A new class of enzymatically cleavable nucleotides for DNA sequencing by synthesis

By

Maria Lopalco

Doctor of Philosophy

January 2009
A new class of enzymatically cleavable nucleotides for DNA sequencing by synthesis

By Maria Lopalco

A new family of nucleotides has been designed, synthesised and evaluated for DNA sequencing by synthesis. Each of the nucleotide analogues had a free 3'-OH group and a fluorophore attached to the base through a small peptide linkage that was designed to stop multiple base additions. Analysis by HPLC and MALDI-TOF proved the modified nucleotides were incorporated into a growing DNA strand by a DNA polymerase and that the peptide linker was cleaved with high efficiency after incubation of the extended primer with a protease.

Enzymatically cleavable nucleotides were successfully applied to chip-based sequencing by synthesis cycles. The fluorescent emission revealed the identity of the incorporated nucleotide and the removal of the fluorophore by a protease ensured the detection of the next base in the following cycle.

Additionally, a practical microwave-mediated solid-phase protocol has been developed for the synthesis of cyanine dyes spanning the visible and the near infrared spectrum to allow the preparation of a series of fluorescent nucleotides for the design of a four-colour DNA sequencing technology.
The research described in this thesis was carried out by the author under the supervision of Prof. Mark Bradley at the University of Southampton (Sept. 2004 – Feb. 2005) and at the University of Edinburgh (Feb. 2005 – May 2008). No part of this thesis has been previously submitted at this or any other university for any other degree or a professional qualification.

Dr. Jin Ku Cho contributed to nucleotide synthesis (Chapter 3), Geraldine Escher contributed to synthesis and in vitro testing of labelled peptoids (Chapter 2), Lois Alexander contributed to microspheres synthesis and labelling (Chapter 2).

Part of this work has been published in the scientific literature:


Signed:

Date: 20/05/09
Acknowledgements

I would like to express my gratitude to Professor Mark Bradley for his patience, trust and for being always present whenever I needed his guidance.

I would like to acknowledge Prof. Cameron Neylon and Peter Roach for their supervision at the very beginning of my PhD and all the members of the 4G group.

I would like to thank the entire Bradley group, both former and current members and, in particular, Dr. Jin Ku Cho for his contribution to this work, for his advices and patience. I would like to thank Dr. Juan Jose Diaz Mochon for his help throughout my PhD and for sharing his knowledge and expertises.

I would like to express my gratitude to Rosario, Carole, Asier, Geraldine, Delphine and Graham for their help during these years and to Jon for allowing me to use the facilities of his lab.

I am especially thankful to all the people I have met during this adventure: Neil, Amaury, Nikos, Maria and all the “greek community” and my flatmates in Montefiore. Thank you for the great time we spent together in Southampton. A special thank goes to Stine and Stephany for their friendship and support in difficults moments. I would like to thank Ellen and Paty, Julian, Nunzio, Sigmund and Cheryl for welcoming me in their lab and for making the writing experience more enjoyable. Thank you all for your friendship.

I would like to thank my family for their love and constant encouragement all these years, especially in hard times, and for being always next to me, despite the distance.

My most profound gratitude goes to Julien. Meeting you has been the biggest surprise of this adventure. Thank you for your love, friendship and for all the happy moments we spent together these years. Thank you for the time and encouragement you have given me during the preparation of this thesis.

iii
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adenine, adenosine, alanine</td>
</tr>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>Ac</td>
<td>acetyl</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>Ahx</td>
<td>6-amino hexanoic acid</td>
</tr>
<tr>
<td>Ala</td>
<td>alanine</td>
</tr>
<tr>
<td>Am-dCTP</td>
<td>aminopropargyl deoxycytidine triphosphate</td>
</tr>
<tr>
<td>Am-dUTP</td>
<td>aminopropargyl deoxyuridine triphosphate</td>
</tr>
<tr>
<td>AMP</td>
<td>adenosine monophosphate</td>
</tr>
<tr>
<td>APS</td>
<td>adenosine phosphosulfate</td>
</tr>
<tr>
<td>aq</td>
<td>aqueous</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BEAMing</td>
<td>beads, emulsion, amplification, magnetics</td>
</tr>
<tr>
<td>Bn</td>
<td>benzyl</td>
</tr>
<tr>
<td>BODIPY</td>
<td>boron difluoride dipyrrromethene</td>
</tr>
<tr>
<td>bp</td>
<td>boiling point, base pair</td>
</tr>
<tr>
<td>br</td>
<td>broad</td>
</tr>
<tr>
<td>Bu</td>
<td>butyl</td>
</tr>
<tr>
<td>Bz</td>
<td>benzoxyl</td>
</tr>
<tr>
<td>C</td>
<td>cytosine, cytidine</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>CAE</td>
<td>capillary array electrophoresis</td>
</tr>
<tr>
<td>calcd</td>
<td>calculated</td>
</tr>
<tr>
<td>CCD</td>
<td>charge-coupled device</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CE</td>
<td>capillary electrophoresis</td>
</tr>
<tr>
<td>cm</td>
<td>centimetre</td>
</tr>
<tr>
<td>CPG</td>
<td>Controlled Pore Glass</td>
</tr>
<tr>
<td>CRT</td>
<td>cyclic reversible termination</td>
</tr>
<tr>
<td>Cy</td>
<td>cyclohexyl, cyanine</td>
</tr>
<tr>
<td>d</td>
<td>doublet, days</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>DCM</td>
<td>dichloromethane</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine triphosphate</td>
</tr>
<tr>
<td>ddATP</td>
<td>dideoxyadenosine triphosphate</td>
</tr>
<tr>
<td>ddCTP</td>
<td>dideoxyctydine triphosphate</td>
</tr>
<tr>
<td>ddGTP</td>
<td>dideoxyguanosine triphosphate</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleoside triphosphate</td>
</tr>
<tr>
<td>ddTTP</td>
<td>dideoxythymidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>DIC</td>
<td>diisopropylcarbodiimide</td>
</tr>
</tbody>
</table>

iv
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIEA</td>
<td>$N,N$-diisopropyl-$N$-ethylamine</td>
</tr>
<tr>
<td>DIPEA</td>
<td>$N,N$-diisopropyl-$N$-ethylamine</td>
</tr>
<tr>
<td>DLP</td>
<td>digital micromirror device</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-(dimethylamino)-pyridine</td>
</tr>
<tr>
<td>DMF</td>
<td>dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DMT</td>
<td>dimethoxytrityl</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleoside triphosphate</td>
</tr>
<tr>
<td>DSC</td>
<td>disuccinimidyld carbonate</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymidine triphosphate</td>
</tr>
<tr>
<td>DVB</td>
<td>divinyl benzene</td>
</tr>
<tr>
<td>EDC</td>
<td>1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide</td>
</tr>
<tr>
<td>ELSD</td>
<td>evaporative light scattering detector</td>
</tr>
<tr>
<td>eq</td>
<td>equivalent</td>
</tr>
<tr>
<td>ES</td>
<td>electrospray mass spectrometry</td>
</tr>
<tr>
<td>ET</td>
<td>energy transfer</td>
</tr>
<tr>
<td>Et</td>
<td>ethyl</td>
</tr>
<tr>
<td>FACS</td>
<td>fluorescence activated cell sorting</td>
</tr>
<tr>
<td>FAM</td>
<td>carboxyfluorescein</td>
</tr>
<tr>
<td>FISSEQ</td>
<td>fluorescent in situ sequencing</td>
</tr>
<tr>
<td>FITC</td>
<td>fluorescein isothiocyanate</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-fluorenylmethoxycarbonyl</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy-transfer</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared spectroscopy</td>
</tr>
<tr>
<td>g</td>
<td>grams</td>
</tr>
<tr>
<td>G</td>
<td>guanine, guanosine</td>
</tr>
<tr>
<td>h</td>
<td>hours</td>
</tr>
<tr>
<td>HGP</td>
<td>human genome project</td>
</tr>
<tr>
<td>HOBt</td>
<td>1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>HPA</td>
<td>3-hydroxypicolinic acid</td>
</tr>
<tr>
<td>HRMS</td>
<td>high resolution mass spectrometry</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>'Pr</td>
<td>isopropyl</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>J</td>
<td>coupling constant</td>
</tr>
<tr>
<td>L</td>
<td>litres</td>
</tr>
<tr>
<td>m</td>
<td>multiplet, metres</td>
</tr>
<tr>
<td>M</td>
<td>moles per litre</td>
</tr>
<tr>
<td>m/z</td>
<td>mass to charge ratio</td>
</tr>
<tr>
<td>MALDI-MS</td>
<td>Matrix-assisted laser desorption ionisation mass spectrometry</td>
</tr>
<tr>
<td>MAS</td>
<td>maskless array synthesiser</td>
</tr>
<tr>
<td>MAS</td>
<td>magic angle spinning</td>
</tr>
</tbody>
</table>
Me  methyl
mer  monomer
min  minutes
mol  moles
mp  melting point
MPSS  massively parallel signature sequencing
MS  mass spectrometry
MW  molecular weight
n.d.  not determined
n/a  not available or not applicable
NBP  nitrobenzyl pyridine
NDP  nucleoside diphosphate
NHS  N-hydroxy succinimide
NIR  near infrared
NMP  nucleoside monophosphate
NMR  nuclear magnetic resonance
NTP  nucleoside triphosphate
PBS  phosphate buffered saline
PCR  Polymerase Chain Reaction
PEG  polyethylene glycol
PGP  personal genome project
Ph  phenyl
PPi  pyrophosphate
ppm  parts per million
Pr  propyl
PS  polystyrene
PS-PEG  polystyrene-polyethylene glycol copolymer
q  quartet
R110  rhodamine 110
R6G  rhodamine 6G
Rf  retention factor (movement of compound/solvent front)
Rho  rhodamine
RNA  ribonucleic acid
ROX  carboxy-X-rhodamine
RP-HPLC  reverse-phase high performance liquid chromatography
rt  room temperature
s  singlet, seconds
SBH  sequencing by hybridisation
SBS  sequencing by synthesis
SDS  sodium dodecyl sulfate
SEM  standard error of the mean
SM  starting material
SMRT  single-molecule real-time
SN2  bimolecular nucleophilic substitution
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNA</td>
<td>single nucleotide addition</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SOLiD</td>
<td>sequencing by oligonucleotide ligation and detection</td>
</tr>
<tr>
<td>SPSC</td>
<td>sodium phosphate saline citrate</td>
</tr>
<tr>
<td>SSC</td>
<td>saline-sodium citrate</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single-stranded DNA</td>
</tr>
<tr>
<td>t</td>
<td>triplet</td>
</tr>
<tr>
<td>T</td>
<td>thymine, thymidine</td>
</tr>
<tr>
<td>TAMRA</td>
<td>tetramethylcarboxyrhodamine</td>
</tr>
<tr>
<td>TBAI</td>
<td>tetrabutylammonium iodide</td>
</tr>
<tr>
<td>'Bu</td>
<td>tert-butyl</td>
</tr>
<tr>
<td>TCA</td>
<td>trichloroacetic acid</td>
</tr>
<tr>
<td>TEA</td>
<td>triethylamine</td>
</tr>
<tr>
<td>TEAA</td>
<td>triethylammonium acetate</td>
</tr>
<tr>
<td>TEAB</td>
<td>triethylammonium bicarbonate</td>
</tr>
<tr>
<td>Tfa</td>
<td>trifluoroacetate</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>tlc</td>
<td>thin layer chromatography</td>
</tr>
<tr>
<td>tR</td>
<td>elution time of compound on HPLC</td>
</tr>
<tr>
<td>TxR</td>
<td>texas red</td>
</tr>
<tr>
<td>U</td>
<td>uracil, uridine, unit</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>δ</td>
<td>chemical shift in ppm downfield from tetramethysilane</td>
</tr>
<tr>
<td>λ</td>
<td>wavelength</td>
</tr>
<tr>
<td>μ0</td>
<td>microwave</td>
</tr>
</tbody>
</table>
# Table of Contents

1  INTRODUCTION ................................................................. 1

1.1 Beyond the Human Genome Project (HGP) ......................... 1

1.2 Deoxyribonucleic Acid (DNA) ............................................. 2
  1.2.1 Nucleotides ..................................................................... 2
  1.2.2 Primary structure of DNA .............................................. 3
  1.2.3 Secondary structure of DNA .......................................... 4
  1.2.4 DNA replication ............................................................ 5

1.3 Sanger DNA Sequencing ...................................................... 6
  1.3.1 Chain Termination method ........................................... 6
  1.3.2 Advances in Sanger sequencing ..................................... 7

1.4 New High Throughput DNA Sequencing technologies ............. 10

1.5 Array-based Cyclic Sequencing Technologies ...................... 12
  1.5.1 Solid Phase Arrays for DNA Sequencing ......................... 12
  1.5.2 Sequencing by Hybridisation ........................................ 15
  1.5.3 Sequencing by Synthesis .............................................. 17
  1.5.4 Reversible Terminators ................................................ 25
  1.5.5 Single Molecule DNA Sequencing ................................. 30

1.6 Aims of the project ........................................................... 33

2 Development of a practical method for the synthesis of cyanine
dyes ......................................................................................... 35

2.1 Introduction ........................................................................ 35
  2.1.1 Cyanine dyes ................................................................. 36
  2.1.2 Synthesis of cyanine dyes .............................................. 37
  2.1.3 Solid phase synthesis of cyanine dyes ............................ 39

2.2 Microwave-mediated synthesis of multicolour cyanine dyes ...... 41
  2.2.1 Synthesis of N-alkyl quaternary ammonium salts ............. 41
  2.2.2 Preparation of polymer-bound polyene-chain precursors ...... 47
  2.2.3 Formation of hemicyanines ............................................ 51
  2.2.4 Formation of mono-reactive cyanine dyes ........................ 54
2.3 Catch and release synthesis of heptamethine dyes .......................... 57

2.4 Synthesis of water soluble cyanine dyes ........................................... 61
  2.4.1 N-Alkylation of sulfonated indolenines ....................................... 62
  2.4.2 Synthesis of sulfonated cyanine dyes ....................................... 64

2.5 Solid phase synthesis of water soluble cyanine dyes ...................... 66
  2.5.1 From polymer-bound imidates.................................................... 66
  2.5.2 Catch and release approach ..................................................... 68

2.6 Summary and Conclusions ............................................................... 69

2.7 Cyanine dyes and their applications ............................................. 72
  2.7.1 Labelled peptoids for in vivo imaging applications ....................... 72
  2.7.2 Labelling of microspheres/doughnuts & in vivo imaging ............... 75
  2.7.3 Labelling of nucleic acids & nucleotides .................................. 76

3 Design & Synthesis of Enzymatically Cleavable Nucleotides ............ 77

3.1 Reporter Nucleotides for DNA Sequencing by Synthesis ............... 77
  3.1.1 Design of a Novel Reporter Nucleotide ..................................... 79
  3.1.2 Synthetic strategy ................................................................. 82

3.2 Design and Synthesis of Labelled Peptides .................................... 84
  3.2.1 Design of the Peptidic Linker .................................................. 84
  3.2.2 Selection of the fluorophore .................................................... 85
  3.2.3 Synthesis of the fluorophores ................................................ 87
  3.2.4 Peptide Labelling ................................................................. 91

3.3 Aminopropargyl nucleoside triphosphates .................................... 94

3.4 Enzymatically Cleavable Nucleotides .......................................... 96
  3.4.1 Coupling of labelled peptides to Am-dUTP ................................ 98
  3.4.2 Coupling of labelled peptides to Am-dCTP ................................ 100
  3.4.3 Coupling of FAM-(Ala)3-OH to Am-dGTP ................................ 103

3.5 Conclusion .......................................................................................... 104

4 Single Nucleotide Primer Extension with Enzymatically Cleavable
  Nucleotides: Proof of the Principle .................................................. 106

4.1 Single Nucleotide Primer Extension .............................................. 106
4.1.1 DNA Polymerase ................................................................. 106
4.1.2 Analysis of Primer Extension ................................................. 108

4.2 Primer Extension with Enzymatically Cleavable Nucleotides..... 109

4.3 Removal of the Reporter Group .................................................. 112

4.4 Investigation of the Termination Properties of Enzymatically Cleavable Nucleotides ................................................................. 114

4.5 Conclusion .................................................................................. 116

5 Chip-Based Interrogation of DNA Sequences with Enzymatically Cleavable Nucleotides ................................................................. 118

5.1 Introduction................................................................................... 118

5.2 DNA Microarrays: Templates ........................................................ 120

5.3 DNA microarrays: Surface and Immobilisation Chemistry...... 120

5.4 DNA microarrays: printing and hybridisation ......................... 124

5.4.1 Design of target and probes sequences ............................... 124

5.4.2 Optimal target and probe density ........................................ 125

5.4.3 Immobilisation of DNA Primers on CodeLink™ slides and Hybridisation ................................................................. 126

5.5 Single Nucleotide Primer extension using Fluorescent Dideoxynucleotides............................................................................ 128

5.5.1 Optical Read-Out .................................................................. 129

5.5.2 Analysis of enzymatic primer extension ................................. 133

5.6 Primer Extension using Protease Cleavable Fluorescent Nucleotides: proof-of-principle experiments ............................................ 134

5.6.1 Incorporation of dUTP-(A)₃- Rho ........................................ 134

5.6.2 Two Steps in One: Hybridisation & Extension ....................... 136

5.6.3 Incorporation of dCTP-(A)₃-FITC ........................................ 138

5.6.4 Incorporation of dCTP-(A)₃-Cy5 (hydrophobic) .................... 139

5.7 Primer Extension using Protease Cleavable Fluorescent Nucleotides: Optimisation ................................................................. 141

5.7.1 Dual color primer extension .................................................. 141

5.8 Removal of the Fluorophore and Second extension .................. 146

5.9 Conclusion .................................................................................. 149
EXPERIMENTAL ................................................................. 150

6.1 Experimental for Chapter 2.................................................. 150

6.2 Experimental for Chapter 3.................................................. 181

6.3 Experimental for Chapter 4.................................................. 207

6.4 Experimental for Chapter 5.................................................. 210

6.4.1 Preparation of DNA Microarrays........................................ 211

6.4.2 Four Color Primer Extension with dideoxynucleotides on CodeLink™ slides.................................................. 213

6.4.3 Primer Extension using Enzymatically Cleavable Nucleotides on CodeLink™ slides.................................................. 217

6.4.4 Two steps in one: annealing and primer extension.............. 218

6.4.5 Dual Color Primer Extension Assay.................................... 220

6.4.6 SBS cycles........................................................................ 221

6.4.7 Dual Color Primer Extension: Optimisation....................... 222

REFERENCES ................................................................................. 232
1 INTRODUCTION

1.1 Beyond the Human Genome Project (HGP)

The human genome contains more than 3 billion base pairs and 20,000-25,000 genes; with only 1.5% of the DNA contained in the genome encoding for proteins. Deciphering the DNA sequences of genes is critical for understanding the function of each human gene and the genetic bases of susceptibility to diseases. This was the primary goal of the Human Genome Project (HGP) an international research project headed by James Watson at the U.S. National Institute of Health.\textsuperscript{1,2} A parallel sequencing project was led by Venter, former president and founder of Celera Genomics. The first working drafts of the human genome, generated by the publicly sponsored Human Genome Project and by Celera Genomics, using two different sequencing technologies, were both released in February 2001.\textsuperscript{1,3} The complete draft of the human genome was released in 2003.\textsuperscript{4}

A deeper understanding of the genome, made possible by the HGP, is having a strong impact on many different areas across life sciences such as molecular medicine, energy production, evolution and forensic. By having a detailed map of all the genes it is possible to investigate the causes of diseases, develop more rapid diagnostic tests and study how genes affect the way individuals respond to drugs. Sequencing of pathogenic microbes and drug resistant strains can give important insights into drug resistance and aid designing more effective drugs.

The realisation of this goal requires the development of fast and cheap technologies which enable scientists to analyse DNA in a massively parallel manner. In the last three decades, advances in DNA sequencing technology have already allowed an impressive reduction of the sequencing cost: from $300 million in 2003 (first draft of the human genome) to $60,000 (March 2008) and the cost of sequencing an entire human genome is expected to drop further to $5000 by 2009.\textsuperscript{5}

Established DNA sequencing technologies face many limitations in terms of speed and costs. Therefore, a "race" to develop faster and cheaper DNA sequencing technologies started, with the aim of reducing the cost of sequencing a human genome to $1000.\textsuperscript{6}
In this chapter the structure of DNA will be first briefly reviewed to understand the chemistry behind sequencing technologies. The history of DNA sequencing technologies will be then presented, with an emphasis on the most common established techniques and recent developments. The chapter will conclude with a statement of aims for the present thesis.

1.2 Deoxyribonucleic Acid (DNA)

Deoxyribonucleic acid (DNA) is a biopolymer, which holds the genetic information necessary for living organisms to function. This information is encoded in the sequence of the building blocks (nucleotides) that form DNA.\(^7\)

1.2.1 Nucleotides

A nucleotide consists of three components covalently bound together: a pentose sugar, a nucleobase and one to three phosphate groups (Figure 1.1). The carbon atoms of the sugar ring are numbered 1’ through 5’ to distinguish them from the atoms of the base.

The base is attached to the C1’ position of the sugar to form what is referred to as nucleoside. When one, two or three phosphate groups are attached to the 5’ carbon of the sugar, the nucleoside is referred to as nucleotide. Nucleotides containing the sugar deoxyribose (like the ones forming DNA) are called deoxynucleotides and they lack of the 2’ hydroxyl group which is normally present in the nucleotides forming RNAs. When the sugar lacks both hydroxyl groups at the 2’ and 3’ position, the nucleotide is referred to as dideoxynucleotide. The latter has found application in the dideoxy-chain termination DNA sequencing method developed by Sanger in 1977.\(^8\)

There are four different types of bases in the nucleotides forming DNA: adenine (A), guanine (G), cytosine (C) and thymine (T). The first two nucleobases are purines and are numbered 1 through 9. The last two nucleobases are pyrimidines and are numbered 1 through 6 (Figure 1.2).
1.2.2 Primary structure of DNA

The DNA backbone consists of nucleotides covalently linked through phosphodiester bonds between the 3' hydroxyl group on the sugar ring of one nucleotide and the phosphate group attached at the 5' position of the adjacent nucleotide. These phosphodiester bonds lead to the formation of the so called “sugar-phosphate backbone” from which the bases project and stack atop each other.

The DNA strand has two distinctive ends referred to as the 5' and the 3' ends, with the 5' end being the one with a terminal phosphate group and the 3' end being the one with a terminal hydroxyl group (Figure 1.3). DNA is synthesised by polymerases in a 5' to 3' direction by adding nucleotides to the 3' end (hydroxyl group) of the previously incorporated nucleotide.
In living organisms, DNA is typically present as a double helix of two antiparallel polynucleotides chains held together by hydrogen bonds between the bases.

### 1.2.3 Secondary structure of DNA

The structure of DNA was published in 1953 by Watson and Crick based on the X-ray diffraction data obtained from Wilkins and Franklin. In the double helix, in which a base forms hydrogen bonds with a specific base on the opposite strand, A and T form two hydrogen bonds while C and G form three hydrogen bonds (Figure 1.4). The stability of DNA is determined by the GC content and by the length of the double helix. Long DNA molecules with a high GC content have stronger interactions between the two strands because of the higher number of hydrogen bonds and stacking interactions between the aromatic bases.

This arrangement of nucleotides in the double helix is the basis for base-pair complementarity and is essential when DNA is copied. In the next section a brief overview of the DNA replication process is presented.
1.2.4 DNA replication

Replication of DNA occurs during cell division cycles and starts at specific sites of the chromosomes, called origins of replication. The DNA strands are unwound by enzymes called helicases and the resulting single-stranded DNA stabilised by the binding of single strand binding proteins to allow access to DNA polymerase. DNA is then synthesised by a polymerase, an enzyme able to bind and move along a single DNA strand using it as a template to extend the 3’ end of a primer by adding nucleotides. The primer extended during replication is a RNA primer (synthesised by the enzyme primase using a DNA strand as a template). The energy necessary for the polymerisation is provided by the removal of two of the phosphate groups attached to the sugar of the incorporated nucleotide. The other phosphate is involved in the formation of a phosphodiester bond. DNA polymerases have proof-reading activity, they correct mistakes in newly synthesised DNA by excising the incorrect base pair. After base excision, the polymerase can insert the correct nucleotide and the replication can continue. At the end of the replication process, two exact copies of the original DNA molecule obtained are divided between the two daughter cells.
DNA replication can be carried out \textit{in vitro} using isolated DNA polymerases and synthetic DNA primers complementary to a known sequence in the DNA template.

1.3 Sanger DNA Sequencing

The sequencing method that enabled scientists to realise the Human Genome Project was the Chain Termination method introduced by Sanger in 1977.

1.3.1 Chain Termination method

The chain termination DNA sequencing method is based on the process of DNA replication and involves the synthesis of new strands of DNA complementary to a DNA strand of unknown sequence that acts as template for the incorporation, by a DNA polymerase, of nucleotides to the 3’ end of a defined DNA primer (Figure 1.6). Firstly, the DNA to be sequenced is fragmented and cloned. After extraction, the fragments are amplified by PCR.

![Figure 1.5 Structure of a deoxynucleotide (dCTP) and a dideoxynucleotide (ddCTP).](image)

Either the primer or the dideoxynucleotides are labelled with different fluorophores in order to be detected. As all four deoxynucleotides are present, chain elongation proceeds until dideoxynucleotides are incorporated. These molecules terminate chain elongation as they lack the 3’ hydroxyl group and therefore they cannot form a phosphodiester bond with another nucleotide (Figure 1.5). In dye-primer sequencing, four sequencing reactions are carried out in parallel. Each reaction mix contains a dye-labelled primer, four dNTPs and one particular ddNTP. In dye-terminator sequencing (currently used in Sanger-based DNA sequencers), the sequencing reaction is carried out using four dNTPs, unlabelled primers and
four labelled ddNTPs. The result is the formation of many DNA fragments of
different lengths and colours emitting light at different wavelengths. All fragments
are separated, by capillary electrophoresis\textsuperscript{14} in an automated DNA sequencer,
according to their size and the identity of the final base of each fragment revealed
by the color detected by the laser at the bottom of the gel. By overlapping, it is
possible to determine the sequence of nucleotides in each fragment and
consequently the sequence of the unknown template.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.6}
\caption{Chain termination DNA sequencing using fluorescent ddNTPs: a) polymerase reaction
with a mixture of dNTPs and labelled ddNTPs; b) labelled fragments generated during polymerase
reaction and terminated after incorporation of a labelled ddNTP; c) separation of all fragments by
capillary electrophoresis; d) detection of fluorescence signals.}
\end{figure}

\subsection{1.3.2 Advances in Sanger sequencing}

Since its introduction in 1977, the Sanger method has undergone many major
innovations which have been essential for the completion of the HGP. Advances in
enzymology,\textsuperscript{15,16} fluorescent dyes,\textsuperscript{17-19} fluorescence detection,\textsuperscript{12} capillary
electrophoresis,\textsuperscript{14} DNA purification, cloning methods,\textsuperscript{20} informatics and robotics
contributed to an increase in the read-length, the accuracy and throughput of the
Sanger method. Most laboratories have routine access to this technology and
sequence DNA with a read-length of up to \textasciitilde1000 \text{ bp.}\textsuperscript{5}
1.3.2.1 \textit{Capillary Electrophoresis}

Capillary electrophoresis (CE) has gained popularity since its introduction in the 1990’s in automated DNA sequencers for the separations of DNA fragments instead of the large vertical slab gels, used in the original Sanger method.\textsuperscript{14,21}

A CE instrument comprises two electrolyte chambers linked by a capillary gel (~50-100 µm in diameter) to separate DNA chains under the application of an electric field and an in-line detector positioned at the capillary outlet to detect the fluorescent signals. This format allows rapid heat dissipation and enables faster separations than in slab gels. In addition it is highly compact and therefore requires low reagent volumes (nanoliter range) while reducing dramatically the cost of sequencing ($0.50 per kilobase).

In 1997, another important innovation contributed to speed up the realisation of the HGP: the introduction of capillary array electrophoresis (CAE) in DNA sequencers.\textsuperscript{22-25} CAE is a multiplexed array of 96 or 384 capillaries arranged on a miniature chip.\textsuperscript{26,27} The DNA fragments are introduced into the array and then separated in parallel inside the capillaries. As the fragments pass in the capillaries, they are irradiated all at once and detected with laser-induced fluorescence. Modern sequencers using CAE, contain 384 or more capillaries and are able to process up to 1000 bp long reads with a production throughput of 2.5 megabases per day.\textsuperscript{5,28} The Sanger method coupled to microchip-based electrophoretic technologies is by far the most competitive platform for sequencing, in relation to read length and accuracy (99.999%); however, it is limited in terms of throughput.

1.3.2.2 \textit{Fluorescent Dyes \& Detection}

In the original chain termination method introduced by Sanger, DNA fragments were tagged using radiolabelled primers made radioactive through incorporation of a phosphorus ($^{32}$P) or sulphur isotope ($^{35}$S), and then loaded in four lanes of a high-resolution polyacrylamide gel for separation. The reaction products were then detected by autoradiography.\textsuperscript{8}

In 1986, the introduction of fluorescent dyes eliminated the need for manipulation of hazardous radioactive compounds and enabled DNA sequencing with higher sensitivity and in a single lane (instead of 4).\textsuperscript{12} Since then, fluorophores with higher sensitivity and better spectral discrimination have been developed for the labelling
of nucleotides together with enzymes able to incorporate them efficiently in DNA.\textsuperscript{18} An ideal set of dyes for DNA sequencing should include fluorophores with a strong absorbance at a common excitation wavelength and strong emission at four different wavelengths. However four-colour sequencing systems have shown some drawbacks as the fluorophores spectra often overlap and the excitation of four fluorophores from a common laser source has proven to be inefficient.\textsuperscript{29,30} This problem has been partially resolved by using multiple lasers\textsuperscript{31} or Förster resonance energy- transfer (FRET) dyes for the labelling of primers or ddNTPs.\textsuperscript{17,32,33} FRET takes place when a donor chromophore in its excited state transfers its excitation energy to an acceptor chromophore by a long-range dipole-dipole coupling mechanism. The efficiency of FRET depends on the distance between donor and acceptor (usually less than 10 nm) and on the degree of overlap between the emission spectra of the donor and the absorption spectra of the acceptor.\textsuperscript{34} In 1995, Ju et al. described the synthesis of universal ET (energy transfer) primers having fluorescein as a donor dye and R110, R6G, TAMRA, ROX as acceptor dyes (Figure 1.7).\textsuperscript{17,32} This set of ET primers has shown good spectral separation of emission wavelengths and good signal strength. Currently, FRET-based dyes are widely used in DNA sequencing for the labelling of primers (such as DYEEnam\textsuperscript{TM} ET primers from GE Healthcare) and nucleotide-terminators (DYEEnam\textsuperscript{TM} ET terminators from GE Healthcare and Big Dye terminators from Applied Biosystems).\textsuperscript{29}

Despite the advantage of using FRET-based dyes, the emission intensities of the corresponding labelled nucleotides cannot compete with those of single dye molecules.\textsuperscript{30} Therefore, to address such limitations a number of strategies have been proposed to improve fluorescence detection, such as fluorescence life-time\textsuperscript{35} and radio frequency modulation.\textsuperscript{36} A promising technology for DNA sequencing, recently proposed, is called pulsed multiline excitation (PME).\textsuperscript{30} This new technology has been applied to capillary electrophoresis for DNA sequencing and is based on the use of sequential laser pulses for the excitation of labelled DNA fragments being separated by CE (capillary electrophoresis). The advantage of this technique is that the emission signals are stronger as the absorption maxima of the dyes match the excitation sources and the temporal separation of the laser pulses eliminate any spectral overlap between fluorophores. Furthermore, stronger signal
intensities are obtained as no dispersing elements (e.g. prisms) are needed for colour separation.

This technique has a great potential for DNA sequencing as it enables the use of lower concentrations of material (because of its high sensitivity) and the simultaneous imaging of high density capillary-arrays and microarrays.

**Figure 1.7 ET primers for DNA sequencing**

1.3.2.3  **Thermostable Enzymes & Polymerase Chain Extension**

Another important advance in DNA sequencing has been the introduction of thermostable DNA polymerases and their use in polymerase chain extension (PCR) reactions.\(^\text{37}\)

PCR is a technique developed by Mullis in 1983 based on the use of thermostable DNA polymerases (such as Taq polymerase, isolated from *Thermus aquaticus*)\(^\text{38}\) for the amplification of DNA.\(^\text{39}\) During PCR cycles, double stranded DNA is alternately heated to a high temperature (~90 °C) and cooled down in order to separate the two strands by denaturation which then enables the DNA polymerase to use them as templates for the synthesis of new complementary DNA strands. With PCR it is possible to amplify one or few copies of DNA to generate millions of copies, allowing thus the analysis of very small amounts of DNA.

1.4  **New High Throughput DNA Sequencing technologies**

With the completion of the HGP, more than 150 billion bp of sequence information was produced.\(^\text{29}\) The availability of this large amount of data paved the way to a new era, the so called “post-genomic era” that sees as main objective a deep
understanding and exploration of the complete human genome for application in pharmacogenomics, disease-gene linkage, gene function studies and finally clinical medicine. Additionally, de novo sequencing and re-sequencing of genomes are required for application in comparative genomics, evolution and epidemiology. There is also a strong need for the identification of all the genetic variations that are related to disease susceptibility and phenotypic diversity. The most common DNA sequence variation is known as Single Nucleotide Polymorphism (SNP) and occurs when a single nucleotide varies among members of a species, affecting their response to diseases, pathogens, drugs, chemicals, etc. 

In response to the need for ultra high throughput DNA sequencing methods, numerous strategies and technologies have been proposed, to reduce the cost of sequencing and to increase the throughput (Table 1.1). They include nanopore sequencing, sequencing by hybridisation and sequencing by synthesis technologies.

Most of these technologies use new ultra-sensitive detection methods (e.g in single molecule DNA sequencing), microfluidic technologies and miniaturised arrays, which contributed to lower the costs and increase the quality of data generated by these new technologies.

These techniques and the most recent DNA sequencing technologies are discussed below.
Table 1.1 New generations of DNA sequencing technologies. Table adapted from ref.544

<table>
<thead>
<tr>
<th>Company</th>
<th>Platform &amp; Feature Generation</th>
<th>Cost per Megabase</th>
<th>Cost per instrument</th>
<th>Read-length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche-454</td>
<td>SBS (prosequencing)</td>
<td>$60</td>
<td>$500 000</td>
<td>250 bp</td>
</tr>
<tr>
<td></td>
<td>Emulsion PCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illumina-Solexa</td>
<td>SBS (reversible terminators)</td>
<td>$2</td>
<td>$430 000</td>
<td>36 bp</td>
</tr>
<tr>
<td></td>
<td>Bridge PCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helicos</td>
<td>SBS-single molecule</td>
<td>$1</td>
<td>$1 350 000</td>
<td>30 bp</td>
</tr>
<tr>
<td>Applied Biosystems</td>
<td>SBS (Ligase-octamers with two-base encoding) │ $2</td>
<td>$591 000</td>
<td>35 bp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Emulsion PCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dover</td>
<td>SBS (Ligase-nonamers)</td>
<td>$1</td>
<td>$155 000</td>
<td>13 bp</td>
</tr>
<tr>
<td></td>
<td>Emulsion PCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Biosciences</td>
<td>Real time SBS-single molecule</td>
<td>-</td>
<td>Under development</td>
<td>&gt;1000 bp*</td>
</tr>
<tr>
<td>Visigen</td>
<td>Real time single molecule</td>
<td>-</td>
<td>Under development</td>
<td>&gt;1000 bp*</td>
</tr>
<tr>
<td>Intelligent Bio-</td>
<td>SBS (reversible terminators)</td>
<td>-</td>
<td>Under development</td>
<td></td>
</tr>
<tr>
<td>systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* expected

1.5 Array-based Cyclic Sequencing Technologies

The need for ultra-high-throughput DNA sequencing technologies encouraged researcher to develop non-electrophoretic-platforms in order to carry out genome analysis in an automated and highly parallel manner.

Array-based sequencing technologies consist of repeated cycles of sequencing performed on an array of spatially separated DNA molecules. The array features are either ordered (DNA microarrays) or randomly dispersed (bead arrays). Each cycle consist in interrogating one or a few bases in parallel for a large number of oligonucleotides (thousands to billions). 45

1.5.1 Solid Phase Arrays for DNA Sequencing

Currently, solid-phase array sequencing devices are used for the production of high density DNA platforms for DNA sequencing applications as they enable scientists
to obtain a large amount of data, using small amount of reagents and miniaturised systems. Typically, DNA probes or targets are immobilised either onto planar surfaces made of glass/polymers (e.g Illumina sequencing technology) or onto arrays of beads located in the microwells of a glass or fiber optic slide (e.g 454 Life Science Picoliter Sequencing System).

1.5.1.1 Bead Arrays

Bead arrays have found applications as highly parallel platforms for DNA cloning in a number of modern DNA sequencing technologies. Clonal beads are discrete colonies of PCR amplicons (PCR-amplified DNA) immobilised onto solid surfaces, such as beads ordered on a patterned planar surface. Polonies can be obtained by solid-phase PCR in polyacrylamide gels, BEAMing (beads, emulsion, amplification, magnetics)-based cloning, or Megaclone technology. All these methods developed for DNA amplification have been used in combination with different sequencing technologies (sequencing by synthesis and sequencing by ligation).

BEAMing PCR (so-called emulsion PCR) is a clonal amplification technique that has been adopted for DNA sequencing platforms, such as 454, Polonator and SOLiD, to amplify sequencing features (these technologies are described in detail in the section 1.5.3). In this technique, magnetic beads coated with streptavidin are bound to oligonucleotides modified at their 5' position with biotin (Figure 1.8). An aqueous mixture containing oligonucleotide-bead complexes, DNA templates, nucleotides and DNA polymerase, is stirred with an emulsifier/oil mix to create a microemulsion. The emulsion is generated such that each droplet or “compartment” contains a bead. The emulsion is then aliquoted into the wells of a 96-well PCR plate and thermal cycled to amplify the DNA templates. In each droplet, where there are a bead and a DNA template, the bead-bound oligonucleotide acts as a primer during the amplification process. After breaking the emulsion, the DNA strands are denatured and the beads captured into the wells of a picotiter plate (e.g 454 sequencing technology) or immobilised on a planar surface to produce high density “random” bead arrays (SOLiD and Polonator). Because each bead can have up to 1 million copies of each DNA template, the signal-to-noise ratio of enzymatic assays (e.g sequencing by synthesis cycles) carried out on these beads is very high.
Another important advantage of this technique is the ability to simultaneously sequence hundreds of thousands of DNA sequences immobilised on microbeads.

The megaclone technology has been adopted in Massively Parallel Signature Sequencing (MPSS). In this approach, DNA templates are conjugated by ligation to 32-mer tags and amplified by PCR to generate thousands of copies of each DNA-tag “hybrid”. Each DNA molecule has a unique tag, complementary to anti-tags sequences attached to microbeads. When amplified DNA and microbeads are mixed, the PCR products are captured by the magnetic beads bearing oligonucleotide sequences complementary to the tag in order to have ~100,000 identical DNA molecules covalently attached to each bead. One million microbeads are then loaded in a specially designed flow cell to form a packed monolayer of beads and used as platform for DNA sequencing by ligation (section 1.5.3.1).

Bead-array-based technologies have also been developed by Illumina for application in DNA genotyping. The arrays are constructed by self assembly of 3 μm silica beads in the microwells of fiber optic chips or silica slides and can reach a density of 40,000 array elements per square millimetre. Hundreds of thousands of oligonucleotides are attached to each bead of the array acting as universal capture sequences in multiplexed genotyping assays. Because the arrays are assembled
randomly, each bead and sequence is identified using a decoding system based on sequential hybridisations to fluorescently labelled oligonucleotides complementary to the capture sequences.\textsuperscript{54}

1.5.2 **Sequencing by Hybridisation**

DNA sequencing by hybridisation consists in using an array of short oligonucleotide probes complementary to different regions of an unknown DNA fragment to reconstruct its sequence from the hybridisation pattern.\textsuperscript{55-57} Two different approaches have been reported. One approach consists of immobilising the DNA to be sequenced onto a surface (membrane or glass chip) and hybridising it to small labelled probes. An alternative strategy consists of hybridising genomic DNA to arrays of immobilised oligonucleotide probes. The binding of one oligonucleotide probe to its complementary strand is sensitive to a single mismatch. Therefore the occurrence of a sequence of nucleotides in a DNA target can be interrogated by using small probes that differ by only one base (A, C, T, G) (Figure 1.9). Only one of the four probes matches fully the target sequence giving the strongest hybridisation signal. In this way, by reading the target sequence with a full set of overlapping probes, it is possible to determine each position in the target. Each region of the array is assigned to a set of probes, complementary to a specific region of the genome to resequence. The hybridisation pattern obtained from the set of overlapping probes is then used to reconstruct the sequence of the original DNA target using mathematical algorithms.

An impressive amount of sequence information can be obtained with this method (>10\textsuperscript{9} bases). Therefore sequencing by hybridisation has found application in massive genome resequencing for the investigation of single-nucleotide polymorphisms. It has been also applied to *de novo* sequencing; however this application still remains a challenge because of the low read-length (limited by the probe length) and the complexity of the method.\textsuperscript{55}
A new high-throughput sequencing-by-hybridisation technology, called 'Shotgun SBH' has been proposed by Pilhak in 2008. This technique overcomes some of the problems of the classic sequencing by hybridisation method, proposing the use of highly parallel sequencing platforms (10 millions sequencing features per square centimetre) containing a library of DNA target colonies (200 bp) obtained by rolling circle amplification. Resequencing is carried out using small universal probes compatible with any genome. Firstly, genomic DNA is fragmented and converted in circular molecules containing a 50-mer universal linker used to hybridise the target DNA to primers immobilised on a surface. The target DNA is amplified by rolling circle PCR generating sub-micrometer structures which can be easily detected using sequence-specific fluorescent probes. A set of 582 fluorescent pentamers (containing locked nucleic acid monomers to increase the melting temperatures) is then sequentially added to hybridise the array of DNA target on the glass slide. At the end of the process, the full spectrum of overlapping DNA fragments is compared to the genome of reference to find genetic variations. The authors reported that, using this approach, 97.3 % of single mutations could be detected in a test genome. This high throughput platform is a cost-effective,
accurate and highly parallel technique for resequencing applications, when a reference genome already exists.

1.5.3 Sequencing by Synthesis

Despite differences in sequencing biochemistry, the work flow of the so-called “second generation DNA sequencing technologies” (454, Polonator, SOLiD, Illumina Genome Analyzer, etc.) consist of three main steps:

1. DNA fragmentation and modification (e.g. \textit{in vitro} ligation of sequence adaptors)
2. DNA amplification
3. array sequencing (e.g. extension of primers by polymerases or ligases and imaging)

These second-generation cyclic array sequencing technologies are all based on the principle of sequencing by synthesis, although primer extension can be performed either by a ligase (Polonator and SOLiD) or by a polymerase (454, Illumina).

1.5.3.1 Sequencing by Ligation

\textit{Multiplex Polony Sequencing}

The method, developed by Church, is based on the amplification of DNA molecules on paramagnetic beads using emulsion PCR\textsuperscript{52} and on the principle of four-colour sequencing by ligation.\textsuperscript{50} Two platforms, SOLiD and Polonator, are based on this system. In the SOLiD platform, commercialised by Applied Biosystems, a two-base encoding system is used in sequencing-by-ligation cycles. In this approach, the beads bearing amplified DNA are immobilised on a planar surface to generate a high density random array. DNA sequences immobilised onto the beads are hybridised to an “anchor” primer that is ligated enzymatically to a population of degenerate octamers labelled with fluorescent dyes such that the identity of two positions (e.g. the first two bases) of the octamers, which are complementary to the query positions, correspond to the identity of one particular fluorophore (Figure 1.10). After those particular positions have been identified by four-colour imaging, the octamer is chemically cleaved to remove the fluorophore and the primer ligated
to another octamer to query other two positions. After the primer has been extended by ligation, the complex anchor primer-octamer is stripped by denaturation and the cycle restarts using a different mixture of octamers, with different positions encoded by the fluorophore, (or another primer that is set back one or few positions from the ligation site) to interrogate other positions. Sequencing data are acquired using a "colour space" mapping system to identify the colour and the position of each bead and consequently the two adjacent bases in the ligated octamer, which were encoded by the colour.

ABI's SOLiD system was introduced on the market in 2007. It has a read-length of 25-35 bp and generates 2-4 Gb per run of DNA sequencing data. The Polonator was launched in 2008 and is currently the most competitive technology in terms of instrument price (Table 1.1). However, the Polonator system is limited by the current read-length (13 bp).

![Figure 1.10 Microbeads generated by emulsion PCR and used in the sequencing method developed by Applied Biosystems (SOLiD). Figure adapted from ref.](ref.61)
Massive Parallel Signature Sequencing

Massive Parallel Signature Sequencing (MPSS), developed by Brenner, is a technology based on the Megaclone technique for the preparation of DNA samples and on cycles of ligation for DNA sequencing. In this method, two million megaclone beads are loaded through the inlet of a flow cell and packed in a planar array so that each bead can be identified individually during sequencing. Like the sequencing method described in the previous section, this approach is based on sequencing by ligation. However, at the end of each cycle, the ligated sequences are cleaved using restriction endonucleases rather than being stripped by denaturation. In this approach, a short adapter containing the recognition sequence for a restriction endonuclease, is ligated to the end of each cDNA on the beads (Figure 1.11). When the restriction enzyme is added to the beads, it binds to the recognition sequence and it cuts another region of the DNA molecule leaving a four base pair overhang. In a ligation cycle, encoded adaptors are used to identify these four bases, one at a time. A set of 1024 encoded adaptors is ligated to the cDNA on the beads such that each microbead has an equal number of four different adaptors, each encoding a different position of the four-base overhang.
The encoded adapters are designed to contain a recognition region for the restriction enzyme (in order to repeat the cycle) and a sequence complementary to fluorescently labelled "decoder" probes used to sequentially identify the desired four-base sequence with an accuracy of 99.9%.

This method has found application mainly in genotyping and gene expression analysis because of the short read-length (20 bp). De novo sequencing application would require an increase in the read-length of the method.\textsuperscript{55}

1.5.3.2 \textit{454 Pyrosequencing}

454 Sequencing technology is a highly parallel sequencing method commercialised by 454 LifeScience Corporation. This technology uses the BEAMing-based cloning technique (described in the section 1.5.1.1) for the preparation of DNA samples and the pyrosequencing approach to generate sequence data.\textsuperscript{47}
Following amplification of DNA by emulsion PCR, DNA-carrying microbeads are loaded on a plate made of $1.6 \times 10^6$ wells, which can be thought as picolitre-scale sequencing reactors (Figure 1.12). The enzymes required for pyrosequencing (ATP-sulfurylase and luciferase) are immobilised on smaller beads.

The sequencing reactions are carried out by delivering the reagents (buffers and nucleotides) into the flow chamber where the PicoTiterPlate is positioned. The four nucleotides are added sequentially over the slide. When a nucleotide is incorporated by a DNA polymerase, the pyrophosphate released is converted to ATP by an ATP-sulfurylase, providing energy to the enzyme luciferase to oxidise the substrate luciferin and to generate light (Figure 1.12). The light generated during each incorporation step, is then recorded by the CCD (charge-coupled device) camera of the instrument in order to detect the DNA sequences in which that particular nucleotide has been incorporated.

The key advantage of the 454 system over modern technologies is in its read length (200-300 bp), throughput and accuracy. However the cost of sequencing is much
higher than that of other platforms (SOLiD and Solexa). The 454 technology is currently the method of choice for the analysis of long reads (e.g. *de novo* assembly).\(^5\)

1.5.3.3  *Fluorescent in situ sequencing*

Another massively parallel sequencing-by-synthesis approach is fluorescent *in situ* sequencing (FISSEQ) developed in Church's lab and licensed to Agencourt Bioscience.\(^5\) In this approach DNA molecules are amplified on microbeads embedded in the polymer matrix of a slide. DNA fragments are sequenced by serial additions of a single fluorescent nucleotide (Figure 1.13) that is incorporated in DNA strands by a DNA polymerase. After each incorporation step, the slide is imaged using CCD optics and the fluorophore is removed either by photocleavage, or chemically from each incorporated nucleotide to allow the next sequencing cycle. The advantage of this approach is that it is inexpensive and has a low error rate (<1/3.3 million bases). It can be used for resequencing and *de novo* sequencing. This approach, like pyrosequencing, is based on the principle of *Single Nucleotide Addition* (SNA).\(^43\) Therefore the major limitation is the analysis of homopolymeric sequences, in which the same base is repeated many times (e.g. AAAA). Because the 3'-OH of the nucleotides used in SNA is not protected, multiple incorporations occur when four nucleotides are added together. To avoid this, in SNA, nucleotides are added sequentially, one at a time. However, when a sequence contains repeats, the same nucleotide can be incorporated more than once, making the sequence analysis difficult.

![Figure 1.13 Structure of Cy5-SS-dCTP](image-url)
To overcome this limitation, nucleotides protected at their 3' position have been designed for DNA sequencing by cyclic reversible termination.

1.5.3.4 Cyclic Reversible Termination

*Cyclic Reversible Termination* is a DNA sequencing method based on cycles of primer extension, imaging and cleavage of a fluorescent reporter group, on a solid surface.\textsuperscript{29,43}

In 2007 Illumina launched on the market a DNA sequencer based on this approach. Illumina genetic analysis technology is based on solid-phase DNA amplification methods\textsuperscript{62} and on the use of reversible terminators (section 1.5.4).\textsuperscript{46,61,63} In this approach, DNA strands are attached onto a planar surface and amplified in order to create a high-density array made of 50 million colonies of DNA molecules (each colony is made of 1000 copies of DNA) (Figure 1.14).

Polymerase reaction is carried out in the presence of four labelled reversible terminator nucleotides with the 3' position blocked to avoid multiple incorporations (Figure 1.16). After imaging the slide to identify the incorporated nucleotides, the fluorophore and the protecting group are removed chemically to allow the next cycle to occur (Figure 1.15).

