Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Towards A Comprehensive System For The Recognition Of Double Helical DNA; Components For T And A

By

Gavin Patrick Dunne

Trinity College Dublin, 1999

A thesis submitted to the University of Dublin in fulfilment of the requirements for the degree of Doctor of Philosophy.

Declaration

This thesis has not been submitted as an exercise for a degree at this or any other University. Except where otherwise stated, the experimental work described herein has been carried out by the author alone. I agree that the library may lend or copy this thesis upon request.

Gavin Patrick Dunne

Acknowledgements

I would like to thank my supervisor Prof. Tony Davis for his guidance throughout this project. His help and tolerance is much appreciated. I am grateful to Prof. Davis and the Health Research Board for financial assistance throughout the course of this project.

I would also like to thank Maija Nissinen for acquiring the x-ray crystal structures of my heterocyclic compounds. Thank you also to Dr. Sylvia Draper for solving the x-ray structures of the squaramide compounds and for all her help with preparing the appendix.

A very big thanks to Dr. John O'Brien for running all those routine NMR spectra, including all the extra ones I didn't ask for. Thanks also for letting me use his beloved DPX-400 whenever I needed it and for asking me to do only one helium fill.

Thanks to all my colleagues in the 'back lab' and 1.7, both past and present. Thanks especially to Richard, Shay and Triona for all their help with binding/dimerisation studies and of course for sharing the secrets of HOSTEST, also to Mary-Rose for supplying me with starting material when I most needed it and for keeping the solvent cupboards full. Thanks to all the rest of the Davis group for their help, and suggestions, too numerous to detail, but greatly appreciated.

Thanks to the Head of Department, Prof. John Kelly, and all the technical staff for making it possible for me to work in the department. Thanks especially to Fred, who can source just about anything including car parts.

Finally, the greatest thanks to my family, without whose support I could not have finished this work. Thank you Isabelle for just being Isabelle and Claire for waiting and waiting still.

For Claire.

Summary

This thesis gives an account of the attempted synthesis of sequence selective double stranded DNA ligands. Two distinctive ligands types were envisaged, a pyrrolo[2',3':4,5]thieno[2,3-*d*]imidazole system (**1.40** and **1.41**) for the binding to the T-A/A-T base pairs, and an *N*-carbamoyl squaramide system (**1.44**) for binding the G-C/C-G base pairs.

Initial work focussed on the synthesis of a 2-aminothieno[2,3-*d*]imidazole (**2.9**) as an intermediate to **1.40** and **1.41**. Reduction or selective deprotection of 2-nitro-3,4-diacetamidothiophene **2.4** failed to provide a route to **2.9**. Similarly, the reduction of the 5-butylamino-3,4-dinitrothiophene **2.11** by a variety of methods failed to lead to **2.9**. The *N*-butylcyanamidothiophene **2.12** was synthesised but was found to be unstable, and so its conversion to **2.9** was not attempted.

Due to the apparent inaccessibility of a 2-aminothieno[2,3-*d*]imidazole, it was decided to pursue the pyrrolo[2',3':4,5]thieno[3,2-*d*]thiazole system **3.1**. The thieno[3,2-*d*]thiazole **3.7** was achieved by oxidative cyclisation of the 1-acetyl-3-(3-thienyl)-thiourea **3.3**. It was found to be necessary to protect the acetamido nitrogen of **3.7** by *N*-benzylation before further derivatisation and cyclisation to give **3.16**. The structure of **3.16** was confirmed by x-ray crystallography. The removal of the *N*-benzyl group of **3.16** was attempted by acid hydrolysis and dissolving metal reduction, but failed in each case.

A number of alternative protecting groups were investigated. Of these the *p*-methoxybenzyl group could be introduced, but as before the protecting group proved impossible to remove.

The synthesis of the potential G-C/C-G ligands **5.7**, **5.9**, **5.12** and **5.13** proved to be uneventful. Preliminary ^{1}H NMR studies in $CDCl_3$ indicated that **5.7** was not self-associating. ^{1}H NMR studies of the nucleoside recognition properties of **5.7** in $CDCl_3$, however, gave unexpected results. A x-ray crystal structure of the related squaramide **5.9** was obtained, demonstrating the dimerisation of **5.9**. A thorough concentration study of **5.9** in $CDCl_3$ indicated that **5.9** was dimerising strongly ($K_d \sim 10^6$ M^{-1}). Dimerisation studies of **5.7**, **5.9**, **5.12** and **5.13** were carried out in a competitive solvent mixture (5% CD_3CN in $CDCl_3$), yielding dimerisation constants in the range of 5,000-10,000 M^{-1}.

The binding of cytidine by **5.9** in mixture 5% CD_3CN in $CDCl_3$ was also investigated using ^{1}H NMR techniques. Analysis of the chemical shift data indicated 1:1 binding with an association constant of 7,000 M^{-1}.

Abbreviations

AcOH	Acetic acid
Ac_2O	Acetic anhydride
^{t}Boc	^{t}Butoxycarbonyl
$(^t\text{Boc})_2O$	Di-^{t}Butyl dicarbonate
DMF	Dimethylformamide
DMSO	Dimethylsulphoxide
DMAP	4-(*N,N*-dimethylamino)pyridine
$^i\text{Pr}_2\text{EtN}$	Diisopropylethylamine
Pd-C	Palladium on activated carbon
$Pd(OAc)_2$	Palladium acetate
$Pd(OH)_2$	Palladium hydroxide
PPh_3	Triphenylphosphine
DNA	Deoxyribonucleic acid
PNA	Peptide nucleic acid
δ	Chemical shift
Hz	Hertz
J	Coupling constant
m/z	mass charge ratio
ppm	Parts per million
DEPT	Distortionless Enhancement by Polarisation Transfer
NOE	Nuclear Overhauser effect
TOCSY	Total Correlation Spectroscopy
NMR	Nuclear Magnetic Resonance
FAB MS	Fast Atom Bombardment Mass Spectroscopy
IR	Infrared
UV	Ultraviolet
*R*f	Retention factor
TLC	Thin Layer Chromatography
bp.	Boiling point
mp.	Melting point

decomp	Decomposed
δ_{dimer}	Dimer chemical shift
δ_{Host}	Host chemical shift
$\delta_{Host\text{-}Guest}$	Host-Guest complex chemical shift
$\delta_{monomer}$	Monomer chemical shift
K_a	1:1 binding (or association) constant
K_d	Dimerisation constant

TABLE OF CONTENTS

Title page (i)
Declaration (ii)
Acknowledgements (iii)
Summary (v)
Abbreviations (vi)
Table of Contents (viii)

CHAPTER 1: INTRODUCTION

1.1 Sequence Selective Binding of Double Helical DNA 1

1.2 Sequence Specific Binding of Double Stranded DNA by Intercalation 2

1.2.1 Sequence specific binding of double stranded DNA by simple intercalation 3
1.2.2 Sequence specific binding of double helical DNA by major groove binding conjugates 4

1.3 Sequence Selective Recognition of Double Helical DNA by Minor Groove Binding Polyamides 7

1.3.1 Sequence selective minor groove binding of DNA by hairpin motif polyamides 10
1.3.2 Sequence specific minor groove binding of DNA by extended polyamides 11
1.3.3 Recognition of T-A base pairs by minor groove binding polyamides 16

1.4 Sequence Selective Recognition of Double Helical DNA by Major Groove Binding Triplex Forming Oligonucleotides 19

1.4.1 Replacement of cytosine with pH independent analogues 22
1.4.2 Recognition of the pyrimidine of T-A and C-G base pairs within the pyrimidine motif 25
1.4.3 Inhibition of quadruplex formation by G-rich TFO's in the purine motif 28
1.4.4 Recognition of the pyrimidine of T-A and C-G base pairs within the purine motif 30
1.4.5 Recognition of T-A and C-G base pairs by extended bases 31

1.5 Aims of this Project 40

CHAPTER 2: The Thieno[2,3-*d*]imidazole system

2.1 Preamble 44

2.2 The Attempted Synthesis of a 6-substituted Thieno[3,2-*b*]pyrrole System 44

2.3 The Attempts to Synthesise the 2-Amino-Thieno[2,3-*d*]imidazole system 47

2.3.1 Investigation of the reactions of 2-nitro-3,4-diacetamidothiophene **2.4** 47
2.3.2 Investigation of the reactions of 5-amino-3,4-dinitrothiophenes **2.11** and **2.12** 49

CHAPTER 3: The *N*-Benzyl Protected Pyrrolo[2',3':4,5]thieno[3,2-*d*] thiazole 3.16

3.1 Preamble 52

3.2 The Synthesis of the 2-Acetamidothieno[3,2-*d*]thiazole system 53

3.3 The Synthesis of enamine derivatives 3.10, 3.11 and 3.12 54

3.4 The Synthesis of 5-Ethoxycarbonylmethylpyrrolo[2',3':4,5]thieno[3,2-*d*] thiazole 3.16 5.9

3.4.1 Deprotection of 5-ethoxycarbonylmethyl-pyrrolothienothiazole **3.16** 61

CHAPTER 4: Alternative Protecting Groups for 3.7

4.1 Preamble 63

4.2 Allylation of 2-acetamido-6-^{t}Butylcarbamates 3.4 and 3.7 63

4.3 Alternative methods of protecting 2-acetamido-6-^{t}Butylcarbamate 3.7 64

4.4 The synthesis of *p*-Methoxybenzyl Protected Tricycle and Attempts to Remove the *p*-Methoxybenzyl Group from 4.5 and 4.7 67

CHAPTER 5: Studies of the *N*-Carbamoyl Squaramide System

5.1 Preamble 69

5.2 Initial Studies of Amino Urea 5.5 and 5.7 69

5.2.1 Initial investigation of the nucleoside recognition properties of **5.7** 71

5.3 Dimerisation studies of *N*-Carbamoyl Squaramides 75

5.3.1 Dimerisation study of *N*-Acetyl-*N*'-Butyl Squaramide **5.15** 86

5.4 Re-examination of Cytidine Binding by 5.7 and 5.9 87

CHAPTER 6: Experimental 92

CHAPTER 7: References 143

APPENDIX A: X-Ray Crystallography Tables 148

Chapter One

INTRODUCTION

1.1 Sequence Selective Binding of Double Helical DNA

It is widely appreciated that the ability to recognise and bind specific sequences of double helical DNA, either covalently or non-covalently, could lead to major advances in medicine and molecular biology. It is possible to envisage a number of ways in which a selective DNA-binding system could be used *in vivo*, e.g. the modulation of gene expression by targeting the sites of transcription factor-binding. This might be achieved by covalent modification (e.g. chemical alkylation[1.1]) of specific DNA sequences. Alternatively, a non-covalently bound system could interfere competitively with DNA-binding proteins. Another possible use of a DNA recognition system would be to direct destructive agents to cells containing specific DNA sequences. The same system *in vitro* could be used in 'synthetic endonucleases' with designed specificities[1.2]. Such molecules would be of great value for molecular biology, e.g. in cloning and genetic analysis.

In order for any of these systems to be effective, a ligand must first be capable of recognising a single sequence in a sample of double helical DNA, and then binding efficiently to that sequence. In the case of the human genome, which contains 3×10^9 base pairs, it has been estimated that a minimum of 17 base pairs must be recognised in order to select a unique base sequence.[1.3] This is based on a statistical distribution of base pairs in the genome and an equal number of adenine-thymine (A-T) and guanine-cytosine (G-C) base pairs.

The structure of B-DNA, the major biologically active form of DNA, is such that the number of ways in which it can be sequence specifically bound is severally limited. B-DNA twists in the form of a right-handed, double stranded helix, with the sequence determining bases buried within its core. If a ligand is to sequence selectively bind to double helical DNA, it must interact with the bases. Access to the bases however, can be had only through either the major or minor grooves, which are on the external surface of the helix (see **Figure 1.1**). This means that specific binding can be achieved only *via* one or more of three motifs; binding within the major groove, binding in the minor groove or intercalation between the base pairs (obviously the intercalator must pass through either the major or minor groove first).

Each of these binding motifs differ in their mode of operation, but all rely on one or more of a variety of co-operative forces, including hydrogen bonding, van der Waals,

electrostatic, hydrophobic and π-aromatic interactions. Steric and thermodynamic factors also play a role in these processes.

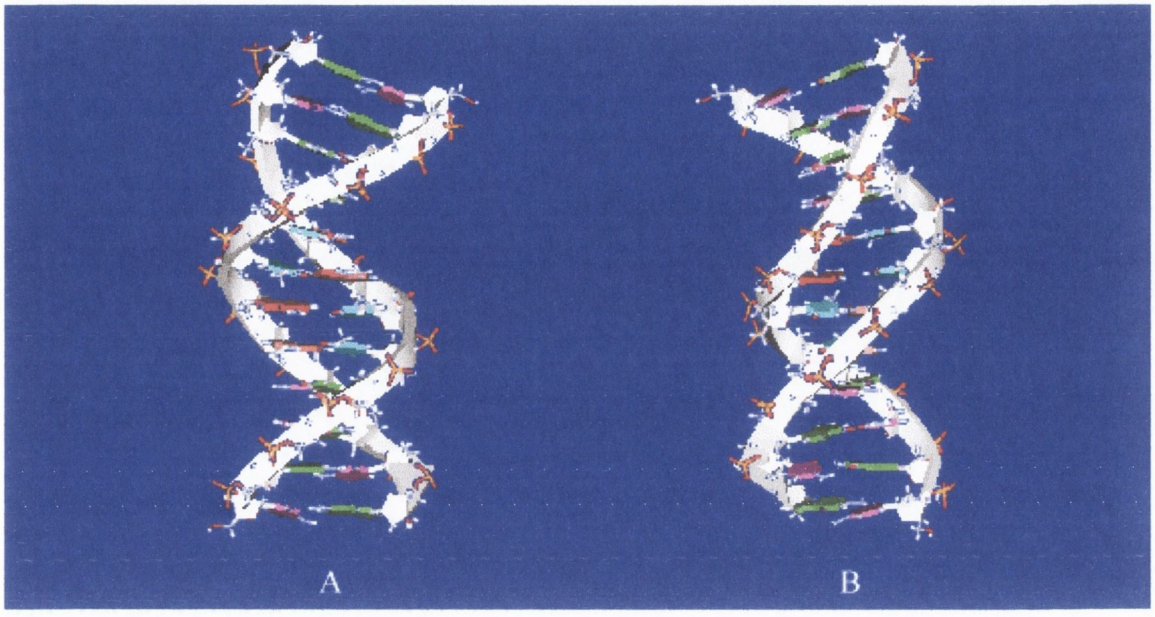


Figure 1.1 B-DNA showing major and minor grooves. Schematics of B-DNA dodecamer CGCGTTAACGCG clearly show the major (A) and minor (B) grooves. (Reprinted from Protein Data Bank)

Attempts have been made, with varying degrees of success, to produce sequence selective, non-covalent binding of double stranded DNA, using each of the above binding motifs and sometimes incorporating more than one motif in the same system. The following pages are a short review of some of the more important and most recent contributions to this area. This review is divided into three sections, each dealing with one of the binding modes mentioned earlier. This is then followed by an outline of the aims of my project: *Towards a comprehensive system for the recognition of double-helical DNA; components for T and A.*

1.2 Sequence Specific Binding of Double Stranded DNA by Intercalation

DNA-intercalators, as a class, are remarkably diverse in structure, but all possess the common structural feature of a planar aromatic ring system. It is this chromophore which inserts (intercalates) between adjacent base pairs. The helix in the vicinity of the insertion site partially unwinds and the base pairs are displaced along the helical axis in order to accommodate the intercalator. The end result is a tightly, yet reversibly bound complex,

stabilised primarily by hydrophobic and π-π interactions between the aromatic rings of the intercalator and the heterocyclic bases.

