am. Chem. Soc.* **2012**, *134* (13), 5887.
- (160) Zatsepin, T. S.; Stetsenko, D. A.; Arzumanov, A. A.; Romanova, E. A.; Gait, M. J.; Oretskaya, T. S. *Bioconjugate Chem.* **2002**, *13* (4), 822.
- (161) Ghosh, S.; Defrancq, E.; Lhomme, J. H.; Dumy, P.; Bhattacharya, S. *Bioconjugate Chem.* **2004**, *15* (3), 520.
- (162) Kalia, J.; Raines, R. T. *Angew. Chem. Int. Ed.* **2008**, *47* (39), 7523.
- (163) Jencks, W. P. *J. Am. Chem. Soc.* **1959**, *81* (2), 475.
- (164) Mezö, G.; Szabó, I.; Kertész, I.; Hegedüs, R.; Orbán, E.; Leurs, U.; Bösze, S.; Halmos, G.; Manea, M. *J. Pept. Sci.* **2011**, *17* (1), 39.
- (165) Marcaurette, L. A.; Shin, Y.; Goon, S.; Bertozzi, C. R. *Org. Lett.* **2001**, *3* (23), 3691.
- (166) Iyer, G.; Pinaud, F.; Xu, J.; Ebenstein, Y.; Li, J.; Chang, J.; Dahan, M.; Weiss, S. *Bioconjugate Chem.* **2011**, *22* (6), 1006.
- (167) Roberts, K. D.; Lambert, J. N.; Ede, N. J.; Bray, A. M. *J. Pept. Sci.* **2004**, *10* (11), 659.
- (168) Galibert, M.; Dumy, P.; Boturyn, D. *Angew. Chem. Int. Ed.* **2009**, *48* (14), 2576.
- (169) Dirksen, A.; Dawson, P. E. *Bioconjugate Chem.* **2008**, *19* (12), 2543.
- (170) Tang, W.; Becker, M. L. *Chem Soc Rev* **2014**, *43* (20), 7013.
- (171) Cavill, J. L.; Peters, J.-U.; Tomkinson, N. C. O. *Chem. Commun.* **2003**, No. 6, 728.
- (172) Fringuelli, F.; Piermatti, O.; Pizzo, F.; Vaccaro, L. *Eur. J. Org. Chem.* **2001**, *2001* (3), 439.
- (173) Chung, Y. S.; Duerr, B. F.; Nanjappan, P.; Czarnik, A. W. *J. Org. Chem.* **1988**, *53* (6), 1334.
- (174) Boutelle, R. C.; Northrop, B. H. *J. Org. Chem.* **2011**, *76* (19), 7994.
- (175) Hill, K. W.; Taunton-Rigby, J.; Carter, J. D.; Kropp, E.; Vagle, K.; Pieken, W.; McGee, D. P. C.; Husar, G. M.; Leuck, M.; Anziano, D. J.; Sebesta, D. P. *J. Org. Chem.* **2001**, *66* (16), 5352.
- (176) Borsenberger, V.; Howorka, S. *Nucleic Acids Res.* **2009**, *37* (5), 1477.
- (177) Pozsgay, V.; Vieira, N. E.; Yergey, A. *Org. Lett.* **2002**, *4* (19), 3191.
- (178) Tischer, T.; Claus, T. K.; Bruns, M.; Trouillet, V.; Linkert, K.; Rodriguez-Emmenegger, C.; Goldmann, A. S.; Perrier, S.; Börner, H. G.; Barner-Kowollik, C. *Biomacromolecules* **2013**, *14* (12), 4340.

- (179) Blackman, M. L.; Royzen, M.; Fox, J. M. *J. Am. Chem. Soc.* **2008**, *130* (41), 13518.
- (180) Venkateswara Rao, B.; Dhokale, S.; Rajamohanan, P. R.; Hotha, S. *Chem. Commun.* **2013**, *49* (92), 10808.
- (181) Hörner, S.; Uth, C.; Avrutina, O.; Frauendorf, H.; Wiessler, M.; Kolmar, H. *Chem Commun* **2015**, *51* (55), 11130.
- (182) Rossin, R.; van Duijnhoven, S. M. J.; Läppchen, T.; van den Bosch, S. M.; Robillard, M. S. *Mol. Pharm.* **2014**, *11* (9), 3090.
- (183) Schoch, J.; Staudt, M.; Samanta, A.; Wiessler, M.; Jäschke, A. *Bioconjugate Chem.* **2012**, *23* (7), 1382.
- (184) Yang, J.; Liang, Y.; Šečkutė, J.; Houk, K. N.; Devaraj, N. K. *Chem. – Eur. J.* **2014**, *20* (12), 3365.
- (185) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. *Nat. Chem.* **2012**, *4* (4), 298.
- (186) Karver, M. R.; Weissleder, R.; Hilderbrand, S. A. *Bioconjugate Chem.* **2011**, *22* (11), 2263.
- (187) Kamber, D. N.; Liang, Y.; Blizzard, R. J.; Liu, F.; Mehl, R. A.; Houk, K. N.; Prescher, J. A. *J. Am. Chem. Soc.* **2015**, *137* (26), 8388.
- (188) Späte, A.-K.; Schart, V. F.; Schöllkopf, S.; Niederwieser, A.; Wittmann, V. *Chem. - Eur. J.* **2014**, *20* (50), 16502.

Chapter 2. On-resin Diels-Alder reactions in water. Conjugates of diene-derivatised polyamides and diene-derivatised oligonucleotides

2.1 Conjugation of diene-derivatised polyamides

2.1.1 Background and objectives

As stated in chapter 1, the Diels-Alder (DA) reaction has attracted attention as a powerful conjugation method due to its water compatibility, orthogonality and no need for catalysts. For biomolecules that are normally synthesised by solid-phase synthesis, for example peptides and oligonucleotides, the most straightforward alternative to use this reaction is to introduce the diene or dienophile while they are still protected and linked to the solid matrix. Yet, the compatibility of these moieties with the cleavage and deprotection conditions of the oligomer has to be considered. While maleimides (the most common dienophile in bioconjugations) are stable to the acidic conditions required for peptide deprotection, they do not withstand the basic treatments necessary for oligonucleotide deprotection and Fmoc deprotection during SPPS. The opposite happens to dienes, which are stable to the final ammonia treatment of oligonucleotides but not to the TFA/scavengers mixture used with peptides. An obvious consequence is that the introduction of maleimides into oligonucleotides and of dienes into peptides is, in principle, precluded. As said in section 1.3.5, some years ago our group developed a methodology based on a DA-rDA sequence that permitted the introduction of maleimides

to oligonucleotides and to internal positions in polyamide chains. A logical step afterwards was, therefore, to find a methodology that allowed the conjugation of diene-derivatised peptides preventing any harm to the diene moiety by the final acidic cleavage and deprotection treatment.

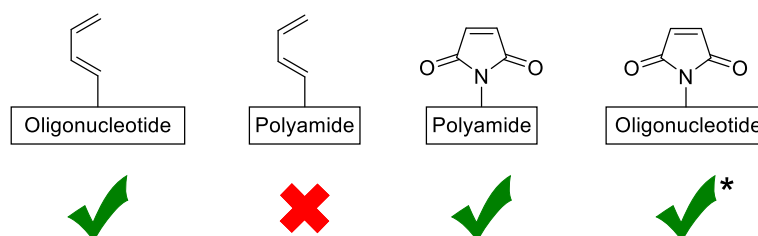


Figure 2.1. Derivatised oligomers for DA reactions that can be obtained by standard solid-phase synthesis. * As long as the maleimide is protected during the cleavage step.

2.1.2 Decomposition of dienes under the cleavage and deprotection conditions

Diene decomposition in the cleavage and deprotection step is due to the ease of carbocation formation and to the high concentration and strong acidity of the TFA employed in this treatment. Protonation of the conjugated diene gives rise to a resonance-stabilised carbocation, which can, afterwards, react with the nucleophilic scavengers typically present in the deprotection cocktail. Madder and co-workers have described the decomposition products arising from the cleavage of a furan-containing peptide that accompanied the desired product.¹ They basically observed products of partial or complete reduction, that result from the reaction with triisopropylsilane (TIS), products arising from hydration and ring-opening, and formation of thioacetals due to reaction with dithiothreitol (DTT) (see **Figure 2.2**).

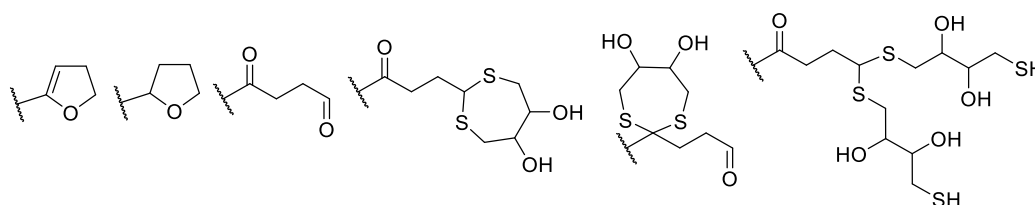


Figure 2.2. Furan decomposition products found by Madder's group upon acidic peptide deprotection.

We decided to evaluate by ourselves the effect of a typical TFA/TIS/H₂O deprotection mixture on a non-aromatic 1,3-diene. For this purpose, the resin-linked, totally protected, diene-modified peptide **2.4** (see section 2.1.8) was treated for two hours with a 95:2.5:2.5 TFA/TIS/H₂O mixture and analysed after a simple work-up aimed to remove any non-peptidic impurity. HPLC/MS Traces showed no presence of the desired product, contrary to that described by Madder, suggesting that the aromaticity of furan might help avoid these secondary reactions (**Figure 2.3**).

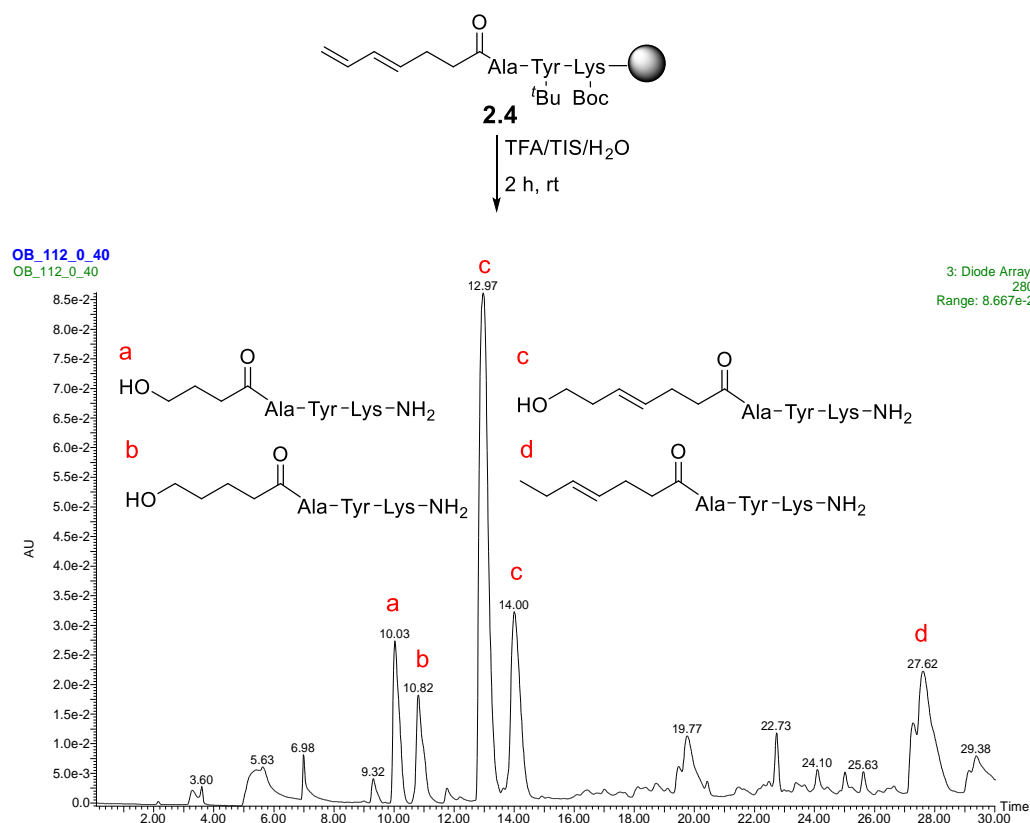
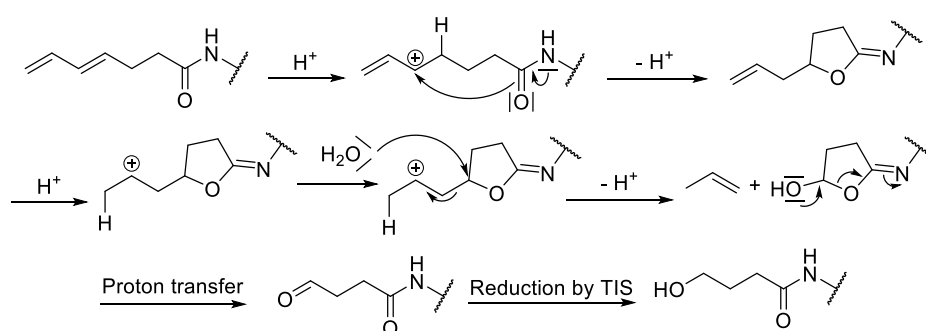


Figure 2.3. HPLC/MS Trace (280 nm) showing the decomposition products of **2.4** after acidic treatment. Regioselectivity of the hydrolysis and partial reduction products c and d has not been determined; only one of the possible regioisomers for each product is depicted here.

As expected from Madder's report, we observed partial reduction products (d in **Figure 2.3**), species arising from double bond hydration (c in **Figure 2.3**), and, quite unexpectedly, products generated by fragmentation and hydration (a and b in **Figure 2.3**). A plausible mechanism for their formation could start with the intramolecular attack of the amide carbonyl to the generated carbocation, rendering a five- or six-membered ring, as depicted in **Scheme 2.1**. Nucleophilic attack of a water molecule on the heterocycle after double bond protonation would eliminate propene or ethene,

respectively, and generate, after rearrangement, a terminal aldehyde. This would finally be reduced by TIS to the terminal alcohol.²



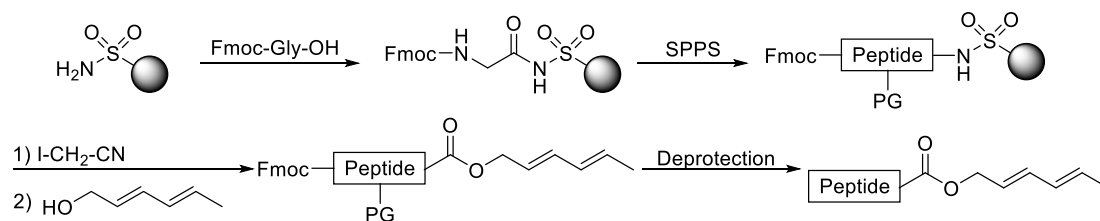
Scheme 2.1. Plausible mechanism for the formation of by-product a in **Figure 2.3**.

The formation of subproduct b is analogous, rendering a six-membered ring and extruding ethane.

2.1.3 Described alternatives to circumvent diene decomposition

Several groups have tried to circumvent the problems imposed by the inherent instability of the diene moiety to the standard acidic conditions used in SPPS. The group of Kahn used 1,3-dienes conjugated to electron withdrawing groups to avoid its decomposition,³ but they are far less reactive than the unstable alkyl-substituted 1,3-dienes and therefore not so interesting.

Waldmann and co-workers employed a completely different strategy that permitted to obtain the desired diene-derivatised peptides.⁴ Their methodology consisted in the use of a nucleophile-cleavable peptide-resin linkage, which allowed to simultaneously cleave the peptide-resin linkage and to introduce the diene by using 2,4-hexadien-1-ol with 4-(dimethylamino)pyridine catalysis. Peptide protecting groups were chosen so that they could be removed under basic or mild-acidic conditions in order not to harm the diene moiety. Specifically, a sulfonamide resin was used, and it was activated with iodoacetonitrile prior to reaction with 2,4-hexadien-1-ol to allow for the nucleophilic bond cleavage (see **Scheme 2.2**). Regarding the protecting groups, the base labile Fmoc group and the ultra acid-labile Trt and Mtt groups were used for internal alcohol and amine protection.



Scheme 2.2. Strategy employed by the Waldmann group to obtain diene-derivatised peptides.

The technique presents several problems, although it allowed for the synthesis of diene-derivatised peptides. First, most of the trifunctional amino acids with this protecting group scheme must be synthesised as they are not commercially available. More importantly, the sulfonamide linker imposes an important sequence limitation. Due to the poor nucleophilicity of sulfonamides, the coupling of the first amino acid needs to be extended for several hours, which may result in racemization. To avoid this problem, the first amino acid introduced when using this methodology must be glycine in all cases. Finally, instead of the typical simultaneous cleavage and deprotection, this strategy requires a two-step procedure to achieve the cleaved and totally unprotected peptide. Furthermore, the cleavage step is also composed of two different processes: an activation step, which requires more than 18 h, and the actual scission of the peptide from the resin, which needs 24 h. This may result in a tedious process lasting more than 2 days, when the standard methodology yields the cleaved and totally unprotected peptide in just 2 h. Finally, the in principle ultra-labile Mtt group could not be removed using the standard 1:5:94 TFA/TIS/DCM treatment, and, as higher percentages of TFA caused degradation of the diene, it had to be left in the final peptide.

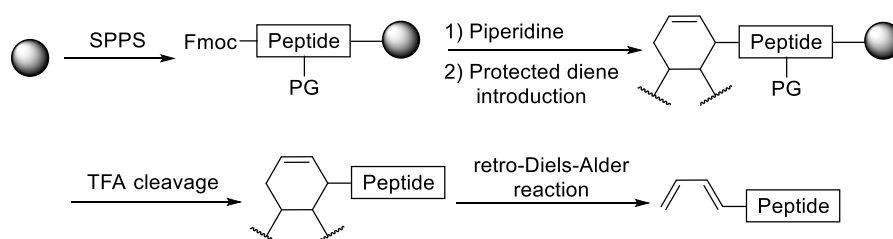
Madder and co-workers published two papers dealing with the DA reaction of furan-derivatised peptides. In one of them,¹ they observed that decomposition of the furan ring only happened when it was in the *N*-terminal position, probably because the more sterically congested internal positions shielded it. Surprisingly, though, this problem could be solved by capping the *N*-terminal position with an aromatic moiety, which was thought to protect the furan ring from degradation through π - π stacking interactions. In the other article,⁵ the same group described the DA reaction of a resin-linked, furan-derivatised peptide with *N*-phenylmaleimide in toluene. Conversions of ca. 85 % were obtained, but complete conversions were impossible to achieve due to the reversible nature of the reaction. Indeed, the labelled peptide was found to completely revert to the original diene and dienophile when heated at 70 °C for 24 hours, a result consistent with the described stability of this type of adducts.⁶ Nevertheless, this group did not provide

a single example in which an actual conjugation reaction was performed, as only the explained model reactions were described. Moreover, the use of toluene as solvent did not seem adequate for our purposes, as biomolecules are not soluble in such medium.

2.1.4 Our proposal to solve the problem

Our first proposal to solve the above-exposed problem was to search for a protecting group for the diene moiety. Our strategy was to find a diene-dienophile couple that would generate a reversible, yet stable to the peptide deprotection conditions, DA adduct. This DA adduct would be introduced into the protected, resin-linked peptide, and, after cleavage and deprotection, be heated to yield a diene-modified peptide and a free dienophile through a rDA reaction (see **Scheme 2.3**). Of course, said DA adduct would need to show a proper balance between stability and heat-lability in order to be handled comfortably and reversed in mild conditions.

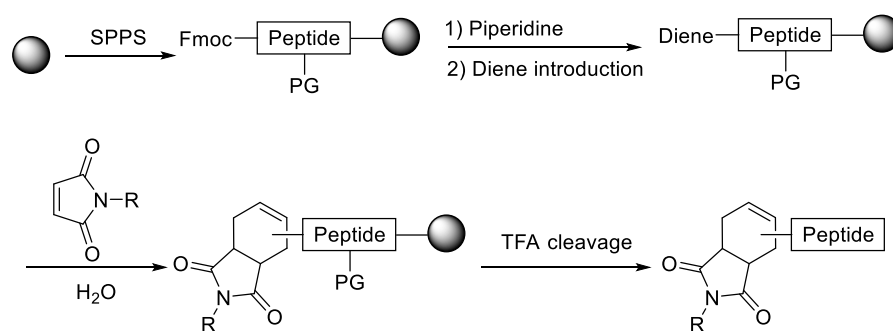
An intensive search into the literature did not provide any example of such a strategy in the field of bioconjugation, so we decided to run some assays. Several diene-dienophile partners were tested for stability to the cleavage and deprotection mixture and for their rDA reaction. Unluckily, all efforts made to find a combination that matched the above-mentioned requisites met only with failure. A more detailed explanation can be found in Appendix 1.



Scheme 2.3. Protection strategy for the incorporation of dienes to peptides through SPPS.

Our second idea for the conjugation of diene-modified peptides, after all our attempts to find an appropriate protecting group for a diene failed, was to perform the DA conjugation reaction on-resin. In this procedure, the diene would not be exposed to the acidic final treatment, as it would have already reacted with the dienophile to furnish the DA cycloadduct. This was expected to be less easily protonated than the conjugated 1,3-diene, as carbocations would not be resonance-stabilised, therefore preventing any further reactions with the scavengers present in the deprotection cocktail. Further, this strategy would not require the use on non-standard SPPS methods, overcoming the

limitations imposed by the Waldmann's methodology, and would allow for removal of any possible dienophile excess by a simple resin wash. It was clear that maleimides would be used as dienophiles, as they incorporate some of the most reactive double bonds and, also, a lot of common labelling agents, such as fluorophores, are commercially available with maleimide modifications. Some tests were performed to decide which diene would perform better (see section 2.1.8). Regarding the reaction conditions, parameters such as concentration, molar ratios and temperature should have to be optimised with model reactions. What we knew for sure was that we would use water as solvent for the on-resin DA reaction, something that, to our knowledge, had never been done before. The reasons for this solvent choice are that it is the best option to dissolve biomolecules and it strongly accelerates the reaction rate (see below). The procedure is schematically depicted in **Scheme 2.4**.



Scheme 2.4. Our strategy for the conjugation of diene-derivatised peptides.

2.1.5 The effect of water in Diels-Alder reactions

The rate enhancement effect of water in DA reactions was discovered in 1980 by Rideout and Breslow, who described reaction rates more than 700 times higher in water than in organic solvents.⁷ Since then, many groups have investigated this phenomenon, both experimentally and computationally, and three main reasons can be held responsible for the observed acceleration: the hydrophobic effect, an H-bond stabilizing effect and the high cohesive energy density (CED) of the solvent.

The hydrophobic effect is the tendency of nonpolar organic molecules dissolved in water to group together. This helps minimize their surface in contact with water, reducing the destabilizing interactions with this highly polar liquid and maintaining the water-water hydrogen-bond pattern as intact as possible. Furthermore, from an entropic point of view, the solvation of organic molecules by water is strongly unfavourable, as water molecules tend to orientate themselves tangential to the apolar solute surface. Aggregation of organic molecules excludes water molecules from their solvation shells, releasing them

to the more disordered bulk aqueous medium, therefore gaining entropy.⁸ The final consequence of this aggregation phenomenon is a higher effective concentration of the organic reagents, which provokes a rate acceleration. Nevertheless, the hydrophobic effect cannot be the only explanation for the observed effects, since intramolecular DA reactions, in which the reagents are already “packed”, are also largely enhanced by water.

The hydrogen-bonding capacity of water also plays a key role in the DA acceleration. Engberts and co-workers found out that the carbonyl groups typically present in electron-poor dienophiles tend to be more polarised in the DA transition state than in the initial state.⁹ This highly polarised states are stabilised by water through H-bonds, therefore lowering the energy of the transition state. The same conclusion was also reached by Jorgensen *et al.* computationally.¹⁰ An indirect evidence for this can be found in examples in which no H-bond accepting groups are present in the reagents. In these cases, the rate enhancement in water is not as pronounced as when polarizable groups are present in the molecule, as can be observed in **Figure 2.4**.¹¹ Of course, polar solvents and H-bonding solvents such as alcohols also tend to accelerate the DA reaction, although not as much as water.

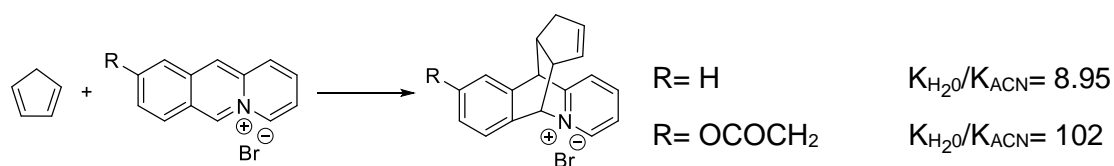


Figure 2.4. Water acceleration of DA reactions with differently substituted reagents.

Finally, there is another important factor that contributes to the acceleration observed in water, the Cohesive Energy Density (CED) of a solvent. This parameter is mathematically described by **Equation 2.1**, and, in qualitative terms, it is the energy necessary to separate a solvent molecule from its neighbours, which can be regarded as the energy necessary to create a cavity in the solvent.¹²

$$CED = \frac{\Delta H_{vap} - RT}{M_r/\rho}$$

Equation 2.1. Mathematical expression of the cohesive energy density. M_r is the relative molecular mass and ρ the density of the solvent.

Water has a high CED, which implies that the generation of a cavity in its core requires a lot of energy. It is due to this fact that the reagent molecules are as packed as possible, in the same way as if they were under the influence of an external pressure. As the DA reaction has a negative volume of activation, that is, the transition state is more compact than the separate reagents, any pressure that brings reagents together will lower the transition state energy.¹³ The fact that this mechanism is actually operating can be confirmed by the *endo* vs *exo* adduct enhanced selectivity: a high CED will favour the most compact transition state, which in the case of the DA reaction is the *endo* one.¹⁴ It is important to mention that all these effects are dramatically reduced when an organic co-solvent is added to the reaction mixture, even in small amounts. Blockzijl *et al.*¹⁵ have shown that increasing molar fractions of organic co-solvents increase the ΔG^\ddagger of the reaction, although in some cases molar fractions lower than 0.05 can induce small decreases in the activation Gibbs energy. The *endo/exo* selectivity is also diminished with increasing molar fractions of the organic co-solvent.

2.1.6 Selection of the solid support for the on-resin Diels-Alder reaction in water

Our choice of water as solvent also affected the choice of the polymeric support we were going to use for the synthesis of the diene-modified polyamide. Since we intended to perform the DA reaction on a resin-linked, diene-modified peptide in water, the use of divinylbenzene cross-linked polystyrene resins was precluded, as they do not swell in polar solvents. Mixed polystyrene/poly(ethyleneglycol) (PS/PEG) resins seemed more adequate, as they are water-compatible and swell in almost every solvent.¹⁶ The inherent low functionality of this kind of resins, which is sometimes viewed as a problem, was in our case an advantage. Low substitution degrees should decrease interchain interactions, thus reducing steric hindrance and swelling-related problems during both the synthesis and the conjugation of the polyamide.¹⁷ This would be particularly important in the conjugation step, where larger molecules, that are more sensitive to steric hindrance, would be attached to the peptide-resin.

Furthermore, this type of resins is nowadays commonly used for the solid-phase synthesis of polyamides, and is commercially available at competitive prices. Our decision was to use the commercial NovaSyn TGR and the TentaGel R RAM resins with the Rink-Amide linker, for the obtention of C-terminal carboxamide polyamides. These resins proved to be adequate for the DA reaction and improved the quality of both the peptides and PNAs synthesised.

2.1.7 Analysis of the reaction outcome

Regarding the conjugation reaction outcome, we decided to assess it by HPLC analyses of the crudes and MALDI-TOF MS of the collected peaks. Reaction yields given are based on the relative area of the product peak at 220 nm in the case of resin-linked peptides, since only at this wavelength the unreacted peptide can be properly quantified. For similar reasons, when working with resin-linked PNAs 260 nm was the wavelength of choice. In the cases where isolated yields are given, they are based on the ratio of the pure, isolated compound to the theoretical amount of product, calculated from the weighed resin and its substitution degree before the introduction of the diene. No attempts to separate the *exo* and *endo* isomers were made, as most of the times, and as expected for the size of the molecules, they co-eluted.

2.1.8 Selection of the diene

After having decided solvent, resin and the monitoring methodology, it was time to select an appropriate diene. Our idea was to use either furan or an alkyl substituted 1,3-diene (at that time the paper by Madder describing the reactivity of furans on solid-phase was still not published). As they needed to be attached to a peptide chain, it was convenient to have a proper carboxyl derivative, and therefore 3-(2-furyl)propanoic acid (**2.1**) and (*E*)-4,6-heptadienoic acid (**2.2**) (**Figure 2.5**) were selected as possible candidates. **2.2** Was chosen over the diene used by Waldmann and co-workers to reduce the number of possible stereoisomers after the DA reaction, although it was envisaged that it could be a little bit less reactive. The stability of the DA cycloadducts generated from maleimides and furans towards the peptide deprotection conditions was already known, and that of the cycloadducts generated from **2.2** and maleimides was confirmed in preliminary experiments (data not shown).

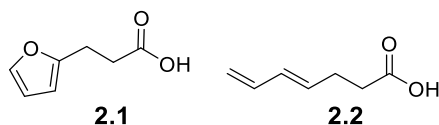
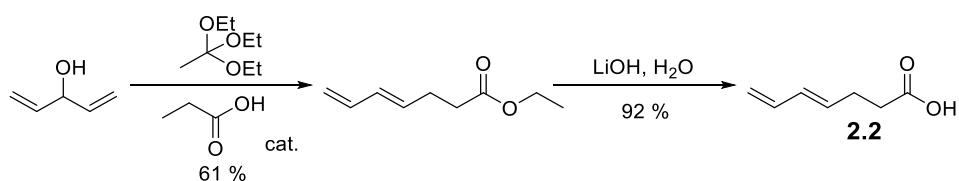


Figure 2.5. Dienes selected to be evaluated in the DA reaction.

Furan derivative **2.1** was commercially available, but **2.2** had to be synthesised as shown in **Scheme 2.5**. Basically, 1,4-pentadiene-3-ol was reacted with 8 equivalents of 1,1,1-triethoxyethane under acid catalysis to render ethyl (*E*)-4,6-heptadienoate, which was further hydrolysed using lithium hydroxide to yield the diene acid derivative **2.2**.¹⁸



Scheme 2.5. Synthetic scheme for the synthesis of diene **2.2**.

With both dienes in hand, it was time to test their reactivity and the stability of the generated DA adducts with a model maleimide. The resin-linked, protected Ala-Tyr(^tBu)-Lys(Boc) sequence was selected as a model peptide due to its water solubility and the fact that it contained a chromophore, which would facilitate the HPLC analyses of the crudes. As a model maleimide, the PNA dinucleotide t-t sequence was used, linked to a Lys residue to ensure solubility (**2.5**). Both **2.1** and **2.2** were attached to the peptide-resin, generating the diene-modified peptides **2.3** and **2.4**, respectively. These were left reacting for 70 h at 40 °C with PNA **2.5** (1.5 equiv., 20 mM concentration) to furnish the corresponding cycloadducts. Conjugates (**2.6** for **2.3** and **2.7** for **2.4**) were obtained after cleavage and deprotection using the typical 95:2.5:2.5 TFA/TIS/H₂O mixture, and analysed by HPLC and MALDI-TOF MS.

As can be seen from the HPLC traces (**Figure 2.6**), the crude from the **2.4** and **2.5** reaction, in which **2.2** was employed as diene, was cleaner than that involving use of the furan derivative **2.1**. While **2.4** rendered 89 % of the conjugate product (**2.7**, $t_R = 15.0$ min), with no major impurities present, the conjugate (**2.6**, $t_R = 12.1 + 12.7$ min) yield for **2.3** was less than 80 %. Partial and total reduction of the furan ring was found to account for approximately 18 % of the crude, with the partially reduced product ($t_R = 12.6$) almost co-eluting with one of the conjugate peaks. The fully reduced product ($t_R = 13.6$ min) could be perfectly separated. Curiously, in the case of **2.7** either only one adduct was generated or the *exo* and *endo* adducts co-eluted, whereas in **2.6** these adducts could be distinguished.

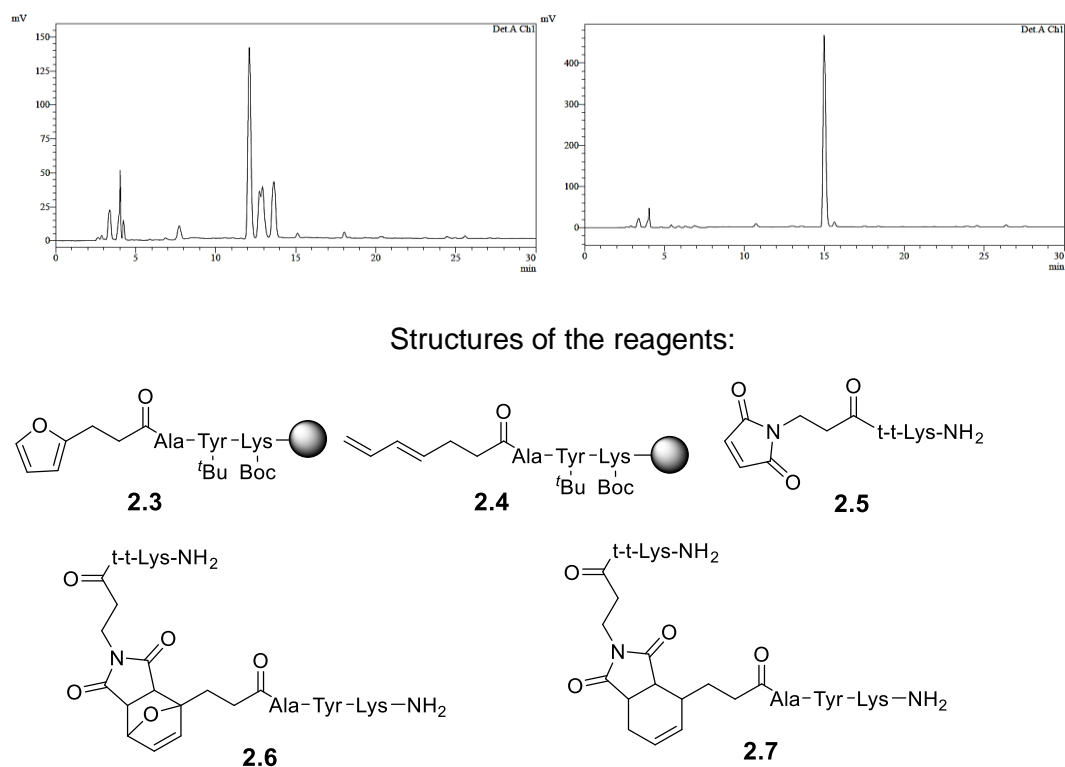


Figure 2.6. HPLC traces (220 nm) of the **2.3 + 2.5** reaction crude (left) and of the **2.4 + 2.5** reaction crude (right). The structures of the reagents are also shown.

Another drawback of the use of a furan derivative as a diene is the reversibility of the cycloaddition. When **2.6** was analysed by MALDI-TOF MS it was soon discovered that it had reverted almost completely to the starting materials (see **Figure 2.7**). On the contrary, this did not happen when analysing **2.7**.

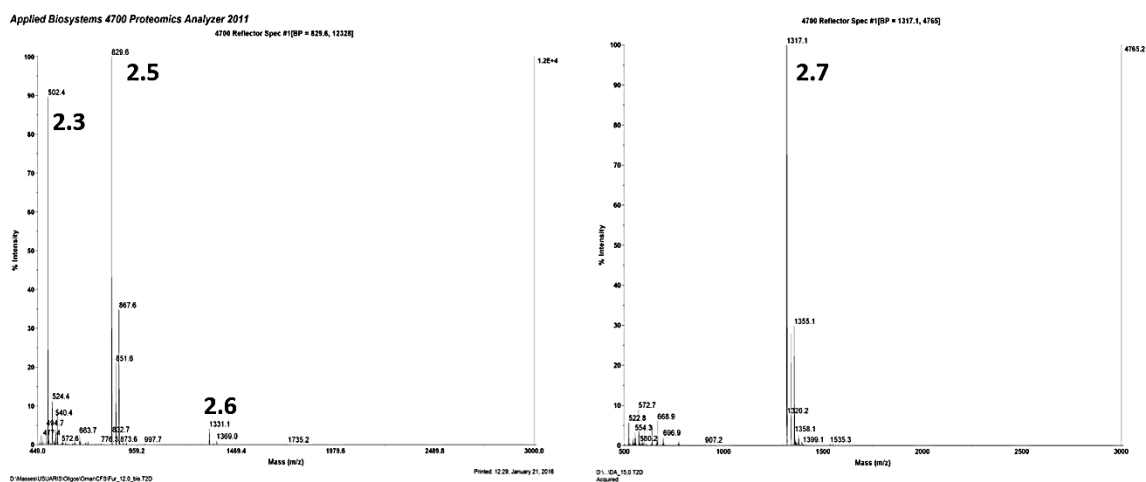


Figure 2.7. MALDI-TOF MS traces of pure **2.6** (left) and **2.7** (right).

For all these reasons the use of **2.1** as a diene was ruled out, and **2.2** became the diene of choice for all the subsequent experiments.

In further experiments, polyamide resins were analysed (by deprotecting and cleaving a small aliquot and analysing the resulting polyamide by HPLC and MALDI-TOF MS) and their functionalisation was determined before diene introduction.

2.1.9 Assessment of the solvent influence on the on-resin Diels-Alder conjugation

We were aware that although pure water was the best solvent in terms of rate acceleration and biomolecule solubility, some important reporter molecules such as fluorophores would not be soluble in such a polar medium. For this reason, the influence of organic co-solvents on this on-resin DA conjugation methodology needed to be assessed. In order to avoid a wide screening of solvents, DMSO was selected as the organic co-solvent, as it is known to solubilize a wide variety of molecules with very different polarities.

For this purpose, the totally protected, resin-linked peptide Gly-Arg-Gly-Ser-Tyr-Glu-Ala-Tyr-Lys was modified with **2.2**, affording the diene-derivatised peptide **2.8**. This sequence was selected because all the different types of amino acid are present: positively charged, negatively charged, aromatic, nonpolar aliphatic and polar non-charged residues.

The reaction of **2.8** with **2.5** was conducted both in pure water and in a 1:1 H₂O/DMSO mixture, using the same conditions as before. HPLC Analyses of the crudes after cleavage and deprotection showed, as expected, a big difference between the two reactions. In pure water the reaction furnished the expected conjugate (**2.9**) in 74 % yield and only little amounts of the diene decomposition products, meaning that only a small percentage of unreacted diene was present at the cleavage step. Other impurities also affecting the HPLC-determined yield arose from the synthesis of the peptide. On the contrary, when the reaction was performed with the DMSO/water mixture, **2.9** was formed in 48 % yield, with larger amounts of diene decomposition products indicating higher amounts of unreacted diene at the moment of the acidic treatment (**Figure 2.8**).

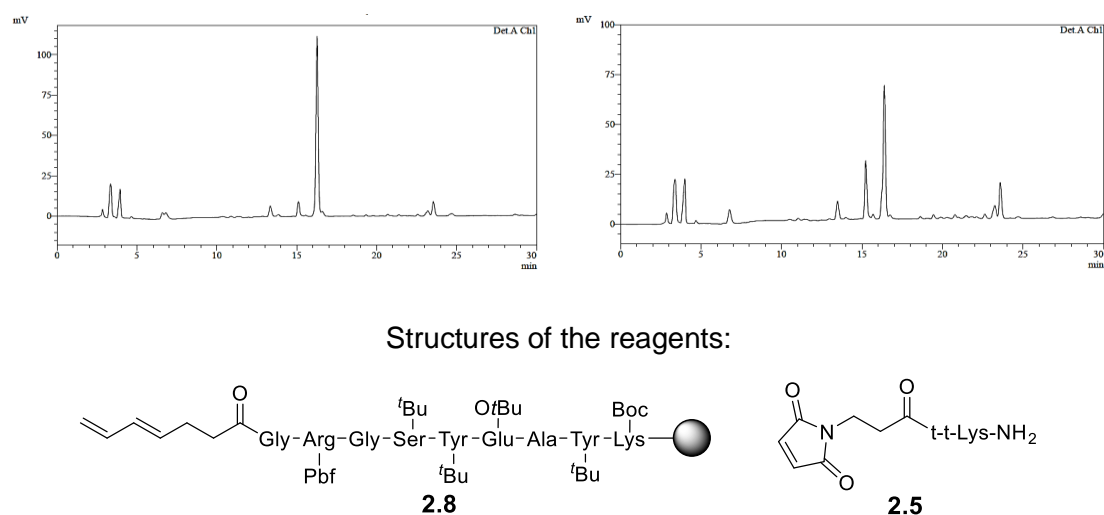


Figure 2.8. Comparison of the **2.5 + 2.8** conjugation reaction in water (left) and in a 1:1 DMSO/water mixture (right). The desired conjugate (**2.9**) is the peak at ~ 16.0 min.

The result is consistent with the described effect that organic co-solvents have in the activation energy of the DA reaction,¹⁵ and pointed out that the use of aqueous/organic mixtures should be avoided as much as possible. Therefore, working with molecules sparingly soluble or insoluble in water was expected to be somewhat troublesome.

2.1.10 Optimizing the reaction conditions

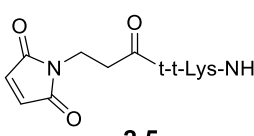
The next step was to find appropriate reaction conditions, allowing to use a minimum excess of the maleimide-derivatised compound and not requiring extremely long reaction times. The conditions reaching the best balance between these variables would be chosen as our standard protocol for reactions in water. Again, **2.8** and **2.5** were used as model compounds.

The first parameter we wanted to change was the long reaction times used in previous assays, employing pure water as solvent due to its positive effects on the reaction yield (**Table 2.1**, entries 1 and 2). For this purpose, temperature was raised from 40 to 65 °C and the reaction left running for 24 h (**Table 2.1**, entry 3). Using these conditions the yield dropped from 74 to 62 %, indicating that the rise in temperature did not fully compensate for the shorter reaction time. As we considered that higher temperatures could start harming the peptide chain or any sensitive moiety to be conjugated, other parameters were changed. Higher conversions were therefore forced by the use of a 3-fold excess of the maleimide derivative and a 2-fold increase in its concentration (**Table 2.1**, entry 4). Thus, the reaction with 3 equivalents of **2.5** (40 mM concentration) for 24 h

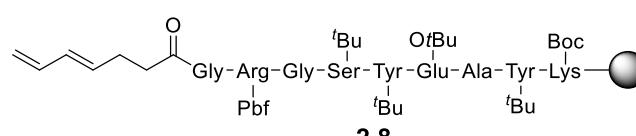
at 65 °C provided a satisfactory 79 % yield. Although the reaction time was somewhat long, the reaction could be performed in fairly mild conditions and only an affordable, small excess of the maleimide derivative was needed. As for the reaction set-up a certain volume of solution is required to fully cover the resin beads, it was not possible to further increase the maleimide derivative concentration without increasing the total amount required. We thought that a shorter reaction time would not compensate for the increase of maleimide derivative needed, especially in cases when this would be a precious compound. Therefore, no more experiments were carried out to further optimize the reaction parameters, and we decided to test the scope of the reaction.

Entry	Equiv. 2.5	Solvent	[Mal.] (mM)	Time (h)	Temp. (°C)	Yield (%)
1	1.5	H ₂ O	20	70	40	74
2	1.5	DMSO/ H ₂ O 1:1	20	70	40	48
3	1.5	H ₂ O	20	24	65	62
4	3	H ₂ O	40	24	65	79

Structures of the reagents:



2.5



2.8

Table 2.1. Reaction conditions tested for the **2.5 + 2.8** reaction. Mal.= maleimide derivative

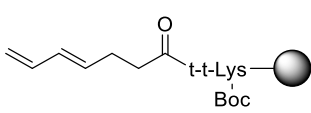
2.1.11 Broadening the scope of the reaction: conjugation of different polyamides

The first question that came to our mind after optimization of the reaction conditions was if other diene-derivatised polyamides could be conjugated employing this methodology. Among them, we focused in the use of PNAs because of their attractive properties as oligonucleotide mimics. For comparison purposes, the already optimised **2.5 + 2.8** reaction was performed interchanging each polyamide role, that is, the di-PNA was derivatised with a diene and the 9-mer peptide was modified with an *N*-terminal maleimide and used as the soluble moiety, generating **2.10** and **2.11**, respectively (**Table 2.2**, entry 1). Their conjugation product (**2.12**) was formed in 92 % yield, proving that the methodology was adequate for the conjugation of resin-linked PNA chains. A biologically

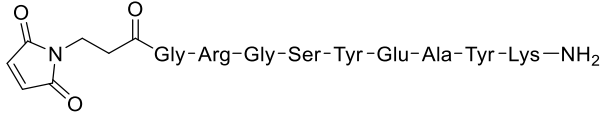
relevant sequence, the anti-*lacZ* 12-mer diene-derivatised PNA **2.13**,¹⁹ was then chosen to test the suitability of this methodology to work in real cases. In a first experiment, **2.13** was reacted with the maleimide-derivatised peptide **2.14** (Table 2.2, entry 2). In this case 3 equivalents of the soluble maleimide at a 40 mM concentration were not enough to fully cover the PNA resin beads, for which reason a larger excess of peptide, 5 equivalents, was used to avoid lowering its concentration. The reaction furnished the expected conjugate **2.15** in 72 % yield, demonstrating the utility of our methodology when working with biologically relevant molecules. The need for extra reagent volume showed by **2.13** must be attributed to the resin volume increase that accompanies the addition of each new monomer to the oligomeric chain.²⁰

Entry	Reaction	Equiv. of Mal.	Solvent	[Mal.] (mM)	Time (h)	Temp. (°C)	Yield (%)
1	2.10 + 2.11 → 2.12	3	H ₂ O	40	24	65	92
2	2.13 + 2.14 → 2.15	5	H ₂ O	40	24	65	72

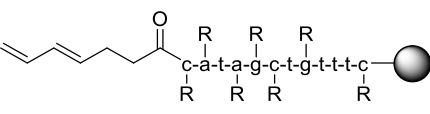
Structures of the reagents:



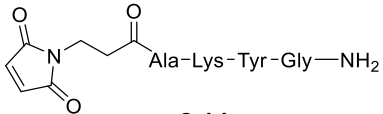
2.10



2.11



R = Bhoc
2.13



2.14

Table 2.2. Conjugation of diene-derivatised PNAs. Mal.= Maleimide derivative.

Another concern was the effect that the chain length would have in the conjugation yield. We expected larger maleimide derivatives to react more slowly, but at first sight it was not clear that this would be the case for the immobilised diene-containing compounds. The fact that in a longer chain the diene is, in principle, more distant from the resin surface could improve its exposure to the other reagents and facilitate its conversion. To answer all these questions only an additional experiment had to be done: the 5-mer, maleimide-derivatised PNA **2.16** was incubated with **2.8** in the optimised conditions, furnishing **2.17** in 69 % yield. The results of all the experiments performed allowing for an evaluation of length effects are summarised in Table 2.3. It must be stated that

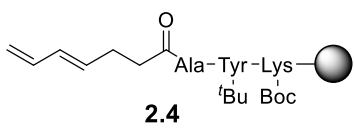
although not all experiments were performed with the same protocol, those used in any comparison were run under identical conditions.

As could be seen, with longer diene-derivatised, fully protected peptides, the yield dropped (compare entries 1 and 2). Long peptide sequences tend to contain more impurities arising from their synthesis than their short counterparts, being consequently normal to observe lower HPLC-based yields even if the extent of the reaction is the same. Nevertheless, a pronounced fall as the one observed cannot be fully ascribed to this fact, pointing to a negative effect of chain length on the conjugation yield.

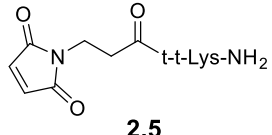
The same undesirable effect was observed when we increased the size of the soluble maleimide moiety (compare entries 3 and 4). A 10 % drop in yield took place when the PNA chain was elongated from 2 to 5 residues. In this case, the purity of the diene-derivatised peptide was the same in both cases as it came from the same synthesis batch, unambiguously demonstrating the influence of chain length on the reaction outcome.

Entry	Reaction	Equiv. of Mal.	Solvent	[Mal.] (mM)	Time (h)	Temp. (°C)	Yield (%)
1	2.5 + 2.4 → 2.7	1.5	H ₂ O	20	70	40	89
2	2.5 + 2.8 → 2.9	1.5	H ₂ O	20	70	40	74
3	2.8 + 2.5 → 2.9	3	H ₂ O	40	24	65	79
4	2.8 + 2.16 → 2.17	3	H ₂ O	40	24	65	69

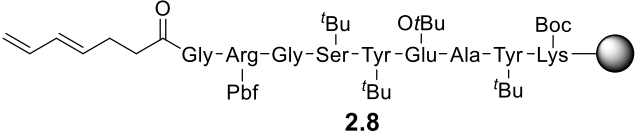
Structures of the reagents:



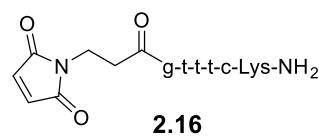
2.4



2.5



2.8



2.16

Table 2.3. Summary table for the evaluation of the effect of chain length on reaction yield. Mal.= Maleimide derivative.

2.1.12 Use of protected maleimides to obtain conjugates with different linking sites

As mentioned in chapter 1, different linking sites may provide conjugates with different stabilities or pharmacokinetic and pharmacodynamic properties. Hence, it is interesting to have a methodology that permits the synthesis of differently linked conjugates.

Stepwise synthesis of peptide-PNA conjugates affords hybrid molecules in which one of the components is linked through its C-terminus to the N-terminus of the other (**Figure 2.9a**), while the methodology developed so far in this work granted access to conjugates linked through their N-terminal positions (**Figure 2.9b**). We envisaged that, as long as the soluble molecule had the maleimide at the C-terminal or at an internal position, we could use the on-resin DA reaction in water to obtain conjugates with different linking patterns like the ones shown in **Figure 2.9a** and **Figure 2.9c**.

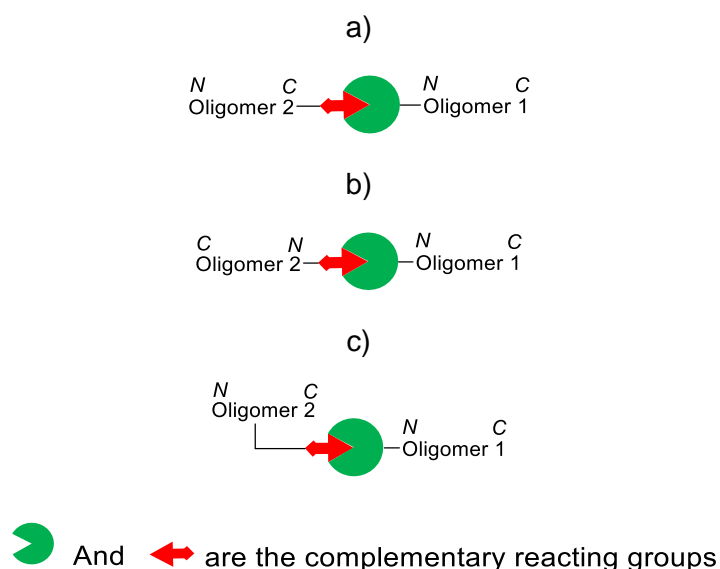


Figure 2.9. Possible linking patterns using the on-resin DA conjugation..

Locating maleimides at the C-terminal or internal positions of a polyamide chain is troublesome as they react with piperidine during the Fmoc/^tBu SPPS. A feasible option, though, is to protect them with 2,5-dimethylfuran as explained in chapter 1, avoiding nucleophile-promoted side-reactions and allowing for the preparation of internally maleimide-modified polyamides after thermal deprotection.

It was therefore of interest to study if the one-pot maleimide deprotection/on-resin DA reaction was feasible. For the proof-of-concept experiment, (2,5-dimethylfuran)-protected 2-hydroxyethylmaleimide (**2.18**) was reacted with **2.8**, furnishing the expected DA adduct (**2.19**) in 82 % yield (**Table 2.4**, entry 1).

This good result prompted us to tackle the synthesis of PNA conjugates with peptides bearing protected maleimides at different positions. In the first place, PNA **2.13a** (see section 2.1.13) was incubated with peptide **2.20** to afford the “T-shaped” conjugate **2.21** in 55 % yield, which was isolated after purification in 30 % overall yield (Table 2.4, entry 2). For the C-terminus to N-terminus linkage we decided to assemble a more complex architecture. The C-terminal maleimide-derivatised cyclic peptide **2.22** was conjugated to **2.13a** to render **2.23**, a peptide-PNA conjugate with a cyclic domain and the same linking pattern as the one obtained by solid-phase synthesis. **2.23** Was furnished in 58% yield, and isolated after purification with 28 % overall yield (Table 2.4, entry 3).

Entry	Reagents	Equiv. of Mal.	Yields (%)	Isolated Yield (%)
1	2.8 + 2.18 → 2.19	3	82	n.d.
2	2.13a + 2.20 → 2.21	5	55	30
3	2.13a + 2.22 → 2.23	5	58	28

Structures of the reagents:

Table 2.4. Conjugation reactions affording different linkage patterns. Mal.= maleimide derivative. All reactions run at 65 °C for 24 h with [Mal.] = 40 mM in water.

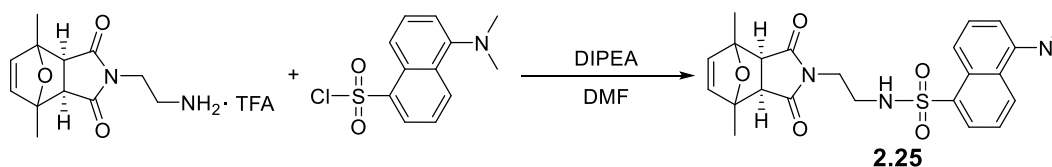
2.1.13 Conjugation with molecules sparingly soluble or insoluble in water

After the good results obtained with complex architectures and the generation of conjugates with different linking sites, it was time to tackle the use of molecules sparingly soluble or insoluble in water. As described in the literature and confirmed by our experiments (see section 2.1.9), the use of water/organic solvent mixtures is not recommended due to its negative effect on the reaction yields. Certain apolar molecules

such as many fluorophores, though, are not soluble in pure water, leaving no other option than using organic co-solvents to deal with their solubility issues.

To assess whether polyamides potentially useful in biochemical studies could be synthesised using our methodology, biotin and dansyl, two well-known reporter molecules, were chosen to be conjugated to the 9-mer peptide **2.8** and the 12-mer PNA **2.13**.

The biotin (**2.24**) and dansyl (**2.25**) derivatives were synthesised using (2,5-dimethylfuran)-protected maleimide instead of the unprotected counterpart, for no other reason than building block availability. **2.24** Had been previously synthesised in the group and had to be only washed with DCM (HPLC-quality) to remove certain impurities. Regarding **2.25**, it was easily obtained from commercially available dansyl chloride by reaction with (2,5-dimethylfuran)-protected 2-aminoethylmaleimide trifluoroacetate in the presence of 2.5 equivalents of DIPEA in DMF (**Scheme 2.6**). The product was obtained, after the appropriate work-up, pure enough to be used in conjugation reactions.



Scheme 2.6. Synthesis of **2.25**.

The reaction of **2.24** with PNA **2.13** was troublesome. A first attempt was performed using an 80:20 water/DMSO mixture as solvent (**Table 2.5**, entry 1) to ensure solubility of **2.24**, but the reaction yielded a complex crude in which it was not possible to identify the product (**2.26**) (**Figure 2.10a**). We thought that this was caused by a dramatic reaction rate decrease triggered by the addition of the organic co-solvent, and, consequently, decided to try to dissolve the biotin derivative in water only. **2.24** Was not completely soluble in pure water at room temperature, but it could be solubilised by incubating its suspension in a sand bath at 150 °C for a few seconds, with no appreciable precipitation upon cooling. When carried out with the usual conjugation protocol, using pure water as solvent (**Table 2.5**, entry 2), the reaction rendered a different but still complex crude (**Figure 2.10b**), pointing to the change of solvent as not the unique cause for those bad results.

Entry	Reagents	Equiv. of Mal.	Solvent	Yield (%)	Isolated Yield (%)
1	2.13 + 2.24 → 2.26	5	80:20 H ₂ O/DMSO	n.d	n.d
2	2.13 + 2.24 → 2.26	5	H ₂ O	n.d	n.d
3	2.13a + 2.24 → 2.26a	5	H ₂ O	60	n.d
4	2.13b + 2.24 → 2.26b	5	H ₂ O	56	n.d
5	2.8 + 2.24 → 2.27	3	H ₂ O	67	n.d
6	2.8 + 2.24 → 2.27	3	80:20 H ₂ O/DMSO	63	n.d
7	2.8 + 2.25 → 2.28	3	65:35 H ₂ O/DMSO	88	48
8	2.13a + 2.25 → 2.29	5	65:35 H ₂ O/DMSO	72	19

Structures of the reagents:

Table 2.5. Conjugation reactions dealing with molecules sparingly soluble or insoluble in water. Mal.= maleimide derivative. All reactions carried out at 65 °C for 24 h with [Mal.] = 40 mM.

We then hypothesised that introduction of a short spacer between the PNA chain and the diene could increase the conjugation yield and help obtain a cleaner crude. Thus, one or two glycine residues were coupled to the 12-mer PNA catagctgttcc prior to the incorporation of the diene, obtaining PNAs **2.13a** and **2.13b**, respectively. These were then tested in the reaction with **2.24** using pure water as solvent to give products **2.26a** and **2.26b**. A clear major peak was observed in both reaction crudes, with similar relative peak areas: 60 % for **2.13a** and 56 % for **2.13b** (Figure 2.10c and d and Table 2.5, entries 3 and 4). This result provided conclusive evidence that the introduction of a spacer between the reacting diene and the PNA chain can help improve reaction yields and obtain cleaner crudes in complicated cases. On the other side, though, the effect of the spacer length cannot be evaluated from these experiments, as such a small difference in yield does not seem important, especially taking into account the error made in the integration of such wide peaks.

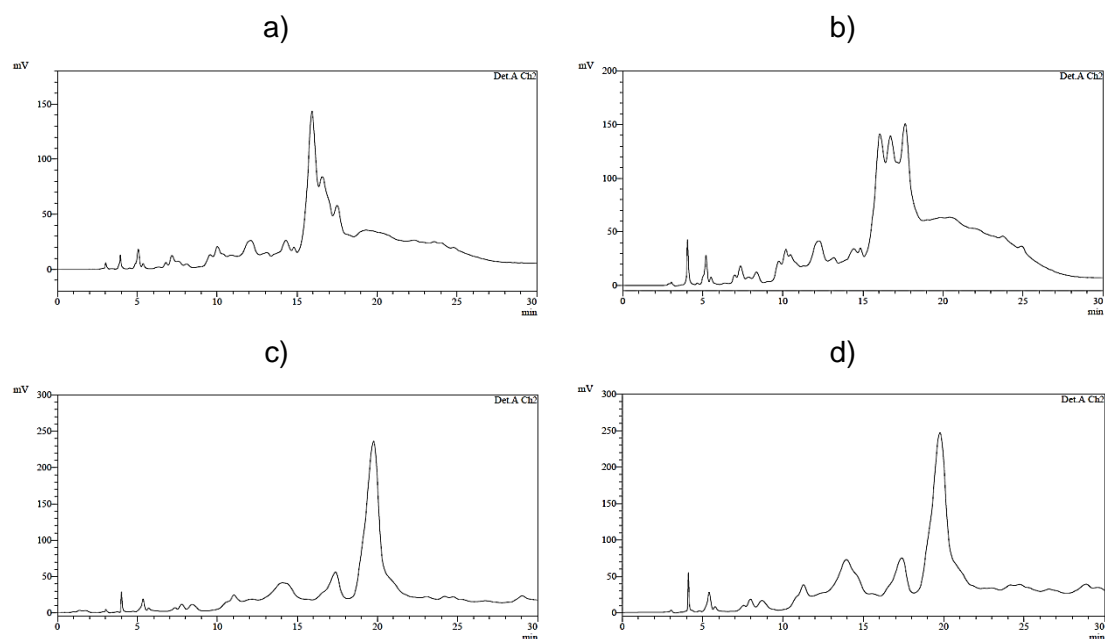


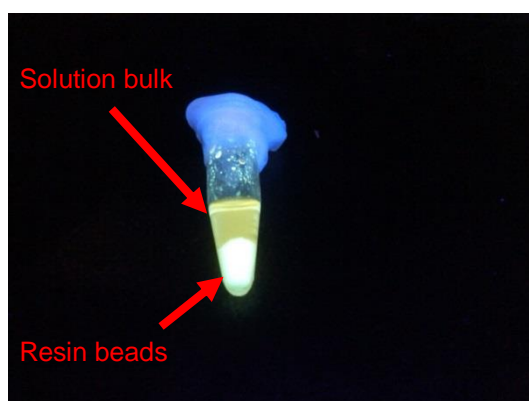
Figure 2.10. HPLC traces of the **2.24 + 2.13** reactions in the 80:20 water/DMSO mixture (a) and in pure water (b), and of **2.24 + 2.13a** (c) and **2.24 + 2.13b** (d) cycloadditions.

2.24 Was also reacted with **2.8** in the typical conditions employed for the diene-modified peptide conjugations, this time using both pure water and the 80:20 water/DMSO mixture as solvents. Clean crudes with similar conjugate (**2.27**) yields, around 65 %, were obtained in both cases (**Table 2.5**, entries 5 and 6). This is surprising as, regarding solvent effects, it does not agree with the previously obtained results (**Table 2.1** entries 1 and 2) and what is described in the literature.

More shocking was the outcome of the reactions performed with **2.25**. The derivatised fluorophore had to be dissolved in a 65:35 water/DMSO mixture (in which it was still sparingly soluble) in order to perform the reactions with **2.8** and **2.13a**. Therefore low conjugation yields were anticipated. However, we were glad to observe an 88 % yield for the dansyl-peptide conjugate **2.28** and a 72 % yield for **2.29**, the dansyl-PNA hybrid (**Table 2.5**, entries 7 and 8). These conjugates were isolated in 48 and 19 % yield, respectively.

Although satisfied with the good results, we were somewhat intrigued by the unexpected high yield of the conjugation reactions in the 65:35 water/DMSO mixture. To try to find an explanation, the **2.25 + 2.8** reaction was repeated and inspected under UV light before incubation at 65 °C.

As can be seen in **Picture 2.1**, the fluorophore seemed to be mainly adsorbed onto the resin surface, presumably due to hydrophobic interactions between the apolar molecule and the carbon-rich chains of the polymeric solid support. This could produce an increase of the effective concentration of **2.25** near the reactive points, the resin surface and inner channels, as depicted in **Figure 2.11**, and likely compensate for the rate decrease expected for a DA reaction performed in a water/organic solvent mixture.



Picture 2.1. Reaction of **2.25** + **2.8** inspected under UV light (254 nm) at the beginning of the process.

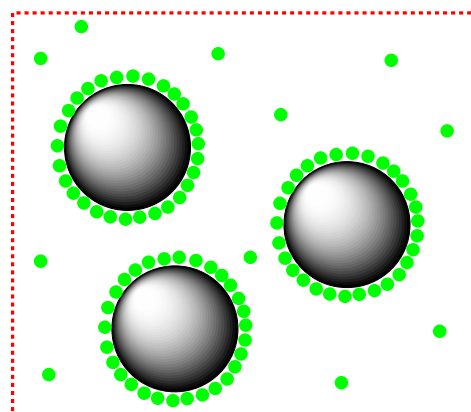


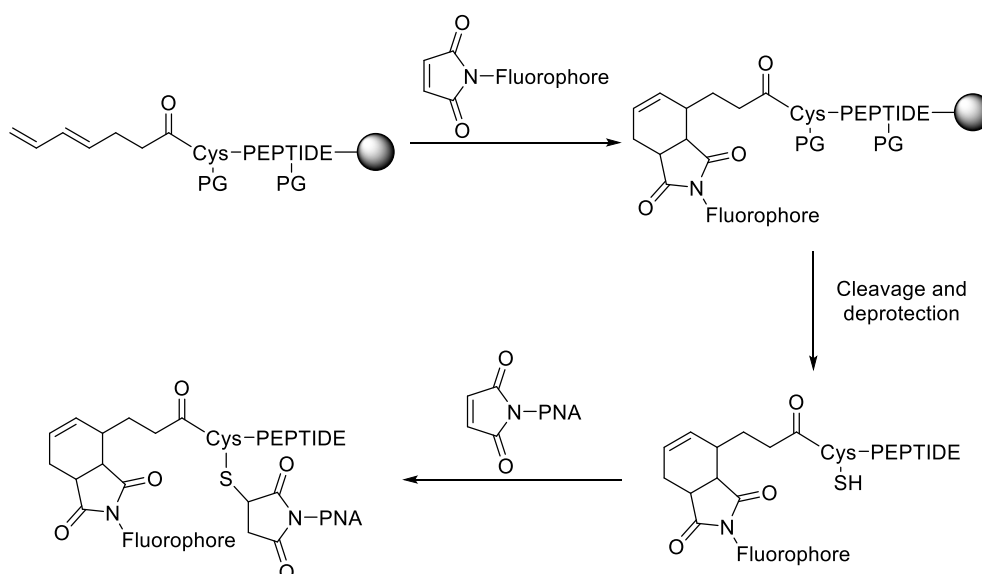
Figure 2.11. Schematic representation of the adsorption of **2.25** onto the resin surface.

Unfortunately, we could not find an explanation accounting for the lack of solvent effects observed in the **2.24** + **2.8** reaction. Developing a theory that could explain the observed results would require a lot of experiments, something completely out of the scope of our project.

2.1.14 Synthesis of double conjugates

The last objective of our work was to test if our methodology would be compatible with other click chemistries and could enable the synthesis of double conjugates. The double derivatization of biomolecules is important in situations where more than one goal is to be achieved. For example, PNAs, which can interact with both DNA and RNA, are unable to penetrate cell membranes. Conjugation to cell-penetrating peptides (CPPs) allows to overcome this obstacle, but the fate of the conjugate cannot be monitored unless a suitable tag is attached to the molecule. PNAs Attached to both CPPs and labelling agents such as fluorophores are thus molecules of interest, as they can cross the cell membrane and, at the same time, be localised by fluorescence microscopy.

We thought that the on-resin DA reaction could be a good platform to obtain this type of double conjugates through tandem DA and Michael-type (MT) reactions. Our idea, as depicted in **Scheme 2.7**, was to add a cysteine residue to the peptide chain prior to the coupling of the diene moiety, thus introducing the reacting thiol for the MT reaction. Although we planned to use a maleimide for both reactions, selectivity should not be a problem as the thiol group of cysteine would be protected during the on-resin DA reaction. After this, the first conjugate would be deprotected and cleaved from the solid support to be, after purification, reacted with a second maleimide moiety to furnish the desired doubly-modified peptide.



Scheme 2.7. General scheme for the synthesis of double conjugates through tandem Diels-Alder and Michael-type reactions.

Regarding the fluorophore, we decided to use the water soluble, commercially available 1-[2-(maleimido)ethyl]-4-[2-(3,4-dihydro-2H-benzopyran-6-yl)-5-oxazolyl]-pyridinium triflate (**2.30**). Because the maleimido-PNA derivative would be more valuable than the commercial fluorophore, the former would be used in the last step of the process, letting us use less equivalents of the precious reagent.

As for the thiol protecting group of cysteine, we thought the Trt group would be highly convenient because its cysteine derivative is commercially available and does not require any additional deprotection treatment, minimizing the number of steps required for the synthesis of the double conjugate.

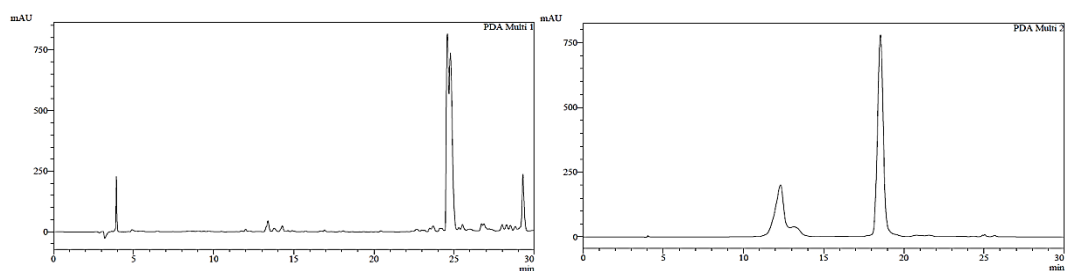
On order to evaluate its stability to the on-resin DA reaction conditions, 3-maleimidopropanoic acid (4 equivalents) was incubated with the resin-linked, protected

peptide Cys-Gly-Arg-Gly-Ser-Tyr-Glu-Ala-Tyr-Lys for 24 h at 65 °C. The peptide was deprotected and cleaved from the resin using a 94:2.5:2.5:1 TFA/EDT/H₂O/TIS mixture and analysed by HPLC. MALDI-TOF MS Analysis of the major peak indicated that the Trt protecting group had remained stable during the thermal treatment, as no incorporation of 3-maleimidopropanoic acid was found.

Having assessed the stability of the Trt group, we started the synthesis of the double conjugate.

Resin-linked peptide Cys-Gly-Arg-Gly-Ser-Tyr-Glu-Ala-Tyr-Lys was derivatised with the diene moiety, affording the diene-derivatised, fully protected peptide **2.31**. In parallel, the maleimide modified PNA Mal-Lys-Lys-catagctgtttc-NH₂ (**2.32**) was synthesised and purified. When assaying the on-resin DA reaction between **2.30** and **2.31**, we discovered that **2.30** was not soluble enough to reach the desired 40 mM concentration in pure water, which forced us to use an organic co-solvent for the reaction. The previously used 65:35 water/DMSO mixture was tested, but even in this conditions **2.30** proved to be less soluble than **2.25**. As we did not want to further increase the DMSO content, we decided to use a suspension of **2.30** in this water/organic solvent mixture. Incubation in the usual peptide conjugation conditions furnished the first conjugate **2.33** in 77 % yield (**Figure 2.12** left) showing that the methodology is flexible and withstands even the use of suspensions. After purification, **2.33** was reacted with 1.3 equivalents of the maleimido-PNA **2.32** in very mild conditions (water, room temperature) to provide the double conjugate **2.34**. The **2.33** to **2.34** conversion was quantitative and, apart from the target double conjugate, only excess PNA **2.32** (and its impurities) could be found in the reaction crude (**Figure 2.12** right).

The yield for the overall process, from the diene-derivatised resin-linked peptide to the isolated final double conjugate **2.34**, was 23 %, demonstrating that our methodology can indeed furnish this type of structures in good overall yields.



Structures of the reagents:

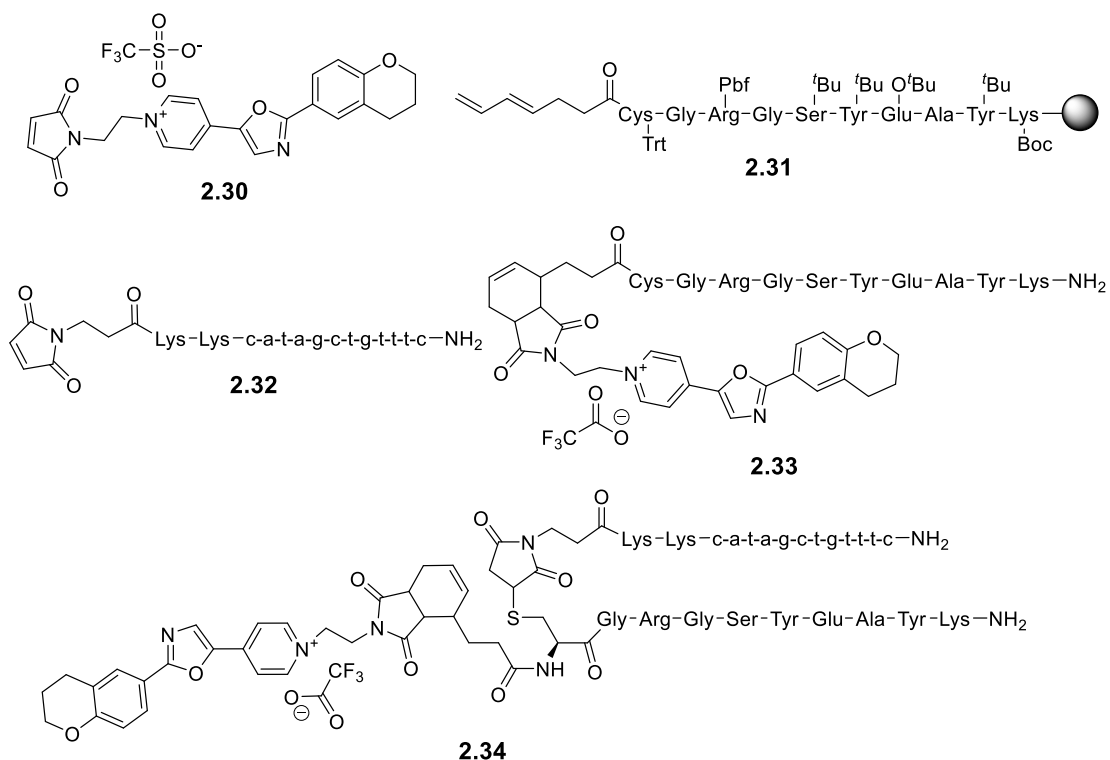


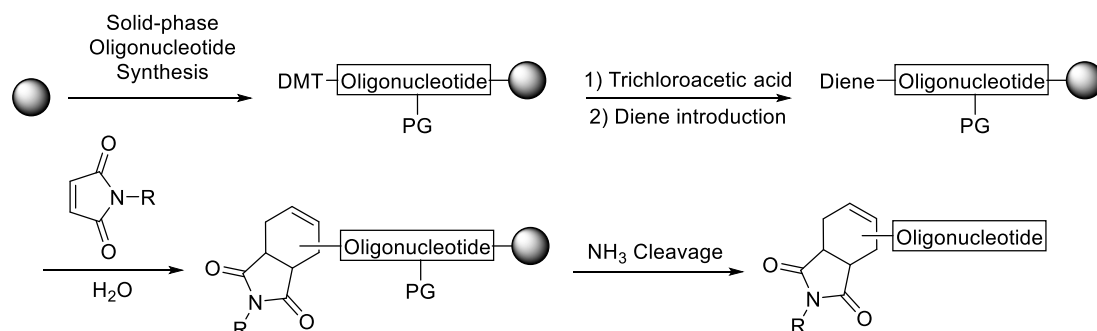
Figure 2.12. HPLC traces of the reaction crudes after Diels-Alder (left) and Michael-type (right) ligations. Peaks around 12 minutes in the Michael-type conjugation are excess PNA and impurities present in the starting material. Structures of the involved molecules are depicted under the HPLC traces.

2.2 Conjugation of diene-derivatised oligonucleotides

2.2.1 Background and objectives

Peptide-oligonucleotide conjugates (POCs) cannot be obtained by reacting maleimido-oligonucleotides with resin-linked diene-peptides because oligonucleotides (ONs)

undergo depurination reactions in acidic medium. Even though maleimido-peptides and diene-oligonucleotides furnish the target conjugate in solution,²¹ motivated by the good results obtained we decided to examine whether the DA reaction could be carried out on a solid support. In other words, our idea was to react maleimide-derivatised peptides with resin-linked, protected ONs modified with a diene moiety (**Scheme 2.8**).



Scheme 2.8. Our proposal for the synthesis of POCs using a solid-phase fragment condensation.

This synthetic scheme would avoid the ON moiety to be in contact with the strong acid, preventing its depurination, and the synthesis of the POC would require less purification steps than the solution conjugation method, increasing the overall synthesis yield. Further, the POC was expected to present an HPLC retention time pretty different from the unconjugated ON, allowing for a better separation from the deletion sequences arising from the stepwise synthesis.

Other groups have investigated the possibility of synthesizing POCs on a solid matrix taking advantage of different reactions. Gait and co-workers have described^{22,23} the synthesis of these structures using amide formation. In their studies, a resin-linked 5'-amino ON was reacted with the C-terminal carboxylic group of a peptide using PyBOP and HOBt as coupling reagents. They found a great influence of the spacer used between the ON chain and the amine functionality, as well as the ON sequence and its hydrophobicity (mostly related to the CNE content) on the coupling yield. Unfortunately, HPLC conjugation yields of mixed ON sequences with relevant peptides were lower than 50 % in all cases. In addition, no mention to peptide sequences containing amino acids requiring protection of the side chain (lysine, aspartic or glutamic acid, for instance) can be found in these papers.

Kubo *et al.* described methodology in which resin-linked, 5'-amino ONs and peptides containing one single reactive amine were linked, using a 1,6-diisocyanatohexane linker, through formation of two urea bonds.²⁴ This methodology allowed for the use of several

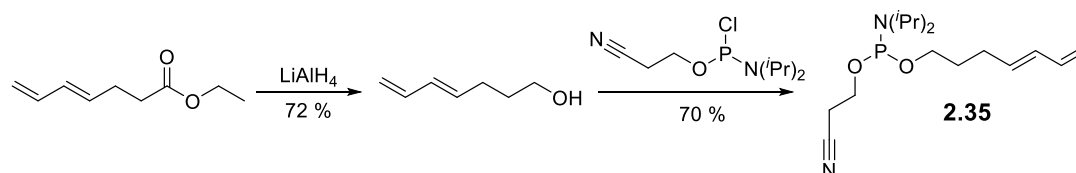
unprotected trifunctional amino acids, including arginine and glutamic acid. The only requirement was to mask lysine side-chains with the Tfa protecting group, which was removed in the final ammonia treatment. Unluckily, the methodology needed quite long reaction times and large excesses of peptide, and rendered the desired conjugates in poor yields (8-18 %) due to formation of large amounts of by-products.

Finally, Strömberg and co-workers used the Cu (I)-catalysed Huisgen reaction to link mixed ON sequences to cell-penetrating peptides.²⁵ In this strategy, basic lysine residues could be used either unprotected or Tfa-protected, with no difference in the reaction outcome, while other trifunctional amino acids (Asp, Arg) were used without protection. As judged from the HPLC traces, the methodology rendered a non-negligible amount of side-products, but the target POC seemed to be easily isolable and could be obtained in good yields after purification.

2.2.2 Reaction parameters and analysis of the reaction outcome

As before, water was chosen as the solvent to be used routinely and maleimide as the dienophile. Sensitivity of the succinimide generated upon on-resin DA reaction to basic media imposed oligonucleotides to be assembled using the UltraMild protecting scheme, which employs *t*-PrPac for the protection of G, Pac to protect A and Ac to protect C, and phenoxyacetic anhydride as capping agent. This protecting scheme allows to perform the cleavage and deprotection of the ON in mild conditions and short reactions times, thus reducing the extent of the succinimide hydrolysis side-reaction. Oligonucleotides are normally synthesised over Controlled Pore Glass (CPG) solid supports, which, contrary to what happens with PS resins, do not shrink in any solvent. Therefore, this enabled us to perform the DA conjugation on this support and in water.

The diene moiety used for the on-resin DA reaction of diene-derivatised oligonucleotides was the same as that employed in the case of diene-derivatised peptides, but in this case its introduction into the oligomeric chain was accomplished using the corresponding phosphoramidite (**2.35**). This was synthesised in two steps from the precursor of **2.2** (see **Scheme 2.5**), as shown in **Scheme 2.9**. Briefly, reduction of the ester moiety by LiAlH₄ afforded the expected alcohol in 72 % yield, which was subsequently phosphitilated using 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite to furnish **2.35** in 70 % yield. This was incorporated into ONs under standard conditions but omitting the capping step, allowing us to repeat the coupling if judged necessary based on the HPLC analysis of a small aliquot of the batch.



Scheme 2.9. Synthesis of phosphoramidite **2.35**.

For the study of the reaction outcome, the same analytical methodology as in section 2.1 was employed.

2.2.3 Work with short ON models

The first assays were performed using **2.36**, a 5-mer poly-dT incorporating the diene moiety, as a model substrate for the on-resin DA reaction with the maleimido-modified tetrapeptide **2.14**. The reaction was performed in pure water as described previously, using 5 equivalents of the soluble reagent due to the low substitution degree of the CPG support. After 24 h incubation at 65 °C, the resin was thoroughly washed and incubated with concentrated aqueous ammonia for 1 h at room temperature. HPLC analysis of the crude and MALDI-TOF MS analysis of the collected peaks allowed us to examine the outcome of the reaction (**Figure 2.13**).

The conjugate (**2.37**) was produced in 68 % yield (HPLC-based), accompanied by a 6 % of unreacted material, a 19 % of ON fragmentation products, a 4 % of hydrolysed conjugate and a 2 % of Pac acylated product. Water addition to the conjugate was anticipated due to the already-known succinimide instability to basic media, but the ON fragmentation and conjugate acylation were not expected beforehand. Curiously, the fragmentation reaction seemed to have taken place mainly on the last phosphate diester bond, cleaving the diene moiety from the rest of the ON.

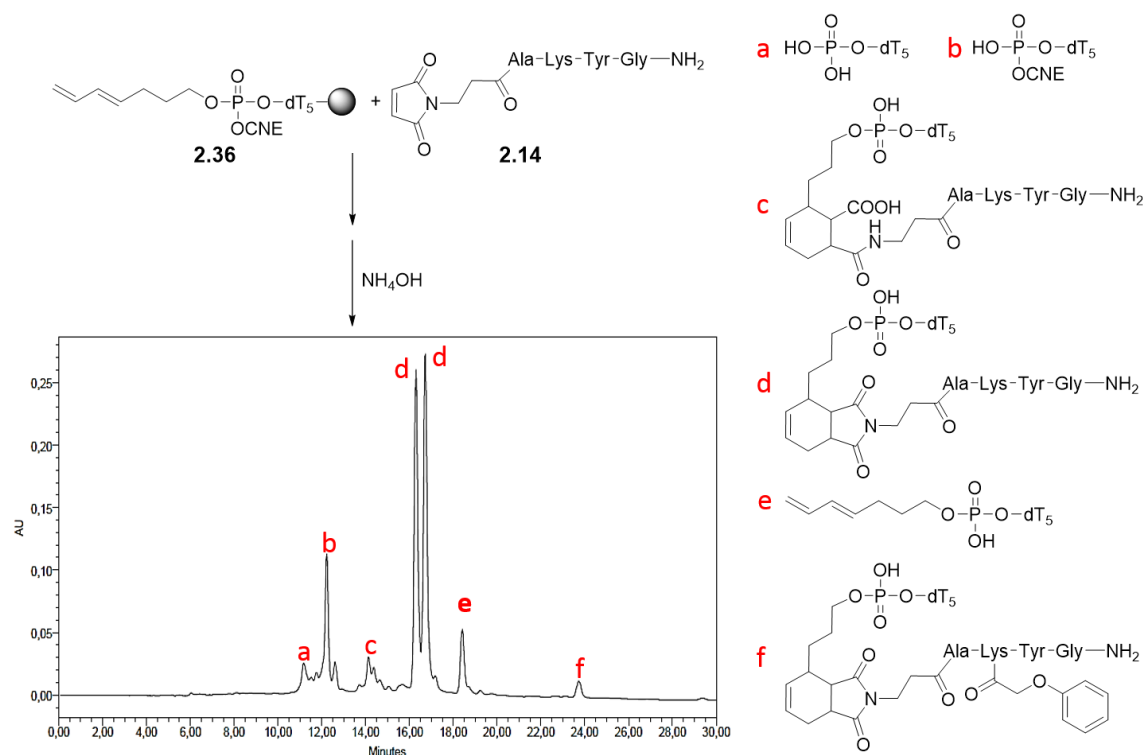
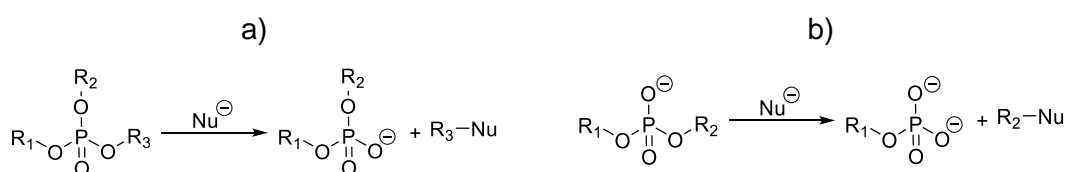


Figure 2.13. HPLC trace (260 nm) of the reaction between **2.36** and **2.14** and observed products. Products were identified by MALDI-TOF MS.

A brief search in the literature, though, made us realize that both the acylation and the fragmentation problems had been described before. Both Richert²⁶ and Strömberg²⁷ had reported the acylation of the terminal amine in 5'-aminomodified ONs when tackling the synthesis of POCs. The former described that *t*-butylphenoxyacetyl exocyclic amine protecting groups could spontaneously migrate to the terminal 5'-amine, but that this acylation problem could be solved with the use of the traditional, less labile, benzoyl and *i*-butyryl or dmf protecting groups. However, Strömberg not only found that these less labile protecting groups also migrated to the 5'-amine terminus, but that nucleobases acylated during the capping steps tended to transfer their acyl groups to the terminal amine position. In our case, as the ON bases used in this first experiment did not require any protection, the source of the conjugate + Pac product had to be the capping reagent. An alternative to circumvent the acylation problem is to avoid the capping step on the last ON synthesis cycle (which we already did) and perform a piperidine or morpholine wash prior to the introduction of the amine functionality.²⁷ This nucleophilic wash removes the most labile acyl groups, either they come from the protecting scheme or from the capping solution, and precludes acylation of the 5'-alkylamine.

Regarding the fragmentation of the ON chain, the Richert group²⁶ had described the cleavage of the phosphodiester backbone to be particularly important when basic peptide

sequences were synthesised. Further, they described the reaction to be apparently absent when non-amino acid appendages were coupled to the amino terminus, all together pointing to the peptide nucleophilic functionalities as the origin of this side-reaction. More importantly, they could not induce the degradation of the purified POC by heating it in ammonium hydroxide, indicating that fragmentation had its origin in the phosphate triester linkage, while the natural diester counterpart was completely stable. This is easily understood: i) when observing the products generated upon cleavage of these bonds (a phosphate triester generates a monoanion, whereas the diester moiety renders a dianion) and ii) taking into account that nucleophilic attack on the phosphorus atom is favoured when the phosphate is not charged (**Scheme 2.10**).



Scheme 2.10. Comparison of the cleavages of a phosphate triester (a) and a phosphate diester (b).

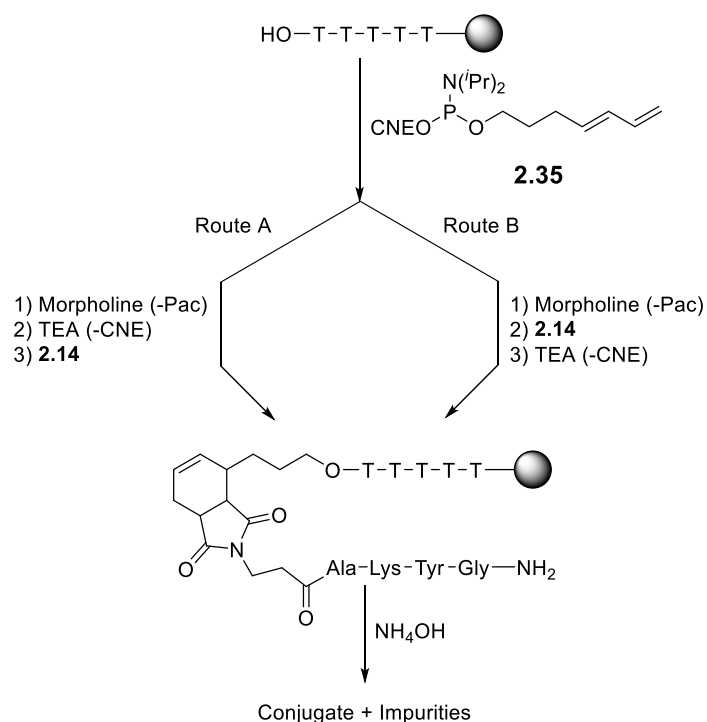
This explanation made us think that a possible way to prevent ON fragmentation was to remove the CNE protecting groups prior to the step where fragmentation seemed to take place. This would generate a phosphate diester linkage that, as mentioned above, is more difficult to cleave. Furthermore, removal of the CNE groups can be done easily and several methods are described, among which we chose the protocol based on TEA washes.²²

As for the hydrolysis reaction, we assumed that nothing could be further done, and we had to accept a 4-5 % of hydrolysis side-products.

With all this information in hand, we then moved to assess if these alternatives would really help us to reduce the formation of side-products.

A first experiment was carried out performing both the morpholine and TEA washes before the conjugation reaction (see **Scheme 2.11** route A). We gladly observed that the conjugate **2.37** was obtained in 83 % yield, that fragmentation decreased from 19 % to less than 2 % and that the acylated product was no longer detectable (see **Figure 2.14a**). However, a new unidentified product (peak at 27.5 min in **Figure 2.14** a and b) accounted for 2 % of the crude, the same percentage as the previously observed acylation. A second experiment, in which the only change was to perform the TEA treatment after the

conjugation reaction (see **Scheme 2.11** route B), yielded very similar results (see **Figure 2.14b**).



Scheme 2.11. Different methods followed to reduce the amount of acylated and fragmentation products.

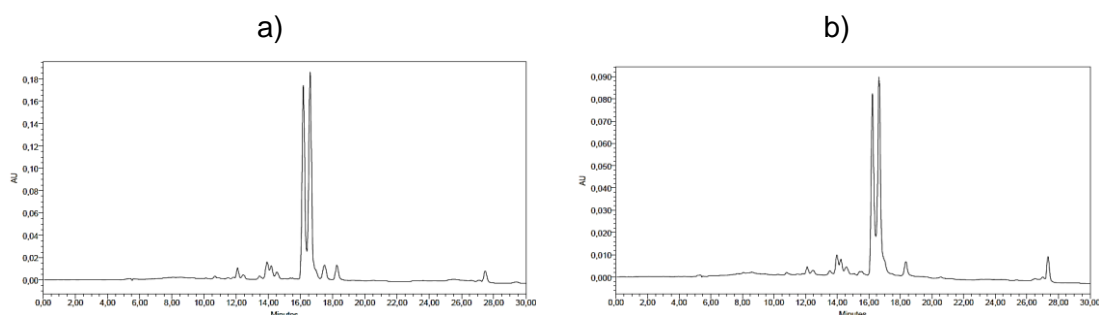
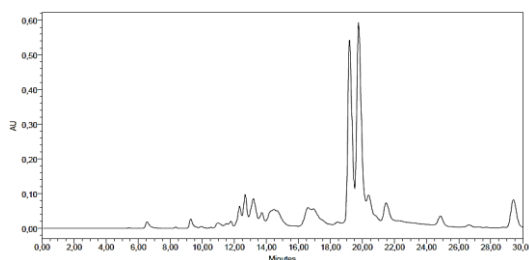


Figure 2.14. HPLC traces (260 nm) of the conjugation reactions of **2.36** + **2.14** when morpholine and TEA treatments are performed before conjugation (a) and when the TEA treatment is performed after conjugation (b).

In both cases, we had managed to suppress most of the fragmentation side-reactions and increased the conjugation yield to a very good percentage, which prompted us to tackle the synthesis of larger POCs of mixed sequences.

2.2.4 Mixed ON sequences

The resin-linked oligonucleotide 5'-d(TGT ACA TCA T)-3' was derivatised with the diene moiety to obtain **2.38**. This was reacted with **2.14** in the previously optimised conditions after a morpholine wash, and CNE groups were removed after the conjugation step using TEA. After a 2 h deprotection with concentrated ammonium hydroxide at room temperature, chromatographic analysis (**Figure 2.15**) revealed that side-reactions were a considerable problem again.



Structures of the reagents:

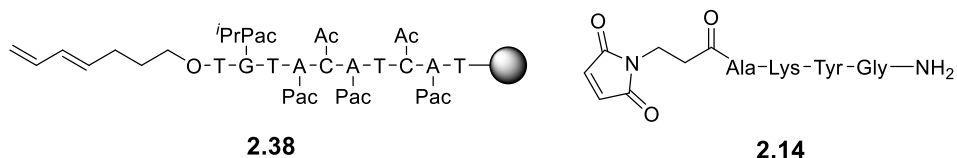


Figure 2.15. HPLC analysis (260 nm) of the **2.14 + 2.38** reaction. Structures of the reagents are also shown.

Hydrolysis had raised to 8 %, while acylation (by Ac and Pac) increased to 6 % despite the morpholine treatment. Even more alarming was fragmentation: almost 20 % of the ON chains were cleaved, generating a large variety of products (peaks between 10 and 16 min in **Figure 2.15**). Remarkably, MALDI-TOF MS analysis of these species revealed that some of them still contained a CNE protecting group. We believed that this was sitting in the cleaved phosphate bond, as preservation of the CNE protecting group at this position would avoid the energetically disfavoured generation of a dianionic phosphate monoester. This indicated that either the TEA treatment was not efficient enough or that fragmentation had taken place during the DA reaction. All in all, the conjugate (**2.39**) yield was 60 %.

As evidences showed that, again, the major problems were basically acylation and fragmentation, we decided to repeat the reaction but doubling the morpholine washes and prolonging the TEA treatment (from a total of 45 minutes to almost 3 h). Despite these efforts, analysis of an aliquot of conjugate resin showed virtually the same

outcome, with no diminution of the fragmentation or the acylation products. We thought that the TEA treatment might not be efficient enough, so we applied an extra treatment with 10% DEA in ACN²⁸ before deprotection with ammonia of the remaining CPG beads. This extra step did not ameliorate the situation at all, and the reaction outcome was, again, the same.

The fact that longer TEA treatments or even the harsher DEA procedure did not induce any change in the reaction outcome made us think that probably fragmentation occurred during the DA conjugation. In order to confirm this hypothesis, we performed the **2.14 + 2.38** reaction again, but this time applying directly the DEA treatment before the conjugation. The reaction yielded a complex crude (**Figure 2.16**) that did not provide any useful information about the problematic of the fragmentation by-products.

