

**NUCLEIC ACID TEMPLATED SYNTHESIS
OF PEPTIDE BONDS**

by

SAMUEL NUÑEZ I PERTIÑEZ



**UNIVERSITY OF
BIRMINGHAM**

A thesis submitted to
The University of Birmingham
for a degree of
DOCTOR OF PHILOSOPHY

School of Chemistry
College of Engineering and Physical Sciences
University of Birmingham
February 2020

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The conventional synthesis of peptides, by reactions performed under highly concentrated conditions or on solid phase, have been widely utilised to produce a variety of oligomers. This has facilitated the discovery of bioactive compounds and the mechanistic study of their function. Recently, nanotechnology has facilitated the synthesis of peptide bonds by directing the interactions between reactive groups, in order to produce the desired peptides. This thesis focuses on the use of nucleic acid templated synthesis (NATS) of peptide bonds, and on overcoming the hydrolysis of activated ester-labelled nucleic acids in aqueous solutions.

Chapter 1 gives a general overview of peptide bond formation strategies, highlighting nanotechnological approaches. The current scope and limitations of nucleic acid templated synthesis are also discussed. Chapter 2 is focused on the use of deep eutectic solvents (DESs) to minimise hydrolysis of activated esters, for the DNA templated synthesis (DTS) of peptide bonds. Chapter 3 describes the use of highly stable phenol ester-conjugated DNA adapters in aqueous buffered solution, and their *in-situ* activation towards aminolysis through oxo-ester mediated native chemical ligation (NCL). Chapter 4 explores if an alternative peptide nucleic acid (PNA) template could form stable duplexes in organic solvent, and thus, potentially template peptide bond synthesis reactions. Selective labelling of oligonucleotides has been a fundamental procedure in this thesis; in Appendix 2, the conjugation of fluorescent dibromomaleimide and dichloromaleimide to DNA has been explored; in addition, reductive amination has been explored as a chemistry for the modification of amine-labelled DNA with phenol esters.

*What we know is a drop,
what we do not know is an ocean.*

Isaac Newton

Acknowledgements

From early on in a PhD, or probably even before, doctoral students are made aware that at some point, their scientific discoveries will have to be compiled into a thesis. However, now that the time to submit my own thesis approaches, I cannot avoid a strange feeling about the fact that it is nearly completed. Along this journey, there has been a few people whose help and support has been key to complete this work. To all of them, I would like to express my sincere gratitude.

Firstly, a big thank you to Prof. Rachel O'Reilly, for giving me the opportunity to work in the fascinating (and sometimes challenging) field of nucleic acids nanotechnology, for being an example of hard work and commitment to scientific discovery, and for letting me follow my own ideas with a proactive and positive approach.

Thanks to the members of the O'Reilly and Dove groups, for their support, for always sparing some time to discuss experimental results, suggesting new experiments, and for providing a steady supply of cake that helped me to overcome even the most bitterly frustrating experimental drawbacks. In particular, thanks to the members of the team DNA: Dr. Thomas Wilks, Lucy Arkinstall, Dr. Charlotte Zammit and Dr. Jennifer Frommer, for their support over this time. Thanks to Lucy for your collegiality and for always being willing to discuss the unforeseen experimental results.

Thanks to Dr. Alexander Roberts (WMG) for his support using the Karl-Fischer titrator. I also would like to say thanks to Prof. Andrew Turberfield and his group, especially Dr. Wenjing Meng, Dr. Jonathan Bath and Dr. Robert Oppenheimer, for their support in helping me to better understand the interactions between oligonucleotides. Thanks to Dr. Robert Oppenheimer for helping me with the electrophoresis analysis, for all the

interesting discussions about science, and for being always happy to look at my data with a fresh perspective.

Thank you for your diligent proofreading to: Dr. Thomas Wilks, Dr. Stefan Lawrenson, Lucy Arkinstall and Jonathan Husband.

I count myself fortunate to have had a wonderful family that have taught me essential things that no lecture can teach. Thanks to Rafael and Carmen for always being there, patiently listening to all my “lab stories”, and always giving me the best counsels. Also, for sending Mediterranean food that gave my, somewhat limited, “culinary repertoire” a bit of variety.

I would like to finish by saying thanks to my dear Laia, for sharing this path, for understanding the highs and lows of a PhD, for believing in my skills even when I questioned them, and for the invaluable support over these years. I am looking forward to our next adventure.

Declaration

This thesis is submitted to The University of Birmingham in support of my application for the degree of Doctor of Philosophy. It has been composed by me and it has not been submitted for any other degree or professional qualification.

The work presented herein has been conducted by myself, except the following:

- The reagent **2-9** was provided by Dr. Wenjing Meng (University of Oxford)
- The DNA melting profile in glycholine / H₂O described in Figure 2-16 was performed by Dr. Thomas Wilks (University of Birmingham)
- The basic considerations for the analysis of the effect of linker length on the rate of templated reactions, described in figure **3-18**, were initially pointed out by Dr. Robert Oppenheimer (University of Oxford). Then, the model was developed independently.
- Yujie Xie (Univesity of Warwick) provided the Adamantyl-dichloromaleimide used in Appendix 2.

Parts of the work described in Chapter 2 have been published in: S. Núñez-Pertíñez and T. R. Wilks, *Front. Chem.*, 2020, **8**:41.

Parts of the work described in Chapter 4 have been published in: S. Núñez-Pertíñez, T. R. Wilks and T. K. O'Reilly, *Org. Biomol. Chem.*, 2019, **17**, 7874-7877.

Contents

Contents	1
List of figures	4
List of Schemes	8
List of tables	10
Abbreviations	11
Chapter 1. Introduction. Templated synthesis of peptides: Reactivity, sequence encoding and programmability.	15
1.1 The peptide bond and its biological relevance	15
1.1.1 Peptide bonds	15
1.1.2 Peptides and proteins: structure and function	16
1.2 Nature's approach to peptide synthesis	19
1.2.1 Encoding the identity of peptides: the genetic code	19
1.2.2 From genetic information to peptides and proteins: ribosomal synthesis of peptides	22
1.3 Chemical synthesis of peptides	25
1.3.1 Solution phase synthesis	25
1.3.2 Solid phase synthesis of peptides	27
1.4 Templated synthesis of peptide bonds	29
1.4.1 Nanotechnology applied to chemical synthesis	29
1.4.2 Peptide templated synthesis of peptide bonds	33
1.4.3 Rotaxane templated peptide bond formation	35
1.4.4 Nucleic acid templated synthesis (NATS) of peptide bonds	37
1.4.5 The potential of NATS of peptide bonds: from encoding the identity of peptides to artificial ribosomes.	45
1.3 Project aims	59
1.4 Bibliography	61
Chapter 2. DNA templated synthesis (DTS) in Deep Eutectic Solvents (DES)	69
2.1 Introduction	69
2.2 Results and discussion	74
2.2.1 Small molecule activated ester stability in DESs	74
2.2.2 DNA activated ester conjugate stability in DES	85
2.2.3 DNA templated peptide bond formation in DES	103
2.2.4 Storage of DNA conjugated activated esters in DES	110
2.2.5 Reductive amination in DES	112
2.3 Conclusions	117
2.4 Experimental section	119
2.4.1 DNA strands and modifications	119
2.4.2 General methods	120
2.4.3 Preparation of DESs	121
2.4.4 Correction of the acidity of DESs	121
2.4.5 HPLC stability assay of activated esters in DESs	122
2.4.6 ¹ H-NMR spectroscopy stability assay of activated esters in DESs	123
2.4.7 Small molecule peptide bond formation in glycoline	124
2.4.8 Synthesis of DNA conjugated activated esters	125

2.4.9 Preparation of FAM labelled <i>L</i> -alanine <i>N</i> -hydroxymaleimide ester	126
2.4.10 Fluorescence stability assay using a DNA hairpin	129
2.4.11 Derivatisation of the equation 1	130
2.4.12 Thermal stability of dsDNA in DES / H ₂ O mixtures	132
2.4.13 DNA templated synthesis in DES	132
2.4.14 Amine-labelled DNA reactivity in DES	133
2.4.15 Storage of DNA conjugated activated esters in DES	134
2.4.16 Synthesis of aldehyde-labelled DNA	135
2.4.17 DNA templated reductive amination in DES	135
2.5 References	137
Chapter 3. DNA templated synthesis of peptide bonds in aqueous solution: in situ activation of phenol esters	143
3.1 Introduction: co-localised catalysts and nucleic acid templated synthesis	143
3.2 Results and discussion	147
3.2.1 Design of the NCL system	147
3.2.2 Synthesis of an <i>L</i> -Ala phenol ester donor strand	149
3.2.3 Synthesis of the catalyst and acceptor group	151
3.2.4. Assay of a templated NCL reaction using a <i>L</i> -Ala phenol ester donor	155
3.2.5. Synthesis of a <i>D</i> -biotin phenol ester donor strand	158
3.2.6 Optimisation of the post-DTS hydrolysis conditions	163
3.2.7 NCL mediated biotin transfer: Cys acceptor	165
3.2.8 Comparison of a co-localised thiol-labelled DNA template and a Cys-labelled DNA acceptor	173
3.3 Conclusions	183
3.4 Experimental section	185
3.4.1 General methods	185
3.4.2 Sequences and modifications	185
3.4.3 Synthesis of 3-3	186
3.4.4 Synthesis of 3-5	186
3.4.5 Synthesis of 3-7	187
3.4.6 Synthesis of 3-11	187
3.4.7 DNA templated NCL (I): <i>L</i> -Ala-Boc transfer	188
3.4.8 Synthesis of 3-19	188
3.4.9 Degradation of maleimide by DMAP	190
3.4.10 Optimisation of the hydrolysis of 3-23	191
3.4.11 DNA templated NCL (II): Hydrolysis of the starting material in the presence of a deprotected thiol	191
3.4.12 Synthesis of 3-28	191
3.4.13 Reverse-NCL: assay of the process and the stability of <i>D</i> -biotin-labelled DNA	192
3.4.14 DNA templated NCL (III): Protection of the thiol before hydrolysis	192
3.4.15 Estimation of the effect of linker length on a bimolecular reaction rate	193
3.5 References	195
Chapter 4. Peptide nucleic acids (PNAs) as potential templates for the synthesis of peptide bonds in organic solvent	199
4.1 Introduction: PNAs and their hybrid properties between peptide and nucleobase materials.	199
4.1.1 PNA hybridisation in the presence of organic solvent	199
4.1.2 Solid phase synthesis of PNA	202
4.1.3 Design of a model double-stranded PNA (dsPNA) system	203
4.1.4 Analytical methods to study the thermal stability of dsPNA in the presence of organic solvent	205

4.2 Results and discussion	209
4.2.1 Synthesis of 10-mer and 15-mer PNA strands	209
4.2.2 Synthesis of fluorophore and quencher labelled PNA	218
4.2.3 Correlation of the thermal stability of dsDNA, PNA·DNA and dsPNA duplexes determined by UV spectroscopy, fluorescence spectroscopy and microcalorimetry in aqueous solution.	227
4.2.4. Micro-differential scanning calorimetry (micro-DSC) study of the thermal stability of dsPNA in aqueous / organic solvent mixture.	229
4.2.5. Isothermal annealing of dsPNA in high organic solution	234
3.2.6 Future directions	238
4.3 Conclusions	241
4.4 Experimental section	243
4.4.1 General methods	243
4.4.2 Sequences and modifications	244
4.4.3 Solid phase synthesis methods	245
4.4.4. Design of 15-mer PNA sequences	246
4.4.5 Synthesis of 4-7	247
4.4.6 Synthesis of 4-10	247
4.4.7 Thermal stability in aqueous solution: comparative study	248
4.4.8 Solubility of 15-mer PNA in organic solvents	248
4.4.9 Micro-DSC raw data	249
4.4.10 Isothermal annealing	254
4.5 References	255
Future prospects	261
Appendix 1 Characterisation methods	263
Reversed phase high performance liquid chromatography (RP-HPLC)	263
Liquid chromatography-mass spectrometry	263
Polyacrylamide gel electrophoresis	264
Nuclear magnetic resonance spectroscopy	264
High resolution mass spectrometry	265
Nanopure water	265
Appendix 2 Additional projects	267
A2.1 Maleimide fluorophore labelling of DNA	267
A2.1.1 Internal labelling of DNA with methyl-dithiomaleimide	267
A2.1.2 Adamantyl-aminochloromaleimide functionalisation of DNA	268
A2.2 Reductive amination and amide coupling of biotin phenol ester to DNA	270

List of figures

Chapter 1. Introduction. Templated synthesis of peptides: Reactivity, sequence encoding and programmability.	15
Figure 1-1. Amide functional group and peptides	16
Figure 1-2. Different levels of peptide and protein structure	18
Figure 1-3. Description of nucleic acids	20
Figure 1-4. Representation of the genetic storage of information and its transcription	21
Figure 1-5. Schematic description of the fundamental aspects of the translation of genetic information into peptide sequences	23
Figure 1-6. Examples of nanotechnology applications	31
Figure 1-7. Coiled-coil templated synthesis of a peptide bond	33
Figure 1-8. Representation of the rotaxane templated synthesis	36
Figure 1-9. Representation of the enabling architectures for NATS	39
Figure 1-10. Schematic description of the split and pool generation of DNA encoded peptide libraries	46
Figure 1-11. Schematic description of DNA routing	48
Figure 1-12. Schematic representation of multistep nucleic acid templated synthesis of peptide bonds through DNA templated synthesis	50
Figure 1-13. Autonomous DNA walker templated synthesis of peptides	52
Figure 1-14. Autonomous hybridisation chain reaction (HCR) templated synthesis of peptides	54
Figure 1-15. Comparison between unnatural amino acids used for autonomous NATS, and natural amino acids	57
Chapter 2 DNA templated synthesis (DTS) in Deep Eutectic Solvents (DES)	69
Figure 2-1. General description of deep eutectic solvents	70
Figure 2-2. Representation of the small molecule 2-1 stability experiment with HPLC analysis	79
Figure 2-3. Results of the RP-HPLC stability study of 2-1 in several solvents	80
Figure 2-4. ¹ H-NMR stability experiment of the activated ester 2-1 in several DES	82
Figure 2-5. Fraction of activated ester after 6 h in several DES - additive mixtures, determined by ¹ H-NMR spectroscopy	83
Figure 2-6. LC-MS analysis of DNA conjugated activated esters using high aqueous mobile phase in RP-UHPLC	87
Figure 2-7. Fluorescence resonance energy transfer (FRET) based assay to follow the solvolysis of a DNA conjugated activated ester in DESs and aqueous solution	88
Figure 2-8. Procedure to transfer a thiol and TAMRA labelled DNA - magnetic nanoparticle (NP) assembly from PBS solution into DES	92

Figure 2-9. Synthesis of a DNA conjugated activated ester hairpin (2-19) and transfer to fresh DES	94
Figure 2-10. Raw data of the fluorescence assay of the stability of the DNA conjugated activated ester (2-19) in glycholine solutions (Gly : ChCl 2:1) with 7, 50 and 90 vol% H ₂ O	95
Figure 2-11. Generic representation of an equation of the type eq. 3 and eq. 4.	97
Figure 2-12. Description of the data from the fluorescence stability assay in DES	98
Figure 2-13. Comparison between the slow hydrolysis in glycholine 7 and 15 vol% H ₂ O (left), and fast aminolysis in 7 and 15 vol% buffer (right)	99
Figure 2-14. Attempts to determine F_{Max} through strand displacement strategies	100
Figure 2-15. Results of the attempts to determine the value of F_{Max}	102
Figure 2-16. Evaluation of the stability of dsDNA in several DES solvents using fluorescence quenching	104
Figure 2-17. Comparison between a DNA templated peptide bond synthesis in Gly _{0.07} and in aqueous buffered solution	107
Figure 2-18. Amine reactivity test in 1 M NaCl aqueous solution and in DES	109
Figure 2-19. Storage of DNA conjugated activated esters in DES and PBS followed by DNA templated amidation reaction	111
Figure 2-20. Effect of the addition of 200 mM NaBH ₃ CN on the T_m of dsDNA	113
Figure 2-21. Synthesis of aldehyde-modified DNA	114
Figure 2-22. DNA templated reductive amination.	115
Figure 2-23. Chemical structure of commercially available DNA modifications	120
Figure 2-24. Evolution of the pH of an ethaline / choline acetate (top) and glycholine / urea 2 (bottom) solutions	122
Figure 2-25. Stacked ¹ H-NMR spectra of the extracted solutions glycholine 50 vol% H ₂ O	124
Figure 2-26. Excitation and emission fluorescence spectra of FAM (2-15) and TAMRA in Gly _{0.07}	130
Figure 2-27. Comparison of the evolution of fluorescence over time in Gly _{0.07} , Gly _{0.5} and Gly _{0.07} / 2 mM Et ₃ N	130
Figure 2-28. Effect of the addition of DMF to Gly _{0.07} on the T_m	132
Chapter 3. DNA templated synthesis of peptide bonds in aqueous solution: in situ activation of phenol esters	143
Figure 3-1. Examples of amino acid esters	144
Figure 3-2. Representation of the local geometry around the Cys thioester intermediate in a NCL reaction	149
Figure 3-3. HPLC analysis of the synthesis of 3-5	151
Figure 3-4. HPLC analysis of the synthesis of 3-7	152
Figure 3-5. HPLC analysis of the synthesis of 3-10	153
Figure 3-6. HPLC analysis of the Fmoc deprotection of 3-10	155

Figure 3-7. Denaturing PAGE analysis of a DNA templated NCL using a co-localised thiol catalyst (3-7, T-SH), a L-Ala phenol ester (3-5, PhOAla) and an amine-labelled acceptor (3-8, NH ₂)	157
Figure 3-8. ¹ H-NMR spectroscopic study of the stability of 3-17 in the presence of the Steglich esterification reagents	160
Figure 3-9. Quantification of the conversion into 3-23	162
Figure 3-10. Optimisation of the conditions for the hydrolysis of 3-23	164
Figure 3-11. Native PAGE analysis of the NCL reaction using a Cys-labelled DNA acceptor	167
Figure 3-12. HPLC analysis of the synthesis of 3-28	170
Figure 3-13. Analysis of the reverse-NCL reaction by native PAGE	171
Figure 3-14. Native PAGE analysis of the stability of a D-biotin-labelled DNA in TCEP solution, at high pH and temperature	173
Figure 3-15. Native PAGE analysis of the NCL process in the absence of a thiol catalyst (lanes 8-9) and with a Cys-labelled DNA acceptor at several pHs (lanes 10-14)	175
Figure 3-16. Native PAGE analysis of a DNA templated NCL with a Cys-labelled DNA acceptor	177
Figure 3-17. Native PAGE analysis of an uncatalysed DNA templated peptide bond formation (lanes 5-10) and with a thiol-labelled DNA template catalyst (lanes 11-14)	178
Figure 3-18. Modelled effect of the linker length on the reaction rate	180
Figure 3-19. Chemical structure of commercially available DNA modifications	185
Figure 3-20. Comparison of the native PAGE analysis of a sample before and after the addition of streptavidin, in the presence of a 100 - 20 nucleotides (nt) ssDNA ladder	190
Chapter 4. Peptide nucleic acids (PNAs) as potential templates for the synthesis of peptide bonds in organic solvent	199
Figure 4-1. Structural comparison between the repeating unit of DNA (left) and PNA (right)	200
Figure 4-2. Representation of evolution of the T _m of a 10-mer dsPNA in the presence of DMF, determined by UV spectroscopy	201
Figure 4-3. Elongation of the reported 10-mer dsPNA using the web application Nupack	204
Figure 4-4. Spectroscopic determination of the T _m of dsPNA by a) UV spectroscopy and b) fluorescence spectroscopy	206
Figure 4-5. Calorimetric determination of the T _m of dsPNA	207
Figure 4-6. Investigation of the coupling efficiencies using oxyma / DIC and PyBOP / 2,6-lutidine : (iPr) ₂ NEt	213
Figure 4-7. RP-HPLC analysis of unlabelled 10-mer and 15-mer strands	218
Figure 4-8. RP-HPLC analysis of the labelled PNA sequences 4-7 and 4-10	221
Figure 4-9. Characterisation of the spectroscopic properties of 4-7 and 4-10 fluorophore / quencher pair	222
Figure 4-10. Thermal denaturalisation of dsPNA 4-7 and 4-10 in 50 vol% DMF observed by micro-DSC	222

Figure 4-11. Synthesis of Trp-labelled PNA (4-11)	224
Figure 4-12. Tryptophan / dansyl labelled PNA quenching	226
Figure 4-13. Comparison of three techniques for the measurement of the T_m of nucleic acid duplexes in aqueous solution	228
Figure 4-14. Solubility study of 15-mer PNA in several organic solvents	230
Figure 4-15. Melting temperatures, determined using micro-DSC	231
Figure 4-16. Conversion of volumetric fraction into molar fraction using eq. 1	232
Figure 4-17. Evolution of the T_m as function of the molar content of DMF	233
Figure 4-18. Isothermal annealing of individual fluorophore and quencher-labelled PNAs, in the absence of a pre-heating step applied to individual PNA solutions	235
Figure 4-19. Isothermal annealing of fluorophore and quencher-labelled PNA in 90 vol% DMF, preceded by a heating cycle on individual stock solutions and thermal denaturalisation and re-annealing of dsPNA	237
Figure 4-20. Chemical structure of DNA and PNA modifications	245
Figure 4-21. Sigmoidal baseline subtracted micro-DSC data of 10-mer dsPNA thermal melting in DMF / H ₂ O solutions	250
Figure 4-22. Integrated molar heat capacity as function of the temperature	250
Figure 4-23. Sigmoidal baseline subtracted micro-DSC data of 15-mer dsPNA thermal melting in DMF / H ₂ O solutions	251
Figure 4-24. Integrated molar heat capacity as function of the temperature	252
Figure 4-25. Sigmoidal baseline subtracted micro-DSC data of 10-mer PNA·DNA thermal melting in DMF / H ₂ O solutions	253
Figure 4-26. Integrated molar heat capacity as function of the temperature	253
Figure 4-27. Superimposition of PNA-Trp and PNA-dansyl fluorescence spectra in DMF	254
Appendix 2 Additional projects	268
Figure A2-1. LC-MS analysis of the synthesis of dithiomaleimide-labelled DNA A2-2	268
Figure A2-2. LC-MS analysis of the synthesis of aminochloromaleimide-labelled DNA A2-4	270
Figure A2-3. LC-MS (Bruker AmazonX Ion Trap) analysis of the reductive amination attempt with A2-5 and 3-4	272

List of Schemes

Chapter 1. Introduction. Templated synthesis of peptides: Reactivity, sequence encoding and programmability.	15
Scheme 1-1. Several approaches to amide bond synthesis	26
Scheme 1-2. Several methods of activation of carboxylic acids towards nucleophilic acyl substitution	27
Scheme 1-3. General representation of the peptide solid phase synthesis stages	29
Scheme 1-4. Simplified mechanism of the splicing of an intein from its N-extein and C-extein, yielding a new peptide bond	34
Scheme 1-5. Representation of the carbodiimide mediated ligation of a poly-A and poly-T nicked double stranded DNA (dsDNA)	38
Scheme 1-6. Schematic description of several nucleic acid templated peptide bond formation	44
Chapter 2 DNA templated synthesis (DTS) in Deep Eutectic Solvents (DES)	69
Scheme 2-1. Several examples of chemical transformations in Deep Eutectic Solvents	72
Scheme 2-2. Comparison between the system of interest composed of a N-hydroxysuccinimide ester conjugated to DNA, and the model small molecule system	75
Scheme 2-3. Synthesis of the dipeptide 2-5 in DESs	84
Scheme 2-4. Synthesis of the DNA conjugated activated esters 2-8 and 2-10 through thiol-Michael addition reaction	86
Scheme 2-5. Synthesis of the FAM labelled L-alanine N-hydroxymaleimide ester (2-15)	89
Scheme 2-6. Scheme of the DTS of peptide bonds in DES	106
Chapter 3. DNA templated synthesis of peptide bonds in aqueous solution: <i>in situ</i> activation of phenol esters	143
Scheme 3-1. General approach to a DNA templated <i>in situ</i> activation of a carbonyl compound (X), by a co-localised catalytic group (Y) that promotes the formation of a peptide bond	145
Scheme 3-2. Representation of templated reactions that require the co-localisation of an activating group	146
Scheme 3-3. Two proposed designs of DNA templated NCL systems showing the thioester intermediate	148
Scheme 3-4. Synthesis of a Boc protected L-Ala phenol ester	149
Scheme 3-5. Synthesis of the donor strand DNA-PhO-Ala-Boc	150
Scheme 3-6. Synthesis of an internal thiol-labelled DNA template	151
Scheme 3-7. Synthesis of a protected DNA-Cys(SS ^t Bu)-(NHFMoc) conjugate	153
Scheme 3-8. Deprotection of the DNA conjugate 3-10	154
Scheme 3-9. DNA templated NCL using a co-localised thiol catalyst (3-7), a Boc protected L-Ala phenol ester (3-5) and an amine-labelled acceptor (3-8)	156

Scheme 3-10. Retrosynthetic analysis of a maleimide phenol ester	159
Scheme 3-11. Synthesis of maleimide intermediate 3-17 followed by attempted Steglich esterification	159
Scheme 3-12. Synthesis of 3-19 utilising D-biotin acyl chloride (3-21) intermediate	161
Scheme 3-13. Thiol-Michael synthesis of a biotin phenol ester donor	161
Scheme 3-14. Optimisation of the hydrolysis conditions of 3-23	163
Scheme 3-15. First attempt of a templated NCL using a Cys-labelled DNA acceptor (3-11), a biotin phenol ester donor (3-23) and a template strand (3-25)	166
Scheme 3-16. Representation of the NCL mediated synthesis of an amide bond, and the competing N → S acyl shift, followed by an irreversible hydrolysis process	169
Scheme 3-17. Synthesis of the expected NCL product from DNA templated synthesis (3-28) with a disulfide protecting group, using conventional carbodiimide coupling	170
Scheme 3-18. Protection of the thiol after DNA templated NCL, by reaction with 3,3'-dithiodipropionic acid	174
Scheme 3-19. Three approaches to DNA templated NCL. a) Cys-labelled DNA acceptor. b) Uncatalysed peptide bond formation. c) Thiol-labelled DNA template catalyst and amine acceptor	174
Chapter 4. Peptide nucleic acids (PNAs) as potential templates for the synthesis of peptide bonds in organic solvent	199
Scheme 4-1. Representation of the SPS cycle of PNA on a rink amide solid phase surface	203
Scheme 4-2. Oxyma / DIC activation and coupling	211
Scheme 4-3. PyBOP activation and coupling	214
Scheme 4-4. Intermediates and products of the Fmoc deprotection by piperidine	215
Scheme 4-5. TFA : m-cresol PNA cleavage and Bhoc deprotection	216
Scheme 4-6. Post-synthetic modification of complementary 10-mer PNA sequences	220
Scheme 4-7. Single step PNA templated synthesis in high organic solvent solutions	239
Scheme 4-8. Toehold mediated PNA strand displacement	240
Appendix 2 Additional projects	267
Scheme A2-1. Synthesis of dithiomaleimide-labelled DNA A2-2 by reaction with methyl dibromomaleimide (MeDBM)	267
Scheme A2-2. Synthesis of adamantly-aminochloromaleimide DNA A2-4 by reaction with adamantly dichloromaleimide	269
Scheme A2-3. Comparison of the reactivity of Boc-L-Ala and D-biotin phenol esters	271

List of Tables

Chapter 1. Introduction. Templated synthesis of peptides: Reactivity, sequence encoding and programmability.	15
Table 1-1. Main features of ribosomal synthesis of peptides	24
Table 1-2. Summary of several NATS coupling reactions reported	40
Table 1-3. Summary of several NATS transfer reactions reported	41
Table 1-4. Summary of several NATS cleavage reactions reported	42
Table 1-5. Highlighted features of split and pool DNA encoded peptides	47
Table 1-6. Highlighted features of DNA routing synthesis of encoded peptides	49
Table 1-7. Highlighted features of nucleic acid templated synthesis of nucleic acid encoded peptides	51
Table 1-8. Highlighted features of DNA walker autonomous templated synthesis of peptides	53
Table 1-9. Highlighted features of the autonomous HCR templated synthesis of peptides	55
Chapter 2 DNA templated synthesis (DTS) in Deep Eutectic Solvents (DES)	69
Table 2-1. DESs used in this work	75
Table 2-2. Base additives to tune the acidity of DESs	77
Table 2-3. Solvent additives and conditions under which DNA templated peptide bond formation was attempted	108
Table 2-4. DNA oligonucleotide sequences and modifications	119
Table 2-5. Experimental conditions	136
Chapter 3. DNA templated synthesis of peptide bonds in aqueous solution: <i>in situ</i> activation of phenol esters	143
Table 3-1. DNA oligonucleotide sequences and modifications	185
Chapter 4. Peptide nucleic acids (PNAs) as potential templates for the synthesis of peptide bonds in organic solvent	199
Table 4-1. Comparison of the different techniques that can be used to study the thermal stability of dsPNA	208
Table 4-2. DNA oligonucleotide sequences and modifications	244
Table 4-3. Solid phase synthesis cycles	246
Table 4-4. Conditions, strands and T_m results of the thermal stability comparative study	248
Table 4-5. 10-mer dsPNA melting temperature results obtained from micro-DSC	249
Table 4-6. 15-mer dsPNA melting temperature results obtained from micro-DSC	251
Table 4-7. 10-mer PNA·DNA melting temperature results obtained from micro-DSC	252

Abbreviations

A	Adenine
AEEA	2-(2-(2-aminoethoxy)ethoxy)amide
APS	Ammonium persulfate
Bhoc	Benzhydryloxycarbonyl
Boc	<i>Tert</i> -butyloxycarbonyl
bp	Base pairs
C	Cytosine
Cbz	Carboxybenzyl
C_p	Molar excess heat capacity at constant pressure
C_t	C-terminus
CuAAC	Copper(I)-catalysed azide alkyne cycloaddition
DabcyI	4-((4-(dimethylamino)phenyl)azo)benzoate
Dansyl	5-(dimethylamino)naphtalene-1-sulfamide
DBM	Dibromomaleimide
DCC	<i>N,N'</i> -dicyclohexylcarbodiimide
DCE	1,2-Dichloroethane
DCM	Dichloromethane
DES	Deep eutectic solvent
DIC	<i>N,N'</i> -diisopropylcarbodiimide
DMAP	<i>N,N</i> -dimethylpyridin-4-amine
DMF	<i>N,N</i> -dimethylformamide
DMSO	Dimethylsulfoxide
DMT-MM	4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
dsDNA	Double-stranded DNA
dsPNA	Double-stranded PNA
DTPA	3,3'-Dithiopropionic acid
DTS	DNA templated synthesis
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EF-Tu	Elongation factor Tu
FAM	Fluorescein
F_i _F	Quantum yield
Fmoc	Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography
FRET	Fluorescence resonance energy transfer
G	Guanine
Gly _{0.07}	Glycine 0.07 vol% H ₂ O

HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HBTU	<i>N,N,N',N'</i> -tetramethyl-O-(1H-benzotriazol-1-yl)uronium
HCR	Hybridisation chain reaction
HFIP	1,1,1,3,3,3-hexafluoroisopropanol
HOBt	1-Hydroxybenzotriazole
HR-MS	High resolution mass spectrometry
IL	Ionic Liquid
INRS	Insulin receptor
L.G.	Leaving group
LC-MS	Liquid chromatography mass spectrometry
LNA	Locked nucleic acid
MeDCM	Methyl-dichloromaleimide
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MOPS	3-(<i>N</i> -morpholino)propanesulfonic acid
mRNA	Messenger RNA
MWCO	Molecular weight cut-off
NATS	Nucleic acid templated synthesis
NCL	Native chemical ligation
NMP	<i>N</i> -methylpyrrolidone
NMR	Nuclear magnetic resonance
nt	Nucleotides
N _t	<i>N</i> -terminus
NVOC	6-Nitroveratryloxycarbonyl
O.N.	Overnight
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDA	Photodiode array
PNA	Peptide nucleic acid
PTFE	Polytetrafluoroethylene
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
r-NCL	Reversed Native Chemical Ligation
R.T.	Room temperature
RNA	Ribonucleic acid
RP-HPLC	Reversed phase high performance liquid chromatography
RPM	Revolutions per minute
S _N 2	Bimolecular nucleophilic substitution
S _N Ac	Nucleophilic acyl substitution
S _N Ar	Nucleophilic aromatic substitution
SPS	Solid phase synthesis

ssDNA	Single-stranded DNA
ssPNA	Single-stranded PNA
SuOH	<i>N</i> -hydroxysuccinimide
T	Thymine
TAMRA	Tetramethylrhodamine
TBE	Tris borate EDTA buffer
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
T_m	Melting temperature
TNA	Threose nucleic acid
tRNA	Transfer RNA
U	Uracil
UV	Ultraviolet

Chapter 1

Introduction

Templated synthesis of peptides: Reactivity, sequence encoding and programmability.

1.1 The peptide bond and its biological relevance

1.1.1 Peptide bonds

Several groups of atoms, covalently bonded within molecules, recurrently appear under the same connectivity in natural and non-natural products; they are known collectively as functional groups. A functional group, whose role is central across chemical sciences, is the amide bond (Figure 1-1 a). In amide bonds, the nonbonding electrons on the nitrogen are delocalised into the carbonyl, generating a planar structure with two possible conformations in equilibrium.

Some of the most commonly found compounds in nature are peptides. Peptides incorporate peptide bonds, which are formed by the carbonyl of an amino acid and the subsequent amino acid amine group, linked by an amide bond (Figure 1-1 b).¹ The native structure of proteinogenic amino acids is not homogeneous, the observed differences between side chains (the R group) of each amino acid confers them a wide range of functions that make them central to natural processes.

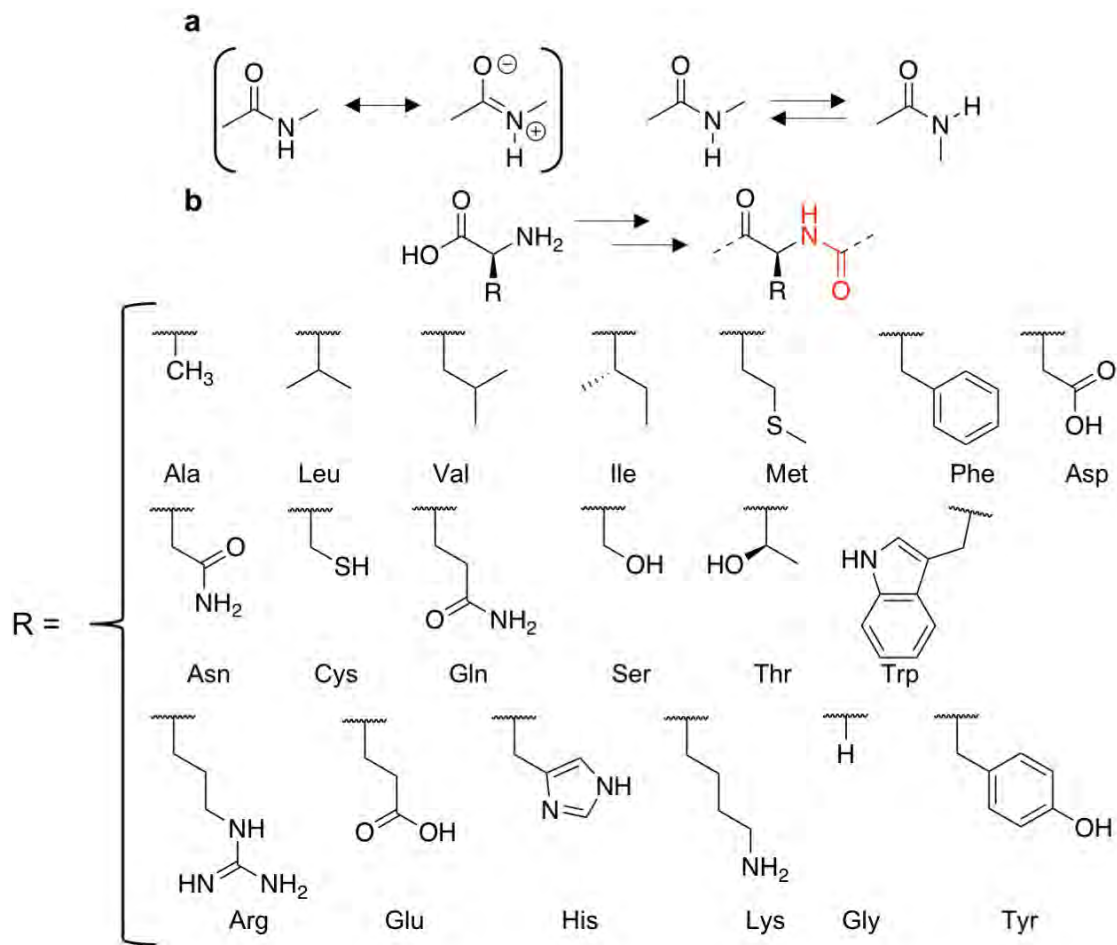


Figure 1-1. Amide functional group and peptides. a) Electronic properties of amide bonds and their interchangeable structures. b) Concatenation of amino acids gives rise to peptide bonds with a large number of functional groups, according to their side chains (R).

1.1.2 Peptides and proteins: structure and function

According to its level of complexity, the structure of peptides can be split into four levels: primary, secondary, tertiary and quaternary.¹ The primary structure describes the linear sequence of amino acids that form the peptide (Figure 1-2 a). The effect of several amino acids, giving rise to a specific primary sequence; in the peptide folding produces the formation of local repetitive conformations, that spontaneously emerge from the primary structure. The most commonly found secondary structures are the α -helix and the β -sheet (Figure 1-2 b). Beyond the amino acid sequence, and

the formation of local structure, the three-dimensional configuration of a peptide is known as the tertiary structure (Figure 1-2 c). Peptides and proteins have a similar structure, they are both formed by amino acids; however, peptides have a smaller number of amino acids, and they are less conformationally restricted. Proteins on the other hand, are large biomolecules with a well-defined tertiary structure that allows them to perform catalytic and structural functions. Finally, proteins are frequently found as assemblies of several subunits (*e.g.* haemoglobin tetramer),² this is known as the quaternary structure.

The tertiary structure of the insulin receptor (Figure 1-2 c) exemplifies the key structure and function relationship in proteins. The binding of insulin to insulin receptor (INSR) leads to a conformational change on the receptor (Figure 1-2 d).³ The conformational change initiates a signal transduction pathway that triggers in the incorporation of glucose channels on the membrane, and thus the metabolism of glucose.⁴

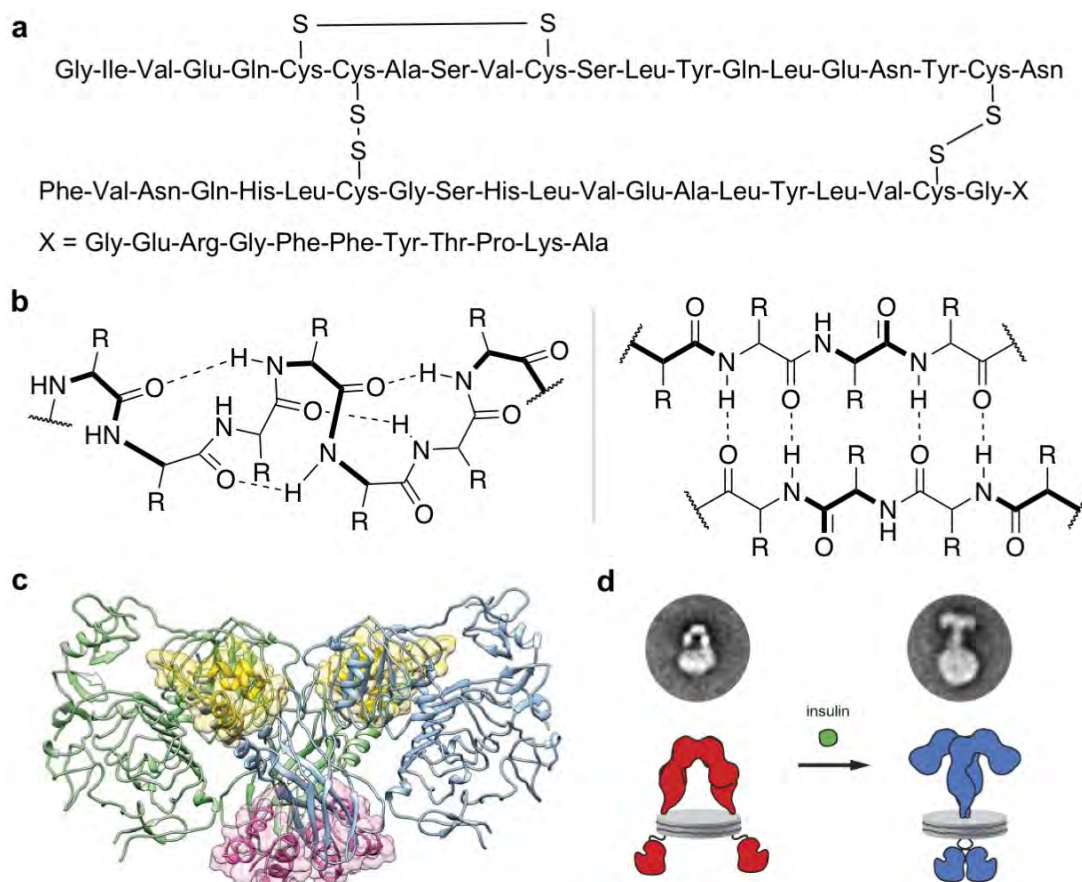


Figure 1-2. Different levels of peptide and protein structure. a) Primary structure of the 52-mer peptide insulin. b) Two examples of peptide secondary structure: α -helix (left) and β -sheet (right). c) High resolution structure of the protein INSR. α -helices are represented by a spiral motif, and β -sheets are represented by an arrow. Reproduced from previous work.⁵ d) Structural change of the INSR upon insulin binding. Reproduced from previous work.³

The insulin / INSR interaction shown in Figure 1-2 exemplifies how the study of protein and peptide structure helped to develop and understanding of biological function, and to devise a therapy in cases where the function is disrupted. For that reason, proteins have been extensively explored as drug targets.⁶ On the other hand, large combinatorial libraries of peptides have been developed in order to find molecules that interact strongly and selectively with the active centres of enzymes and receptors.⁷

1.2 Nature's approach to peptide synthesis

1.2.1 Encoding the identity of peptides: the genetic code

Proteins perform crucial functions to sustain life, for example, enzymes provide catalysis, antibodies mediate immune responses, receptors participate in signalling processes and proteins such as actin, myosin or collagen perform a structural function. There is therefore a continuous synthesis of peptide bonds in any given cell. The primary structure of the vast majority of proteins is preserved across generations (*i.e.* the proteins that perform the functions essential for a given organism to survive, will be the same in its descendants). It has been found that the storage and transcription of the primary sequences of peptides and proteins relies on natural nucleic acid polymers deoxyribonucleic acid (DNA) and ribonucleic acid (RNA, Figure 1-3 a).

DNA is a polymer formed by alternating units of deoxyribose sugar and phosphodiester. Each deoxyribose is linked to an aromatic nucleobase in the 1' position, namely adenine (A), thymine (T), cytosine (C) and guanine (G). The combination of a nucleobase, sugar and phosphate is known as a nucleotide, and the sugars and phosphates are often referred to as the backbone. The hydroxyl group at the 3' position is known as the 3' terminus, and the hydroxyl group at the 5' position is known as the 5' terminus. RNA is structurally related to DNA, with ribose sugars instead of deoxyribose. The nucleobase T is also replaced by uracil (U).

The nucleobases in DNA and RNA can interact selectively, leading to the formation of antiparallel nucleic acid duplexes, due to a balance between attractive hydrophobic, π - π and hydrogen bonding interactions, and repulsive electrostatic interactions (Figure 1,3 b).⁸ The specificity of this supramolecular interaction has been shown as crucial in biological processes and will be discussed in detail below.

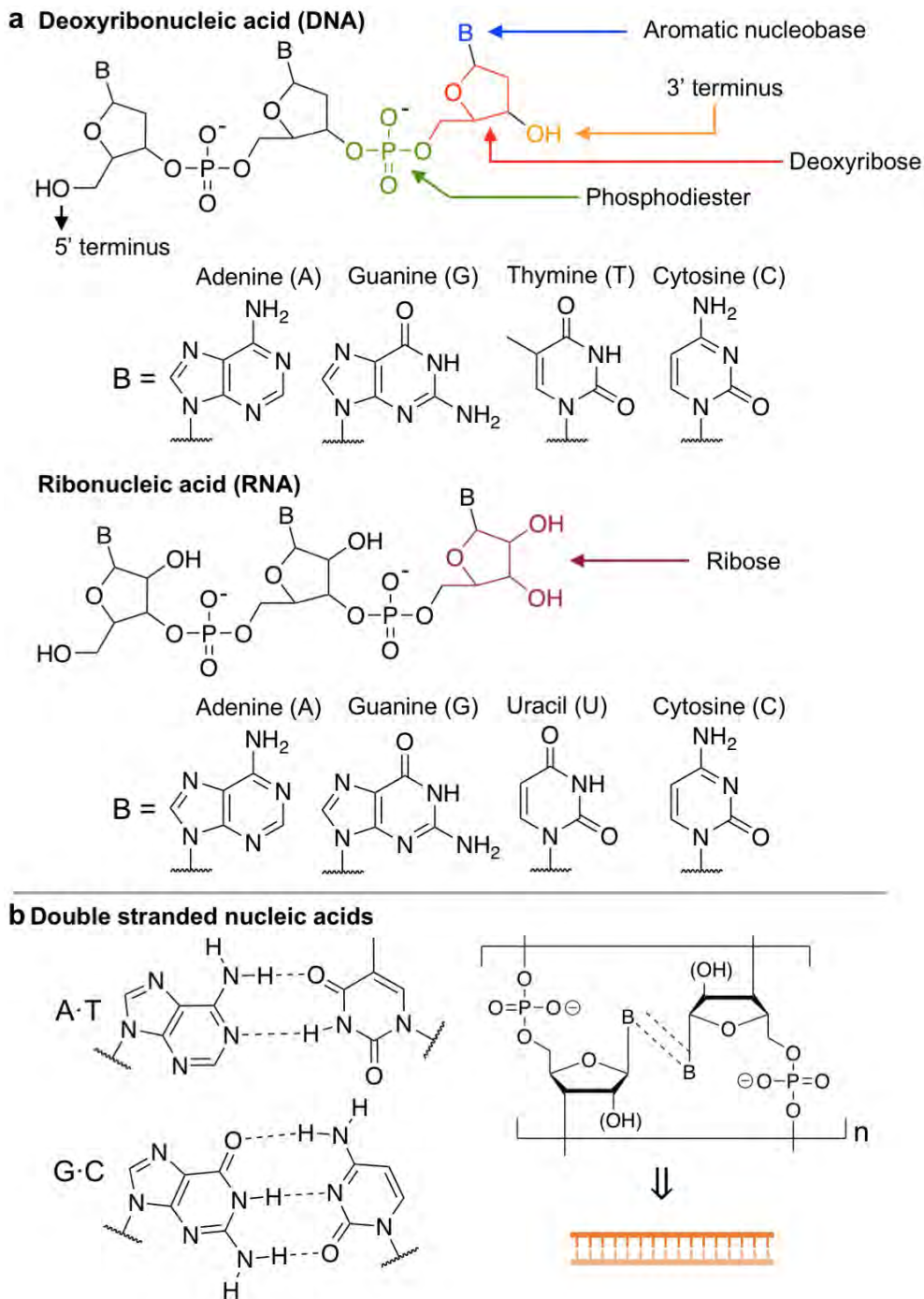


Figure 1-3. Description of nucleic acids. a) Chemical structure of deoxyribonucleic acid (DNA) and ribonucleic acid RNA. b) Watson-crick base pairs lead to the formation of double stranded nucleic acids.

The mechanism behind the preservation of protein primary sequences is based on encoding it in the form DNA genes (DNA, Figure 1-4).⁹ DNA is then transcribed into an intermediary messenger, called messenger ribonucleic acid (mRNA, Figure 1-4). Both

DNA and RNA have a modular nature, with a primary sequence of nucleobases that can be described with the letters A, T, G, C and U. The nucleobase sequence provides a means to encode the primary sequence of proteins, in a conceptually similar manner as written letters encode physical concepts. Therefore, the primary nucleobase sequence acts as a code.

Nucleobases have specific supramolecular interactions with complementary strands, for example, in the so-called Watson-Crick base pairing, A interacts with T or U, and G interacts with C (Figure 1-3 b)^{8,10}. As it was famously pointed out by the authors: *“It has not escaped our notice, that the specific pairing we have postulated immediately suggest a possible copying mechanism for the genetic material”*.⁸ In other words, the transcription of DNA into mRNA is performed according to the selective hybridisation rules; therefore, the primary sequence of the encoded protein is transcribed into the messenger strand mRNA.

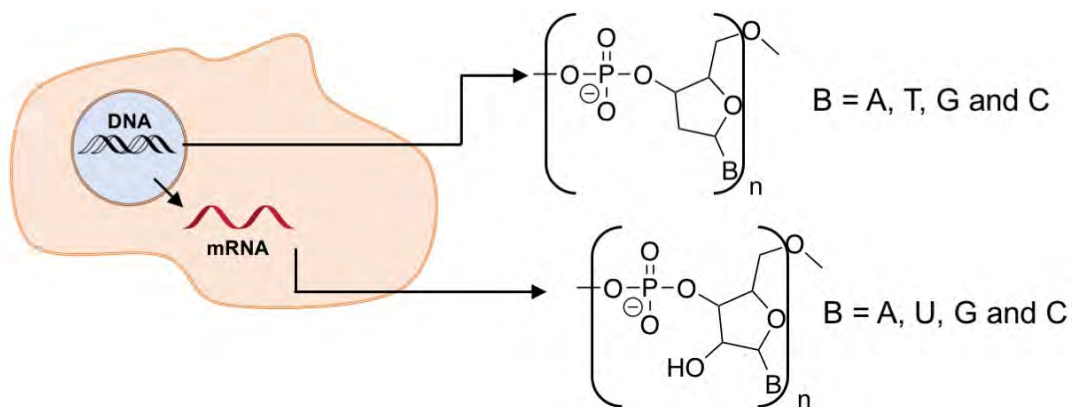


Figure 1-4. Representation of the genetic storage of information and its transcription into a primary messenger: transcription of the deoxyribonucleic acid (DNA) based genome into a ribonucleic acid (RNA) based messenger RNA (mRNA). The structural difference between DNA and RNA is found on the backbone sugar moiety. B = nucleobase.

1.2.2 From genetic information to peptides and proteins: ribosomal synthesis of peptides

The modular nature of DNA provides a means to store the primary structure of proteins, and the selective hybridisation between nucleobases provides a mechanism for the sequence-dependent transcription of DNA into a messenger intermediated strand mRNA. However, in order to produce functional peptides and proteins, the mRNA sequence has to be translated into peptide bonds (Figure 1-5 a).

The correlation between genetic sequence and protein sequence is called the genetic code.¹¹ It was found that each amino acid was encoded by a set of three mRNA nucleobases, called codons. Individual amino acids are covalently attached through a ribose ester bond to a RNA sequence called transfer RNA (^tRNA) forming an aminoacyl-^tRNA. The three-nucleobase codon in mRNA is complementary to the three nucleotides anticodon in the aminoacyl-^tRNA. Altogether, this provides a mechanism for the genetic programming of the synthesis of peptide bonds (Figure 1-5 b).

Aminoacyl-^tRNAs contain the chemical functional groups required for the synthesis of peptide bonds: a primary amine and an ester. However, the successful synthesis of peptides at a high rate requires very precise catalysis.¹² In a cellular environment, this catalysis is performed by a large ribozyme (*i.e.* a RNA catalytic biomolecule) called ribosome (Figure 1-5 c). The specific mechanism of ribosomal peptide synthesis has been extensively studied by multiple structural analysis techniques, and it has been subject to a very intense research.^{13,14} Briefly, the 2 subunits that compose the ribosome assemble around the mRNA, aminoacyl-^tRNA then binds to a specific site in the complex and incorporates the growing peptide chain through an nucleophilic acyl substitution reaction. The aminolysed ^tRNA is then eliminated and the next aminoacyl-

tRNA is incorporated into the complex for the next peptide bond synthesis step.¹ This is a very condensed description, focused only on the main elements, and there are other catalytic proteins that are required for the synthesis, such as initiation and elongation factors. Altogether, this process is known as ribosomal synthesis of proteins. The main features of ribosomal synthesis of proteins are summarised in Table 1-1.

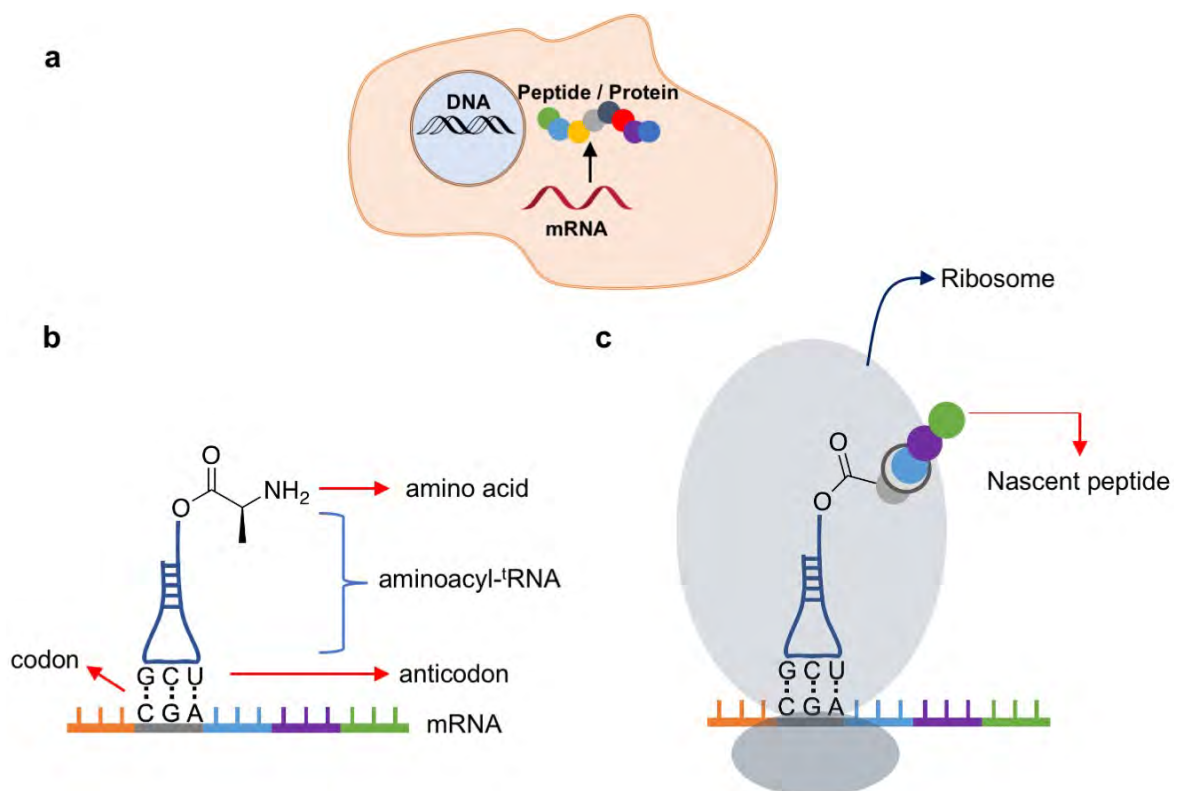


Figure 1-5. Schematic description of the fundamental aspects of the translation of genetic information into peptide sequences. a) The mRNA encodes the sequence of proteins in a series of codons. b) The codons in mRNA are complementary to the anticodons of aminoacyl-tRNAs. c) The ribosome catalyses the formation of the mRNA•aminoacyl-tRNA complex, and the transfer of the growing peptide chain to the last aminoacyl-tRNA, through the formation of a new peptide bond.

Table 1-1. Main features of ribosomal synthesis of peptides.

Feature	Mechanism
Sequence fidelity - The primary structure of proteins is preserved over time	The primary sequence of peptides and proteins is encoded in the genome DNA. In turn, this is protected inside the cell nucleus. The DNA sequence is transcribed into mRNA, which migrates into the cytoplasm. ¹
Encoded synthesis - The synthesis of proteins is performed according to the genetic information	The ribosomal translation of the genetic information into protein material is performed according to the well-defined interactions between mRNA and tRNA nucleobases in the genetic code. ¹¹
High efficiency - Long peptides and proteins synthesised rapidly	The ribosomal synthesis of peptides has been reported to display coupling rates between 12 to 20 amino acids per second. ^{15,16}
High efficiency - The aminoacyl-tRNA precursors are protected from hydrolysis	In principle, the ester bond in aminoacyl-tRNA could be hydrolysed in the highly aqueous environment of the cytoplasm. It has been demonstrated that elongation factors (e.g. EF-Tu) greatly reduce the hydrolysis rate. ¹⁷
High fidelity - Hydrolysed tRNAs do not cause deletions in the final peptide	The average error rate of ribosomal synthesis (i.e. the incorporation of an amino acid that does not correspond with the mRNA codon), has been estimated to be between 1:10 ³ and 1:10 ⁴ . ¹⁸⁻²¹ Importantly, hydrolysed tRNAs do not participate in the elongation mechanism, and thus, do not cause deletions in the peptide sequence.
Autonomous - The multistep synthesis proceeds in the absence of external stimuli	As opposed to most artificial multistep syntheses, the ribosomal synthesis of peptides proceeds in the absence of external stimuli (e.g. it does not require the addition of reagents or the removal of side products).
Reaction-dependent - The multistep synthesis proceeds upon completion of the previous peptide bond formation	The ribosomal synthesis mechanism requires that, in order to incorporate an amino acid to the nascent peptide, the previous synthesis cycle must have been completed. This dependency between the multistep synthesis and the successful completion of couplings reduces the incorporation of deletions in the final peptide. ²²

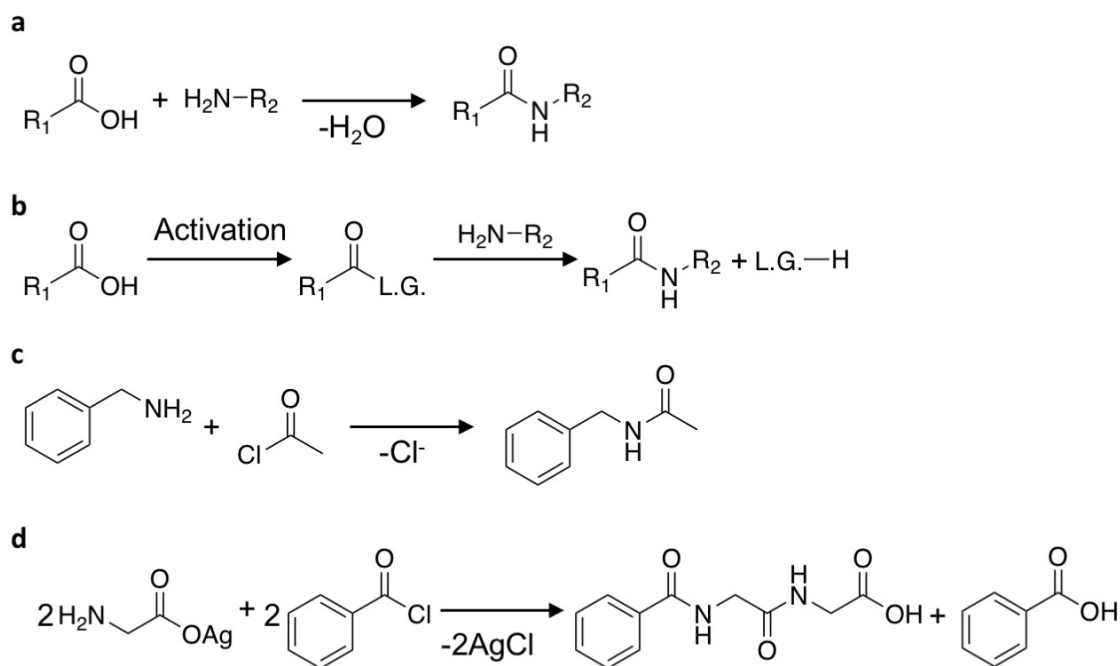
Overall, the ribosomal biosynthesis of peptide bonds can be considered a genetically encoded reaction, because the primary sequence of the peptide is encoded in the nucleic acid sequence. It is also a nucleic acid templated synthesis (NATS) as the

nucleic acid codon / anticodon supramolecular interaction, in addition to the ribosome and other macromolecules, contribute to direct the peptide bond formation.

1.3 Chemical synthesis of peptides

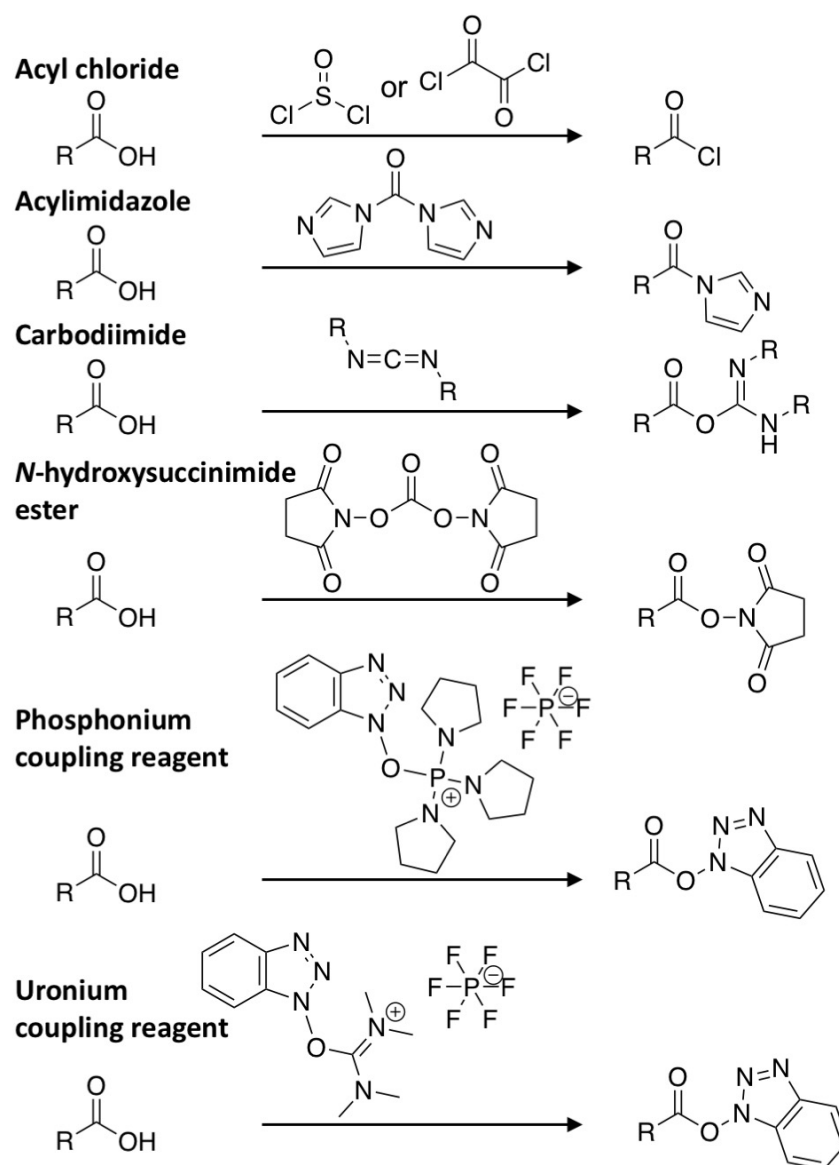
1.3.1 Solution phase synthesis

The synthesis of amides has been extensively studied as a consequence of the broad range of applications of such biomolecules. Early reports of amide coupling can be traced back to as early as the late 19th century. It was mentioned before that peptides are formally the condensation product between the carboxylic acid of an amino acids, and the amine group of another (Scheme 1-1 a). This reaction actually takes place at high temperature (180-190 °C, 12 h).²³ The thermal condensation of amino acids requires harsh conditions that are not compatible with heat sensitive products. A widely applied methodology to overcome this problem is the activation of the carboxylic acid towards nucleophilic acyl substitution (Scheme 1-1 b).²⁴ The Schotten-Baumann reaction,^{25,26} reported in 1884, consisted of the coupling between an acyl chloride and an amine to afford the corresponding *N*-benzylacetamide (Scheme 1-1 c). The coupling of two glycine amino acids was reported by T. Curtius in 1882, by reacting together a silver salt of glycine and benzoyl chloride to produce benzoylglycylglycine (Scheme 1-1, d).²⁷



Scheme 1-1. Several approaches to amide bond synthesis. a) Carboxylic acid / amine condensation.²³ b) Carboxylic acid activation followed by acyl nucleophilic substitution by a primary amine.²⁴ c) Schotten-Baumann reaction.^{25,26} d) Acylated glycyglycine synthesis.²⁷

To date, there are dozens of commercially available reagents for the activation of carboxylic acids.²⁴ Here, a non-exhaustive list, that includes some of the most widely used methodologies, has been compiled (Scheme 1-2 a-f). It includes the formation of an acyl chloride, acylimidazole, carboxydiimide activation, *N*-hydroxysuccinimide ester formation and oxybenzotriazole formation using a phosphonium and an uronium reagent.²⁴ The chemical methods to prepare amides has provided a mild and selective approach for the total synthesis of natural products,²⁸ the synthesis of bioactive compounds,²⁹ applications in materials science³⁰ and bioconjugation applications³¹ among others.



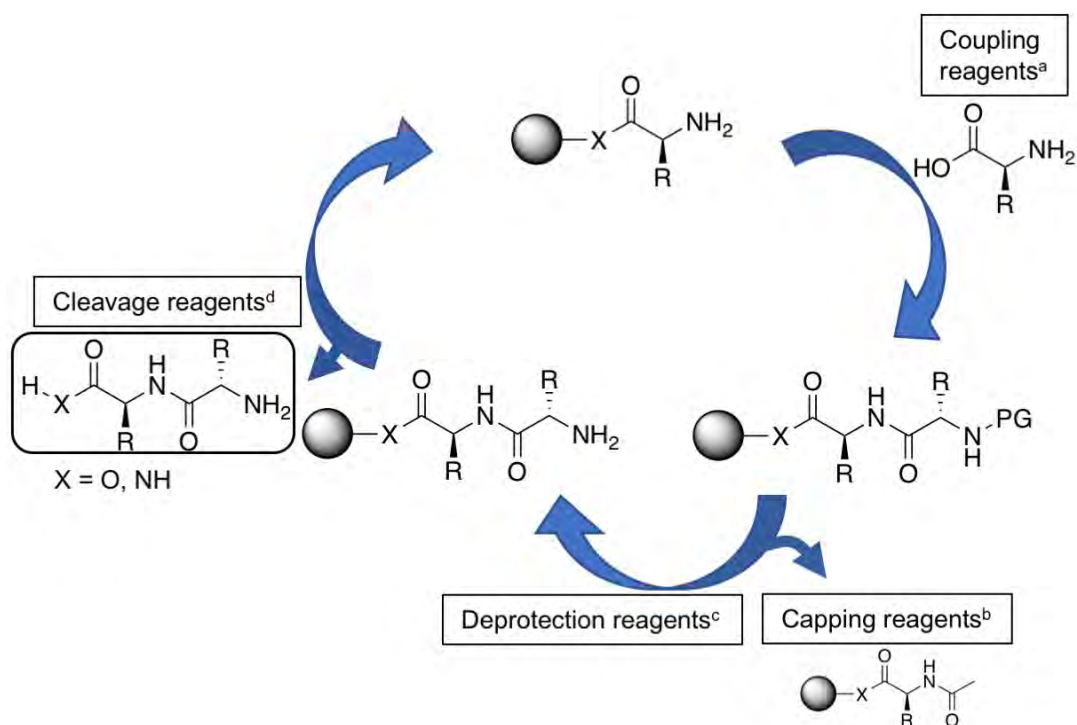
Scheme 1-2. Several methods of activation of carboxylic acids towards nucleophilic acyl substitution.

1.3.2 Solid phase synthesis of peptides

A breakthrough in the field of chemical synthesis of peptides was in the development of solid phase synthesis reported by Merrifield in 1963.³² This innovation allowed for the synthesis of the molecule of interest while it was attached to a solid support, and adding the reagents in solution; therefore, replacing the costly intermediate purification

steps by simple washing steps. This methodology has been extensively applied to the synthesis of peptides, and further integrated in automated synthesis devices.³³

The solid phase synthesis of peptides is achieved through a series of repeating cycles, according to the number of amino acids, followed by the cleavage of the product from the solid support (Scheme 1-3). The cycles are coupling, capping, deprotection, and cleavage. The coupling consists of the activation of the carbonyl on an *N*-protected amino acid. This is possible through a wide variety of strategies as mentioned above (Scheme 1-2), followed by coupling to the primary amine of the growing peptide on the solid phase surface. To prevent the addition of amino acids to truncated sequences, the unreacted primary amines are acetylated (or capped), usually with a mixture of acetic anhydride and base. To prevent undesired reactions the primary amines of activated amino acids are protected with several protecting groups, typically fluorenylmethyloxycarbonyl (Fmoc) or *tert*-butyloxycarbonyl (Boc), the deprotection reagents are piperidine and trifluoroacetic acid (TFA) respectively. The previous stages are iterated until the peptide chain, bonded to the solid phase, is completed; then, the chemical bond with the solid phase is cleaved by the addition of different reagents such as TFA or hydrofluoric acid (HF), which depend on the surface chemistry of the resin.



Scheme 1-3. General representation of the peptide solid phase synthesis stages. a) Many coupling reagents are compatible (Scheme 1-2). b) Ac_2O / base. c) For Fmoc temporary protecting group, piperidine; for Boc temporary protecting group, trifluoroacetic acid (TFA). d) For Fmoc temporary protecting group, TFA; for Boc temporary protecting group, HF.

1.4 Templated synthesis of peptide bonds

1.4.1 Nanotechnology applied to chemical synthesis

Chemical synthesis and nanotechnology have developed a symbiotic relationship where synthesis provide the means to prepare nanosized objects, which in turn can facilitate chemical synthesis.

Nanotechnology requires the precise control of physical objects at the molecular and atomic scale. Since the famous lecture entitled “There is plenty of room at the bottom”, by Feynman in 1959,³⁴ there has been extensive research into the properties of nanoscale objects, and the development of novel technical applications that nanosized objects can perform.³⁵

Thirty years after Feynman's lecture, Eigler placed several Xe atoms displaying the letters "IBM" on a nickel surface using a scanning tunnelling microscope (Figure 1-6 a).³⁶ On a larger scale, it is now possible to synthesise metallic nanoparticles with sizes ranging from 5 to 50 nm and moderate to low dispersion, using now well-established protocols (Figure 1-6 b).³⁷ Carbon based nanostructures have also been developed, showing a high degree of control on the morphology, for example the Buckminsterfullerene or C₆₀ has a nearly spherical shape,³⁸ while graphene is a planar pure carbon sheet (Figure 1-6 c-d).³⁹

One of the most interesting properties of nanosized objects is the development of specific interactions that leads to what is known as molecular recognition. A relatively early example of molecular recognition is the selectivity towards the incorporation of alkaline and alkaline earth cations into crown ethers and cryptands (Figure 1-6 e).⁴⁰ Molecular recognition in other small molecules define inner cavities upon self-assembly, behaving as molecular capsules (Figure 1-6 f).^{41,42} Molecular recognition can also be used to synthesise molecules with the ability to perform cyclic movements in a microscopic analogy of macroscopic engines, for example, a non-directional oscillation of a macrocycle between two equilibrium positions on the extremes of a rotaxane (Figure 1-6 g).⁴³

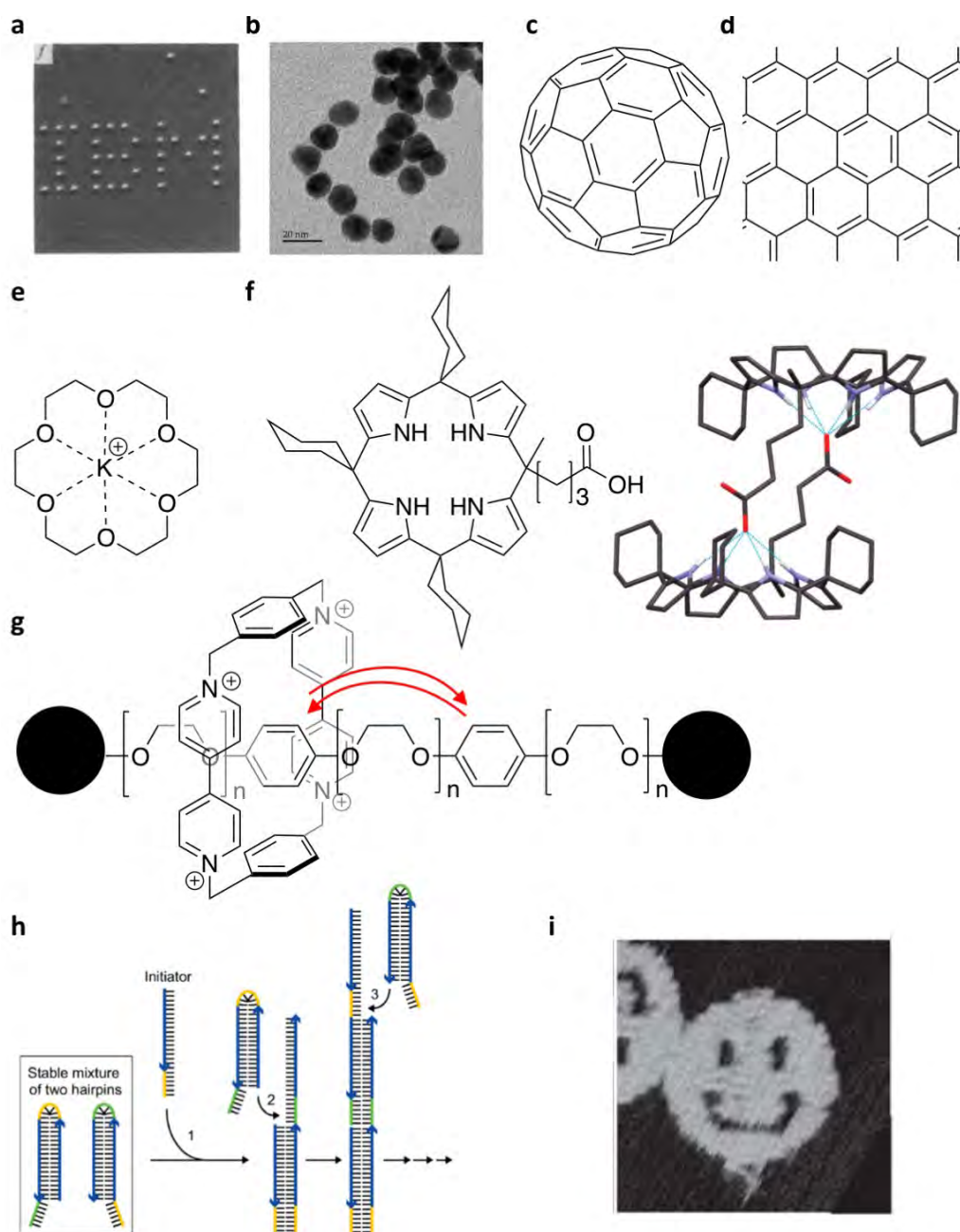


Figure 1-6. Examples of nanotechnology applications. a) Individual atoms pattern. Reproduced from previous work.³⁶ b) Gold nanoparticles. Reproduced from previous work.³⁷ c) C₆₀ molecule. d) Graphene sheet. e) 18-crown-6 complex with K⁺. f) Calix[4]pyrrole derivative and self-assembled dimer. Reproduced from previous work.⁴² g) Rotaxane molecular shuttle. h) DNA hybridisation chain reaction. Reproduced from previous work.⁴⁴ i) DNA origami “smiley face”. Reproduced from previous work.⁴⁵

All the previous examples are artificially generated nanoobjects; however, living organisms are full of nanosized functional objects such as enzymes and nucleic acids

that exquisitely exploit weak interactions at the nanoscale, for example the high affinity between a ligand and a protein, to perform defined functions. Some of these biologically originated tools have been reproduced and altered to generate new nanotechnology. For example, DNA origami and DNA computing nanotechnology have been developed based on the recognition between DNA strands, and repurposed to display the desired motifs (Figure 1-6 h-i).^{44,45}

In the aforementioned lecture, Feynman proposed the “physical synthesis” of molecules, as opposed to the chemical synthesis (which he describes as follows: “[the chemist] *mixes this and that, and he shakes it, and he fiddles around*”),³⁴ through the direct spatial rearrangement of atoms, as potentially feasible through nanotechnology. While nanotechnology has not, so far, developed general methodologies to perform “physical synthesis” of molecules, it has certainly played an increasingly evident role in directing and catalysing chemical synthesis of molecules. Chemical reactions are based on the rearrangement of valence electrons. In non-unimolecular reactions this requires the spatial co-localisation between two molecules. The encapsulation of the reagents in molecular capsules has been found to catalyse many reactions, displaying rates that would correspond to much higher concentrations for the homologous uncatalysed reaction (this is known as the high local concentration effect, and can be evaluated using the effective concentration).⁴⁶

The previous examples show that there has been a considerable development in the understanding of the nanoscale to engineer molecular processes. Chemical synthesis is one of the areas where nanotechnology has been exploited to direct and catalyse the formation of new compounds, providing tools to manipulate, with different levels of precision, molecules at the nanoscale. In subsequent sections, it will be shown how

this approach has been used to template the synthesis of peptides through a variety of approaches.

1.4.2 Peptide templated synthesis of peptide bonds

Several peptides are known to self-assemble through weak interactions.⁴⁷ Coiled-coils for example, are assemblies of 2 to 7 α -helices stabilised through hydrophobic and electrostatic interactions.⁴⁸ The formation of these assemblies has the inherent capacity of co-localising functional groups, therefore templating chemical reactions.⁴⁹ The formation of a peptide assembly can be exploited to co-localise chemically reactive groups. This has been successfully demonstrated by labelling of membrane proteins with fluorophores, through a peptide bond synthesis upon coiled-coil formation (Figure 1-7).^{50,51}

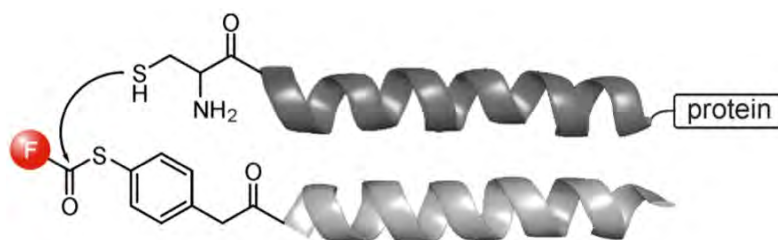
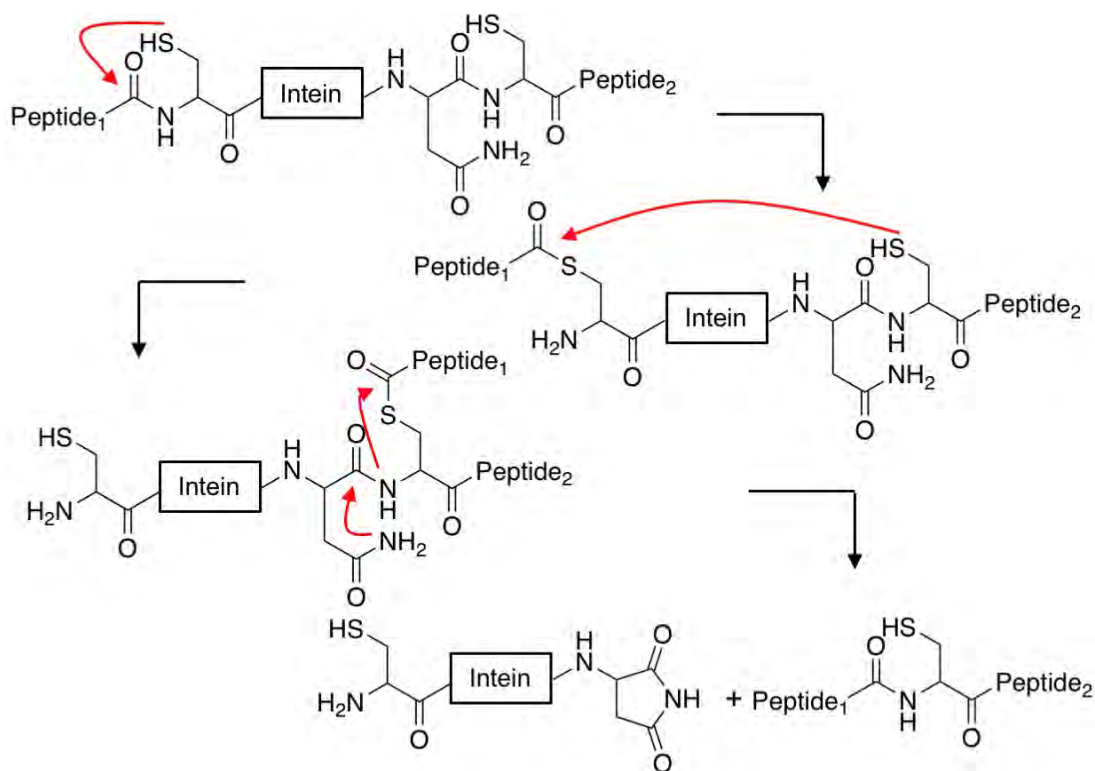


Figure 1-7. Coiled-coil templated synthesis of a peptide bond, for the fluorescence labelling of membrane proteins. Reproduced from previous work.⁵⁰

Interestingly, the fact that the assembly of peptides can trigger the formation of a peptide bond, has been exploited to develop self-replicating peptide systems.^{52,53} The formation of amyloid fibres has been shown to catalyse the chemical condensation of amino acids and the amyloid-forming peptides.⁵⁴ The formation of β -sheet and amyloid aggregates has also templated the formation of peptide bonds through chemical ligation reactions.^{55,56}

There are several examples of specific catalytic peptide bond formation carried out by proteins outside the ribosomal polypeptide. For example, isopeptide bond formation has been reported between a short peptide (SpyTag 13-mer) and the naturally occurring protein SpyCatcher.⁵⁷ This selective interaction has been further exploited to template the formation of peptide homocatenanes⁵⁸ and heterocatenanes.⁵⁹ Inteins are natural proteins with the ability to self-splice from their N-terminus and C-terminus (also known as N-extein and C-extein). In the process of self-splicing, a new peptide bond is formed between the two exteins, yielding a newly formed peptide (Scheme 1-4).⁶⁰ The intein can be considered a template, as it catalyses the formation of a specific product through a well-defined spatial co-localisation of the reagents.



