

Open Research Online

The Open University's repository of research publications and other research outputs

Molecular Recognition of Damaged DNA Using Synthetic Affinity Reagents

Thesis

How to cite:

Turner, Jayne Susan (2004). Molecular Recognition of Damaged DNA Using Synthetic Affinity Reagents. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2004 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Molecular Recognition of Damaged DNA Using Synthetic Affinity Reagents

Thesis submitted for the degree of
Doctor of Philosophy
to The Open University

by

Jayne Susan Turner

Department of Chemistry
The Open University

October 2004



AUTHOR NO. U6382239
DATE OF SUBMISSION 29 OCTOBER 2004
DATE OF AWARD 11 AUGUST 2005

Abstract

Potentially carcinogenic DNA damage that may be derived from dietary sources can be quantified using a number of methods. Some of the best of these methods are based on the use of antibodies that have high affinity and specificity for modified DNA bases. However, given the sometimes limited availability of antibodies there is a requirement for novel reagents that mimic the properties of the best antibodies. From a chemical perspective antibodies are 'over engineered' and synthetic affinity reagents would recreate the optimal properties of the binding site with respect to selectivity and affinity.

Phage display was used to identify amino acid residues that contribute to recognition of a diet derived DNA adduct O⁶-carboxymethyl-2'-deoxyguanosine. However, this information was not sufficient to synthesise a novel peptide reagent directly and it was decided to develop an approach whereby a library of compounds was synthesised which includes the characteristics expected within the binding site of an antibody. This library was based around cholic acid, a readily available scaffold molecule with suitable functionality compatible with linking amino acids and a fluorescent tag.

A synthetic procedure based on the formation of stable oxime derivatives of cholic acid was optimised and a small library of pyrene-tagged trisubstituted cholic acid derivatives was characterised. The library was examined for its ability to bind to affinity columns containing immobilised O⁶-carboxymethylguanosine and there was some evidence that indicated specific binding of a subset of library molecules.

Acknowledgements

I would like to express my sincere gratitude to Professor D.E.G. Shuker for his supervision, advice and guidance throughout this project. I thank the Open University for financial support. My gratitude must go to Mr G. Jeffs for his never ending optimism and Mr P. Patel for his precise laboratory practice. I also thank the Department of Biology at the Open University, particularly Dr M. Hirst and Mrs Julie Bone. Finally to my husband, Mark, and daughters, Jodi and Jenifer, thank you for your encouragement and emotional support.

Dedicated to the memory of Mr Paul E. Wright.

Who will appear in the mortality statistics for oesophageal cancer, 2004, and who reminded me that behind the numbers are people.

Abbreviations

2D-TLC	2 dimensional thin layer chromatography
A	adenine
APS	ammonium persulphate
ATase	O ⁶ -alkylguanine-DNA alkyltransferase
BSA	bovine serum albumin
C	cytosine
°C	degrees centigrade
CH ₂ Cl ₂	dichloromethane
CHCl ₃	chloroform
D	Daltons
DBU	1,8-diazabicyclo[5.4.0.]undec-7-ene
DCC	<i>N,N</i> -dicyclohexylcarbodiimide
DCM	dichloromethane
DMAP	dimethylaminopyridine
DMF	<i>N,N</i> -dimethylformamide
DNA	deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetra-acetic acid (disodium salt)
ES	electrospray
Et ₃ N	triethylamine
EtOAc	ethyl acetate
EtOH	ethanol
fmol	femtomoles
FMOC	<i>N</i> - α -(9-fluorenylmethyloxycarbonyl)
g	gram
G	guanosine
GlyOBz	glycine benzyl ester
HCl	hydrochloric acid
HPLC	high performance liquid chromatography
HOMO	highest occupied molecular orbital
id	internal diameter
IPTG	isopropyl- β -D-thiogalactoside
KBr	potassium bromide
KCl	potassium chloride
KOH	potassium hydroxide

L	litre
LB	Luria-Bertani
LiAlH ₄	lithium aluminium hydride
MeOH	methanol
μg	microgram
mg	milligram
min	minute
mL	millilitre
μL	microlitre
mm	millimetre
mM	millimolar
μM	micromolar
mmol	millimoles
μmol	micromoles
MIP	molecularly imprinted polymer
MHz	megahertz
mp	melting point
MS	mass spectrometry
<i>m/z</i>	mass to charge ratio
NaIO ₄	sodium periodate
Na ₂ SO ₄	sodium sulphate
NaCl	sodium chloride
NaOH	sodium hydroxide
NDMA	<i>N</i> -nitrosodimethylamine
NER	nucleotide excision repair
nmol	nanomoles
NMR	nuclear magnetic resonance
NOC	<i>N</i> -nitroso compounds
O ⁶ CMdG	<i>O</i> ⁶ -carboxymethyl-2'-deoxyguanosine
O ⁶ CMG	<i>O</i> ⁶ -carboxymethyl guanosine
O ⁶ MG	<i>O</i> ⁶ -methyl guanosine
OD	optical density
OV	ovalbumin
PBS	phosphate buffered saline
PBS-T	PBS-Tween 20
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming unit
PyCA	1'-pyrenyl 3,7,12 trioxo-(5β)-cholan-24-amide
rf	ratio value of distance between applied spot and eluted spot divided by distance between applied spot and solvent front
RNA	ribonucleic acid
rpm	revolutions per minute
SELEX	systematic evolution of ligands by exponential enrichment

S_N	nucleophilic substitution
SPPS	solid phase peptide synthesis
T	thymine
TASP	template-assembled synthetic protein
TBE	tris borate ethylenediaminetetraacetic acid
TBS	tris-buffered saline
TBST	tris-buffered saline Tween 20
TEMED	tetramethylethylenediamine
tet	tetracycline
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TrCl	triphenylmethyl chloride
Tris	tris[hydroxymethyl]aminomethane
Tween 20	polyoxyethylenesorbitan monolaurate
UV	ultraviolet
Xgal	5-bromo-4-chloro-3-indolyl- β D-galactoside

Contents

CHAPTER 1	- 1 -
INTRODUCTION	- 1 -
1.1 Introduction	- 2 -
1.2 Cancer.....	- 3 -
1.2.1 Statistics of Cancer Cases	- 4 -
1.2.2 Etiology of cancer	- 4 -
1.2.3 Gastrointestinal Cancers	- 5 -
1.2.4 Biomarkers.....	- 8 -
1.3 <i>The Alkylation of DNA by N-Nitroso Compounds</i>	- 9 -
1.3.1 N-Nitroso compounds	- 9 -
1.3.2 Exogenous Sources of NOCs.....	- 10 -
1.3.3 Endogenous Nitrosation.....	- 10 -
1.3.4 Nitrosated Glycine Derivatives.....	- 12 -
1.3.5 Mechanism of Adduct Formation	- 13 -
1.4 <i>Repair Mechanisms and Consequences of Non Repair</i>	- 15 -
1.4.1 Consequences of Adduct Persistence	- 17 -
1.5 <i>Methods of detecting O⁶-Carboxymethyl deoxyguanosine</i>	- 19 -
1.5.1 Immunoaffinity/HPLC	- 19 -
1.5.2 Immunoblot-technique	- 20 -
1.6 <i>Molecular Recognition</i>	- 22 -
1.6.1 Antibody recognition	- 23 -
1.6.2 Intermolecular forces	- 24 -
1.7 <i>Chemical Genetics</i>	- 26 -
1.8 <i>Chemical diversity</i>	- 28 -
1.8.1 Peptide synthesis.....	- 29 -
1.8.2 Template-assembled synthetic protein (TASP)	- 31 -
1.8.3 Non-peptide libraries	- 32 -
1.8.4 Biological based libraries.....	- 33 -
1.8.5 Steroid based libraries.....	- 34 -
1.8.6 Dynamic combinatorial libraries	- 34 -
1.9 <i>Directed Molecular Imprinting</i>	- 35 -
1.9.1 Molecular Imprinting Polymers.....	- 35 -

1.10	<i>Steroids as scaffold molecules for diversity-oriented synthesis</i>	- 36 -
1.10.1	An Introduction to Steroids and Bile Acids.....	- 36 -
1.10.2	The Uses of Cholic Acid and its Derivatives.....	- 38 -
1.11	<i>The Suitability of Cholic Acid as a Scaffold</i>	- 40 -
1.12	<i>Aims of this thesis</i>	- 41 -
CHAPTER 2		- 44 -
PHAGE DISPLAY.....		- 44 -
2.1	<i>Introduction</i>	- 45 -
2.2	<i>An introduction to Phage Display</i>	- 45 -
2.3	<i>History</i>	- 47 -
2.4	<i>Background</i>	- 48 -
2.5	<i>Application</i>	- 49 -
2.6	<i>Indication of phage presence</i>	- 50 -
2.7	<i>Method of Use</i>	- 53 -
2.7.1	Selection.....	- 53 -
2.7.2	Titering.....	- 53 -
2.7.3	Amplification	- 55 -
2.8	<i>DNA determination</i>	- 56 -
2.9	<i>Results</i>	- 59 -
2.10	<i>Discussion</i>	- 61 -
CHAPTER 3		- 63 -
CONSTRUCTION AND CHARACTERISATION OF CHOLATE-AMINO ACID LIBRARIES FOR RECOGNITION OF O ⁶ -CARBOXYMETHYL-2'-DEOXYGUANOSINE		- 63 -
3.1	<i>Introduction</i>	- 64 -
3.2	<i>Synthesis of immobilised O⁶-carboxymethylguanine</i>	- 64 -
3.2.1	Synthesis	- 64 -
3.2.2	Conjugation of the ligand to matrices.....	- 67 -
3.2.2.1	Preparation of O ⁶ -carboxymethyl guanine bound to ovalbumin	- 67 -
3.2.2.2	Determination of the hapten-carrier ratio	- 70 -
3.2.2.3	Preparation of O ⁶ -carboxymethyl guanine bound to glass beads	- 70 -

3.3	<i>Synthesis and characterisation of cholate-amino acid libraries</i>	- 71 -
3.3.1	The development of amide bonded cholic acid derivatives.....	- 71 -
3.3.2	The development of a general synthesis of oxime bonded cholic acid derivatives	- 73 -
3.3.2.1	Incorporation of a fluorescent marker	- 73 -
3.3.2.2	The incorporation of a spacer unit	- 75 -
3.3.3	Oxime formation.....	- 79 -
3.3.3.1	The α -effect	- 80 -
3.3.4	Oxime bond formation.....	- 80 -
3.3.5	The scaffold molecule.....	- 82 -
3.4	<i>Synthesis of library components</i>	- 83 -
3.4.1	Synthesis of aminooxy compounds	- 83 -
3.5	<i>Synthesis of tri-substituted 1'-pyrenyl methyl-3,7,12 trioximo-(5β)-cholan-24-amide (PyCA) compounds</i>	- 87 -
3.5.1	Pilot studies on library formation.	- 89 -
3.5.1.1	PyCA Mix 1	- 89 -
3.5.1.2	PyCA Mix 2	- 92 -
3.6	<i>Synthesis and analysis of the library</i>	- 95 -
3.7	<i>Regioselective oxime formation</i>	- 101 -

CHAPTER 4..... - 102 -

ANALYSIS OF CHOLATE-AMINO ACID LIBRARIES FOR RECOGNITION OF O⁶-

	CARBOXYMETHYL-2'-DEOXYGUANOSINE	- 102 -
4.1	<i>Chromatography</i>	- 103 -
4.2	<i>HPLC development for the analysis of the cholic acid libraries</i>	- 104 -
4.3	<i>Two dimensional (2D)-TLC method for the visualization of cholic acid libraries</i>	- 108 -
4.4	<i>Affinity chromatography</i>	- 118 -
4.4.1	Introduction.....	- 118 -
4.4.2	Column matrix	- 119 -
4.4.3	Eluant system	- 121 -
4.4.4	Column capacity	- 121 -
4.4.5	Antibody interaction with the column matrices.....	- 122 -
4.5	<i>Fluorescence detection</i>	- 125 -

4.5.1.	The Fluorimeter	- 125 -
4.5.2	Quenching.....	- 126 -
4.5.3	Excitation and emission	- 126 -
4.5.4	Background fluorescence.....	- 127 -
4.5.5	Columns	- 128 -
CHAPTER 5.....		- 134 -
DISCUSSION AND FUTURE PERSPECTIVES		- 134 -
5.1	<i>Discussion and Future Perspectives</i>	- 135 -
CHAPTER 6.....		- 140 -
EXPERIMENTAL.....		- 140 -
6.1	<i>General and Instrumentation</i>	- 141 -
6.1.1	Physical Measurements.....	- 141 -
6.1.2	Reagents and Reactants	- 141 -
6.2	<i>Synthesis of DNA Adduct and conjugation to carrier matrices</i>	- 142 -
6.2.1	2',3',5'-tri- <i>O</i> -isobutyryl guanosine.....	- 142 -
6.2.2	<i>O</i> ⁶ -[(2,4,6-trimethylphenyl)sulphonyl]-2',3',5'-tri- <i>O</i> - isobutyrylguanosine	- 143 -
6.2.3	<i>O</i> ⁶ -(carboxymethyl, methyl ester)-2',3',5'-tri- <i>O</i> - isobutyrylguanosine	- 144 -
6.2.4	<i>O</i> ⁶ -(carboxymethyl)guanosine	- 145 -
6.2.5	<i>O</i> ⁶ -carboxymethylguanosine and Ovalbumin conjugate	- 146 -
6.2.6	Determination of the adduct to protein ratio.....	- 146 -
6.2.7	<i>O</i> ⁶ -carboxymethylguanosine and glass bead conjugate.....	- 147 -
6.2.8	<i>O</i> ⁶ -carboxymethylguanosine and agarose conjugate	- 147 -
6.2.9	Guanosine and agarose conjugate.....	- 148 -
6.3	<i>Phage Display Experimental</i>	- 148 -
6.3.1	Equipment and reagents.....	- 148 -
6.3.2	Phage titer of library	- 149 -
6.3.3	Panning and titer	- 149 -
6.3.4	Amplification	- 150 -
6.3.5	Subsequent pannings.....	- 150 -
6.3.6	DNA purification	- 151 -

6.3.7	PCR amplification.....	- 152 -
6.3.8	DNA sequencing.....	- 152 -
6.3.9	Media and solutions used with phage display.....	- 153 -
6.4	<i>Synthesis of cholic acid derivatives</i>	- 154 -
6.4.1	3 α ,7 α ,12 α -trihydroxy-(5 β)-cholan-24-oic acid methyl ester...	- 154 -
6.4.2	3 α ,7 α ,12 α -trihydroxy-(5 β)-cholan-24-ol.....	- 155 -
6.4.3	3,7,12 trioxo-(5 β)-cholan-24-oic acid methyl ester.....	- 156 -
6.4.4	3,7,12 trioximo-(5 β)-cholan-24-oic acid methyl ester.....	- 157 -
6.4.5	Succinimido-(<i>N</i>)-3,7,12 trioxo-(5 β)-cholan-24-oate	- 158 -
6.4.6	1'-pyrenyl methyl-3 α ,7 α ,12 α trioxo-(5 β)-cholan-24-amide ...	- 159 -
6.4.7	1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide.....	- 160 -
6.4.8	1'-pyrenyl methyl-3,7,12 trioximomethyl-(5 β)-cholan-24- amide.....	- 161 -
6.4.9	1'-pyrenyl methyl-3,7,12 trioximocarboxymethyl-(5 β)-cholan-24- amide.....	- 162 -
6.4.10	1'-pyrenyl methyl-3,7,12 trioximobenzyl-(5 β)-cholan-24-amide	- 163 -
6.5	<i>Synthesis of N-(aminooxy acetyl) amino acids</i>	- 164 -
6.5.1	((1',1'-dimethylethoxycarbonyl)aminooxy)ethanoic acid.....	- 164 -
6.5.2	Succinimido-(<i>N</i>)-((1',1'-dimethylethoxycarbonyl)aminooxy) ethanoate	- 165 -
6.5.3	Glycine benzyl ester toluene-4-sulphonate.....	- 166 -
6.5.4	Glycine benzyl ester,.....	- 166 -
6.5.5	N-(1''-glycine benzyl ester)-((1',1'- dimethylethoxycarbonyl)aminooxy) ethanamide	- 166 -
6.5.6	N-(1''-glycine benzyl ester)aminooxy ethanamide	- 167 -
6.5.7	N-(1''- glycine methyl ester)-((1',1'- dimethylethoxycarbonyl)aminooxy) ethanamide	- 168 -
6.5.8	N-(1''-glycine methyl ester)aminooxy ethanamide.....	- 168 -
6.5.9	Phenylalanine methyl ester	- 169 -
6.5.10	N-(1''-phenylalanine methyl ester)-((1',1'- dimethylethoxycarbonyl)aminooxy) ethanamide	- 169 -
6.5.11	N-(1''-phenylalanine methyl ester)aminooxy) ethanamide.....	- 170 -
6.5.12	Lysine (Fmoc) methyl ester	- 170 -

6.5.13	N-(1''-lysine (FMOC) methyl ester)-((1',1'-dimethylethoxycarbonyl) aminoxy) ethanamide.....	- 170 -
6.5.14	N-(1''-lysine (FMOC) methyl ester)aminoxy) ethanamide....	- 171 -
6.5.15	N-(1''-arginine (NO ₂) methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide	- 172 -
6.5.16	N-(1''-arginine (N ^G -NO ₂) methyl ester)aminoxy) ethanamide	- 173 -
6.5.17	N-(1''-arginine methyl ester)aminoxy) ethanamide	- 173 -
6.6	<i>Synthesis of tri substituted 1'-pyrenyl methyl-3,7,12 trioxo-(5β)-cholan-24-amide</i>	- 174 -
6.6.1	1'-pyrenyl methyl-3,7,12 tri(N-iminoxymethylenecarbonyl (glycine benzyl ester).....	- 174 -
6.6.2	1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine-(5β)-cholan-24-amide	- 175 -
6.6.3	1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) phenylalanine-(5β)-cholan-24-amide	- 176 -
6.6.4	1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) lysine-(5β)-cholan-24-amide	- 178 -
6.6.5	1'-pyrenylmethyl-3((N-iminoxymethylenecarbonyl) arginine) 7,12(dioxo)-(5β)-cholan-24-amide	- 179 -
6.6.6	1'-pyrenyl methyl-3,7,12 trioximo-(5β)-cholan-24-amide tri substituted with a mixture of methyl and benzyl hydroxyalmines -	180 -
6.6.7	1'-pyrenyl methyl-3,7,12 trioxo-(5β)-cholan-24-amide library containing all aminoxy amino acid dervitatives	- 180 -
6.7	<i>Sakaguchi Reagent</i>	- 181 -
CHAPTER 7.....		- 182 -
REFERENCES		- 182 -
7.1 References		- 183 -

Chapter 1

Introduction

**‘Let us hope that this Association (American Association for Cancer Research) will be out of business by the year 2000!’ – James A. Miller
March 1969**

1.1 Introduction

In spite of James A. Miller's hopeful words in 1969, the American Association for Cancer Research is very much still in business. At their 95th Annual Meeting, in 2004, 5,700 abstracts were submitted for presentation and about 15,000 researchers attended, giving an indication of how active the area of cancer research is in the 21st century. This meeting brought together oncologists, students, biologists, chemists, as well as cancer survivors to facilitate a cross-disciplinary exchange of ideas and knowledge. The more that we learn about cancer, the more questions we find that require answers.

The critical difference between a normal and a cancer cell is a discrete change in specific genes that control proliferation and tissue homeostasis, by the activation of oncogenes and/or the inactivation of tumour suppressor genes. More than 100 cancer related genes have been identified and they are consistently found in a mutated or in a different form in tumours¹. The most striking is the p53 tumour suppressor gene which is found, mutated, in about half of all human cancer cases².

p53 mutations usually occur at the pre-invasive stage events of many cancers¹ with the G-C to A-T transition, caused by DNA adducts, being the most frequent change. DNA adducts are covalent modifications of DNA, which if not repaired can be converted into mutations during cellular replication, leading to initiation and progression of cancer tumours.

One major source of DNA adducts are *N*-nitroso compounds which can come from exogenous or endogenous sources. Many *N*-nitroso compounds have been studied

and their carcinogenic properties identified. *N*-nitroso compounds can interact with DNA to result in the formation of modified bases. These modifications can exert a detrimental effect on cell reproduction, especially when their associated repair mechanisms fail. A DNA base that is modified by attachment of an alkyl residue is called an adduct, and adducts have been associated with risk of tumour development. One area of major interest in cancer research is the detection of normal background levels of adducts so that quantitative comparisons can be made with the high levels associated with tumours.

Various methods of detecting DNA adducts, have been developed based on the recognition properties of antibodies. These rely on the presence of intermolecular forces which allow for specific binding of two molecules that have a high affinity for each other. Mimicking this action, but between a chemically synthesised molecule and a biological one has formed the basis of the work of many research chemists and is an ever expanding field. One of the key aspects of this is the formation of chemical libraries, broadly analogous to antibody repertoires, which have high levels of diversity. This population can then be screened to reveal any members within it that may be suitable candidates for molecular recognition studies.

1.2 Cancer

The first disease causing mutation in a human gene was identified in 1978³ and as more genes and their products are being identified so we increase our understanding of the consequences of genetic mutation. The differences in mutational frequency among all the different human cancer types suggest that both endogenous and exogenous factors contribute to the etiology of cancer and have implications for human cancer risk assessment¹.

1.2.1 Statistics of Cancer Cases

In 2000, 320,000 new cases of cancer were recorded in England; the cost of cancer was £1 billion being spent on the NHS and £260 million on research (£25 million from the Government). Less than two percent of cases occur in women under 40, and men under 50, and there is then a steady increase in cases until a peak is reached at the age of 70-75 in both sexes. The most common cancer in men is prostate cancer, followed by cancers of the lung and the colon/rectum. In women the most common is breast cancer, followed by colorectal and then lung. A total of 28,300 cases of colorectal cancer were confirmed for both sexes in 2000. Colorectal cancer is therefore the most significant cancer of the gastrointestinal tract, followed by gastric cancer and then oesophageal cancer. Interestingly, while the incidences of gastric cancer are decreasing by around 20% per year, those of the oesophagus are steadily increasing particular with regards to male patients⁴.

1.2.2 Etiology of cancer

Carcinogenesis is a multi-step processes, which is why it has been difficult for researchers to be able to say definitively what causes cancer and what can be used to 'cure' cancer. A recognised way of discussing cancer is in terms of risk, and the minimisation of this risk⁵. Studies on animals have lead to a traditional view of the etiology of cancer development which is demonstrated in Figure 1.1. The first step is the 'initiation' step, in which damage or modification to the cellular DNA occurs as a response to chemical, physical or microbial carcinogens. These may alter the ability of cells to respond to their microenvironment and give them a proliferative advantage⁶. The 'promotion' stage may result in the survival of the initiated cell and its proliferation, which may be greater than that of the surrounding normal cells. The probability of further genetic damage is increased and further exposure to DNA-

agents may result in the activation of proto-oncogenes and/or the inactivation of tumour suppressor genes, and these events are also passed on through the expanding population of cells⁷. This is now recognised as being the ‘progression’ step⁶.

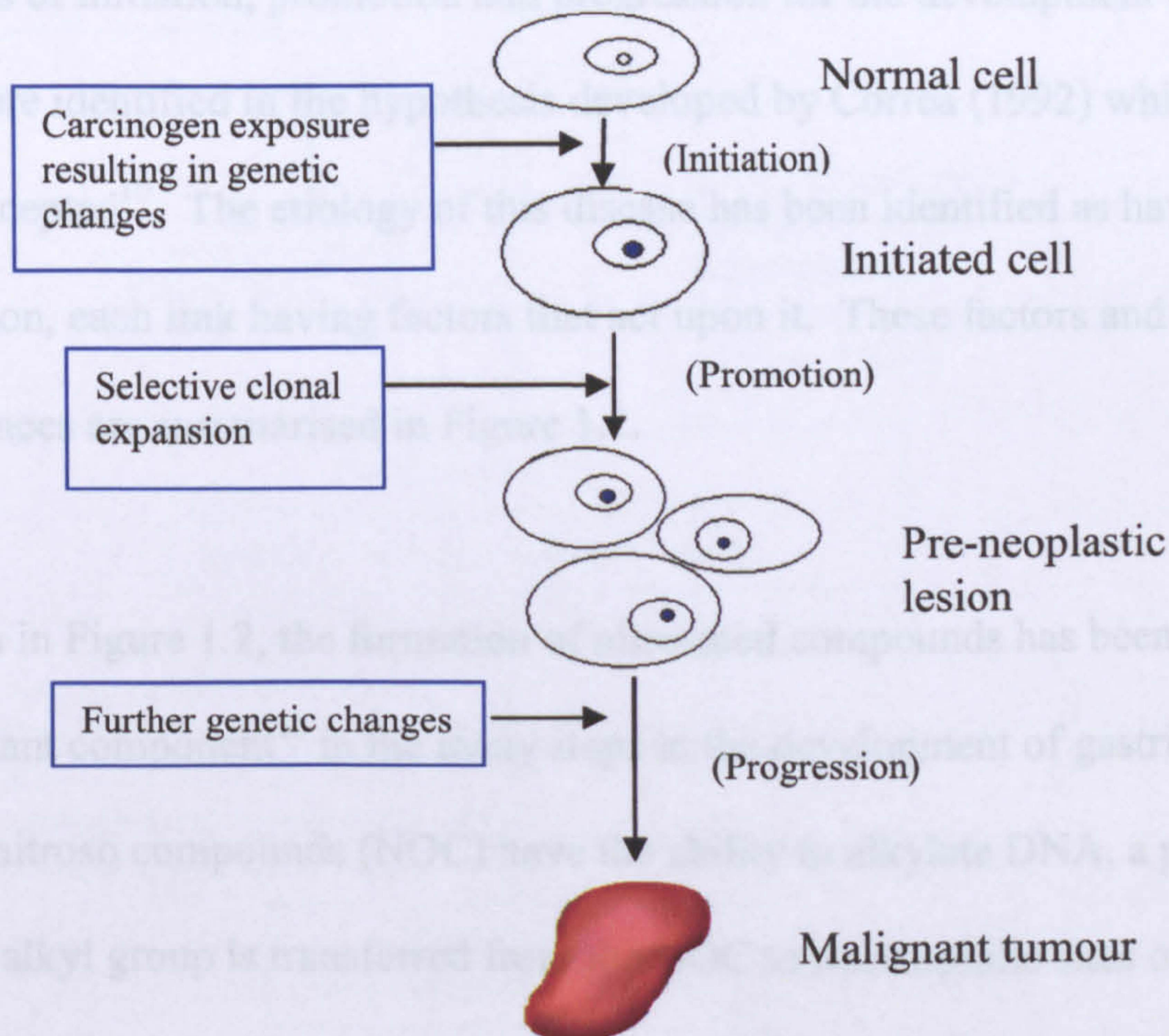


Figure 1-1. Carcinogenesis is a multi-step process which involves genetic events within key genes. These genetic changes result in an initiated cell, which if it survives cellular replication, passes on these changes to all subsequent daughter cells. This can lead to clonal expansion of pre-neoplastic lesions. Further events result in the activation of proto-oncogenes and/or inactivation of tumour suppressor genes resulting in a malignant tumour and clinical cancer. (Adapted from reference 6).

1.2.3 Gastrointestinal Cancers

The high incidences of colorectal cancer and the steady increase of oesophageal cancer have made gastrointestinal cancers the centre of much research and although the rates of gastric cancer have declined worldwide, probably due to better food

preservation and storage⁸, they still remain high in Japan, Eastern Europe and in areas of Latin America⁹.

The stages of initiation, promotion and progression for the development of gastric cancer were identified in the hypothesis developed by Correa (1992) which has been widely accepted¹⁰. The etiology of this disease has been identified as having a chain of causation, each link having factors that act upon it. These factors and their consequences are summarised in Figure 1.2.

As shown in Figure 1.2, the formation of nitrosated compounds has been identified as an important component¹⁰ in the many steps in the development of gastric cancer.

These *N*-nitroso compounds (NOC) have the ability to alkylate DNA, a process in which an alkyl group is transferred from the NOC to nucleophilic sites on the DNA, forming a DNA adduct. The presence of such adducts are considered to be a key step in the induction of cancer^{7,11}.

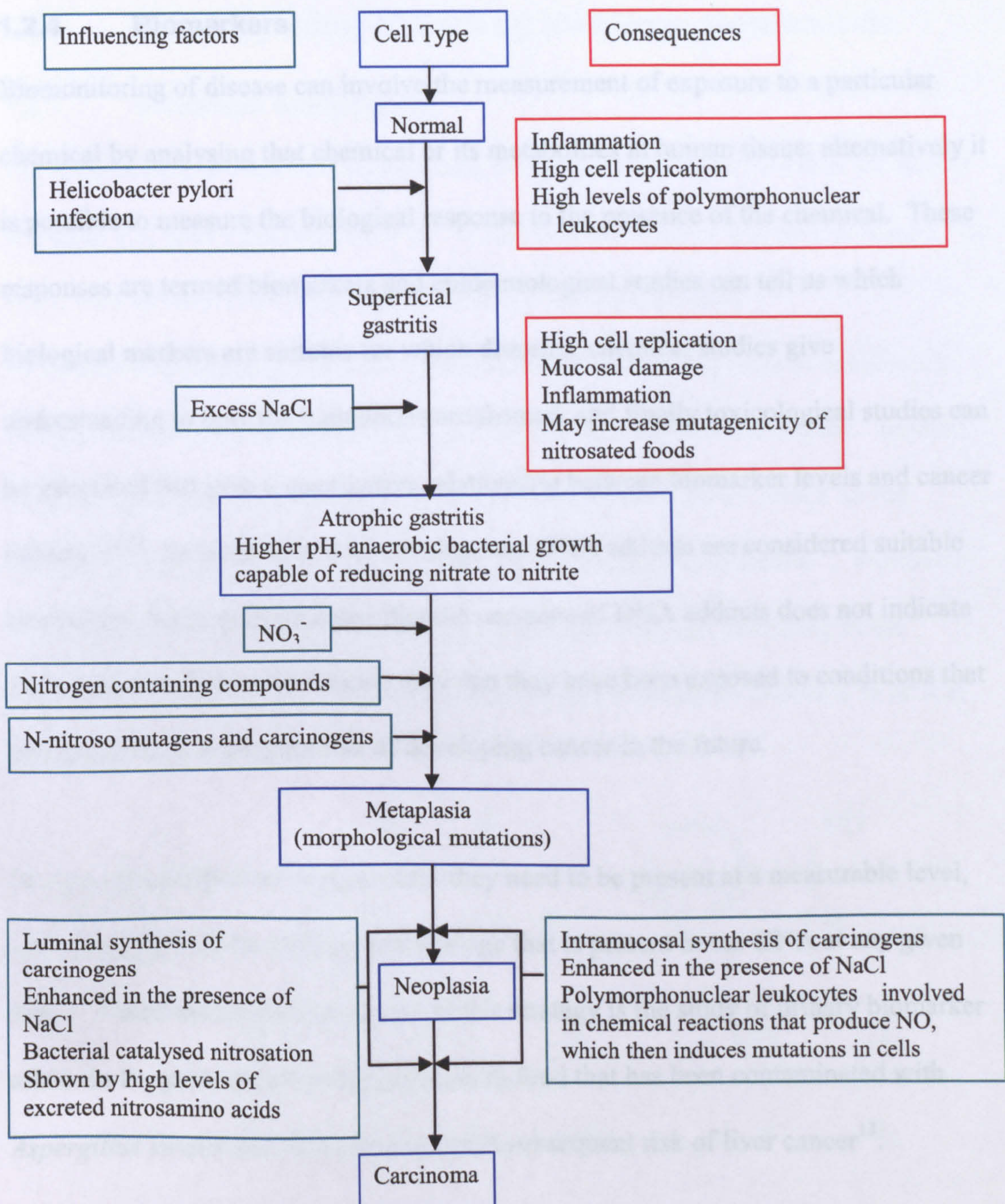


Figure 1-2. A summary of the etiology of gastric cancer as a multistep process.

(Based on Reference 10.)

1.2.4 Biomarkers

Biomonitoring of disease can involve the measurement of exposure to a particular chemical by analysing that chemical or its metabolites in human tissue; alternatively it is possible to measure the biological response to the presence of the chemical. These responses are termed biomarkers and epidemiological studies can tell us which biological markers are suitable for which diseases; chemical studies give understanding to how the biomarkers are formed, and finally toxicological studies can be generated that give a quantitative relationship between biomarker levels and cancer induction^{6,12}. In terms of cancer development DNA adducts are considered suitable biomarkers, but it must be noted that the presence of DNA adducts does not indicate that a person will develop cancer, only that they have been exposed to conditions that may place them at a higher risk of developing cancer in the future.

For these biomarkers to be significant they need to be present at a measurable level, and above those of the background damage that is present in our DNA at any given time⁵. A well-documented example of this strategy is the study of urinary biomarker aflatoxin B₁-guanine following exposure to food that has been contaminated with *Aspergillus* strains and its association with subsequent risk of liver cancer¹³.

Exposure assessment can be carried out at different stages in the process that can lead to the development of tumours but the measurement of biomarkers, particularly those that result in the alkylation of DNA are of considerable importance. One group of the most extensively studied alkylating agents are the NOCs.

1.3 The Alkylation of DNA by N-Nitroso Compounds

1.3.1 N-Nitroso compounds

The toxic effects of *N*-nitroso compounds, the generic structures of which are shown in Figure 1.3, have long been known¹⁴. Following the work of Barnes and Magee¹⁵ in 1956, who induced liver cancer in rats with the administration of nitrosodimethylamine, much work has been carried out on this class of compounds and this work continues today such is the importance of these carcinogens. Humans have been shown to be susceptible to the carcinogenic actions of NOCs on the grounds of the biochemical and histopathological data that is available¹⁵. Particular work has concentrated on NOCs as possible factors in cancers occurring in the gastrointestinal tract¹⁶, research has therefore focused on human exposure to NOCs.

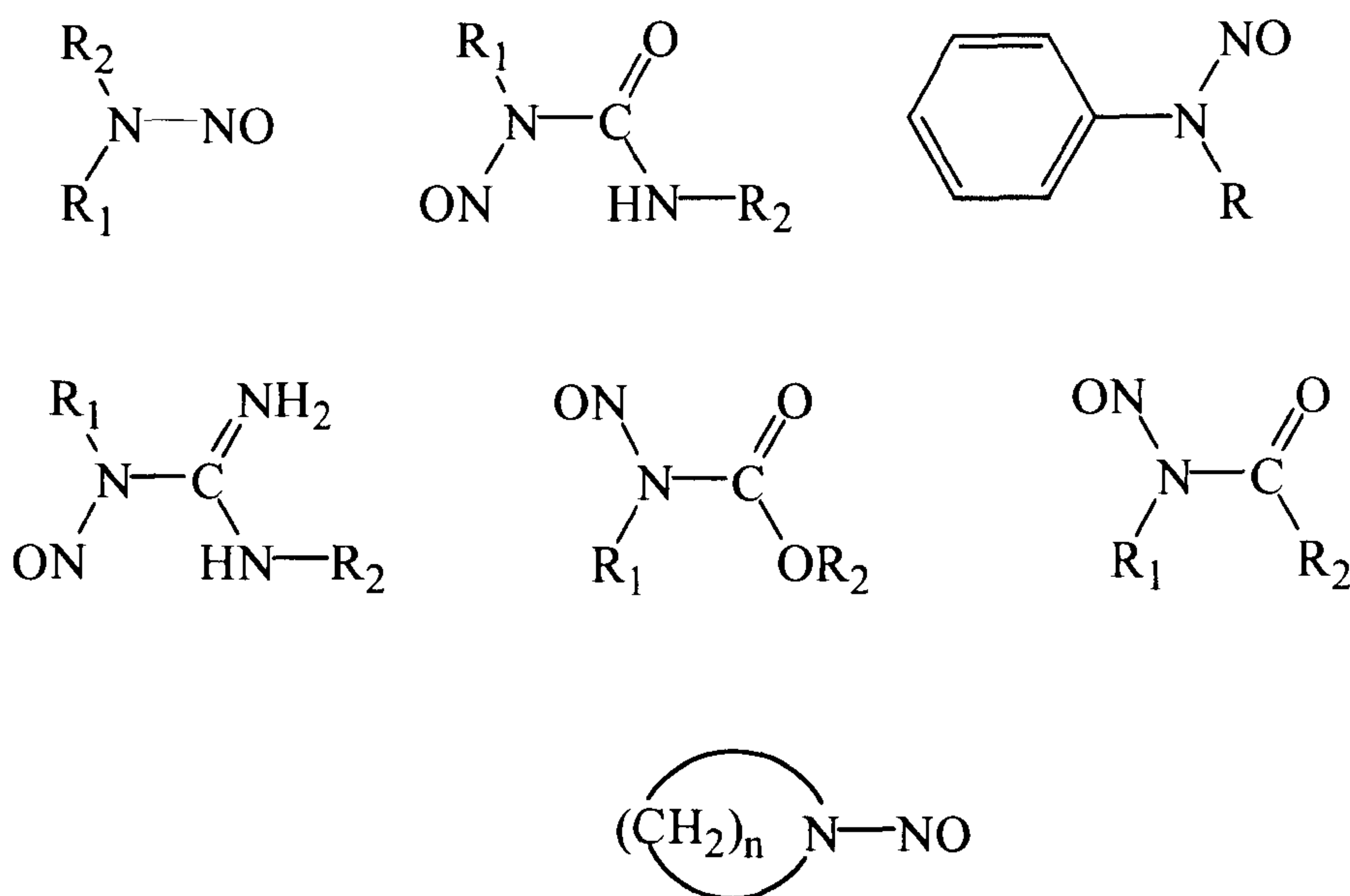


Figure 1-3. The generic structures of the main types of *N*-nitroso compounds. (From Lijinsky, 1992, *Chemistry and biology of N-nitroso compounds*, p2¹⁴)

1.3.2 Exogenous Sources of NOCs

Humans are exposed to a number of NOCs that are exogenous in origin. Tobacco products contain 4-(*N*-methylnitrosamino)-1-(3-pyridyl)-1-butanone and *N*'-nitrosornicotine, both potent carcinogens related to human cancers of the mouth, oesophagus and lungs¹⁷. Several industrial processes have been identified as being sources of occupational exposure for workers including rubber production and leather tanning¹⁸. NOCs have also been found as contaminants of food, either due to the processing procedures or from packaging⁷. The diet of the Chinese and Japanese, which is high in salted fish and pickled vegetables have been inferred as a correlation with the high incidences of oesophageal and gastric cancers found in those areas⁹ and¹⁹. Alcohol consumption in Western countries is thought to be a major contributing factor to the development of oesophageal cancer. Anderson *et al*²⁰ reproduced the effects of low to moderate alcohol consumption in humans in a study using primates. They showed that the co-administration of ethanol and nitrosamine *N*-nitrosodimethylamine (NDMA) resulted in a significant increase of methyl alkylation at the O⁶ position of 2'-deoxyguanosine in the DNA of the digestive tract²¹.

1.3.3 Endogenous Nitrosation

As well as being exposed to preformed NOCs humans are exposed to NOCs that are formed inside the body in endogenous processes. The principal agent is the nitrite ion, which is primarily found in the saliva as a result of the reduction of nitrate, which is in turn based on dietary intake¹⁴, and nitrite ion concentration is one of the factors influencing the rate of nitrosation. The optimal pH for nitrosation depends on which precursor is available but the resting pH of the stomach (~ 1.5-2) rises after the ingestion of food to around pH 5²², this allows the formation of intermediates responsible for nitrosation¹⁰.

Helicobacter pylori infection also raises the pH levels in the stomach and when these are higher than pH 4 the stomach can be colonised by bacteria capable of the reduction of nitrate to nitrite which in turn gives rise to the possibility of an increased source of precursors of NOCs. Bacterially mediated nitrosation has been shown to be inhibited by the presence of ascorbate²³. This shows that nitrosation is under the influence of both catalysts and inhibitors.

A standard human diet contains many components that can be nitrosated in the gastrointestinal tract to form NOCs and precursors include:

- i) classes of alkylamines^{24,25}
- ii) aromatic amines^{26,25}
- iii) guanidines²⁵
- iv) ureas^{24,25}
- v) bile acids^{27,28-30}
- vi) amino acids, amides and peptides²⁴

These precursors differ greatly in the amount consumed and rate of nitrosation, and their products also differ markedly in carcinogenicity²⁵. Many NOCs decompose spontaneously to generate alkyl or aromatic diazonium ions. These are strong reactive electrophiles and as such can react with less nucleophilic centres on DNA such as exocyclic base oxygens, forming a DNA adduct, via S_N1 mechanisms. With regards to the gastrointestinal tract, interest has turned to the availability of amino acids due to their high concentrations within the digestive tract from the proteolytic breakdown of proteins of a dietary source²⁵. In humans who have an increased risk of

gastrointestinal tumour formation have been shown to have higher levels of NOCs in gastric juice and *O*⁶-methyldeoxyguanosine in their DNA³¹.

1.3.4 Nitrosated Glycine Derivatives

Glycine is one of the most common amino acids present in the diet and has the simplest structure; it would therefore be a likely candidate as a major source of alkylating agents in the gastrointestinal tract following nitrosation.

Stable *N*-nitrosoglycine derivatives have been shown to produce a potent alkylating agent, the carboxymethyldiazonium ion, in a number of studies. These include *N*-nitrosoglycocholic acid²⁷ and *N*-(*N*'-acetyl-L-prolyl)-*N*-nitrosoglycine²⁹ (Figure 1.4). Interestingly, a natural form of a nitrosated glycine, azaserine exists and was tested as an anticancer agent. However, it is now regarded as a carcinogen and is widely used to produce experimental pancreatic cancer in mammals.

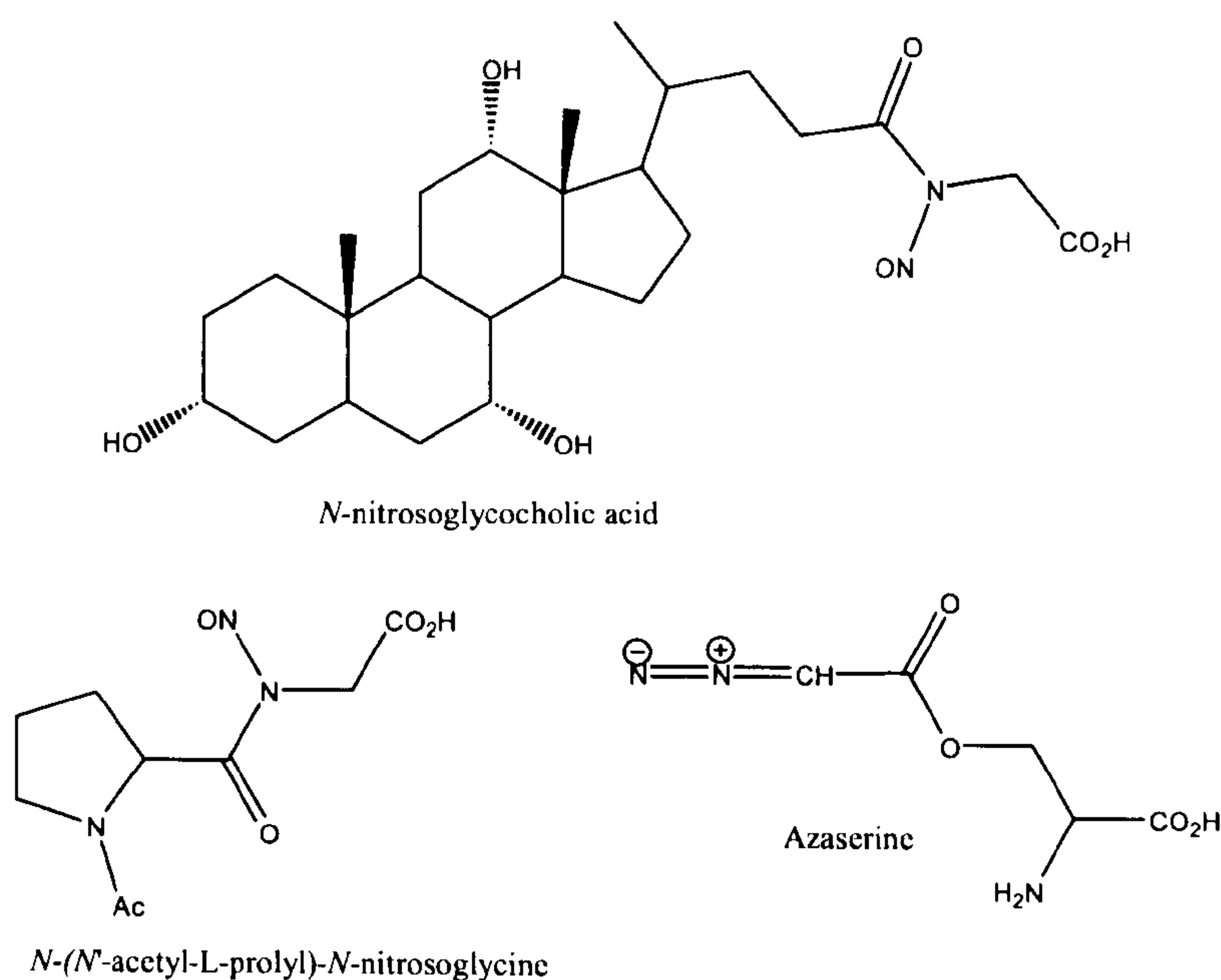


Figure 1-4. The structures of some nitrosated glycine derivatives.

1.3.5 Mechanism of Adduct Formation

In confirming the importance of nitrosated glycine derivatives, it has been shown that key intermediates can be isolated and lead to the expected products. The proposed mechanism for nitrosation of glycine itself is shown in Figure 1.5, which shows that nitric oxide firstly combines with oxygen to form dinitrogen trioxide. This species can react with glycine in neutral and alkaline conditions and its hydrolysis yields the diazoacetate anion which is stable enough to be detected by HPLC in alkali conditions. Diazoacetate reacts with deoxyguanosine to give the expected products, in physiological conditions³². The products can lead to a characteristic p53 mutation spectrum found in tumours³³.

The electrophilic species formed by the NOCs are classed as being relatively hard acids and the theory of Pearson proposes that hard acids have a preference to react with hard bases, while soft acids react preferentially with soft bases. Hard acids have high positive charge and low polarizability, with polarizability being defined as the ease with which the electrons of a molecule are distorted³⁴. Alkylating species have been classified as being intermediate, as have ring nitrogen and exocyclic oxygen atoms on DNA (see Figure 1.7)²². Therefore, although nitrogen and oxygen atoms are alkylated by many alkylating agents, NOCs give significant adducts on oxygen atoms.

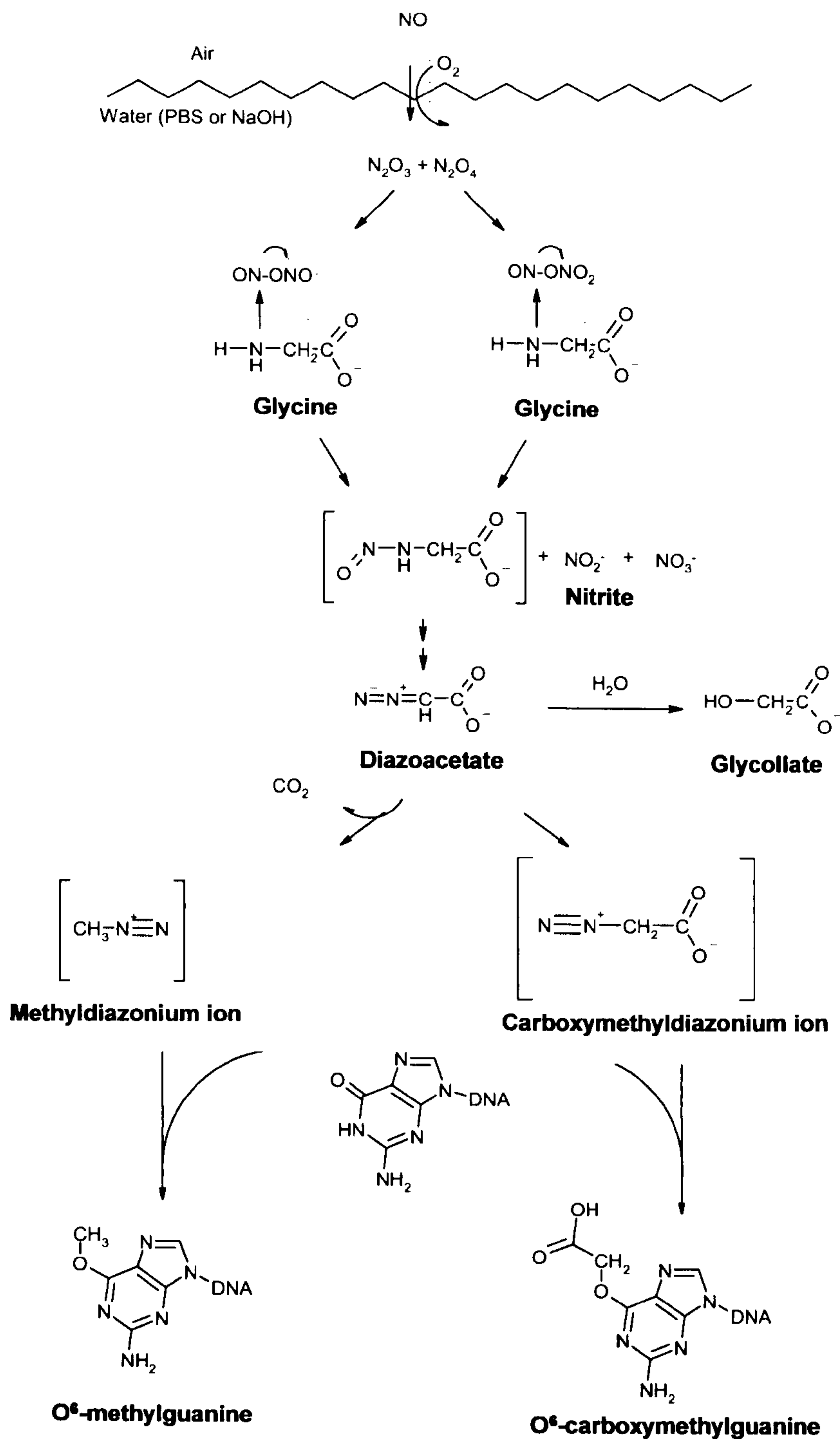


Figure 1-5. The reaction pathway of the formation of O⁶-methyl and O⁶-carboxymethylguanine adducts, following the formation of glycine derived diazoacetate, and subsequently the alkyldiazonium ion³².

DNA alkylating intermediates have been detected following a reaction between nitric oxide and glycine³³ and it is known that these react with DNA to form O^6 -methyldeoxyguanosine (O^6 MdG) as a minor product and O^6 -carboxymethyldeoxyguanosine (O^6 CMdG) as a major product³², structures as shown in Figure 1.6. With the latter being present at ten times the concentration of O^6 MdG in experiments that have been done to simulate those that occur in the gastrointestinal tract³⁵.

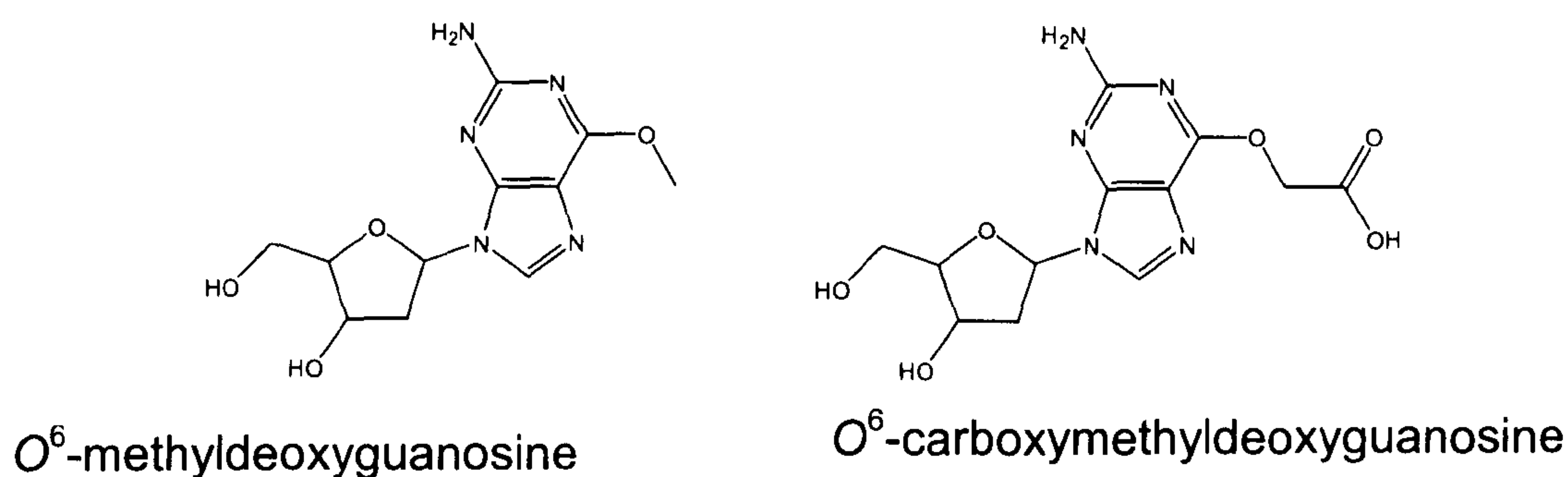


Figure 1.6. The structures of DNA adducts known to form in the presence of alkylating agents.

1.4 Repair Mechanisms and Consequences of Non Repair

Before DNA damage can be regarded as mutagenic and subsequently carcinogenic, the damage has to be carried through cell replication and survive in the daughter cells³⁶. The carcinogenic effects of alkyl adducted deoxyguanosine have a greater significance if they are persistent; this suggests that the DNA repair process is an important consideration when looking at the mutagenicity and carcinogenicity effects of an alkylating agent.

Studies on *Escherichia coli* (*E. coli*) have shown that over one hundred genes are responsible for the production of a remarkable multitude of enzymes involved in

DNA repair mechanisms³⁷. One of the main repair pathways for DNA is nucleotide excision repair (NER), which involves the recognition of large adducts that have resulted in the distortion of the helix. A length of DNA that is about 30 nucleotides long containing the damaged base is first excised. The excised area is then filled by the action of DNA polymerase I which works in concert with other proteins. The ends of the new section are sealed to the strand by DNA ligase³⁸. Smaller adducts that do not cause helical distortions are repaired by base excision repair in which the N-glycosidic bond between the damaged base and the deoxyribose is cleaved by DNA glycosylase; the nucleotide gap is recognized by DNA polymerase β and filled with assistance of other proteins, and finally ligated by the action of DNA ligase III³⁶.

Overall the simplest methods of repair in these small adducts is by the direct action of O⁶-alkylguanine-DNA alkyltransferase (ATase). ATase directly removes the adduct and thus restores the guanine back to its original state³⁹. ATase genes have been found in a great many different species including *Archaea* bacteria, this suggests that this process is of particular significance if it has survived throughout evolution, especially as it is metabolically expensive. This is because ATase is not a true enzyme as it is not recoverable after it has exerted its action on the damaged base. The action is a multi-step process that involves non-specific binding to DNA, followed by binding to alkylated guanine which is then flipped out of the helix and finally the alkyl group is transferred to a cysteine residue on ATase, possibly by the mechanisms proposed in Figure 1.7⁴⁰. The fate of the protein is a conformational change that ultimately leads to its proteolytic degradation. The binding motif is highly conserved and recognized in thirty one DNA sequences coding for ATase in many species from *Archaeobacteria* and *Eubacteria* through to eukaryotes³⁹. The crystal structure of

ATase reveals that the active site is buried deep inside the protein, which also contains a groove into which DNA could potentially fit. Therefore it is reasonable to suggest that the surrounding DNA sequence could affect the rate of alkyl transfer⁴¹.

There is evidence to show that when modified DNA is incubated with an excess of ATase, O⁶-MdG was completely repaired but the amount of O⁶-CMdG remained unchanged. This suggests that O⁶-CMdG is resistant to repair by ATase³⁵.

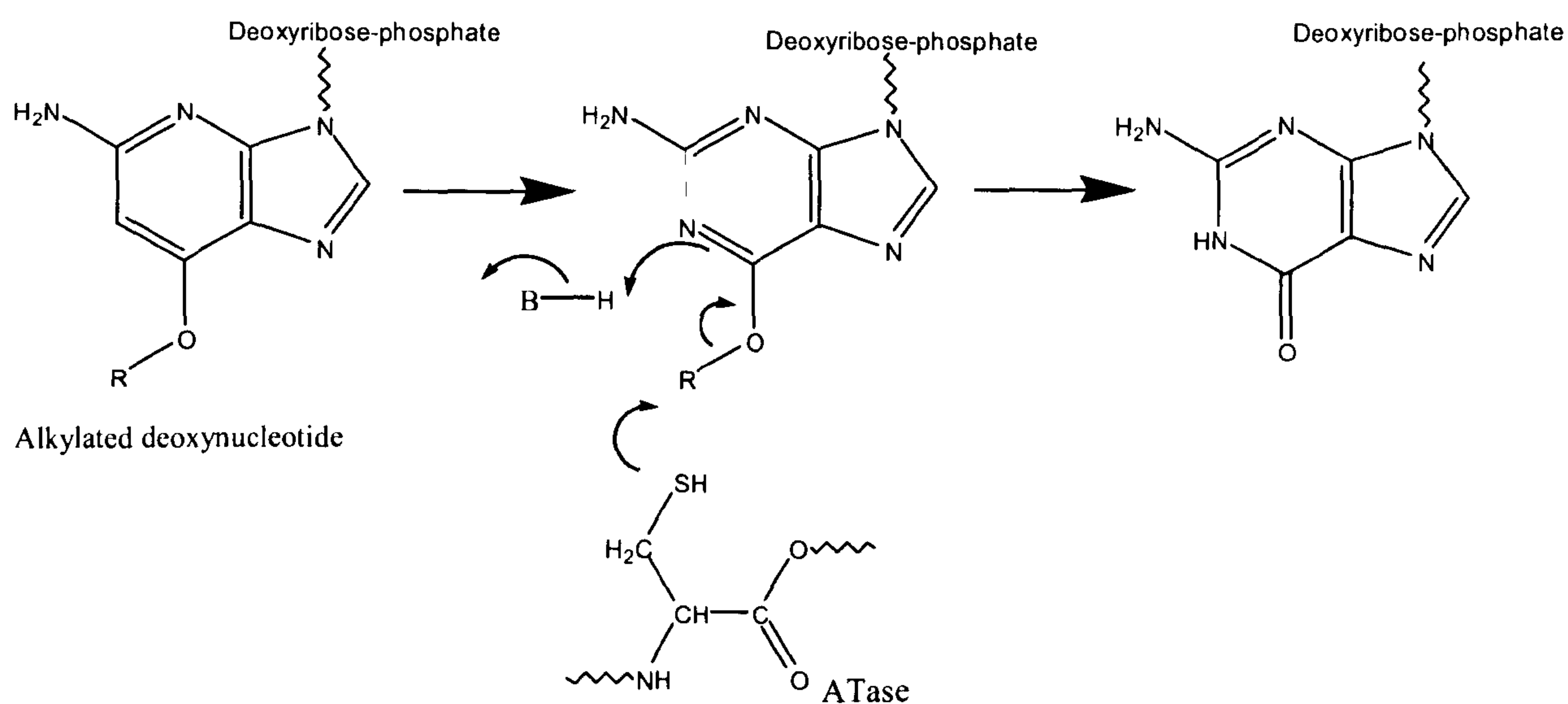


Figure 1-7. The mechanism of the reaction of ATase with alkylated guanine. The above reaction mechanism repairs many types of alkylation but reactivity decreases with increasing size of adduct, with the exception of O⁶benzylguanine⁴⁰.

1.4.1 Consequences of Adduct Persistence

DNA is a polymer that is made up of deoxyribonucleotide units, with each unit consisting of a nitrogenous base, a sugar and a phosphate group as seen in Figure 1.8.

The specific folding pattern of DNA is dominated by base pairing, as nucleic acid chains will tend to fold in such a way as to maximise base pairing. The stability of the double helix is largely due to base stacking, as the planar aromatic rings of adjacent

base pairs lay vertically one on top of the other; this maximises the van der Waals interaction between them.

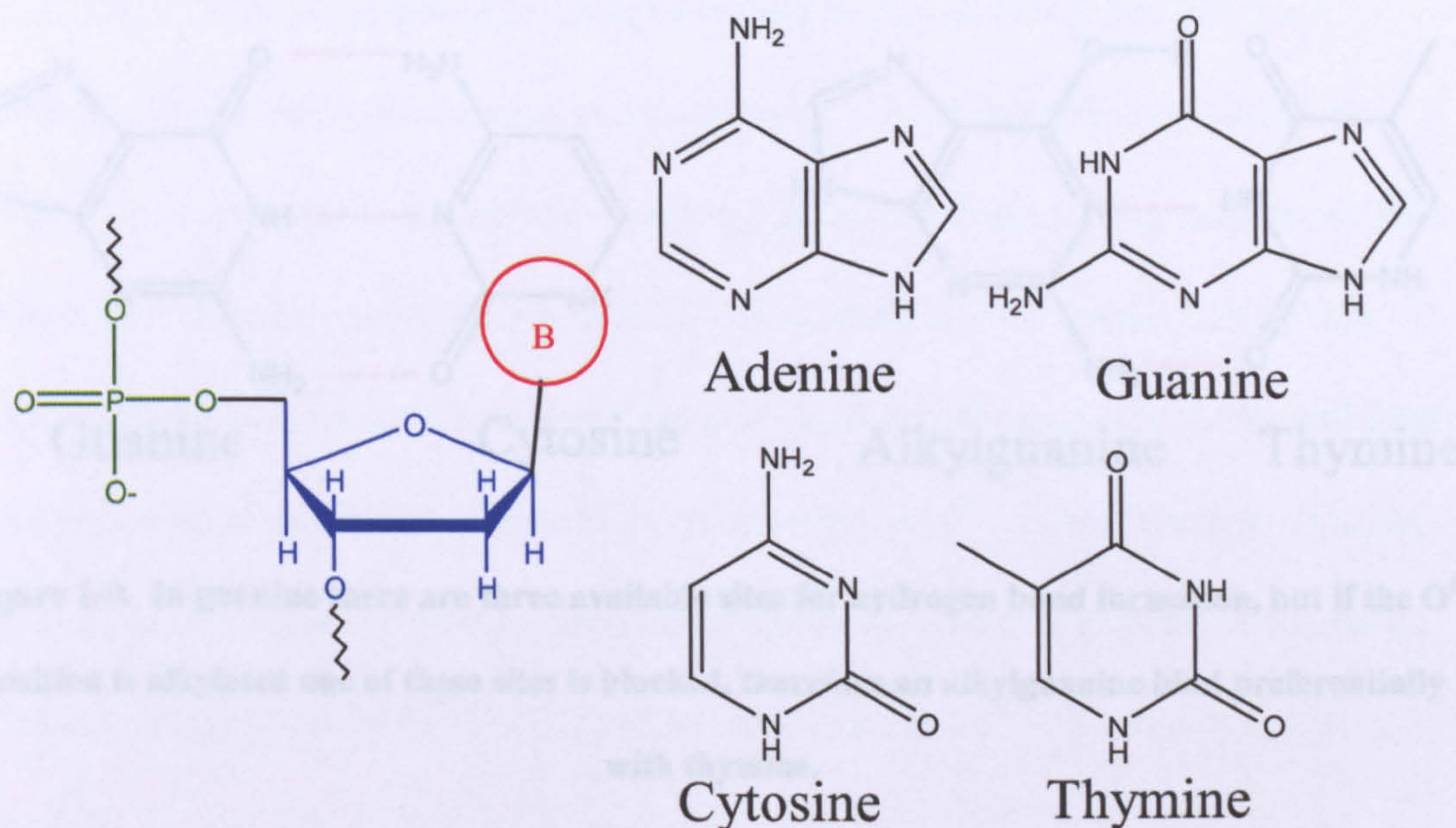


Figure 1-8. The structure of a nucleotide building block, consisting of a base (red), a sugar (blue) and phosphate group (green). The B represents one of the four available bases, adenine, guanine, thymine or cytosine. Hydrogen bonding between complementary bases dominate the folding pattern of DNA, as chains tend to fold in a way that maximises base pairing.

Guanine is usually found in its keto tautomer (Figure 1.9) and base pairs with cytosine. However, if the O⁶ position is alkylated the enol tautomer is present and this blocks a hydrogen bonding site and the base is more likely to base pair with thymine. If the cell replicates the guanine is replaced by adenine in the daughter cell, therefore ATase blocks the transition mutation of GC → AT which can occur if the adduct is not repaired³⁵. However, as ATase is not regenerated there is a limited supply of ATase at any given time in the cell, and at high levels of alkylation some alkylation escapes this repair process. As already mentioned, the carboxymethyl adduct is resistant to repair by this pathway. In addition, the biological significance of adducts depend on where they occur. In general, adducts that occur on atoms on DNA bases that are

involved in base pairing are more likely to be promutagenic; they can result in misincorporation during DNA replication.

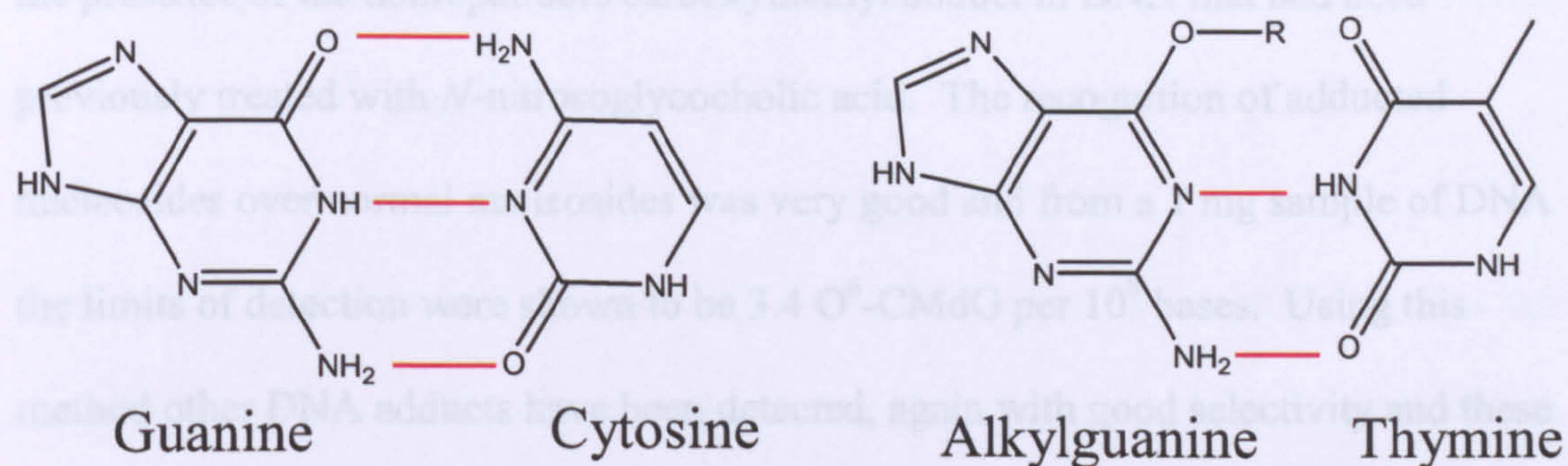


Figure 1-9. In guanine there are three available sites for hydrogen bond formation, but if the O⁶ position is alkylated one of these sites is blocked, therefore an alkylguanine bind preferentially with thymine.

1.5 Methods of detecting O⁶-Carboxymethyl deoxyguanosine

Antibodies, both monoclonal and polyclonal, have been found to be very useful for the detection of alkyldeoxynucleosides in small samples of hydrolysed DNA⁴². A polyclonal antiserum specific to O⁶-CMdG was raised by Harrison and her co-workers in 1997 and was prepared by injecting rabbits with an O⁶-CMdG-bovine serum albumin (BSA) conjugate. Thirteen weeks later blood was removed, serum isolated and the antibodies characterised⁴³. These antibodies have provided the basis of the two methods of detecting O⁶-CMdG.

1.5.1 Immunoaffinity/HPLC

An immunoaffinity/HPLC method uses the antibodies after they have been linked to a Protein A-Sepharose column. A DNA sample, containing the adduct, is hydrolysed and the resulting mixture passed through the column. The modified DNA binds within the column and is later eluted by treatment with aqueous 1M trifluoroacetic

acid (TFA). When the eluate was analysed by HPLC varying amounts of O⁶-CMdG were detectable. O⁶-CMdG was synthesised and used as a standard which confirmed the presence of the nonrepairable carboxymethyl adduct in DNA that had been previously treated with *N*-nitrosoglycocholic acid. The recognition of adducted nucleosides over normal nucleosides was very good and from a 1 mg sample of DNA the limits of detection were shown to be 3.4 O⁶-CMdG per 10⁸ bases. Using this method other DNA adducts have been detected, again with good selectivity and these used both monoclonal antibodies and polyclonal antisera⁴⁴. The advantage of using monoclonal antibodies is their continuous availability. Despite numerous efforts a monoclonal antibody for O⁶-CMdG has not been identified⁴⁵.

The amount of DNA required in this assay for low levels of adducts is more than can usually be obtained from small biopsies; this limits the use of this assay with regards to the development of a clinical useful assay³².

1.5.2 Immunoslot-blot technique

By combining the antibodies with the immunoslot-blot technique developed by Nehls *et al* (1984)⁴⁶, it has been possible to detect O⁶-CMdG in DNA extracted from small human gastric mucosa samples. The immunoslot-blot assay involves attaching adducted DNA to a nitrocellulose filter. This is exposed to antibodies that have been raised against O⁶-CMdG. Secondary antibodies, which have binding affinity with the first, are added and they bind to the primary antibody. These secondary antibodies possess a light generating system. Therefore the light intensity can indirectly be recorded as a measure of the amount of DNA and is summarized in Figure 1.10. The limits of detection were shown to be 15 O⁶-CMdG per 10⁸ bases, however a much smaller sample size is required, 1 µg of DNA²¹, less than for the immunoaffinity

HPLC method described above. This technique has also been used to measure the amount of malondialdehydeoxyguanosine, an adduct formed by malondialdehyde, a product of lipid peroxidation and prostaglandin biosynthesis, with a detection limit of 2.5 adducts per 10^8 bases⁴⁷.

Detection of background levels of the DNA adduct needs to be established so that the contribution of a particular adduct to the overall risk of cancer can be determined.

The immunoslot blot method uses as little as 1 μg of DNA, suitable for a clinical sized sample but the limits of detection are 15 adducts/ 10^8 nucleosides⁴³. The HPLC method can detect as few as 3.7 adduct/ 10^8 but a large sample size of DNA is required (1 mg)⁴⁸.

It was proposed that a new method of detecting $\text{O}^6\text{-CMdG}$ could be developed, based on a chemically synthesised affinity reagent with the binding properties of antibodies. It is suggested that a non-antibody reagent will have the ability to detect, from a small clinical DNA sample, the adduct at very low concentrations. Damage detected above a predetermined background level can be considered as an indication of the potential that disease may occur⁴⁹, in other words a quantitative measure of risk. Being a chemically synthesised reagent, with a known structure, it will be possible to make large quantities as required and also, if necessary, improve on the binding abilities by selective synthesis.

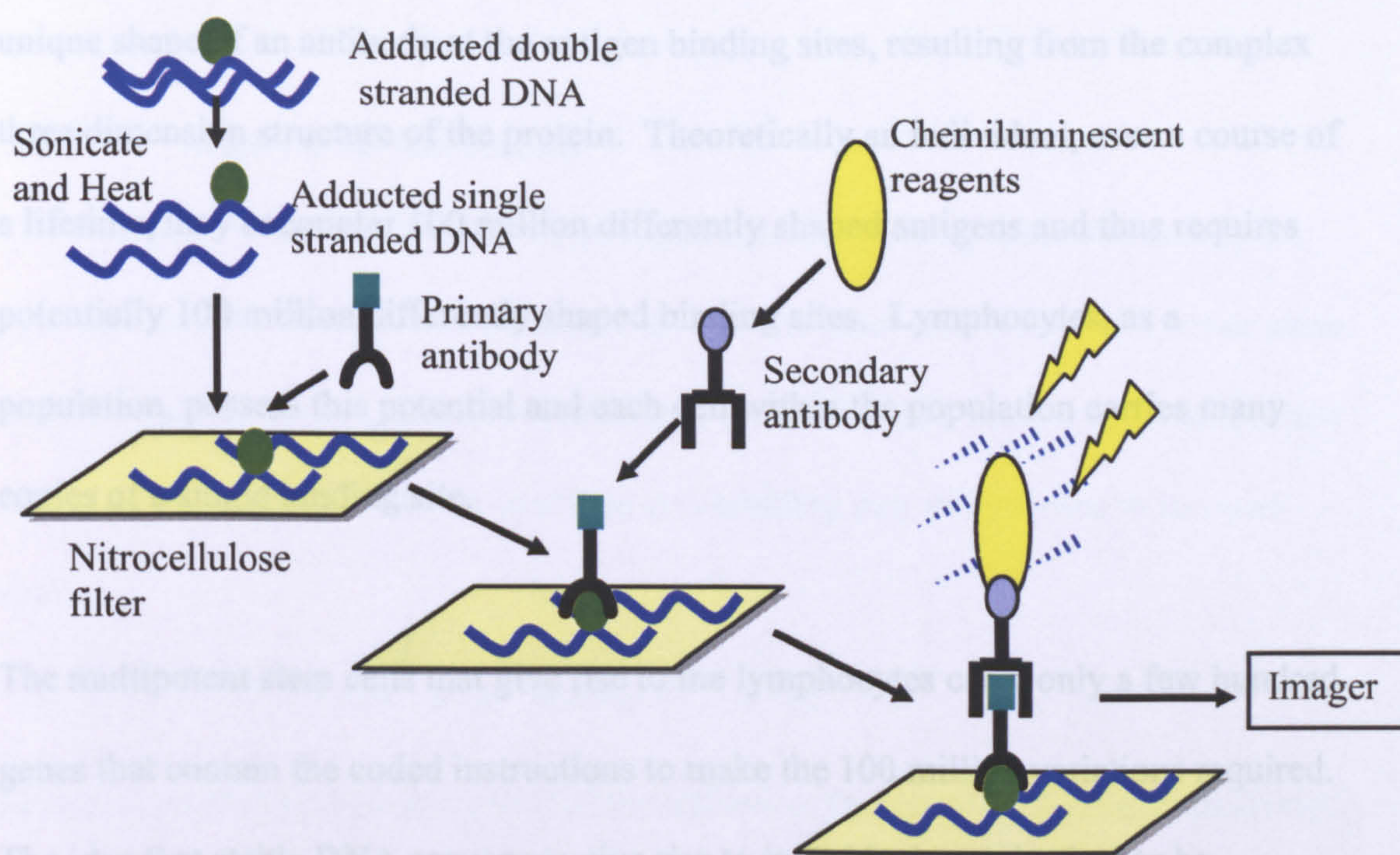


Figure 1-10. A diagrammatic representation of the immunoslot blot assay that requires a sample of 1 μg of DNA. The double strand of DNA that contains the adduct is sonicated and heated, this disrupts the binding between the strands and the single stranded DNA is fixed upon a nitrocellulose filter. The primary antibody is then exposed to the filter and binds in any location along the DNA strand that contains the adduct. A commercially available secondary antibody, horseradish peroxidase-conjugated goat anti-rabbit, is then exposed to the filter, this binds to the primary antibody and in the presence of chemiluminescent reagents a sensitive hyperfilm records the intensity of the light generated³².

1.6 Molecular Recognition

Both the detection methods mentioned above were based on the principle that an antibody has the properties that allow it to bind to a specific molecule, which is the underlying basis of the immune system.

A key component of the immune system is recognition; of the presence of an invader and the ability to discriminate between self and non-self. This is achieved by the

unique shape of an antibody at the antigen binding sites, resulting from the complex three dimension structure of the protein. Theoretically an individual, over a course of a lifetime, may encounter 100 million differently shaped antigens and thus requires potentially 100 million differently shaped binding sites. Lymphocytes, as a population, possess this potential and each cell within the population carries many copies of a single binding site.

The multipotent stem cells that give rise to the lymphocytes carry only a few hundred genes that contain the coded instructions to make the 100 million variations required. The idea that stable DNA sequences give rise to individual proteins had to be reconsidered to account for this paradox. Somatic recombination is the random allocation of a few genes, from the few hundred available, to ensure that each lymphocyte receives a unique gene sequence and a unique antigen receptor⁵⁰. This method accounts for the vast diversity required by the body to protect it against the whole host of antigens that it may encounter.

1.6.1 Antibody recognition

The major function performed by an antibody is high affinity binding to an antigen³⁷. This is achieved by hydrophobic forces, ionic forces and van der Waals forces but not covalent bonds. These intermolecular forces are individually weak and are of a transient existence but when many interactions are being used they produce highly stable structures.

1.6.2 Intermolecular forces

The strongest of these forces is the **hydrogen bond**, especially when the donor, the hydrogen and the acceptor are co-linear. In order to form such a bond the donor atom must be electronegative, therefore polarising, the D-H bond. The acceptor atom must also be electronegative, with at least one non-bonding pair of electrons in its outer shell.



Ionic interactions occur as a result of the attraction that occurs between a positively charged cation and a negatively charged anion, these interactions do not have specific or fixed geometry because the electrostatic field surrounding the ions is uniform in all directions.

Van der Waals interactions arise from the transient dipoles, momentary random fluctuations in the distribution of the electrons in the atoms. When two non bonded atoms come together, the transient dipole of one atom perturbs the electron cloud of the other, causing a second transient dipole, these two dipoles then weakly attract each other. Stacking interaction can occur between aromatic amino acids and nucleic bases⁵¹.