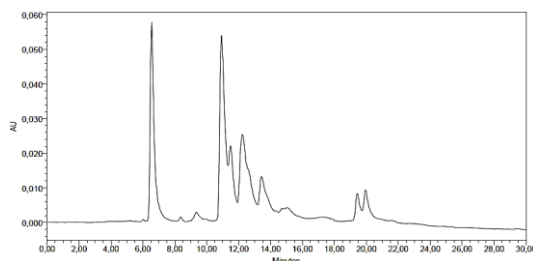


Figure 2.16. HPLC analysis (260 nm) of the **2.14 + 2.38** reaction when the DEA treatment was performed prior to conjugation.

2.2.5 MALDI-TOF MS monitoring of the conjugation reaction

As the previous experiment did not allow us to verify that fragmentation was taking place during the DA reaction, we decided to look for alternative ways to proof or refute this hypothesis.

Vasseur and co-workers published in 2000²⁹ and 2002³⁰ two papers describing the MALDI-TOF analysis of fully protected, resin-linked ONs to monitor progress of solid-phase synthesis. This technique relies on the fact that the laser source of the MALDI mass spectrometer not only ionizes the ON chain but also promotes the cleavage of the phosphate bond, generating anions detached from the solid support that can be detected. The key concept is that the detached anion does not suffer any further fragmentations and, as phosphate bond cleavage takes place only once, information about the ON sequence can be obtained. In other words, all fragments detected contain the entire sequence from the 5' end until the phosphate position where the cleavage has taken place, as shown in **Figure 2.17**.

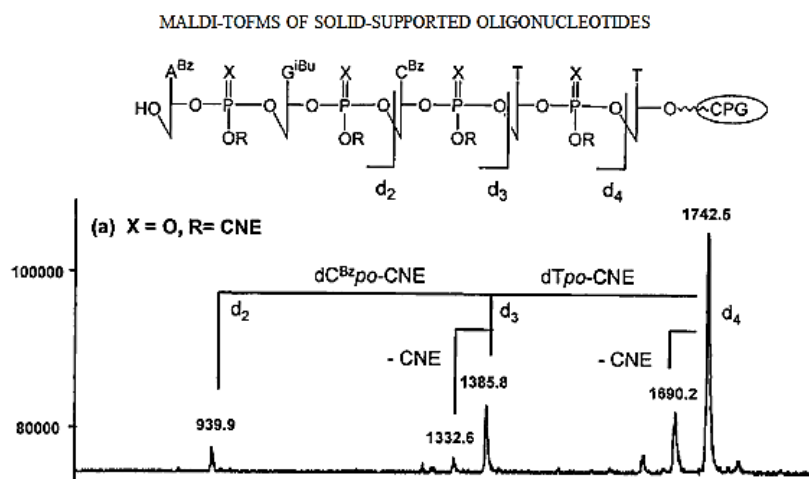


Figure 2.17. MALDI-TOF analysis of a protected ON linked to the solid support.

Reproduced from reference 29.

All the observations made by Vasseur's group showed that, in the negative mode, among all the possible bond breakings (**Figure 2.18**) in phosphate triester unions, only the "d" fragmentation, the one between the 5'-hydroxyl and the sugar ring actually takes place. This leads to clean and easily interpretable mass spectra. Apart from the peaks corresponding to these ions, other peaks of minor intensity, attributed to the loss of labile groups such as CNE or DMT (when the analysis was performed on 5'-DMT ONs) were also observed. No loss of the usual Bz (for A and C protection) and ⁱBu (for G protection) protecting groups was observed.

On the other hand, when mass spectral analyses were performed in positive mode, results were more complex, although they did not provide much extra useful information.

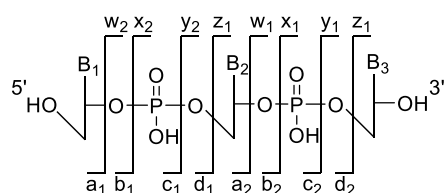


Figure 2.18. Possible ON fragmentations and their nomenclature.

We envisaged that this could be a good method to determine if fragmentation products were already present after the DA reaction and before the final deprotection: any mass corresponding to the lack of nitrogenous bases and the peptide and cycloadduct fragments would indicate that the cleavage process had occurred during the conjugation step.

Our first step was to assay this technique (and small modifications of this) with ONs bearing UltraMild protecting groups, in particular the 7-mer ON chains 5'-d(TAGCTCT)-resin (**2.40**) and 5'-diene-d(TCATGCT)-resin (**2.41**), as they would allow us to perform the MALDI analyses in the reflector mode, giving us more accuracy than the linear mode (see structures in **Figure 2.19**).

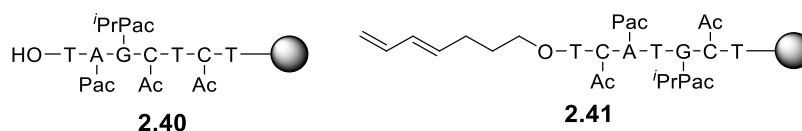


Figure 2.19. Structures of products **2.40** and **2.41**.

Samples were prepared as described by the Vasseur group, as small modifications of their protocol did not give better results. On the other hand, we found that a change in the matrix composition (from 92:8 w/w THAP/CA to 17:83) ameliorated the results obtained with the one described, as high masses were observed with higher intensity and using lower laser energies.

As can be seen in **Figure 2.20**, not only the mass of a given fragment was observed, but also three other main peaks could be distinguished. One, with mass a 42 Da lower, could correspond to the loss of an Ac protecting group. However, the fact that ON fragments that did not contain dC also showed this peak while fragments not containing dG did not present this behaviour pointed to the ⁱPrPac-dG protecting group being replaced by Pac, as a result of an exchange with the capping agent.³¹ The other two main peaks present corresponded to the loss of a CNE group from both the original ON and that in which ⁱPrPac-Pac exchange had taken place. Occasionally, loss of other protecting groups such as Pac or ⁱPrPac could be detected in very low intensity, and addition of an extra Pac group, arising from nucleobase acylation with the capping agent, was sometimes observed in low intensity (not observed in **Figure 2.20**).

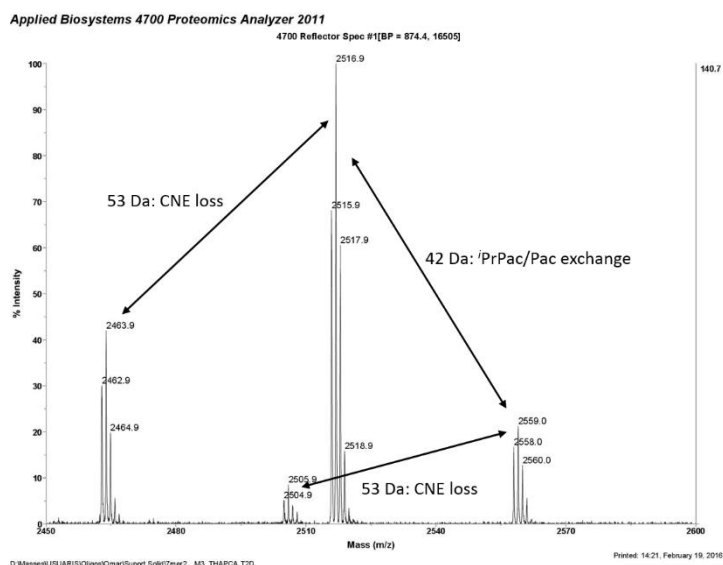


Figure 2.20. Mass spectrum showing the four major peaks observed for the TAGCTC “d” fragment when analysing the solid support by MALDI-TOF MS. The peak with a mass of 2558.0 Da corresponds to the fragment with all the protecting groups (and no iPrPac-Pac exchange).

As we planned to control the process also after the CNE removal treatment, we decided to perform the same experiment with an ON that had been subjected to the DEA treatment. We observed almost complete removal of the CNE protecting groups, as only peaks corresponding to fragments bearing 0 or 1 CNE could be detected. Curiously, removal of the exocyclic amine protecting groups was not detected. This was rather intriguing because Strömberg had described²⁷ that washings with morpholine, reduced the percentage of acylation by *i*Bu and Bz, meaning that at least some of this groups had to be removed during said treatment. Based on this report and the fact that we were using much more labile protecting groups, we expected to observe partial loss of Pac and *i*PrPac groups even though DEA is somewhat less nucleophilic than morpholine. With all this information in hand, we moved to analyse by MALDI-TOF MS of the solid support the outcome of the conjugation reaction between **2.14** and **2.41**. This time, the morpholine treatment was not performed in order to prevent any loss of protecting groups, which would generate a very complicated mass spectrum. As can be seen in **Figure 2.21**, many peaks whose mass corresponded to ON backbone cleavage could be detected (red marks). This was the first direct evidence that cleavage of the oligonucleotide backbone had taken place during conjugation, explaining the fact that the different CNE removal processes performed after the DA reaction did not provide different results.

Aside from the previously observed exchange of *i*PrPac for Pac and loss of a CNE group, transamidation reactions between the ON and the peptide chains could be detected. This was indicated by the fact that several Pac-acylated conjugate fragments (violet marks in **Figure 2.21**) were detected, along with the intact conjugate (green marks) and ON cleavage products lacking the Pac group. This result was also confirmed after cleavage and inspection of the HPLC and the MALDI-TOF MS of the collected peaks, which corroborated that the conjugate (**2.42**) had been partially acylated in a permanent manner.

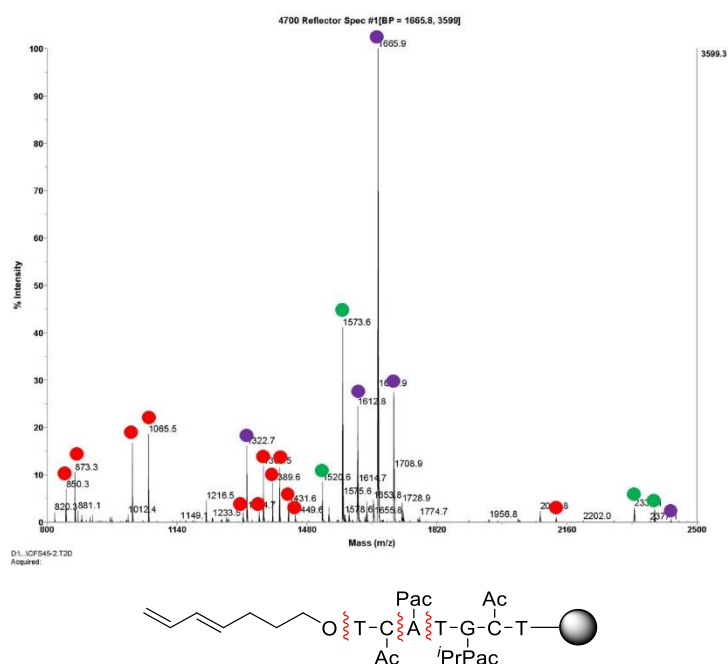


Figure 2.21. MALDI-TOF mass spectrum of the solid support after conjugation reaction. Peaks marked in red correspond to ON backbone cleavage during the DA reaction. Green and violet marked peaks are the conjugate and Pac-acylated conjugate, respectively. The fragmentation points are represented by a red wavy line.

On the other hand, no acetylation products were observed by MALDI-TOF MS of the solid support, but their presence could be detected after cleavage.

2.2.6 Assessment of the fragmentation in the conjugation reaction conditions

To further confirm the results obtained with the mass spectrometric analyses of the solid support, that is, that fragmentation was taking place during the DA reaction, two different experiments were carried out. First of all, the resin-linked diene-derivatised oligonucleotide **2.41** was incubated without **2.14** in the reaction conditions to ensure that the maleimido-peptide was not the cause of fragmentation. This experiment was carried out at slightly acidic, neutral and slightly basic pHs (5.5, 7.0 and 8.3) in order to discard any base or acid catalysed mechanism. The results of these experiments, in which the fragmentation of the ON chains accounted for 15-25 % of the crude, suggested that there was not a clear relationship between pH and fragmentation, as both the slightly basic and slightly acidic pHs yielded higher fragmentation levels than neutral pH. More importantly, they confirmed that the conditions employed for the DA conjugations were harming the oligonucleotide even in the absence of the basic amine groups of **2.14**. To ensure that fragmentation was not induced by the CPG support, the cleaved and totally deprotected d(TCATGCT) (**2.43**) was incubated for 24 h at 65 °C in the presence and absence of CPG beads, observing in both cases that fragmentation had taken place (actually CPG beads seemed to protect the ON from decomposition). Temperature was found to be the actual problem leading to the ON fragmentation, as incubation of the same ON at 37 °C did not produce any appreciable side-reaction after 24 h.

However, a clear sequence dependence was found when comparing the fragmentation of a diene-derivatised, resin-linked dT₁₀ (**2.44**) and **2.41**, showing that poly-dT ONs were more stable than those containing all the bases.

A last attempt to reduce the ON backbone cleavage, consequently, was performed lowering the reaction temperature and increasing the time span. Again, **2.41** was incubated with **2.14** in water, this time at 40 °C for 48 h. Unfortunately, the outcome of the reaction did not vary significantly, with fragmentation products still accounting for almost 15 % of the crude.

In view of these results, all pointing to the impossibility of avoiding the fragmentation reaction, we decided to stop working on this project. We thought that the fact that not even the deprotected ON could withstand the reaction conditions indicated that it would not be possible to circumvent the fragmentation problem. Although the ON backbone cleavage was not extremely high, one must take into account that we had been working with really short sequences, and that for commonly used sequences, about or longer than 20-mer, the problem would probably increase.

2.3 Critical overview of the methodology developed

The developed methodology for polyamide conjugation has several positive aspects and offers certain advantages over other processes:

- 1- The reaction tolerates a wide range of molecule polarities and can be performed either in pure water or in mixtures with organic solvents.
- 2- Good conjugate yields are obtained using fairly mild conditions and little reagent excesses.
- 3- Compared to methodologies carried out in solution, one purification step is avoided, as the diene-derivatised polyamide is reacted directly after its synthesis. This, apart from reducing the number of these tedious and time-consuming processes, entails an increase in the overall yield.
- 4- All side-chain functional groups of the resin-linked polyamide are protected, avoiding any undesired reaction with the soluble maleimide-derivative.
- 5- The polyamide synthesis is performed using standard methods and commercial reagents, eliminating all restrictions imposed by the previously described methodologies and saving the time and resources needed for the preparation of special amino acid derivatives.
- 6- As the conjugate will have a very different molecular weight and size compared to the unreacted diene-derivatised polyamide and all the by-products generated during solid-phase synthesis, purification of the conjugate should be, in principle, quite easy.
- 7- The adduct formed is stable and does not reverse during routine handling of the product or under the mass spectrometric analyses conditions.

On the negative side, though, some drawbacks must be taken into account:

- 1- More than one isomer is formed during the DA reaction, and their separation is normally not possible due to the size of the conjugated molecules.
- 2- The methodology is limited to the use of molecules compatible with the final acidic treatment required for polyamide deprotection and cleavage. Among the molecules that would not withstand this process, oligonucleotides (ONs), which tend to undergo depurination reactions in acidic media, are the most relevant ones, impeding the synthesis of peptide-oligonucleotide conjugates (POCs).
- 3- Relatively long reaction times are needed.
- 4- The synthesis of POCs cannot be performed employing this strategy. Neither the use of resin-linked diene-derivatised oligonucleotides nor that of diene-modified peptides is adequate to furnish the desired conjugates in good yields.

2.4 Abbreviations

Ac: Acetyl

AC: Ammonium citrate

CED: Cohesive Energy Density

CNE: Cyanoethyl

CPG: Controlled Pore Glass

CPP: Cell-Penetrating Peptide

DCM: Dichloromethane

DEA: Diethylamine

DIPEA: *N,N*-Diisopropylethylamine

DMF: Dimethylformamide

DTT: Dithiothreitol

EDT: Ethanedithiol

HOBt: 1-hydroxybenzotriazole

HPLC: High Performance Liquid Chromatography

HPLC/MS: High Performance Liquid Chromatography-Mass Spectrometry

ⁱPrPac: 4-isopropyl-phenoxyacetyl

LPFC: Liquid-Phase Fragment Conjugation

MALDI-TOF MS: Matrix-Assisted Laser Desorption Ionization-Time Of Flight Mass Spectrometry

MT: Michael-Type

ON: Oligonucleotide

Pac: phenoxyacetyl

PEG: Polyethylenglycol

PS: Polystyrene

PyBOP: benzotriazole-1-yl-oxy-trispyrrolidinophosphonium hexafluorophosphate

SPFC: Solid-Phase Fragment Conjugation

SPPS: Solid-Phase Peptide Synthesis

TEA: Triethylamine

THAP: 2,4,6-trihydroxyacetophenone

Tfa: Trifluoroacetate

TFA: Trifluoroacetic acid

TIS: Triisopropylsilane

2.5 Bibliography

- (1) Hoogewijs, K.; Deceuninck, A.; Madder, A. *Org. Biomol. Chem.* **2012**, *10* (20), 3999.
- (2) Doyle, M. P.; DeBruyn, D. J.; Donnelly, S. J.; Kooistra, D. A.; Odubela, A. A.; West, C. T.; Zonnebelt, S. M. *J. Org. Chem.* **1974**, *39* (18), 2740.
- (3) Blaskovich, M. A.; Kahn, M. *J. Org. Chem.* **1998**, *63* (4), 1119.
- (4) de Araújo, A. D.; Palomo, J. M.; Cramer, J.; Seitz, O.; Alexandrov, K.; Waldmann, H. *Chem. – Eur. J.* **2006**, *12* (23), 6095.
- (5) Hoogewijs, K.; Buyst, D.; Winne, J. M.; Martins, J. C.; Madder, A. *Chem. Commun.* **2013**, *49* (28), 2927.
- (6) Boutelle, R. C.; Northrop, B. H. *J. Org. Chem.* **2011**, *76* (19), 7994.
- (7) Rideout, D. C.; Breslow, R. *J. Am. Chem. Soc.* **1980**, *102* (26), 7816.
- (8) Otto, S.; Engberts, J. B. F. N. *Org. Biomol. Chem.* **2003**, *1* (16), 2809.
- (9) Otto, S.; Blokzijl, W.; Engberts, J. B. F. N. *J. Org. Chem.* **1994**, *59* (18), 5372.
- (10) Blake, J. F.; Lim, D.; Jorgensen, W. L. *J. Org. Chem.* **1994**, *59* (4), 803.
- (11) Otto, S.; Engberts, J. B. F. N. *Pure Appl. Chem.* **2000**, *72* (7), 1365.
- (12) Pirrung, M. C. *Chem. – Eur. J.* **2006**, *12* (5), 1312.
- (13) Lubineau, A. *J. Org. Chem.* **1986**, *51* (11), 2142.
- (14) Engberts, J. B. F. N. *Pure Appl. Chem.* **1995**, *67* (5).
- (15) Blokzijl, W.; Blandamer, M. J.; Engberts, J. B. F. N. *J. Am. Chem. Soc.* **1991**, *113* (11), 4241.
- (16) Sherrington, D. C. *Chem. Commun.* **1998**, No. 21, 2275.
- (17) Tam, J. P.; Lu, Y.-A. *J. Am. Chem. Soc.* **1995**, *117* (49), 12058.
- (18) Baillie, L. C.; Batsanov, A.; Bearder, J. R.; Whiting, D. A. *J. Chem. Soc. [Perkin 1]* **1998**, No. 20, 3471.
- (19) Good, L.; Awasthi, S. K.; Dryselius, R.; Larsson, O.; Nielsen, P. E. *Nat. Biotechnol.* **2001**, *19* (4), 360.
- (20) Rodionov, I. L.; Peshenko, I. A.; Baidakova, L. K.; Ivanov, V. T. *Int. J. Pept. Res. Ther.* **2007**, *13* (1–2), 161.
- (21) Marchán, V.; Ortega, S.; Pulido, D.; Pedroso, E.; Grandas, A. *Nucleic Acids Res.* **2006**, *34* (3), e24.
- (22) Peyrottes, S.; Mestre, B.; Burlina, F.; Gait, M. J. *Tetrahedron* **1998**, *54* (41), 12513.
- (23) Peyrottes, S.; Mestre, B.; Burlina, F.; Gait, M. J. *Nucleosides Nucleotides* **1999**, *18* (6–7), 1443.
- (24) Kubo, T.; Morikawa, M.; Ohba, H.; Fujii, M. *Org. Lett.* **2003**, *5* (15), 2623.

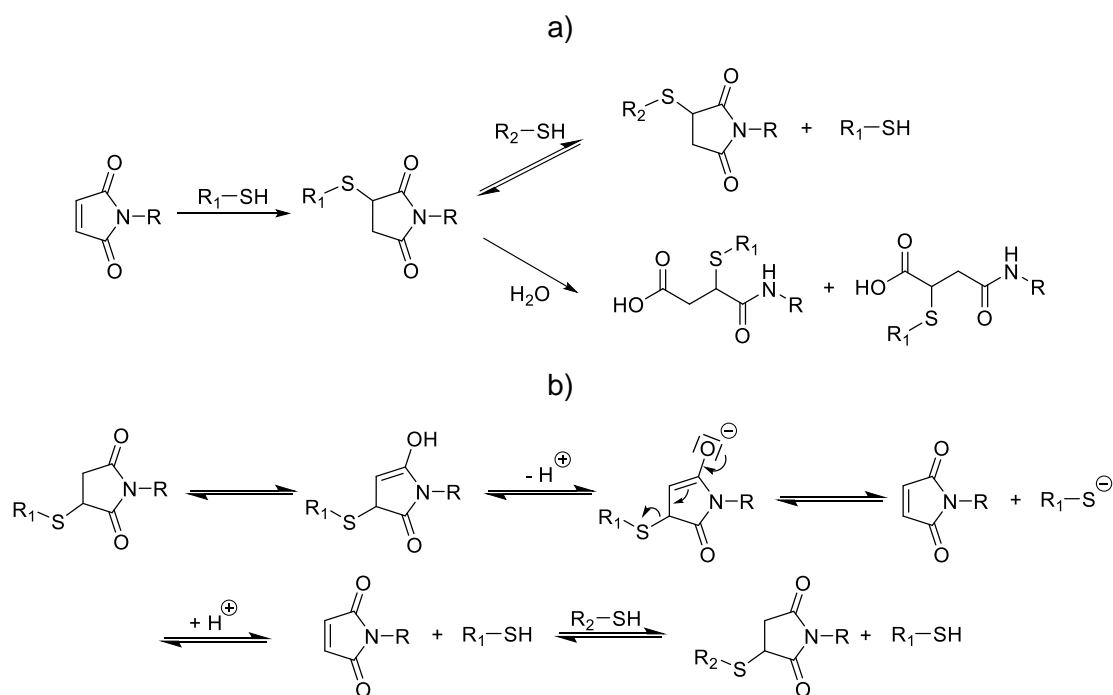
- (25) Wenska, M.; Alvira, M.; Steunenberg, P.; Stenberg, A.; Murtola, M.; Stromberg, R. *Nucleic Acids Res.* **2011**, *39* (20), 9047.
- (26) Tetzlaff, C. N.; Schwope, I.; Bleczinski, C. F.; Steinberg, J. A.; Richert, C. *Tetrahedron Lett.* **1998**, *39* (24), 4215.
- (27) Zaramella, S.; Yeheskiely, E.; Strömberg, R. *J. Am. Chem. Soc.* **2004**, *126* (43), 14029.
- (28) Glen Research Product Information
http://www.glenresearch.com/Technical/TB_Avoid_Amine_Alkylation.html.
- (29) Meyer, A.; Spinelli, N.; Imbach, J.-L.; Vasseur, J.-J. *Rapid Commun. Mass Spectrom.* **2000**, *14* (4), 234.
- (30) Guerlavais, T.; Meyer, A.; Debart, F.; Imbach, J.-L.; Morvan, F.; Vasseur, J.-J. *Anal. Bioanal. Chem.* **2002**, *374* (1), 57.
- (31) Zhu, Q.; Delaney, M. O.; Greenberg, M. M. *Bioorg. Med. Chem. Lett.* **2001**, *11* (9), 1105.

Chapter 3. 2,2-Disubstituted cyclopent-4-ene-1,3-diones: more than simple Michael acceptors

3.1 Background and objectives

The Michael-type reaction between a thiol and a maleimide is one of the most commonly used reactions for bioconjugations (see section 1.3.5). Its usefulness relies on the specificity of maleimides for thiols, the fast kinetics of the reaction, its water compatibility, lack of by-products and the stability of the generated Michael-type adduct (MTA).¹ The reaction has been widely applied in the biochemistry field due to the natural presence of thiols in biomolecules such as peptides, proteins and antibodies, allowing the use of these entities without modification and avoiding the need to introduce a suitable reacting group. The thiol-maleimide reaction has been applied for PEGylation,² fluorescent labelling,³ conjugation,⁴ etc. Remarkably, it has also been used for the synthesis of several ADCs, some of which are commercial or in clinical trials.^{5,6}

Despite its positive aspects, there are two main problems with the thiol-maleimide reaction: thiol-exchange and hydrolysis of the generated succinimide (see **Scheme 3.1a**).



Scheme 3.1. The two side reactions taking place on maleimide Michael-type adducts (MTAs) (a) and mechanism for the retro Michael-type reaction and thiol exchange (b).

Thiol exchange is an important problem, especially when the synthesised conjugates are intended to have a therapeutic application, as for example ADCs. Antitumor activity may diminish due to reduced exposure to the antibody-conjugated form of the drug, while toxicity may increase due to the premature cleavage of the drug-linker union, which may liberate the cytotoxic drug far from the targeted tissue.^{7,8} The mechanism of the thiol exchange reaction is depicted in **Scheme 3.1b**, and occurs through the retro-Michael reaction of the generated MTA. Reversion of said product is not detectable unless other thiols are present, indicating that the position of the equilibrium in the Michael-type reaction is largely displaced towards product formation.¹ Reversion is strongly influenced by the thiol pK_a : the more acidic the thiol, the more easily the MTA will revert (as more stable the leaving thiolate will be).¹ Also, the concentration of thiols will affect the half-life of the MTA, according to the general rules of an equilibrium reaction. Practically, this means that thiol exchange will be important only inside cells, where concentrations of free thiols such as glutathione and cysteine are high, whereas during circulation MTAs will be essentially stable.¹

A possible way to avoid thiol exchange is to favour the other naturally occurring side-reaction, that is, hydrolysis. As the pK_a of the hydrolysed succinimide (see structure in

Scheme 3.1a) is considerable higher than that of the cyclic counterpart, formation of the enolate that leads to the retro-Michael reaction is even less favourable.⁷ Several groups have studied this process and proposed different strategies to favour the hydrolysis reaction and avoid thiol exchange. Raines and co-workers, for example, studied several anions capable of catalysing the ring-opening reaction.⁹ Other groups decided to modify the structure of the linker at the *N* position of the maleimide ring^{8,10} or to directly incubate the generated MTAs at basic pH to ensure hydrolysis.⁷

All these studies demonstrated that, basically, the stability of the succinimide ring towards hydrolysis depends on the substituent at the *N* position.¹⁰

Albeit hydrolysis can solve the thiol exchange problem, it also presents several problems on its own. Firstly, intentional hydrolysis of MTAs and can be difficult to drive to completion, it is normally time-consuming and the time needed is substrate-dependent.¹⁰ On the second place, side-reactions can occur during this process, as for example deamidation of asparagine residues,⁷ and OH⁻ anions present in the basic media can trigger a decrease in the drug-to-antibody ratio (DAR).⁷ Furthermore, intentional hydrolysis of the MTA increases the number of synthetic and purification steps required for the synthesis of a given compound, with the corresponding negative effects on the final yield. Importantly, as the nucleophilic attack of the hydroxy anion is not regioselective, the opening of the succinimide ring furnishes two constitutional isomers that might complicate analysis and purification of the target compound in addition to having different biochemical properties. Finally, ring-opening of the succinimide generates a carboxylic function that will be negatively charged at physiological pH, which can negatively affect cellular internalization in those cases where this depends on the cationic charge.

Given the inherent problems of hydrolysis and that thiol exchange may not be intrinsically negative (it can be regarded as a platform for controlled drug release), we decided to develop a non-hydrolyzable maleimide analog and study its stability and thiol exchange behaviour.

Our idea was to use cyclopent-4-ene-1,3-diones, which can be formally regarded as maleimide analogues according to their structural resemblance, as an alternative to maleimides (**Figure 3.1a**). Substituting the nitrogen in the maleimide ring by a carbon atom should avoid addition of water to the MTA and, in principle, not worsen the reactivity properties regarding the Michael-type and Diels-Alder (DA) transformations. Actually, we believed that both the Michael-type and the DA reactions should be more favoured in the case of cyclopent-4-ene-1,3-diones due to the more electro-withdrawing effect of the two keto groups compared to the imido functionality. On the other hand, though, replacement of nitrogen with a carbon atom also entailed the introduction of additional substituents.

The first step was, then, to decide whether one of these substituents would be a hydrogen atom or not, that is, if it was better to use mono- or di-substituted cyclopent-4-ene-1,3-diones (**Figure 3.1b**).

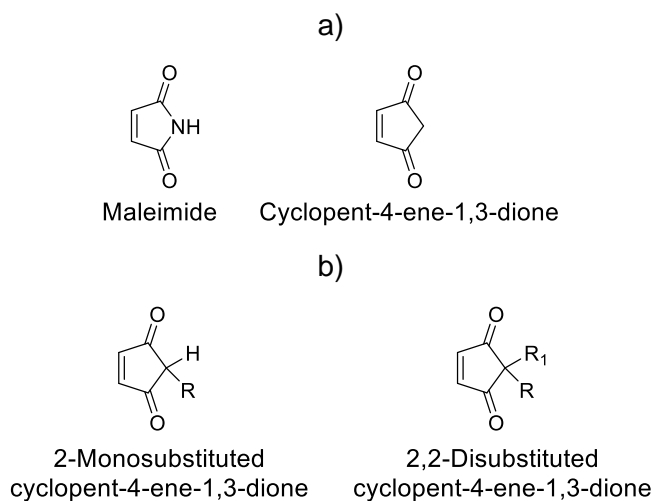


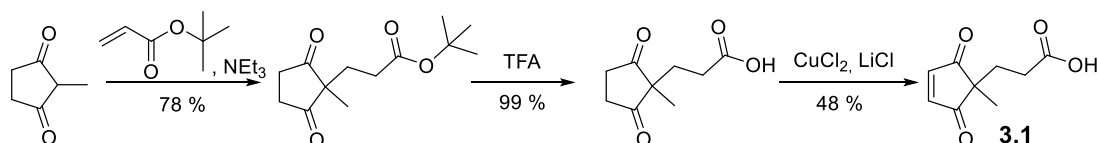
Figure 3.1. Basic structures of maleimides and cyclopent-4-ene-1,3-diones (a), and the corresponding mono- and disubstituted cyclopent-4-ene-1,3-diones (b).

We surmised that using 2,2-disubstituted cyclopent-4-ene-1,3-diones (CPDs) would be advantageous as it would avoid enolization of the 2 position, which was envisaged as somewhat problematic due to the possibility of triggering the polymerization of the cyclopentendione and secondary reactions of the Michael-type and DA adducts.

Since our intention was to use CPDs for the conjugation of peptides and possibly other polyamides, the molecule had to incorporate a carboxylic derivative to allow for its introduction into polyamide chains by standard SPPS methods. Although previous work described the synthesis of some CPDs,¹¹ we decided to synthesise derivative **3.1** (see **Scheme 3.2**) because its synthesis started from cheap and commercial materials, and seemed easy and short. The two methylene groups spacer between the carboxylic acid and the five-membered ring was considered long enough to avoid steric hindrance on the conjugated double bond, as judged from the experience with 3-maleimidopropanoic acid, and allowed to perform a simple Michael reaction to introduce the masked acid functionality. As expected, the synthesis of **3.1** was simple and, after small adjustments, could be performed with only one chromatographic purification step.

Previous experiments to obtain CPDs similar to **3.1** involving alkylation of 2-methylcyclopentan-1,3-dione with haloalkanes rendered mixtures of *O*- and *C*-alkylated products and were, consequently, abandoned (data not shown).

As depicted in **Scheme 3.2**, 2-methyl-1,3-cyclopentanedione was reacted with *tert*-butyl acrylate in refluxing TEA to yield the Michael adduct in 78 % yield. This was reacted with TFA to generate the carboxylic acid from the *tert*-butyl ester in 99 % yield prior to oxidation of the α position with CuCl_2 .¹² Generation of the conjugated double bond in **3.1** was the only step requiring chromatographic separation, and was also the process with the lowest yield of the sequence (48 % yield). Inversion of the oxidation and acidolysis steps did not render better results.



Scheme 3.2. Synthesis of the CPD carboxylic acid derivative **3.1**.

The synthesis of derivative **3.1** is more straightforward than that of the CPD analogue previously described and does not require any special equipment (such as an ozone generator).¹¹

3.2 Examining CPDs reactivity

With **3.1** in our hands, we moved to assess its reactivity.

We first assayed the DA reaction of **3.1** with diene **2.2** (see section **2.1.8**) in water. Albeit the reaction was left stirring at 37 °C for a week, no product could be detected by NMR. A possible explanation for this was found in the literature: a report dealing with the use of CPDs pointed towards steric hindrance precluding the DA reaction.¹³ The kinetically favoured *endo* approximation, which was even more favoured by the solvent (section **2.1.5**), was congested due to the interaction of the substituents at the 2 position of the CPD ring with the dienophile, as depicted in **Figure 3.2**. Furthermore, most probably the *exo* approximation, which would circumvent this problem, required too much energy to take place in the tested conditions.

As judged from the mentioned report, the reaction would require too harsh conditions to be used for the conjugation of biomolecules, which made us rule out any further attempts to make the reaction happen.

After this result, we moved to the Michael-type reaction, which was, actually, our main objective.

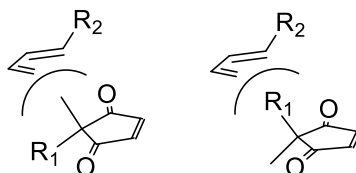


Figure 3.2. The endo approximations of a CPD towards a diene are sterically hindered.

The analytical methodology to study the reaction outcome was, as in chapter 2, HPLC analysis of the crude and MALDI-TOF MS of the collected peaks.

In a couple of preliminary experiments, peptide H-Lys-Tyr-Ala-Tyr-Cys-Gly-NH₂ (**3.2**) containing a cysteine at an internal position and peptide H-Cys-Tyr-Gly-NH₂ (**3.3**) containing this amino acid at the *N*-terminal position were reacted with 5 equivalents of CPD **3.1** at pH 7.8. To our surprise, the outcome of the reaction with peptide **3.2** was completely different from that of peptide **3.3**, as judged from the HPLC and MALDI-TOF MS analyses, and, therefore, we decided to study each situation independently.

3.2.1 Reaction of CPDs with peptides containing internal or C-terminal cysteines

Peptide **3.2** was reacted at room temperature with five equivalents of the CPD derivative **3.1** in 0.1 M phosphate buffer at pH 7.8 (typical conditions for the Michael-type addition of a thiol to a maleimide). HPLC Analysis of the reaction crude after 1 h showed two different peaks instead of the four expected for four MTAs (**3.4**,* see **Scheme 3.3**), likely resulting from the co-elution of some of the products. More importantly, only a half of the starting peptide had reacted (**Figure 3.3a**), which, as mentioned before, was surprising as the Michael-type reaction of maleimides is fast. For this reason, temperature was increased to 45 °C after a total of two hours reaction time.

4.5 h Later, the crude was analysed again to find that considerable formation of the peptide disulfide dimer had taken place, and that there was still an important amount of unreacted reduced peptide (**Figure 3.3b**). More surprising than the fact that conversion was still low was to find, after integration of the corresponding peaks, that the ratio of MTAs to peptide (monomer + dimer) had decreased with time.

* As the stereoisomeric MTAs were not separated, we refer to all the MTAs formed in the reaction with the same number. This applies to all the MTAs appearing in this chapter.

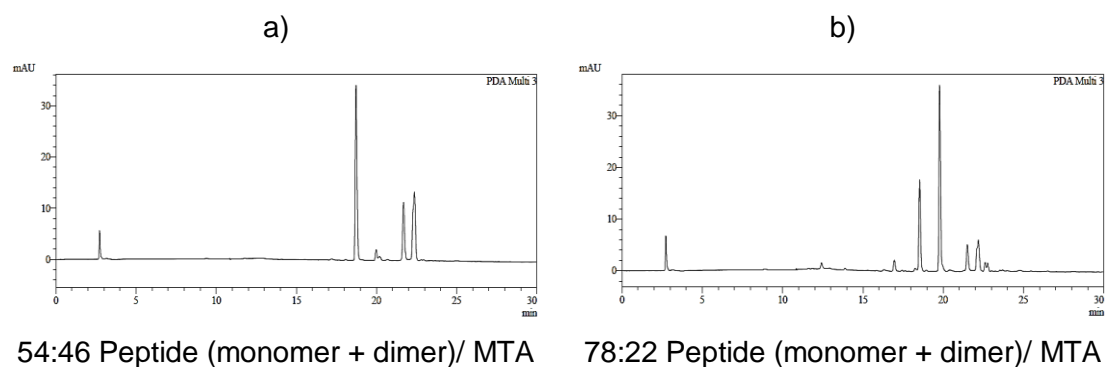
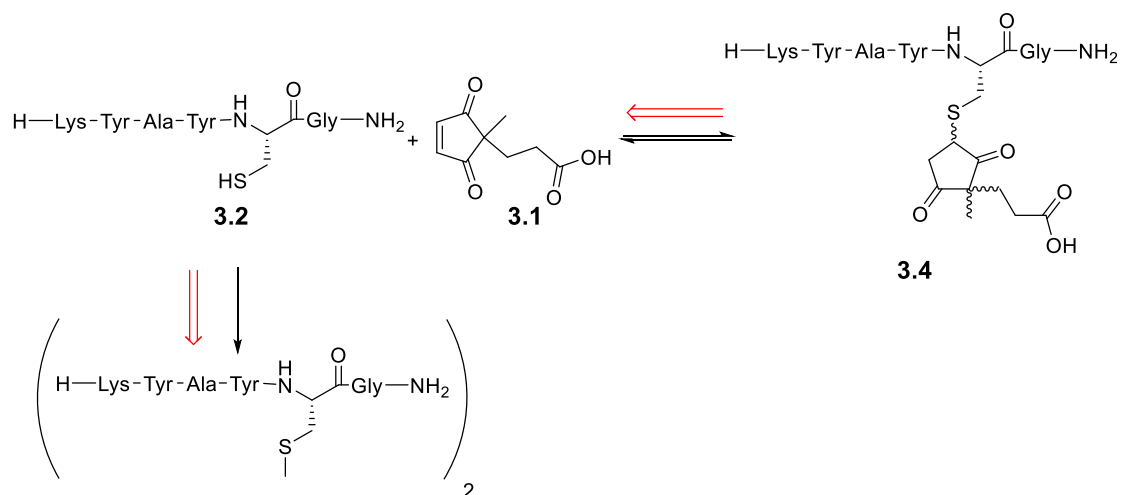


Figure 3.3. HPLC Traces (280 nm, **3.1** is not detected) of the Michael-type reaction between **3.1** and **3.2** after 1 h (a) and 6.5 h (b). Peak at 18.6 min is peptide monomer, peak at 19.8 min is the peptide dimer and the two twin peaks at 21.5 and 22.2 min the MTAs.

As no decomposition products were found in the reaction crude, our only explanation to this fact was to consider that this Michael-type reaction was reversible. According to this hypothesis, decrease of the MTA concentration would have happened, following the principle of Le Châtelier, to compensate for the reduction in peptide concentration caused by peptide disulfide dimer formation (**Scheme 3.3**). The low conversions found in the first analysis could then be explained by the equilibrium position as being displaced towards the starting materials.



Scheme 3.3. The disulfide dimeric form of the peptide increases with time, which triggers reduction of the MTA concentration (following the principle of Le Châtelier).

Another evidence supporting this hypothesis was found when analysing the HPLC peaks corresponding to **3.4** by MALDI-TOF MS. Mass spectra of these peaks revealed the presence of the parent free peptide along with the expected MTA, indicating that **3.4** had reversed either during the process of isolation, lyophilisation and sample preparation, under the MALDI conditions, or both.

If our hypothesis was correct, pure **3.4** should not be stable when removed from the reaction crude, and therefore reversion to the parent compounds (**3.1** + **3.2**) should be observed when incubating any of the isomers after isolation. Indeed, it was observed that after independently collecting each of the peaks, lyophilisation and reinjection in the liquid chromatograph, equilibration to the other peak had taken place (**Figure 3.4a** and c). Free peptide **3.2** and CPD **3.1** could be detected at that point, but only in small amounts. However, incubation of the isolated MTAs at room temperature for 4.5 h triggered their practically total reversion to the starting materials (**Figure 3.4b** and d), corroborating the fact that the Michael-type reaction of thiols and CPD **3.1** is reversible, and that the equilibrium position is greatly displaced to the left.

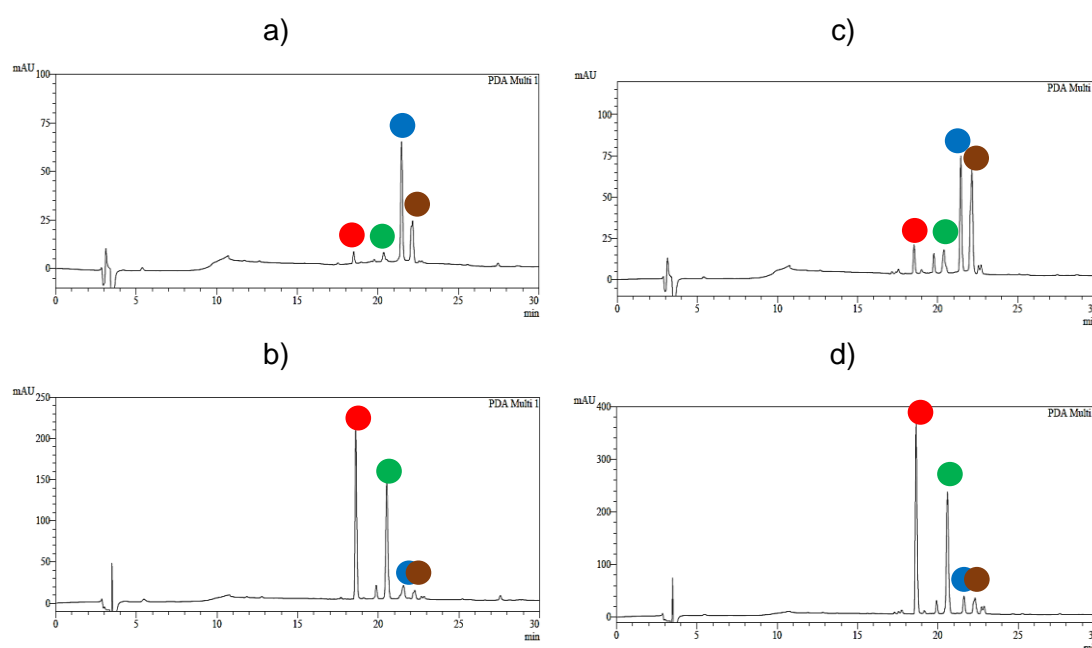


Figure 3.4. HPLC Traces (220 nm, **3.1** is detected) obtained upon analysing the collected MTA peak eluting at 21.5 min (blue mark, see **Figure 3.3**) immediately after lyophilisation (a) and after 4.5 h incubation in water (b). Profiles (c) and (d) show the corresponding results, respectively, for the MTA peak eluting at 22.2 min (brown mark). Red and green marks are the starting peptide and **3.1**, respectively.

To further confirm and gain insight into the reversibility of the MTAs obtained from CPDs and peptides containing non *N*-terminal cysteines, reactions of **3.2** and H-Trp-Gly-Arg-Gly-Cys-NH₂ (**3.5**) with 2,2-dimethylcyclopent-4-ene-1,3-dione (**3.6**), as well as the stability of the corresponding MTAs (**3.7** for peptide **3.2** and **3.8** for peptide **3.5**), were studied. Use of CPD **3.6** was expected to be advantageous because it would render less diastereomeric MTAs and, at the same time, discard any influence of the substituent at the 2 position of the CPD. This time, both the direct and inverse reactions were conducted in pure water (pH~6) to reduce the extent of disulfide formation. In both cases peptides were only partially consumed, as expected for an equilibrium reaction, and **3.7** and **3.8** were found to reverse to the parent compounds to reach the equilibrium point in less than 1 h (although reversion might have started during the isolation process).

To summarize, all these experiments demonstrated that the Michael-type reaction of peptides containing internal or *C*-terminal cysteines with CPDs is reversible and that, contrary to the case of maleimides, whose MTAs can be regarded as basically irreversible if no thiols are present, the equilibrium is greatly displaced to the starting materials.

3.2.2 Reaction of CPDs with peptides containing *N*-terminal cysteines

The reaction of peptide **3.3**, containing an *N*-terminal cysteine, with 5 equivalents of **3.1** neither reproduced the results found with peptides **3.2** and **3.5**, nor those of thiol-maleimide reactions.

HPLC Analysis of the crude at 280 nm, where CPD cannot be detected, revealed the presence of peptide dimer, which was already observed in the starting material, and the appearance of four new peaks, but this time no unreacted peptide could be detected.

These four peaks seemed to be split in two different sets (* and #), whose relative ratios surprisingly changed with time (**Figure 3.5**) and depending on the reaction conditions (data not shown). We discovered that higher temperatures favoured the set with a shorter retention time (set * in **Figure 3.5**), and that made us wonder if we were facing a reaction that could give either kinetically or thermodynamically favoured products. More shocking was to find, after MALDI-TOF MS of the collected products, that none of them had the mass expected for a MTA (further referred to as **M** Da), but one corresponding to a product with a mass 20 Da lower (**M-20** Da).

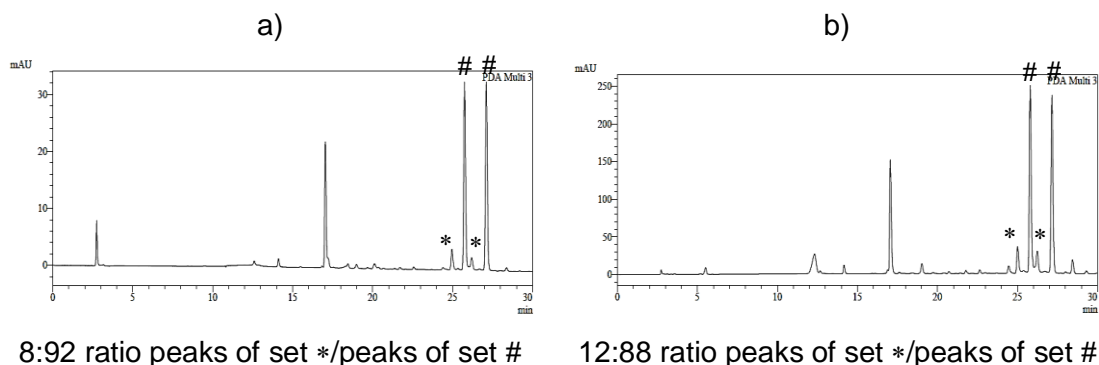
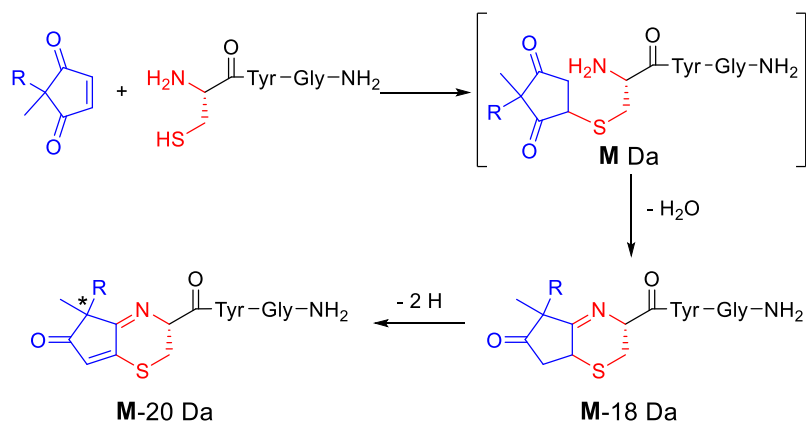


Figure 3.5. HPLC traces (280 nm) of the **3.1** + **3.3** reaction after 2 h (a) and 6 h (b).

Peak set *: $t_R=25.0$ and 26.2 min; peak set #: $t_R=25.7$ and 27.1 min. Peak at 17.0 min is peptide dimer, **3.1** is not detected at this wavelength.

A fact we had not previously considered was that replacement of the nitrogen atom of the maleimide ring with a carbon would also entail that the carbonyl groups of the resulting CPD would be more prone to nucleophilic attacks. In CPDs there is no electron delocalisation from the imide N atom, and thus the carbonyl groups are more electrophilic. Close inspection of the CPD-generated MTA structure revealed that intramolecular imine formation would be highly favoured since it would generate a six-membered ring, as depicted in **Scheme 3.4**. This would displace the equilibrium of the Michael-type reaction into a new product, accounting for quantitative conversions. Assuming that this reaction was potentially feasible, only two H atoms should be eliminated to explain the observed **M-20** Da mass. Oxidation of the 4 and 5 positions of the CPD ring would generate a highly conjugated adduct whose formation would be, in principle, energetically favoured (see **Scheme 3.4**). Therefore, we hypothesised that the observed **M-20** Da products could be the result of the sequential Michael-type reaction, intramolecular imine formation and oxidation of the CPD 4 and 5 positions. These last reactions would eliminate a water molecule (18 Da) and two hydrogen atoms (2 Da), explaining the loss of 20 Da in relation to the MTA. Of course, an alternative reaction path in which oxidation took place before cyclization could also explain the formation of an **M-20** Da adduct.



Scheme 3.4. Hypothetical reaction path to explain formation of the observed **M-20** Da adduct.

Unfortunately, this mechanistic proposal was not able to explain the formation of four products with the same **M-20** Da mass, as only one new chiral centre was generated upon formation of the **M-20** Da adduct (marked with an asterisk in **Scheme 3.4**) and therefore only two diastereomers should be formed and only two peaks should be observed. Furthermore, the fact that the relative ratios of the generated peaks changed with time was still puzzling for us.

To assess whether this behaviour could be reproduced with a different CPD, **3.6** was used instead of **3.1**. The reaction between **3.3** and **3.6** was performed, again with an excess of CPD, to find equivalent results: formation of two peaks with a mass of **M-20** Da when analysed by MALDI-TOF MS. This time, analysis of the reaction crude at different reaction times, that is, after 15 min, 1, 3, 6 and 24 h, allowed us to observe the complete conversion of one of the peaks into the other (**Figure 3.6**). Further confirmation of this behaviour was found when the two peaks were independently collected, lyophilised and reincubated in water. The peak with the smaller retention time was found to be completely stable after a 24 h incubation, while that appearing at 23.6 min was slowly converted into the stable one over the same time period.

Given that these results were puzzling, we decided to change our analytical methodology to gain insight into the reaction evolution. Instead of analysing the crude by HPLC and collecting the peaks to characterize them by MALDI-TOF MS after lyophilisation, we started working with HPLC/MS (ESI). This implied a change in the sample ionization process (ESI rather than MALDI) and allowed us to characterize by MS each peak without the need for isolation, lyophilisation and sample preparation, avoiding any reaction taking place during these processes.

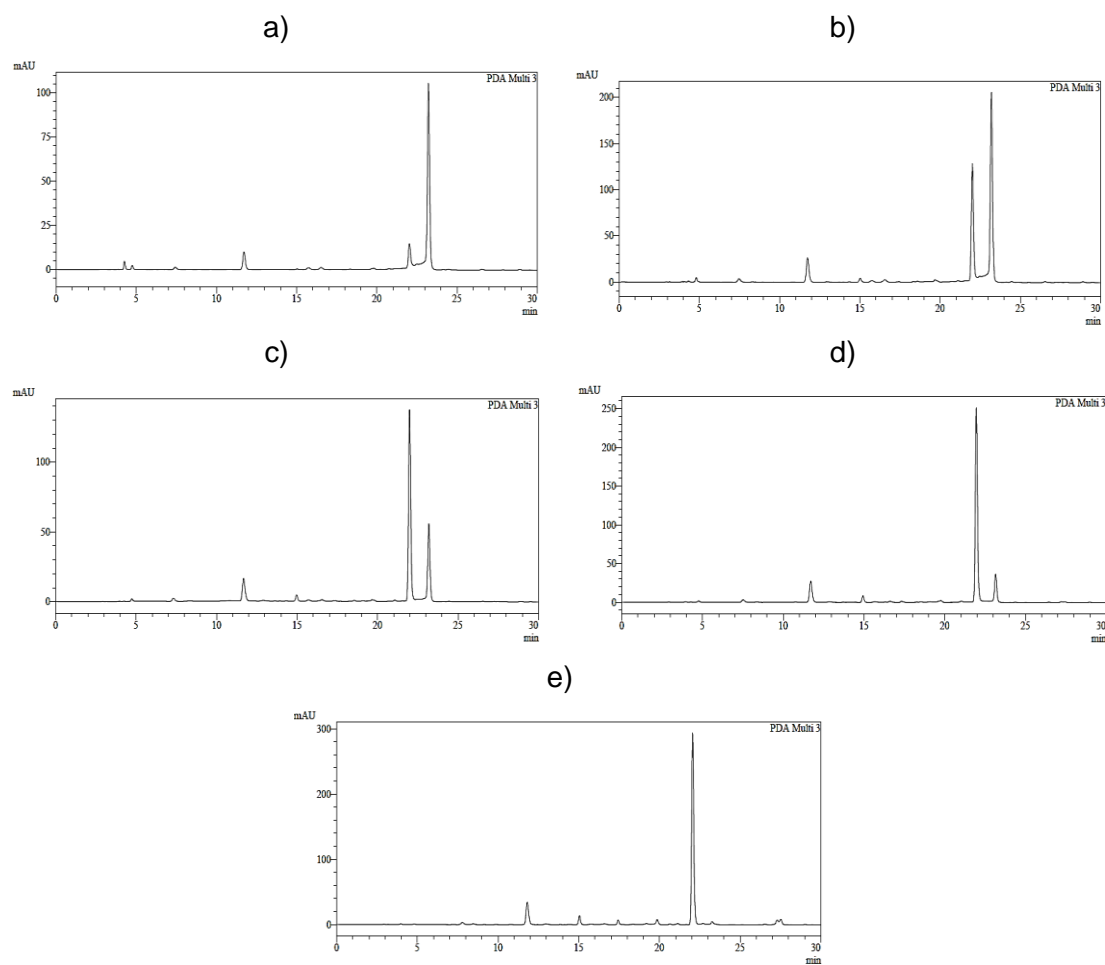


Figure 3.6. HPLC traces (280 nm) of the **3.6+3.3** reaction at 15 min (a), 1 h (b), 3 h (c), 6 h (d) and 24 h (e).

3.3 Was again reacted with **3.6** and the crude analysed after 2 h by HPLC/MS. This was, actually, the experiment that shed light into the process that was taking place. ESI MS analyses indicated that, contrary to what we had previously observed with the MALDI-TOF MS technique, both peaks did not have the same mass. While the stable peak was found to be the expected **M-20 Da** adduct (**3.9**), the other one did not possess the previously observed **M-20 Da** mass but instead a mass of **M-18 Da** (**3.10**). These results were in perfect agreement with our proposed reaction pathway. The **M-18 Da** adduct corresponded to the dehydration product furnished by intramolecular imine formation, and slowly evolved into the final **M-20 Da** adduct through loss of two H atoms. The **M-20 Da** adduct was, then, the final stable compound. This finding explained why we were observing in the case of both CPDs double the amount of expected products: one peak set (one product for **3.6** and two for **3.1**) corresponded to the diastereomeric final products and the other set to the diastereomeric intermediates (again, one product for

3.6 and two for **3.1**). Noteworthy, the MTA was at no time detected, suggesting that it had a really short half-life due to fast intramolecular imine formation.[†]

The disagreement found upon between the two analysis methodologies could be explained by the fact that the MALDI conditions can trigger certain chemical reactions, and, in our case, could be promoting the oxidation of the **M-18** Da adduct. In addition to the possibility that the dehydrated intermediate could undergo oxidation during the isolation, lyophilisation and sample preparation processes, this explains why the **M-18** Da adduct was never detected using the HPLC + MALDI-TOF methodology.

Additional confirmation for the results obtained with the HPLC/MS methodology was found in the UV-Vis profiles of both peaks. The final **M-20** Da adduct exhibited a characteristic broad maximum around 330 nm, which fitted with such a conjugated structure,¹⁴ while the **M-18** Da adduct only showed the typical profile of a tyrosine-containing peptide. This also indicated that they were, indeed, different products and, therefore, fully supports the results obtained by ESI MS.

3.3 Structural determination of the **M-20** Da adduct

After confirming that our hypothesis regarding the reaction path and the structures of the adducts fitted with the experimental HPLC/MS results, it was time to fully characterize the final, stable adduct by NMR.

We decided to work with the simplest possible model of a peptide containing an *N*-terminal cysteine, methyl L-cysteinate, because it would be easier to obtain larger amounts of the corresponding **M-20** Da product and, above everything, the resulting NMR spectra would be easier to interpret. The ester derivative was selected over the amide version due to its commercial availability and low price.

After ensuring that methyl L-cysteinate reacted in the same way as peptide **3.3** with **3.6** by HPLC/MS, the reaction was repeated at a higher scale and the final **M-20** Da adduct (**3.11**) isolated (silica gel chromatography) and characterised by NMR. The ¹H NMR spectrum (**Figure 3.7a**) showed the presence of a single proton in the olefinic zone, which fitted with the sequential Michael-type and oxidation reactions. The chemical shift found for the α carbon moved almost 2 ppm downfield in comparison to that of methyl *N,S*-dimethyl-L-cysteinate in the same solvent,¹⁵ suggesting that the nitrogen atom was participating in a more electron-withdrawing functional group. More informative was the ¹³C NMR spectrum (**Figure 3.7b**). Imine formation was confirmed by a signal around 170 ppm (the other signal in this area corresponded to the ester) and the presence of only

[†] The MTA of a similar reaction could be detected by direct ESI MS analyses as those performed in the case of the reaction between cysteine and **3.6**, although, as in this case, not by HPLC/MS.

one ketone peak at 204 ppm. Oxidation of the 4 and 5 positions of the CPD was evident from the two olefinic carbons situated at 130 and 150 ppm, of which only one had a hydrogen atom (that appearing around 130 ppm, as assessed by DEPT experiments).

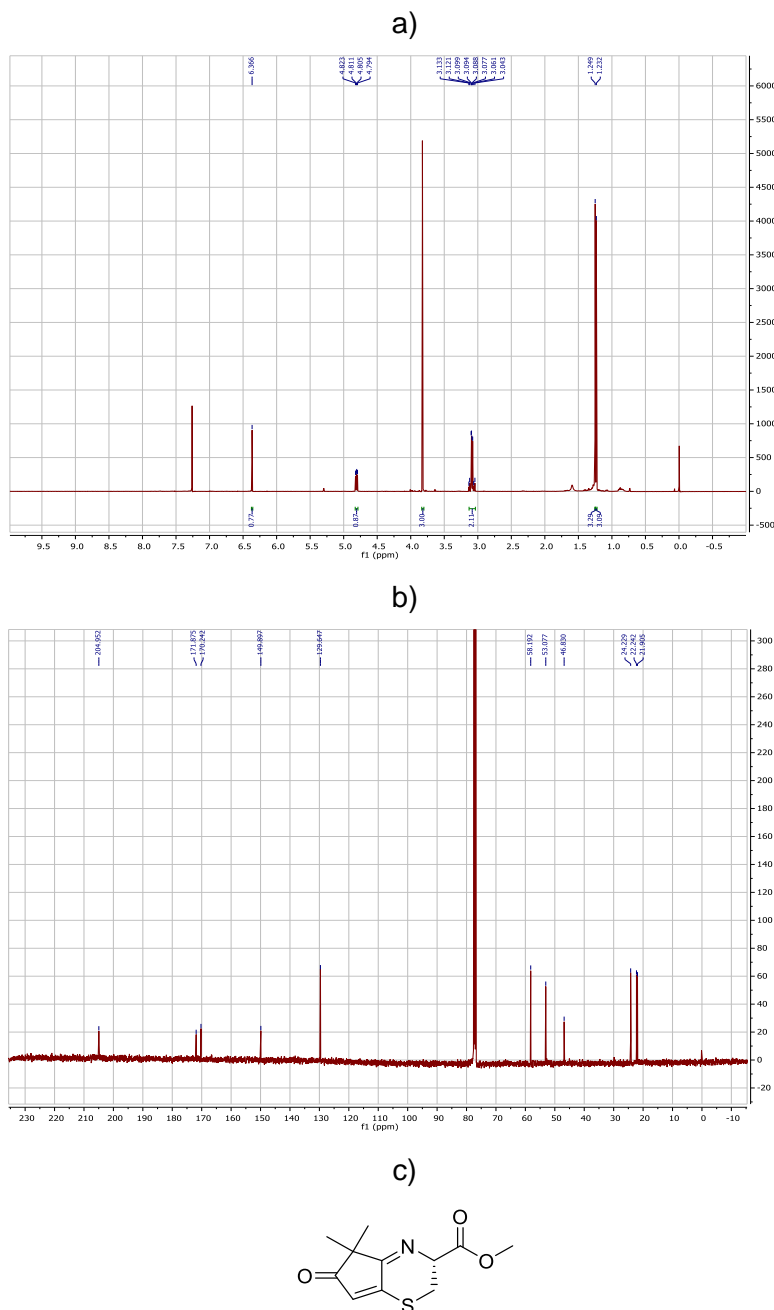


Figure 3.7. ^1H NMR (400 MHz, a) and ^{13}C NMR (101 MHz, b) of the **3.11** M-20 Da adduct in CDCl_3 . The structure of the adduct is shown in (c).

HRMS analysis of the adduct showed the expected mass for **3.11**, and the UV-Vis profile again showed the characteristic maximum around 330 nm.

3.4 Summary of the information available at this point

At this point, it was clear that CPDs do not react as maleimides, which is what we had initially foreseen, but actually present a completely different behaviour towards the DA and Michael-type reactions. The former does not take place in mild conditions probably because of steric hindrance, as mentioned before. Regarding the latter, it definitely takes place but, contrary to what happens with maleimides, the outcome of the reaction depends on the nature of the reacting thiol. Internal or C-terminal cysteines react with CPDs in a reversible reaction to furnish a MTA that is in equilibrium with the parent compounds. All evidences point out to a fast retro-Michael-type reaction and the position of the equilibrium being displaced to the left. In the case of 1,2-aminothiols with the amine free (that is, N-terminal cysteines), the Michael-type reaction does take place but the corresponding adduct is not observed by HPLC. Instead, it immediately evolves into a dehydrated product through intramolecular imine formation, rendering a product with two fused rings and explaining the quantitative conversions observed in this case. This dehydrated intermediate then undergoes oxidation to a final adduct with a mass 20 Da lower than that expected for the MTA, generating a conjugated system with a characteristic maximum around 330 nm.

3.5 Screening of reaction conditions favouring the formation of the M-20 Da adduct

When working with a CPD excess, the **M-18 Da** adduct was always observed after the limiting reagent had completely disappeared together with variable amounts of the **M-20 Da** adduct. This indicated that intramolecular cyclisation to yield the **M-18 Da** adduct was quick and that oxidation to the **M-20 Da** adduct was the rate-limiting step in the overall process. We considered it would be interesting to study which factors were affecting the oxidation process, in order to gain control over it and, hopefully, be able to accelerate it. For this reason, the influence of the solvent composition, temperature, concentration and reagents ratio were studied. General trends for the effect of these variables were obtained by following the reaction between **3.1** and methyl cysteinate by HPLC at 240 nm, where areas of the **M-18 Da (3.12)** and the **M-20 Da (3.13)** peaks were comparable. Due to the tendency of methyl cysteinate to dimerize, the concentration of free thiol in a stock solution of this compound was determined by the Ellman's test, and said solution used within the next two days after its preparation (always stored in the freezer while not in use). The results obtained from these experiments are presented in **Table 3.1**.

Entry	[Methyl cysteinyl] (mM)	[3.1] (mM)	Temp. (°C)	Solvent	3.13 to 3.12 relative ratio				
					Time (min)				
					15	60	120	240	360
1	2	2	37	H ₂ O	6:94	26:74	59:41	93:7	-
2	1	1	37	H ₂ O	4:96	25:75	52:48	87:13	100:0
3	0.5	0.5	37	H ₂ O	13:87	36:64	65:35	96:4	-
4	0.25	0.25	37	H ₂ O	19:81	49:51	82:18	98:2	-
5	1	1	60	H ₂ O	100:0	100:0	-	-	-
6	1	1	5	H ₂ O	18:82	21:79	37:63	69:31	83:17
7	1	1	37	H ₂ O/ACN 1:1	0:100	1:99	3:97	15:85	48:52
8	1	1	37	H ₂ O/MeOH 1:1	1:99	3:97	10:90	36:64	69:31
9	1	1	37	H ₂ O/DMSO 1:1	1:99	2:98	6:94	55:45	83:17
10	1	5	37	H ₂ O	17:83	92:8	100:0	-	-
11	1	3	37	H ₂ O	11:89	81:19	100:0	-	-
12	5	1	37	H ₂ O	5:95	20:80	54:46	95:5	-
13	3	1	37	H ₂ O	5:95	20:80	45:55	82:18	-

Structures of the reagents:

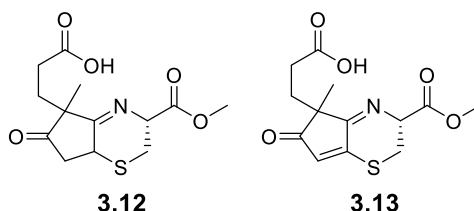


Table 3.1. 3.13 (M-20 Da adduct) to 3.12 (M-18 Da adduct) ratios at different times under different reaction conditions. Temp. = Temperature

The effect of reagents concentration, when mixed in a 1:1 molar ratio, was surprising. If we do not take into account the data obtained at a 2 mM concentration (entry 1), the oxidation process seemed to be slowed down when concentration was increased (entries 2-4), surprisingly in contrast to the normal tendency of chemical reactions to be accelerated at high concentrations. Yet, should oxidation be oxygen-promoted (which seems reasonable), more diluted reaction mixtures would contain a higher oxygen excess, which could account for the higher reaction rate.

The effect of temperature fitted with the normal tendency of chemical reactions (entries 2, 5 and 6), with high temperatures fastening the oxidation process. This effect was very

pronounced: while the oxidation of the **M**-18 Da adduct was completed after 15 min at 60 °C, 6 h were needed to reach a 83:17 **3.13/3.12** ratio when the reaction was performed at 5 °C. A similar conversion was obtained after 4 h when the temperature was 37 °C. The solvent influence was also clear (entries 2, 7-9). Addition of organic co-solvents slowed down the oxidation reaction, with acetonitrile having the most marked effect. DMSO And methanol performed better in terms of reaction rates, with the former being the one with the best behaviour.

Finally, regarding the reagent ratios, we found a quite curious effect. While an excess of methyl cysteinate did not substantially affect the oxidation step (entries 2, 12 and 13), the presence of a CPD excess greatly enhanced the oxidation kinetics (entries 2, 10 and 11). The reaction between equimolar amounts of reagents at 1 mM concentration presented an approximately 1:1 HPLC ratio between the intermediate and the final product after 2 h, whereas the same reaction with a 3- or 5-fold CPD excess was finished after the same time. In the case where 5 equivalents of CPD were present, the reaction was almost finished after 1 h.

3.6 Attempts to gain insight into the oxidation process

In spite of having information on how the **M**-18 to **M**-20 Da adduct transformation could be accelerated, we did not know which was the reaction mechanism that lead to the final **M**-20 Da adduct. A radical mechanism was likely operating, as we could not imagine how two hydrogen atoms could be removed from the **M**-18 Da adduct without generating species with unpaired electrons. Furthermore, formation of a radical species through H abstraction from the five-membered ring carbon adjacent to the sulfur atom should be favourable thanks to resonance stabilization, as it would be delocalised over four atoms. A second abstraction process would then form the conjugated double bond between the imine and ketone groups, that is, the **M**-20 Da adduct. However, we had no evidence to support or reject this proposal nor any idea why an excess of CPD could accelerate the oxidation step. Since no oxidiser had been added to the reaction medium, the **M**-18 to **M**-20 Da adduct transformation could be oxygen-mediated. Yet, attempts to perform the reaction in the absence of oxygen were unsuccessful. In other words, the **M**-20 Da adduct was also formed.

To try to gain insight into the oxidation step and detect the formation of radical species, we decided to monitor the reaction by EPR spectroscopy. Our intention was detect the formation of radical species and to identify them based on their spectra.

With this purpose, the reaction between methyl cysteinate and **3.6** was performed both at a 1 mM concentration with a 1:1 stoichiometry and at a 5 mM concentration with a 2:1

methyl cysteinate/**3.6** ratio. EPR Measurements were carried out for 20-30 minutes from 2000 to 4000 G. Unfortunately, no radical species could be detected in any case.

It is important to highlight that not detecting a radical using EPR measurements does not allow to reject the radical oxidation hypothesis, in the same way as not detecting a species by MS does not mean it is not present. Several factors, such as low concentration or extremely short half-lives of the radicals could be responsible for the absence of signal observed.

In view of these results and that other issues were more urgent, no further experiments dealing with the oxidation mechanism were performed.

3.7 Computational analyses

Intrigued by the unexpected results obtained with CPDs, we started a collaboration with Carme Rovira and Lluís Raich to try to understand the reasons behind the reversibility of the Michael-type reaction between internal or C-terminal cysteines and CPDs using quantum chemical calculations.

In the first place, standard Gibbs free energies of the Michael-type reactions of methyl cysteinate, which was used as model thiol, with *N*-methylmaleimide (NMMal) and **3.6** were calculated by means of DFT methods. As can be seen in **Figure 3.8a**, the reaction turned out to be highly exergonic in the case of NMMal, with a $\Delta G^{\circ} = -10.6$ kcal/mol. In contrast, the same reaction with **3.6** was more than 4 times less favoured, with a $\Delta G^{\circ} = -2.4$ kcal/mol. This is translated to an equilibrium constant for the maleimide reaction 10^6 times larger than that for the CPD reaction, confirming the experimental observation that the Michael-type reaction for maleimides is basically irreversible, whilst in that with CPDs an equilibrium is established.

Computational analysis of the structures of both Michael acceptors revealed important differences that can account for the different reactivity. Analysis of the structure of NMMal and **3.6** revealed that the C=C bond is slightly longer for **3.6** than for NMMal (1.34 Å vs. 1.33 Å, respectively) and the CO-CH bond is shorter in **3.6** than in NMMal (1.49 Å vs. 1.50 Å). Actually, the value for the NMMal C=C bond was found to be the same as for the pure localised π -system of cyclopentene, while its CO-CH bond was only 0.01 Å shorter in comparison with the same molecule. This was not the case of **3.6**, whose bond distances were half way between cyclopentene and cyclopentadiene, indicating a major delocalization (**Figure 3.8b**). Probably, then, the Michael-type reaction involving CPDs is less favoured than that involving maleimides to preserve the major delocalization found in the case of the former class of molecules.

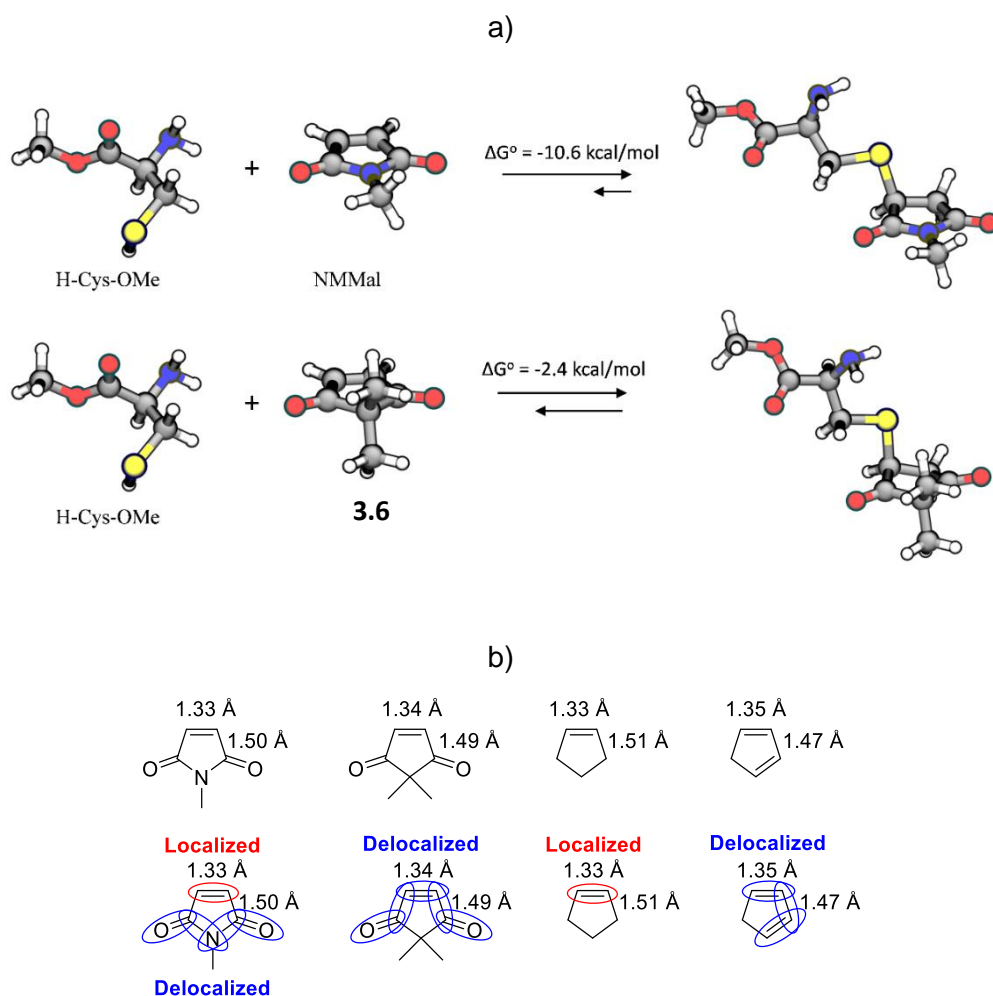


Figure 3.8. a) Calculated energies for the Michael-type reaction between methyl cysteinate and NMMal or **3.6**. b) Bond distances of NMMal and **3.6** compared to localised and delocalised systems. H-Cys-OMe stands for methyl cysteinate.

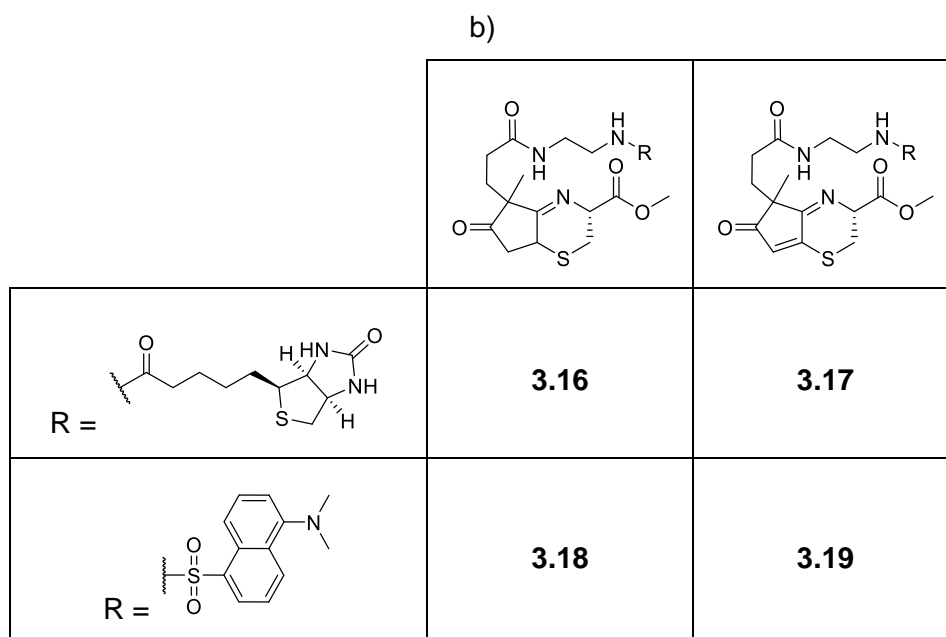
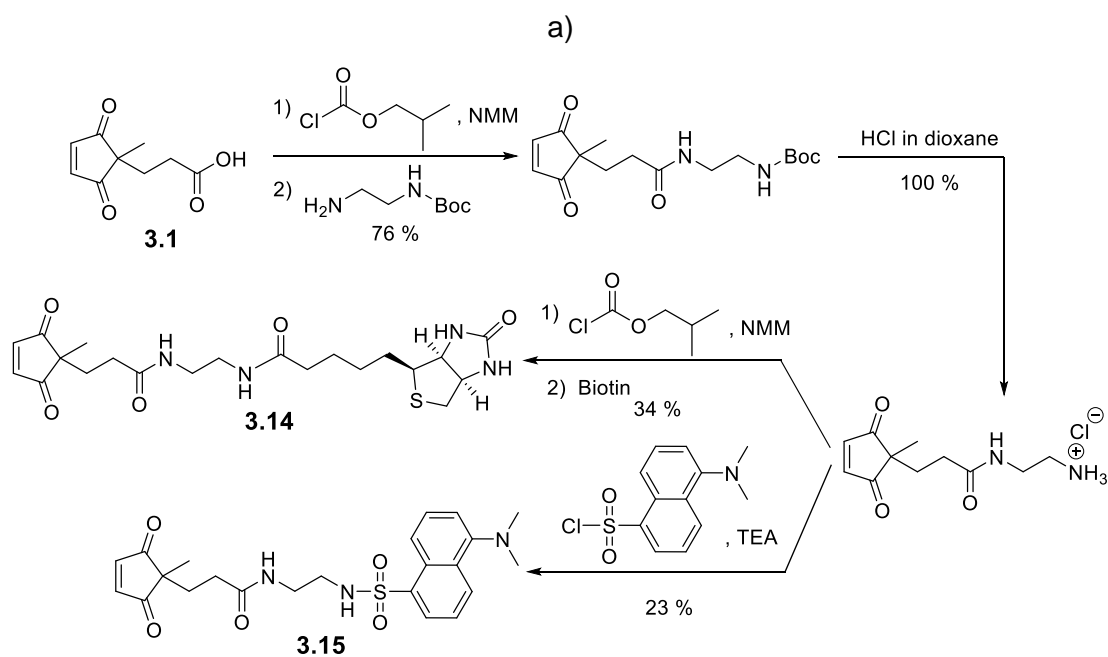
3.8 Reactions with other CPDs

Jordi Agramunt, at that time a master student in our group, was in charge of the synthesis of new CPD derivatives that could be useful for bioconjugation. For this purpose, the reporter molecules biotin and dansyl were linked to the carboxylic derivative of **3.1** through an ethylenediamine linker. Their synthesis was accomplished using the synthetic route depicted in **Scheme 3.5a**. In short, **3.1** was activated through formation of the mixed anhydride with isobutylchloroformate. This was *in situ* reacted with *tert*-butyl (2-aminoethyl)carbamate to render the desired amide in 76 % yield.

Other methods for acid activation, such as activation with carbodiimides and formation of the *N*-hydroxysuccinimidyl ester were not successful. The first alternative failed to obtain the desired product, while the formation of the NHS ester was low yielding and

subsequent nucleophilic attack by the amine required long reaction times and rendered only small amounts of the desired product.

After quantitative acidic amine deprotection (using hydrochloric acid in dioxane), biotin was introduced by the method of the mixed anhydride described above and dansyl using its chloride derivative. Both the biotin-derivatised (**3.14**) and the dansyl-derivatised (**3.15**) CPDs were then studied in the reaction with methyl cysteinate.



Scheme 3.5. Synthesis of compounds **3.14** and **3.15** (a) and structures of the products after reaction a) with methyl cysteinate.

3.14 and **3.15** were independently incubated with five equivalents of methyl cysteinate at 37 °C and 60 °C, respectively. Although the oxidation step required longer reaction times for **3.14**, as expected for a lower reaction temperature, both reactions evolved as previously observed for **3.1** and **3.6**. Reaction of the biotin derivative was analysed after 15 min, 1 h and 24 h, allowing the conversion of the **M-18 Da** intermediate (**3.16**, see **Scheme 3.5b**) to the final **M-20 Da** adduct (**3.17**) to be observed. Curiously, in this case both isomers co-eluted in one peak (**Figure 3.9a**). On the other hand, complete oxidation of the **M-18 Da** adduct derived from CDP-dansyl (**3.18**, see **Scheme 3.5b**) after 2.5 h rendered two peaks fitting with the expected mass for the final oxydised product (**3.19**) (**Figure 3.9b**).

In conclusion, we proved that regardless of the substitution at the 2 position of the CPD ring, the final **M-20 Da** adduct was always obtained and that the reaction path leading to its formation was the same.

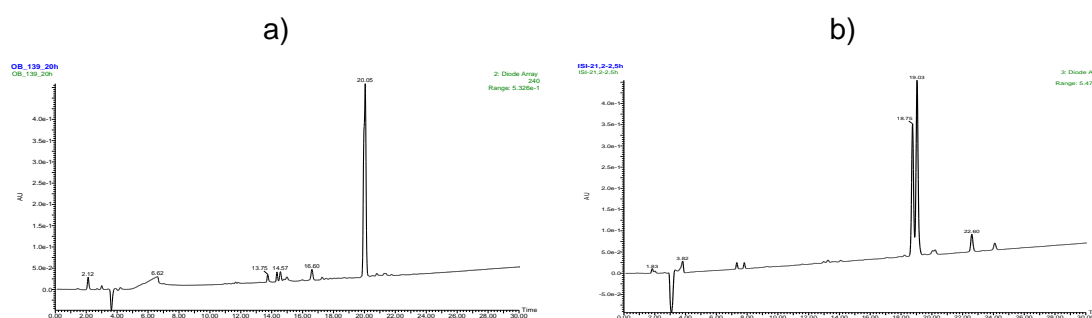


Figure 3.9. HPLC/MS traces (240 nm) of the reaction of methyl cysteinate with **3.14** after 24 h (a) and with **3.15** after 2.5 h (b).

3.9 Reactions between CPDs and other amino acids: cysteine and homocysteine

We then assayed the reactions of CPDs with cysteine and homocysteine (Hcy). These two thiols are biologically relevant and are known to be related to several diseases.

Cysteine low levels are related to slow growth in children, hair depigmentation, liver damage, weakness and skin lesions.¹⁶ Moreover, cysteine is one of the main reductants present in living cells and a precursor of glutathione, which is known to act as a defensive reagent against the action of toxic xenobiotics, to behave as antioxidant, and to be involved in protein regulatory functions through formation of disulfide bridges with proteins.¹⁷

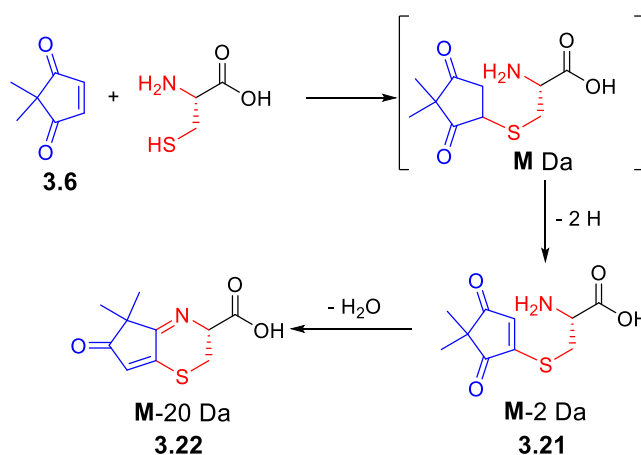
Hyperhomocysteinemia plays an important role in the development of cardiovascular diseases, as it has been positively correlated to myocardial infarction.¹⁸ Further,

abnormal levels of Hcy are also related to Diabetes Mellitus types 1 and 2,^{19,20} and it is known that bone health is also related to this non-proteinogenic amino acid levels.²¹ Hcy levels are also correlated with cognitive decline in patients with Alzheimer²² and with late-life dementia.²³

For these reasons the study of Hcy and Cys levels has gained interest over the past decade, and many groups have developed different methodologies for the determination of these amino acids.^{24,25,26} Therefore, we believed it would be of interest to study the reaction of CPDs with these thiols.

3.9.1 Reaction of CPDs with Cys

CPD **3.6** Was used as a model for the reaction with cysteine. A single HPLC/MS analysis of the reaction crude indicated a situation different from that encountered in the case of methyl cysteinate, as the expected **M-18 Da** adduct (**3.20**), which could not be detected, seemed to have been replaced by an adduct with an **M-2 Da** mass (**3.21**). At first we believed that the reaction with cysteine evolved through a different path towards the final **M-20 Da** adduct (**3.22**), that is, through the oxidation of the MTA to the **M-2 Da** adduct, which would then cyclize to yield the final **M-20 Da** product, as depicted in **Scheme 3.6**.



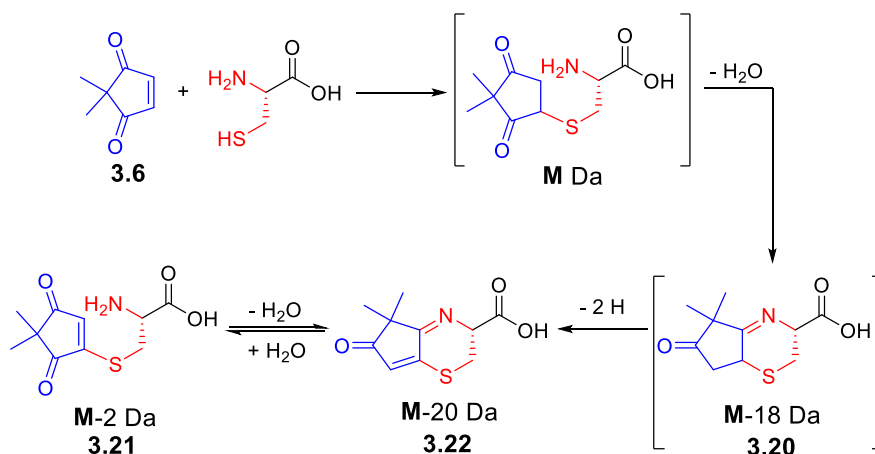
Scheme 3.6. Plausible reaction path to explain the formation of **3.21**.

However, in the experiment described above we had not monitored the crude evolution, and therefore we could not support our hypothesis. For this reason, the reaction was repeated but this time analysed at different incubation times. Analysis after running the reaction for 5 minutes showed the presence of **3.22** and **3.21** in an 80:20 ratio, as determined by HPLC/MS at 300 nm (where areas were comparable). Contrary to the expected trend for an intermediate species, the amount of the **M-2 Da** adduct was found

to increase with time and quantitative transformation to **3.22** never took place. As can be seen in **Table 3.2**, the amount of **3.21** grew until an equilibrium point was reached in which the ratio had changed to 62:38 **3.22/3.21**. The fact that the relative amount of the **M-2 Da** adduct increased with time pointed towards this not being an intermediate but actually arising from the **M-20 Da** product through imine hydrolysis. Furthermore, because the relative amounts of **3.21** and **3.22** stabilised after 2 h, we inferred that the hydrolysis reaction had to be reversible (**Scheme 3.7**).

Time (min)	M-2 Da adduct 3.21 (%)	M-20 Da adduct 3.22 (%)
5	20	80
15	24	76
30	29	71
60	36	64
120	38	62
240	38	62

Table 3.2. Relative ratios of **3.21** and **3.22** at different times, as determined by HPLC (300 nm).



Scheme 3.7. Mechanistic alternative for the formation of the **M-2 Da** adduct.

To confirm that **3.21** and **3.22** can be transformed into each other, their corresponding peaks were independently collected and reanalysed after a 3 h incubation in water at 37 °C. In this experiment it was observed that each pure compound had evolved to furnish a mixture of **3.21** and **3.22**, demonstrating the reversibility of the water addition to the **M-20 Da** adduct.

Quantum chemical calculations, performed by Carme Rovira and Lluís Raich (**Figure 3.10**) showed that intramolecular attack of the cysteine amine to the CPD carbonyl to generate the corresponding hemiaminal was more favoured than in the case of methyl cysteinate. Yet, in the cysteine reaction the **M-18 Da** adduct was not observed, which was intriguing because, as previously stated, oxidation to the **M-20 Da** adduct is typically slow.

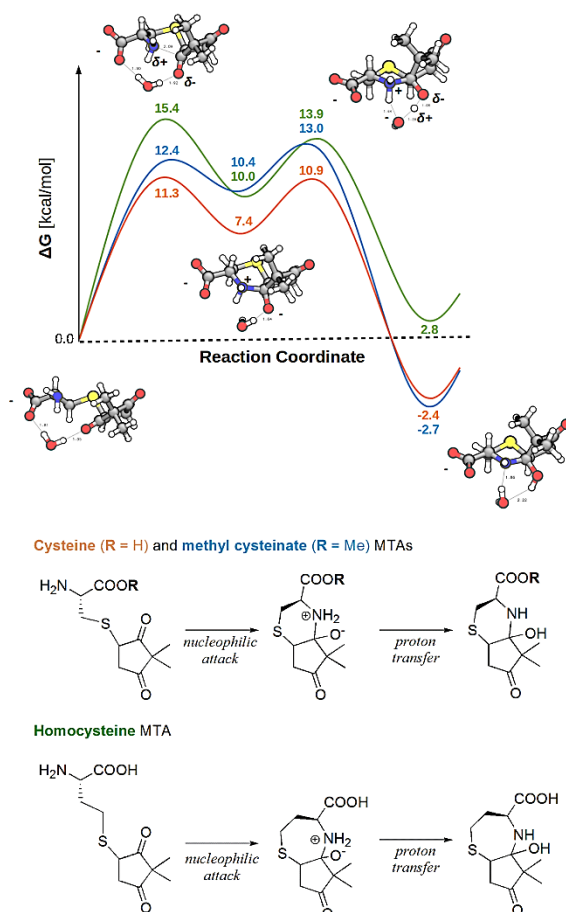


Figure 3.10. Energy profiles for the generation of the hemiaminal that leads to imine formation for methyl cysteinate (blue), cysteine (orange) and homocysteine (green).

Detection of the expected **M-18 Da** adduct **3.20** would be important to support the mechanistic proposal depicted in **Scheme 3.7**, although the results obtained by quantum mechanical calculations and the **3.21/3.22** relative ratios evolution clearly agreed with it. To confirm formation of **3.20** we believed that an analytical method allowing for the study of the system with a minimal perturbation was needed. NMR would be adequate in this

sense, but we were not sure the NMR time scale would be suitable to detect what seemed to be a very short-lived precursor. Further, NMR experiments require amounts of products much above our stock of **3.6**, and we envisaged rather difficult-to-interpret spectra. For these reasons we decided to work with ESI MS analyses, a technique that requires very small amounts of products and produces a minimal perturbation in the system. At the same time, no problems regarding the time scale of the technique were expected, and it would let us distinguish between the three expected products easily (**3.20**, **3.21** and **3.22**).

Luckily, direct perfusion of the reaction crude on the ESI analyser (see spectra in **Figure 3.11**) allowed us to detect the mass calculated for **3.20**. Its role as an **M-18** Da adduct was confirmed by the fact that its signal intensity decreased with time until total disappearance. All in all, these results confirmed our hypothesis that **3.20** led to the formation of **3.22**, which, through a reversible hydrolysis reaction, generated the newly detected **M-2** Da adduct **3.21**.

Gratifyingly, the presence of the MTA (**3.23**) could also be confirmed in this experiment, being the first time that this elusive species was observed in our experiments. As in the case of **3.20**, its signal intensity decreased with time until disappearance.

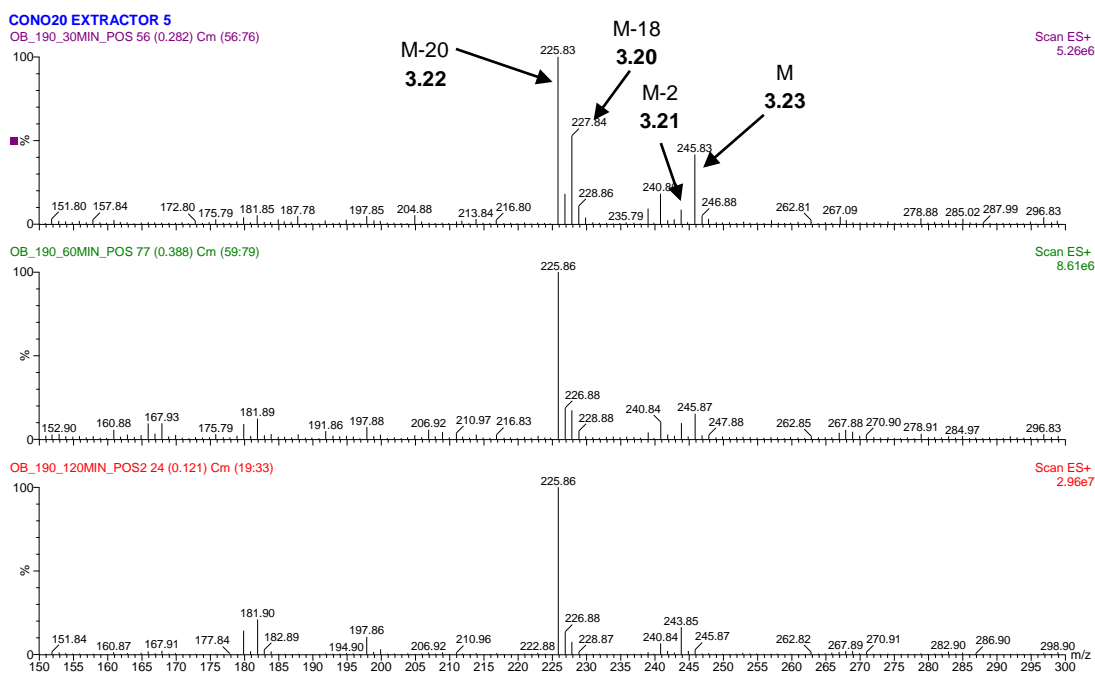


Figure 3.11. ESI MS analyses of the **3.6** + cysteine reaction crude (direct perfusion, 20 V, positive mode) after 30, 60 and 120 min, from top to bottom.