Scheme 1-4. Simplified mechanism of the splicing of an intein from its N-extein and C-extein, yielding a new peptide bond.

In summary, there are several peptide and protein systems, such as coiled-coils, amyloid aggregates, α -helices or β -sheets aggregates and catalytic proteins that template the synthesis of peptide bonds. While the condensation of amino acids onto amyloid aggregates is not programmable, coiled-coil templated synthesis can be programmable, in the sense that the specificity of the interaction allows to define the identity of the product, according to the identity of the template. However, by contrast to the ribosome templated synthesis, the previous examples are limited to single-step synthesis, and there is not an individual template that can be translated multiple times such as the mRNA which it is independent from the amino acid building blocks, in coiled-coil templated synthesis, the template is consumed during the reaction.

1.4.3 Rotaxane templated peptide bond formation

Rotaxanes were briefly mentioned before (Section 1.4.1), they are synthetically made interlocked molecules, in which an elongated molecule is threaded through a macrocycle. The system is topologically stabilised by the addition of steric groups, sometimes called “stoppers”, at each end of the long molecule.⁶¹

An elegant example of a molecular machine capable of templated peptide bond formation was described by Lewandowsky *et. al.*^{62,63} In this design, the stochastic movement of the macrocyclic ring is limited by the presence of α -amino acids attached to the elongated molecule through ester bonds. The incorporation of the amino acids to the macrocyclic moiety, through chemical ligation reactions, allows the macrocycle to progress to the next step and to incorporate the subsequent amino acid, thus yielding a sequence-defined product (Figure 1-8).

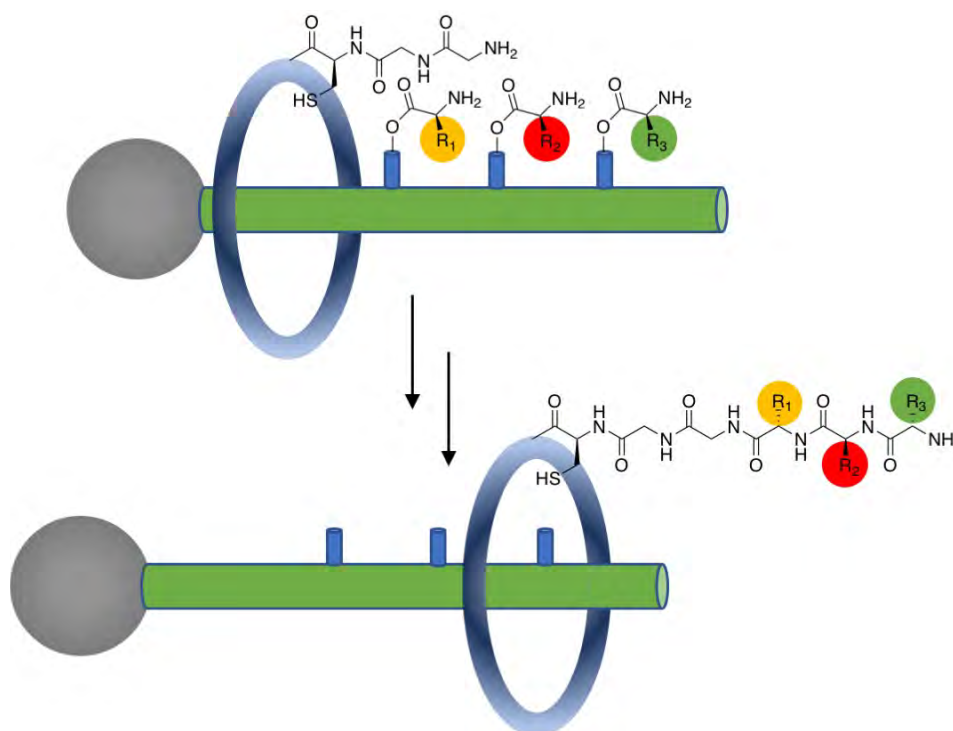


Figure 1-8. Representation of the rotaxane templated synthesis of a tripeptide.⁶²

Initially, a 3-step synthesis was performed, and the expected, sequence-defined peptide was detected by mass spectrometry.⁶² Later on, an analogous 4-step rotaxane-templated synthesis was reported with an overall isolated yield of 53%.⁶³ The topology of this assembled nanomachine produced a well-defined short peptide oligomer; however, as opposed to the ribosomal synthesis of peptides, the chemical ligation approach suffers from slow kinetics, requiring up to 48 h to complete 4 peptide bonds. Furthermore, the addition of new amino acids to the oligomer increased the distance between the nucleophile amine and the catalytic thiol, thus requiring the formation of unfavourable larger intermediate macrocyclic states according to the widely accepted native chemical ligation mechanism.⁶⁴ This would presumably increase progressively the kinetic barriers to overcome in each synthetic step, ultimately limiting the maximum length of the peptide oligomer. This is a very general

problem found in templated multistep synthesis, that can be understood as the lack of a constant reaction environment across all the synthetic steps; usually the template defines a precise reaction environment during the initial steps; however, as the oligomer is extended, the required configuration becomes less favourable.

Despite the aforementioned limitations, the rotaxane templated synthesis was applied to the formation of a β -peptide, thus showing that the rotaxane templated synthesis is more tolerant to variations on the identity of the amino acid building blocks, as opposed to the ribosome, which has a restricted range of compatible amino acid building blocks.⁶⁵ The rotaxane templated peptide bond synthesis was also reported to produce the synthesis of a poly-Leu α -helix with a distribution of lengths between 3 and 12 residues. The oligopeptide was later used as an asymmetric catalyst for the epoxidation of a chalcone.⁶⁶ This exemplified the formation of a functional material, with catalytic activity, as a result of the templated peptide synthesis.

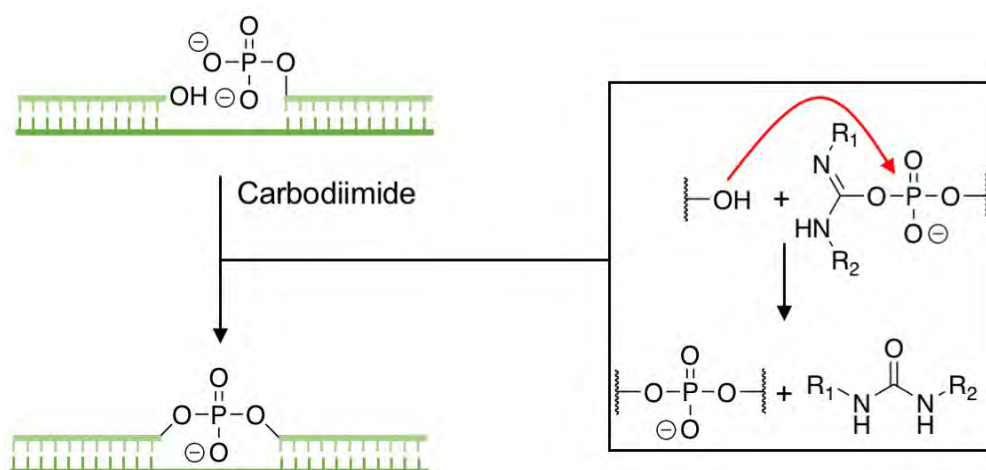
These reports are excellent examples of bottom-up nanotechnology applied to the preparation of synthetic nanomachines: conventional organic synthesis is applied to the preparation of the self-assembled nanomachine that, in a subsequent step, produce a pre-defined material. The synthesis is programmed, as the final peptide sequence is a result of the template sequence (the rotaxane axis bearing amino acyl esters). However, reprogramming the synthesis would require synthesising a new template that, in turn, will be consumed, as opposed to the mRNA in ribosomal synthesis.

1.4.4 Nucleic acid templated synthesis (NATS) of peptide bonds

Nucleic acids comprise the natural nucleobase materials DNA and RNA, and also other non-natural analogues such as locked nucleic acids LNA,⁶⁷ threose nucleic acids

(TNA)⁶⁸ or peptide nucleic acids (PNA)⁶⁹ among others. Individual nucleic acid strands can hybridise, therefore, they share with molecular capsules, self-assembling peptides and rotaxanes, the ability to spatially co-localise chemical groups, and therefore to template the synthesis of peptide bonds, if the reactive functional groups are attached to complementary strands, this is known as nucleic acid templated synthesis (NATS).⁷⁰ While the untemplated reaction would still be possible, NATS is usually performed at very low concentration which results in extremely slow kinetics. It is the annealing of complementary nucleic acids that accelerates the reaction through a high local concentration effect.⁷¹

The first example of NATS was nonenzymatic ligation of two DNA strands annealed to their complementary template (Scheme 1-5).⁷² In this case, the phosphoester group was activated by reaction with a carbodiimide, the activated species then reacted with the adjacent hydroxyl group producing the ligated double-stranded DNA (dsDNA).



Scheme 1-5. Representation of the carbodiimide mediated ligation of a poly-A and poly-T nicked double stranded DNA (dsDNA).⁷²

The previous example of NATS was performed on a “nicked” dsDNA, the templated reaction in this architecture is often described as across-the-nick. Later on, many more

hybridisation architectures have been utilised with NATS (Figure 1-9). The end-of-helix architecture consists of the localisation of the reactive groups at the complementary 5' and 3' terminus.⁷³ In the T architecture, one of the modifications is attached to a nucleobase instead of the strand terminus,⁷⁴ it has the advantage of leaving a non-hybridised overhang that can be used to trigger more complex nucleic acid displacement steps. The Ω -architecture also leaves a non-hybridised nucleic acid loop region, with the reactive groups attached to complementary 3' and 5' terminus.⁷⁴ Minor-groove coupling was recently described, with the distinct feature that both reactive groups can be attached to 3' or 5' terminus, therefore displaying the same directionality. Furthermore, it also allows a very close spatial co-localisation of the reactive moieties.⁷⁵ Finally, the yoctoreactor design makes use of a three-way junction to template chemical reactions through internal modification of nucleic acid strands.⁷⁶

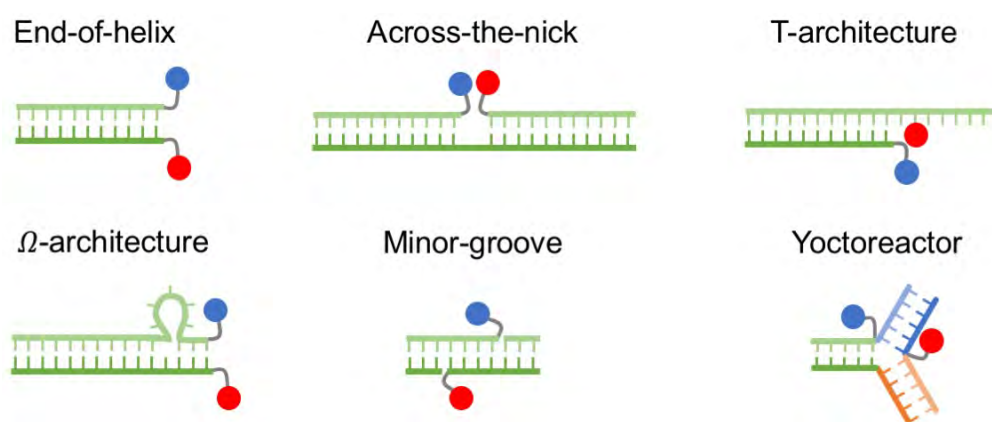


Figure 1-9. Representation of the enabling architectures for NATS: 1) End-of-helix;⁷³ 2) Across-the-nick;⁷⁴ 3) T-architecture; 4) Ω -architecture;⁷⁴ 5) Minor groove;⁷⁵ 6) Yoctoreactor.⁷⁶

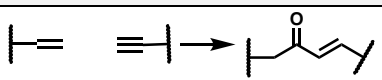
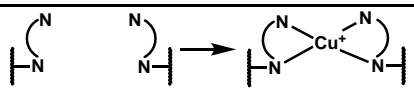
The previous architectures have been implemented in a range of nucleic acid templated reactions. Reactions can be divided into coupling, transfer and cleavage chemistries, depending on the connectivity between the nucleic acid strands after the

NATS. Here, a non-exhaustive summary of NATS compatible reactions has been composed to illustrate the versatility of this approach to chemical synthesis.

Coupling NATS results in the two strands connected through a covalent bond (Table 1-2). This has been applied in detection⁷⁷ and programmed synthesis⁷⁸ among others.

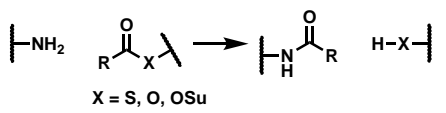
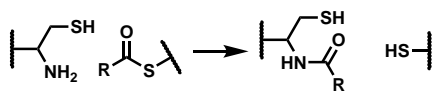
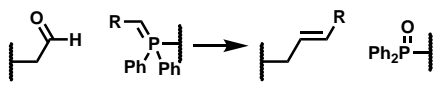
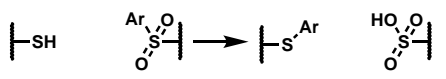
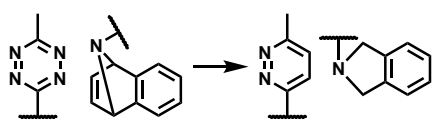
Table 1-2. Non-exhaustive summary of NATS coupling reactions reported. Each reagent is attached to a DNA oligonucleotide. A covalent bond is formed between the DNA reagent strands

Reaction	Structures	Architecture	Ref.
S _N 2		Across-the-nick	79
Conjugate addition		End-of-helix	79
Reductive Amination		End-of-helix Ω -architecture	74,80
Oxazolidine formation		End-of-helix Ω -architecture	74,80,81
Nitro-aldol		End-of-helix	80
Thiol-Michael addition		End-of-helix Across-the-nick	79
Nitro-Michael addition		End-of-helix	80
Wittig olefination		End-of-helix Ω -architecture	74,80
Amide bond formation		End-of-helix Ω -architecture	74,80,82
Native Chemical Ligation		Across-the-nick	83,84
Selenocysteine selenoester ligation		Across-the-nick End-of-helix	77
CuAAC		Across-the-nick End-of-helix	85-87

Reaction	Structures	Architecture	Ref.
Pd-cross coupling		End-of-helix	88
Metal coordination		End-of-helix	89

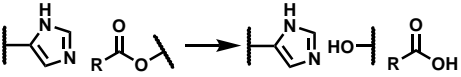
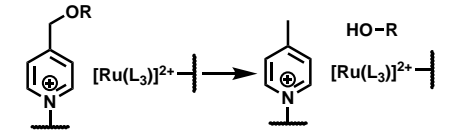
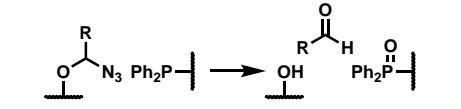
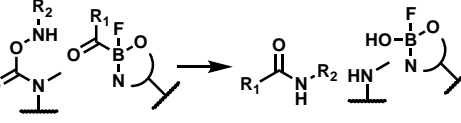
In transfer chemistries, one of the reactive groups attached to the strand, is transferred across to the complementary strand (Table 1-3). It is interesting to note that the same chemistry, for example peptide bond formation, can result in a coupling or a transfer reaction, depending on whether the attachment of the ester to DNA is through the leaving group (transfer) or through a functional group that does not participate in the templated reaction (coupling). Transfer chemistries have been applied to detection, for example through the change in fluorescence upon transfer,⁹⁰ and for the programmed synthesis of oligomers.⁹¹

Table 1-3. Summary of several NATS transfer reactions reported. A small chemical group is transferred between DNA oligonucleotides, without generating a covalent bond between them.

Reaction	Structures	Architecture	Ref.
Acyl transfer		End-of-helix Across-the-nick Minor-groove	75,92-94
Native chemical ligation		Across-the-nick	95,96
Wittig olefination		End-of-helix Across-the-nick	97,98
S _N Ar		Across-the-nick	99
Tetrazine transfer		Across-the-nick	100

Finally, cleavage NATS selectively releases one or two of the groups attached to the nucleic acids (Table 1-4). Cleavage NATS has previously been applied to the selective release of an ester, upon hybridisation of complementary strands,¹⁰¹ and for the release of a small-molecule after reduction (e.g. pyridinium photoreduction or Staudinger reduction).

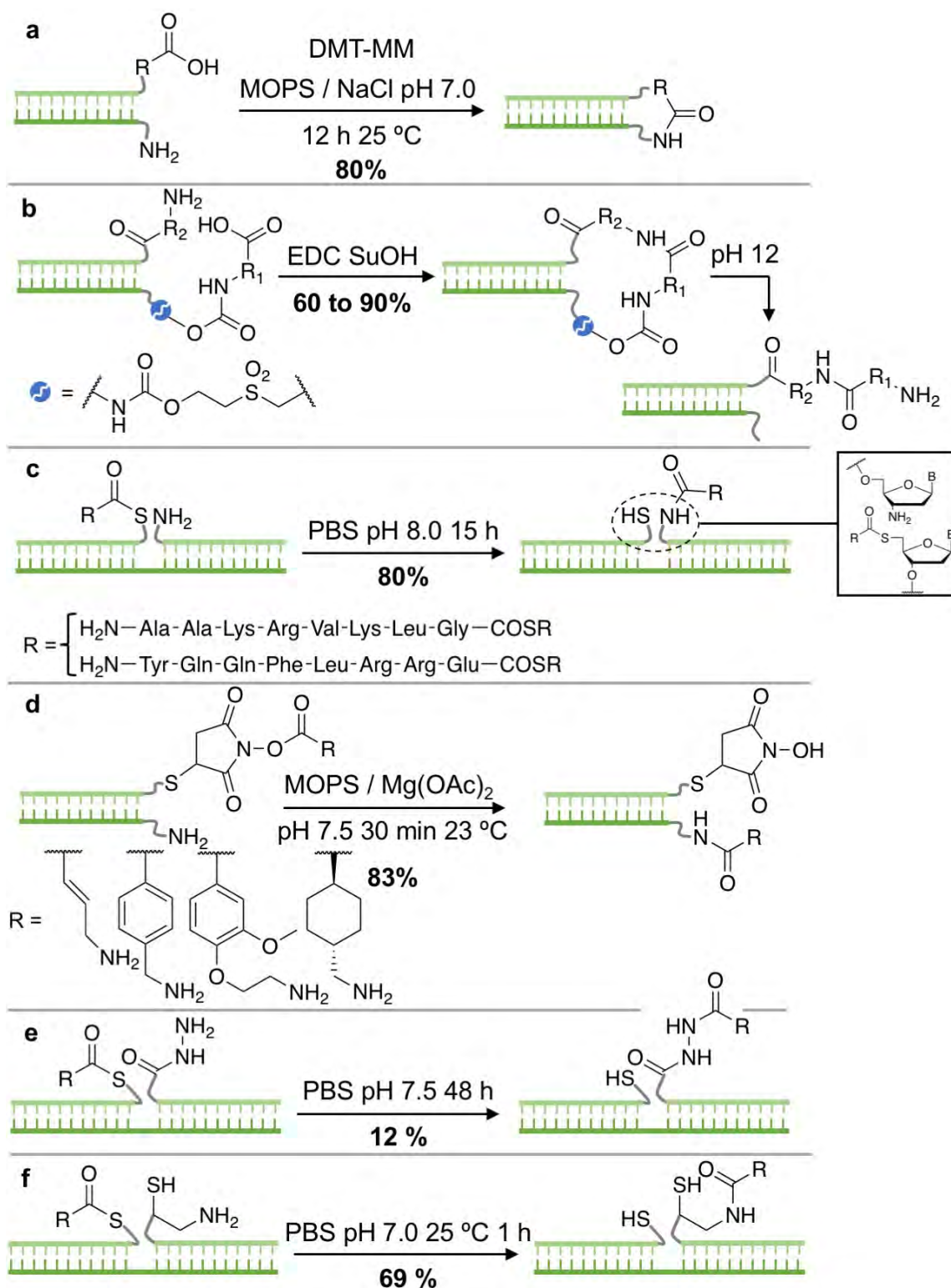
Table 1-4. Summary of several NATS cleavage reactions reported. *Acylboronate hydroxylamine ligation was performed using a tetrameric streptavidin template instead of nucleic acids.

Reaction	Structures	Architecture	Ref.
Ester hydrolysis		Across-the-nick	
Pyridinium photoreduction		End-of-helix Across-the-nick	102,103
Staudinger reduction		Across-the-nick	104
Acylboronate-hydroxylamine ligation		*Tetrameric streptavidin	105

Overall, there is a wide range of chemistries that are compatible with nucleic acid templates. These have been applied to both the detection of nucleic acids, and the templated synthesis of compounds of interest.

Focusing on templated peptide bond synthesis, this has been demonstrated through coupling and transfer NATS (Scheme 1-6 a,c).⁷⁴ Multistep NATS of peptide bonds through coupling reactions requires the addition of cleavage steps (Scheme 1-6 b).⁸² Interestingly, one of the early examples of NATS of peptide bonds, using thioester aminolysis, displayed high conversion, up to 80 % (Scheme 1-6 c).⁹² NATS was performed in the absence of any linker; therefore, the co-localisation of the ribose

thioester and the acceptor amine was in optimal space proximity. Subsequent examples of NATS of peptide bonds have used highly activated DNA conjugated N-hydroxysuccinimide esters, these provide high NATS yield, as long as the reaction is performed quickly (Scheme 1-6 d).⁹⁴ However, the DNA conjugated activated esters display a high hydrolysis rate, and thus, cannot remain in solution for a long time.¹⁰⁶ Peptidomimetic NATS systems, with flexible linkers between nucleic acid strands and the reacting groups have been used, showing a moderate increase in peptidomimetic bond synthesis (12% after 48 h, hydrazide acceptor, Scheme 1-6 e) compared to peptide bond synthesis (3% after 48 h, amine acceptor).⁹³ High peptide bond NATS yield was achieved (~70%) through *iso*-cysteine native chemical ligation (Scheme 1-6 f).⁹⁶



Scheme 1-6. Schematic description of several nucleic acid templated peptide bond formation. a) Reagent mediated coupling.⁷⁴ b) Reagent mediated coupling followed by cleavage.^{82,107} c) Thioester mediated transfer.⁹² d) *N*-hydroxysuccinimide ester mediated transfer.⁹⁴ e) Thioester / hydrazide mediated peptidomimetic transfer.⁹³ f) Native chemical ligation mediated transfer.⁹⁶

1.4.5 The potential of NATS of peptide bonds: from encoding the identity of peptides to artificial ribosomes.

The previous examples of NATS of peptides shown in Scheme 1-6, focused on single-step templated synthesis; however, nucleic acids have the potential to mediate multistep templated synthesis of peptide oligomers through the addition of toehold mediated nucleic acid strand displacement steps.¹⁰⁸ Here, some of the strategies previously applied to multistep NATS of peptides will be briefly described and compared to ribosomal peptide synthesis.

One of the applications of multistep NATS of oligopeptides is the preparation of encoded combinatorial libraries, where the primary sequence of the peptide is encoded on the nucleic acid sequence.¹⁰⁹ The split and pool approach to generate encoded peptide libraries (Figure 1-10) alternated rounds of peptide solid phase synthesis with oligonucleotide solid phase synthesis. This method was successfully applied in the identification of a highly specific antibody ligand, which was identified through the DNA encoding sequence.¹¹⁰

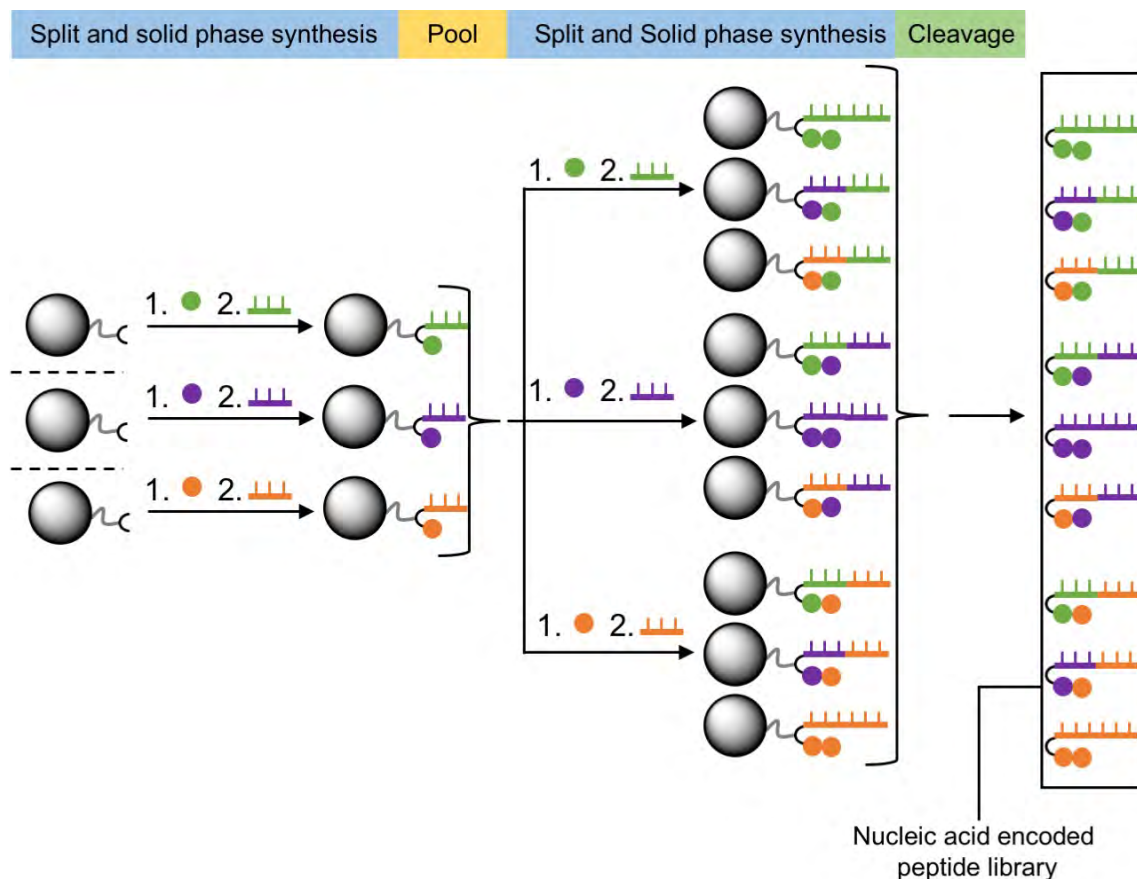


Figure 1-10. Schematic description of the split and pool generation of DNA encoded peptide libraries. Grey spheres represent solid phase beads.¹⁰⁹

This approach cannot be considered templated, as the function of the oligonucleotide sequence was to encode the identity of the peptide, and not to direct in any way the synthesis itself. Similar to ribosomal synthesis, the split and pool approach produces a peptide sequentially related to an oligonucleotide. However, it differs in the fact that the sequence of the oligonucleotide does not direct the synthesis (Table 1-5).

Table 1-5. Highlighted features of split and pool DNA encoded peptides.¹⁰⁹

Feature	Split and pool	Ribosomal synthesis
Peptide sequence encoded in the nucleic acid sequence?	Yes	Yes
Genetic code and peptide covalently bonded?	Yes	No
Nucleic acid hybridisation directs the synthesis?	No	Yes
Nucleic acid hybridisation during the peptide bond formation?	No	Yes
Nucleic acid template can be reused?	No	Yes
One pot synthesis of multiple products?	No	Yes
Autonomous multistep synthesis?	No	Yes
If autonomous: The progress of the multistep synthesis is conditional to the previous peptide bond formation?	-	Yes

DNA routing approach took advantage of the selective hybridisation of DNA to selectively separate individual members of a DNA combinatorial library. Then, each isolated oligonucleotide was functionalised with the appropriate amino acid encoded in the DNA sequence (Figure 1-11).¹¹¹

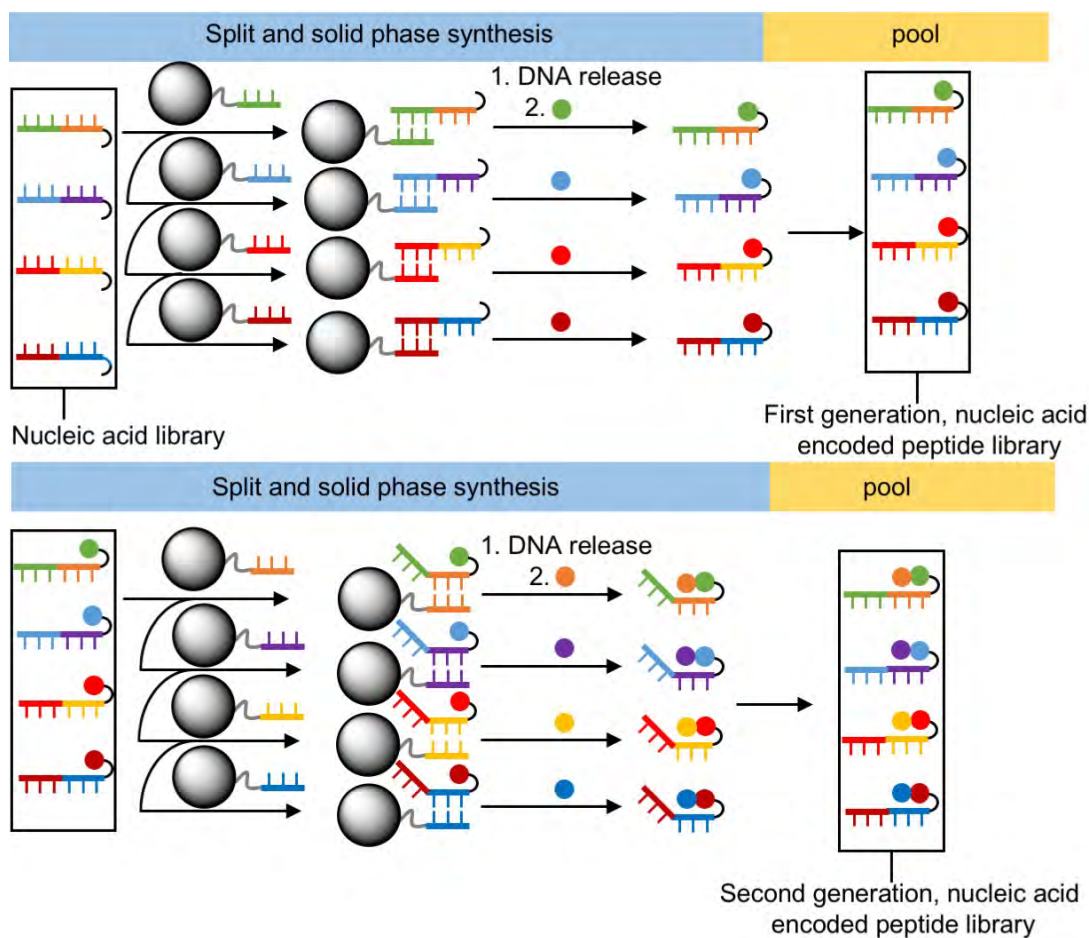


Figure 1-11. Schematic description of DNA routing.¹¹¹

In DNA routing, the conditions required for peptide bond formation were not compatible with the hybridisation of DNA, and therefore the selective hybridisation between complementary strands was exploited to isolate each individual DNA sequence. However, the functionalisation with the corresponding amino acid was not a templated reaction, as opposed to ribosomal synthesis, where the mRNA / tRNA interaction directs the peptide bond formation (Table 1-6).

Table 1-6. Highlighted features of DNA routing synthesis of DNA encoded peptides.¹¹¹

Feature	DNA routing	Ribosomal synthesis
Peptide sequence encoded in the nucleic acid sequence?	Yes	Yes
Genetic code and peptide covalently bonded?	Yes	No
Nucleic acid hybridisation directs the synthesis?	Yes	Yes
Nucleic acid hybridisation during the peptide bond formation?	No	Yes
Nucleic acid template can be reused?	No	Yes
One pot synthesis of multiple products?	No	Yes
Autonomous multistep synthesis?	No	Yes
If autonomous: The progress of the multistep synthesis is conditional to the previous peptide bond formation?	-	Yes

Multistep DNA templated synthesis of peptide bonds was performed due to sequential NATS, followed by strand displacement (Figure 1-12).⁹⁴ In this case, the nucleic acid hybridisation was used to encode the identity of the peptide product, and to direct the synthesis of the peptide bond by co-localising the activated ester donor and the amine acceptor. The highly reactive *N*-hydroxysuccinimide ester adapters displayed fast reaction kinetics only requiring 30 min incubation. In addition, the fact that activated ester adapters were added sequentially potentially contributed to minimising the time that they remained in solution, and thus limited hydrolysis.

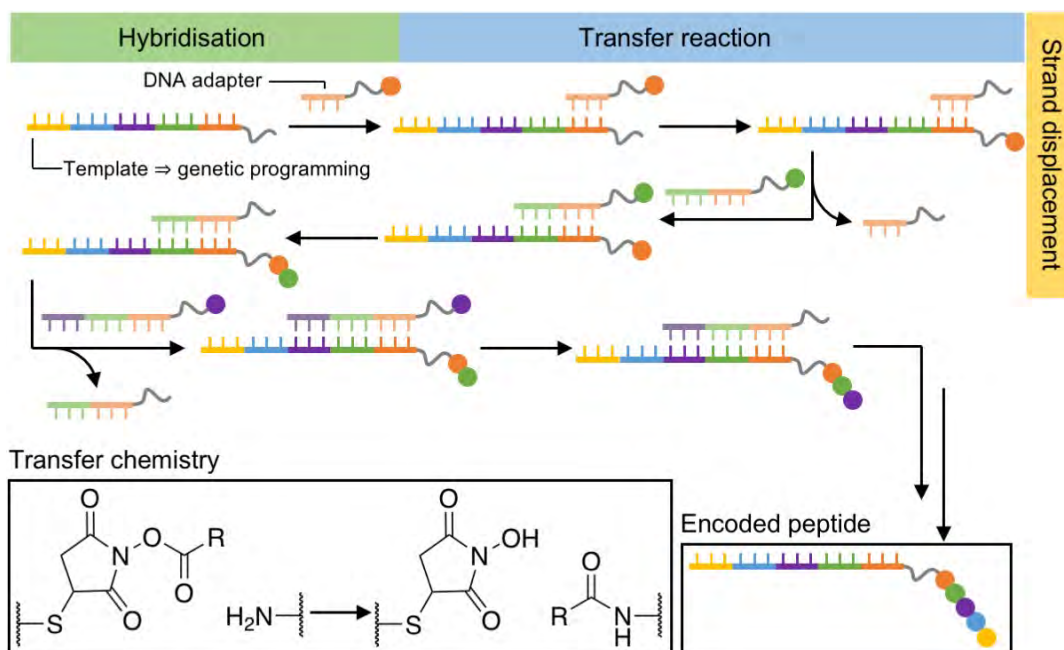


Figure 1-12. Schematic representation of multistep nucleic acid templated synthesis of peptide bonds through DNA templated synthesis.⁹⁴

In multistep NATS of peptide bonds, the sequence encoded the identity of the amino acids, and also facilitated the chemical reaction providing high local concentration conditions (Table 1-7). A major difference with ribosomal synthesis is that it required the sequential addition of the different adapters in order to produce the encoded peptide.

Table 1-7. Highlighted features of nucleic acid templated synthesis of nucleic acid encoded peptides. *One pot synthesis of multiple products would require the addition of positional information to each adapter.⁹⁴

Feature	DNA templated synthesis	Ribosomal synthesis
Peptide sequence encoded in the nucleic acid sequence?	Yes	Yes
Genetic code and peptide covalently bonded?	Yes	No
Nucleic acid hybridisation directs the synthesis?	Yes	Yes
Nucleic acid hybridisation during the peptide bond formation?	Yes	Yes
Nucleic acid template can be reused?	Yes	Yes
One pot synthesis of multiple products?	Yes*	Yes
Autonomous multistep synthesis?	No	Yes
If autonomous: The progress of the multistep synthesis is conditional to the previous peptide bond formation?	-	Yes

In 2010, the first autonomous multistep NATS of peptides was reported by He and Liu.¹¹² In this case, a DNA molecular walker was used to perform the templated reaction at each step of the “walk”. The acceptor strand annealed to a nearby DNA adapter bearing an activated ester. This configuration allowed for the templated peptide bond formation to take place. After dsDNA hybridisation, a restriction site was enzymatically cleaved, thus generating a metastable assembly, allowing the acceptor strand to progress to the next step (Figure 1-13). This successfully demonstrated how an artificial DNA assembly could be used to perform multistep programmed synthesis of oligomers. However, the maximum length of the product was a 3-mer. Similarly to the system in Figure 1-11, *N*-hydroxysuccinimide activated ester adapters were used.

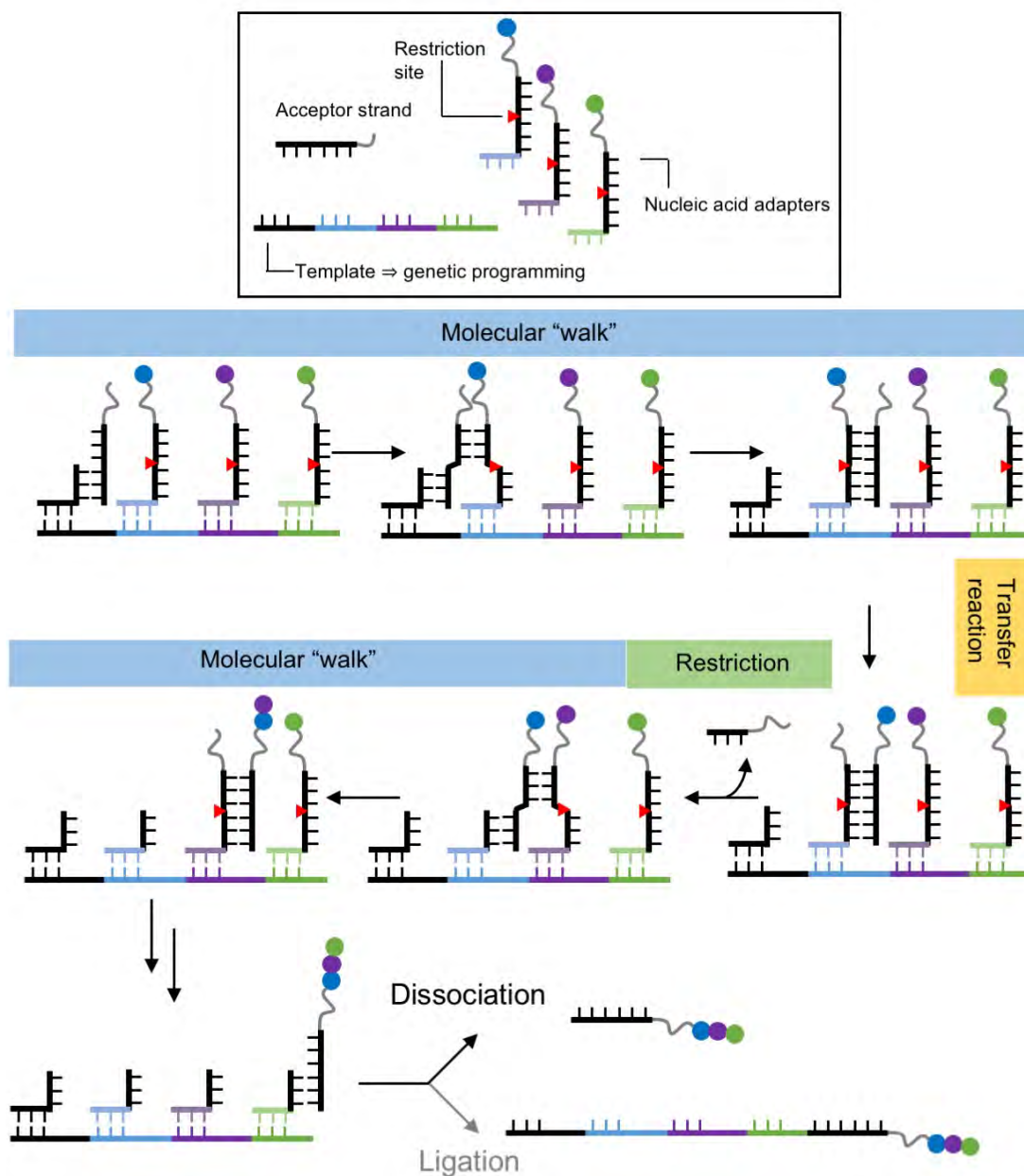


Figure 1-13. Autonomous DNA walker templated synthesis of peptides.¹¹²

In this example of autonomous multistep NATS, there is a nucleic acid motor, that performs a series of steps, and there are a series of chemical reactions that take place. By contrast to ribosomal synthesis, the progress of the molecular motor is not conditional to a successful bond formation; therefore, if the kinetics of the reaction were slower than the kinetics of the DNA motor, the final oligomer would not contain all the

expected amino acids (Table 1-8). Interestingly, in this system the encoding template and the adapters are individual entities analogous to the mRNA and tRNA in ribosomal synthesis. Therefore, the same pool of reagents could be programmed to synthesise a different product by simple addition of a different template.

Table 1-8. Highlighted features of DNA walker autonomous templated synthesis of peptides. *It would be possible through the incorporation of a DNA ligation (Figure 1-12, greyed section).¹¹²

Feature	DNA templated synthesis	Ribosomal synthesis
Peptide sequence encoded in the nucleic acid sequence?	Yes	Yes
Genetic code and peptide covalently bonded?	No*	No
Nucleic acid hybridisation directs the synthesis?	Yes	Yes
Nucleic acid hybridisation during the peptide bond formation?	Yes	Yes
Nucleic acid template can be reused?	Yes	Yes
One pot synthesis of multiple products?	Yes	Yes
Autonomous multistep synthesis?	Yes	Yes
If autonomous: The progress of the multistep synthesis is conditional to the previous peptide bond formation?	No	Yes

In 2016, Meng *et. al.* reported a new autonomous nucleic acid motor capable of programmable templated synthesis of polyolefins and polypeptides.¹¹³ In this case, a previously described hybridisation chain reaction (HCR)⁴⁴ was modified to allow the deterministic synthesis of oligomers. HCR is based on the fact that, due to their strained conformation, the nucleobases in a DNA hairpin loop do not form stable duplexes with complementary strands. However, upon hairpin opening with a single stranded DNA (ssDNA) initiator, the hairpin loop is exposed which allows the opening of another hairpin and the process is repeated iteratively following an unbranched chain reaction scheme. In the HCR NATS, the initiator was annealed to an acceptor

strand modified with an amine functionality, and the hairpins were divided into instruction hairpins and chemistry hairpins (Figure 1-14). Instruction hairpins contains ssDNA overhangs capable of recognising the previous synthetic step, and the hairpin loop determines the next synthetic step. Chemistry hairpins were designed to recognise the previous instruction hairpin and were modified with an activated ester. Upon mixing, the initiator / acceptor duplex forms a Holliday junction¹¹⁴ with the first instruction hairpin progressing into a duplex, exposing the hairpin loop; then, a chemistry hairpin recognising the loop, is incorporated into the dsDNA, and the activated ester is transferred to the acceptor strand. This process was iteratively repeated according to the set of instruction hairpins.

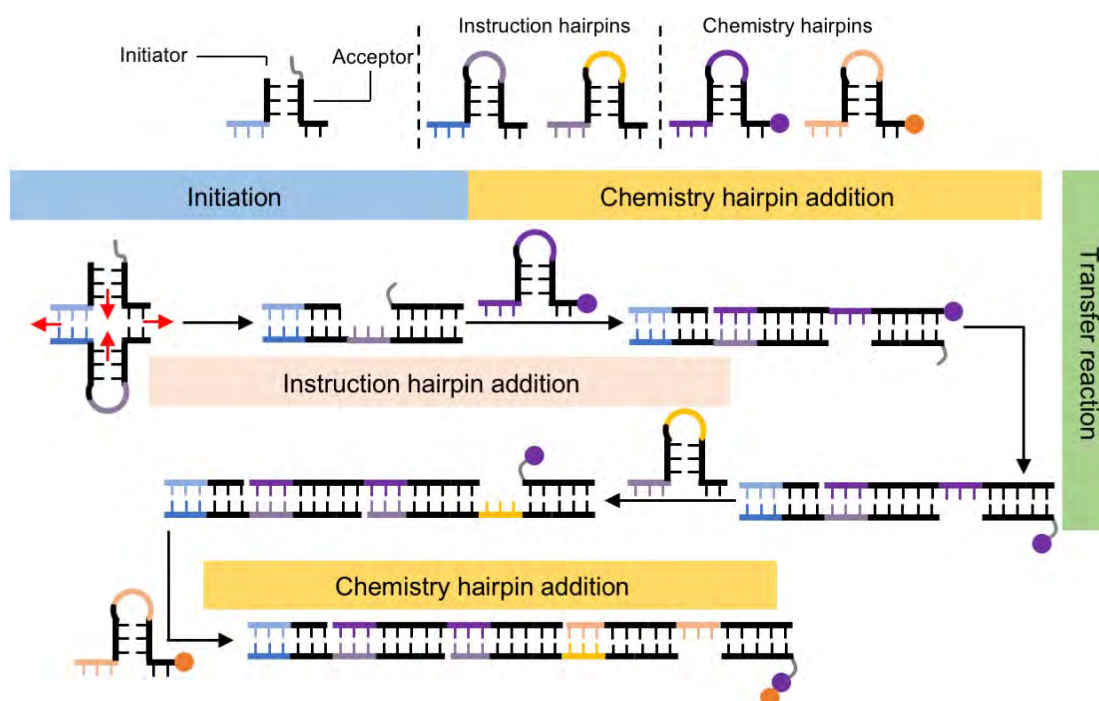


Figure 1-14. Autonomous hybridisation chain reaction (HCR) templated synthesis of peptides.¹¹³

Similarly to the DNA walker system, the autonomous HCR NATS of peptides allows a one pot multistep synthesis, with all the necessary components in solution from the beginning of the reaction. The genetic program that determines the outcome of the

synthesis (*i.e.* the set of chemistry hairpins) is decoupled from the nucleic acid adapters, similarly to the mRNA and the tRNA. A deterministic combinatorial synthesis can thus be performed by simply changing the set of instruction hairpins (Table 1-9). The progress of the DNA motor was again not conditioned by the successful performance of synthetic steps; therefore, slow reaction kinetics would introduce errors in the correspondence program / oligomer sequence.

Table 1-9. Highlighted features of the autonomous HCR templated synthesis of peptides.¹¹³

Feature	Hybridisation chain reaction	Ribosomal synthesis
Peptide sequence encoded in the nucleic acid sequence?	Yes	Yes
Genetic code and peptide covalently bonded?	Yes	No
Nucleic acid hybridisation directs the synthesis?	Yes	Yes
Nucleic acid hybridisation during the peptide bond formation?	Yes	Yes
Nucleic acid template can be reused?	Yes	Yes
One pot synthesis of multiple products?	Yes	Yes
Autonomous multistep synthesis?	Yes	Yes
If autonomous: The progress of the multistep synthesis is conditional to the previous peptide bond formation?	No	Yes

The HCR NATS was successfully applied to the synthesis of oligopeptides, showing a maximum of 7 consecutive couplings in low yield.¹¹³

The previous two examples of autonomous multistep NATS of peptide bonds highlight the potential that nucleic acids hold to program and catalyse the synthesis of peptides, emulating to a certain point, the ribosomal synthesis of peptides. However, while the ribosome is capable of catalysing the incorporation of a wide range of α -amino acids at high rate, these artificial ribosomes, have only been compatible with a very limited

range of unnatural amino acid, at least partially due to their resistance to rapid hydrolysis in aqueous buffered solution (Figure 1-15 a). Autonomous multistep peptide synthesis requires the dissolution of intrinsically reactive activated esters at the beginning of the process, and the activated esters remain in solution throughout the process, which makes them prone to hydrolysis. This has limited the potential of autonomous NATS to produce natural peptides, which could perform functions such as enzyme inhibition or catalysis. Some of the amino acids widely found in interesting natural peptides have a low steric hindrance in the vicinity of the carbonyl, which makes them prone to hydrolysis (Figure 1-15 b). This is exemplified in the therapeutic peptides *Abarelix*,¹¹⁵ *Enfuvirtide*,¹¹⁶ *Ziconotide*,¹¹⁷ *Exenatide*¹¹⁸ and *Pepstatin A*,¹¹⁹ which contain Ala, Gly, Cys and Ser amino acids.

While the development of NATS arguably started more than 60 years ago, and there are still unknowns in the knowledge of this phenomenon, the desire to learn from nature, and even to outcompete it, will certainly fuel further development towards a true artificial ribosome.

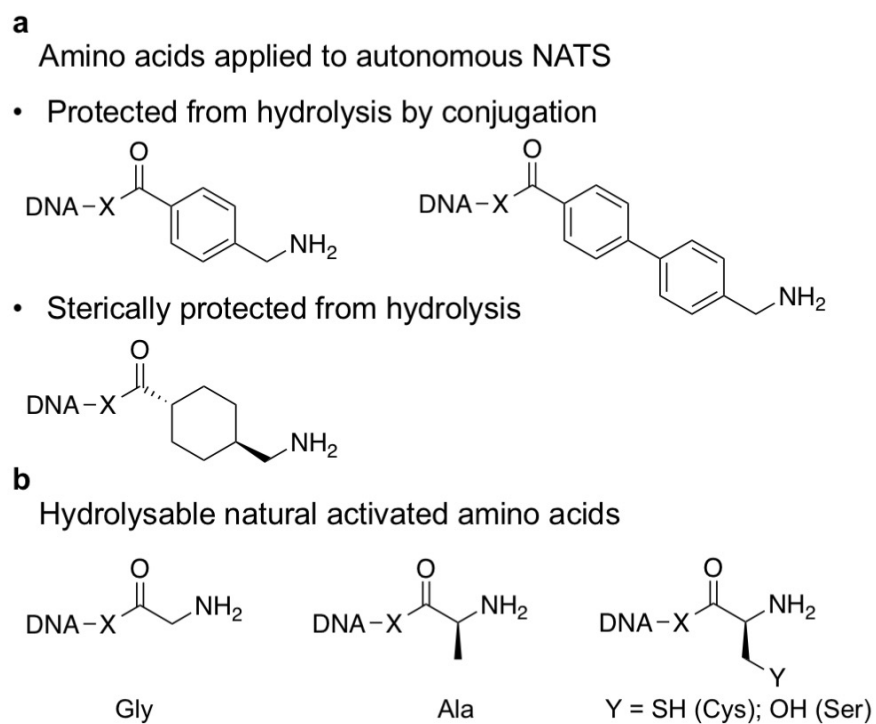


Figure 1-15. Comparison between unnatural amino acids used for autonomous NATS, and natural amino acids. a) Unnatural amino acids previously applied to autonomous NATS of peptides. b) Natural amino acids susceptible of rapid hydrolysis. X = *N*-hydroxysuccinimide ester.

1.3 Project aims

The previous examples of the templated synthesis of peptide bonds demonstrated how NATS can be used to program peptide synthesis. By comparison to the ribosomal synthesis of peptides, many approaches differ in the fact that they require a sequential addition of the DNA adapters,^{82,94} while autonomous NATS of peptides facilitates the multistep peptide synthesis in one pot, obtaining a genetically encoded libraries of peptides.^{112,113}

Autonomous NATS of peptide bonds required the dissolution of DNA adapters that remained in solution over the entire multistep synthesis. The conceptual demonstration of autonomous NATS of peptide bonds was performed with a limited range of activated esters, chosen due to their stability towards hydrolysis in aqueous solution. However, to date, it has not been possible to apply autonomous NATS to the synthesis of natural peptides, due to the likelihood of natural amino acid activated esters to hydrolyse.

Here, the aim is to develop strategies to prevent the hydrolysis of natural amino acid activated ester DNA adapters, and to perform NATS of natural peptide bonds. To that end, in the first chapter, the DNA templated synthesis (DTS) of peptide bonds mediated by highly reactive activated esters was assayed in a low water activity solvent compatible with dsDNA formation, namely Deep Eutectic Solvents (DESs). In the second chapter, highly stable phenol esters were activated towards peptide bond formation *in situ*, only after hybridisation with the DNA template, through native chemical ligation. Finally, in the third chapter, the DNA template was replaced by a PNA template, and the hybridisation in a high organic solvent solution was determined.

1.4 Bibliography

- 1 J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochemistry, 7th Edition*, 2012.
- 2 M. Paoli, R. Liddington, J. Tame, A. Wilkinson and G. Dodson, *J. Mol. Biol.*, 1996, **256**, 775–792.
- 3 T. Gutmann, K. H. Kim, M. Grzybek, T. Walz and Ü. Coskun, *J. Cell Biol.*, 2018, **217**, 1643–1649.
- 4 L. M. Furtado, R. Somwar, G. Sweeney, W. Niu and A. Klip, *Biochem. Cell Biol.*, 2002, **80**, 569–578.
- 5 E. Uchikawa, E. Choi, G. Shang, H. Yu and X. Bai, *Elife*, 2019, **8**, e48630.
- 6 R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea and J. P. Overington, *Nat. Rev. Drug Discov.*, 2016, **16**, 19–34.
- 7 A. Tavassoli, *Curr. Opin. Chem. Biol.*, 2017, **38**, 30–35.
- 8 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
- 9 F. Crick, *Nature*, 1970, **227**, 561–563.
- 10 R. E. Franklin and R. G. Gosling, *Acta Crystallogr.*, 1953, **6**, 673–677.
- 11 M. Nirenberg, *Trends Biochem. Sci.*, 2004, **29**, 46–54.
- 12 F. Vanzi, S. Vladimirov, C. R. Knudsen, Y. E. Goldman and B. S. Cooperman, *Rna*, 2003, **9**, 1174–1179.
- 13 A. Yonath, *Angew. Chemie Int. Ed.*, 2010, **49**, 4340–4354.
- 14 Y. Hashem, A. des Georges, J. Fu, S. N. Buss, F. Jossinet, A. Jobe, Q. Zhang, H. Y. Liao, R. A. Grassucci, C. Bajaj, E. Westhof, S. Madison-Antenucci and J. Frank, *Nature*, 2013, **494**, 385–389.
- 15 N. Bilgin, F. Claesens, H. Pahverk and M. Ehrenberg, *J. Mol. Biol.*, 1992, **224**,

1011–1027.

- 16 S. Proshkin, A. Rachid Rahmouni, A. Mironov and E. Nudler, *Science.*, 2010, **328**, 504–508.
- 17 J. R. Peacock, R. R. Walvoord, A. Y. Chang, M. C. Kozlowski, H. Gamper and Y. M. Hou, *RNA*, 2014, **20**, 758–764.
- 18 R. F. Rosenberger and G. Foskett, *MGG Mol. Gen. Genet.*, 1981, **183**, 561–563.
- 19 E. B. Kramer and P. J. Farabaugh, *Rna*, 2007, **13**, 87–96.
- 20 P. Edelmann and J. Gallant, *Cell*, 1977, **10**, 131–137.
- 21 H. S. Zaher and R. Green, *Cell*, 2009, **136**, 746–762.
- 22 V. Ramakrishnan, *Cell*, 2002, **108**, 557–572.
- 23 P. Melius and J. Yon-Ping Sheng, *Bioorg. Chem.*, 1975, **4**, 385–391.
- 24 E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606–631.
- 25 C. Schotten, *Berichte der Dtsch. Chem. Gesellschaft.*, 1884, **17**, 2544–2547.
- 26 E. Baumann, *Berichte der Dtsch. Chem. Gesellschaft*, 1886, **19**, 3218–3222.
- 27 T. Curtius, *J. für Prakt. Chemie*, 1882, **26**, 145–208.
- 28 S. Fuse, Y. Mifune, H. Nakamura and H. Tanaka, *Nat. Commun.*, 2016, **7**, 13491.
- 29 T. B. Trinh, P. Upadhyaya, Z. Qian and D. Pei, *ACS Comb. Sci.*, 2016, **18**, 75–85.
- 30 D. Hetemi and J. Pinson, *Chem. Soc. Rev.*, 2017, **46**, 5701–5713.
- 31 Y. Li, E. Gabriele, F. Samain, N. Favalli, F. Sladojevich, J. Scheuermann and D. Neri, *ACS Comb. Sci.*, 2016, **18**, 438–443.
- 32 R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149–2154.
- 33 V. Mäde, S. Els-Heindl and A. G. Beck-Sickinger, *Beilstein J. Org. Chem.*, 2014,

- 10**, 1197–1212.
- 34 R. P. Feynman, *Caltech Eng. Sci.*, 1960, **23**, 22–36.
- 35 G. Vives and J. M. Tour, *Acc. Chem. Res.*, 2009, **42**, 473–487.
- 36 D. M. Eigler and E. K. Schweizer, *Nature*, 1990, **344**, 524–526.
- 37 C. Daruich De Souza, B. Ribeiro Nogueira and M. E. C. M. Rostelato, *J. Alloys Compd.*, 2019, **798**, 714–740.
- 38 H. Kroto, *Angew. Chemie Int. Ed. English*, 1997, **36**, 1578–1593.
- 39 K. S. Novoselov, *Angew. Chemie Int. Ed.*, 2011, **50**, 6986–7002.
- 40 J. M. Lehn, *Acc. Chem. Res.*, 1978, **11**, 49–57.
- 41 T. Heinz, D. M. Rudkevich and J. Rebek, *Nature*, 1998, **394**, 764–766.
- 42 D. S. Kim and J. L. Sessler, *Chem. Soc. Rev.*, 2015, **44**, 532–546.
- 43 P. L. Anelli, N. Spencer and J. Fraser Stoddart, *J. Am. Chem. Soc.*, 1991, **113**, 5131–5133.
- 44 R. M. Dirks and N. A. Pierce, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15275–15278.
- 45 P. W. K. Rothmund, *Nature*, 2006, **440**, 297–302.
- 46 M. Yoshizawa, J. K. Klosterman and M. Fujita, *Angew. Chemie Int. Ed.*, 2009, **48**, 3418–3438.
- 47 C. J. C. Edwards-Gayle and I. W. Hamley, *Org. Biomol. Chem.*, 2017, **15**, 5867–5876.
- 48 D. N. Woolfson, *Adv. Protein Chem.*, 2005, **70**, 79–112.
- 49 S. Cavalli, F. Albericio and A. Kros, *Chem. Soc. Rev.*, 2010, **39**, 241–263.
- 50 U. Reinhardt, J. Lotze, K. Mörl, A. G. Beck-Sickinger and O. Seitz, *Bioconjug. Chem.*, 2015, **26**, 2106–2117.

- 51 U. Reinhardt, J. Lotze, S. Zernia, K. Mörl, A. G. Beck-Sickinger and O. Seitz, *Angew. Chem. Int. Ed.*, 2014, **53**, 10237–10241.
- 52 R. Issac, Y. W. Ham and J. Chmielewski, *Curr. Opin. Struct. Biol.*, 2001, **11**, 458–463.
- 53 I. Ghosh and J. Chmielewski, *Curr. Opin. Chem. Biol.*, 2004, **8**, 640–644.
- 54 S. K. Rout, M. P. Friedmann, R. Riek and J. Greenwald, *Nat. Commun.*, 2018, **9**, 234–241.
- 55 B. Rubinov, N. Wagner, H. Rapaport and G. Ashkenasy, *Angew. Chem. Int. Ed.*, 2009, **48**, 6683–6686.
- 56 Y. Takahashi and H. Mihara, *Bioorganic Med. Chem.*, 2004, **12**, 693–699.
- 57 B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth, *Proc. Natl. Acad. Sci.*, 2012, **109**, 690–697.
- 58 D. Liu, W. H. Wu, Y. J. Liu, X. L. Wu, Y. Cao, B. Song, X. Li and W. Bin Zhang, *ACS Cent. Sci.*, 2017, **3**, 473–481.
- 59 X. Di Da and W. Bin Zhang, *Angew. Chem. Int. Ed.*, 2019, **58**, 11097–11104.
- 60 N. H. Shah and T. W. Muir, *Chem. Sci.*, 2014, **5**, 446–461.
- 61 I. T. Harrison and S. Harrison, *J. Am. Chem. Soc.*, 1967, **89**, 5723–5724.
- 62 B. Lewandowski, G. De Bo, J. W. Ward, M. Papmeyer, S. Kuschel, M. J. Aldegunde, P. M. E. Gramlich, D. Heckmann, S. M. Goldup, D. M. D'Souza, A. E. Fernandes and D. A. Leigh, *Science*, 2013, **339**, 189–193.
- 63 G. De Bo, S. Kuschel, D. A. Leigh, B. Lewandowski, M. Papmeyer and J. W. Ward, *J. Am. Chem. Soc.*, 2014, **136**, 5811–5814.
- 64 Q. Wan, J. Chen, Y. Yuan and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2008, **130**, 15814–15816.

- 65 G. De Bo, M. A. Y. Gall, M. O. Kitching, S. Kuschel, D. A. Leigh, D. J. Tetlow and J. W. Ward, *J. Am. Chem. Soc.*, 2017, **139**, 10875–10879.
- 66 G. De Bo, M. A. Y. Gall, S. Kuschel, J. De Winter, P. Gerbaux and D. A. Leigh, *Nat. Nanotechnol.*, 2018, **13**, 381–385.
- 67 H. Kaur, A. Arora, J. Wengel and S. Maiti, *Biochemistry*, 2006, **45**, 7347–7355.
- 68 K.-U. Schoning, *Science*, 2000, **290**, 1347–1351.
- 69 P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497–1500.
- 70 M. Madsen and K. V. Gothelf, *Chem. Rev.*, 2019, **119**, 6384–6458.
- 71 M. J. Catalano, N. E. Price and K. S. Gates, *Bioorganic Med. Chem. Lett.*, 2016, **26**, 2627–2630.
- 72 R. Naylor and P. T. Gilham, *Biochemistry*, 1966, **5**, 2722–2728.
- 73 Z. J. Gartner, M. W. Kanan and D. R. Liu, *Angew. Chem. Int. Ed.*, 2002, **41**, 1796–1800.
- 74 Z. J. Gartner, R. Grubina, C. T. Calderone and D. R. Liu, *Angew. Chem. Int. Ed.*, 2003, **42**, 1370–1375.
- 75 R. Oppenheimer, University of Oxford, 2019.
- 76 M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Hoøjfeldt, K. V. Gothelf and N. J. V. Hansen, *J. Am. Chem. Soc.*, 2009, **131**, 1322–1327.
- 77 J. Sayers, R. J. Payne and N. Winssinger, *Chem. Sci.*, 2018, **9**, 896–903.
- 78 Y. Brudno, M. E. Birnbaum, R. E. Kleiner and D. R. Liu, *Nat. Chem. Biol.*, 2010, **6**, 148–155.
- 79 Z. J. Gartner and D. R. Liu, *J. Am. Chem. Soc.*, 2001, **123**, 6961–6963.
- 80 Z. J. Gartner, M. W. Kanan and D. R. Liu, *Angew. Chem. Int. Ed.*, 2002, **41**,

1796–1800.

- 81 X. Li, Z. J. Gartner, B. N. Tse and D. R. Liu, *J. Am. Chem. Soc.*, 2004, **126**, 5090–5092.
- 82 Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder and D. R. Liu, *Science*, 2004, **305**, 1601–1605.
- 83 D. Li, X. Wang, F. Shi, R. Sha, N. C. Seeman and J. W. Canary, *Org. Biomol. Chem.*, 2014, **12**, 8823–8827.
- 84 A. Roloff and O. Seitz, *Chem. Sci.*, 2013, **4**, 432–436.
- 85 A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli and T. Brown, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11338–11343.
- 86 M. Kukwikila, N. Gale, A. H. El-Sagheer, T. Brown and A. Tavassoli, *Nat. Chem.*, 2017, **9**, 1089–1098.
- 87 A. H. El-Sagheer, R. Kumar, S. Findlow, J. M. Werner, A. N. Lane and T. Brown, *ChemBioChem*, 2008, **9**, 50–52.
- 88 M. W. Kanan, M. M. Rozenman, K. Sakural, T. M. Snyder and D. R. Liu, *Nature*, 2004, **431**, 545–549.
- 89 H. Yang, A. Z. Rys, C. K. McLaughlin and H. F. Sleiman, *Angew. Chem. Int. Ed.*, 2009, **48**, 9919–9923.
- 90 S. Barluenga and N. Winssinger, *Acc. Chem. Res.*, 2015, **48**, 1319–1331.
- 91 M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, A. J. Turberfield and R. K. O'Reilly, *Angew. Chem. Int. Ed.*, 2010, **49**, 7948–7951.
- 92 R. K. Bruick, P. E. Dawson, S. B. H. Kent, N. Usman and G. F. Joyce, *Chem. Biol.*, 1996, **3**, 49–56.
- 93 M. L. McKee, A. C. Evans, S. R. Gerrard, R. K. O'Reilly, A. J. Turberfield and E.

- Stulz, *Org. Biomol. Chem.*, 2011, **9**, 1661–1666.
- 94 Y. He and D. R. Liu, *J. Am. Chem. Soc.*, 2011, **133**, 9972–9975.
- 95 J. Michaelis, G. J. Van Der Heden Van Noort and O. Seitz, *Bioconjug. Chem.*, 2014, **25**, 18–23.
- 96 T. N. Grossmann and O. Seitz, *J. Am. Chem. Soc.*, 2006, **128**, 15596–15597.
- 97 M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, A. J. Turberfield and R. K. O'Reilly, *Angew. Chem. Int. Ed.*, 2010, **49**, 7948–7951.
- 98 X. H. Chen, A. Roloff and O. Seitz, *Angew. Chem. Int. Ed.*, 2012, **51**, 4479–4483.
- 99 A. Shibata, T. Uzawa, Y. Nakashima, M. Ito, Y. Nakano, S. Shuto, Y. Ito and H. Abe, *J. Am. Chem. Soc.*, 2013, **135**, 14172–14178.
- 100 H. Wu, B. T. Cisneros, C. M. Cole and N. K. Devaraj, *J. Am. Chem. Soc.*, 2014, **136**, 17942–17945.
- 101 K. Gorska and N. Winssinger, *Angew. Chem. Int. Ed.*, 2013, **52**, 6820–6843.
- 102 D. Chang, E. Lindberg and N. Winssinger, *J. Am. Chem. Soc.*, 2017, **139**, 1444–1447.
- 103 K. T. Kim, D. Chang and N. Winssinger, *Helv. Chim. Acta*, 2018, **101**, e1700295.
- 104 R. M. Franzini and E. T. Kool, *J. Am. Chem. Soc.*, 2009, **131**, 16021–16023.
- 105 A. Osuna Gálvez and J. W. Bode, *J. Am. Chem. Soc.*, 2019, **141**, 8721–8726.
- 106 P. Cuatrecasas and I. Parikh, *Biochemistry*, 1972, **11**, 2291–2299.
- 107 D. L. Usanov, A. I. Chan, J. P. Maianti and D. R. Liu, *Nat. Chem.*, 2018, **10**, 704–714.
- 108 N. Srinivas, T. E. Ouldrige, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye and E. Winfree, *Nucleic Acids Res.*, 2013, **41**, 10641–10658.
- 109 R. A. Goodnow, C. E. Dumelin and A. D. Keefe, *Nat. Rev. Drug Discov.*, 2017,

- 16, 131–147.
- 110 J. Nielsen, S. Brenner and K. Janda, *J. Am. Chem. Soc.*, 1993, **115**, 9812–9813.
- 111 D. R. Halpin and P. B. Harbury, *PLoS Biol.*, **2**, e173.
- 112 Y. He and D. R. Liu, *Nat. Nanotechnol.*, 2010, **5**, 778–782.
- 113 W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes, A. H. El-Sagheer, J. Bath, B. G. Davis, T. Brown, R. K. O'Reilly and A. J. Turberfield, *Nat. Chem.*, 2016, **8**, 542–548.
- 114 Y. Liu and S. C. West, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 937–944.
- 115 F. Debruyne, G. Bhat and M. B. Garnick, *Futur. Oncol.*, 2006, **2**, 677–696.
- 116 J. P. Lalezari, K. Henry, M. O'Hearn, J. S. G. Montaner, P. J. Piliero, B. Trottier, S. Walmsley, C. Cohen, D. R. Kuritzkes, J. J. Eron, J. Chung, R. DeMasi, L. Donatucci, C. Drobnes, J. Delehanty and M. Salgo, *N. Engl. J. Med.*, 2003, **348**, 2175–2185.
- 117 J. G. McGivern, *Neuropsychiatr. Dis. Treat.*, 2007, **3**, 69–85.
- 118 Lonza Ltd., Switz.; Lonza Braine S. A. ., *PCT Int. Appl.*, 2011, 92pp.
- 119 H. Umezawa, T. Aoyagi, H. Morishima, M. Matsuzaki, M. Hamada and T. Takeuchi, *J. Antibiot.*, 1970, **23**, 259–262.

Chapter 2

DNA templated synthesis (DTS) in Deep Eutectic Solvents (DES)

2.1 Introduction

Ionic liquids (ILs) were first reported by Paul Walden in 1914 when he discovered that ethylammonium nitrate had a melting point of 12 °C.¹ Since then, a large number of ILs have been described.^{2,3} ILs are liquid phases formed by a combination of positively and negatively charged species. In addition to the low melting temperature, ionic liquids display high viscosity and low vapour pressure.⁴⁻⁷ Due to the ability to tune their properties by choosing the appropriate combination of ions, ILs have received great attention from the chemistry community. They have been widely studied as reaction solvents in both a research context and in the industrial scale production of chemicals.⁸⁻¹¹

Deep Eutectic Solvents (DESs) are a closely related family of solvents, that can contain non-ionic components, with low toxicity and formed from a simple mixture of renewable and cheap components. They were initially described by Abbott and co-workers upon the realisation that some quaternary ammonium halide salts and hydrogen bond donor mixtures had a remarkably low melting temperature (as low as 12 °C for urea / choline chloride in a 2:1 ratio) when mixed in the right proportion to reach the eutectic point (Figure 2-1 a).¹² Due to their intriguing physicochemical properties, together with their potential applications in industrial processes, DESs have been the subjects of intense

study despite their relatively recent discovery. There were over 750 publications featuring the concept “Deep Eutectic Solvents” in 2018, and the original article from Abbott and co-workers¹³ has been cited more than 1400 times.[†]

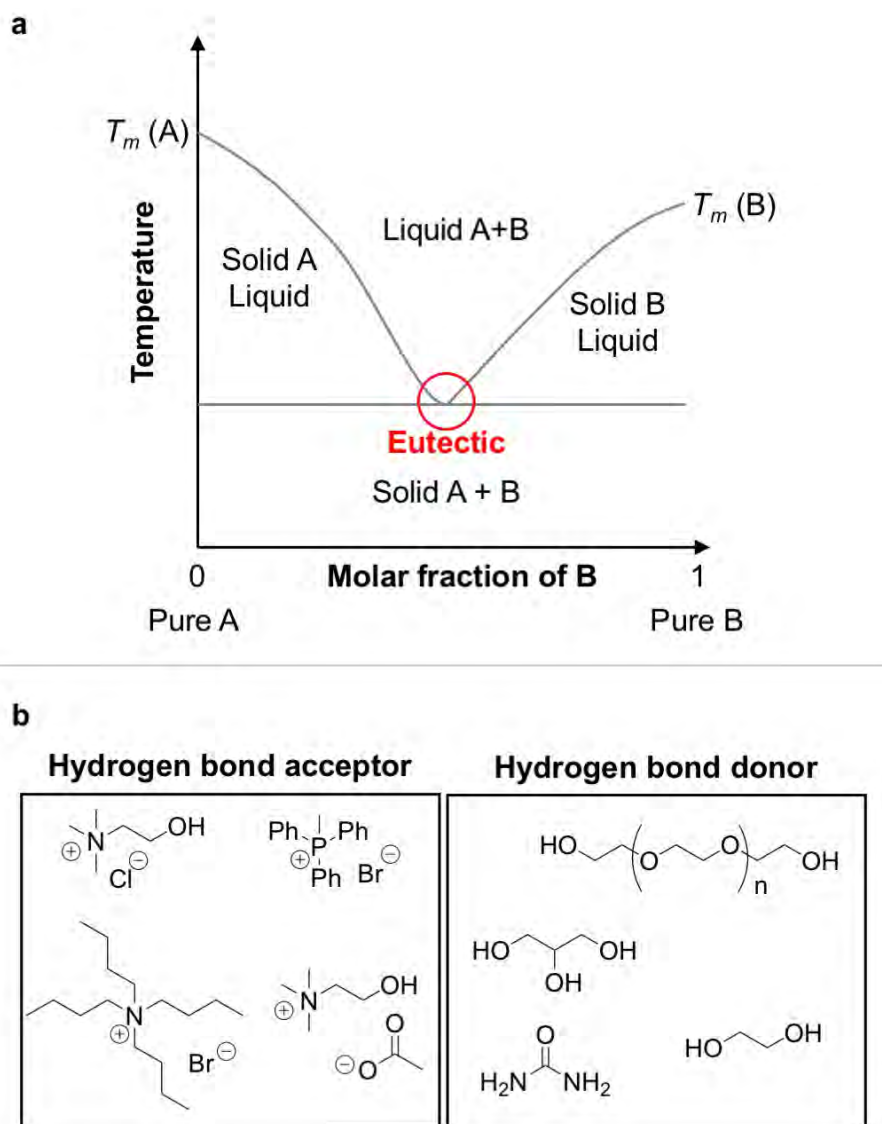
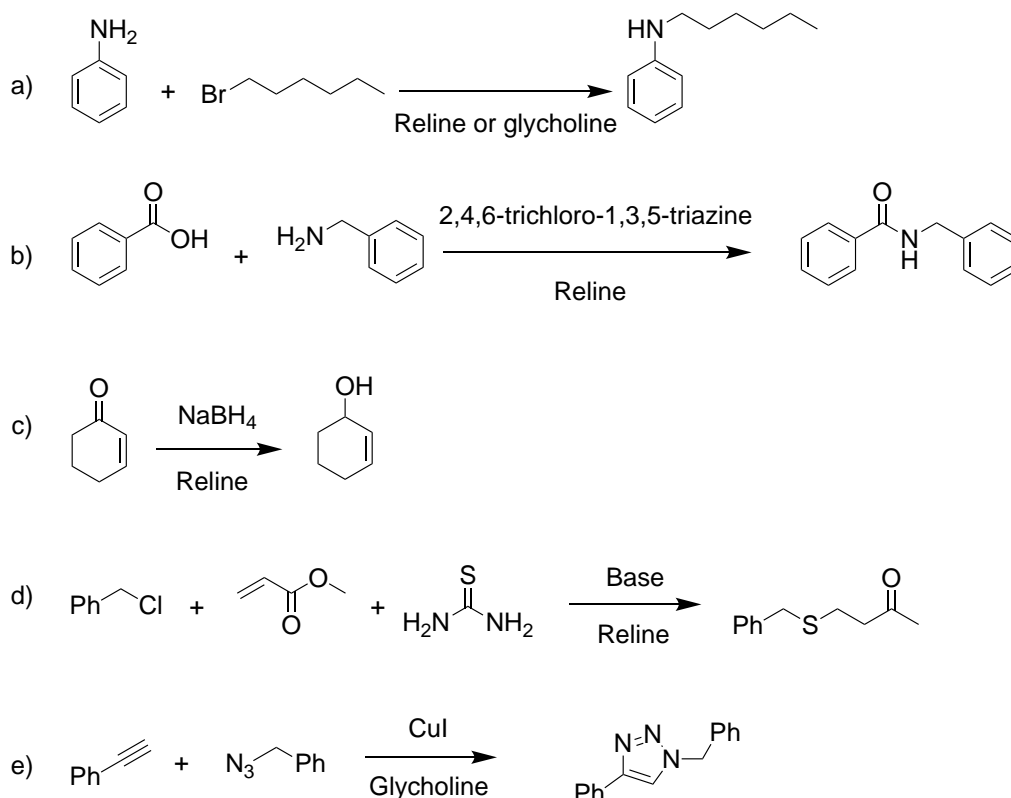


Figure 2-1. General description of deep eutectic solvents. a) Schematic binary phase diagram showing the different states as a function of the temperature and the content of individual components. b) Examples of hydrogen bond acceptors and hydrogen bond donors that have been used to produce deep eutectic solvents.¹⁴

[†] Bibliometrics from Web of Science (Reuters, October 2019)

DESs retain most properties of ILs such as low vapour pressure, high viscosity, high ionic content and high conductivity. However, while the concept IL and DES might seem interchangeable, they describe fundamentally different systems: DESs are formed by a eutectic mixture of a Lewis or Brønsted acid and a base (Figure 2-1 b), which can contain a high concentration of charged species,^{13,15,16} while ILs are formed exclusively by a mixture of cationic and anionic species.⁴ In DESs, the acidic component is usually described as the hydrogen bond donor (HBD) and the basic component the hydrogen bond acceptor (HBA). DESs they can be classified into 4 types according to their components.¹⁶ Type I are formed by quaternary ammonium salts and metal salts; Type II are formed by quaternary ammonium salts and hydrated metal salts; Type III are formed by quaternary ammonium salts and hydrogen bond donors and Type IV are formed by metal halides and hydrogen bond donors. Due to their low toxicity, simple preparation and reported compatibility with biomolecules, this study focused on the use of Type III DESs.^{14,17,18}

A wide range of reactions have been reported to proceed efficiently in DESs.¹⁹⁻²¹ In Scheme 2-1, a selection of chemistries relevant for the synthesis of bioactive molecules or for bioconjugation are shown.



Scheme 2-1. Several examples of chemical transformations in Deep Eutectic Solvents. a) Nucleophilic substitution²² b) Peptide bond formation²³ c) Enone reduction²⁴ d) Thiol-Michael click reaction²⁵ e) Copper catalysed azide-alkyne click reaction.²⁶ Glycholine = glycerol / choline chloride; Reline = urea / choline chloride.

Type III DESs display low toxicity and have been found to be compatible with biomacromolecules such as nucleic acids and proteins.^{14,17,27} A large number of enzymatic reactions have been reported to proceed in DESs^{14,17} such as transesterification,^{28,29} aldol and nitro-aldol reactions,^{30,31} oxidation³² and peptide bond formation³³ among many others. While it has been reported that a small degree of conformational change of proteins takes place upon dissolution in DES,³⁴ the fact that the vast majority of the reported biocatalysed transformations were performed in choline chloride or choline acetate based Type III DESs is an indication that this particular type of solvent does not substantially disrupt the conformation of large proteins.

The addition of water to choline chloride based DESs in enzymatic reactions was particularly relevant to the present research, as water can be used to reduce the viscosity and melting point of the DES mixture.^{35,36} However, the main objective in the present work was to reduce the hydrolysis of activated esters to allow higher yielding nucleic acid templated synthesis of peptides, and it would be expected that competing hydrolysis would dominate over peptide bond synthesis in solutions containing a large proportion of water. In this regard, it was reported that enzyme catalysed esterifications in pure DES produced a very low conversion, the addition of 8 to 10 vol% H₂O increased the conversion up to 100%, and further addition of H₂O decreased the yield, most likely due to hydrolysis of the product.^{37,38} These two examples show how the properties of DESs could be tailored by the addition of a small volume of water while not adding significant competing hydrolysis.