Nonpolar molecules do not possess ions and do not become hydrated. Like molecules aggregate together not because they have an attraction for each other but because water has a high affinity for itself.

It is the combined affects of these interactions that bring about stability when two molecules come together as illustrated in Figure 1.11, a hypothetical representation of two proteins that have an affinity for each other.

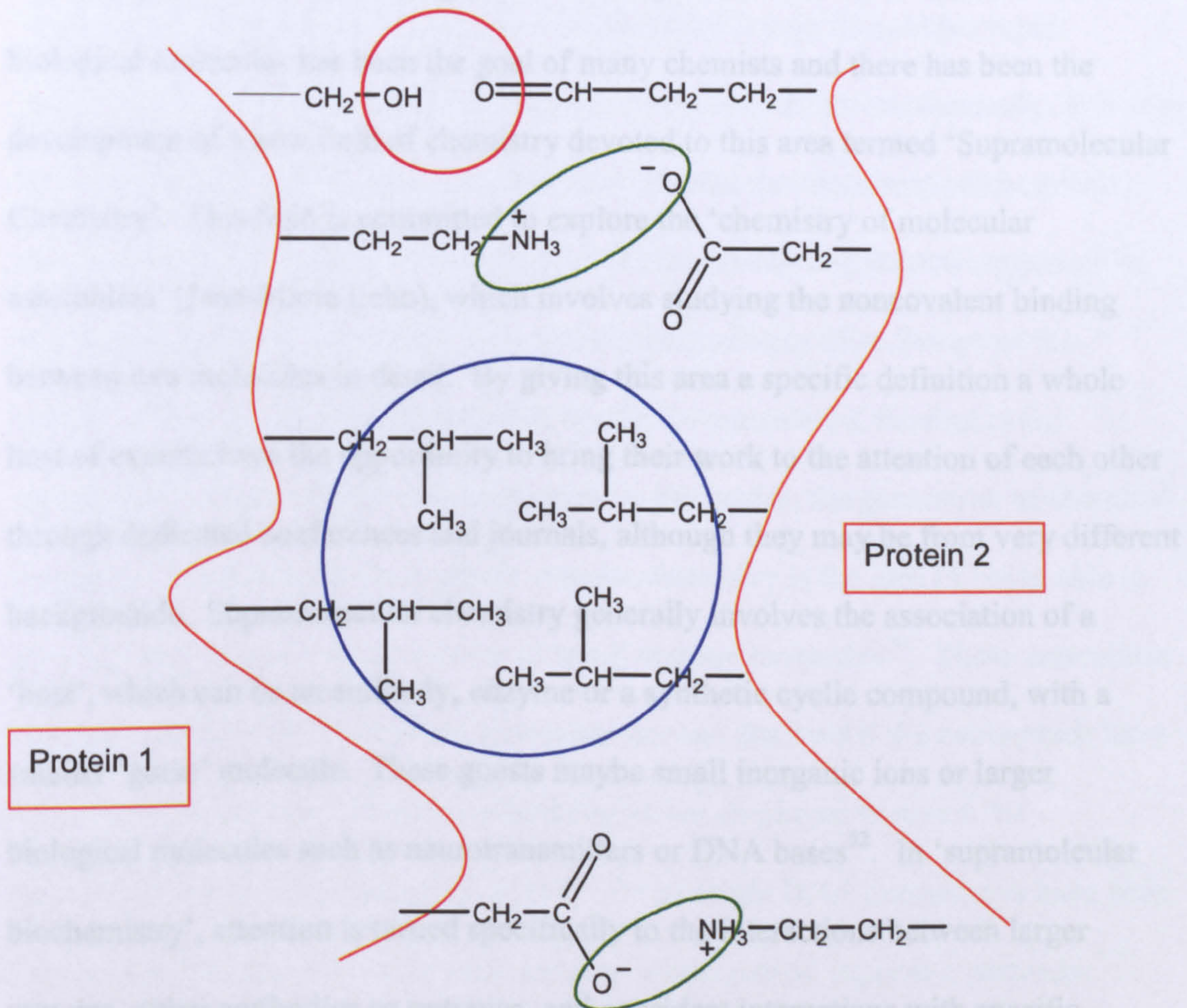
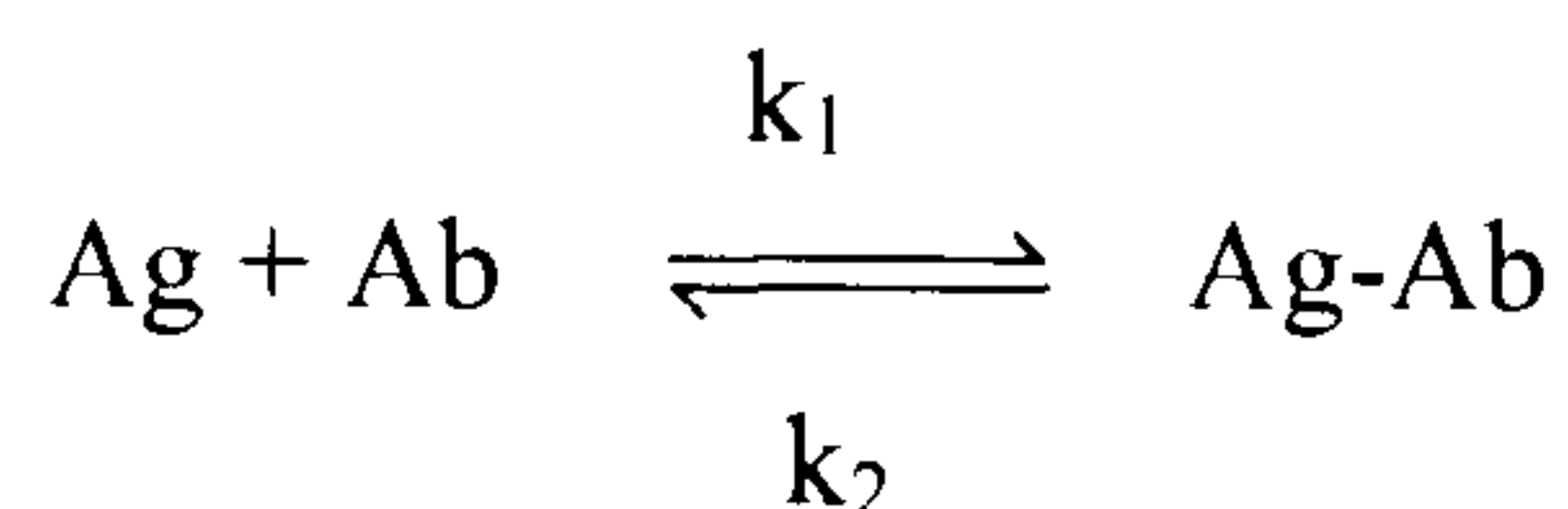


Figure 1-11. The hypothetical binding of two proteins by two ionic bonds (green), one hydrogen bond (red) and one combination of hydrophobic and van der Waals interactions (blue). The complementarity of the two surfaces and the overall force of the interactions give rise to the specificity of binding between any two given molecules.

The binding of an antibody to its antigen is a simple bimolecular reversible reaction:



The ability to mimic this binding and to make synthetic molecules that interact with biological molecules has been the goal of many chemists and there has been the development of a new field of chemistry devoted to this area termed ‘Supramolecular Chemistry’. This field is committed to explore the ‘chemistry of molecular assemblies’ (Jean-Marie Lehn), which involves studying the noncovalent binding between two molecules in detail. By giving this area a specific definition a whole host of experts have the opportunity to bring their work to the attention of each other through dedicated conferences and journals, although they may be from very different backgrounds. Supramolecular chemistry generally involves the association of a ‘host’, which can be an antibody, enzyme or a synthetic cyclic compound, with a smaller ‘guest’ molecule. These guests maybe small inorganic ions or larger biological molecules such as neurotransmitters or DNA bases⁵². In ‘supramolecular biochemistry’, attention is turned specifically to the interactions between larger proteins, either antibodies or enzymes, and considers interactions with specific ‘guests’ or ligand molecules. One consideration is that both types of ‘hosts’ are derived from very large families of molecules, each with slightly differing properties that allow for unique specificity.

1.7 Chemical Genetics

Chemical genetics is a term that was first used in the inaugural issue of ‘Chemistry and Biology’ (1994) by Stuart L. Schreiber⁵³. Chemical genetics has also been called chemical genomics, chemogenomics or chemical biology; these terms are used to describe the study of the effects that an exogenous ligand has on proteins (gene-

products) in a cellular environment. Geneticists use gene mutations to perturb cellular function, whereas the chemist can use low molecular weight organic molecules that are capable of altering protein function; these alterations are reversible and conditional⁵⁴. These experiments are usually carried out on flies with a rapid reproduction rate it is harder to use these types of experiments on mammals. It is with the development of new techniques that have allowed the usefulness of chemical genetics to be explored more fully and allowed there to be a systematic approach to screening compounds for biological activity immediately after dosage of the ligand⁵⁵. There are three main steps that need to be considered, the first is the assembly of a large collection of small organic molecules, the generation of chemical diversity. Unprecedented in synthetic organic chemistry is the idea of being able to generate large searchable assemblies of small organic molecules⁵⁶. These assemblies can then undergo high throughput screening that can discover if the compounds have any biological function. Finally methodologies are employed to reveal the mechanistic basis of the biological activity, for example DNA microarrays have been used, and in one case the entire yeast genome was screened against a particular compound and it was found to induce the expression of five genes, two of unknown function⁵³.

Small molecules are usually designed to a specific shape in order to carry out a specific purpose on a gene product and as such have been made by retrosynthetic analysis in which a 'compound undergoes many rounds of individualized hand-crafted modifications'⁵⁶. At each stage there needs to be careful control of stereoselectivity and regioselectivity; the desired diastereoisomer needs to be produced or purified with the exclusion of all other products. This paradigm has served the organic chemist well

and led to the widespread availability of collections of highly diverse small molecules. However the need to create larger collections of compounds has allowed the chemist to employ a diverse set of blocking blocks, working to generic schemes that yield a multitude of related products. Chemical genetics that require such diversity have shown small molecules, however produced, can be used to identify previously unknown gene products⁵³ and modulate the action of other gene products⁵⁴. More recently chemical genetics have been used to find molecules that also bind directly to DNA⁵⁷, RNA⁵⁸ and even the directly inhibit specific gene expression⁵⁹. The limits for medicinal purposes appear to be endless and the ultimate aim of the chemical geneticist is that it will be possible to find a small molecular partner for every gene product⁵⁷.

1.8 Chemical diversity

There are many millions of small organic molecules that exist, these have been collated and stored in research institutions and pharmaceutical companies. The origin of these compounds range from tens of thousands that are derived from nature, e.g. from plants, bacteria, fungi and marine sponges⁵⁵, compounds that have been made by synthetic chemists via the traditional routes over many years, and also by the combinatorial chemist, whose aim it is to create populations of molecular structures. The goal of such chemists is to continually strive to make more libraries, with greater populations and this area of research is still ongoing with the challenge to do so in a way that is high yielding and reliable. There are many ways in which this has been already been achieved, and combinatorial library synthesis, in its various forms, is now considered to be the major routes in providing diversity⁶⁰. The identification of biologically active parent molecules has given the opportunity for libraries to be built, using both solid- and solution-phase synthesis, around different scaffolds, allowing

natural product mimicry to be established⁶¹. Whilst many libraries still employ solid phase synthesis the percentage of solution phase libraries had increased to 30% by 1998⁶⁰. Terpenoid and steroid based libraries are better suited to solution phase synthesis⁶¹.

1.8.1 Peptide synthesis

The most diverse library of compounds must be those that are used by Nature in the assembly of the proteins that are utilized in the biochemistry of the body. Since the chemical methods for assembling individual peptides was developed over thirty years ago improvements have been made in both the efficiency and the minimization of deleterious side products⁶². There are now three recognised methods of peptide synthesis:-

Stepwise Synthesis – the attachment of a single amino acid at a time to build up a linear strand. The most widely used is the Solid Phase Peptide Synthesis (SPPS) developed by Merrifield.

Fragment Assembly – separate strands are built and then linked together by either convergent synthesis of fully protected fragments, or by chemoselective ligation of unprotected fragments.

Directed Assembly – individual peptide strands are allowed to associate non-covalently⁶²

A bicyclic decapeptide has been used as a scaffold to introduce tetrapodal libraries in solution phase combinatorial synthesis, this involved the attachment of four different protected sites. Following a split synthesis protocol, it proved to be a relative easy way to produce a library⁶³.

A small library, 39 members of pentapeptides, were screened and a peptide was found that mimicked kinase activity by selecting a specific area of a molecule for phosphorylation⁶⁴. Oxime bridges have been used in numerous ways to attach a variety of substances to proteins including in vivo applications. Peptide fragments with N-terminal aminoxyacetyl groups have been attached to a poly aldehyde scaffold via oxime bridges resulting in the formation of homogeneous branched polypeptides with a molecular weight of almost 20 000 D, the first protein to be made of this size⁶⁵. Synthetic receptors were isolated from a combinatorial library that had been synthesized by solid phase split synthesis, in which specific individual members were encoded by the attachment of chemical markers or tags⁶⁶.

The peptidic synthesis of a targeted sequence depends on the folding patterns within the peptide chain and understanding this protein folding is by no means trivial. The complex, and not wholly understood, folding pattern of a linear strand of amino acids into a three dimensional architecture which is ultimately responsible for a particular protein, is the main challenge for researchers making non native but biological potent proteins⁶⁷. Computer assisted drug design programmes such as DESIGN developed by Wilson (1993) has aided the understanding of the spatial requirements of a protein mimic⁶⁸. However, this problem can be somewhat by-passed by attaching the biologically active component on to a pre-designed backbone template or scaffold molecule. By the late 1990's with the improvements in methodologies there was an extension of solid phase peptide synthesis which allowed for the development of efficient access to template based protein mimics of high structural diversity⁶⁹.

1.8.2 Template-assembled synthetic protein (TASP)

A peptidomimetic can be made via a template-assembled synthetic protein (TASP) method, a two-step process. The first is the chemical synthesis of a topological template, and the second the chemoselective ligation of amino acid side chain derivatives or functional groups via covalent bond formation⁷⁰, for example, attaching peptide binding loops with terminal aminoxy groups to aromatic templates via oximes⁷¹. A summary of some of the possible alternatives is given in Figure 1.12.

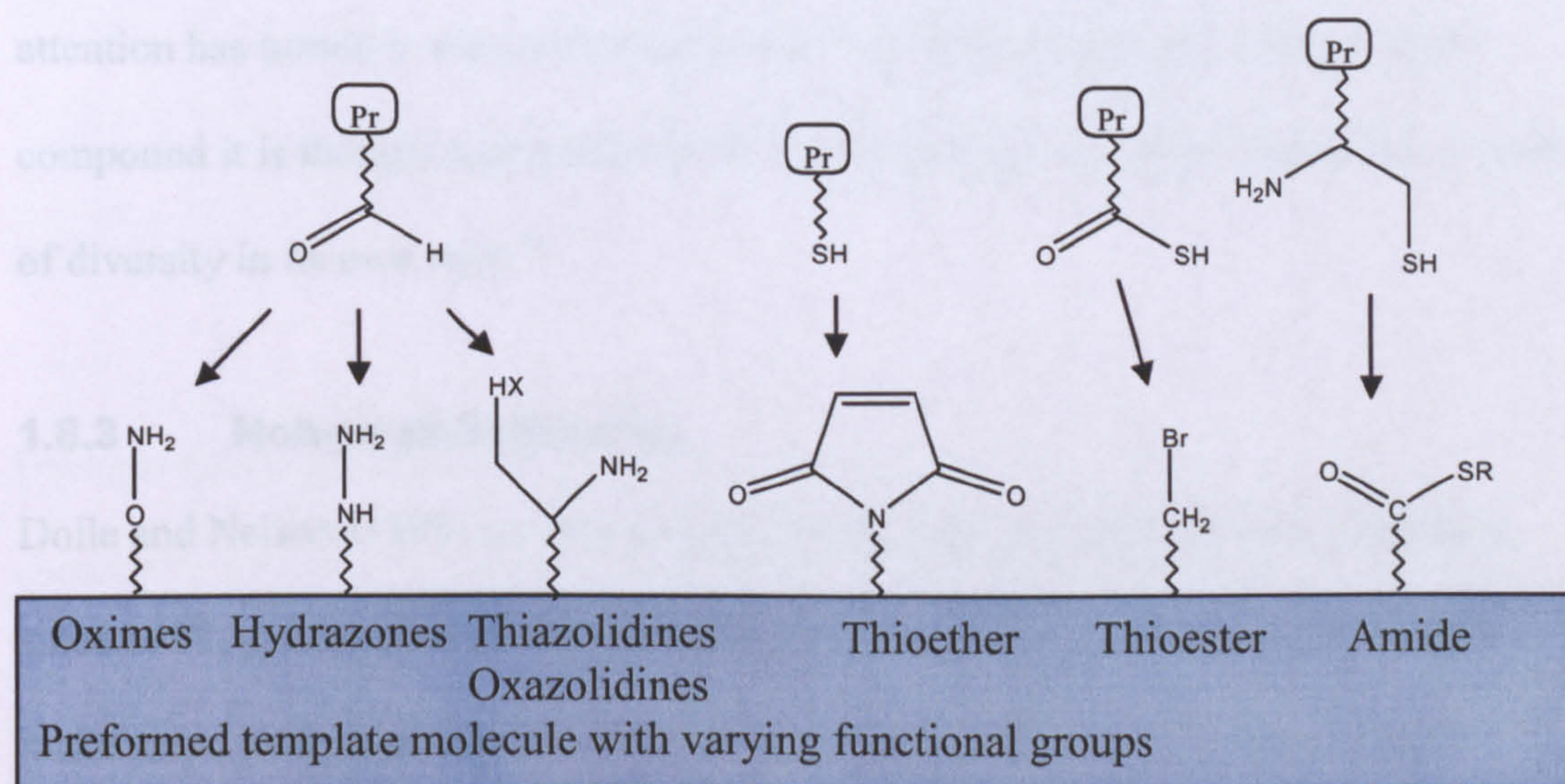


Figure 1-12. A summary of the ligation methods that have been employed using the TASP method of producing protein mimetics, following condensation of unprotected peptide fragments (Pr) in aqueous solutions^{62,67}.

In TASP the overall peptidic molecule is not a folded linear protein but an array of branched peptides. However both stepwise solid and protected segment condensation up to the early 1990's proved to be arduous and synthetically challenging. However the work of Dawson and Kent successfully used the 4-Helix TASP molecule designed by Mutter using thioester nucleophilic ligation between the bromoacetyl

functionalities on the template and COSH moiety on the peptide fragment C terminus⁷².

In addition to the design of single molecules TASP also provides a route on which a combinatorial library can be based and these templates were first used to make constrained peptides, cyclodextrins and polycyclic systems⁷³.

Chemists are continually striving to increase the diversity within their libraries, and attention has turned to the scaffold molecule. As this forms a part of the overall compound it is thought that instead of an inert template it can also be used as a source of diversity in its own right⁷⁴.

1.8.3 Non-peptide libraries

Dolle and Nelson (1999) reviewed the 321 new libraries that have been published since 1997, giving information on the size of the library and the reactions used to make them, as well as biological activity, if known. The vast majority of these libraries are nonpeptidic, some are acyclic, but most range from macrocyclic to monocyclic; there are also scaffold based libraries⁶⁰. This review highlighted the area of combinatorial chemistry that is being researched by a great number of people from many areas, it also emphasises the way that the conventional boundaries between chemistry and biology are being broken down.

The main problem with nonpeptidic libraries is the lack of stereocentres, which is in direct contrast to the complexity that is found in nature⁵⁵. However, when a nonpeptidic library of 126 members, based on a D-xylose scaffold with ester linked function groups, was screened it revealed a compound that mimicked the binding

properties of the Arg-Gly-Asp sequence and displayed biological activity. Many other saccharide based libraries are now being investigated and this has led to the discovery of new carbohydrate based antibodies⁷⁵. Carbohydrates readily lend themselves to combinatorial chemistry as a scaffold molecule, a conformationally rigid structure, with available reactive groups that can be readily functionalized, their versatility is described by Hirschmann in 1993 who acknowledged them as 'privileged platforms'. Kemp's acid has three axial carboxyl groups, which has also been used to build diversity upon resulting in a library with, theoretically, 75,000 members⁷⁶. A whole host of molecules have been made based on the idea that synthetic carbohydrate receptors can be used as anti-infectives or synthetic antibodies⁷⁷.

Peptoids (oligo-N-substituted glycines) are suited to the construction of large chemically diverse libraries and have been used to make libraries with, theoretically up to, 531 441 members⁷⁸.

1.8.4 Biological based libraries

Aptamers are specific RNA or DNA oligonucleotides that have been selected from by a 'systematic evolution of ligands by exponential enrichment' (SELEX) process. The selection process starts with single stranded DNA oligonucleotides with a central region containing randomised sequences. This undergoes a series of selection, with amplification of an anticipated binding aptamer⁷⁹. Genetically encoded peptide libraries can also be displayed as a library of unconstrained peptides on the coat of bacteriophage in phage display⁸⁰, or as constrained peptides libraries in yeast that can be screened *in vivo*⁸¹. Woiwode and his colleagues have developed a way of

hybridizing phage display and combinatorial chemistry by using hydrazone attachment to link organic molecules to a primary amine on the phage protein coat⁸².

1.8.5 Steroid based libraries

Steroids have been used in a variety of ways and in 1993 they were recognized as structural components in molecular engineering⁸³ and in 1997 a review of reactions and partial synthesis covered the extensive work that had been done on steroids⁸⁴.

More recently work has covered molecular recognition⁸⁵, cage formation⁸⁶ and as a base for protonated aminosteroids that act as anion receptors^{87,88}. In addition, Gemini surfactants which contain a polycationic head based on cholic acid, have turned out to be effective anti-bacterial agents of both Gram-positive and Gram-negative bacteria⁸⁹.

A number of crystal structures of cholic acid have been described⁹⁰ and the host-guest interaction between cholic acid and acetonitrile has been examined in detail⁹¹.

Chenodeoxycholic acid was first used by Wess and co-workers to link short amino acid chains, two-five in length, to the C3 and C5 positions, via ethers⁹², to construct mini libraries. Steroidal cores (*AB-cis* and *AB-trans*) have been used to attach peptidic appendages^{93 and 94 respectively}. Cholic acid libraries will be discussed in further detail in section 1.10.

1.8.6 Dynamic combinatorial libraries

Dynamic combinatorial chemistry elegantly combines combinatorial chemistry with molecular evolution, as the library of compounds is generated by the assembly of building blocks via reversible bonds. Therefore the product distribution within the library is under thermodynamic control and influenced by the presence of a ligand.

The molecules that bind to the ligand become amplified in concentration. A receptor for *N*-methyl alkylammonium salts was isolated from such a library⁹⁵.

1.9 Directed Molecular Imprinting

In addition to the many ways of making libraries of compounds and then screening such libraries, based on dynamic combinatorial chemistry, there was proposed a way of manufacturing a molecule that has the binding properties built in during the reaction mechanism. The required molecule is a result of the ligand being present during synthesis allowing the molecule to be built around it.

1.9.1 Molecular Imprinting Polymers

The idea that a host builds itself around a ligand was first proposed by Linus Pauling in 1940. He suggested that a specific antibody is produced as a response to being exposed to a particular ligand. He thought that an unfolded polypeptide interacted with a ligand and it was the presence of the ligand that allowed the polypeptide to be moulded into the shape that is required to bind to the ligand⁹⁶. Although his hypothesis was subsequently proved to be wrong, the idea of a polymer being able to be imprinted with a particular shape by the presence of a ligand has led to the development of molecularly imprinted polymers (MIPs).

MIPs are formed when the monomers of the polymer are assembled in the presence of a ligand which acts as a template molecule. The monomers form a complex with the template and following polymerization form a highly cross linked polymer. The template is removed, leaving a cavity within the polymer that has the complementary shape and functional groups that facilitate selectivity and affinity. By the nature of their synthesis these polymers are robust and very stable; they can be used in extreme

conditions, including organic solvents and in basic or acidic conditions⁹⁷. Due to the mode of action of MIP they have been called 'antibody mimics'.

MIPs have been used in the generation of biomimetic sensors, in which the key element of efficacy is specific recognition. With a few exceptions they remain inferior to biological based sensors as the MIP tend to have heterogeneous binding sites on the resulting polymer⁹⁷. These form as a consequence of the need to push the equilibrium of the reaction towards the desired product by having an excess of monomers, as such many sites are formed in the absence of the template and these sites have much lower affinity and selectivity for the required ligand⁹⁸. This heterogeneity is not insignificant with regards to developing this technique and its removal remains high on the 'wish list' of researchers. Other problems that have not been fully explored include the complete removal of the template following MIP assembly and the ability to spectroscopically characterise binding sites.

Although theoretically MIPs have come along way since the early days, and many research groups are working on developing these ideas further, this technique was not considered further in relation to this project.

1.10 Steroids as scaffold molecules for diversity-oriented synthesis

1.10.1 An Introduction to Steroids and Bile Acids

The basic building block of steroids is acetyl CoA, which builds into a C₃₀ structure, squalene containing six isoprene units. It is from this that cholesterol is derived³⁸ or smaller amounts are ingested⁸⁵. First isolated from gallstones in 1784, it has been the

basis of the work of thirteen Nobel Prize winners³⁸. Cholesterol provides us with vitamin D and its derivatives include lipophilic hormones that are able to cross the cell membrane to interact with receptors inside the cell³⁷, and bile acids that form micelles with water-insoluble fats allowing their uptake into the bloodstream^{85,86}. All these molecules have the same basic chemical skeleton (see Figure 1.13), in which the carbon atoms are numbered around the rings, and each ring is lettered. The ring fusion between A and B is either *cis* or *trans*, all steroid hormones that have a hydrogen atom at C5, have *trans* fusion, whereas all bile acids have *cis* fusion³⁸. Converting cholesterol into bile acids, a fifteen step enzymatic process, effectively removes cholesterol from the body. This is achieved by either the classical (neutral) pathway which involved conjugation to either glycine or taurine or by the alternative (acidic) pathway that results in the production of chenodeoxycholic and cholic acids⁹⁹.

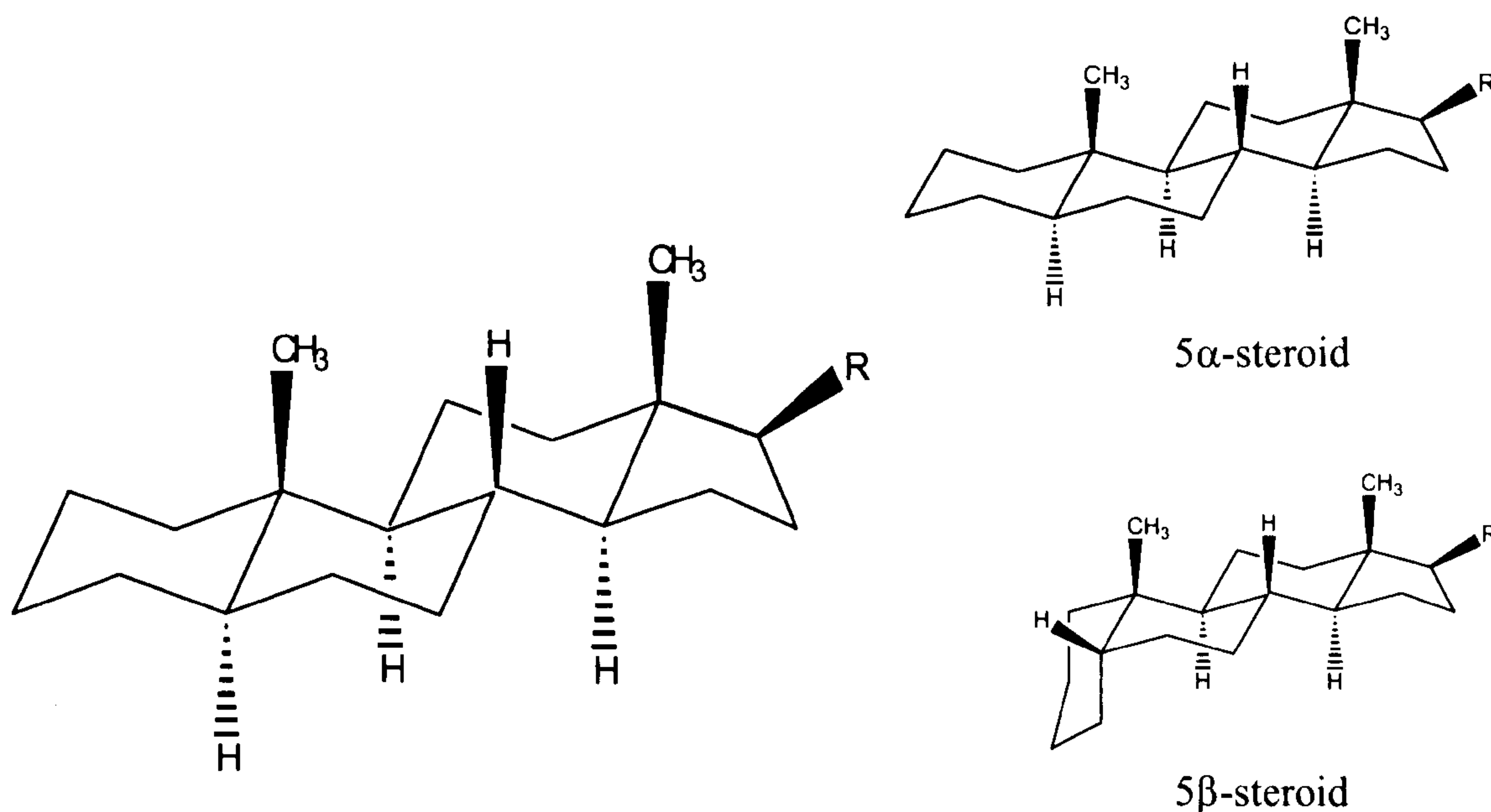


Figure 1-13. The basic structure and nomenclature of steroids, the rings are lettered from left to right, and the numbering starts at the top of ring A. B and C, as well as C and D ring junctions are always in the *trans* configuration, but the A and B junction is possible as either *cis* or *trans*.

1.10.2 The Uses of Cholic Acid and its Derivatives

The unique properties of cholic acid are due to the presence of a hydrophobic face on one side of the molecule and a hydrophilic face on the other (Figure 1.14). Due to this cholic acid has the ability to complex with aromatic moieties to form bilayered structures made up of repeating hydrophilic and hydrophobic layers, with channels. These crystal lattices are held together via hydrogen bonds, van der Waals and electrostatic interactions⁹⁰. The hydrophobic face of cholic acid has also been used in the recognition of polar molecules in non-polar solvents acting via the formation of hydrogen bonds¹⁰⁰. Cholic acid and its derivatives have been used extensively - a review by Tamminen in 2001 highlights the many ways in which they have been used as building blocks for supramolecular hosts including the development as hosts for acyclic appendages and combinatorial chemistry¹⁰¹. Primary amines and/or guanidine groups linked to a methyl cholate scaffold have been used to synthesise a potent sensitiser of Gram negative bacteria¹⁰² and has also been used in the chemical ligation of two proteins via aminoxy acetyl and ketone functionalities on the protein N-terminals to produce a biochemically active heterodimer¹⁰³. The oxime linkage has also been used in the construction of a library based on 5 α ,14 α -androstandane scaffold from which Na⁺,K⁺-ATPase inhibitors have been found¹⁰⁴.

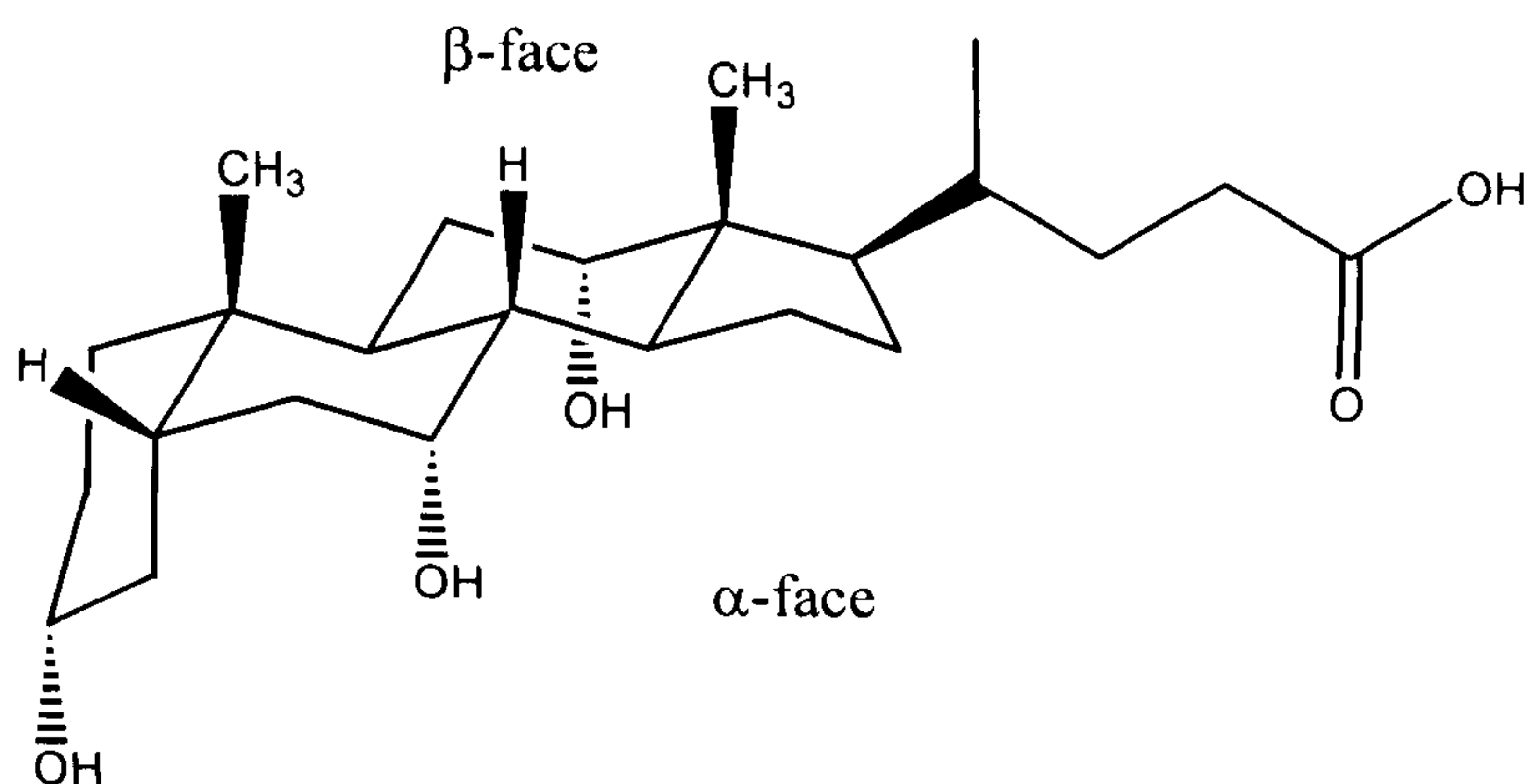


Figure 1-14. Structure of cholic acid showing the hydrophilic α -face and the hydrophobic β -face.

The DNA binding properties of seven new amino and/or guanidine substituted deoxycholic or lithocholic acid were explored, following the conversion of the α -hydroxyls to α -amino groups. Hsieh and his co-workers (1995)¹⁰⁵ tried to follow a scheme proposed by Davis and Orchard (1992)¹⁰⁶, which was successful for conversion at the C12 position on the β face. This proved unsuccessful for conversion at both the C3 and C12 positions. Therefore they employed a slightly different method which included reduction of oximes¹⁰⁵. Davis improved the C3 conversion in 1997 by using Mitsunobu reactions with methanesulphonic acid¹⁰⁷, and by 1998, with Broderick and Williams, Davis published the first synthesis of a tris-deoxa-tris-aza analogue of methyl cholate¹⁰⁸. A fifteen step synthesis, giving an overall yield of <8%, did allow for the possibility of selectivity at each of the hydroxyl sites. This technique was explored further by the same group in 1999, with the publication of a trifunctional methyl cholate, which was later used as a scaffold for combinatorial chemistry¹⁰⁹ as each of the three hydroxyl sites were protected with groups that are used in standard orthogonal peptide chemistry. Li *et al* suggested a much shorter synthesis of only four steps, but this route, via a trioxime compound,

required more purification steps and resulted in an overall yield of <8% of the desired isomer¹¹⁰.

As the potential of cholic acid became more apparent the idea of selecting one of the available sites for modification came to the attention of other groups; the selective one-pot esterification of the 3 α hydroxyl of cholic and deoxycholic acid was published in 1996¹¹¹. Cholic acid can also undergo hydroxyl conversion to keto groups via indirect electrochemical oxidation with PbO₂, platinum and carbon; oxidation occurs at C12 or C7, depending on conditions, followed by C7 or C12, again dependant on conditions, and then finally at C3¹¹². Microbial oxidation has also been explored. The selectivity of bacteria was such that two bacteria were found to oxidise C7, one for C7 and C3 positions, three for C7 and C12 and finally two bacteria that oxidised all three positions¹¹³.

1.11 The Suitability of Cholic Acid as a Scaffold

There are key issues that need to be addressed, before a synthetic receptor, with antibody-like binding properties, can be made. Effective receptors generally have concave binding sites that are complementary in size and electrical charge distribution to the ligand molecule; with as many opportunities for hydrogen bonds to form as possible.

Cholic acid has been proven to be a suitable scaffold for the assembly of combinatorial chemistry¹⁰⁸, it has three evenly spaced hydroxyl groups that have co-directed functionality ready for the presentation of appendages. It also has a side chain carboxylic group available for derivatization that allows for the attachment to a solid phase¹⁰⁹ or a fluorescent marker molecule, that is remote from the binding area. This

is an important consideration as the coloured or fluorescent tag should not alter the receptor-substrate binding¹¹⁴. The steroid nucleus is asymmetrical and the groups are not equivalent, which allows for selectivity. Thus, when appendages are ligated to the steroid a readily available pocket is formed, due to the structure of the scaffold¹⁰⁰. Finally cholic acid is readily available at a minimal cost of £9.90 for 25 g (Aldrich 2004).

The positive attributes of cholic acid have already been explored with regard of the construction of combinatorial libraries. Li *et al* used cholic acid as the scaffold for a range of mini libraries in which short peptide chains were linked to the support via ethers¹⁰². This is the approach was used previously by Still and his colleagues, who synthesised a library of short peptide chains attached also by ethers on two of the three hydroxyl sites. The synthesis of the tri amino analogue of methyl cholate really opened the way for the development of combinatorial libraries. The three sites were shown to be available for functionalisation utilizing the more stable amide bond, either all at one time¹⁰⁸ or selectively¹⁰⁹.

1.12 Aims of this thesis

A key step in the etiology of cancer is the alkylation of DNA. A major mutagenic lesion is the alkylation at the O⁶ position of 2'-deoxyguanosine. This is particularly significant when it involves the carboxymethyl, as this lesion appears to be resistant to repair by the usual pathway³⁵. Such adducts can be brought about by the action of N-alkyl-N-nitroso derivatives of glycine-containing peptides or related compounds and these modifications can be effectively recognised by antibodies. However, the antibodies available have been produced as rabbit sera⁴³ and a finite amount is available for use.

The first part of the first part

Antibodies are large complex macromolecules, of which only a small part is necessary for the recognition and binding to a particular ligand. In principle a simpler peptide based reagent could be isolated from a library of compounds - a library that has been made in such a way as to maximize the possibility of a candidate compound being present. Whilst constructing such a library a number of considerations have to be made, for example, which method will be used to make the library? Should a scaffold molecule be used as a base, if so, which is the most suitable? A 'tag' that allows for easy detection also needs to be introduced – what is the best way to do this? Is it possible to add enough diversity? Finally, can a candidate molecule be isolated from the library, identified and synthesised in sufficient quantities to be useful?

Finally the answer to the first part

Ideally, a library of compounds could be built upon a stable backbone, a scaffold molecule that allows the introduction of diversity in equal amounts at each of the specified sites, and to have a functional group remote from this diversity that is available for the attachment of a molecule that allows for easy analysis (Figure 1.15).

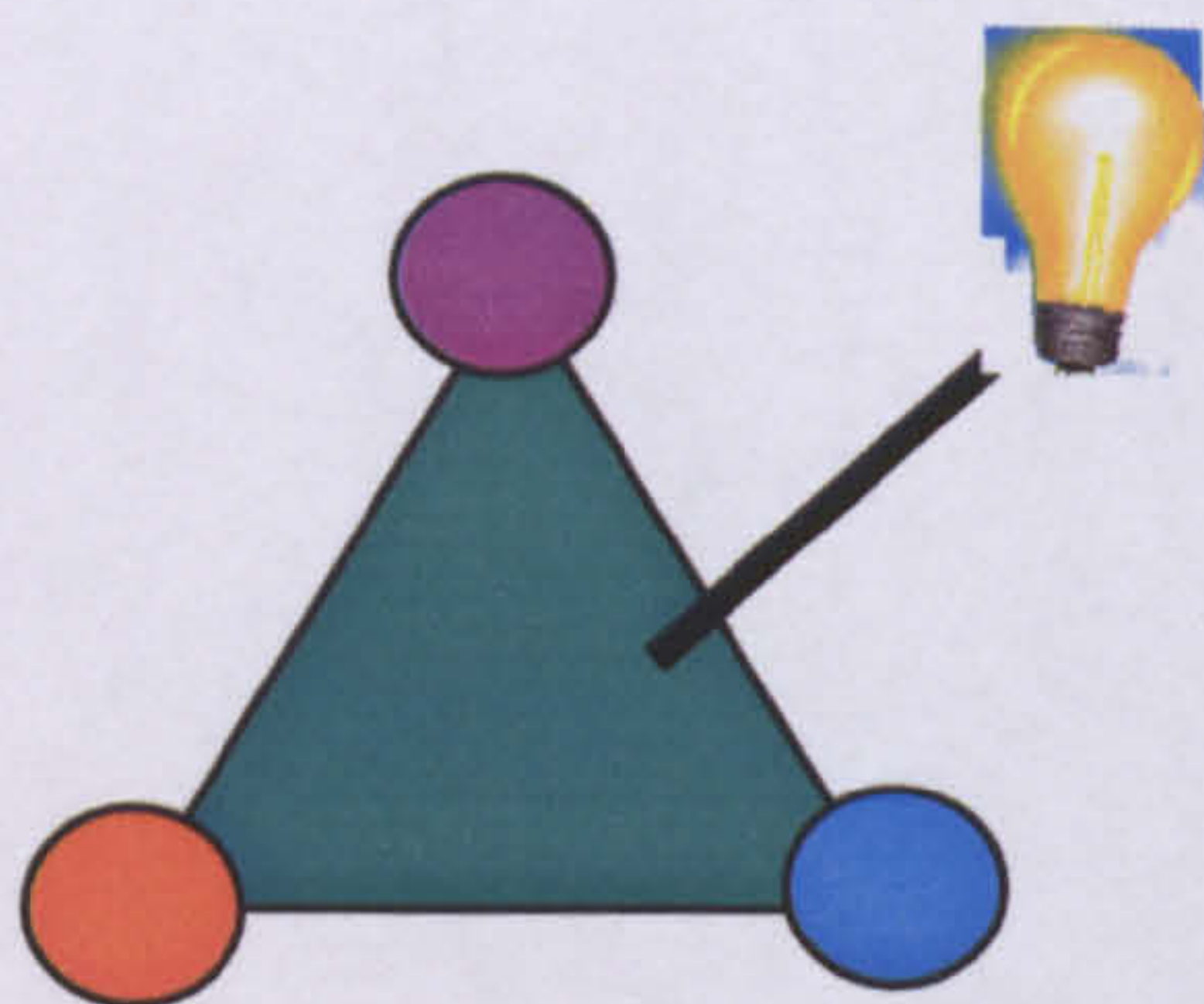


Figure 1-15. A cartoon of the ideal affinity reagent, a rigid backbone on which diversity can be built, with a site remote from the binding area, on which a tag can be positioned to aid identification.

The first part of this thesis is aimed at establishing which amino acids could be present in the antigen binding site of the antibodies known to recognize O⁶CMdG, those that may contribute to selection and affinity, and so enhance the binding potential of a novel reagent. Phage display was used for this purpose and an explanation given as to why the method does not result in a direct method for making a reagent.

The second part is to establish a method by which diversity can be introduced to a suitably tagged scaffold molecule by a method that is both reliable and repeatable. Preparation of the suitable building blocks was then carried out and these building blocks were used to construct a diverse library. The diversity in the resulting mixture was then investigated.

Finally the exploration of the library using affinity columns, on which O⁶CMG has been immobilised to the agarose matrix, was used to see if a suitable candidate molecule can be extracted and identified.

Chapter 2

Phage Display

‘Design must have had a designer’ – William Paley (1743-1805)

2.1 Introduction

Antibodies are large molecules that are, as far as chemists are concerned, over-engineered. They contain many hundreds of amino acids, most of which contribute to the three dimensional structure of the protein. Many of the functions of antibody molecules are related to their role in the immune system. The site of interest with regards to designing a novel affinity reagent is the site of binding and being able to discern which amino acids directly contribute to selectivity and affinity to the carboxymethyl DNA adduct. A technique that can be used to identify these amino acids is phage display.

2.2 An introduction to Phage Display

William Paley was a theologian who stated that if a watchmaker had designed the components for a watch, to have a defined structure for a particular purpose, then living organisms, being far more intricate than a watch, must have had a benevolent designer. One that adapted an organism to its environment by designing features that it would require to survive¹¹⁵. It is thought that Darwin was intrigued by Paley's idea of adaptation, and that organisms have characteristics that make them able to function in the environment in which they live, but that these were fashioned by a different process. The book, *The Blind Watchmaker*, written by Richard Dawkins proposes that evolution 'has no vision, no foresight' and therefore it is a process that occurs without a plan within given constraints¹¹⁶. If a chemist can introduce the components of evolution into a test tube then he could become, not just a designer of molecules but a 'blind' overseer of chemical evolution¹¹⁵.

Molecules do not possess the ability to 'evolve' per se but techniques such as phage display can be used, as the molecule in question is in fact a gene product and there is a direct link between the gene product and the gene¹¹⁷. Phage display can be thought of as analogous to natural selection, the differential survival and replication of a percentage of a population¹¹⁸. Another key component of evolution, mutability (the ability to pass on mutations to offspring) is not, as yet, incorporated into the phage display technique but this should be possible in the future with the development of 'sexual polymerase chain reaction', when error prone conditions are used during the reaction to induce mutations¹¹⁷.

The application of phage display begins with a phage-display library, which is a heterogeneous mix of phage clones. All the clones have a foreign DNA insert of random sequence, with the result that the make up of the insert differs in individuals. When expressed each phage has a differing peptide within one of their coat proteins. A form of chemical evolution occurs when individuals from the overall population are selected against a user defined criterion to give rise to a 'fitter' subpopulation¹¹⁵. This subpopulation replicates and further selection occurs, thus there is an ever increasing amplification of the selected phage. It is the ultimate combination of cloning and combinatorial chemistry which results in the production of great molecular diversity that can mimic the characteristics of the immune system¹¹⁹. The key to the process is that there is a direct link between the protein displayed on the surface of the phage and the gene that encodes for it, and that these are both packaged within the same virus¹¹⁹.

2.3 History

The term phage display was coined to encompass the technology that was developed from the pioneering work of George Smith in 1985. Smith used vectors to introduce a foreign DNA segment into the gene for a minor coat protein of a filamentous bacteriophage (phage for short). When this gene was expressed a foreign amino acid sequence is fused within the wild type protein. Smith cloned a restriction enzyme digest of plasmid DNA into the middle of the gene III. Gene III codes for pIII, a minor protein, in the coat. The N terminus of the protein is important in the infectivity of bacteria and so phage replication, whereas the C terminus participates in morphogenesis. Smith was able to insert a gene sequence that coded for an endonuclease in between these two regions. When Gene III was expressed the resulting pIII contained a foreign sequence that did not affect the function of the overall protein. He went on to expose an antibody to collection of cloned and wild type phage. Following affinity purification he showed that the cloned phage had been selected a thousand fold over the wild type phage with no insert¹²⁰.

Work published in 1990¹²¹ built on the foundations laid down by Smith to give us a technique that is more like the ones that are in use today. He and his colleagues randomly synthesized DNA sequences that were then incorporated into the genome of the phage via an expression vector and so creating a large, diverse library, in which the tip of the wild type pIII displayed a short variable amino acid sequence. Taking this diverse population they exposed it to a monoclonal antibody against L-enkephalin and found that from the vast phage population only those that carried a certain amino acid sequence were selected. He showed that it was possible to home in on a particular combination of amino acids that give a best fit to a molecule without having

prior knowledge of the ligand or antibody¹²¹. This early work was carried out on the *Escherichia coli* (*E. coli*) – specific bacteriophage M 13. Its success led to the development of systems using other phage such as those based on λ phage¹²² and T7 (Novagen, Darmstadt, Germany) but M13 phage display remains dominant¹²³.

2.4 Background

The most common type of M13 based phage libraries are constructed from DNA inserts that have been made by adding a mixture of nucleotides to a growing nucleotide chain, as opposed to the addition of single nucleotides. For example, taking the sequence XXYXXY, X is an equal mixture of A, C, G and T, with Y being an equal mix of G and T. Each XXY produced this way represents one of thirty two triplets that could code for any of the twenty amino acids¹¹⁵.

The M13 phage is a flexible rod like organism about 1 μm long and 6 nm in diameter and is lysogenic, therefore undergoes its life cycle without destroying the host bacterium. It contains a single strand of DNA, containing 11 genes enclosed in a coat made up of five different proteins¹¹⁵. There are about 2700 copies of the gene 8, encoding for pVIII, the major capsid protein which is fifty amino acids long¹²⁴. In addition to pVIII, the ends of the filament are capped by the minor coat proteins. There are five copies each of pVII and pIX at one end, and five copies each of pIII and pVI at the other¹²³, shown diagrammatically in Figure 2.1. The pIII is a 406 amino acid protein that initiates infection by attaching to bacterium that display the f pilus, a threadlike appendage that is used for the transmission of genetic information between *E. coli* cells¹²⁵. pIII is the most suitable protein to incorporate foreign peptide sequences up to twelve amino acids long, larger peptides cannot be fused to

this site as they can compromise infectivity¹²¹. Although the pIII gene can be successfully modified at the C terminal the more common methods involved variation is displayed at the N terminal¹²³.

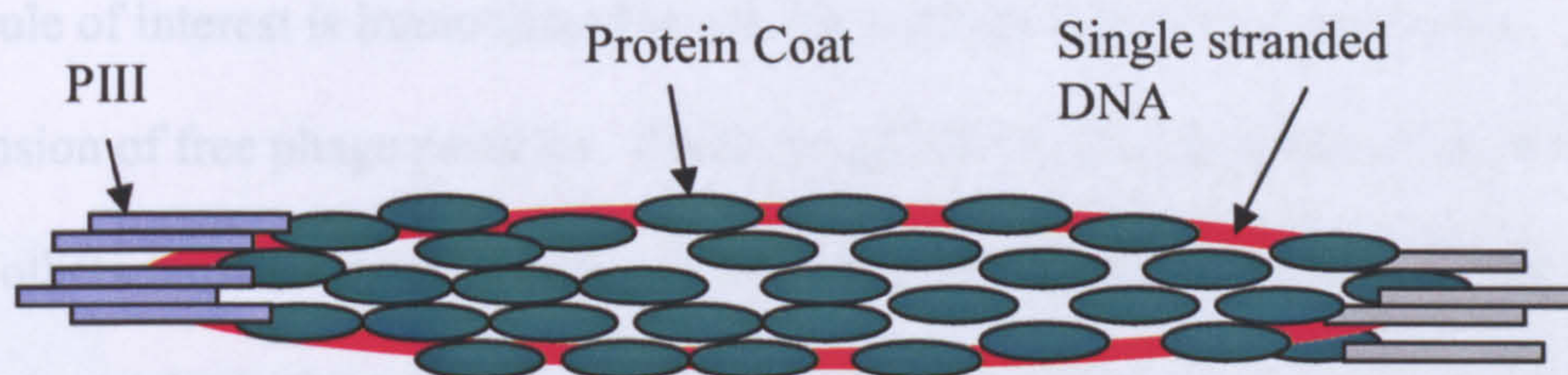


Figure 2-1. A representation of a phage, showing the single stranded DNA surrounded by many copies of the PVIII protein, with each end tipped with copies of the proteins PIII and PVI at one end and PVI and PIX the other. The overall length is 1000 nm and width is 6 nm, approximately.

The success of any technique that involves selection from a library depends on the diversity and the quality of the said library. The development of M13 phage libraries has been such that they are now considered extremely diverse and reliable¹²³.

2.5 Application

Phage display has been used in many different applications. These include identifying peptides that preferentially bind to double stranded oligodeoxyribonucleotides¹²⁶, isolating peptides that bind to immunoglobulins¹²⁷, to identify non competitive enzyme inhibitors¹²⁸ and to identify proteins with novel enzymatic activity¹²⁹. These are all examples of using phage to identify a ligand that fits into a known binding site. However, more recently phage have been used to identify partial amino acid sequences that catalyse the inorganic nanostructures made of calcium carbonate and silica that are found in many diatoms¹³⁰.

Two important parameters of selection are stringency and yield which often have an inverse relationship. That is, the greater stringency (conditions likely to perturb or

disrupt affinity for the receptor) the lower the yield (the number of phage that survive the selection process). In the early rounds of selection a high yield of clones is important as it is from these that stringency can be improved¹¹⁵. In general the molecule of interest is immobilized to a surface which is then covered with a suspension of free phage particles. Some phage bind to the immobilised molecule, while others remain free in solution. Probing a phage library against a specific molecule results in the selection of a sub population, effectively removing phage that do not have any binding qualities. This is achieved by panning or screening the entire population against the molecules of interest.

2.6 Indication of phage presence

The Ph.D.-12™ Kit (New England Biolabs) uses phage that are derived from M13mp19 and possess a region of *E. coli* operon *lac*. This operon consists of three genes, two of which, Z and Y are required for the metabolism of the milk sugar lactose, and the third which encodes for thiogalactoside transacetylase (as shown in Figure 2.2). The phage contains DNA that encodes the N-terminal fragment of β -galactosidase³⁷ when induced by the presence of isopropyl- β -D-thiogalactoside (IPTG). This is capable of intra-allelic (α) complementation with the defective host β -galactosidase, resulting in a functioning enzyme¹³¹. The natural substrate is lactose and results in the formation of galactose and glucose but it will also catalyse the hydrolysis of 5-bromo-4-chloro-3-indolyl- β D-galactoside (Xgal) as a substrate with the release of a blue indicator. Figure 2.3 shows which bonds are broken to give rise to the release of the indicator molecule. Bacterial colonies show on IPTG/Xgal plates as blue spots, as shown in Figure 2.4. Wild type virus that may have infected the *E. coli* show as white plaques and can be avoided. Figure 2.4 also shows a dilution

series, this was done on a single plates and 10 μL of each dilution was placed in each square, this allowed for a rough count of which dilution range would give the most accurate readings on which to base the titer.

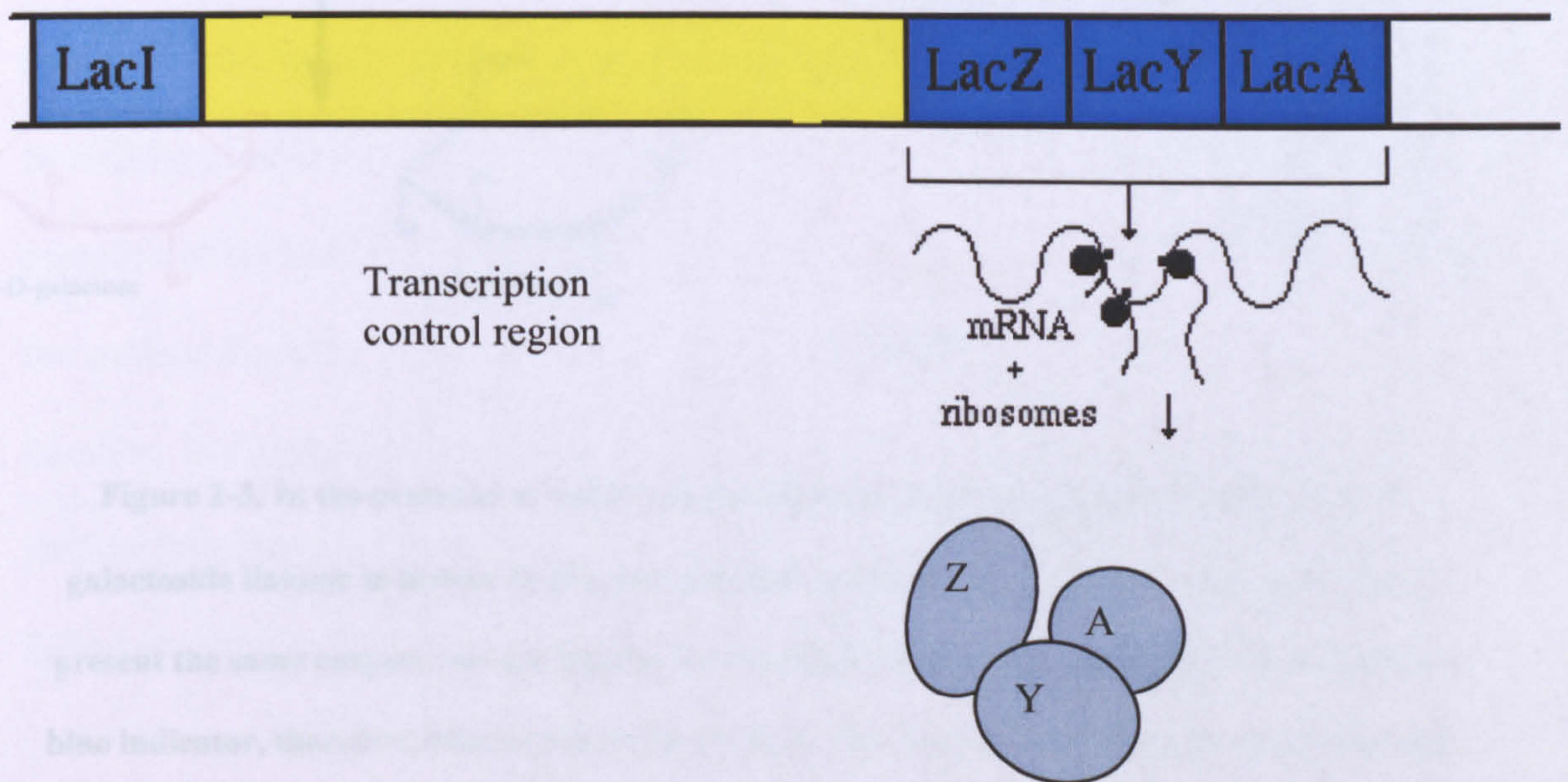


Figure 2-2. The *lac* operon consists of three genes: *lacZ* encodes β -galactosidase (Z), *lacY* encodes lactose permease (Y), and *lacA* encodes thiogalactoside transacetylase (A). Transcription is controlled by the binding of proteins within the transcription control region, coloured yellow, and *lacI* encodes the lac repressor. (Adapted from³⁷).

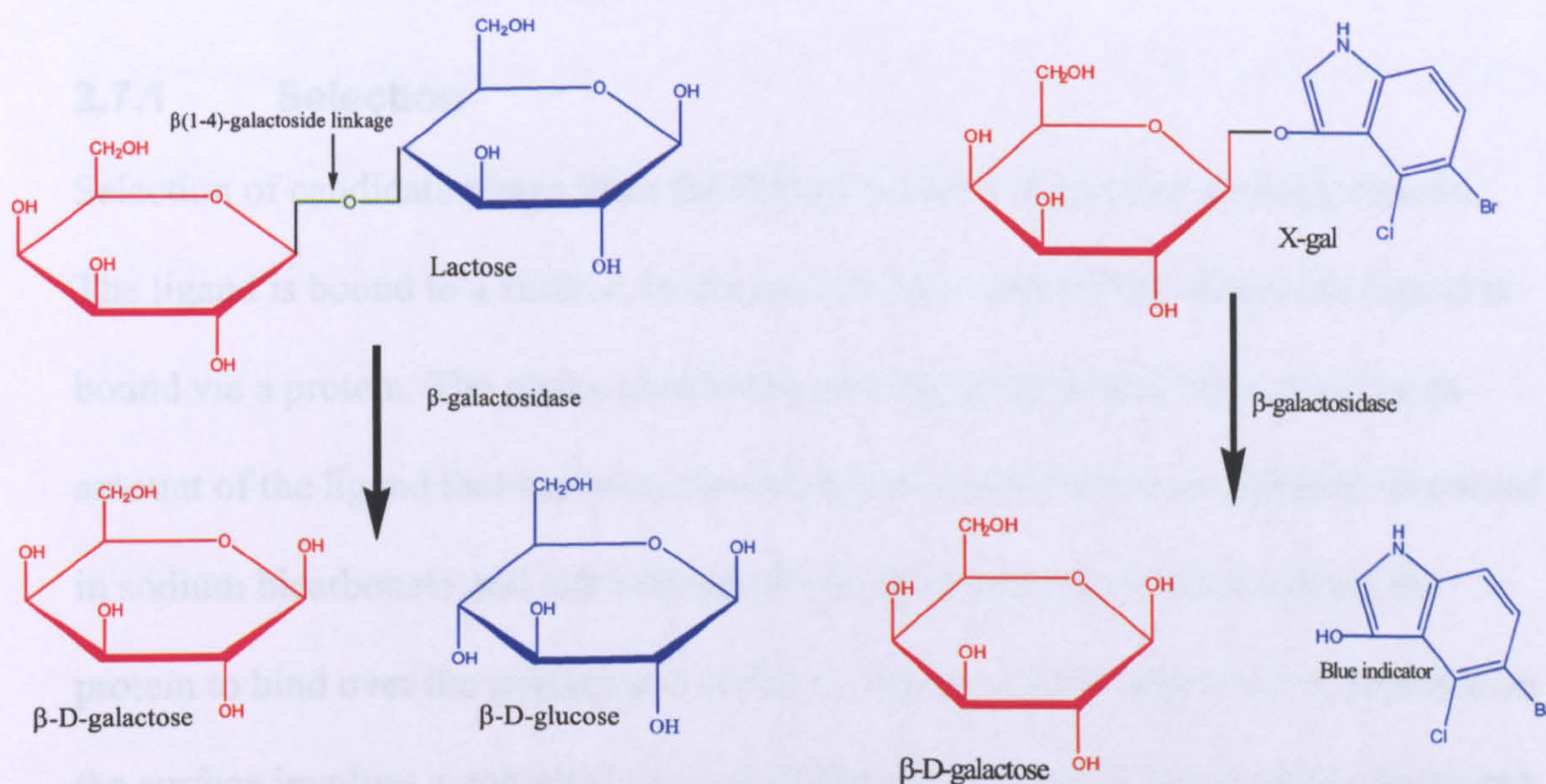


Figure 2-3. In the presence of water galactosidase catalyses the hydrolysis of the $\beta(1-4)$ -galactoside linkage in lactose to give two products galactose and glucose. When no lactose is present the same enzyme can use Xgal as an alternative substrate, resulting in the release of a blue indicator, therefore infected bacterial colonies show up on IPTG/Xgal plates as blue spots

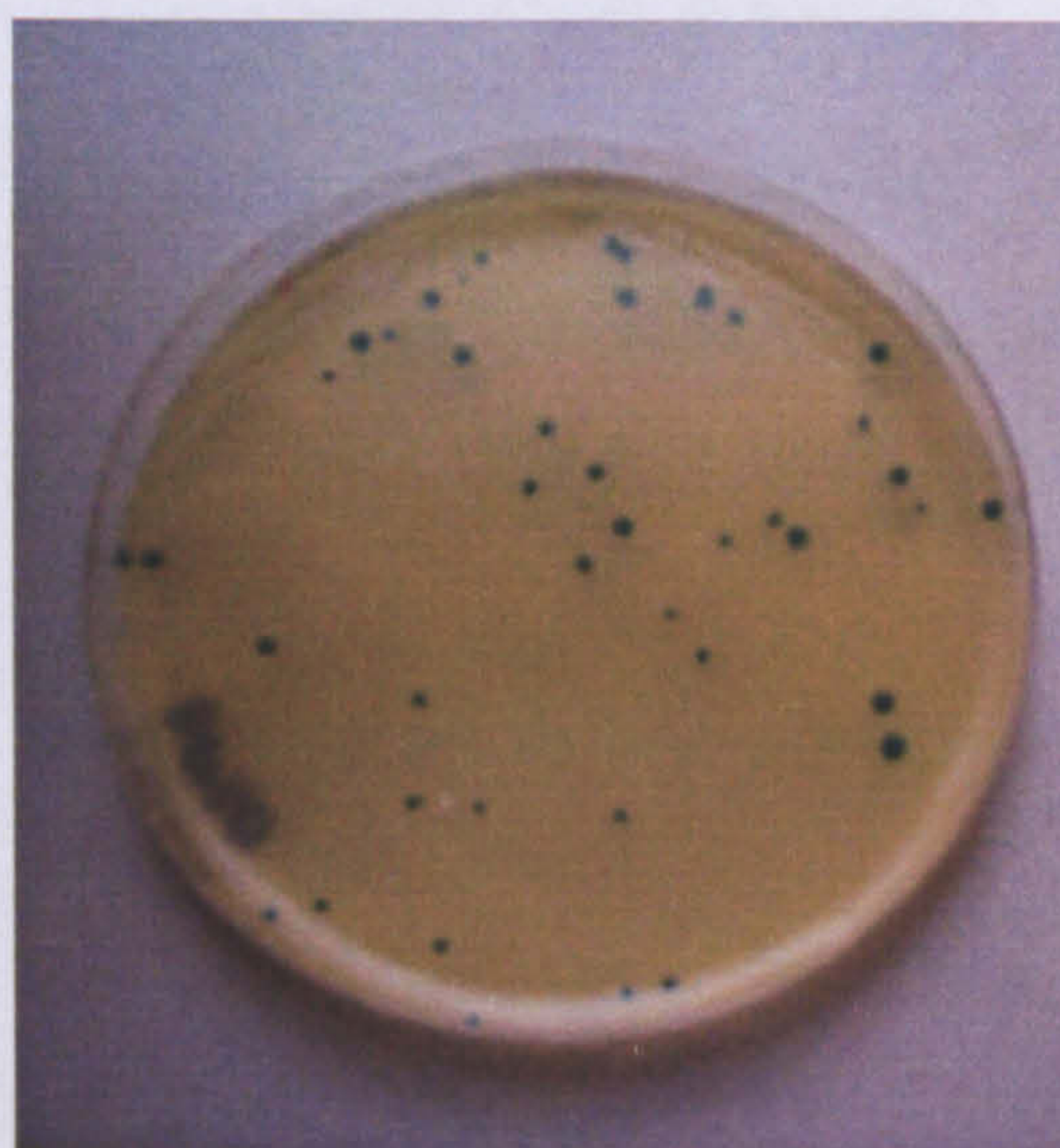
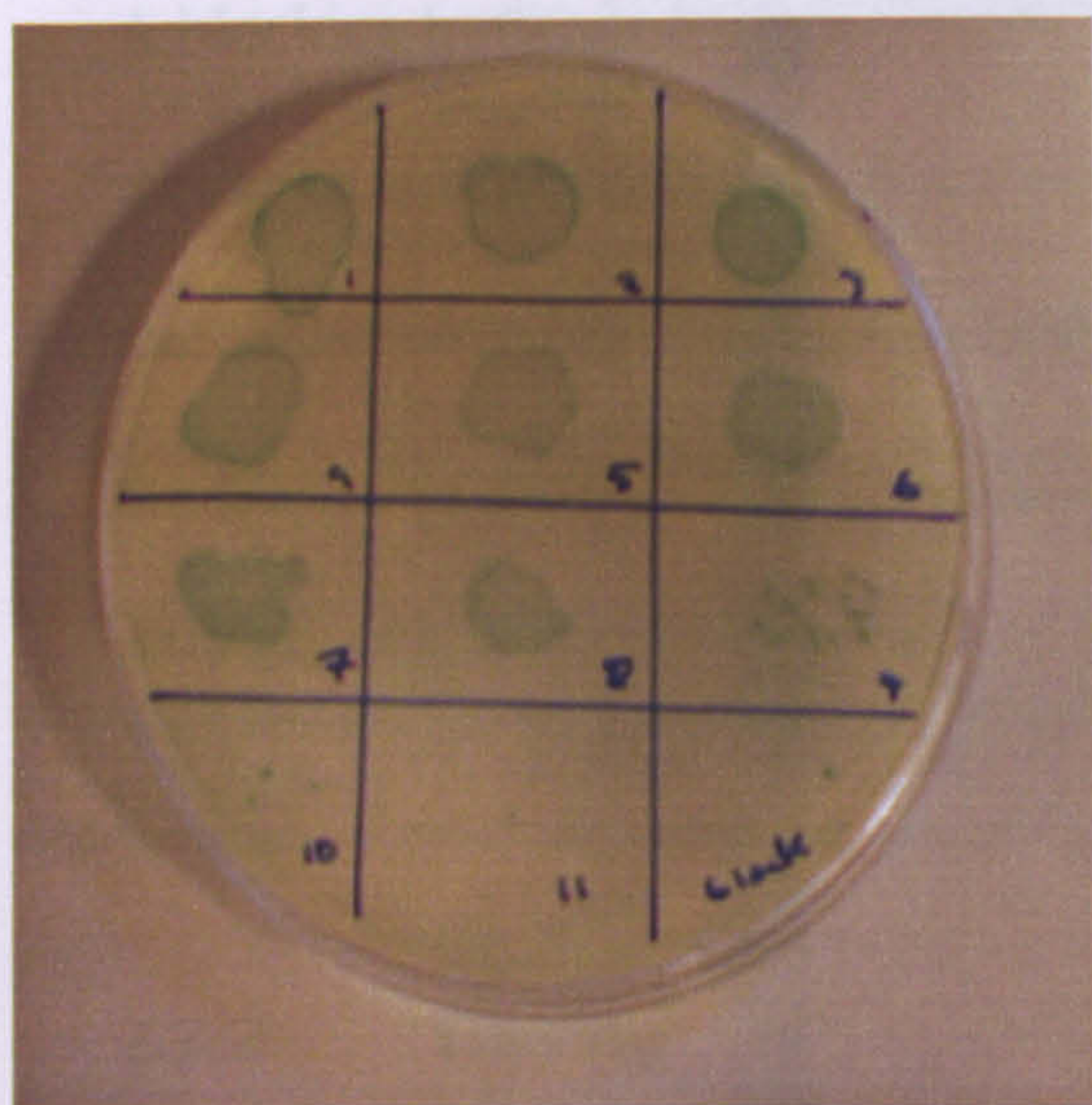


Figure 2-4. A dilution range is established by placing a small amount of each dilution into each square of the sample plate, at high phage concentration the spot appears totally blue and as the dilution decrease the amount of blue decreases until discrete spots are visible. The second photograph is a typical plate clearly showing the blue plaques that show single colonies of *E. coli* infected with the bacteriophage.

2.7 Method of Use

2.7.1 Selection

Selection of candidate phage from the library occurs via repeated panning rounds.

The ligand is bound to a surface, in the case of polystyrene Petri dishes the ligand is bound via a protein. The plates used in the panning are prepared by dissolving an amount of the ligand that has been previously covalently bound to a protein dissolved in sodium bicarbonate and left with gentle swirling movement which allows the protein to bind over the polystyrene surface. The interfacial behaviour of proteins on the surface involves a complex array of interactions including electrostatic, hydrogen bonding, hydrophobic and van der Waals that involve not only the protein and the surface but also the solvent and other solutes. These result in the protein and any ligand molecule covalently bound to it being adsorbed on to the Petri dish surface.

A blocking buffer is washed over the surface of the plate to fill any sites on the plate that have not reacted the protein and so prevent non-specific binding. The ligand is then exposed to the library with gentle agitation . The phage that have not bound to the plate are removed and the binding between the phage and the ligand is disrupted with glycine-HCl and BSA which act as a general buffer for non-specific disruption of binding.

2.7.2 Titering

Titering is a method by which the number of phage present in a given volume is determined, usually measured as the number of plaque forming units (pfu) per μL . A small quantity of the phage undergoes a series of dilutions and each dilution is used to infect a culture of *E. coli* and incubated overnight. The plaques formed are

multiplied by the dilution factor and the pfu determined. A phage titer was carried out to determine the number of plaque forming units that exist within the library. A series of dilutions of the original library were prepared as below (shown in Table 2.1). 10 μL of each dilution of these were added to 200 μL of bacterial culture at the mid log phase (OD_{600} 0.5). Each was then added to 3 mL of agarose top at 47 $^{\circ}\text{C}$ and poured over prepared IPTG/Xgal plates (100 mL LB/agar containing 4 mg Xgal and 4.5 mg IPTG dissolved in 100 μL DMF is sufficient for 6 plates), warmed to 37 $^{\circ}\text{C}$. These plates were incubated overnight.

Sample number	LB medium μL	Phage μL	Dilution	Plaques formed
1	49	1 of the original library	1/50	Blue plate
2	45	5 of sample 1	1/500	Blue plate
3	45	5 of sample 2	1/5000	Blue plate
4	45	5 of sample 3	$1/5 \times 10^4$	Blue plate
5	45	5 of sample 4	$1/5 \times 10^5$	Too many to count
6	45	5 of sample 5	$1/5 \times 10^6$	517
7	45	5 of sample 6	$1/5 \times 10^7$	48
8	45	5 of sample 7	$1/5 \times 10^8$	4
9	45	5 of sample 8	$1/5 \times 10^9$	1

Table 2-1. The dilutions were made and used to establish the titer of the original library. The number of blue colonies produced with each dilution from which the number of phage per μL was determined.

By multiplying the number of plaques by the dilution factor the pfu can be determined. The pfu per 1 μL of the library is approximately 2.5×10^8 , this compares to the suggested pfu of 4×10^9 from the company literature. The phage collected after each panning are also titered in the same way and this reveals the percentage of the library that has been selected.

2.7.3 Amplification

The phage collected from the first panning round can then be increased in number by amplification. When the phage infects it uses the host's metabolism to replicate. The phage is extracted from the bacteria and the titer is again measured to give the number of phage per μL which is used to calculate the amount required to input the phage in the second panning round.

Subsequent pannings can take on a variety of forms, exposing the ligand to further rounds aim to increase the affinity of the phage to the ligand. Also it is possible to carry out a negative pan, that is, to expose the phage to another, structurally different, molecule and take the phage that show no binding affinity. This can be used to screen out phage that have bound to the carrier protein molecule rather than to the ligand in question.

The first panning round used the O⁶CMG-ovalbumin conjugate (see section 3.2.2.1) which selected phage that showed affinity for this moiety. The second panning round was against ovalbumin only, this time the non-binding phage were collected.

Therefore the phages that were used in the third panning had affinity for O⁶CMG-ovalbumin but not for ovalbumin. The final pan was against O⁶CMG linked to glass

(see section 3.2.2.3) beads, the phage amplified in this round were assumed to have a high affinity to O⁶CMG in order to have survived the successive panning rounds.

2.8 DNA determination

The optimum balance between stringency and yield occurs after two or three pannings¹³². When the appropriate pannings have been completed the phage are plated so that there are less than one hundred plaques are on one plate, this ensures that each plaque contains a single phage. Phage can 'leak' from one plaque to another if they are too close together. These plaques are used to determine the DNA sequence of the phage and this will ultimately give the amino acid sequence that was incorporated into the vector in the original library.

The DNA concentration from each plaque is determined by UV as measuring the OD at 260 nm as 1 OD₂₆₀ is equivalent to 33 µg per µl of single stranded DNA. Between 100-250 fmol of DNA template is required for the polymerase chain reaction (PCR).

PCR can be used to amplify the short DNA sequence that contains the variation giving rise to the amino acid sequence so that it can be sequenced. This can be done manually or automatically depending on the method used in the PCR. For automated sequencing a -96 primer end labelled with a fluorescent dye (IRD800™ Li-Cor) was used to allow for detection of the sequence products. The primer is introduced to the DNA template along with a temperature resistant DNA polymerase and a supply of dideoxynucleotides and deoxynucleotides. The PCR involves heating the template to 95°C to denature the template. The temperature is reduced to room temperature to allow the primer to anneal to the template, increasing the temperature to 65-70°C

allows the polymerase to extend from the annealed primer. This process is repeated as many times as considered necessary, as shown in Figure 2.5. When the reaction is complete a Stop/Loading buffer is added and then template is ready for sequencing.

