UNIVERSITY OF SOUTHAMPTON

NON-GEL BASED ENZYME ASSAYS AND COMBINATORIAL SCREENING OF LINKER-DYE COMBINATIONS

by

Karin Johnson

Doctor of Philosophy

Department of Chemistry
Faculty of Science

July 2003

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE

DEPARTMENT OF CHEMISTRY

Doctor of Philosophy

NON-GEL BASED ENZYME ASSAYS AND COMBINATORIAL SCREENING OF LINKER-DYE COMBINATIONS

by Karin Johnson

An understanding of the function and role of nucleic acids has grown rapidly since the structure of DNA was first postulated in 1953. Modified nucleosides and nucleotides now play an important role both as therapeutic agents and as probes for obtaining information concerning the structure and sequence of DNA. This project set out to develop a new method of evaluating target compounds designed for DNA sequencing. Combinatorial chemistry and solid phase synthesis have revolutionised the process of drug discovery since their introduction but have not been widely applied in nucleoside chemistry. Realising the value of modified nucleosides and the advantages of solid phase chemistry it was an attractive proposition to combine the two to allow the preparation of libraries of nucleoside analogues. A thymidine-based resin was therefore developed and successfully used for the synthesis of libraries of nucleoside analogues. The resin proved to be well suited for solid phase peptide synthesis when libraries of compounds were easily synthesised using this resin.

A novel DNA sequencing-by-synthesis strategy has been proposed in which single stranded DNA fragments are immobilised on a solid support. The primer is extended one base at a time using nucleotides that carry a reporter moiety attached through a peptide spacer. To optimise the peptide functionality for cleavage and incorporation a library was synthesised on the thymidine resin and screened for efficient substrates. Several good substrates were identified using this approach. To further improve the throughput of the method a biocompatible thymidine resin was also prepared on which a smaller library was synthesised. On-bead screening proved successful and target sequences could rapidly be identified.

ACKNOWLEDGEMENTS

First I would like to thank my supervisors Professor Mark Bradley, Dr Jon Cummins and my advisor Dr Richard Brown for advice and guidance from the start of this project. I would also like to thank all enthusiastic staff working on the Single Molecule Detection project at Amersham Biosciences, especially Dr Pickering for proof reading and many useful discussions.

I would like to thank Miss Hernimann and Dr Langley for MS analysis and Mrs Street and Dr Wells for NMR.

I would like to thank Julia and the Bradley Group, both past and present members. I would especially like to thank Dr Mittoo for help with proof reading.

A big thank you goes to my husband Neil for the huge support during all those tough times. Finally, I would like to thank my family, my mother Siv and my brothers Erik and Gunnar who I hope to see more often now and my father Bo who encouraged me at the start of this journey but is unfortunately no longer with us, *thanks dad*.

Table of Contents

Abbreviations

δ	Chemical shift (ppm)
A	Adenosine
A	Alanine
Ac	Acetyl
ACCA	*cis*-4-amino-1-cyclohexane carboxylic acid
Ala	Alanine
Ahx	Aminohexanoic acid
Asp	Aspartic acid
ATP	Adenosine triphosphate
AU	Absorbance units
AZT	3'-Azido-2',3'-dideoxythymidine
Boc	*tert*-Butoxycarbonyl
Bu	Butyl
C	Cytidine
CPG	Controlled pore glass
D	Aspartic acid
d	Doublet
dATP	Deoxyadenosine triphosphate
DBU	1,8-diazabicycloundec-7-ene
DCC	Dicyclohexylcarbodiimide
dCTP	Deoxycytidine triphosphate
ddNTP	Di-deoxynucleoside triphosphate
dGTP	Deoxyguanidine triphosphate
DHAdT	5,6-Dihydro-5-azathymidine
DIC	Diisopropylcarbodiimide
DIPEA	Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethylsulphoxide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
dTTP	Deoxythymidine triphosphate
dU	Deoxyuridine
dUTP	Deoxyuridine triphosphate
DVB	Divinylbenzene
ES	Electrospray
Et_2O	Diethylether
FAM	Fluorescein
Fmoc	Fluoren-9-ylmethoxycarbonyl
FRET	Fluorescence Resonance Energy Transfer
G	Glycine
G	Guanidine
Gly	Glycine
HMBA	Hydroxymethylbenzoic acid
HOBt	1-Hydroxybenzotriazole
HPLC	High performance liquid chromatography
IR	Infrared

J	Coupling constant (Hz)
K	Lysine
L	Leucine
Leu	Leucine
LPCS	Liquid phase combinatorial synthesis
Lys	Lysine
m	Multiplet
Me	Methyl
min	minutes
MOM	Methoxymethylidene
MP	Macroporous
MS	mass spectrometry
MTr	Methoxytrityl
NMR	Nuclear magnetic resonance
P	Proline
PEG	Poly(ethylene glycol)
PEGA	Poly(ethylene glycol) dimethylacrylamide
ppm	Parts per million
Pro	Proline
PS	Polystyrene
R_f	Retention factor
RNA	Ribonucleic acid
R_t	Retention time
rt	Room temperature
s	Second
S	Serine
s	Singlet
Ser	Serine
SPOS	Solid phase organic synthesis
SPPS	Solid phase peptide synthesis
T	Thymidine
t	Triplet
TBAN	Tetrabutylammonium nitrate
TFA	Trifluoroacetic acid
TFAA	Trifluoroacetic acid anhydride
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TOPCAT	2-pyridiylthiocarbonate
TsOH	Tosic acid
U	Uridine
UV	Ultra violet

1 Introduction

The purpose of this chapter is to provide an introduction to solid phase chemistry and nucleosides. Emphasis has been placed on the necessary properties of biocompatible resins and nucleoside chemistry is focused on the modification of the nucleobase.

1.1 Biocompatible Solid Supports

1.1.1 Introduction

Combinatorial and solid phase synthesis has its origin in the synthesis of peptides. Peptide synthesis, by solution methods, often requires purification after each step making the method of synthesis long and laborious. Solid phase synthesis, first introduced by Bruce Merrifield[1] proved, in a revolutionary way, that methods for peptide synthesis could be greatly simplified by the solid phase method where the purification steps only became a matter of extensive washing of the resin. An immense achievement in 1971 was the solid phase synthesis of the enzyme Ribonuclease A (124 amino acid residues long, MW = 20kDa).[2-4] Polystyrene resin was the basis of initial solid phase synthesis but has since been subjected to continuous refinement. With the synthesis of longer peptides it soon became clear that the cross-linking of the resin had to be optimised. The best results were obtained with 1% cross-linked resin and this is still used today. Combinatorial and solid phase syntheses have also greatly increased the speed at which organic compounds can be produced such as in the area of drug discovery where libraries of compounds, as opposed to single compounds, are synthesised and then tested for bioactivity. Furthermore, if compounds could be screened for desired activities whilst still attached to the synthesis support the route to successful drug candidates would be shortened even further. In addition, since the cleavage step would be eliminated, the often necessary purification step after cleavage would no

longer be required and compound numbers generated by solid phase methods can be huge.

The recognised advantages of assaying and screening of compounds whilst still attached to their synthesis support, has led to investigations into the biocompatibility of many solid supports. Controlled pore glass,[5] carbohydrate,[6] polystyrene,[7] polystyrene grafted with other materials,[8] cellulose paper,[9] cotton and polyethylene glycol-acryl amide (PEGA) are all supports that have been explored with enzymes. However, the most commonly used synthesis supports are polystyrene (PS), TentaGel, and polyethylene glycol-polyacryl amide (PEGA), not all of which are "biofriendly".

1.1.2 Some Important Properties of Solid Supports

For successful solid phase organic synthesis (SPOS) or solid phase peptide synthesis (SPPS) the appropriate choice of solid support is of greatest importance. Resins used for organic synthesis need to be chemically robust and compatible with a wide range of reaction conditions to allow a diverse range of synthetic steps. To allow easy resin handling and / or automation the resin also needs to be mechanically robust. There have been many reports concerning resin break down in both strongly basic and acidic media while the mechanical instability of resins has been reported to clog up filters used in the synthesis process. Resins are, however, continuously being improved to minimise these problems.[10-12] Synthesis resins must possess good swelling properties and swell in a variety of solvents. Swelling is a property associated with almost every polymeric material and is related to the cross-linking in the polymer. A resin expands when solvent enters its structure and solvates it. The level of cross-linking limits the amount of swelling possible. The greater the swelling of a polymeric support the greater the space for access and diffusion of substrates to the active sites and hence greater reaction rates. Cross-linking is an important property and there is an inherent relationship between rigidity, mechanical stability, swelling properties and cross-linking parameters.[13]

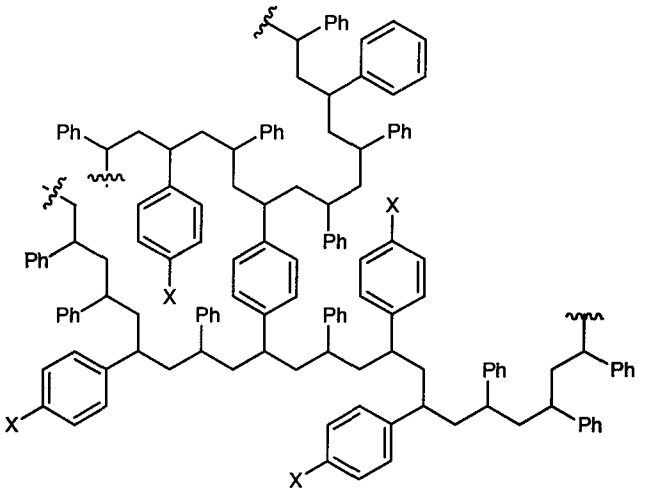
In terms of solid phase enzymatic assays it is necessary that the resins are biocompatible *i.e.* swell in aqueous media, show little non-specific binding to biomolecules such as proteins, allow enzymes access to the interior of the resin and do not interfere with enzyme activity.

1.1.3 Development and Evaluation of Biocompatible Resins

1.1.3.1 Polystyrene

Resins based on the polystyrene/ divinylbenzene (DVB) copolymers **1** are the most frequently used synthesis resins. They are cheap to prepare and have a relatively high loading level (0.5-1.5 mmol/ g). They are resistant to physical shock, which allows PS to be used in stirring or vortex mixing conditions. There is no real limit to the size of the molecules which can be synthesized on PS supports, peptides of >100 amino acid residues have been synthesised.[2] If harsh conditions such as high temperature or organometallic reagents are required more rigid 2% crossed linked PS is preferred. Of all reaction sites on the PS resin 99% are embedded inside the resin while only 1% is located on the surface of resin, therefore, for efficient reactions, proper swelling is very important. The hydrophobic properties of polystyrene make the resin swell very well in non-polar organic solvents, swelling up to five times its dry volume. However, the hydrophobic matrix results in poor swelling in polar solvents, in fact solvents such as water collapse the matrix completely and hence reactions are unable to proceed.

Polystyrene resins with high cross-linking, over 20%, have what is termed a macroporous (MP) structure and the properties are quite different from general PS. The high cross-linking gives the resin a very rigid structure whose shape is not affected by solvents. Therefore, various solvents, which cannot be used with the low cross-linked resin, can be used to perform reactions on MP resins. The higher cross-linking also means that the resin is more sensitive to mechanical shock as the shape of the resin interior is rigid hence macromolecule synthesis can be limited.



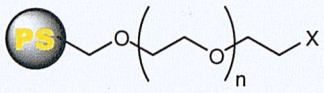
1

Figure 1.1.1 *The structure of polystyrene resin. X represents any suitable functional group which is normally obtained from Merrifield resin where X = chloromethyl*

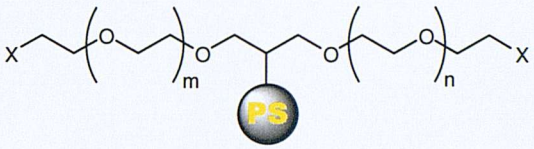
1.1.3.2 PS-PEG

Poly(styrene-oxyethylene) graft copolymers (PS-PEG) are combination-polymers developed to improve the defects of PS by introducing the more polar properties of PEG. The most eminent of these polymers, TentaGel **2**, consists of polyethylene glycol attached to cross-linked polystyrene through an ether linkage and combines the benefits of both the soluble PEG and the insoluble PS polymers. The PEG content in TentaGel is generally around 70-80% (by weight) and thus PS-PEG beads display relatively uniform swelling properties in a variety of solvents from toluene to water. As the reactions occur at the end of the PEG chains and there is no cross-linking between the long and flexible PEG chains, access of reagents to the reaction sites on PS-PEG is easier. This usually makes the reaction rate of PS-PEG higher than that of polystyrene. Disadvantages of PS-PEG resins are the relatively low loading levels compared to polystyrene, the potential of the PEG chains to complex Lewis acids and the potential instability of PEG. The polymer also has the tendency of resins to become sticky and difficult to handle as the syntheses progress. TentaGel **2** and ArgoGel **3,** developed to improve acid

stability, loading and purity of final compound, are examples of two commercial PS-PEG resins.



2



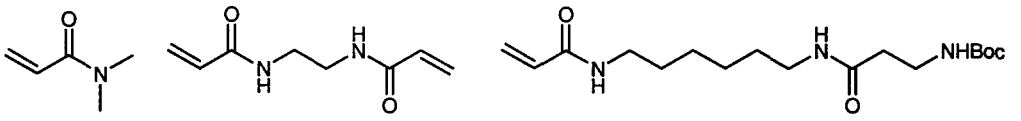
3

An evaluation of three resins for on-bead bioassays was carried out by Lowe[8] and co-workers. Papain, having a MW of 23 kDa, and one known peptide substrate, GGFGLGGG, was chosen for the investigation of TentaGel (TG), ArgoGel (AG) and PEGA, section *1.1.3.4*. The peptide was synthesised by SPPS on all resins and labelled with dansyl at the last residue. All resins were treated with papain and the changes in fluorescence of the beads were recorded. The peptide attached to TentaGel and ArgoGel beads appeared largely unaffected even after prolonged treatment with papain. However, analysis of the assay solution, by fluorescence spectroscopy, showed partial hydrolysis by both resins, which was in accordance with work by Lebl and co-workers,[14] indicating that PS-PEG exhibit poor biocompatibility.

Investigations into the biocompatibility of TentaGel and ArgoGel in protease assays[8,25] have mainly shown that the beads are only partially accessible to enzymes, which seem to be able to carry out their actions on the surface of the beads only. The fact that PS-PEG was only partially accessible to macromolecules presented another use of solid supports in enzymatic screenings. Lebl[14] presented a method of enzyme-mediated spatial segregation or "shaving" in 1996 where the selective hydrolysis provides a resin with dual functionalities and hence allows synthesis of two unique compounds per resin bead. TentaGel was treated with chymotrypsin, elastase and pepsin. Chymotrypsin accessed 2-2.5% of the sites whereas elastase and pepsin accessed 10-15% of the sites. Despite the general poor biocompatibility of TentaGel it has been found successful when applied to enzymatic phosphorylations.[15,16]

1.1.3.3 Polyamide

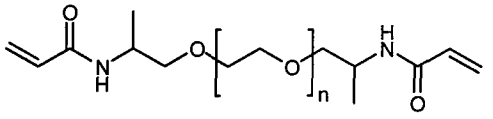
Ideally both the supporting polymer and the synthesised compound should be equally solvated under reaction conditions to enhance the rate and yields of synthesis. A dimethyl acrylamide resin suitable for peptide synthesis, where the support is structurally close to peptides, was developed in Sheppard's laboratory.[17] The combination of dimethylacrylamide **4**, *N,N'*-bisacryloylethylene-diamine **5** and *N*-Boc-alanyl-*N'*-acrylolhexamethylenediamine **6** gave a resin that possessed good swelling properties in water and polar organic solvents but unfortunately suffered from very poor swelling in less polar solvents such as dichloromethane.



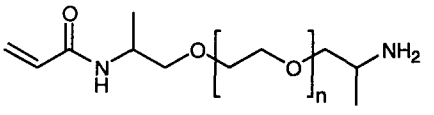
4 **5** **6**

1.1.3.4 Polyamide-PEG (PEGA)

Following on from the polyamide resins the hydrophilic and flexible PEGA resins were developed. PEGA, initially constructed for continuous flow solid phase synthesis, was developed in 1992 by Meldal.[18] The resin was designed to be highly polar assisting peptide solvation and allowing penetration of polar components into the interior of the bead. These beads displayed good swelling properties in dimethylformamide, dichloromethane, TFA, alcohols and water, hence showing a scope for many applications. PEGA resin differed from the earlier acrylamide resins in that it was obtained from bis-acrylamidopropyl-PEG **7**, acrylamidopropyl-aminopropyl-PEG **8** and dimethylacrylamide **4**.



7



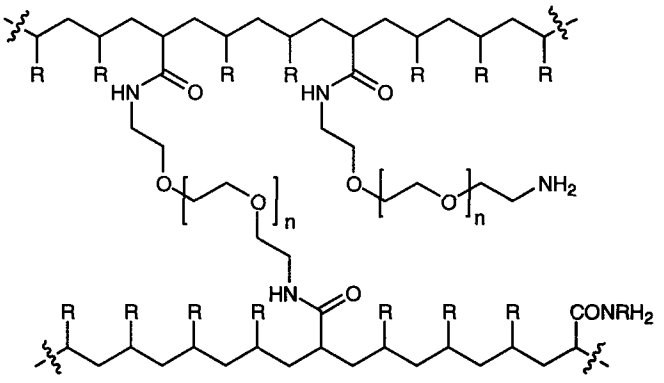
8

The high content of PEG in the polymer made it highly flexible and biocompatible. The PEGA matrix was designed and synthesised in such a way that macromolecules such as enzymes could enter the interior of the polymer freely and thereby facilitating solid-phase enzymatic reaction.[19]

In one assay development process Meldal[20] treated an amino-PEGA resin with a mixture of hydroxymethyl benzoic acid and the first amino acid residue, Fmoc-Lys(Boc)-OH, of a known substrate, deprotected the lysine side chain, introduced an intramolecular resonance energy transfer donor, and deprotected the Fmoc to reveal the N-terminus. This provided beads functionalised with two different nucleophiles, amine- and hydroxy-nucleophiles, with different reactivity. Utilising the selectivity of active esters for amino nucleophiles a known, fluorescently-labelled, peptidic substrate was assembled by conventional solid phase peptide synthesis (SPPS). The hydroxyl group was then esterified by Fmoc-Val-OH and a split synthesis library was assembled in a library generator. Attached to the same bead both the substrate and the potential inhibitor are contained in a confined space and therefore compete for the binding site on the same protein. The substrate was labelled with a fluorescent dye and as the library was subjected to a solution of subtilisin (MW = 27 kDa) the efficiency of the inhibitor could be measured by the change in fluorescence. A similar experiment was carried out for cruzipain[21] (MW = 57 kDa) and the results from both solid phase assays were identical when compared to the solution phase assays. PEGA was also found fully accessible to cathepsins B and D (MW = 42 kDa) where a comparison was made between libraries of peptides with an Fmoc protected N-terminus, trifluoro-acetylated N-terminus and free N-terminus. The two different enzymes produced a selection of substrates with clearly different sequences suggesting that the resin did not mask the specificity of the enzymes.[22]

1.1.3.5 Extended PEG chain lengths PEGA supports

It was shown that enzymes as large as 50 kDa[23] could enter the PEGA resin but in 1998 when Renil[24] *et al.* wanted to optimise substrates for MMP-9 it was necessary to increase the pore size of PEGA due to the molecular weight of MMP-9 (92 kDa). A series of PEGA supports, aimed at combinatorial peptide synthesis and solid-phase enzymatic library assays, were therefore prepared. Two types of resin with four different length chains $PEGA_{1900}$, $PEGA_{4000}$, $PEGA_{6000}$ and $PEGA_{8000}$ were synthesised, *Figure 1.1.2* and *Figure 1.1.3.*

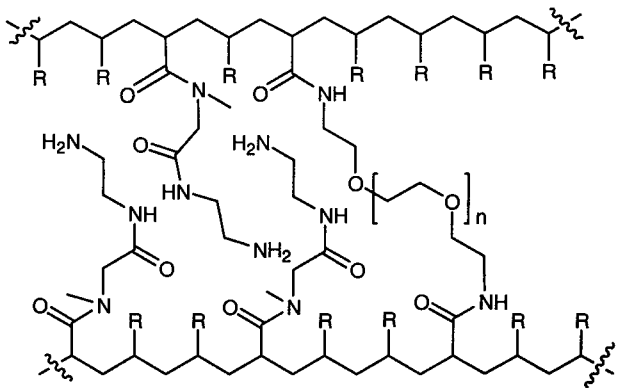


9

Figure 1.1.2 *PEGA Type I*[24]

R = CONH₂

n = 45, 89, 135, 180



10

Figure 1.1.3 *PEGA Type II*[24]

R = $CON(CH_3)_2$ or $CON(CH_3)CH_2CONHCH_2CH_2NH_2$

n = 45, 89, 135, 180

To obtain type I PEGA resin, partially acryloylated bis-amino-$PEG_{1900, 4000, 6000 \text{ or } 8000}$ were copolymerised with acrylamide. Type I resins have high swelling capacities and are suited for enzyme reactions. However, they have comparatively low amino group capacities of 0.08 - 0.13 mmol/g. The type II resin was prepared to increase the amino group capacity of the PEGA support by introducing sarcosin ethyl ester in the polymerisation mixture. The type II resin had a larger loading of 0.22 - 1.0 mmol/ g. These resins are more highly cross-linked and have therefore a lower swelling capacity. The swelling capacity was found to increase with the increase in the PEG chain length. The high degree of solvation of PEG provides a porous gel matrix completely filled with solvent. This gives a firm mechanically robust appearance of the swollen resin showing the most favourable swelling properties. Effective permeation of the resin matrix occurs under the influence of a swelling solvent leading to good exposure of reactive groups for peptide bond formation or enzymatic cleavage. It was found that the long PEG chain containing PEGA resins could efficiently be used in the synthesis and assays of fluorescently labelled peptide libraries. Solid phase enzymatic cleavage reactions using matrix metalloproteinase MMP-9 (molecular weight of the active form 67-83 kDa) was

optimised using intramolecular resonance energy transfer labelled libraries prepared on $PEGA_{4000}$ and $PEGA_{6000}$.

1.1.3.6 Controlled Pore Glass

Controlled pore glass (CPG) is a rigid glass-derived solid support compatible with any type of solvent, stable to aggressive reagents and extremes of pressure and temperature. Its rigid structure makes it suitable for continuous flow synthesis and it has therefore been used for the combinatorial synthesis of oligonucleotides. CPG is available in different bead and pore sizes; however, loading levels of glass resins are generally low.

The biocompatibility of CPG was recently compared to that of $PEGA_{1900}$ and TentaGel.[25] In this study one general peptide substrate was synthesised and subsequently loaded onto the three different supports. The eight-residue peptide substrate was terminated with 4-cyanobenzoic acid to permit analysis of the presence of stretching frequency of the cyano group by Raman microscopy.[26] All resins were incubated overnight with a number of enzymes, with molecular weights ranging from 22 kDa to 90 kDa. The study showed that TentaGel was inaccessible to all enzymes studied although this method would not detect cleavages lower than 5%. The PEGA resin was found to be accessible to MMP 12 (22 kDa) and thermolysin (35 kDa) but not to MMP 13 (42.5 kDa). CPG, with the large pore size of 100 nm, was accessible even to largest enzyme (90 kDa).

1.1.3.7 Conclusion

In conclusion there seem to be some confusion about the biocompatibility of TentaGel presented in the literature but two resins have reliable enzyme compatibility, CPG (large pore size) and PEGA. PEGA, which is the most exploited, could comfortably be used with enzymes up to 40 kDa and CPG is compatible with even larger enzymes.

A fundamental question, when considering solid phase assays, is that of the correlation between enzymatic activities on solid phase compared to that in solution. The activity observed from solid phase combinatorial library assays generally does not exhibit a direct, linear correlation to the activity observed in solution phase assays.[27] However, published studies show strong evidence of similarity in the overall ranking of the substrates.

1.1.4 Methods of Solid Phase Synthesis

Combinatorial synthesis is based on efficient common synthetic steps and libraries are either created by: a.) parallel synthesis, were milligrams of small numbers of compounds (~100) are obtained or, b.) split synthesis where large numbers of compounds are obtained but in very small quantities. With the increased interest in solid phase combinatorial synthesis different synthesis methodologies have been developed.

1.1.4.1 Multi-Pins

The *Multi-Pin*[28,29] method, where hundreds of peptides are synthesised in parallel, was introduced using polyethylene rods grafted with polyacrylates. The pins were arranged in a microtitre plate format by attachment to a supporting block. This format allowed 96 different peptides to be synthesised simultaneously and also allowed microtitre plates to be used as reactors. Peptides were synthesised on the pins and the products could either be cleaved into a microtitre plate or tested whilst still attached to the support used for their synthesis. Pins had their greatest impact on the synthesis of peptides for epitope mapping, the process of making and testing every short peptide along the complete sequence of an active protein. Pins were advantageous over traditional solid supports as they were easily manipulated but suffered from rather low loadings and were no longer feasible when the number of required compounds exceeds several hundred.

1.1.4.2 Tea-Bags

The *tea-bag method*[30] provides a rapid way of synthesising a large number of peptides (>100) in quantities that satisfy most needs (>10 mg). Houghten, who developed this method in 1985, described a synthesis of 247 individual peptides, each peptide being obtained in 10-20 mg quantities *i.e.* plenty for conventional characterisation and screening. Synthesis was carried out on 'bundles' of beads contained within polypropylene mesh teabag-like packets, *Figure 1.1.4*. HPLC data showed that no cross-contamination occurred for different peptides prepared simultaneously and purity was found to be as good as, or better than, the purity of peptides prepared by traditional free resin based syntheses.

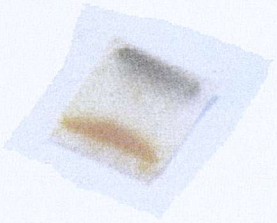


Figure 1.1.4 *Photo of a "tea-bag" containing resin and a microchip for identification*

Using the tea-bag methodology one can combine all washing, deprotection and common coupling steps during synthesis and hence achieve improved efficiency.

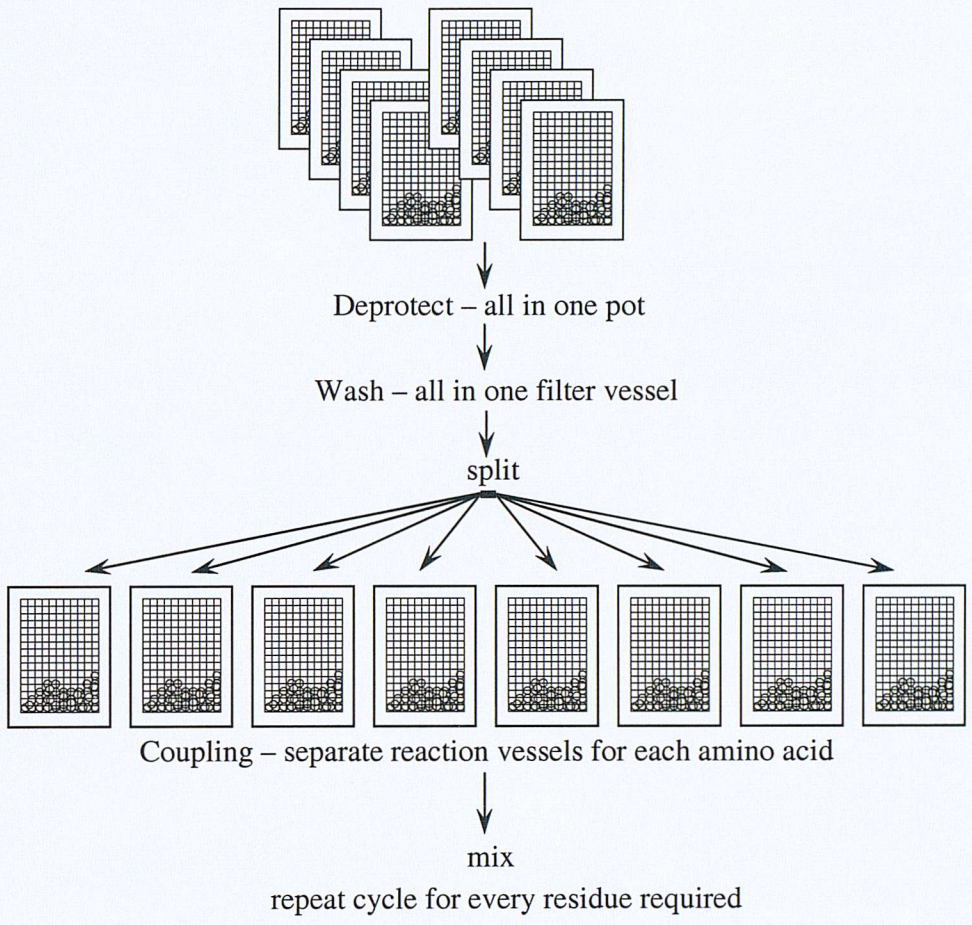


Figure 1.1.5 *Illustration of the combinatorial tea-bag method developed by Houghten*

Later more refined systems were introduced for the synthesis of larger number of compounds in multi-milligram quantities. Microreactors[31] such as MicroKans™, also based on 'bundles' of resin, have been introduced. These microreactors are prepared from polypropylene mesh with, radio frequency tags used for labelling and an automated directed sorting system.

1.1.4.3 Spot Synthesis

Spot synthesis,[9] a simple and practical approach for the simultaneous parallel synthesis of large numbers of peptides, was described in 1992 by Frank. Spot synthesis is simple and makes economic use of reagents and allows nanomolar to micromolar scale syntheses. In this method a sheet of paper was derivatised by esterification of an acid to the hydroxyl groups on cellulose. The carboxylic acid could be an amino acid or a cleavable handle to allow peptide cleavage. Reagents were spotted onto paper manually or automatically by pipette workstations. Amino acid couplings go to completion in soaked cellulose paper. Hence, there is no need for a large excess of solutions or reagents. Volumes and positions can be calibrated not to overlap, while the spot size is determined by the volume dispensed, the absorptive properties of the paper material and the volatility of the solvent. Peptides can be assayed immobilised or punched out and cleaved for use in solution and the ready availability of different papers allow a choice to meet the specific properties necessary.

1.1.4.4 Split Synthesis

A further innovation that allowed increased synthetic productivity was the method described as *split synthesis,* first reported for the synthesis of libraries of equimolar peptide mixtures by Furka.[32] The *one-bead-one compound*[33] concept was recognised by Lam and co-workers in 1991 and was based on the fact that combinatorial bead libraries contain single beads displaying only one type of compound. This methodology involves creating large peptide libraries consisting of potentially millions of compounds but in very small quantities. In the first cycle beads are distributed into separate reaction vessels, with a single amino acid in each vessel. The beads are pooled and then split again. The result is a library of bead bound peptides where each bead contains a single peptide, although with many copies. The complete collection represents all possible random peptides in roughly equimolar ratios, *Figure 1.1.6.*

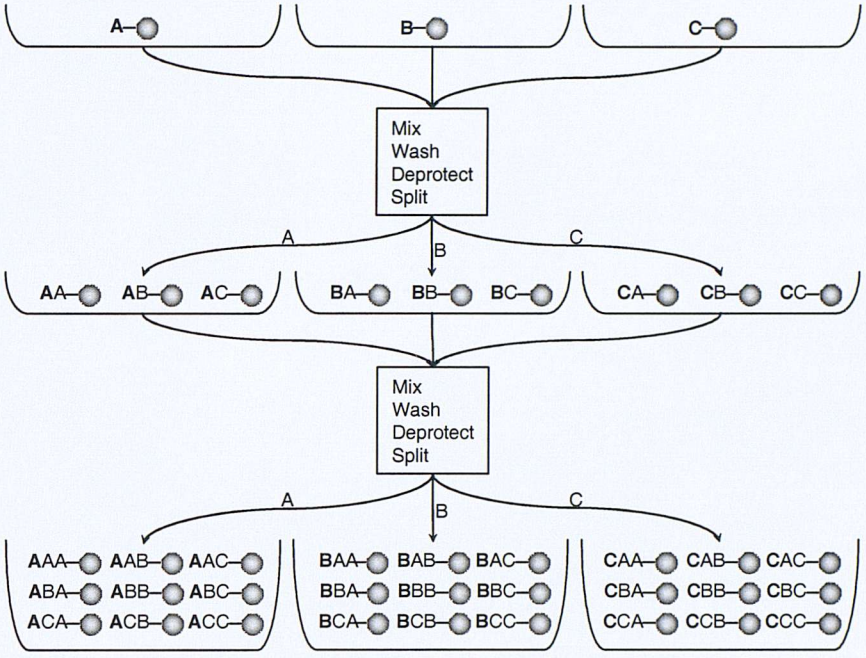


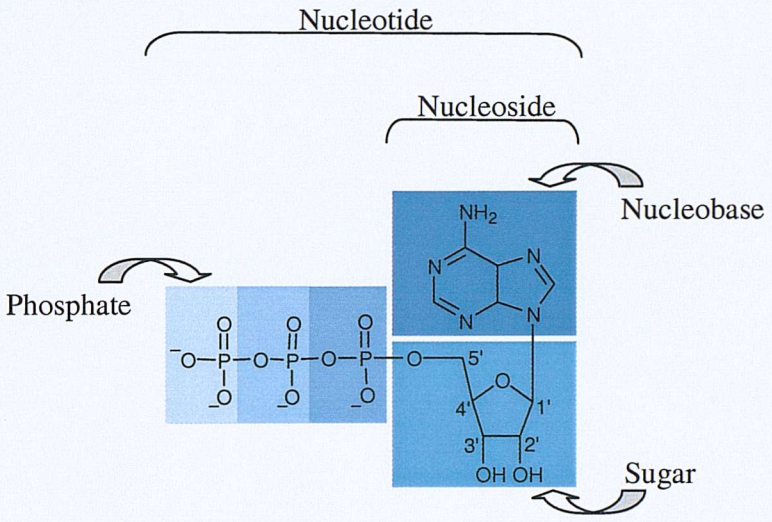
Figure 1.1.6 The Split Synthesis Method

1.2 Introduction to Nucleic Acids

Nucleic acids are essential for the storage of genetic information, the passing of that genetic information to the next generation and the expression of that genetic information in the organism. Deoxyribonucleic acid (DNA) is the permanent storage deposit of genetic information in the nucleus of the cell specifying the synthesis of ribonucleic acids (RNA). The amino acid sequence of every protein in a cell, and the nucleotide sequence of every RNA fragment, is specified by a nucleotide sequence in the cell's DNA. Close relatives of DNA and RNA are nucleotides such as adenosine triphosphate (ATP), much smaller molecules that are the chief energy storage molecules for all life processes. Nucleic acids and nucleoside mimetics have played an important role for advances in medicine and life sciences. The potential of nucleic acids as therapeutic agents was realised in the 1950s and 1960s when naturally occurring, and later synthetic, nucleosides with anticancer activity were discovered.[34]

1.2.1 Primary Structure of DNA

There are two characteristic components to a nucleoside: a pentose sugar, a nitrogenous base and in the case of nucleotides, one or more phosphate groups.



11

Figure 1.2.1 *A nucleotide molecule contains a phosphate group, a pentose sugar group and an organic base group. Illustrated here by the energy storage molecule adenosine triphosphate, ATP,* **11**.

1.2.1.1 Sugar

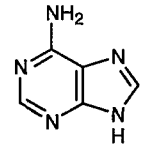
Nucleotides fall into two groups, those that contain a ribose sugar group, like ATP in *Figure 1.2.1*, and those that contain a deoxyribose sugar group, which lacks the oxygen on carbon 2'. The base component of the nucleotide is connected to the asymmetric carbon 1' of the sugar, called the anomeric centre, and the phosphate group of the nucleotide is joined to carbon 5' of the sugar.

1.2.1.2 Base

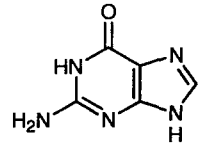
There are five principal organic bases found in nucleotides, three of them, *i.e.* adenine, cytosine, and guanine, are shared by RNA and DNA, but the fourth is

specific. Uracil occurs only in RNA and 5-methyluracil, *i.e.* thymine, occurs only in DNA. The naturally occurring nucleobases can be divided into two main groups: purines and pyrimidines. Adenine **12** and guanine **13** are purines, which means that they have a double ring structure, while cytosine **14**, thymine **15** and uracil **16** are pyrimidines, which means that they have a single ring structure.

Purines

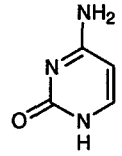


Adenine
12

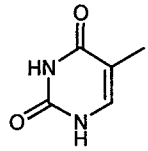


Guanine
13

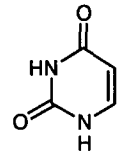
Pyrimidines



Cytosine
14



Thymine
15

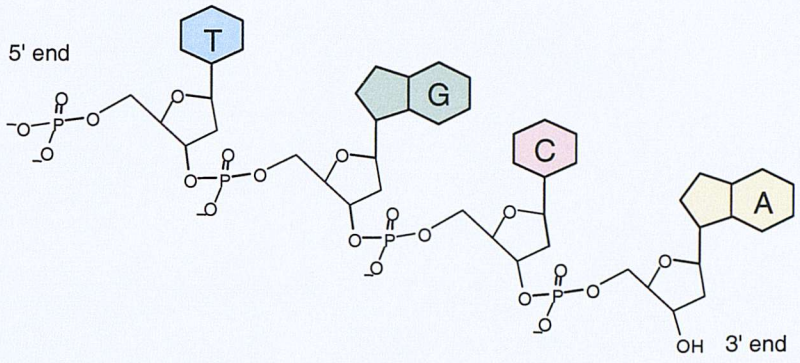


Uracil
16

1.2.1.3 Phosphate

One, two or three phosphates can be joined to the sugar molecule through the 5' hydroxyl. At basic pH the monophosphate form of the nucleotide has two negative charges, the diphosphate form has three and the triphosphate form has four. Nucleotides are joined together in a condensation polymerisation reaction to form covalent bonds between the hydroxyl on carbon 3' of the pentose sugar group and the phosphate group, attached to the 5' hydroxyl, of the next nucleotide. These phosphodiester bonds link successive nucleotides to form polynucleotides, DNA or RNA, comprising a sugar-phosphate backbone with the bases projecting from one side of this backbone. These 3'-5' phophodiester bonds all have the same

orientation and the phosphate groups, are completely ionised and negatively charged at neutral pH, giving the nucleic acid strand its direction.



Scheme 1.2.1 *Primary structure of DNA*

1.2.2 Secondary Structure of DNA

In 1953 Watson and Crick[35] postulated that DNA was composed of two polymeric strands coiled together into a *double helix*. This was based on, and agreed with, x-ray diffraction data obtained by Franklin[36,37] and Wilkins[38] and Chargaff's rules.

The base units of each strand are pointed into the interior of the helix, and pairs of bases from the two strands are linked together by hydrogen bonds. The critical aspect of the Watson-Crick model is that hydrogen-bonding can best occur between specific bases. A purine base of one strand generally pairs with a pyrimidine base of the other strand. Watson and Crick found that the complementary hydrogen-bonded base pairs, G with C and A with T, are those that fit best within the structure, providing support for Chargaff's rule which states that in DNA, the amount of A is the same as that of T and the amount of G is the same as that of C. In order for the base pairs to form, the polynucleotides have to run in opposite directions, 5'-3' and 3'-5', *i.e.* the strands must be antiparallel. When the base pairs form they twist the polynucleotide into a double helix stabilised by the base pair hydrogen bonds and stacking of the bases. The complementary base pairing in the double-stranded helix is a kind of 'back-up system' nature has devised to support accurate replication in new DNA.

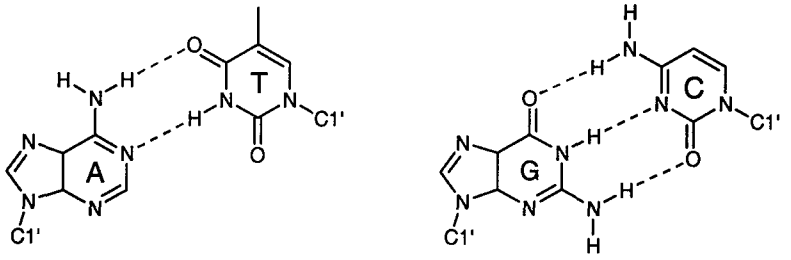


Figure 1.2.2 *Hydrogen-bonding in the base pairs defined by Watson and Crick. Thymine forms two hydrogen bonds with adenine and cytosine forms three hydrogen bonds with guanine. Each base acts both as a hydrogen bond donor and acceptor within the base pair.*

1.2.3 The Sequence of DNA

The most important property of a DNA molecule is its nucleotide sequence. In 1977 two techniques, which allow analysis of genes at the nucleotide level, were independently developed by Maxam and Gilbert[39] and Sanger.[40]

The most commonly used method, the Sanger or chain termination method, utilises 2',3'-dideoxynucleotide triphosphates (ddNTPs), molecules that differ from deoxynucleotides by having a hydrogen atom attached to the 3' carbon rather than an OH group. These molecules terminate DNA chain elongation because they cannot form a phosphodiester bond with the next deoxynucleotide, *Figure 1.2.3*. Enzymatic strand synthesis is carried out four times, in parallel, in the prescence of all the four different deoxynucleoside triphosphates dATP, dCTP, dGTP, dTTP, a DNA polymerase and 1% of the respective dideoxynucleotide. DNA to be sequenced acts as a template for the enzymatic synthesis of new DNA (starting at a defined primer) and incorporation of a dideoxynucleotide interrupts further chain elongation.

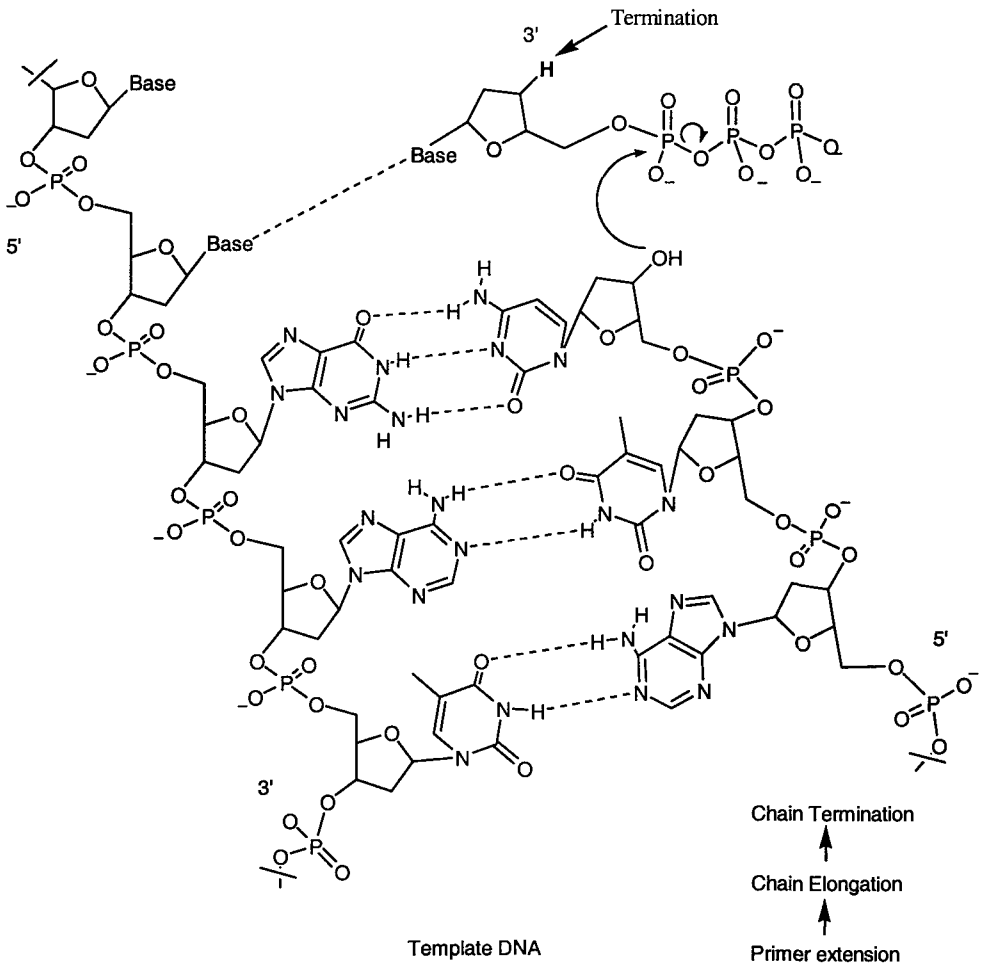


Figure 1.2.3 *DNA strand termination achieved by introducing 1% ddNTPs in enzymatic strand synthesis reaction, once a ddNTP has been incorporated chain extension is no longer possible.*

The result from the enzymatic extension of the oligonucleotide primer in the presence of chain terminating ddNTPs is four different families of DNA fragments having primer defined 5' ends and variable 3' ends. The location of the specific bases can be determined by electrophoresis, the smallest piece of DNA moving furthest.

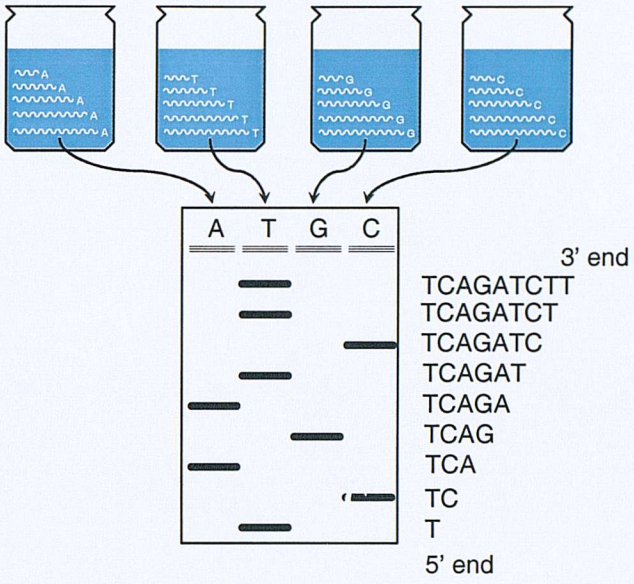


Figure 1.2.4 *Illustration showing how electrophoresis can provide information about the sequence of nucleotides in DNA. Each family terminated by ddATP, ddTTP, ddGTP or ddCTP reveal information about the DNA sequence.*

1.2.3.1 Detection of oligonucleotide bands

Detection of oligonucleotide bands on the electrophoresis gel was typically achieved by incorporating radio labelled dNTPs and visualisation by autoradiography[41] but advances in spectroscopic methods have changed the preferences in DNA sequencing.[42,43] Labelling of oligonucleotides by fluorescence has not only eliminated use of hazardous radioactive chemistry but also increased sensitivity and allowed single lane electrophoresis. To allow single-lane electrophoresis to replace the traditional four-lane electrophoresis, *Figure 1.2.4*, each family of oligonucleotides must be labelled with one of four distinguishable dyes. For sensitive DNA sequencing, one would ideally like each of the four dyes to:

- have an emission maximum at a distinctly different wavelength,
- absorb strongly at a common wavelength,
- be highly fluorescent to provide sufficient detection sensitivity,
- not significantly impair the hybridisation of the oligonucleotide primer; and,
- introduce the same relative electrophoretic mobility shift of the DNA sequencing fragment.

These requirements, though not by single fluorescent molecules, are fulfilled by energy transfer primers. In 1995, Ju *et al.* presented the use of fluorescence resonance energy transfer, FRET, technology in DNA sequencing where FRET was used to optimise the absorption and emission properties of the primer label.[44] Four different primers each with FAM as a common donor dye and four different acceptor dyes were synthesised **17, 18, 19** and **20**. Strand synthesis reactions were performed, one for each specific dye labelled primer/ dideoxynucleoside triphosphate combination. The reaction mixtures were then combined and electrophoresed, now only using one lane. The common donor dye allows only one laser to be used for excitation. Good spectral separation on emission and increased signal strengths were obtained.

17 **18** **19** **20**

DYE^1 = 5-carboxyfluorescein or 2'-7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein for **17** and **18** respectively.

DYE^2 = *N, N, N',N'*-tetramethyl-6-carboxyrhodamine or 6-carboxy-rhodamine for **19** and **20** respectively.

1.2.4 Nucleoside Analogues

Over the years many different chemical/ structural modifications have been made to nucleosides. Modifications have been carried out on all the different

characteristic parts of the nucleic acids. The sugar moiety can be modified so that the heterocyclic base is attached to the 2'- or 3'-position. Substituting the ring oxygen with other heteroatoms has also shown promising results regarding biological activity where increased stability and decreased toxicity have made it superior of its 4'-*O* analogue. The sugar moiety can be further modified by introducing two heteroatoms in the ring, either the same or different, *e.g.* 2',3'-dideoxy-4'-thiacytidine which possess anti-HIV activity.[45] The anomeric carbon can be attached to the nitrogen on the base moiety, *N*-nucleosides, or a carbon on the base moiety, *C*-nucleosides.

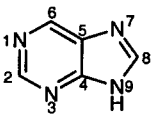
Herein focus will be limited to heterocyclic base modifications of *N*-nucleosides.

1.2.5 The Heterocyclic Base - Modifications

Modifications of the base generally concern ring substituents, aza or deaza modifications. When the nitrogen atoms, in purine **21** or pyrimidine **22**, are replaced or added the resultant analogues are called *deaza-* or *aza*-analogues, respectively.

1.2.5.1 Purine Modifications

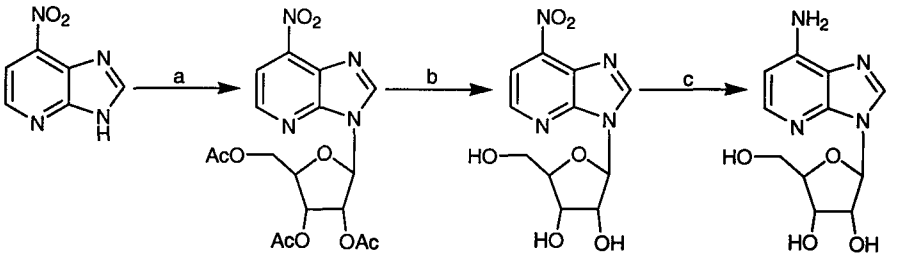
The conventional numbering of the purine structure is shown on the right. This numbering system will be referred to throughout the text.



21
Purine

1-Deaza

1-Deazaadenosine **26** has been found to have a wide spectrum of biological properties. Anti-tumour activity, inhibitor of adenosine deaminase and agonist interaction with adenosine receptors are examples of activities reported.[46] The 1-deaza analogue can be obtained from the nitroimidazo-pyridine derivative **23** by glycosylation of tetra-*O*-acetylated ribofuranose (TAR), removal of the acetate protecting groups and finally reduction of the nitro group.



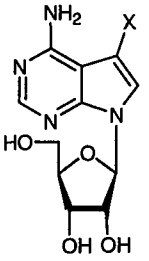
23 **24** **25** **26**

a.) TAR, $SnCl_4$ b.) NH_3 c.) H_2, Pd/C

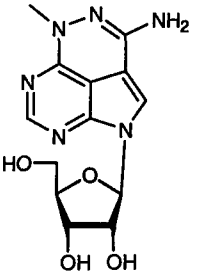
Scheme 1.2.2 Synthesis of 1-deazaadenosine **26**[46]

7-deaza

7-Deazaadenosine **27**, also known as tubercidine, is a naturally occurring nucleoside antibiotic. The synthesis and properties of this compound as well as those of 7-deaza-7-substituted analogues such as toyocamycin **28** and sangivamycin **29**,[47,48] have been explored as potential drugs for the treatment of human cytomegalovirus (HCMV). HCMV is a virus generally not harmful to immuno-competent individuals. However, it is dangerous to individuals with an immunodeficiency since they are unable to control the replication of HCMV. Triciribine **30** and its derivatives comprise another group of 7-deaza nucleosides that exhibit activity against HIV.[49]

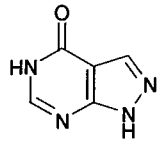


X = H (tubercidine) **27**
X = CN (toyocamycin) **28**
X = $CONH_2$ (sangivamycin) **29**

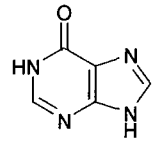


30

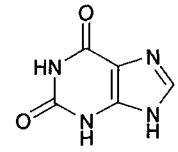
Gout is perhaps the oldest known type of arthritis, caused by an excess of uric acid in the body, but today, mainly thanks to the 7-deaza compound allopurinol **31**, gout also remains one of the most treatable forms of arthritis. Purine metabolism expels hypoxanthine **32**, xanthine **33** and uric acid **34** and these last two steps in the purine breakdown pathway are catalysed by an enzyme, xanthine oxidase. Allopurinol **31** is an inhibitor of xanthine oxidase and so reduces the build up of urates and increases the levels of the more soluble hypoxanthine and xanthine.