In addition to protein biomolecules, the supramolecular interactions of nucleic acids in ILs and DESs have been a matter of interest, and it has been comprehensively reviewed in the past.^{18,27,39–42} DNA has been shown to be chemically stable in DES for up to 6 months.⁴³ The formation of DNA duplex, triplex and quadruplex in DES was initially reported in 2010 by Mamajanov, Hud and co-workers,⁴⁴ although the melting temperature of the double strand was reduced from 73 °C in 100 mM NaCl (aq) to 37 °C in glycholine. Furthermore, the folding of a highly complex DNA origami has also been reported in hydrated DES.⁴⁵

Altogether, these findings show that DESs provide a suitable anhydrous or low water medium for nucleic acid hybridisation while preserving their chemical stability. Due to the limited solubility of DNA in organic solvents and the even further limited tolerance of the double stranded supramolecular structure, mostly limited to ethylene glycol and

glycerol solutions,^{46–48} these results represent an exciting opportunity to explore water-sensitive templated chemistries, incompatible with aqueous solution, in DESs.

2.2 Results and discussion

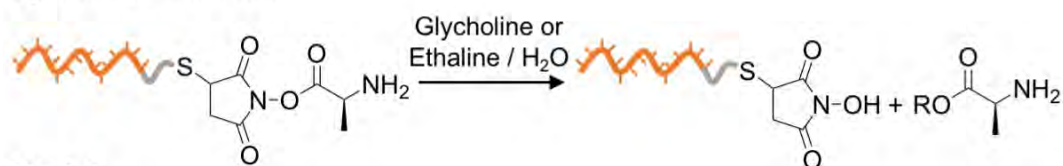
2.2.1 Small molecule activated ester stability in DESs

The DNA molecular assemblers described previously (Section 1.4.5) used *N*-hydroxysuccinimide activated esters conjugated to DNA in order to template the synthesis of peptides.^{49,50} The use of highly reactive esters allows for fast amidation kinetics, compatible with a fast multistep DNA annealing and displacement mechanism. However, due to their high reactivity these activating groups hydrolyse quickly in aqueous solution. The hydrolysis of *N*-hydroxysuccinimide DNA adapters has previously limited the multistep DTS of peptide bonds.⁵⁰

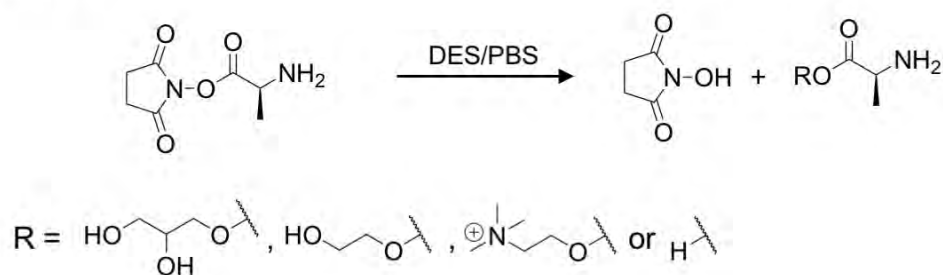
One of the key objectives in this work was to generalise the DTS to natural amino acids, moving away from the unnatural amino acids used previously.^{49–51} Based on the stability of several aminoacyl-^tRNA conjugates,^{52,53} alanine represented a good test candidate: it undergoes fast hydrolysis,⁵³ is one of the most commonly found amino acids in natural peptides,⁵⁴ and there are many commercially available derivatives and synthetic protocols.

Before conducting a direct investigation on DNA conjugates, stability and reactivity tests were performed on a small molecule model, alanine *N*-hydroxysuccinimide ester (Scheme 2-2). This approach allowed to use nuclear magnetic resonance (NMR) spectroscopy to study the solvolysis process, in addition to chromatographic analysis.

System of interest



Model



Scheme 2-2. Comparison between the system of interest composed of a N-hydroxysuccinimide ester conjugated to DNA, and the model small molecule system. PBS = buffered aqueous phosphate solution containing NaCl, derived from the “phosphate-buffered saline”.

Ethaline (ethylene glycol / choline chloride), reline (urea / choline chloride) and glycoline (glycerol / choline chloride) type III DESs were selected (Table 2-1) for this study based on their low toxicity,⁵⁵ simple preparation protocol¹⁵ and the well-reported stabilisation of double-stranded DNA (dsDNA).⁴⁴

Table 2-1. DESs used in this work. HBD = hydrogen bond donor. HBA = hydrogen bond acceptor.

DES name	HBD	HBA	HBD : HBA molar ratio
Ethaline	<chem>OCCO</chem>	<chem>[N+](C)(C)CCO.[Cl-]</chem>	2:1
Glycoline	<chem>OCC(O)CO</chem>	<chem>[N+](C)(C)CCO.[Cl-]</chem>	2:1
Reline	<chem>NC(=O)N</chem>	<chem>[N+](C)(C)CCO.[Cl-]</chem>	2:1

The properties of DESs can be tailored by the incorporation of additives. In the current work, a small amount of H₂O was added to reduce the high viscosity of DESs, which is for example in the case of anhydrous reline 632 cP (as a comparison point, the viscosity of isopropanol is 2.1 cP).⁵⁶ The addition of water has been demonstrated to

progressively alter the properties of DESs such as the conductivity and hydrogen bond strength.⁵⁷ However, there is evidence that the microstructure of DES is preserved with water contents as high as 25 to 42 wt%.^{57,58} In the present work, 7 vol% water was added to many DESs mixtures. A DES formed by glycholine with a 7 vol% H₂O will be denoted by Gly_{0.07}.

With the objective in mind of using DESs as a medium for peptide bond formation reactions, it was very important to adjust the acidity of these solutions close to neutrality to minimise hydrolysis. The acidity of aqueous solutions is ubiquitously measured using the pH scale, this is facilitated by internationally recognised pH standards such as the silver / silver chloride electrode or the hydrogen electrode. Determining the acidity of ILs and DESs is more challenging due to the lack of internationally recognised standards. Previous attempts to quantify the acidity or basicity of DESs in the past have used either a standard pH sensitive glass electrode^{59,60} or a pH indicator such as bromophenol blue or phenol red.⁶¹ These results have shown that ethaline is slightly acidic (pH 4.7), glycholine is neutral (pH 7.5) and reline is considerably basic (pH 10.4).

One of the most common strategies to determine the pH of aqueous solutions is to use a pH sensitive glass electrode. However, using a glass electrode to determine the acidity in anhydrous media is not straightforward, due to complex effects of the dehydration of the glass membrane and the formation of water layers between the solution and the electrode.⁶² As the current study was not focused on the physicochemical properties of DESs, but rather oriented to build up upon previous knowledge to use them as a solvent medium for DNA templated peptide bond

formation, here the acidity was studied by dilution of 1 mL of DES in 4 mL of water, and the pH was determined using a calibrated pH sensitive electrode (Table 2-2).

Table 2-2. Base additives to tune the acidity of DESs. The pH was determined using a selective glass electrode, the instrumental error was quantified by triplicate measurements of independent solutions, and it did not exceed 0.1 pH units (Experimental section 2.4.4). ^aThe HBA choline chloride was replaced by choline acetate in several molar fraction of the HBA up to 100%

DES name	Base additive	Concentration	pH
Ethaline	none	-	6.1
	choline acetate	0 to 1 as HBA mol fraction ^a	6.1 to 6.3
	sodium acetate	0.5 M	7.8
Glycholine	none	-	6.8
	urea	0.3 to 3.3 M	6.6 to 6.8
	sodium acetate (NaOAc)	0.1 to 0.2 M	7.2 to 7.5
	triethylamine (Et ₃ N)	1 to 100 mM	7.3 to 12
Reline	none	-	9.6

It was shown that ethaline was slightly acidic with a pH of 6.1, in good agreement with the literature.⁵⁹ It was anticipated that the substitution of the chlorine counter anion by acetate would reduce this acidity, however, when increasing molar fractions of choline chloride were replaced with choline acetate, this only produced a mild reduction of the acidity to a maximum pH of 6.3 (Experimental Section 2.4.4). This surprising result contrasts with the drastic increase in the pH upon the addition of sodium acetate (NaOAc) to the aqueous solution (*e.g.* the solution with only choline acetate contains 2.9 M acetate showing a pH = 6.3, however a 0.1 M NaOAc solution displays a pH = 8.9). This highlights how the formation of the eutectic mixture deeply influences the chemical activity of the species in DESs. The addition of NaOAc to an ethaline solution did increase the pH to 7.8.

Upon dilution in water, glycholine produced a pH very close to neutral. The addition of urea did not allow accurate adjustment of the acidity (Experimental Section 2.4.4), while the addition of triethylamine (Et₃N) had a larger impact than the addition of NaOAc. Finally, reline was measured moderately basic, in good agreement with reported data.⁵⁹

The hydrolysis rate of a *N*-hydroxysuccinimide (SuO) ester of alanine in several DESs was assessed using reverse-phase high performance liquid chromatography (RP-HPLC) with a UV absorption detector (Figure 2-2). A set of 5 mM solutions of **2-1** were prepared in several solvents, incubated and analysed at 6, 24 and 30 h. The elution of standards showed a good separation between the free SuOH (**2-2**) and the starting material (**2-1**) which overlapped with the hydrolysis product (**2-3**). The progress of hydrolysis was determined using the absolute peak area corresponding to SuOH (**2-2**), compared to the peak area of a 5 mM SuOH solution as a reference (Experimental section 2.4.5).

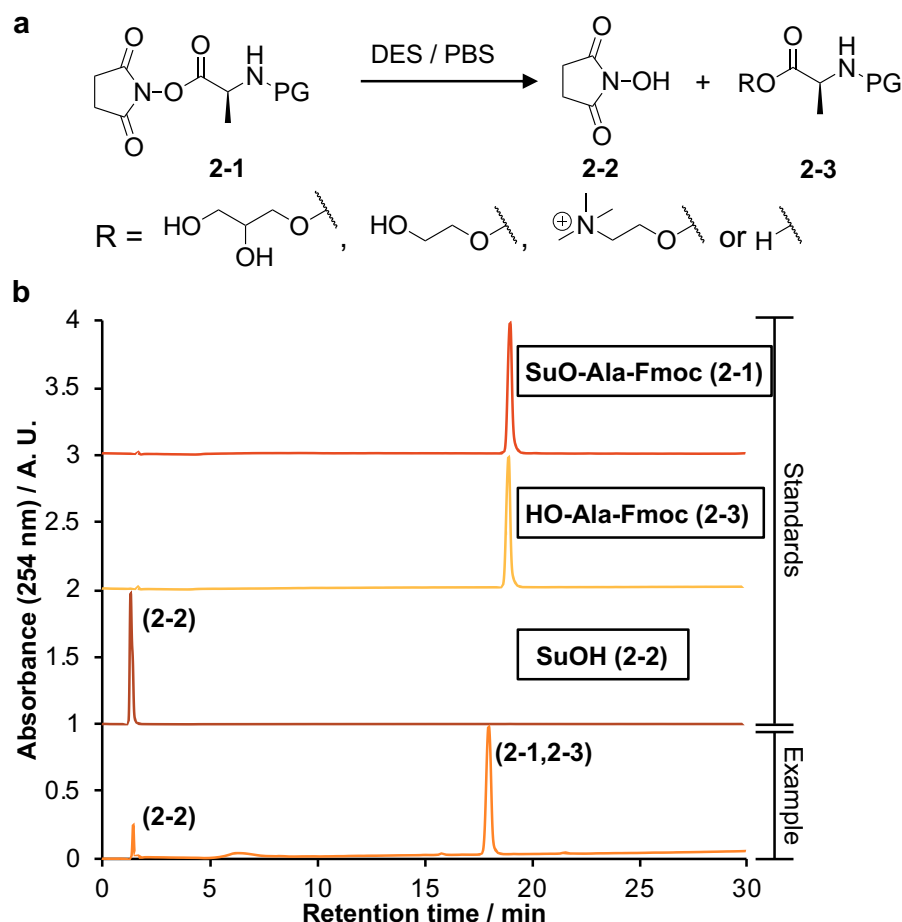


Figure 2-2. Representation of the small molecule **2-1** stability experiment with HPLC analysis. The protecting group in **2-1** and **2-3** (PG) corresponds to Fmoc. a) Scheme of the solvolysis process that cleaves the ester bond. Due to the presence of several nucleophiles in DES solutions, a few decomposition products are possible, represented by the generic R group b) Example of HPLC chromatograms. The standards showed that N-hydroxysuccinimide (**2**) eluted at a very distinct time from compounds **2-1** and **2-3**. The example chromatogram corresponds to 4 h solvolysis in Gly_{0.07}, the retention time of **2-1** and / or **2-3** in the example varied slightly compared to the standards most likely due to insufficient equilibration time between chromatographic analyses.

In order to scope a broad range of solvents and bases, based on the previous pH tests (Table 2-2), glycholine was combined with NaOAc and Et₃N. As reline was basic by itself, it was used without any additive (Figure 2-3).

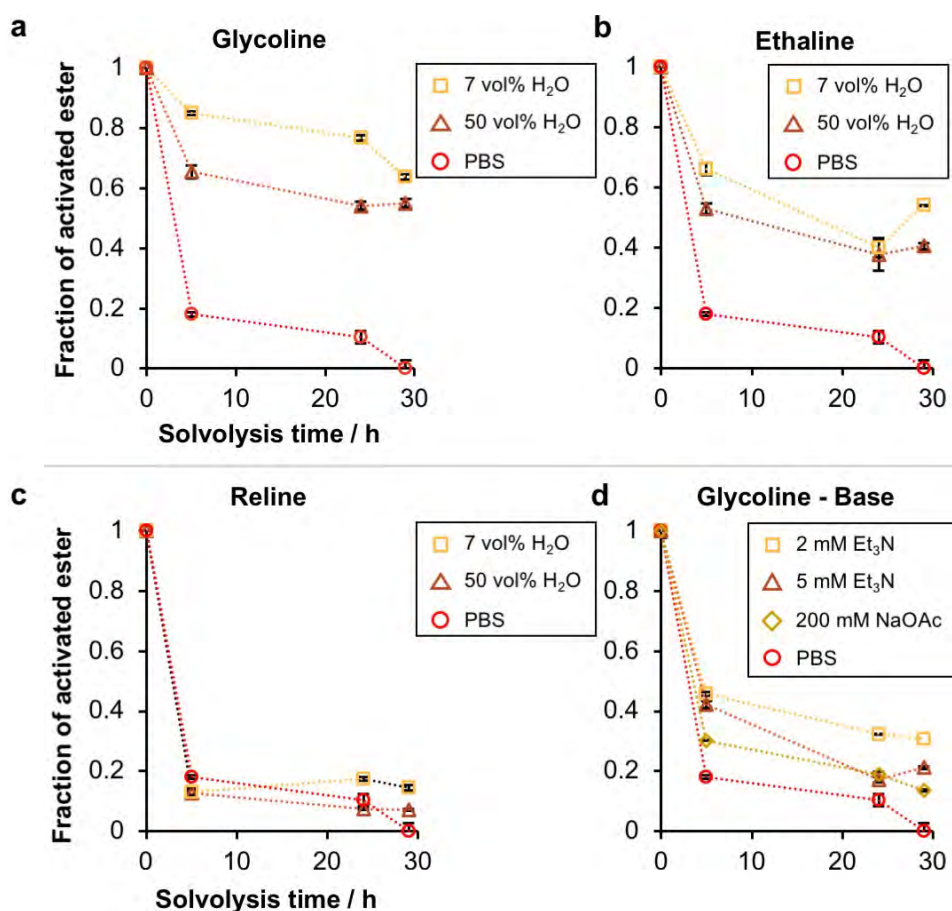


Figure 2-3. Results of the RP-HPLC stability study of **2-1** in several solvents. a) Glycoline / H₂O solutions compared to PBS. b) Ethaline / H₂O solutions compared to PBS. c) Reline / H₂O solutions compared to PBS. d) Gly_{0.07} - base additive solutions compared to PBS (0.1 M, pH = 7.5). Time points 4, 24 and 29 h. Due to the slow chromatographic elution, no replicates were performed, the error was evaluated as the standard deviation on triplicate determination of the peak areas.

DESs that do not contain basic additives, only 7 vol% water, show that the stability follows the trend (from highest to lowest stability) glycoline > ethaline >> reline. The large difference between reline and other DESs is probably related to the basic character of reline that can accelerate the hydrolysis of activated esters (Table 2-2). The addition of 50 vol% of water to glycoline and ethaline only induced a moderate reduction in the stability. In reline, the solvolysis of the activated ester was rapid in both 7 and 50 vol% water. Finally, the addition of base accelerated the solvolysis process. The stability of the activated ester showed the following trend: 2 mM Et₃N > 5 mM Et₃N

>> 200 mM NaOAc. The fastest solvolysis corresponded to a PBS solution (pH 7.5), with 20 mol% activated ester remaining after 4 h. The stability of the activated ester in PBS was very low compared with glycholine and ethaline DESs. DES solutions with a pH very similar to PBS, such as glycholine with 2 mM Et₃N, displayed an enhanced ester stability compared to PBS, demonstrating a stabilisation due to the DES.

In RP-HPLC, the mobile phase is highly aqueous. The DES samples also had to be diluted into aqueous solution prior to analysis. It was therefore possible that some of the solvolysis took place during the analysis. A complementary nuclear magnetic resonance (NMR) stability analysis experiment was designed (Figure 2-4 a). A series of solutions of **1** in DESs were prepared, but this time the solutions were extracted with deuterated acetone, which is immiscible with DESs, and the relative amount of activated ester was determined by proton NMR spectroscopy (¹H-NMR, Figure 2-4 a, Experimental section 2.4.6).[‡] It is worth noting at this point that different partition coefficients of **2-1**, **2-2** and **2-3** would affect the validity of this study. As a starting point, it was assumed that all compounds had similar partition coefficients. Later on, the correlation between the RP-HPLC and the NMR stability experiments will be analysed to ensure the consistency of this hypothesis.

The NMR stability required independent samples for each condition and timepoint, as one sample could not be extracted multiple times. This limited the number of conditions that could be probed; therefore, the study was focused only on glycholine DESs, which showed the highest stability results (Figure 2-4 b).

[‡] DES solutions were extracted with acetone-*d*₆. PBS samples were extracted with chloroform-*d*₃ due to the immiscibility with aqueous solution.

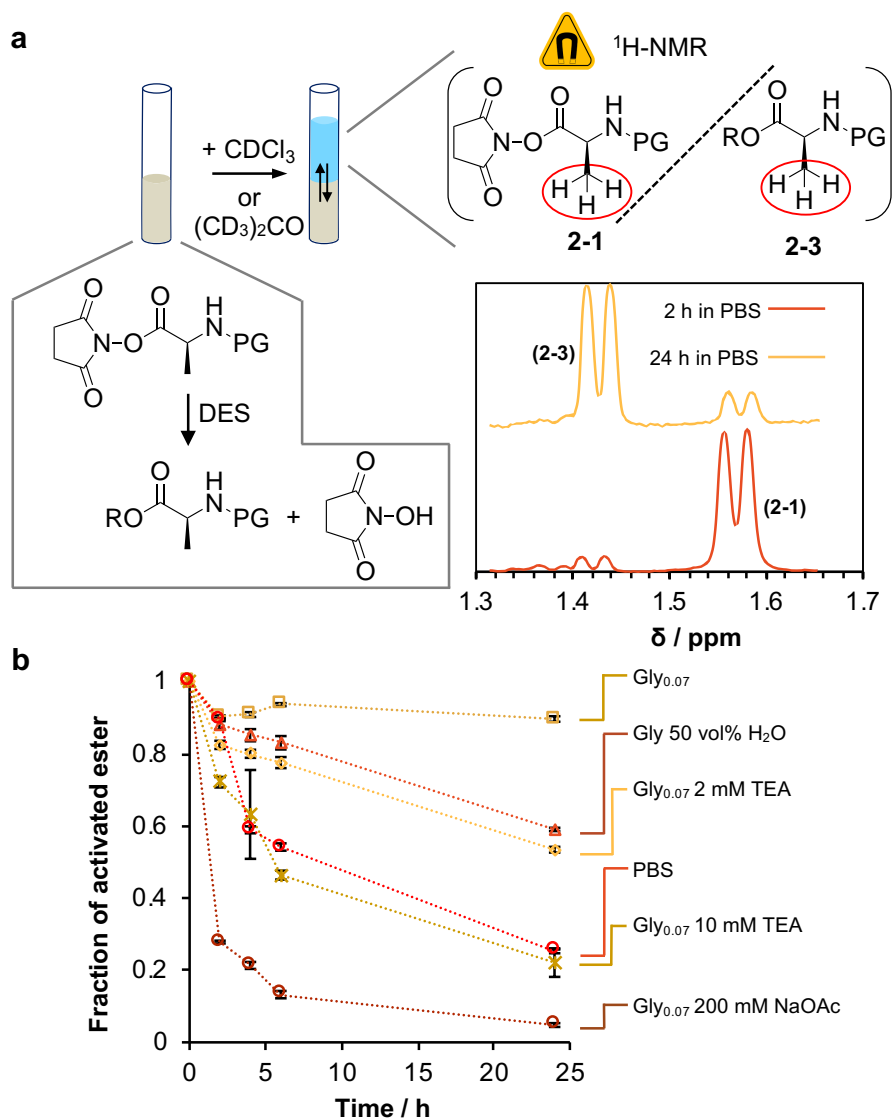


Figure 2-4. $^1\text{H-NMR}$ stability experiment of the activated ester **2-1** in several DES solution. a) Schematic representation of the process showing the solvolysis and subsequent extraction with CDCl_3 or $(\text{CD}_3)_2\text{CO}$. The hydrogen atoms that produce distinguishable signals are highlighted in red. b) Evolution of the fraction of activated ester over time. Gly_{0.07} = glycoline with 7 vol% H_2O . PBS (0.1 M, pH 7.5). Error bars correspond to the standard deviation over triplicate integration events.

In terms of activated ester stability, the increase of the water content in glycoline, from 7 to 50 vol% had a moderate destabilising effect. The stability in glycoline/base mixtures followed the trend 2 mM Et_3N > 10 mM Et_3N >> 200 mM NaOAc, analogous to the previous results (Figure 2-3 d). One notable difference between assays was that the solvolysis in PBS studied by RP-HPLC was considerably faster than in 200 mM

NaOAc glycholine, while in the $^1\text{H-NMR}$ study the solvolysis was moderately slower in PBS than in 200 mM NaOAc glycholine. This might be due to a higher affinity of the hydrolysed product for the aqueous layer that affected the $^1\text{H-NMR}$ study.

Overall, the stability in Gly_{0.07} was the highest while the addition of Et₃N was shown to have less of a negative impact than NaOAc for equally basic glycholine solutions (Figure 2-5).

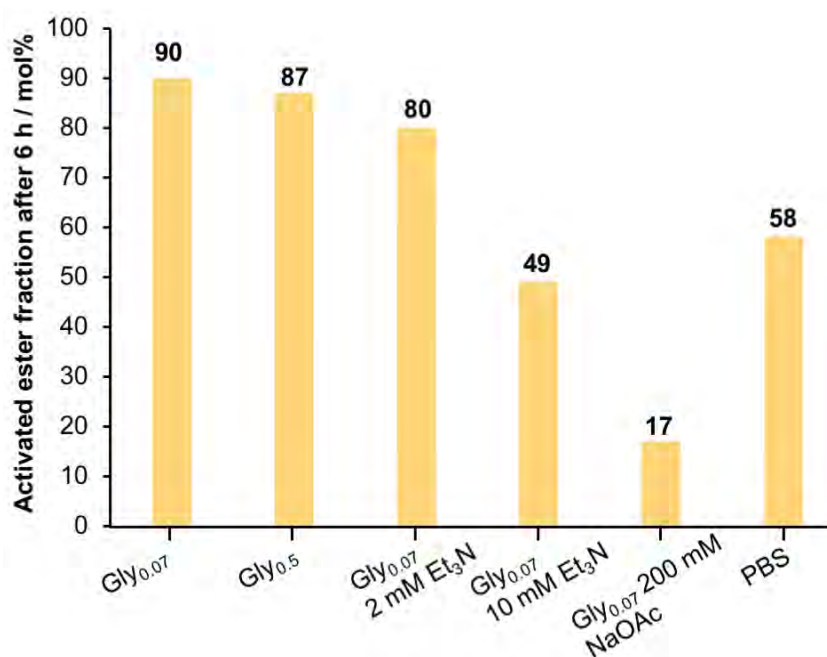
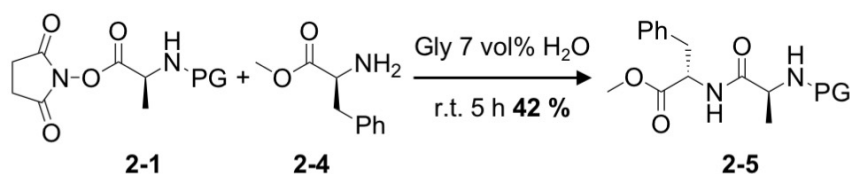


Figure 2-5. Fraction of activated ester after 6 h in several DES - additive mixtures, determined by $^1\text{H-NMR}$ spectroscopy.

Both the RP-HPLC and the $^1\text{H-NMR}$ spectroscopy studies confirmed that the stability of activated esters was higher in glycholine than in PBS. The addition of H₂O to the DES reduced the stability. Upon the addition of base, for equally basic glycholine solution, Et₃N had a lower negative impact on the stability than NaOAc. Finally, the stability was inversely proportional to the concentration of base. Despite their individual drawbacks, the combination of both assays allowed to confidently conclude that glycholine DES was an advantageous solvent over conventional aqueous solution.

Having established that activated esters are more stable in glycholine DESs than in aqueous solution, the next logical step was to demonstrate that peptide bond formation could take place in DESs. The formation of an amide bond between an N-protected *L*-alanine *N*-hydroxysuccinimide ester and a methyl ester of *L*-phenylalanine was assayed in Gly_{0.07} (Scheme 2-3, Experimental section 2.4.7).



Scheme 2-3. Synthesis of the dipeptide **2-5** in DESs.

The synthesis produced the desired amidation product (**2-5**) in a moderate 42% yield after the precipitation of the dipeptide by the addition of water to the DES solution and isolation by column chromatography. The low yield was a reason for concern; however, the formation of the expected dipeptide showed the feasibility of peptide bond formation in DES. It was later reported that the use of cyanuric chloride can allow the synthesis of amide bonds in reline with high yield (up to 90%).²³

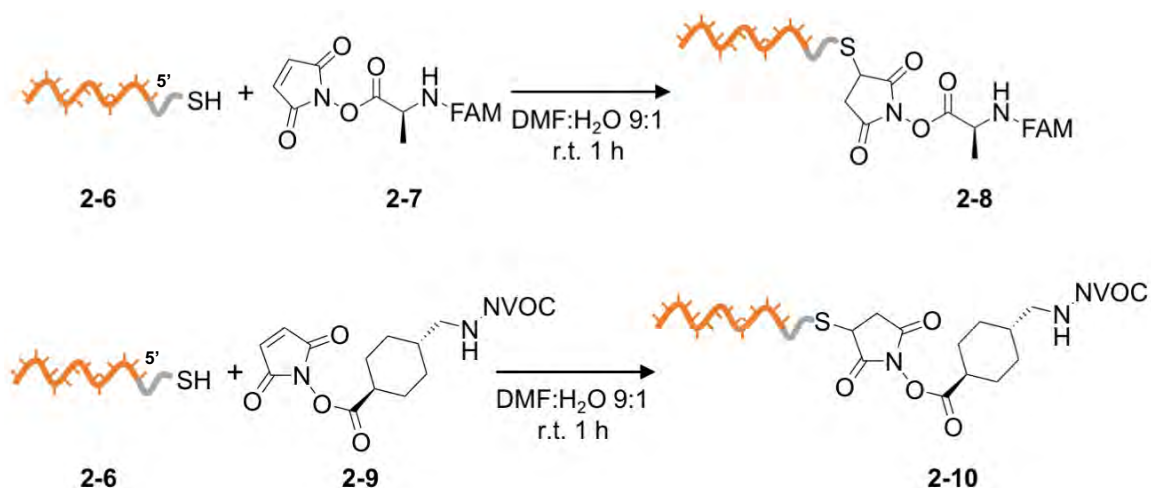
To conclude, using a small molecule model, it was shown that DESs provide better stability for activated esters than neutral aqueous PBS solutions. Glycholine provided higher stability to activated esters than reline, which correlates to the basic pH of reline that promotes basic hydrolysis; however, it remains unclear why the stability in ethaline was lower than in glycholine. The acidity of DESs can be corrected by the addition of bases such as Et₃N and NaOAc, and the addition of Et₃N has a smaller negative impact on the stability of activated esters than NaOAc while producing the same pH in diluted DES / aqueous solution. Finally, it has been shown that the synthesis of peptide bonds using small molecules proceeds in moderate yield in hydrated glycholine DES.

2.2.2 DNA activated ester conjugate stability in DES

The direct analysis of the stability of DNA conjugated esters in DESs represented a considerable challenge as these compounds can only be analysed in very small concentrations. In addition, most analytical methods used with DNA are optimised in aqueous conditions (*e.g.* RP-HPLC and gel electrophoresis) and this could facilitate hydrolysis during the analysis. However, the positive results using a small molecule model were encouraging enough to proceed into designing an assay to determine the stability of DNA conjugated activated esters.

While the DTS of peptide bonds mediated by activated esters has already been reported,⁴⁹⁻⁵¹ the key limitation is that in previous reports the amino acid building blocks had to be chosen so that their activated esters did not undergo rapid hydrolysis in aqueous solution. The reported amino acids previously used in DTS (Scheme 2-4) have very different features compared to many common natural amino acids. Steric hindrance as well as conjugation seems to favour longer half-life in solution, but some of the most common amino acids such as glycine, alanine or lysine have a lower steric hindrance in the vicinity of the ester group, and no conjugation, therefore their activated esters rapidly decompose.⁵³

In order to demonstrate this difference between unnatural amino acids previously used for DTS, and a natural amino acid such as alanine, the synthesis of the DNA conjugated esters **2-8** and **2-10** was performed (Scheme 2-4 and Experimental Section 2.4.8) using a *N*-hydroxymaleimide ester small molecule and a thiol-labelled DNA, using the literature reported procedure.⁵⁰



Scheme 2-4. Synthesis of the DNA conjugated activated esters **2-8** and **2-10** through thiol-Michael addition reaction. FAM = fluorescein. NVOC = 6-nitroveratryl.

The DNA conjugates were isolated, transferred into aqueous solution minimising the time in solution as much as possible and immediately analysed by LC-MS (Figure 2-6). The LC-MS results showed that while the vast majority of the alanine activated ester (Figure 2-6 a) was hydrolysed, the tranexamic acid activated ester conjugate (Figure 2-6 b) reported in DTS showed greater stability. This experiment highlights that with such a low stability in aqueous solution, it would not be possible to perform the long multistep DTS of natural peptides.

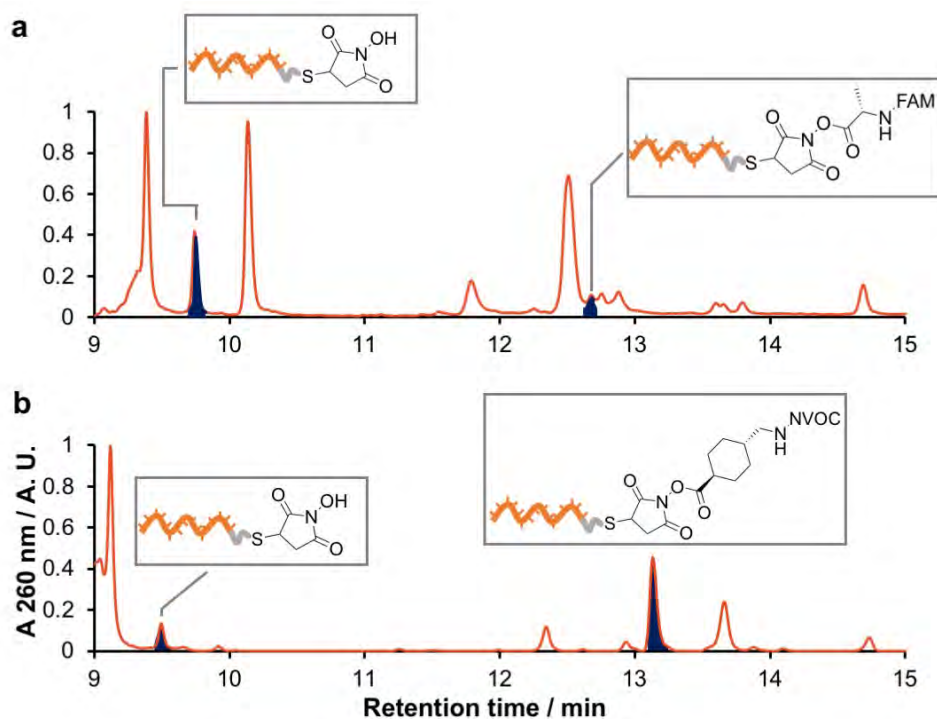


Figure 2-6. LC-MS analysis of DNA conjugated activated esters using high aqueous mobile phase in RP-UHPLC. a) DNA conjugated FAM-labelled alanine activated ester. LC-MS characterisation (Waters Xevo-G2-XS): DNA-SuOH $[M]^0$ m/z calcd. 9558.632 found 9558.557 DNA-SuO-Ala-FAM $[M]^0$ m/z calcd. 9943.727 found 9943.557 b) DNA conjugated tranexamic acid activated ester. LC-MS characterisation product $[M]^0$ m/z calcd. 9936.775 found 9936.685. Unlabelled peaks correspond to small molecule residues (based on HRMS) remaining due to the shorter aqueous DNA isolation protocol to prevent further hydrolysis.

In order to quantify the stability difference between DNA conjugated activated esters of alanine in aqueous and DES solutions, an assay based on Förster resonance energy transfer (FRET) was designed (Figure 2-7). FRET is a phenomenon where the photons emitted by a fluorophore, which is known as the donor, are absorbed by an acceptor group due to the spectral overlap of the donor emission and the acceptor absorption, and the close spatial co-localisation of both groups.⁶³ The FRET intensity depends on the inverse of the sixth power of the distance between the donor and the acceptor, which makes it extremely sensitive to distance changes, for that reason

FRET has been used in the past to determine conformational changes in biomacromolecules,⁶³ and DNA circuits⁶⁴ among others.

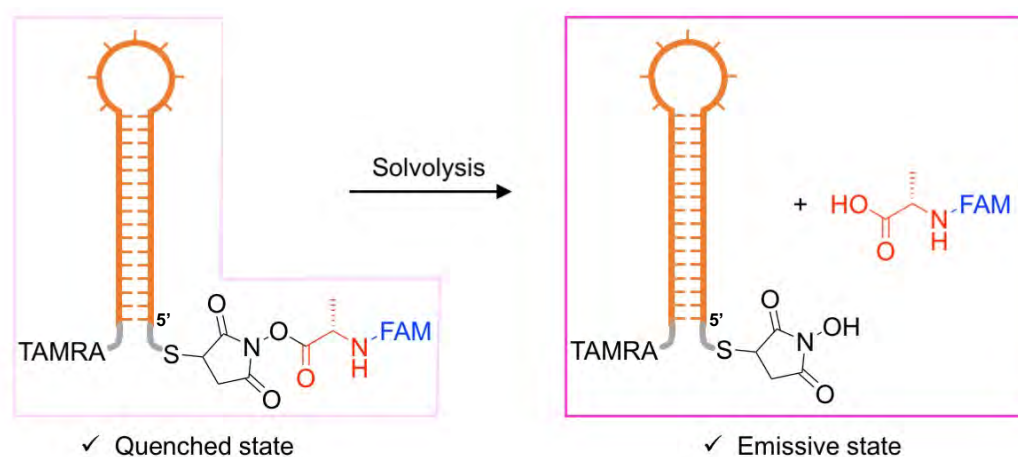
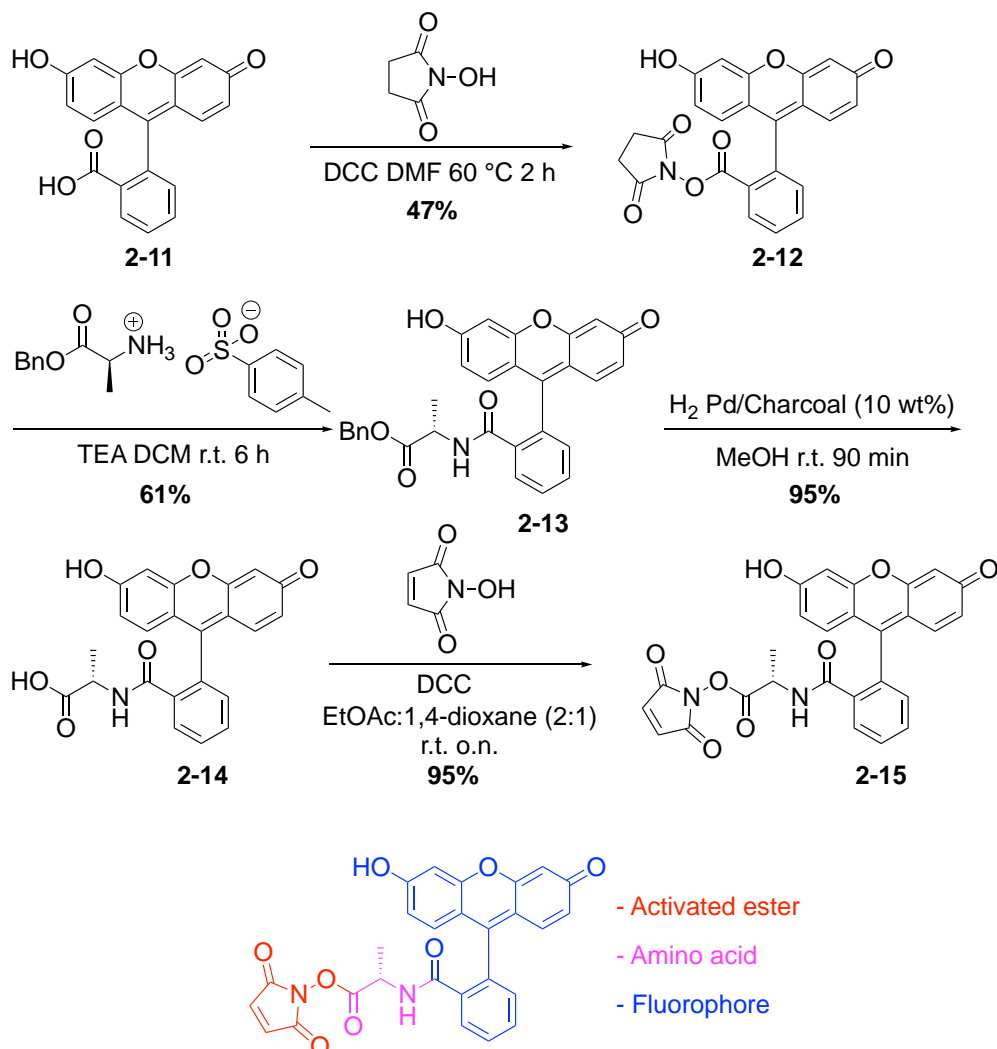


Figure 2-7. Fluorescence resonance energy transfer (FRET) based assay to follow the solvolysis of a DNA conjugated activated ester in DESs and aqueous solution. Representation of the DNA hairpin system containing a fluorescein-labelled (FAM, donor) alanine activated ester in the 5' end of helix, and a 6-carboxytetramethylrhodamine (TAMRA, acceptor) in the 3' end of helix. As the solvolysis progresses, the system evolves from a highly quenched state (left) to a non-quenched state (right).

In this assay (Figure 2-7), the alanine activated ester conjugated to the 5' DNA end was labelled with a fluorescein donor in the *N* position (FAM), while the 3' end of the helix contained a 6-carboxytetramethylrhodamine (TAMRA) acceptor. As the solvolysis of the activated ester happens, the resulting hydrolysed small molecule will diffuse away from the DNA hairpin, therefore reducing the efficiency of the FRET and causing a measurable difference in the fluorescence intensity. This assay would allow the solvolysis of a DNA conjugated activated ester to be followed over time without involving transfer into aqueous solution. FAM and TAMRA are a well-studied FRET pair. Additionally, FAM derivatives have a moderate cost compared to many fluorophores therefore increasing the feasibility of scale-up, and there were available

procedures to functionalise the carboxylic acid functionality with amines that could be adapted to the current purpose (Scheme 2-5).⁶⁵



Scheme 2-5. Synthesis of the FAM labelled L-alanine N-hydroxymaleimide ester (**2-15**)

A *N*-hydroxymaleimide ester of FAM labelled *L*-alanine ester that could be conjugated to DNA through Thiol-Michael coupling was devised (Scheme 2-5, **2-15**). The simplest possible synthesis of **2-15** consists of the in-situ activation of the carboxylic acid (**2-11**) by dicyclohexylcarbodiimide (DCC) followed by the addition of *L*-alanine: however, only a small fraction of the product **2-15** was detected, most likely due to the potential side

reactions such as alanine-alanine couplings, fluorescein-fluorescein anhydride formation and the low solubility of zwitterionic alanine in apolar solvent.

A step-wise synthesis was devised (Scheme 2-5, Experimental Section 2.4.9). The *N*-hydroxysuccinimide ester of fluorescein, **2-12**, was prepared and isolated by column chromatography according to a reported procedure.⁶⁵ **2-12** was reacted with a benzyl protected *L*-alanine *p*-toluenesulfonate salt. The benzyl protecting group was removed by catalytic hydrogenation. Finally, the free carboxylic acid (**2-14**) was activated by DCC and subsequently reacted with *N*-hydroxysuccinimide to yield the final product **2-15**. This molecule had all the necessary structural components required for the desired function: The *N*-hydroxymaleimide group which is a precursor to an *N*-hydroxysuccinimide activated ester, the alanine amino acid and the fluorescein fluorophore.

The spectral overlap in Gly_{0.07} between compound **2-15** and TAMRA was ensured by fluorescence spectroscopy (Experimental Section 2.4.10) showing overlap between the FAM fluorescence emission and the TAMRA absorption band. In a conventional FRET experiment, the sample is irradiated in the maximum absorbance wavelength of the acceptor, and the conjugated emission of the donor is recorded, however in this case the acceptor (FAM) also partially emitted in the donor (TAMRA) emission wavelengths. To circumvent this problem, the setup was slightly changed to a quenching experiment in which the reduction on the emission of the acceptor (FAM) was determined.

Usually, the reaction between thiol-labelled DNA and compound **2-15** to produce the conjugate described in Figure 2-7 would be performed in a non-hydrolysing solvent such as DES or a DMF:H₂O mixture, followed by precipitation, ultrafiltration, size

exclusion or RP-HPLC isolation of the desired product from the excess of reagent. Unfortunately, the previous isolation techniques require transfer of the product into a highly aqueous solution; as it was demonstrated in previous activated ester stability studies, this would result in hydrolysis of the product. To circumvent this issue, here the conjugation of **2-15** to DNA was performed on the surface of magnetic beads in glycholine solution, this approach had the advantage that the excess of small molecules could be removed by accumulating the beads on the side of a microcentrifuge tube, discarding the supernatant containing small molecules, and subsequently transferring the beads into fresh glycholine without requiring contact with aqueous solution at any point. In order to release the labelled DNA conjugate from the beads into the DES solution, an indirect immobilisation followed by toehold displacement was devised: a biotin-labelled ssDNA fully complementary to the DNA conjugate was annealed to the thiol-labelled DNA, and the dsDNA was immobilised on streptavidin coated magnetic nanoparticles (Dynabeads);⁶⁶ biotin-streptavidin immobilisation is based on a highly resilient supramolecular interaction that could be exploited in DES. The conjugate was then released from the magnetic nanoparticles surface through a toehold mediated strand displacement. This approach relying only on DNA-DNA supramolecular interactions to immobilise and then release the conjugate had the advantage that it did not require the addition of small molecules such as an excess of biotin to streptavidin, imidazole (if a His tag had been used) or to use low wavelength UV radiation on photocleavable linkers, that can cause moderate DNA damage (Figure 2-8, Experimental section 2.4.10).

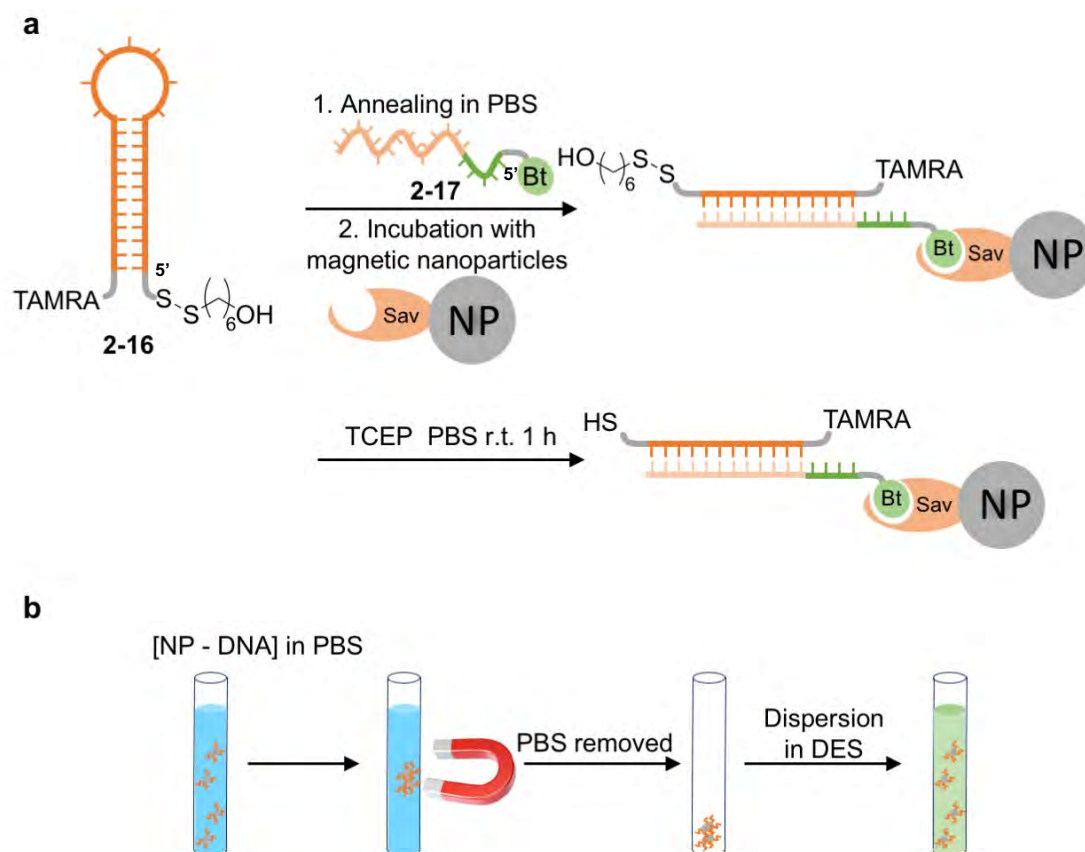


Figure 2-8. Procedure to transfer a thiol and TAMRA labelled DNA - magnetic nanoparticle (NP) assembly from PBS solution into DES. a) Process to open a disulfide and TAMRA labelled hairpin by annealing to a fully complementary biotinylated DNA strand attached to magnetic nanoparticles. The reduction of the disulfide group into a thiol was followed by conjugation. b) The nanoparticles in PBS solution were then magnetically accumulated on the side of a microcentrifuge tube, the solvent was removed while the nanoparticles were not disturbed due to the magnetic interactions, and finally re-dispersed in DES solution.

The hairpin containing the TAMRA acceptor group and a disulfide (**2-16**), was thermally annealed in PBS with a biotinylated fully complementary DNA strand (**2-17**) (Figure 2-8 a). The disulfide was then reduced to a thiol by the addition of TCEP and the dimer was conjugated to streptavidin coated magnetic nanoparticles (*Dynabeads*). Finally, using a strong neodymium magnet, the nanoparticles were magnetically accumulated on the side of a microcentrifuge tube, the solvent was removed using a micropipette and replaced with fresh Gly_{0.07} solvent (Figure 2-8 b).

Once the DNA-coated nanoparticles had been transferred to DES, they were reacted with the *N*-hydroxymaleimide ester **2-15** (Figure 2-9 a), the excess of reagent was removed by magnetically accumulating the nanoparticles and adding fresh glycholine DES. Finally, the hairpin was released by toehold mediated DNA strand displacement⁶⁷ with a fully complementary strand (**2-18**). In order to isolate the hairpin **2-19** in Gly_{0.07}, the nanoparticles were magnetically concentrated on the side of the microcentrifuge tube and the DESs solution was transfer to a clean vial, that solution constituted the stock, which was in turn diluted to the final analysis conditions (Figure 2-9 b).

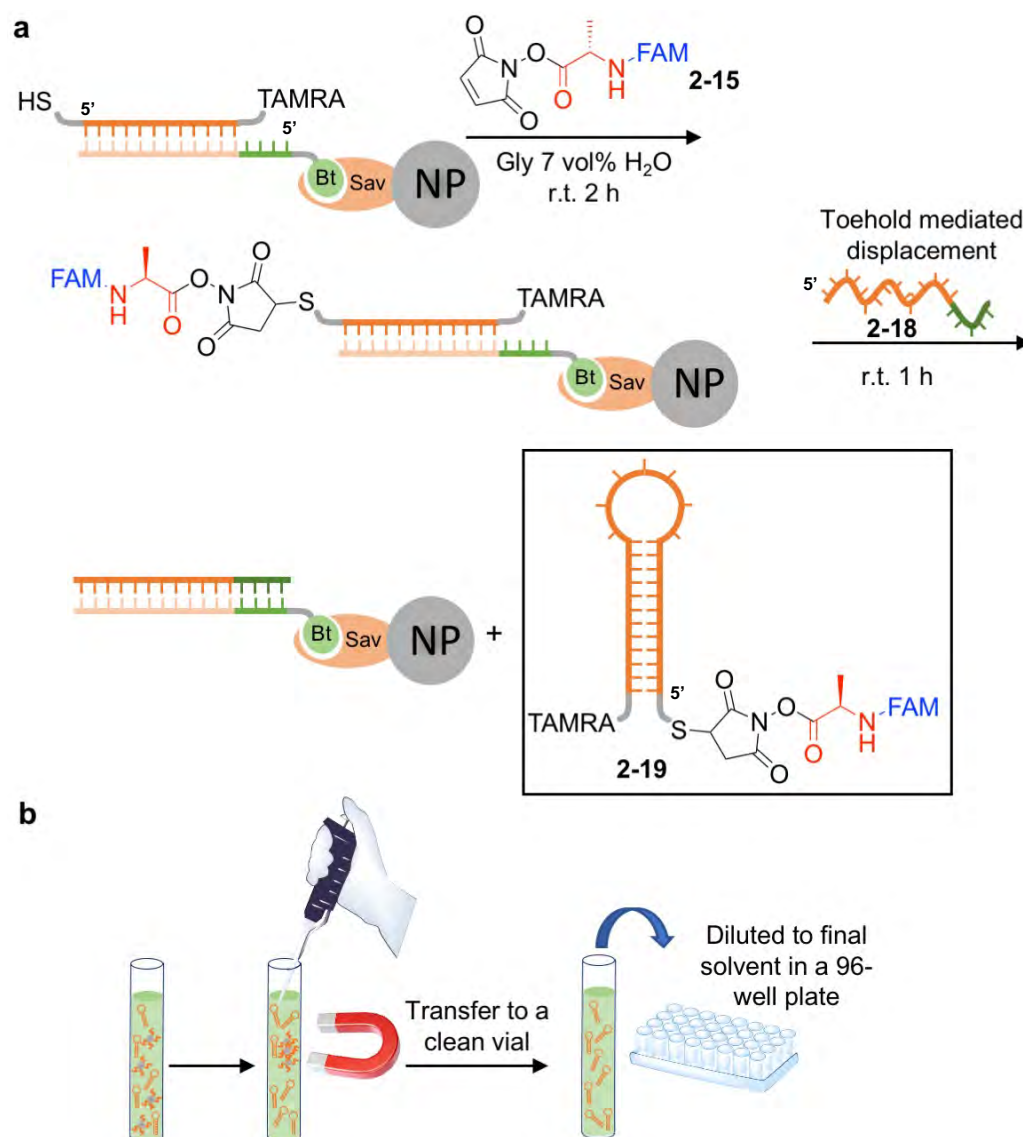


Figure 2-9. Synthesis of a DNA conjugated activated ester hairpin (**2-19**) and transfer to fresh DES. a) Formation of the N-hydroxysuccinimide ester while the DNA strand is conjugated to the nanoparticles, transfer to fresh DES followed by toehold mediated release of the hairpin. b) Transfer of the hairpins to a fresh DES solution prior to dilution into the final solvent conditions.

The stock of the DNA conjugated activated ester hairpin was diluted 20-fold in the final solvent conditions and the sample was placed in a real-time PCR (or quantitative PCR) instrument at 25 °C. The fluorescence intensity of the fluorescein group was measured by triplicate every 3 min. Using this procedure, an initial test on the effect of increasing water content DESs was performed (Figure 2-10). According to the small molecule

test, it was anticipated that the higher water content DESs would promote faster hydrolysis than the lower ones.

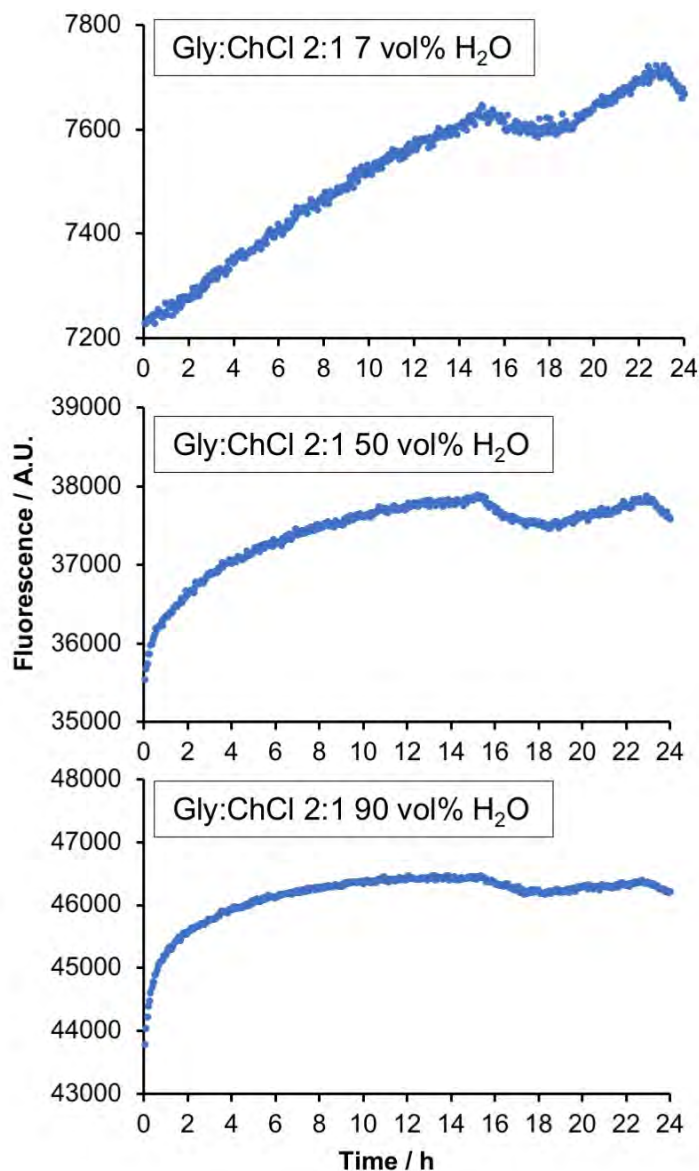


Figure 2-10. Raw data of the fluorescence assay of the stability of the DNA conjugated activated ester (**2-19**) in glycoline solutions (Gly : ChCl 2:1) with 7, 50 and 90 vol% H₂O. A fluctuation on the signal was observed after 16 h, this was attributed to an instrumental issue such as decrease in the light intensity of the bulb.

This first stability test showed that the fluorescence values in low H₂O (7 vol%) glycoline were much lower and changed less than in high H₂O (50 and 90 vol%)

glycholine, and the change in fluorescence over time, related to the progress of the hydrolysis, was slower in 7 vol% H₂O than in 50 and 90 vol% H₂O.

The raw fluorescence data can be correlated to the fraction of activated ester by eq. 1 (Experimental Section 2.4.11 for the derivatisation of eq. 1):

$$\chi_{Ae} = \frac{F_{Max} - F(t)}{F_{Max} - F_{Min}} \quad \text{Eq. 1}$$

Where χ_{Ae} is the fraction of activated ester, F_{Max} is the maximum fluorescence corresponding to the fully hydrolysed system; and F_{Min} is the minimum fluorescence corresponding to the non-hydrolysed system.

In acid and base-catalysed hydrolysis processes, as the reagents (H₂O, OH⁻ and H₃O⁺) are in a large excess compared to the substrate, their concentrations are approximately constant over the hydrolysis process and can be included into an apparent rate constant (k_{obs}), therefore the kinetics follow a pseudo-first order rate law.

$$[Ae](t) = [Ae]_0 \cdot e^{-k_{obs} \cdot t} \quad \text{Eq. 2}$$

Eq. 2 can be re-arranged and combined with eq. 1 to express the evolution of fluorescence over time.

$$F(t) = F_{Max} - e^{-k_{obs}t}(F_{Max} - F_{Min}) \quad \text{Eq. 3}$$

Qualitatively, the representation of eq. 13 as function of time shows a shape as that of eq. 4 (Figure 2-11).

$$y = 1 - e^{-x} \quad \text{Eq. 4}$$

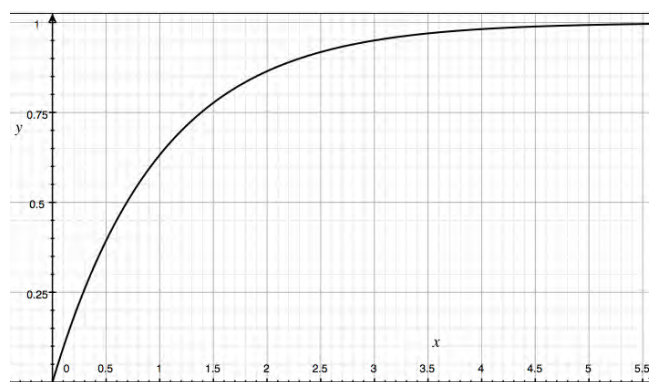


Figure 2-11. Generic representation of an equation of the type eq. 3 and eq. 4.

It can be seen that the evolution of fluorescence for samples where hydrolysis is rapid (glycholine 50 and 90 vol% H₂O) in Figure 2-10 matches well the expected shape, thus reassuring that the observed trend corresponds to hydrolysis. It is worth noting that the experimental data in Gly_{0.07} seems to show a linear increase, this most likely caused by a comparatively slow hydrolysis rate; therefore, only a small fraction of the total process is observed.

Critically, converting the fluorescence raw data into meaningful kinetic data, requires a good determination of the F_{Max} and F_{Min} values. In solutions where the hydrolysis was slow, such as in Gly_{0.07}, the fluorescence corresponding to complete hydrolysis (*i.e.* F_{Max}) was not reached, in solutions where the hydrolysis was very fast, the initial fluorescence (*i.e.* F_{Min}) could not be confidently determined as the hydrolysis could have progressed a lot in the few minutes before introducing the samples in the PCR instrument and starting the measurements. This was worsened by the fact that, due to the viscosity of DESs, the well-plates containing the samples had to be centrifuged before starting the measurements in order to obtain a stable signal. These limitations are discussed below (Figure 2-12).

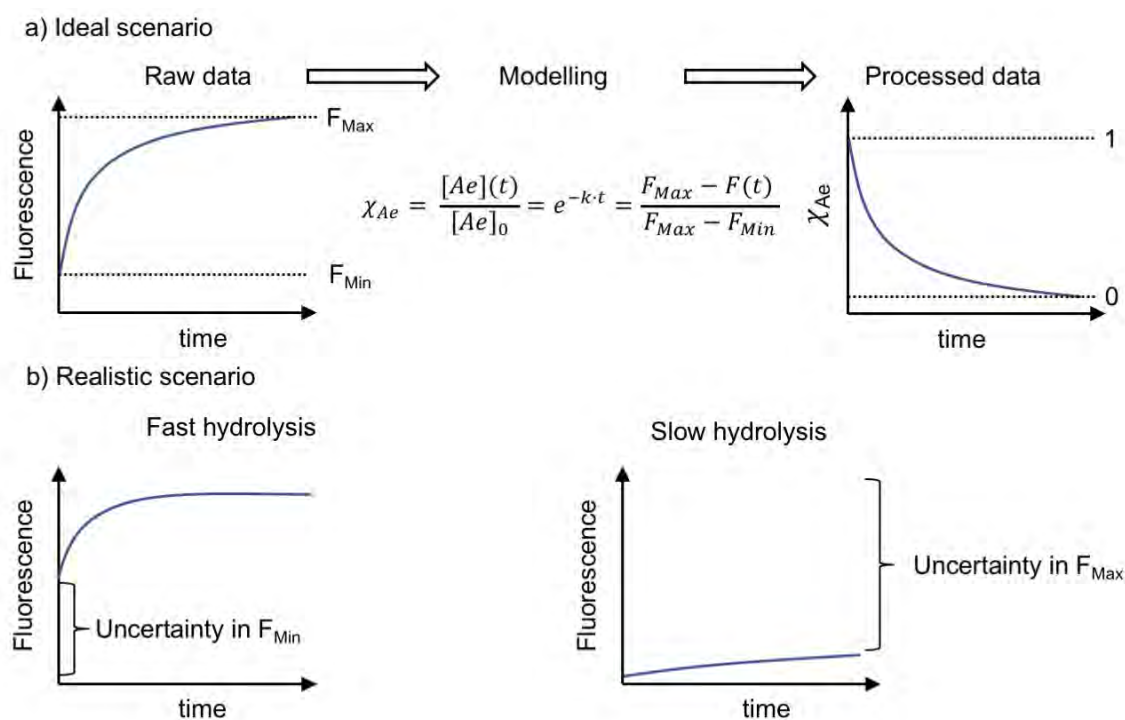


Figure 2-12. Description of the data from the fluorescence stability assay in DES. a) Description of the ideal data, where the minimum fluorescence (at time 0) and the maximum fluorescence (ideally at infinite time) are precisely determined. A pseudo-first order hydrolysis rate law could be applied, and the kinetic parameters determined. b) Realistic data where, for fast hydrolysis processes the initial fluorescence cannot be determined due to hydrolysis before the measurement started and in slow hydrolysis the maximum fluorescence is not achieved.

It is worth clarifying that, due to changes in the quantum yield of fluorophores in different DES / H₂O mixtures, the F_{Max} and F_{Min} values were individual for each sample, which prevented simply determining F_{Min} in a slow hydrolysis solution and F_{Max} in a fast hydrolysis solution and apply these values to all samples. This major inconvenience prevented eq.1 and eq. 3 being applied and therefore a rigorous kinetic analysis. It could only be qualitatively concluded that, due to the slower change in fluorescence for 7 vol% H₂O solution compared to 50 and 90 vol%, it seems that indeed the lower water content in DES allowed for a higher stability of the activated ester.

In order to determine if the fluorescence in 7 vol% H₂O solution would change upon full hydrolysis, parallel experiments using 7 vol% of a buffer that accelerated the

solvolysis of **2-19** were performed (Figure 2-13). This buffer consisted of a solution containing 10 mM Tris (Tris·HCl pH 7.5) primary amine, which performed a rapid aminolysis of the ester bond, thus allowing the change in fluorescence for a glycholine 7 vol% aqueous solution to be observed.

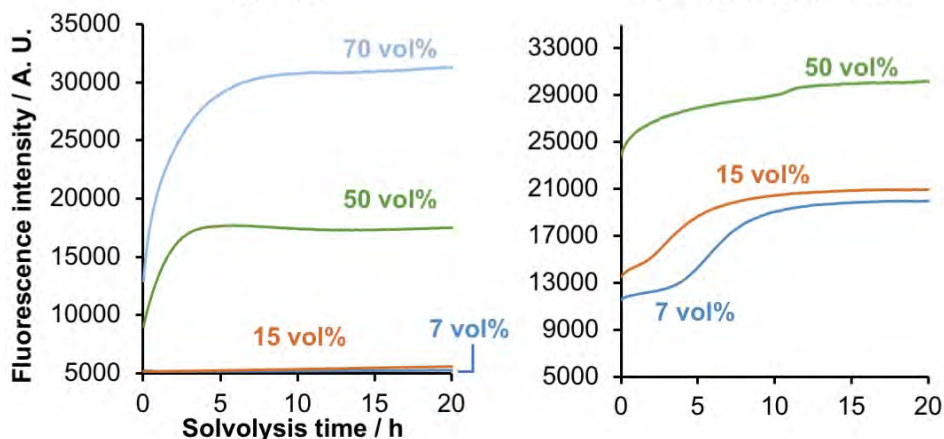


Figure 2-13. Comparison between the slow hydrolysis in glycholine 7 and 15 vol% H₂O (left), and fast aminolysis in 7 and 15 vol% buffer (right). The aminolysis buffer contained: 10 mM Tris·HCl and the pH was adjusted to 7.5.

By performing an aminolysis of the activated ester, it can be seen that the fluorescence signal for both 7 and 15 vol% samples increases drastically. While it was still not possible to determine the kinetic parameters described in eq. 3, as the F_{Max} in aminolysis conditions is not necessary the exact value for the slow hydrolysis conditions, it allowed to reiterate the qualitative argument that in 7 and 15 vol% H₂O / glycholine solutions the hydrolysis progress was very slow. Further stability tests were attempted in glycholine / Et₃N. The addition of base resulted in a fast increase of the fluorescence, indicating a rapid hydrolysis, in good agreement with the small molecule stability trends (Experimental section 2.7.10 and Figure 2-26).

A strand displacement strategy was attempted to determine F_{Max} (Figure 2-14).

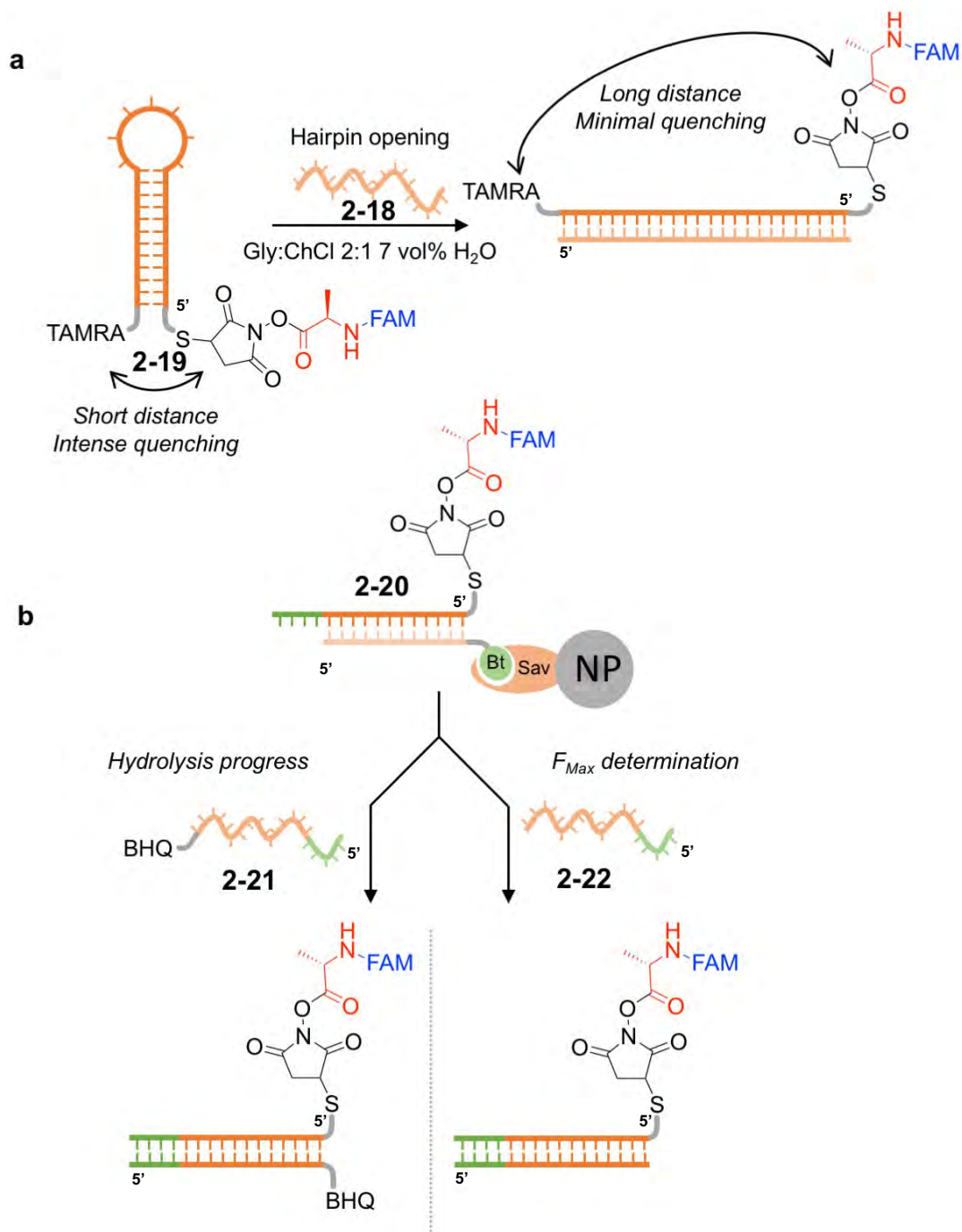


Figure 2-14. Attempts to determine F_{Max} through strand displacement strategies. a) Blunt-end⁶⁸ hairpin opening in order to reduce the quenching effect between FAM and TAMRA. b) Parallel assay using one aliquot of the DNA conjugated activated ester to determine the hydrolysis progress, using a complementary quencher strand (BHQ = black hole quencher) and using another aliquot to determine the maximum fluorescence in the absence of quenching.

This strategy was based on approximating the value of F_{Max} by the fluorescence values of similar systems. The first approach (Figure 2-14 a) was to open the hairpin using a fully complementary DNA strand, in this case the resulting duplex would be 37 base-pairs (bp), and therefore approximately 12.6 nm⁶⁹ plus linkers. The increased distance between dye and quencher should reduce the efficiency of the quenching therefore displaying higher fluorescence. Unfortunately, the fluorescence did not substantially increase compared to the quenched system (Figure 2-15 a), which was hypothesised to be due to the slow kinetics of blunt-end displacement⁶⁸ worsened by the large viscosity of DESs.

In a second attempt (Figure 2-14 b), the approach was based on splitting the DNA conjugated activated ester into two aliquots. One of the aliquots was annealed to a complementary DNA labelled with a Black Hole Quencher, which has the ability of absorbing emitted radiation across a wide range of wavelengths and then relaxing to the ground state non-radiatively. This would result in a quenched system that would evolve into a non-quenched one as hydrolysis progresses. The second aliquot was annealed to a complementary DNA strand with no quencher; therefore, the maximum fluorescence should be detected from time 0. The strand displacement kinetics were expected to be much faster due to the toehold mediated displacement, as it has been demonstrated that a relatively short 4 bases toehold provides an acceleration of a 10^4 factor.⁷⁰ Perhaps unsurprisingly after the multiple failed attempts to determine the hydrolysis kinetics in low water content DES, this second approach did not provide the intended definitive data (Figure 2-15 b). It can be seen that in all solutions, the fluorescence of the sample containing the black hole quencher was lower than the non-quenched solution, and even in 90 vol% H₂O, where the hydrolysis was expected

to be fast, the quenched solution did not reach the fluorescence emission intensity of the non-quenched samples.

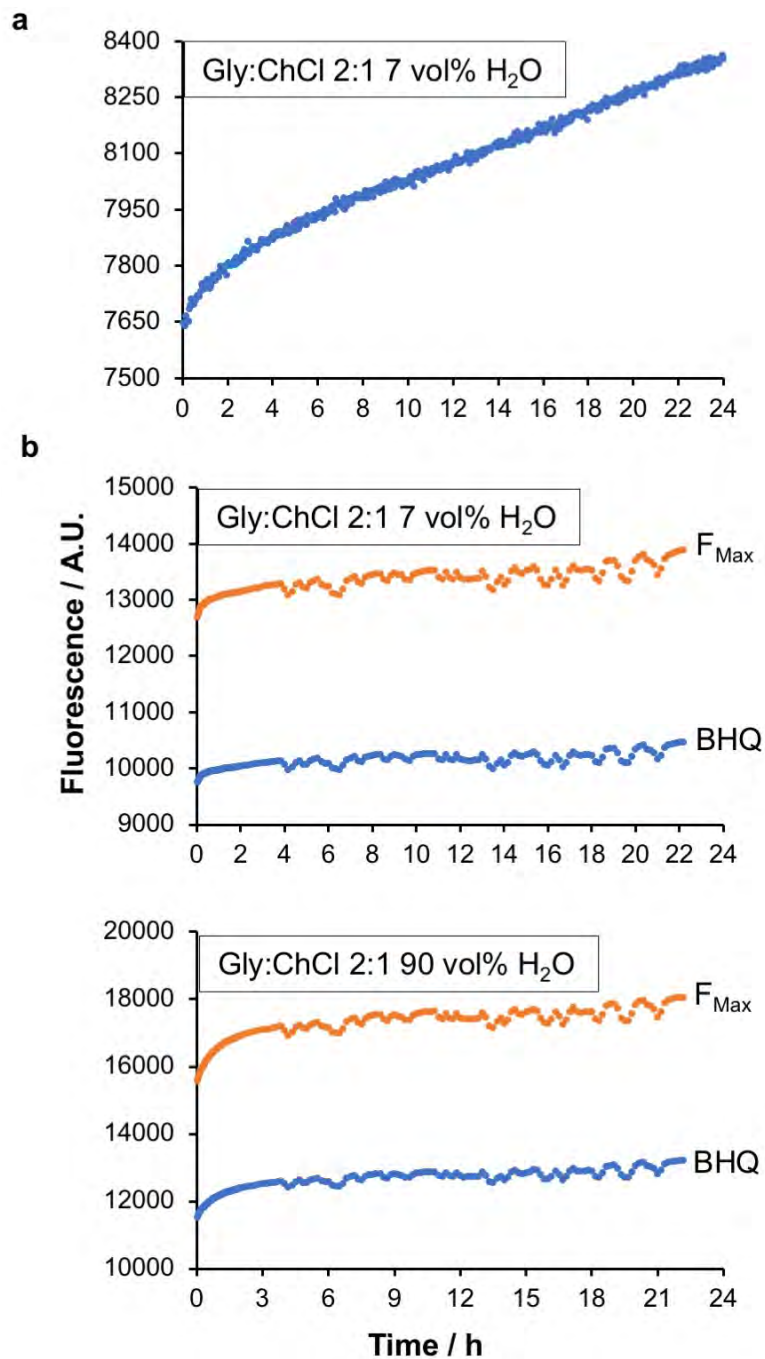


Figure 2-15. Results of the attempts to determine the value of F_{Max} a) Blunt-end mediated hairpin opening. b) Quenched and unquenched dsDNA.

In conclusion, these set of experiments oriented to quantify the stability of DNA conjugated activated esters using a fluorescence quenching phenomenon did not afford the quantitative kinetic data expected. However, the extremely slow change in fluorescence over time that was observed in glycholine with 7 and 15 vol% H₂O compared to 50 or 90 vol% H₂O or in the presence of aminolysis buffer strongly suggest that DESs indeed allow a much larger stability of esters than aqueous solution.

2.2.3 DNA templated peptide bond formation in DES

Prior to attempting a DTS of peptides in DES, a test to quantify the thermal stability of the DNA sequences in DES was carried out. The thermal stability of dsDNA and dsRNA is usually quantified using the melting temperature (T_m) of the duplex, defined as the temperature at which half of the possible duplexes are formed. From the several available methods to determine the T_m , the fluorescence quenching fluorescence approach was employed, using commercially available DNA strands (Figure 2-16 a).

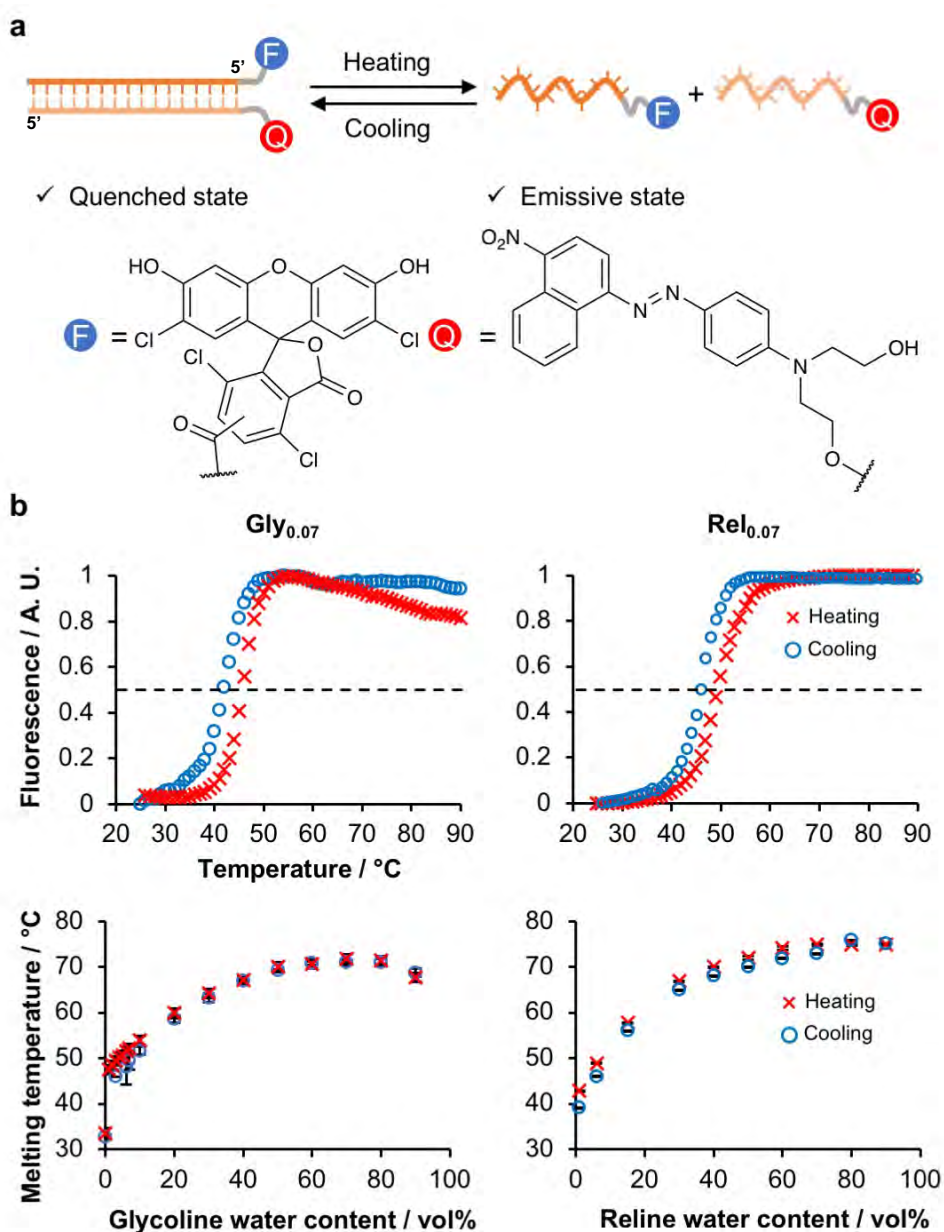


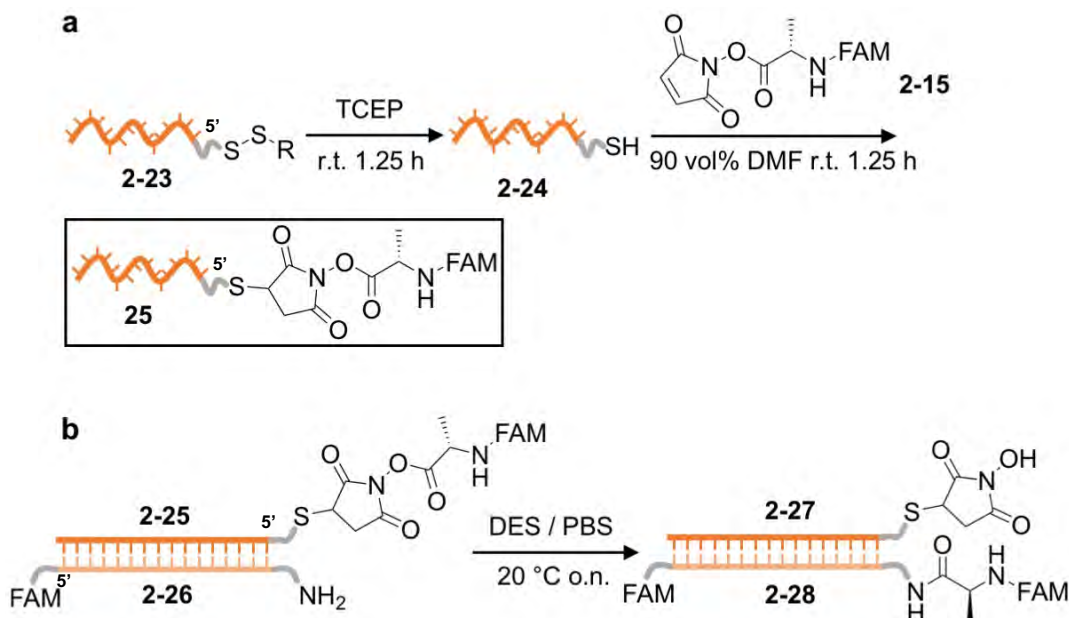
Figure 2-16. Evaluation of the stability of dsDNA in several DES solvents using fluorescence quenching. a) Representation of progression towards an unquenched state by thermal denaturalisation. b) Two examples of the fluorescence change as function of temperature: Gly_{0.07} and Rel_{0.07} (reline 7 vol% H₂O). The 50% fluorescence (dashed line) determines the T_m (heating = red cross, cooling = blue circle). Evolution of the T_m as function of the water content in glycoline and reline (heating = red cross, cooling = blue circle). Measurements were made in triplicate. The error bars correspond to the standard deviation. Sequences **F2-16D** and **F2-16Q** (Experimental section 2.4.12).