It was clear that, since the rest of the molecule was the same as for methyl cysteinate, the free carboxylic moiety of cysteine had to be responsible for the differential behaviour

between the **M**-20 Da adducts **3.22** and **3.11**. As the carboxylic group of **3.22** was expected to be deprotonated in the pH of work, the first step of the hydrolysis reaction, which implies the protonation of the imine nitrogen, would generate a zwitterion (see **Figure 3.12**) and therefore ease this reaction. On the contrary, the same process was not expected to be so favoured in the case of **3.11** as the first reaction intermediate would be a positively charged molecule instead of a neutral one (**Figure 3.12**).

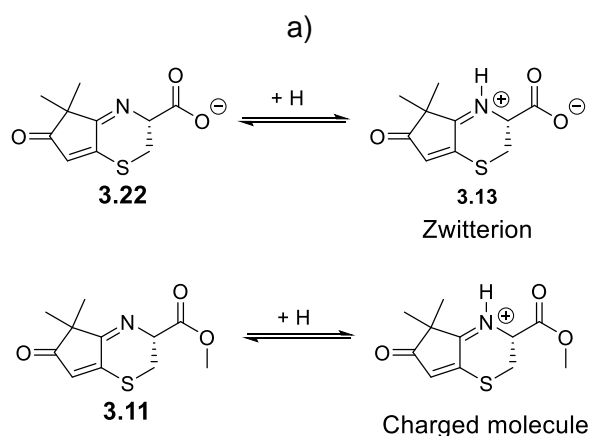


Figure 3.12. Intermediates formed upon protonation of the imine nitrogen of the **M**-20 Da adducts of cysteine and methyl cysteinate.

3.9.2 Reaction of CPDs with Hcy

In the case of Hcy (**3.24**), we anticipated that formation of the corresponding **M**-18 Da adduct would be more difficult, as cyclisation would generate a seven-membered ring, which is energetically less favourable than formation of a six-membered one. Hence, we anticipated that in case this adduct was formed, the reaction would be slower than in the case of reactions involving cysteine and methyl cysteinate. Possible hydrolysis of the corresponding **M**-20 Da adduct was not considered because at the time experiments with Hcy were performed this reaction had not yet been observed.

To find out whether Hcy would behave as a 1,2-aminothiol or as a simple thiol, its reaction with CPD **3.14** (at 37 °C) was followed by HPLC/MS. We could observe formation of the MTA between **3.14** and Hcy (**3.25**), but not the generation of the **M**-20 and **M**-18 Da adducts, at least at short reaction times (less than 20 h). Curiously, small amounts of the corresponding **M**-2 Da adduct were observed, but the fact that the **M**-20 Da adduct was at no time detected made us conclude that there might be another mechanism operating for its formation. These results indicated that Hcy was behaving as a simple thiol due to the difficulty to furnish the seven-membered ring through imine formation (see **Figure 3.13**).

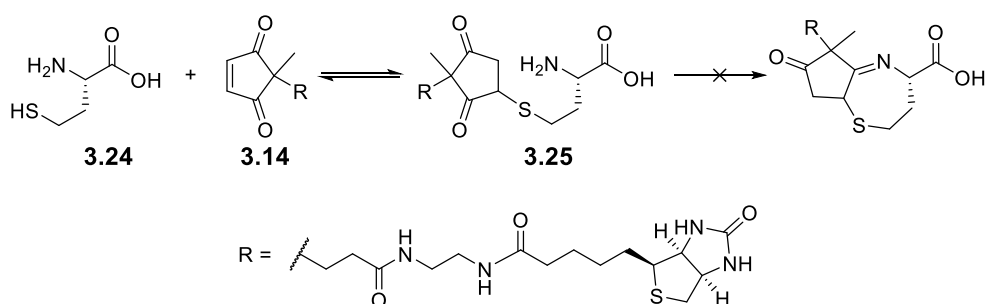


Figure 3.13. Reaction of **3.14** with Hcy furnishes the expected MTA, although this does not further evolve to furnish the **M-18 Da** adduct.

Indeed, DFT calculations showed that hemiaminal generation, which leads to imine formation, was energetically disfavoured in comparison with cysteine and methyl cysteinate, in agreement with the experimental results. As can be seen in **Figure 3.10**, the most energetically demanding step was found to be the attack of the Hcy amine to the CPD ketone, while proton transfer to render the hemiaminal was not significantly different in terms of energy from the same step for cysteine and methyl cysteinate. A difference in, at least, 3.0 kcal/mol for the intramolecular attack of the amine explains the absence of the **M-18 Da** adduct formation, and, therefore, that the **M-20 Da** adduct could not be detected.

Reversibility of the reaction was, again, confirmed by collecting the peaks corresponding to the MTA **3.25** and incubating them in water after lyophilisation. Reversion to the parent **3.14** was clearly observed, as in the cases in which MTAs had not evolved to furnish the bicyclic **M-18 Da** adduct structure.

3.10 Attachment of CPDs to peptide chains

To assess whether CPDs could be incorporated at the *N*-terminal position of peptide chains without any damage during the cleavage step, **3.1** was coupled to the 9-mer Gly-Arg-Gly-Ser-Tyr-Glu-Ala-Tyr-Lys peptide-resin by carbodiimide activation. Totally standard coupling conditions were used, but no additives such as HOBt were used to prevent any possible conjugate addition to the CPD double bond. After coupling **3.1**, the derivatised peptide was cleaved and deprotected using the standard 95:2.5:2.5 TFA/TIS/H₂O mixture (see **Figure 3.14a**). HPLC analysis of the crude and MALDI-TOF MS analysis of the collected major peak revealed that the expected CPD-derivatised peptide (**3.26**) had been successfully obtained. As can be seen in **Figure 3.14b**, no major

by-products were observed by HPLC, indicating that no important side-reactions had occurred during both the coupling and cleavage steps.

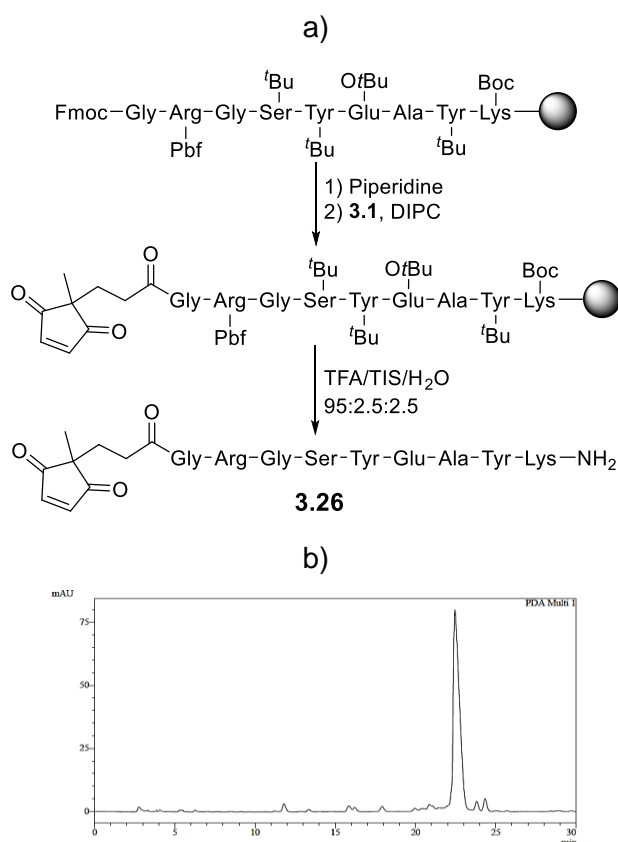


Figure 3.14. Synthesis (a) and HPLC trace (b, 280 nm) of the crude **3.26**.

3.11 Synthesis of bioconjugates exploiting the formation of an M-20 Da adduct and assessment of the stability of the generated products

The next logical step after proving that CPDs were suitable to be incorporated into peptide chains was to demonstrate their usefulness for bioconjugation through formation of the corresponding **M-20 Da** adduct with *N*-terminal cysteines. For this proof-of-principle experiment we decided to synthesize a peptide-PNA conjugate.

CPD-Peptide **3.26** was selected as the peptide moiety in this experiment. Then, a cysteine residue was incorporated to the resin-linked PNA Lys-catagctgtttc (using Fmoc-Cys(Trt)-OH), yielding the *N*-terminal cysteine PNA **3.27** after cleavage, deprotection and isolation (see structures in **Figure 3.15**).

3.27 Was incubated with 1.2 equivalents of **3.26** at 37 °C using water as solvent (0.85 mM PNA concentration). Although the reaction seemed to be completed after 3 h, as no

M-18 Da adduct could be observed by HPLC, the mixture was left stirring at 37 °C for a total of 5 h (**Figure 3.15a**). This decision was taken because one of the **M-18 Da** isomers practically co-eluted with one of the isomers of the **M-20 Da** adduct, and we preferred to ensure that the oxidation step was complete. HPLC Analysis of the reaction crude after 5 h demonstrated that practically no by-products were formed during the conjugation reaction. The only peaks observed in the chromatogram aside from those corresponding to the **M-20 Da** adduct were impurities already present in the starting PNA reagent and, therefore, not created during the reaction. The final conjugate **3.28** could be isolated, as a mixture of isomers, in 51 % yield after HPLC purification (**Figure 3.15b**).

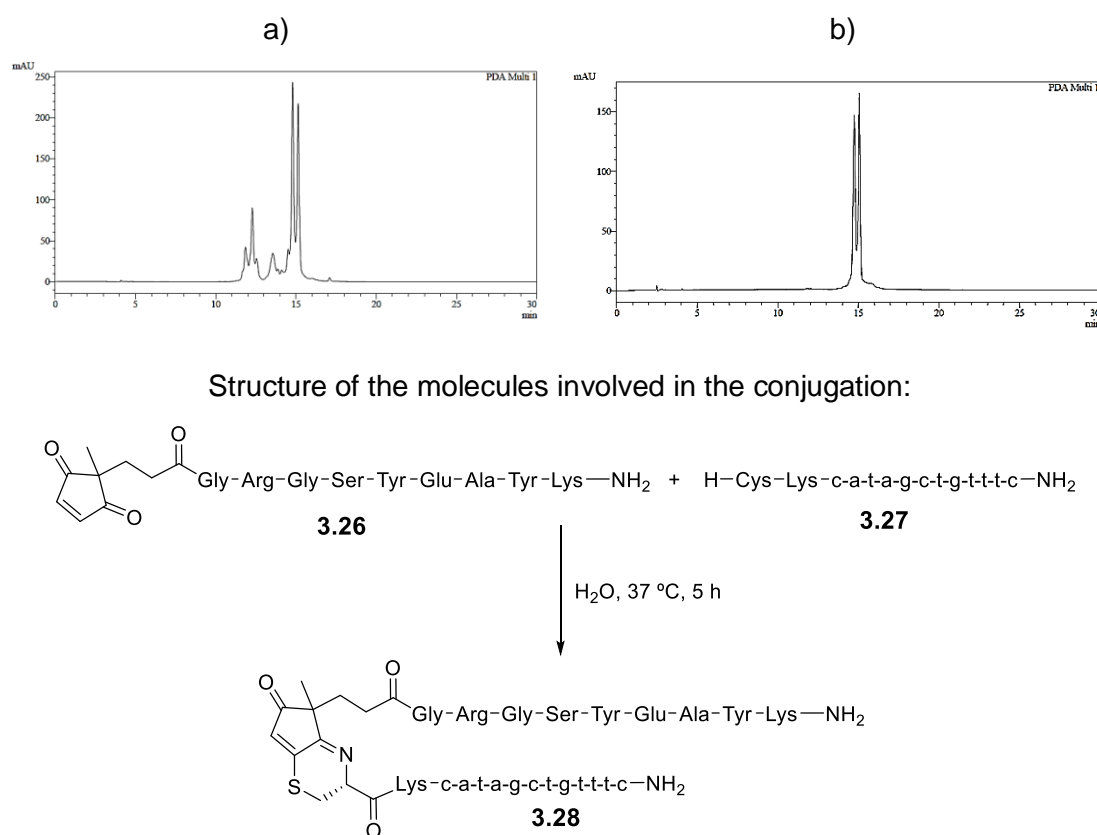


Figure 3.15. HPLC traces (260 nm) of the crude and pure conjugate **3.28** and structures of the molecules involved.

After purification, **3.28** was used to study the stability of the linkage between the two conjugate units. For this purpose, different aliquots of the conjugate were incubated at different pHs and analysed over time. The results of these experiments are summarised in **Table 3.3**.

	Remaining 3.28 percentage		
	pH = 6.0	pH = 7.0	pH = 8.0
2 h	100	100	100
6 h	99	98	87
24 h	97	86	65

Table 3.3. Remaining percentage of **3.28** (HPLC determined) after incubation at different pHs.

3.28 Was found to be almost completely stable over 24 h in a slightly acidic medium (pH=6.0) and over 6 h at neutral pH. Incubation at pH 7.0 for 24 h reduced the amount of unaltered conjugate to 86 %, indicating that the stability of the conjugated bicycle was fairly good in these conditions. On the contrary, basic pH induced a faster alteration of the conjugate, leaving 87 % of unaltered **3.28** after 6 h at pH 8.0 and a 65 % after 24 h. The effect of external thiols (100-fold excess) was also tested (**Table 3.4**). This time, all experiments were performed at pH 7.0, and a comparison between the results in the presence or absence of an external thiol was used to evaluate the effect of said entities on the conjugate stability. The effect of glutathione could only be noticeable (and in a low extent) after 6 h, but the results obtained after 24 h indicated that this thiol has little or no effect on the conjugate stability. On the other hand, methyl cysteinate proved to induce a clearly faster alteration of **3.28**. As in the case of glutathione, most of the conjugate modification happened in the first 6 h, while the remaining time until 24 h seemed to cause a rather little effect.

	Remaining 3.28 percentage	
	+ Glutathione	+ H-Cys-OMe
2 h	100	79
6 h	93	62
24 h	86	53

Table 3.4. Effect of external thiols on the stability of **3.28** at pH 7.0. H-Cys-OMe stands for methyl cysteinate.

All in all, we found that conjugates formed upon reaction of an *N*-terminal cysteine and a CPD are stable at slightly acidic and neutral pH for moderately long time periods. It was also found that the *in vivo* abundant glutathione had virtually no effect on conjugate

stability, while other 1,2-aminothiols, such as cysteine methyl ester, significantly altered the conjugate structure.

3.12 Exploiting the selectivity of CPDs towards *N*-terminal cysteines

There are not many reports in the literature in which cysteines are selectively tagged depending on their position in peptide chains. Rao's group reported that 2-cyanobenzothiazole (CBT) selectively reacts with 1,2- and 1,3-aminothiols in the presence of other thiols, and took advantage of this reactivity to label proteins,²⁷ for the controlled assembly of nanostructures in living cells²⁸ and for the imaging of protease activity.²⁹ They found out, however, that the reactivity of aromatic cyano compounds was difficult to predict, as structures similar to CBT failed to reproduce its reactivity or provided the same results with very different rate constants.²⁹ Other groups have exploited CBTs for the immobilization of proteins and other biomolecules.^{30,31}

Thiazolidine formation and native chemical ligation also allow to selectively tag *N*-terminal cysteines in the presence of other thiols (see chapter 1).

The experiments described in this chapter have shown that CPDs react irreversibly with 1,2-aminothiols, yielding stable adducts, and reversibly with thiols lacking a free amine at the β position. Hence, we wondered whether it would be possible to exploit this differential behaviour to selectively derivatize cysteines at the *N*-terminal position of peptide chains in the presence of *N*-acylated cysteines. We believed that it would be interesting, from a biochemical point of view, to selectively label peptides containing an *N*-terminal cysteine in the presence of peptides that contain internal cysteines. Furthermore, the regioselective double derivatization of a peptide incorporating two differently placed cysteines, one of which is at the *N*-terminal position, also seemed relevant.

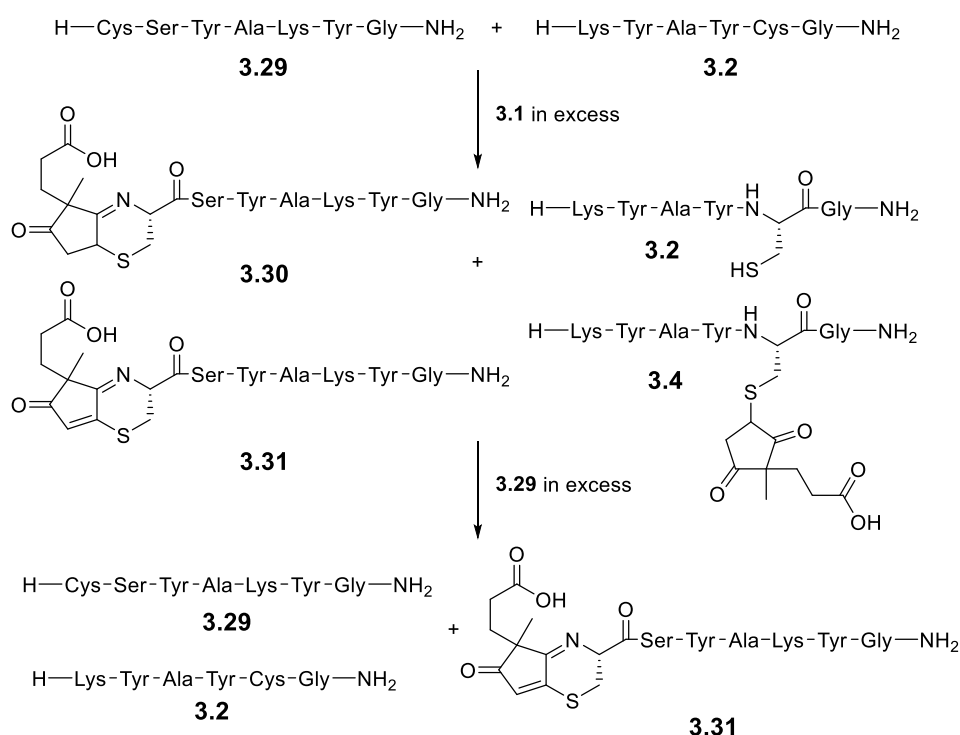
3.12.1 Selective derivatization of a peptide containing an *N*-terminal cysteine in a mixture of cysteine-containing peptides

To assess whether a CPD could selectively react with a peptide containing an *N*-terminal cysteine in the presence of another cysteine containing peptide, we decided to use peptides H-Cys-Ser-Tyr-Ala-Lys-Tyr-Gly-NH₂ (**3.29**) and **3.2** and CPD **3.1**. Three different variants of the experiment were performed. In each of them, both peptides were mixed in equimolar amounts and their ratios verified by HPLC before the experiment.

The difference between the three experiments was that **3.1** was added in defect, equimolar amounts or excess to the peptide mixture.

In the first place, when **3.1** was added to the peptide mixture in defect (0.7 equiv.), the only peptide derivatised was **3.29**, while **3.2** remained unaltered. Of course, some unreacted **3.29** was found because the amount of CPD added was not enough to fully consume it (**Figure 3.16a**). Further addition of 0.3 equivalents of **3.1** totally consumed the remaining **3.29** without reacting with **3.2**. Total transformation of the **M-18 Da** intermediate (**3.30**) to the corresponding **M-20 Da** adduct (**3.31**) was observed after a 2.5 h incubation time (**Figure 3.16b**).

Secondly, when 3.5 equiv. of **3.1** were added to the equimolar peptide mixture, complete derivatisation of peptide **3.29** and partial labelling of **3.2** could be detected, according to a quantitative reaction in the first case and an equilibrium reaction in the second (**Scheme 3.8** and **Figure 3.16c**). Reversibility of the MTA adducts generated by the reaction of **3.1** and **3.2** was confirmed by adding an excess of **3.29**. After this was done, it was observed that the **3.1 + 3.2** MTA disappeared from the reaction mixture, which was accompanied by an increase in the amount of **3.31** as a result of the reaction between the liberated CPD and the excess of peptide **3.29** added (**Figure 3.16d**).



Scheme 3.8. Representation of the experiment performed with an initial excess of **3.1** aimed at the selective derivatisation of a peptide containing an *N*-terminal cysteine in a mixture of cysteine-containing peptides.

Finally and as expected, when **3.1** was added in equimolar amounts, complete modification of **3.29** to furnish **3.31** was observed after 3 h, although some **3.30** was still detected (**Figure 3.16e**). **3.2** Essentially did not react and, although minor amounts of labelled **3.2** were detected on the HPLC traces of the reaction, these must be attributed to small errors in the quantification of all the products and thus product ratios not being exactly 1:1:1.

In all experiments the two different isomers of **3.31** could be separated and isolated in ~45% yield.

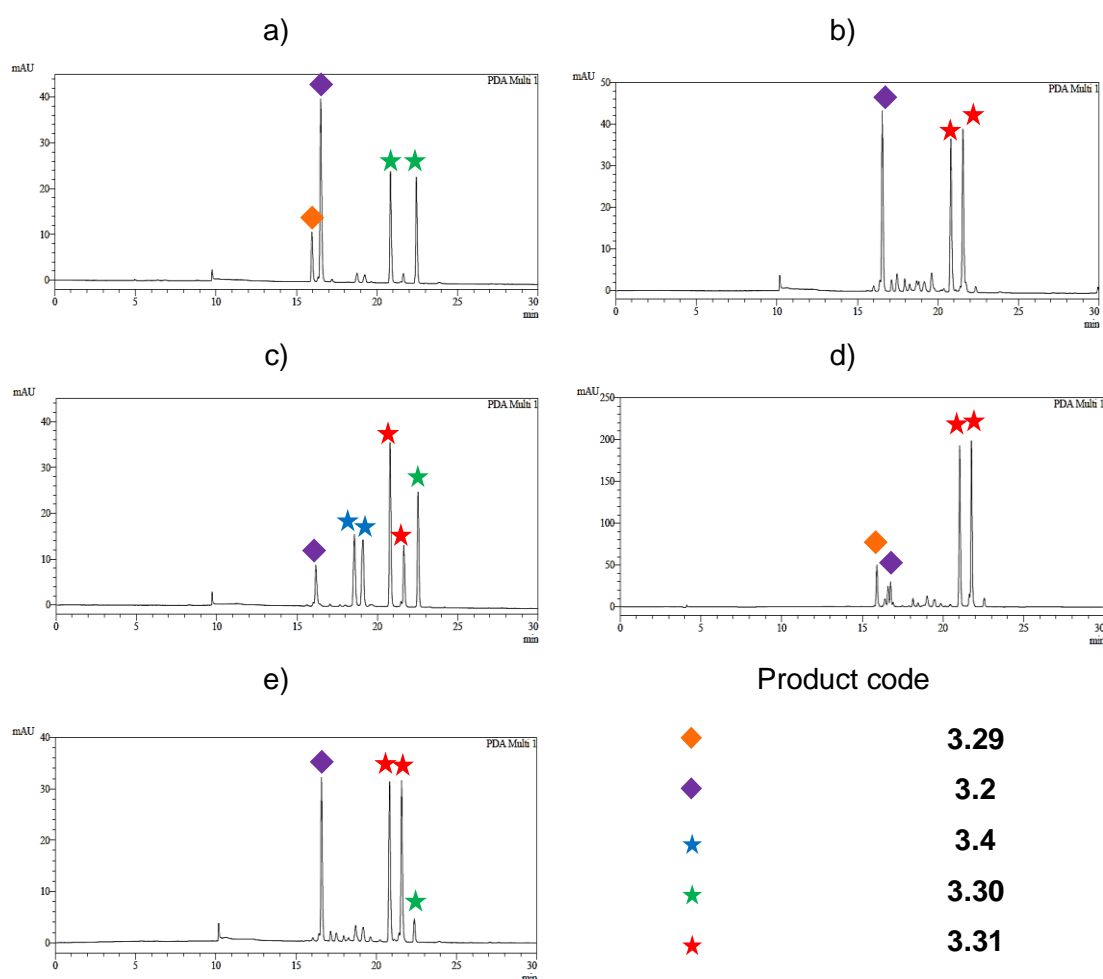


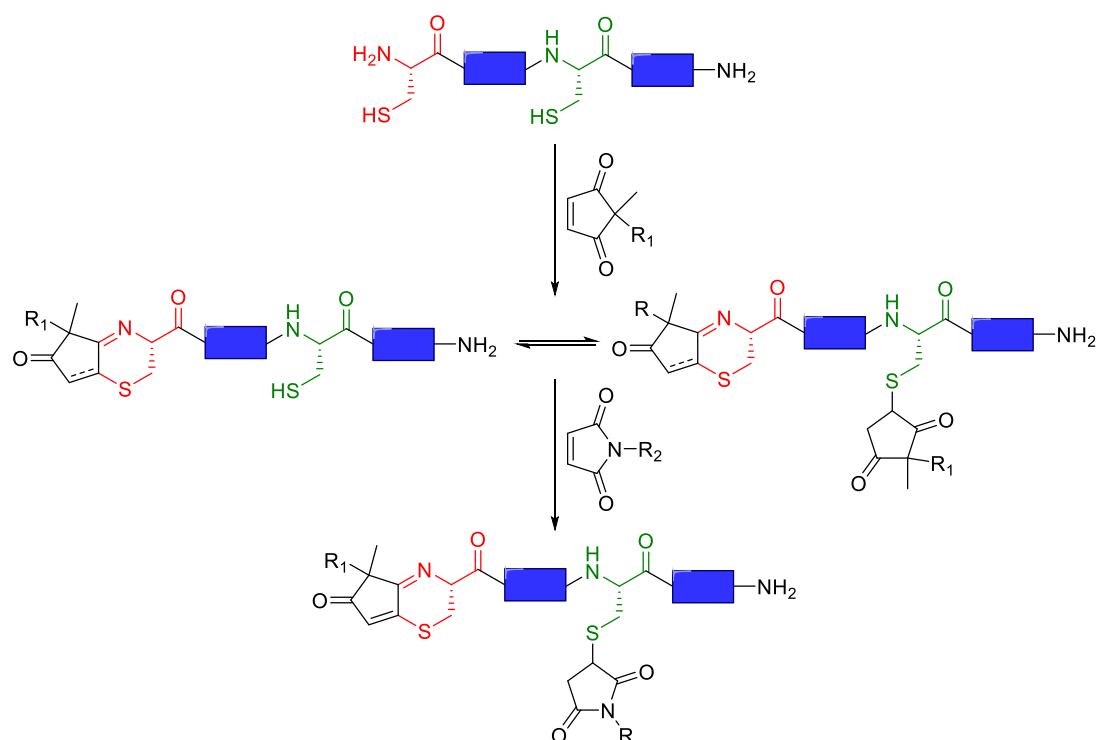
Figure 3.16. HPLC traces of the experiments aiming at the selective derivatization of a peptide containing an *N*-terminal cysteine in a mixture of peptides. a) reaction with a defect of **3.1** after 30 min; b) reaction with a defect of **3.1** after 180 min. c) reaction with an excess of **3.1** after 30 min; d) reaction with an excess of **3.1** 1 h after the addition of excess **3.29**. e) reaction with equimolar amounts of **3.1** after 180 min.

In summary, these experiments showed that the selective labelling of a peptide containing an *N*-terminal cysteine in the presence of other cysteine-containing peptides is possible and easy. As long as the CPD is not added in excess with respect to the peptide containing the *N*-terminal cysteine, the other cysteine-containing peptides in the mixture will remain unaltered. This offers a platform to discriminate between different peptides or proteins in complex mixtures, and can find application to separate compounds incorporating with a free amine from other thiols.

3.12.2 Regioselective double labelling of a peptide containing *N*-terminal and internal cysteines

Another possibility to take advantage of the selectivity of CPDs towards 1,2-aminothiols was to use a peptide containing an *N*-terminal cysteine and a cysteine placed in an internal position for incubation with a CPD. If the peptide/CPD ratio was 1:1, only the *N*-terminal cysteine should react. However, addition of slightly larger amounts of CPD should not be a problem because the MTA generated upon reaction with the internal cysteine would be reversible. Therefore, subsequent addition of another thiol labelling molecule, such as a maleimide, would promote derivatisation of the only available thiol, that is, the internally placed cysteine, independently of the amount of CPD initially added. This would generate a regioselectively doubly derivatised peptide (**Scheme 3.9**).

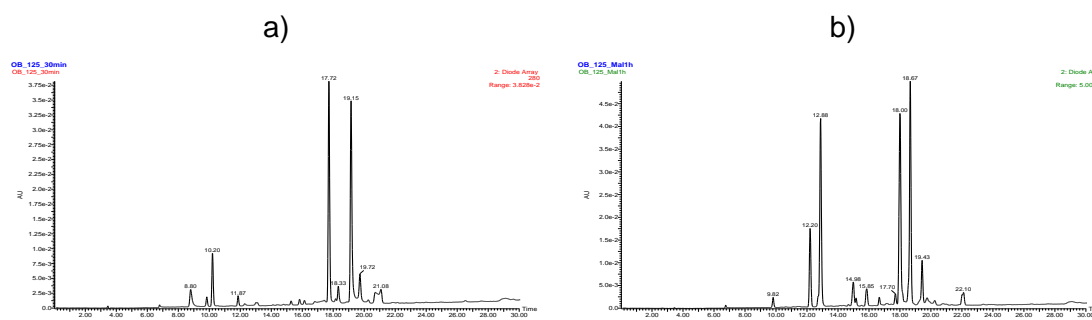
To confirm that this was possible peptide H-Cys-Ser-Tyr-Ala-Cys-Lys-Tyr-Gly-NH₂ (**3.32**) was synthesised. In a first experiment **3.32** was reacted sequentially with 5 equiv. of **3.1** and 3-maleimidopropanoic acid but, to our surprise, another product was formed together with the expected doubly derivatised peptide. This experiment, which did not produce satisfactory amounts of the desired product, is described in section **4.1**.



Scheme 3.9. Regioselective double derivatization of a peptide containing an *N*-terminal and an internal cysteine.

In a second experiment, **3.32** was first reacted with 1.1 equiv. of **3.1**. HPLC/MS Analysis of the crude (30 min after the addition of **3.1**, see **Figure 3.17a**) revealed the presence of the cyclic disulfide structure as a minor side-product and the major formation of the corresponding **M-18 Da** adduct (**3.33**), while formation of products containing two CPDs was only marginal. Addition of 5 equiv. of 3-maleimidopropanoic acid (at a 90 min total reaction time) furnished, after 1 h, a product with a UV-Vis profile containing a maximum around 330 nm and a mass that fitted with formation of the **M-20 Da** adduct and the Michael-type addition of 3-maleimidopropanoic acid on the internal cysteine (see **Figure 3.17b**). This was, of course, the desired doubly derivatised peptide (**3.34**), which was isolated in a 44 % yield.

Together with the target compound, a product corresponding to the double Michael-type addition of 3-maleimidopropanoic acid to the peptide was also detected. As the disulfide cyclic product had disappeared, possibly this product was formed upon reduction of the disulfide bridge and reaction of the generated thiols with the excess maleimide.



Structures of the reagents and products:

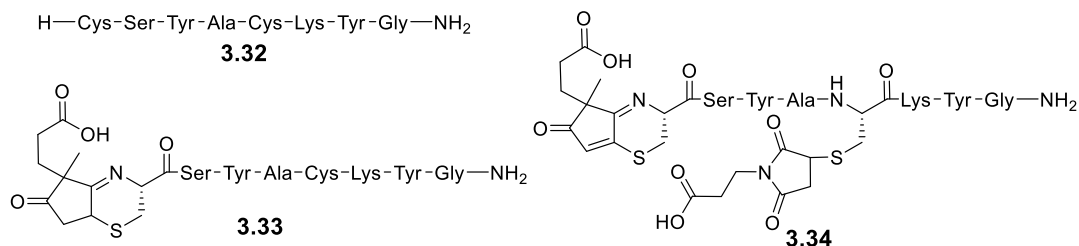


Figure 3.17. HPLC/MS traces (280 nm) of the reaction crude 30 minutes after the addition of **3.1** (a) and 1 h after the addition of 3-maleimidopropanoic acid (b). The peak at 10.2 min in (a) is the cyclic disulfide and those at 17.7 and 19.1 min are **3.33**. Peaks at 18.0 and 18.7 min in (b) are **3.34**, whilst the by-product incorporating two maleimide moieties corresponds to the peak at 12.2 min. Excess of 3-maleimidopropanoic acid appears at 12.9 min (b).

3.13 Estimation of the M-20 Da adduct molar extinction coefficient

Finally, the last group of experiments related with the studies and applications of CPDs reactions was to estimate the molar extinction coefficient or ϵ of the **M-20 Da** adducts. To determine the ϵ value of a given compound one needs to know the concentration of a solution of the sample being analysed and its absorbance. To compensate for any possible error in the sample preparation, it is advisable to prepare a series of solutions of known concentration and measure their absorbance. Applying the Lambert-Beer law the molar extinction coefficient can be determined from the gradient of the line of best fit obtained when plotting the absorbance values in front of the concentration of the solutions. All absorbance values must be between 0.1 and 1.0, as lower or higher values tend to turn aside from said law.

The problem, though, is to determine the concentration of a given sample. This is normally performed by weighing the product and dissolving it to reach the desired concentration range. However, the scale at which we were working precluded this option,

since weighing such small amounts of compounds would result in large errors. Thus, we decided to make an estimation of the ϵ value.

A possible way to approximate to the ϵ value for the **M-20** Da adduct was to measure the absorbance of a peptide sample before and after the reaction with a CPD. Of course, this peptide would need to possess an *N*-terminal cysteine and, at least, one chromophore.

Assuming that formation of the **M-20** Da adduct was quantitative, comparison of the sample initial absorbance at 280 nm and its final absorbance at 280 and 332 nm would give us the ratio between the peptide molar extinction coefficient and that of the **M-20** Da adduct. Since the value of a peptide molar extinction coefficient can be known from its sequence (from the sum of each amino acid ϵ), we should be able to establish that of the final, oxidised product.

On this basis, we used the previously synthesised peptide **3.29**, which contained two tyrosine residues that would act as a good chromophore. A stock solution of this peptide was prepared and the absorbance of three different aliquots, diluted to 1000 μ L, measured at 280 nm.

After quantifying the peptide, 3 identical aliquots were left to react with CPD **3.1** at 37 °C for 2.5 h. Reaction completion end was confirmed by measuring the absorbance of the three reactions mixtures (2 replicates each) after 1.5 and 2.5 h and observing no changes.

From each of the three reaction mixtures, three aliquots were taken and their UV-Vis profile recorded three times. As the difference between the highest and the lowest absorbance values was less than 6 %, which fits the typical errors of a UV-Vis quantification (10 %), all samples were considered comparable. The mean absorbance found at 332 nm had a relative error of less than 2 %, and the estimated molar absorption coefficient was $9140 \text{ cm}^{-1} \text{ M}^{-1} \pm 2 \%$. The small differences found in the UV-Vis measurements and the small relative error of the final ϵ value made us believe that the results obtained were trustworthy.

The same process was done to establish the ϵ value of the **M-20** Da adduct at 280 nm. Again, as the difference between the highest and the lowest value was around 8.5 %, all 27 absorbance results were considered comparable. The mean absorbance found at 280 nm had a relative error around 2.5 %, and, after subtracting the tyrosines contribution, the molar absorption coefficient was found to be $4210 \text{ cm}^{-1} \text{ M}^{-1} \pm 2 \%$.

Of course, the error made in this approximation is to consider the reaction to quantitatively form the **M-20** Da adduct. Although from previous experiments we were sure that no other products absorbing at 332 nm are formed (and if so, in negligible

amounts), we had evidence that in some cases reactions small amounts of by-products were formed (typically below 10 %).

3.14 Critical overview of the methodology

The new reactivity discovered in this work presents several advantages and positive aspects:

- 1- Adducts generated from the reaction between CPDs and *N*-terminal cysteines are stable and can be used for conjugation.
- 2- Synthesis of the **M**-20 Da adducts can be performed in short times and in very mild conditions without the formation of noticeable amounts of by-products.
- 3- In contrast with the MTAs generated by reaction between maleimides and thiols, the **M**-20 Da adduct isomers are normally separable by HPLC, allowing for the isolation of stereochemically pure compounds.
- 4- The synthesis of CPDs with different substituents is feasible and does not present major problems, and these compounds have shown to furnish the desired **M**-20 Da adducts satisfactorily.
- 5- The different reactivity of CPDs towards Cys and Hcy offers the possibility to develop new methodologies for their quantification
- 6- The different reactivity of CPDs towards *N*-terminal and *C*-terminal or internal cysteines can be exploited to selectively derivatise peptides containing an *N*-terminal cysteine in a mixture of cysteine-containing peptides.
- 7- Regioselective double derivatisation of peptides containing an *N*-terminal and internal cysteines is possible thanks to the combination of CPDs and other thiol-reactive molecules such as maleimides.

However, some negative aspects must also be taken into consideration:

- 1- The reversibility of the MTAs between CPDs and simple thiols (those not possessing a 1,2-aminothiol functionality with the amine free) precludes the general applicability of these maleimide analogs to label cysteine-containing peptides.
- 2- CPDs, contrary to maleimides, cannot be used in Diels-Alder reactions in mild conditions.
- 3- The stability of the **M**-20 Da adducts is very good at slightly acidic pHs, but presents problems at basic pHs and at neutral pH for long periods of time.
- 4- The stability of the **M**-20 Da adducts is largely affected by the presence of cysteine, which is naturally found in high concentrations inside cells.

3.15 Abbreviations

ACN: Acetonitrile

ADC: Antibody-Drug Conjugate

CBT: 2-cyanobenzotriazole

CPD: 2,2-disubstituted cyclopent-4-ene-1,3-dione

DAR: Drug-to-Antibody Ratio

DEPT: Distortionless Enhancement by Polarisation Transfer

DFT: Density Functional Theory

DMSO: Dimethyl Sulfoxide

EPR: Electron Paramagnetic Resonance

ESI: Electrospray Ionisation

HOBt: 1-hydroxybenzotriazole

HPLC: High Performance Liquid Chromatography

HPLC/MS: High Performance Liquid Chromatography/ Mass Spectrometry

HRMS: High Resolution Mass Spectrometry

MALDI: Matrix-Assisted Laser Desorption Ionization

MS: Mass Spectrometry

MTA: Michael-Type Adduct

NHS: *N*-hydroxysuccinimide

NMMal: *N*-methylmaleimide

NMR: Nuclear Magnetic Resonance

PEG: Polyethylene Glycol

ppm: parts per million

SPPS: Solid-Phase Peptide Synthesis

TEA: Triethylamine

TFA: Trifluoroacetic acid

TIS: Triisopropylsilane

TOF: Time-Of-Flight

3.16 Bibliography

- (1) Baldwin, A. D.; Kiick, K. L. *Bioconjugate Chem.* **2011**, 22 (10), 1946.
- (2) Roberts, M. J.; Bentley, M. D.; Harris, J. M. *Adv. Drug Deliv. Rev.* **2012**, 64, 116.
- (3) Chan, P.-H.; So, P.-K.; Ma, D.-L.; Zhao, Y.; Lai, T.-S.; Chung, W.-H.; Chan, K.-C.; Yiu, K.-F.; Chan, H.-W.; Siu, F.-M.; Tsang, C.-W.; Leung, Y.-C.; Wong, K.-Y. *J. Am. Chem. Soc.* **2008**, 130 (20), 6351.

- (4) Eberhard, H.; Diezmann, F.; Seitz, O. *Angew. Chem. Int. Ed.* **2011**, *50* (18), 4146.
- (5) Peddi, P. F.; Hurvitz, S. A. *Future Oncol. Lond. Engl.* **2013**, *9* (3).
- (6) Senter, P. D.; Sievers, E. L. *Nat. Biotechnol.* **2012**, *30* (7), 631.
- (7) Tumey, L. N.; Charati, M.; He, T.; Sousa, E.; Ma, D.; Han, X.; Clark, T.; Casavant, J.; Loganzo, F.; Barletta, F.; Lucas, J.; Graziani, E. I. *Bioconjugate Chem.* **2014**, *25* (10), 1871.
- (8) Lyon, R. P.; Setter, J. R.; Bovee, T. D.; Doronina, S. O.; Hunter, J. H.; Anderson, M. E.; Balasubramanian, C. L.; Duniho, S. M.; Leiske, C. I.; Li, F.; Senter, P. D. *Nat. Biotechnol.* **2014**, *32* (10), 1059.
- (9) Kalia, J.; Raines, R. T. *Bioorg. Med. Chem. Lett.* **2007**, *17* (22), 6286.
- (10) Fontaine, S. D.; Reid, R.; Robinson, L.; Ashley, G. W.; Santi, D. V. *Bioconjugate Chem.* **2015**, *26* (1), 145.
- (11) Billington, S.; Mann, J.; Quazi, P.; Alexander, R.; Eaton, M. A. W.; Millar, K.; Millican, A. *Tetrahedron* **1991**, *47* (28), 5231.
- (12) Kreiser, W.; Wiggermann, A.; Krief, A.; Swinnen, D. *Tetrahedron Lett.* **1996**, *37* (39), 7119.
- (13) Liu, P.-Y.; Wu, Y.-J.; Pye, C. C.; Thornton, P. D.; Poirier, R. A.; Burnell, D. J. *Eur. J. Org. Chem.* **2012**, *2012* (6), 1186.
- (14) Pretsch, E.; Bühlmann, P.; Affolter, C.; Herrera, A.; Martínez, R. *Determinación estructural de compuestos orgánicos*; Masson, 2002.
- (15) Vippila, M. R.; Ly, P. K.; Cuny, G. D. *J. Nat. Prod.* **2015**, *78* (10), 2398.
- (16) Shahrokhian, S. *Anal. Chem.* **2001**, *73* (24), 5972.
- (17) Pompella, A.; Visvikis, A.; Paolicchi, A.; Tata, V. D.; Casini, A. F. *Biochem. Pharmacol.* **2003**, *66* (8), 1499.
- (18) Woodward, M.; Rumley, A.; Rumley, A.; Rumley, C.; Lewington, S.; Morrison, C. E.; Lowe, G. D. O. *Blood Coagul. Fibrinolysis Int. J. Haemost. Thromb.* **2006**, *17* (1), 1.
- (19) Poirier, L. A.; Brown, A. T.; Fink, L. M.; Wise, C. K.; Randolph, C. J.; Delongchamp, R. R.; Fonseca, V. A. *Metabolism.* **2001**, *50* (9), 1014.
- (20) Robillon, J. F.; Canivet, B.; Candito, M.; Sadoul, J. L.; Jullien, D.; Morand, P.; Chambon, P.; Freychet, P. *Diabète Métabolisme* **1994**, *20* (5), 494.
- (21) Herrmann, M.; Tami, A.; Wildemann, B.; Wolny, M.; Wagner, A.; Schorr, H.; Taban-Shomal, O.; Umanskaya, N.; Ross, S.; Garcia, P.; Hübner, U.; Herrmann, W. *Bone* **2009**, *44* (3), 467.
- (22) Oulhaj, A.; Refsum, H.; Beaumont, H.; Williams, J.; King, E.; Jacoby, R.; Smith, A. D. *Int. J. Geriatr. Psychiatry* **2010**, *25* (1), 82.
- (23) Zylberstein, D. E.; Lissner, L.; Björkelund, C.; Mehlig, K.; Thelle, D. S.; Gustafson, D.; Ostling, S.; Waern, M.; Guo, X.; Skoog, I. *Neurobiol. Aging* **2011**, *32* (3), 380.

- (24) Rusin, O.; St. Luce, N. N.; Agbaria, R. A.; Escobedo, J. O.; Jiang, S.; Warner, I. M.; Dawan, F. B.; Lian, K.; Strongin, R. M. *J. Am. Chem. Soc.* **2004**, *126* (2), 438.
- (25) Tanaka, F.; Mase, N.; Barbas III, C. F. *Chem. Commun.* **2004**, No. 15, 1762.
- (26) Wang, W.; Rusin, O.; Xu, X.; Kim, K. K.; Escobedo, J. O.; Fakayode, S. O.; Fletcher, K. A.; Lowry, M.; Schowalter, C. M.; Lawrence, C. M.; Fronczek, F. R.; Warner, I. M.; Strongin, R. M. *J. Am. Chem. Soc.* **2005**, *127* (45), 15949.
- (27) Ren, H.; Xiao, F.; Zhan, K.; Kim, Y.-P.; Xie, H.; Xia, Z.; Rao, J. *Angew. Chem. Int. Ed.* **2009**, *48* (51), 9658.
- (28) Liang, G.; Ren, H.; Rao, J. *Nat. Chem.* **2010**, *2* (1), 54.
- (29) Ye, D.; Liang, G.; Ma, M. L.; Rao, J. *Angew. Chem. Int. Ed.* **2011**, *50* (10), 2275.
- (30) Wang, P.; Zhang, C.-J.; Chen, G.; Na, Z.; Yao, S. Q.; Sun, H. *Chem. Commun.* **2013**, *49* (77), 8644.
- (31) Wang, H.-C.; Yu, C.-C.; Liang, C.-F.; Huang, L.-D.; Hwu, J.-R.; Lin, C.-C. *ChemBioChem* **2014**, *15* (6), 829.

Chapter 4. Cyclisation and derivatisation of peptides using 2,2-disubstituted cyclopent-4-ene-1,3-diones

4.1 First indication for the formation of a cyclic structure

In a preliminary experiment aimed at the regioselective double derivatisation of a peptide containing an internal and an *N*-terminal cysteine (see section 3.12.2), peptide H-Cys-Ser-Tyr-Ala-Cys-Lys-Tyr-Gly-NH₂ (**3.32**) was reacted for 15 minutes with 5 equivalents of **3.1**. After this time, HPLC/MS analysis showed the complete consumption of **3.32** and the formation of a product resulting from the reaction of **3.1** both with the *N*-terminal (**M-18**) and the internal cysteine (four peaks). At this point, 5 equiv. of 3-maleimidopropanoic acid were added to the reaction mixture, and HPLC/MS analyses after 3 and 5.5 h showed identical crude compositions, evidencing the formation of the desired doubly derivatised peptide **3.34** together with an unexpected adduct (**4.3**, precursors **4.1** and **4.2** appear below in **Scheme 4.2**) in an almost 1:1 ratio (see **Figure 4.1**). Noteworthy, and as expected for the reversible reaction between CPDs and internal cysteines, the products resulting from addition of **3.1** to both the *N*-terminal and internal cysteines were not observed at the end of the reaction, indicating that they had completely evolved to any of the final products detected. The crude was also analysed by HPLC/HRMS to obtain more accurate information.

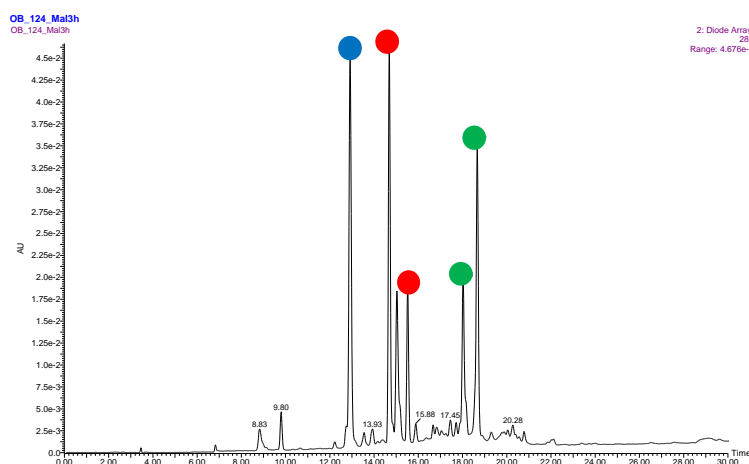
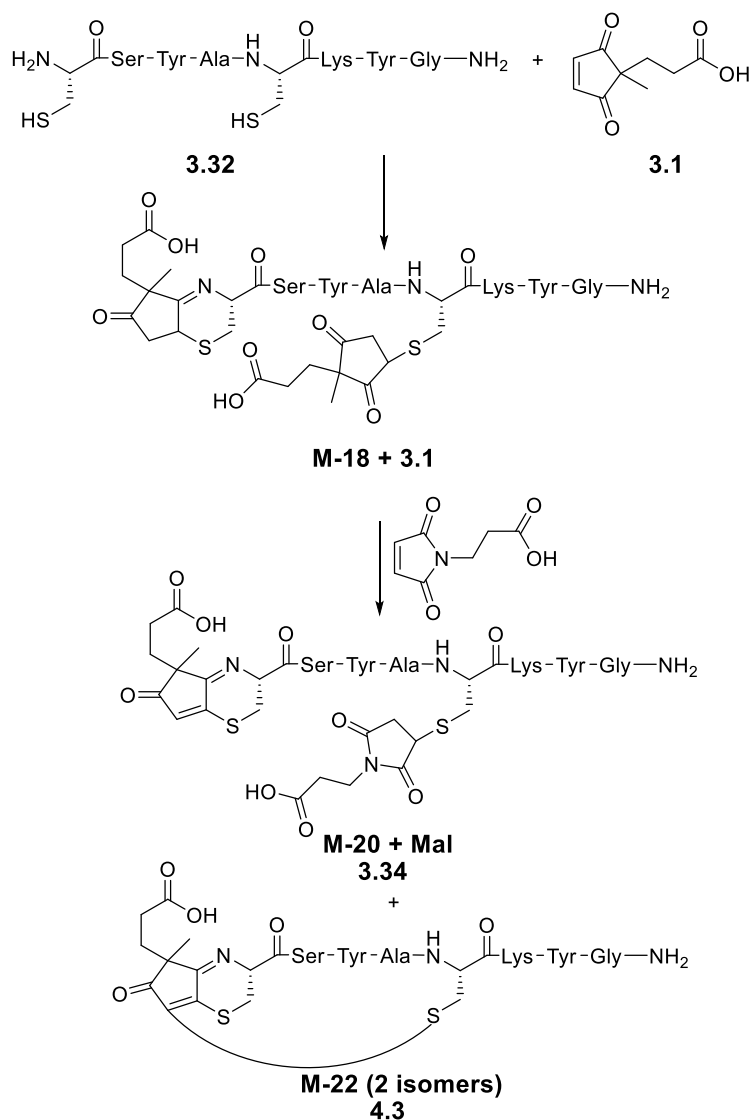


Figure 4.1. HPLC/MS (280 nm) Of the first double derivatisation experiment of **3.32**, 3 h after the addition of 3-maleimidopropanoic acid. The peak marked with a blue dot corresponds to excess 3-maleimidopropanoic acid. Peaks marked with the red and green dots are the hypothesised cyclic adduct **4.3** and the regioselectively double derivatised peptide, respectively.

The data obtained indicated that the previously unobserved adduct **4.3**, composed by two isomers, had a mass 22 Da lower than the Michael-type adduct resulting from a single CPD-thiol reaction, and a UV-Vis profile showing a maximum around 375 nm. The mass recorded was coherent with the **M-20** Da adduct (**4.2**) losing two extra H atoms, and the absorption spectrum indicated a change in the substitution pattern of the conjugated bicyclic structure of this type of adducts. We surmised that this new product could formally be the result of the substitution of the olefinic H atom of the *N*-terminal conjugated system (**M-20**) by the internal cysteine thiol, which would furnish a cyclic structure as depicted in **Scheme 4.1**. The mechanism by which the second thiol addition to the CPD ring could take place was not clear to us, and we thought that two different reaction paths could explain the formation of such cyclic motif. Either the conjugate or the radical addition of the internal cysteine thiol to the α,β -unsaturated imine with posterior oxidation could lead to the hypothesised structure.

Given the importance of cyclic peptides, as well as of their conjugates (that is, hybrid molecules incorporating cyclic peptides and another moiety), we decided to explore the use of CPDs for simultaneous peptide cyclisation and conjugation.

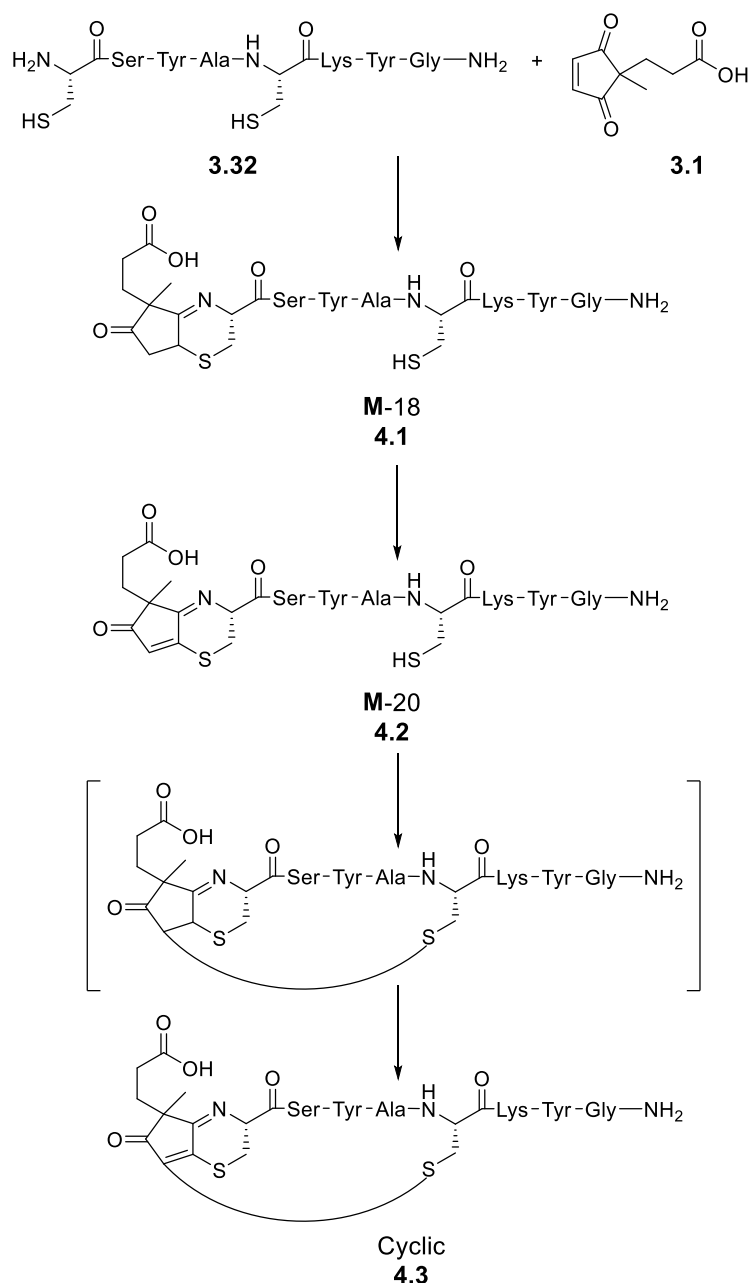


Scheme 4.1. Representation of the main steps in the reaction that furnished the hypothesised cyclic peptide for the first time.

4.2 First steps to set a procedure for the cyclisation reaction

To prove that a second thiol addition to the CPD ring was feasible and therefore that the synthesis of conjugated cyclic peptides could be achieved using CPDs, we needed to confirm the structure of the product obtained by NMR. Of course this entailed the need for large enough amounts of product, which, in turn, meant that we needed to obtain this type of compounds in good yields. Therefore, some attempts to optimise the synthesis of **4.3** were performed.

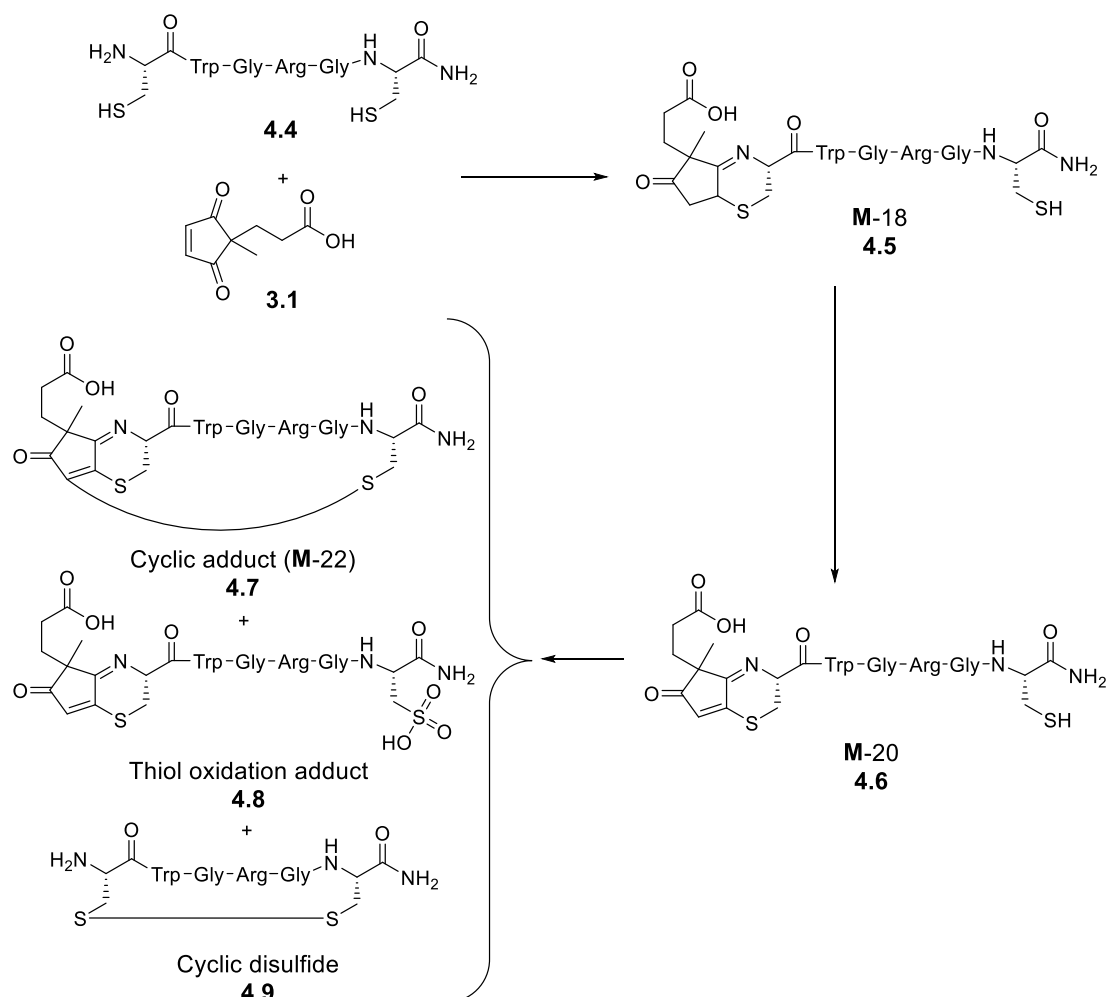
First, we hypothesised that it would be better to obtain the **M**-20 Da adduct **4.2** and, once this had been formed, study how to induce what we believed was the cyclisation process (see structures in **Scheme 4.2**). With this idea in mind, **3.1** and **3.32** were incubated in a 1.1:1 ratio at 60 °C and the outcome of the reaction followed by HPLC/MS. After 15 minutes all peptide had been consumed to furnish the corresponding **M**-18 Da adduct **4.1** and small quantities of **4.2**. Little amounts of products arising from the reversible addition of the CPD to the internal cysteine were also found due to the slight excess of **3.1** (not shown in **Scheme 4.2**). The reaction was analysed again at 45 minutes to see that this time **4.2** was the major product, but some traces of **4.3** and **4.1** could be detected along with unidentified impurities. 45 Minutes later **4.3** was the major product as judged by peak integration (280 nm), but several impurities and still important amounts of remaining **4.2** were also detected. Importantly, and as expected for a reversible reaction, the CPD double addition products observed at the early reaction stages had disappeared.



Scheme 4.2. Structures of the products generated upon reaction of **3.32** with **3.1**.

This experiment revealed that controlling the reaction evolution was not easy and that completely separating the **M-18** Da oxidation from the supposed cyclisation was not feasible. Further, it seemed that there was no need to add any external agent to induce the formation of **4.3** and, therefore, that synthesis of this type of structures could be temperature-promoted. However, further tuning of the reaction needed to be done to reduce the amount of side-products.

The same experiment was repeated with peptide H-Cys-Trp-Gly-Arg-Gly-Cys-NH₂ (**4.4**) but this time the reaction was followed only by HPLC (as the UV-Vis profile of each peak was enough to establish its identity, see structures of the products in **Scheme 4.3**).



Scheme 4.3. Structures proposed for the products generated upon reaction of **4.4** with **3.1**.

After 45 minutes no trace of the **M-18 Da** adduct (**4.5**) was found, and the conversion to the **M-20 Da** product (**4.6**) had taken place cleanly. Analyses at 90 and 180 min revealed the presence of the presumed cyclic adduct **4.7** together with remaining **4.6**, indicating a slower conversion as compared to that of **4.2** to **4.3**. It was not until 5 h that **4.6** had completely disappeared, but at this point the presence of 4 peaks with an absorption maximum corresponding to what we believed was the cyclic adduct was evident.*

* This time the absorption maximum was at 362 nm due to the change of solvents performed when moving from HPLC/MS to HPLC. The influence of the solvent on the absorption maximum has been confirmed in independent experiments.

HPLC/MS Confirmed that all four peaks had the same mass, that corresponding to the cyclic peptide. Since only two isomers were expected for the cyclic entity, we inferred that the other two peaks had to be the result from a side reaction not producing any global change on the number of atoms but an isomerisation. Our hypothesis was that the *N*-terminal cysteine had undergone epimerisation at the α -carbon (for a detailed explanation see section 4.11).

Whatever the cause for the appearance of four peaks with the same mass and UV-Vis profile was, it seemed clear that heating the reaction at 60 °C was not the best way to obtain **4.3** or **4.7**. Further, we thought it would be more practical to work with peptide **4.4** in the next experiments rather than with **3.32** because, as stated above, the former showed to be less prone to form the corresponding **M**-22 Da product (hypothesised cyclic peptide). Consequently, any method to properly generate **4.7** would be applicable for the formation of **4.3**.

Prompted by the idea that some additions of thiols to alkenes can be radical reactions depending on the nature of the reagents,^{1,2} we decided to assay the use of radical oxidising species to induce the formation of **4.7**. TEMPO Was selected as an additive to the reaction for its commercial availability, its known stability and its ability to participate in oxidation processes.³

In a first experiment, **4.4** was reacted with **3.1** (1.2 equiv.) at 60 °C until the formation of **4.6** was almost complete (2 h, assessed by HPLC). At this point the temperature was lowered to 37 °C and the reaction split in four aliquots to test the effect of different amounts of TEMPO. 0.2 Equiv. of TEMPO were added to aliquot A, 1.0 equiv. to aliquot B, 5.0 equiv. to aliquot C and, finally, 10.0 equiv. to aliquot D.

The positive effect of this radical on the cyclisation reaction was proven by the fact that **4.7** had been formed faster than in the previous experiment even though the reaction temperature was lower. However, in aliquots A, B and C **4.6** had not disappeared completely, and analyses performed within one or two hour difference revealed lack of evolution during these time points. Further, increasing the amount of TEMPO from 0.2 to 1.0 or 5.0 equiv. had not induced relevant changes in the observed **4.6** to **4.7** ratios. Aliquot D was the only showing complete consumption of **4.6**, but the formation of **4.7** was accompanied, almost in equal amounts, by a product (**4.8**, see **Scheme 4.3**) whose mass was 48 Da higher than that of **4.6**. As its UV-Vis profile indicated that the structure of conjugated bicycle had been conserved, we attributed **4.8** to the species resulting from oxidation of the thiol group of the *C*-terminal cysteine to sulfonic acid.

Although the effect of TEMPO was clearly positive, the fact that no major changes were observed when using amounts of TEMPO up to 5.0 equivalents after a certain time-point suggested that the half-life of this species in the reaction medium was somewhat short.

This experiment also showed that side-reactions caused by a large excess of TEMPO could take place, meaning that it would be better to use amounts as tiny as possible of said radical.

Based on these evidences, we asked ourselves if it would be possible to achieve complete conversion of **4.6** to **4.7** using sub-stoichiometric amounts of TEMPO, in particular, whether the continuous addition of small amounts of this radical would compensate for its short half-life in the reaction crude and help avoiding the formation of **4.8**.

To answer these questions and, at the same time, check the effect of peptide concentration during the supposed cyclisation, **4.4** was reacted with 1.2 equiv. of **3.1** until almost complete formation of **4.6**. At this point, the reaction was split in two different aliquots. 0.2 Equiv. of TEMPO were added to each aliquot, but the peptide concentration of aliquot A was adjusted to 0.1 mM while that of aliquot B remained at 0.5 mM. Both aliquots received three extra TEMPO additions (again 0.2 equiv.) 1, 2 and 2.5 h after the first one, totalling 0.8 equiv. of reagent. Aliquot A showed almost complete disappearance of **4.6** after 2 h (only 5 % remaining, as shown by the HPLC trace at 280 nm) and total conversion after 3. The cyclic product yield determined by HPLC at 280 nm was 76 %, while **4.8** and the cyclic disulfide (**4.9**) accounted for an 8-9 % each. Other minor products were those arising from the supposed isomerization of the *N*-terminal cysteine. On the other hand, consumption of **4.6** was complete after 1 h in aliquot B and formation of **4.7** had taken place in a slightly better yield (80 %). Both experiments proved that formation of **4.7** could be accomplished using sub-stoichiometric amounts of TEMPO and the results obtained from aliquot A demonstrated that continuous additions of this radical helped compensate for its fast disappearance from the reaction crude. Further, the positive effect of increasing peptide concentration was clear.

After these experiments we conjectured if it was completely necessary to wait for the complete formation of the **M-20 Da** adduct before performing the first TEMPO addition. We believed that it would be advantageous to carry out this addition when the **M-18/M-20** ratio was approximately 1:1 (as determined at 280 nm), as this would save time and a less strict control of the oxidation reaction would be required.

For this purpose, **4.4** and **3.1** were again reacted at 60 °C but this time for 45 minutes. At that point HPLC analysis of the crude at 280 nm revealed that the reaction had reached the desired **M-18/M-20 Da** adducts ratio. Temperature was then lowered to 37 °C and 0.2 equiv. of TEMPO were added. After 1 h the crude still showed small amounts of **4.6**, so 0.2 equiv. of TEMPO were again added and the reaction run for 30 more min. Analysis after this time revealed complete disappearance of **4.6** and a clean crude containing almost 85 % of **4.7** (**Figure 4.2a**), confirming that it was not necessary to wait

for the complete formation of the **M-20 Da** adduct to start adding the radical species. Repetition of this experiment revealed that some times, and for unknown factors,[†] it was necessary to keep performing additions of 0.2 equiv. of TEMPO every 30 minutes to see the reaction completely finished. In all the cases, the total amount of TEMPO used remained equal or below 0.8 equivalents.

Peptide **3.32** gave similar results when reacted with **3.1**, as expected for a peptide that had shown to cyclise easier (**Figure 4.2b**).

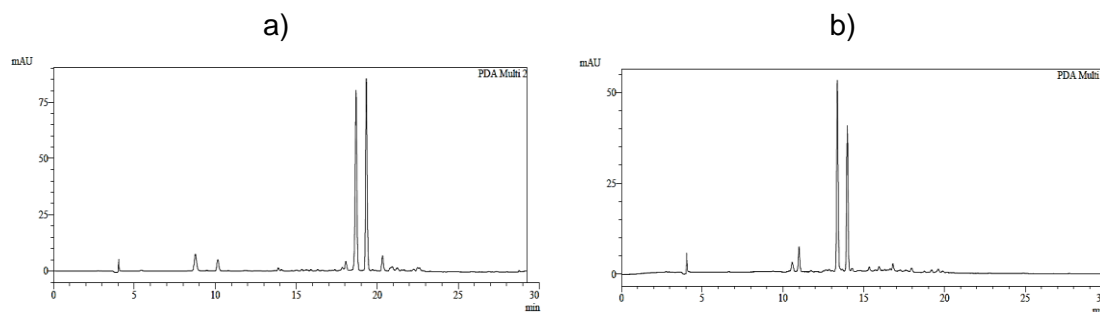


Figure 4.2. HPLC traces (280 nm) of the **4.4 + 3.1** reaction (a) and of **3.32 + 3.1** (b). The two major twin peaks in each chromatogram correspond to the desired products.

Given the good yields obtained with this protocol, we believed that the cyclisation reaction was optimized enough to move to next step, that is, the structural characterization of the supposed cyclic adduct. The protocol to be followed for the cyclisation was, then, to react the peptide (0.5 mM) and the CPD (1.2 equivalents) at 60 °C until an almost 1:1 **M-18/M-20 Da** adducts ratio was accomplished (45 minutes). Subsequently, temperature had to be lowered to 37 °C and 0.2 equivalents of TEMPO added. After 1 h, 0.2 equiv. of TEMPO were to be added to the crude every 30 minutes until complete disappearance of the **M-20 Da** adduct was observed (1 or 2 additions were normally enough).

4.3 Confirmation of the cyclic structure by NMR

As in the case of the **M-20 Da** adduct (chapter 3), we needed to confirm cyclisation of the peptide by NMR.

At this moment, Lewis Archibald joined the group and this particular project. Experiments described from now on have been carried out in collaboration with him. We have teamed

[†] Most probably small errors in the quantification of the CPD slowed down the oxidation of the **M-18 Da** adduct, consequently diminishing the rate of the overall process.

up to study the CPD-mediated cyclisation and derivatisation of peptides (see appendix 2).

In order to simplify the NMR spectra of the, in principle, cyclic product, it would be best to work with a peptide containing only a few amino acids. This peptide should possess two cysteine residues, one of which at the *N*-terminal position, and a chromophore in order to be able to quantify it by UV-Vis spectroscopy before the reaction. Other amino acids were chosen so that they yielded easily identifiable signals in the NMR spectrum. Based on these considerations, we chose the peptide H-Cys-Ser-Tyr-Ala-Cys-NH₂ (**4.10**) as a model to confirm formation of the cyclic structure by NMR. Tyrosine was selected as a chromophore due to its lower tendency to undergo side-reactions (in this sense tryptophan is more sensitive), and serine and alanine because of their side-chain simplicity and, at the same time, due to their characteristic methyl and methylene signals in the NMR spectra. **3.1** Was therefore reacted with **4.10** to furnish **4.13** (**4.11** and **4.12** appear in section **4.7**).

Confirmation of the formation of a cyclic compound would come upon observation of the following (see **Figure 4.3** for support):

- 1- In the ¹H NMR spectrum:
 - a) Absence of the olefin signal present in the **M**-20 Da adduct.
- 2- In the ¹³C and DEPT NMR spectra:
 - a) Absence of the CH olefin signal corresponding to the **M**-20 Da adduct.
 - b) Presence of two quaternary carbon signals in the olefin area.
- 3- In the HMBC spectrum:
 - a) Correlation of the *C*-terminal cysteine β-hydrogens with one of the above-mentioned quaternary carbon signals in the olefin area.

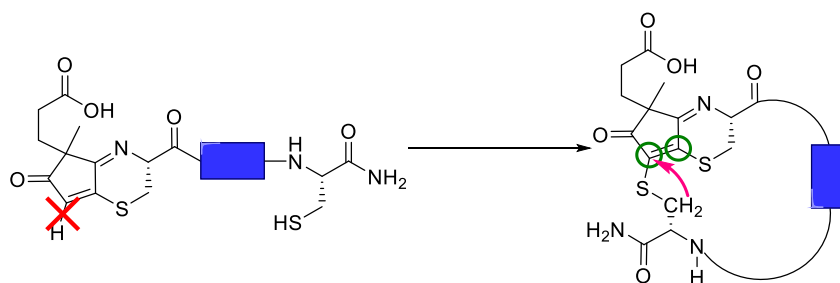


Figure 4.3. Diagnostic signals confirming the cyclisation process. Red marks are signals that should disappear and green marks signals that should appear. The magenta arrow indicates the HMBC correlation mentioned in point 3.

The synthesis of the hypothesised cyclic adduct was accomplished following the protocol previously established with minor variations.

To avoid complications or ambiguities during the NMR characterization, the two **4.13** diastereomers were purified and separated by HPLC. Each of them was subjected to ^1H , ^{13}C , DEPT, COSY, HSQC and HMBC NMR analyses. D_2O was used as solvent for the NMR experiments, although it was observed that **4.13** was not very soluble and tended to precipitate in this medium. ^1H and ^{13}C showed that the purification process had not been completely effective, as some CPD- and TEMPO-derived impurities not connected to the peptide could be detected. However, these contaminations, which were different for isomer 1 and isomer 2, did not suppose a problem to identify the signals corresponding to each **4.13** isomer. Furthermore, ^1H , HSQC and HMBC experiments revealed that these impurities were present in sub-stoichiometric amounts with respect to the cyclic adduct and that they did not correlate with it in any way.

Assignment of each peak was achieved as follows. Each ^{13}C peak was assigned as a primary, secondary, tertiary or quaternary carbon using the DEPT spectrum. Afterwards, HSQC was used to link each of them to its corresponding H atoms. The tyrosine aromatic system was used as a starting point to ensemble the adduct structure due to its ease to be identified. HMBC Correlations of the aromatic ring allowed for the identification of its α and β carbons. The carbonyl amide signal correlating with both the α and β hydrogens was that of the own amino acid, while the carbonyl correlating only with the α hydrogens was that of the amino acid at its *N*-terminus. In a third step and from these amide signals, HMBC correlations permitted to identify the α hydrogens of the amino acids located at the tyrosine *C*- and *N*-termini. From these signals, the side-chains of said amino acids could be pinpointed (see **Figure 4.4**). Expansion of this methodology towards both the *N* and *C*-terminal positions permitted to link each C and H atom of the peptide structure to an NMR signal.

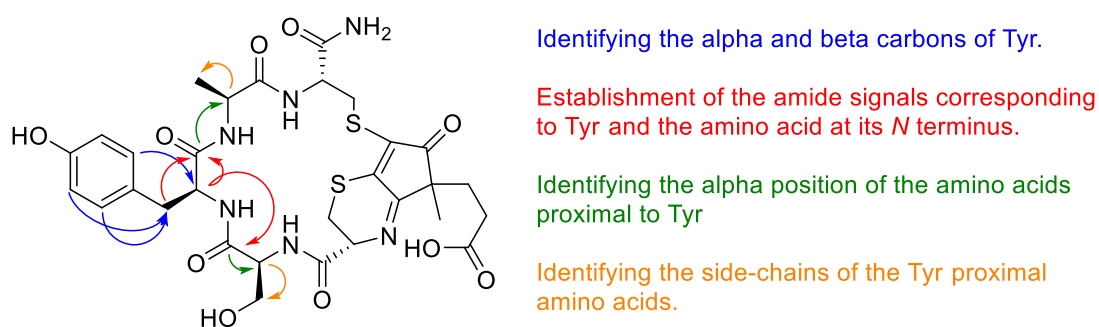


Figure 4.4. Schematic representation of the method used to assign the NMR signals of the peptide using the HMBC experiment.

Then, starting from the acid carbon of the CPD, identified because it is the only carbonyl correlating with two methylenes, the CPD moiety was constructed. Differentiation between the ketone and imine carbons was not only achieved from their different chemical shift, but also because the latter showed a correlation with the *N*-terminal cysteine α hydrogen. This, ultimately, allowed to unambiguously differentiate between the α carbons of the serine and *N*-terminal cysteine residues. HMBC Also helped to discard those signals corresponding to impurities, as correlations between said signals were observed while no correlations with the **4.13** signals could be found. The only unidentified carbons were two signals around 143 and 148 for isomer 1 and around 143 and 150 ppm for isomer 2, corresponding to quaternary carbons, which were assigned to the tetrasubstituted olefin of the CPD ring. The COSY spectrum and the knowledge of the peptide sequence were used post-assignment to confirm that the proposed structure was correct.

As mentioned in points 1 and 2 at the beginning of this section, we did not observe any signal in the olefin zone, nor in the ^1H neither in the ^{13}C spectra, belonging to a CH, ensuring that there was no **M-20** Da adduct. Finally, and in agreement with the previously established in point 3, a clear cross peak correlating the C-terminal cysteine β hydrogens and the carbon at 148 or 150 ppm was observed, confirming that cyclisation had taken place.

4.4 Determination of the molar absorption coefficient

As for the **M-20** Da adduct (see chapter 3), the molar absorption coefficient of the cyclic adducts needed to be determined. We decided to use NMR titration to determine the amount of cyclic product in a given sample and then measure the absorbance of solutions prepared from that sample to determine the molar absorption coefficient.

In this technique, a known amount of titrating agent or standard is added to a solution containing the compound to be quantified. NMR Integration of both species provides information about their relative ratio once the amount of ^1H nuclei integrated for each molecule is taken into account, allowing for the quantification of the compound of interest. This technique permits to work with tiny amounts of compound and still be accurate in the quantification.

After quantification, removal of the solvent and the standard gives us a known amount of pure compound ready to be dissolved to obtain a solution of known concentration, from which the molar absorption coefficient can be determined after measuring its absorbance.

However, several factors needed to be taken into account to be able to use NMR titration for our purposes:

- 1- The compound of interest had to be quantitatively recovered from the NMR tube. This could be achieved by performing several washes of the NMR tube and carefully collecting all of them.
- 2- Both the solvent and the titrating agent employed in the quantification had to be removable under conditions that did not imply any risk of losing sample. From our point of view, this was to be best achieved by lyophilisation or smooth evaporation. Therefore, only volatile or lyophilisable solvents and titrating agents were to be used.
- 3- The titrating agent had to possess an easily identifiable signal appearing in a clear region of the spectrum. Further, said signal should ideally have a well-defined shape to allow for an accurate integration.
- 4- The relative signal ratio between titrating and titrated compounds should not be very dissimilar in order to minimize errors, as comparable signals will always provide more accurate results. For this purpose, we decided to work with relative signal ratios ranging from 1:5 to 5:1.
- 5- In order to minimize accuracy-affecting errors, all solutions of the titrating agent had to be prepared by weighing the titrating agent.
- 6- The amount of titrating agent solution added would be determined by measuring mass differences instead of volumes added, as weighing is more accurate and no approximations to the titrating agent solution density should be done.
- 7- Since the relative integrations of the titrating and titrated compounds had to reflect their relative ratios, there should be no influence on the ability of each H nucleus to relax after the electromagnetic pulse. To be sure that this was the case, the NMR experiments needed to be performed with long relaxation delays to ensure complete relaxation.

Furthermore, and to be sure that the method was linear (that is, that integrations of the titrating and titrated compounds were directly proportional to their relative ratios), we decided to perform several additions of the titrating agent to the compound to be quantified. A linear relationship between the amount of added titrating agent or standard and the relative integration of the two compounds would indicate that this method was adequate to quantify the cyclic adduct. The amount of cyclic adduct would be obtained from **Equation 4.1** after representing the relative integrations of standard and cyclic compound versus the amount of standard added.

$$\frac{I_{St}}{I_X} = \frac{H_{St}}{H_X} \frac{1}{n_X} \times n_{St}$$

Equation 4.1. The amount of the cyclic adduct can be obtained from the gradient of the line of best fit obtained after representing the relative integration of the standard to the cyclic adduct in front of the amount of standard added. St and X subindexes indicate the standard or titrating agent and the titrated compound, respectively. I, H and n represent the integration of each signal, the number of nuclei integrated and the number of mols, respectively.

The cyclic adducts generated (**4.13**) upon reaction of peptide **4.10** and CPD **3.1** used for the structural elucidation were also employed for the ϵ determination. Each of the isomers generated (referred to as isomer 1 or isomer 2 according to their elution order in the HPLC) was independently titrated and its molar absorption coefficient determined. Since there is no way to ensure that the elution order of the two different isomers will be the same for every CPD-peptide combination, at least without complete structural analyses, the mean value of the two independently obtained molar absorption coefficients can be used for the global yield determination of each cyclic product.

Regarding NMR quantification, solubility problems encountered in the structural determination forced us to use a mixture of D₂O and d₆-DMSO as solvent to ensure complete dissolution of **4.13** over the titration process. As titrating agent we used 1,4-dioxane because it appeared in a blank region of the spectrum and its signal was a singlet. This was selected of the best quality available and filtered through basic alumina to remove trace amounts of peroxides; then, it was stored over calcium hydride prior to its use.

The proton signal observed at 4.3 ppm for both isomers was used to establish the standard/cyclic adduct ratio. The reason behind this election was that this signal appeared isolated from all the others, had a defined shape and was close enough to the standard signal to minimize possible integration errors due to the phase correction. Further, when this proton was used as reference, the integration value of the other signals was very close to the theoretical, indicating that it was a good indicator of the amount of cyclic product. Unfortunately, as the addition of DMSO caused a shift in the signals in comparison to the spectra recorded in pure water, it is not possible to assign the signal integrated. Comparison with the spectra used for the structural determination points towards this signal being one of the β hydrogens of the C-terminal cysteine or the α hydrogen of the serine residue.

To minimise human errors in integration, an integration template was used in the MestReNova software.

After quantification, 1,4-dioxane and deuterated DMSO were removed by adding a large excess of water to the NMR sample and lyophilising the resulting aqueous solution.

An example of the NMR profile obtained upon titration is shown in **Figure 4.5a**. Representation of the NMR relative integration versus of the amount of 1,4-dioxane added gave, as expected, a linear relationship for both isomers. The line of best fit gave as a result R^2 coefficients of 0.9989 and 0.9999, indicating a good precision in the titration process and a completely linear behaviour (**Figure 4.5b**). The good R^2 coefficients and the points dispersion along the line of best fit showed that random errors were done within an acceptable range. All in all, these facts indicate that the quantification values obtained are trustworthy. The only point previously mentioned that was not completely fulfilled was the ratio of titrating agent to cyclic compound, that in the case of isomer 2 reached a 6:1 ratio and in the case of isomer 1 a 5.3:1 ratio. However, this was considered merely anecdotic as it was only a small deviation from our planned range and we could not observe any negative effect in the titration experiment due to this fact. As to the ϵ determination, each isomer was dissolved in DMSO and diluted with water to a known concentration reaching a maximum DMSO value of 2 % (v/v), and the absorbance of different dilutions of these solutions was measured.

Molar absorption coefficients were independently determined in water at 373 nm (the absorbance maximum) for each isomer from the gradient of the lines of best fit shown in **Figure 4.6**. As can be seen, the R^2 coefficients showed good precision and the obtained ϵ values were $13023 \text{ L mol}^{-1} \text{ cm}^{-1}$ for isomer 1 and $11408 \text{ L mol}^{-1} \text{ cm}^{-1}$ for isomer 2. Again, the R^2 coefficients indicated acceptable random errors.

Therefore, it was concluded that the ϵ value to be used for the quantification of the different cyclic products in water was $12215 \text{ cm}^{-1} \text{ M}^{-1}$ at 373 nm. However, as experiments performed at a later stage showed that cyclic derivatives obtained with the dansyl and biotin CPD derivatives were mostly insoluble in water, the molar absorption coefficient of cyclic products was also determined in methanol. For this purpose, the already quantified isomer 1 was dissolved in methanol and the absorption of different dilutions made thereof measured (isomer 2 was lost during its manipulation and consequently its molar absorption coefficient in methanol could not be determined). Solvent change induced a hypsochromic shift of the maximum, which was displaced to 365 nm. Again, the gradient of the line of best fit yielded the ϵ value in methanol, which was found to be $12643 \text{ M}^{-1} \text{ cm}^{-1}$ at 365 nm (**Figure 4.7**).

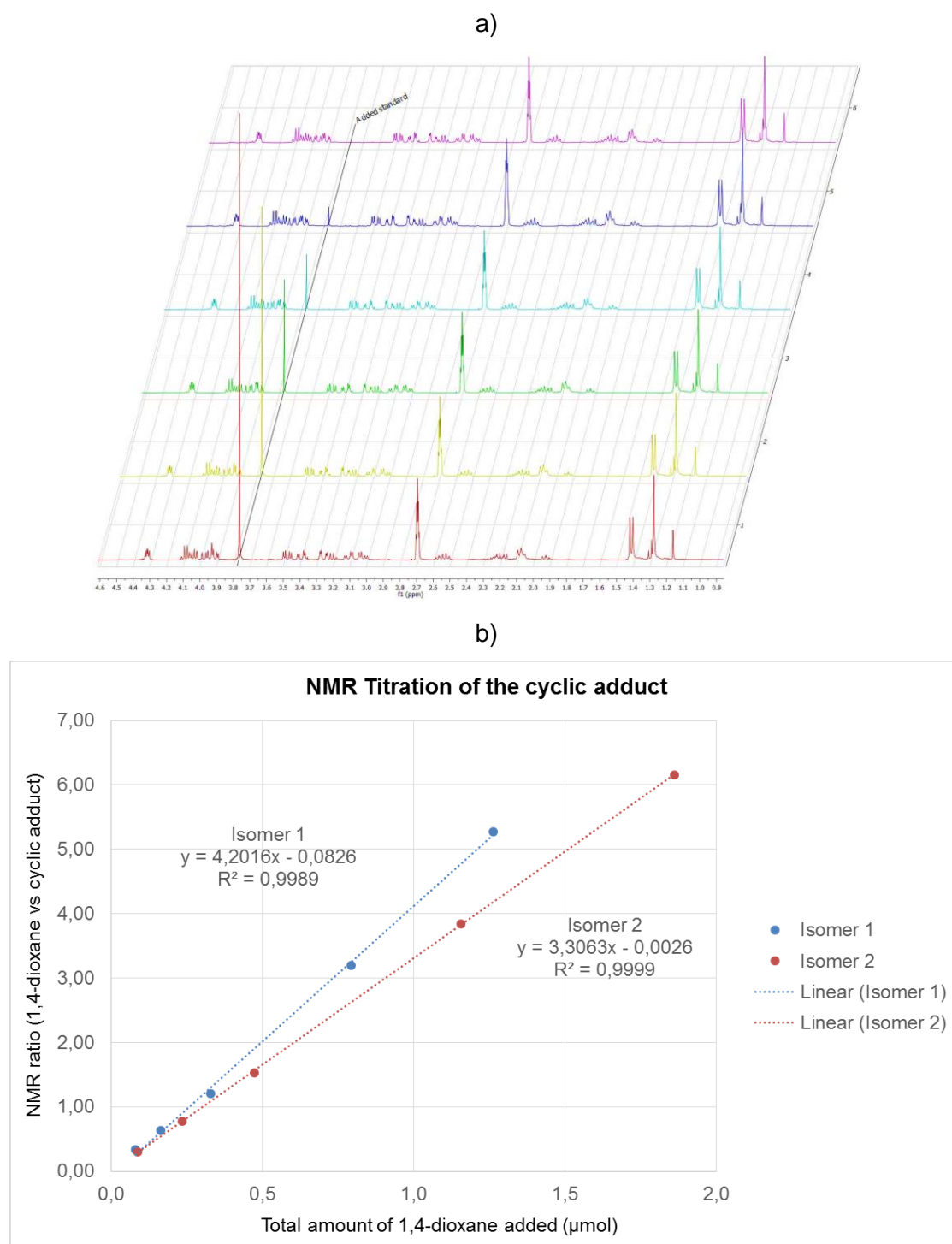


Figure 4.5. Overlapped NMR profiles from the titration experiments of cyclic isomer 1, with increasing amounts of 1,4-dioxane from top to bottom (a). The amount of cyclic product was obtained from the gradient of the line of best fit upon representation of the NMR ratio of 1,4-dioxane to cyclic adduct in front of the amount of 1,4-dioxane added (b).

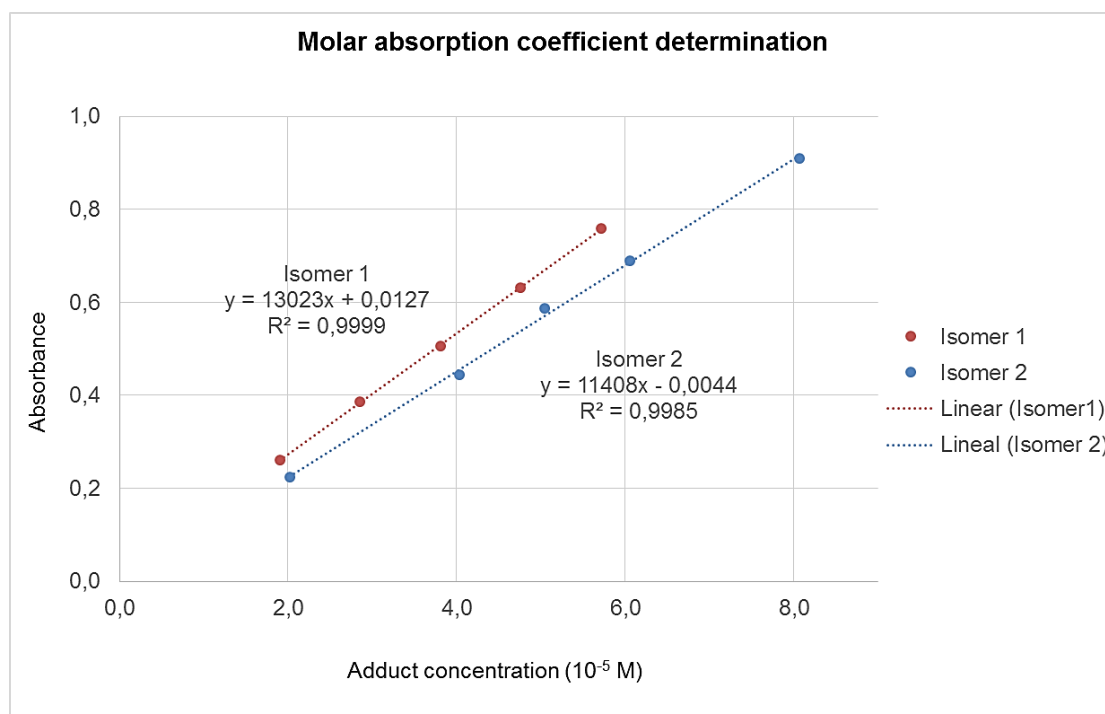


Figure 4.6. Determination of the molar absorption coefficient in water (at 373 nm) for both cyclic isomers.

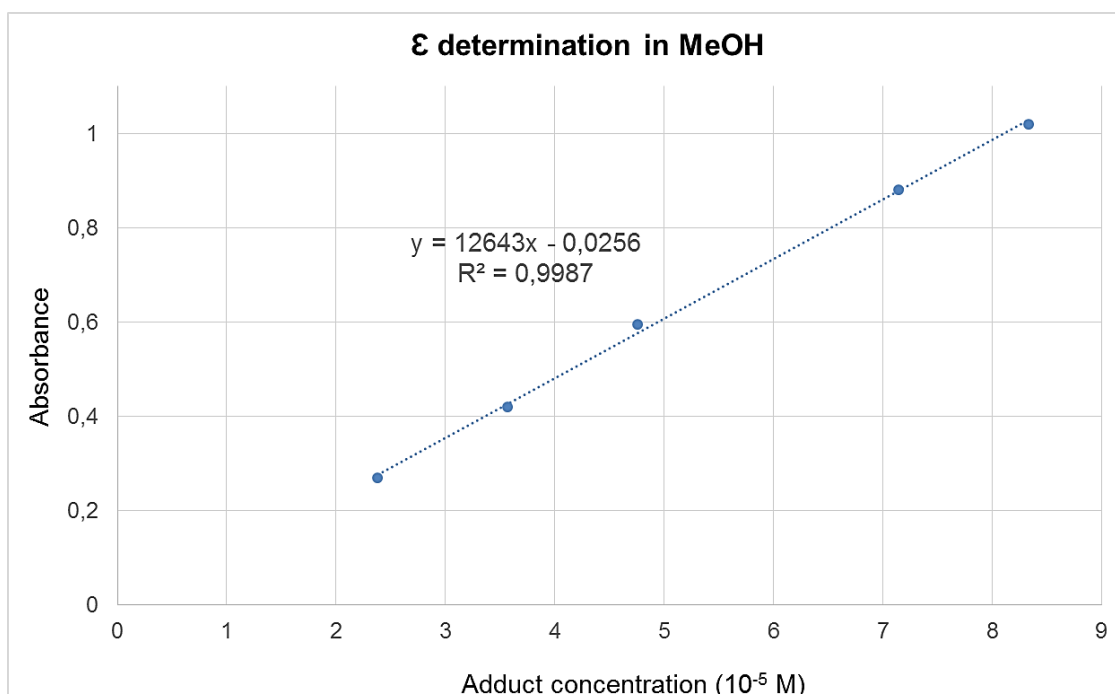


Figure 4.7. Determination of the molar absorption coefficient of the cyclic adduct in methanol (at 365 nm).

4.5 Search of alternatives to the use of TEMPO

After confirming the cyclic structure and determining the molar absorption coefficient of the linking moiety, we decided to go back to the cyclisation process. In order to avoid the use of radical reagents (which as seen before could trigger the formation of oxidation by-products) and assess whether the cyclisation reaction could be performed through the nucleophilic addition of the free thiol to the conjugated bicyclic structure, several experiments were performed with the **4.4** + **3.1** reaction.

To begin, peptide **4.4** was reacted with **3.1** at pH 8.0 and 5.0, using a 0.1 M phosphate buffer and a 0.1 M acetate buffer, respectively. The outcome of both reactions was analysed by HPLC to find that, in the first case, the peptide had formed the cyclic disulfide without reacting with **3.1** and, in the second, that the crude contained a very complex mixture of products. It seemed clear that basic pH values precluded the reaction with the CPD by favouring the intramolecular disulfide formation, and that either pH 5.0 or the buffer selected were not adequate to run the reaction.

To keep evaluating the possibility of catalysing a possible polar addition of the C-terminal cysteine thiol to the conjugated double bond, the effect of two different bases was studied. For this purpose, the **3.1** + **4.4** reaction was left at 60 °C until the formation of **4.6** was complete (no more **4.5** detected). The reaction temperature was lowered to 37 °C and the effect of 1 and 2 equiv. of both NaOH and NEt₃ was assessed by HPLC/MS. Prior to the addition of base, though, the concentration was lowered to 0.1 mM instead of using the usual 0.5 mM to avoid base-catalysed intermolecular disulfide formation.