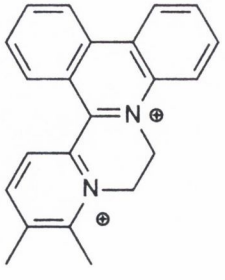
There is little scope for a system binding DNA solely by intercalation, to do so in a sequence selective manner. This is due to the non-specific nature of the forces involved. A small degree of selectivity may be acquired by targeting sites which are particularly easily unwound or are highly polarizable.[1.4] In general, however, intercalation is used as a strong, non-specific form of binding and the selectivity of an intercalating system is derived from the interaction of substituents which it bears with the duplex. These interactions may include hydrogen bonding, electrostatic forces, van der Waals contacts and steric hindrance. The following short review [1.2.1-1.2.2] is a summary of some of the most recent and more important publications in this area.

1.2.1 *Sequence specific binding of double stranded DNA by simple intercalation*

Contribution of Orellana and co-workers

The phenanthridine dicationic derivative **1.1** was prepared and binding studies with polynucleotides carried out by visible absorption titration studies at 25 °C.[1.5] From this, association constants were calculated (**Table 1.1**). The highest binding constant was for poly[d(GC)]$_2$, indicating that **1.1** has a preference for alternating G-C base pairs. Binding to poly[d(AT)]$_2$ was ~25% as strong as that to poly[d(GC)]$_2$, while binding to calf thymus DNA, with 42% G-C base pairs, was in an intermediate position. The homopolynucleotides afforded the weakest binding constants.

The preference of **1.1** for binding to alternating G-C base pairs was ascribed to the higher polarity of the G-C pair, which through interacting with the polarizable **1.1**, is able to stabilise the complex.



1.1

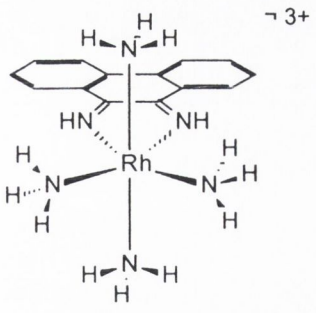
Table 1.1 Binding constants for interaction of **1.1** with calf thymus DNA and polynucleotides in pH 5.5 sodium phosphate buffer at 25 °C

Polynucleotide	K_a (M^{-1})
Poly[d(GC)]$_2$	2.42×10^5
Calf Thymus DNA	1.90×10^5
Poly[d(AT)]$_2$	6.3×10^4
Poly(dG)·poly(dC)	2.6×10^4
Poly(dA)·poly(dT)	1.8×10^4

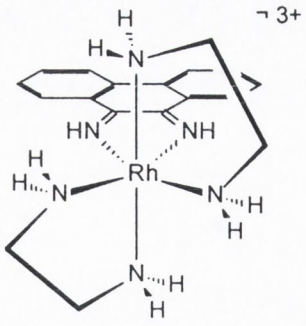
1.2.2 Sequence specific binding of double helical DNA by intercalator-major groove binding conjugates

Contribution of Barton and co-workers

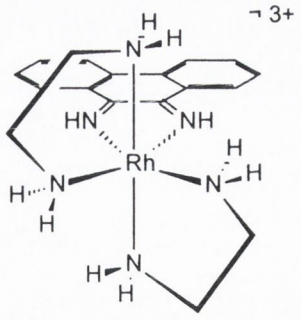
It has been shown by DNA-binding and photocleavage studies, that rhodium (III) 9,10-phenanthrenequinone diimine polyamine complexes **1.2** and **1.3** bind strongly in the major groove of DNA, *via* intercalation of the aromatic system, with a marked preference for 5′-GC-3′ steps ($K_a \geq 10^6$ M^{-1}).[1.6] This 5′-GC-3′ preference is attributed to hydrogen bond interactions between the axial amine ligands and the guanine C-6 carbonyl group. Cleavage studies of DNA with **1.4**, the enantiomer of **1.3**, showed poorer sequence selectivity, but enantioselective cleavage of 5′-TX-3′ (X = A, T, C or G) was observed.[1.6] Photocleavage of an oligonucleotide containing the substitution of 5′-UA-3′ for 5′-TA-3′ showed no enantioselectivity. The enantioselectivity was thus accredited to chiral recognition of the 5′-TA-3′ step by van der Waals contacts between the methylene groups of **1.4** and the thymine methyl groups displayed in the major groove.



1.2

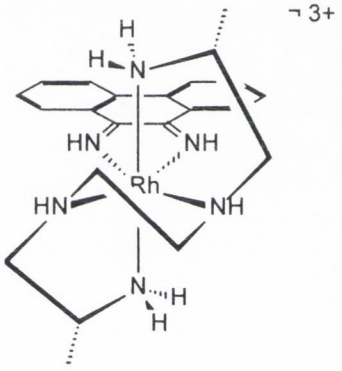


1.3

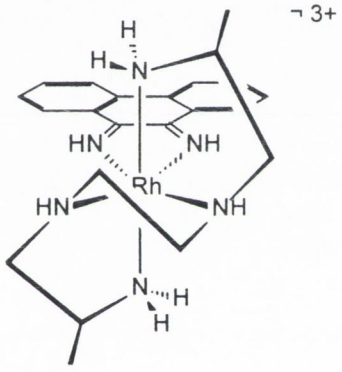


1.4

Building on these findings, Barton *et al.* designed and synthesised the metallointercalator **1.5-(R,R)** and **1.5-(S,S)**. It was thought that site recognition might be extended to 5′-TGCA-3′ by placement of methyl groups adjacent to the axially co-ordinated amines on **1.3,** in order to make specific van der Waals contacts with the thymine methyl groups in the major groove (see **Figure 1.2**).



1.5-(R,R)



1.5-(S,S)

The site selectivity of **1.5-(S,S)** and **1.5-(R,R)** was investigated by DNA photocleavage studies.[1.7] **1.5-(S,S)** was found to show comparable selectivity as **1.3**, but **1.5-(R,R)** exhibited increased cleavage at 5′-TGCX-3′ steps and significantly so at a 5′-TGCA-3′ site. Quantification showed cleavage was almost doubled at this site compared to the neighbouring 5′-GGCA-3′ site. These results are in agreement with that predicted by the molecular model, the structural details of which is also supported by high resolution ^{1}H NMR studies.[1.8] The formation of a single bound complex was observable in NMR titration studies. Two-dimensional NOSEY data was consistent with the model and identified intercalation at 5′-GC-3′. Intermolecular NOEs between the thymine methyl and **1.5-(R,R)** methyl groups was also detected.

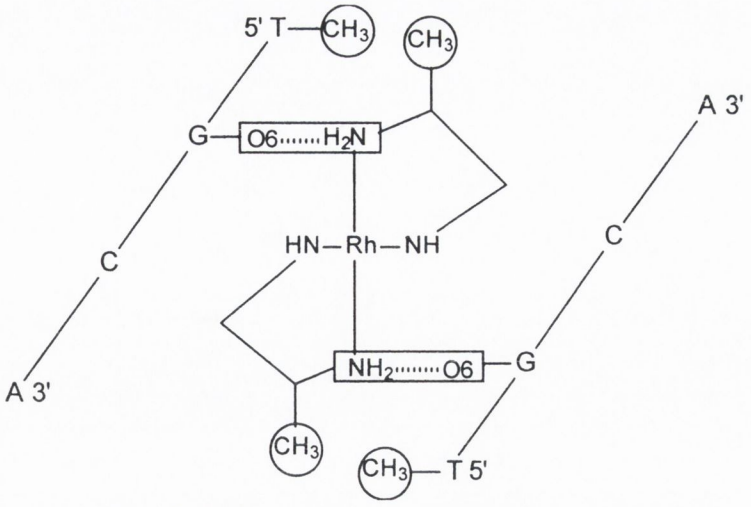
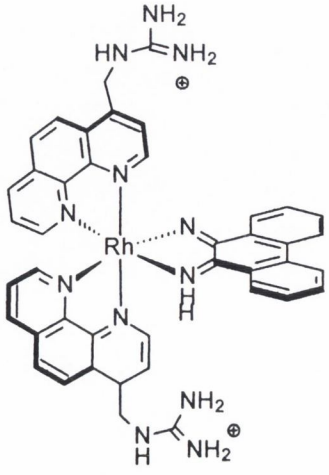


Figure 1.2 Schematic of **1.5-(R,R)** bound to 5′-TGCA-3′ showing proposed hydrogen bonds and van der Waals contacts

Another family of phenanthrenequinone based metallointercalator systems, represented by **1.6**, also shows sequence selectivity on binding DNA. [1.9, 1.10] Photocleavage studies showed that **1.6** selects the 5′-CATATG-3′ site on binding.[1.9] Cleavage, DNA-cyclisation and NMR[1.11] studies suggest that this sequence selectivity is due to the chirality of **1.6**. **1.6** incorporates two 4-(guanidylmethyl)-1,10-phenanthroline ligands, arranged in a left-handed helix. On intercalation of the phenanthrenequinone within the duplex, the guanidyl groups are directed into the major groove, where they can form hydrogen bonds with the N-7 and C-6 carbonyl of guanine at the first and sixth base pair positions. It is necessary however for the duplex to unwind by 70 ° at the binding site, in order for these guanidyl-guanine contacts to occur. This requirement naturally selects the central AT-rich stretch, as such stretches are known to be quite flexible and so unwind without a large energetic penalty.



1.6

1.3 Sequence Selective Recognition of Double Helical DNA by Minor Groove Binding Polyamides

The minor groove of B-DNA is structurally well defined and imposes a number of criteria which can be used to control binding. The minor groove is a narrow, deep channel between the two deoxyribose-phosphate backbones. The floor of this channel is made up of A-T and G-C base pairs, orientated edge on so that hydrogen bond donor and acceptor groups are displayed.

If a system is to bind sequence selectively in the minor groove, it must satisfy the conditions that the structure of the groove dictates. Firstly, a binder must be a complementary size and shape, it must be narrow enough so as to fit between the two sugar-phosphate spines, but also capable of filling the groove, thus stabilising the complex by van der Waals and hydrophobic interactions with the groove walls. It is also necessary that the ligand either be flexible, or else be preorganised in a concave shape, so that it can follow the line of the groove. Secondly, it should be either neutral or cationic in order to avoid electrostatic repulsion with the negatively charged spines. Thirdly, it must be capable of inserting deep enough within the minor groove so as to interact with the base pairs. Finally, the binder must be able to recognise and discriminate between the structural information presented by the four canonical base pairs, G-C, C-G, A-T and T-A.

The G-C base pair presents a hydrogen bond donor/acceptor pattern (N-3 and C-2 amine of guanine and C-2 carbonyl of cytosine) which is different to the A-T/T-A base pair, and due to the guanine exocyclic amine being off-centre, is non-superimposable on that of the C-G base pair (see **Figure 1.3**). The amine of guanine also protrudes up from the floor of the minor groove, allowing for further distinction between the G-C/C-G and A-T/T-A base pairs on steric grounds. The A-T/T-A base pairs display two hydrogen bond acceptors only (N-3 of adenine and C-2 carbonyl of thymine), which are almost identically placed in the minor groove. This makes the discrimination between A-T and T-A base pairs very challenging.

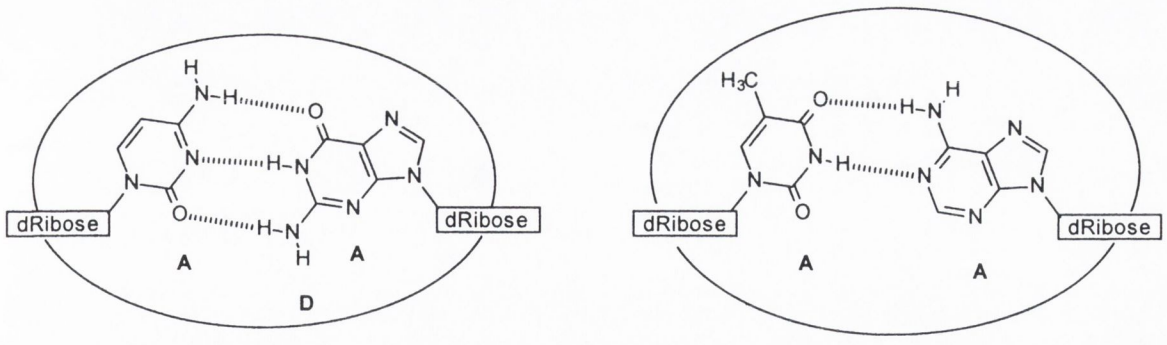


Figure 1.3 C-G (left) and T-A (right) base pairs. Free hydrogen bond donor (**D**) and acceptor (**A**) sites in the minor groove of double-stranded DNA.

There are many examples of minor groove binders, both naturally occurring and man-made (e.g. netropsin, distamycin and hoechst 33258, see **Figure 1.4a-c**). All of these however show a marked preference for AT sequences.[1.12] This is attributed to complementary hydrogen bonding (interaction of the hydrogen bond donors on the ligands with the N-3 of adenine and C-2 carbonyl of thymine) and, perhaps more importantly, van der Waals, hydrophobic and steric interactions. The exocyclic amino group of guanine, which protrudes from the floor of the minor groove, might sterically hinder a ligand from binding deeply in the groove of a sequence containing a G-C/C-G base pair. In addition, the minor groove of an A/T tract is narrower than that of a GC segment, so a ligand such as netropsin fits more tightly within an A/T sequence and can make a full complement of non-covalent interactions with the groove. This idea was supported by the attempts to extend recognition to include a G-C base pair by substituting one of the pyrrole rings of netropsin for an imidazole. The netropsin analogue was expected to recognise G-C by the formation of a specific hydrogen bond from the N-3 of imidazole to the exocyclic amine of guanine. The result however was a ligand which while being more tolerant of a G-C base pair, now suffered an overall decrease in binding affinity and specificity.[1.13]

The breakthrough in netropsin analogues with expanded recognition codes came from Dervan and co-workers. 1-Methyl-imidazole-2-carboxamide-netropsin (**Figure 1.4d**) was found to recognise a 5′-TGACT-3′ tract with high specificity.[1.14] Structural studies indicated that the netropsin analogue formed an anti-parallel dimer, which bound to DNA in the minor groove (see **Figure 1.5**). The adoption of a dimeric form means that the ligand is now large enough to fill the wider G-C/C-G minor groove. The G-C/C-G base pair recognition is attributed to the formation of a hydrogen bond from the N-3 of

imidazole to the exocyclic amine of guanine. It appears that the neighbouring pyrrole packs the imidazole to one side of the groove, resulting in the correct positioning of the imidazole. The NH of the polyamide participate in hydrogen bonding with the acceptor groups on C, A and T.