It can be seen that the T_m upon heating is approximately 2 to 3 °C higher than the T_m upon cooling (Figure 2-16 b) in low-water content DES. This hysteresis effect was

attributed to the high viscosity of the DES; however, this effect was moderate and did not affect the reproducibility of the T_m measurements over several $1\text{ }^\circ\text{C} \cdot \text{min}^{-1}$ heating analyses. A similar effect was found on the kinetics of DNA origami folding in anhydrous and hydrated DES.⁴⁵ The effect of the addition of water to DES has not been widely studied, the only precedent was the very modest increase in the T_m of a 32 bp dsDNA from 52 to 53 $^\circ\text{C}$ upon the addition of 10 weight% of H_2O .⁴⁵ In a broader study, here it was found that the addition of water progressively increased the T_m , the maximum thermal stability of dsDNA was reached and then it progressively decreased (Figure 2-16 b - T_m in glycholine) or reached a plateau (Figure 2-16 b - T_m in reline) before a sharp decrease in the T_m (note that the T_m in $>90\text{ vol}\%$ H_2O could not be determined in glycholine or reline). The interactions that influence the formation of dsDNA are the electrostatic repulsion between negatively charged groups, hydrogen bonding, hydrophobic, π - π stacking and van der Waals interactions.⁷¹ Based on that, the formation of dsDNA in low H_2O DES was likely facilitated by the charge screening produced by choline cations, that reduced the inter-strand electrostatic repulsion,⁷² then, the addition of water further stabilised the double-strand through hydrophobic effects, however; while the increase of water content favoured hydrophobic interactions it also reduced the concentration of choline cations and caused the T_m reduction observed in reline. In summary, the thermal stability of a 24 bp dsDNA was 48.0 $^\circ\text{C}$ in Gly_{0.07}. This stability was considered adequate as at room temperature most DNA would be in the duplex form.

The potential of DESs as a solvent medium to allow DTS of peptide bonds was tested using the same alanine activated ester synthesis described in Scheme 2-5. The activated

ester was then diluted in an excess of DES, followed by the addition of the complementary amine-DNA acceptor (Scheme 2-6).



Scheme 2-6. Scheme of the DTS of peptide bonds in DES. a) The synthesis of the N-hydroxysuccinimide adapters was performed by the TCEP mediated reduction of a disulfide-labelled DNA followed by thiol-Michael addition in 90 vol% DMF solution. b) The DNA templated transfer reaction was subsequently performed in several solvent and base additive conditions through the annealing of activated ester donor, and the amine acceptor strands. The templated reaction was performed at 2 μM .

The previous procedure relied on diluting a DMF solution in Gly_{0.07}. As DMF is a well-known dsDNA denaturing agent, the T_m of DNA in several Gly_{0.07} / DMF solutions was determined. It was found that as long as the content of DMF remained <1 vol%, the effect on the T_m was very modest (Experimental Section 2.4.12)

The reaction solutions in Gly_{0.07} were then transferred into nanopure water and analysed by denaturing RP-HPLC and LC-MS (Figure 2-17, Experimental Section 2.4.13). Initially, the idea was that only the alanine amino acid would be fluorescently labelled, therefore only the amide product (**2-28**) would be fluorescent. However, before the LC-MS methodology was optimised, it was difficult to distinguish the starting material **2-25** from **2-26**. To overcome this the amine functionalised DNA was also

labelled with a fluorophore, therefore the amine starting material (**2-26**) and the potential product (**2-28**) appeared fluorescent in the HPLC fluorescence detector.

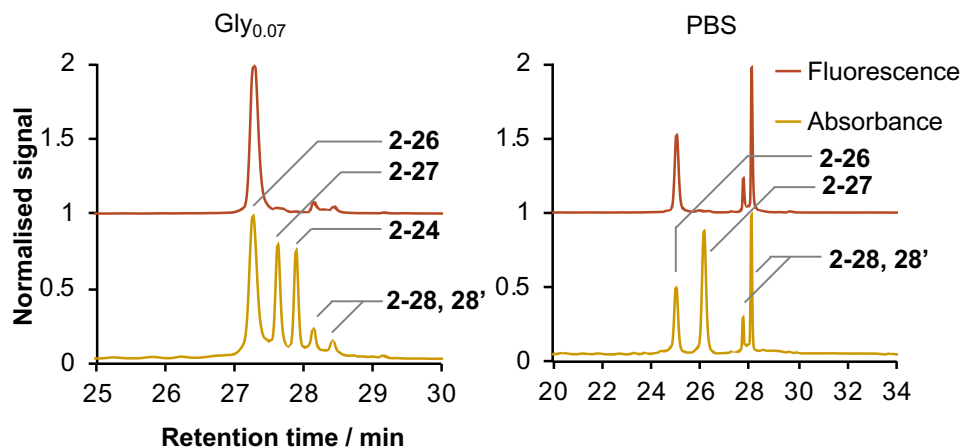


Figure 2-17. Comparison between a DNA templated peptide bond synthesis in Gly_{0.07} and in aqueous buffered solution. LC-MS (Bruker AmazonX): **2-26** [M-H]⁻ *m/z* calcd. 9337.2 found 9336.0; **2-27** [M-H]⁻ *m/z* calcd. 10712.0 found 10712.4 **2-24** [M-H]⁻ *m/z* calcd. 10615.2 found 10599.6 [M-OH]⁻; **2-28, 2-28'** (isomers) [M-H]⁻ *m/z* calcd. 9723.6 found 9719.9. Absorbance was determined at 260 nm, fluorescence excitation 495 nm and emission 520 nm.

As a positive control of the templated reaction, after the synthesis of the activated ester (**2-25**) in 90 vol% DMF, was dissolved in PBS solution (0.1 M, pH 7.5) containing the complementary amine-labelled DNA (**2-26**). Under these conditions, a substantial amount of product (**2-28**) was detected by both HPLC and LC-MS (48% based on peak area HPLC, Figure 2-17). It is important to note that the DNA templated synthesis of **2-28** in aqueous buffer can only be achieved if the activated ester was immediately annealed to the complementary amine-labelled DNA, otherwise the ester undergoes fast hydrolysis, and as a consequence, this approach does not allow multistep synthesis or storage of the adapter even for short periods of time. Using DES on the other hand, only produced a very moderate amount of the product (**2-28**) as it can be seen in the chromatogram (<5% conversion based on peak area, Figure 2-17). In order

to investigate the cause of this low yield in DES, and to improve it, a set of additives and conditions were assayed (Table 2-3).

Table 2-3. Solvent additives and conditions under which DNA templated peptide bond formation was attempted.^a The reactive groups were placed across a nicked DNA duplex.⁷³

Entry	Solvent	Additives	Result
1	Glycholine	7 vol% H ₂ O	No product detected
2	Glycholine	20 vol% H ₂ O	No product detected
3	Glycholine	40 vol% H ₂ O	No product detected
4	Glycholine	7 vol% H ₂ O and 500 mM NaOAc	Product detected in very low yield (< 5%)
5	Glycholine	7 vol% H ₂ O and 100 mM Et ₃ N	Product detected in very low yield (< 5%)
6	Glycholine	7 vol% H ₂ O and 5 mM Et ₃ N	No product detected
7	Glycholine	7 vol% H ₂ O and 5 mM Et ₃ N ^a	Product detected in very low yield (< 5%)
8	Reline	7 vol% H ₂ O	No product detected

The DTS in glycholine / 7 to 40 vol% H₂O did not yield any product (Table 2-3 entries 1 to 3). Glycholine was very slightly acidic (Table 2-2), while the addition of high concentration of Et₃N and NaOAc did allow to detect a small amount of product, it did not allow to obtain the amide product (**2-28**) in more than a 5% yield (Table 2-4 entries 4-7). Reline, which showed to be basic itself (Table 2-2), did not allow to obtain a detectable amount of the product (Table 2-3 entry 8). Most transfer reactions were performed in an end-of-helix dsDNA design;⁵¹ however, an across-the-nick architecture⁷³ did not substantially improve the conversion.

The stability tests performed on small molecule activated ester by RP-HPLC and ¹H-NMR spectroscopy, combined with the qualitative data obtained from the fluorescence

stability assay performed on DNA conjugated activated esters, all point in the direction that activated esters have a higher stability in glycholine than in aqueous solution. However, the addition of base seemed to accelerate the solvolysis process. It was hypothesised that there was a lack of reactivity of the amine as a nucleophile. In order to test this, an experiment where an amine-labelled DNA strand was dissolved in 10 mM small molecule activated ester solution was performed (Figure 2-18). A series of solutions with increasing concentrations of Et₃N in 1 M aqueous NaCl and glycholine were prepared. The DNA strand was shaken overnight and then the conversion to an amide was quantified by HPLC and LC-MS (Experimental Section 2.4.14).

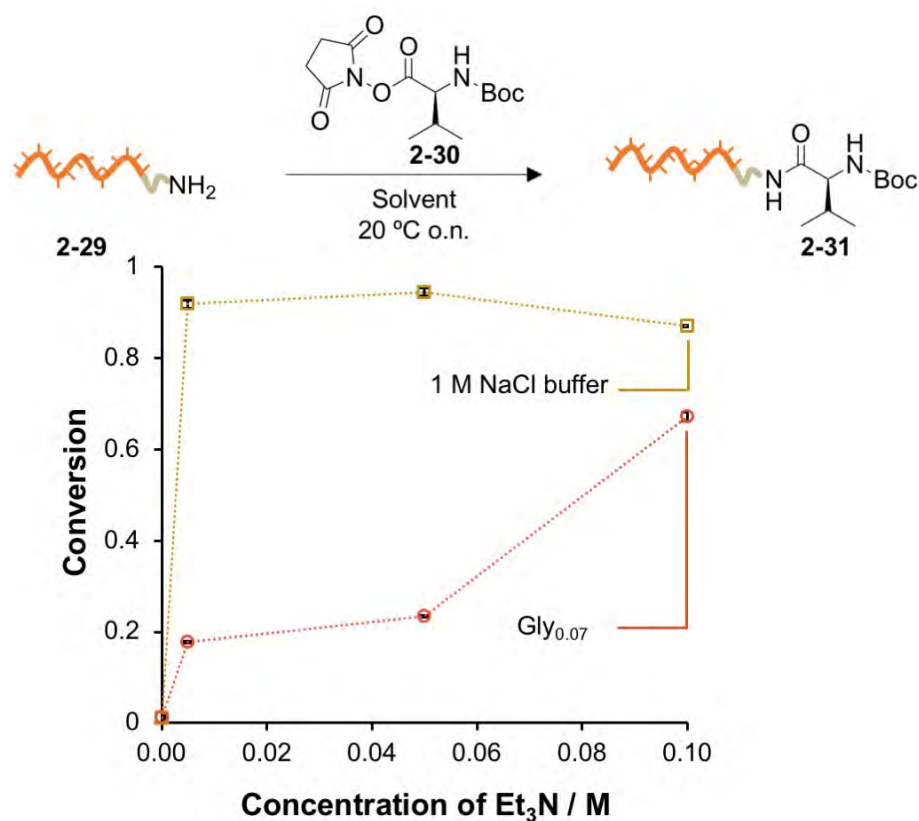


Figure 2-18. Amine reactivity test in 1 M NaCl aqueous solution and in DES. LC-MS (Bruker AmazonX): 2-29 [M-H]⁻ m/z calcd. 3897.7 found 3898.6; 2-31 [M-H]⁻ m/z calcd. 4096.8 found 4097.7. Error bars correspond to the standard deviation on triplicate integration events.

The results show how in the absence of base, there is almost no product formation in either 1 M NaCl (aq) and Gly_{0.07}. However, after the addition of a minimal volume of Et₃N, the conversion of amine-labelled DNA into an amide was nearly quantitative. On the other hand, the addition of large volumes of Et₃N produced a progressive and moderate increase on the conversion. This simple experiment highlights that while the stability of the activated ester in Gly_{0.07} was higher than in an aqueous solution, the reactivity of the amine was very low, and required the addition of a large excess of a tertiary amine to enhance its nucleophilicity. The stability experiments (Figures 2-3, 2-4 and 2-10) showed that the addition of Et₃N to Gly_{0.07} considerably reduced the stability of the activated ester. It can be therefore concluded that it is not possible to combine the high stability conditions and high reactivity in DES.

As a consequence of the previous considerations, the DTS of peptide bonds cannot happen in the studied DES systems. However, it would be intriguing to discover if DESs can play a role in the stabilisation of activated esters for storage purposes.

2.2.4 Storage of DNA conjugated activated esters in DES

It has been previously shown that DESs can facilitate the formation of activated organophosphates, which would be a route to obtain the necessary monomers for nonenzymatic replication of nucleic acids.^{74,75} The absence of water would reduce the solvolysis rate. Here, DNA conjugated activated esters were stored in Gly_{0.07}, and subsequently diluted in PBS containing the complementary amine-labelled DNA (Figure 2-19). If the activated ester is preserved in the DES, then a significant fraction of peptide product (**2-35**) should be formed, otherwise the major product would be hydrolysed to DNA-SuOH (**2-33**, Figure 2-19 a).

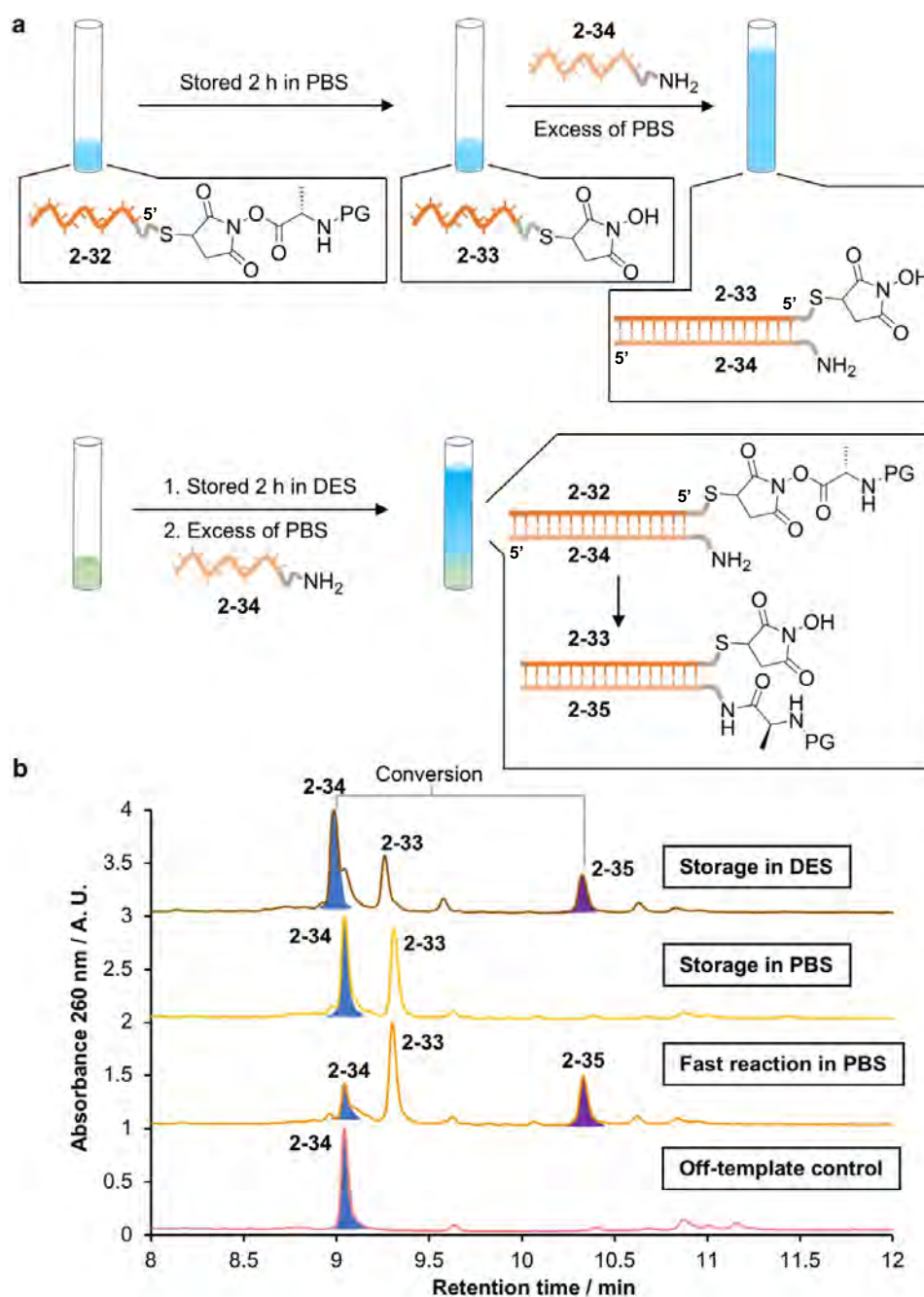


Figure 2-19. Storage of DNA conjugated activated esters in DES and PBS followed by DNA templated amidation reaction. a) Scheme of the chemical processes during the storage of the activated ester in PBS and in DES, followed by the DNA templated amidation. b) LC-MS analysis after the templated reaction. As a positive control, an aliquot of the activated ester was reacted with the complementary amine-labelled DNA immediately after its synthesis. The off-template reaction between **2-34** and the excess of activated ester was also tested. The conversion was determined from the LC peak area of **2-34** over the combined area of **2-35** and **2-34**. (Experimental section 2.4.15) LC-MS (Waters Xevo-G2-XS): **2-34** [M]⁰ *m/z* calcd. 10891.855 found 10891.522; **2-33** [M]⁰ *m/z* calcd. 9558.632 found 9558.581; **2-35** [M]⁰ *m/z* calcd. 11201.935 found 11201.489. DTS was performed at 0.67 μM.

The LC-MS analysis (Figure 2-19 b) demonstrates that: 1) If the activated ester is immediately reacted with the complementary amine-labelled DNA, a large fraction of the amine (**2-34**) is transformed into an amide (64%); 2) If the activated ester is stored in PBS, after 2 h the hydrolysis is complete and no amide product (**2-35**) is obtained; 3) If the activated ester is stored in Gly_{0.07} DES for 2 h, a considerable fraction (30%) is converted into the amide product. These results show that the compartmentalisation of the activated ester synthesis in a glycholine DES, followed by transfer into aqueous solution improves the efficiency of nonenzymatic DTS of peptide bonds.

In conclusion, while it has been demonstrated that the DTS of peptide bonds is hampered in DESs, even in the presence of varying amounts of water and basic additives, it has also been shown that DESs allow a better stability of *N*-hydroxysuccinimide activated esters. Together with the reported spontaneous formation of activated species in DES,⁷⁵ this suggests that DES could facilitate the nonenzymatic translation of genetic information into functional peptides.

2.2.5 Reductive amination in DES

In the previously attempted DTS, a relatively small chemical group was transferred from a DNA strand to its complementary, and therefore the two DNA strands remained bonded by weak interactions. It would be however interesting to establish a comparison with a templated chemical reaction that connects two hybridised strands by a covalent bond. To that effect, a reductive amination reaction was assayed in Gly_{0.07}. Reductive amination was an interesting comparison point with amidation because it also relies on the reactivity of an amine group. However, it is compatible with more acidic conditions.⁷⁶

Firstly, as reductive amination requires a considerable concentration of NaBH_3CN , the effect of this additive on the T_m was tested, showing that it did only cause a moderate decrease in the stability of dsDNA with low or moderate PBS content (Figure 2-20).

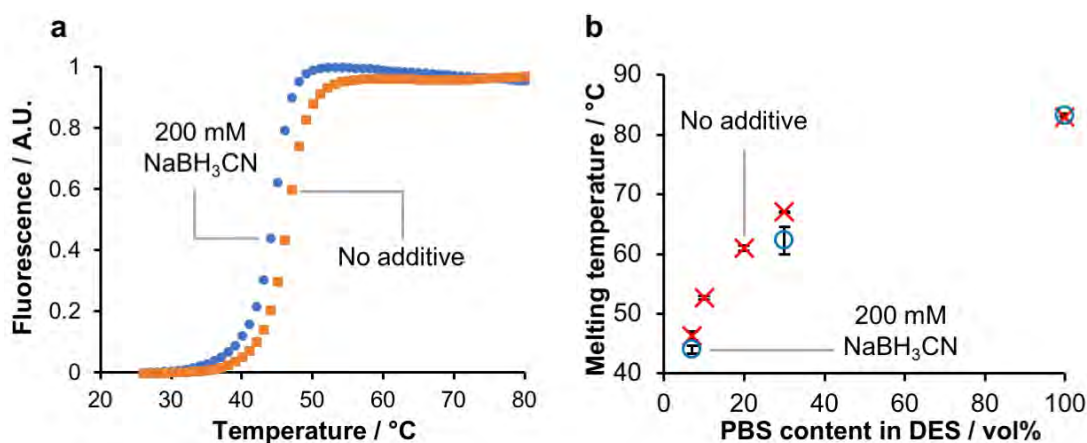


Figure 2-20. Effect of the addition of 200 mM NaBH_3CN on the T_m of dsDNA. Heating profiles analysed. a) Example of the melting trace in DES with 7 vol% PBS (0.1 M, pH 7.0) with and without reducing agent. Cooling profiles. b) T_m of dsDNA in glycoline, as function of the PBS content, with and without reducing agent. Measurements were done in triplicate. The error bars correspond to the standard deviation.

In order to prepare the aldehyde-DNA conjugate, an amine-labelled DNA strand (**2-36**) was reacted with *p*-formylbenzoic acid in the presence of EDC, *N*-hydroxysuccinimide and a tertiary amine, to furnish the amide **2-37** (Figure 2-21, Experimental Section 2.4.16).

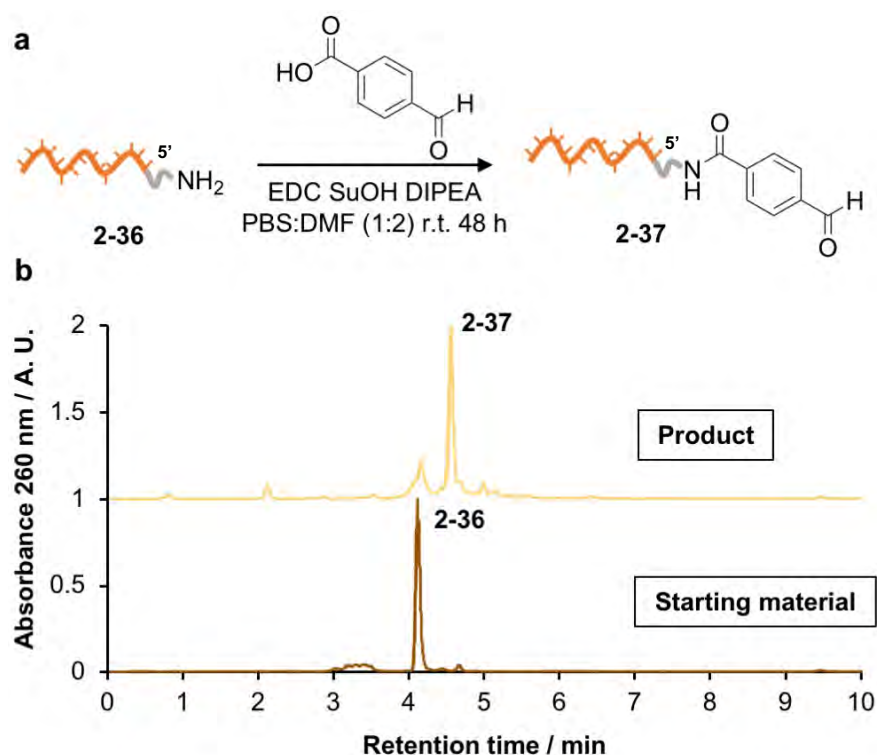
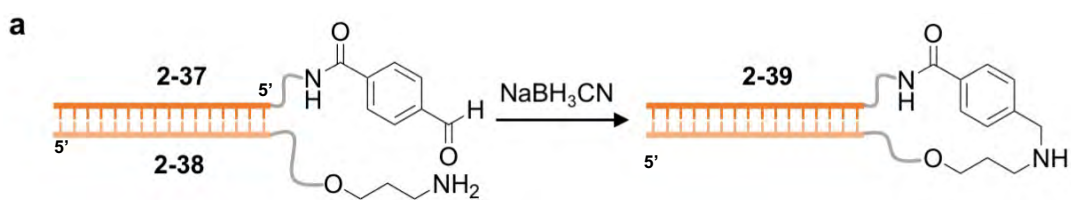


Figure 2-21. Synthesis of aldehyde-modified DNA. a) Scheme of the amide coupling reaction between p-formylbenzoic acid and amine-labelled DNA. b) HPLC of the starting material (**2-36**) and the product (**2-37**, Experimental Section 2.4.16) LC-MS (Bruker AmazonX): **2-36** $[M+Cl]^-$ m/z calcd. 10233.7, found 10229.1; **2-37** $[M-H]^-$ m/z calcd. 10330.8, found 10333.12.

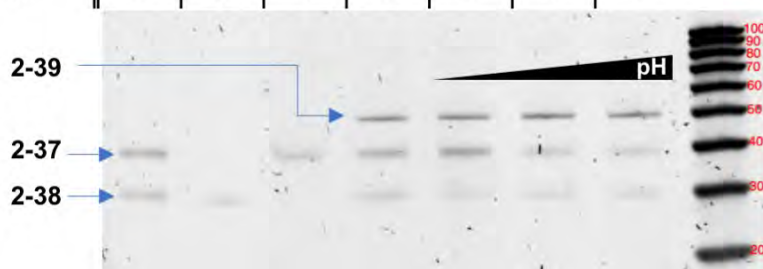
The previous reaction showed nearly quantitative conversion of **2-36** into **2-37**. Furthermore, as the excess of reagents and side products are water-soluble, **2-37** could be isolated by successive centrifugal ultrafiltration and dilution with H₂O.

Then, the DNA templated reductive amination was assayed in a series of solvents and analysed by denaturing polyacrylamide gel electrophoresis (PAGE, Figure 2-22). Firstly, the acidity of glycholine was adjusted by the addition of a 7 vol% of PBS with different pHs, and the conversion into **2-38** was estimated by optical densitometry.



b

Lane	1	2	3	4	5	6	7
NaBH ₃ CN	-	-	-	+	+	+	+
7 vol% solvent	H ₂ O	H ₂ O	H ₂ O	H ₂ O	PBS pH 4	PBS pH 5	PBS pH 6
DNA-NH ₂	+	+	-	+	+	+	+
DNA-CHO	+	-	+	+	+	+	+



c

Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14
NaBH ₃ CN (mM)	-	-	-	-	*+	+	+	+	+	-	+	+	+	+
Gly _{0.07}	+	+	+	+	+	+	+	+	+	-	-	-	-	-
MES	-	-	-	-	-	-	-	-	-	+	+	+	+	+
DNA-NH	+	-	-	+	+	+	+	+	+	+	+	+	+	+
DNA-CHO	-	+	-	+	+	+	+	+	+	+	+	+	+	+
DNA (ref.)	-	-	+	-	+	+	+	+	+	+	+	+	+	+

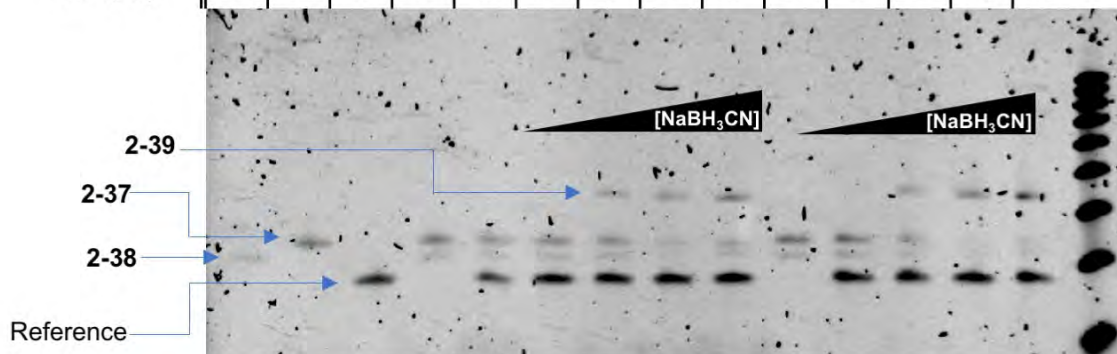


Figure 2-22. DNA templated reductive amination. a) Scheme of the reaction. b) Denaturing PAGE analysis of the reaction in glycoline / H₂O and glycoline / PBS mixtures. c) Denaturing PAGE analysis of the reaction in Gly_{0.07} and MES buffer with increasing concentration of NaBH₃CN. An unfunctionalised DNA standard was added (lane 3). In lane 5, the NaBH₃CN was added during the denaturing of the sample and loading onto the gel (Experimental section 2.4.17).

The reaction (Figure 2-22 b) proceeded only in the presence of NaBH_3CN , in its absence only starting material was detected (lane 1). The conversion in 7 vol% H_2O was very similar to 7 vol% pH 4 PBS (34%, lanes 4 and 5), while the conversion in 7 vol% pH 5 and pH 6 PBS was slightly higher (42 and 47% respectively, lanes 6 and 7). No difference in conversion was detected after 22 h compared to 4 h.

The effect of the concentration of NaBH_3CN on the conversion was determined in both $\text{Gly}_{0.07}$ and 2-(*N*-morpholino)ethanesulfonic acid (MES, 0.1 M NaCl, pH 6.0) buffer (Figure 2-23 c). In $\text{Gly}_{0.07}$, with 0.1 mM NaBH_3CN no product was observed (lane 6) and between 1 and 100 mM the yield progressively increased (lanes 7-9). The same trend, but with higher yields was observed in MES buffer (lanes 10 to 11 to 14). A control showed that the reductive amination product was not formed during the dilution and denaturing step prior to loading the samples onto the gel (lane 5).

Overall, these results show that the reductive amination conversion in glycholine DES was higher than the peptide bond formation. The compatibility of the reductive amination chemistry with acidic conditions is the most likely cause, but it has also been reported that reactions that eliminate H_2O , such as reductive amination, are favoured in DES.⁷⁷

2.3 Conclusions

The stability of a small molecule *N*-hydroxysuccinimide activated ester was improved in glycholine and ethaline DESs compared to buffered aqueous solutions at pH 7.5. Approximately 80 mol% of activated ester was remaining after 24 h in Gly_{0.07}, determined by ¹H-NMR and RP-HPLC. No improvement in stability was observed in reline DES compared to a buffered aqueous solution. The addition of moderate amounts of water, up to 50 vol% increased slightly the hydrolysis rate. The addition of triethylamine and sodium acetate had a large negative impact on the stability of the ester in DES.

While largely qualitative, an analogous study of the stability of a DNA conjugated *N*-hydroxysuccinimide activated ester showed the same trends as the small molecule study, with maximum stabilisation achieved in Gly_{0.07}. The addition of triethylamine and sodium acetate had a large destabilising effect on the stability of the conjugate.

DNA templated synthesis of peptide bonds did not proceed to more than a 5% conversion in DES, even in the presence of basic additives. It was shown that, while the reduced reactivity of amine-labelled DNA in glycholine DES could be slightly improved by the addition of a tertiary amine, this largely decreased the stability of the ester, thus preventing a good reaction yield. However, it was demonstrated that glycholine DES can provide a better storage medium for DNA conjugated activated esters, which in turn could be later diluted into aqueous buffered solution for templated peptide synthesis.

Finally, DNA templated reductive amination, which is a reaction compatible with mildly acidic pH, proceeded to a moderate conversion of 34% determined by HPLC.

2.4 Experimental section

2.4.1 DNA strands and modifications

DNA oligonucleotide stands were purchased from Integrated DNA Technologies (IDT), their sequences were designed, and their supramolecular assemblies were analysed using the web application NUPACK.⁷⁸

Table 2-4. Chapter 2 DNA oligonucleotide sequences and modifications.

Identification number	Sequence and modifications (5' → 3')
2-6	/5ThioMC6-D/ ATG TAA GTA AGT CAA GTC CAG GTC GTT CAA
2-16	/5ThioMC6-D/ ACG TTG CTG CAT TTT ACT CTT CTC CCC TCG GCA ACG T/36-TAMTSp/
2-17	/5Biosg/ TTT TTT TTT TTT TTT ACG TTG CTG CCG AGG GGA GAA GAG TAA AAT GCA ACG T
2-18	ACG TTG CTG CAT TTT ACT CTT CTC CCC TCG GCA ACG TAA AAA AAA AAA A
2-21	GGA TAA GGG CAG TGG AGG TAG GTA AGG TAG TAG GGT GAG AGT /3AIBkFQ/
2-22	GGA TAA GGG CAG TGG AGG TAG GTA AGG TAG TAG GGT GAG AGT
F2-16D	/5TET/ GAG GAG GGC AGC AAA CGG GAA GAG
F2-16Q	CTC TTC CCG TTT GCT GCC CTC CTC /3AIBkFQ/
2-23	5'-/5ThioMC6-D/ GAA CTA TGT CGT TTC CGA TGG GCA CA TTT TTT T-3'
2-26	/56-FAM/ TGT GTG CCC ATC GGA AAC GAC ATA GTT C /3AmMO/
2-29	GCC GGC CGG GCG /3AmMO/
2-S32	/5ThioMC6-D/ ATG TAA GTA AGT CAA GTC CAG GTC GTT CAA
2-34	TTT ATT TGA ACG ACC TGG ACT TGA CTT ACT TAC AT /3AmMO/

2-36

/5AmMC6/ GGA CTC CCA TTT CAA ACT CGA TGC ATT GCC
GCT

2-38

CAA TGC ATC GAG TTT GAA ATG GGA GTC C/3AmMO/

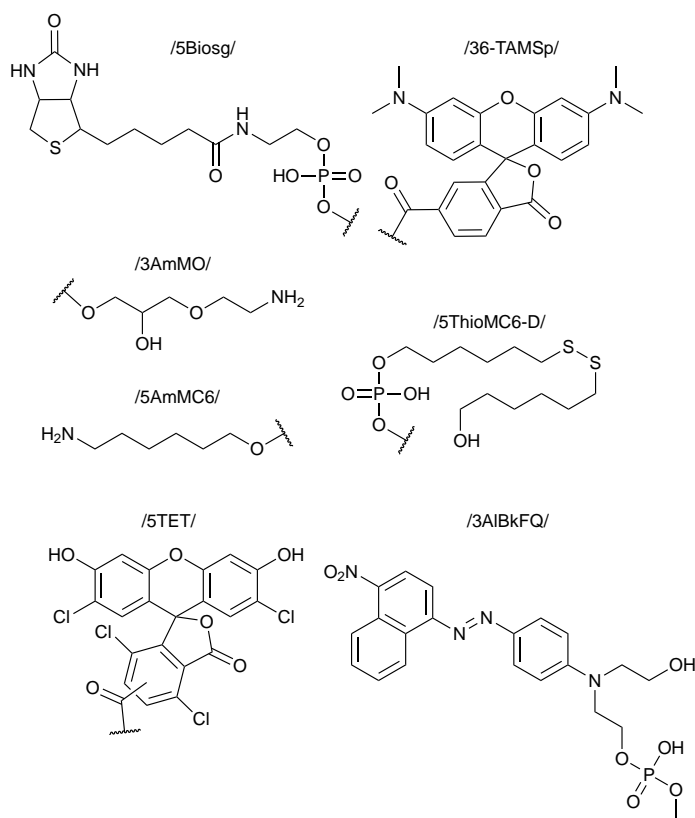


Figure 2-23. Chemical structure of commercially available DNA modifications.

2.4.2 General methods

For details on reversed phase high performance liquid chromatography (RP-HPLC), liquid chromatography-mass spectrometry (LC-MS), polyacrylamide gel electrophoresis (PAGE), nuclear magnetic resonance (NMR) spectroscopy, high resolution mass spectrometry (HR-MS) and nanopure water, see Appendix 1.

Fluorescence spectroscopy: Fluorescence spectral data was recorded on an Agilent Cary Eclipse fluorescence spectrophotometer equipped with a photomultiplier tube (PMT) detector. Quartz cuvettes from Starna scientific (Type 3/Q/10) with four polished

sides were used for fluorescence. The emission and excitation spectra were recorded using Cary Eclipse v.1.2.0.0 software.

Double-stranded DNA (dsDNA) thermal stability: dsDNA melting temperatures (T_m) were determined on a Mx3005P quantitative PCR (qPCR) instrument (Agilent) equipped with CY5, CY3, ROX, HEX/JOE and FAM fluorescence filters set. The data was recorded and processed using MxPro (Agilent) v.4.10 software.

2.4.3 Preparation of DESs

Choline chloride, urea, glycerol and ethylene glycol were purchased from Sigma Aldrich. DESs were prepared using the literature reported procedure.¹³ Very briefly, the hydrogen bond donor and the hydrogen bond acceptor were mixed in the desired molar ratio, inside a round bottom flask. Then, under positive nitrogen pressure, the heterogeneous mixture was heated up to 100 °C with constant stirring until the mixture was homogeneous. The sample was allowed to cool down, and then it was dried by different methods. The water content was determined by Karl Fischer titration: 1) Freshly prepared glycholine, no drying method, kept sealed on the bench for 1 week: 0.5 vol% H₂O; 2) Glycholine solution kept on a reagent desiccator for 12 weeks: 1.3 vol% H₂O; 3) Dried in a vacuum centrifuge for 24 h at 45 °C: 0.2 vol%; 4) Dried using a high vacuum Schlenk line for 9 h at 80 °C: 0.2 vol% H₂O; 5) Lyophilised for 4 days: 0.1 vol% H₂O. Based on the previous results, the DESs were regularly dried using a vacuum Schlenk line prior to use.

2.4.4 Correction of the acidity of DESs

Choline acetate was added to ethaline during the synthesis of the DES. Sodium acetate was added to ethaline and glycholine, and the solution was heated to 60 °C and stirred for approximately 1 h. Et₃N was added to glycholine and the solution was

shaken in a vortex mixer. 1 mL of DES was diluted in 4 mL of H₂O and the pH was determined using a pH sensitive glass electrode. The error was evaluated in three different series of samples, by performing triplicated measurements and determining the standard deviation. Then, single measurements were performed.

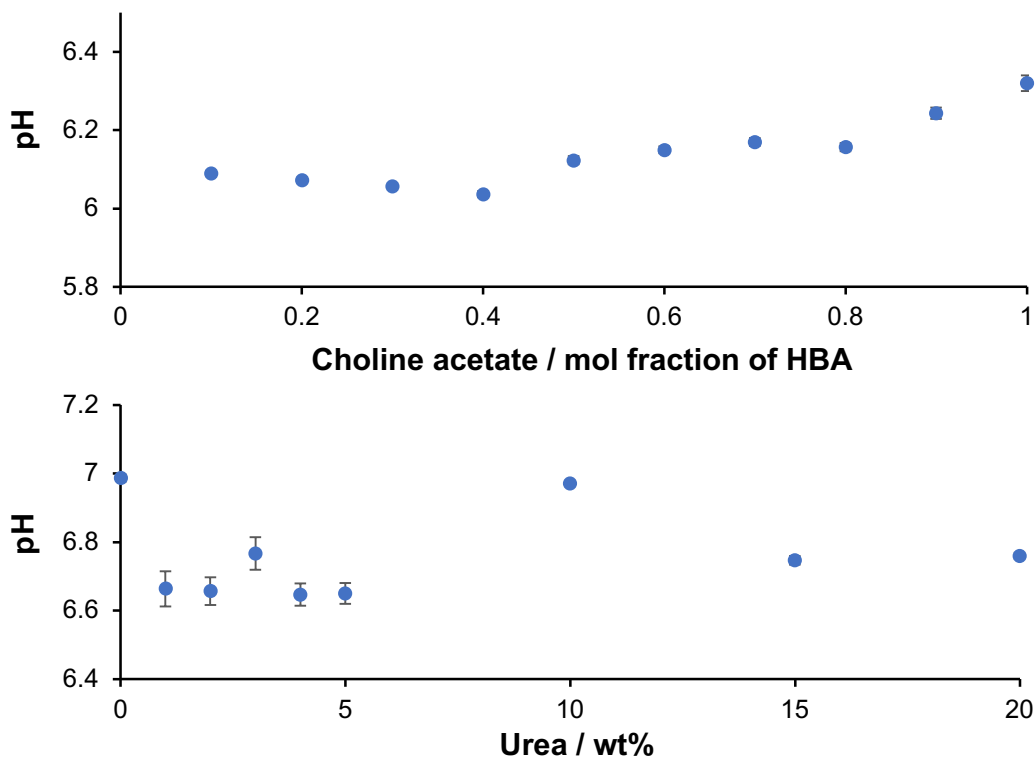


Figure 2-24. Evolution of the pH of an ethaline / choline acetate (top) and glycholine / urea 2 (bottom) solutions. Triplicate measurements were performed. The error bars correspond to the standard deviation.

2.4.5 HPLC stability assay of activated esters in DESs

4.1 mg (0.01 mmol) of activated ester (**2-1**) were dissolved in 2 mL of DES. Each solution was prepared 40 min after the previous one. The solutions were stirred over a period of 29 h. After stirring for 5, 24 and 29 h, a 100 μ L aliquot of each sample was diluted in 900 μ L of H₂O. Then, the sample was analysed by RP-HPLC.

Method: column Discovery C18 (Sigma Aldrich, 5 μm , 10x4.6 cm), flow 1 $\text{mL}\cdot\text{min}^{-1}$, temperature 40 $^{\circ}\text{C}$ and injection volume 10 μL . Solvent A: H_2O , 0.05 vol% TFA. Solvent B: 70 vol% ACN, 0.05 vol% TFA.

Time / min	Solvent B / vol%
0.0	0
27.0	75
27.5	75
28.5	0
30.0	0

2.4.6 $^1\text{H-NMR}$ spectroscopy stability assay of activated esters in DESs

4.1 mg of activated ester (**1**) were dissolved in 2 mL of DES. An individual solution was prepared for each solvent and timepoint. The solutions were stirred over a period of 24 h. The samples were extracted with 0.7 mL of acetone- d_6 and subsequently analysed by $^1\text{H-NMR}$ spectroscopy. The proportion of ester to carboxylic acid in the mixture was determined from the integrations of the methyl protons (Figure 2-5). The progression of the NMR signal is shown in Figure 2-24. For PBS solutions, CDCl_3 was used.

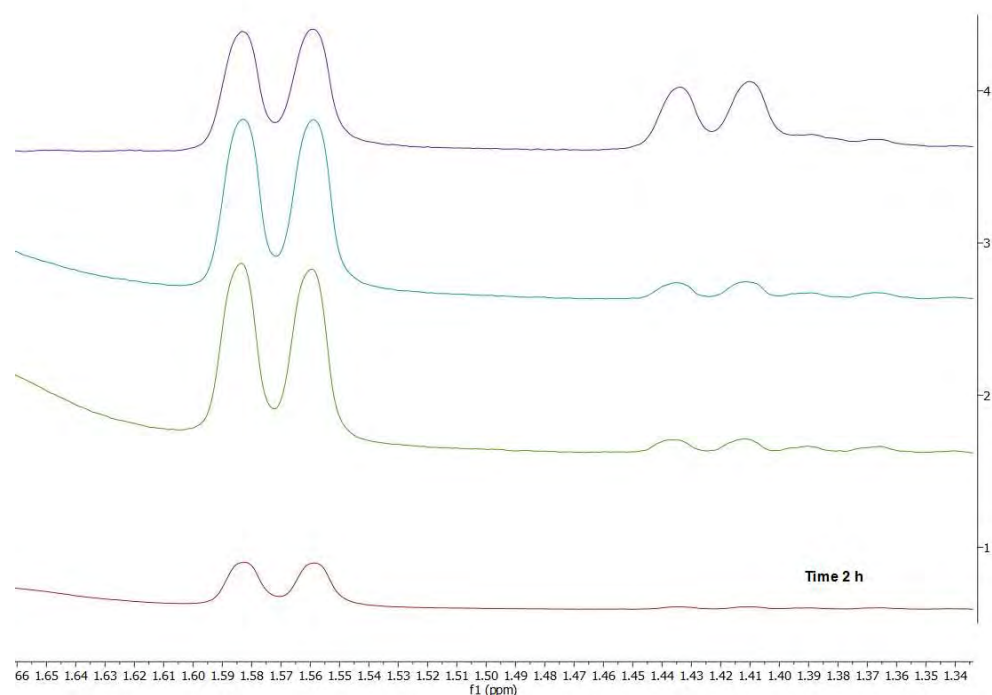


Figure 2-25. Stacked $^1\text{H-NMR}$ spectra of the extracted solutions glycoline 50 vol% H_2O . From top to bottom, the traces correspond to 24 h (4), 6 h (3), 4 h (2) and 2 h (1) solvolysis time.

2.4.7 Small molecule peptide bond formation in glycoline

51 mg of H-L-Phe-OMe·HCl (**2-4**, 0.24 mmol) were dissolved in 1 mL of THF with 40 μL of Et_3N (0.29 mmol). The $\text{Et}_3\text{N}\cdot\text{HCl}$ precipitate was filtered off and the organic solvent was removed under reduced pressure. The resulting product was dissolved in 4 mL of $\text{Gly}_{0.07}$, and 100 mg of activated ester **1** were added (0.24 mmol). The mixture was stirred at room temperature for 5 h. Finally, the glycoline solution was diluted with 12 mL of H_2O and the product was extracted three times with 10 mL of EtOAc. The product was isolated by SiO_2 column chromatography with DCM:MeOH (100:0 to 95:5) to produce 47 mg (42%) of **5** as a white powder.

$^1\text{H-NMR}$ (300 MHz, chloroform-*d*) δ (ppm): 7.70 (1H, s, Ar-*H*), 7.23 (3H, m, Ar-*H*), 7.07 (2H, m, Ar-*H*), 7.00 (1H, s, Ar-*H*), 6.56 (1H, d $J = 7.9$ Hz, NH), 6.56 (1H, d $J = 7.9$ Hz,

NH), 5.61 (1H, d $J = 8.0$ Hz, NH) 5.49 (2H, s^{br}, OCH₂), 4.84 (1H, q $J = 6.0$ Hz, CH-Bn), 4.26 (1H, q $J = 7.4$ Hz, CH-CH₃), 3.97 (3H, s, OCH₃), 3.94 (3H, s, OCH₃), 3.72 (3H, s, OCH₃), 3.11 (2H, m, CH₂-Ph), 1.37 (3H, d $J = 12.0$ Hz, CH-CH₃). ¹³C-NMR (75 MHz, chloroform-*d*) δ (ppm): 171.98 (C=O), 171.76 (C=O), 155.44 (C=O), 153.77 (ArC), 148.13 (ArC), 139.62 (ArC), 135.68 (ArC), 129.28 (ArC-H), 128.65 (ArC-H), 128.13 (ArC), 127.24 (ArC-H), 109.84 (ArC-H), 108.19 (ArC-H), 63.90 (OCH₂), 56.58 (OCH₃), 56.46 (OCH₃), 53.38 (OCH₃), 52.49 (NCH), 50.53 (NCH), 37.85 (CH₂-Ph), 18.79 (CH-CH₃). **HRMS** (ESI+) m/z [M+H]⁺ calcd. 512.1640 found 512.1639

2.4.8 Synthesis of DNA conjugated activated esters

The synthesis of DNA conjugated activated esters was adapted from the literature reported procedure.⁵⁰ Very briefly, 60 μ L of DNA-SS-(CH₂)₆-OH solution (1 mM in H₂O, 60 nmol) were combined with 13.3 μ L of a tris-(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) solution (489 mM, 6.5 μ mol) previously neutralised to pH 4 with 5.0 M NaOH. The resulting solution was shaken at 18 °C for 1.25 h. 660 μ L of a solution of the *N*-hydroxymaleimide activated ester in DMF (**2-7** or **2-9**, 36.4 mM, 400 eq) was added and the resulting solution was shaken at 18 °C for 1.0 h. The organic solvent was diluted with 1 vol. of H₂O, the precipitate was centrifuged at 4 °C and 21 kRCF for 5 min. The supernatant was collected, frozen in liquid nitrogen and the total volume of the sample was reduced by lyophilisation over 3 to 4 h, while keeping the sample refrigerated to prevent rapid melt. The solution was allowed to thaw and, using ultrafiltration, it was washed three times with water and concentrated to ~240 μ L using 3 kDa molecular weight cut-off (MWCO) amicon (Sigma Aldrich) spin-filters.

2.4.9 Preparation of FAM labelled L-alanine N-hydroxymaleimide ester

All reagents were directly purchased from Sigma Aldrich unless otherwise stated. Dry solvents were directly purchased from Sigma Aldrich.

Synthesis of 2-12: It was synthesised according to a reported method.⁶⁵ 3.322 g of fluorescein free carboxylic acid (**11**, 10 mmol) and 1.151 g of *N*-hydroxysuccinimide (10 mmol) were dissolved in 19 mL of dry DMF under positive nitrogen pressure in a dry Schlenk tube. Then, 2.063 g of *N,N'*-dicyclohexylcarbodiimide (DCC, 10 mmol) were added and the resulting solution was stirred at 60 °C for 2 h under positive nitrogen pressure. *N,N'*-dicyclohexylurea (DCU) was removed by cooling down the reaction mixture to -20 °C for 2 h and the white precipitate was filtered off. The solvent was removed under reduced pressure and the solid crude was isolated by SiO₂ column chromatography using an EtOAc : acetone gradient. The desired fractions were combined to produce 2.002 g (47%) of **2-12** as a bright orange solid. The ¹H-NMR analysis was in good agreement with the reported characterisation.⁶⁵

TLC (EtOAc:MeOH 6:4) R_f = 0.1 **¹H-NMR** (300 MHz, DMSO-*d*₆) δ (ppm): 8.36 (1H, d *J* = 7.7 Hz, Ar-*H*), 7.96 (2H, m, Ar-*H*), 7.65 (1H, d *J* = 7.5 Hz, Ar-*H*), 6.79 (2H, d *J* = 9.6 Hz, Ar-*H*), 6.53 (4H, m, Ar-*H*), 2.73 (4H, s^{br}, (CH₂)₂ succinimide). **HRMS** (ESI-) *m/z* [M-H]⁻ calcd. 428.0770 found 428.0763

Synthesis of 2-13: 1.000 g of **2-12** (2.23 mmol) and 1.637 g of TsOH·H-L-Ala-OBn were dissolved in 15 mL of dry DMF under positive nitrogen pressure in a dry round bottom Schlenk flask. Then, 1.1 mL of Et₃N (8.16 mmol) were added drop wise over 10 min and the resulting solution was stirred at room temperature for 6 h. The solvent was removed under reduced pressure and the crude was isolated by SiO₂ column

chromatography using a DCM : MeOH (100:0 to 75:15) gradient to produce 704 mg (61%) of **13** as a bright orange solid.

TLC (EtOAc:MeOH 6:4) $R_f = 0.6$ **¹H-NMR** (400 MHz, methanol- d_4) δ (ppm): 7.76 (1H, m, Ar-H), 7.55 (1H, m, Ar-H), 7.41 (1H, m, Ar-H), 7.21 (2H, m, Ar-H), 7.11 (4H, m, Ar-H), 6.88 (1H, m, Ar-H), 6.56 (1H, m, Ar-H), 6.51 (1H, m, Ar-H), 6.45 (1H, m, Ar-H), 6.20 (1H, d $J = 8.7$ Hz, Ar-H), 5.97 (1H, dd $J_1 = 8.7$ Hz $J_2 = 2.5$ Hz, Ar-H), 4.80 (2H, m, OCH₂), 3.59 (1H, q $J = 7.2$ CH-CH₃), 1.02 (3H, d $J = 7.2$ Hz, CH-CH₃). **¹³C-NMR** (100 MHz, methanol- d_4) δ (ppm): 171.75 (C=O), 168.79 (C=O), 160.29 (C=O), 160.13 (ArC), 154.54 (ArC), 154.36 (ArC), 154.24 (ArC), 136.95 (ArC), 134.40 (ArC-H), 132.13 (ArC-H), 130.71 (ArC-H), 130.60 (ArC-H), 129.89 (ArC-H), 129.53 (ArC), 129.37 (ArC-H), 129.35 (ArC-H), 129.15 (ArC), 129.09 (ArC), 125.13 (ArC-H), 123.55 (ArC-H), 113.26 (ArC-H), 112.95 (ArC-H), 110.26 (ArC), 109.24 (ArC), 103.59 (ArC-H), 103.08 (ArC-H), 68.08 (OCH₂), 51.55 (NCH), 15.44 CH₃). **HRMS** (ESI-) m/z [M-H]⁻ calcd. 492.1447 found 492.1457

Synthesis of 2-14: 50 mg of Pd/charcoal (10 wt%) were placed in a dry round bottom Schlenk flask under positive nitrogen pressure. 500 mg of **2-13** (1.01 mmol) were added and dissolved with 50 mL of degassed MeOH. Nitrogen atmosphere was replaced by hydrogen by bubbling the solution for 3 min with a hydrogen balloon. The positive hydrogen pressure was maintained for 1.5 h with constant stirring at room temperature. The solid supported catalyst was removed by filtration through celite. The solvent was removed under reduced pressure to produce 375 mg (92%) of **2-14** as a bright yellow solid.

TLC (DCM:MeOH 9:1) $R_f = 0.2$ **¹H-NMR** (500 MHz, methanol- d_4) δ (ppm): 7.05 (1H, m, Ar-H), 6.61 (3H, m, Ar-H), 6.48 (2H, m, Ar-H), 6.41 (1H, dd $J_1 = 8.7$ Hz $J_2 = 2.4$ Hz,

Ar-H), 3.67 (1H, q $J = 7.2$ Hz, CHCH₃), 1.20 (3H, d $J = 7.2$ Hz CHCH₃). **¹³C-NMR** (160 MHz, methanol-*d*₄) δ (ppm): 173.6 (C=O), 169.1 (C=O), 160.6 (C=O), 160.4 (ArC), 154.9 (ArC), 154.6 (ArC), 154.4 (ArC), 134.5 (ArC), 132.5 (ArC), 132.4 (ArC-H), 130.9 (ArC-H), 130.0 (ArC-H), 113.4 (ArC-H), 113.0 (ArC-H), 110.6 (ArC), 109.6 (ArC), 103.7 (ArC-H), 103.2 (ArC-H), 51.7 (CH-CH₃), 15.7 (CH-CH₃). **HRMS** (ESI-) m/z [M-H]⁻ calcd. 402.0983 found 402.0982

Synthesis of 2-15: 50 mg of **2-14** (0.124 mmol), 25 mg of DCC (0.121 mmol) and 14 mg of *N*-hydroxymaleimide (0.124 mmol) were dissolved in 2 mL of EtOAc:1,4-dioxane (2:1) and stirred at room temperature overnight. Then, the solvent was lyophilised in the Schlenk line, the crude was dissolved in 2.5 mL of EtOAc and cooled down to 4 °C for 1 h. The white precipitate of DCU was centrifuged at 4 °C and 21 kRCF for 15 min. The supernatant was dried to produce 58 mg (95%) of **2-15** as a bright yellow powder. The NMR spectra showed extra peaks which were attributed to atropoisomers and the presence of trace amounts of DCU.

¹H-NMR (500 MHz, DMSO-*d*₆) δ (ppm): 9.99 (1H, s, OH or NH), 9.89 (1H, s, OH or NH), 7.83 (1H, m, Ar-H), 7.57 (2H, m, Ar-H), 7.17 (2H, s^{br}, maleimide Ar-H), 7.05 (1H, m, Ar-H), 6.61 (2H, m, Ar-H), 6.55 (1H, m, Ar-H), 6.49 (2H, m, Ar-H), 6.38 (1H, m, dd $J_1 = 8.7$ Hz $J_2 = 2.4$ Hz, Ar-H), 4.11 (1H, q $J = 7.1$ Hz, CH-CH₃), 1.26 (3H, d $J = 7.1$ Hz, CH-CH₃). **¹³C-NMR** (160 MHz, DMSO-*d*₆) δ (ppm): 167.28 (C=O), 165.96 (C=O), 164.38 (C=O), 158.71 (C=O), 157.36 (ArC), 153.20 (ArC), 152.13 (ArC), 133.43 and 133.09 (maleimide CH=CH), 131.91 (ArC-H), 130.07 (ArC-H), 129.68 (ArC), 129.45 (ArC-H), 128.78 (ArC-H), 123.85 (ArC-H), 122.58 (ArC-H), 112.37 (ArC-H), 112.11 (ArC-H), 108.47 (ArC), 107.62 (ArC), 102.41 (ArC-H), 102.06 (ArC-H), 48.16 (CH-CH₃), 15.05 (CH-CH₃). **HRMS** (ESI-) m/z [M-H]⁻ calcd. 497.0985 found 497.0992

2.4.10 Fluorescence stability assay using a DNA hairpin

1 μL of the DNA hairpin (**2-16**) (1 mM in H_2O , 1 nmol) was combined with 1.1 μL (1 mM, 1.1 nmol) of the fully complementary biotinylated DNA strand **2-17** and 10 μL of PBS. The strands were thermally annealed in a PCR mastercycler instrument (Eppendorf) by fast heating to 90 $^\circ\text{C}$ and cooling to 20 $^\circ\text{C}$ at 2 $^\circ\text{C}\cdot\text{min}^{-1}$. 1 μL of 0.5 M TCEP·HCl solution (pH 4 corrected with 5 M NaOH, 500 nmol) was subsequently added, and the solution was shaken at 18 $^\circ\text{C}$ for 1.25 h. 275 μL of streptavidin coated magnetic nanoparticles Dynabeads (Thermofisher, 400 pmol / 100 μL , 1.1 nmol capacity) were washed with 3 \times 500 μL with PBS and dispersed in 100 μL of PBS. The previous dsDNA solution was added over the Dynabeads solution and the resulting mixture was incubated at 18 $^\circ\text{C}$ for 0.5 h to allow the biotin-streptavidin binding. The beads were magnetically concentrated on the side of a microcentrifuge tube, washed with 500 μL of Gly_{0.07} and suspended in 200 μL of Gly_{0.07} with intense shaking. 1 mg of the maleimide activated ester **2-15** (2.5 μmol) was added to the solution and it was shaken at 18 $^\circ\text{C}$ for 2 h. The beads were magnetically concentrated on the side of the microcentrifuge tube, the excess of reagent was removed, and the beads were washed with 2 \times 500 μL of fresh Gly_{0.07}. Finally, the beads were suspended in 200 μL of Gly_{0.07}. Then, 1.1 μL (1 mM, 1.1 nmol) of the displacement strand (**2-18**) were added to trigger the toehold mediated displacement of the doubly functionalised hairpin. The solution was incubated at 18 $^\circ\text{C}$ for 2 h. The nanoparticles were magnetically concentrated on the side of the microcentrifuge tube and the free hairpin (at a maximum possible concentration of 5 μM) was collected and diluted 5:95 in optical grade PCR microcentrifuge tubes, with the solvents of interest for the stability assay. It was very

important to centrifuge the PCR tubes for 60 s at 14 kRPM in order to stabilise the fluorescence signal. The tubes were placed in the qPCR instrument described above, and the fluorescence emission of the FAM group was monitored for 24 h.

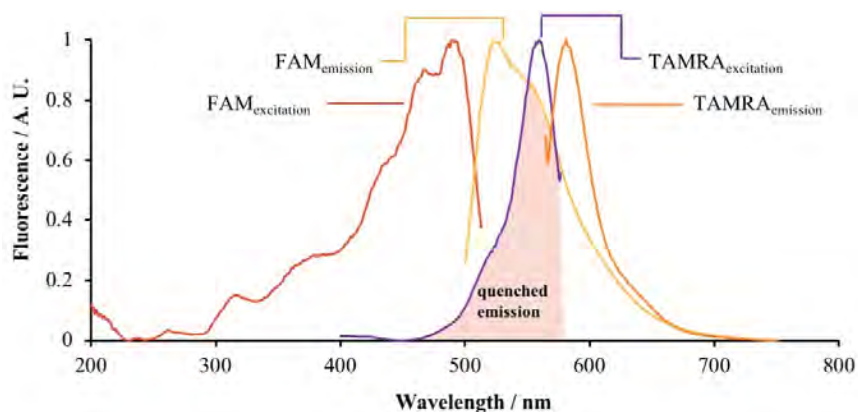


Figure 2-26. Excitation and emission fluorescence spectra of FAM (2-15) and TAMRA in Gly_{0.07}.

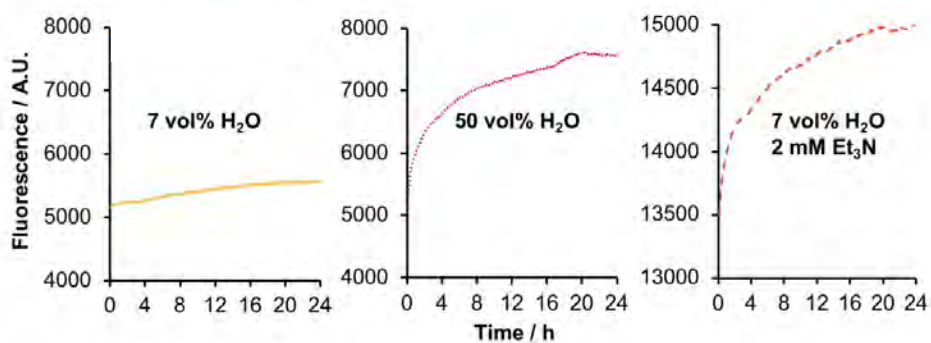


Figure 2-27. Comparison of the evolution of fluorescence over time in Gly_{0.07}, Gly_{0.5} and Gly_{0.07} / 2 mM Et₃N.

2.4.11 Derivatisation of the equation 1

The fraction activated ester and the fluorescence can be correlated based on: 1) the system will evolve from fully quenched (before hydrolysis) to fully unquenched (after hydrolysis), 2) each one of the possible states, hydrolysed and activated ester, has a distinctive fluorescence emission at a given wavelength, 3) the hydrolysed system has a high emission while the activated ester has a low emission, 4) the total fluorescence

intensity of the sample will be a combination of the fluorescence intensity of the hydrolysed system and the activated ester system.

$$F(t) = [COOH](t) \cdot k_{COOH} + [Ae](t) \cdot k_{Ae} \quad \text{Eq. S1}$$

$[COOH](t)$ = Concentration of free carboxylic acid as function of time

$[Ae](t)$ = Concentration of activated ester as function of time

k_{COOH} = Proportionality constant between the concentration of free carboxylic acid and its fluorescence

k_{Ae} = Proportionality constant between the concentration of activated ester and its fluorescence

$$[Ae]_0 = [COOH](t) + [Ae](t) \quad \text{Eq. S2}$$

$[Ae]_0$ = Concentration of activated ester at time 0

$$F(t = 0) = F_{Min} = [Ae]_0 \cdot k_{Ae} \quad \text{Eq. S3}$$

F_{Min} = Absolute fluorescence of the fully quenched system, which is the only significant state at the beginning of the experiment, time 0.

$$F(t \rightarrow \infty) = F_{Max} = [COOH](t \rightarrow \infty) \cdot k_{COOH} = [Ae]_0 \cdot k_{COOH} \quad \text{Eq. S4}$$

$[COOH](t \rightarrow \infty)$ = Concentration of acid once the hydrolysis has been completed

Substituting eq. S3 and S4 in eq. S1:

$$F(t) = [COOH](t) \cdot \frac{F_{Max}}{[Ae]_0} + [Ae](t) \cdot \frac{F_{Min}}{[Ae]_0} \quad \text{Eq. S5}$$

Substituting eq. S2 in eq. S5 and re-arranging:

$$F(t) = \{[Ae]_0 - [Ae](t)\} \cdot \frac{F_{Max}}{[Ae]_0} + [Ae](t) \cdot \frac{F_{Min}}{[Ae]_0} \quad \text{Eq. S6}$$

Eq. S6 can be re-arranged into Eq. S7 as follows:

$$F(t) = F_{Max} - \frac{[Ae](t)}{[Ae]_0} \{F_{Max} - F_{Min}\} \quad \text{Eq. S7}$$

The fraction of activated ester (χ_{Ae}), is defined in eq. S8

$$\chi_{Ae} = \frac{[Ae](t)}{[Ae]_0} \quad \text{Eq. S8}$$

Introducing eq. S8 into eq. S7 and re-arranging:

$$F(t) = F_{Max} - \chi_{Ae} \cdot \{F_{Max} - F_{Min}\} \quad \text{Eq. S9}$$

This can be re-arranged, and signs inverted as follows:

$$\chi_{Ae} = \frac{F_{Max} - F(t)}{F_{Max} - F_{Min}} \quad \text{Eq. 1}$$

2.4.12 Thermal stability of dsDNA in DES / H₂O mixtures

The thermal stability of dsDNA in DES solutions was determined using fluorophore and quencher labelled DNA strands (**F2-16D** and **F2-16Q**). Solutions at 100 nM were briefly centrifuged and placed in a qPCR instrument. Samples were heated and cooled at 1 °C·min⁻¹. The T_m was determined as the middle-point of the sigmoidal trace defined by the fluorescence emission intensity as function of the temperature.

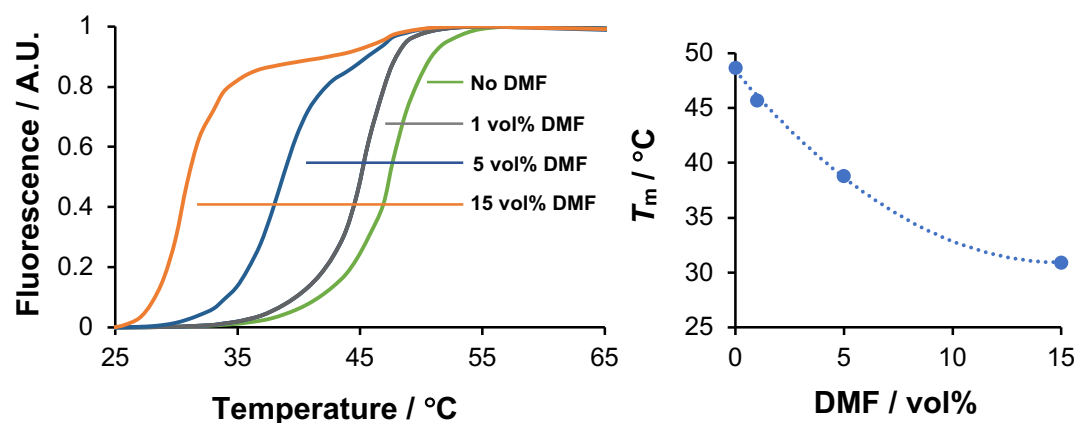


Figure 2-28. Effect of the addition of DMF to Gly_{0.07} on the T_m .

2.4.13 DNA templated synthesis in DES

10 μ L of disulfide-labelled DNA (**2-23**) (1 mM in H₂O, 10 nmol) were mixed with 2 μ L of TCEP·HCl solution (pH 4 corrected with 5 M NaOH, 1 μ mol) in a microcentrifuge

tube. The solution was shaken at 18 °C for 1.25 h. 108 µL of a DMF solution of the maleimide activated ester **2-15** (37 mM, 4 µmol) were added and the combined solution was shaken at 18 °C for 1.0 h. Then, the volume of the solution was reduced to approximately 50 µL by freeze-drying at low temperature. The previous solution was diluted with 4950 µL of Gly_{0.07}, 10 µL of the amine-labelled DNA (**2-26**, 1 mM in H₂O, 10 nmol), the resulting mixture was vigorously shaken until the solution was homogeneous and it was shaken at 18 °C overnight. The DES solution was diluted with an excess of H₂O and the DNA strands were transferred into water by successive centrifugal ultrafiltrations using a 3 kDa MWCO amicon spin filter. The product was analysed by denaturing HPLC: 63 mM 1,1,1,3,3,3-hexafluoroisopropanol (HFIP), 16.3 mM Et₃N, 5 vol% MeOH (buffer A), 63 mM HFIP, 16.3 mM Et₃N, 80 vol% MeOH (buffer B). 0.8 mL·min⁻¹, XBridge OST C18 column (135 Å, 2.5 µM, 4.6x50 mm), 50 °C.

The analogous templated reaction in aqueous buffered solution was performed by synthesising the DNA conjugated activated ester as described in Section 2.4.8, then the complementary amine-labelled DNA strand (**26**) was added minimising the time to transfer the activated ester into buffered solution.

2.4.14 Amine-labelled DNA reactivity in DES

1 µL of amine-labelled DNA (**2-29**, 1 mM in H₂O, 1 nmol) was diluted in 100 µL of the appropriate DES / Et₃N or NaCl (1 M in H₂O) / Et₃N solvent. Approximately 0.3 mg of the *N*-hydroxysuccinimide ester (**2-30**, 1 µmol) were added, and the mixture was shaken to produce a 10 mM ester and 10 µM DNA solution. The solution was shaken at 20 °C, overnight. The product was transferred into H₂O by dilution of the samples

and successive centrifugal ultrafiltrations using a 3 kDa MWCO amicon spin filter. The product was analysed by HPLC and LC-MS.

2.4.15 Storage of DNA conjugated activated esters in DES

3 μL of disulfide-labelled DNA (**S32**) (1 mM in H_2O , 3 nmol) were mixed with 3 μL of TCEP·HCl solution (pH 4 corrected with 5 M NaOH, 1.5 μmol) in a microcentrifuge tube. The solution was shaken at 18 °C for 1.25 h. 54 μL of a DMF solution of the activated ester NVOC-*L*-Ala-(*N*-hydroxymaleimide) (22 mM, 1.2 μmol) were added and the combined solution was shaken at 18 °C for 1.0 h. This sample was divided into 3 aliquots.

1) Storage in Gly_{0.07}: 20 μL of the previous activated ester solution were diluted with 180 μL of Gly_{0.07}, and the resulting solution was shaken at 18 °C for 2.0 h. Then, the solution was diluted with 1.3 mL of PBS (0.1 M, pH 7.5) containing 1 μL of complementary amine-labelled DNA **2-34** (1 mM in H_2O , 1 nmol). The solution was shaken at 18 °C for 5 h. Finally, the solution was diluted with 1 vol. of H_2O , the precipitate was centrifuged, and the DNA products were transferred into H_2O by successive steps of dilution of the supernatant with H_2O and concentration by centrifugal ultrafiltration using a 3 kDa MWCO amicon spin filter. The product was analysed by LC-MS.

2) Storage in PBS: 20 μL of the activated ester solution were diluted with 180 μL of PBS (0.1 M, pH 7.0), and the resulting solution was shaken at 18 °C for 2.0 h. Then, the sample was treated as in 1.

3) Fast reaction in PBS: 20 μL of the activated ester solution were diluted with 1480 μL of PBS (0.1 M, pH 7.5) containing 1 nmol of complementary amine-labelled DNA.

The solution was shaken at 18 °C for 5.0 h. The products were transferred into H₂O for LC-MS analysis as described previously.

4) Off-template control: 18 µL of maleimide activated ester NVOC-*L*-Ala-(*N*-hydroxymaleimide) (22 mM in DMF, 400 nmol) were diluted in 1482 µL of PBS (0.1 M, pH 7.5) containing 1 µL of amine-labelled DNA (1 mM in H₂O, 1 nmol). The solution was shaken at 18 °C for 5.0 h and the product was transferred into H₂O for LC-MS analysis as described previously.

2.4.16 Synthesis of aldehyde-labelled DNA

Stock solutions: 300 mM 4-formylbenzoic acid in DMF (45 mg · mL⁻¹), 300 mM EDC·HCl in DMF (dispersion, 57.6 mg · mL⁻¹), 300 mM *N*-hydroxysuccinimide in DMF (34.6 mg · mL⁻¹). 67 µL of each stock solution were combined in a microcentrifuge tube, and 3.5 µL of ⁱPr₂EtN were added. The mixture was allowed to react for 10 min. In a separate microcentrifuge tube 20 µL of amine-labelled DNA (1 mM in H₂O, 20 nmol) were diluted with 180 µL of PBS (0.1 M, pH 7.5) and 200 µL of DMF. The small molecule solution was added over the DNA solution and the resulting mixture was shaken at 20 °C for 48 h. The solution was then diluted with 1 mL of H₂O, the precipitate was centrifuged, and the DNA was transferred into H₂O by successive steps of dilution of the supernatant with H₂O and concentration by centrifugal ultrafiltration using a 3 kDa MWCO amicon spin filter. The product was analysed by HPLC and LC-MS (85% conversion based on HPLC peak area).

2.4.17 DNA templated reductive amination in DES

Stock solutions and reagents: NaBH₃CN (aq, 2 M with heating), anhydrous glycoline, amine-labelled DNA (1 mM in H₂O) and aldehyde-labelled DNA (1 mM in H₂O), PBS

(0.1 M, pH 4), PBS (0.1 M, pH 5), PBS (0.1 M, pH 6) and H₂O. Glycholine was mixed with the appropriate volumes of PBS, H₂O and DNA strands were combined, vigorously shaken, briefly centrifuged and the appropriate volume of NaBH₃CN was finally added. The volumes are described in Table 2-5

The solution was shaken at 18 °C overnight. Then, 1 µL of the sample was denatured by heating at 75 °C, diluted with 50 µL of H₂O and 50 µL of PAGE denaturing loading buffer. The denatured sample was analysed by 15% denaturing PAGE (see Appendix 1 polyacrylamide electrophoresis).

Table 2-5. Experimental conditions. Volumes of reagents and solvents in µL.

Lane	1	2	3	4	5	6	7
DNA-NH ₂	0.7	0.7	-	0.7	0.7	0.7	0.7
DNA-CHO	0.8	-	0.8	0.8	0.8	0.8	0.8
NaBH ₃ CN	-	-	-	1	1	1	1
PBS	-	-	-	-	1 (pH 4.0)	1 (pH 5.0)	1 (pH 6.0)
H ₂ O	2	2.8	2.7	1	-	-	-
Glycholine	46.5	46.5	46.5	46.5	46.5	46.5	46.5

2.5 References

- 1 P. Walden, *Bull. l'Academie Imp. des Sci. (St. Petersburg)*, 1914, **1800**, 405–422.
- 2 S. Zhang, N. Sun, X. He, X. Lu and X. Zhang, *J. Phys. Chem. Ref. Data*, 2006, **35**, 1475–1517.
- 3 Q. Dong, C. D. Muzny, A. Kazakov, V. Diky, J. W. Magee, J. A. Widegren, R. D. Chirico, K. N. Marsh and M. Frenkel, *J. Chem. Eng. Data*, 2007, **52**, 1151–1159.
- 4 T. Welton, *Chem. Rev.*, 1999, **99**, 2071–2084.
- 5 J. P. Hallett and T. Welton, *Chem. Rev.*, 2011, **111**, 3508–3576.
- 6 H. Niedermeyer, J. P. Hallett, I. J. Villar-Garcia, P. A. Hunt and T. Welton, *Chem. Soc. Rev.*, 2012, **41**, 7780.
- 7 K. Padászyński and U. Domańska, *J. Chem. Inf. Model.*, 2014, **54**, 1311–1324.
- 8 N. V. Plechkova and K. R. Seddon, *Chem. Soc. Rev.*, 2008, **37**, 123–150.
- 9 M. Armand, F. Endres, D. R. MacFarlane, H. Ohno and B. Scrosati, *Nat. Mater.*, 2009, **8**, 621–629.
- 10 P. Kubisa, *Prog. Polym. Sci.*, 2009, **34**, 1333–1347.
- 11 T. Torimoto, T. Tsuda, K. I. Okazaki and S. Kuwabata, *Adv. Mater.*, 2010, **22**, 1196–1221.
- 12 A. P. Abbott, G. Capper, D. L. Davies, R. K. Rasheed and V. Tambyrajah, *Chem. Commun.*, 2003, **9**, 70–71.
- 13 A. P. Abbott, D. Boothby, G. Capper, D. L. Davies and R. K. Rasheed, *J. Am. Chem. Soc.*, 2004, **126**, 9142–9147.

- 14 P. Xu, G. W. Zheng, M. H. Zong, N. Li and W. Y. Lou, *Bioresour. Bioprocess.*, 2017, **4**, 34–52.
- 15 Q. Zhang, K. De Oliveira Vigier, S. Royer and F. Jérôme, *Chem. Soc. Rev.*, 2012, **41**, 7108–7146.
- 16 E. L. Smith, A. P. Abbott and K. S. Ryder, *Chem. Rev.*, 2014, **114**, 11060–11082.
- 17 M. Pätzold, S. Siebenhaller, S. Kara, A. Liese, C. Syldatk and D. Holtmann, *Trends Biotechnol.*, 2019, **37**, 943–959.
- 18 H. Tateishi-Karimata and N. Sugimoto, *Nucleic Acids Res.*, 2014, **42**, 8831–8844.
- 19 S. Handy. *Ionic Liquids - Current State of the Art*, 1st Ed. 2015.
- 20 D. A. Alonso, A. Baeza, R. Chinchilla, G. Guillena, I. M. Pastor and D. J. Ramón, *European J. Org. Chem.*, 2016, **2016**, 612–632.
- 21 M. Obst and B. König, *European J. Org. Chem.*, 2018, **2018**, 4213–4232.
- 22 A. K. Sanap and G. S. Shankarling, *Catal. Commun.*, 2014, **49**, 58–62.
- 23 K. Salimiyani and D. Saberi, *ChemistrySelect*, 2019, **4**, 3985–3989.
- 24 N. Azizi, E. Batebi, S. Bagherpour and H. Ghafuri, *RSC Adv.*, 2012, **2**, 2289–2293.
- 25 N. Azizi, Z. Yadollahy and A. Rahimzadeh-Oskooee, *Tetrahedron Lett.*, 2014, **55**, 1722–1725.
- 26 C. Vidal and J. García-Álvarez, *Green Chem.*, 2014, **16**, 3515–3521.
- 27 A. Benedetto and P. Ballone, *ACS Sustain. Chem. Eng.*, 2016, **4**, 392–412.
- 28 J. T. Gorke, F. Srienc and R. J. Kazlauskas, *Chem. Commun.*, 2008, **10**, 1235–1237.

- 29 M. Cvjetko Bubalo, A. Jurinjak Tušek, M. Vinković, K. Radošević, V. Gaurina Srček and I. Radojčić Redovniković, *J. Mol. Catal. B Enzym.*, 2015, **122**, 188–198.
- 30 D. González-Martínez, V. Gotor and V. Gotor-Fernández, *European J. Org. Chem.*, 2016, **2016**, 1513–1519.
- 31 X. Tian, S. Zhang and L. Zheng, *J. Microbiol. Biotechnol.*, 2015, **26**, 80–88.
- 32 B. P. Wu, Q. Wen, H. Xu and Z. Yang, *J. Mol. Catal. B Enzym.*, 2014, **101**, 101–107.
- 33 S. L. Cao, H. Xu, X. H. Li, W. Y. Lou and M. H. Zong, *ACS Sustain. Chem. Eng.*, 2015, **3**, 1589–1599.
- 34 A. Sanchez-Fernandez, K. J. Edler, T. Arnold, D. Alba Venero and A. J. Jackson, *Phys. Chem. Chem. Phys.*, 2017, **19**, 8667–8670.
- 35 C. Ma, A. Laaksonen, C. Liu, X. Lu and X. Ji, *Chem. Soc. Rev.*, 2018, **47**, 8685–8720.
- 36 P. J. Smith, C. B. Arroyo, F. Lopez Hernandez and J. C. Goeltz, *J. Phys. Chem. B*, 2019, **123**, 5302–5306.
- 37 N. Guajardo, H. P. Domínguez de María, K. Ahumada, R. A. Schrebler, R. Ramírez-Tagle, F. A. Crespo and C. Carlesi, *ChemCatChem*, 2017, **9**, 1393–1396.
- 38 E. Durand, J. Lecomte, B. Baréa, E. Dubreucq, R. Lortie and P. Villeneuve, *Green Chem.*, 2013, **15**, 2275–2282.
- 39 H. Zhao, *J. Chem. Technol. Biotechnol.*, 2015, **90**, 19–25.
- 40 M. Sivapragasam, M. Moniruzzaman and M. Goto, *Biotechnol. J.*, 2016, **11**, 1000–1013.