Before the addition of bases the crude showed no trace of the cyclic product. In this particular case the degree of **4.6** isomerisation, likely due to the prolonged heating time, was 10 %, as inferred from the observation of four peaks with its mass and UV/Vis spectra.

Analysis of the reactions 2 h after the addition of base revealed that neither NaOH nor NEt₃ had triggered the formation of the cyclic adduct significantly, but instead, and as judged from the integration of the four **4.6** peaks, they had a negative effect on its diastereomeric purity (see **Table 4.1**).

Base	NaOH		NEt ₃	
Equivalents	1	2	1	2
Isomerisation (%)	20	37	19	26

Table 4.1. Percentage of isomerisation as determined by HPLC/MS (280 nm) 2 h after the addition of base.

NaOH induced more isomerization than NEt₃, both when using 1 or 2 equivalents, although in this last case the difference was more evident. Albeit this process started to take place before the addition of any base, it was clear that increasing amounts of base caused a major isomerisation of **4.6**.

Some months after these experiments we discovered that the cyclic adducts are also prone to isomerise in basic media, especially when they are synthesised using L-cysteine at the *N*-terminal position (see section **4.10** and **4.11**). Therefore, we can conclude that even if the addition of base had successfully triggered the formation of the cyclic product, it would probably have had a negative effect on the final product stability.

No more experiments employing other bases were performed, as it was considered that the amount or the nature of the base necessary to effectively induce the cyclisation through a polar mechanism would, at the same time, cause a large degree of isomerisation.

After the failed attempts to cyclise **4.6** by the addition of bases, we decided to try DMSO as mild oxidising agent. DMSO is commonly used for the formation of disulfide bridges in peptides or proteins,⁴ and as our cyclisation is, formally, an oxidation reaction, we thought it was a good idea to try DMSO as additive in our reaction. The same factor that made it worth trying, though, also implied a problem. Cyclisation by disulfide formation had proven to be one of the side-reactions taking place in CPD-mediated cyclisation and, therefore, adding DMSO at the beginning of the reaction seemed a bad idea. For this reason, **4.4** and **3.1** were incubated at 60 °C until almost complete conversion to **4.6** (verified by HPLC), at which point the temperature was lowered to 37 °C, the reaction concentration lowered from 0.5 to 0.1 mM and a 10 % DMSO (v/v) was added to the reaction mixture. Analysis of this crude after 3 h by HPLC/MS revealed a complex crude, with a lot of products present in it (see **Figure 4.8**). The reasons for this bad result are unclear because we could not identify most of the side-products, but subsequent experiments revealed fast changes in the **M**-20 Da adduct upon addition of DMSO (instantaneous colour change), which would fit with an incompatibility of this type of structures with the mentioned sulfoxide (data not shown).

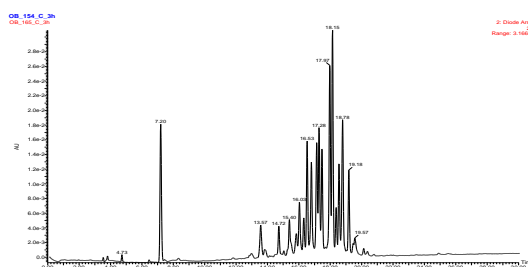


Figure 4.8. HPLC/MS trace (280 nm) of the **4.4 + 3.1** reaction crude 3 h after the addition of DMSO.

Finally, O₂ was tried as an oxidising agent. In this case, though, the reaction conditions were different than those used so far. **4.4** Was reacted with **3.1** at 37 °C in a 2 molal LiCl solution (see section **4.8**) in the presence of oxygen, and the effect of this diradical on the reaction outcome was analysed by HPLC/MS. A first attempt was carried out bubbling oxygen to the reaction mixture for 10 minutes and maintaining an O₂ atmosphere during the reaction progress. Analysis of the reaction after 2 h showed complete disappearance of **4.6** and the formation of **4.7** (see **Figure 4.9**). However, a side-product accounting for the 14 % of the crude (determined at 280 nm) was detected. Its identity could not be elucidated because it did not ionise and its UV-Vis profile had not been previously observed (two maxima, one at 274 and another at 313 nm). We believed said species appeared due to secondary oxidation reactions triggered by the O₂ atmosphere, although the above-mentioned properties confirmed that it was not **4.8**, the previously observed by-product generated by oxidation of the C-terminal cysteine thiol (which does ionise and absorbs at 330 nm). Further, two peaks eluting at 3.0 and 3.9 min retention times were also detected, accounting for a 21 and a 7 % of the crude, respectively. The peak at 3.0 min was identified as the cyclic disulfide derivative **4.9** (by comparison with other experiments), but it was not possible to establish the identity of that eluting at 3.9 min because the MS detector had to be disconnected during the first 5 min of the analysis to avoid LiCl deposition in it.[‡]

[‡] Carefully examining the reaction outcome made us realise that we should have used a different gradient for the HPLC analysis to avoid this ambiguity. However, the experiment was not repeated because the final result was, anyway, not good enough.

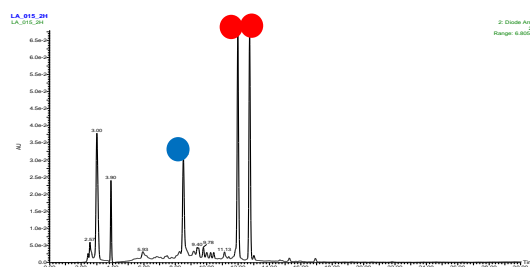


Figure 4.9. HPLC/MS trace (280 nm) of the first attempt employing O_2 as oxidant instead of TEMPO. Red marks correspond to the desired cyclic adduct, the blue mark is cyclic disulfide product.

A second experiment with lower amounts of O_2 was performed to try to diminish the percentage the newly generated by-product. In this case, O_2 was bubbled through the reaction for only 2 minutes, and the reaction vessel sealed without maintaining any special atmosphere. Analysis of the reaction mixture after 2 h revealed a slower **4.6** to **4.7** conversion, and 3 h were found necessary to complete this process. The amount of the new by-product generated had slightly decreased, although it still accounted for the 12 % of the reaction crude, while that of **4.9** and the peak eluting at 3.9 minutes were considerably reduced, accounting now for a 15 and a 2 % of the crude, respectively. We assumed that the amount of O_2 was still too high, and therefore a third experiment was carried out bubbling air to the reaction mixture for 10 minutes, which in principle should reduce even more the concentration of these species. Thanks to this, the peak at 3.9 min had almost disappeared (1 %) and the amounts of **4.9** and new by-product were 15 and 10 %, respectively.

However, although the reaction worked well enough in these last conditions (~ 70 % yield of cyclic product) the use of TEMPO was considered better because it provided better yields of cyclic peptide and it was easier to control the equivalents added to the reaction. However, it is worth considering air as a possible alternative to TEMPO in cases where its use is not recommended.

4.6 Use of labelling CPDs in cyclisation reactions - 1

After optimisation of the reaction using a model peptide and CPD **3.1** it was time to start working with more complex peptides and CPDs. Therefore, peptide **4.4** was reacted with **3.14** and **3.15**, two CPDs that may be useful in various kinds of studies (see **Table 4.2**).

CPD	Peptide	M-18	M-20	Cyclic
3.14	4.4	4.14	4.15	4.16
3.15	4.4	4.17	4.18	4.19
3.1	4.20	4.21	4.22	4.23

Table 4.2. Summary of products formed during the cyclisation reactions of peptides **4.4** and **4.20** with different CPDs. Structures of the different adducts can be seen in **Scheme 4.3** for the case of **3.1 + 4.4**.

In the first place, the reaction between **4.4** and **3.14** was assayed using the conditions described in section **4.2**.

It was rapidly discovered that the cyclisation process was slower than with CPD **3.1**. While the cyclisation reaction of peptide **4.4** with CPD **3.1** was finished after 2 h, at this time point the **4.4 + 3.14** crude showed only a 76:24 ratio of the cyclic product (**4.16**) to the **M-20** adduct (**4.15**). Analysis after 4 h revealed completion of the cyclisation reaction, but a quite large amount of cyclic disulfide (21 %) and products of C-terminal cysteine oxidation (14 %) accompanied the desired **4.16**. The reaction yield determined by HPLC/MS was 60 %, considerably lower than that obtained with the simpler CPD **3.1** in the same conditions (85 %).

More discouraging were the results obtained with **3.15**. In this case, oxidation of the **M-18** Da adduct (**4.17**) to the **M-20** Da species (**4.18**) was slower and a total of 2.5 h at 60 °C were needed to reach an acceptable **4.18/4.17** ratio (impossible to evaluate correctly due to co-elution issues). Oxidation was not the only process affected by the nature of the CPD, as HPLC/MS analyses revealed. The first 0.2 equiv. of TEMPO were totally invested in pushing the conversion of **4.17** to **4.18** to completion, and none of the desired cyclic product (**4.19**) was observed after 1 h. The next 0.4 equiv. only triggered a small change in the reaction crude that could only be evaluated qualitatively due to the co-elution of one of the **4.19** peaks with one of the **4.18** isomers. Again, only little evolution was observed with the addition of the next 0.4 equivalents, and the reaction did not reach its end until a total of 4 h after the first TEMPO addition (6 h for the overall process). At this point, however, the yield of cyclic peptide **4.19** was under 50 % and formation of the thiol oxidation by-product was alarmingly high, reaching a 0.8:1 ratio to **4.19**. Surprisingly, though, the extent of cyclic disulfide formation was, more or less, the same as in the previous reaction. As for other impurities, it was also observed that reactions with **3.1** and **3.14** were cleaner than with **3.15**.

These results clearly indicated that the method established before was not good enough when dealing with CPDs other than **3.1**, and that the CPD substituents clearly influenced the outcome of the cyclisation reaction.

When moving to work with other peptides, other unsatisfactory results were obtained. The oxytocin (**4.20**, H-Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH₂) sequence was also employed due to its biological relevance in human beings.^{5,6,7} Its reaction with **3.1** was studied using the essentially same protocol but changing the amount of **3.1** from 1.2 to 1.6 equivalents to facilitate the oxidation of the **M**-18 Da adduct (**4.21**) to the **M**-20 Da adduct (**4.22**), which, in previous experiments (data not shown), was found to be slow. In this experiment, it was observed by HPLC/MS that the formation of the final cyclic species (**4.23**) was laboured and, as previously seen in slow reactions, yielded a large amount of side-products. The conversion of **4.22** towards **4.23** had not evolved to completion even after 3 h, and at that time point oxidation of the internal cysteine and other side-reactions had taken place to an absolutely unacceptable extent (determined to be almost 50 % by HPLC/MS).

Observing these results, it was clear that both the sequence of the peptide and the CPD substituents had an important effect on the reaction outcome. It was also evident from the HPLC/MS results that the cyclisation reaction conditions had to be optimised.

4.7 Effect of aromatic residues close to the *N*-terminal cysteine

In the previous experiments, it was observed that peptides **4.4** and oxytocin (**4.20**) were more difficult to cyclise when compared to the first peptide studied, **3.32** or its shorter version **4.10**. Although the number of peptides tested was small, it seemed that having an aromatic residue proximal to the *N*-terminal cysteine was the cause of those difficulties.

To assess whether this hypothesis was correct, we decided to compare two peptides that would form a ring of identical size and with the same amino acid composition, differing only in the position of the aromatic residue. For this purpose, peptide H-Cys-Tyr-Ser-Ala-Cys-NH₂ (**4.24**) was synthesised and its promptness to cyclise compared to that of H-Cys-Ser-Tyr-Ala-Cys-NH₂ (**4.10**).

4.7.1 Description of the experiments

A total of three experiments were performed to evaluate the cyclisation rate. In the first two, both peptides were reacted with **3.1** under the same conditions: 0.2 equiv. of

TEMPO were added to each reaction mixture at t=0, 60, 90, 120 and 150 min (indications that this methodology worked well had already been obtained, data not shown). HPLC/MS Analyses were performed after 30, 60, 120, 180 and 240 min reaction times. In the experiment involving peptide **4.24**, an additional TEMPO addition at 240 min and an HPLC/MS analysis after a total of 300 min were performed to prove, again, the need for continuous TEMPO additions. A third experiment was run to follow the progression of the early stages of the reaction. It was performed as the other two, but it was only carried out over 120 min.

The structure of each of the products involved in the cyclisation reaction can be seen in **Table 4.3**.

A quantitative analysis of all the results obtained in these experiments turned out to be impossible due to the co-elution of one of the **4.11** isomers and one of the **4.12** isomers. For this reason, no **M-18** to **M-20** Da adduct ratios can be given for **4.10**. Also, to facilitate the discussion of results and evaluate the ease to form the final cyclic product, the term precursors will be employed to encompass the **M-18**, **M-20** Da and non-oxidised cyclic intermediate adducts. These data are assembled in **Table 4.4**.

It can be observed that some reproducibility problems were encountered, in particular when comparing the data for stages where cyclisation was taking place in a considerable extent (first 60 minutes of the reaction of **4.10** and between 60 and 120 min in the case of **4.24**). Nevertheless, several conclusions can be drawn.

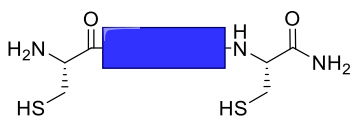
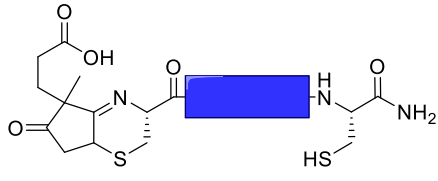
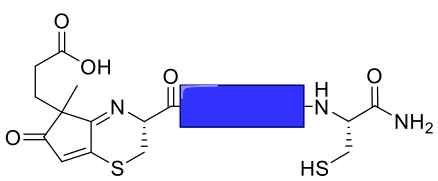
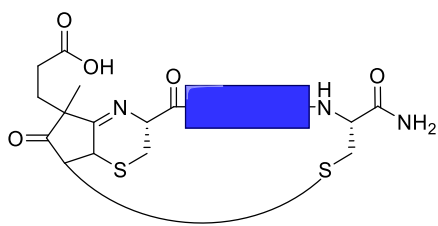
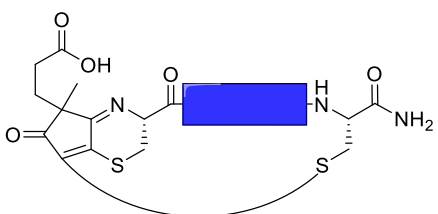
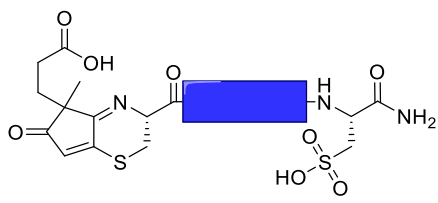
					
Sequence	Numbering	t _R (min)	Sequence	Numbering	t _R (min)
-Ser-Tyr-Ala-	4.10	-	-Ser-Tyr-Ala-	4.11	19.6-19.7 + 21.4-21.5
-Tyr-Ser-Ala-	4.24	-	-Tyr-Ser-Ala-	4.25	20.1-20.2 + 21.3-21.4
					
Sequence	Numbering	t _R (min)	Sequence	Numbering	t _R (min)
-Ser-Tyr-Ala-	4.12	19.6-19.7 + 20.4-20.5	-Ser-Tyr-Ala-	4.28	17.5-17.6
-Tyr-Ser-Ala-	4.26	19.7-19.8 + 20.7-20.8	-Tyr-Ser-Ala-	4.29	17.1-17.2
					
Sequence	Numbering	t _R (min)	Sequence	Numbering	t _R (min)
-Ser-Tyr-Ala-	4.13	15.3-15.4 + 16.3-16.4	-Ser-Tyr-Ala-	-	-
-Tyr-Ser-Ala-	4.27	16.5-16.6 + 17.9-18.0	-Tyr-Ser-Ala-	4.30	24.5 + 25.8

Table 4.3. Structures, numbering and retention times of the compounds formed over the cyclisation experiment to assess the effect of an aromatic residue next to the *N*-terminal cysteine.

Time	Product percentages							
	Reactions of 4.10			Reactions of 4.24				
	4.11+4.12	4.28	4.13	4.25+4.26	4.25/4.26	4.29	4.27	4.30
30min	79	14	7	91	38/55	5	2	n.d
	38	17	45	92	41/51	5	3	n.d
	90	7	3	91	35/56	6	3	n.d
1 h	43	21	36	91	31/60	5	4	n.d
	42	20	38	92	31/61	5	3	n.d
	56	24	20	91	28/63	6	3	n.d
2 h	19	10	71	66	0/66	7	27	n.d
	17	10	73	47	0/47	1	52	n.d
	18	12	70	57	0/57	5	38	n.d
3 h	5	3	92	22	0/22	0	68	10
	8	6	86	17	0/17	0	69	14
4 h	0	0	100	19	0/19	0	70	11
	0	0	100	15	0/15	0	70	15
5 h	n.d.	n.d.	n.d.	0	0/0	0	77	23

Table 4.4. Percentages of the different products present in the reaction mixture as a function of time, as determined by HPLC/MS at 280 nm. The three replicates of each reaction have been included on the table to show the reproducibility problems encountered. To facilitate interpretation and discussion of the results, only the peaks corresponding to the products of interest are listed.

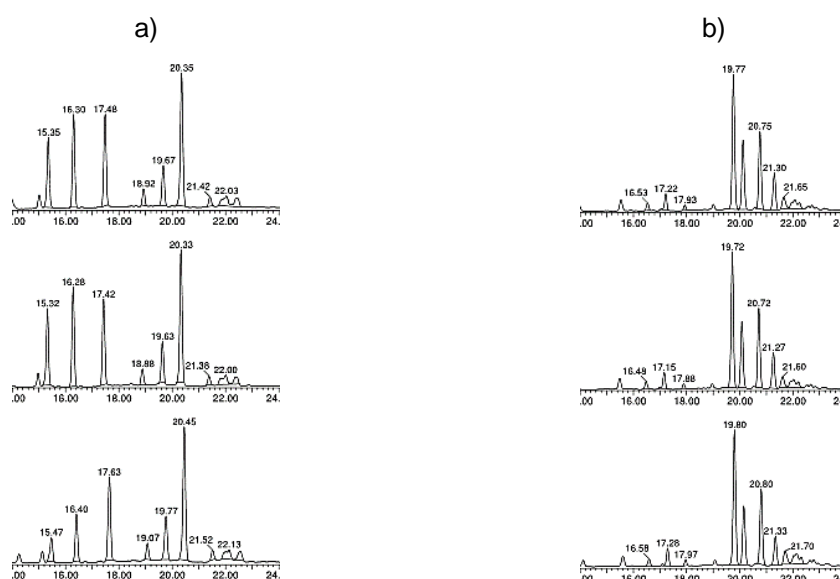
4.7.2 Detection of a previously overlooked intermediate

Curiously, in the case of **4.10**, a previously unobserved peak could be detected. This product was present in fairly high amounts (~ 15-20 %) at the first stages of the reaction, it had practically disappeared after 3 h (~ 5 %) and after 4 h it could no longer be detected. Its mass fitted with that of the **M**-20 Da adduct, although the UV-Vis profile did not possess the characteristic maximum around 330 nm. For these reasons, we surmised that **4.28** was the structure of this product, previously not perceived in cyclisation reactions. This finding made us revise the results of the reaction with peptide

4.24. We could identify a product (**4.29**) with the same properties and behaviour, previously ignored due to the low extent of its formation. Quite possibly this precursor of the final cyclic peptide was also present in other reaction mixtures but in low amounts, which made it go unnoticed.

4.7.3 Oxidation of the M-18 to M-20 Da adduct

Analysis of the reaction crudes after 60 min clearly shows that the **M-18** to **M-20** Da adduct oxidation was hindered when an aromatic residue was placed next to the *N*-terminal cysteine (see **Table 4.5**).



Product	Retention time (min)	Area percentage		
		Replicate 1	Replicate 2	Replicate 3
4.11	19.6-19.7 + 21.4-21.5	9.4	9.2	12.8
4.25	20.1-20.2 + 21.3-21.4	26.3	25.4	23.1

Table 4.5. HPLC/MS traces of the reaction between **4.10** and **3.1** (a) and between **4.24** and **3.1** (b) after 60 min. Area percentages of the **M-18** Da adducts of both peptides are also shown.

As stated above, the exact **M-18/M-20** (**4.11/4.12**) ratio cannot be determined for **4.10**. Yet, this ratio can be roughly calculated by assuming that the peak at 19.6-19.7 min, in which one isomer of **4.11** and one of **4.12** co-elute, only contains **4.11** (which in fact largely overestimates the amount of **4.11** present in the mixture). With this approximation, the percentage of **M-18** Da in the **4.10** reaction crude after 1 h varied

between 9 to 13 %. By contrast, the percentage of the **M**-18 Da adduct of **4.24**, **4.25**, is approximately 25 % after the same reaction time. This means that the **M**-18 Da adduct present in the **4.24** crude is approximately double the amount of that considered to be present in the **4.10** crude, revealing that the oxidation process is hindered when an aromatic residue is placed next to the *N*-terminal cysteine.

4.7.4 Extent of by-product formation

The detrimental effect of an aromatic residue next to the *N*-terminal cysteine also affected the reaction outcome in terms of side-reactions. The cleaner outcome of the **4.10** reaction was evident after both reactions were completely finished, which required 4 h for **4.10** and 5 h for **4.24**. **4.13** Was almost the only product present in the **4.10** crude, with only traces of minor impurities, but a considerable amount (20-25 %) of the thiol oxidation product (**4.30**) accompanied the formation of **4.27** (see **Figure 4.10**). In other words, the slower the cyclisation rate, the higher the amount of undesired sulfonate peptide (**4.30**). The correlation amount of undesired sulfonated peptide-slow cyclisation rate has also been observed with other difficult to cyclise peptide sequences (see section **4.6**).

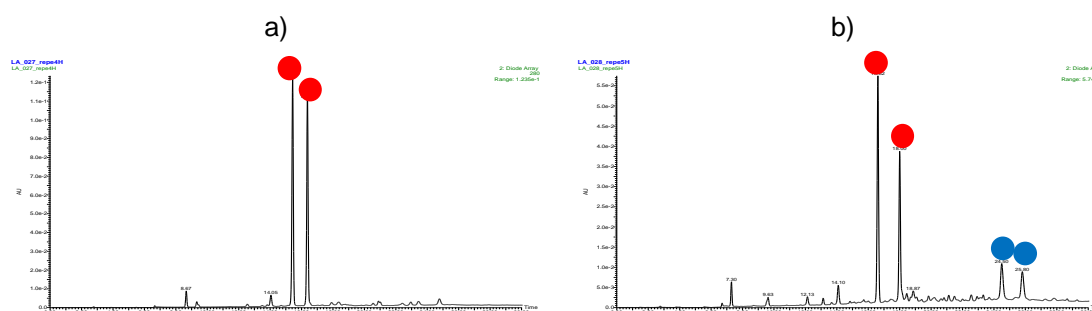


Figure 4.10. HPLC/MS traces (280 nm) of the crudes of **4.10** and **4.24** at the end of each reaction (4 and 5 h, respectively). Peaks marked with a red dot are the desired cyclic products. Peaks marked with the blue dot are the **4.30** isomers.

A considerable amount of thiol oxidation product was found at the end of the reaction of **4.24** and not in that of **4.10**. It is tempting to infer that this is due to the fact that **4.24** required more TEMPO and longer reaction times to cyclise, in other words, to the presence of larger amounts of TEMPO in the reaction mixture. Yet, it is worth mentioning that formation of **4.30** is already observed at 180 min in the **4.24** crude, whilst the corresponding by-product is not present in the **4.10** crude at the same reaction time (where both reactions had received the same amount of TEMPO).

In conclusion, the main reason accounting for the formation of thiol oxidation by-products is the difficulty to cyclise. When this is the case, the internal thiol remains in contact with TEMPO for a longer time, producing the sulfonate in high extent.

4.7.5 Conclusion

In conclusion, the overall cyclisation progress is slower when an aromatic residue is placed next to the *N*-terminal cysteine. This can be already observed after 1 h, as well as in the data of subsequent analyses. For example, after 2 h the cyclic/precursors ratio was approximately 70:30 for **4.10**, while for **4.24** remained between 52:48 and 27:73. After 3 h the crudes did not show such a striking difference in the cyclic/precursors ratio, probably due to the usual slower conversions at the final stages of a reaction, but the difference was still considerable, roughly 90:10 for **4.10** and 78:22 for **4.24**. Furthermore, the detrimental effect of an aromatic residue proximal to the *N*-terminal cysteine had been proven also in terms of by-product formation. It was found that the presence of an aromatic residue next to the *N*-terminal cysteine triggered the oxidation of the internal cysteine thiol to sulfonic acid, most probably due to the larger exposure of this thiol to TEMPO due to the slower cyclisation rate.

4.8 Effect of lithium chloride on the reaction progress

The previous experiments had showed that there was an influence of the peptide sequence on the reaction progress and that peptides with an aromatic residue next to the *N*-terminal cysteine cyclised more slowly. The reaction between the internal cysteine and the conjugated system could be hampered by conformational effects.

With the aim of assessing whether a chaotropic agent could overcome this problem and accelerate the cyclisation, the effect of lithium chloride in the reaction progress was tested. We chose oxytocin (**4.20**) as the peptide to perform these tests with because it was the most difficult to cyclise and, therefore, a clearer effect would be observed in case LiCl could enhance the cyclisation rate.[§]

3.1 And **4.20** (2:1 ratio) were reacted for 45 minutes at 60 °C, at which time the **M-18/M-20** Da adducts ratio was almost 1:1. After this, the reaction was split in four aliquots and LiCl was added to three of them to reach a final 5 molal concentration (aliquot A), a 2

[§] To accelerate the formation of the **M-20** Da (**4.22**) adduct 2 equivalents of CPD **3.1** were employed instead of the lower amounts normally used. This, of course, entailed that more double addition products were formed and that the HPLC profiles were more complicated, but as we only wanted to observe the effect of LiCl in the cyclisation process we assumed that was a price worth paying to ensure a fast transformation to **4.22**.

molal concentration (aliquot B) and a 1 molal concentration (aliquot C). No LiCl was added to the fourth aliquot (aliquot D) to serve as control. At the same time, 0.2 equiv. of TEMPO were added to each of the aliquots (also aliquot D) prior to be left shaking at 37 °C.

The effect of LiCl in facilitating the cyclisation process was evaluated from the HPLC/MS integrations of the peaks corresponding to cyclic product (**4.23**) and **4.22**. Unfortunately, although a very flat gradient was used in these analyses, one of the **4.23** isomers co-eluted with one of **4.22**'s. This entailed that the real area percentage of both products could not be determined. However, and given the comparison we needed to make, the relative product ratios described in this section were approximated to the relative integrations of the **4.23** and **4.22** peaks that did not co-elute, *in lieu* of using the integration of the two HPLC peaks of **4.23** and the two of **4.22**.

Analysis of aliquot A after 30 minutes showed a 97:3 **4.23/4.22** ratio, which indicated that conversion to the final cyclic adduct had proceeded almost quantitatively (see **Table 4.6**). Aliquots B and C were analysed after 90 minutes and showed ratios of 88:12 and 70:30, respectively. At the same reaction time aliquot D, used as a control, showed a 35:65 ratio.

Aliquot:	A ^a (5 molal LiCl)	B ^b (2 molal LiCl)	C ^b (1 molal LiCl)	D ^b (no LiCl)
4.23/4.22 ratio:	97:3	88:12	70:30	35:65

Table 4.6. Relative **4.23/4.22** ratios at different LiCl concentrations as determined by HPLC/MS (280 nm). ^aAnalysed 30 min after the addition of TEMPO. ^bAnalysed 90 minutes after the addition of TEMPO.

From these results, the positive effect of lithium chloride seemed clear. Not only comparison of aliquot D with the others revealed that cyclisation was much slower in the absence of LiCl, but also that there was a correlation between the amount of salt used and the cyclisation rate.

Our decision was, finally, to use LiCl (2 molal) as additive in every reaction to ensure a proper cyclisation rate. It was clear that higher concentrations of LiCl accelerated the reaction more, but we were afraid that they might pose a problem during the HPLC purification.

4.9 Optimisation of the cyclisation procedure

At this point, we knew that both TEMPO (or O₂) and LiCl accelerated the cyclisation reaction.

From a practical point of view, a one-pot protocol in which TEMPO would be added at the same time as the CPD seemed more convenient for various reasons. In addition to accelerating the reaction since the very beginning, the addition of TEMPO would be always made at the exact same reaction point, entailing that more reproducible results should be obtained.

CPDs would be added in slight excess to ensure complete conversion of the starting peptide and minimise formation of the cyclic disulfide. LiCl concentration would be 2 molal, and peptide concentration 0.5 mM as in most of the already described experiments. Regarding the TEMPO additions, 0.2 equivalents would be added at the beginning of the reaction (after peptide and CPD were mixed) and no further additions would be made within the next hour. From that point, 0.2 equiv. would be added every 30 minutes until complete conversion to the cyclic product.

A first round of experiments was carried out to evaluate the optimal amount of CPD to be added. Peptide **4.4** and CPD **3.1** were again used as model compounds, and their reaction carried out at 37 °C, following the guidelines given above. Amounts between 1.25 and 1.75 equiv. of **3.1** were tested. Results depicted in **Table 4.7** describe the composition of the crude after the reaction went to completion (as judged by HPLC/MS at 280 nm), which in all cases was after 1 or 2 h.

	Cyclic disulfide (%)	Unknown by-products (%)	Cyclic adduct (%)	Thiol oxidation by-product (%)
1.25 equiv.	11	13	69	5
1.50 equiv.	7	12	74	6
1.50 equiv.*	14	6	75	5
1.75 equiv.	6	13	75	6

Table 4.7. Summary of the outcome of the reactions between **3.1** and **4.4** at 37 °C using different amounts of the CPD. Product ratios were obtained by integration of the corresponding peaks at 280 nm. * Reaction carried out at 60 °C.

As can be seen, when **3.1** was added below 1.50 equivalents, formation of the cyclic disulfide product was **4.9** above 10 %, which entailed lower **4.7** final yields. Increasing

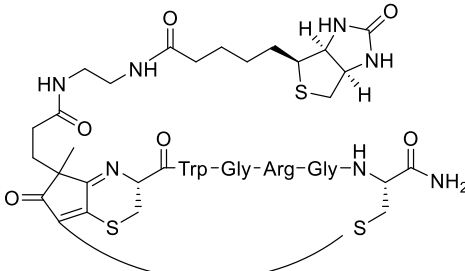
the amount of **3.1** above 1.50 equivalents did not induce major changes in the reaction crude as compared to when this exact amount was added, and therefore 1.50 equiv. of CPD were deemed the best option.

However, the reaction was, from our point of view, still generating too many by-products. We envisaged that a major cyclisation rate would help avoid this, and therefore the reaction with 1.5 equiv. of **3.1** was repeated at 60 °C. This time the amount of unknown by-products was lower than in previous experiments (6 %) and, although the amount of cyclic disulfide was higher (14 %), the yield of the final product was 75 %, the same as when the reaction was performed at 37 °C. Although increasing the temperature at this stage did not seem beneficial, it was not worse in terms of final **4.7** yield. Since higher temperatures favour the formation of the **M-20 Da** adduct (see section **3.5**) and substituted CPDs (**3.14** and **3.15**) hinder the cyclisation reaction, we considered that it would be beneficial to work at 60 °C. The protocol for the cyclisation + derivatisation (in case the CPD is derivatised) reaction was established as follows: the peptide, at a 0.5 mM concentration, would be incubated with 1.5 equiv. of the CPD and 0.2 equivalents of TEMPO at 60 °C for 1 h in the presence of 2 molal LiCl. After this time, 0.2 equiv. of TEMPO would be added every 30 minutes until completion of the reaction. The solvent of the reaction would need to be optimised for each CPD on the basis of its solubility.

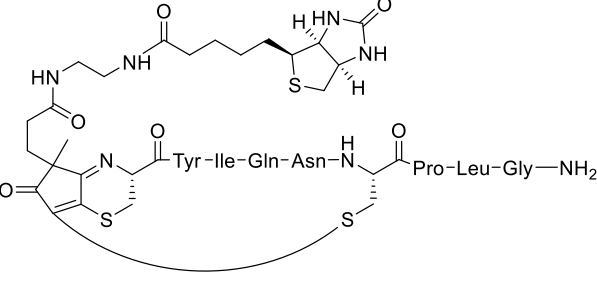
4.10 Use of labelling CPDs in cyclisation reactions - 2. Epimerisation of the *N*-terminal cysteine

At this stage it was time to assess whether this protocol was suitable to cyclise any peptide making use of differently derivatised CPDs. For this reason, the peptides previously found to cyclise with difficulty (**4.4** and **4.20**) were reacted with **3.14** and **3.15** (100 nmol scale) in two different solvent systems. The results of these experiments, in the form of HPLC/MS-based yields, and the structures of the products can be seen in **Table 4.8**, while the chromatograms for the worst- and best-performing cases are shown in **Figure 4.11**.

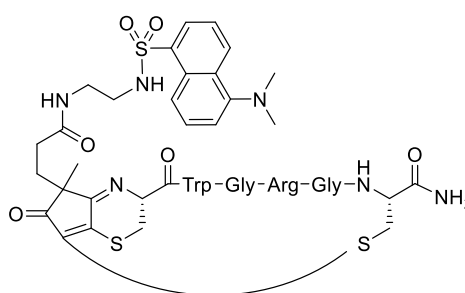
	4.4		4.20	
	H ₂ O	1:1 H ₂ O/MeOH	H ₂ O	1:1 H ₂ O/MeOH
3.14	75 %	68 %	52 %	76 %
3.15	n.d	73 %	n.d.	73 %



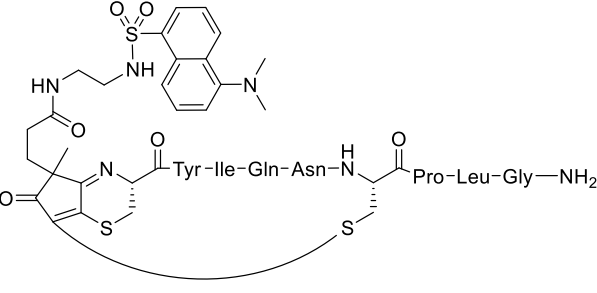
4.16



4.31



4.19



4.32

Table 4.8. Yield of cyclic adduct as determined by HPLC/MS (280 nm). The structure and numbering of the different products is also indicated.

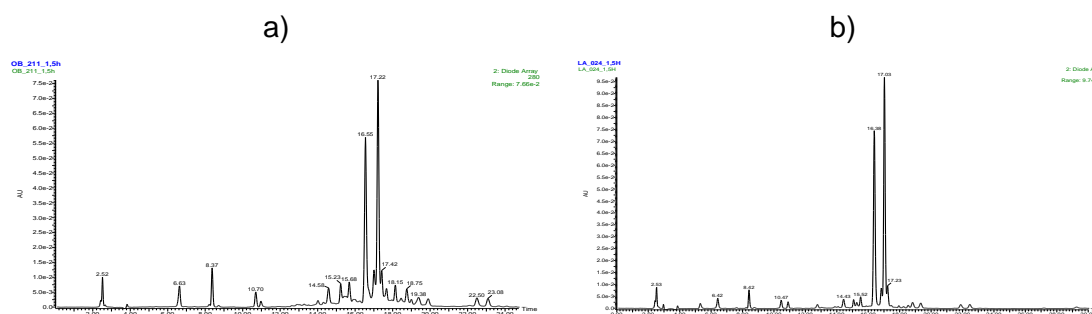


Figure 4.11. HPLC/MS traces (280 nm) of the worst-performing (**4.20 + 3.14** in water, a) and the best-performing (**4.20 + 3.14** in H₂O/MeOH, b) reactions.

Regarding the solvent composition, it was soon observed that it was not possible to work with **3.15** in pure water. Also, in some cases the reaction with **3.14** worked better in a 1:1 water/MeOH mixture than in pure water, although that was not always the case.

Yields above 60 % were always obtained, and the major side-reactions, as previously observed, were the formation of the cyclic disulfide product and what we believed was the product resulting from epimerisation of the *N*-terminal L-cysteine to its D- counterpart. Cyclic disulfide formation was particularly important in the case of peptide **4.4**, while in the case of **4.20** it was less pronounced. The hypothesized epimerisation of the *N*-terminal cysteine, on the other hand, was basically observed for **4.20** and not for **4.4**. Other unidentified products, all in small amounts, accompanied the formation of the mentioned species.

After confirming that the reactions worked well, they were repeated at a 500 nm scale to isolate the final products. Scaling-up the process implied the need to prolong the reaction times in almost all the cases, entailing a total of 2 h except in the case of the **4.4 + 3.14** reaction, which was complete after 1 h. Again, good HPLC/MS-based yields were obtained in these reactions (see **Table 4.9**). Curiously, reactions with **4.20** were more affected by the scaling-up process than those of **4.4**, and, as can be seen, a better yield at the 500 nmol scale was obtained with the dansyl-derivatised CPD (**3.15**) while that of the biotin-derivatised CPD (**3.14**) was lower in comparison to the reaction at the 100 nmol scale.

Reagents	Solvent	Expected product	HPLC/MS yield (280 nm)
4.4 + 3.14	H ₂ O	4.16	76 %
4.4 + 3.15	1:1 H ₂ O/MeOH	4.19	75 %
4.20 + 3.14	1:1 H ₂ O/MeOH	4.31	65 %
4.20 + 3.15	1:1 H ₂ O/MeOH	4.32	84 %

Table 4.9. HPLC/MS yields (at 280 nm) of the reactions of **4.4** and **4.20** with **3.14** and **3.15** at the 500 nmol scale.

All products were purified by HPLC and the two cyclic isomers could be separated with no difficulties. All products had to be dissolved in methanol to proceed to their analysis and quantification because they were not soluble enough in pure water or water containing small amounts of organic solvents (albeit they were soluble in the purification medium after eluting from the HPLC). This is most probably due to the solubility of the CPDs, especially in the case of **3.15**. However, it must also be taken into account that

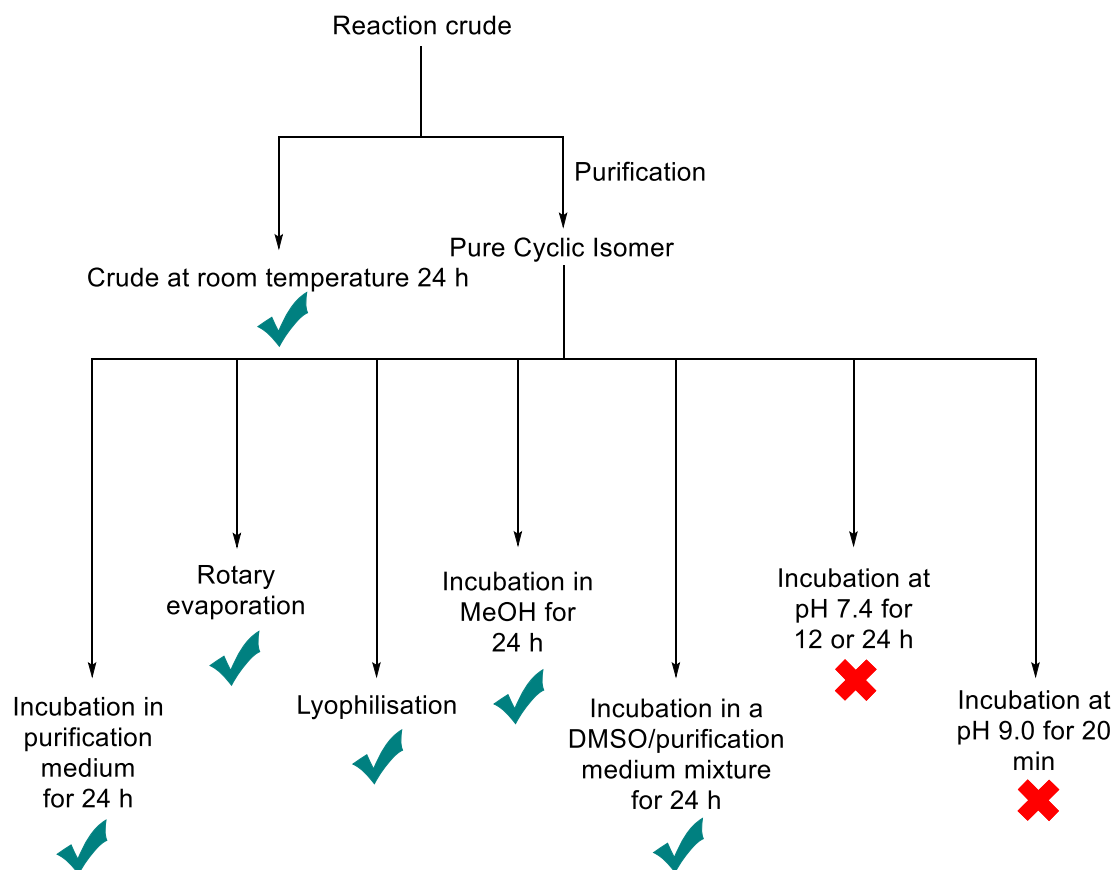
upon the cyclisation + derivatisation steps the positively charged *N*-terminal amine is converted to an imine, and that two thiol groups give rise to two thioethers, losing 3 H-bond donors and one charge in the overall cyclisation process. This, most probably, also has a negative effect on the final products solubility in water.

Unfortunately, analysis of the products obtained in these reactions revealed that they all had decomposed. The analysed products contained either a mixture of the desired product and a by-product with the same mass and absorption profile or only the latter. This by-product was theorised to arise from the isomerization of the α carbon of the *N*-terminal cysteine.

On the positive side, though, we realised that the supposed “D version” of the cyclic adduct (for convenience this nomenclature is adopted to refer to an adduct containing all L-amino acids except for the *N*-terminal D-cysteine) was formed, at least in some cases, in an extent above 50 %. That seemed to point towards the D isomer at the *N*-terminal position being more reluctant to epimerisation.

4.11 Confirmation of *N*-terminal cysteine epimerisation

The first thing we did after discovering that the cyclic products were not stable was to try to discover why their decay had taken place and if that had happened with previously obtained products. The two isomers of **4.13**, which had been prepared and purified two months before, were analysed by HPLC/MS again. They both showed a side-product with the same mass and UV-Vis profile as the target cyclic peptide, indicating that **4.13** had also undergone the supposed epimerisation at the *N*-terminal cysteine position. To discover at which step this process had taken place, product **4.16** was again synthesised (this time at the 1 μ mol scale) and the two expected isomers were independently isolated. Several tests were performed to each cyclic diastereomer to identify when the hypothesised epimerisation had taken place (see schematic representation in **Scheme 4.4**).



Scheme 4.4. Summary of the procedures carried out to assess when the presumed epimerisation had taken place. A green tick represents that only the pure, desired cyclic isomer was found while a red cross indicates the theorised epimerisation.

First, to confirm that the reaction media did not trigger the decay of the product, a small aliquot of the reaction crude was taken before the purification step and left stand for 24 h at room temperature. HPLC/MS Analysis after this time did not reveal any substantial change in the crude, indicating that the reaction conditions did not induce the changes observed. The second test performed was aimed to assess whether the purification medium could trigger the formation of side-products. For this reason, after collecting each isomer from the HPLC in the purification process, an aliquot was stored at room temperature and analysed 24 h later. Only the pure, expected peaks could be detected in this case, ensuring that the purification media was not causing the products' decay. The next step performed after purification of the products was removing the solvent. To confirm that this normally considered benign process was not involved in the observed decay, each isomer was analysed after removing the purification media both by rotary

evaporation and by lyophilisation.** Again, only the desired products were found after HPLC/MS analyses.

Given that none of the steps involved in the reaction or purification process had been able to induce the hypothesised epimerisation, we continued testing other possibilities. As mentioned above, all products synthesised in section 4.10 were found to be insoluble in water, which implied that some unsuccessful attempts to dissolve them in a DMSO/water mixture were performed before using methanol. To discard any possible influence of these organic solvents in the observed side reaction, each isomer was both incubated in methanol (after lyophilisation) and left standing in a 1:1 DMSO/purification medium mixture for 24 h. Analysis of these solutions allowed us to reject any effect of these solvents, as only the expected pure isomers were observed.

Finally, we decided to check if a basic pH would harm 4.16. For this purpose, and due to the solubility issues found with these products, each isomer, directly after eluting from the HPLC, was brought to pH 7.4 or 9.0 using a phosphate buffer or a diluted NaOH solution, respectively. Analyses at different times (12 and 24 h for pH 7.4 and 20 min for pH 9.0) revealed that basic pHs altered the product (data shown in section 4.12).

Most probably the reason behind the isomerisation of the *N*-terminal cysteine relies on the large stabilisation of the anion generated upon abstraction of the α proton due to electronic delocalisation.

However, why could our products decomposed due to basic pH if they were never exposed to such conditions? The answer most probably is at the lyophiliser. In our group several people were working with oligonucleotides at that time, which are freeze-dried from a 0.1 M buffer of triethylammonium acetate. This liberates triethylamine during the lyophilisation, which might have reacted with our compounds. Another explanation would be that residual amounts of ammonia present in the cleavage crudes of oligonucleotides were present in the lyophiliser atmosphere, triggering the detected decomposition.

In any case, we had to confirm that we were dealing with the epimerisation of the *N*-terminal cysteine and not with other issues. For this purpose, we synthesised H-D-Cys-Trp-Gly-Arg-Gly-Cys-NH₂ (4.33). This peptide was reacted with 3.14 under the same conditions as its all-L counterpart and the peaks corresponding to the two cyclic diastereomers (4.34) independently isolated. Said compounds were found to exhibit the same retention times, mass and UV-Vis profiles as the previously observed “decomposition” products. To further confirm that they were the same exact product, the pure D isomer 1 was co-injected in the HPLC/MS with a sample of the L isomer 1

** Both isomers were redissolved in methanol and injected on the HPLC/MS immediately after, in order to avoid any detrimental effects due to this organic solvent. Posterior analyses revealed that methanol did not trigger the decay of the purified products.

incubated at pH 9.0, and the same was performed for the corresponding isomers 2 (numbers given on the basis of their elution order in the HPLC and HPLC/MS). In both cases de decomposition product was found to co-elute with the pure D isomer, definitively confirming that we had been observing epimerisation of cysteine at the *N*-terminal position (see **Figure 4.12**).

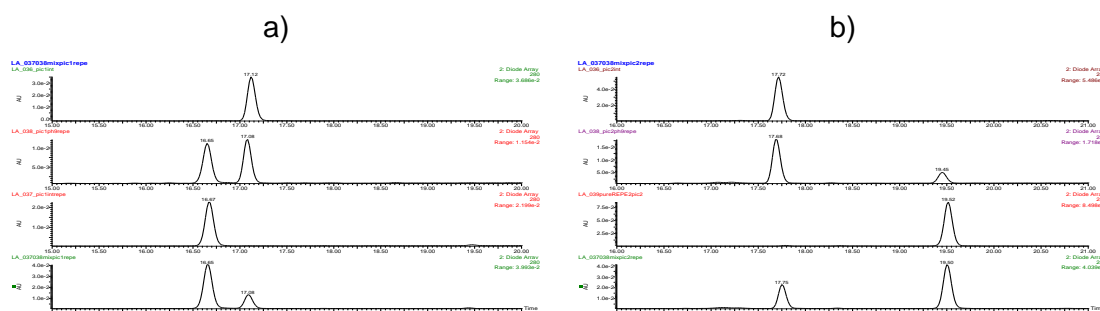


Figure 4.12. HPLC/MS traces (280 nm) demonstrating the epimerisation of isomer 1 (a) and isomer 2 (b) of **4.16**. From top to bottom: pure **4.16**, **4.16** at pH 9.0, pure **4.34**, and coinjection of the crude obtained after incubation of **4.16** at pH 9.0 and **4.34**.

4.12 Comparison of the stability of the D and L epimeric cyclic peptides

After having evidence that the “decomposition” products observed previously were caused by epimerisation of the *N*-terminal cysteine, it was time to see if replacing this amino acid by its enantiomer could help minimise the epimerisation reaction. With this idea in mind, pure **4.16** and **4.34** (the two isomers of each product), just collected from the HPLC, were brought to pH 7.4 and pH 9.0 and their integrity checked after different times (once again, each isomer of the D and L versions was treated independently). As seen in **Table 4.10**, a D amino acid at the *N*-terminal position provided the product with a marked improved stability. While the isomerisation of **4.16** after 12 h at pH 7.4 laid between 16 and 33 %, that of **4.34** reached a maximum of 8 %. The dissimilarity after 24 h was, also, remarkable.

Product	Epimerisation percentage		
	pH 7.4 12 h	pH 7.4 24 h	pH 9.0 20 min
4.16 isomer 1	33	45	64
4.16 isomer 2	16	21	35
4.34 isomer 1	6	9	5
4.34 isomer 2	8	14	5

Table 4.10. Epimerisation percentages of **4.16** and **4.34** at different times and conditions.

More pronounced was the difference when the products were incubated at pH 9.0 for 20 minutes. The stability towards epimerisation of the product containing a D-cysteine at the *N*-terminal position was found to be between 12 and 7 times higher than that of its L-counterpart. While the isomerisation degree did not exceed a 5 % for the two diastereomers of **4.34**, it reached a 64 % for isomer 1 and a 35 % for isomer 2 of **4.16**. It was also curious to see that the two diastereomers of the “L version” of the product presented very different stabilities towards epimerisation while, on the contrary, both isomers containing a D-cysteine at the *N*-terminal positions displayed very similar degrees of epimerisation in all the conditions tested.

4.13 Use of labelling CPDs in cyclisation reactions - 3

After the discovery that the non-proteinogenic version of cysteine at the *N*-terminal position was less prone to undergo isomerisation, analogs of products **4.16**, **4.19**, **4.31** and **4.32** were again synthesised placing D-cysteine at the *N*-terminal position. For this purpose, **4.33** was reacted with **3.14** and **3.15** following basically the same protocol as for the synthesis of cyclic peptides described in section **4.9**. The only difference was that, in order to reduce the amount of cyclic disulfide by-product, TEMPO was not added from the very beginning but 30 min later. Although this change did not alter the amount of cyclic disulfide generated, which was our goal, and HPLC-based yields were only slightly superior, it surprisingly eased in the purification process, allowing to attain better isolated yields (compare entries 1 and 2 with 3 and 4 in **Table 4.11**) and purer compounds. The cyclic product arising from the reaction between **4.33** and **3.14**, **4.34**, was formed in a 76 % yield as determined by HPLC/MS at 280 nm, and obtained in a final 40 % isolated yield (see **Table 4.11** entry 3). For its dansyl counterpart, **4.35**, the HPLC/MS determined

yield was 66 % and the product was obtained in 40 % yield (see **Table 4.11** entry 4). The crudes of both reactions are shown in **Figure 4.13**.

Entry	Product	HPLC/MS yield (% , 280 nm)	Isolated Yield (%)
1	4.33 + 3.14 → 4.34	73	22
2	4.33 + 3.15 → 4.35	60	12
3	4.33 + 3.14 → 4.34*	76	40
4	4.33 + 3.15 → 4.35*	66	40
5	4.36 + 3.14 → 4.37	77	33
6	4.36 + 3.15 → 4.38	82	45

Table 4.11. HPLC/MS (280 nm) and isolated yields for the different products synthesised with peptides containing an *N*-terminal D-cysteine.* In these cases TEMPO was added 30 min after the reaction had started.

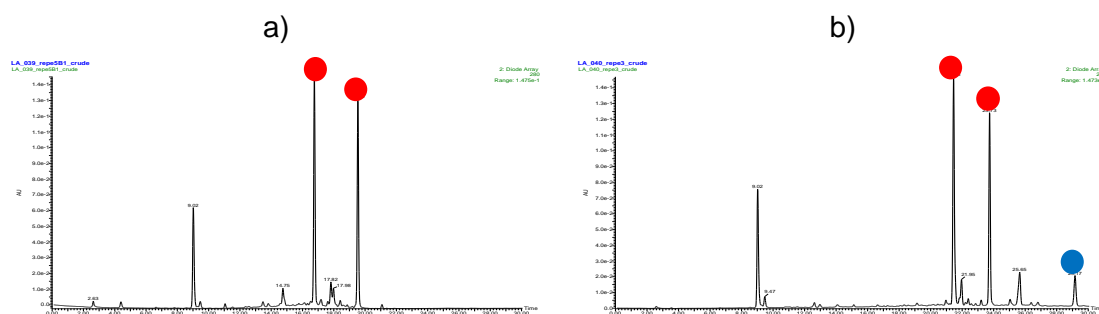


Figure 4.13. HPLC/MS traces (280 nm) of the **4.34** crude (a) and of the **4.35** crude (b) at the end of the reaction. **3.15** is marked with a blue dot while **3.14** is not observed at this wavelength. Cyclic products are the peaks marked with a red dot.

The *N*-terminal D version of oxytocin (H-D-Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH₂, **4.36**) was also synthesised and used with both **3.14** and **3.15** to furnish the corresponding cyclic products (**4.37** and **4.38**, respectively). In this case, TEMPO was added since the beginning of the reaction, as no major problems regarding the formation of cyclic disulfide were found with **4.36**. As can be seen in **Table 4.11** (entries 5 and 6), good HPLC-determined yields, both around 80 %, and good final yields, between 30 and 45 % were obtained. Clean crudes were obtained in both cases and, as determined from the HPLC/MS traces shown in **Figure 4.14**, no major side-reactions had taken place.

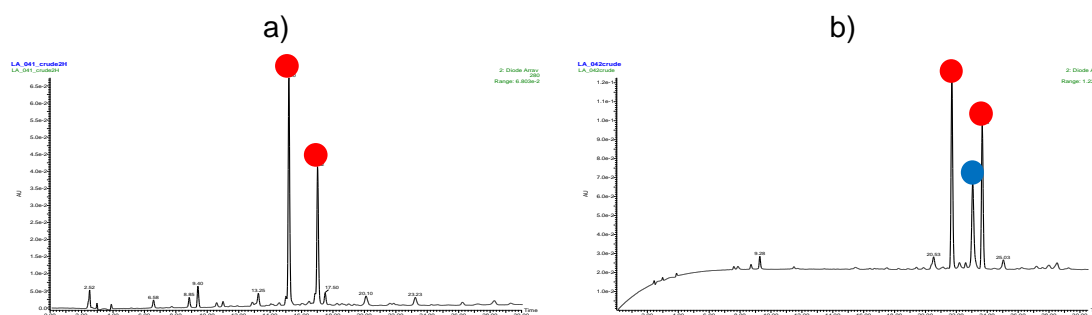


Figure 4.14. HPLC/MS traces (280 nm) of the **4.37** crude (a) and of the **4.38** crude (b) at the end of the reaction. **3.15** is marked with a blue dot while **3.14** is not observed at this wavelength. Cyclic products are marked with a red dot.

4.14 Critical overview of the methodology developed

The methodology developed in this work offers several advantages:

- 1) Peptides can be cyclised and conjugated using a single reaction.
- 2) Formation of the cyclic conjugate is fast and requires only a slight excess of CPD.
- 3) Although TEMPO must be employed as catalyst, this can be used in sub-stoichiometric amounts (normally between 0.2 and 0.6 equivalents).
- 4) The two cyclic isomers generated are perfectly separable by HPLC, allowing for the isolation of stereoisomerically pure compounds.
- 5) Limitations imposed by the peptide sequence or by the substituents in the CPD can be overcome using LiCl as additive in the reaction mixture, ensuring clean and fast reactions.

However, some negative aspects must be taken into account:

- 1) Peptides to be cyclised must possess an *N*-terminal cysteine.
- 2) The *N*-terminal L-cysteine has a high tendency to epimerise in neutral or basic pHs, and D-cysteine must be used to overcome this problem.

4.15 Abbreviations

COSY: Correlation Spectroscopy

CPD: 2,2-disubstituted cyclopent-4-ene-1,3-dione

DEPT: Distortionless Enhancement by Polarisation Transfer

DMSO: Dimethyl Sulfoxide

HMBC: Heteronuclear Multiple Bond Correlation

HPLC: High Performance Liquid Chromatography

HPLC/MS: High Performance Liquid Chromatography/ Mass Spectrometry

HPLC/HRMS: High Performance Liquid Chromatography/ High Resolution Mass Spectrometry

HSQC: Heteronuclear Single Quantum Correlation

MS: Mass Spectrometry

NMR: Nuclear Magnetic Resonance

ppm: parts per million

TEMPO: 2,2,6,6-Tetramethylpiperidine 1-oxyl radical

4.16 Bibliography

- (1) Stolz, R. M.; Northrop, B. H. *J. Org. Chem.* **2013**, *78* (16), 8105.
- (2) Northrop, B. H.; Coffey, R. N. *J. Am. Chem. Soc.* **2012**, *134* (33), 13804.
- (3) Zhang, G.; Yang, C.; Liu, E.; Li, L.; Golen, J. A.; Rheingold, A. L. *RSC Adv* **2014**, *4* (106), 61907.
- (4) Tam, J. P.; Wu, C. R.; Liu, W.; Zhang, J. W. *J. Am. Chem. Soc.* **1991**, *113* (17), 6657.
- (5) Zak, P. J.; Stanton, A. A.; Ahmadi, S. *PLOS ONE* **2007**, *2* (11), e1128.
- (6) Viero, C.; Shibuya, I.; Kitamura, N.; Verkhatsky, A.; Fujihara, H.; Katoh, A.; Ueta, Y.; Zingg, H. H.; Chvatal, A.; Sykova, E.; Dayanithi, G. *CNS Neurosci. Ther.* **2010**, *16* (5), e138.
- (7) Kemp, A. H.; Guastella, A. J. *Curr. Dir. Psychol. Sci.* **2011**, *20* (4), 222.

Conclusions

1. The experiments carried out to identify a suitable group for the protection of 1,3-dienes have been unsuccessful. A different alternative allowing diene-derivatised polyamides to be used in Diels-Alder conjugation reactions has been explored and developed. In this methodology, conjugation takes place between a fully protected, resin-linked diene-derivatised polyamide and a soluble maleimide-containing compound (3-5 equivalents). The acid-labile diene is thus not exposed to acids, and the acidic deprotection treatment that follows does not affect the cycloadduct. In contrast with previously described alternatives, this methodology has no limitations in terms of sequence and does not require special protecting groups.

2. As to the details of the on-resin conjugation process, it is worth noting that:

- i) (*E*)-4,6-Heptadienoyl-polyamide-resins gave better results than 3-(2-furyl)propanoyl-polyamide-resins, both in terms of the Diels-Alder reaction yield and regarding the stability of the conjugate.
- ii) Polyamides are assembled using the standard, commercially available building blocks on water-swelling solid matrixes, and no special building blocks need to be synthesised.
- iii) Both water and organic solvent/water mixtures can be used for the on-resin cycloaddition
- iv) Sparingly soluble maleimide derivatives can be used, even in suspension.

v) Simultaneous maleimide deprotection and on-resin Diels-Alder cycloaddition is feasible. This allows protected maleimides to be used and conjugates with different linking sites to be prepared.

3. Double conjugates can be prepared from diene-derivatised, cysteine-containing peptide-resins. The on-resin Diels-Alder reaction allows for the first derivatisation, and after the deprotection treatment a Michael-type reaction with a maleimide provides the second derivatisation.

4. Resin-linked, diene-derivatised oligonucleotides do not withstand the reaction conditions required for the solid-phase Diels-Alder conjugation. Although preliminary studies carried out with small poly-dT sequences gave good results, the methodology failed to provide conjugates of longer, mixed sequences. This indicates that poly-dT sequences are not necessarily good models of oligonucleotides.

5. 2,2-Disubstituted cyclopent-4-ene-1,3-diones (CPDs), which were chosen as non-hydrolysable maleimide analogs, have been found to possess an unexpected reactivity. While maleimides react in an irreversible manner with all types of thiols, the Michael-type reaction between CPDs and cysteines placed at internal or C-terminal positions, which do not possess a free amine, is reversible. This result is supported by calculations performed by Lluís Raich and Dr. Carme Rovira. In contrast, cysteines with a free amine (at the N-terminus of peptides) react with CPDs to end up furnishing a stable product with a mass 20 Da lower (**M-20 Da**) than the Michael-type adduct (weighing **M Da**).

6. Formation of the **M-20 Da** adduct plausibly takes place through the following steps: first, conjugate addition of the N-terminal cysteine thiol to the CPD double bond yields a Michael-type adduct. This product immediately undergoes intramolecular imine formation, by reaction between one of the CPD keto groups and the N-terminal amine, giving an intermediate with two fused rings and a mass 18 Da lower (**M-18 Da**) than the initially formed Michael-type adduct. Subsequent oxidation of this **M-18 Da** adduct provides the final, stable **M-20 Da** adduct. This is the rate-limiting step, and can be accelerated by increasing temperature or adding an excess of CPD. NMR Has confirmed the structure of the **M-20** conjugated system, which absorbs around 330 nm.

7. The method employed to monitor the reactions between peptides containing N-terminal cysteines and CPDs is critical to correctly interpret the results of these

transformations. HPLC Coupled to an ESI MS detector is the method of choice to follow their progress. The protocol using HPLC and MALDI-TOF mass spectrometric analysis of the collected peaks can be misleading, because oxidation of the **M**-18 Da adduct to **M**-20 Da one takes place either during the ionization conditions or during the product isolation process, or both.

8. Both methyl cysteinate and cysteine (1,2-aminothiols) react with CPDs to yield the **M**-20 Da adduct. In the case of cysteine, and probably due to the presence of the free carboxylic acid, this adduct undergoes imine hydrolysis. The reaction with homocysteine, a 1,3-aminothiol, only furnishes the reversible Michael-type adduct.

9. CPD-derivatised peptides can be prepared making use of standard solid-phase methodologies, and used to synthesise conjugates by reaction with cysteine-derivatised PNAs. The resulting conjugates remain stable under acidic conditions, but increasing the pH has a detrimental effect on their integrity.

10. The different reactivity of CPDs towards 1,2-aminothiols and other thiols has been exploited: i) to selectively tag a peptide containing an *N*-terminal cysteine in the presence of peptides with cysteines in other positions, and ii) to double-derivatise a peptide that contains an *N*-terminal and an internal cysteine.

11. Peptides containing an *N*-terminal and an internal (or *C*-terminal) cysteine can be cyclised by reaction with CPDs. In case the latter incorporate a tagging unit (biotin, fluorophore) both cyclisation and labelling are simultaneously accomplished. Formation of the **M**-20 Da adduct (which takes place first) is followed by Michael-type addition of the internal thiol to this adduct, and finally by oxidation to yield a conjugated system absorbing around 370 nm. Cyclisation has been assessed by NMR experiments.

12. The cyclisation reaction is accelerated by the presence of oxidants, namely O₂ or TEMPO. A continuous addition of TEMPO in sub-stoichiometric amounts is the best method to obtain the derivatised cyclic peptides in good yields. Addition of LiCl has also a beneficial effect, in particular with difficult to cyclise sequences. In this respect, it has been found that aromatic residues proximal to the *N*-terminal cysteine hinder the cyclisation.

13. The *N*-terminal cysteine of CPD-cyclised peptides undergoes epimerisation at the α carbon when exposed to bases, as confirmed by synthesis of the cyclic analogs from

peptides with a D-cysteine at the *N*-terminus. Curiously, the latter are less prone to epimerise than the L-cysteine counterparts, for which reason use of the non-proteinogenic residue is recommended.

Appendix 1. Attempts to identify a diene protecting group

A.1.1 Attempts performed

As stated in section 2.1.4 of this dissertation, the experiments carried out to find an appropriate diene protecting group are summarised in this appendix. The different dienes and dienophiles tested here were chosen with the aim to obtain Diels-Alder (DA) cycloadducts that could be easily reversed and that were stable to the typical polyamide cleavage and deprotection conditions (a 2 h treatment with 95:2.5:2.5 TFA/H₂O/TIS, see section 2.1.4). Those dienes (like anthracene) that were really hydrophobic or bulky were immediately discarded to avoid problems when performing the conjugation in aqueous media or the formation of bulky cycloadducts. However, due to the typical lack of information on the DA adducts reversibility and the total absence of knowledge about their stability under the deprotection conditions, most of the times the dienophiles were selected from those most commonly used.

As a starting point we decided to work with (*E*)-*N*-(2-phenoxyethyl)hepta-4,6-dienamide (**A.1**). The phenoxyethyl moiety was linked as an amide to the diene in order to introduce a chromophore into the molecule and allow for its easy UV monitoring during all the experiments. Our first attempt to develop a protecting group for the diene relied on the use of 4-phenyl-1,2,4-triazole-3,5-dione (PTAD), which was described to be very reactive

towards dienes^{1,2} and to render cycloadducts with anthracene that can be reversed under mild conditions.²

It is known that 1,2,3,6-tetrahydropyridazines can be air-oxidised to render an unstable azocompound that extrudes nitrogen to generate a 1,3-diene (see **Figure A.1.1**),³ and that these 1,2,3,6-tetrahydropyridazines can be obtained after basic hydrolysis of the DA adduct generated by reaction with PTAD. Therefore, PTAD furnishes [4+2] cycloadducts that can, in principle, be reversed by basic hydrolysis and air oxidation. Another possible way to remove the PTAD protecting group is with the use of DIBAL-H,⁴ although this method seemed, at least at first sight, incompatible with peptide solubilisation, which normally requires water, and most of peptide functionalities.

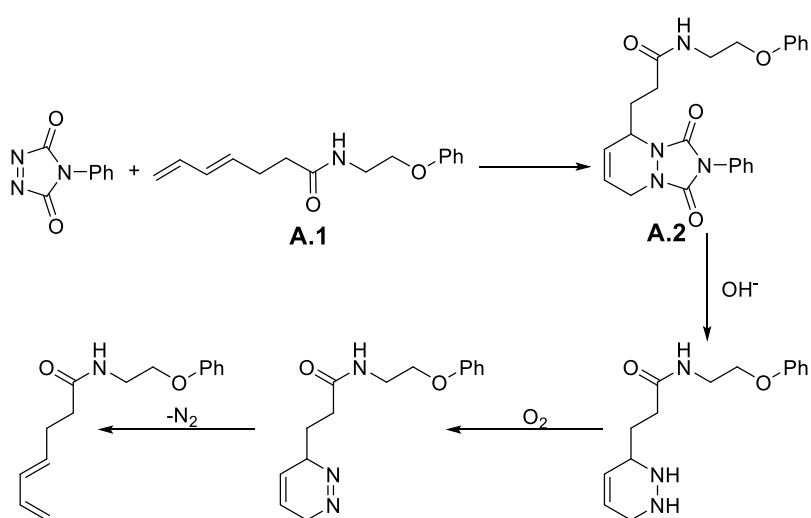


Figure A.1.1. Full protection/deprotection sequence for diene **A.1** using PTAD .

With this information in hand we synthesised the DA adduct of **A.1** + PTAD (**A.2**). Its stability towards a typical cleavage cocktail was verified, but there was no way we could satisfactorily perform the rDA reaction. All attempts to perform the rDA reaction thermally failed to render the desired **A.1**, as it also happened with the assays aimed to reverse the cycloadducts through hydrolysis of **A.2**. We even tried the use of DIBAL-H in different amounts, but only unsatisfactory results were observed.

In some of the cases the cycloadduct did not react at all, whilst in the others it decomposed to give several by-products. In no case was **A.1** generated. The rDA was also carried out in the presence of a diene scavenger (*N*-methylmaleimide, NMMal), but these conditions also failed to induce the reversion of the cycloadduct.

In view of this, we decided to evaluate other dienes. Both naftoquinone and benzoquinone were used because of their high reactivity towards dienes and their availability in our lab. The DA adduct of naftoquinone with **A.1** (**A.3**, see **Figure A.1.2a**)

was easily synthesised by thermal treatment of the diene/dienophile mixture, but its unstability to the typical cleavage and deprotection mixture was soon discovered. However, it withstood a 10 minute treatment, which we believed would be enough to cleave and deprotect peptides using the Sieber linker⁵ and protecting groups removable under very mild conditions.⁶ For this reason, we subsequently moved on and tried to reverse the DA adduct. Several attempts were performed using organic solvents and water/organic solvents mixtures (pure water was deemed not worth trying due to the product solubility), but they only met with failure. Even using cysteamine as scavenger did not produce any valuable result.

When benzoquinone and **A.1** were reacted to furnish the expected DA adduct (**A.4**) the result was disappointing. The desired product was not isolated but instead we obtained an oxidation by-product. NMR Analyses (¹H, ¹³C, HSQC and COSY) revealed that the newly generated six-membered ring had underwent oxidation and furnished an aromatic structure (see **Figure A.1.2b**). Most probably this was due to the ability of benzoquinone, which had been added in excess in the reaction, to act as an oxidising agent. For this reason, attempts to obtain **A.4** using lower amounts of benzoquinone were made, although they all ended in failure.

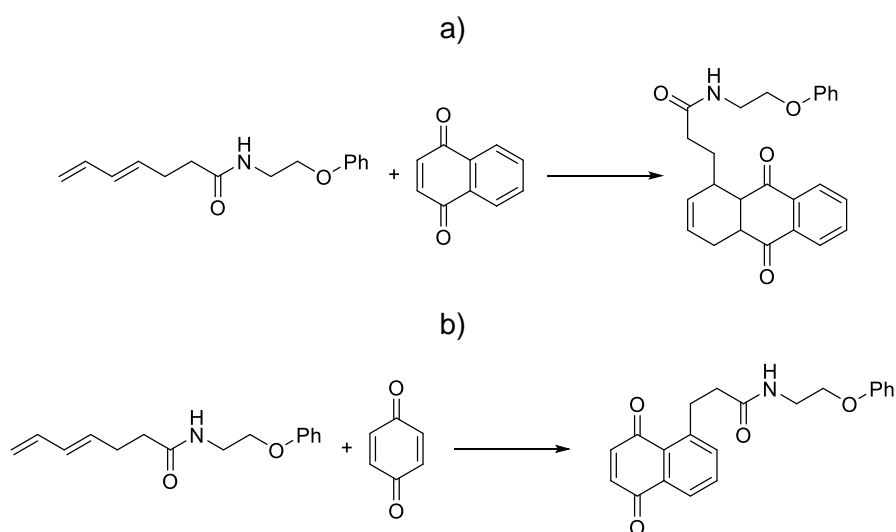


Figure A.1.2. Synthesis of adduct **A.3** (a) and structure of the product obtained upon reaction of **A.1** and excess benzoquinone (b).

After these observations, we decided to work with pentamethylcyclopentadiene (**A.5**). It had been described that substitution of the cyclopentadiene skeleton results in DA cycloadducts that can be more easily reversed. Grieco and Abood described long ago that cycloadducts of **A.5** could be reversed in organic solvents using Lewis acids under

mild conditions.⁷ Similar results had been found by Revés *et al.* when working with tetramethylcyclopentadiene, whose DA adducts with α,β -unsaturated ketones could be reversed, in the presence of Lewis acids and a scavenger, in dichloromethane at 0°C in less than 1 hour.⁸ We reasoned that the positive effect of a Lewis acid in an apolar solvent such as dichloromethane could be maintained by a changing the solvent to water, which, as happens with the DA reaction, accelerates the rDA transformation.⁹

On this basis, we synthesised a small library of DA adducts with standard dienophiles to test their stability to the acidic cleavage and deprotection mixture and their rDA reaction. As dienophiles we used benzoquinone, NMMal, naftoquinone, cyclopent-2-en-1-one, divinylsulfone and PTAD. All DA adducts (see structures and numbering of the products in **Figure A.1.3**) were obtained by simple thermal treatment, except that of cyclopent-2-en-1-one, which needed iminium catalysis to render the desired cycloadduct.¹⁰

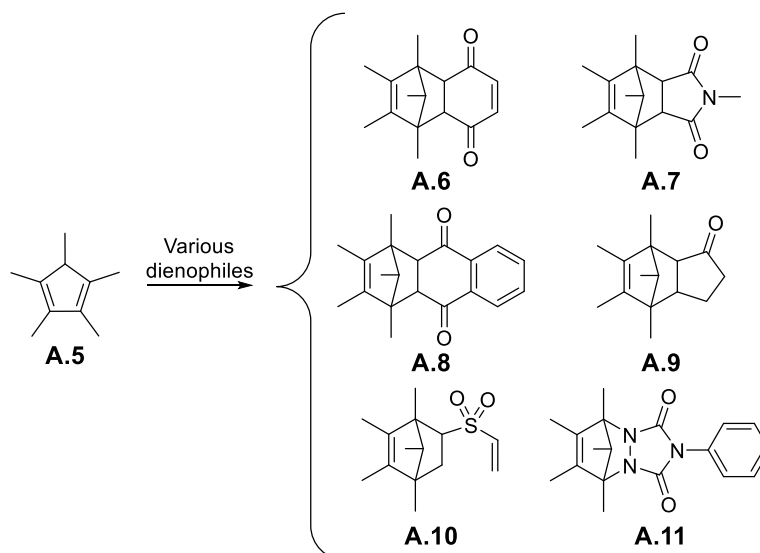


Figure A.1.3. Structures and numbering of the products synthesised with diene **A.5**.

Of all the products synthesised, only **A.7** and **A.10** showed stability to the cleavage and deprotection mixture, and therefore they were the only ones tested for the rDA reaction (the others did not withstand the cleavage cocktail even for 10 minutes).

Unfortunately, the first experiments performed aroused suspicions concerning the thermal stability of **A.5**. This fact had not been assessed in the papers by Grieco and Revés because in both papers the diene (penta- or tetramethylcyclopentadiene) was used as a protecting group, and therefore its fate after the rDA was not relevant. However, our situation was exactly the opposite: we were interested not in the dienophile but in the diene. Therefore, we performed several tests to check the thermal stability of **A.5**. It was found (by NMR) that **A.5** had decomposed after 2 h in toluene at reflux, 90

°C and 75 °C, and even some decay was found after an overnight incubation in this solvent at room temperature. Therefore, use of pentamethylcyclopentadiene was ruled out.

After the unsuccessful results obtained with dienes **A.1** and **A.5**, we believed it was time to slightly change the strategy. As finding a cycloadduct stable to the acidic cleavage conditions of polyamides that could undergo a rDA had been impossible, we decided to start working with a diene/dienophile combination whose stability and reversibility had already been assessed.

The 2,5-dimethylfuran/maleimide couple had previously been used in the group in several projects that involved sequential DA/rDA reactions, and therefore the reversibility of the adducts generated upon their reaction as well their stability to the cleavage cocktail employed for polyamide deprotection was well-known.^{11,12,13}

For these reasons, we believed that an appropriate furan, suitable to be introduced into polyamide chains, could be used as diene and a maleimide employed for its temporary protection. A dienophile still to determine that could render a completely stable DA adduct would be used for the final conjugation reaction.

The challenge here was to find the appropriate furan, assess its reaction with maleimides and check that the resulting product would exhibit the reversibility and stability to the cleavage cocktail previously observed for 2,5-dimethylfuran cycloadducts. After this, a dienophile rendering a stable [4+2] cycloaddition product with this furan would need to be found.

3-(2-Furyl)propanoic acid (**A.12**) seemed to fit our purposes. It could be easily introduced into polyamide chains by means of standard solid-phase peptide synthesis (SPPS) methods and was commercially available. Its cycloadduct with NMMal was synthesised with no major problems, and its stability towards a typical peptide deprotection and cleavage cocktail verified. A microwave-assisted rDA reaction was assayed using a 1:1 water/MeOH mixture as solvent. 20 Minutes at 120 °C triggered a 40 % reversion, indicating a feasible rDA reaction after some optimisation (probably requiring the use of scavengers and longer reaction times).

The next step was, then, to find a dienophile that would render a stable cycloadduct upon reaction with **A.12** and that could be easily derivatised. Upon thinking of molecules that would possess a good dienophile motif, a structure such as the one depicted in **Figure A.1.4a** seemed suitable. As a result, we decided to synthesise diethyl maleate, ethyl 4-oxopent-2-enoate and hex-3-ene-2,5-dione. No special attention to the stereochemistry of the double bond was paid, as it was deemed secondary at this stage of the process. For this reason, no efforts to control the stereoselectivity in the oxidation of ethyl levulinate and ethyl 4-oxopent-2-enoate were made, and the products were used in the

E:Z ratio in which they were obtained. Diethyl maleate (**Figure A.1.4b**) was synthesised by acidic esterification of monoethyl maleate, which had been prepared by ethanolic ring-opening of maleic anhydride under mild conditions. Ethyl 4-oxopent-2-enoate (**Figure A.1.4c**) was obtained as the *E* isomer from ethyl levulinate by a bromination-elimination sequence. Ethyl levulinate, in turn, had been prepared by acidic esterification of levulinic acid.

Finally, hex-3-ene-2,5-dione was obtained in a 1:0.75 *E:Z* ratio by reaction of its reduced precursor with *m*-chloroperbenzoic acid (**Figure A.1.4d**).

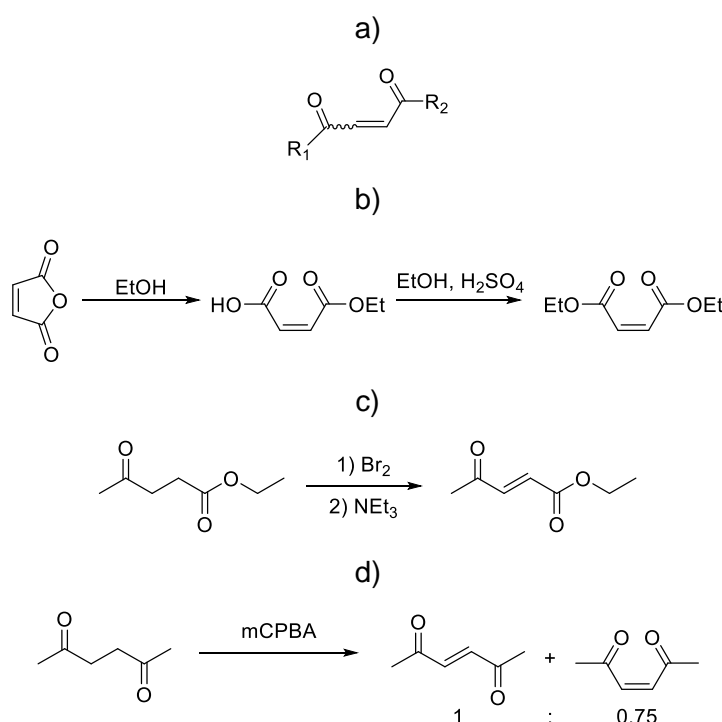


Figure A.1.4. Synthesis of the dienophiles used for the experiments with 3-(2-furyl)propanoic acid.

Diethyl maleate was found to react slowly with **A.12**, and less than a 50 % conversion was obtained after long reaction times (yielding product **A.13**), even though **A.12** was used in a five-fold excess. The two other products, surprisingly, did not undergo a DA reaction but instead a S_EAr at the 5 position of the furan ring (**Figure A.1.5**), as discovered after analysis of the 1H NMR spectra of the products.

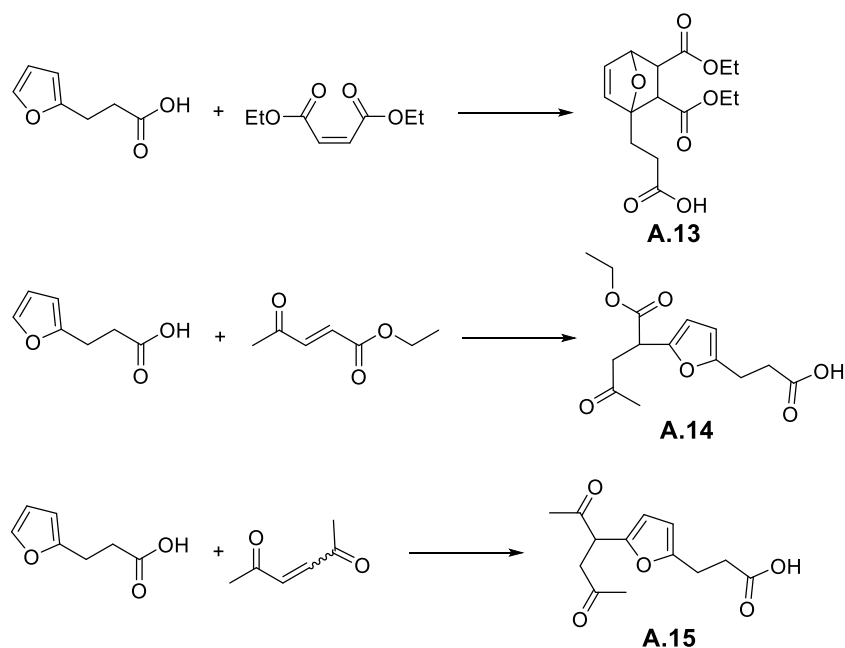


Figure A.1.5. Products generated after reaction of **A.12** with the different dienophiles. The reaction crude of **A.12** + ethyl 4-oxopent-2-enoate showed only a new product, which was inferred to be **A.14**.

A comparison between the DA adduct obtained after reaction of **A.12** and NMMal and the crude obtained after reaction of **A.12** with hex-3-ene-2,5-dione can be observed in **Figure A.1.6**.

The first evidence indicating a $S_{E}Ar$ reaction was that the products of **A.12** + ethyl 4-oxopent-2-enoate (**A.14**) and hex-3-ene-2,5-dione (**A.15**) did not present the typical signal between 5.0 and 5.5 ppm corresponding to the bridgehead hydrogen. Instead, a signal integrating one proton around 4.25 ppm was found for both products, fitting the expected chemical shift for a hydrogen placed between an aromatic and a carbonyl group. Although in the case of ethyl 4-oxopent-2-enoate its multiplicity could not be determined because it appeared together with the methylene group of the ethyl ester, the experiment with hex-3-ene-2,5-dione clearly showed a double doublet shape, fitting with the $S_{E}Ar$ product. The two doublets around 6.0 ppm did not fit neither with the multiplicity nor with the chemical shift expected for a DA cycloadduct double bond (one signal is a doublet, the other a double doublet) but they were in agreement with two proximal furan aromatic protons, supporting the hypothesis of an $S_{E}Ar$ reaction. The small signal at 4.9 ppm could indicate the minor formation of the desired cycloadduct, which would have been formed in a 1:20 ratio to the $S_{E}Ar$ product.

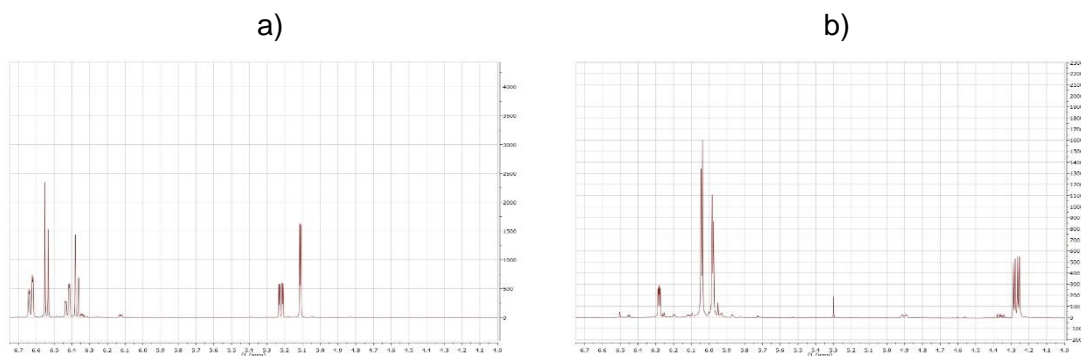


Figure A.1.6. NMR Spectra of the pure DA adduct (mixture of *exo* and *endo* isomers) of **A.12** + NMMal (a) and of the crude obtained after reaction of **A.12** with hex-3-ene-2,5-dione (b). The peak at 6.28 ppm in (b) corresponds to unreacted **A.12**, which also has a doublet at 6.05 ppm that appears overlapped with one of the product signals.

After revising the bibliography, we realised that the $S_{E}Ar$ reaction of furans with activated double bonds had previously been described by many groups^{14,15,16} and that actually, several DA adducts tended to isomerise to $S_{E}Ar$ products with time.¹⁷ This is due to the major thermodynamic stability of the $S_{E}Ar$ species, which tend to appear after decay of the DA adducts through rDA reaction.

Due to the fact that at this point we were simultaneously working with the on-resin DA reaction (described in chapter 2) and that the results with the latter seemed more promising, this research line was abandoned.

A.1.2 Abbreviations

COSY: Correlation Spectroscopy

DA: Diels-Alder

DIBAL-H: Diisobutylaluminium hydride

HSQC: Heteronuclear Single Quantum Correlation

NMMal: *N*-Methylmaleimide

NMR: Nuclear Magnetic Resonance

ppm: parts per million

PTAD: 4-Phenyl-1,2,4-triazole-3,5-dione

rDA: retro-Diels-Alder

SPPS: Solid-Phase Peptide Synthesis

TFA: Trifluoroacetic acid

TIS: Triisopropylsilane

A.1.3 Bibliography

- (1) Jensen, F.; Foote, C. S. *J. Am. Chem. Soc.* **1987**, *109* (21), 6376.
- (2) Roy, N.; Lehn, J.-M. *Chem. – Asian J.* **2011**, *6* (9), 2419.
- (3) Berson, J. A.; Olin, S. S. *J. Am. Chem. Soc.* **1969**, *91* (3), 777.
- (4) Kondo, F.; Miyashita, M.; Konno, K.; Takayama, H. *J. Chem. Soc. [Perkin 1]* **1995**, No. 21, 2679.
- (5) Sieber, P. *Tetrahedron Lett.* **1987**, *28* (19), 2107.
- (6) Ocampo, S. M.; Albericio, F.; Fernández, I.; Vilaseca, M.; Eritja, R. *Org. Lett.* **2005**, *7* (20), 4349.
- (7) Grieco, P. A.; Abood, N. *J. Chem. Soc. Chem. Commun.* **1990**, No. 5, 410.
- (8) Revés, M.; Lledó, A.; Ji, Y.; Blasi, E.; Riera, A.; Verdaguer, X. *Org. Lett.* **2012**, *14* (13), 3534.
- (9) Wijnen, J. W.; Engberts, J. B. F. N. *J. Org. Chem.* **1997**, *62* (7), 2039.
- (10) Cavill, J. L.; Peters, J.-U.; Tomkinson, N. C. O. *Chem. Commun.* **2003**, No. 6, 728.
- (11) Sánchez, A.; Pedroso, E.; Grandas, A. *Org. Lett.* **2011**, *13* (16), 4364.
- (12) Elduque, X.; Pedroso, E.; Grandas, A. *Org. Lett.* **2013**, *15* (8), 2038.
- (13) Elduque, X.; Sánchez, A.; Sharma, K.; Pedroso, E.; Grandas, A. *Bioconjug. Chem.* **2013**, *24* (5), 832.
- (14) Zuzack, J. W.; Hufker, W. J.; Donahue, H. B.; Meek, J. S.; Flynn, J. J. *J. Org. Chem.* **1972**, *37* (26), 4481.
- (15) Koerner, M.; Rickborn, B. *J. Org. Chem.* **1990**, *55* (9), 2662.
- (16) Itoh, K.; Kitoh, K.; Kishimoto, S. *Can. J. Chem.* **2006**, *84* (3), 392.
- (17) Itoh, K.; Kitoh, K.; Sera, A. *HETEROCYCLES* **1999**, *51* (2), 243.

**Appendix 2. Experiments performed by
Lewis Archibald**

Lewis Archibald worked in the cyclic project under my direct supervision. All results were inspected by both of us and decisions were made by me in collaboration with him. He personally carried out the following experiments and syntheses.

- Synthesis of peptides **4.10**, **4.20**, **4.24**, **4.33**, **4.36** and **4.39**.
- Experiments in section **4.5** dealing with the use of O₂ as oxidant.
- All experiments described in sections **4.7**, **4.8**, **4.9**, **4.11**, **4.12**, **4.13** and **4.14**.
- Reaction between **4.20** and **3.15** at the 100 nmol scale described in section **4.10**.

Experimental section

Experimental section: materials and methods

E.M.M.1 Reagents and solvents

1) Synthesis of oligomers

Fmoc-protected amino acids used for the synthesis of peptides were obtained from Novabiochem or Iris Biotech, Fmoc-protected PNA monomers from Link Technologies and commercial phosphoramidites from Glen Research. Regarding solid supports, NovaSyn TGR resin was obtained from Novabiochem, TentaGel R RAM resin from Rapp Polymere and CPG (controlled pore glass) supports from Glen Research. All solutions and solvents employed for the synthesis of oligonucleotides were purchased from Glen Research. DIPC (*N,N'*-Diisopropylcarbodiimide) and COMU (1-[(1-(cyano-2-ethoxy-2-oxoethylideneaminoxy)-dimethylamino-morpholinomethylene)] methanaminium hexafluorophosphate) employed for the synthesis of polyamides were from Sigma-Aldrich, and HOBt (1-hydroxybenzotriazole) from Iris Biotech. 3-Maleimidopropanoic acid was from Alfa Aesar. Piperidine and NMP (*N*-methylpyrrolidone) were from Sigma-Aldrich, acetic anhydride from Scharlau and DMF (*N,N*-dimethylformamide) from Carlo Erba. TFA (Trifluoroacetic acid) and TIS (triisopropylsilane) were from Sigma-Aldrich, and concentrated aqueous ammonia from Scharlau.

2) Synthesis of small organic molecules

Reagents employed for the synthesis of small organic molecules were purchased from different suppliers (Sigma-Aldrich, Fluka, Alfa Aesar, Acros, etc.) and employed without further purification. Salts and other common products usually employed in the work-up were from Jescuder or Panreac.

3) Anhydrous solvents

DCM: Distilled from CaH_2 and stored under an Ar atmosphere over CaH_2 .

DMF: Made anhydrous using 4 Å activated molecular sieves.

ACN: Bought anhydrous. Trap Pack was used always to avoid wetting.

E.M.M.2 Chromatographic techniques

1) Thin Layer Chromatography (TLC)

Thin layer chromatographic analyses were performed on aluminium foils coated with silica gel (60 F, 0.2 mm) from Merck. To visualize the spots several methods and reagents were used depending on the nature of the analytes: UV light (254 nm), potassium permanganate, iodine, etc.

2) High Performance Liquid Chromatography (HPLC)

Different instruments were used depending on the nature of the analyte. Analysis of oligonucleotides or oligonucleotide conjugates was performed in a Waters 2695 instrument equipped with a Waters 2996 Photodiode Array Detector. Analysis of peptides, PNAs and their conjugates (except those containing oligonucleotides) were performed in Shimadzu systems equipped with either a SPD-M20A diode array detector or with a SPD-20A UV/Vis detector. The same systems were used for purification. Conditions for both analysis and purification are described for each particular case.

3) High Performance Liquid Chromatography/Mass spectrometry (HPLC/MS)

HPLC/MS Analyses were performed using a Waters 2695 equipped with a Waters 2996 Photodiode Array Detector coupled to a Waters Micromass ZQ 4000 mass spectrometer. In case reaction crudes contained salts, the MS detector was disconnected from the system for the first five minutes of the analysis. Conditions used for analysis are described for each particular case.

E.M.M.3 Spectroscopic techniques

1) Mass Spectrometry (MS)

Matrix-Assisted Laser Desorption/Ionisation Time-Of-Flight (MALDI-TOF)

Analyses were carried out in a 4800 Plus MALDI TOF/TOF from Applied Biosystems, using both reflector and linear modes. All samples were prepared by mixing 1 or 2 μL of sample solution with 1 μL of matrix solution, the latter being chosen depending on the nature of the analyte. The matrix/sample mixture was prepared directly on the plate, and allowed to air-dry. Oligonucleotides were analysed using a 1:1 THAP/CA mixture, while DHB or THAP were chosen as matrices when analysing peptides or PNAs. Matrix solutions were prepared as follows:

- THAP (2,4,6-trihydroxyacetophenone): 10 mg/mL in a 1:1 $\text{H}_2\text{O}/\text{ACN}$ mixture (v/v).
- AC (ammonium citrate): 10 mg/mL in water

- DHB (2,5-dihydroxybenzoic acid): 10 mg/mL in a 1:1 H₂O/ACN (v/v) mixture containing 0.1% of TFA.

ElectroSpray Ionisation (ESI)

High resolution MS experiments were carried out by the *Unitat d'Espectrometria de Masses (CCiT, UB)* using an ESI-MS LC/MSD-TOF mass spectrometer from Agilent Technologies.

2) Nuclear Magnetic Resonance (NMR)

¹H And ¹³C spectra were recorded in a Varian-Mercury 400 or a Bruker 400 MHz Avance III spectrometer from the NMR unit of *CCiT (UB)*. ³¹P Experiments were carried out in the Varian-Mercury 400. Routine analyses were performed by dissolving approximately 5 mg of sample in 0.7 mL of the deuterated solvent of choice. Chemical shifts were referred to an internal standard (TMS) or to the solvent signal.

3) UV-Vis: quantification of peptides, PNAs, oligonucleotides and assessment of the resin substitution degree

Peptides, PNAs, oligonucleotides and their conjugates were quantified using either a Uvikon XS or a Jasco V-550 spectrophotometer. In all cases quartz cuvettes with a 10 mm pathlength were used. Molar absorption coefficients of the different oligomers were obtained from the sum of each monomer absorption coefficient. Oligonucleotides and PNAs were quantified at 260 nm and peptides at 280 nm (see below).

When necessary, the substitution degree of a peptide-resin was determined by taking a carefully weighed aliquot, removing the Fmoc group with piperidine, and quantitating the amount of 9-fluorenylmethylpiperidine formed ($\lambda_{\max} = 301 \text{ nm}$, $\epsilon = 7800 \text{ cm}^{-1} \text{ M}^{-1}$). This value was taken as the substitution degree of the resin-linked diene-polyamides in conjugation experiments, without any further correction for the addition of the diene, and used to determine the overall synthesis and purification yield for the isolated conjugates.

$$\epsilon_{oligomer} = \sum_{i=1}^{i=n} \epsilon_i$$

	Monomer	ϵ_i ($M^{-1} \text{ cm}^{-1}$)	Wavelength (nm)
PNAs (ref 1)	A	13700	260
	G	11700	
	T	8600	
	C	6600	
Oligonucleotides (ref 2)	dA	15400	260
	dG	11500	
	dT	8700	
	dC	7400	
Peptides (ref 3)	Tyr	1490	280
	Trp	5500	

E.M.M.4 Oligomer synthesis

1) Polyamide synthesis

Polyamides were manually assembled in a polyethylene syringe fitted with a polypropylene disk. The DCM used in polyamide synthesis was acid-free (passed through a basic alumina column).