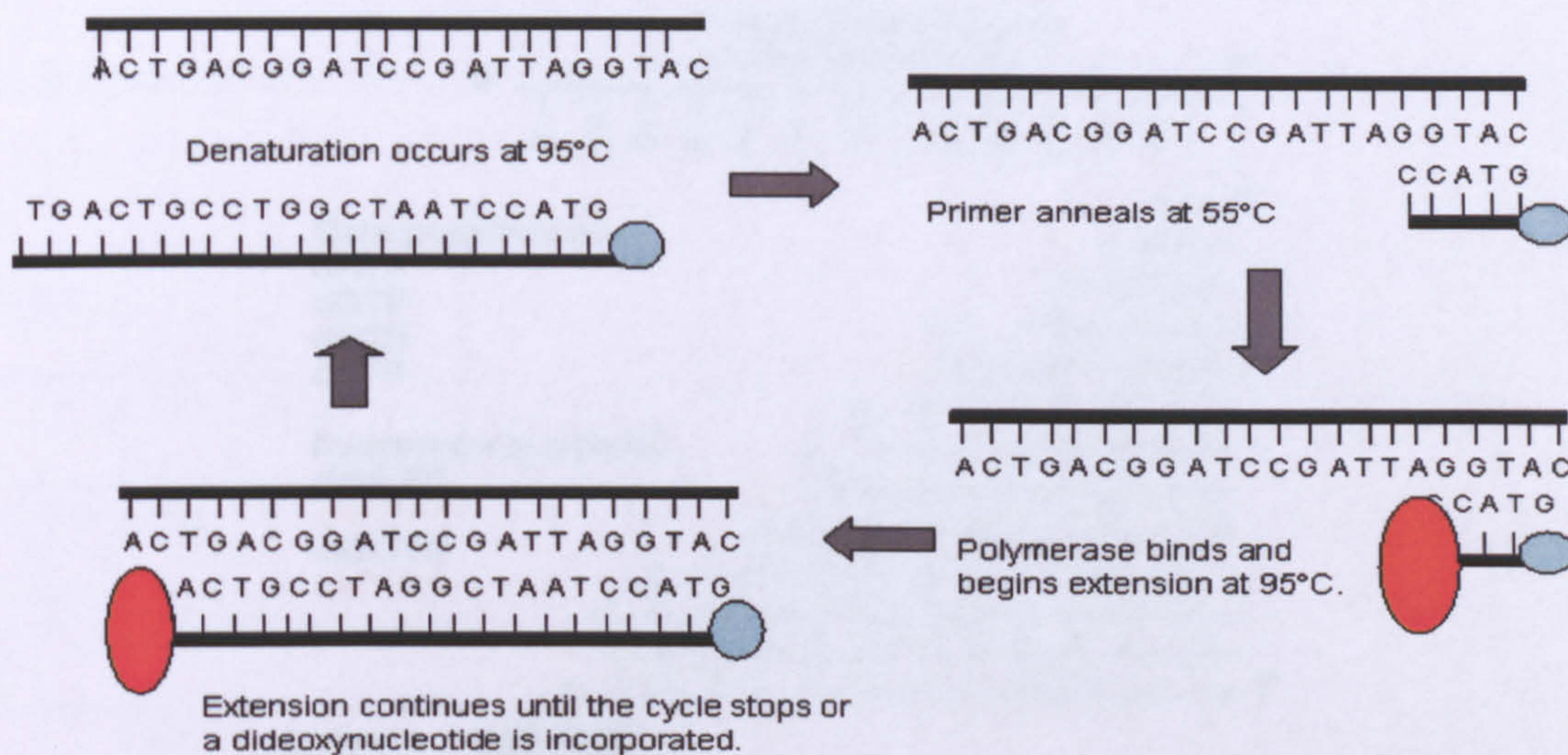


Figure 2-5. A summary of the PCR in which the DNA is denatured, the labelled primer anneals to a specific sequence, at the higher temperature the polymerase extends from the primer incorporating nucleotides until the cycle stops or when a dideoxynucleotide is included in the growing chain which prevents further nucleotides being attached. The cycle is repeated numerous times.

Sequencing is possible because of the introduction of dideoxynucleotides which do not have a hydroxyl group on the 3' and so prevents the addition of further deoxynucleotides (summarised in Figure 2.6). Separation of the fragments occurs via gel electrophoresis on an agarose or polyacrylamide gel that has been loaded with the DNA fragments and a current is applied. DNA is negatively charged and will migrate towards the positive pole however the larger pieces of DNA are held within the gel

matrix and are slowed down. Smaller pieces of DNA move through more quickly.

The DNA sequence of the non coding strand is determined and using the Crick and

Watson base pairing rule the coding strand from the phage is found.

diluted plate and the DNA coding for the ...
following amino acid sequence

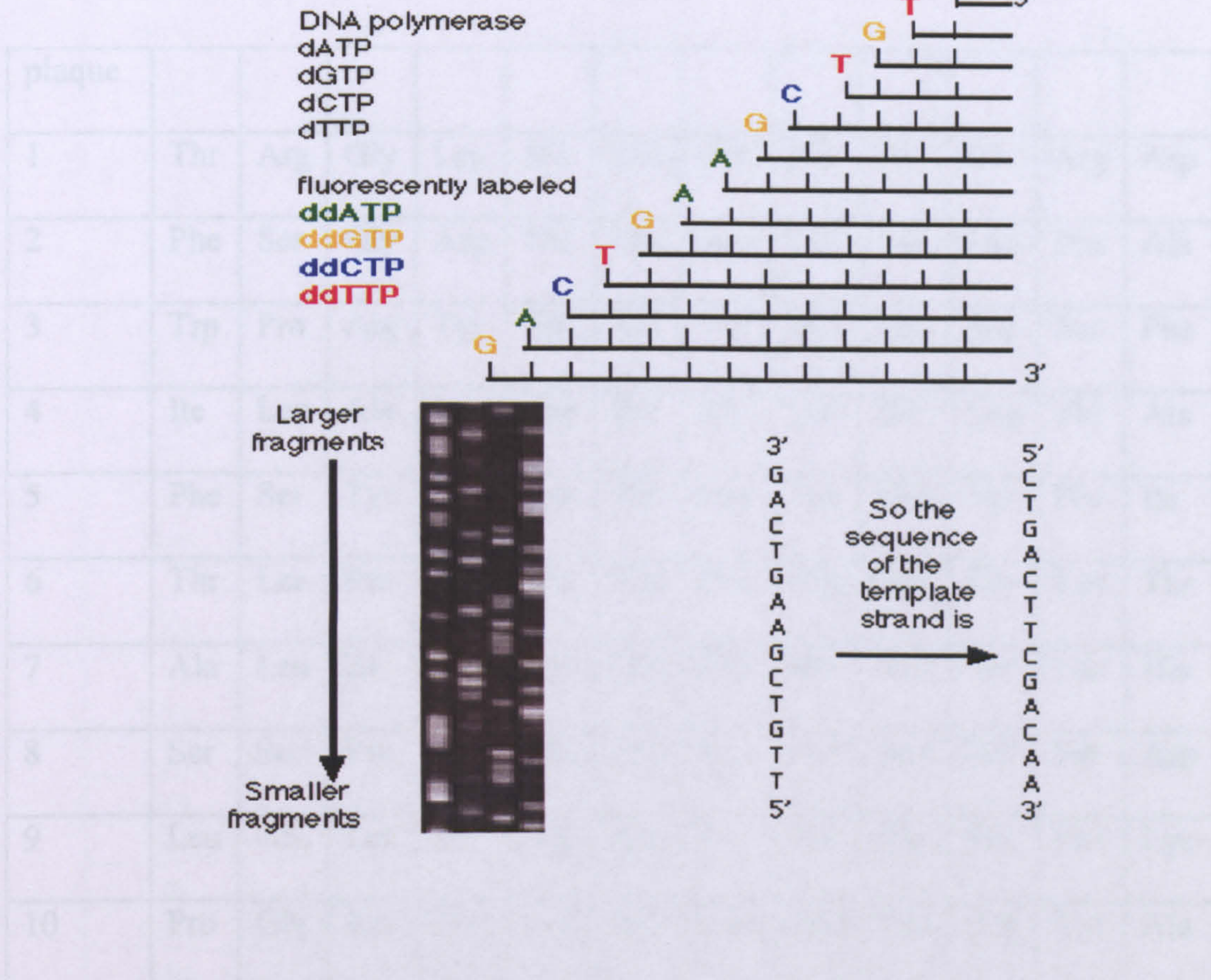


Figure 2-6. A brief explanation of how the PCR products are read in the automated sequencer.

The complementary strand of the DNA to be sequenced is formed during the PCR as fragments

as the extension is stopped when a ddnucleotide is incorporated into the strand. This results in

the formation of differing lengths of DNA; these are separated on the gel whilst an electrical

current is applied. The bands appeared in lanes in the order that they were loaded, ACGT, and

so the sequence can be read. The corresponding sequence is determined and from this the amino

acid sequence.

pattern in the types of ...
more prevalent. An ...

2.9 Results

It is suggested in the literature that ten blue plaques need to be sequenced to give a consensus between clones¹³³. Therefore ten plaques were taken from a suitably diluted plate and the DNA coding for the end of the pIII protein was sequenced. The following amino acid sequences were found:

plaque												
1	Thr	Arg	Gly	Leu	His	Leu	Ser	His	Ser	His	Arg	Asp
2	Phe	Ser	Gly	Asp	His	Leu	Arg	Glu	Lys	Leu	Pro	Ala
3	Trp	Pro	Arg	Tyr	Thr	Ser	Thr	His	Leu	Pro	Ser	Phe
4	Ile	Leu	Ala	Leu	Gly	Pro	Ser	Lys	Ser	Leu	Thr	Ala
5	Phe	Ser	Tyr	Glu	Lys	Thr	Leu	Lys	Thr	Tyr	Pro	Ile
6	Thr	Leu	Pro	Thr	Thr	Phe	Pro	Arg	Ala	Thr	Leu	Thr
7	Ala	Leu	Ile	His	Ser	Leu	His	Ser	Asn	Ser	Leu	His
8	Ser	Ser	Pro	Arg	Trp	Phe	Ile	Pro	Asn	Ser	Ser	Asp
9	Leu	Gln	Leu	Ile	Arg	Leu	Ile	Ala	His	Pro	Pro	Lys
10	Pro	Gly	Tyr	Ser	Pro	Ala	Thr	Glu	Leu	Arg	Tyr	Ala

Table 2-2. The amino acids identified from ten phage plaques following the three panning rounds.

Although a discrete consensus sequence of amino acids was not found, there was a pattern in the types of amino acids present with those having basic side chains being more prevalent. An amino acid distribution was available in the product literature in

which 104 clones from the native library were sequenced. Six of these did not contain the insert but from the other 98 the distribution pattern was obtained as a frequency percentage, this is given in the table below along with the distribution pattern achieved from the selected ten clones.

Amino acid	type	reported frequency %	observed frequency %	ratio
Gly	aliphatic	2.6	3.3	1.3
Ala	aliphatic	6	6.7	1.1
Pro	aliphatic	12.2	10.8	0.9
Val	aliphatic	3.9	0	0
Leu	aliphatic	9.3	15	1.6
Ile	aliphatic	3.4	5	1.5
Met	aliphatic	2.6	0	0
Phe	aromatic	3.3	4.1	1.2
Trp	aromatic	2.2	1.6	0.7
Tyr	aromatic	3.6	4.2	1.2
Ser	polar uncharged	10	13.3	1.3
Thr	polar uncharged	11.1	10	0.9
Cys	polar uncharged	0.5	0	0
Lys	Basic	2.8	4.17	1.5
Arg	Basic	4.7	6.7	1.4
His	Basic	6.3	7.5	1.2
Asp	Acidic	2.8	2.5	0.9
Glu	Acidic	3.1	2.5	0.8
Asn	Acidic amide	2.8	1.7	0.6
Gln	Acidic amide	5.1	0.8	0.2

Table 2-3. The distribution of amino acids found in ten phages selected following the three panning rounds, shown as a frequency percentage and the frequency percentage from 98 native phage (supplied by the manufacturer of the Ph.D.12 library). The final column shows the ratio between the two.

2.10 Discussion

Three amino acids, valine, methionine and cysteine are not represented at all in the selected clones. Although cysteine is under represented in the native library because its presence interferes with phage infectivity¹³², the other two are present in the native library and appear to have been excluded. The acidic amide amino acids asparagine and glutamine also appear to have been excluded. Others that are underrepresented include glutamic acid and the aromatic tryptophan. All the basic amino acids appear more often in the selected clones compared to the native library with lysine and arginine appearing 50% and 40% more respectively. The aliphatic amino acids leucine and isoleucine also are over represented.

It was expected the negatively charged amino acids might contribute to the binding of the purine base however, the results suggested that this is not the case, and these can be excluded from any library. It was also expected that positively charged amino acids would be included in a binding site that is complementary to the negatively charged carboxymethyl moiety and this appears to be borne out by the results. The aromatic amino acids phenylalanine and tyrosine are also slightly in excess of the normal representation of the library, suggesting that the planar ring-ring interactions are also contributing to the overall binding affinity.

It is difficult to draw further conclusions from this work. It was hoped that a consensus between the clones would give a definitive pattern of amino acids on which a reagent could have been based. However, with hindsight it is perhaps unlikely that this would have been possible. Past literature shows many examples of a ligand being found that fits into a preformed binding pocket, either peptidic or RNA in origin¹²⁶⁻¹²⁸,

but there is no evidence that phage display has been used to find the binding area of a given ligand even though, theoretically, it should be possible.

Attention was therefore turned to synthesising a combinatorial library of compounds based on the amino acids suggested in the conclusion of the phage display work.

Making a whole host of slightly differing molecules would increase the possibility that a potential novel reagent could be isolated from the library.

Chapter 3

Construction and characterisation of cholate-amino acid libraries for recognition of O⁶-carboxymethyl-2'-deoxyguanosine

'Science is as an edged tool, with which men play like children and cut their own fingers'. Sir Arthur Eddington (1882-1944)

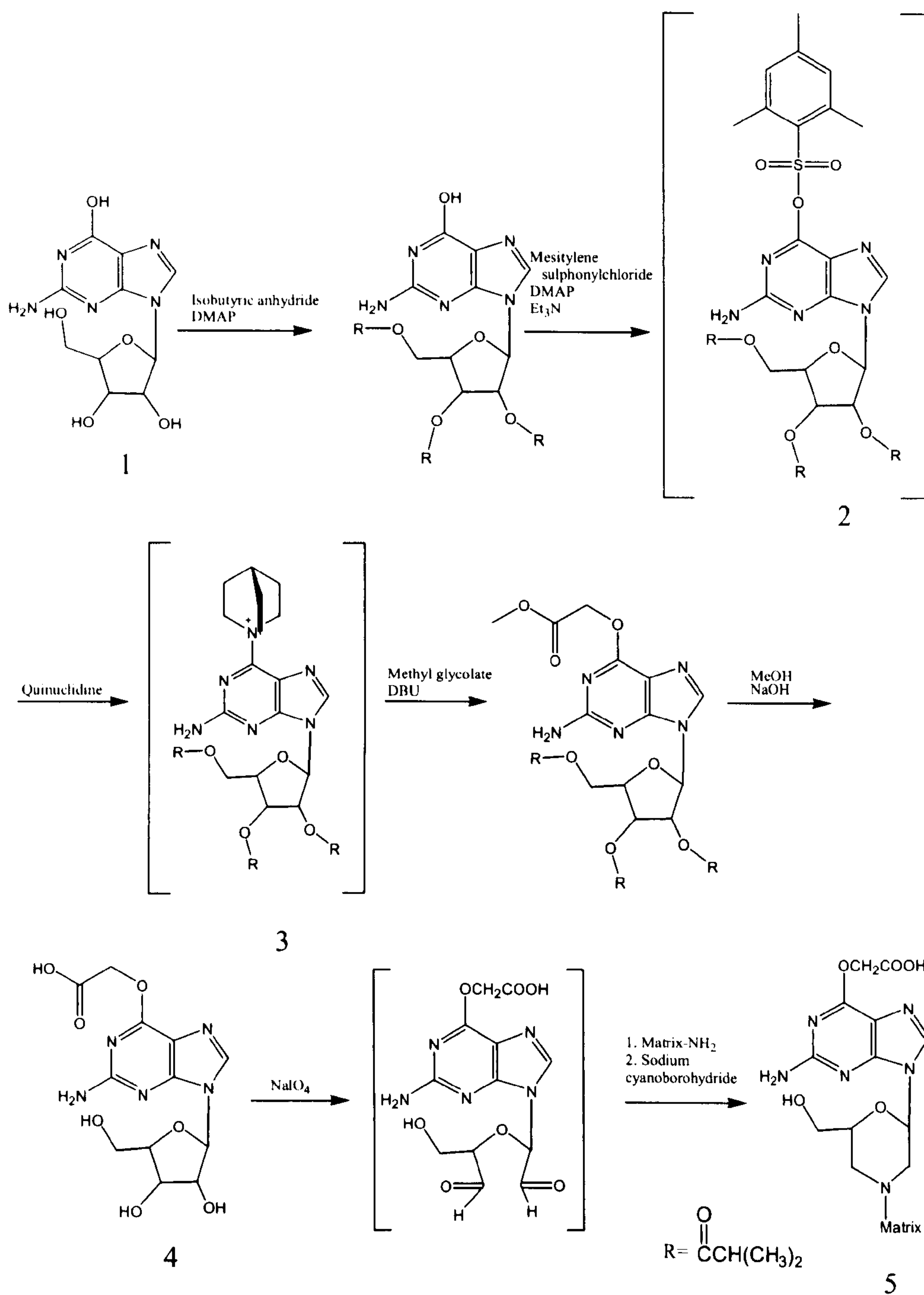
3.1 Introduction

Interest in the detection of O⁶carboxymethyl guanine in DNA of gastrointestinal cells has led to a need for a novel reagent. The reagent must fulfil certain criteria: it must be as good at detecting the ligand, if not better, than the pre-existing techniques; its results need to be reproducible and the reagent needs to be synthesised in large quantities. It was decided that the best method of finding such a reagent was to construct a library of compounds, with a putative binding site, the synthesis of which was directed so as to maximise the number intermolecular forces that may participate in the binding between the DNA adduct and the reagent. Candidate reagents would be identified by screening this library. With these prerequisites in mind, the first stage was the synthesis of the ligand molecule in a form that could readily be immobilised.

3.2 Synthesis of immobilised O⁶-carboxymethylguanine

3.2.1 Synthesis

The synthetic scheme for preparation of immobilised O⁶-carboxymethylguanine is shown in Scheme 3.1. The 2',3'5' hydroxyl groups on the ribose ring of guanosine (**1**) were protected and to ensure that the more labile ester at the 2' remains protected throughout the synthesis¹³⁴, isobutyryl protection was used. The O⁶ oxygen atom is unaffected by the protection step as it is less nucleophilic than the hydroxyl groups on the ribose ring, due to delocalisation of the non bonding pair of electron into the adjoining cyclic system. Based on previous work it was not considered necessary to protect the exocyclic N²-amino group¹³⁵ especially as the removal of protection from the site using standard ammonia techniques, may result in substitution of the O⁶-alkyl by an amino group¹³⁶. The protection step was completed with an overall yield of 74%.



Scheme 3-1. General synthetic scheme for the preparation of immobilised O⁶-carboxymethylguanine.

The next step involved the introduction of mesitylene sulphonyl chloride to give O^6 -[(2,4,6-trimethylphenyl)sulphonyl]-2',3',5'-tri-*O*-isobutyryl guanosine (**2**), this was readily followed by TLC as the starting material shows as a baseline spot and the product appears on the solvent front as it is considerably less polar. This was followed by the introduction of quinuclidine which displaces the mesitylene sulphonyl group to give a quinuclidium salt (**3**). This appeared, on TLC, as a blue, fluorescent, baseline spot, and again it is easy to see when the reaction reached completion. Addition of methyl glycolate yielded the carboxymethyl methyl ester at the O^6 position. Deprotection, using sodium hydroxide, resulted in the formation of the carboxymethyl group and regenerated the hydroxyl groups on the ribose ring, to give O^6 -carboxymethyl guanosine (O^6 CMG) (**4**). The difference between O^6 CMG and guanosine can clearly be seen by UV (Figure 3.1).

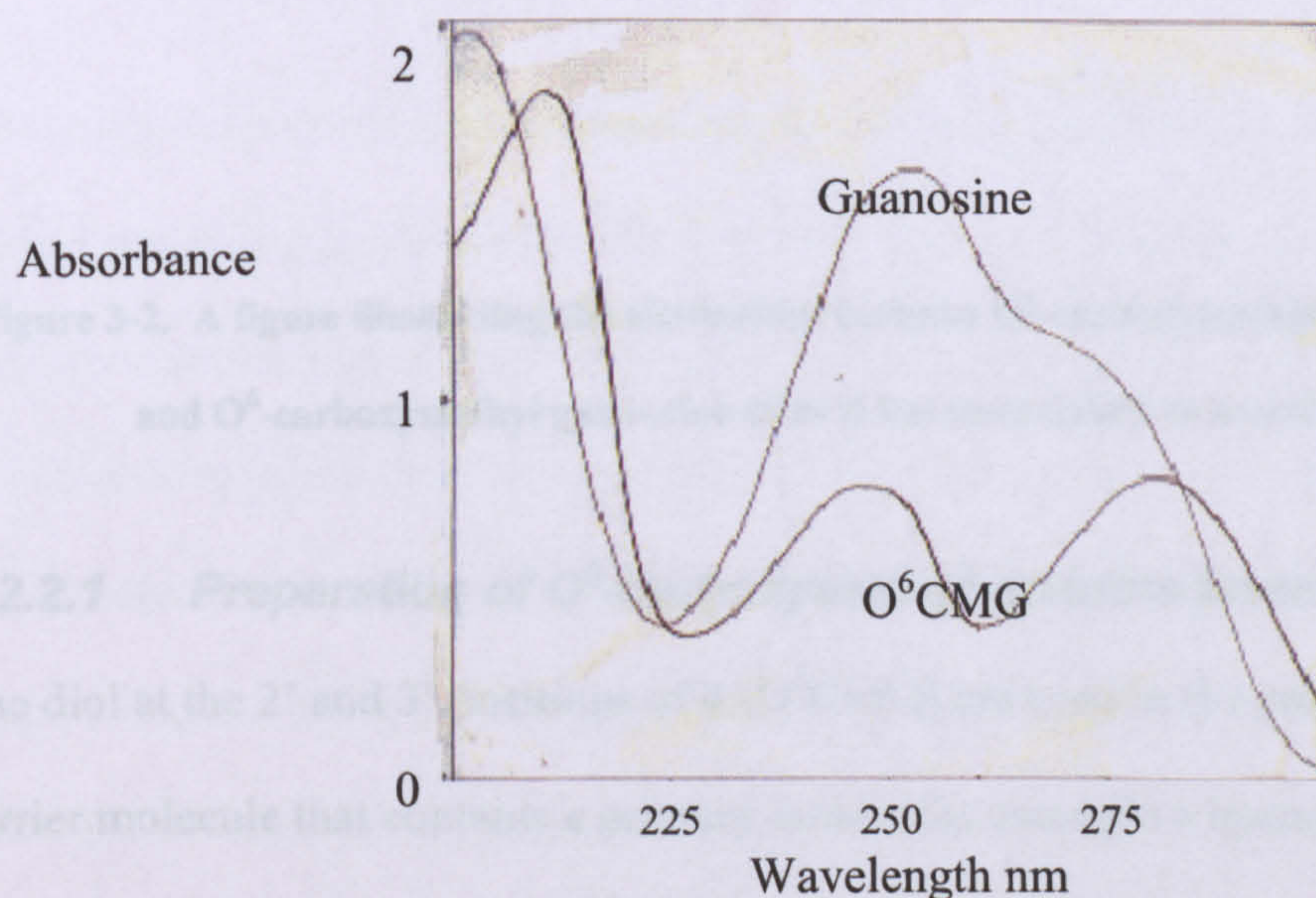


Figure 3-1. The UV traces comparing guanosine, the upper trace, overlaid with the UV trace of O^6 CMG at the same concentration.

3.2.2 Conjugation of the ligand to matrices

Attaching O⁶-carboxymethyl guanosine to a carrier molecule or matrix results in the formation of a moiety that closely resembles O⁶-carboxymethyl-2'-deoxyguanosine.

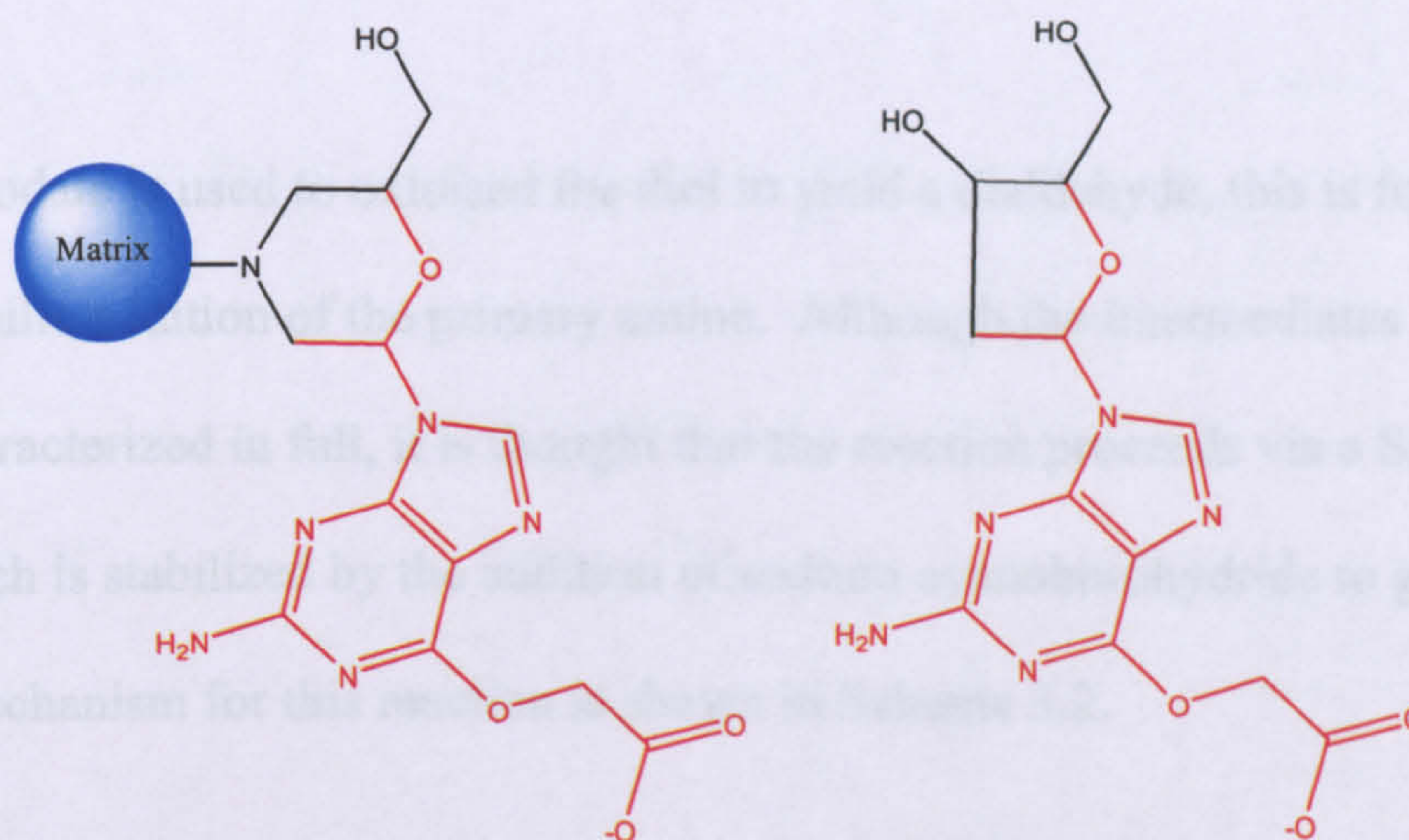


Figure 3-2. A figure illustrating the similarities between O⁶-carboxymethyl-2'-deoxyguanosine and O⁶-carboxymethyl guanosine after it has been linked to a carrier matrix.

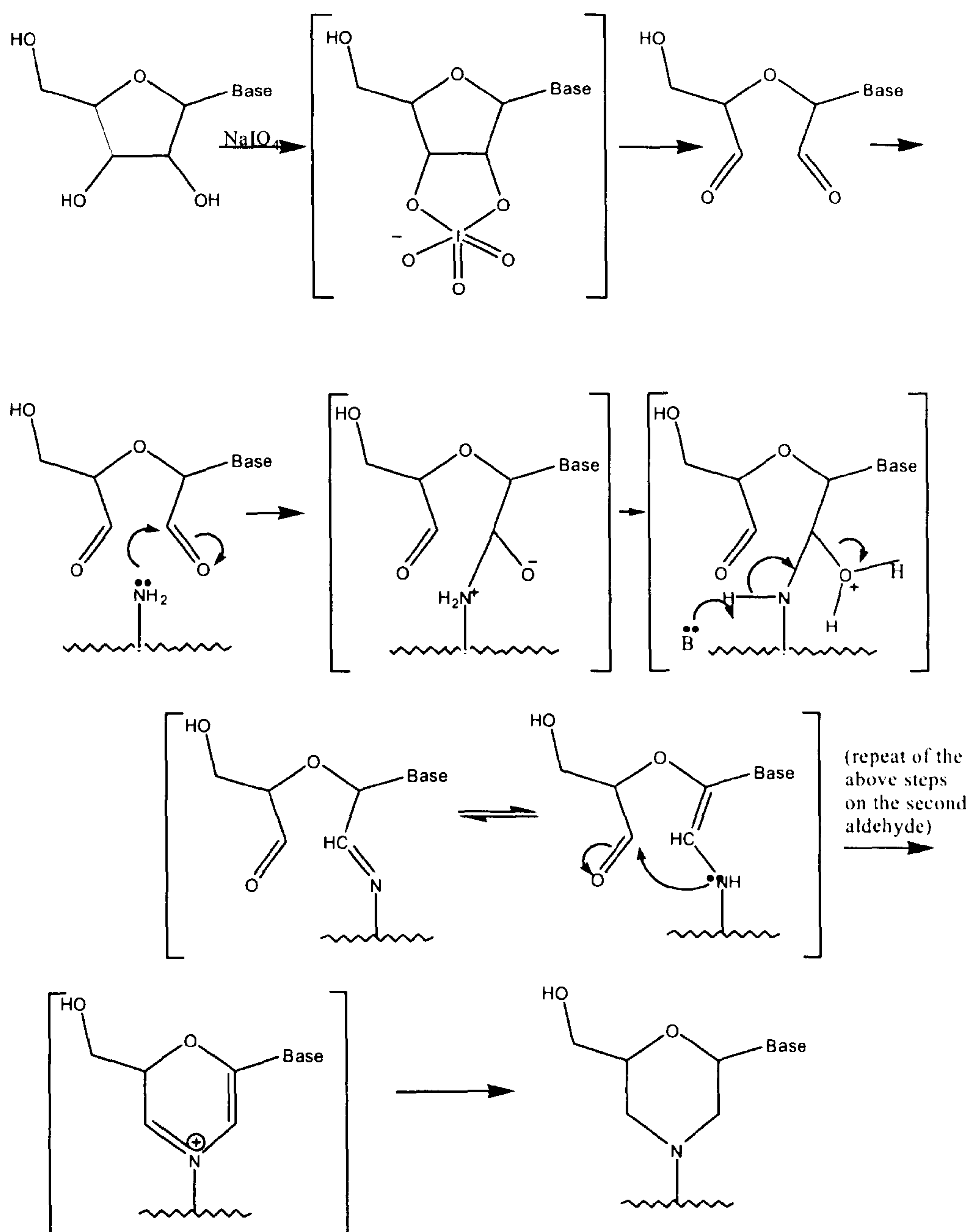
3.2.2.1 Preparation of O⁶-carboxymethyl guanine bound to ovalbumin

The diol at the 2' and 3' positions of **4** (O⁶CMG) are used in the conjugation to a carrier molecule that contains a primary amine, for example a lysine side chain residue. Ovalbumin is a protein that has 385 amino acids, seventeen of which are lysine. Therefore, theoretically, seventeen modified guanosine molecules may be attached to an ovalbumin protein. In reality only one molecule was linked, possibly due to the folded structure of the protein in which many of the lysine amino acids are likely to be inaccessible.

Erlanger and Beiser (1964)¹³⁷ developed a general method for attaching purines and pyrimidines to protein carriers. In the original paper the carrier in question was bovine serum albumin (BSA) but the method has been used for other carriers. While this method of coupling may not be suitable for some oxidized nucleosides¹³⁸, it was previously used to couple O⁶CMG to both BSA and ovalbumin¹³⁹.

Sodium periodate is used to oxidise the diol to yield a dialdehyde, this is followed by the nucleophilic addition of the primary amine. Although the intermediates have not yet been characterized in full, it is thought that the reaction proceeds via a Schiff base¹³⁹, which is stabilized by the addition of sodium cyanoborohydride to give **5**. A proposed mechanism for this reaction is shown in Scheme 3.2.

Following the procedure in Scheme 3.2, the resulting mixture was dialysed overnight against phosphate buffer saline (PBS) to remove salts and unbound nucleoside. The mixture was then eluted through a Sephadex G50 column with water. Sephadex G50 is a column matrix in which the swollen beads of the matrix are large enough to allow small molecules to enter in to the bead¹⁴⁰. When a mixture of varying sized molecules is applied to the top of the column the larger molecules are able to pass relatively quickly down the column between the gel particles. However, the smaller molecules enter the beads and as a consequence they take a longer route down the column and require a greater elution volume before being released from the column. Monitoring the UV at 210 nm the first absorbing fraction was collected and freeze dried.



Scheme 3-1. The proposed reaction mechanism and structural intermediate formed during the linkage of O⁶CMG to the primary amine side chain residue of lysine found in the ovalbumin protein. The resulting molecule is reduced with sodium cyanoborohydride to give 5.

3.2.2.2 Determination of the hapten-carrier ratio

A standard curve (Figure 3.3) of hapten-carrier ratio was determined at 280 nm and the UV absorbance for the adduct-protein was 0.691 which gives an average ratio of 1.44 adducts per protein. This is consistent with previously published results¹³⁹.

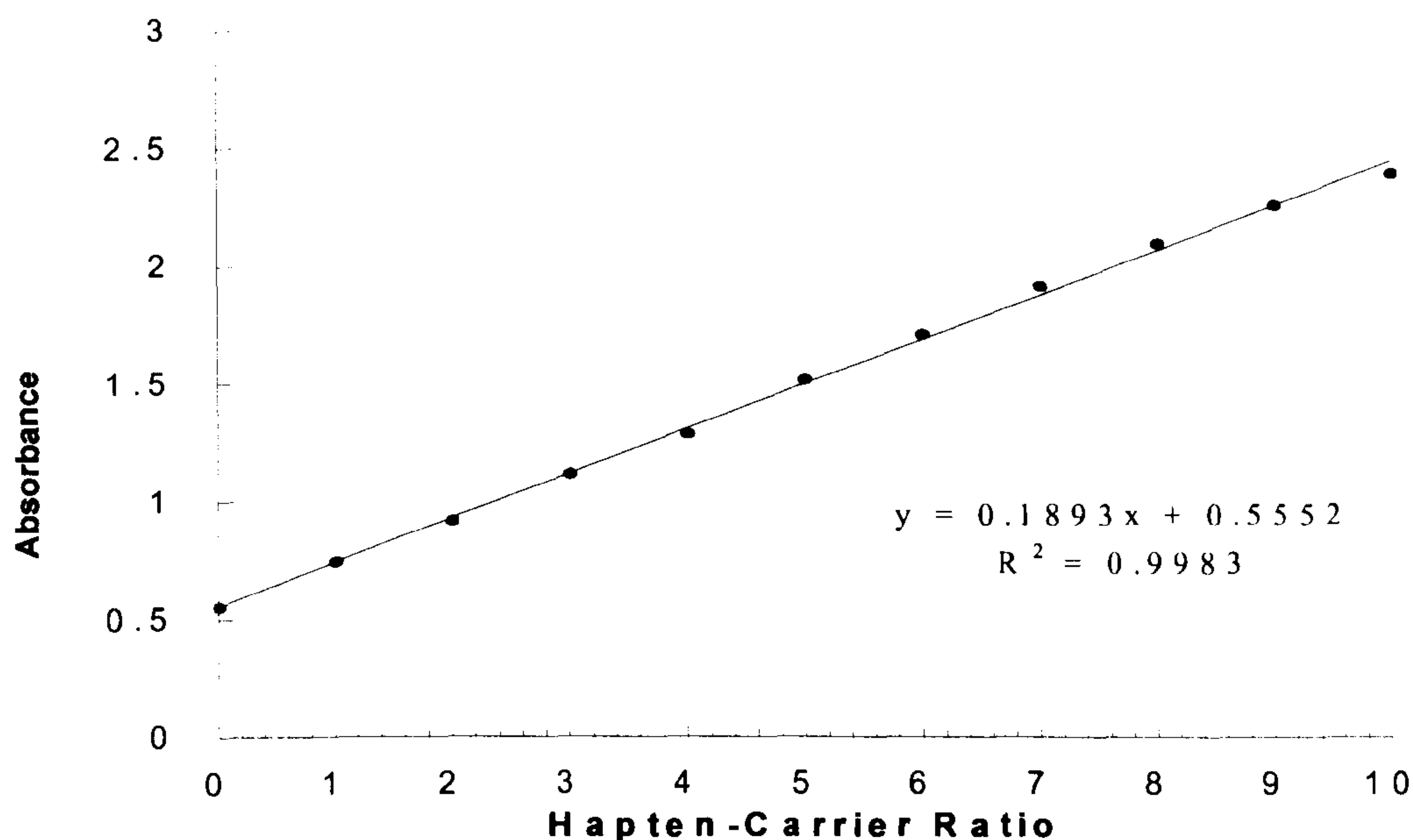


Figure 3-3. A standard curve showing the relationship between the UV absorbance, A_{280} , and the number of haptens present on a carrier protein.

3.2.2.3 Preparation of O^6 -carboxymethyl guanine bound to glass

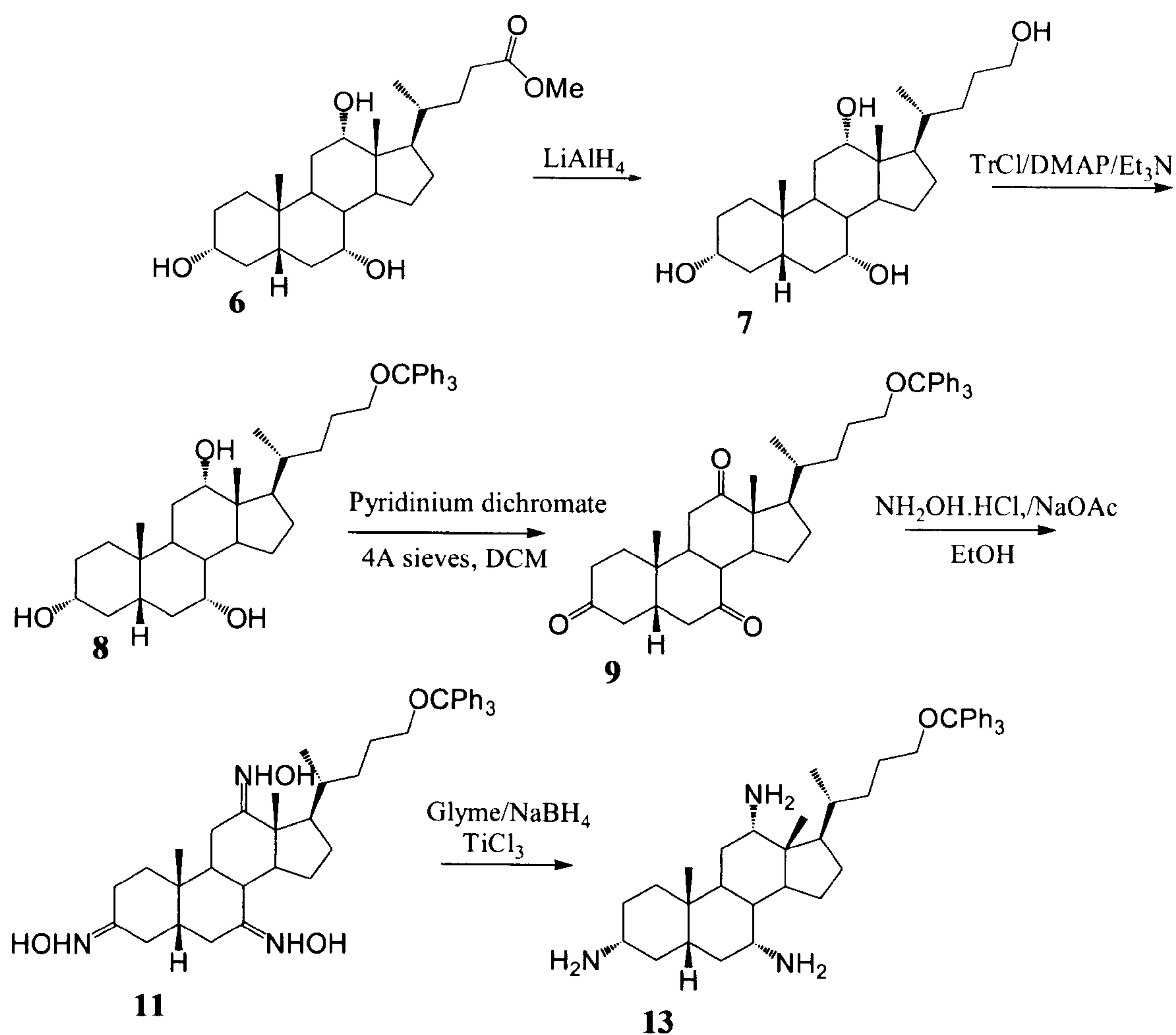
O^6 CMG was also linked to amino propyl glass beads that were used in phage display as previously discussed in section 2.6.3. The same technique was used in the preparation of affinity columns that were used in probing the library and is discussed in detail in section 4.3.

3.3 Synthesis and characterisation of cholate-amino acid libraries

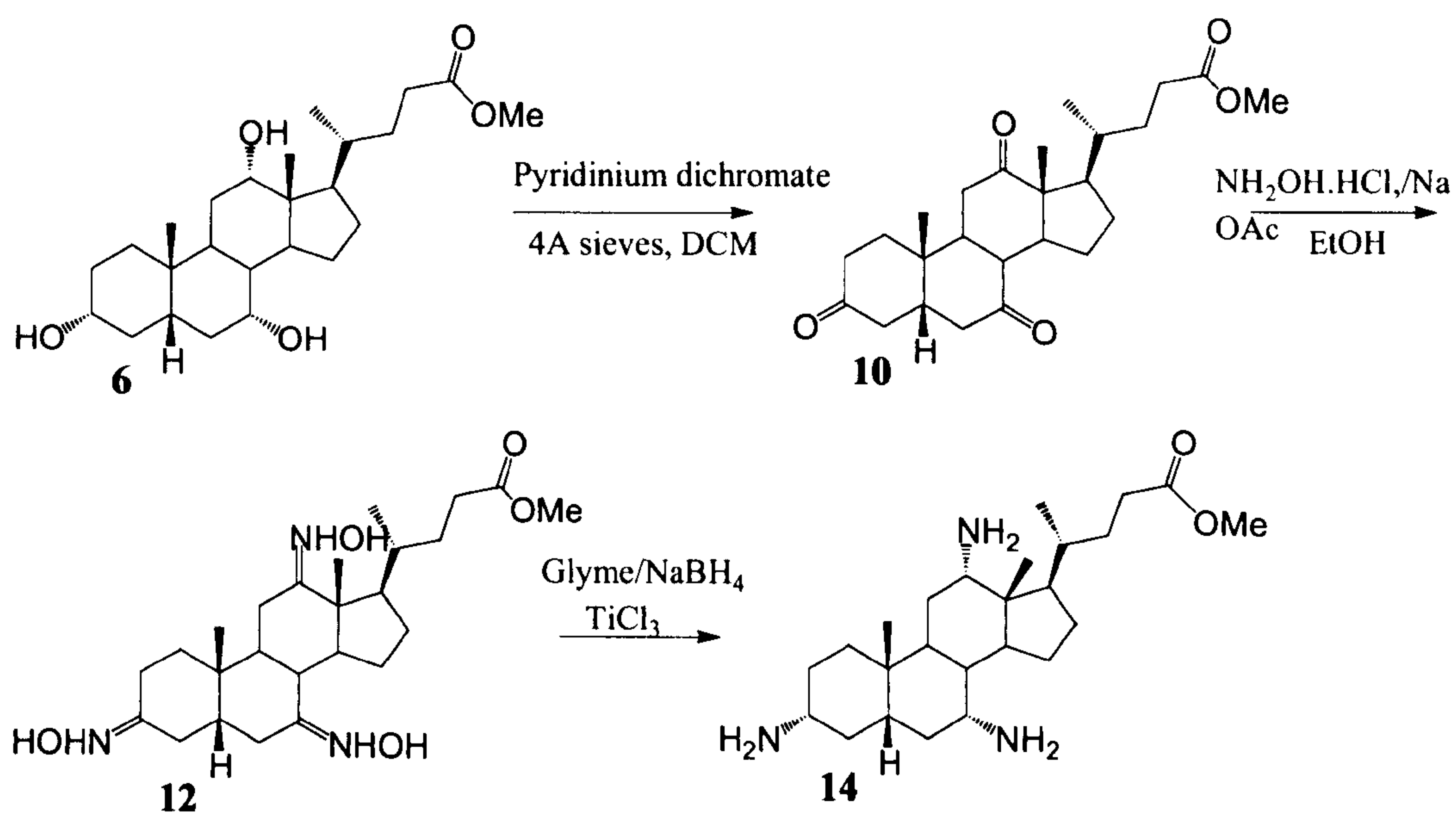
3.3.1 The development of amide bonded cholic acid derivatives

The original plan was to follow the scheme shown in Scheme 3.2. The first two steps were taken directly from Li *et al*¹⁴¹ and although the literature¹⁴²⁻¹⁴⁴ suggests that it would be possible to selectively address each hydroxyl site individually it was decided that this would not be necessary in the initial construction of a library of compounds. Cholic acid, 3 α ,7 α ,12 α trihydroxy-(5 β)-cholan-24-oic acid, was supplied by Aldrich and converted into methyl cholate. Methylation was achieved simply with methanol and a catalytic amount on concentrated sulphuric acid in almost quantitative yield to give 3 α ,7 α ,12 α trihydroxy-(5 β)-cholan-24-oic acid methyl ester, also known as methyl cholate (**6**). The C24 ester is reduced with lithium aluminium hydride, to give 3 α ,7 α ,12 α trihydroxy-(5 β)-cholan-24-ol (**7**) (Scheme 3.2). The subsequent reactions involving trityl chloride (**8**, **9**, **11** and **13**) proved to be difficult to take to completion in a satisfactory manner and the lowest yielding step was the tritylation. It was decided to miss out this step and the associated LiAlH₄ reduction, as Davis and his co-workers had made the triamino analogue without it¹⁴².

Pyridinium dichromate was found to oxidize all three hydroxyl groups on methyl cholate to ketones to give 3,7,12 trioxo-(5 β)-cholan-24-oic acid methyl ester (**10**) in very good yield (Scheme 3.3). In turn this was treated with hydroxylamine hydrochloride to give 3,7,12 trioximo-(5 β)-cholan-24-oic acid methyl ester (**12**), again in good yield. However the reduction of the oxime groups by the addition of sodium borohydride and titanium chloride in 1,2-dimethoxyethane could not be repeated in this laboratory with success.



Scheme 3-2. The proposed scheme for the formation of a triamino analogue of methyl cholate.

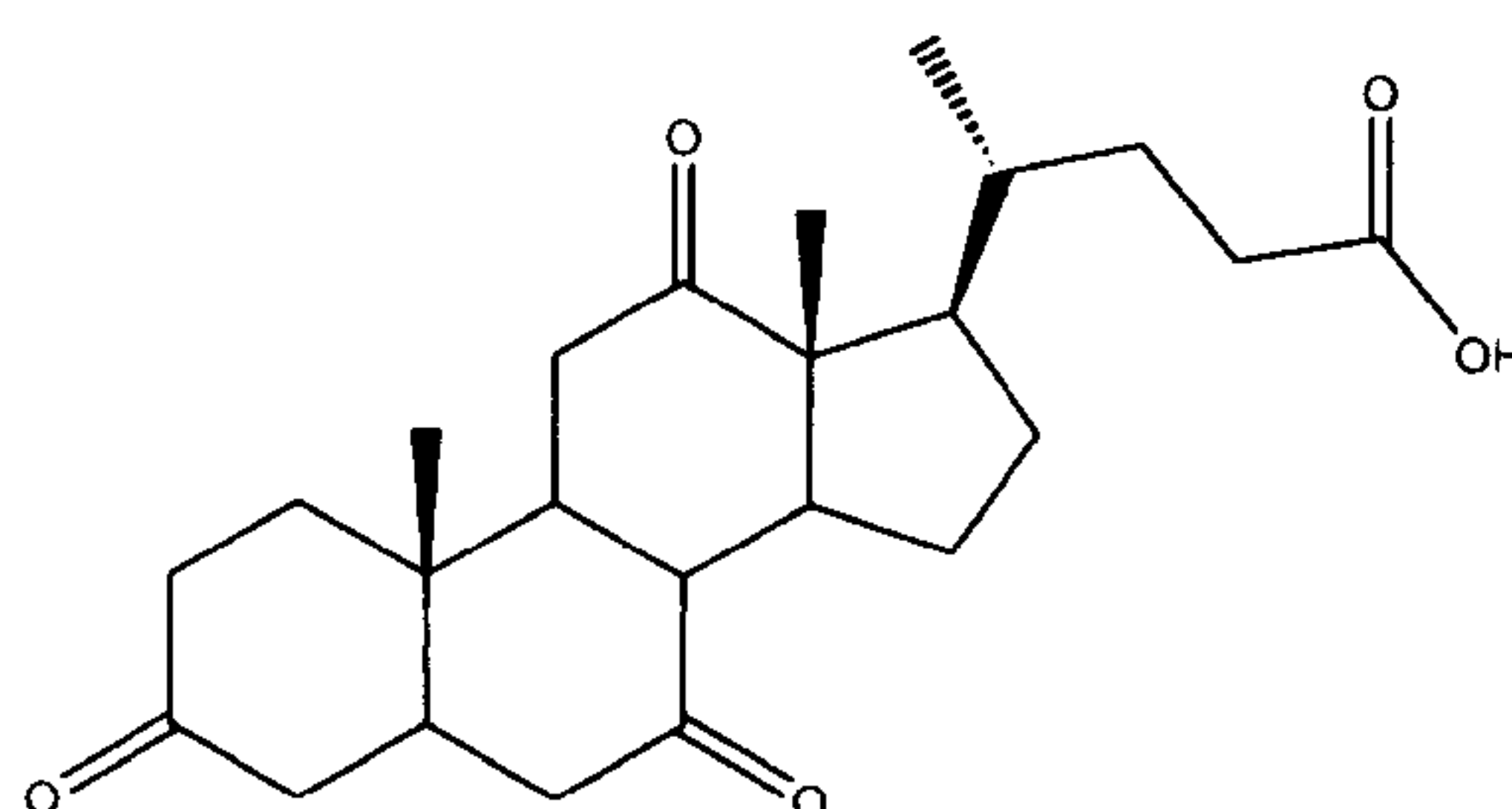


Scheme 3-3. The revised scheme for the formation of a triamino analogue of methyl cholate.

However, the facility by which trioximo derivatives of trioxocholate could be prepared suggested a more convenient route to combinatorial libraries based on cholic acid. This approach became the focus of subsequent work.

3.3.2 The development of a general synthesis of oxime bonded cholic acid derivatives

A commercial laboratory in the USA was located who could supply steroid derivatives for research purposes and 500 g of 3,7,12 trioxo-(5 β)-cholan-24-oic acid (**15**), for approximately £55, was obtained. This compound became the starting point of subsequent reactions. The opportunity was taken at this point to incorporate the marker that would aid subsequent detection.



3,7,12 trioxo-(5 β)-cholan-24-oic acid (15)

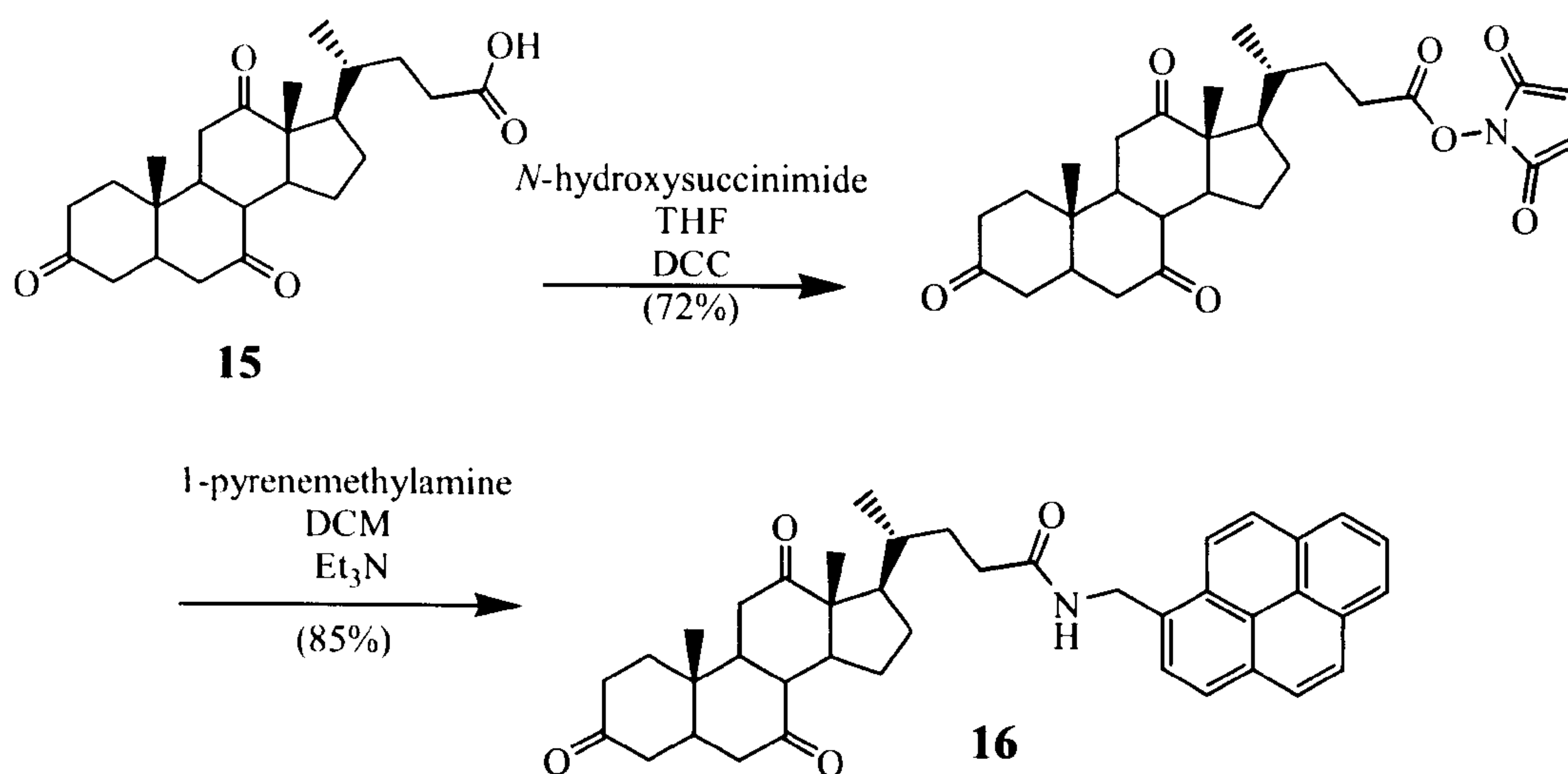
3.3.2.1 Incorporation of a fluorescent marker

Fluorescence was considered to be a readily measurable and sensitive characteristic that would increase the detection potential of the ultimate reagent. Fluorescence spectra are a plot of fluorescent intensity against wavelength¹⁴⁵. Light absorbed by the molecule is of a specific wavelength with discrete photon energies, these induce electronic excitation and molecular vibrations which make the molecule thermodynamically less stable relative to its surroundings. When the energy is lost via a $S_1 \rightarrow S_0$ transition, from a high energy state to one of a low vibrational level, the energy is released in the form of light, fluorescence¹⁴⁶. Fluorophores, substances that are fluorescent, usually possess delocalized electrons present in conjugated double

bonds¹⁴⁵ and it was decided that pyrene would be a suitable molecule to introduce fluorescence into the reagent.

Initial attempts to introduce pyrene into the cholate molecule were based on the synthesis of pyrene methyl esters with the C24 carboxylic acid group¹⁴⁷. Based on published methods, 1-pyrenyldiazomethane was prepared from 1-pyrenecarboxyaldehyde hydrazone and reacted with cholic acid to give the required ester^{148,149}. However this approach was abandoned due to low yields of the ester as well as concerns about the stability of the pyrenyl ester linkage. It was decided that a better approach would be to explore an amide linkage.

The *N*-hydroxysuccinimide ester of cholic acid has been used in the preparation of bile acid-based Gemini surfactants¹⁵⁰ and this method was followed using **15** (3,7,12 trioxo-(5 β)-cholan-24-oic acid). This product was coupled with 1-pyrenylmethylamine to give the pyrenyl amide of 3,7,12 trioxo-(5 β)-cholanoic acid (PyCA), (Figure 3.6 compound **16**), by the proposed method shown in scheme 3.3 with overall yield of >60%. This gave a robust, highly detectable, scaffold molecule upon which a library could be based.



Scheme 3-4. The scheme that was followed during the synthesis of *N*-(1'-pyrenylmethyl)- 3,7,12-trioxo-(5 β)-cholan-24-amide (PyCA) Compound 16.

3.3.2.2 The incorporation of a spacer unit

Simple modelling of the scaffold molecule, using ball and stick, showed that there may be considerable restriction around the putative binding pocket if amino acids were bound directly to amino groups formed by possible reduction of the trioxime intermediate. This can be seen in Figure 3.4, a WebLab Viewer™ representation. WebLab Viewer™ is a basic programme that allows ball and stick models to be viewed on a computer. Figure 3.4 is a model of the scaffold molecule (the lower molecule) with the pyrene appendage shown in the lower right hand corner. The upper molecule is that of ligand with the carboxymethyl adduct being presented to the 'pocket' formed by amide linked glycine amino acids.

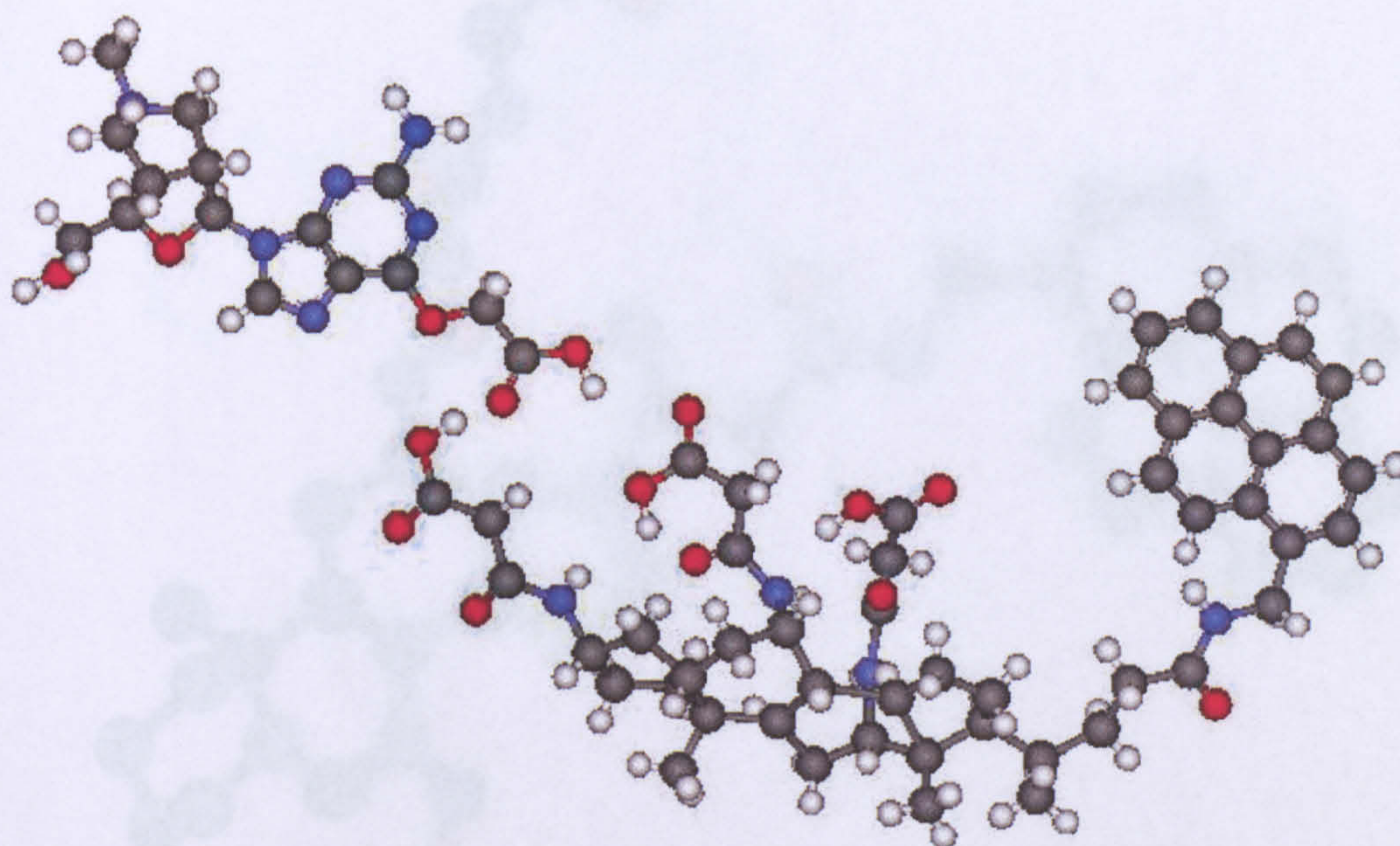


Figure 3-4. A WebLab Viewer™ representation of the ligand, the upper left hand molecule, which has the carboxymethyl adduct close to the binding portion of the scaffold molecule (the lower molecule). This molecule contains three glycine amino acids that have been attached to the scaffold by amide bonds.

By replacing the amide linked glycine with aminooxy acetic acid derivatives the oxime bond opens out the pocket (represented in Figure 3.5). Thus, by introducing diversity through amide bond formation with the introduced aminooxyacetyl function the putative binding pocket is likely to be more accommodating to O⁶-carboxymethyl-2'-deoxyguanosine (Figure 3.6).

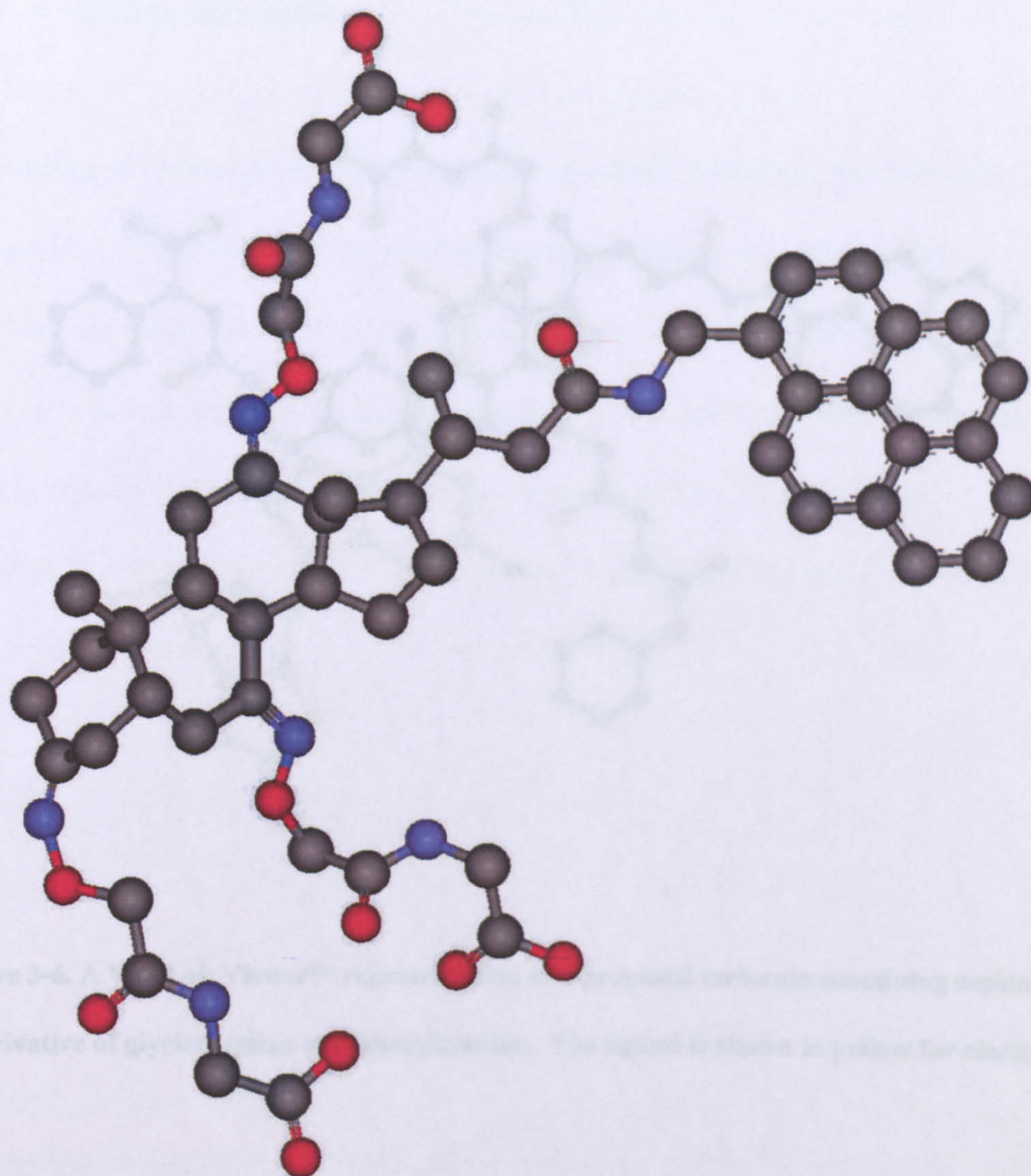


Figure 3-5. A WebLab™ representation of the scaffold with three oxime linked aminoxy glycine derivatives.

3.3.3 Oxime formation

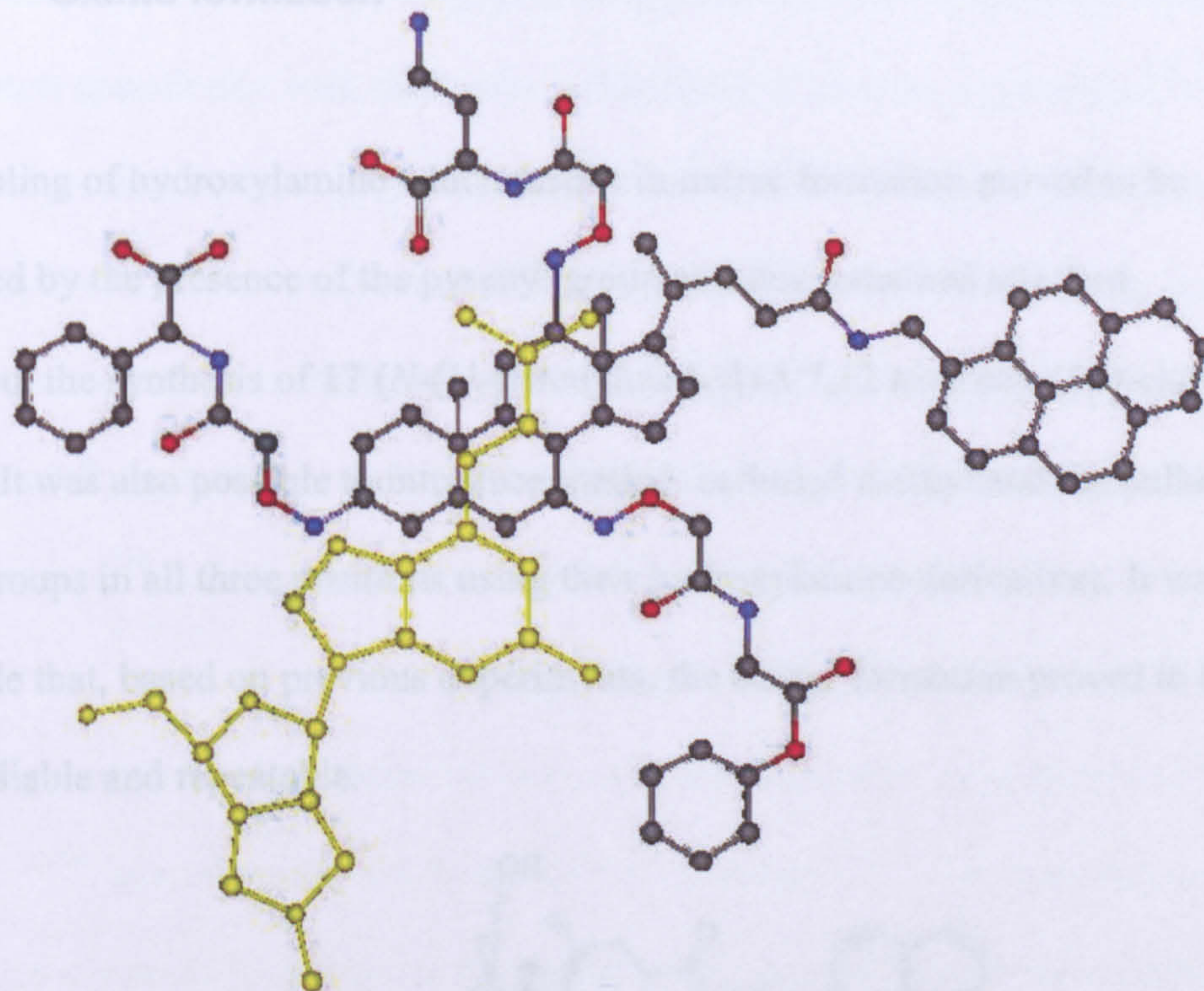
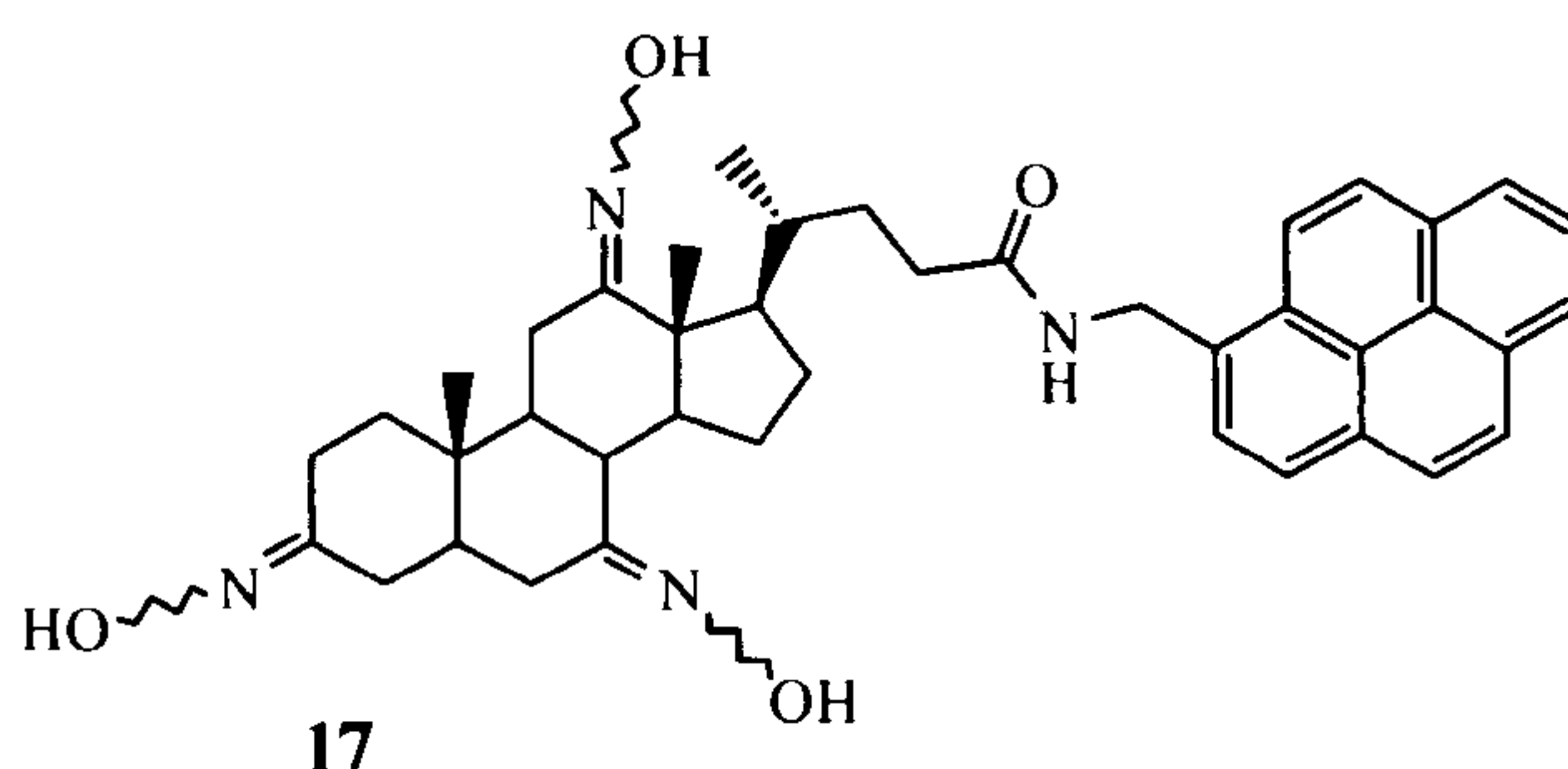


Figure 3-6. A WebLab Viewer™ representation of a proposed molecule containing aminoxy derivative of glycine, lysine and phenylalanine. The ligand is shown in yellow for clarity.

3.3.3 Oxime formation

The coupling of hydroxylamine which results in oxime formation proved to be unaffected by the presence of the pyrenyl group and this remained attached throughout the synthesis of **17** (*N*-(1'-pyrenylmethyl)-3,7,12 trioximo-(5 β)-cholan-24-amide). It was also possible to introduce methyl, carboxyl methyl and the bulkier benzyl groups in all three positions using their hydroxylamine derivatives. It was also noticeable that, based on previous experiments, the oxime formation proved to be facile, reliable and repeatable.



The coupling of two mutually and uniquely reactive functional groups in an aqueous environment has been given the term 'chemoselective ligation' and such reactions have been explored by both biologists and chemists in their search for selective chemical couplings that may occur in living organisms¹⁵¹. Traditional stepwise syntheses are limited by by-product accumulation particularly with regards to peptide synthesis. Chemoselective ligation is considered to be a way of bypassing some of the synthetic steps and thus improve yield. Many of the chemoselective ligation partners have been identified, some of which were discussed in section 1.8.2. One set of partners were the (aminoxy)acetyl functionality and its reaction with ketones or aldehydes.

The literature shows that under very mild conditions the (aminoxy)acetyl terminal group reacts specifically with aldehydes and ketones to form an oxime bond^{71,104,152-161}, in a way that is rapid, highly efficient and stable¹⁶². This ligation has been used in a whole host of different purposes and the compounds that result from these couplings have been shown to have inhibitory potential of Na⁺, K⁺-ATPase¹⁶³, *in vivo* applications¹⁶⁰, protein dimerisation¹⁶¹ and template-assembled synthetic protein (TASP) design¹⁵⁸, to name but a few.

It is also known that oxime formation occurs in the presence of unprotected side chain residues^{153,156}, guanidino, amino, hydroxyl and carboxylic¹⁵⁸. This is probably due to the physico-chemical properties of the aminoxy group, the oxygen atom affects the nucleophilic character of the neighbouring amino group, the α -effect¹⁵⁴.

3.3.3.1 The α -effect

The α -effect is thought to come about because the adjacent electron pair can act to stabilize any charge deficiencies at the transition state¹⁶⁴, or because they raise the highest occupied molecular orbital (HOMO) energy that participates in bond formation and enhances 'soft' base behaviour¹⁶⁵. However the α -effect is brought about, atoms which are directly bonded to another with unshared electron pairs tend to be stronger nucleophiles¹⁶⁶, for example, hydroxylamines (NH₂OH) and hydrazines (NH₂NH₂) are both more nucleophilic than ammonia¹⁶⁴.

3.3.4 Oxime bond formation

An oxime is the condensation product of hydroxylamine and an aldehyde or a ketone, which can be divided into two distinct stages, of addition followed by elimination proceeding through a tetrahedral intermediate, shown in Figure 3.7:

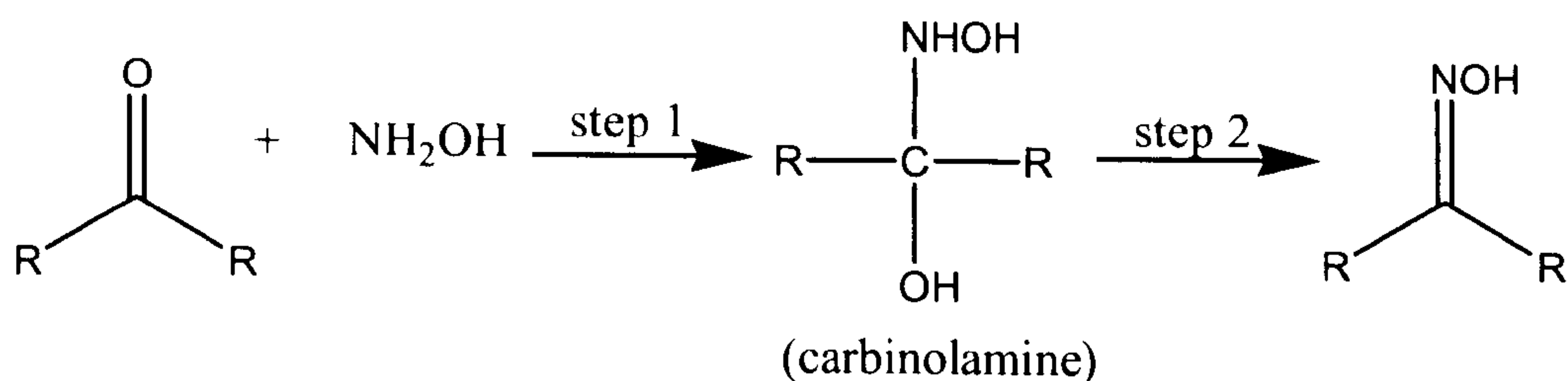


Figure 3-5. The two step reaction that results in oxime formation, via a tetrahedral intermediate.