- 41 D. V. Wagle, H. Zhao and G. A. Baker, *Acc. Chem. Res.*, 2014, **47**, 2299–2308.
- 42 T. El Achkar, S. Fourmentin and H. Greige-Gerges, *J. Mol. Liq.*, 2019, **288**, 11028-11030.
- 43 D. Mondal, M. Sharma, C. Mukesh, V. Gupta and K. Prasad, *Chem. Commun.*, 2013, **49**, 9606–9608.
- 44 I. Mamajanov, A. E. Engelhart, H. D. Bean and N. V. Hud, *Angew. Chem. Int. Ed.*, 2010, **49**, 6310–6314.
- 45 I. Gállego, M. A. Grover and N. V. Hud, *Angew. Chem. Int. Ed.*, 2015, **54**, 6765–6769.
- 46 G. Bonner and A. M. Klibanov, *Biotechnol. Bioeng.*, 2000, **68**, 339–344.
- 47 B. Hammouda and D. Worcester, *Biophys. J.*, 2006, **91**, 2237–2242.
- 48 D. D. Albergo and D. H. Turner, *Biochemistry*, 1981, **20**, 1413–1418.
- 49 Y. He and D. R. Liu, *Nat. Nanotechnol.*, 2010, **5**, 778–782.
- 50 W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes, A. H. El-Sagheer, J. Bath, B. G. Davis, T. Brown, R. K. O'Reilly and A. J. Turberfield, *Nat. Chem.*, 2016, **8**, 542–548.
- 51 Y. He and D. R. Liu, *J. Am. Chem. Soc.*, 2011, **133**, 9972–9975.
- 52 V. Stepanov and J. Nyborg, *Extremophiles*, 2002, **6**, 485–490.
- 53 J. R. Peacock, R. R. Walvoord, A. Y. Chang, M. C. Kozlowski, H. Gamper and Y. M. Hou, *RNA*, 2014, **20**, 758–764.
- 54 K. F. Dyer, *J. Biol. Educ.*, 1971, **5**, 15–24.
- 55 I. P. E. Macário, H. Oliveira, A. C. Menezes, S. P. M. Ventura, J. L. Pereira, A. M. M. Gonçalves, J. A. P. Coutinho and F. J. M. Gonçalves, *Sci. Rep.*, 2019, **9**, 3932–3940.

- 56 Y. Xie, H. Dong, S. Zhang, X. Lu and X. Ji, *J. Chem. Eng. Data*, 2014, **59**, 3344–3352.
- 57 F. Gabriele, M. Chiarini, R. Germani, M. Tiecco and N. Spreti, *J. Mol. Liq.*, 2019, **291**, 111301.
- 58 O. S. Hammond, D. T. Bowron and K. J. Edler, *Angew. Chem. Int. Ed.*, 2017, **56**, 9782–9785.
- 59 F. S. Mjalli and O. U. Ahmed, *Asia-Pacific J. Chem. Eng.*, 2016, **11**, 549–557.
- 60 F. S. Mjalli and O. U. Ahmed, *Korean J. Chem. Eng.*, 2016, **33**, 337–343.
- 61 A. P. Abbott, S. S. M. Alabdullah, A. Y. M. Al-Murshedi and K. S. Ryder, *Faraday Discuss.*, 2018, **206**, 365–377.
- 62 C. C. Westcott, *pH Measurements*, Academic Press, New York, 1978.
- 63 A. J. P. Teunissen, C. Pérez-Medina, A. Meijerink and W. J. M. Mulder, *Chem. Soc. Rev.*, 2018, **47**, 7027–7044.
- 64 J. Huang, H. Wang, X. Yang, K. Quan, Y. Yang, L. Ying, N. Xie, M. Ou and K. Wang, *Chem. Sci.*, 2016, **7**, 3829–3835.
- 65 J. Gao, P. Wang and R. W. Giese, *Anal. Chem.*, 2002, **74**, 6397–6401.
- 66 A. A. Neurauter, M. Bonyhadi, E. Lien, L. Nøkleby, E. Ruud, S. Camacho and T. Aarvak, *Cell Separation*, 2007, **106**, 41–73.
- 67 D. Y. Zhang and E. Winfree, *J. Am. Chem. Soc.*, 2009, **131**, 17303–17314.
- 68 N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye and E. Winfree, *Nucleic Acids Res.*, 2013, **41**, 10641–10658.
- 69 J. C. Wang, *Proc. Natl. Acad. Sci.*, 1979, **76**, 200–203.
- 70 B. Yurke and A. P. Mills, in *Genetic Programming and Evolvable Machines*, 2003, **4**, 111–122.

- 71 M. Sundaralingam and P. K. Ponnuswamy, *Biochemistry*, 2004, **43**, 16467–16476.
- 72 A. Chandran, D. Ghoshdastidar and S. Senapati, *J. Am. Chem. Soc.*, 2012, **134**, 20330–20339.
- 73 M. L. McKee, A. C. Evans, S. R. Gerrard, R. K. O'Reilly, A. J. Turberfield and E. Stulz, *Org. Biomol. Chem.*, 2011, **9**, 1661–1666.
- 74 M. Gull, M. Zhou, F. M. Fernández and M. A. Pasek, *J. Mol. Evol.*, 2014, **78**, 109–117.
- 75 B. Burcar, M. Pasek, M. Gull, B. J. Cafferty, F. Velasco, N. V. Hud and C. Menor-Salván, *Angew. Chem. Int. Ed.*, 2016, **55**, 13249–13253.
- 76 S. Ogo, K. Uehara, T. Abura and S. Fukuzumi, *J. Am. Chem. Soc.*, 2004, **126**, 3020–3021.
- 77 B. Burcar, M. Pasek, M. Gull, B. J. Cafferty, F. Velasco, N. V. Hud and C. Menor-Salván, *Angew. Chem. Int. Ed.*, 2016, **55**, 13249–13253.
- 78 J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks and N. A. Pierce, *J. Comput. Chem.*, 2011, **32**, 170–173.

Chapter 3

DNA templated synthesis of peptide bonds in aqueous solution: *in situ* activation of phenol esters

3.1 Introduction: co-localised catalysts and nucleic acid templated synthesis

The two examples of autonomous DNA molecular assemblers known to date (Chapter 1, Section 1.4.5) with the ability to template the synthesis of peptides, rely on *N*-hydroxysuccinimide activated esters.^{1,2} These nanomachines do not incorporate the chemical reaction as part of the required stimuli in order to proceed to the next synthesis step, but instead, they perform two simultaneous processes: the strand displacement via DNA motor and the chemical synthesis. As a direct consequence, the chemical synthesis has to be faster than the DNA strand displacement in order to avoid missing the incorporation of chemical blocks into the peptide product. In this regard, *N*-hydroxysuccinimide activated esters are advantageous, as their high reactivity allows a fast aminolysis reaction during the limited time the ester and the amine are closely co-localised in space. As discussed before, the limitation that *N*-hydroxysuccinimide activated esters, and similar highly reactive groups, impose is that they rapidly hydrolyse in aqueous buffered solutions.

Nature's approach to ribosomal peptide and protein synthesis differs considerably from this approach. Aminoacyl-tRNAs contain ribose esters of amino acids, these are considerably less reactive than *N*-hydroxysuccinimide esters, which also provides

them with high hydrolytic stability (Figure 3-1). This is further increased by the presence of elongation factors (e.g. EF-Tu) that protect the ester bond from hydrolysis.³ Comparing features of esters in aqueous solution, there is an inverse relationship between the reactivity and the resistance to hydrolysis in the absence of catalysis.⁴

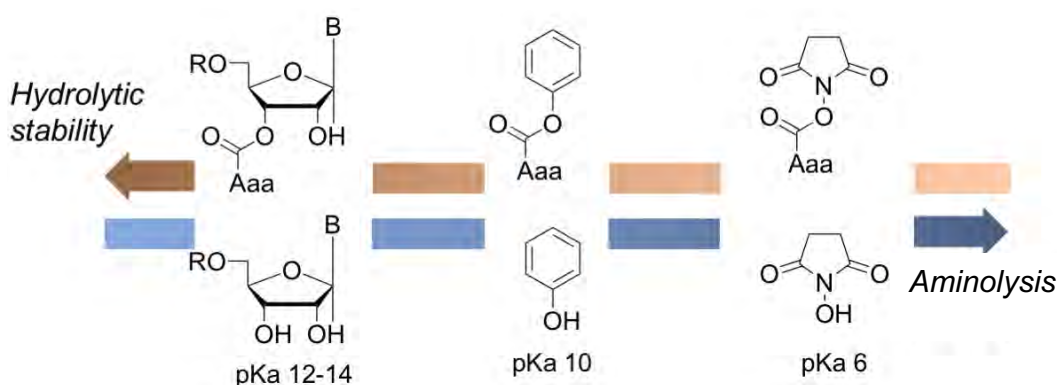


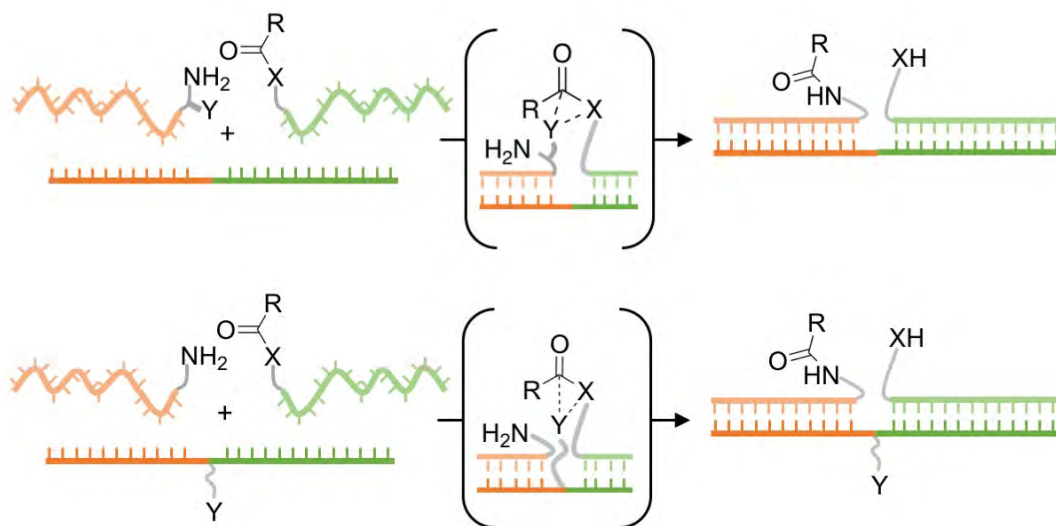
Figure 3-1. Examples of amino acid esters, from low to high reactivity (left to right): ribose, phenol and *N*-hydroxysuccinimide ester.

The ribosome increases the reactivity of ribose esters only when the reactive groups (*i.e.* the acceptor amine and the donor ribose ester) are in close proximity, herein referred to as *in situ* activation.⁵ By its nature, this *in situ* activation does not promote the hydrolysis of the activated ribose esters in solution. The ribosome only catalyses the hydrolysis between the wholly formed peptide and the last ^tRNA as a termination sep.⁶

In addition to the *in situ* activation, the multistep synthesis mechanism in the ribosome also has the remarkable property of being reaction-dependent,^{5,7} this implies that the multistep peptide synthesis does not proceed to the addition of the next amino acid unless the previous peptide bond has been formed. The combination of these effects

means that ribosomal peptide and protein synthesis is catalytic, has a low error frequency and is not affected by the hydrolysis of aminoacyl-tRNAs in solution.

A general approach that incorporates the *in situ* activation principle to DNA templated synthesis of peptides can be generally described as follows (Scheme 3-1).

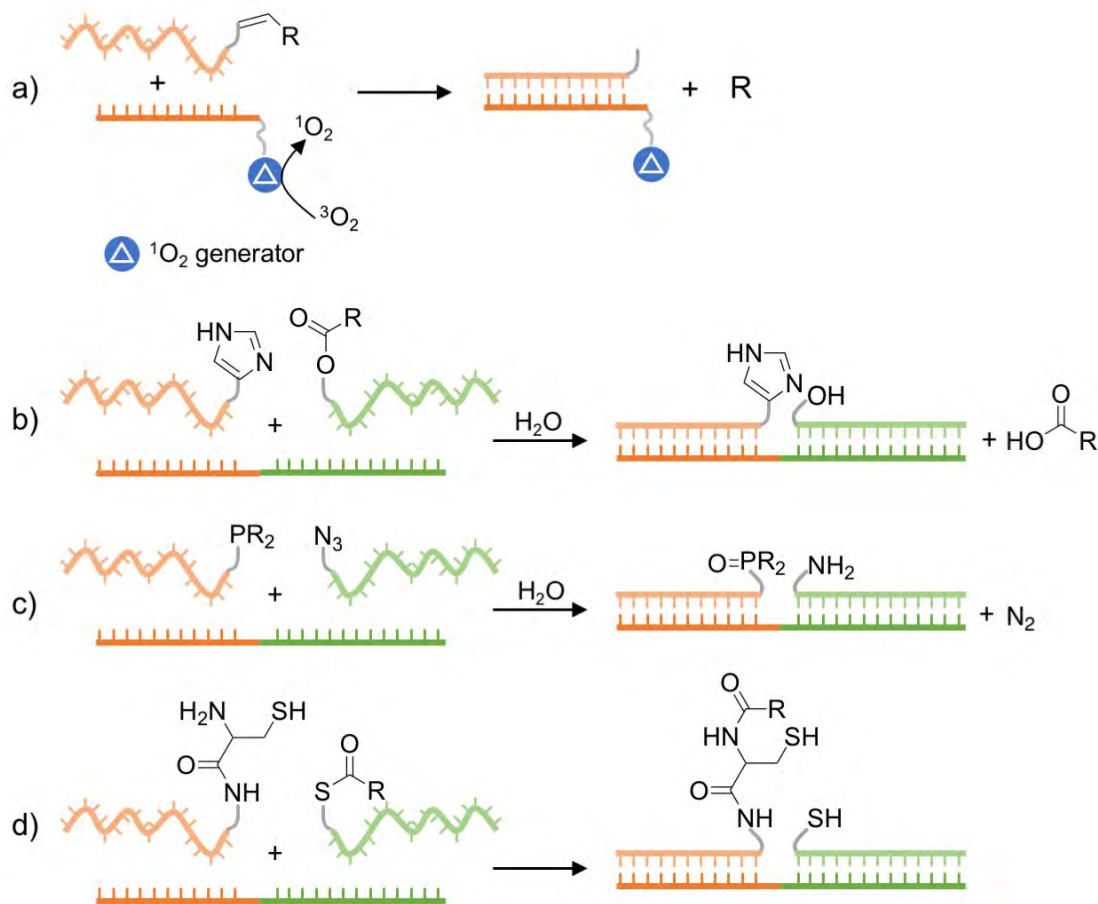


Scheme 3-1. General approach to a DNA templated *in situ* activation of a carbonyl compound (X), by a co-localised catalytic group (Y) that promotes the formation of a peptide bond. The catalytic group can be placed in the acceptor strand (top) or in an independent template strand (bottom).

In this approach, an unreactive acyl donor (group X, Scheme 3-1) would become activated upon hybridisation to a complementary DNA template (group Y, Scheme 3-1). At the same time, the amine-labelled DNA acceptor would be co-localised, ready to react with the activated acyl donor before non-templated hydrolysis takes place.

There are several examples of *in situ* activation of chemical groups towards a particular type of reaction upon co-localisation by DNA templates. A non-exhaustive list includes the co-localisation of a singlet O₂ generator that cleaves an olefin upon DNA hybridisation (Scheme 3-2 a),⁸ the promotion of ester hydrolysis by imidazole-labelled DNA (Scheme 3-2 b),⁹ a Staudinger reaction (Scheme 3-2 c),¹⁰ and the synthesis of peptide bonds by native chemical ligation (NCL, Scheme 3-2 d).¹¹ All these examples

have in common that the chemical groups attached to DNA are stable until the hybridisation of complementary DNA strands co-localises them in close proximity.



Scheme 3-2. Representation of templated reactions that require the co-localisation of an activating group. a) Singlet O_2 generator that cleaves an olefin.⁸ b) Imidazole catalysed ester hydrolysis.⁹ c) Staudinger reaction.¹⁰ d) Native chemical ligation.¹¹

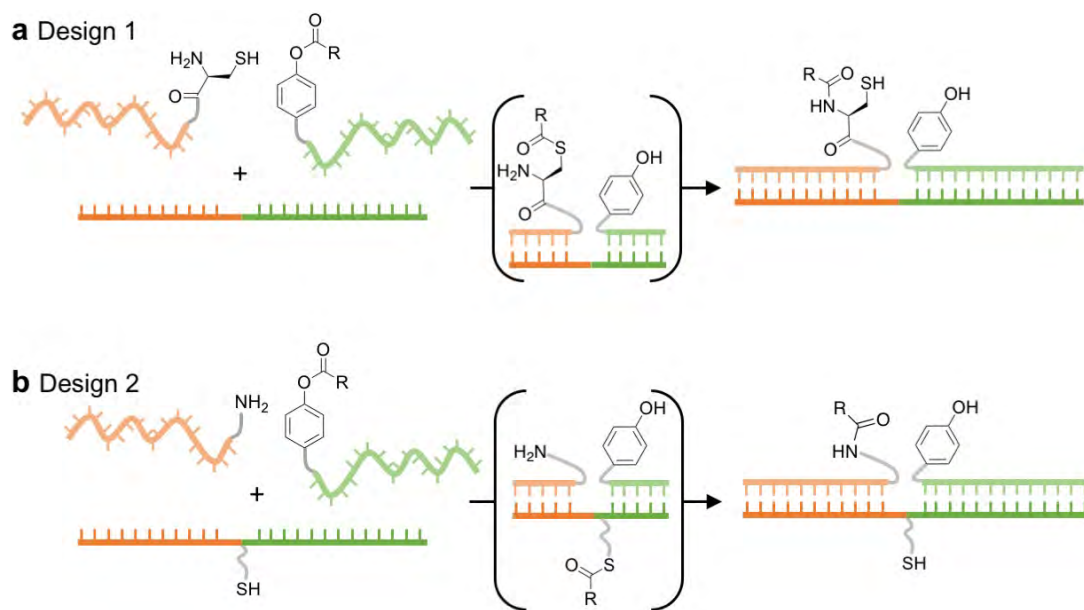
The previous examples show an unreactive functional group that, upon co-localisation to an activating group, becomes highly reactive and produces a chemical output. Templated native chemical ligation (NCL, Scheme 3-2 d) is an example of applying this principle to the DNA templated synthesis (DTS) of peptide bonds. However, NCL uses thioesters, which are prone to undergoing hydrolysis in aqueous solution, much like oxo-esters.¹² Interestingly, oxo-ester mediated NCL using phenol esters, which are very resistant to hydrolysis, has been previously described.^{13–15} Oxo-ester mediated

NCL is an interesting methodology for the synthesis of peptide bonds because working with alcohols is usually more straightforward from the synthetic perspective, than working with thiols. Furthermore, oxo-ester mediated NCL was successfully applied to a rotaxane templated synthesis of α and β peptides in organic solvent, using phenol ester amino acid donors.¹⁶⁻¹⁸ This approach is more similar to nature's approach for the biosynthesis of peptides, as it uses a highly stable and unreactive ester, comparable to ribose esters in aminoacyl-tRNAs (Figure 3-1). This precursor becomes activated by the co-localisation of a catalyst that promote the peptide bond formation. While NCL has limitations in terms of slow kinetics,¹⁹ strong dependence on the identity of the amino acid ester,¹⁹ and a progressive reduction of the kinetics upon peptide growth,¹⁹ the rotaxane templated NCL was an encouraging starting point that showed the potential for activating *in situ* chemically stable precursors. While the previous results were only reported in organic solvent, a DNA templated oxo-ester mediated NCL using stable phenol esters in aqueous buffer would overcome the hydrolysis of the DNA adapters.

3.2 Results and discussion

3.2.1 Design of the NCL system

Typically, NCL relies on a cysteine (Cys) amino acid acceptor (Scheme 3-2 d), which contains a thiol and an amine functional group. The thiol has a catalytic role, which is to convert the oxo-ester into a thioester, which is then able to form the desired amide bond (Chapter 1 Section 1.4.2). Where the final acceptor group is the primary amine in the amino acid. In the design of a DNA NCL catalytic system, it is, however, possible to separate these two groups onto different DNA strands (Scheme 3-3 b).



Scheme 3-3. Two proposed designs of DNA templated NCL systems showing the thioester intermediate. a) “Design 1” Cys acceptor and unfunctionalised DNA template b) “Design 2” amine acceptor and thiol-labelled template.

A priori, it would seem that the Cys acceptor group in design 1 has an advantageous geometrical disposition of the functional groups in the space, so that the thioester intermediate can evolve into the amide final product through a 5-membered ring conformation (Figure 3-2), which would favour faster reaction kinetics.²⁰ However, localisation of the catalytic thiol group on an independent template strand in design 2 (Scheme 3-3 b) would have the advantage of being separate from the acceptor strand, thus facilitating the design of multistep hybridisation mechanisms. For this reason, both approaches were explored in this work.

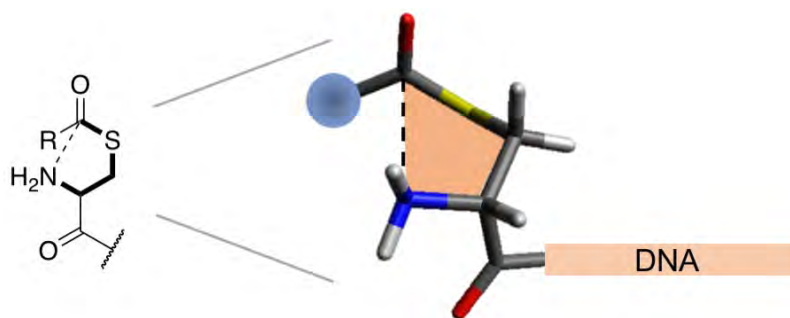
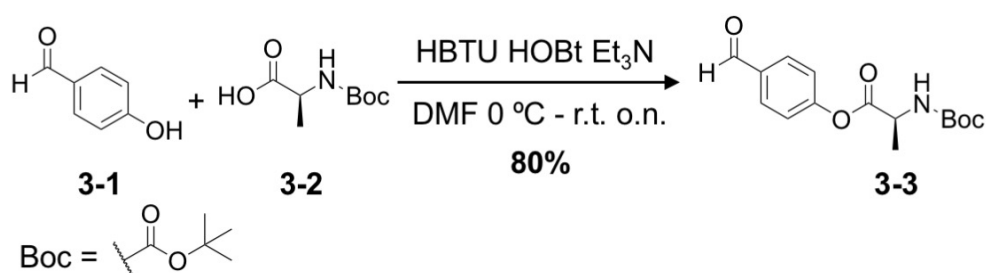


Figure 3-2. Representation of the local geometry around the Cys thioester intermediate in a NCL reaction.[§]

In order to assay DNA templated *in situ* activation, design 1 requires Cys-labelled DNA acceptor, a phenol ester-labelled DNA donor and unlabelled DNA template; design 2 requires commercially available amine-labelled DNA acceptor, a phenol ester-labelled DNA donor and a thiol-labelled template. The subsequent sections will focus on the preparation of these components prior to exploiting the reactivity of designs 1 and 2.

3.2.2 Synthesis of an *L*-Ala phenol ester donor strand

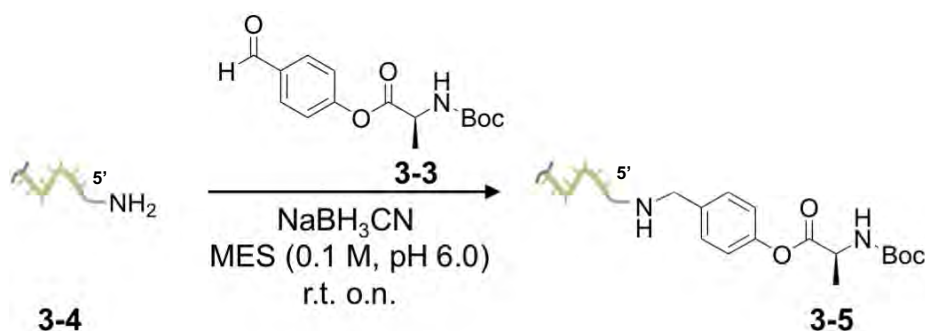
The first transferable group that was assayed was a Boc protected *L*-Ala phenol ester (**3-3**). An aldehyde functional group was placed at the *para* position of the ester, to facilitate the attachment to amine-labelled DNA using reductive amination. This small molecule reagent was prepared by carbodiimide coupling (Scheme 3-4).



Scheme 3-4. Synthesis of a Boc protected *L*-Ala phenol ester (**3**, Experimental section 3.4.3)

[§] Optimisation of the geometry was performed using Avogadro v1.1.1 by molecular mechanics (ghemical force field).

An amine-labelled DNA strand (**3-4**) was then reacted with **3-3**, in the presence of the mild reducing agent NaBH₃CN to produce the DNA conjugated ester **3-5** (Scheme 3-5). Importantly, the resulting secondary amine is conformationally restricted and therefore, it cannot undergo self aminolysis intramolecularly. The intermolecular self aminolysis can be neglected due to the very low rate of untemplated reaction. It is worth noting that, together with the carbodiimide mediated amide coupling, reductive amination was a very useful approach to DNA functionalisation using readily available amine-labelled DNA.



Scheme 3-5. Synthesis of the donor strand DNA-PhO-Ala-Boc (**3-5**, Experimental section 3.4.4) by reductive amination.

This reaction proceeded in very high conversion (79%, HPLC signal integration), and the stability of the phenol ester, as opposed to *N*-hydroxysuccinimide esters, allowed HPLC elution and fraction-collection to generate a pure starting material **3-5** (Figure 3-3).

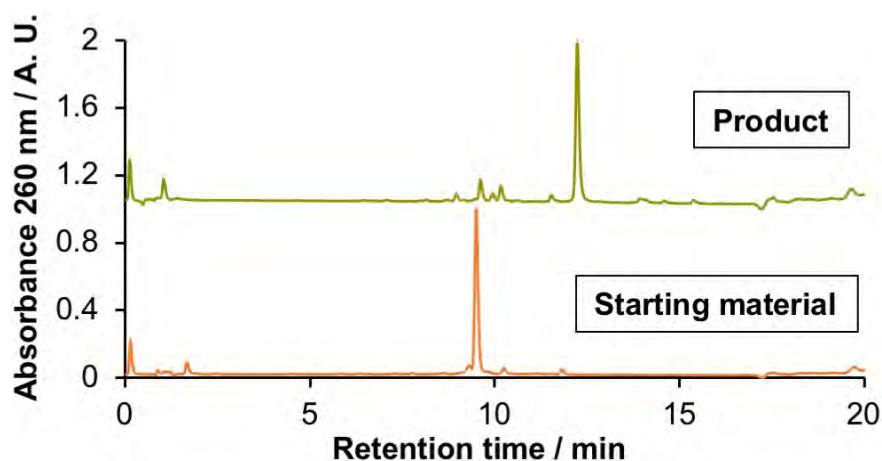
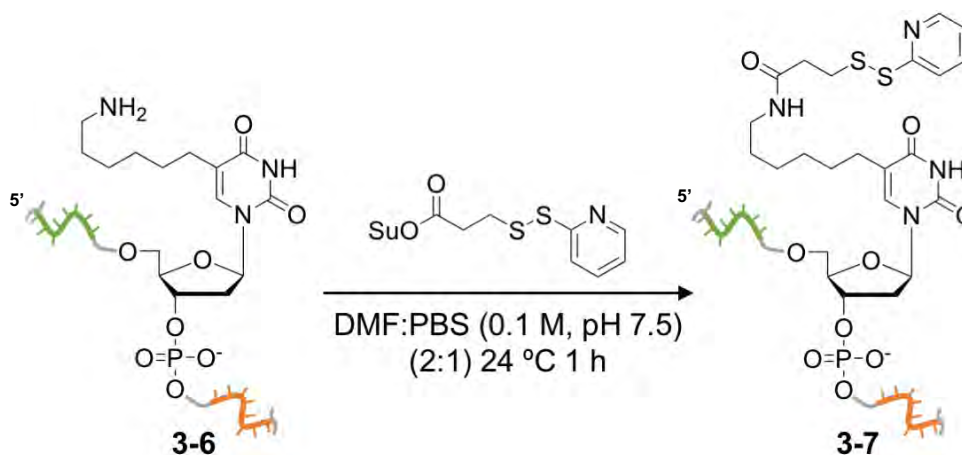


Figure 3-3. HPLC analysis of the synthesis of **3-5**. The product (**3-5**) was identified by LC-MS (Bruker AmazonX): **3-5** $[M-H]^-$ m/z calcd. 9705.8 found 9698.4.

3.2.3 Synthesis of the catalyst and acceptor group

Both Cys-labelled acceptor for design 1, and thiol-labelled catalyst for design 2 require the addition of a protected thiol group either in the form of a cysteine amino acid or an aliphatic thiol. In order to synthesise a thiol-labelled DNA template (Scheme 3-3 b), an internally labelled amino-DNA (**3-6**) was used. In this design, each half of the template was fully complementary to the corresponding donor and acceptor strands, with a central, non-hybridised T base modified with an amine. This amine was reacted with a succinimidyl ester of 3-(2-pyridyldithio)propionate (Scheme 3-6).



Scheme 3-6. Synthesis of an internal thiol-labelled DNA template (**3-7**, Experimental section 3.4.5).

The pyridyldithio group was particularly suitable protecting group, as it is easily cleaved in the presence of diluted tris(2-carboxyethyl)phosphine (TCEP), which is compatible with DNA. This reaction proceeded in high conversion (>80%, HPLC integration). The DNA conjugate **3-7** was analysed and HPLC fraction-collected to produce a pure sample (Figure 3-4).

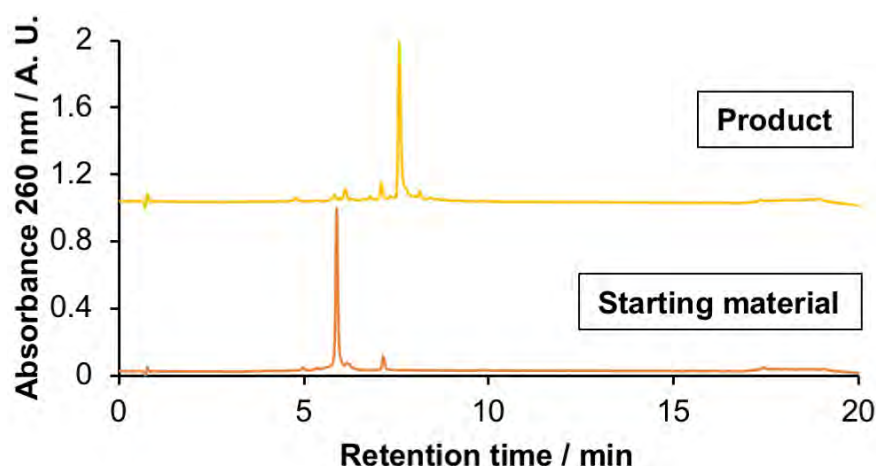
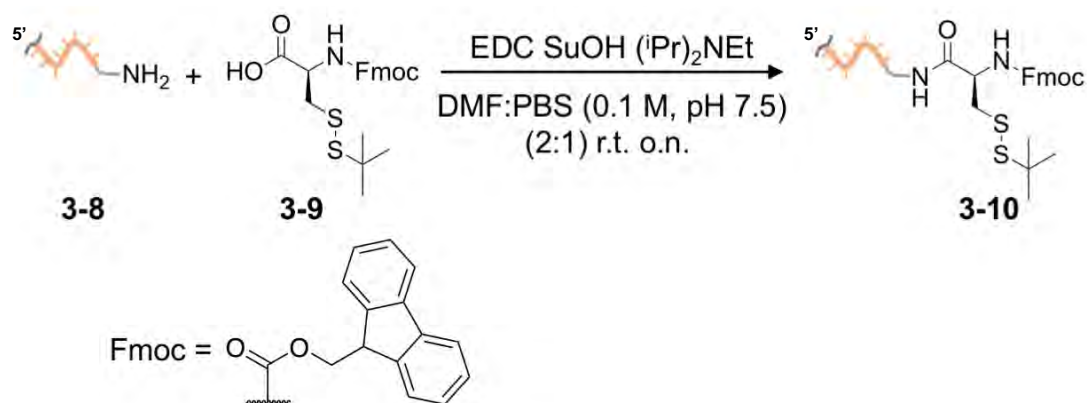


Figure 3-4. HPLC analysis of the synthesis of **3-7**. The product (**3-7**) was identified by LC-MS (Waters Xevo-G2-XS): **3-7** [M]⁰ m/z calcd. 14443.433 found 14444.028.

Both the side chain and the backbone amine in Cys require protection. The combination of a disulfide / fluorenylmethyloxycarbonyl (Fmoc) protecting groups (Scheme 3-7, **3-9**) was identified as the best strategy. This made it possible to isolate a thiol-protected conjugate that could be deprotected *in situ* during the DNA templated NCL. The small molecule amino acid (**3-9**) was conjugated to DNA by carbodiimide coupling, and formed the DNA conjugate **3-10** (Scheme 3-7).



Scheme 3-7. Synthesis of a protected DNA-Cys(SS'Bu)-(NHFmoc) conjugate (**3-10**, Experimental section 3.4.6).

The DNA conjugate **3-10** was considerably hydrophobic, and it displayed a long retention time in reversed-phase HPLC analysis (Figure 3-5). The chromatogram indicated quantitative conversion, and the presence of a hydrophobic small molecule corresponding to one of the possible isomers of HO-Cys-Cys(S'Bu)-Cys(S'Bu)-Cys(S'Bu)-NH-Fmoc (LC-MS analysis). In order to isolate **3-10** from the small molecule, the DNA product was HPLC fraction-collected.

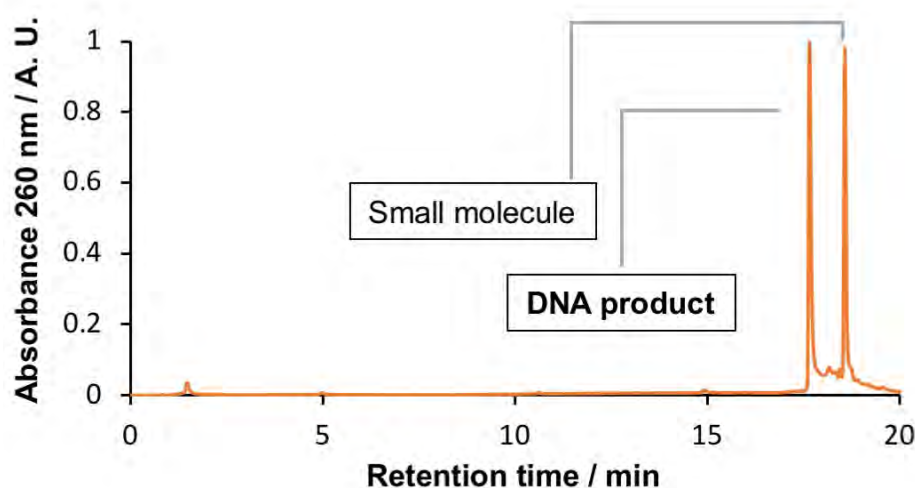
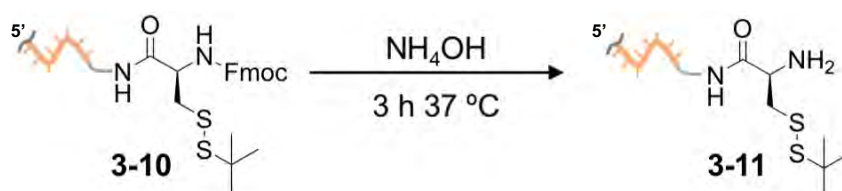


Figure 3-5. HPLC analysis of the synthesis of **3-10**. The product (**3-10**) and the small molecule side product were identified by LC-MS (Waters Xevo-G2-XS): **3-10** [M]⁰ m/z calcd. 6734.242 found 6734.232; **small molecule** [M-H]⁻ m/z calcd. 861.3429 found 861.2935.

Finally, the Fmoc protected amine was deprotected, adapting a reported literature procedure.^{21,22} The use of concentrated aqueous ammonia is a conventional cleavage reagent in DNA solid phase synthesis, meaning it is not expected to damage the DNA.²³



Scheme 3-8. Deprotection of the DNA conjugate 3-10.

In the first deprotection attempts, a 6 h deprotection in concentrated aqueous ammonia at 37 °C, as per the original procedure, generated a large fraction of the thiol-deprotected DNA conjugate, up to 22% (Figure 3-6 a, HPLC integration). Reducing the deprotection time from 6 to 3 h, allowed to completely suppress the thiol deprotection, while still quantitatively completing the Fmoc cleavage (Figure 3-6 b).

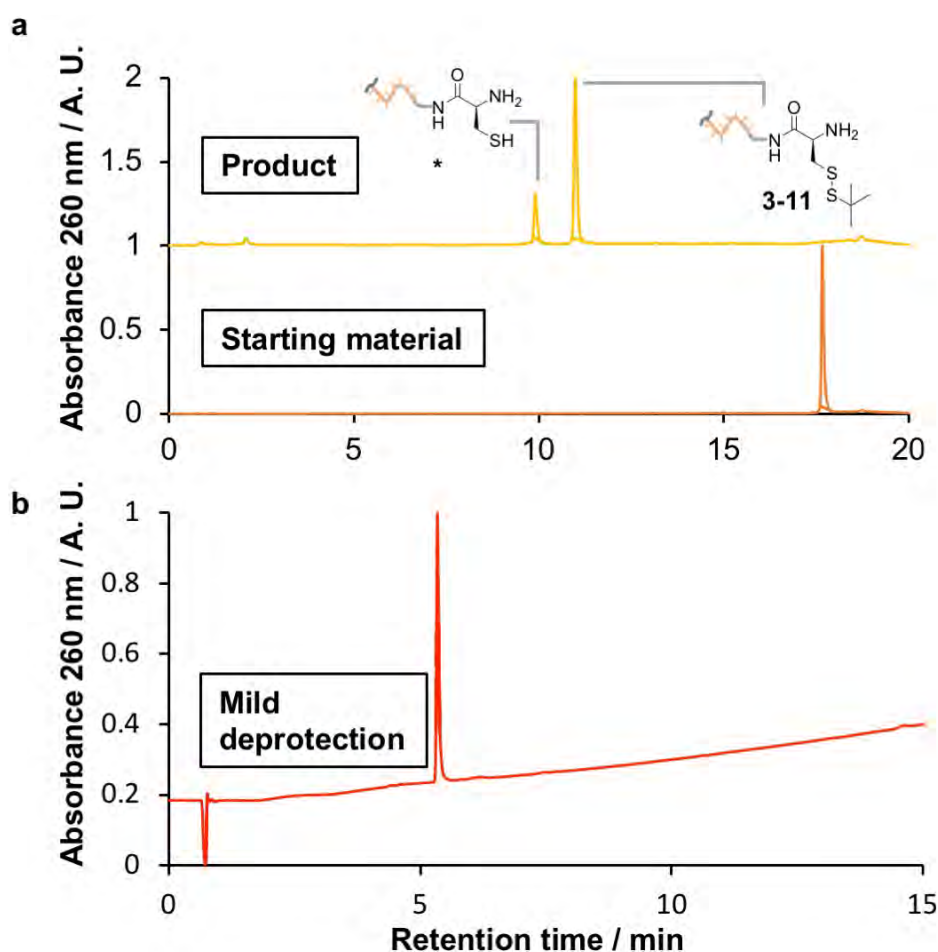
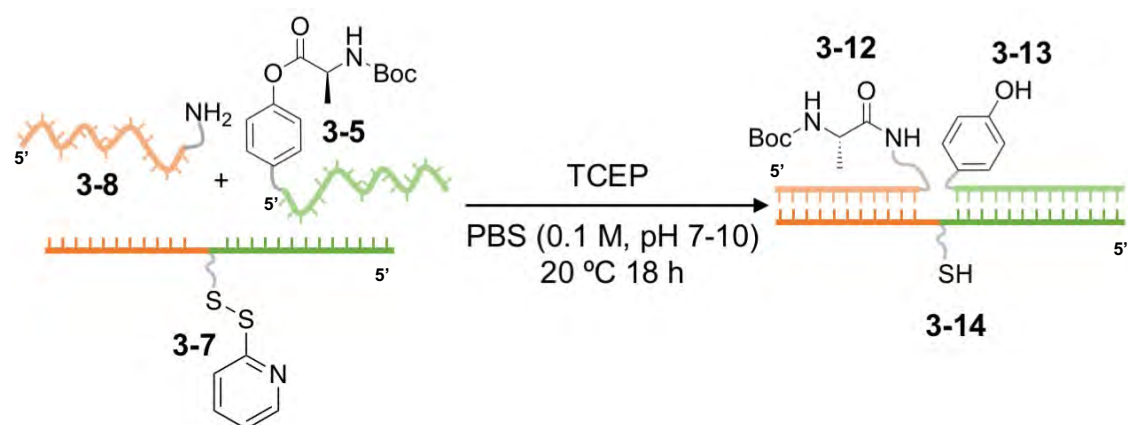


Figure 3-6. HPLC analysis of the Fmoc deprotection of **3-10**. a) The main product (**3-10**) and the side, fully deprotected side product (*) was identified by LC-MS (Waters Xevo-G2-XS): **3-11** [M^0 m/z calcd. 6512.164 found 6512.356; * [M^0 m/z calcd. 6446.198 found 6446.390]. b) Mild deprotection 3 h at 37 °C showed conversion exclusively into **3-11** (LC-MS).

3.2.4. Assay of a templated NCL reaction using a *L*-Ala phenol ester donor

Initially, the DNA templated NCL was attempted using the catalytic template approach (Scheme 3-9), with *L*-Ala-Boc transferable group. The DNA strands were mixed in equimolar amounts, in the presence of TCEP at varying pHs (7.0 to 10.0). This is because pH is one of the key factors that can affect the conversion in NCL,²⁴ as it can alter the fraction of protonated amine, deprotonated thiol, and the mechanism of peptide bond formation.



Scheme 3-9. DNA templated NCL using a co-localised thiol catalyst (**3-7**), a Boc protected *L*-Ala phenol ester (**3-5**) and an amine-labelled acceptor (**3-8**, Experimental section 3.4.7). DTS was performed at 10 μ M.

The outcome of the assay was studied by denaturing PAGE (Figure 3-7). It was expected that, if the product **3-12** was present in the sample, it would migrate slightly slower than the starting material **3-8** due to the slower mobility through the gel pores of larger products, and therefore the conversion could be estimated by densitometry upon staining of the gel with a DNA selective stain.

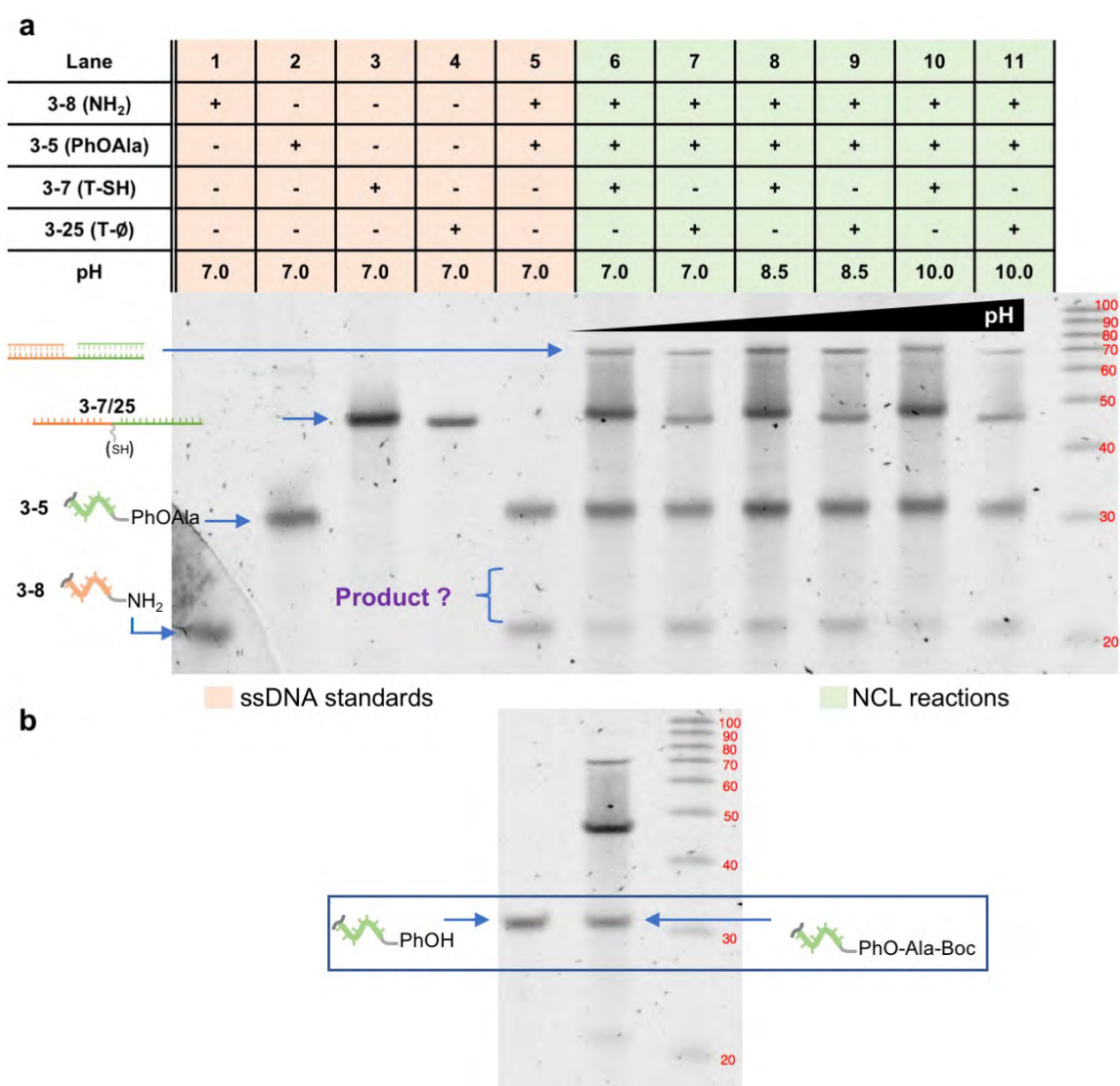


Figure 3-7. Denaturing 20% PAGE analysis of a DNA templated NCL using a co-localised thiol catalyst (**3-7**, T-SH), a *L*-Ala phenol ester (**3-5**, PhOAla) and an amine-labelled acceptor (**3-8**, NH₂). a) Analysis of the single-stranded DNA (ssDNA) standards and reaction mixtures containing thiol catalyst (**3-8**) and catalyst-free template (**3-25**, T-∅) at several pHs. b) Demonstration of the same migrating properties of the alanine phenol ester-labelled DNA (**3-5**) and the same strand in the absence of the modification (**3-4**). ssDNA ladder was added, with 20 to 100 nucleotides.

The denaturing PAGE analysis (Figure 3-7 a) showed clear ssDNA standards (lanes 1-4), and a small fraction of DNA that does not denature and remains in the dsDNA form (lanes 6-11, top band). Based on the assumption that product **3-12** would produce a band close and higher than **3-8**, it would seem that no product had been formed.

However, by comparing the starting material (**3-4**) and the same strand containing the *L*-Ala phenol ester (**3-5**) it was evident that they are not distinguishable even using a high acrylamide % resolving gel (Figure 3-7 b). This observation made it clear that the transferable group should be a chemical probe with a specific property that made it distinguishable from the other DNA strands and small molecules.

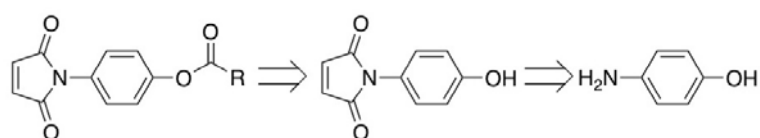
3.2.5. Synthesis of a *D*-biotin phenol ester donor strand

In the previous chapter, a chemical probe was prepared by labelling an *L*-Ala amino acid with a fluorophore. Whilst fluorescence is a property that would make it possible to determine if the NCL had taken place, it would not allow separation of the DNA product strand from the starting material by gel electrophoresis, due to the similarities in size between the starting material and the product. An alternative approach was therefore going to be necessary.

D-biotin is a small molecule, sometimes referred to as vitamin B₇, which is present in a number of natural sources, such as egg yolk and dried fruits, amongst others. It has one of the strongest known non-covalent interactions with the tetrameric proteins avidin and streptavidin ($K_d \sim 10^{-15}$ M).²⁵ The biotin-streptavidin interaction has been commonly used in chemical biology to determine the outcome of an experiment thanks to the robustness of the interaction and the chemical stability of biotin to the reaction conditions.²⁶

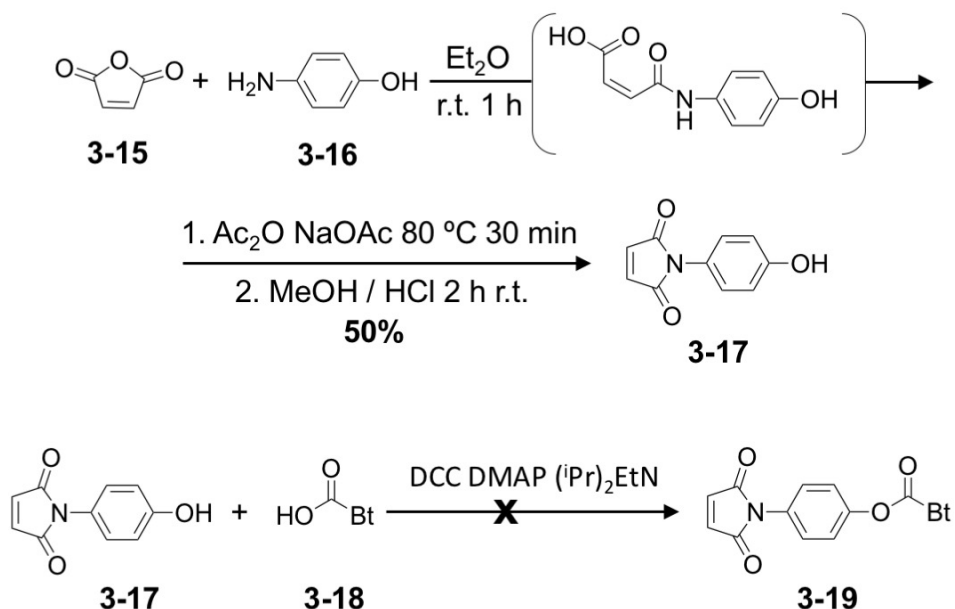
As streptavidin is a considerably large protein (~53 kDa),²⁷ in comparison to the DNA system (dsDNA formed from strands **3-5**, **3-7** and **3-8**) the conjugation of streptavidin to biotin, after DNA templated NCL, would provide a very clear size difference between the unreacted product and the desired peptide bond in a PAGE experiment. Therefore, it was proposed that the extent of the templated reaction could be estimated by

densitometric analysis, analogously to a previously reported procedure.²⁸ Here, a retrosynthesis to prepare a *D*-biotin phenol ester was performed (Scheme 3-10). The structure includes a maleimide moiety, that would react with thiol-labelled DNA in a thiol-Michael addition; and a phenol ester. The proposed disconnections were: 1) The ester bond, which can be prepared by reaction between the free phenol and a *D*-biotin acyl donor; 2) The maleimide ring, which can be prepared by reaction between an amine and maleic anhydride.²⁹



Scheme 3-10. Retrosynthetic analysis of a maleimide phenol ester (*i.e.* R contains biotin).

The synthesis of the maleimide moiety was performed according to the literature procedure (Scheme 3-11).²⁹ Maleic anhydride (**3-15**) was reacted with *p*-hydroxyaniline (**3-16**) to produce the desired product in moderate yield (**3-17**).



Scheme 3-11. Synthesis of maleimide intermediate **3-17** followed by attempted Steglich esterification (Experimental section 3.4.8).

D-biotin free carboxylic acid was reacted with **3-17** in Steglich esterification conditions;³⁰ however, the expected product (**3-19**) could not be detected by ¹H-NMR spectroscopy. After a series of failed syntheses, to prepare **3-19**, the effect of individual reagent and solvent on the stability of **3-17** was studied by ¹H-NMR spectroscopy (Figure 3-8).

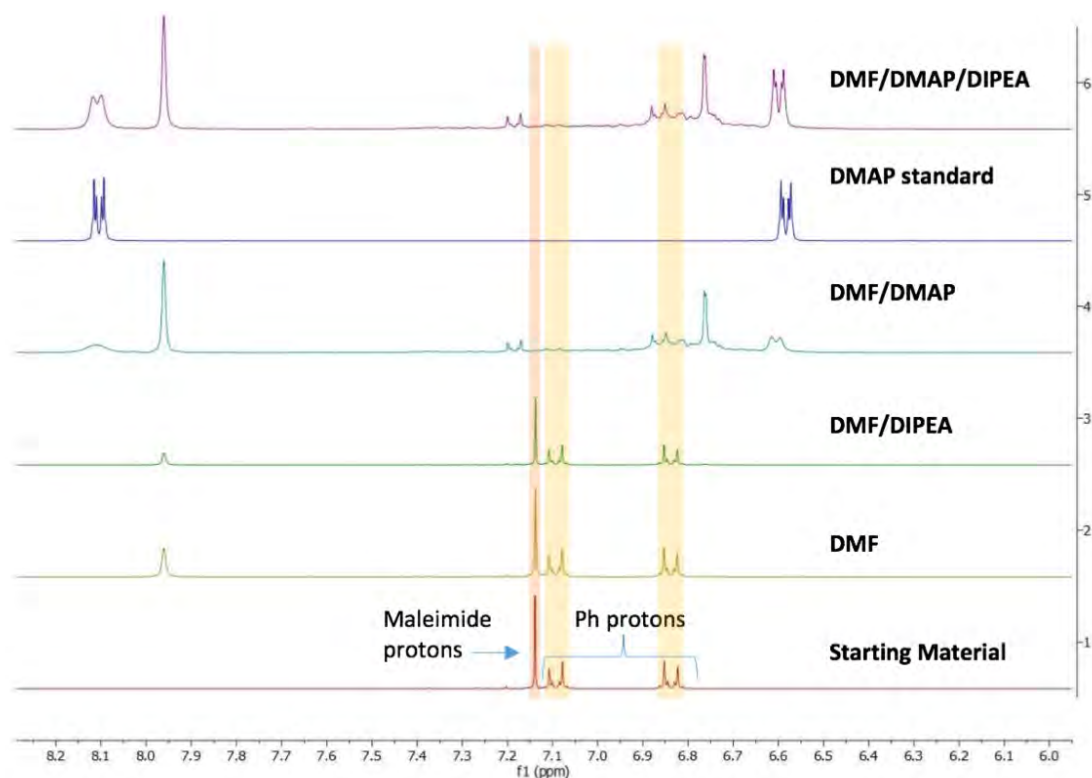
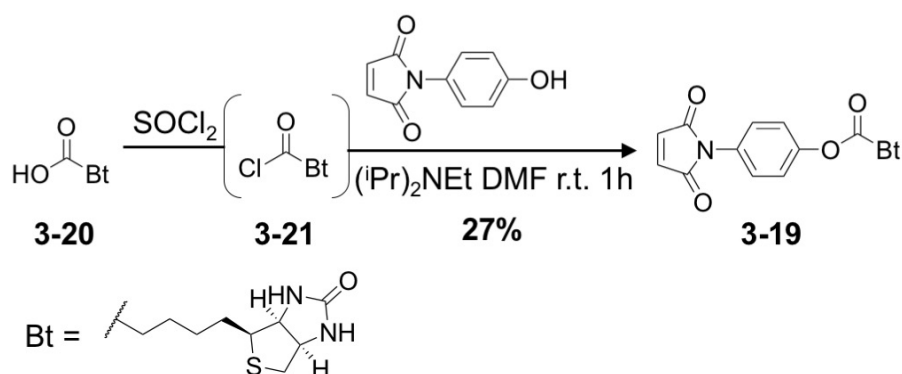


Figure 3-8. ¹H-NMR spectroscopic study of the stability of **3-17** in the presence of the Steglich esterification reagents. 1) **3-17** pure in deuterated solvent; 2) **3-17** dissolved in DMF, dried and dissolved in deuterated solvent; 3) **3-17** dissolved in a DMF / (ⁱPr)₂NEt mixture, dried and dissolved in deuterated solvent; 4) **3-17** dissolved in a DMF / 4-dimethylamino pyridine (DMAP), dried and dissolved in deuterated solvent; 5) DMAP standard dissolved in deuterated solvent; 6) **3-17** dissolved in a DMF / (ⁱPr)₂NEt mixture, dried and dissolved in deuterated solvent (Experimental section 3.4.9).

From the previous data, it became clear that a catalytic amount of DMAP, which is essential for a successful esterification, was causing decomposition of the starting material. The maleimide H signal (δ 7.12 ppm, singlet) disappears in the presence of

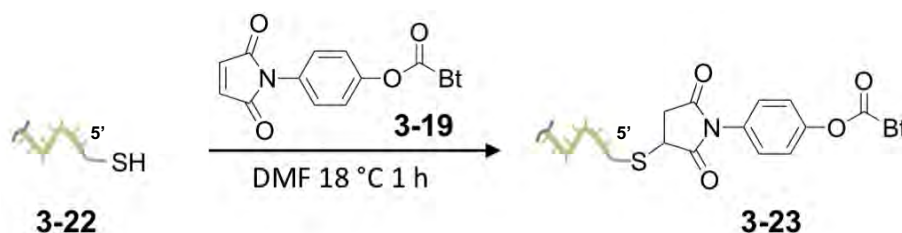
DMAP in DMF. Interestingly, no previous reports of maleimide decomposition in the presence of DMAP were found in the literature.

In an alternative approach, *D*-biotin acyl chloride (**3-21**) was synthesised with an excess of thionyl chloride, the excess of reagent was removed under vacuum and the resulting crude containing the activated acyl donor was directly reacted with **3-17** to afford the expected product (**3-19**, Scheme 3-12).



Scheme 3-12. Synthesis of **3-19** utilising *D*-biotin acyl chloride (**3-21**) intermediate (Experimental Section 3.4.8).

The product **3-19** was directly reacted with a thiol-labelled DNA (**3-22**, Scheme 3-13) without further purification, to produce the phenol ester-labelled DNA conjugate **3-23**.



Scheme 3-13. Thiol-Michael synthesis of a biotin phenol ester donor (**3-23**, Experimental section 3.4.8)

The resulting biotin phenol ester (**3-23**) was analysed by native PAGE after binding to an excess of streptavidin (Figure 3-9 a). The specificity of the Sybr Gold (Thermofisher) staining towards nucleic acids allowed visualisation of DNA bands by PAGE, in the presence of an excess of streptavidin. The densitometric analysis revealed an

approximate 28% of expected product. The sample was then analysed by HPLC and LC-MS, and the chromatographic analysis revealed a 37% of expected product (**3-23**) and a 63% of phenol (*, Figure 3-9 b). The discrepancy between the two analyses can be explained by different degrees of hydrolysis during the electrophoresis and the chromatography. The lower accuracy of densitometric analysis might also be a factor to consider.

The product **3-23** was HPLC fraction-collected, this allowed isolation of enough material to perform the required experiments; however, it also limited the maximum amount of material that could be isolated due to the scale limitations of analytical HPLC. The extremely high sensitivity of PAGE analysis, combined with the clear differentiation of the product after binding to streptavidin proved an advantageous strategy to study DNA templated NCL; however, it did not allow analysis of the reaction outcome by LC-MS as the available amount of product was below the limit of detection.

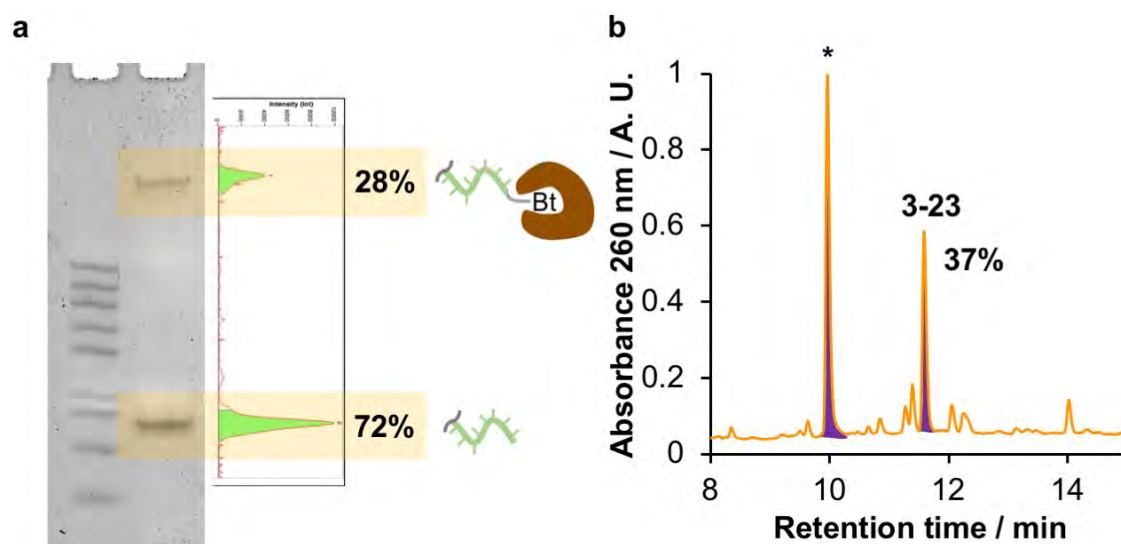
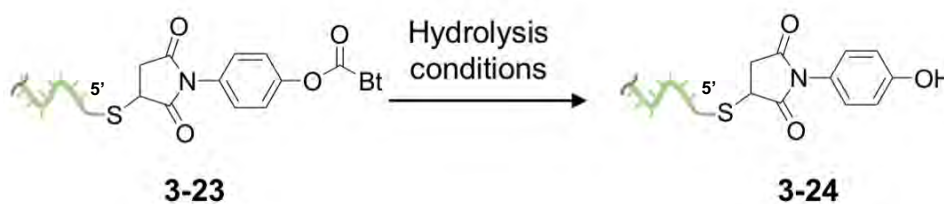


Figure 3-9. Quantification of the conversion into **3-23**. a) Native PAGE analysis of the synthesis of **3-23** followed by conjugated to streptavidin. The conversion was estimated by densitometry. b) HPLC analysis of the synthesis of **3-23**. The products were identified by LC-MS (Waters Xevo-G2-XS): * DNA-Ph-OH [M]⁰ m/z calcd. 9634.663 found 9634.339; **3-23** [M]⁰ m/z calcd. 9860.741 found 9860.394.

With the DNA conjugate **3-23**, the Cys-labelled DNA **3-10**, and the catalyst template **3-7**, all the necessary elements to perform templated reactions were ready. However, in order to confidently determine the conversion after DTS, the unreacted **3-23** had to be hydrolysed.

3.2.6 Optimisation of the post-DTS hydrolysis conditions

Even if the templated NCL did not proceed to full conversion, the resistance to hydrolysis of phenol esters would allow the detection of a large fraction of biotin-labelled DNA starting material (**3-23**) instead of the expected amide. As the starting material would bind streptavidin, it would be indistinguishable from the amide product. In order to confidently determine the peptide bond formation, the sample would have to be subject to hydrolysis conditions (Scheme 3-14).



Scheme 3-14. Optimisation of the hydrolysis conditions of **3-23** (Experimental Section 3.4.10). Hydrolysis was performed at 0.1 μ M.

In order to ensure the stability of DNA under hydrolysis conditions,³¹ as well as to determine the optimal temperature and time, the isolated sample **3-23** was subject to increasingly harsh hydrolysis conditions and analysed by native PAGE (Figure 3-10 a and b). After several attempts, it was found that hydrolysing the DNA conjugate for 3 h at 70 °C and in pH 12, cleaved the vast majority of the ester (Figure 3-10 a, lane 3). In addition, these conditions did not have a negative impact on the length of the DNA strands indicating no hydrolysis of the DNA backbone, and did not reduce the ability of DNA to hybridise (Figure 3-10 a, lanes 8-9).

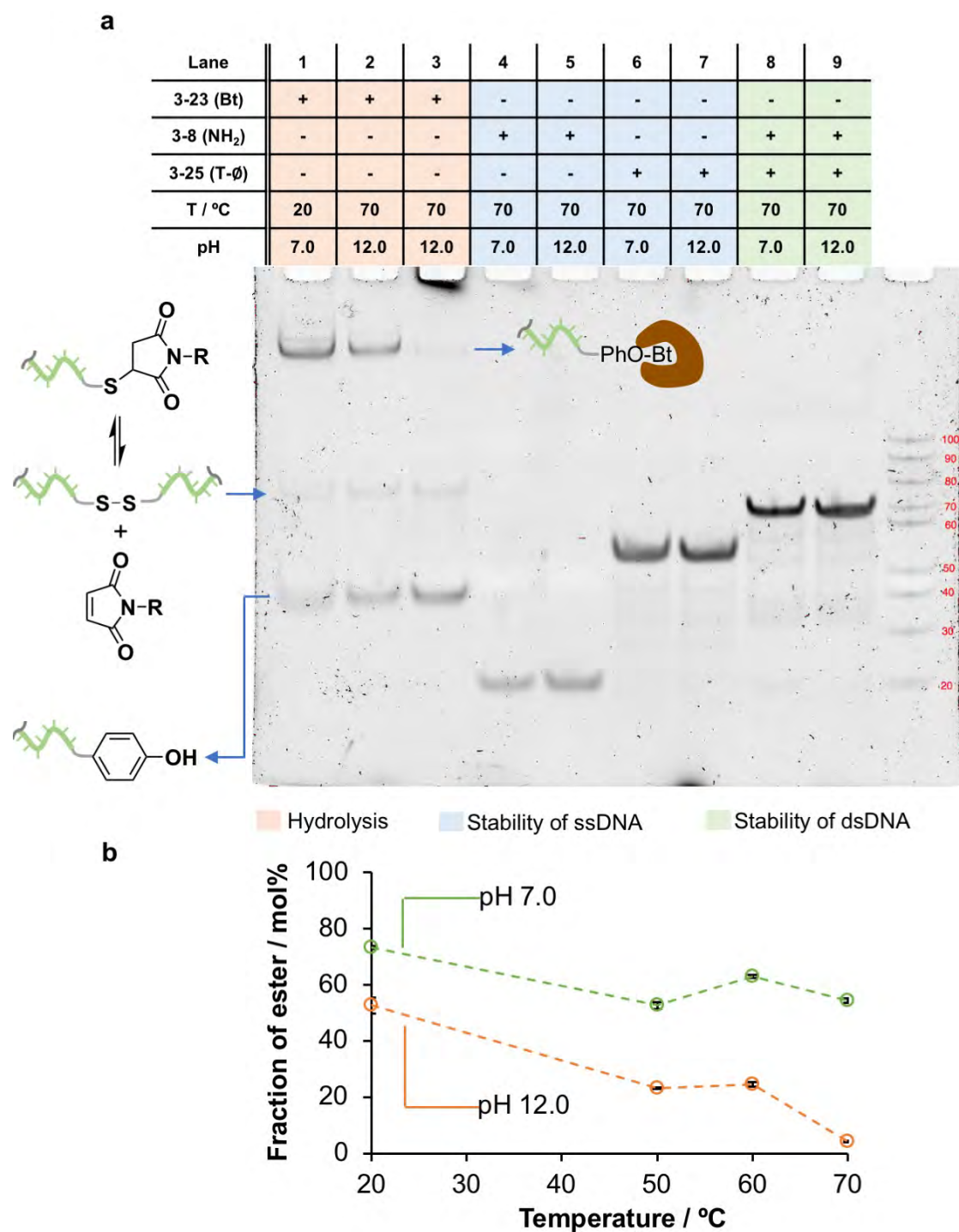
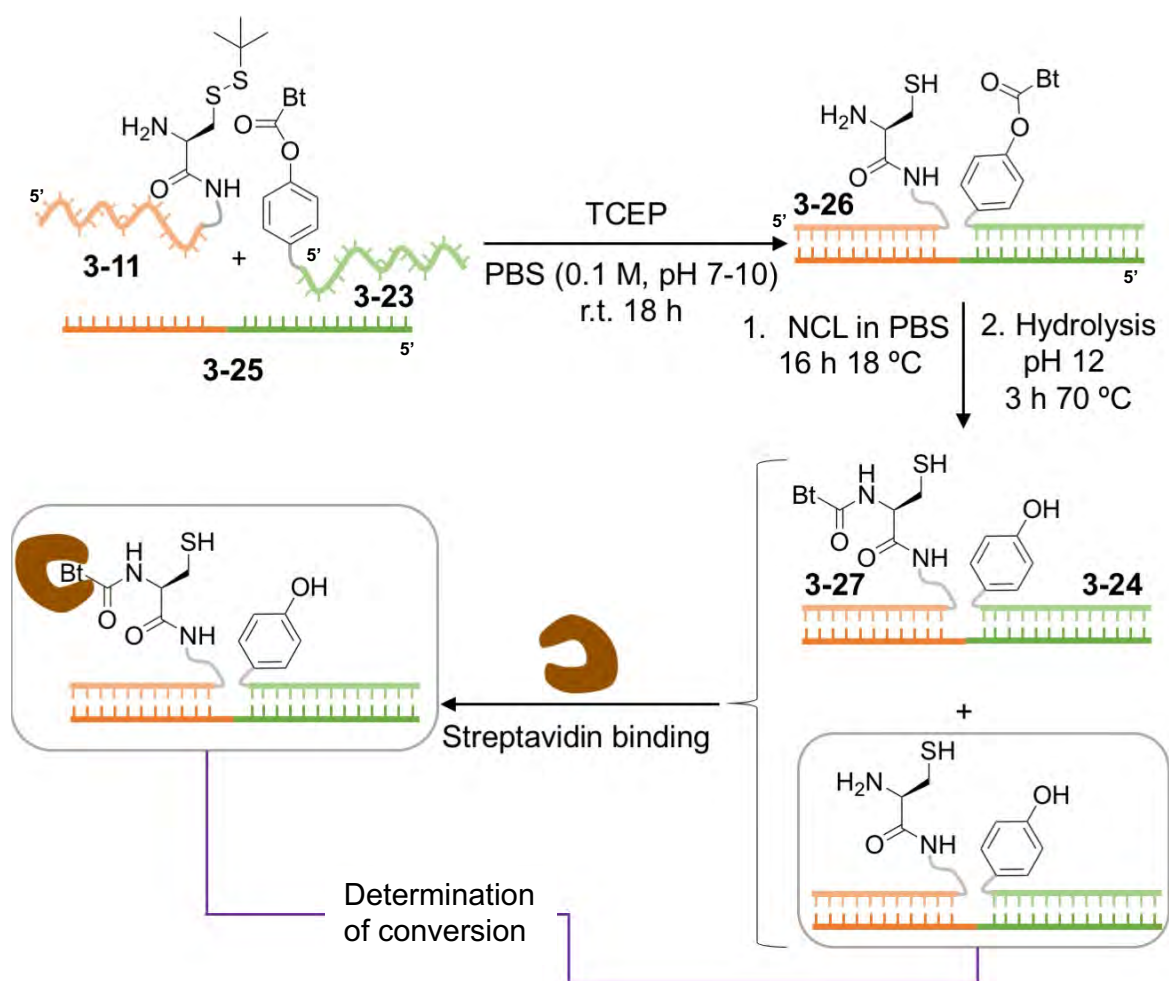


Figure 3-10. Optimisation of the conditions for the hydrolysis of **3-23**. a) Native PAGE analysis after hydrolysis at 70 °C for 3 h. b) Fraction of remaining ester after 3 h hydrolysis at different temperatures and pH. The fraction of ester was estimated by densitometry, as the proportion between the streptavidin-bound **3-23** (top band, lanes 1-3) and the hydrolysed product **3-24** (bottom band, lanes 1-3). Multiple integration events were performed, the error bars correspond to the standard deviation.

It was therefore concluded that a post-reaction hydrolysis step could be safely applied. The streptavidin-bound material was unambiguously characterised by comparing the PAGE gel in the presence or in the absence of streptavidin; with DNA-streptavidin conjugates showing a distinct relative mobility (R_f) lower than the 100 nucleotides (nt) band in the DNA ladder (Experimental section 3.4.8 and Figure 3-20). A faint band of DNA corresponding to a strand with twice the length of the starting material was also detected. It was hypothesised that a retro-Michael addition followed by disulfide formation could have taken place (Figure 3-10 lanes 1-3).

3.2.7 Design 1: NCL mediated biotin transfer

With the set of tools developed in the previous sections, it was possible to do a first attempt of the DNA templated NCL. In this case, as opposed to the attempt using the *L*-Ala transferable group (Section 3.2.4), a Cys-labelled DNA acceptor was used. The catalysis of Cys was expected to be more efficient than the catalysis of a co-localised thiol (e.g. DNA template **3-7**) according to the argument made above (Figure 3-2). The thiol-protected acceptor **3-11**, the donor **3-23** and the unfunctionalised DNA template **3-25** were isothermally annealed, and the NCL reaction was triggered by *in situ* deprotection of the thiol group in the presence of TCEP. After an overnight reaction, the remaining excess of phenol ester was subject to hydrolysis and the product was bound to streptavidin (Scheme 3-15).



Scheme 3-15. First attempt of a templated NCL using a Cys-labelled DNA acceptor (**3-11**), a biotin phenol ester donor (**3-23**) and a template strand (**3-25**). The excess of starting material (**3-23**) was hydrolysed using the previously optimised conditions, and the remaining product (**3-26**) was conjugated to streptavidin (Experimental section 3.4.11).

The results of the reaction were analysed by native PAGE (Figure 3-11). The ssDNA controls (Figure 3-11, lanes 1-3), and half duplex dsDNA (either **3-25**·**3-23** or **3-25**·**3-11**, lanes 4-5) allowed identification of partial denaturalisation of the duplex and the position of the fully formed duplex.

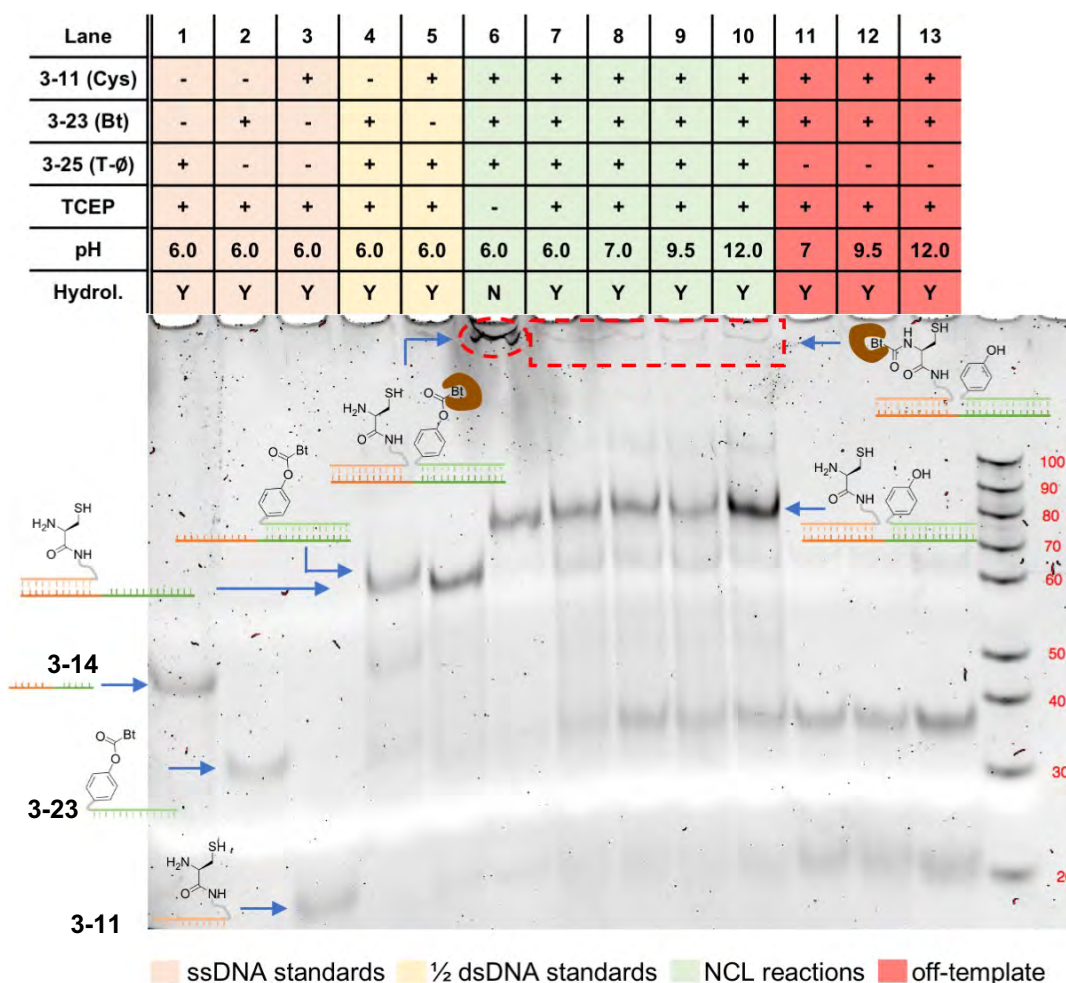


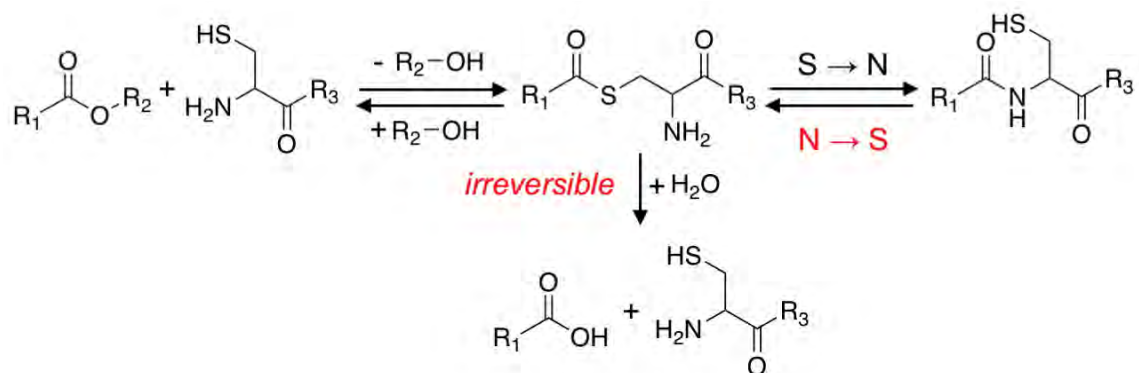
Figure 3-11. Native PAGE analysis of the NCL reaction using a Cys-labelled DNA acceptor (**3-11**). The sample in lane 6 was not subject to hydrolysis, to provide a standard of the expected migration of the product. The conversion was estimated by densitometry between the product and the hydrolysed side product bands.

The sample in lane 6 was purposely not subjected to hydrolysis, therefore the abundant biotin phenol ester-labelled DNA provided a standard of the expected position of the product in the gel. The DNA templated NCL assays at different pHs (lanes 7-10) showed a very faint band at the expected position for the product (Figure 3-11, dashed red label). This was the first indication that maybe a small fraction of the initial sample was being converted into the amide product (**3-26**). By comparison, the untemplated reactions (lanes 11-13) did not show any trace of the biotin-labelled DNA.