Peptide Synthesis

Peptides were assembled on the NovaSyn TGR resin derivatized with a Rink amide linker, which was washed with DCM (3 x), DMF (3 x), MeOH (3 x), and DCM (3 x). This resin is supplied without an Fmoc protecting group on the amine, so no piperidine treatment is needed prior to chain elongation. Incorporation of the first amino acid onto the resin and peptide assembly were accomplished by using 3 equiv of both Fmoc-amino acid, HOBt·H₂O and DIPC, all dissolved in a minimal amount of DCM and a few drops of NMP, for 90 min at room temperature. Coupling was followed by washing with DCM, DMF, and MeOH (3 x each). In case the coupling was not complete, as assessed by the Kaiser test,⁴ it was repeated using the same conditions. The resin was always allowed to swell in DCM for 2 min before Fmoc removal. Removal of the Fmoc groups was effected by reaction with 20% piperidine/DMF (1 x 3 min + 1 x 10 min), followed by

washing with DCM (3 x), DMF (3 x), and DCM (3 x). No capping steps were performed, except after incorporation of the last amino acid and prior to incorporation of (*E*)-4,6-heptadienoic acid or 3-maleimidopropanoic acid. The capping step was effected by reaction with Ac₂O/2,6-lutidine/DMF (v/v/v 5:6:89 mixture, 2 x 5 min) and followed by washing with DMF (3 x) and DCM (3 x).

Activation of 3-maleimidopropanoic acid and of 3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanoic acid were carried out with DIPC only (90 min reaction time) to avoid any possible conjugate addition of HOBt. (*E*)-4,6-Heptadienoic acid was coupled in the same conditions as amino acid residues (DIPC + HOBt). (*E*)-4,6-Heptadienoic acid, 3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanoic acid and 3-maleimidopropanoic acid were coupled to the peptide by using 10 equiv of the appropriate reagents. A 95:2.5:2.5 (v/v/v) TFA/H₂O/TIS mixture (1-2 mL) at room temperature was used for the deprotection and cleavage from the resin of *N*-terminal maleimido-peptides, *N*-terminal CPD peptides and unmodified peptides. In case of cysteine-containing peptides, a 94:2.5:1:2.5 (v/v/v/v) TFA/H₂O/TIS/EDT mixture (EDT = 1,2-ethanedithiol) was used. Filtrate and washings (TFA, no scavengers) were collected and concentrated by blowing N₂ over the mixture, and diethyl ether was added to the resulting oil or semisolid (3-5 mL). Water was added to dissolve the peptide (2-3 mL), the biphasic mixture was shaken, and the two phases were allowed to separate. The organic phase was discarded, and new diethyl ether was added. This washing procedure was repeated 3 times. In cases where the amount of peptide cleaved allowed to do so, peptide was precipitated in cold diethyl ether and washed with the same solvent instead of being extracted with water. After this, the peptide was dissolved in water.

Whatever was the method employed, the final aqueous phase was then lyophilised and the peptide analysed and purified by HPLC. *N*-Terminal diene-peptide-resins were stored at 8 °C until conjugation was effected.

PNA Synthesis

PNA's were assembled on the TentaGel R RAM resin (derivatized with a Rink amide linker). Prior to chain elongation the resin was washed with DCM (3 x), DMF (3 x), MeOH (3 x), and DCM (3 x), treated with 20% piperidine/DMF (1 x 3 min + 1 x 15 min), and washed with DCM (3 x), DMF (3 x), and DCM (3 x). Incorporation of the PNA monomers was effected by reaction with 4 equiv of both the Fmoc monomer and COMU and 8 equiv of DIPEA (*N,N*-diisopropylethylamine), all dissolved in the minimal amount of NMP. This mixture was preactivated for 1 min before being poured onto the resin and allowed to react for 90 min at room temperature. Coupling was followed by washing with DCM, DMF, and MeOH (3 x each). In case the coupling was not complete, as assessed by the

Kaiser test,⁴ it was repeated as before. The resin was always allowed to swell in DCM for 2 min before Fmoc removal. Removal of the Fmoc groups was performed by reaction with 20% piperidine/DMF (1 × 3 min + 1 × 10 min), followed by washing with DCM (3 ×), DMF (3 ×), and DCM (3 ×). A capping step was carried out after incorporation of each monomer, as described above, to ensure that no trace of free amines was left.

Activation of 3-maleimidopropanoic acid, and (E)-4,6-heptadienoic acid was effected as for the synthesis of peptides. Conditions for storage of the diene-PNA-resins and cleavage of the maleimido-PNAs were also the same as for the peptide-resins.

2) Oligonucleotide synthesis

Oligonucleotides were synthesised in an automatic 3400 Applied Biosystems synthesiser using the standard phosphoramidite methodology and at the 1 μmol scale. Standard β-cyanoethyl phosphoramidites bearing ultra-mild nucleobase protecting groups were used (dA^{Pac}, dC^{Ac}, dG^{PrPac}). The solid support employed was controlled pore glass (CPG). Phenoxyacetyl anhydride was used as capping agent, tetrazole as activator and iodine as oxidising agent. The required amounts of phosphoramidites were left in a desiccator with P₂O₅ and KOH overnight before using them. The synthesis cycle is summarised below, together with the different solutions employed.

Step	Operation	Reagent/Solvent	Time (s)
1 (only 1 st base)	Wash	ACN	3 × 15 ^a
2	Detrytilation	3 % TCA in DCM	140 ^a
3	Wash	ACN	3 × 15 ^a
4	Dry	Argon	10 ^a
5	Coupling	0.1 M phosphoramidite in ACN + 0.45 M tetrazole in ACN	8 ^a + 120 ^b
6	Wash	ACN	3 × 15 ^a
7	Capping	85:10:5 THF/Pyr/Pac ₂ O (Cap A) + 16 % <i>N</i> -methylimidazole in THF (Cap B)	15 ^a + 6 ^b
8	Oxidation	0.02 M I ₂ in 7:2:1 THF/Pyr/H ₂ O	24 ^a + 60 ^b
9	Wash	ACN	3 × 15 ^a + 2 × 20 ^a

^a Continuous flow. ^b Incubation

Cleavage of oligonucleotides or oligonucleotide conjugates was carried out by incubating the CPG support for 1 or 2 h with concentrated aqueous ammonia at room temperature. Ammonia was removed in vacuo (miVac) before analysing or lyophilising the crudes.

E.M.M.5 Other instruments used

Lyophilisation of aqueous solutions was accomplished in either a Labconco Freezone 6 or a Virtis Freezemobile 25EL freeze-dryer.

Suspensions were centrifuged in either a 5430R Eppendorf Centrifuge or an IKA Mini G. pH Was measured with a Crison micrPH Basic 20.

Solutions containing ammonia were evaporated (partially or to dryness) in a miVac duo concentrator from GeneVac.

Mixing and heating of reactions performed in water (or water mixtures) of experiments described in chapter 3 and 4 was performed in an Eppendorf Thermomixer Compact device.

E.M.M.6 Abbreviations

AC: ammonium citrate

ACN: Acetonitrile

COMU: 1-[(1-(cyano-2-ethoxy-2-oxoethylideneaminoxy)-dimethylamino-morpholinomethylene)] methanaminium hexafluorophosphate

CPG: controlled pore glass

DCM: Dichloromethane

DHB (2,5-dihydroxybenzoic acid):

DIPC: *N,N'*-Diisopropylcarbodiimide

DMF: *N,N*-dimethylformamide

EDT : 1,2-ethanedithiol

ESI: ElectroSpray Ionisation

HOBt: 1-hydroxybenzotriazole

HPLC/MS: High Performance Liquid Chromatography/Mass spectrometry

HPLC: High Performance Liquid Chromatography

MALDI-TOF: Matrix-Assisted Laser Desorption/Ionisation-Time Of Flight

MS: Mass Spectrometry

NMP: *N*-methylpyrrolidone

NMR :Nuclear Magnetic Resonance

PNA: Peptide Nucleic Acid

TFA: trifluoroacetic acid

THAP: 2,4,6-trihydroxyacetophenone

TIS: Triisopropylsilane

TLC: Thin Layer Chromatography

TMS: tetramethylsilane

E.M.M.7 Bibliography

- (1) Guide for Handling PNA - Brochures\Oligonucleotides\PNA-guide.pdf
<http://www.eurogentec.com/uploads/FileBrowse/Brochures%5COligonucleotides%5CPNA-guide.pdf> (accessed May 31, 2016).
- (2) Oligo Quantification <http://www.sigmaaldrich.com/technical-documents/articles/biology/quantitation-of-oligos.html> (accessed May 31, 2016).
- (3) Schmid, F.-X. In *eLS*; John Wiley & Sons, Ltd, 2001.
- (4) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34* (2), 595.

Experimental section: chapter 2

E.2 General methods

On-resin DA conjugation procedure

The appropriate amount (2-17 mg) of resin-linked, fully protected diene-polyamide (PNA or peptide) or diene-oligonucleotide was carefully weighed into a 500 μ L Eppendorf tube (the μ mol of diene were calculated on the basis of the weigh and the substitution degree of the polyamide- or oligonucleotide-resin, see below). The maleimide-containing reactant was dissolved (or suspended in the solvent), poured into the Eppendorf and allowed to react. Solvent, reaction time, temperature, molar excess, and concentration of the reagent are summarised below and indicated in each particular reaction description. The Eppendorf tube was sonicated immediately after the addition of the dienophile to ensure homogeneity of the mixture (until it was observed that all the beads were covered by the maleimide solution), and three or four times over the reaction course. The Eppendorf tube was placed in a sand bath heated at the desired temperature, and the reaction was allowed to proceed with no additional stirring.

After the appropriate reaction time, the Eppendorf content was filtered through a 2 mL syringe equipped with a filter, and the resin washed with H₂O (3 \times), ACN (3 \times), DMF (3 \times), H₂O (3 \times), ACN (3 \times), and DCM (3 \times), and allowed to air dry. In case of sparingly soluble maleimide-containing reagents (biotin and dansyl derivatives), 0.1% of TFA was added to the H₂O and ACN washings to increase their solubilizing properties. After washing the resin, the conjugate was cleaved as described in the polyamide and oligonucleotide synthesis sections and then analysed by HPLC and characterised by MALDI-TOF MS.

Method	Equivalents of maleimide derivative	Solvent	Concentration (mM)	Time (h)	Temp. (°C)
A	1.5	H ₂ O	20	70	40
B	1.5	H ₂ O/DMSO 1:1	20	70	40
C	1.5	H ₂ O	20	24	65
D	3	H ₂ O	40	24	65
E	5	H ₂ O	40	24	65
F	5	H ₂ O/DMSO 80:20	40	24	65
G	3	H ₂ O/DMSO 80:20	40	24	65
H	3	H ₂ O/DMSO 65:35	40	24	65
I	5	H ₂ O/DMSO 65:35	40	24	65
J	5	H ₂ O	40	48	40

HPLC or HPLC/MS conditions

Analysis conditions

Set 1: Shimadzu instrument, Jupiter Proteo column (4 μ m, 90 Å, 250 \times 4.6 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.045 % TFA in water, solvent B: 0.036 % TFA in ACN, flow: 1 mL/min, detection wavelength: 220 and 260 nm.

Set 2: Waters 2695 instrument, Jupiter column (10 μ m, 300 Å, 250 \times 10 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 M TEAA in water, solvent B: ACN, flow: 3 mL/min, detection wavelength: 260 nm.

Set 3: Waters 2695 instrument, Jupiter column (10 μ m, 300 Å, 250 \times 4.6 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 M TEAA in water, solvent B: ACN, flow: 1 mL/min, detection wavelength: 260 nm.

Purification conditions

Set 1: Jupiter Proteo column (10 μ m, 90 Å, 250 \times 10.0 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 % TFA in water, solvent B: 0.1 % TFA in ACN, flow: 4 mL/min, detection wavelength: 220 and 260 nm.

Set 2: Jupiter Proteo column (10 μ m, 90 Å, 250 \times 10.0 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 % TFA in water, solvent B: 0.1 % TFA in ACN, flow: 3 mL/min, detection wavelength: 220 and 260 nm.

HPLC/MS analysis conditions

Set 1: Waters 2695 instrument coupled to a Waters Micromass ZQ 4000 mass spectrometer, Jupiter Proteo column (4 μm , 90 \AA , 250 \times 4.6 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 % formic acid in water, solvent B: 0.1 % formic acid in ACN, flow: 1 mL/min, detection wavelength: 220 and 280 nm.

Treatments employed for CNE and Pac removal

Morpholine treatment: 5 mL of a 2 % morpholine solution in ACN (v/v) were passed, at an approximate 1.5 mL/min flow rate, through an oligonucleotide synthesis column in which the oligonucleotide-resin has been placed. After this, the resin was washed with ACN (20 mL at a 4 mL/min flow and 20 mL at a 10 mL/min flow) and dried under a N_2 stream.

TEA treatment 1: 2 mL Of a 40 % TEA solution in ACN (v/v) were passed at an approximate 1 mL/min flow rate, through an oligonucleotide synthesis column containing the oligonucleotide- or conjugate-resin. After this, the synthesis column was placed between two syringes and incubated with a 5 mL fresh aliquot of the TEA/ACN solution for 20 minutes, with occasional circulation performed with a push-pull movement of the syringes. This treatment was performed twice before washing the resin with a large amount of ACN.

TEA treatment 2: the oligonucleotide- or conjugate-resin was incubated with fresh aliquots (of the synthesis column volume) of a 40 % TEA solution in ACN (v/v) as follows: 3 \times 1.5 min, 3 \times 5 min, 3 \times 10 min and 4 \times 30 min. After this, the resin was washed with a large amount of ACN.

DEA treatment: The oligonucleotide- or conjugate-resin was placed inside an oligonucleotide synthesis column and incubated with a 10 % DEA solution in ACN (v/v) for 5 min. This treatment was performed a total of 3 times with fresh aliquots of said solution before washing the resin with a large amount of ACN.

E.2 Analysis of the polyamide- and oligonucleotide-resins before diene introduction

Assessment of the purity of the [9-mer peptide]-resin precursor of peptide-resins **2.8** and **2.31** and the [12-mer PNA]-resin precursor of PNA-resins **2.13**, **2.13a** and **2.13b** was achieved by deprotecting a small aliquot of the polyamide- or oligonucleotide-resin before introduction of an additional amino acid or the diene. Cleavage and deprotection was performed as described in section **E.M.M.4**.

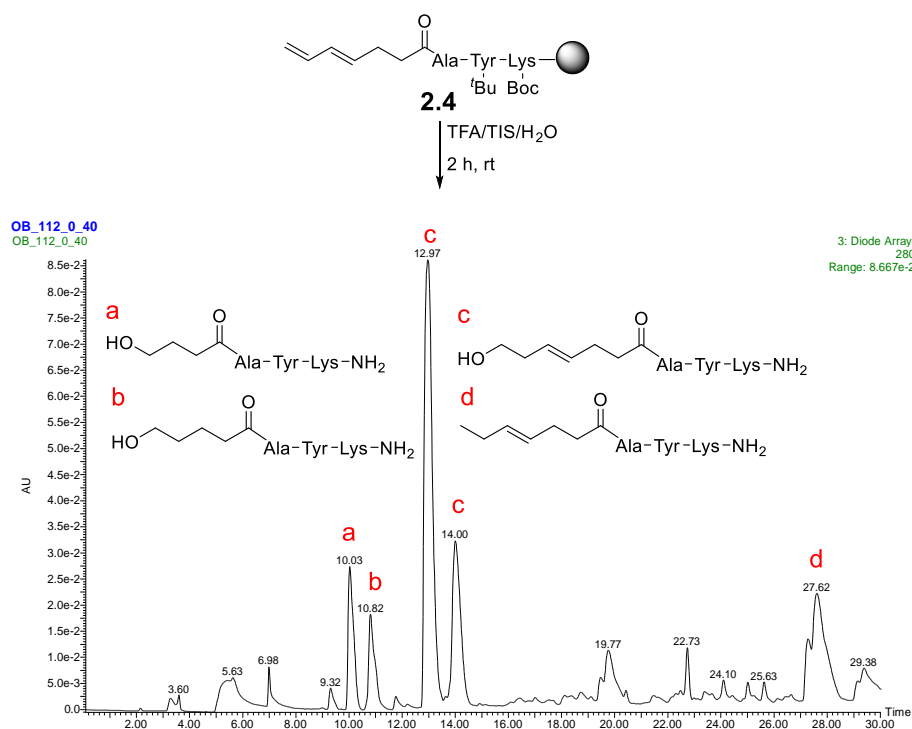
The substitution degree of polyamide-resins was determined by taking a carefully weighed aliquot of the resin, removing the Fmoc group with piperidine and quantitating the amount of 9-fluorenylmethylpiperidine formed ($\lambda_{\max}=301$ nm, $\varepsilon=7800$).¹ This value was taken as the substitution degree of the resin-linked diene-polyamides in conjugation experiments, without any further correction for the addition of the diene or extra amino acid, and used to determine the overall synthesis and purification yield for the isolated conjugates. Fmoc-Gly-Arg(Pbf)-Gly-Ser(^tBu)-Tyr(^tBu)-Glu(O^tBu)-Ala-Tyr(^tBu)-Lys(Boc)-resin, substitution degree: 0.166 mmol/g, Fmoc-t-t-Lys(Boc)-resin, substitution degree: 0.17 mmol/g; Fmoc-c(Bhoc)-a(Bhoc)-t-a(Bhoc)-g(Bhoc)-c(Bhoc)-g(Bhoc)-t-t-t-c(Bhoc)-resin, substitution degree: 73 μ mol/g.

No corrections to the substitution degree of the CPG supports was performed upon oligonucleotide synthesis and diene incorporation.

E.2.1 Conjugation of diene-derivatised polyamides

E.2.1.2 Decomposition of dienes under the cleavage and deprotection conditions

1,3-Diene-Ala-Tyr(^tBu)-Lys(Boc)-Rink amide resin (**2.4**) was reacted with a 95:2.5:2.5 TFA/TIS/H₂O mixture for 2 h at r. t. Filtrate and washings (TFA) were collected and concentrated by blowing N₂ over the mixture, and diethyl ether was added to the resulting oil or semi-solid. Water was added to dissolve the peptide, the biphasic mixture shaken, and the two phases allowed to separate. The organic phase was discarded and new diethyl ether was added. This washing procedure was repeated 3 times. The aqueous phase was then lyophilised and the crude analysed by HPLC/MS (analysis conditions Set 1, 0-40 % B).



Product	Mass calculated	<i>m/z</i> Found for [M+H] ⁺
a	C ₂₂ H ₃₅ N ₅ O ₆ 465.3	466.1
b	C ₂₃ H ₃₇ N ₅ O ₆ 479.3	480.1
c	C ₂₅ H ₃₉ N ₅ O ₆ 505.3	506.1
d	C ₂₅ H ₃₉ N ₅ O ₅ 489.3	490.1

The structures drawn fit with the found *m/z* ratios, but the positions of double bonds and hydroxyl groups have not been unequivocally established. For instance, the peaks eluting at 13.0 and 14.0 min have the same mass, but in addition to the structure shown on the HPLC/MS profile, other possible structures are H₃C-CH(OH)-CH=CH-CH₂-CH₂-CO-Ala-Tyr-Lys-NH₂, H₂C=CH-CH(OH)-CH₂-CH₂-CH₂-CO-Ala-Tyr-Lys-NH₂ and H₂C=CH-CH₂-CH(OH)-CH₂-CH₂-CO-Ala-Tyr-Lys-NH₂. The same for the product eluting at 19.8 min, for which all we know is that only one of the double bonds has been retained.

E.2.1.8 Selection of the diene

Diene **2.2** was synthesised as described in reference 2. Both **2.1** and **2.2** were coupled to the peptide-resin as the other amino acids, but using a 10-fold excess of reagents (instead of the normal 3-fold excess). On-resin DA reactions were carried out as follows:

Reagents: **2.3 + 2.5**

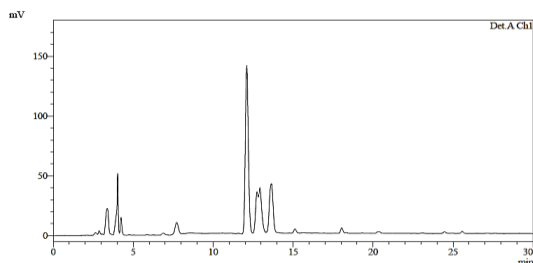
Product: **2.6**

Conjugation method: **A**

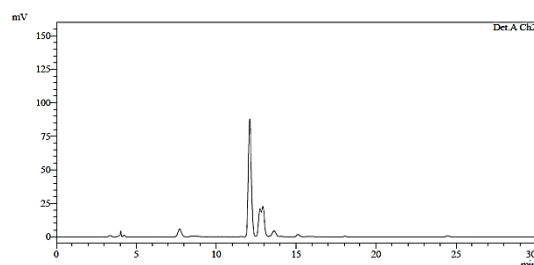
HPLC yield (220 nm): **77 %**

HPLC traces of the crude:

220 nm



260 nm



Reagents: **2.4 + 2.5**

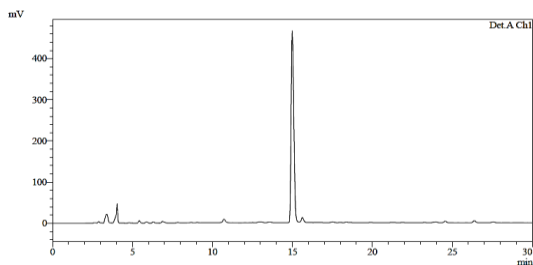
Product: **2.7**

Conjugation method: **A**

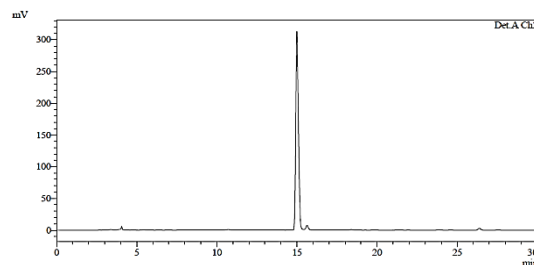
HPLC yield (220 nm): **89 %**

HPLC traces of the crude:

220 nm



260 nm



E.2.1.9 Assessment of the solvent influence on the on-resin Diels-Alder conjugation

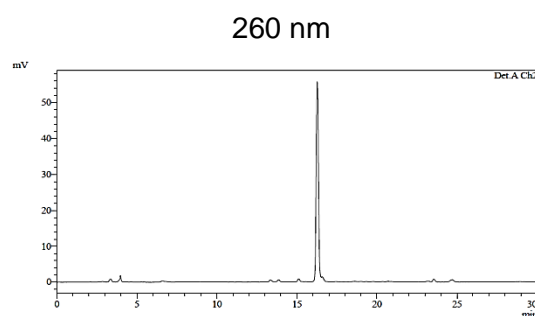
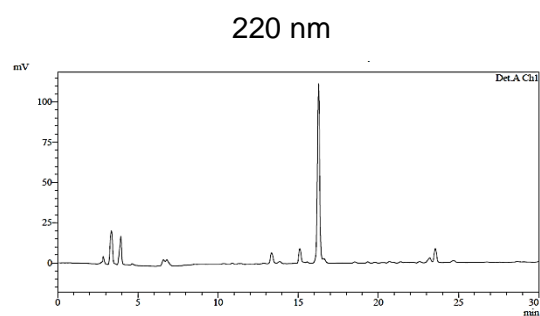
Reagents: **2.8 + 2.5**

Product: **2.9**

Conjugation method: **A**

HPLC yield (220 nm): **74 %**

HPLC traces of the crude:



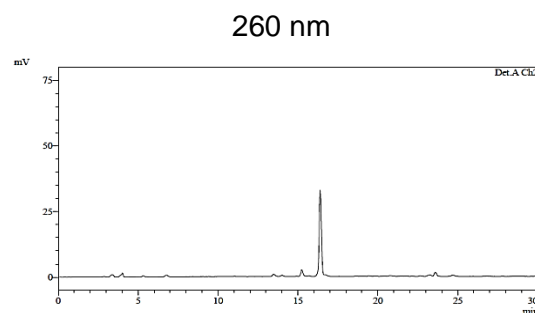
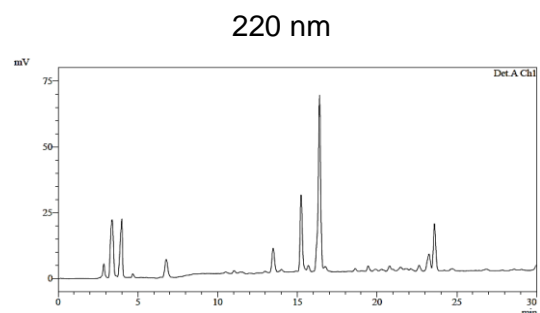
Reagents: **2.8 + 2.5**

Product: **2.9**

Conjugation method: **B**

HPLC yield (220 nm): **48 %**

HPLC traces of the crude:



E.2.1.10 Optimizing the reaction conditions

For reactions between **2.8** and **2.5** carried out using conjugation methods A and B see section **E.2.1.9**.

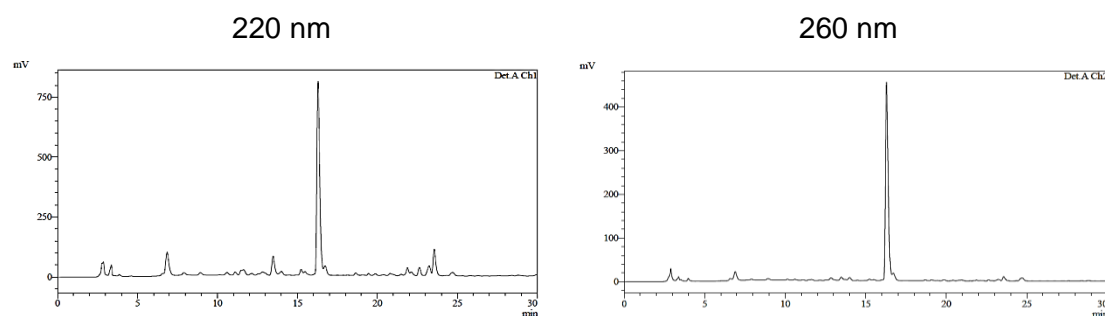
Reagents: **2.8 + 2.5**

Product: **2.9**

Conjugation method: **C**

HPLC yield (220 nm): **62 %**

HPLC traces of the crude:



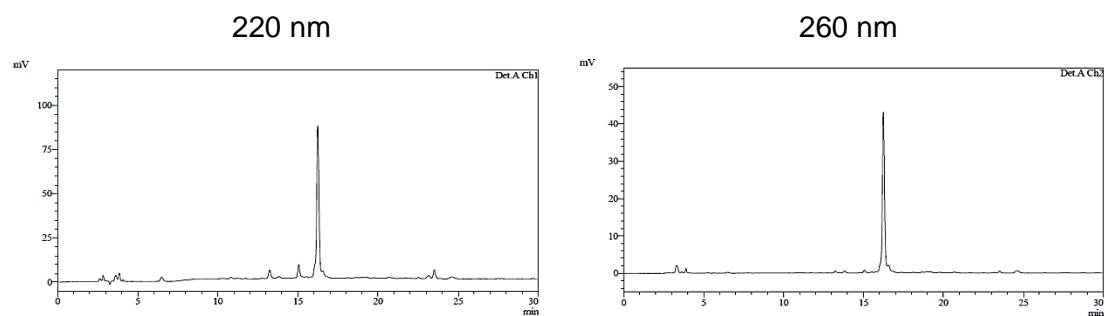
Reagents: **2.8 + 2.5**

Product: **2.9**

Conjugation method: **D**

HPLC yield (220 nm): **79 %**

HPLC traces of the crude:



E.2.1.11 Broadening the scope of the reaction: conjugation of different polyamides

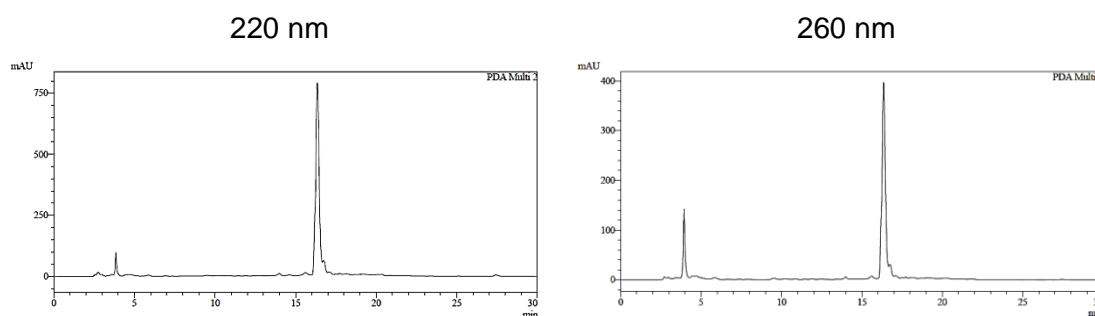
Reagents: **2.10 + 2.11**

Product: **2.12**

Conjugation method: **D**

HPLC yield (260 nm): **92 %**

HPLC traces of the crude:



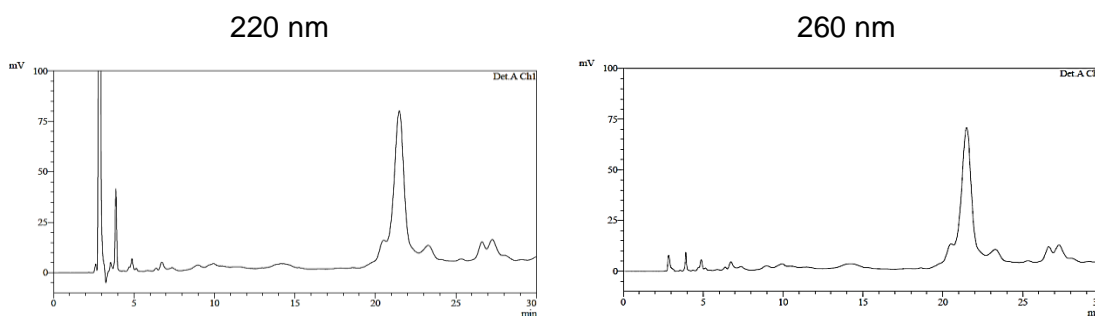
Reagents: **2.13 + 2.14**

Product: **2.15**

Conjugation method: **E**

HPLC yield (260 nm): **72 %**

HPLC traces of the crude:



For evaluation of the chain length effect on the reaction outcome most of the experiments performed have already been described in sections **E.2.1.8** and **E.2.1.10**. The remaining experiment is described below:

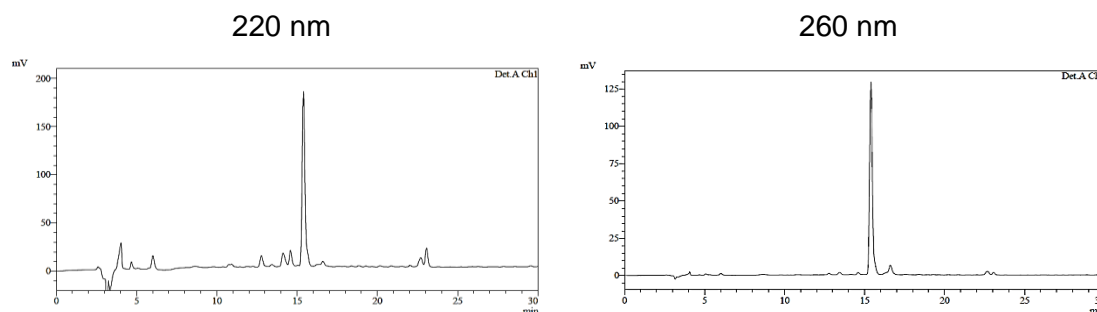
Reagents: **2.8 + 2.16**

Product: **2.17**

Conjugation method: **D**

HPLC yield (220 nm): **69 %**

HPLC traces of the crude:



E.2.1.12 Use of protected maleimides to obtain conjugates with different linking sites

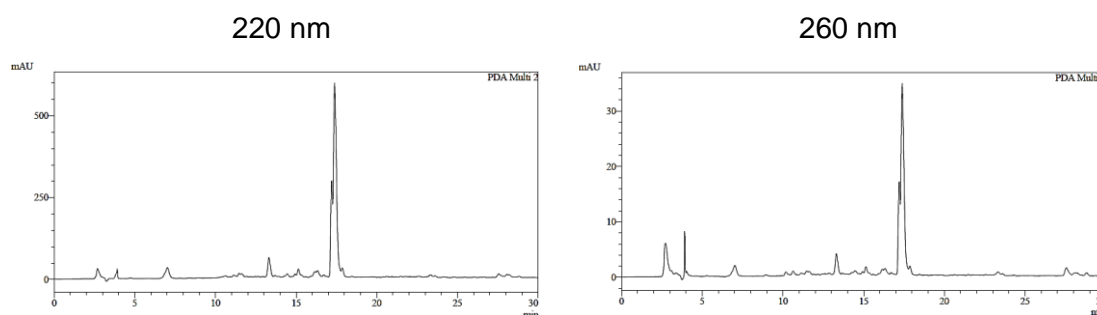
Reagents: **2.8 + 2.18**

Product: **2.19**

Conjugation method: **D**

HPLC yield (220 nm): **82 %**

HPLC traces of the crude:



Reagents: **2.13a + 2.20**

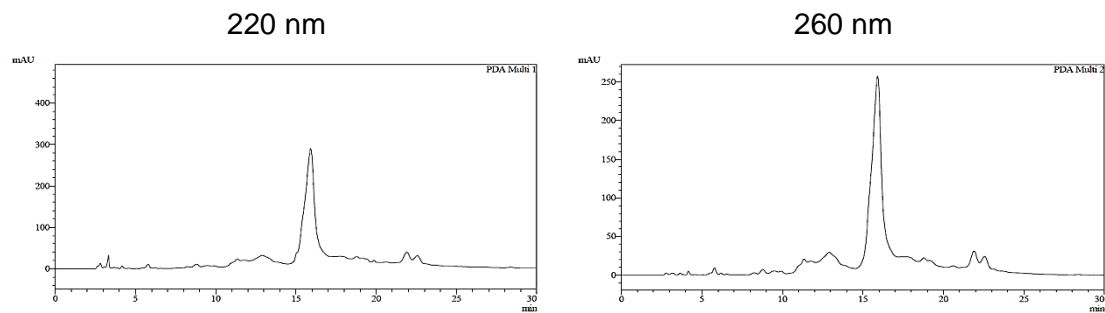
Product: **2.21**

Conjugation method: **E**

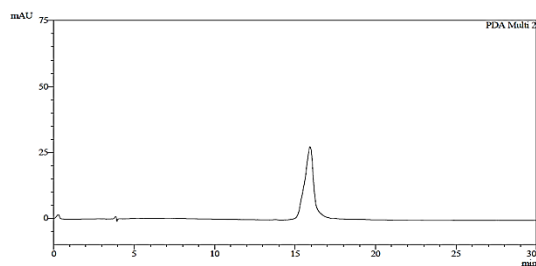
HPLC yield (260 nm): **55 %**

Isolated yield: **30 %**

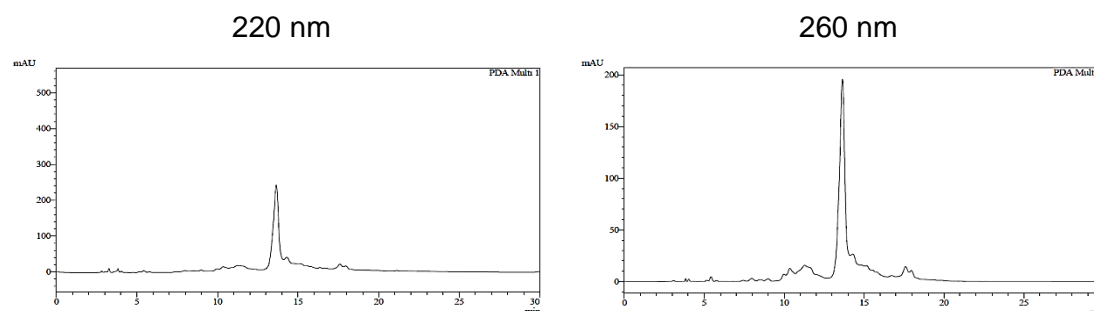
HPLC traces of the crude:



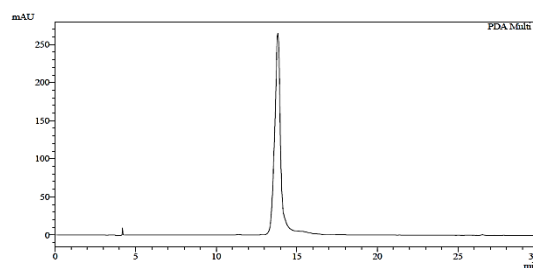
HPLC trace of the pure compound (260 nm):



Reagents: **2.13a + 2.22**
Product: **2.23**
Conjugation method: **E**
HPLC yield (260 nm): **58 %**
Isolated yield: **28 %**
HPLC traces of the crude:

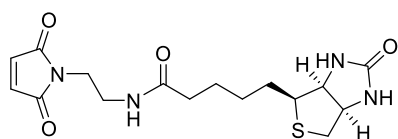


HPLC trace of the pure compound (260 nm):



E.2.1.13 Conjugation with molecules sparingly soluble or insoluble in water

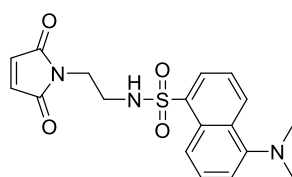
exo-(2,5-Dimethylfuran)-protected biotin-maleimide (2.24)



Biotin-NHS ester (154 mg, 0.49 mmol) was dissolved in DMF (2 mL) and *exo*-(2,5-dimethylfuran)-protected 3-maleimidoethylamine-TFA (207 mg, 0.59 mmol, 1.2 equiv) was added. After complete dissolution of all products TEA (83 μ L, 0.59 mmol, 1.2 equiv) was added and the mixture stirred at 40 $^{\circ}$ C for 18 h. The solvents were evaporated in vacuo and the crude product was washed twice with DCM (HPLC quality) to obtain a white powder solid (104 mg, 45%).

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ : 7.78 (t, $J = 5.9$ Hz, 1H), 6.41 (s, 1H), 6.36 (s, 2H), 6.35 (s, 1H), 4.30 (m, 1H), 4.13 (m, 1H), 3.40 (t, $J = 6.3$ Hz, 2H), 3.16 (q, $J = 6.1$ Hz, 2H), 3.12 – 3.06 (m, 1H), 2.86 (s, 2H), 2.81 (dd, $J = 12.4, 5.1$ Hz, 1H), 2.57 (d, $J = 12.4$ Hz, 1H), 1.99 (t, $J = 7.4$ Hz, 2H), 1.64 – 1.56 (m, 1H), 1.53 (s, 6H), 1.50 – 1.39 (m, 3H), 1.35 – 1.22 ppm (m, 2H); ^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ : 174.8, 172.2, 162.7, 140.6, 86.9, 61.0, 59.2, 55.4, 52.2, 39.8, 37.7, 35.9, 35.1, 28.2, 28.0, 25.0, 15.7 ppm; HRMS (ESI, positive mode): M calc. for $\text{C}_{22}\text{H}_{30}\text{N}_4\text{O}_5\text{S}$ 462.1937, m/z found 463.2000 $[\text{M}+\text{H}]^+$, 485.1815 $[\text{M}+\text{Na}]^+$.

exo-(2,5-Dimethylfuran)-protected dansyl-maleimide (2.25)



Dansyl chloride (31 mg, 0.115 mmol, 1.33 equiv) was dissolved in DMF (6.5 mL) and DIPEA (38 μL , 0.215 mmol, 2.5 equiv) and *exo*-(2,5-dimethylfuran)-protected 3-maleimidoethylamine-TFA (30 mg, 0.086 mmol) were added. The reaction mixture was stirred for 1 h at rt, and then taken to dryness in vacuo. DCM (20 mL) was added to dissolve the resulting oil, and this organic phase was washed with 1% aq HCl (3 \times 15 mL). The organic layer was dried over anhydrous MgSO_4 , filtered and concentrated in vacuo (33 mg, 82 %). The product was used without further purification.

^1H NMR (400 MHz, CDCl_3) δ : 8.51 (d, $J = 8.5$ Hz, 1H), 8.21 (dd, $J = 12.9, 8.6$ Hz, 2H), 7.52 (m, 2H), 7.16 (d, $J = 7.5$ Hz, 1H), 6.25 (s, 2H), 5.29 - 5.26 (m, 1H), 3.57 - 3.51 (m, 2H), 3.13 (q, $J = 6.0$ Hz, 2H), 2.86 (s, 6H), 2.55 (s, 2H), 1.62 ppm (s, 6H); ^{13}C NMR (101 MHz, CDCl_3) δ : 187.2, 175.0, 151.9, 140.7, 134.4, 130.4, 129.8, 129.4, 128.4, 123.3, 119.0, 115.2, 87.6, 52.4, 45.4, 41.1, 38.1, 15.8 ppm (impurities detected at 1.26 ppm in the ^1H NMR correlate with impurities at 29.7 ppm and 21.8 ppm as seen by gHSQC); HRMS (ESI, positive mode): M calc. for $\text{C}_{24}\text{H}_{27}\text{N}_3\text{O}_5\text{S}$ 469.1671, m/z found 470.1745 $[\text{M}+\text{H}]^+$, 939.3406 $[2\text{M}+\text{H}]^+$.

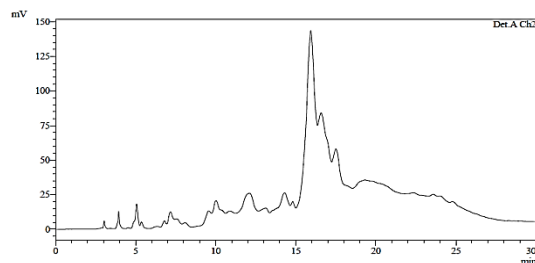
Reagents: **2.13 + 2.24**

Product: **2.26**

Conjugation method: **F**

HPLC yield (220 nm): **n.d.**

HPLC traces of the crude (shown only at 260 nm):



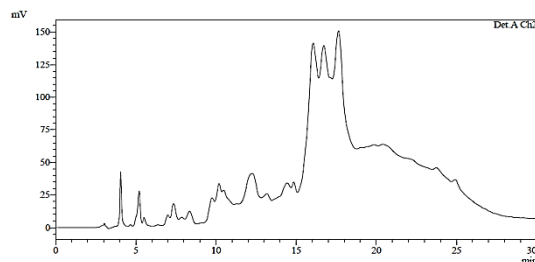
Reagents: **2.13 + 2.24**

Product: **2.26**

Conjugation method: **E**

HPLC yield (220 nm): **n.d.**

HPLC traces of the crude (shown only at 260 nm):



Reagents: **2.13a + 2.24**

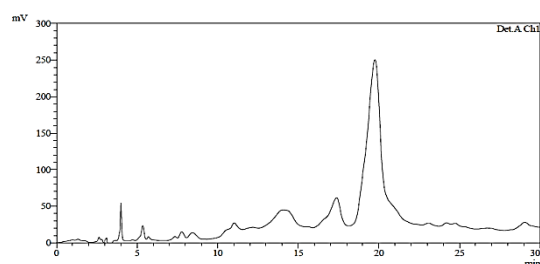
Product: **2.26a**

Conjugation method: **E**

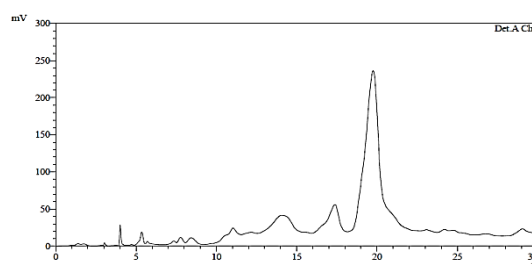
HPLC yield (260 nm): **60 %**

HPLC traces of the crude:

220 nm



260 nm



Reagents: **2.13b + 2.24**

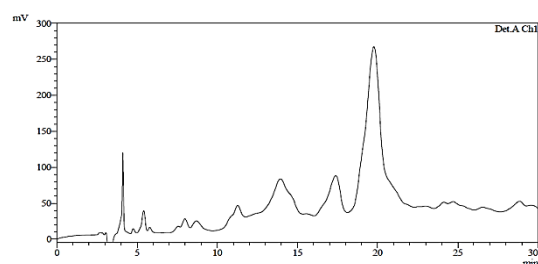
Product: **2.26b**

Conjugation method: **E**

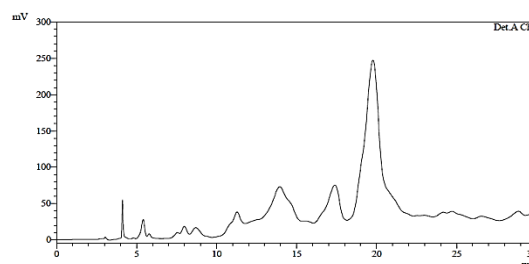
HPLC yield (260 nm): **56 %**

HPLC traces of the crude:

220 nm



260 nm



Reagents: **2.8 + 2.24**

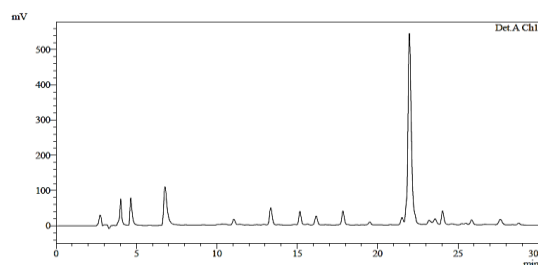
Product: **2.27**

Conjugation method: **D**

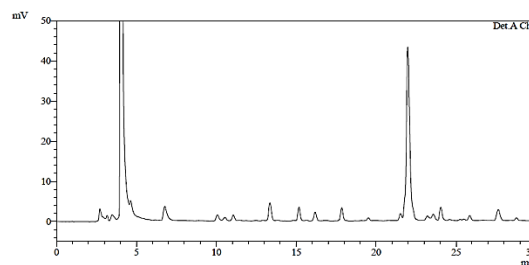
HPLC yield (220 nm): **67 %**

HPLC traces of the crude:

220 nm



260 nm



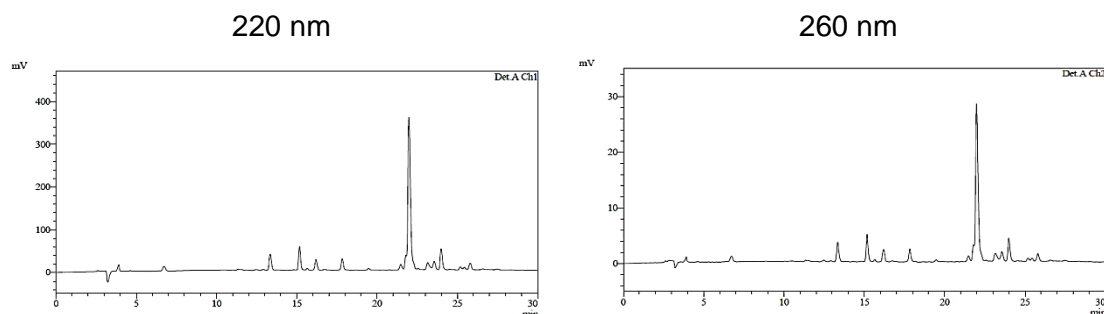
Reagents: **2.8 + 2.24**

Product: **2.27**

Conjugation method: **G**

HPLC yield (220 nm): **63 %**

HPLC traces of the crude:



Reagents: **2.8 + 2.25**

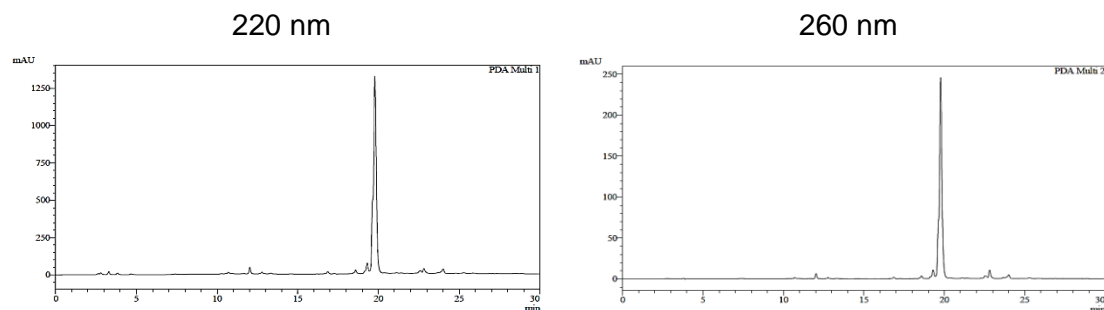
Product: **2.28**

Conjugation method: **H**

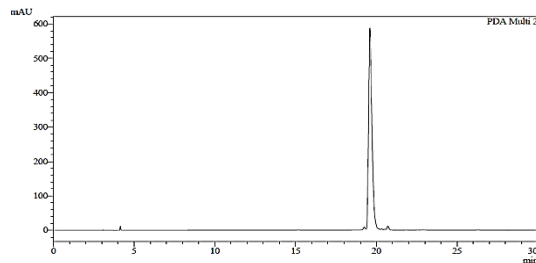
HPLC yield (220 nm): **88 %**

Isolated yield: **48 %**

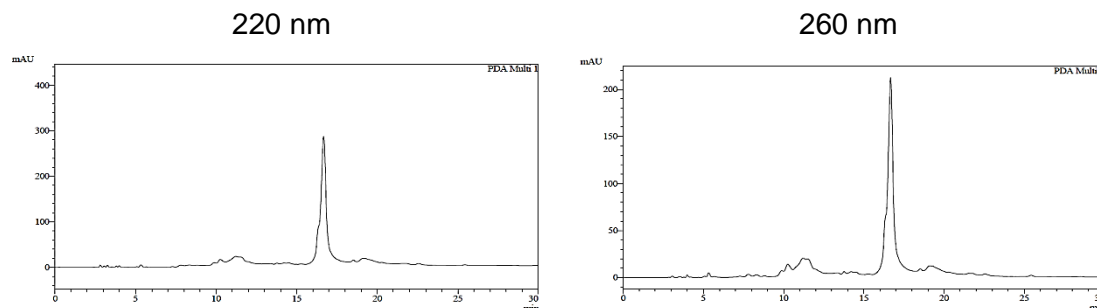
HPLC traces of the crude:



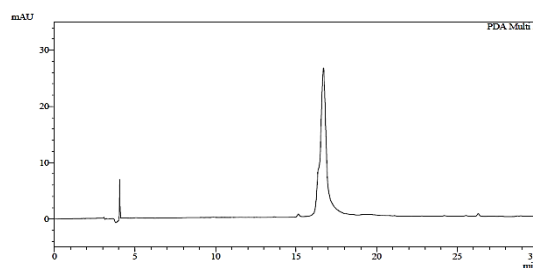
HPLC trace of the pure compound (220 nm):



Reagents: **2.13a + 2.25**
Product: **2.29**
Conjugation method: I
HPLC yield (260 nm): 72 %
Isolated yield: 19 %
HPLC traces of the crude:



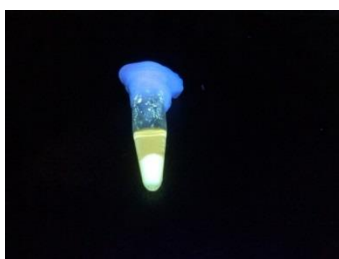
HPLC trace of the pure compound (260 nm):



Inspection of the reaction mixture of 2.8 + 2.25 under UV light

As can be seen below, at the beginning of the reaction the resin surface presents a higher fluorescence intensity than the solution, which suggests that the fluorophore is essentially near the reactive points (picture on the left). Once the reaction has taken place and the resin has been extensively washed with both water and organic solvents it is fluorescent due to the functionalization of the polyamide with the dansyl fluorophore (picture on the right).

Reaction mixture exposed to UV light at the beginning of the reaction.



Resin after conjugation reaction and extensive washings.



E.2.1.14 Synthesis of double conjugates

1) Assessment of the stability of the Trt group to the on-resin DA reaction conditions

12.1 mg of Fmoc-Cys(Trt)-Gly-Arg(Pbf)-Gly-Ser(^tBu)-Tyr(^tBu)-Glu(O^tBu)-Ala-Tyr(^tBu)-Lys(Boc)-resin were incubated with 3-maleimidopropanoic acid (0.2 mL of a 40 mM aqueous solution, 4 equiv.) at 65 °C for 24 h. After filtration and washing the peptide was deprotected and cleaved from the solid support by using a TFA/TIS/EDT/H₂O mixture (94:1:2.5:2.5) for 2 h. The crude peptide was precipitated with diethyl ether, extracted with water and this solution analysed by HPLC. The major peak was collected and analysed by MALDI-TOF mass spectrometry. Its mass (m/z 1355.0 [M+H]⁺) fitted with that of the peptide with the free cysteine (M calcd. for C₆₃H₈₃N₁₅O₁₇S 1353.6) and therefore the Trt protecting group was deemed adequate.

2) Synthesis of the double conjugate

The on-resin DA reaction to furnish **2.33** was performed as the others with the exception that a suspension of **2.30** was used instead of the completely dissolved product.

Reagents: **2.30 + 2.31**

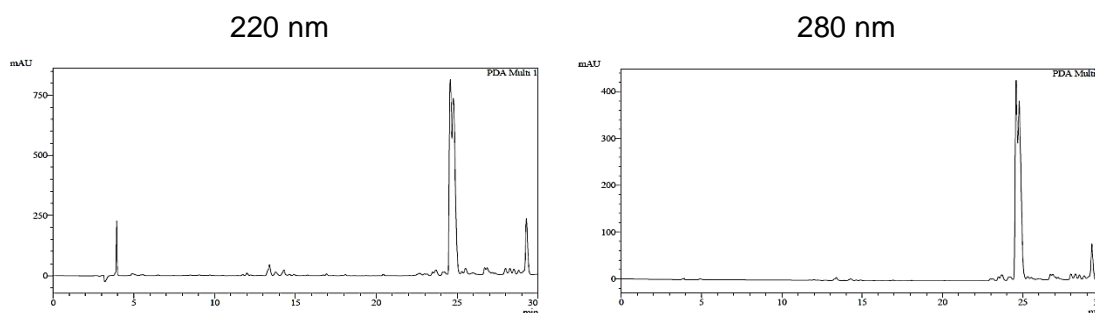
Product: **2.33**

Conjugation method: H

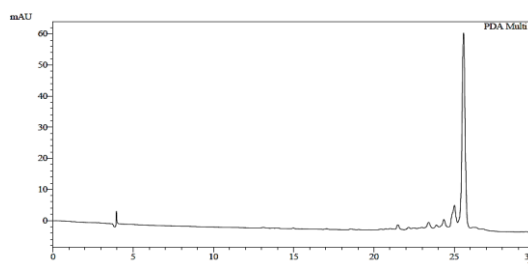
HPLC yield (220 nm): 77 %

Isolated yield: n.d.

HPLC traces of the crude:

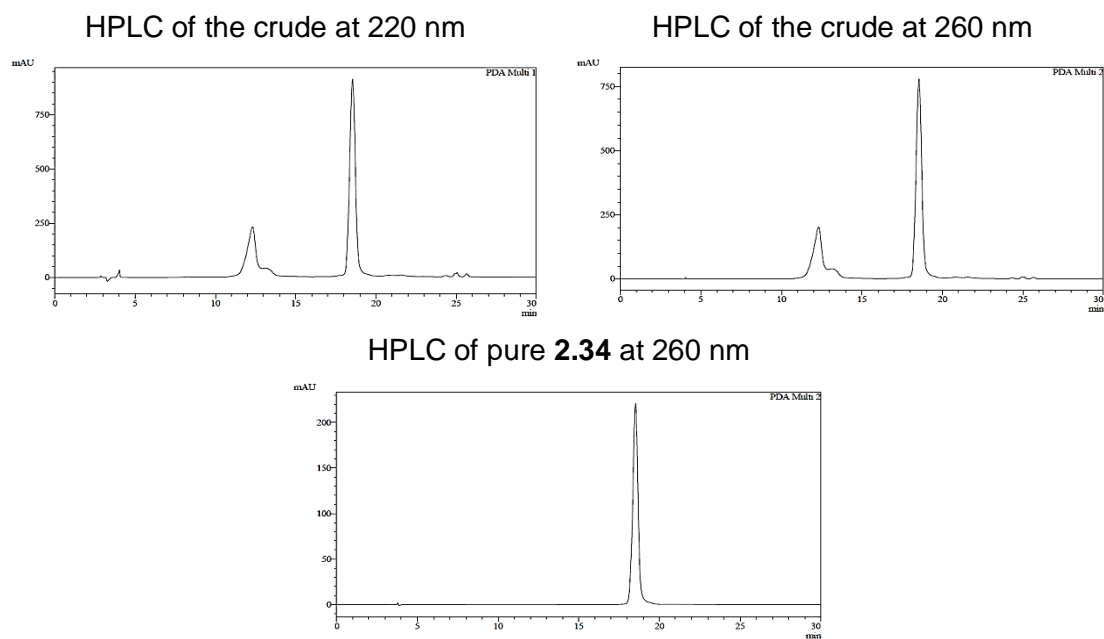


HPLC trace of the pure compound (280 nm):



The Michael-type reaction between **2.33** and **2.32** was performed as follows:

1.3 equiv.* of Mal-Lys-Lys-catagctgtttc-NH₂ (**2.32**, 4 mM) were incubated with **2.33** for 2.5 h at 37 °C in water. After that, the crude was purified by HPLC and the final product lyophilised, analysed and quantified. The final product **2.34** was obtained in 23 % overall yield.



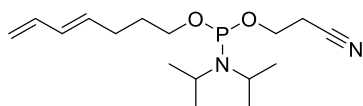
E.2.2 Conjugation of diene-derivatised oligonucleotides

E.2.2.2 Reaction parameters and analysis of the reaction outcome

(*E*)-hepta-4,6-dien-1-ol

Synthesised as in reference 3.

* HPLC analysis after the 2.5 h reaction time showed that 95-99 % of compound **2.33** had been consumed, so the ratio of the unreacted PNA peak area to double conjugate peak area at t=2.5 h allowed to determine that the actual **2.32:2.33** ratio in the Michael-type conjugation reaction had been 1.3:1.

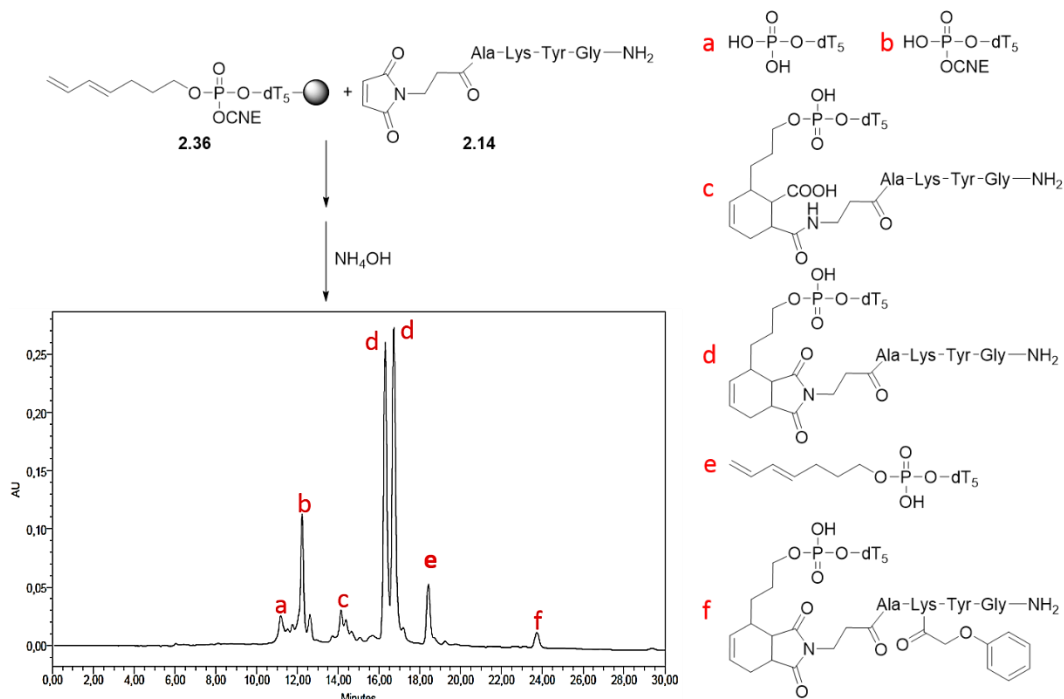
(E)-2-cyanoethyl hepta-4,6-dien-1-yl diisopropylphosphoramidite (2.35)

(E)-Hepta-4,6-dien-1-ol (107 mg, 0.96 mmol) was co-evaporated once with dry acetonitrile before being dissolved in a mixture of 380 μL of TEA and 1450 μL of DCM (both dried using molecular sieves). 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (250 mg, 1.1 equiv.) was slowly added to the reaction mixture and the reaction left at room temperature for 45 min. The reaction mixture was concentrated to dryness in vacuo and redissolved in hexane. The white precipitate that had been formed during the reaction was removed by centrifugation and decantation, and the rest of the crude purified by silica gel chromatography (isocratic 95:5 hexanes/TEA) to obtain a colourless oil (210 mg, 70 % yield).

^1H NMR (400 MHz, CDCl_3) δ : 6.30 (dt, $J = 17.0, 10.3$ Hz, 1H), 6.07 (dd, $J = 14.9, 10.7$ Hz, 1H), 5.70 (dt, $J = 14.5, 6.8$ Hz, 1H), 5.09 (d, $J = 17.2$ Hz, 1H), 4.96 (d, $J = 10.1$ Hz, 1H), 3.90 – 3.76 (m, 2H), 3.69 – 3.55 (m, 4H), 2.64 (t, $J = 6.5$ Hz, 2H), 2.17 (q, $J = 7.1$ Hz, 2H), 1.72 (p, $J = 6.9$ Hz, 2H), 1.23 – 1.17 ppm (m, 13H); ^{13}C NMR (101 MHz, cdcl_3) δ : 137.1, 134.3, 131.5, 117.6, 115.0, 63.1, 62.9, 58.4, 58.2, 43.1, 43.0, 30.7, 30.7, 28.9, 24.7, 24.6, 24.6, 24.5, 20.4, 20.3 ppm; ^{31}P NMR (162 MHz, CDCl_3) δ : 147.4 ppm; HRMS (ESI, positive mode): M calc. for $\text{C}_{16}\text{H}_{29}\text{N}_2\text{O}_2\text{P}$ 312.1967, m/z found 313.2032 $[\text{M}+\text{H}]^+$.

E.2.2.3 Work with short ON models

A representative HPLC trace of the reaction between **2.36** and **2.14** is shown below, in which the typical by-products generated can be observed.

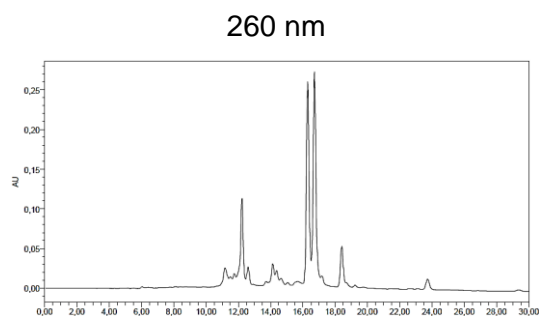


Product	Mass calculated	m/z found for $[\text{M}-\text{H}]^-$
a	$\text{C}_{50}\text{H}_{67}\text{N}_{10}\text{O}_{36}\text{P}_5$ 1538.2	1537.6
b	$\text{C}_{53}\text{H}_{70}\text{N}_{11}\text{O}_{36}\text{P}_5$ 1591.3	1590.6
c	$\text{C}_{84}\text{H}_{116}\text{N}_{17}\text{O}_{45}\text{P}_5$ 2237.6	2237.0
d	$\text{C}_{84}\text{H}_{114}\text{N}_{17}\text{O}_{44}\text{P}_5$ 2219.6	2218.9
e	$\text{C}_{57}\text{H}_{77}\text{N}_{10}\text{O}_{36}\text{P}_5$ 1632.3	1631.6
f	$\text{C}_{92}\text{H}_{120}\text{N}_{17}\text{O}_{46}\text{P}_5$ 2353.6	2353.0

Summary of the experiments:

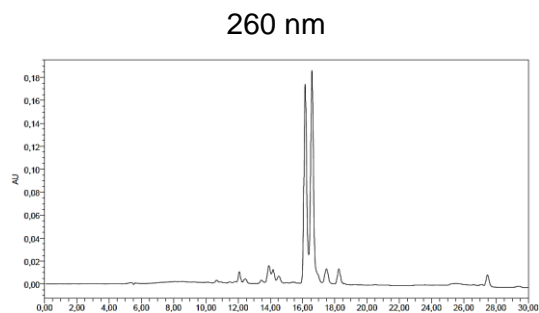
Reagents: **2.36 + 2.14**
Product: **2.37**
Conjugation method: E
HPLC yield (260 nm): 68 %
By-products (260 nm): 6 % of unreacted material, 19 % of ON fragmentation, 4 % of hydrolysed conjugate and 2 % of Pac-acylated product.

HPLC traces of the crude:



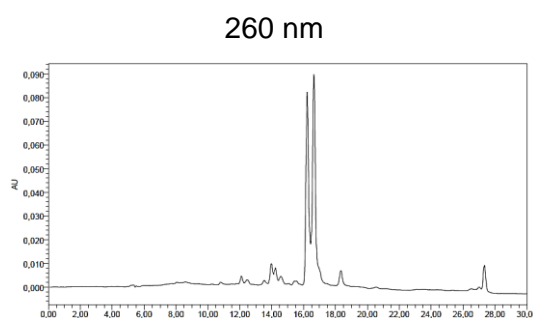
Reagents: **2.36 + 2.14**
Product: **2.37**
Conjugation method: Morpholine treatment and TEA treatment 1, then DA following method E
HPLC yield (260 nm): 83 %
By-products (260 nm): 3 % of unreacted material, 2 % of ON fragmentation, 6 % of hydrolysed conjugate and 3 % of unknown by-product.

HPLC traces of the crude:



Reagents: **2.36 + 2.14**
Product: **2.37**
Conjugation method: Morpholine treatment, then DA following method E and finally TEA treatment 1.
HPLC yield (260 nm): 80 %
By-products (260 nm): 3 % of unreacted material, 3 % of ON fragmentation, 8 % of hydrolysed conjugate and 5 % of unknown by-product.

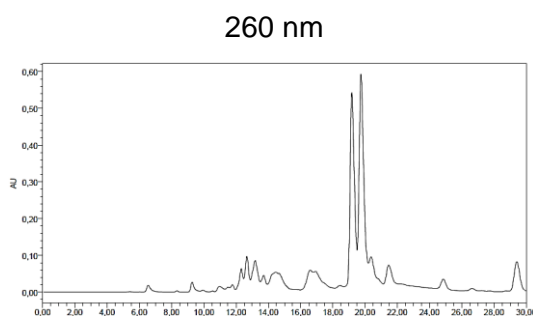
HPLC traces of the crude:



E.2.2.4 Mixed ON sequences

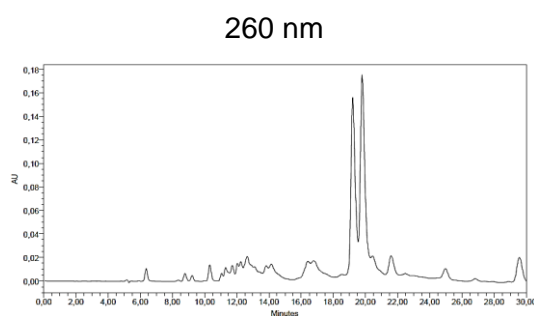
Reagents: **2.38 + 2.14**
Product: **2.39**
Conjugation method: Morpholine treatment, then DA following method E and finally TEA treatment 1.
HPLC yield (260 nm): 60 %
By-products (260 nm): 3 % of unreacted material, 20 % of ON fragmentation, 8 % of hydrolysed conjugate and 6 % of Pac-acylated product.

HPLC traces of the crude:



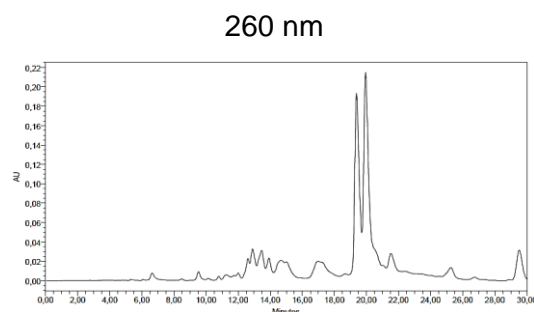
Reagents: **2.38 + 2.14**
Product: **2.39**
Conjugation method: Morpholine treatment (2x), then DA following method E and finally TEA treatment 2.
HPLC yield (260 nm): 59 %
By-products (260 nm): 3 % of unreacted material, 19 % of ON fragmentation, 6 % of hydrolysed conjugate and 6 % of Pac-acylated product.

HPLC traces of the crude:



Reagents: **2.38 + 2.14**
Product: **2.39**
Conjugation method: Morpholine treatment (2x), then DA following method E, TEA treatment 2 and DEA treatment.
HPLC yield (260 nm): 57 %
By-products (260 nm): 3 % of unreacted material, a 23 % of ON fragmentation, 7 % of hydrolysed conjugate and 6 % of Pac-acylated product.

HPLC traces of the crude:



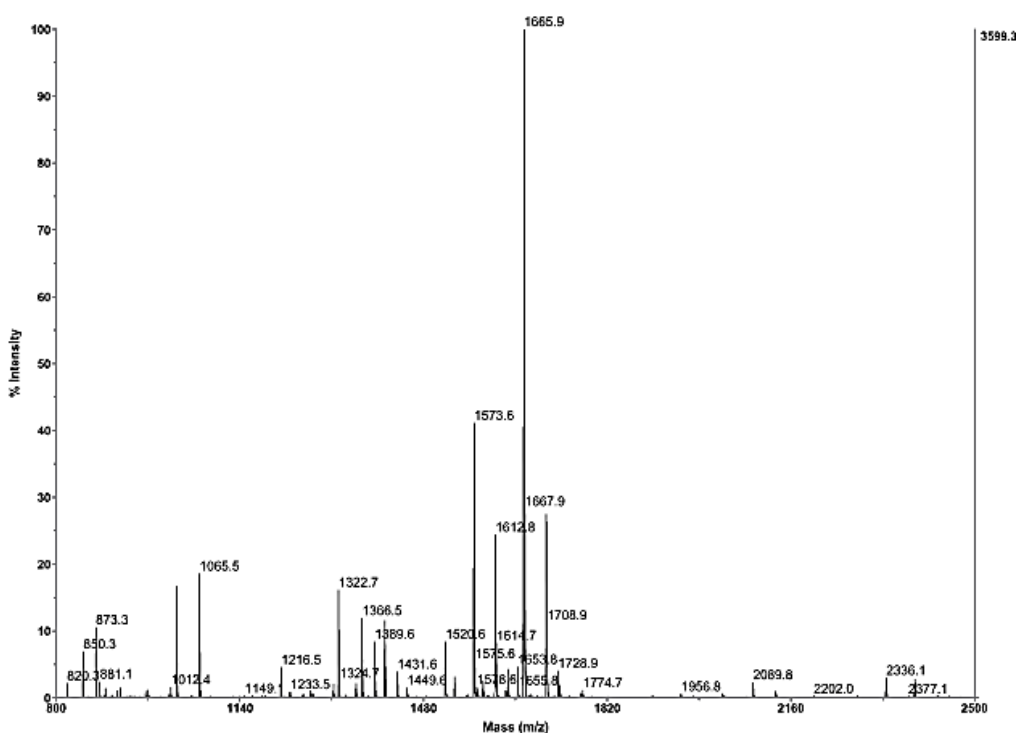
E.2.2.5 MALDI-TOF MS monitoring of the conjugation reaction

1) Method for sample preparation

Samples for MALDI-TOF MS analysis of oligonucleotide- or conjugate-resins were prepared as described by Vasseur and co-workers,⁴ with the only difference that the matrix mixture used was prepared by combining 0.5 μL of a 10 mg/mL THAP solution in a 1:1 $\text{H}_2\text{O}/\text{ACN}$ mixture and 0.5 μL of a 50 mg/mL AC solution in H_2O .

2) Analysis of the resin-linked conjugate 2.42

The resin-linked conjugate formed upon reaction between **2.14** and **2.41**, to which no morpholine, TEA or DEA treatments had been performed, was subjected to MALDI-TOF MS analysis. The spectrum obtained is shown below, and the processed data summarised in the following tables.



Ions corresponding to oligonucleotide backbone cleavage found during MALDI-TOF MS analysis of the resin-linked conjugate.		
m/z found [M-H]⁻	Structure	Mass calculated for M
849.3	p-TC-p + 3 CNE	850.1
873.3	p-AT-p + 3 CNE	874.1
1023.4	p-TG-p + 3 CNE + Pac	1024.2
1065.5	p-TG-p + 3 CNE + ⁱ PrPac	1066.2
1354.5	p-TGC- + 3 CNE + ⁱ PrPac (or Pac + Ac)	1355.3
1365.5	p-TGC-p + 4 CNE + Pac	1366.2
1389.6	p-ATG-p + 4 CNE + Pac	1390.2
1407.6	p-TGC-p + 4 CNE + ⁱ PrPac (or Pac + Ac)	1408.3
1431.6	p-ATG-p + 4 CNE + ⁱ PrPac	1432.3
1449.6	p-TGC-p + 4 CNE + ⁱ PrPac + Ac	1450.3
2088.8	p-TCATG-p + 6 CNE + Pac	2089.4
2130.8	p-TCATG-p + 6 CNE + ⁱ PrPac (or Pac + Ac)	2131.4
<p>Note: All species shown here are “d” fragments. The one-letter code refers to unprotected nucleosides (obviously linked by phosphate diesters). Additional phosphates and protecting groups, if found, are indicated.</p>		

Ions corresponding to the peptide-oligonucleotide conjugate found during MALDI-TOF MS analysis of the resin-linked conjugate.		
m/z found [M-H]⁻	Structure	Mass calculated for M
1322.7	Peptide-T + 2 CNE + Pac	1323.6
1519.6	Peptide-TC + 2 CNE + Ac	1520.6
1572.6	Peptide-TC + 3 CNE + Ac	1573.7
1611.8	Peptide-TC + 2 CNE + Pac	1612.6
1664.9	Peptide-TC + 3 CNE + Pac	1665.7
1706.9	Peptide-TC + 3 CNE + Pac + Ac	1707.7
2335.1	Peptide-TCAT + 4 CNE + Pac	2336.0
2388.1	Peptide-TCAT + 5 CNE + Pac	2389.1
2430.1	Peptide-TCAT + 5 CNE + Pac + Ac	2431.1
<p>Note: All species shown here are “d” fragments. The one-letter code refers to unprotected nucleosides (obviously linked by phosphate diesters). Additional phosphates and protecting groups, if found, are indicated. The 6-membered ring resulting from the DA cycloaddition is not indicated.</p>		

E.2.2.6 Assessment of the fragmentation in the conjugation reaction conditions

1. Effect of the pH

Resin-linked, diene-derivatised oligonucleotide **2.41** was incubated at pH 5.5, 7.0 and 8.3 for 24 h at 65 °C (using phosphate buffers). After this the resin was thoroughly washed with water and ACN before being incubated in conc. aqueous ammonia at room temperature for 2 h. Crudes were analysed by HPLC (analysis conditions set 3, 5-30 % B) and the extent of the fragmentation side-reactions determined from the integration of the HPLC traces at 260 nm. The outcome of each experiment is summarised in the following table.

pH	Percentage of fragmentation products in the crude
5.5	21
7.0	15
8.3	25

2) Effect of the solid support on the fragmentation process

A solution (0.75 mM) of **2.43** was incubated for 24 h at 65 °C both in the presence or the absence of CPG resin beads. After this, both solutions were analysed by HPLC (analysis conditions set 3, 5-15 % B) and the extent of oligonucleotide fragmentation determined as above. The outcome of each experiment is summarised in the following table.

CPG resin beads	Percentage of fragmentation products in the crude
Yes	30
No	55

3) Effect of the sequence

The sequence dependence in the oligonucleotide backbone fragmentation was studied by independently incubating **2.44** and **2.41** in water at 65 °C for 24 h. After this time both resins were subjected to conc. aqueous ammonia for 2 h at room temperature and the resulting crudes analysed by HPLC (analysis conditions set 3, 5-30 % B) . Extent of the fragmentation was determined as above, and the results are shown below.

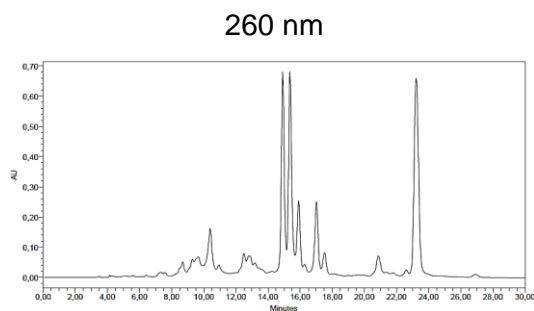
Product	Percentage of fragmentation products in the crude
2.41	27
2.44	19

4) Effect of lower temperature on the reaction outcome

This time the conjugation was performed without any morpholine, TEA or DEA treatments, as the only goal was to study the effect that a lower conjugation temperature would have in the oligonucleotide backbone fragmentation. Hence, a large amount of acylated conjugate was found upon deprotection and cleavage from the resin. To simplify analysis of the reaction outcome acylated products have been included for yield determination.

Reagents: **2.41 + 2.14**
Product: **2.42**
Conjugation method: DA conjugation following method J
HPLC yield (260 nm): 69 %.
By-products (260 nm): 7 % of unreacted material, a 14 % of ON fragmentation,
7 % of hydrolysed conjugate.

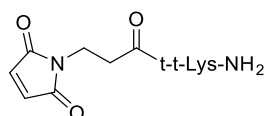
HPLC traces of the crude:



E.2. Characterisation of the products

Product: **2.5**

Structure:



HPLC characterisation

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 7.1$ min

MS characterisation

Technique: MALDI-TOF, positive mode, DHB as matrix

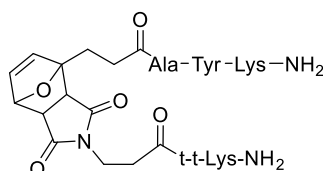
Mass calculated: C₃₅H₄₈N₁₂O₁₂ 828.4

Mass found: m/z 829.7 [M+H]⁺, 851.7 [M+Na]⁺

Purification conditions: Set 1, 10-40 % B, $t_R = 8.6$ min

Product: **2.6**

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 12.1+12.8+12.9$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

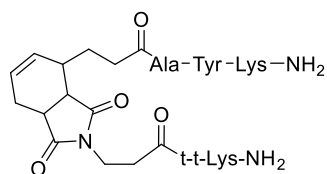
Mass calculated: C₆₁H₈₄N₁₆O₁₈ 1329.6

Mass found: m/z 1331.0 [M+H]⁺

Purification conditions: -

Product: **2.7**

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 15.0$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

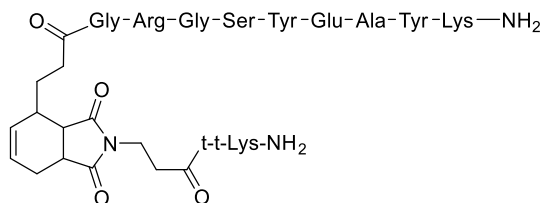
Mass calculated: $C_{61}H_{86}N_{16}O_{17}$ 1315.6

Mass found: m/z 1317.0 $[M+H]^+$

Purification conditions: -

Product: **2.9**

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 16.3$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

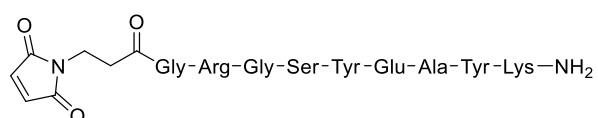
Mass calculated: $C_{87}H_{124}N_{26}O_{27}$ 1964.9

Mass found: m/z 1966.6 $[M+H]^+$

Purification conditions: -

Product: **2.11**

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 0-30 % B, $t_R = 20.0$ min;

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

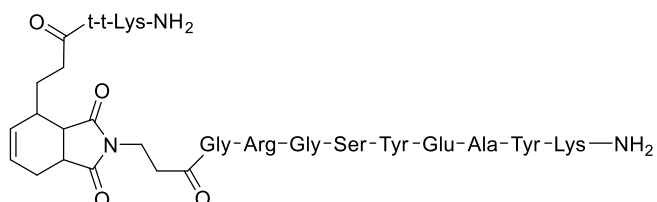
Mass calculated: $C_{67}H_{88}N_{28}O_{22}$ 1636.6

Mass found: m/z 1660.5 $[M+Na]^+$

Purification conditions: Set 1, 10-20 % B, $t_R = 12.1$ min

Product: **2.12**

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 16.3$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

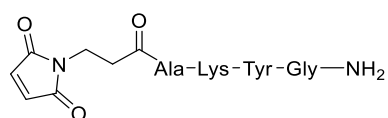
Mass calculated: $C_{87}H_{124}N_{26}O_{27}$ 1964.9

Mass found: m/z 1966.5 $[M+H]^+$

Purification conditions: -

Product: **2.14**

Structure:



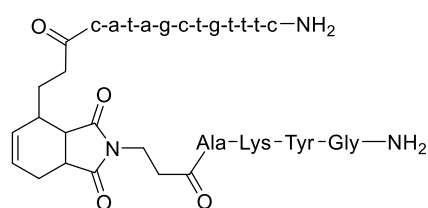
2.14

Reference: 5

Product:

2.15

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-20 % B, $t_R = 21.5$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: C₁₆₃H₂₀₉N₇₁O₄₈ 3928.6Mass found: m/z 3929.6 [M+H]⁺

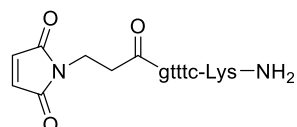
Purification conditions:

-

Product:

2.16

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 0-30 % B, $t_R = 20.0$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: C₆₇H₈₈N₂₈O₂₂ 1636.6Mass found: m/z 1660.5 [M+Na]⁺

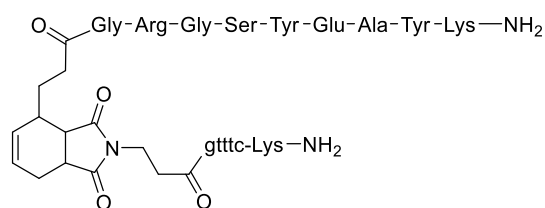
Purification conditions:

Set 1, 10-20 % B, $t_R = 12.1$ min

Product:

2.17

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 15.4$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: $C_{119}H_{164}N_{42}O_{37}$ 2773.2

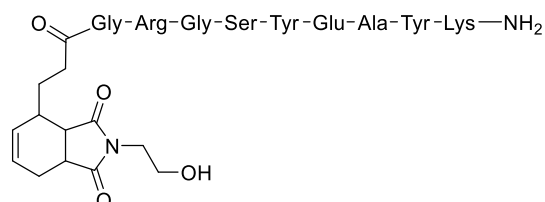
Mass found: m/z 2774.1 $[M+H]^+$

Purification conditions: -

Product:

2.19

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 17.2+17.4$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: $C_{58}H_{83}N_{15}O_{18}$ 1277.6

Mass found: For both peaks, m/z 1279.2 $[M+H]^+$

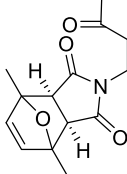
Purification conditions: -

Product:

2.20

Structure:

Ac—Ser—Lys—Tyr—Gly—OH



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 5-50 % B, $t_R = 12.4$ min

MS characterisation:

Technique: ESI, positive mode

Mass calculated: $C_{35}H_{46}N_6NaO_{12}$ 765.3

Mass found: m/z 765.3 $[M+Na]^+$

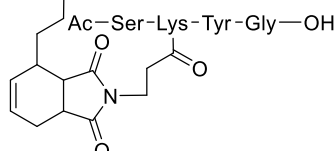
Purification conditions: Set 2, 15-60 % B, $t_R = 7.2$ min

Product:

2.21

Structure:

Gly—c-a-t-a-g-c-t-g-t-t-t-c—NH₂



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-40 % B, $t_R = 15.9$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, THAP as matrix

Mass calculated: $C_{167}H_{213}N_{71}O_{52}$ 4044.6

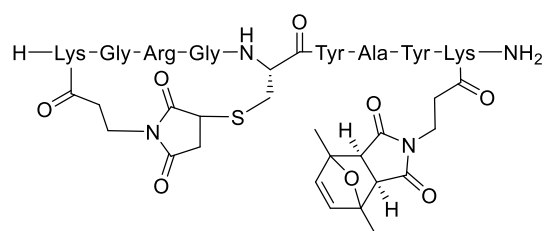
Mass found: m/z 4046.3 $[M+H]^+$, 4068.3 $[M+Na]^+$

Purification conditions: Set 2: 10-50 % B, $t_R = 13.0$ min

Product:

2.22

Structure:



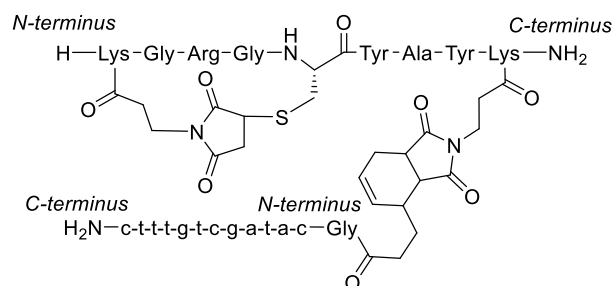
Reference:

6

Product:

2.23

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-50 % B, $t_R = 13.6$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: $C_{198}H_{258}N_{82}O_{58}S$ 4744.0

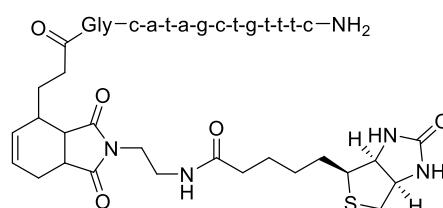
Mass found: m/z 4745.9 $[M+H]^+$, 4767.9 $[M+Na]^+$

Purification conditions:

Set 2: 10-60 % B, $t_R = 11.7$ min

Product: **2.26a**

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 19.8$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

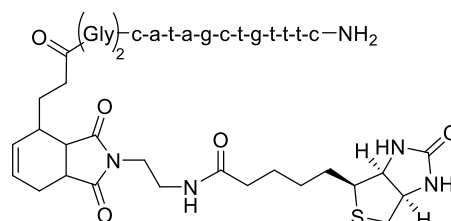
Mass calculated: C₁₅₄H₁₉₇N₆₉O₄₅S 3764.5

Mass found: m/z 3766.6 [M+H]⁺, 3788.6 [M+Na]⁺

Purification conditions: -

Product: **2.26b**

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 19.8$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: C₁₅₆H₂₀₀N₇₀O₄₆S 3821.5

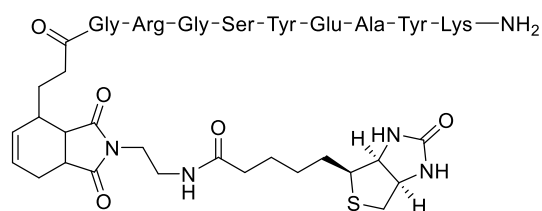
Mass found: 3822.5

Purification conditions: -

Product:

2.27

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-30 % B, $t_R = 22.0$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: C₆₈H₉₈N₁₈O₁₉S 1502.7

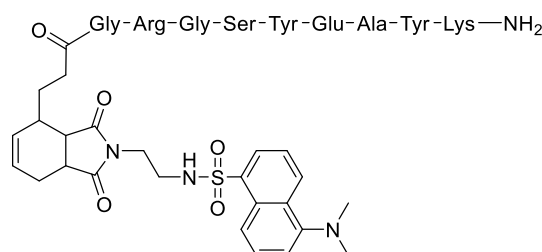
Mass found: m/z 1504.1 [M+H]⁺

Purification conditions: -

Product:

2.28

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-50 % B, $t_R = 19.6$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: C₇₀H₉₅N₁₇O₁₉S 1509.7

Mass found: m/z 1510.9 [M+H]⁺, 1532.9 [M+Na]⁺

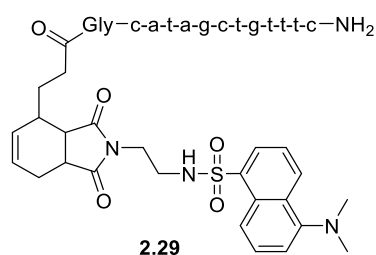
Purification conditions:

Set 2 20-60 % B, $t_R = 11.2$ min

Product:

2.29

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-50 % B, t_R = 16.7 min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: C₁₅₆H₁₉₄N₆₈O₄₅S 3771.5

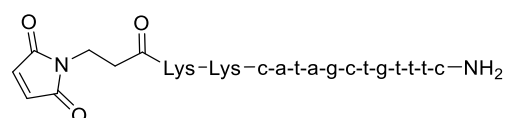
Mass found: m/z 3773.1 [M+H]⁺, 3795.1 [M+Na]⁺

Purification conditions: Set 2: 10-60 % B, t_R = 13.9 min

Product:

2.32

Structure:



HPLC characterisation:

Analysis conditions: Set 1

Gradient and retention time: 10-40 % B, t_R = 12.2 min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

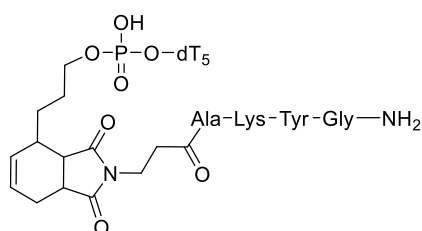
Mass calculated: C₁₄₈H₁₉₃N₆₉O₄₄ 3642.5

Mass found: m/z 3642.1 [M+H]⁺

Purification conditions: Set 2, 10-40 % B, t_R = 11.8 min

Product: **2.37**

Structure:



HPLC characterisation:

Analysis conditions: Set 2

Gradient and retention time: 5-35 % B, $t_R = 16.2 + 16.6$ min

MS characterisation:

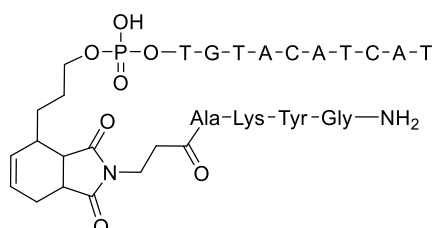
Technique: MALDI-TOF, negative mode, THAP/AC as matrix

Mass calculated: $C_{84}H_{114}N_{17}O_{44}P_5$ 2219.6

Mass found: m/z 2218.9 [M-H]⁻

Product: **2.39**

Structure:



HPLC characterisation:

Analysis conditions: Set 2

Gradient and retention time: 5-25 % B, $t_R = 19.3 + 19.9$ min

MS characterisation:

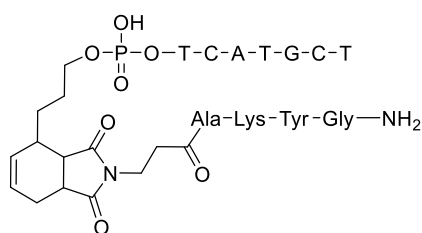
Technique: MALDI-TOF, negative mode, THAP/AC as matrix

Mass calculated: $C_{132}H_{173}N_{41}O_{70}P_{10}$ 3763.8

Mass found: m/z 3760.8

Product: **2.42**

Structure:



HPLC characterisation:

Analysis conditions: Set 3

Gradient and retention time: 5-30 % B, $t_R = 15.2 + 15.7$ min

MS characterisation:

Technique: MALDI-TOF, negative mode, THAP/AC as matrix

Mass calculated: C₁₀₂H₁₃₆N₂₉O₅₃P₇ 2831.7

Mass found: m/z 2831.2 [M-H]⁻

Product: **2.43**

Structure: HO-T-C-A-T-G-C-T

HPLC characterisation:

Analysis conditions: Set 3

Gradient and retention time: 5-15 % B, $t_R = 15.6$ min

MS characterisation:

Technique: MALDI-TOF, negative mode, THAP/AC as matrix

Mass calculated: C₆₈H₈₇N₂₂O₄₂P₆ 2069.4

Mass found: m/z 2070.0 [M-H]⁻

E.2 Abbreviations

AC: Ammonium Citrate

ACN: Acetonitrile

DA: Diels-Alder

DCM: Dichloromethane

DEA: Diethylamine

DHB: 2,5-Dihydroxybenzoic acid

DMF: N,N-Dimethylformamide

DMSO: Dimethyl sulfoxide

HPLC/MS: High Pressure Liquid Chromatography/Mass Spectrometry

HPLC: High Pressure Liquid Chromatography

MALDI-TOF: Matrix-Assisted Laser Desorption/Ionization – Time-Of-Flight

MS: Mass Spectrometry

Pac: phenoxyacety

PNA: Peptide Nucleic Acid

TEA: Triethylamine

TEAA: Triethylammonium acetate

TFA: Trifluoroacetic acid

THAP: 2,4,6-trihydroxyacetophenone

TIS: Triisopropylsilane

E.2 Bibliography

- (1) Meienhofer, J.; Waki, M.; Heimer, E. P.; Lambros, T. J.; Makofske, R. C.; Chang, C. D. *Int. J. Pept. Protein Res.* **1979**, *13* (1), 35.
- (2) Baillie, L. C.; Batsanov, A.; Bearder, J. R.; Whiting, D. A. *J. Chem. Soc. [Perkin 1]* **1998**, No. 20, 3471.
- (3) Ware, R. W.; Day, C. S.; King, S. B. *J. Org. Chem.* **2002**, *67* (17), 6174.
- (4) Meyer, A.; Spinelli, N.; Imbach, J.-L.; Vasseur, J.-J. *Rapid Commun. Mass Spectrom.* **2000**, *14* (4), 234.
- (5) Elduque, X.; Sánchez, A.; Sharma, K.; Pedroso, E.; Grandas, A. *Bioconjug. Chem.* **2013**, *24* (5), 832.
- (6) Elduque, X.; Pedroso, E.; Grandas, A. *J. Org. Chem.* **2014**, *79* (7), 2843.