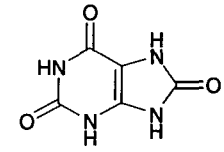
31



32



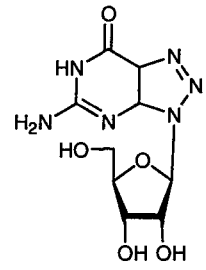
33



34

The importance of N7 of the purine moiety varies. The purine N7 is known to take part in triple helix base pairing and could therefore be of importance in association with ribosomal proteins. However, a study comparing the nucleoside antibiotic puromycin versus the 7-deaza analogue proved that the N7 played little, if any, part at all when no loss, or indeed increase, in bio-activity was observed.[50] In addition to the potential of 7-deaza nucleoside analogues as drugs they are also useful in DNA sequencing as chain terminators, which will be discussed in more detail in section *3.1.1.2*. The glycosyl bond of N-7 alkylated purine nucleosides has been found to be more labile than that of the 7-deaza-7-substituted.[51] 7-deaza-7-substituted purines are directed toward the major groove of DNA whereas 8-substituted purines interfere sterically with the sugar-phosphate backbone and hence destabilise DNA duplexes.[52,53] DNA duplexes have been significantly stabilized by 7-deaza-7-substituents such as chloro-, bromo-, iodo- and alkynyl-groups.[54,55]

8-aza

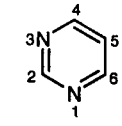


35
8-Azaguanosine

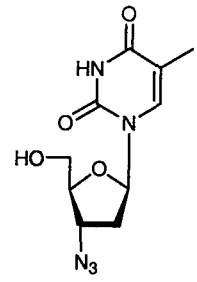
8-Aza purines have attracted interest due to the antifungal, antiviral and anticancer activities found for naturally occurring 8-azaguanine.[56,57] As well as 8-azaguanine **35** several other 8-azapurine ribonucleosides have been prepared by chemical and enzymatic methods. The synthesis is difficult due to the many possible glycosylation products brought on by the additional N(8)-atom. The glycosylation step yields not only the three different regioisomers but also both the α and β anomers. The glycosylation reaction was improved by Seela who found that nucleobase-anion glycosylation was stereoselective giving only the desired β anomer.[58] The reduced basicity of N(7) in 8-azapurine analogues compared to that of purines makes them more difficult to protonate. This in turn means that 8-aza nucleosides do not form triplex DNA. Furthermore, oligonucleotides containing 8-azaguanine/ cytosine base pairs form more stable duplex structures than those containing pairs of guanine/ cytosine pairs.[59]

1.2.5.2 Pyrimidine Modifications

The conventional numbering of the pyrimidine structure is shown on the right. This numbering system will be referred to throughout the text. The first drug to be approved for clinical use against HIV was a pyrimidine nucleoside, 3'-Azido-2',3'-dideoxythymidine (AZT) **36**, which was initially synthesised as an anticancer drug in 1964.[60]



22
Pyrimidine



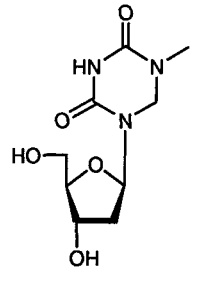
36

3-deaza

3-deaza nucleoside analogues have been noted to have anti-cancer activity. The 3-deaza pyrimidine derivatives can be obtained from various pyridine starting materials.[61,62] The pyridine ring is modified as far as possible to provide the desired aglycone. Glycosylation with the pyridine derivative and a hydroxyl protected 1'-halo sugar forms the nucleoside.

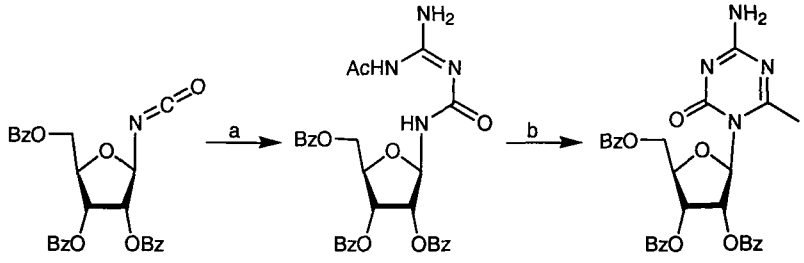
5-aza

5,6-Dihydro-5-azathymidine, DHAdT, **37** possesses activity against a variety of DNA viruses. DHAdT is a naturally occurring nucleoside and it has also been prepared chemically.[63] The triazine moiety of 5-aza nucleosides can either be directly coupled to the sugar[63,64] moiety or prepared from the 1'-isocyanate derivatives of the sugar moiety. 5-Azacytidine and 2'-deoxy-5-azacytidine are used in treatment of acute leukaemia and variations of these compounds are therefore very interesting.



37

Preparation of a 6-substituted-5-azacytidine derivative by the 1'-isocyanate method is shown in *Scheme 1.2.3*.[65]

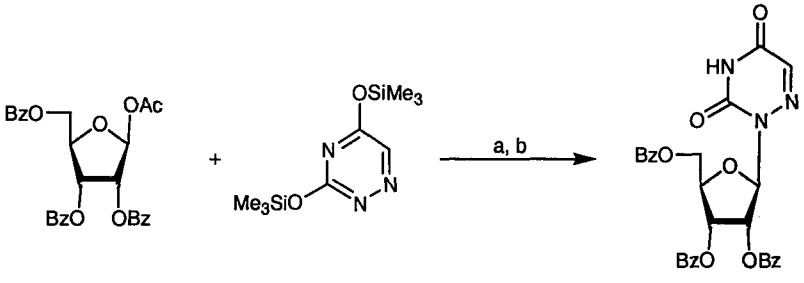


a.) N-acetylguanidine, Acetone
b.) Chlorotrimethylsilane, NEt_3, Acetone

Scheme 1.2.3 Synthesis of 2',3',5'-Tri-O-benzoyl-6-methyl-5-azacytidine

6-aza

6-aza nucleoside analogues of uridine, thymidine and cytidine are found to be antiviral agents. The potential carcinogenicity of 6-aza compounds compared to 5-aza compounds is very low.[66] 6-aza nucleosides can be prepared by the Hilbert-Johnson reaction as modified by Vorbrüggen.[67,68] The glycosylation reaction, which uses $SnCl_4$ as catalyst, proceeds smoothly with a series of pyrimidine compounds. The reaction gives exclusively β products when using sugars with 2'-acetate substituents whereas a mixture of α and β is observed for 2'-deoxy sugars. Good yields are observed for the 6-aza compounds as well as other pyrimidine analogues. 6-azauridine 2',3',5'-tri-*O*-benzoate **38** has been prepared in 93% yield.

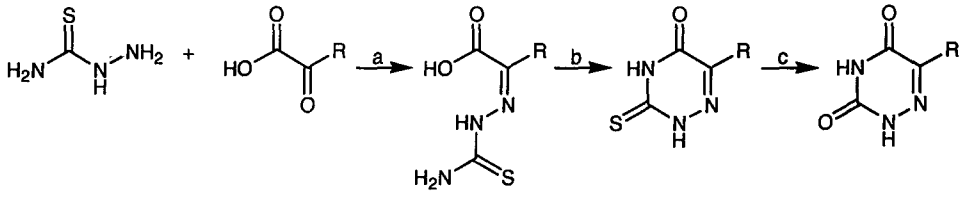


38

a.) $SnCl_4$, $ClCH_2CH_2Cl$, 4 hr, 22°C
b.) H_2O/ $NaHCO_3$

Scheme 1.2.4 Synthesis of 6-azauridine 2',3',5'-tri-O-benzoyl **38**[67]

The 6-aza moiety can be prepared from glyoxylic acids and thiosemicarbizide as shown in *Scheme 1.2.5*.[69]

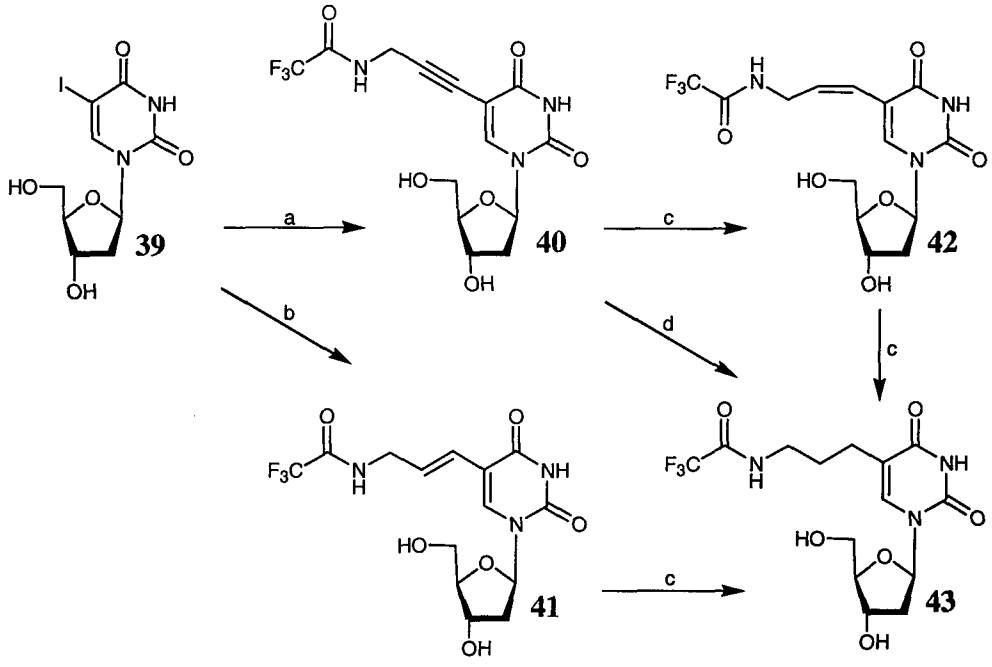


a.) H_2O hot, R = cyclopropyl, thienyl
b.) NaOH/ H_2O, Δ
c.) ClH_2CCOOH/ H_2O, Δ

Scheme 1.2.5 Preparation of a 6-aza pyrimidine moiety

5-substituted

Carbonyl substituents have been introduced into the 5-position by Stille-type reactions[70] and Sonogashira and Heck reaction. 5-Alkynyl substituted nucleosides have been synthesised as precursors for chain-terminating, fluorescence-tagged substrates for DNA polymerases.[71] The principle is the same as that of Sanger in that the chain-terminator does not carry a 3'-OH group but has the additional advantage of the fluorescence to distinguish between chain-terminating bases.[72] As well as the alkynyl substituted nucleosides the alkenyl- and the alkyl-substituted nucleosides have been synthesised.[73] The *trans*-alkenyl- and the alkynyl-substituted derivatives can be obtained by direct coupling via the Heck reaction or the Sonogashira reaction, respectively. The *cis*-alkene- and the alkyl-derivatives are accessible through reduction of the alkynyl derivative. The partial reduction of the alkyne was first carried out using Lindlar's catalyst[74] but $NiCl_2$/ $NaBH_4$[73] was found to produce more reproducible results. Further reduction using $NiCl_2$/ $NaBH_4$ yielded the fully saturated product, which can also be produced from the alkyne using H_2/ PtO_2[75] directly with improved yields.



a.) *N*-propynyltrifluoroacetamide, $Pd(PPh_3)_4$, CuI, NEt_3, DMF

b.) *N*-propenyltrifluoroacetamide, Na_2PdCl_4, NaOAc aq.

c.) $NiCl_2$/ $NaBH_4$, MeOH, -78°C

d.) H_2, PtO_2, MeOH

Scheme 1.2.6 Preparation of 5-substituted pyrimidine nucleoside analogues

A vital property of these compounds is that they are accepted by DNA/ RNA polymerases. Incorporation of 5-modified nucleosides is dependent on the flexibility, stereochemistry and the length of the linker group.[75] Nucleosides **40**, **41**, **42** and **43**, derived from 5-iodo-2'-deoxyuridine **39**, have been investigated as potential substrates for DNA and RNA polymerases. The three-carbon spacer has different oxidation states, and therefore different rigidity, and conformations. To investigate the incorporation by polymerases the triphosphate derivatives of compound **40**, **41**, **42** and **43** were prepared and the trifluoroacetate groups were removed. Additionally the carbon spacer was extended by coupling of either imidazole 4-acetic acid or urocanic acid to the triphosphates of **40**, **41** and **43**. Surprisingly, all attempts to make derivatives of the *cis*-alkene triphosphate **42** failed. Of the four different amine functionalised triphosphates of **40**, **41**, **42** and **43**, only the alkyne **40** and *trans*-alkene **41** analogues were found to be DNA polymerase substrates. Analogues of **40** and **41** with imidazole extended spacer were also found to be substrates whereas that of **43** was not.[75]

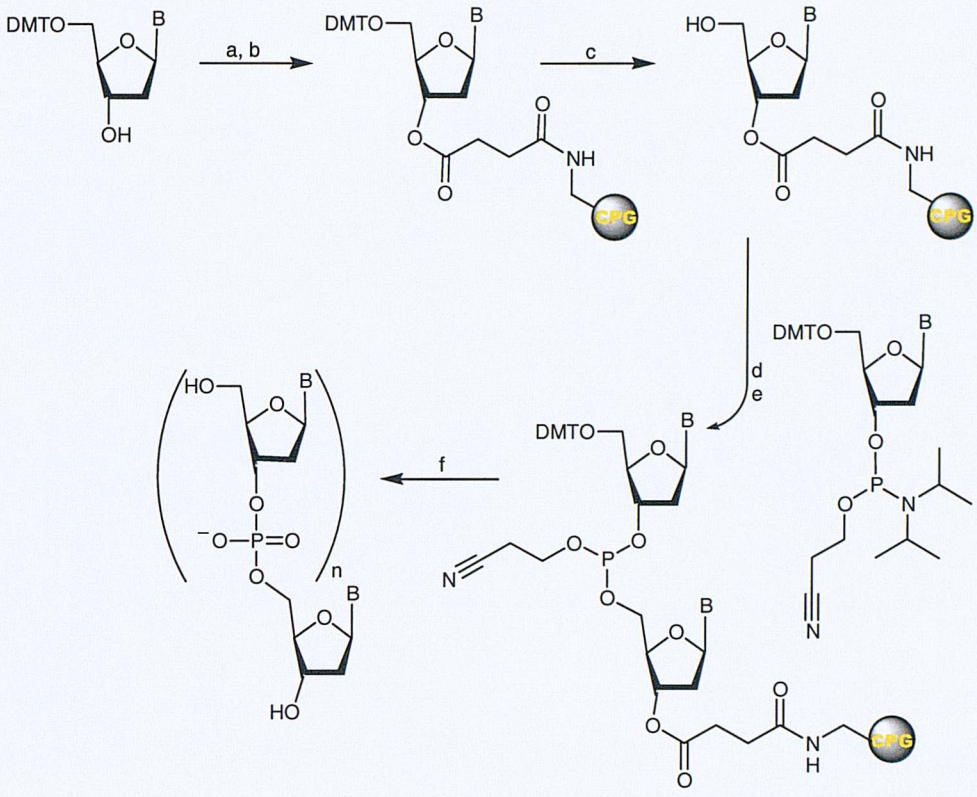
1.3 Nucleic Acids and Solid Phase Chemistry

The extent of solid phase nucleoside synthesis reported in the literature is largely restricted to that of DNA- and RNA-oligomers. Despite the great importance of nucleoside analogues and the fast developing methods of solid phase/ combinatorial chemistry few nucleosides have been synthesised by solid phase methods.

1.3.1 Oligonucleotides by Solid Phase Methods

Single stranded DNA and RNA oligomers can be synthesised chemically by the use of nucleoside phosphoramidite building blocks.[76-78] Automated solid phase synthesisers make use of 5'-protected nucleosides attached to controlled pore glass beads at the 3'-end *via* a chemically cleavable linker. A typical process for extension of one nucleotide requires the following steps:

- attaching the 5'-dimethoxytrityl protected nucleoside to the support, typically a silica support or CPG,
- removal of the 5'-protecting group (commonly dimethoxytrityl),
- condensation with the appropriately protected deoxynucleoside 3'-phosphoramidite,
- acylation or capping of unreactive deoxynucleoside,
- oxidation of the phosphite triester to the phosphate triester; and,
- removal of base and phosphate protecting groups and cleavage from the support.



a.) 3'-derivatisation; succinic anhydride, pyridine, DMAP
b.) Attachment to solid support; aminomethyl-CPG, NEt_3, HOBt/ DIC
c.) Trityl deprotection; dichloroacetic acid
d.) Condensation; phosphoramidite derivative
e.) Phosphite to phosphate oxidation; I_2
f.) Deprotection and cleavage; dichloroacetic acid then NH_3 / MeOH

Scheme 1.3.1 *Chemical synthesis of oligonucleotides by the solid phase phosphoramidite method where step c and d are repeated n times*

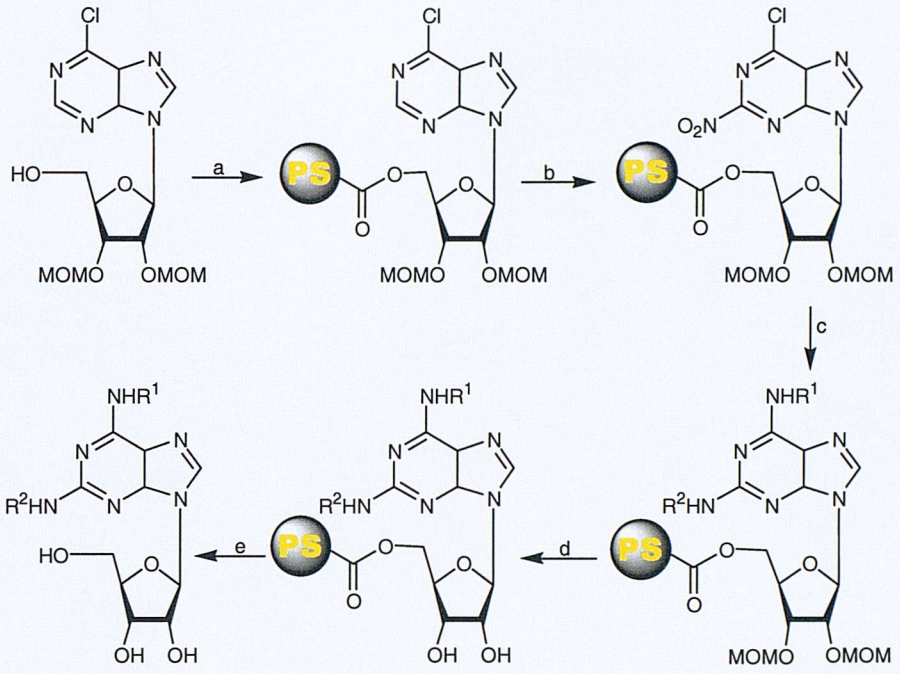
After phosphite oxidation the 5'-position can then be deprotected and the process repeated to prepare the desired oligonucleotide, *Scheme 1.3.1*. The final step is to treat the resin with ammonium hydroxide to remove protecting groups from the phosphorus and the base and to hydrolyse the ester linkage attaching the oligonucleotide to the solid support. β-Cyanoethyl diisopropylamino-phosphoramidites and iodine oxidation are usually utilised in solid phase oligonucleotide synthesis but alternative phosphorus protecting groups,[76] phosphoramidite synthons,[79] dialkylamino activating groups,[80] and oxidation

procedures[79] have also been explored. The exocyclic amino groups on the phosphoramidite building blocks are normally protected[81,82] but have also been used unprotected.[83,84] The use of unprotected nucleoside phosphoramidites is attractive as it reduces the number of steps in the synthesis but also the risks of depurination. The standard phosphoramidite procedure typically involves protection of the exocyclic amines and phosphate linkages with base labile groups. Pd(0)-removable allyl group derivatives have been successfully synthesised and incorporated by solid phase methods.[85,86]

Synthetic oligonucleotides have advanced the understanding of DNA, its structure and properties and therefore also many biochemical processes. The ease of synthesis has made it possible to provide many synthetic DNA oligomers that have been used to explore genes and gene control by mutagenesis and hence produced changes in protein structures, the dideoxy sequencing method and labelling to locate unique DNA sequences.[87] Backbone, sugar and nucleobase modifications have been achieved by synthetic methods and found to stabilise/ destabilise the duplex structure. The sugar backbone has been replaced with peptides to obtain so-called PNA or peptide nucleic acids. Peptide nucleic acids, which can easily be prepared by solid phase methods, bind with high affinity to complementary nucleic acids.[88]

1.3.2 Solid Supported Nucleoside Modifications

Today's literature contains only a few examples of nucleosides that have been prepared by solid phase methods although many modified nucleosides have been used in the solid phase preparation of nucleic acids, section *1.3.1.* One such example is the preparation of disubstituted ribonucleosides, *Scheme 1.3.2.*[89] Carboxypolystyrene was the solid support of choice as alcohols are easily coupled to this resin by standard esterification methods. Nitration of the nucleobase at the 2-position was accomplished by treatment with tetrabutylammonium nitrate (TBAN) and trifluoroacetic acid anhydride (TFAA).

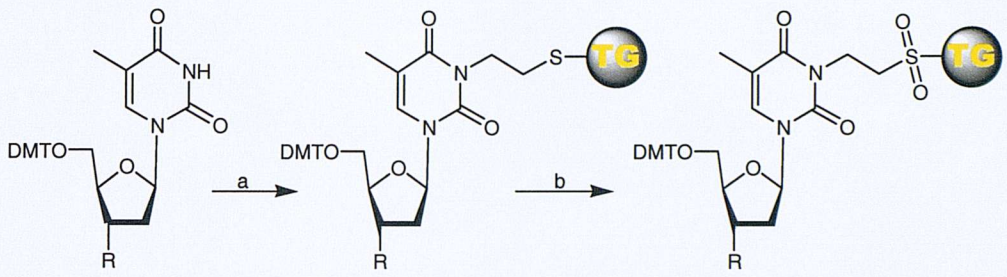


a.) Carboxypolystyrene/ DIC/ DMAP, DCM (MOM = methoxymethylidene protecting group)
b.) TBAN-TFAA, DCM
c.) *Chloro displacement*: R^1-NH_2, DIPEA, DCM; *Nitro displacement*: R^2-NH_2, DIPEA, NMP, 80-90ºC
d.) *p*TsOH·H_2O, DCM-MeOH (97:3)
e.) NaOMe, MeOH, THF

Scheme 1.3.2 Solid phase synthesis of 2,6-disubstituted ribonucleosides

Chloro displacement by aliphatic or aromatic amines in the presence of DIPEA occurred at room temperature without affecting the nitro group. Nitro group displacement was achieved using a variety of amine nucleophiles at elevated temperature. Deprotection of the hydroxyl groups followed by cleavage from the resin afforded the 2,6-disubstituted ribonucleoside.

Thymidine derivatives attached to the solid support through the base have also been reported, *Scheme 1.3.3*.[90] Thus, thymidine derivatives **44** and **45** were loaded onto β-hydroxyethylthioether TentaGel resin under Mitsunobu conditions. During this synthesis the 3'- and 5'-positions of the sugar moiety are both available for modification.



44 R = N_3

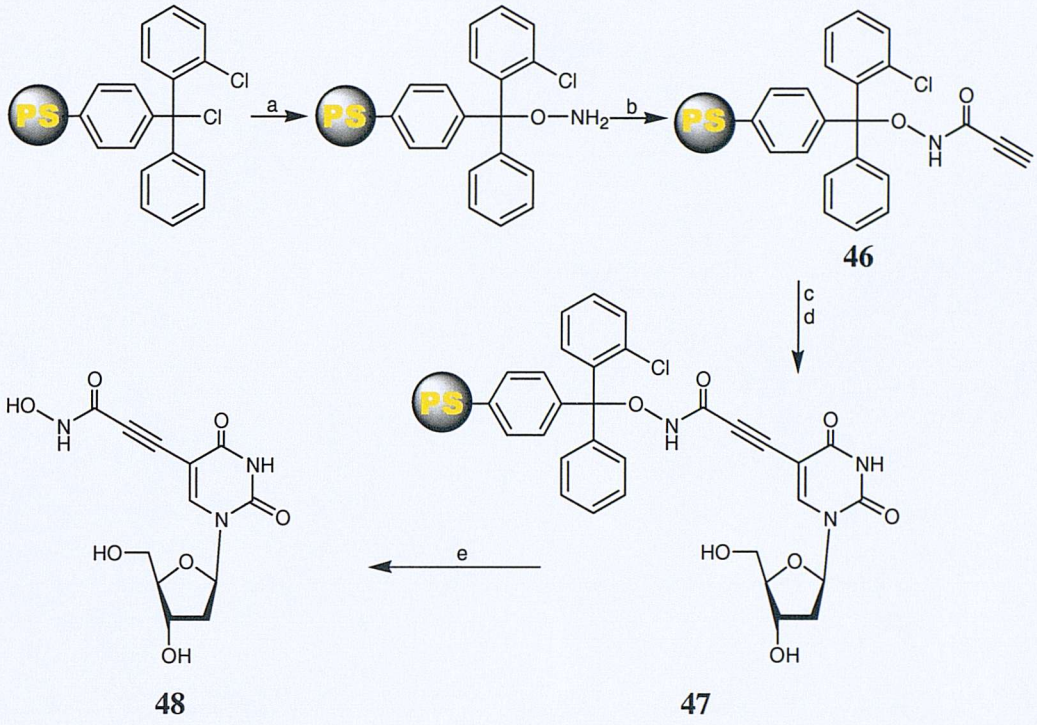
45 R = OTBDMS

a.) PPh_3, DEAD, THF/ DCM
b.) *m*CPBA, DCM

Scheme 1.3.3 *Preparation of Thymidine based supports*

Thus, dinucleotides can be prepared by phosphoramidite synthesis and the 3'-azido group can be reduced and extended by peptide coupling. Efficient release of the nucleosidic material *via* β-elimination could be achieved by alkaline treatment on the alkylthioethyl function following oxidation to the sulfone.

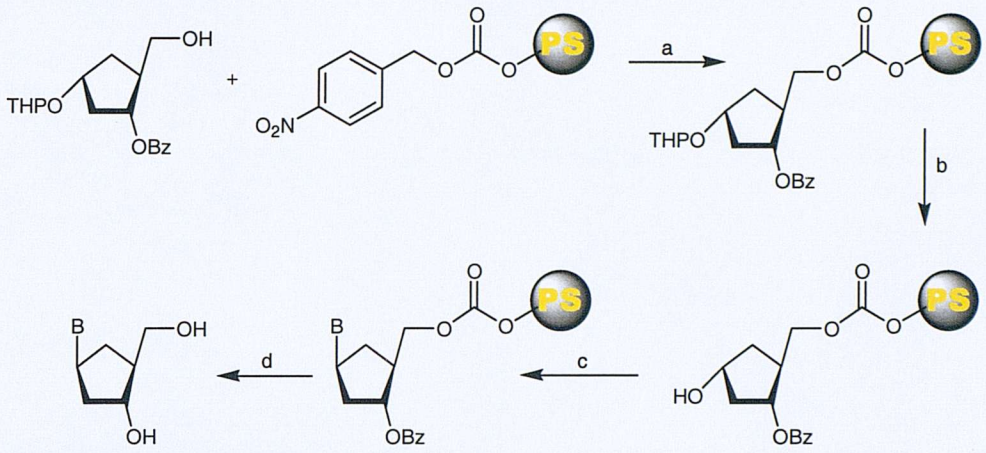
Grinstaff and co-workers developed a facile solid phase strategy for the preparation of nucleoside hydroxamic acids, *Scheme 1.3.4*.[91] This synthesis was developed by first introducing the new functionality on the support and subsequent introduction of the nucleoside. Solid phase *O*-linked hydroxylamine was treated with standard amide bond forming coupling conditions, DCC/ HOBt and a carboxylic acid. In this case propiolic acid was used to allow the subsequent carbon-carbon bond forming reaction with the nucleoside. The solid supported alkyne **46** and 3',5'-dibenzoyloxy-2'-deoxy-5-iodouridine could then be treated with $Pd(PPh_3)_4$/ CuI to provide the solid supported nucleoside analogue **48** after cleavage from the resin.



a.) N-hydroxyphthalimide, NEt_3, DMF; then, $H_2NNH_2.H_2O$, THF
b.) DCC, HOBt, DIPEA, propiolic acid, -10°C
c.) 5-iodo-3'-5'-*O*-benzoyl-2'-deoxyuridine, $Pd(PPh_3)_4$, CuI, DMF, NEt_3
d.) MeOH-NH_3, 25ºC
e.) 5% TFA in DCM

Scheme 1.3.4 Solid Phase preparation of nucleoside hydroxamic acids

In addition to the synthesis of classical nucleosides, solid phase synthesis of carbocyclic nucleosides has been reported in the literature. Thus, pyrimidine and purine carbocyclic nucleoside derivatives have been prepared by solid phase synthesis, *Scheme 1.3.5*.[92] The carbocyclic moiety was loaded onto a *p*-nitrobenzyl carbonate resin *via* its free 5'-hydroxyl. Subsequent deprotection of the 1'-hydroxyl allowed coupling to a purine or pyrimidine moiety under Mitsunobu conditions and carbocyclic nucleosides were obtained by treatment with K_2CO_3.



a.) DMAP, DIPEA, CH_2Cl_2, 40ºC
b.) PPTS, 1-butanol/ 1,2-dichloroethane, 60ºC
c.) B-H, DEAD, PPh_3, DMF, (B = protected Uracil, Thymine, Cytosine, Adenine or Guanine)
d.) K_2CO_3, THF/ MeOH

Scheme 1.3.5 Solid phase synthesis of carbocyclic nucleosides

Realising the value of the modified nucleosides and the advantages of solid phase chemistry it was an attractive proposition to combine the two to allow the preparation of libraries of nucleoside analogues.

2 Preparation of a Novel Nucleoside Based Resin

The purpose of this chapter is to give a detailed description of the preparation of a Thymidine-linked resin, which forms the starting point for most of the syntheses relevant to this project.

2.1 Introduction

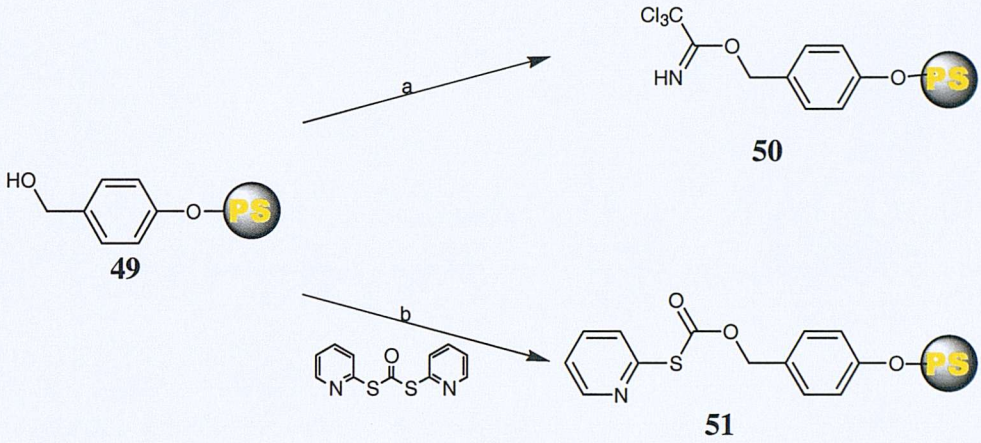
Solid phase chemistry methodologies have long been essential to the rapid assembly of synthetic long chain peptides and have more recently led to the synthesis of huge libraries of diverse small organic molecules.[93] The choice of solid support and compound attachment strategies are important and have to be established for each specific synthesis, with issues such as synthesis orthogonality and analysis having to be carefully considered.

2.1.1 Attachment of Alcohols to Solid Supports

For the purposes of this project, a nucleoside was to be immobilised on a solid support to act as a scaffold for further modification. As modification of the heterocyclic base was the aim, attachment of the nucleoside via the hydroxyl groups of the sugar seemed an attractive strategy.

Alcohols have been loaded onto solid supports by various methods, such as esterification, silylation and etherification;[94] diols have also been immobilised by ketal linkages.[95] Derivatisation of the 3'-position by *esterification* is the most common method of attaching a nucleoside to the solid support and is used in the chemical synthesis of DNA.[96] Esterification of the 3'-hydroxyl is carried out using succinic anhydride on a 5'-protected nucleoside. Attachment to aminomethyl resin through the 3'-succinyl ester is then normally effected using HOBt/ DIC.[76,97] *Silyl linkers,* which can be formed from a range of hydroxy nucleophiles including sugars and hindered 3° alcohols,[98] provide an alternative to acid- or base-mediated cleavage owing to the ability to cleave by fluoridolysis.[99-101]

Immobilisation of alcohols by *etherification* has been achieved by trityl- or benzyl-ether formation.[94] Activated forms of the well-known Wang-type linkers allow immobilisation of alcohols.



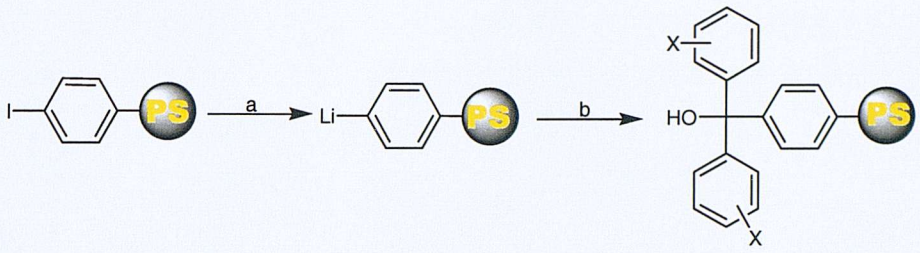
a.) Cl_3CCN, DBU, 0°C, 20-40 min
b.) Et_3N, CH_2Cl_2, di-(2-pyridyl)-thiocarbonate, CH_2Cl_2, 24 hrs

Scheme 2.1.1 Modified Wang linkers for immobilisation of alcohols

Preparation of the Wang trichloroacetimidate resin[102,103] can be achieved by addition of DBU to a cooled suspension of **49** and trichloroacetonitrile. The Wang TOPCAT resin **51**[104] can be prepared by the addition of di-(2-pyridyl)-thiocarbonate to a suspension of resin **49** and alcohols can be immobilised onto either resin using Lewis acid catalysis under anhydrous conditions.

2.1.2 Trityl Functionalised Solid Supports

The trityl group is very useful as a solid supported protecting group for a wide range of functionalities especially alcohols and acids. Polymer-bound trityl chloride displays comparable advantages to its solution phase analogue, namely its reactivity and selectivity for primary *vs.* secondary or tertiary hydroxyl groups and the ease of cleavage of the subsequent trityl ethers under acidic conditions.[105-108] Two basic schemes, *Scheme 2.1.2*[109] and *Scheme 2.1.3*,[110] have been used to introduce the integral trityl linker onto polystyrene resin.

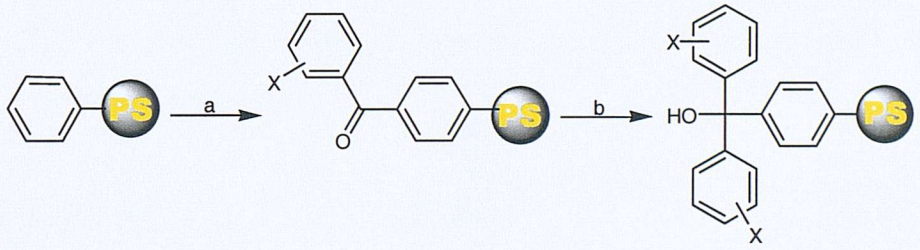


52

a.) *n*-BuLi, benzene
b.) $XH_4C_6COC_6H_4X$ (X = OCH_3 or H)

Scheme 2.1.2 *Preparation of trityl resin by lithiation followed by treatment with a benzophenone*

Treatment of iodo-polystyrene with *n*-butyl lithium in benzene yields the lithiated polystyrene. Subsequent addition of a benzophenone gives the polymer-bound trityl alcohol **52**.



a.) $AlCl_3$, XH_4C_6COCl (X = CH_3, OCH_3 or H)
b.) XH_4C_6MgBr (X = CH_3, OCH_3 or H)

52

Scheme 2.1.3 *Preparation of trityl resin by Friedel Crafts acylation and subsequent treatment with a Grignard reagent*

The polymer bound trityl alcohol **52** is also accessible by subjecting unfunctionalised polystyrene to Friedel-Crafts acylation followed by treatment with aryl magnesium bromide.

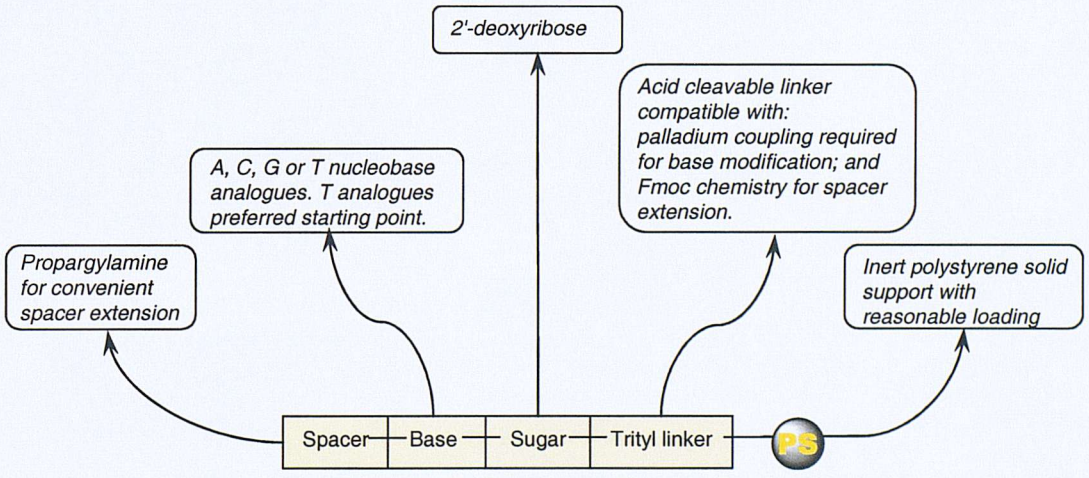
Conversion of trityl alcohol **52** to a trityl chloride is normally achieved by refluxing with acetyl chloride in benzene or thionyl chloride in CH_2Cl_2.[111]

Loading of these resins with alcohols or carboxylic acids is carried out with an amine base and cleavage is realised by acid treatment. Trityl linkers are acid labile, the degree of acid lability depending upon the aryl substituents hence 4-methoxytrityl is more acid labile than 2-chlorotrityl.[112,113]

2.2 Initial Loading Attempts and Cleavage Studies

2.2.1 Synthetic Strategy

For this project a trityl functionalised polystyrene support was the preferred choice and, since the aim was to develop a technique for the swift synthesis and evaluation of a large number of targets for DNA labelling, 2'-deoxy nucleoside analogues were selected for attachment. The most straightforward nucleosides to handle are thymidine and uridine analogues as they, generally, do not need protection. From earlier work, as described in section *1.2.5.2*,[75] it has been shown that pyrimidines with 5-substituents are well accommodated in the DNA double helix, vital for efficient incorporation of the target into a growing DNA strand in sequencing reactions.



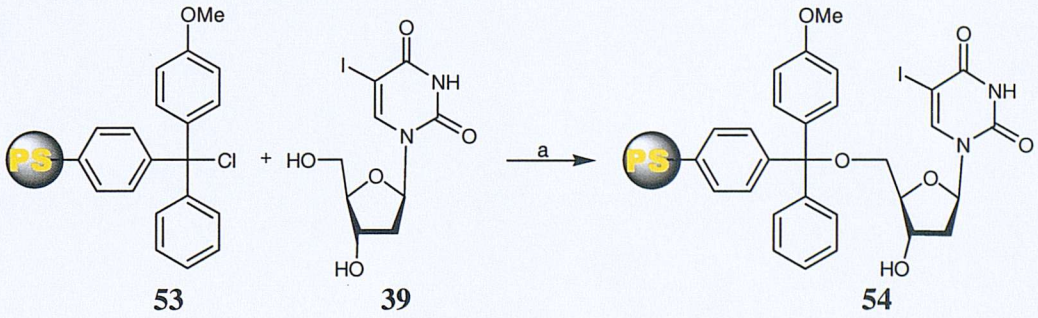
Scheme 2.2.1 *Illustration outlining the basic components of the synthetic strategy of the novel nucleoside resin*

2.2.2 Loading of the Nucleoside

5-Iodo-2'-deoxyuridine **39** appeared to be a good starting point for synthesis providing the required sugar and base components where the iodide on the 5-position of the base would allow simple modification by palladium chemistry to be developed.

Loading materials on, and cleavage from, a solid support are fundamental for any solid phase synthesis. Preparation of a nucleoside based resin with facile methods for loading and cleavage were therefore considered to be critical and methods for optimum loading and cleavage were explored. First and foremost, coupling of nucleoside **39** using anhydrous pyridine was investigated.[111] 5-Iodo-2'-deoxyuridine **39** in dry pyridine was added to resin **53**, and the coupling analysed by cleavage from the resin followed by HPLC analysis of the residue. However, no evidence of coupling was detected.

As early as 1968 Cramer and Köster[110] described the coupling of a nucleoside to a trityl resin. Their method was similar to that described above for simple alcohols but carried out at a higher temperature.



53 **39** **54**

a.) Dry pyridine, N_2, 70°C, 48 hrs

Scheme 2.2.2 *Loading of nucleoside* **39** *onto trityl polystyrene resin*

This was repeated with **39** at 70°C and after cleavage with 4M HCl in dioxane, a loading of 0.01 mmol/ g was determined. Such a low loading was not suitable for further synthesis and thus alternative methods were sought.

2.3 Modified Successful Loading Procedure

It seemed sensible to try to improve the loading using dry pyridine as solvent since this reaction worked at elevated temperatures, although not very efficiently. A catalytic amount of DMAP[114,115] was tried and encouragingly gave a much more efficient coupling, showing a greater than twentyfive-fold increase in loading to 0.271 mmol/ g (corresponding to a 51% yield).

With an increased loading achieved, cleavage reactions were investigated. 4M HCl in dioxane produced similar results to that of 5% TFA in CH_2Cl_2 using triisopropyl silane (TIS) as a scavenger and it was thought that a trace amount of moisture present was acting as a scavenger in the case of HCl/ dioxane. Hence, it was thought that water or methanol could be used in conjunction with TFA instead of TIS although methanol would be preferred as it could be more easily removed under reduced pressure.

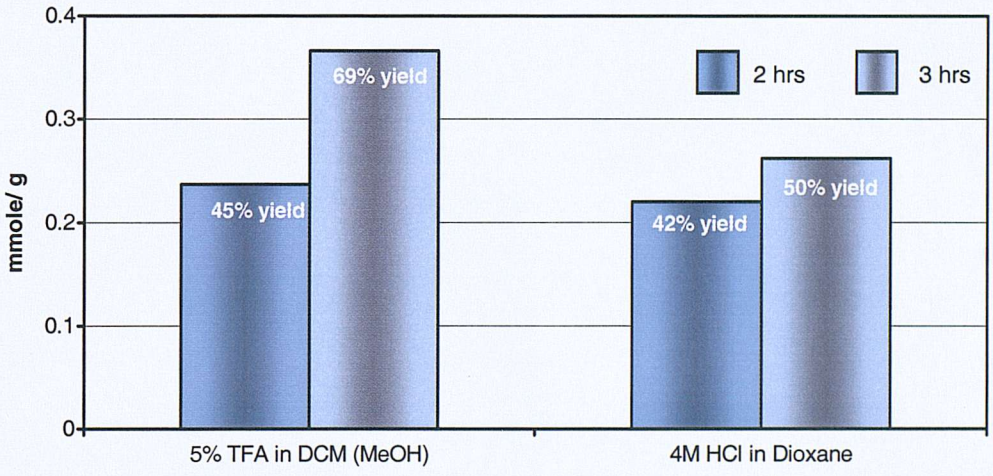


Figure 2.3.1 *Graphical representation of the cleavage of resin* **54**. *A comparison between TFA and HCl as acid cleavage medium.*

Methanol worked very well as a scavenger of the trityl cation with 5% TFA in CH_2Cl_2 (MeOH) being marginally better than 4 M HCl in dioxane. For practical reasons (storage) 5% TFA in CH_2Cl_2 (MeOH) was the preferred cleavage condition. It was later also realised that repeated acid treatments were preferred over longer exposure times. Shorter times of exposure of nucleosides to acidic

conditions were preferred since the conformation at C1' was prone to anomerisation. Thus, typically the resin was subjected three to five times to 2 minute treatments with 5% TFA in CH_2Cl_2 and a drop of methanol.

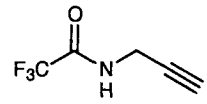
2.4 Nucleoside Modification

2.4.1 Introduction

Having achieved an acceptable loading of 5-iodo-2'-deoxyuridine **39** onto a solid support the nucleoside was ready for modification. 5-Alkynyl-modified nucleic acid monomers are used as chain terminating tags but have other uses such as a replacement for the natural bases to alter the properties of DNA or RNA. 5-Propargylamino substituted uridines have already been shown to stabilise DNA duplexes and triplexes whereas aliphatic substituents lead to duplex destabilisation.[116-119] The amino function on the 5-propargylamino nucleoside analogue also provides a functionality available for modification. A 5'-*O*-attached-2'-deoxyuridine resin that possesses the propargylamino handle on position 5 of the uracil moiety would provide a very useful tool for a facile synthesis of novel nucleosides. Modifications at the 5-position have, by solution phase palladium-catalysed coupling, provided nucleoside analogues such as those described in section *1.2.5.2.* Numerous carbon-carbon bond forming reactions with acetylenes and aromatic iodo compounds have successfully been carried out on the solid phase[120-134] making it promising to carry out palladium-catalysed modification at the 5-position on the solid phase.

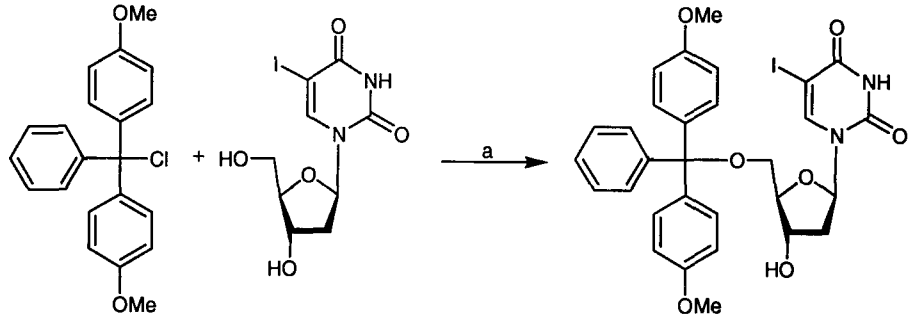
2.4.2 Synthesis

Propargylamine was protected as described in the literature using ethyl trifluoroacetate in anhydrous methanol providing *N*-trifluoroacetylpropargylamine **55** in 61% yield.[116]



55

The 5'-hydroxyl group on 5-iodo-2'-deoxyuridine **39** was protected by the addition of 4,4'-dimethoxytrityl chloride **56** to **39**, *Scheme 2.4.1.*

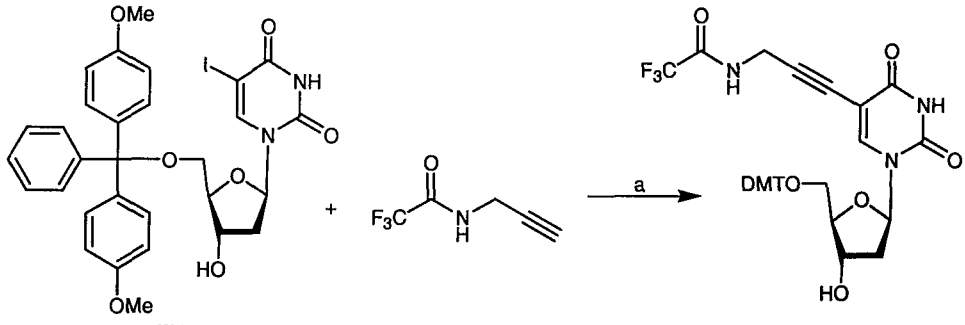


56 **39** **57**

a.) dry NEt_3, dry Pyridine

Scheme 2.4.1 *Preparation of 5-iodo-5'-O-(4,4'-dimethoxytrityl)-2'-deoxyuridine* **57**

Exposure of *N*-trifluoroacetylpropargylamine **55** to 5-iodo-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuridine **57** under $Pd(PPh_3)_4$ and CuI catalysis in DMF and NEt_3 afforded nucleoside **58** in 71% yield.



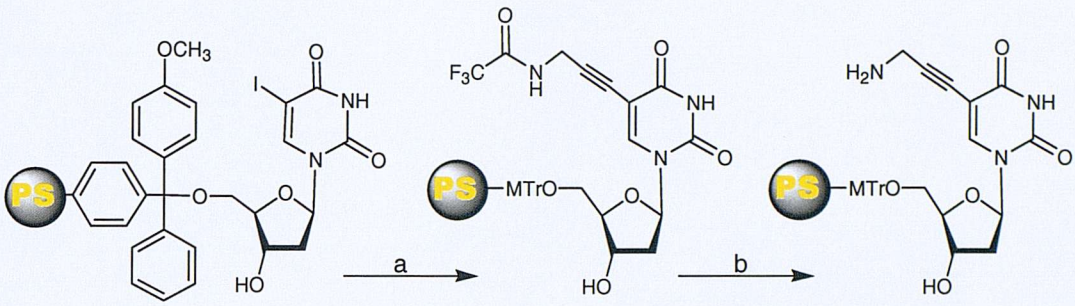
57 55 58

a.) $Pd(PPh_3)_4$, CuI, NEt_3, DMF

Scheme 2.4.2 *Preparation of 5-N-trifluoroacetylpropargylamino-5'-O-(4,4'-dimethoxytrityl)-2'-deoxyuridine* **58**

Before attempting the solid phase Sonogashira reaction with **54** the reaction was assessed using a solid phase bound 3-iodobenzyl alcohol. Solid phase synthesis made it possible to remove excess reagents by simply washing the resin repeatedly. The trial reaction was validated by HPLC and mass spectrometry before attempting the Pd(0)/ CuI catalysed reaction on nucleoside **54** and the coupling between on-bead iodo-benzyl-alcohol with the terminal acetylene **55** proceeded smoothly. Analysis suggested quantitative conversion with complete disappearance of the starting material. Hydrolysis of the trifluoroacetate group with 1M aqueous KOH in dioxane gave the free amine.

With these promising results in hand the reaction was carried out using nucleoside resin **54**.