---

*Figure 1.14* Bridge PCR technique used for the amplification of DNA: a) genomic DNA is fragmented and ligated to adaptors at both ends (in orange and blue) and then attached to a planar surface; b) and c) unlabelled nucleotides are used to amplify each sequence; d) and e) double stranded DNA is denatured and clusters (or colonies) of DNA sequences are formed on the surface. Adapted from ref.\textsuperscript{5}
This technology has a shorter sequence read-length (~35 bp) than capillary-based methods (>700 bp). Therefore, to date it has been applied mainly to the analysis of genetic variations and re-sequencing rather than whole genome de novo sequencing. However, the low sequencing cost ($2 per megabase) and throughput (up to 90 million bases analysed in one run) make this technology one of the most
competitive DNA sequencing platforms. In addition, the use of reversible terminators allows for the analysis of homopolymeric repeats that are difficult to analyse with other methods (e.g. the pyrosequencing technology developed by 454 Life Science). Another advantage of cyclic reversible termination methods over single nucleotide addition methods is the possibility of using four different nucleotides simultaneously rather than four separate steps and thus reducing the number of cycles required for sequence analysis.

Because the read-length of cyclic reversible termination technologies is determined by the efficiency of incorporation and deprotection steps, many research groups have focused on the optimisation of the chemistry of reversible terminators.

Recent advances in the chemistry of reversible terminators are described in the section 1.5.4.

1.5.4 Reversible Terminators

The first example of a reversible terminator was reported by Metzker in 1994. Since then, a number of fluorescently labelled reversible terminators with cleavable linkers have been developed by academic groups and biotech companies for application in DNA sequencing by synthesis.

Typically, a reversible terminator consists of a nucleotide with a small group capping the 3'-OH and a cleavable group linking the fluorophore to the base (generally at the 5 position of pyrimidines and at the 7 position of purines) (Figure 1.16).

Figure 1.16 Structure of a reversible terminator with a protecting group in 3'-position and a cleavable linker connecting the fluorophore (orange star) to the base.
During a polymerase reaction, the reversible terminator is incorporated into the growing DNA strand causing temporary termination of the DNA synthesis. After identification of the incorporated nucleotide (e.g. by optical imaging), the reporter and the capping groups are removed and another polymerase reaction is allowed to proceed with another nucleotide analogue.

Ideal reversible terminators should be recognised by a DNA polymerase, efficiently incorporated into DNA and rapidly cleaved to release the fluorophore. It is also important that the cleavage is complete to avoid residual fluorescent signals in the extended primer that would interfere with the next incorporation and detection steps. Cleavage of the linker and deprotection should be carried out in mild conditions compatible with the stability of DNA. Cleavable groups used to date for the synthesis of reversible terminators include disulfide bonds, allyl groups, azidomethyl groups and photo-cleavable 2-nitro-benzyl linkers (sections 1.5.4.1 and 1.5.4.2). Several small cleavable groups have been described for the protection of the 3' hydroxyl group. Metzker reported that nucleotides protected at the 3' position with groups containing ester linkages (acetyl or benzoyl) were not incorporated by any polymerase. On the contrary, small protecting groups containing ether linkages (e.g. allyl, methyl, nitrobenzyl, tetrahydropyranyl) have proved to be well tolerated by some DNA polymerases.

Another class of reversible terminators, described by Burgess, includes nucleotides labelled with fluorescent molecules attached to the 3' position through a cleavable linker. The advantage of using these nucleotides is the fact that the fluorescent cleavable reporter group is used to block the 3' position and therefore only one deprotection step is required after incorporation. However, this second type of nucleotide has been less exploited by researchers as polymerases do not tolerate modifications at the 3' position with bulky groups.

The reversible terminators described in the literature include chemically cleavable and photocleavable linkers. The properties of these two classes of reversible terminators are discussed in the sections 1.5.4.1 and 1.5.4.2.
1.5.4.1 Chemically Cleavable Reversible Terminators

In 2003 Church et al. proposed the use of a disulphide group as a chemically cleavable linker for reversibly dye-labelled nucleotides for application in fluorescent in situ sequencing on polymerase colonies (FISSEQ) (Figure 1.13).\(^{51}\) The nucleotides were incorporated by a DNA polymerase in DNA strands immobilised on a surface; after imaging the slide, the linker was cleaved using a buffer solution containing β-mercaptoethanol. 26-34 consecutive cycles of sequencing by synthesis were performed on a solid surface. Like the pyrosequencing method, fluorescent in situ sequencing is limited in the analysis of repeated bases (e.g. AAAA) as the hydroxyl group in the 3' position is not blocked. In addition, the nucleotides must be added sequentially to avoid multiple incorporations. In 2008 Turcatti et al. showed that reversible terminators bearing a disulfide linker could be used in 10 consecutives cycles of SBS. They also reported that, with an appropriate choice of the fluorophore, these nucleotide analogues could terminate the extension without the need for a blocking group at the 3' position.\(^{66}\)

Linkers containing disulfide bridges have the advantage of being cleaved in relatively mild conditions, using a thiol or a water soluble trialkylphosphine. The main limitation of this cleavable group is the fact that it can be reversed or destabilised in particular reaction conditions.\(^{68-70}\)

An alternative chemically cleavable linker has been reported by Ju et al. at the Columbia Genome Center. A full set of four fluorescent 3'-O-allyl-dNTPs-allyl-fluorophore has been synthesized by protecting the 3'-OH with an allyl group and linking the fluorophore to the base through an allyl moiety (Figure 1.17). The allyl group and the fluorophore were removed in one step in 30 s by Pd-catalysed deallylation.\(^{71}\) These nucleotides were efficiently incorporated by a DNA polymerase in 20 consecutive cycles of SBS on a microarray chip. Currently, this chemistry is being used in the sequencing platform of Intelligent Biosystems, a company co-founded by Ju and Gordon.\(^{44}\) They claim that this technology will reduce the cost of sequencing to $ 5000 by 2009 and to $ 1000 within the next five years.\(^{44}\)
The Genome Analyzer, commercialised by Illumina (see section 1.5.3.4), uses reversible terminators 3'-O-azidomethyl 2' deoxynucleoside triphosphates, each labelled with a different removable fluorophore (with a disulfide bridge). To improve the incorporation of these substrates, the DNA polymerase 9°Nm was modified.\textsuperscript{46}

An alternative method for removing the fluorescent signal after incorporation of the labelled nucleotide into a DNA strand has been proposed by Quake et al.\textsuperscript{73} This method consists in quenching the fluorescence of the fluorophore so that a new signal can be detected after incorporation of the next nucleotide. The main limitation to the application of this method is the fact that the photobleached fluorophore remains within the DNA template and therefore interferes with the next cycles of extension.

1.5.4.2 \textit{Photocleavable Reversible Terminators}

Photocleavable groups have been used as an alternative to chemically cleavable linkers for the synthesis of reversible terminators. The advantage is in the possibility of carrying out the cleavage in milder conditions and without the need for chemicals or for additional purifications. The nitrobenzyl group is cleaved within 10s by irradiation with light (>300 nm) after the identity of the incorporated nucleotide has been determined by fluorescence imaging.

The use of a photocleavable moiety for the synthesis of reversible terminators has been reported by many research groups.\textsuperscript{43} Ruparel et al. have described in 2005 the synthesis and application of 3'-O-allyl-dNTPs labelled with fluorophores connected to the bases through a nitrobenzyl linker (Figure 1.20).\textsuperscript{74} These reversible terminators required two deprotection steps in order to cleave the fluorophore by
irradiation at 340 nm to deprotect the 3' hydroxyl group by Pd-catalysed deallylation.

---

**Figure 1.20** Structure of the photocleavable reversible terminator 3'-O-allyl-dCTP-PC-Bodipy

### 1.5.5 Single Molecule DNA Sequencing

All DNA sequencing methods, described in the previous sections, require amplification of DNA to achieve the required signal during sequence analysis. To reduce the cost of sequencing and avoid bias (due to sequences that amplify poorly) and errors introduced during the amplification process, single-molecule DNA sequencing (SMS) technologies have been developed by a number of companies such as Helicos, Pacific Biosciences, Visigen and Genovoxx.

These new DNA sequencing platforms, so-called "third-generation DNA sequencing technologies" include sequencing by synthesis methods, nanopore sequencing, real-time electrical detection and transmission electron microscopy (TEM) of DNA synthesis.

In this section, SMS methods which are already commercially available (True Single-Molecule Sequencing, by Helicos) or that are expected to be available in the next few years (e.g. SMRT sequencing by Pacific Biosciences and FRET-based SMS by Visigen) are discussed.

#### 1.5.5.1 Real-Time Single Molecule Monitoring of DNA Synthesis

The true single molecule sequencing (tSMS) technology developed by Quake and commercialised by Helicos Biosciences, relies on the cyclic interrogation of DNA sequences (bearing poly-(dA) "tails") hybridised to universal primers (with poly-(dT) sequences) immobilised on a planar surface. DNA analysis is performed
by adding fluorescent nucleotides (one a time) and a DNA polymerase to the array. The position of each template on the array is determined by fluorescence imaging. After primer extension, the array is imaged and then the fluorophore cleaved to allow next extension.

This technology enables the analysis of billions of sequences per run (1h), providing a huge amount of data and sequence informations. However, this sequencing platform is still very expensive and limited in terms of read-length. Other technologies (e.g. SMRT), which are currently under development, seem more promising in terms of speed, cost and potential to analyse longer reads.\textsuperscript{75} Furthermore, like the pyrosequencing approach, the method developed by Helicos uses nucleotides that are not protected in the 3' position, therefore analysis of homopolymeric sequences is less accurate.

One of the most exciting and promising DNA sequencing technologies is the single-molecule real time sequencing (SMRT) technology, developed by Webb and Craighead at Cornell University and by Pacific Biosciences.\textsuperscript{78,79} This technology is based on two proprietary innovations: the use of fluorescent nucleotides labelled on the γ phosphate group and zero-mode waveguide chips. The nucleotide analogues used in this technology are labelled on the phosphate group rather then on the base (Figure 1.21). Therefore, when a nucleotide is incorporated in a polymerase reaction, the fluorophore is cleaved from the nucleotide and diffuses in the small reaction-chamber volume (20 \times 10^{-21} \text{liters}) where each strand is localised with a single molecule of a DNA polymerase. It is possible to detect each incorporation event in real time against the background fluorescence of the unincorporated nucleotides, which float in the un-illuminated part of the chamber. In fact each zero-mode waveguide (ZMW) is a cylindrical hole (few tens of nanometers in diameter) illuminated through a transparent surface by a laser light. The wavelength of the light is too big to pass through the waveguide’s aperture, therefore only the volume directly in contact with the transparent “floor” of the ZMWs is illuminated.
Figure 1.21 Fluorescent nucleotide used in SMRT technology

Pacific Biosciences claims that, by 2013, this technology will enable the analysis of the human genome in just 15 min. Thanks to its speed, high data quality and the ability to analyse long reads, this platform will be used in a broad range of applications, such as de novo sequencing and whole genome resequencing.75

A FRET-based single molecule sequencing technology is being currently developed by Visigen.75 Like the SMRT approach, fluorescent nucleotides labelled on the γ phosphate group are used to monitor DNA synthesis by a polymerase in real time. Each of the four nucleotides has a different acceptor fluorophore; the donor fluorophore is attached to the polymerase. During extension, energy transfer from the polymerase to the nucleotide, generates a fluorescence signal, which reveals the identity of the incorporated nucleotide. After incorporation, the pyrophosphate labelled with the dye is released in solution and the FRET signal quenched.

This “third generation sequencing method” is expected to analyse DNA with a speed of 1 million bases per second and a whole genome in less than one hour.

1.5.5.2 Nanopore Sequencing

Nanopore sequencing is a single molecule approach based on the measurement of fluctuations in the electric conductance of a nanopore as DNA passes through it.80 A nanopore can be represented by a protein pore in a membrane (biological pore) or by an artificial pore (solid state pore).75,81,82 When a nanopore is immersed in a conducting fluid and a voltage is applied, a current of ions is generated through the pore. Because the intensity of this current depends on the size of the pore, when a DNA molecule pass through it, the differences in the chemical and physical properties of the nucleotides are translated in changes of current.
This technique, introduced for the first time in 1989, is still at a proof-of-the-principle stage. The interest for this technique is mainly due to the possibility of sequencing single molecule DNA without the need for sample preparation and with high speed (PCR amplification or labelling). Significant advances in pore engineering will be necessary to identify single bases with a high accuracy.

1.6 Aims of the project

As shown in the previous sections, a variety of DNA sequencing technologies have been proposed as an alternative to the classic Sanger chain termination method to satisfy the demand for rapid, accurate and cheap DNA sequencing platforms.

In this context, sequencing by synthesis technologies based on cyclic reversible termination have proven to be so far among the most competitive methods in terms of throughput, accuracy and analysis of homopolymeric sequences. Current reversible terminator consist of chemically cleavable or photocleavable nucleotides and require the use of chemicals or irradiations for the cleavage which can result in damage to DNA. Therefore, there is a strong interest in developing new reversible terminators that have efficient incorporation and deprotection kinetics and that can be cleaved in mild conditions.

The aim of the project was to design and synthesise a new class of enzymatically cleavable nucleotides and to verify the feasibility of using them in primer extension reactions on a solid surface for application in DNA sequencing by synthesis.

To reach this objective, nucleoside triphosphate-peptide conjugates have been prepared and labelled with different fluorescent dyes in order to identify the base incorporated during primer extension by virtue of the colour. After detection, the fluorophore will be removed by enzymatic cleavage using a protease (Chapters 3 and 4).

In order to be able to interrogate a large number of DNA molecules in parallel, this method will be adapted to a solid phase platform, such as a DNA microarray (Chapter 5).

The application of dNTP-Peptide-Dye conjugates in DNA sequencing by synthesis offers several advantages. Firstly, all steps (incorporation and cleavage) are mediated by enzymes (polymerases and proteases), allowing for rapid, fast and
specific reactions. Additionally, the peptide moiety is relatively stable, easy to handle and can be easily modified to control each step.

The synthesis of fluorescently labelled peptides and nucleotides requires the availability of fluorescent dyes with different excitation and emission spectra. For this purpose, the cyanine dyes have been considered. To overcome the limitation of the high cost and lack of high-yielding protocols for their preparation, a practical method for the synthesis of cyanine dyes has been developed by combining solid phase chemistry and microwave heating (Chapter 2).
Development of a practical method for the synthesis of cyanine dyes

The purpose of this chapter is to provide a detailed description of the synthesis of fluorescent cyanine dyes for the labelling of enzymatically cleavable nucleotides.

2.1 Introduction

The development of a novel four-colour DNA sequencing technology, based on the detection of fluorescent dye-labelled nucleotides, requires the availability of highly fluorescent dyes with different excitation and emission spectra and chemical/physical stability. Ideal fluorophores for application in microarray-based DNA sequencing technologies must also be pH insensitive and display non-specific binding to biomolecules and surfaces in order to reduce the signal-to-noise ratio. Cyanine dyes have all these requirements and are widely used in fluorescent applications, such as microarray detection, DNA sequencing, flow cytometry and in vivo imaging. Therefore they were selected for the preparation of our fluorescent nucleotides. However, the main limitation to their use is the high cost (175 £.mg⁻¹) and the lack of practical and high-yielding synthetic methodologies.

The idea behind the work described in this chapter was to develop a practical and easy method for the synthesis of cyanine dyes spanning the whole colour range. Cyanine dyes of different colours and optical properties can be prepared by modulating the length of the polymethine chain (building block II) and the heterocycles at its ends (building blocks I and III), (Figure 2.1).

For the purposes of the project, the synthesis of trimethine dyes with $\lambda_{\text{max}} = 540$-600 nm and pentamethine dyes $\lambda_{\text{max}} = 640$-700 nm was investigated. The solid phase synthetic strategy was also applied to the preparation of heptamethine dyes, commonly used for in vivo imaging applications, and to sulfonated dyes, widely used for biomolecule labelling because of their water solubility and biocompatibility.
2.1.1 Cyanine dyes

The cyanine dyes are a family of highly fluorescent molecules belonging to the group of the polymethine dyes (Figure 2.2).\textsuperscript{91} They were first developed as spectral sensitisers for silver halide emulsions films for the photographic industry.\textsuperscript{92} Afterwards they have found application in optical data storage and as fluorescent labels for biomolecules.\textsuperscript{88,93,94} Cyanine dyes contain electron donor (D) and acceptor (A) groups connected by a polyene chain and can be classified in cyanine, hemicyanine or streptocyanine, depending on whether or not the N atoms are part of a ring (Table 2.1).\textsuperscript{91}

![Figure 2.1 General structure of cyanine dyes](image)

\[ R^1, R^2 = H, -(\text{CH} = \text{CH})_n, R^1 = \text{SO}^3; R^2 = H; \]
\[ n = 1, 2, 3. \]

![Figure 2.2 Generic structure of polymethine dyes](image)

The groups A and D are part of a highly conjugated system where the positive charge is delocalised, as shown by the two mesomeric structures 1a and 1b (Scheme 2.1). By varying the length of the polymethine bridge along with the terminal heteroaromatic rings, it is possible to tune their absorption profile. The absorption maxima of cyanine dyes increase, almost linearly, by 100 nm with \( n \) (so-called vinylene shift) while spectral fine-tuning is obtained by varying the substituents A and D.\textsuperscript{95}
The cyanine dyes have tunable wavelengths across the visible and near-Infrared (NIR) spectrum and intense colours, therefore they are ideal probes for highly sensitive multicolour detection. Their long fluorescence wavelengths are remote from the natural autofluorescence of biomolecules and their small molecular size is a desirable feature for fluorescent labels in order not to interfere with the system to be probed. In addition, cyanine dyes have excellent photophysical properties and narrow spectral widths.

<table>
<thead>
<tr>
<th>Group Name</th>
<th>General Formula</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyanine</td>
<td><img src="image" alt="Cyanine" /></td>
<td><img src="image" alt="Example" /></td>
</tr>
<tr>
<td>Hemicyanine</td>
<td><img src="image" alt="Hemicyanine" /></td>
<td><img src="image" alt="Example" /></td>
</tr>
<tr>
<td>Streptocyanine</td>
<td><img src="image" alt="Streptocyanine" /></td>
<td><img src="image" alt="Example" /></td>
</tr>
</tbody>
</table>

### Table 2.1 Groups of polymethylene dyes. Table adapted from ref.\(^9\)

#### 2.1.2 Synthesis of cyanine dyes

The first synthetic cyanine dye was obtained by Williams in 1856.\(^9\) However, it was only in 1912 that structure elucidation\(^9\) paved the way for a rational approach to the synthesis of polymethylene dyes.

Conventional synthetic methods for the preparation of unsymmetrical cyanine dyes are based on the condensation of two different heterocyclic quaternary salts 2 and 8 with a polyene-chain precursor such as orthoformate, malonaldehyde or vinylogous homologues of diphenyl formamidine 3, to give cyanine dyes with \(1 < n < 3\), Scheme 2.2.\(^9\)
Indolenine quaternary ammonium salts 2 (Scheme 2.2) undergo elimination of hydrogen halide HX, even in neutral conditions, to afford the stable Fisher's base, which behaves as nucleophile in the methylene base form.\textsuperscript{100} Generally, after \textit{in situ} deprotonation by a tertiary base, they react with an electrophilic polyene-precursor, activated by acetylation or with a Lewis acid, to give a hemicyanine intermediate 4. The isolation and purification of such intermediate is vital for the synthesis of unsymmetrical cyanine dyes, in order to avoid contamination of the final product 6 by hemicyanine 4 and the symmetrical derivative 5. The reaction can be easily monitored by UV spectrometry to reduce the formation of the symmetrical dyes. However, purification by chromatography is always required to afford unsymmetrical dyes in high purity and often is a difficult and time-consuming process since the impurities have similar properties to the desired products.
2.1.3 Solid phase synthesis of cyanine dyes

Fluorescent probes have attracted a large interest because of their high sensitivity and ease of handling, compared to radioactive labels.\textsuperscript{101} However, they are often difficult to synthesise or purify and consequently have high costs, when commercially available.

The use of solid supports avoids extensive work-up and chromatographic purifications, facilitates slow or difficult reactions by the ability to use high concentrations of reagents and allows for automation of the synthetic process.\textsuperscript{102} Combinatorial chemistry is widely applied in medicinal chemistry or chemical biology for the discovery of drug candidates or biologically active compounds and has been also applied to the discovery of dyes.\textsuperscript{103-106} Two solid-phase approaches have been recently reported in the literature for the synthesis of cyanine dyes (Scheme 2.3).

The first example was by Isacsson \textit{et al.} in 2001,\textsuperscript{107} and based on the loading of a carboxy-functionalised heterocycle onto the solid support 9 followed by condensation with a hemicyanine, such as 10, to give a trimethine dye 11, Scheme 2.3. The final product 12 was then isolated by acidic cleavage from the resin while the excess of hemicyanine was washed away from the resin in the previous step, avoiding purification of the final product by chromatography.

In 2005 Balasubramanian proposed an alternative solid-phase method for the preparation of cyanine dyes based on attack of the heterocyclic carbon nucleophiles, such as 14a, on polyene-chain precursors immobilised onto a solid support 13b to form unsymmetrical trimethine or pentamethine dyes, such as 16, Scheme 2.3.\textsuperscript{108} In this approach the hemicyanine intermediate 15c is prepared on a solid support. Therefore, libraries of cyanine dyes with different functional groups and properties can be easily synthesised without the need for isolation and purification of the hemicyanine intermediates.
However, the compounds described in this work were not suitable for the labelling of biomolecules since they did not have functional groups for conjugation. In addition, this method cannot be applied to the preparation of cyanine dyes bearing sulfonated groups on both heterocycles because of the low reactivity of the sulfonated heterocycles during the loading reaction step.

Therefore, there was still a need for the development of a facile and general synthetic method for the preparation of hydrophobic and hydrophilic cyanine dyes amenable to biomolecule conjugation and available in different colours.
2.2 Microwave-mediated synthesis of multicolour cyanine dyes

The main focus of the work described in this chapter was to develop a common synthetic pathway for the preparation of a wide-colour-range of hydrophobic and hydrophilic cyanine dyes, Scheme 2.4, suitable for bioconjugation, in large scale (≥ 100 mg scale) and without the need for chromatographic purification. The method described by Balasubramanian\textsuperscript{108} was optimised to allow the parallel synthesis of carboxyl functionalised cyanine dyes in a microwave, Scheme 2.4.

The synthetic procedure reported, herein starts with the \textit{N}-alkylation of indolenines (Section 2.2.1).

2.2.1 Synthesis of \textit{N}-alkyl quaternary ammonium salts

\textit{N}-alkyl substituted quaternary ammonium salts derived from benzothiazole, benzoazole, benzoselenazoles, indole and quinoline bearing an activated methyl group in 2-position are generally synthesised by heating the corresponding heteroaromatic base, at 100-120 °C, with an excess of alkylating agent such as, alkyl iodide or bromide (3-5 eq.) in aprotic solvents such as 1,2-dichlorobenzene and acetonitrile, Scheme 2.5.\textsuperscript{53,109,110} The \textit{N}-alkylation requires 1-3 days to be completed,\textsuperscript{111,112} depending on the solvent, alkylating agents and substituents on the aromatic ring.

Consequently, the molecular diversity of cyanine dyes and the combinatorial approach to the discovery of new fluorescent probes are limited by the long reaction times required for the synthesis of these precursors.
Scheme 2.4 Parallel synthesis of cyanine dyes
N-Alkylation of 23a-b with bromohexanoic acid or methyl iodide was carried out by refluxing the reaction mixture at 100-120 °C, in acetonitrile. Yields and reaction time are shown in Table 2.2.

During our work on the synthesis of cyanine dyes, we were interested in investigating whether microwave heating could lead to an improved protocol for the N-alkylation of indolenines, especially in terms of reaction time. Microwave heating and its application to the N-alkylation of indolenines are discussed in the sections 2.2.1.1 and 2.1.1.2.

Table 2.2 N-alkylation of indolenines. *Precipitation of the product from the reaction mixture was observed after 2h.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Compound</th>
<th>T (°C)</th>
<th>t</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14a</td>
<td>100-120</td>
<td>64h</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>14b</td>
<td>100-120</td>
<td>16h*</td>
<td>77</td>
</tr>
<tr>
<td>3</td>
<td>14c</td>
<td>100-120</td>
<td>16h</td>
<td>84</td>
</tr>
<tr>
<td>4</td>
<td>14d</td>
<td>100-120</td>
<td>48</td>
<td>72</td>
</tr>
</tbody>
</table>

2.2.1.1 Microwave heating

High-speed synthesis with microwaves was reported by Gedye and Giguere/Majetich in 1986.\textsuperscript{113,114} Microwaves are a form of electromagnetic radiations defined as having frequencies between 300 MHz and 300 GHz.\textsuperscript{115}
Irradiation of a sample, with microwave frequencies, results in the dipoles or ions aligning with the applied electric field. When the applied field oscillates, the dipoles or ion try to realign itself with an alternating electric field, the molecule thus try to continually “realign” itself and consequently, heat is generated through molecular friction and dielectric loss.\textsuperscript{116}

The conversion of electromagnetic energy into heat depends on the dielectric properties of a particular substance and its so-called $\tan\delta = \varepsilon''/\varepsilon'$, where $\varepsilon''$ is the dielectric loss, which depends on the efficiency with which the electromagnetic energy is converted into heat and $\varepsilon'$ is the dielectric constant.\textsuperscript{116}

Solvents can be classified according to their $\tan\delta$, as high ($\tan\delta > 0.5$), medium ($\tan\delta 0.1 - 0.5$), and low microwave absorbing ($\tan\delta < 0.1$). However, a low $\tan\delta$ value does not preclude the use of a particular solvent, if the reagents or substrates are polar, or if polar additives such as ionic liquids are added to the reaction mixture.

The advantage of microwave heating is that the sample is heated uniformly, whereas in the oil-heated vessels the solution in contact with the walls is heated first and unevenly.\textsuperscript{116}

2.2.1.2 \textit{N-Alkylation of indolenines and microwave heating}

The alkylation of indolenines with alkyl halides is a $S_N2$ reaction, which involves the displacement of the halide by the nucleophile (indolenine) and the formation of a quaternary ammonium salt.

\textit{N-Alkylation of indolenines 23a and 23b}, Figure 2.3, is expected to be accelerated by the use of microwave irradiations, since the reaction results in the formation of an ion pair, in addition to thermal microwave effects.\textsuperscript{115,117,118}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{indolenines.png}
\caption{Indolenines 23a and 23b}
\end{figure}

\begin{align*}
23a: & \quad R_1, R_2 = H \\
23b: & \quad R_1, R_2 = -(\text{CH} = \text{CH})_2
\end{align*}

\textit{SN2} reactions generally require the use of polar aprotic solvents (Table 2.3), since polar protic solvents stabilise the nucleophile. However, the use of water as a reaction media in conjunction with microwave irradiation is an emerging method for
the development of environmentally friendly organic reactions and has been applied to various organic reactions and to nucleophilic substitutions, therefore water was considered for our experiments.

In order to find the best conditions for the microwave-assisted alkylation of indolenines, polar aprotic solvents with high dielectric constant and loss factors were selected: dimethylformamide, and acetonitrile. DMSO was not tested, due to its high boiling point and the difficulty of removing it from the reaction mixture by evaporation.

Water, DMF and acetonitrile were tested for the alkylation of 2,3,3-trimethylindolenine 23a with bromo-hexanoic acid at 150 °C for 30 min, Figure 2.4. We found 150 °C to be the optimal temperature for this reaction since higher temperatures cause formation of unidentified by-products (HPLC analysis), while at lower temperatures the reaction rate was too low and consequently low conversion values were obtained for all solvents. As shown in Figure 2.4, the highest conversions were obtained in DMF and acetonitrile at 150 °C for 30 min.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Dielectric costant</th>
<th>tanδ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,4-Dioxane</td>
<td>2.3</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tetrahydrofuran (THF)</td>
<td>7.5</td>
<td>0.047</td>
</tr>
<tr>
<td>Dichloromethane (DCM)</td>
<td>9.1</td>
<td>0.042</td>
</tr>
<tr>
<td>Acetone</td>
<td>21</td>
<td>0.054</td>
</tr>
<tr>
<td>Acetonitrile (MeCN)</td>
<td>37</td>
<td>0.062</td>
</tr>
<tr>
<td>Dimethyl formamide (DMF)</td>
<td>38</td>
<td>0.161</td>
</tr>
<tr>
<td>Dimethylsulfoxide (DMSO)</td>
<td>47</td>
<td>0.825</td>
</tr>
<tr>
<td>1,2 Dichlorobenzene</td>
<td>2.8</td>
<td>0.280</td>
</tr>
</tbody>
</table>

Table 2.3 Dielectric costants and tanδ of polar aprotic solvents

These two solvents were further tested to push to completion the conversion of indolenine into the quaternary ammonium salt (Figure 2.5).

The alkylations were carried out in sealed microwave vials by applying 3 cycles of microwave irradiation, of 30 minutes each. The highest conversion (73%) was
obtained using acetonitrile as solvent. After prolonged heating unidentified by-products were observed when DMF was used as solvent, (Figure 2.5).

Figure 2.4 Microwave-assisted alkylation of 23a with bromohexanoic acid at 150 °C for 30 min in various solvents.

Alkylation of 2,3,3-tetramethylindolenine 23a with bromohexanoic acid at 150 °C for 1h, in acetonitrile, under microwave irradiation, afforded the quaternary salt 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (14c) in 66% yield. After evaporating acetonitrile in vacuo, the final compound 14c was precipitated with diethyl ether and then washed several times with diethyl ether and dichloromethane to remove starting materials.

Figure 2.5 Optimisation of the synthesis of 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (14c) under microwave irradiation

Alkylation of 2,3,3-tetramethylindolenine with methyl iodide was also performed under microwave irradiation. The use of microwaves allowed the preparation of the
desired product (14a) in only 30 minutes and in 94% yield, while the classical oil bath heating afforded the product in 84% yield and required 3 days of refluxing.

2.2.2 Preparation of polymer-bound polyene-chain precursors

2.2.2.1 Loading of tert-butyl phenylcarbamate and removal of the Boc group

The starting point for the preparation of the range of cyanine dyes was the loading of N-butyloxy carbonyl 4-hydroxy aniline onto chloromethyl resin (1 % DVB cross-linking, 75-150 μm, 1.75 mmol g⁻¹) by nucleophilic substitution (Scheme 2.6).¹⁰⁸,¹²⁰

![Scheme 2.6 Formation of 4-aminophenol PS: a) N-butyloxy carbonyl 4-hydroxy aniline, Cs₂CO₃, KI, acetone, reflux, 16h; b) 20% TFA in DCM](image)

The protocol reported by Balasubramanian was modified by refluxing the reaction mixture at 70 °C in acetone (instead of the reported heating at 50 °C in dimethylformamide). These reaction conditions were preferred since a more efficient stirring of the resin was expected to be obtained by reflux. In addition, the exchange of the chloride on the resin with iodide, known as the Finkelstein reaction,¹²¹ is more efficient in acetone where KCl and KI have different solubilities. While KI dissolves readily in acetone, the solubility of KCl is low, thus the equilibrium of the reaction is shifted by the precipitation of the insoluble salt. The reaction was monitored by a colorimetric test based on the reaction of 4-(4-nitrobenzyl) pyridine (NBP) with the chloromethyl groups on the polystyrene resin.¹²² Formation of the desired product was also confirmed by gel-phase NMR,¹²³ and elemental analysis (quantitative by N analysis, 96% by Cl analysis, 94% by I analysis).

Removal of the Boc protecting group to form 18 was carried out with TFA (20% in DCM) for 2h. Merrifield resin is stable to treatment with this moderately strong acid.¹²⁴ The completion of the deprotection was confirmed by gel-phase NMR, while the loading of the resin was determined by elemental analysis.
Formation of polymer-bound polyene-precursors \((n=1 \text{ and } n=3)\)

Conversion of 18 into the corresponding polyene-chain precursors 13 was achieved following the method proposed by Balasubramanian et al.\(^{108}\) The polymer-bound imidate ethyl 4-hydroxyphenylimidoformate \((n=1)\), intermediate of the trimethine dye synthesis, was obtained by reacting 18 with triethylorthoformate in the presence of \(\text{BF}_3\text{Et}_2\text{O}\) (Scheme 2.7). The Lewis acid was added first to a solution of triethylorthoformate in DCM. Afterwards the reaction mixture was added to 18 and left under agitation, at room temperature, for 6 hours in a solid phase extraction (SPE) cartridge. The reaction mixture was then neutralised with 10 % triethylamine in DCM.

The polymer-bound imidate \((n=3)\), intermediate of the pentamethine dye synthesis, was obtained by reacting 18 with 1,1,3,3-tetramethoxypropane, under the same reaction conditions (Scheme 2.7). The formation of polyene-precursors was confirmed by the change in the colour of the resin (orange for 13a and dark blue for 13b).

Imidate formation was also confirmed by a sharp characteristic peak at 1645 cm\(^{-1}\) (C=N stretch) in the FTIR spectrum and by gel-phase NMR.

\[
\begin{align*}
\text{H}_2\text{N} & \quad \text{c)} \\
\text{O} & \quad \text{N} \\
18 & \quad \text{R} = \text{OEt} \\
13a: & \quad \text{R} = \text{OEt} \\
13b: & \quad \text{R} = -\text{CH} = \text{CH-OMe}
\end{align*}
\]

Scheme 2.7 Formation of PS-polyene precursors: c) triethylorthoformate or 1,1,3,3-tetramethoxypropane, \(\text{BF}_3\text{Et}_2\text{O}, \text{DCM}, 6\text{h}\).

The reaction of formation of imidates is a variation of the solution-phase method proposed by Roberts.\(^{125}\) The postulated mechanism for the reaction in solution involves two steps: A-formation of diphenylformamidine (rapid) and B-reaction with another mole of orthoformate (slow) to give the N-phenylformimidate. Step B occurs only in the presence of acids, Scheme 2.8.
Scheme 2.8 Formation of imidate in solution

Scheme 2.9 Formation of imidates on solid support

On the solid phase, the nucleophilic attack of a second molecule of aniline does not occur, and the imidate $13a$ is thus directly obtained. The proposed mechanism for the formation of imidates on the solid phase is described in Scheme 2.9.

2.2.2.3 Formation of polymer-bound polyene-precursors ($n = 5$)

Further studies were carried out to apply this solid phase method to the synthesis of heptamethine cyanine dyes. In recent years, heptamethine dyes have attracted great interest as fluorescent probes for in vivo imaging applications as they have excellent penetrating ability through the tissues and absorption/emission spectra in the near infrared (NIR) region, where biomolecules and tissue autofluorescence is low.89,90
In our initial investigation, the polymer-bound polyene intermediate 13c was obtained by reacting 18 with glutaconaldehyde hydrochloride I in DCM, in the presence of BF$_3$Et$_2$O. However, 13c was obtained in only 45% yield (by N elemental analysis), probably due to low solubility of the amidino pentadienylium salt in the solvent. Moreover, it was difficult to wash away the excess of reagent from the resin. Neither DMF, neither the presence of a tertiary base could improve the solubility of the reagent, since they form complexes with BF$_3$. Hence, it was investigated the use of acetic anhydride and diisopropylethylamine (DIEA) to activate glutaconaldehyde hydrochloride I to the nucleophilic attack of 18, as shown in Scheme 2.10. It was expected that phenylacetamide would be eliminated in preference to the more basic resin-bound aniline, during the loading step. However, also in this case, the yield was still low (49% by N elemental analysis). The extent of reaction was unclear from gel-phase NMR but was confirmed by FTIR and by formation of the final compound in the last step (Section 2.2.4).

An alternative pathway to the preparation of polymer-bound unsymmetrical polyene-intermediates for heptamethine dyes, would be the use of precursors of glutaconaldehyde hydrochloride, such as 1,1,3,5,5-pentaethoxy pentane $^{126}$ in the presence of Lewis acids. However, such an intermediate is not commercially available and additional steps would be required for the preparation of supported imidate intermediates. Therefore, a catch-and-release strategy was considered for
the preparation of heptamethine dyes on a solid support. The description of this approach will be discussed in the section 2.3.

2.2.3 Formation of hemicyanines

2.2.3.1 Hemicyanines with \( n = 1 \) and \( n = 3 \)

Preparation of immobilised hemicyanines from precursors 13a-b was performed by condensation with quaternised indolenine derivative 14a-b at 120 °C for 15 min under microwave heating (Scheme 2.11), in a variation of the method originally proposed by Balasubramanian.\(^{108}\)

The formation of hemicyanines involves the reactivity of nucleophilic indoleninium salts 14 a-b with 13 a-b to form another ion pair where the positive charge is highly delocalised. The interaction of microwaves with the solvent (DMF) and the reagents determines an increment of the reaction rate due to thermal effects (dielectric heating, higher temperatures and increased diffusion).\(^{117,118}\)

\[
\begin{align*}
13a: & \quad R = OEt \\
13b: & \quad R = -CH=CH-OMe
\end{align*}
\]

\[
\begin{align*}
14a: & \quad R_1, R_2 = H \\
14b: & \quad R_1, R_2 = -(CH=CH)_2
\end{align*}
\]

\[
\begin{align*}
15a: & \quad n=1; R_1, R_2 = H \\
15b: & \quad n=1; R_1, R_2 = -(CH=CH)_2 \\
15c: & \quad n=2; R_1, R_2 = H \\
15d: & \quad n=2; R_1, R_2 = -(CH=CH)_2
\end{align*}
\]

Scheme 2.11 Formation of hemicyanines by microwave heating: a) 14a-b, 120 °C, \( \mu \), 15 min.

Formation of hemicyanines was confirmed by the disappearance of an IR peak at 1630 cm\(^{-1}\), indicative of the presence of the conjugated bond (C=N) in the imidate precursor. The polymer bound dimethine hemicyanine 15a was obtained in 66% yield (N elemental analysis), while the tetramethine hemicyanine 15c was obtained
in quantitative yield. This result was in agreement with the yields of trimethine 20a and pentamethine 20c dyes 68% and 94% respectively (yields of the cleavage step) (see section 2.2.4).

The use of microwave heating, during formation of the hemicyanine, leads to higher yields and purity of the final compound (68% yield for trimethine dye 20a, Figure 2.6) compared to the classical oil bath heating (20% yield), Figures 2.7-2.8.

Any symmetrical impurity obtained by nucleophilic attack of two equivalents of indolenium would be automatically cleaved from the resin and removed by simple washing, thus avoiding difficult separation by chromatography.

The use of microwaves instead of classical heating offers also the advantage of shorter reaction times (15 min at 120 °C instead of 4 h at 80 °C), allowing for easy automation of the synthetic process.

![Figure 2.6 Trimethine dye 20a](image)

![Figure 2.7 Effect of temperature/microwave heating on the yield of the trimethine dye 20a (cleaved from the resin in solution).](image)
2.2.3.2 Hemicyanine with \( n = 5 \)

Formation of the hemicyanine intermediate \( 15e \) (with \( n = 5 \)) was carried out by acetylating \( 13c \) in the presence of DIEA and acetic anhydride followed by reaction with \( 14a \) in DMF under microwave heating at 120 °C for 15 min. However, the final heptamethine dye was obtained in low yield (<10%). Additionally, by-products absorbing at lower wavelength (550-650nm) were also observed. It is known that NIR heptamethine dyes have poor chemical and photochemical stability, due to photooxidation. Interaction of an excited dye molecule with dioxygen can lead to the formation of singlet oxygen (\( ^1O_2 \)) and superoxide anion (\( O_2^- \)) which accelerate the photofading of dyes by oxidation. To avoid thermal degradation of the heptamethine dye, the formation of the hemicyanine intermediate was carried out at room temperature, in the presence of DIEA, and acetic anhydride to activate the polyene-precursor \( 13c \) to nucleophilic attack of the quaternised heterocycle \( 14a \) (Scheme 2.12). However, this alternative protocol did not give the final dye in high yields. In fact, reaction of the indolenium salt \( 14a \) with the intermediate \( 13c \) forms the desired supported hemicyanine \( 15e \) by loss of \( N \)-phenylacetamide, but is also likely to result in the release of the hemicyanine \( 19e \) in solution by cleavage from the resin and formation of the symmetrical dye, with consequent decrease in the yield of the final compound (Scheme 2.12).
Because the use of the intermediate 13c is not ideal for the solid phase synthesis of heptamethine dyes, an alternative strategy to afford these dyes in higher purity and yield was developed and a detailed description of this method is reported later in this chapter.

### 2.2.4 Formation of mono-carboxylated cyanine dyes

Purification of asymmetrical mono-reactive cyanine dyes synthesised in solution is a time-consuming process, because the formation of the desired compound is often complicated by the formation of impurities such as unreacted hemicyanine 19, symmetrical dyes 21 and 22 and unreacted heterocycles 14a,c which are difficult to remove by chromatography (Scheme 2.13).

![Diagram of formation of hemicyanine and by-products](image)

**Scheme 2.12** Formation of hemicyanine and by-products: a) Ac₂O, DIEA, pyridine; b) 14a.
Solid phase synthesis allowed the preparation of cyanine dyes by reaction of a nucleophilic heterocycle $14c-d$ with the hemicyanine intermediates $15a-f$, activated by acetylation in the presence of pyridine, DIEA and acetic anhydride, Scheme 2.13. The reaction was carried out at room temperature for 1-2h and the fluorescent dyes were cleaved from the resin directly in solution, allowing removal of the unreacted hemicyanine $15a-f$ by filtration and avoiding the presence of symmetrical dye such as $21$.

Formation of $22$ was prevented by ensuring that hemicyanines intermediates $15a-f$ were formed in good yields from the imidates $13a-c$ and using substoichiometric amounts of heterocycles $14c-d$ in the last step. Any unreacted heterocycle was removed from the reaction mixture by aqueous washings while the desired product

\[
\begin{align*}
\text{Scheme 2.13 Synthesis of trimethine dye in solution:} \\
a) & \text{Ac}_2\text{O, AcOH, reflux, 2h;} \\
b) & \text{Ac}_2\text{O, pyridine, 80 °C, 1h.}
\end{align*}
\]
Scheme 2.14 Cleavage of cyanine dyes from the resin: a)14c-d, Ac₂O, DIEA, pyridine

20a-f was extracted into DCM. The fluorescent dyes were then precipitated and washed with diethyl ether, removing any residual pyridine.

This synthetic pathway allowed the preparation of trimethine and pentamethine fluorescent cyanine dyes in good yield (Table 2.5), purity (Figures 2.9-2.10) and on a 100 mg scale (~15 mmol). Further purification by column chromatography was carried out, when necessary, and only to ensure full characterisation of all the dyes.

As already mentioned, the heptamethine dye 20e (n = 5) was obtained in low yield and purity. In fact, the low loading level of the hemicyanine 15e meant there was a higher percentage of unreacted heterocycle 14a in the last step.

Various attempts were carried out in order to increase the yield by varying the equivalents of heterocycle and acetic anhydride and the reaction temperature during hemicyanine formation and cleavage. However no significant improvement was observed.

The synthetic protocol for the formation of heptamethine dyes required further improvement in order to obtain the final product in higher yield and purity and avoid thus further purification. More details will be given in section 2.3.
Figure 2.9 HPLC analysis of the crude pentamethine dye 20c. Detection: ELSD and at 650 nm

Figure 2.10 $^1$H NMR of the pentamethine dye 20c

2.3 Catch and release synthesis of heptamethine dyes

Heptamethine dyes synthesised by stepwise condensation of nucleophilic heterocycles with polystyrene bound glutaconaldehyde were obtained in poor yield and low purity, as previously reported in the section 2.2.4.
This limitation led to the development of an alternative catch-and-release-strategy, see Scheme 2.15, based on the formation of the hemicyanine intermediate in solution 19e and on its loading onto the solid support to form a polymer-bound hemicyanine 15e. This procedure allowed the separation of the final dye 20e from the hemicyanine by simple filtration.

Balasubramanian et al. reported a similar strategy, which involves formation of hemicyanine in solution, activation with sulfonyl chloride resin and then nucleophilic attack of a nucleophilic heterocycle to give the desired dye (Scheme 2.16). The advantage of the protocol presented herein is the possibility to use a common polymer-bound precursor, such as 18, to synthesise all the range of cyanine dyes, including heptamethine dyes. The hemicyanine 19e was synthesised by firstly activating glutaconaldehyde dianilido hydrochloride to the nucleophilic attack of a quaternary indoleninium salt 14a by treatment with a mixture 1:1 of acetic anhydride and acetic acid, Scheme 2.15. The reaction was monitored by UV spectrometry, in order to avoid or reduce the formation of symmetrical dye. After removal of acetic acid and acetic anhydride in vacuo and removal of starting material by aqueous washings, the hemicyanine 19e was dissolved in dichloromethane and added to the polymer bound aniline 18. The reaction mixture was left under agitation for 1h at room temperature to allow the polymer bound aniline 18 to attack the hemicyanine and displace a molecule of acetanilide. It is
important, during this time, to avoid prolonged exposure of such intermediates to light.

\[
\begin{align*}
\text{MeO} & \quad \text{OMe} \\
\text{a} & \quad \text{MeO} \\
\text{OMe} & \quad \text{OMe}
\end{align*}
\]

Scheme 2.16 Catch and Release strategy described by Balasubramanian et al.\textsuperscript{129} Reagents and Conditions: a) (EtO)\textsubscript{3}CH, EtOH, 80 °, 2h; b) DIEA, DCM, rt, 4h; c) DIEA, pyridine, rt, 30 minutes.

Afterwards the resin was washed extensively with dichloromethane, DMF and methanol, to remove the unreacted hemicyanine and then dried \textit{in vacuo} to give the polymer bound hemicyanine 15e in 83% yield (nitrogen elemental analysis).

The desired dark-green dye was cleaved from the resin by nucleophilic attack of a second heterocycle 14c onto the hemicyanine in the presence of pyridine and acetic anhydride, Scheme 2.17. Following this protocol, the unsymmetrical heptamethine dye 20e (Cy7) was obtained in high yield (86%) and purity (≈95%), as determined by HPLC (Figure 2.11) and \textsuperscript{1}H NMR analysis (Figure 2.12), without the need for further purifications by chromatography.

\[
\begin{align*}
\text{Cy7} & \quad \text{OH} \\
\text{15e} & \quad \text{20e}
\end{align*}
\]

Scheme 2.17 Cleavage of heptamethine dye 20e in solution: a)14c, Ac\textsubscript{2}O, pyridine
The catch-and-release method was also successfully applied to the synthesis of the heptamethine dye 20f, Figure 2.13. This dye has excitation and emission wavelengths in the NIR region, above 800 nm and therefore ideally suited for multicolour imaging in tissues.

Figure 2.11 HPLC analysis of the crude heptamethine dye 20e. Detection: ELSD and at 750 nm.

Figure 2.12 $^1$H NMR of crude 20e (Cy7)
2.4 Synthesis of water soluble cyanine dyes

The use of microwave heating and solid phase chemistry was successfully applied to the preparation of hydrophobic cyanine dyes with excitation and emission spectra ranging from 547 to 807 nm. These fluorophores were prepared for the labelling of modified nucleotides to be used in our four-colour DNA sequencing technology. However, preliminary experiments carried out on microarray chips (see Chapter 5) showed that nucleotides labelled with hydrophobic cyanine dyes can bind in a non-specific manner to DNA giving misleading results.

The non-specific binding was probably caused by a combination of hydrophobic interactions with the DNA base-pairs and ionic binding to the negatively charged phosphate backbone. The presence of negative charges onto the fluorophore should reduce this phenomenon. Therefore, water soluble cyanine dyes bearing a sulfonate group on the heterocyclic moieties were synthesised. In order to prove the principle, two different colours were initially needed. The cyanine dyes Cy3 and Cy5 absorbing at 550 nm and 650 nm, respectively, were ideal fluorophores since modern microarray readers can detect both dyes and distinguish between them.

These compounds are generally synthesised by following the procedure described by Waggoner et al.
2.4.1 \textit{N}-Alkylation of sulfonated indolenines

The use of microwave heating for the alkylation of indolenines proved to be very useful for the preparation of quaternary salts 26a-b from 2,3,3-trimethyl-3H-indole-5-sulfonate 25, see Scheme 2.18.

These compounds are important starting material for the preparation of water soluble cyanine dyes, widely used for biomolecules labelling. Sulfonated indolenines are less reactive than the non-sulfonated, because of the electron-withdrawing properties of the sulfonate groups. Consequently, longer reaction times are often required to afford the desired products in good yield. Classical oil bath heating usually requires 48h to ensure completion of \textit{N}-alkylation.\cite{1882}

Scheme 2.18 Synthesis and alkylation of sulfonated indolenine: a) 3-methyl-2-butanone, acetic acid, 3h, reflux; b) MeI, acetonitrile, 150 °C, MW, 60-90 min for 26a or 6-bromohexanoic acid, acetonitrile, 150 °C, MW, 3h for 26b

Compound 25 was prepared by conventional Fisher indole synthesis,\cite{1883,1884} by refluxing hydrazinobenzene sulfonic acid 24 with 3-methyl-2-butanone in acetic acid, for 3 hours. The cyclisation involves the formation of the phenylhydrazone of isopropyl methyl ketone followed by ring closure at the tertiary carbon to give 2-methylindolenines. Compound 25 was precipitated as the potassium salt with a saturated solution of KOH in 2-propanol and afterwards alkylated to give the quaternary salt 26. By applying microwave heating it was possible to reduce the reaction time for \textit{N}-alkylation of indolenines to 1-3 hours, Table 2.4. Acetonitrile, dimethylformamide (DMF) and 1,2 dichlorobenzene were tested as solvents for the microwave mediated \textit{N}-alkylation of 25 (Figure 2.14). The alkylations were carried out at 150 °C in sealed vials and the conversions measured by HPLC analysis. As expected, longer reaction times were necessary for the alkylation of sulfonated indolenines. The highest conversion (88%) was obtained by performing the reaction
in acetonitrile at 150 °C for three hours (three cycles of one hour each). Small amounts of DMF were also added to increase the solubility of the starting materials.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Compound</th>
<th>Temperature (°C)</th>
<th>Time (h)</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26a</td>
<td>150</td>
<td>1.5 (µl)</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150</td>
<td>24 (reflux)</td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>26b</td>
<td>150</td>
<td>3 (µl)</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150</td>
<td>48 (reflux)</td>
<td>89</td>
</tr>
</tbody>
</table>

Table 2.4 N-Alkylation of sulfonated indolenines

In order to avoid contamination of sulfonated dyes by various intermediates, it is important to ensure that the sulfonated heterocycles 26 are not contaminated by starting materials, that would be difficult to remove by chromatographic column and that could react with imidates or hemicyanines to form other undesired cyanine dyes. Methyl iodide and bromo-hexanoic acid can be easily removed by washings with diethylether. However, the removal of 25 is more challenging since sulfonated indolenine and alkylated products are both, water soluble. Removal of 25 from the crude product requires washings with 2-propanol/diethylether (1:1 v/v).