Experimental section: chapter 3

E.3 General methods

HPLC:

Analysis conditions

Shimadzu instrument, Jupiter Proteo column (4 μm , 90 \AA , 250 \times 4.6 mm) from Phenomenex. Linear gradients of 30 min were always used unless otherwise stated. Solvent A: 0.045 % TFA in water, solvent B: 0.036 % TFA in ACN, flow: 1 mL/min. Detection wavelength: 200-400 nm. All analysis were performed at room temperature unless otherwise indicated.

Purification conditions

Jupiter Proteo column (10 μm , 90 \AA , 250 \times 10.0 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 % TFA in water, solvent B: 0.1 % TFA in ACN, flow: 3 mL/min. Detection wavelength: 200-400 nm. All purifications were performed at room temperature unless otherwise indicated.

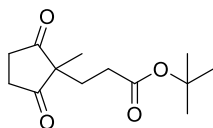
HPLC/MS analysis conditions:

Waters 2695 instrument coupled to a Waters Micromass ZQ 4000 mass spectrometer, Jupiter Proteo column (4 μm , 90 \AA , 250 \times 4.6 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 % formic acid in water, solvent B: 0.1 % formic acid in ACN, flow: 1 mL/min. Detection wavelength: 200-600 nm.

Important note: To ease interpretation and discussion of the results obtained, the major compounds appearing in each HPLC or HPLC/MS trace shown have been labelled. Retention times indicated are those obtained in the experiment shown, and might slightly vary from other experiments with the same products and analysis conditions due to small differences in the eluent composition, fluctuations in the temperature of the lab, etc. The labelling system employed for HPLC traces is different from that used for HPLC/MS traces, for no other reason than the smaller size of the numbers in the HPLC/MS traces. For a detailed characterisation of the numbered products see section **Product characterisation**. The same applies to the experimental section corresponding to chapter 4.

E.3.1 Background and objectives

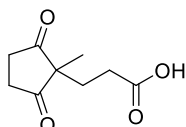
tert-Butyl 3-(1-methyl-2,5-dioxocyclopentyl)propanoate



2-Methylcyclopentane-1,3-dione (1.22 g, 10.9 mmol) and *tert*-butyl acrylate (5.9 mL, 26.8 mmol, 2.5 equiv.) were dissolved in 20.0 mL of triethylamine. The mixture was refluxed overnight with fast stirring. Triethylamine and excess *tert*-butyl acrylate were removed in vacuo to render the title compound (2.50 g, 78%) as a dark brown solid that was no further purified.

^1H NMR (400 MHz, CDCl_3) δ : 2.87 – 2.70 (m, 4H), 2.20 (t, J = 7.5 Hz, 2H), 1.92 (t, J = 7.6 Hz, 2H), 1.40 (s, 9H), 1.12 (s, 3H) ppm. ^{13}C NMR (101 MHz, CDCl_3) δ : 215.9, 172.3, 81.1, 55.4, 35.0, 30.0, 29.3, 28.2, 19.8 ppm. HRMS (ESI, positive mode): m/z 241.1446 $[\text{M}+\text{H}]^+$, M calcd. for $\text{C}_{13}\text{H}_{20}\text{O}_4$ 240.1362.

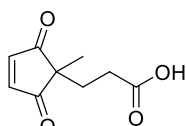
3-(1-Methyl-2,5-dioxocyclopentyl)propanoic acid



tert-Butyl 3-(1-methyl-2,5-dioxocyclopentyl)propanoate (1.0 g, 4.16 mmol) was dissolved in 26.0 mL of a 1:1 TFA/DCM mixture and stirred for 2 h at 25 °C. Toluene (10-15 mL) was added to the reaction mixture, which was concentrated and dried in vacuo to give the free acid as a brown solid (760 mg, 99%). No purification was necessary.

^1H NMR (400 MHz, CD_3OD) δ : 2.78 (s, 4H), 2.26 (t, J = 8.0 Hz, 2H), 1.89 (t, J = 8.0 Hz, 2H), 1.10 (s, 3H) ppm. ^{13}C NMR (101 MHz, CD_3OD) δ : 217.9, 176.4, 56.5, 35.6, 30.2, 29.5, 19.6 ppm. HRMS (ESI, negative mode): m/z 183.0661 $[\text{M}-\text{H}]^-$, M calcd. for $\text{C}_9\text{H}_{12}\text{O}_4$ 184.0736.

3-(1-Methyl-2,5-dioxocyclopent-3-en-1-yl)propanoic acid (3.1)



3-(1-Methyl-2,5-dioxocyclopentyl)propanoic acid (500 mg, 2.72 mmol) was dissolved in 20.0 mL of ethyl acetate, and CuCl_2 (803 mg, 5.98 mmol, 2.2 equiv.) and LiCl (253 mg, 5.98 mmol, 2.2 equiv.) were added to the mixture. After refluxing overnight, ca. 20 mL of ethyl acetate were added to the crude, and the resulting mixture was subsequently

washed with 5 % aqueous HCl (ca. 40 mL × 3). The organic fraction was dried over anhydrous MgSO₄, and the solvent removed in vacuo. The resulting crude was purified by silica gel column chromatography (98:2 DCM/AcOH, isocratic), and a yellow powder was obtained (240 mg, 48%).

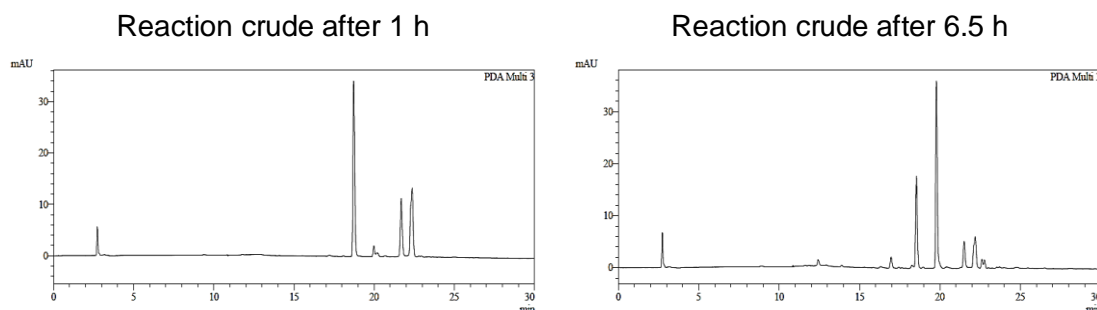
Note: sometimes it is advisable to re-extract the aqueous phase with ethyl acetate (once or twice) to recover small amounts of **3.1** that may have remained at the aqueous phase. ¹H NMR (400 MHz, CDCl₃) δ: 7.25 (s, 2H), 2.33 – 2.24 (m, 2H), 2.01 – 1.93 (m, 2H), 1.18 (s, 3H) ppm. ¹³C NMR (101 MHz, CDCl₃) δ: 206.9, 176.9, 148.1, 49.2, 28.8, 28.5, 18.6 ppm. HRMS (ESI, negative mode): *m/z* 181.0500 [M-H]⁻, M calcd. for C₉H₁₀O₄ 182.0579.

E.3.2.1 Reaction of CPDs with peptides containing internal or C-terminal cysteines

1) Reaction between **3.1** and **3.2** to furnish **3.4**

H-Lys-Tyr-Ala-Tyr-Cys-Gly-NH₂ (**3.2**, 20 nmol) was reacted with 5 equiv. of **3.1** in 0.1 mM phosphate buffer (pH=7.8, 0.1 mM peptide concentration) at room temperature for 2 h, after which time the mixture was stirred for an additional 4.5 h at 45 °C. Reaction progress was followed by HPLC, analysing aliquots removed at 1 and 6.5 h reaction times.

HPLC details: 0-40 % B, traces shown at 280 nm



Product retention times (min):

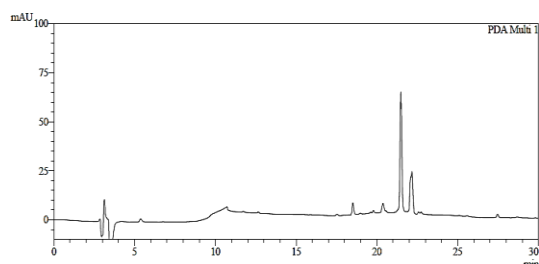
3.1: not detected at this wavelength; **3.2**: 18.6; dimer of **3.2**: 19.8; **3.4**: 21.5 + 22.2

2) Reversion of **3.4** in water

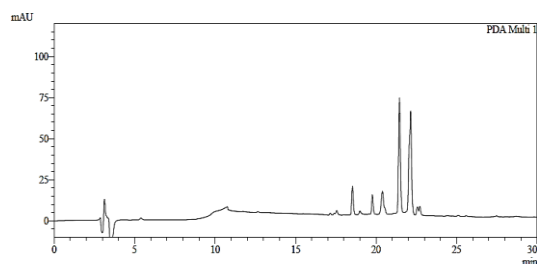
Both isomers of **3.4** (referred to as isomer 1 or isomer 2 according to their elution order in the HPLC) were independently collected from the HPLC and lyophilised. After this, water (300 µL) was added to each isomer and the resulting solutions were analysed by HPLC immediately after being prepared and after a 4.5 h incubation at room temperature.

HPLC details: 0-40 % B, traces shown at 220 nm

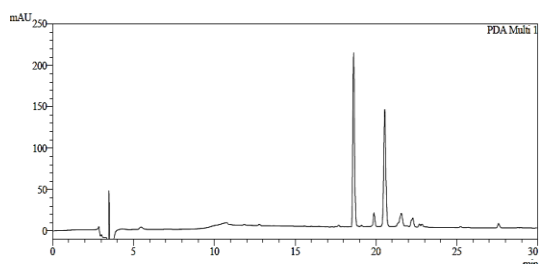
Isomer 1 immediately after redissolution



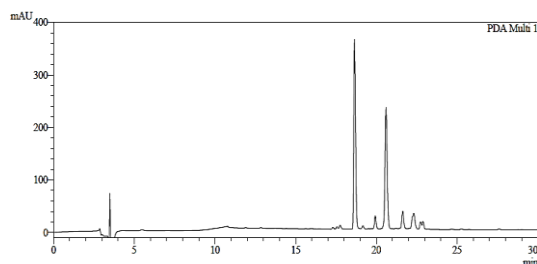
Isomer 2 immediately after redissolution



Isomer 1 after 4.5 h at room temperature



Isomer 2 after 4.5 h at room temperature



Product retention times (min):

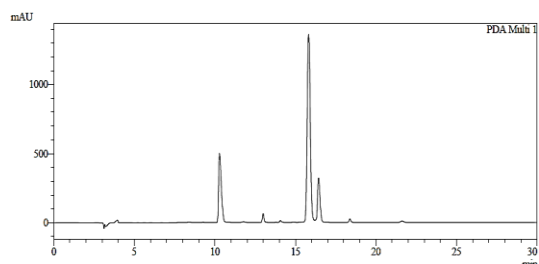
3.1: 20.4; **3.2:** 18.6; dimer of **3.2:** 19.8; **3.4:** 21.5 + 22.2

3) Reaction between 3.6 and 3.2 to furnish 3.7

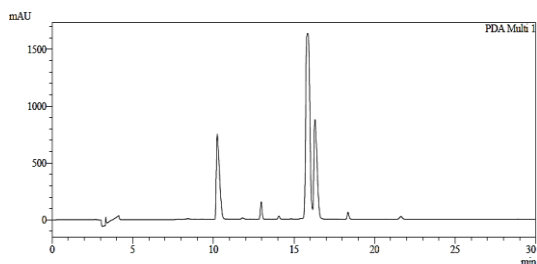
Peptide **3.2** (50 nmol, 0.1 mM concentration) was reacted with 5 equiv. of **3.6** in water at 37 °C. Reaction crude was analysed by HPLC after 15 min, 1 h and 3 h.

HPLC details: 10-40 % B, traces shown at 220 nm

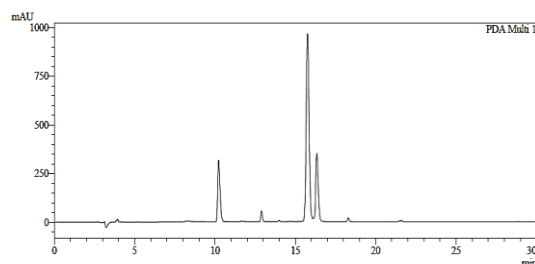
Reaction crude after 15 min



Reaction crude after 1 h



Reaction crude after 3 h



Product retention times (min):

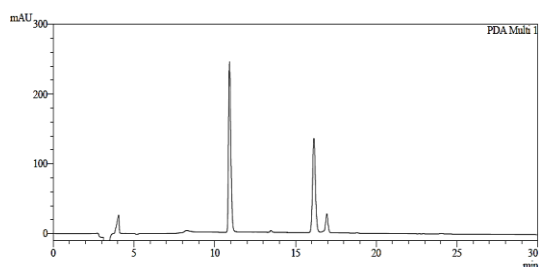
3.6: 15.5; **3.2:** 10.3; **3.7:** 16.5

4) Reversion of 3.7 in water

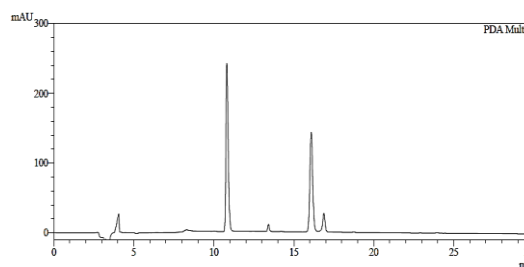
3.7 Was collected from the HPLC and lyophilised. After this, water (400 μ L) was added and the resulting solution was incubated at 37 $^{\circ}$ C and analysed by HPLC after 1, 3 and 6 h.

HPLC details: 10-40 % B, traces shown at 220 nm

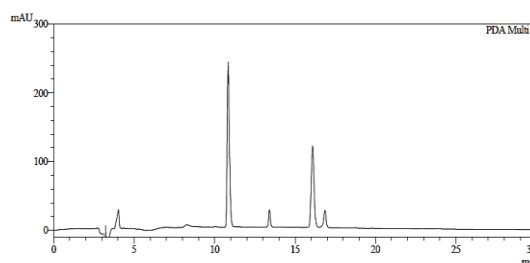
Product after 1 h



Product after 3 h



Product after 6 h



Product retention times (min):

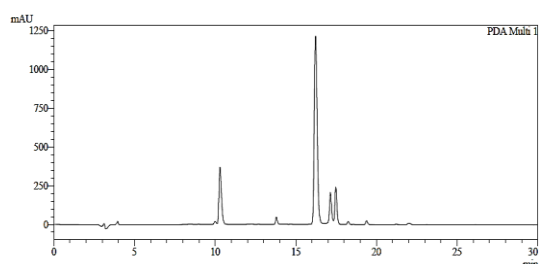
3.6: 16.0; **3.2:** 10.3; **3.7:**17.0

5) Reaction between 3.6 and 3.5 to furnish 3.8

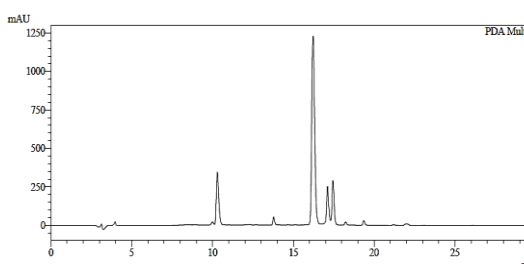
Peptide **3.5** (50 nmol, 0.1 mM concentration) was reacted with 5 equiv. of 2,2-dimethyl-4-cyclopentene-1,3-dione (**3.6**) in water at 37 $^{\circ}$ C. Reaction progress was followed by HPLC, analysing aliquots removed at 15 min, 1 h and 3 h.

HPLC details: 10-40 % B, traces shown at 220 nm

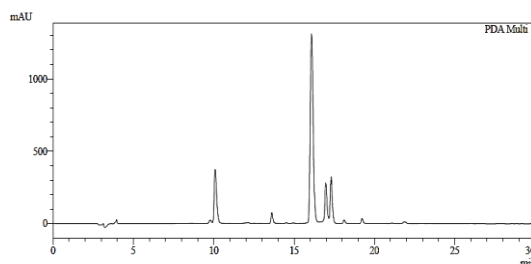
Reaction crude after 15 min



Reaction crude after 1 h



Reaction crude after 3 h



Product retention times (min):

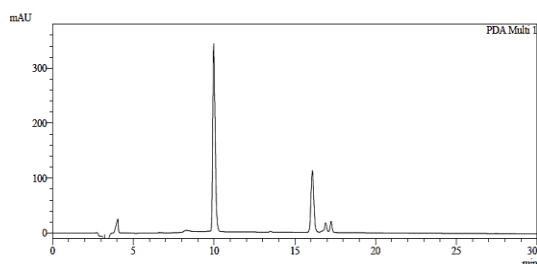
3.6: 16.0; **3.5:** 10.3; **3.8:** 17.0 + 17.4

6) Reversion of 3.8 in water

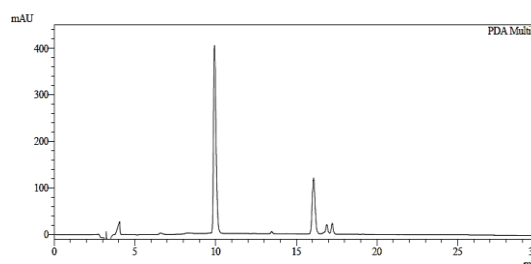
Both isomers of **3.8** were collected together from the HPLC and lyophilised. After this, water (400 μ L) was added to the isomer mixture and the resulting solution was incubated at 37 $^{\circ}$ C and analysed by HPLC after 1, 3 and 6 h.

HPLC details: 10-40 % B, traces shown at 220 nm

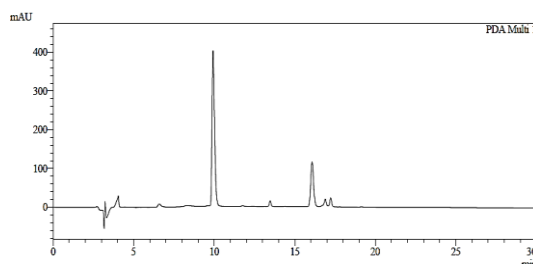
Product after 1 h



Product after 3 h



Product after 6 h



Product retention times (min):

3.6: 16.0; **3.5:** 10.0; **3.8:** 16.9 + 17.2

E.3.2.2 Reaction of CPDs with peptides containing *N*-terminal cysteines

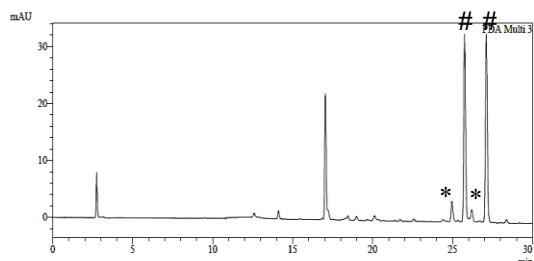
1) Reaction between 3.3 and 3.1

H-Cys-Tyr-Gly-NH₂ (**3.3**, 40 nmol) was reacted with 5 equiv. of **3.1** in 0.1 mM phosphate buffer (pH=7.8, 0.08 mM peptide concentration) at room temperature for 2 h, after which time the crude was analysed by HPLC. 4 New peaks appeared whose ratio was shown

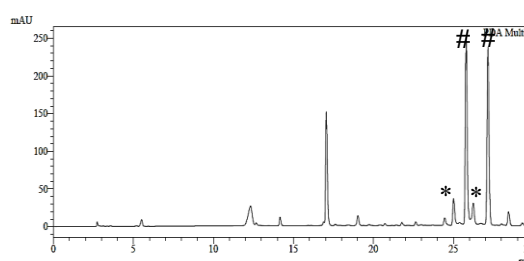
to change with time: the areas of the peaks eluting at 25.0 and 26.2 min were observed to increase while those of the other 2 peaks decreased. The 4 new peaks were collected after 6 h reaction time, lyophilised and characterised by MALDI-TOF MS.

HPLC details: 0-40 % B, traces shown at 280 nm

Reaction crude after 2 h



Reaction crude after 6 h



8:92 ratio peaks of set */peaks of set #

12:88 ratio peaks of set */peaks of set #

Product retention times (min):

3.1: not detected at this wavelength; **3.3**: completely consumed; dimer of **3.3**

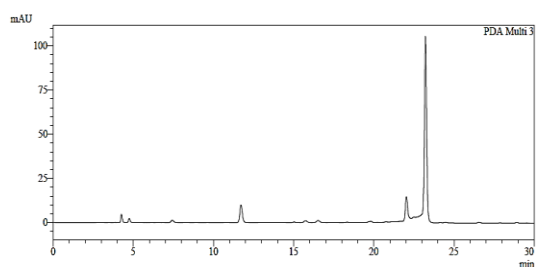
(already present in the starting material): 17.0

2) Reaction of 3.6 with 3.3

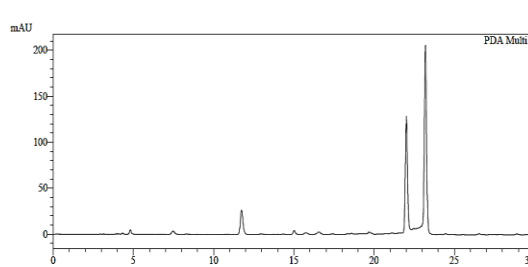
Peptide H-Cys-Tyr-Gly-NH₂ (**3.3**, 100 nmol) was reacted with 5 equiv. of **3.6** in water (0.2 mM peptide concentration) at 37 °C, and the crude analysed by HPLC at 15 min, 1 h, 3 h, 6 h and at 24 h reaction times.

HPLC details: 0-40 % B, traces shown at 280 nm

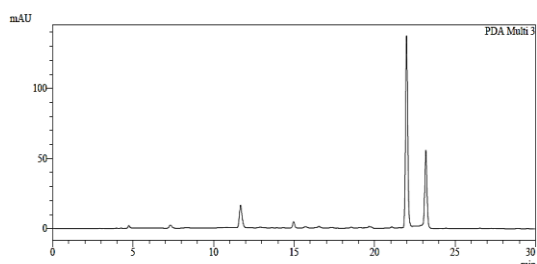
Reaction crude after 15 min



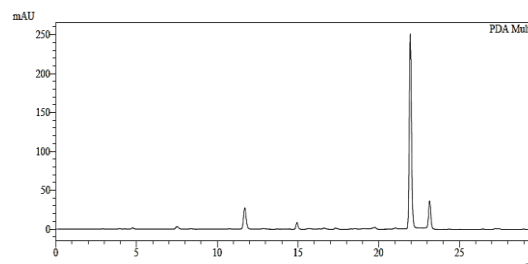
Reaction crude after 1 h



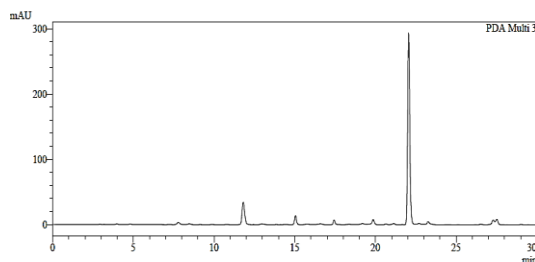
Reaction crude after 3 h



Reaction crude after 6 h



Reaction crude after 24 h



Product retention times (min):

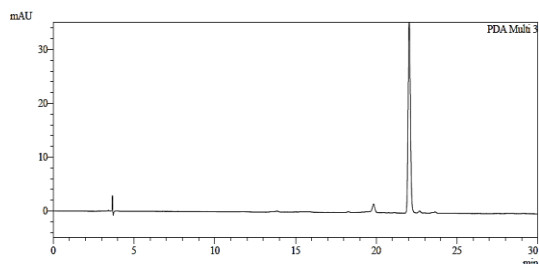
3.6 not detected at this wavelength; **3.3**: completely consumed; products: 22.4 + 23.6

3) Stability of the adducts generated after the 3.6 + 3.3 reaction

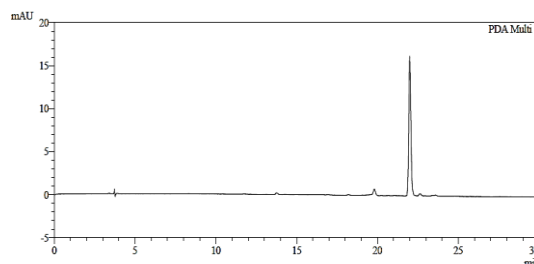
Peaks corresponding to the two newly generated products were independently collected, lyophilised, and dissolved in water. They were left with stirring at 37 °C and the peak eluting at 22.4 min analysed by HPLC after 15 h and 24 h, and the peak eluting at 23.6 min analysed after 15 min, 1 h, 6 h and 24 h.

HPLC details: 10-40 % B, traces shown at 220 nm

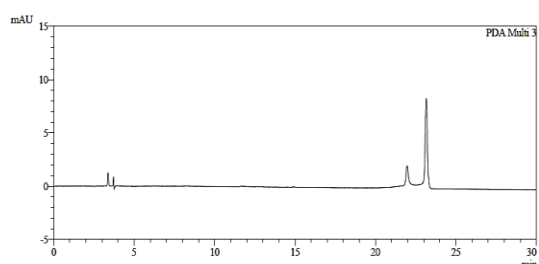
Peak eluting at 22.4 min after 15 h



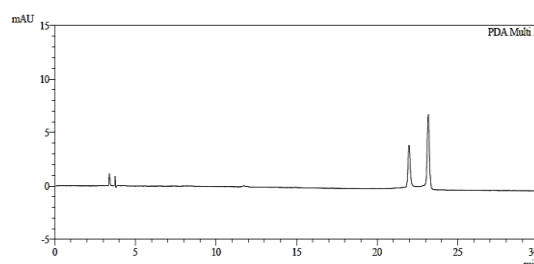
Peak eluting at 22.4 min after 24 h



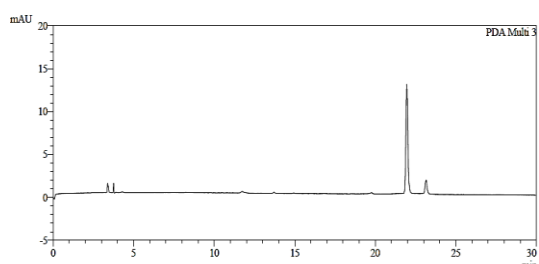
Peak eluting at 23.6 min after 15 min



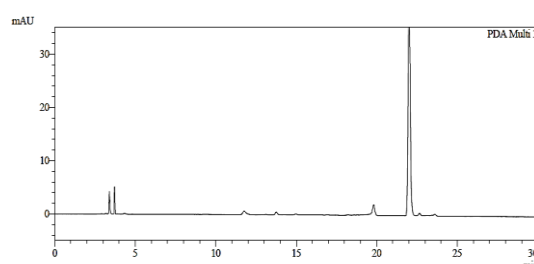
Peak eluting at 23.6 min after 1 h



Peak eluting at 23.6 min after 6 h

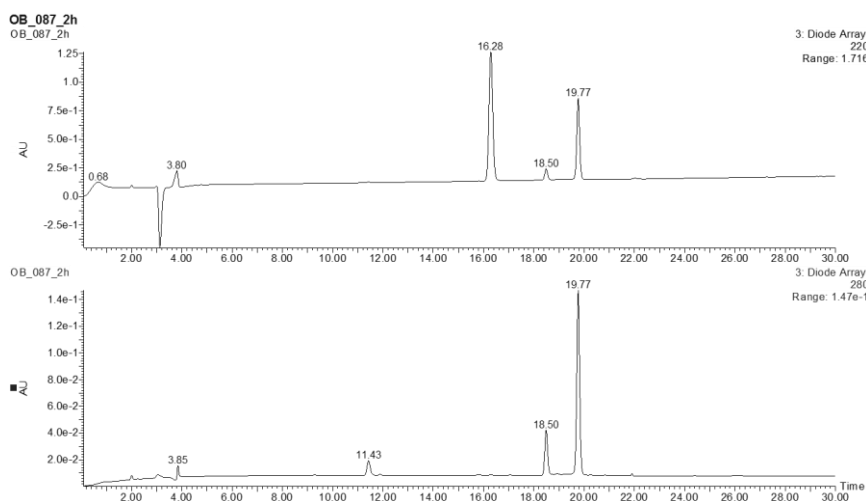


Peak eluting at 23.6 min after 24 h



The same experiment was repeated and the reaction mixture analysed after 2 h by HPLC/MS instead of carrying out HPLC analyses of the mixture and characterizing the different components by MALDI-TOF MS. It was then discovered that two different products were being formed (**3.9** and **3.10**).

HPLC/MS details: 10-40 % B, traces shown at 220 nm (top) and 280 nm (bottom)



Product retention times (min):

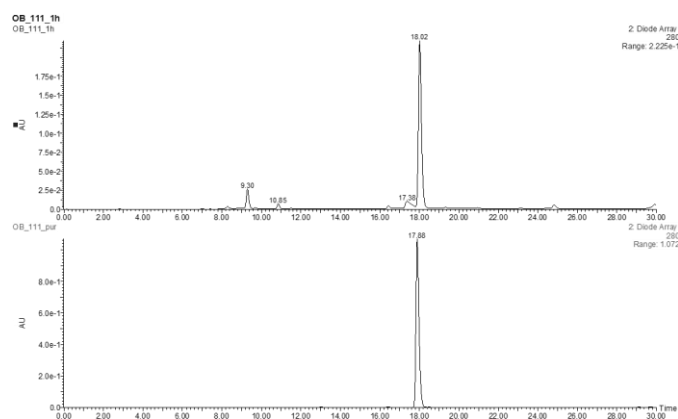
3.6: 16.3 (not observed at 280 nm); **3.3**: completely consumed; **3.9**: 18.5; **3.10**: 19.8

E.3.3 Structural determination of the M-20 Da adduct

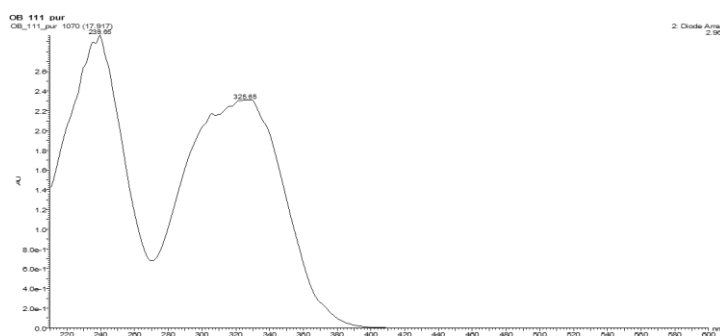
2,2-Dimethyl-4-cyclopentene-1,3-dione (**3.6**) (8.0 μ L, 0.071 mmol) was dissolved in 1.64 mL of ACN and mixed with an aqueous solution of methyl L-cysteinate hydrochloride (24.2 mg in 4.92 mL of water, 0.141 mmol, 2 equiv.). After 1 h stirring at 25 $^{\circ}$ C (the reaction had gone to completion, as assessed by HPLC/MS), the mixture was diluted with 60 mL of water, brought to pH=1 with 10 % aq HCl, and extracted with DCM (3 \times 60 mL). The organic phase was dried over anhydrous $MgSO_4$, and the solvent removed under reduced pressure. The resulting yellow-to-orange solid was purified by silica gel column chromatography eluting with DCM and increasing amounts of ethyl acetate (\rightarrow 40%) to render **3.11** as a pale yellow solid (7.2 mg, 42%).

1H NMR (400 MHz, $CDCl_3$) δ : 6.37 (s, 1H), 4.81 (dd, J = 6.9, 4.6 Hz, 1H), 3.83 (s, 3H), 3.13 – 3.04 (m, 2H), 1.25 (s, 3H), 1.24 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ : 205.0, 171.9, 170.2, 149.9, 129.7, 58.2, 53.1, 46.8, 24.2, 22.2, 21.9 ppm. HRMS (ESI, positive mode): m/z 240.0694 $[M+H]^+$, M calcd. for $C_{11}H_{13}NO_3S$ 239.0616. HPLC/MS: 10-50 % B, t_R = 17.9 min

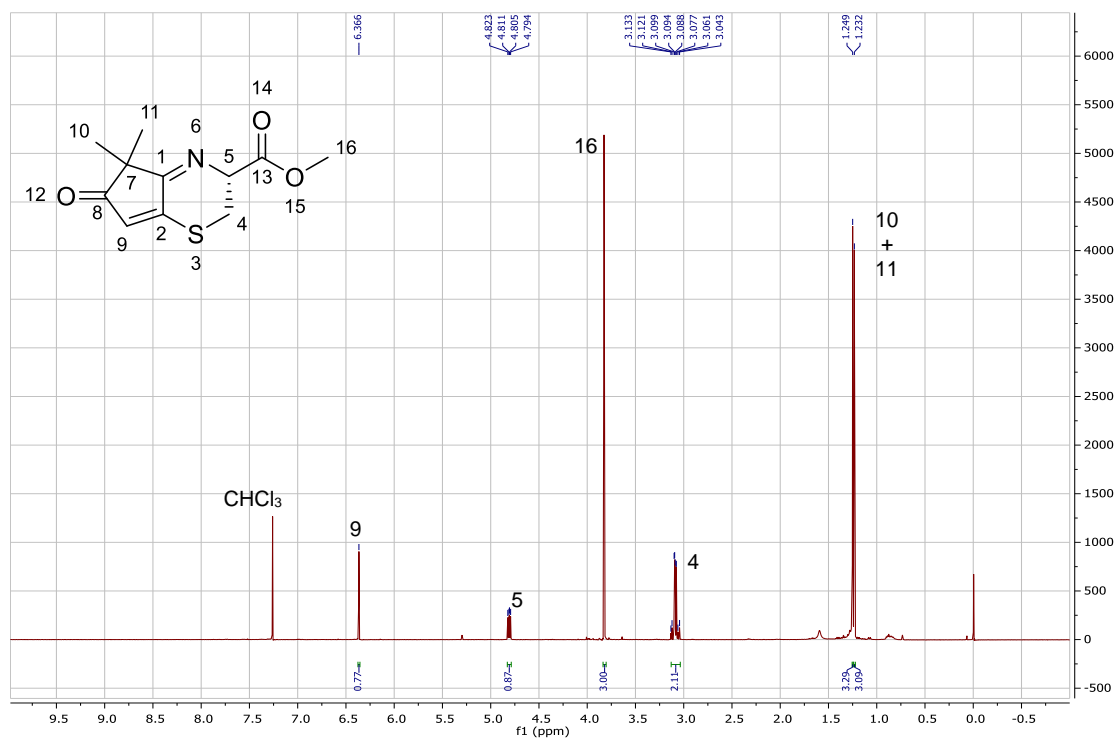
HPLC/MS details: 10-50 % B, traces shown at 280 nm. Top: reaction crude after 1 h;
bottom: pure **3.11**.



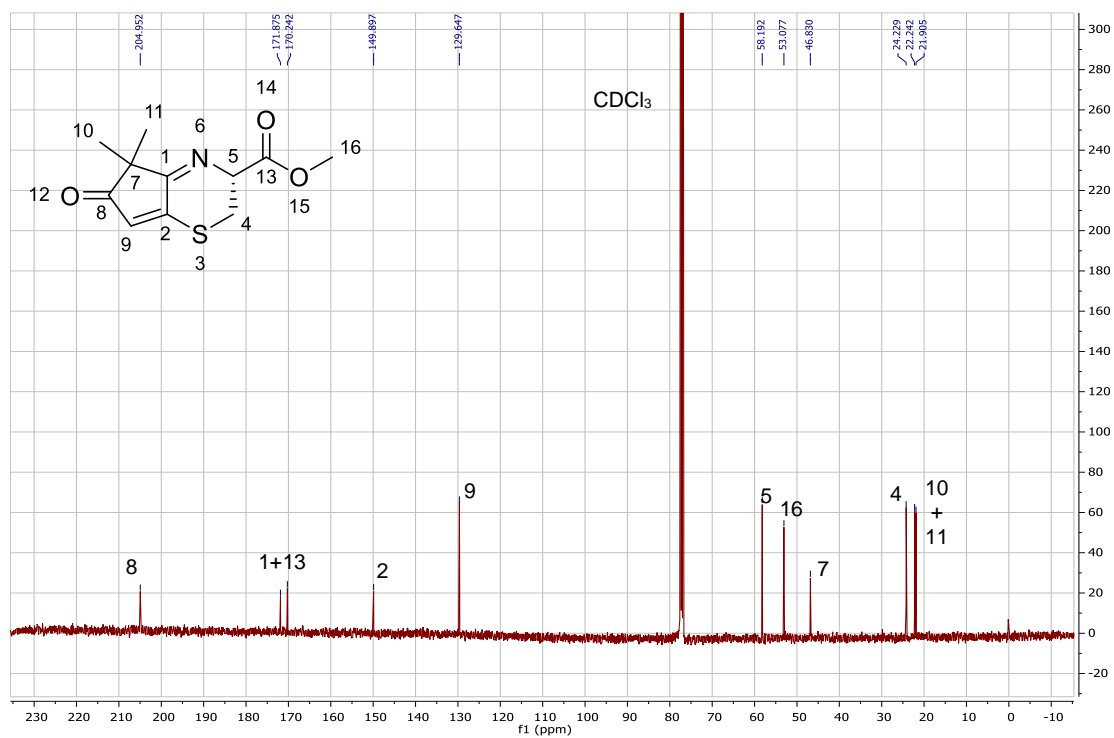
UV-Vis profile of the **M-20 Da** adduct



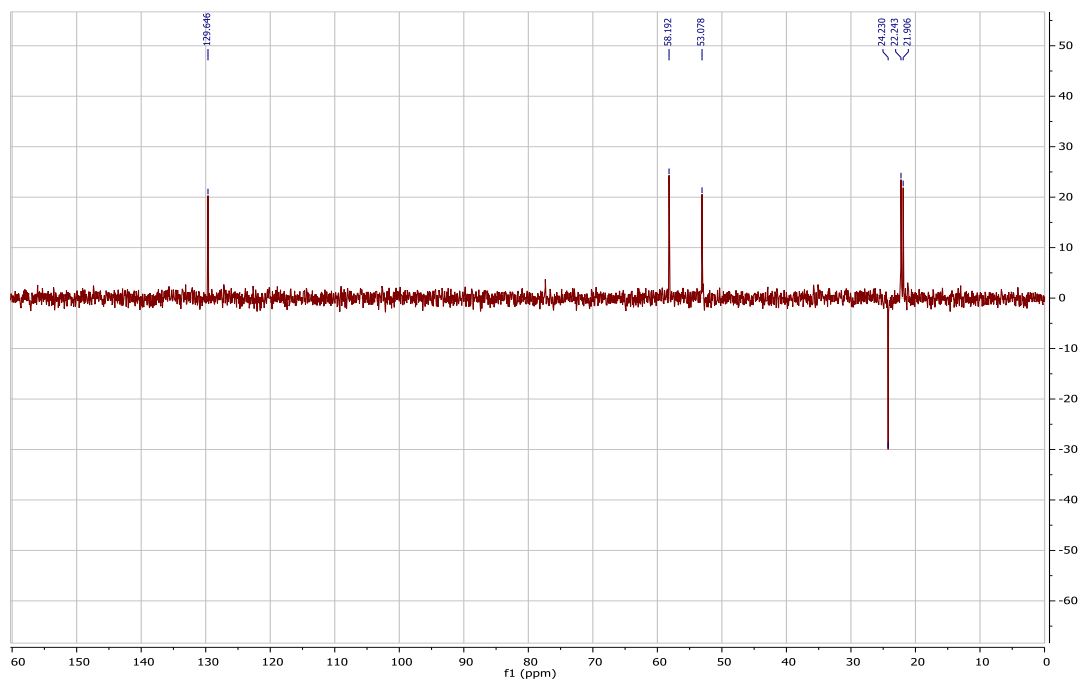
NMR Spectra:



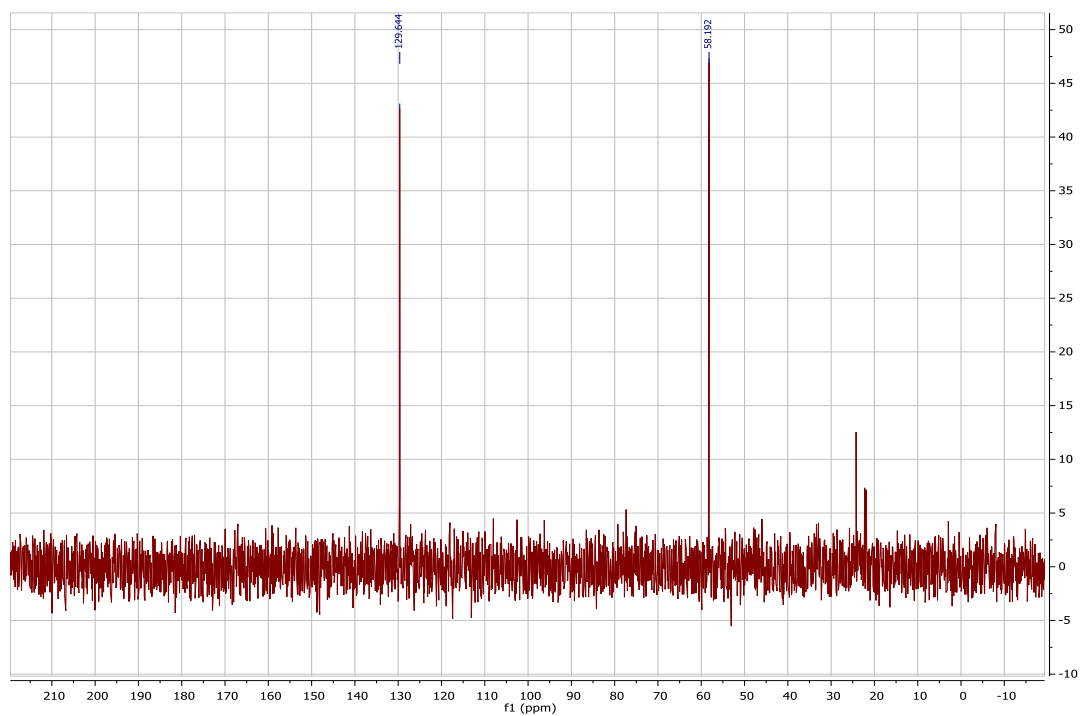
¹H NMR spectrum (400 MHz, CDCl₃)



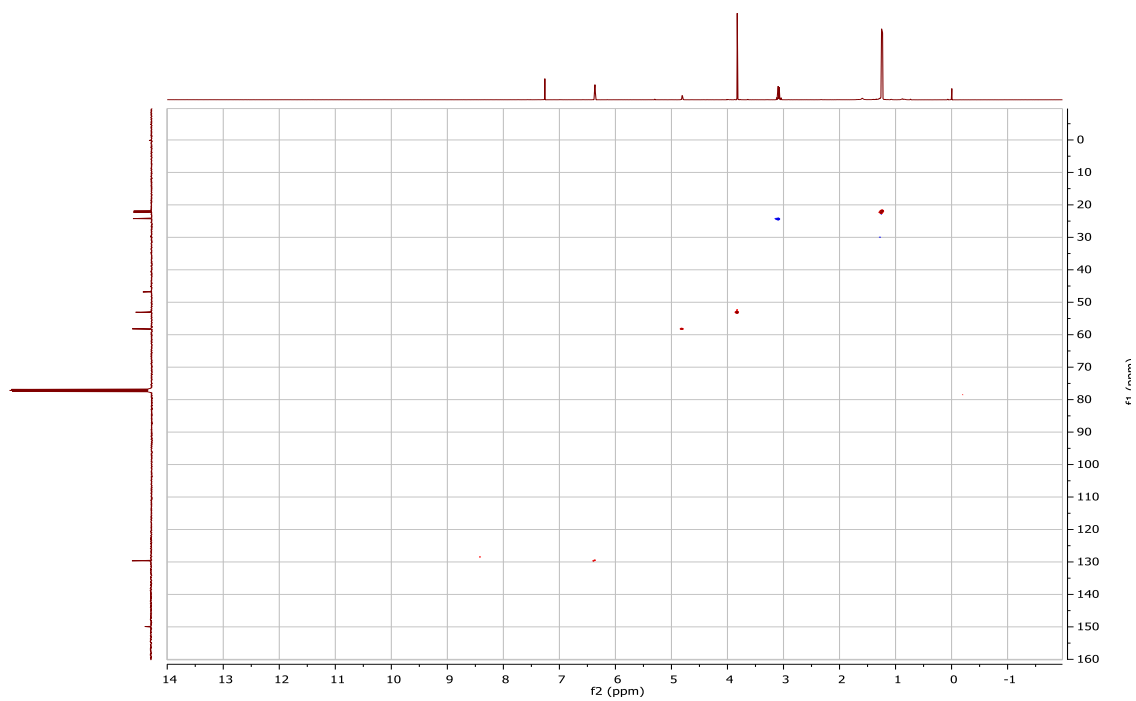
¹³C NMR spectrum (101 MHz, CDCl₃)



DEPT at 135 degrees spectrum (101 MHz, CDCl₃)



DEPT at 90 degrees spectrum (101 MHz, CDCl₃)



gHSQC spectrum (400 MHz, CDCl₃)

E.3.5 Screening of reaction conditions favouring the formation of the M-20 Da adduct

Due to the tendency of methyl cysteinate to dimerize, the concentration of free thiol in a stock solution of this compound was determined by the Ellman's test,¹ and this solution used within the next two days after its preparation (always stored in the freezer while not in use).

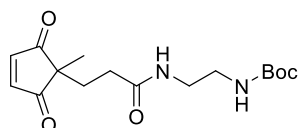
For the preparation of the reaction mixtures, aqueous solutions of methyl cysteinate (24.05 mM) and **3.1** (11.42 mM) were independently preincubated at the reaction temperature for 10 min. Afterwards, the required amounts of both reagents were mixed. Aliquots of the reaction mixtures (10 % of the total volume each) were withdrawn after 15, 60, 120, 240 and 360 min and stored in liquid nitrogen until being analysed by HPLC. Those samples containing an organic co-solvent were diluted with water until the organic solvent content was about 15% (in volume). The **M-20:M-18** relative ratios (**3.13** to **3.12**) were determined by HPLC (measuring the absorbance at 240 nm) and are shown in the following table. **3.1** Had completely reacted after 1 hour in all cases except when the reaction was performed with organic co-solvents or in pure water at 5 °C, where 2 hours were necessary for complete reaction.

Entry	[Methyl cysteinate] (mM)	[3.1] (mM)	Temp. (°C)	Solvent	3.13 to 3.12 relative ratio				
					Time (min)				
					15	60	120	240	360
1	2	2	37	H ₂ O	6:94	26:74	59:41	93:7	-
2	1	1	37	H ₂ O	4:96	25:75	52:48	87:13	100:0
3	0.5	0.5	37	H ₂ O	13:87	36:64	65:35	96:4	-
4	0.25	0.25	37	H ₂ O	19:81	49:51	82:18	98:2	-
5	1	1	60	H ₂ O	100:0	100:0	-	-	-
6	1	1	5	H ₂ O	18:82	21:79	37:63	69:31	83:17
7	1	1	37	H ₂ O/ACN 1:1	0:100	1:99	3:97	15:85	48:52
8	1	1	37	H ₂ O/MeOH 1:1	1:99	3:97	10:90	36:64	69:31
9	1	1	37	H ₂ O/DMSO 1:1	1:99	2:98	6:94	55:45	83:17
10	1	5	37	H ₂ O	17:83	92:8	100:0	-	-
11	1	3	37	H ₂ O	11:89	81:19	100:0	-	-
12	5	1	37	H ₂ O	5:95	20:80	54:46	95:5	-
13	3	1	37	H ₂ O	5:95	20:80	45:55	82:18	-

E.3.8 Reactions with other CPDs

1) Synthesis of the CPDs

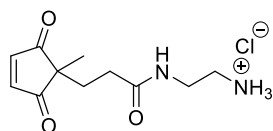
tert-Butyl (2-(3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanamido)ethyl)carbamate



A solution of **3.1** (50 mg, 0.274 mmol) in DCM (1.0 mL) was cooled to -15 °C. Subsequently, NMM (30 μ L, 0.274 mmol, 1.0 equiv.) and isobutyl chloroformate (35 μ L, 0.274 mmol, 1.0 equiv.) were added. After 10 minutes with heavy stirring, *tert*-butyl (2-aminoethyl)carbamate (48 mg, 0.301 mmol, 1.1 equiv.) was added. The suspension was kept at -15 °C for another 15 minutes and afterwards was reacted for 3 h at room temperature. After removing the solvent under reduced pressure, the crude was dissolved in EtOAc (20 mL) and extracted with water (2 \times 10 mL), 5% HCl (2 \times 10 mL) and 5% NaHCO₃ (2 \times 10 mL). The organic layer was dried over MgSO₄, filtered and evaporated under reduced pressure to afford the title compound (67 mg, 76% yield) as a yellow solid.

¹H NMR (CDCl₃, 400 MHz): δ 7.24 (s, 2H), 6.31 (bs, 1H), 5.05 (bs, 1H), 3.32-3.23 (m, 4H), 2.13- 2.09 (m, 2H), 1.99- 1.96 (m, 2H), 1.44 (s, 9H, CH₃), 1.16 (s, 3H, CH₃) ppm. ¹³C NMR (CDCl₃, 101 MHz): δ 206.9, 171.8, 156.9, 147.8, 79.8, 49.3, 40.8, 40.1, 31.1, 29.5, 28.4, 18.4 ppm. HRMS (ESI, positive mode): *m/z*: 325.1767 [M+H]⁺, 225.1232 [M-Boc+H]⁺, 247.1053 [M-Boc+Na]⁺, 347.1577 [M+Na]⁺, 649.3430 [2M+H]⁺, 671.3249 [2M+Na]⁺; M calcd. for C₁₆H₂₄N₂O₅ 324.1685.

***N*-(2-Aminoethyl)-3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanamide**

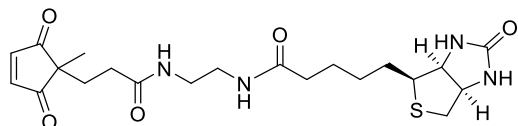


4 M Hydrochloric acid in dioxane (0.5 mL, 7.39 mmol) was added to a solution of *tert*-butyl (2-(3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanamido)ethyl)carbamate (60 mg, 0.184 mmol) in DCM (0.5 mL) and stirred at room temperature for 2 h. The solvent was removed under reduced pressure to afford the title compound (42 mg, quantitative yield) as a red solid.

¹H NMR (CD₃OD, 400 MHz): δ 7.34 (s, 2H), 3.38 (t, *J* = 8.0 Hz, 2H), 3.02 (t, *J* = 8.0 Hz, 2H), 2.15-2.11 (m, 2H), 1.93-1.89 (m, 2H), 1.12 (s, 3H) ppm. ¹³C NMR (CD₃OD, 101 MHz): δ 207.3, 174.4, 148.1, 49.0, 39.5, 36.8, 30.2, 29.1, 17.5 ppm. HRMS (ESI, positive

mode): m/z : 225.1234 $[M+H]^+$, 247.1053 $[M+Na]^+$, 471.2220 $[2M+Na]^+$; M calcd for $C_{11}H_{16}N_2O_3$ 224.1161.

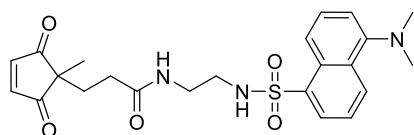
***N*-(2-(3-(1-Methyl-2,5-dioxocyclopent-3-en-1-yl)propanamido)ethyl)-5-((3*aS*,4*S*,6*aR*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (3.14)**



D-Biotin (50 mg, 0.204 mmol) was dissolved in anhydrous DMF (3.0 mL) under a nitrogen atmosphere and stirred for 30 min at $-10\text{ }^{\circ}\text{C}$. Once dissolved, NMM (56 μL , 0.510 mmol, 2.5 equiv.) and isobutyl chloroformate (26.5 μL , 0.204 mmol, 1.0 equiv.) were added, and the mixture was reacted under heavy stirring at $-10\text{ }^{\circ}\text{C}$ for 1 h. Thereafter, a solution of *N*-(2-aminoethyl)-3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanamide (46 mg, 0.204 mmol, 1.0 equiv.) in anhydrous DMF (2.0 mL) was added, and the mixture was stirred for 1 additional hour at $-10\text{ }^{\circ}\text{C}$ and at room temperature for 3 h. Afterwards, the solvent was evaporated in vacuo and the crude was purified by reversed-phase chromatography, using a 10 mL-syringe filled with 2.0 mL of C18-derivatised silica (Vydac). Elution was carried out with a gradient of 8.0 mL of H_2O , 8.0 mL of 3:1 (v/v) $\text{H}_2\text{O}/\text{MeOH}$ and 8.0 mL of 1:1 (v/v) $\text{H}_2\text{O}:\text{MeOH}$. The title compound (28 mg, 34%) was obtained as a yellow solid.

^1H NMR (CD_3OD , 400 MHz): δ 7.38 (s, 2H), 4.53 (dd, $J = 4.4\text{ Hz}$, $J = 7.6\text{ Hz}$, 1H), 4.35 (dd, $J = 4.8\text{ Hz}$, $J = 8.0\text{ Hz}$; 1H), 3.26 (m, 4H); 3.26 (m, 1H), 2.96 (dd, $J = 4.8\text{ Hz}$, $J = 12.4\text{ Hz}$, 1H), 2.74 (d, $J = 12.8\text{ Hz}$, 1H), 2.22 (t, $J = 6.4\text{ Hz}$, 2H), 2.09 (m, 2H), 1.91 (m, 2H), 1.81-1.57 (m, 4H), 1.45 (q, $J = 7.6\text{ Hz}$, 2H), 1.15 (s, 3H) ppm. ^{13}C NMR (CD_3OD , 101 MHz): δ 207.3, 174.9, 173.4, 164.67 7, 148.1, 61.9, 60.2, 55.5, 49.0, 39.7, 38.6, 38.5, 35.4, 30.5, 29.6, 28.3, 28.1, 25.4, 17.5 ppm. HRMS (ESI, positive mode): m/z : 451.2015 $[M+H]^+$, 473.1830 $[M+Na]^+$; 923.3712 $[2M+Na]^+$ M calcd. for $\text{C}_{21}\text{H}_{30}\text{N}_4\text{O}_5\text{S}$ 450.1937.

***N*-(2-(5-(dimethylamino)naphthalene-1-sulfonamido)ethyl)-3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanamide (3.15)**



Dansyl chloride (54.67 mg, 0.202 mmol) was added to a solution of *N*-(2-aminoethyl)-3-(1-methyl-2,5-dioxocyclopent-3-en-1-yl)propanamide (50 mg, 0.223 mmol) in anhydrous ACN (1 mL). Subsequently, TEA (0.5 mL, 3.59 mmol) was added and the mixture left for

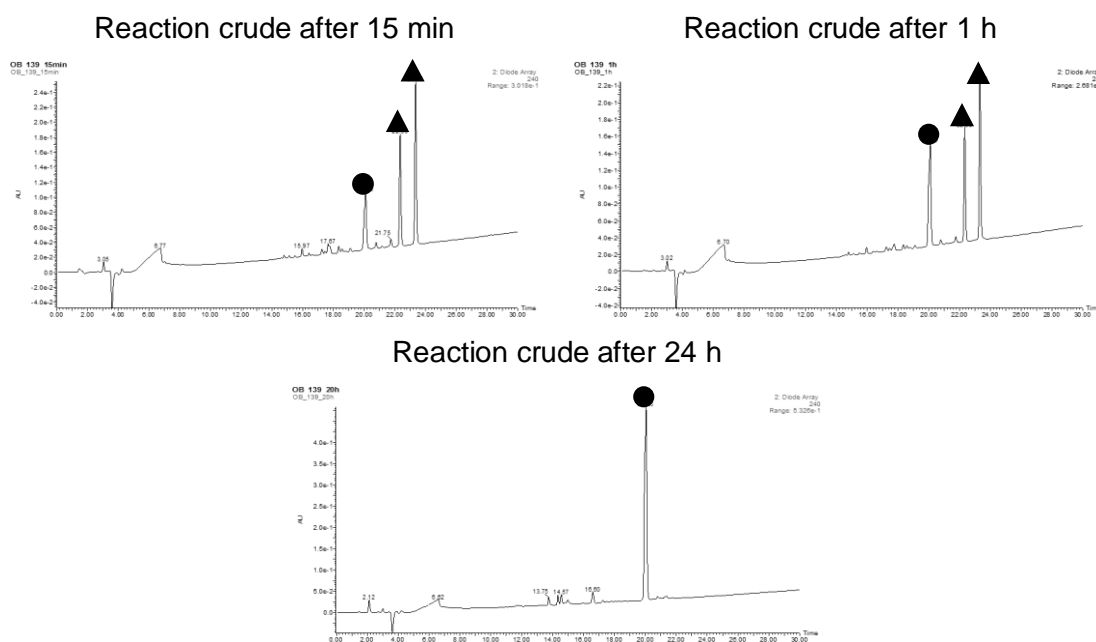
3 hours under heavy stirring at room temperature. Afterwards, the solvent was evaporated under reduced pressure, and the resulting residue dissolved in DCM (15 mL) and extracted with water (3 × 15 mL). The organic layer was extracted with 5% HCl (3 × 15 mL). The aqueous layer was basified to pH 6-7 with 5% NaHCO₃ and extracted with DCM (3 × 15 mL). The organic layer was dried over MgSO₄, filtered and the solvent evaporated under reduced pressure. The title compound (21.10 mg, 23% yield) was obtained as a yellow solid.

¹H NMR (CDCl₃, 400 MHz): δ 8.54 (d, *J* = 8.4 Hz, 1H); 8.28 (d, *J* = 8.4 Hz, 1H); 8.22 (d, *J* = 7.2 Hz); 7.58-7.50 (m, 2H); 7.20 (s, 2H), 7.19 (d, *J* = 6.8 Hz); 6.04 (t, *J* = 2.4 Hz, 1H); 5.82 (t, *J* = 5.2 Hz, 1H); 3.24-3.20 (m, 2H); 3.02-2.98 (m, 2H); 2.89 (s, 6H); 2.03-1.98 (m, 2H); 1.92-1.89 (m, 2H); 1.13 (s, 3H) ppm. ¹³C NMR (CDCl₃, 101 MHz): δ 207.02, 172.20, 152.06, 147.93, 134.44, 130.61, 129.89, 129.56, 129.44, 128.50, 123.19, 118.60, 115.27, 49.30, 45.39, 42.75, 39.48, 30.87, 29.34, 18.42 ppm. ESI HRMS (positive mode): *m/z*: 458.1730 [M+H]⁺, 480.1571 [M+Na]⁺, 937.3224 [2M+Na]⁺; M calcd. for C₂₃H₂₈N₃O₅S 457.1671.

2) Reaction of 3.14 with methyl cysteinate to furnish 3.16 and 3.17

3.14 (100 nmol) Was incubated with 5 equiv. of methyl cysteinate (500 nmol) in water (0.2 mM CPD concentration) at 37 °C, and the crude was analysed by HPLC/MS after 15 min, 1 h and 24 h reaction times.

HPLC/MS details: 0-50 % B, traces shown at 240 nm



Product code:

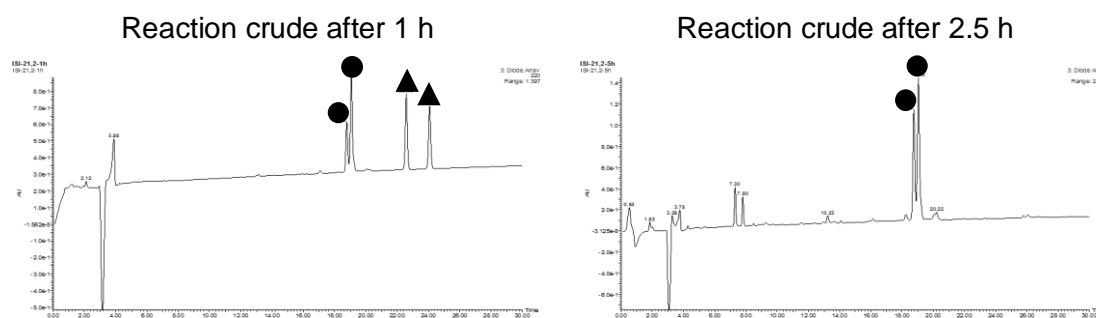
▲ : 3.16

● : 3.17

3) Reaction of 3.15 with methyl cysteinate to furnish 3.18 and 3.19

CPD-Dansyl (**3.15**, 100 nmol) was incubated with 5 equiv. of methyl cysteinate (500 nmol) in water (0.2 mM CPD concentration) at 60 °C, and the crude was analysed by HPLC/MS after 1 and 2.5 h.

HPLC/MS details: 0-50 % B, traces shown at 240 nm



Product code:

▲ : 3.18

● : 3.19

E.3.9 Reactions between CPDs and other amino acids: cysteine and homocysteine

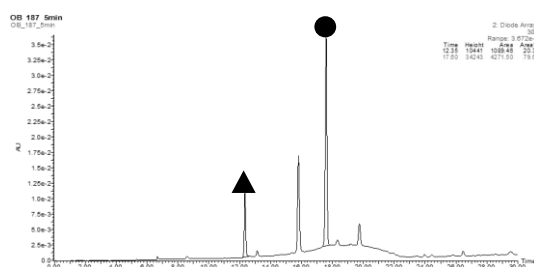
E.3.9.1 Reaction of CPDs with Cys

1) Reaction between cysteine and 3.6 to furnish 3.20, 3.21, 3.22 and 3.23

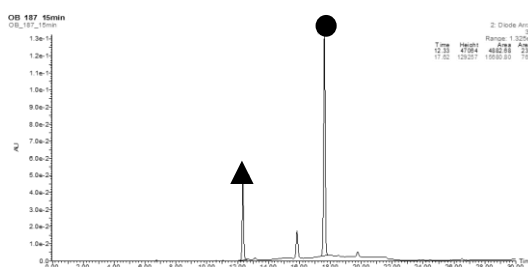
L-Cysteine (300 nmol, 1 mM concentration) was reacted with 5 equiv. of **3.6** in water at 37 °C. After 5, 15, 30, 60, 120 and 240 min aliquots of the reaction mixture were removed and analysed by HPLC/MS. HPLC/MS Traces and a table with the product ratios are shown below.

HPLC/MS details: 0-50 % B, traces shown at 300 nm

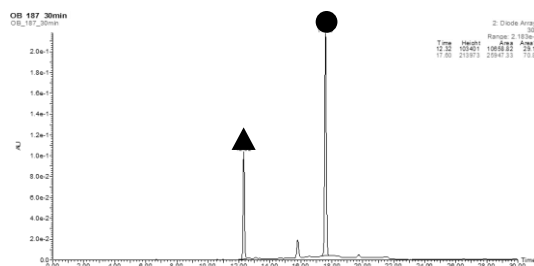
Reaction crude after 5 min



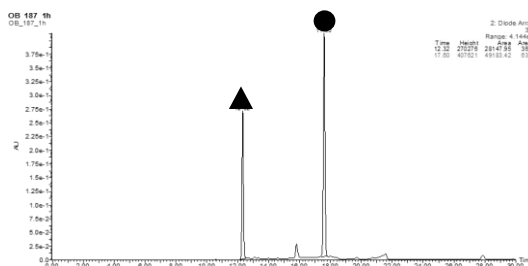
Reaction crude after 15 min



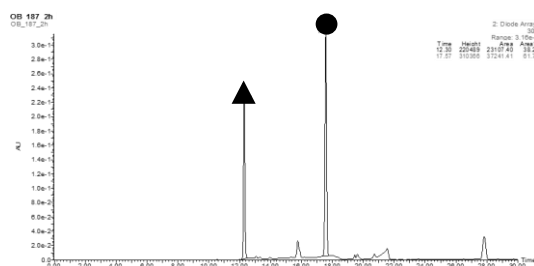
Reaction crude after 30 min



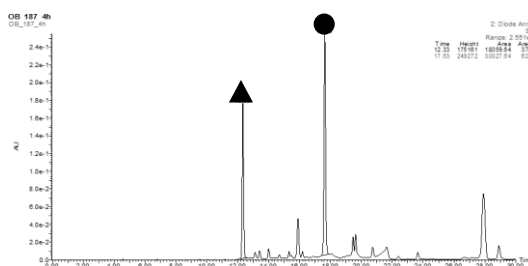
Reaction crude after 1 h



Reaction crude after 2 h



Reaction crude after 4 h



Product code:

▲ : 3.21

● : 3.22

Time (min)	3.21 (M-2 Da adduct, %)	3.22 (M-20 Da adduct, %)
5	20	80
15	24	76
30	29	71
60	36	64
120	38	62
240	38	62

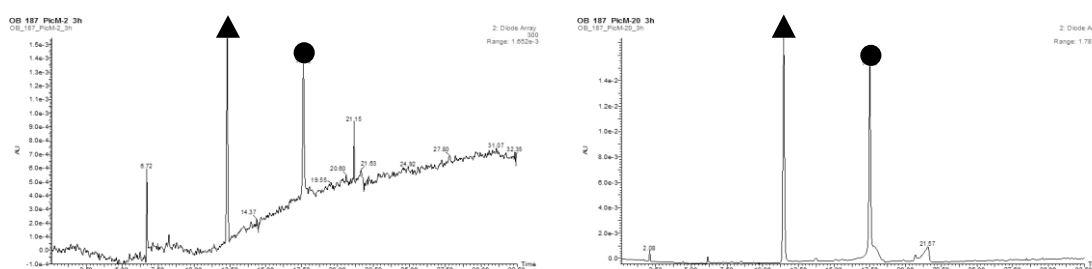
2) Stability of the products

Isolated peaks corresponding to 3.21 and 3.22 were incubated at 37 °C in water and reanalysed after 3 h to assess their stability.

HPLC/MS details: 0-50 % B, traces shown at 300 nm

Isolated **3.21** after 3 h at 37 °C

Isolated **3.22** after 3 h at 37 °C



Product code:

▲ : **3.21**

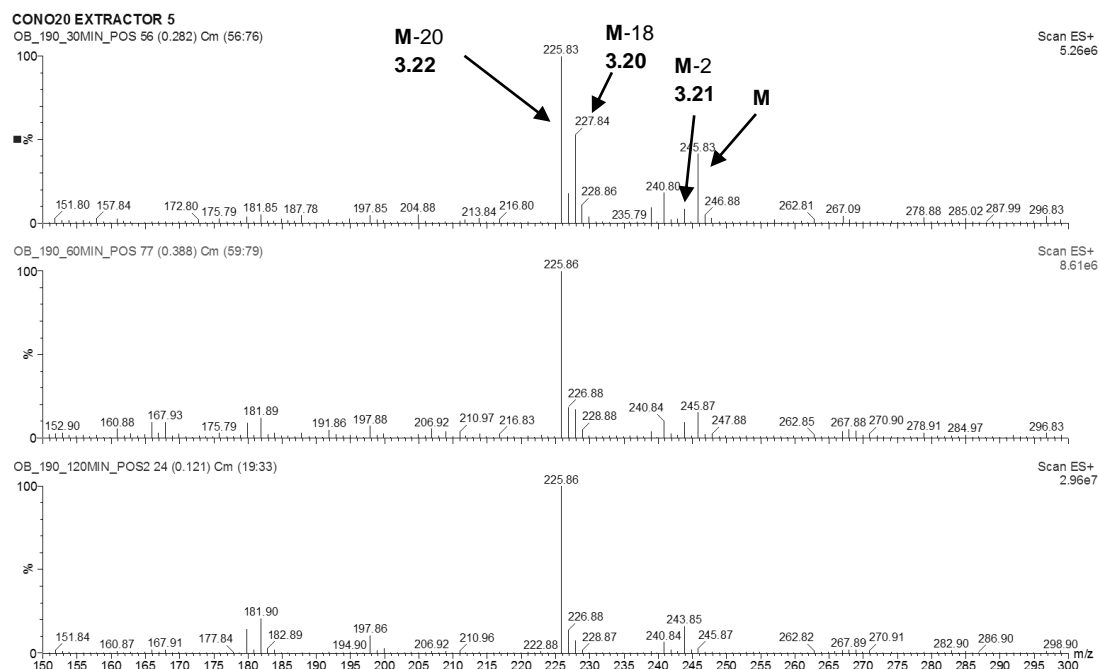
● : **3.22**

3) Reaction between 3.6 and cysteine followed by ESI-MS

In a separate experiment under the same conditions as those described above, aliquots of the reaction mixture were removed after 30, 60 and 120 min and analysed by ESI-MS (direct perfusion).

ESI-MS details: direct perfusion, 20 V, positive mode.

Traces shown after 30, 60 and 120 min reaction times, from top to bottom.

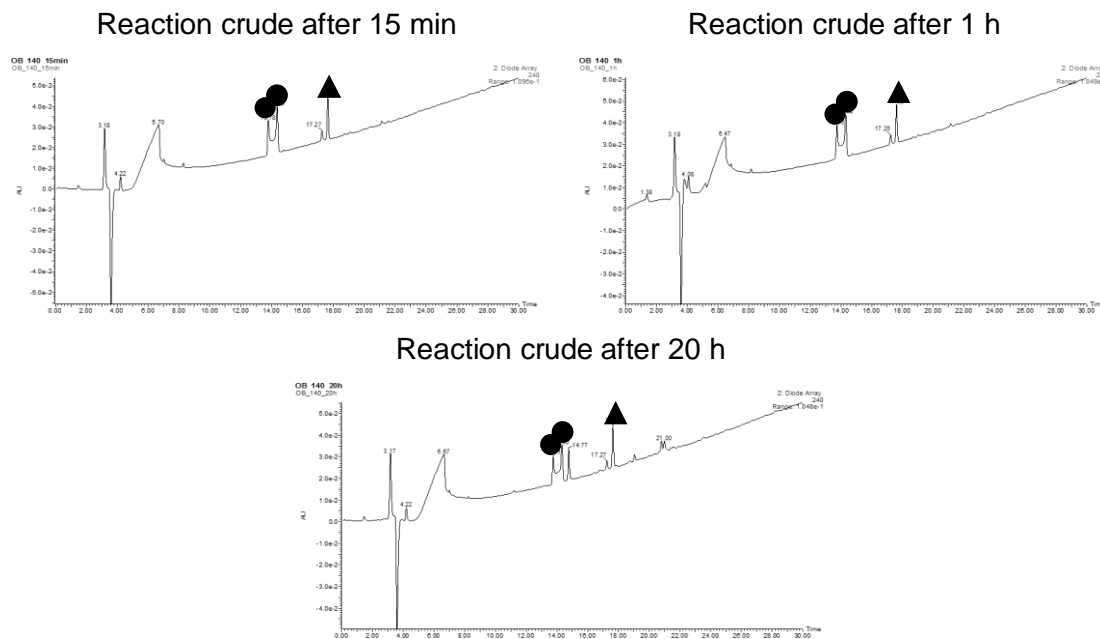


E.3.9.2 Reaction of CPDs with Hcy

1) Reaction between 3.14 and Hcy (3.24) to furnish 3.25

3.14 (100 nmol, 0.2 mM concentration) Was incubated with 5 equiv. of homocysteine in water at 37 °C, and the crude was analysed by HPLC/MS after 15 min, 1 h and 20 h reaction times.

HPLC/MS details: 0-50 % B, traces shown at 240 nm



Product code:

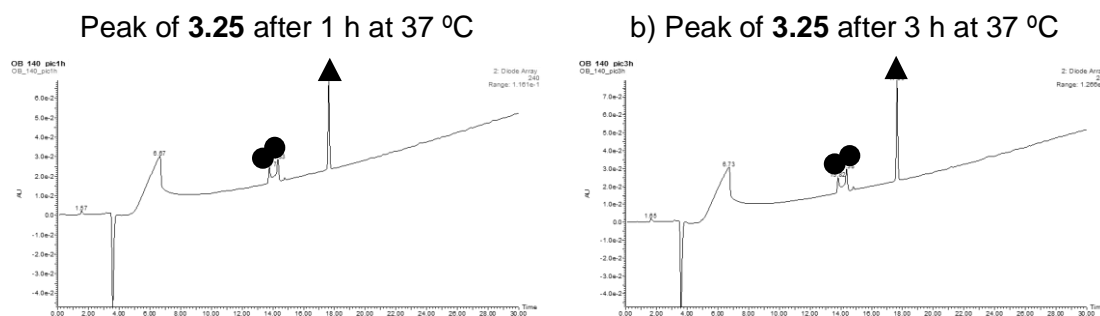
▲ : 3.14

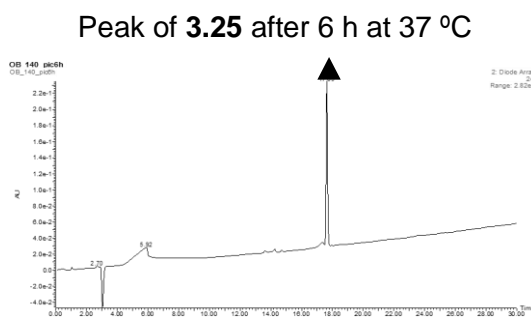
● : 3.25

2) Stability of 3.25 in water

Peaks corresponding to **3.25** were collected (together) and lyophilised. Afterwards, they were dissolved in water (400 µL) and reanalysed 1, 3 and 6 h after incubation at 37 °C.

HPLC/MS details: 0-50 % B, traces shown at 240 nm





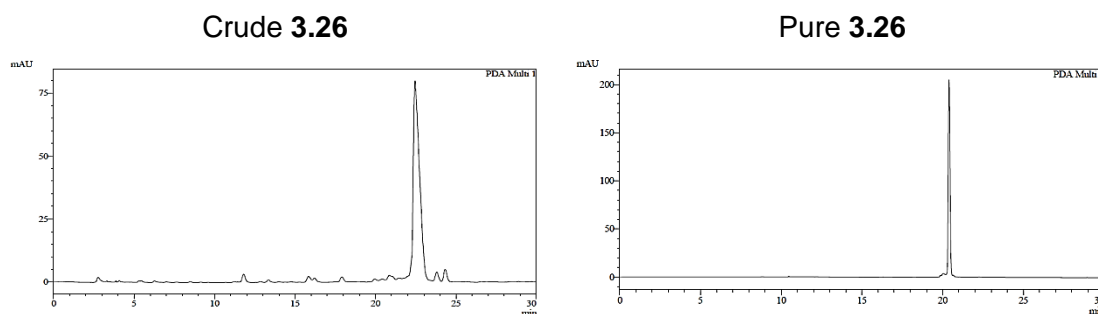
Product code:

▲ : **3.14**● : **3.25**

E.3.10 Attachment of CPDs to peptide chains

After chain elongation, **3.1** was incorporated to 3 μ mol of protected-Gly-Arg-Gly-Ser-Tyr-Glu-Ala-Tyr-Lys-resin by reacting 5 equivalents of **3.1** and 5 equivalents of DIPC with the peptide-resin for 90 min. Cleavage and deprotection of the CPD-derivatised peptide was carried out by treatment with a 95:2.5:2.5 TFA/H₂O/TIS mixture for 2 h at room temperature. The CPD-peptide **3.26** was purified and isolated in 21% yield (incorporation of **3.1**, cleavage and deprotection and purification steps).

HPLC details: 10-20 % B for the crude, 0-40 % B for the pure compound, traces shown at 280 nm



E.3.11 Synthesis of bioconjugates exploiting the formation of an M-20 Da adduct and assessment of the stability of the generated products

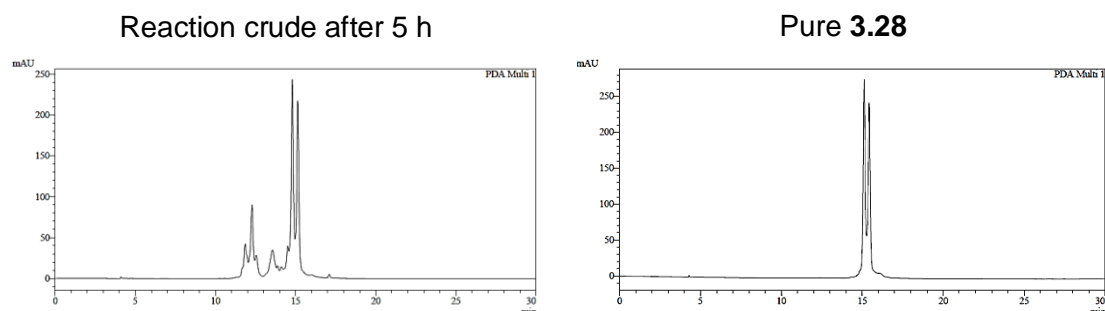
1) Reaction between **3.26** and **3.27** to generate **3.28**

PNA H-Cys-Lys-catagctgtttc-NH₂ (**3.27**, 130 nmol, 0.85 mM concentration) was reacted with 1.2 equivalents of **3.26** in water at 37 °C. Progress of the reaction was monitored by

HPLC. Although the reaction seemed to have finished after 3 h, it was left stirring for a total of 5 h to make sure that all the **M-18** Da adducts had evolved to the **M-20** Da ones (one of the **M-18** Da adducts coeluted with one of the **M-20** Da adducts).

The final conjugate (**3.28**) was purified and isolated in 51% yield.

HPLC details: 5-35 % B, t= 60 °C, traces shown at 260 nm



Product retention times:

3.28: 15.1 + 15.4, PNA impurities present in the starting material: 11.5-13.0 min

2) Stability of the conjugate **3.28** at different pHs

Aliquots of 5 nmol of **3.28** were incubated at 37 °C at pH 6.0, 7.0 and 8.0 using 500 µL of 0.2 M phosphate buffers (10 µM conjugate concentration). Each sample was analysed after 2, 6 and 24 h by HPLC and the percentage of unaltered conjugate determined by integration of the different peaks at 260 nm. Data obtained are shown below:

	Remaining 3.28 percentage		
	pH = 6.0	pH = 7.0	pH = 8.0
2 h	100	100	100
6 h	99	98	87
24 h	97	86	65

3) Stability of the conjugate **3.28** at pH 7.0 in the presence of thiols

Two aliquots of 5 nmol of **3.28** were incubated at 37 °C at pH 7.0 (0.2 M phosphate buffer, 10 µM conjugate concentration) in the presence of a 100-fold excess of methyl L-cysteinate or glutathione. Each sample was analysed after 2, 6 and 24 h by HPLC and the percentage of unaltered conjugate determined by integration of the different peaks at 260 nm. Data obtained is shown below:

	Remaining 3.28 percentage	
	+ Glutathione	+ Methyl cysteinate
2 h	100	79
6 h	93	62
24 h	86	53

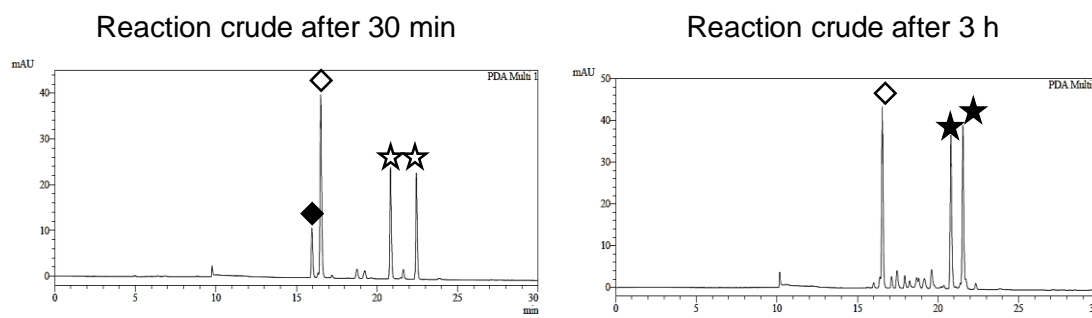
E.3.12 Exploiting the selectivity of CPDs towards *N*-terminal cysteines

E.3.12.1 Selective derivatization of a peptide containing an *N*-terminal cysteine in a mixture of cysteine-containing peptides

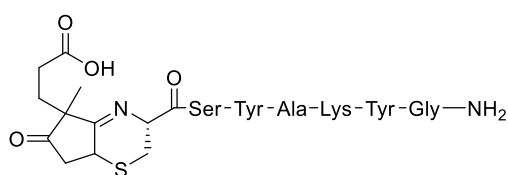
1) Experiment starting with a defect of CPD

A mixture containing equimolar amounts of **3.29** and **3.2** (100 nmol each, ratio checked by HPLC) was mixed with 0.7 equiv. of **3.1** to reach a final 0.5 mM peptide concentration. This solution was incubated at 37 °C for 30 min and analysed by HPLC (the mixture was kept in liquid N₂ during the HPLC analysis of an aliquot). After this first analysis, 0.3 additional equiv. of **3.1** were added to the reaction mixture and this was left at 37 °C until a 3 h total reaction time. The mixture was again analysed and the final **3.31** isolated in 41 % yield.

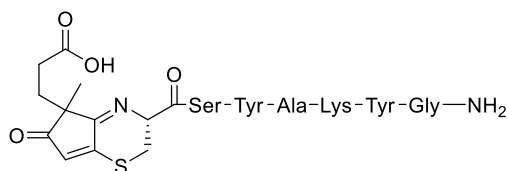
HPLC details: 0-50 % B, traces shown at 280 nm



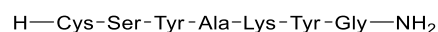
Product code



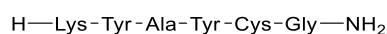
★ 3.30



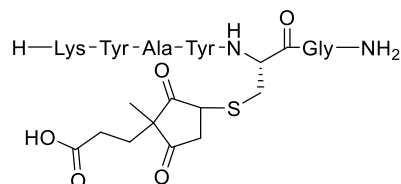
★ 3.31



◆ 3.29



◇ 3.2

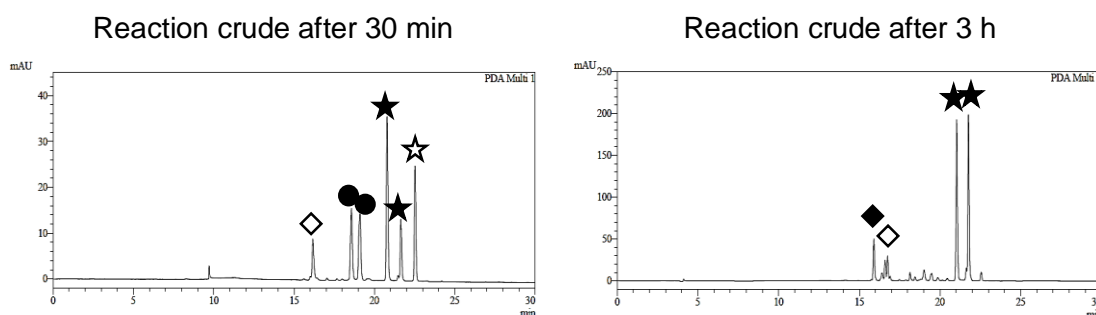


● 3.4

2) Experiment starting with an excess of CPD:

An equimolar mixture of **3.29** and **3.2** (100 nmol each, ratio checked by HPLC) was mixed with 3.5 equiv. of **3.1** (350 nmol) to reach a final 0.5 mM peptide concentration. This solution was incubated at 37 °C for 30 min and analysed by HPLC (the mixture was kept in liquid N₂ during the HPLC analysis of an aliquot). After a total of 2 h at 37 °C (when we deemed complete oxidation of the **M**-18 Da adduct would have taken place) 400 nmol of **3.29** were added to the reaction mixture. 60 Minutes after this addition the reaction crude was analysed and purified to obtain **3.31** in a 45 % yield.

HPLC details: 0-50 % B, traces shown at 280 nm



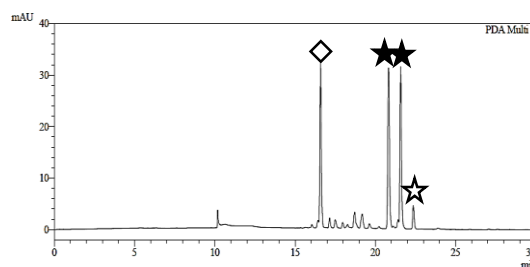
Product code as above

3) Experiment starting with equimolar amounts of CPD

An equimolar mixture of **3.29** and **3.2** (100 nmol each, ratio checked by HPLC) was mixed with 1 equiv. of **3.1** (100 nmol) to reach a final 0.5 mM peptide concentration. This solution was incubated at 37 °C for 3 h and analysed by HPLC. **3.31** Was isolated in 47 % yield.

HPLC details: 0-50 % B, traces shown at 280 nm

Reaction crude after 3 h

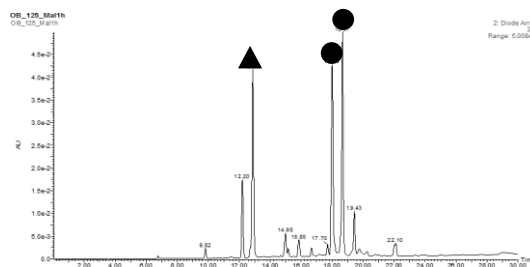


Product code as above

E.3.12.2 Regioselective double labelling of a peptide containing *N*-terminal and internal cysteines

Peptide **3.32** (100 nmol, 0.5 mM concentration) was incubated with 1.1 equiv. of **3.1** in water, at 37 °C for 90 min. 5 Equiv. of 3-maleimidopropanoic acid were then added, and the mixture was left stirring for an additional hour. After purification, **3.34** was obtained in 44 % yield (as two peaks, numbered after their elution order in the HPLC).

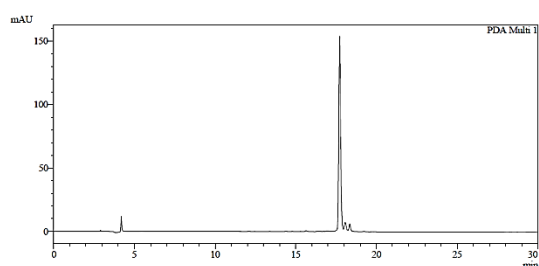
HPLC/MS details: 0-40 % B, trace shown at 280 nm.



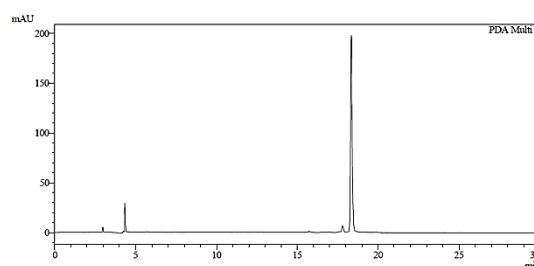
▲ : excess of 3-maleimidopropanoic acid.

● : **3.34**

Pure peak 1 (280 nm)



Pure peak 2 (280 nm)



E.3.13 Estimation of the M-20 Da adduct molar absorption coefficient

A stock solution of peptide **3.29** was prepared and quantified by measuring the absorbance of four different aliquots at 280 nm. Its concentration was determined to be 1.275 mM.

Three aliquots of 400 μL of this peptide solution were mixed with 5 equivalents of **3.1** to reach a final volume of 480 μL (1.063 mM peptide concentration) and left shaking at 37 $^{\circ}\text{C}$.

After 1.5 h the absorbance of two aliquots of each reaction mixture was measured (48 μL diluted to 1000 μL). After an additional hour, three aliquots were taken from each reaction mixture and their UV-Vis profile recorded three times (48 μL diluted to 1000 μL). As there was no difference between the absorbance values recorded at 1.5 h and those recorded at 2.5 h, the reaction was deemed finished. Also, because the difference between the highest and the lowest absorbance values recorded after 2.5 h was less than 6 %, which fits the typical errors of a UV-Vis quantification (~ 10 %), all 27 values were considered comparable. The mean absorbance found at 332 nm had a relative error of less than 2 %, and the estimated molar absorption coefficient was $9140 \text{ cm}^{-1} \text{ M}^{-1} \pm 3$ %. The small differences found in the UV-Vis measurements and the small relative error of the final ϵ value made us believe that the results obtained were trustworthy.

The same process was performed to establish the ϵ value of the **M-20 Da** adduct at 280 nm. Again, as the difference between the highest and the lowest value was around 8.5 %, all 27 absorbance results were considered comparable. The mean absorbance found at 280 nm had a relative error around 2.5 %, and, after subtracting the tyrosines contribution, the molar absorption coefficient was found to be $4210 \text{ cm}^{-1} \text{ M}^{-1} \pm 2$ %.

		Absorption values at 332 nm			ε values at 332 nm		
		Mesure 1	Mesure 2	Mesure 3	Mesure 1	Mesure 2	Mesure 3
Reaction 1	Replicate 1	0,4809	0,4794	0,4786	9424,9	9396,3	9381,6
	Replicate 2	0,4554	0,4556	0,4572	8926,1	8930,8	8962,1
	Replicate 3	0,4785	0,4781	0,4786	9378,9	9371,8	9380,6
Reaction 2	Replicate 1	0,4676	0,4661	0,4681	9165,2	9135,0	9174,2
	Replicate 2	0,4615	0,4618	0,4607	9046,4	9050,5	9029,6
	Replicate 3	0,4723	0,4712	0,4709	9257,5	9236,4	9230,7
Reaction 3	Replicate 1	0,4670	0,4672	0,4676	9153,1	9156,6	9165,2
	Replicate 2	0,4605	0,4603	0,4587	9025,3	9022,1	8990,8
	Replicate 3	0,4569	0,4571	0,4564	8955,9	8958,6	8945,1

		Absorption values at 280 nm			ε values at 280 nm		
		Mesure 1	Mesure 2	Mesure 3	Mesure 1	Mesure 2	Mesure 3
Reaction 1	Replicate 1	0,3856	0,3814	0,3797	7558,8	7476,0	7442,9
	Replicate 2	0,3590	0,3555	0,3559	7037,0	6967,8	6976,6
	Replicate 3	0,3774	0,3767	0,3771	7397,6	7383,7	7391,6
Reaction 2	Replicate 1	0,3712	0,3692	0,3701	7275,9	7236,5	7254,5
	Replicate 2	0,3585	0,3596	0,3583	7025,8	7048,9	7022,5
	Replicate 3	0,3714	0,3703	0,3676	7280,2	7257,3	7204,6
Reaction 3	Replicate 1	0,3702	0,3708	0,3700	7255,7	7267,3	7251,2
	Replicate 2	0,3589	0,3584	0,3586	7035,4	7025,0	7029,5
	Replicate 3	0,3593	0,3580	0,3570	7042,9	7017,0	6997,4

ε values at 332 nm		
Mean	Std. Deviation	% error
9142,6	162,4	1,8

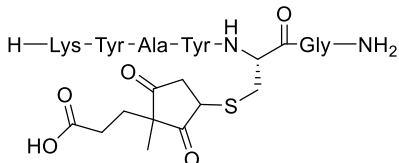
ε values at 280 nm		
Mean	Std. Deviation	% error
7191,1	172,4	2,4
- 2980 = 4211,1		

E.3 Product characterisation

Product: **3.2**
Structure: H—Lys-Tyr-Ala-Tyr-Cys-Gly—NH₂
HPLC characterisation:
Gradient and retention time: Analysis conditions 10-40 % B, t_R= 10.5 min
MS characterisation:
Technique: MALDI-TOF, positive mode, DHB as matrix
Mass calculated: C₃₂H₄₆N₈O₈S 702.3
Mass found: *m/z* 703.5 [M+H]
Purification conditions: 15-20 % B, t_R= 8.1 min

Product: **3.3**
Structure: H—Cys-Tyr-Gly—NH₂
HPLC characterisation:
Gradient and retention time: Analysis conditions 0-40 % B, t_R= 12.8 min
MS characterisation:
Technique: MALDI-TOF, positive mode, DHB as matrix
Mass calculated: C₁₄H₂₀N₄O₄S 340.1
Mass found: *m/z* 341.2 [M+H]⁺
Purification conditions: 5-50 % B, t_R= 9.8 min

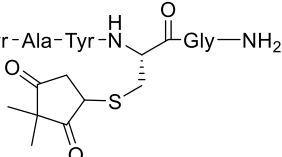
Product: **3.4**
Structure:



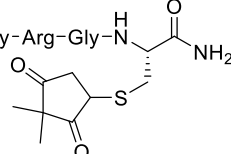
The structure shows a peptide chain H-Lys-Tyr-Ala-Tyr-NH-CO-Gly-NH₂. The fourth tyrosine residue is modified with a cyclic structure. This structure consists of a five-membered ring containing a sulfur atom (S) and a carbonyl group (C=O). A carboxylic acid group (HO-C=O) is attached to the ring. The nitrogen atom of the tyrosine residue is connected to the ring via a dashed bond, and the sulfur atom is connected to the ring via a solid wedge bond.