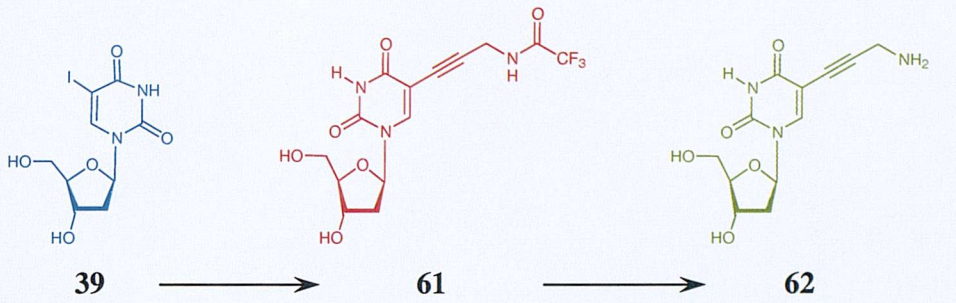
54 **59** **60**

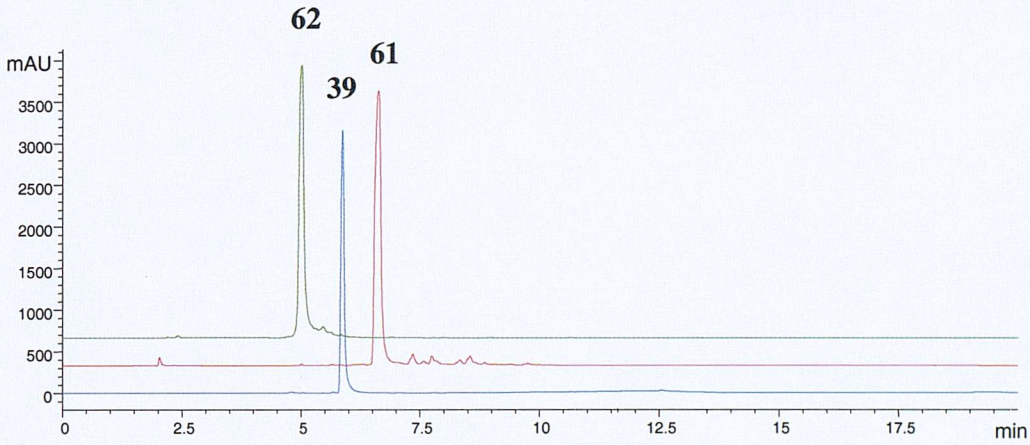
a.) *N*-trifluoroacetylpropargylamine **55,** $Pd(PPh_3)_4$, CuI, NEt_3, DMF

b.) 1M KOH aq./ Dioxane (1:2)

Scheme 2.4.3 *Preparation of 5-N-trifluoroacetylpropargylamino-5'-O-(4-methoxytrityl polystyrene)-2'-deoxyuridine* **59**

The Sonogashira reaction proceeded smoothly on resin **54** by the addition of CuI, *N*-trifluoroacetylpropargylamine **55,** NEt_3, and $Pd(PPh_3)_4$ at 60°C, and allowed the complete conversion to **59**. Hydrolysis, to provide the free amine, by agitating with 1M aqueous KOH in dioxane proceeded cleanly to give, after cleavage from the solid support **62**, *Scheme 2.4.4*.





Scheme 2.4.4 *HPLC traces of compounds cleaved off the resin during preparation of thymidine resin* **59** *and* **60**

For parallel library synthesis, a new batch of loaded resin **59** was prepared. However, the reaction progress was not as smooth as previously observed and *Figure 2.4.1* shows the HPLC analysis of an identical palladium coupling reaction carried out on the old and the new resin in parallel.

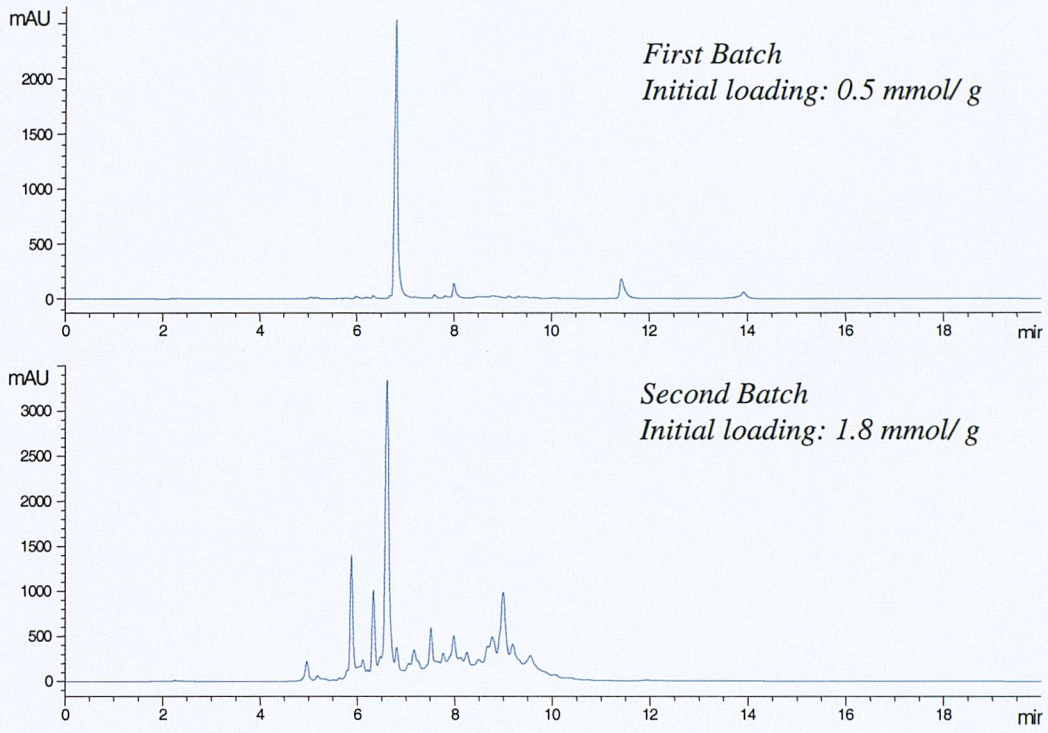
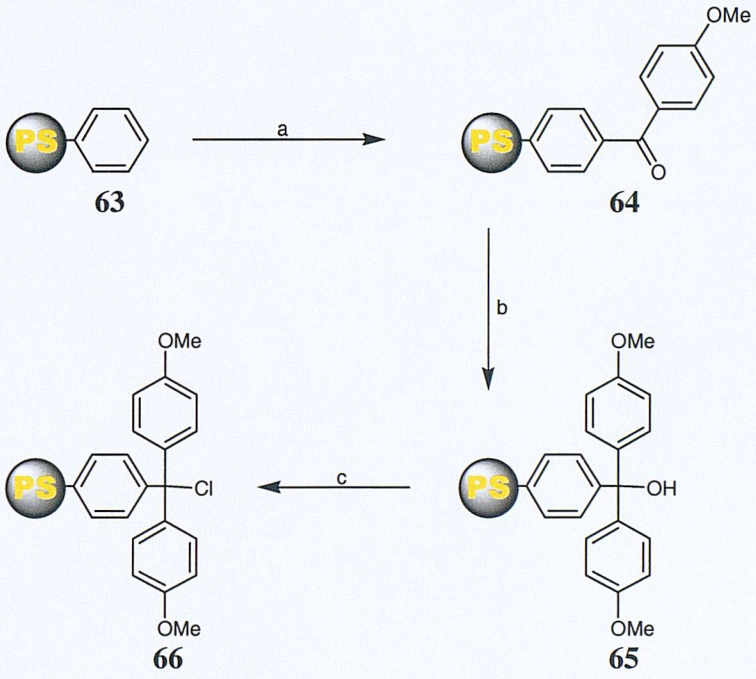


Figure 2.4.1 *HPLC analysis of the initial Sonogashira product* **61** *vs. that using the higher loaded resin.*

The most obvious difference between the two resins was the large deviation in initial loading, 1.8 mmol/ g *vs.* 0.5 mmol/ g. A notable physical difference was that the resin with the lower loading floated in CH_2Cl_2 whereas the resin with the higher loading did not. In order to obtain consistency the resin was prepared in-house.

2.5 Resin Functionalisation

Commercially available resin **63** from styrene and 1% divinylbenzene was subjected to a Friedel Crafts reaction with *p*-methoxybenzoyl chloride. The resulting ketone **64** was subjected to a Grignard reaction with *p*-methoxyphenylmagnesium bromide. Resin **65** was a light yellow colour which turned deep red upon treatment with acid. Conversion of the hydroxyl trityl to the trityl chloride according to published procedures using $SOCl_2$ afforded resin **66**.



a.) *p*-Anisoyl chloride, $AlCl_3$, CS_2, Δ
b.) 4-methoxyphenylmagnesium bromide, dry THF
c.) $SOCl_2$, pyridine, dry CH_2Cl_2

Scheme 2.5.1 Preparation of Polystyrene trityl chloride resin **66**

The loading determined, by loading/ cleavage of 5-iodo-2'-deoxyuridine and HPLC analysis, corresponded to 0.125 mmol/ g.

2.6 Spacer Extension and Library Synthesis

The extensive need for modified nucleotides and nucleosides in the areas of DNA sequencing and therapeutic agents prompted the synthesis of nucleosidic resin **60**. The idea was that this resin would provide a starting material for syntheses of large numbers of 5-modified nucleosides in a combinatorial fashion. To prove that this was an attainable objective a smaller trial library was synthesised.

2.6.1 Library Design

A simple trial library was designed. In order to evaluate the mobility/ rigidity of potential spacer units, six unnatural amino acids both cyclic, acyclic and aromatic, *Table 2.6.1*, possessing to side chain functionality to complicate synthesis, were chosen.

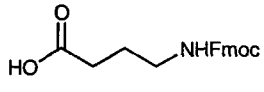
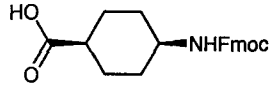
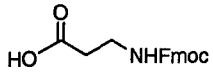
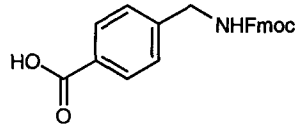
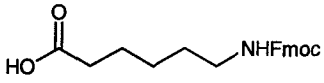
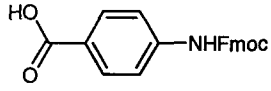
 **67**	 **70**
 **68**	 **71**
 **69**	 **72**

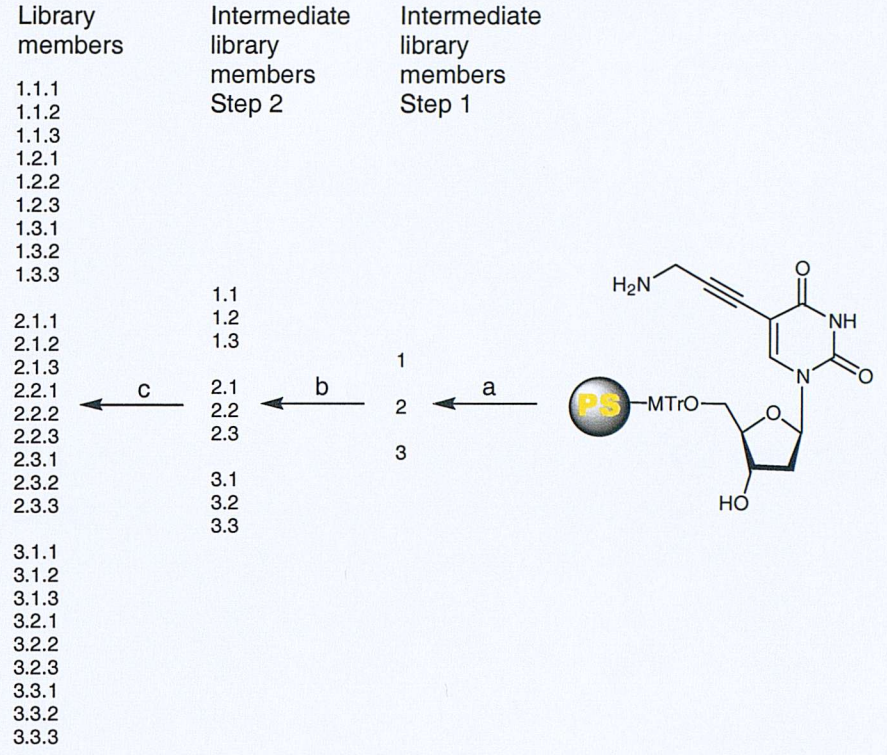
Table 2.6.1 *Structures of spacers chosen for synthesis of the trial library*

The acyclic monomers were to be coupled to the resin first, each batch would then be split in three and then the cyclic monomers would be coupled making a total of nine members. Each batch of resin would again be split in three and three different dyes would be coupled to the end of the spacer making 27 members.

2.6.2 Library synthesis

Fmoc protection of the cyclic non-aromatic amino acid required modified procedures as described in the literature.[135] *cis*-4-Amino-1-cyclohexanecarboxylic acid (ACCA) was Fmoc protected using Fmoc-succinimide and gave **70** in 83% yield. Protection of 4-aminomethylbenzoic acid was accomplished using Fmoc chloroformate in dioxane and aqueous sodium bicarbonate according to a procedure described in the literature[136] and the product **71** was obtained in 79% yield. Protection of *p*-aminobenzoic acid was effected in dioxane/ aqueous sodium bicarbonate using 9-fluorenylmethyl chloroformate according to a literature procedure.[137] The Fmoc protected aminobenzoic acid **72** was obtained in a much lower yield (36%) than the other spacers. This is presumably due to the lower reactivity of the aniline. All other spacer moieties were commercially available.

The library was assembled using standard solid phase peptide synthesis methodology. Equimolar solutions of reactants were prepared by the addition of DIC to a solution of HOBt and Fmoc-amino acid in CH_2Cl_2 / DMF (6:1). Resin **60** was split into three and reacted with each of the three reaction solutions, each containing **67**, **68** or **69** respectively. All resins were subjected to 20% piperidine in DMF to remove the Fmoc protecting group. Each batch was then split into three vessels again and the procedure was repeated using the cyclic spacers **70**, **71** and **72**. Each step was monitored by HPLC and mass spectroscopy and all steps proceeded smoothly. The resin, now in nine vessels, was again each split into three for the coupling of the dye. Three different dyes, fluorescein, rhodamine, and erythrosin that possessed an isothiocyanate functional group for attachment to amines were chosen for the labelling of the library. The resins were treated with solutions of the dye in DMF, in the presence of triethylamine.



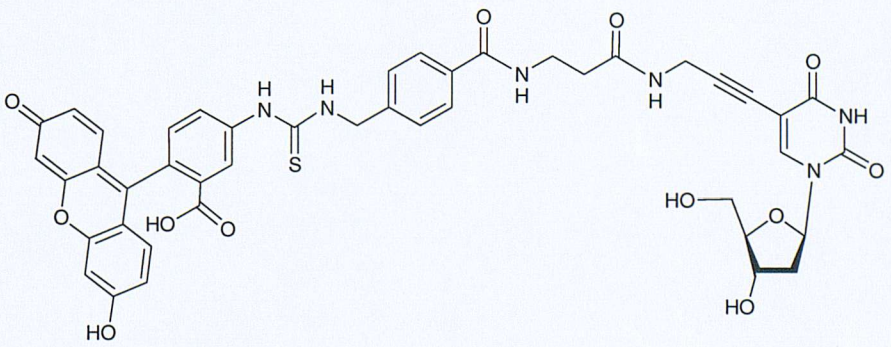
a.) HOBt/ DIC and **67**(1), **68**(2) or **69**(3), respectively; then 20% piperidine in DMF

b.) HOBt/ DIC and **70**(1), **71**(2) or **72**(3), respectively; then 20% piperidine in DMF

c.) NEt_3, FITC (1), RITC (2) or EITC (3), respectively

Scheme 2.6.1 *27 Membered Library Synthesis*

Unfortunately the coupling of the dyes was not totally successful. The rhodamine dye coupling resulted in clean HPLC traces, though it proved difficult to obtain adequate mass spectral analysis for these compounds, and fluorescein and erythrosine couplings resulted in mixtures of compounds being formed. The synthesis of one member (shown in *Scheme 2.6.1* as compound 2.2.1) of the library, compound **73,** was scaled up to allow full characterisation.



73

As determined from weight of resin and the mass of compound cleaved from the solid support compound **73** was obtained in quantitative yield and high purity (95%) as determined by HPLC (ELS, UV = 282 nm).

2.7 Conclusion

The loading and cleavage of 5-iodo-2'-deoxyuridine onto a trityl chloride polystyrene resin was successfully performed. The loading level of the initial resin was found to be an important factor as resins with high loading (1.8 mmol/ g) complicated further synthetic modifications to the resin bound nucleoside. Attachment of propargylamine to the base of the resin bound nucleoside was successfully achieved under Sonogashira conditions and provided a resin with a convenient handle for further modification.

Synthesis of a small library has shown that nucleoside-resin **60** can be used to synthesise 5-modified thymidine analogues. Although the coupling of the dye was not completely successful in all cases the repeat synthesis suggested that this was probably due to the batch of dye used and hence does not appear to be a general problem.

Using this resin the generation of a nucleosidic library is shown to be a fast and simple procedure.

3 Synthesis of Reporter Nucleoside Analogues

The purpose of this chapter is to describe the synthesis of a 57 membered library of derivatised nucleosides. Syntheses were accomplished using the nucleoside based resin developed in Chapter 2. Enzymatic screening of the library with proteases will be discussed in detail.

3.1 Introduction

Resins such as that developed in chapter 2 have potential uses in the generation of nucleosidic drugs and in the area of DNA sequencing reagents. New methods of DNA sequencing are constantly being explored and subsequently new modified nucleoside triphosphates need to be developed to fit the requirements of these methods. A nucleoside resin provides an easy route to libraries of nucleosides that could significantly shorten the route to compounds with the necessary properties. The work described in this chapter was aimed toward the synthesis of reporter nucleotides required for a potential new DNA single molecule sequencing-by-synthesis methodology. Solid phase combinatorial synthesis techniques were utilised to prepare target compounds necessary to develop this sequencing method.

3.1.1 Four Colour DNA Sequencing

The current state of the art technology for high-throughput DNA sequencing, such as that used in the Human Genome Project,[138] is capillary array DNA sequencing using laser-induced fluorescence detection.[139-142] However, the basis for the technique is still that developed by Sanger *et al.*[40] The new concept of "single-lane" fluorescence detection was applied to the enzymatic Sanger-concept in 1986,[143] where the use of hazardous and costly radioisotopes was replaced by fluorophores for detection. Fluorophores have since become an important and powerful tool in DNA detection protocols.[144-146] The use of fluorescence was extended to four distinguishable fluorophores where each base was represented by one of the four fluorophores, as discussed in Chapter 1.

Automated four-colour DNA sequencing with fluorescent primers or terminators are the most commonly used methods in high-throughput DNA sequencing laboratories today. Sanger sequencing is performed as a set of four base specific reactions, the reaction products are combined and electrophoresed down a single polyacrylamide gel tube. The separated fluorescent bands of DNA are detected as they elute at the bottom of the tube as one specific colour per base. Two variations of automated DNA sequencing have been developed: dye-labelled primer sequencing, in which the fluorescent dyes are attached to the 5' end of the primer oligonucleotide, and dye-labelled terminator sequencing, in which the dyes are attached to the terminating dideoxynucleoside triphosphates.

Fluorescence energy transfer primers (ET primers),[147] see *Figure 3.1.1*, and fluorescent chain terminators are constantly being developed to improve their spectral and chemical properties such as those listed in section *1.2.3.1.*

3.1.1.1 Energy Transfer Primers

Donor and acceptor fluorophores may be incorporated into the primer by the use of two modified nucleotides in the primer synthesis[44] or as a tail at the start of the primer sequence, *Figure 3.1.1.*[148]

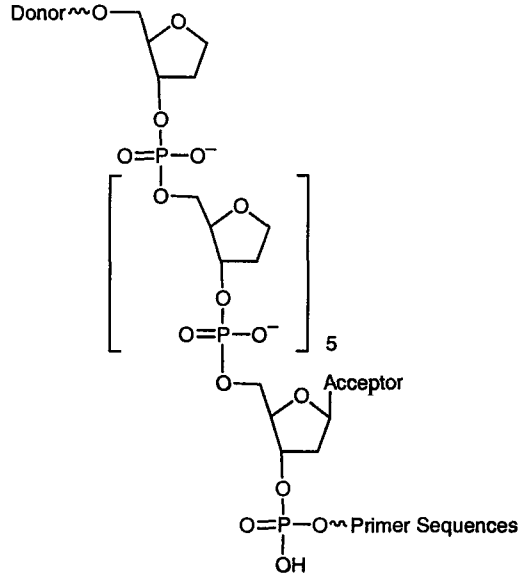


Figure 3.1.1 Design of ET cassette-labelled primers

These universal primers, called ET cassette-labelled primers, provide a general approach for labelling oligonucleotide primers of all sequences. This technology addresses the throughput and read length requirements of large-scale DNA-sequencing projects but there is still a need for improvement when, for example, band compressions cause inaccuracy. Band compressions arise for sequences with GC-rich regions and are due to non-Watson-Crick type G-A base pairing and have been improved upon by replacing adenosine and guanine with their 7-deaza-analogues. The major disadvantage with the use of dye-labelled primers is the requirement of four different primers and the necessary four extension reactions.

3.1.1.2 Fluorescent Chain Terminators

Chain terminators are commonly used in DNA sequencing. Fluorescent chain terminators work on the same principle as normal terminators, as described in Chapter 1 see *Figure 1.2.3*, *i.e.* they lack the 3'-hydroxyl group necessary for extension. The major advantages of dye-labelled terminators are:

- only one extension reaction is required for each template,
- the synthesis of a labelled primer is unnecessary hence allowing all preferred hybridisation sites to be used; and,
- false terminations, where fragments are terminated by a deoxy-nucleotide rather than a dideoxynucleotide, are not observed as these products are unlabelled.

The major disadvantage of dye-labelled terminators is that the terminations are uneven compared to using dye-labelled primers. Very small or very large signal peaks can complicate the base sequence reading.

3.1.2 Alternative DNA Sequencing Methods

DNA sequencing by electrophoresis continuously receives attention directed toward improvements such as ET primers, new terminators and technical equipment.[152,153] Simultaneously alternative methods of DNA sequencing are being developed. Alternative methods that have been exploited for DNA sequencing include sequencing by mass spectrometry, sequencing by hybridisation and sequencing by synthesis.

3.1.2.1 DNA Sequencing by Mass Spectrometry

The recent method of sequencing by *mass spectrometry*[150] makes use of the high affinity between the small biotin molecule and the protein streptavidin to isolate pure DNA-sequencing fragments. The fragments are generated by enzymatic methods and are terminated by a 5-biotin-modified nucleotide. Use of biotinylated termination molecules overcomes the problem of falsely stopped sequences. False stops are generated when the sequence terminates by a deoxynucleotide rather than a dideoxynucleotide. Although by this method falsely stopped sequences will still be generated, they are effectively removed by the streptavidin capture/ release procedure. The mass difference between each peak in the mass spectrum determines the identity of the nucleotide by which this fragment is longer than the previous. Band compression, due to GC rich regions with folding structures, is no longer a problem since the sequencing by laser desorption mass spectrometry is based on the measurement of molecular weights. PCR is not required due to more sensitive analysis and hence non-faithful replication by PCR due to the secondary structures or a large number of repeats are be eliminated. Mass spectrometry is a fast method that provides a high-throughput means for DNA sequencing for up to 100 bp. However, mass resolution and detection sensitivity still require improvement for sequencing longer DNAs.

3.1.2.2 Sequencing by Hybridisation

Sequencing by hybridisation[154] uses a set of short oligonucleotide probes of defined sequences to search for complementary sequences on a longer target strand of DNA. This hybridisation method can be applied to solid phase chemistry allowing large arrays of probes to be synthesised on a chip. The active area of the chip is structured as a matrix, each region of which is assigned to a specific oligonucleotide (a short sequence of nucleotides). The matrix of solid supported probes is then treated with a solution of the target DNA sequence, suitably labelled. The target DNA will bind to probes with complementary sequences and hence hybridisation is identified in a single experiment. The hybridisation pattern, also called the spectrum of the sequence, is then used to reconstruct the target DNA sequence by mathematical methods. [155,156]

3.1.2.3 Pyrosequencing

Pyrosequencing[157-159] is a method of *sequencing-by-synthesis* and is based on the release of light as a result of numerous enzymatic reactions, *Figure 3.1.2.* Enzymatic DNA synthesis is accompanied by release of pyrophosphate in equal molarity to that of the incorporated deoxynucleotide. The pyrophosphate released is subsequently converted to ATP, by ATP sulfurylase, which provides energy. Luciferase makes use of the energy to oxidise luciferin which subsequently emits light. The amount of light produced due to the presence of ATP, luciferin and firefly luciferase is an estimate of the number of nucleotides incorporated. *e.g.* if the reaction was run using dTTP and light was produced it means that one, or more, T were incorporated. The reaction is then repeated with a different dNTP. Pyrosequencing is advantageous in its speed and reduction of costs. There is no need for labelled nucleotides, labelled primers or tedious electrophoresis. However, although it is accurate it still suffers from difficulty in the proportionality of the signal intensity when more than 5 nucleotides, carrying the same base in a row are incorporated.

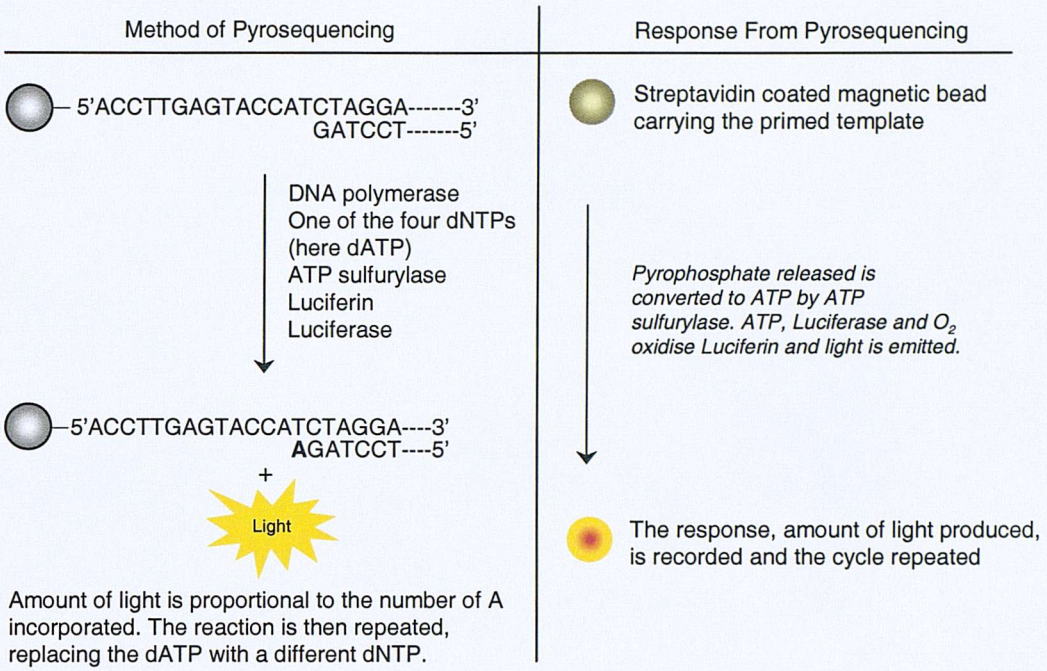


Figure 3.1.2 *Solid-phase pyrosequencing. The four different nucleotides are added stepwise to the immobilized primed DNA template and the incorporation is followed by the emission of light produced by the row of enzymatic reactions carried out by DNA polymerase, ATP sulfurylase and luciferase. Although pyrosequencing is illustrated here using solid supported DNA it can also be carried out in solution.*

3.1.2.4 Base Addition Sequencing Scheme

Another example of sequencing-by-synthesis is the *Base Addition Sequencing Scheme or BASS*.[160,161] This approach uses four nucleoside triphosphates that have a 3'-*O*-blocking group that is both labile and spectroscopically unique. A primer is annealed to a solid supported DNA template. Chain extension is carried out using 3'-*O*-labelled dNTPs, each with a unique label for base identification. Imaging of the support identifies the nucleotide incorporated and subsequent cleavage of the reporter group results in a 'free' 3'-hydroxyl group allowing the addition of the next complementary base.

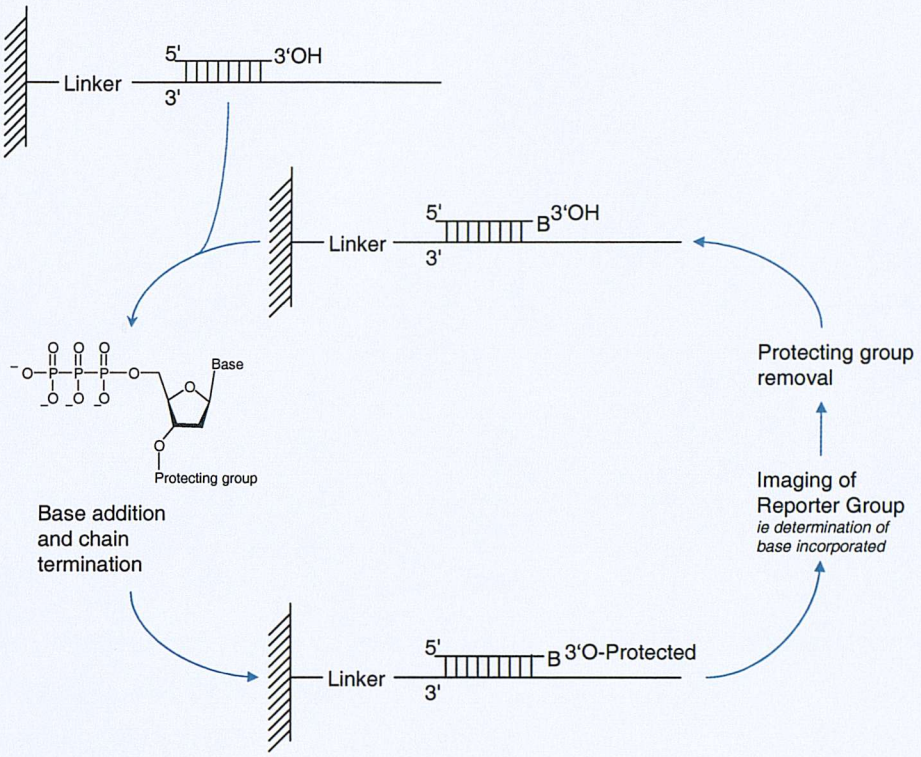
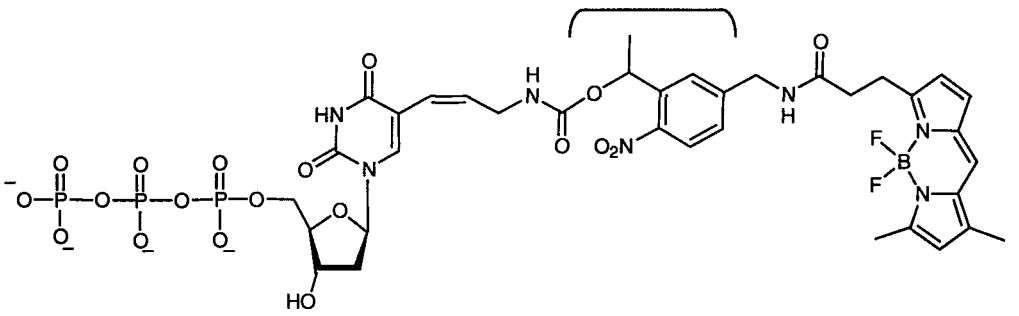


Figure 3.1.3 Illustration of the Base Addition Sequencing Scheme (BASS)

Initial studies of BASS made use of 3'-*O*-(2"-nitrobenzyl)-adenosine triphosphate as sequence terminating nucleotide. It was found that 3'-*O*-(2"-nitrobenzyl)-adenosine triphosphate could be incorporated by a DNA polymerase and that photo-deprotection of the 3'-hydroxy was possible facilitating DNA replication. However, four distinguishable nucleotides are necessary and although they could be synthesised they were not accepted by polymerases. Molecular simulations have suggested that large moieties situated on the 3'-position are too big to fit into the active site of DNA polymerases.

Recently, a modified version of this strategy was described in the literature.[162,163] This approach involved linking the reporter group to the 5 position by a photocleavable spacer and the 3'-hydroxyl would be capped by a smaller entity purely for chain termination purposes. The photocleavable fluorescent nucleotide analogue **74**, bearing a fluorescent dye, Bodipy, on its 5 position via a photocleavable 2-nitrobenzyl spacer, was synthesised and base specific incorporation was obtained in a polymerase reaction.

Photocleavable spacer moiety 2-nitrobenzyl carbamate



74

Fluorescent dye BODIPY

Sequences generated by the incorporation assay were terminated by a biotinylated ddNTP and isolated by streptavidin-coated magnetic beads. So far it has been established that **74** can be incorporated efficiently into the growing DNA strand by a DNA polymerase and its incorporation does not hinder the addition of subsequent nucleotides. Modifications of the 3'-hydroxyl of the nucleotide, which stop further incorporation but are compatible with DNA polymerases, are therefore necessary and are yet to be investigated.

3.1.3 A Novel Sequencing-by-Synthesis Strategy

A novel DNA sequencing-by-synthesis strategy has been proposed[164,165,166] in which single stranded DNA fragments are immobilised on a solid support *i.e.* a glass slide. A primer is then annealed to the template and the slide subjected to DNA chain extension conditions using 2'-deoxy fluorescent-dye labelled nucleotides. Certain dye-spacer combinations are known to inhibit some DNA mutant polymerases[167] and a carefully chosen dye-spacer-2'-deoxy nucleotide matched with a mutant polymerase could act as a DNA strand synthesis terminator. In this way single base addition synthesis products would be generated as represented in *Figure 3.1.4* and the incorporated nucleotides would be determined by the unique fluorescent label attached to it where each unique colour represents either A, T, C or G.

After the first chain extension reaction the extended DNA chains would be fluorescent and the slide would be read to determine the identity of the new base incorporated onto the template. The reporter moiety would then be cleaved off resulting in slides containing no fluorescent signals. Removal of the reporter group would then allow further chain extension, to allow further sequence information to be obtained.

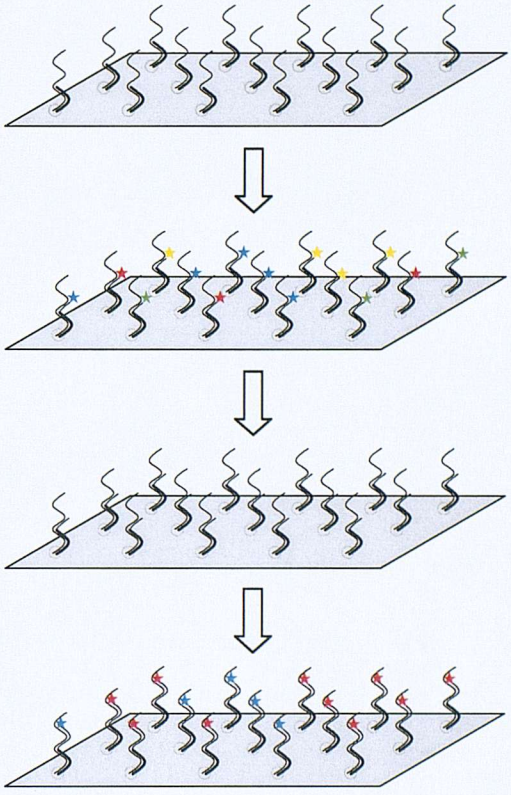


Figure 3.1.4 *Illustration of the new DNA sequencing-by-synthesis on glass slides*

3.1.3.1 Microarray or Single Molecule Detection

It is possible to immobilise single stranded DNA templates on glass slides in the form of a microarray by dispensing nanolitre volume spots of multiple copies of the single stranded DNA templates onto the glass slide. The spots could then simply be spatially encoded.

However, more interestingly, DNA could be immobilised in the form of a random array of single DNA molecules. In this case either the primer, linker or the template itself would need to carry a fluorescent label for detection. Once the slide has been subjected to DNA chain extension conditions incorporation would be determined by looking for the two dyes together, one from the template and one from the specific nucleotide incorporated. It would then also be possible to identify non-specific binding of labelled nucleotides since non-specific bound nucleotides would appear as a single dye only.

3.1.3.2 Attachment to the solid support

An object of the invention was to provide a method that would permit an essentially random distribution of biomolecules, *i.e.* nucleic acids, on a solid surface whilst allowing a degree of control of the density of molecules obtained. The solid support may be massive but would initially be limited to microscope glass slides. To allow detection of single molecules it is required that the molecules are distributed with sufficient separation between the molecules to enable each molecule to be individually resolved by optical methods. A coating to provide a sparse distribution of reactive groups comprising has been proposed where a reactive group, *e.g.* epoxide, has been attached to the solid surface through a Si-O bond. A polymer such as PEG would act as a linker group and would be added as a mixture of HS-PEG-SH and HS-PEG-OCH_3 at a ratio of $1 : 6.25 \times 10^6$ to achieve the desired density.

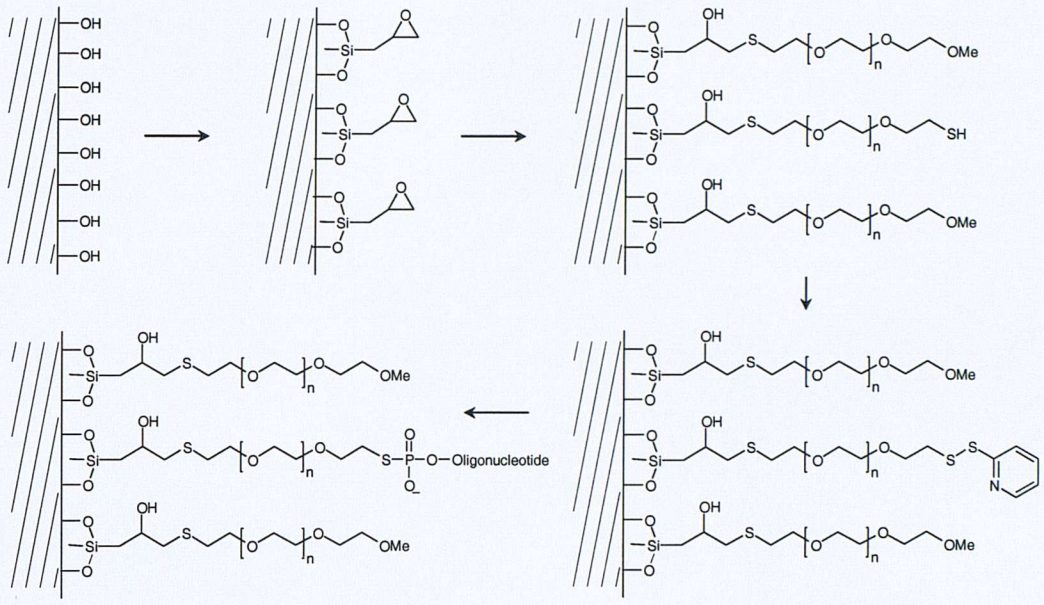


Figure 3.1.5 *An illustration of phosphorothioate oligonucleotides to surfaces grafted with sparsely distributed reactive groups.*

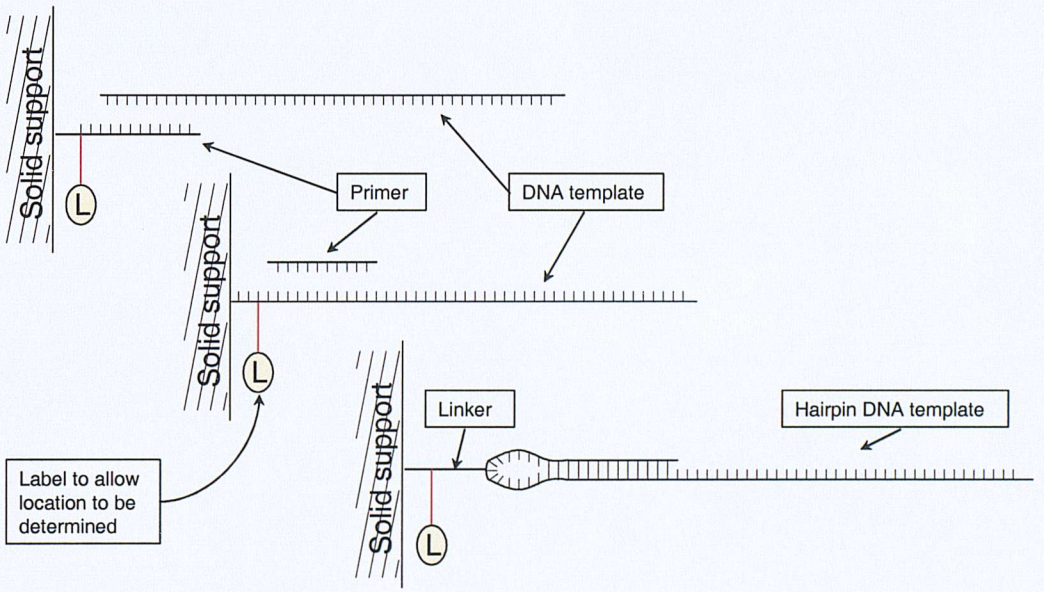


Figure 3.1.6 *Illustration showing alternative template/ primer attachments suggested*

The length of the oligonucleotide template would depend upon the genome being sequenced and the restriction enzyme used but an approximate 100 base pairs has

been suggested. The sequence information would come from the number of oligonucleotides on the slide rather than the number of extension reactions carried out as it is estimated that only 6 or 7 cycles would be necessary.

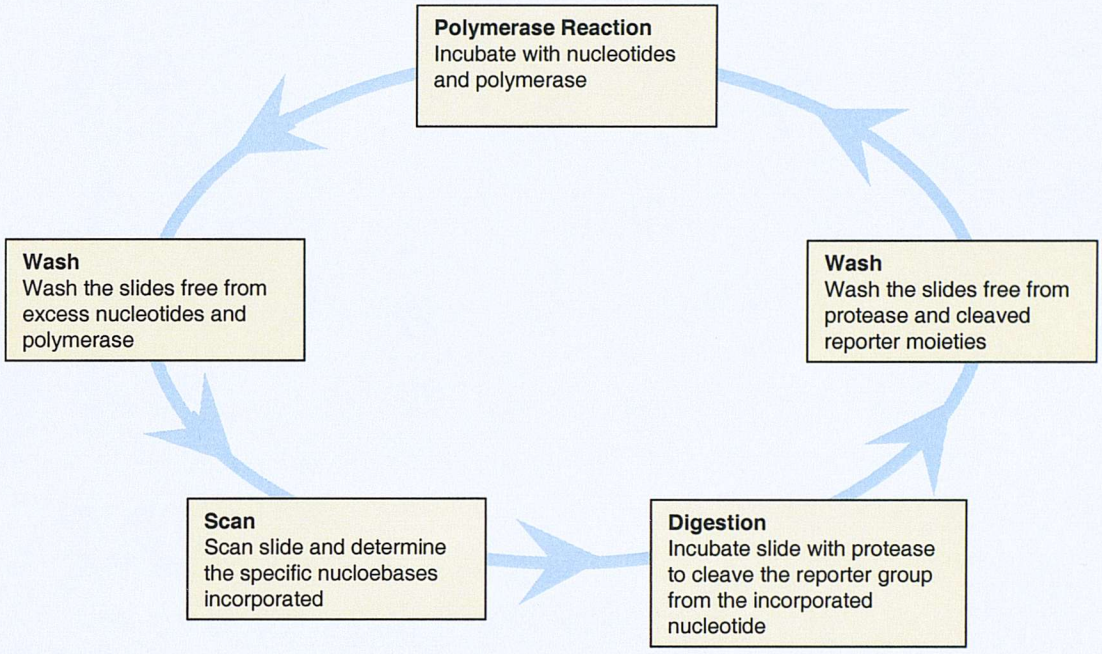


Figure 3.1.7 *Illustration of the cycle of base extensions proposed in the novel DNA sequencing system*

It is estimated that one would achieve a distribution where 10^8 oligonucleotide molecules are distributed over 1 cm^2. Hence, information of 10^8base pairs would be obtained in one 45 minutes cycle. The array of information from say 6 cycles would make up the sequence. This is of course an estimate and may have to be altered due to for example extended scans needed to improve signal to noise ratios.

3.1.3.3 Accuracy

The efficiency of cleavage would most probably not be 100% and one would therefore expect some false base calls. However, this can be over come by using more than one set of reporters.

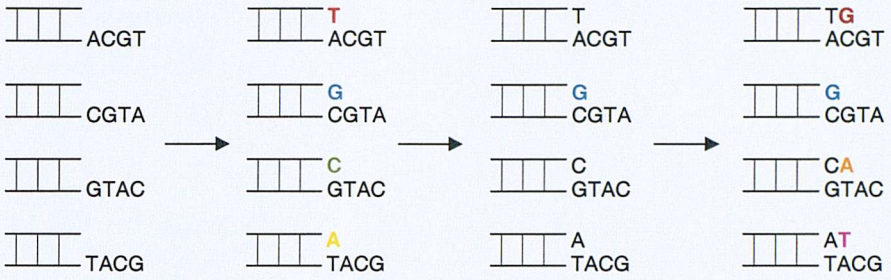


Figure 3.1.8 *Illustration of how the use of two (or more) sets of reporters can overcome the problem of false base calls.*

Figure 3.1.8 shows how the sequencing system would work if all four individually labelled bases were added simultaneously. After the first cycle the slide is read according to the first set of reporters. The reporters are then removed and base extension is repeated. *Figure 3.1.8* shows how the cleavage of the first set of reporters has failed where a reporter moiety from the first set of reporter nucleotides remains. However, by using a different set of reporters in the second cycle the failed cleavage can be readily identified.

3.1.3.4 Terminating or Non-terminating Labelled Nucleotides

2',3'-Dideoxynucleoside triphosphates are normally used as terminators in DNA sequencing. However, when using ddNTPs chain extension is permanently terminated. Here it was necessary that the 3'-hydroxyl group was reversibly protected or left unprotected. If dNTPs with unprotected 3' hydroxyl groups are used then it is not a true terminator and additional incorporations may be possible. This being the case, the sequencing slide would therefore have to be treated sequentially with the four fluorescently labelled dNTPs allowing the imaging system to detect possible double incorporations. If dNTPs with a reversible protecting group on the 3' position are used then the sequencing slide could be treated with all four dNTPs labelled with different fluorophores simultaneously. The imaging system would identify each colour as a different base in a single experiment but it includes an extra deprotection step in the sequencing system. This thesis is focused upon the development of "non-terminating nucleotides".

Non-terminating nucleotides terminate due to bulk in the base region of the triphosphate, which prevents additional incorporation beyond the first base.

This method of sequencing places certain requirements on the labelled 2'-deoxynucleotide used, *i.e.*:

- the reporter nucleotides used must possess a cleavable spacer between the nucleotide and the reporter moiety,
- the spacer must be cleaved as close to the nucleotide as possible in order to minimise disturbance of the efficiency of the DNA polymerases during subsequent chain extensions; and,
- the spacer and the fluorophore must inhibit polymerase activity. However, some limited double incorporation could possibly also be accounted for by optical methods (*i.e.* CCD camera).

To investigate these requirements the 5-propargyl aminopyrimidine (and 7-deaza-7-propargyl aminopurine) nucleotides, such as those described in section *1.2.5*, seemed to be a good starting point as they proved to incorporate well during DNA incorporation assays. The cleavable spacer could be a light cleavable, chemical cleavable or enzyme cleavable moiety. Chemical cleavage of common spacer groups generally involve the use of acid or base that could cause degradation of the DNA. A light cleavable spacer would be more desirable since only mild conditions would be necessary to separate the reporter from the growing DNA strand. However, the dye could be affected, but this would only be a problem if the cleavage was incomplete and photobleaching occurred as this could give the appearance of a fully cleaved spacer. In this case the dye moiety may not be fluorescent giving misleading results. Enzymatic cleavage of a spacer could provide a mild, fast and efficient cleavage of the spacer linking the nucleotide to the reporting fluorophore. Cleavage of peptides using proteases has been studied extensively and the propargylamine spacer is an excellent handle for peptide synthesis hence peptides as enzymatically cleavable spacers were investigated. There is a wide range of proteases available for hydrolysis of peptides and there is a large number of peptides that could be considered as potential spacer groups. Work in this chapter describes how a library of peptidic nucleosides were synthesised and evaluated using a variety of enzymes.

3.1.4 Strategy for the Optimisation of Peptidic Spacers

It is critical for the protease to withstand the close proximity of the nucleoside such that cleavage as close as possible to the nucleoside maybe achieved. As discussed earlier, this is very important as bulky groups may cause disturbance to subsequent chain extensions. However, the N-terminus of the peptide (the fluorophore bearing segment) is not as critical and therefore any undesirable interactions maybe alleviated by incorporating a spacer between the dye and the peptide. In order to find an effective reporter-peptide-nucleotide / protease combination a library synthesis of nucleoside analogues using resin **60**, *Figure 3.1.9*, seemed an attractive approach. Solid phase peptide synthesis could then be employed to create a library of peptidic nucleosides. The peptides would then be labelled with a dye and the nucleosides cleaved from the support and assayed in solution.

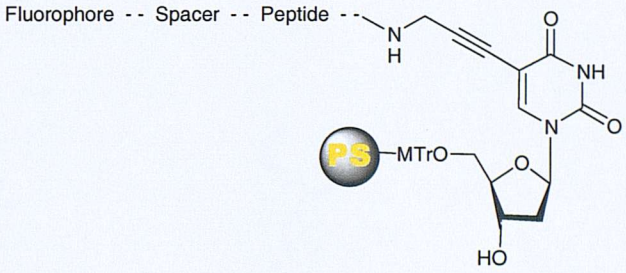


Figure 3.1.9 *Illustration of the type of reporter nucleosides that would be prepared by solid phase methods using resin* **60**

The aims of the library synthesis were:

- to verify that a peptidic spacer, between a nucleoside and a fluorophore, could be digested by enzymatic methods, and
- to obtain information about the most efficient protease / nucleoside combination.

3.1.4.1 Design Considerations

The library would include a number of different peptides varying in length and functionality. In some cases hydrophilic residues were incorporated to aid solubility of the final products in an aqueous environment. It was important that all protecting groups used could be removed under the same conditions to minimise the number of steps needed. Whilst the peptide would be variable, the reporting fluorophore would be fluorescein, for simplicity, and an aminohexanoic acid spacer between the peptide and the fluorophore was chosen throughout this study.

3.1.5 Protease Selection

3.1.5.1 Terminology

The terminology used in describing the specificity of proteases depends on a model in which the catalytic site is considered to be flanked on one or both sides by specificity subsites, each able to accommodate the side chain of a single amino acid residue. These sites are numbered from the catalytic site, S1…Sn towards the N-terminus of the substrate, and S1'…Sn' towards the C-terminus. The residues they accommodate are numbered P1…Pn, and P1'…Pn', respectively, as follows (the catalytic site of the enzyme being marked '*'):

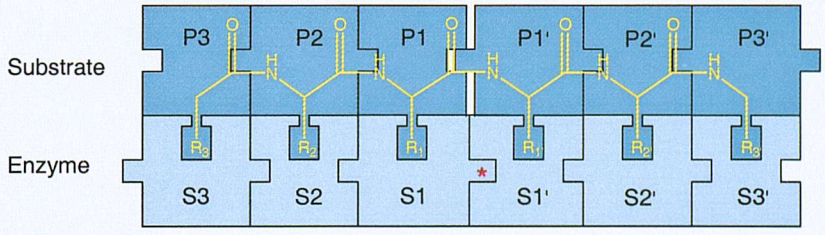


Figure 3.1.10 Illustration of the terminology used for an enzyme active site

3.1.5.2 Properties Considered for Selection of Protease

The choice of proteases was based on ease of handling, the sizes of the enzymes, availability and costs. Size of the enzyme, although not such an important property in solution phase screening, could be crucial if applied to surface screening methodologies where arrays of molecules are highly dense. For this reason, enzymes of a molecular weight (lower than 40 kDa) were focussed upon, although one larger enzyme, enterokinase (MW = 160 kDa), was also considered. Enzymes with optimum activity at a pH in the region where DNA is stable were desirable and all enzymes chosen fulfilled this requirement. Many proteases are relatively non-specific but a few facts, summarised in *Table 3.1.1*, regarding some properties guided the library synthesis.[168]

	Protease	**MW kDa**	**Catalytic type**	**Specificity**
1	Proteinase K	29	Serine	Relatively unspecific. However, aromatic and hydrophobic amino acid residues such as Leu are preferred at P1 while it does not hydrolyse substrates with Pro at P1'.
2	Elastase	26	Serine	A primary specificity for non-bulky residues such as Ala, Ser, Gly and Val in P1 has been observed. The catalytic activity has been shown to increase with the chain length of the synthetic substrate.
3	Subtilisin	27	Serine	Subtilisin is relatively non-specific but subsites S1 and S4 bear some preferences. In S1 non-β-branched hydrophobic side chains are preferred. The S4 subsite strongly prefers hydrophobic side chains.
4	Papain	23	Cysteine	Papain is considered to possess a fairly broad specificity but subsite S2 prefers bulky non-polar side chains and a weak preference for Lys in P1 has been observed.
5	Thermolysin	35	Metallo	Thermolysin prefers to cut the scissile bond at the N-terminal side of Leu, Phe, Ile and Val and a hydrophobic residue is preferred in P1 position.
6	Trypsin	24	Serine	Trypsin strongly prefers to cleave amide substrates following P1 Arg or Lys residues.
7	Enterokinase	160	Serine	Enterokinase cleaves at sites that closely resemble the cleavage of cattle trypsinogen: Asp-Asp-Asp-Asp-Lys-Ile.