Figure 2.14 Microwave-assisted N-alkylation of compound 25 with 6-bromohexanoic acid.
2.4.1.1 \(N\)-Alkylation: Summary

In conclusion, alkylation of sulfonated and non-sulfonated trimethylindolenines was efficiently performed by microwave heating. The great advantage of the use of microwaves instead of the classical oil bath heating, is the possibility to easily and rapidly synthesise in parallel and in shorter time (Figure 2.15) various nucleophilic heterocycles useful in discovery chemistry for the screening of dyes with new optical and chemical properties.

![Figure 2.15 N-alkylation: microwave heating vs reflux.](image)

2.4.2 Synthesis of sulfonated cyanine dyes

The trimethine dye Cy3 was prepared by heating to reflux quaternary salt 26a and \(N,N'\)-diphenylformamidine in a mixture 1:1 of acetic acid and acetic anhydride (Scheme 2.19). The reaction was monitored by UV to avoid/reduce the formation of any symmetrical dye and drive to completion the synthesis of the hemicyanine intermediate 28.

The reaction mixture was left under agitation until the signal at 286 nm (heterocycle) disappeared and the signal at 380 nm (hemicyanine) arose. After evaporation of acetic acid and acetic anhydride and removal of the excess \(N,N'\)-diphenylformamidine 27a (by washing with ethyl acetate), the hemicyanine was dried \textit{in vacuo} and used directly for the next step. The hemicyanine was activated in the presence of pyridine and acetic anhydride and reacted at 80 °C with 26b. The
reaction mixture was heated until the signal at 380 nm completely vanished. After removing the solvent *in vacuo*, the product was precipitated in diethyl ether and centrifuged.

![Scheme 2.19](image)

Scheme 2.19 Synthesis of sulfonated cyanine dyes: a) 26a, Ac₂O, AcOH, reflux, 2h; b) 26c,Ac₂O, pyridine, 80 °C, 20 min-1h

The dye obtained contained various impurities, such as symmetrical dyes and unreacted quaternary salt. Therefore, chromatography was performed in order to obtain the desired compound in high purity. However, this procedure was laborious and often required two consecutive purifications to obtain the dye in > 95% purity. Water soluble trimethine dye 29a was obtained in 36% yield.

Water soluble pentamethine dye Cy5 was prepared following the same synthetic procedure. Quaternary salt 26a was refluxed with malonaldehyde dianyl hydrochloride to form the hemicyanine 28b, in a mixture 1:1 of acetic acid and acetic anhydride. After isolation, 28b was condensed with a different quaternary salt 26b in the presence of pyridine and acetic anhydride. The signal at 445 nm in the UV spectra vanished while the absorption at 650 nm rose correspondingly. After 20 minutes, the solvent was removed and the product precipitated in diethyl ether.
ether and then centrifuged. Purification by chromatography was performed eluting with a DCM: MeOH gradient (100% DCM → DCM:MeOH 1:1).

The pentamethine dye 29b was obtained as a blue glassy solid in 91% purity (by ELSD) (Figure 2.16) and in 40% yield.

![Figure 2.16 HPLC trace of 29b prepared in solution. Detection at λ = 650 nm.](image)

2.5 Solid-phase synthesis of water-soluble cyanine dyes

Purification by chromatography of cyanine dyes is difficult, in particular when they are purified in larger amounts (≥ 100 mg) and often, more then one purification is required to afford them in high purity. This is the reason for their very high cost. Solid phase chemistry represents an attractive strategy for the practical synthesis of water-soluble cyanine dyes since chromatography can be avoided for the purification of the desired dye. The application of solid phase methods to the preparation of water-soluble cyanine dyes is discussed in the section 2.5.1.

2.5.1 From polymer-bound imidates

Our first attempts to synthesise water-soluble cyanine dyes, were carried out by following the method proposed by Balasubramanian et al.\textsuperscript{108} This synthetic protocol involves the condensation of water-soluble trimethylindoleninium salts 26a-b with the imidate immobilised onto polystyrene resin 13b, Scheme 2.20. As reported in
published work, sulfonated heterocycles do not react well during condensation with immobilised imidates, but they reacted well during the cleavage of the dye from the resin.

We postulated that microwave heating would be useful to afford cyanine dyes with sulfonated groups on both heterocycles.

Following this protocol, water-soluble cyanine dyes bearing sulfonate groups on both heterocyclic moieties were generated. However, yield (10%) and purity of the final dye were not satisfactory. One main limitation was the low solubility of the sulfonated heterocycle in DMF and other organic solvents and consequently, the difficulty to react with the imidate on the resin and to be washed away afterwards from the beads. To overcome this limitation, Jiang et al., recently, suggested the use of poly(ethylene glycol), a soluble polymer support which combines homogenous reaction and therefore, high reactivity with solid phase techniques (easier isolation and purification of products). However, this type of support is hard to handle (difficulty in completely separating product and impurities) and low yields, due to loss of the product during precipitation and to the low loading of the support.

Therefore we considered the possibility of applying an alternative catch-and-release strategy (Scheme 2.21). The hemicyanine is more hydrophobic than the sulfonated

\[13b\] 

\[30b\] 

\[29b\]

Scheme 2.20 Synthesis of sulfonated cyanine dyes on solid phase: a) 26a, DMF, 150 °C, μw, 15 min; b) 26b, Ac₂O, Pyridine.

Therefore we considered the possibility of applying an alternative catch-and-release strategy (Scheme 2.21). The hemicyanine is more hydrophobic than the sulfonated
heterocycles and therefore more soluble in organic solvents, where the resin has good swelling properties. This should improve the loading reaction and the removal of any unreacted hemicyanine from the resin.

2.5.2 Catch and release approach

Hemicyanine 28b was prepared as described in section 2.4.2. After completion of the reaction, the hemicyanine was precipitated with diethyl ether and then washed with ethylacetate to remove excess N-[(1E,3E)-3-(phenylimino)-1-propenyl] aniline hydrochloride. After drying in vacuo, 28b was used directly for the next step, without further purification, avoiding excessive exposure to light and prolonged storage at room temperature.

The hemicyanine 28b was added to the resin 18 and shaken for 1 hour at room temperature in a mixture 3:1 of DCM and DMF (Scheme 2.20). The immobilised hemicyanine 30b obtained was washed extensively with DCM and MeOH to remove any residual hemicyanine or symmetrical dye absorbed onto the beads before being reacted at room temperature with 26b in the presence of pyridine and acetic anhydride until all the sulfonated heterocycle 26b was consumed. It was important to avoid the contamination of the dye with such compound since a chromatographic purification would be necessary to separate them. The sulfoindocyanine 29b was obtained in excellent yield (97%) and purity (~98%, by ELSD and 94% at λ = 650 nm) (Figure 2.17) without purification.
Scheme 2.21 Catch and release synthesis of water soluble pentamethine dye 29b: a) 18, DCM:DMF 3:1, rt, 1h; b) 26b, Ac₂O, pyridine.

Figure 2.17 HPLC trace of crude pentamethine dye 48. Detection: ELSD and at λ = 650 nm. Peak at 0.75 min is not attributable to the sample 48.

2.6 Summary and Conclusions

In conclusion, two practical approaches to the solid phase synthesis of unsymmetrical cyanine dyes suitable for bioconjugation have been reported in this chapter. The catch-and-release approach and the application of microwave heating to the synthesis of cyanine dyes, allowed the preparation of cyanine dyes spanning
the whole colour range (Table 2.5 and Figure 2.18), in a few steps, overcoming the limitations of the method described by Balasubramanian.

Through a catch-and-release strategy we have also been able to synthesise, on the solid phase, heptamethine and water soluble cyanine dyes in high yields and purity without the need for chromatographic purification, thus providing an easy route to the preparation of these fluorescent probes.

![Figure 2.18 Normalised absorbance spectra of cyanine dyes prepared on solid phase.](image-url)
Table 2.5 Cyanine dyes prepared on solid phase. a Yield of isolated product in the final step (based on the amount of heterocycle used in the cleavage step

<table>
<thead>
<tr>
<th>Dye</th>
<th>$\lambda_{\text{max, abs}}$ (nm)</th>
<th>$\lambda_{\text{max, em}}$ (nm)</th>
<th>Yield$^{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20a</td>
<td>547</td>
<td>561</td>
<td>68%</td>
</tr>
<tr>
<td>20b</td>
<td>587</td>
<td>603</td>
<td>49%</td>
</tr>
<tr>
<td>20c</td>
<td>640</td>
<td>660</td>
<td>84%</td>
</tr>
<tr>
<td>20d</td>
<td>678</td>
<td>704</td>
<td>92%</td>
</tr>
<tr>
<td>20e</td>
<td>743</td>
<td>765</td>
<td>86%</td>
</tr>
<tr>
<td>20f</td>
<td>781</td>
<td>808</td>
<td>51%</td>
</tr>
<tr>
<td>29b</td>
<td>647</td>
<td>667</td>
<td>97%</td>
</tr>
</tbody>
</table>
2.7 **Cyanine dyes and their applications**

In the past few years, cyanine dyes have found wide application in biology and biomedical imaging, in multiplexed experiments requiring multicolour fluorescent probes.

Today, they are the most commonly used dyes for the labelling of nucleic acids,\(^\text{135}\) in DNA sequencing,\(^\text{84}\) SNPs analysis\(^\text{136,137}\) and gene expression profiling experiments.\(^\text{138,139}\) They are also widely used for the labelling of proteins and *in vivo* experiments because of their absorbance in the NIR region.\(^\text{140}\)

Biological tissues exhibit high photon absorbance in the visible wavelength range (350-700 nm, due to haemoglobin and tissue pigments) and in the infrared range (>900 nm, due to lipids and water). However, the absorbance of biomolecules reach minima in the NIR region. Hence, NIR dyes (such as Cy5.5, Cy7 and Cy7.5) offer a great advantage for the imaging of physiological and pathological states.\(^\text{141}\)

Thanks to development of our synthetic protocol for the easy preparation of cyanine dyes, a number of projects in our laboratory took advantage of the availability of multicoloured fluorescent molecules. In this part of the chapter, some of the applications of the dyes synthesised by microwave-mediated solid-phase chemistry are reported.

*In vitro* cellular uptake of labelled peptoids and microspheres shows the applicability of these labels in flow citometry and *in vivo* imaging while the use of cyanine dyes in research projects, based on SNP's analysis and DNA sequencing, shows the importance of multicolour labels in microarray analysis for the labelling of nucleotides, oligonucleotides and PNA.

### 2.7.1 Labelled peptoids for *in vivo* imaging applications

Peptoids are oligomeric peptidomimetics that consist of \(N\)-substituted glycines.\(^\text{3,142}\)

They have a polypeptide backbone with side chains bounded to amidic nitrogen rather than to the \(\alpha\)-carbon. They have the advantage of protease-resistance, biocompatibility and easy incorporation of diverse side chains. Nine-mer peptoids, prepared by Geraldine Escher were labelled with the pentamethine dye 20c and the heptamethine dye 20e (Figure 2.19) and tested (by Geraldine Escher) for cellular uptake *in vitro*, in mammalian cells such as HeLa (human ovarian cancer) and B16F10 (mouse melanoma) cells and *in vivo* (in primary murine monocytes and...
primary human monocytes) in collaboration with the Centre for Inflammation Research at Edinburgh. HeLa and B16F10 cells were incubated with the 9-mer peptoid 31b and analysed by FACS (Fluorescence Activated Cell Sorting). Preliminary results have shown that the cellular uptake of 31b is very efficient in both cell lines and even the dye alone seems to have a great penetrability. Encouraged by these preliminary results, in vivo imaging studies of cells labelled with 31b were also carried out.

Primary murine monocytes incubated with 31b were instilled into the lungs of a mouse. After one day, the labelled cells were visualised in the mouse by a charge coupled device which showed their accumulation in the lung, as expected (Figure 2.20-a). Bioluminiscence studies also proved the non toxicity of the labelled peptoid. Labelled cells were transfected with adenovirus encoding firefly luciferase and instilled into the lungs of a mouse injected intraperitoneally with luciferin. Whole body imaging of the mouse showed that cells labelled with 31b are also bioluminescent and that consequently the viability of the cells is not affected by the peptoid (Figure 2.20-b).

![Figure 2.19 Labelled 9-mer peptoid](image)

**Figure 2.19** Labelled 9-mer peptoid
Another experiment was carried out *in vivo*, using primary human monocytes labelled with $31b$. These cells were injected via the tail vein into a mouse with peritoneal inflammation. Visualisation of the cells by optical imaging revealed that they accumulated predominantly in the thymus, liver, bone marrow and peritoneum (Figure 2.21).
Labelling of microspheres/ doughnuts & \textit{in vivo} imaging

Another important application of NIR cyanine dyes is the labelling of nano and microparticles for \textit{in vivo} imaging applications.\textsuperscript{144} Polystyrene beads (of 0.1-5 \textmu m) are widely used in immunoassays and for cellular labelling, often in combination with flow cytometry. In recent years, beads have been fluorescently dyed for purposes of encoding.\textsuperscript{145} It is possible to either covalently bind the fluorescent dyes onto the modified surface of the microparticles or to dissolve the dyes into the particles. For covalent coupling, the beads and the fluorophore must have functional groups for binding. Incorporation of dyes inside the beads requires the use of lipophilic fluorophores. This method is usually preferred because it reduces the interference of the dye with the surface chemistry and additionally there is more dye incorporated since there is more space available inside the bead for the fluorophore. Hydrophobic cyanine dyes described in this chapter are ideal for this purpose. In a recent work, carried out in the group, \textit{20a} and \textit{20b} were used for the labelling of microspheres and micro-donughts, in order to visualise the accumulation of these materials inside the cells (\textit{in vitro}) and in the liver of a mouse after intravenous injection (\textit{in vivo}) (Figure 2.22).\textsuperscript{146}

![Figure 2.22 Optical whole body image of Cy7 labelled doughnuts (yellow) in the liver.](image)

In another experiment, primary human neutrophils incubated with 0.5 \textmu m microspheres labelled with \textit{20 e} were injected into a thigh of a mouse. Optical imaging of the whole mouse showed how the microparticles labelled with \textit{20e} could be easily discriminated from a 9-mer peptoid labelled with \textit{20c}. This experiment shows the importance of the availability of cyanine dyes in different
colours for the simultaneous investigation of various biological processes in vivo (Figure 2.23).

![Figure 2.23 Primary human neutrophils incubated with 0.5 μm microspheres labelled with Cy7(left) or Cy5-labelled peptoid (right) and injected in two different thighs](image)

### 2.7.3 Labelling of nucleic acids & nucleotides

There is a great interest in labelling oligonucleotides and nucleotides with fluorophores, because of their application in DNA sequencing and DNA hybridisation studies. Functional groups able to react with fluorophores are usually absent in oligonucleotides. A common method is to introduce an amino reactive group into a synthetic oligonucleotide as the last step in a synthetic process. Then the label can be introduced into the oligonucleotide via a reactive group such as a NHS ester which binds to the amino group.\(^{147}\) The most important method to label DNA is the use of fluorescent phosphoroamidites obtained by reacting a phosphine derivative with a fluorophore containing an hydroxyl group.\(^{148}\)

Labelled nucleotides are generally synthesised by coupling allylamine or propargylamine-dNTP to succinimidy1 ester derivatives of fluorescent dyes.\(^{84,135}\) Some companies include also a spacer between the dye and the nucleotide to improve enzymatic incorporation of the modified nucleotides. More details on the labelling and applications of nucleotides will be provided in Chapters 3-5.
3 Design & Synthesis of Enzymatically Cleavable Nucleotides

The purpose of this chapter is to describe in details the synthesis of a new family of reversible terminators for DNA sequencing by synthesis.

3.1 Reporter Nucleotides for DNA Sequencing by Synthesis

The read-length of a DNA sequencing by synthesis (SBS) technology is determined by 1) the efficiency of incorporation of reversible terminators, 2) the efficiency of cleavage of the fluorophore and 3) the efficiency of deprotection of the 3' hydroxyl group from the extension product (when the 3' hydroxyl group is blocked to avoid multiple extensions). If one cycle of SBS, comprising incorporation, imaging, deprotection/cleavage has a low efficiency, the read length (the number of times the cycle can be repeated) of the method will be low (see Figure 3.1).43

Therefore, research efforts for the development of new SBS methods have mostly focused on the improvement of cycle efficiency through the design of adequate reversible terminators.

![Figure 3.1 Left: a SBS cycle. Right: the maximum read-length decreases dramatically as the cycle efficiency lowers. Table and scheme adapted from ref.1.](image)
In order to obtain a high sequencing by synthesis cycle efficiency, the following requirements for the cleavable moiety of the reversible terminators has to be considered:

1. It must be stable during the reaction with DNA polymerase.
2. It must be easily conjugated with nucleotides and fluorophores.
3. It must not prevent nucleotide recognition by DNA polymerases.
4. It should be cleaved under mild conditions compatible with the stability of DNA and biochip surfaces.
5. It must be cleaved as close to the nucleotide as possible to minimize distortions in the structure of the growing DNA strand that could impair the ability of DNA polymerase to further extend the primer.29,149

Another important issue that must be addressed to sequence DNA using reversible terminators is the termination of the polymerase reaction after the incorporation of a single nucleotide. If the 3' hydroxyl group is left unprotected, multiple extensions might occur in the presence of all four nucleotides making difficult to determine the sequence of the DNA template. It is possible to cap the 3'-OH with a protecting group. However the protecting group must be stable during primer extension and also be efficiently and rapidly removed after incorporation of the nucleotide to regenerate the 3'-OH to allow the next primer extension to take place. DNA strands that have a remaining 3'-OH blocking group are not extended and will not contribute to the detection of the incorporated base, reducing thus the intensity of the fluorescent signal. However, if that blocking group is removed in the next deprotection cycle, it will generate a false signal which will contribute to increase the noise.29

Alternatively, the 3'-OH can be left unprotected and the cleavable moiety on the base designed such that multiple incorporations of nucleotides by DNA polymerase do not occur.

With these considerations in mind it was sought to design a series of modified nucleotides that would incorporate a cleavable fluorescent moiety for imaging but that were nonetheless still able to be recognised by a DNA polymerase.
3.1.1 Design of a Novel Reporter Nucleotide

To design reporter nucleotides, it is important to consider the structure of a DNA polymerase complexed with a double stranded DNA and an incoming nucleotide. In Figure 3.2, it can be seen that the 5-position of the nucleotide points away from the catalytic pocket of the polymerase, while the 3'-position of the ribose ring is very crowded because it is close to a number of amino acid residues in the active site.

![Figure 3.2](image)

**Figure 3.2** The active site of DNA polymerase I from *Thermus Aquaticus* (grey surface) in complex with a double stranded DNA terminated by a dideoxynucleotide (orange sticks) and a dideoxynucleotide (colored sticks) bound to the active site. The orange and blue arrows indicate the 5-position and 3'-position of the free dideoxynucleotide. This structure was downloaded from the Protein Data Bank (PDB), (code 2ktq) and the image created with the program VMD.

Thus, any group that is attached at the 3'-position must be small enough not to inhibit the polymerase activity or reduce the fidelity of the enzyme. On the contrary, large bulky groups at the 5 position of pyrimidines (C and T) and 7 positions of purines (A and G) are well tolerated by modified polymerases. Numerous works published in the literature supports this fact.

We proposed to link the fluorophore to the nucleotide at the 5 position while the 3' hydroxyl group is left unprotected. We postulated that, with an adequate choice of linker and fluorophore, the steric hindrance of the reporter moiety may confer...
termination properties to the nucleotide and that protection of the hydroxyl group in 3' position may not be necessary in this case. This would avoid carrying out an additional protection step during the synthesis and one deprotection step during SBS cycles.

Recently, the use of fluorescent nucleotides with unprotected 3' hydroxyl groups as terminators of DNA polymerisation has indeed been reported, supporting the idea that the protection of the 3' site might be not necessary in particular reaction conditions and with an appropriate reporter group.

Several types of cleavable linkers have been used in the past for the synthesis of reversible terminators, including chemically cleavable and photocleavable linkers (Chapter 1). However, removal of reporter/terminator groups by chemicals or irradiation can cause damage to DNA (crosslinking, depurination, transformation).

As an alternative to chemical and photocleavable reporters, it was thought to use a small peptide to link the fluorophore to the base (Figure 3.3). It was envisioned that the peptide would be cleaved in a mild, fast and efficient way using proteases. Additionally, by varying its sequence of amino acids it should be possible to modulate the properties (e.g. steric hindrance, hydrophilicity/hydrophobicity, polarity) of the peptide in order to find a reversible terminator able to stop multiple base additions without inhibiting the polymerase.

![Figure 3.3 Design principles for enzymatically cleavable fluorescent nucleotides](image)

For incorporation by DNA polymerase

For cleavage by protease

Tag for sequencing
Given a series of modified nucleotides as shown in Figure 3.3 the sequencing by synthesis technology here proposed would therefore consist in the following steps described in Figure 3.4.

1. Unknown single stranded DNA’s bearing complementary sequences to a primer are hybridised onto the array of chip-immobilised primers.

2. Chain extension then takes place with a DNA polymerase and four modified nucleotide analogues (A, T, C, G) each bearing a fluorescent label with a unique colour connected through the peptide substrate.

3. Only one matched nucleotide out of the four nucleotides is incorporated into the growing DNA strand and chain extension is temporarily terminated due to the structure of the unnatural nucleotide analogue. Therefore, the first unknown base next to the primer can be identified by the “colour” of the fluorophore.

4. The fluorescent reporter is cleaved by a protease to allow a second polymerase reaction to occur.

Steps 2-4 are repeated until the unknown DNA template has been sequenced entirely.

The rest of this chapter will discuss the synthesis of the modified nucleotides shown in Figure 3.3.

Figure 3.4 DNA Sequencing by Synthesis on chips using protease cleavable nucleotides
3.1.2 Synthetic strategy

The major challenge in the preparation of reversible terminators is the synthesis of nucleoside triphosphates. They are difficult to synthesise, isolate and characterize. Therefore, it is always preferable to carry out the triphosphorylation reaction in the very last step of a synthetic pathway that involves nucleotides.

In view of this consideration, the best approach for the preparation of enzymatically cleavable nucleotides would be 1) to couple the reporter group (peptide labelled with a fluorophore) to a nucleoside modified on the base with an aminopropargyl group 2) to carry out the triphosphorylation of the dye-peptide-nucleoside conjugate 1 (route A) (Scheme 3.1).

Previous efforts to develop a synthetic protocol based on this approach have been reported. However, the low yield of the triphosphorylation step led to an insufficient amount of product making difficult the isolation and characterization of

Scheme 3.1 Retrosynthesis of enzymatically cleavable nucleotides
the final compound. In addition, the amide bonds of the cleavable moiety are not stable to triphosphorylation agents.

Therefore an alternative route B (Scheme 3.1) based on the coupling of the aminopropargyl nucleotide 2 with the reporter group 3 was considered. This approach is commonly used for the preparation of fluorescent nucleotide analogues and is well documented in the literature.

In summary, the synthetic protocol for the preparation of enzymatically cleavable nucleotides (Scheme 3.2) comprises the following steps, which are described in details in sections 3.3, 3.4 and 3.5:

1. Peptide synthesis and labelling with a fluorescent dye
2. Synthesis and triphosphorylation of aminopropargyl nucleosides
3. Final coupling between the nucleotide and the labelled peptide to form the reporter nucleotide.

Scheme 3.2 General scheme for the synthesis of protease-cleavable nucleotides. R = resin
3.2 Design and Synthesis of Labelled Peptides

As mentioned earlier, the reporter group (peptide and fluorophore) of the cleavable nucleotides must satisfy a number of requirements to ensure a high efficiency in the sequencing by synthesis cycle.

- The peptide used to link the base to the fluorophore must be designed such that its bulkiness does not inhibit its incorporation in a DNA strand by a DNA polymerase but does terminate the extension after the first incorporation.
- The digestion of the peptide by a protease must be fast, complete and as close as possible to the aminopropargyl group.
- The protease and the polymerase must tolerate the close proximity of the nucleotide and the fluorophore to be able to digest efficiently the peptide.

3.2.1 Design of the Peptidic Linker

Previous work was carried out in our lab by Dr. Karin Johnson, to find the best combination of fluorescent nucleoside-peptide conjugates and proteases by screening a library of nucleoside analogues constructed on a solid support (Figure 3.5).

![Figure 3.5 General structure of fluorescent peptide-nucleoside conjugates](image)

A library of 57 compounds was prepared by solid phase parallel synthesis, from resin-bound nucleosides, and screened with seven different types of proteases (proteinase K, elastase, subtilisin, papain, thermolysin, trypsin, enterokinase). The cleavage efficiency for each combination of peptide and enzyme are represented in Figure 3.6.

The library included peptides varying in length, polarity, bulkiness and hydrophobicity/hydrophilicity.
Time course experiments proved that subtilisin and proteinase K exhibited high digestion efficiency for all the peptides and fast cleavage (within 10 seconds) for the following sequences: Ala-Ala-Ala, Ala-Ala-Leu, Gly-Ala-Gly-Leu. These sequences were selected as possible cleavable linkers for the synthesis of our reporter nucleotides.

![Figure 3.6](image)

**Figure 3.6** Protease assay for the selection of the best peptide sequences. The cleavage efficiency was measured after 2h of incubation. This figure was reproduced from Karin Johnson PhD thesis.

### 3.2.2 Selection of the fluorophore

Fluorescence detection is among the technical innovations that contributed, in the last twenty years, to an increase in the throughput of DNA sequencing technologies. Because fluorescent dyes have different spectral properties, they can be mixed and discriminated in virtue of their colours. The selection of fluorescent dyes is critical in the design of fluorescent nucleotides and in particular of reversible terminators because their properties (size, hydrophobicity and polarity, number of charges) can influence the fidelity and the efficiency of the polymerase, the termination properties of the nucleotide, and the solubility of the final compound in aqueous buffers. Fluorescent dyes used in DNA sequencing must also have distinct excitation and emission spectra and good
photophysical properties to ensure high sensitivity during detection. Moreover, they must tolerate high temperatures (>90°C) when used with thermostable DNA polymerases in thermal cycle sequencing.

Xanthine dyes (fluorescein and rhodamine derivatives) have been used for the labelling of primers and dye-terminators for DNA sequencing (Figure 3.7).\textsuperscript{29}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{dyes.png}
\caption{Dyes used in DNA sequencing}
\end{figure}

The common dye set used in DNA sequencers is the REG, R110, TAMRA and ROX series.\textsuperscript{29} These fluorophores have been used in combination with FAM for FRET-based detection (see Chapter 1). Despite their popularity, xanthine dyes suffer from quite a few drawbacks (e.g. pH sensitivity, photobleaching and broad emission spectrum) and therefore have been recently replaced by other classes of dyes such as the cyanine and BODIPY dyes.\textsuperscript{162}
In recent years, cyanine dyes (Figure 3.8) have been attracting growing interests for the labelling of biological compounds thanks to their high photostability, quantum yield and solubility in water. They are available in different colours; therefore they are routinely used in multiplexed experiments for the labelling of nucleic acids and antibodies and are the most popular choice in microarray applications.\textsuperscript{94,162}

BODIPY dyes (Figure 3.8) have also found application in DNA sequencing, as substitutes of xanthine dyes, because of their narrow emission bandwidths that reduce cross-talking between different detection channels and because they exhibit minimal influence on the mobility of fragments in gel electrophoresis.\textsuperscript{18}

![Cyanine Dyes](image)

![BODIPY Dyes](image)

**Figure 3.8 Cyanine and BODIPY dyes**

### 3.2.3 Synthesis of the fluorophores

The cyanine dyes were selected for the labelling of protease-cleavable fluorescent nucleotides. A detailed description of the synthesis of cyanine dyes used for this purpose has been reported in the Chapter 2. However, during our early investigations on the synthesis of fluorescent nucleotides, the preparation of cyanine dyes was still in progress and therefore rhodamine and fluorescein derivatives were selected. Despite the disadvantages mentioned earlier, they are available at relatively low prices and can be prepared or derivatised on large scale from inexpensive precursors.
The derivatisation of carboxyfluorescein and rhodamine B dyes is described in the sections 3.3.3.1 and 3.3.3.2.

To ensure peptide accessibility to the protease and minimize the steric hindrance caused by the fluorophore to the polymerase, it was thought to incorporate an additional spacer between the dye and the peptide. The presence of a spacer offered also the advantage of facilitating the labelling of the peptide during synthesis. Dyes are in general difficult to purify by chromatography, therefore practical synthetic protocols were sought, using solid supports and microwave heating whenever possible, in order to be able to purify them by filtration or extraction instead of chromatography.

3.2.3.1 Synthesis of Ahx-Carboxyfluorescein

5(6)-Carboxyfluorescein is a fluorescent dye with an excitation maximum of 494 nm (in the blue-green region) and an emission maximum of 515 nm (in the green region). It is commercially available and it usually sold as a mixture (40:60) of two isomers: 5 and 6 carboxyfluorescein. Isomerically pure carboxyfluorescein is also available, although in milligram quantities and at a higher cost (125 fold higher than the mixture). 163

5(6)-Carboxyfluorescein display a pH-dependent equilibrium between the lactone and the open form. The open form (carboxylic acid group in position 2 is free) has fluorescence properties due to its conjugated π-system and is present in aqueous solutions with pH above 5. By increasing the pH, the absorbance of carboxyfluorescein is shifted from 450 nm to 495 nm due to the ionization of the phenol and the carboxylic groups. In the lactone form, normally present in organic solvents, the carboxylic group in position 2 is involved in the formation of a lactone and the dye is not fluorescent (Figure 3.9). 34

![Figure 3.9 Tautomerism of carboxyfluorescein](image-url)
In order to be coupled to the peptides, 5(6)-carboxyfluorescein was derivatised with a spacer by coupling the carboxylic group in position 5(6) with aminohexanoic acid methylester. The use of PyBOP as coupling reagent for the derivatisation of 5(6)-carboxyfluorescein has been previously reported. Here the coupling was performed in the presence of PyBOP and DIEA under microwave heating at 60 °C for 20 min (Scheme 3.3). The product was purified by extracting the excess of amine and the coupling reagent in the aqueous layer and the product in a mixture 2:1 of 2-propanol and DCM. Afterwards, the methyl ester was hydrolised with LiOH in a mixture 1:1 of THF and water. After evaporation of the solvents, the residue was dissolved in water and precipitated with 2M HCl to afford the fluorescent compound 47 which was used for the labelling of peptides without purification by chromatography. The HPLC trace of 47 (at λ = 495 nm) is shown in Figure 3.10.

Scheme 3.3 Synthesis of Ahx-carboxyfluorescein. a) PyBOP (1.5 eq.), DIEA, 6-aminohexanoic acid methyl ester (2 eq.), 60 °C, 20 min; b) LiOH in water/THF, room temperature.

Figure 3.10 HPLC trace of 47 (detection at λ = 495 nm)
3.2.3.2 Functionalisation of Rhodamine B

Because 5(6)-tetramethylcarboxyrhodamine is a relatively expensive dye (£629/g), Rhodamine B (£1/g) was selected as starting material for the preparation of a functionalised fluorophore to use for peptide labelling. Rhodamine B base has only one carboxylic group in the 2 position and has the advantage of not having two isomers. However, the use of this dye has been limited by the fact that this carboxylic group (and any amide derivative) rapidly cyclize and is difficult to couple to amines.\textsuperscript{165}

To address this problem, Francis et al. proposed a practical method for the preparation of tertiary amides from Rhodamine B (Scheme 3.4). This approach allows one to prepare rhodamine derivatives in multigram scale and high purity without any chromatographic purification.\textsuperscript{165}

Rhodamine B was first treated with piperazine (4 eq.) and AlMe\textsubscript{3} (2 eq.) in refluxing DCM. After purification by simple workup, the dye was functionalised using succinic anhydride, NE\textsubscript{T} and DMAP to afford the product 43 in high purity (95\%) and with an overall yield of 70\% (Figure 3.11).

The procedure described by Francis was slightly modified, as it was found 1 M HCl to be too strong for the acidic workup of the dye. Indeed, these conditions were causing hydrolysis of the product forming the starting material rhodamine B. This problem was overcome by using a 0.1 M solution of HCl. Rhodamine B was shown to be stable to TFA/DCM (95:5) and 1 M HCl in dioxane. The HPLC trace of 43 is shown in Figure 3.11.
Fluorescein and Rhodamine derivatives were chosen for the first attempts of nucleotide labelling. The development of this project required the use of four dyes with different excitation/emission spectra, therefore cyanine dyes were also considered. As mentioned in chapter 2, this class of dyes had excitation/emission spectra that could be tuned by modifying the length of the polymethine chain and the heterocycles. In addition, their fluorescence is not pH dependent and their non-specific binding to biomolecules is low. Most importantly, all the cyanine dyes have similar properties. Cyanine dyes are commercially available. However, because of their high cost, an efficient way to synthesise them by following a single protocol it was sought (see Chapter 2).

3.2.4 Peptide Labelling

The peptide was assembled on a Wang resin\textsuperscript{166,167} by Fmoc chemistry (Scheme 3.5). The first amino acid was loaded using DIC (3 eq.) and a catalytic amount of DMAP in DMF. For the other couplings, equimolar solutions of reactants (3 eq. relatively to the resin) were prepared by adding DIC to a mixture of HOBT and Fmoc amino acid in a solution 1:2 of DMF and DCM. Deprotection of Fmoc amino acids was carried out with a 20% solution of piperidine in DMF. Coupling and deprotection steps were monitored by qualitative ninhydrin test\textsuperscript{168} carried out on small portions of resin. After successful preliminary enzymatic tests carried out in solution using the nucleotide dUTP-(A)\textsubscript{3}-Rho, it was thought to focus mainly on the labelling and nucleotide conjugation of the sequence H-Ala-Ala-Ala-OH.
The resin 32 was separated in five different portions (100 mg each). Each portion was treated with a solution of the dye previously activated as NHS ester. The dye, dissolved in a mixture 1:1 of DMF and pyridine, was activated by reaction with disuccinimidyl carbonate (DSC) at 60 °C for 1-3h. It was important to purify the NHS ester from the excess of DSC before labelling in order to avoid any reaction with the amino group of the peptide.

Scheme 3.5 Synthesis of Dye-Ala-Ala-Ala-OH. Reagents and conditions: a) Solid Phase Peptide Synthesis (SPPS): i) H-Ala-OH, DIC (3 eq), DMAP (0.3 eq.) in DCM/DMF (2:1, v/v); ii) piperidine in DMF (1:4, v/v); iii) H-Ala-OH, DIC (3 eq.), HOBt (3 eq), in DCM/DMF (2:1, v/v); repeat step ii) for Fmoc-deprotection and step iii) for coupling; b) Activation of the dye: DSC (~1.5 eq.), pyridine/DMF (1:1, v/v), 60 °C, 1-2h; then coupling to 32 in DMF; c) TFA/water or TFA/DCM (95:5, v/v), 15 min (twice).
Therefore, after removing the solvents from the product, the residue was washed with Et₂O and then centrifuged to remove the supernatant containing DSC and traces of pyridine. This activation method was preferred to the use of DIC and HOBt as unreacted or hydrolysed dyes could be reused after coupling. This was particularly important in the case of the valuable cyanine dyes.

The coupling was carried out by mixing the activated dye (3 eq.) dissolved in DMF with the resin 32 for 1 h (or longer when necessary). The excess of unbound dye was usually removed using DCM and DMF. However in the case of the dye peptide FAM-Ala-Ala-Ala-OH 46, the resin was also treated with a 20% solution of pyridine in DMF in order to cleave any side product formed from the acylation of the phenolic groups with the activated carboxyfluorescein derivative. The efficiency of the washings was confirmed by the absence of dyes (and colour) in the filtrate.

Because of the availability of large amounts of dyes, it was possible to push the reaction of labelling to completion often without the need for purifications by RP-HPLC.

![HPLC traces](image)

**Figure 3.12** HPLC traces (overlapped) of the labelled peptides. From left to right: Cy3(S)-(Ala)₃-OH, FAM-(Ala)₃-OH, Cy5(S)-(Ala)₃-OH, Rho-(Ala)₃-OH, Cy5(H)-(Ala)₃-OH. Detection at λ = 490 nm, 550 nm, 650 nm.

Unbound dyes washed away from the resin, were recovered by evaporating the solvent and by precipitating the residue, dissolved in the minimum amount of DMF, in a 25 or 50 mL Falcon tube filled with Et₂O. After centrifugation, the supernatant was removed; the precipitate was dried and used again for other labelling reactions.
Aminopropargyl nucleoside triphosphates

Aminopropargyl nucleoside triphosphates are important precursors of fluorescent nucleotides. However, the high cost and the challenging chemical synthesis place a limitation to their use.

To address this issue, two different solid phase methods have been developed in parallel in our lab for the triphosphorylation of nucleosides. One method is based on the use of a solid supported triphosphorylation reagent: the Ludwig-Eckstein reagent. The second is based on the triphosphorylation of supported nucleosides using the classic Ludwig-Eickstein reagent. The latter approach was developed by Dr. Jin Ku Cho, who collaborated in the realisation of this project, and was applied to the preparation of enzymatically cleavable nucleotides.

The synthetic scheme for the preparation of aminopropargyl nucleoside triphosphates is described in Scheme 3.6 for the amino-modified nucleotide dCTP. The primary 5'-OH was protected selectively with a DMT group followed by a Sonogashira reaction with trifluoroacetylpropargyl amine to introduce the aminopropargyl linker on the base. After succinylation of 3'-OH, the nucleoside was loaded onto an amino-methylated silica resin. The unreacted amino groups on the resin were capped by acetylation.

After removal of the DMT group, solid phase triphosphorylation was carried out using the Ludwig-Eckstein method: the nucleoside derivative II was reacted with 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one to afford an activated phosphite III. Reaction with pyrophosphate gave the cyclic intermediate IV that was oxidised and hydrolysed to give the correspondent triphosphate V. Treatment of IV with aqueous ammonia cleaved also the product from the resin in solution and deprotected the aminopropargyl group.

The advantage of this approach is the feasibility of carrying out the triphosphorylation reaction on a solid support: the use of excess of reagents and removal of side-products from the resin by filtration ensure that the final triphosphate is obtained in good yields and purity.
Scheme 3.6 Synthesis of aminopropargyl nucleoside triphosphates on solid phase. a) DMT-Cl (1 eq.), DMAP, pyridine; b) succinic anhydride, TEA, DMAP, 1,2-DCE; c) CHCCH₂NHCOCF₃, Pd(PPh₃)₄, CuI, Amberlite-IRA 67, DMF; d) CPG-NH₂, DIPCDI, HOBt, DMF; e) 3% TCA in DCM; f) 1M chloro-4H-1,3,2-benzodioxaphosphorin-4-one in THF; g) 0.5 M bis(tri-n-butylammonium) pyrophosphate in DMF; h) 0.5 M I₂ in pyridine/water (98/2); i) NH₄OH.

This method was efficiently applied by Dr. Jin Ku Cho to the synthesis of all four amino-modified nucleotides (dUTP, dCTP, dATP and dGTP) which were all obtained in good yields (53-61%) after HPLC purification.
3.4 Enzymatically Cleavable Nucleotides

The position of the reporter group on the base is critical for the successful incorporation of the nucleotide into DNA. For the pyrimidines (C and T), the attachment of a linker at the 5 position is optimal as it is less likely to destabilise the duplex secondary structure. For the purines, the linker can be attached either at the 7 or at the 8 position although it is known that substituents at the 8 position can destabilise the DNA secondary structure (Scheme 3.8).\textsuperscript{172,173}

The amino-modified nucleotides used for the synthesis of enzymatically cleavable nucleotides were mainly Am-dCTP V and Am-dUTP VI (Scheme 3.7). These compounds were obtained in larger amounts than Am-dGTP and Am-dATP and were easier to obtain as they did not require additional protection steps for the bases.

To ensure that the succinimidyl ester of the dye and the nucleotide were soluble and stable in the solvent used for the coupling, TEAB buffer was selected as solvent for the coupling reaction and for the purification, as offered the advantage of being easily removed from the product by liophilisation and to solubilise both dyes and nucleotides. Furthermore, pH 8.5 of the buffer ensured that the amino group linked to the base was in the unprotonated form and avoided the hydrolysis of the triphosphates and of the succinimidyl ester, which would normally occur in aqueous buffers. The labelled peptides activated as NHS esters, were coupled to the aminopropargyl nucleoside triphosphates, V or VI, in 0.1 M TEAB buffer at pH 8.5. Reaction times varied from 30 min to 2h depending from the properties of the fluorophore used for the labelling.
Different combinations of dye-labelled peptides and nucleotides were prepared. The detailed preparation and characterisation of these compounds are described in the sections 3.5.1-3.5.3.

This part of the project was carried out in collaboration with Dr. Jin Ku Cho who worked on the synthesis and labelling of the amino-modified nucleotides.
3.4.1 Coupling of labelled peptides to Am-dUTP

3.4.1.1 Synthesis of dUTP-(Ala)₃-Rho 46

The dye peptide 45 was activated as NHS ester by reaction with DSC (1.5 eq.) in a mixture 1:1 of dry pyridine and dry DMF for 1h at 60 °C. The NHS ester thus obtained was precipitated in Et₂O and then washed several times with dry ethyl acetate or dry Et₂O. Afterwards, the amino-modified nucleotide Am-dTUP VI (3 μmol) dissolved in 100 μL of 0.1M TEAB buffer was mixed with 100 μL of the labelled peptide 45 (3 eq.) previously dissolved in the same buffer. The coupling reaction, carried out at room temperature, was monitored by TLC and the product purified by semi-preparative RP-HPLC using 0.1 M TEAB buffer and acetonitrile as eluents. Analysis by HPLC of the reaction mixture showed the presence of the product 46 (peak 1) (Figure 3.13), the unreacted labelled peptide 45 (peak 2) and additional small peaks 3, 4, 5, which were formed during the coupling reaction and during the purification by hydrolysis of the rhodamine dye and of the triphosphate.
Fraction I was isolated and analysed by MALDI-TOF giving a peak at \( m/z \) 1326. 50 nmoles of cleavable nucleotide (1.5% yield) were obtained, as confirmed by UV-visible absorbance measurement in 0.1 M TEAB buffer.

![HPLC trace of dUTP-(Ala)3-Rho](image)

**Figure 3.14** HPLC trace of dUTP-(Ala)3-Rho 46 before (green) and after (blue) purification. Detection at \( \lambda = 567 \) nm. Dr. Jin Ku Cho's analysis.

The yield of the reaction could be eventually improved by prolonging the reaction time; because it was important avoiding the decomposition of the triphosphate and eventually of the dye in the buffer it was preferred to keep the reaction time to 30 min-2h.

### 3.4.1.2 Synthesis of dUTP-(Ala)3-Cy3 35

The synthesis and the purification of dUTP-(Ala)3-Cy3 were carried out analogously to dUTP-(Ala)3-Rho. 35 was obtained in 2.1% yield (~50 nmoles). The HPLC chromatograms of the product before and after purification are reported in Figure 3.15.

![HPLC trace of dUTP-(Ala)3-Cy3](image)

**Figure 3.15** HPLC trace of dUTP-(Ala)3-Cy3 35 before (green) and after (blue) purification. Detection at \( \lambda = 567 \) nm. Dr. Jin Ku Cho's analysis.
It is important to notice that the peptide labelled with the sulfonated cyanine dyes Cy3 and activated as NHS esters in pyridine/DMF was used for the coupling with the nucleotide as pyridinium salts without further purification by ion exchange chromatography. Because the purification by RP-HPLC of the nucleotides labelled with water soluble cyanine dyes is more difficult to carry out as the dye derivatives have very similar retention time, it is critical to use high purity nucleotides and labelled peptides in order to avoid complicated and time-consuming purifications. The isolated fraction containing the desired nucleotide 35 was analysed by MALDI-TOF. A peak at \( m/z \) 1333 corresponding to the mass of nucleotide 35 was found.

### 3.4.2 Coupling of labelled peptides to Am-dCTP

![Chemical structure](image)

**DYE:**

- Cy5-hydrophobic = ![Chemical structure](image)
- Cy5-sulfonated = ![Chemical structure](image)
- 5(6)-Ahx-carboxyfluorescein = ![Chemical structure](image)

Figure 3.16 Structures of dCTP-(Ala)\(_3\)-Cy5(H) 41, dCTP-(Ala)\(_3\)-Cy5(s) 38 and dCTP-(Ala)\(_3\)- FAM 51.
3.4.2.1 **Synthesis of dCTP-(Ala)₃- FAM 51**

The nucleotide dCTP-(Ala)₃- FAM 51 (Figure 3.16) was prepared analogously to the other nucleotides described in the sections 3.5.1.1 and 3.5.1.2. However, the activation of the peptide 49 as NHS ester required a prolonged reaction time. Furthermore, the NHS ester seemed to be more labile than other NHS esters in TLC. Because the dye has two isomers 5 and 6, the HPLC trace showed 4 peaks, the first and the third being the dye-peptide 49 and the second and the forth being the product 51 (Figure 3.17). The isolated product was analysed by MALDI-TOF giving three peaks at \( m/z \) 1202 (M-2H), 1204 M⁻ and 1224 (M-3H+Na)⁺.

![HPLC chromatogram of the crude nucleotide dCTP-(Ala)₃- FAM 51](image)

**Figure 3.17** HPLC chromatogram of the crude nucleotide dCTP-(Ala)₃- FAM 51. Detection at \( \lambda = 490 \) nm. Dr. Jin Ku Cho's analysis.

3.4.2.2 **Synthesis of dCTP-(Ala)₃- Cy5 38 and 41**

Activation of the dye peptide Cy5-(Ala)₃-OH as NHS ester was carried out as for the other dye peptides although the reaction seemed far more efficient than for the other dye-peptides. Also the coupling with the Am-dCTP was faster: analysis by TLC showed formation of the product within the first 10 mins. The cleavable nucleotide was purified as usual by semi-preparative RP-HPLC. The HPLC chromatograms of the crude reaction mixture and of the purified product are shown in Figure 3.18. The isolated fraction analysed by MALDI-TOF MS gave two peaks at \( m/z \) 1197 (M-3H)⁻ and 1199 M⁻. 60 nmoles (1.4% yield) of dCTP-(Ala)₃- Cy5 41 were obtained, as confirmed by UV-visible analysis.
The nucleotide dCTP-(Ala)$_3$-Cy5 41 was initially prepared to perform dual colour primer extension experiments on microarray chips. HPLC analysis of primer extension using dCTP-(Ala)$_3$-Cy5 41 in solution (Chapter 4) showed that the nucleotide was efficiently incorporated by DNA polymerase into a growing DNA strand. However, single nucleotide primer extension conducted on microarray chips showed non-specific binding of this nucleotide to DNA due to the hydrophobic nature of the dye and to the presence of a positive charge (Chapter 5).

Consequently, the water soluble cyanine dye Cy5 with sulfonate groups (negative charges) on the heterocycles was considered for the labelling of Am-dCTP V in order to reduce the non-specific binding to DNA. dCTP-(Ala)$_3$-Cy5 38 was obtained in 1% yield (30 nmol), as confirmed by UV-visible analysis. The HPLC chromatograms of dCTP-(Ala)$_3$-Cy5 38 before and after purification by semipreparative HPLC are shown in Figure 3.19. The presence of a peak at m/z 1358 (M-2H), 1360 (M), 1380 (M-2H+Na) in the correspondent MALDI-TOF MS spectrum confirmed the formation of the desired product 38.

Figure 3.18 HPLC chromatogram of Cy5-AAA-dCTP before (green) and after (blue) purification. Detection at $\lambda = 650$ nm. Dr. Jin Ku Cho’s analysis.

Figure 3.19 HPLC chromatogram of dCTP-(Ala)$_3$-Cy5-sulfonated before (green) and after (blue) purification. Detection at $\lambda = 650$ nm. Dr. Jin Ku Cho’s analysis.
3.4.3 Coupling of FAM-(Ala)₃-OH to Am-dGTP

Despite the fact that purines labelled at the 8 position can destabilise DNA, the labelling and incorporation of the amino-modified nucleotide Am-dGTP, modified with a linker in position 8, was also investigated. Indeed, the preparation of the nucleotide VII requires fewer steps than the nucleotide analogue modified at the 7-position VIII therefore it was attractive synthesising and testing such a substrate. Synthesis and purification of the nucleotide dGTP-(Ala)₃-FAM 50 (Figure 3.20) were carried out analogously to the other nucleotides described in the section 3.5.

![Figure 3.20 Structure of dGTP-(Ala)₃-FAM 50](image)

All the fractions isolated after HPLC purification (Figure 3.21) of the reaction mixture were analysed by MALDI-TOF MS; a peak at $m/z$ 1258 (M-4H+NH₄)⁻ was found. However, when tested in a primer extension reaction by a DNA polymerase on a microarray chip, no extension product was detected using this nucleotide.

![Figure 3.21 HPLC trace of dGTP-(A)₃-FAM before purification. Detection at $\lambda = 490$ nm. Dr. Jin Ku Cho's analysis.](image)
3.5 Conclusion

The design, the synthesis and the characterisation of enzymatically cleavable nucleotides have been described in detail in this chapter. Amino modified nucleoside triphosphates, prepared with a novel solid phase triphosphorylation method, were successfully coupled to a small peptide (Ala-Ala-Ala-H) labelled with diverse fluorophores, Table 3.1.

Fluorophores and peptide derivatives were obtained in relative high amounts (30 mg scale for peptides, 100mg-1g scale for the dyes) and often without the need for purification by chromatography.

Further work will be necessary in the future for the preparation of a set of four labelled nucleotides.

As a practical solid phase method has been optimised (Chapter 2) for the synthesis of water soluble cyanine dyes, the use of Cy3, Cy3.5, Cy5 and Cy5.5 is proposed for the labelling of the four bases.

The synthesis and use of cleavable nucleotides containing Am-dGTP and Am-dATP needs further investigation.

Table 3.1 Fluorescent Protease-Cleavable Nucleotides
4 Single Nucleotide Primer Extension with Enzymatically Cleavable Nucleotides: Proof of the Principle

The major challenges facing any novel sequencing by synthesis technology is the design and the synthesis of reversible terminators with the desired properties (Chapter 3) and the identification of a DNA polymerase which can recognise these substrates with high selectivity/fidelity.

The aim of the work described in this chapter was to assess the ability of DNA polymerases to incorporate the enzymatically cleavable nucleotides described in the Chapter 3 and verify the feasibility of using proteases for the cleavage of the reporter group linked to the base.

4.1 Single Nucleotide Primer Extension

Initial efforts focused on the selection of a DNA polymerase that would tolerate and efficiently incorporate nucleotides modified on the base with a labelled peptide and evaluation of the extent of incorporation in a primer extension reaction in solution.