There is an optimum pH at which the rate of reaction is at its maximum, at a pH higher or lower than this the rate falls. This is caused by changes in the rate determining step. At a pH lower than the optimum step 2 is rapid and step 1 is the slow rate determining step. Most of the hydroxylamine molecules have been converted to the conjugated $+\text{NH}_3\text{OH}$ which cannot attack the substrate. As the pH increases, the fraction of free hydroxylamine molecules increases and the rate of step 1 increases.

As the pH increases beyond optimal step 2 becomes the rate determining. Although the rate of step 1 is still increasing it is now step 2 which determines the overall rate as it is the rate of this step that is decreased in increasing basic conditions.

Following the addition of the nucleophile the carbinolamine intermediate is formed, dehydration of this intermediate involves nitrogen deprotonation concerted with the elimination of water.

A general method of oxime formation that was repeatable and reliable was developed. This was based on the oxime formation that was used to connect testosterone, containing one cyclic ketone, to varying alkylaminoxy compounds¹⁵⁵. The method

involved dissolving the reactants in ethanol, with sodium acetate to buffer the reaction. Therefore it was decided to develop a general coupling process based on these reactions that will bring about the formation of an oxime bond between the tagged scaffold molecule and suitable compounds to introduce diversity around this scaffold.

3.3.5 The scaffold molecule

The framework of a novel reagent was fully prepared which had an inherently rigid backbone with an attached fluorescent marker molecule. The presence of three cyclic ketones allowed for the chemoselective ligation of aminoxy derivatives that formed the stable oxime bond at all three sites. The next consideration was the attachment of diversity at these sites, synthesising a library of differing compounds.

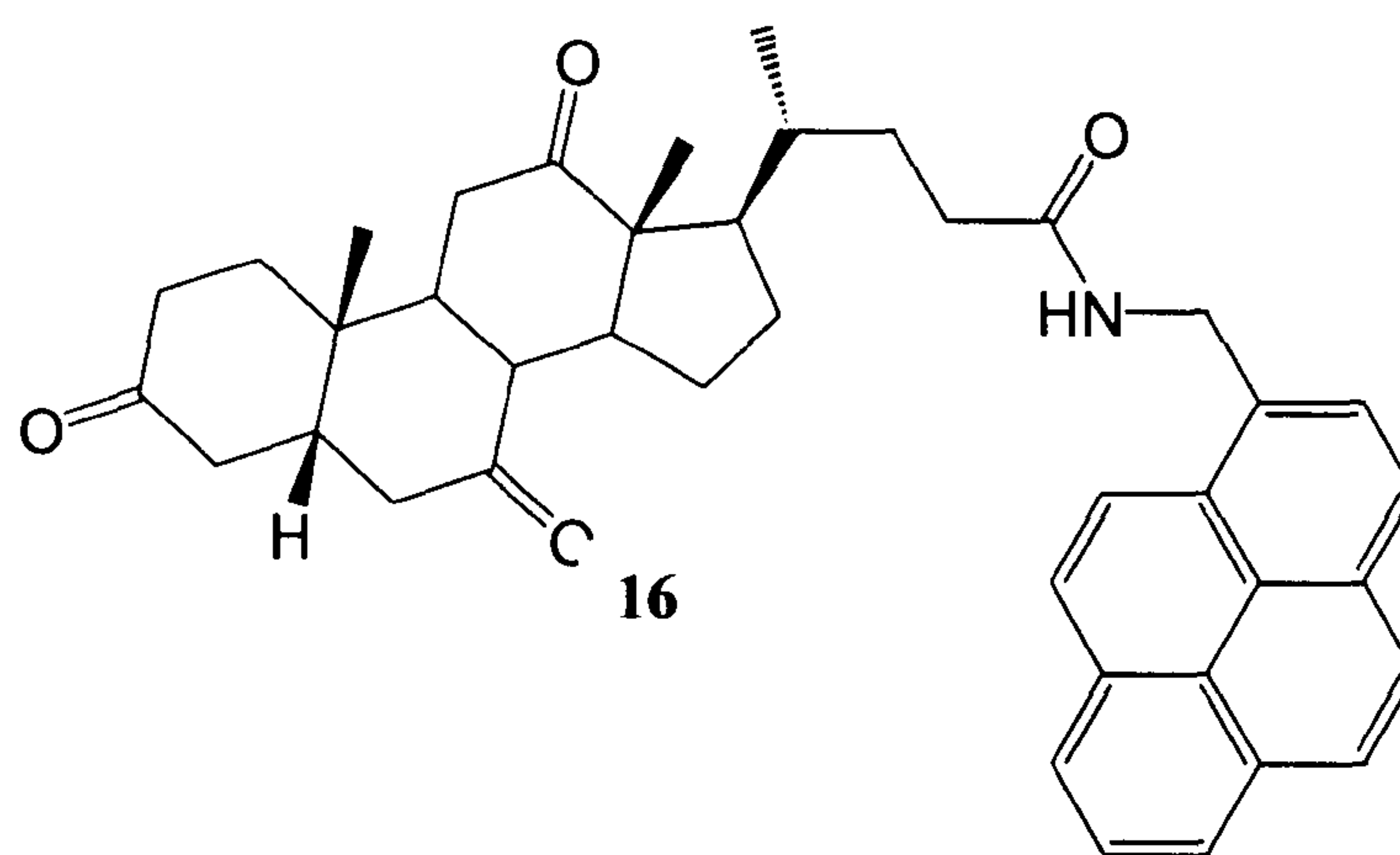
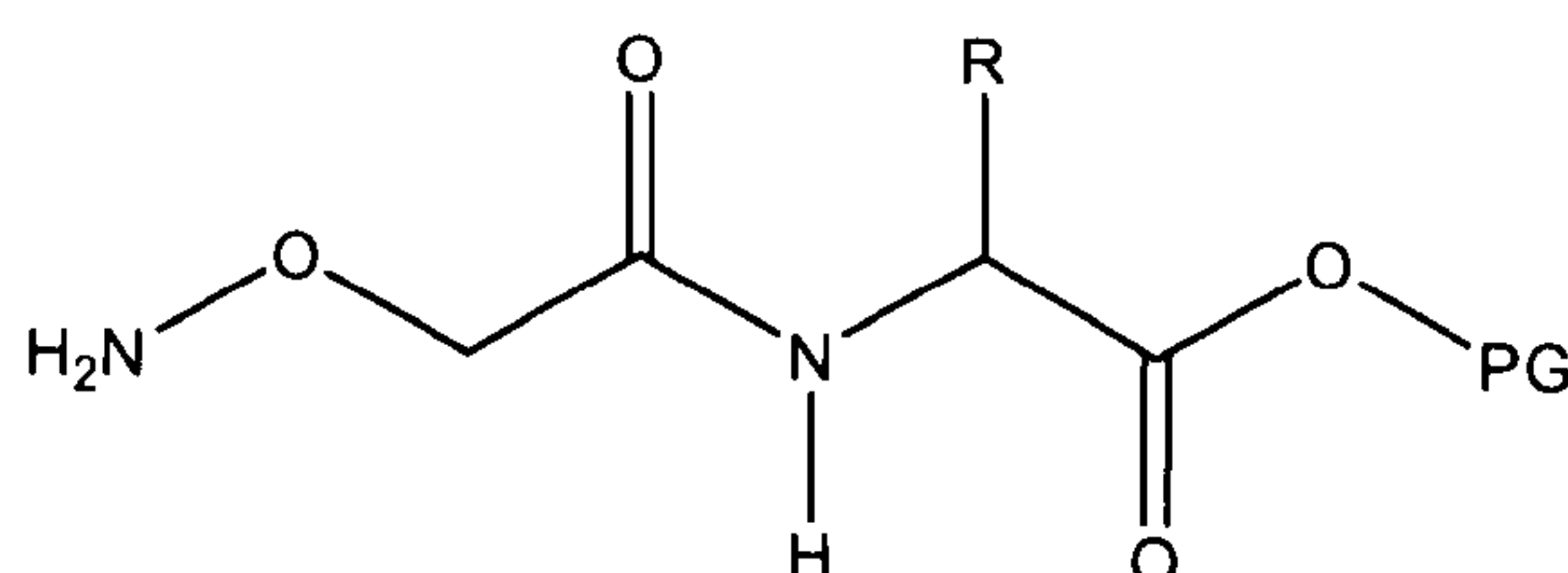


Figure 3-6. Compound 16. 1'-pyrenylmethyl-3,7,12 trioximo-(5β)-cholan-24-amide (PyCA). The molecule that became the scaffold of the library, it has three ketone groups on a rigid backbone with a remote, highly fluorescent component.

3.4 Synthesis of library components

When it was considered that the oxime bond would allow for slightly more flexibility and there was a requirement for spacers, this suggested a molecule such as:

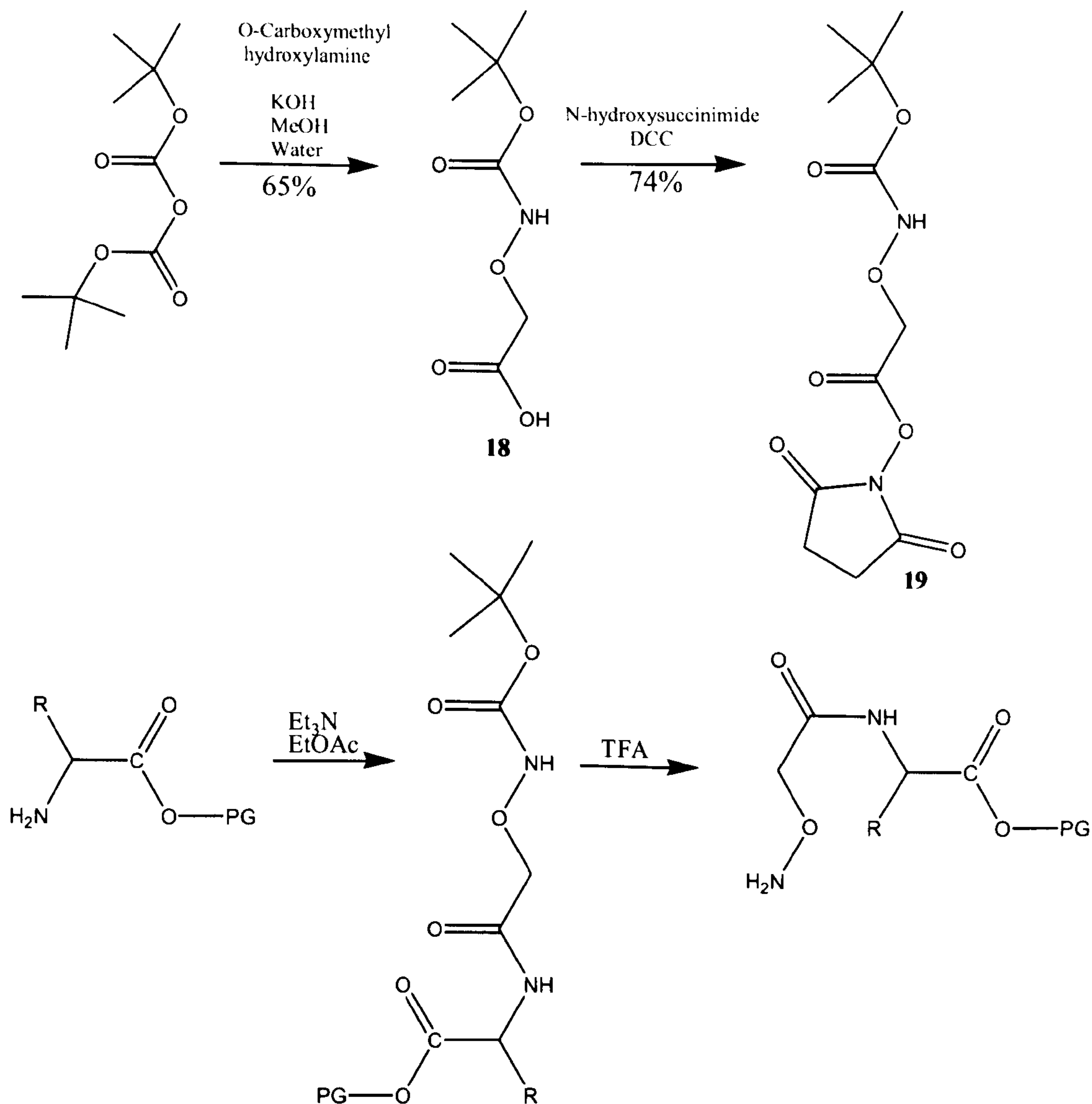


in which the hydroxylamine functionality would be available after the attachment of suitably protected amino acids. It was decided that it would be possible to make a range of compounds based on the above molecule that would be used to couple to the scaffold molecule. An advantage of this method would be that the coupling process would not be influenced by the side chain amino acid residues as they are situated somewhat remotely from the coupling site. A range of (aminoxy)acetyl derivatives of amino acids were therefore synthesised.

3.4.1 Synthesis of aminoxy compounds

((1,1-dimethylethoxycarbonyl)aminoxy ethanoic acid (**18**, Scheme 3.5) was first used in the formation of *O*-alkyl aminoxy ethers in 2000, with alkyl halides¹⁶⁷ and more specifically with regard to peptides¹⁵². Succinimido-(*N*)-((1,1-dimethylethoxycarbonyl)aminoxy) ethanoate (**19**) was prepared by the method previously published¹⁶⁸. This was then used to couple with various protected amino acids according to Scheme 3.4, including glycine benzyl ester, glycine methyl ester, phenylalanine methyl ester, lysine (Fmoc) methyl ester and arginine (ϵ -NO₂) methyl ester. These aminoxy derivatives of amino acids were considered to be the most

desirable compounds to introduce diversity to the scaffold molecule because they covered a range of side chain structures thought to be critical for forming a binding site for O⁶-carboxymethyl-2'-deoxyguanosine. The products were mainly white fluffy powders which were hygroscopic and had to be stored under vacuum.



Scheme 3-5. The Scheme that was followed during the preparation of aminoxy compounds that were used to couple to the scaffold molecule. R represents the various side chain residues for the different amino acids. PG stands for protecting group which was either benzyl or methyl.

The yields of these reactions (Table 1) were modest and although some effort was spent trying to improve them, there was little success in doing this. The amino acids were freed from their salts before coupling to succinimido-(*N*)-((1,1-dimethylethoxy)carbonyl)aminoxy ethanoate (**19**), it was later found that this was not necessary and the addition of triethylamine to the coupling reaction freed the amine *in situ*, with no difference in yield between the two methods. Although small quantities of the required amino acid precursors were made, with other amino acids, serine and alanine, the yield was too low to be useful. It has been noted previously¹⁶⁷ that the synthesis of aminoxy compounds give unsatisfactory yields and the availability of aminoxy precursors are very limited, with only the simplest being commercially available. This seems to be a major limitation and improvement in the synthetic methodologies for such aminoxy compounds is a must before the full potential of this chemistry can be achieved.

Amino acid precursor	Aminoxy compound	Overall yield
Glycine	<p>20 AOA-NH-CH₂-COOCH₃</p>	33%
Glycine	<p>21 AOA-NH-CH₂-COOCH₂C₆H₅</p>	34%
Phenylalanine	<p>22 AOA-NH-CH(CH₂C₆H₅)-COOCH₃</p>	24%
Arginine	<p>23 AOA-NH-CH(CH₂CH₂CH₂NH-C(=NH)-NH₂)-COOCH₃</p>	35%
Lysine	<p>24 AOA-NH-CH(CH₂CH₂CH₂CH₂CH₂NH-Fmoc)-COOCH₃</p>	32%

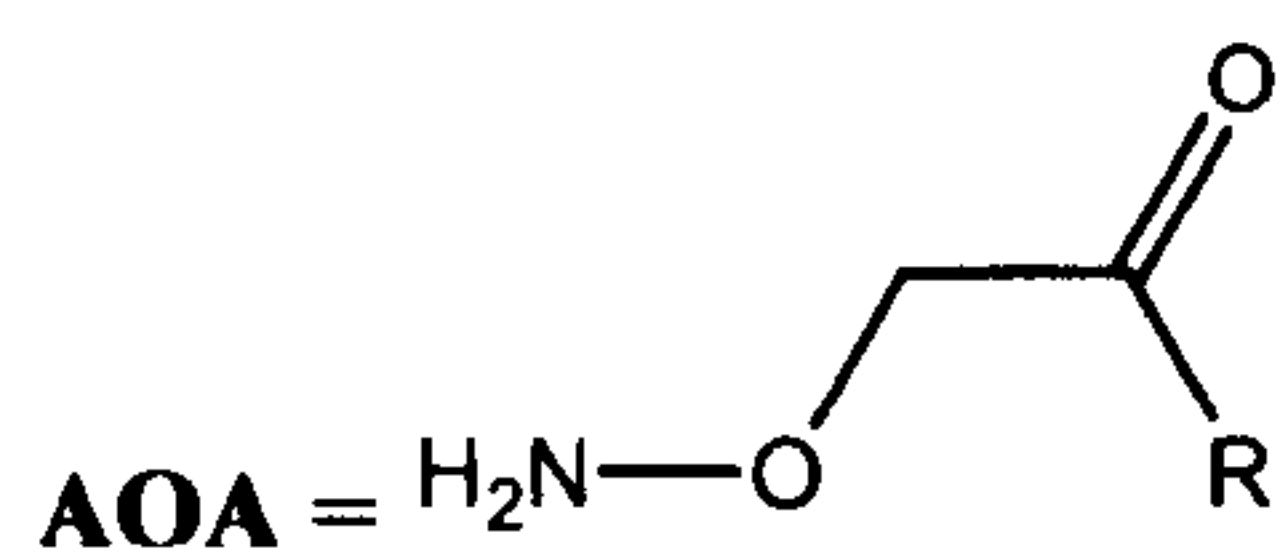


Table 3-1. A table of the aminoxy compounds that were used in the synthesis of the library, showing the precursor amino acid and the structure of the compound with the overall yield.

3.5 Synthesis of tri-substituted 1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide (PyCA) compounds

The aminoxy glycine derivative, which was protected on at the carboxylic acid group by a benzyl group, was coupled to PyCA. This reaction was undertaken in the same manner as with the hydroxylamines, to give a tri-substituted compound. The aminoxy compounds were reacted in a slight excess of 3:1 ratio to ensure that all three sites reacted. The reaction was followed by TLC and was considered to be complete when a single spot was observed on the TLC plate. The benzyl group was then removed by hydrogenolysis; this also removed the pyrenyl group and so another protecting group had to be used. The facile removal of methyl esters by sodium hydroxide proved that these were a suitable alternative and were subsequently used. However, as the aminoxy acetyl glycine benzyl ester had already been synthesised it was decided to incorporate this into the library as well. The benzyl group could act to encourage ring-ring interaction with the O⁶-carboxymethyl-2'-deoxyguanosine analogue. For each of the aminoxy compounds the tri-substituted PyCA molecule was made (shown in Figure 3.7) and, although the molecular ion was not detected, ion fragments of these compounds were clearly seen. The Sakaguchi reagent¹⁶⁹ was used to confirm that the guanidine derivatives had indeed survived the coupling reaction in the case of arginine derivatives. Mass spectrometry was used but in many cases it was difficult to locate a molecular ion. However it was apparent that the oxime bond was susceptible to cleavage and the fragments corresponded to the loss of an 'arm' on the scaffold with the break occurring between the N and the O atoms in the bond.

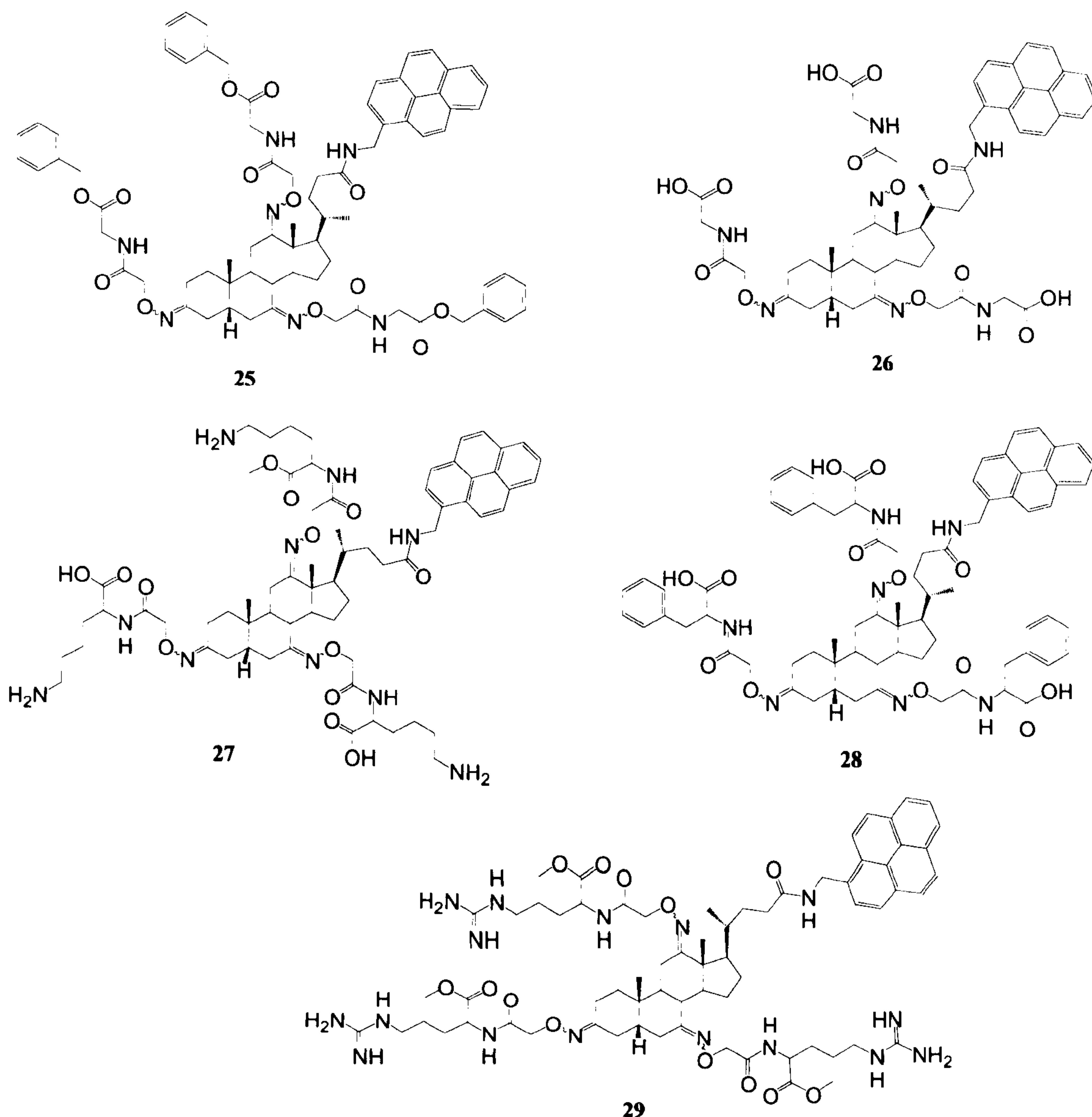


Figure 3.7. Digrams of the structure of the tri-substituted PyCA compounds made with each of the aminoxy derivatives of four different amino acids. Compound 25, 1'-pyrenyl methyl-3,7,12 tri(N-inoxymethylene carbonyl) (glycine benzyl ester). Compound 26, 1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylene carbonyl) glycine-(5 β)-cholan-24-amide. Compound 27, 1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylene carbonyl) lysine-(5 β)-cholan-24-amide. Compound 28, 1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylene carbonyl) phenylalanine-(5 β)-cholan-24-amide. Compound 29, 1'-pyrenylmethyl-3,7,12(tri(N-inoxymethylene carbonyl) arginine)-(5 β)-cholan-24-amide.

3.5.1 Pilot studies on library formation.

3.5.1.1 PyCA Mix 1

In order to demonstrate that structurally diverse aminoxy derivatives could react *O*-carboxymethyl and *O*-benzyl (to give products represented in Figure 3.8) were reacted in equal molar concentrations and coupled to PyCA in the way previously discussed. Using 5% methanol and 0.5% diethylamine in chloroform a TLC showed that the starting material PyCA had an *rf* value of 0.63, the trisubstituted carboxyl methyl compound *rf* = 0.04, and the tri benzyl substituted compound *rf* = 0.79. The mixed reaction gave eight spots that corresponded to *rf* = 0.79, 0.46, 0.41, 0.36, 0.30, 0.22, 0.13 and 0.03, shown diagrammatically in Figure 3.9. This clearly shows that a range of compounds were formed with a mix of characteristics between the fully tri substituted compounds, it also indicated that there was no remaining starting material.

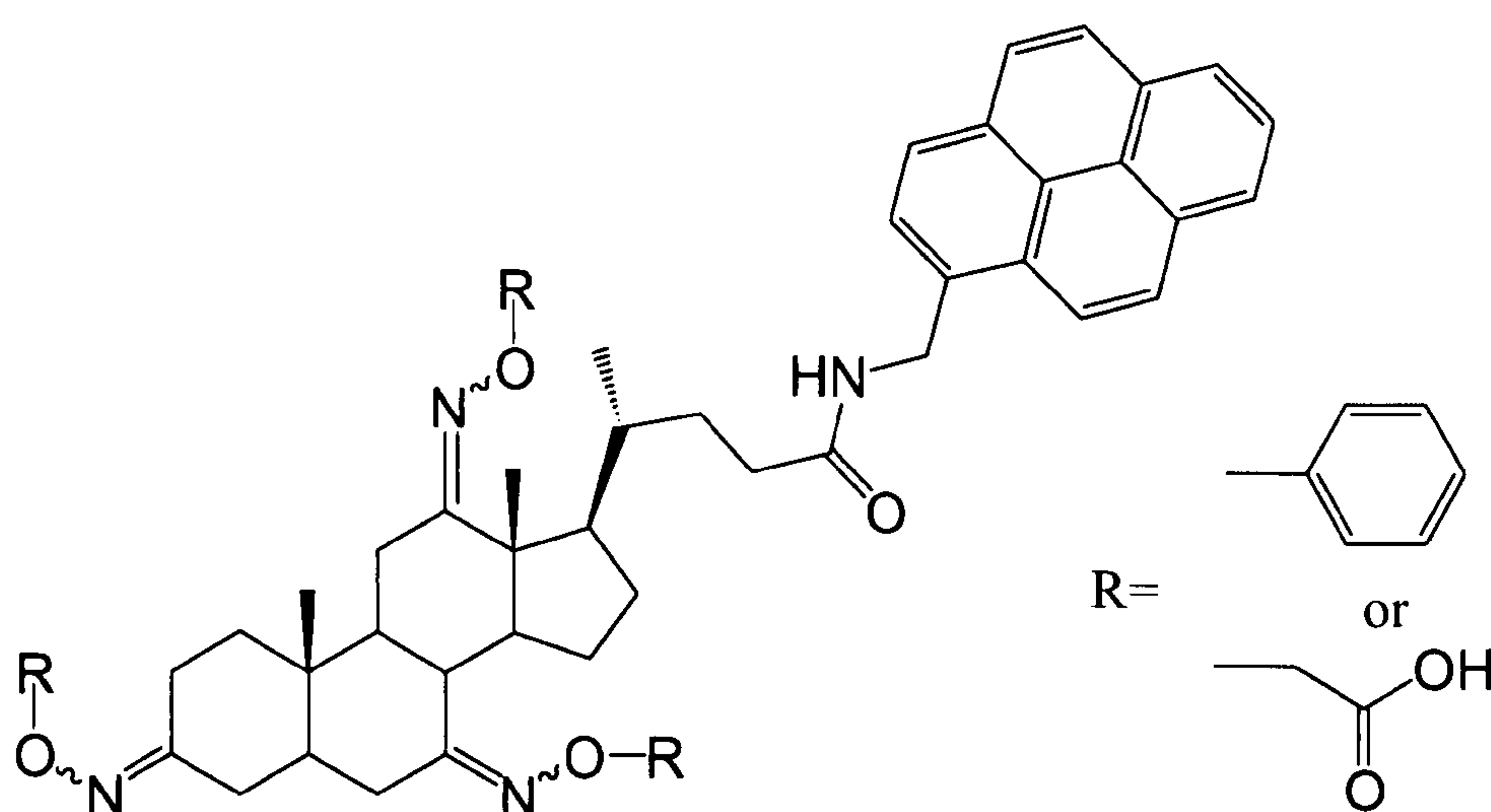


Figure 3.8. A diagram showing the possible products from the PyCA Mix 1 reaction.

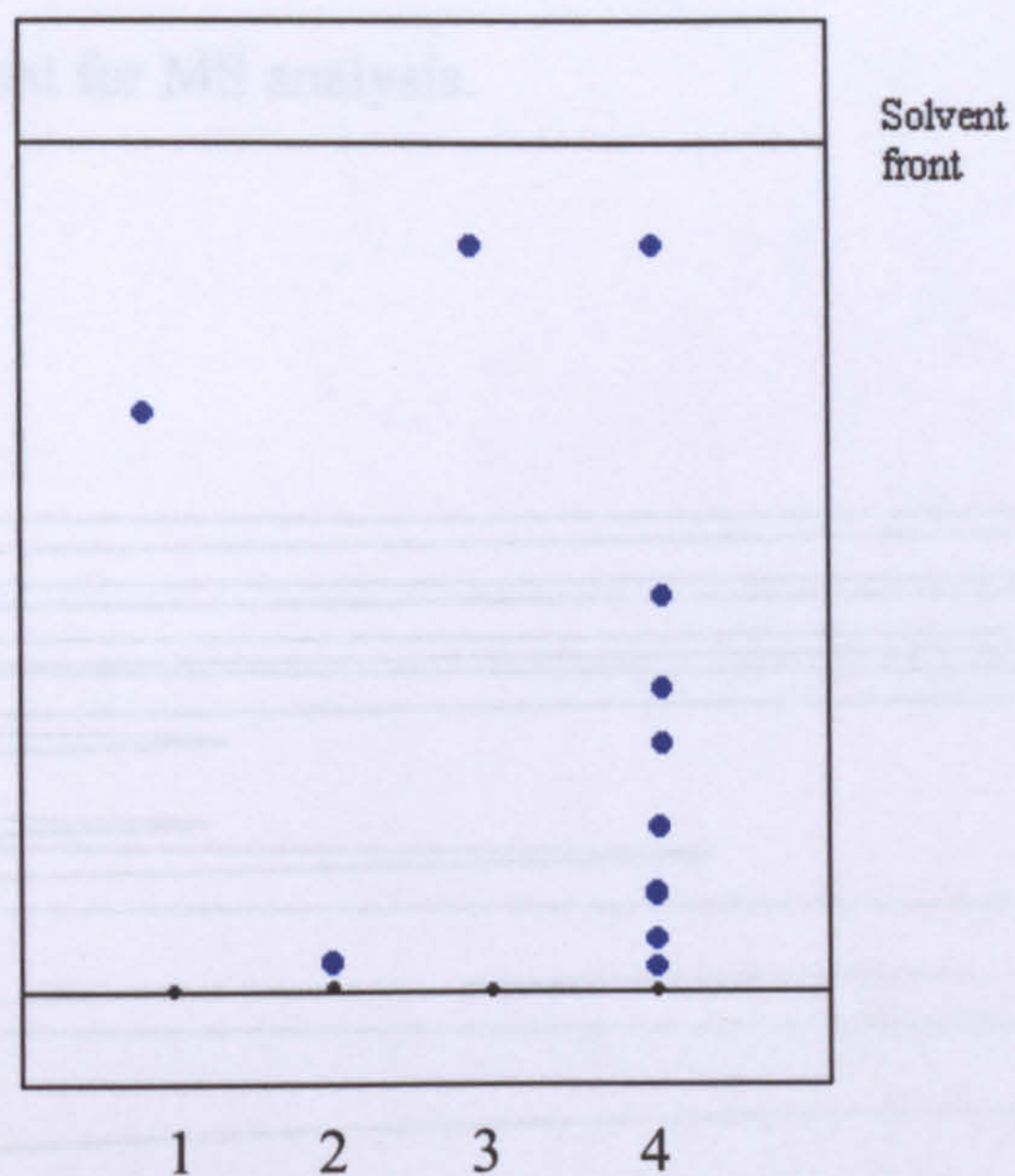


Figure 3-9 A diagram to represent the TLC plate of the first pilot study to show that a mixture of compounds are found in the presence of different aminoxy compounds. 1 indicates the position of the PyCA starting material, 2 indicates the position of the tri-carboxymethyl substituted PyCA compound and 3 indicates the position of the tri-benzyl substituted PyCA compound. 4 is the range of spots produced as a result of reacting both aminoxy compounds with PyCA.

The NMR of the mix show a multiplet centred around 8 ppm (pyrene), a multiplet centred around 1 ppm (steroidal core) as well as evidence for benzyl protons and carbons and also carbonyl carbons. Varian normal phase HPLC using Jones Apex II silica 5 micron column at a flow rate of 1 mL/min of 50/50 methanol/dichloromethane was used to attempt to separate the compounds in the mixture which was monitored at 342 nm, resulting in the trace in Figure 3.10. Although peaks were present with a retention time of 3.5 mins and 4.6 mins, these did not coincide with the starting

material as on the same system PyCA had a retention time of 3.95 mins. The peaks 1-5 were collected and sent for MS analysis.

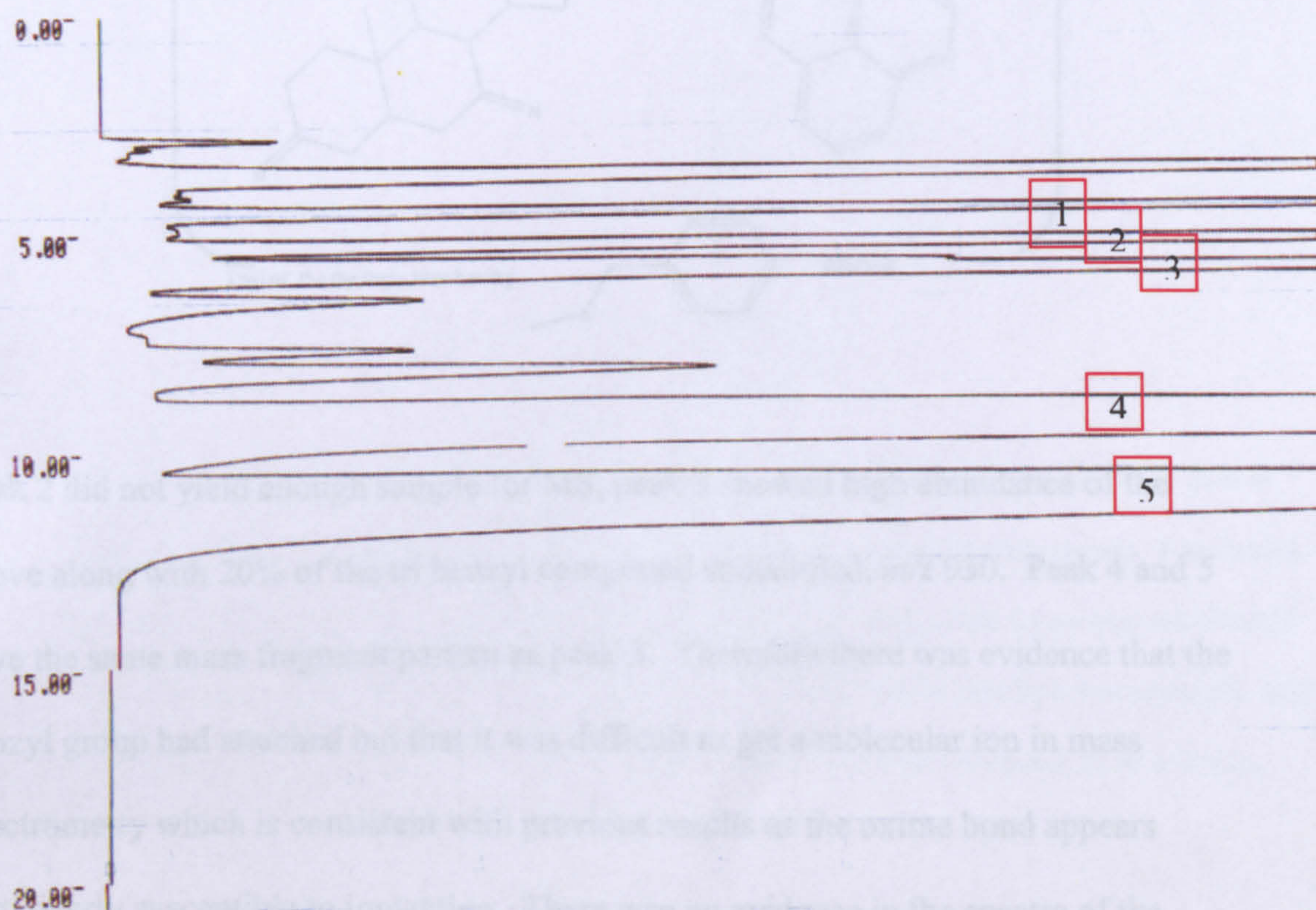
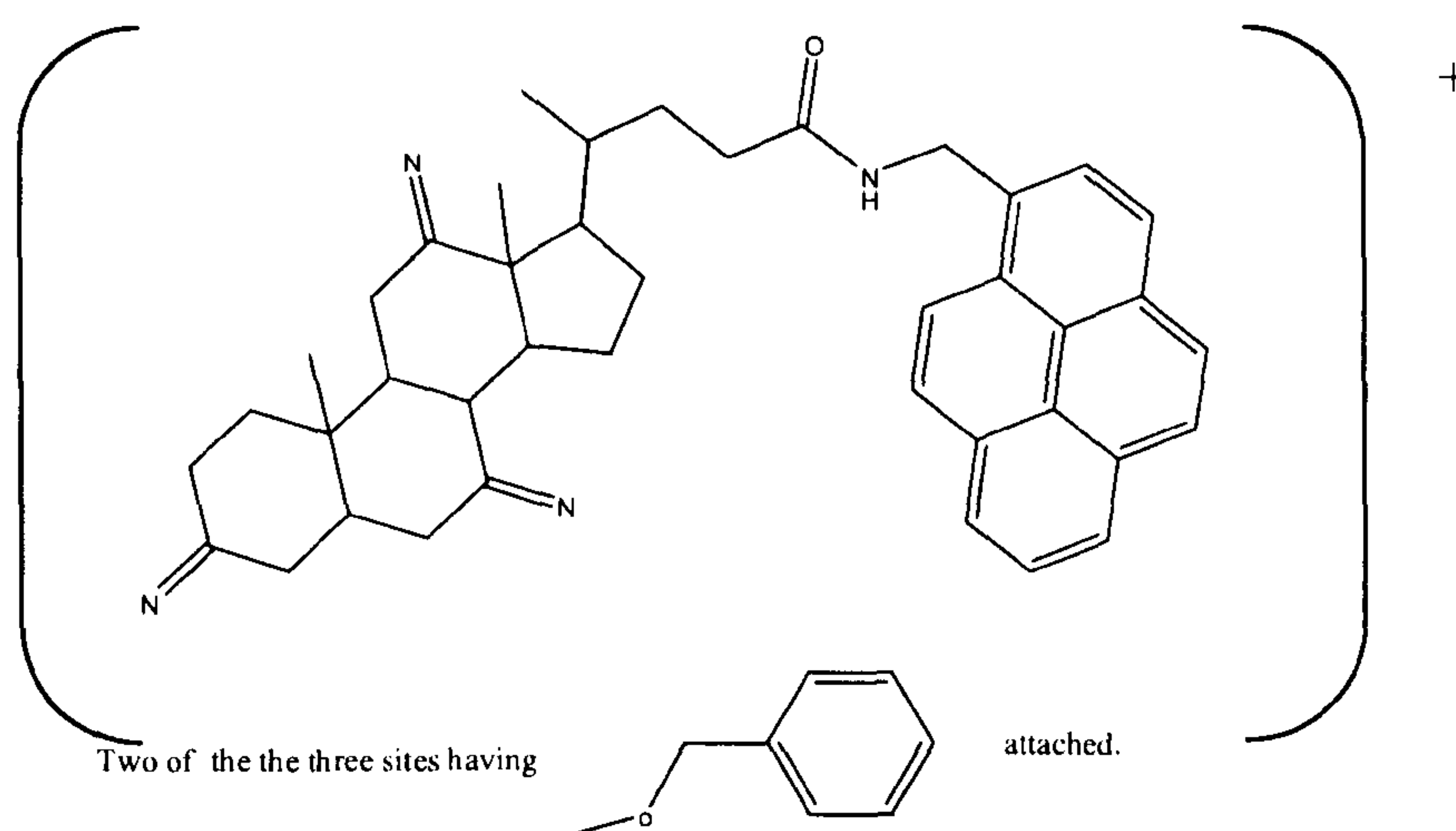


Figure 3-10. RP-HPLC trace of the mixed PyCA compounds found in the reaction mixture. The peaks 1-5 were collected and sent for MS.

Peak 1 showed 100% abundance of a fragment m/z 826 that coincided with:



Peak 2 did not yield enough sample for MS, peak 3 showed high abundance of the above along with 20% of the tri benzyl compound substituted, m/z 930. Peak 4 and 5 gave the same mass fragment pattern as peak 3. Therefore there was evidence that the benzyl group had attached but that it was difficult to get a molecular ion in mass spectrometry which is consistent with previous results as the oxime bond appears particularly susceptible to ionization. There was no evidence in the spectra of the carboxyl methyl substituted compounds and it was thought that any polar compounds may not have eluted under these conditions, the column was reversed and washed out. The mass spectra of this gave the molecular ion for the compounds substituted with 1 benzyl and 2 carboxy methyl, m/z 866, and 3 carboxyl methyl, m/z 835.

3.5.1.2 PyCA Mix 2

It was thought that replacing carboxymethyl with a methyl would make the above reaction easier to follow and therefore it was repeated by replacing the *O*-carboxymethyl hydroxylamine with *O*-methyl hydroxylamine. Using the same TLC conditions as above five spots were seen in the reactions mixture, $r_f = 0.79, 0.71, 0.69, 0.55$ and 0.39 . These compare to the starting material ($r_f = 0.64$), tri methyl substituted ($r_f = 0.4$), and tri benzyl substituted ($r_f = 0.8$) (Figure 3.11).

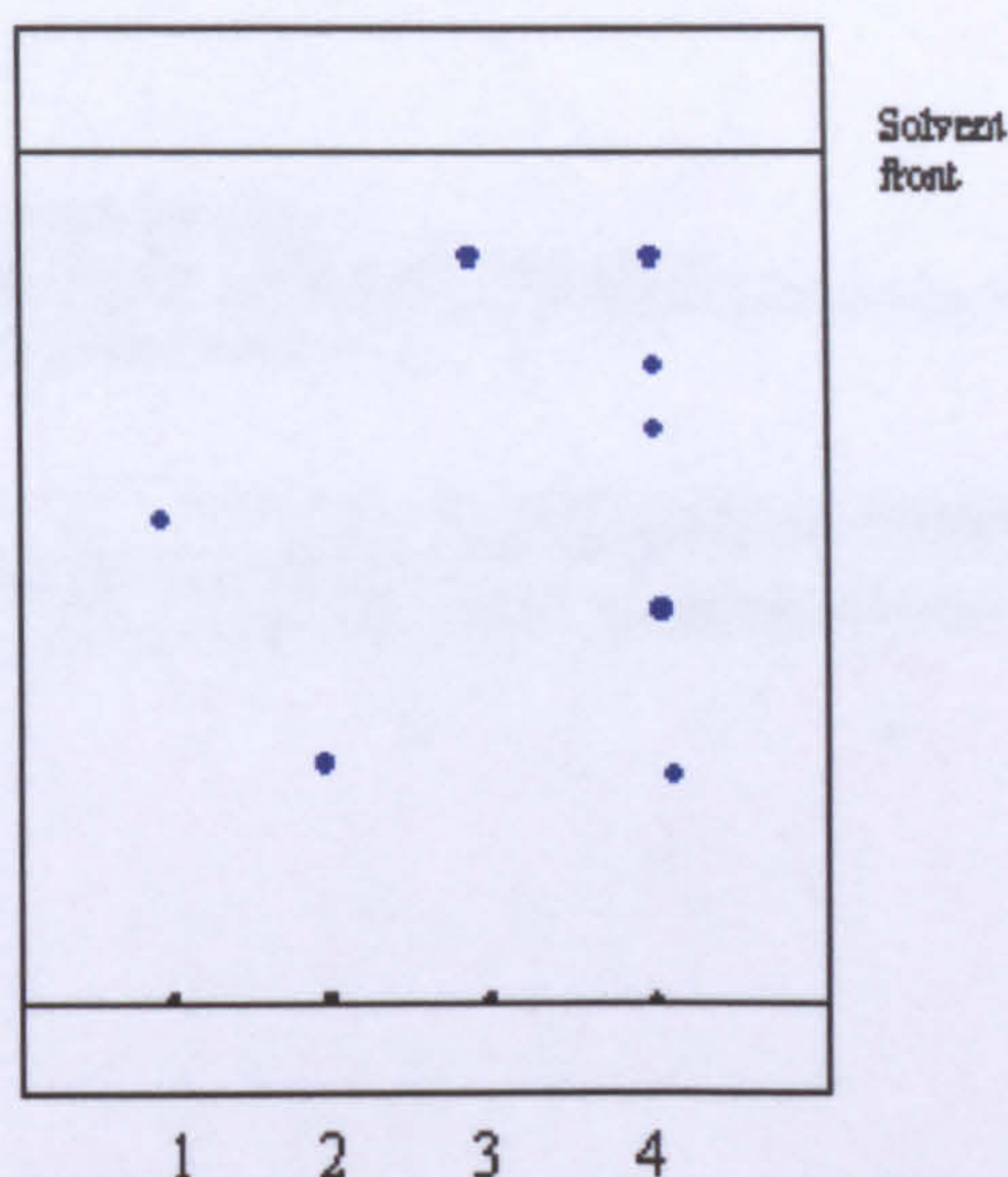


Figure 3-11. A diagram to represent the TLC plate of the second pilot study to show that a mixture of compounds are found in the presence of different aminoxy compounds. 1 indicates the position of the PyCA starting material, 2 indicates the position of the tri-methyl substituted PyCA compound and 3 indicates the position of the tri-benzyl substituted PyCA compound. 4 is the range of spots produced as a result of reacting both aminoxy compounds with PyCA.

It was not possible to use the previous RP-HPLC conditions as although the sample was soluble in methanol, when water was added a precipitate formed, and an aqueous sample is essential to run reverse phase HPLC. Thus a normal phase HPLC system was developed using 5% methanol in dichloromethane and again peaks were collected.

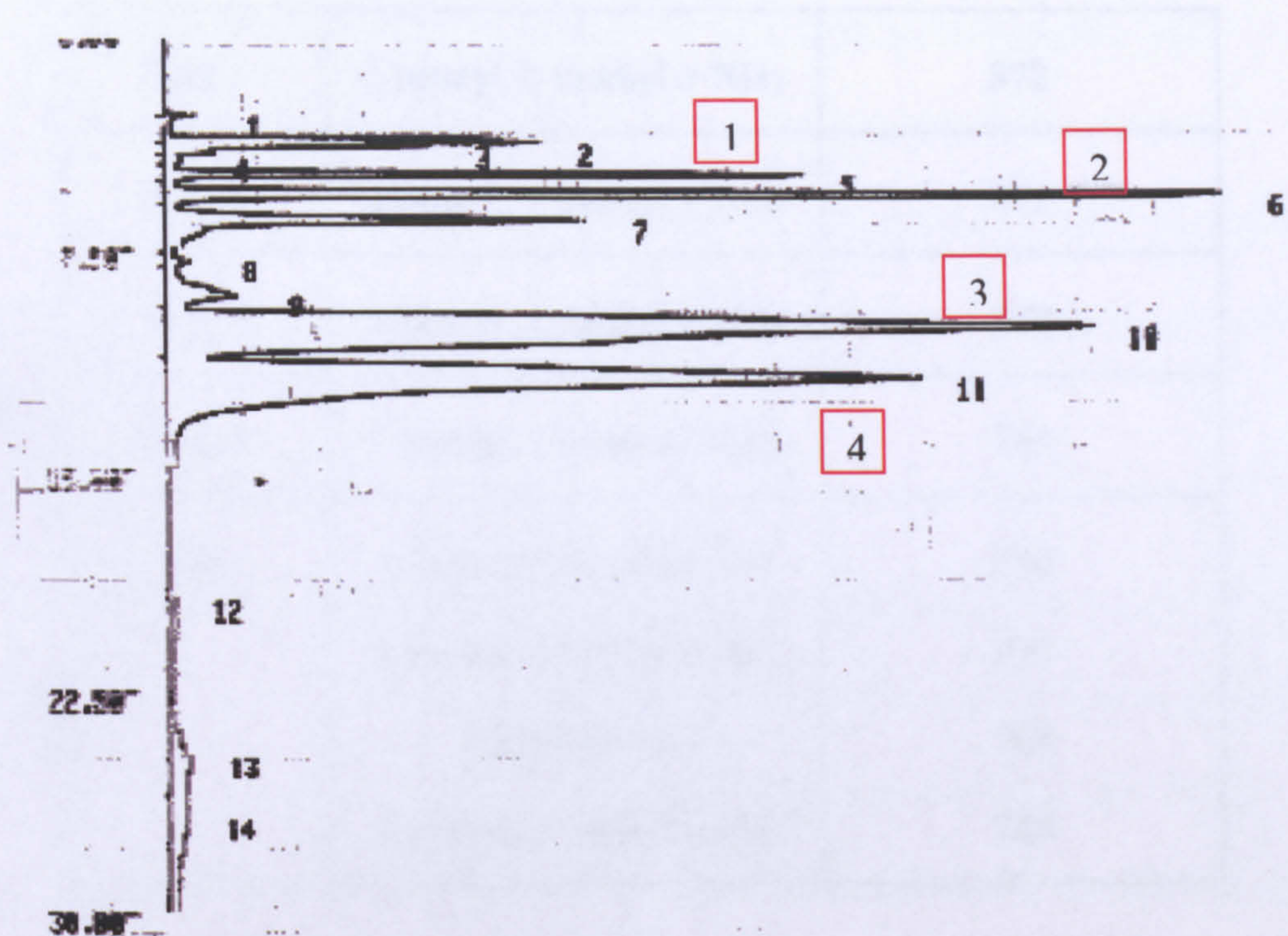


Figure 3-12. RP-HPLC trace of the mixed PyCA compounds found in the reaction mixture. The peaks 1-4 were collected and sent for MS.

Peak 1 gave the molecular ion for a di-benzyl, mono-methyl compound, as did peaks 2 and 3, this gives three isomers of the compound and it is assumed that the methyl group could be positioned at each of the three available sites. Peak 4 gave the molecular ion for the di-methyl, mono-benzyl compound. MS spectral data of the mixture showed the molecular ions of the tri-benzyl, mono-benzyl di-methyl, di-benzyl mono-methyl compounds.

Peak 1	2 benzyl, 1 methyl + NH ₃	872
Peak 2	2 benzyl, 1 methyl + NH ₃	872
Peak 3	2 benzyl, 1 methyl + NH ₃	872
Peak 4	1 benzyl, 2 methyl – CH ₂	765
Mix	1 benzyl, 2 methyl + H	780
	2 benzyl, 1 methyl + NH ₃	872
	3 methyl + H	704
	1 benzyl, 2 methyl – CH ₂	765

Table 3-2. A table showing the compounds in which samples and their *m/z* ratios.

3.6 Synthesis and analysis of the library

Based on the above pilot studies, a larger library was made using the prepared *N*-(aminooxy acetyl) amino acid compounds in a molar excess to the three functional groups on the PyCA scaffold. TLC was carried out after the overnight reflux, which showed that there was no starting material present and a range of fluorescent spots between $r_f = 0.56 - 0.70$. This was followed by the deprotection steps that removed the Fmoc group from the lysine side chain. Normal phase HPLC on the PyCA tri-arginine aminooxy compound and tri-lysine aminooxy compounds did not give a clear peak however the PyCA tri-glycine derivative did and this peak was present when it was co-eluted with the library mixture. When a sample of PyCA tri-phenylalanine and the library mix were co-eluted there was a sharp increase in the peak sizes that had retention times 37.8 and 38.9. Again the column was reversed and it was found that a number of compounds had become stuck on the column. After base hydrolysis to

remove the methyl ester from the C-termini of the aminoxy compounds, TLC showed a baseline fluorescent spot. At pH 11 the library was soluble in water. Following addition of acid, pH 4, the resulting solid was spun in a centrifuge and the supernatant removed. After repeated re-suspension in water and centrifugation, the solid was dried under vacuum.

Co-injection of standard compounds with the library revealed that there was no PyCA starting material in the library as PyCA had a retention time of 28.6 and this did not correspond to any peaks in the library. However, there was evidence that the tri-phenylalanine derivative was present, as were peak patterns for the tri-arginine, tri-glycine and tri-lysine derivatives. The HPLC trace shows that while there were many compounds present, the UV trace shows that they all have the characteristic profile of pyrene containing compounds as shown in the following Figures 13 and 14.

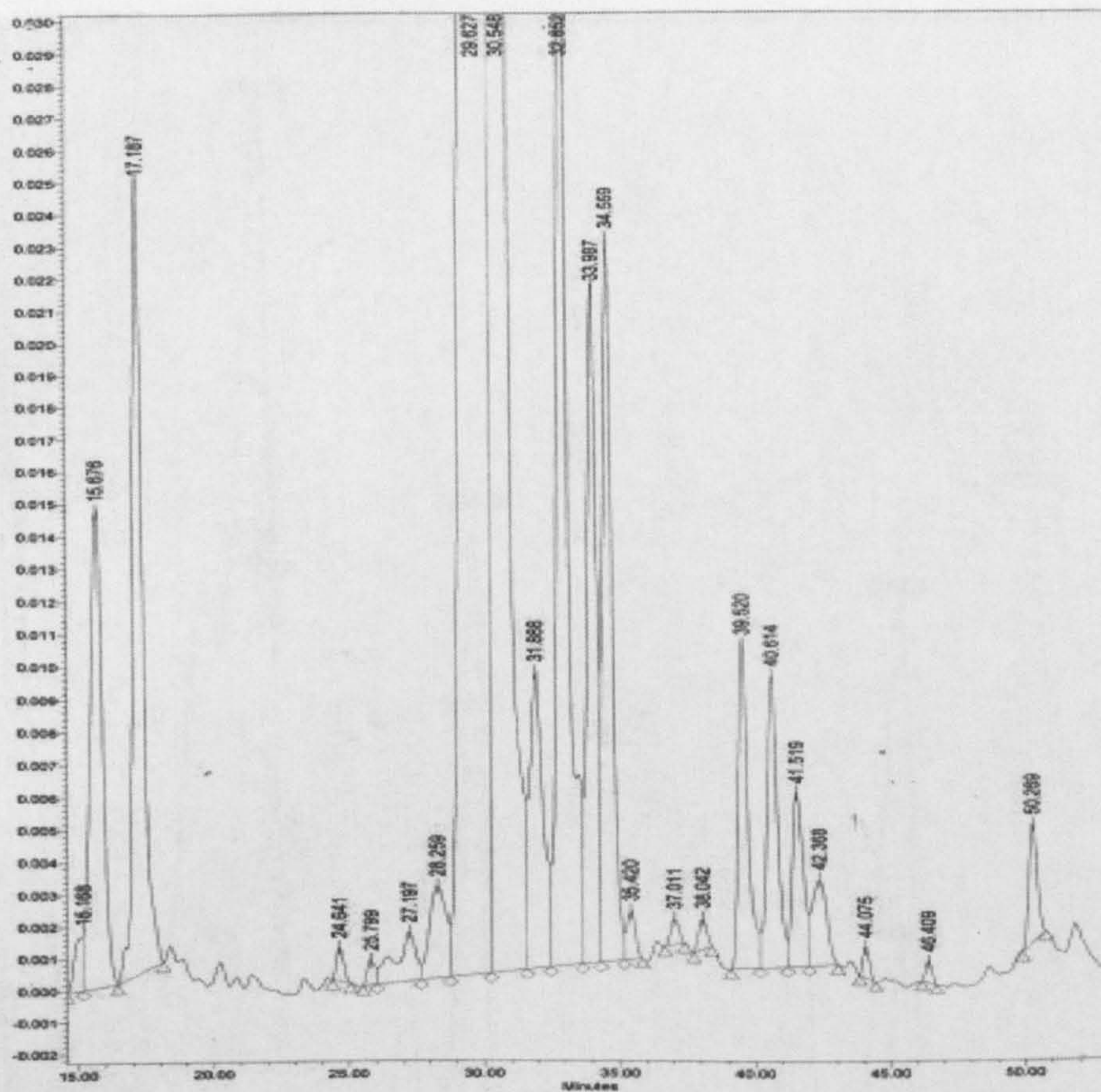


Figure 3-13. The HPLC profile of the library mix of compounds, showing that there is a wide variety of compounds in the library. The retention times range from 15 minutes to 50 minutes. When the starting material was coeluted there was no correspondence with any of the peaks in the shown trace. A cross section of peaks, range from early to late eluting compounds, were taken and a UV trace recorded, shown in Figure 3.14.

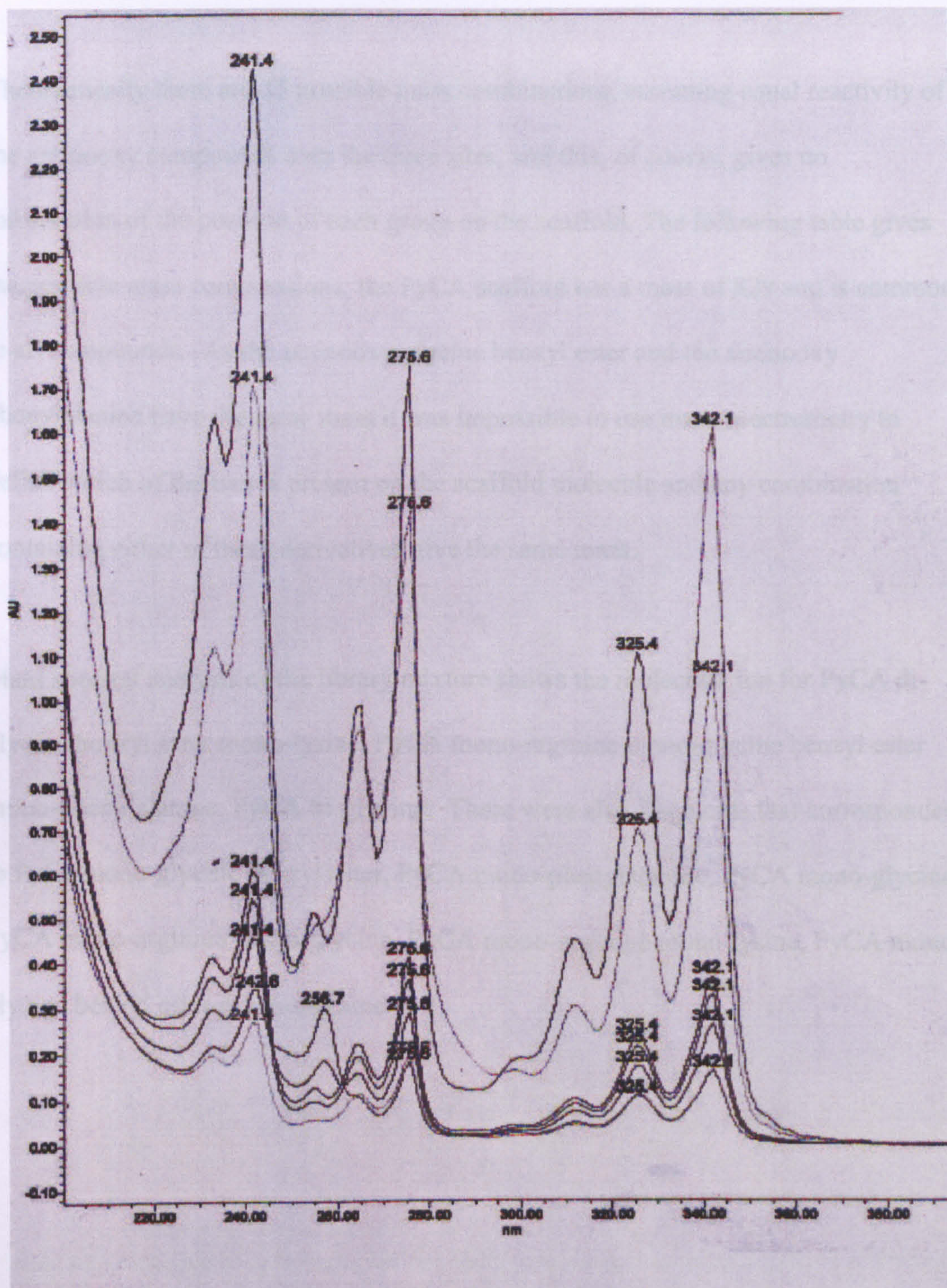


Figure 3-14. A cross section of peaks analysed in UV. This shows that irrespective of being an early or late eluting compound the UV was the same and indicative of pyrene containing compounds.

Theoretically there are 35 possible mass combinations, assuming equal reactivity of the aminoxy compounds over the three sites, and this, of course, gives no information of the position of each group on the scaffold. The following table gives the possible mass combinations; the PyCA scaffold has a mass of 829 and is common to all compounds. As the aminoxy glycine benzyl ester and the aminoxy phenylalanine have the same mass it was impossible to use mass spectrometry to define which of the two is present on the scaffold molecule and any combination containing either of these derivatives give the same mass.

Mass spectral analysis of the library mixture shows the molecular ion for PyCA di-glycine benzyl ester mono-lysine, PyCA mono-arginine mono-glycine benzyl ester mono-phenylalanine, PyCA tri-glycine. There were also fragments that corresponded to PyCa mono-glycine benzyl ester, PyCA mono-phenylalanine, PyCA mono-glycine, PyCA mono-arginine mono-glycine, PyCA mono-arginine mono-lysine, PyCA mono-glycine benzyl ester mono-arginine

Combination	PyCA	Mass 1	Mass 2	Mass 3	Total
aaa	829	159	159	159	1306
aab	829	159	159	149	1296
aag	829	159	159	59	1206
aal	829	159	159	130	1277
aap	829	159	159	149	1296
abb	829	159	149	149	1286
abg	829	159	149	59	1196
abl	829	159	149	130	1267
abp	829	159	149	149	1286
agg	829	159	59	59	1106
agl	829	159	59	130	1177
agp	829	159	59	149	1196
all	829	159	130	130	1248
alp	829	159	130	149	1267
app	829	159	149	149	1286
bbb	829	149	149	149	1276
bbg	829	149	149	59	1186
bbl	829	149	149	130	1257
bbp	829	149	149	149	1276
bgg	829	149	59	59	1096
bgp	829	149	59	149	1186
bgl	829	149	59	130	1167
bll	829	149	130	130	1238
blp	829	149	130	149	1257
bpp	829	149	149	149	1276
ggg	829	59	59	59	1006
ggl	829	59	59	130	1077
ggp	829	59	59	149	1096
gll	829	59	130	130	1148
glp	829	59	130	149	1167
gpp	829	59	149	149	1186
lll	829	130	130	130	1219
llp	829	130	130	149	1238
lpp	829	130	149	149	1257
ppp	829	149	149	149	1276

Table 3-2. The possible mass combinations; PyCA is common to all the compounds and Mass 1 represents the possible aminoxy compound that could be present at one of the three keto sites of the scaffold, Mass 2 and Mass 3 therefore represent the other two sites that are available for reaction. The total mass is give in the final column along with the average mass of the library compounds. a = the aminoxy arginine derivative, b = aminoxy glycine benzyl ester derivative, g = aminoxy glycine derivative, l = aminoxy lysine derivative, p = aminoxy phenylalanine derivative. The average mass of a compound within the library is 1216.6.

3.7 Regioselective oxime formation

The ability to select one of the three ketones over the other would definitely be a synthetic advantage. Previous work has shown that gentle conditions may be employed to give selectivity at the C3 position¹⁴⁷, and it is believed that the C12 position is more reactive than the C7¹⁷⁰. Therefore it was thought that it would be possible to control the reaction first at the C3 and then at the C12 by using a limited amount of reactants and gently increasing the temperature. It is known that refluxing the PyCA scaffold overnight with a >3 excess of aminoxy compounds results in a reaction occurring at all three keto sites, so it seemed logical that by controlling the time, temperature and aminoxy equivalents that the sites on the scaffold can be manipulated individually. A simple small scale reaction was carried out in which PyCA and O-benzylhydroxylamine were reacted together in the presence of appropriate amounts of sodium acetate and ethanol. A heating block was used to heat individual reaction tubes and the reaction monitored by TLC, eluant 5% methanol in chloroform. In conclusion it was possible to carry out the reaction in 1 hour at room temperature and get a single product over the three ratios, this indicates that one site, possibly C3, acquires the benzyl group. After 5 hours at 70°C all three sites had gone to completion in the three to one ratio mixture, however there was no clear evidence that one of the other sites was reacted before the other as after 3 hours trace spots were noted in the two to one and the three to one mixes which were identical. The two to one mix had a total four spots after 5 hours. It should be possible to investigate this in more detail, on a larger scale so that enough material is made to take samples after each hour and explore the structure of the compounds by NMR which should show which position is being reacted and possibly in what order.

Chapter 4.

Analysis of cholate-amino acid libraries for recognition of O⁶-carboxymethyl-2'-deoxyguanosine

'Science.... never solves a problem without creating ten more'.

(George Bernard Shaw 1856-1950)

4.1 Chromatography

Chromatography was coined from the Greek of two words *chroma*, for colour, and *graphein*, to write, by a Russian botanist M. S. Tswett in 1903. He first used the technique for the separation of coloured plant pigments and, although he did state that it could also be used for the separation of colourless substances, chromatography was the term he used in all his papers. Since these early beginnings chromatography has become a universal, versatile technique, in its many guises¹⁷¹.

Despite of all the variations in the application of chromatography the principle remains the same, separation occurs by the distribution of the components between two phases, one of which is stationary and a mobile phase¹⁷². The mixture to be separated is dissolved in the mobile phase and passed over an adsorbent material, the stationary phase. Due to the differing properties of the compounds in the mixture they adsorb to different extents and migrate through the stationary phase at different rates. They are eluted and emerge as (usually) separate compounds¹⁷³.

In liquid chromatography the mobile phase is a liquid which passes over a liquid or, more usually, a solid stationary phase of a large surface area. This can be an adsorbent surface, found in column chromatography, thin-layer chromatography (TLC) and high performance liquid chromatography (HPLC); a resin, used in ion exchange; a gel, used in size exclusion, such as that mentioned in section 3.2, and affinity columns.

In TLC the stationary phase is planar and consists of a thin layer of the stationary phase being spread uniformly over a flat sheet of either glass, plastic or aluminium.

In the other techniques the stationary phase is packed in to a column.

Affinity chromatography involves the immobilization of either a ligand or an affinity reagent, such as an antibody, on to a support matrix and separation relies on biospecific interactions, and it is this specificity that allows some molecules to be retained on the column, whilst unbound molecules are eluted in the mobile phase. As the interaction between the molecule and ligand are reversible molecules are eluted by changing the pH or composition of the mobile phase. A demonstration of this technique has already been discussed in section 1.5.1 concerning the immunoaffinity/HPLC detection of O⁶CMdG. Briefly, the total immunoglobulin/G fraction from the rabbit antiserum were covalently linked to a Protein-A-Sepharose matrix. O⁶CMdG was then passed down the column, the binding between the modified matrix and O⁶CMdG being so strong that 1 M trifluoroacetic acid (TFA) was required to release the ligand from the matrix. Under these conditions the glycosidic bond was hydrolysed and the eluting fractions contained the alkylated base O⁶CMguanine, and it was this moiety that was detected by HPLC fluorescence^{43,44}.

4.2 HPLC development for the analysis of the cholic acid libraries

It was thought that it would be possible to design a HPLC system that could be used to give a series of retention times corresponding to each of the trisubstituted compounds. These could then have been used to identify any trisubstituted compounds that appear in the library. Any intermediate peaks would be tentatively assigned their amino acids residues based on their retention times. To achieve this it

would be necessary to find a system that worked on all compounds simultaneously so that a comparison could be made.

A suitable gradient was developed that used an increase of acetonitrile from 15% to 90% over 20 minutes with water and a constant 5% 0.1 M phosphoric acid and this gave a clear peak for **16** (PyCA) with a retention time of 10.2 minutes.

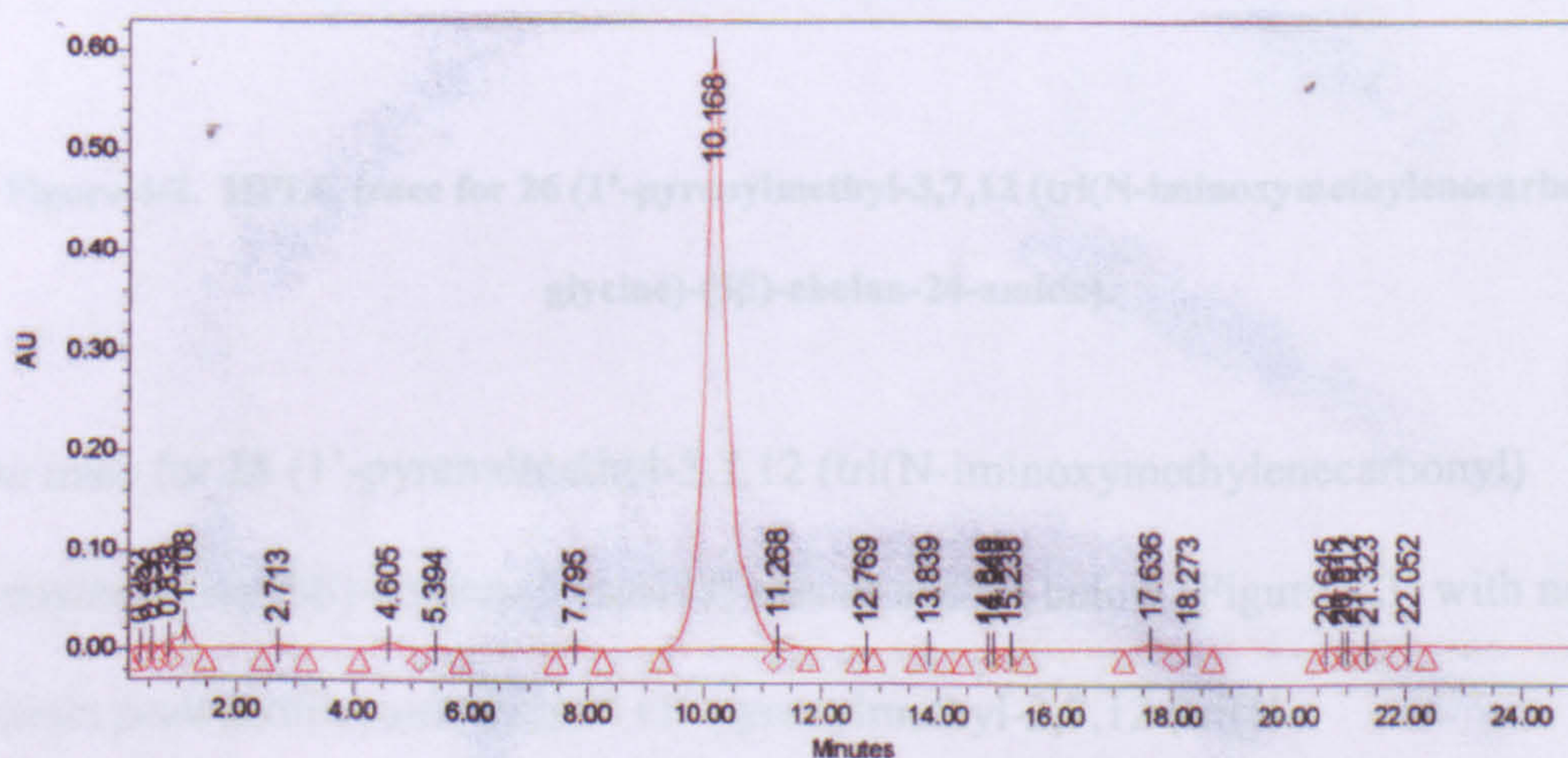


Figure 4-1. The HPLC trace of PyCA

The same system gave a peak for **26** (1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine)-(5 β)-cholan-24-amide) centered at 9.5 minutes. The peak was very broad indicative of there being very similar compounds being eluted over time (8.71- 10 minutes).

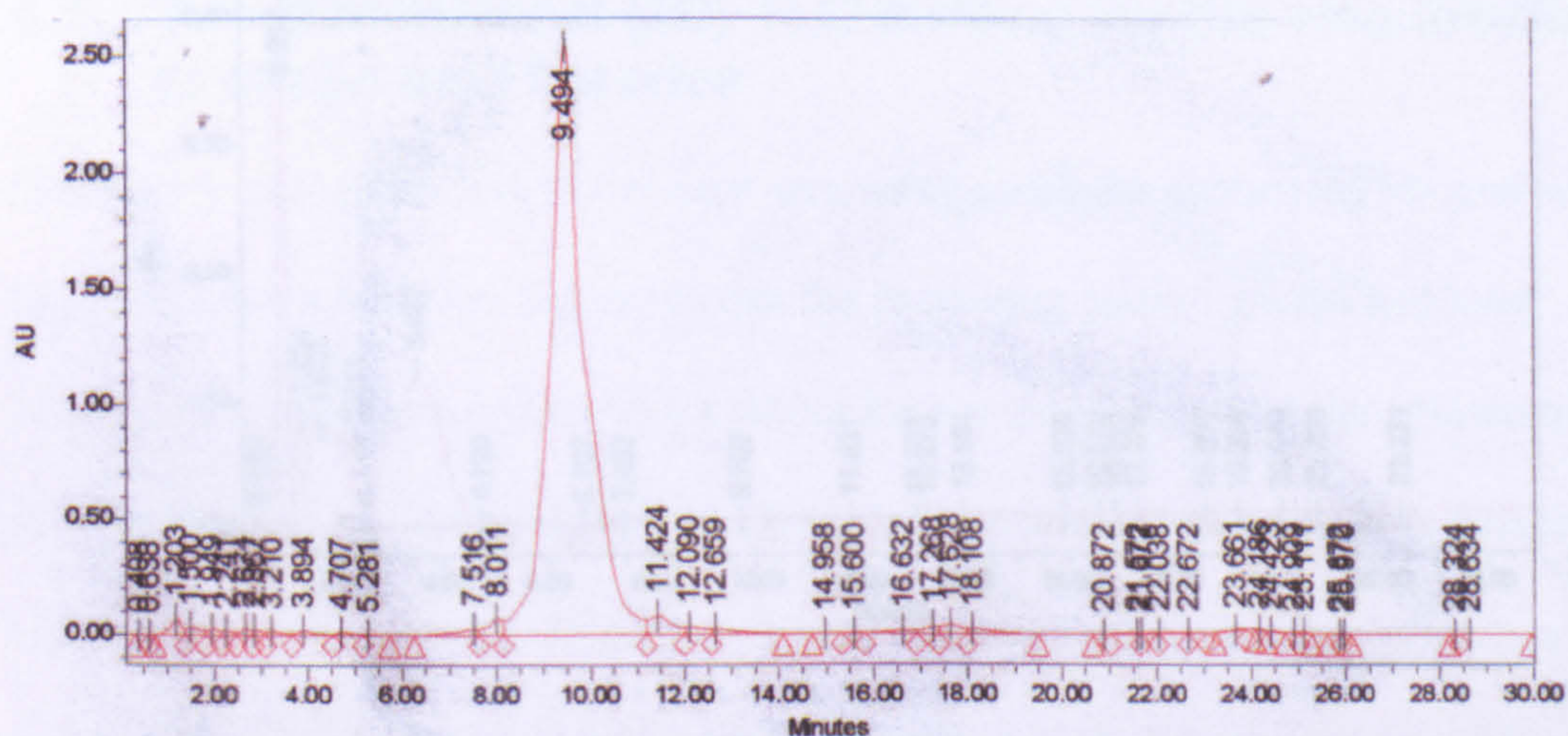


Figure 4-2. HPLC trace for 26 (1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine)-(5 β)-cholan-24-amide).

The trace for **28** (1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) phenylalanine)-(5 β)-cholan-24-amide) was as shown below (Figure 4.3) with no distinct peak profile, and with **27** (1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) lysine)-(5 β)-cholan-24-amide) no stable baseline could be established.