Table 3.1.1 *Selected information about the seven enzymes that were to be employed when assaying the nucleoside library*

3.1.6 Selection of Peptide Sequences

Nine amino acids, alanine, aspartic acid, glycine, lysine, leucine, methionine, proline, serine and valine were used to construct the library. Three of these amino acids (aspartic acid, lysine and serine) required protection of their respective side chain functional groups. Aspartic acid, lysine and serine were used protected with the *tert*-butyl ester-, Boc- and *tert*-butyl ether group, respectively. This allowed removal of the protecting groups by acid treatment and concomitant release of the nucleoside analogue from the resin. The peptide sequences chosen are shown in *Table 3.1.2.*

Selected Sequences		
AA	APKS	GSAAG
AAA	APLS	GSAAK
AAAA	APMS	GSAAL
AAAS	APVS	GSAAPA
AAG	DDDDK	GSAAPK
AAK	DDDDL	GSAAPL
AAL	DDDKS	GSAAPV
AAPA	DDDLS	GSAGL
AAPK	GAGL	GSAIPM
AAPL	GGGL	GSALK
AAPV	GGL	GSDDDDK
AAS	GGLS	GSDDDDL
AGL	GK	GSGGGL
AGS	GKS	GSGK
AIPM	GL	GSLGL
AKS	GLS	GSVLK
ALK	GSAA	LGL
ALS	GSAAA	LKS
APAS	GSAAAA	VLK

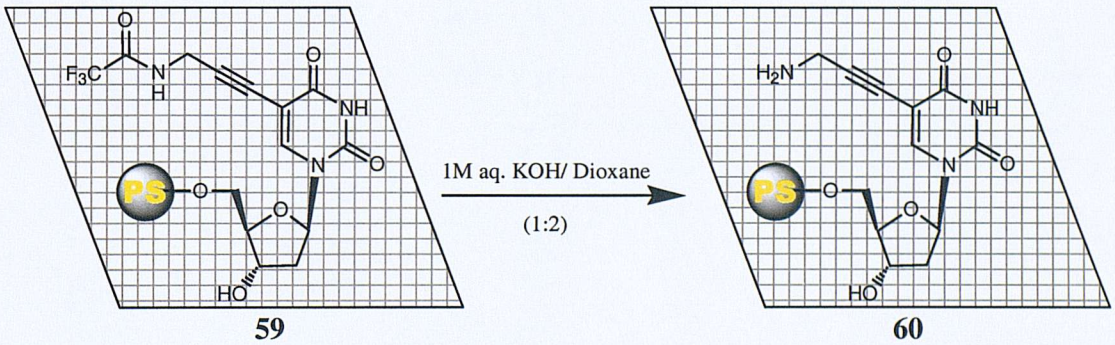
Table 3.1.2 *Library design. The table displays the peptides chosen to make up the spacer between the fluorophore and the nucleobase, here represented using single letter codes.*

In total 57 unique sequences were chosen, *Table 3.1.2*. Simple sequences, which did not possess side chain functional groups that needed protection, were desirable as synthetic targets would otherwise be more complicated. However, it was thought that the compounds may become too hydrophobic and serine was therefore incorporated either at the nucleoside end or the fluorophore end of the peptide.

3.1.7 Synthetic Methodology

Combinatorial tea-bag methodology suited this type of library synthesis over the alternative more laborious parallel synthesis using disposable filter tubes. The advantage of the teabag method over parallel synthesis is that only one pot of reactants is needed for each amino acid coupling (*i.e.* a maximum of nine reaction vessels). Moreover, all washing/ deprotection steps could be combined reducing the time immensely, see *Figure 1.1.5*.

To prepare the tea-bags polypropylene mesh was cut out, folded and sealed at three ends leaving one open for filling. Approximately 150 mg of resin **59** was distributed between ten polypropylene bags. A microchip transponder was placed in each bag for identification purposes and the bags were subsequently sealed by welding the ends together.



59 **60**

Scheme 3.1.1 *Hydrolysis of resin* **59** *in tea-bags*

The polypropylene bags containing resin **59** were treated with 1M aq. KOH/ dioxane (1:2) overnight, *Scheme 3.1.1*. During this procedure an issue concerning

the compatibility of the tea-bag methodology with this resin arose as the beads escaped from the enclosed polypropylene tea-bags.

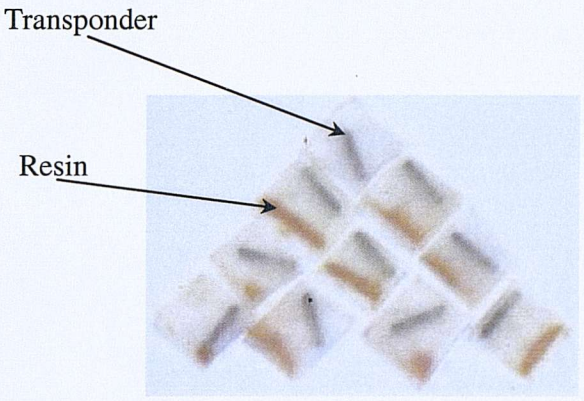
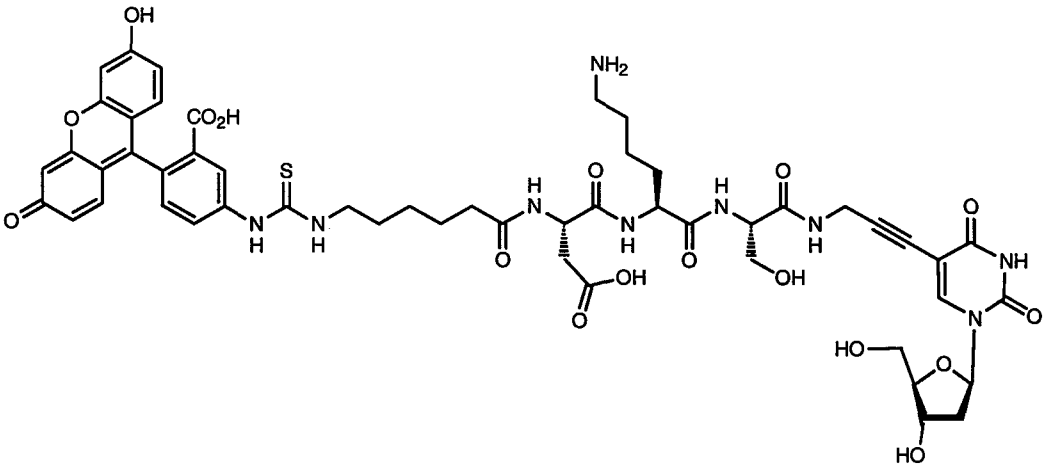


Figure 3.1.11 *Tea bags each initially filled with 10 – 20 mg of resin* **59** *plus a transponder for identification*

Figure 3.1.11 shows the tea-bags after overnight hydrolysis and it can clearly be seen that the amount of resin varies considerably from bag to bag. A visual inspection of the triangle array of tea-bags immediately suggests that the tea-bag at the top of the triangle was completely empty of resin beads. At this stage it was not possible to determine whether this was due to one or two badly sealed bags or whether all bags leaked some material. Obviously if all bags leaked some material this would not be a feasible method as the synthesis would include up to 18 on-resin synthesis steps and most of the resin could be lost. It was therefore decided to use these polypropylene tea-bags for the synthesis of a library control compound to evaluate whether the loss of resin was negligible or whether another method had to be used.

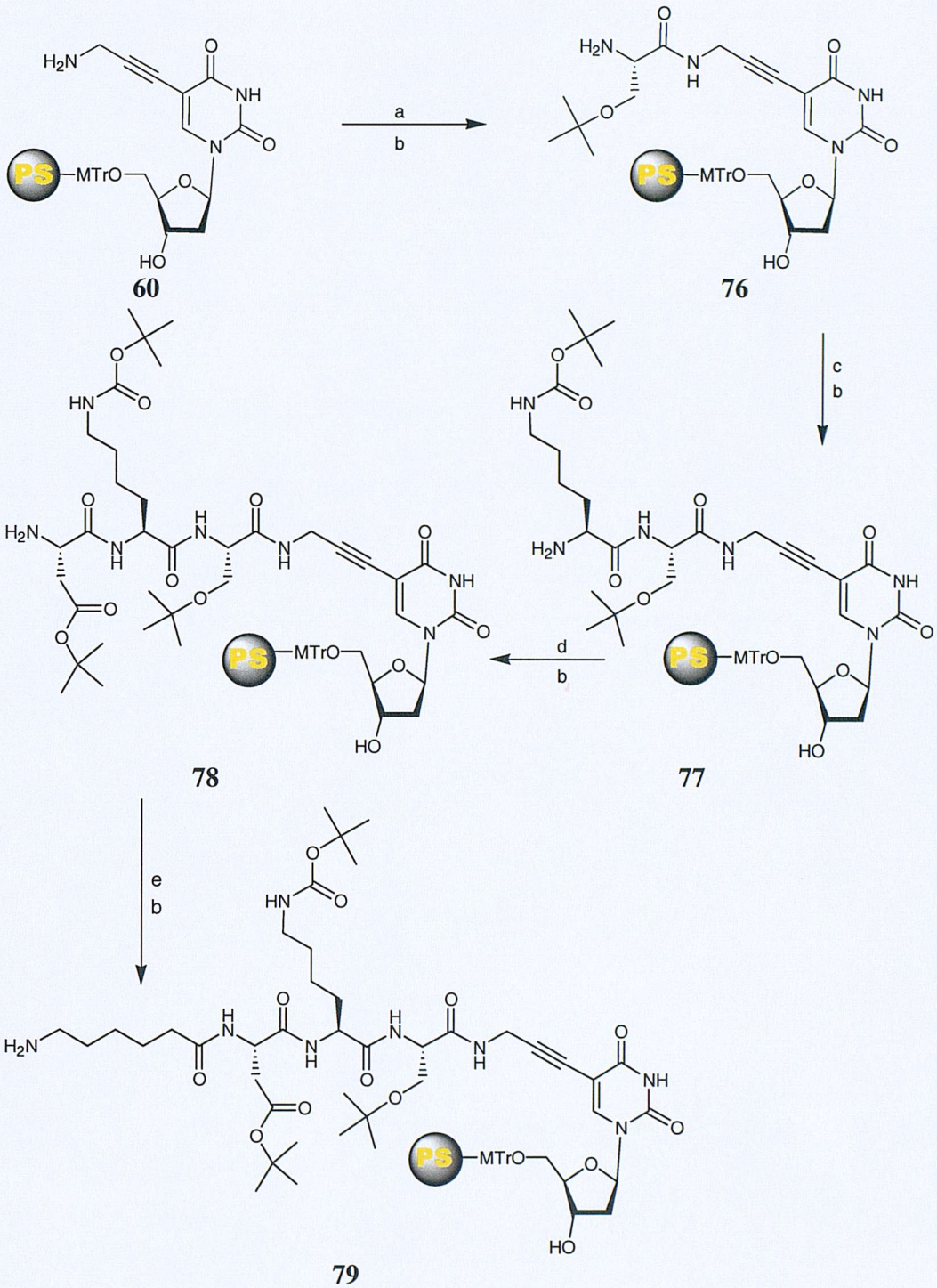
3.2 Monitored Synthesis of Library Control Compound

Synthesis of one library member was carried out to verify the synthetic strategy before tackling synthesis of the larger library.



75

Fam-Ahx-DKS-propargylamino-dU **75**, which includes three amino acids that require protection, was chosen as a target library control compound. The synthesis of this compound was monitored by HPLC and mass spectroscopy at each step and the tea-bag methodology was employed for synthesis. During the synthesis of the control compound both the method and the chemistry was monitored and evaluated for the larger library synthesis.

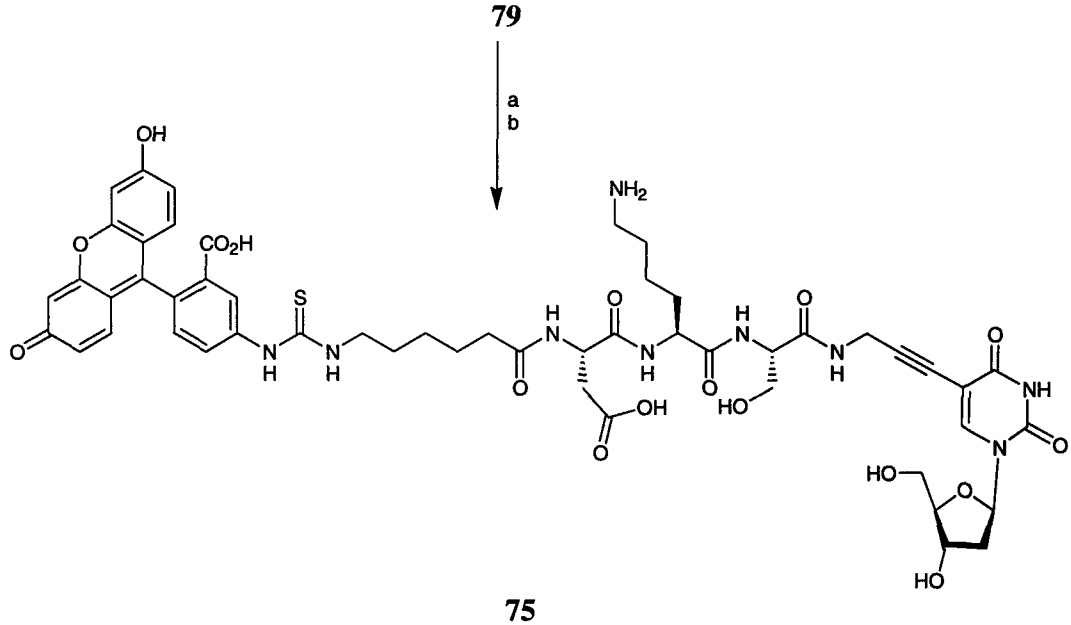


a.) Fmoc-Ser(^{t}Bu)-OH, DIC, HOBt, DMF/ CH_2Cl_2 (3:7)
b.) 20% Piperidine in DMF
c.) Fmoc-Lys(Boc)-OH, DIC, HOBt, DMF/ CH_2Cl_2 (7:3)
d.) Fmoc-Asp(O^{t}Bu)-OH, DIC, HOBt, DMF/ CH_2Cl_2 (7:3)
e.) Fmoc-Ahx-OH, DIC, HOBt, DMF/ CH_2Cl_2 (7:3)

Scheme 3.2.1 *Synthesis of* **79** *using tea-bag methodology, toward the synthesis of library control* **75**

Six tea bags containing resin **60**, 5-*N*-propargylamino-5'-*O*-(4-methoxytrityl polystyrene)-2'-deoxyuridine, were added to a 2M solution of Fmoc-Ser(^{t}Bu)-OH, DIC and HOBt in DMF/ CH_2Cl_2 (3:7). The reaction vessel was shaken on a flat bed shaker for 3 hours and the tea-bags were subsequently washed with CH_2Cl_2, DMF, MeOH and ether. One tea-bag was subjected to 5% TFA in DCM and the cleaved compound was collected for analysis. The remaining tea-bags were subjected to 20% piperidine in DMF for 1 hour to remove the Fmoc protecting group, one tea-bag was again cleaved and analysed by HPLC and mass spectroscopy.

The procedure was repeated with Fmoc-Lys(Boc)-OH, Fmoc-Asp(O^{t}Bu)-OH and Fmoc-aminocaproic acid analysing each step by HPLC and mass spectrometry. Purity was determined by HPLC and was typically 100% by ELS and 75% by UV_{282} detection. Once the amino acid couplings were complete the remaining resin **79** was placed into a filter vessel and treated with fluorescein isothiocyanate, *Scheme 3.2.2.*



75

a.) Fluorescein isothiocyanate isomer I, NEt_3, DMF, overnight
b.) 5% TFA in CH_2Cl_2 and then neat TFA

Scheme 3.2.2 Coupling of fluorescein isothiocyanate to resin **79**

The nucleoside derivatives were cleaved from the resin using 5% TFA in CH_2Cl_2 and further treated with neat TFA to remove the side chain protecting groups to give compound **75**. Treating a nucleoside with strong acid could potentially result in anomerisation thus a study to determine the time needed for deprotection yet to minimise anomerisation was carried out. The nucleoside was subjected to TFA for 1, 3, 30 and 60 minutes, *Table 3.2.1.*

Minutes	Protecting groups removed	Anomerisation
1	Boc deprotection	No
3	Only partially	No
30	Yes	No
60	Yes	Yes

Table 3.2.1 *NMR study of side-chain deprotection showed that 30 minutes treatment with neat TFA allowed deprotection without anomerisation.*

Although the NMR of this compound has a very complicated aliphatic region the 1' hydrogen is well separated from other signals in the aliphatic and aromatic regions and anomerisation can hence be detected when the peak corresponding to H-1' becomes broad. The shift was not large enough to obtain accurate integration of the amount of each anomer but it did allow a qualitative analysis. The NMR of the cleaved compound also confirmed that purification of the final compound was necessary. Control compound **75** was purified by semi-preparative HPLC and as a result provided a much improved NMR spectrum, especially in the aromatic region which was better resolved, *Figure 3.2.1.* Hence it was decided to purify all library compounds after cleavage from the resin.

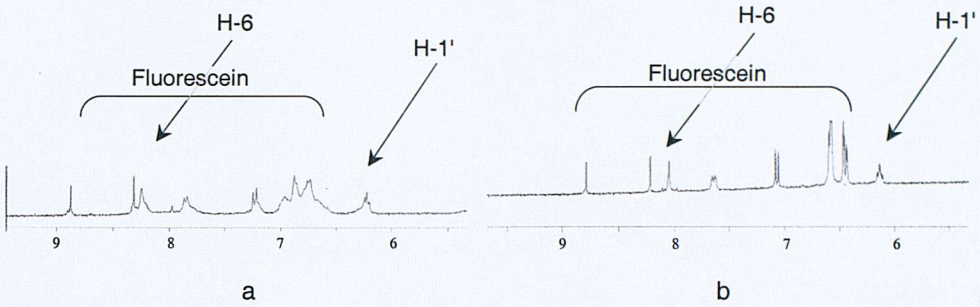


Figure 3.2.1 *Comparison of NMR data of control compound* **75**
a.) directly after cleavage from resin and
b.) after purification by HPLC

3.2.1 Evaluation of Synthesis Methodology

Unfortunately only one milligram of material was obtained after the monitored synthesis. This was due mainly to the great loss of resin during the synthesis. Once synthesis of the control compound, using the tea-bag methodology, was completed it was clear that the tea-bag method was unsuitable as the resin escaped through the polypropylene mesh. The library synthesis was therefore carried out in parallel using filter vessels. The control compound was resynthesised on a larger scale using a single filter vessel to obtain material sufficient for characterisation.

Microscope analysis of the beads and mesh revealed that the beads were too small for combinatorial synthesis by the tea bag method, particularly when dry.

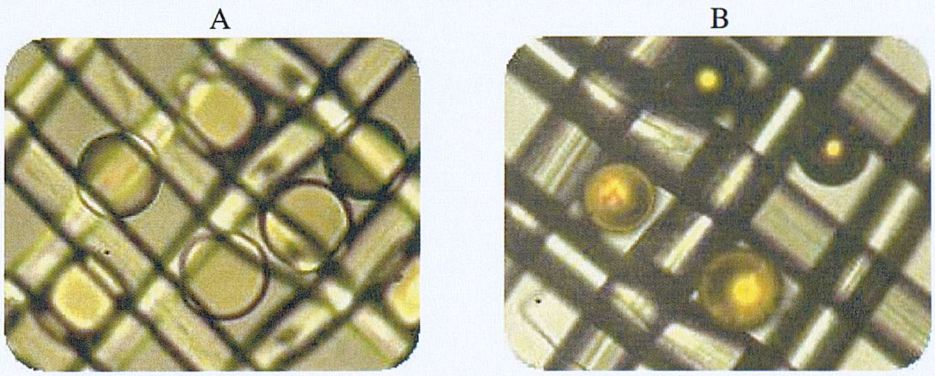


Figure 3.2.2 *A.)* *Beads swollen in DMF under the polypropylene mesh*
B.) *Dry beads under the polypropylene mesh*

3.3 Library Synthesis

Since the tea-bag method proved unsuited for synthesis it was decided to carry out the synthesis in parallel using single filter vessels. The entire library was too large to handle manually in a parallel sense and was therefore divided into three libraries, A, B and C. Sequences that did not contain serine were included in library A. For simplicity, all sequences for which syntheses started with serine and all sequences for which syntheses ended with serine followed by glycine were included in library B and C, respectively. The division of compounds is shown in *Table 3.3.1*, *Table 3.3.2* and *Table 3.3.3*, respectively.

Library A					
1	AA	8	AIPM	15	GL
2	AAA	9	AAPK	16	AAG
3	AAL	10	VLK	17	GGGL
4	AAAA	11	GK	18	AGL
5	AAPA	12	AAK	19	LGL
6	AAPV	13	ALK	20	GAGL
7	AAPL	14	GGL	21	DDDDL
				22	DDDDK

Table 3.3.1 *Library A*

Library B					
1	AAS	6	APLS	11	AKS
2	ALS	7	APMS	12	GLS
3	AAAS	8	APKS	13	AGS
4	APAS	9	LKS	14	GGLS
5	APVS	10	GKS	15	DDDLS
				16	DDDKS

Table 3.3.2 *Library B*

Library C					
1	GSAA	7	GSAAPL	13	GSALK
2	GSAAA	8	GSAIPM	14	GSAGL
3	GSAAL	9	GSAAPK	15	GSAAG
4	GSAAAA	10	GSVLK	16	GSGGGL
5	GSAAPA	11	GSGK	17	GSLGL
6	GSAAPV	12	GSAAK	18	GSDDDDL
				19	GSDDDDK

Table 3.3.3 Library C

The enzymes available directed the choice of peptides and hence A1-8, B1-7 and C1-8 would be assayed with elastase, which prefers non-bulky residues in P1. A9-13, B8-11 and C9-13 possessed lysine in the P1 position and would therefore be potential substrates for trypsin. Thermolysin could be assayed with a number of compounds but was limited to A15 and 20 and B12. Enterokinase was used for the assessment of aspartic acid rich sequences, A21-22, B15-16 and C18-19. Library members A14-15, A17-19, B12, B14, C14 and C16-17 would be assayed with proteinase K, which prefers Leu in P1. A1-3, B1-2 and C1-3 would be assayed with subtilisin as it prefers non-branched residues. A selection of library members with non-bulky side chains were screened against papain since it has been described as fairly non-specific but with a preference for non-bulky groups, *Table 3.1.1.*

The spacers were extended using standard solid phase peptide synthesis (SPPS). Coupling reagents were used as an equimolar mixture of HOBt/DIC and Fmoc amino acid (0.2M) in CH_2Cl_2 / DMF (1:2). Where glycine was used the amount of CH_2Cl_2 had to be replaced by DMF due to poor solubility in CH_2Cl_2. After each coupling reaction, each vessel was treated with 20% piperidine in DMF to remove the Fmoc protecting group and washed thoroughly. This cycle of peptide extension and Fmoc removal was repeated for each amino acid in the chosen sequence and finally Fmoc-amino caproic acid was coupled to every resin bound compound in the library followed by Fmoc group removal. Each vessel was then treated with fluorescein isothiocyanate in DMF and NEt_3. The resins were washed

thoroughly and subjected to cleavage using 5% TFA in CH_2Cl_2. Any compound released from the resin was highly coloured and the release of compound could therefore easily be monitored visually. Compounds in library B were cleaved reasonably efficiently by employing two to four cleavage cycles whereas Library A and especially Library C required up to ten cleavage cycles. The cleavage mixtures were collected in 48 well plates, solvents being removed using a Genevac™ evaporator before final treatment with neat TFA for removal of any side chain protecting groups. TFA was removed using the Genevac™ evaporator and all compounds were purified by semi-preparative HPLC.

3.3.1 Alternative Protecting Group Used in Library C

To ensure minimum exposure of the library of nucleosides to TFA an alternative protected serine was employed in the synthesis of library C (Fmoc-Ser(TBDMS)-OH). Synthesis was carried out as described above but before cleaving library C from the resin it was treated with 0.1M TBAF in DMF at 50°C for 3 hours. The resin was thoroughly washed and subsequently cleaved in the same manner as library A and B. However, upon HPLC analysis of the crude products a significant by-product was observed and NMR analysis revealed it to be a tetra-butyl ammonium salt. This peak was well separated from the product and although the by-product should have been completely removed by HPLC purification small amounts of the tetra butyl ammonium salt was still detectable after HPLC purification.

3.3.2 Library Analysis

All compounds were analysed by HPLC and mass spectrometry. All predicted masses were found and HPLC purity ranged from 24% to 100% of the crude and 79% to 100% of the purified products. ^1H NMR was carried out on 10% of the compounds *i.e.* A1, A2, A3, A15, B1, C1, C2 and C13.

3.4 Library Screening

All compounds prepared were assayed with a selection of proteases to determine the most suitable spacer/ protease combinations. Considering the unspecific nature of many proteases, assays of the total protease – library matrix could be of interest, however this was not feasible due to the number of manual assays required and the assays were therefore limited to the selection shown in *Table 3.4.1.*

Library / Protease	A	B	C
Proteinase K	14, 15, 17-20	12, 14	14, 17, 19
Elastase	1-8, 16	1-7, 13	1-8, 15
Subtilisin	2-3, 14	1,2, 12	1-3, 14
Papain	3, 6, 8, 13, 18, 19	2, 5, 7, 9, 12	3, 6, 8, 10, 13, 14, 17
Thermolysin	15, 20	12	
Enterokinase	21, 22	15,16	18, 19
Trypsin	9-13	8-11	9-13

Table 3.4.1 *The table displays all library compound / protease combinations assayed.*

Each compound was incubated with the specified enzyme and appropriate buffer for two hours at the enzyme optimum temperature. The enzyme was separated from the buffer solution containing the compound by filtration through a size exclusion membrane and hence the digestion was interrupted. All solutions were analysed by HPLC and digested fragments lacking the fluorophore moiety could be detected by the UV profile. Generally the digestions resulted in one new fragment containing the fluorophore and two or more new fragments containing the nucleoside moiety only. All assays included a blank run, *i.e.* where water replaced the enzyme, and by comparing the HPLC of this sample with that treated with an enzyme it was possible to determine whether or not the peptide had been

digested. For example, B12 containing the peptide GLS had an R_t of 23.4 minutes and the UV spectrum shows absorbance in the 290 and 440 nm region, *Figure 3.4.1 and Figure 3.4.2.*

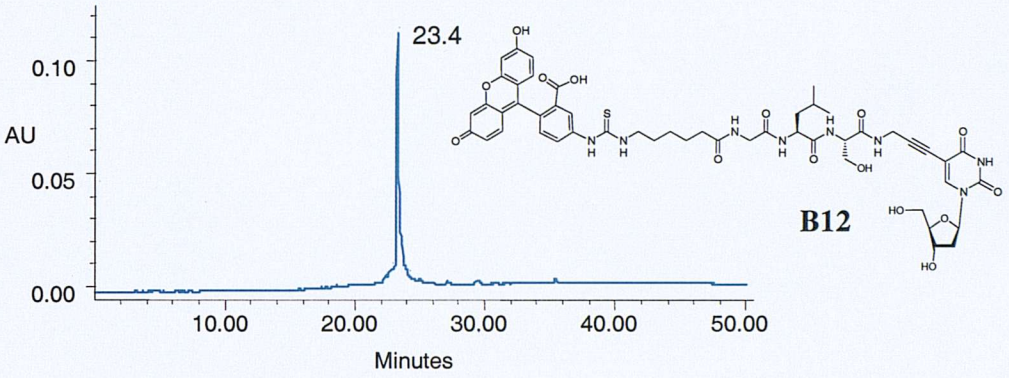


Figure 3.4.1 *HPLC trace of library compound B12 containing the peptide sequence GLS*

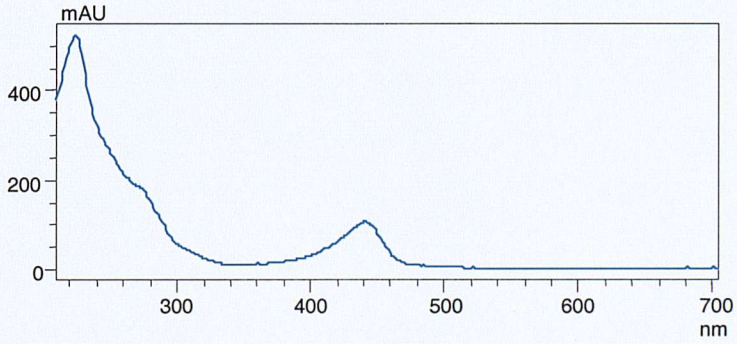


Figure 3.4.2 *UV spectrum of library compound B12 containing the peptide sequence GLS*

After the assay, using proteinase K, the spectrum displayed two new major peaks, $R_t = 6.4$ and $R_t = 26.3$, suggesting that proteinase K was reasonably specific about the cleavage site. However, minor peaks can also be detected and it is evident that a small amount is also cleaved at a different scissile bond. As expected, one major

peak had a slightly longer retention time, 26.3 minutes, compared to the starting material and the other major peak had a much shorter retention time, 6.4 minutes.

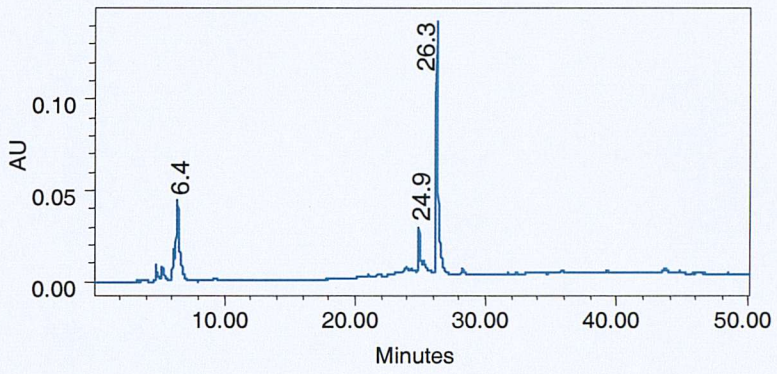


Figure 3.4.3 *HPLC trace of library compound B12 after proteinase K assay. The peptide has been digested by the enzyme to provide 2 major (and two minor) fragments.*

The UV spectra of the peaks with higher retention times displayed an absorbance in the 290 and 490 nm region, strongly suggesting that these fragments contained the fluorophore. UV spectra of peaks with shorter retention times lacked absorbance in the 490 nm region but did exhibit absorbance at 290 nm, *Figure 3.4.4*.

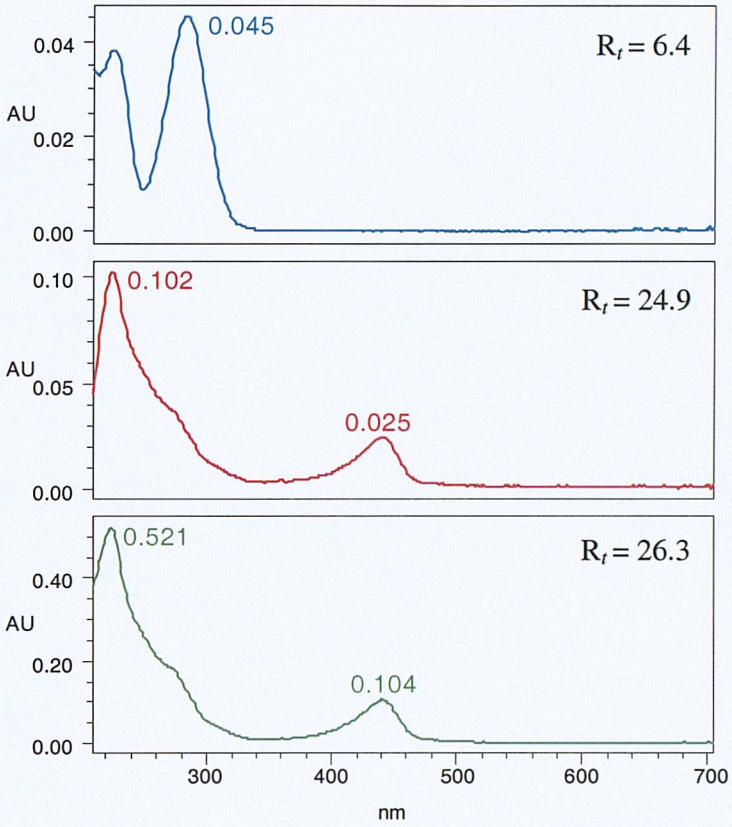


Figure 3.4.4 *UV spectra of individual peaks from library compound B12 after proteinase K assay.*

To determine the actual cleavage site it would be necessary to employ mass spectrometric analysis. This was not successful in many cases due to interference from the buffer used in the assay. However all assays carried out with trypsin were carried out using the volatile ammonium bicarbonate buffer and analysis by mass spectrometry proved possible.

3.4.1 Cleavage Site Determination

A selection of samples from the trypsin assay were analysed by LCMS. Compounds A9-13, B8-11 and C9-13 were assayed using trypsin since these included a lysine residue and the LCMS data suggested that trypsin exhibited a strong preference for hydrolysing peptides on the C-terminal side of lysine. This could be very interesting as the enzyme would cleave the peptidic spacer between the propargylamine spacer and the first lysine residue. However, cleavages of library C compounds, which do not include the serine at the nucleoside-end of the peptide, were not complete whilst compounds from library B, having serine at the nucleoside-end (C-terminus) of the peptide, were fully digested. Trypsin displayed poor digestion of compounds from library A suggesting that trypsin may prefer slightly longer peptides / spacers and more polar residues. By HPLC nucleoside-containing fragments were observed with a shorter retention time. Unfortunately these fragments were hidden in the solvent front of the LCMS system and therefore could not be analysed and hence only data regarding the fluorophore moiety could be obtained. Typically the fragments could be identified both by negative and positive electrospray mass spectrometry where the ions $[M-H]^-$ and $[(M+2H)/2]^+$, respectively were observed, *Figure 3.4.5*.

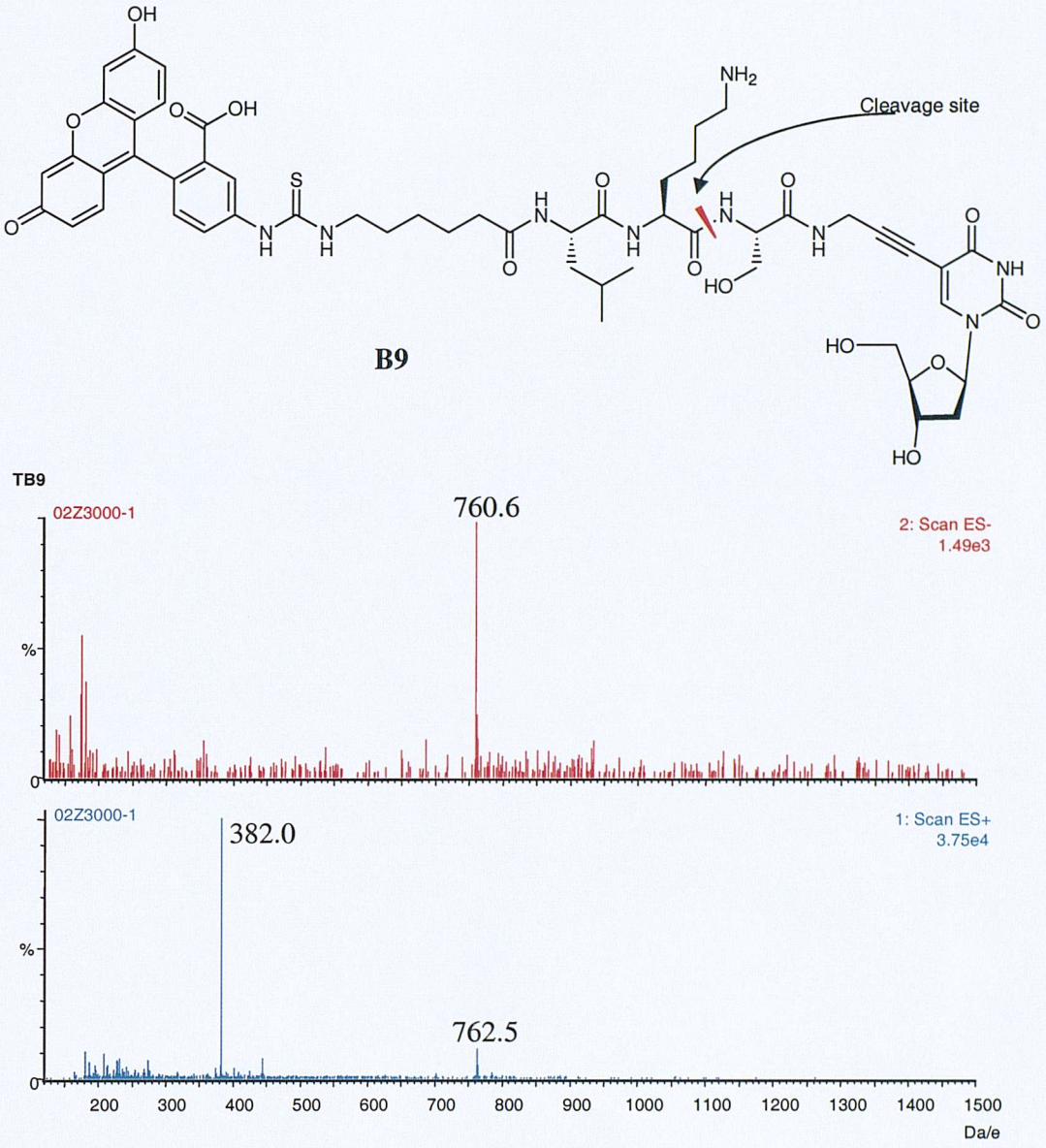


Figure 3.4.5 *Mass spectra (obtain by LCMS) of compound 9 library B containing the peptide sequence LKS after digestion with trypsin*

Proteinase K and subtilisin completely digested the selected sequences whereas papain and enterokinase displayed poor digestion efficiency. Trypsin, elastase and thermolysin were more selective but displayed poor cleavage of some sequences whereas other sequences were wholly cleaved, *Figure 3.4.6*.

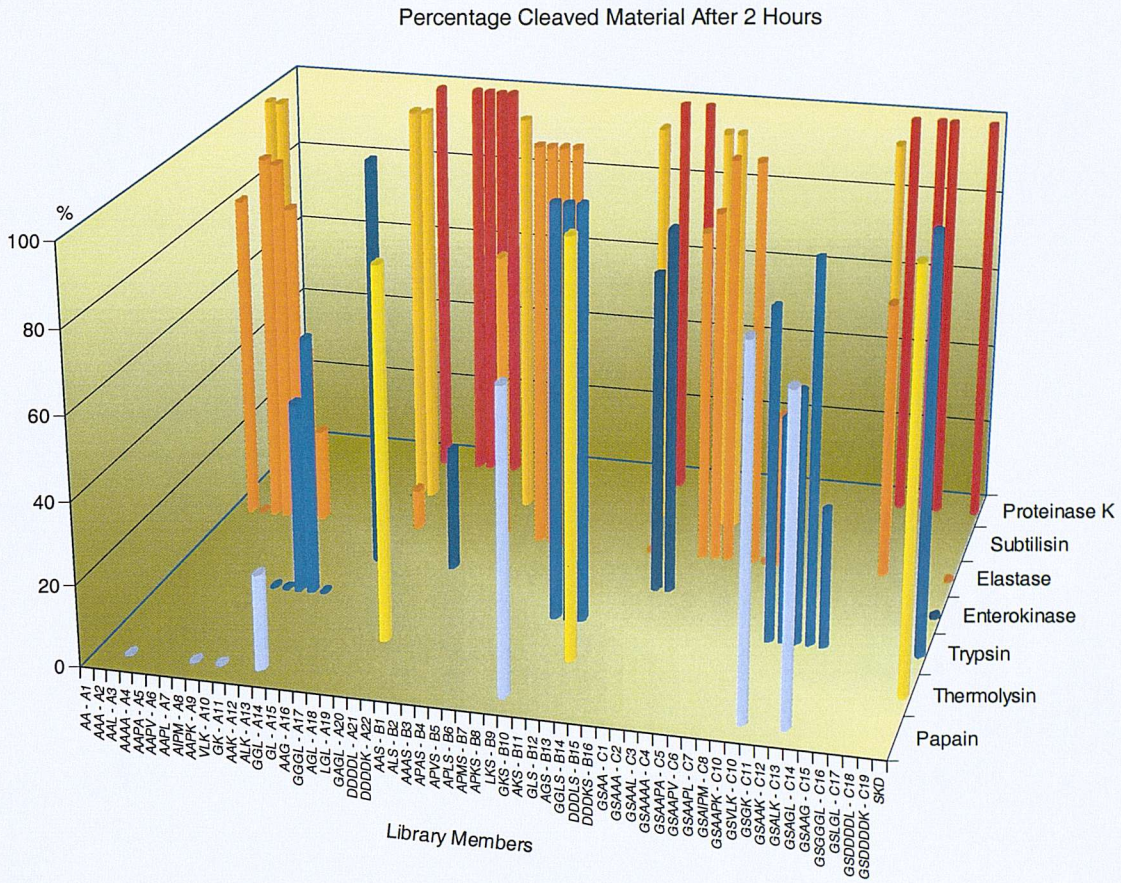


Figure 3.4.6 *Solution screening of library A, B and C with a selection of the preferred proteases to determine the most suitable spacer/ protease combinations.*

To obtain information about the digestion reaction rate, a selection of the more successful assays were repeated over time. The most interesting enzymes were subtilisin and proteinase K since they showed high digestion efficiency for all selected compounds. Elastase, trypsin and thermolysin which displayed a more selective digestion were also of interest. The following combinations of compounds and enzymes were selected for time course assays:

	Proteinase K	Subtilisin	Elastase
A2		✓	
A3		✓	
A5			✓
A14		✓	
A20	✓		
B12	✓		
C2		✓	
C5			✓
C14	✓	✓	

Table 3.4.2 Combinations selected for time course assays

Assays were set up following the same principle as for the initial screen, but aliquots were removed at various time points and the enzymatic activity quenched. The enzyme was separated from the solution containing the compound by filtration through a size exclusion membrane and the samples were analysed by HPLC.

3.4.2 Results from the Subtilisin Time Course Experiment

A2 (AAA) and A3 (AAL) were cleaved within the first 10 seconds and all enzymatic activity seemed to be over in approximately 8 minutes. A14 (GGL) was cleaved in approximately 3 min. C2 (GSAAA) and C14 (GSAGL) were probably cleaved within 10s but complicated HPLC data made it difficult to judge. The complicated HPLC traces are due the fact that A14, C2 and C14 were not stable to the quenching conditions and therefore had complex 0s analysis. However, the cleaved products were stable and the analysis became simpler as the digestion progressed.

3.4.3 Results from the Proteinase K Time Course Experiment

A20 (GAGL) and B12 (GLS) were digested in 10 and 30 seconds, respectively. Assay solutions from B12 and C14 (GSAGL) were colourless once filtered (this was also observed in the subtilisin assay of C14). This suggested that the compounds had precipitated on the size exclusion filter. Time point 0s displayed no starting material and subsequent aliquots display fragments containing only the nucleoside chromophore. To recover possible material precipitated the filters were treated with a 1M solution of triethylammonium bicarbonate for 3 hours, filtered and yellow solutions were obtained. Analysis of these solutions showed that B12 was digested after 30 seconds when analysis of the triethylammonium solution displayed one fluorophore fragment with R_t = 7.7 min at 30s and a different fluorophore fragment at all other time points R_t = 8.4 min. With C14 it was not possible to analyse when the starting fluorophore displayed identical retention time to that of the digested fragment.

3.4.4 Results from the Elastase Time Course Experiment

Digestion of A5 (AAPA) by elastase was rapid and by 8 minutes it was completely cleaved. C5 (GSAAPA) was stable toward the protease for the first two minutes but then started to show digested fragments. Even after 16 minutes there was still a considerable amount of starting material left. In comparison to subtilisin and proteinase K, elastase displayed slower rates of digestion and was therefore not considered further.

From these time course experiments it was clear that the simpler and shorter peptidic spacer sequences were as good as, and in most cases better than, the longer and more complex peptidic spacers. From the rapidly digested peptide sequences, AAA, GSAAA, AAL, GLS and GAGL, the aliphatic peptides AAA, AAL and GAGL were the simplest to synthesise and therefore were of most interest for further research/ development.

3.5 Conclusion

The resin developed in Chapter 2 was successfully used to synthesise a library consisting of nucleoside-spacer-dye conjugates. The resin was successfully applied to standard solid phase peptide synthesis as the spacers were introduced as peptides. The spacers were two to seven amino acids in length and were all terminated with aminohexanoic acid and a fluorescein dye. Library members were screened, in solution, for activity as substrates against a selection of proteases. Several library members were successfully digested by the proteases and a selection was chosen for time course experiments. Time course experiments confirmed that AAA, GSAAA, AAL, GLS and GAGL were rapidly digested and would therefore be good candidates for the Sequencing-by-Synthesis scheme suggested, section *3.1.3*.

The successful library synthesis and screening substantiate the fact that resin **60**, and indeed equivalent resins of the other three bases, is a very useful type of resin. Compound synthesis is fast with only a single purification step at the end of the synthesis being necessary.

4 Synthesis of a Biocompatible Nucleoside Resin and On Bead Enzymatic Assays

The purpose of this chapter is to discuss the details of the synthesis and on-bead enzymatic assays of a biocompatible nucleoside based resin.

4.1 Introduction

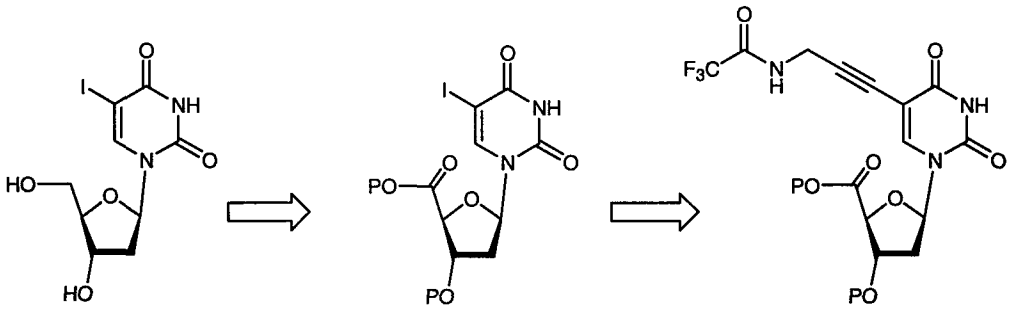
Solid phase chemistry has greatly facilitated synthesis of a vast number of compounds for biological screening. Of course, if the compounds were synthesised on a biocompatible resin they could be screened whilst still attached to the solid support. This would eliminate the tedious yet necessary cleavage and purification steps and ultimately shorten preparation time. As discussed in Chapter 1, to date only two resins with reliable enzyme compatibility, *i.e.* CPG (large pore size) and PEGA, have been developed.

This chapter describes the synthesis of a nucleoside based PEGA resin, analogous to that described in Chapter 2, and further derivatisation of four resins, analogous to the nucleosides described in Chapter 3.

4.2 Synthetic Strategy

To prepare nucleosides analogous to those described in Chapter 3, it was necessary to find an alternative attachment strategy. Cleavage of the side chain protecting groups requires acid treatment and thus the integral acid labile trityl linker was therefore not an option for library synthesis where the final products were to remain resin bound. Since detachment of the nucleosides from the solid support was not necessary it was decided to use an amino methyl PEGA support. Modification of the nucleoside was required as the attachment point on the sugar was the 5'-position, *Scheme 4.2.1.* If the 5'-hydroxyl group of a compound such as **80** could be oxidised to the corresponding 5'-carboxylic acid **81** it would be possible to anchor the nucleoside *via* a 5'-amide bond. Solid phase palladium chemistry always results in discoloured (slightly brown) resin possibly due to

some of the palladium catalyst remaining within the resin matrix. As the assays were to be carried out on the solid support it seemed a safer option to extend the solution phase chemistry to include the Sonogashira reaction and hence exclude palladium contamination of the resin matrix. The synthetic strategy developed hence eliminated the possibility of palladium impurities by preparing **82** prior to immobilisation.



80 **81** **82**

Scheme 4.2.1 Synthetic Strategy (P = protecting group)

Once **82** had been prepared, the ester could be hydrolysed to give the necessary 5'-carboxylic acid for immobilisation. Once immobilised, peptide extension and fluorescent labelling would be carried out in the same fashion as for the polystyrene based analogue described in Chapter 3.

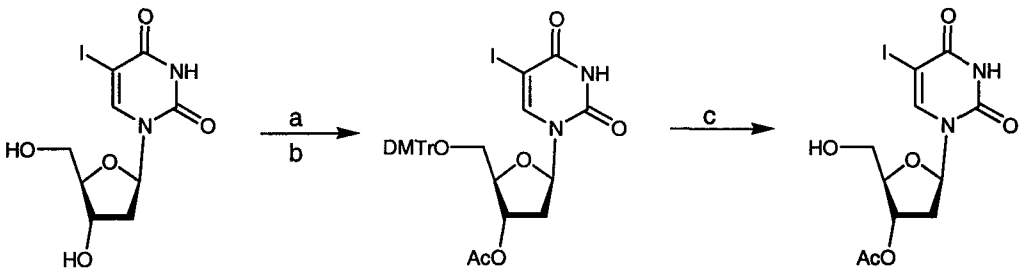
In summary the synthetic strategy comprised of four main sections:

- Sugar modification,
- Base modification,
- Immobilisation, and
- Spacer extension.