4.1.1 DNA Polymerase

4.1.1.1 Introduction

Since the discovery of the first DNA polymerase in 1958 by Kornburg, many DNA polymerases able to catalyse the incorporation of deoxyribonucleotides into a growing DNA strand have been identified. Based on their sequence homology, they have been sorted into seven different families: A, B, C, D, X, Y, RT. Most of the DNA polymerases used with labelled nucleotides are of the so-called A-family and include Klenow DNA pol I, T7, Taq DNA polymerases, and B-family DNA polymerases such as T4, Tli/Vent, Pfu, Tgo. The original enzyme used by Sanger in DNA sequencing was the Klenow fragment of Escherichia Coli DNA polymerase I, an enzyme which retains 5’ to 3’ polymerase activity but lacks 3’ to 5’ exonuclease activity (proof-reading activity). Indeed, enzymes retaining exonuclease activity cannot be used in DNA sequencing because they could degrade the extended primers, removing the dye labelled terminator and allow further unwanted growth by incorporating unmodified dNTPs.
Klenow has been routinely used in DNA sequencing; however more recently it has been replaced by other enzymes because of its low processivity (average length of DNA synthesized before the enzyme dissociates from the template), long elongation rate and lack of thermal stability.

The polymerases currently used in DNA sequencing are mostly genetically modified T7 and Taq DNA polymerases. T7 DNA polymerase, developed by Tabor and Richardson, is a chemically modified polymerase able to preserve the high processivity and elongation rate of native polymerases. It lacks a 5'→3' exonuclease domain but has a 3'→5' exonuclease activity that is 1000-fold higher than Klenow; therefore it is not recommended for DNA sequencing.

However, a genetically engineered form of T7 DNA polymerase able to retain polymerase activity with almost no 3'→5' exonuclease activity was developed. This polymerase (marketed as Sequenase version II) is often used for high quality manual DNA sequencing; it is highly processive and able to incorporate nucleotide analogues for sequencing (dideoxy NTPs, α-thio dATP, dITP, 7-deaza-dGTP, etc.).

However, currently thermostable DNA polymerases are principally used as they enable rapid amplification of extremely small amounts of DNA in a polymerase chain reactions (PCR) and thermal cycling.

Taq polymerase is a thermostable DNA polymerase isolated from the thermophilic bacterium Thermus aquaticus and is often used in polymerase chain reactions (PCR). It does not have significant 3'→5' exonuclease activity and in part because of this, has low replication fidelity. Indeed, it is often used in combination with the polymerase Pfu in PCR reactions in order to combine its high processivity with the proofreading activity of Pfu polymerases. In 1995 Tabor and Richardson developed two genetically modified Taq polymerases (marketed as AmpliTaq® and Thermo Sequenase) containing single point mutations that reduced significantly the exonuclease activity of the polymerase and the discrimination against ddNTPs and labelled nucleotides. For these reasons AmpliTaq and Thermo Sequenase are generally preferred for DNA sequencing methods using dye-terminators.
4.1.1.2 Selection of a DNA Polymerase

For the purpose of the project, commercially available DNA polymerases were tested in order to find one able to recognise our substrate. The first polymerase assayed was Sequenase v. II. As mentioned earlier, this polymerase does not have proofreading activity and is able to recognise modified nucleotides.

Sequenase v. II was first assayed using unlabelled dNTPs in order to find the best conditions for the analysis and purification of the extended primer by HPLC (see section 4.1.2). However, when tested with the protease-cleavable fluorescent nucleotides, primer extension did not occur.

As alternative, DNA polymerase Thermo Sequenase, a thermostable DNA polymerase specifically engineered for DNA sequencing, was considered. The thermostability of this polymerase makes cycle sequencing possible, combining the processivity of T7 Sequenase DNA polymerase with the sensitivity and reliability of cycle sequencing. It is well known that Thermo Sequenase can recognize and efficiently incorporate fluorescently labelled ddNTPs and that is highly tolerant to extensive modifications to the base of a nucleotide with large groups, such as energy transfer dyes. For all these reasons, Thermo Sequenase seemed very promising for the realization of our project and was thus tested with our reporter nucleotides (section 4.2).

4.1.2 Analysis of Primer Extension

The oligonucleotide sequences, generated during primer extension reactions using fluorescent nucleotides, are usually separated by polyacrylamide gel electrophoresis and visualised by laser-induced fluorescence detection (oligonucleotides can be also labelled with isotopes and examined by autoradiography or can be stained with ethidium bromide or SYBR-Gold and visualised under UV light, but these methods are now used infrequently).

Analysis of extended oligonucleotides by high performance liquid chromatography (HPLC) using 0.1 M TEAB buffer and acetonitrile was perfomed. The advantage of this method is the possibility to easily purify the desired sequence from unreacted nucleotides or oligonucleotide primer and template and to monitor the incorporation
of the fluorophore by UV/Vis detectors. In addition, the isolated fraction containing the extended primer can be analysed by MALDI-TOF to identify the peak.

Oligonucleotides were analysed by MALDI-TOF MS, using 3-hydroxypicolinic acid (3-HPA) as matrix. To obtain good quality MALDI-TOF MS spectra, it was critical for the sample to be concentrated, free of non-volatile cations and agents which disturb crystallization (detergents, urea, DMSO). Indeed, the presence of salts can make MALDI-TOF analysis extremely difficult because of the formation of adducts between the negatively charged nucleic acids and the cations. Moreover, the matrix tends to form crystals at the edge of the sample spotted onto the plate causing an inhomogeneous distribution of DNA and matrix, which requires a time-consuming search for the hot spots of the sample. To improve the sensitivity of MALDI-TOF MS analysis, the extension products, purified by RP-HPLC and lyophilised, were also desalted using the commercially available C$_{18}$ ZipTips.

4.2 Primer Extension with Enzymatically Cleavable Nucleotides

To verify the feasibility of using enzymatically cleavable fluorescent nucleotides as reversible terminators in Sequencing by Synthesis, the nucleotide dUTP-(A)$_3$-Rho was tested in a primer extension reaction performed in solution using Thermo Sequenase DNA polymerase (Scheme 4.1).

The oligonucleotide sequences and the reaction conditions reported by Ju et al. were considered for this proof-of-the-principle experiment. A synthetic 60-mer oligonucleotide corresponding to a portion of exon 7 of the human p53 gene was used as template I to perform the primer extension reaction. The first nucleotide in the template adjacent to the annealing site of the 18-mer primer I was “A”, meaning that the complementary nucleotide to be incorporated during polymerase reaction was “T” (or “U”). The polymerase extension reaction mixture consisted of 200 pmol of primer, 320 pmol of p53-exon7 template, 640 pmol of the nucleotide analogue dUTP-(Ala)$_3$-Rho 46, 16 units of Thermo Sequenase DNA Polymerase and 1 × Thermo Sequenase reaction buffer in a total volume of 50 μL.
The incorporation of the modified nucleotide 46 by Thermo Sequenase DNA polymerase was investigated by HPLC analysis using two different wavelengths to follow the progress of the reaction: 260 nm to measure the UV absorbance of DNA and 567 nm to measure the absorbance of the dye rhodamine B (Figure 4.1). A peak at 6.4 min absorbing only at 260 nm, corresponded to the oligonucleotide template while a peak at 7.2 min, which was detected in both channels, corresponded to the extension product II. The nucleotide 46, which was added in excess, was detected at 13.7 min. MALDI-TOF MS analysis of the purified extended primer II confirmed that the labelled nucleotide 46 was successfully incorporated, into the primer, by Thermo Sequenase. Primer extension was repeated using a reporter nucleotide labelled with a different fluorophore dCTP-(A)$_3$- Cy5 41 (Scheme 4.2).
In this case, the nucleotide in the template II adjacent to the annealing site of the 18-mer primer I was "G".

Template II and primer I sequence:

3'-GGCCGTGTGTATCTCTGCTGTTGGCTGTACCACCATCCACTACAACTACATGTG-5'
5'-AGAGGATCCAACCGAGAC-3'

The incorporation of the modified nucleotide 41 by Thermo Sequenase was investigated by HPLC analysis using again two different wavelengths to follow the progress of the reaction (260 nm and 641 nm). The HPLC chromatogram is shown in Figure 4.2. A peak at 6.9 min absorbing only at 260 nm, corresponded to the oligonucleotide template II while the peak at 9.2 min, which was detected at both wavelengths, corresponded to the extension product IV; the nucleotide 41, which was added in excess, was detected at 15.6 min (Figure 4.2).

HPLC analysis of the primer extension reaction, shown in Figure 4.2, indicated that the modified nucleotide dCTP-(Ala)3-Cy5 41 was recognised and successfully incorporated by DNA polymerase in the primer I.
Figure 4.2 HPLC chromatogram of primer extension by Thermo Sequenase using dCTP-(A)$_3$-Cy5. Detection at 260 nm (red) and at 641 nm (blue).

4.3 Removal of the Reporter Group

Because the read length of a SBS technology is highly dependent on the efficiency of removal of the fluorophore from the extended primer, the enzymatic cleavage of the labelled peptide from the extension product II (Scheme 4.1 and 4.3) was investigated by HPLC and MALDI-TOF MS analysis (Figure 4.3).

Scheme 4.3 Left: Digestion of primer II by subtilisin. Right: Structure of primer III.

10 min of incubation at 37 °C of the primer II with the protease subtilisin A resulted in the disappearance of the peak at 7.9 min from the HPLC chromatogram and in the formation of two new peaks at 6.3 min (detected at 260 nm) and at 15.7 min (detected at 567 nm). The peak at 6.3 min was isolated and analysed by MALDI-TOF MS to produce three fragments (m/z 5865, 5936, 6007). The main peak (m/z 5865) resulted from cleavage between the aminopropargyl linker and the
labelled peptide while two small peaks were obtained from the cleavage at the sites P1' (m/z 5936) and P1 (m/z 6007), Figure 4.4. The peak at 15.7 min in the HPLC chromatogram was the dye-peptide cleaved from the extended primer. These data demonstrated that the reporter group (fluorophore and peptide) was completely released from the extended primer II during incubation with the protease subtilisin and consequently that no residual signal from the fluorophore was left on the extended primer III.

Figure 4.3 a) HPLC chromatogram of the digestion of primer II by subtilisin. Detection at 260 nm (blue) and at 565 nm (red); b) MALDI-TOF MS of oligonucleotide III.

Figure 4.4 Cleavage sites of the enzyme subtilisin for the fluorescent nucleotide 35.
4.4 Investigation of the Termination Properties of Enzymatically Cleavable Nucleotides

To verify the termination properties of the enzymatically cleavable nucleotides, primer extension was performed on a DNA template III containing a homopolymeric region with three consecutive bases "A" adjacent to the annealing site of the primer, Scheme 4.4. The primer and the template were incubated with the nucleotide dUTP-(Ala)_3-Rho and Thermo Sequenase, as reported in the section 4.2.

Template III and primer I sequences:

3' -GGCCTGTGTATCTCCTAGGTTGGCTCTGAAAGTACCACCATCCACTACAACTACATGTG - 5'

5' -AGAGGATCCAACCGAGAC-3'

Primer I

Template III

3' -AAAGTACC5'

Primer II

Template III

3' -AAAGTACC5'

Scheme 4.4 Left: Primer extension with dUTP-(A)_3-Rho 46 and the homopolymeric template III. Right: structure of primer II.

HPLC analysis of the reaction mixture revealed the presence of a peak at 7.1 min, which was postulated to be the primer extended with one nucleotide dUTP-(A)_3-Rho 46 as the retention time was similar to that of primer II (see figures 4.1 and 4.5). If multiple incorporations would have occurred, three peaks corresponding to the incorporation of 1, 2 and 3 nucleotides all absorbing at 565 nm, would be expected in the HPLC chromatogram.

Figure 4.5 HPLC chromatogram of a primer extension using dUTP-(A)_3-Rho and a homopolymeric template. Detection at 260 nm (blue) and at 565 nm (red)
Attempts to identify the extended oligonucleotide by MALDI-TOF MS were carried out; however no peak could be detected corresponding to any extension product due, likely, to the presence of salts and matrix-crystals in the sample spotted on the MALDI plate. To investigate the termination properties of the enzymatically cleavable nucleotides, another experiment was carried out by incubating a primer and the corresponding template with Thermo Sequenase and two protease cleavable nucleotides dUTP-(Ala)₃-Rho 46 and dCTP-(Ala)₃-Cy5 41 (Scheme 4.5). Template IV contained the nucleotides “A” and “G” next to the annealing site of the primer.

Template IV and primer I sequences:

3' -GGCCTGTGTTATCTCCTAGGTTGGCTCTGAGAGAGAGAGTCCACTACAACTACATGTG - 5'
5' -AGAGGATCCAACCGAGAC-3'

In the case of multiple incorporations, two peaks would be expected to be visualised in the HPLC chromatogram for the extension products: an extended primer corresponding to II absorbing at 565 nm and another extended primer absorbing at both 565 and 641 nm corresponding to V (Figure 4.6). In the case of termination by the first incorporated nucleotide, only one extension product absorbing at 565 nm and corresponding to the primer II should be found in the HPLC chromatogram.

Scheme 4.5 Primer extension with dUTP-(A)₃-Rho and dCTP-(A)₃-Cy5. Primer II is obtained when only the nucleotide dUTP-(A)₃-Rho is incorporated. Primer V is obtained when both nucleotides dUTP-(A)₃-Rho and dCTP-(A)₃-Cy5 are incorporated.

In the HPLC chromatogram of the extension reaction represented in Scheme 4.5, the major peak consisted of the termination product II obtained by incorporation of
the nucleotide dUTP-(A)₃-Rho (Figure 4.6). A small peak at 7.4 min absorbing at 641 nm was also found however the identity of this peak could not be determined by MALDI-TOF MS as the amount obtained was not sufficient. This peak might correspond to an impurity of the starting material 41 rather than to the extension product V. In this case a peak corresponding to the incorporation of both nucleotides would be expected to be found at a higher retention time due to the hydrophobicity of both dyes. In fact, the primer extended with 41, shown in Figure 4.2 had a retention time of 9.2 min.

Figure 4.6 HPLC chromatogram of DNA chain extension by Thermo Sequenase using Cy5-(A)₃-dCTP and dUTP-(A)₃ Rho B. Detection at 260 nm (grey), at 565 nm (red), at 641 nm (blue).

4.5 Conclusion

In summary, it has been proved that enzymatically cleavable nucleotides (such as dUTP-(Ala)₃-Rho and dCTP-(Ala)₃-Cy5) could be successfully incorporated by Thermo Sequenase DNA polymerase into a DNA primer and could be completely cleaved (within 10 min) by a protease to release the fluorophore. The termination properties of enzymatically cleavable nucleotides are important when four fluorescent nucleotides need to be used at the same time in sequencing by synthesis cycles. The termination properties of the reporter nucleotides were investigated by HPLC analysis. Preliminary results have shown that in the presence of homopolymeric templates or of a mixture of cleavable nucleotides, the major extension product consisted of a primer extended by a single nucleotide. Further tests need to be carried out to identify this extension product by MALDI-TOF. Encouraged by the results of primer extension and cleavage experiments in solution, presented in this chapter, enzymatically cleavable nucleotides were tested
in polymerase reactions on DNA microarrays. The results of these experiments will be described in the next chapter.
5 Chip-Based Interrogation of DNA Sequences with Enzymatically Cleavable Nucleotides

The purpose of this chapter is to describe the use of enzymatically cleavable nucleotides in chip-based polymerase reactions. Incorporation of these modified nucleotides into DNA primers immobilised on a solid support and protease-cleavage of the fluorophore are presented.

5.1 Introduction

The demand for ultra-high-throughput re-sequencing for personal medicine encouraged the development of non-electrophoretic platforms to enable scientists to sequence DNA in a massively parallel and highly automated fashion. DNA sequencing by synthesis (SBS) approaches combined with microfluidic systems, microarray technology and ultra-sensitive optics (e.g. SMRT technology developed by Pacific Biosciences and the SBS technologies developed by Helicos and Illumina) contributed to lower the cost of whole genome sequencing.

Therefore, once it had been proven that the nucleotide analogues were incorporated by a DNA polymerase in a primer extension reaction in solution (Chapter 4), it was necessary to implement the method on a microarray format for sequencing applications.

In the novel SBS approach, here proposed, unknown single stranded DNA's would be hybridised to an array of DNA primers immobilised on a chip (Figure 5.1-a). Four nucleotide analogues designed such that each is labelled with a unique fluorescent dye linked to the base through a peptide would be utilised (Figure 5.1-b). Upon addition of the four nucleotide analogues and a DNA polymerase, only the nucleotide analogue complementary to the next nucleotide on the template would be incorporated by the polymerase on each spot of the array. In this manner the first unknown nucleotide adjacent to the primer would be identified by virtue of the "colour" of the incorporated base. At the same time chain extension is temporarily terminated by inhibition of the DNA polymerase induced by the bulk of the fluorophore and the peptide.
After incorporation and detection, the “colour” is removed and the base “unblocked”. At this stage the DNA primer is ready for the next cycle of the incorporation to identify the next nucleotide sequence of the template DNA.

![Diagram](image)

**Figure 5.1** a) DNA SBS on chips using protease cleavable nucleotides; b) Structure of a protease cleavable nucleotide.

The experiments presented in this chapter were aimed to prove the principle of our new technique and optimise chip-based enzymatic assays by addressing the following issues:

1. Construction of a DNA microarray and optimisation in relation to DNA sequences (templates and primers) and reaction conditions.
2. Evaluation of enzyme activity and specificity on the array.
3. Evaluation of enzymatically cleavable nucleotides as reversible terminators.

5.2 DNA Microarrays: Templates

Most sequencing methods require an amplification step, so that the DNA to be sequenced is present in sufficient copy numbers to achieve the required signal. A recent advance in DNA template preparation is the emulsion-PCR technique which has been implemented in a number of SBS approaches (e.g. the pyrosequencing technology developed by 454 Life Science) (see Chapter 1).

The advantage of DNA sequencing by synthesis technologies is that sequence information can be also obtained from single DNA molecules. A number of companies (e.g. Helicos, Pacific Biosciences, Visigen Biotechnologies) have developed single molecule DNA sequencing technologies, reducing time, costs and avoiding potential biases (due to sequences that amplify poorly).

In both cases, the DNA fragments to be sequenced are physically separated in an array and subjected to repeated cycles of enzymatic manipulation to produce the sequence.

Because the main goal of the work here reported was to characterise fluorescent cleavable dNTPs in polymerase-mediated extension reactions for their ability to behave as DNA polymerization transient terminators, proof-of-the-principle experiments have been carried out using synthetic model oligonucleotides, instead of cloned DNA.

More details on the design of DNA primers (or “targets”) and DNA templates (or “probes”) will be provided in the section 5.4.1.

5.3 DNA microarrays: Surface and Immobilisation Chemistry

The design of our novel SBS approach required, first of all, the construction of DNA microarrays on which to perform polymerase reactions.

The quality of the microarrays is determined by various factors:

1. Immobilisation chemistry.
2. Target and probe density.
3. Target and probe design (length, melting temperature, etc.).
4. Buffer composition (pH, salt concentration, etc.).
5. Surface-compatibility with enzyme functionality.
6. Easy removal of enzyme and labelled compounds from the surface. The primary requirement in the immobilisation of DNA is that the coupling reaction must occur in high yields and in conditions compatible with DNA handling. Another ideal property required for the functional groups to be coupled (one from the DNA and the other on the surface) is the stability of the groups under aqueous conditions.

A variety of attachment chemistries have been used for the immobilisation of DNA on surfaces. Lindroos et al. have compared the immobilisation chemistries of different surface-oligonucleotide interactions and found that amino-modified nucleotides immobilised on the commercially available CodeLink™ slides (Figure 5.2) provided the best environment for solid phase minisequencing (Table 5.1).

<table>
<thead>
<tr>
<th>Slide type (glass slides)</th>
<th>5'-modification on primer</th>
<th>Functional group on slide</th>
<th>Relative signal intensity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isothiocyanate</td>
<td>NH$_2$-(CH$_2$)$_6$-O-(PO)$_2$-oligonucleotide</td>
<td>-N=C=S</td>
<td>1.0</td>
</tr>
<tr>
<td>'Superaldehyde'</td>
<td>NH$_2$-(CH$_2$)$_6$-O-(PO)$_2$-oligonucleotide</td>
<td>-CHO</td>
<td>0.09-0.20</td>
</tr>
<tr>
<td>Mercaptosilane</td>
<td>NH$_2$-(CH$_2$)$_6$-O-(PO)$_2$-oligonucleotide</td>
<td>-SH</td>
<td>1.9</td>
</tr>
<tr>
<td>EZ-RAYS™</td>
<td>CH$_2$=CH-CO-NH-R-oligonucleotide</td>
<td>3D with -SH groups</td>
<td>0.19</td>
</tr>
<tr>
<td>CodeLink™</td>
<td>NH$_2$-(CH$_2$)$_6$-O-(PO)$_2$-hydroxysuccinimidyl ester</td>
<td></td>
<td>7.7</td>
</tr>
<tr>
<td>Unmodified (glass slide)</td>
<td>none</td>
<td></td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 5.1 Efficiencies of the reaction between the 5' functional group on the oligonucleotide and the reactive groups on the slide. Table adapted from ref. 6. *After incorporation of labelled nucleotides

Experimental results (shown later in this chapter) confirmed the high quality and performance of these slides in minisequencing and SBS cycles.
The more affordable aldehyde slides (Genetix), in which DNA molecules were covalently bound to the glass surface through formation and reduction of a Schiff's base, showed poor results and lower quality data after primer extension reactions (Figure 5.3).

One main drawback of these slides is the difficulty of washing off fluorescent compounds sticking onto the surface. In particular, the enzymatically cleavable nucleotide dUTP-(A)₃-Rho, used in primer extension reactions, was difficult to remove from the surface of the chip, because of the hydrophobic nature of the dye. Since the dye Rhodamine B is soluble in aqueous solutions at acidic pH, the slides were treated with citrate/phosphate buffers ranging from pH 6.5 to pH 5. However, the dye could not be efficiently removed. The use of an organic solvent (such as 2-
propanol) in which the dye is highly soluble, did not improve the quality of the background.

This limitation was easily overcome using CodeLink™ slides which, thanks to the hydrophilicity of the coating, avoided non-specific dye binding and provided microarray data with high signal to noise ratios. An example of the quality of the images obtained with CodeLink™ slides is shown in Figure 5.4.

![Figure 5.4](image)

**Figure 5.4** Fluorescence images (scanned using a Cy3 filter) of CodeLink™ slides after primer extension with dUTP-(A)₃-Rho; a) before image processing; b) after image processing

Single nucleotide primer extension was performed on both CodeLink™ and aldehyde slides, and fluorescence intensities (average values for all the oligonucleotides on each slide) after single nucleotide primer extension were measured at $\lambda_{em} = 585$ nm using the FIPS software (LaVision BioTech). The comparison, shown in Figure 5.5, confirmed the results reported by Syvanen.¹⁸⁴

![Figure 5.5](image)

**Figure 5.5** Diagram of normalised fluorescence intensities acquired for CodeLink™ and aldehyde slides ($\lambda_{em} = 585$ nm) after primer extension with dUTP-(A)₃-Rho. The error bars represent the standard deviation of the mean (SEM).

In the course of this work, nucleotides labelled with various fluorescent dyes (carboxyfluorescein, rhodamine derivatives, cyanine dyes) having different
properties (hydrohobicity/hydrophilicity) have been tested on CodeLink™ slides. All dyes could be easily removed without the need for extensive washings.

5.4 DNA microarrays: printing and hybridisation

5.4.1 Design of target and probes sequences

The design of target and probe sequences is crucial for the success of primer extension reactions onto a solid surface. To evaluate and optimise our method, DNA sequences previously used in minisequencing experiments published in the literature were selected. Target/probe physical characteristics were analysed using the Oligo Analyzer 3.1 program and modified, where necessary, to improve the performance of hybridisation and incorporation assays.

The following DNA properties were considered for the design of DNA sequences:

1. Primer and template length.
2. Stability of DNA duplexes.
3. Cross-homology.

When designing oligonucleotides for DNA analysis on a microarray slide, it is important to ensure that all probes and targets hybridise at the same temperature. Additionally, to allow the formation of DNA duplexes, the hybridisation must occur at temperatures lower than the melting temperature (Tm) that is the temperature at which a double stranded DNA molecule dissociates into single strands. The Tm of DNA duplexes depends on the length and on the sequence of a DNA molecule. Generally, long heteroduplexes with a high GC content have higher Tm because of the higher number of interactions between base pairs.

DNA primers and templates chosen for primer extension experiments on microarray, consisted of 25-31mer oligonucleotides (with overlapping regions of 20-25 bp) with a GC content of ~40-60%. All selected probes and targets had a Tm in the range of 65-85 °C.

The hybridisation temperature must be carefully selected as a too low hybridisation temperature may lead to the formation of non-specific hybridisation products.
caused by a high number of base pair mismatches while a too high hybridisation temperature can result in low extension yields.\textsuperscript{186,187}

Taking into account all these factors, hybridisation and primer extensions were carried out at $\sim$55 °C. This temperature ensured hybridisation specificity and at the same time was well tolerated by the coating of CodeLink\textsuperscript{TM} slides.

When designing probes and targets for primer extension reactions, it is also important to avoid cross-hybridisation and the formation of secondary structures (e.g. homo- and heterodimers and hairpins) produced by intermolecular or intramolecular interactions of DNA molecules.\textsuperscript{187} The formation of secondary structures can be predicted using specific softwares, such as the Oligo Analyzer 3.1 program, and must be avoided. In fact, if primers and templates form secondary structures stable above the hybridisation temperature, they will be unable to hybridise and therefore primer extension will not occur.\textsuperscript{186}

5.4.2 Optimal target and probe density

The quality of DNA microarrays depends on two important factors: the optimal density of bound target molecules on the microarray surface and the optimal concentration of probe molecules that provide the strongest signal in a primer extension microarray assay.\textsuperscript{187}

When target and probe molecules hybridise each other on the microarray surface, there are important physical and/or steric properties that must be taken into account. High target concentration produces too many target molecules on the microarray surface causing quenching of fluorescence or preventing incoming probes from hybridising or DNA polymerase from extending the primer, thus reducing the fluorescent signal. On the other hand, insufficient target density determines a weaker fluorescent signal.\textsuperscript{187}

Probe and target concentration together with enzyme and nucleotide concentration, can also affect the specificity of nucleotide incorporation by DNA polymerases. To reduce the probability of misincorporations, dual colour primer extension reactions have been carried out using different concentrations of primers, templates, enzyme and nucleotides (see Appendix).
5.4.3 Immobilisation of DNA Primers on CodeLink™ slides and Hybridisation

5.4.3.1 Printing

DNA microarrays were constructed with synthetic DNA primers modified in their 5’-ends with amino groups in order to be linked to NHS ester-derivatised slides. DNA primers were first dissolved in a 50 mM sodium phosphate buffer at a concentration of 20 μM and then coupled to CodeLink™ slides at pH 8.5 in a humid environment, after contact printing with a Genetix QArray robot. Because the NHS ester groups onto the surface are moisture sensitive, printing was performed in an environment with relative humidity < 50 %. This humidity ensures stability of NHS esters and, at the same time, prevents samples from drying due to excessive evaporation of water.

After immobilisation of DNA, CodeLink™ slides were treated with a solution 50 mM of ethanolamine to “block” the un-reacted NHS groups on the surface and then washed at 50 °C with a 4 × SSC (saline-sodium citrate) buffer containing SDS (sodium dodecyl sulfate). SDS and washings at higher temperature are commonly used to reduce non-specific binding of DNA. After washing, the slides were dried by centrifugation at 4000 rpm for 5 min. The protocol proposed by the manufacturer, suggests to centrifuge at 800 rpm for 3 min. However, this centrifugation time was not sufficient to completely dry the surface of the slides, while higher speed resulted in their breakage.

A 5'-amino modified oligonucleotide, labelled in the 3’ position with fluorescein, was used as a control in order to check the quality of printing on the surface. Fluorescence images of the slides were scanned using the appropriate filter for the detection of fluorescein in order to show the fluorescence signal of the labelled DNA primers. An example of CodeLink™ slide printed with labelled DNA primers is shown in Figure 5.6.
5.4.3.2 Hybridisation

DNA primers immobilised onto CodeLink™ slides were hybridised to complementary DNA templates with known sequences. A preheating step at 85 °C was carried out in order to remove any secondary structure formed in the probe sample. As mentioned earlier, a hybridisation temperature of 55 °C was found to provide good results for all the primers selected for primer extension experiments. Therefore, the hybridisation was performed at 55 °C for 3 h in an automated hybridisation system using the following hybridisation buffer (Genetix): 50% formamide, 2 mg/mL herring sperm DNA, 0.4% SDS and 8.4 × SSC. The buffered herring sperm DNA is composed of small DNA fragments which blocks non-specific binding during prehybridisation and nucleic acid hybridisation. Successful hybridisation could be determined by using an oligonucleotide template labelled in the 5’ position with fluorescein, as hybridisation control. Fluorescence images of the slide were scanned using a FITC filter to verify the success of the hybridisation step.

The DNA primers printed onto the slide shown in Figure 5.6 were hybridised to a DNA template labelled in the 5’ position with fluorescein. The fluorescence image of this slide after hybridisation is shown in Figure 5.7. Although, primers and templates of the subarray B were both labelled with fluorescein, the average fluorescence intensity value (5.2 $10^5$) of the correspondent signals was very similar to the values obtained for the array A (4.9 $10^5$).
Table 5.2 Oligonucleotides sequences and their positions as shown in Figure 5.6 and 5.7. Fl = fluorescein

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-H2N- TTT TAT GAC ACC GTC ATC AGC AG-3'</td>
<td>A</td>
</tr>
<tr>
<td>3'-ATA CTG TGG CAG TAG TCG TCA TCA G-Fl-5'</td>
<td></td>
</tr>
<tr>
<td>5'-H2N- TTT TAT GAC ACC GTC ATC AGC AG-Fl-3'</td>
<td>B</td>
</tr>
<tr>
<td>3'-ATA CTG TGG CAG TAG TCG TCA TCA G-Fl-5'</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.7 Left: Fluorescence image (scanned using a FITC filter) of a CodeLink™ slide after hybridisation. A) unlabelled primers hybridised to templates labelled in the 5’ position, B) DNA primers labelled with fluorescein in the 3’ position and hybridised to DNA templates labelled with fluorescein in the 5’ position; Right: a) schematic representation of a DNA primer (white bar) immobilised onto a microarray surface (in grey) and hybridised to a DNA template labelled in the 5’ position with fluorescein (white bar with a green star).

5.5 Single Nucleotide Primer extension using Fluorescent Dideoxynucleotides

Before testing the enzymatically cleavable fluorescent nucleotides in sequencing by synthesis assays, solid phase minisequencing experiments were carried out to test the reaction conditions for primer extension and gain important informations in relation to:

1. Discrimination, by our detection system, of differently labelled oligonucleotides after incorporation of fluorescent nucleotides.
2. Specificity of hybridisation/ incorporation.
3. Efficiency of incorporation of fluorescently labelled nucleotides by Thermo Sequenase DNA polymerase.

A total of 102 DNA primers were coupled to the surface and used to prime 102 synthetic DNA templates corresponding to gene regions containing SNPs. Each
slide contained four duplicates of an array of 102 heteroduplexes. The primed templates were incubated in an automated hybridisation system with Thermo Sequenase DNA polymerase (1U/μL) and four fluorescent dideoxynucleotides (2.5 μM) labelled with four fluorophores, each coding for a different base (A, C, T, G): Texas Red-ddATP, Fluorescein-ddCTP, Cy5-ddUTP and TAMRA-ddGTP. After incubation and washing, scans of the microarray surface with a microarray scanner equipped with four different filter settings allowed for detection of the incorporated base for each of the DNA templates (Figures 5.8-5.15).

The four labelled dideoxynucleotides were chosen on the basis of the following criteria: 1) four distinguishable colors as fluorescent labels; 2) commercial availability; 3) previous works published in the literature.

5.5.1 Optical Read-Out

Biochips were scanned using a microarray reader featuring a white light source, four colour filters and a CCD camera detection system. Unlike lasers, which have a defined monochromatic emission, white light sources emit many colours simultaneously and are thus said polychromatic.

The following filters were used to scan the microarray slides (excitation wavelength ± bandwidth of transmission/emission wavelength ± bandwidth of transmission):

- FITC: (480 nm ± 40/ 535 nm ± 50);
- Cy 3: (546 nm ± 12×/ 585 nm ±40);
- Texas-Red: (560 nm ± 55/ 645 nm ± 75);
- Cy 5: (640 nm ± 20/ 680 nm ± 30).

Because the emission wavelength for a given dye is longer than the excitation wavelength and because the emission spectra often extends over 50-100 nm, emitted light or signal from one channel can be erroneously detected in a second channel.

A common standard for cross-talk is ≤ 10% from channel to channel, absolutely necessary for the detection of different labels. As shown in Figures 5.13-5.16, cross-talk occurred when the microarray slides were scanned using TAMRA and Texas Red filters because TAMRA and Texas Red dyes have very similar spectra. Normalised fluorescence intensities were measured for all the extended primers using the FIPS software (LaVision BioTech) after scanning the slides with four different filters (Cy5, FITC, Texas Red and TAMRA).

Figures 5.9 and 5.11 show that the normalised fluorescence intensities measured for the primers extended with Cy5-ddUTP and Fluorescein-ddCTP are distinguishable.
from the signals given by the primers extended with Texas Red-ddATP or TAMRA-ddGTP, which have shown normalised fluorescence intensities $\leq 0.2$.

**Figure 5.8** Left: Fluorescence image (scanned using a Cy5 filter) of DNA duplexes after incorporation of Cy5-ddUTP. Right: expected pattern for primer extension (the letters A, C, G, T indicate the base expected to be incorporated in the DNA duplex immobilised onto the slide). F = fluorescein (printing control).

**Figure 5.9** Normalised Fluorescence Intensities ($\lambda_{em} = 680$ nm) of 102 heteroduplexes after primer extension with four fluorescent dideoxynucleotides. The blue squares correspond to the fluorescence intensities of primers extended with Cy5-ddUTP. The grey diamonds correspond to the fluorescence intensities of primers extended with the other three fluorescent dideoxynucleotides.
Figure 5.10 Left: Fluorescence image (scanned using a FITC filter) of DNA duplexes after incorporation of FITC-ddCTP. Right: expected pattern for primer extension (the letters A, C, G, T indicate the base expected to be incorporated in the DNA duplex immobilised onto the slide). F = fluorescein (printing control).

Figure 5.11 Normalised Fluorescence Intensities ($\lambda_{\text{exc}} = 480$ nm) of 102 heteroduplexes after primer extension with four fluorescent dideoxynucleotides. The green squares correspond to the fluorescence intensities of primers extended with FITC-ddCTP. The grey diamonds correspond to the fluorescence intensities of primers extended with the other three fluorescent dideoxynucleotides.

On the other hand, Figures 5.13 and 5.15 show that cross-talk occurred when the slide was scanned using the Texas Red and TAMRA filters. Cross-talk is usually minimized using a number of strategies, including a careful choice of excitation/emission filters, use of lasers (monochromatic emission instead of polychromatic light) and appropriate dye sets. \(^{187}\)
Figure 5.12 Left: Fluorescence image (scanned using a Cy3 filter) of DNA duplexes after incorporation of TAMRA-ddGTP. Right: expected pattern for primer extension (the letters A, C, G, T indicate the base expected to be incorporated in the DNA duplex immobilised onto the slide).

Figure 5.13 Normalised Fluorescence Intensities ($\lambda_{em} = 585$ nm) of 102 heteroduplexes after primer extension with four fluorescent dideoxynucleotides. The orange squares correspond to the fluorescence intensities of primers extended with TAMRA-ddGTP. The grey diamonds correspond to the fluorescence intensities of primers extended with the other three dideoxynucleotides.

To reduce cross-talk between TAMRA and Texas Red channels, a different filter set emitter ($602\pm13\times/631\pm23$nm) was selected for Texas Red. However, the two dyes could still not be discriminated. As confirmed by the filter manufacturer (Chroma), TAMRA and Texas-Red are difficult to distinguish using microarray imagers with white light source.

A similar dye set (Cy5, Texas Red, TAMRA, R-110 G) was used by Syvanen in a four color multiplexed minisequencing experiment in 2002\textsuperscript{136}. In this work, cross-talk between different channels was reduced using lasers instead of a light source.
imager. However, they also reported that the fluorescence signals of primers extended with a ddNTP labelled with TAMRA could be detected when a Texas Red filter was used.

![Fluorescence Image (scanned using a Texas Red filter) of DNA duplexes after incorporation of TxR-ddATP. Right: expected pattern for primer extension (the letters A, C, G, T indicate the base expected to be incorporated in the DNA duplex immobilised onto the slide).](image)

**Figure 5.14** Fluorescence image (scanned using a Texas Red filter) of DNA duplexes after incorporation of TxR-ddATP. Right: expected pattern for primer extension (the letters A, C, G, T indicate the base expected to be incorporated in the DNA duplex immobilised onto the slide).

![Normalised Fluorescence Intensities (λ_em= 645 nm) of 102 heteroduplexes after primer extension with four fluorescent dideoxynucleotides. The pink squares correspond to the fluorescence intensities of primers extended with TxR-ddATP. The grey diamonds correspond to the fluorescence intensities of primers extended with the other three dideoxynucleotides.](image)

**Figure 5.15** Normalised Fluorescence Intensities (λ_em= 645 nm) of 102 heteroduplexes after primer extension with four fluorescent dideoxynucleotides. The pink squares correspond to the fluorescence intensities of primers extended with TxR-ddATP. The grey diamonds correspond to the fluorescence intensities of primers extended with the other three dideoxynucleotides.

### 5.5.2 Analysis of enzymatic primer extension

The single nucleotide primer extension assay was primarily performed to gain experience on primer extension on chips and to provide information useful for our
next goal: primer extension using protease cleavable nucleotides. The enzymatic assay showed that:

1. The hybridisation was specific
2. Our microarray imager was able to distinguish Cy5, Fluorescein and TAMRA but that cross-talk occurred between Texas Red and TAMRA channels.
3. Thermo Sequenase DNA polymerase was able to incorporate ddCTP and ddUTP accurately. Incorporation of ddGTP and ddATP was more difficult to evaluate due to cross-talk between channels.

As mentioned in Chapter 2 and 3, the dyes originally selected for labelling our protease cleavable nucleotides were Cy3, Cy3.5, Cy5, Cy5.5. These dyes have different absorption/emission spectra, similar chemical/physical properties and high fluorescence intensity. In addition, they could be easily synthesized following the protocol described in Chapter 2. However, the single nucleotide primer extension assay suggested that this set of cyanine dyes would require mathematical corrections and post-processing analysis due to overlap of their signals in different channels (Cy3 and Cy3.5 have similar spectra to TAMRA and Texas Red). The same observation was also reported in the work published by Richert where the same dye set was used for labelling specially modified phosphoramidates.137

5.6 Primer Extension using Protease Cleavable Fluorescent Nucleotides: proof-of-principle experiments.

5.6.1 Incorporation of dUTP-(A)₃-Rho

Encouraged by the results obtained in the solution tests (Chapter 4) and on the microarray using fluorescent ddNTPs, enzymatically cleavable nucleotides were tested in chip-based primer extension reactions. Arrays of four short synthetic primer/template duplexes with known sequences were used to demonstrate the feasibility of the technique. Each primer (Table 5.3) was spotted in a 10 × 10 pattern, in duplicate. The fluorescent primer used as printing control was detected using a FITC filter (Figure 5.16-a). Hybridisation with complementary synthetic DNA templates was performed in an automated hybridisation chamber. The
fluorescence signals of fluorescently labelled templates used as hybridisation controls are shown in Figure 5.16-b. Then, the primed templates were incubated at 55 °C for 15 min with Thermo Sequenase DNA polymerase (1U/1μL) and dUTP-(A)₃-Rho (2.5 μM) in Thermo Sequenase reaction buffer (1×).

At the beginning of our investigation, the incubation time for primer extension in our experiments was 1h. However, we found that an incubation time of 15 min was adequate for all primer extension experiments.

Scanning using the appropriate filter for the detection of rhodamine (Figure 5.16-c) showed that all primers were efficiently extended except the controls in position P1 and P3 (primers labelled with fluorescein in the 3' position). The data shown in Figure 5.16 demonstrated qualitatively that Thermo Sequenase polymerase was
able to incorporate our modified nucleotide in all primers and that, as expected, the labelled nucleotide did not bind in a non-specific manner to the DNA duplexes.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3'</td>
<td>P1/P3</td>
</tr>
<tr>
<td>3'-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5'</td>
<td></td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT CAG GAC CTA GAA CGG GCA GC-CU-3'</td>
<td></td>
</tr>
<tr>
<td>3'-GTC CTG GAT CTT GCC GTT CGT CGA TCA G-5'</td>
<td>P2/P4</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3'</td>
<td></td>
</tr>
<tr>
<td>3'-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5'</td>
<td>P5/P7</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT CCT TCT TGC TGG CAC CCA ATU-3'</td>
<td></td>
</tr>
<tr>
<td>3'-GGA AGA ACG ACC GTG GGT TAA TCA G-5'</td>
<td>P6/P8</td>
</tr>
</tbody>
</table>

Table 5.3 Oligonucleotide Sequences and their positions as shown in Figure 5.17. Fl = fluorescein.

5.6.2 Two Steps in One: Hybridisation & Extension

All polymerase reactions, previously described, were performed with double stranded DNA already immobilised onto the chip. Each primer extension experiment consisted of three steps:

1. printing of primers onto the chip
2. hybridisation of DNA templates
3. primer extension by DNA polymerase.

To reduce the number of steps necessary for each experiment, steps 2 and 3 were carried out together by adding to the slide a reaction mixture containing DNA templates, enzymatically cleavable nucleotides and Thermo Sequenase DNA polymerase in its reaction buffer. DNA primers and templates used in this experiment are shown in Table 5.4.

The reaction mixture was first preheated at 90 °C and then added to the slide. Annealing of DNA templates and incorporation of dUTP-(A)₂-Rho were carried out at 55 °C for 1h. Afterwards, the slide was cooled down to 35 °C and scanned using the appropriate filter for the detection of rhodamine (Figure 5.17). The primers printed in the subarray P1 were modified in the 3' position with fluorescein in order to control the printing step; therefore they were not extended during polymerase reaction.
Figure 5.17 Left: fluorescence image of a CodeLink slide (scanned using a Cy3 filter) after primer extension with dUTP-(A)₃-Rho. Right: expected pattern. The primers are immobilised on the slide (in brown) and hybridised to complementary DNA templates. Primers and templates used as printing and hybridisation controls are labelled with fluorescein (green star). Rhodamine is represented by a red star.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-H₂N-TTT TTT GAC ACC GTC ATC AGC AG-Fl-3' 3'-ATA CTG TGG CAG TAG TCG TCA TCA G-Fl-5'</td>
<td>P1</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT CAG GAC CTA GAA CGG GCA GCU-3' 3'-GTC CTG GAT CTT GCC CGT CGA TCA G-5'</td>
<td>P2</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT GAC ACC GTC ATC AGC AGU-3' 3'-ATA CTG TGG CAG TAG TCG TCA TCA G-Fl-5'</td>
<td>P3</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT CCT TCT TGC TGG CAC CCA ATU-3' 3'-GGA AGA ACG ACC GTG GGT TAA TCA G-5'</td>
<td>P4</td>
</tr>
</tbody>
</table>

Table 5.4 Oligonucleotide Sequences and their positions as shown in Figure 5.17. Fl = fluorescein

Generally, DNA Sequencing by Synthesis requires all the nucleotide analogues labelled with different dyes to be added simultaneously to the slide for decoding the sequence of the DNA templates.

Before reaching this phase, it is critical to prove that each nucleotide analogue is incorporated faithfully by DNA polymerases. Experiments were therefore carried out aimed at evaluating the performance of the sequencing by synthesis method using a unique label and adding one base at the time. After proving that dUTP-(A)₃-Rho could be incorporated into a growing DNA strand, single nucleotide primer extension was performed using nucleotide analogues labelled with different dyes (Cy5, Cy3, and FITC). All these nucleotides could be efficiently incorporated by the polymerase Thermo Sequenase.
5.6.3 Incorporation of dCTP-(A)₃-FITC

The polymerase reaction was also carried out with the nucleotide dCTP-(A)₃-FITC. The primers (Table 5.5) immobilised onto the slide and hybridised to complementary DNA templates were extended by Thermo Sequenase DNA polymerase, as described in the section 5.6.2. After washing the slide with appropriate buffers, the extension of the primer by the fluorescent nucleotide dCTP was confirmed by scanning the slide, using a FITC filter (Figure 5.18). The primers immobilised in the sub-array P1 were not extended during the polymerase reaction. The fluorescent signals corresponding to this sub-array were attributed to the primers used as printing control (primers modified in the 3' position with fluorescein) and to the complementary DNA templates (labelled in the 5' position).

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3' 3'-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5'</td>
<td>P1</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT GGT GGA GAA CTG CCA GCT C-3' 3'-CCA CCT CTT GAC GGT CGA GAC TCG GAC G-5'</td>
<td>P2</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AGC-3' 3'-ATA CTG TGG CAG TAG TCG TCG TCA G-F1-5'</td>
<td>P3</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT TCT GGA TGA ATA CTG CAG CTG C-3' 3'-AGA CCT ACT TAT GAC GTC GAC GAC ATC ATC G-5'</td>
<td>P4</td>
</tr>
</tbody>
</table>

Table 5.5 Oligonucleotide Sequences and their positions as shown in Figure 5.18. Fl = fluorescein

Figure 5.18 Left: fluorescence image of a CodeLink slide (scanned using a FITC filter) after primer extension with dCTP-(A)₃-FITC. Right: expected pattern. The primers are immobilised on the slide (in brown) and hybridised to complementary DNA templates. The primers printed in the sub-array P1 were used as printing controls and were modified in the 3' position with fluorescein. Fluorescein is represented by a green star.
5.6.4 Incorporation of dCTP-(A)₃-Cy5 (hydrophobic)

Primer extension was carried out also using dCTP-(A)₃-Cy5 and Thermo Sequenase DNA polymerase, in the same conditions described above. This experiment was aimed at evaluating the incorporation of dCTP-(A)₃-Cy5 by a DNA polymerase for application in dual color polymerase reactions and multiple extensions. dCTP-(A)₃-Cy5 was preferred to dCTP-(A)₃-FITC as it was available in higher amounts. In addition, pH sensitivity and photo bleaching made 5(6)-carboxyfluorescein less suitable than Cy dyes for this application.

Preliminary tests carried out using dCTP-(A)₃-Cy5 (labelled with Cy5 hydrophobic) showed that non specific binding to DNA occurred during incorporation.

Figure 5.19 Left: fluorescence image of a CodeLink slide (scanned using a Cy5 filter) after primer extension with dCTP-(A)₃-Cy5. Right: expected pattern. The primers are immobilised on the slide (in brown) and hybridised to complementary DNA templates. Primers and templates used as printing and hybridisation controls are labelled with fluorescein (green star). Cy5 is represented by a blue star.

Scanning of the chip, using the filter for the detection of Cy5, revealed that the primers in the subarray P1, used as printing control, were also fluorescent, though they could not be extended because of their modification with fluorescein at the 3' end (Figure 5.19). The average fluorescence intensities of all primers are shown in Figure 5.20.
After incorporation, the slides were washed with buffers at different stringency (salt concentration), but non-specific binding could not be completely removed. It is known that cyanine dyes are able to interact with DNA by intercalation but also by groove binding. Although, the presence of two methyl groups in the indolenine ring creates steric hindrance and therefore is expected to make difficult the intercalation of the dye in the minor groove. The dye used for the labelling of the nucleotide Cy5(A)₃-dCTP had a single positive charge which might have favoured electrostatic interactions with the negatively charged DNA. In addition, the hydrophobic nature of the dye might have favoured the interaction with the bases.
5.7 Primer Extension using Protease Cleavable Fluorescent Nucleotides: Optimisation

To avoid non-specific interaction of dyes with immobilised DNA, water soluble cyanine dyes bearing sulfonate groups (negative charges) on the indolenine ring were prepared. Their synthesis has been extensively discussed in Chapter 2. As mentioned, the water soluble cyanine dyes Cy3, Cy3.5, Cy5 and Cy5.5 were chosen as fluorescent labels for our nucleotide analogues.

However, the synthesis and purification of all the range of water soluble cyanine dyes is difficult and time-consuming and, at this stage of the project, the solid phase protocol for their preparation was not yet optimised. Therefore, before proceeding to the synthesis of all hydrophilic cyanine dyes needed for the labelling of the four nucleotides, it was decided first to carry out preliminary tests to verify the feasibility of using two enzymatically cleavable nucleotides labelled with the sulfonated cyanine dyes Cy3 and Cy5 and to optimise the protocol for dual color primer extension reactions on chip.

5.7.1 Dual color primer extension

A full SBS-based approach requires all four reversible terminators to be added simultaneously to the slide for decoding the DNA sequences by identifying the dye detected at each cycle of the DNA extension.

Fluorescent cleavable dNTPs were evaluated for their ability to behave as DNA polymerization “deblockable” terminators in dual color extension experiments. These experiments would give important insights towards the improvements needed for reducing misincorporation and other undesired events.

It was verified first that the enzymatically cleavable nucleotides labelled with sulfonated cyanine dyes were not binding to DNA in a non-specific way.

Two DNA primers (Table 5.6) were printed onto a CodeLink™ slide (6 spots for each primer) and then hybridised to complementary DNA templates.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-H₂N-TTT TTT CAG GAC CTA GAA CGG GCA GCA-F1-3' 3'-GTC CTG GAT CCT GCC GGT CGA GAG AGA GAG-F1-5'</td>
<td>P1</td>
</tr>
<tr>
<td>5'-H₂N-TTT TTT CAG GAC CTA GAA CGG GCA GCU-3' 3'-GTC CTG GAT CCT GCC GGT CGA GAG AGA GAG-F1-5'</td>
<td>P2</td>
</tr>
</tbody>
</table>

Table 5.6 DNA sequences and their positions as shown in Figure 5.21. Fl = fluorescein.
A DNA primer and a DNA template labelled with fluorescein were used as printing and hybridisation controls. Figures 5.21-A and 5.21-B show the fluorescence signals of the labelled nucleotides after printing and after hybridisation. The polymerase reaction was carried out in the presence of Thermo Sequenase DNA polymerase and the protease cleavable nucleotide dUTP-(A)₃-Cy3, labelled with sulfonated Cy3. The slide was then scanned using a Cy3 filter. The correspondent fluorescence image (Figure 5.21-C) shows that dUTP-(A)₃-Cy3 was correctly incorporated into the primer of the subarray P2 producing a red signal when excited at 546 nm. The primer of the subarray P1, labelled in the 3' position with fluorescein and used as control for the printing, was not extended and therefore was not detected in the red channel.