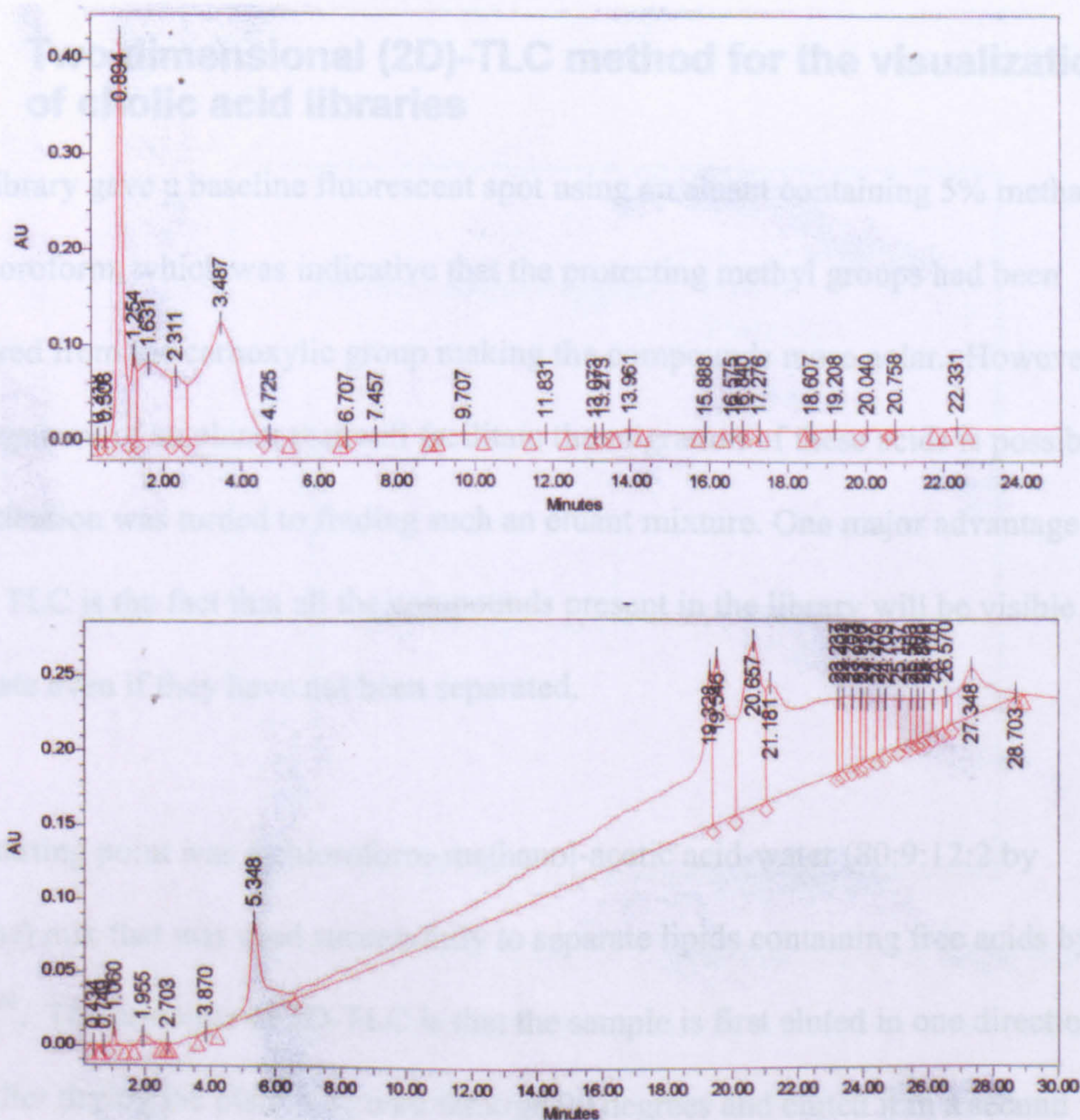


Figure 4-3. HPLC traces showing the problems in the chromatography of trisubstituted compounds.

Considerable time was spent trying to modify the method set but by gaining resolution for one compound resulted in the loss of resolution for another. Without the ability to run all compounds on the same method set, HPLC could not be used consistently against the library.

4.3 Two dimensional (2D)-TLC method for the visualization of cholic acid libraries

The library gave a baseline fluorescent spot using an eluant containing 5% methanol in chloroform, which was indicative that the protecting methyl groups had been removed from the carboxylic group making the compounds more polar. However the development of an eluant that will facilitate the migration of these acids is possible and attention was turned to finding such an eluant mixture. One major advantage of using TLC is the fact that all the compounds present in the library will be visible on the plate even if they have not been separated.

The starting point was a chloroform-methanol-acetic acid-water (80:9:12:2 by volume) mix that was used successfully to separate lipids containing free acids by 2D-TLC¹⁷⁴. The principle of 2D-TLC is that the sample is first eluted in one direction and, after drying the plate is rotated through 90 degrees and eluted it in a second direction, thus separating the sample over a larger area (Figure 4.4).

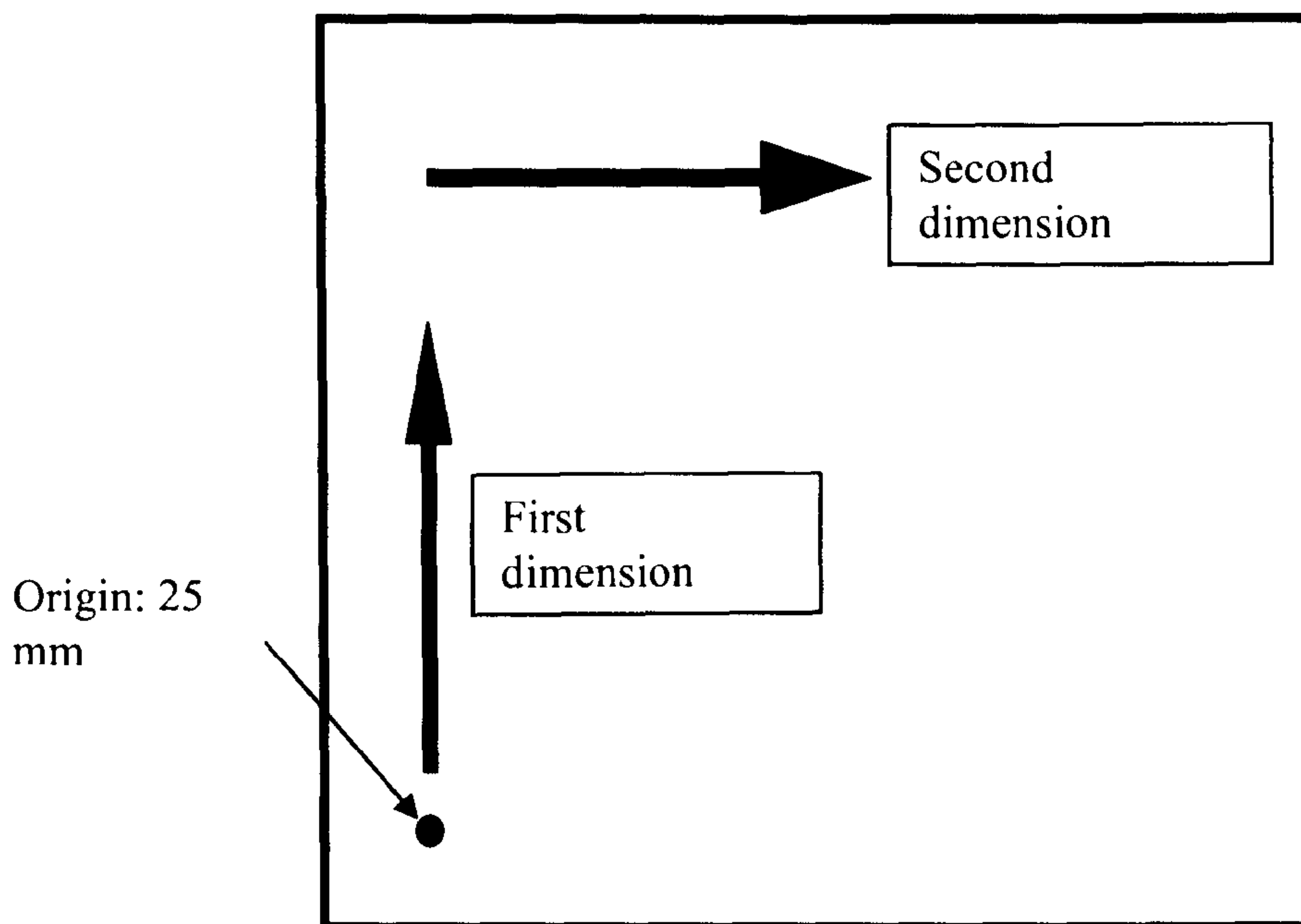


Figure 4-4. A representation of 2D-TLC, in which the sample is spotted at the origin which is positioned 25 mm from both edges of the plate. The plate is run in an appropriate eluant in the direction of the first dimension. The plate is then dried and run again, in either the same or a different eluant.

The eluant system that proved the most successful in giving the greatest spread of compounds was with the first dimension consisting of 94% chloroform, 6% methanol with a trace (50 μL in 500 mL) of TFA. This was complemented with the second dimension being eluted 65% chloroform, 26% ethyl acetate, 9% methanol and again a trace amount of TFA (100 μL in 768 mL). The preliminary work was carried out on plastic backed Macherey-Nagel™ TLC plates, which were replaced by glass backed SIL G-25 plates. High performance TLC (HPTLC) plates have a smaller particle size which should give a tighter particle distribution; these were tried (Analtech HPTLC-HL silica gel 150 microns) but they did not give a significant improvement in

resolution. The plates of choice proved to be Merck glass backed silica gel 60 20 cm x 20 cm.

Figure 4.5 shows the main problem in developing the TLC system, namely, a secondary line that appeared after the solvent front had moved up the plate. Figure 4.6 shows the consequence of the presence of this line. This was thought initially to be due to one or more of the following phenomena: (a) a problem with the polyvinyl alcohol binder of the silica, (b) a result of solvents de-mixing in the vapour state or (c) changes in the hydration state of the silica. The latter was later proved to be inconsequential as added trace amounts of water to the eluting system still resulted in the formation of the line. The exact nature of the binder on the Merck plates is a trade secret and no information could be obtained from the manufacturer. However, it was found that this problem could be minimised by adopting the following procedure. The plates were first run in the first dimension system and then left overnight in a vacuum dessicator. After sample application and the first dimension run, plates were removed from the tank and allowed to dry under vacuum for 2-3 hours before commencing with the second dimension.

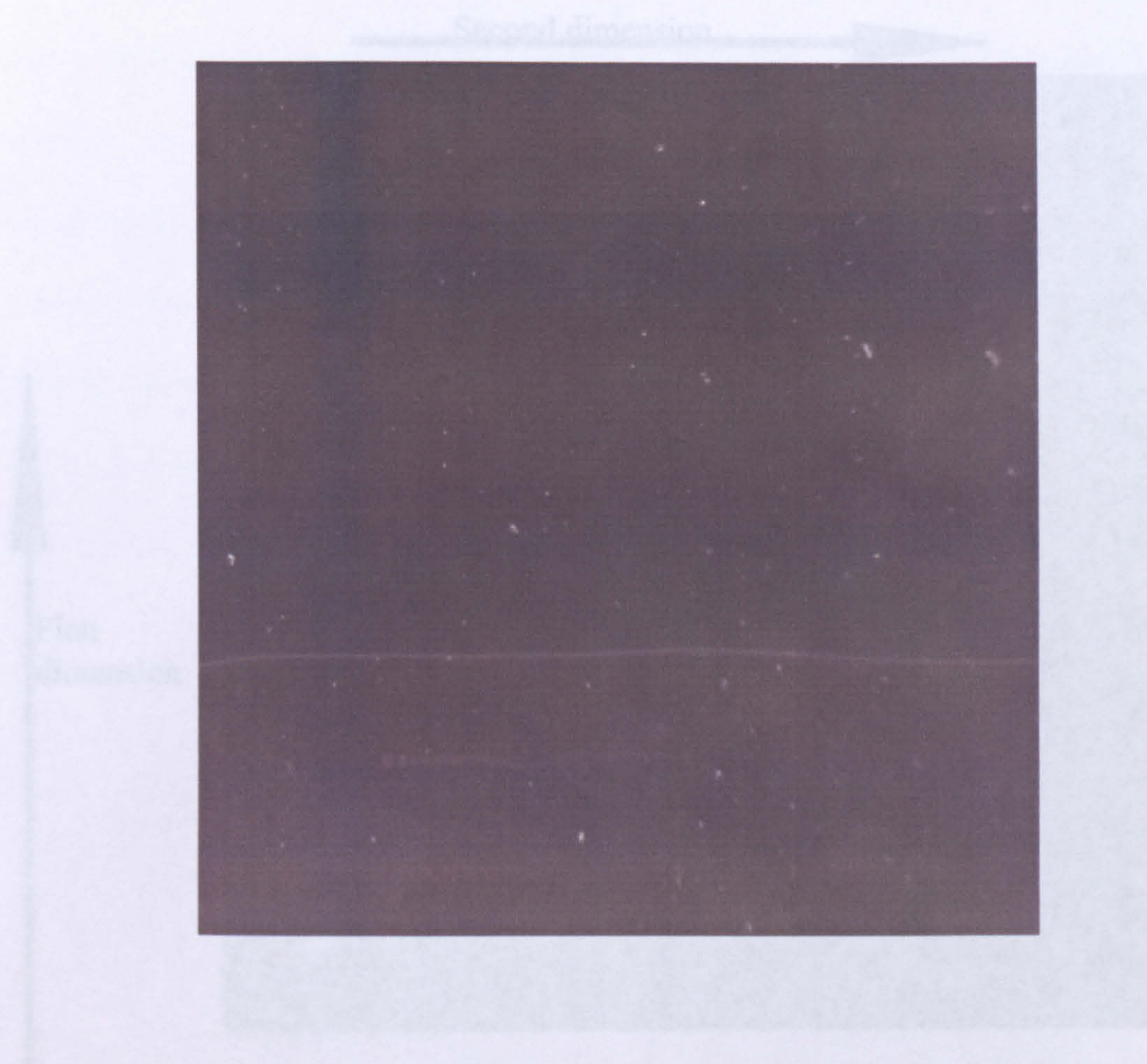


Figure 4-5. A photograph of a TLC plate that has been washed in the first dimension eluting system showing a line across the plate. Image taken on Kodak imager.

Figure 4-5 is a photograph of a TLC plate that was run, after sample application, in the first dimension and then in the second dimension. It can be seen how the line that appears across the plate in the first dimension wash and causes a distortion of the spots. The red circle indicates the origin.

Samples were applied as a 10 μL spots of 5 mg mL⁻¹. A stream of N₂ was used between each spot to thoroughly dry the spot before the next addition, this was done to keep the spot as tight as possible. It was found the fluorescence was visible on the plate suggesting that some of the residue was not being applied to the plate. Changing to a syringe with a plastic tipped needle reduced this residue. Following these procedures it was possible to obtain a reproducible pattern of spots on the TLC

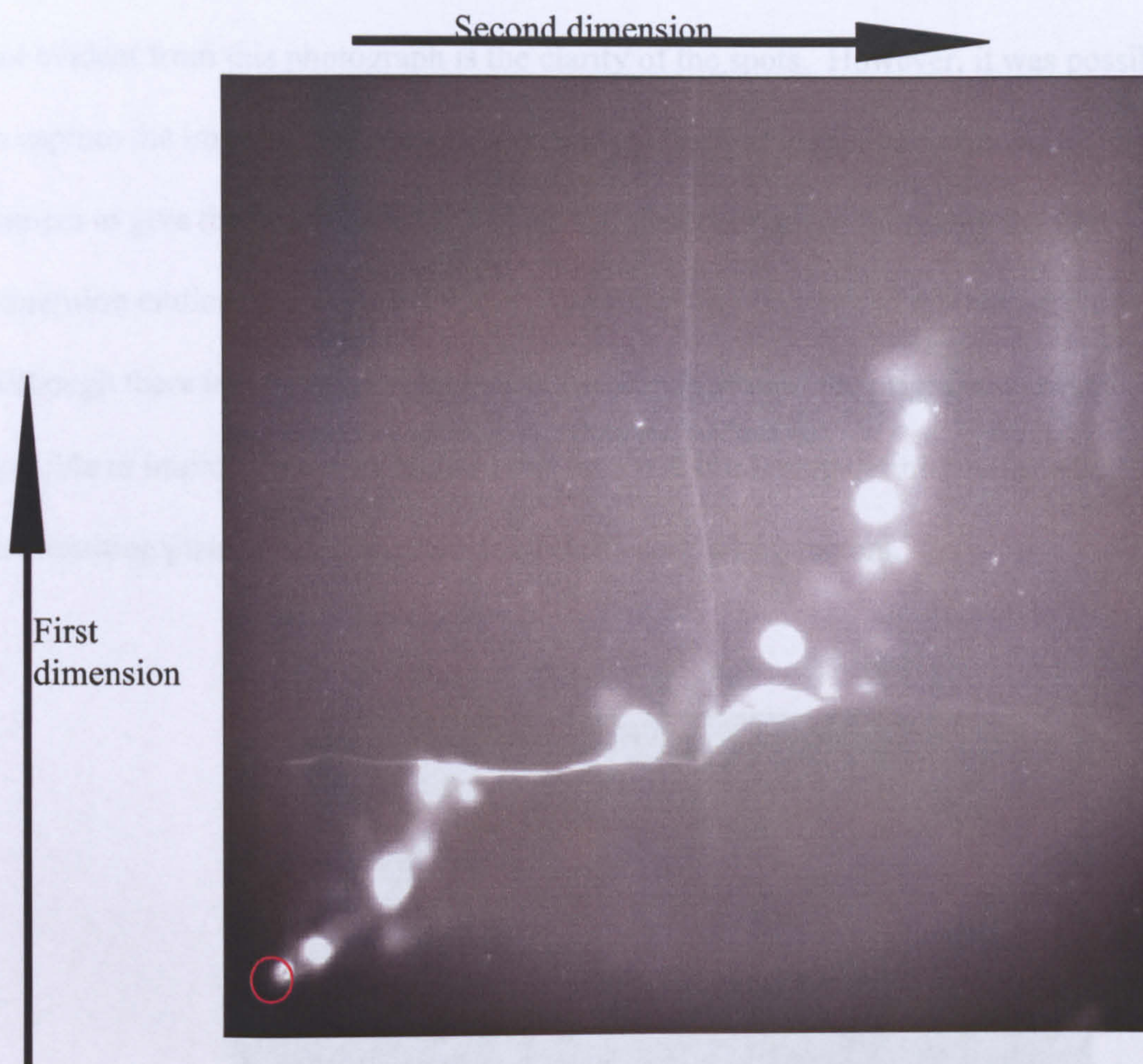


Figure 4-6. A photograph of a TLC plate that was run, after sample application, in the first dimension and then in the second dimension. It can be seen how the line that appears across the plate carries the sample along and causes a distortion of the spots. The red circle indicates the origin.

Samples were applied as a 10 μL spots of 5 mg mL^{-1} . A stream of N_2 was used between each spot to thoroughly dry the spot before the next addition, this was done to keep the spot as tight as possible. It was found the fluorescence was visible on the needle tip suggesting that some of the residue was not being applied to the plate. Changing to a syringe with a plastic tipped needle reduced this residue. Following these procedures it was possible to obtain a reproducible pattern of spots on the TLC

plate as shown in Figure 4.7, taken on a Kodak imager in fluorescence mode. What is not evident from this photograph is the clarity of the spots. However, it was possible to capture the image of the plate as seen under UV light in the darkroom with a digital camera to give the images seen in Figure 4.8, the spot pattern following the first dimension elution, and Figure 4.9, the plate following the second dimension elution. Although there is a significant area of fluorescence around the origin is was not possible in improve the resolution in this area without losing definition elsewhere and the resulting plate was a compromise of these conflicting factors.



Figure 4-7. A photograph of the TLC following the first and second dimension eluting systems, which gave the best spread of spot pattern that was reproducible.

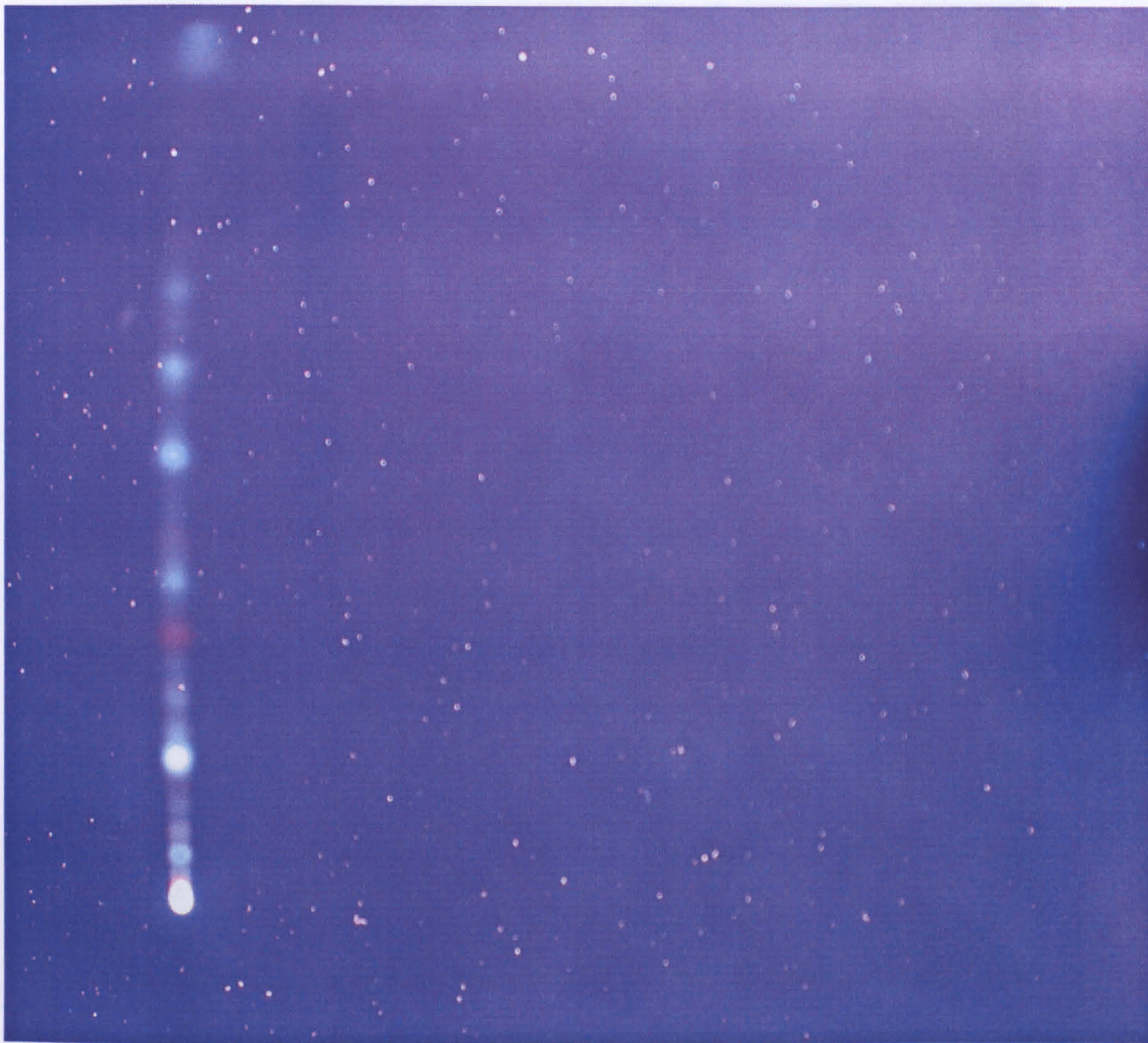


Figure 4-8. A photograph of the TLC plate following some chemical analysis showing the spot spread, taken with a digital camera in the darkroom under UV light.

Figure 4-8. A photograph of the first dimension showing the spot spread, taken with a digital camera in the darkroom under UV light.

By simply laying an acetate sheet over the plate and marking the position of the spots, a map of the pattern was made that could be used to compare the position of known compounds with those within the library mixture (Figure 4.10). The library contained elements within it which were recognized as being the trisubstituted compounds. There was also a spread of compounds between those that have intermediate polarities. The starting material is not evident in the library.

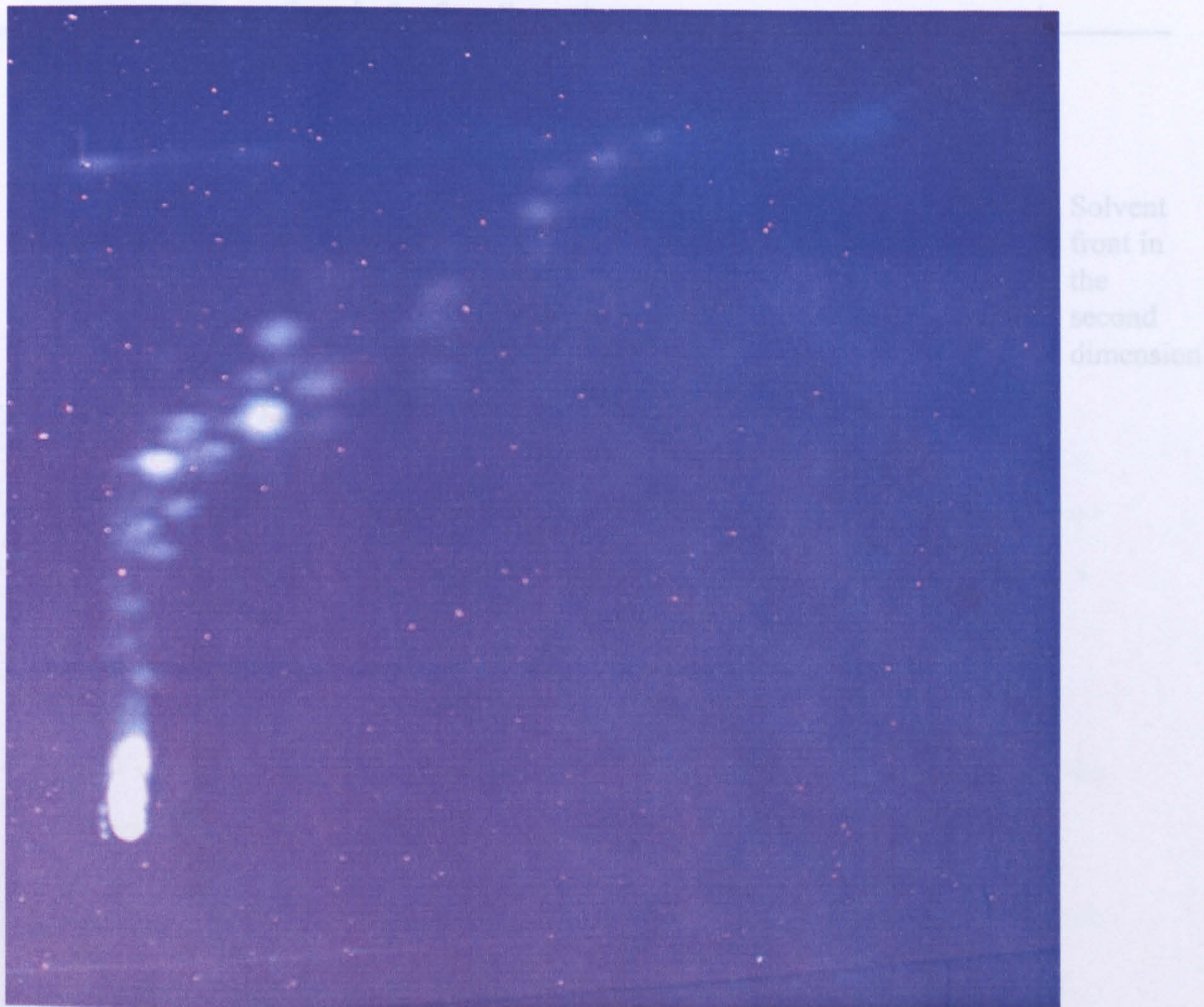


Figure 4-9. A photograph of the TLC plate following both dimension elutions showing the spot spread, taken with a digital camera in the darkroom under UV light.

By simply laying an acetate sheet over a plate and marking on the position of the spots, a map of the pattern was made that could be then used to compare the position of known compounds with those within the library mixture (Figure 4.10). The library contained elements within it which were recognised as being the trisubstituted compounds. There was also a spread of compounds between these that have intermediate polarities. The starting material is not evident in the library.

Solvent front in the first dimension

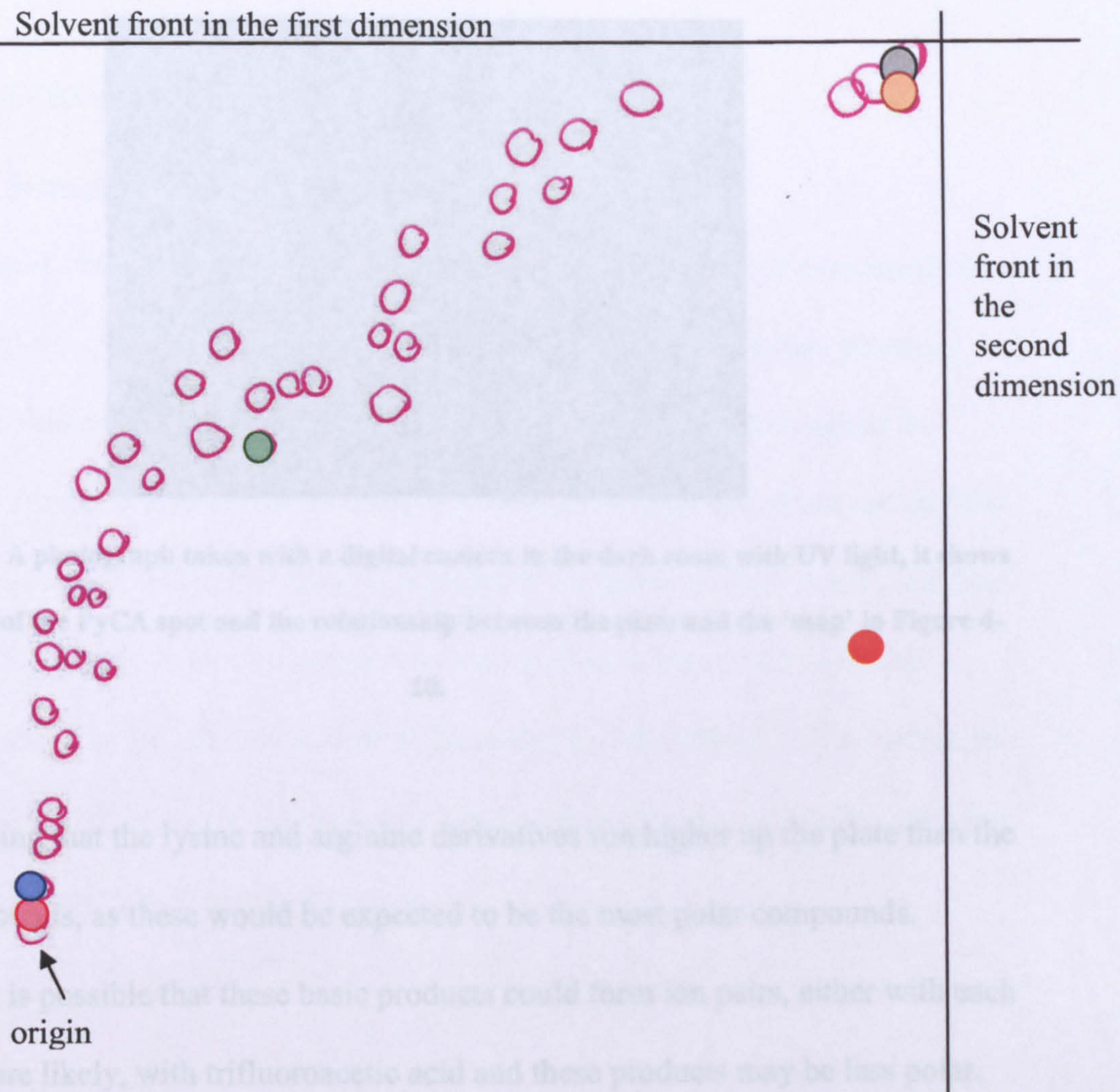


Figure 4-10. A map of the position of the spots that were produced following 2D-TLC, made by laying a sheet of acetate over the plate and marking outline of the highly visible spots. The coloured circles indicate the position of a small cluster of spots that were found when differing compounds were eluted in the system as that used for the library. The brown and black circles in the upper right hand corner indicate the position of 27 (1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) lysine)-(5 β)-cholan-24-amide) and 29 (1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) arginine)-(5 β)-cholan-24-amide), respectively. 26 (1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) glycine)-(5 β)-cholan-24-amide) is indicated by the green circle. The blue circle marks the position of 25 (1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) glycine benzyl ester) -(5 β)-cholan-24-amide) and 28 (1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) phenylalanine)-(5 β)-cholan-24-amide) is shown as a red circle. Finally the orange spot is where 16 (PyCA) starting material is situated.

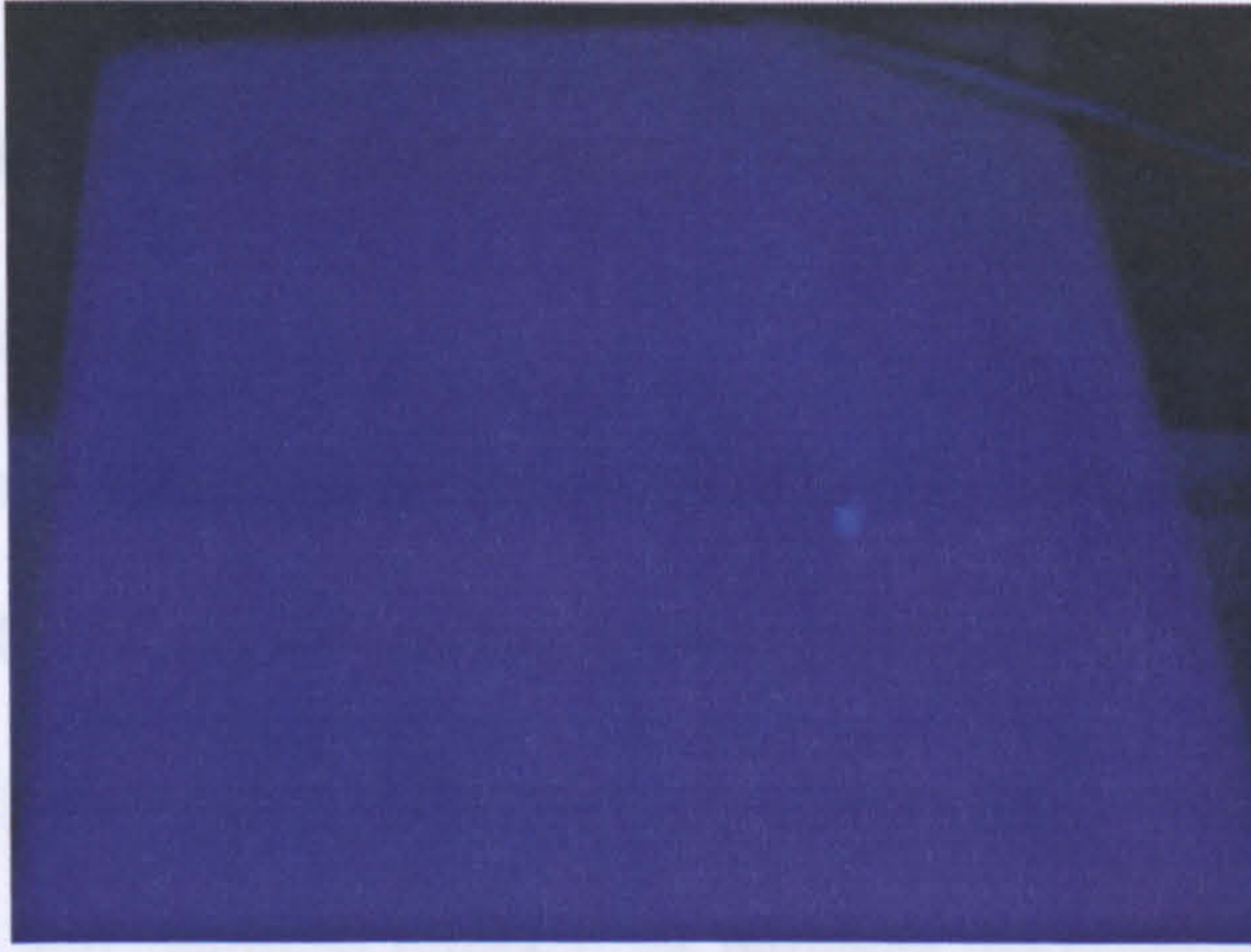


Figure 4-11. A photograph taken with a digital camera in the dark room with UV light, it shows the position of the PyCA spot and the relationship between the plate and the 'map' in Figure 4-10.

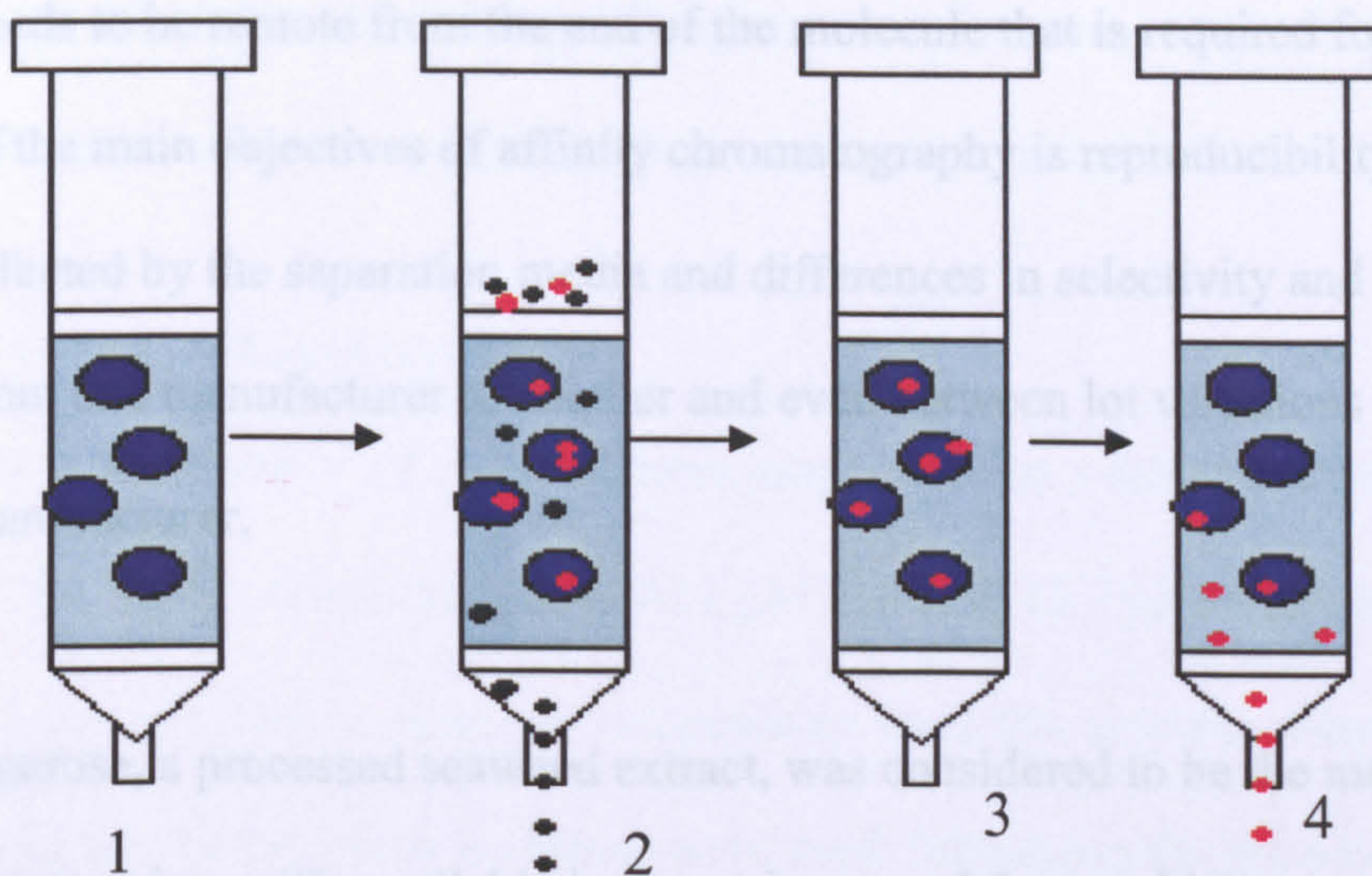
It is surprising that the lysine and arginine derivatives run higher up the plate than the other compounds, as these would be expected to be the most polar compounds.

However, it is possible that these basic products could form ion pairs, either with each other, or more likely, with trifluoroacetic acid and these products may be less polar.

4.4 Affinity chromatography.

4.4.1 Introduction

Affinity columns have been used for many years in the purification of proteins from complex mixtures. In affinity chromatography separation occurs because proteins interact with various matrices in different ways. These matrices are designed to exploit the physical and chemical interactions between the proteins of interest and the matrix. For example, antibodies may be coupled to a suitable column matrix and then packed into a column. A protein solution is then applied to the column and any antigenic proteins in the solution adhere to the immobilised antibodies. The adherent proteins are then eluted by the addition of a solution that disrupts the binding between protein and antibody, as shown in Figure 4.12.



1. A column matrix is prepared with covalently linked antibodies/ligands
2. A mix of ligands/antibodies is applied to the column, non binding molecules are washed out
3. The binding molecules are retained within the matrix, due to specific interactions.
4. A change in the eluate cause a disruption of the binding releasing the molecules

Figure 4-12. The basic principle of affinity chromatography. Antibodies or ligands are immobilised on a matrix and packed into a suitable column. A mixture of ligands or antibodies are applied to the column: those with specific interactions are retained within the matrix; non-binding residues are washed through the column. The adherent antibodies or ligands are released from the matrix by a change in the eluant solution which disrupts the binding between the two and the purified ligand or antibody is released from the column.

4.4.2 Column matrix

Affinity chromatography utilises an affinity support which is packed into the column, it consists of the antibody or ligand (usually of a known molecular configuration) covalently attached to a matrix of a solid insoluble substance. The matrix used is dependent upon the ligand and the method of ligand attachment. Ligand orientation needs to be considered. The end of the ligand that is used to bind it to the matrix

needs to be remote from the end of the molecule that is required for recognition. One of the main objectives of affinity chromatography is reproducibility, this can be affected by the separation media and differences in selectivity and capacity differ from one manufacturer to another and even between lot variations from the same manufacturer.

Agarose, a processed seaweed extract, was considered to be the most suitable matrix to use, it is readily available in the aminopropyl form, which was used in the linking of the ligand, **4** (O⁶CMG), in such a way as to not hinder the binding end of the molecule within the column. Agarose is a polysaccharide with alternating residues of D-galactose and 3,6-anhydro-D-galactose which provide an uncharged hydrophilic matrix with an abundance of primary and secondary hydroxyl groups. Secondary and tertiary structures form when the single fibres are twisted into multiple fibres and become folded into a fabric with large accessible pores.

Agarose has a relative narrow particle size range; resulting in a more efficient column by reducing the number of cracks that appear. There are also minimal 'fines', these are smaller than average particles that can restrict the column flow. Thorough washing of the matrix before use can remove the majority of the fines and gentle handling thereafter should prevent the formation of new ones.

The matrix is supplied as a pre-swollen thick paste that was thoroughly washed in buffer to remove preservatives and small fines. Three types of matrices were prepared, two were linked with ligands, guanosine and O⁶carboxymethyl guanosine as discussed in 3.2, and a third control column was prepared using just agarose.

Following extensive washing, the agarose matrices were degassed for three hours to prevent the formation of channels or cracks and packed into the columns.

4.4.3 Eluant system

The eluant used in the column chromatography was a dilute phosphate buffered saline (PBS), 1 μM phosphate buffer, 0.27 μM KCl and 13.7 μM NaCl, to which methanol was added to give a range of concentration from 0% to 90% methanol. High purity water and HPLC grade methanol were used. Increasing the methanol concentration changes the polarity of the solvent and increasingly disrupts the binding between any compounds which have shown binding to the column matrix. The column were eluted with increasing amounts of methanol in PBS and one cycle was 15 mL of each concentration starting with 0% and increasing to 90% in 10% increments.

4.4.4. Column capacity

Column capacity, that is, how much target was selectively bound per unit volume of support material, gives an indication of the amount of library that can be added to the column without it being overloaded and allows selection to occur. The column capacity also influences the efficiency of the column, for example if a column has a high capacity but takes a long time to go through one cycle, it is not efficient. For O^6CMG linked agarose the column capacity was determined by hydrolysing the adducted guanine base from the column matrix and comparing the HPLC trace against a standard for O^6CMG . A range of concentrations of standard from 0.5 to 100 pmol per 10 μL injection were used to give the following calibration curve.

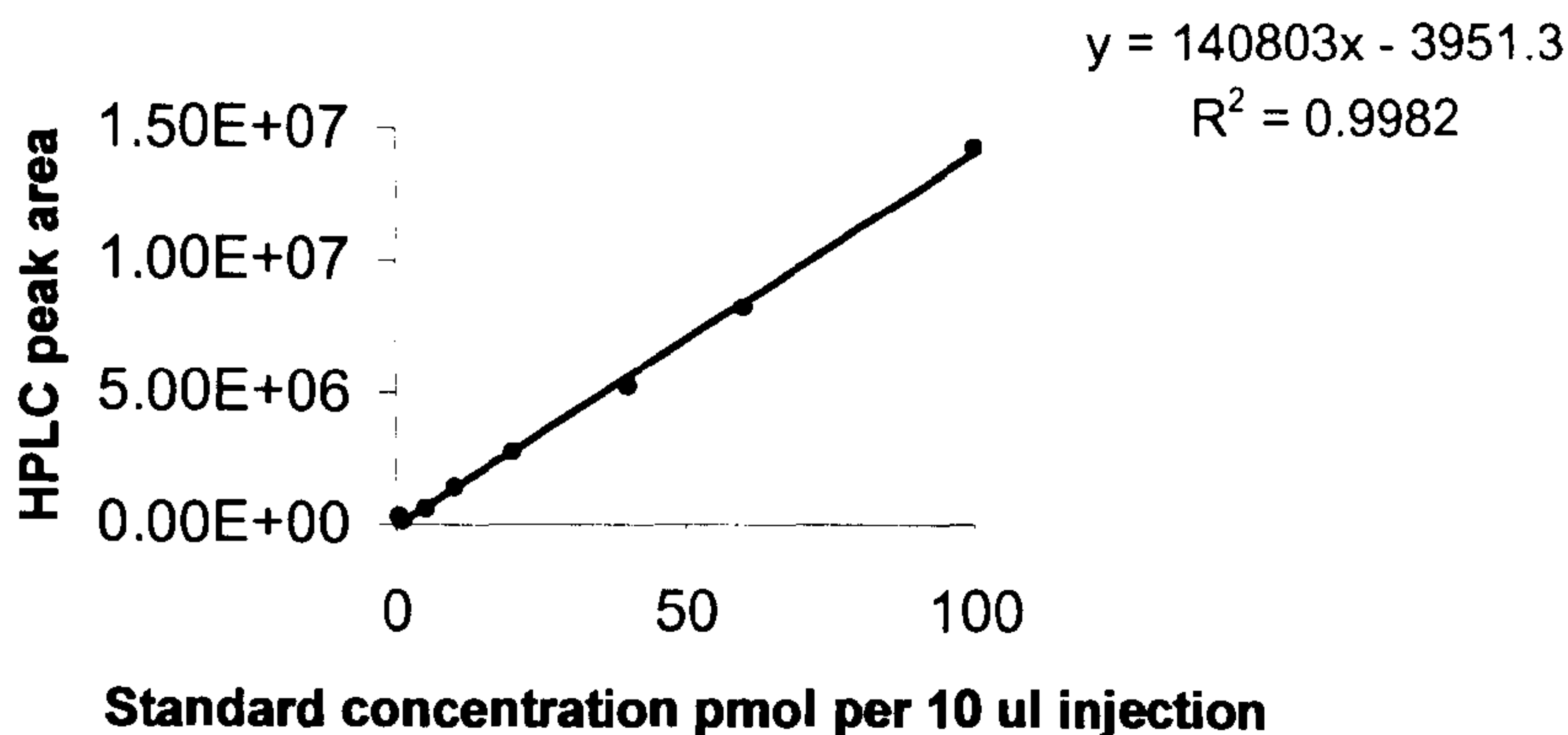


Figure 4-13. A standard line to show the relationship between the peak area and concentration of standards of O⁶CMG.

A known concentration of 50 nmol mL⁻¹ was eluted in this system and the peak area corresponded to 54 nmol mL⁻¹ in the standard. The samples from the hydrolysed matrix were then run and repeated runs showed that the affinity matrix contained 30.1 +/- 0.2 nmol O⁶CMG per mL of gel.

4.4.5 Antibody interaction with the column matrices

A sample of O⁶-CMdG antiserum from previous work⁴³ was found to have an absorbance at 280 nm of 0.9 per 100 μL. 250 μL of this stock were put onto two columns, one had a plain agarose matrix and the other contained the O⁶CMG-agarose matrix. These were then both eluted with PBS and increasing methanol concentrations, starting at 10% and going up to 90%. Three UV readings of each sample was taken and the average given in the following results.

Fraction Number	Methanol %	Agarose column	O ⁶ CMG column
1	0	0.0190	0.0792
2	0	0.2824	0.1935
3	0	0.3037	0.0566
4	0	0.3492	0.0567
5	0	0.2628	0.1398
6	0	0.2322	0.1352
7	10	0.3959	0.1049
8	10	0.1078	0.0514
9	20	0.1433	0.0910
10	20	0.0742	0.2484
11	30	0.1599	0.0238
12	30	0.1117	0.0881
13	40	0.0549	0.0813
14	40	0.1369	0.2761
15	50	0.0850	0.0746
16	50	0.0309	0.0420
17	60	0.0939	0.1109
18	60	0.0022	0.1148
19	70	0.0403	0.2310
20	70	0.0340	0.0995
21	80	0.0562	0.0837
22	80	0.0473	0.1109
23	90	0.0394	0.0595
24	90	0.0559	0.0568

Table 4-1. A table of results showing the UV absorbance at 280 nm over the 24 different fraction numbers and the corresponding methanol concentrations. Results for both the agarose only matrix and the O⁶CMG-agarose matrix are shown.

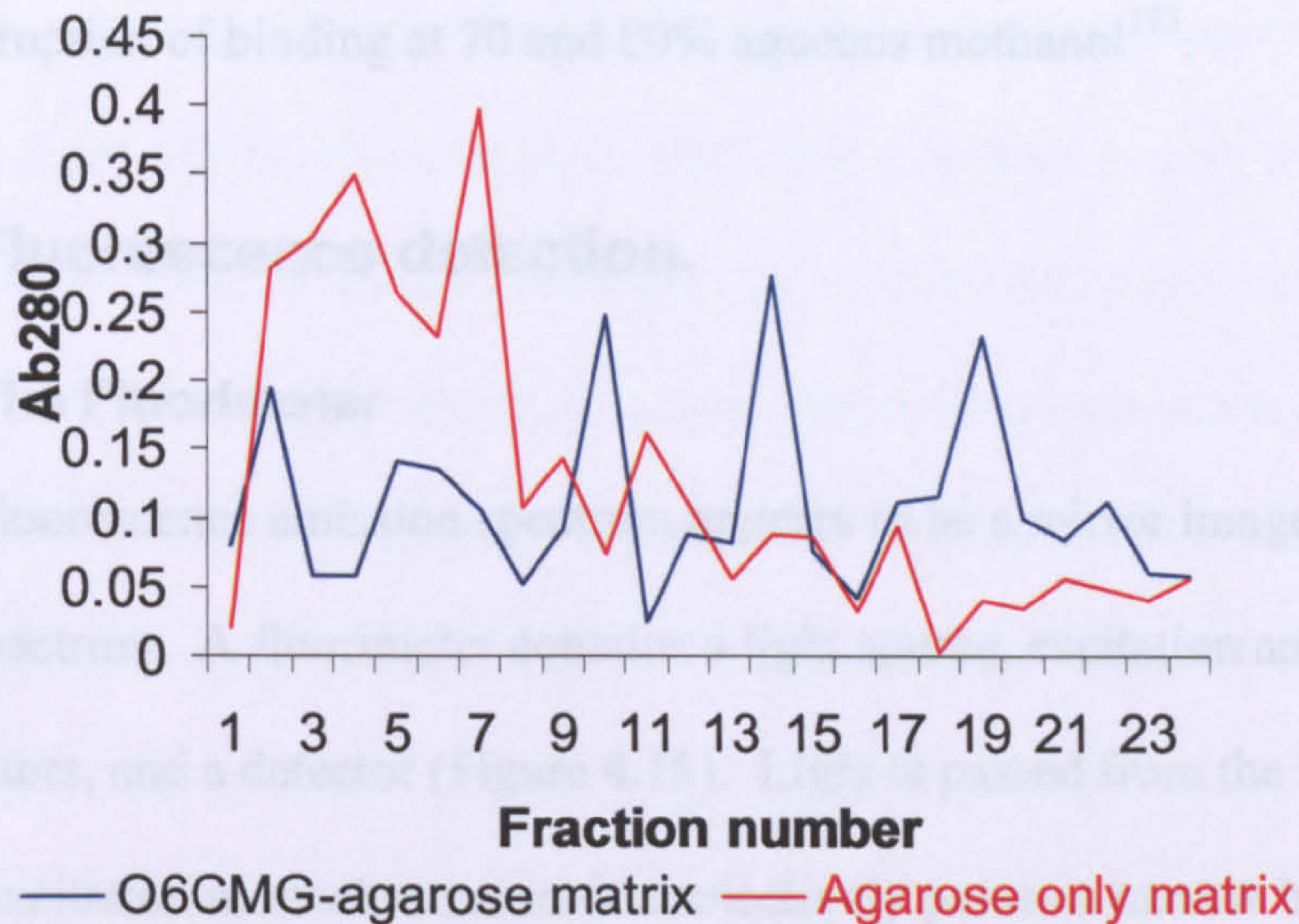


Figure 4-14. A graph to show the amount of absorbance at 280 nm, indicative of the amount of antibody being eluted from both types of column.

In the agarose only column the majority of the antibody washed out of the column in PBS, although some was washed out in the fraction containing 30% methanol and an even smaller amount at 50%. This suggests that some of the antibodies could be retained by another method for example size. The O⁶CMG column excluded small amount of the antibody early in the elution cycle although the majority was retained until higher methanol concentrations were reached. These antibodies were eluted later in three peaks around 20, 40 and 70 %. The antiserum used was a polyclonal mix in which different antibodies have been selected due to their ability to bind to the DNA adduct, however they are not identical and as so will have differing properties, it is possible that the three peaks represent three differing antibodies within the mix. This result compares with previously published results in which the antibody was covalently linked to the column matrix and a mixture of DNA bases, including

O⁶carboxymethylated guanine, was passed down the column. Results showed that there was disruption of binding at 70 and 80% aqueous methanol¹⁷⁵.

4.5 Fluorescence detection.

4.5.1. The Fluorimeter

Generally a fluorescence emission spectrum appears to be a mirror image of the absorption spectrum. A fluorimeter contains a light source, excitation and emission monochromators, and a detector (Figure 4.15). Light is passed from the light source through the excitation monochromator, this selectively passes a narrow band of the spectrum, focused around the excitation wavelength, to the sample. Light given off by the sample then passes through to the emission monochromator which again selects a narrow band of the spectrum, this then passes to the detector.

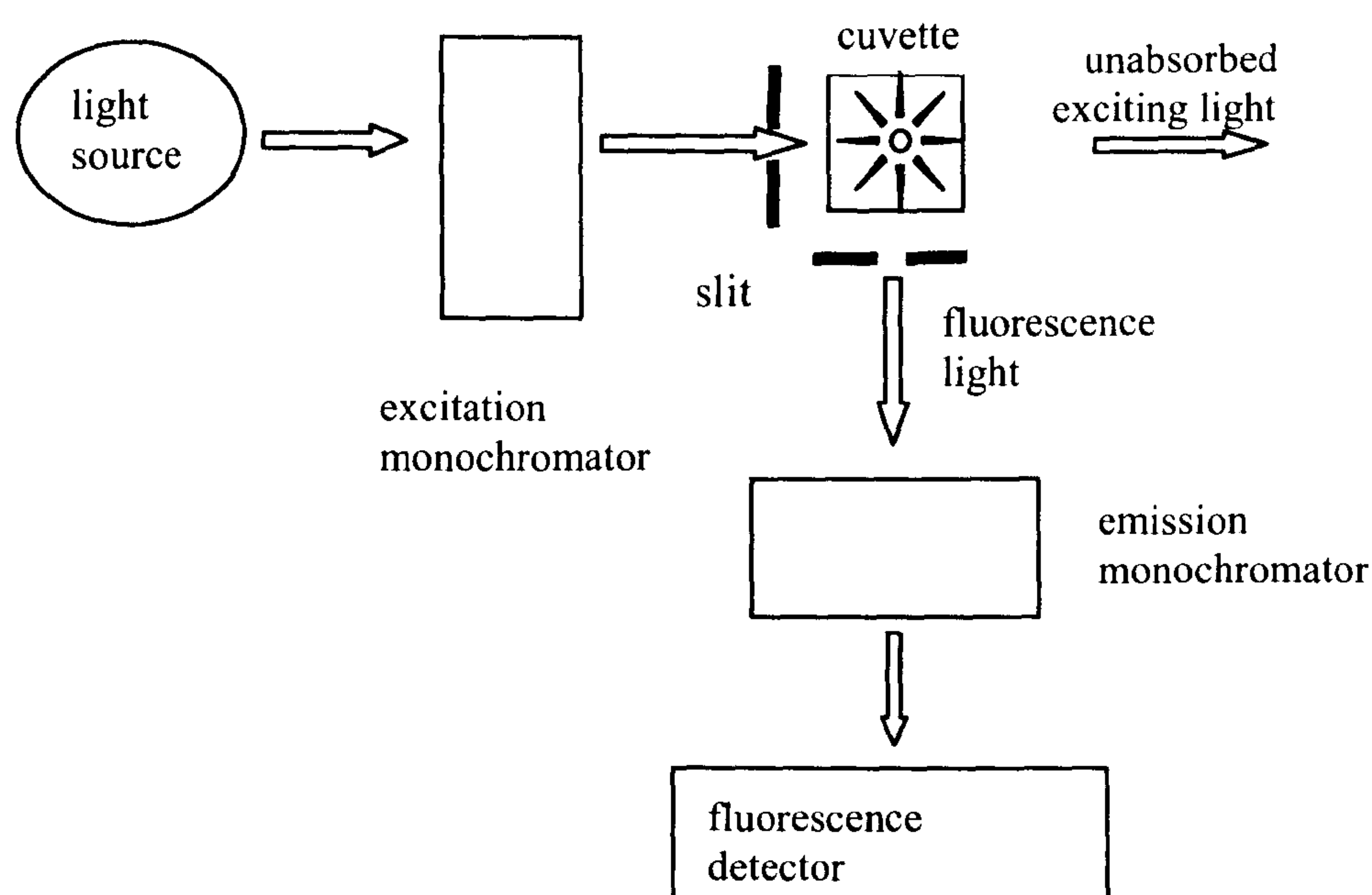


Figure 4-15. A diagram of the basic principles of the inside of a fluorimeter, containing a light source, monochromators and a detector, also indicates the position of the sample cuvette.

4.5.2 Quenching

Fluorescence quenching refers to any process which decreases the fluorescence intensity of a given substance including excited state reactions, energy transfer, complex formation and collisional quenching. One of the best known collisional quenchers is molecular oxygen which quenches almost all known fluorophores.

There is much debate of how this phenomena actually occurs but it is known that contact between the oxygen molecule and the fluorophore is required for quenching.¹⁴⁵

The effect of increasing concentration of methanol on the fluorescence emission of the pyrene tag in the library was examined (Figure 4.16). Significant quenching of fluorescence was seen at methanol concentration >50%.

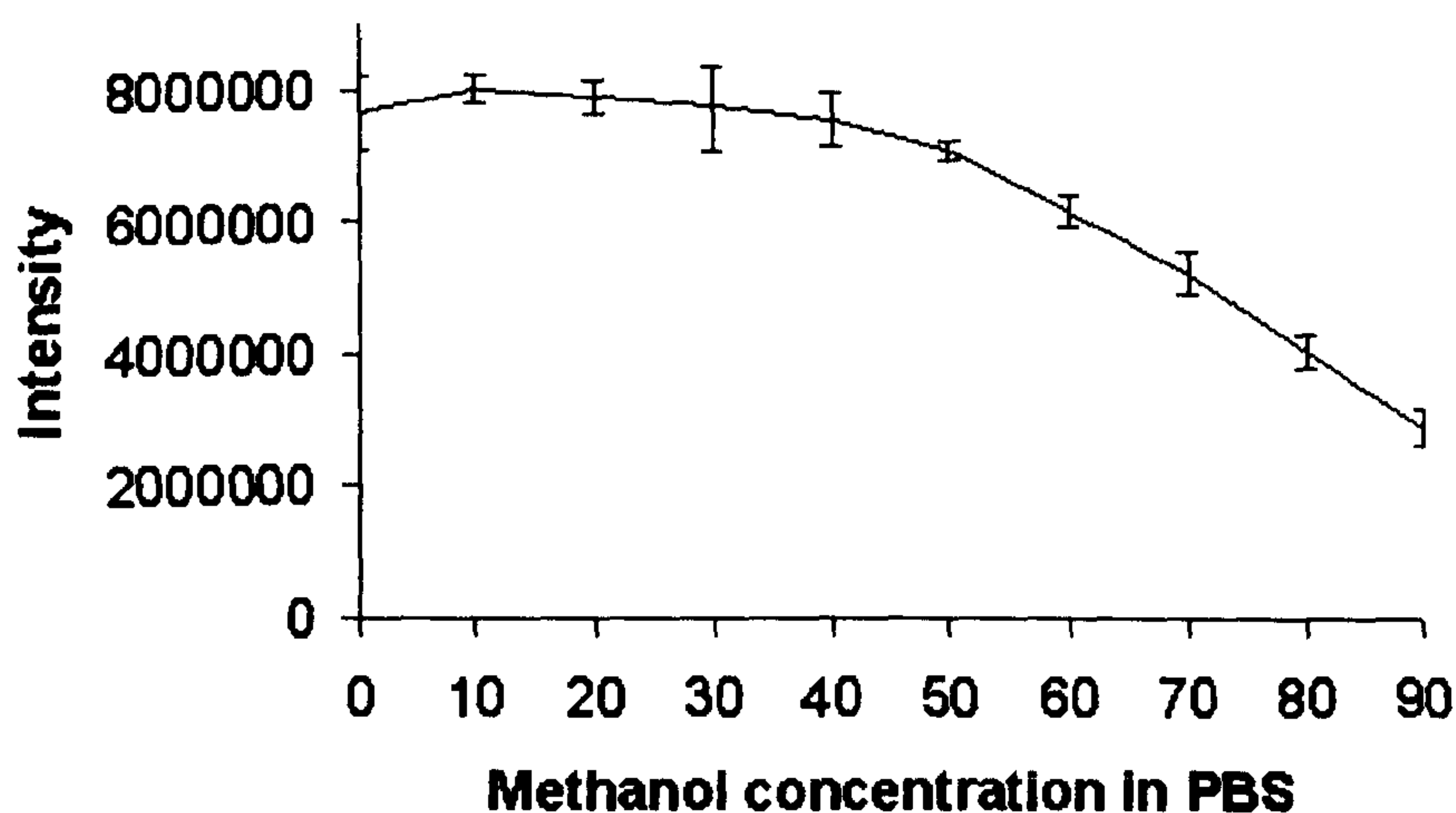


Figure 4-16. A graph showing the effect that increasing methanol concentration as a percentage has on the fluorescence intensity of the library.

4.5.3 Excitation and emission

In order to determine which wavelengths to monitor the following scans were conducted on different concentrations of the library, taking 3 mL of each sample, the scans were conducted at excitation of 342 nm and the emission spectra recorded between 350 and 450 nm.

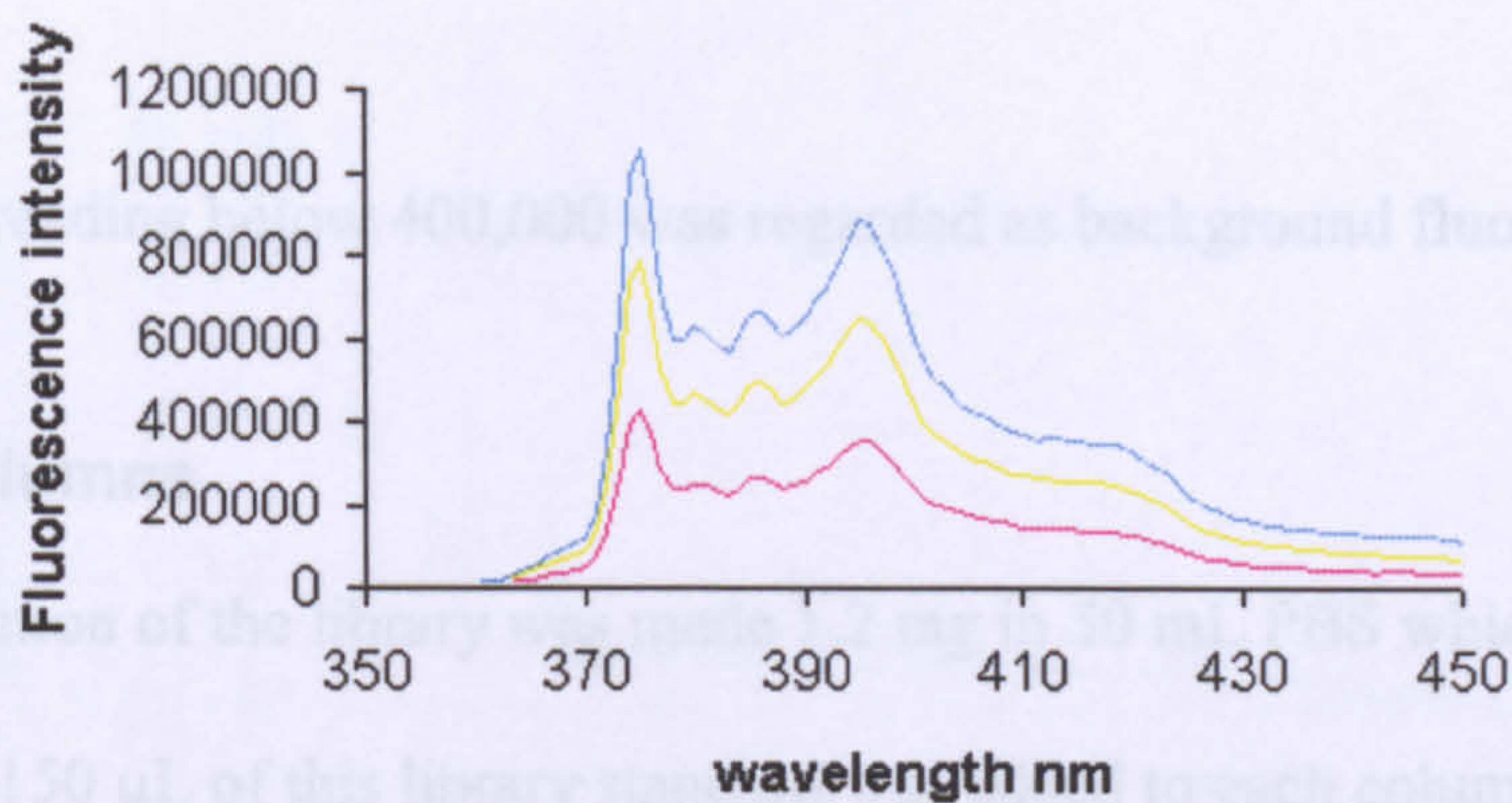


Figure 4-17. The emission spectra of three samples of the library. Each sample was 3 mL and the pink line represents 5 nmol mL^{-1} , the yellow line represents 10 nmol mL^{-1} and the blue line 15 nmol mL^{-1} .

It was decided to record with an excitation wavelength of 342 nm and an emission wavelength of 395 nm.

4.5.4 Background fluorescence

Before use all the columns were copiously washed with PBS to remove storage solution and then each underwent one cycle of the eluant system, the effects of this on the intensity of the fluorescence were recorded below.

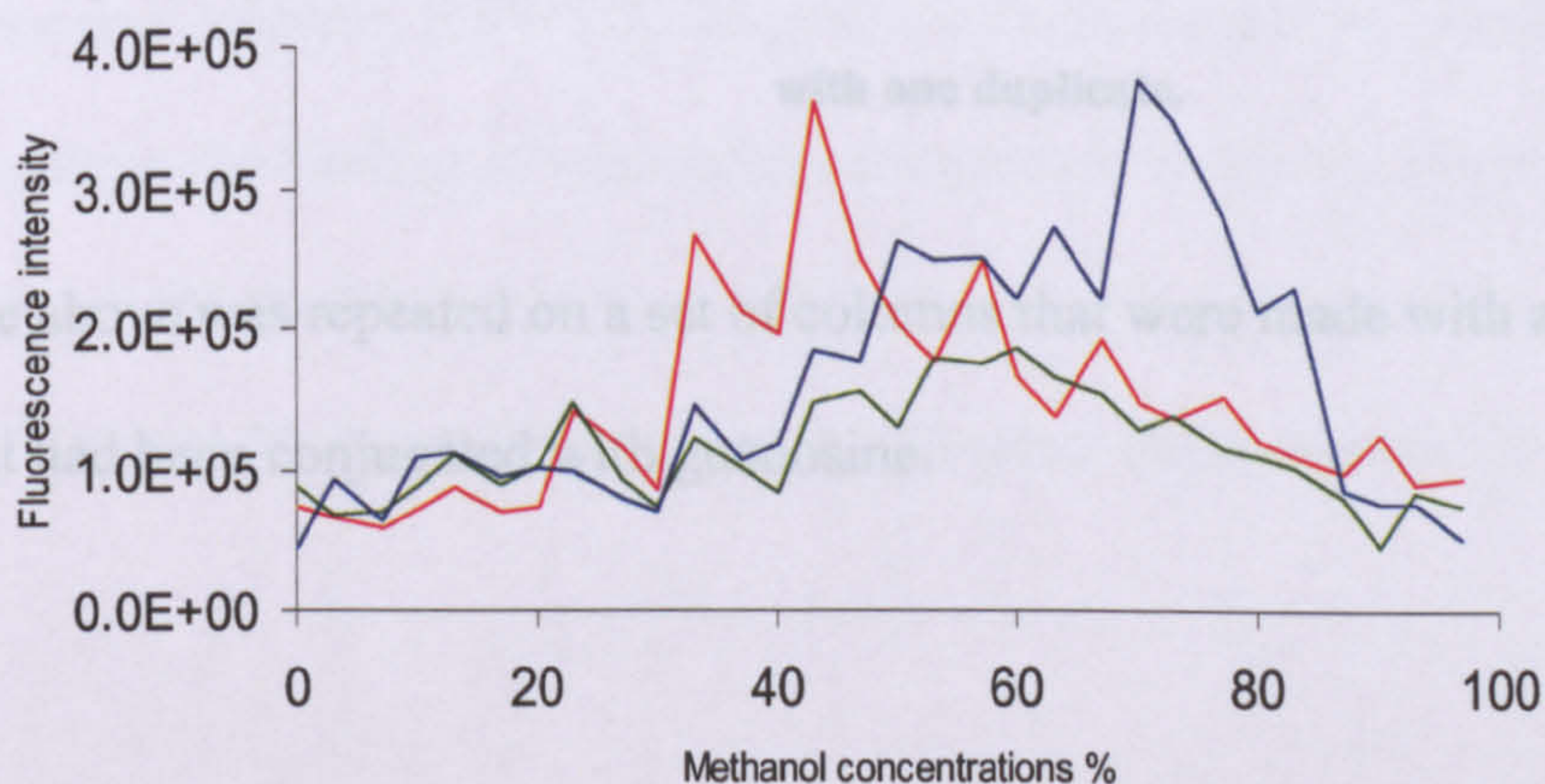


Figure 4-18. A graph showing the fluorescence intensity of the columns as they undergo a cycle of increasing methanol concentrations. Blue represents an agarose only column, red represents an agarose-guanosine column and green represents an agarose- O^6CMG column.

Therefore any reading below 400,000 was regarded as background fluorescence.

4.5.5 Columns

A standard solution of the library was made 1.2 mg in 50 mL PBS which equals $0.02 \text{ nmol } \mu\text{L}^{-1}$ and $150 \mu\text{L}$ of this library standard was added to each column containing an unmodified agarose matrix. These columns then underwent one cycle of the eluant system.

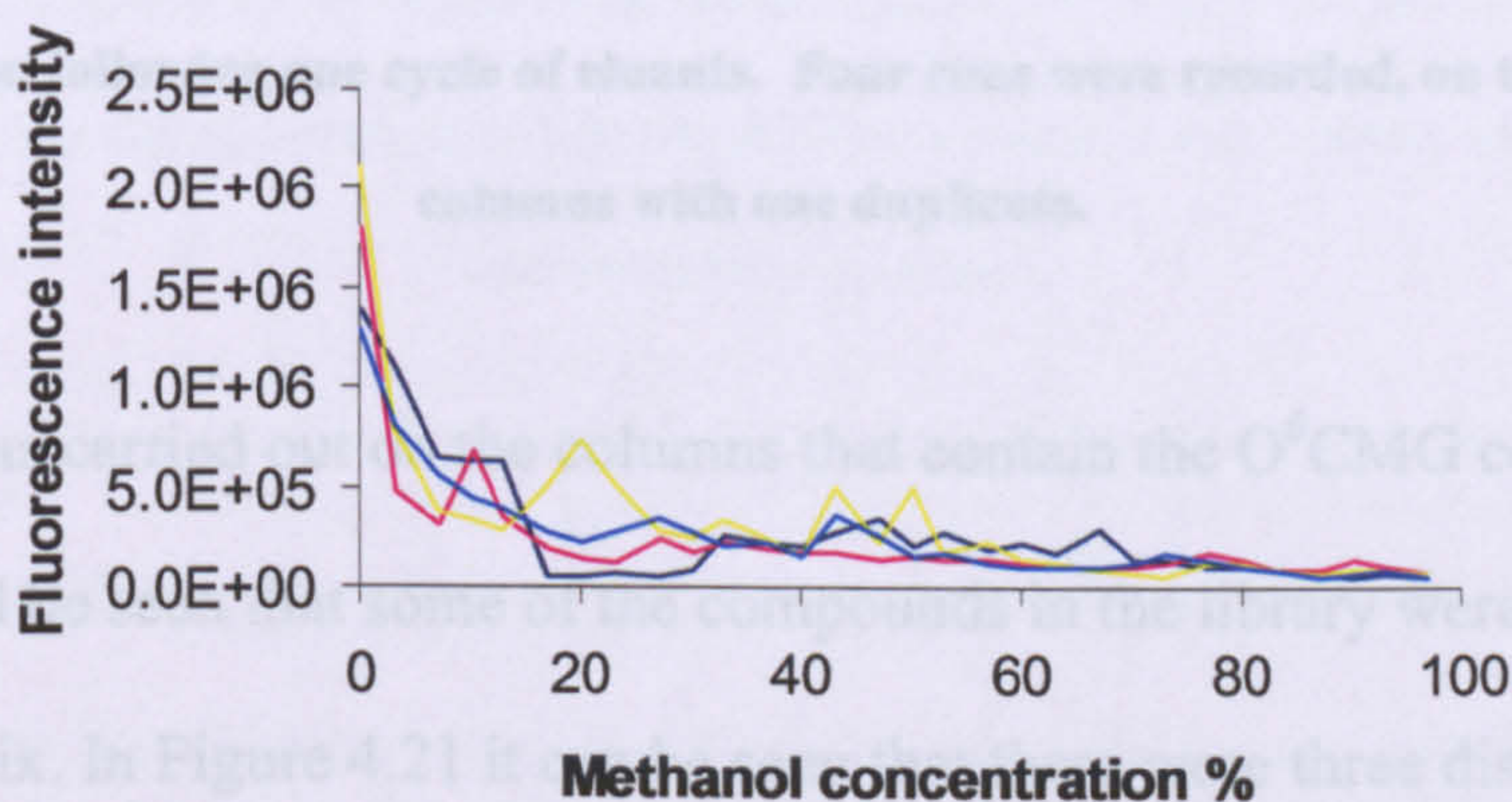


Figure 4-19. Elution of fluorescence from unmodified agarose after the addition of $150 \mu\text{L}$ of library solution following one cycle of eluants. Four separate runs, on three different columns with one duplicate.

The above was repeated on a set of columns that were made with an agarose matrix that had been conjugated with guanosine.

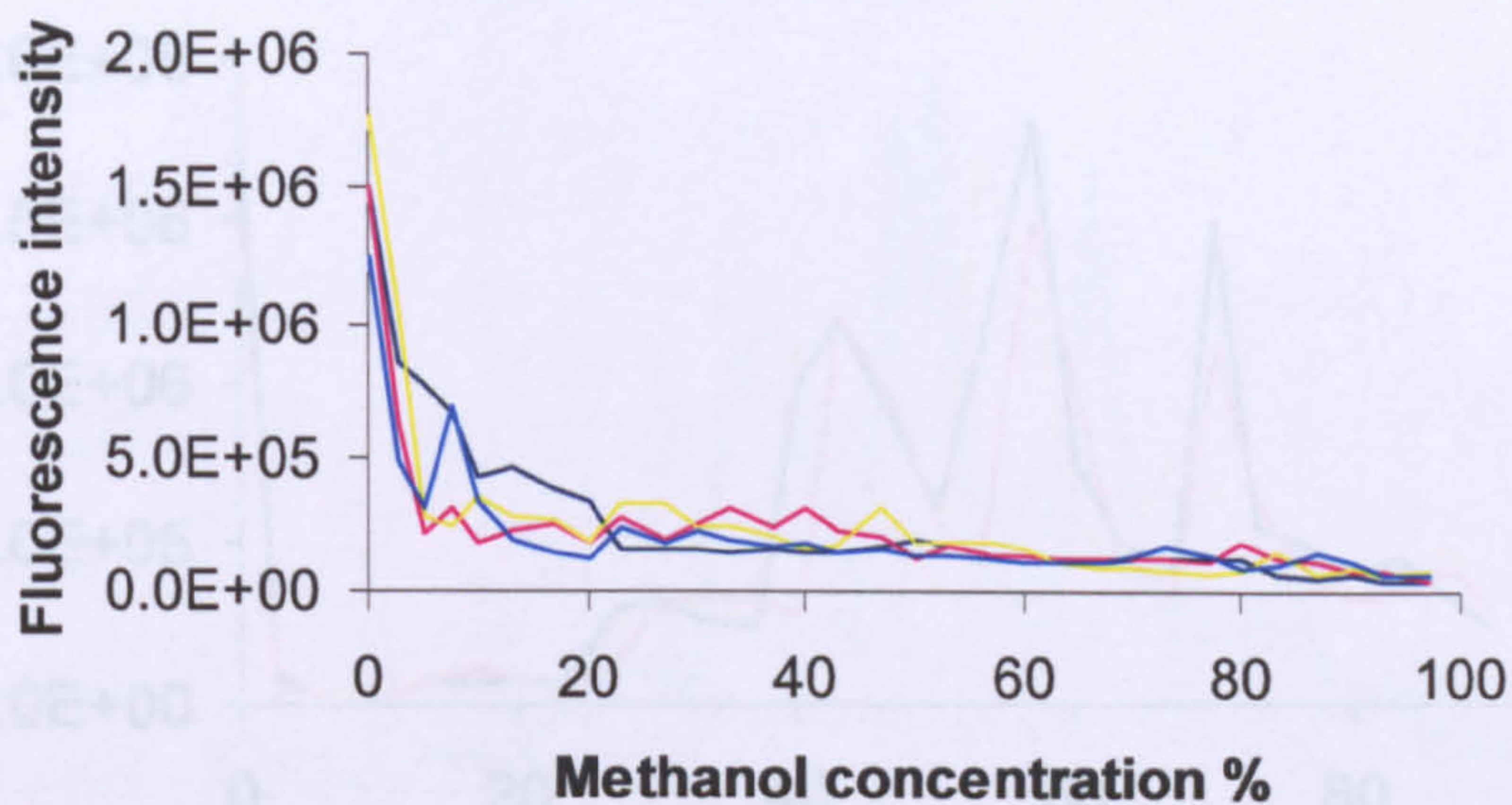


Figure 4-20. Elution of fluorescence from guanosine-agarose after the addition of 150 μ L of library solution following one cycle of eluants. Four runs were recorded, on three different columns with one duplicate.