4.3 Synthesis

4.3.1 Sugar Modification

Ketonucleoside derivatives are useful synthetic intermediates for the synthesis of sugar-modified nucleosides.[169-172] Many reports on the oxidation of the primary hydroxyl group as well as the secondary hydroxyls have been reported in the literature where successful reagents are mainly those derived from chromium(VI)-based oxidants. A particularly attractive modification was that presented by Classon who reported the conversion of the 5'-hydroxyl group into a *tert*-butyl ester group in one pot.[173,174] To modify the 5'-hydroxyl group it was first necessary to synthesise the 3'-*O*-protected 5-iodo-2'-deoxyuridine, *Scheme 4.3.1.*



39 **83** **84**

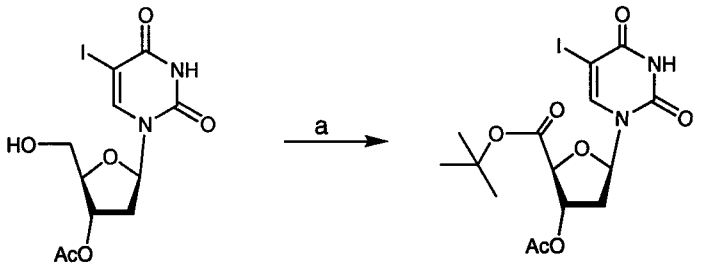
a.) Dimethoxytritylchloride, pyridine, 0°C
b.) Acetic anhydride
c.) 50% Trifluoroacetic acid

Scheme 4.3.1 Preparation of 3'-O-acetyl-2'-deoxyuridine

5-Iodo-(3'-*O*-acetyl-5'-*O*-4,4'-dimethoxytrityl)-2'-deoxyuridine **83** was prepared directly in one pot in 34% yield by the addition of 4,4'-dimethoxytrityl chloride to a cooled solution of 5-iodo-2'-deoxyuridine **39** in pyridine followed by the addition of an excess of acetic anhydride. Compound **83** was also prepared stepwise with isolation of 5-iodo-(5'-*O*-4,4'-dimethoxytrityl)-2'-deoxyuridine **57** followed by conversion to **83** in 81% and 84% yields, respectively. Removal of

the dimethoxytrityl protecting group (by dissolving the product in 50% TFA) was quantitative. At this stage it was important to keep the product in the dark and to keep the TFA treatment to a minimum otherwise loss of iodide and anomerisation were observed.

Oxidation of the hydroxymethyl group of **84** was attained using pyridinium dichromate and acetic anhydride,[173] *Scheme 4.3.2*. The direct addition of *t*-butyl alcohol to the oxidative reaction mixture provided a convenient route to the *t*-butyl ester protected compound **85** in 88% yield.



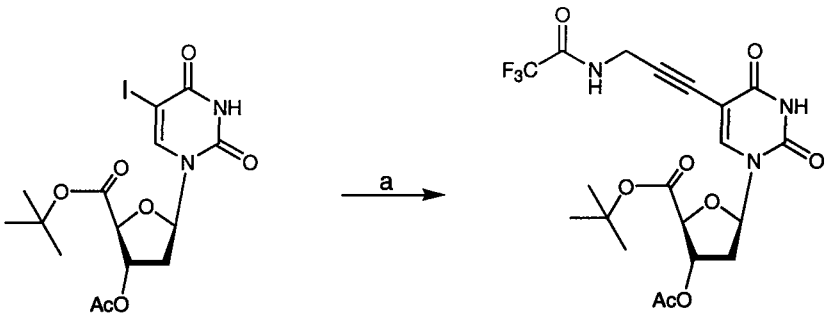
84 **85**

a.) PDC, *t*-butyl alcohol, acetic anhydride, DCM

Scheme 4.3.2 *Preparation of 1-(t-butyl 3'-O-acetyl-2'-deoxy-β-d-ribo-furanosyluronate)-5-iodouracil*

4.3.2 Base Modification

With the 3'-hydroxyl and 5'-carboxylic acid protected as acetate and *t*-butyl esters, respectively, the compound was ready for base modification under Sonogashira conditions.



85 **86**

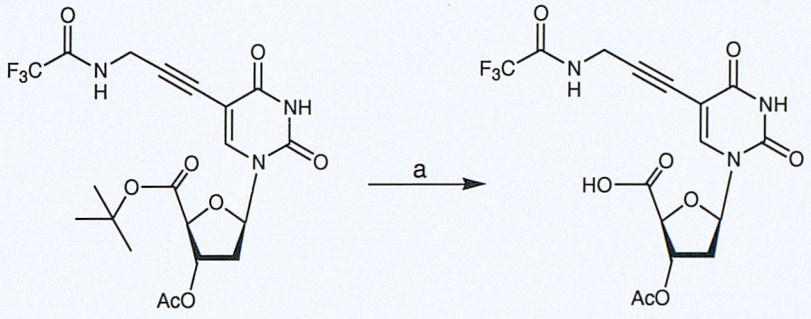
a.) CuI, NEt_3, *N*-trifluoroacetate propargylamine, $Pd(PPh_3)_4$, DMF, 50°C

Scheme 4.3.3 *Preparation of 1-(t-butyl 3'-O-acetyl-2'-deoxy-β-D-ribo furanosyluronate)-5-N-trifluoroacetylpropargylaminouracil* **86**

Preparation of 1-(*t*-butyl 3'-*O*-acetyl-2'-deoxy-β-D-ribofuranosyluronate)-5-*N*-trifluoroacetylpropargylaminouracil **86** was achieved by the addition of copper iodide, triethylamine, *N*-trifluoroacetate propargylamine and *tetrakis* triphenyl phosphine palladium(0) to a degassed solution of **85** in DMF. The reaction, which was monitored by TLC and HPLC, was complete within 50 minutes and furnished **86** in 78% yield.

4.3.3 Immobilisation

The *t*-butyl protected carboxylic acid was deprotected by acid hydrolysis by subjecting **86** to 80% aqueous trifluoroacetic acid for 30 minutes and afforded **87** in 83% yield, *Scheme 4.3.4*.

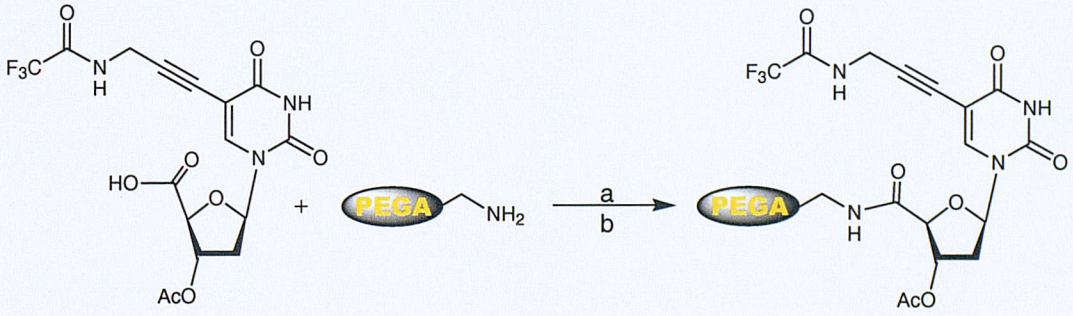


86 **87**

a.) 80% aq. TFA

Scheme 4.3.4 *Preparation of 1-(3'-O-acetyl-2'-deoxy-β-D-ribofuranosyluronic acid)-5-N-trifluoroacetylpropargylaminouracil* **87**

The carboxylic acid functionality of compound **87** provided a convenient attachment handle for resin immobilisation. The chosen support for immobilisation was polyethylene glycol-poly-dimethylacrylamide (PEGA), an aminomethyl functionalised resin to which **87** could conveniently be coupled. The immobilisation was monitored by qualitative ninhydrin tests, which were essentially negative after two hours, however, to ensure that no residual free amines were present on the resin the coupling was repeated followed by a capping step using acetic anhydride.



87 **88**

a.) HOBt/ DIC, DMF

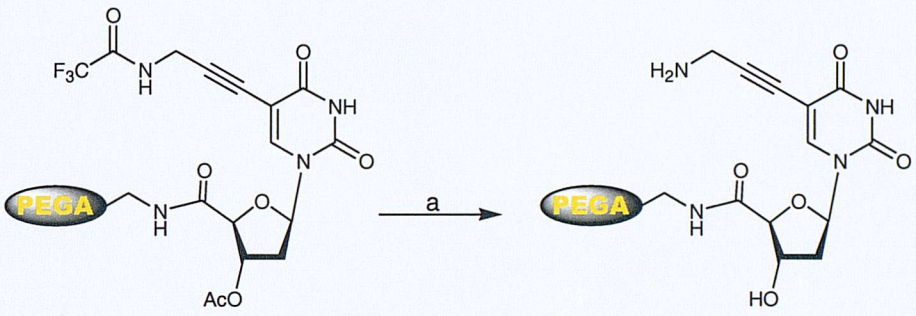
b.) Ac_2O, Pyridine, CH_2Cl_2

Scheme 4.3.5 *Immobilisation of 1-(3'-O-acetyl-2'-deoxy-β-D-ribofuranosyl-uronic acid)-5-N-trifluoroacetylpropargylaminouracil* **87**

Having efficiently prepared **87** and immobilised it onto PEGA resin, the subsequent Fmoc chemistry was carried out in a similar fashion to the polystyrene based library as described in Chapter 3.

4.3.4 Spacer Extension

In preparation for spacer extension the *N*-trifluoroacetamide protecting group was hydrolysed using a mixture of 1M aqueous potassium hydroxide / dioxane (1:2) with concomitant removal of the 3'-acetate protecting group.

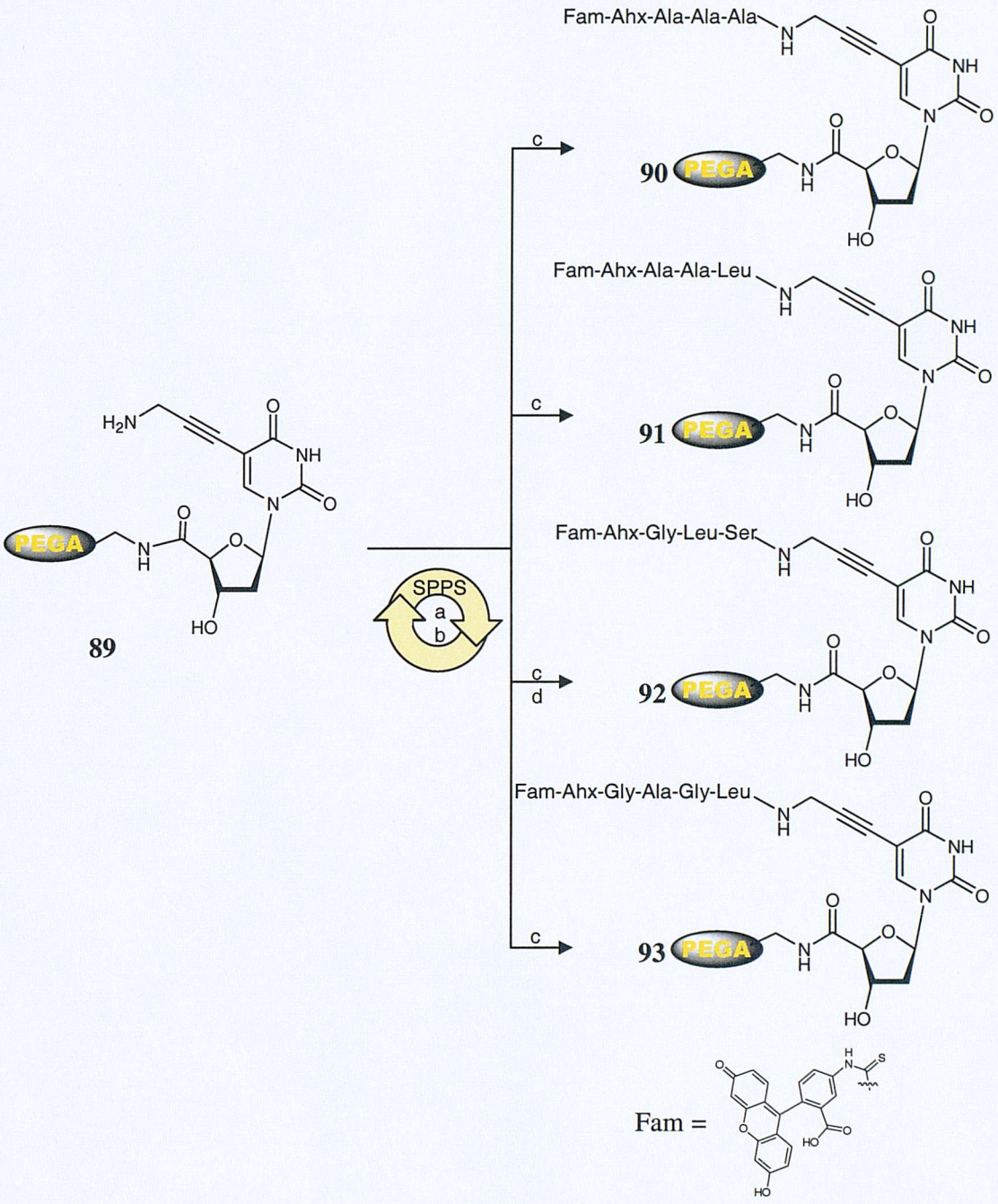


88 **89**

a.) 1M aq. KOH/ Dioxane (1:2)

Scheme 4.3.6 *Preparation of PEGA immobilised 1-(3'-hydroxy-2'-deoxy-β-D-ribofurnaosyluronic acid)-5-N-propargylaminouracil* **89**

Four fluorescein labelled peptidic spacer sequences, (AAA, AAL, GLS, and LGAG) selected from the library described in Chapter 3, were sequentially built up on PEGA resin **89**. Spacer extension was performed by standard Fmoc solid phase peptide synthesis as described in, *Scheme 4.3.7*. All sequences were terminated with Fmoc aminohexanoic acid before labelling the resins with fluorescein isothiocyanate.



a.) HOBt/ DIC, Fmoc-AA-OH, DMF/ CH_2Cl_2

b.) 20% Piperidine in DMF

c.) FITC, NEt_3, DMF

d.) neat TFA

Scheme 4.3.7 *Spacer extension carried out in parallel to give four nucleoside analogues immobilised on biocompatible PEGA resin*

92 contained a side chain protected serine residue and was therefore treated with neat TFA to remove the *tert*-butyl ether protecting group before analysis.

4.4 On Bead Assays

The resins described in this chapter are similar to those in Chapter 3, the only difference being the nature of the resin and the attachment strategy. Whereas solution phase compounds had to be screened using HPLC, compounds permanently attached to this PEGA resin could be screened using the much faster method of fluorescence spectrometry. In this manner, on bead analysis offered significant advantages over solution phase analysis in terms of through-put.

4.4.1 Assay Procedure

All fluorescent resins, **90**, **91**, **92**, **93**, were thoroughly washed with the assay buffer to avoid release of compound independent of protease treatment. The resins were assayed with an assay reaction mixture consisting of one mg of resin, 1 mL of reaction buffer and 10 units of enzyme. Aliquots (3 μL) were collected at various time intervals and transferred to a pH 9 borate buffer (3 mL) for analysis by fluorescence spectroscopy.

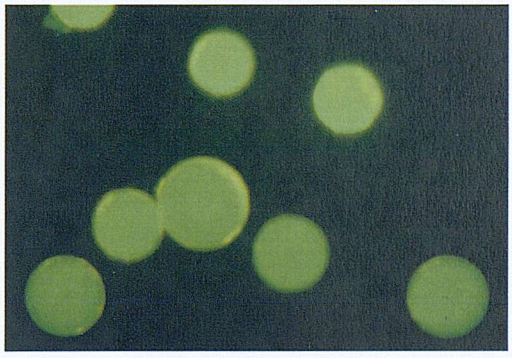


Figure 4.4.1 Photo of fluorescent PEGA resin **90**

4.4.2 Assay Results

As expected the proteases were able to digest the peptides, which resulted in an increase of fluorescence of the assay solution. A typical response from the fluorimeter is shown in *Figure 4.4.2*.

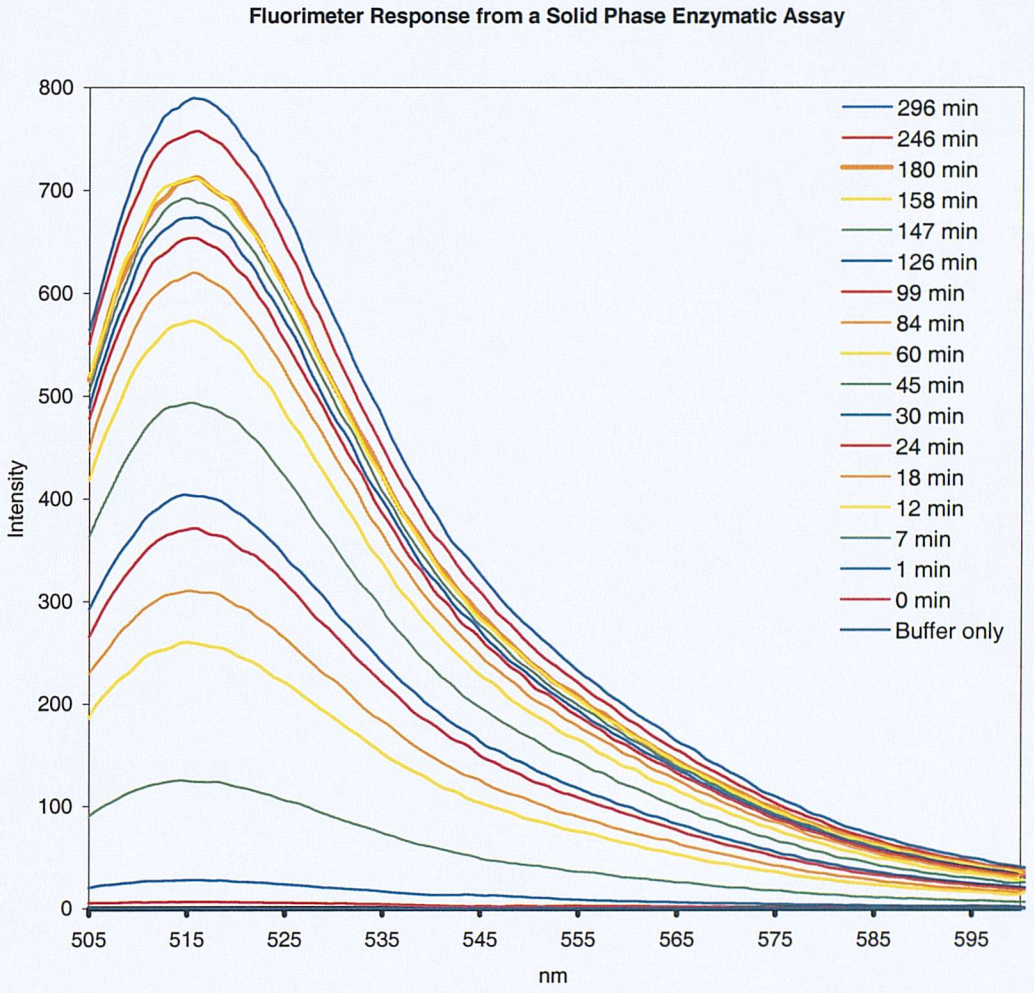


Figure 4.4.2 *Fluorescence spectra obtained from the digestion, by Subtilisin, of PEGA immobilised Fam-Ahx-Ala-Ala-Ala-5-N-propargyl-aminouracil 1-(3'-hydroxy-2'-deoxy-β-D-ribofuranosyluronic acid)* **90**

Digestion of the peptide on resin **90** using subtilisin was rapid for the first 20 minutes and by 60 minutes the reaction had reached 75% completion. The digestion then proceeded slowly and reached a maximum at approximately 160 minutes, *Figure 4.4.3*. Similar to the enzymatic assay of the solution phase analogue (*Figure 3.4.6*, section *3.4.1*, **A2**-AAA), the fluorescence spectra obtained from the solid phase assay displayed faster hydrolysis by subtilisin compared to elastase. The solution phase analogue **A2** (AAA) was not assayed using proteinase K but the solid phase assay suggested that the AAA analogue was a good substrate for proteinase K, *Figure 4.4.3*.

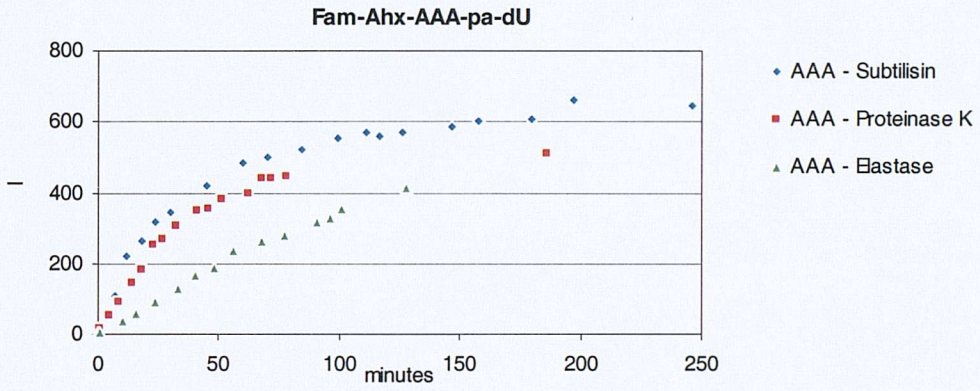


Figure 4.4.3 *Solid phase enzymatic digestion assays of resin* **90** *using subtilisin, proteinase K and elastase.*

The enzymatic digestion of the AAL analogue (resin **91**) displayed the only significant deviation from the solution phase assays carried out in Chapter 3. The fluorescence spectra indicated that elastase hydrolysed the resin bound nucleosidic peptide at an equivalent rate to subtilisin, *Figure 4.4.4*, yet elastase failed to hydrolyse the solution phase conjugate, *Figure 3.4.6* section *3.4.1*.

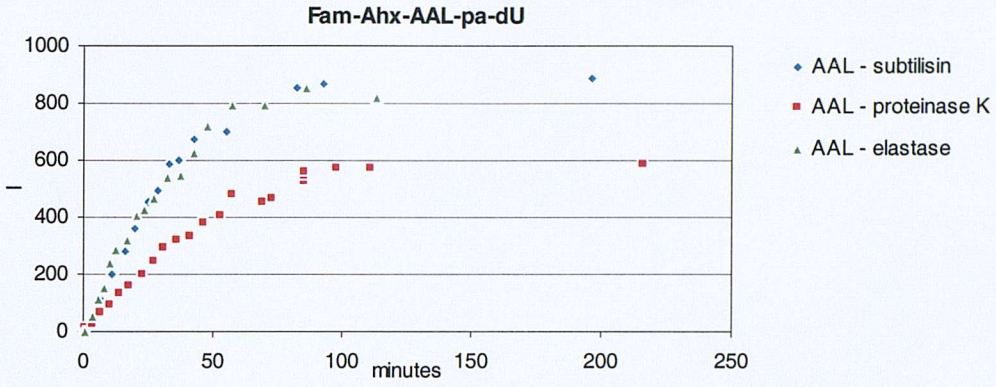


Figure 4.4.4 *Solid phase enzymatic digestion assays of resin* **91** *using subtilisin, proteinase K and elastase.*

The solution phase analogue of resin **92**, **B12** (GLS), was digested by proteinase K within five minutes. Resin **92** itself was digested by proteinase K and subtilisin at a reasonable rate whereas hydrolysis by elastase was much slower, *Figure 4.4.5*.

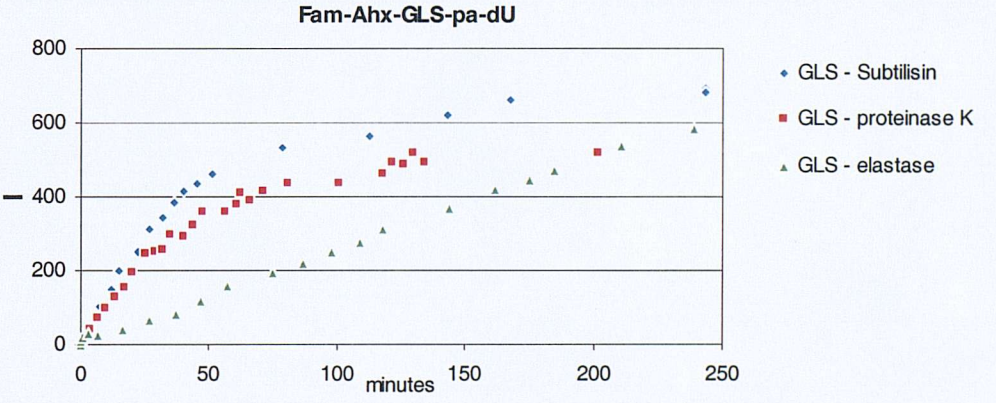


Figure 4.4.5 *Solid phase enzymatic digestion assays of resin* **92** *using subtilisin, proteinase K and elastase.*

Solution phase assay of the GAGL analogue **A20** using proteinase K displayed complete hydrolysis within four minutes. Subtilisin and elastase were not employed for the solution phase assays but resin **93** was digested by all enzymes

tested, *Figure 4.4.6*. In fact, both elastase and in particular subtilisin which were not employed for the solution screening were superior to proteinase K.

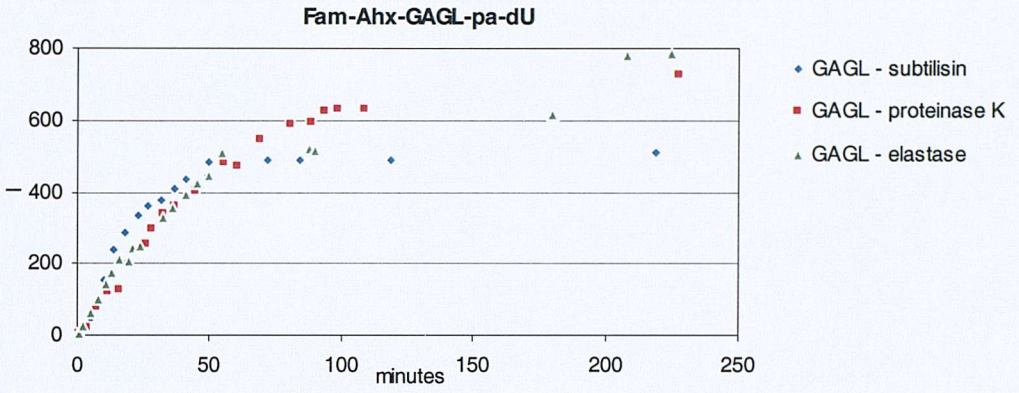


Figure 4.4.6 *Solid phase enzymatic digestion assays of resin* **93** *using subtilisin, proteinase K and elastase.*

The enzymatic hydrolyses of **90, 91, 92** and **93** followed first order kinetics. However, some deviation was observed toward the end of the hydrolysis and this was assumed to be due to the nature of the bead *i.e.* accessibility. The rate constants *k* and half lives were determined by plots of $\ln(I_\infty - I_t)$ *vs.* time.

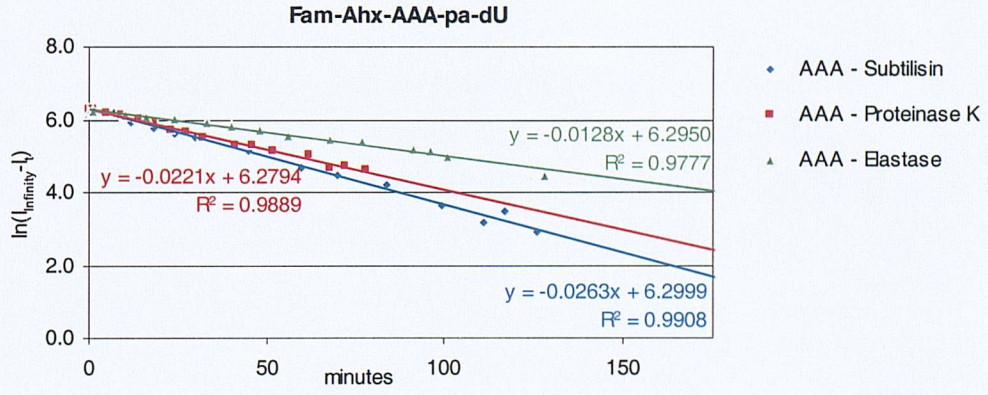


Figure 4.4.7 *The figure shows one of the four logarithmic plots produced to obtain rate constants and half lifes for the enzymatic hydrolysis of the solid phase bound peptidic nucleosides* **90**, **91**, **92** and **93**.

Rate constants k(min^{-1})

	AAA	AAL	GLS	GAGL
Subtilisin	0.0263	0.0311	0.0199	0.0595
Proteinase K	0.0221	0.0152	0.0157	0.0177
Elastase	0.0128	0.0305	0.0049	0.0350

Table 4.4.1 *Rate constants for enzymatic hydrolysis of the solid phase bound peptidic nucleosides* **90**, **91**, **92** and **93**.

Half live (minutes)

	AAA	AAL	GLS	GAGL
Subtilisin	120	110	164	53
Proteinase K	142	217	199	188
Elastase	246	110	1273	92

Table 4.4.2 *Half lives for enzymatic hydrolysis of the solid phase bound peptidic nucleosides* **90**, **91**, **92** and **93**.

Values for rate constants and half lives indicate that elastase perhaps has the most varied activity, suggesting it is the most selective enzyme. Elastase / GAGL was one of the best protease / peptide combinations while GLS was only hydrolysed very slowly indeed. Proteinase K displayed medium to good hydrolysis for all resins analysed. Subtilisin displayed good hydrolysis and the overall best protease / peptide combination was subtilisin and resin **93** (GAGL).

4.5 Conclusion

The rate constants and the half lives suggest that the slowest hydrolysis was obtained using the elastase / GLS combination and the fastest was obtained using the subtilisin / GAGL combination. Subtilisin displayed the fastest hydrolysis both in solution and on the solid phase. In solution the AAA / subtilisin combination was the best whereas on solid phase GAGL/ subtilisin was slightly better than AAA/ subtilisin. It is highly probable that the GAGL/ subtilisin would be the best combination also in solution however this was not investigated. For further research the subtilisin / GAGL combination should definitely be considered.

This study has shown that a solid phase assay of a library of labelled peptidic nucleosides can shorten the route to a successful target. It offers additional advantages to those obtained by utilising solid phase library synthesis and solution screening. The very tedious purification of final compounds was eliminated and the analysis time was immensely reduced. Changing the support to something more similar to the support that would be used in the real DNA sequencing method could possibly give even more accurate information.

In this thesis work was carried out on the uridine nucleoside but could of course be applied to the equally important cytidine, guanidine and adenosine.

5 Towards a Peptide-Nucleoside-Triphosphate

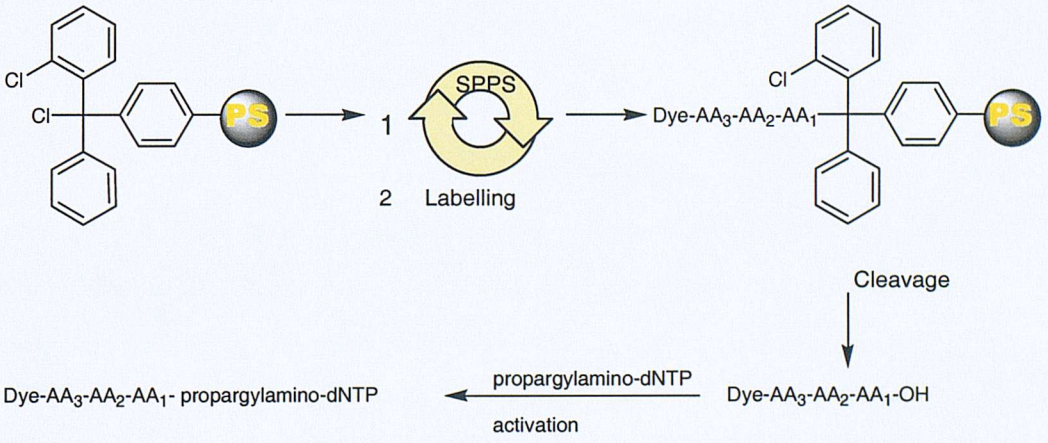
The purpose of this chapter is to describe the attempts of converting nucleosides such as those selected in chapter 3 and 4 to nucleoside triphosphates.

5.1 Introduction

5.1.1 Synthetic Strategy

It was initially thought that the modified peptidic nucleosides prepared in the previous chapters could be converted to triphosphates using literature methods.[175] However, this was not possible and a different strategy had to be developed.

Instead of synthesising the nucleoside triphosphate from the nucleoside the molecule would be divided into two parts, a reporter part and a nucleoside triphosphate part. Nucleoside triphosphates modified with a propargylamine spacer on the 5-position are described in the literature[75] and although no triphosphate synthesis is very efficient it should be a feasible route to the triphosphate. The reporter part of the molecule would be synthesised separately from the nucleoside by solid phase methods, cleaved from the solid support and coupled to the propargylamine spacer on the nucleoside triphosphate.

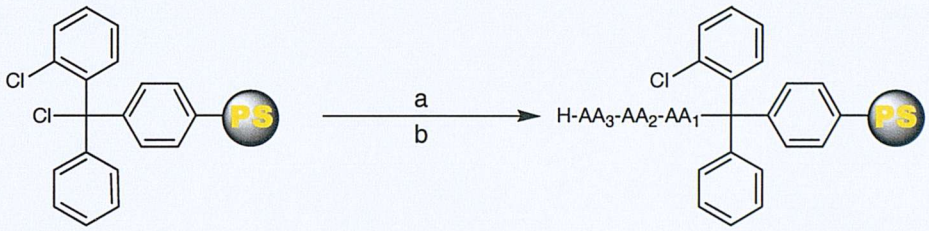


Scheme 5.1.1 *Synthesis strategy for the preparation of reporter nucleoside triphosphates containing peptidic spacers suitable for enzymatic cleavage.*

5.2 Synthesis

5.2.1 Synthesis of the Peptide Spacer

The first Fmoc amino acids in the sequences were loaded onto 2-chloro-tritylchloride polystyrene resin using dry NEt_3 and freshly distilled CH_2Cl_2.



94

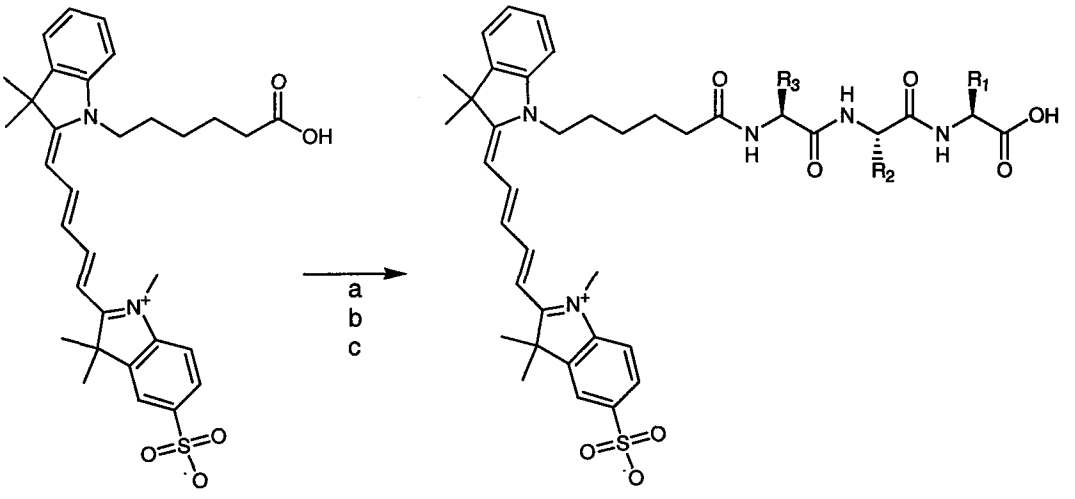
	AA_1	AA_2	AA_3
95	Ala	Ala	Ala
96	Ser	Leu	Gly

a.) Dry Net_3, dry CH_2Cl_2, AA_1 = Fmoc-Ala or Fmoc(^{t}Bu)Ser

b.) [20% Piperidine in DMF, FmocAAOH (3 eq), HOBt (3 eq), DIC (3 eq), DMF/ DCM (1:4)] x 2

Scheme 5.2.1 *Loading of the first amino acid onto 2-chlorotritylchloride polystyrene resin and subsequent SPPS gave solid phase bound peptides* **95** *and* **96**

The peptides were then extended using standard solid phase peptide synthesis. The Fmoc protecting group was removed using 20% piperidine in DMF and the coupling reactions was carried out using equimolar amounts of Fmoc-amino acid, HOBt and DIC, *Scheme 5.2.1*. Due to its superior spectral properties Cy5™ was used as the reporter moiety. Cy5™, which already possesses a six-carbon spacer, was directly coupled to the peptide using HOBt/ DIC.



97

	AA_1	AA_2	AA_3
98	Ala	Ala	Ala
99	Ser	Leu	Gly

a.) HOBt (3 eq), DIC (3 eq), DMF/ DCM (1:4), Resin bound peptide **95** or **96**

b.) 5% TFA in DCM

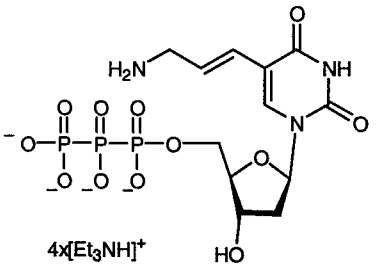
c.) neat TFA (only for compound **99**)

Scheme 5.2.2 Synthesis of **98** *Cy5-AAA and* **99** *Cy5-GLS*

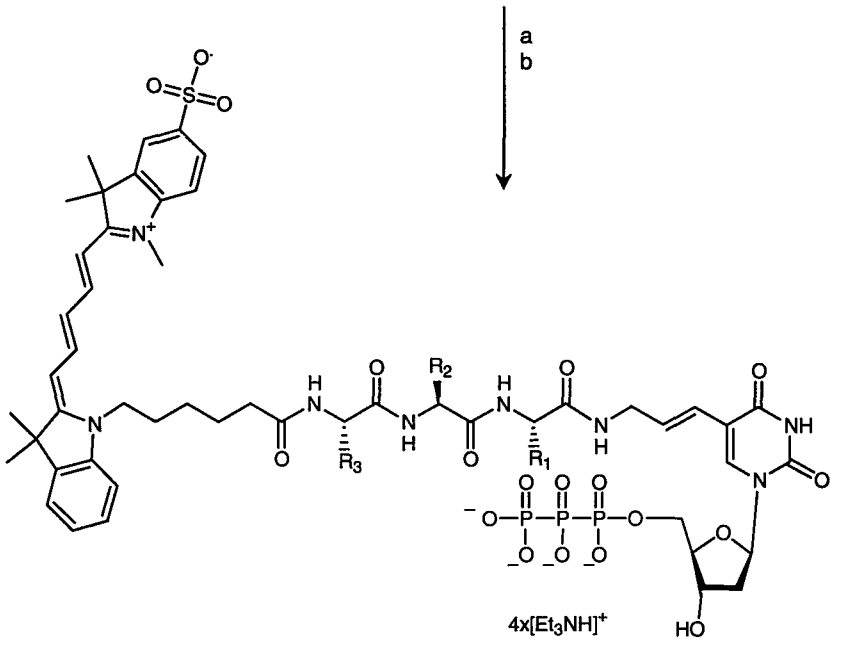
The compounds were cleaved from the resin using 5% TFA in CH_2Cl_2 and the cleaved product (Cy5™-GLS) was treated with neat TFA to remove the serine protecting group. The products were then purified by semi-preparative HPLC and the products **98** Cy5™-AAA and **99** Cy5™-GLS were lyophilised and obtained as bright blue solids in 40% and 70% yield, respectively.

5.3 Attempted Preparations of Triphosphates

The strategy was to couple the labelled peptides to the commercially available dNTP, 5-allyl amino-2-deoxy-uridine triphosphate.



100



	AA_1	AA_2	AA_3
101	Ala	Ala	Ala
102	Ser	Leu	Gly

a.) **98** (Cy5™-AAA) or **99** (Cy5™-GLS)

b.) Coupling Reagents (see discussion below)

Scheme 5.3.1 *Strategy for the preparation of reporter nucleoside triphosphates* **101** *and* **102**

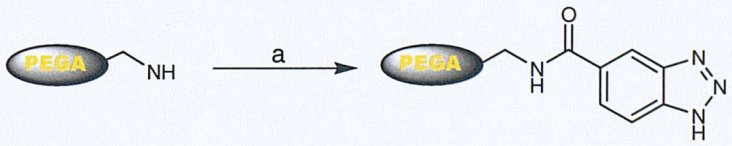
Five different coupling conditions were investigated for the coupling of the dye labelled peptide **99** to triphosphate **100** but no successful conditions were found. Reaction progress was monitored by HPLC and no new peak or only a very small new peak was observed (max 3%).

Reagents	Solvent	Temperature	HPLC observation
TSTU/ DIPEA	DMSO	Ambient	No new peak
TSTU/ DIPEA	DMSO	50°C	No new peak
TSTU/ NEt_3	DMSO	Ambient	Possible product peak 3%
HOBt/ DIC	DMSO	Ambient	Possible product peak 2%
HOBt/ DIC	DMF/ Water	Ambient	Possible product peak 1%

Table 5.3.1 *Conditions investigated for the coupling of the dye labelled peptide* **99** *to triphosphate* **100**

Factors such as large steric hindrance, between the dye and the triphosphate, and low nucleophilicity of the allyl amine relative to other amines were thought to be the reasons for failure. Since the solution phase couplings were found to be unsuccessful the use of resin bound active esters was investigated.

Benzotriazole-5-carboxylic acid was coupled to aminomethyl PEGA resin to give benzotriazole resin **103**.



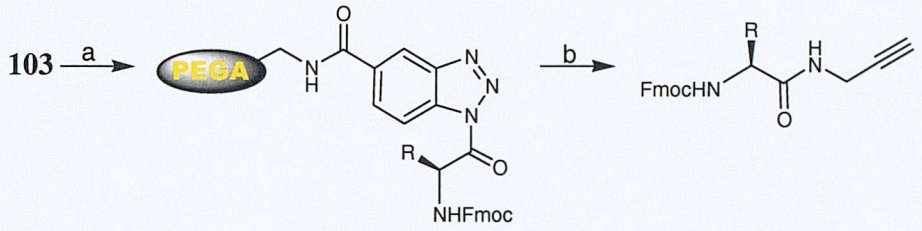
103

a.) Benzotriazole 5-carboxylic acid, HOBt, DIC, DMF

Scheme 5.3.2 *Preparation of resin* **103**

A control study was performed with all three amino acids, *i.e.* Fmoc-Ser-OH, Fmoc-Gly-OH and Fmoc-Ala-OH, using propargylamine as the nucleophile and the reactions were monitored by LCMS. Fmoc-Gly-OH and Fmoc-Ala-OH

performed well whilst Fmoc-Ser-OH displayed two products suggesting that the hydroxyl on the side chain of serine acted as a nucleophile and displaced the amino acid as well as propargylamine.



104 **105**

R = H, CH_2OH, CH_3 or $CH_2OC(CH_3)_3$

a.) DIC, DMF/ CH_2Cl_2 (1:2)

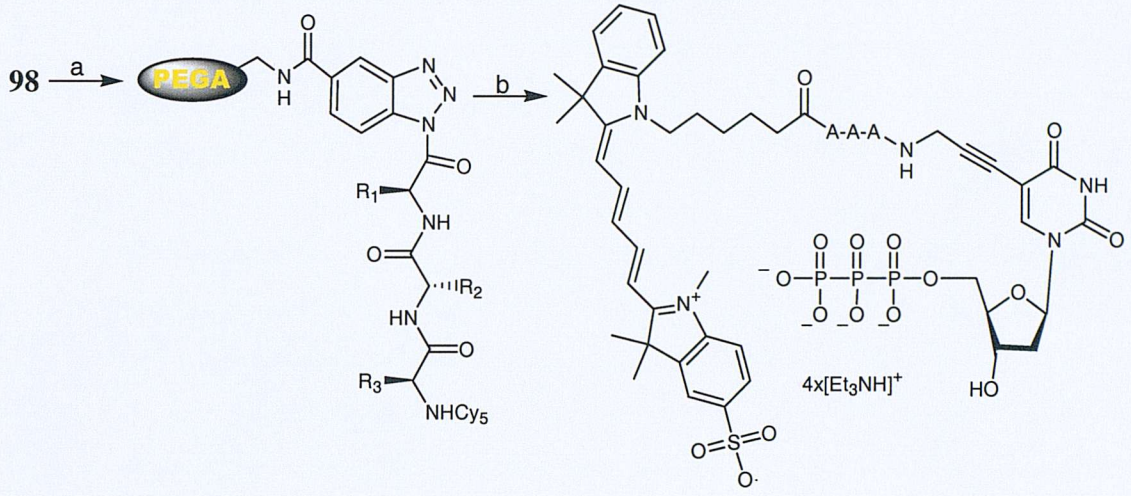
b.) Propargylamine, DMF/ Water

Scheme 5.3.3 *Studies verified that displacement was feasible*

All reactions displayed a single peak and the correct mass by LCMS analysis apart from R = CH_2OH (*i.e.* serine) which displayed an additional peak with the mass corresponding to Fmoc-Ser(FmocSer)-OH, suggesting that the serine primary alcohol was able to carry out the nucleophilic displacement. When using the protected Fmoc-Ser(^{t}Bu)-OH only one product was observed, which reinforced the hypothesis that the unprotected primary alcohol could not be used.

Coupling of 5-propargylamino-deoxyuridine triphosphate (supplied by Amersham) to Cy5™-AAA **98** using the active ester resin **103** was attempted, *Scheme 5.3.4*. Neither the loading reaction nor the displacement reaction proved to be very efficient. The resin was used in excess (3 eq.) and should have loaded most of the Cy5™AAA **98** but analysis of the reaction mixture showed that there was still a lot of unconsumed starting material present. The reaction mixture was drained off and the resin washed. 5-Propargylamino-deoxyuridine triphosphate in DMSO was added to the resin to displace the dye labelled peptide. This reaction proved to be very slow and the solution did not start to turn blue until after 2 hours. The reaction was agitated overnight and the solution was collected and analysed by HPLC. A small amount of product could be detected but in very

small quantities. (Comparison of the HPLC trace of the product to a sample of this compound prepared by Amersham suggests that the product formed is the desired product **107**).



106 **107**

a.) **103**, DMF/ CH_2Cl_2 (1:4)

b.) 5-propargylamino-deoxyuridine triphosphate, DMSO

Scheme 5.3.4 *Attempted preparation of* **107** *using PEGA active ester resin*

5.4 Conclusion

It was not possible to convert nucleosides modified with labelled peptides by conventional triphosphate synthesis methods described in the literature. Two peptides, AAA and GLS, were synthesised and labelled with Cy5™ using solid phase methodology. Attempts to couple these to a triphosphate were largely unsuccessful, although some product was observed.

It is believed that allyl- and propargyl-amines are poor nucleophiles compared to other amines and that there is a sufficient steric hindrance between the dye and the triphosphate to prevent successful reaction. A better synthetic route would be to split the synthesis once more having the peptide on an active ester resin and the dye on another and displacing one at a time. The dye would then be coupled on as a final step.

6 Overall Conclusion

The importance of modified nucleosides and nucleotides is evident and the power of combinatorial chemistry has increased as the methodology keeps expanding. Oligonucleotide synthesis aside, hitherto there has only been very few examples where combinatorial chemistry or solid phase chemistry has been utilised for synthesis of modified nucleosides and nucleotides.

In this thesis the successful preparation of a nucleside-based polystyrene resin is described. The resin proved to be well suited for SPPS when libraries of compounds were easily synthesised using this resin. The libraries were used to rapidly evaluate compounds as potential targets for use in a new type of DNA sequencing by the method of Sequencing-by-Synthesis, section *3.1.3.* Several suitable compounds were found and these are now used in the early stages of the proposed sequencing system.

To further utilise the power of solid phase chemistry an equivalent resin was synthesised on the biocompatible support PEGA. This displayed how the equivalent nucleoside compound could be evaluated whilst still attached to the support. By evaluating the compounds whilst resin bound both the synthetic route and the assay time are greatly shortened.

In this thesis work was carried out on the uridine nucleoside but could of course be applied to the equally important cytidine, guanidine and adenosine. This will be essential to complete the synthesis for the modified triphosphates required in the DNA sequencing system required.

7 Experimental

The purpose of this chapter is to provide detailed information of the experimental procedures relating to chapters 2 to 5.

7.1 Instrumentation

Infra-red spectra were recorded on a Bio-Rad FT IR spectrometer fitted with a Golden Gate Accessory. Fluorescence measurements were recorded on a Perkin-Elmer LS 50 B Luminescence spectrometer. Mass spectra were obtained on a VG Platform single quadrupole mass spectrometer in electrospray mode or on a Tofspec by MALDI-TOF. High resolution ES mass spectra were run on an FT-ICR ES-MS Bruker Apex III. Low resolution EI mass spectra were recorded using a ThermoQuest Trace MS gas chromatography mass spectrometer configured for open access operation. ^{1}H NMR and ^{13}C NMR spectra were recorded on a Bruker DPX400 (400 and 100 MHz, respectively) or on a Bruker AC300 (300 and 75 MHz, respectively) at 298 K. All chemical shifts are quoted in ppm on the δ scale using the residual protonated solvent as the internal standard. Coupling constants (*J* values) were measured in Hz. Spectra interpretation was aided by DEPT, H-H correlations, C-H correlations and long range C-H correlations.

High performance liquid chromatography was carried out on a Hewlett Packard 1100 HPLC Chemstation or Waters HPLC system. Solvents were A: H_2O (0.10% TFA) and B: MeCN (0.04% TFA).

HPLC System I

Column: Phenomenex Prodigy, C18, 150 x 4.6 mm, 5 μm. Gradient: 100% A to 100% B over 15 min. Detection: UV detection at 282 nm or 236 nm.

HPLC System II

Column: LUNA C18, 150 x 5 mm, 3 μm. Gradient: 100% A to 100% B over 15 min. Detection: UV and or ELS detection.

HPLC System III

Column: Prodigy 250 mm x 5 mm, 5μm. Gradient: 0-5 min 95% A, 5 - 30 min 95% A to 100% B, 30 - 35 min 100%B. Detection: DAD UV detection.

HPLC System IV

Semi-preparative HPLC was carried out on a Hewlett Packard 1100 HPLC Chemstation with a Phenomenex Prodigy Column 5μm. Gradient: 90% A to 90% B over 20 min.

Where necessary, loadings were determined using a HPLC calibration line obtained from a series of known concentrations of the compound.

7.1.1 Materials

Cy5™ was a gift from Amersham Biosciences. 4-Methoxytrityl chloride polystyrene resins were purchased from novabiochem. Unfunctionalised polystyrene was purchased from Advanced ChemTech. The polymer matrix was in both cases copoly(styrene-1% DVB), 200-400 mesh. Aminomethyl PEGA resin was purchased from novabiochem the bead size was 50-100 mesh (water swollen). For details of enzymes see table below:

Enzyme	Supplier	Id numbers	Unit definition
Trypsin	Promega	V511A, 13103206	1U produces a ΔA_{253} of 0.001 per min using *N*-benzoyl-L-arginine ethyl ester as substrate
Subtilisin	Sigma	P-5380 EC 3.4.21.62	1U hydrolyses casein to produce 1μmole of tyrosine per min
Thermolysin	Sigma	T-7902 EC 3.4.24.27	*As Subtilisin*
Proteinase K	Sigma	EC 3.4.21.64	1U hydrolyses urea-denatured hemoglobin to produce 1μmole tyrosin pre min
Papain	Fluka	76222, 11119/1	1U hydrolyses 1μmole of *N*-benzoyl-L-arginine ethyl ester per min
Elastase	Sigma	EC 3.4.21.36	1U hydrolyse 1μmole of N-succinyl-L-AAA-*p*-nitroanilide
Enterokinase	Sigma	EC 3.4.21.9	1U produce 1nmole of trypsin from trypsinogen

Table 7.1.1 Table of enzyme suppliers and unit definitions

7.2 General Procedures

7.2.1 General Cleavage Procedure

5% TFA in DCM (0.2 mL) was added to the resin (typically 10 mg) and left to stand for 2 minutes. DCM (0.3 mL) was added and the resin was left to stand for 3 minutes before the DCM/ TFA mixture was collected. 5% TFA in DCM (0.2 mL) was added and the resin left to stand for 2 minutes. A drop of MeOH and DCM (0.3 mL) was added and the resin was left to stand for 3 minutes before the mixture was collected.

On average the cleavage procedure was repeated three times to ensure maximum release of compound. However, an extended number of cleavage cycles were sometimes necessary.