HPLC characterisation:
Gradient and retention time: Analysis conditions 0-40 % B, t_R= 21.5 min + 22.2 min
MS characterisation:
Technique: MALDI-TOF, positive mode, DHB as matrix
Mass calculated: C₄₁H₅₆N₈O₁₂S 884.4
Mass found: *m/z* 885.7 [M+H]⁺
Purification conditions: Purified using analysis conditions

Product: **3.5**
 Structure: H—Trp—Gly—Arg—Gly—Cys—NH₂
 HPLC characterisation:
 Gradient and retention time: Analysis conditions 0-50 % B, t_R= 16.5 min
 MS characterisation:
 Technique: MALDI-TOF, positive mode, DHB as matrix
 Mass calculated: C₂₄H₃₆N₁₀O₅S 576.3
 Mass found: *m/z* 577.5 [M+H]⁺
 Purification conditions: 15-50 % B, t_R= 7.9

Product: **3.7**
 Structure: H—Lys—Tyr—Ala—Tyr—N^H—CH₂—C(=O)—Gly—NH₂


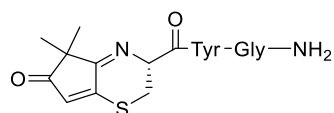
HPLC characterisation:
 Gradient and retention time: Analysis conditions 10-40 % B, t_R= 16.5 min
 MS characterisation:
 Technique: MALDI-TOF, positive mode, DHB as matrix
 Mass calculated: C₃₉H₅₄N₈O₁₀S 826.4
 Mass found: *m/z* 827.6 [M+H]⁺
 Purification conditions: Purified using analysis conditions

Product: **3.8**
 Structure: H—Trp—Gly—Arg—Gly—N^H—CH₂—C(=O)—NH₂


HPLC characterisation:
 Gradient and retention time: Analysis conditions 10-40 % B, t_R= 17.0 + 17.4 min
 MS characterisation:
 Technique: MALDI-TOF, positive mode, DHB as matrix
 Mass calculated: C₃₁H₄₄N₁₀O₇S 700.3
 Mass found: *m/z* 701.5 [M+H]⁺
 Purification conditions: Purified using analysis conditions

Product: **3.9**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 10-40 % B, $t_R = 18.5$ min

MS characterisation:

Technique: ESI, positive mode

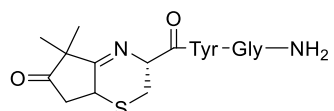
Mass calculated: $C_{21}H_{24}N_4O_5S$ 444.2

Mass found: m/z 444.6 $[M+H]^+$

Purification conditions: -

Product: **3.10**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 10-40 % B, $t_R = 19.8$ min

MS characterisation:

Technique: ESI, positive mode

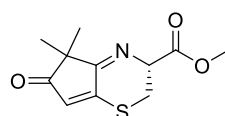
Mass calculated: $C_{21}H_{26}N_4O_5S$ 446.2

Mass found: m/z 446.6 $[M+H]^+$

Purification conditions: -

Product: **3.11**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 10-50 % B, $t_R = 17.9$ min

MS characterisation:

Technique: HRMS ESI, positive mode

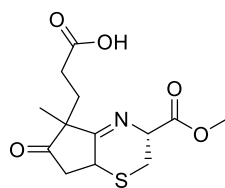
Mass calculated: $C_{11}H_{13}NO_3S$ 239.0616

Mass found: m/z 240.0694 $[M+H]^+$

Purification conditions: -

Product: **3.12**

Structure:

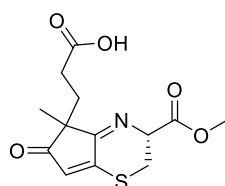


HPLC characterisation:

Gradient and retention time: Analysis conditions 20-50 % B in 20 minutes, $t_R = 16.2 + 17.7$ min

Product: **3.13**

Structure:



HPLC characterisation:

Gradient and retention time: Analysis conditions, 20-50 % B in 20 minutes, $t_R = 11.3 + 11.6$ min

MS characterisation:

Technique: ESI, positive mode

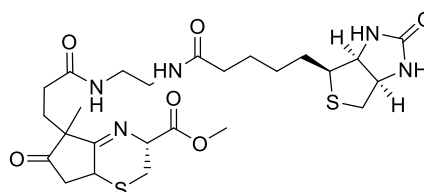
Mass calculated: $C_{13}H_{15}NO_5S$ 297.0

Mass found: m/z 298.2 $[M+H]^+$

Purification conditions: -

Product: **3.16**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 22.4$ min + 23.4 min

MS characterisation:

Technique: ESI, positive mode

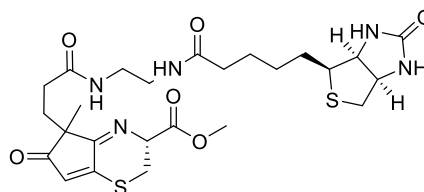
Mass calculated: $C_{25}H_{37}N_5O_6S_2$ 567.2

Mass found: m/z 567.9 $[M+H]^+$

Purification conditions: -

Product: **3.17**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 20.1$ min

MS characterisation:

Technique: ESI, positive mode

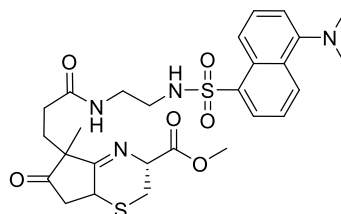
Mass calculated: $C_{25}H_{35}N_5O_6S_2$ 565.2

Mass found: m/z 565.9 $[M+H]^+$

Purification conditions: -

Product: **3.18**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 20-70 % B, $t_R = 22.6$ min + 24.1 min

MS characterisation:

Technique: ESI, positive mode

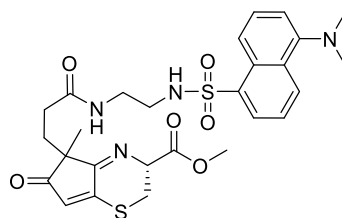
Mass calculated: $C_{27}H_{34}N_4O_6S_2$ 574.2

Mass found: m/z 574.8 $[M+H]^+$

Purification conditions: -

Product: **3.19**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 20-70 % B, $t_R = 18.8 \text{ min} + 19.1 \text{ min}$

MS characterisation:

Technique: ESI, positive mode

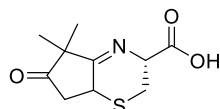
Mass calculated: $C_{27}H_{32}N_4O_6S_2$ 572.2

Mass found: m/z 572.8 $[M+H]^+$

Purification conditions: -

Product: **3.20**

Structure:



MS characterisation:

Technique: ESI, positive mode

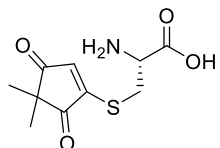
Mass calculated: $C_{10}H_{13}NO_3S$ 227.1

Mass found: m/z 227.8 $[M+H]^+$

Purification conditions: -

Product: **3.21**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 12.3 \text{ min}$

MS characterisation:

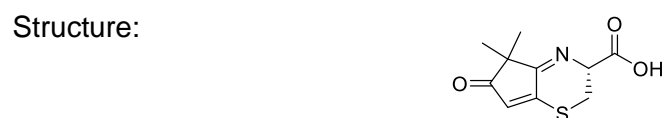
Technique: ESI, positive mode

Mass calculated: $C_{10}H_{13}NO_4S$ 243.0

Mass found: 243.7 $[M+H]^+$

Purification conditions: -

Product: **3.22**



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 17.6$ min

MS characterisation:

Technique: ESI, positive mode

Mass calculated: $C_{10}H_{11}NO_3S$ 225.0

Mass found: m/z 225.7 $[M+H]^+$

Purification conditions: -

Product: **3.23**



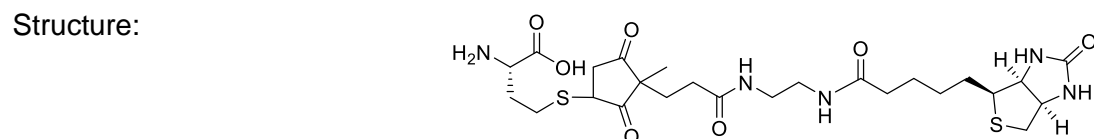
MS characterisation:

Technique: ESI, positive mode

Mass calculated: $C_{10}H_{15}NO_4S$ 245.1

Mass found: m/z 245.8 $[M+H]^+$

Product: **3.25**



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 13.8$ min + 14.4 min

MS characterisation:

Technique: ESI, positive mode

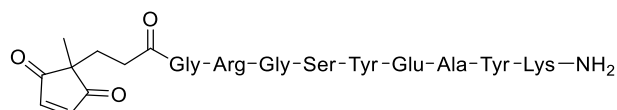
Mass calculated: $C_{25}H_{39}N_5O_7S_2$ 585.2

Mass found: m/z 586.0 $[M+H]^+$

Purification conditions: -

Product: **3.26**

Structure:



HPLC characterisation:

Gradient and retention time: Analysis conditions, 0-40 % B, $t_R = 21.0$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

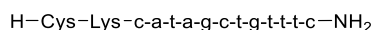
Mass calculated: $C_{54}H_{76}N_{14}O_{17}$ 1192.6

Mass found: m/z 1193.5 $[M+H]^+$

Purification conditions: -

Product: **3.27**

Structure:



HPLC characterisation:

Gradient and retention time: Analysis conditions, 5-35 % B, 60°C, $t_R = 12.4$ min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

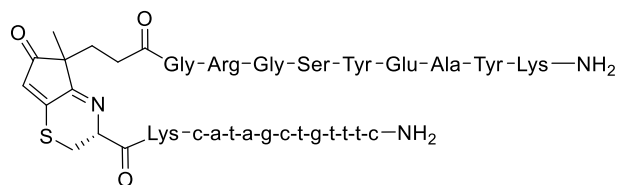
Mass calculated: $C_{138}H_{181}N_{67}O_{41}S$ 3464.4

Mass found: m/z 3464.9 $[M+H]^+$

Purification conditions: 5-35 % B, 60°C, $t_R = 12.5$ min

Product: **3.28**

Structure:



HPLC characterisation:

Gradient and retention time: Analysis conditions, 5-35 % B, 60 °C, $t_R = 15.1$ min + 15.4 min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

Mass calculated: $C_{192}H_{253}N_{81}O_{57}S$ 4636.9

Mass found: m/z 4639.7 $[M+H]^+$

Purification conditions: 10-30 % B, 60°C, $t_R = 12.5$ min

Product: **3.29**

Structure: H—Cys—Ser—Tyr—Ala—Lys—Tyr—Gly—NH₂

HPLC characterisation:

Gradient and retention time: Analysis conditions, 0-50 % B, t_R= 16.0 min

MS characterisation:

Technique: MALDI-TOF, positive mode, DHB as matrix

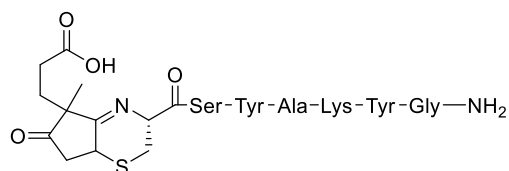
Mass calculated: C₃₅H₅₁N₉O₁₀S 789.9

Mass found: m/z 790.5 [M+H]⁺

Purification conditions: 10-30 % B, t_R= 11.9 min

Product: **3.30**

Structure:

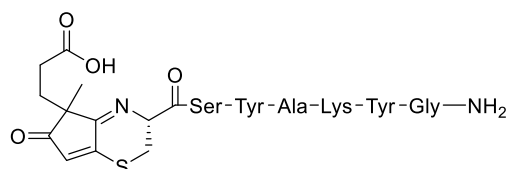


HPLC characterisation:

Gradient and retention time: Analysis conditions, 0-50 % B, t_R= 22.5 min

Product: **3.31**

Structure:



HPLC characterisation:

Gradient and retention time: Analysis conditions, 0-50 % B, t_R= 20.9 min and 21.7 min.

MS characterisation:

Technique: ESI, positive mode

Mass calculated: C₄₄H₅₇N₉O₁₃S 951.4

Mass found: m/z 952.7 [M+H]⁺

Purification conditions: 20-60 % B, t_R= 10.0 + 10.8 min

Product: **3.32**

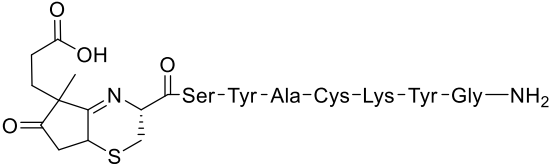
Structure: H—Cys—Ser—Tyr—Ala—Cys—Lys—Tyr—Gly—NH₂

HPLC characterisation:
 Gradient and retention time: Analysis conditions, 0-50 % B, t_R = 17.4 min

MS characterisation:
 Technique: MALDI-TOF, positive mode, DHB as matrix
 Mass calculated: C₃₈H₅₆N₁₀O₁₁S₂ 892.4
 Mass found: m/z 893.5 [M+H]⁺

Purification conditions: 15-45 % B, t_R = 8.2 min

Product: **3.33**

Structure: 

HPLC/MS characterisation:
 Gradient and retention time: 0-40 % B, t_R = 17.7 + 19.1 min

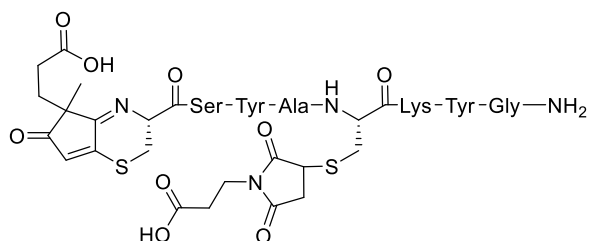
MS characterisation:
 Technique: ESI, positive mode
 Mass calculated: C₄₇H₆₄N₁₀O₁₄S₂ 1056.4
 Mass found: m/z 1057.7 [M+H]⁺

Purification conditions: -

Product:

3.34

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-40 % B, $t_R = 18.0$ min and 18.7 min

MS characterisation:

Technique: HRMS ESI, positive mode

Mass calculated: $C_{54}H_{69}N_{11}O_{18}S_2$ 1223.4263.

Mass found: m/z 1224.4310 $[M+H]^+$

Purification conditions: -

E.3 Abbreviations

ACN: Acetonitrile

CPD: 2,2-disubstituted cyclopent-4-ene-1,3-dione

DCM: Dichloromethane

DMF: *N,N*-dimethylformamide

DMSO: Dimethyl Sulfoxide

ESI: ElectroSpray Ionisation

HPLC/MS: High Performance Liquid Chromatography/Mass spectrometry

HPLC: High Performance Liquid Chromatography

MS: Mass Spectrometry

NMR :Nuclear Magnetic Resonance

ppm: parts per milion

TEA: Triethylamine

TFA: trifluoroacetic acid

E.3 Bibliography

(1) Ellman, G. L. *Arch. Biochem. Biophys.* **1959**, 82 (1), 70.

Experimental section: chapter 4

E.4 General methods

HPLC

Analysis conditions

Shimadzu instrument, Jupiter Proteo column (4 μm , 90 \AA , 250 \times 4.6 mm) from Phenomenex. Linear gradients of 30 min were always used unless otherwise stated. Solvent A: 0.045 % TFA in water, solvent B: 0.036 % TFA in ACN, flow: 1 mL/min. Detection wavelength: 200-400 nm. All analysis were performed at room temperature unless otherwise indicated.

Purification conditions

Jupiter Proteo column (10 μm , 90 \AA , 250 \times 10.0 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 % TFA in water, solvent B: 0.1 % TFA in ACN, flow: 3 mL/min. Detection wavelength: 200-400 nm. All purifications were performed at room temperature unless otherwise indicated.

HPLC/MS analysis conditions

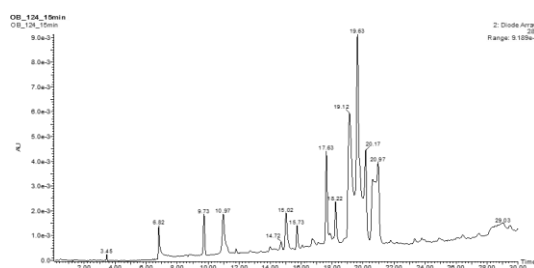
Waters 2695 instrument coupled to a Waters Micromass ZQ 4000 mass spectrometer, Jupiter Proteo column (4 μm , 90 \AA , 250 \times 4.6 mm) from Phenomenex. Linear gradients of 30 min were always used. Solvent A: 0.1 % formic acid in water, solvent B: 0.1 % formic acid in ACN, flow: 1 mL/min. Detection wavelength: 200-600 nm.

E.4.1 First indication for the formation of a cyclic structure

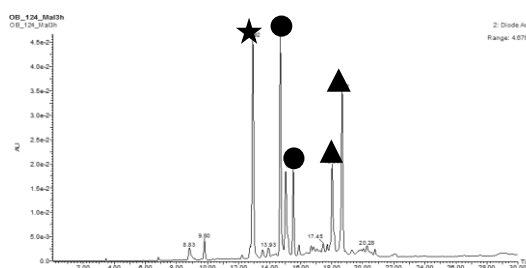
Peptide H-Cys-Ser-Tyr-Ala-Cys-Lys-Tyr-Gly-NH₂ (**3.32**, 100 nmol, 0.5 mM peptide concentration) was reacted for 15 min with 5 equiv. of **3.1** at 37 °C in water, after which time the crude was analysed by HPLC/MS (the reaction crude was kept in liquid nitrogen in the meanwhile). The reaction forming the **M**-18 Da adduct **4.1** (as a mixture of diastereomers) was shown to be complete. The crude was subsequently melted and 5 equiv. of 3-maleimidopropanoic acid were added to the reaction mixture, which was left stirring at 37 °C for 3 h. A new HPLC/MS analysis at this point revealed the formation of **3.34** together with **4.3**. Posterior analysis at 5.5 h did not show any appreciable change in the reaction crude.

HPLC/MS details: 0-40 % B, traces shown at 280 nm

Reaction crude after 15 min



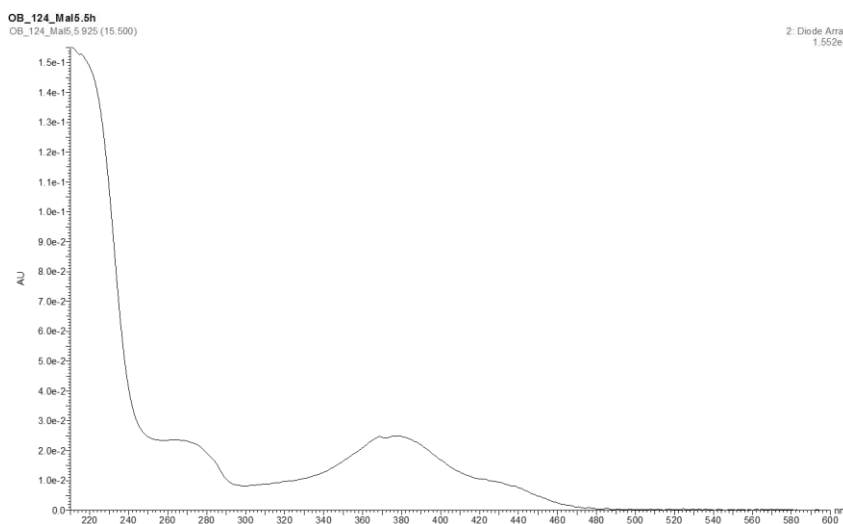
Reaction crude 3 h after the addition of
3-maleimidopropanoic acid



Product code:

- ★ :excess 3-maleimidopropanoic acid; ● :hypothesised cyclic adduct **4.3**
- ▲ :regioselectively double derivatised peptide **3.34**

UV-Vis profile of **4.3**

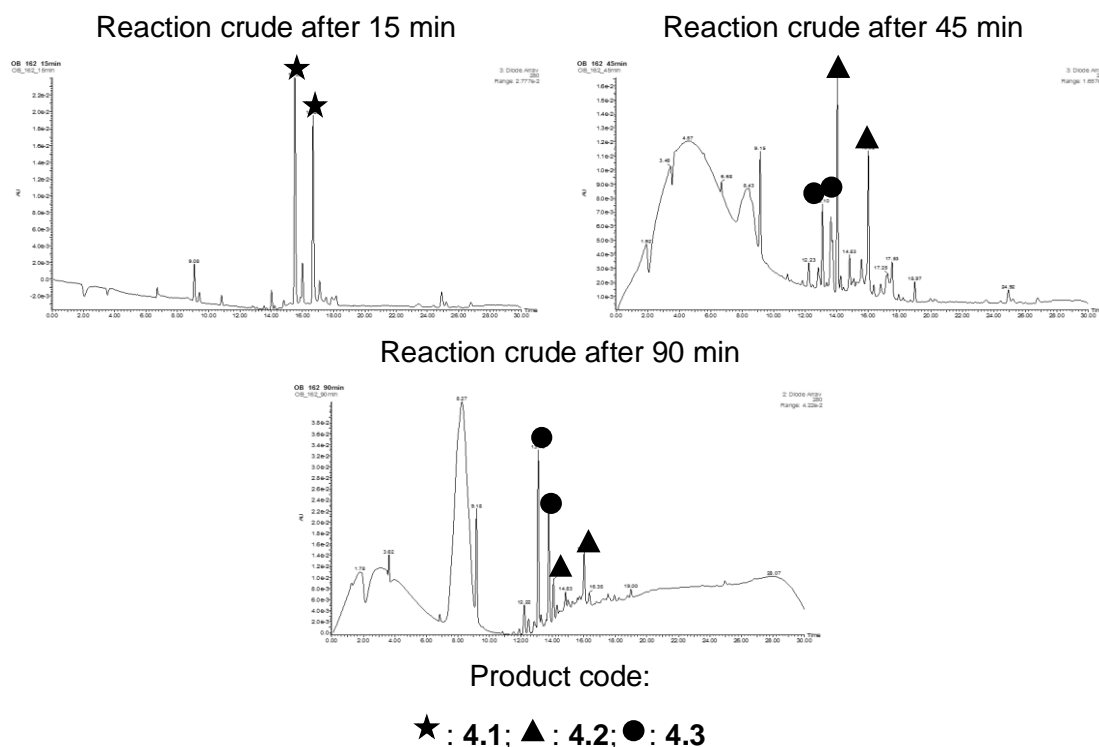


E.4.2 First steps to set a procedure for the cyclisation reaction

1) Reaction with peptide **3.32**

Peptide **3.32** (50 nmol, 0.5 mM concentration) was mixed with 1.1 equiv. of **3.1** and left reacting in water at 60 °C. HPLC/MS Analyses were performed at 15, 45 and 90 min reaction times.

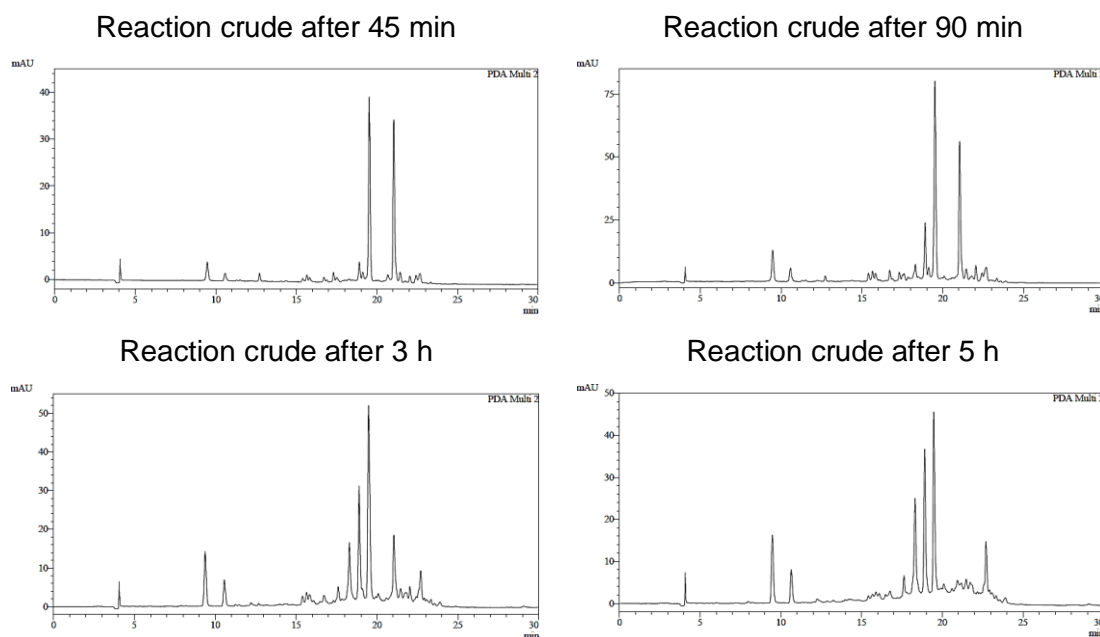
HPLC/MS details: 0-50 % B, traces shown at 280 nm



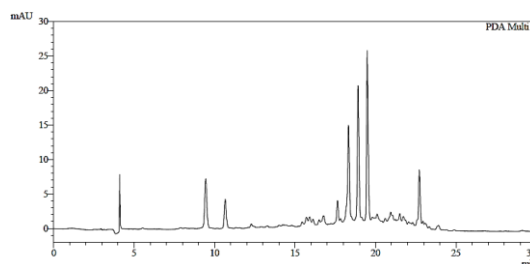
2) Reaction with peptide 4.4

Peptide 4.4 (50 nmol, 0.5 mM concentration) was mixed with 1.1 equiv. of 3.1 and left reacting in water at 60 °C. HPLC Analyses were performed at 45 min, 90 min, 3, 5 and 7 h reaction times. An additional HPLC/MS analysis was performed at 7 h.

HPLC details: 10-50 % B, traces shown at 280 nm



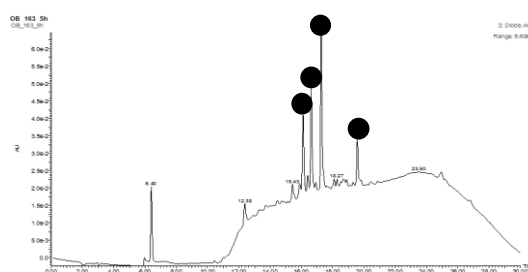
Reaction crude after 7 h



Product retention times (min):

4.5: not detected; **4.6**: 19.5 + 21.1; **4.7**: 18.9 + 19.5 and 18.3 + 22.7 for the theorised epimerised counterpart.

HPLC/MS details: 0-50 % B, traces shown at 280 nm



Product code:

●: products with the mass and UV-Vis profile of **4.3**

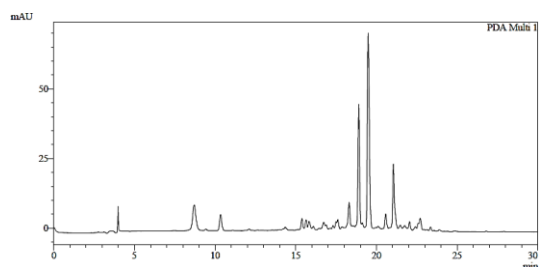
3) Reaction between **4.4** and **3.1** using TEMPO - 1

Peptide **4.4** (100 nmol, 0.5 mM concentration) was incubated with 1.1 equiv. of **3.1** in water at 60 °C for 2 h, to ensure the practically complete formation of **4.6** (checked by HPLC, data not shown). At this point temperature was lowered to 37 °C and the reaction split into four aliquots (A, B, C and D) to test the effect of different amounts of TEMPO. 0.20 Equiv. of TEMPO were added to aliquot A, 1.0 equiv. to aliquot B, 5.0 equiv. to aliquot C and, finally, 10.0 equiv. to aliquot D. Each aliquot was diluted until peptide concentration was 0.1 mM. Each reaction was analysed by HPLC after 1 and 2 h (except aliquot C, which was analysed after 1 and 3 h by mistake). No important changes between these time points were found for any of the reactions, and therefore only the analyses at the last time point are shown. Aliquots A and D were also analysed by HPLC/MS after 2 h.

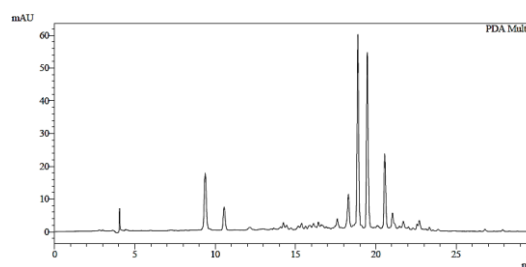
For comparison purposes, HPLC traces of aliquots A, B and C are shown. For clarity, only the HPLC/MS trace of aliquot D is shown (HPLC analysis gave puzzling results due to co-elution problems) and compared to that of aliquot A.

HPLC details: 10-50 % B, traces shown at 280 nm

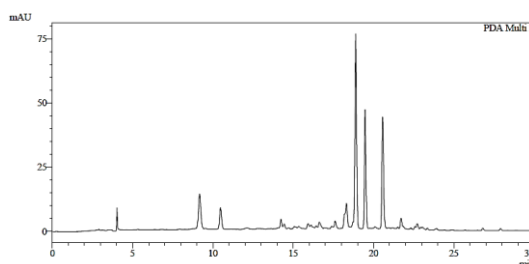
Aliquot A after 2 h



Aliquot B after 2 h



Aliquot C after 3 h

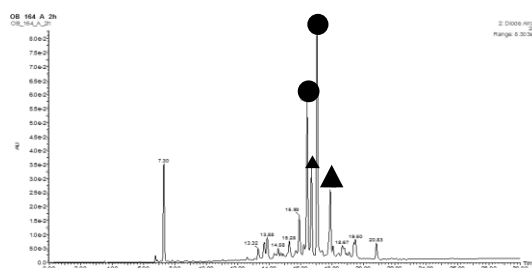


Product retention times (min):

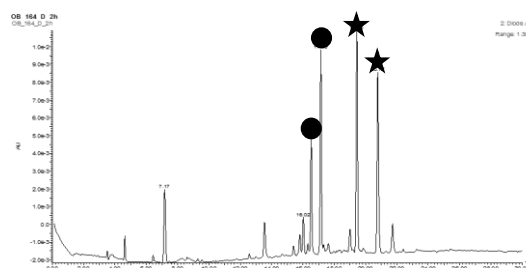
4.6: 19.5 + 20.8; **4.7:** 18.9 + 19.4

HPLC/MS details: 0-50 % B, traces shown at 280 nm

Aliquot A after 2 h



Aliquot D after 2 h



Product code:

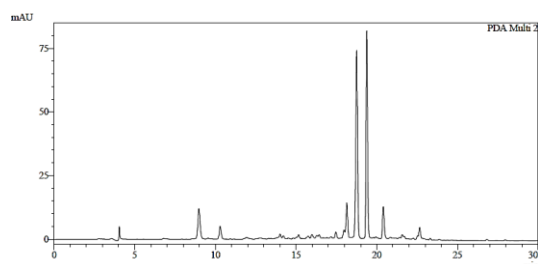
▲ : 4.6; **●** : 4.7; **★** : 4.8

4) Reaction between 4.4 and 3.1 using TEMPO - 2

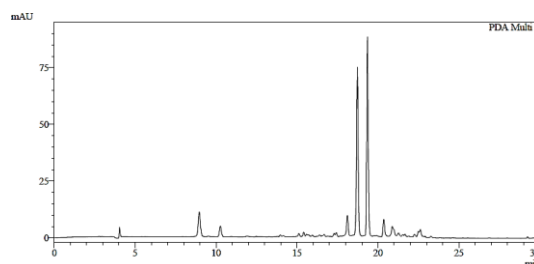
Peptide **4.4** (100 nmol, 0.5 mM concentration) was reacted with 1.2 equiv. of **3.1** in water for 90 min at 60 °C until the oxidation of **4.5** to **4.6** was almost complete (as assessed by HPLC, data not shown). At this point temperature was lowered to 37 °C and the reaction split into two different aliquots. 0.20 Equiv. of TEMPO were added to each aliquot, and peptide concentration of aliquot A was adjusted to 0.1 mM while that of aliquot B remained at 0.5 mM. Both aliquots received three extra TEMPO additions (again 0.20 equiv.) 1, 2 and 2.5 h after the first one, totalling 0.8 equiv. of reagent. Reaction of aliquot B was finished after 1 h, while that of aliquot A required 3 h.

HPLC details: 10-50 % B, traces shown at 280 nm

Aliquot A after 3 h



Aliquot B after 1 h



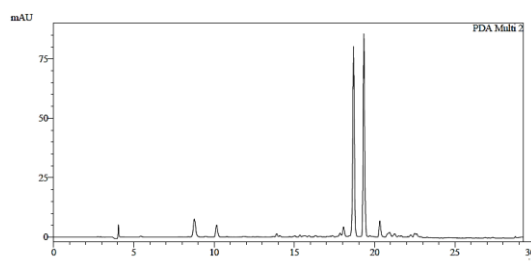
Product retention times (min):

4.6: 19.5 + 21.0; **4.7:** 18.8 + 19.4

4) Reaction between 4.4 and 3.1 using TEMPO - 3

Peptide **4.4** (100 nmol, 0.5 mM concentration) was reacted with 1.2 equiv. of **3.1** in water for 45 min at 60 °C, when the **4.5** to **4.6** ratio was approximately 1:1 (as assessed by HPLC, data not shown). At this point, temperature was lowered to 37 °C and 0.20 equiv. of TEMPO were added to the reaction mixture. HPLC Analysis 1 h later revealed that some unreacted **4.6** was still present in the crude, so 0.20 equiv. of TEMPO were again added and the reaction left running for an additional 30 min. Analysis after this time indicated completion of the reaction.

HPLC details: 10-50 % B, traces shown at 280 nm



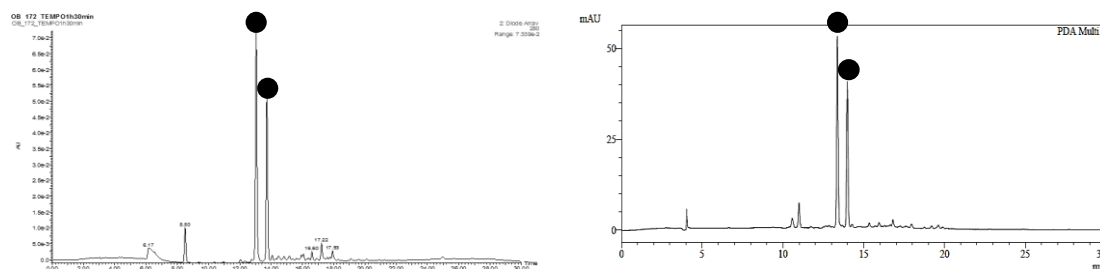
Product retention times (min):

4.6: completely consumed; **4.7:** 18.8 + 19.4

Exactly the same procedure was repeated for peptide **3.32**. HPLC/MS And HPLC traces are shown below.

HPLC/MS details: 0-50 % B. HPLC details: 10-50 % B. Traces shown at 280 nm

Reaction 1.5 h after the addition of TEMPO



Product code:

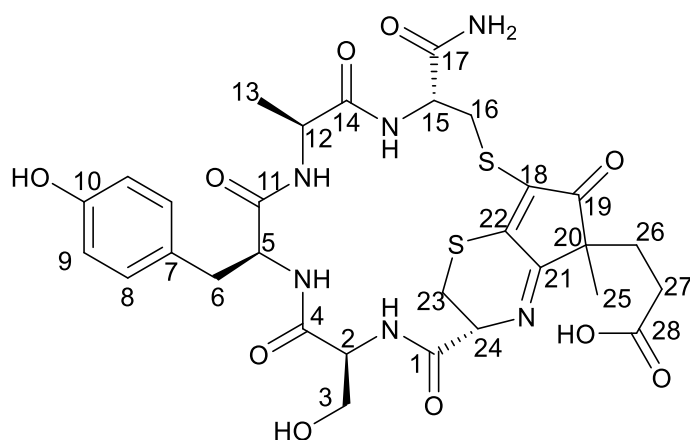
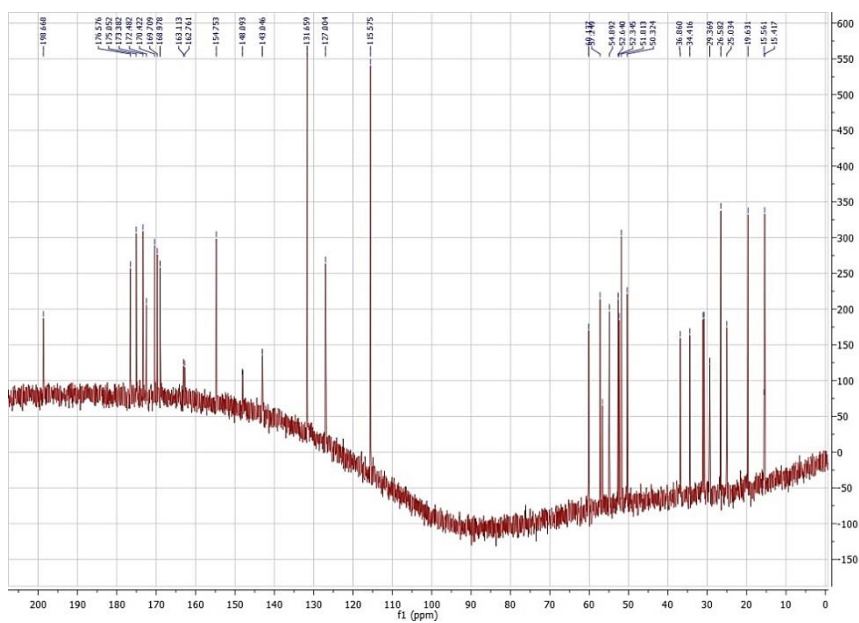
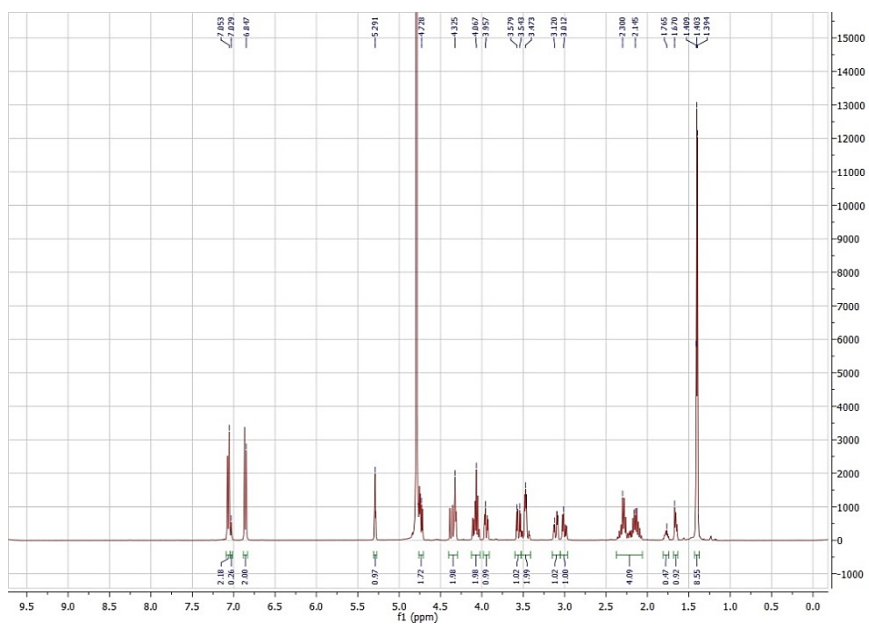
●: 4.3

E.4.3 Confirmation of the cyclic structure by NMR

Peptide **4.10** (1.0 mmol, 2.0 mM concentration) was incubated with 1.4 equiv. of **3.1** in water at 60 °C for 10 min. Subsequently, temperature was lowered to 37 °C and 0.20 equiv. of TEMPO (20 μ L of a 10.0 mM solution) were added to the reaction mixture. The reaction was left running for three hours with additions of 0.20 equiv. of TEMPO 1 and 2 h after the first one.

Twenty identical reactions were run in parallel to obtain enough amount of peptide to perform the NMR studies. All crudes were combined and lyophilised before being purified by HPLC, separating the two diastereomers (isomers 1 and 2, numbering based on the elution order on the HPLC) of **4.13**. The pure isomers, which then could not be quantified (the ϵ of the cyclic adducts still had not been determined), were directly subjected to NMR (^1H , ^{13}C , DEPT, COSY, HSQC and HMBC).

^1H And ^{13}C spectra are shown below.



Isomer 1		
Position	C δ (ppm)	H δ (ppm) and multiplicity
1	169.0	-
2	57.2	4.29-4.40 (m)
3	60.1	4.02-4.12 (m) + 3.95 (dd, $J_1 = 11.6$ Hz, $J_2 = 3.6$ Hz)
4	169.7	-
5	52.3	4.70-4.77 (m)
6	36.9	3.12 (dd, $J_1 = 14.0$ Hz, $J_2 = 3.6$ Hz) + 3.00 (dd, $J_1 = 14.0$ Hz, $J_2 = 6.0$ Hz)
7	127.0	-
8	131.7	7.06 (d, $J = 8.8$ Hz)
9	115.6	6.86 (d, $J = 8.8$ Hz)
10	154.8	-
11	170.4	-
12	50.3	4.02-4.12 (m)
13	15.4	1.37-1.43 (m)
14	175.0	-
15	52.6	4.70-4.77 (m)
16	30.8	4.29-4.40 (m) + 3.55 (dd, $J_1 = 14.0$ Hz, $J_2 = 4.4$ Hz)
17	173.4	-
18	148.1	-
19	198.7	-
20	51.8	-
21	172.5	-
22	143.0	-
23	25.0	3.42-3.52 (m)
24	54.9	5.29 (t, $J = 2.8$ Hz)
25	19.7	1.37-1.43 (m)
26	31.1	2.06-2.22 (m)
27	29.4	2.22-2.38 (m)
28	176.6	-

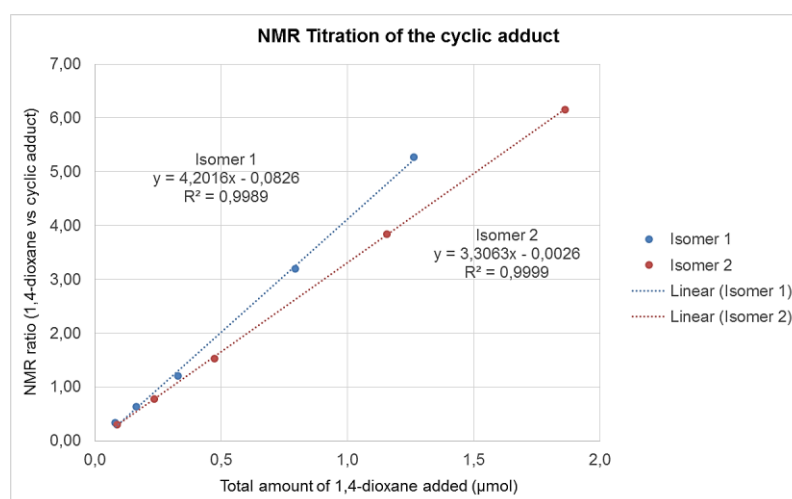
Isomer 2		
Position	C δ (ppm)	H δ (ppm) and multiplicity
1	168.9	-
2	57.2	4.33 (dd, $J_1 = 6.0$ Hz, $J_2 = 3.2$ Hz)
3	60.3	4.05-4.12 (m) + 3.93 (dd, $J_1 = 12.0$ Hz, $J_2 = 3.2$ Hz)
4	169.5	-
5	52.3	Under the HDO signal
6	37.0	3.11 (dd, $J_1 = 14.0$ Hz, $J_2 = 4.0$ Hz) + 3.00 (dd, $J_1 = 14.0$ Hz, $J_2 = 6.0$ Hz)
7	127.0	-
8	131.8	7.07 (d, $J = 8.8$ Hz)
9	115.5	6.86 (d, $J = 8.8$ Hz)
10	154.9	-
11	170.1	-
12	50.3	4.05-4.12 (m)
13	15.5	1.39 (d, $J = 7.2$ Hz)
14	175.4	-
15	52.7	Under the HDO signal
16	30.7	4.44 (dd, $J_1 = 14.0$ Hz, $J_2 = 11.2$ Hz) + 3.55 (dd, $J_1 = 14.0$ Hz, $J_2 = 4.0$ Hz)
17	173.5	-
18	150.0	-
19	198.1	-
20	51.7	-
21	173.0	-
22	143.0	-
23	25.2	3.44-3.54 (m)
24	54.9	5.31 (t, $J = 2.8$ Hz)
25	20.6	1.37 (s)
26	30.2	2.10-2.34 (m)
27	28.7	2.10-2.34 (m) + 2.46-2.56 (m)
28	176.9	-

E.4.4 Determination of the molar absorption coefficient

The already purified and characterised isomers 1 and 2 were dissolved in an approximately 1:1 (v/v) d_6 -DMSO/ D_2O mixture and titrated by NMR. 1,4-Dioxane was used as titrating agent. This was taken from the best quality available, passed through a basic alumina column to remove peroxides and stored over calcium hydride before use. For the titration process, carefully weighed amounts of a 1,4-dioxane solution in D_2O were sequentially added to each of the isomers' solutions, and the integration of both the cyclic adduct and the titrating agent measured. Integration of the 1,4-dioxane signal divided by that of the cyclic adduct was represented versus the total amount of 1,4-dioxane added. The slope of the line of best fit obtained after this representation corresponded to the number of H nuclei present in the 1,4-dioxane signal divided by that of the H nuclei present in the integrated cyclic adduct signal multiplied by the inverse of the amount of cyclic adduct. As in our case the integrated cyclic signal corresponded to a single H nucleus and the signal of 1,4-dioxane corresponds to 8 H nuclei, the amount of cyclic adduct was calculated by dividing 8 by the slope of the line of best fit.

As to the NMR details, the relaxation delay was set to 35 seconds to allow for the complete relaxation of all the H nuclei present in the sample. Both isomers were, after the titration process, quantitatively recovered from the NMR tube and lyophilised.

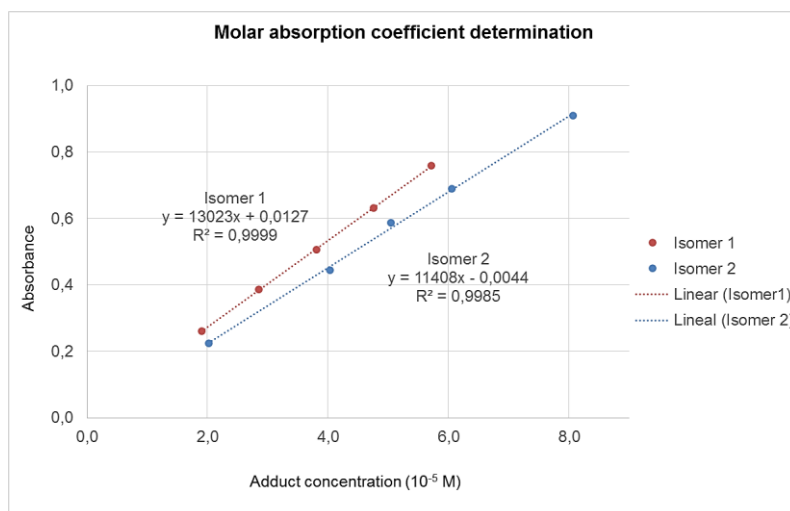
Data obtained are shown below. This allowed to quantify the amount of cyclic adduct as follows:



Amount of cyclic adduct:

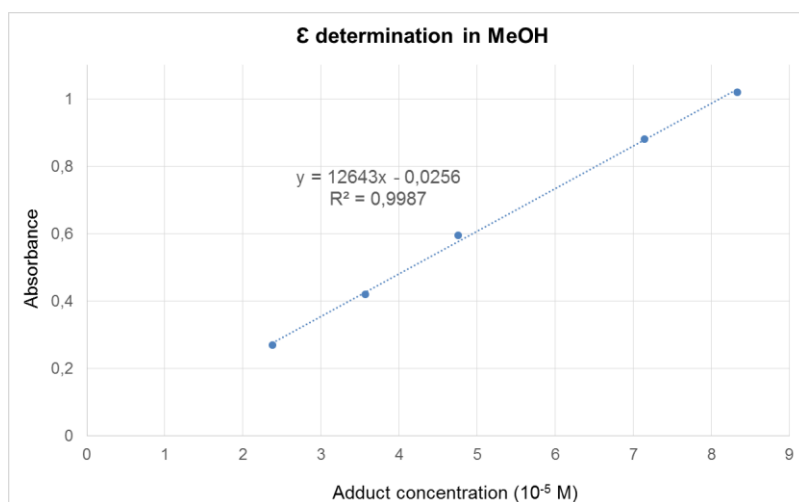
$$\text{Isomer 1: } \frac{8}{1} x \frac{1}{4.202} = 1.904 \text{ } \mu\text{mol} \quad \text{Isomer 2: } \frac{8}{1} x \frac{1}{3.306} = 2.419 \text{ } \mu\text{mol}$$

Once this had been done, each isomer was dissolved in DMSO and diluted with water to a known concentration, reaching a maximum DMSO value of 2 % (v/v), and the absorbance of these solutions was measured. Molar absorption coefficients in water were obtained from the slope of the line of best fit obtained when representing the absorbance of the different solutions versus their concentration (cuvette pathlength was 1 cm). The mean value, which was rounded to the closest five or ten, was calculated from the two independently obtained molar absorption coefficients. Data obtained are shown below.



$$\epsilon = 12215 \text{ M}^{-1} \text{ cm}^{-1}$$

To determine the molar absorption value in methanol only isomer 1 was used (the sample of isomer 2 was lost during its manipulation). This isomer was dissolved in methanol and the absorbance of different dilutions (in methanol) measured. Data obtained can be seen below. The slope of the line of best fit was, again, the molar absorption value, which was rounded to the closest five or ten.



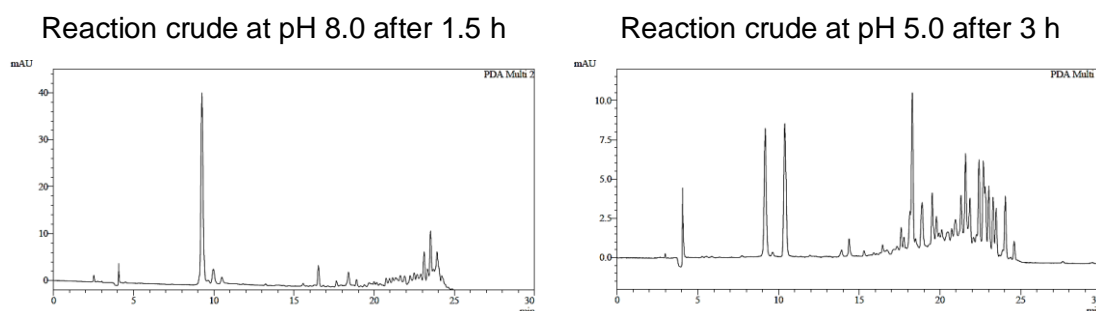
$$\epsilon = 12645 \text{ M}^{-1} \text{ cm}^{-1}$$

E.4.5 Search of alternatives to the use of TEMPO

1) Effect of different pHs in the cyclisation reaction

Peptide **4.4** (50 nmol) was dissolved both in a 0.1 M phosphate buffer at pH 8.0 and in a 0.1 M acetate buffer at pH 5.0, and mixed with 1.2 equiv. of **3.1** (final peptide concentration was 0.5 mM). Both reaction mixtures were left shaking at 60 °C and analysed after 1.5 or 3 h.

HPLC details: 10-50 % B, traces shown at 280 nm



Product retention times (min):

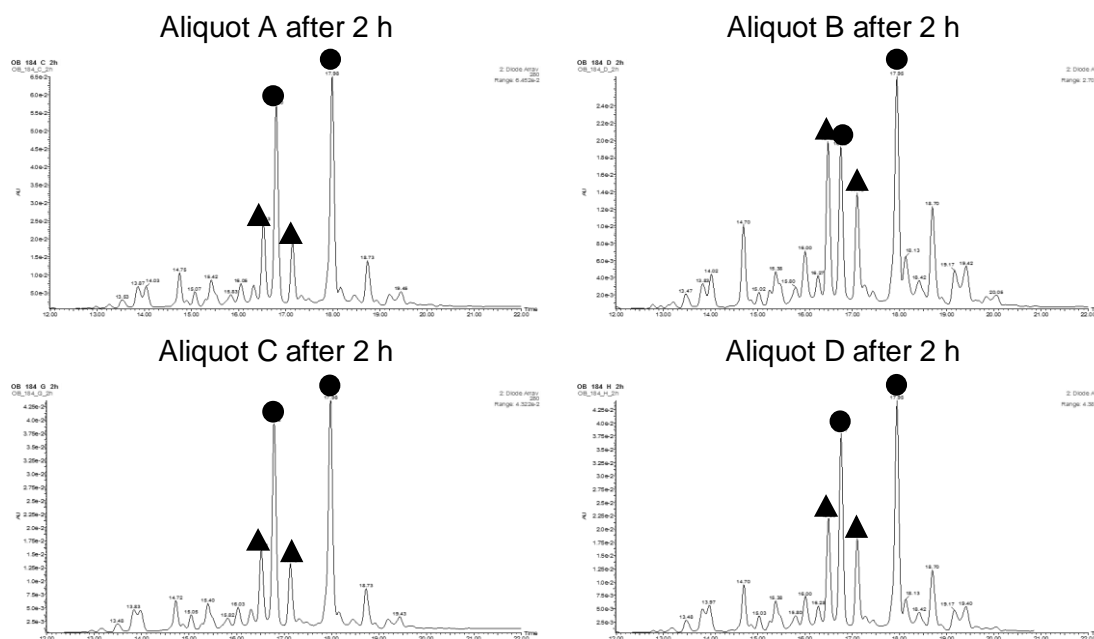
4.9: 9.3 min

2) Effect of different bases in the cyclisation reaction

Peptide **4.4** (200 nmol, 0.5 mM concentration) was incubated with 1.3 equiv. of **3.1** in water for 2.5 h at 60 °C. Analysis of the reaction crude showed the total disappearance of **4.5** and the formation of **4.6** together with a product with the same UV-Vis profile and mass that was inferred to arise from the epimerisation of the *N*-terminal cysteine of **4.6**

(10 %, data not shown). At this point, the reaction was split into four different aliquots. Aliquot A received 1.0 equiv. of NaOH, aliquot B 2.0 equiv. of NaOH, aliquot C 1.0 equiv. of NEt_3 and aliquot D 2.0 equiv. of NEt_3 . All aliquots were diluted with water until peptide concentration was 0.1 mM before incubating them at 37 °C for 2 h.

HPLC/MS details: 0-50 % B, traces shown at 280 nm (only from 12 to 22 min)



Product code:

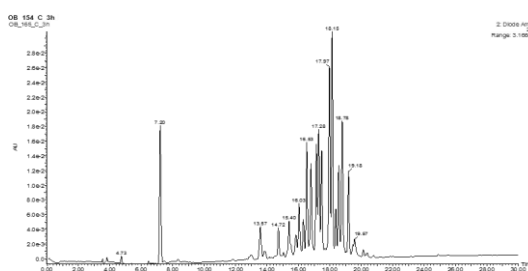
▲ : Proposed epimerisation; ● : 4.6

3) Effect of DMSO in the cyclisation reaction

Peptide **4.4** (100 nmol, 0.5 mM concentration) was reacted with 1.2 equiv. of **3.1** in water for 90 min at 60 °C until the oxidation of **4.5** to **4.6** was almost complete (as assessed by HPLC, data not shown). At this point temperature was lowered to 37 °C and DMSO and water were added so that the final peptide concentration was 0.1 mM and the amount of DMSO 10 % (v/v). The crude was analysed after 3 h by HPLC/MS.

HPLC/MS details: 0-50 % B, trace shown at 280 nm

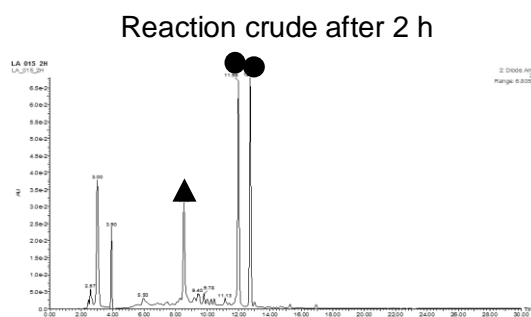
Reaction crude 3 h after the addition of DMSO



4) Effect of O₂ in the cyclisation reaction

In a first experiment, peptide **4.4** (100 nmol, 0.5 mM concentration) and 1.2 equiv. of **3.1** were dissolved in a 2 molal LiCl solution in water (see section **4.8**). Oxygen was bubbled for 10 min through the reaction mixture and an oxygen atmosphere maintained during the reaction process (using an oxygen balloon). The reaction was run for 2 h at 37 °C.

HPLC/MS details: 10-50 % B, trace shown at 280 nm

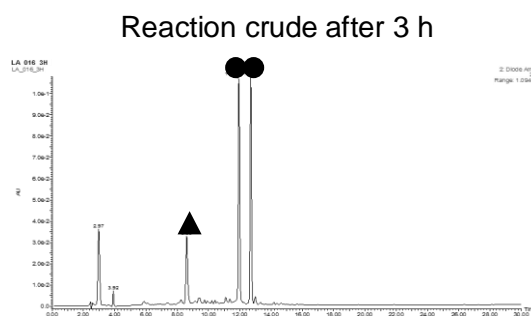


Product code:

▲ : Unknwon product; ● : **4.7**

In a second experiment the same procedure was followed but oxygen was only bubbled for 2 min and no efforts to maintain an oxygen atmosphere were performed, so the reaction vessel was closed without an oxygen balloon. This time the reaction needed 3 h to reach completion.

HPLC/MS details: 10-50 % B, trace shown at 280 nm



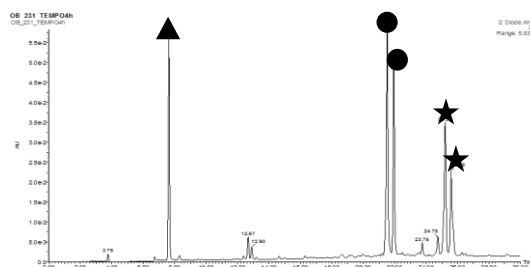
Product code:

▲ : Unknwon product; ● : **4.7**

Finally, a third experiment in which air was bubbled through the reaction mixture for 10 min was carried out (other parameters were kept the same).

One hour after this first addition, 0.20 equiv. of TEMPO were added every 30 min until 4 h had passed since the first addition. Reaction had not finished at that time, but no more TEMPO additions or analyses were performed.

HPLC/MS details: 0-50 % B, trace shown at 280 nm
Reaction crude 4 h after the first addition of TEMPO



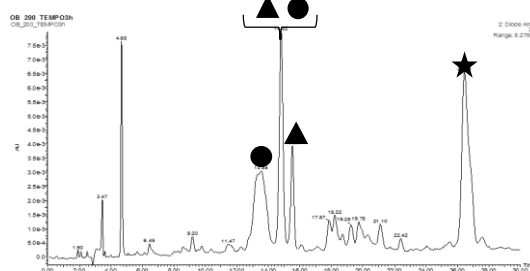
Product code:

▲ : 4.9; ● : 4.19; ★ : Product of thiol oxidation

3) Reaction between 4.20 and 3.1

Peptide **4.20** (100 nmol, 0.5 mM concentration) was incubated with 1.6 equiv. of **3.1** in water at 60 °C for 1 h. Subsequently, temperature was lowered to 37 °C, 0.20 equiv. of TEMPO were added, and the reaction was left running for 1 h. After this time, 0.20 equiv. of TEMPO were added every 30 min.

HPLC/MS details: 25-35 % B, trace shown at 280 nm
Reaction crude 3 h after the first addition of TEMPO



Product code:

● : 4.23; ★ : Internal cysteine oxidation by-product; ▲ : 4.22

E.4.7 Effect of aromatic residues close to the *N*-terminal cysteine

Peptides **4.10** and **4.24** (100 nmol, 0.5 mM concentration) were independently incubated with 1.3 equiv. of **3.1** in water at 37 °C. 0.20 Equiv. of TEMPO were added to each reaction mixture at t=0, 60, 90, 120 and 150 min. HPLC/MS Analyses were performed after 30, 60, 120, 180 and 240 min reaction times. These experiments were run in duplicate, and in one of the experiments involving peptide **4.24**, an additional TEMPO addition at 240 min and an HPLC/MS analysis after a total of 300 min were performed to prove, again, the need for continuous TEMPO additions. A third experiment was run to follow the progression of the early stages of the reaction. It was performed as the other two, but it was only carried out over 120 min.

Those peaks corresponding to products **4.11**, **4.12**, **4.13**, **4.25**, **4.26**, **4.27**, **4.28**, **4.29** and **4.30** were integrated at 280 nm to evaluate and compare both reactions.

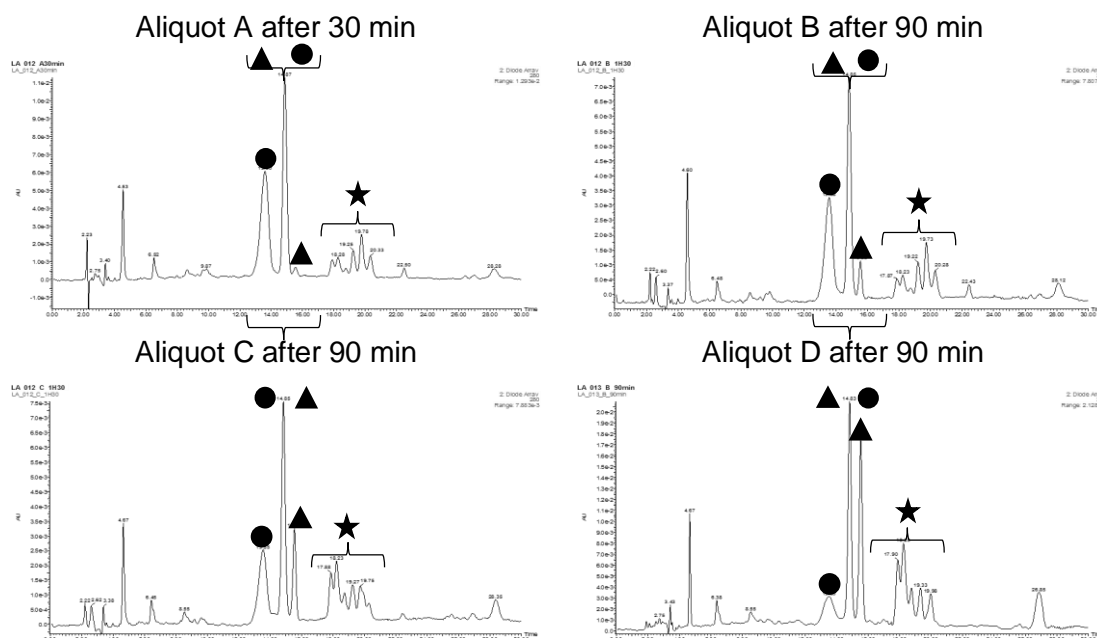
Due to co-elution problems, the ratio between **4.11** and **4.12** could not be determined. Data obtained are shown in the following table.

Time	Product percentages							
	Reactions of 4.10			Reactions of 4.24				
	4.11+4.12	4.28	4.13	4.25+4.26	4.25/4.26	4.29	4.27	4.30
30min	79	14	7	91	38/55	5	2	n.d
	38	17	45	92	41/51	5	3	n.d
	90	7	3	91	35/56	6	3	n.d
1 h	43	21	36	91	31/60	5	4	n.d
	42	20	38	92	31/61	5	3	n.d
	56	24	20	91	28/63	6	3	n.d
2 h	19	10	71	66	0/66	7	27	n.d
	17	10	73	47	0/47	1	52	n.d
	18	12	70	57	0/57	5	38	n.d
3 h	5	3	92	22	0/22	0	68	10
	8	6	86	17	0/17	0	69	14
4 h	0	0	100	19	0/19	0	70	11
	0	0	100	15	0/15	0	70	15
5 h	n.d.	n.d.	n.d.	0	0/0	0	77	23

E.4.8 Effect of lithium chloride on the reaction progress

Peptide **4.20** (500 nmol, 0.5 mM concentration) and 2 equiv. of **3.1** were reacted for 45 min at 60 °C, at which time the **M-18/M-20 Da** adducts ratio was almost 1:1 (as seen by HPLC/MS, data not shown). After this, temperature was lowered to 37 °C and the reaction was split into four aliquots. Lithium chloride was added to three of them to reach a final 5 molal concentration (aliquot A), a 2 molal concentration (aliquot B) and a 1 molal concentration (aliquot C). No LiCl was added to the fourth aliquot (aliquot D) to serve as control. At the same time, 0.20 equiv. of TEMPO were added to each of the aliquots (also aliquot D) prior to be left shaking at 37 °C.

HPLC/MS details: 25-35 % B, traces shown at 280 nm



Product code:

● : **4.23**; ★ : Double addition products; ▲ : **4.22**

Aliquot:	A ^a (5 molal LiCl)	B ^b (2 molal LiCl)	C ^b (1 molal LiCl)	D ^b (no LiCl)
4.23/4.22 ratio:	97:3	88:12	70:30	35:65

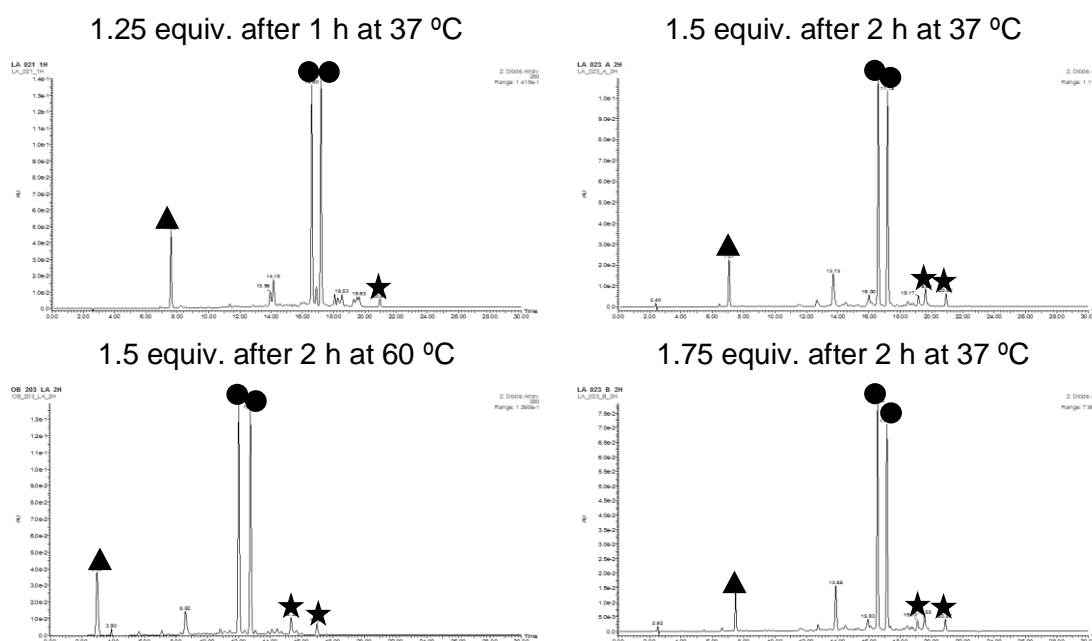
Relative **4.23/4.22** ratios at different reaction times were determined by HPLC/MS (280 nm) on the basis of the integrations of the **4.23** and **4.22** peaks that did not co-elute, *in lieu* of using the integration of the two HPLC peaks of **4.23** and the two of **4.22**. ^aAnalysed 30 min after the addition of TEMPO. ^bAnalysed 90 min after the addition of TEMPO.

E.4.9 Optimisation of the cyclisation procedure

Peptide **4.4** (100 nmol, 0.5 mM concentration), TEMPO (0.20 equiv.) and CPD **3.1** (1.25, 1.50 or 1.75 equiv.) were dissolved in an aqueous 2 molal LiCl solution and reacted at 37 °C. After 1 h, 0.20 equiv of TEMPO were added every 30 min. Crudes were analysed after 1 or 2 h by HPLC/MS. An additional experiment using 1.50 equiv. of **3.1** was run at 60 °C. Results obtained are summarised below.

	Cyclic disulfide (%)	Unknown by-products (%)	Cyclic adduct (%)	Thiol oxidation by-product (%)
1.25 equiv. (37 °C)	11	13	69	5
1.50 equiv. (37 °C)	7	12	74	6
1.50 equiv. (60 °C)	14	6	75	5
1.75 equiv. (37 °C)	6	13	75	6

HPLC/MS details: 0-50 % B when temperature was 37 °C, 10-50 % B when temperature was 60 °C, traces shown at 280 nm



Product code:

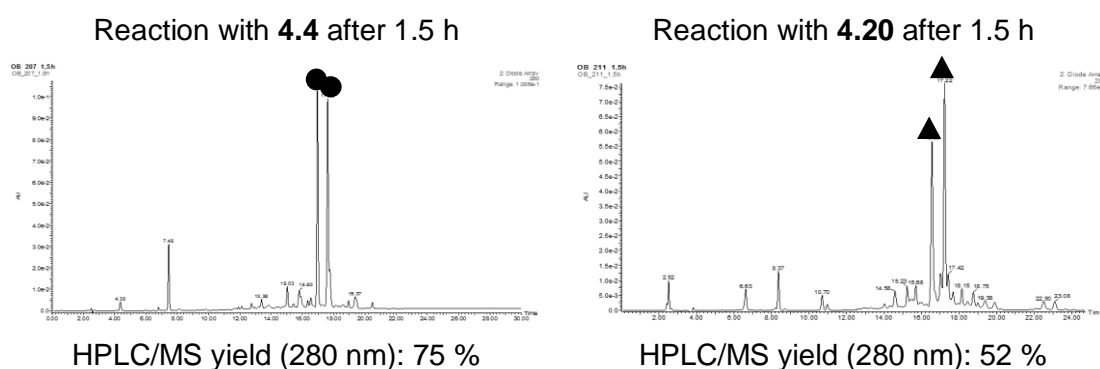
● : 4.7; ★ : 4.8; ▲ : 4.9

E.4.10 Use of labelling CPDs in cyclisation reactions - 2. Epimerisation of the *N*-terminal cysteine

1) Reactions involving CPD 3.14 at the 100 nmol scale in water

Peptides **4.4** and **4.20** (100 nmol, 0.5 mM concentration) were independently mixed with 1.50 equiv. of **3.14** and 0.20 equiv. of TEMPO in 2 a molal LiCl solution in water. Reaction mixtures were left shaking at 60 °C for 1 h, after which time 0.2 equiv. of TEMPO were again added. HPLC/MS analyses were performed after 1.5 h.

HPLC/MS details: 0-50% B for **4.4** and 10-60 % B (60 °C) for **4.20**, traces shown at 280 nm



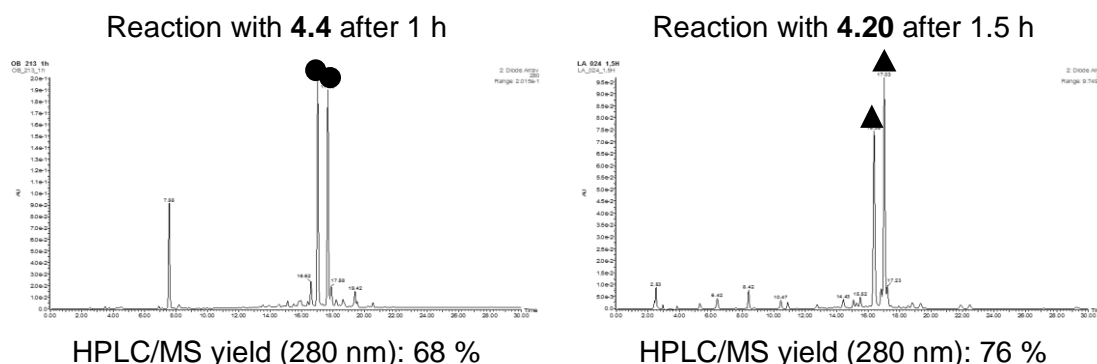
Product code:

●: **4.16**; ▲: **4.31**

2) Reactions involving CPD 3.14 at the 100 nmol scale in 1:1 MeOH/H₂O

Peptides **4.4** and **4.20** (100 nmol, 0.5 mM concentration) were independently mixed with 1.50 equiv. of **3.14** and 0.20 equiv. of TEMPO in a 2 molal LiCl solution in a 1:1 water/methanol mixture. Reaction mixtures were left shaking at 60 °C for 1 h, after which time 0.2 equiv. of TEMPO were again added. HPLC/MS analysis were performed after 1 and 1.5 h (only those in which the reaction had evolved to completion are shown).

HPLC/MS details: 0-50% B for **4.4** and 10-60 % B (60 °C) for **4.20**, traces shown at 280 nm



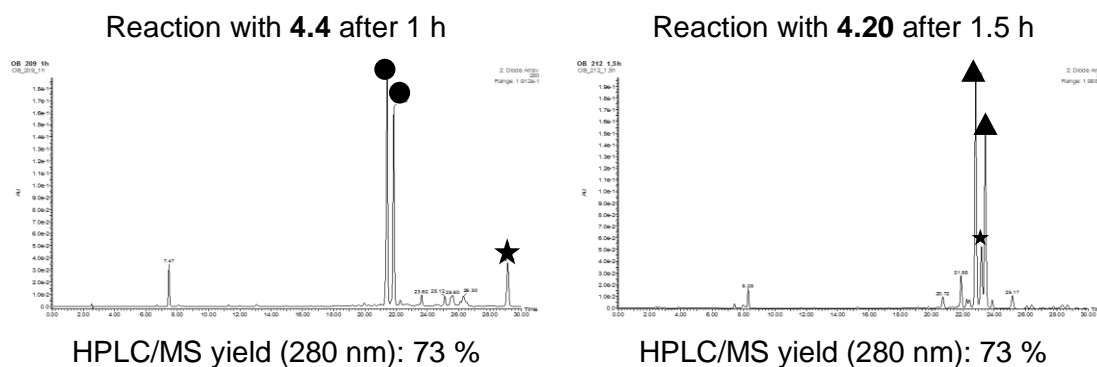
Product code:

● : **4.16**; ▲ : **4.31**

3) Reactions involving CPD **3.15** at the 100 nmol scale in 1:1 MeOH/H₂O

Peptides **4.4** and **4.20** (100 nmol) were independently mixed with 1.50 equiv. of **3.15** and 0.20 equiv. of TEMPO in 2 molal LiCl solution in a 1:1 water/methanol. Reaction mixtures were left shaking at 60 °C for 1 h, after which time 0.2 equiv. of TEMPO were again added. HPLC/MS Analysis were performed after 1 and 1.5 h (only those where the reaction had evolved to completion are shown).

HPLC/MS details: 0-50% B for **4.4** and 10-60 % B (60 °C) for **4.20**, traces shown at 280 nm



Product code:

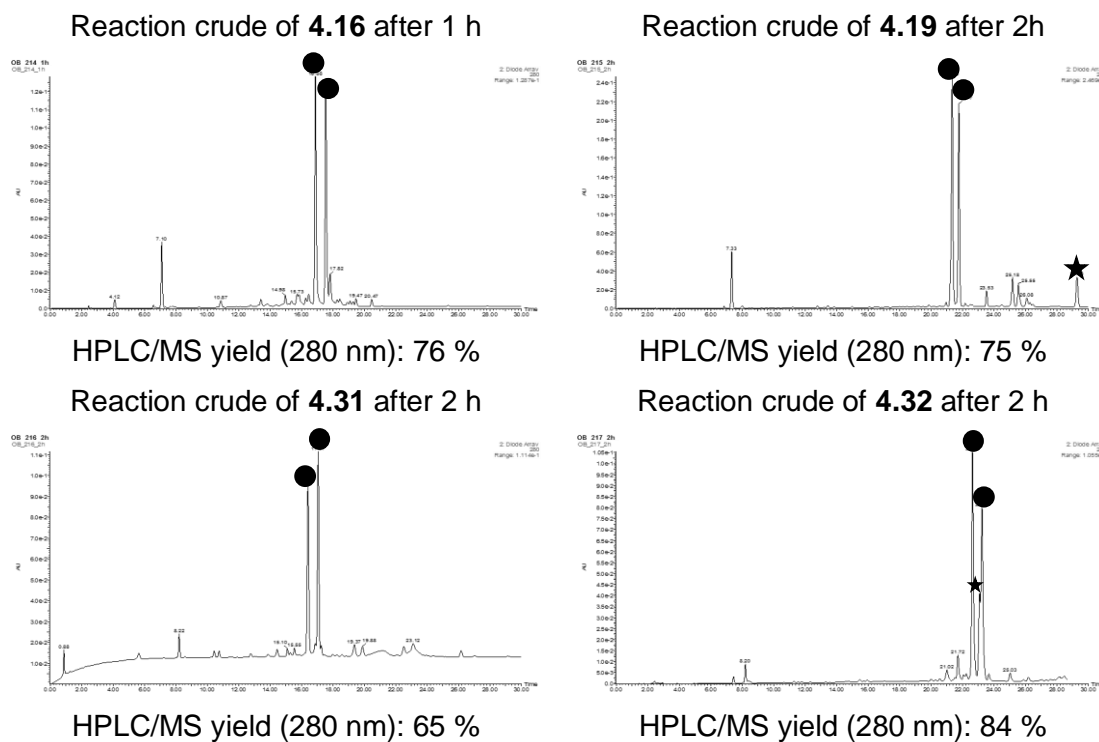
● : **4.19**; ▲ : **4.32**; ★ : Excess **3.15**

4) Repetition of the reactions at the 500 nmol scale

Products **4.16**, **4.19**, **4.31** and **4.32** were synthesised again at the 500 nmol scale using the same procedure as described above. Water was used as solvent in the synthesis of **4.16** and the 1:1 water/methanol mixture in the other syntheses. All reactions required 2 h to reach completion, except that of **4.16**, which ended after 1 h. Each product was

purified and their isomers independently isolated by HPLC (numbered on the basis of their elution order in the HPLC).

HPLC/MS details: 0-50% B for **4.16** and **4.19** and 10-60 % B (60 °C) for **4.31** and **4.32**, traces shown at 280 nm



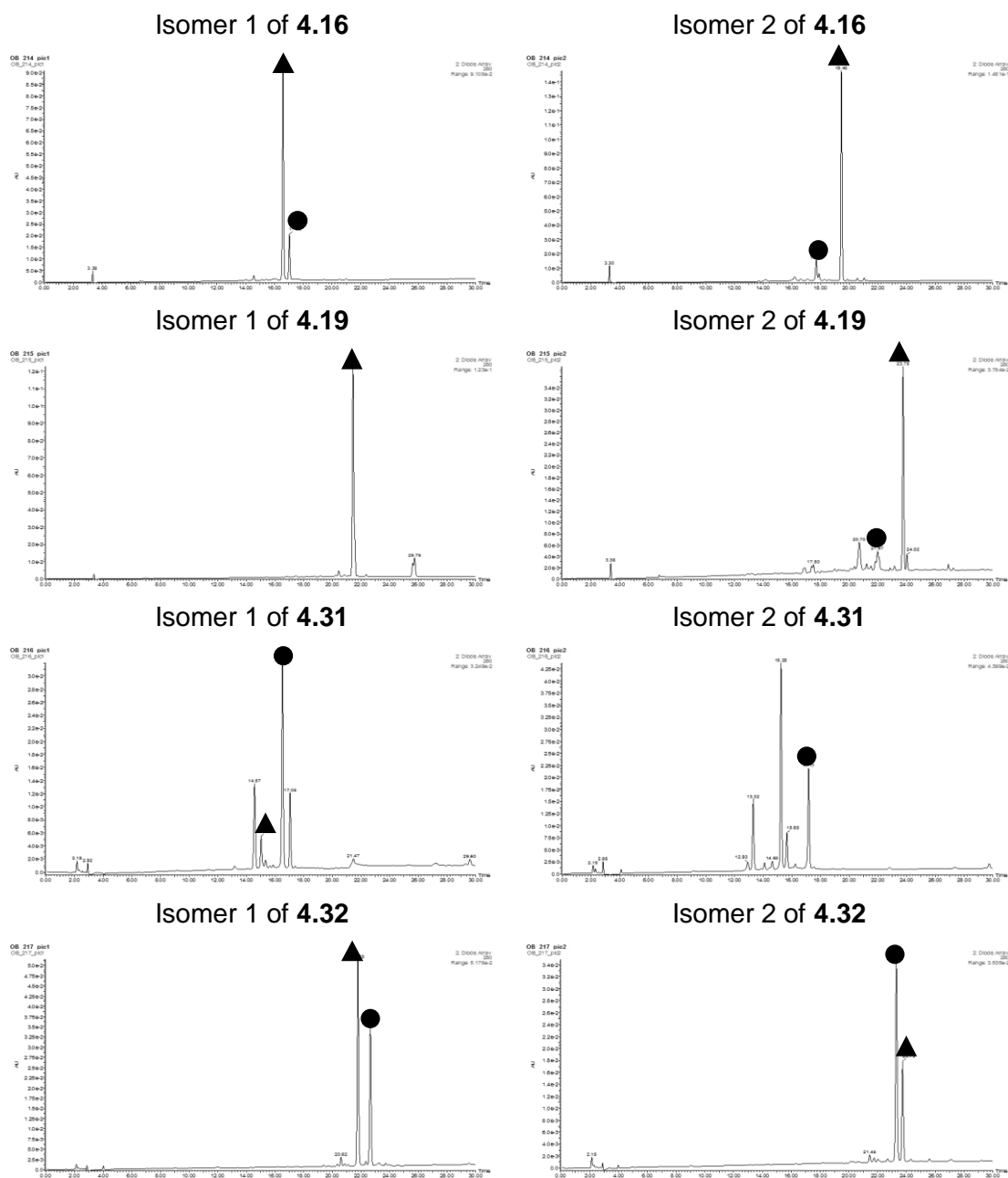
Product code:

● : Cyclic adduct; ★ : Excess **3.15**

5) Observation of the “decomposition” of the products

Products **4.16**, **4.19**, **4.31** and **4.32** were dissolved in methanol after their purification and reanalysed. HPLC/MS Traces showed the formation of a product with a different retention time (not that of the other isolated isomer) and with the same mass and UV-Vis profile.

HPLC/MS details: 0-50% B for **4.16** and **4.19** and 10-60 % B (60 °C) for **4.31** and **4.32**, traces shown at 280 nm. Isomer numbering performed on the basis of their elution order in the HPLC



Product code:

● : Desired product

▲ : Product hypothesised to arise from the epimerisation of the *N*-terminal Cys

E.4.11 Confirmation of *N*-terminal cysteine epimerisation

1) Study of the processes that could have lead to the imagined epimerisation

Product **4.16** was synthesised again and purified by HPLC (separating both isomers).

Each isomer, immediately after its elution from the HPLC, was:

- 1) Left in the purification medium for 24 h at room temperature and reanalysed by HPLC/MS.
- 2) Subjected to rotary evaporation (to dryness) and reanalysed by HPLC/MS (immediately after being redissolved in methanol).
- 3) Lyophilised and reanalysed by HPLC/MS (immediately after being redissolved in methanol).
- 4) Lyophilised and reanalysed by HPLC/MS after 24 h incubation in methanol.
- 5) Left in a 1:1 DMSO/purification medium mixture for 24 h and reanalysed by HPLC/MS.
- 6) Brought to pH 7.4, using a phosphate buffer, and analysed by HPLC/MS after 12 and 24 h.
- 7) Brought to pH 9.0, using diluted NaOH, and analysed by HPLC/MS after 20 min.

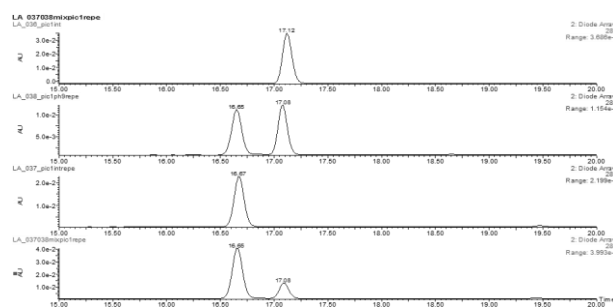
A small aliquot of the reaction crude was reanalysed after 24 h at room temperature.

Only after incubation at pH 7.4 or pH 9.0 the product was found to have been altered.

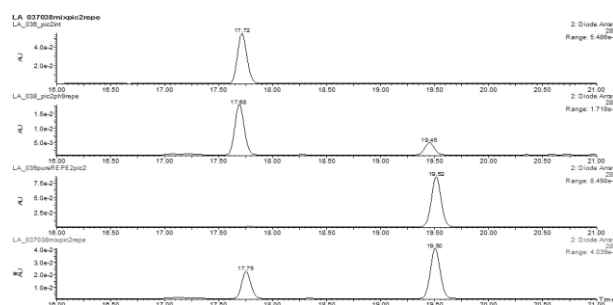
2) Confirmation of the hypothesised isomerisation by independent synthesis of the epimer at the *N*-terminal position

Product **4.34** was synthesised following the procedures described in section **E.4.10** and its two isomers independently purified and isolated. Their retention times in the HPLC/MS were compared with those of the products proposed to arise from epimerisation of the *N*-terminal cysteine of **4.16**. Also, each isomer of **4.16**, after being incubated at pH 9.0 (which induced the imagined epimerisation), was coinjected in the HPLC/MS with a pure sample of the corresponding isomer of **4.34**.

Isomer 1:



Isomer 2:



From top to bottom: pure **4.16**, **4.16** at pH 9.0, pure **4.34**, and coinjection of the crude obtained after incubation of **4.16** at pH 9.0 and **4.34**.

E.4.12 Comparison of the stability of the D and L epimeric cyclic peptides

Both isomers of **4.16** and **4.34** (independently isolated), just collected from the HPLC, were brought to pH 7.4 and pH 9.0, and their integrity was checked by HPLC/MS after 12 and 24 h at pH 7.4 and after 20 min at pH 9.0. Epimerisation percentages are shown below, as determined by integration of the corresponding peaks at 280 nm.

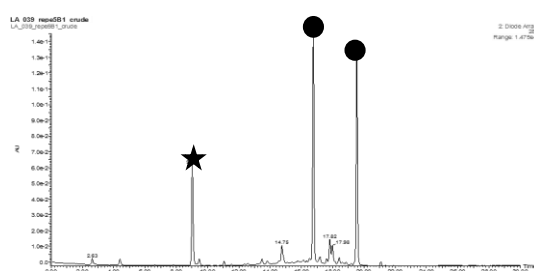
Product	Epimerisation percentage		
	pH 7.4 12 h	pH 7.4 24 h	pH 9.0 20 min
4.16 isomer 1	33	45	64
4.16 isomer 2	16	21	35
4.34 isomer 1	6	9	5
4.34 isomer 2	8	14	5

E.4.13 Use of labelling CPDs in cyclisation reactions - 3

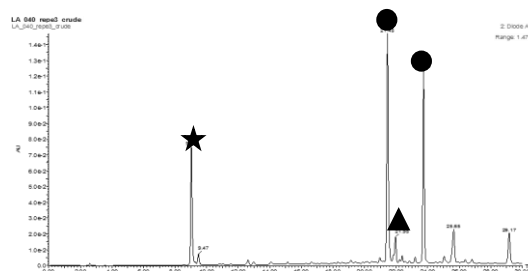
The syntheses of **4.37** and **4.38** was performed as described in part 4 of section **E.4.10**. Both reactions required 2 h to reach completion. In the case of products **4.34** and **4.35**, the synthesis protocol was slightly modified so that the first TEMPO addition was performed after the reaction had run for 30 min. They required 1.5 and 2.5 h to reach completion, respectively.