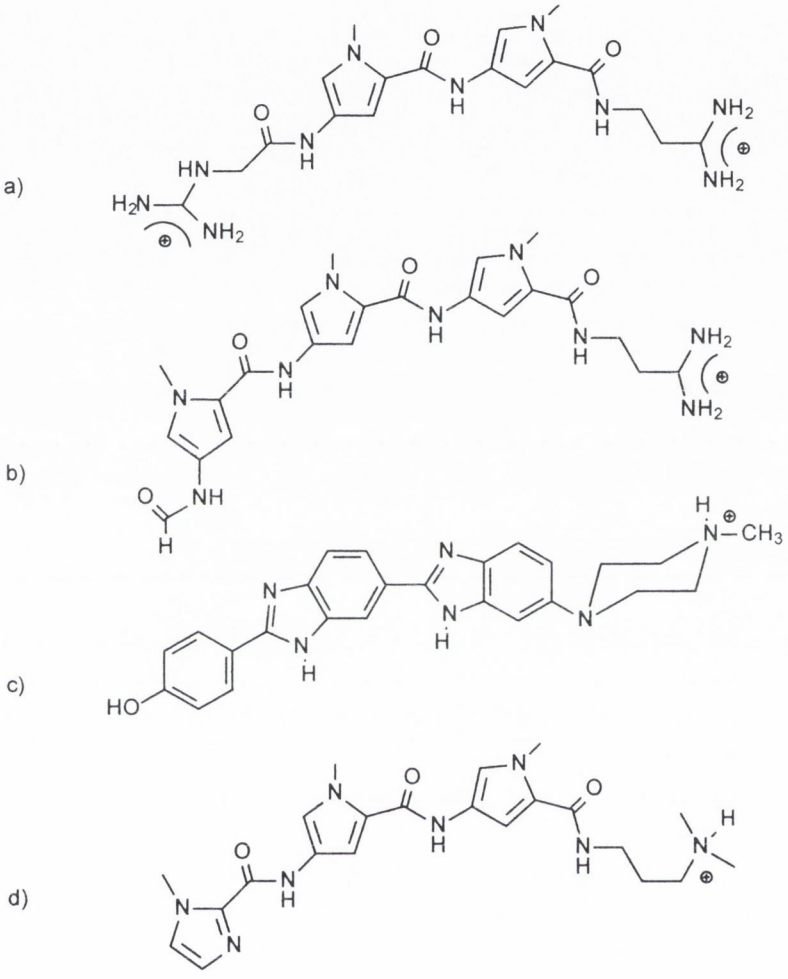


Figure 1.4 Structure of natural and synthetic DNA minor groove-binders. a) Netropsin, b) distamycin, c) hoechst 33258, d) 1-methyl-imidazole-2-carboxamide-netropsin.

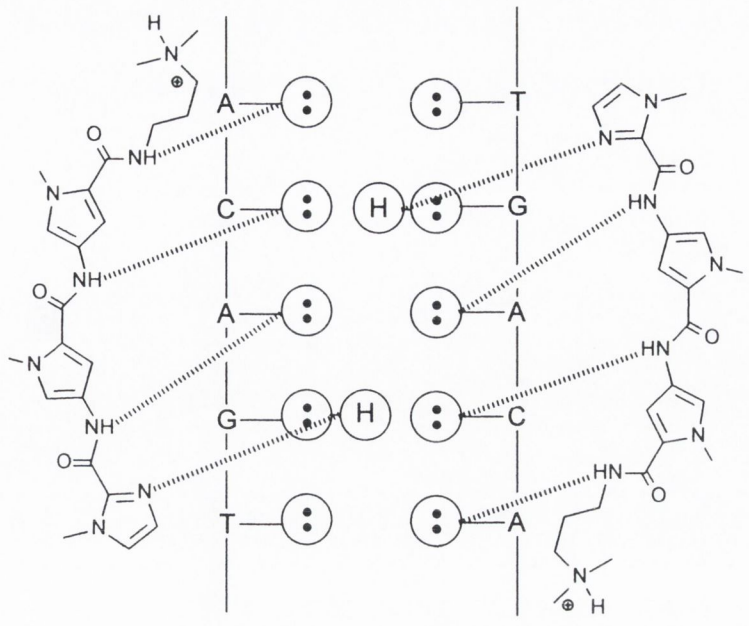


Figure 1.5 Schematic of sequence-specific binding of 1-methyl-imidazole-2-carboxamide-netropsin in the minor groove of a DNA double helix. Circles with dots represent lone pairs on N-3 of purines and C-2 carbonyl of pyrimidines. Circles with an H represent the exocyclic amine hydrogen of guanine.

This result suggested that it is possible to formulate pairing rules for base pair recognition using two polyamide units, one from each molecule of the dimer. An imidazole ring paired with a pyrrole ring recognises a G-C base pair, while a pyrrole/imidazole pairing recognises a C-G base pair. The A-T and T-A base pairs are degenerate and both are recognised by a pyrrole/pyrrole pair. Since the development of these pairing rules, extensive work has been carried out, mostly by Dervan, to extend the recognition beyond five base pairs to larger DNA sequences. The following pages are a short review [1.3.1-1.3.3] of some of the major advancements made in this direction.

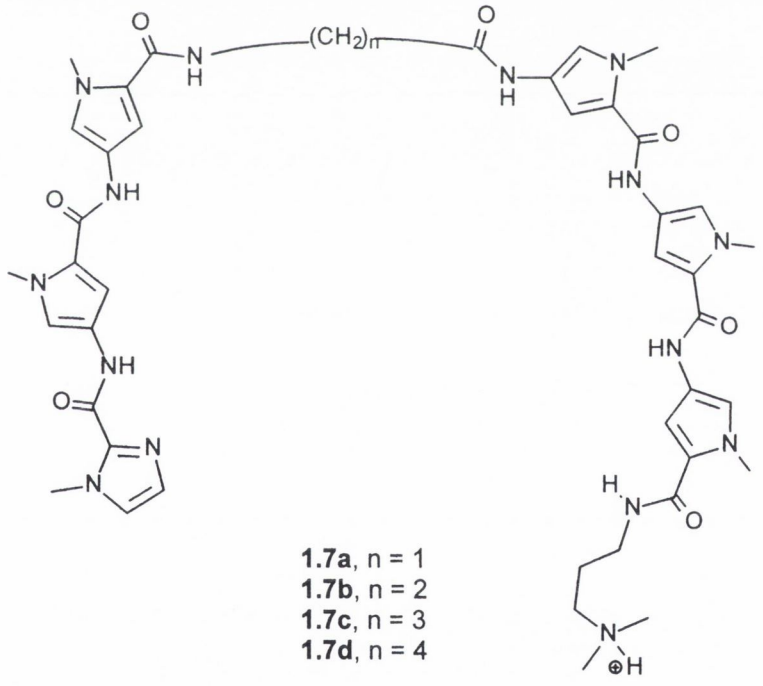
1.3.1 Sequence selective minor groove binding of DNA by hairpin motif polyamides

Contribution of Dervan and co-workers

While it was possible to bind the DNA sequence 5′-TGACT-3′ by formation of a homodimer from a single polyamide, it was recognised that other sequences would necessitate the use of heterodimer. However, when two different peptides are mixed to form a heterodimer, there is the extra complexity due to the formation of the two possible homodimers. Thus, there would be three dimers, each binding a different DNA sequence. In order to avoid this complication, Dervan and co-workers designed a hairpin motif whereby two peptides are covalently linked head to tail by use of a flexible tether so that the molecule can fold, bringing the peptides side by side in an anti-parallel geometry.

The series of hexapeptides **1.7a-d** were synthesised and the binding affinities to three sites, 5′-TTTTT-3′, 5′-TGTTA-3′ and 5′-TGACA-3′, on a 135 base pair DNA fragment were determined by quantitative DNase I footprint titration experiments.[1.15] The γ-aminobutyric acid (γ) linked peptide **1.7c**, binds the target sequence, 5′-TGTTA-3′, with the greatest affinity (~10^8 M^{-1}) indicating that it adopts the hairpin motif as designed. Importantly the use of the linker was found to improve sequence specificity, as the single mismatch site, 5′-TGACA-3′, was bound with 24 fold lower affinity. The increase in affinity of the linked peptides over the uncoupled peptides was estimated to be at least two orders of magnitude.

Studies on the effect of varying the location of γ (on the pyrrole or imidazole ends of the peptide) indicated that the position of γ did not dramatically affect the binding affinity or specificity of the polyamides.[1.16]



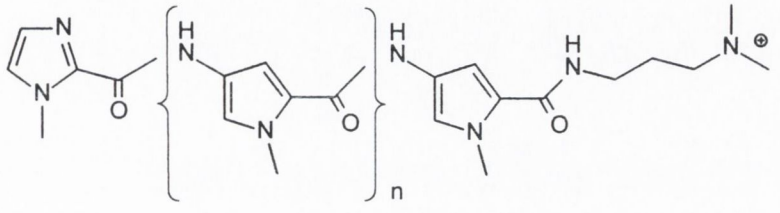
1.7a, n = 1
1.7b, n = 2
1.7c, n = 3
1.7d, n = 4

1.3.2 Sequence specific minor groove binding of DNA with extended polyamides

Contribution of Dervan and co-workers

The series of polyamides **1.8a-f** were synthesised, and the binding affinities and specificities to target DNA sequences were determined using quantitative DNase I footprint titration experiments.[1.17] The binding affinity was found to increase as the number of heterocyclic rings increased, peaking at six rings. The affinity decreased

slightly on going to seven rings and then dropped dramatically for the eight ring polyamide (see **Table 1.2**). The binding specificity, defined as the ratio of match site binding affinity to the affinity for the single base pair mismatch site, remains approximately stable (~6) for the polyamides up to a total of five rings and then decreases dramatically as polyamide length increases (see **Table 1.2**).



1.8a, n = 1
1.8b, n = 2
1.8c, n = 3
1.8d, n = 4
1.8e, n = 5
1.8f, n = 6

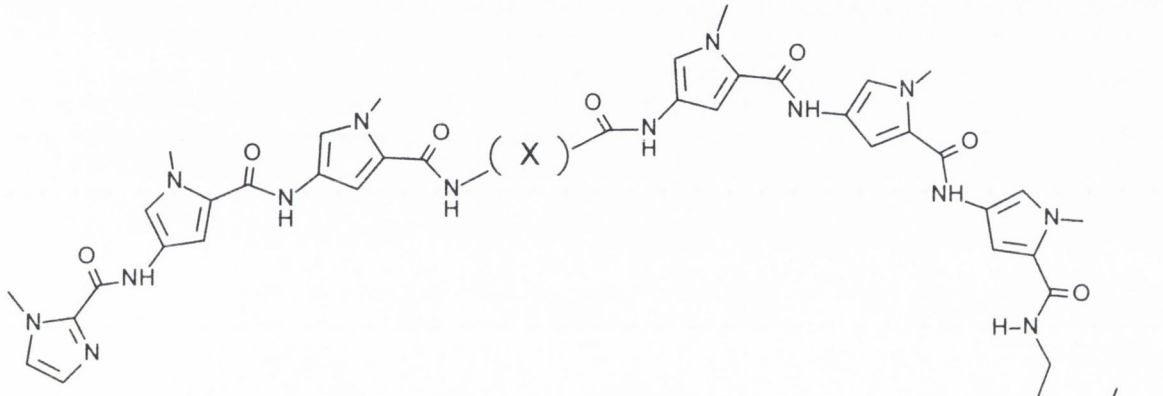
Table 1.2 Binding constants and specificity for interaction of polyamides **1.8a-f** with DNA

Polyamide	Number of rings	Binding site size, bp	K_a (M^{-1}) Match site	K_a (M^{-1}) Mismatch site	Binding specificity
1.8a	3	5	1.3×10^5	$<2 \times 10^4$	>6.5
1.8b	4	6	8.5×10^6	1.6×10^6	5.3
1.8c	5	7	4.5×10^7	7.9×10^6	5.7
1.8d	6	8	5.3×10^7	$<2 \times 10^7$	>2.7
1.8e	7	9	4.7×10^7	1.7×10^7	2.8
1.8f	8	10	$<2 \times 10^6$	$<2 \times 10^6$	≈1

It was considered that the observed drop in affinity was due to the longer ligands falling out of register with the DNA base pairs. This has since been confirmed by x-ray crystallography which indicated that the curvature of the ligands does not perfectly match that of DNA, the ligand is over-wound relative to the pitch of the DNA helix.[1.18] It was evident that a new class of polyamides was necessary in order to extend recognition beyond nine base pairs. It was thought that replacement of a central pyrrole or imidazole amino acid with a more flexible amino acid subunit would allow the dimer to reset the register with DNA and so continue to bind efficiently.

With the above results in mind, a series of polyamides, **1.9a-c** were synthesised. The polyamides are essentially two three-ring polyamides (imidazole-pyrrole-pyrrole and

pyrrole-pyrrole-pyrrole), linked by different amino acid subunits, 4-amino-2-carboxypyrrole, glycine and β-alanine. The binding affinity to the nine base pair site 5′-TGTTAAACA-3′ in a 281 base pair DNA fragment was determined for each polyamide.[1.19] **1.9a** and **1.9b** bound the target site with similar affinities(1 x 10^8 M^{-1}), while the β-alanine linked peptide **1.9c** bound the same site with approximately 8 fold higher affinity, indicating that β-alanine is the optimal linker for forming extended polyamide. Extensive NMR studies by Wemmer and co-workers, characterising these polyamide-DNA complexes also supported these conclusions.[1.20]



1.9a, X = pyrrole
1.9b, X = CH_2
1.9c, X = $(CH_2)_2$

Besides binding to the nine base pair site, each of the polyamides **1.9a-c** also displayed specificity for two thirteen-base pair sites, 5′-AAAAAGACAAAAA-3′ and 5′-ATATAGACATATA-3′. The nine base pair site was bound *via* the usual 'overlapped' 2:1 binding mode (as in **Figure 1.5**), while the thirteen base pair sites were presumably recognised by a 'slipped' binding mode. In this second binding mode the imidazole-pyrrole-pyrrole moieties of two polyamides bind the central 5′-AGACA-3′ tract in a 2:1 manner and the pyrrole-pyrrole-pyrrole moieties of the same polyamides bind the A/T flanking sequences in a 1:1 manner similar to netropsin-DNA complexes (see **Figure 1.6**).

In the nine base pair 'overlapped' site and the thirteen base pair 'slipped' site, the G-C and C-G base pairs are separated by five and one A/T base pairs, respectively. In theory, the polyamides could also bind to 'partially slipped' sites of ten, eleven and twelve base pairs, where the G-C and C-G base pairs are separated by four, three and two

A/T base pairs respectively. This multiplicity of binding modes would obviously lead to a reduction in the specificity of these systems.

As a means of avoiding these 'slipped' binding modes, the asymmetric γ-linked hairpin polyamides **1.10** and **1.11** were synthesised. Due to the self-complementary nature of the pyrrole-imidazole overhang on these peptides, two molecules can associate, to form a co-operative hairpin dimer with no risk of 'slipped' binding occurring (see **Figure 1.7**). **1.10** bound the target DNA sequence 5′-AGCAGCTGCT-3′, with an association constant of 1.9 x 10^8 M^{-1}, and the single base pair mismatch site 5′-AGATGCTGCA-3′ with 9-fold lower affinity.[1.21] **1.11** bound the target 12 base pair 5′-AAGCAGCTGCTT-3′ with 10-fold higher affinity than **1.10**.