Figure 5.21 Left: A) Fluorescence image of the slide, scanned using a FITC filter, after primer printing. The fluorescent primers in P1 were modified with fluorescein in the 3' position; B) Fluorescence image of the slide, scanned using a FITC filter, after hybridisation. The fluorescent spots correspond to the primer and to the template used as printing and hybridisation controls; C) Fluorescence image of the slide, scanned using a Cy3 filter, after incorporation of dUTP-(A)₃-Cy3. Right: expected pattern. Fluorescein is represented by a green star and Cy3 by a red star.

The fluorescence intensities of the primers of the subarrays P1 and P2, measured after scanning the slide with a Cy3 filter, are shown in Figure 5.22.

To verify the ability of Thermo Sequenase DNA polymerase to incorporate the correct nucleotide complementary to the corresponding base on the template, seven different primers (Table 5.7) were immobilised on the CodeLink™ slide, hybridised to complementary DNA templates and extended with a mixture of dUTP-(A)₃-Cy3 and dCTP-(A)₃-Cy5 labelled with sulfonated cyanine dyes.
The incorporation of dUTP-(A)$_3$-Cy3 into the DNA primers was detected using a Cy3 filter (Figure 5.23-A). The incorporation of dCTP-(A)$_3$-Cy5 was detected using a Cy5 filter (Figure 5.23-B). The DNA polymerase Thermo Sequenase incorporated the base U with high specificity; only the primers (in P1, P2, P4, P6) hybridised to DNA templates containing the base “A” next to the annealing site were extended, as expected (Figure 5.23-A). The sequences of the DNA duplexes in P1-P7 are shown in Table 5.7 (the base expected to be incorporated in each primer is shown in colour).

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-H$_3$N-TTT TTT CAG GAC CTA GAA CGG GCA GC-3’</td>
<td>P1</td>
</tr>
<tr>
<td>3’-GTC CTG GAT CTT GCC GTG CGA GAG AGA GAG-5’</td>
<td></td>
</tr>
<tr>
<td>5’-H$_3$N-TTT TTT CCT TCT TGG CAC CCA ATU-3’</td>
<td>P2</td>
</tr>
<tr>
<td>3’-GGA AGA ACG ACC GTG GTT TAA GAG AGA GAG-5’</td>
<td></td>
</tr>
<tr>
<td>5’-H$_3$N-TTT TTT GGT GGA GAA CTG CCA GCT C-3’</td>
<td>P3</td>
</tr>
<tr>
<td>3’-CCA CCT GAC GGT CGA GAC TCG GAC G-5’</td>
<td></td>
</tr>
<tr>
<td>5’-H$_3$N-TTT TTT AAA TAC TCA CCA AAC TGC CCT CU-3’</td>
<td>P4</td>
</tr>
<tr>
<td>3’-TTT ATG AGT GGT TTG ACG GGA GAG ACT ACT ACG A-5’</td>
<td></td>
</tr>
<tr>
<td>5’-H$_3$N-TTT TTT TTT TTT GGT GGT GGA C-3’</td>
<td>P5</td>
</tr>
<tr>
<td>3’-AAA AAG AGC GAA CCA CCT GCA CAT CGT-5’</td>
<td></td>
</tr>
<tr>
<td>5’-H$_3$N-TTT TTT TTT TGG TGG GAA TTC CCA GCA TTC U-3’</td>
<td>P6</td>
</tr>
<tr>
<td>3’-ACC AAC CTT AAG GGT CGT AAA ACT GAT ACT G-5’</td>
<td></td>
</tr>
<tr>
<td>5’-H$_3$N-TTT TTT TTT TTT CGG GGA AGA GTG GAA CAA TTT TTC G-3’</td>
<td>P7</td>
</tr>
<tr>
<td>3’-GCG CCT TCT CAC CTT GGT AAA AAG CTT CGT TGC T-5’</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7 DNA sequences used in dual colour primer extension and their positions as shown in Figure 5.23.

The fluorescence intensities of the extended primers, measured using the FIPS software (LaVision BioTech) are shown in Figure 5.24 ($\lambda_{em} = 585$ nm) and 5.25 ($\lambda_{em} = 680$ nm).
The only primers that were expected to incorporate the nucleotide dCTP-(A)_3-Cy5 during polymerase reaction and to produce a fluorescence signal with \( \lambda_{em} = 680 \) nm (Cy5 emission) were the ones printed in P3 and P5. However, also the primers in P1 and P2, produced a fluorescence signal when the slide was scanned using a Cy5 filter (Figure 5.23-B). The average fluorescence intensities calculated for the primers in P1 and P2 (1.44 \(10^5\) and 8.8 \(10^4\)) at \( \lambda_{em} = 680 \) nm are high, considering that they were not expected to be fluorescent in the blue channel and that the fluorescence intensities of the primers in P3 and P5 are relatively similar (1.3 \(10^5\) and 1.6 \(10^5\)) (Figure 5.24 and 5.25).

**Figure 5.23** Fluorescence image of a CodeLink slide after dual colour primer extension. A) scanned using a Cy3 filter to show the incorporation of dUTP-(A)_3-Cy3; B) scanned using a Cy5 filter to show the incorporation of dCTP-(A)_3-Cy5. P1-P7 indicate the primers immobilised onto the slide. Each primer is printed in 6 duplicates on the same row.

There are different possible explanations to this result. Reagent purity, polymerase fidelity, mispriming, non specific binding of dyes, and multiple incorporations (see sequences in P1 and P2, Table 5.7) can all be source of false-positive signals in primer extension experiments.

One possible explanation is that multiple incorporations occurred for the primers in P1 and P2. Indeed, these two primers were hybridised to DNA templates containing repeats of identical adjacent bases (AG) (Table 5.7), therefore there is a possibility that they might have incorporated both nucleotides dUTP-(A)_3-Cy3 and dCTP-(A)_3-Cy5.
Figure 5.24 Normalised fluorescence intensities ($\lambda_{em} = 585$ nm) of the primers in P1-P7. The error bars represent SEM.

Figure 5.25 Normalised fluorescence intensities ($\lambda_{em} = 680$ nm) of the primers in P1-P7. The error bars represent SEM.

However, the results reported in the Chapter 4 have shown that the enzymatically cleavable nucleotides have termination properties. In addition, the primers P3 and P4 (Table 5.7), hybridised to DNA templates containing the bases “AG” or “GA” next to the annealing site, were detected predominantly either at $\lambda_{em} = 585$ nm or at $\lambda_{em} = 680$ nm, proving that they did not incorporate both nucleotides.

An alternative explanation to the result obtained in the dual colour primer extension experiment is that misincorporation of nucleotides by DNA polymerase occurred for the primers in P1 and P2. Misincorporations are very frequent in polymerase reactions and they can be sequence-dependent.\textsuperscript{190,191}
During primer extension reactions, DNA polymerases incorporate preferably complementary bases rather than non-complementary bases. In addition, they have exonuclease activity; therefore they can remove misincorporated nucleotides. However, DNA polymerases with exonuclease proofreading activity cannot be used in SBS\textsuperscript{66} (and in general, in DNA sequencing) because sequencing primers would be digested and labelled nucleotides would not be efficiently incorporated. Therefore Thermo Sequenase,\textsuperscript{192,193} an exonuclease-deficient DNA polymerase\textsuperscript{194} with mutations evolved to efficiently incorporate labelled nucleotides, was selected. Thermo Sequenase is generally used in combination with a thermostable inorganic pyrophosphatase, which is needed to remove the pyrophosphate produced during extension reactions. Published works showed that, by optimising reaction time and conditions, the likelihood of misincorporation can be reduced and polymerase fidelity improved.\textsuperscript{66,189,195} In addition, literature\textsuperscript{66,189} and our experimental results suggested that non-specific events, such as misincorporation can be reduced when multiple DNA sequences are used or when all dNTPs are added simultaneously.

To reduce the probability of misincorporations, dual color extension experiments were carried out using different concentrations of primers, templates, nucleotides and enzyme. These factors\textsuperscript{195} are known to affect the enzyme fidelity. In addition, as Thermo Sequenase is relatively expensive and our nucleotides are not commercially available, it was desirable to find the minimum amount of enzyme and nucleotides that provided acceptable results. The following conditions provided the best results in terms of signal intensity and specificity: 20 μM solution of primer for printing, 2 μM solution of DNA template, 1U/μL of Thermo Sequenase and 2.5 μM solution of nucleotide. Experimental results of concentration gradient experiments are reported in the Appendix.

5.8 Removal of the Fluorophore and Second extension

Complete removal of the fluorophore from extension products is critical for the successful application of enzymatically cleavable nucleotides in SBS.\textsuperscript{43} HPLC and MALDI-TOF analyses proved that the fluorophore Rhodamine B could be released completely after incubation of the extended primer with the protease subtilisin at 37 °C for 10 min (Chapter 4).
Encouraged by this result, the enzymatic cleavage of the fluorophore from the labelled nucleotides, after their incorporation in DNA primers immobilised on CodeLink™ slides, was investigated. Our aim was to prove that the enzyme subtilisin:

1. was compatible with the surface of CodeLink™ slides and the reaction conditions
2. could efficiently and rapidly remove the fluorophore from extended primers to allow the interrogation of next positions in the template
3. could be efficiently removed from the slide by washing, to avoid digestion of Thermo Sequenase during the next cycle of the polymerase reaction

DNA duplexes (Table 5.8) immobilised on CodeLink™ slides were treated with Thermo Sequenase DNA polymerase and dUTP-(A),Cy3. The primers in P2, P3, P4, P5 hybridised to DNA templates, having the base “A” next to the annealing site, were extended producing a fluorescent signal (λem = 585 nm) when scanned with a Cy3 filter (Figure 5.26-A). The primer P1 was not extended because it was modified in the 3’ position with fluorescein (printing control) while the primer P6 had the base “C” next to the annealing site, therefore could not incorporate the base “U” (Table 5.8).

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3’ 3’-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5’</td>
<td>P1</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AGUA-3’ 3’-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5’</td>
<td>P2</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT TTT AAA TAC TCA CCA AAC TGC CCT CGG 3’ 3’-TTT ATG AGT GGT TGG AGC GGA GAC TAC TAC GA-5’</td>
<td>P3</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT GTG GAG TGT TGG TTT ACC CCC CCU A-3’ 3’-CAC CTC ACA AAC AAG TGT GGG GGA TAC GAC TAC-5’</td>
<td>P4</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT CCT TCT TGC TGG CAC CCA ATU C-3’ 3’-GGA AGA ACG ACC GTG GGT TAA GAG AGA GAG-5’</td>
<td>P5</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC G-3’ 3’-GCC CCT TCT CAC CTT GTT AAA AAG CTT CTG TGC T-5’</td>
<td>P6</td>
</tr>
</tbody>
</table>

Table 5.8 DNA sequences and their positions as shown in Figure 5.26. The bases expected to be incorporated are shown in colour. Fl = fluorescein.

The slide was then incubated with the protease subtilisin, in an automated system (GenHyb4) for 15 min at 37 °C. The fluorophore was efficiently removed (≥ 95 % efficiency) and the red signal from Cy3 was almost completely removed, producing an image with fluorescence levels close to background (Figures 5.26-B and 5.27).

To verify that primer extension could take place again after incubation of the slide with a protease and that the protease treatment could not interfere with next
incorporation, a polymerase reaction was carried out with dCTP-(A)$_3$-Cy5 (labelled with Cy5 sulfonated) and Thermo Sequenase. The nucleotide was incorporated in a specific manner only into the primer P5 (as expected) producing a strong signal in the blue channel (Figure 5.26-C). Afterwards, the fluorophore was removed by protease cleavage (Figure 5.26-D).

![Fluorescence image of a CodeLink slide after primer extension with dUTP-(A)$_3$-Cy3 ($\lambda_{em} = 585$ nm); B) after protease cleavage ($\lambda_{em} = 585$ nm); C) after primer extension with dCTP-(A)$_3$-Cy5 ($\lambda_{em} = 680$ nm); D) protease cleavage ($\lambda_{em} = 680$ nm).](image)

The fluorescence intensities were measured for all primers at each step using the FIPS software (La Vision BioTech) (Figure 5.27).

![Fluorescence intensities of the primers in P1-P6: after the first extension (red bars) and after the first protease cleavage (pink bars) ($\lambda_{em} = 585$ nm); after the second extension (light blue bars) and after the second cleavage (dark blue bars) ($\lambda_{em} = 680$ nm).](image)
5.9 Conclusion

In order to obtain an efficient SBS sequencing technology, the following requirements must be met:

1. Amplification of DNA with high throughput methods rather than standard cloning techniques.
2. Use of a format that allows each template to be probed multiple times and in parallel.
3. Nucleotides must be reversible terminators so that only a single nucleotide is added each step during Sequencing by Synthesis technology.
4. The reversible terminator must be efficiently incorporated and the fluorescent label easily removed after detection for subsequent nucleotide addition.
5. Washing and reagents additions between detection cycles must be rapid and simple.

In this chapter, we have proved that enzymatically cleavable nucleotides could be efficiently incorporated into growing DNA strands during primer extensions on chip, independently from the fluorophore used for the labelling. Furthermore, we have shown that the fluorescent label could be easily removed by incubation of the extension products with a protease to allow the next incorporations to take place. Two SBS cycles were run successfully, allowing the nucleotides dUTP-(A)3-Cy3 and dCTP-(A)3-Cy5 to be incorporated in two consecutive steps.

In addition, we have optimised the enzymatic assay through an opportune choice of the chip surface, fluorophores and DNA sequences.

These preliminary experiments proved that our novel SBS method is a promising technology but showed also that many challenges need to be addressed for the implementation of this technique. Additional work will be required to test our nucleotides with many different DNA sequences, to increase the read-length, reduce the likelihood of side-reactions and implement the technique in a real sequencing experiment.
6 EXPERIMENTAL

6.1 Experimental for Chapter 2

General Information
Unless otherwise noted, all solvents and reagents were obtained from commercial sources and used without further purification. $^1$H NMR and $^{13}$C NMR spectra were recorded at 298 K on a Bruker ARX250 spectrometer (at 250 MHz and 62.9 MHz), on a Bruker DPX360 spectrometer (at 360 MHz and 90.5 MHz) and on a Bruker DMX500 (at 500 MHz and 126 MHz), respectively. Gel phase $^{13}$C NMR spectra were recorded on a Bruker DPX 360 spectrometer. Chemical shifts are reported on the $\delta$ scale in ppm, using the residual protio solvent as the internal standard. Coupling constants ($J$) are reported in Hz. Symbols such as t (dd), and td (ddd) indicate apparent patterns with the results of detailed analyses in parentheses. Polymer-bound imidates and hemicyanine could not be analysed by gel-phase $^{13}$C NMR. Only the imidate 6a could be analysed. Melting points were determined using a Gallenkamp melting point apparatus and are uncorrected. Microanalysis were performed by MEDAC Ltd., Surrey. Elemental analysis was carried out for all the imidates and for one hemicyanine for each class (tri-, penta- and heptamethine) to determine the loading efficiency. IR spectra were recorded on a FTIR Bruker Tensor series spectrometer with ATR Golden Gate. All samples were run neat. Microwave-assisted reactions were performed on either a Biotage Initiator 2.0 or a Biotage Smith Synthesizer. Absorption spectra were recorded on an 8453 Agilent UV-Visible spectrophotometer. Fluorescence emission spectra were recorded on a Jobin Yvon Fluoromax spectrofluorimeter. Analytical thin layer chromatography (TLC) was performed on Merck silica Gel 60-F254 plates (0.25 mm) with visualization by ultraviolet (UV) irradiation at 254 nm. Purifications by flash chromatography were performed using Merck silica gel 60 (230-400 mesh). ESI-MS analyses were carried out on an Agilent Technologies LC/MSD 1100 Series quadrupole mass spectrometer (QMS) equipped with an electrospray ionisation source. Major peaks are reported with percentage intensities of the base peak.
Electrospray high resolution mass spectrometry was performed on a Finnigan MAT 900 XTL spectrometer.

High performance liquid chromatography was carried out on an Agilent Technologies HP1100 Chemstation system coupled to a Polymer Lab 100 ES evaporative light scattering detector, eluting with (A) 0.1% TFA/H₂O, (B) 0.04% TFA/MeCN, (C) 0.1% formic acid/H₂O, (D) 0.10% formic acid/MeOH using the methods described below.

**HPLC System I**

Column: Phenomenex Gemini, C18, 110 Å, 100 mm x 4.6 mm, 5 µm. Flow rate: 1.0 mL min⁻¹. Eluents C and D. Gradient: 95% to 5% C over 6 min, then 5% C for 3 min, then 5% to 95% C over 1 min. Detection: UV and ELS detection.

**HPLC System II**

Column: Discovery (Supelco) C18, 50 mm x 4.6 mm, 5 µm. Flow rate: 1.0 mL min⁻¹. Eluents A and B. Gradient: 95% A to 5% A over 5 min, then 5% A for 3 min, then 5% to 95% A over 1 min. Detection: UV and or ELS detection.

**HPLC System III**

Column: Phenomenex Luna, C18(2), 150 mm x 4.6 mm, 100 Å. Flow rate: 1.0 mL min⁻¹. Eluents A and B. Gradient: 95% A to 5% A over 8 min, then 5% A for 6 min, then 5% to 95% A over 1 min. Detection: UV and or ELS detection.

**PREPARATION OF INDOLENIUM SALTS**

1,2,3,3-tetramethyl-3H-indolium iodide¹¹⁰,¹⁹⁶

Classical method

A mixture of 1,1,2-trimethylindole (40.10 mL, 0.25 mol) and methyl iodide (38.90 mL, 0.625 mol) in acetonitrile (150 mL) was heated under reflux (110 °C) for three days. The reaction mixture was cooled to room temperature and evaporated. The residue obtained was washed with diethyl ether several times and then with ethanol.
until removal of the starting material. The residue was dried \textit{in vacuo} to give a white/pink solid (64.3g, 84%).

\textit{Microwave-assisted alkylation}

A mixture of 2,3,3-trimethyl-3H-indole (0.25 mL, 1.6mmol) and methyl iodide (0.6 mL, 9.6 mmol) in acetonitrile (1 mL) was heated in a sealed vial at 150 °C for 30 min, under microwave irradiation. After cooling to room temperature, the pink precipitate was collected by filtration, washed with diethyl ether, ethanol and then dried \textit{in vacuo} to give a white/pink solid (457 mg, 94%).

$R_f$: 0.50 (DCM:MeOH 9/1)

\textbf{HPLC} $t_R = 4.54$ min, 100% (ELSD)

\textbf{MS} (ES): $m/z$ (%) 174.2 [M+, 100], 175.2 [(M+1)$^+$, 14]

$^1$H NMR $\delta$ (250 MHz, DMSO): $\delta$ 7.90-7.86 (m, 1H), 7.82-7.78 (m, 1H), 7.63-7.54 (m, 2H), 3.94 (s, 3H), 2.74 (s, 3H), 1.49 (s, 6H)

$^{13}$C NMR (62.9 MHz, DMSO): $\delta$ 196.34 (C=N), 142.45 (C), 141.95 (C), 129.66 (CH), 129.16 (CH), 123.65 (CH), 115.48 (CH), 54.28 (C), 35.11 (CH$_3$), 22.06 (CH$_3$), 14.56 (CH$_3$).

\textbf{FTIR} 3013, 2964, 2360, 1630, 1480, 988, 938 cm$^{-1}$

\textbf{Mp} 219 °C (lit. 219-220 °C)$^{197}$
A mixture of 1,2,2-trimethyl-1H-benz[e]indole (55.1 g, 0.26 mol) and methyl iodide (40.50 mL, 0.65 mol) in acetonitrile (250 mL) was heated under reflux at 100 °C for 2h. The reaction solution was cooled to room temperature and evaporated. The residue obtained was washed with diethyl ether several times and then with ethanol until removal of the starting material. The residue was dried in vacuo to give white/yellow solid (66g, 72%).

Rf: 0.31 (DCM:MeOH 9/1)

MS (ES): m/z (%) 224.2 [M⁺, 100], 225.1 [(M+1)+, 17]

¹H NMR δ (250 MHz, DMSO): δ 8.36 (d, 1H, J = 8.28 Hz, Ar), 8.29 (d, 1H, J = 8.96 Hz, Ar), 8.22 (d, 1H, J = 7.58 Hz, Ar), 8.11 (d, 1H, J = 8.90 Hz, Ar), 7.82-7.68 (m, 2H), 4.19 (s, 3H), 2.97 (s, 3H), 1.84 (s, 6H)

¹³C NMR (62.9 MHz, DMSO): δ 196.26 (C=N), 139.83 (C), 136.86 (C), 133.38 (C), 130.87 (CH), 130.10 (CH), 128.75 (CH), 127.48 (CH), 123.77(CH), 113.53 (CH), 55.62 (C), 35.57 (CH₃), 21.65(CH₃), 14.50 (CH₃)

FTIR 3484, 3419, 2978, 1639, 1583, 1524, 1464, 1133, 800, 736, 704 cm⁻¹

HPLC t_R = 4.80 min, 100% (ELSD)

Mp 221-223 °C (lit. 199 °C)
1-(5-carboxypentyl)-2,3,3-trimethyl-3\textit{H}-indolium bromide$^{196}$

![Chemical structure](image)

**Classical Method$^{10}$**

A mixture of 2,3,3-trimethyl-3\textit{H}-indole (12.1 mL, 75.4 mmol) and 6-bromoexanoic acid (30 g, 189 mmol) in acetonitrile (50 mL) was heated at 100 °C for 16h. The solution was cooled to room temperature and evaporated. The residue obtained was washed with diethyl ether and dichloromethane several times until removal of the starting material to give a white-grey solid (18g, 68%).

**Microwave-assisted alkylation**

A mixture of 2,3,3-trimethyl-3\textit{H}-indole (0.25 mL, 1.6 mmol) and 6-bromoexanoic acid (936 mg, 4.8 mmol) in acetonitrile (1 mL) was heated in a sealed vial at 150 °C for 1 hour under microwave irradiation. After cooling to room temperature, the solvent was evaporated. The residue obtained was washed with diethyl ether and dichloromethane several times until removal of the starting material and then dried in vacuo to give a white-grey solid (374 mg, 66%).

**Rf:** 0.26 (DCM/MeOH 3:2)

**HPLC**

$t_R = 3.74$ min, 100% (ELSD, $\lambda = 220, 254, 280$ nm)

**MS (ES):** $m/z$ (%) 274.2 [M$^+$, 100], 275.3 [(M+1)$^+$, 19]

**$^1$H NMR** (250 MHz, DMSO): $\delta$ 12.34 (br s, 1H, COOH), 8.23-8.19 (m, 1H), 8.09-8.06 (m, 1H), 7.87-7.80 (m, 2H), 4.69 (t, 2H, $J = 7.6$ Hz), 3.91 (s, 3H), 2.44 (t, 2H, $J = 7.1$ Hz), 2.12-2.00 (m, 2H), 1.83-1.61 (m, 10H).

**$^{13}$C NMR** (62.9 MHz, DMSO): $\delta$ 196.88 (C=N), 174.66 (C=O), 142.23 (C), 141.40 (C), 129.75 (CH), 129.30 (CH), 123.88 (CH), 115.87 (CH), 54.52 (C), 47.82 (CH$_2$), 33.73 (CH$_2$), 27.31 (CH$_2$), 25.77 (CH$_2$), 24.38 (CH$_2$), 22.37 (CH$_3$), 14.44 (CH$_3$)

**IR** (neat) 3449, 3195, 2817, 1693, 1624, 1481, 1459, 1290, 1174, 872, 775 cm$^{-1}$.

**Mp** 138 °C (lit. 127-129 °C$^{196}$)
3-(5-carboxypentyl)-1,1,2-trimethyl-1H-benzo[e]indolium bromide

A mixture of 1,2,2-trimethyl-1H-benz[e]indole (15.8 g, 75.4 mmol) and 6-bromoexanoic acid (30 g, 189 mmol) in acetonitrile (50 mL) was heated at 120 °C for 2 days. The solution was cooled to room temperature and evaporated. The residue obtained was washed with diethyl ether and dichloromethane several times until removal of the starting material to give a pink-grey solid (23.4 g, 77%)

Rf: 0.14 (DCM/MeOH 3:2)

HPLC \( t_R = 4.54 \) min, 100% (ELSD)

MS (ES): \( m/z \) 324.3 [M⁺, 100], 325.2 [(M+1)⁺, 35]

\(^1\)H NMR (250 MHz, DMSO): \( \delta \) 12.38 (br s, 1H, COOH), 8.46 (d, 1H, \( J = 8.2 \) Hz), 8.37 (d, 1H, \( J = 8.9 \) Hz), 8.30 (dd, 1H, \( J = 8.7, 1.4 \) Hz), 8.26 (d, 1H, \( J = 8.9 \) Hz), 7.89-7.77 (m, 2H), 4.68 (t, 2H, \( J = 7.5 \) Hz), 3.05 (s, 3H), 2.31 (t, 2H, \( J = 7.0 \) Hz), 2.01-1.93 (m, 2H), 1.68-1.54 (m, 10H).

\(^13\)C NMR (62.9 MHz, DMSO): \( \delta \) 196.74 (C=N), 174.67 (C=O), 138.86 (C), 137.31 (C), 133.39 (C), 131.03 (CH), 130.07 (CH), 128.75 (CH), 127.58 (CH), 123.78 (CH), 113.74 (CH), 55.85 (C), 48.07 (CH₂), 33.74 (CH₂), 27.52 (CH₂), 25.75 (CH₂), 24.42 (CH₂), 21.97 (CH₃), 14.27 (CH₃)

IR (neat) 2864, 1707, 1474, 1389, 1156, 824, 715 cm⁻¹

Mp 234 °C (lit. 211-213 °C)
PREPARATION OF SULFONATED INDOLENIUM SALTS

Potassium 2,3,3-trimethyl-3H-indole-5-sulfonate

\[ \text{Hydrazinobenzene sulfonic acid (15g, 79.5 mmol) and 3-methyl-2-butanone (25.2 mL, 240 mmol) were dissolved in 45 mL of acetic acid. The mixture was heated to reflux for 3 h. Acetic acid was removed under vacuum. The residue obtained was dissolved in MeOH and then stirred with a saturated solution of potassium hydroxide in 2-propanol. The alkaline solution turned yellow and the potassium salt of the sulfoindole precipitated as yellow solid almost quantitatively (16.5 g, 75\%).} \]

\[ \text{Rf: } 0.40 \text{ (RP, water)} \]

\[ \text{HPLC } t_R = 2.6 \text{ min} \]

\[ \text{MS (ES): } m/z \text{ (%) } 238.0 [(M-1)^-, 100], 238.9 [M^-, 12], 239.9 [(M+1)^-, 5], 499.0 [(2M+Na)^-, 10]. \]

\[ ^1\text{H NMR (360 MHz, DMSO)} \delta: 7.63 (d, 1H, J = 1.5 Hz, H-4), 7.56 (dd, 1H, J = 7.9, 1.5 Hz, H-6), 7.34 (d, 1H, J = 7.9 Hz, H-7), 2.21 (s, 3H), 1.24 (s, 6H). \]

\[ ^13\text{C NMR (90.5 MHz, DMSO)} \delta: 188.67 (C=N), 153.12 (C), 146.91 (C), 144.68 (C), 124.75 (CH), 118.80 (CH), 117.76 (CH), 52.89 (C), 22.12 (CH\text{$_3$}), 14.78 (CH\text{$_3$}). \]

\[ \text{IR (neat) } 3377, 2966, 2360, 1607, 1572, 1421, 1388, 1180, 1121, 1063, 1029, 833, 726 \text{ cm}^{-1} \]

\[ \text{Mp: } 292-293 \degree \text{C (lit. 290-295 \degree \text{C})} \]
1,2,3,3-tetramethyl-3H-indolium-5-sulfonate

![Chemical structure](image)

26a

The potassium salt of 2,3,3-Trimethyl-3H-indole-5-sulfonate (1 g, 3.6 mmol) and methyl iodide (2.2 mL, 18 mmol) were heated in acetonitrile under microwave irradiation at 150 °C in a sealed vial for 1h. After cooling to room temperature, the solvent was decanted. The yellow-brown precipitate was suspended in diethyl ether, filtered, washed with diethylether and dichloromethane and dried in vacuo to yield a beige solid (0.86 g, 95%).

Rf: 0.20 (RP, water)

HPLC $t_R = 2.8$ min

MS (ES): $m/z$ (%) 252.1[(M-1), 100], 253.1 [M, 10]

$^1$H NMR (360 MHz, DMSO) $\delta$: 7.97 (s, 1H, H-4), 7.82 (d, 1H, $J = 8.3$ Hz, H-7), 7.77 (dd, 1H, $J = 8.3$, 1.3 Hz, H-6), 3.96 (s, 3H), 2.76 (s, 3H), 1.55 (s, 6H).

$^{13}$C NMR (90.5 MHz, DMSO): $\delta$ 198.32 (C=N), 150.63 (C), 143.45 (C), 142.69 (C), 127.62 (CH), 119.29 (CH), 116.06 (CH), 55.47 (C), 36.37 (CH$_3$), 23.06 (CH$_3$), 15.85 (CH$_3$).

IR (neat) 3010, 2972, 2354, 1629, 1593, 1470, 1421, 1195, 1118, 1057, 1031 cm$^{-1}$

Mp 273-274 °C
1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium-5-sulfonate

Classical method
A mixture of potassium 2,3,3-trimethyl-3H-5-indolesulfonate (2.00 g, 7.2 mmol), 6-bromohexanoic acid (4.20 g, 21.6 mmol) and potassium iodide (0.12 g, 0.7 mmol) in 1,2-dichlorobenzene (20 mL) was refluxed at 150 °C for 48 h. The reaction was monitored by TLC (DCM:MeOH 4:1). After cooling the reaction mixture to room temperature, the flask was placed in the fridge for 1 h. After this time, a beige solid had formed in the purple solution. The solid was collected by filtration, then washed with dichlorobenzene, 2-propanol/diethylether (1:1 v/v) and finally with diethylether. After drying in vacuo, overnight, a beige solid was obtained (3.0 g, 89%) 

Microwave-assisted alkylation
A mixture of potassium 2,3,3-trimethyl-3H-5-indolesulfonate (0.25 g, 0.9 mmol) and 6-bromohexanoic acid (0.53 g, 2.7 mmol) in a mixture (9:1) of acetonitrile and DMF (1 mL) was heated in a sealed vial at 150 °C for 3hs (3 cycles of 1h each), under microwave irradiation. After cooling to room temperature, the solvent was evaporated. The residue obtained was washed with 2-propanol/diethylether (1:1 v/v) and finally with diethylether and then dried in vacuo to give a beige solid (254 mg, 80 %). 

Rf 0.24 (RP, water) 
HPLC = 3.6min
$^1$H NMR (360 MHz, DMSO) δ 8.02 (s, 1H), 7.91 (d, 1H, $J = 8.3$ Hz), 7.82 (d, 1H, $J = 8.3$, 1.1 Hz), 4.44 (t, 2H, $J = 7.5$ Hz), 2.83 (s, 3H), 2.22 (t, 2H, $J = 7.1$ Hz), 1.88-1.77 (m, 2H), 1.57-1.38 (m, 10H).

$^{13}$C NMR (90.5 MHz, DMSO) δ 198.8 (C=N), 175.8 (CO), 151.0 (C), 143.0 (C), 142.4 (C), 127.8 (CH), 122.2 (CH), 116.4 (CH), 55.8 (C), 49.0 (CH2), 34.8 (CH2), 28.4 (CH2), 26.9 (CH2), 25.5 (CH2), 23.4 (CH3), 15.6 (CH3).

MS (ES): m/z (%) 352.0 [(M-2)$^-$, 100], 353.0 [(M-1)$^-$, 18], 354.0 [M$^+$, 5], 705.2 [(2M-3)$^-$, 21].

IR (neat) 3039, 2949, 1704, 1474, 1381, 1236, 1154, 1117, 1024, 795 cm$^{-1}$

Mp 235 °C

N-('butoxycarbonyl)-4-hydroxylaniline PS$^{108}$

A mixture of N-('butoxycarbonyl)-4-hydroxylaniline (6.27 g, 30 mmol), Cs$_2$CO$_3$ (9.78 g, 30 mmol), KI (0.17 g, 1 mmol) and 1% DVB cross-linked chloromethyl polystyrene (5 g, 10 mmol) in acetone (45 mL) was heated at 70 °C for 16 h. The resin was isolated by filtration, washed with water (4 x 100 mL), DMF (4 x 50 mL), DCM (3 x 50 mL), Et$_2$O (3 x 50 mL) and dried overnight in vacuo at 40 °C to give the product as a beige resin (quantitative by N analysis, 96% by Cl analysis). Found: N, 2.13; Cl, 0.27. Calcd: N, 2.08; Cl, 0%. IR (neat) 3399, 3024, 2919, 1722, 1511, 1217, 1154, 698 cm$^{-1}$. $^{13}$C NMR (90.5 MHz, CDCl$_3$) δ 155.5 (C=O), 154.0 (C-O), 131.5-125.5 (Ar-resin + C-N), 119.6 (Ar), 114.2 (Ar), 80.1 (C), 70.2 (CH2), 40.2 (CH-resin), 28.4 (CH3).

Colorimetric test for the detection of chloromethyl groups in Merrifield resin$^{122}$

A 75 mM solution of 4-(4-nitrobenzyl)pyridine in toluene with 5% of triethylamine (1mL) was added to 2mg of resin 17. No coloration was observed (naked eye). The colorimetric test was then repeated on 2 mg of Merrifield resin (1% DVB, 2mmol Cl/g). The resin turned to pink.
4-aminophenol PS\textsuperscript{108}

![Image of 4-aminophenol PS]

Resin 17 (5g, 7.5 mmol) was shaken at room temperature with a 20\% solution of TFA in DCM (75 mL) for 2 h, filtered and washed with DCM (2 × 50 mL). The resin was then shaken with a 10\% solution of triethylamine in DCM (75 mL) for 15 min, filtered, washed with DCM (4 × 50 mL), and dried overnight \textit{in vacuo} at 40 °C to give the product as a beige resin. Found: N, 2.56; Calcd: N, 2.33\%. \textbf{IR} (neat) 3360, 3023, 2917, 1601, 1508, 1217, 698 cm\textsuperscript{-1}. \textbf{\textsuperscript{13}C NMR} (90.5 MHz, CDCl\textsubscript{3}) \( \delta \) 134.2-121.3 (Ar-resin, C-O, C-N), 116.1 (Ar), 115.8 (Ar), 70.5 (CH\textsubscript{2}), 40.2 (CH-resin).

\textbf{TRIMETHINE DYES}

Ethyl 4-hydroxyphenylimidoformate PS\textsuperscript{108}

![Image of Ethyl 4-hydroxyphenylimidoformate]

To 4-aminophenol PS 18 (2 g, 3.3 mmol) was added a solution of triethylorthoformate (7.50 mL, 45.1 mmol) and BF\textsubscript{3}·OEt\textsubscript{2} (0.46 mL, 3.7 mmol) in dry DCM (10 mL) and the solution stirred at room temperature for 6h. Dry DIEA (0.94 mL, 5.51 mmol) was added and the mixture stirred for 5 min. The resin was isolated by filtration, washed several times with DCM (20 mL), and dried \textit{in vacuo} to give the product as a brown resin.

Found: N, 2.29; Calcd: N, 2.14. \textbf{IR} (neat) 3024, 2920, 1639, 1503, 1184, 698 cm\textsuperscript{-1}. \textbf{\textsuperscript{13}C NMR} (90.5 MHz, CDCl\textsubscript{3}) \( \delta \) 155.0 (CH), 123.1 (Ar), 115.6 (Ar), 71.3 (CH\textsubscript{2}), 63.3 (CH\textsubscript{2}-Et), 41.5 (CH-resin), 15.4 (CH\textsubscript{3}-Et).
2-\{(E)-2-(4-hydroxyanilino)ethenyl\}-1,3,3-trimethyl-3\textit{H}-indolium iodide PS$^{108}$

\begin{center}
\includegraphics[width=0.2\textwidth]{image}
\end{center}

To ethyl 4-hydroxyphenylimidoformate PS 13a (1 g, 1.53 mmol) and 1,2,3,3-tetramethyl-3\textit{H}-indolium iodide (2.3 g, 7.65 mmol) was added DMF (8 mL) and the mixture stirred at 120 °C for 15 min under microwave irradiation. After cooling, the resin was isolated by filtration, washed with DMF (3 × 10 mL) and DCM (3 × 10 mL), and dried \textit{in vacuo} to give the product as an orange resin (0.72 mmol\textsuperscript{-1}, 66\% by N analysis). Found: N, 2.03; Calcd: N, 3.08. IR (neat) 3024, 2916, 1668, 1506, 1451, 1225, 1014, 907, 824, 754 cm\textsuperscript{-1}.

\begin{center}
\includegraphics[width=0.2\textwidth]{image}
\end{center}

2-\{(E)-2-(4-hydroxyanilino)ethenyl\}-1,1,3-trimethyl-1\textit{H}-benzo[e]indolium iodide PS

To ethyl 4-hydroxyphenylimidoformate PS 13a (1 g, 1.53 mmol) and 1,1,2,3-tetramethyl-1\textit{H}-benzo[e]indolium iodide (2.7 g, 7.65 mmol) was added DMF (8 mL) and the mixture stirred at 120 °C for 15 min under microwave irradiation. After cooling, the resin was isolated by filtration, washed with DMF (3 × 10 mL) and DCM (3 × 10 mL), and dried \textit{in vacuo} to give the product as a dark red-brown resin. IR (neat) 3024, 2916, 1685, 1506, 1451, 1206, 1014, 823, 751, 704 cm\textsuperscript{-1}.
To PS-bound hemicyanine 15a (1.1 g, 0.80 mmol) were added dry pyridine (5 mL), DIEA (1.3 mL, 7.4 mmol), Ac₂O (0.7 mL, 7.4 mmol) and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (35 mg, 0.10 mmol). The mixture was stirred at room temperature for 2 h. Afterwards, the resin was removed by filtration and washed several times with DCM (10 mL). The filtrates were combined and evaporated and the resulting solid was dissolved in DCM (15 mL) and washed several times with water (10 mL). After removal of the solvent in vacuo, the crude product was precipitated in Et₂O, washed with Et₂O and then purified by chromatography (DCM → DCM/MeOH 1:1) to give 20a as a dark glassy red solid (31 mg, 68%).

Rᵣ 0.42 (DCM/MeOH 9:1)

HPLC \(t_R = 10.7 \text{ min}, 100\%\) (ELSD, \(\lambda = 550 \text{ nm}\)).

MS (ES): \(m/z (\%) 457.2 ([M^+1], 100), 458.2 ([M+1]^+, 31), 459.2 ([M+2]^+, 4)\)

\(^1\)H NMR (360 MHz, CD₃OD) \(\delta 8.55 (t (dd), 1H, J_{1,2} \approx J_{2,3} = 13.5 \text{ Hz}, \text{CH-H2}), 7.55 (d, 2H, J = 7.4 \text{ Hz}, \text{Ar-H4, Ar-H4'}), 7.53-7.48 (m, 2H, Ar-H6, Ar-H6'), 7.36 (d, 2H, J = 7.8 \text{ Hz}, \text{Ar-H7, Ar-H7'}), 7.31 (td (ddd) 1H, \(J_{4,5} \approx J_{5,6} \approx 7.4, J_{5,7} = 1.2 \text{ Hz}, \text{Ar-H5 or Ar-H5'}), 7.30 (td (ddd), 1H, \(J_{4,5} \approx J_{5,6} = 7.4, J_{5,7} = 1.2 \text{ Hz}, \text{Ar-H5 or Ar-H5'}), 6.48 (d, 1H, J = 13.5 \text{ Hz}, \text{CH-H1 or H3}), 6.47 (d, 1H, J = 13.5 \text{ Hz}, \text{CH-H1 or H3}), 4.16 (t, \text{d, } J = 7.5 \text{ Hz}, \text{CH}_2-\text{CH}_2), 3.69 \text{ (s, 3H, NCH}_3), 2.21 \text{ (t, 2H, } J = 7.5 \text{ Hz, CH}_2-\text{CH})\).

\(^1^3\)C NMR (90.5 MHz, CDCl₃) \(\delta 174.5 \text{ (COOH), 173.6 (C=N), 150.5 (CH), 142.7 (C-Ar), 141.9 (C-Ar), 140.5 (C-Ar), 140.4 (C-Ar), 128.9 (CH-Ar), 125.4 (CH-Ar), 125.3 (CH-Ar), 122.1 (CH-Ar), 121.9 (CH-Ar), 111.0 (CH-Ar), 110.9 (CH-Ar), 105.8 (CH-Ar), 101.7 (CH-Ar), 100.8 (CH-Ar), 98.0 (CH-Ar), 97.0 (CH-Ar), 96.0 (CH-Ar), 94.0 (CH-Ar), 92.0 (CH-Ar), 90.0 (CH-Ar), 88.0 (CH-Ar), 86.0 (CH-Ar), 84.0 (CH-Ar), 82.0 (CH-Ar), 80.0 (CH-Ar), 78.0 (CH-Ar), 76.0 (CH-Ar), 74.0 (CH-Ar), 72.0 (CH-Ar), 70.0 (CH-Ar), 68.0 (CH-Ar), 66.0 (CH-Ar), 64.0 (CH-Ar), 62.0 (CH-Ar), 60.0 (CH-Ar), 59.0 (CH-Ar), 58.0 (CH-Ar), 57.0 (CH-Ar), 56.0 (CH-Ar), 55.0 (CH-Ar), 54.0 (CH-Ar), 53.0 (CH-Ar), 52.0 (CH-Ar), 51.0 (CH-Ar), 50.0 (CH-Ar), 49.0 (CH-Ar), 48.0 (CH-Ar), 47.0 (CH-Ar), 46.0 (CH-Ar), 45.0 (CH-Ar), 44.0 (CH-Ar), 43.0 (CH-Ar), 42.0 (CH-Ar), 41.0 (CH-Ar), 40.0 (CH-Ar), 39.0 (CH-Ar), 38.0 (CH-Ar), 37.0 (CH-Ar), 36.0 (CH-Ar), 35.0 (CH-Ar), 34.0 (CH-Ar), 33.0 (CH-Ar), 32.0 (CH-Ar), 31.0 (CH-Ar), 30.0 (CH-Ar), 29.0 (CH-Ar), 28.0 (CH-Ar), 27.0 (CH-Ar), 26.0 (CH-Ar), 25.0 (CH-Ar), 24.0 (CH-Ar), 23.0 (CH-Ar), 22.0 (CH-Ar), 21.0 (CH-Ar), 20.0 (CH-Ar), 19.0 (CH-Ar), 18.0 (CH-Ar), 17.0 (CH-Ar), 16.0 (CH-Ar), 15.0 (CH-Ar), 14.0 (CH-Ar), 13.0 (CH-Ar), 12.0 (CH-Ar), 11.0 (CH-Ar), 10.0 (CH-Ar), 9.0 (CH-Ar), 8.0 (CH-Ar), 7.0 (CH-Ar), 6.0 (CH-Ar), 5.0 (CH-Ar), 4.0 (CH-Ar), 3.0 (CH-Ar), 2.0 (CH-Ar), 1.0 (CH-Ar), 0.0 (CH-Ar)).

162
104.4 (CH), 103.6 (CH), 51.5 (C), 44.3 (CH₂), 40.7 (CH₂), 31.8 (CH₃), 29.7 (CH₂),
28.1 (CH₃), 27.9 (CH₃), 26.3 (CH₂), 25.7 (CH₂).

UV/vis (MeOH) λₐbs = 547 nm, λₑm = 561 nm, ε = 1.06 × 10⁵ M⁻¹ cm⁻¹

IR (neat) 2927, 2360, 1686, 1558, 1456, 1414, 1199, 1114, 926 cm⁻¹

HRMS (ES) for C₃₀H₃₇N₂O₂⁺ (M⁺): calc 457.28496, found 457.28661

2-{(1E,3E)-3-[3-(5-carboxypentyl)-1,1-dimethyl-1,3-dihydro-2H-
benzo[e]indol-2-ylidene]-1-propenyl}-1,1,3-trimethyl-1H-benzo[e]indolium
Inner Salt

To PS-bound hemicyanine 15b (0.5 g, 0.50 mmol)* were added dry pyridine (3.5
mL), DIEA (1.0 mL, 5.8 mmol), Ac₂O (0.54 mL, 5.8 mmol) and 3-(5-
carboxypentyl)-1,1,2-trimethyl-H-benzo[e]indolium bromide (24 mg, 0.06 mmol).
The mixture was stirred at room temperature for 2 h. Afterwards, the resin was
removed by filtration and washed several times with DCM (10 mL). The filtrates
were combined and evaporated and the resulting solid was dissolved in DCM (15
mL) and washed with water (3 × 10 mL). After removal of the solvent in vacuo,
the crude product was precipitated in Et₂O, washed with Et₂O and then purified by
chromatography (DCM → DCM/MeOH 1:1) to give 20b as a purple solid (16 mg,
49%).

*based on theoretical loading of the resin 15b

Rf 0.35 (DCM/MeOH 9:1)

HPLC tR = 5.65 min, 100% (ELSD)

MS (ES): m/z (%) 557.2 [M⁺, 100], 558.2 [(M+1)⁺, 40], 559.2 [(M+2)⁺, 10]
**1H NMR** (360 MHz, CD$_2$OD) δ 8.78 (t (dd), 1H, $J_{1,2}$ = $J_{2,3}$ = 13.5 Hz, CH-H2), 8.29 (d, 2H, $J$ = 8.2, Hz, Ar-H9, Ar-H9'), 8.10-8.04 (m, 4H, Ar-H4, Ar-H4', Ar-H6, Ar-H6'), 7.70-7.63 (m, 4H, Ar-H5, Ar-H5', Ar-H8, Ar-H8'), 7.55 (bt (dd), 2H, $J_{6,7}$ = $J_{7,8}$ = 7.2 Hz, Ar-H7, Ar-H7'), 6.50 (d, 1H, $J$ = 13.5 Hz, CH-H1 or H3), 6.49 (d, 1H, $J$ = 13.5 Hz, CH-H1 or H3), 4.30 (t, 2H, $J$ = 7.2 Hz, CH$_2$-e), 3.82 (s, 3H, NCH$_3$), 2.31 (t, 2H, $J$ = 7.2 Hz, CH$_2$-a), 2.10 (s, 12H, CH$_3$), 1.94 (quintet, 2H, $J$ = 7.2 Hz, CH$_2$-δ), 1.74 (quintet, 2H, $J$ = 7.2 Hz, CH$_2$-β), 1.61-1.52 (m, 2H, CH$_2$-γ).

**13C NMR** (90.5 MHz, CD$_2$OD) δ 182.0 (COO), 177.5 (C=N), 176.9 (C=N), 150.3 (CH), 141.2 (C-Ar), 140.5 (C-Ar), 134.5 (C-Ar), 134.3 (C-Ar), 133.2 (C-Ar), 131.6 (CH), 131.4 (CH-Ar), 130.8 (CH-Ar), 128.9 (CH-Ar), 128.4 (CH-Ar), 125.9 (CH-Ar), 122.9 (CH-Ar), 111.9 (CH-Ar), 111.7 (CH-Ar), 110.6 (CH-Ar), 102.9 (CH), 102.8 (CH), 52.0 (C), 51.9 (C), 45.0 (CH$_2$), 38.5 (CH$_2$), 31.7 (CH$_3$), 28.1 (CH$_2$), 27.6 (CH$_3$), 27.5 (CH$_3$), 27.3 (CH$_2$), 26.8 (CH$_2$).

**UV/vis** (MeOH) $\lambda_{abs}$ = 587 nm, $\lambda_{em}$ = 603 nm, $\varepsilon$ = 1.05 $\times 10^5$ M$^{-1}$ cm$^{-1}$

**IR** (neat) 3400, 2924, 1720, 1558, 1515, 1483, 1447, 1414, 1227, 1157, 1013, 935, 750 cm$^{-1}$.

**HRMS** (ES) for C$_{38}$H$_{41}$N$_2$O$_2$ (M)$^+$: calcld 557.31626, found 557.31623.

**PENTAMETHINE DYES**

4-{{[(E,2E)-3-methoxy-2-propenylidenel amino}phenol PS$^{108}$

![Image of 13b]

To 4 aminophenol PS 18 (2 g, 3.34 mmol) was added a solution of 1,1,3,3-tetramethoxypropane (7.4 mL, 45.1 mmol) and BF$_3$·OEt$_2$ (0.46 mL, 3.71 mmol) in dry DCM (9 mL) and the solution stirred at room temperature for 6h. Dry DIEA (0.94 mL, 5.51 mmol) was added and the mixture stirred for 5 min. The resin was isolated by filtration, washed several times with DCM (20 mL), and dried in vacuo to give the product as a dark blue-black resin. Found: N, 2.19; Calcd: N, 2.10. **IR** (neat) 3023, 2917, 2360, 2341, 1626, 1507, 1450, 1218, 1172, 698, cm$^{-1}$.
2-[(1E,3E)-4-(4-hydroxyanilino)-1,3-butadienyl]-1,3,3-trimethyl-3H-indolium iodide PS

To 4-{{[(E,2E)-3-methoxy-2-propenylidene]amino}phenol PS 13b (1 g, 1.56 mmol) and 1,2,3,3-tetramethyl-3H-indolium iodide (2.3 g, 7.6 mmol) was added DMF (8 mL) and the mixture stirred at 120 °C for 15 min under microwave irradiation. After cooling, the resin was isolated by filtration, washed with DMF (3 × 10 mL) and DCM (3 × 10 mL), and dried in vacuo to give the product (0.96 mmol, 87 % by N analysis) as a blue-black resin. Found: N, 1.87; Calcd: N, 2.14. IR (neat) 3023, 2917, 1588, 1506, 1451, 1220, 1171, 1014, 822, 757, 733, 698 cm⁻¹ The signal at 1626 (C=N) disappeared.