HPLC/MS details: 0-50% B for **4.34** and **4.35** and 10-60 % B (60 °C) for **4.37** and **4.38**, traces shown at 280 nm

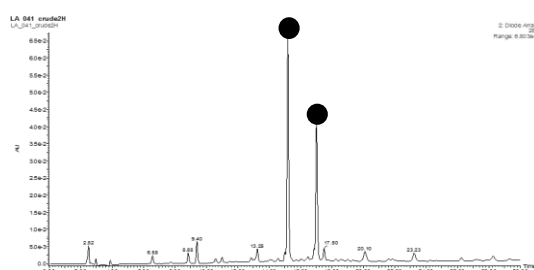
Reaction crude of **4.34** after 1.5 h



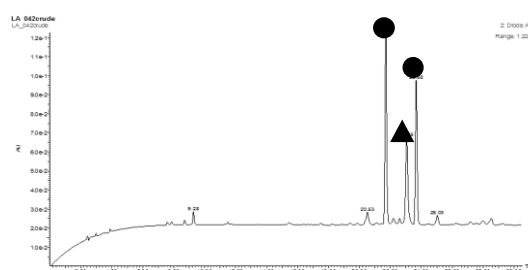
Reaction crude of **4.35** after 2.5 h



Reaction crude of **4.37** after 2 h



Reaction crude of **4.38** after 2 h



Product code:

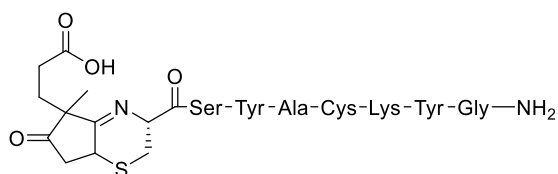
● : Cyclic adduct; ▲ : Excess **3.15**; ★ : Cyclic disulfide of **4.33**

Product	HPLC/MS yield (% , 280 nm)	Isolated Yield (%)
4.33 + 3.14 → 4.34	76	40
4.33 + 3.15 → 4.35	66	40
4.36 + 3.14 → 4.37	77	33
4.36 + 3.15 → 4.38	82	45

E.4 Characterisation of the products

Product: **4.1**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 15.5 + 16.7$ min

MS characterisation:

Technique: ESI, positive mode

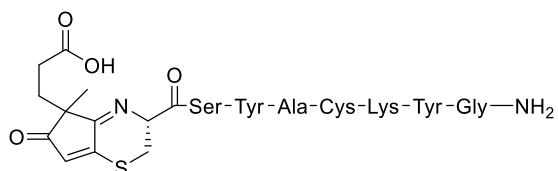
Mass calculated: C₄₇H₆₄N₁₀O₁₄S₂ 1056.4

Mass found: m/z 1057.7 [M+H]⁺

Purification conditions: -

Product: **4.2**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 14.0 + 16.0$ min

MS characterisation:

Technique: ESI, positive mode

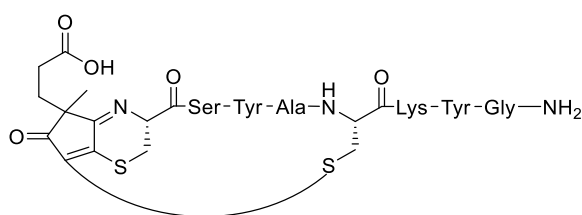
Mass calculated: C₄₇H₆₂N₁₀O₁₄S₂ 1054.4

Mass found: m/z 1055.6 [M+H]⁺

Purification conditions: -

Product: **4.3**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50 % B, $t_R = 13.1 + 13.8$ min

MS characterisation:

Technique: ESI, positive mode

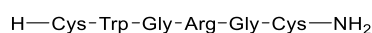
Mass calculated: $C_{47}H_{60}N_{10}O_{14}S_2$ 1052.4

Mass found: m/z 1053.6 $[M+H]^+$

Purification conditions: -

Product: **4.4**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-30 % B, $t_R = 8.5$ min

MS characterisation:

Technique: ESI, positive mode

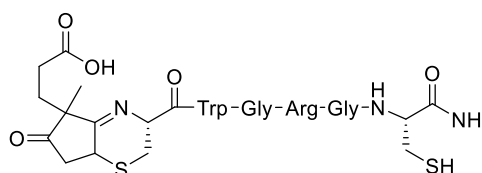
Mass calculated: $C_{27}H_{41}N_{11}O_6S_2$ 679.3

Mass found: m/z 680.0 $[M+H]^+$

Purification conditions: 15-30% B, $t_R = 8.7$ min

Product: **4.5**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 17.7 + 18.1$ min

MS characterisation:

Technique: ESI, positive mode

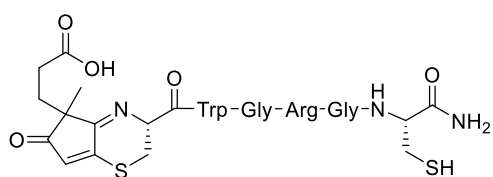
Mass calculated: $C_{36}H_{49}N_{11}O_9S_2$ 843.3

Mass found: m/z 844.3 $[M+H]^+$

Purification conditions: -

Product: **4.6**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 16.7 + 17.9$ min

MS characterisation:

Technique: ESI, positive mode

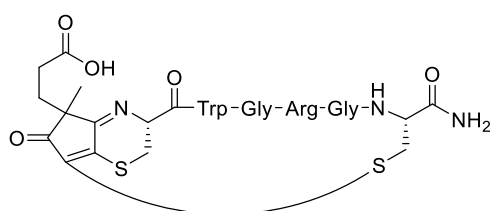
Mass calculated: $C_{36}H_{47}N_{11}O_9S_2$ 841.3

Mass found: m/z 842.5 $[M+H]^+$

Purification conditions: -

Product: **4.7**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 16.3 + 17.0$ min

MS characterisation:

Technique: ESI, positive mode

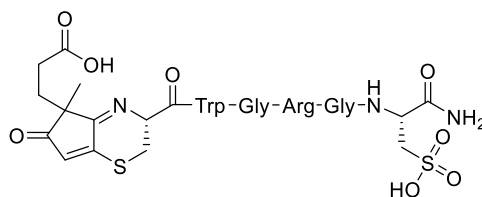
Mass calculated: $C_{36}H_{45}N_{11}O_9S_2$ 839.3

Mass found: m/z 840.0 $[M+H]^+$

Purification conditions: -

Product: **4.8**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 19.4 + 20.8$ min

MS characterisation:

Technique: ESI, positive mode

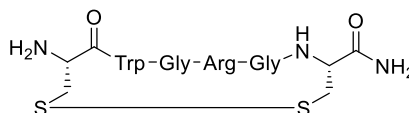
Mass calculated: $C_{36}H_{47}N_{11}O_{12}S_2$ 889.3

Mass found: m/z 890.4 $[M+H]^+$

Purification conditions: -

Product: **4.9**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 7.2$ min

MS characterisation:

Technique: ESI, positive mode

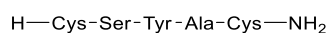
Mass calculated: $C_{27}H_{39}N_{11}O_6S_2$ 677.3

Mass found: m/z 678.3 $[M+H]^+$

Purification conditions: -

Product: **4.10**

Structure:



HPLC/MS characterisation

Gradient and retention time: 0-30%B, $t_R = 9.8$ min

MS characterisation

Technique: ESI, positive mode

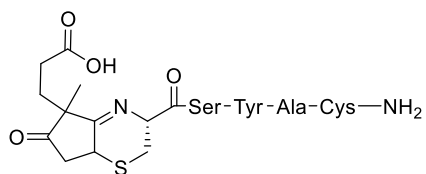
Mass calculated: $C_{21}H_{32}N_6O_7S_2$ 544.2

Mass found: m/z 544.8 $[M+H]^+$

Purification conditions: 15-50 % B, $t_R = 7.1$ min

Product: **4.11**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 19.7 + 21.5$ min

MS characterisation:

Technique: ESI, positive mode

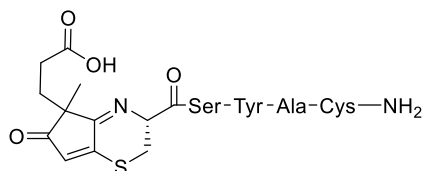
Mass calculated: $C_{30}H_{40}N_6O_{10}S_2$ 708.2

Mass found: m/z 709.0 [M+H]⁺

Purification conditions: -

Product: **4.12**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 19.7 + 20.3$ min

MS characterisation:

Technique: ESI, positive mode

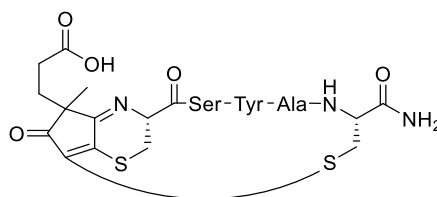
Mass calculated: $C_{30}H_{38}N_6O_{10}S_2$ 706.2

Mass found: m/z 707.0 [M+H]⁺

Purification conditions: -

Product: **4.13**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 15.3 + 16.3$ min

MS characterisation:

Technique: ESI, positive mode

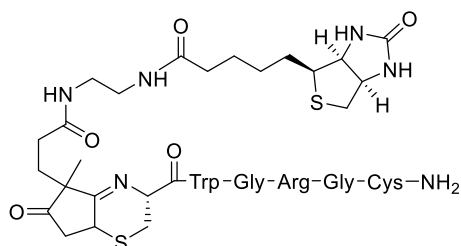
Mass calculated: $C_{30}H_{36}N_6O_{10}S_2$ 704.2

Mass found: m/z 705.0 $[M+H]^+$

Purification conditions: 15-30% B, $t_R = 9.0 + 10.4$ min

Product: **4.14**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 18.1$ min

MS characterisation:

Technique: ESI, positive mode

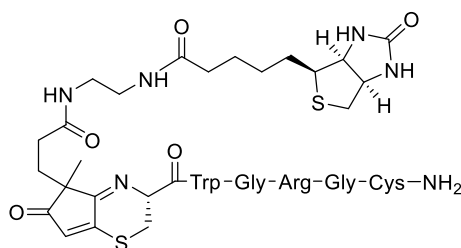
Mass calculated: $C_{48}H_{69}N_{15}O_{10}S_3$ 1111.5

Mass found: m/z 1112.6 $[M+H]^+$

Purification conditions: -

Product: **4.15**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 17.4 + 18.2$ min

MS characterisation:

Technique: ESI, positive mode

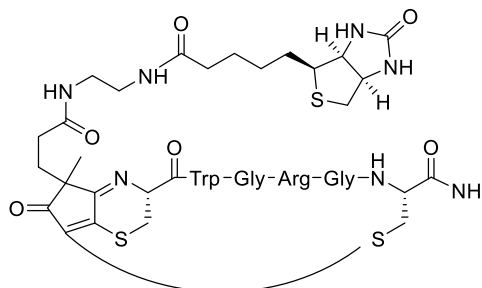
Mass calculated: $C_{48}H_{67}N_{15}O_{10}S_3$ 1109.4

Mass found: m/z 1110.4 $[M+H]^+$

Purification conditions: -

Product: **4.16**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 16.9 + 17.5$ min

MS characterisation:

Technique: ESI, positive mode

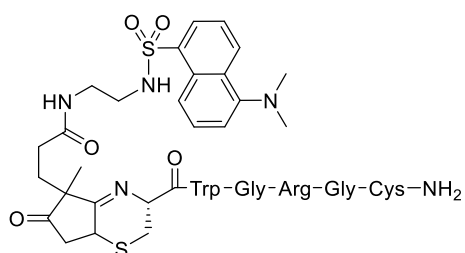
Mass calculated: $C_{48}H_{65}N_{15}O_{10}S_3$ 1107.4

Mass found: m/z 1108.7 $[M+H]^+$

Purification conditions: 20-50% B, $t_R = 13.6 + 14.2$ min

Product: **4.17**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 22.2 + 22.4$ min

MS characterisation:

Technique: ESI, positive mode

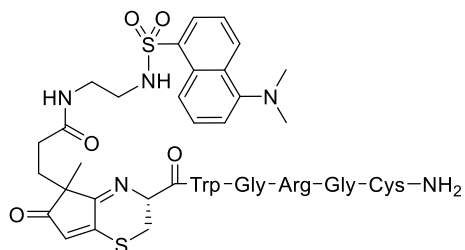
Mass calculated: $C_{50}H_{66}N_{14}O_{10}S_3$ 1118.4

Mass found: m/z 1119.4 [M+H]⁺

Purification conditions: -

Product: **4.18**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 21.9 + 22.3$ min

MS characterisation:

Technique: ESI, positive mode

Mass calculated: $C_{50}H_{64}N_{14}O_{10}S_3$ 1116.4

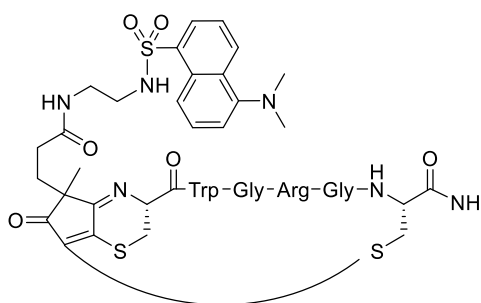
Mass found: m/z 1117.3 [M+H]⁺

Purification conditions: -

Product:

4.19

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 21.5 + 21.9$ min

MS characterisation:

Technique: ESI, positive mode

Mass calculated: $C_{50}H_{62}N_{14}O_{10}S_3$ 1114.4

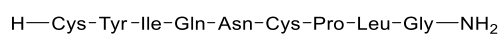
Mass found: m/z 1115.3 $[M+H]^+$

Purification conditions: 25-35% B, $t_R = 15.0 + 16.2$ min

Product:

4.20

Structure:



HPLC/MS characterisation:

Gradient and retention time: 10-60 % B, $t_R = 9.6$ min

MS characterisation:

Technique: ESI, positive mode

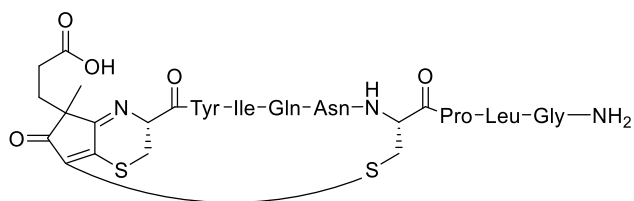
Mass calculated: $C_{43}H_{68}N_{12}O_{12}S_2$ 1008.5

Mass found: m/z 1009.2 $[M+H]^+$

Purification conditions: 25-50%B $t_R = 6.8$ min

Product: **4.23**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 25-35% B, $t_R = 13.6 + 14.8$ min

MS characterisation:

Technique: ESI, positive mode

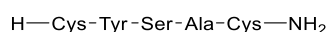
Mass calculated: $C_{52}H_{72}N_{12}O_{15}S_2$ 1168.5

Mass found: m/z 1169.7 $[M+H]^+$

Purification conditions: -

Product: **4.24**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-30% B, $t_R = 8.5$ min

MS characterisation:

Technique: ESI, positive mode

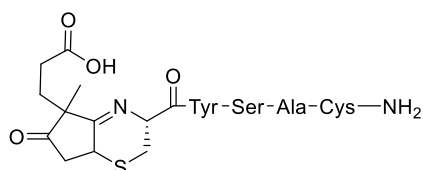
Mass calculated: $C_{21}H_{32}N_6O_7S_2$ 544.2

Mass found: m/z 544.9 $[M+H]^+$

Purification conditions: 15-40%B, $t_R = 6.7$ min

Product: **4.25**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 20.1 + 21.3$ min

MS characterisation:

Technique: ESI, positive mode

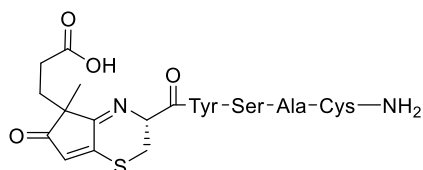
Mass calculated: $C_{30}H_{40}N_6O_{10}S_2$ 708.2

Mass found: m/z 709.0 $[M+H]^+$

Purification conditions: -

Product: **4.26**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 19.8 + 20.8$ min

MS characterisation:

Technique: ESI, positive mode

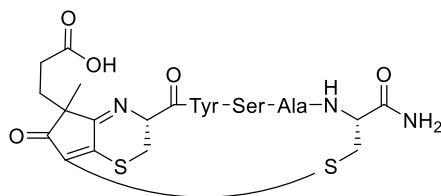
Mass calculated: $C_{30}H_{38}N_6O_{10}S_2$ 706.2

Mass found: m/z 707.1 $[M+H]^+$

Purification conditions: -

Product: **4.27**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 16.6+17.9$ min

MS characterisation:

Technique: ESI, positive mode

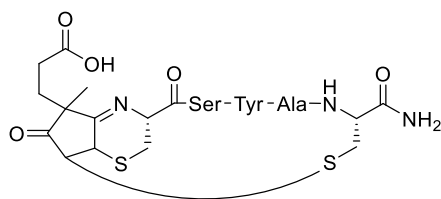
Mass calculated: $C_{30}H_{36}N_6O_{10}S_2$ 704.2

Mass found: m/z 705.0 $[M+H]^+$

Purification conditions: -

Product: **4.28**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 17.6$ min

MS characterisation:

Technique: ESI, positive mode

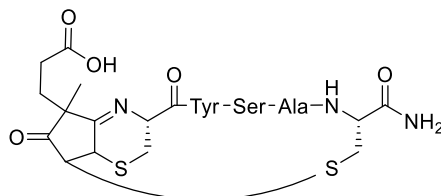
Mass calculated: $C_{30}H_{38}N_6O_{10}S_2$ 706.2

Mass found: m/z 707.2 $[M+H]^+$

Purification conditions: -

Product: **4.29**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 17.2$ min

MS characterisation:

Technique: ESI, positive mode

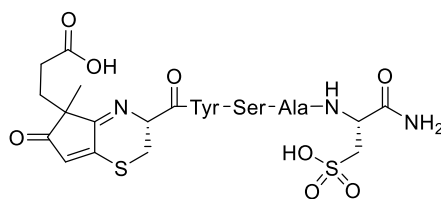
Mass calculated: $C_{30}H_{38}N_6O_{10}S_2$ 706.2

Mass found: m/z 707.1 $[M+H]^+$

Purification conditions: -

Product: **4.30**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 24.5 + 25.8$ min

MS characterisation:

Technique: ESI, positive mode

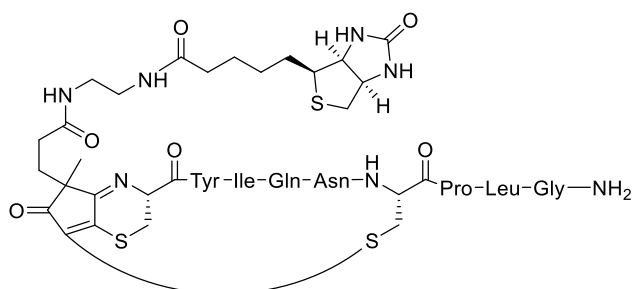
Mass calculated: $C_{30}H_{38}N_6O_{13}S_2$ 754.2

Mass found: m/z 755.3 $[M+H]^+$

Purification conditions: -

Product: **4.31**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 10-60% B, $t_R = 16.4 + 17.1$ min

MS characterisation:

Technique: ESI, positive mode

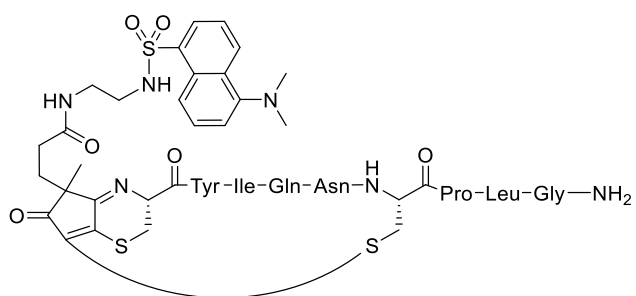
Mass calculated: $C_{64}H_{92}N_{16}O_{16}S_3$ 1436.6

Mass found: m/z 1437.8 $[M+H]^+$

Purification conditions: 25-45% B (60 °C), $t_R = 12.7 + 14.2$ min

Product: **4.32**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 10-60% B, $t_R = 22.6 + 23.2$ min

MS characterisation:

Technique: ESI, positive mode

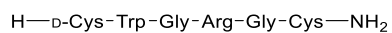
Mass calculated: $C_{66}H_{89}N_{15}O_{16}S_3$ 1443.6

Mass found: m/z 1444.6 $[M+H]^+$

Purification conditions: 25-45% B (60 °C), $t_R = 17.9 + 18.6$ min

Product: **4.33**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-30% B, $t_R = 9.5$ min

MS characterisation:

Technique: ESI, positive mode

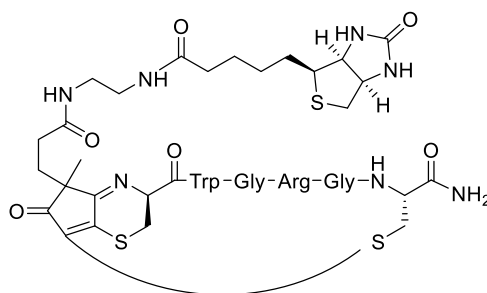
Mass calculated: $C_{27}H_{41}N_{11}O_6S_2$ 679.3

Mass found: m/z 680.0 $[M+H]^+$

Purification conditions: 15-30% B, $t_R = 10.2$ min

Product: **4.34**

Structure:



HPLC/MS characterisation:

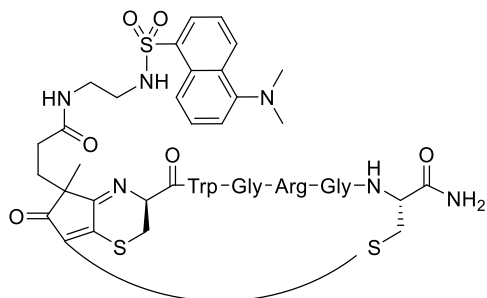
Gradient and retention time: 0-50% B, $t_R = 16.8 + 19.5$ min

MS characterisation:

Technique: ESI, positive mode

Mass calculated: $C_{48}H_{65}N_{15}O_{10}S_3$ 1107.4Mass found: m/z 1108.7 $[M+H]^+$ Purification conditions: 20-50% B, $t_R = 13.0 + 17.6$ minProduct: **4.35**

Structure:



HPLC/MS characterisation:

Gradient and retention time: 0-50% B, $t_R = 21.4 + 23.7$ min

MS characterisation:

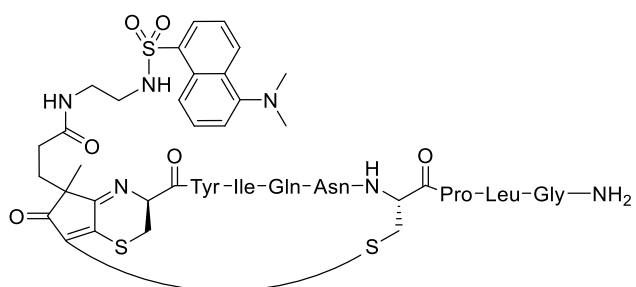
Technique: ESI, positive mode

Mass calculated: $C_{50}H_{62}N_{14}O_{10}S_3$ 1114.4Mass found: m/z 1115.5 $[M+H]^+$ Purification conditions: 30-40% B, $t_R = 7.5 + 13.5$ min

Product:

4.38

Structure:



HPLC/MS characterisation:

Gradient and retention time: 10-60% B, $t_R = 21.7 + 23.7$ min

MS characterisation:

Technique: ESI, positive mode

Mass calculated: $C_{66}H_{89}N_{15}O_{16}S_3$ 1443.6

Mass found: m/z 1444.5 $[M+H]^+$

Purification conditions: 25-45% B (60 °C), $t_R = 16.0 + 20.1$ min

Resum en català

Introducció i objectius

Tant oligonucleòtids com pèptids i proteïnes estan implicats en la majoria dels processos cel·lulars. Aquestes molècules no només fan possible la vida, sinó que algunes d'elles estan directament relacionades amb el desenvolupament de malalties. Per exemple, algunes proteïnes estan implicades en malalties com l'Alzheimer, el Huntington o el Parkinson,^{1,2} mentre que repeticions aberrants d'algunes seqüències en el genoma estan relacionades amb una àmplia varietat d'afeccions conegudes com a TRED (Trinucleotide Repeat Expansion Disorders).^{3,4}

D'altra banda, però, aquestes biomolècules tenen un gran potencial com a possibles medicaments. Fins avui, unes 180 proteïnes han estat aprovades com a fàrmacs o agents de diagnòstic per l'FDA,⁵ i tot i que no molts oligonucleòtids han arribat al mercat, sí que n'hi ha força que estan en fase d'assajos clínics.⁶ Els pèptids també tenen un lloc important en el camp de la medicina: les ciclosporines actuen com a immunosupressors,⁷ els anàlegs de somatostatina s'usen per al tractament de tumors neuroendocrins,⁸ etc.

Tot i això, tant oligonucleòtids com pèptids i proteïnes necessiten, normalment, ser modificats per millorar les seves propietats farmacocinètiques i farmacodinàmiques. D'entre les modificacions normalment practicades, en aquesta tesi doctoral s'ha treballat amb la conjugació i la ciclació, posant especial èmfasi en la seva aplicació a pèptids i PNAs.

La ciclació de pèptids i oligonucleòtids resulta especialment interessant per a incrementar la seva estabilitat als enzims que els degraden i per a obtenir estructures pre-organitzades. Quant a la conjugació, s'usa normalment per a conferir propietats noves a les biomolècules. L'objectiu de la conjugació pot ser millorar la biodistribució o estabilitat de les biomolècules, afavorir la seva internalització cel·lular, dotar-les de marques radioactives o fluorescents per a permetre el seu seguiment dins l'organisme o dins la cèl·lula, o servir com a plataforma per a la síntesi de pro-fàrmacs.

Malgrat que hi ha una alta varietat de reaccions que permeten dur a terme aquests dos processos, cada una d'elles presenta algun inconvenient o limitació, fent que el desenvolupament de noves reaccions i metodologies de ciclació i conjugació segueixi essent necessari. Factors com la formació de més d'un estèreo- o regioisòmer en la generació de nous enllaços o l'estabilitat dels conjugats encara imposen limitacions a l'hora de generar conjugats de biomolècules.

En base als fets exposats, ens vàrem plantejar dos objectius clars per a aquesta tesi:

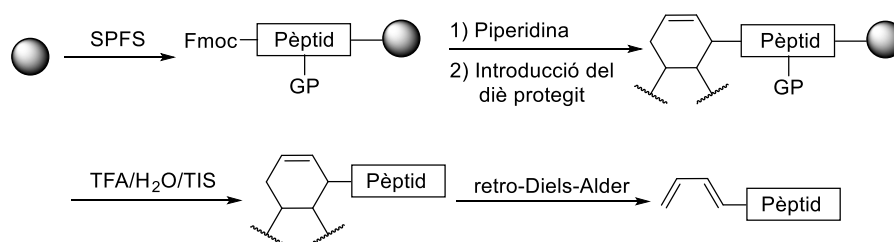
- 1- Trobar una alternativa que permetés usar poliamides derivatitzades amb un 1,3-diè per a sintetitzar conjugats fent servir la reacció de Diels-Alder.
- 2- Explorar l'ús de ciclopent-4-en-1,3-diones 2,2-disubstituïdes com a anàlegs de maleimida en reaccions de tipus Michael amb tiols per a la conjugació i ciclació de pèptids.

Reacció de Diels-Alder sobre resina i en aigua. Conjugats de poliamides i oligonucleòtids derivatitzats amb un diè

Si bé anteriorment ja s'havia usat la reacció de Diels-Alder (DA) per a la conjugació de pèptids generats per síntesi en fase sòlida, encara hi havia una limitació important. Degut a la inestabilitat dels alquil-1,3-diens a la mescla típicament usada per al desancoratge i desprotecció de poliamides (95:2.5:2.5 TFA/TIS/H₂O), la seva introducció en aquest tipus de cadenes per síntesi en fase sòlida no era possible o bé implicava un gran nombre de limitacions.^{9,10}

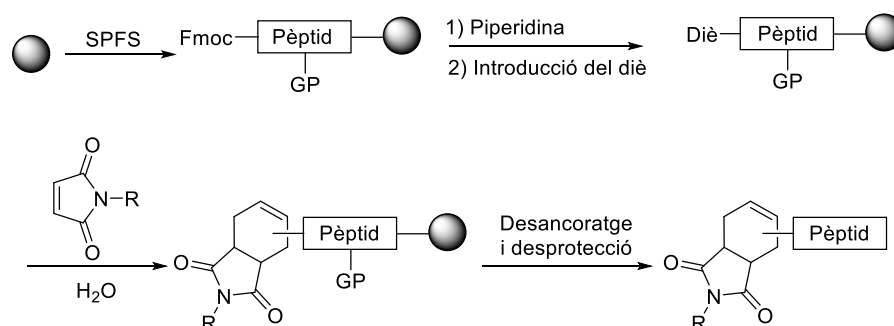
Per tal de solucionar aquest problema, la primera possibilitat que vam decidir estudiar fou el desenvolupament d'un grup protector per al diè. L'objectiu era aconseguir protegir el diè mitjançant una reacció de DA, i, un cop efectuat el desancoratge i desprotecció de la poliamida, desprotegir-lo mitjançant una reacció de retro-DA. Per tant, el cicloadducte format havia de ser resistent al tractament de desprotecció però suficientment làbil per a ser revertit en condicions relativament suaus (**Esquema R.C.1**). Malauradament, els

diferents intents duts a terme emprant diverses combinacions diè/dienòfil foren infructuosos.



Esquema R.C.1. Estratègia de protecció del diè mitjançant la formació d'un adducte de Diels-Alder. SPFS = Síntesi de pèptids en fase sòlida, GP = grup protector.

Posteriorment ens vàrem plantejar la possibilitat de realitzar la reacció de DA entre una poliamida derivatitzada amb un diè quan aquesta encara es trobés totalment protegida i unida a la resina. Aquesta estratègia, representada en l'**Esquema R.C.2**, havia de solucionar el problema de la descomposició del diè en les condicions de desprotecció i desancoratge ja que aquest hauria reaccionat per formar l'adducte de DA abans de ser tractat amb la mescla àcida.



Esquema R.C.2. Reacció de Diels-Alder sobre resina i en aigua per a evitar la descomposició de l'alquil-1,3-diè en la mescla de desancoratge i desprotecció.

Per a dur a terme tal estratègia, però, era necessari tenir en compte certs aspectes. En primer lloc, i donat que les reaccions de DA són molt més ràpides en aigua que en mescles aigua/dissolvent orgànic o en dissolvents orgànics purs, vàrem pensar que el millor dissolvent per a la nostra proposta seria aigua pura. L'efecte positiu de l'aigua en les reaccions de DA recau en l'estabilització dels estats de transició polaritzats d'aquesta cicloaddició,¹¹ en l'anomenat efecte hidrofòbic,¹² i en el fet que l'energia de cohesió d'aquest dissolvent té un efecte similar sobre la reacció al que pot tenir l'aplicació de pressió externa.^{13,14} A més, aquesta elecció permetria treballar amb biomolècules que

poguessin ser d'interès per a la síntesi de conjugats, que típicament són solubles en aigua. En segon lloc, i com a conseqüència d'aquesta elecció de dissolvent, va sorgir la necessitat de treballar amb resines formades a partir de mescles polietilenglicol/poliestirè per tal d'evitar l'encongiment que pateixen les resines de poliestirè en dissolvents polars, que impedeix una reacció eficient. Curiosament, no vàrem poder trobar cap exemple a la literatura on es dugués a terme una reacció de DA sobre resina emprant aigua com a dissolvent.

Com a pas previ, es va decidir avaluar l'efecte que la mescla TFA/H₂O/TIS 95:2.5:2.5 tenia sobre un pèptid derivatitzat amb un alquil-1,3-diè. Això va permetre confirmar alguns resultats prèviament descrits¹⁰ així com observar nous productes de descomposició que no havien estat mai descrits, per la formació dels quals es va suggerir un mecanisme. Bàsicament, la mescla de desancoratge i desprotecció va provocar la reducció i hidratació parcials del doble enllaç conjugat, així com la fragmentació d'aquest (**Figura R.C.1**). Contràriament al que s'havia descrit per a diens aromàtics com el furà,¹⁰ no es va detectar cap fragment que conservés el diè totalment inalterat.

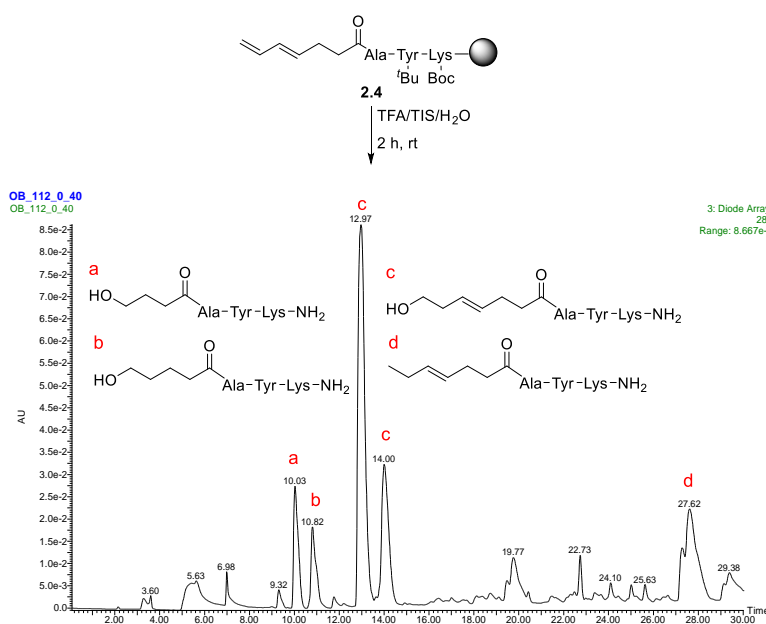


Figura R.C.1. Productes de descomposició del diè obtinguts després de tractar un tripèptid derivatitzat amb un diè amb la mescla TFA/H₂O/TIS 95:2.5:2.5. Nota: tant la reducció com la hidratació parcial del doble enllaç poden donar lloc a diferents regioisòmers dels quals, a fi de simplificar la figura, aquí només se'n representa un.

Després d'aquest estudi previ, es van iniciar els experiments per desenvolupar la metodologia proposada en l'**Esquema R.C.2**. Els primers passos que es van fer van anar encaminats a elegir un diè adequat per a la reacció i a confirmar que el cicloadducte generat entre aquest i una maleimida, que van ser triades com a dienòfils, era estable a les condicions de desprotecció i desancoratge. Quant als diens, es van introduir en un tripèptid el derivat de furà àcid 3-(2-furil)propanoic i l'àcid (*E*)-4,6-heptadienoic. Aquest últim va generar els cicloadductes desitjats amb un rendiment superior que el derivat de furà quan un maleimido-PNA es va usar com a dienòfil, per la qual cosa va ser elegit com a diè a ser usat en la resta d'experiments. A més, els cicloadductes generats amb el derivat de furà van demostrar ser reversibles en les condicions d'anàlisi per espectrometria de masses MALDI-TOF, fet que donava una indicació de la seva menor estabilitat a la reacció de retro-Diels-Alder.

Un cop es va haver corroborat que els cicloadductes eren estables a la mescla àcida usada per a la desprotecció i es va haver escollit el diè més adequat, l'efecte que la composició del dissolvent tenia sobre la conversió del material de partida unit a resina es va avaluar. Tal com està àmpliament descrit en la bibliografia per a reaccions on els dos reactius estan en solució,¹⁵ l'efecte negatiu que els co-solvents orgànics tenen en la velocitat d'aquesta cicloaddició va quedar clarament demostrat. Quan la mateixa reacció es va dur a terme en aigua pura o en una mescla 1:1 aigua/DMSO es va observar que en aquestes últimes condicions el rendiment de la reacció era molt més baix. A més, degut a que quedava diè sense reaccionar i a la seva descomposició en diferents subproductes quan aquest es sotmet a les condicions de desprotecció i desancoratge de la resina, el cru de reacció era molt més brut quan el DMSO es va usar com a co-solvent.

Posteriorment, la reacció es va optimitzar per tal d'aconseguir els cicloadductes desitjats amb bon rendiment i usant només petits excessos del reactiu soluble. Les condicions de reacció que es van considerar més adequades usaven entre 3 i 5 equivalents del reactiu derivatitzat amb una maleimida, i fornien els conjugats objectiu després de 24 hores de reacció a 65 °C (veure **Figura R.C.2a**). Aquest marge entre 3 i 5 equivalents es va adoptar degut a que les resines amb una substitució més baixa requerien un major volum de dissolució per a quedar completament cobertes i, per tant, calia un excés més gran per a que la reacció tingués lloc amb bon rendiment.

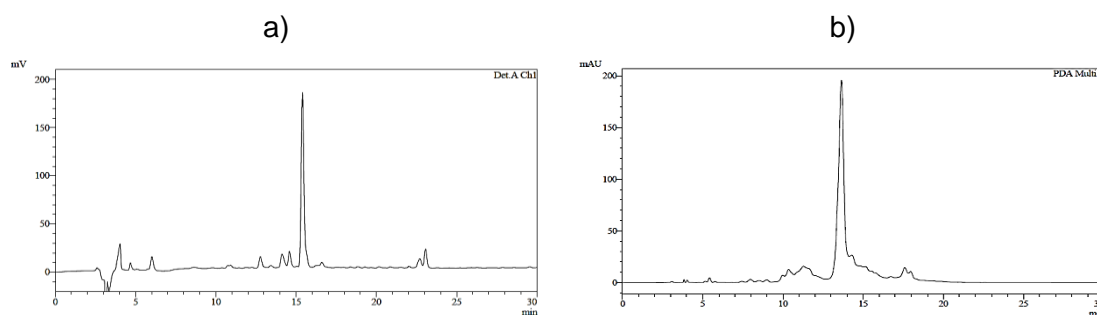


Figura R.C.2. Crus de reacció entre un diè-pèptid unit a resina i un maleimido-PNA (a) i entre un diè-PNA unit a resina i un pèptid cíclic que conté una maleimida protegida.

Es va comprovar l'efecte que la grandària de les molècules implicades en la reacció, tant la que està unida a suport sòlid com la soluble, tenia sobre el rendiment final. Com era d'esperar, a mesura que la mida de qualsevol dels dos reactius augmentava el rendiment de la reacció disminuïa, tot i que això no va impedir que s'obtinguessin conjugats de poliamides, tant pèptids com PNA, de longituds que puguin resultar útils per assajos biològics.

En aquest punt es va assajar l'ús de maleïmides protegides, àmpliament usades en el nostre laboratori,^{16,17,18,19} per a la síntesi de conjugats amb diferents punt d'unió. Si bé fins aquest punt els conjugats sintetitzats s'havien generat unint els extrems *N*-terminals de les dues poliamides conjugades, aquest tipus de reactius va permetre la síntesi de conjugats formats mitjançant la unió entre l'extrem *N*-terminal de la poliamida unida al suport sòlid i l'extrem *C*-terminal o posicions internes de les poliamides en solució (**Figura R.C.3**). A més, gràcies a l'ús d'aquests reactius es van poder obtenir conjugats amb estructures força complexes, com per exemple un pèptid amb un motiu cíclic unit a un PNA (veure **Figura R.C.2b**).

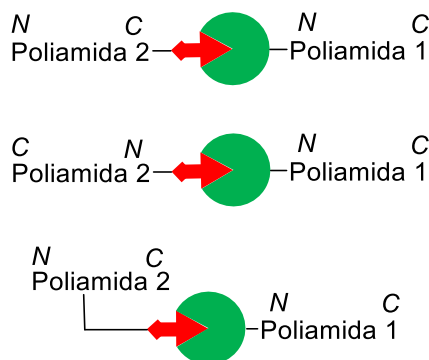


Figura R.C.3. Diferents patrons d'unió obtinguts usant la reacció de Diels-Alder sobre resina.

Una de les nostres majors preocupacions era el fet de treballar amb molècules poc solubles o insolubles en aigua, com per exemple la majoria de fluoròfors, degut al ja esmentat efecte negatiu que té l'addició de co-solvents orgànics a la mescla de reacció. Sorprenentment, però, aquells reactius més apolars que requerien DMSO per a ser completament o parcialment dissolts van donar els cicloadductes corresponents amb bons rendiments. La inspecció d'un dels cruds de reacció en les etapes més inicials de la transformació va rebel·lar el que semblava l'adsorció del fluoròfor sobre la superfície de la resina, incrementant la concentració efectiva d'aquesta espècie prop dels centres reactius. Probablement aquest fet va compensar la disminució en la velocitat de reacció causada per l'addició del co-solvent orgànic, i va permetre d'obtenir poliamides modificades amb aquestes molècules amb bons rendiments, com es pot veure en la **Figura R.C.4**. A més, es va poder comprovar també que aquesta metodologia tolerava bé l'ús de suspensions en mescles aigua/DMSO, fet que amplia força la seva utilitat a l'hora de treballar amb molècules apolars.

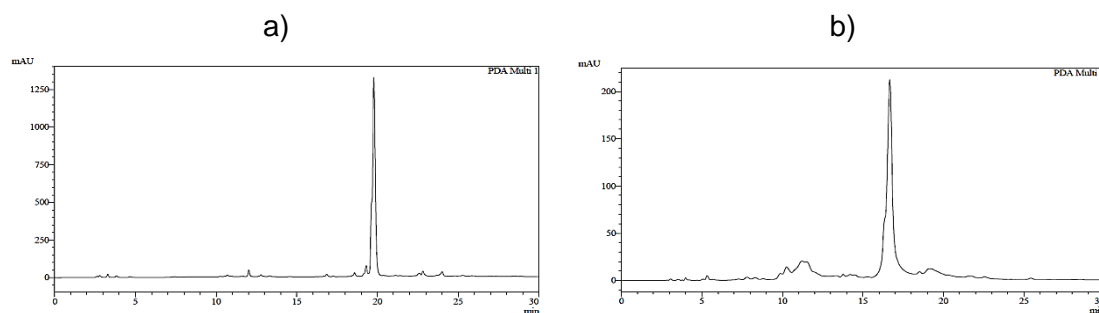
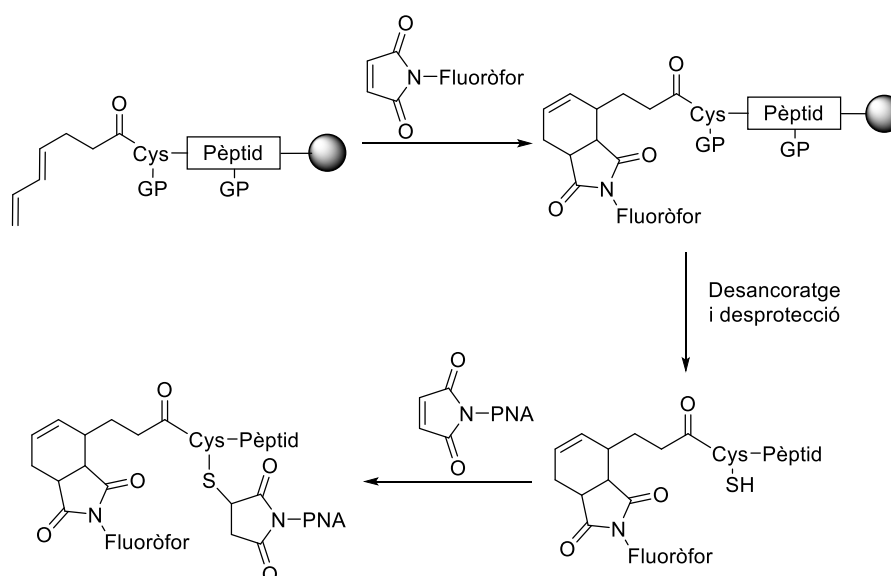


Figura R.C.4. Reacció entre el fluoròfor dansil derivatitzat amb una maleimida protegida i un diè-peptid unit a resina (a) i entre el mateix fluoròfor i un diè-PNA unit a resina (b).

Finalment, la possibilitat de combinar la reacció de DA en aigua i sobre resina amb una de les reaccions “click” més emprades, aquella entre un tiol i una maleimida, va ser demostrada. Gràcies al fet que el protector de tiols làbil a àcids Trt és estable a les condicions emprades per a la cicloaddició, es va poder obtenir un conjugat pèptid-fluoròfor, generat usant la metodologia aquí descrita, que incorporava una cisteïna (**Esquema R.C.3**). El tiol d'aquest aminoàcid va ser usat posteriorment en una reacció de tipus Michael amb un maleimido-PNA per a generar un doble conjugat amb un rendiment global, després de dues reaccions i purificacions per HPLC, del 23 %.



Esquema R.C.3. Síntesi d'un conjugat doble emprant la reacció de Diels-Alder sobre resina en aigua i una posterior reacció de tipus Michael.

Donats els bons resultats obtinguts amb la reacció de DA entre poliamides derivatitzades amb un diè unides a resina i diversos tipus de maleimides, vàrem decidir intentar adaptar aquesta metodologia per a la síntesi de conjugats pèptid-oligonucleòtid. Com que aquests últims no són estables al tractament àcid emprat per a la desprotecció de poliamides, era necessari derivatitzar l'oligonucleòtid amb el diè i dur a terme la reacció de conjugació mentre aquest encara estigués protegit i unit a resina.

Els primers experiments que es van realitzar usant un oligonucleòtid model, que només contenia timidines, ens van permetre identificar i solucionar, pràcticament del tot, les reaccions secundàries que van tenir lloc. Aquestes foren, bàsicament, la fragmentació de l'esquelet oligonucleotídic i l'acilació del conjugat per part de l'agent de “capping” usat durant la síntesi en fase sòlida. Quan vàrem passar a treballar amb seqüències mixtes,

però, aquestes reaccions secundàries varen tornar a guanyar importància, especialment aquella que provocava l'escissió de l'oligonucleòtid. Com que els diferents intents per minimitzar aquest problema van resultar inútils, vam procedir a estudiar amb detall en quin moment exacte del procés es produïa aquest trencament dels enllaços fosfodi- o fosfotrièster. Per a tal fi es va usar l'anàlisi per espectrometria de masses (MALDI-TOF) del suport sòlid sobre el qual s'havia fet la reacció de conjugació, usant els protocols descrits pel grup del Dr. Vasseur.^{20,21} Aquesta metodologia ens va permetre descobrir que la fragmentació es donava durant la reacció de DA i que el problema era que les condicions de reacció eren massa dures per aquests substrats. Altres intents de conjugació emprant condicions més suaus requerien allargar els temps de reacció, cosa que conduïa a resultats igualment decebedors. Per aquests motius, el treball amb oligonucleòtids en fase sòlida es va abandonar.

Ciclopent-4-en-1,3-diones 2,2-disubstituïdes: més que simples acceptors de Michael

Tot i que la reacció entre tiols i maleïmides s'havia considerat sempre exempta de problemes, en els últims anys s'ha posat de manifest que les 4-alkilsulfonilsuccinimides generades després de la reacció de tipus Michael no són tant estables com es pensava. A part que el tiol unit a la succinimida pot intercanviar-se per d'altres presents al medi,^{22,23} les succinimides poden experimentar reaccions d'hidròlisi en medis aquosos (**Figura R.C.5**).^{24,25} Aquesta segona reacció genera dos nous regioisòmers que poden tenir propietats diferents, cosa que pot acabar essent un problema tant des del punt de vista purament químic com des del biològic.

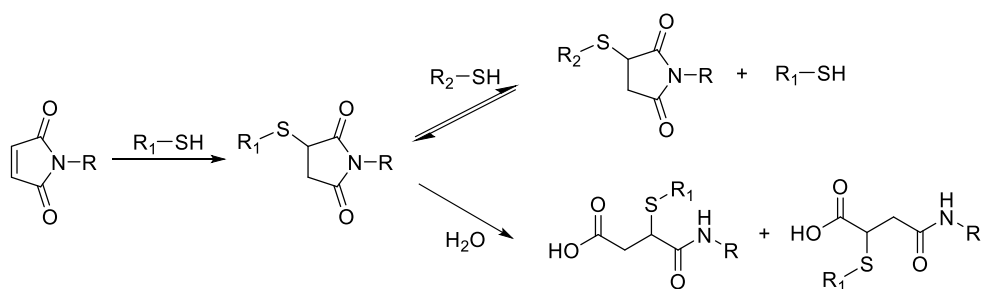


Figura R.C.5. Reaccions secundàries que pot experimentar una 4-alkilsulfonilsuccinimida en medi aquós.

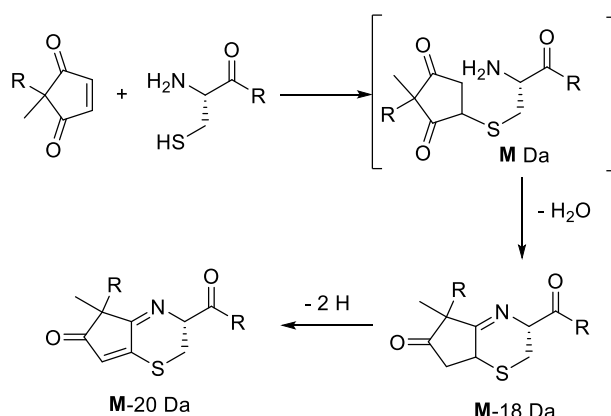
Donat que la reacció de tipus Michael entre tiols i malemides és molt útil com a mètode de conjugació i que alguns conjugats sintetitzats d'aquesta manera han arribat al mercat o a fases avançades en assajos clínics, vàrem pensar que seria interessant treballar amb un anàleg de maleimida que no fos hidrolitzable.

Aquests anàlegs foren les ciclopent-4-en-1,3-diones 2,2-disubstituídes (CPDs), unes molècules que, com que el grup imido ha estat reemplaçat per una dicetona, no poden donar reaccions d'hidròlisi, ja sigui abans o després de l'addició d'un tiol.

Quan es va procedir a avaluar la reacció d'aquest tipus de compostos amb pèptids que contenien cisteïna en posicions internes o en la C-terminal es va comprovar que el seu comportament és diferent al de les maleimides. Contràriament a aquestes últimes, que reaccionen quantitativament amb qualsevol tipus de tiol per generar adductes de tipus Michael (ATMs) que no reverteixen als reactius de partida si no hi ha altres tiols presents, la reacció entre una CPD i un tiol que no posseeixi una amina lliure en la posició 2 no es va produir mai de manera quantitativa. Tot i que la reacció de tipus Michael era ràpida, es va poder comprovar que els ATMs formats a partir de CPDs eren reversibles a temperatura ambient i que l'equilibri estava desplaçat cap a la formació dels reactius de partida. Es va obtenir evidència directa d'aquest fet quan els esmentats ATMs es van aïllar i reanalitzar després d'una breu incubació en aigua, experiment que també va permetre d'observar que el procés de retro-Michael era molt ràpid.

D'altra banda, aquells pèptids on la cisteïna ocupava la posició N-terminal sí que van reaccionar completament quan es van mesclar amb CPDs. Sorprenentment, però, l'adducte que es va generar no era l'esperat ATM (de massa M Da), sinó un amb una massa 20 Da inferior ($M-20$ Da). Diferents experiments van permetre postular una

possible estructura per aquest nou producte, així com un camí de reacció plausible per explicar la seva formació (**Esquema R.C.4**).



Esquema R.C.4. Camí de reacció proposat per a la formació de l'adducte final **M-20 Da**.

Breument, després de la reacció de tipus Michael entre la CPD i el tiol, l'amina de la posició *N*-terminal i un dels grups ceto de la CPD condensen per formar una imina, donant lloc a un producte amb un cycle de sis baules que presenta una massa 18 unitats menor que l'ATM esperat (**M-18 Da**). Aquesta reacció és tant ràpida que l'ATM generat en primer lloc no es va detectar mai per HPLC (tot i que sí que es va poder detectar, en alguna ocasió, per injecció directa dels crús de reacció en un espectròmetre de masses). Aquest producte intermediari generat per deshidratació s'oxida, en una etapa més lenta, per donar l'adducte estable final **M-20 Da**, que posseeix un sistema conjugat amb un màxim d'absorció al voltant de 330 nm. L'estructura d'aquest tipus de compost es va poder confirmar mitjançant tècniques de RMN emprant cisteïnat de metil com a 1,2-aminotiol model.

Després d'establir que les CPDs reaccionen amb 1,2-aminotols de forma quantitativa per donar uns adductes prèviament no descrits, però que les reaccions amb altres tiols són reversibles, es va procedir a estudiar les seves reaccions amb cisteïna i homocisteïna. Tots dos aminoàcids es van fer reaccionar amb CPDs i es va observar que ho feien de manera diferent. La cisteïna va generar l'adducte **M-20 Da** tal com s'esperava per un 1,2-aminotiol. Tot i així, i segurament degut a la presència de l'àcid carboxílic de l'aminoàcid, es va observar que aquest adducte patia, de forma reversible, hidròlisi de l'imina, donant lloc a un producte de massa **M-2 Da** prèviament no observat. La reversibilitat d'aquesta reacció es va comprovar tal i com s'havia fet per comprovar la dels ATM. Quant a la reacció amb homocisteïna, i com a conseqüència de la superior mida de l'anell que es generaria, de set baules en comptes de sis, la formació de la imina

no té lloc. Això comporta que la reacció entre homocisteïna i CPDs sigui reversible, com en el cas de pèptids que contenen cisteïna en posicions internes o en la C-terminal.

Afortunadament, els resultats experimentals es van poder confirmar mitjançant els càlculs teòrics que va dur a terme en Lluís Raich en el grup de la Dra. Carme Rovira. Els seus resultats recolzen el diferent comportament de maleïmides i CPDs en reaccions de tipus Michael observat experimentalment. A més, l'anàlisi dels orbitals i de les distàncies d'enllaç de les dues estructures va permetre suggerir que la raó d'aquesta diferent reactivitat és la major deslocalització de l'enllaç C=C en les CPDs. Els seus resultats també van permetre confirmar la nostra hipòtesi que la raó rere la reversibilitat de la reacció entre l'homocisteïna i les CPDs era la major dimensió de l'anell que es formaria. Els càlculs indicaren que la formació de l'hemiaminal que conduiria a la formació de la imina, és a dir, el primer pas en l'atac de l'amina sobre la cetona, requeria d'una major energia d'activació que quan la reacció genera un anell de sis baules.

També es va avaluar el comportament de CPDs amb diferents substituents. Jordi Agramunt, que en aquell moment estava realitzant el seu treball de màster al nostre grup, va sintetitzar dos derivats de CPD, un incorporant el fluoròfor dansil i l'altre incorporant biotina. Totes dues CPDs es van fer reaccionar amb cisteïnat de metil, i es va comprovar que seguien el mateix camí de reacció i que fornien el mateix tipus d'adducte que les CPDs anteriorment emprades amb bons rendiments (veure **Figura R.C.6**).

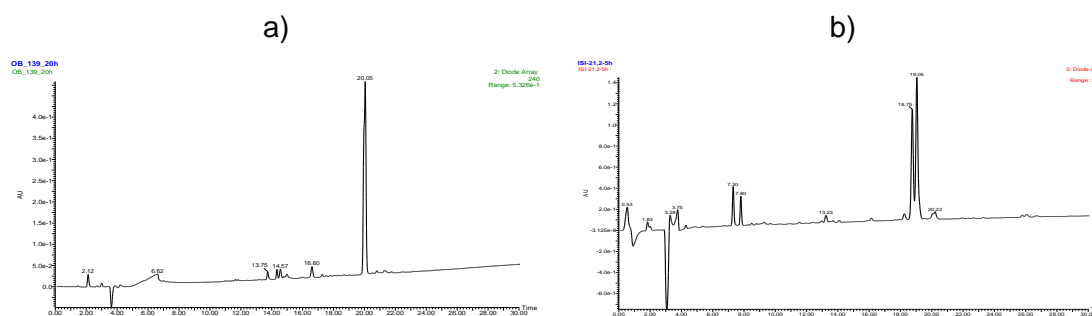


Figura R.C.6. Reaccions del cisteïnat de metil amb la CPD derivatitzada amb biotina (a, els dos isòmers coelueixen) i amb la CPD derivatitzada amb dansil (b, els dos isòmers tenen temps de retenció diferents).

Tot i els bons resultats, durant la realització dels experiments aquí resumits, s'havia anat observant que hi havia força factors que afectaven la rapidesa de la oxidació de l'adducte intermedi **M-18 Da** cap al producte final **M-20 Da**, que és l'etapa limitant de

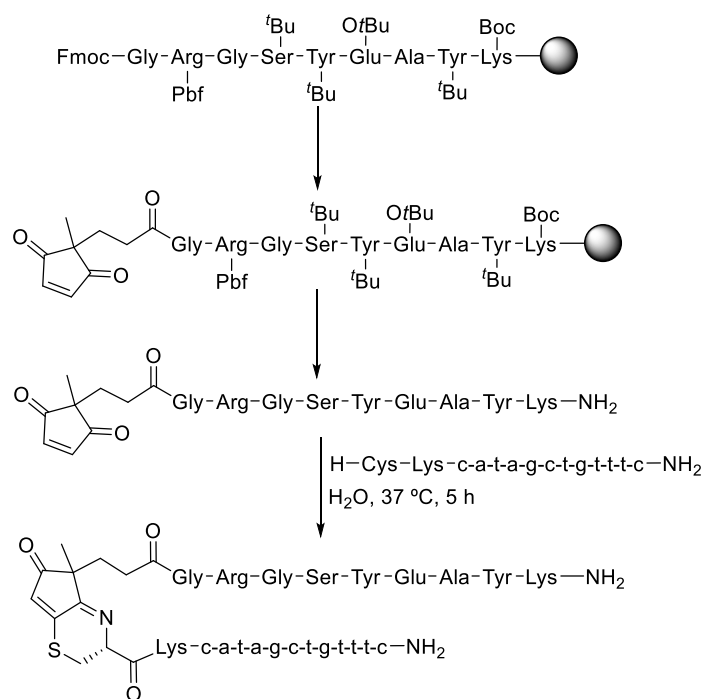
velocitat del procés global. Per tal d'aprofundir en aquest aspecte, es va dur a terme un estudi sistemàtic de les variables que típicament afecten la velocitat de les reaccions químiques. Amb aquest fi, es va avaluar l'efecte de la temperatura, concentració, dissolvent i relacions molars dels reactius en la velocitat de l'esmentada oxidació. Es va veure que, tal com passa en la majoria de les reaccions químiques, un augment en la temperatura produïa un augment en la velocitat de l'etapa limitant, observant-se un efecte força pronunciat (mentre la reacció a 60 °C tenia lloc de forma quantitativa en menys de 15 min, la formació de l'adducte **M-20** a 5 °C no era completa a les 4 h). Pel que fa a la concentració, la tendència va resultar ser la inversa a la normal, doncs en les reaccions més diluïdes l'oxidació va ser més ràpida. Quant als dissolvents, es va veure que el millor és fer les reaccions en aigua pura, doncs els co-solvents orgànics van demostrar disminuir la velocitat d'oxidació. Finalment i de manera molt sorprenent, es va veure que un excés de CPD en la reacció afavoria l'oxidació de l'adducte **M-18**, mentre que un excés de tiol no tenia cap influència destacable.

Donat que l'adducte **M-20** posseeix un màxim d'absorció prop de 330 nm i que aquest també absorbeix a 280 nm, vàrem decidir determinar el coeficient d'absorció molar (ϵ) d'aquest tipus d'estructures. Donada la mancança de quantitats suficients de reactius per a poder determinar l' ϵ per pesada, es va decidir fer una aproximació al seu valor. Amb aquest objectiu, es van dur a terme tres reaccions idèntiques amb una quantitat coneguda de pèptid i un excés de CPD i es van deixar a 37 °C fins que van arribar al seu punt final. En aquest moment, de cada reacció se'n van prendre 3 aliquotes idèntiques i es va mesurar l'absorbància de cada una d'elles 3 cops. Aquests 27 valors, que es va comprovar que eren estadísticament comparables (desviació estàndard per sota de l'error típic d'una tècnica basada en espectroscòpia UV-Vis), es van usar per determinar el coeficient d'absorció molar. Es va obtenir un valor de $9140 \text{ cm}^{-1} \text{ M}^{-1}$ a 332 nm amb un error del 2 %, i un de $4210 \text{ cm}^{-1} \text{ M}^{-1}$ a 280 nm amb el mateix error. Vam considerar que aquest valors, donat el seu error associat, eren suficientment bons com per a poder quantificar els productes. Tot i així, i de cara a un futur pròxim, caldria determinar el valor de l' ϵ dels adductes **M-20** Da usant altres tècniques per a la seva quantificació, com per exemple la valoració per RMN que es va realitzar per a la determinació del coeficient d'absorció molar dels compostos cíclics explicat en el capítol 4 d'aquest treball.

Finalment, tot i que encara a nivell d'experiments preliminars, vàrem poder demostrar l'aplicabilitat de la recentment descoberta reactivitat de les CPDs. Com a primer pas es va comprovar que era possible introduir CPDs a la posició *N*-terminal d'una cadena

peptídica usant mètodes estàndard de síntesi de pèptids en fase sòlida. Tant la introducció de CPDs usant com a reactiu d'acoblament una carbodiimida com la seva resistència a la mescla àcida usada per al desancoratge i desprotecció de poliamides van quedar clares, podent-se obtenir un pèptid derivatitzat amb aquests tipus d'estructures (**Esquema R.C.5**, primera reacció).

En una segona etapa, aquest pèptid derivatitzat amb una CPD es va usar per a formar un conjugat amb un PNA que contenia una cisteïna en la posició *N*-terminal. La síntesi d'aquest híbrid pèptid-PNA es va dur a terme sense cap problema, usant només un lleuger excés del reactiu que contenia la CPD (1.2 equivalents) i en condicions molt suaus, podent-se obtenir el producte final amb un 51 % de rendiment després de purificació (veure **Figura R.C.7**).



Esquema R.C.5. Introducció d'una CPD a una cadena peptídica usant mètodes estàndard de síntesi en fase sòlida, i conjugació amb un PNA derivatitzat amb una cisteïna en la posició *N*-terminal.

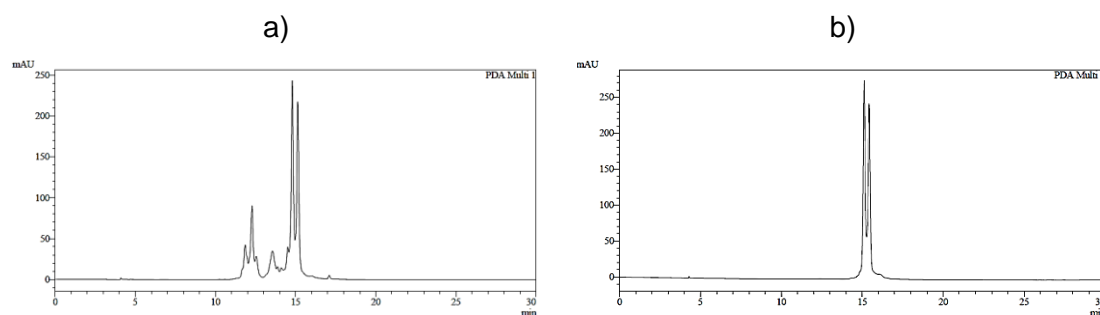
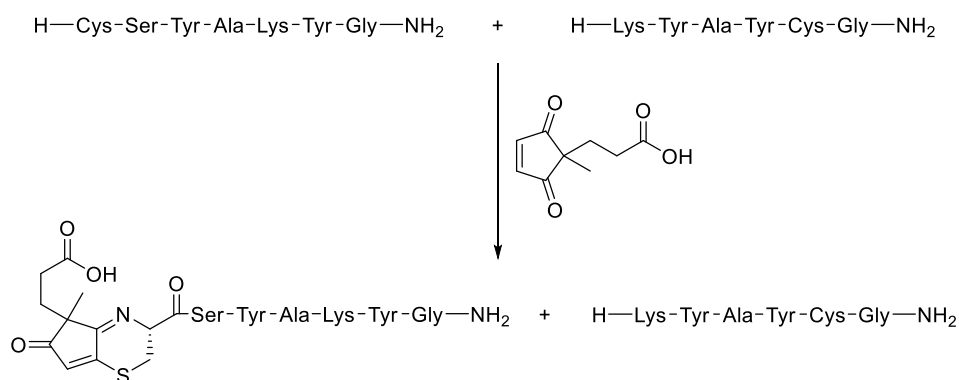


Figura R.C.7. Cru de reacció per a la síntesi del conjugat pèptid-PNA (a, els pics que apareixen entre 11 i 13 min són impureses del PNA de partida) i del producte purificat (b).

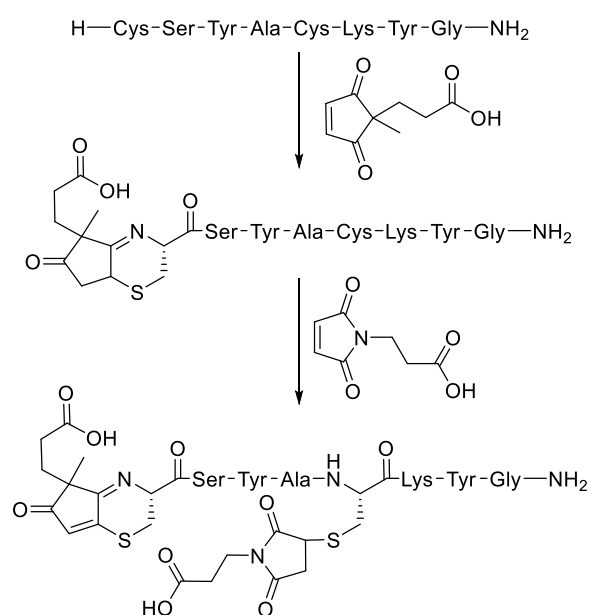
L'estabilitat d'aquest híbrid es va estudiar tant a diferents pHs com en presència d'altres tiols. Es va descobrir que aquest híbrid era molt estable a pHs lleugerament àcids (pH= 6.0) i força estable a pH neutre, mentre que un pH lleugerament bàsic (pH= 8.0) afavoria la seva alteració. Pel que fa a la presència d'altres tiols, es va veure que mentre que els 1,2-aminotiols sí que influeixen negativament en l'estabilitat d'aquest tipus de compostos (almenys en un gran excés), aquells tiols que no presentaven una amina lliure en la posició β , com per exemple el glutatió, no tenien cap tipus d'influència.

A continuació es va procedir a realitzar el marcatge selectiu d'un pèptid que contenia una cisteïna en la posició *N*-terminal en presència d'un altre pèptid que la tenia en una posició interna. Mentre que una maleimida seria incapaç de diferenciar els dos pèptids i de reaccionar de manera selectiva amb un d'ells, l'ús de CPDs va permetre marcar selectivament només aquell que contenia la cisteïna en la posició *N*-terminal (**Esquema R.C.6**).



Esquema R.C.6. Marcatge selectiu d'un pèptid que conté cisteïna en la posició *N*-terminal en front d'un que la conté en una posició interna.

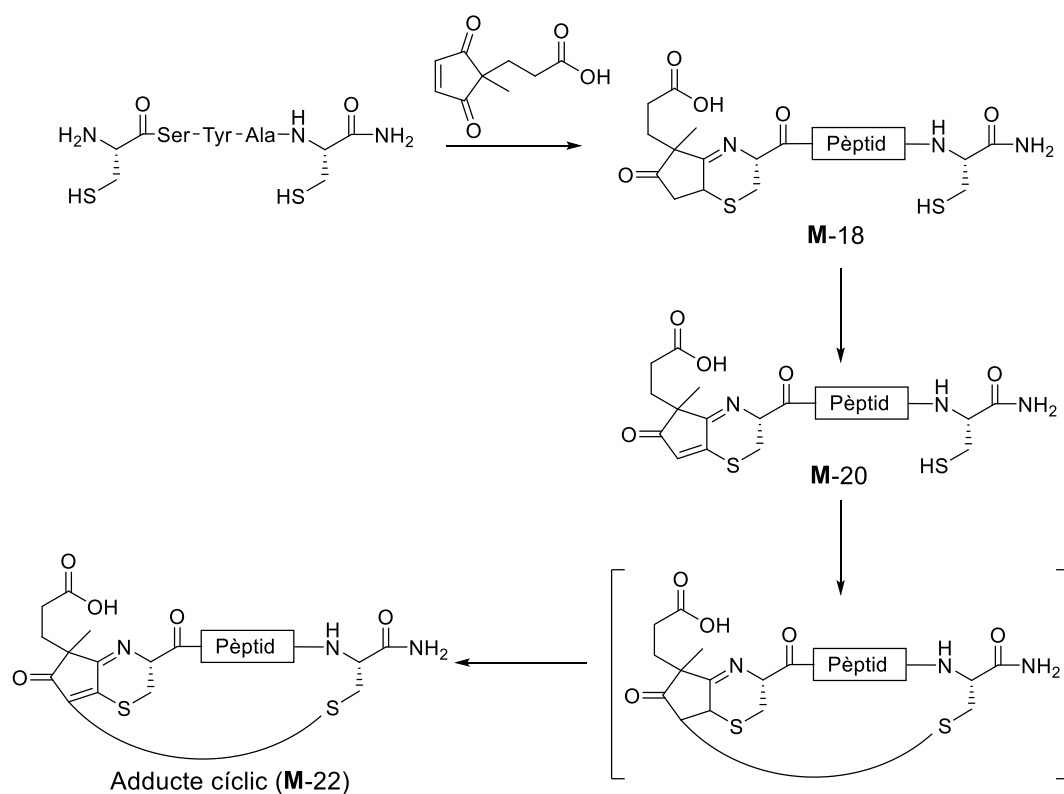
En un experiment en què també s'aprofitava la diferent reactivitat de les CPDs en front de 1,2-aminotiols respecte a altres tipus de tiols, un pèptid que contenia dues cisteïnes, una en posició *N*-terminal i l'altre en una posició interna, va ser doblement marcat de manera regioselectiva. Per a aconseguir-ho, el pèptid es va incubar amb 1.1 equivalents de CPD fins que la formació de l'adducte **M-18 Da** fou completa. Tot seguit es van afegir 5 equivalents de maleimida, i al cap de només una hora ja es va obtenir, amb bon rendiment, el producte desitjat. Aquest estava marcat en la posició *N*-terminal amb una CPD i en la cisteïna interna amb una maleimida (**Esquema R.C.7**). Tot i que aquests experiments es van realitzar amb CPDs i maleïmides sense derivatitzar, van servir per establir les bases per a realitzar aquests processos amb molècules més complexes.



Esquema R.C.7. Doble derivatització regioselectiva d'un pèptid que conté dues cisteïnes, una en la posició *N*-terminal i l'altre en una posició interna.

Ciclació i derivatització de pèptids usant ciclopent-4-en-1,3-diones 2,2-disubstituïdes

En un experiment previ a l'anteriorment comentat, també enfocat al doble marcatge regioselectiu d'un pèptid, es va observar que la formació del producte desitjat anava acompanyada de la d'un altre producte amb una massa de **M-22** Da i que tenia un perfil d'absorció UV-Vis amb un màxim a aproximadament 370 nm. A partir d'aquesta informació, vàrem pensar que aquesta nova espècie podria ser el resultat de l'addició del tiol de la cisteïna interna sobre l'adducte **M-20**, i que, per tant, s'estava formant un producte cíclic (**Esquema R.C.8**).



Esquema R.C.8. Etapes en la formació d'un adducte cíclic de massa **M-22** Da.

Donada la importància dels pèptids cíclics així com l'interès en trobar metodologies que permetin ciclar i derivatitzar o conjuguar pèptids en una sola etapa, vàrem decidir investigar aquesta reacció. En aquest projecte va estar involucrat en Lewis Archibald, que va realitzar una part important dels experiments descrits en aquest treball sota la meva supervisió directa.

En primer lloc es van intentar trobar unes condicions de reacció que permetessin l'obtenció d'aquests productes amb bons rendiments. Basats en el fet que moltes reaccions d'addició de tiols a dobles enllaços pobres en electrons succeeixen a través de mecanismes radicalaris,^{26,27} vàrem decidir afegir el radical TEMPO un cop ja havia tingut lloc l'oxidació de l'adducte **M-18** Da al **M-20**. Aquest procediment ens va permetre obtenir els productes desitjats amb uns rendiments suficientment bons com per afrontar la síntesi a una escala major d'un d'aquests productes per a la seva determinació estructural. Per tal de simplificar al màxim el procés d'elucidació estructural del presumpte adducte cíclic, es va usar el pentapèptid H-Cys-Ser-Tyr-Ala-Cys-NH₂. Els adductes **M-22** Da generats a partir de la seva reacció amb una CPD simple (dos diastereòmers) es van aïllar per separat i sotmetre a diferents experiments d'RMN per tal de confirmar la nostra hipòtesi. Totes les evidències van confirmar l'addició del tiol de

la cisteïna interna sobre l'adducte **M-20 Da**, és a dir, la ciclació del pèptid. En concret, l'experiment d'HMBC va permetre observar l'acoblament dels hidrògens β de la cisteïna C-terminal amb un dels carbonis olefínics de l'anell de cinc baules, prova irrefutable de la unió d'aquest tiol a l'esmentat cicle (**Figura R.C.8**).

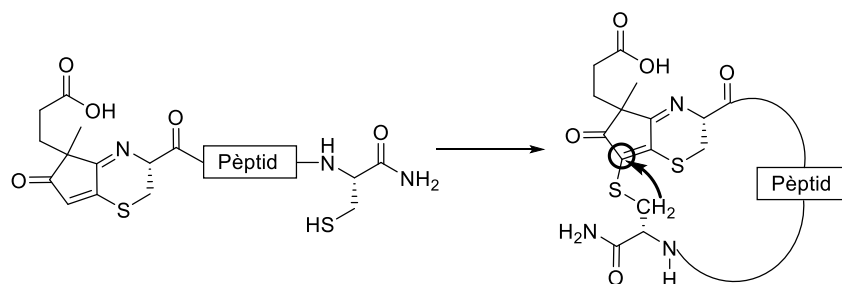


Figura R.C.8. Representació de la correlació entre els hidrògens β de la cisteïna C-terminal i un dels carbonis olefínics del sistema de cinc baules que confirma la ciclació del pèptid.

Tot seguit, i aprofitant que aquest producte ja havia estat aïllat i caracteritzat, es va procedir a determinar-ne el coeficient d'absorció molar ϵ . En aquest cas, una alíquota de cada un dels dos isòmers de l'adducte cíclic es va quantificar mitjançant una valoració per RMN. En aquest procés la quantitat de producte d'una mostra concreta es determina fent addicions de quantitats conegudes d'una substància patró i representant la proporció d'àrees entre producte i patró (la seva integració en l'espectre d'RMN) enfront de la quantitat de patró afegida. Els punts obtinguts s'ajusten a una línia recta, el pendent de la qual està relacionat amb la quantitat de producte que hi ha a la mostra. Un exemple dels resultats obtinguts en la quantificació d'un dels isòmers es mostra en la **Figura R.C.9a**, mentre que l'ajust lineal obtingut per als dos isòmers es pot veure en la **Figura R.C.9b**.

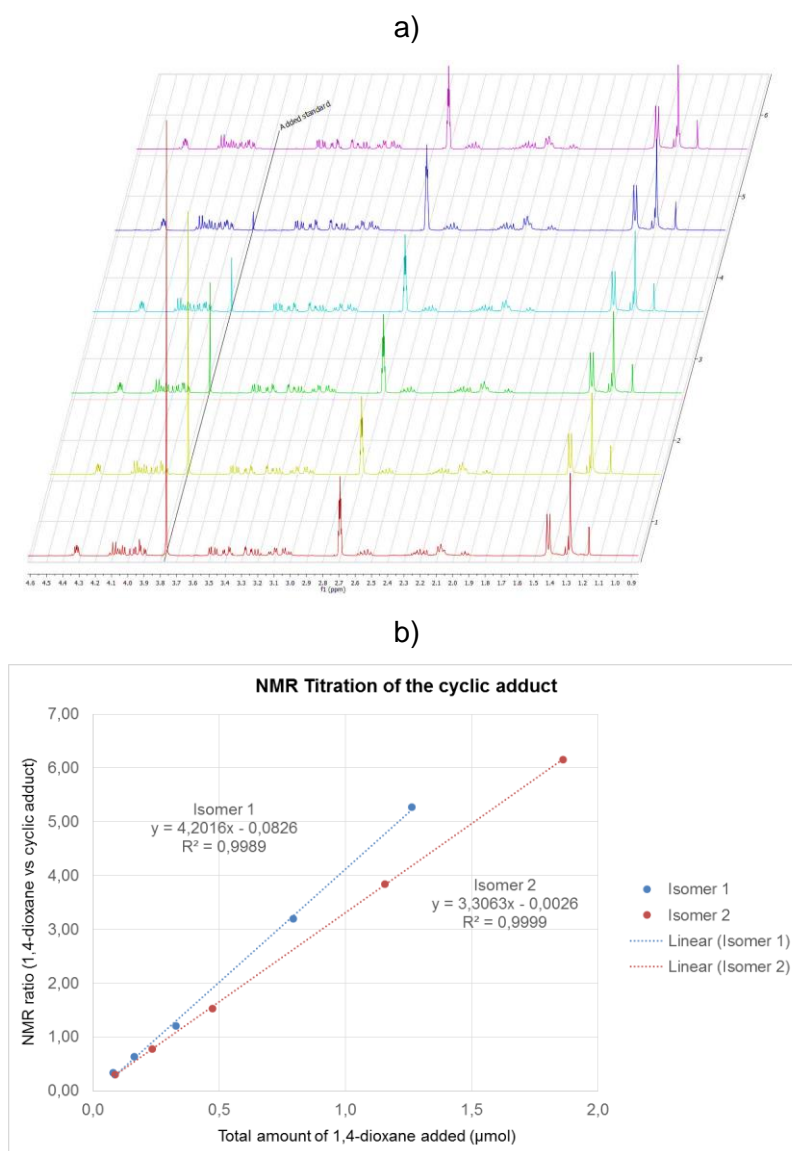


Figura R.C.9. Espectres obtinguts en la valoració per RMN d'un dels isòmers de l'adducte cíclic (a) i gràfics obtinguts en la quantificació dels dos isòmers del producte cíclic per RMN (b).

Un cop els dos isòmers del producte van estar quantificats d'aquesta manera, se'n van preparar diverses dissolucions i l'absorbància d'aquestes va ser mesurada. Això va permetre establir un ϵ de $12215 \text{ cm}^{-1} \text{ M}^{-1}$ a 373 nm en aigua (valor mitjà dels dos isòmers) i un de $12645 \text{ cm}^{-1} \text{ M}^{-1}$ a 365 nm en metanol (usant només l'isòmer 1 degut a la pèrdua del 2 durant la seva manipulació). Els coeficients R^2 obtinguts en fer els ajustos lineals de les diferents mesures, així com la dispersió dels punts experimentals al voltant de la recta d'ajust, van indicar que els errors aleatoris comesos eren petits. Els resultats de la determinació del valor del coeficient d'absorció molar en aigua es poden veure en la figura **Figura R.C.10**.

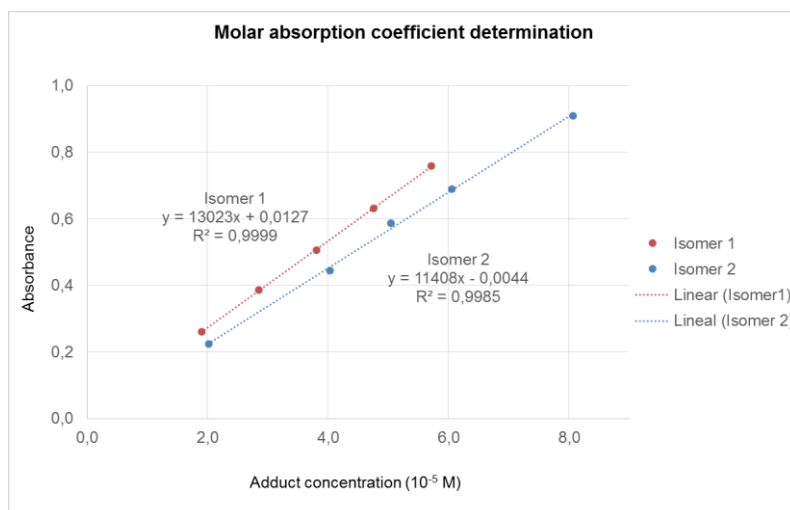


Figura R.C.10. Gràfics obtinguts en la determinació del valor del coeficient d'absorció molar en aigua (373 nm).

Havent confirmat l'estructura d'aquests adductes i determinat el seu coeficient d'absorció molar, es van realitzar un seguit d'experiments per intentar aconseguir la ciclació de diversos pèptids usant diferents CPDs. En aquest procés es va descobrir que la reacció de derivatització i ciclació era molt sensible a la seqüència del pèptid, concretament a si hi havia un aminoàcid aromàtic o no al costat de la cisteïna *N*-terminal, i a la substitució de la CPD. Les diferències de reactivitat eren tant grans que, mentre que en alguns casos la ciclació tenia lloc de manera ràpida i neta, en altres aquest procés mai es va produir de manera satisfactòria. A més, la no ciclació provocava l'exposició del tiol de la cisteïna interna al TEMPO durant temps llargs, la qual cosa desencadenava l'oxidació d'aquest primer a àcid sulfònic. Experiments posteriors on es va comparar la reactivitat de dos pèptids la única diferència entre els quals era l'ordre dels aminoàcids van demostrar que la ciclació, quan hi ha un residu aromàtic contigu a la cisteïna *N*-terminal, té lloc de forma més lenta i dóna lloc a una major quantitat de producte amb el tiol oxidat a àcid sulfònic (veure **Figura R.C.11**).

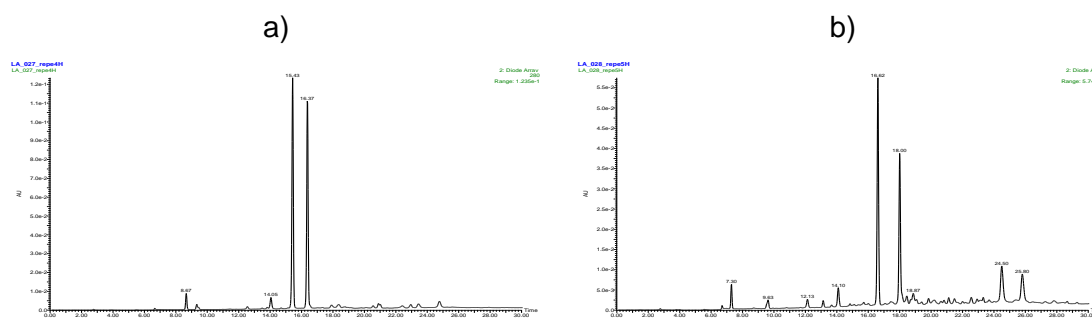


Figura R.C.11. Comparació de la reacció de ciclació de dos pèptids que només difereixen en l'ordre dels seus aminoàcids. (a) Perfil del cru de la reacció de ciclació de la seqüència H-Cys-Ser-Tyr-Ala-Cys-NH₂ a les 4 h i (b) de la reacció de ciclació de la seqüència H-Cys-Tyr-Ser-Ala-Cys-NH₂ a les 5 h.

Justament per intentar evitar l'aparició d'aquest subproducte, es van assajar altres mètodes per intentar promoure la ciclació sense haver d'usar TEMPO. Amb aquesta finalitat es va assajar l'efecte que diferents bases en varies proporcions tenien sobre la reacció de ciclació, però de seguida es va veure que aquestes no promovien la reacció de ciclació i que, malauradament, semblaven provocar l'epimerització de l'adducte **M-20**. També es va avaluar l'efecte del DMSO en la mescla de reacció, donat que aquest és un conegut oxidant en les reaccions de formació de ponts disulfur. El resultat d'aquestes reaccions també va ser decebedor, doncs els crus que es van obtenir presentaven molts subproductes que no es van identificar. Finalment, es va provar de substituir el TEMPO per oxigen. Es va comprovar que si bé també provocava la ciclació del pèptid, generava un subproducte desconegut en unes quantitats massa altes per a poder reemplaçar el TEMPO com a agent oxidant (veure **Figura R.C.12**).

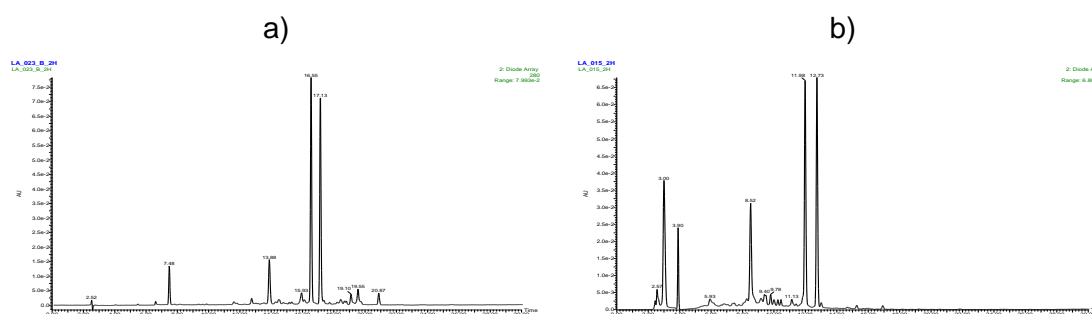


Figura R.C.12. Comparació entre la mateixa reacció duta a terme usant TEMPO (a) i O₂ (b). Tot i les diferents condicions d'anàlisi, la diferència en la quantitat de subproductes generats és clara.

Per tal de superar les limitacions que la seqüència peptídica imposava sobre la reacció de ciclació i de resoldre el que semblava ser un efecte estructural, vàrem decidir afegir clorur de liti al medi de reacció degut al seu efecte caotrópic. Proves efectuades amb diferents concentracions d'aquesta sal van demostrar que afavoria la ciclació dels pèptids alhora que disminuïa la formació de subproductes, bàsicament el d'oxidació del tiol de la cisteïna interna, dos processos que estaven mútuament relacionats.

Després d'aquest descobriment es van seguir fent experiments per a posar a punt una metodologia que permetés ciclar i derivatitzar simultàniament qualsevol pèptid usant CPDs amb diferents substituents. Finalment, i després de força optimització, es va establir un protocol en què la reacció de ciclació es duia a terme a 60 °C en presència de LiCl 2 molal, i on l'addició de TEMPO es realitzava en el mateix moment en que CPD i pèptid es mesclaven. En referència a l'addició d'aquest radical, totes les proves obtingudes en el transcurs d'aquest treball indiquen que el seu temps de vida mitja en el cru de reacció no és molt llarg, i que per tant són necessàries addicions continuades d'aquest reactiu en quantitats sub-estequiomètriques. Aquest mètode permet la formació dels conjugats de manera neta i amb bons rendiments, tal com es pot veure en la **Figura R.C.13**. Alguns experiments han demostrat que l'addició d'un excés de TEMPO respecte al pèptid pot produir l'oxidació del tiol de la cisteïna interna a àcid sulfònic abans de la ciclació, reduint el rendiment de la reacció.

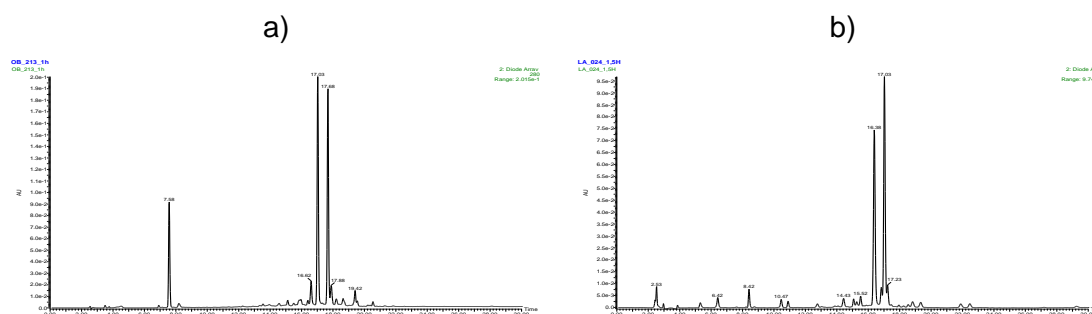


Figura R.C.13. Reacció entre la CPD derivatitzada amb biotina i un pèptid model (a) i entre aquesta mateixa CPD i un pèptid amb la seqüència de l'oxitocina (b).

Desgraciadament, els primers productes sintetitzats i aïllats (deixant de banda el model emprat per a la determinació estructural i del coeficient d'absorció) es van fer malbé en el procés de purificació. Gràcies a síntesis independents, es va poder demostrar que el problema rere la seva “descomposició” era l'epimerització del carboni α de la cisteïna *N*-terminal, deguda, molt probablement, a l'estabilització per ressonància del carbaní que es pot generar en aquesta posició. Es va observar que pHs bàsics o molt lleugerament

bàsics eren capaços de promoure l'epimerització de la L-cisteïna en posició *N*-terminal, però que, curiosament, el seu isòmer D no era tan propens a la isomerització. Per aquest motiu, es va considerar adequat usar D-cisteïna en la posició *N*-terminal dels pèptids a ciclar, per tal de minimitzar la possible epimerització d'aquest aminoàcid.

Usant aquesta estratègia, es van poder derivatitzar i ciclar diversos pèptids amb rendiments d'entre el 60 i 85 % determinats mitjançant anàlisi per HPLC/MS del cru i amb rendiments d'aïllament entre el 30 i el 40 %. Tots ells van resultar ser poc solubles en aigua pura, fet pel qual la seva quantificació i anàlisi es va haver de fer en metanol.

Conclusions

1. Els experiments duts a terme per identificar un grup protector adequat per a la protecció d'1,3-diens han sigut infructuosos. S'ha explorat i desenvolupat una estratègia diferent que permet que pèptids derivatitzats amb diens siguin usats en reaccions de Diels-Alder per a la obtenció de conjugats. En aquesta metodologia la reacció de conjugació es duu a terme entre un pèptid derivatitzat amb un diè, totalment protegit i unit a resina, i un derivat de maleimida soluble (3-5 equivalents). El diè, inestable a les condicions de desancoratge i desprotecció de poliamides, no s'exposa a aquest tractament, i el cicloadducte generat hi roman estable. En contrast amb metodologies prèviament descrites, aquesta alternativa no presenta limitacions en quant a seqüència o a l'elecció dels grups protectors a usar per a les cadenes laterals dels aminoàcids.
2. Pel que fa als detalls del procés de conjugació sobre resina, cal dir que:
 - i) Les peptidil-resines derivatitzades amb àcid (*E*)-4,6-heptadienoic han donat millors resultats que aquelles derivatitzades amb àcid 3-(2-furil)propanoic, tant en

termes de rendiment en la reacció de Diels-Alder com quant a l'estabilitat dels cicloadductes generats.

ii) Les poliamides es sintetitzen usant monòmers comercials i amb protectors estàndard sobre resines compatibles amb l'ús d'aigua. No cal sintetitzar cap monòmer especial.

iii) La reacció de Diels-Alder sobre resina tolera tant l'ús d'aigua com de mesclures aigua/dissolvent orgànic.

iv) Es poden usar maleïmides poc solubles, fins i tot en suspensió.

v) La desprotecció de maleïmides protegides amb 2,5-dimetilfurà es pot dur a terme simultàniament amb la reacció de Diels-Alder sobre resina. Això permet l'ús de maleïmides protegides per a la preparació de conjugats amb diferents punts d'unió.

3. Es poden preparar dobles conjugats a partir de peptidil-resines derivatitzades amb diens i que continguin cisteïna. La reacció de Diels-Alder permet la primera derivatització, i una segona reacció entre el tiol de la cisteïna i una maleïmida, després del tractament de desprotecció i desancoratge, forneix el doble conjugat.
4. Els oligonucleòtids derivatitzats amb un diè i units a resina no resisteixen les condicions necessàries per dur a terme la reacció de Diels-Alder en fase sòlida. Tot i que estudis preliminars duts a terme amb seqüències poli-dT van donar bons resultats, la metodologia va fallar a l'hora de generar conjugats de seqüències mixtes i més llargues. Això indica que les seqüències poli-dT no són necessàriament bons models d'oligonucleòtids.
5. S'ha posat de manifest que les ciclopent-4-en-1,3-diones 2,2-disubstituïdes (CPDs), que es van escollir com a anàlegs no hidrolitzables de maleïmida, tenen una reactivitat fins ara no descrita. Mentre que les maleïmides reaccionen de manera irreversible amb qualsevol tipus de tiol, les reaccions de tipus Michael entre CPDs i cisteïnes que ocupen una posició interna o C-terminal en la cadena peptídica, que no posseeixen una amina lliure, és reversible. Aquest resultat va ser corroborat pels càlculs teòrics realitzats per en Lluís Raich i la Dra. Carme Rovira. Per contra, aquelles cisteïnes que tenen l'amina lliure (que ocupen la posició N-terminal del pèptid) reaccionen amb les CPDs per acabar formant un adducte estable que té una massa 20 Da menor (**M**-20 Da) que la de l'adducte de tipus Michael (que té una massa de **M** Da).

6. La formació de l'adducte **M-20 Da** té lloc, plausiblement, a través dels següents passos: en primer lloc, l'addició conjugada del tiol de la cisteïna *N*-terminal a la CPD genera un adducte de tipus Michael. Aquest producte experimenta, de manera immediata, una ciclació intramolecular per reacció entre l'amina *N*-terminal del pèptid i un dels grups ceto de la CPD, donant lloc a un intermedi amb dos anells fusionats i una massa 18 Da per sota (**M-18 Da**) de la de l'adducte de Michael inicialment format. Una posterior oxidació d'aquest adducte **M-18 Da** genera el producte final **M-20 Da**. Aquesta és l'etapa limitant de la reacció, i es pot accelerar incrementant la temperatura o afegint un excés de CPD. L'estructura de l'adducte **M-20 Da**, que absorbeix al voltant de 330 nm, ha estat confirmada per RMN.
7. El mètode usat per a fer el seguiment de les reaccions entre pèptids que contenen una cisteïna en la posició *N*-terminal i CPDs és crític per a poder interpretar de manera correcta els resultats d'aquestes transformacions. La tècnica d'HPLC acoblada a espectrometria de masses amb ionització per electrospray és el mètode elegit per seguir aquest tipus de reaccions. El protocol que usa HPLC i anàlisi per espectrometria de masses MALDI-TOF dels pics col·lectats pot portar a interpretacions errònies, doncs l'adducte **M-18 Da** s'oxida al **M-20 Da** durant l'aïllament del producte, en les condicions de ionització, o ambdues.
8. Tant la cisteïna com el cisteïnat de metil (ambdós 1,2-aminotiols) reaccionen amb CPDs per donar els corresponents adductes **M-20 Da**. En el cas de la cisteïna, i segurament degut a la presència de l'àcid carboxílic, aquest producte experimenta la hidròlisi de l'imina. La reacció entre homocisteïna, un 1,3-aminotiol, i CPDs només genera els adductes de tipus Michael, que són reversibles.
9. És possible derivatitzar pèptids amb CPDs mitjançant mètodes estàndard de síntesi de pèptids en fase sòlida i usar-los per a la formació de conjugats amb PNAs derivatitzats amb cisteïna. Els conjugats resultants són estables a condicions àcides, però pHs bàsics tenen un efecte negatiu sobre ells.
10. La diferent reactivitat de les CPDs front a 1,2-aminotiols i altres tiols s'ha explotat per: 1) marcar selectivament un pèptid que conté una cisteïna en posició *N*-terminal en una mescla que conté pèptids amb cisteïna en una altra posició, i 2)

derivatitzar doblement un pèptid que conté una cisteïna interna i una en la posició *N*-terminal.

11. És possible ciclar pèptids que contenen una cisteïna en la posició *N*-terminal i una en una posició interna (o *C*-terminal) mitjançant la reacció amb CPDs. En el cas que aquestes incorporin algun substituent (com biotina o un fluoròfor), la ciclació i conjugació d'aquests pèptids té lloc de manera simultània. La formació de l'adducte **M**-20 Da, que té lloc en primera instància, és seguida per una addició de tipus Michael del tiol intern a aquest adducte i per una posterior oxidació que genera un sistema conjugat que absorbeix al voltant de 370 nm. La ciclació ha estat corroborada per RMN.
12. La reacció de ciclació es veu accelerada per la presència d'oxidants com O₂ o TEMPO. S'ha demostrat que el millor mètode per obtenir pèptids cíclics amb bons rendiments és l'addició contínua de quantitats sub-estequiomètriques de TEMPO. L'addició de LiCl també té un efecte beneficiós, especialment sobre aquelles seqüències més difícils de ciclar. Respecte a aquest punt, s'ha vist que la presència d'un residu aromàtic al costat de la cisteïna *N*-terminal té un efecte negatiu en el procés de ciclació.
13. La cisteïna *N*-terminal dels pèptids ciclats usant una CPD té tendència a epimeritzar en la seva posició α en presència de bases, tal com s'ha pogut comprovar per síntesi independent dels anàlegs cíclics amb una D-cisteïna a la posició *N*-terminal. Curiosament, aquests últims tenen menys tendència a epimeritzar que els seus epímers amb L-cisteïna, per la qual cosa es recomana l'ús de la versió no proteïnogènica d'aquest aminoàcid.

Abreviacions

ATM: Adducte de tipus Michael

CPD: Ciclopent-4-en-1,3-diona 2,2-disubstituïda

DA: Diels-Alder

DMSO: Dimetilsulfòxid

HMBC: Correlació multienllaç heteronuclear

MALDI-TOF: Ionització per desorció làser assistida per matriu-Temps de vol

PNA: Àcid nucleic peptídic

RMN: Ressonància magnètica nuclear

TEMPO: Radical (2,2,6,6-tetrametilpiperidin-1-il)oxil

TFA: Àcid trifluoroacètic

TIS: Triisopropilsilà

Bibliografia

- (1) Iqbal, K.; del C. Alonso, A.; Chen, S.; Chohan, M. O.; El-Akkad, E.; Gong, C.-X.; Khatoon, S.; Li, B.; Liu, F.; Rahman, A.; Tanimukai, H.; Grundke-Iqbal, I. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **2005**, 1739 (2–3), 198.
- (2) Rubinsztein, D. C. *Nature* **2006**, 443 (7113), 780.
- (3) Ranum, L. P. W.; Cooper, T. A. *Annu. Rev. Neurosci.* **2006**, 29 (1), 259.
- (4) Orr, H. T.; Zoghbi, H. Y. *Annu. Rev. Neurosci.* **2007**, 30 (1), 575.
- (5) DrugBank: Biotech Drugs
<http://www.drugbank.ca/drugs?approved=1&c=name&d=up&type=biotech> (accessed Oct 22, 2015).
- (6) Tse, M. T. *Nat. Rev. Drug Discov.* **2013**, 12 (3), 179.
- (7) Hebert, M. F. *Adv. Drug Deliv. Rev.* **1997**, 27 (2), 201.
- (8) Wolin, E. M. *Gastrointest. Cancer Res. GCR* **2012**, 5 (5), 161.
- (9) de Araújo, A. D.; Palomo, J. M.; Cramer, J.; Seitz, O.; Alexandrov, K.; Waldmann, H. *Chem. – Eur. J.* **2006**, 12 (23), 6095.
- (10) Hoogewijs, K.; Deceuninck, A.; Madder, A. *Org. Biomol. Chem.* **2012**, 10 (20), 3999.

- (11) Otto, S.; Blokzijl, W.; Engberts, J. B. F. N. *J. Org. Chem.* **1994**, *59* (18), 5372.
- (12) Otto, S.; Engberts, J. B. F. N. *Org. Biomol. Chem.* **2003**, *1* (16), 2809.
- (13) Lubineau, A. *J. Org. Chem.* **1986**, *51* (11), 2142.
- (14) Pirrung, M. C. *Chem. – Eur. J.* **2006**, *12* (5), 1312.
- (15) Blokzijl, W.; Blandamer, M. J.; Engberts, J. B. F. N. *J. Am. Chem. Soc.* **1991**, *113* (11), 4241.
- (16) Sánchez, A.; Pedroso, E.; Grandas, A. *Org. Lett.* **2011**, *13* (16), 4364.
- (17) Sánchez, A.; Pedroso, E.; Grandas, A. *Org. Biomol. Chem.* **2012**, *10* (42), 8478.
- (18) Elduque, X.; Pedroso, E.; Grandas, A. *Org. Lett.* **2013**, *15* (8), 2038.
- (19) Elduque, X.; Sánchez, A.; Sharma, K.; Pedroso, E.; Grandas, A. *Bioconjugate Chem.* **2013**, *24* (5), 832.
- (20) Guerlavais, T.; Meyer, A.; Debart, F.; Imbach, J.-L.; Morvan, F.; Vasseur, J.-J. *Anal. Bioanal. Chem.* **2002**, *374* (1), 57.
- (21) Meyer, A.; Spinelli, N.; Imbach, J.-L.; Vasseur, J.-J. *Rapid Commun. Mass Spectrom.* **2000**, *14* (4), 234.
- (22) Baldwin, A. D.; Kiick, K. L. *Bioconjugate Chem.* **2011**, *22* (10), 1946.
- (23) Ryan, C. P.; Smith, M. E. B.; Schumacher, F. F.; Grohmann, D.; Papaioannou, D.; Waksman, G.; Werner, F.; Baker, J. R.; Caddick, S. *Chem. Commun.* **2011**, *47* (19), 5452.
- (24) Lyon, R. P.; Setter, J. R.; Bovee, T. D.; Doronina, S. O.; Hunter, J. H.; Anderson, M. E.; Balasubramanian, C. L.; Duniho, S. M.; Leiske, C. I.; Li, F.; Senter, P. D. *Nat. Biotechnol.* **2014**, *32* (10), 1059.
- (25) Kalia, J.; Raines, R. T. *Bioorg. Med. Chem. Lett.* **2007**, *17* (22), 6286.
- (26) Northrop, B. H.; Coffey, R. N. *J. Am. Chem. Soc.* **2012**, *134* (33), 13804.
- (27) Stolz, R. M.; Northrop, B. H. *J. Org. Chem.* **2013**, *78* (16), 8105.