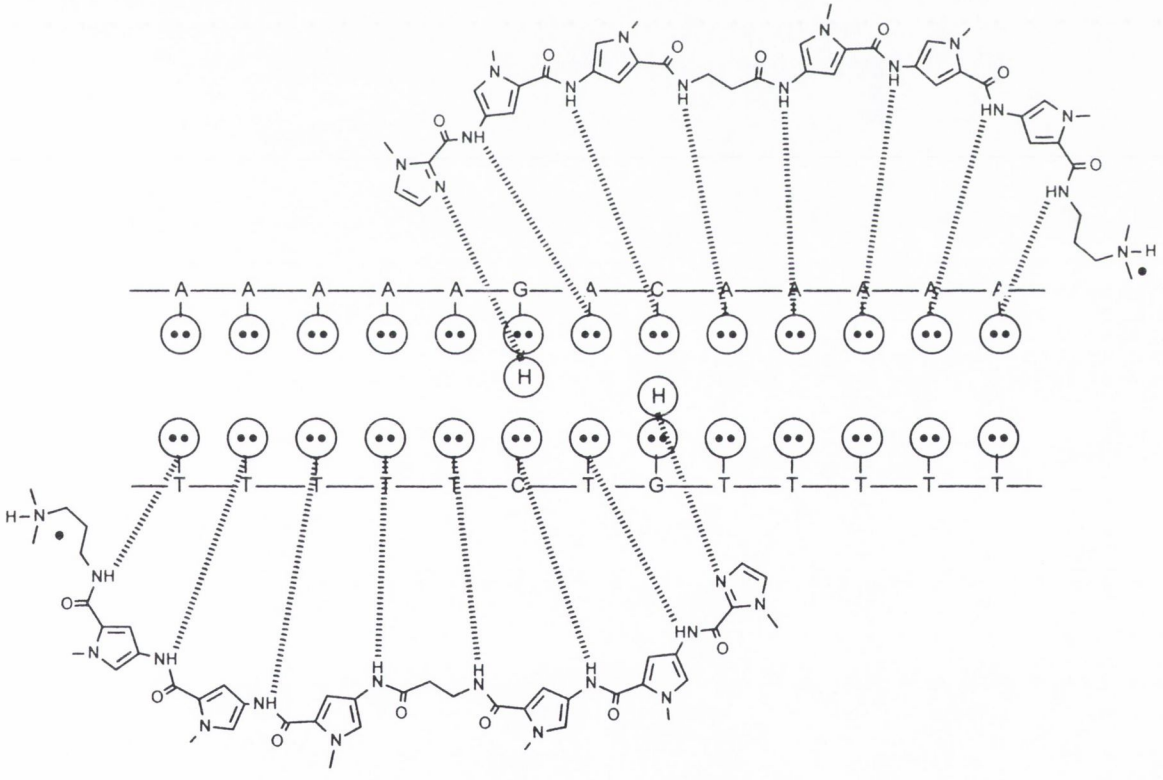
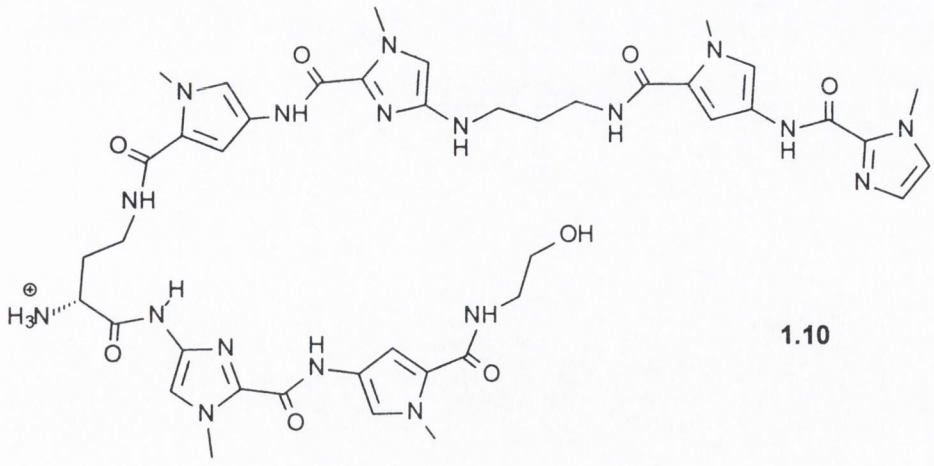
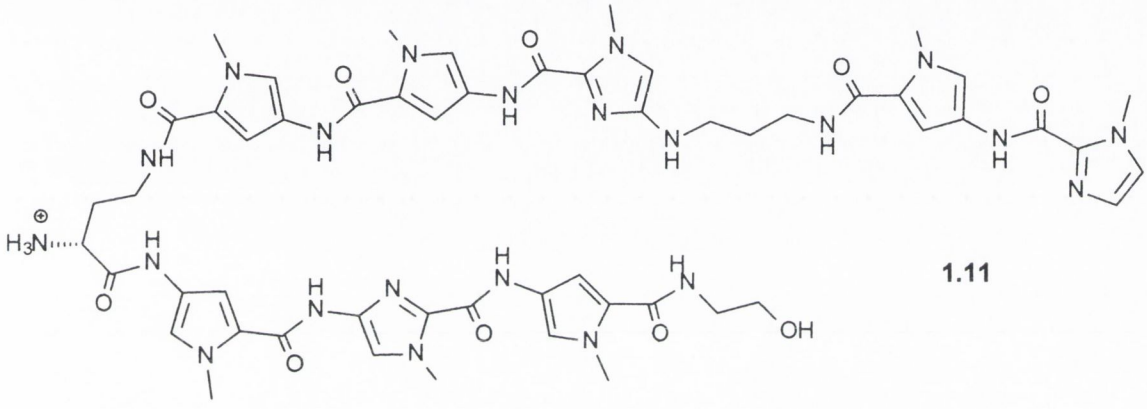


Figure 1.6 Schematic of 'slipped' mode binding of **1.9c** in the minor groove of a DNA double helix. Circles with dots represent lone pairs on N-3 of purines and C-2 carbonyl of pyrimidines. Circles with an H represent the exocyclic amine hydrogen of guanine.



1.10



1.11

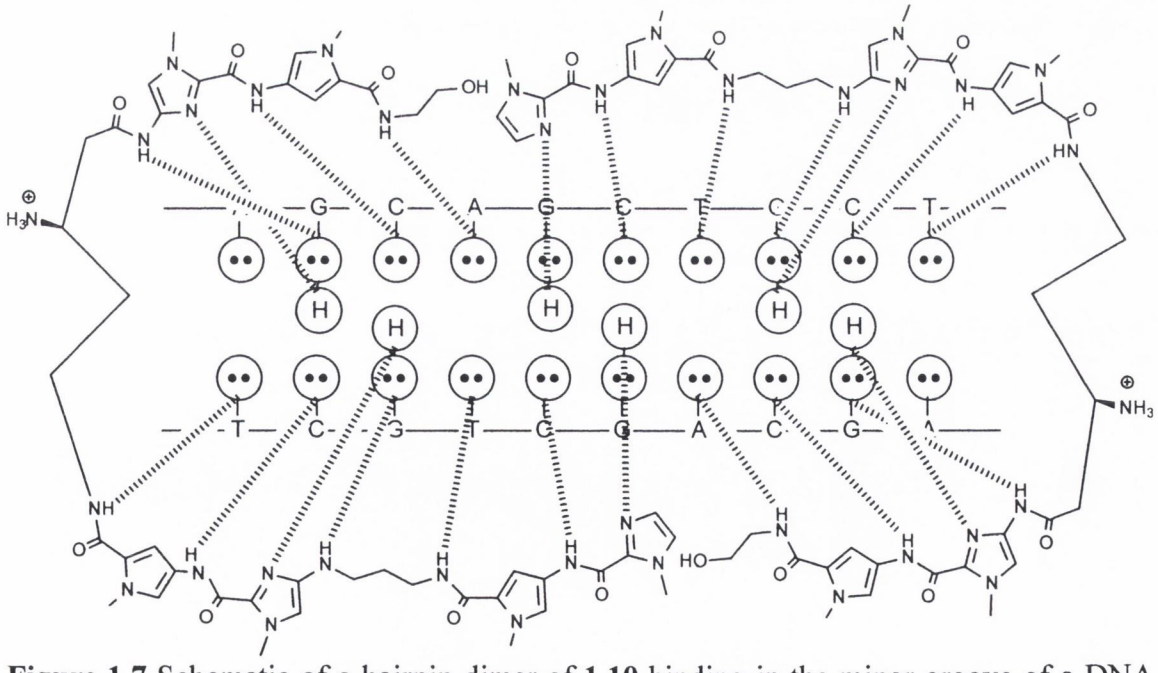
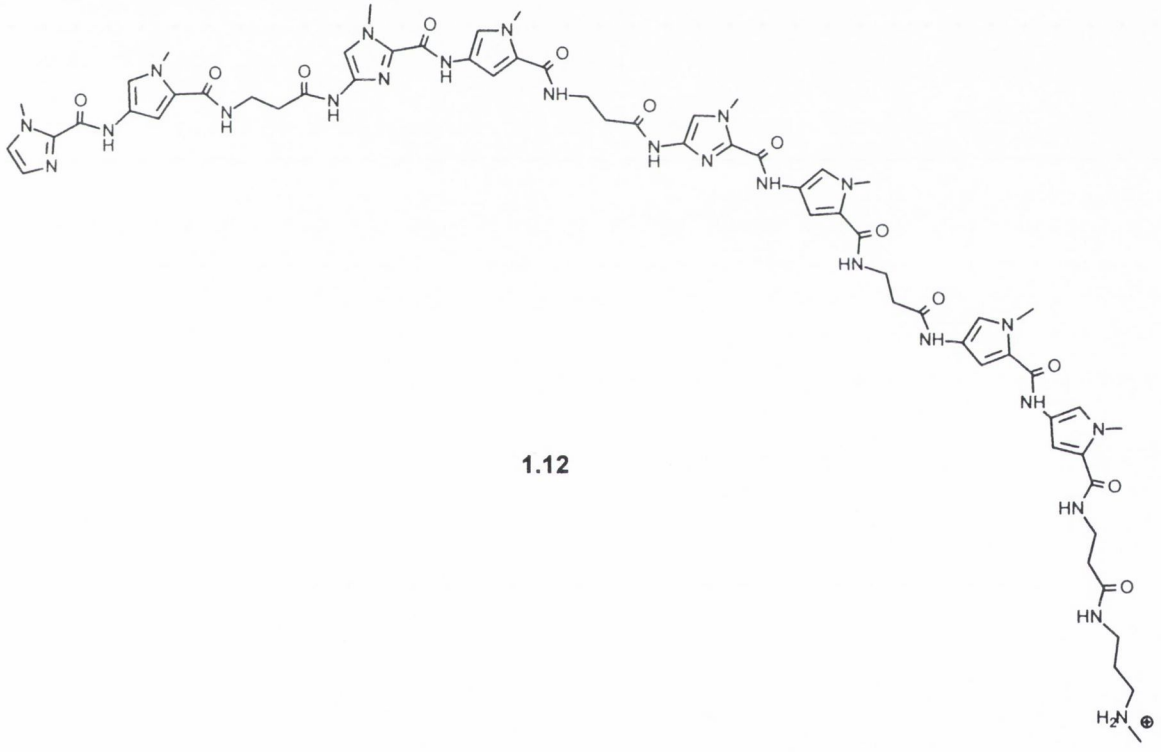


Figure 1.7 Schematic of a hairpin dimer of **1.10** binding in the minor groove of a DNA double helix. Circles with dots represent lone pairs on N-3 of purines and C-2 carbonyl of pyrimidines. Circles with an H represent the exocyclic amine hydrogen of guanine.

The efforts to design extended DNA-sequence specific polyamides has to date culminated in the recognition of a sixteen base pair segment. This represents a significant milestone in DNA recognition, as a minimum of 16-17 base pairs are necessary to specify any site within the human genome. The eight ring polyamide **1.12** binds the target sequence 5′-ATAAGCAGCTGCTTTT-3′ in a 245 base pair DNA fragment with an association constant of >3.5 x 10^{10} M^{-1}.[1.22] Specificity was not very high however as a number of base pair mismatch sequences were bound with 10-20-fold lower affinity relative to the match site. The polyamide recognised the target sequence *via* the 'slipped' dimer binding mode; the central 5′-GCAGCTGC-3′ was bound by the six imidazole/pyrrole pairs and two β/β pairs in a 2:1 manner, while the flanking A/T base pairs were bound by the remaining two β-alanine and two pyrrole in a 1:1 manner.



1.12

1.3.3 Recognition of T-A base pairs minor groove binding polyamides

Contribution of Dervan and co-workers

The degenerate recognition of the A-T/T-A base pairs by pyrrole/pyrrole pairings, represents a significant limitation in the usefulness of polyamides for sequence selective binding of DNA. To break this degeneracy, Dervan designed a polyamide incorporating a 3-hydroxypyrrole unit. It was reasoned that while the hydrogen bond acceptors of

adenine and thymine may be almost identically placed within the minor groove (making discrimination between the two difficult), the asymmetrically placed H-2 of adenine would allow a shape selective mechanism for A-T/T-A recognition (see **Figure 1.8**). This selectivity could arise from destabilisation of polyamide binding by steric clash on placing 3-hydroxypyrrole opposite adenine, or stabilisation of binding due to hydrogen bonding between the C-2 carbonyl of thymine and the C-3 OH of hydroxypyrrole.