7.2.2 Method for calculating resin loading by quantitative Fmoc determination

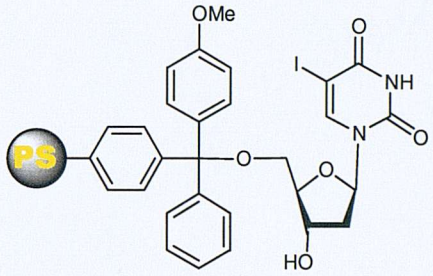
The resin sample (ca 5 mg) was added to a 25 mL volumetric flask. 20% piperidine in DMF (5 mL) was added and the mixture shaken for 15 min. The volume was made up to 25 mL with 20% piperidine in DMF and the absorbance at λ = 302 nm measured against a 20% piperidine in DMF blank at t = 20 min (value of ε for fullvene adduct is 7800).

Formula:

$$\text{Loading of Fmoc (mmol/g)} = \frac{\text{Abs}_{300} \times 25 \times 10^3}{\text{Weight of resin (mg)} \times 7800}$$

7.3 Experimental Procedures for Chapter 2

7.3.1 Preparation of 5-Iodo-5'-*O*-(4-methoxytrityl polystyrene)-2'-deoxyuridine 54



Loading Procedure

All reagents were pre-dried *in vacuo* over P_2O_5. Glassware was oven dried and flushed with N_2. 5-Iodo-2'-deoxyuridine **39** (1.275 g, 3.6 mmol, 10 eq), DMAP (4.45 mg, 0.036 mmol, 0.1 eq), 4-methoxytrityl chloride polystyrene resin **53** (805 mg, 0.36 mmol, 1 eq) and dry pyridine (35 mL) were placed in a jacketed peptide vessel and shaken for 2 days at 70°C. The reaction mixture was removed by filtration and the resin washed with: DMF (3 x 50 mL), DCM (3 x 50 mL), MeOH (3 x 50 mL), DCM (3 x 50 mL) and Et_2O (3 x 5 mL) and dried under reduced pressure.

Cleavage

a.) 4M HCl in dioxane (1 mL) was added to **54** (12 mg) and shaken for 3 hours. The resin was filtered, washed with 4M HCl in dioxane (3 x 1 mL), 4M HCl in dioxane (0.3 mL) plus MeOH (2 drops) plus DCM (0.3 mL) (3x), DCM (3 x 1 mL), MeOH (3 x 1 mL) and Et_2O (3 x 1 mL).

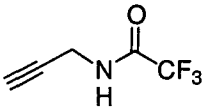
b.) 5% TFA in DCM (1 mL) was added to **54** (13.5 mg) and shaken for 3 hours. The resin was filtered, washed with 5% TFA in DCM (3 x 1 mL), 5% TFA in DCM (0.3 mL) plus MeOH (2 drops) plus DCM (0.3 mL) (3x), DCM (3 x 1 mL), MeOH (3 x 1 mL) and Et_2O (3 x 1 mL).

An average loading of 0.271 mmol/g was determined by HPLC.

HPLC System I (λ_{282}) Rt/min: 6.2

IR$_{resin}$ ν(cm^{-1}) (neat) 1685 (C=O str, s, uridine ring)

7.3.2 Preparation of *N*-trifluoroacetylpropargylamine 55



Propargylamine (2.60 g, 47.0 mmol) was added dropwise to a solution of ethyltrifluoroacetate (8.00 g, 56.3 mmol, 1.2 eq) in anhydrous methanol (50 mL), stirring under N_2 at 0°C (ice bath). The reaction mixture was allowed to warm to room temperature and stirred for a further 8 hours. The solvent was removed under reduced pressure and the residue redissolved in dichloromethane (150 mL) and extracted with saturated aqueous sodium bicarbonate (2 x 100 mL) and saturated aqueous calcium chloride (2 x 100 mL). The organic phase was dried over anhydrous magnesium sulphate and concentrated *in vacuo*. After purification of the brown oil by Kugelrohr distillation, product **55** was obtained as a colourless liquid in 61% yield (4.35 g, 28.8 mmol).

GCMS (EI): 151 (M, 100%) R_t = 3.4 min

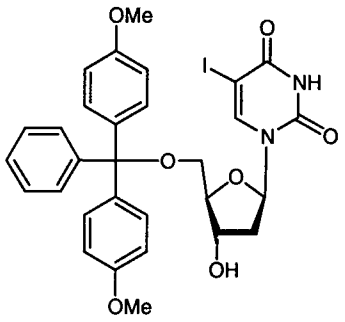
IR ν(cm^{-1}) (neat) 1704 (C=O str, s), 3300 (N-H str, m)

δ ^{1}H NMR (300 MHz, Chloroform-d_1): 2.33 (t, 1H, ≡C<u>H</u>, *J* = 3), 4.15 (dd, 2H, C<u>H</u>$_2$, *J* = 3, *J* = 2), 7.05 (broad s, 1H, N-<u>H</u>)

δ ^{13}C NMR (75 MHz, Chloroform-d_1): 29.8 (<u>C</u>H$_2$), 73.2 (≡<u>C</u>H), 77.2 (<u>C</u>≡CH), 115.7 (q, CO<u>C</u>F$_3$, *J* = 286), 157.3 (q, <u>C</u>OCF$_3$, *J* = 38)

Agrees with literature [176]

7.3.3 Preparation of 5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuridine 57



5-Iodo-2'-deoxyuridine **39** (3.00 g, 8.47 mmol, 1 eq) was dissolved in dry pyridine (30 mL), NEt_3 (1.3 mL, 9.32 mmol, 1.1 eq) was added along with 4,4'-dimethoxytrityl chloride **56** (3.16 g, 9.32 mmol, 1.1 eq) which was added portion-wise over 3 hours with DMAP (0.21 g, 1.69 mmol, 0.2 eq) being added with the first portion. The reaction was not complete after 17 hours and additional 4,4'-dimethoxytrityl chloride **56** (0.29 g, 0.85 mmol, 0.1 eq) was added over 1 hour. After 22 hours, the reaction had not reached completion so additional 4,4'-dimethoxytrityl chloride **56** (0.57 g, 1.69 mmol, 0.2 eq) was added over 2 hours. After 24 hours the reaction was complete and methanol (30 mL) was added to the reaction mixture. Solvents were removed under reduced pressure, the residue redissolved in DCM (200 mL) and extracted with saturated potassium chloride (3 x 100 mL). The organic phase was dried over anhydrous $MgSO_4$ and concentrated *in vacuo*. The crude product (7.82 g) was purified by column chromatography using silica pre-eluted with 1% NEt_3 in DCM, eluting with a gradient of 0-5% MeOH in 1% NEt_3/ DCM. Purification afforded compound **57** as a white solid (4.51 g, 6.86 mmol, 81%).

MS *m/z* (ES+) 695 (M+K, 10%)$^+$

HPLC System I (λ_{282}) Rt/min: 10.7

IR ν(cm^{-1}) (neat) 1687 (C=O str, m, uridine ring)

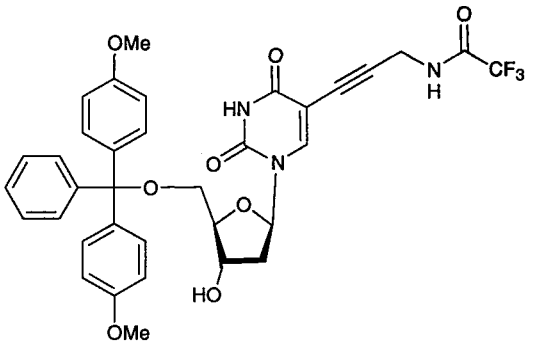
δ ^{1}H NMR (400 MHz, Pyridine-d_6): 2.54 - 2.66 (m, 1H, H-2'), 2.75 (ddd, 1H, H-2', J = 14, J = 6, J = 4), 3.56-3.66 (m, 2H, H-5'), 3.68 (s, 6H, $OC\underline{H}_3$), 4.50 - 4.56 (m, 1H, H-4'), 4.84 - 4.92 (m, 1H, H-3'), 6.84 (t, 1H, H-1', J = 7), 6.96-7.02 (m, 4H, H-DMT), 7.30 - 7.24 (m, 1H, H-

DMT), 7.38 - 7.44 (2H, H-DMT), 7.58-7.64 (m, 4H, H-DMT), 7.72-7.78 (m, 2H, H-DMT), 8.46 (s, 1H, H-6)

δ ^{13}C NMR (100 MHz, Pyridine-d_6): 44.0 (C-2') 57.6 ($O\underline{C}H_3$), 67.0 (C-5'), 73.3 (C-5), 74.0 (C-3'), 88.5 (C-1'), 89.4 (C-4'), 89.8 (C-Ar_3) 116.2, 129.6, 130.8, 131.0, 133.0, 138.7, 147.0 (C-DMT), 148.0 (C-2), 153.8 (C-6), 161.5 (C- DMT), 164.0 (C-4)

Agrees with literature [116]

7.3.4 Preparation of 5-*N*-trifluoroacetylpropargylamino-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuridine 58



5-iodo-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuridine (1.00 g, 1.52 mmol, 1 eq) **57** was dissolved in DMF (3 mL) and the solution stirred under N_2. Copper iodide (58.0 mg, 0.31 mmol, 0.2 eq.), anhydrous triethylamine (1 mL), *N*-trifluoroacetylpropargylamine **55** (690 mg, 4.57 mmol, 3 eq.) and *tetrakis* (triphenyl phosphine) palladium(0) (176 mg, 0.15 mmol, 0.1 eq.) were added sequentially. The mixture was stirred for 24 hours, DMF was removed under reduced pressure and the residue redissolved in DCM (130 mL), extracted with aqueous disodium EDTA (5% w:v solution, 2 x 100 mL) followed by aqueous sodium bisulphate (5% w:v, 2 x 100 mL). The organic phase was dried over anhydrous magnesium sulphate and concentrated under reduced pressure. The product was isolated by column chromatography using silica pre-eluted with 1% triethylamine in DCM, eluting with a gradient of 0-3% of methanol in 1% triethylamine in DCM. Removal of solvent afforded **58** as a brown solid (738 mg, 1.09 mmol, 71%).

MS *m/z* (ES+) 680 (M+H, 20%)$^+$

HPLC System I (λ_{282}) Rt/min: 10.5

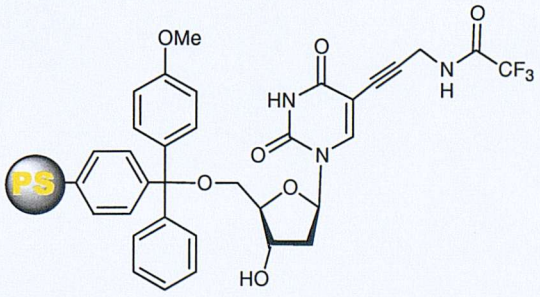
IR ν(cm^{-1}) (neat) 1710 (C=O str, s, $COCF_3$), 1687 (C=O str, m, uridine ring)

δ ^{1}H NMR (400 MHz, $CDCl_3$): 2.25 - 2.10 (m, 1H, H-2'), 2.50 - 2.40 (m, 1H, H-2'), 3.25 (s, 2H, H-5'), 3.70 (s, 6H, $OC\underline{H}_3$), 3.75 – 3.95 (m, 2H, $C{\equiv}CC\underline{H}_2$), 4.05 – 4.15 (m, 1H, H-4'), 4.45 – 4.55 (m, 1H, H-3'), 6.25 (t, 1H, H-1', J = 7), 6.72 - 6.80 (m, 4H, H DMT), 7.05 - 7.40 (m, 9H, H DMT), 8.11 (s, 1H, H-6)

δ ^{13}C NMR (100 MHz, $CDCl_3$): 29.3 ($C{\equiv}C\underline{C}H_2$), 54.3 ($O\underline{C}H_3$), 40.7 (C-2'), 54.3 ($O\underline{C}H_3$), 62.6 (C-5'), 71.0 (C-3'), 74.7 ($C{\equiv}\underline{C}CH_2$), 84.9 ($\underline{C}{\equiv}CCH_2$), 85.7 (C-4'), 86.0 ($C\text{-}Ar_3$), 97.9 (C-5), 112.3 (DMT), 113.2 ($\underline{C}F_3$, J = 287), 126.0, 127.0, 129.0 134.5 (DMT), 142.5 (C-6), 143.5 (DMT), 148.9 (C-2), 155.9, 155.5 ($\underline{C}OCF_3$, J = 38), 157.6 (DMT), 162.4 (C-4)

Agrees with literature [176]

7.3.5 Preparation of resin bound 5-*N*-trifluoroacetylpropargylamino-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuridine 59

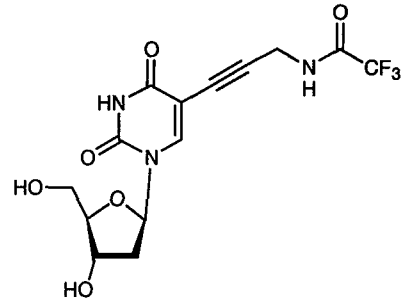


Resin **54** (100 mg, 0.003 mmol eq) was placed in a round bottomed flask and DMF (1.2 mL) was added. The mixture was stirred under N_2, copper iodide (0.11 mg, 0.6 μmol, 0.2 eq), anhydrous triethylamine (0.15 mL), *N*-trifluoroacetyl propargylamine (9.10 mg, 60.0 μmol, 20 eq) and *tetrakis* (triphenylphosphine)

palladium(0) (1.73 mg, 1.50 μmol, 0.5 eq) were added in sequence. The mixture was stirred at 60°C under N_2 overnight. The resin mixture was transferred into a peptide vessel and washed with DMF (3 x 4 mL), DCM (3 x 4 mL), MeOH (3 x 4 mL), DCM (3 x 4 mL) and Et_2O (3 x 4 mL). The resin was dried under vacuum.

IR $\nu(cm^{-1})$ (neat) 1718 (C=O str, w, $COCF_3$), 1670 (C=O str, m, uridine ring)

7.3.6 Preparation of 5-*N*-trifluoroacetylpropargylamino-2'-deoxyuridine 61



5% TFA in DCM was added to resin **59** (104 mg, corresponding to 0.028 mmol based on starting resin **54**) and shaken for 24 hours. The resin was washed with 5% TFA in DCM (3 x 5 mL), 5% TFA in DCM (2 mL) plus MeOH (2 drops) plus DCM (2 mL) (5x), DCM (3 x 5 mL), 5% TFA in DCM (2 mL) plus MeOH (2 drops) plus DCM (2 mL) (5x) and DCM (3 x 5 mL). The filtrate was collected and the solvent removed under reduced pressure. The crude product was purified by semi-preparative HPLC. The three fractions (from three injections) were combined and, upon freeze-drying, gave compound **61** an off-white solid (2 mg, 20 %).

MS *m/z* (ES+) 378 ($M+H^+$, 70%)

HPLC System I (λ_{282}) Rt/min: 6.6

IR $\nu(cm^{-1})$ (neat) 1718 (C=O str, s, $COCF_3$), 1670 (C=O str, m, uridine ring)

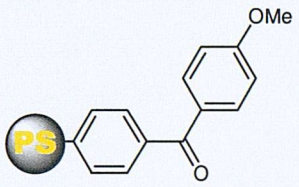
δ ^{1}H NMR (400 MHz, DMSO-d_6): 2.24 - 2.08 (m, 2H, H-2'), 3.64 (m, 2H, H-5'), 3.86 (m, 1H, H-4'), 4.29 (m, 3H, C≡CC$\underline{H_2}$, H-3'), 5.14 (m,

1H, H-5'-OH), 5.29 (m, 1H, H-3'-OH), 6.17 (t, 1H, J = 7, H-1'), 8.25 (s, 1H, H-6)

δ ^{13}C NMR (100 MHz, DMSO-d_6): 30.2 (C≡C<u>C</u>H$_2$), 40.7 (C-2'), 61.7 (C-5'), 70.9 (C-3'), 76.1 (C≡<u>C</u>CH$_2$), 85.5 (C-1'), 88.2, 88.4 (C-4', <u>C</u>≡CCH$_2$), 98.4 (C-5), 144.8 (C-6), 150.2 (C-2), 162.3 (C-4) (CF_3 and $COCF_3$ could not be observed probably due to small amount of material)

Agrees with literature[177]

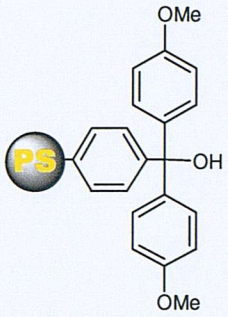
7.3.7 Preparation of *p*-methoxybenzoyl polystyrene 64



Polystyrene (1% DVB, 200-400 mesh, 5.00 g), 4-methoxybenzoyl chloride (0.85 g, 5.00 mmol) and $AlCl_3$ (0.67 g, 5.00 mmol) were refluxed in CS_2 (40 mL) under N_2 overnight. The solvent was removed by filtration and the resin was washed with THF/ water (3 x 50 mL), THF (1 x 50 mL), CH_2Cl_2 (5 x 50 mL), MeOH (5 x 50 mL) and Et_2O (3 x 5 mL). The light yellow resin was dried under reduced pressure.

IR ν(cm^{-1}) (neat) 1651 (C=O str, m), 1600 (ArC=C str, s), 1250 (ArC-O-CH_3, s)

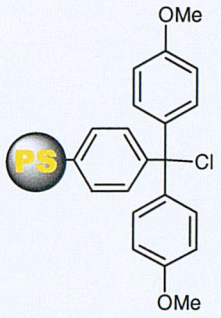
7.3.8 Preparation of dimethoxytrityl polystyrene 65



To a suspension of **64** (5.27 g) in dry THF (60 mL) was added 4-methoxyphenylmagnesium bromide (10.5 mL, 0.5 M solution in dry THF, 5.27 mmol) over 1 hour. The reaction was shaken for an additional 1 hour under N_2. THF/ water (1:1) (20 mL) was added to the reaction, the solvent was removed by filtration and the resin was washed with: water (10 x 20 mL), ethanol (5 x 10 mL), dioxane (5 x 10 mL), ether (5 x 10 mL) and dried under reduced pressure. The slightly orange resin showed a deep red colouration, characteristic of the trityl cation, when in contact with TFA.

IR $\nu(cm^{-1})$ (neat) 3226 (O-H str, broad very strong) 1648 (C=O str, m), 1609 (ArC=C str, m), 1258 (ArC-O-CH_3, m)

7.3.9 Preparation of dimethoxytrityl chloride polystyrene 66



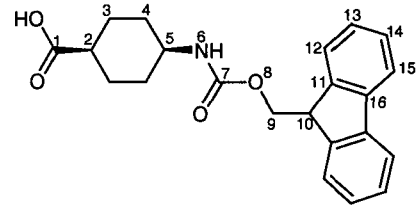
Resin **65** (5.80 g, 5.8 mmol) placed in a peptide vessel under N_2, was swollen in dry CH_2Cl_2 (55 mL) and cooled with water. Thionyl chloride (0.51 mL, 0.83 g, 7.0 mmol, 1.2 eq) and dry pyridine (1.1 mL, 1.10 g, 13.9 mmol, 2.4 eq) were added by syringe addition. The resin was agitated for 4 hours, the solvent was removed by filtration and the resin was washed with dry CH_2Cl_2 (6 x 50mL) and dried under reduced pressure.

IR analysis indicated a large number of OH groups free on the resin and therefore the procedure was repeated. Identical quantities were used but the apparatus were changed. The resin was placed in a round bottomed flask fitted with a condenser. Thionyl chloride and pyridine was added by syringe followed by reflux for 3 hours. The slightly orange resin was washed with CH_2Cl_2 as described above and dried in a desiccator over P_2O_5.

IR $\nu(cm^{-1})$ (neat) 1608 (ArC=C str, w), 1253 (ArC-O-CH_3, s)

The loading was determined as being 0.125 mmol/ g., by loading/ cleavage of 5-iodo-2'-deoxyuridine and proportion relation HPLC analysis.

7.3.10 Preparation of *cis*-4-(Fmoc-amino)-cyclohexanecarboxylic acid 70



Amino cyclohexyl carboxylic acid (200 mg, 1.40 mmol) was dissolved in a mixture of dioxane (10 ml) and 10% aq. Na_2CO_3 (15 ml) and cooled to 0°C with vigorous stirring. 9-Fluorenyl methyl chloroformate (361 mg, 1.40 mmol) was then added in small portions. The mixture was then stirred overnight in room temperature. Water (100 ml) was added and the pH was adjusted to pH 2 by addition of 6M HCl. The white precipitate was extracted into ethyl acetate (150 ml). The ethyl acetate layer was washed with dilute HCl (2 x 25 ml), H_2O (2 x 25 ml) and sat. NaCl aq (2 x 25 ml), dried ($MgSO_4$) and concentrated *in vacuo.* Purification by silica column chromatography, using ethyl acetate / Hexane (1:4) as eluent, afforded **70** as a glassy solid (350 mg, 0.958 mmol, 68%).

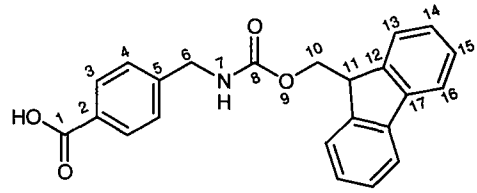
MS *m/z* (ES+) 366 ($M+H^+$, 20%)

δ ^{1}H NMR (300 MHz, d_6-DMSO): 1.40 - 1.70 (m, 6H, H-3, H-4), 1.90 (m, 2H, H-3), 2.38 (m, 1H, H-2), 3.40 (m, 1H, H-5), 4.23 (m, 3H, H-9, H-10), 7.31 (t, 2H, Fmoc-H, J = 8), 7.41 (t, 2H, Fmoc-H, J = 8), 7.71 (d, 2H, Fmoc-H, J = 7), 7.88 (d, 2H, Fmoc-H, J = 7)

δ ^{13}C NMR (75 MHz, $CDCl_3$): 24.9 (C-3), 29.3 (C-4), 30.8 (C-10), 46.8 (C-2), 47.8 (C-5), 65.3 (C-9), 120.2 (C-14), 125.4 (C-13), 127.1 (C-15), 127.7 (C-12), 140.8 (C-16), 144.0 (C-11), 155.5 (C-7), 176.1 (C-1)

Agrees with literature[135]

7.3.11 Preparation of 4-Fmoc-aminomethyl-benzoic acid 71



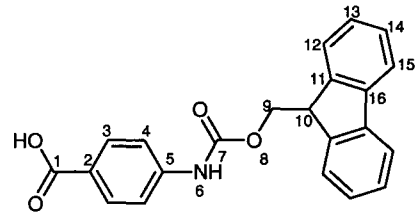
4-Aminomethyl benzoic acid (1.5 g, 10 mmol) was dissolved in a mixture of dioxane (25 ml) and 10% sodium bicarbonate (100 ml) and cooled to 0°C with vigorous stirring. A solution of 9-fluorenylmethyl chloroformate (2.72 g, 10 mmol) in dioxane (25 ml) was added dropwise over 20 minutes. The ice bath was removed and the suspension was allowed to stir for 3 days. Water (200 ml) was added and the mixture was extracted with ether (3 x 80 ml). The aqueous layer was cooled to 0°C and carefully acidified with concentrated HCl to pH 3. The resulting white precipitate was extracted into ethyl acetate (5 x 100 ml). The organic layer was dried with $MgSO_4$, filtered and the solvent was removed under reduced pressure to afford compound **71** as a white solid (2.60 g, 7.9 mmol, 79%).

δ ^{1}H NMR (300 MHz, d_6-DMSO): 4.25 (m, 3H, H-10, H-11), 4.39 (d, 2H, H-10, J = 7), 7. 25 - 7. 45 (m, 6H, Ar-H), 7.70 (d, 2H, Ar-H, J = 7), 7.80 – 8.00 (m, 4H, Ar-H)

δ ^{13}C NMR (75 MHz, $CDCl_3$): 43.6 (C-11), 46.9 (C-6), 65.4 (C-10), 120.2, 125.2, 127.0, 127.0, 127.1, 127.7, 129.5 (C-3, C-13, C-2, C-14, C-16, C-4, C-15), 140.9 (C-17), 143.9 (C-12), 145.0 (C-5), 156.9 (C-8)

Agrees with literature[136]

7.3.12 Preparation of 4-Fmoc-aminobenzoic acid 72



A solution of 4-aminobenzoic acid (1.37 g, 10 mmol) in 10% Na_2CO_3 (25 ml) and dioxane (15 ml) was cooled to 0°C with stirring. 9-Fluorenylmethyl chloroformate (2.60 g, 10 mmol, 1.02 eq) was then added in small portions. The mixture was stirred for 3 hours at 0°C and then overnight at room temperature. The white suspension was transferred into a separating funnel. Water (800 ml) was added and the resulting solution was extracted with ether (3 x 100 ml). The aqueous phase was cooled to 0°C under vigorous stirring and pH was adjusted to 3 by addition of concentrated HCl. The resulting white precipitate was extracted into ethyl acetate (8 x 100 ml). The combined organic layers were washed with water (4 x 100 ml), dried ($MgSO_4$), filtered, and concentrated *in vacuo* to yield a white solid (1.27 g, 3.53 mmol, 36 %).

δ ^{1}H NMR (300 MHz, d_6-DMSO): 4.33 (t, 1H, H-10), 4.55 (d, 2H, H-9), 7.30 – 8.00 (m, 12H, Ar-H), 10.10 (br s, 1H, N-H)
δ ^{13}C NMR (75 MHz, $CDCl_3$): 46.7 (C-10), 65.9 (C-9), 117.5, 120.3, 124.5, 125.2, 127.2, 127.8, 130.5 (C-4, C-2, C-14, C-15, C-13, C-12, C-3), 140.9 (C-16), 143.3 (C-5), 143.8 (C-11), 153.3 (C-7), 167.0 (C-1)

Agrees with literature[137]

7.3.13 Preparation of the 27 Membered Library

The spacers were assembled by standard solid phase peptide synthesis. Reactant solutions were made up using three equivalents (to resin reactive sites) of Fmoc-

amino acid (**67**, **68**, **69**, **70**, **71** or **72**) / HOBt / DIC in CH_2Cl_2 / DMF (6:1, 1 mL per reaction vessel). For the first coupling resin **62** was distributed between three filter vessels, (batch 1, 153 mg, 0.073 mmol), (batch 2, 157 mg, 0.075 mmol) and (batch 3, 131 mg, 0.063 mmol). To batch 1, 2 and 3 was added reactant solution containing Fmoc-amino acid **67**, **68**, **69**, respectively (1 mL per vessel). The resin shaken at room temperature overnight and then washed with DMF (5 x 1 ml), CH_2Cl_2 (5 x 1 ml), MeOH (5 x 1 ml) and Et_2O (5 x 1 ml) and thereafter dried under reduced pressure. All filter vessels were treated with 20% piperidine in DMF for 1 hour and washed with DMF (5 x 1 ml), CH_2Cl_2 (5 x 1 ml), MeOH (5 x 1 ml) and Et_2O (5 x 1 ml) and thereafter dried under reduced pressure. Each batch was then split in three equal portions and the procedure was repeated using the cyclic spacer moieties. All steps were monitored by HPLC and mass spectrometry.

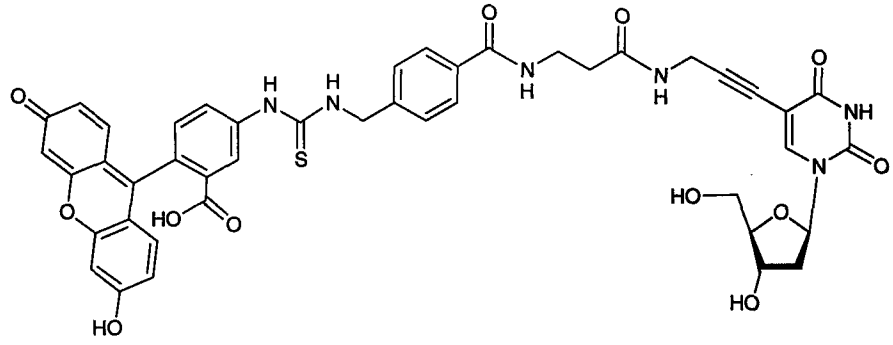
Library analysis

	Coupling spacer	*HPLC Rt/ min*	*HPLC purity %*	*MS $(M+H)^+$ *$(M+Na)^+$*
1	Fmoc-γ-Abu-OH	8.5	86	589
2	Fmoc-β-Ala-OH	8.4	78	575
3	*N*-Fmoc-Caproic acid	9.1	93	617
1.1	Fmoc-*cis*-ACCA	8.8	98	714
1.2	*N*-Fmoc-methylbenzoic acid	9.0	94	*744
1.3	*N*-Fmoc-benzoic acid	9.3	96	*730
2.1	Fmoc-*cis*-ACCA	8.7	100	*722
2.2	*N*-Fmoc-methylbenzoic acid	8.9	96	*730
2.3	*N*-Fmoc-benzoic acid	9.1	95	*716
3.1	Fmoc-*cis*-ACCA	9.1	96	*764
3.2	*N*-Fmoc-methylbenzoic acid	9.3	78	*772
3.3	*N*-Fmoc-benzoic acid	9.6	88	*758

Each batch of resin was again distributed between three filter vessels, making a total of 27 vessels. Three reactant solutions containing NEt_3 (2 eq.) and FITC, RITC, EITC (1.1 eq.), respectively in DMF (9 mL, respectively) were prepared. The solutions was distributed to the appropriate reaction vessel (1 mL to each vessel) and the resin shaken at room temperature overnight and then washed with

DMF (5 x 1 ml), CH_2Cl_2 (5 x 1 ml), MeOH (5 x 1 ml) and Et_2O (5 x 1 ml) and dried under reduced pressure.

Synthesis of one member **73** was repeated on a larger scale to allow characterisation. Synthesis proceeded quantitatively from resin **62** and gave after cleavage and repeated precipitation from ether a yellow solid.



73

MS *m/z* (ES^-) 873.2 ($M-H^-$, 30%)

HRMS $C_{44}H_{39}N_6O_{12}S$: 875.2341 (calc.); 875.2350 (found)

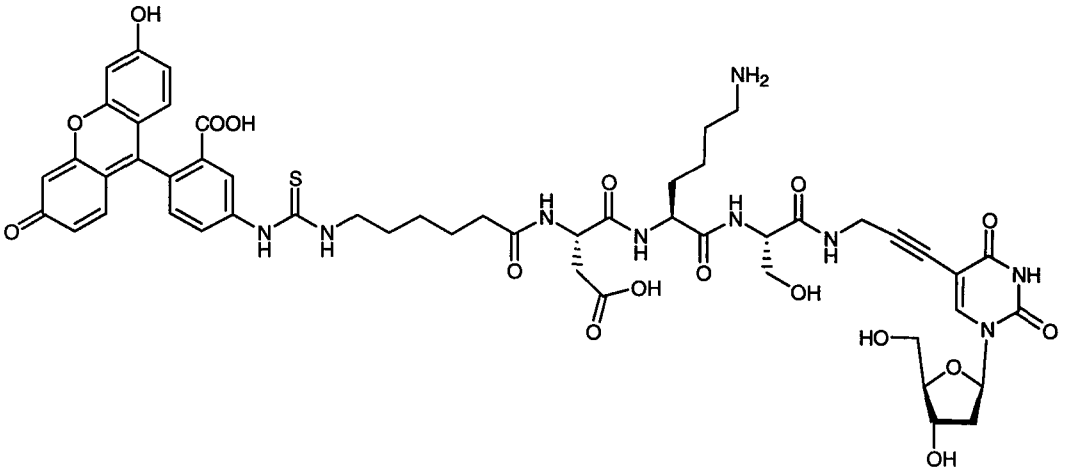
HPLC System II Rt/min: 6.7; Purity ($UV_{\lambda=282}$) 95%, ($UV_{\lambda=440}$) 95%

δ ^{1}H NMR (400 MHz, DMSO-d_6): 2.02 - 2.08 (m, 2H, H-2'), 2.32 - 2.39 (m, 2H, β-Ala-α-C$\underline{H}_2$), 3.32 - 3.58 (m, 4H, β-Ala-β-C$\underline{H}_2$, H-5'), 3.70 – 3.76 (m, 2H, H-4'), 4.00 - 4.60 (m, 2H, C≡CC$\underline{H}_2$), 4.12 - 4.20 (m, 1H, H-3'), 4.74 – 4.82 (m, 2H, Ar-C$\underline{H}_2$), 4.92 – 5.60 (m, 1H, O$\underline{H}$-5'), 5.14 – 5.24 (m, 1H, O$\underline{H}$-3'), 6.02 – 6.08 (m, 1H, H-1'), 6.43 – 6.63 (m, 6H, H-Ar), 7.10 – 7.16 (d, 1H, H-Ar, J = 8), 7.32 – 7.39 (d, 2H, H-Ar, J = 8), 7.68 – 7.76 (m, 3H, H-Ar), 8.10 (s, 1H, H-6), 8.17 - 8.22 (m, 0.7H, H-Ar), 8.59 – 8.61 (m, 0.3H, H-Ar)

δ ^{13}C NMR (100 MHz, DMSO-d_6): 26.4, 26.9, 32.9, 33.8, 38.0, 58.9, 68.1, 68.2, 72.2, 81.0, 82.7, 83.6, 85.5, 85.8, 87.4, 96.1, 100.1, 107.7, 109.6, 110.3, 114.8, 117.9, 121.9, 124.6, 124.9, 125.1, 126.9, 127.7, 131.0, 139.0, 139.9, 141.6, 144.7, 145.4, 147.3, 149.9, 157.3, 159.0, 159.5, 164.1, 166.4, 168.4, 168.2, 178.9, 192.1

7.4 Experimental Procedures for Chapter 3

7.4.1 Solid Phase Preparation of Fam-Ahx-Asp-Lys-Ser-propargylamino-2'-deoxyuridine – Library Control Compound



Resin hydrolysis

1M aqueous KOH / dioxane (1:2, 40 mL) was added to resin **59** (ten tea-bags containing 10-15 mg of resin per tea-bag) and the mixture was shaken overnight. The resin was washed with dioxane : water (1:1) (20 mL), dioxane (5 x 20 mL), MeOH (5 x 20 mL), DCM (5 x 20 mL), DMF (5 x 20 mL), DCM (5 x 20 mL) and MeOH (5 x 20 mL).

General Amino Acid Coupling Procedure

DIC was added to equimolar amounts of Fmoc-AA-OH, HOBt (concentration of reactants were always kept at 0.2 M) in DCM/ DMF (3:2) (20 mL for 10 tea-bags and 2 mL when only one tea-bag was left). After 5 to 10 minutes activation time the reaction mixture was added to the pre-swollen (DCM) resin **60**. The resin was shaken for 3 hours and the resin was washed with DMF (5 x 10 mL), DCM (5 x 10 mL), MeOH (5 x 10 mL) and Et_2O (5 x 10 mL).

General Deprotection Procedure

To the pre-swollen resin (DMF) was added a solution of 20% Piperidine in DMF (20 mL for 10 tea-bags and down to 2 mL when only one tea-bag was left) and the

resin was shaken for 1 hour. The resin was washed with DMF (5 x 10 mL), DCM (5 x 10 mL), MeOH (5 x 10 mL), Et_2O (5 x 10 mL), MeOH (2 x 5 mL), DCM (2 x 5 mL) and DMF (2 x 5 mL).

The procedures were repeated for all three amino acid residues, Fmoc-Ser(^{t}Bu)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Asp(O^{t}Bu)-OH and Fmoc-Ahx-OH.

Fluorophore Coupling Procedure

A solution of fluorescein isothiocyanate isomer I (95 mg, 0.24 mmol, 10 eq) and NEt_3 (34 µL, 25 mg, 0.24 mmol, 10 eq) in DMF (1 mL) was added to resin **79** (54 mg). The resin was shaken overnight and then washed with DMF (5 x 1 mL), DCM (5 x 1 mL), MeOH (5 x 1 mL) and Et_2O (5 x 1 mL). The compound was cleaved off the resin using the standard cleavage procedure, *Section 7.2.1*, and purified by HPLC. Each step was analysed by MS and HPLC.

	Fmoc-protected		deprotected	
	Rt	$[M+H]^+$	Rt	$[M+H]^+$
Fmoc-Ser(^{t}Bu)-propargylamino-2'-dU	27.6	647.2	15.1	425.2
Fmoc-Lys(Boc)-Ser(^{t}Bu)-propargylamino-2'-dU	24.1	875.2	15.0	653.3
Fmoc-Asp(O^{t}Bu)-Lys(Boc)-Ser(^{t}Bu)-propargylamino-2'-dU	33.2	1046.2	22.4	824.2
Fmoc-Ahx-Asp(O^{t}Bu)-Lys(Boc)-Ser(^{t}Bu)-propargylamino-2'-dU	32.5	1159.2	22.6	937.2
Fam-Ahx-Asp(O^{t}Bu)-Lys(Boc)-Ser(^{t}Bu)-propargylamino-2'-dU	27.6	1326.1	-	-
Fam-Ahx-Asp-Lys-Ser-propargylamino-2'-dU	-	-	19.0	1114.1

MS *m/z* (ES+) 1114.1 ($M+H^+$, 5%)

MS *m/z* (ES+) 1114.2 ($M+H^+$, 60%) 558.0 ((M+2)/2, 100%)

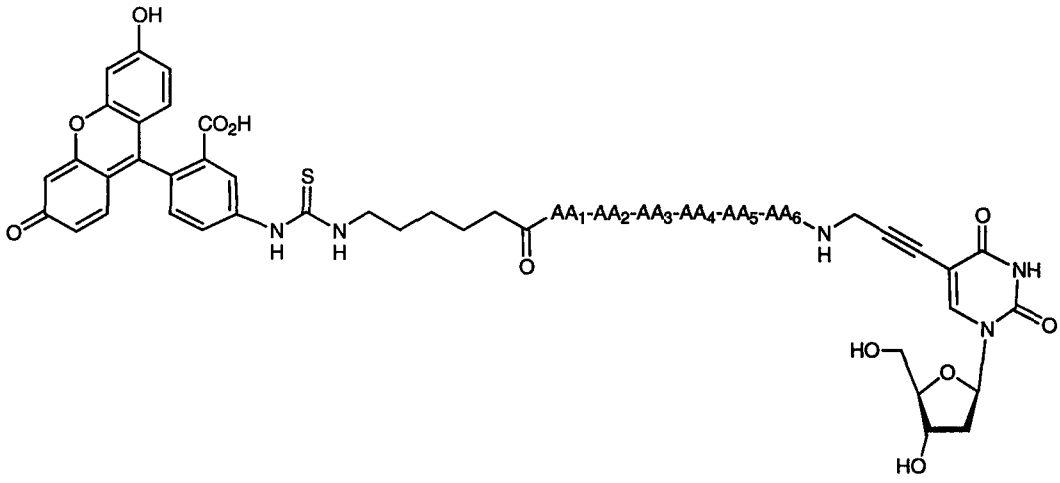
HPLC System II Rt/min: 6.0; Purity ($UV_{\lambda=282}$) 82%, ($UV_{\lambda=440}$) 88%

δ ^{1}H NMR (300 MHz, MeOH): 1.00 – 1.80 (m, 12H, CH_2-Lys, CH_2-Ahx), 2.00 – 2.40 (m, 4H, H-2', $COC\underline{H}_2$), 2.55 – 2.75 and 2.75 – 2.90 (m, 4H, CH_2-Lys, CH_2-Asp), 3.40 – 4.30 (m, 9H, H-4', H-5', $C\underline{H}_2$-Asp, $C\underline{H}_2$-Lys, $COCH_2CH_2CH_2CH_2C\underline{H}_2$), 4.30 – 4.40 (m, 3H,

CH-Ser, CH-Lys, CH-Asp), 4.50 – 4.65 (m, 1H, H-3'), 6.05 – 6.20 (m, 1H, H-1'), 6.40 – 6.50 (m, 2H, H-Ar), 6.55 – 6.65 (m, 4H, H-Ar), 7.10 (d, 1H, H-Ar, J = 8), 7.60 – 7.70 (m, 1H, H-Ar), 8.05 (s, 1H, H-6), 8.20 - 8.22 (m, 0.5H, H-Ar), 8.78 - 8.80 (m, 0.5H, H-Ar)

δ ^{13}C NMR (100 MHz, DMSO-d_6): 22.4, 25.5, 26.6, 27.0, 28.6, 29.5, 31.6, 34.5, 35.6, 36.4, 44.2, 50.1, 52.8, 55.5, 61.6, 62.1, 70.8, 75.1, 83.5, 85.3, 86.2, 88.2, 88.5, 98.6, 102.8, 110.2, 112.1, 113.1, 116.8, 124.4, 126.9, 129.5, 129.8, 142.1, 144.3, 147.3, 149.9, 150.1, 152.4, 158.6, 160.0, 161.8, 162.2, 169.1, 170.1, 170.2, 171.4, 171.7, 172.3, 173.1, 180.9, 194.7

7.4.2 **Parallel Solid Phase Library Synthesis:** Solid Phase Preparation of 57 Nucleoside-Peptide-Dye Conjugates



General Library Amino Acid Coupling Procedure

Resin **60** was distributed between 22 (for library A, 16 for B and 19 for C) disposable filter vessels (10-20 mg per vessel). DCM was added to swell the resin and the excess was drained off. A microtag was added to the vessel for identification purposes. DIC was added to equimolar solutions of Fmoc-AA-OH and HOBt (0.2 M) in DCM/ DMF. The appropriate solution of the activated amino acid was added to the reaction vessels. The reaction vessels were stoppered and placed horizontally on a flat bed shaker. The vessels were agitated for 3-3.5 hours. The vessels were drained of the reaction mixture and the resin was washed

with DMF (5 x 1 mL), DCM (5 x 1 mL), MeOH (5 x 1 mL), Et_2O (3 x 1 mL) (volume of washings are shown per filter tube).

General Library Fmoc Deprotection Procedure

DMF was added to the vessels to allow the resin to swell. After draining 20% piperidine in DMF was added and the vessels left to agitate for 1 hour. The deprotection mixture was drained off and the resin was washed with DMF (5 x 1 mL), DCM (5 x 1 mL), MeOH (5 x 1 mL), Et_2O (3 x 1 mL).

General Library Fluorophore Coupling Procedure

As described for library control compound section *7.4.1.*

General Library Cleavage Procedure

For library B cleavage was performed according to the general procedure outlined in section *7.2.1.* For library A and C the general procedure was repeated 10 times for complete cleavage.

General Library Purification Procedure

Each compound was purified by semi-preparative RP-HPLC (System IV).

HPLC Data on Library A

A	Compound	Rt/ min	% Purity		% Purity after HPLC purification		
		HPLC	282 nm	440 nm	ELSD	282 nm	440 nm
1	Fam-Ahx-AA-propargylamino-2'-dU	22.2	71	-	-	-	-
2	Fam-Ahx-AAA-propargylamino-2'-dU	14.9	80	66	100	100	100
3	Fam-Ahx-AAL-propargylamino-2'-dU	15.7	64	62	100	100	100
4	Fam-Ahx-AAAA-propargylamino-2'-dU	14.9	53	78	100	79	76
5	Fam-Ahx-AAPA-propargylamino-2'-dU	14.9	61	92	100	100	93
6	Fam-Ahx-AAPV-propargylamino-2'-dU	13.7	55	61	100	84	86

A	Compound	Rt/ min	% Purity		% Purity after HPLC purification		
		HPLC	282 nm	440 nm	ELSD	282 nm	440 nm
7	Fam-Ahx-AAPL-propargylamino-2'-dU	14.2	50	57	100	100	100
8	Fam-Ahx-AIPM-propargylamino-2'-dU	14.7	37	55	100	100	100
9	Fam-Ahx-AAPK-propargylamino-2'-dU	16.1	76	83	100	100	77
10	Fam-Ahx-VLK-propargylamino-2'-dU	17.5	48	72	100	100	77
11	Fam-Ahx-GK-propargylamino-2'-dU	16.2	43	46	100	90	87
12	Fam-Ahx-AAK-propargylamino-2'-dU	16.3	27	40	100	86	79
13	Fam-Ahx-ALK-propargylamino-2'-dU	14.8	54	64	100	100	100
14	Fam-Ahx-GGL-propargylamino-2'-dU*	15.6	70	82	100	100	100
15	Fam-Ahx-GL-propargylamino-2'-dU	16.0	90	100	100	100	92
16	Fam-Ahx-AAG-propargylamino-2'-dU	15.1	77	84	100	100	90
17	Fam-Ahx-GGGL-propargylamino-2'-dU**	15.4	89	100	100	100	73
18	Fam-Ahx-AGL-propargylamino-2'-dU	15.9	78	100	100	100	98
19	Fam-Ahx-LGL-propargylamino-2'-dU	16.8	100	100	100	100	54
20	Fam-Ahx-GAGL-propargylamino-2'-dU	10.3	34	41	100	94	93
21	Fam-Ahx-DDDDL-propargylamino-2'-dU	15.2	72	100	100	100	100
22	Fam-Ahx-DDDDK-propargylamino-2'-dU	14.4	68	100	100	100	100

* broad peak

** not representative (broad base and split top)

Mass Spec Data on Library A

A	Compound	MS ES$^+$		MS ES$^-$	
		m/z	Ion	m/z	Ion
1	Fam-Ahx-AA-propargylamino-2'-dU	-	-	-	-
2	Fam-Ahx-AAA-propargylamino-2'-dU*	997.6	$[M+H]^+$	-	-
3	Fam-Ahx-AAL-propargylamino-2'-dU	1061.6	$[M+Na]^+$	1037.6	$[M-H]^-$
4	Fam-Ahx-AAAA-propargylamino-2'-dU	1068.4	$[M+H]^+$	1066.4	$[M-H]^-$
5	Fam-Ahx-AAPA-propargylamino-2'-dU	1116.6	$[M+Na]^+$	1092.7	$[M-H]^-$
6	Fam-Ahx-AAPV-propargylamino-2'-dU	1122.8	$[M+H]^+$	1120.8	$[M-H]^-$
7	Fam-Ahx-AAPL-propargylamino-2'-dU	1136.9	$[M+H]^+$	1134.9	$[M-H]^-$
8	Fam-Ahx-AIPM-propargylamino-2'-dU	1218.7	$[M+Na]^+$	1194.7	$[M-H]^-$
9	Fam-Ahx-AAPK-propargylamino-2'-dU*	1151.9	$[M+H]^+$	-	-
10	Fam-Ahx-VLK-propargylamino-2'-dU*	1125.0	$[M+H]^+$	-	-
11	Fam-Ahx-GK-propargylamino-2'-dU	969.6	$[M+H]^+$	967.6	$[M-H]^-$
12	Fam-Ahx-AAK-propargylamino-2'-dU	1072.6	$[M+H+18]^+$	1052.8	$[M-H]^-$
13	Fam-Ahx-ALK-propargylamino-2'-dU	1096.9	$[M+H]^+$	1094.8	$[M-H]^-$
14	Fam-Ahx-GGL-propargylamino-2'-dU	1033.7	$[M+Na]^+$	1009.7	$[M-H]^-$
15	Fam-Ahx-GL-propargylamino-2'-dU*	954.6	$[M+H]^+$	-	-
16	Fam-Ahx-AAG-propargylamino-2'-dU	1005.7	$[M+Na]^+$	981.6	$[M-H]^-$
17	Fam-Ahx-GGGL-propargylamino-2'-dU	1090.7	$[M+Na]^+$	1066.8	$[M-H]^-$
18	Fam-Ahx-AGL-propargylamino-2'-dU	1047.8	$[M+Na]^+$	1023.7	$[M-H]^-$
19	Fam-Ahx-LGL-propargylamino-2'-dU	1089.9	$[M+Na]^+$	1065.8	$[M-H]^-$
20	Fam-Ahx-GAGL-propargylamino-2'-dU	1104.8	$[M+Na]^+$	1080.7	$[M-H]^-$
21	Fam-Ahx-DDDDL-propargylamino-2'-dU	1375.7	$[M+H+18]^+$	1356.0	$[M-H]^-$

		MS ES$^+$		MS ES$^-$	
22	Fam-Ahx-DDDDK-propargylamino-2'-dU	1390.8	$[M+H+18]^+$	1371.0	$[M-H]^-$

* Obtained by MALDI-TOF

HPLC Data on Library B

B	Compound	Rt/ min	% Purity		% Purity after HPLC purification		
		HPLC	282 nm	440 nm	ELSD	282 nm	440 nm
1	Fam-Ahx-AAS-propargylamino-2'-dU*	14.5		70	100	100	99
2	Fam-Ahx-ALS-propargylamino-2'-dU	14.5	55	38	-	-	-
3	Fam-Ahx-AAAS-propargylamino-2'-dU	14.5	48	69	-	-	-
4	Fam-Ahx-APAS-propargylamino-2'-dU*	14.5	64	81	100	100	100
5	Fam-Ahx-APVS-propargylamino-2'-dU	14.8	44	63	100	84	86
6	Fam-Ahx-APLS-propargylamino-2'-dU	15.2	34	55	100	100	90
7	Fam-Ahx-APMS-propargylamino-2'-dU	15.0	27	26	-	-	-
8	Fam-Ahx-APKS-propargylamino-2'-dU*	13.8	37	53	53	42	56
9	Fam-Ahx-LKS-propargylamino-2'-dU	14.4	34	45	100	95	95
10	Fam-Ahx-GKS-propargylamino-2'-dU	13.8	55	64	100	100	77
11	Fam-Ahx-AKS-propargylamino-2'-dU*	13.8	24	42	85	100	81
12	Fam-Ahx-GLS-propargylamino-2'-dU	15.1	29	39	100	100	78
13	Fam-Ahx-AGS-propargylamino-2'-dU	14.5	70	47	76	100	80
14	Fam-Ahx-GGLS-propargylamino-2'-dU*	14.9	32	51	100	98	53
15	Fam-Ahx-DDDLS-propargylamino-2'-dU	14.7	35	43	100	89	89
16	Fam-Ahx-DDDKS-propargylamino-2'-dU	13.8	30	55	88	87	83

* broad peak

Mass Spec Data on Library B

B	Compound	MS MALDI	
		m/z	Ion
1	Fam-Ahx-AAS-propargylamino-2'-dU	1013.7	$[M+H]^{1+}$
2	Fam-Ahx-ALS-propargylamino-2'-dU	-	-
3	Fam-Ahx-AAAS-propargylamino-2'-dU	1084.7	$[M+H]^+$
4	Fam-Ahx-APAS-propargylamino-2'-dU	1110.7	$[M+H]^+$
5	Fam-Ahx-APVS-propargylamino-2'-dU	1138.5	$[M+H]^+$
6	Fam-Ahx-APLS-propargylamino-2'-dU	1152.6	$[M+H]^+$
7	Fam-Ahx-APMS-propargylamino-2'-dU	1171.7	$[M+H]^+$
8	Fam-Ahx-APKS-propargylamino-2'-dU	1167.9	$[M+H]^+$
9	Fam-Ahx-LKS-propargylamino-2'-dU	1112.7	$[M+H]^+$
10	Fam-Ahx-GKS-propargylamino-2'-dU	1056.5	$[M+H]^+$
11	Fam-Ahx-AKS-propargylamino-2'-dU	1071	$[M+H]^+$
12	Fam-Ahx-GLS-propargylamino-2'-dU	1041.7	$[M+H]^+$
13	Fam-Ahx-AGS-propargylamino-2'-dU	999.7	$[M+H]^+$
14	Fam-Ahx-GGLS-propargylamino-2'-dU	1098.9	$[M+H]^+$
15	Fam-Ahx-DDDLS-propargylamino-2'-dU	1329.9	$[M+H]^+$
16	Fam-Ahx-DDDKS-propargylamino-2'-dU	1344.6	$[M+H]^+$