When above was carried out on the columns that contain the O⁶CMG containing agarose it could be seen that some of the compounds in the library were retained within the matrix. In Figure 4.21 it can be seen that there were three discrete peaks that were eluted after 40% methanol concentration had been introduced on to the column. As these peaks do not appear in either the agarose only or in the agarose-guanosine columns, the O⁶CMG-agarose column must have inherent properties that encourage the binding of some of the compounds within the library which are then eluted as the methanol increases which changes the polarity of the eluting solution and disrupts the intermolecular forces that retain the compound in the matrix.

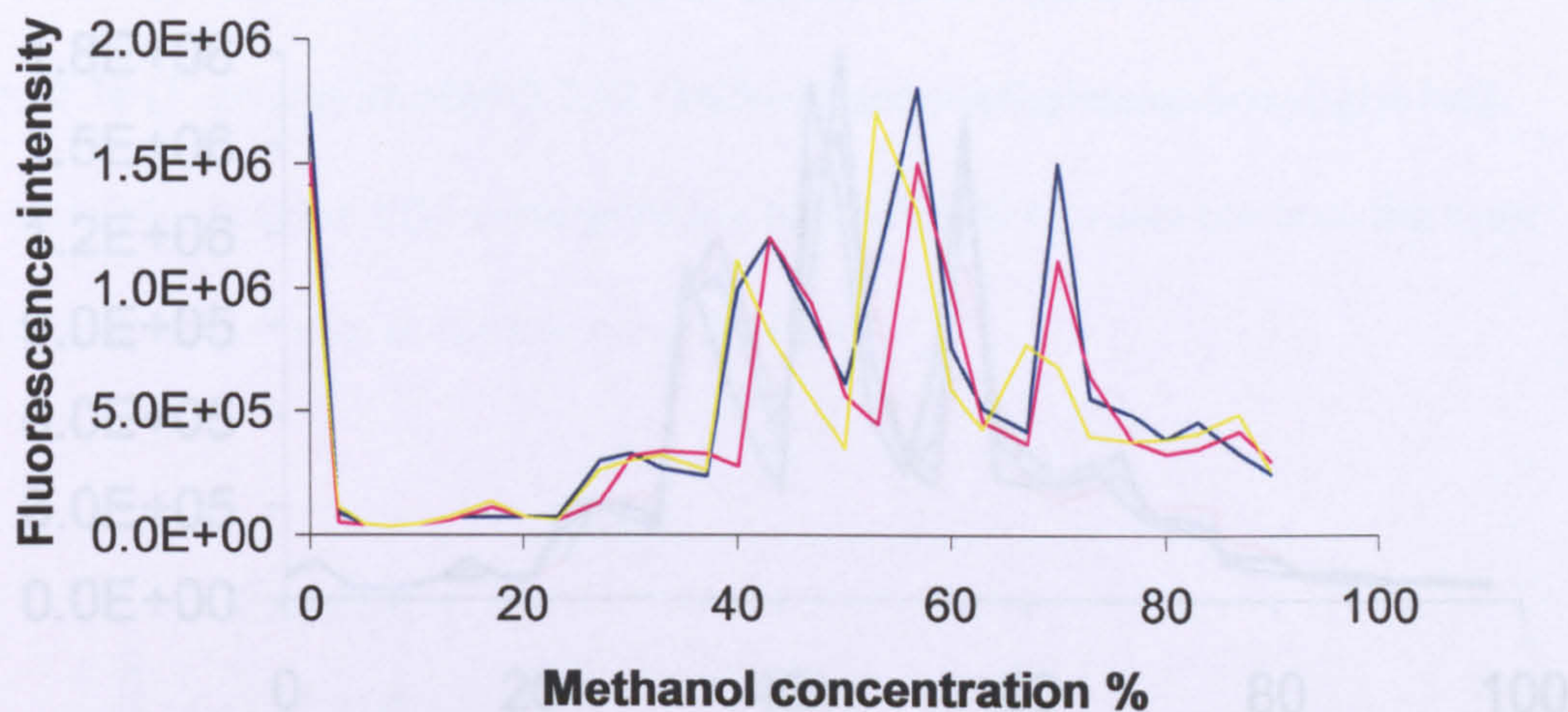


Figure 4-21. Elution of fluorescence from three different columns with the O⁶CMG-agarose matrix following the application of 150 μ L of library solution and elution with increasing methanol concentration.

As compounds within the library could be showing affinity to the carboxylate ion on the ligand and exhibiting no other selection processes, it was decided to add 1 mM sodium acetate to the eluant which would effectively remove any compounds that are interacting solely on an ionic basis. The results are shown in Figure 4.22.

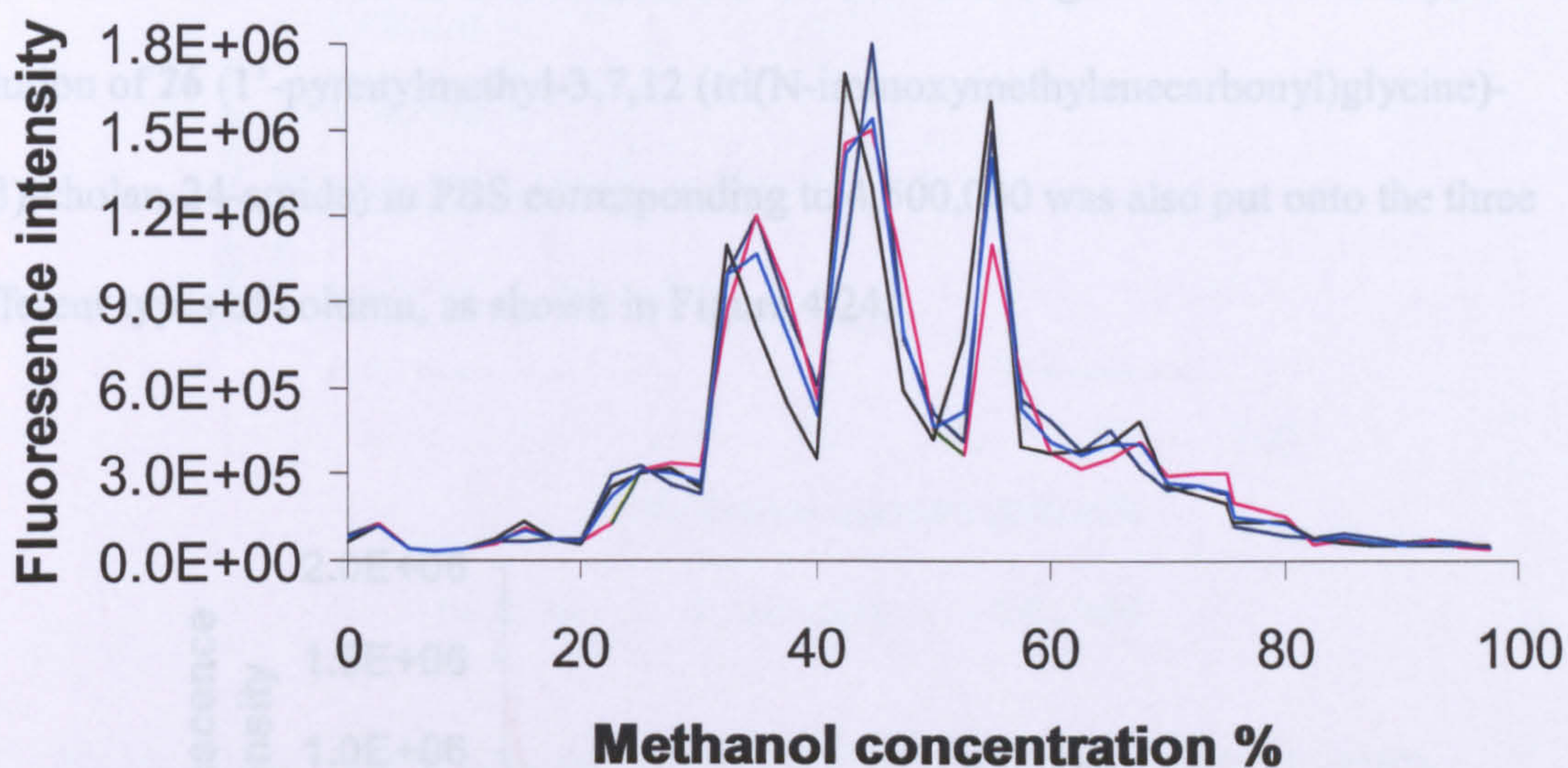


Figure 4-22. A graph showing the results of three separate O⁶CMG columns in which 150 μ L of the library solution was added to each column and then they were eluted with an increasing methanol gradient. The fluorescence intensity of aliquots of the eluant were recorded. The pale blue line represents an average of the three.

In the presence of sodium acetate the peaks eluted from the column had the same profile, with the three discrete peaks being evident. However they appear to be eluted at a lower concentration of methanol with the earliest peak coming off the column at 30-35% methanol compared to 40-45% in the columns that were run without sodium acetate. Therefore, it can be concluded that the presence of sodium acetate does not appear to affect the binding of compounds in the library to the matrix but it appears to have an influence in the release of these compounds from the matrix.

In the previous column experiments 150 μ L of a dilute solution of the library was added to each column, giving a fluorescence count of approximately 3,000,000. A solution of **29** (1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl)arginine)-(5 β)-cholan-24-amide) in PBS corresponding to 2,800,000 fluorescence units was

added to the three different columns, results are shown in Figure 4.23. Similarly, a solution of **26** (1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl)glycine)-(5 β)-cholan-24-amide) in PBS corresponding to 4,500,000 was also put onto the three different types of column, as shown in Figure 4.24.

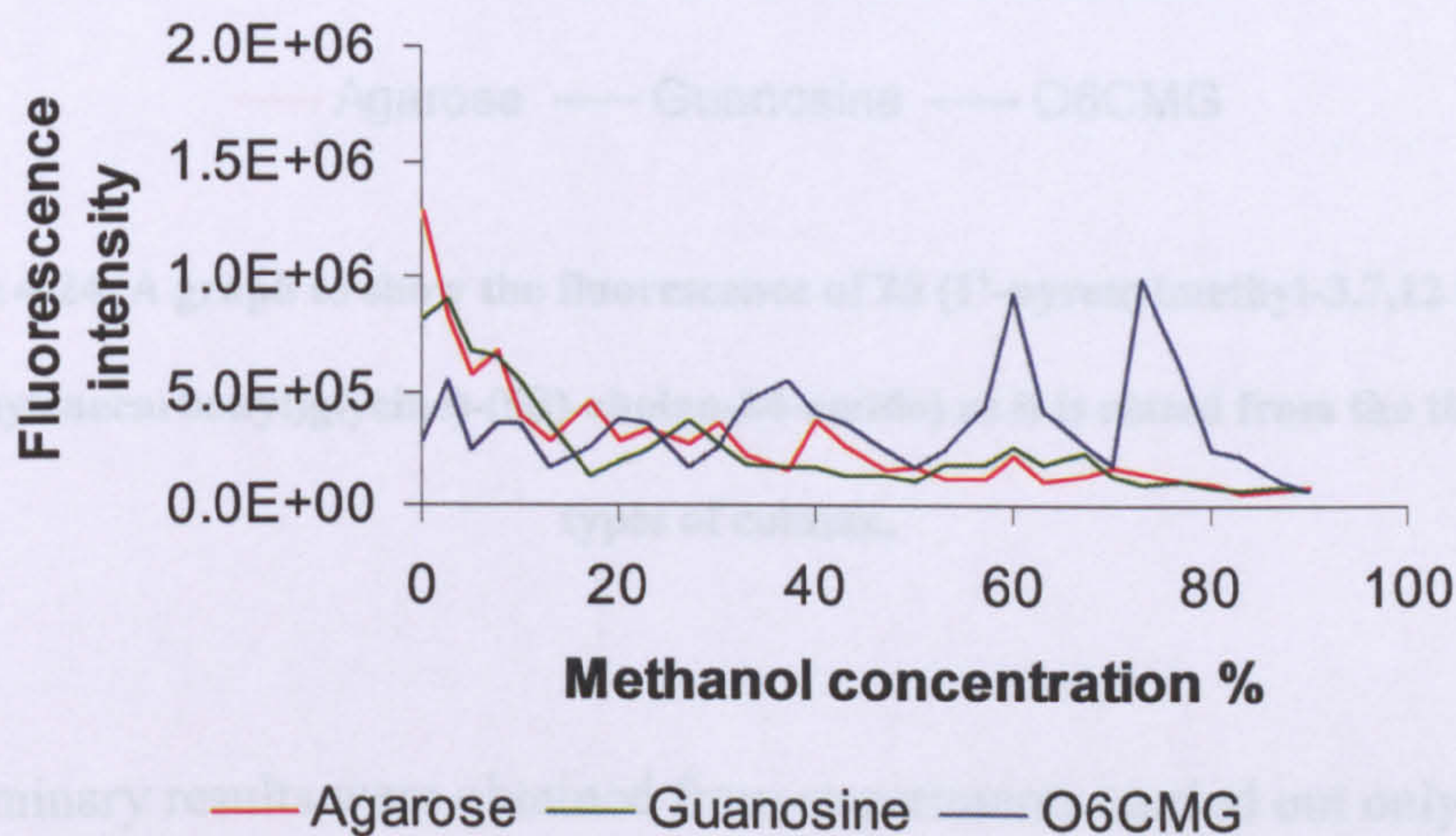


Figure 4-23. A graph to show the fluorescence intensity of **29** (1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl)arginine)-(5 β)-cholan-24-amide) as it is eluted from the three different columns.

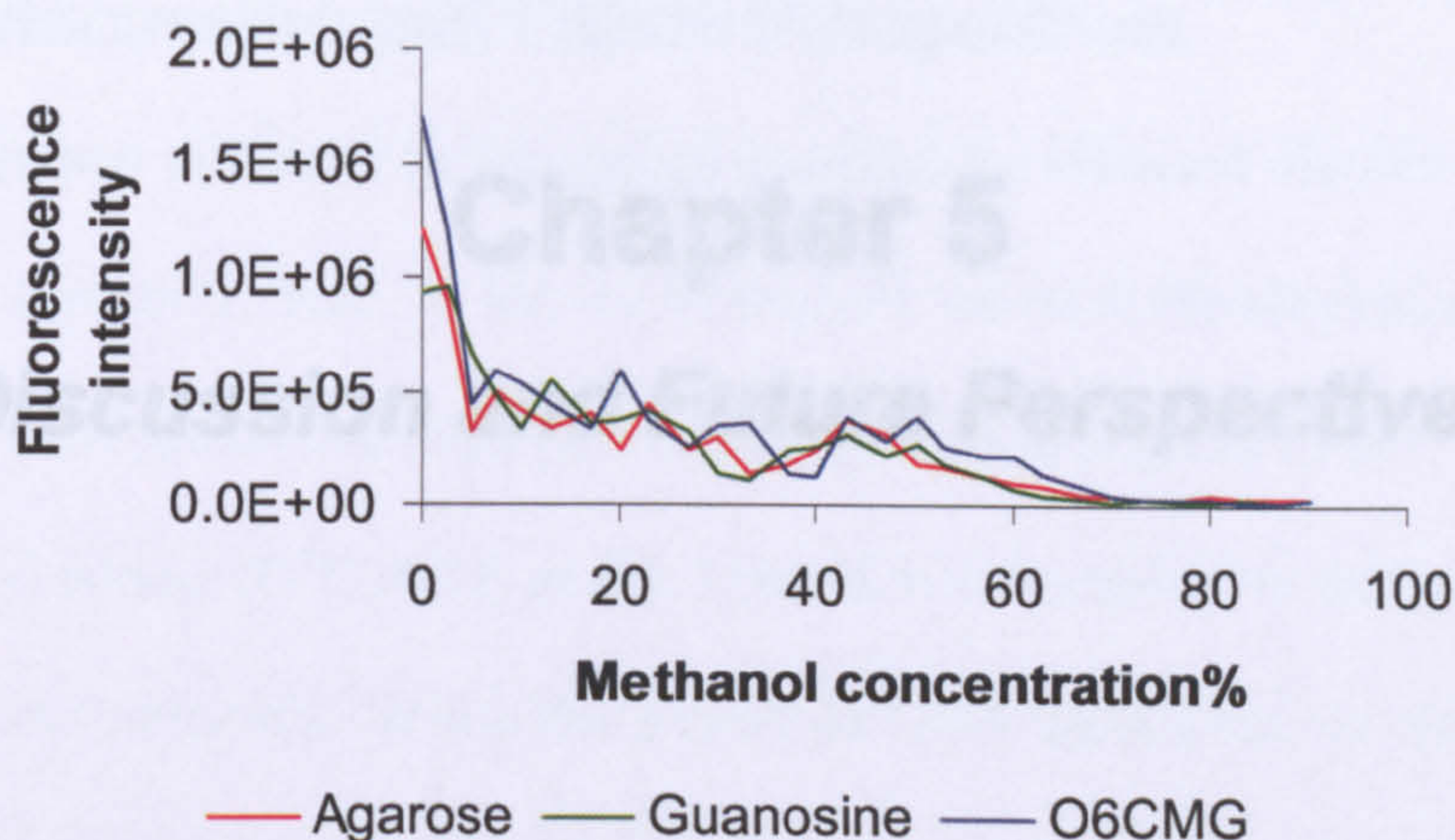


Figure 4-24. A graph to show the fluorescence of 26 (1'-pyrenylmethyl-3,7,12 (tri(N- iminoxymethylenecarbonyl)glycine)-(5 β)-cholan-24-amide) as it is eluted from the three different types of column.

These preliminary results were obtained from experiments carried out only once.

However, it appears that **26** (1'-pyrenylmethyl-3,7,12 (tri(N- iminoxymethylenecarbonyl) glycine)-(5 β)-cholan-24-amide) was not retained on any of the column types. **29** (1'-Pyrenylmethyl-3,7,12 (tri(N- iminoxymethylenecarbonyl)arginine)-(5 β)-cholan-24-amide) did not bind to the agarose and the agarose-guanosine columns. Interestingly, there were two distinct peaks for the O⁶CMG-agarose column with the arginine derivative of PyCA, eluted at 60 and 75% methanol concentrations. These results are further evidence for the importance of arginine moieties in the recognition of O⁶CMdG.

Chapter 5

Discussion and Future Perspectives

5.1 Discussion and Future Perspectives

In the development of cancer in the oesophagus and the stomach alkylation of the DNA plays a significant role. A known mutagenic lesion is the alkylation of the O⁶ position of 2'-deoxyguanosine, and particular importance is given to the carboxymethyl adduct (O⁶CMdG) as this appears to be resistant to the usual biological repair pathways. While this particular DNA adduct can be effectively recognised by antibodies there are a finite amount available for use. It would therefore be advantageous if a novel affinity reagent could be synthesised so that an unlimited amount would be available from the chemical laboratory.

As antibodies are large complex macromolecules with the site of binding being a consequence of the whole structure it was thought that it would be possible to identify amino acids contributing strongly to molecule recognition and to synthesis a library of compounds based on these. The phage display technique was used to this end. It was initially thought that the phage display would give a clearer indication of the exact amino acids present. From the sequences identified in phages that bound to immobilised O⁶-CMdG there was evidence that a binding pocket required contribution from positively charged amino acids and the presence of aromatic side chains.

Based on previously published work from a number of groups cholic acid was used as a scaffold molecule to host not only the diversity required by a library but also a highly fluorescent pyrene molecule to be used to detect binding. The pyrene was attached through an amide linkage that was not affected by subsequent reactions.

Diversity around the scaffold was via the reaction of acetyl amino acid derivatives with keto groups at the 3, 7 and 12 positions. The synthesis of the amino acid derivatives was no means trivial and much time was spent trying to improve the methods and yields. However the use of the oxime chemistry binding to the scaffold was a straightforward procedure. By refluxing in ethanol overnight all three keto functional groups reacted, in a simple and reproducible reaction, in the presence of sodium acetate as a base. Having made a library mass spectrometry confirmed that the library did indeed contain the expected molecular ions and fragments. HPLC and then 2D-TLC methods were used to establish that a range of compounds with varying properties were present in the library.

An affinity column was made in which the adducted DNA was immobilised on to the column matrix, the preparation of this matrix was carefully considered so that the resulting exposed moiety had a similar structure to O⁶CMdG. These columns were tested against a previously prepared antibody and it was shown that the matrix coupled with O⁶carboxymethylguanosine did in fact retain the antibody added to the column and that these antibodies would elute at increasing levels of methanol.

Affinity columns have been used previously¹⁷⁵ in which the antibody was immobilised within the matrix, hydrolysed DNA passed down the column and a similar range of methanol percentage was used to elute off the adducted DNA. This result confirmed that the columns had O⁶carboxymethyldeoxyguanosine-like residue capable of retaining known binding antibodies.

These columns were then used to probe the library of compounds that had been synthesised. Results showed that there are compounds within the library which bound

strongly to the O⁶carboxymethylguanosine-agarose matrix and not to the control agarose or guanosine-agarose columns. As the methanol concentrations increased three distinct peaks were detected by fluorescence and these were not due solely to the presence of a negative charge on the O⁶carboxymethylguanosine-agarose matrix as the result was repeatable in the presence of sodium acetate. Finally the columns were tested with trisubstituted compounds, the trisubstituted glycine derivative passed through both control columns and the O⁶carboxymethylguanosine-agarose column. However the trisubstituted arginine derivative was strongly retained on the O⁶carboxymethylguanosine-agarose matrix. These preliminary results show that the arginine derived compound shows high affinity for the negatively charged carboxymethyl residue of O⁶CMdG.

Ideally a single fraction from the O⁶carboxymethylguanosine-agarose column would be collected and the compound/s within the peak could have been identified and from this individual compounds synthesised and tested for their affinity for O⁶-carboxymethyl-2'-deoxyguanosine. Such an approach could involve repeating the O⁶carboxymethylguanosine-agarose column experiment but modifying the elution method by washing nonbinding and weakly bound compounds with copious amounts of 60% methanol before increasing the methanol concentration to limit the number of different compounds coming off the column. However it maybe more useful to do this on a larger column, say 5 mL, and have the option of loading more of the library on to the column and hopeful have enough material for mass spectrometric identification. This would give information on the groups present on the scaffold but not the position.

For the position of each of the three groups could be determined by selectively synthesising each of the three keto groups. It has been shown that a single position, possibly C3, reacts first under mild conditions. The other two sites, possibly C7 and C12, are not as easily controlled. However, this can be overcome by using either 5 β -cholanic acid 3,7 dione, 12 hydroxy or 5 β -cholanic acid 3,12 dione, 7 hydroxy starting materials available from Steraloid Inc. and manipulating the synthesis accordingly.

Finally, consideration needs to be given to the orientation of the groups with regard to the oxime bond, as each of these could be in either the *E* or *Z* form. The predominant isomer will be selected from the library by affinity for the ligand, however for future synthesis the correct isomer will need to be characterised. There are many references to how X-ray crystallography has given detailed information about the structure of cholic acid derivatives, e.g. the configuration at C25 in 3 α ,7 α ,12 α -trihydroxy-5 β -cholestan-26-oic acid was unequivocally confirmed following TLC purification, the growth of crystals and X-ray crystallography¹⁷⁶. If the reagent could be induced to co-crystallise with the ligand moiety, again X-ray crystallography could be used to determine the precise structural arrangement¹⁷⁷. If enough of the correct compound can be isolated the spatial arrangement of the H and C atoms adjacent to the oxime could be explored by NMR. Another consideration could be to reduce the oxime to an amine and explore if this has any effect on the binding properties, either positively or negatively.

The main objective of this project was to identify a novel affinity reagent for O⁶CMdG and although this was not accomplished much has been achieved with

regards to developing a novel approach. The methods laid down by this project have the potential to be used for other DNA adducts. If a large cholic acid based diversity library can be synthesised which is optimal for recognition of alkylated DNA bases, the affinity approaches for isolation of candidate reagents can be applied to many other modified DNA bases.

Chapter 6

Experimental

6.1 General and Instrumentation

6.1.1 Physical Measurements

Infra-red spectra were recorded as thin films in nujol mulls or as potassium bromide discs using a Perkin-Elmer 1710 infrared fourier transform spectrometer. Ultra violet/visible spectra were run in solvents stated on a Uvikon XL spectrophotometer (BioTek Instruments). Nuclear magnetic resonance spectra, in the solvents indicated, were recorded at 300 MHz on a Jeol JNM-Lambda 300 spectrometer, with an internal standard of tetramethylsilane. Elemental analyses were supplied by Medac Ltd. at Brunel Science Centre, Surrey. Mass spectra were obtained by Mr Graham Jeffs using a VG Quattro instrument in both positive and negative electrospray (ES) mode. All melting points were measured on a Electrothermal digital melting point apparatus. Thin layer chromatography was carried out on Machery-Nagel Polygram SIL G/UV254 plates on a plastic backing, Fluka silica gel aluminium backed plates or Machery-Nagel pre-coated glass backed TLC plate SIL G25, as indicated.

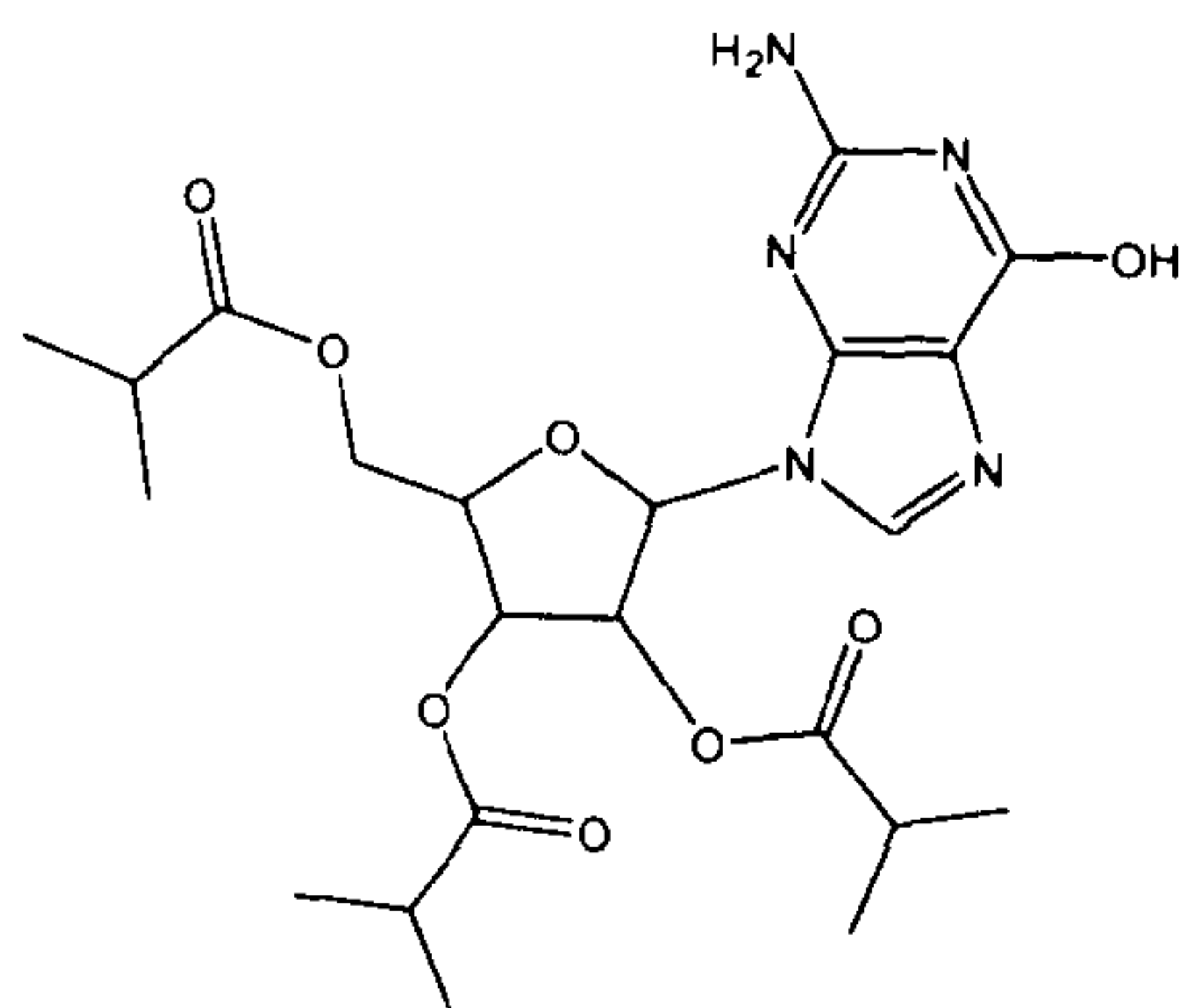
6.1.2 Reagents and Reactants

Chemicals were used as supplied unless otherwise stated and were obtained as follows. Guanosine, mesitylenesulphonyl chloride, 1, 8-diazabicyclo [5.4.0]-undec-7-ene (DBU), and hydroxylamine hydrochloride from Acros. Isobutyric anhydride and methyl glycolate were supplied by Fluka. Dimethylaminopyridine (DMAP), triethylamine, quinuclidine, amino propyl glass beads, lithium aluminium hydride, pyridinium dichromate, *N*-hydroxysuccinimide, pyrenylmethylamine hydrochloride, methylhydroxylamine hydrochloride, carboxymethylhydroxylamine hydrochloride benzylhydroxylamine hydrochloride *O*-carboxymethyl hydroxylamine hemi hydrochloride, *N,N*-dicyclohexylcarbodiimide, glycine benzyl ester, glycine methyl

ester, phenylalanine methyl ester, piperidine and sodium cyanoborohydride were purchased from Aldrich. Tetrahydrofuran, anhydrous sodium sulphate, sodium carbonate, di-*tert*-butyl dicarbonate, sodium acetate trihydrate and triethylamine were supplied by Fisher. Sigma supplied cholic acid and 5 β -cholanic acid-3,7,12, trione (dehydrocholic acid) was purchased from Steraloid Inc, Newport, USA. The more unusual amino acids N^G-Nitro-arginine methyl ester hydrochloride and N- ϵ -Fmoc-lysine methyl ester hydrochloride were supplied by NovaBioChemTM.

6.2 Synthesis of DNA Adduct and conjugation to carrier matrices

6.2.1 2',3',5'-tri-*O*-isobutyryl guanosine (2)



Guanosine (5g, 18.25mmol) was dried, twice, by the evaporation of toluene. This was then suspended in dry tetrahydrofuran (50ml), to which isobutyric anhydride (9.7 mL, 58.2mmol) and DMAP (100mg, 0.8mmol) was added, the resultant mixture was stirred, at room temperature, overnight. Sodium bicarbonate was dissolved in distilled water and added to the reaction mixture, the mixture was maintained above pH 7.5 by the continuous addition of sodium bicarbonate. The mixture was stirred for two hours and then filtered, washed three times with distilled water and then dried under vacuum. The product was recrystallized by suspending in dichloromethane and

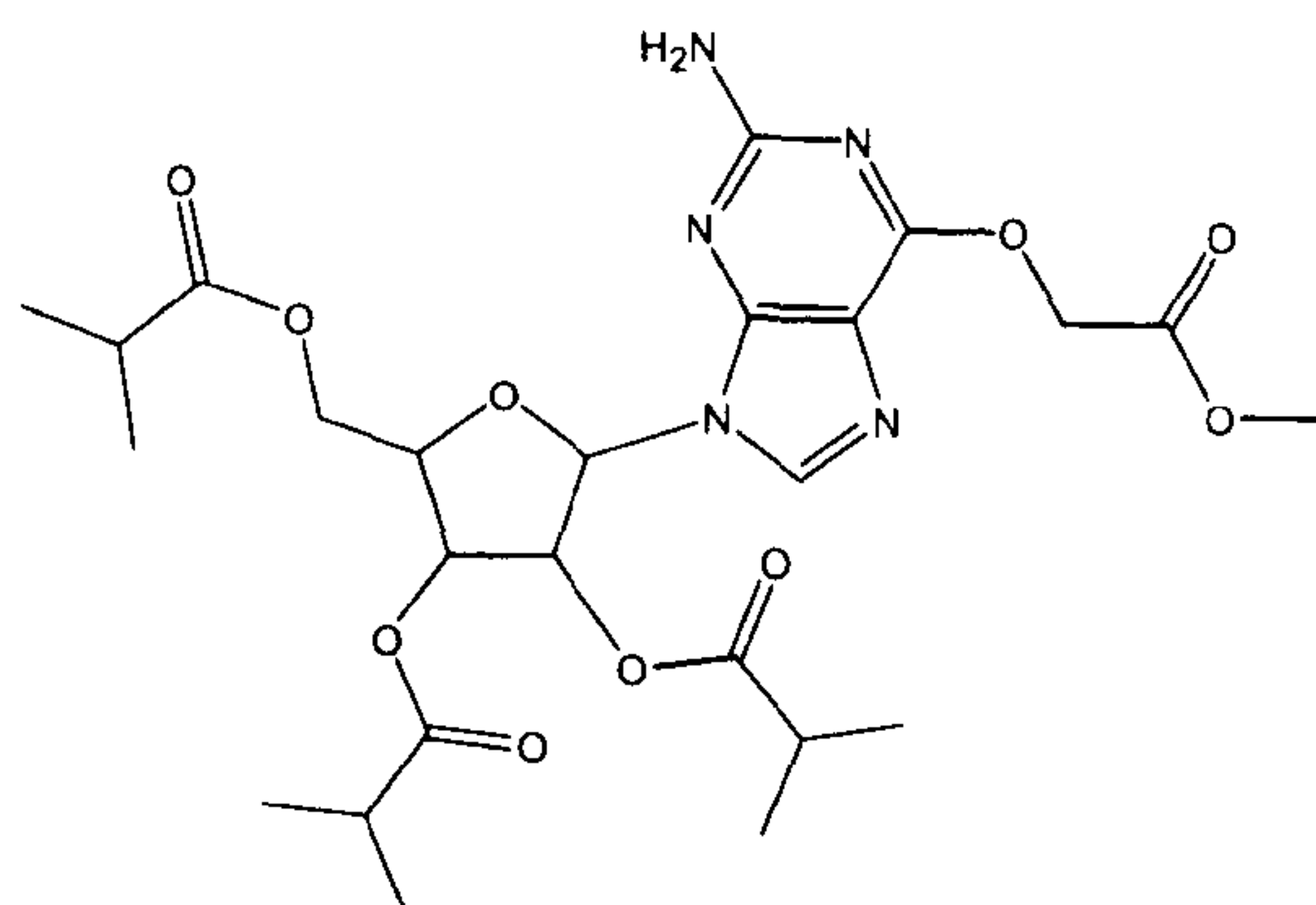
adding enough methanol to dissolve the compound, this was gently heated and agitated. The solution was then placed in the freezer with a few drops of hexane. After the product had crystallised, the solvent was removed by pipette and the crystals blown dry with nitrogen gas.

Yield 8.5g (74%); mp 232-234 °C; λ_{\max} 248 nm; MS m/z (ES+) 494 (M + 1) 100, 343 (M - guanine + 1) 40; MSMS m/z (ES+) 494 (M + 1) 31, 343 (M - guanine + 1) 100, 167 (M - guanine - COCH(CH₃)₂ + 1) 85, 97 (M - guanine - (CH₂OCOCH(CH₃))₂ - COCH(CH₃)₂ + 1) 68, 70 (COHC(CH₃)₂) 68, ¹H NMR δ 1.3 (m, 18H, (CH(CH₃)₂)), 2.5 (m, 3H, (CH(CH₃)₂)), 4.45 (m, 1H, 1'-H), 5.60 (m, 1H, 3'-H), 5.95 (t, 1H, 2'-H), 6.05 (d, 1H, ³J 5.1, 1'-H), 7.75 (s, 1H, NCH).

6.2.2 **O⁶-[(2,4,6-trimethylphenyl)sulphonyl]-2',3',5'-tri-O-isobutyrylguanosine**

2 (1.0 g, 2.4 mmol) was suspended in dry dichloromethane, to this mesitylenesulphonyl chloride (1.05 g, 4.8 mmol) and DMAP (40 mg, 0.32 mmol) was added. This was followed by the dropwise addition of triethylamine (1 mL, 7 mmol), the resulting solution was left stirring at room temperature. The reaction followed by TLC (2% methanol in dichloromethane) and after 1 hour was used in the subsequent reaction without purification.

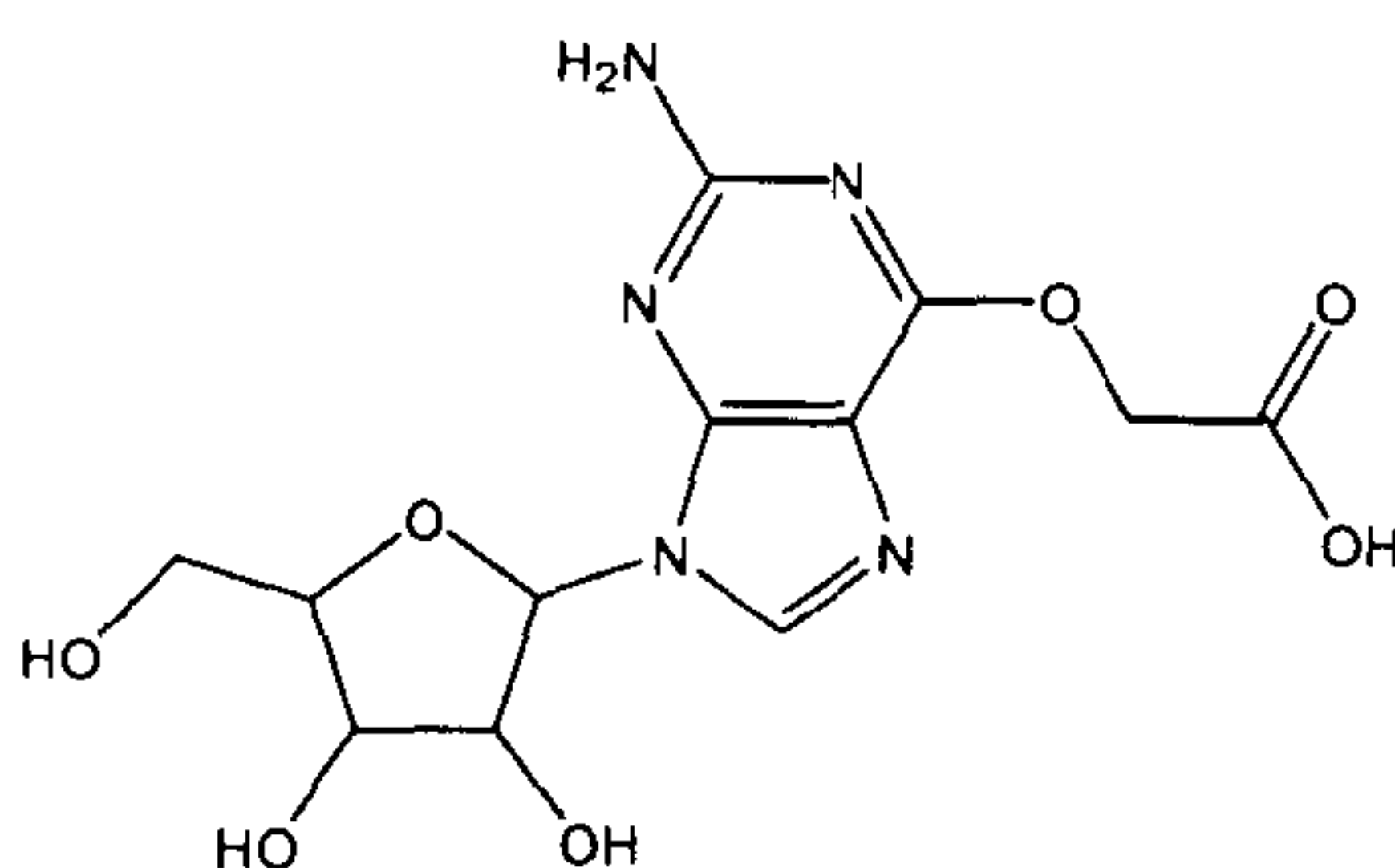
6.2.3 O^6 -(carboxymethyl, methyl ester)-2',3',5'-tri-*O*-isobutyrylguanosine



To the reaction vessel above quinuclidine (1g, 10mmol) was added and after a further 1½ hours there was full conversion to a baseline blue fluorescent spot (presumably the quinuclidinium salt). Methyl glycolate (1ml, 14mmol) and DBU (1ml, 6mmol) was added to the reaction mixture and the reaction was left stirring overnight, at room temperature. The resulting mixture was rotary evaporated for 3 hours, then dissolved in 10ml of dichloromethane and separated using flash chromatography with methanol and dichloromethane (98:2 v/v). An automated 5ml fraction collection was set up and each fraction was run through a normal phase LC (Varian 5060 ternary gradient LC system). Varian 5060 Diode array detector was used with a 25cm x 4.6 mm id Phenomenex LUNA C18 (2) + 3cm matching guard column at 30 °C, with a 50µl sample loop. The product containing fractions were pooled and rotary evaporated to dryness to give a pale yellow solid, which was then recrystallised from MeOH and hexane.

Yield 1.05g (93%); mp 120-122 °C; λ_{\max} 248, 284 nm; $^1\text{H NMR}$ δ 1.1 (m, 18H, $(\text{CH}(\text{CH}_3)_2)$), 2.5 (m, 3H, $(\text{CH}(\text{CH}_3)_2)$), 3.67 (s, 3H, OCH_3), 4.11 (s, 2H, $\text{OCH}_2\text{OOCH}_3$), 4.45 (m, 1H 1'-H), 5.69 (m, 1H, 3'-H), 5.85 (t, 1H, 2'-H), 5.96 (d, 1H, ^3J 5.1, 1'-H), 7.74 (s, 1H, NCH).

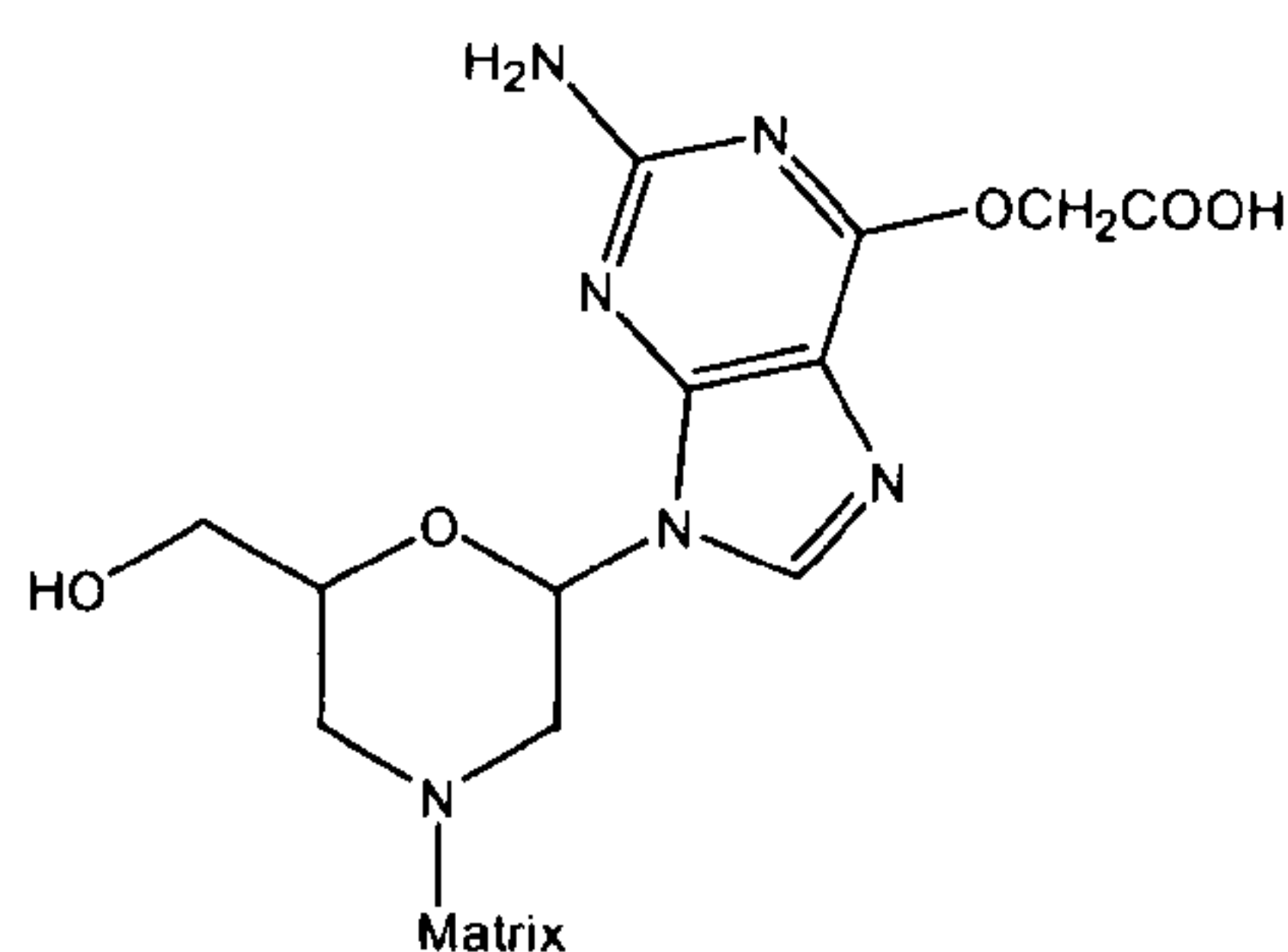
6.2.4 O^6 -(carboxymethyl)guanosine (4)



O^6 -(Carboxymethyl, methyl ester)-2',3',5'-tri-*O*-isobutyryl guanosine (1g, 1.78mmol) was dissolved in methanol and sodium hydroxide (0.02M) added until the reaction went to completion. HCl (0.02M) was added dropwise until the mixture was pH7 and the methanol removed by rotary evaporation. Preparative LC was used for separation the major peak of each injection was collected and pooled. Excess water was removed by cold finger and the product was freeze dried, to give a white solid.

Yield 243 mg, 33%; 236 °C (lit. >250 °C⁴³) λ_{\max} 248, 282 nm; ν_{\max} (KBr) 3300 (O-H, N-H str), 2880 (C-H str) 1650 (C=N str), 1620 (C=O str), 1580 (C=C str), 1250 (C-O str), 1065 cm^{-1} (C-O str); MS m/z (ES+) 342 (M+1) 100; $^1\text{H NMR}$ 3.7 (m, 2H, 5'- CH_2), 4.1 (m, 1H, 1'-H), 4.3 (m, 1H, 3'-H), 4.7 (m, 1H, 2'-H), 4.80 (s, 2H, OCH_2COOH), 6.85 (d, 1H, ^3J 6.3, 1'-H), 7.95 (s, 1H NCH).

6.2.5 O⁶-carboxymethylguanosine and Ovalbumin conjugate



O⁶-CMG (10.6 mg, 0.03 mmol) was dissolved in water (0.5 mL), to which NaIO₄ (21.6 mg, 1 mmol) was added, this was stirred at room temperature for 15 minutes. The reaction was stopped by the addition of 5 µl of ethylene glycol. Ovalbumin (OV) (10 mg) was dissolved in 1 mL of water and the pH adjusted to 9.5 (Na₂CO₃, 0.02M), the reaction mixture was left at room temperature for 4 minutes. A fresh solution of NaBH₃CN (30 mg in 1 mL of water) was prepared and 0.5 mL of this solution was added to the reaction mixture, along with 1 drop of octanol, to prevent foaming. The mixture was then left at 4 °C for 3 hours. The resulting mixture was dialyzed overnight in dialyzing tubing against phosphate buffer saline, also at 4 °C. The mixture was then eluted through a Sephadex G-50 column, using water, and the first UV-absorbing fractions were collected, pooled and concentrated by freeze drying.

6.2.6 Determination of the adduct to protein ratio

1 mg of ovalbumin was dissolved in 1 mL phosphate buffer and the UV at 280 nm was recorded, to this was added a molar equivalent of O⁶-CMG and the UV recorded. This was repeated with concentrations that gave the ratios of adduct to protein of 0:1 to 10:1. The resulting standard curve was used to determine the adduct to protein ratio of the product.

6.2.7 O⁶-carboxymethylguanosine and glass bead conjugate

O⁶-CMG (10 mg, 0.028mmol) was dissolved in water as above and the same protocol followed except that the OV was replaced with aminopropyl glass beads (22 mg). After dialyzing against phosphate buffer saline, the beads were washed with water and stored in storage solution at 4 °C. It was not possible to measure the level of modification directly: the UV profile was the same for O⁶CMG but this was not quantifiable due to refraction from the glass beads.

6.2.8 O⁶-carboxymethylguanosine and agarose conjugate

The supplied agarose suspension (6 mL) was transferred to a Falcon tube and allowed to settle, most of the solution was poured off, leaving a little on top. Coupling buffer (10 mL) was added and mixed gently to prevent the formation of fines. This was repeated twice more and as much buffer as possible was decanted without drying out the agarose, which was suspended in coupling buffer (5 mL). A O⁶-CMG solution was prepared by dissolving O⁶-CMG (11 mg, 0.031 mmol) in ice cold water (2 mL) and to this was added sodium periodate (10 mg, 0.46 mmol). The mixture was allowed to stir in an ice bath for 30 minutes. Following this 0.1 volume ethylene glycol (200 µL) was used to remove excess sodium periodate. The resulting mixture was used immediately to couple with the aminopropyl agarose.

The agarose gel was suspended in coupling buffer (10 mL) and added to the periodate-oxidised O⁶-CMGuo solution. This was then allowed to stir at room temperature for 6 hours. Sodium cyanoborohydride (50 mg, 0.80 mmol) was dissolved in 2 mL of the coupling buffer, added to the mixture and left to stir for 30 minutes. The O⁶-CMG-agarose was then extensively washed with coupling buffer and

water, and stored in 10mM Tris-Cl buffer, pH 7.2 containing 0.15M NaCl and 0.02% sodium azide, at 4 °C.

6.2.9 Guanosine and agarose conjugate

The above procedure was repeated with guanosine (10 mg, 0.18 mmol) to give an agarose gel to which guanosine has been linked.

6.3 Phage Display Experimental

6.3.1 Equipment and reagents

The incubation of bacteria was carried out in Jencons-PLS Orbital Incubator 5150, all UV data was collected on UV Spectronic Genesys 5, the PCR work was done on Techne Genofuge 16M and the DNA was sequenced in a DNA analyser (4200 LICOR gene reader).

The Ph.D.-12™ Phage Display Peptide Library Kit from New England Biolabs was used and the supplied *E. coli* host strain ER2738 was well suited for M13 propagation. It is a male-specific coli phage and infectivity occurs via the f-pilus that ER2738 produces to allow recombination of DNA, as such it is advised that the inoculation occurs to colonies that have been grown on a tetracycline containing medium. ER2738 that contain the F factor are tetracycline resistant. Therefore the supplied host was streaked out on Luria-Bertani (LB) -tetracycline (tet) plates and incubated, resulting in the growth of ER2738 that contain the F factor that are then susceptible to M13 infection. The procedure was carried out using standard LB media, tris buffered saline Tween -20 (TBST) and an eluting solution for non-specific disruption as described in 'media and solutions' (Section 6.3.9).

6.3.2 Phage titer of library

A phage titer was carried out to determine the number of plaque forming units that exist within the library. A series of dilution of the original library were prepared as shown in Table 2.1, 10 μL of each dilution of these were added to 200 μL of bacterial culture at the mid log phase (OD_{600} 0.5). Each was then added to 3 mL of agarose top at 47 $^{\circ}\text{C}$ and poured over prepared IPTG/Xgal plates (100 mL LB/agar containing 4 mg Xgal and 4.5 mg IPTG dissolved in 100 μL DMF is sufficient for 6 plates), warmed to 37 $^{\circ}\text{C}$. These plates were incubated overnight. By multiplying the number of plaques by the dilution factor the number of plaque forming units (pfu) can be determined. The pfu per 1 μL of the library is approximately 2.5×10^8 , this compares to the suggested pfu of 4×10^9 from the company literature.

6.3.3 Panning and titer

A solution of the target molecule linked to ovalbumin, 100 $\mu\text{L mL}^{-1}$ of the target in 0.1 M NaHCO_3 , pH 8.6, was prepared and 1.5 ml was repeatedly swirled on to a sterile polystyrene Petri dish. This was incubated overnight at 4 $^{\circ}\text{C}$ with gentle agitation in a humidified container (a sealable plastic box lined with damp paper towels), to allow maximum adsorption of the molecule on to the plate.

The coating solution was poured off the plate and it was firmly slapped down onto a clean paper towel to remove residual solution. The plate was filled with blocking buffer and left at 4 $^{\circ}\text{C}$ for 1 hour. After the hour the plate was rapidly washed six times with TBST. 10 μL of the original library, containing approximately 2.5×10^9 phage, was diluted in 1 ml of TBST, and this was pipetted onto the coated plate, and rocked gently at room temperature for one hour. The non-binding phage was

discarded by pouring off and the plate was washed ten times with TBST. The eluting solution for non-specific disruption of the binding interactions was used to release binding phage from the target molecule. These were collected and 1 μL was titered as shown in table 2.1. The titer showed that an average of 635 pfu μL^{-1} was selected, corresponding to a total of 730 250 phage, thus 0.029 % of the available phage were selected in the first panning.

6.3.4 Amplification

The remaining phage were amplified by the addition of the eluate to 20 ml of ER2738 culture at the early log phase and incubated at 37 °C for 7.5 hours. After this time the culture was spun for 15 mins at 10,000 rpm at 4°C, twice. The upper 80 % of the supernatant was transferred to a fresh tube and 1/6 volume of PEG/NaCl added and the phage was allowed to precipitate overnight. This was centrifuged at 10 000 rpm (4 °C) for 10 mins and the supernatant removed, the residual supernatant was removed after a second brief spin. The pellet was suspended in 1 mL of TBS and the suspension spun as before for 5 mins to pellets any residual cells. The supernatant was transferred to a fresh microcentrifuge tube and re-precipitated with 1/6 volume of PEG/NaCl, on ice for 1 hour. This was spun as before for 10 minutes and the supernatant discarded, the pellet was suspended in 200 μL TBS with 0.02% Na N₃, this is the amplified eluate which was titered. The resulting titer gave an average of 1.96×10^{10} pfu μL^{-1} which equates to a total phage population of 3.91×10^{13} .

6.3.5 Subsequent pannings

The second panning round was carried out in the same manner as the first, but the molecule attached to the Petri dish was ovalbumin on its own, the phage that did not bind were collected and titered as well as the binding phage. This would allow phage

that showed affinity to the binding protein to be removed from the population.

Taking 10 μL of the amplified phage from the first, representing 1.96×10^{11} phage were exposed to the plate. The titer showed that 1.81×10^{11} phage (92.6%) had bound to the protein and 1.46×10^{10} phage (7.4%) were unbound and showed little affinity for the protein. Amplification, as before, resulted in 2.1×10^{11} phage which were used in the final panning.

The final panning round was carried out by linking the ligand to glass beads and exposing the phage population selected so far to these in solution, all other criteria were the same. 9.8×10^{10} phage were exposed to the beads and 1.05×10^{10} were collected after being released.

6.3.6 DNA purification

After the final panning approximately 0.00004% of the original library survived, corresponding to about 1000 different phage. Ten different blue plaques were characterized; this was done by diluting an ER2738 overnight culture to 1:100 LB, 1 mL was dispensed into culture tubes and a blue plaque added to each. These were incubated with vigorous shaking for five hours. After centrifugation for 30 seconds at 4 °C the supernatant was removed and re-spun, the upper 80% was transferred to a fresh tube. 500 μL of this was taken and 200 μL PEG/NaCl was added. After being left to stand at room temperature for 10 minutes the solution was spun for 12 minutes at 4 °C and 12 000 rpm. The supernatant was discarded and the remaining pellet suspended in 100 μL iodine buffer and 250 μL ethanol. A short incubation of 10 minutes preferentially precipitates the single-stranded phage DNA leaving most of the phage protein in solution. Further centrifugation as before left a DNA pellet, this was

washed twice with 70% ethanol and dried. The remaining pellet was finally resuspended in 30 μL TE buffer and the DNA concentration measured via absorbance at OD_{260} .

6.3.7 PCR amplification

From these concentrations it was determined that the maximum amount of DNA template was required for the PCR reaction. A suitable primer was obtained from MWG Oligo Synthesis with a 50% GC content and a -96 primer end labelled with the fluorescent dye IRD 800. A SequiTherm EXCEL™ II DNA Sequencing Kit-LC was obtained from Epicentre® and a premix was made containing 1 μL IR-labelled primer, 7.5 μL DNA template, 7.2 μL sequencing buffer and 1 μL DNA polymerase. For each template four microfuge tubes were labelled A, C, G and T and to each the corresponding SequiTherm EXCEL™ II-LC termination mix (2 μL) was added along with 4 μL of the premix. These were then placed in the thermal cycler and allowed to undergo the following steps:-

Step 1 95 °C for 5 minutes

Step 2 95 °C for 10 seconds

Step 3 55 °C for 15 seconds

Step 4 70 °C for 30 seconds

Steps 2-4 were repeated for a total of 30 cycles.

6.3.8 DNA sequencing

The gel was prepared as follows, urea (28 g) was dissolved in water (28 mL), tris borate ethylenediaminetetraacetic acid (TBE) (5 mL), formamide (4 mL) and acrylamide (7.5 mL) were added and the mixture allowed to stir. When ready to pour

the gel, ammonium persulphate (APS) (350 μ L) and tetramethylethylenediamine (TEMED) (75 μ L) was swirled into the solution. The gel was then poured and allowed to set. The plates were then pre heated to 37 °C, the samples added and allowed to run overnight.

6.3.9 Media and solutions used with phage display

LB Medium /litre: 10 g Bacto-Tryptone, 5 g Yeast and 5 g NaCl, autoclave and store at room temperature.

LB plates: add agar to LB medium (15 mL per L) autoclave and pour. Store plates at 4 °C in the dark.

Isopropyl- β -D-thiogalatoside (IPTG) 5-bromo-4-chloro-3-indolyl- β D-galactoside

(Xgal) Mix: 1.25 g IPTG and 1 g Xgal dissolved in dimethyl formamide (25 mL).

Store at -20 °C in the dark.

LB/IPTG/Xgal plates: 10 g Bacto-Tryptone, 5 g yeast extract, 5 g NaCl, 1 g

MgCl₂.H₂O and 7 g agarose. Autoclave the solution and cool to less than 70 °C add 1 mL IPTG/Xgal and put into 50ml aliquots. Store solid at room temperature, melt in microwave when needed.

Tetracycline Stock: Tetracycline (20 mg/mL in EtOH). Store at -20 °C in the dark.

Vortex before use.

LB-Tet Plates: add agar to LB medium (15 mL per L) autoclave and cool to < 70 °C,

add 1 ml of tetracycline stock and pour. Store plates at 4 °C in the dark and do not use if plates are black or brown.

Blocking Buffer: 0.1 M NaHCO₃ (pH 8.6), 5 mg/mL BSA and 0.02% NaN₃. Filter

sterilize and store at 4 °C.

Tris-buffered saline (TBS): 50 mM Tris-HCl (pH 7.5) and 150 mM NaCl. Autoclave, store at room temperature

Tris-buffered saline Tween 20 (TBST) first round: TBS and 0.1% (v/v) Tween-20

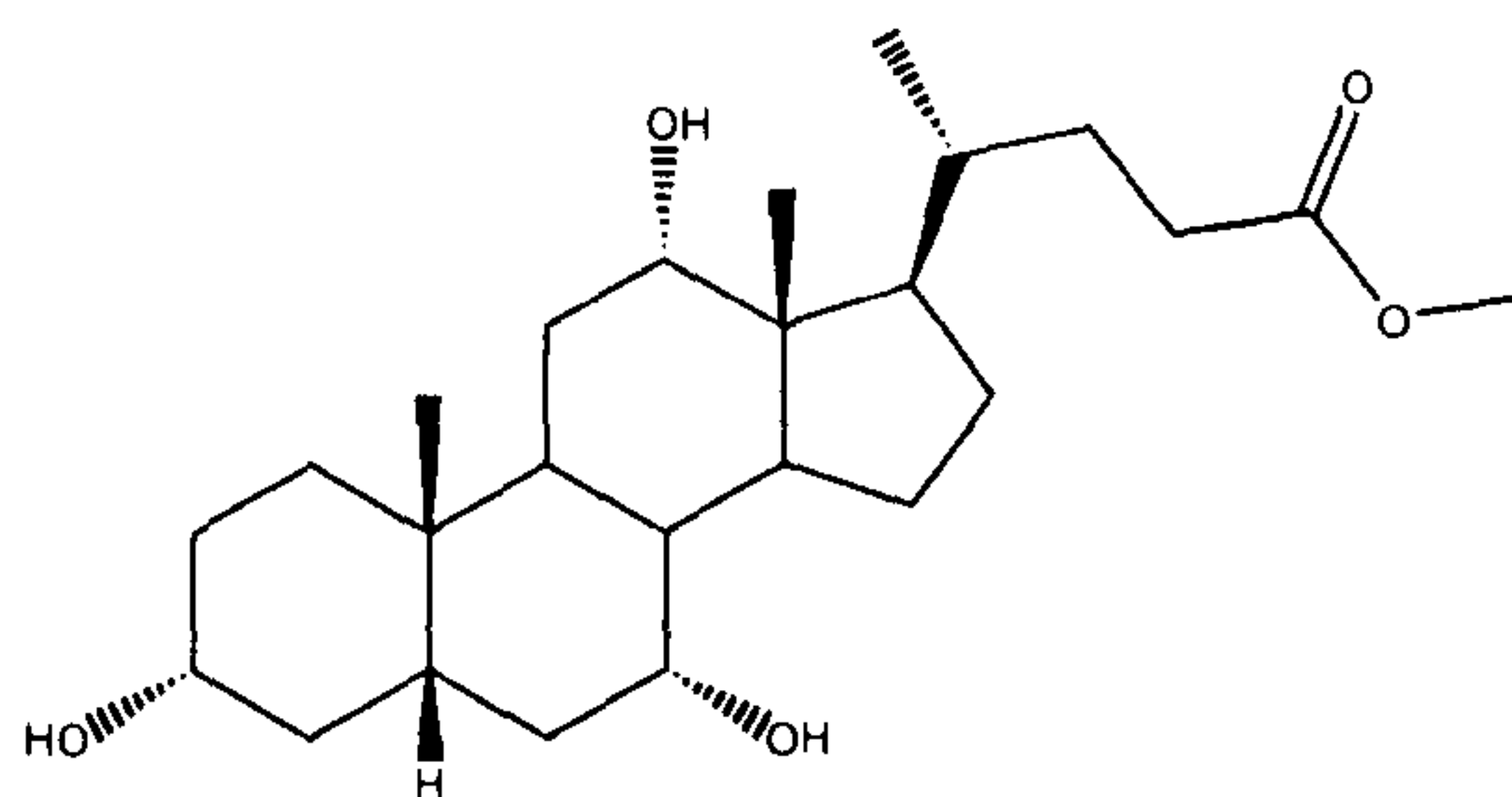
TBST subsequent rounds: TBS and 0.5% (v/v) Tween-20

Eluting Solution: 0.2 M Glycine-HCl (pH 2.2) and 1 mg/mL BSA

Polyethylene glycol (PEG)/NaCl: 20% (w/v) polyethylene glycol-8000 and 2.5M NaCl. Autoclave and store at room temperature.

6.4 Synthesis of cholic acid derivatives

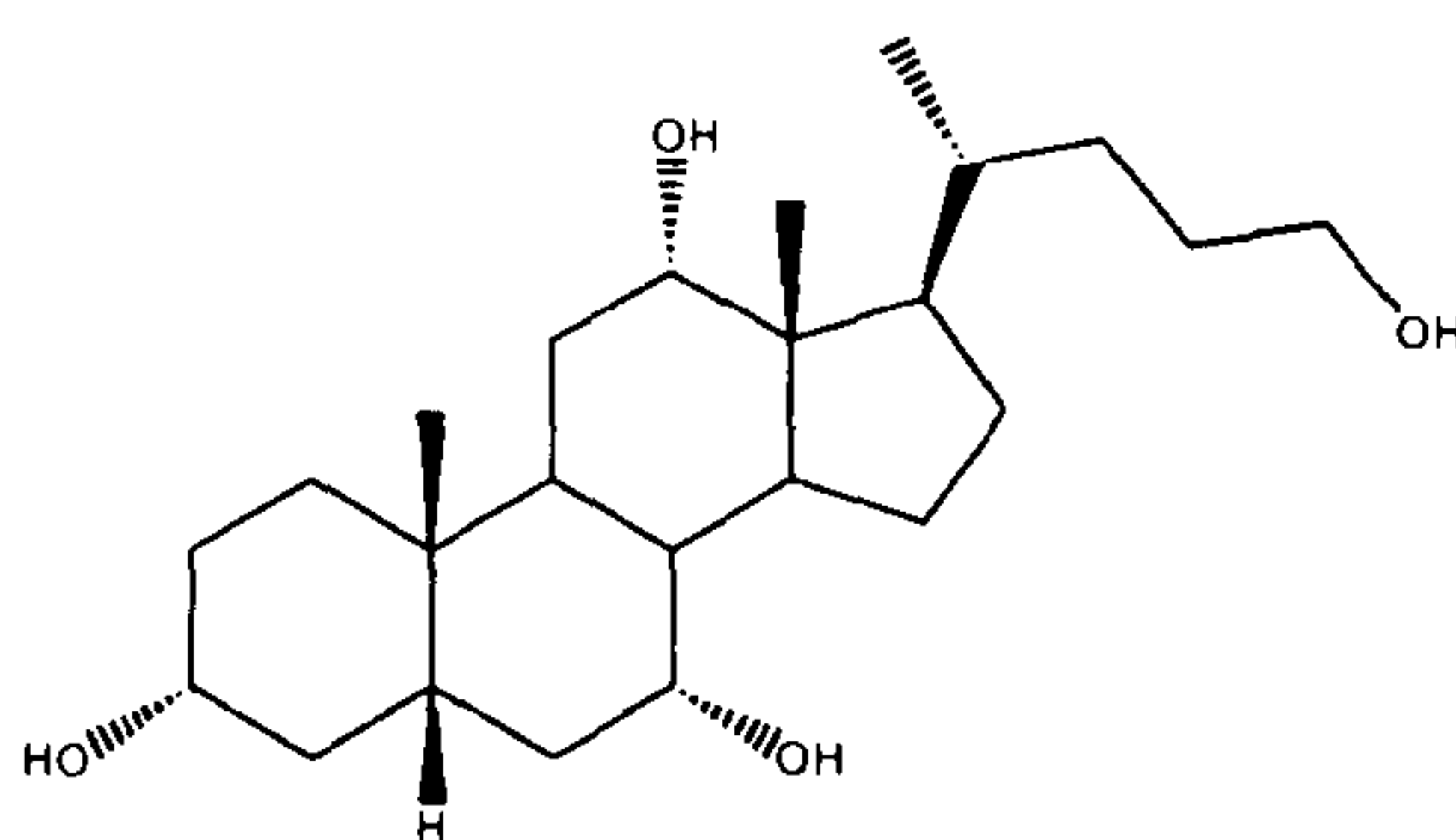
6.4.1 $3\alpha,7\alpha,12\alpha$ -trihydroxy-(5β)-cholan-24-oic acid methyl ester (6)



$3\alpha,7\alpha,12\alpha$ -trihydroxy-(5β)-cholan-24-oic acid (20.12 g, 49.24 mmol) was dissolved in methanol (200 mL) and a catalytic amount of concentrated sulphuric acid (0.5 mL) was added. The mixture was refluxed for 3.5 hours, after which the methanol was removed under vacuum. The solid was dissolved in chloroform and repeatedly washed with sodium bicarbonate and water. The solution was then dried with magnesium sulphate and the chloroform removed by rotary evaporation.

Yield 20.71g (99.7%); mp 154-7 °C (lit 155-156¹⁷⁸); ν_{max} (KBr) 1750 (ester C=O), 1350 cm^{-1} (secondary alcohol); MS m/z (ES+) 423 (M+1)₁₀, 440 (M+NH₄⁺)₁₀₀; ¹H NMR 0.6 (s, 3H, C²¹H₃), 0.8 (s, 3H, C¹⁸ or ¹⁹H₃), 0.9 (d, 3H, C¹⁸ or ¹⁹H₃), 1.4 (m, 27H, steroid), 3.4 (m, 3H, C³H, C⁷H, C¹²H), 3.7 (s, 3H, OCH₃); ¹³C NMR δ 13 (C¹⁸), 17 (C¹⁹), 23 (C²³), 24-51 (C, steroid), 70 (C³), 73 (C⁷), 74 (C¹²), 170 (C²⁴).

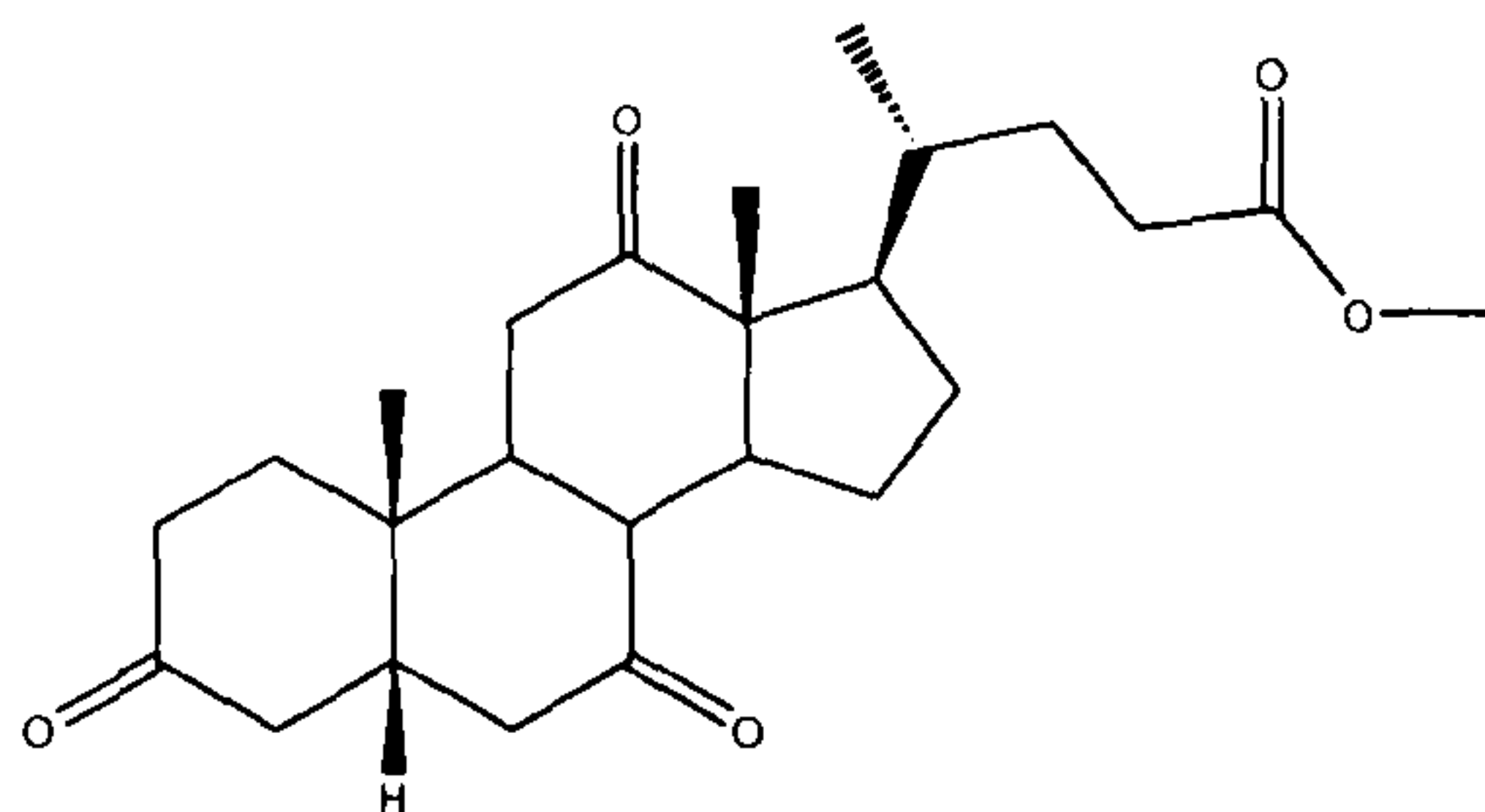
6.4.2 3 α ,7 α ,12 α -trihydroxy-(5 β)-cholan-24-ol (7)



Using dry THF, suspend **6** (3 α ,7 α ,12 α -trihydroxy-(5 β)-cholan-24-oic acid methyl ester) (30.67 g, 67.45 mmol) in a 1L round bottom flask, add LiAlH₄ (4.13 g, 109 mmol) and reflux for 48 hours. Saturated Na₂SO₄ (100 mL) was introduced slowly and the resulting mixture filtered. The insoluble precipitate was washed with a mixture of hot THF and MeOH. The solvents were evaporated of the resulting white solid was re-dissolved in hot methanol and allowed to crystallise in the freezer.

Yield 14.69 g (51%); mp 233-235 °C (lit. value 236.5-238 °C¹⁴¹); TLC dichloromethane and methanol (5:1 v/v) r_f = 0.65; ¹H NMR 0.2 (s, 3H, C²¹H₃), 0.4 (s, 3H, C¹⁸ or ¹⁹H₃), 0.5 (d, 3H, C¹⁸ or ¹⁹H₃), 1.1 (m, 25H, steroid), 3.0 (s, 3H, OCH₃), 3.3 (t, 2H, C²⁴H₂), 3.5 (m, 3H, C³H, C⁷H, C¹²H); ¹³C NMR δ 12 (C¹⁸), 17 (C¹⁹), 22 (C²³), 22-46 (C, steroid), 61 (C²⁴), 66 (C³), 70 (C⁷), 71 (C¹²), 170 (C²⁴).

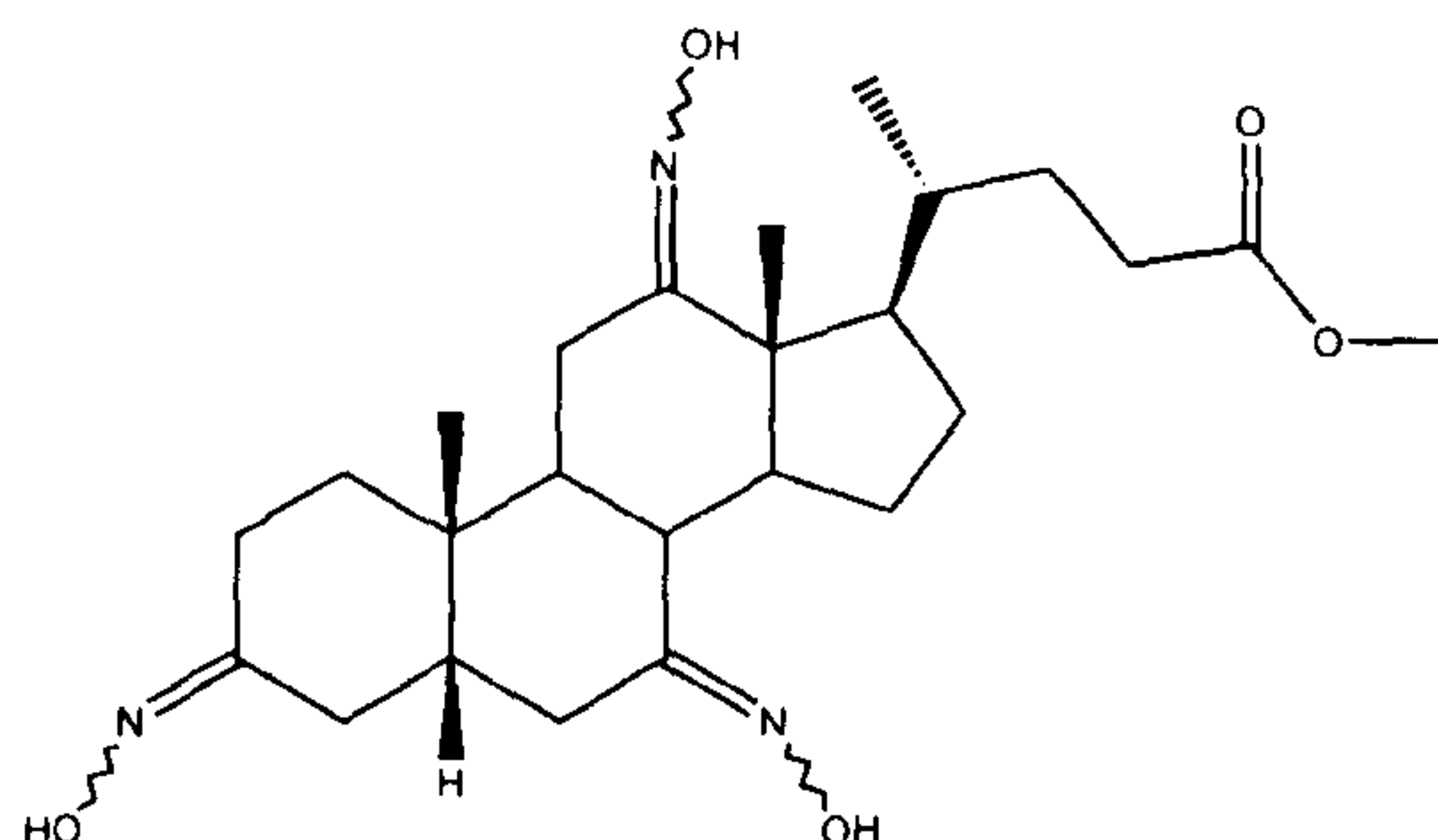
6.4.3 3,7,12 trioxo-(5 β)-cholan-24-oic acid methyl ester (10)



6 (3 α ,7 α ,12 α -trihydroxy-(5 β)-cholan-24-oic acid methyl ester) (500 mg, 1.2 mmol) was suspended in dry CH₂Cl₂ (18 mL) with 4 Å sieves, to which pyridinium dichromate (1.8 g, 677 mmol) was added and left stirring at room temperature. After 20 hours the solvent was removed under vacuum and redissolved in ethyl acetate and hexane (1:2) and passed through a silicon column using the same solvent mix as eluant. The product containing fractions were pooled and again the solvent was removed under vacuum to yield a white powder.