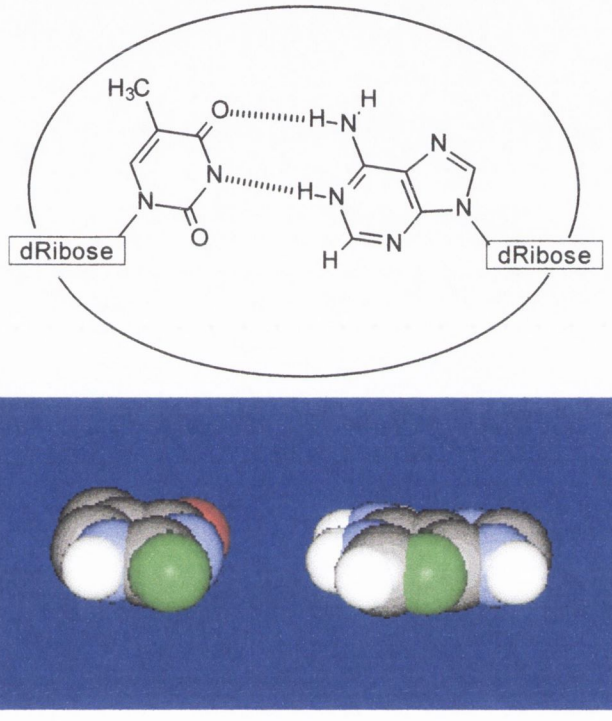
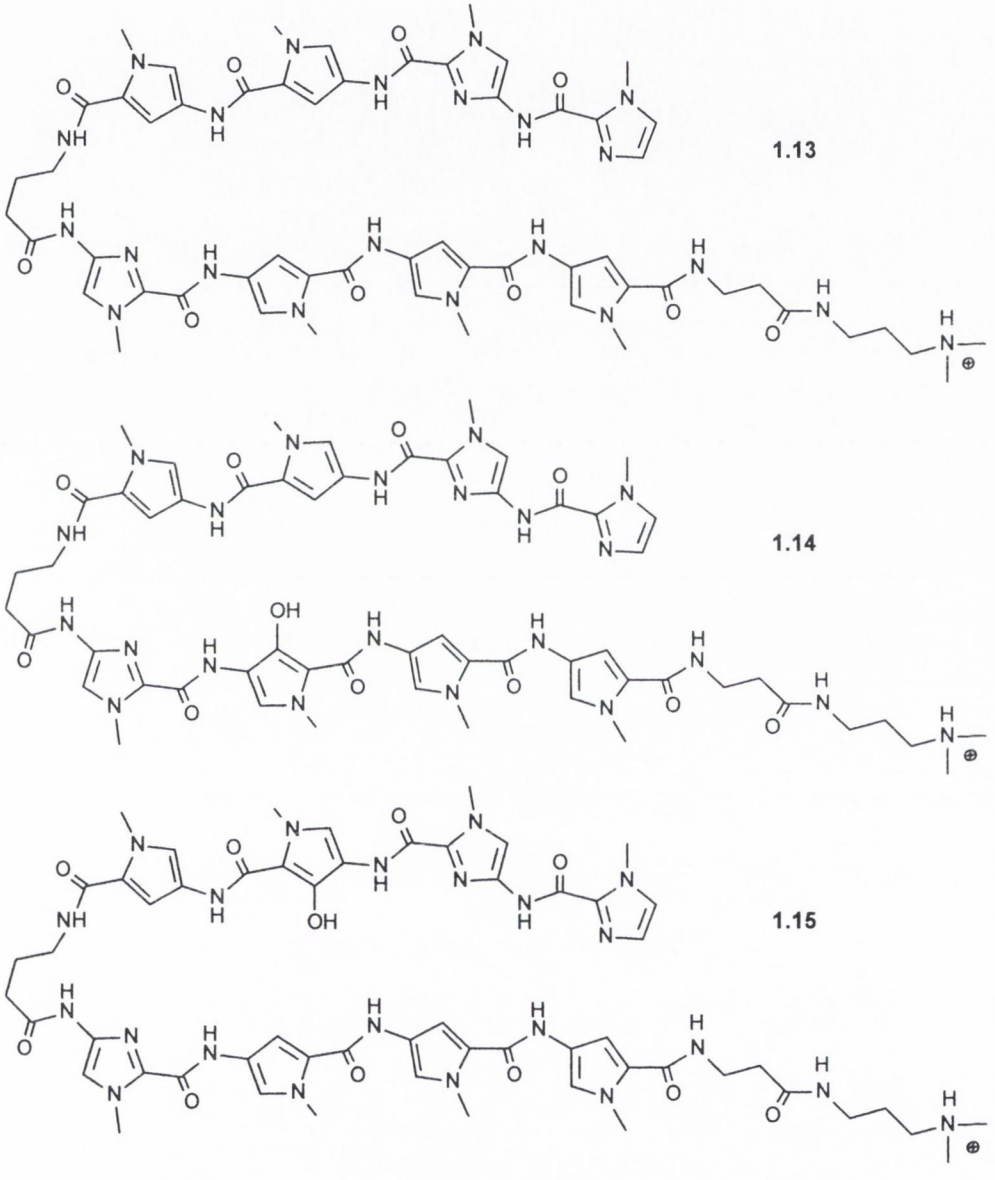


Figure 1.8 Chemical structures and space-filling models of the T-A base pair as viewed from the minor groove of DNA. The hydrogen bond acceptors (N-3 of adenine and C-2 carbonyl of thymine) are in green.

The three 'hairpin' polyamides **1.13**, **1.14** and **1.15** were synthesised and the equilibrium dissociation constants (K_D) determined by quantitative DNase I footprint titration experiments on a 250 base pair DNA fragment containing the two target sites, 5′-TGGACA-3′ and 5′-TGGTCA-3′ which differ by a single A-T/T-A base pair in the fourth position (see **Table 1.3**).[1,23] Based on the pairing rules for polyamide-DNA interactions, **1.13** should bind both sites with near equal affinity, as is the case. The polyamide **1.14** shows a marked preference for binding to 5′-TGGACA-3′, which would involve placing a pyrrole-3-hydroxypyrrole pair opposite an A-T base pair. **1.15**, in which the pairing is reversed (a 3-hydroxypyrrole-pyrrole pair), binds strongest to the 5′-TGGTCA-3′ site. Binding this site would place the 3-hydroxypyrrole-pyrrole pair

opposite a T-A base pair. An x-ray study of a cocrystal of a 3-hydroxypyrrole-containing polyamide bound in the minor groove of DNA fragment has since been published[1.24] and confirms the structural basis of the A-T/T-A base pair recognition by 3-hydroxypyrrole-pyrrole pairs.



1.13

1.14

1.15

Table 1.3 Dissociation constants K_D for interaction of polyamides **1.13-1.15** with DNA

Polyamide	K_D (M) 5′-TGGTCA-3′	K_D (M) 5′-TGGACA-3′
1.13	7.7×10^{-11}	1.5×10^{-10}
1.14	1.5×10^{-8}	8.3×10^{-10}
1.15	4.8×10^{-10}	3.7×10^{-8}

1.4 Sequence Selective Recognition of Double Helical DNA by Major Groove Binding Triplex Forming Oligonucleotides

The idea of selectively binding double stranded DNA at the major groove currently attracts a great deal of attention. This is perhaps inspired by the abundance of natural ligands such as transcription factors and restriction endonucleases, which bind sequence specifically *via* this groove. To the synthetic chemist, the main attractions of targeting the major groove are, a) the ease of access to the four base pairs, due to the wide, shallow nature of the groove, and b) the number and variety of base pair precognition features available by this route. The four base pairs can be unambiguously identified by the pattern and position of the hydrogen bond donor and acceptor sites. The methyl group of thymine, which projects into the major groove, allows further discrimination based on van der Waals or steric interactions (see **Figure 1.9**).

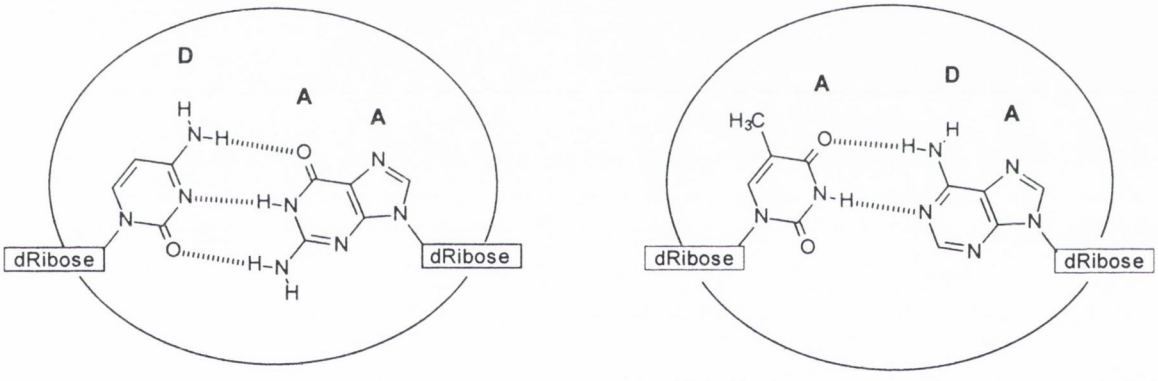


Figure 1.9 C-G (left) and T-A (right) base pairs. Free hydrogen bond donor (**D**) and acceptor (**A**) sites in the major groove of double-stranded DNA.

While the structural basis of many protein-DNA complexes has been elucidated, the knowledge gained from this does not lend itself very well to the design of new DNA receptors. Most proteins bind DNA by adopting a tertiary structure which is complementary in 'shape' to the target DNA sequence. The difficulty in mimicking such systems however, is that the protein folding involved is poorly understood. A more attractive approach would be to build a ligand which is more modular, in just the same way as DNA is.

It has long been recognised that, under the correct conditions, polynucleotides are able to form triple helices.[1,25] Double helical DNA containing a homopurine strand can bind a third polynucleotide, either a homopyrimidine strand (e.g. poly(dGA)-poly(dCT) binding poly(dCT)) or a homopurine strand (e.g. poly(dG)-poly(dC) binding poly(dG)).

More recently, the groups of Dervan and Hélène showed that short homopyrimidine oligonucleotides could form triplexes with specific purine sequences of a DNA duplex target.[1.26] These developments have stimulated extensive research into DNA triple helices and the possibility of using triplex forming oligonucleotides (TFOs) as general DNA recognition devices.

Two main types of DNA triple helices have been identified, the pyrimidine motif and the purine motif, which differ in the composition and orientation of the third strand. In the pyrimidine motif, a homopyrimidine third strand binds in the major groove, parallel to the purine strand of the duplex. The resultant triplexes are held together by interactions between the hydrogen bond donors/acceptors on the purine strand of the duplex and the pyrimidines of the third strand. Thus thymine binds the adenine of an A-T base pair to give a T*A-T base triplet ('*' represents hydrogen bonds of third strand with purine strand of duplex), while protonated cytosine binds the guanine of a G-C base pair affording a C^+*G-C base triplet. (see **Figure 1.10**). Guanine is also able to adopt the pyrimidine motif, binding the guanine from a G-C base pair. This however occurs only under special conditions and guanine generally prefers to adopt the purine motif. There are two important points to note about the pyrimidine motif. Firstly, the recognition of G-C base pairs is pH dependent and ideally requires a pH lower than physiological conditions (*c.a.* pH 7.3). Secondly, triplexes containing contiguous C^+*G-C triplets are destabilised by electrostatic repulsion between adjacent C^+ residues. Thirdly, the T*A-T and C^+*G-C base triplets are both isomorphous, i.e. the C-1' atom of the nucleotides are superimposable. Non-isomorphism results in distortion of the backbones and hence instability of the triplex.

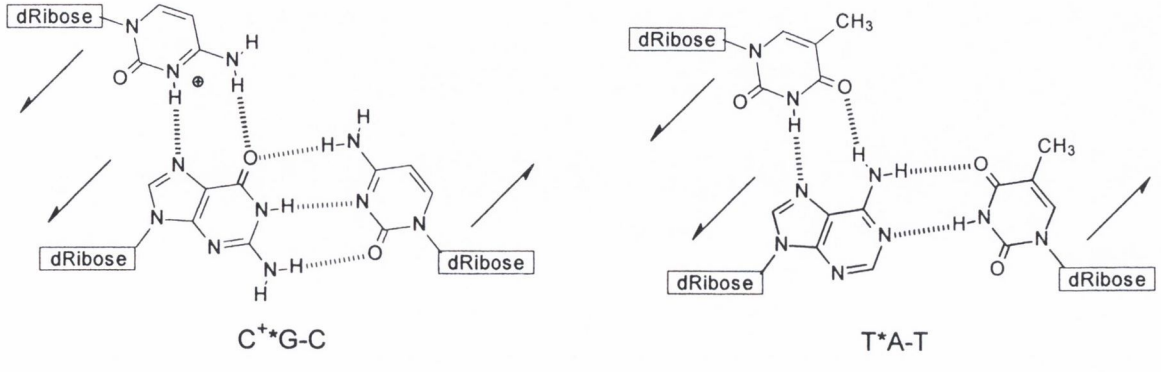


Figure 1.10 Pyrimidine motif homopurine-targeted triple helices. Half-arrows indicate the relative deoxyribose-phosphate (dRibose) backbone orientation.

The purine motif is formed by a homopurine third strand binding in the major groove, antiparallel to the purine strand of the duplex. Again, the triplex is maintained by hydrogen bonding between the third strand and the purine strand of the duplex. Guanine recognises the guanine of a G-C base pair to yield a G*G-C base triplet while adenine can bind to the adenine of an A-T base pair to give an A*A-T triplet. Thymine can also adopt the purine motif and recognises the adenine of an A-T base pair to afford a T*A-T base triplet (see **Figure 1.11**). This T*A-T triplet is related to the pyrimidine motif T*A-T triplet by rotating the third strand thymine through 180 °. The important features to note about the purine motif are, firstly, triplex formation is pH independent. Secondly, the G*G-C and A*A-T triplets are not isomorphous and so A*A-T triplets are generally found only in homopolymers. Thirdly, the T*A-T and G*G-C triplets are not isomorphous and as both can adopt either the pyrimidine or the purine motif (T*A-T preferring the pyrimidine motif and G*G-C preferring the purine motif), a poly(dGT) third strand will adopt the purine motif only if it is sufficiently rich in G-residues. Finally, triplexes containing G-rich third strands are destabilised by the propensity of G-Rich strands to form quadruplexes in the presence of monovalent cations.[27]

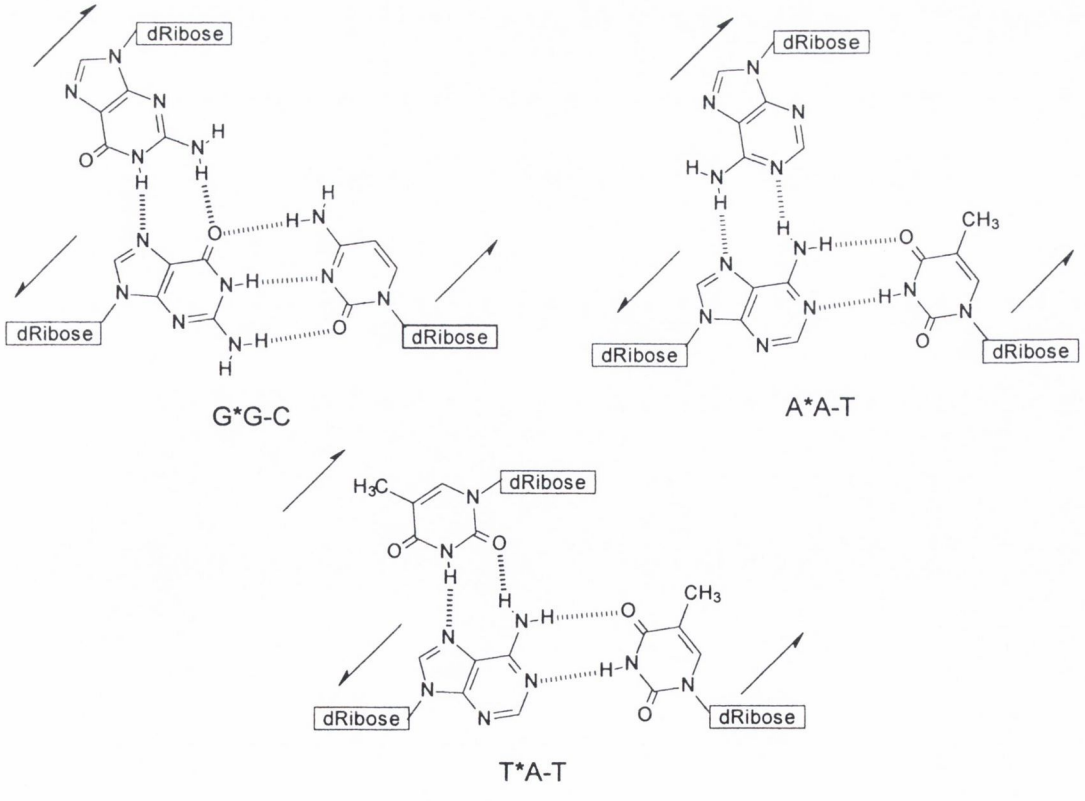


Figure 1.11 Purine motif homopurine-targeted triple helices. Half-arrows indicate the relative deoxyribose-phosphate (dRibose) backbone orientation.