2-[(1E,3E)-4-(4-hydroxyanilino)-1,3-butadienyl]-1,3,3-trimethyl-1H-benzo[e]indolium iodide PS

To 4-{{[(E,2E)-3-methoxy-2-propenylidene]amino}phenol PS 13b (1 g, 1.56 mmol) and 1,2,3,3-tetramethyl-3H-indolium iodide (2.7 g, 7.8 mmol) was added DMF (8 mL) and the mixture stirred at 120 °C for 15 min under microwave irradiation. After cooling, the resin was isolated by filtration, washed with DMF (3 × 10 mL) and DCM (3 × 10 mL), and dried in vacuo to give the product as a blue-black resin. IR (neat) 3023, 2917, 2360, 1506, 1455, 1168, 1012, 939, 824, 752, 698 cm⁻¹.
To PS-bound hemicyanine 15c (1.1 g, 1.1 mmol) were added dry pyridine (7 mL), DIEA (1.9 mL, 11 mmol), Ac₂O (1.0 mL, 11 mmol) and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (50 mg, 0.14 mmol). The mixture was stirred at room temperature for 1-2 h. Afterwards, the resin was removed by filtration and washed with DCM (4 × 10 mL). The filtrates were combined and evaporated and the resulting solid was dissolved in DCM (15 mL) and washed several times with water (10 mL). After removal of the solvent in vacuo, the crude product was precipitated in Et₂O, washed with Et₂O and then purified by chromatography (DCM → DCM/MeOH 1:1) to give 20c as a blue solid (57 mg, 84%).

Rf 0.32 (DCM/MeOH 9:1)
IR (neat) 2929, 2359, 1684, 1478, 1436, 1092, 919, 897, 786, 728, 708 cm⁻¹
HPLC tR = 6.25 min, 100% (ELSD, λ = 650 nm)
purity of crude product: 97% (ELSD), 90% (λ = 650 nm)
MS (ES): m/z (%) 483.3 [M⁺, 100], 484.3 [(M+1)⁺, 36], 485.3 [(M+2)⁺, 7].

¹H NMR (360 MHz, MeOD) δH 8.25 (t (dd), 2H, J₁,₂ ≈ J₂,₃ ≈ J₃,₄ = J₄,₅ = 13.1 Hz, CH-H2 + CH-H4), 7.49 (bd, 2H, J = 7.4 Hz, Ar-H4, Ar-H4'), 7.44 (td (ddd), 2H, J₅,₆ ≈ J₆,₇ ≈ J₇,₈ = 1.1 Hz, Ar-H6, Ar-H6'), 7.31-7.23 (m, 4H, Ar-H5, Ar-H5', Ar-H7, Ar-H7'), 6.64 (t (dd), 1H, J₂,₃ ≈ J₃,₄ = 12.4 Hz, CH-H3), 6.30 (d, 1H, J = 13.7, Hz, CH-H1 or H5), 6.27 (d, 1H, J = 13.7, Hz, CH-H1 or H5), 4.10 (t, 2H, J = 7.5 Hz, CH₂-δ), 3.62 (s, 3H, NCH₃), 2.19 (t, 2H, J = 7.4 Hz, CH₂-α), 1.87-1.79 (m, 2H, CH₂-δ), 1.72-1.65 (s + m, 14H, CH₃-β + CH₃), 1.54-1.44 (m, 2H, CH₂-γ).

¹³C NMR (90.5 MHz, MeOD) δC 182.2 (COO⁻, salt form), 180.3 (COOH), 174.7 (CN), 174.3 (CN), 155.2 (CH), 155.0 (CH), 143.9 (C-Ar), 143.1 (C-Ar), 142.3 (C-
Ar), 142.0 (C-Ar), 129.4 (CH-Ar), 129.3 (CH-Ar), 126.3 (CH), 125.8 (CH-Ar), 125.7 (CH-Ar), 122.9 (CH-Ar), 122.8 (CH-Ar), 111.7 (CH-Ar), 111.3 (CH-Ar), 103.9 (CH), 103.5 (CH), 50.1 (C), 50.0 (C), 44.5 (CH2), 38.2 (CH2), 31.0 (CH3), 27.9 (CH2), 27.5 (CH3), 27.4 (CH3), 26.8 (CH2), 23.6 (CH3).

UV/vis (MeOH) $\lambda_{abs} = 640$ nm, $\varepsilon = 1.77 \times 10^5$ M$^{-1}$cm$^{-1}$, $\lambda_{em} = 660$ nm

HRMS for C$_{32}$H$_{39}$N$_2$O$_2$ $^+$ (M): calcd 483.30061, found 483.30063

2-|{(1E,3E,5E)-5-[3-(5-carboxypentyl)-1,1-dimethyl-1,3-dihydro-2H-benzo[e]indol-2-ylidene]-1,3-pentadienyl}-1,1,3-trimethyl-1H-benzo[e]indolium Inner Salt

To PS-bound Hemicyanine 15d (1.5 g, 1.56 mmol) were added dry pyridine (11 mL), DIEA (2.6 mL, 15 mmol), Ac$_2$O (1.4 mL, 15 mmol) and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (80 mg, 0.20 mmol). The mixture was stirred at room temperature for 1-2 h. Afterwards, the resin was isolated by filtration and washed several times with DCM (15 mL). The filtrates were combined and evaporated and the resulting solid was redissolved in DCM (20 mL) and washed several times with water (10 mL). After removal of the solvent in vacuo, the crude product was precipitated in Et$_2$O, washed with Et$_2$O and then purified by chromatography (DCM $\rightarrow$ DCM/MeOH 1:1) to give 20d as a dark blue-green solid (107 mg, 92%).

HPLC $t_R = 11.5$ min, 100% (ELSD, $\lambda = 675$ nm)

MS (ES): $m/z$ (%) 583.5 [M$^+$, 100], 584.5 [M+1$^+$, 45], 585.5 [M+2$^+$, 10].

$^1$H NMR (360 MHz, MeOD) $\delta$ 8.33 (t, (dd) 2H, $J_{1,2} \approx J_{2,3} = J_{3,4} = J_{4,5} = 12.9$ Hz, CH-H2 + CH-H4), 8.24 (d, 2H, $J = 8.5$ Hz, Ar-H9, Ar-H9!), 8.00 (d, 1H, $J = 8.8$ Hz, Ar-H6 or Ar-H6!), 7.99 (d, 1H, $J = 8.8$ Hz, Ar-H6 or Ar-H6!), 7.97 (d, 2H, $J = 8.2$ Hz, Ar-H4, Ar-H4!), 7.59-7.67 (m, 4H, Ar-H5, Ar-H5!, Ar-H8, Ar-H8!), 7.47
(bt, (dd) 2H, \( J_{6.7} \approx J_{7.8} = 7.6 \) Hz, Ar-H7, Ar-H7'), 6.68 (t, (dd) 1H, \( J_{2.3} \approx J_{3.4} = 12.4 \) Hz, CH-H3), 6.35 (d, 2H, \( J = 13.7 \) Hz, CH-H1 or H5), 6.34 (d, 2H, \( J = 13.8 \) Hz, CH-H1 or H5), 4.23 (t, 2H, \( J = 7.3 \) Hz, CH2-β), 3.76 (s, 3H, NCH3), 2.33 (t, 2H, \( J = 7.2 \) Hz, CH2-α), 1.98 (s, 12H, NCH3), 1.86-1.92 (m, 2H, CH2-δ), 1.71-1.79 (m, 2H, CH2-β), 1.53-1.61 (m, 2H, CH2-γ).

\(^{13}\)C NMR (90.5 MHz, MeOD) δ 179.4 (COOH), 176.4 (C=N), 175.8 (C=N), 154.3 (CH), 141.6 (C), 140.9 (C), 135.1 (C), 134.9 (C), 133.4 (C), 131.7 (CH), 131.1 (CH), 129.4 (CH), 129.4 (CH), 128.7 (CH), 126.6 (CH), 126.1 (CH), 126.0 (CH), 123.3 (CH), 112.1 (CH), 111.9 (CH), 104.1 (CH), 103.9 (CH), 52.4 (C), 52.3 (C), 45.0 (CH2), 36.5 (CH2), 31.9 (CH3), 28.5 (CH2), 27.6 (CH2), 27.5 (CH3), 26.4 (CH2).

UV/vis (MeOH) \( \lambda_{\text{abs}} = 678 \) nm, \( \varepsilon = 1.42 \times 10^5 \) M\(^{-1}\) cm\(^{-1}\), \( \lambda_{\text{em}} = 704 \) nm

R\(_f\) 0.35 (DCM/MeOH 9:1)

IR (neat) 3023, 2918, 2360, 1552, 1501, 1450, 1159, 1012, 828, 755 cm\(^{-1}\).

HRMS for C\(_{40}\)H\(_{43}\)N\(_2\)O\(_2\)\(^+\) (M)

HEPTAMETHINE DYSES

4-[[1(E,3E,5E)-5-(phenylimino)-1,3-pentadienyl]amino]phenol PS

![Chemical Structure](image)

Method 1

A suspension of glutaconaldehyde dianil monohydrochloride (2.85 g, 10 mmol) and BF\(_3\)-Et\(_2\)O (0.9 mL, 7.5 mmol) in DCM (20 mL) was stirred at room temperature for 15 min and then added to PS-Bound aniline \( \mathbf{18} \) (2.8 g, 5 mmol). The reaction mixture was stirred at room temperature for 6h and then filtered. The resin was washed extensively with DMF, then with DCM and dried \textit{in vacuo} to give the product as a dark red resin (0.6 mmol g\(^{-1}\), 45 % by N analysis). Found: N, 1.66; Calcd: N, 3.71. IR (neat) 3023, 2918, 2360, 1552, 1501, 1450, 1159, 1012, 828, 755 cm\(^{-1}\).
Method 2

A solution of Ac₂O (0.9 mL, 9.6 mmol) in DCM (4.1 mL) was added dropwise to a stirring suspension of glutaconaldehyde dianilide monohydrochioride (2.5 g, 8.7 mmol) and DIEA (2.8 mL, 17.4 mmol) in DCM (20 mL). The resulting clear solution was stirred for 1h and evaporated. The residue was dissolved in DMF (20 mL) and added to PS-Bound aniline 18 (2.5 g, 4.5 mmol). The mixture was stirred at room temperature for 4h and then filtered. The resin was washed extensively with DMF, then with DCM and dried in vacuo to give the product as a dark red resin (0.6 mmol g⁻¹, 49 % by N analysis). Found: N, 1.73; Calcd: N, 3.51. IR (neat) 3023, 2918, 2360, 1552, 1501, 1450, 1159, 1012, 828, 755 cm⁻¹.

2-[(1E,3E,5E)-6-(acetylanilino)-1,3,5-hexatrieny]-1,3,3-trimethyl-3H-indolium iodide¹⁹⁹

A suspension of 1,2,3,3-tetramethyl-3H-indolium iodide (2.5 g, 8.3 mmol) and N-[(E,2E,4E)-5-anilino-2,4-pentadienylidene]aniline hydrochloride (2 g, 7 mmol) in a (1:1) mixture of acetic acid and acetic anhydride (30 mL) was heated at 120 °C for 2.5 h. The completion of the reaction was carefully monitored by UV-visible spectrometry: hemicyanine intermediate and symmetrical cyanine dye have an absorption maximum around 489 and 750 nm, respectively. After cooling, the reaction mixture was evaporated and the resulting solid was dissolved in DCM (20 mL) and repeatedly washed with water (15 mL). After removal of the solvent in vacuo, the crude product was precipitated in Et₂O, collected by filtration and dried to give the compound 19e as a dark red solid.

Rᵣ: 0.36 (DCM/MeOH 9:1)
IR (neat) 3318, 3050, 1684, 1593, 1547, 1259, 1156, 1106, 995, 917, 760 cm⁻¹.
HPLC tᵣ = 8.5 min
MS (ES): m/z (%) 371.1 [M⁺, 100], 372.1 [(M+1)⁺, 25], 373.1 [(M+2)⁺, 3].
¹H NMR (360 MHz, CDCl₃) 8.17 (d, J = 14.0, Hz, 1H, CH), 8.15-8.12 (m, 1H, CH), 7.92 (dd, J = 14.9, 11.4 Hz, 1H, CH), 7.64-7.12 (m, 9H, Ar), 7.00-6.96 (m, 2H, CH), 6.61 (dd, J = 14.2, 11.4 Hz, 1H, CH), 3.78 (s, 3H, CH₃), 1.74 (s, 3H, CH₃), 1.48 (s, 6H, CH₃)
UV = (MeOH) \( \lambda_{\text{abs}} = 489 \text{ nm} \)

2-[(1E,3E,5E)-6-(acetyl-4-hydroxyanilino)-1,3,5-hexatrienyl]-1,3,3-trimethyl-3\(H\)-indolium iodide PS

\[
\begin{align*}
\text{Method 1-HEMICYANINE 1} \\
\text{To the resin 13e (2.9 g, 1.7 mmol) were added Ac}_2\text{O (1.6 mL, 17 mmol), DIEA (3.0 mL, 17 mmol), pyridine (5 mL) and 1,2,3,3-tetramethyl-3\(H\)-indolium iodide (0.6 g, 2 mmol) and the mixture was stirred at room temperature for 1.5h. The resin was isolated by filtration, washed with DMF and DCM and dried in vacuo to give the product as a dark resin. IR (neat) 3025, 2919, 2359, 1717, 1507, 1490, 1225, 1017, 821, 756, 724 cm}^{-1}. \\
\text{Method 2-HEMICYANINE 2} \\
\text{To 4-aminophenol PS 18 (0.5 g, 0.8 mmol) was added a solution of compound 19e (0.9 g, 1.8 mmol) in DCM (5 mL) and the mixture stirred at room temperature for 1 h. Afterwards, the resin was isolated by filtration, washed extensively with DMF, MeOH, DCM to remove the symmetrical dye and dried in vacuo to give the product as a dark resin (0.77 mmolg\(^{-1}\), 83 % by N analysis). Found: N, 2.17; Calcd: N, 2.60. IR (neat) 3025, 2919, 2359, 1717, 1507, 1490, 1225, 1017, 821, 756, 724 cm}^{-1}. 
\end{align*}
\]
A suspension of 1,2,3,3-tetramethyl-3H-indolium iodide (3.0 g, 8.4 mmol) and glutaconaldehyde dianil hydrochloride (2 g, 7 mmol) in a (1:1) mixture of acetic acid and acetic anhydride (30 mL) was heated at 120 °C for 2.5 h. The completion of the reaction was carefully monitored by UV-visible spectrometry: hemicyanine intermediate and symmetrical dicarboxylic acid cyanine dye have an absorption maximum around 493 and 780 nm, respectively. After cooling, the reaction mixture was evaporated and the resulting solid was dissolved in DCM (20 mL) and repeatedly washed with water (15 mL). After removal of the solvent in vacuo, the crude product was precipitated in Et₂O, collected by filtration and dried to give the compound 19f as a dark red-purple solid.

\[ R_f : 0.36 \text{ (DCM/MeOH 9:1)} \]

\[ \text{IR (neat)} : 3315, 2930, 1705, 1666, 1242, 1091, 923, 756 \text{ cm}^{-1} \]

\[ \text{HPLC } t_R = 6.91 \text{ min} \]

\[ \text{MS (ES): } m/z (\%) \begin{array}{ll}
421.1 [M^+] & 100, \\ 422.1 [(M+1)^+] & 30, \\ 423.1 [(M+2)^+] & 6
\end{array} \]

\[ \text{^1H NMR (500 MHz, CDCl}_3) \delta 8.18-8.14 (m, 2H, Ar + CH), 8.06-8.00 (m, 3H, Ar), 7.72-7.52 (m, 6H, Ar + CH), 7.40 (d, 1H, J = 15.1 Hz, CH), 7.29 (t, 1H, J = 9.0 Hz, CH), 7.17 (d, 2H, J = 7.0 Hz, Ar), 6.83 (t, 1H, J = 12.7 Hz, CH), 5.38 (t, 1H, J = 12.6 Hz, CH), 1.98 (s, 6H, CH₃), 1.96 (s, 3H, CH₃), 1.88 (s, 3H, CH₃).
\]

\[ \text{UV } (\text{MeOH}) \lambda_{abs} = 492 \text{ nm} \]
2-[(1E,3E,5E)-6-(acetyl-4-hydroxyanilino)-1,3,5-hexatrienyl]-1,1,3-trimethyl-1H-benzo[e]indolium iodide

To 4-aminophenol PS 18 (1 g, 1.7 mmol) was added a solution of compound 19f (2 g, 3.6 mmol) in DCM (10 mL) and the mixture stirred at room temperature for 1 h. Afterwards, the resin was isolated by filtration, washed extensively with DMF, MeOH, DCM to remove the symmetrical dye and dried in vacuo to give the product as a dark blue-black resin. IR (neat) 3024, 2915, 1555, 1494, 1157, 994, 941, 827, 745, 704 cm⁻¹

2-{[(1E,3E,5E,7E)-7-(5-carboxypentyl)-3,3-dimethyl-1,3-dihydro-2H-indol-2-ylidene]-1,3,5-heptatrienyl]-1,3,3-trimethyl-3H-indolium Inner Salt

Method 1-20e
To PS-bound Hemicyanine 15e-1 (0.4 g, 0.2 mmol) were added dry pyridine (2.5 mL), and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (71 mg, 0.2 mmol). The mixture was stirred at room temperature for 2 h. Afterwards, the resin was isolated by filtration and washed with DCM. The filtrates were combined and evaporated and the resulting solid was dissolved in DCM and washed several times with water. The product was precipitated in Et₂O, washed with Et₂O and then purified by chromatography (DCM → DCM/MeOH 1:1) to give 20e as a green solid (14 mg, 12%).

HPLC \( t_R = 5.9 \text{ min} \), 100% (ELSD, \( \lambda = 760 \text{ nm} \)) Method 1

*based on theoretical loading of resin 15e-1

172
Method 2-20e

To PS-bound Hemicyanine 15e-2 (0.84 g, 0.64 mmol) were added dry pyridine (4 mL), Ac₂O (0.6 mL, 6.3 mmol) and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (40 mg, 0.11 mmol). The mixture was stirred at room temperature for 2 h, until completion. Afterwards, the resin was isolated by filtration and washed (2 × 10 mL) with DCM. The reaction was then repeated in the same conditions described above. The filtrates were then combined and evaporated and the resulting solid was dissolved in DCM (10 mL) and washed several times with water (8 mL). The product was precipitated in Et₂O, washed with Et₂O and dried in vacuo giving 8c as a green solid (96 mg, 86%).

Rf: 0.32 (DCM/MeOH 9:1)

IR (neat) 2931, 2362, 1719, 1436, 1399, 1352, 1060, 1006, 915, 724, 704 cm⁻¹

HPLC tr = 5.89 min, 96 % (ELSD)

MS (ES): m/z (%) 509.5 [M⁺, 100], 510.5 [(M+1)⁺, 37], 511.5 [(M+2)⁺, 8]

¹H NMR (500 MHz, MeOD) δ 7.93 (t, (dd) 2H, J₁,₂ = J₂,₃ = J₅,₆ = J₆,₇ = 13.1 Hz, CH-H6 + CH-H2), 7.60 (t, (dd) 1H, J₃,₄ = J₄,₅ = 12.0 Hz, CH-H4), 7.49 (d, 1H, J = 7.5 Hz, Ar-H4 or Ar-H4'), 7.48 (d, 1H, J = 7.5 Hz, Ar-H4 or Ar-H4'), 7.39 (td (ddd), 2H, J₅,₆ = J₆,₇ = 7.7, J₄,₆ = 1.1 Hz, Ar-H6, Ar-H6'), 7.27-7.22 (m, 4H, Ar-H5, Ar-H5', Ar-H7, Ar-H7'), 6.61 (t, (dd), 1H, J = 12.6 Hz, CH-H3 or H5), 6.58 (t, (dd) 1H, J = 12.5 Hz, CH-H3 or H5), 6.30 (d, 1H, J = 13.7 Hz, CH-H1 or CH-H7), 6.27 (d, 1H, J = 13.7 Hz, CH-H1 or CH-H7), 4.07 (t, 2H, J = 7.5 Hz, CH2-e), 3.59 (s, 3H, NCH₃), 2.29 (t, 2H, J = 7.3 Hz, CH2-α), 1.82 (quintet, 2H, J = 7.7 Hz, CH2-δ), 1.70-1.67 (m + s, 14H, CH2-β + CH3), 152-1.46 (m, 2H, 2x CH2-γ).

¹H NMR (360 MHz, MeOD) δ 7.93 (t, 2H, J = 13.1 Hz, CH), 7.60 (t, 1H, J = 12.0 Hz, CH), 7.51-7.39 (m, 4H, Ar), 7.34-7.23 (m, 4H, Ar), 6.61 (t, 1H, J = 12.4 Hz, CH), 6.58 (t, 1H, J = 12.4 Hz, CH), 6.35 (d, 1H, J = 13.7 Hz, CH), 6.26 (d, 1H, J = 13.6 Hz, CH), 4.12 (t, 2H, J = 7.6 Hz, CH2), 3.61 (s, 3H, NCH3), 2.21 (t, 2H, J = 7.3 Hz, CH2), 1.89-1.67 (m + s, 16H, 2x CH2 + CH3), 158-1.49 (m, 2H, CH2).

¹³C NMR (90.5 MHz, MeOD) δ 178.6 (COOH), 173.8 (C=N), 173.2 (C=N), 153.0 (CH), 144.4 (C), 143.7 (C), 143.3 (C), 142.5 (C), 142.3 (C), 141.8 (C), 129.8 (CH), 129.7 (CH), 127.0 (CH), 126.0 (CH-Ar), 124.7 (CH-Ar), 123.4 (CH-Ar), 123.3 (CH-Ar), 111.9 (CH-Ar), 111.6 (CH-Ar), 104.8 (CH), 50.4 (C), 50.2 (C), 44.9
To PS-bound Hemicyanine 15f (0.7 g, 0.7 mmol)* were added dry pyridine (3.2 mL), Ac₂O (0.8 mL, 8.4 mmol) and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium bromide (34 mg, 0.08 mmol). The mixture was stirred at room temperature for 2 h, until completion. Afterwards, the resin was isolated by filtration and washed with DCM (2 x 10 mL). The reaction was then repeated in the same conditions described above. The filtrates were combined and evaporated, the resulting solid was dissolved in DCM (15 mL) and washed several times with water (10 mL). The product was then precipitated in Et₂O, washed with Et₂O and dried in vacuo giving 20f as a green solid (50 mg, 51%).

*based on theoretical loading of the resin 15f

**Rf**: 0.39 (DCM/MeOH 9:1)

**IR** (neat) 2924, 2457, 1725, 1410, 1087, 1008, 920, 897, 729 cm⁻¹

**HPLC** \( t_R = 4.20 \text{ min}, 86\% (\lambda_{abs}= 800 \text{ nm}) \)

**MS** (ES): \( m/z (\%) 609.2 [M^+, 100], 610.2 [(M+1)^+, 40], 611.2 [(M+2)^+, 10] \)

**¹H NMR** (500 MHz, MeOD) \( \delta 8.18 (d, 2H, J = 8.6 \text{ Hz, Ar-H9, Ar-H9'}) \), 8.03-8.00 (m, 2H, CH-H₂ + CH-H₆), overlapping with 8.00 (d, 1H, \( J = 8.8 \text{ Hz, Ar-H6 or Ar-H6'} \)), 7.99 (d, 1H, \( J = 8.8 \text{ Hz, Ar-H6 or Ar-H6'} \)), 7.95 (d, 2H, \( J = 8.30 \text{ Hz, Ar-H4, Ar-H4'} \)), 7.65-7.62 (m, 1H, CH-H₄) overlapping with 7.59 (td, (ddd) 2H, \( J = 8.3 \),...
1.1 Hz, Ar-H8, Ar-H8'), 7.58 (d, 1H, J = 8.9 Hz, Ar-H5), 7.56 (d, 1H, J = 9.0 Hz, Ar-H5'), 7.42 (bt (dd), 2H, J6,7 ≈ J7,8 = 7.5 Hz, Ar-H7, Ar-H7'), 6.56 (t (dd), 1H, J = 12.4 Hz, CH-H3 or H5), 6.55 (t (dd), 1H, J = 12.4 Hz, CH-H3 or H5), 6.27 (d, 1H, J = 13.7 Hz, CH-H1 or H7), 6.26 (d, 1H, J = 13.7 Hz, CH-H1 or H7), 4.19 (t, 2H, J = 7.5 Hz, CH2-ε), 3.72 (s, 3H, CH3), 2.33 (t, 2H, J = 7.3 Hz, CH2-α), 1.92 (s, 3H, NCH3), 1.85 (quintet, 2H, J = 7.7 Hz, CH2-δ), 1.70 (quintet, 2H, J = 7.3 Hz, CH2-β), 1.52 (quintet, 2H, J = 7.8 Hz, CH2-γ).

13C NMR (126 MHz, MeOD) δ 182.4 (COO), 177.9 (C=N), 177.4 (C=N), 150.8 (CH), 141.7 (C), 140.9 (C), 135.0 (C), 134.8 (C), 133.7 (C), 132.1 (C), 131.9 (CH-Ar), 131.3 (CH-Ar), 131.2 (CH-Ar), 129.4 (CH), 129.3 (CH), 128.9 (CH), 126.4 (CH-Ar), 126.3 (CH-Ar), 123.4 (CH-Ar), 112.4 (CH-Ar), 112.2 (CH-Ar), 103.4 (CH), 103.3 (CH), 52.5 (C), 52.4 (C), 45.4 (CH2), 36.7 (CH2), 33.2 (CH3), 29.8 (CH2), 28.9 (CH3), 28.8 (CH3), 28.7 (CH2), 27.3 (CH2).

UV/vis (MeOH) λabs = 781, ε = 1.3 × 10^5 M^{-1} cm^{-1}, λem = 808 nm

HRMS for C42H45N2O2^+ (M)^+: calcld 609.34756, found 609.34752

2-[(E)-2-(acetylanilino)ethenyl]-1-(5-carboxypentyl)-3,3-dimethyl-3H-indolium-5-sulfonate

A mixture of 26b (500 mg, 1.41 mmol) and N,N'-diphenylformamidine (306 mg, 1.56 mmol) was heated at 100 °C in a solution (1:1) of acetic acid and acetic anhydride (3 mL) for 20 min until an orange-red colour was noticed. The reaction was carefully monitored by UV-visible spectrometry to avoid the formation of the symmetrical dye: hemicyanine intermediate 28a and symmetrical dicarboxylic acid have an absorption maximum at 380 and 550 nm, respectively. After cooling the reaction mixture, the solvent was evaporated and the solid was washed with ethyl
acetate to remove the excess of N,N'-diphenylformamidine. After drying *in vacuo*,
the intermediate 28a, isolated as an orange-red solid, was directly used for the next step.

**UV** = (MeOH) $\lambda_{\text{abs}} = 380$ nm

**$^1$H NMR** (360 MHz, D$_2$O) $\delta$: 9.12 (d, 1H, $J = 13.7$ Hz, CH), 7.98 (s, 1H, Ar), 7.86 (d, 1H, $J = 8.5$ Hz, Ar), 7.64 (d, 2H, $J = 7.1$ Hz, Ar), 7.57 (d, 1H, $J = 8.5$ Hz, Ar), 7.39 (d, 2H, $J = 7.0$ Hz, Ar), 7.19-7.15 (m, 1H, Ar), 5.43 (d, 1H, $J = 13.4$ Hz, CH$_2$), 3.97 (t, 2H, $J = 7.6$ Hz, CH$_2$), 2.22 (t, 2H, $J = 7.6$ Hz, CH$_2$), 1.99 (s, 3H, CH$_3$), 1.71-1.57 (m, 8H, CH$_2$ + (CH$_3$)$_2$), 1.36-1.28 (m, 2H, CH$_2$), 1.03 (quintet, 2H, $J = 7.0$ Hz, CH$_2$).

**MS** (ES): $m/z$ (%) 455.0 [(M-1)$^+$, 100], 456.0 [M$, 25$, 457.0 [(M+1)$^+$, 5], 497.0 [(M+COCH$_3$)$^+$, 50], 498.0 [(M+COCH$_3$)$^+$, 14],

**HPLC** = 2.4 min

**R$_f$:** 0.33 (DCM: MeOH 9:1)

**Potassium 1-(5-carboxypentyl)-3,3-dimethyl-2-[(1E,3E)-3-(1,3,3-trimethyl-5-sulfonato-1,3-dihydro-2H-indol-2-ylidene)-1-propenyl]-3H-indolium-5-sulfonate**

The intermediate 28a (654 mg, 1.06 mmol) was dissolved in Ac$_2$O/pyridine (1:1) (2-3 mL). Indolenine salt 26a (444 mg, 1.06 mmol) was added and the mixture heated at 80 °C for 1h. After cooling the reaction mixture and removing the solvent *in vacuo*, the residue was washed with diethyl ether, dried, and then purified by silica gel chromatography eluting with a DCM: MeOH gradient (100% DCM to 50% DCM/MeOH). Product 25a was obtained as a red solid.

**Yield:** 248 mg (0.38 mmol, 36 %)

**$^1$H NMR** (360 MHz, CD$_3$OD) $\delta$: 8.57 (t (dd), 1H, $J_{1,2} \approx J_{2,3} = 13.5$ Hz, CH-H2), 7.95 (d, 2H, $J = 1.5$ Hz, H-16, Ar-H4), 7.94 (dd, 1H, $J = 8.0, 1.3$ Hz, Ar-H6 or Ar-
H6'), 7.93 (dd, 1H, J = 8.0, 1.6 Hz, Ar -H6, Ar-H6'), 7.43 (d, 1H, J = 8.3 Hz, Ar -H7 or Ar-H7'), 7.40 (d, 1H, J = 8.3 Hz, Ar -H7 or Ar-H7'), 6.57 (d, 1H, J = 13.4 Hz, CH-H1 or H3), 6.55 (d, 1H, J = 13.4 Hz, CH-H1 or H3), 4.17 (t, 2H, J = 7.5 Hz, CH-H1 or H3), 3.72 (s, 3H, NCH3), 2.20 (t, 2H, J = 7.3 Hz, CH2-a), 1.87-1.79 (quintet + s, 14H, J = 7.3 Hz, 2 × CH3 + CH2-δ), 1.70 (quintet, 2H, J = 7.5 Hz, CH2-β), 1.56-1.47 (m, 2H, CH2-γ).

13C NMR (90.5 MHz, MeOD): δ 181.7 (COO'), 177.1 (C=N), 176.4 (C=N), 152.4 (CH), 145.0 (C-Ar), 144.3 (C-Ar), 143.6 (C-Ar), 141.8 (C-Ar), 141.7 (C-Ar), 127.9 (CH-Ar), 127.8 (CH-Ar), 121.1(CH-Ar), 120.1 (CH-Ar), 111.8 (CH-Ar), 111.6 (CH-Ar), 104.5 (CH), 104.3 (CH), 50.4 (C), 50.3 (C), 45.1 (CH2), 38.0 (CH2), 31.7 (CH3), 27.8 (CH3), 27.7 (CH3), 27.3 (CH2), 26.5 (CH2).

MS ES': 307.0 ((M/2)², 100 %), 615.1 (M⁻¹, 20%)

HPLC tR = 7.14 min.

FTIR 3446 (m), 2970 (w), 2360 (w), 1738 (s), 1556 (s), 1445 (s), 1365 (s), 1216 (s), 1107 (s), 1018 (s), 871 (m) cm⁻¹

Rt: 0.38 (DCM: MeOH 1:1)

Mp >300 °C

UV/vis (H2O) λex = 548 nm, λem = 561 nm

HRMS for C30H35N2O8S2 (M)': calcd 615.18293, found 615.18444.

177
2-[(1E,3E)-4-(acetylaniino)-1,3-butadienyl]-1,3,3-trimethyl-3H-indolium-5-sulfonate

A mixture of 1,2,3,3-tetramethyl-3H-indolium-5-sulfonate (1.0 g, 2.4 mmol) and N-[(1E,3E)-3-(phenylimino)-1-propenyl]aniline hydrochloride (0.7 g, 2.7 mmol) was heated at 120 °C in a solution (1:1) of acetic acid and acetic anhydride (4 mL) for 2 h until an orange colour was noticed. The reaction was carefully monitored by UV-visible spectrometry: hemicyanine intermediate 3c and symmetrical dicarboxylic acid have an absorption maximum at 445 and 650 nm, respectively. After cooling, the solvent was evaporated and the solid precipitated with diethyl ether, filtered and washed several times with ethyl acetate. After drying *in vacuo*, the intermediate, isolated as an orange solid, was directly used for the next step.

**Rf**: 0.63 (DCM/MeOH 4:1)

**IR (neat)** 3475, 3069, 2978, 2358, 1694, 1574, 1512, 1468, 1167, 1111, 1027, 995 cm⁻¹

**HPLC tR = 4.74 min**

**MS (ES): m/z (%) 423.0 [(M-1), 100], 424.1 [M, 23], 847 [(2M-1), 7]**

**1H NMR** (360 MHz, D₂O) δ 8.55 (d, 1H, J = 12.8 Hz, CH), 8.00-7.90 (m, 2H, CH), 7.78-7.53 (m, 5H, Ar), 7.38-7.30 (m, 3H, Ar), 6.42 (d, 1H, J = 15.1 Hz, CH), 5.55 (t, 1H, J = 12.8 Hz, CH), 1.66-1.64 (m, 12H).

**UV = (MeOH) λabs = 445 nm**
2-[(1E,3E)-4-(acetyl-4-hydroxyanilino)-1,3-butadienyl]-1,3,3-trimethyl-3H-indolium-5-sulfonate PS

To 4-aminophenol PS 18 (0.35 g, 0.6 mmol) was added a solution of 28b (0.50 g, 1.2 mmol) in DCM/DMF (3:1) (4 mL) and the mixture stirred at room temperature for 1 h. Afterwards, the resin was isolated by filtration, washed extensively with DMF, MeOH, DCM and dried in vacuo to give the product as a dark resin (0.75 mmol g⁻¹, 75% by nitrogen elemental analysis). Found: N, 2.10; Calcd: N, 2.80.

IR (neat) 3024, 2917, 1667, 1538, 1507, 1451, 1169, 1026, 822, 755, 704 cm⁻¹

Potassium 2-[(1E,3E,5E)-5-[1-(5-carboxypentyl)-3,3-dimethyl-5-sulfonato-1,3-dihydro-2H-indol-2-ylidene]-1,3-pentadienyl]-1,3,3-trimethyl-3H-indolium-5-sulfonate

Classical Method
A solution of compound 28b (225 mg, 0.53 mmol) and compound 26b (187 mg, 0.53 mmol) in Ac₂O/pyridine (1:1) (2-3 mL) was stirred at 80 °C 1h. After cooling and removing the solvent in vacuo, the residue was washed with diethyl ether, dried, and then purified by chromatography (DCM → DCM/MeOH 1:1) to give the product 25b as a blue solid (142 mg, 40%).

Solid Phase Method
To PS-bound Hemicyanine 30b (0.35 g, 0.26 mmol) were added dry pyridine (1.6 mL), Ac₂O (0.4 mL, 4.2 mmol) and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium-5-sulfonate 26b (20 mg, 0.06 mmol). The mixture was stirred at room temperature...
temperature for 2 h, until completion. Afterwards, the resin was isolated by filtration, washed with a solution 9:1 DCM/DMF (2 × 5 mL) and the reaction repeated in the same conditions described above. The resin was isolated by filtration and then washed with a mixture 9:1 DCM/DMF (3 × 5 mL). The filtrates were combined and evaporated. Afterwards, the resulting solid was washed and precipitated with Et₂O and centrifuged. The supernatant was removed and the precipitate dried in vacuo to give 25b as a blue solid (77 mg, 94%).

Rᵣ: 0.40 (DCM/MeOH 3:2)

IR (neat) 3413, 2929, 1716, 1494, 1459, 1368, 1330, 1153, 1099, 1016, 926, 817 cm⁻¹

HPLC tᵣ = 3.6 min, 94% (λₘₐₓ = 650 nm)

MS (ES): m/z (%) 320.0 [(M/2)², 100], 641.0 [(M-2), 40], 642.0 [(M-1)ᵣ, 21], 643.0 [M-R, 8]

¹H NMR (360 MHz, MeOD) δ: 8.31 (t (dd), 2H, J₁,₂ ≈ J₂,₃ ≈ J₃,₄ ≈ J₄,₅ = 13.4 Hz, CH-H₂ + CH-H₄), 7.90-7.85 (m, 4H, Ar-H₄, Ar-H₄', Ar-H₆, Ar-H₆'), 7.36 (d, 1H, J = 8.2 Hz, Ar-H₇ or Ar-H₇'), 7.33 (d, 1H, J = 8.8 Hz, Ar-H₇ or Ar-H₇'), 6.70 (t (dd), 1H, J₂,₃ ≈ J₃,₄ = 12.2 Hz, CH-H₃), 6.36 (d, 1H, J = 13.1 Hz, CH-H₁ or CH-H₅), 6.32 (d, 1H, J = 13.2 Hz, CH-H₁ or CH-H₅), 4.12 (t, 2H, J = 7.3 Hz, CH₂-α), 3.64 (s, 3H, NCH₃), 2.20 (t, 2H, J = 7.4 Hz, CH₂-α), 1.20-1.56(m, 16H, CH₂-β + CH₂-γ + CH₂-β + 2 × CH₃), 1.56-1.47 (m, 2H, CH₂-γ).

¹³C NMR (90.5 MHz, MeOD): δ 182.4 (COO⁻), 175.9 (C=N), 175.4 (C=N), 156.2 (CH), 145.7 (C-Ar), 144.9 (C-Ar), 143.4 (C-Ar), 143.3 (C-Ar), 142.7 (C-Ar), 142.5 (C-Ar), 128.1 (CH-Ar), 128.0 (CH-Ar), 127.8 (CH), 121.4 (CH-Ar), 121.3 (CH-Ar), 111.7 (CH-Ar), 111.4 (CH-Ar), 104.4 (CH), 105.3 (CH), 50.6 (C), 50.5 (C), 45.2 (CH₂), 38.8 (CH₂), 31.8 (CH₃), 30.7 (CH₃), 28.3 (CH₂), 27.9 (CH₃), 27.8 (CH₃), 27.7 (CH₂), 27.2 (CH₂).

UV/vis (MeOH) λₘₐₓ = 647, λₑₘᵋₑ = 667 nm

HRMS for C₃₂H₃₇N₂O₂S₂⁻ (M⁻): calcd 641.19858, found 641.19658
6.2 Experimental for Chapter 3

High performance liquid chromatography was carried out on an Agilent Technologies HP1100 Chemstation system coupled to a Polymer Lab 100 ES evaporative light scattering detector, eluting with (A) 0.1% TFA/H$_2$O, (B) 0.04% TFA/MeCN.

**HPLC System III**

Column: Phenomenex Luna, C18(2), 150 mm × 4.6 mm, 100 Å. Flow rate: 1.0 mL min$^{-1}$. Eluents A and B. Gradient: 95% A to 5% A over 8 min, then 5% A for 6 min, then 5% to 95% A over 1 min. Detection: UV and or ELS detection.

Preparative RP-HPLC was performed on a Hewlett Packard HP1100 Chemstation eluting with (A) 0.1 M TEAB in H$_2$O, (B) MeCN. The fractions were collected using a Gilson G2250A robot.

**HPLC System IV**

Column: Waters (Xterra) RP18 OBD preparative column, 19 mm × 150 mm i.d. 5 μm at a flow rate of 10 mL min$^{-1}$. Gradient: 100 % A to 80 % A over 4 min, 80 % A to 50 % A over 12 min, 50 % A to 0 % A over 2 min, 0% A for 1 min, 0% to 100% A over 1 min. Detection by UV at 230, 254, 260 and 282 nm.
**Preparation of TEAB buffer**

TEAB buffer was prepared by bubbling CO₂ through a solution of deionised water (650 mL) and TEA (278 mL) over an ice-bath until pH 8.5. The solution was diluted to 1 L to give a 2 M TEAB buffer and was stored at 4 °C.

**Preparation of Solid Phase Bound Peptide 32.**

![Peptide Structure](image)

**Attachment of the 1st amino acid to the Wang resin**

Wang resin (1g, 0.93 mmol) was placed in a disposable peptide vessel and left to swell in 10 mL of DMF for 30 min. In another flask, a solution of Fmoc-Ala-OH (869 mg, 2.79 mmol, 3 eq.) and DIC (352 mg, 2.79 mmol, 3 eq.) in DCM/DMF (2:1) was left under agitation for 10 min. DMAP (34 mg, 0.28 mmol, 0.1 eq.) was added to the solution and the mixture was then added to the resin and left under agitation for 1h. The progress of the coupling was monitored by a qualitative ninhydrin test and the coupling procedure was repeated to push the reaction to completion. The resin was washed with DMF (3 ×10 mL), CH₂Cl₂ (3 ×10 mL), MeOH (3 ×10 mL), CH₂Cl₂ (3 ×10 mL), Et₂O (3 ×10 mL). Afterwards, the resin was dried in vacuo and the loading of Fmoc-Alanine, was measured by a standard Fmoc test, 0.91 mmol/g (98%). The resin was swollen in CH₂Cl₂ and the solution drained off, 20% piperidine in DMF was added, the resin was agitated for 30 minutes (twice) and washed with DMF (3 ×10 mL), CH₂Cl₂ (3 ×10 mL), MeOH (3 ×10 mL), CH₂Cl₂ (3 ×10 mL) Et₂O (3 ×10 mL) and dried under reduced pressure.
**Solid Phase Peptide Synthesis**

To a solution of Fmoc-Ala-OH (3 eq.) and HOBt (3 eq.) in DCM/DMF (2:1) was added DIC (3 eq.). After 10 min, the reaction mixture was added to the preloaded resin and stirred for 2 h. The progress of the coupling was monitored by a qualitative ninhydrin test and the coupling procedure was repeated if necessary. The resin was washed with DMF (3 x 20 mL), CH₂Cl₂ (3 x 20 mL), MeOH (3 x 20 mL), CH₂Cl₂ (3 x 20 mL), and dried under reduced pressure. A solution of 20% piperidine in DMF was added, the resin was stirred for 30 minutes (twice) and washed with DMF (3 x 10 mL), MeOH (3 x 10 mL), CH₂Cl₂ (3 x 10 mL), CH₂Cl₂ (2 x 10 mL), MeOH (2 x 10 mL), DMF (2 x 10 mL) and dried under reduced pressure. The procedure was repeated for each aminoacid in the sequence.

**Cleavage Procedure (for HPLC and MS analysis)**

95% TFA in DCM (2 mL) was added to the resin (~ 10 mg) and left to stand for 10 minutes. The resin was washed with DCM (3 x 3mL). After combining the filtrates, the mixture TFA/DCM was removed by evaporation under reduced pressure and cold diethyl ether was added to achieve a good precipitation.

ES⁺ m/z (%): 232.3 ((M+H)⁺, 100), 254.3 ((M+Na)⁺, 30), 463.5 ((2M+H)⁺, 10), 485.5 ((2M+Na)⁺, 25).

**HPLC** (λ220) tR 11.1 min

¹H NMR (360 MHz, D₂O) δ: 4.26 (q, 1H, J = 7.3 Hz, CH-Ala), 4.25 (q, 1H, J = 7.2 Hz, CH-Ala), 3.99 (q, 1H, J = 7.1 Hz, CH-Ala), 1.44 (d, 3H, J = 7.1 Hz, CH₃-Ala), 1.35 (d, 3H, J = 7.3 Hz, CH₃-Ala), 1.32 (d, 3H, J = 7.2 Hz, CH₃-Ala).
Succinimidyl ester of pyridinium 2-\{(1E,3Z)-3-\{(1-(5-carboxypentyl)-3,3-dimethyl-5-sulfonato-1,3-dihydro-2H-indol-2-ylidene)-1-propenyl\}-1,3,3-trimethyl-3H-indolium-5-sulfonate

Dye 25a (117 mg, 0.17 mmol) was dissolved in a mixture of dry DMF (2 mL) and dry pyridine (2 mL). Disuccinimidyl carbonate (DSC) (56 mg, 0.21 mmol) was added, and the mixture was stirred at 60 °C for 1h, under a nitrogen atmosphere. The reaction was monitored by TLC (DCM:MeOH 3:2). 20 mg (0.08 mmol) of DSC were added to the reaction mixture to push the reaction to completion. After dilution of the mixture with dry ethyl acetate, the suspension was centrifuged and the supernatant decanted. The product, obtained as pyridinium salt, was washed repeatedly with dry ethyl acetate (2 mL) and dried \textit{in vacuo}. Nearly quantitative yields of active esters were obtained. The active ester 33 was used without further purification.

\textbf{MS ES}: 355.5 ((M-2H)^2, 7 \%), 367.1 ((M-H+Na)^2, 100 \%), 368.1 ((M+Na)^2, 23 \%), 712.0 ((M-2H)^1, 25 \%), 713.0 ((M-H)^1, 12 \%), 714.0 (M^1, 5 \%).

\textbf{R}_f: 0.80 (DCM: MeOH 4:1)
The active ester 33 (106 mg, 0.14 mmol) was dissolved in DMF (2 mL) and allowed to react with the solid-phase bound peptide Ala-Ala-Ala-H (100 mg, 0.051 mmol), under agitation, for 3 h. The reaction was monitored by qualitative ninhydrin test and repeated if required. After completion of the labelling reaction, the resin was washed several times with water (4 mL), DMF (4 mL) and DCM (4 mL) to remove the excess of the unbound dye. The labelled peptide was then cleaved from the resin with (95:5) TFA/water (8 mL) for 30 min; the cleavage was repeated twice. The labelled peptide was precipitated with cold diethyl ether, centrifuged and dried in vacuo to give a red solid. No further purification was required for use in the next step.

Yield: 32 mg (0.034 mmol, 67 %)

HPLC $t_R = 5.67$ min

ES$^-$ m/z (%): 413.8 ((M-3H/2)$^2$, 100), 567.2 ((M/2+CF$_3$COO$^-+K^+$)$^2$, 15), 568.1 ((M/2+H+ CF$_3$COO$^-+K^+$)$^2$, 10), 829.3 ((M-H)$^-$, 80)

$^1$H NMR (360 MHz, D$_2$O) δ: 8.30 (t (dd), 1H, J$_{1,2}$ ≈ J$_{2,3}$ = 13.5 Hz, CH-H2), 7.74 (s, 2H, Ar-H4, Ar-H4'), 7.66 (d, 2H, J = 8.2 Hz, Ar-H6, Ar-H6'), 7.15 (d, 1H, J = 8.2 Hz, Ar-H7 or Ar-H7'), 7.12 (d, 1H, J = 8.3 Hz, Ar-H7 or Ar-H7'), 6.20 (d, 1H, J = 12.7 Hz, CH-H1 or CH-H3), 6.17 (d, 1H, J = 13.4 Hz, CH-H1 or CH-H3), 4.16 (q, 1H, J = 7.2 Hz, CH-Ala), 4.15 (q, 1H, J = 7.1 Hz, CH-Ala), 4.04 (q, 1H, J = 7.2 Hz, CH-Ala), 3.88 (m, 2H, H-23, CH$_2$-e), 3.40 (s, 3H, NCH$_3$), 2.12 (t, 2H, J = 6.6 Hz,
CH2-α), 1.68-1.45 (m, 18H, CH2-δ + CH2-β + 2 × CH3, CH2-γ), 1.25 (d, 3H, J = 7.3 Hz, CH3-Ala), 1.23 (d, 3H, J = 7.1 Hz, CH3-Ala), 1.18-1.12 (m, 3H, CH3-Ala).

13C NMR (90 MHz, D2O): δ 177.4 (C = N), 177.2 (COOH), 176.6 (CONH), 175.7 (CONH), 175.5 (CONH), 152.7 (CH, C-2), 145.6 (C-Ar), 145.1 (C-Ar), 142.4 (C-Ar), 142.3 (C-Ar), 140.6 (C-Ar), 127.7 (CH, Ar-H6, Ar-H6'), 120.8 (CH, Ar-H4 or Ar-H4'), 120.6 (CH, Ar-H4 or Ar-H4'), 112.4 (CH, Ar-H7 or Ar-H7'), 112.2 (CH, Ar-H7 or Ar-H7'), 105.6, 104.6 (CH, C-1 and C-3), 50.5 (C) 50.2 (C) 50.1 (CH-Ala), 49.6 (CH-Ala), 47.7 (CH-Ala), 45.0 (CH2-ε), 35.9 (CH2-α), 32.2 (CH3, NCH3), 28.1 (CH3), 27.9 (CH3), 27.6 (CH2-δ), 26.4 (CH2-β), 25.8 (CH2-γ), 17.7 (CH3-Ala), 17.6 (CH3-Ala), 17.1 (CH3-Ala).

FTIR 3295 (m), 2980 (m), 1652 (s), 1552 (s), 1447 (s), 1412 (s), 1176 (s), 1114 (s), 1066 (s) cm⁻¹

Mp at 248 °C (change of colour and bubbling)

Rf: 0.39 (DCM: MeOH 13:7)

UV/vis (H2O) λex= 548 nm, λem= 561 nm


**Synthesis of dUTP-(Ala)₃-Cy3**

N, N-disuccinimidyl carbonate (1.7 mg, 6.6 μmol) was added to a solution of 34 (4.0 mg, 4.2 μmol) in a mixture of dry DMF (95 μL) and dry pyridine (5 μL). The reaction mixture was stirred under nitrogen at 50-60 °C for 1.5 h and then diluted with dry ethyl acetate (2 mL). After centrifugation, the supernatant was decanted and the precipitate washed repeatedly with dry ethyl acetate. To the resulting red solid, dissolved in 100 μL of 0.1M TEAB buffer (pH 8.5) were added 100 μL of 5-
(3-aminopropargyl)-2′-deoxyuridine 5′-triphosphate (2.0 mg, 2.2 μmol) in the same buffer solution. The reaction mixture was stirred at room temperature for 30 min-1 h and then purified by reverse-phase HPLC to afford the nucleotide 35 dUTP-(Ala)_3-Cy3 in 2.1% yield (50 nmol). The purified fraction was freeze-dried and the lyophilised red solid obtained characterised by MALDI-TOF analysis and stored at −20 °C.

MALDI-TOF m/z: anal.calcd, 1333; found 1333 (M), 1356 (M+Na).

HPLC (λ_{565}) {t_R} = 16.5 min

UV/vis (1 × TE buffer) λ_{max}= 550 nm

Succinimidyl ester of pyridinium 2-[(1E,3E,5Z)-5-11-(5-carboxypentyl)-3,3-dimethyl-5-sulfonato-1,3-dihydro-2H-indol-2-ylidene]-1,3-pentadienyl]-1,3,3-trimethyl-3H-indolium-5-sulfonate

Dye 25b (70 mg, 0.10 mmol) was dissolved in a mixture of dry DMF (1 mL) and dry pyridine (1 mL). Disuccinimidyl carbonate (DSC) (33 mg, 0.13 mmol) was added, and the mixture was stirred at 60 °C for 1 h, under a nitrogen atmosphere. 10 mg (0.04 mmol) of DSC were added to the reaction mixture to push the reaction to completion. After dilution of the mixture with dry ethyl acetate, the suspension was centrifuged and the supernatant decanted. The product was washed repeatedly with dry ethyl acetate (2 mL) and dried in vacuo. Nearly quantitative yields of active esters were obtained.