The faint bands attributed to the product in Figure 3-11 indicated that the conversion was very low. In comparison, the rotaxane-templated oxo-ester mediated NCL mentioned before (Section 3.1) displayed a 53% conversion over 4 steps.¹⁷ Furthermore, a similar DNA templated thio-ester mediated NCL has been reported in moderately high yield.³²

While trying to rationalise the low conversion in the DNA templated oxo-ester mediated NCL, it was found that in some conditions, a reverse-NCL process can be predominant (*i.e.* a N \rightarrow S acyl shift),^{33,34} which followed by hydrolysis of the thioester can lead to the cleavage of the peptide bond.³⁵ This effect has also reported to be able to reverse the ligation of DNA strands through NCL.³⁶ This last report was key to understand the current results: it was reported that the peptide bond used to ligate two adjacent DNA strands hydrolysed at moderately high temperature (37 °C), while the addition of an excess of exogenous thiol promoted the protection of the internal Cys thiol in the form of a disulfide, thus suppressing the reverse-NCL.

The previous results could be clearly extrapolated to the current work. As the samples are subject to high temperature and high pH, this could clearly promote the irreversible cleavage of the newly formed amide bond, through an intramolecular N \rightarrow S acyl shift followed by irreversible hydrolysis (Scheme 3-16).



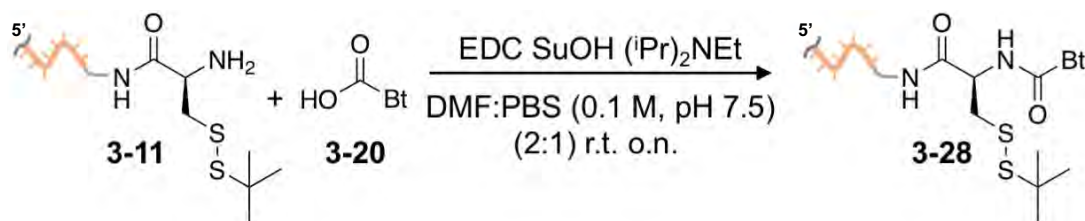
Scheme 3-16. Representation of the NCL mediated synthesis of an amide bond, and the competing **N** → **S** acyl shift, followed by an irreversible hydrolysis process.

The driving force that promotes the **S** → **N** acyl shift is usually attributed to the enhanced stability of the amide bond over the thioester.³⁷ However, there is a thermodynamic argument to understand why in this experiment the process is reversed: while the enthalpic component of the free energy of reaction favours the formation of the amide, the entropic component favours the hydrolysis, due to the larger degrees of freedom of a bimolecular system over a unimolecular one. The relative enthalpic and entropic contribution to the overall free energy is determined by the Gibbs free energy equation (eq. 3-1).

$$\Delta G = \Delta H - T\Delta S \quad \text{Eq. 3-1}$$

From eq. 3-1, it is clear that the temperature will increase the entropic component of the free energy, hence the hydrolysis of the thioester intermediate becomes thermodynamically favoured over the amide formation at high temperature.

In order to test the reverse-NCL hypothesis, the expected DNA conjugated product (**3-28**) was prepared by carbodiimide coupling of free carboxylic acid biotin (**3-20**) to Cys-labelled DNA (**3-11**, Scheme 3-17).



Scheme 3-17. Synthesis of the expected NCL product from DNA templated synthesis (**3-28**, Scheme 3-15) with a disulfide protecting group, using conventional carbodiimide coupling (Experimental section 3.4.12).

The reaction afforded a complex mixture of products, from which the main one was identified to be the expected product (**3-28**). This was then analysed by HPLC and LC-MS and purified by chromatographic fraction-collection (Figure 3-12).

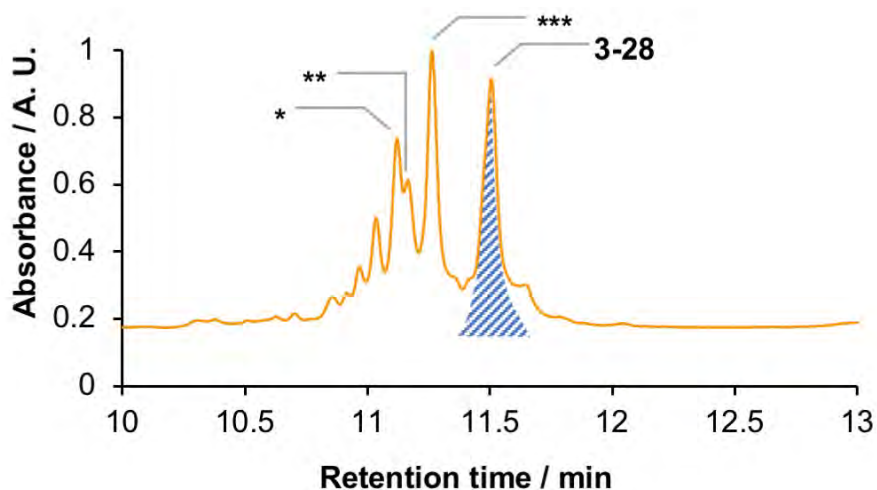


Figure 3-12. HPLC analysis of the synthesis of **3-28**. The highlighted peak, corresponding to the expected product, was fraction-collected. The expected product and several unidentified side-products were detected by LC-MS (Waters Xevo-G2-XS): **3-28** $[M]^0$ m/z calcd. 6738.435 found 6738.412; * $[M]^0$ m/z found 6609.412 (**3-28** -S^tBu); ** $[M]^0$ m/z found 6880.525 (**3-28** + 142.09); *** $[M]^0$ m/z found 6809.471 (**3-28** + 71.06).

The DNA conjugate **3-28** was deprotected *in situ* and subject to high and low temperature under moderately acidic and highly basic hydrolysis conditions (Figure 3-13 a). The results were analysed by native PAGE after binding to streptavidin (Figure 3-13 b and c).

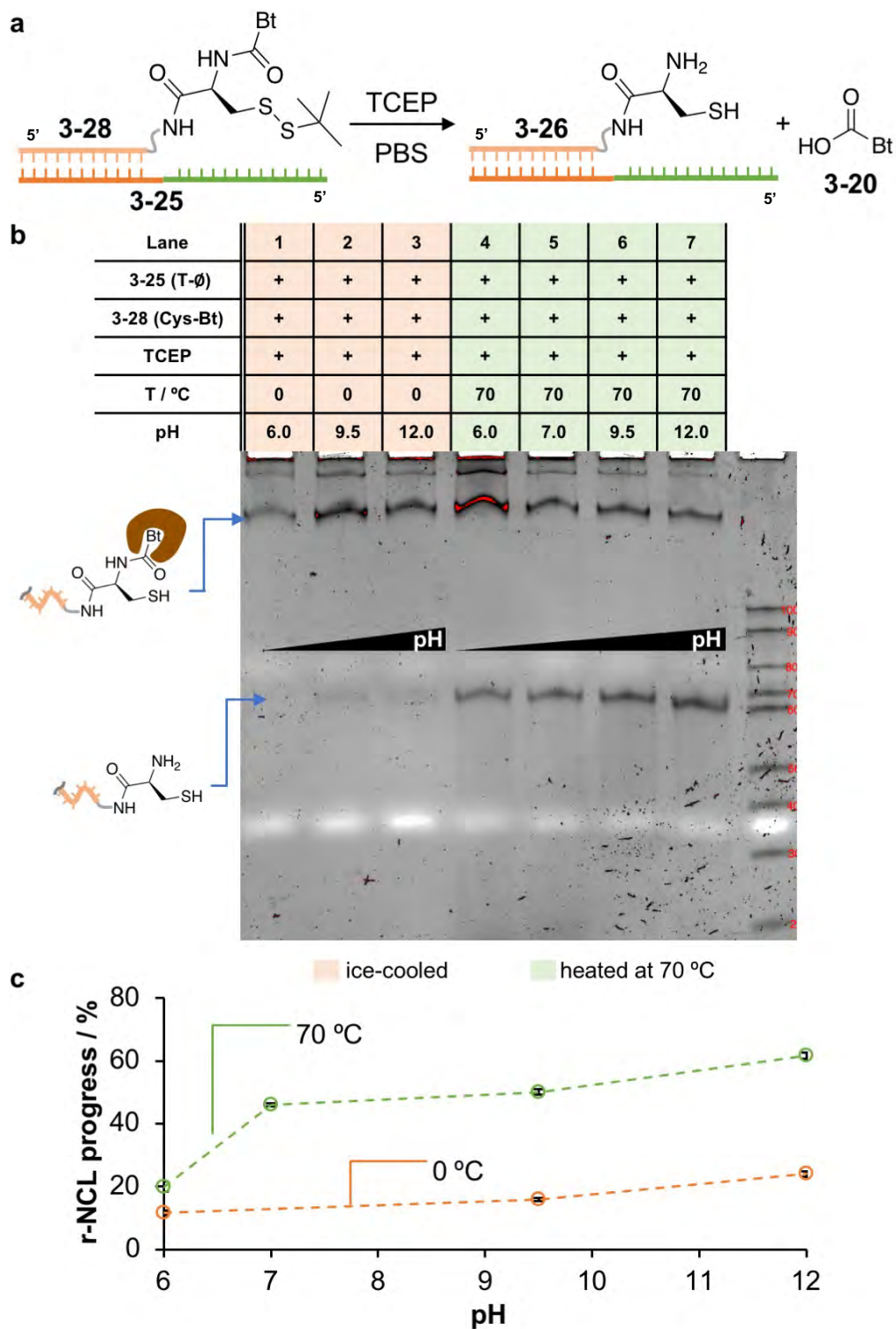


Figure 3-13. Analysis of the reverse-NCL reaction by native PAGE. a) Scheme of the reverse-NCL process (Experimental section 3.4.13). b) Native PAGE analysis of the reverse-NCL at different temperatures and pHs. c) Plot of the fraction of completed reverse-NCL (r-NCL) as function of the pH, at 0 °C and 70 °C. The reaction was performed at 83 nM.

After the addition of streptavidin, the low mobility bands in the PAGE gel corresponded to the amide, while the reverse-NCL produced a higher mobility band due to the loss of the biotin moiety. This appears to be in good agreement with the thermodynamic prediction that reverse-NCL is favoured only at high temperature, the samples that were subject to hydrolysis at moderate to high pH and low temperature (0 °C) only displayed a low (< 25%) reverse-NCL with a small promoting effect of the high pH. At high temperature (70 °C) increasing the pH had a dramatic effect promoting the reverse-NCL reaction up to ~60%.

As the PAGE analysis only reports binding to streptavidin, one could argue that the previous data (Figure 3-13 b) could be interpreted as degradation of the biotin moiety itself, rather than the cleavage of the amide bond. To rule out this possibility, a biotin-labelled DNA strand through an amide bond** was subject to the same high temperature and high pH (Figure 3-14 a and b).

From this experiment, it can be concluded that the addition of TCEP, high pH and high temperature does not decompose the biotin moiety and all the DNA binds streptavidin. We concluded that the lack of streptavidin binding (Figure 3-13 b) must be due to a reverse-NCL process followed by hydrolysis of the thioester.

** Integrated DNA technologies (IDT, US) biotin modification /5Biosg/

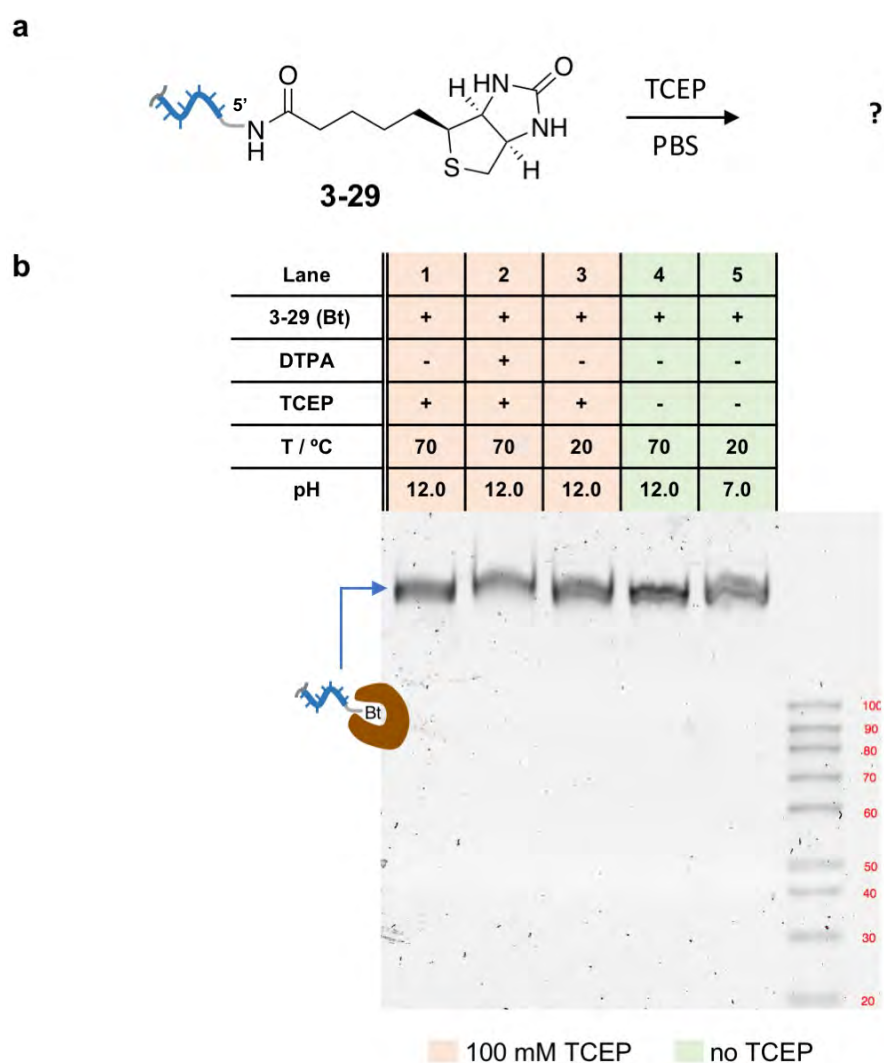
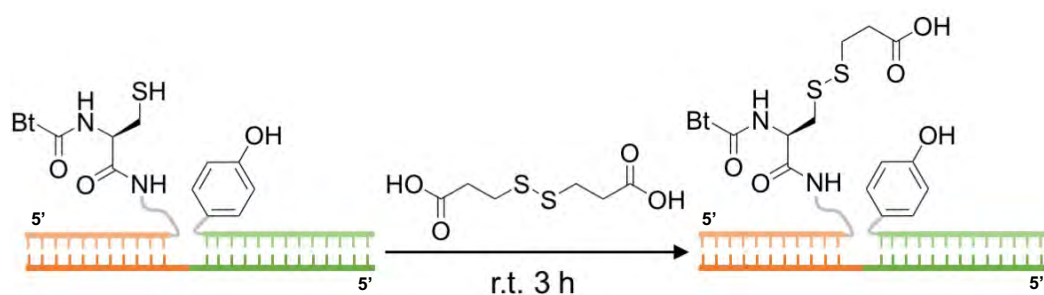


Figure 3-14. Native PAGE analysis of the stability of a D-biotin-labelled DNA in TCEP solution, at high pH and temperature (Experimental section 3.4.13).

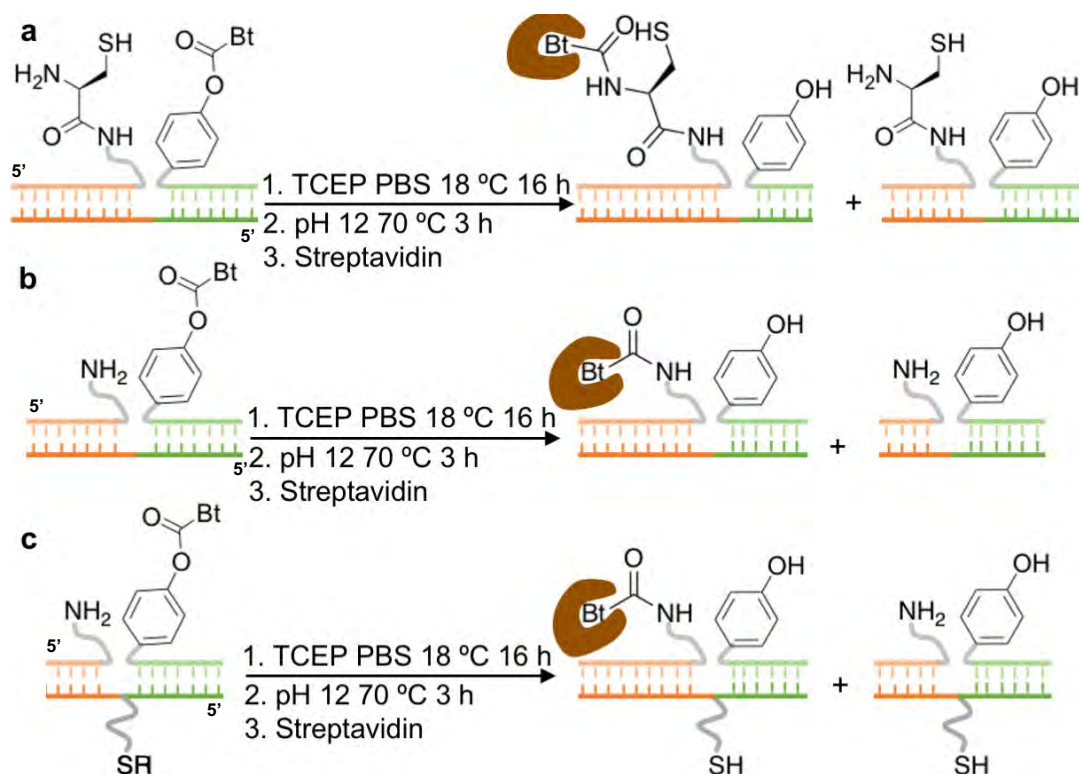
3.2.8 Comparison of a co-localised thiol-labelled DNA template and a Cys-labelled DNA acceptor

In order to suppress the reverse-NCL and determine the conversion of the several DNA templated reaction designs, the thiol was protected prior to the hydrolysis step, by reaction with 3,3'-dithiodipropionic acid (DTPA, Scheme 3-18). The protection of the thiol as a disulfide, was implemented in all subsequent DNA templated reactions prior to the high temperature hydrolysis, in order to prevent the reverse-NCL reaction.



Scheme 3-18. Protection of the thiol after DNA templated NCL, by reaction with 3,3'-dithiodipropionic acid.

In order to quantify the catalytic function of the co-localised thiol, and to compare it to the catalysis of a Cys acceptor group, DTS reactions were performed in parallel with a DNA-Cys acceptor, DNA-NH₂ acceptor, and a thiol-modified DNA template (Scheme 3-19).



Scheme 3-19. Three approaches to DNA templated NCL. a) Cys-labelled DNA acceptor. b) Uncatalysed peptide bond formation. c) Thiol-labelled DNA template catalyst and amine acceptor (Experimental section 3.4.14). DTSs were performed at 0.1 μ M.

Initially, the Cys catalysed NCL (Scheme 3-18 a) was performed at several pH values (Figure 3-15).

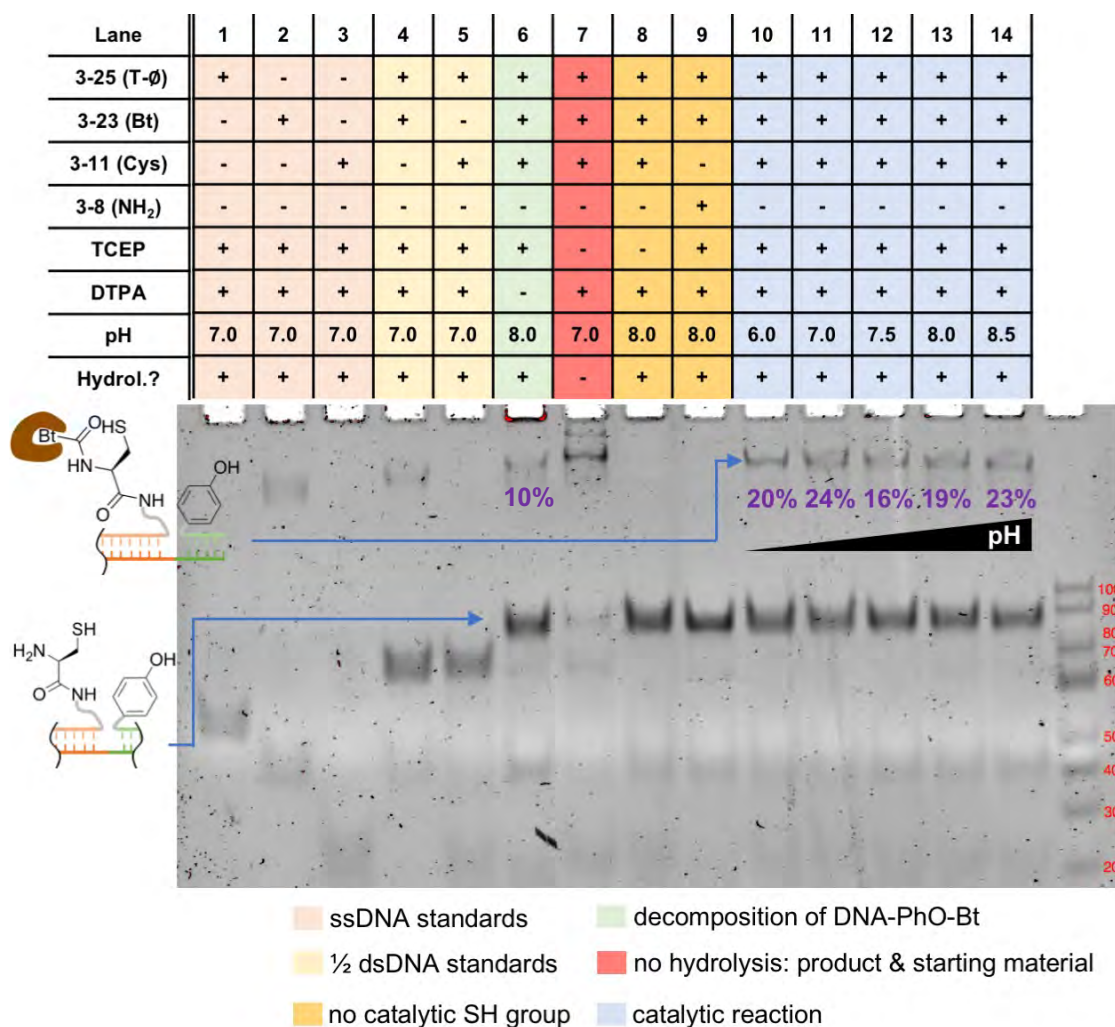


Figure 3-15. Native PAGE analysis of the NCL process in the absence of a thiol catalyst (lanes 8-9, Scheme 3-19 b) and with a Cys-labelled DNA acceptor at several pHs (lanes 10-14, design 1, Scheme 3-19 a).

In the absence of thiol protection prior to hydrolysis, a small fraction of streptavidin-binding DNA was detected (lane 6). If no hydrolysis was performed, there was a combined 77% of starting material and product, over a 23% of hydrolysis (lane 7), this highlights the potential of phenol esters to be used in long molecular assembly processes in where the DNA adapters remain in solution for a long period. If the

catalytic thiol was not available, either because it was in the disulfide form in the absence of TCEP (lane 8) or because the acceptor was a primary amine (lane 9), no product was observed. On the contrary, when all the components for a DNA templated NCL were present, using a Cys-labelled DNA acceptor, an average 20% conversion was detected (lanes 10-14). Interestingly, very minor conversion differences were detected between a pH range 6.0 to 8.5, which indicated that this particular NCL reaction was not very pH sensitive within this range. The differences were within the quantification error (Figure 3-15).

The previous experiment clearly indicated that the presence of a thiol catalytic group in a Cys amino acid increased the conversion of the DNA templated NCL from 0% to approximately 20%. As the densitometric analysis has certain limitations for quantitative analysis,³⁸ the reaction at pH 8.5 was performed in triplicate and the results were analysed by native PAGE, consistently showing the expected product (Figure 3-16). In this experiment, the lack of reactivity in the absence of a catalytic thiol group was reproduced (lanes 6-7). The formation of the product also showed to be reproducible (lanes 8-10), expressing the error as the standard deviation, the conversion of this reaction is $(20\pm 4)\%$.

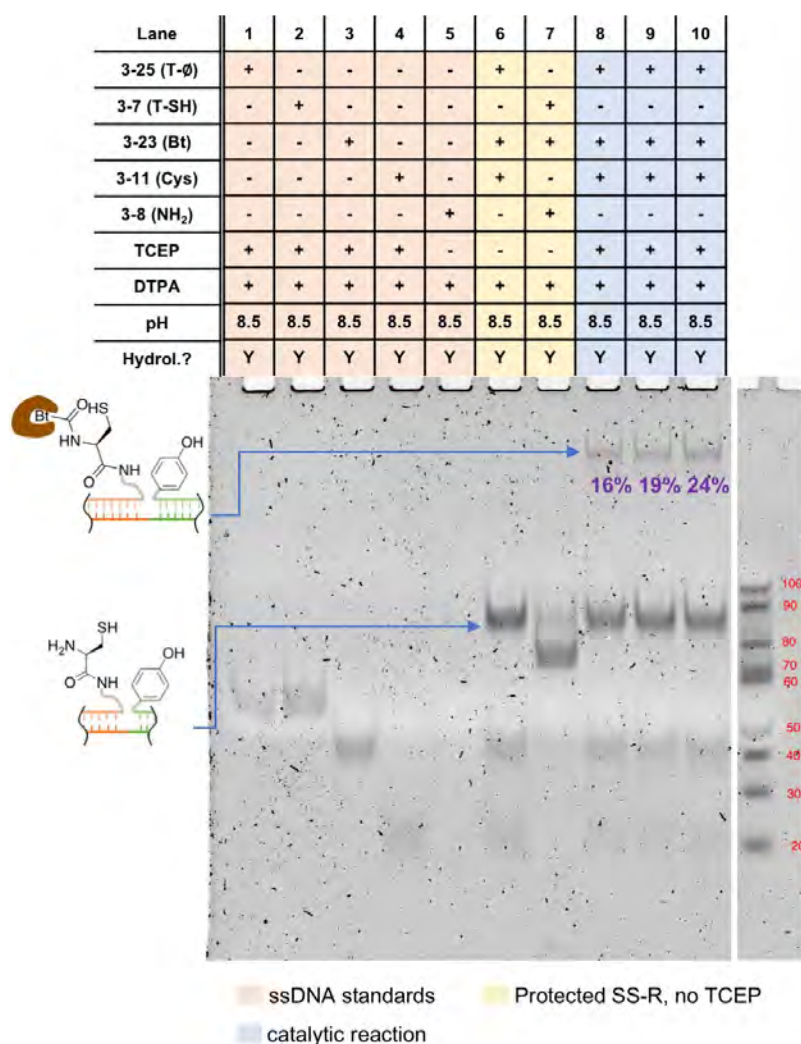


Figure 3-16. Native PAGE analysis of a DNA templated NCL with a Cys-labelled DNA acceptor. Three independent reactions were performed (lanes 8-10). Redundant lanes were cropped from the gel.

Finally, design 2 was tested by using thiol-labelled template and an amine-labelled acceptor (Scheme 3-18 b and c). Four independent replicates of the reaction were performed and compared to four independent replicates of the uncatalysed DNA templated reaction. After the reaction, the samples were subject to thiol protection with DTPA, the excess of starting material phenol ester was hydrolysed, and the results were analysed by native PAGE (Figure 3-17).

Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14
3-25 (T- \emptyset)	+	-	-	-	-	-	+	+	+	+	-	-	-	-
3-7 (T-SH)	-	+	-	-	+	+	-	-	-	-	+	+	+	+
3-23 (Bt)	-	-	+	-	+	+	+	+	+	+	+	+	+	+
3-8 (NH ₂)	-	-	-	+	+	+	+	+	+	+	+	+	+	+
TCEP	+	+	+	+	-	-	+	+	+	+	+	+	+	+
DTPA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
pH	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5
Hydrolysis?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

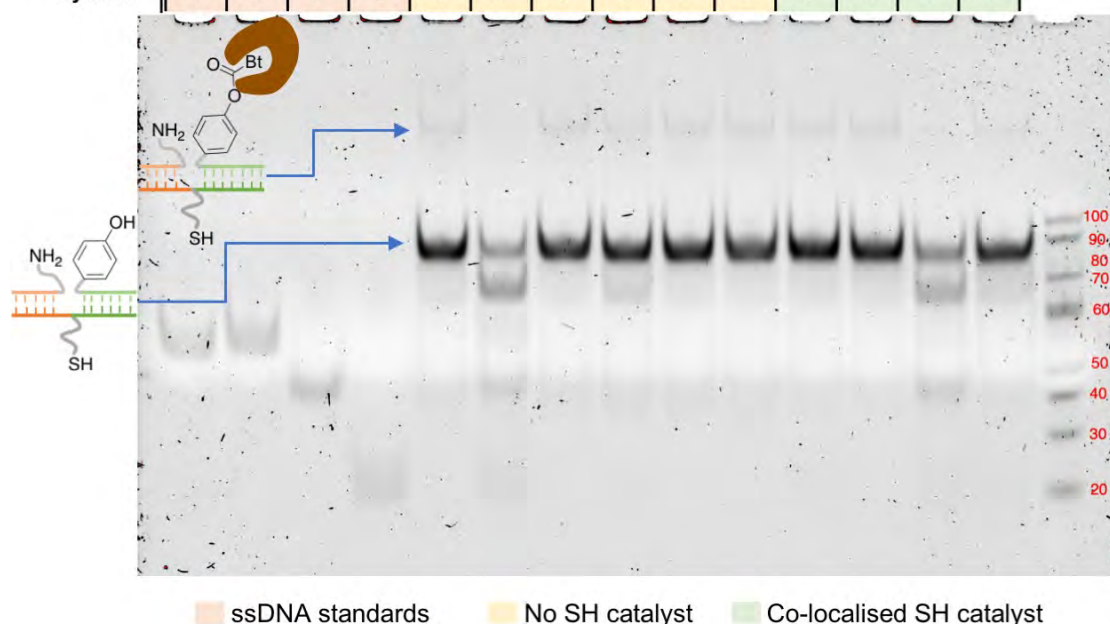


Figure 3.17. Native PAGE analysis of an uncatalysed DNA templated peptide bond formation (lanes 5-10) and with a thiol-labelled DNA template catalyst (lanes 11-14, design 2, Scheme3-19 c).

Using a protected disulfide-labelled template (**3-7**) in the absence of TCEP (lanes 5-6), an unfunctionalised DNA template (**3-25**, lanes 7-10) or a deprotected thiol-labelled template after the addition of TCEP (**3-14**, lanes 11-14) did not produce a substantial amount of product, and therefore, there appears to be no catalytic effect from the thiol. A very faint band (< 3% integrated optical density) corresponding to streptavidin-binding DNA was observed in all lanes (5-14), this was attributed to partial hydrolysis. The large difference between the conversion when the amine and thiol functional groups are closely co-localised in a Cys residue (Figure 3-16), and in separated

positions (Figure 3-17) sheds some light on the mechanistic requirements of DNA templated NCL: the transition from the thioester intermediate into the amide bond in Cys takes place through a low energy intermediate and transition state (Figure 3-2), while in the thiol-labelled template system, this step requires the co-localisation of long and flexible linkers forming macrocyclic intermediates and transition states. The less favourable energetics of the second system does not enable any catalytic activity of the thiol group. This highlights the importance of a precise geometrical design that chemical templates require to perform an efficient catalytic activity. This is especially important when working with stable chemical groups such as thioesters (reported DTS yield < 40%),^{36,39-41} as opposed to highly reactive *N*-hydroxysuccinimide esters or phosphorous ylide/aldehyde templated chemistry (reported DTS yield > 80%).^{42,43} There is one notable example of high DNA templated acyl transfer yield (80%) using an unreactive thioester.⁴⁴ Crucially, in this work the reactive groups were directly attached to DNA in the absence of linkers. Presumably, this had the effect of inducing a very high local concentration which accelerated the kinetics of the DNA templated reaction. Long commercially available linkers have been typically used in DNA templated synthesis;^{1,2} however, only a limited number of studies were found in the literature.⁴⁵ To try to visualise the effect of increasing the length of flexible linkers on the reaction rate of a bimolecular templated reaction, a very simplistic model was applied here (Figure 3-18).

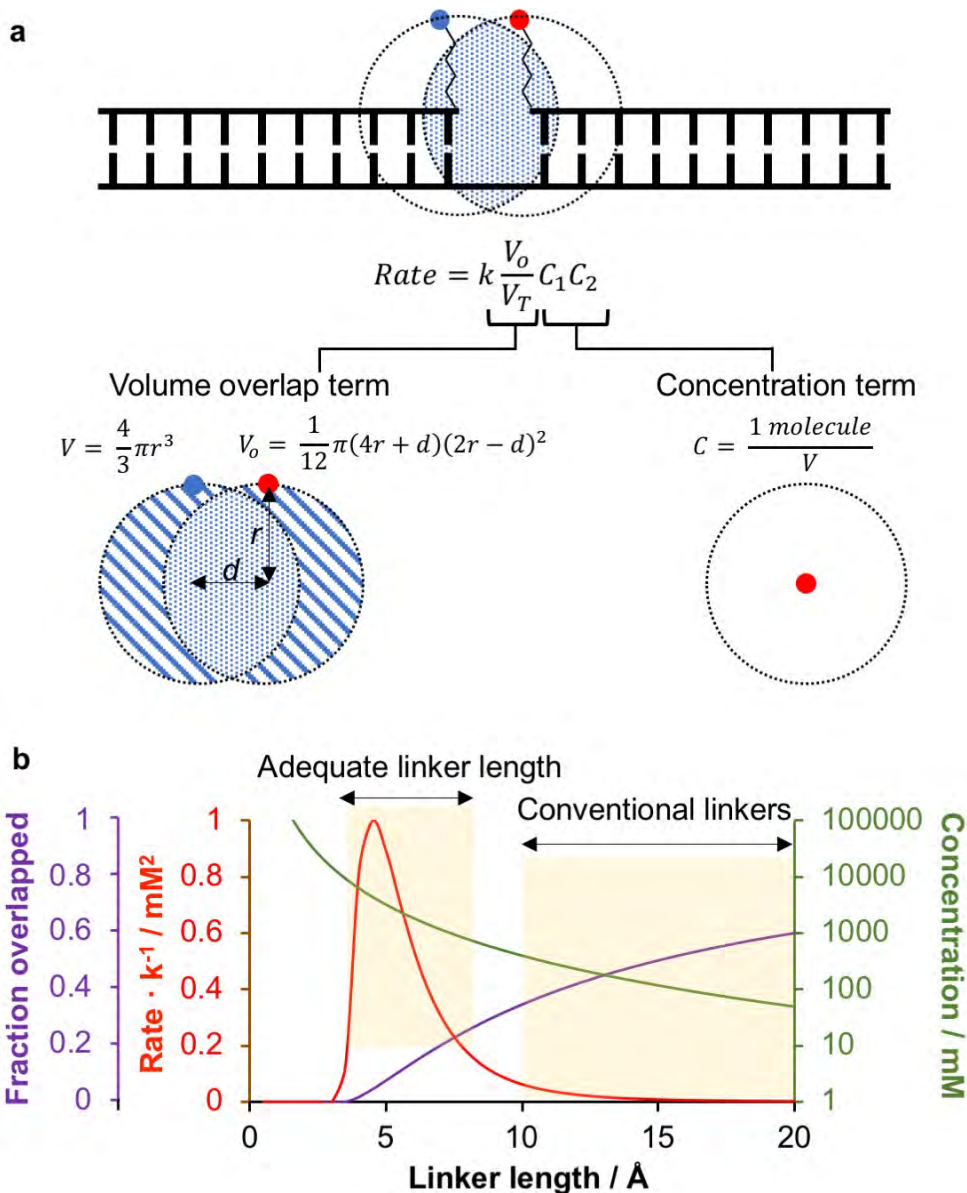


Figure 3-18. Modelled effect of the linker length on the reaction rate. The 3D model was represented using closely related 2D geometry to simplify the diagrams. a) Schematic representation of the intersection between the volume defined by two flexible linkers attached to DNA strands. The fraction of overlapped volume, and the concentration of the molecule attached to the linker were approximated based on simple geometric considerations. The distance between the origin of the spheres “d” was determined from the B-DNA crystallographic structure.⁴⁶ b) Plot of the Rate / k; Fraction of overlapped volume and concentration as function of the linker length (Experimental section 4.3.15).

In this model, it was assumed that the linkers could adopt any conformation within the defined spheres with linker length radii (Figure 3-18 a). To simplify the calculations, it

was imposed that the two linkers would have the same length. The reaction could only happen in the region of the space where the two spheres overlap. The effective concentration of each reactant was determined as the proportion between one molecule (converted into mM using Avogadro constant) and the volume of the sphere determined by the linker radius.

To examine the impact of linker length on the reaction rate, a second order kinetic rate law was applied (Figure 3-18 b). This only relates to a single reaction step, under the condition that this analysis only applies to situations where the linker length limits the overall reaction rate. The concentrations were corrected by the fraction of overlapped volume over the total volume (*i.e.* V_0 / V_T). As no absolute rates were determined, the $Rate / k$, where k is the second order rate constant, was plotted as function of the linker length (Figure 3-18 b). The change on the concentration and the fraction of volume overlapped were represented as well. This showed that, due to the sharp decrease on the effective concentration with the linker length, and the progressive increase on the overlapped volume, the reaction rate was optimal (> 20% of the maximum rate) over a narrow range of 4 to 8 Å, and it decreased very sharply outside this length window, thus implying that an optimisation of linkers is crucial for high reaction rate.

This model contains many simplifications: the space occupied by the DNA template would not be available for the reactive groups; binding to the DNA groove could facilitate the interaction between reactive groups thus reducing the effect of the linker flexibility; not all linker conformations will be equally likely, thus the space occupied will not be a sphere; an effective interaction between two reactants requires not only close proximity but also the right spatial orientation, this could improve with linker flexibility. Regarding the effective concentrations, from the minimal linker length that allows the

reactive groups to interact (3.4 Å per linker), the concentration (corrected by the overlapped volume) is estimated to be 1.6 M, while this is a very high concentration, it is not too different from reported effective concentrations in DNA templated synthesis.⁴⁷ Beyond all its limitations, this approach allowed to visualise the importance of the design of appropriate linkers. In this work, the linkers used were approximately 11.7 and 22.2 Å.^{††} The low NCL reaction yield was partially attributed to the excessive linker length.

This chapter began by discussing how ribosomal synthesis of peptides relies on stable ribose esters and their activation in the ribosome active centre (Figure 3-1). It is currently believed that the catalytic activity of the ribosome is largely entropic, through a very precise co-localisation of reagents in the space.^{48–50} An effective activation of unreactive chemical groups for DNA templated synthesis will require a precise spatial localisation in order to reduce the activation entropy, this could be achieved by a precise design of the linkers, or by evolution of biopolymers (*i.e.* proteins and nucleic acids) that fold appropriately to provide an active centre. Significant success using the later approach has been already achieved for different applications.^{51–53}

^{††} The linker lengths were estimated in an extended conformation optimised by molecular mechanics (MM2 force field) using Chem3D (v 15.0).

3.3 Conclusions

Two DNA systems containing the chemical groups required to template a NCL reaction (*i.e.* a phenol ester, an amine acceptor and a catalytic thiol) have been prepared: in the first one, the catalytic thiol is placed in a complementary template strand, while in the second system both the amine and the thiol were present in a Cys residue.

An *L*-Ala-Boc transferable group did not allow to unambiguously determine the conversion of DNA templated NCL; however, the use of a *D*-biotin transferable group allowed a clear identification of the reaction product by native PAGE upon binding to streptavidin.

It was found that at high temperature and high pH, a reverse-NCL process followed by hydrolysis of the intermediate thioester cleaved the majority of the peptide bond. The suppression of the reverse-NCL was achieved by protection of the thiol as a disulfide. This made it possible to apply high temperature and pH to hydrolyse the excess of phenol ester starting material prior to PAGE analysis.

The co-localisation of a thiol in a complementary DNA template did not produce any improvement in the conversion of DNA templated NCL compared to the same DNA system in the absence of the thiol, and no product formation was detected. Conversely, when the acceptor group was a Cys-labelled DNA, the formation of the product increased to a $20\pm 4\%$. While this moderate conversion would not allow multistep DNA templated synthesis of peptides, it exemplifies the crucial role that precise localisation of reactants in the space plays in DNA templated chemistry.

It would be possible in the future to improve the yield of DNA templated NCL by improving the spatial co-localisation of the reactive groups either by evolving biopolymers, such as aptamers, to determine the spatial co-localisation of the reactive

groups, by the design of a supramolecular interaction (e.g. simultaneous coordination of the reactive groups to a metallic centre) or the optimisation of rigid linkers that co-localise the reactive groups in close proximity. In addition, the reported higher catalytic performance of selenocysteine in NCL reactions⁵⁴ could further improve the kinetics of this reaction thus allowing multistep DNA templated peptide synthesis.

3.4 Experimental section

3.4.1 General methods

For details on RP-HPLC, LC-MS, PAGE and nanopure water, see Appendix 1.

All reagents were directly purchased from Sigma Aldrich unless otherwise stated.

Streptavidin solution was purchased from New England Biolabs (1 mg·mL⁻¹). A 20-100 nt ssDNA ladder was purchased from Integrated DNA Technologies.

3.4.2 Sequences and modifications

Table 3-1. DNA oligonucleotide sequences and modifications.

Identification number	Sequence and modifications (5' → 3')
3-4	/5AmMC6/ ATG TAA GTA AGT CAA GTC CAG GTC GTT CAA
3-6	GCT ACT GGA CTT GAC TTA CTT ACA T/iAmMC6T/C GGC TGA CTG CTG GCT CGG C
3-8	GCC GAG CCA GCA GTC AGC CG /3AmMO/
3-18	/5ThioMC6-D/ ATG TAA GTA AGT CAA GTC CAG GTC GTT CAA
3-21	GCT ACT GGA CTT GAC TTA CTT ACA TTC GGC TGA CTG CTG GCT CGG C
3-26	/5Biosg/ TTT TTT TTT TTT TTT ACG TTG CTG CCG AGG GGA GAA GAG TAA AAT GCA ACG T

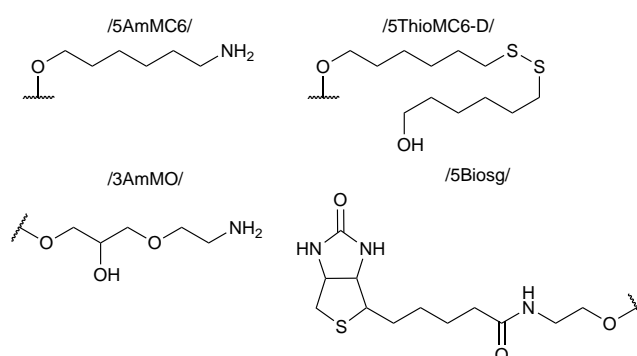


Figure 3-19. Chemical structure of commercially available DNA modifications.

3.4.3 Synthesis of 3-3

774 mg of Boc-L-Ala-OH (4.09 mmol), 1551 mg of HBTU (4.09 mmol) and 277 mg of HOBT (2.05 mmol) were dissolved in 2 mL of DMF in a dry Schlenk tube under positive nitrogen pressure and the resulting solution was stirred at room temperature for 5 min. 500 mg of 4-hydroxybenzoic acid (4.09 mmol) were dissolved with 2 mL of DMF in a dry Schlenk and they were added via cannula into the first solution. 627 μ L of Et₃N (4.50 mmol) were added and the mixture was stirred for 1 h at room temperature. The solution was diluted in 20 mL of EtOAc and washed with 3 \times 20 mL of brine, 3 \times 20 mL of NaHCO₃ and 5 \times 20 mL of LiCl (aq, 1.0 M).

¹H-NMR (300 MHz, chloroform-*d*) δ (ppm): 9.92 (1H, s, CHO), 7.86 (2H, d J = 7.8 Hz, Ar-H), 7.22 (2H, d J = 7.9 Hz, Ar-H), 5.03 (1H, s^{br}, NH), 4.48 (1H, *m*, CH), 1.49 (3H, d J = 7.1 Hz, CH₃), 1.40 (9H, s, C(CH₃)₃). **¹³C-NMR** (75 MHz, chloroform-*d*) δ (ppm): 190.98 ppm (CHO), 171.65 (C=O), 155.30 (C=O), 134.32 (ArC), 131.38 (ArC-H), 122.28 (ArC-H), 116.05 (ArC), 82.53 (C(CH₃)₃), 49.68 (CH), 28.43 (C(CH₃)₃), 18.29 (CHCH₃). **HRMS** (ESI+) m/z [M+Na]⁺ calcd. 316.1155 found 316.1157.

3.4.4 Synthesis of 3-5

133 μ L of amine-labelled DNA (**3-4**, 75 μ M in H₂O, 10 nmol) were diluted with 133 μ L of MES buffer (0.1 M, pH 6). 12 mg of **3-3** (40 μ mol) were dissolved in 266 μ L of DMF and added over the DNA solution. The mixture was shaken for 10 min. Then, 12.5 mg of NaBH₃CN (200 μ mol) were added and the reaction mixture was shaken at 20 °C for 4 h. The mixture was then eluted through a NAP-10 size exclusion column (Biorad) and the product was transferred into H₂O and concentrated using 3 kDa molecular weight cut-off (MWCO) amicon (Sigma Aldrich) spin-filter. The product was analysed

by HPLC and LC-MS (79% conversion based on HPLC peak area). The product was isolated by HPLC fraction-collection.

3.4.5 Synthesis of 3-7

25 μL of internal amine-labelled DNA (**3-6**, 2 mM, 50 nmol) were diluted with 125 μL of PBS (0.1 M, pH 7.6). 100 μL of 3-(2-pyridyldithio)propionate (40 mM in DMSO, 4 μmol) were added over the DNA solution and the mixture was shaken at 24 $^{\circ}\text{C}$ for 1 h. Finally, the solution was diluted with 250 μL of H_2O . DNA was precipitated by the addition of 1/10 vol of NaOAc (3.0 M, pH 6.0) and 2 vol of EtOH, the mixture was incubated at 0 $^{\circ}\text{C}$ for 30 min and centrifuged at 4 $^{\circ}\text{C}$ and 21 kRCF for 30 min. The excess of solvent was removed, and the DNA pellet was resuspended in H_2O . The product was analysed by HPLC and LC-MS (82% conversion based on HPLC peak area). The product was HPLC fraction-collected.

3.4.6 Synthesis of 3-11

Stock solutions: 300 mM Fmoc-Cys(S^tBu)-OH in DMF (**3-9**, 129.5 $\text{mg} \cdot \text{mL}^{-1}$), 300 mM EDC·HCl in DMF (dispersion, 57.6 $\text{mg} \cdot \text{mL}^{-1}$), 300 mM *N*-hydroxysuccinimide in DMF (34.6 $\text{mg} \cdot \text{mL}^{-1}$). 67 μL of each stock solution were combined in a microcentrifuge tube, and 3.5 μL of ⁱPr₂NEt were added. The mixture was allowed to react for 10 min. In a separate microcentrifuge tube 20 μL of amine-labelled DNA (**3-8**, 1 mM in H_2O , 20 nmol) were diluted with 180 μL of PBS (0.1 M, pH 7.5) and 200 μL of DMF. The small molecule solution was added over the DNA solution and the resulting mixture was shaken at 20 $^{\circ}\text{C}$ for 48 h. The solution was then diluted with 1 mL of H_2O , the precipitate was centrifuged, and the DNA was transferred into H_2O by successive steps of dilution of the supernatant with H_2O and concentration by centrifugal ultrafiltration using a 3 kDa MWCO amicon spin filter. The product was analysed by HPLC and LC-MS

(complete conversion and small molecule contaminant). The product was HPLC fraction-collected. The previous product (**3-10**) was subsequently diluted in concentrated aqueous ammonia and shaken at 36 °C for 3 h. The solution was neutralised with an excess of PBS (0.1 M, pH 6) washed and concentrated using a 3 kDa MWCO amicon spin filter. The product was analysed by LC-MS showing full conversion (Figure 3-6 b).

3.4.7 DNA templated NCL (I): L-Ala-Boc transfer

Solvent: PBS (1.0 M NaCl, 0.1 M NaH₂PO₃, pH 7-10)

Stock solutions: TCEP (0.5 M in PBS, pH 7.0), DNA **3-5**, **3-7** and **3-8** 100 µM.

In a microcentrifuge tube, 1 µL of each DNA solution was added to 96 µL of PBS and 1 µL of TCEP. Final concentrations dsDNA 1 µM and TCEP 5 mM. The mixture was shaken at 20 °C for 18 h. Then, samples were diluted 1:10 in denaturing loading buffer, heated at 80 °C for 5 min and analysed by 30% acrylamide denaturing PAGE.

3.4.8 Synthesis of 3-19

1.960 g of maleic anhydride (**3-15**, 20 mmol) were dissolved in 25 mL of Et₂O. 2.180 g of *p*-aminophenol (**3-16**, 20 mmol) dissolved in 5 mL of Et₂O were added to the previous solution and the resulting mixture was stirred at r.t. for 1 h. Then, it was cooled in an ice bath for a further 30 min. The yellow precipitate was filtered, dried and dissolved in a solution of 650 mg of NaOAc in 6.7 mL of Ac₂O. The solution was stirred at 80 °C for 30 min, and then allowed to cool down to r.t. The solution was poured over ice, the resulting white precipitate was filtered, washed with cold water, and subject to hydrolysis in 90 mL of MeOH / 5 mL HCl (aq, ~12 M) at r.t. for 2 h. The progress of the hydrolysis was monitored by TCL (R_f product 0.6 in hexane / EtOAc 1:1). The excess

of solvent was removed under vacuum and the resulting yellow solid (**3-17**) was analysed by $^1\text{H-NMR}$, showing good correlation with the reported chemical shifts.²⁹

$^1\text{H-NMR}$ (300 MHz, chloroform-*d*) δ (ppm): 7.30 (2H, m, Ar-*H*), 7.12 (2H, m, Ar-*H*), 6.77 (2H, s, maleimide-*H*).

In a dry Schlenk flask under positive nitrogen pressure, 2 mL of SOCl_2 were added dropwise to 150 mg of *D*-biotin (**3-20**, 0.61 mmol). The mixture was stirred at room temperature for 1 h, then the excess of SOCl_2 was removed under reduced pressure. 115 mg of **3-17** (0.61 mmol) were dissolved in 1 mL of DMF and slowly added over *D*-biotin acyl chloride intermediate (**3-21**). 310 μL of $(i\text{Pr})_2\text{NEt}$ (3.2 mmol) were added dropwise and the resulting solution was stirred at r.t. for 1 h. The excess of solvent and base was removed under reduced pressure, and the solids were dissolved in 2 mL of CHCl_3 . 0.2 mL of EtOAc were added, the precipitate was centrifuged, and the supernatant was washed with HCl (aq, 1.0 M). Finally, the organic layer was dried with a small amount of anhydrous MgSO_4 , filtered and the solvent was removed under reduced pressure. This product was conjugated to thiol-labelled DNA without further purification.

Conjugation to thiol-labelled DNA and LC-MS characterisation: 1 μL of DNA (**3-22**, 1 mM in H_2O , 1 nmol) was mixed with 2 μL of TCEP (100 mM, pH 7 adjusted with 5 M NaOH) and shaken at 18 $^\circ\text{C}$ for 1 h. Then, 1 mg of **3-19** was dissolved in 10 μL of DMF and added over the DNA sample. The resulting solution was shaken at 18 $^\circ\text{C}$ for 1 h. The product was diluted with 30 μL of H_2O and eluted through Bio-spin 6 column (Biorad) previously equilibrated with H_2O . The product was analysed by HPLC, LC-MS and native 15% PAGE after binding to streptavidin (37% conversion based on HPLC

peak area, 27% conversion based on PAGE densitometry). The product was HPLC fraction-collected. **LC-MS** (ESI-): $[M]^0$ m/z calcd. 9860.741 found 9860.394.

The effect of streptavidin binding on the relative mobility of DNA was studied by native PAGE analysis of a streptavidin bound DNA, and the same free DNA (Figure 3-20).

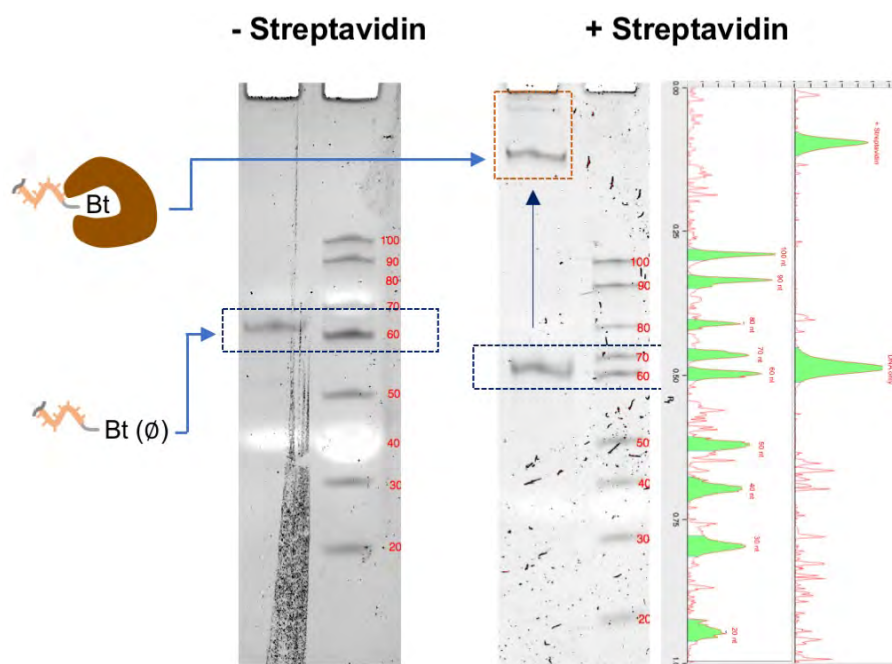


Figure 3-20. Comparison of the native PAGE analysis of a sample before and after the addition of streptavidin, in the presence of a 100 - 20 nucleotides (nt) ssDNA ladder. Streptavidin-DNA conjugates appear before the 100 nt band, while all the sequences in the present work were designed within 20 to 100 nt length.

3.4.9 Degradation of maleimide by DMAP

19 mg of **3-17** (74 μmol) were dissolved in 1 mL of : a) DMF. b) DMF and 18 μL of $(i\text{Pr})_2\text{NEt}$. c) DMF and 4 mg of *N,N*-dimethylaminopyridine (DMAP, 35 μmol). d) DMF, 18 μL of $(i\text{Pr})_2\text{NEt}$ and 4 mg of DMAP. The samples were stirred at r.t. for 3 h, the solvent was removed under reduced pressure and the resulting solid was dissolved in chloroform-*d* and analysed by $^1\text{H-NMR}$ spectroscopy.

3.4.10 Optimisation of the hydrolysis of 3-23

1 μL of biotin phenol ester-labelled DNA (**3-23**, 1 μM) was mixed with 6 μL of PBS (0.1 M, pH 7.0). Then, 3 μL of PBS (0.2 M, pH 14) were added and the solution was heated at the desired temperature for 3 h. Then, the pH was reduced by the addition of 3 μL of PBS (0.1 M, pH 4). 1 μL of streptavidin solution (1 $\text{mg} \cdot \text{mL}^{-1}$ *D*-biotin binding capacity 62 $\text{pmol} \cdot \mu\text{L}^{-1}$) was added and the solution was incubated for 15 min at 18 $^{\circ}\text{C}$. Finally, the sample was mixed with 1 vol of native PAGE loading buffer, and analysed by 15% native PAGE.

3.4.11 DNA templated NCL (II): Hydrolysis of the starting material in the presence of a deprotected thiol

Stock solutions: PBS (0.1 M, pH 6.0 to 12.0), TCEP (50 mM, pH 7.0 adjusted with 5 M NaOH), DNA solutions (**3-25 T**, **3-11 Cys** and **3-19 Bt**) 1 μM in the reaction pH, PBS (0.2 M, pH 14).

1 μL of each DNA strand was added to a microcentrifuge tube, the volume was adjusted to 9 μL with PBS at the desired pH, and finally 1 μL of TCEP was added and the combined solution was shaken at 18 $^{\circ}\text{C}$ for 16 h. 3 μL of PBS (0.2 M, pH 14) were added, and the solution was shaken at 70 $^{\circ}\text{C}$ for 3 h. The pH was then neutralised with 3 μL of PBS (0.1 M, pH 4) and the sample was divided into two aliquots: the first aliquot was combined with 1 μL of streptavidin solution (1 $\text{mg} \cdot \text{mL}^{-1}$) and shaken at 18 $^{\circ}\text{C}$ for 15 min and then analysed by native PAGE; the second aliquot was directly analysed by native PAGE.

3.4.12 Synthesis of 3-28

Stock solutions: 300 mM *D*-biotin in DMF (**3-20**, 73 $\text{mg} \cdot \text{mL}^{-1}$), 300 mM EDC·HCl in DMF (dispersion, 57.6 $\text{mg} \cdot \text{mL}^{-1}$), 300 mM *N*-hydroxysuccinimide in DMF (34.6 $\text{mg} \cdot$

mL⁻¹). 67 μ L of each stock solution were combined in a microcentrifuge tube, and 3.5 μ L of ⁱPr₂NEt were added. The mixture was allowed to react for 10 min. In a separate microcentrifuge tube 20 μ L of Cys(S^tBu)-labelled DNA (**3-11**, 1 mM in H₂O, 20 nmol) were diluted with 80 μ L of PBS (0.1 M, pH 7.5) and 200 μ L of DMF. The small molecule solution was added over the DNA solution and the resulting mixture was shaken at 20 °C for 16 h. The solution was then diluted with 1 mL of H₂O, the precipitate was centrifuged, and the DNA was transferred into H₂O by successive steps of dilution of the supernatant with H₂O and concentration by centrifugal ultrafiltration using a 3 kDa MWCO amicon spin filter. The product was analysed by HPLC and LC-MS (32% conversion based on HPLC peak area). The product was HPLC fraction-collected.

3.4.13 Reverse-NCL: assay of the process and the stability of *D*-biotin-labelled DNA

Stock solutions: PBS (0.1 M, pH 6.0 to 12.0). DNA solutions (**3-28** Cys(S^tBu)Bt and **3-25** T) 1 μ M in H₂O and TCEP (50 mM, pH 7.0 adjusted with 5 M NaOH).

1 μ L of each DNA strand was added to a microcentrifuge tube, the volume was adjusted to 9 μ L with PBS at the desired pH, and finally 1 μ L of TCEP was added. The solutions were placed in an ice bucket or a thermal shaker at 70 °C and were shaken for 3 h. The PAGE analysis was performed as described previously.

In order to ensure the stability of the biotin moiety, a commercially available DNA strand (**3-29**) labelled with the modification /5Biosg/ was subject to analogous conditions.

3.4.14 DNA templated NCL (III): Protection of the thiol before hydrolysis

The procedure was analogous to the previously described (Experimental section 3.4.11). Before the addition of PBS (0.2 M, pH 14), 1 μ L of 100 mM 3,3'-

dithiodipropionic acid dispersion (PBS, pH 12), was added to the samples and they were shaken at 18 °C for 3 h. Then the previous procedure was resumed.

3.4.15 Estimation of the effect of linker length on a bimolecular reaction rate

Concentration of each molecule:

$$C \text{ (mM)} = \frac{1 \text{ molecule} \cdot 10^3}{N_A} \quad \text{Eq. 3-2}$$

C = concentration

N_A = Avogadro number

Fraction of overlapped volume:

$$V_o = \frac{1}{12} \pi \cdot (4r + d)(2r - d)^2 \quad \text{Eq. 3-3}$$

$$V_T = 2 \cdot \left[\frac{4}{3} \pi \cdot r^3 \right] - V_o \quad \text{Eq. 3-4}$$

$$f_o = \frac{V_o}{V_T} \quad \text{Eq. 3-5}$$

V_o = Overlap volume

V_T = Total volume

f_o = fraction of overlapped volume

$$\frac{\text{Rate}}{k} = \frac{V_o}{V_T} \cdot C \cdot C = f_o \cdot C^2 \quad \text{Eq. 3-6}$$

k = kinetic constant

3.5 References

- 1 Y. He and D. R. Liu, *Nat. Nanotechnol.*, 2010, **5**, 778–782.
- 2 W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes, A. H. El-Sagheer, J. Bath, B. G. Davis, T. Brown, R. K. O'Reilly and A. J. Turberfield, *Nat. Chem.*, 2016, **8**, 542–548.
- 3 J. R. Peacock, R. R. Walvoord, A. Y. Chang, M. C. Kozlowski, H. Gamper and Y.M. Hou, *RNA*, 2014, **20**, 758–764.
- 4 D. F. Detar and C. Delahunty, *J. Am. Chem. Soc.*, 1983, **105**, 2734–2739.
- 5 V. Ramakrishnan, *Cell*, 2002, **108**, 557–572.
- 6 M. D. Erlacher and N. Polacek, *RNA Biol.*, 2008, **5**, 5–12.
- 7 B. Alberts; A. Johnson; J. Lewis; M. Raff; K. Roberts, *Molecular biology of the cell*, Garland Science, New York, Fourth Edition., 2002.
- 8 N. V. Voigt, T. Tørring, A. Rotaru, M. F. Jacobsen, J. B. Ravnsbæk, R. Subramani, W. Mamdouh, J. Kjems, A. Mokhir, F. Besenbacher and K. V. Gothelf, *Nat. Nanotechnol.*, 2010, **5**, 200–203.
- 9 Z. Ma and J. S. Taylor, *Proc. Natl. Acad. Sci.*, 2000, **97**, 11159–11163.
- 10 J. Cai, X. Li, X. Yue and J. S. Taylor, *J. Am. Chem. Soc.*, 2004, **126**, 16324–16325.
- 11 A. Roloff and O. Seitz, *Chem. Sci.*, 2013, **4**, 432–436.
- 12 W. Yang and D. G. Drueckhammer, *J. Am. Chem. Soc.*, 2001, **123**, 11004–11009.
- 13 Q. Wan, J. Chen, Y. Yuan and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2008, **130**, 15814–15816.
- 14 G. M. Fang, H. K. Cui, J. S. Zheng and L. Liu, *ChemBioChem*, 2010, **11**, 1061–

- 1065.
- 15 P. Botti, M. Villain, S. Manganiello and H. Gaertner, *Org. Lett.*, 2004, **6**, 4861–4864.
 - 16 B. Lewandowski, G. De Bo, J. W. Ward, M. Papmeyer, S. Kuschel, M. J. Aldegunde, P. M. E. Gramlich, D. Heckmann, S. M. Goldup, D. M. D'Souza, A. E. Fernandes and D. A. Leigh, *Science*, 2013, **339**, 189–193.
 - 17 G. De Bo, S. Kuschel, D. A. Leigh, B. Lewandowski, M. Papmeyer and J. W. Ward, *J. Am. Chem. Soc.*, 2014, **136**, 5811–5814.
 - 18 G. De Bo, M. A. Y. Gall, M. O. Kitching, S. Kuschel, D. A. Leigh, D. J. Tetlow and J. W. Ward, *J. Am. Chem. Soc.*, 2017, **139**, 10875–10879.
 - 19 V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J. C. M. Monbaliu and O. Melnyk, *Chem. Rev.*, 2019, **119**, 7328–7443.
 - 20 H. M. Burke, L. McSweeney and E. M. Scanlan, *Nat. Commun.*, 2017, **8**, 15655.
 - 21 M. Lovrinovic and C. M. Niemeyer, *Biochem. Biophys. Res. Commun.*, 2005, **335**, 943–948.
 - 22 M. Lovrinovic, M. Spengler, C. Deutsch and C. M. Niemeyer, *Mol. Biosyst.*, 2005, **1**, 64–67.
 - 23 A. V. Azhayev and M. L. Antopolsky, *Tetrahedron*, 2001, **57**, 4977–4986.
 - 24 R. J. Hondal, B. L. Nilsson and R. T. Raines, *J. Am. Chem. Soc.*, 2001, **123**, 5140–5141.
 - 25 N. M. Green, *Adv. Protein Chem.*, 1975, **29**, 85–133.
 - 26 G. Mayer, *Nucleic Acid and Peptide Aptamers*, Humana Press, Totowa, NJ, 2009, 535.
 - 27 L. Chaiet and F. J. Wolf, *Arch. Biochem. Biophys.*, 1964, **106**, 1–5.

- 28 H. Murakami, D. Kourouklis and H. Suga, *Chem. Biol.*, 2003, **10**, 1077–1084.
- 29 N. Matuszak, G. G. Muccioli, G. Labar and D. M. Lambert, *J. Med. Chem.*, 2009, **52**, 7410–7420.
- 30 B. Neises and W. Steglich, *Angew. Chemie Int. Ed.*, 1978, **17**, 522–524.
- 31 T. Lindahl, *Angew. Chemie Int. Ed.*, 2016, **55**, 8528–8534.
- 32 T. N. Grossmann and O. Seitz, *J. Am. Chem. Soc.*, 2006, **128**, 15596–15597.
- 33 D. MacMillan, A. Adams and B. Premdjee, *Isr. J. Chem.*, 2011, **51**, 885–899.
- 34 B. Cowper, L. Shariff, W. Chen, S. M. Gibson, W. L. Di and D. MacMillan, *Org. Biomol. Chem.*, 2015, **13**, 7469–7476.
- 35 T. T. Mihaylov, T. N. Parac-Vogt and K. Pierloot, *Chem. A Eur. J.*, 2014, **20**, 456–466.
- 36 D. Li, X. Wang, F. Shi, R. Sha, N. C. Seeman and J. W. Canary, *Org. Biomol. Chem.*, 2014, **12**, 8823–8827.
- 37 A. C. Conibear, E. E. Watson, R. J. Payne and C. F. W. Becker, *Chem. Soc. Rev.*, 2018, **47**, 9046–9068.
- 38 L. Van Oeffelen, E. Peeters, P. Nguyen Le Minh and D. Charlier, *PLoS One*, 2014, e85146.
- 39 J. Michaelis, G. J. Van Der Heden Van Noort and O. Seitz, *Bioconjug. Chem.*, 2014, **25**, 18–23.
- 40 M. L. McKee, A. C. Evans, S. R. Gerrard, R. K. O'Reilly, A. J. Turberfield and E. Stulz, *Org. Biomol. Chem.*, 2011, **9**, 1661–1666.
- 41 K. M. Wilcoxon, L. J. Leman, D. A. Weinberger, Z. Z. Huang and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2007, **129**, 748–749.
- 42 Y. He and D. R. Liu, *J. Am. Chem. Soc.*, 2011, **133**, 9972–9975.

- 43 P. J. Milnes, M. L. McKee, J. Bath, L. Song, E. Stulz, A. J. Turberfield and R. K. O'Reilly, *Chem. Commun.*, 2012, **48**, 5614–5616.
- 44 R. K. Bruick, P. E. Dawson, S. B. H. Kent, N. Usman and G. F. Joyce, *Chem. Biol.*, 1996, **3**, 49–56.
- 45 J. Šečkute, J. Yang and N. K. Devaraj, *Nucleic Acids Res.*, 2013, **41**, 148–157.
- 46 F. A. Hays, A. Teegarden, Z. J. R. Jones, M. Harms, D. Raup, J. Watson, E. Cavaliere and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7157–7162.
- 47 M. J. Catalano, N. E. Price and K. S. Gates, *Bioorganic Med. Chem. Lett.*, 2016, **26**, 2627–2630.
- 48 E. K. Y. Leung, N. Suslov, N. Tuttle, R. Sengupta and J. A. Piccirilli, *Annu. Rev. Biochem.*, 2011, **80**, 527–555.
- 49 M. Beringer, C. Bruell, L. Xiong, P. Pfister, P. Bieling, V. I. Katunin, A. S. Mankin, E. C. Böttger and M. V. Rodnina, *J. Biol. Chem.*, 2005, **280**, 36065–36072.
- 50 A. Sievers, M. Beringer, M. V. Rodnina and R. Wolfenden, *Proc. Natl. Acad. Sci.*, 2004, **101**, 7897–7901.
- 51 A. R. Hesser, B. M. Brandsen, S. M. Walsh, P. Wang and S. K. Silverman, *Chem. Commun.*, 2016, **52**, 9259–9262.
- 52 A. J. Camden, S. M. Walsh, S. H. Suk and S. K. Silverman, *Biochemistry*, 2016, **55**, 2671–2676.
- 53 J. Morimoto, Y. Hayashi, K. Iwasaki and H. Suga, *Acc. Chem. Res.*, 2011, **44**, 1359–1368.
- 54 J. Sayers, R. J. Payne and N. Winssinger, *Chem. Sci.*, 2018, **9**, 896–903

Chapter 4

Peptide nucleic acids (PNAs) as potential templates for the synthesis of peptide bonds in organic solvent

4.1 Introduction: PNAs and their hybrid properties between peptide and nucleobase materials.

4.1.1 PNA hybridisation in the presence of organic solvent

The interactions between nucleic acid strands consist of a delicate balance between attractive forces; such as hydrogen-bonding, π - π stacking and hydrophobic interactions; and electrostatic repulsion due to the anionic character of the poly-phosphate backbone. As a result, the secondary structure of nucleic acids is very sensitive to changes in solvent conditions.¹⁻³

With the exception of glycerol and ethylene glycol,⁴ the addition of organic co-solvent to an aqueous buffer can have a very detrimental impact on the stability of DNA duplexes (dsDNA).⁵ Indeed, it was previously shown in this thesis that the addition of small volumes of DMF (> 2 vol%) had a negative impact on the thermal stability of dsDNA (see Chapter 2, Experimental Section 2.4.12).

Precise control over the secondary structure of nucleic acids is crucial in order to exploit it in genetic programming and templation of chemical reactions. To date, there are only very limited examples of the use of organic solvent in DNA templated synthesis (DTS). McKee, Turberfield, O'Reilly *et. al.* reported an improvement on the yield of peptidomimetic DTS upon the addition of 10 vol% THF.⁶ DTS has also been

reported on pre-annealed DNA duplexes transferred from aqueous buffered solution into organic solvents.⁷ Furthermore, polyethylene glycol (PEG)-conjugated G-quadruplex DNA pre-annealed in aqueous buffer showed some degree of preservation of the secondary structure in organic solvent by circular dichroism.⁸ These previous strategies highlight the complexity of developing DNA nanotechnology in organic solvents. However, there are nucleic acid analogues that have been reported to tolerate high proportions of organic solvent while preserving their secondary structure. Peptide nucleic acids (PNAs) are nucleic acid analogues, originally described by Nielsen *et. al.* in 1991.⁹ Structurally, they are very similar to DNA (Figure 4-1), their building blocks incorporate conventional DNA and RNA nucleobases (*i.e.* A, T, G, C and U); however, instead of an anionic polyphosphate backbone, they contain a neutral polyamide backbone.

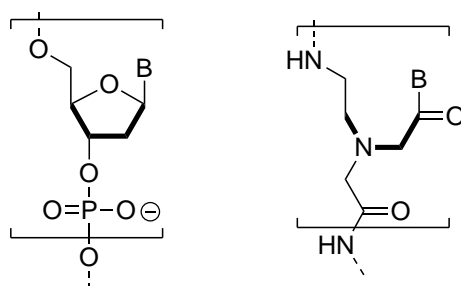


Figure 4-1. Structural comparison between the repeating unit of DNA (left) and PNA (right).

PNAs can form duplexes (dsPNA) analogous to their DNA counterparts, following Watson-Crick hybridisation modes.¹⁰⁻¹² Due to the lack of repulsive electrostatic interactions between strands, dsPNA display higher stability than dsDNA and do not require the presence of counterions in the solvent.¹³ Due to this, PNAs are excellent candidates for the formation of supramolecular assemblies in organic solvents. Sen and Nielsen reported the detection of dsPNA in 50¹⁴ and 70¹⁵ vol% DMF / H₂O solution

using UV spectroscopy, constituting the only systematically studied organic solvent for dsPNA that could be found.

Upon the addition of organic solvent, the reduction in positive hydrophobic interactions was compensated by stronger hydrogen-bonding interactions, which to a large extent preserved the stability of the duplex. Based on the observed linear decrease of dsPNA melting temperature (T_m) as function of the DMF vol%, it was expected that the T_m could be around 50 °C in pure DMF for the studied 10-mer dsPNA (Figure 4-2). While these results were extremely promising, the T_m of dsPNA in solutions with more than 70 vol% DMF could not be determined by UV spectroscopy due to the intense absorbance of the solvent in the wavelengths where the hyperchromic effect is measurable (~260 nm).¹⁶

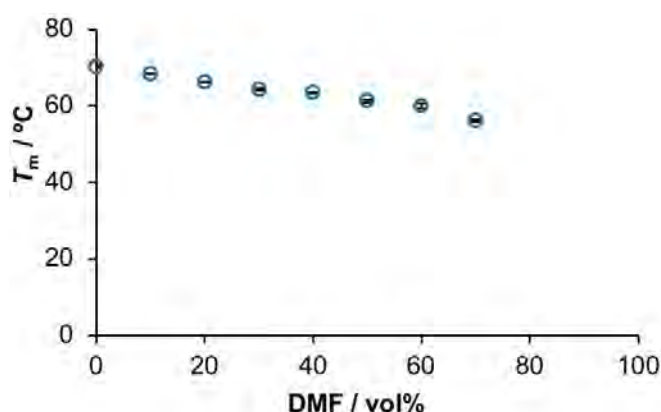


Figure 4-2. Representation of evolution of the T_m of a 10-mer dsPNA in the presence of DMF, determined by UV spectroscopy.¹⁵

The use of PNAs in nucleic acid templated synthesis has been largely restricted to the use of PNA·DNA and PNA·RNA hybrid duplexes.^{17–21} However, there is a preliminary study using dsPNA and non-activated esters to template the formation of peptide bonds resulting in low yield templated reactions (0.5 to 5%).²²

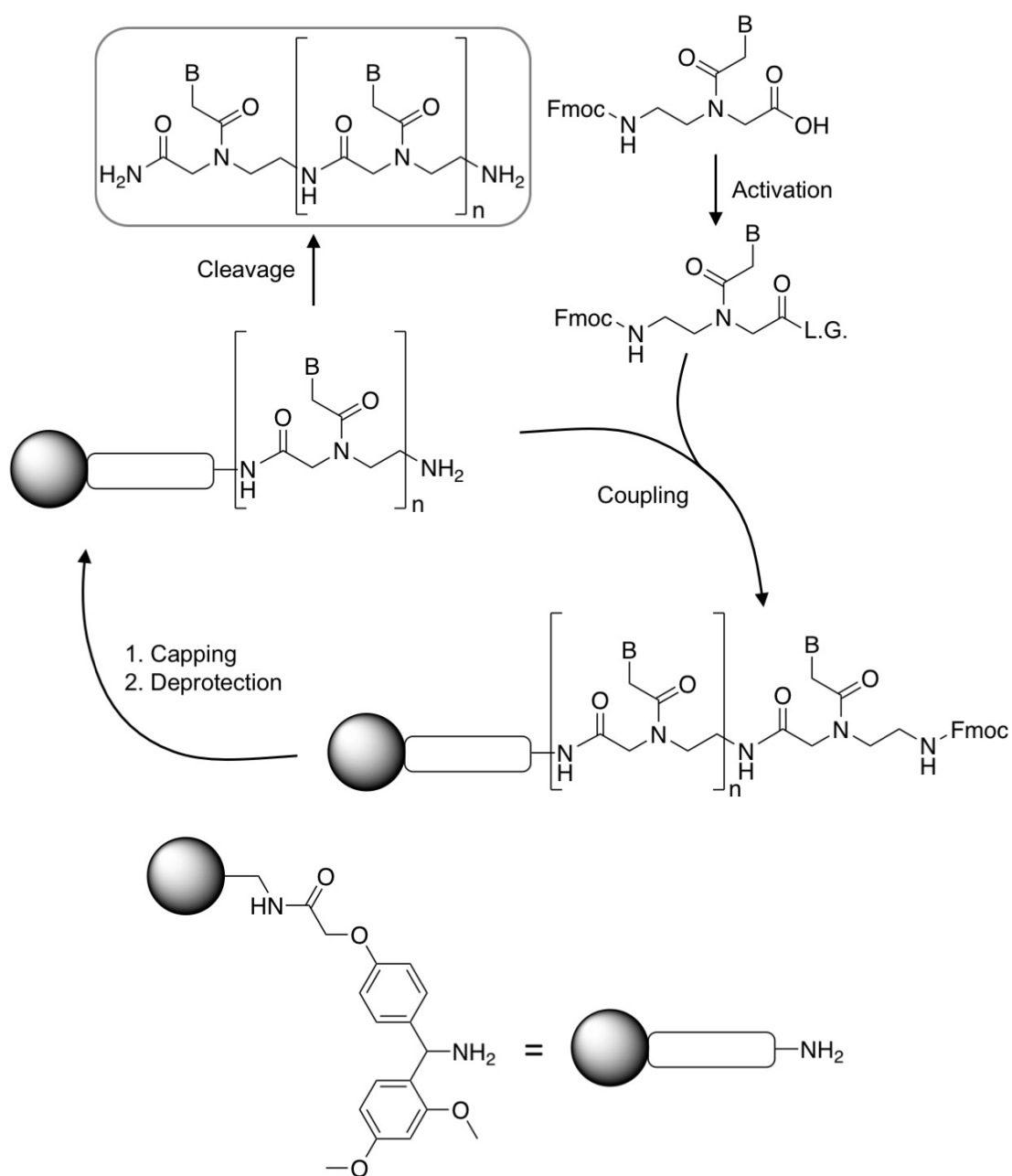
4.1.2 Solid phase synthesis of PNA

The introduction of solid phase synthesis (SPS) techniques by Merrifield,²³ has allowed the preparation of long peptides²⁴ through using a large excess of reagents, followed by simple purification through washing steps. Analogously to peptides, PNAs are formed by a polyamide backbone; therefore, they can be effectively prepared by SPS using analogous chemistry.

A generic scheme of the SPS of PNAs is shown below (Scheme 4-1). A detailed description of SPS can be found in the introductory chapter (Chapter 1, Section 1.3.2). Two general strategies have been described for the SPS of PNAs: Fmoc / Bhoc^{‡‡} and Boc / Cbz^{§§} protecting group strategies.²⁵ Due to the compatibility with different Fmoc-based syntheses performed in the SPS automated setup used here, the Fmoc strategy was followed for the preparation of PNA strands in this work.

^{‡‡} Bhoc: benhydryloxycarbonyl. Required to protect the exocyclic amines in A, G and C bases.

^{§§} Cbz: carboxybenzyl. Required to protect the exocyclic amines in A, G and C bases.



Scheme 4-1. Representation of the SPS cycle of PNA on a rink amide solid phase surface. L.G. = leaving group.

4.1.3 Design of a model double-stranded PNA (dsPNA) system

The relationship between DNA primary structure (or sequence) and their secondary structure has been widely studied.^{26–29} The models to predict the secondary structure of short oligonucleotides have been shown extremely useful and powerful for

nanotechnology applications such as the design of DNA origami³⁰ and DNA motors.³¹ Due to the usefulness of such techniques, open source software has been developed to facilitate the design of nucleic acids assemblies.^{32,33} DNA secondary structure design tools can be applied, with limitations, to the design of simple PNA structures, due to the presence of purine and pyrimidine bases. In this work, the toolkit known as: nucleic acid package (Nupack) has been used for the analysis and design of PNA duplexes (Figure 4-3).³⁴

It is important to stress this result has been determined from tools optimised for DNA assemblies and cannot be directly extrapolated for PNA assemblies, as the interaction parameters will be different. However, this is a useful approach when assuming simple Watson-Crick PNA·PNA interactions in the design of short strands.³⁵

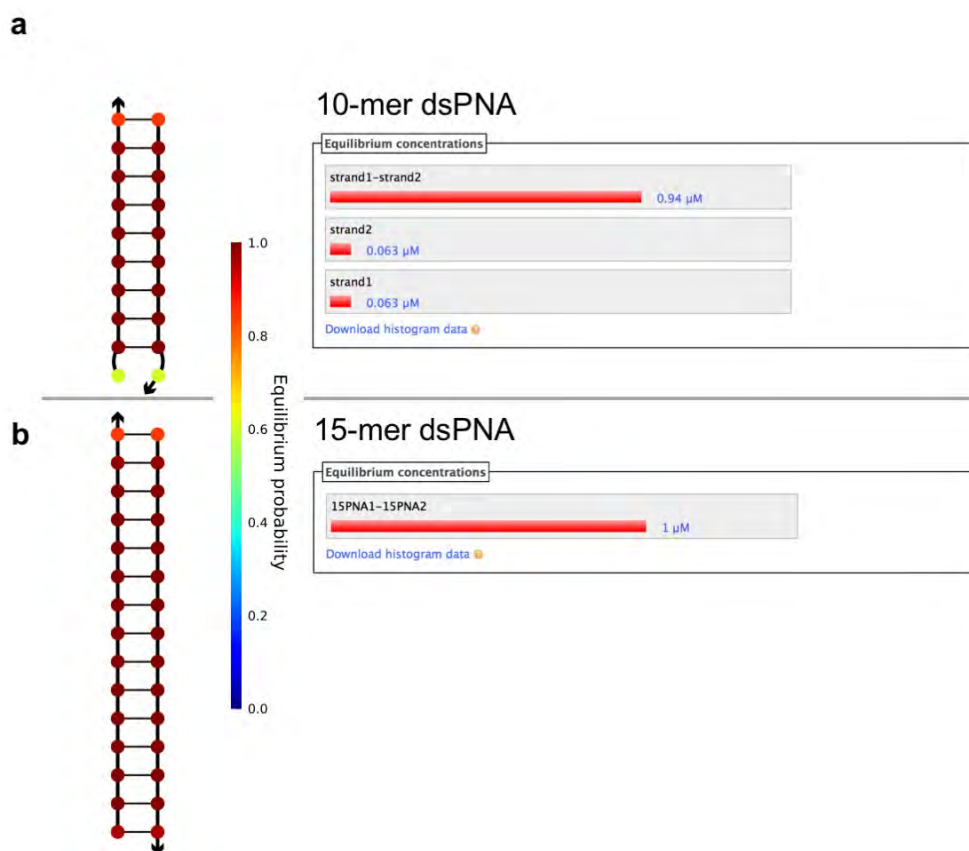


Figure 4-3. Elongation of the reported 10-mer dsPNA using the web application Nupack.³⁴ a) 10-mer dsPNA sequence previously studied.^{14,15} b) Elongation of the previous 10-mer dsPNA into a 15-mer

dsPNA. All possible 1, 2 and 3-strand complexes were considered. The code for the elongation of the strands is available (Experimental section 4.4.4).