The triple helices as described above are of limited use for sequence specific binding of DNA. This is principally, although not solely, due to the restriction to the recognition of homopurine-homopyrimidine sequences. If TFOs are to be of general application then this problem, along with issues such as pH dependence, non-isomorphism, *etc.* must be overcome. However this is to be achieved, the new binding models must satisfy not just the hydrogen bonding requirements of triplexes, but also questions such as aromatic stacking, van der Waals and electrostatic interactions, tautomerisation and base pair mismatch stability. The following short review [1.41-1.45] is a summary of some of the more important efforts to tackle these problems.

1.4.1 *Replacement of cytosine with pH independent analogues*

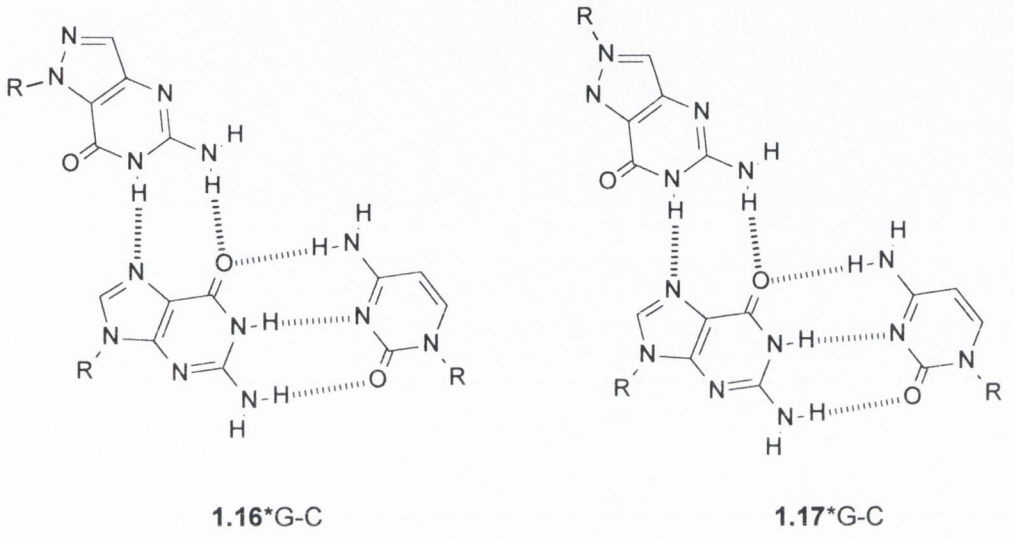
The two main drawbacks of the pyrimidine motif are the lack of pyrimidine recognition and the necessity to protonate cytosine. In an effort to address the second of these problems, cytosine can be replaced with 5-methylcytosine (m^5C).[1.28] This stabilises the triplex, possibly by forming a hydrophobic spine of methyl groups from m^5C and thymine,[1.29] but as protonation of the N-3 of cytosine is still necessary, contiguous m^5C^+*G-C triplets are unfavourable. To avoid this problem, a number of neutral cytosine analogues, which display the same hydrogen bonding pattern as protonated cytosine, have since been investigated.

Contribution of Dervan and co-workers

The two cytosine analogues **1.16** and **1.17** were synthesised and the sequence specificity towards G-C base pairs (see **Figure 1.11**) determined for each.[1.30] DNA affinity cleavage studies using four different oligonucleotides $d(T_7ZT_7)$ (Z = **1.16**, **1.17**, C,) and a 30 base pair duplex containing the target site $d(A_7XA_7)$-$d(T_7YT_7)$ (XY = AT, TA, GC, CG) were carried out at pH 7.4. The most intense cleavage was observed for XY = GC, Z = **1.16** and XY = GC, Z = C, while moderate cleavage was seen for the **1.17***G-C base triplet.

Affinity cleavage studies were also carried out over a range of pH (6.2-7.8), targeting a duplex containing the fifteen base pair purine tract $d(A_5(GA)_5)$ Four oligonucleotides $d(T_5(ZT)_5)$ (Z = **1.16**, **1.17**, C, m^5C) were tested. The affinities of the C-containing and m^5C-containing oligonucleotides dropped to near zero, by pH 7.0 and 7.8 respectively, while the oligonucleotide containing **1.16** bound the duplex strongly at all pH tested. The **1.17**-containing oligonucleotide failed to bind the duplex at any pH. This is probably due to the non-isomorphism of the T*A-T and **1.17***G-C base triplets.

Finally a duplex containing a $(G)_6$ tract was targeted by oligonucleotides containing a $d(C_6)$, $d(m^5C_6)$ or $d(\mathbf{1.16}_6)$ tract over the pH range 6.2-7.4. Affinity cleavage studies indicated that binding by the $d(C_6)$ and $d(m^5C_6)$ tracts, while observable at pH 6.2, was dramatically decreased at pH 6.6 or higher. Binding by the $d(\mathbf{1.16}_6)$ tract was still very strong at pH 7.4.



1.16*G-C **1.17***G-C

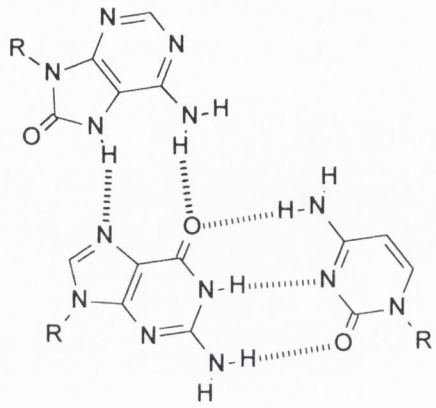
Figure 1.11 Base triplets **1.16***G-C and **1.17***G-C. R = deoxyribose-phosphate backbone

Contribution of Miller and co-workers

It had previously been shown by a number of groups that the cytosine analogue 8-oxo-adenine **1.18**, is capable binding the G-C base pair by forming two hydrogen bonds with the N-7 and C-6 carbonyl of guanine at physiological pH (see **Figure 1.12**).[1.31] In order to further explore the scope of this triplet, comparative binding studies of **1.18**-containing and cytosine-containing oligonucleotides with double stranded DNA were carried out.[1.32]

Triplex formation, by the binding of the oligonucleotides $d(TZ_3TZ_3T_5ZT)$ (Z = **1.18**, C) to a purine tract $d(AG_3AG_3A_5GA)$ of a 28 base pair DNA duplex, was studied by UV thermal denaturation. A melting curve of the cytosine-containing oligonucleotide-DNA mixture at pH 6.0, showed two transitions, the first at 26 °C corresponding to triplex dissociation and a second at 73 °C for the duplex dissociation. At the same pH, a mixture of DNA and the strand containing **1.18** showed only one transition corresponding to dissociation of the duplex. The same denaturation experiments, run at pH 7.0, showed the opposite trend. A mixture of the duplex and the oligonucleotide containing **1.18,** showed a two transition melting curve with a triplex

dissociation temperature of 16 °C. The cytosine-containing strand however failed to form a triplex with the DNA target.

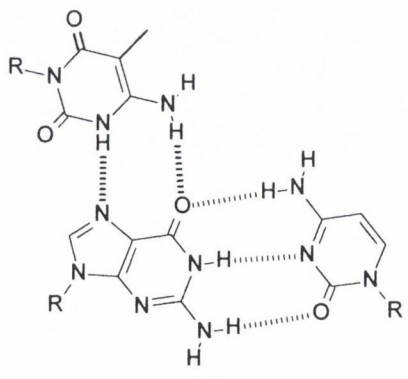


1.18*G-C

Figure 1.12 Base triplets **1.18***G-C. R = deoxyribose-phosphate backbone

Contribution of McLaughlin and co-workers

The use of the cytosine analogue m^5C in TFOs results in increased triplex stability in comparison to using cytosine. This stability is presumably due to the presence of the methyl group at the C-5 position of cytosine. The system however still suffers from poor stability at physiological pH. In order to address this problem, while retaining the advantages of m^5C, **1.19** was synthesised and its affinity with double stranded DNA tested (see **Figure 1.13**). The binding affinities of the 15-mer oligonucleotides $d(T_3ZT_4ZTZTZT_2)$ (Z = C, m^5C, **1.19**) with a 25-mer duplex containing the purine tract $d(A_3GA_4GAGAGA_2)$ and in the presence or absence of spermine, were examined over a range of pH values (6.4-8.5), by UV thermal denaturation studies.[1.33] The results of these studies are shown in **Table 1.4.**



1.19*G-C

Figure 1.13 Base triplets **1.19***G-C. R = deoxyribose-phosphate backbone

Table 1.4 Triplex transition temperature T_m for 15-mers containing C, m^5C and **1.19**, targeting isolated G-C base pairs

		Triplex transition temperature T_m (°C)				
Z	**[spermine]**	**pH 6.4**	**pH 7.0**	**pH 7.5**	**pH 8.0**	**pH 8.5**
C	-	35.4	25.0	19.5	12.1	10.8
C	0.5 mM	45.1	32.6	26.9	15.8	-
m^5C	-	42.9	34.6	33.0	28.7	26.3
m^5C	0.5 mM	54.6	42.6	37.9	27.2	14.4
1.19	-	26.9	27.0	27.5	27.3	26.3
1.19	0.5 mM	39.1	40.7	39.7	38.0	30.8

It was found that at low pH, triplexes formed with TFOs containing C or m^5C were more stable than those containing **1.19**. The strand containing **1.19** however was pH independent and formed triplexes with moderate stability at all pH values. A second study was carried out, targeting a duplex containing five contiguous G-C base pairs. Two oligonucleotides were tested, $d(T_3ZT_4Z_5T_2)$ (Z = m^5C, **1.19**) and both failed to give stable triplexes. A third oligonucleotide was prepared containing alternating m^5C and **1.19** residues and this formed a triplex with the target at pH 6.4-8.0. The inability of m^5C to form a triplex was assumed to be due to electrostatic repulsion of adjacent bases, while poor base stacking or undesirable steric effects of the additional carbonyl group were proposed to be the failing of **1.19**.

1.4.2 Recognition of the pyrimidine of T-A and C-G base pairs within the pyrimidine motif

The principal failing of the pyrimidine motif is the poor recognition of T-A and C-G base pairs. This is because the duplex pyrimidine bases are capable of forming only one hydrogen bound with an incoming third base, thus preventing both discrimination of

bases and efficient binding. Devising a means of overcoming these difficulties is a challenging task, which has yielded little progress to date.

Contribution of Fox and co-workers

A study was carried out on the binding affinities of each of the natural bases A, T, C and G, with single and multiple insertions of the base pair inversions, T-A and C-G, within a purine tract. Duplexes were prepared with the target site $d(A_2GA_4XA_2GA_4)$ (X = T, C) and the affinity for each of the oligonucleotides $d(T_2CT_4ZT_2CT_4)$ (Z = A, C, G, T) was examined by DNase I footprinting in the presence or absence of 10 μM naphthylquinoline (NQ), a triplex stabilising intercalator.[1.34] In the absence of NQ the triplexes containing the non-canonical triplets, G*T-A, C*T-A and T*C-G, were the most stable, producing footprints which persisted to a concentration of 1μM of TFO. This represents a 30-fold reduction in stability relative the same triplex containing a C^+*G-C triplet in place of the non-canonical triplets. The triplets C*C-G and A*C-G give less clear footprints at 1 μM indicating lower stability. The presence of NQ was found to stabilise all eight possible triplets, giving footprints which persisted to 0.3 μl.

Similar studies were carried out on duplexes containing two and three contiguous insertions of the T-A or C-G base pairs. These studies indicated that the G*T-A and T*C-G triplets were the most stable. The presence of NQ stabilised all possible triplets for both double insertions, but a preference for the G*T-A and T*C-G triplets was still observable. In the case of the duplexes with triple insertions, NQ was necessary in order to form the triplexes, with G*T-A and T*C-G being the only triplets formed.

Contribution of McLaughlin and co-workers

The m^5C-intercalator conjugate **1.20** was designed to recognise the C-G base pair by forming a single hydrogen bond between the N-3 of m^5C and the C-4 amine of cytosine. This weak interaction would be supplemented by the intercalation of the tethered naphthalene-based system (see **Figure 1.14**). **1.20** was synthesised and incorporated in to an oligonucleotide for duplex binding studies.[1.35] Two oligonucleotides were prepared $d(T_3CT_4CTZTCT_2)$ (Z = C, **1.20**) and were targeted at four DNA duplexes, each containing $d(A_3GA_4GAXAGA_2)$-$d(T_3CT_4CTYTCT_2)$ (X-Y = A-T, T-A, C-G, G-C). Triplex formation was monitored by UV thermal denaturation studies, the results of which are shown in **Table 1.5**.

The TFO containing cytosine at Z, shows a pH dependent preference for the G-C base pair as expected. The conjugate base **1.20** appears to bind all four base pairs, while showing increased stability when binding the C-G base pair. This may be due to the C-G base pair benefiting from both intercalation and hydrogen bond formation, while the other base pairs benefit from intercalation only.

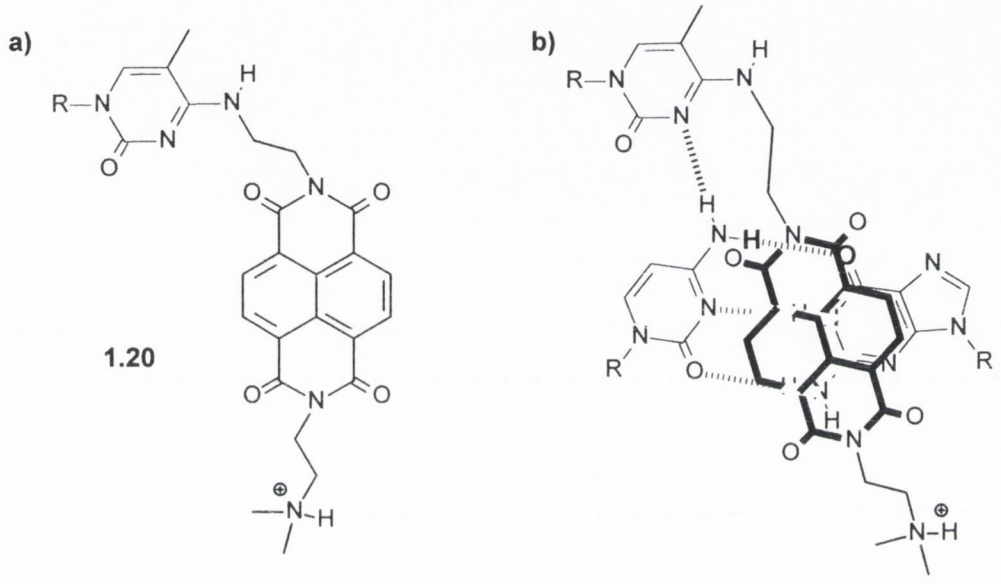


Figure 1.14 a) Structure of analogue **1.20**. b) Potential hydrogen bonding and intercalation interactions between **1.20** and a C-G base pair.