Yield 200 mg (40%); mp 238-240 °C; ν_{max} 1720 (cyclic ketone), 1750 (ester C=O), 1200 cm⁻¹ (ester C-O); MS m/z (ES+) 417 (M-methyl ester) 100; ¹H NMR 0.8 (d, 3H, C²¹H₃), 0.9 (s, 3H, C^{18 or 19}H₃), 1.0 (s, 3H, C^{18 or 19}H₃), 1.7 (m, 32H, steroid), 3.0 (s, 3H, OCH₃);

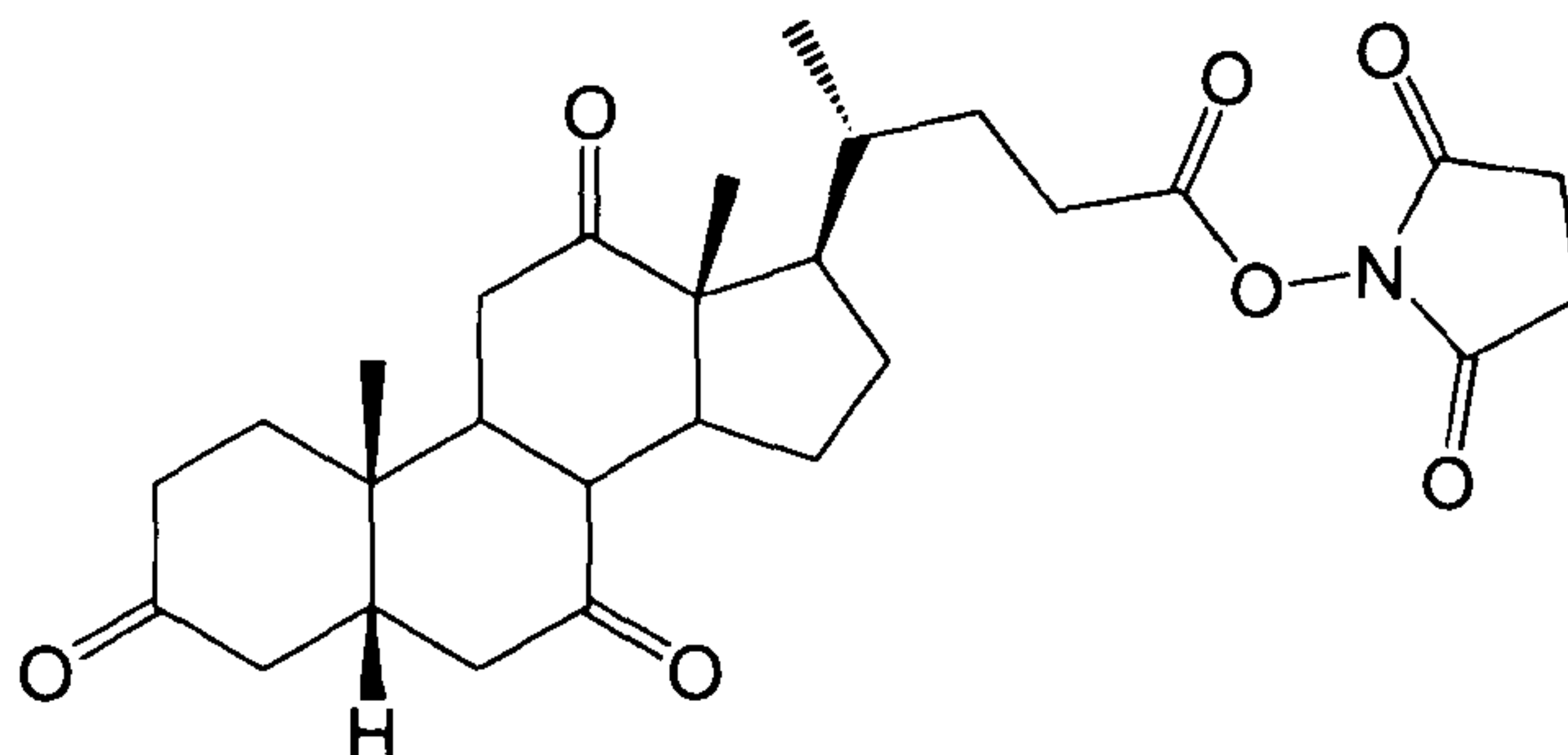
6.4.4 3,7,12 trioximo-(5 β)-cholan-24-oic acid methyl ester (12)



10 (3,7,12 trioxo-(5 β)-cholan-24-oic acid methyl ester) (110 mg, 24 mmol) was stirred into ethanol (10 mL) to which hydroxylamine hydrochloride (60 mg, 87 mmol) was added. After the addition of sodium acetate (120 mg, 146 mmol), the mixture stirred at room temperature for 1 hour followed by 12 hours of reflux. The solvent was removed and dissolved in HCl (0.1 M), NaCO₄ (0.1 M) was added and the solids filtered and washed with water. This was then extracted thrice with chloroform and the solvent removed under vacuum.

Yield 100mg (83%); mp 236 °C; ν_{max} 3500 (oxime), 1750 (ester C=O), 1660 (oxime), 1250 cm⁻¹ (ester C-O); MS m/z (ES+) 461 (M) 100; ¹H NMR 1.3 (s, 3H, C²¹H₃), 1.7 (m, 30H, steroid), 4.0 (s, 3H, OCH₃);

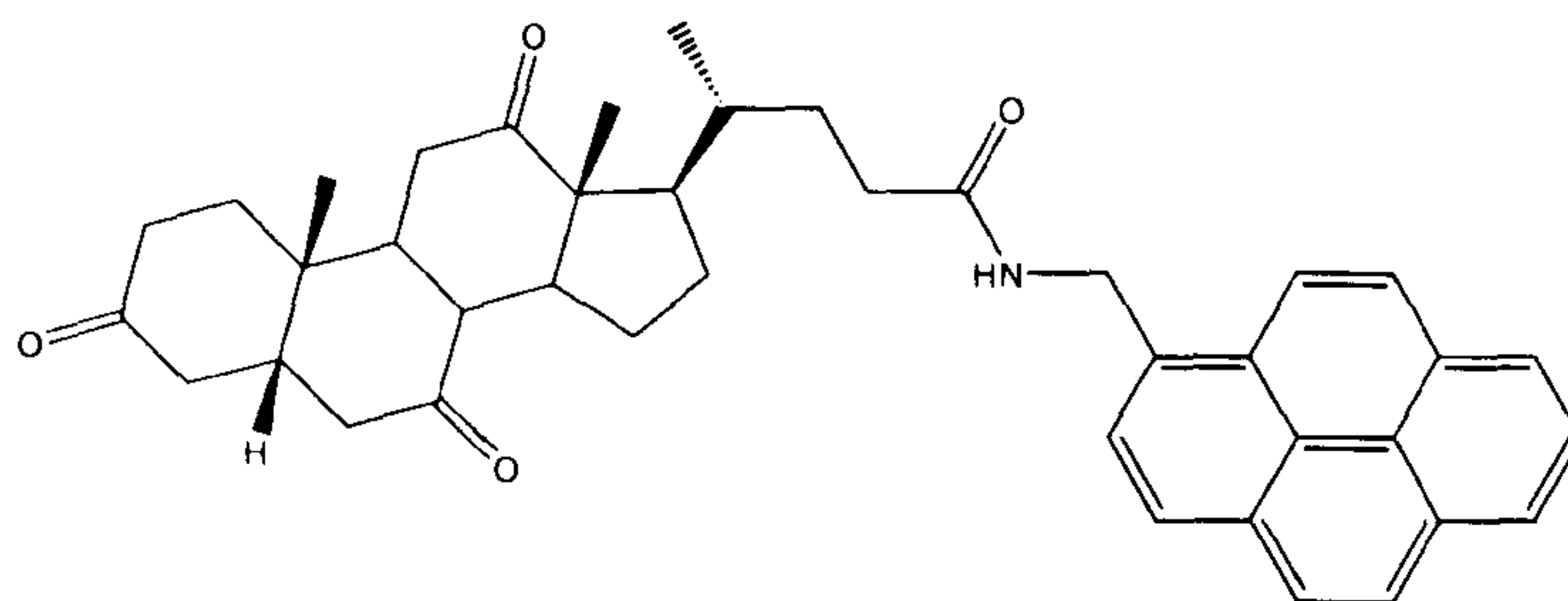
6.4.5 Succinimido-(*N*)-3,7,12 trioxo-(5 β)-cholan-24-oate



To a solution of **15** (3,7,12 trioxo-(5 β)-cholan-24-oic acid acid) (6.02 g, 15 mmol) in dry THF (100 mL), add *N*-hydroxysuccinimide (1.80 g, 15.45 mmol, 1.03 eq.). After 5 minutes, a solution of DCC (3.2g 15.45 mmol 1.03 eq.) in THF (20 mL) was added. The reaction mixture was stirred at room temperature for 28 hours. The dicyclohexyl urea formed was filtered and washed with ethyl acetate. The solvent was removed and the residue was dissolved in ethyl acetate, filtered and concentrated to dryness. The resulting oil was dissolved in a small amount of CHCl₃ and Et₂O was added to induce crystallisation. The solid was filtered off and dried to give the product.

Yield 5.4 g (72%); mp 236 °C; ν_{max} (nujol) 1784 (ester C=O), 1739 (ester C=O/cyclic ketone), 1708 (cyclic ketone), 1377 (cyclic ketone), 1206 (ester C-O), 1069 cm⁻¹ (C-N); MS m/z (ES+) 517 (M+NH₄⁺)₁₀₀; TLC dichloromethane and methanol (95:5 v/v) r_f = 0.36 +/- 0.05; elemental analysis, theoretical C 67.32% H 7.46% N 2.80%, found C 66.82% H 7.29% N 3.13%; ¹H NMR 0.9 (d, 3H, C²¹H₃), 1.9 (m, 30H, steroid), 2.8 (m, 4H, succinimide), 4.1 (q, 1H, succinimide); ¹³C NMR δ 11 (C¹⁸), 19 (C¹⁹), 22 (C²³), 24-56 (C, steroid), 168 (C²⁴), 169 (succinimide), 208 (C³), 209 (C⁷), 212 (C¹²).

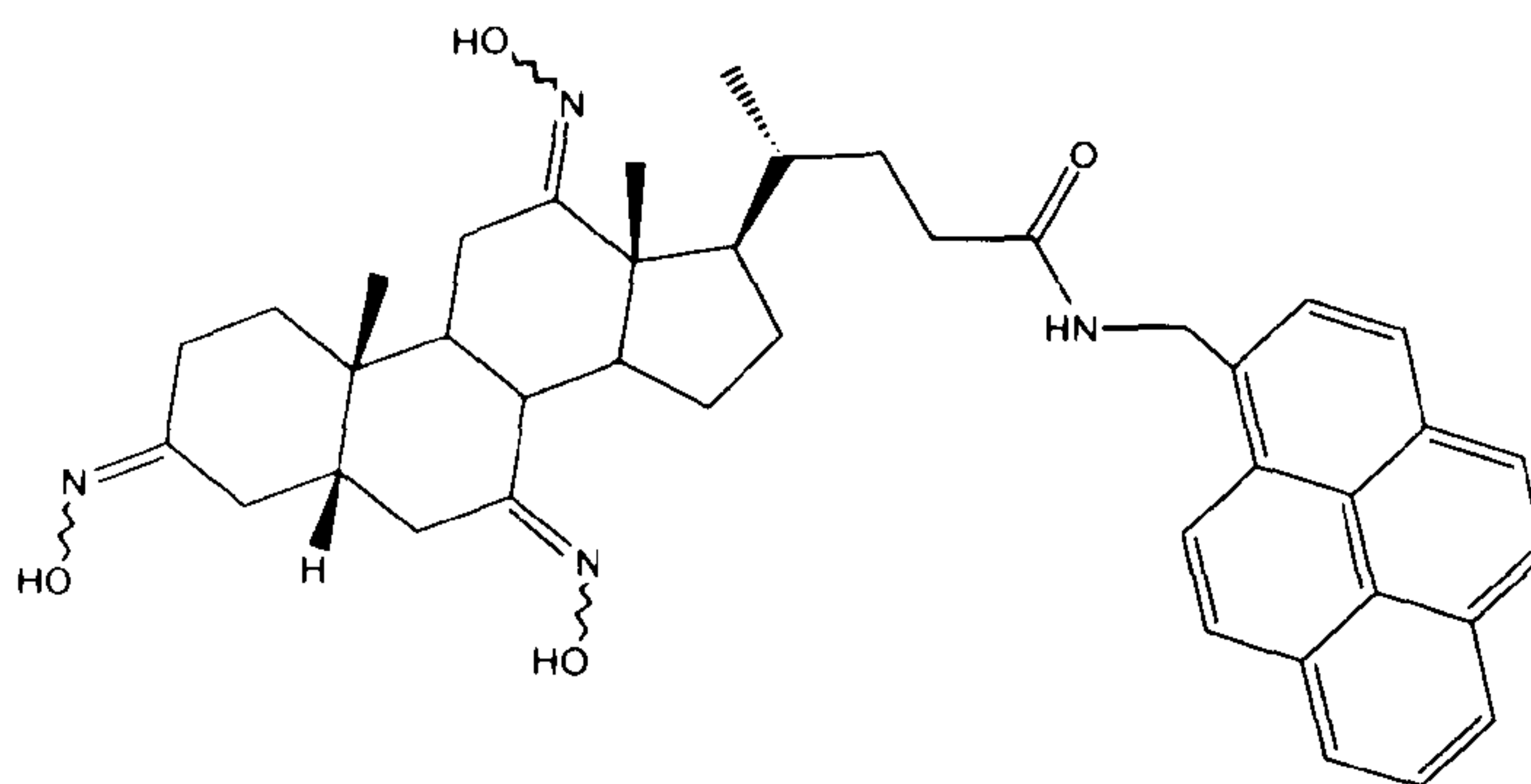
6.4.6 1'-pyrenyl methyl-3 α ,7 α ,12 α trioxo-(5 β)-cholan-24-amide (16)



Succinimido-(N)-3,7,12 trioxo-(5 β)-cholan-24-oate (1 g, 2 mmol) was dissolved in 25 mL DCM, to which triethylamine (0.2 g, 2 mmol) and pyrenylmethlyamine (0.5 g, 2 mmol) was added and allowed to stir at room temperature overnight. The reaction mixture was filtered and washed with HCl (0.1 M), then NaOH (0.1 M) and twice with water. Finally the solvent was dried with sodium sulphate and removed under vacuum. Recrystallization was from CHCl₃ with a small amount of Et₂O.

Yield 1.06 g (85%); mp 207 °C; λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES⁺) 616 (M+1) 55, 1232 (2M+2) 80; accurate mass, calculated 616.3421, measured 616.3427; TLC dichloromethane and methanol (95:5 v/v) r_f = 0.32; ¹H NMR 0.6 (d, 3H, C²¹H₃), 0.7 (s, 3H, C^{18 or 19}H₃), 1.1 (s, 3H, C^{18 or 19}H₃), 1.6 (m, 21H, steroid), 2.3 (m, 3H, steroid) 5.0 (m, 2H, NHCH₂), 8.0 (m, 9H, pyrene); ¹³C NMR δ 11 (C¹⁸), 18 (C¹⁹), 21 (C²³), 24-56 (C, steroid), 123-131 (C, aromatic), 172 (C²⁴), 208 (C³), 209 (C⁷), 212 (C¹²).

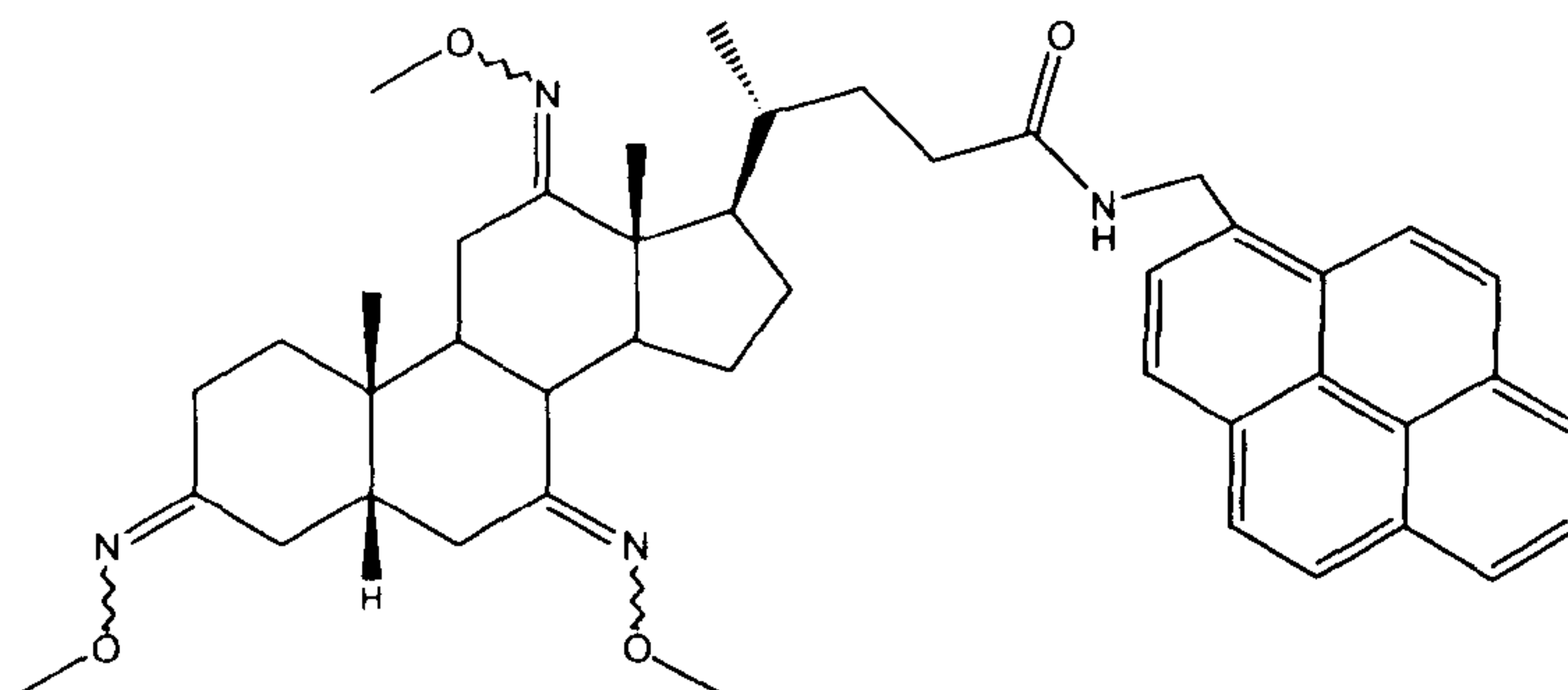
6.4.7 1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide (17)



16 (1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide) (200 mg 0.32 mmol) was dissolved in ethanol (20 mL) to which NaOAc (160 mg, 1.94 mmol) and hydroxylamine hydrochloride (81 mg, 1.17 mmol) was added. The mixture was stirred at room temperature for 1 hour and then refluxed overnight. The solvent was removed under vacuum, HCl (0.1 M) was added to the flask and this was washed three times with chloroform (20 mL each), the extract was dried with sodium sulphate, and the solvent removed under vacuum to give a pale yellow glass.

Yield 200 mg, (93%); λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES+) 661 (M+1) 100, ^1H NMR 0.7 (d, 3H, C²¹H₃), 0.9 (s, 3H, C^{18 or 19}H₃), 1.4 (m, 30H, steroid), 5.1 (d, 2H, ^3J 5.1, NHCH₂), 8.1 (m, 9H, pyrene);

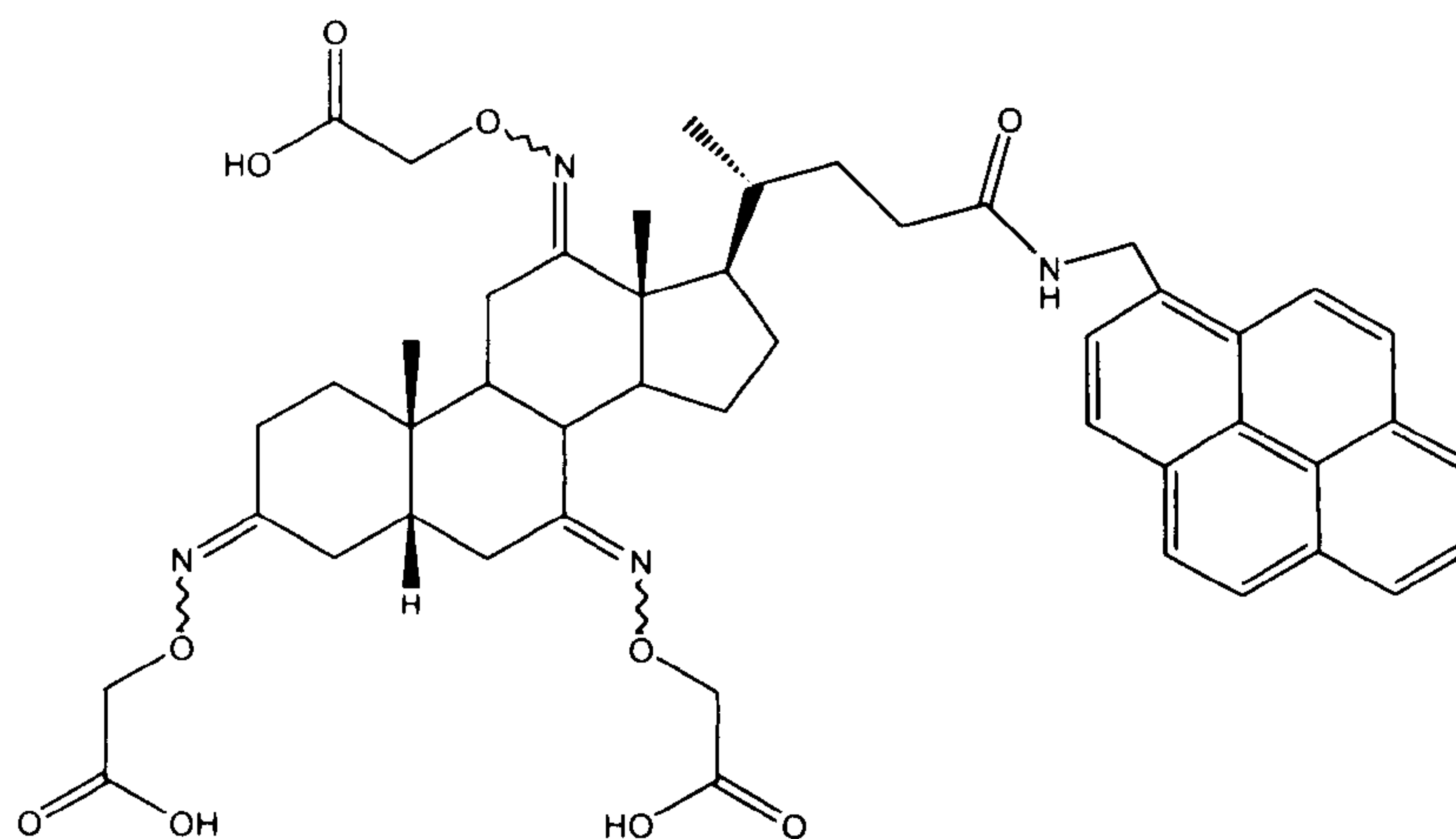
6.4.8 1'-pyrenyl methyl-3,7,12 trioximomethyl-(5 β)-cholan-24-amide



1'-pyrenyl methyl-3,7,12 trioximomethyl-(5 β)-cholan-24-amide was prepared as in 1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide, using 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (200 mg 0.32 mmol), methylhydroxylamine hydrochloride (73 mg, mmol).

Yield 160 mg (71%); λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES+) 720 (M + NH₃)³⁰, 689 (M – O methyl + NH₃)²⁰, ¹H NMR 1.7 (m, 33H, steroid), 4.2 (m, 9H, (CH₃)₃), 5.1 (m, 2H, NHCH₂), 8.2 (m, 9H, pyrene);

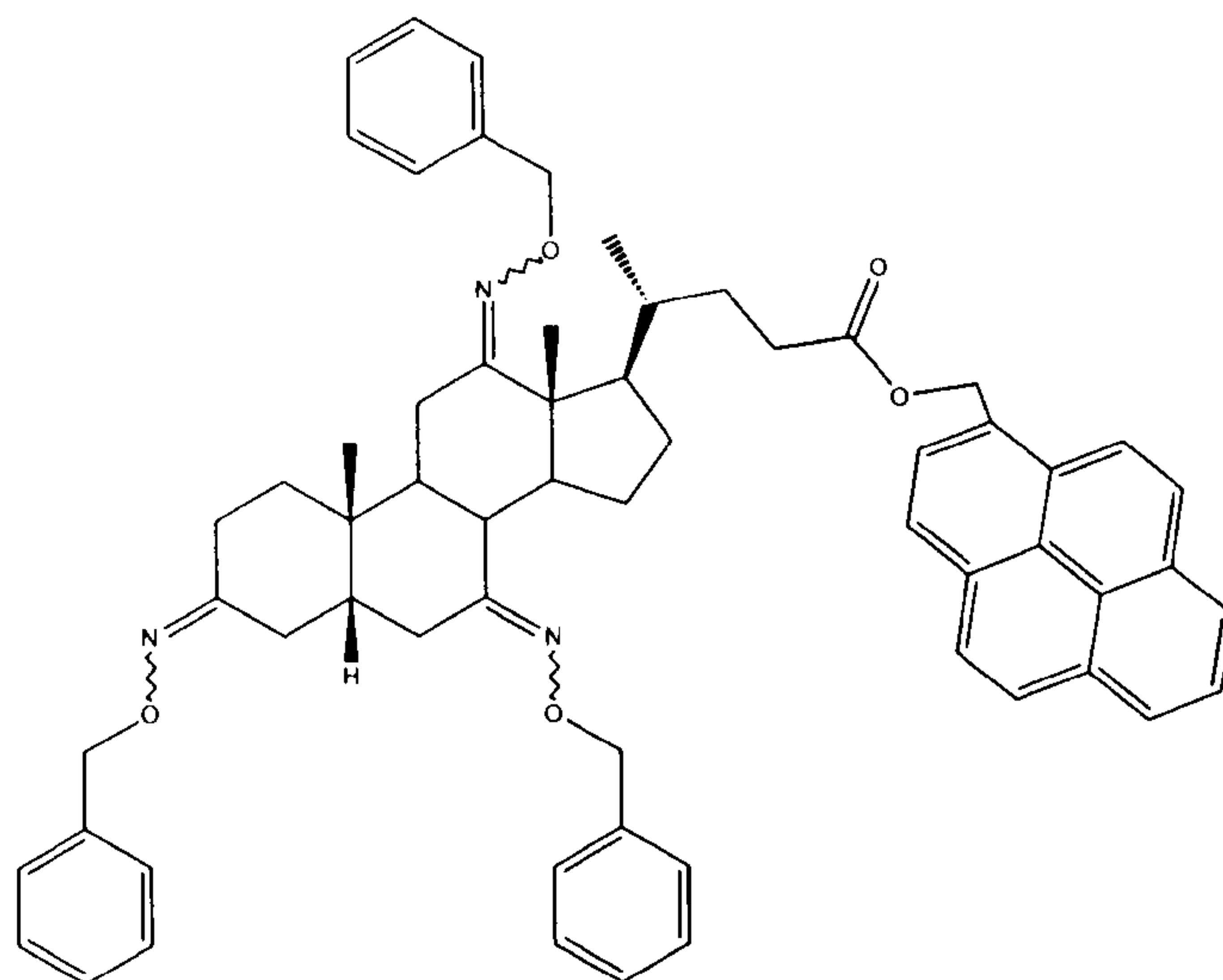
6.4.9 1'-pyrenyl methyl-3,7,12 trioximocarbonylmethyl-(5 β)-cholan-24-amide



1'-pyrenyl methyl-3,7,12 trioximocarbonylmethyl-(5 β)-cholan-24-amide was prepared as in 1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide, using 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (200 mg, 0.32 mmol), with carbonylmethylhydroxylamine hydrochloride (150 mg, mmol) and sodium acetate (160 mg, 1.94 mmol).

Yield 240 mg (89%); λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES+) 834 (M) 100, 760 (M-O-carboxymethyl) 55, ^1H NMR 1.7 (m, 31H, steroid), 4.5 (m, 9H, (CH₃)₃), 4.9 (m, 2H, NHCH₂), 7.9 (m, 9H, pyrene); ^{13}C NMR δ 11 (C¹⁸), 14 (C¹⁹), 15 (C²³), 21-56 (C, steroid), 124-131 (C, aromatic), 70 ((NOCH₂)₃), 157 (C³, C⁷ and C¹²).

6.4.10 1'-pyrenyl methyl-3,7,12 trioximobenzyl-(5 β)-cholan-24-amide

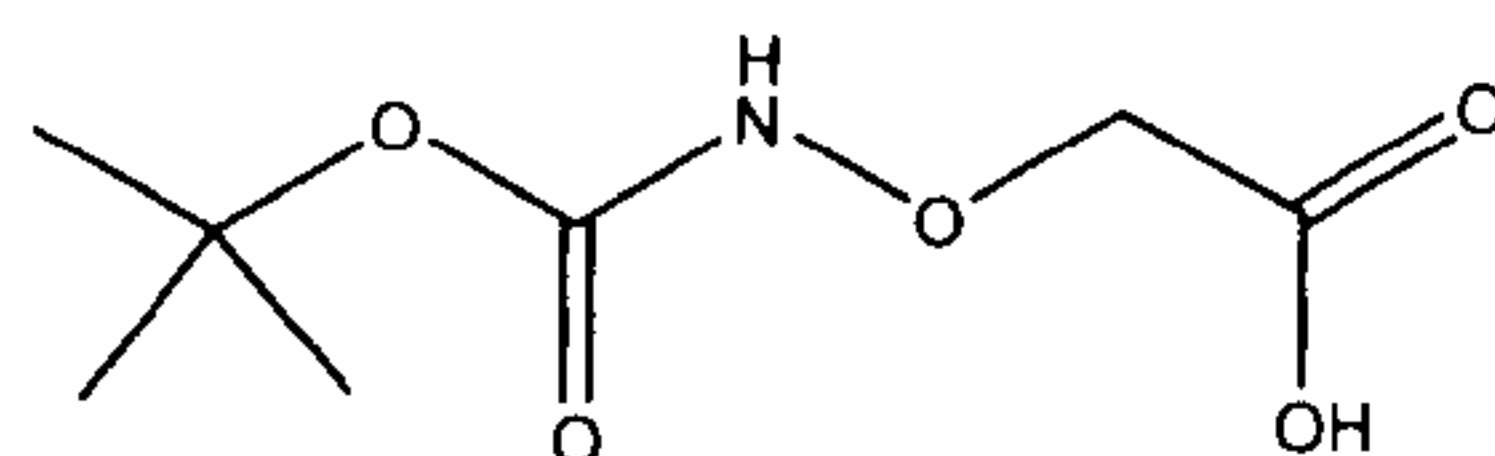


1'-pyrenyl methyl-3,7,12 trioximobenzyl-(5 β)-cholan-24-amide was prepared as in 1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide, using 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (200 mg, 0.32 mmol), benzylhydroxylamine hydrochloride (195 mg, 1.17 mmol) and sodium acetate (160 mg, 1.94 mmol)

Yield 241 mg, 80%; TLC dichloromethane and methanol (95:5 v/v) r_f = 0.7; λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES+) 932 ($M + 1$)₅₀, 823 ($M - O$ -benzylmethyl)₁₀₀; ^1H NMR 1.5 (m, 44H, steroid), 4.9 (m, 6H, (CH₂benzyl)₃), 7.2 (m, 15H, (benzyl)₃), 8.0 (m, 9H, pyrene); ^{13}C NMR δ 12 (C¹⁸), 14 (C¹⁹), 19 (C²³), 21-53 (C, steroid), 75-77 (C, (benzyl)₃), 122-128 (C, aromatic), 156 (C³, C⁷ and C¹²).

6.5 Synthesis of *N*-(aminooxy acetyl) amino acids

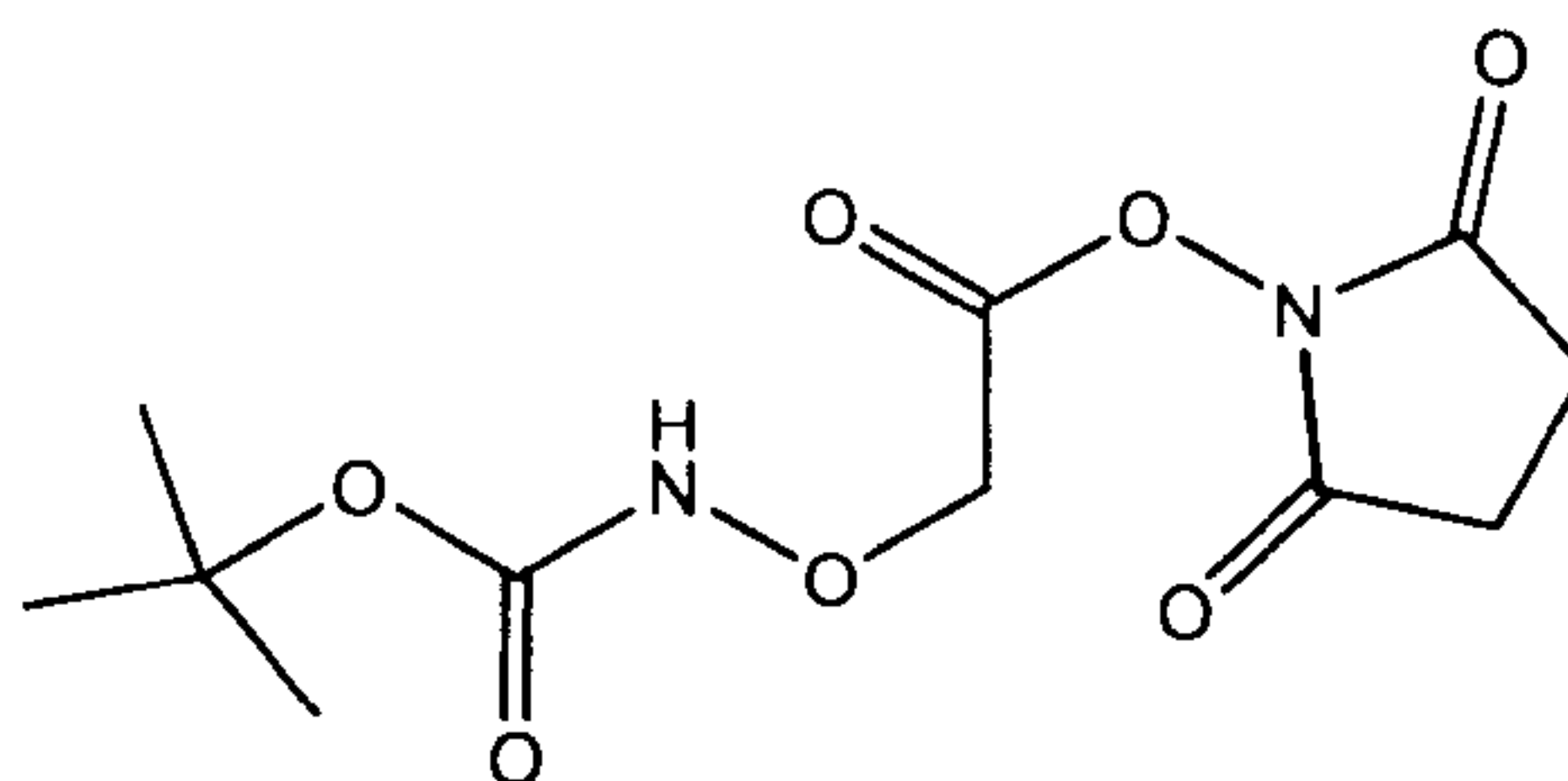
6.5.1 ((1',1'-dimethylethoxycarbonyl)aminooxy)ethanoic acid (18)



Solution A was prepared as follows, KOH (2.24 g) was dissolved in MeOH (40 mL) and H₂O (60 mL). O-Carboxymethyl hydroxylamine hemihydrochloride (1.00 g, 7.87 mmol) was dissolved in 15 mL of solution A and adjusted to pH 9. Di-*tert*-butyl dicarbonate (3.43 g, 15.74 mmol) was added in one portion, the mixture was then stirred at room temperature for 16 hrs, maintaining pH 9 by the dropwise addition of solution A. The mixture was then concentrated to dryness by rotary evaporation to give a white solid which was dissolved in H₂O (10 mL). This was cooled to 0°C and the dropwise addition of 6 N HCl lowered the pH to 3, whilst being stirred vigorously, resulted in the formation of a white precipitate. This was filtered and freeze dried to give the desired product as a white solid.

Yield 1.1g (73%); mp 99-102 °C, mp (lit. 111-112 °C¹⁷⁹); MS *m/z* (ES⁺) 192 (M + H)³⁰; elemental analysis, theoretical C 43.98% H 6.85% N 7.33%, found C 43.75% H 6.96% N 7.26%; ¹H NMR δ 1.5 (s, 9H, (CH₃)₃), 4.0 (s, 2H, NHCH₂), 4.3 (s, 2H, OCH₂); ¹³C NMR δ 23 (C(CH₃)₃), 74 (C(CH₃)₃), 83 (NOC), 159 (OC=O), 172 (COOH)

6.5.2 Succinimido-(N)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanoate (19)



A solution of *N*-hydroxysuccinimide (0.6 g, 5 mmol) in ethyl acetate (15 mL) was prepared and to it was added a solution of **18** ((1,1-dimethylethoxycarbonyl)-aminoxy) ethanoic acid (0.95 g, 5 mmol) and ethyl acetate (15 mL). This was allowed to stir at room temperature for a few minutes and a solution of *N*-*N'*-dicyclohexylcarbodiimide (1.03 g, 5 mmol) in ethyl acetate (1.5 mL) was added. The mixture was stirred overnight. Dicyclohexylurea forms as a white precipitate which is removed by filtration. The solvent is removed by rotary evaporation; the product is a white solid.

Yield 1.2 g (85%); MS *m/z* (ES+) 274 (M+1) ₁₀₀; IR (KBr) 3343 (N-H), 1786 (ester C=O), 1737 (ester C=O/cyclic ketone), 1377 (cyclic ketone), 1208 cm⁻¹ (ester C-O); MS *m/z* (ES+) 576 (Mx2) ₃₀, 215 (M – O tert butyl) ₁₀₀; ¹H NMR δ 1.4 (s, 9H, (CH₃)₃), 2.8 (s, 4H, succinimide), 4.7 (s, 2H, OCH₂); ¹³C NMR δ 24 (succinimide), 28 ((CH₃)₃), 70 (C(CH₃)₃), 83 (NOC), 156 (OC=O), 165 (COOH), 169 (succinimide)

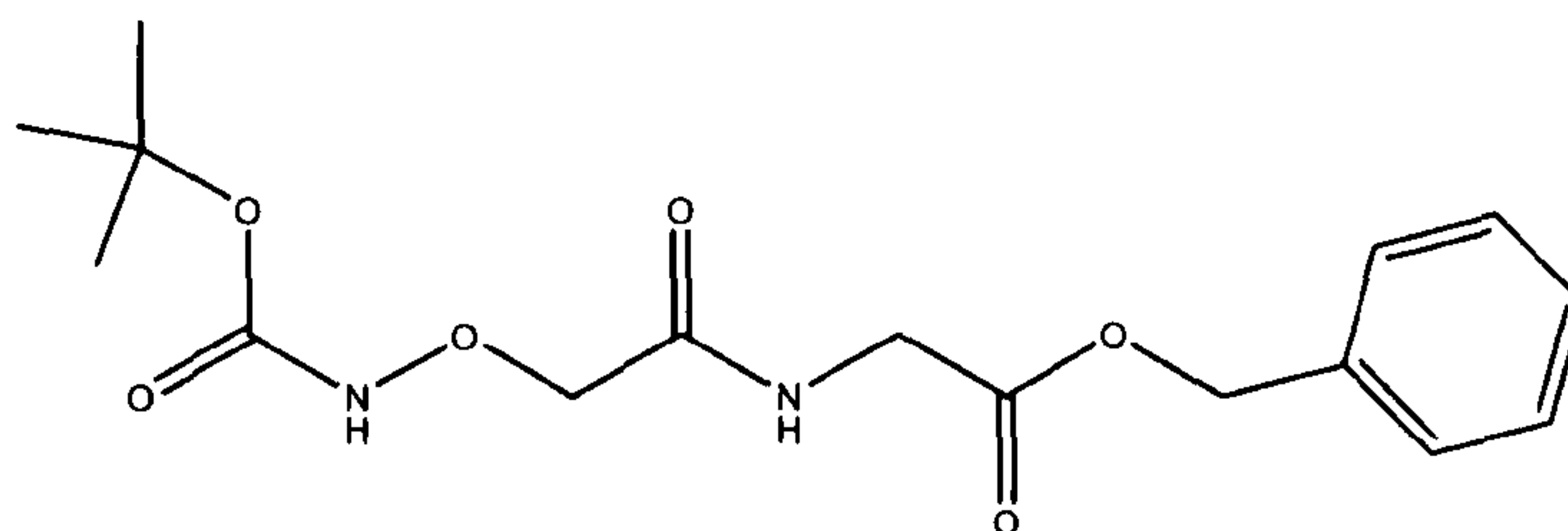
6.5.3 Glycine benzyl ester toluene-4-sulphonate

Glycine (1 g, 13.3 mmol) was added to toluene-4-sulphonic acid (3.2 g, 16.7 mmol) in toluene (20 mL) and benzyl alcohol (20 mL). This was heated and kept at reflux for 3 hours using a Dean-Stark head until it became apparent that no more water was being produced. The mixture was placed in an ice bath and when cool diethyl ether (100 mL) was added to the flask which was then kept on ice for a further two hours. After filtration, the white solid was washed well with ether and allowed to dry.

6.5.4 Glycine benzyl ester,

Glycine benzyl ester toluene-4-sulphonate (1.03 g, 3.0 mmol) was dissolved in dichloromethane (50 mL) with triethylamine (0.6 g, 6.0 mmol), this was stirred and left to stand for 10 minutes. The mixture was washed with water (20 mL) and dried with sodium sulphate and the solvent removed to give an oil that was used immediately in the coupling reaction.

6.5.5 N-(1''-glycine benzyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide

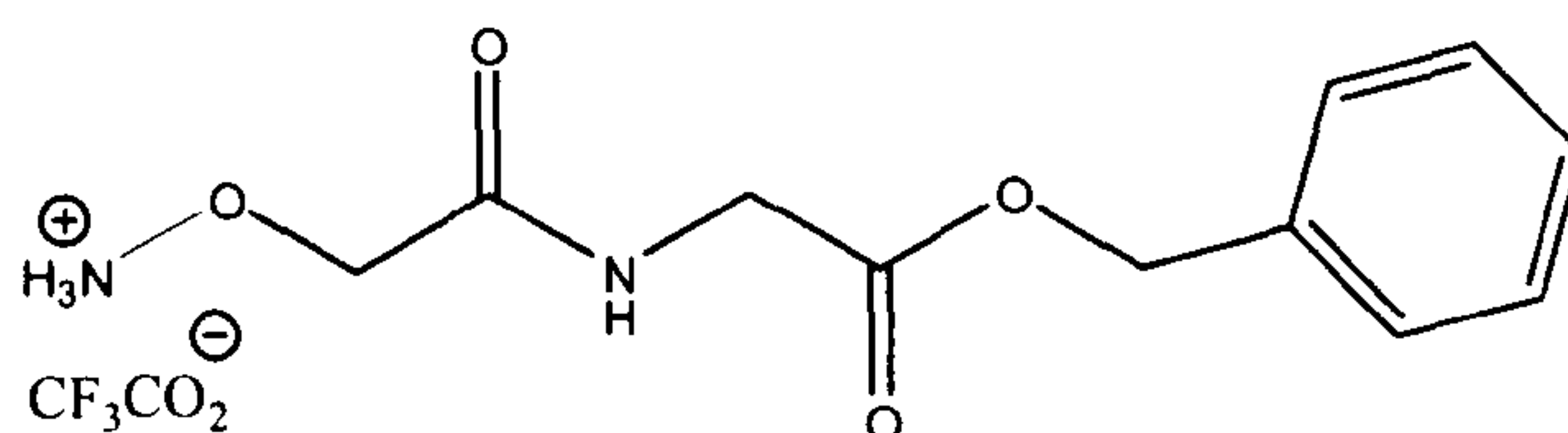


Glycine benzyl ester (128 mg, 0.78 mmol), prepared as above, was dissolved in ethyl acetate (1 mL), this was then added to a solution of ethyl acetate (20 mL) and **19** (succinimido-(N)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanoate) (225 mg,

0.78 mmol). This solution was left to stir at room temperature and the reaction followed by TLC (chloroform and methanol (95:5 v/v)). The reaction was complete after three hours and the solvent removed under vacuum. The semi-solid was dissolved in a little ethyl acetate (2-3 mL), washed twice with water and dried with sodium sulphate. The solvent evaporated and recrystallised from warm ethyl acetate with the addition of diethyl ether.

Yield 153 mg, 45%; $^1\text{H NMR}$ δ 1.5 (s, 9H, $(\text{CH}_3)_3$), 4.2 (d, 2H, $^3J = 4.74$ Hz, NHCH_2), 4.4 (s, 2H, OCH_2), 5.5 (s, 2H, OCH_2), 7.3 (s, 5H, C_6H_5); $^{13}\text{C NMR}$ δ 28 ($(\text{CH}_3)_3$), 52 (OCH_3), 53 (NHCH), 75 ($\text{C}(\text{CH}_3)_3$), 83 (NOC), 158 ($\text{OC}=\text{O}$), 172 (CH_2CONH), 179 ($\text{CHC}=\text{O}$), 126 ($\text{CH}_2\text{C}_6\text{H}_5$), 128 ($\text{CH}_2\text{C}_6\text{H}_5$), 129 ($\text{CH}_2\text{C}_6\text{H}_5$);

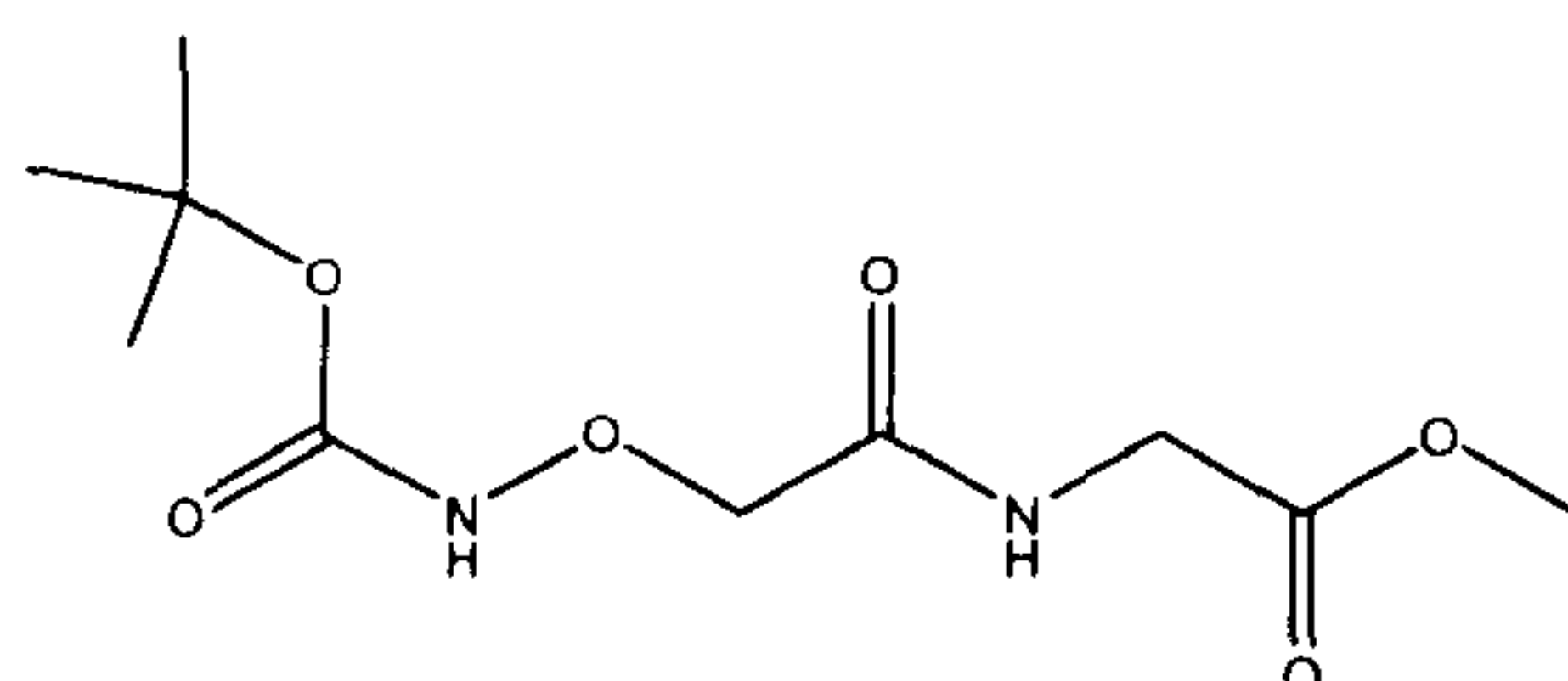
6.5.6 N-(1''-glycine benzyl ester)aminoxy ethanamide (21)



Taking **21** (N-(1''-glycine benzyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide) (51 mg, 0.15 mmol), TFA was added dropwise until the solid had dissolved, this was then left for 1 hour. The acid was removed under vacuum overnight which found an oily residue, to which 5mL of diethyl ether was added and fine powder formed overnight in the freezer, this was filtered and dried

Yield 40 mg, 75%; MS m/z (ES+) 239 (M+1) 45, 477 (Mx2+1) 100, ^1NMR δ 4.0 (s, 2H, OCH_2), 4.2 (s, 2H, NHCH_2O) 5.2 (m, 2H, $\text{OCH}_2\text{C}_6\text{H}_5$), 7.4 (s, 5H, C_6H_5),

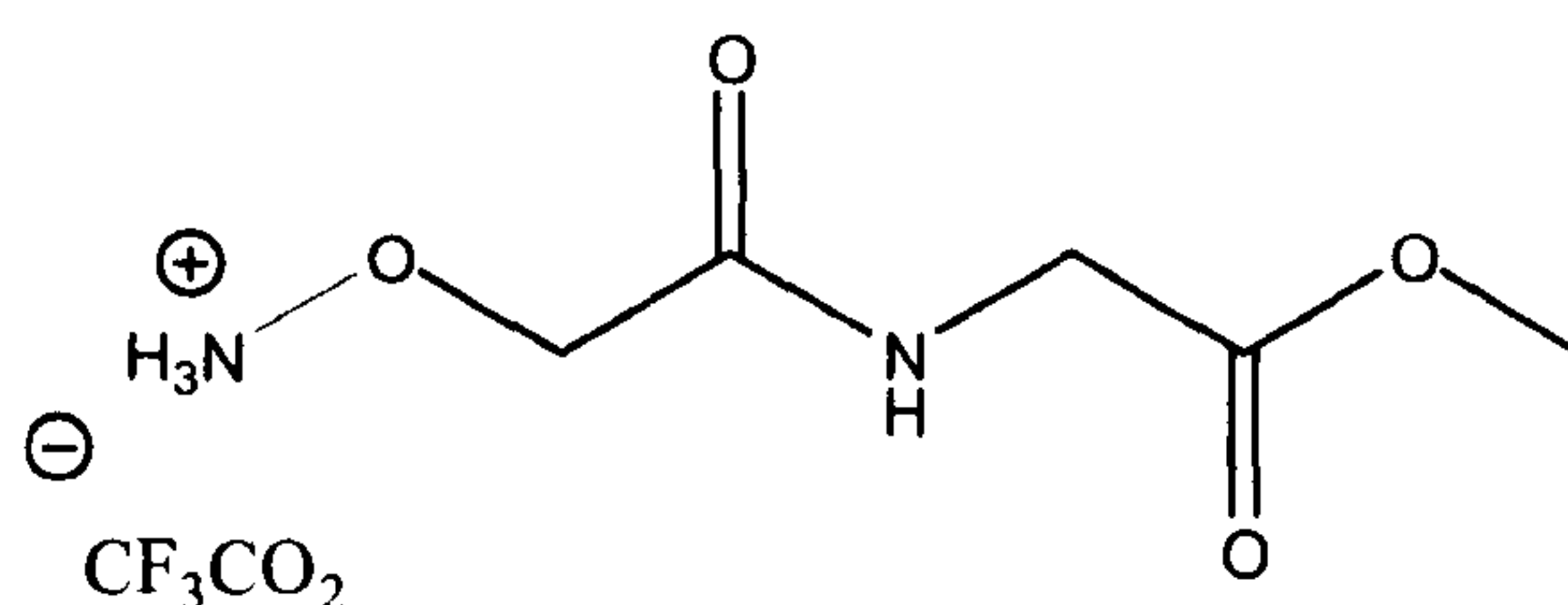
6.5.7 N-(1''- glycine methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide



Glycine methyl ester (325 mg, 2.6 mmol) was suspended in DCM (25 mL) and triethylamine (600 mg, 6 mmol) was added. To this was added succinimido-(N)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanoate (750 mg, 2.6 mmol) dissolved in DCM (10 mL) and the mixture was stirred at room temperature for 48 hours. This was then washed trice with water and dried with sodium sulphate, solvent removed under vacuum to give a clear glass.

Yield 270 mg, 37%; MS m/z (ES+) 239 (M+1) 45, 476 (Mx2) 100, ^1NMR δ 1.4 (s, 9H, (CH₃)₃), 3.7 (s, 3H, methyl), 4.0 (s, 2H, NHCH₂), 4.4 (s, 2H, OCH₂);

6.5.8 N-(1''-glycine methyl ester)aminoxy ethanamide (20)



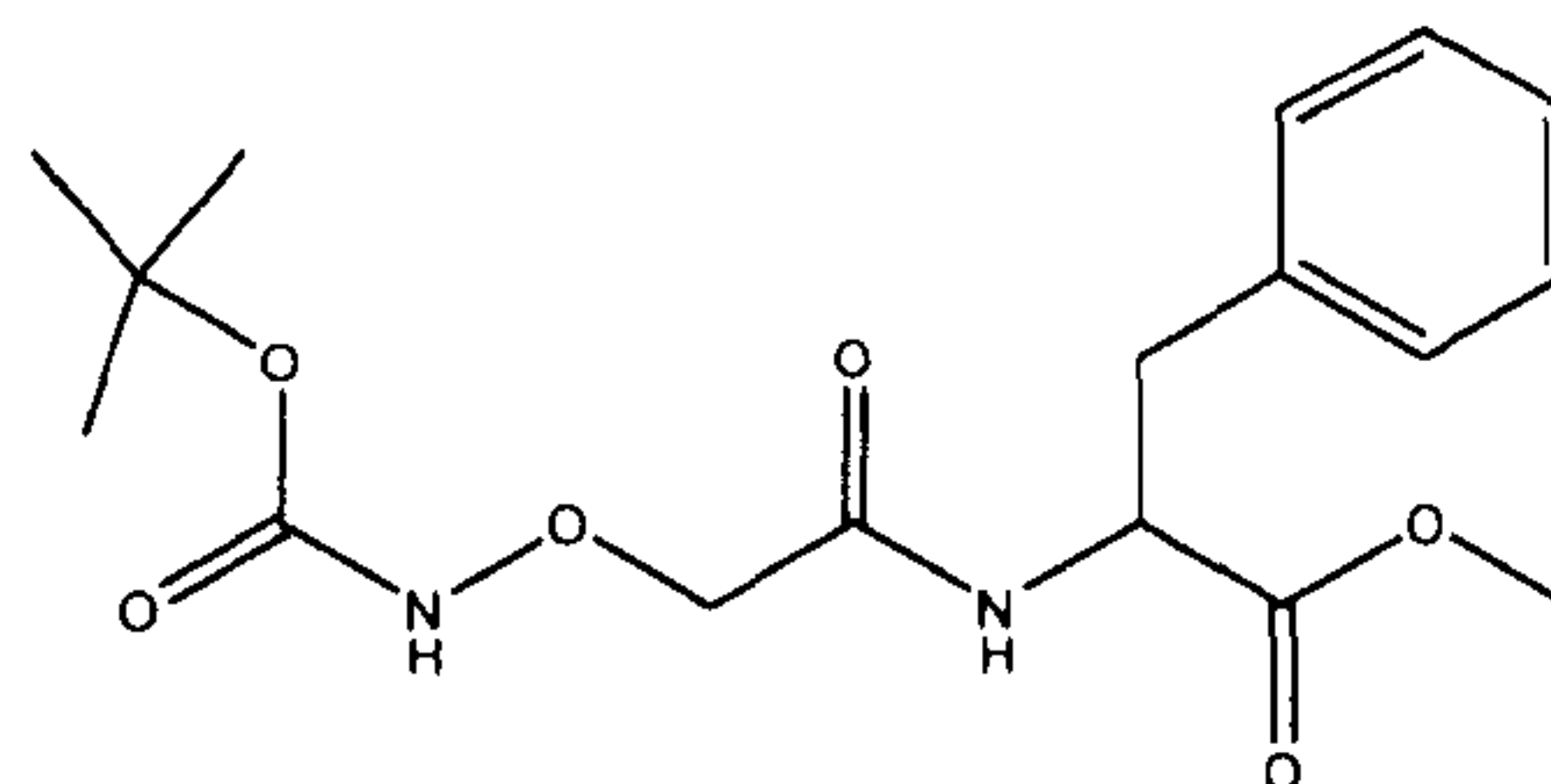
N-(1''-glycine methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide (200 mg, 0.76 mmol) was treated the same as for N-(1''-glycine benzyl ester) aminoxy ethanamide.

Yield 178 mg, 89%; δ 3.6 (s, 3H, methyl), 3.8 (s, 2H, NHCH₂), 4.2 (s, 2H, OCH₂);

6.5.9 Phenylalanine methyl ester

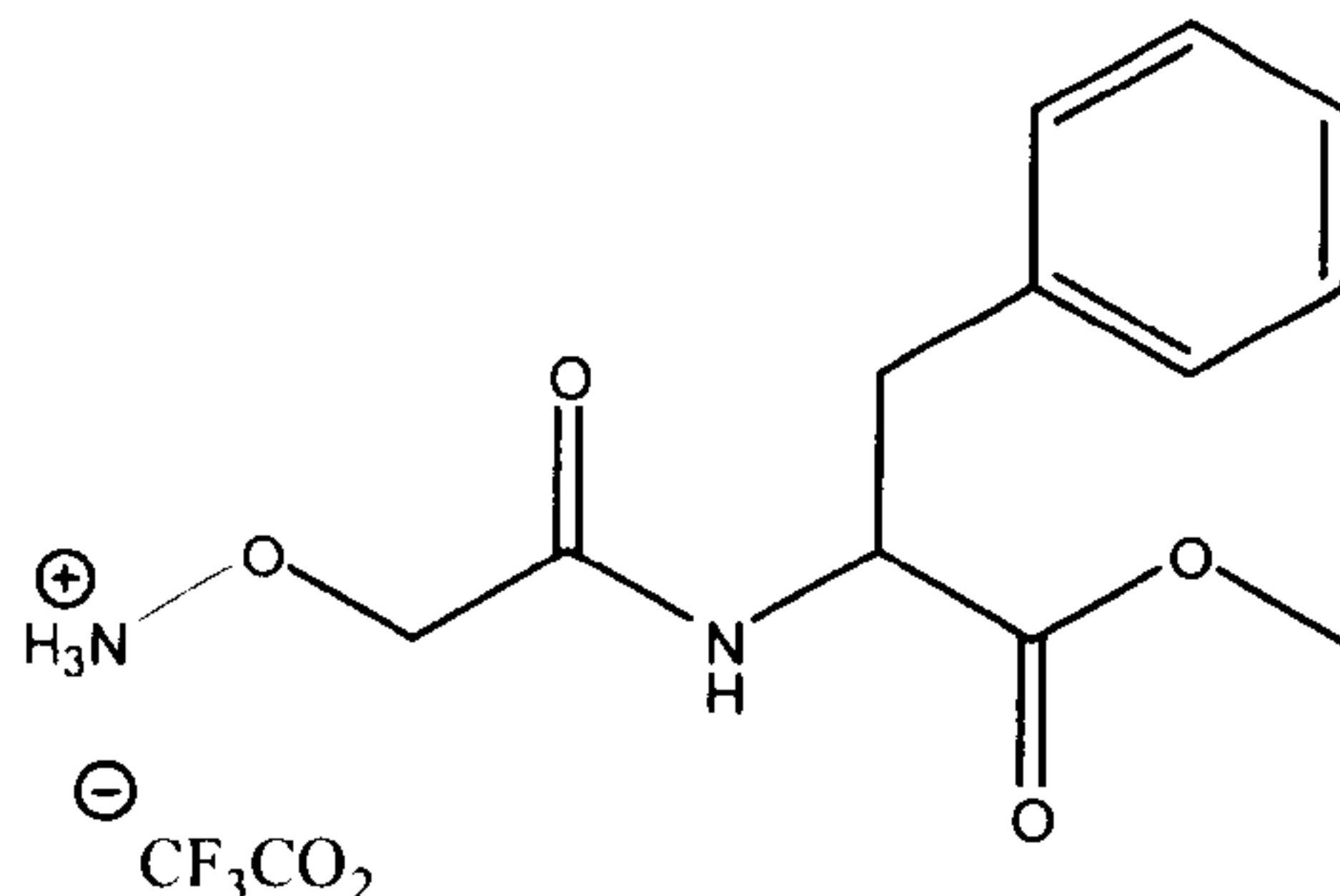
Phenylalanine methyl ester hydrochloride (1.1 g, 5.1 mmol) was converted into the free base as for glycine benzyl ester and also used immediately in the following reaction.

6.5.10 N-(1''-phenylalanine methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide



Phenylalanine methyl ester (210 mg, 1.2 mmol) was prepared the same as N-(1''-glycine methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide, using succinimido-(N)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanoate (346 mg, 1.2 mmol)

Yield 93 mg, 25%; ¹NMR δ 1.4 (s, 9H, (CH₃)₃), 3.8 (s, 3H, methyl), 4.3 (s, 2H, OCH₂) 4.6 (s, 2H, NHCH₂), 4.8 (m, 1H, CH₂CHNH), 7.2 (s, 5H, phenyl);

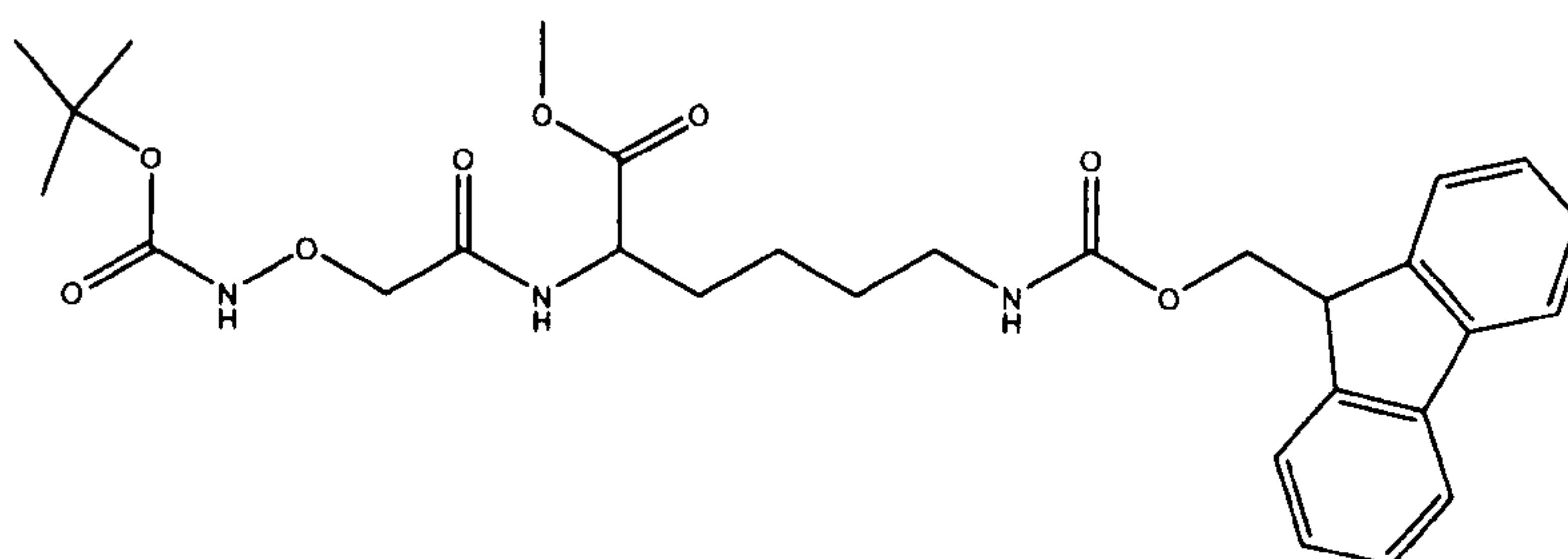
6.5.11 N-(1''-phenylalanine methyl ester)aminoxy) ethanamide (22)

N-(1''-phenylalanine methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide (90 mg, 0.24 mmol) was treated the same as N-(1''-glycine benzyl ester)aminoxy ethanamide and the product was a pale yellow oil.

Yield 86 mg, 95%; 3.6 (s, 3H, methyl), 4.3 (s, 2H, OCH₂) 4.6 (s, 2H, NHCH₂), 4.8 (m, 1H, CH₂CHNH), 7.1 (s, 5H, phenyl);

6.5.12 Lysine (Fmoc) methyl ester

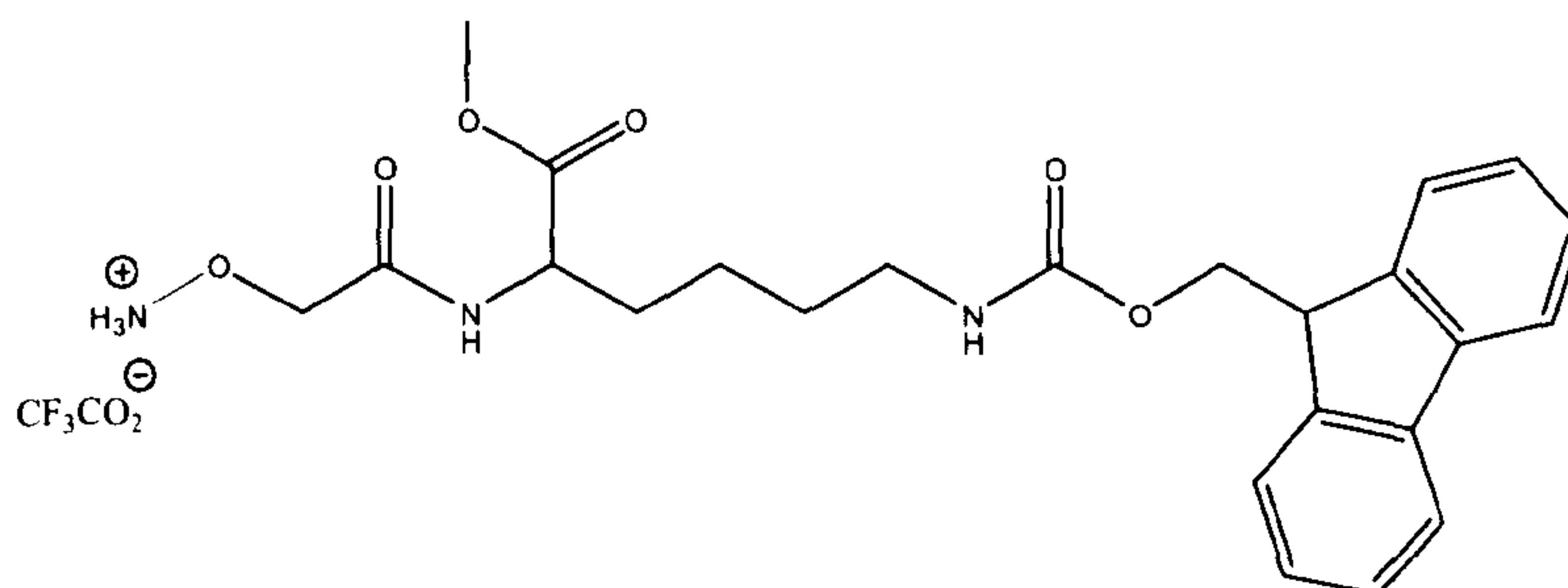
N-ε-Fmoc-lysine methyl ester hydrochloride (1.1 g, 2.6 mmol) was converted into the free base in the same way as for glycine benzyl ester and also used immediately in the following reaction.

6.5.13 N-(1''-lysine (FMOC) methyl ester)-((1',1'-dimethylethoxycarbonyl) aminoxy) ethanamide

N- ϵ -Fmoc-lysine methyl ester (992 mg, 2.6 mmol) treated in the same way as phenylalanine methyl ester, using succinimido-(N)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanoate (750 mg, 2.6 mmol).

Yield 510 mg, 47%; ^1NMR δ 1.3 (m, 2H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 1.4 (m, 2H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 1.6 (m, 2H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 1.5 (s, 9H, $(\text{CH}_3)_3$), 3.6 (m, 2H, $\text{CH}_2\text{CH}_2\text{NH}$), 3.7 (s, 3H, methyl), 4.3 (m, 1H, NHCH), 4.4 (m, 1H, NHCH), 4.5 (s, 2H, OCH_2), 4.7 (m, 2H, OCH_2CH), 7.4 (m, 4H, Fmoc), 7.6 (m, 4H, Fmoc);

6.5.14 N-(1''-lysine (FMOC) methyl ester)aminoxy) ethanamide (24)

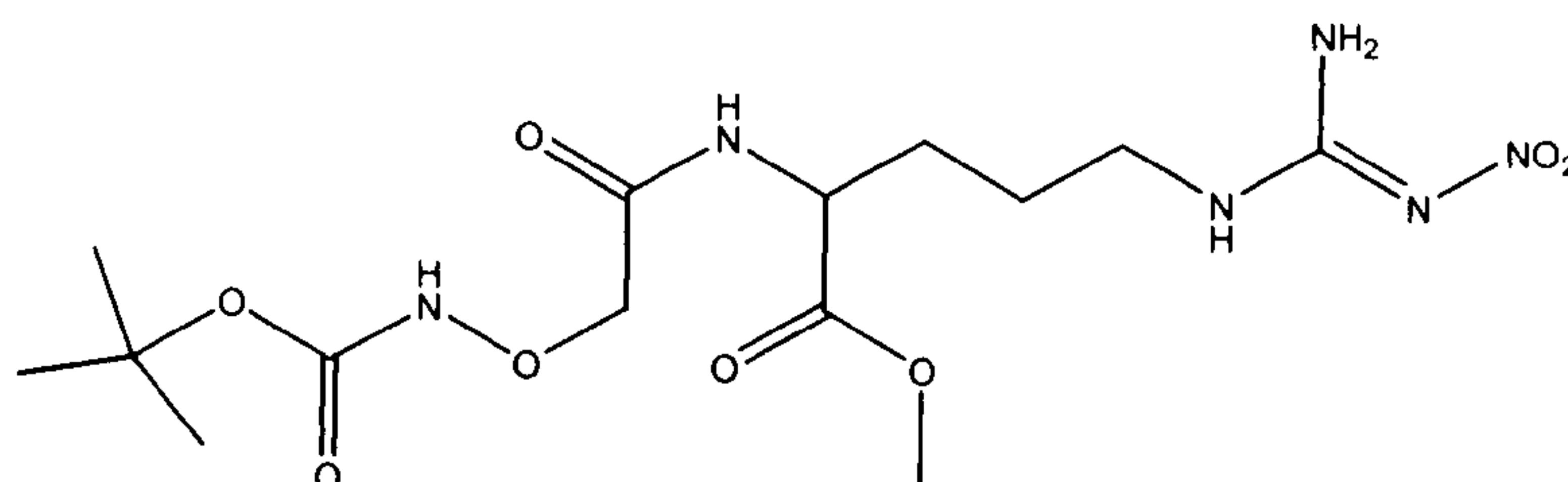


N-(1''-lysine (FMOC) methyl ester)-((1',1'-dimethylethoxycarbonyl) aminoxy) ethanamide (488 mg, 1.2 mmol) was treated in the same way as was treated the same as N-(1''-glycine benzyl ester)aminoxy ethanamide. The product was a very sticky semi solid which was dissolved in chloroform and placed under vacuum, which, when removed gave a pale yellow solid.

Yield 460 mg, 69%; $^1\text{HNMR}$ δ 1.2 (m, 2H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 1.4 (m, 4H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 1.7 (m, 2H $\text{CH}_2\text{CH}_2\text{CH}_2$), 3.4 (m, 2H, $\text{CH}_2\text{CH}_2\text{NH}$), 3.7 (s, 3H, methyl), 4.3 (s, 2H,

OCH₂) 4.4 (m, 1H, NHCH), 4.7 (m, 2H, OCH₂CH) 7.2 (m, 4H, Fmoc), 7.8 (m, 4H, Fmoc),

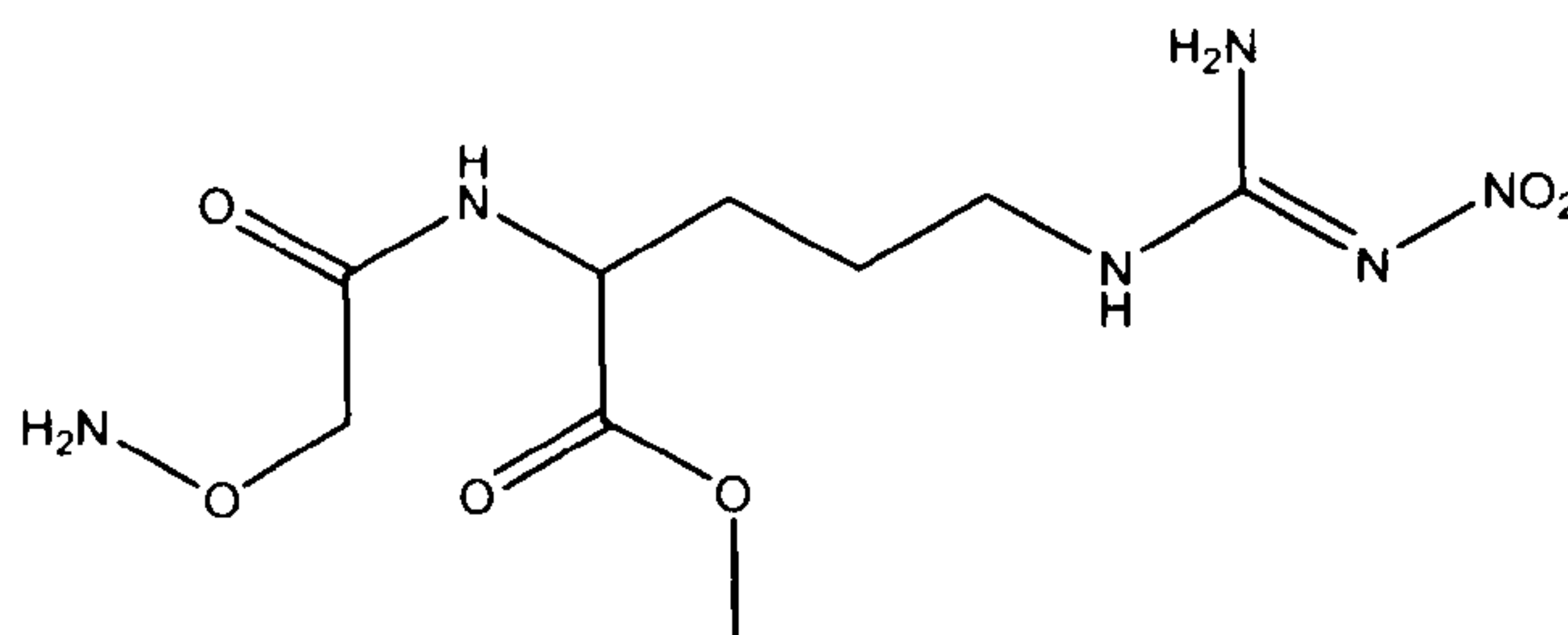
6.5.15 N-(1''-arginine (NO₂) methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide



N^G-Nitro-arginine methyl ester hydrochloride (250 mg, 0.93 mmol) was prepared in the same way as for N-(1''-glycine methyl ester)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide, using triethylamine (2 g, 2 mmol) and succinimido-(N)-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanoate (288mg, 1 mmol). In the clean up the organic layer was yellow but after concentrating and drying under vacuum a white powder was formed.