Here, two sets of dsPNA strands were designed and used. The initial studies were based on previously reported 10-mer sequences,^{14,15} and for further studies, an extra 5 bases were added using Nupack,³⁴ resulting in a 15-mer dsPNA sequence. Several designs were obtained, and these were analysed to ensure that no alternative folding was likely. Despite the limitations previously mentioned, this approach prevents potential unforeseen manual design errors.

Several PNA secondary structures have been described, such as duplex,^{10,12,36–38} PNA·PNA·PNA triplex³⁹ and hybrid DNA·PNA G-quadruplex⁴⁰ among other hybrid structures.³⁸ However, in comparison to the DNA and RNA secondary structures, there is less previous research on the secondary structures of PNA; therefore, there is a certain degree of uncertainty on the prediction of duplex forming PNA sequences.

4.1.4 Analytical methods to study the thermal stability of dsPNA in the presence of organic solvent

It was previously mentioned that the thermal stability of dsPNA in H₂O / DMF mixtures had been determined using UV spectroscopy; however, this approach was limited to up to 70 vol% DMF due to the intense absorption of DMF in the UV region.¹⁵ Alternative robust methodologies have been previously used to determine the thermal stability of nucleic acid duplexes in aqueous solutions: the stability of nucleic acid duplexes has been previously studied using UV and fluorescence spectroscopy (Figure 4-4).

UV thermal stability studies rely upon the hyperchromic effect, which is the higher absorbance that DNA displays at 260 nm upon melting (*i.e.* the dehybridisation of the complementary strands).⁴¹ This can be used to determine the melting temperature (T_m), along with the thermodynamic parameters (*i.e.* ΔH , ΔS and ΔG ; Figure 4-4 a).¹⁵

To determine the stability of nucleic acid duplexes using fluorescence spectroscopy, the complementary strands have to be labelled with either a fluorophore / quencher, or a donor / acceptor, FRET pair.⁴² In this approach, the nucleic acid system transitions from a high quenching or FRET state to a low quenching or FRET state, due to the increased distance between the fluorophore and quencher upon melting (Figure 4-4 b).

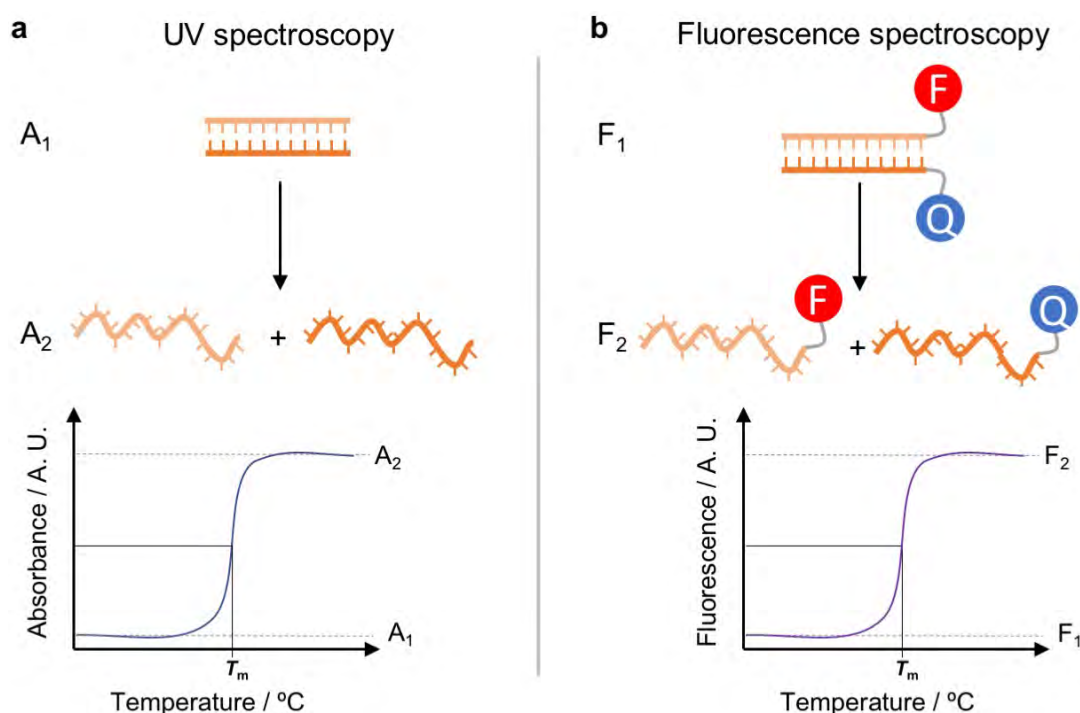


Figure 4-4. Spectroscopic determination of the T_m of dsPNA by a) UV spectroscopy and b) fluorescence spectroscopy. F = fluorophore, Q = quencher.

In addition to spectroscopic methods, calorimetry has been used to study the thermal stability of nucleic acid duplexes for example, the stability of PNA·DNA and DNA·DNA duplexes have been determined by differential scanning calorimetry (DSC).⁴³ DSC is based on applying a constant heat flow to a sample solution containing the analyte, and a blank solvent solution, simultaneously. The different changes in temperature between the blank and the analyte containing solution can be used to determine the

molar excess heat at constant pressure (C_p), defined as the amount of heat required to increase the temperature of the solutions. The representation of the C_p as function of the temperature allows to study the thermal denaturing transition of nucleic acids, analogously to UV and fluorescence spectroscopy (Figure 4-5).

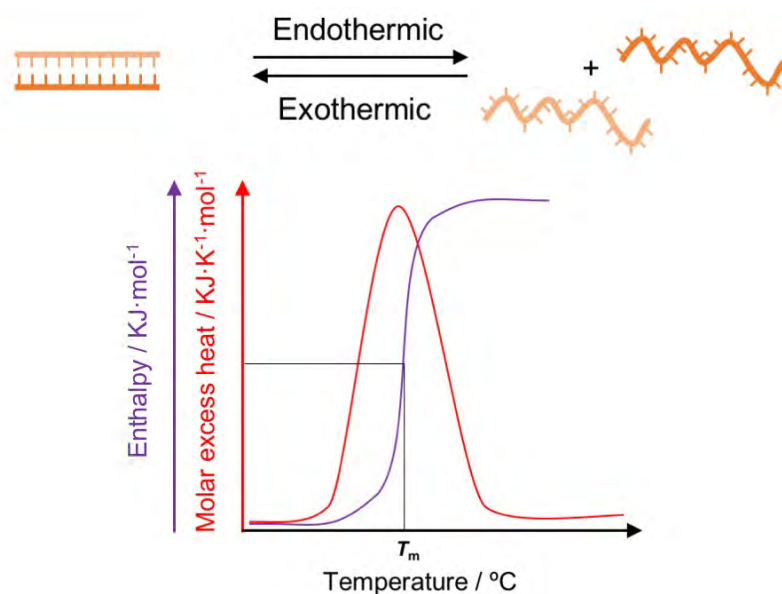


Figure 4-5. Calorimetric determination of the T_m of dsPNA.

In principle, the previous three techniques can all be used to determine the T_m of dsPNA, used as a proxy of the thermal stability of the duplex, in H_2O / DMF mixtures (Table 4-1). These methods however, have distinct advantages and disadvantages. UV absorbance is severely limited by the spectroscopic properties of the solvent; however, it does not require the chemical modification of the strands and it has a moderate to high sensitivity. Fluorescence is one of the most sensitive and specific analytical techniques available,⁴⁴ but it requires the labelling of complementary PNA strands with a fluorophore and a quencher; in addition, while fluorescence *per se* is compatible with a wide range of solvents, the fluorophores must be optimised for the solvent of interest.⁴⁵ Finally, calorimetry is a technique sensitive to any transitions that

involve a heat exchange, and therefore have a non-zero enthalpy, which is the vast majority of processes involving biomacromolecules. In addition, it is not affected by the spectroscopic properties of the solvent and it does not require the labelling of the strands. The main drawback of calorimetry is its low sensitivity (e.g. experiments on dsDNA have been reported between $1 \text{ mg}\cdot\text{mL}^{-1}$ and up to $55 \text{ mg}\cdot\text{mL}^{-1}$).^{46,47} With the advent of microcalorimetry (*i.e.* measurements of heat flows in the order of $\mu\text{J}\cdot\text{s}^{-1}$ or μW) this limitation was partially overcome, and multiple applications have been reported analysing biomacromolecules⁴⁸ and even microorganisms.⁴⁹

Table 4-1. Comparison of the different techniques that can be used to study the thermal stability of dsPNA. *Fluorescence spectroscopy require dyes with high quantum yield in each individual solvent.

Feature	UV spectroscopy	Fluorescence spectroscopy	Calorimetry
Sensitivity	Moderate	High	Low
Chemical modification	None required	Required	None required
Solvent range	Narrow	Wide*	Wide

The reported studies of the thermal stability of dsPNA duplexes in H_2O / DMF mixtures were limited by the UV spectroscopy methodology, and its compatibility with DMF.^{14,15} Extending the previous investigations with fluorescence and microcalorimetry would presumably allow to determine the thermal stability of dsPNA duplexes in up to 100% organic solvent; microcalorimetry will be used to systematically determine the thermal stability of dsPNA in several H_2O / DMF mixtures, while fluorescence will be used to study the isothermal hybridisation in high organic solvent. The discovery of a genetically programmable template such as PNAs, capable of displaying sequence-

dependent hybridisation in organic solvent could potentially incorporate highly reactive activated esters for nucleic acid templated peptide synthesis, without incurring in the degradation issues which are characteristic of aqueous solutions (Chapter 2, Section 2.2.2).

4.2 Results and discussion

4.2.1 Synthesis of 10-mer and 15-mer PNA strands

As a starting point, the previously reported 10-mer complementary PNA sequences were synthesised.^{14,15} In order to study the effect of elongating the PNA strands on the duplex formation in organic solvent, 15-mer strands were designed (Section 4.1.3) and synthesised. This was performed on an automated microwave SPS LibertyBlue instrument (CEM corporation). The availability of an automated SPS setup was crucial for the synthesis of long strands; a manual synthesis of a 15-mer PNA would require 15 × 1 h couplings, 15 × 5 min coupling washes, 30 × 10 min deprotections, and finally, 15 × 5 min deprotection washes. This would amount to a total of ~24 h of manual intervention. On an automated instrument, in syntheses that are started early on the day, the final resin-bound product can be ready for cleavage on the next working day. Conventional amino acids for the SPS of peptides are available in large scale however, PNA monomers are only available on very small scale.^{***} Therefore, PNA synthesis was performed at 10 µmol scale and, to allow the use of reproducible volumes by the automated SPS synthesiser at such a small scale, PNA monomer solutions were prepared at 0.1 M in DMF, while conventional amino acids are usually prepared at 0.2 M.⁵⁰ This necessitated 1 h coupling steps, due to the slower kinetics. The smaller

^{***} *E.g.* Fmoc-Trp(Boc)-OH £ 3.44 · g⁻¹ - Merck Millipore catalogue (2019). Fmoc-PNA-A(Bhoc)-OH £ 150.00 · g⁻¹ - LGC genomics catalogue (2019).

synthesis scale also required other minor changes such as turning off the solvent drain detectors and to adjust the volume of DMF washes to ensure the complete removal of piperidine from the reaction vessel. From the chemistry perspective, the SPS was set up according to previously reported procedures.^{25,51}

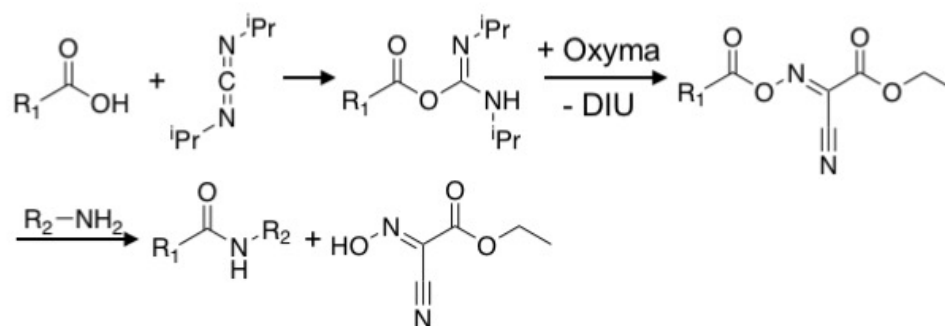
4.2.1.1 Resin chemistry

PNAs have been traditionally synthesised on rink amide resins (Scheme 4-1). Upon cleavage, this produces a terminal carboxamide in the C terminus (C_t), which is stable to most chemical reagents, and non-pH responsive.⁵²⁻⁵⁵

4.2.1.2 Coupling of monomers

Both Fmoc / Bhoc^{†††} and Boc / Cbz protected PNA monomers are commercially available, and have been extensively studied.²⁵ Here, the Fmoc / Bhoc method was used due to the similarity with other syntheses performed in the automated SPS instrumentation, reducing the risk of cross-contamination. A widely used methodology to activate carboxylic acids for SPS is the *N,N'*-diisopropylcarbodiimide (DIC) / ethyl cyano(hydroxyimino)acetate (oxyma) activation.⁵⁶ Despite a safety concern regarding the evolution of hydrogen cyanide,⁵⁷ this methodology has shown high coupling efficiency,^{51,58} low optical purity loss⁵¹ and versatility,⁵⁶ while being compatible with mild conditions (e.g. 1 to 2 h coupling times at room temperature).^{51,59} The proposed intermediates are shown below (Scheme 4-2).⁵⁶

^{†††} Bhoc: benzhydroxycarbonyl



Scheme 4-2. Oxyima / DIC activation and coupling proposed intermediates and products.⁵⁶ DIU = 1,3-diisopropylurea.⁵⁶

The oxyima / DIC coupling approach produced a series of positive results: the coupling efficiencies were high, which translated into a high yield of the full length PNA sequence, and a very minor fraction of truncated products. However, over time the purity of newly synthesised sequences decreased, which the reasons behind this were unknown and required a thorough investigation. Initially, two batches of oxyima from different sources were compared.^{‡‡} Batch 1 produced only the desired PNA product, while batch 2 produced a complex mixture of products (Figure 4-6 a and b; analysis based on HPLC and LC-MS). The syntheses were continued using batch 1; however, after storing the oxyima batch 1 for ~4 months, the initial PNA synthesis could not be reproduced (Figure 4-6 c). It was hypothesised that some sort of degradation process was taking place. However, in order to rule out a problem with the solid phase synthesiser, the oxyima / DIC PNA synthesis was repeated manually on a small-scale column, which again produced a complex mixture of products (Figure 4-6 d). The automated SPS of a conventional 6-mer peptide (H-Val-Phe-Lys-Phe-Gly-Lys-NH₂) using the same batch (1) of oxyima / DIC produced a highly pure peptide (Figure 4-6 e); however, the high yielding SPS could not be reproduced for PNA strands. Finally,

^{‡‡} Batch 1: CEM corporation Oxyima Pure® #S001-C. Batch 2: Merck Millipore Oxyima Pure® 8510860025.

a PyBOP (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate) / 2,6-lutidine : (iPr)₂NEt previously reported PNA synthesis method⁶⁰ was used both manually and in the automated synthesiser. This successfully produced the desired PNA product (Figure 4-6 f). To summarise, oxyma / DIC produced excellent SPS results initially, however, over time the purity of the products decreased, most likely due to a decomposition of the oxyma reagent. This did not affect the quality of conventional peptide synthesis containing Val, Phe, Gly and Lys amino acids.

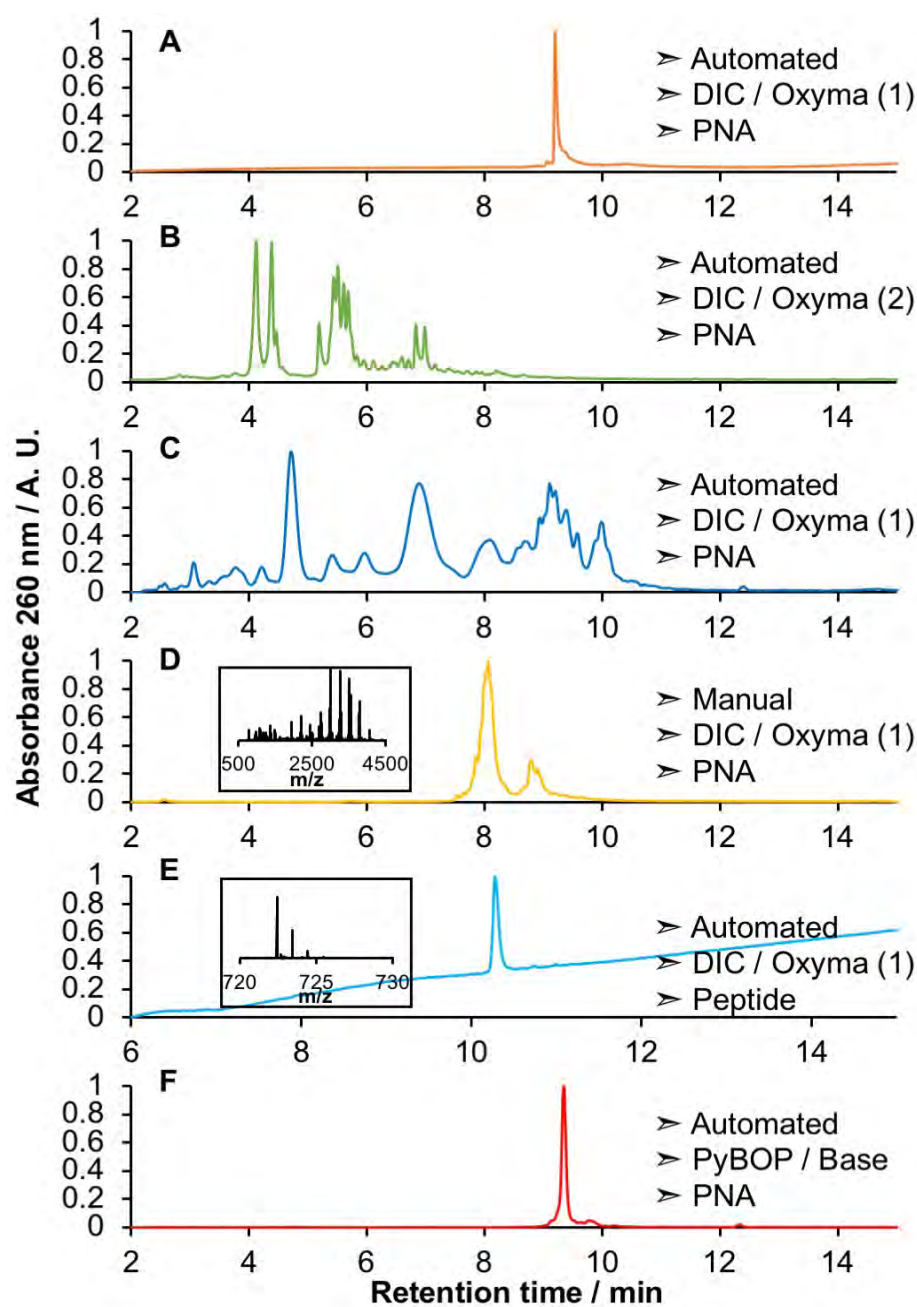
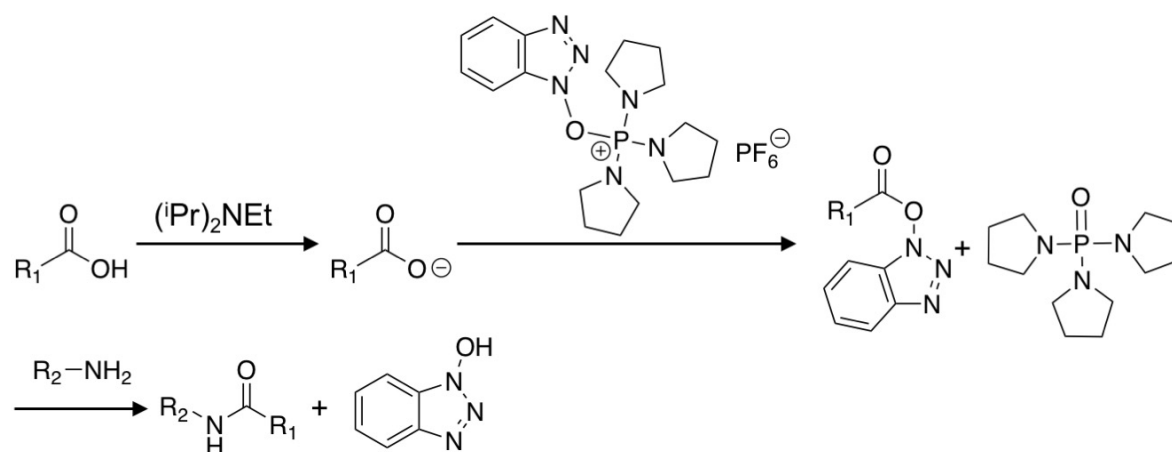


Figure 4-6. Investigation of the coupling efficiencies using oxyma / DIC and PyBOP / 2,6-lutidine : $(i\text{Pr})_2\text{NEt}$. a) Initial automated SPS results using batch 1 oxyma. b) Failed automated SPS using batch 2 oxyma. c) Failed automated SPS using batch 1 oxyma. d) Failed manual SPS using batch 1 oxyma. e) Successful synthesis of a 6-mer peptide (H-Val-Phe-Lys-Phe-Gly-Lys-NH₂) using batch 1 oxyma. f) Successful automated SPS of PNA (**4-1**) using PyBOP / 2,6-lutidine : $(i\text{Pr})_2\text{NEt}$ (Experimental Section 4.4.3).

The current data do not fully explain the decrease in efficiency of oxyma / DIC mediated couplings over time - oxyma is known for its high stability over a range of temperatures⁵¹ - and the present results indicate that conventional peptides did not exhibit the drift towards low coupling efficiencies.

After the unsuccessful SPS using oxyma / DIC, a PyBOP coupling strategy was adopted. PyBOP is part of the phosphonium coupling agents class, typically used in substitution of the BOP reagent (benzotriazol-1-yloxytris(dimethylamino) phosphoniumhexafluorophosphate) due to its lower toxicity.⁶¹ The proposed intermediates and products for the coupling are shown below (Scheme 4-3).⁶²



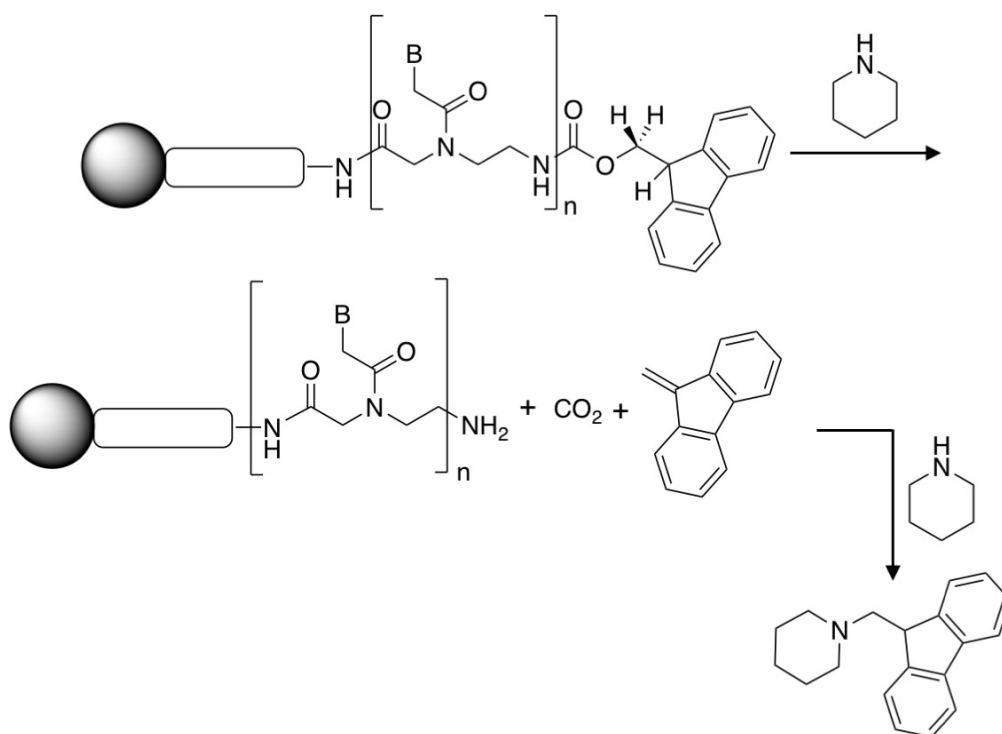
Scheme 4-3. PyBOP activation and coupling. Proposed intermediates and products.⁶²

The syntheses of 15-mer PNAs (**4-3** and **4-4**) were completed using PyBOP / $(iPr)_2NEt$: 2,6-lutidine couplings in high purity (>90% HPLC integration at 260 nm). Over time, the build-up of an exclusively water-soluble precipitate (not soluble in MeOH, EtOH, DMF and NMP) in the synthesiser lines and in the UV cell was observed, this was attributed to the formation of $(iPr)_2NH_2^+ \cdot PF_6^-$ salt. At the time that this issue was discovered, the PNA SPS required for the current experiments had already been completed; however, this remains an outstanding issue that will require optimising an

alternative set of reagents compatible with the automated solid phase synthesiser as this issue could damage the instrument. The identity of the base could be varied in an attempt to reduce the precipitate formation.

4.1.1.3 Fmoc deprotection

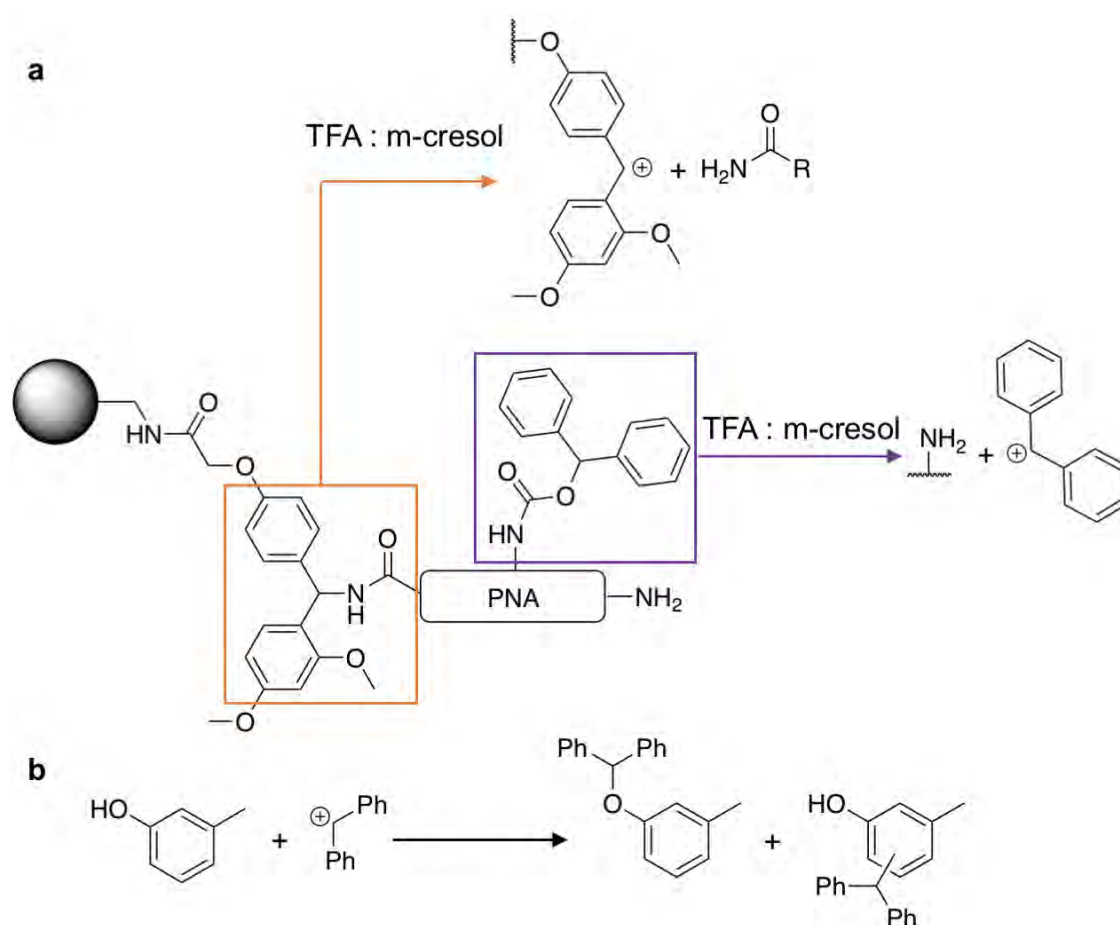
20 vol% piperidine in DMF is a conventional reagent for fast deprotection of the Fmoc group (Scheme 4-4). However, if an excess of piperidine is mixed with the subsequent activated monomer solution, it would greatly reduce the yield of the following coupling step. At the current SPS synthesis scale (10 μmol), a small residue of 20 vol% piperidine, in contact with the monomer-containing coupling solution, could potentially deactivate a large fraction of the PNA monomer. In order to reduce this risk, diluted 10 vol% piperidine solution in DMF was used instead.



Scheme 4-4. Intermediates and products of the Fmoc deprotection by piperidine.

4.2.1.4 Cleavage

Based on literature procedure, the rink amide resin was treated with 95 vol% trifluoroacetic acid (TFA) : 5 vol% m-cresol at room temperature for 90 min (Scheme 4-5 a).²⁵ This accomplished both the deprotection of the Bhoc groups and the cleavage from the solid support. m-Cresol performs a scavenging function, preventing side alkylation reactions between the carbocations produced during the treatment with TFA and the aromatic nucleobases (Scheme 4-5 b).²⁵



Scheme 4-5. TFA : m-cresol PNA cleavage and Bhoc deprotection. A) Simultaneous deprotection of the Bhoc-protected exocyclic amine groups and rink amide cleavage. B) Scavenging function of m-cresol: reaction with carbocations resulting from the acidic treatment.

The syntheses of unlabelled 10-mer PNA strands **4-1** and **4-2** were completed using DIC / oxyma mediated automated SPS (**4-1** 77% purity, HPLC integration; **4-2** 69%

purity HPLC integration). The synthesis of 15-mer PNA strands were prepared by automated SPS using a PyBOP / base methodology (**4-3** 91% purity, HPLC integration; **4-4** 93% purity, HPLC integration). The same surface chemistry, deprotection and cleavage approaches were applied to all syntheses.

The products were analysed by RP-HPLC and LC-MS (Figure 4-7). As micro-DSC required a substantial amount of PNA material, and the analysis showed high purity of the crude products, unlabelled strands were directly used after purification by precipitation into Et₂O. Fluorescently labelled PNA strands were HPLC fraction-collected, as in that second case, high purity was more crucial than quantity, due to the higher sensitivity of fluorescence analysis.

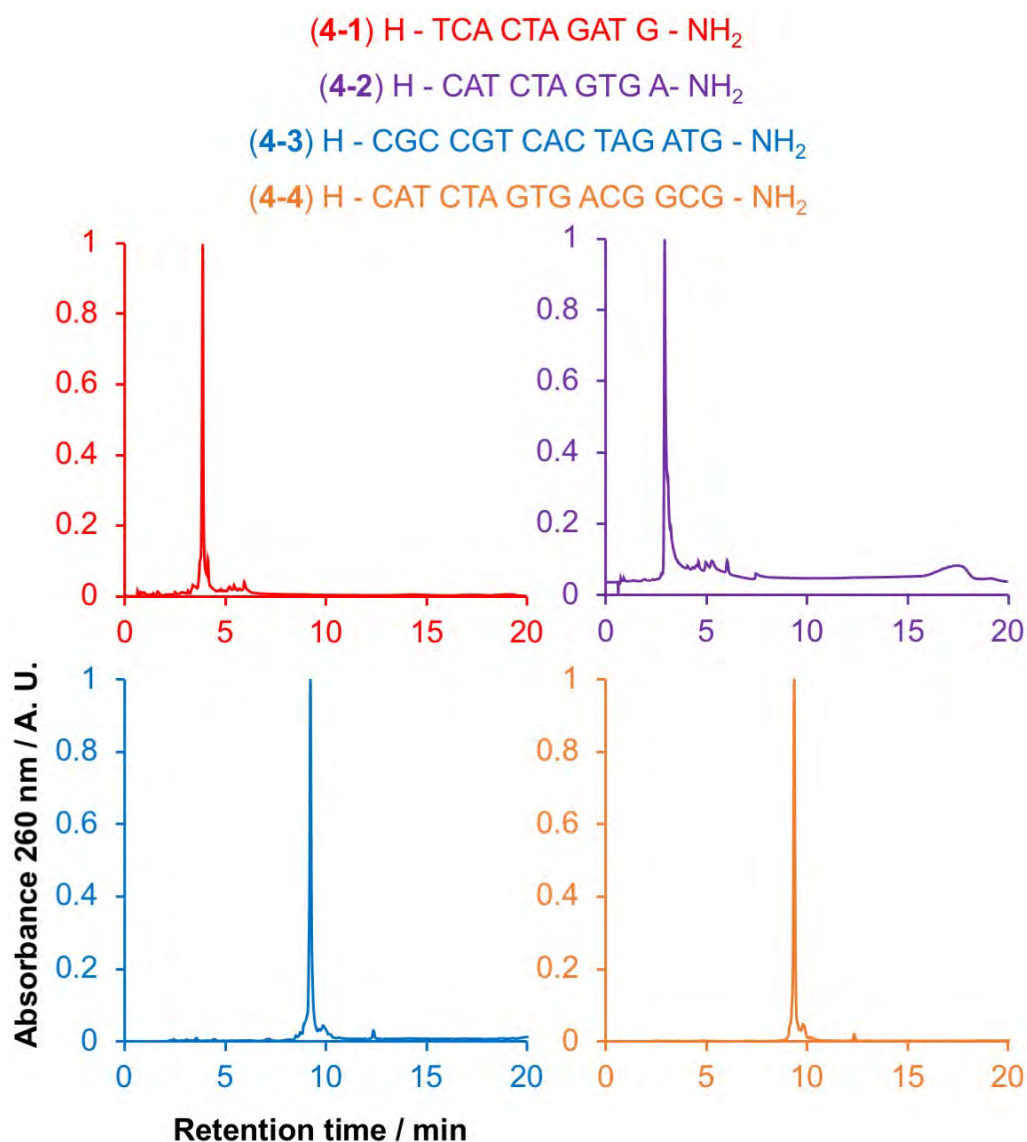


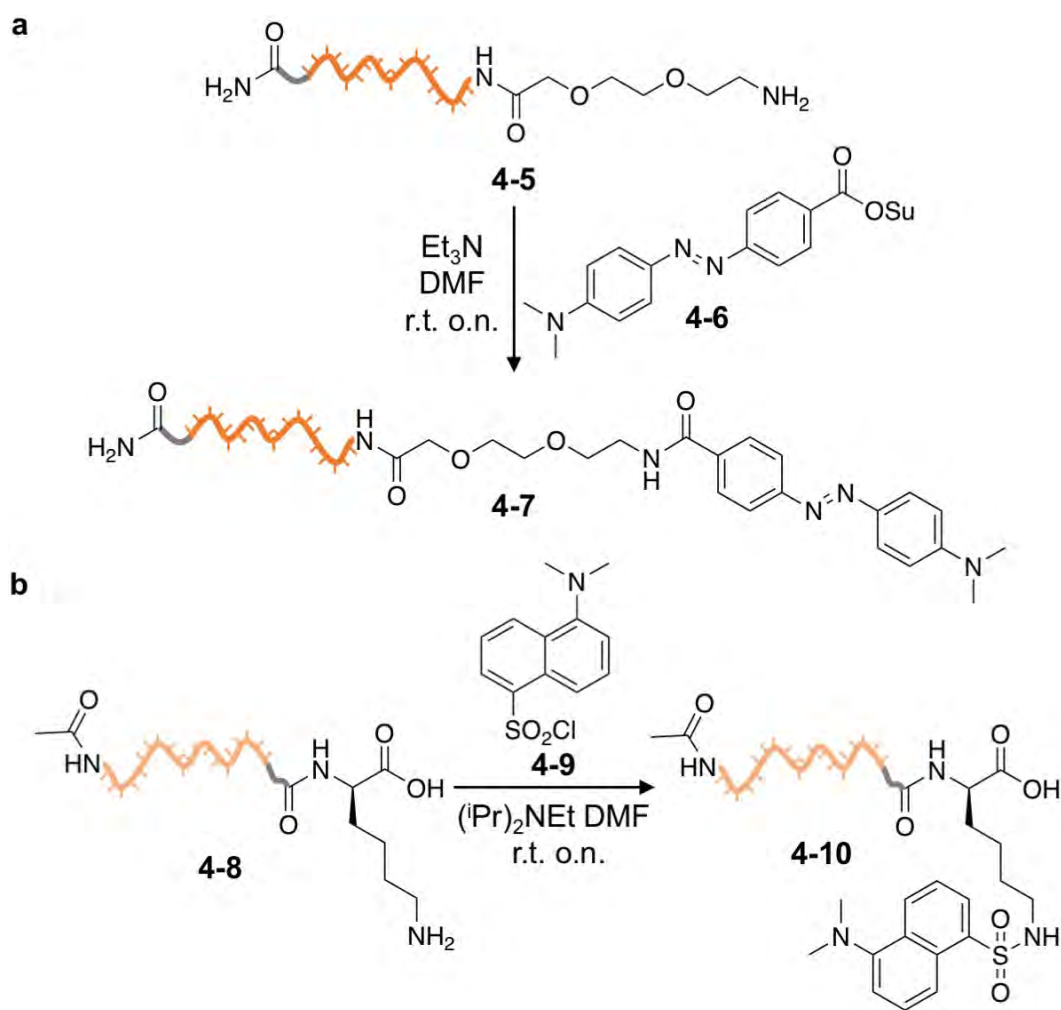
Figure 4-7. RP-HPLC analysis of unlabelled 10-mer and 15-mer strands. a) 10-PNA1 (1). b) 10-PNA2 (4-2). c) 15-PNA1 (4-3). d) 15-PNA2 (4-4). The products were analysed by LC-MS (Waters Xevo-G2-XS): **4-1** [M]⁰ *m/z* calcd. 2725.090 found 2725.130; **4-2** [M]⁰ *m/z* calcd. 2725.090 found 2725.119; **4-3** [M]⁰ *m/z* calcd. 4060.611 found 4060.660; **4-4** [M]⁰ *m/z* calcd. 4100.618 found 4100.451.

4.2.2 Synthesis of fluorophore and quencher labelled PNA

In order to simultaneously apply calorimetry and fluorescence spectroscopy to determine the T_m of dsPNA in high DMF / H₂O solutions, a suitable fluorophore and quencher-labelled dsPNA system was required. Dansyl fluorophore (5-(dimethylamino)naphthalene-1-sulfonamide) had a reported high quantum yield in

several organic solvents such as DMF ($\phi_F = 0.59$),⁶³ and there were precedents of bioconjugation of dansyl to biomacromolecules.⁶⁴ Dabcyl (4-((4-(dimethylamino)phenyl)azo)benzoate dark quencher had been previously used in conjunction with dansyl to generate a fluorophore / quencher pair⁶⁵ and its spectroscopic properties were reported in several organic solvents, including DMF.⁶⁶ Due to the potential incompatibility between both dansyl and dabcyl moieties with the cleavage conditions (95 vol% TFA : 5 vol% m-cresol), the two 10-mer PNA strands (**4-1** and **4-2**) were labelled post-synthetically (*i.e.* after the cleavage from the solid support; Scheme 4-6).

As opposed to dsDNA, dsPNA can be designed parallel (*i.e.* C_t complementary to C_t and N_t complementary to N_t); however, the assembly of antiparallel complementary PNA sequences is strongly favoured (*i.e.* C_t complementary to N_t).²⁵ In the current antiparallel design, the fluorophore and the quencher would be attached to the C_t and the N_t of complementary strands. Strand **4-5** bearing an ethylene glycol spacer was labelled at the N_t with a dabcyl-succinimide ester (**4-6**, Scheme 4-6 a) to produce the quencher strand **4-7**. The labelling of the C_t with dansyl required the acetylation of the N-terminus (N_t) and the addition of a Lys residue to the C_t, to provide a free primary amine (**4-8**); then, it was reacted with dansyl chloride (**4-9**, Scheme 4-6 b) to produce the fluorophore strand **4-10**.



Scheme 4-6. Post-synthetic modification of complementary 10-mer PNA sequences. a) Labelling of **4-5** with dabcyl to produce **4-7**. b) Labelling of **4-8** with dansyl to produce **4-10** (Experimental Section 4.4.5 and 4.4.6).

The results of these previous reactions were analysed by LC-MS. The synthesis of **4-7** showed the presence of several small molecules and a large fraction of the expected product (Figure 4-7), therefore the product was HPLC fraction-collected. The synthesis of **4-10** produced mainly the expected product (Figure 4-8) and the product was again HPLC fraction-collected.

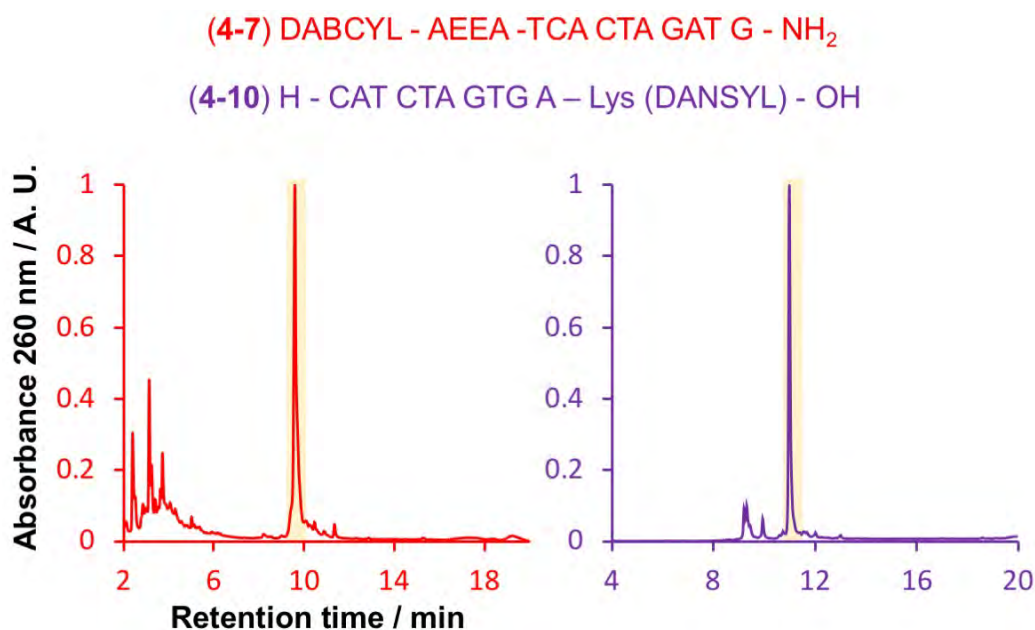


Figure 4-8. RP-HPLC analysis of the labelled PNA sequences **4-7** and **4-10**. The highlighted peaks were fraction-collected. AEEA = 2-[2-(2-aminoethoxy)ethoxy]amide linker. The products were analysed by LC-MS (Waters Xevo-G2-XS): **4-7** [M]⁰ *m/z* calcd. 3121.262 found 3121.297; **4-10** [M]⁰ *m/z* calcd. 3128.247 found 3128.140

The fluorescence of strand **4-10** and the UV absorbance of strand **4-7** (Figure 4-9) in DMF were determined individually. A good degree of spectral overlap between the emission of the fluorophore and the absorbance of the quencher was observed, confirming good FRET potential.

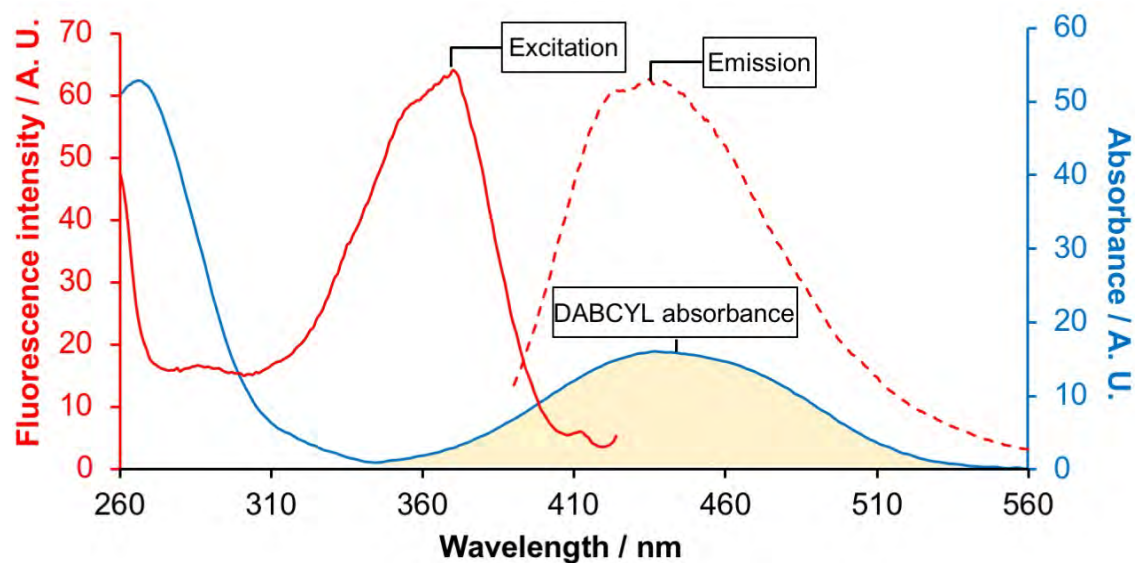


Figure 4-9. Characterisation of the spectroscopic properties of **4-7** and **4-10** fluorophore / quencher pair. Fluorescence excitation (red trace; $\lambda_{\text{emission}} = 434 \text{ nm}$) and emission (red dashed trace; $\lambda_{\text{excitation}} = 370 \text{ nm}$) spectra of **4-10** in DMF. UV absorbance of **4-7** in DMF (blue trace).

Despite the spectral overlap, a thermal denaturalisation of dsPNA duplex of **4-7** and **4-10** did not produce a clear melting transition in the fluorescence emission intensity. (Figure 4-10). Calorimetric measurements, later on, displayed in 50 vol% DMF a melting transition at 63 °C; as fluorescence was aimed to study the isothermal hybridisation in organic solvent mixtures, having determined by calorimetry that PNA would anneal high such conditions, it was important to obtain a reliable fluorescence and calorimetry correspondence. The possible reasons leading to this negative result were: 1) The spectroscopic properties of the donor or quencher were drastically affected by the temperature change, leading to a thermal drift; 2) The AEEA linker used reduced the spatial co-localisation of the fluorophore and the quencher, causing an inefficient quenching; 3) The presence of this fluorophore / quencher greatly destabilised the duplex producing a melting transition with a much lower T_m of 22 °C

(Figure 4-9, green shadowed area). This result indicated that dansyl / dabcyI were not a suitable fluorophore / quencher system to determine the T_m in high DMF solutions.

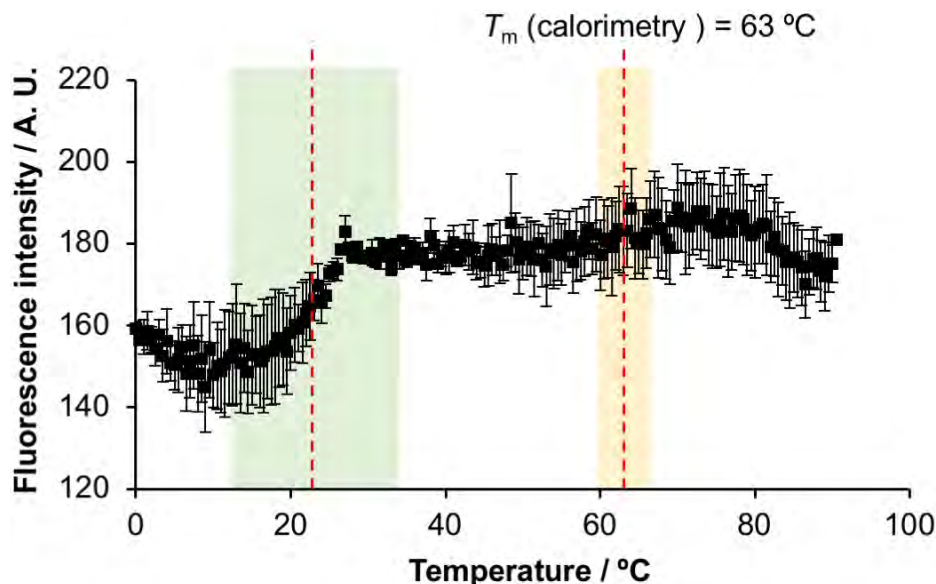


Figure 4-10. Thermal denaturalisation of dsPNA **4-7** and **4-10** in 50 vol% DMF observed by fluorescence. Transition observed (green shadowed area) $T_m = 22$ °C, expected $T_m = 63$ °C (orange shadowed area; unlabelled 10-mer dsPNA). $\lambda_{\text{excitation}} = 370$ nm; $\lambda_{\text{emission}} = 434$ nm. Two consecutive cycles of heating and cooling were performed. The error bars correspond to the standard deviation.

Post-cleavage modification is usually more challenging than on-resin modification of PNA strands (*i.e.* while the strand is still attached to the solid support), as in the on-resin approach a large excess of reagent can be applied and, once the reaction has been completed, removed by the addition of an excess of solvent. Beyond the unexpected lack of quenching activity in the previous dansyl / dabcyI system, a major limitation was that both conjugates were prepared by post-cleavage modification, and only small quantities of labelled-PNAs were produced. Replacing one of the probes with an on-resin compatible synthesis would greatly reduce the difficulty of the synthesis and also would allow to increase the scale. As dansyl chloride (**4-9**) could be

obtained in larger amounts than dabcyyl succinimidyl ester (**4-6**),^{§§§} it was decided to replace the latter and keep using the dansyl-labelled PNA.

Tryptophan (Trp) is a natural amino acid with intrinsic fluorescence,⁶⁷ and it is known to form a good fluorophore / quencher pair with dansyl.^{68,69} It can be attached to PNA strands during the solid phase synthesis stages, by simply adding an extra coupling step. For the previous reasons, it was decided to use to a Trp-labelled fluorophore and a dansyl-labelled quencher.

The Trp-labelled PNA strand was prepared by SPS (Figure 4-11 a), analysed by HPLC and LC-MS (Figure 4-11 b), and a semi-preparative scale chromatographic isolation was performed using fast protein liquid chromatography (FPLC) instrumentation.

^{§§§} Dansyl chloride: Sigma-Aldrich catalogue, £74.60 · g⁻¹. Dabcyyl-SE: Sigma-Aldrich catalogue, £157.00 · 100 mg⁻¹ (December 2019).

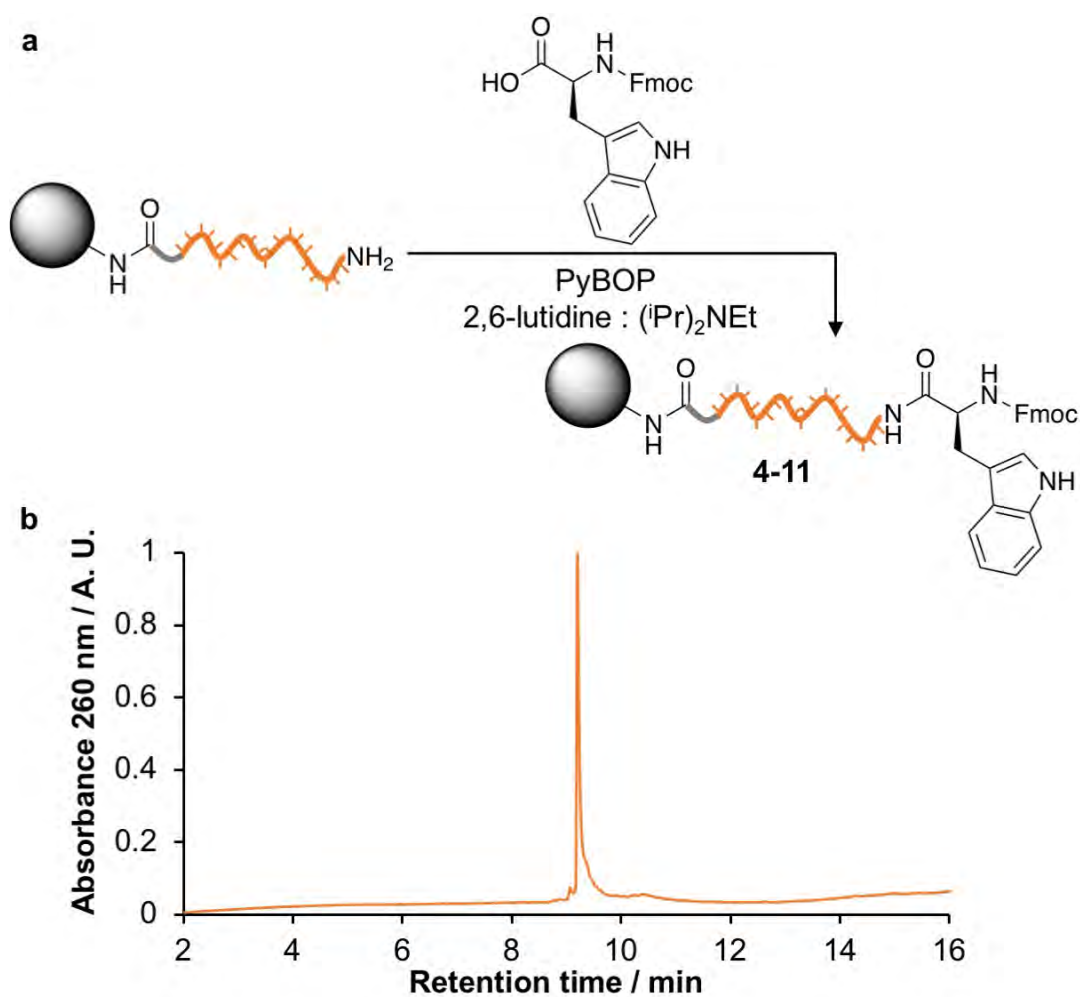


Figure 4-11. Synthesis of Trp-labeled PNA (**4-11**). a) Coupling of Fmoc-Trp-OH monomer to a PNA strand (only coupling SPS shown). b) RP-HPLC analysis of **4-11**. The product was identified by LC-MS (Waters Xevo-G2-XS): **4-11** [M]⁰ *m/z* calcd. 2911.169 found 2911.189

The isothermal hybridisation in H₂O of **4-10** and **4-11** was performed, and the change in fluorescence was monitored (Figure 4-12). A solution of **4-11** was placed in a fluorometer and its fluorescence was determined for 10 min. Then, 1 equiv. of the complementary quencher **4-10** was added, the cuvette was shaken, and the fluorescence of the resulting solution was measured for a further 15 min. The results of the unquenched and the quenched solutions were averaged to determine the extent of quenching. This gave a reduction in fluorescence intensity of 31%. It is important to mention that this experiment was initially performed in H₂O, as dsPNA formation was

well-documented in H₂O; however, this fluorophore / quencher system is expected to display higher fluorescence and quenching efficiency in high DMF solutions.

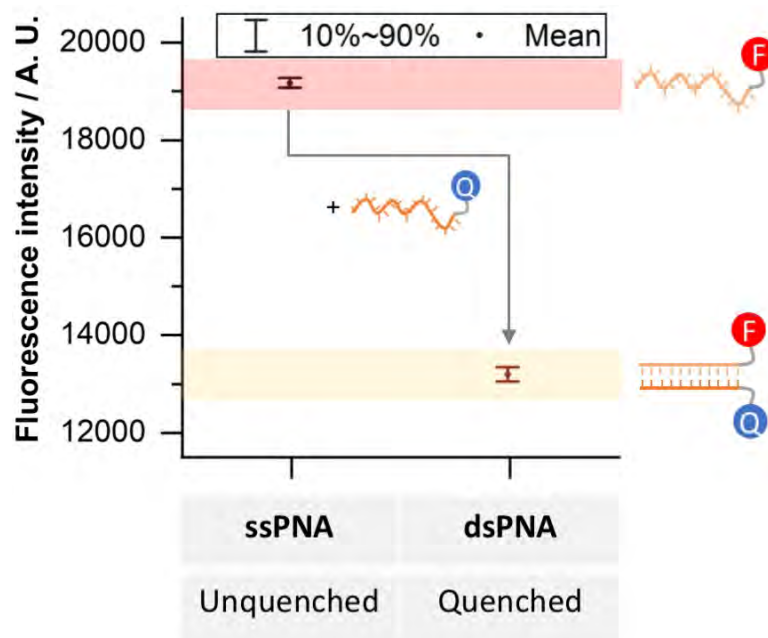


Figure 4-12. Tryptophan / dansyl labelled PNA quenching.

The superimposition of the fluorescence emission spectrum of PNA-Trp (4-11) and PNA-dansyl (4-10) in DMF also shows a good degree of overlap between the emission of tryptophan and the excitation of dansyl; thus highlighting the potential of these strands to form a good FRET donor / quencher pair (Experimental Section 4.4.10, Figure 4-27).

In summary, using a SPS approach, two sets of complementary 10-mer and 15-mer PNA strands were prepared in a moderate scale (10 μ mol). In addition, the 10-mer PNA strands were labelled with either a Trp fluorophore or a dansyl quencher. These fluorophore / quencher labelled strands were used in an isothermal annealing experiment, which showed a decrease in fluorescence upon annealing. This was an indication that the Trp / dansyl system could potentially be used to determine the thermal stability of dsPNA by fluorescence quenching.

4.2.3 Correlation of the thermal stability of dsDNA, PNA·DNA and dsPNA duplexes determined by UV spectroscopy, fluorescence spectroscopy and microcalorimetry in aqueous solution.

UV spectrophotometry has been widely used to determine the thermal stability of double-stranded nucleic acids. In order to establish a comparison between the thermal stability determined by UV spectrophotometry, fluorescence and micro-DSC, a comparative study was performed. The melting temperature of dsDNA, PNA·DNA and dsPNA 10-mer duplexes were determined in aqueous solution using the three methodologies and the results were compared to previously reported T_m values (Figure 4-13).¹⁵

Overall, a good agreement was found across the different techniques: the thermal stability showed the trend T_m dsPNA \approx PNA·DNA \gg dsDNA. However, a maximum discrepancy of 7 °C was found in the T_m values across different techniques for dsDNA, 8 °C for PNA·DNA and 5 °C for dsPNA. Fluorescence measurements are inherently different from micro-DSC and UV spectrophotometry measurements, as it requires fluorophore / quencher labelled strands and these have been shown to produce a change in the T_m results in the past.⁷⁰ In addition, as the fluorophore / quencher system designed for dsPNA was optimised for DMF solution, the fluorescence melting profile of PNA·DNA showed an unstable trace, and a large thermal drift; while no transition was observed for dsPNA. Therefore, fluorescence analysis was only applied to high DMF solution at a later stage. Finally, the lower sensitivity of micro-DSC required higher concentration of PNA (40 μ M compared to 2.5 μ M in UV spectrophotometry and fluorescence), which may have an effect as the T_m of nucleic acids is dependent on the overall concentration.⁷¹

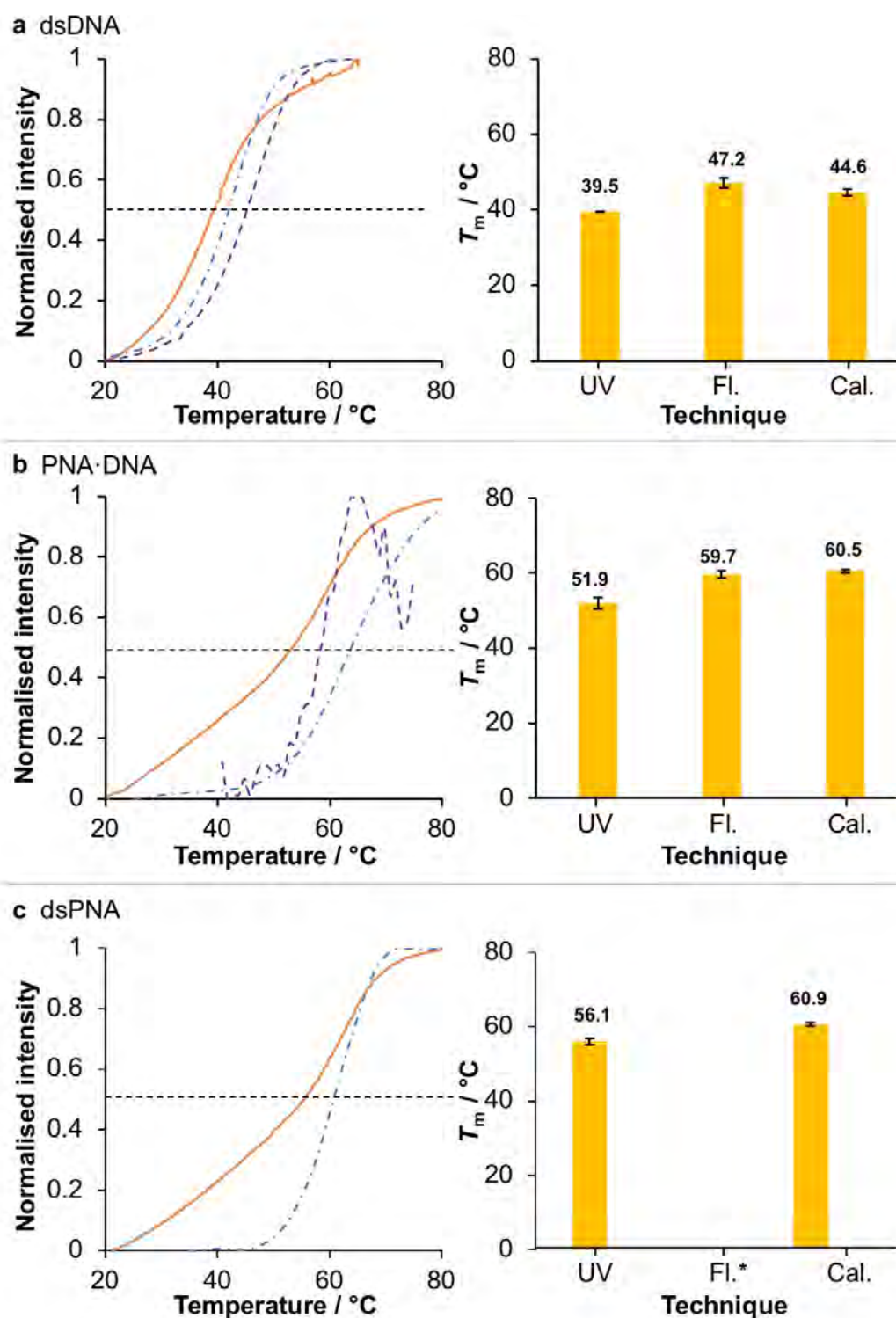


Figure 4-13. Comparison of three techniques for the measurement of the T_m of nucleic acid duplexes in aqueous solution: UV absorbance (UV, orange solid line), fluorescence (FI., violet dashed line) and calorimetry (Cal., blue dashed-dotted line). The black dashed line indicates the 50% maximum intensity from which the T_m s were calculated. a) dsDNA. b) DNA / PNA. c) dsPNA; *fluorescence did not produce a transition (Experimental Section 4.4.7).

Despite the considerable differences on the measured T_m across techniques, the expected stability trends were observed, showing transitions from single-stranded to double stranded nucleic acids. It was therefore concluded that the results displayed meaningful results correlating to the thermal stability of nucleic acid duplexes.

4.2.4. Micro-differential scanning calorimetry (micro-DSC) study of the thermal stability of dsPNA in aqueous / organic solvent mixture.

The primary focus of this work was to extend the previously reported thermal stability studies of dsPNA in H₂O / DMF to a larger concentration range (i.e. higher concentrations of DMF). However, in order to scope other compatible organic solvents, a solubility study of 15-mer PNA (strands **4-3** and **4-4**) was performed in a range of organic solvents such as apolar solvents, alcohol and amide solvents among others (Figure 4-14).

To achieve this, equal volumes of a stock solution of dsPNA in H₂O were placed in glass containers and dried under vacuum. Then, the organic solvent of interest was added to yield a solution up to 200 μ M PNA. The solutions were shaken, sonicated for 15 min and allowed to settle. A fraction of the solvent was transferred to a clean glass vial, dried under vacuum and re-dissolved in H₂O. The final concentration of PNA was determined by UV spectrophotometry, based on the calculated molar extinction coefficient.²⁵

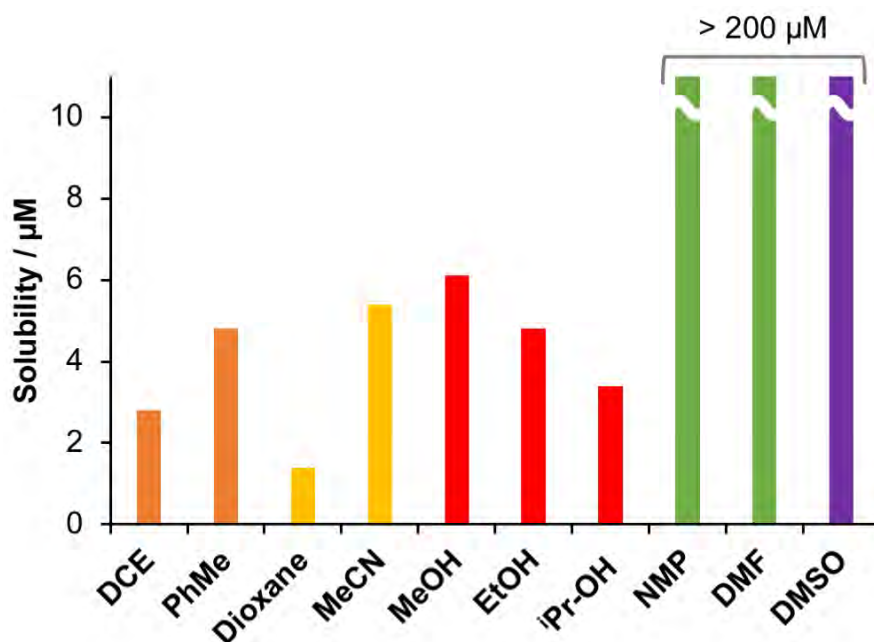


Figure 4-14. Solubility study of 15-mer PNA in several organic solvents. DCE = 1,2-dichloroethane; NMP = *N*-methyl-2-pyrrolidone; DMF = *N,N*-dimethylformamide; DMSO = dimethyl sulfoxide (Experimental Section 4.4.8).

The previous results displayed that the dsPNA had limited solubility (2 to 6 μM) in most organic solvents tested, with the exception of NMP, DMF and DMSO, where the solubility exceeded the 200 μM limit. The solvents that provide high solubility are non-nucleophilic, moderately polar and have a high boiling point. The subsequent experiments were performed in DMF, in order to compare the results with previous work; however, the combination of properties of NMP and DMSO makes them potentially good solvents to study the thermal stability of dsPNA and to attempt templated chemical reactions in the future.

The thermal stability of short (10-mer, **4-1-4-2**) and long (15-mer, **4-3-4-4**) dsPNA in H_2O / DMF solutions was determined using micro-DSC. Micro-DSC scans were performed between 0 and 90 $^\circ\text{C}$ in 0 to 100 DMF vol% solvent mixtures (Figure 4-15).

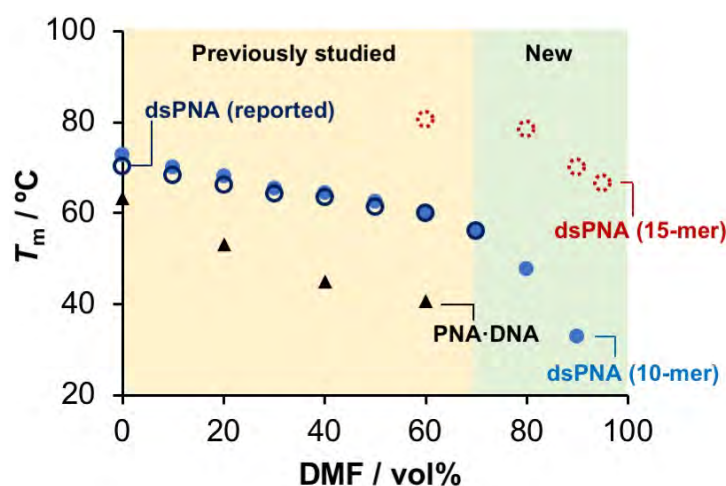


Figure 4-15. Melting temperatures, determined using micro-DSC, of long dsPNA (red dotted circles, 250 μ M), short dsPNA (blue solid dots, 500 μ M) and hybrid PNA·DNA 10-mer duplexes (black solid triangles, 500 μ M) in several H₂O / DMF solutions. The reported melting temperatures of short dsPNA, determined by UV spectrophotometry are shown (blue open circles).¹⁵ Each sample was subject to two heating and cooling cycles, and the T_m was determined as the average temperature at the maximum heat capacity; further data and the standard deviation are shown below (Experimental Section 4.4.9). The current micro-DSC results showed a good agreement with the reported T_m values by Sen and Nielsen, using UV spectrophotometry,¹⁵ with only small differences in the T_m found at low DMF solutions (0 to 30 vol%), which could be attributed to the fact that the reported thermal stability studies were performed using 5 μ M solutions. As it was previously reported by Sen and Nielsen, the T_m decreased linearly within 0 to 70 vol% DMF; however, a marked deviation from linearity took place beyond 70 vol% DMF, and no transition was detected in 100% DMF. The long dsPNA showed higher T_m values and a sharp decrease in the thermal stability of the duplex at high DMF vol%. Notably, stable long dsPNA was detected in 95 vol% DMF. The current experimental results highlight that a minimal fraction of H₂O is required to stabilise the PNA duplex, and the fraction of H₂O required decreases with the increase on the PNA strand length. In addition, the effect of DMF was more pronounced on hybrid PNA·DNA duplexes. This

highlights the higher compatibility of dsPNA with organic solvent, compared to alternative nucleobase materials.

The dramatic change in the T_m of dsPNA in the high DMF vol% range (70 to 100) raised the question of why the thermal stability of dsPNA was preserved over a large DMF vol% range (0 to 70) and then, it changed so drastically. While H₂O and DMF have very similar densities, their molecular weights are very different.^{****} This produces a non-linear relationship between the vol% and the mol% of DMF (Eq. 1 and Figure 4-16).

$$f_{n,DMF} = \frac{\rho_{DMF} \cdot MW_{H_2O} \cdot f_{v,DMF}}{\rho_{DMF} \cdot MW_{H_2O} \cdot f_{v,DMF} + MW_{DMF} \cdot \rho_{H_2O} \cdot (1 - f_{v,DMF})} \quad \text{Eq. 1}$$

$f_{n,DMF}$ = molar fraction of DMF

$f_{v,DMF}$ = volumetric fraction of DMF

ρ_{DMF} = density of DMF ; ρ_{H_2O} = density of H₂O

MW_{DMF} = molecular weight of DMF; MW_{H_2O} = molecular weight of H₂O

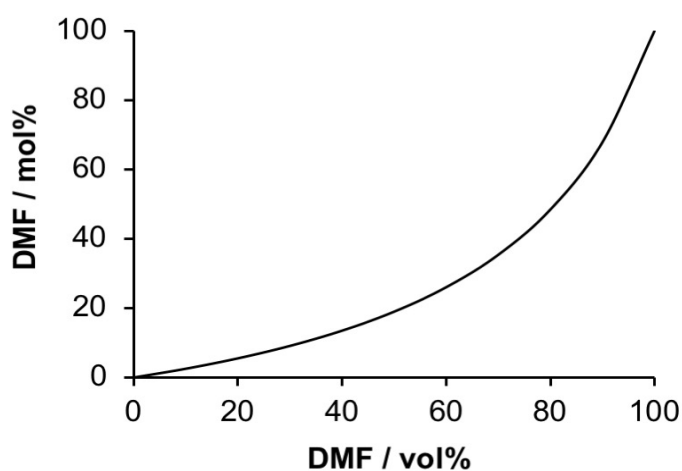


Figure 4-16. Conversion of volumetric fraction into molar fraction using eq. 1.

^{****} $\rho_{DMF}(25\text{ }^\circ\text{C}) = 0.944\text{ g} \cdot \text{mL}^{-1}$; $\rho_{H_2O}(25\text{ }^\circ\text{C}) = 0.997\text{ g} \cdot \text{mL}^{-1}$; $MW_{DMF} = 73.09\text{ g} \cdot \text{mol}^{-1}$ $MW_{H_2O} = 18.02\text{ g} \cdot \text{mol}^{-1}$

The plot of DMF (mol%) as function of DMF (vol%) shows that in the low vol% region, there is a nearly linear change of the mol%; however, the slope of the function changes drastically in the high vol% range. At the high DMF vol%, a small volumetric change implies a large molar change.

If the T_m of dsPNA is represented as function of the DMF mol%, a quasi-linear trend appears (Figure 4-17).

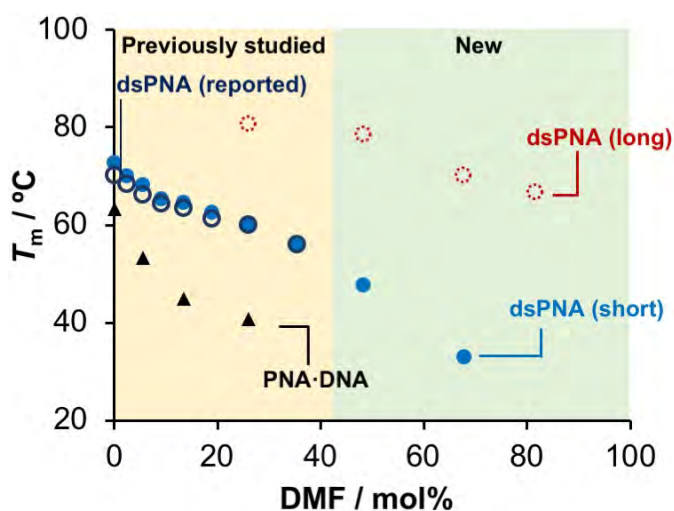


Figure 4-17. Evolution of the T_m of long dsPNA (red dotted circles, 250 μM), short dsPNA (blue solid dots, 500 μM) and hybrid PNA·DNA 10-mer duplexes (black solid triangles, 500 μM) in several H_2O / DMF solutions as function of the molar content of DMF in DMF / H_2O mixtures.

It can be concluded that the effect of the addition of DMF to aqueous solutions of dsPNA on its thermal stability is linearly dependent on the number of molecules of organic solvent, rather than the volume that these molecules occupy. This is quite intuitive, as the dsPNA·solvent interactions are based on a number of functional groups that can interact with a number of solvent molecules, rather than with a volume of solvent.

These results show that even using long dsPNA, a small fraction of H_2O will be required to provide stable duplexes. While this poses a limitation on the range of compatible

water-sensitive chemistries, it has been previously reported that the yield of a water-sensitive DNA templated pyrrolidine catalysed aldol reaction, greatly improved through reduction of the H₂O content to 5 vol%,⁷ thus highlighting the potential that PNA templates in 95 vol% DMF solution have to improve the templated synthesis of peptide bonds.

4.2.5. Isothermal annealing of dsPNA in high organic solution

The formation of DNA and RNA duplexes can be achieved through thermal annealing (*i.e.* by applying a heating cycle, followed by a slow cooling that allows the transition to double stranded nucleic acids). However, the isothermal mixing of short, complementary, nucleic acids strands also leads to duplex formation.⁷²

Isothermal annealing is a key challenge for DNA templated synthesis of peptide bonds, as a heating cycle could lead to the decomposition of highly reactive activated ester-labelled DNA adapters. This isothermal annealing approach has been successfully demonstrated in multistep DNA templated synthesis,^{73–75} and autonomous DNA hybridisation mechanisms.^{76–79}

Here, it was of utmost importance to determine whether complementary PNA strands, analogously to DNA strands, would isothermally hybridise in a high percentage organic solution. Together with toehold-mediated strand displacement; this would allow the design of multistep PNA templated synthesis of peptide bonds.

While micro-DSC was a suitable technique to study the thermal stability of dsPNA in high DMF solutions, fluorescence spectroscopy was the best approach to study the isothermal annealing of labelled PNA strands, as it allows monitoring the formation of PNA duplex at a constant temperature. To achieve this, to a solution of the fluorophore-labelled PNA strand (**4-11**) in 90 vol% DMF, the complementary, quencher-labelled

PNA strand (4-10) was added. As a consequence of the co-localisation of the fluorophore and quencher in close proximity, upon hybridisation, the fluorescence emission decreased over time (Figure 4-18).

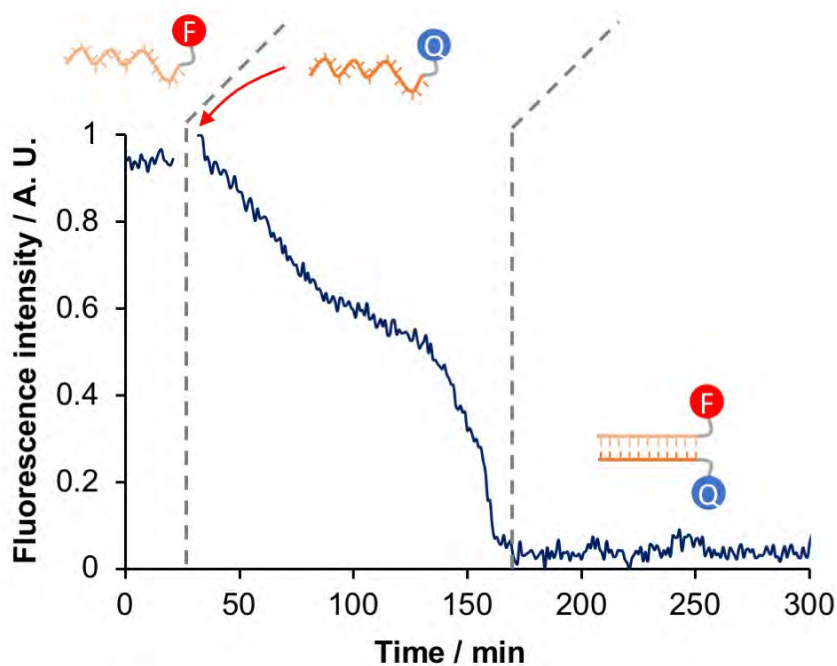


Figure 4-18. Isothermal annealing of individual fluorophore and quencher-labelled PNAs, in the absence of a pre-heating step applied to individual PNA solutions. Initially, the fluorescence of the fluorophore-labelled PNA (4-11), was monitored. Then, the complementary quencher-labelled PNA (4-10) was added (red arrow). Finally, the fluorescence of the duplex was measured during ~2 h. Solvent 90 vol% DMF (Experimental Section 4.4.10).