Table 1.5 Triplex transition temperature T_m for 15-mers containing C, and **1.20**, targeting various base pairs.

	Triplex transition temperature T_m (°C)		
Base triplet	**pH 6.4**	**pH 7.0**	**pH 7.5**
X-Y = G-C			
Z = C	30	22	15
Z = 1.20	28	22	-
X-Y = C-G			
Z = C	13	-	-
Z = 1.20	36	30	27
X-Y = A-T			
Z = C	9	-	-
Z = 1.20	26	19	11
X-Y T-A			

Z = C	-	-	-
Z = 1.20	28	22	broad[a]

[a] Broad transition below 20 °C

1.4.3 *Inhibition of quadruplex formation by G-rich TFOs in purine motif*

One of the major limitations of the purine motif is the tendency of G-rich strands to self-associate in quadruplexes, thus reducing the efficiency of the TFO. These structures are formed by eight intermolecular hydrogen bonds and are stabilised by co-ordination of monovalent cations, especially K^+ and Na^+ (see **Figure 1.15**). The significant presence of K^+ and Na^+ in the physiological environment would suggest that the formation of such structures may occur to some degree *in vivo*.

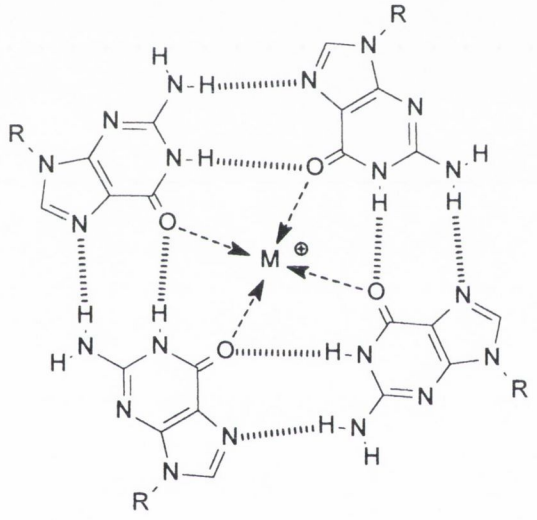
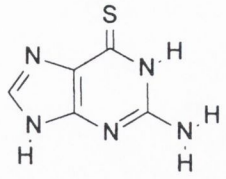


Figure 1.15 Structure of hydrogen bonded G-tetrad, with an internally co-ordinated monovalent cation M^+

Contribution of Revankar and co-workers

In the hope of inhibiting quadruplex formation, 6-thioguanine **1.21** was substituted for guanine. It was reasoned that the sulphur atom would reduce the strength of the hydrogen bonding and also disrupt ion chelation. In order to assess the effect of this replacement, a number of TFOs with varying degrees of **1.21** substitution were prepared. The dissociation constant (K_D) of a duplex, containing the purine-rich target tract d(GAG_2AG_2 $AG_2CGCAG_2AG_2AG_2A_3G_6CG_4$), binding the TFO d($G_4TG_6T_3G_2TG_2TG_2T_2GTG_2TG_2T$ G_2TG) was determined to be 6 x 10^{-9} M.[1.36] The K_D in the presence of 50 mM KCl was found to increase to 1 x 10^{-6} M. The same duplex, targeted with TFOs which have increased substitution of guanine with **1.21**, yielded

increased K_D values (K_D = 3 x 10^{-8} M for ~50% substitution). The inhibitory effect of adding KCl was however greatly reduced as substitution increased (K_D = 3 x 10^{-8} M for ~50% substitution in presence of 50 mM KCl).



1.21

Contribution of Maher and co-workers

The effect of substituting **1.21** for guanine was investigated by Maher also. A duplex containing the purine target tract d($G_3AG_3AG_2AG_3$) was complexed by each of the two TFOs d(GZGTGZGTZGTGZG) (Z = G, **1.21**) and the dissociation constant K_D determined with or without K^+ or Na^+ present.[1.37] The results of these studies are summarised in **Table 1.7**. The triplex containing the **1.21***G-C base triplets was almost completely resistant to inhibition by M^+. The triplex showed no sign of dissociation until the KCl concentration was at 200 mM, whereas the unsubstituted triplex was completely dissociated at 50 mM KCl.

Table 1.7 Effect of M^+ on triplex dissociation constants K_D

Z	10 mM M^+	K_D (M)
G	-	1.4 x 10^{-7}
G	Na^+	2.1 x 10^{-7}
G	K^+	5.3 x 10^{-6}
1.21	-	2.2 x 10^{-7}
1.21	Na^+	2.0 x 10^{-7}
1.21	K^+	1.9 x 10^{-7}

1.4.4 *Recognition of the pyrimidine of T-A and C-G base pairs within the purine motif*

The lack of pyrimidine recognition is another major deficiency within the purine motif. As with the pyrimidine motif, the poor recognition is due to the lack of distinguishing structural information presented by the pyrimidine.

Contribution of Dervan and co-workers

A systematic study was carried out, of the binding affinities of each of the four bases for the four possible base pairs, within the purine motif. A 15 base pair site $d(AG_4AG_4XG_3A)$ (X = A, C, G, T) in a 648 base pair DNA fragment was targeted with a purine rich third strand $d(TG_3ZG_4TG_4T)$ (Z = A, C, G, T) and the equilibrium association constants for the resulting triplexes determined using quantitative DNase I footprining.[1.38] The results of these experiments are summarised in **Table 1.6**.

Table 1.6 Association constants K_T (M^{-1}) for the formation of 16 triple helical complexes containing the Z*X-Y triplets at 37 °C

	Z =			
X-Y	**A**	**G**	**C**	**T**
A-T	5.3×10^7	1.9×10^6	4.0×10^6	8.9×10^7
G-C	8.2×10^6	5.5×10^7	1.7×10^6	2.9×10^6
C-G	1.1×10^6	5.8×10^5	8.5×10^5	1.3×10^7
T-A	$8.5\ 10^5$	2.7×10^5	8.3×10^5	1.7×10^6

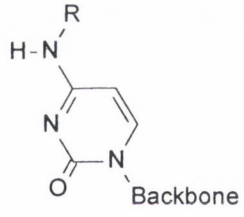
The three canonical base triplets G*G-C, A*A-T and T*A-T have the highest association constants, as would be expected. There are also two moderately stable base triplets, A*G-C and T*C-G. The T*C-G triplet may allow recognition of the C-G base pair, but due to the degeneracy of the recognition properties of T, it is expected that specificity would be compromised.

1.4.5 Recognition of T-A and C-G base pairs by extended bases

As stated before, the recognition of C-G and T-A base pairs by binding the pyrimidine base is hindered by the paucity of structural information carried by the pyrimidine. An improvement in selectivity might be achieved by using an extended base, which would be capable of spanning the major groove so as to interact with and bind, not just the pyrimidine, but also the purine bases. It is possible to distinguish between each of the four base pairs, by the hydrogen bonding donor/acceptor pattern presented at the major groove. Thus, it should be feasible to design systems which by spanning the entire groove can discriminate between all four base pairs.

Contribution of Miller and co-workers

The series of cytosine derivatives **1.22a-d** was synthesised and incorporated into oligonucleotides. UV thermal denaturation studies of the affinities of these oligonucleotides, d($m^5CT_2m^5CT_6ZT_4$) (Z = C, **1.22a-d**), with the duplexes d(GA_2GA_6 XA_4)-d($T_4YT_6CT_2C$) (X-Y = C-G, G-C, T-A, A-T) were carried out.[1.39] The results of these studies are shown in **Table 1.7**.



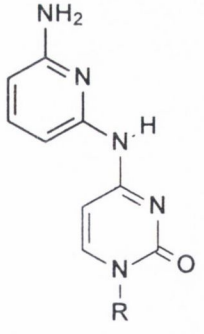
1.22a, R = $(CH_2)_3CH_3$
1.22b, R = $(CH_2)_3COOH$
1.22c, R = $(CH_2)_3NH_2$
1.22d, R = $(CH_2)_3NHCOCH$

The results would suggest that the 3-aminopropyl and 3-acetamidopropyl side chains interact in a specific manner with the C-G base pair. Molecular modelling indicated that the side chain is long enough to span the major groove, and allow formation of a hydrogen bond between the NH of the 3-aminopropyl or 3-acetamidopropyl group, and the C-6 carbonyl of guanine.

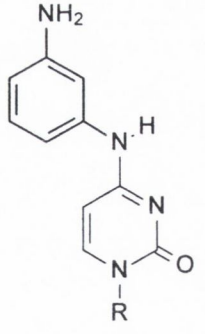
Table 1.7 Triplex transition temperatures T_m for 15-mers containing C and **1.22a-d**, targeting various base pairs

X	Y	Z	T_m (°C)
C	G	C	<0
C	G	**1.22a**	<0
C	G	**1.22b**	<0
C	G	**1.22c**	10
C	G	**1.22d**	20
T	A	**1.22d**	<0
A	T	**1.22d**	<0
G	C	**1.22d**	8
G	C	C	32

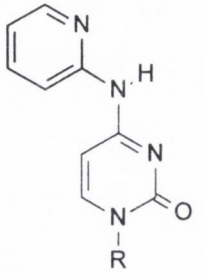
Reasoning that by replacing the flexible propyl chain of **1.22c-d** with a more rigid structure, the binding efficiency might be improved, the N^4-(aryl)cytosine derivatives **1.23a-c** were synthesised.[1.40]



1.23a

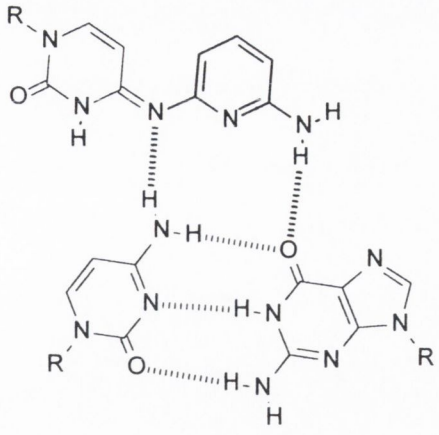


1.23b

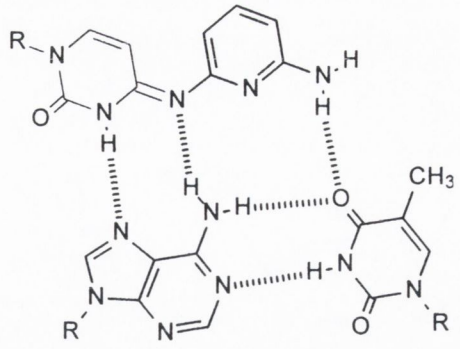


1.23c

Molecular modelling indicated that the pyridine ring of **1.23a** can span the major groove and place the amino group within hydrogen bonding range of the N-7 or C-6 carbonyl of guanine. If **1.23a** was to tautomerise to an imine as in **Figure 1.16**, an additional hydrogen bond between the C-4 amino group of cytosine and the imino nitrogen could also form. UV and NMR studies were carried out and indicated that **1.23a** does undergo tautomerisation to the imino structure.



1.23a*C-G



1.23a*A-T

Figure 1.16 Hydrogen bonding scheme for **1.23a***C-G and **1.23a***A-T base triplets

The novel bases were incorporated into oligonucleotides d($m^5CT_2m^5CT_6ZT_4$) (Z = **1.23a-c**) and their affinities to the duplexes d($GA_2GA_6XA_4$)-d($T_4YT_6CT_2C$) (X-Y = A-T, T-A, G-C, C-G) was assessed by UV thermal denaturation studies. The results of these studies are summarised in **Table 1.8**.

Table 1.8 Triplex transition temperature T_m (°C) for 15-mers containing **1.23a-c**, with various base pairs

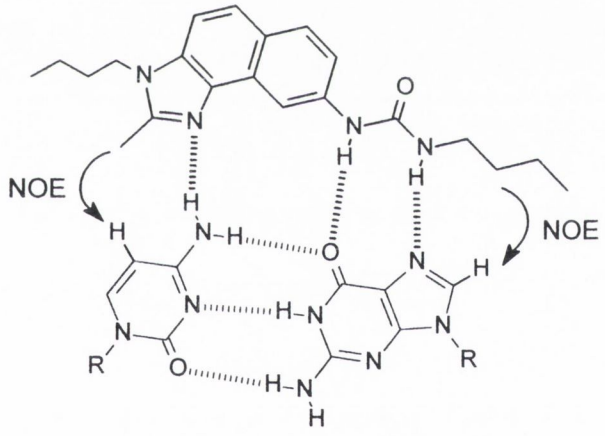
		T_m (°C)		
X	**Y**	**Z = 1.23a**	**Z = 1.23b**	**Z = 1.23c**
A	T	32	29	-
T	A	23	23	<5
G	C	20	18	-
C	G	25, 38	23, 36	<5

The C-G base pair gave two different triplex transition temperatures with each of the TFOs incorporating **1.23a** or **1.23b**, depending on the rate of heating. This could indicate the presence of two different modes of binding of the TFO, possibly intercalation and hydrogen bonding. Thermal denaturation studies of the these TFOs with a 45 base pair duplex containing the same target tract showed one T_m at 28 °C. Besides binding to the C-G base pair, **1.23a** and **1.23b** also form triplets with the A-T base pair. This is probably by the hydrogen bonding pattern shown in **Figure 1.16**. **1.23a** and **1.23b**

are also able to form moderately stable triplets with other base pairs, this lack of specificity may be due to the aromatic pyridinyl ring stabilising these 'mismatches' by π-stacking with adjacent bases.

Contribution of Zimmerman and Schmitt

The novel base **1.24** was designed to span the major groove and recognise a C-G base pair by specific hydrogen bond formation as shown in **Figure 1.17**. **1.24** was synthesised and its binding to a C-G base pair was examined by ^{1}H NMR titration experiments.[1.41] A 1:1 mixture of tri-*O*-acetylguanosine and tri-*O*-acetylcytidine in $CDCl_3$ was titrated with **1.24** and the association constant K_a calculated to be 160 M^{-1}. The orientation of the binding was confirmed by difference NOE experiments on a 1:1:1 mixture of 9-octylguanine-1-propylcytosine-**1.24**. Weak intermolecular contacts were observable for the methyl group of **1.24** with H-5 of cytosine and between the β-methylene protons of **1.24** with H-8 of guanine.