HPLC Data on Library C

C	Compound	Rt/ min	% Purity		% Purity after HPLC purification		
		HPLC	282 nm	440 nm	ELSD	282 nm	440 nm
1	Fam-Ahx-GSAA-propargylamino-2'-dU*	7.2	39	53	97	100	100
2	Fam-Ahx-GSAAA-propargylamino-2'-dU	7.2	55	74	100	95	86
3	Fam-Ahx-GSAAL-propargylamino-2'-dU	7.6	43	56	94	83	75
4	Fam-Ahx-GSAAAA-propargylamino-2'-dU	7.2	41	54	68	81	64
5	Fam-Ahx-GSAAPA-propargylamino-2'-dU*	7.2	52	63	97	100	100
6	Fam-Ahx-GSAAPV-propargylamino-2'-dU*	7.3	43	52	94	100	100
7	Fam-Ahx-GSAAPL-propargylamino-2'-dU*	7.5	36	48	93	100	100
8	Fam-Ahx-GSAIPM-propargylamino-2'-dU	7.8	42	46	100	100	100
9	Fam-Ahx-GSAAPK-propargylamino-2'-dU	6.9	48	65	100	100	100

C	Compound	Rt/ min	% Purity		% Purity after HPLC purification		
		HPLC	282 nm	440 nm	ELS D	282 nm	440 nm
10	Fam-Ahx-GSVLK-propargylamino-2'-dU*	7.4	39	50	49	100	92
11	Fam-Ahx-GSGK-propargylamino-2'-dU*	6.9	37	49	92	100	100
12	Fam-Ahx-GSAAK-propargylamino-2'-dU*	6.9	42	57	96	100	100
13	Fam-Ahx-GSALK-propargylamino-2'-dU	7.2	42	54	100	100	100
14	Fam-Ahx-GSAGL-propargylamino-2'-dU	7.5	30	35	100	100	100
15	Fam-Ahx-GSAAG-propargylamino-2'-dU*	7.2	51	68	94	100	95
16	Fam-Ahx-GSGGGL-propargylamino-2'-dU	7.6	31	44	100	87	100
17	Fam-Ahx-GSLGL-propargylamino-2'-dU	7.9	28	32	100	100	100
18	Fam-Ahx-GSDDDDL-propargylamino-2'-dU**	7.4	42	36	-	-	-
19	Fam-Ahx-GSDDDDK-propargylamino-2'-dU**	6.9	24	32	-	-	-

* Analysis by ELS showed that the purification did not result in complete removal of TBAF

** Purification not possible

Mass Spec Data on Library C

C	Compound	MS ES^+		MS ES^-	
		m/z	Ion	m/z	Ion
1	Fam-Ahx-GSAA-propargylamino-2'-dU*	1070.5	$[M+H]^+$	-	-
2	Fam-Ahx-GSAAA-propargylamino-2'-dU*	1141.7	$[M+H]^+$	-	-
3	Fam-Ahx-GSAAL-propargylamino-2'-dU	-	-	-	-
4	Fam-Ahx-GSAAAA-propargylamino-2'-dU	-	-	-	-
5	Fam-Ahx-GSAAPA-propargylamino-2'-dU	1238.7	$[M+H]^+$	1237.0	$[M-H]^-$
6	Fam-Ahx-GSAAPV-propargylamino-2'-dU	1266.7	$[M+H]^+$	1265.1	$[M-H]^-$
7	Fam-Ahx-GSAAPL-propargylamino-2'-dU	1280.8	$[M+H]^+$	1278.8	$[M-H]^-$

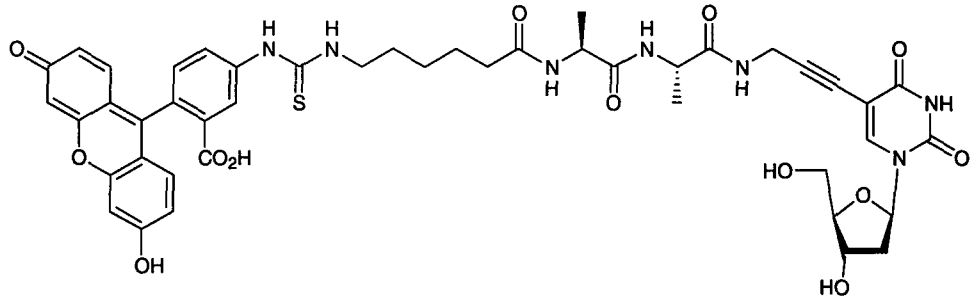
C	Compound	MS ES$^+$		MS ES$^-$	
		m/z	Ion	m/z	Ion
8	Fam-Ahx-GSAIPM-propargylamino-2'-dU*	1340.9	$[M+H]^+$	-	-
9	Fam-Ahx-GSAAPK-propargylamino-2'-dU	1295.9	$[M+H]^+$	1293.8	$[M-H]^-$
10	Fam-Ahx-GSVLK-propargylamino-2'-dU	1268.9	$[M+H]^+$	1266.9	$[M-H]^-$
11	Fam-Ahx-GSGK-propargylamino-2'-dU	1135.8	$[M+Na]^+$	1111.8	$[M-H]^-$
12	Fam-Ahx-GSAAK-propargylamino-2'-dU	1198.8	$[M+H]^+$	1196.8	$[M-H]^-$
13	Fam-Ahx-GSALK-propargylamino-2'-dU	1240.8	$[M+H]^+$	1238.9	$[M-H]^-$
14	Fam-Ahx-GSAGL-propargylamino-2'-dU	1191.8	$[M+Na]^+$	1167.9	$[M-H]^-$
15	Fam-Ahx-GSAAG-propargylamino-2'-dU	1127.5	$[M+H]^+$	1125.9	$[M-H]^-$
16	Fam-Ahx-GSGGGL-propargylamino-2'-dU	-	-	1096.9	$[M-H]^-$
17	Fam-Ahx-GSLGL-propargylamino-2'-dU	1211.9	$[M+H]^+$	1209.9	$[M-H]^-$
18	Fam-Ahx-GSDDDDL-propargylamino-2'-dU**	-	-	-	-
19	Fam-Ahx-GSDDDDK-propargylamino-2'-dU**	-	-	-	-

* Obtained by MALDI-TOF

** Purification not possible

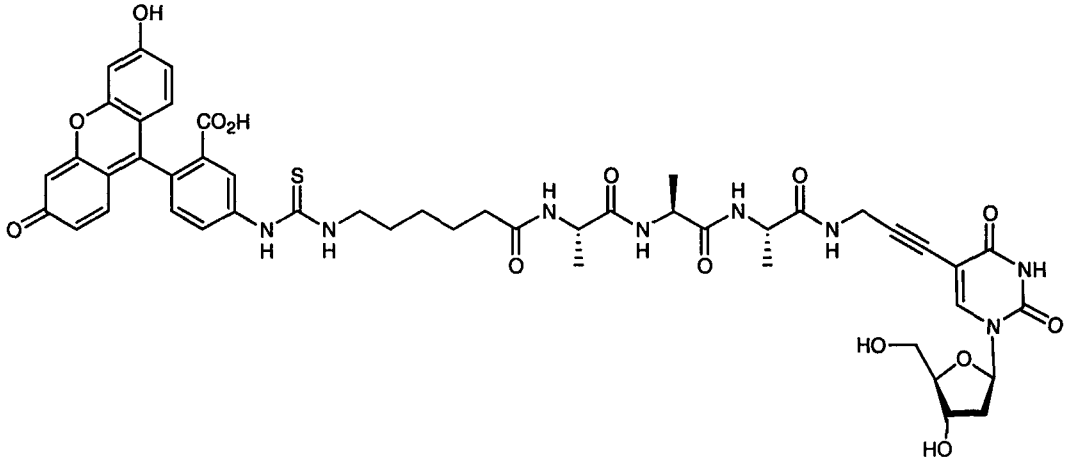
^{1}H NMR of selected library members

A1 Fam-Ahx-AA-propargylamino-2'-dU



δ ^{1}H NMR (300 MHz, MeOH): 1.25 – 1.50 (m, 8H, 2 x C$\underline{H}_3$-Ala, $COCH_2CH_2C\underline{H}_2CH_2CH_2$), 1.55 – 1.80 (m, 4H, $COCH_2C\underline{H}_2CH_2C\underline{H}_2$-$CH_2$), 2.1 – 2.4 (m, 4H, H-2', $COC\underline{H}_2CH_2CH_2CH_2CH_2$), 3.50 – 3.65 (m, 2H, $COCH_2CH_2CH_2CH_2C\underline{H}_2$), 3.65 – 3.85 (m, 2H, H-5'), 3.85 – 4.00 (m, 1H, H-4'), 4.05 – 4.15 (m, 2H, C≡C$C\underline{H}_2$), 4.20 – 4.45 (m, 3H, H-3', 2 x CH-Ala), 6.21 (t, 1H, H-1', *J* = *7*), 6.50 – 6.60 (m, 2H, H-Ar), 6.65 – 6.68 (m, 4H, H-Ar), 7.15 (d, 1H, H-Ar, *J* = *8*), 7.70 – 7.80 (m, 1H, H-Ar), 8.10 (s, 1H, H-6), 8.29 (s, 0.75H, H-Ar), 8.86 (s, 0.25H, H-Ar)

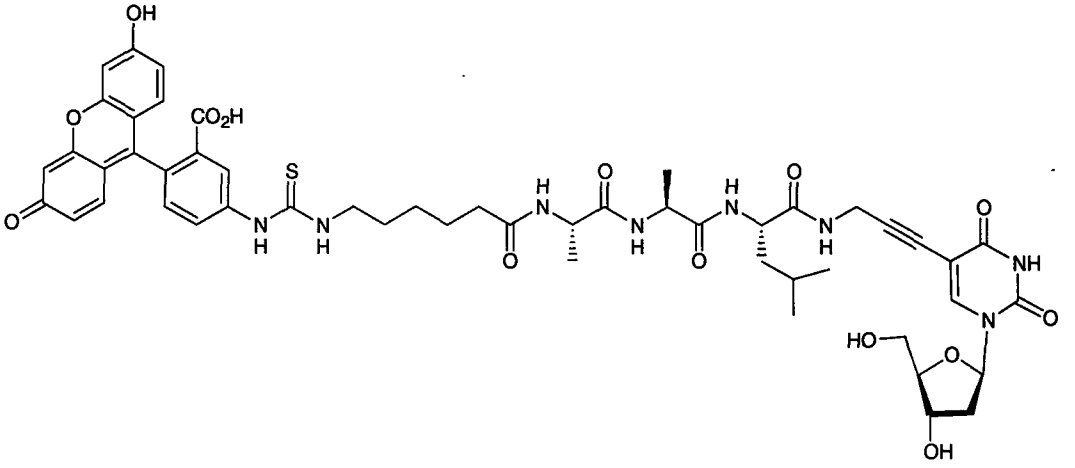
A2 Fam-Ahx-AAA-propargylamino-2'-dU



δ ^{1}H NMR (400 MHz, MeOH): 1.16 – 1.30 (m, 9H, C$\underline{H}_3$-Ala), 1.30 – 1.40 (m, 2H, $COCH_2CH_2C\underline{H}_2CH_2CH_2$), 1.50 – 1.64 (m, 4H,

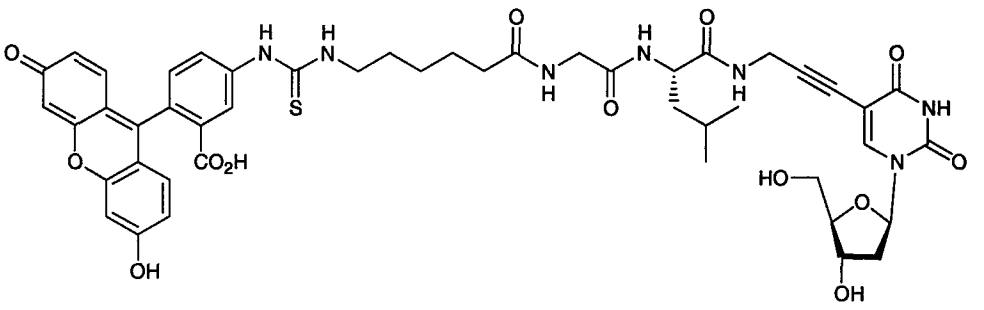
$COCH_2C\underline{H}_2CH_2C\underline{H}_2CH_2$), 2.04 -2.32 (m, 4H, H-2', $COC\underline{H}_2CH_2CH_2$-$CH_2CH_2$), 3.44 – 3.58 (m, 2H, $COCH_2CH_2CH_2CH_2C\underline{H}_2$), 3.58 – 3.74 (m, 2H, H-5'), 3.78 – 3.92 (m, 1H, H-4'), 4.02 – 4.12 (m, 2H, $C{\equiv}CC\underline{H}_2$), 4.10 – 4.25 (m, 3H, 3 x CH-Ala), 4.25 – 4.34 (m, 1H, H-3'), 6.08 – 1.16 (m, 1H, H-1'), 6.45 – 6.57 (m, 2H, H-Ar), 6.57 – 6.68 (m, 3H, H-Ar), 7.07 (d, 2H, H-Ar, *J* = *8*), 7.64 – 7.70 (m, 1H, H-Ar), 8.05 (s broad, 1H, H-6), 8.20 (s, 0.6H, H-Ar), 8.8 (s, 0.4H, H-Ar)

A3 Fam-Ahx-AAL-propargylamino-2'-dU



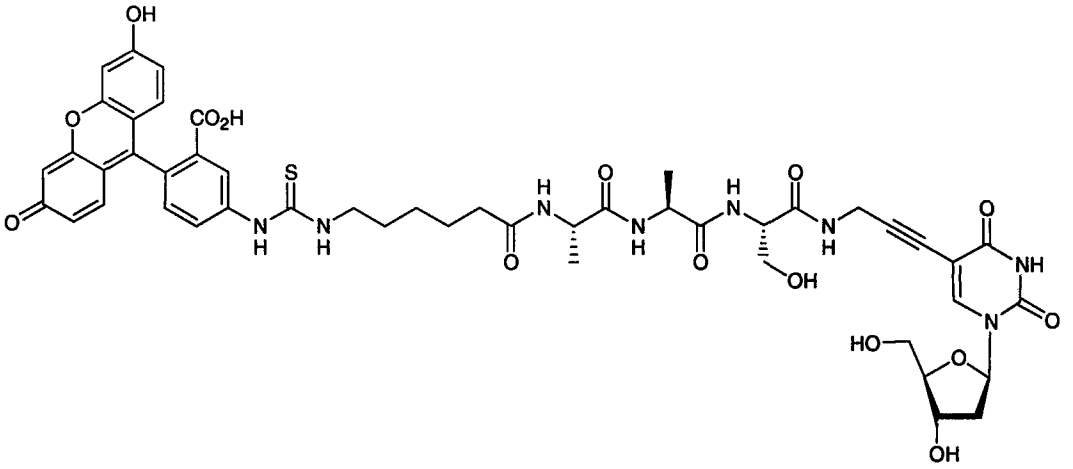
δ ^{1}H NMR (400 MHz, MeOH): 0.74 – 0.86 (m, 6H, $C\underline{H}_3$-Leu), 1.18 – 1.38 (m, 8H, 2 x $C\underline{H}_3$-Ala, $COCH_2CH_2C\underline{H}_2CH_2CH_2$), 1.46 – 1.64 (m, 7H, $COCH_2C\underline{H}_2CH_2C\underline{H}_2CH_2$, $C\underline{H}_2$-Leu, $C\underline{H}$-Leu), 2.06 – 2.32 (m, 4H, H-2', $COC\underline{H}_2CH_2CH_2CH_2CH_2$), 3.46 – 3.56 (m, 2H, $COCH_2CH_2CH_2CH_2C\underline{H}_2$), 3.58 – 3.74 (m, 2H, H-5'), 3.80 – 3.92 (m, 1H, H-4'), 4.00 - 4.10 (m, 2H, $C{\equiv}CC\underline{H}_2$), 4.10 – 4.33 (m, 4H, H-3', 2 x CH-Ala, CH-Leu), 6.09 – 6.16 (m, 1H, H-1'), 6.42 – 6.52 (m, 2H, H-Ar), 6.56 – 6.68 (m, 4H, H-Ar), 7.04 – 7.10 (d, 1H, H-Ar, *J* = *8*), 7.65 – 7.70 (m, 1H, H-Ar), 8.03 – 8.07 (m, 1H, H-6), 8.20 (s, 0.6H, H-Ar), 8.80 (s, 0.4H, H-Ar)

A15 Fam-Ahx-GL-propargylamino-2'-dU



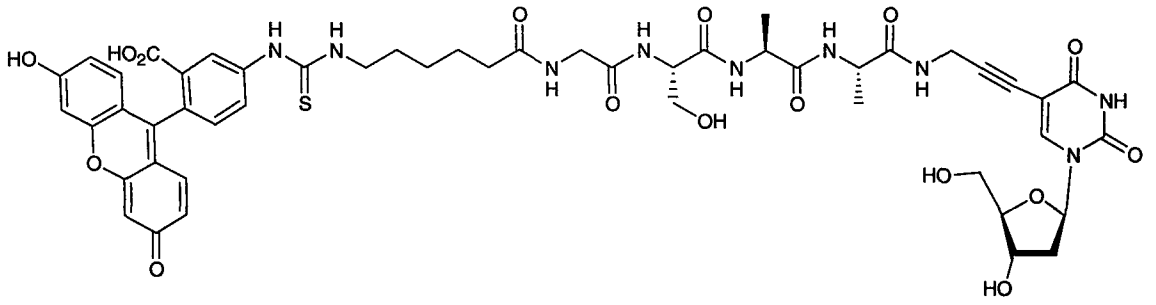
δ ^{1}H NMR (400 MHz, MeOH): 0.70 – 0.90 (m, 6H, C<u>H</u>$_3$-Leu), 1.26 - 1.40 (m, 2H, C<u>H</u>$_2$-Leu), 1.40 – 1.68 (m, 7H, COCH$_2$C<u>H</u>$_2$C<u>H</u>$_2$-C<u>H</u>$_2$CH$_2$, C<u>H</u>-Leu), 2.04 – 2.34 (m, 4H, H-2', COC<u>H</u>$_2$CH$_2$CH$_2$-CH$_2$CH$_2$), 3.44 – 3.58 (m, 2H, COCH$_2$CH$_2$CH$_2$CH$_2$C<u>H</u>$_2$), 3.60 – 3.72 (m, 2H, H-5'), 3.73 – 3.79 (m, 2H, C≡CC<u>H</u>$_2$), 3.79 – 3.91 (m, 1H, H-4'), 4.02 (s, 2H, CH$_2$-Gly), 4.17 – 4.25 (m, 1H, CH-Leu), 4.25 – 4.33 (m, 1H, H-3'), 6.08 – 6.16 (m, 1H, H-1'), 6.4 – 6.49 (m, 2.5H, H-Ar), 6.54 – 6.64 (m, 3.5H, H-Ar), 7.06 (d, 1H, H-Ar, *J* = *8*), 7.66 (d, 1H, H-Ar, *J* = *8*), 8.03 (s, 1H, H-6), 8.18 – 8.20 (m, 0.6H, H-Ar), 8.75 – 8.78 (m, 0.4H, H-Ar)

B1 Fam-Ahx-AAS-propargylamino-2'-dU



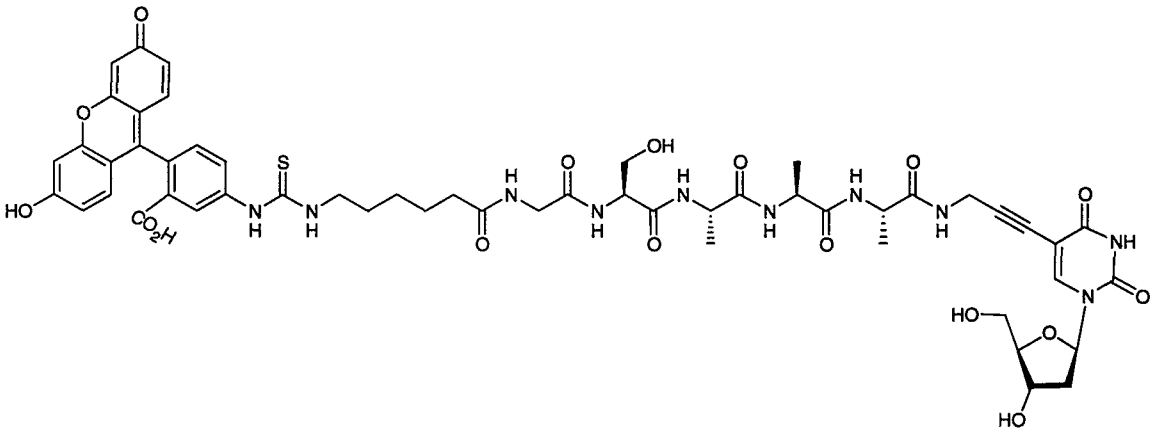
δ ^{1}H NMR (300 MHz, MeOH): 1.30 – 1.40 (m, 8H, C<u>H</u>$_3$-Ala, COCH$_2$CH$_2$C<u>H</u>$_2$CH$_2$CH$_2$), 1.52 – 1.75 (m, 4H, COCH$_2$C<u>H</u>$_2$CH$_2$C<u>H</u>$_2$-CH$_2$), 2.25 – 2.40 (m, 4H, H-2', COC<u>H</u>$_2$CH$_2$CH$_2$CH$_2$CH$_2$), 3.40 – 4.50 (m, 13H, H-4', H-5', C≡CC<u>H</u>$_2$, COCH$_2$CH$_2$CH$_2$CH$_2$C<u>H</u>$_2$, C<u>H</u>$_2$-Ser, H-3', CH-Ser, 2 x CH-Ala), 6.19 – 6.25 (m, 1H, H-1'), 6.50 – 6.60 (m, 2H, H-Ar), 6.60 -6.75 (m, 4H, H-Ar), 7.16 (d, 1H, H-Ar, *J* = *8*), 7.70 - 7.80 (m, 1H, H-Ar), 8.13 (s, 1H, H-6), 8.29 – 8.33 (m, 0.5H, H-Ar), 8.84 – 8.88 (m, 0.5H, H-Ar)

C1 Fam-Ahx-GSAA-propargylamino-2'-dU



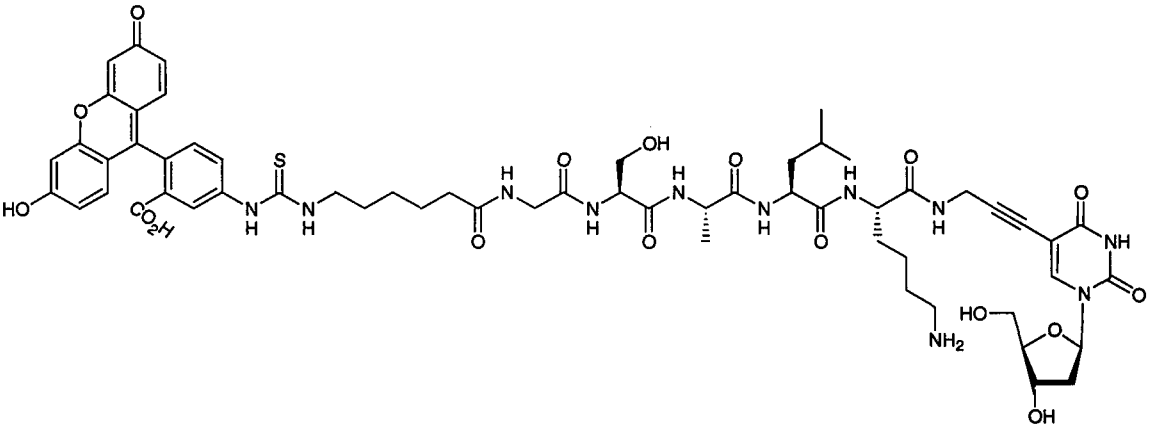
δ ^{1}H NMR (400 MHz, MeOH): 1.06 – 1.42 (m, 8H, C<u>H</u>$_3$-Ala, COCH$_2$CH$_2$C<u>H</u>$_2$CH$_2$CH$_2$), 1.50 – 1.66 (m, COCH$_2$C<u>H</u>$_2$CH$_2$C<u>H</u>$_2$CH$_2$), 2.06 – 2.32 (m, 4H, H-2', COCH$_2$CH$_2$CH$_2$CH$_2$C<u>H</u>$_2$), 3.46 – 3.56 (m, 2H, COCH$_2$CH$_2$CH$_2$CH$_2$C<u>H</u>$_2$), 3.58 – 4.08 (m, 9H, H-4', H-5', C≡CC<u>H</u>$_2$, CH-Gly, C<u>H</u>$_2$-Ser), 4.10 – 4.40 (m, 4H, H-3', 2 x CH-Ala, CH-Ser), 6.08 – 6.18 (m, 1H, H-1'), 6.40 – 6.75 (m, 6H, H-Ar), 7.08 (d, 1H, H-Ar, *J* = *8*), 7.62 – 7.72 (m, 1H, H-Ar), 8.06 (s, 1H, H-6), 8.18 – 8.22 (m, 0.4H, H-Ar), 8.74 – 8.81 (m, 0.6H, H-Ar)

C2 Fam-Ahx-GSAAA-propargylamino-2'-dU



δ ^{1}H NMR (400 MHz, MeOH): 1.08 – 1.40 (m, 11H, C<u>H</u>$_3$-Ala, COCH$_2$CH$_2$C<u>H</u>$_2$CH$_2$CH$_2$), 1.50 – 1.66 (m, 4H, COCH$_2$C<u>H</u>$_2$CH$_2$C<u>H</u>$_2$-CH$_2$), 2.06 – 2.32 (m, 4H, H-2', COC<u>H</u>$_2$CH$_2$CH$_2$CH$_2$CH$_2$), 3.46 – 3.56 (m, 2H, COCH$_2$CH$_2$CH$_2$CH$_2$C<u>H</u>$_2$), 3.58 – 3.92 (m, 9H, H-4', H-5', C≡CC<u>H</u>$_2$, C<u>H</u>$_2$-Ser, C<u>H</u>$_2$-Gly), 3.96 – 4.40 (m, 5H, H-3', 3 x CH-Ala, CH-Ser), 6.08 – 6.16 (m, 1H, H-1'), 6.40 – 6.70 (m, 6H, H-Ar), 7.08 (d, 1H, H-Ar, *J* = *8*), 7.64 – 7.72 (m, 1H, H-Ar), 8.06 (s, 1H, H-6), 8.18 – 8.22 (s, 0.4H, H-Ar), 8.74 – 8.80 (m, 0.6H, H-Ar)

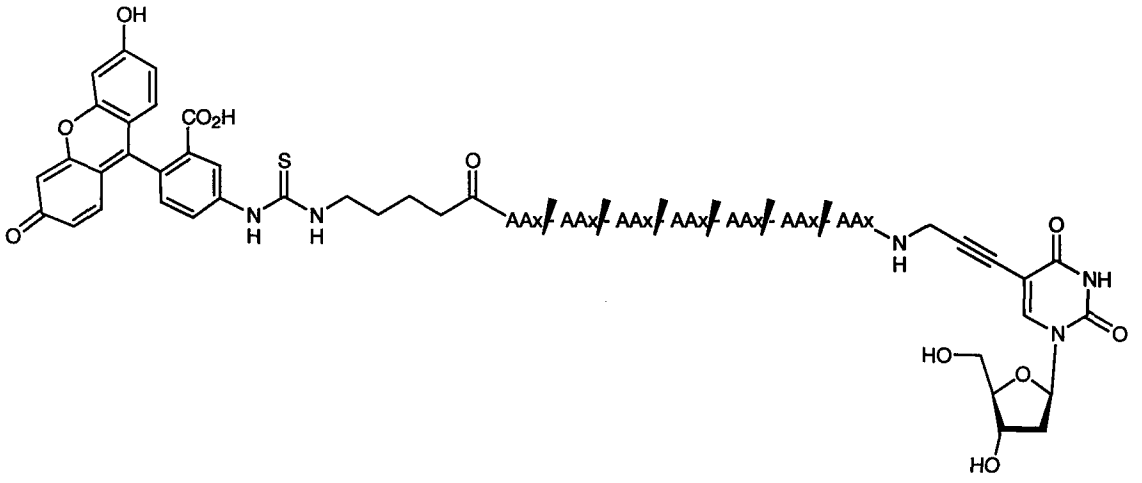
C13 Fam-Ahx-GSALK-propargylamino-2'-dU



δ ^{1}H NMR (400 MHz, MeOH): 0.74 – 0.88 (m, 6H, 2 x C<u>H</u>$_3$-Leu), 1.20 – 1.40 (m, 9H, C<u>H</u>$_3$-Ala, 2 x C<u>H</u>$_2$-Lys, COCH$_2$CH$_2$C<u>H</u>$_2$CH$_2$-

CH_2), 1.44 – 1.82 (m, 9H, $COCH_2C\underline{H}_2CH_2C\underline{H}_2CH_2$, $C\underline{H}_2$-Lys, $C\underline{H}_2$-Leu, $C\underline{H}$-Leu), 2.06 – 2.32 (m, 4H, H-2', $COC\underline{H}_2CH_2CH_2CH_2CH_2$), 2.78 – 2.88 (m, 2H, $C\underline{H}_2$-Lys), 3.46 – 3.56 (m, 2H, $COCH_2CH_2CH_2$-$CH_2\underline{C}H_2NH$), 3.58 – 3.96 (m, 9H, H-4', H-5', CH_2-Ser, CH_2-Gly $C{\equiv}CC\underline{H}_2$), 4.10 – 4.36 (m, 5H, H-3', CH_2-Lys, CH_2-Leu, CH_2-Ala, CH_2-Ser), 6.08 – 6.16 (m, 1H, H-1'), 6.40 – 6.64 (m, 6H, H-Ar), 7.07 (d, 1H, H-Ar, J = 8), 7.62 – 7.68 (m, 1H, H-Ar), 8.04 (s, 1H, H-6), 8.20 – 8.21 (m, 0.2H, H-Ar), 8.77 – 8.81 (m, 0.7H, H-Ar)

7.4.3 Protease Cleavage Assay



Nucleoside Assay Solutions

Each nucleoside was dissolved in water, in some cases Et_3NHCO_3 buffer was added to aid solubility, and the concentration and total mass was determined by UV detection of the fluorescein chromophore and on average 1 mg of compound was obtained. Appropriate aliquots were taken out to prepare 2 mM aqueous assay solutions from all nucleoside stock solutions.

Enzymes

All compounds made were assayed with a selection of proteases: proteinase K (*Tritirachium album*), elastase (pancreatic elastase), subtilisin (*Bacillus licheniformis*), papain (*Carica papaya*), thermolysin (*Bacillus thermo-*

proteolyticus), trypsin (*Porcine*) and enterokinase (*Porcine*). Enzyme stock solutions were prepared to contain 0.33, 0.5 or 1 unit per μL. For details of concentration of enzymes and buffers used see table below.

Enzyme	Units per μL	Buffer
Elastase	0.50	0.1 M Tris pH 8.0
Proteinase K	0.33	0.1 M Tris pH 7.5, 5 mM $CaCl_2$, 0.5% SDS
Subtilisin	0.50	0.1 M Tris pH 8.0, 5 mM $CaCl_2$
Papain	1.00	0.2 M NaCl, 60 mmol Mercaptoethanol, 5 mmol EDTA, 50 mmol L-cysteine hydrochloride hydrate
Thermolysin	1.00	0.1 M Tris pH 8.0, 10 mM $CaCl_2$
Enterokinase	1.00	0.1 M Tris pH 8.0
Trypsin	1.00	Ammonium bicarbonate 50 mM

Standard Assay Method

The buffer (to make a total of 200 μl assay solution), nucleoside (5 μl, 10 nmol), and the enzyme (or water for controls) (1 unit, 1, 2, or 3 μl, depending on enzyme stock) was placed in an eppendorf tube. Solutions were then held at the enzyme's optimal temperature* for two hours and centrifuged through a 10 000 Da molecular weight size exclusion membrane [Amicon Microcon YM-10] to remove the enzyme and the nucleoside solutions analysed by HPLC System IV.

Solutions from the Trypsin assay, which utilised an ammonium bicarbonate buffer, were also analysed by mass spectrometry.

Selected compounds were also analysed over time.

** Trypsin, Papain, Elastase, Enterokinase and Proteinase K (25 °C), Subtilisin and Thermolysin (37 °C)*

Standard Time Course Assay Method

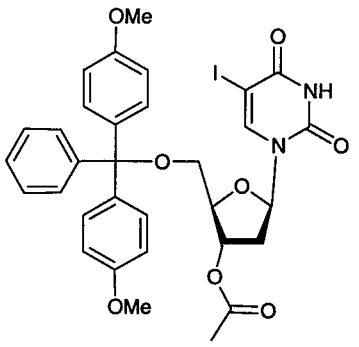
A 2mM solution of the nucleoside substrate (25 μl) was added to the buffer (965 μl). The solution was vortexed and 100 μl was taken out and added to 0.3 M TFA (200 μl). A solution of the enzyme (10 μl, 1unit per μl) was added to the solution,

vortexed and, after approximately 10 seconds, an aliquot of the reaction mixture (100 μl) was added to 0.3 M TFA (200 μl), the first time point. Aliquots were taken at approximately 0 s, 10 s, 30 s, 1 min, 2 min, 4 min, 8 min, 16 min and 2 hours for each compound. All solutions were analysed by HPLC system IV. The following time course assays were carried out.

Enzyme	**Compounds**
Elastase	A5 (Fam-Ahx-AAPA-propargylamino-2'-dU)
	C5 (Fam-Ahx-GSAAPA-propargylamino-2'-dU)
Proteinase K	B15 (Fam-Ahx-GLS- propargylamino-2'-dU)
	C14 (Fam-Ahx-GSAG-propargylamino-2'-dU)
	C15 (Fam-Ahx-GAGL-propargylamino-2'-dU)
Subtilisin	A2 (Fam-Ahx-AAA-propargylamino-2'-dU)
	A3 (Fam-Ahx-AAL-propargylamino-2'-dU)
	A14 (Fam-Ahx-GGL- propargylamino-2'-dU)
	C2 (Fam-Ahx-GSAAA- propargylamino-2'-dU)
	C14 (Fam-Ahx-GSAGL- propargylamino-2'-dU)

7.5 Experimental Procedures for Chapter 4

7.5.1 Preparation of 5-Iodo-(5'-dimethoxytrityl-3'-acetyl-2'-deoxy)uridine 83



Dimethoxytrityl chloride (2.72 g, 8.04 mmol, 1.1 eq) in pyridine (25 mL) was added to a cooled (ice-bath) solution of 5-iodo-2'-deoxyuridine (2.59 g, 7.31 mmol) in pyridine (30 mL) with molecular sieves 3Å. The reaction progress was monitored by TLC and three more equivalents of dimethoxytrityl chloride (7.42 g, 21.9 mmol) was added over two days and the reaction heated to 40°C. When only a trace of starting material could be seen by TLC acetic anhydride (5 mL) was added and the reaction mixture stirred for an additional 3 hours. The mixture was diluted with $CHCl_3$ (100 mL) and the organic layer washed with water (2 x 75 mL), dried with $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by column chromatography using toluene/ethyl acetate (3:1) as eluent to yield the desired product as a white solid (1.79 g, 2.55 mmol, 35%).

MS *m/z* (ES+) 721($M+Na^+$, 20%) 737 ($M+K^+$, 30%)

R_f 0.18 Toluene / Ethyl Acetate (3:1)

HPLC System II Rt/min: 8.7; Purity ($UV_{\lambda=282}$) 100%, (ELS) 100%

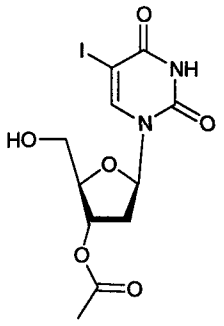
δ ^{1}H NMR (400 MHz, $CDCl_3$): 1.99 (s, 3H, $COC\underline{H}_3$), 2.65 – 2.75 (m, 2H, H-2'), 3.58 – 3.65 (m, 1H, H-5'), 3.70 (s, 6H, 2 x $OC\underline{H}_3$), 3.71 – 3.76 (m, 1H, H-5'), 4.38 – 4.42 (m, 1H, H-4'), 4.92 (broad s,

1H, H-3, NH), 5.65 – 5.70 (m, 1H, H-3'), 6.73 (t, 1H, H-1', *J* = 7), 7.06 (d, 4H, H-DMT, *J* = 9), 7.16 – 7.34 (m, 3H, H-DMT), 7.47 (t, 2H, H-DMT, *J* = 8), 7.65 (d, 4H, H-DMT, *J* = 8), 7.78 (d, 2H, H-Ar, *J* = 7), 8.43 (s, 1H, H-6)

δ ^{13}C NMR (100 MHz, $CDCl_3$): 22.7 ($CO\underline{C}H_3$), 40.4 (C-2'), 57.2 (2 x $O\underline{C}H_3$), 66.3 (C-5'), 73.6 (C-5), 77.4 (C-3'), 86.6 (C-4'), 87.6 (C-1'), 89.4 (C-Ar_3), 115.9, 127.7, 129.4, 130.5, 131.4, 132.6 (C-DMT), 146.2 (C-6), 147.6 (C-DMT), 153.4 (C-2), 161.2 (C-DMT), 163.5 (C-4), 172.4 ($\underline{C}OCH_3$)

Agrees with literature[173]

7.5.2 Preparation of 5-(Iodo)-3'-acetyl-2'-deoxyuridine 84



A 50% solution of TFA in DCM (5 mL) was added to compound **83** (168 mg, 0.24 mmol) and stirred at 45°C for 30 min (the solution turned red immediately upon addition). The solvent was evaporated and 1 mL of water was added. The sample was concentrated under reduced pressure and the compound was purified by column chromatography using toluene/ethyl acetate (3:1) as eluent. The desired compound was obtained as a white solid (94 mg, 2.37 mmol, 99%).

MS *m/z* (ES+) 815 ($2M+Na^+$, 30%)

R_f 0.74 Toluene / Ethyl Acetate (3:1)

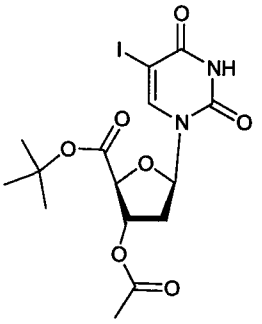
HPLC System II Rt/min: 7.6; Purity (ELS) 93%, ($UV_{\lambda=282}$) 92%

δ ^{1}H NMR (400 MHz, Pyridine-d_5): 1.72 (s, 3H, $COC\underline{H}_3$), 2.25-2.45 (m, 2H, H-2'), 3.85 (d, 2H, H-5', J = 2), 4.08 – 4.12(m, 1H, H-4'), 5.39 – 5.41 (m, 1H, H-3'), 6.50 – 6.56 (m, 1H, H-1'), 8.44(s, 1H, H-6)

δ ^{13}C NMR (100 MHz, Pyridine-d_5): 23.2 ($CO\underline{C}H_3$), 40.9 (C-2'), 64.6 (C-5'), 73.3 (C-5), 78.5 (C-3'), 88.2 (C-1'), 88.8 (C-4'), 147.9 (C-6), 154.0 (C-2), 164.0 (C-4), 172.8 ($\underline{C}OCH_3$)

Agrees with literature[173]

7.5.3 Preparation of 1-(*t*-butyl 3'-*O*-acetyl-2'-deoxy-β-D-ribofuranosyluronate)-5-iodouracil 85



t-Butyl alcohol (397 mg, 505 μL, 5.36 mmol, 20 eq), pyridinium dichromate (202 mg, 0.536 mmol, 2 eq) and acetic anhydride (274 mg, 2.68 mmol, 10 eq) were added to a stirred solution of 5-iodo-(3'-acetyl-2'-deoxy)uridine (106 mg, 0.268 mmol) in DCM (2 mL). The mixture was stirred for 1 hour and then loaded on a 3 x 3 cm plug of silica using EtOAc. After 15 min the product was filtered through using 3 x 25 mL of EtOAc. The solvent was evaporated and the compound was filtered through another silica plug eluted with EtOAc/ $CDCl_3$ (1:1). Solvent was removed under reduced pressure and the product was obtained as an off white solid (80 mg, 0.172 mmol, 64%).

MS *m/z* (ES+) 484 ($M+NH_4^+$, 5%) 505 ($M+K^+$, 5%) 954 ($2M+Na^+$, 100%)

IR $\nu(cm^{-1})$ (neat) 1675 - 1725 (C=O str, $COOC(CH_3)_3$, $COCH_3$, uridine ring)

R_f 0.69 Ethyl acetate : Hexane (1:1)

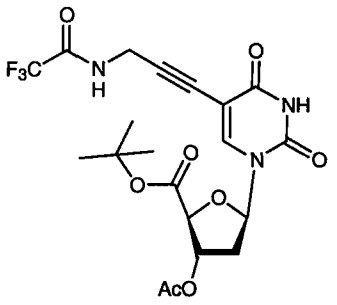
HPLC System II Rt/min: 10.5; Purity ($UV_{\lambda=282}$) 80%, (ELS) 99%

δ ^{1}H NMR (400 MHz, Pyridine-d_5): 1.26 (s, 9H, $COOC(C\underline{H}_3)_3$), 1.77 (s, 3H, $C\underline{H}_3$-acetate), 2.30 – 2.60 (m, 2H, H-2'), 4.62 – 4.67 (m, 1H, H-4'), 5.57 (d, 1H, H-3', *J* = *5*), 6.57 - 6.64 (m, 1H, H-1'), 9.83 (s, 1H, H-6)

δ ^{13}C NMR (100 MHz, Pyridine-d_5): 23.0 ($\underline{C}H_3$-acetate), 30.2 ($COOC(\underline{C}H_3)_3$), 39.5 (C-2'), 73.6 (C-5), 79.0 (C-3'), 85.6 ($COO\underline{C}(CH_3)_3$), 85.9 (C-4'), 89.8 (C-1'), 147.5 (C-6), 153.8 (C-2), 163.8 (C-4), 172.1 ($\underline{C}OCH_3$), 172.2 (C-5')

Agrees with literature[173]

7.5.4 Preparation of 1-(*t*-butyl 3'-*O*-acetyl-2'-deoxy-β-D-ribofuranosyluronate)-5-*N*-trifluoroacetylpropargylaminouracil 86



1-(*t*-butyl 3'-*O*-acetyl-2'-deoxy-β-D-ribofuranosyluronate)-5-iodouracil **85** (438 mg, 0.94 mmol) was dissolved in dry DMF (20 mL) and degassed with N_2 for 10 minutes. Copper iodide (36.0 mg, 0.188 mmol, 0.2 eq.), anhydrous triethylamine (476 mg, 4.70 mmol, 655 μL, 5 eq.), *N*-trifluoroacetylpropargylamine (426 mg,

2.82 mmol, 3 eq.) and *tetrakis* (triphenyl phosphine) palladium(0) (109 mg, 9.40 mmol, 0.1 eq.) were added sequentially. The mixture was stirred under N_2 at 50°C for 50 minutes. The solvent was reduced to a minimum under reduced pressure and the residual oil was redissolved in CH_2Cl_2 (20 mL). The organic phase was washed with 5% disodium EDTA (2 x 50 mL) and aqueous sodium bisulphate 5% (w:v, 2 x 50 mL). The organic phase was dried ($MgSO_4$) and the crude product was purified by silica gel column chromatography using ethyl acetate/hexane (1:1) as eluent to afford the desired product **86** (357 mg, 0.73 mmol) in 78% yield.

MS *m/z* (ES+) 512 ($M+Na^+$, 100%)

HPLC System II Rt/min: 11.0; Purity ($UV_{\lambda=282}$ ELS) 100%

HRMS $C_{20}H_{22}N_3O_8F_3Na$: 512.1257 (calc.), 512.1251 (found)

IR $\nu(cm^{-1})$ (neat) 1724 (C=O str, s, $COOC(CH_3)_3$),1714 (C=O str, s, $COCF_3$), 1709 (C=O str, s, $COCH_3$), 1665 (C=O str, s, uridine ring)

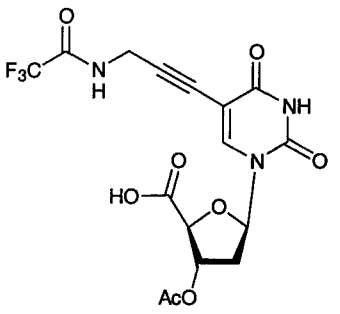
R_f 0.23 Hexane / Ethyl Acetate (1:1)

Mp 166°C

δ ^{1}H NMR (400 MHz, Pyridine-d_5): 1.22 (s, 9H, $COOC(C\underline{H}_3)_3$), 1.73 (s, 3H, $C\underline{H}_3$-acetate), 2.27 - 2.34 (m, 1H, H-2'), 2.51 - 2.56 (m, 1H, H-2'), 4.32 (d, 2H, $C{\equiv}CC\underline{H}_2$, *J = 5*), 4.63 (s, 1H, H-4'), 5.52 - 5.57 (m, 1H, H-3'), 6.52 - 6.58 (m, 1H, H-1'), 8.54 (s, 1H, H-6), 11.21 - 11.29 (m, 1H, $C{\equiv}CCH_2N\underline{H}$)

δ ^{13}C NMR (100 MHz, Chloroform-d_1): 19.9 ($\underline{C}H_3$-acetate), 26.9 ($COOC(\underline{C}H_3)_3$), 29.6 ($C{\equiv}C\underline{C}H_2$), 36.3 (C-2'), 74.7 (C-3'), 75.0 (C-5), 82.0 (C-4'), 83.1 ($COO\underline{C}(CH_3)_3$), 86.1 (C-1'), 86.3 ($C{\equiv}\underline{C}CH_2$), 98.3 ($\underline{C}{\equiv}CCH_2$), 114.7 ($C{\equiv}CCH_2NHCO\underline{C}F_3$, *J = 287*), 143.2 (C-6), 148.3 (C-2), 155.9 ($C{\equiv}CCH_2NH\underline{C}OCF_3$, *J* = 38), 160.9 (C-4), 168.0 (C-5'), 168.8 ($\underline{C}OCH_3$)

7.5.5 Preparation of 1-(3'-*O*-acetyl-2'-deoxy-β-D-ribofuranosyluronic acid)-5-*N*-trifluoroacetylpropargylaminouracil 87



1-(*t*-butyl 3'-*O*-acetyl-2'-deoxy-β-D-ribofuranosyluronate)-5-*N*-trifluoroacetylpropargylaminouracil **86** (91.4 mg, 0.187 mmol) was dissolved in aqueous trifluoroacetic acid (80%, 1 mL) and stirred for 45 minutes. The solvent was removed under reduced pressure and the crude product was purified by silica gel column chromatography using a gradient of ethyl acetate/MeOH (10:1 to 4:1) as eluent. Product **87** was obtained as a white solid (67 mg, 0.147 mmol, 83%).

MS *m/z* (ES-) 432 (M^-, 100%)

HPLC System II Rt/min: 6.5; Purity ($UV_{\lambda=282}$, ELS) 100%

HRMS $C_{16}H_{14}F_3N_3O_8Na$: 456.0631 (calc.), 456.0625 (found)

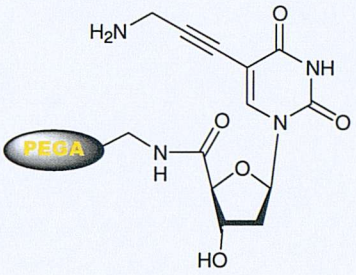
R_f 0.2 Ethyl Acetate / MeOH (10:1)

Mp 210°C

δ ^{1}H NMR (400 MHz, MeOH-d_4): 2.02 (s, 3H, C$\underline{H}_3$-acetate), 2.10 – 2.20 (m, 1H, H-2'), 2.40 – 2.50 (m, 1H, H-2'), 4.19 (s, 2H, C≡CC$\underline{H}_2$), 4.55 (s, 1H, H-4'), 5.42 – 5.46 (m, 1H, H-3'), 6.20 – 6.27 (m, 1H, H-1'), 8.56 (s, 1H, H-6)

δ ^{13}C NMR (100 MHz, MeOH-d_4): 20.0 (C$\underline{H}_3$-acetate), 29.9 (C≡C-C$\underline{H}_2$), 36.9 (C-2'), 75.1 (C-3'), 77.0 (C-5), 82.8 (C-4'), 87.6 (C-1'), 87.8 (C≡$\underline{C}$CH$_2$), 99.0 ($\underline{C}$≡CCH$_2$), 116.5 (C≡CCH$_2$NHCO$\underline{C}$F$_3$, J = 287 145.0 (C-6), 150.3 (C-2), 157.6 (C≡CCH$_2$NH$\underline{C}$OCF$_3$, J = 38), 163.6 (C-4), 170.6 ($\underline{C}$OCH$_3$), 172.2 (C-5')

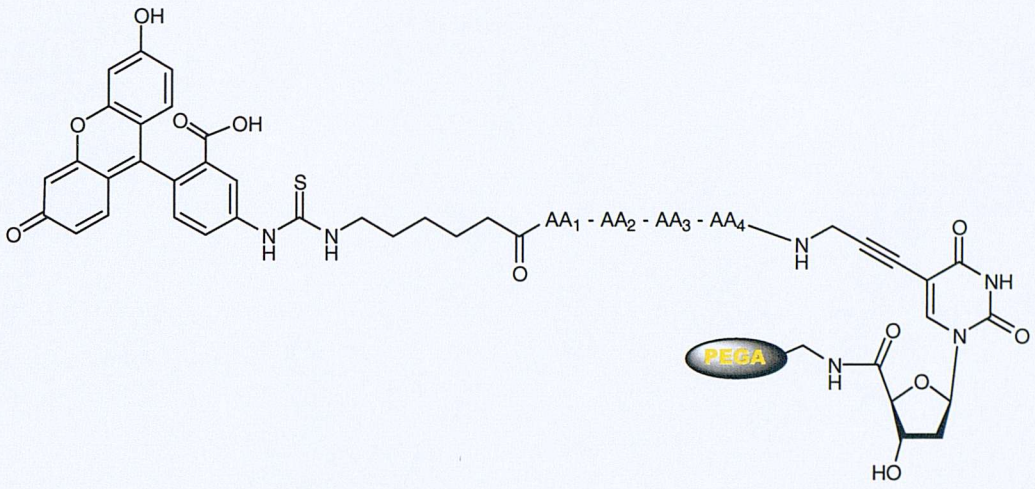
7.5.6 Preparation of PEGA immobilized 1-(3'-hydroxy-2'-deoxy-β-D-ribofurano-syluronic acid)-5-*N*-propargylaminouracil 89



Compound **87** (120 mg, 0.276 mmol, 1.2 eq), HOBt (37 mg, 0.276 mmol, 1.2 eq) and DIC (43 mg, 54 μL, 0.344 mmol, 1.5 eq) were dissolved in DMF (10 ml). After 10 minutes the reaction mixture was added to the pre-swollen amino methyl $PEGA_{800}$ resin (0.574 g, 0.230 mmol) (pre-swollen in DMF) and agitated overnight. The reaction progress was monitored by qualitative ninhydrin test and to ensure complete coupling the loading procedure was repeated with a fresh reaction mixture for 12 hours. Then the resin was washed with DMF (3 x 6 mL), CH_2Cl_2 (5 x 6 mL), MeOH (3 x 6 mL) and the remaining sites were capped by shaking the resin with CH_2Cl_2 (5.5 mL), pyridine (0.2 mL) and acetic anhydride (1.4 mL) for 20 minutes. The resin was washed with DMF (3 x 6 mL), CH_2Cl_2 (5 x 6 mL), MeOH (3 x 6 mL) and CH_2Cl_2 (3 x 6 mL).

Dioxane (6 mL) was added to the resin (0.616 g), when swollen 1M KOH aq. (3 mL) was added and the resin shaken overnight. The resin was washed with dioxane/ water (1:1) (3 x 10 mL), water (3 x 10 mL), dioxane/ water (1:1) (1 x 10 mL), dioxane (3 x 10mL), MeOH (3 x 10 mL) and CH_2Cl_2 (3 x 10 mL).