Yield 202 mg, 52%; ¹NMR δ 1.2 (m, 2H, CH₂CH₂CH₂), 1.5 (m, CH₂CH₂CH₂), 1.5 (s, 9H, (CH₃)₃), 2.9 (m, 2H, CH₂CH₂NH), 3.7 (s, 3H, methyl), 4.3 (s, 2H, OCH₂) 4.5 (d, 1H, NHCH); ¹³C NMR δ 28 ((CH₃)₃), 30 (CH₂CH₂), 41 (CH₂CH₂C), 46 (OCH₃), 53 (NHCH), 76 (C(CH₃)₃), 83 (NOC), 158 (OC=O), 170 (NCN), 171 (CH₂CONH), 173 (CHC=O),

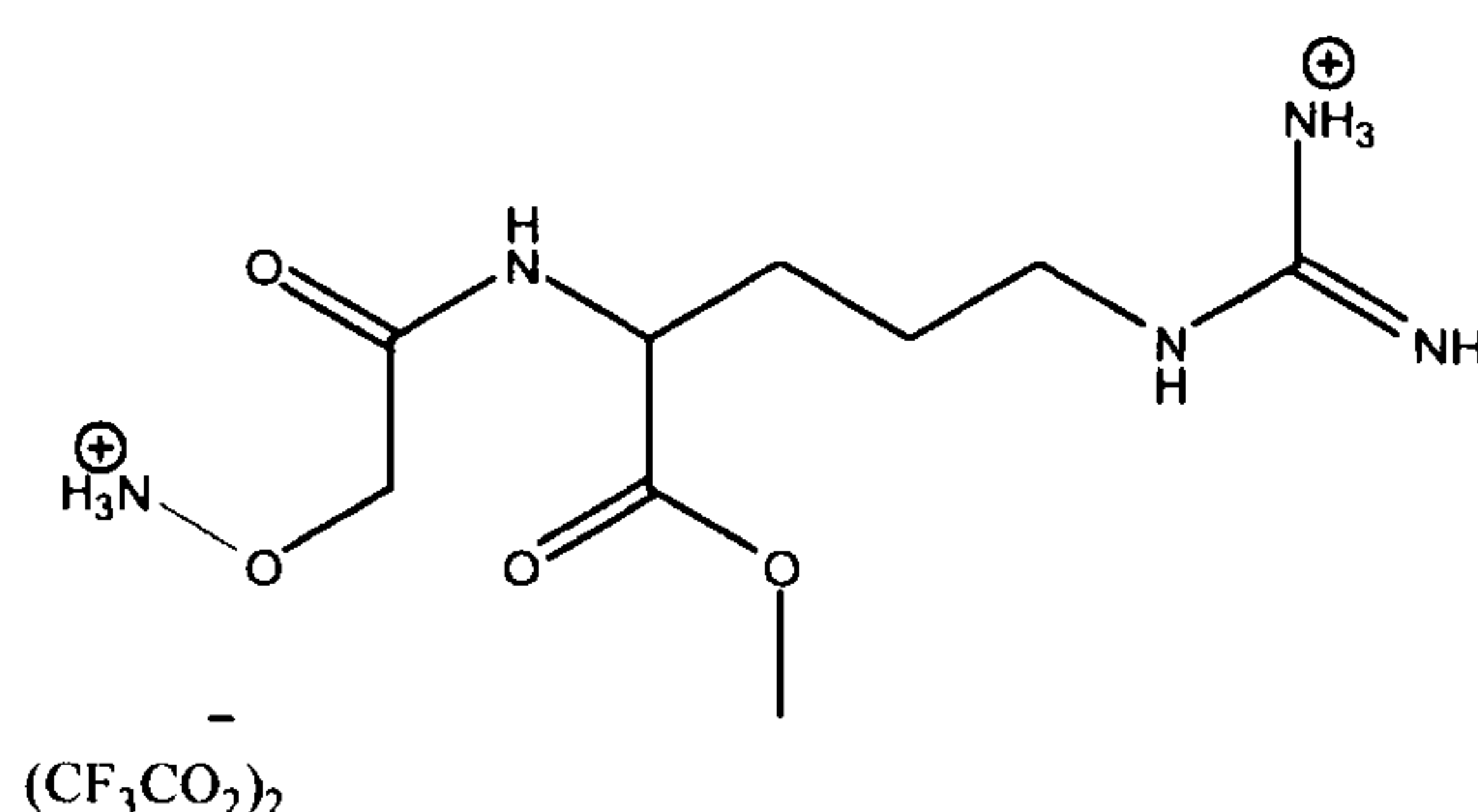
6.5.16 N-(1''-arginine (N^G-NO₂) methyl ester)aminoxy) ethanamide



The t-butyl protecting group was removed from N-(1''-carboxymethyl arginine (NO₂))-((1',1'-dimethylethoxycarbonyl)aminoxy) ethanamide (760 mg, 1.9 mmol) in the same way as in N-(1''-carboxymethyl phenylalanine)aminoxy) ethanamide.

Yield 580 mg, 73%; ¹NMR δ 1.3 (t, 2H, CH₂CH₂CH₂), 1.8 (m, CH₂CH₂CH₂), 3.2 (m, 2H, CH₂CH₂NH), 3.8 (s, 3H, methyl), 4.5 (s, 2H, OCH₂) 4.6 (m, 1H, NHCH); ¹³C NMR δ 20 (CH₂CH₂), 41 (CH₂CH₂C), 52 (OCH₃), 53 (NHCH), 72 (OCH₂CO), 161 (CH₂OC), 170 (NCN), 173 (OCO),

6.5.17 N-(1''-arginine methyl ester)aminoxy) ethanamide (23)



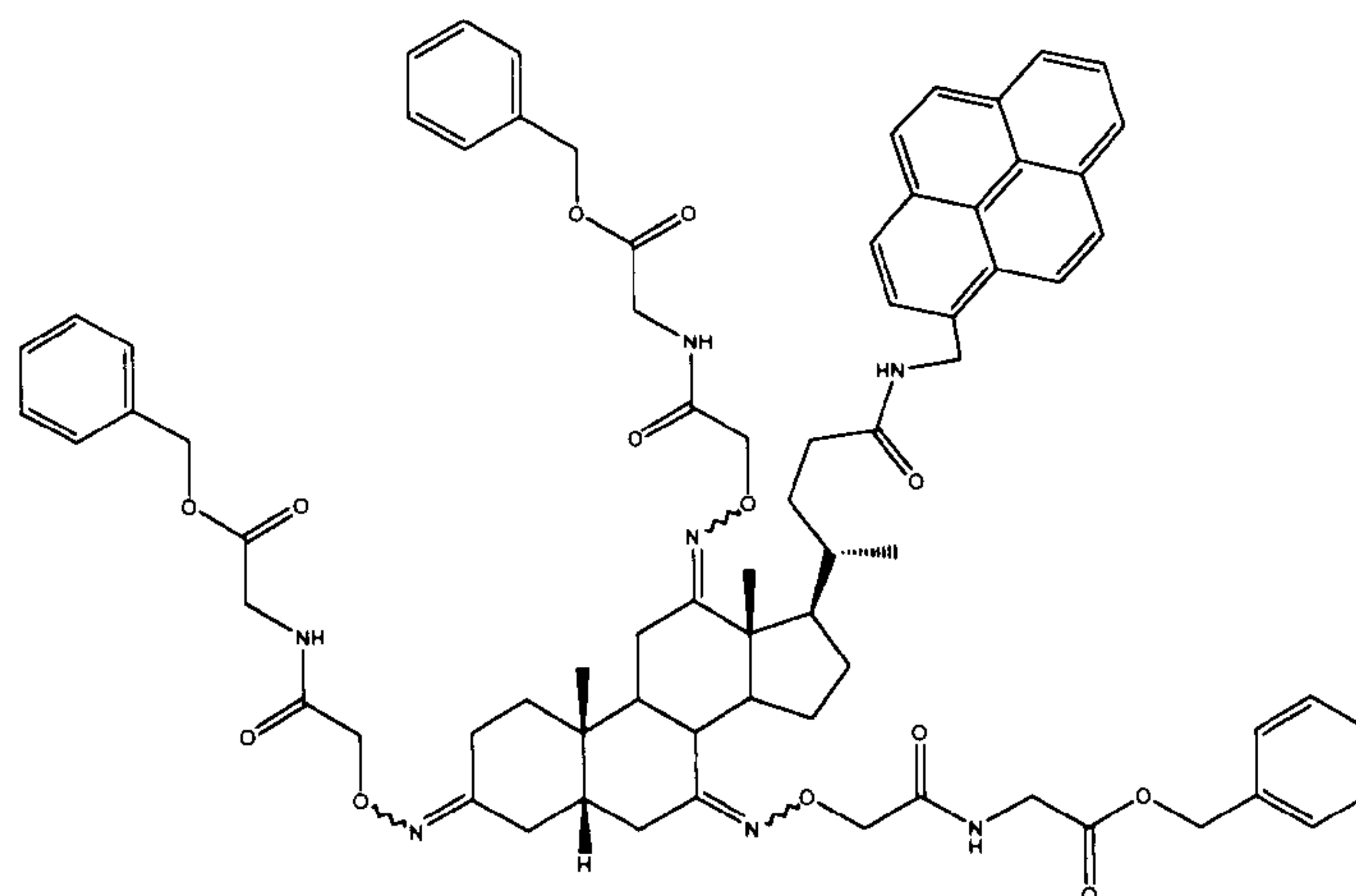
The NO₂ protecting group on the arginine side chain was removed by hydrogenation with palladium/charcoal catalyst. N-(1''-arginine (NO₂) methyl ester)aminoxy) ethanamide trifluoroacetate salt (500 mg, 1.2 mmol) was dissolved in methanol and

the reaction was complete after 3 hours. The mixture was filtered and the solvent removed under vacuum.

Yield 412 mg, 91%; ^1H NMR 1.3 (t, 2H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 2.0 (m, $\text{CH}_2\text{CH}_2\text{CH}_2$), 3.4 (m, 2H, $\text{CH}_2\text{CH}_2\text{NH}$), 3.8 (s, 3H, methyl), 4.5 (s, 2H, OCH_2) 4.6 (s, 1H, NHCH); ^{13}C NMR δ 25 (CH_2CH_2), 41 ($\text{CH}_2\text{CH}_2\text{C}$), 49 (OCH_3), 53 (NHCH), 75 (OCH_2CO), 158 (CH_2OC), 160 (NCN), 172 (OCO),

6.6 Synthesis of tri substituted 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide

6.6.1 1'-pyrenyl methyl-3,7,12 tri(N-iminomethylenecarbonyl (glycine benzyl ester)) (25)

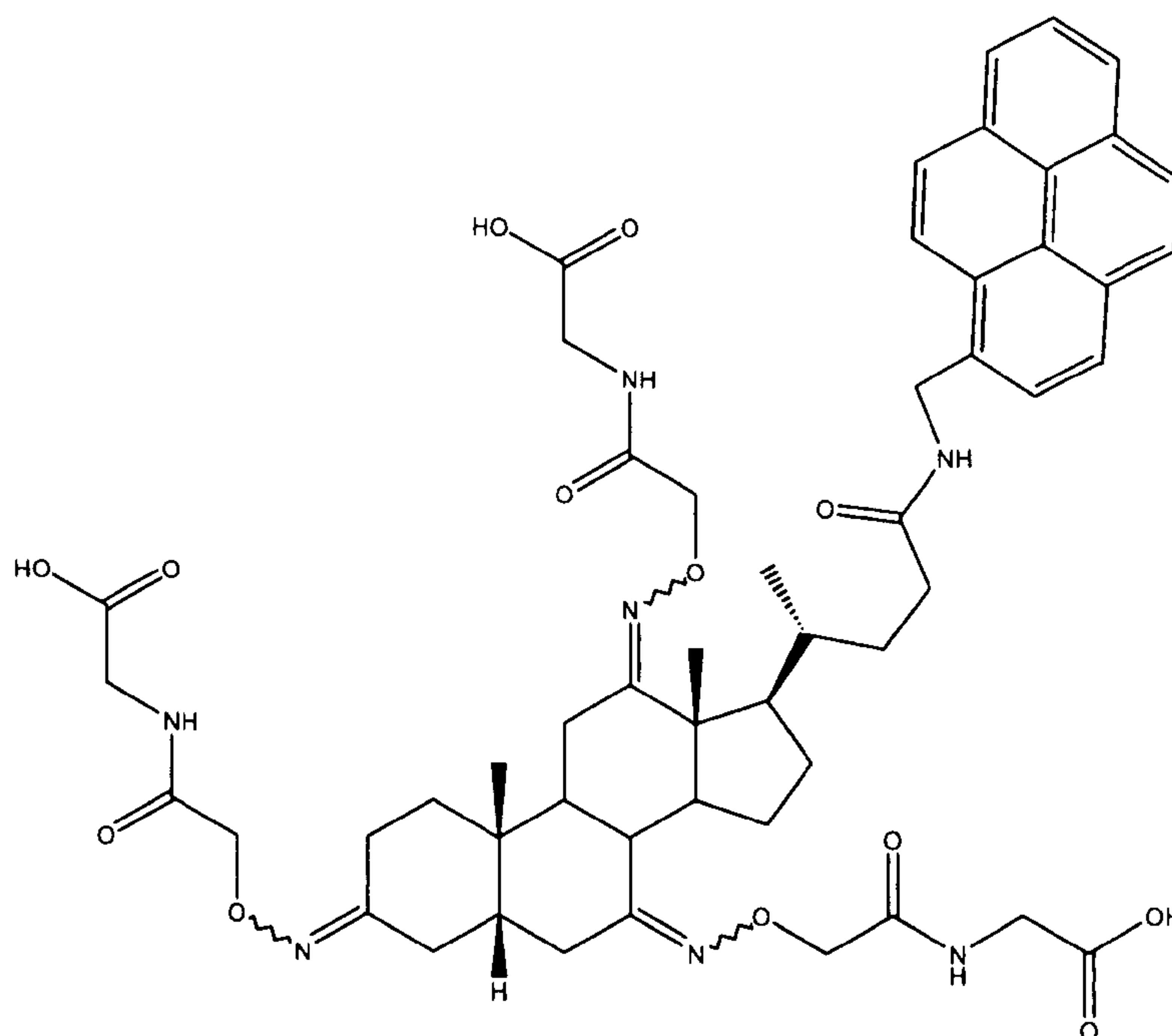


1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (20 mg 0.033 mmol) was suspended in ethanol (2 mL), N-(1''-glycine benzyl ester)aminoxy ethanamide (40 mg, 0.12 mmol) and sodium acetate (16 mg, 0.194 mmol) were added and the mixture was allowed to stir at room temperature for 1 hour. The pale yellow solution was then refluxed overnight. The solvent was then removed under vacuum and enough 0.1M HCl added to dissolved the solid. This was extracted thrice with chloroform, which was then dried with sodium sulphate and the solvent removed again under vacuum. The resulting yellow glass was dissolved in chloroform and a few drops of diethyl

ether added, this was left in the freezer and the pale yellow solid was filtered off and dried.

Yield 22 mg, 52%; λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES-) 1274 (M-1) ³⁵, ¹H NMR δ 0.8 (m, 22H, steroid), 1.3 (m, 6H, steroid), 4.0 and 4.6 (m, 14H, (OCH₂CONHCH₂)₃ and NHCH₂), 5.1 (m, 6H, CH₂C₆H₅), 7.2 (m, 15H, C₆H₅), 8.0 (m, 9H, pyrenyl).

6.6.2 1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine-(5 β)-cholan-24-amide (26)

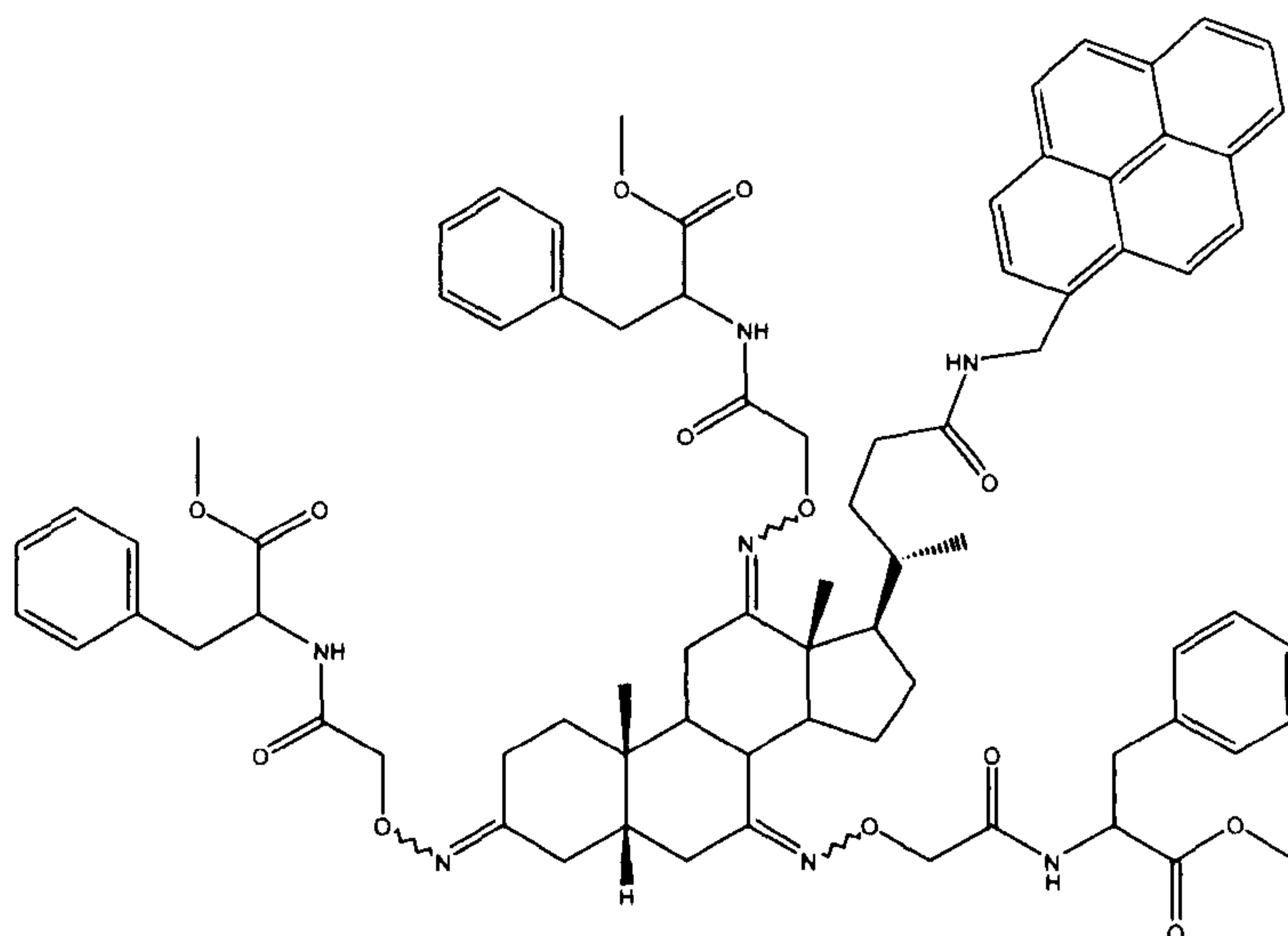


1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine-(5 β)-cholan-24-amide was prepared in the same way as for 1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine benzyl ester-(5 β)-cholan-24-amide, using 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (167 mg, 0.27 mmol), N-(1''-

glycine methyl ester) aminoxy ethanamide (250 mg, 0.9 mmol) and sodium acetate (220 mg, 1.62 mmol). The resulting compound gave m/z ES⁺ of 1049 (M+1)₁₀₀, the methyl protecting groups were removed by the addition of sodium hydroxide (0.1 M, 1 mmol) until there was full conversion to a baseline spot by TLC.

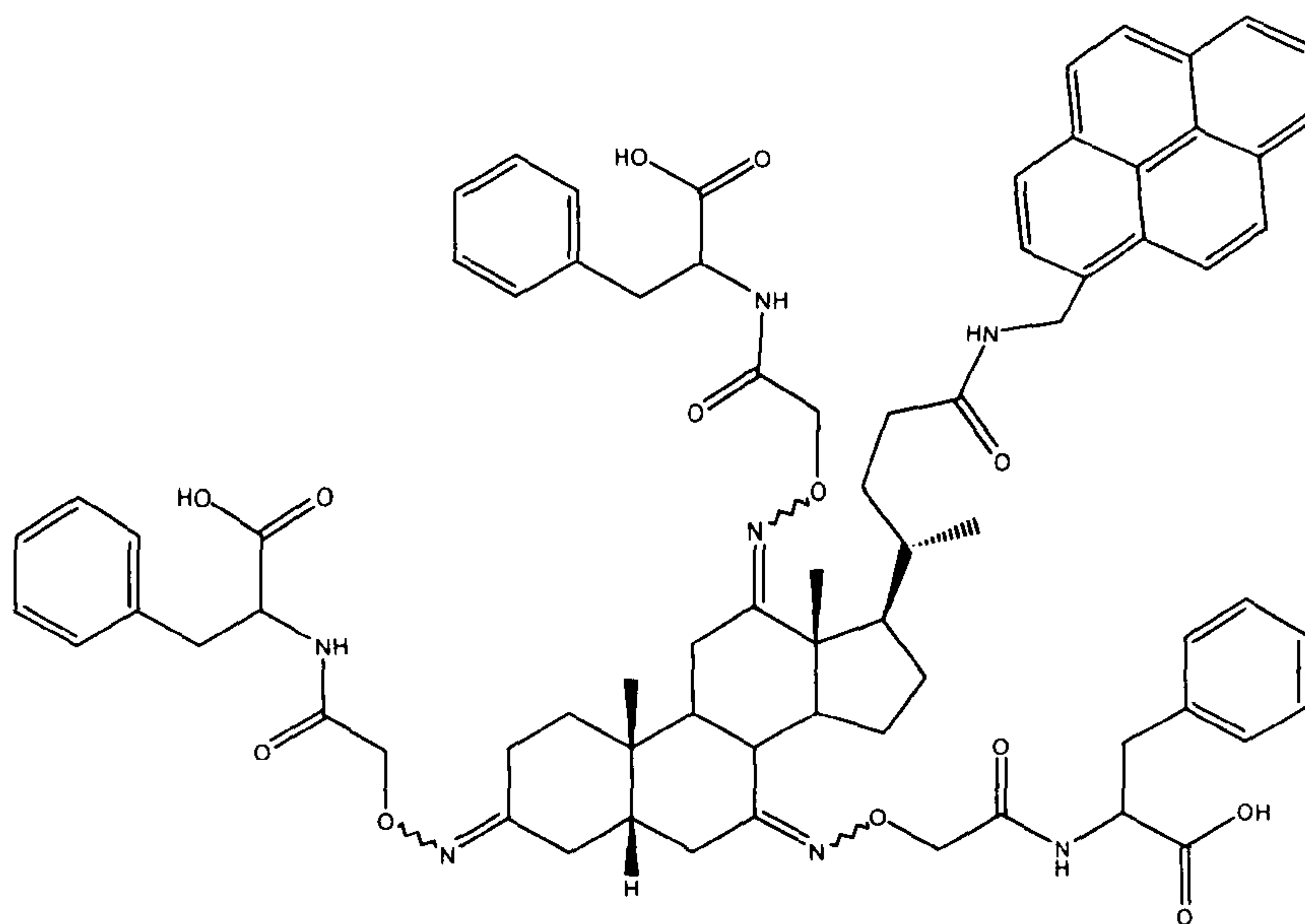
Yield 220 mg, 78%; λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES⁺) 1007 (M+1)₂₀, 874 (M- OCH₂CONHCH₂COOH +1)₃₀, 743 (M - 2(OCH₂CONHCH₂COOH) +1)₁₀₀; ¹H NMR δ 1.8 (M, 33H, steroid), 4.4 and 5.0 (m, 14H, (OCH₂CONHCH₂)₃ and NHCH₂), 8.1 (m, 9H, pyrenyl).

6.6.3 1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) phenylalanine)-(5 β)-cholan-24-amide (28)



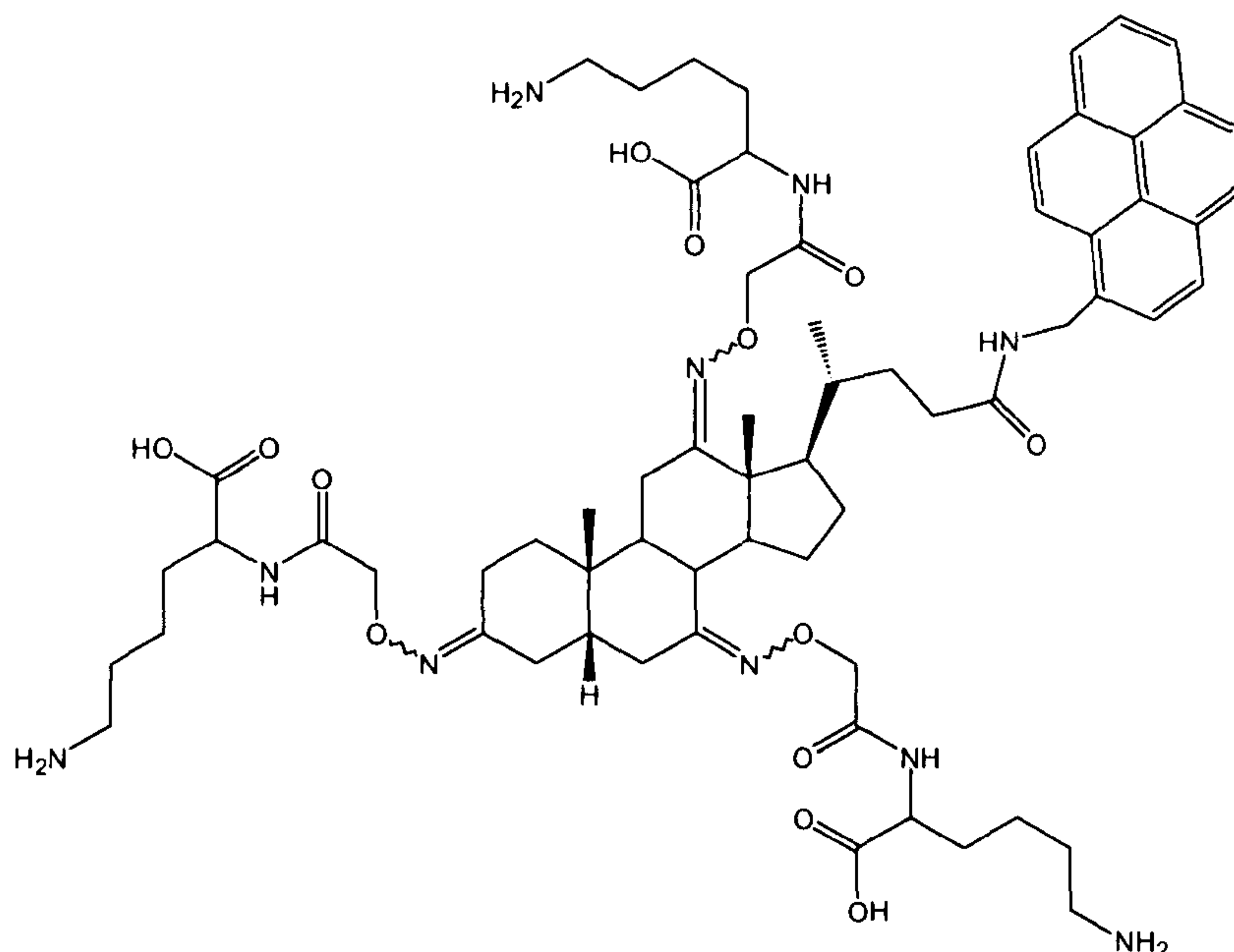
1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) phenylalanine)-(5 β)-cholan-24-amide was prepared in the same way as for 1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine benzyl ester)-(5 β)-cholan-24-amide, using 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (100 mg, 0.16 mmol), N-(1''-phenylalanine methyl ester)aminoxy) ethanamide (212 mg, 0.58 mmol) and sodium

acetate (131 mg, 0.96 mmol). This compound gave the following data: MS m/z (ES+) 1318 (M+1)₃₀, 1083 (M- OCH₂CONHCHCH₂C₅H₆COOH +1)₆₀, 847 (M-(OCH₂CONHCHCH₂C₅H₆COCH₃)₂ + 1)₁₀₀; ¹H NMR δ 1.3 (M, 33H, steroid), 3.2 (m, 6H, CH₂C₆H₅), 3.6 (m, 9H, (CH₃)₃), 3.8 and 4.8 (m, 14H, (OCH₂CONHCH₂)₃ and NHCH₂), 7.1 (m, 15H, C₆H₅), 8.1 (m, 9H, pyrenyl), and also underwent deprotection by base treatment as before.



Yield 200 mg, 94%; λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES-) 1275 (M-1)_{<10}, 1054 (M- OCH₂CONHCHCH₂C₅H₆COOH -1)₃₀, 833 (M- 2(OCH₂CONHCHCH₂C₅H₆COOH) -1)₇₀, ¹H NMR δ 1.8 (M, 36H, steroid), 3.3 and 4.8 (m, 14H, (OCH₂CONHCH₂)₃ and NHCH₂), 7.1 (m, 15H, C₆H₅), 8.1 (m, 9H, pyrenyl),

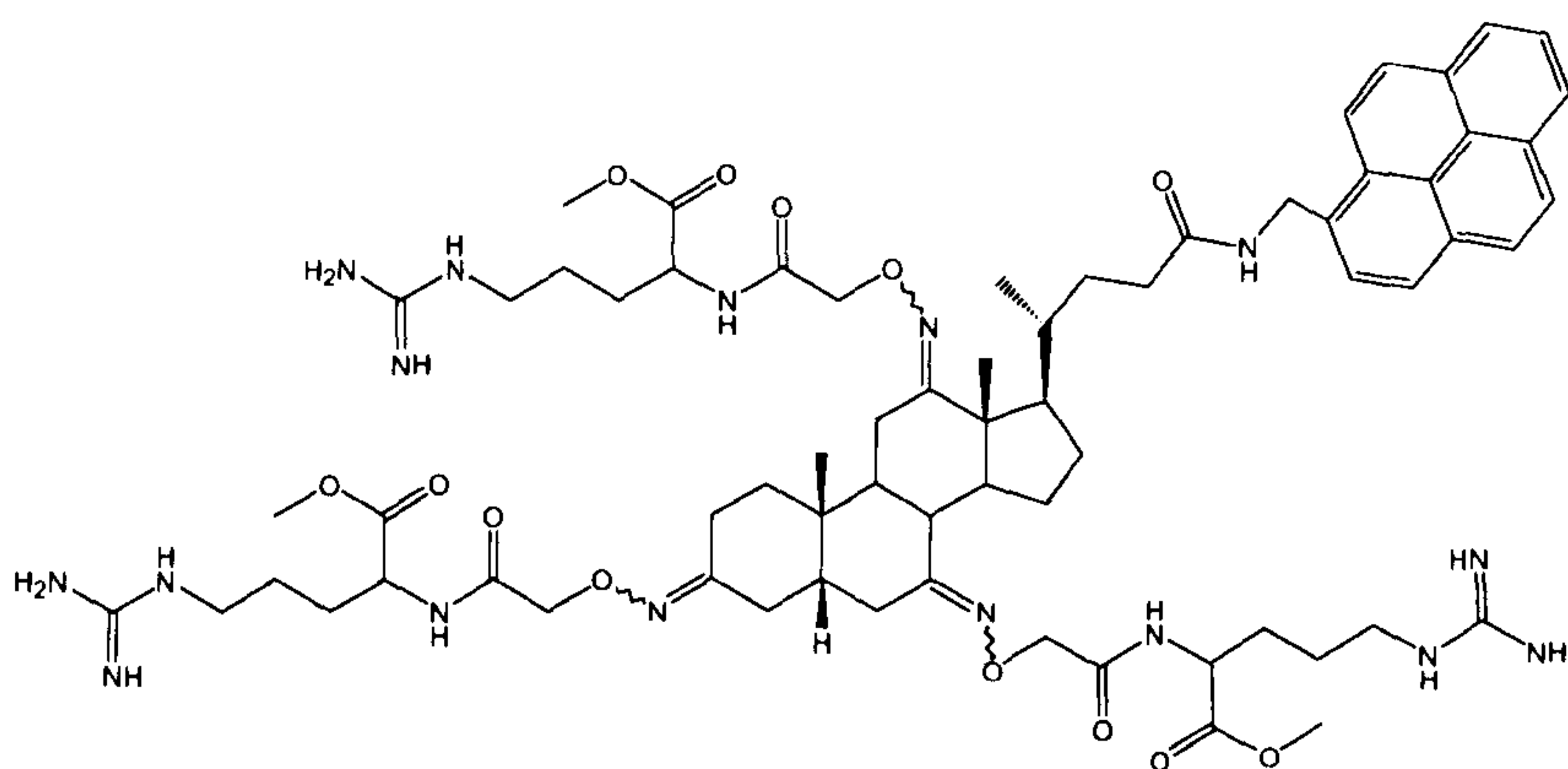
6.6.4 1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) lysine-(5 β)-cholan-24-amide (27)



1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) lysine (Fmoc))-(5 β)-cholan-24-amide was prepared in the same way as 1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) phenylalanine)-(5 β)-cholan-24-amide, using 1'-pyrenylmethyl-3,7,12 trioxo-(5 β)-cholan-24-amide (20 mg, 0.033 mmol), N-(1''-lysine (Fmoc))aminoxy) ethanamide (50 mg, 0.09 mmol) and sodium acetate (73 mg, 0.54 mmol). 1'-pyrenylmethyl-3,7,12 (tri(N-inoxymethylenecarbonyl) lysine (Fmoc))-(5 β)-cholan-24-amide (50 mg, 0.024 mmol) was dissolved in piperidine and DMF (2:8 v/v) and stirred at room temperature for 5 minutes, this was repeated two more times and washed in DMF which was removed under vacuum, followed by sodium hydroxide to remove the methyl protecting group as before.

Yield 30 mg, 83%; λ_{\max} 242, 276, 326, 342 nm; MS m/z (ES⁺) 837 (M-(OCH₂CONHCH(CH₂)₄NH₂COOH)₂ + 1)₃₀; ¹H NMR δ 1.3 (m, 71H, steroid), 2.6 (m, 6H, CH₂NH₂), 4.8 (m, 14H, (OCH₂CONHCH₂)₃ and NHCH₂), 8.1 (m, 9H, pyrenyl),

6.6.5 1'-pyrenylmethyl-3((N-iminoxymethylenecarbonyl) arginine) 7,12(dioxo)-(5 β)-cholan-24-amide (29)



This was prepared in the same way as for 1'-pyrenylmethyl-3,7,12 (tri(N-iminoxymethylenecarbonyl) glycine benzyl ester)-(5 β)-cholan-24-amide, using 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (58 mg, 0.095 mmol), N-(1''-arginine)aminoxy) ethanamide (37 mg, 0.1 mmol) and sodium acetate (82 mg, 0.6 mmol)

Yield 61 mg, 75%; λ_{\max} 242, 276, 326, 342 nm; ¹H NMR 1.5 (m, 42H, steroid), 1.8 (m, 12H, 3(CH₂CH₂)), 3.5 (m, 6H, 3(NOCH₂)), 4.5 (m, 3H, 3(NHCH₂CH₂)), 5.1 (d, 6H, ³J 5.3, 3(NHCH₂)), 8.1 (m, 9H, pyrenyl).

6.6.6 **1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide tri substituted with a mixture of methyl and benzyl hydroxylamines**

1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide tri substituted with a mixture of methyl and benzyl hydroxylamines was prepared as in 1'-pyrenyl methyl-3,7,12 trioximo-(5 β)-cholan-24-amide, using 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (150 mg, 0.24 mmol), sodium acetate (200mg, 1.46 mmol) methylhydroxylamine hydrochloride (40 mg, 0.45 mmol) benzylhydroxylamine hydrochloride (78 mg, 0.49 mmol)

yield 200mg; λ_{\max} 242, 276, 326, 342 nm;

6.6.7 **1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide library containing all aminooxy amino acid derivatives**

1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide (0.5 g, 0.81 mmol) was suspended in ethanol (150 mL) and left to stir at room temperature. N-(1''-lysine (FMOC) methyl ester)aminooxy) ethanamide (278 mg, 0.5 mmol), N-(1''-arginine methyl ester)aminooxy) ethanamide (224 mg, 0.6 mmol), N-(1''-phenylalanine methyl ester)aminooxy) ethanamide (183 mg, 0.5 mmol), N-(1''-glycine benzyl ester)aminooxy) ethanamide (176 mg, 0.5 mmol) and N-(1''-glycine methyl ester)aminooxy) ethanamide (138 mg, 0.5 mmol) were all dissolved in ethanol (0.5 mL) each and added to the 1'-pyrenyl methyl-3,7,12 trioxo-(5 β)-cholan-24-amide suspension. After the addition of sodium acetate (660 mg, 4.9 mmol) the mixture was allowed to stir at room temperature for one hour, it was then heated and left refluxing overnight. TLC, chloroform and methanol (95:5 v/v), showed the conversion of all the starting material. The clean up was as before to give 0.91 g. The deprotection of the Lys side chains was carried out as before. The methyl groups protecting the carboxyl

functionality were removed by adding sodium hydroxide (0.1M, 2.5 mL) and left to stir overnight. The methanol was removed under vacuum and the pH reduced to 7 by the dropwise addition of HCl (0.1M). The precipitate was washed thrice with water and centrifuged each each, with the supernatant being removed, finally the product was freeze dried.

Yield 610 mg, mmol; λ_{\max} 242, 276, 326, 342 nm;

6.7 Sakaguchi Reagent

Solution A: 16% Urea, 0.2% 1-naphthol in ethanol/water 1:1

Solution B: 500 mL aq. NaOH, 3.3 mL bromine

Spray solution A on the sample, dry at room temperature and spray solution B.

Positive for arginine and other guandine derivatives appear as pink/red spots

Chapter 7

References

7.1 References

- (1) Harris, C. C. p53 Tumor Suppressor gene: at the Crossroads of Molecular Carcinogenesis, Molecular Epidemiology and Cancer Risk Assessment. *Toxicology Letters* **1995**, 82/83, 1-7
- (2) Hollstein, M.; Sidransky, D.; Vogelstein, B.; Harris, C. C. p53 Mutations in Human Cancer. *Science* **1991**, 253, 49-53.
- (3) Smooker, P. M.; Cotton, R. G. H. The use of Chemical Reagents in the Detection of DNA Mutations. *Mutation Research* **1993**, 288, 65-77.
- (4) www.statistics.gov.uk National Statistics Online; National Statistics, 2003.
- (5) Farmer, P. B. Monitoring of Human Exposure to Carcinogens through DNA and Protein Adduct Determination. *Toxicology Letters* **1995**, 82/83, 757-762.
- (6) Harris, C. C. Chemical and Physical Carcinogenesis: Advances and Perspectives for the 1990's. *Cancer Research* **1991**, 51, 5023-5044.
- (7) Miller, E. C. Some Current Perspectives on Chemical Carcinogenesis in Humans and Experimental Animals: Presidential Address. *Cancer Research* **1978**, 38, 1479-1496.
- (8) www.cancernet.nci.nih.gov CancerNet Stomach - US Racial/ethnic Cancer Patterns; National Cancer Institute, 2000.
- (9) www.who.int/en World Health Organization Mortality Database; WHO, 2000.
- (10) Correa, P. Human Gastric Carcinogenesis: A Multistep and Multifactorial Process - First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Research* **1992**, 52, 6735-6740.
- (11) Seidel, W.; Pischetsrieder, M. Reaction of Guanosine with Glucose under Oxidative Conditions. *Bioorganic and medicinal chemistry letters* **1998**, 8, 2017-2022.

- (12) Henderson, R. F. Strategies for use of Biological Markers of Exposure. *Toxicology Letters* **1995**, 82/83, 379-383.
- (13) Shuker, D. E. G. The Enemy at the Gates? DNA adducts as biomarkers of exposure to exogenous and endogenous genotoxic agents. *Toxicology Letters* **2002**, 134, 51-56.
- (14) Lijinsky, W. *Chemistry and Biology of N-nitroso Compounds*; Cambridge University Press: Cambridge, 1992.
- (15) Preussmann, R. *Occurrence and Exposure to N-nitroso Compounds and Precursors*; IARC Scientific Publications, Lyon; 1983 pp 3-7.
- (16) Lu, S. H.; Ohshima, H.; Bartsch, H. Recent Studies on N-Nitroso Compounds as Possible Etiological Factors in Oesophageal Cancer. *N-nitroso compounds. Occurrence, biological effects and relevance to human cancer* **1983**, 57, 947-953.
- (17) Hecht, S. S.; Hoffman, D. N-Nitroso Compounds and Tobacco-Induced Cancers in Man. *Relevance to Human Cancer of N-Nitroso Compounds, Tobacco Smoke and Mycotoxins*; IARC Scientific Publications: Lyon, 1991; pp 55-63.
- (18) Preussmann, R. Occurrence and Exposure to N-Nitroso Compounds and Precursors. *Relevance to Human Cancer of N-Nitroso Compounds, Tobacco Smoke and Mycotoxins*; IARC Scientific Publications: Lyon, 1991; pp 3-14.
- (19) Lu, S. H.; Chui, S. X.; Yang, W. X.; Hu, X. N.; Guo, L. P. et al. Relevance of N-Nitroamines to Oesophageal Cancer in China. *Relevance to Human Cancer of N-Nitroso Compounds, Tobacco Smoke and Mycotoxins*; IARC Scientific Publications: Lyon, 1991; pp 11-17.

- (20) Anderson, L. M.; Souliotis, V. L.; Chhabra, S. K.; Moskal, T. J.; Harbaugh, S. D. et al. N-Nitrosodimethylamine-derived O6-methylguanine in DNA of Monkey Gastrointestinal and Urogenital Organs and Enhancement by Ethanol. *Int. J. Cancer* **1996**, *66*, 130-134.
- (21) Shuker, D. E. G. DNA Adducts in Mammalian Cells as Indicators of Exposure to Carcinogens. *IARC Scientific Publications* **1999**, *146*, 287-308.
- (22) Gottschalg, E. Detection, Stability and Factors Influencing the Formation of Promutagenic Endogenous DNA Damage. In *PhD Thesis*; University of Leicester: Leicester, 2002; pp 232.
- (23) Leach, S. A.; Mackerness, C. W.; Hill, M. J.; Thompson, M. H. Inhibition of Bacterially Mediated N-Nitrosation by Ascorbate: Therapeutic and Mechanistic Considerations. *Relevance to Human Cancer of N-Nitroso Compounds, Tobacco Smoke and Mycotoxins*; IARC Scientific Publications: Lyon, 1991; pp 571-578.
- (24) Miller, J. A. Carcinogenesis by Chemicals: An Overview - G. H. A. Clowes Memorial Lecture. *Cancer Research* **1970**, *30*, 559-576.
- (25) Shephard, S. E.; Lutz, W. K. Nitrosation of Dietary Precursors. *Cancer Surveys* **1989**, *8*, 402-423.
- (26) Wiestler, o. D.; Uozumi, A.; Kleihues, P. DNA Methylation by N-nitrosomethylbenzylamine in Target and Non-target Tissues of Laboratory Rodents. Comparison with Carcinogenicity. *N-nitroso compounds. Occurrence, biological effects and relevance to human cancer* **1983**, *57*, 595-601.

- (27) Shuker, D. E. G.; Tannenbaum, S. R.; Wishnok, J. S. N-Nitroso Bile Acid Conjugates. Synthesis, Chemical Reactivity and Mutagenic Activity. *The Journal of Organic Chemistry* **1981**, *46*, 2092-2096.
- (28) Dayal, B.; Bhojawala, J.; Rapole, K. R.; Pramanik, B. N.; Ertel, N. H. et al. Chemical Synthesis, Structural Analysis and Decomposition of N-Nitroso Bile Acid Conjugates. *Bioorganic and Medicinal Chemistry* **1996**, *4*, 885-890.
- (29) Paju, S.; Shuker, D. E. G.; Bishop, W. W.; Falchuk, K. R.; Tannenbaum, S. R. et al. Mutagenicity of N-Nitroso Bile Acid Conjugates in *Salmonella typhimurium* and Diploid Human Lymphoblasts. *Cancer Research* **1982**, *42*, 2601-2604.
- (30) Shuker, D. E. G.; Howell, J. R.; Street, B. W. Formation and Fate of Nucleic Acid and Protein Adducts Derived from N-nitroso-bile acid Conjugates. *IARC Scientific Publications* **1986**, *84*, 187-190.
- (31) EUROGAST O⁶-Methylguanine in Blood Leucocyte DNA: an association with the geographic prevalence of gastric cancer and with low levels of serum pepsinogen A, a marker of severe chronic atrophic gastritis. *Carcinogenesis* **1994**, *15*, 1815-1820.
- (32) Cupid, B. C.; Zeng, Z.; Singh, R.; Shuker, D. E. G. Detection of O⁶-Carboxymethyl-2'-deoxyguanosine in DNA Following Reaction of Nitric Oxide with Glycine and in Human Blood DNA Using a Quantitative Immunoslot Blot Assay. *Chem Res Toxicol* **2004**, *17*, 294-300.

- (33) Gottschalg, E.; Scott, G. B.; Harrison, K. L.; Bandaletova, T.; Bailey, N. et al. Endogenous nitrosation of glycine in the etiology of human gastrointestinal tract cancers: DNA adduct detection and p53 mutation spectra. *In press*. **2004**.
- (34) Mondal, P.; Hazarika, K. K.; Deka, R. C. Reactivity of α,β -unsaturated Carbonyl Compounds Towards Nucleophilic Addition Reaction: a local hard-soft acid-base approach. *Phys. Chem. Comm* **2003**, *6*, 24-27.
- (35) Shuker, D. E. G.; Margison, G. P. Nitrosated Glycine Derivatives as a Potential Source of O⁶-Methylguanine in DNA. *Cancer Research* **1997**, *57*, 366-369.
- (36) Pegg, A. E. DNA Repair Pathways and Cancer Prevention. *Advances in Nutrition and Cancer* **1999**, 253-267.
- (37) Lodish, H.; Baltimore, D.; Berk, A.; Zipursky, S. L.; Matsudaira, P. et al. *Molecular Cell Biology*; 3rd ed.; Scientific American Books: New York, 1995.
- (38) Stryer, L. *Biochemistry*; Freeman, 1988.
- (39) Pegg, A. E. Repair of O⁶-alkylguanine by Alkyltransferases. *Mutation Research* **2000**, *462*, 83-100.
- (40) Pegg, A. E.; Byers, T. L. Repair of DNA Containing O⁶-alkylguanine. *The FASEB Journal* **1992**, *6*, 2302-2310.
- (41) Meyer, A. S.; McCain, M. D.; Fang, Q.; Pegg, A. E.; Spratt, T. E. O⁶-Alkylguanine-DNA Alkyltransferases Repair O⁶-Methylguanine in DNA with Michaelis-Menten-like Kinetics. *Chem. Res. Toxicol.* **2003**, *16*, 1405-1409.
- (42) Adamkiewicz, J.; Ahrens, O.; Huh, N.; Rajewsky, M. F.; Spiess, E. High Affinity Monoclonal Antibodies for the Specific Recognition and Quantification of Deoxynucleosides Structurally Modified by N-nitroso

Compounds. *N-nitroso compounds. Occurrence, biological effects and relevance to human cancer* **1983**, 57, 581-587.

- (43) Harrison, K. L.; Fairhurst, N.; Challis, B. C.; Shuker, D. E. G. Synthesis, Characterization, and Immunochemical Detection of O⁶-(Carboxymethyl)-2'-deoxyguanosine: A DNA Adduct Formed by Nitrosated Glycine Derivatives. *Chem. Res. Toxicol.* **1997**, 10, 652-659.
- (44) Harrison, K. L.; Jukes, R.; Cooper, D. P.; Shuker, D. E. Detection of concomitant formation of O⁶-carboxymethyl- and O⁶-methyl-2'-deoxyguanosine in DNA exposed to nitrosated glycine derivatives using a combined immunoaffinity/HPLC method. *Chem Res Toxicol* **1999**, 12, 106-111.
- (45) Shuker, D. E. G. Personal communication, 2001.
- (46) Nehls, P.; Adamkiewicz, J.; Rajewsky, M. F. Immunoslot-blot: A Highly Sensitive Immunoassay for the Quantification of Carcinogenic-modified Nucleosides in DNA. *J. Cancer Res. clin. Oncol.* **1984**, 108, 23-29.
- (47) Leuratti, C.; Singh, R.; Lagneau, C.; Farmer, P. B.; Plastaras, J. P. et al. Determination of Malondialdehyde-induced DNA Damage in Human Tissues Using an Immunoslot Blot Assay. *Carcinogenesis* **1998**, 19, 1919-1924.
- (48) Cupid, B. C.; Zeng Z.; Singh, R.; Shuker, D. E. G. Detection of O⁶carboxymethyl-2'-deoxyguanosine in DNA Following Reaction of Nitric Oxide with Glycine and in Human Blood Using Quantitative Immunoslot Blot Assay. *Chem. Res. Toxicol.* **2004**, 17, 294-300.
- (49) Farmer, P. B.; Shuker, D. E. G. What is the Significance of Increases in Background Levels of Carcinogen-derived Protein and DNA adducts? Some consideration for incremental risk. *Mutation Research* **1999**, 424, 275-286.

- (50) Bunker, J.; Davey, B.; Stewart, M.; Toates, F. *Growing and Responding*; The Open University: London, 1997; 195-197.
- (51) Helene, C.; Maurizot, J.-C. Interactions of Oligopeptides with Nucleic Acids. *CRC Critical Reviews in Biochemistry* **1981**, 213-258.
- (52) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; John Wiley and Sons Ltd: New York, 2000.
- (53) Wedin, R. The Promise of Chemical Genetics. In *Chemistry C and E News*: Washington, 2002; pp 15-17.
- (54) Shuker, D. E. G. Chemical Genetics. *Annual Reports on the Progress of Chemistry* **2004**, 100.
- (55) Stockwell, B. R. Chemical Genetics: ligand-based discovery of gene function. *Nature Reviews Genetics* **2000**, 1, 116-125.
- (56) Gordon, E. M.; Bennett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies and Future Directions. *Medicinal Chemistry* **1994**, 37, 1385-1401.
- (57) Schreiber, S. L. Chemical Genetics Resulting from a Passion for Synthetic Organic Chemistry. *Bioorganic and Medicinal Chemistry* **1998**, 6, 1127-1152.
- (58) Agrawal, S.; Zhao, Q. Antisense Therapeutics. *Current Opinions in Chemical Biology* **1998**, 2, 519-528.
- (59) Dervan, P. B.; Burli, R. W. Sequence-specific DNA Recognition by Polyamides. *Current opinions in Chemical Biology* **1999**, 3, 688-693.
- (60) Dolle, R. E.; Jr., K. H. N. Comprehensive Survey of Combinatorial Library Synthesis: 1998. *Journal of Combinatorial Chemistry* **1999**, 1, 235-282.

- (61) Nielson, J. Combinatorial Synthesis of Natural Products. *Current opinions in Chemical Biology* **2002**, *6*, 297-305.
- (62) Borgia, J. A.; Fields, G. B. Chemical Synthesis of Proteins. *Trends Biotechnol.* **2000**, *18*, 243-251.
- (63) Xu, Q.; Borremans, F.; Devreese, B. A Useful Topological Decapeptide Template for Solution-phase Combinatorial Synthesis of Tetrapodal Libraries. *Tetrahedron Letters* **2001**, *42*, 7261-7263.
- (64) Sculimbrene, B. R.; Miller, S. J. Discovery of a Catalytic Asymmetric Phosphorylation Through Selection of a Minimal Kinase Mimic: A D-myo-Inositol-1-Phosphate. *J. Am. Chem. Soc.* **2001**, *123*, 10125-10126.
- (65) Rose, K. Facile Synthesis of Homogeneous Artificial Proteins. *J. Am. Chem. Soc.* **1994**, *116*, 30-33.
- (66) Still, W. C. Discovery of Sequence-Selective Peptide Binding by Synthetic Receptors Using Encoded Combinatorial Libraries. *American Chemical Society* **1996**, 155-163.
- (67) Mutter, M.; Tuschcherer, G. Non-native Architectures in Protein Design and Mimicry. *Cell. mol. life sci.* **1997**, *53*, 851-863.
- (68) Wilson, S. R.; Tam, W. K.; Grandi, M. J. D.; Cui, W. The Calculation and Synthesis of a Template Molecule. *Tetrahedron* **1993**, *49*, 3655-3663.
- (69) Peluso, S.; Dumy, P.; Nkubana, C.; Yokokawa, Y.; Mutter, M. Solid-Phase Strategies for the Assembly of Template-Based Protein Mimetics. *J. Org. Chem.* **1999**, *64*, 7114-7120.
- (70) Mathieu, M.; Lehmann, C.; Razaname, A.; Tuschcherer, G. Engineering of Zinc Finger and MHC Motifs to Locked-in Tertiary Folds. *Letters to Peptide Science* **1997**, *4*, 95-100.

- (71) Peri, F.; Grell, D.; Dumy, P.; Yokokawa, Y.; Welzenbach, K. et al. Assembly of Binding Loops on Aromatic Templates as VCAM-1 Mimetics. *Journal of Peptide Science* **1999**, *5*, 313-322.
- (72) Dawson, P. E.; Kent, S. B. H. Convenient Total Synthesis of a 4-Helix TASP Molecule by Chemoselective Ligation. *J. Am. Chem. Soc.* **1993**, *115*, 7263-7266.
- (73) Baltzer, L. Functionalization and Properties of Designed Folded Polypeptides. *Topics in Current Chemistry* **1999**, *202*, 39-76.
- (74) Spring, D. R. Diversity-orientated Synthesis; a challenge for synthetic chemists. *Org. Biomol. Chem.* **2003**, *1*, 3867-3870.
- (75) Marcaurelle, L. A.; Seeberger, P. H. Combinatorial Carbohydrate Chemistry. *Current opinions in Chemical Biology* **2002**, *6*, 289-296.
- (76) Kocis, P.; Issakova, O.; Sepetov, N. F.; Lebl, M. Kemp's Triacid Scaffolding for Synthesis of Combinatorial Nonpeptide Uncoded Libraries. *Tetrahedron Letters* **1995**, *36*, 6623-6626.
- (77) Davis, A. P.; Wareham, R. S. Carbohydrate Recognition Through Noncovalent Interactions: A challenge for biomimetic and supramolecular chemistry. *Angew. Chem. Int. Ed.* **1999**, *39*, 2978-2996.
- (78) Alluri, P. G.; Reddy, M. M.; Bachhawat-Sikder, K.; Olivos, H. J.; Kodadek, T. Isolation of Protein Ligands from Large Peptoid Libraries. *J. Am. Chem. Soc.* **2003**, *125*, 13995-14004.
- (79) White, R. R.; Sullenger, B. A.; Rusconi, C. P. Developing Aptamers into Therapeutics. *J. Clin. Invest.* **2000**, *106*, 929-934.
- (80) Scott, J. K.; Smith, G. P. Searching for Peptide Ligands with an Epitope Library. *Science* **1990**, *249*, 386-390.

- (81) Colas, P.; Cohen, B.; Jessen, T.; Grishina, I.; McCoy, J. et al. Genetic Selection of Peptide Aptamers that Recognize and Inhibit Cyclin-Dependent Kinase 2. *Nature* **1996**, *380*, 548-550.
- (82) Woiwode, T. F.; Haggerty, J. E.; Katz, R.; Gallop, M. A.; Barrett, R. W. et al. Synthetic Compound Libraries Displayed on the Surface of Encoded Bacteriophage. *Chemistry and Biology* **2003**, *10*, 847-858.
- (83) Davis, A. P. Cholaphanes et al.; Steroids as Structural Components in Molecular Engineering. *Chemical Society Reviews* **1993**, 243-253.
- (84) Hanson, J. R. Steroids: reactions and partial synthesis. *Natural Products Reports* **1997**, *14*, 373-386.
- (85) Wallimann, P.; Marti, T.; Furer, A.; Diederich, F. Steroids in Molecular Recognition. *Chem. Rev.* **1997**, *97*, 1567-1608.
- (86) Hirayama, Y.; Iwamura, M.; Furuta, T. Design, Synthesis and Photochemical Properties of Caged Bile Acids. *Bioorganic and Medicinal Chemistry* **2003**, *13*, 905-908.
- (87) Davis, A. P.; Joos, J.-B. Steroids as Organising Elements in Anion Receptors. *Coordination Chemistry Reviews* **2003**, *240*, 143-156.
- (88) Sisson, A. L.; Clarke, J. P.; Taylor, L. H.; Charmant, J. P. H.; Davis, A. P. Pertubing the Hofmeister Series: a steroid-based anion receptor with preorganised quaternaty ammonium and H-bond donor groups. *Chem Comm* **2003**, 2246-2247.
- (89) Ronsin, G.; Kirby, A. J.; Rittenhouse, S.; Woodnutt, G.; Cammilleri, P. Structure and antimicrobial activity of new bile acid-based gemini surfactants. *J. Chem. Soc., Perkin Trans. 2* **2002**, *2*, 1302-1306.

- (90) Shibakami, M.; Tamura, M.; Sekiya, A. Topological and Nontopological Rearrangement in Crystal Lattice of Cholic Acid Inclusion Complex Induced by Guest Exchange. *J. Am. Chem. Soc.* **1995**, *117*, 4499-4505.
- (91) Scott, J. L. Solid-Vapour Reactions of Cholic Acid and Methyl Cholate with Acetonitrile: Structures and Reaction Kinetics. *J. Chem. Soc., Perkin Trans. 2* **1995**, 495-502.
- (92) Wess, G.; Bock, K.; Kleine, H.; Kurz, M.; Guba, W. et al. The Design and Synthesis of a Scaffold for Combinatorial Chemistry Based on Bile Acid. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2222-2224.
- (93) Boyce, R.; Li, G.; Nestler, H. P.; Suenaga, T.; Still, W. C. Peptidosteroidal Receptors for Opioid Peptides. Sequence Selective Binding Using a Synthetic Receptor Library. *J. Am. Chem. Soc.* **1994**, *116*, 7955-7956.
- (94) Cheng, Y.; Toshiro Suenaga; Still, W. C. Sequence Selective Peptide Binding with a Peptido- α,β -trans-steroidal Receptor Selected from an Encoded Combinatorial Library. *J. Am. Chem. Soc.* **1996**, *118*, 1813-1814.
- (95) Cousins, G. R. L.; Furlan, R. L. E.; Ng, Y.-F.; Redman, J. E.; Sanders, J. K. M. Identification and Isolation of a Receptor for N-Methyl Alkylammonium Salts: Molecular Amplification in a Pseudo-peptide Dynamic Combinatorial Library. *Angew. Chem. Int. Ed.* **2001**, *40*, 423-428.
- (96) Pauling, L. A Theory of the Structure and Process of Formation of Antibodies. *J. Am. Chem. Soc.* **1940**, *62*, 2643-2657.
- (97) Haupt, K.; Mosbach, K. Molecularly Imprinted Polymers and Their Use in Biomimetic Sensors. *Chem. Rev.* **2000**, *100*, 2495-2504.

- (98) Zimmerman, S. C.; Lemcoff, N. G. Synthetic Hosts via Molecular Imprinting - Are universal synthetic antibodies realistically possible? *Chem. Commun.* **2004**, 5-14.
- (99) Chiang, J. Y. L. Regulation of Bile Acid Synthesis. *Frontiers in Bioscience* **1998**, 3, 176-193.
- (100) Davis, A. P.; Perry, J. J.; Wareham, R. S. Anion Recognition by Alkyl Cholates: Neutral Anionphores Closely Related to a Natural Product. *Tetrahedron Letters* **1998**, 39, 4569-4572.
- (101) Tamminen, J.; Kolehmainen, E. Bile Acid as Building Blocks of Supramolecular Hosts. *Molecules* **2001**, 6, 21-46.
- (102) Li, C.; Peters, A. S.; Meredith, E. L.; Allman, G. W.; Savage, P. B. Design and synthesis of potent sensitizers of Gram-Negative bacteria on a cholic acid scaffolding. *J. Am. Chem. Soc.* **1998**, 120, 2961-2962.
- (103) Canne, L. E.; Ferre-D'Amare, A. R.; Burley, S. K.; Kent, S. B. H. Total Chemical Synthesis of a Unique Transcription Factor-Related Protein: cMyc-Max. *J. Am. Chem. Soc.* **1995**, 117, 2998-3007.
- (104) Munari, S. D.; Cerri, A.; Gobbini, M.; Almirante, N.; Banfi, L. et al. Structure-Based Design and Synthesis of Potent Na⁺,K⁺-ATPase Inhibitors Derived from 5a,14a-Androstane Scaffold as Positive Inotropic Compounds. *J. Med. Chem* **2003**, 46, 3644-3654.
- (105) Hsieh, H.-P.; Muller, J. G.; Burrows, C. J. Synthesis and DNA Binding Properties of C3-, C12-, and C25- Substituted Amino-Steroid Derived from Bile Acids. *Bioorganic and Medicinal Chemistry* **1995**, 3, 823-838.

- (106) Davis, A. P.; Orchard, M. G. Stereocontrolled Synthesis of Cholic Acid Derivatives with N-protected 7 β and/or 12 β - amino Substituents. *Tetrahedron Letters* **1992**, *33*, 5111-5112.
- (107) Davis, A. P.; Dresen, S.; Lawless, L. J. Mitsunobu Reactions with Methanesulphonic Acid: The Replacement of Equatorial Hydroxyl Groups by Azide with Net Retention of Configuration. *Tetrahedron Letters* **1997**, *38*, 4305-4308.
- (108) Broderick, S.; Davis, A. P.; Williams, R. P. The 'Triamino-analogue' of Methyl Cholate; A Facial Amphiphile and Scaffold with Potential for Combinatorial and Molecular Recognition Chemistry. *Tetrahedron Letters* **1998**, *39*, 6083-6086.
- (109) Barry, J. F.; Davis, A. P.; Perez-Payan, M. N. A Trifunctional Steroid-Based Scaffold for Combinatorial Chemistry. *Tetrahedron Letters* **1999**, *40*, 2849-2852.
- (110) Li, C.; Rehman, A.-u.; Dalley, N. K.; Savage, P. B. Short Synthesis of Triamine Derivatives of Cholic Acid. *Tetrahedron Letters* **1999**, *40*, 1861-1864.
- (111) Kuhajda, K.; Kandrak, J.; Cirin-Novta, V.; Miljkovic, D. One-pot Esterification and Selective 3 α -Acetylation of Cholic and Deoxycholic acid. *Collect. Czech. Chem. Commun.* **1996**, *61*, 1073-1076.
- (112) Medici, A.; Pedrini, P.; Battisti, A. D.; Fantin, G.; Fogagnolo, M. et al. Anodic Electrochemical Oxidation of Cholic Acid. *Steroids* **2001**, *66*, 63-69.
- (113) Fantin, G.; Ferrarini, S.; Medici, A.; Pedrini, P.; Poli, S. Regioselective Microbial Oxidation of Bile Acids. *Tetrahedron* **1998**, *54*, 1937-1942.

- (114) Wennemers, H.; Still, W. C. Peptide Complexation in Water. Sequence-Selective Binding with Simple Dye Molecules. *Tetrahedron Letters* **1994**, *35*, 6413-6416.
- (115) Smith, G. P.; Petrenko, V. A. Phage Display. *Chemical Review* **1997**, *97*, 391-410.
- (116) Dawkins, R. *The Blind Watchmaker*; W. W. Norton: New York, 1987.
- (117) Hall, B. G. Toward an understanding of evolutionary potential. *Mircobiology Letters* **1999**, *178*, 1-6.
- (118) Collins, J.; Rottgen, P. Evolutive Phage-display - a Method as Basis for a Company; GBF Braunschweig, 1997.
- (119) Swanson, D. The Ties That Bind: Peptide Display Technology. *The Scientist* **1999**, *13*.
- (120) Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985**, *228*, 1315-1317.
- (121) Cwirla, S. E.; Peters, E. A.; Barrett, R. W.; Dower, W. J. Peptides on phage: A vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci* **1990**, *87*, 6378-6382.
- (122) Santini, C.; Brennan, D.; Mennuni, C.; Hoess, R. H.; Nicosia, A. et al. Efficient Display of an HCV cDNA Expression Library as C-terminal Fusion to the Capsid Protein D of Bacteriophage Lambda,. *Journal of Molecular Biology* **1998**, *282*, 125-135.
- (123) Sidhu, S. S. Phage display in pharmaceutical biotechnolgy. *Current Opinions in Biotechnology* **2000**, *11*, 610-616.

- (124) Azzazy, H. M. E.; W. Edward Highsmith, J. Phage display technology: clinical applications and recent innovations. *Clinical Biochemistry* **2002**, *35*, 425-445.
- (125) Stryer, L. *Biochemistry*; Third ed.; W.H. Freeman and Company: New York, 1988; 686.
- (126) Cheng, X.; Kay, B. K.; Juliano, R. L. Identification of a biologically significant DNA-binding peptide motif by use of a random phage display library. *Gene* **1996**, *171*, 1-8.
- (127) Delano, W. L.; Ultsch, M. H.; Vos, A. M. d.; Wells, J. A. Convergent solutions to binding at a protein-protein interface. *Science* **2000**, *287*, 1279-1283.
- (128) Dennis, M. S.; Eigenbrot, C.; Skelton, N. J.; Ultsch, M. H.; L.Santell et al. Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* **2000**, *404*, 465-470.
- (129) Forrer, P.; Jung, S.; Pluckthun, A. Beyond Binding: Using phage display to select for structure, folding, and enzymatic activity in proteins. *Current Opinion in Structural Biology* **1999**, *9*, 514-520.
- (130) Whaley, S. R.; English, D. S.; Hu, E. L.; Barbara, P. F.; Belcher, A. M. Selection of Peptides with Semiconductor Binding Specificity for Directed Nanocrystal Assembly. *Nature* **2000**, *405*, 665-668.
- (131) Slilaty, S. N.; Lebel, S. Accurate Insertional Inactivation of *lacZ* α . construction of pTrueBlue and M13TrueBlue cloning vectors. *Gene* **1998**, *213*, 83-91.

- (132) Rodi, D. J.; Soares, A. S.; Makowski, L. Quantitative Assessment of Peptide Sequence Diversity in M13 Combinatorial Peptide Phage Display Libraries. *J. Mol. Bio.* **2002**, *322*, 1039-1052.
- (133) Biolabs, N. E. PhD-12 Phage Display Peptide Library Kit; New England Biolabs Inc: Cambridge MA, 1998; pp 2-4.
- (134) Bridson, P. K.; Markiewicz, W. T.; Reese, C. B. Acylation of 2',3',5'-Tri-O-acylguanosine. *Chem. Commun.* **1977**, 791-792.
- (135) Xu, Y.-Z. Synthesis and Characterization of DNA Containing O⁶-Carboxymethylguanosine. *Tetrahedron* **2000**, *56*, 6075-6081.
- (136) Gaffney, B. L.; Jones, R. A. Synthesis of O⁶-Alkylated Deoxyguanosine Nucleosides. *Tetrahedron Letters* **1982**, *23*, 2253-2256.
- (137) Erlanger, B. F.; Beiser, S. B. Antibodies specific for ribonucleosides and ribonucleotides and their reaction with DNA. *Proc. N. A. S.* **1964**, *52*, 68-74.
- (138) Girault, I.; Shuker, D. E. G.; Cadet, J.; Molko, D. Use of Morpholinonucleosides to Conjugate Oxidized DNA bases to Proteins. *Bioconjugate Chem.* **1996**, *7*, 445-450.
- (139) Harrison, K. L.; Fairhurst, N.; Challis, B. C.; Shuker, D. E. G. Synthesis, Characterization, and Immunochemical Detection of O⁶-(Carboxymethyl)-2'-deoxyguanosine: A DNA Adduct Formed by Nitrosated Glycine Derivatives. *Chem. Res. Toxicol.* **1997**, *10*, 652-659.
- (140) Challis, B.; Furth, A.; Hill, R.; Hubbard, T.; Rose, S. *The Organization of Living Matter*; The Open University: Milton Keynes, 1995.
- (141) Li, C.; Peters, A. S.; Meredith, E. L.; Allman, G. W.; Savage, P. B. Design and Synthesis of Potent Sensitizers of Gram-Negative Bacteria Based on a Cholic Acid Scaffolding. *J. Am. Chem. Soc* **1998**, *120*, 2961-2962.

- (142) Davis, A. P.; Perez-Payan, M. N. The 'Triamino-analogue' of Methyl Cholate; A practical, large-scale synthesis. *Synlett* **1999**, *S1*, 991-993.
- (143) Kasal, A.; Kohout, L.; Lebl, M. Synthesis of a Scaffold for the Creation of Non-Peptide Libraries. *Collect. Czech. Chem Comm.* **1995**, *60*.
- (144) Zhou, X.-T.; Atiq-u-Rehman; Li, C.; Savage, P. B. Preparation of a Protected Triamino Analogue of Cholic Acid and Sequential Incorporation of Amino Acids in Solution and on Solid Support. *Organic Letters* **2000**, *2*, 3015-3018.
- (145) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Plenum Press: New York, 1984.
- (146) Archer, M.; Barber, J.; Hill, R.; Ho, M.-W.; Iley, J. *Bioenergetics*; The Open University: Milton Keynes, 1995.
- (147) Kuhajda, K.; Kandrak, J.; Cirin-Novta, V.; Miljkovic, D. One-Pot Esterification and Selective 3 α -Acetylation of Cholic and Deoxycholic Acid. *Collect. Czech. Chem Comm.* **1996**, *61*, 1073-1076.
- (148) Iwamura, M.; Ishikawa, T.; Sakuma, K.; Iwamura, H. 1-Pyrenylmethyl esters, photolabile protecting groups for carboxylic acids. *Tetrahedron Letters* **1987**, *28*, 679-682.
- (149) Nimura, N.; Kinoshita, T. 1-Pyrenyldiazomethane as fluorescent labeling reagent for liquid chromatographic determination of carboxylic acids. *Anal. Chem.* **1988**, *60*, 2067-2070.
- (150) Ronsin, G.; Kirby, A. J.; Rittenhouse, S.; Woodnutt, G.; Camilleri, P. Structure and Antimicrobial Activity of New Bile Acid-based Gemini Surfactants. *J. Chem. Soc., Perkin Trans.* **2002**, *2*, 1302-1306.
- (151) Lemieux, G. A.; Bertozzi, C. R. Chemoselective Ligation Reactions with Proteins, Oligosaccharides and Cells. *Trends Biotechnol.* **1998**, *16*, 506-512.

- (152) Vilaseca, L. A.; Rose, K.; Werlen, R.; Meunier, A.; Offord, R. E. et al. Protein Conjugates of Defined Structure: Synthesis and Use of a New Carrier Molecule. *Bioconjugate Chem.* **1993**, *4*, 515-520.
- (153) Ingallinella, P.; Marco, A. D.; Taliani, M.; Fattori, D.; Pessi, A. A New Method of Chemoselective Conjugation of Unprotected Peptides to Daunorubicin and Doxorubicin. *Bioorganic and Medicinal Chemistry Letters* **2001**, *11*, 1343-1346.
- (154) Lang, I.; Donze, N.; Garrouste, P.; Dumy, P.; Mutter, M. Chemoselectively Addressable HCan Building Blocks in Peptide Synthesis: L-Homocanaline Derivatives. *Journal of Peptide Science* **1998**, *4*, 72-80.
- (155) Mikola, H.; Hanninen, E. Introduction of Aliphatic Amino and Hydroxy Groups to Keto Steroids Using O-Substituted Hydroxylamines. *Bioconjugate Chem.* **1992**, *3*, 182-186.
- (156) Wahl, F.; Mutter, M. Analogues of Oxytocin with an Oxime Bridge using Chemoselectively Addressable Building Blocks. *Tetrahedron Letters* **1996**, *37*, 6861-6864.
- (157) Adamczyk, M.; Gebler, J. C.; Reddy, R. E.; Yu, Z. A Chemoselective Method for Site-Specific Immobilization of Peptides via Aminooxy Group. *Bioconjugate Chem.* **2001**, *12*, 139-142.
- (158) Tuchscherer, G. Template Assembled Synthetic Proteins: Condensation of a Multifunctional Peptide to a Topological Template via Chemoselective Ligation. *Tetrahedron Letters* **1993**, *34*, 8419-8422.
- (159) Williams, D. R.; Benlow, J. W.; Sattleberg, T. R.; Ihle, D. C. Zirconium cation coordination in the borohydride-mediated synthesis of β -hydroxy-N-alkoxylamines. *Tetrahedron Letters* **2001**, *42*, 8597-8601.

- (160) Ryser, J. E.; Jones, R. M. L.; Egeli, R.; Pelegrin, A.; Rose, K. et al. Colon Carcinoma Immunoscintigraphy by Monoclonal Anti-CEA Antibody Labeled with Gallium-67-Aminooxyacetylferrooxamine. *The Journal of Nuclear Medicine* **1992**, *33*, 1766-1773.
- (161) Canne, L. E.; Ferre-D'Amare, A. R.; Burley, S. K.; Kent, S. B. H. Total Chemical Synthesis of a Unique Transcription Factor-Related Protein: cMyc-Max. *J. Am. Chem. Soc* **1994**, *117*, 2998-3007.
- (162) Kochendoerfer, G. G.; Tack, J. M.; Cressman, S. Total Chemical Synthesis of a 27 kDa TASP Protein Derived from the MscL Ion Channel of *M. tuberculosis* by Ketoxime-Forming Ligation. *Bioconjugate Chem.* **2002**, *13*, 474-480.
- (163) Munari, S. D.; Cerri, A.; Gobbini, M.; Almirante, N.; Banfi, L. et al. Structure-Based Design and Synthesis of Potent Na⁺,K⁺-ATPase Inhibitors Derived from 5 α ,14 α -Androstane Scaffold as Positive Inotropic Compounds. *J. Med. Chem.* **2003**, *46*, 3644-3654.
- (164) Carey, F. A.; Sunberg, R. J. *Advanced Organic Chemistry. Part A: Structure and Mechanisms*; 4th ed.; Plenum Publishers: New York, 2000.
- (165) Isaacs, N. *Physical Organic Chemistry*; Longman Scientific and Technical: New York, 1995.
- (166) March, J. *Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*; McGraw-Hill Kogakusha Ltd.: Tokyo, 1977.
- (167) Jones, D. S.; Hammaker, J. R.; Tedder, M. E. A convenient synthesis of *N*-(tert-butyloxycarbonyl)aminoxy ether. *Tetrahedron Letters* **2000**, *41*, 1531-1533.

- (168) Kurth, M.; Pelegrin, A.; Rose, K.; Offord, R. E.; Pochon, S. et al. Site-Specific Conjugation of a Radioiodinated Phenethylamine Derivative to a Monoclonal Antibody Results in Increased Radioactivity Localization in Tumor. *J. Med. Chem.* **1993**, *36*, 1255-1261.
- (169) Zweig, G.; Sherma, J. Handbook of Chromatography; CRC Press: Ohio; pp 119.
- (170) Davis, A. P.; Joos, J.-B. Steroids as Organising Elements in Anion Receptors. *Coordination Chemistry Reviews* **2003**, *204*, 143-156.
- (171) Robards, K.; Haddad, P. R.; Jackson, P. E. *Principles and Practice of Modern Chromatographic Methods*; Academic Press Limited: London, 1994.
- (172) Freemantle, M. *Chemistry in action*; Macmillian Press: London, 1995.
- (173) McMurry, J. *Organic Chemistry*; Brookes/Cole Publishing Company: Pacific Grove, 1996.
- (174) Christie, W. W.; Dobson, G. Thin-layer chromatography-revisited. *Lipid Technology* **1999**, *11*, 64-66.
- (175) Harrison, K. Biomarkers of Exposure to Nitrosation Products of Amino Acids and Peptides. In *PhD Thesis*; University of Leicester, 1998; pp 64.
- (176) Batta, A. K.; Salen, G.; Shefer, J. F. B. S. Configuration at C-25 in 3 α ,7 α ,12 α -trihydroxy-5 β -cholestan-26-oic acid by X-ray Crystallography. *Journal of Lipid Research* **1979**, *20*, 935-940.
- (177) Yoswathananont, N.; Sada, K.; Miyata, M.; Akita, S.; Nakano, K. Dependence of Selective Enclathration on Types of Cholic Acid Crystals. *Org. Biomol. Chem.* **2003**, *1*, 210-214.
- (178) Lancaster Research Chemicals 2002-2003. A Clariant Group Company. United Kingdom, 2002

- (179) Kurth, M.; Pelegrin, A.; Rose, K.; Offord, R. E.; Pochon, S. et al. Site-Specific Conjugation of a Radioiodinated Phenethylamine Derivative to a Monoclonal Antibody Results in Increased Radioactivity Localization in Tumor. *J. Med. Chem* **1993**, *36*, 1255-1261.