HPLC {t_R} = 7.26 min

ES' m/z 738.2 (M-2H)^−, 739.1 (M-1H)^−, 740.1 (M).
Synthesis of a Cy5 Labelled Peptide: Cy5-(Ala)_3-OH

The active ester 36 (70 mg, 0.090 mmol) was dissolved in DMF (2 mL) and allowed to react with the solid-phase bound peptide PS-Ala-Ala-Ala-H (80 mg, 0.045 mmol), under agitation, for 3 h. The reaction was monitored by qualitative ninhydrin test and repeated if required. After completion of the labelling reaction, the resin was washed several times with water (4 mL), DMF (4 mL) and DCM (4 mL) to remove the excess of the unbound dye. The labelled peptide was then cleaved from the resin with (95:5) TFA/water (8 mL) for 30 min; the cleavage was repeated twice. The labelled peptide was precipitated with cold diethyl ether, centrifuged and dried in vacuo to give 38 mg (98%) of a blue solid which was purified by reverse-phase HPLC.

Yield: 26 mg (0.027 mmol, 60%)

HPLC \( t_R = 5.7 \) min

ES \( m/z \) 426.7 (M-3H/2)^2, 854.1 (M-2H)^-

\(^1\)H NMR (360 MHz, D\(_2\)O) \( \delta \): 7.97 (t (dd), 2H, J\(_{1,2} \approx J_{2,3} \approx J_{3,4} \approx J_{4,5} = 13.0 \) Hz, CH-H2+ CH-H4), 7.83 (d, 2H, \( J = 1.5 \) Hz, Ar-H4, Ar-H4'), 7.79 (dd, 2H, \( J = 8.4, 1.5 \) Hz, Ar-H6, Ar-H6'), 7.29 (d, 2H, \( J = 8.1 \) Hz, Ar-H7, Ar-H7'), 6.46 (t (dd), 1H, \( J_{2,3} \approx J_{3,4} = 12.5 \) Hz, CH-H3), 6.16 (d, 1H, \( J = 13.4 \) Hz, CH-H1 or CH-H5), 6.14 (d, 1H, \( J = 13.4 \) Hz, CH-H1 or CH-H5), 4.30 (q, 1H, \( J = 7.0 \) Hz, CH-Ala), 4.28 (q, 1H, \( J = 7.1 \) Hz, CH-Ala), 4.18 (q, 1H, \( J = 7.1 \) Hz, CH-Ala), 4.03 (t, 2H, \( J = 5.7 \) Hz, CH2-e), 3.56 (s, 3H, NCH3), 2.26 (t, 1H, \( J = 6.7 \) Hz, CH2-\( \alpha \)), 2.25 (t, 2H, \( J = 6.4 \) Hz, CH2-\( \alpha \)).
Hz, CH$_2$-$\delta$), 1.81-1.76 (m, 2H, CH$_2$-$\beta$), 1.65-1.60 (m, 14H, 2 x CH$_3$, CH$_2$-$\gamma$), 1.38 (d, 6H, J = 7.2 Hz, 2 x CH$_3$-Ala), 1.29 (d, 3H, J = 7.1 Hz, CH$_3$-Ala).

$^{13}$C NMR (90 MHz, D$_2$O): $\delta$ 177.7 (C = N), 177.4 (COOH), 175.8 (CONH), 175.4 (CONH), 174.7 (CONH), 152.8, 152.6 (CH, C-2, C-4), 145.9 (C-Ar), 145.4 (C-Ar), 142.7 (C-Ar), 140.1 (C-Ar), 140.0 (C-Ar), 127.5 (CH, Ar-H6, Ar-H6'), 126.9 (CH-3), 120.7 (CH, Ar-H4 or Ar-H4'), 120.6 (CH, Ar-H4 or Ar-H4'), 112.0 (CH, Ar-H7 or Ar-H7'), 111.9 (CH, Ar-H7 or Ar-H7'), 105.4, 104.9 (CH-1, CH-5), 50.5 (C), 50.3 (C), 50.1 (CH-Ala), 50.0 (CH-Ala), 49.9 (CH-Ala), 44.9 (CH$_2$-$\varepsilon$), 36.0 (CH$_2$-$\alpha$), 32.1 (NCH$_3$), 27.9 (CH$_2$-$\delta$), 27.8 (CH$_3$), 27.7 (CH$_3$), 27.6 (CH$_2$-$\beta$), 25.9 (CH$_2$-$\gamma$), 17.7 (CH$_3$-Ala), 17.6 (CH$_3$-Ala) 17.3 (CH$_3$-Ala).

FTIR 3276 (w), 2933 (w), 2362 (w), 1652 (m), 1497 (s), 1462 (s), 1367 (m), 1316 (m), 1192 (s), 1154 (s), 1098 (s), 1015 (s), 920 (s), 817 (s), 738 (s) cm$^{-1}$

Mp at 240 °C (change of colour and bubbling)

Rf: 0.46 (DCM: MeOH 13:7)

UV/vis (H$_2$O) $\lambda_{ex}$ = 647, $\lambda_{em}$ = 661 nm

HRMS for C$_{41}$H$_{53}$N$_5$O$_{11}$S$_2^-$ (M-H): calcd 854.3099, found 854.3100.
Synthesis of dCTP-(Ala)₃-Cy5 (sulfonated)

N,N-disuccinimidyl carbonate (2.0 mg, 7.4 μmol) was added to a solution of 37 (6.0 mg, 6.2 μmol) in a mixture of dry DMF (95 μL) and dry pyridine (5 μL). The reaction mixture was stirred under nitrogen at 50-60 °C for 1.5 h and then diluted with dry ethyl acetate (2 mL). After centrifugation, the supernatant was decanted and the precipitate washed repeatedly with dry ethyl acetate. To the resulting blue solid, dissolved in 100 μL of 0.1M TEAB buffer (pH 8.5) were added 100 μL of 5-(3-aminopropargyl)-2'-deoxycytidine 5'-triphosphate (3.0 mg, 3.2 μmol) in the same buffer solution. The reaction mixture was stirred at room temperature for 30 min-1 h and then purified by reverse-phase HPLC to afford the nucleotide 38 dCTP-(Ala)₃-Cy5 in 1% yield (30 nmol). The purified fraction was freeze-dried and the lyophilised blue solid obtained characterised by MALDI-TOF analysis and stored at −20 °C.

MALDI-TOF m/z: anal.calcd, 1360; found 1358 (M-2H), 1360 (M), 1380 (M-2H+Na).

HPLC (λ₆49) tᵣ = 17.9 min

UV/vis (1 × TE buffer) λₑₓₜₐₓ = 648 nm
Succinimidyl ester of 2-\{(1E,3E,5Z)-5-[1-(5-carboxypentyl)-3,3-dimethyl-1,3-dihydro-2H-indol-2-ylidene][1,3-pentadienyl]-1,3-trimethyl-3H-indolium bromide

Dye 20c (60 mg, 0.11 mmol) was dissolved in a mixture of dry DMF (1 mL) and dry pyridine (1 mL). Disuccinimidyl carbonate (DSC) (54 mg, 0.20 mmol) was added, and the mixture was stirred at 60 °C for 1 h, under a nitrogen atmosphere. After dilution of the mixture with dry diethyl ether, the suspension was centrifuged and the supernatant decanted. The product was washed repeatedly with diethyl ether (2 mL) and dried in vacuo. The active ester 39 was used without further purification.

\textbf{HPLC} t_R = 6.85 \text{ min}.

\textbf{ES}^+ m/z (%) 580.5 (M^+, 100), 581.5 ((M+H)^+, 40), 582.5 ((M+H)^2+, 10).
Synthesis of a Cy5 (Hydrophobic) Labelled Peptide: Cy5(H)-(Ala)3-OH

The active ester 39 (70 mg, 0.11 mmol) was dissolved in DMF (2 mL) and allowed to react with the solid-phase bound peptide PS-Ala-Ala-Ala-H (100 mg, 0.056 mmol), under agitation, for 3 h. The reaction was monitored by qualitative ninhydrin test and repeated if required. After completion of the labelling reaction, the resin was washed several times with DMF (4 mL), MeOH (4 mL) and DCM (4 mL) to remove the excess of the unbound dye. The labelled peptide was then cleaved from the resin with (95:5) TFA/DCM (8 mL) for 30 min; the cleavage was repeated twice. The labelled peptide was precipitated with cold diethyl ether, centrifuged and dried in vacuo to give 44 mg (98%) of a blue solid which was used for the next step without further purification.

$^1{\text{H}}$ NMR (500 MHz, MeOD) $\delta_H$ 8.24 (t, (dd), 2H, $J_{1,2} \approx J_{2,3} \approx J_{3,4} = J_{4,5} = 13.0$ Hz, CH-H2 + CH-H4), 7.49 (d, 2H, $J = 7.4$ Hz, Ar-H4, Ar-H4'), 7.41 (t (dd), 2H, $J_{5,6} \approx J_{6,7} \approx 7.5$ Hz, Ar-H6, Ar-H6'), 7.29 (d, 1H, $J = 7.5$ Hz, Ar-H7 or Ar-H7'), 7.28 (d, 1H, $J = 7.5$ Hz, Ar-H7 or Ar-H7'), 7.26 (t (dd), 2H, $J_{4,5} \approx J_{5,6} \approx 7.5$ Hz, Ar-H5, Ar-H5'), 6.62 ((t (dd), 1H, $J_{2,3} \approx J_{3,4} = 12.4$ Hz, CH-H3), 6.33 (d, 1H, $J = 13.8$ Hz, CH-H1 or H5), 6.29 (d, 1H, $J = 13.8$ Hz, CH-H1 or H5), 4.38 (q, 1H, $J = 7.1$ Hz, CH-Ala), 4.35 (q, 1H, $J = 7.1$ Hz, CH-Ala), 4.32 (q, 1H, $J = 7.1$ Hz, CH-Ala), 4.11 (t, 1H, $J = 7.33$ Hz, CH2, CH2-o), 3.63 (s, 3H, NCH3), 2.30 (t, 1H, $J = 7.4$ Hz, CH2-o),
2.28 (t, 1H, \( J = 7.1 \) Hz, CH2-\( \delta \)), 1.87-1.69 (m + s, 16H, CH2-\( \beta \), CH2-\( \gamma \) + 2 \( \times \) CH3.), 1.44-1.38 (m, 10 H, H-3\(5 \), 2 \( \times \) CH3-Ala), 1.36 (d, 3H, \( J = 7.2 \) Hz, CH3-Ala).

\( ^{13} \)C NMR (126 MHz, MeOD) \( \delta_{C} \) 176.0 (CN), 175.8 (CN), 175.3 (COOH), 174.8 (CONH), 174.7 (CONH), 174.5 (CONH), 155.6 (CH-2 or CH-4), 155.5 (CH-2 or CH-4), 144.3 (C-Ar), 143.6 (C-Ar), 142.7 (C-Ar), 142.6 (C-Ar), 129.8 (CH, Ar-H6 or Ar-H6'), 129.7 (CH, Ar-H6 or Ar-H6'), 126.8 (CH, H-3), 125.8, 125.7 (CH, Ar-H5, Ar-H5'), 123.4 (CH, Ar-H4 or Ar-H4'), 123.3 (CH, Ar-H4 or Ar-H4'), 112.1 (CH, Ar-H7 or Ar-H7'), 111.8 (CH, Ar-H7 or Ar-H7'), 104.4 (CH-1, CH-5), 51.0 (C), 50.6 (C), 50.5 (CH-Ala), 50.4 (CH-Ala), 50.1 (CH-Ala), 44.9 (CH2-\( \alpha \)), 36.2 (CH2-\( \alpha \)), 31.8 (NCH3), 28.2 (CH2-\( \delta \)), 27.9 (CH3), 27.8 (CH3), 27.2 (CH2-\( \beta \)), 26.4 (CH2-\( \gamma \)), 18.3 (CH3-Ala), 18.0 (CH3-Ala), 17.7 (CH3-Ala).

ESI± m/z (%) 696.6 (M+, 100), 413.4 (M12/2, 25)

HPLC \( t_{R} \) = 8.71 min

FTIR 3267 (w), 2930 (w), 1628 (s), 1481 (s), 1449 (s), 1368 (m), 1332 (m), 1146 (s), 1090 (s), 920 (s), 829 (s), 795 (s) cm\(^{-1}\)

\( \text{Mp} \) 123-125 °C

\( \text{Rf} \): 0.24 (DCM: MeOH 13:7)

UV/vis (MeOH) \( \lambda_{ex} = 642, \lambda_{em} = 661 \) nm

HRMS (ES) for C41H54N5O5+: calcd 696.41195, found 696.41173

Synthesis of dCTP-(Ala)3-Cy5 (Hydrophobic)

A solution of N, N-disuccinimidyl carbonate (2.8 mg, 11.0 \( \mu \)mol) in DMF (50 \( \mu \)L) was added to a solution of 40 (6.0 mg, 7.4 \( \mu \)mol) in a mixture of dry DMF (95 \( \mu \)L) and dry pyridine (5 \( \mu \)L). The reaction mixture was stirred under nitrogen at 50-60 °C for 1-2 h and then diluted with dry ethyl acetate (2 mL). After centrifugation, the
supernatant was decanted and the precipitate washed repeatedly with dry ethyl acetate. To the resulting blue solid, dissolved in 100 μL of a solution (1:9) acetonitrile/0.1M TEAB buffer (pH 8.5) were added 100 μL of 5-(3-aminopropargyl)-2'-deoxycytidine 5'-triphosphate (4.0 mg, 4.3 μmol) in the same buffer solution. The reaction mixture was stirred at room temperature for 30 min-1 h and then purified by reverse-phase HPLC to afford the nucleotide 41 dCTP-(Ala)₃-Cy5 in 1.4% yield (60 nmol). The purified fraction was freeze-dried and the lyophilised blue solid obtained characterised by MALDI-TOF analysis and stored at −20 °C.

**MALDI-TOF m/z:** anal.calcd, 1199; found 1197 (M-3H), 1199 (M).

**HPLC** ($\lambda_{469}$) $t_R = 28.8$ min

**UV/vis (1 × TE buffer) $\lambda_{max}$ = 642 nm**

$N$-{6-(diethylamino)-9-[2-(1-piperazinylcarbonyl)-phenyl]-3H-xanthen-3-ylidene}-$N$-diethyl-ammonium chloride$^{165}$

To a 2.0 M solution of trimethyl aluminium in toluene (2.30 mL, 4.5 mmol), a solution of piperazine (0.78 g, 9.1 mmol) in CH₂Cl₂ (3.5 mL) was added dropwise. After one hour of stirring, a white precipitate was observed. A solution of Rhodamine B base (1.0 g, 2.3 mmol) in 2.0 mL of CH₂Cl₂ was added dropwise to the heterogeneous solution. Gas evolution was observed during the adding period. After stirring at reflux for 24 h, 0.1 M aqueous solution of HCl was added dropwise until gas evolution cessed. The heterogeneous solution was filtered and the solid was rinsed with CH₂Cl₂ and a solution 4:1 CH₂Cl₂: MeOH. The filtrate was concentrated and the residue was dissolved in CH₂Cl₂, filtered to remove insoluble salts and concentrated again. The resulting glassy solid was then portioned between dilute aqueous NaHCO₃ and EtOAc. After isolation, the aqueous layer was washed
with 3 additional portions of EtOAc to remove residual starting material. The aqueous layer was saturated with NaCl, acidified with 0.1 M HCl and then extracted with 1:2 CH2Cl2: iPrOH. The organic layer was dried over anhydrous Na2SO4, filtered and the solvent was removed in vacuo. The glassy purple solid was then precipitated in Et2O and dried under high vacuum. The product 42 was obtained as a dark purple solid 0.76 g (67%).

**1H NMR** (360 MHz, CD3OD): δ 7.81-7.78 (2H, m, Ar-H3’, Ar-H5’), 7.70 (1H, m, Ar-H6’), 7.54 (1H, m, Ar-H4’), 7.32 (2H, d, J= 9.5 Hz, Ar-H1, Ar-H8), 7.07 (2H, dd, J= 9.6 Hz, J= 2.5 Hz, Ar-H2, Ar-H7), 7.02 (2H, d, J= 2.4 Hz, Ar-H4, Ar-H5), 3.73 (8H, q, J= 7.1 Hz, CH2-Et), 3.37 (4H, br s, CH2-2, CH2-6), 2.38 (4H, br s, CH2-3, CH2-5), 1.33 (12H, t, J= 7.1 Hz, CH3-Et).

**13C NMR** (90 MHz, CD3OD): δ 169.1 (C=O), 158.9 (C-Ar), 156.8, 156.6 (C-Ar), 136.4 (C-Ar), 132.9 (Ar-H1, Ar-H8), 131.7 (Ar-H4’), 131.4 (Ar-H3’, Ar-H5’), 131.0 (C-Ar), 130.8 (Ar-H6’), 128.5 (C-Ar), 115.1 (Ar-H2, Ar-H7), 114.5 (C-Ar), 97.0 (Ar-H4, Ar-H5), 46.6 (CH2-Et), 46.2 (CH2-2, CH2-6), 45.6 (CH2-3, CH2-5), 12.6 (CH3-Et).

**HPLC** (λ₂₆₅) tR 5.56 min (method 1).

**ES** m/z (%): 256.3 ((M+H)²⁺/2, 30), 511.6 (M⁺, 100) 512.7 (M+H⁺, 30).

**FTIR** 3331 (w), 2979 (w), 1626 (s), 1586 (s), 1445 (s), 1412 (s), 1334 (s), 1275 (s), 1252 (s), 1178 (s), 1132 (s), 1075 (m), 1046 (m), 977 (m), 848 (sm), 795 (m) cm⁻¹

**Mp** 223-225 °C (lit. 218-220 °C)₁⁶⁵

**Rf:** 0.25 (iPrOH:H2O:NH4OH 11:2:7)

**HRMS** (ES) for C32H39N₄O₂⁺ (M⁺): calcd 511.30675, found 511.30728

**UV/vis** (MeOH) λₓₑₐ = 561, λₑₘ = 581 nm
To a stirred solution of 42 (450 mg, 0.88 mmol), succinic anhydride (90 mg, 0.88 mmol) and DMAP (110 mg, 0.88 mmol) in CH₂Cl₂, triethylamine was added. After stirring at room temperature for 24 h, the reaction mixture was portioned between AcOEt and 1 M aqueous K₂CO₃. The aqueous layer was washed with additional portions of AcOEt. NaCl was added to the isolated aqueous layer until saturation was achieved and the solution was extracted with ¹PrOH: CH₂Cl₂ 2:1. The organic layer was dried over Na₂SO₄ and concentrated under reduced pressure. The resulting solid was dissolved in CH₃Cl₃ and filtered to remove insoluble salts. Upon concentration, the product 43 was obtained as a dark-red solid, 526 mg (98%).

**1H NMR** (360 MHz, CD₃OD): δ 7.78-7.76 (2H, m, Ar-H₃', Ar-H₅'), 7.71-7.69 (1H, m, Ar-H₆'), 7.53-7.50 (1H, m, Ar-H₄'), 7.28 (2H, d, J= 9.5 Hz, Ar-H₁, Ar-H₈), 7.08 (2H, d, J= 9.6, Ar-H₂, Ar-H₇), 7.00 (2H, d, J= 2.5 Hz, Ar-H₄, Ar-H₅), 3.69 (8H, q, J= 7.2 Hz, CH₂-Et), 3.44 (4H, br s, CH₂-2, CH₂-6), 3.38 (4H, br s, CH₂-3, CH₂-5), 2.56 (2H, t, J= 7.3 Hz, CH₂-α), 2.40 (2H, t, J= 7.1 Hz, CH₂-β), 1.30 (12H, t, J= 7.1 Hz, CH₃-Et).

**13C NMR** (90 MHz, CD₃OD): δ 180.2 (COOH), 173.5 (C=O), 169.2 (C=O), 158.9 (C-Ar), 158.8 (C-Ar), 136.2 (C-Ar), 132.9 (Ar-H₁, Ar-H₈), 131.9 (Ar-H₄'), 131.4 (Ar-H₃', Ar-H₅'), 130.9 (Ar-H₆'), 128.6 (C-Ar), 115.1 (Ar-H₂, Ar-H₇), 114.5 (C-Ar), 97.0 (Ar-H₄, Ar-H₅), 46.6 (CH₂-Et), 42.8 (CH₂-2, CH₂-6), 42.5 (CH₂-3, CH₂-5), 33.4 (CH₂, CH₂-α), 30.2 (CH₂-β), 12.6 (CH₃-Et).

**HPLC** (λ₅₆₅) tᵣ 6.76 min (method 1).

**ES⁺ m/z (%)** 611.7 (M⁺, 100).

**FTIR** 2977 (w), 1631 (s), 1588 (s), 1446 (s), 1411 (s), 1336 (s), 1274 (s), 1246 (s), 1222 (s), 1178 (s), 1131 (s), 1075 (m), 1005 (m), 976 (m), 820 (m), 750 (m) cm⁻¹
Mp 167-169 °C (lit. 166-168 °C)\textsuperscript{165}

$R_f$: 0.52 (DCM:MeOH 9:1)

HRMS (ES) for $C_{36}H_{43}N_4O_5^+$: calcd 611.32280, found 611.32208

UV/vis (MeOH) $\lambda_{ex} = 560$, $\lambda_{em} = 580$ nm

Succinimidyle ester of (9-[(2-[4-(3-carboxy-propionyl)-piperazine-1-carbonyl]-phenyl]-6-diethylamino-xanthen-3-ylidene)-diethyl-ammonium; chloride

![Chemical Structure](image)

Dye 43 (116 mg, 0.18 mmol) was dissolved in a mixture of dry DMF (1 mL) and dry pyridine (1 mL). Disuccinimidyl carbonate (DSC) (69 mg, 0.27 mmol) was added, and the mixture was stirred at 60 °C for 1h, under a nitrogen atmosphere. After dilution of the mixture with dry diethyl ether, the suspension was centrifuged and the supernatant decanted. The product was washed repeatedly with diethyl ether, dried \textit{in vacuo} and used for the next step without further purification.

HPLC ($\lambda_{550}$) $t_R$ 5.07 min

MS (ES): $m/z$ (%): 708.3 [M$^+$, 100], 709.3 [(M+1)$^+$, 40], 709.3 [(M+2)$^+$, 10]

$R_f$: 0.68 (DCM:MeOH 9:1)
Synthesis of a Rho B Labelled Peptide: Rho -(Ala)₃-OH

The active ester 44 (134 mg, 0.18 mmol) was dissolved in DMF (2 mL) and allowed to react with the solid-phase bound peptide PS-Ala-Ala-Ala-H (100 mg, 0.056 mmol), under agitation, for 3 h. The reaction was monitored by qualitative ninhydrin test and repeated if required. After completion of the labelling reaction, the resin was washed several times with DMF and DCM to remove the excess of the unbound dye. The labelled peptide was then cleaved from the resin with TFA/water (95:5) for 30 min; the cleavage was repeated twice. The labelled peptide was precipitated with cold diethyl ether, centrifuged and dried in vacuo to give 43 mg (93%) of a blue solid which was used for the next step without further purification.

**¹H NMR** (360 MHz, D₂O): δ 7.83-7.72 (m, 2H, Ar-H₃', Ar-H₅'), 7.58 (d, 1H, J = 7.7 Hz, Ar-H₆'), 7.56-7.50 (m, 1H, Ar-H₄'), 7.09-7.04 (m, 2H, Ar-H₁, Ar-H₈), 6.87-6.80 (m, 2H, Ar-H₂, Ar-H₇), 6.65-6.64 (m, 2H, Ar-H₄, Ar-H₅), 4.25 (q, 1H, J = 7.3 Hz, CH-Ala), 4.20 (q, 1H, J = 7.1 Hz, CH-Ala), 4.12 (q, 1H, J = 7.1 Hz, CH-Ala), 3.44-3.55 (m, 8H, CH₂-Et), 3.23 (br s, 2H, CH₂-2, CH₂-6), 3.23 (br s, 2H, CH₂-3, CH₂-5), 2.57-2.40 (m, 4H, CH₂-α), 1.28 (m, 9H, CH₃-Ala), 1.14 (t, 12H, J = 6.9 Hz, CH₃-Et).

**¹³C NMR** (90 MHz, D₂O): δ 176.0 (COOH), 175.3 (C=O), 173.7, 173.6, 173.4 (CONH), 170.8 (C=O), 157.9 (C-Ar), 156.2 (C-Ar), 154.3 (C-Ar), 135.0 (C-Ar), 133.3 (Ar-H₁, Ar-H₈), 132.3 (Ar-H₄'), 131.6 (Ar-H₃', Ar-H₅'), 131.1 (Ar-H₆'), 128.9 (C-Ar), 115.0 (Ar-H₂, Ar-H₇), 113.4 (C-Ar), 97.2 (Ar-H₄, Ar-H₅), 50.9, 50.0, 49.4 (3 × CH-Ala), 46.8 (CH₂-Et), 42.8, 42.6, 42.4, 41.5 (CH₂-2, CH₂-3, CH₂-5, CH₂-6), 30.8 (CH₂, CH₂-α), 28.9 (CH₂-β), 17.3, 17.2, 17.1 (3 × CH₃-Ala), 12.9 (CH₃-Et).
HPLC ($\lambda_{565}$) $t_R$ 7.54 min  
ES$^+$ m/z (%) 824.9 (M$^+$, 100), 413.4 (M$^+$/2, 47)  
FTIR 3291 (w), 2977 (w), 1644 (s), 1587 (s), 1465 (s), 1411 (s), 1336 (s), 1274 (s), 1246 (s), 1177 (s), 1128 (s), 1072 (m), 1006 (m), 921 (m), 822 (m), 796 (m) cm$^{-1}$  
Mp 113-115 °C.  
Rf: 0.30 (DCM:MeOH 9:1)  
UV/vis (MeOH) $\lambda_{ex}= 560$, $\lambda_{em}= 581$ nm  
HRMS (ES) for C$_{45}$H$_{58}$N$_7$O$_{10}$ ($M^+$): calcd 824.43414, found 824.43470

**Synthesis of dUTP-(Ala)$_3$-Rho**

A solution of N, N-disuccinimidyl carbonate (3.6 mg, 11.0 µmol) in DMF (50 µL) was added to a solution of 45 (8.0 mg, 9.3 µmol) in a mixture of dry DMF (95 µL) and dry pyridine (5 µL). The reaction mixture was stirred under nitrogen at 50-60 °C for 1-2 h and then diluted with dry ethyl acetate (2 mL). After centrifugation, the supernatant was decanted and the precipitate washed repeatedly with dry ethyl acetate. To the resulting blue solid, dissolved in 100 µL of 0.1M TEAB buffer (pH 8.5) were added 100 µL of 5-(3-aminopropargyl)-2'-deoxyuridine 5'-triphosphate (3.0 mg, 3.2 µmol) in the same buffer solution. The reaction mixture was stirred at room temperature for 30 min- 1 h and then purified by reverse-phase HPLC to afford the nucleotide 46 dUTP-(Ala)$_3$-Rho in 1.5% yield (50 nmol). The purified fraction was freeze-dried and the lyophilised red solid obtained characterised by MALDI-TOF analysis and stored at −20 °C.  
MALDI-TOF m/z: anal.calcd, 1327; found 1326 (M-H)$^+$  
HPLC ($\lambda_{565}$) $t_R$ = 23.9 min  
UV/vis (1 × TE buffer) $\lambda_{max}= 567$ nm
5(6)-Carboxyfluorescein (100 mg, 0.27 mmol) in DMF (3 mL) was treated with PyBOP (211 mg, 0.41 mmol) for 2 min, then 6-aminohexanoic acid methyl ester hydrochloride (95 mg, 0.52 mmol) and DIEA (185 µL, 1.08 mmol) were added and the reaction mixture heated at 60 °C for 20 min under microwave irradiation. The reaction mixture was evaporated under reduced pressure, dissolved in CH₂Cl₂ (3 mL) and treated with 2M HCl (3 mL). The brown precipitate was dissolved in (2:1) PrOH/CH₂Cl₂ (3 mL) and washed with brine (3 mL × 3), The organic layer was dried over Na₂SO₄, and concentrated under reduced pressure. The resulting solid was dissolved in water/THF (2 mL) and stirred with 2M LiOH (2 mL) for 2 h. The reaction mixture was evaporated under reduced pressure, dissolved in H₂O (2 mL) and treated with 2 M HCl. The precipitate was then washed with water, filtered and freeze-dried. Product 47 was obtained as orange-brown solid 114 mg (85%).

¹H NMR (360 MHz, CD₃OD): δH 8.41 (s, 1H, H-7, isomer 6), 8.19 (dd, 1H, J = 8.0, 1.3 Hz, H-5, isomer 6), 8.12 (dd, 1H, J = 8.1, 0.9 Hz, H-6, isomer 5), 8.07 (d, 1H, J = 8.2 Hz, H-7, isomer 5), 7.60 (s, 1H, H-4, isomer 5), 7.29 (d, 1H, J = 7.9 Hz, H-4, isomer 6), 6.69 (d, 2H, J = 1.8 Hz, H-α, isomer 5 & 6), 6.61 (d, 1H, J = 8.7 Hz, H-μ, isomer 5), 6.60 (d, 1H, J = 8.7 Hz, H-μ, isomer 6), 6.54 (td, 2H, J = 8.7, 2.7 Hz, H-ϕ, isomer 5 & 6), 3.44 (t, 2H, J = 7.1 Hz, CH₂-e, isomer 6), 3.33 (t, 2H, J = 6.9 Hz, CH₂-e, signal partially obscured by CD₃OD peak, isomer 5), 2.33 (t, 2H, J = 7.3 Hz, CH₂-α, isomer 6), 2.26 (t, 2H, J = 7.3 Hz, CH₂-α, isomer 5), 1.68 (quintet, 4H J = 7.0 Hz, CH₂-β, CH₂-δ, isomer 6), 1.58 (quintet, 4H J = 7.7 Hz, CH₂-β, CH₂-δ, isomer 5),
Hz, CH₂-β, CH₂-δ, isomer 5), 1.46 (quintet, 2H, J = 6.9 Hz, CH₂-γ, isomer 6), 1.35 (quintet, 2H J = 7.2 Hz, CH₂-γ, isomer 5).

¹³C NMR (90 MHz, DMSO): δC 176.0 (COOH), 169.7 (C=O), 169.6 (C=O), 166.0 (C=O), 165.8 (C=O), 161.2 (C-Ar), 156.0 (C-Ar), 154.0 (C-Ar), 153.3 (C-Ar), 142.2 (C-Ar), 137.9 (C-Ar), 136.1 (CH, C-5, isomer 6), 130.7 (CH, C-6, isomer 5), 130.6 (CH, m), 129.7 (C-Ar), 128.0 (C-Ar), 126.4 (CH, C-7, isomer 5), 125.7 (CH, C-4, isomer 6), 124.7 (CH, C-7, isomer 6), 123.8 (CH, C-4, isomer 5), 114.3 (CH, p), 114.2 (CH, p), 110.7 (C-Ar), 110.6 (C-Ar), 103.8 (CH, o), 83.4 (C), 40.7 (CH₂-ε), 35.2 (CH₂-α), 35.1 (CH₂-α), 30.2 (CH₂-δ), 30.1 (CH₂-δ), 27.5 (CH₂-β), 25.8 (CH₂-γ), 25.7 (CH₂-γ).

FTIR 2930 (w), 1710 (m), 1585 (s), 1537 (s), 1454 (m), 1265 (s), 1239 (s), 1208 (s), 1112 (s), 849 (s).

HPLC (λ₄40) tR 5.05 min

ES⁻ m/z (%) 488.2 (M⁻⁻⁻)

Mp 180-185 °C

HRMS (ES) for C₂₇H₂₄N₁O₈⁺ (M⁺): calcd 490.14964, found 490.14961

Rf: 0.20 (DCM:MeOH 4:1)

UV/vis (MeOH) λₑₓ= 492, λₑₘ= 517 nm
Succinimidyl ester of \((4\text{S}-\{(5\text{-carboxypentyl})\text{amino}\}\text{carbonyl}\)\)-2-(3,6-dihydroxy-9H-xanthen-9-yl)benzoic acid

\[
\text{HO'OOH}
\]

Dye 47 (118 mg, 0.24 mmol) was dissolved in a mixture of dry DMF (1 mL) and dry pyridine (1 mL), 200 µL of DMSO were added to aid the solubilisation of 47. Disuccinimidyl carbonate (DSC) (123 mg, 0.48 mmol) was added, and the mixture was stirred at 60 °C for 1h, under a nitrogen atmosphere. After dilution of the mixture with dry diethyl ether, the suspension was centrifuged and the supernatant decanted. The product was washed repeatedly with diethyl ether and dried in vacuo to afford a yellow solid.

\textbf{MS (ES): } m/z (\%) 585.1 [(M-1)], 100, 586.1 [M, 35].

\textbf{Rf: } 0.73 (DCM:MeOH 4:1)
Synthesis of a Carboxyfluorescein Labelled Peptide: FAM-(Ala)₃-OH

The active ester 48 (106 mg, 0.18 mmol) was dissolved in DMF (2 mL) and allowed to react with the solid-phase bound peptide PS-Ala-Ala-Ala-H (100 mg, 0.056 mmol), under agitation, for 3 h. The reaction was monitored by qualitative ninhydrin test and repeated if required. After completion of the labelling reaction, the resin was washed several times with DMF (4 mL), 20 % piperidine in DMF (4 mL), DMF (4 mL), and DCM (4 mL) to remove the excess of the unbound dye. The labelled peptide was then cleaved from the resin with (95:5) TFA/water (8 mL) for 30 min; the cleavage was repeated twice. The labelled peptide was precipitated with cold diethyl ether, centrifuged and dried *in vacuo* to give 43 mg (93%) of a yellow solid which was used for the next step without further purification.

¹H NMR (500 MHz, CD₃OD): δH 8.82 (t, 1H, J = 5.6 Hz, NH, isomer 6), 8.69 (t, 1H, J = 5.5 Hz, NH, isomer 5), 8.45 (s, 1H, H-7, isomer 6), 8.22 (dd, 1H, J = 8.0, 1.6 Hz, H-5, isomer 6), 8.16 (dd, 1H, J = 8.1, 1.4 Hz, H-6, isomer 5), 8.11 (d, 1H, J = 7.9 Hz, H-7, isomer 5), 7.63 (s, 1H, H-4, isomer 5), 7.33 (d, 1H, J = 8.0 Hz, H-4, isomer 6), 6.73 (d, 2H, J = 1.7 Hz, H-o, isomer 5 & 6), 6.65 (d, 1H, J = 8.2 Hz, H-m, isomer 5), 6.64 (d, 1H, J = 8.1 Hz, H-m, isomer 6), 6.60-6.57 (m, 2H, H-p, isomer 5 & 6), 4.38 (q, 1H, J = 7.1 Hz, CH-Ala), 4.35 (q, 1H, J = 7.2 Hz, CH-Ala), 4.31 (q, 1H, J = 7.2 Hz, CH-Ala), 3.47 (t, 2H, J = 7.0 Hz, CH₂-e, isomer 6), 3.33 (t, 2H, J = 6.9 Hz, CH₂-e, isomer 5, signal obscured by CD₃OD peak), 2.30 (t, 2H, J =
7.3 Hz, CH2-α, isomer 6), 2.23 (t, 2H, \( J = 7.6 \) Hz, CH2-α, isomer 5), 1.75-1.67 (m, 4H, CH2-β, CH2-δ, isomer 6), 1.60 (ddd, 4H \( J = 7.2 \) Hz, CH2-β, CH2-δ, isomer 5), 1.47 (quintet, 2H, \( J = 7.6 \) Hz, CH2-γ, isomer 6), 1.42-1.32 (m, 11H, CH2-γ, isomer 5, 3 x CH3-Ala).

\(^{13}\text{C} \text{ NMR} \) (126 MHz, CD3OD): δC 175.7 (COOH), 175.3 (C=O), 174.5 (C=O), 174.1 (C=O), 170.1 (C=O), 167.9 (C=O), 167.7 (C=O), 161.4 (C-Ar), 153.9 (C-Ar), 142.0 (C-Ar), 137.7 (C-Ar), 135.0 (CH, C-5, isomer 6), 130.0 (CH, \( m \)), 129.9 (CH, C-6, isomer 5), 128.4 (C-Ar), 126.0 (CH, C-7, isomer 5), 125.5 (CH, C-4, isomer 6), 124.6 (CH, C-7, isomer 6), 123.8 (CH, C-4, isomer 5), 113.6 (CH, \( p \)), 110.7 (C-Ar), 103.2 (CH, \( o \)), 50.0, 49.9, 49.7 (3 x CH-Ala), 40.7 (CH2-ε), signals partially obscured by CD3OD peak, 45.5 (C), 45.4 (C), 36.2, 36.1 (CH2-α, isomer 5 and 6), 29.7, 29.6 (CH2-δ, isomer 5 and 6), 27.2, 27.1 (CH2-β, isomer 5 and 6), 26.4, 26.0 (CH2-γ, isomer 5 and 6), 17.8, 17.6, 17.5, 17.2, 17.1 (3 x CH3-Ala, isomer 5 and 6).

**HPLC** \( (λ_{440}) \) \( t_R \) 7.07

**ES** \( m/z \) (%) 701.4 (\((\text{M}-\text{H})^-\), 100), 723.4 (\((\text{M}-2\text{H}^+ \text{ Na})^-\), 47), 350.8 (\((\text{M}-\text{H})^2/2\), 37),

**Rf**: 0.77 (DCM:MeOH 4:1)

**UV/vis** (MeOH) \( \lambda_{\text{ex}} = 492 \), \( \lambda_{\text{em}} = 517 \) nm

**IR** (neat) 3467, 2928, 2861, 2361, 2342, 1737, 1655, 1407, 1254, 1062, 864, 819 cm\(^{-1}\).

**HRMS** (ES) for \( \text{C}_{36}\text{H}_{39}\text{N}_4\text{O}_{11}^+ \) (M\(^+\)): calcd 703.26098, found 703.25936

204
Synthesis of dGTP-(Ala)_3-FAM

A solution of N, N-disuccinimidyl carbonate (3.6 mg, 11.0 μmol) in DMF (50 μL) was added to a solution of 49 (8.0 mg, 9.3 μmol) in a mixture of dry DMF (95 μL) and dry pyridine (5 μL). The reaction mixture was stirred under nitrogen at 50-60 °C for 1-2 h and then diluted with dry ethyl acetate (2 mL). After centrifugation, the supernatant was decanted and the precipitate washed repeatedly with dry ethyl acetate. To the resulting yellow solid, dissolved in 100 μL of 0.1M TEAB buffer (pH 8.5) were added 100 μL of 8-(3-aminopropargyl)-2'-deoxyguanosine 5'-triphosphate (4.0 mg, 4.1 μmol) in the same buffer solution. The reaction mixture was stirred at room temperature for 30 min-1h and then purified by reverse-phase HPLC to afford the nucleotide 50 dGTP-(Ala)_3-FAM. The purified fraction was freeze-dried and the lyophilised yellow solid obtained characterised by MALDI-TOF analysis and stored at −20 °C.

MALDI-TOF m/z: anal.calcd, 1244; found, 1258 (M-4H+N1-L1)

HPLC (λ_{d1}) {t}_R = 24.3 and 25.5 min

UV/vis (1 × TE buffer) λ_{max} = 498 nm
Synthesis of dCTP-(Ala)₃-FAM

A solution of N, N-disuccinimidyl carbonate (3.6 mg, 11.0 µmol) in DMF (50 µL) was added to a solution of 50 (5.0 mg, 5.8 µmol) in a mixture of dry DMF (95 µL) and dry pyridine (5 µL). The reaction mixture was stirred under nitrogen at 50-60 °C for 1-2 h and then diluted with dry ethyl acetate (2 mL). After centrifugation, the supernatant was decanted and the precipitate washed repeatedly with dry ethyl acetate. To the resulting yellow solid, dissolved in 100 µL of 0.1M TEAB buffer (pH 8.5) were added 100 µL of 5-(3-aminopropargyl)-2'-deoxycytidine 5'-triphosphate (2.0 mg, 2.2 µmol) in the same buffer solution. The reaction mixture was stirred at room temperature for 30 min-1h and then purified by reverse-phase HPLC to afford the nucleotide 51 dCTP-(Ala)₃-FAM. The purified fraction was freeze-dried and the lyophilised yellow solid obtained characterised by MALDI-TOF analysis and stored at —20 °C.

MALDI-TOF m/z: anal.calcd, 1204; found 1202 (M-2H)⁺, 1204 (M)⁺, 1224 (M-3H+Na)⁺.

HPLC (λ₄₀₁) tᵣ = 22.7 and 25.4 min
UV/vis (1 x TE buffer) λₘₐₓ = 498 nm
6.3 Experimental for Chapter 4

Materials and Methods

RP (reverse phase) purified synthetic oligonucleotides were purchased from Sigma Genosys Ltd. (UK). The lyophilised DNA samples (50 nmol) were reconstituted using 1×TE buffer (10 mM Tris, pH 7.5-8.0, 1mM EDTA) and stored at -20 °C. Fluorescently labelled ddNTPs were obtained from Perkin-Elmer Life Science (Boston, MA). Thermo Sequenase DNA polymerase supplied with 10× reaction buffer was purchased from GE Healthcare (Chalfont St. Giles, UK). Subtilisin Carlsberg was obtained from Sigma Aldrich, UK. ZipTipsC18 were purchased from Millipore (Bedford, USA). All other chemicals were purchased from Sigma Aldrich. Primer extensions and protease cleavages were performed on a Eppendorf MasterCycler Personal thermal cycler with a heated lid.

HPLC Analysis

RP-HPLC was performed on a Hewlett Packard HP1100 Chemstation eluting with (A) 50% Acetonitrile in 0.1 M TEAB, (B) 0.1 M TEAB and (C) Acetonitrile on a Luna (Phenomenex) C18 ODS (2) analytical column, 100 mm × 4.60 mm, i.d. 5 μm, flow rate 1.0 mL min⁻¹, using the method described below. The fractions were collected using a Gilson G2250A.

Method: gradient: 100 % (A) to 0 % (A) over 15 min, 50 % (C) to 100 % (C) over 1 min, constant at 100 % (B) for 6.5 min, injection volume 100 μL, detection by UV at 565, 260, 254, 220 nm, 641 nm.

MALDI-TOF Analysis

Oligonucleotide MS spectra were obtained by matrix-assisted laser desorption ionization time-of-flight (MALDI) on an Applied Biosystems Voyager DE STR MALDI-TOF mass spectrometer (MS) equipped with a pulsed nitrogen laser (20 Hz, 337 nm). A total of 400 shots were summed for each spectrum with an acceleration voltage of 20 kV.

Preparation of the matrix

9 parts of 5 mg/100 μL of 3-hydropicolinic acid in 50% acetonitrile/water were mixed with 1 part 5 mg/100 μL of ammonium citrate in water.
MALDI Plate Spotting

Oligonucleotides were desalted using ZipTipsc_{18} according to the manufacturer’s protocol and dissolved in 10 μL of a solution of 10% acetonitrile/0.1 M TEAA buffer. 1 μL of the oligonucleotide solution was mixed with 1 μL of the matrix solution directly on the stainless steel MALDI sample target plate.

Primer extension using dUTP-(Ala)_{3}-Rho 46

The polymerase extension reaction mixture consisted of 200 pmol of primer I, 320 pmol of p53-exon7 template I, 640 pmol of dUTP-(Ala)_{3}-Rho 46, 16 units of Thermo Sequenase DNA Polymerase and 1 x Thermo Sequenase reaction buffer in a total volume of 50 μL. The reaction consisted of 25 cycles at 94 °C for 20 s, 46 °C for 40 s and 60 °C for 90 s. After reaction, the extension product II was purified by RP-HPLC and freeze-dried. A portion of the desired DNA product was desalted by using ZipTipsc_{18} and analyzed by MALDI-TOF MS. The rest of the product was used for the cleavage step.

MALDI-TOF MS (linear, negative) m/z: anal.calcd, 6674; found, 6673 (M-1).

HPLC (λ_{565, 260}) t_R 7.21 min

Primer I sequence:
5' - AGAGGATCCAACCGAGAC-3'

p53-exon7 Template I sequence:
5' - GTGTACATCAACATCACCTACCACCATGTCAGTCTCGGTTGGATCCTCTATTGTGTCCGG-3'

Primer extension using dCTP-(Ala)_{3}-Cy5 41

The polymerase extension reaction mixture consisted of 200 pmol of primer I, 320 pmol of p53-exon7 template II, 640 pmol of dCTP-(Ala)_{3}-Cy5 41, 16 units of Thermo Sequenase DNA Polymerase and 1 x Thermo Sequenase reaction buffer in a total volume of 50 μL. The reaction consisted of 25 cycles at 94 °C for 20 s, 46 °C for 40 s and 60 °C for 90 s. After reaction, the extension product IV was purified by RP-HPLC and freeze-dried.

HPLC (λ_{641, 260}) t_R 9.29 min

Primer I sequence:
5' - AGAGGATCCAACCGAGAC-3'

p53-exon7 Template II sequence:
5' - GTGTACATCAACATCACCTACCACCATGTCGGTCTCGGTTGGATCCTCTATTGTGTCCGG-3'
Primer extension using dUTP-(Ala)3-Rho on a homopolymeric sequence

The polymerase extension reaction mixture consisted of 200 pmol of primer I, 320 pmol of p53-exon7 template IV, 640 pmol of dUTP-(Ala)3-Rho 46, 16 units of Thermo Sequenase DNA Polymerase and 1 x Thermo Sequenase reaction buffer in a total volume of 50 µL. The reaction consisted of 25 cycles at 94 °C for 20 s, 46 °C for 40 s and 60 °C for 90 s. After reaction, the extension product was purified by RP-HPLC and freeze-dried.

HPLC (λ641, 260) tR 7.21 min

Primer sequence I:

5' - AGAGGATCCAACCGAGAC-3'

ST p53-exon7 Template IV sequence:

5' - GTGTACATCAACATCACCTACCACCATCAAAGTCTCGGTTGGATCCTCTATTGTGTCCGG-3'

Primer extension using dUTP-(Ala)3-Rho and dCTP-(Ala)3-Cy5

The polymerase extension reaction mixture consisted of 200 pmol of primer I, 320 pmol of p53-exon7 template V, 640 pmol of dCTP-(Ala)3-Cy5 41, 640 pmol of dUTP-(Ala)3-Rho 46, 16 units of Thermo Sequenase DNA Polymerase and 1 x Thermo Sequenase reaction buffer in a total volume of 50 µL. The reaction consisted of 25 cycles at 94 °C for 20 s, 46 °C for 40 s and 60 °C for 90 s. After reaction, the extension product was purified by RP-HPLC and freeze-dried.

HPLC major peak (λ565, 260) tR 6.70 min

HPLC small peak (λ641, 260) tR 7.40 min

Note: all the peaks (template and product extensions) were shifted of 0.5 min.

Primer sequence I:

5' - AGAGGATCCAACCGAGAC-3'

p53-exon7 Template (AG) V sequence:

5' - GTGTACATCAACATCACCTGAGAGAGAGGTCTCGGTTGGATCCTCTATTGTGTCCGG-3'

Cleavage by protease of the extension product

The extended primer II was dissolved in 7 µL of a solution (2:5) of 0.1 M TEAB buffer/ H2O. 15 units of Subtilisin A were dissolved in 10 µL of Subtilisin A buffer (10 mM Sodium Acetate buffer with 5 mM Calcium Acetate, pH 7.5), added to the solution of the extended primer and left in incubation at 37 °C for 10 min. The resulting DNA product III was purified by RP-HPLC and freeze-dried. A portion
of the desired DNA product was desalted using ZipTips and analyzed by MALDI-TOF MS.

MALDI-TOF MS (linear, negative) m/z: anal.calcd, 5866; found, 5865 (M-1), 5936 (M+71), 6007 (M+142).

HPLC (λ565, 260) tR 6.3 min

Primer after cleavage (III):

3′-UCAGAGCCAACCTAGGAGA-5′

6.4 Experimental for Chapter 5

Materials and Instruments

RP (reverse phase) purified synthetic oligonucleotides, 50 nmol scale, were purchased from Sigma Genosys Ltd. (UK). The 5’ end of DNA primers was amino modified. The lyophilised DNA samples were reconstituted using 1×TE buffer (10 mM Tris, pH7.5-8.0, 1mM EDTA) and stored at -20 °C. Fluorescently labelled ddNTPs were obtained from Perkin-Elmer Life Science (Boston, MA). GenHyb hybridisation buffer and aldehyde slides were purchased from Genetix Ltd. (New Milton, UK). CodeLink™ glass slides (25 mm x 75 mm x 1mm) and Thermo Sequenase DNA polymerase supplied with 10× reaction buffer were purchased from GE Healthcare (Chalfont St. Giles, UK). Subtilisin Carlsberg was obtained from Sigma Aldrich, UK. 20× Standard saline citrate (SSC) buffer was obtained from Fisher Scientific UK (Loughborough, UK). Other buffers and solutions for microarray experiments were prepared following standard recipes using DNA grade water (Fisher Scientific UK). Hybridisation and primer extension reactions were performed on a Hyb4 hybridisation station from Genomic Solutions (Huntingdon, UK). DNA printing was performed using a Robot Microarrayer (Genetix QMini, U.K.).

Design of Primers & Templates

DNA primers and templates were designed with the Oligo Analyzer v. 3.1 program.

Signal Detection and Data Analysis

All slides were scanned using a LaVision BioTech Bioanalyzer 4F/4S Scanner featuring a white-light source, four color filters and a CCD camera. The following
filters were used (excitation wavelength ± bandwidth of transmission/emission wavelength ± bandwidth of transmission): FITC: (480 ±40 / 535 ± 50); Cy3: (546 ±12 / 585 ± 40); Texas Red: (560 ± 55 / 645 ± 75); Cy 5: (640 ± 20 / 680 ± 30). The filters were from Chroma Technology Corporation (Rockingham, USA).

The intensity of the fluorescence signals was determined with the FIPS 2.0 analysis software supplied with the chip reader and data files were imported in Microsoft Excel for further analysis.

6.4.1 Preparation of DNA Microarrays

Printing on CodeLink™ slides

5'-amino modified DNA primers were diluted to a final concentration of 20 µM in 1× printing buffer (50 mM sodium phosphate, pH 8.5). A 5'-amino modified DNA primer labelled in 3' position with fluorescein was used as control.