7.5.7 Preparation of Solid Phase Bound Reporter Nucleosides 90, 91, 92 and 93



	AA_1	AA_2	AA_3	AA_4
90	Ala	Ala	Ala	-
91	Ala	Ala	Leu	-
92	Gly	Leu	Ser	-
93	Gly	Ala	Gly	Leu

General procedure for SPPS on PEGA

DIC (3 eq.) was added to a solution of HOBt (3 eq.) and Fmoc-AA-OH (3 eq.) in DMF/ DCM (2:3) (0.5 mL). After 10 minutes the solution was added to resin **89** and shaken for 1.5 hours. The resin was washed with DMF (5 x 1 mL) and the coupling repeated to ensure complete loading. The resin was washed with DMF (3 x 5 mL), CH_2Cl_2 (3 x 5 mL), MeOH (3 x 5 mL), CH_2Cl_2 (3 x 5 mL) and MeOH (3 x 5 mL). The Fmoc group was removed by treatment with 20% piperidine in DMF for 1 hour followed by washing the resin with 20% piperidine in DMF (2 x 5 mL), DMF (3 x 5 mL), CH_2Cl_2 (3 x 5 mL) and MeOH (3 x 5 mL). This procedure was repeated for each amino acid residue in the desired sequence. For quantities used see *Table 7.5.1.*

Peptide Sequence / Reagents *	**90**	**91**	**92**	**93**
	AAA	AAL	GLS	GAGL
Resin	88	100	83	83
HOBt	14	16	14	14
DIC	13	15	13	13
AA_1	(A) 33	(L) 43	(S) 38	(L) 35
AA_2	(A) 33	(A) 37	(L) 35	(G) 30
AA_3	(A) 33	(A) 37	(G) 30	(A) 31
AA_4	n/a	n/a	n/a	(G) 30
Ahx	37	42	35	35

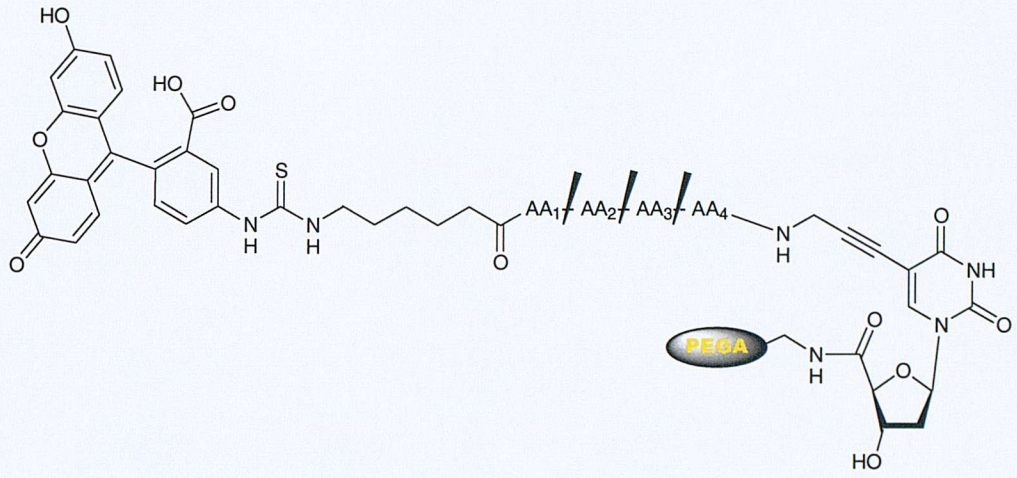
* all quantities in mg

Table 7.5.1 *Quantities of reagents used in SPPS toward the preparation of solid phase bound reporter nucleosides* **90, 91, 92** and **93**

Fluorescein Labelling

NEt_3 (68 µL, 0.48 mmol) was added to a solution of FITC (188 mg, 0.48 mmol) in DMF (8 mL). To each resin, **90, 91, 92, 93** was added 2 mL of the FITC solution and the mixture agitated overnight. The resin was washed with DMF (5 x 4 mL), CH_2Cl_2 (5 x 4 mL), MeOH (5 x 4 mL), DMF (2 x 4 mL), CH_2Cl_2 (2 x 4 mL) and MeOH (5 x 4 mL). Resin **92** was treated with trifluoroacetic acid (1 mL) for 30 minutes and the washing procedure was repeated for this resin.

7.5.8 Solid Phase Protease Assay



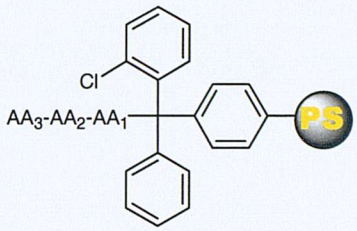
Resins **90, 91, 92** and **93** (1.1 mg, respectively) were washed with the appropriate buffer for the enzyme, see *Table 7.5.2*, (5 x 1 mL or until no release of colour could be observed) and placed in an eppendorf tube. The appropriate reaction buffer (1 mL) for each enzyme was added to each resin and the solution (solution A) was kept at the enzyme optimal temperature, *Table 7.5.2*. The analysis buffer (2000μL) (pH 9 sodium borate/ HCl buffer) and solution A(3 μL) was added to the cuvette followed by the analysis buffer (1000 μL). The solution was mixed and the background fluoresence (blank) was recorded. The appropriate enzyme (subtilisin, proteinase K or elastase) (10 units) was added to solution A and aliquots were taken out for analysis every 5 to 15 minutes for fluorescence analysis as described for the blank. The assay was repeated for each enzyme / resin combination.

Enzyme	**Buffer**	**Optimal T °C**
Elastase	0.1M TRIS pH 8	25
Subtilisin	0.1M TRIS pH 8, 5mM $CaCl_2$	37
Proteinase K	0.1M TRIS pH 7.5, 5mM $CaCl_2$, 0.5% SDS	25

Table 7.5.2 *Table of enzymes and appropriate buffers*

7.6 Experimental Procedures for Chapter 5

7.6.1 Preparation of Solid Phase Bound Peptides 95 and 96



	AA_1	AA_2	AA_3
95	Ala	Ala	Ala
96	Ser	Leu	Gly

Loading of the Trityl Chloride Polystyrene Resin

The apparatus was oven dried overnight prior to use. FmocSer(^{t}Bu)OH, FmocAlaOH, FmocGlyOH and FmocLeuOH were dried under vacuum over P_2O_5 overnight. DIPEA stored over KOH was used. Anhydrous DMF and freshly distilled DCM was used.

2-Chlorotrityl chloride polystyrene resin, for quantities see *Table 7.6.1*, was placed in a disposable peptide vessel and N_2 was continously purged through the vessel. FmocAAOH (1.2 eq) was dissolved in dry DMF / CH_2Cl_2 (1:20) and DIPEA (4 eq to carboxylic acid). The mixture was then added to the resin and left to stand overnight under a continous flow of nitrogen. The resin was washed with CH_2Cl_2/ MeOH/ DIPEA (10:1:0.5) (3 x 10 mL), CH_2Cl_2 (3 x 10 mL), DMF (3 x 10 mL), CH_2Cl_2 (3 x 10 mL). The resin was dried under vacuum and the loading of Fmoc-Alanine and Fmoc-(^{t}Bu)Serine was measured by a standard Fmoc test, 0.913 mmol/g (92%) and 0.686 mmol/g (74%), respectively. The resin was swollen in CH_2Cl_2 and the solution drained off, 20% piperidine in DMF was added, the resin was agitated for 40 minutes and washed with 20% piperidine in DMF (2 x 10 mL), DMF (3 x 10 mL), CH_2Cl_2 (3 x 10 mL), MeOH (3 x 10 mL), CH_2Cl_2 (3 x 10 mL) and dried under reduced pressure.

Resin (g)	mmol resin	Amino acid	Mass (g)	mmol AA	eq	DIPEA (mL)	Loading FmocAA mmol
1.05	1.47	FmocAlaOH	0.523	1.68	1.2	1.2	0.913
1.00	1.40	FmocSer(^{t}Bu)OH	0.644	1.68	1.2	1.2	0.686

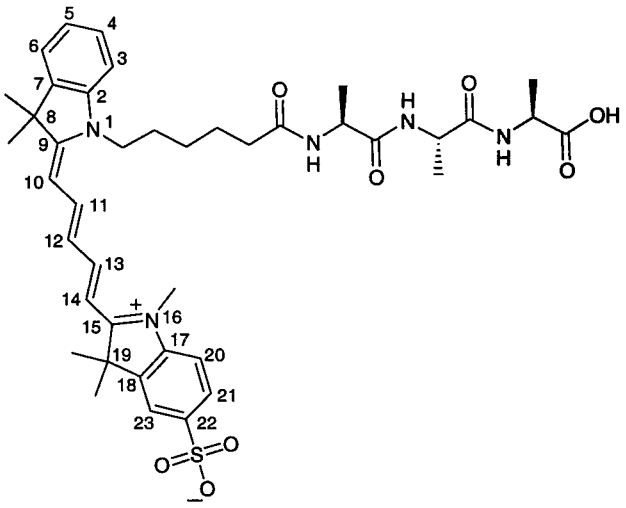
Table 7.6.1 Quantities used for loading

Solid Phase Peptide Synthesis

To a solution of FmocAAOH (3 eq), HOBt (3 eq) in DMF/ DCM (6 mL, 1:4) was added DIC (3 eq). After 5 minutes activation, the reaction mixture was added to the preloaded resin and agitated overnight. The progress of coupling was monitored by a qualitative ninhydrin test and the coupling procedure was repeated if necessary. The resin was washed with DMF (3 x 10 mL), CH_2Cl_2 (3 x 10 mL), MeOH (3 x 10 mL) and CH_2Cl_2 (3 x 10 mL) and dried under reduced pressure. The substitution was measured by a standard Fmoc test. A solution of 20% piperidine in DMF was added, the resin was agitated for 40 minutes and washed with 20% piperidine in DMF (2 x 10 mL), DMF (3 x 10 mL), CH_2Cl_2 (3 x 10 mL), MeOH (3 x 10 mL) and CH_2Cl_2 (3 x 10 mL) and dried under reduced pressure. The procedure was repeated for each aminoacid in the sequence. Fmoc tests after the last amino acid coupling displayed substitutions of 0.551 mmol/g (68%) and 0.424 mmol/g (69%) for **95** (resin-AAA) and **96** (resin-SLG), respectively.

(For coupling of Fmoc-Gly-OH DMF only was used as a solvent.)

7.6.2 Preparation of a Cy5™ Labelled Peptide: Cy5-Ala-Ala-Ala 98



HOBt (14.4 mg, 0.107 mmol) and Cy5™ (60 mg, 0.107 mmol) were dissolved in DMF (1 mL) and DCM (1.5 mL). DIC (13.4 mg, 0.107 mmol, 16.6 μL) and DIPEA (13.8 mg, 0.107 mmol, 18.6 μL) were added. After 15 min the reaction mixture was added to resin **95** (468 mg, 0.160 mmol) and the resin was agitated overnight. The resin was washed with DMF (8 x 3 mL), DCM (8 x 3 mL), MeOH/ DCM (1:1) (8 x 3 mL) and DCM (8 x 3 mL).

The standard cleavage procedure outlined in section *7.2.1* was applied. The crude material was precipitated from ice cold Et_2O (30 mL) in a centrifuge tube. The tube was centrifuged for 15 min, solvent was decanted and the procedure was repeated three times. Drying the product under reduced pressure afforded the crude product as a blue solid (175 mg). The crude product, also containing unreacted peptide was purified by semi-preparative HPLC. The appropriate fractions were pooled and lyophilized solid to afford a blue solid (33 mg, 40 %).

MS *m/z* (ES+) 776.8 ($M+H^+$, 100%) 798.4 ($M+Na^+$, 80%)
HRMS $C_{41}H_{54}N_5O_8S$: 776.3690 (calc.); 776.3688 (found)

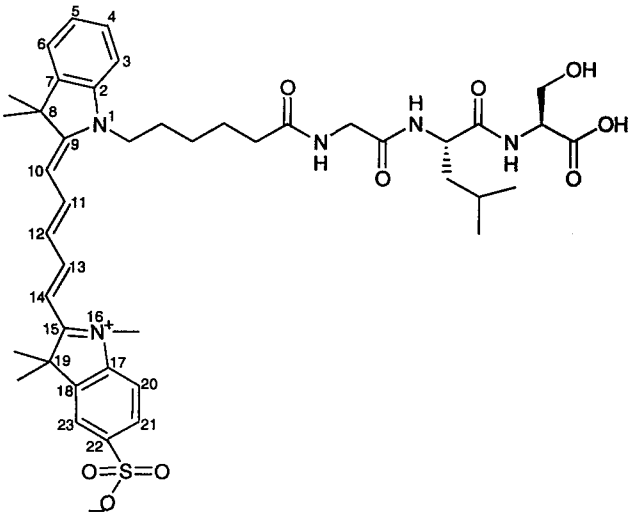
Mp 202°C

HPLC Rt/min: 6.9; Purity ($UV_{(\lambda = 220, 282, 649\ nm)}$, ELS = 100%)

δ ^{1}H NMR (400 MHz, $CDCl_3$): 1.24, 1.27, 1.28 (d x 3, 9H, 3 x $C\underline{H}_3$-Ala, *J* = 7), 1.35 - 1.47 (m, 2H, $COCH_2CH_2C\underline{H}_2CH_2CH_2$), 1.60 - 1.66 (m, 14H, 4 x $C\underline{H}_3$-Cy5, $COCH_2C\underline{H}_2CH_2CH_2CH_2$), 1.67 - 1.82 (m, 2H, $COCH_2CH_2CH_2C\underline{H}_2CH_2$), 2.09 - 2.24 (m, 2H, $COC\underline{H}_2CH_2CH_2CH_2CH_2$), 3.48 (s, 3H, N-$C\underline{H}_3$), 4.07 (t, 2H, $COCH_2CH_2CH_2CH_2C\underline{H}_2$, *J* = 7), 4.25 (q x 3, 3H, 3 x CH-Ala, *J* = 7), 6.14 (d, 1H, H-10, *J* = 14), 6.34 (d, 1H, H-14, *J* = 14), 6.55 (t, 1H, H-12, *J* = 13), 7.15 (d, 1H, H-20, *J* = 9), 7.21 (t, 1H, H-5, *J* = 7), 7.27 (d, 1H, H-3, *J* = 8), 7.34 (d, 1H, H-4, *J* = 8), 7.43 (d, 1H, H-6, *J* = 7), 7.73 – 7.78 (m, 2H, H-21, H-23), 8.16 (m, 2H, H-11, H-13)

δ ^{13}C NMR (100 MHz, $CDCl_3$): 18.0, 18.4, 18.6 (3 x $\underline{C}H_3$-Ala) 26.7 ($COCH_2CH_2\underline{C}H_2CH_2CH_2$), 27.5 ($COCH_2\underline{C}H_2CH_2CH_2CH_2$), 28.2, 28.2, 28.3, 28.3 (4 x $\underline{C}H_3$-Cy5), 28.7 ($COCH_2CH_2CH_2\underline{C}H_2CH_2$), 31.8 (N-$\underline{C}H_3$), 36.5 ($CO\underline{C}H_2CH_2CH_2CH_2CH_2$), 45.6 ($COCH_2CH_2$-$CH_2CH_2\underline{C}H_2$), 49.4, 49.7, 49.7, 50.5, 50.8, (3 x $\underline{C}$H-Ala, 2 x $\underline{C}(CH_3)_2$), 104.5 (C-10), 106.1(C-14), 111.2 (C-20), 113.0 (C-3), 121.6 (C-21), 123.9 (C-6), 127.3 (C-5), 127.7 (C-12), 128.4 (C-23), 130.3 (C-4), 142.5 (C-2), 143.2, 143.4, 143.7 (C-9, C-15, C-17), 146.3 (C-22), 155.3 (C-11), 156.9 (C-13), 174.7, 174.9, 175.2 (2 x $\underline{C}$=O-Ala, C=O-Ahx), 176.1, 176.2, 176.6 ($\underline{C}$OOH, C-7, C-18)

7.6.3 Preparation of a Cy5™ Labelled Peptide: Cy5-Gly-Leu-Ser 99



HOBt (12 mg, 0.086 mmol) and Cy5™ (49 mg, 0.086 mmol) was dissolved in DMF (1 mL) and DCM (1.5 mL). DIC (11 mg, 0.086 mmol, 14 µL) and DIPEA (14 mg, 0.108 mmol, 19 µL) was added. After 15 min the reaction mixture was added to resin **96** (120 mg, 0.072 mmol) and the resin was agitated overnight. The resin was washed with DMF (8 x 3 mL), DCM (8 x 3 mL), MeOH/ DCM (1:1) (8 x 3 mL) and DCM (8 x 3 mL). The standard cleavage procedure was applied. The crude product (55 mg, 91%) was purified by semi-preparative HPLC, the fractions were pooled and freeze dried, a blue lyophilised solid was obtained (29 mg, 0.035 mmol, 70 %).

MS *m/z* (ES+) 820.5 ($M+H^+$, 95%) 842.5 ($M+Na^+$, 100%)

HRMS $C_{43}H_{57}N_5O_9SNa$: 842.3775 (calc.); 842.3791 (found)

Mp 189°C

HPLC System II Rt/min: 7.2; Purity ($UV_{(\lambda = 220, 282, 649\ nm}$, ELS = 100 %)

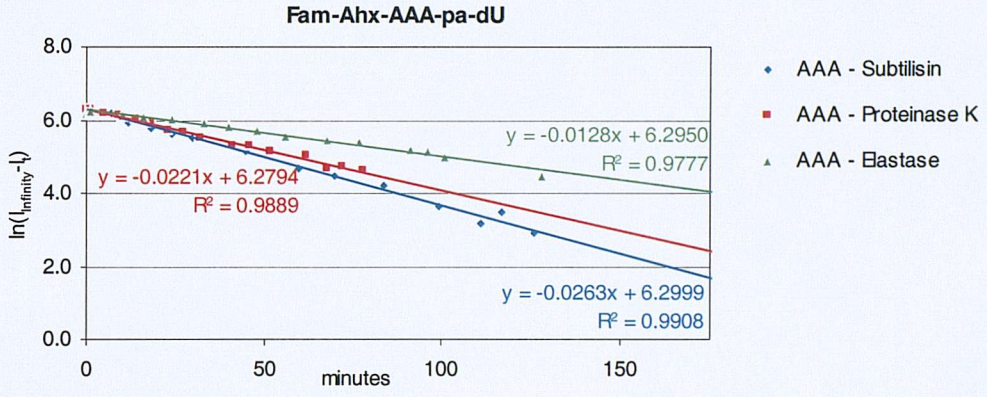
δ ^{1}H NMR (400 MHz, $CDCl_3$): 0.79 - 0.87 (m, 6H, 2 x CH_3-Leu), 1.35 - 1.45 (m, 2H, $COCH_2CH_2C\underline{H}_2CH_2CH_2$), 1.45 - 1.67 (m, 17H, $COCH_2C\underline{H}_2CH_2CH_2CH_2$, 4 x $C\underline{H}_3$-Cy5, CH_2-Leu, $C\underline{H}(CH_3)$), 1.68 - 1.79 (m, 2H, $OCH_2CH_2CH_2C\underline{H}_2CH_2$), 2.20 (t, 2H, $OC\underline{H}_2$-

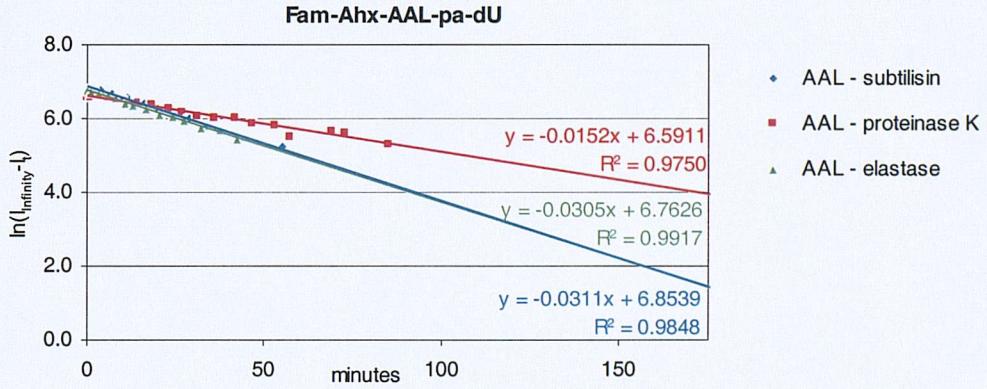
CH2CH2CH2CH2, J = 7), 3.48 (s, 3H, N-C$\underline{H}_3$), 3.68 – 3.81 (m, 4H, CH_2-Gly, CH_2-Ser), 4.07 (t, 2H, $OCH_2CH_2CH_2CH_2C\underline{H}_2$, J = 8), 4.20 – 4.26 (m, 1H, CH-Ser), 4.34 – 4.41 (m, 1H, CH-Leu), 6.13 (d, 1H, H-10, J = 17), 6.31 (d, 1H, H-14, J = 17), 6.55 (t, 1H, H-12, J = 12), 7.15 (d, 1H, H-20, J = 9), 7.21 (t, 1H, H-5, J = 7), 7.27 (d, 1H, H-3, J = 8), 7.30 – 7.37 (m, 1H, H-4), 7.42 (d, 1H, H-6, J = 7), 7.72 – 7.79 (m, 2H, H-21, H-23), 8.15 (dd x 2, 2H, H-11, H-13, J = 13)

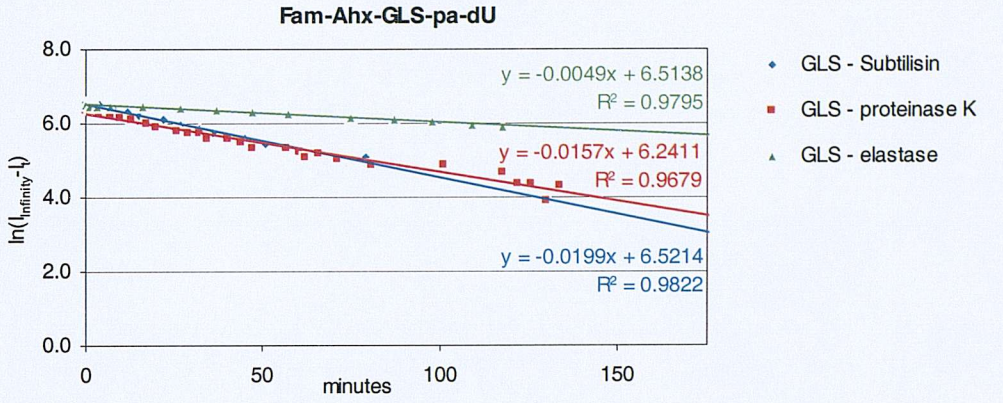
δ ^{13}C NMR (100 MHz, $CDCl_3$): 22.0, 23.5 (2 x $\underline{C}H_3$-Leu), 25.9 ($COCH_2CH_2\underline{C}H_2CH_2CH_2$), 26.2, 27.2 ($CH_2$-Leu, CH-Leu), 27.8, 27.9 (4 x $\underline{C}H_3$-Cy5), 28.3 ($COCH_2CH_2CH_2\underline{C}H_2CH_2$), 31.4 (N-$\underline{C}H_3$), 36.3 ($CO\underline{C}H_2CH_2CH_2CH_2CH_2$), 41.8 ($COCH_2\underline{C}H_2CH_2CH_2CH_2$), 43.7 ($CH_2$-Ser), 45.2 ($COCH_2CH_2CH_2CH_2\underline{C}H_2$), 56.3 (CH-Leu), 56.9, 60.0 (C-8, C-19), 57.1 (CH-Ser), 62.9 (CH_2-Gly), 104.1 (C-10), 105.7 (C-14), 110.8 (C-20), 112.7 (C-3), 121.2 (C-21), 123.5 (C-6), 126.9 (C-5), 127.2 (C-12), 128.0 (C-23), 129.0 (C-4), 142.1 (C-2), 142.8, 143.1, 143.3 (C-9, C-17, C-18), 145.8 (C-22), 155.0 (C-11), 156.6 (C-13), 171.7, 173.0, 174.4, 174.7 ($\underline{C}OOH$, $\underline{C}$=O-Leu, $\underline{C}$=O-Gly, $\underline{C}$=O-Ahx), 176.2, 176.6 (C-7, C-18)

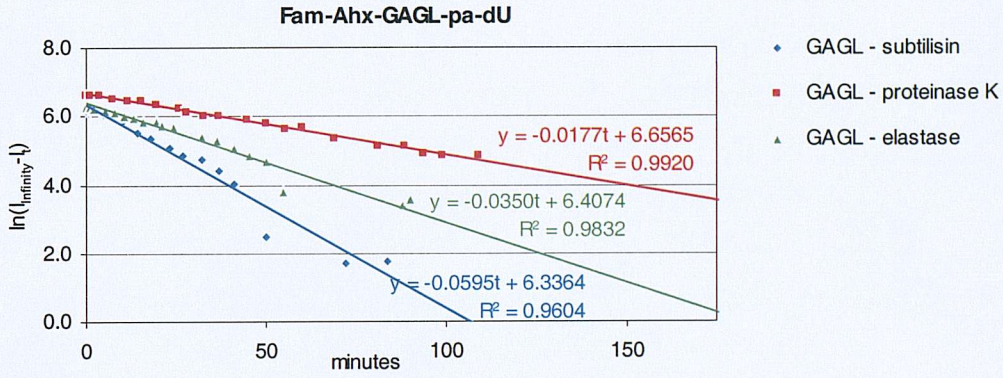
8 Appendix

8.1 Logarithmic plots of the hydrolysis of 90, 91, 92 and 93






Fam-Ahx-GLS-pa-dU
8.0
6.0
4.0
2.0
0.0
0
50
100
150
minutes
ln(I_Infinity-I_t)
y = -0.0049x + 6.5138
R² = 0.9795
y = -0.0157x + 6.2411
R² = 0.9679
y = -0.0199x + 6.5214
R² = 0.9822
GLS - Subtilisin
GLS - proteinase K
GLS - elastase


Fam-Ahx-GAGL-pa-dU
8.0
6.0
4.0
2.0
0.0
0
50
100
150
minutes
ln(I_Infinity-I_t)
y = -0.0177t + 6.6565
R² = 0.9920
y = -0.0350t + 6.4074
R² = 0.9832
y = -0.0595t + 6.3364
R² = 0.9604
GAGL - subtilisin
GAGL - proteinase K
GAGL - elastase

9 References

1 Merrifield R.B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.

2 Gutte, B; Merrifield, R.B. *J. Am. Chem. Soc.* **1969**, *91*, 501-502.

3 Denkewalter, R.G.; Veber, D.F.; Holly, F.W.; Hirshmann, R. **1969**, *91*, 502-503.

4 Gutte, B.; Merrifield, R.B. *J. Biol. Chem.* **1971**, *246*, 1922-1941.

5 Damha, M.J.; Giannaris, P.A.; Zabarylo, S.V. *Nucleic Acids Res.* **1990**, *13*, 3813-3821.

6 Tegge, W.; Frank, R. *J. Peptide Res.* **1997**, *49*, 355-362.

7 Meldal, M.; Svendsen, I., Breddam, K.; Auzanneau, F.-I. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3314-3318.

8 Leon, S.; Quarrell, R.; Lowe, G.; *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2997-3002.

9 Frank, R. *Tetrahedron* **1992**, *42*, 9217-9232.

10 Larhed, M.; Lindeberg, G.; Hallberg, A. *Tetrahedron Lett.* **1996**, *37*, 8219-8222.

11 Gooding, O.W.; Baudart, S.; Deegan, T.L.; Heisler, K.; Labadie, J.W.; Newcomb, W.S.; Porco, J.A.; van Eikeren, P. *J. Comb. Chem.* **1999**, *1*, 113-122.

12 Cho, J.K.; Park, B.-D.; Lee, Y.-S. *Tetrahedron Lett.* **2000**, 7481-7485.

13 Li, W.; Xiao, X.; Czarnik, A.W. *J. Comb. Chem.* **1999**, *1*, 127-129.

14 Vagner, J.; Barany, G.; Lam, K.S.; Krchnak, V.; Sepetov, N.F.; Ostrem, J.A.; Strop, P.; Lebl, M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8194-8199.

15 Lou, Q.; Leftwich, M.E.; Lam, K.S. *Bioorg. Med. Chem.* **1996**, *5*, 677-682.

16 Martin, S.E.; Peterson, B.R. *J. Peptide Sci.* **2002**, *8*, 227-233.

17 Atherton, E.; Clive, D.L.J.; Sheppard, R.C. *J. Am. Chem. Soc.* **1975**, *97*, 6584-6585.

18 Meldal, M. *Tetrahedron Lett.* **1992**, *33*, 3077-3080.

19 Rossé, G.; Kueng, E.; Page, M.G.P.; Schauer-Vukasinovic, V.; Giller, T.; Lahm, H.-W.; Hunziker, P.; Schlatter, D. *J. Comb. Chem.* **2000**, *2*, 461-466.

20 Meldal, M.; Svendsen, I. *J. Chem. Soc. Perkin Trans. I* **1995**, 1591-1596.

21 Meldal, M.; Svendsen, I.; Juliano, L.; Juliano, M.; Del Nery, E.; Scharfstein, *J. Peptide Sci.* **1998**, *4*, 83-91.

22 Peterson, J.J.; Meares, C.F. *Bioconjugate Chem.* **1998**, *9*, 618-626.

23 Meldal, M.; Auzanneau, F.-I.; Hindsgaul, O.; Palcic, M.M. *J. Chem. Soc., Chem. Commun.* **1994**, 1849-1850.

24 Renil, M.; Ferreras, M.; Delaisse, J.M.; Foged, N.T.; Meldal, M. *J. Peptide Sci.* **1998**, *4*, 195-210.

25 Kress, J.; Zanaletti, R.; Amour, A.; Ladlow, M.; Frey, J.G.; Bradley, M. *Chem. Eur. J.* **2002**, *8*, 3769-3772.

26 Altava, B.; Burguete, I.; García-Verdugo, E.; Luis, S.V.; Vicent, M.J. *Tetrahedron Lett.* **2001**, *42*, 8459-8462.

27 St. Hilaire, P.M.; Willert, M.; Juliano, M.A.; Juliano, L.; Meldal, M. *J. Comb. Chem.* **1991**, *1*, 509-523.

28 Geysen, H.M.; Meloen, R.H.; Barteling, S.J. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 3998-4002.

29 Takahashi, T.; Inoue, H.; Tomida, S.; Doi, T. *Tetrahedron Lett.* **1999**, *40*, 7843-7846.

30 Houghten, R.A. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 5131-5135.

31 Xiao, X.-Y.; Li, R.; Zhuang, H.; Ewing, B.; Karunaratne, K.; Lillig, J.; Brown, R.; Nicolaou, K.C. *Biotechnol. Bioeng.* **2000**, *71*, 44-50.

32 Furka, A.; Sebestyen, F.; Major, F.; Dibo, F. *Int. J. Peptide Protein Res.* **1991**, *37*, 98-101.

33 Lam, K.S.; Salmon, S.E.; Hersh, E.M.; Hruby, V.J.; Kazmierski, W.M.; Knapp, R.J. *Nature* **1991**, *354*, 82-84.

34 Montgomery, J.A. *Med. Res. Rev.* **1982**, 271-308.

35 Watson, J.D.; Crick, F.H.C. *Nature* **1953**, *171*, 737-738.

36 Franklin, R.E.; Gosling, R.G. *Nature* **1953**, *171*, 740-741.

37 Franklin, R.E.; Gosling, R.G. *Nature* **1953**, *172*, 156-157.

38 Wilkins, M.H.F.; Stokes, A.R.; Wilson, H.R. *Nature* **1953**, *171*, 738-740.

39 Maxam A.M., Gilbert W., *Proc. Natl. Acad. Sci.* USA **1977**, *74*, 560-564.

40 Sanger F., Nicklen S., Coulson A.R., *Proc. Natl. Acad. Sci.* USA **1977**, *74*, 5463-5467.

41 Mizusawa, S.; Nishimura, S.; Seela, F. *Nucleic Acids Res.* **1986**, *14*, 1319-1324

42 Ansorge, W.; Sproat, B.S.; Stegemann, J.; Schwager, C. *J. Biochem. Biophys. Methods* **1986**, *13*, 315-323

43 Ansorge, W.; Sproat, B.S.; Stegemann, J.; Schwager, C.; Zenke, M. *Nucleic Acids Res.* **1986**, *15*, 4593-4602

44 Ju, J; Ruan, C.; Fuller, C.W.; Glazer, A.N.; Mathies, R.A. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 4347-4351.

45 Cui, L.; Schinazi, R.F.; Gosselin, G.; Imbach, J.-L.; Chu, C.K.; Rando, R.F.; Revankar, G.R. Sommadossi, J.-P. *Biochem. Pharmacol.* **1996**, *52*, 1577-1584.

46 Cristalli, G.; Franchetti, P.; Grifantini, M.; Vittori, S.; Bordoni, T.; Geroni, C. *J. Med. Chem.* **1987**, *30*, 1686-1688.

47 Krawczyk, S.H.; Renau, T.E.; Nassiri, M.R.; Westerman, A.C; Wotring, L.L.; Drach, J.C.; Townsend, L.B. *J. Med. Chem.* **1995**, *38*, 4115-4119.

48 Krawczyk, S.H.; Nassiri, M.R.; Kucera, L.S.; Kern, E.R.; Ptak, R.G.; Wotring, L.L.; Drach, J.C.; Townsend, L.B. *J. Med. Chem.* **1995**, *38*, 4106-4114.

49 Porcari, A.R.; Ptak, R.G.; Borysko, K.Z; Breitenbach, J.M.; Drach, J.C; Townsend, L.B. *J. Med. Chem.* **2000**, *43*, 2457-2463.

50 Robins, M.J.; Miles, R.W.; Samano, M.C.; Kaspar, R.L. *J. Org. Chem.* **2001**, *66*, 8204-8210.

51 Cocuzza, A.J. *Tetrahedron Lett.* **1988**, *33*, 4061-4068.

52 Seela, F.; Becher, G. *Helv. Chim. Acta* **1999**, *82*, 1640-1655.

53 Seela, F.; Zulauf, M. *J. Chem. Soc., Perkin Trans. 1*, **1998**, 3233-3239.

54 Seela, F.; Zulauf, M. *Chem. Eur. J.***1998**, *9*, 1781-1790.

55 Seela, F.; Wei, C.; Becher, G.; Zulauf, M.; Leonard, P. *Bioorg. Med. Chem. Lett.* **2000**, 289-292.

56 Seela, F.; Mersmann, K. *Helv. Chim. Acta* **1993**, *76*, 2184-2193.

57 Seela, F.; Lampe, S. *Helv. Chim. Acta* **1993**, *76*, 2388-2397.

58 Kazimierczuk, Z.; Binding, U.; Seela, F. *Helv. Chim. Acta* **1989**, *72*, 1527-1535.

59 Seela, F.; Münster, I.; Löchner, U.; Rosemeyer, H. *Helv. Chim. Acta* **1998**, *81*, 1139-1155.

60 Horwitz, J.P.; Chua, J.; Noel, M. *J. Org. Chem.* **1964**, *29*, 2076-2078.

61 Eschenhof, H.; Strazewski, P.; Tamm, C. *Tetrahedron* **1992**, *30*, 6225-6230.

62 Nesnow, S.; Miyazaki, T.; Khwaja, T.; Meyer Jr, R.B.; Heidelberger, C. *J. Med. Chem.* **1973**, *5*, 524-528.

63 Skulnick, H. *J. Org. Chem.* **1978**, *16*, 3188-3194.

64 Cappellacci, L.; Tiwari, K.N.; Montgomery, J.A.; Secrist III, J.A. *Nucleosides & Nucleotides* **1999**, *18*, 613-614.

65 Hanna, N.B.; Zajíček, J.; Pískala, A. *Nucleosides & Nucleotides* **1997**, *16*, 129-144.

66 Novotny, L.; Vachalkova, A.; Piskala, A. *Bioelectrochem. Bioenerg.* **1999**, *48*, 129-134.

67 Niedballa, U.; Vorbrüggen, H. *J. Org. Chem.* **1974**, *25*, 3654-3660.

68 El-Barbary, A.A.; Khodair, A.I.; Pedersen, E.B. *J. Org. Chem.* **1993**, *58*, 5994-5999.

69 Basnak, I.; Coe, P.L.; Walker, R.T. *Nucleosides & Nucleotides* **1994**, *13*, 163-175.

70 Dewey, T.M.; Mundt, A.A.; Crouch, G.J.; Zyzniewski, M.C.; Eaton, B.E. *J. Am. Chem. Soc.* **1995**, *117*, 8474-8475.

71 Prober, J.M.; Trainor, G.L.; Dam, R.J.; Hobbs, F.W.; Robertson, C.W.; Zagursky, R.J.; Cocuzza, A.J.; Jensen, M.A.; Baumeister, K. *Science* **1987**, *238*, 336-341.

72 Hobbs Jr, F.W. *J. Org. Chem.* **1989**, *54*, 3420-3422.

73 Lee, S.E.; Vyle, J.S.; Williams, D.M.; Grasby, J.A. *Tetrahedron Lett.* **2000**, *41*, 267-270.

74 Robins, M.J.; Barr, P.J. *J. Org. Chem.* **1983**, *48*, 1854-1862.

75 Lee, S.E.; Sidorov, A.; Gourlain, T.; Mignet, N.; Thorpe, S.J.; Brazier, J.A.; Dickman, M.J.; Hornby, D.P.; Grasby, J.A.; Williams, D.M. *Nucleic Acids Res.* **2001**, *7*, 1565-1573.

76 Caruthers, M.H. *Science* **1985**, *230*, 281-285.

77 Graham, S.M.; Pope, S.C. *Org. Lett.***1999**, *5*, 733-736.

78 Verma, S.; Eckstein, F. *Annu. Rev. Biochem.* **1998**, *67*, 99-134

79 Hartsel, S.A.; Marshall, W.S. *Bioorg. Med. Chem. Lett.* **1996**, *24*, 2993-2998.

80 Schwarz, M.W.; Pfleiderer, W. *Tetrahedron Lett.* **1984**, *48*, 5513-5516.

81 Khorana, H.G.; Turner, A.F.; Vizsolyi, J.P. *J. Am. Chem. Soc.* **1961**, *83*, 686-698.

82 Tetzlaff, C.N.; Schwope, I.; Bleczinski, C.F.; Steinberg, J.A.; Richert, C. *Tetrahedron Lett.* **1998**, *39*, 4215-4218.

83 Gryaznov, S.M.; Letsinger, R.L. *J. Am. Chem. Soc.* **1991**, *113*, 5876-5877.

84 Hayakawa, Y.; Kataoka, M. *J. Am. Chem. Soc.* **1998**, *120*, 12395-12401.

85 Bogdan, F.M.; Chow, C.S. *Tetrahedron Lett.* **1998**, *39*, 1897-1900.

86 Greenberg, M.M.; Matray, T.J.; Kahl, J.D.; Yoo, D.J.; McMinn, D.L. *J. Org. Chem.* **1998**, *63*, 4062-4068.

87 Krider, E.S.; Miller, J.E.; Meade, T.J. *Bioconjugate Chem.* **2002**, *13*, 155-162.

88 Seitz, O. *Angew. Chem. Int. Ed.* **2000**, *39*, 3249-3252.

89 Rodenko, B; Wanner, M.J.; Koomen, G.-J. *J. Chem. Soc. Perkin Trans.* **2002**, *1*, 1247-1252.

90 de Champdoré, M.; De Napoli, L.; Di Fabio, G.; Messere, A.; Montesrchio, D.; Piccialli, G. *Chem. Commun.* **2001**, 2598-2599.

91 Khan, S.I.; Grinstaff, M.W. *Tetrahedron Lett.* **1998**, *39*, 8031-8034.

92 Choo, H.; Chong, Y.; Chu, C.K. *Org. Lett.* **2001**, *10*, 1471-1473.

93 Früchtel, J.S.; Jung, G. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 17-42.

94 Guiller, F.; Orain, D.; Bradley, M. *Chem. Rev.* **2000**, *100*, 2091-2057.

95 Furman, B.; Thürman, R.; Kaluża, Z.; Voelter, W.; Chmielewski, M. *Tetrahedron Lett.* **1999**, *40*, 5909-5912.

96 Pon, R.T.; Yu, S.; Sanghvi, Y.S. *Bioconjugate Chem.* **1999**, *10*, 1051-1057.

97 Walsh, A.J.; Clark, C.; Fraser, W. *Tetrahedron Lett.* **1997**, *9*, 1651-1654.

98 Meloni, M.M.; Brown, R.C.D.; White, P.D.; Armour, D. *Tetrahedron Lett.* **2002**, *43*, 6023-6026.

99 Nakamura, K.; Hanai, N.; Kanno, M.; Kobayashi, A.; Ohnishi, Y.; Ito, Y.; Nakahara, Y. *Tetrahedron Lett.* **1999**, *40*, 515-518.

100 Zheng, C.; Seeberger, P.H.; Danishefsky, S.J.; *J. Org. Chem.* **1998**, *63*, 1126-1130.

101 Roberge, J.Y.; Beebe, X.; Danishefsky, S.J. *J. Am. Chem. Soc.* **1998**, *120*, 3915-3927.

102 Hanessian, S.; Xie, F. *Tetrahedron Lett.* **1998**, *39*, 733-736.

103 Yan, L.Z.; Mayer, J.P. *J. Org. Chem.* **2003**, *68*, 1161-1162.

104 Hanessian, S.; Huynh, H.K. *Tetrahedron Lett.* **1999**, *40*, 671-674.

105 Fréchet, J.M.J.; Nuyens, L.J. *Can. J. Chem.* **1976**, *54*, 926-934.

106 Fyles, T.M.; Leznoff, C.C. *Can. J. Chem.* **1976**, *54*, 935-942.

107 Chen, C.; Ahlberg Randall, L.A.; Miller, R.B.; Jones, A.D.; Kurth, M.J. *J. Am. Chem. Soc.* **1994**, *116*, 2661-2662.

108 Wenschuh, H.; Beyermann, M.; Haber, H.; Seydel, J.K.; Krause, E.; Bienert, M. *J. Org. Chem.* **1995**, *60*, 405-410.

109 Melby, L.R.; Strobach, D.R. *J. Am. Chem. Soc.* **1967**, *89*, 450-453.

110 Cramer, F.; Köster, H. *Angew. Chem. Int. Ed.* **1968**, *6*, 473-474.

111 Ref 105, 106 and suggested method in *Advanced ChemTech Handbook of Combinatorial & Solid Phase Organic Chemistry* **1998**, ISBN 0-9665883-0-4

112 Barlos, K.; Gatos, D.; Kallitsis, J.; Papaphotiu, G.; Sotiriu, P.; Wenqing, Y.; Schäfer, W. *Tetrahedron Lett.* **1989**, *30*, 3943-3946.

113 Eleftheriou, S.; Gatos, D.; Panagopoulos, A.; Stathophoulos, S.; Barlos, K. *Tetrahedron Lett.* **1999**, *40*, 2825-2828.

114 Chaudhary, S.K.; Hernandez, O. *Tetrahedron Lett.* **1979**, *2*, 95-98

115 Chaudhary, S.K.; Hernandez, O. *Tetrahedron Lett.* **1979**, *2*, 99-102

116 Davison, A.; McClean, J.; Vaman Rao, V.; Brown, T. *Biomed. Pept. Proteins Nucleic Acids* **1996**, *1*, 1-6.

117 McClean, J.; Davison, A.; Vaman Rao, V.; Brown, T. *Biomed. Pept. Proteins Nucleic Acids* **1996**, *1*, 7-12.

118 Bijapur, J., Keppler, M.S.; Bergqvist, S.; Brown, T.; Fox, K.R. *Nucleic Acids Res.* **1999**, *27*, 1802-1809.

119 Kypr, J.; Sági, J.; Szakonyi, E.; Ebinger, K.; Peňázová, H.; Chládková, J.; Vorlíčková, M. *Biochemistry* **1994**, *33*, 3801-3806.

120 Yu, K.-L.; Deshpande, M.S.; Vyas, D.M. *Tetrahedron Lett.* **1994**, *35*, 8919-8922.

121 Hiroshige, M.; Hauske, J.R.; Zhou, P. *Tetrahedron Lett.* **1995**, *36*, 4567-4570.

122 Hiroshige, M.; Hauske, J.R.; Zhou, P. *J. Am. Chem. Soc.* **1995**, *117*, 11590-11591.

123 Nelson, J.C.; Young, J.K.; Moore, J.S. *J. Org. Chem.* **1996**, *61*, 8160-8168.

124 Crisp, G.T.; Gore, J. *Tetrahedron* **1997**, *53*, 1523-1544.

125 Dyatkin, A.B.; Rivero, R.A. *Tetrahedron Lett.* **1998**, 3647-3650 .

126 Berteina, S.; Wendeborn, S.; Brill, W. K.-D.; De Mesmaeker, A. *Synlett* **1998**, 676-678.

127 Khan, S.I.; Grinstaff, M.W. *J. Am. Chem. Soc.* **1999**, *121*, 4704-4705.

128 Stieber, F.; Grether, U.; Waldmann, H. *Angew. Chem. Int. Ed.* **1999**, *38*, 1073-1077.

129 Davies, M.J.; Shah, A.; Bruce, I.J. *Chem. Soc. Rev.* **2000**, *29*, 97-107.

130 Bräse, S.; Dahmen, S.; Heuts, J. *Tetrahedron Lett.* **1999**, *40*, 6201-6203.

131 Böhm, V.P.; Herrmann, W.A. *Eur. J. Org. Chem.* **2000**, 3679-3681.

132 Liao, Y.; Fathi, R.; Reitman, M.; Zhang, Y.; Yang, Z. *Tetrahedron Lett.* **2001**, *42*, 1815-1818.

133 Dyatkin, A.B.; Rivero, R.A. *Tetrahedron Lett.* **1998**, *39*, 3647-3650.

134 He, J.; Seela, F. *Nucleic Acids Res.* **2002**, *30*, 5485-5496.

135 Snyder, K.R.; Murray, T.F.; DeLander, G.E.; Aldrich, J.V. *J. Med. Chem.*, **1993**, *36*, 1100-1103.

136 Schneider, S.E.; Bishop, P.A.; Salazar, M.A.; Bishop, O.A.; Anslyn, E.V. *Tetrahedron* **1998**, *54*, 15063-15086.

137 Masiero, S.; Fini, F., Mangini, A., *J. Chem. Res. (M)* **1998**, 2737-2745.

138 Hood, L.; Galas, D. *Nature* **2003**, *421*, 444-448.

139 Huang, X.C.; Quesada, M.A.; Mathies, R.A. *Anal. Chem.* **1992**, *64*, 967-972.

140 Soper, S.A.; Legendre, B.L.; Williams, D.C. *Anal. Chem.* **1995**, *67*, 4358-4365.

141 Lieberwirth, U.; Arden-Jacob, J.; Drexhage, K.H.; Herten, D.P.; Muller, R.; Neumann, M.; Schulz, A.; Siebert, S.; Sagner, G.; Klingel, S.; Sauer, M.; Wolfrum, J. *Anal. Chem.* **1998**, *70*, 4771-4779.

142 Liu, S.; Shi, Y.; Ja, W.W.; Mathies, R.A. *Anal. Chem.* **1999**, *71*, 566-573.

143 Smith, L.M.; Sanders, J.Z.; Kaiser, R.J.; Hughes, P.; Dodd, C.; Connell, C.R.; Heiner, C.; Kent, S.B.H.; Hood, L.E. *Nature* **1986**, *321*, 674-679.

144 Wojczewski, C.; Stolze, K.; Engels, J.W. *Synlett* **1999**, *10*, 1667-1678.

145 Metzker, M.L.; Lu, J.; Gibbs, R.A. *Science* **1996**, *271*, 1420-1422.

146 Nunnally, B.K.; He, H.; Li, L.-C.; Tucker, S.A.; McGown, L.B. *Anal. Chem.* **1997**, *69*, 2392-2397.

147 Ju, J.; Glazer, A.N.; Mathies, R.A. *Nucleic Acids Res.* **1996**, *24*, 1144-1148.

148 Yamakawa, H.; Ohara, O. *Nucleic Acids Res.* **1997**, *25*, 1311-1312.

149 Finn, P.J.; Sun, L.; Nampalli, S.; Xiao, H.; Nelson, J.R.; Mamone, J.A.; Grossmann, G.; Flick, P.K.; Fuller, C.W.; Kumar, S. *Nucleic Acids Res.* **2002**, *30*, 2877-2885.

150 Edwards, J.R.; Itagaki, Y.; Ju, J. *Nucleic Acids Res.* **2001**, *29*, e104, 1-6.

151 Rosenblum, B.B.; Lee, L.G.; Spurgeon, S.L.; Khan, S.H.; Menchen, S.M.; Heiner, C.R.; Chen, S.M. *Nucleic Acids Res.* **1997**, *25*, 4500-4504.

152 Zhang, J.; Yang, M.; Puyang, X.; Fang, Y.; Cook, L.M.; Dovichi, N.J. *Anal. Chem.* **2001**, *73*, 1234-1239.

153 Soper, S.A.; Williams, D.C.; Xu, Y.; Lassiter, S.; Zhang, Y.; Ford, S.M. *Anal. Chem.* **1998**, *70*, 4036-4043.

154 Caviani Pease, A.; Solas, D.; Sullivan, E.J.; Cronin, M.T.; Holmes, C.P.; Fodor, S.P.A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 5022-5026.

155 Preparata, F.P.; Frieze, A.M.; Upfal, E. *Third Annual Int. Conference on Computational Molecular Biology Lyon* **1999**, 295-301.

156 Frieze, A.M.; Preparata, F.P.; Upfal, E. *J. Comput. Biol.* **1999**, *6*, 361-368.

157 Ronaghi, M.; Uhlén, M.; Nyrén, P. *Science* **1998**, *281*, 363-365.

158 Pourmand, N.; Elahi, E.; Davis, R.W.; Ronaghi, M. *Nucleic Acids Res.* **2002**, *30*, e31, 1-5.

159 Ronaghi, M. *Genome Res.* **2001**, *11*, 3-11.

160 Metzker, M.L.; Raghavachari, R.; Richards, S.; Jacutin, S.E.; Civitello, A.; Burgess, K.; Gibbs, R.A. *Nucleic Acids Res.* **1994**, *22*, 4259-4267.

161 Welch, M.B.; Martinez, C.I.; Zhang, A.J.; Jin, S.; Gibbs, R.; Burgess, K. *Chem. Eur. J.* **1999**, *5*, 951-960.

162 Bai, X.; Li, Z.; Jockusch, S.; Turro, N.J.; Ju, J. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 409-413.

163 Li, Z.; Bai, X.; Ruparel, H.; Kim, S.; Turro, N.J.; Ju, J. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 414-419.

164 Odedra, R. Patent WO 02/02813 A2, **2002**.

165 Odedra, R.; Pickering, L.; Simmonds, A.; Johnson, K. Patent WO 03/048178 A2, **2003**.

166 Odedra, R; Pickering, L. Patent WO 03/027677 A2, **2003**.

167 Personal communication with Lea Pickering Amersham Biosciences

168 Barrett, A.J.; Rawlings, N.D.; Woessner, J.F. Handbook of protelytic enzymes, Academic Press **1998** ISBN 0-12-079370-9

169 Hansske, F.; Robins, M.J. *Tetrahedron Lett.* **1983**, *24*, 1589-1592.

170 Andersson, F.; Samuelsson, B. *Carbohydrate* **1984**, *129*, C1-C3.

171 Hansske, F.; Madej, D.; Robins, M.J. *Tetrahedron* **1984**, *40*, 125-135.

172 Kim, J.N.; Ryu, E.K. *Tetrahedron Lett.* **1992**, *33*, 3141-3144.

173 Classon, B.; Samuelsson, B. *Acta Chem. Scand., Ser B* **1985**, *39*, 501-504.

174 Corey, E.J.; Samuelsson, B. *J. Org. Chem.* **1984**, *49*, 4735-4735.

175 Personal communication with L. Pickering, Amersham Biosciences

176 Braven, H. PhD Thesis, **1999**, University of Southampton

177 Battersby, T.R.; Ang, D.N.; Burgstaller, P.; Jurczyk, S.C.; Bowser, M.T.; Buchanan, D.D.; Kennedy, R.T.; Benner, S.A. *J. Am. Chem. Soc.* **1999**, *121*, 9781-9789.