Typical PNA·DNA annealing processes are extremely fast, with second order kinetic constants on the order of $10^4 \text{ M}^{-1}\cdot\text{s}^{-1}$,⁸⁰ this lead to complete annealing, under typical concentration conditions (e.g. 200 nM), in under 10 min. By contrast, the current PNA annealing was complete after ~2 h (Figure 4-17). As the stock solutions of individual fluorophore-labelled (4-11) and quencher-labelled (4-10) PNAs were prepared by isothermal dissolution of solid pellets in 90 vol% DMF, the remarkably slow hybridisation kinetics detected here, were attributed to the presence of kinetically persistent interactions between PNA strands that originated in the solid state.

To overcome this slow kinetics, the isothermal annealing experiment was repeated, preceded by a heating and cooling cycle on the individual PNA stock solutions, with the aim of dissociating any kinetically trapped PNA structures. Then, the fluorescence of **4-11** fluorophore-labelled PNA was measured for ~8 min, followed by the addition of an equimolar amount of the complementary quencher-labelled PNA (**4-10**) in a small volume of 90 vol% DMF. The emission of the resulting solution was followed for another 8 min (Figure 4-19 a). The data showed how, by subjecting the stock PNA solutions to a heating and cooling cycle, the isothermal annealing, analysed by fluorescence spectroscopy, took place rapidly (*i.e.* faster than the time required to reinstall the cuvette in the spectrofluorometer holder). This data highlights that, in order to perform future isothermal annealing of PNA in the absence of heating, long annealing times are required as a result of trapped PNA assemblies.

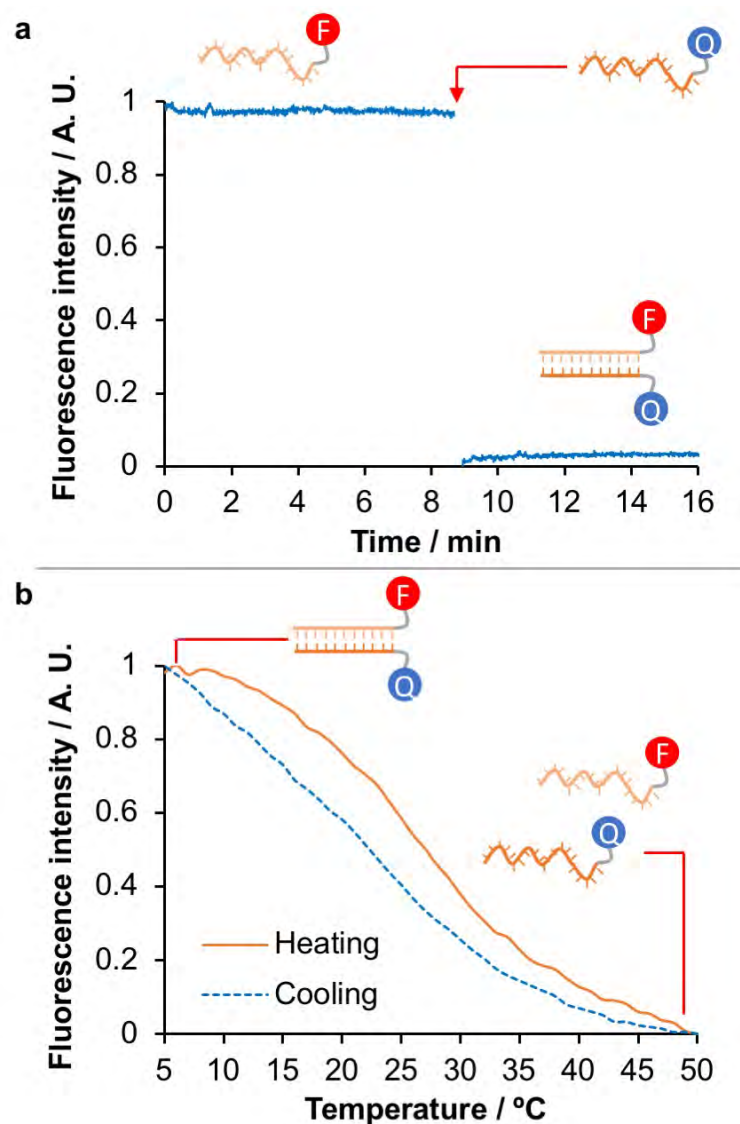


Figure 4-19. a) Isothermal annealing of fluorophore and quencher-labelled PNA in 90 vol% DMF, preceded by a heating cycle on individual stock solutions. b) Thermal denaturalisation and re-annealing of dsPNA studied by FRET ($\lambda_{\text{excitation}} = 340 \text{ nm}$, $\lambda_{\text{emission}} = 550 \text{ nm}$, Experimental Section 4.4.10).

In order to corroborate the isothermal formation of dsPNA, the solution was subject to a heating and cooling ramp, while monitoring the fluorescence intensity. In this case, a FRET experiment irradiating the fluorophore ($\lambda_{\text{excitation}} = 340$, Trp) and measuring the emission of the acceptor ($\lambda_{\text{emission}} = 550 \text{ nm}$, dansyl) displayed a high sensitivity to the annealed / denatured state over the thermal drift, and showed a clear melting transition

(Figure 4-19 b). The PNA duplex in 90 vol% DMF reversibly transitioned from a denatured to a re-annealed state.

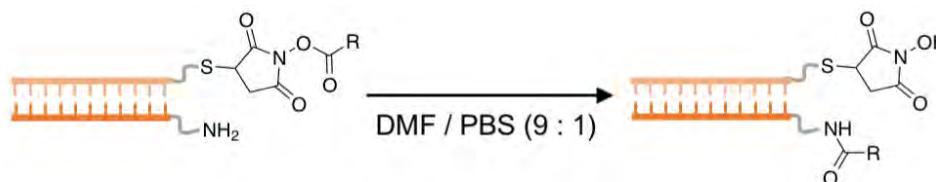
With these results, it was clearly demonstrated that isothermal hybridisation of PNAs took place in high vol% DMF solution, both in the presence and in the absence of a heating cycle on the individual PNA strands. In the absence of a heating cycle, the kinetics of isothermal hybridisation were considerably slower attributed to the presence of kinetically trapped PNA structures. Finally, it was demonstrated that PNA hybridisation in such conditions was a reversible process which is essential to perform subsequent hybridisation cascade processes.

3.2.6 Future directions

So far, PNA strands have mainly been utilised for templated synthesis in combination with DNA and RNA templates, with very limited precedents in the use of pure PNA duplexes.²² The ultimate goal of the present work was to utilise PNA strands as templates to encode, program and template the synthesis of peptide bonds in organic solvent, in an analogous manner to DNA templated synthesis of peptide bonds in aqueous buffered solution. The current results show that the hybridisation of PNA strands in up to 95 vol% DMF would be compatible with the requirements for PNA templated synthesis; however, there are two more requirements to apply this concept: successful single-step synthesis of peptide bond and the demonstration of toehold-mediated PNA displacement.

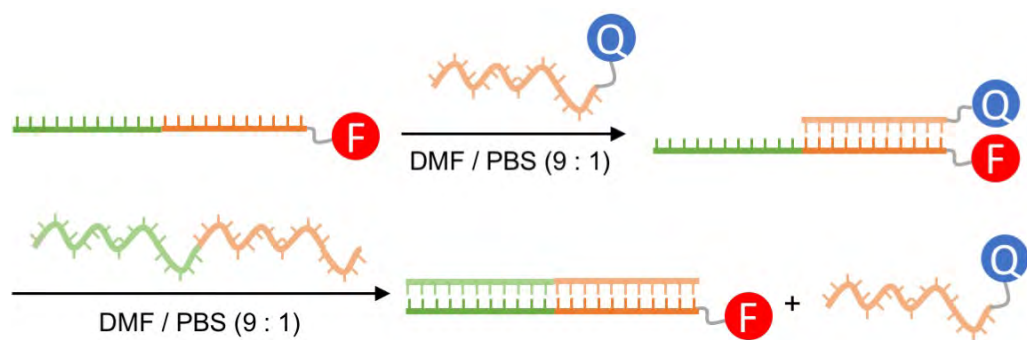
In order to demonstrate single-step PNA templated synthesis of peptide bonds, an activated-ester labelled, and a complementary amine-labelled PNA strands could be isothermally annealed over a long period (~2 h) in 90 vol% DMF (Scheme 4-7). Several pH buffers could be assayed to complete the 10 vol%, in order to maximise the yield

while reducing hydrolysis. The outcome of these reactions could be studied by denaturing LC-MS, for example using HFIP / Et₃N additives, or high temperature, to determine the ratio of amine-labelled PNA to amide-labelled PNA.



Scheme 4-7. Single step PNA templated synthesis in high organic solvent solutions.

Toehold mediated PNA strand displacement could be demonstrated by displacing a partially complementary fluorophore-labelled and quencher-labelled duplex using a fully complementary unlabelled PNA strand (Scheme 4-8). The expected evolution of fluorescence would be: 1) High fluorescence of the fluorophore-labelled ssPNA; 2) Low fluorescence of the fluorophore and quencher-labelled dsPNA; 3) High fluorescence recovered after the displacement of the short quencher-labelled PNA with a long unlabelled PNA strand. In addition, it would be very interesting to compare the displacement efficiency with mismatched, long PNA strands, in order to evaluate the specificity of PNA-PNA interactions. If PNA-PNA interactions are strictly sequence-dependent (*i.e.* the duplex formation between fully complementary strands is favoured over unspecific, non-complementary PNA-PNA duplexes), this would facilitate designing complex templated syntheses with several strands coexisting in the same solution.



Scheme 4-8. Toehold mediated PNA strand displacement.

While this falls only in the domain of speculation, based on the fact that the formation of dsPNA in 90 vol% DMF was fully reversible, and proceeded isothermally, it would be expected that a strand displacement would take place in this system, thanks to the higher thermodynamic stability of the long dsPNA and, even if there are kinetic barriers that slow down the process, it would eventually lead to successful strand displacement.

4.3 Conclusions

It has been previously demonstrated that PNA duplexes have high tolerance to the addition of DMF and dioxane^{14,15} among other organic solvents.⁸¹ However, previous thermal stability studies were constrained to a maximum 70 vol% DMF due to the limitations of UV spectroscopy methods.

Here, the thermal stability of PNA duplexes has been studied in H₂O / DMF mixtures using microcalorimetry. This technique recorded reliable data in 0 to 100 vol% DMF solutions, showing that 10-mer dsPNA was stable in up to 90 vol% DMF with a T_m of 33.0 ± 1.1 °C. The elongation of the duplex to a 15-mer increased the T_m to 67.7 ± 0.3 °C in the same solvent, and allowed to detect PNA duplex in up to 95 vol% DMF. The requirement of a small volume of H₂O to allow the PNA duplex formation indicates the importance of hydrophobic interactions in the formation of such duplexes. Despite the requirement for a small percentage of water, it has been previously demonstrated that the yield of a DNA templated water-sensitive reaction dramatically increased when the water content was reduced to 5 vol%,⁷ highlighting the potential of the current results to improve PNA templated reaction yields.

By contrast to previous assumptions,¹⁵ the T_m of dsPNA as function of the vol% of DMF deviates from linearity at high DMF content, displaying a drastic reduction of the thermal stability beyond 70 vol% DMF. The evolution of T_m is quasilinear, as function of the molar percentage of DMF (*i.e.* mol%), rather than the volumetric percentage.

Using a Trp / dansyl-labelled PNA duplex, the isothermal annealing of PNA strands was demonstrated using fluorescence spectroscopy. The addition of a heating and cooling cycle on individual PNA strands prior to isothermal annealing was required for rapid hybridisation kinetics. Following the previous experiment, the thermal

denaturalisation and re-annealing of these strands demonstrated the reversibility of the hybridisation transition.

Altogether, the current findings suggest that PNA strands are promising candidates to template peptide bond formation reactions, using highly reactive activated esters, in high DMF solution, thus facilitating high conversion through the stabilisation of the activated esters.

4.4 Experimental section

4.4.1 General methods

For RP-HPLC, LC-MS, and nanopure water, see Appendix 1.

UV absorbance measurements were performed on 2.5 μM solutions of dsDNA, PNA·DNA and dsPNA at 260 nm on an Evolution 360 UV-Vis (Thermoscientific) coupled to a PCCU1 Peltier control and cooling unit. The instrument was controlled by ThermoInsight2 software. The sample was heated and cooled in 1 $^{\circ}\text{C}$ steps, with 40 s stabilisation time, and with continuous stirring, while the temperature was determined using a compact transition joint thermocouple probe (Omega).

Fluorescence measurements were performed on labelled dsDNA, PNA·DNA and dsPNA strands as 2.5 μM solutions at the optimal excitation and emission wavelengths for each fluorophore / quencher pair, on a FS5 steady state spectrofluorometer (Edinburgh instruments) with a SC-25 temperature controlled holder coupled to an Integrated TC1 Temperature controller (Quantum Northwest). The instrument was controlled by the Fluoraclessoftware. The sample was heated and cooled in 0.5 $^{\circ}\text{C}$ steps with 40 s stabilisation time before measurement and a ± 0.1 $^{\circ}\text{C}$ tolerance.

DSC measurements were performed on a Nano Differential Scanning Calorimeter 602000 (TA instruments) at 3 atm pressure and 1 $^{\circ}\text{C}\cdot\text{min}^{-1}$ heating and cooling rate for diluted solutions (< 50 μM) or 0.5 $^{\circ}\text{C}\cdot\text{min}^{-1}$ for concentrated solutions. A volume of 600 μL of solution was introduced into the sample and reference capillary cells. The instrument was controlled by the DSCrun software. The data was analysed with the NanoAnalyze software, corrected with a sigmoidal baseline and a two-state model was fitted to determine the melting temperature.⁴³

4.4.2 Sequences and modifications

DNA sequences were purchased from Integrated DNA Technologies (IDT) with standard desalting and HPLC purification for the quencher labelled strands. A 5,6-carboxyfluorescein / Black Hole Quencher 1 donor-acceptor system was chosen for the fluorescence melting experiments of dsDNA.

Table 4-2. DNA oligonucleotide sequences and modifications.

Identification number	Sequence (PNA) and modifications (N → C)
4-1	H - TCA CTA GAT G - NH ₂
4-2	H - CAT CTA GTG A- NH ₂
4-3	H - CGC CGT CAC TAG ATG - NH ₂
4-4	H - CAT CTA GTG ACG GCG - NH ₂
4-7	DabcyI-AEEA- TCA CTA GAT G - NH ₂
4-10	Ac - CAT CTA GTG A - Lys(Dansyl) - NH ₂
4-11	H - Trp - TCA CTA GAT G - NH ₂
Identification number	Sequence (DNA) and modifications (3' → 5')
4-12	TCA CTA GAT G
4-13	CAT CTA GTG A
4-14	/56-FAM/ TCA CTA GAT G
4-15	CAT CTA GTG A /3BHQ_1/

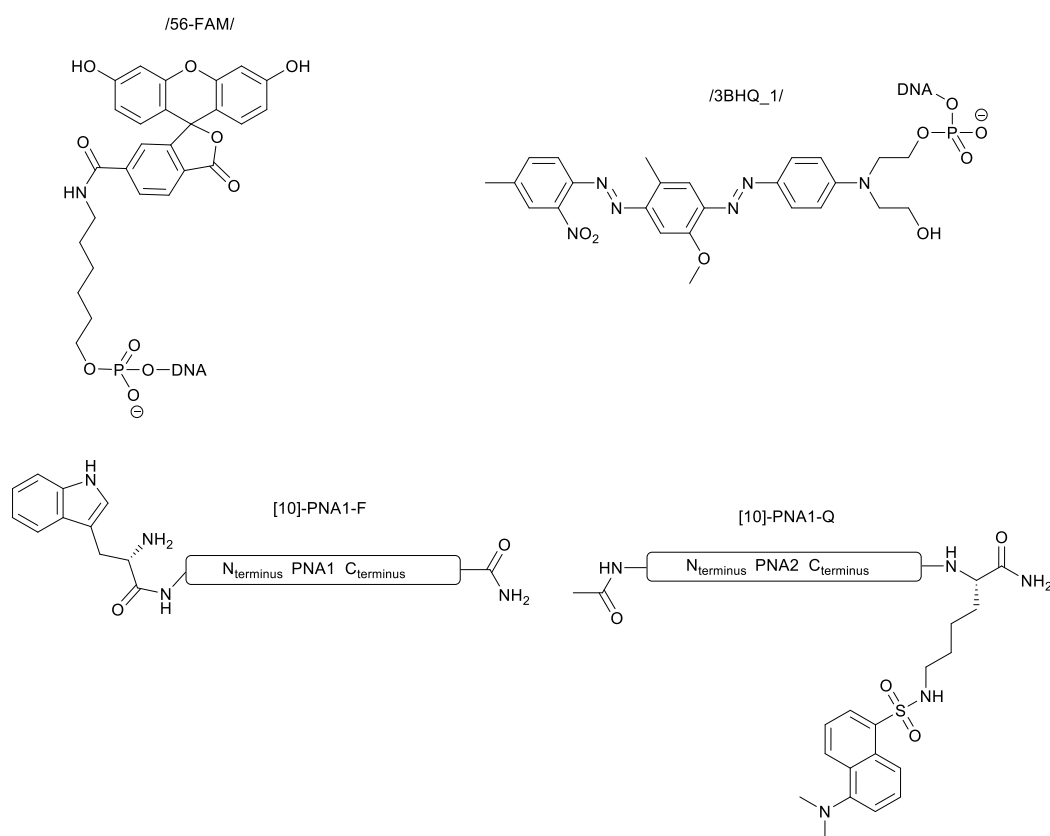


Figure 4-20. Chemical structure of DNA and PNA modifications.

4.4.3 Solid phase synthesis methods

PNAs were prepared using a microwave-assisted Liberty Blue (CEM) automated synthesiser at a 10 μmol scale with a 10 mL PTFE reaction vessel. Fmoc/Bhoc-protected monomers were purchased from Link technologies. A PAL-novaPEG (Novabiochem) 0.44 $\text{mmol}\cdot\text{g}^{-1}$ loading capacity or Rink amide ProTide(LL) (CEM) 0.19 $\text{mmol}\cdot\text{g}^{-1}$ loading capacity solid supports were interchangeably used. Peptide synthesis grade DMF was used as the main solvent (VWR). Oxyma was purchased from CEM corporation and Merck and the remaining reagents were directly purchased from Sigma-Aldrich. The synthetic cycles are summarised in table 4-3:

Table 4-3. Solid phase synthesis cycles.

Cycle	Description
1. Resin swelling	The resin was swollen with 3 mL of DMF for 5 min at 20 °C.
2. Fmoc deprotection	The resin was treated with 2 mL of 10 vol% piperidine in DMF for 10 min at 20 °C. Extensive DMF washes were applied after the Fmoc UV _{301 nm} quantification.
3. Coupling	0.5 mL of 0.1 M solution of Fmoc-PNA-OH monomer in DMF, 0.5 mL of 0.1 M Solution of PyBOP (or Oxyma) in DMF and 0.5 mL of 0.1 M <i>N,N</i> -diisopropylethylamine 0.1 M 2,6-lutidine (or DIC) in DMF were added to the reaction vessel. The coupling step was performed for 1 h at 50 °C followed by two DMF washes.
4. Cleavage	The resin was transferred to a manual solid phase synthesis device, washed with 4 mL of MeOH, MeOH, DCM and DCM. Then, it was air dried and briefly dried under vacuum. The dry resin was transferred to a vial with 1 mL of 95 vol% TFA and 5 vol% of <i>m</i> -cresol and the solution was shaken at 20 °C for 1.5 h. The resin was removed by filtration and the supernatant was collected. The cleaved product was precipitated by addition of 2 volumes of diethyl ether and centrifuged at 4 °C for 30 min. The white pellet was rinsed further with 1 mL of diethyl ether and centrifuged at 4 °C for 15 min.

4.4.4 Design of 15-mer PNA sequences

The 10-mer PNA sequences were designed similarly to previously reported relevant studies^{14,15} and extended into 15-mer PNAs using Nupack web application.³⁴

```
# CODE FOR NUPACK ELONGATION OF DSPNA
# Samuel Nunez, May 2017
# parameters for Nupack design
material = dna
temperature = 23.0
dangles = some
trials = 4
# target structure defined in dot-parens-plus notation
structure assemble =(((((((((((((((((((+))))))))))))))))))))
# define sequence domains
domain a = N5
domain b = TCACTAGATG
```

```
domain a* = N5
domain b* = CATCTAGTGA
# thread domains onto strands
strand PNA1 = a b
strand PNA2 = b* a*
# thread strands onto target structures
assemble.seq = PNA1 PNA2
# prevent sequence patterns
prevent = AAAA, CCCC, GGGG, TTTT, UUUU, CCGG, GGCC, CGCG, GCGC, KKKKKK, MMMMMM,
RRRRRR, SSSSSS, WWWWWW, YYYYYY
```

4.4.5 Synthesis of 4-7

DabcyI succinimide ester (3.7 mg, 10 μmol) was dissolved in 20 μL of DMF. 20 μL of **4-5** solution (1 mM in DMF, 20 nmol) was then added with 1.4 μL of Et_3N (10 μmol) and the solution was then shaken at 25 $^\circ\text{C}$ overnight. The excess of small molecule was precipitated with the addition of 20 μL of H_2O , and then the sample was centrifuged, and the supernatant was concentrated under vacuum. Finally, the resulting solid pellet was dissolved in 200 μL of H_2O and the product was analysed by LC-MS showing complete conversion (Figure 4-7). The product was HPLC fraction-collected.

4.4.6 Synthesis of 4-10

200 nmol of **4-8** was dissolved in 200 μL of DMF in a microcentrifuge tube. 11 mg of 5-(dimethylamino)naphthalene-1-sulfonyl chloride (dansyl chloride, 40 μmol) was then dissolved in 200 μL of DMF and added to the PNA solution. 7 μL of $(i\text{Pr})_2\text{NEt}$ (40 μmol) was then added and the resulting mixture was shaken in a thermal shaker overnight at 20 $^\circ\text{C}$. The solvent was removed by freezing the solution with liquid nitrogen and placing under high vacuum. The excess of reagent was removed by two consecutive precipitations of the product into 200 μL of DMF and 400 μL of diethyl ether. The product was pelleted by centrifugation at 4 $^\circ\text{C}$ and $21 \cdot 10^3$ Rcf. Finally, the product was

analysed by LCMS and was HPLC fraction-collected. The conversion based upon the chromatographic peak area was 66% (Figure 4-7).

4.4.7 Thermal stability in aqueous solution: comparative study

Table 4-4. Conditions, strands and T_m results of the thermal stability comparative study.

		dsDNA	DNA·PNA	dsPNA
UV Absorbance	Concentration	2.5 μ M	2.5 μ M	2.5 μ M
	Solvent	PBS (100 mM, pH 7.2)	18.2 M Ω ·cm water	18.2 M Ω ·cm water
	Sequences	4-12·4-13	4-1·4-13	4-1·4-2
	Heating rate	0.5 $^{\circ}$ C / step – 60 s	0.5 $^{\circ}$ C / step – 60 s	0.5 $^{\circ}$ C / step – 60 s
	Melting temperature	39.5 \pm 1.4 $^{\circ}$ C	51.9 \pm 3 $^{\circ}$ C	56.1 \pm 1.5 $^{\circ}$ C
Fluorescence	Concentration	2.5 μ M	2.5 μ M	
	Solvent	PBS (100 mM, pH 7.2)	18.2 M Ω ·cm water	
	Sequences	4-14·4-15	4-11·4-15	Not determined
	Heating rate	0.5 $^{\circ}$ C / step – 60 s	0.5 $^{\circ}$ C / step – 60 s	
	Melting temperature	47.2 \pm 2.6 $^{\circ}$ C	59.7 \pm 2.0 $^{\circ}$ C	
Calorimetry	Concentration	20 μ M	40 μ M	40 μ M
	Solvent	PBS (100 mM, pH 7.2)	18.2 M Ω ·cm water	18.2 M Ω ·cm water
	Sequences	4-12·4-13	4-1·4-13	4-1·4-2
	Heating rate	1 $^{\circ}$ C / min	1 $^{\circ}$ C / min	1 $^{\circ}$ C / min
	Melting temperature	44.6 \pm 1.8 $^{\circ}$ C	60.5 \pm 0.7 $^{\circ}$ C	60.9 \pm 1.0 $^{\circ}$ C

4.4.8 Solubility of 15-mer PNA in organic solvents

77 μ L of 15-mer PNA solution (H₂O, 1.3 mM, 100 nmol) were placed in glass vials and lyophilised. Then, 500 μ L of organic solvent (Figure 4-14) was added to the vials, and the mixtures were shaken and sonicated. The solutions were allowed to settle for 30 min, and then, 100 μ L of each solution (carefully taken from the top of the vials) was transferred into clean glass vials and concentrated to dryness under vacuum. Finally, the solid was dissolved in 50 μ L of H₂O and the concentration was determined by UV spectroscopy. (260 nm).

4.4.9 Micro-DSC raw data

In order to prevent instability of the micro-DSC baseline, the DMF / H₂O solutions were degassed by sonication for 10 minutes. The 10-mer dsPNA and PNA·DNA duplexes were prepared at 500 μM while the 15-mer dsPNA solutions were prepared at 200 μM. Micro-DSC was performed across a temperature range of 0 to 90 °C, at a heat flux of 0.5 °C·min⁻¹. Each sample was subjected to two heating-cooling cycles. The T_m was determined as the average temperature at the maximum molar excess heat capacity of the melting transition, for the first cooling and the next heating and cooling ramps. The molar excess heat capacity was numerically integrated and normalised showing the progress of the melting process.

Table 4-5. 10-mer dsPNA melting temperature results obtained from micro-DSC. The standard deviation was calculated from three consecutive runs,

DMF (vol%)	T_m (°C)	SD (°C)
0	72.8	0.2
10	70.0	0.1
20	68.2	0.1
30	65.3	0.2
40	64.5	0.1
50	62.5	0.4
60	59.9	0.4
70	55.7	0.3
80	47.7	0.5
90	32.9	0.5

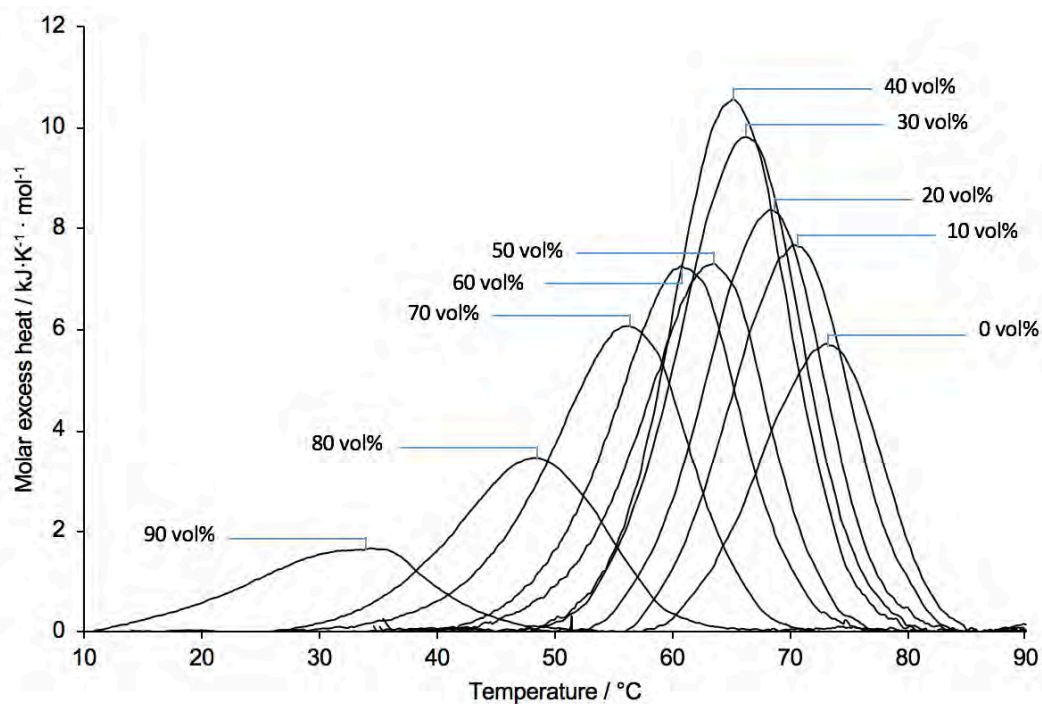


Figure 4-21. Sigmoidal baseline subtracted micro-DSC data of 10-mer dsPNA thermal melting in DMF / H₂O solutions.

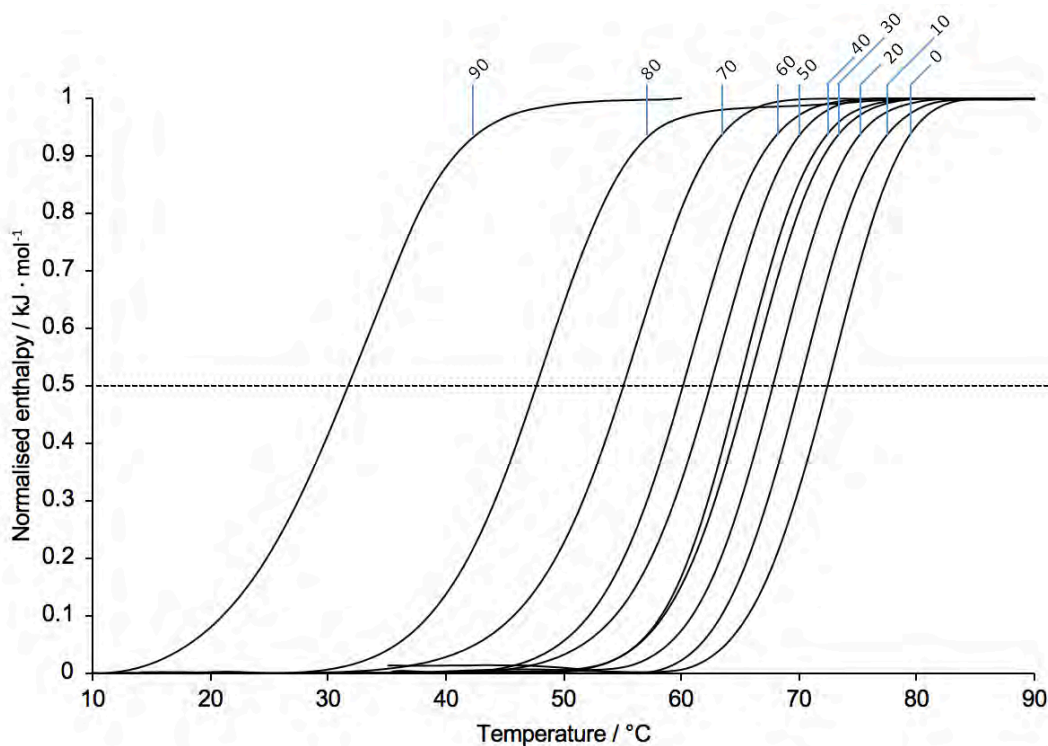


Figure 4-22. Integrated molar heat capacity as function of the temperature. DMF vol% indicated.

Table 4-6. 15-mer dsPNA melting temperature results obtained from micro-DSC. The standard deviation was calculated from three consecutive runs.

DMF (vol%)	T_m (°C)	SD (°C)
95	66.7	0.6
90	70.1	1.0
80	78.5	1.4
60	80.6	0.6

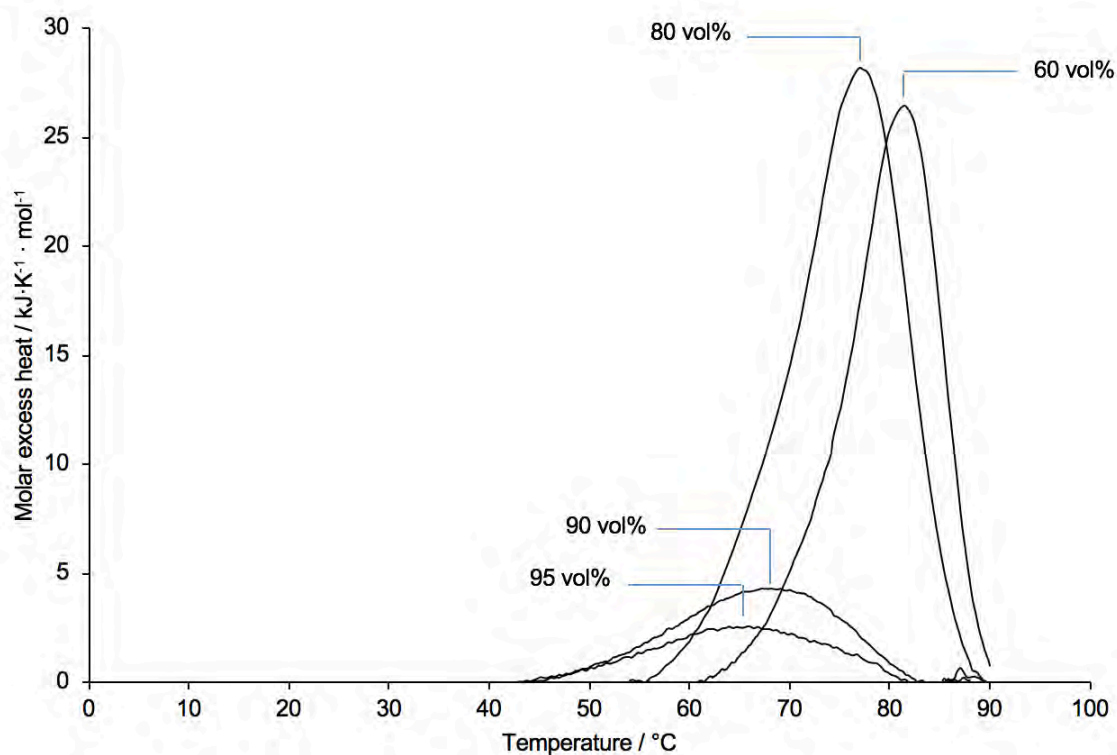


Figure 4-23. Sigmoidal baseline subtracted micro-DSC data of 15-mer dsPNA thermal melting in DMF / H₂O solutions.

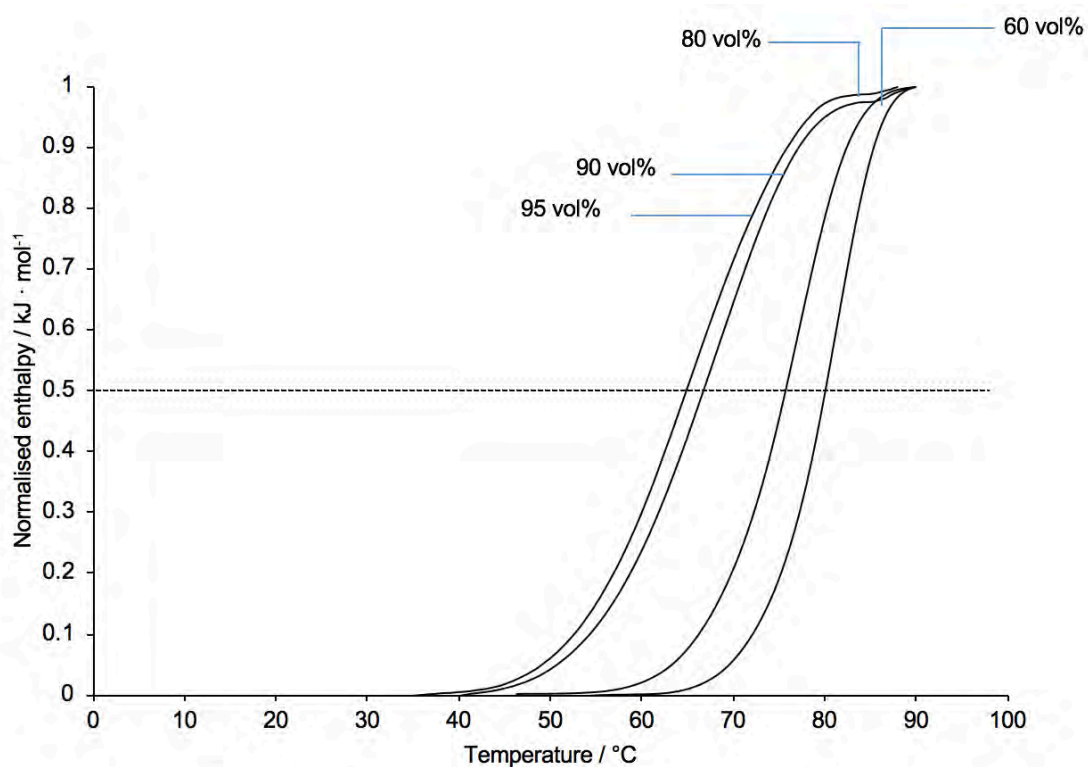


Figure 4-24. Integrated molar heat capacity as function of the temperature. DMF vol% indicated.

Table 4-7. 10-mer PNA-DNA melting temperature results obtained from micro-DSC. The standard deviation was calculated from three consecutive runs.

DMF (vol%)	T_m (°C)	SD (°C)
0	63.3	0.2
20	53.2	0.4
40	45.0	1.0
60	40.7	1.1

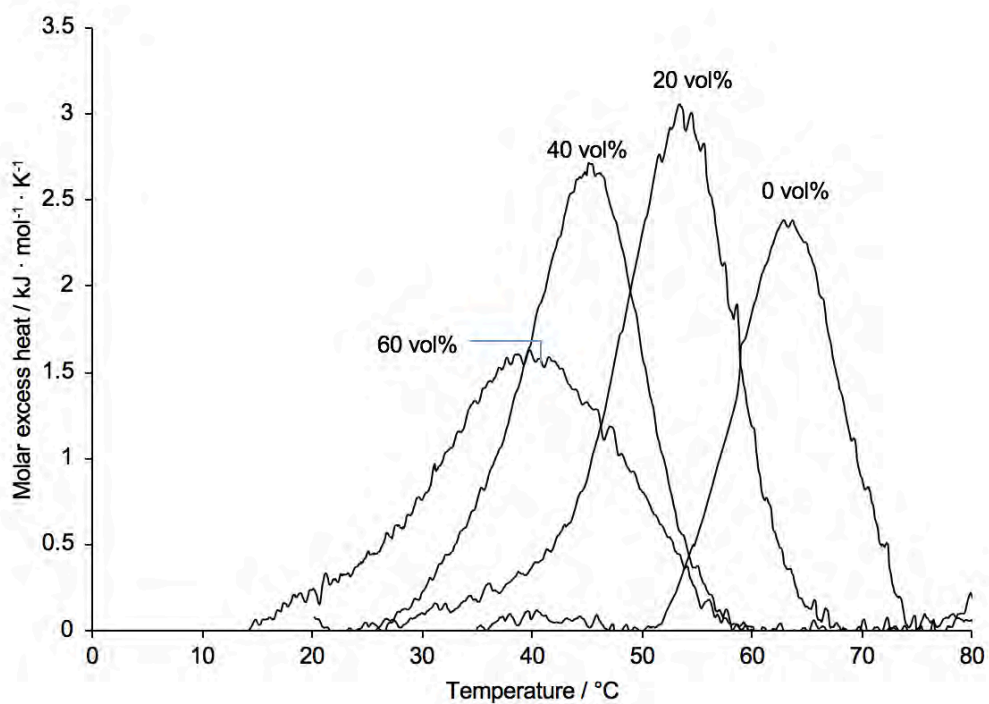


Figure 4-25. Sigmoidal baseline subtracted micro-DSC data of 10-mer PNA·DNA thermal melting in DMF / H₂O solutions.

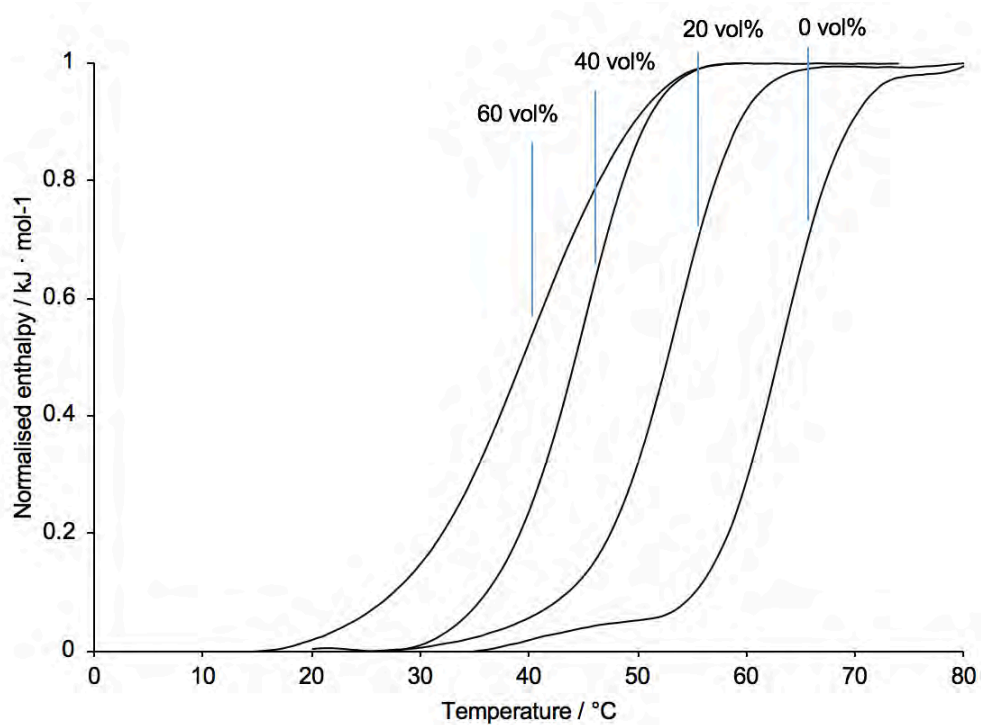


Figure 4-26. Integrated molar heat capacity as function of the temperature. DMF vol% indicated.

4.4.10 Isothermal annealing

After dissolving **4-11** as a solid pellet in 90 vol% DMF, it was found that a heating and cooling cycle, between 80 and 10 °C, was required before the addition of **4-10** for faster annealing kinetics (for comparison, see Figure 4-17 and 4-18). This was performed in a thermal cycler with a heated lid to prevent evaporation. Then, the solution was placed in a fluorescence quartz cuvette and monitored until a stable baseline was obtained, corresponding to the unquenched fluorescence emission. Finally, the cuvette was removed from the spectrofluorometer, the complementary quencher strand was added and the mixture was briefly shaken before returning the cuvette to the cell holder to monitor the annealing process. During the analysis, the temperature of the solution was adjusted to 10 °C with a pre-equilibrated circulating water bath. The final concentration of dsPNA was 10 μ M.

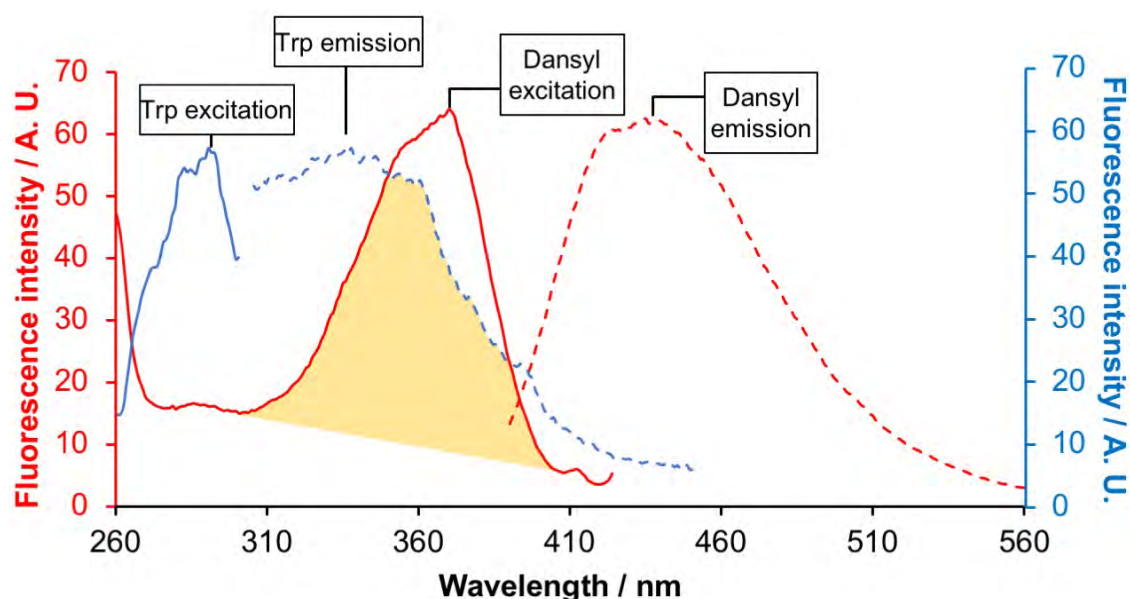


Figure 4-27. Superimposition of the fluorescence excitation and emission spectra of PNA-dansyl (**4-10**) and PNA-Trp (**4-11**) recorded in DMF. Dansyl fluorescence (red trace; $\lambda_{\text{emission}} = 434$ nm, $\lambda_{\text{excitation}} = 370$ nm). Tryptophan fluorescence (blue trace, $\lambda_{\text{emission}} = 336$ nm, $\lambda_{\text{excitation}} = 290$ nm)

4.5 References

- 1 S. I. Nakano, M. Fujimoto, H. Hara and N. Sugimoto, *Nucleic Acids Res.*, 1999, **27**, 2957–2965.
- 2 S. Benabou, A. Aviñó, R. Eritja, C. González and R. Gargallo, *RSC Adv.*, 2014, **4**, 26956–26980.
- 3 S. I. Nakano, H. Karimata, T. Ohmichi, J. Kawakami and N. Sugimoto, *J. Am. Chem. Soc.*, 2004, **126**, 14330–14331.
- 4 G. Bonner and A. M. Klibanov, *Biotechnol. Bioeng.*, 2000, **68**, 339–344.
- 5 N. Dave and J. Liu, *J. Phys. Chem. B*, 2010, **114**, 15694–15699.
- 6 M. L. McKee, A. C. Evans, S. R. Gerrard, R. K. O'Reilly, A. J. Turberfield and E. Stulz, *Org. Biomol. Chem.*, 2011, **9**, 1661–1666.
- 7 M. M. Rozenman and D. R. Liu, *ChemBioChem*, 2006, **7**, 253–256.
- 8 H. Abe, N. Abe, A. Shibata, K. Ito, Y. Tanaka, M. Ito, H. Saneyoshi, S. Shuto and Y. Ito, *Angew. Chemie - Int. Ed.*, 2012, **51**, 6475–6479.
- 9 P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science (80-.)*, 1991, **254**, 1497–1500.
- 10 H. Rasmussen, S. J. Kastrup, J. N. Nielsen, J. M. Nielsen and P. E. Nielsen, *Nat. Struct. Biol.*, 1997, **4**, 98–101.
- 11 M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature*, 1993, **365**, 566–568.
- 12 P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm and B. Nordén, *Nature*, 1994, **368**, 561–563.
- 13 S. Tomac, M. Sarkar, T. Ratilainen, P. Wittung, P. E. Nielsen, B. Norden and A. Graslund, *J. Am. Chem. Soc.*, 1996, **118**, 5544–5552.

- 14 A. Sen and P. E. Nielsen, *Biophys. J.*, 2006, **90**, 1329–1337.
- 15 A. Sen and P. E. Nielsen, *Nucleic Acids Res.*, 2007, **35**, 3367–3374.
- 16 A. O. Nwokeoji, P. M. Kilby, D. E. Portwood and M. J. Dickman, *Anal. Chem.*, 2017, **89**, 13567–13574.
- 17 S. Ficht, A. Mattes and O. Seitz, *J. Am. Chem. Soc.*, 2004, **126**, 9970–9981.
- 18 A. Roloff and O. Seitz, *Chem. Sci.*, 2013, **4**, 432–436.
- 19 F. Diezmann, H. Eberhard and O. Seitz, *Biopolymers*, 2010, **94**, 397–404.
- 20 T. N. Grossmann and O. Seitz, *J. Am. Chem. Soc.*, 2006, **128**, 15596–15597.
- 21 S. Barluenga and N. Winssinger, *Acc. Chem. Res.*, 2015, **48**, 1319–1331.
- 22 A. Singhal, V. Bagnacani, R. Corradini and P. E. Nielsen, *ACS Chem. Biol.*, 2014, **9**, 2612–2620.
- 23 R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149–2154.
- 24 R. Behrendt, P. White and J. Offer, *J. Pept. Sci.*, 2016, **22**, 4–27.
- 25 P. E. Nielsen, *Peptide Nucleic Acids: methods and protocols*, Humana Press, Totowa, NJ, Second Edi., 2014, vol. 1050.
- 26 K. J. Breslauer, R. Frank, H. Blocker and L. A. Marky, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 3746–3750.
- 27 J. X. Zhang, J. Z. Fang, W. Duan, L. R. Wu, A. W. Zhang, N. Dalchau, B. Yordanov, R. Petersen, A. Phillips and D. Y. Zhang, *Nat. Chem.*, 2018, **10**, 91–98.
- 28 M. Sundaralingam and P. K. Ponnuswamy, *Biochemistry*, 2004, **43**, 16467–16476.
- 29 J. SantaLucia and D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, **33**, 415–440.

- 30 P. W. K. Rothmund, *Nature*, 2006, **440**, 297–302.
- 31 B. Yurke and A. P. Mills, in *Genetic Programming and Evolvable Machines*, 2003, vol. 4, pp. 111–122.
- 32 S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church and W. M. Shih, *Nucleic Acids Res.*, 2009, **37**, 5001–5006.
- 33 P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye and A. A. Louis, *J. Chem. Phys.*, 2012, **137**, 135101.
- 34 J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks and N. A. Pierce, *J. Comput. Chem.*, 2011, **32**, 170–173.
- 35 F. Xuan, T. W. Fan and I. M. Hsing, *ACS Nano*, 2015, **9**, 5027–5033.
- 36 P. Wittung, R. Lyng, B. Nordén, M. Eriksson and P. E. Nielsen, *J. Am. Chem. Soc.*, 1995, **117**, 10167–10173.
- 37 A. Kiliszek, K. Banaszak, Z. Dauter and W. Rypniewski, *Nucleic Acids Res.*, 2015, **44**, 1937–1943.
- 38 F. Totsingan, V. Jain, W. C. Bracken, A. Faccini, T. Tedeschi, R. Marchelli, R. Corradini, N. R. Kallenbach and M. M. Green, *Macromolecules*, 2010, **43**, 2692–2703.
- 39 P. Wittung, P. Nielsen and B. Norden, *J. Am. Chem. Soc.*, 1997, **119**, 3189–3190.
- 40 B. Datta, M. E. Bier, S. Roy and B. A. Armitage, *J. Am. Chem. Soc.*, 2005, **127**, 4199–4207.
- 41 M. D’Abramo, C. L. Castellazzi, M. Orozco and A. Amadei, *J. Phys. Chem. B*, 2013, **117**, 8697–8704.
- 42 L. E. Morrison, T. C. Halder and L. M. Stols, *Anal. Biochem.*, 1989, **183**, 231–

244.

- 43 M. C. Chakrabarti and F. P. Schwarz, *Nucleic Acids Res.*, 1999, **27**, 4801–4806.
- 44 D. J. Nieves, K. Gaus and M. A. B. Baker, *Genes (Basel)*., 2018, **9**, 621–635.
- 45 A. B. Mabire, M. P. Robin, W. D. Quan, H. Willcock, V. G. Stavros and R. K. O'Reilly, *Chem. Commun. (Camb)*., 2015, **51**, 9733–9736.
- 46 F. Manyanga, *J. Anal. Pharm. Res.*, , DOI:10.15406/japlr.2016.02.00013.
- 47 J. G. Duguid, V. A. Bloomfield, J. M. Benevides and G. J. Thomas, *Biophys. J.*, 1996, **71**, 3350–3360.
- 48 P. L. Privalov and A. I. Dragan, *Biophys. Chem.*, 2007, **126**, 16–24.
- 49 X. Qiu, *PLoS One*, 2012, **7**, e39793.
- 50 F. Kates, Steven; Albericio, *Solid-phase synthesis: a practical guide*, CRC press, New York, First edit., 2000.
- 51 R. Subirós-Funosas, R. Prohens, R. Barbas, A. El-Faham and F. Albericio, *Chem. - A Eur. J.*, 2009, **15**, 9394–9403.
- 52 D. Chouikhi, M. Ciobanu, C. Zambaldo, V. Duplan, S. Barluenga and N. Winssinger, *Chem. - A Eur. J.*, 2012, **18**, 12698–12704.
- 53 S. Pothukanuri, Z. Pianowski and N. Winssinger, *European J. Org. Chem.*, 2008, 3141–3148.
- 54 J. J. Díaz-Mochón, L. Bialy and M. Bradley, *Org. Lett.*, 2004, **6**, 1127–1129.
- 55 C. Zambaldo, S. Barluenga and N. Winssinger, *Curr. Opin. Chem. Biol.*, 2015, 26, 8–15.
- 56 R. Subirós-Funosas, S. N. Khattab, L. Nieto-Rodríguez, A. El-Faham and F. Albericio, *Aldrichimica Acta*, 2013, **46**, 21–44.
- 57 A. D. McFarland, J. Y. Buser, M. C. Embry, C. B. Held and S. P. Kolis, *Org.*

- Process Res. Dev.*, 2019, **23**, 2099–2105.
- 58 A. Pernille Tofteng, S. L. Pedersen, D. Staerk and K. J. Jensen, *Chem. - A Eur. J.*, 2012, **18**, 9024–9031.
- 59 M. Hachisu, H. Hinou, M. Takamichi, S. Tsuda, S. Koshida and S. I. Nishimura, *Chem. Commun.*, 2009, 1641–1643.
- 60 J. J. Turner, G. D. Ivanova, B. Verbeure, D. Williams, A. A. Arzumanov, S. Abes, B. Lebleu and M. J. Gait, *Nucleic Acids Res.*, 2005, **33**, 6837–6849.
- 61 J. Coste, D. Le-Nguyen and B. Castro, *Tetrahedron Lett.*, 1990, **31**, 205–208.
- 62 T. I. Al-Warhi, H. M. A. Al-Hazimi and A. El-Faham, *J. Saudi Chem. Soc.*, 2012, **16**, 97–116.
- 63 R. F. Chen, *Arch. Biochem. Biophys.*, 1967, **120**, 609–620.
- 64 G. T. Hermanson, *Bioconjugate Techniques: Third Edition*, 2013.
- 65 M. Taliani, E. Bianchi, F. Narjes, M. Fossatelli, A. Urbani, C. Steinkühler, R. De Francesco and A. Pessi, *Anal. Biochem.*, 1996, **240**, 60–67.
- 66 M. E. Moustafa and R. H. E. Hudson, *Nucleosides, Nucleotides and Nucleic Acids*, 2011, **30**, 740–751.
- 67 A. B. T. Ghisaidoobe and S. J. Chung, *Int. J. Mol. Sci.*, 2014, **15**, 22518–22538.
- 68 P. G. Wu and L. Brand, *Anal. Biochem.*, 1994, **218**, 1–13.
- 69 M. M. Lee and B. R. Peterson, *ACS Omega*, 2016, **1**, 1266–1276.
- 70 B. G. Moreira, Y. You, M. A. Behlke and R. Owczarzy, *Biochem. Biophys. Res. Commun.*, 2005, **327**, 473–484.
- 71 J. Applequist and V. Damle, *J. Chem. Phys.*, 1963, **39**, 2719–2721.
- 72 Y. Zhao, F. Chen, Q. Li, L. Wang and C. Fan, *Chem. Rev.*, 2015, **115**, 12491–12545.

- 73 Y. He and D. R. Liu, *J. Am. Chem. Soc.*, 2011, **133**, 9972–9975.
- 74 E. Stulz, A. J. Turberfield, M. L. McKee, P. J. Milnes, R. K. O'Reilly and J. Bath, *J. Am. Chem. Soc.*, 2012, **134**, 1446–1449.
- 75 M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, A. J. Turberfield and R. K. O'Reilly, *Angew. Chemie Int. Ed.*, 2010, **49**, 7948–7951.
- 76 Y. He and D. R. Liu, *Nat. Nanotechnol.*, 2010, **5**, 778–782.
- 77 W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes, A. H. El-Sagheer, J. Bath, B. G. Davis, T. Brown, R. K. O'Reilly and A. J. Turberfield, *Nat. Chem.*, 2016, **8**, 542–548.
- 78 R. M. Dirks and N. A. Pierce, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15275–15278.
- 79 D. Y. Zhang, A. J. Turberfield, B. Yurke and E. Winfree, *Science*, 2007, **318**, 1121–1125.
- 80 H. Park, A. Gemini, S. Sforza, R. Corradini, R. Marchelli and W. Knoll, *Biointerphases*, 2006, **1**, 113–122.
- 81 A. Singhal, V. Bagnacani, R. Corradini and P. E. Nielsen, *ACS Chem. Biol.*, 2014, **9**, 2612–2620.

Future prospects

DNA encoding has been established as a reliable methodology to identify chemical entities, this has been exploited to detect molecules that have been selected due to an activity such as binding or catalysis. Further to this encoding capabilities, there is extensive evidence showing that selective recognition between complementary DNA strands can be exploited to template chemical reactions. DNA templated reactions usually require sequential addition of the adapters, but more recently autonomous DTS has been demonstrated. While the DNA encoding strategies could be understood as conventional chemical synthesis using molecules labelled with DNA; the latter autonomous DTS strategies emulate (with limitations) the autonomous, encoded and templated ribosomal synthesis of proteins.

From a chemical perspective, DNA encoding is compatible with a large set of reaction conditions, such as organic solvents and denaturing reagents. DNA templated synthesis is limited to reaction conditions that preserve DNA hybridisation, which as discussed in this thesis, are very restricted. Finally, autonomous DTS requires not only reaction conditions that are compatible with DNA hybridisation, but also chemistries with a higher rate than the progression of the DNA displacement mechanism.

This thesis shows that transferring DNA templated synthesis to deep eutectic solvents is not compatible with amide forming chemistries; however, the DNA templated formation of amines through reductive aminations highlights that it is still possible to develop alternative transfer DNA templated chemistries in DES. The following two questions could be the basis of conceptually meaningful results: 1) Is it possible to exploit the high viscosity of DESs for DTS? For example, it is possible that DESs high

viscosity reduces the rate of autonomous DNA displacement mechanisms, thus allowing slower chemistries to proceed to completion between displacement stages.

2) Do DESs allow to perform DNA templated chemistries that would not take place in aqueous solution? It is possible that water evolving reactions are favoured in such environments due to their dehydrating properties.

The challenges highlighted in this thesis to perform efficient DNA templated native chemical ligation in aqueous solution, together with simple modelling of the effect of linker length on the bimolecular reaction rate, shows that in order to exploit the reactivity of stable groups (such as phenol esters) a very precise control over the spatial disposition of chemical groups is required. This could be achieved through supramolecular interactions, in addition to DNA hybridisation, that co-localise the reagents with a high degree of precision. One specific way to achieve this could be achieved through evolved aptamers that co-localise phenol esters and acceptor groups.

Peptide nucleic acids have very strong interactions and limited solubility in aqueous solvent, this is probably one of the reasons why complex displacement PNA mechanisms have not been developed, as opposed to DNA analogues. The reduction on the T_m of dsPNA in high DMF solutions, together with the isothermal annealing shown in this thesis, open up the possibility of performing multistep displacement cascades in DMF / H₂O solvents; the solvent ratio could be finely tuned to allow rapid toehold displacement. With a genetically programmable polymer capable of performing multistep displacement mechanisms in organic or high organic solvent, the window of compatible templated chemistries would open up dramatically.

Appendix 1

Characterisation methods

Reversed phase high performance liquid chromatography (RP-HPLC)

RP-HPLC was performed on a Varian 920LC system with a photodiode array UV detector (PDA), and a fluorescence detector. The solvent system, gradient, temperature and column are detailed in the experimental section dedicated to each individual experiment. The data was processed using Galaxie (Varian) v.1.9 software.

Liquid chromatography-mass spectrometry

LC-MS was performed on an Agilent 1200 HPLC system coupled to a Bruker AmazonX high resolution ion trap, in negative ion mode. The desalted oligonucleotide samples were eluted through a XBridge oligonucleotide BEH C18 column (130 Å, 2.5 µm, 4.6 x 50 mm) using a 5 vol% MeOH, 10 mM ammonium acetate (buffer A) and a 70 vol% MeOH, 10 mM ammonium acetate (buffer B) solvent system. The data was processed using Compass Data Analysis (Bruker) v.4.1 software, and the MaxEnt integrated deconvolution algorithm. Alternatively, LC-MS was performed on a Waters ACQUITY UPLC system coupled to a Xevo GS2-XS qToF mass spectrometer in negative sensitivity mode with leucine-enkephalin [M-H]⁻ 554.2620 Lockspray. The oligonucleotides were eluted through an AQUITY UPLC oligonucleotide BEH C18 column (130 Å, 1.7 µm, 2.1 x 50 mm) using a 50 mM triethylammonium acetate (TEAA, pH 7.0) solution in H₂O (buffer A) and a 50 mM TEAA solution in MeCN (buffer B) at 60 °C and a 0.2 mL·min⁻¹ flow. Source temperature = 100 °C Desolvation temperature

= 250 °C, capillary voltage = 2.5 kV, cone voltage = 40 V, source offset = 80 V, cone gas = 50 L · h⁻¹ and desolvation gas = 600 L · h⁻¹.

Polyacrylamide gel electrophoresis

15% denaturing polyacrylamide gel electrophoresis (PAGE) was performed in 1xTBE (Tris 100 mM, boric acid 100 mM, EDTA 2 mM, pH = 8.3) running buffer, using a vertical nucleic acid electrophoresis cell connected to a PowerPack basic power supply (Biorad). Denaturing gels contained 7 M urea, 25 vol% formamide, a 3:8 dilution of 40 vol% acrylamide solution (19:1 acrylamide:*bis*-acrylamide, Biorad), a 1:10 dilution of 10xTBE buffer and the volume was adjusted with water. The gels were polymerised by addition of 0.0120 vol% of 10 wt% ammonium persulfate (APS) solution and 0.0012 vol% of *N,N,N',N'*-tetramethylethylenediamine. Finally, the combined solution was rapidly poured into a 1 mm glass gel casting system (Biorad) and left to polymerise for 30-45 min. The oligonucleotide samples were diluted with water to approximately 100 nM, and then diluted 1:1 with a formamide solution containing 0.02 wt% bromophenol blue, the resulting solution was heated at 75 °C for approximately 10 min, and loaded onto the gel, prior to performing the electrophoresis. The gel was stained by using a (1:10000) aqueous SybrGold nucleic acid gel stain solution (ThermoFisher), washed repeatedly with water and finally imaged using a ChemiDoc MP transilluminator system (Biorad). The images were processed using ImageLab v.6.0.0 software.

Nuclear magnetic resonance spectroscopy

One-dimensional ¹H and ¹³C NMR, and two-dimensional NMR spectra were recorded on a Bruker Advance 300 MHz, Bruker Advance III 400 MHz or a Bruker Advance III 500 MHz at 25 °C. Deuterated solvents chloroform-*d* (99.8 D atom%), methanol-*d*₄

(99.8 D atom%), dimethyl sulfoxide- d_6 (DMSO- d_6 , 99.8 D atom%) acetone- d_6 (99.9 D atom%) were purchased from Sigma Aldrich. The residual nondeuterated solvent peak was used as a chemical shift (δ , ppm) internal standard. The data was processed using Mestrenova (Mestrelab research) v.12.0.2 and ADC/NMR software.

High resolution mass spectrometry

HRMS was recorded in a Bruker Q-ToF Maxis Plus spectrometer or on a Waters Xevo GS2-XS qToF system. A relative error under 5 ppm was ensured.

Nanopure water

18.2 M Ω ·cm water was produced by processing denoised water using a Millipore simplicity system.

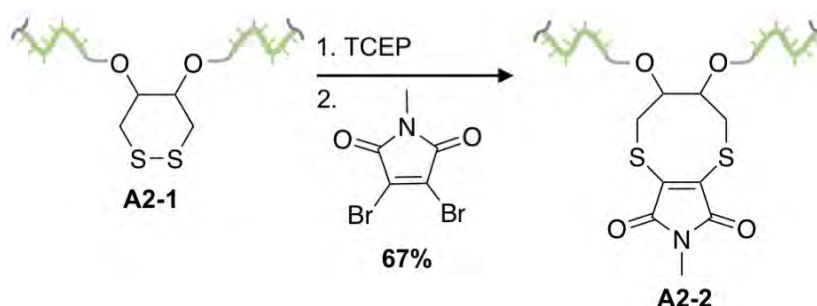
Appendix 2: Additional projects

A2.1 Maleimide fluorophore labelling of DNA

Several substituted maleimides have fluorescence properties and could potentially be used as environmentally sensitive dyes (Y. Xie, J. T. Husband, M. Torrent-Sucarrat, H. Yang, W. Liu and R. K. O'Reilly, *Chem. Commun.*, 2018, **53**, 3339-3342). Here, several methods to prepare maleimide-labelled DNA were explored.

A2.1.1 Internal labelling of DNA with methyl-dithiomaleimide

The functionalisation of a disulfide-labelled DNA oligonucleotide with dibromomaleimide produced the expected dithiomaleimide-labelled DNA (Scheme A2-1).



Scheme A2-1. Synthesis of dithiomaleimide-labelled DNA **A2-2** by reaction with methyl dibromomaleimide (MeDBM).

30 μL of **A2-1** DNA (1 mM, 30 nmol in H_2O), 30 μL of TCEP (100 mM, 3 μmol , pH 9.0 in PBS) and 30 μL PBS (pH 9.0) were mixed together and shaken at r.t. for 10 min. Then, 300 μL of DMF and 30 μL MeDBM (100 mM, 3 μmol , DMF) were added and the solution was shaken overnight. The product was transferred into H_2O by successive steps of centrifugal ultrafiltration and dilution with H_2O using 3 kDa MWCO

amicon spin filters. The product was analysed by LC-MS, showing a 67% conversion (LC-MS integration, Figure A2-1).

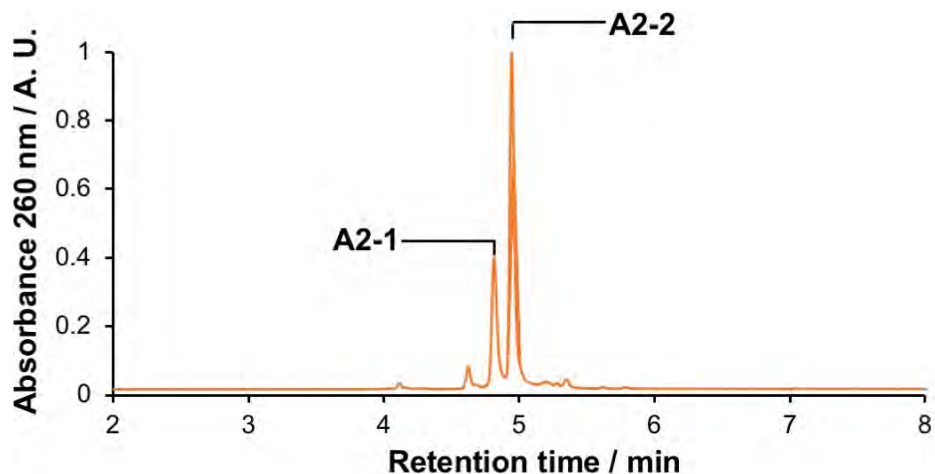
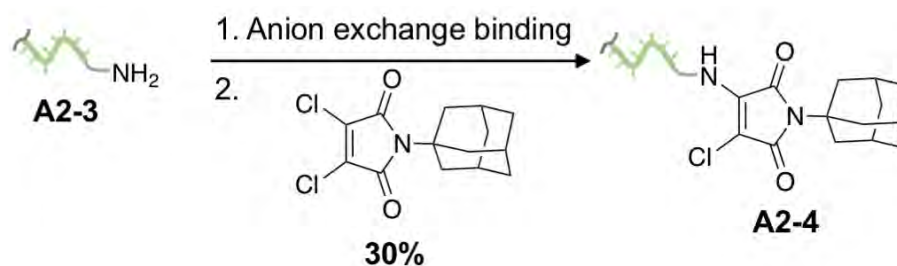


Figure A2-1. LC-MS analysis of the synthesis of dithiomaleimide-labelled DNA **A2-2**. The starting material and the product were identified by LC-MS (Waters Xevo-G2-XS): **A2-1** [M]⁰ *m/z* calcd. 8558.622 found 8558.167; **A2-2** [M]⁰ *m/z* calcd. 8667.393 found 8667.232.

A2.1.2 Adamantyl-aminochloromaleimide functionalisation of DNA

Anion exchange resins have been previously utilised to immobilise and functionalise DNA oligonucleotides using an excess of reagents dissolved in several organic solvents (D. R. Halpin, J. A. Lee, S. J. Wrenn and P. B. Harbury, *PLoS Biol.*, **2004**, *2*, e175; D. T. Flood, S. Asai, X. Zhang, J. Wang, L. Yoon, Z. C. Adams, B. C. Dillingham, B. B. Sanchez, J. C. Vantourout, M. E. Flanagan, D. W. Piotrowski, P. Richardson, S. A. Green, R. A. Shenvi, J. S. Chen, P. S. Baran, P. E. Dawson, *J. Am. Chem. Soc.*, 2019, **141**, 9998-10006). This protocol was applied to the functionalisation of amine-labelled DNA with adamantyl-dichloromaleimide (Scheme A2-2).



Scheme A2-2. Synthesis of adamantly-aminochloromaleimide DNA **A2-4** by reaction with adamantly dichloromaleimide.

Binding buffer: 10 mM AcOH, 0.005 vol% Triton X-100

Eluting buffer: 1.5 M NaCl, 0.005 vol% Triton X-100, 50 mM Tris·HCl, NaOH (5.0 M)
pH = 8.0.

Resin: DEAE Sepharose (Global life sciences solutions)

250 μ L of DEAE sepharose dispersion were placed in an empty Glen Research column. The resin was washed with 20 mL of H₂O and 12 mL of binding buffer. 10 μ L of DNA **A2-3** (1 mM in H₂O, 10 nmol) were diluted in 1 mL of binding buffer, and slowly loaded onto the resin. The column was then washed with 3 mL of binding buffer and 3 mL of DMF. Approximately, 5.6 mg of adamantly-dichloromaleimide (20 μ mol) were dissolved in 500 μ L of DMF and 5 μ L of Et₃N, and mixed with the resin overnight using two syringes. Then, the process was repeated with freshly dissolved adamantly-dichloromaleimide, with occasional mixing during 4 h. The resin was washed with 3 mL of DMF and 3 mL of binding buffer. Finally, the product was recovered by slow elution with 1 mL of eluting buffer and transferred into H₂O by spin filtration with a 3 kDa MWCO amicon spin filter, showing a 30% conversion (LC-MS integration, Figure A2-2).

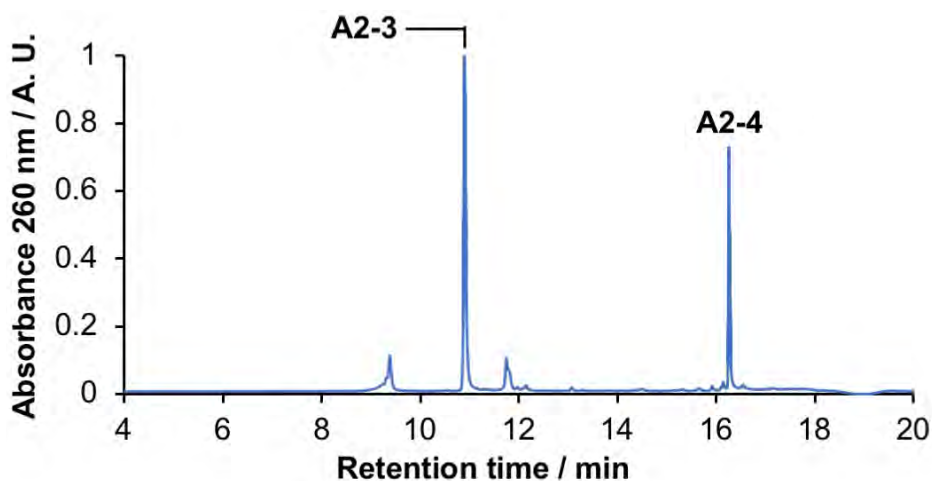
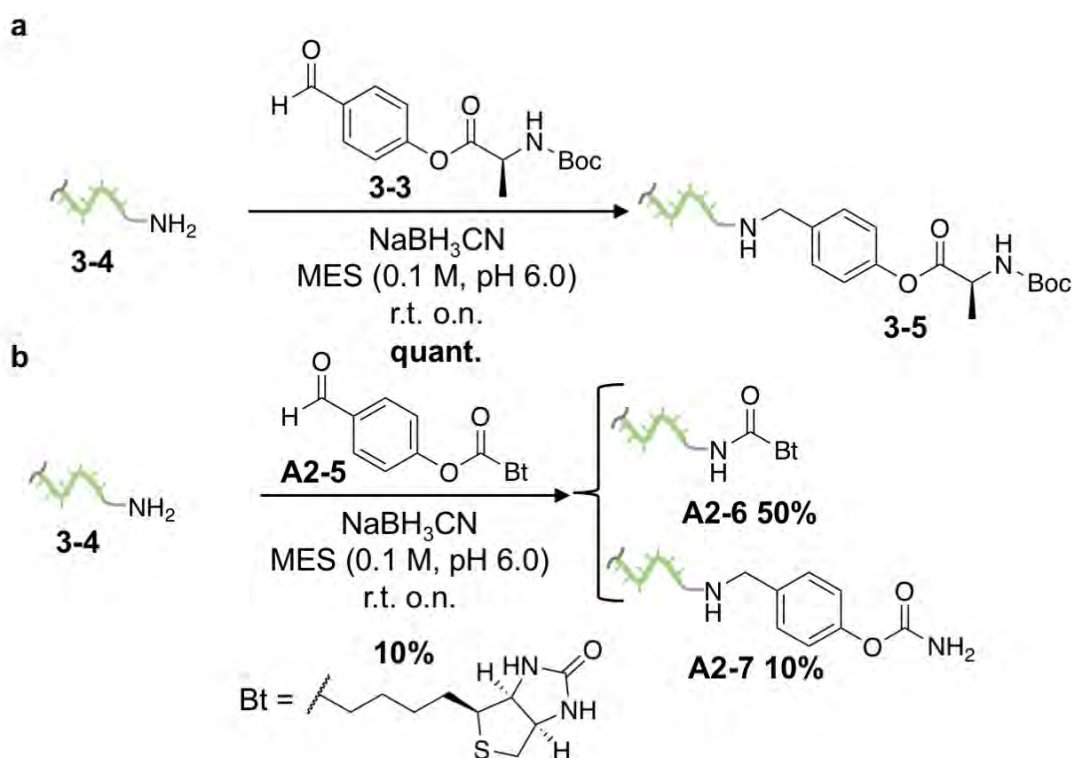


Figure A2-2. LC-MS analysis of the synthesis of aminochloromaleimide-labelled DNA **A2-4**. The starting material and the product were identified by LC-MS (Waters Xevo-G2-XS): **A2-4** $[M]^0$ m/z calcd. 7366.349 found 7366.303; **A2-5** $[M+HCO_2H]^0$ m/z calcd. 7675.426 found 7675.315.

A2.2 Reductive amination and amide coupling of biotin phenol ester to DNA

Reductive amination was a successful strategy for the conjugation of *L*-Ala phenol ester to amine-labelled DNA (Chapter 3, Section 3.2.2). By contrast, the same reaction conditions applied to synthesise a *D*-biotin phenol ester-labelled DNA produced the amide product (Scheme A2-3 a-b). Increasing the equivalents of reducing agent, and reducing the pH to 4 did reduce the amide formation; however, it also reduced the reductive amination product formation.



Scheme A2-3. Comparison of the reactivity of Boc-*L*-Ala and *D*-biotin phenol esters. a) The reductive amination using **3-3** and **3-4** produced the expected secondary amine **3-5**. b) The attempted reductive amination using **A2-5** and **3-4** produced the amide **A2-6** and a small amount of **A2-7**.

Synthesis of **A2-5**

Biotin (244 mg, 1 mmol) and *p*-hydroxybenzaldehyde (122 mg, 1 mmol) were dissolved in 8 mL of DMF and stirred for 5 minutes. EDC (211 mg, 1.1 mmol) and DMAP (61 mg, 0.5 eq) were added and the solution was stirred at r.t. for 24 h. The volume of DMF was reduced to approx. $\frac{1}{2}$ under vacuum and 20 mL of HCl (1 M) were added. The product was precipitated by cooling at $-78\text{ }^{\circ}\text{C}$ for 5 minutes, and the resulting crystals were filtered with a Büchner funnel and washed with 20 mL of HCl (1 M), 20 mL of H_2O and finally 20 mL of EtOH to produce 181 mg of pure product as a white powder (52%).

$^1\text{H-NMR}$ (400 MHz, $\text{DMSO-}d_6$) δ ppm: 10 (s, 1H, CHO), 7.98 (d, $J = 7.6$ Hz, 2H, Ar-*H*), 7.37 (d, $J = 7.7$ Hz, 2H, Ar-*H*), 6.47 (s, 1H, NH), 6.38 (s, 1H, NH), 4.32 (m, 1H, CH Bt), 4.16 (1H, m, CH Bt), 3.14 (m, 1H, CH Bt), 2.84 (dd, $J_1 = 7.0$ Hz $J_2 = 6.4$ Hz, 2H, CH_2)

Bt), 2.62 (m, 2H, CO-CH₂), 1.56 (6H, m, CH₂CH₂CH₂). ¹³C-NMR (100 MHz, DMSO-*d*₆) δ ppm: 192.03 (CHO), 171.34 (C=O), 162.73(C=O), 155.08 (ArC), 133.84 (ArC), 131.09 (ArC-H), 122.75 (ArC-H), 61.02 (C-X, X=N,S), 59.21 (C-X, X=N,S), 55.30 (C-X, X=N,S), 39.88 (C-X, X=N,S), 33.32 (CO-CH₂), 27.99 (CH₂), 27.89 (CH₂), 24.27 (CH₂).

Conjugation of **A2-5** to amine-labelled DNA

3.5 mg of **A2-5** (10 μmol) dissolved in 150 μL of DMF, and 5 μL of **3-4** amine-labelled DNA (2 mM in H₂O, 10 nmol) diluted in 45 μL of MES buffer (0.1 M, pH 6.0) were combined. 6.4 mg of NaBH₃CN (100 μmol) were added, and the solution was shaken at 25 °C o.n. The reaction mixture was diluted with 300 μL of H₂O, the precipitate was centrifuged and the supernatant was transferred into H₂O by successive steps of centrifugal ultrafiltration and dilution with H₂O using 3 kDa MWCO amicon spin filters. The product was analysed by LC-MS, showing approximately a 40% of starting material, 50% of **A2-6**, and a 10% of the desired product (Figure A2-3).

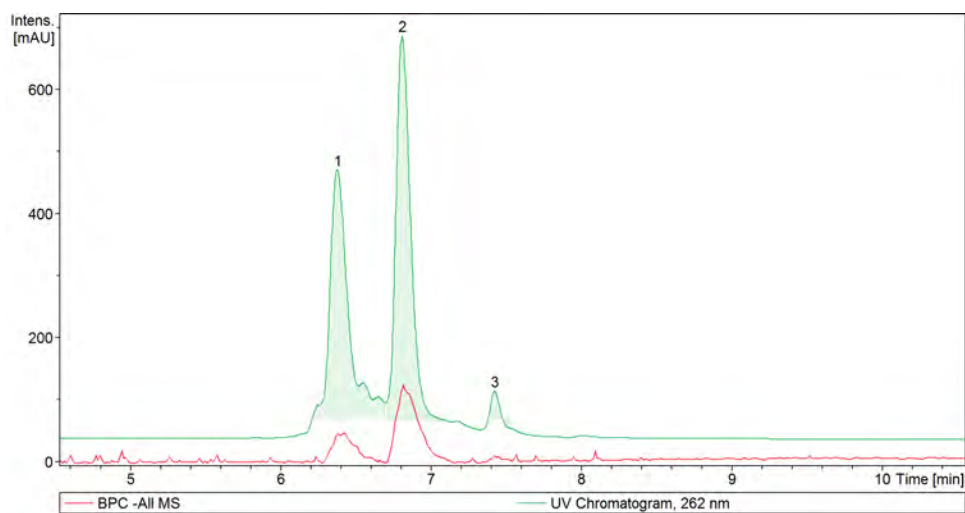


Figure A2-3. LC-MS (Bruker AmazonX Ion Trap) analysis of the reductive amination attempt with **A2-5** and **3-4**. Compound 1 = starting material (**3-4**, [M-H]⁻ *m/z* calcd. 9428.65 found 9430.04); compound 2 = **A2-6** ([M-H]⁻ *m/z* calcd. 9654.73 found 9657.56); compound 3 = **A2-7** ([M-H]⁻ *m/z* calcd. 9761.78 found 9762.86).