The oligonucleotides were printed onto CodeLink™ slides at 25 °C and 35-40% relative humidity by contact printing. Solid pins (150 μm diameter, K2783, Genetix) were used for printing with the following conditions: 1 stamp per spot, 10 ms stamping time and 10 ms inking time. The same pattern was reproduced twice: a 10×10 subarray for a 3'-amino modified and 5' fluorescein labelled oligonucleotide and three 10×10 subarrays for 3'-amino modified DNA oligonucleotides. The slides were kept overnight in a chamber containing a 3M NaCl solution. Residual NHS ester groups on the surface were blocked using pre-warmed blocking solution (0.1 M Tris, 50 mM ethanolamine, pH 9.0, 20 mL) at 50 °C for 30 min. Afterwards, the slides were rinsed with deionised water (2 × 20 mL), then with 4 × SSC, 0.1 % SDS (20 mL, pre-warmed to 50 °C) for 30 min and with deionised water (2 × 20 mL). The slides were spin-dried at 3000 rpm for 6 min and scanned using a FITC filter. DNA sequences of primers and their fluorescence intensities (average) are shown in Table 6.1.
Table 6.1 DNA sequences and correspondent fluorescence intensities ($\lambda_{em} = 480$ nm). Exposure time: 100 ms.

*Hybridisation on CodeLink™ slides*

To 50 $\mu$L of a 2 $\mu$M solution of oligonucleotide templates in 1 x TE buffer, were added 50 $\mu$L of 2 x GenHyb buffer (50% formamidine, 2 mg/mL herring sperm DNA, 0.4 % SDS, 8.4 x SSC). Prior to hybridisation the solution was heated to 85 °C for 3 minutes to denaturise the oligonucleotides. The solution was then added to the slide surface and incubated for 3 h at 55 °C in an automated hybridisation system (GenHyb4). The slide was rinsed briefly with 4 x SSC (20 mL), then with 2 x SSC, 0.1 % SDS at 55 °C for 5 min (2 x 20 mL), with 0.2 x SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using a FITC filter. An oligonucleotide labelled in the 5' position with fluorescein was used as hybridisation control. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.2.
Table 6.2 DNA sequences and correspondent fluorescence intensities ($\lambda_{em} = 480$ nm). Exposure time: 100 ms.

### 6.4.2 Four Color Primer Extension with dideoxynucleotides on CodeLink™ slides

CodeLink™ slides were spotted with four identical arrays of 102 amino 3’ amino-modified DNA primers and then treated with a solution of complementary DNA templates. Printing and hybridisation were performed as described in the previous sections 6.4.1. A reaction mix of 100 µL was prepared with a final concentration of 0.8 U/µL of Thermo Sequenase DNA Polymerase and 2 µM of labelled ddNTPs (Texas Red-ddATP, Fluorescein-ddCTP, Cy5-ddUTP and TAMRA-ddGTP) in 1× Thermo Sequenase reaction buffer (26 mM TrisHCl, pH 9.5 and 6.5 mM MgCl₂). The solution was added to the slide and incubated at 55 °C for 1h in an automated hybridisation chamber (GenHyb4). After incubation, the slide was rinsed briefly with 4×SSC (20 mL), then with 2× SSC, 0.1 % SDS (2×20 mL) at 55 °C for 5 min, with 0.2×SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at r.t for 1 min. The slide was spin-dried at 3000 rpm for 6 min and then scanned. Fluorescence intensities were measured for all DNA duplexes at four different wavelengths (Table 6.3).

<table>
<thead>
<tr>
<th>DNA sequences</th>
<th>Pattern 1</th>
<th>Pattern 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3’ 3’-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5’</td>
<td>574700</td>
<td>601924</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AG-3’ 3’-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5’</td>
<td>606584</td>
<td>574326</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT CAG GAC CTA GAA CGG GCA GC-3’ GTC CTG CAT TT TGC TGG CAC CCA TCA G-5’</td>
<td>1877</td>
<td>2044</td>
</tr>
<tr>
<td>5’-H₂N-TTT TTT CCT TCT TGC TGG CAC CCA AT 3’ GGA AGA ACG ACC GTG GGT TAA TCA G-5’</td>
<td>1834</td>
<td>1815</td>
</tr>
</tbody>
</table>
Table 6.3 DNA sequences and correspondent fluorescence intensities.

Fl = fluorescein, S = \textasciitilde\text{HN-TTT TTT-}

<table>
<thead>
<tr>
<th>Position</th>
<th>FITC</th>
<th>Cy3</th>
<th>Texas Red</th>
<th>Cy5</th>
<th>Primer/Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>390870250</td>
<td>22451000</td>
<td>24480751</td>
<td>2028250</td>
<td>5'-S-TAT GAC ACC GTC ATP AG-F1-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-ATA CTG TAG TGG TCA TCA G-F1-3'</td>
</tr>
<tr>
<td>B1</td>
<td>106500000</td>
<td>19690750</td>
<td>19450720</td>
<td>20895608</td>
<td>5'-S-TAT GAC ACC GTC ATP AG-F1-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-ATA CTG TAG TGG TCA TCA G-F1-3'</td>
</tr>
<tr>
<td>C1</td>
<td>16662000</td>
<td>88496252</td>
<td>34100000</td>
<td>1239500</td>
<td>5'-S-CAG CCC CGA TGG TCT CAC TC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GTC GGG GCT AGG AGA GTA GTC CTA GCA GTC</td>
</tr>
<tr>
<td>D1</td>
<td>24860500</td>
<td>45806249</td>
<td>15531750</td>
<td>15036000</td>
<td>5'-S-GCA AAG TGA TCC CCC AAA GAA C 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GAT TTC ACT AGG GGG TTT ATT GAC TCG TCC AC</td>
</tr>
<tr>
<td>E1</td>
<td>21789746</td>
<td>20929500</td>
<td>19690750</td>
<td>88596249</td>
<td>5'-S-Ctg TCC CGA GAG GAT ATG GT 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GAC AGG GGT CTC CTA CAG GAC TCG 3'</td>
</tr>
<tr>
<td>F1</td>
<td>21172500</td>
<td>23974499</td>
<td>18776750</td>
<td>31560000</td>
<td>5'-S-ACT CCT CCT CCG GCC GAA TCA C 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GCA ACT CCT TGG GCT GGA ATC CTA GTC 3'</td>
</tr>
<tr>
<td>G1</td>
<td>12280999</td>
<td>59803750</td>
<td>14376750</td>
<td>21085000</td>
<td>5'-S-TTC TGG TGC TCA ATP AGT GTC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GCC GGA ATC AGG GGG GCT CTA GTC 3'</td>
</tr>
<tr>
<td>H1</td>
<td>18799500</td>
<td>105816498</td>
<td>37190000</td>
<td>15050000</td>
<td>5'-S-Att TCT CCT CCT GCA TCC ATG 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-CTG ACC TCT CTT GCA GCG AGC TCA GTC</td>
</tr>
<tr>
<td>I1</td>
<td>35530000</td>
<td>17473750</td>
<td>38425000</td>
<td>42502750</td>
<td>5'-S-ATT ACA GGC AGG AAC CAC CA 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GTA TAT CGG YAC TAC TGG GTA TCA GTC TGT</td>
</tr>
<tr>
<td>J1</td>
<td>14108250</td>
<td>69390000</td>
<td>29518500</td>
<td>1229500</td>
<td>5'-S-CAG GCA TGA GCC ACC AGC CC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GTC GTG ACT CGG TGG TGC GGT GCG ACT GGT</td>
</tr>
<tr>
<td>K1</td>
<td>14267500</td>
<td>65266000</td>
<td>26318750</td>
<td>1496500</td>
<td>5'-S-AGG AGG GCC ACT TGT GAC CTA G 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GAC TTT TGG GTA CAT TGG GTA TCA GTC</td>
</tr>
<tr>
<td>L1</td>
<td>23869250</td>
<td>21955501</td>
<td>57817500</td>
<td>40306749</td>
<td>5'-S-AGG CGG ACT TGT GCT GCA ACA CCC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GCC TGA TAA AGC GGT TTT ATT GAC GTC</td>
</tr>
<tr>
<td>M1</td>
<td>28501500</td>
<td>35119251</td>
<td>62275000</td>
<td>42579750</td>
<td>5'-S-GCC TGG AAA GGC AGC TGCG 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GGG TCT GGC AGC TCG ACA AGC TAC GTC</td>
</tr>
<tr>
<td>N1</td>
<td>18531000</td>
<td>85151752</td>
<td>41517750</td>
<td>1776750</td>
<td>5'-S-AGA CCC CAC CTC TCC GAA CT 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-CTG CGG GAG AGC CAT CCT GCA GTC</td>
</tr>
<tr>
<td>O1</td>
<td>27832750</td>
<td>29480250</td>
<td>60790000</td>
<td>51081751</td>
<td>5'-S-CTA TCT CAG TTT TGG CTT TAA TGA 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GTG GCA TGG GAG GGC CAT CCT GCA GTC</td>
</tr>
<tr>
<td>P1</td>
<td>19959250</td>
<td>97416250</td>
<td>37450751</td>
<td>1683500</td>
<td>5'-S-AGC TGC CAC TCT CCT GGG GCA TGG 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-AGC CAT GGT CAT GGA GAT TGG GTA TCA GTC</td>
</tr>
<tr>
<td>A2</td>
<td>36502500</td>
<td>36669750</td>
<td>55065000</td>
<td>78877000</td>
<td>5'-S-TTT CTG CCA TGG CAC TTT GTC CTA GTC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GAC TTT TGG GTA CAT TGG GTA TCA GTC</td>
</tr>
<tr>
<td>B2</td>
<td>13579500</td>
<td>14836250</td>
<td>94825000</td>
<td>13736250</td>
<td>5'-S-ACG TCG AAG ACT CAT GGA AAG AGT G 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GAT TGG ACA TGG GCA AGT TGG GTA TCA GTC</td>
</tr>
<tr>
<td>C2</td>
<td>21645250</td>
<td>20889000</td>
<td>55777500</td>
<td>38997750</td>
<td>5'-S-TCT TGG TGC CCA AGC ATT TCT TCA 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GAG TAT ATT TCT GAC GGA ATG GAC TGG</td>
</tr>
<tr>
<td>D2</td>
<td>14520249.75</td>
<td>47499500</td>
<td>27374501</td>
<td>1125500</td>
<td>5'-S-TGA TCG AAA AAG TCA TGG GGC TCC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-GTC CAT GGC CAC TTT AGT AGG TAG TTA 3'</td>
</tr>
<tr>
<td>E2</td>
<td>21224500</td>
<td>23524000</td>
<td>49635000</td>
<td>47880999</td>
<td>5'-S-CTG CGA TCT CIC TCT TAC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-AGC GGT AGG GAG GAT GGC AGA TAT GAC TCG TAC</td>
</tr>
<tr>
<td>F2</td>
<td>16075000</td>
<td>13202250</td>
<td>29855000</td>
<td>52255500</td>
<td>5'-S-TCC ATA TAA AGC CAT ACC TCA 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-AGG TAT TCT TAC TGG GTA GTC TCA GGC</td>
</tr>
<tr>
<td>G2</td>
<td>297693750</td>
<td>36987750</td>
<td>23818750</td>
<td>46558000</td>
<td>5'-S-CTG CCA TCC TCT GTC ATG 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-AGA ACT CAT TAT GAC GTC GAC ATC TAC G</td>
</tr>
<tr>
<td>H2</td>
<td>31690750</td>
<td>32912750</td>
<td>62310000</td>
<td>50904251</td>
<td>5'-S-GCC TCC AGT TGG CTA AGG 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-CCC AGG ATG CAT GTC GA TCA GTC TCT</td>
</tr>
<tr>
<td>I2</td>
<td>31152750.5</td>
<td>131383998</td>
<td>42914500</td>
<td>5245000</td>
<td>5'-S-GGT TTT TGG AGA TCA CTA GTC 3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-CCA AAA ACT AAA CGT AGT CTA GCA GTC 3'</td>
</tr>
<tr>
<td>J2</td>
<td>22125250</td>
<td>27158999</td>
<td>61762500</td>
<td>50680000</td>
<td>5'-S-GAG TCC AGA GGA GTA TAA TGG G 3'</td>
</tr>
<tr>
<td>K2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3'-CTC CAG TCT CTC TCT TAT TAT AAT CAT CAT CAT CAT 3'</td>
</tr>
</tbody>
</table>

214
6.4.3 Primer Extension using Enzymatically Cleavable Nucleotides on CodeLink™ slides

**Primer Extension with dUTP-(A)$_3$-Rho**

CodeLink™ slides with immobilised DNA duplexes were prepared as described in the sections 6.4.1. Four 3'-amino modified DNA primers were printed in four 6×6 subarrays. 100 µL of a solution consisting dUTP-(A)$_3$-Rho (2.5 µM), and Thermo Sequenase DNA polymerase (1 U/µL) in 1 × Thermo Sequenase reaction buffer were added to the surface of a CodeLink™ slide. After incubation at 55 °C for 1 h, the slide was rinsed briefly with 4×SSC (20 mL), then with 2× SSC, 0.1 % SDS (2×20 mL) at 55 °C for 5 min, with 0.2×SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using a Cy3 filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.4.

Initial primer extension experiments were performed on an automated hybridisation chamber at 55 °C for 1 h. An incubation time of 15 min proved to be adequate for a
successful incorporation of enzymatically cleavable nucleotides in growing DNA primers.

<table>
<thead>
<tr>
<th>DNA sequences</th>
<th>Fluorescence Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-H2N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F3' 3'-ATA CTG TGG CAG TAG TCG TCA TCA G-F5'</td>
<td>17,005</td>
</tr>
<tr>
<td>5'-H2N-TTT TTT TAT GAC ACC GTC ATC AGC AG-3' 3'-ATA CTG TGG CAG TAG TCG TCA TCA G-F5'</td>
<td>690,911</td>
</tr>
<tr>
<td>5'-H2N-TTT TTT CAG GAC CTA GAA CGG GCA GC-3' 3'-GTC CTG GAT CIT GCC CGT CGA TCA G-5'</td>
<td>605621</td>
</tr>
<tr>
<td>5'-H2N-TTT TTT CCT TCT TGC TGG GAC CCA AT 3' 3'-GGA AGA ACG ACC GTG GCT TAA TCA G-5'</td>
<td>713,966</td>
</tr>
</tbody>
</table>

Table 6.4 DNA sequences and correspondent fluorescence intensities (λem = 585 nm). Exposure time: 100 ms.

6.4.4 Two steps in one: annealing and primer extension

Primer Extension with dUTP-(A)₃-Rho.

5'-amino modified DNA primers were printed as described in the section 6.4.1. 100 μL of a solution containing three DNA templates (2 μM), Thermo Sequenase polymerase (1U/μL) and dUTP-(A)₃-Rho (2.5 μM) in 1× Thermo Sequenase reaction buffer, were heated to 85 °C for 3 min and added to CodeLink slides™. The slide was incubated at 55 °C for 1h and then cooled down to 35 °C over 1h. Then the slide was rinsed briefly with 4×SSC (20 mL), with 2× SSC, 0.1 % SDS (2×20 mL) at 55 °C for 5 min, with 0.2×SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using a Cy3 filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.5.
Table 6.5 DNA sequences and correspondent fluorescence intensities ($\lambda_{em} =$ 585 nm). Exposure time: 200 ms.

**Primer Extension with dCTP-(A)$_3$-FAM.**

Primer extension was carried out with dCTP-(A)$_3$-FAM as described in the section 6.44. The slide was scanned using a FITC filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.6.

<table>
<thead>
<tr>
<th>DNA sequences</th>
<th>Fluorescence Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5'$-H$_2$N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3'</td>
<td>200382</td>
</tr>
<tr>
<td>$3'$-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5'</td>
<td></td>
</tr>
<tr>
<td>$5'$-H$_2$N-TTT TTT TAT GAC ACC GTC ATC AGC AG-3'</td>
<td>1086064</td>
</tr>
<tr>
<td>$3'$-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5'</td>
<td></td>
</tr>
<tr>
<td>$5'$-H$_2$N-TTT TTT CAG GAC CTA GAA CGG GCA GC-3'</td>
<td>1375772</td>
</tr>
<tr>
<td>$3'$-GTC CTG GAT CTG CTC GCC GCT GCA TCA G-5'</td>
<td></td>
</tr>
<tr>
<td>$5'$-H$_2$N-TTT TTT CCT TCT TCG TGG CAC AT 3'</td>
<td>1219389</td>
</tr>
<tr>
<td>$3'$-GGA AGA ACG ACC GTG GGT TAA TCA G-5'</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.6 DNA sequences and correspondent fluorescence intensities ($\lambda_{em} =$ 480 nm). Exposure time: 100 ms.

**Primer Extension with dCTP-(A)$_3$-Cy5.**

Primer extension was carried out with dCTP-(A)$_3$-Cy5 as described in the section 6.44. The slide was scanned using a Cy5 filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.7.
6.4.5 Dual Color Primer Extension Assay

Dual color primer extension was carried out following the procedure described in the section 6.4.3 using dCTP-(A)$_3$-Cy5 (2.5 μM), dUTP-(A)$_3$-Cy3 (2.5 μM) and Thermo Sequenase DNA polymerase (1U/μL) in 1 × Thermo Sequenase reaction buffer for a total volume of 100 μL. The slide was scanned using Cy3 and Cy5 filters. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.8.

<table>
<thead>
<tr>
<th>DNA sequences</th>
<th>Fluorescence Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-H$_2$N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3' 3'-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5'</td>
<td>199410</td>
</tr>
<tr>
<td>5'-H$_2$N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3' 3'-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5'</td>
<td>395602</td>
</tr>
<tr>
<td>5'-H$_2$N-TTT TTT GGT GGA GAA CTG CCA GCT C-3' 3'-CCA CCT CTT GAC GGT CGA GAC TCG GAC G-5'</td>
<td>335376</td>
</tr>
<tr>
<td>5'-H$_2$N-TTT TTT TCT GGA ATA CTG CAG CAG C-3' 3'-AGA CCT ACT TAT GAC GTC GAC GAC ATC GAC-5'</td>
<td>322561</td>
</tr>
</tbody>
</table>

Table 6.7 DNA sequences and correspondent fluorescence intensities ($\lambda_{em}$ = 680 nm). Exposure time: 500 ms.

Table 6.8 DNA sequences and correspondent fluorescence intensities ($\lambda_{em}$ = 585 nm and 680 nm). Exposure time: 200 ms.
6.4.6 SBS cycles

Cycle 1: Extension with \(dUTP-(A)_3\)-Cy3

Primer extension was carried out with \(dUTP-(A)_3\)-Cy3 as described in the section 6.4.3. The slide was scanned using a Cy3 filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.9.

Cycle 1: Removal of the fluorophore

0.11 \(\mu\)moles of Subtilisin Carlsberg were dissolved in 200 \(\mu\)L of 10 mM Sodium Acetate buffer (pH7.5). The solution, pre-warmed to 37 \(^\circ\)C, was applied on the surface of the chip and left in incubation at 37 \(^\circ\)C for 15 min. The cleavage was repeated, if necessary. Afterwards, the slide was rinsed briefly with 4×SSC (20 mL), then with 2× SSC, 0.1 % SDS (2×20 mL) at 55 \(^\circ\)C for 5 min, with 0.2×SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using a Cy3 filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.9.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Extension</th>
<th>Cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′-H$_2$N-TTT TTT TAT GAC ACC GTC ATC AGC AG-F1-3′   3′-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5′</td>
<td>11997</td>
<td>26233</td>
</tr>
<tr>
<td>5′-H$_2$N-TTT TTT TAT GAC ACC GTC ATC AGC AGUA-3′   3′-ATA CTG TGG CAG TAG TCG TCA TCA G-F1-5′</td>
<td>745216</td>
<td>36803</td>
</tr>
<tr>
<td>5′-H$_2$N-TTT TTT AAA TAC TCA CCA AAC TGC CCT CUG 3′   3′-TTT ATG AGT GGT TTG ACG GGA GAC TAC TAC GA-5′</td>
<td>660282</td>
<td>1346</td>
</tr>
<tr>
<td>5′-H$_2$N-TTT TTT GTG GAG TGG TGG TGG TGG TCC ACA CCC CCG A-3′   3′-CAC CTC ACA AAG AGT TGT GGG GGA TAC GAC TAC-5′</td>
<td>1255773</td>
<td>65148</td>
</tr>
<tr>
<td>5′-H$_2$N-TTT TTT CCT TCT TGC TGG CAC CCA ATU C-3′   3′-GGA AGA ACG ACC GTG GGT TAA GAG AGA GAG-5′</td>
<td>1083526</td>
<td>38767</td>
</tr>
<tr>
<td>5′-H$_2$N-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTG G-3′   3′-GGC CCT TCT TAC CTG TCT G2T AAA AAG TCT CGG TGC T-5′</td>
<td>4981</td>
<td>14850</td>
</tr>
</tbody>
</table>

Table 6.9 DNA sequences and correspondent fluorescence intensities after the first cycle of extension and cleavage (\(\lambda_{em}=585\) nm). Exposure time: 100 ms.
Cycle 2: Extension with dCTP-(A)₃-Cy5

Primer extension was carried dCTP-(A)₃-Cy5 as described in the section 6.4.3. The slide was scanned using a Cy5 filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.10.

Cycle 2: Removal of the fluorophore

Removal of the fluorophore was carried out as described in the section 6.4.3. The slide was scanned using a Cy5 filter. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.10.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>Extension</th>
<th>Cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' - H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AG - F1-3' 3' - ATA CTG TGG CAG TAG TCG TCA TCA G - F1-5'</td>
<td>1032</td>
<td>1134</td>
</tr>
<tr>
<td>5' - H₂N-TTT TTT TAT GAC ACC GTC ATC AGC AGA - F1-3' 3' - ATA CTG TGG CAG TAG TCG TCA TCA G - F1-5'</td>
<td>1417</td>
<td>1205</td>
</tr>
<tr>
<td>5' - H₂N-TTT TTT AAA TAC TCA CCA AAC TGC CCT C G 3' 3' - TTT ATG AGT GGT TTG ACG GGA GAC TAC TAC GA-5'</td>
<td>683</td>
<td>1067</td>
</tr>
<tr>
<td>5' - H₂N-TTT TTT GTG GAG TGT TGG TTC ACA CCC CCG A - 3' 3' - CAC CTC ACA ACG TGT GGG GGA TAC GAC TAC-5'</td>
<td>676</td>
<td>1036</td>
</tr>
<tr>
<td>5' - H₂N-TTT TTT CCT TCT TGC TGG CAC CCA ATU C - 3' 3' - GGA AGA AGG ACC GTC GGT TAA GAG AGA GAG-5'</td>
<td>37782</td>
<td>2840</td>
</tr>
<tr>
<td>5' - H₂N-TTT TTT CCT TCT TGC TGG CAC CCA ATU C - 3' 3' - GGC CCT TCT CAC CTG TGT AAA AAG CTT CTG TCG TGC T-5'</td>
<td>1656</td>
<td>919</td>
</tr>
</tbody>
</table>

Table 6.10 DNA sequences and correspondent fluorescence intensities after the second cycle of extension and cleavage (λₑm= 680 nm). Exposure time: 100 ms.

6.4.7 Dual Color Primer Extension: Optimisation

Primer Concentration

5' amino-modified DNA primers were prepared at a final concentration of 30, 20, 10 and 5 μM in 1× printing buffer (50 mM sodium phosphate, pH 8.5) and printed on a CodeLink™ slide in four different subarrays. All primers were printed in three replicates as described in the section 6.4.1 and then hybridised to their complementary DNA templates. After washing and drying the slide, DNA primers were extended at 55 °C for 15 min using 100 μL of a reaction mix containing
Thermo Sequenase DNA polymerase (1U/μL), dCTP-(A)₃-Cy5 (2.5 μM) and dUTP-(A)₃-Cy3 (2.5 μM) in 1× reaction buffer (26 mM Tris-HCl, pH 9.5 and 6.5 mM MgCl₂). Then the slide was rinsed briefly with 4×SSC (20 mL), with 2× SSC, 0.1 % SDS (2×20 mL) at 55 °C for 5 min, with 0.2×SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using Cy3 and Cy5 filters. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.11 and 6.12.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>5 μM</th>
<th>10 μM</th>
<th>20 μM</th>
<th>30 μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>1969</td>
<td>1498</td>
<td>2639</td>
<td>1815</td>
</tr>
<tr>
<td>GGC CCT TCT CAC CTT GTT AAA AAG CTT CTG TGC T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-2HN-TTT TTT ATA CGA CTG AGG CGC CTA CCT 3'</td>
<td>11683</td>
<td>16412</td>
<td>19366</td>
<td>12303</td>
</tr>
<tr>
<td>TAT GCT GAC TCC GCG GAT AGA AGA CTC TCC T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-2HN-TTT TTT GGG TCG GTC AGT TCA CTG TGA AG 3'</td>
<td>14589</td>
<td>20901</td>
<td>30783</td>
<td>20552</td>
</tr>
<tr>
<td>CCC AGG TCA AGT GAC ACT TCA GCG TCT CTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-2HN-TTT TTT TCT GGA TGA ATA CTG CAG CTG 3'</td>
<td>26209</td>
<td>33900</td>
<td>90900</td>
<td>68276</td>
</tr>
<tr>
<td>GAG TAT ATT TAT CCG GUA TGG GTA GTC GTC ACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.11 DNA sequences and correspondent fluorescence intensities (λₑm= 680 nm). Exposure time: 100 ms.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>5 μM</th>
<th>10 μM</th>
<th>20 μM</th>
<th>30 μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>1338</td>
<td>1371</td>
<td>2173</td>
<td>1407</td>
</tr>
<tr>
<td>GGC CCT TCT CAC CTT GTT AAA AAG CTT CTG TGC T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-2HN-TTT TTT ATA CGA CTG AGG CGC CTA CCT 3'</td>
<td>294611</td>
<td>291356</td>
<td>642849</td>
<td>586908</td>
</tr>
<tr>
<td>TAT GCT GAC TCC GCG GAT AGA AGA CTC TCC T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-2HN-TTT TTT GGG TCG GTC AGT TCA CTG TGA AG 3'</td>
<td>134990</td>
<td>165346</td>
<td>324123</td>
<td>277849</td>
</tr>
<tr>
<td>CCC AGG TCA AGT GAC ACT TCA GCG TCT CTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-2HN-TTT TTT TCT GGA TGA ATA CTG CAG CTG 3'</td>
<td>27777</td>
<td>33343</td>
<td>62128</td>
<td>60055</td>
</tr>
<tr>
<td>AGA CCT ACT TAT GAT GTA GTC GAC ATC ATC G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-2HN-TTT TTT CTC ATA TAA ATA GCC CAT ACC CA 3'</td>
<td>7681</td>
<td>13363</td>
<td>36454</td>
<td>33720</td>
</tr>
<tr>
<td>GAG TAT ATT TAT CCG GTA TGG GTA GTC GTC ACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.12 DNA sequences and correspondent fluorescence intensities (λₑm= 585 nm). Exposure time: 100 ms.

**Template Concentration**

5’ amino-modified DNA primers were printed on a CodeLink slide at a final concentration of 20 μM. Each primer was printed in three replicates and in three identical subarrays. Afterwards, three solutions (10 μL each) of complementary
DNA templates prepared at different concentrations: 0.5, 2 and 5 μM were pipetted on each sub-array on the slide, covered by cover slips and incubated at 55 °C for 3h. After washing, the slide was spin-dried at 3000 rpm for 6 min, scanned and incubated at 55 °C for 15 min with 100 μL of a reaction mixture containing Thermo Sequenase DNA polymerase (1U/μL), dCTP-(A)₃-Cy5 (2.5 μM) and dUTP-(A)₃-Cy3 (2.5 μM) in 1× reaction buffer.

Then the slide was rinsed briefly with 4×SSC (20 mL), with 2× SSC, 0.1 % SDS (2×20 mL) at 55 °C for 5 min, with 0.2×SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using Cy3 and Cy5 filters. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.13 and 6.14.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>0.5 μM</th>
<th>2 μM</th>
<th>5 μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>230</td>
<td>1240</td>
<td>4618</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT ATA CGA CTG AGG GGC CTA CCT 3'</td>
<td>4053</td>
<td>34248</td>
<td>65431</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT GGG TCC AGT TCA CTG TGA AG 3'</td>
<td>10065</td>
<td>57565</td>
<td>89635</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT TCT GGA TGA ATA CTG CAG CTG 3'</td>
<td>66073</td>
<td>301994</td>
<td>236441</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT CTC ATA TAA ATA GGC CAT ACC CA 3'</td>
<td>5530</td>
<td>26197</td>
<td>13749</td>
</tr>
</tbody>
</table>

Table 6.13 DNA sequences and correspondent fluorescence intensities (λₑₒₑ= 680 nm). Exposure time: 100 ms.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>0.5 μM</th>
<th>2 μM</th>
<th>5 μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>1323</td>
<td>2365</td>
<td>2269</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT ATA CGA CTG AGG GGC CTA CCT 3'</td>
<td>473191</td>
<td>2159232</td>
<td>1424391</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT GGG TCC AGT TCA CTG TGA AG 3'</td>
<td>252602</td>
<td>54844</td>
<td>465582</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT TCT GGA TGA ATA CTG CAG CTG 3'</td>
<td>63551</td>
<td>111982</td>
<td>72616</td>
</tr>
<tr>
<td>5'-2HN-TTT TTT CTC ATA TAA ATA GGC CAT ACC CA 3'</td>
<td>49884</td>
<td>104373</td>
<td>105108</td>
</tr>
</tbody>
</table>

Table 6.14 DNA sequences and correspondent fluorescence intensities (λₑₒₑ= 585 nm). Exposure time: 100 ms.

224
Nucleotide Concentration

5’ amino-modified DNA primers were printed on CodeLink slides at a final concentration of 20 µM. Each primer was printed as three replicates and in three identical subarrays. After hybridisation with a 2 µM solution of DNA templates, to each sub-array on the slide were added three different solutions (10 µL each), containing Thermo Sequenase DNA polymerase at a constant concentration of 1U/µL and the labelled nucleotides dCTP-(A)₃-Cy5 and dUTP-(A)₃-Cy3 at a final concentration of 2.5, 5 and 25 µM. The three subarrays, each covered by a cover slip, were incubated at 55 °C for 15 min. Then the slide was rinsed briefly with 4xSSC (20 mL), with 2x SSC, 0.1 % SDS (2x20 mL) at 55 °C for 5 min, with 0.2xSSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using Cy3 and Cy5 filters. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.15 and 6.16.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>2.5 µM</th>
<th>5 µM</th>
<th>25 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' -2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>2132</td>
<td>2089</td>
<td>1949</td>
</tr>
<tr>
<td>GGC CCT TCT CAC CTT GTC AAA AAG CTT CTG TGC T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT ATA CCA CTG AGG CGC CTA CCT 3'</td>
<td>2613</td>
<td>6132</td>
<td>3065</td>
</tr>
<tr>
<td>TAT GCT GAC TCC GCC GAT GGA AGA ACG TCC T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT GGG TCC AGT TCA TGT TGA AG 3'</td>
<td>3377</td>
<td>5599</td>
<td>3599</td>
</tr>
<tr>
<td>CCC AGG TCA AGT GAC ACT TCA GCG TCT CTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT TCT GGA TGA ATA CGA CAG CTC 3'</td>
<td>5794</td>
<td>9933</td>
<td>11332</td>
</tr>
<tr>
<td>AGA CCT ACT TAT GAC GAC GAC ATC ATT G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT CTC ATA TAA ATA GGC CAT ACC CA 3'</td>
<td>4773</td>
<td>6851</td>
<td>5239</td>
</tr>
<tr>
<td>GAG TAT ATT TAT CCG GTA TGG GTA GTC GTC ACG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.15 DNA sequences and correspondent fluorescence intensities (λₑm = 680 nm). Exposure time: 100 ms.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>2.5 µM</th>
<th>5 µM</th>
<th>25 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' -2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>8681</td>
<td>10570</td>
<td>7633</td>
</tr>
<tr>
<td>GGC CCT TCT CAC CTT GTC AAA AAG CTT CTG TGC T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT ATA CCA CTG AGG CGC CTA CCT 3'</td>
<td>1084237</td>
<td>964024</td>
<td>1118387</td>
</tr>
<tr>
<td>TAT GCT GAC TCC GCG GAT GGA AGA ACG TCC T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT GGG TCC AGT TCA TGT TGA AG 3'</td>
<td>627764</td>
<td>556675</td>
<td>607813</td>
</tr>
<tr>
<td>CCC AGG TCA AGT GAC ACT TCA GCG TCT CTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT TCT GGA TGA ATA CGA CAG CTC 3'</td>
<td>206803</td>
<td>356655</td>
<td>167447</td>
</tr>
<tr>
<td>AGA CCT ACT TAT GAC GAC GAC ATC ATT G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' -2HN-TTT TTT CTC ATA TAA ATA GGC CAT ACC CA 3'</td>
<td>996853</td>
<td>842773</td>
<td>856402</td>
</tr>
<tr>
<td>GAG TAT ATT TAT CCG GTA TGG GTA GTC GTC ACG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.16 DNA sequences and correspondent fluorescence intensities (λₑm = 585 nm). Exposure time: 100 ms
**Enzyme Concentration**

Printing and hybridisation were performed as described in the section 5.10.2.1 and 5.10.3.1. Three solutions of 10 µL each, containing the labelled nucleotides dCTP-(A)3-Cy5 and dUTP-(A)3-Cy3 at a final concentration of 2.5 µM and Thermo Sequenase DNA polymerase at different concentrations (2, 1 and 0.25 U/µL) were pipetted on each sub-array and then covered by cover slips. After incubation at 55 °C for 15 min, the slide was rinsed briefly with 4×SSC (20 mL), with 2× SSC, 0.1 % SDS (2×20 mL) at 55 °C for 5 min, with 0.2×SSC (20 mL) at room temperature for 1 min and finally with 0.1 SSC (20 mL) at room temperature for 1 min. The slide was spin-dried at 3000 rpm for 6 min and scanned using Cy3 and Cy5 filters. DNA sequences of primers and templates and their fluorescence intensities (average) are shown in Table 6.17 and 6.18.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>0.25U/µL</th>
<th>1U/µL</th>
<th>2U/µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-'2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>1,458</td>
<td>1,885</td>
<td>9,143</td>
</tr>
<tr>
<td>GGC CCT TCT CAC CTT GTT AAA AAG CTT CTG T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT ATA CGA CTG AGG CGC CTA CCT 3'</td>
<td>4,018</td>
<td>8,354</td>
<td>23,804</td>
</tr>
<tr>
<td>TAT GCT GAC TCC GCG GAT GGA AGA CTC TCC T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT GGG TCC AGT TCA CTG TGA AG 3'</td>
<td>6,016</td>
<td>12,259</td>
<td>23,804</td>
</tr>
<tr>
<td>CCC AGG TCA AGT CAC TCA GCG TCT CTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT CCT CTA ATA ATG AGA CTC TCC T</td>
<td>6,334</td>
<td>21,467</td>
<td>55,604</td>
</tr>
<tr>
<td>AGA CCT ACT TAT GAC GAC GAC ATC ATC G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT CTC ATA TAA ATA GGC CAT ACC CA 3'</td>
<td>4,344</td>
<td>5,298</td>
<td>19,241</td>
</tr>
<tr>
<td>GAG TAT ATT TAT CCG GTA TGG GTA GTC GTC ACG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.17** DNA sequences and correspondent fluorescence intensities (λem= 680 nm). Exposure time: 100 ms.

<table>
<thead>
<tr>
<th>Primer/Template duplexes</th>
<th>0.25U/µL</th>
<th>1U/µL</th>
<th>2U/µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-'2HN-TTT TTT CCG GGA AGA GTG GAA CAA TTT TTC 3'</td>
<td>11,863</td>
<td>8532</td>
<td>9662</td>
</tr>
<tr>
<td>GGC CCT TCT CAC CTT GTT AAA AAG CTT CTG T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT ATA CGA CTG AGG CGC CTA CCT 3'</td>
<td>647524</td>
<td>979411</td>
<td>840812</td>
</tr>
<tr>
<td>TAT GCT GAC TCC GCG GAT GGA AGA CTC TCC T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT GGG TCC AGT TCA CTG TGA AG 3'</td>
<td>343279</td>
<td>563527</td>
<td>806383</td>
</tr>
<tr>
<td>CCC AGG TCA AGT CAC TCA GCG TCT CTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT CCT GGA TGA ATA CTC TCC T</td>
<td>193303</td>
<td>134855</td>
<td>337659</td>
</tr>
<tr>
<td>AGA CCT ACT TAT GAC GAC GAC ATC ATC G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-'2HN-TTT TTT CTC ATA TAA ATA GGC CAT ACC CA 3'</td>
<td>581772</td>
<td>674433</td>
<td>1010308</td>
</tr>
<tr>
<td>GAG TAT ATT TAT CCG GTA TGG GTA GTC GTC ACG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.18** DNA sequences and correspondent fluorescence intensities (λem= 585 nm). Exposure time: 100 ms

226
Primer Extension on aldehyde slides

Printing

Aldehyde slides were printed according to the manufacturer protocol. 5'-amino modified DNA primers were diluted to a final concentration of 20 µM in a solution 1:1 of 1× TE buffer and microarray spotting solution (0.3 M NaCl, 0.2 M NaHCO₃, pH 8.5). Printing was conducted at 75% humidity using a contact printer. Slides were washed twice with 0.2% SDS (20 mL) for 2 min and twice with deionised water for 2 min. Schiff bases were reduced at room temperature with a sodium borohydride (100 mg of sodium borohydride dissolved in 30 mL of PBS buffer and 10 mL of absolute ethanol). Slides were then washed 3 times with 0.2% SDS for 1 min, rinsed twice with deionised water and spin-dried at 4000 rpm for 6 min.

Hybridisation

Hybridisation was performed as for CodeLink™ slides. Slides were washed twice at 37 °C for 10 min with a SPSC buffer (50 mM sodium phosphate/1 M NaCl, pH 7.5), then at 37 °C for 1 min with a SPSC buffer at pH 8.5. Slides were spin-dried at 4000 rpm for 6 min.

Primer Extension

Primer extension was performed as for CodeLink™ slides. Washing was carried out using SPSC (50 mM sodium phosphate/1 M NaCl) buffers at pH 7.5 and 8.5. After incorporation of dUTP-(A)₃-Rho, slides were washed with SPSC buffer at pH 6.5.
Appendix: Optimisation of dual color primer extension

Primer concentration
Primer concentration gradient experiments were carried out to find the optimal primer concentration for dual color primer extension. Four subarrays of 5 DNA primers were printed at 5, 10, 20 and 30 μM concentrations, in duplicate.

![Graph A](image)

![Graph B](image)

**Figure 6.1** Primer Gradient Experiment: A) Normalised Fluorescence Intensities in the blue channel; B) in the red channel. Solutions: A=5 μM, B=10 μM, C= 20 μM, D= 30 μM.

After annealing with a solution containing complementary DNA templates (2 μM), primer extension was performed with Thermo Sequenase (1U/μL) and a combination of Cy3-(A)₃-dUTP and Cy5-(A)₃-dCTP (2.5 μM). Afterwards the slide was scanned in the blue and in the red channel.
Experimental data have shown that 20 and 30 μM solutions of DNA primer provided good signal intensity and specificity. Decreasing of fluorescence intensity at 30 μM concentration is due to dye quenching at high density of DNA primers.

**Template concentration**

A template concentration gradient experiment was run on a CodeLink™ slide. 20 μM solutions of 5 DNA primers were printed onto the slide in three subarrays. Afterwards, the primers were annealed to DNA templates added to different areas of the slide at different concentrations: 5, 2 and 0.5 μM. The result of dual color primer extension has shown that a concentration of 2 μM provided good signal intensity in the red and blue channel. Decreasing of fluorescence intensity at 5 μM concentration is due to dye quenching at higher density of labelled primers.

![Figure 6.2 Template Gradient Experiment: Normalised Fluorescence Intensities in the red channel. Solutions: A=0.5 μM, B=2 μM, C= 5 μM.](image1)

![Figure 6.3 Template Gradient Experiment: Normalised Fluorescence Intensities in the blue channel. Solutions: A=0.5 μM, B=2 μM, C= 5 μM.](image2)
**Effect of enzyme & nucleotide concentration**

Six subarrays of 5 DNA primers were printed on CodeLink™TM slides. After annealing, three subarrays of double stranded DNA were treated with three solutions containing the enzyme at different concentrations: 2, 1 and 0.25 U/μL and the nucleotides Cy3-(A)₃-dUTP and Cy5-(A)₃-dCTP at a constant concentration of 5 μM.

The other four subarrays were treated with three solutions containing the nucleotides Cy3-(A)₃-dUTP and Cy5-(A)₃-dCTP at different concentrations (2.5, 5 μM, 25 μM) while the enzyme concentration was kept constant (1U/μL).

![Graph A](image1.png)

![Graph B](image2.png)

**Figure 6.4** Enzyme Gradient Experiment: A) Normalised Fluorescence Intensities in the red channel B) in the blue channel. Solutions: A=0.25 U/ μL, B=1 U/ μL, C= 2 U/ μL.

This experiment was carried out in order to see if, higher concentrations of nucleotides would favorite multiple incorporations or mis-incorporation and to find
the minimum amount of enzyme needed to provide acceptable results, as Thermo Sequenase is relatively expensive.

As expected, higher signal intensities were obtained at higher concentrations of nucleotides and enzyme. No substantial difference was noted in terms of specificity at different concentrations of enzyme and nucleotides.

Figure 6.5 Nucleotide Gradient Experiment: A) Normalised Fluorescence Intensities in the red channel; B) in the blue channel. Solutions: A=2.5 μM, B=5 μM, C= 25 μM.
7 REFERENCES


(40) Kwok, P. Y. *Single Nucleotide Polymorphisms: Methods and Protocols*; Humana Press Inc.: Totowa, NJ.


(45) Shendure, J. A.; Porreca, G. J.; Church, G. M. *Curr Protoc Mol Biol 2008, Chapter 7, Unit 7.1*.


(149) Johnson, K., University of Southampton, UK, 2003.
Catch and release microwave mediated synthesis of cyanine dyes†

Maria Lopalco,* Eftychia N. Koini, Jin Ku Cho and Mark Bradley* 

Received 19th November 2008, Accepted 12th December 2008

Unsymmetrical functionalised cyanine dyes, covering the whole colour range, were readily synthesised (in 100 mg amounts) by a combination of microwave and solid-phase methodologies.

Fluorescent probes represent a versatile tool for visualising specific molecular targets and events in vitro and increasingly in vivo. Multicolour fluorescence detection has proved to be useful for multiplexed assays on a variety of microarray formats and for the simultaneous investigation of multiple biological processes in living cells. As a consequence, an increasingly large number of fluorescent probes, available in different colours, with the ability to be straightforwardly conjugated to biomolecules, are required.

The cyanine dyes (Fig. 1), a class of highly fluorescent compounds, have all the key requirements necessary for highly sensitive multicolour detection, with wavelengths which are tunable, by synthesis, across the visible spectrum. They also display excellent photophysical properties, including high extinction coefficients and quantum yields, while having fluorescence wavelengths remote from the natural autofluorescence of biomolecules. They are thus routinely used as fluorescent probes in a wide-range of applications such as DNA sequencing, cellular analysis, flow cytometry and in vivo imaging.

Conventional synthetic methods for the preparation of unsymmetrical cyanine dyes are based on the condensation of quaternised indolenine derivatives with vinylogous homologues of diphenyl formamidine and typically affording mixtures of the desired compound along with unreacted hemicyanine and undesired symmetrical analogues.

Scheme 1 Microwave-mediated alkylation of indolenines. (i) methyl iodide, acetonitrile, 150°C, 30 min to 1 h, microwave; (ii) 6-bromohexanoic acid, acetonitrile, 150°C, 1 to 3 h, microwave.

The heterocyclic quaternary ammonium salts 2 and 3 are usually obtained by heating the corresponding heteroaromatic base 1 with an excess of alkyllating agent in an aprotic solvent over several days. In the first simplification of the synthesis it was possible to reduce the reaction time by microwave heating the reaction mixture at 150°C in acetonitrile (Scheme 1) to give rapid access to a broad range of desired heterocyclic quaternary ammonium salts.

The electronic supplementary information (ESI) available: Detailed experimental procedures and characterisation data of the products. See DOI: 10.1039/b820719b
Scheme 2  Solid phase synthesis of the cyanine dyes: (a) triethylorthoformate for 6a, malonaldehyde bis dimethyl-acetal for 6b, BF$_3$Et$_2$O, DCM, 6 h; glutaraldehyde (2-pentenedial) dianilide hydrochloride, Ac$_2$O, DIEA, and pyridine for 6c; (b) DMF, 120 °C, 15 min, microwave, for trimethine and pentamethine dyes; Ac$_2$O, DIEA, and pyridine, 1.5 h; (c) DCM, 1 h.

Table 1  Properties of dyes synthesised on solid phase

<table>
<thead>
<tr>
<th>Dye</th>
<th>$\lambda_{\text{max, obs}}$ (nm)</th>
<th>$\lambda_{\text{max, en}}$ (nm)*</th>
<th>Yield$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8a</td>
<td>547</td>
<td>561</td>
<td>68%</td>
</tr>
<tr>
<td>8b</td>
<td>640</td>
<td>660</td>
<td>84%</td>
</tr>
<tr>
<td>8c</td>
<td>743</td>
<td>765</td>
<td>86%</td>
</tr>
<tr>
<td>8d</td>
<td>587</td>
<td>603</td>
<td>49%</td>
</tr>
<tr>
<td>8e</td>
<td>678</td>
<td>704</td>
<td>92%</td>
</tr>
<tr>
<td>8f</td>
<td>781</td>
<td>808</td>
<td>51%</td>
</tr>
<tr>
<td>8g</td>
<td>647</td>
<td>667</td>
<td>94%</td>
</tr>
</tbody>
</table>

* Measured in MeOH (except for compound 8g which was dissolved in H$_2$O).$^a$ Yield of isolated product in the final step (based on the amount of heterocycle used in the cleavage step).

Scheme 3  Cleavage of the cyanine dyes from the resin. (i) Ac$_2$O, (DIEA), and pyridine, 2 h.

in good yields (66-95%), including sulfonated (2c, 3c) and/or carboxylated (3a-c) variants.

To synthesise cyanine dyes the alkylated indolenines (2a-c) were reacted with the resin bound polymethine imines (6a-c) (Scheme 2, route A) as originally reported by Mason et al., although again using microwave irradiation (15 mins, 120 °C). Though microwave heating accelerates the attack of a second molecule of the indolenium salt 2 onto the polymer-bound imine 6, the symmetrical impurity obtained is cleaved from the resin and it was observed that microwave heating in this step gave rise to higher yields and purities of the final compound. Cleavage from the solid support was then accomplished using the carboxylated indolium salt (3a-c) to give functionalised cyanine dyes suitable for conjugation with biomolecules (Scheme 3). The unsymmetrical trimethine and pentamethine cyanine dyes were obtained in good yield (Table 1) and purity (85-100% by ELSD detection) and in preparative scale (up to 100 mg).

Further studies were carried out to extend this protocol to the synthesis of heptamethine dyes, as they have attracted great interest as fluorescent probes for in vivo imaging.$^{11-12}$ In an initial approach the monofunctionalised heptamethine dye 8c was obtained following the route A. Glutaconaldehyde dianilide hydrochloride was reacted with the supported aniline 4, in the presence of Ac$_2$O, to form the polymer-bound polymethine 6c.

The reaction of the indolenium salt 2a with 6c produced the desired supported hemicyanine 7c via loss of N-phenylacetamide,
but resulted also in the release of the hemicyanine in solution by cleavage from the resin, with a consequent decrease in the yield of the hemicyanine 7c and of the final compound 8c (12%) (Scheme 4).

Scheme 4  Formation of hemicyanine 7c. (i) 2a, Ac₂O, DİEA, pyridine, 1.5 h.

This led to the development of an alternative catch-and-release-strategy (Scheme 2, route B) based on the formation of the hemicyanine intermediates 5a,b in solution and the captive loading onto the solid support (4) to form 7c,f. The loading of the hemicyanine 5a onto the resin in this manner was carried out in DCM for 1 h, with an 86% yield as determined by nitrogen elemental analysis. The desired dyes were then cleaved from the resin by nucleophilic attack of 3a,b onto the intermediates 7c,f in pyridine:Ac₂O (10:1) to afford the unsymmetrical heptamethine dyes 8c and 8f in good yield (Table 1) and purity (>90%), as determined by HPLC analysis and ¹H NMR (Fig. 2).

The catch-and-release protocol described above has proven to be effective also for the synthesis of sulfoindocyanine dyes. The presence of sulfonate groups on the ring systems improves water solubility, prevents aggregation in water and reduces non-specific binding to biomolecules and cellular constituents. However, purification by chromatography of these dyes is difficult.

To overcome this limitation, Jiang et al. recently suggested the use of poly(ethylene glycol) as a soluble support, which has its own handling problems and low loading. In addition, Mason et al. described the synthesis of sulfonated cyanine dyes on polystyrene resin; however they reported that the sulfonated heterocycles did not react during the formation of the hemicyanine intermediate on solid phase but they did react in the dye formation step allowing the synthesis of mono-sulfonated cyanine dyes.

Microwave heating enabled the solid-phase synthesis of the pentamethine sulfoindocyanine dye 8g from polystyrene resin (route A), although in low yield (10%) and purity.
The catch-and-release method (route B) proved to be far more effective, resulting in the formation of 8g in excellent yield and purity (94%), without the need for further purification by chromatography (Fig. 3).

Fig. 3  HPLC chromatogram of the crude sulfonated dye 8g.

In conclusion, two practical approaches to the solid phase synthesis of unsymmetrical cyanine dyes suitable for bioconjugation have been reported. The method proposed is straightforward and enables, in a few steps, the synthesis of cyanine dyes spanning the whole colour range. Through a catch-and-release strategy we have also been able to synthesise, on solid phase, heptamethine and water soluble cyanine dyes in high yield and purity without the need for chromatographic purification, thus providing an easy route to the preparation of these invaluable fluorescent probes in amounts sufficient to satisfy the most hungry of biological appetites.

Acknowledgements

We thank the Basic Technology Programme of the RCUK. ENK thanks EU for a Marie Curie Host Fellowship for the Transfer of Knowledge.

Notes and references

Purification by column chromatography of the dyes obtained from the supported imidates was carried out to provide their detailed characterisation but may not be necessary for application of the dyes as labeling reagents.