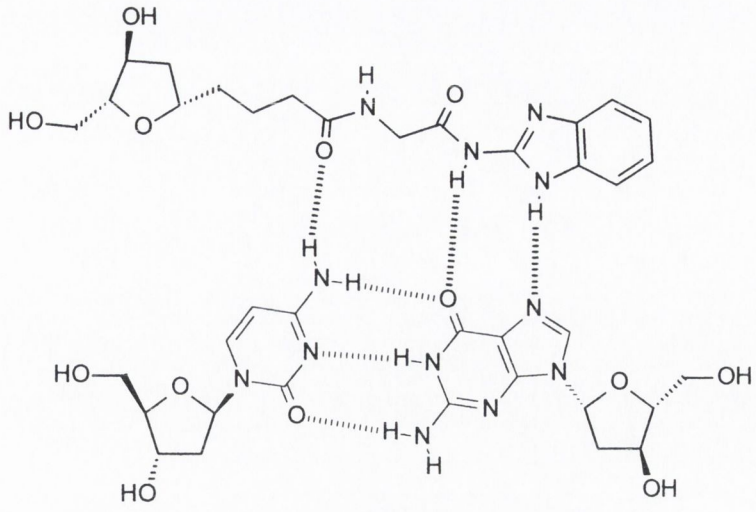


1.24*C-G

Figure 1.17 Hydrogen bonding scheme and observed NOEs for **1.24***C-G base triplet

Contribution of Sasaki, Maeda and co-workers

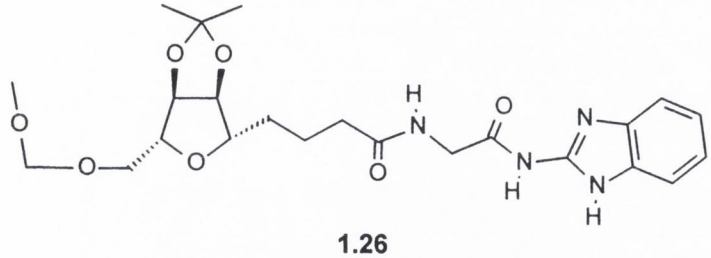
A semi-empirical molecular orbital calculation was used to design **1.25**, an artificial nucleobase which by spanning the major groove can form three specific hydrogen bonds with a C-G base pair as shown in **Figure 1.18**. Molecular modelling of this triplet suggested that the deoxyribose groups can adopt positions isomorphous with those of T*A-T and C^+*G-C base triplets.



1.25*C-G

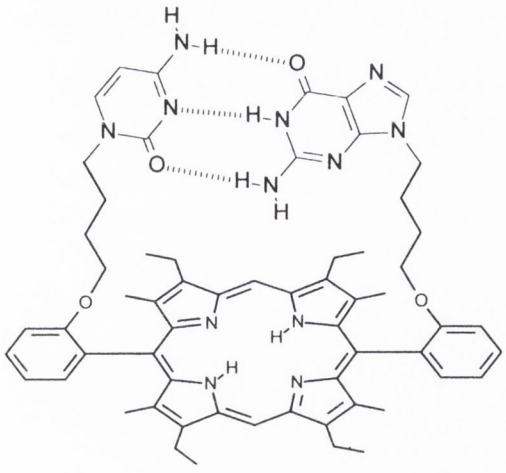
Figure 1.18 Hydrogen bonding scheme for **1.25***C-G base triplet

1.26, a protected form of 2'-oxy-**1.25,** was synthesised and the recognition properties with the four natural deoxynucleotides examined by ^{1}H NMR titration experiments in $CDCl_3$.[1.42] **1.26** was found to bind cytidine in a 1:1 complex with an association constant K_a of 2.6 x 10^3 M^{-1}, approximately half that of the G-C base pair. **1.26** also bound guanosine in a 1:2 ratio with a K_a value of 3 x 10^2 M^{-1}. As the association of **1.26** with the C and G bases precluded the use of ^{1}H NMR titration techniques to calculate K_a for the formation of a **1.26***C-G base triplet, intermolecular NOE experiments were used to probe triplet formation.

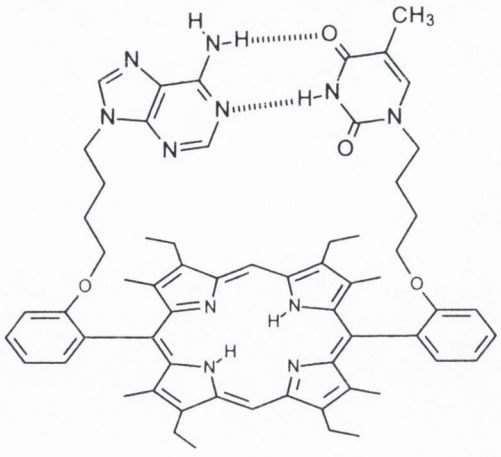


1.26

The compounds **1.27** and **1.28**, in which the C-G and T-A bases are fixed in close proximity to each other, on a porphyrin skeleton, were prepared as models of stable base pairs. The NOESY spectrum of a 1:1 mixture of **1.26** and **1.27** showed NOE cross peaks indicative of formation of a **1.26***C-G base Triplet. No such cross peaks were seen for the a 1:1 mixture of **1.26** and **1.28**.



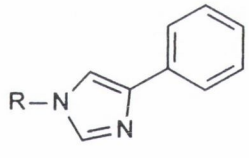
1.27



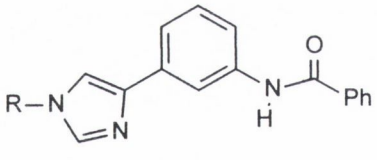
1.28

Contribution of Dervan and co-workers

From model building, two non-natural bases **1.29** and **1.30** were chosen as candidates for recognition of a C-G base pair. It was proposed that **1.29** would form a single hydrogen bond between the imidazole N-3 and the amino group of cytosine. **1.30** was designed to span both strands of the C-G base pair and form two hydrogen bonds.



1.29



1.30

Triplex formation by a 15 base pair purine tract with one variable base pair site d(A_7XA_7)-d(T_7YT_7) (X-Y = A-T, C-G, G-C, T-A) within a 30 base pair duplex, and a series of 15-mer oligonucleotides d(T_7ZT_7) (Z = T, C, **1.29**, **1.30**) was examined using affinity cleavage studies.[1.43] The studies indicated that the most stable base triplets formed were the T*A-T, C^+*G-C, **1.30***C-G and **1.30***T-A base triplets. This demonstrated that **1.30** discriminates for pyrimidine-purine base pairs over purine-pyrimidine. However, a mechanism for this discrimination could not be formulated on the basis of the hydrogen bonding interactions between **1.30** and the four base pairs.

In an attempt to determine the binding mode of **1.30** with the T-A base pair, the oligonucleotide d(AGATAGAACCCCTTCTATCTTATATCT**1.30**TCTT), which folds

to produce an intramolecular triplex incorporating a **1.30***T-A base triplet (see **Figure 1.19**), was synthesised and studied.[1.44] The ^{1}H NMR spectra of this molecule is consistent with an intramolecular triplex. In the NOESY spectrum however, a number of expected NOE cross peaks between T_4 and A_{17}, between T_{16} and A_{17} and between T_{27} and $\mathbf{1.30}_{28}$ were missing. There were also some unexpected NOE cross peaks which could only be explained if **1.30** was intercalated between the T_{29}*A_5-T_{16} base triplet and the T_4-A_{17} base pair. An extensive NMR study of this **1.30**-containing intramolecular triplex, has since been published, and indicates that **1.30** preferentially stacks over pyrimidine-purine base pairs.[1.45]

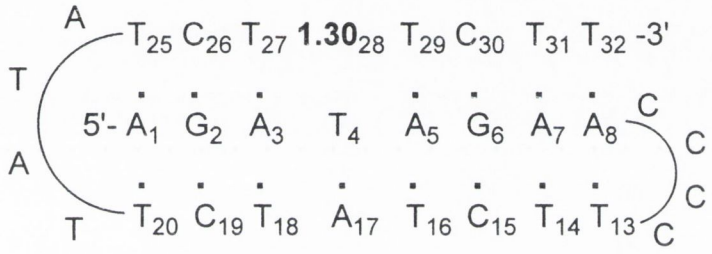
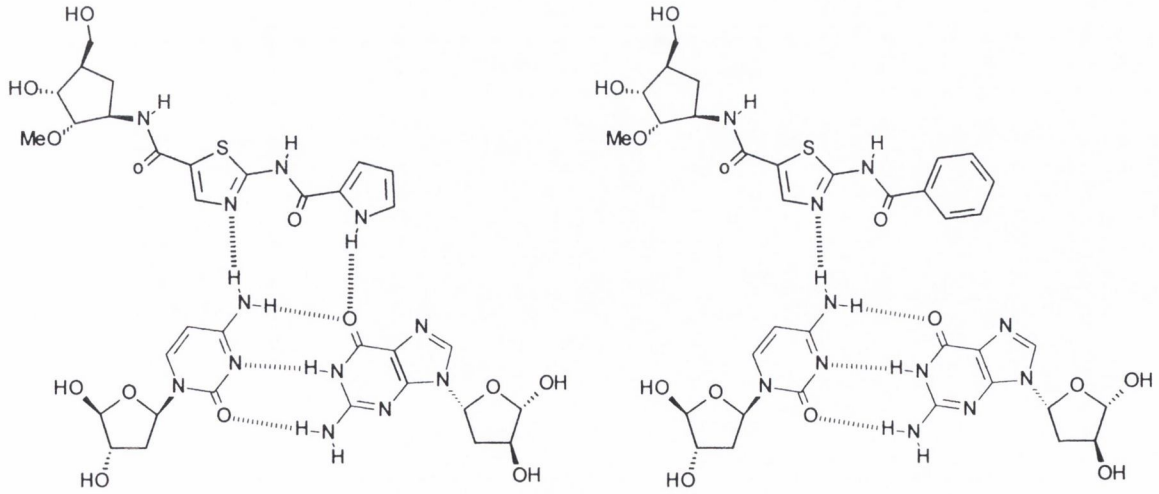


Figure 1.19 Schematic of the proposed folding of the 32-base intramolecular triplex with the numbering system used in text. Hydrogen bonds are indicated by (.).

Two additional systems, **1.31** and **1.32**, were designed using molecular modelling, with the intention of binding C-G base pairs. It was thought that **1.31** would form two hydrogen bonds, one with guanine and the other with cytosine, while **1.32** would bind cytosine with a single hydrogen bond (see **Figure 1.20**).



1.31a*C-G **1.32a***C-G

Figure 1.20 Hydrogen bonding schemes for **1.31***C-G and **1.32***C-G base triplets

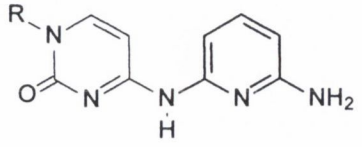
Using quantitative footprinting titration experiments, the binding affinities and selectivities of two oligonucleotides d(TTm^{5}CTm^{5}CZTm^{5}CTm^{5}Cm^{5}CTTT) (Z = **1.31**, **1.32**) for a 314 base pair DNA fragment containing all four possible 15 base pair double stranded binding sites d(AAGAGAXAGGAAA) (X = A, C, G, T) were determined.[1.46] The results of these experiments are summarised in **Table 1.9**.

Table 1.9 Association constants K_T (M^{-1}) 15-mers containing **1.31** and **1.32**, with various base pairs

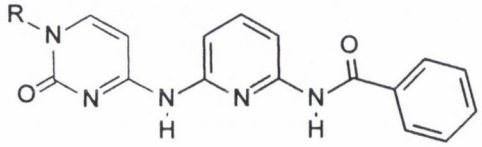
X	Z = 1.31	Z = 1.32
C	9.2×10^5	2.4×10^6
T	1.6×10^6	2.6×10^6
G	1.4×10^5	1.0×10^5
A	1.5×10^5	1.9×10^5

1.31 and **1.32** both show the same selectivity for pyrimidine-purine base pairs as **1.30** does. As **1.31** and **1.32** possess a system of aromatic rings fused by amide bonds similar to **1.30**, it is probable that **1.31** and **1.32** also bind base pairs by an intercalation mechanism.

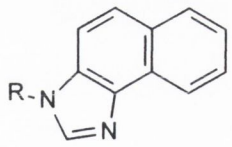
In order to compare these results with other nucleobases possessing extended aromatic structures, a number of additional systems were synthesised. The nucleobase described by Miller and co-workers[1.40] **1.23a**, the *N*-benzoylated derivative of **1.23a**, **1.33**, the imidazoloquinoline **1.34**, the 8-aminoimidazoloquinoline **1.35**, and the deoxyribose analogue of the nucleobase **1.24** described by Zimmerman and co-workers,[1.41] **1.36**, were synthesised and incorporated into the same TFOs as **1.31** and **1.32** had been. The binding affinities and specificities of oligonucleotides were determined by quantitative footprinting. The results of these studies are given in **Table 1.10**.



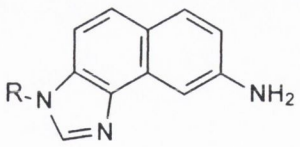
1.23a



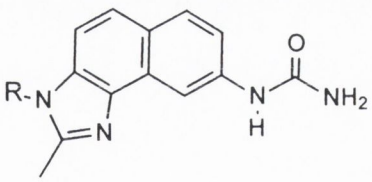
1.33



1.34



1.35



1.36

Table 1.10 Association constants K_T (M^{-1}) 15-mers containing **1.23a** and **1.31-1.36**, with various base pairs

Compound	Base pairs			
	C-G	T-A	G-C	A-T
1.32	9.2×10^5	1.6×10^6	1.4×10^5	1.5×10^5
1.32	2.4×10^6	2.6×10^6	1.0×10^5	1.9×10^5
1.33	8.5×10^6	5.8×10^6	5.7×10^5	8.8×10^5
1.23a	$\sim 1 \times 10^5$	$< 1 \times 10^5$	7.0×10^5	$< 1 \times 10^5$
1.34	$< 1 \times 10^5$	$< 1 \times 10^5$	$< 1 \times 10^5$	$< 1 \times 10^5$
1.35	$< 1 \times 10^5$	$< 1 \times 10^5$	2.6×10^5	$< 1 \times 10^5$
1.36	$< 1 \times 10^5$	$< 1 \times 10^5$	1×10^5	$< 1 \times 10^5$

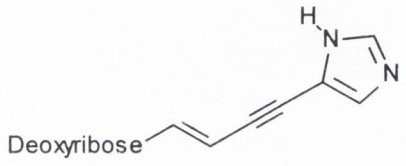
The extended nucleobase **1.33** shows similar selectivity as **1.30**, **1.31** and **1.32**, suggesting that **1.35** also binds by intercalation. The nucleobases **1.23a**, **1.34**, **1.35** and **1.36**, all of which have shorter aromatic systems, show low affinity for each of the four base pairs. Considering the NMR studies of Zimmerman *et al.* on **1.24**, it appears that models of triplex systems in organic solvents may bear little resemblance to the same system under aqueous conditions.

Contribution of Richards and Rothman

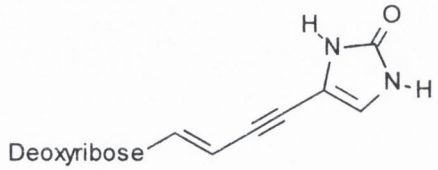
To this date, there have been no publications of extended nucleobases selectively binding T-A base pairs. There has been some work in this direction however, with a report by Richards and Rothman[1.47] on the molecular modelling of a series of T-A base pair ligands.

The four structures **1.37a-d** were proposed as non-natural nucleobases capable of binding T-A base pairs. The binding of the 11-mer oligonucleotides $d(T_5ZT_5)$ (Z = **1.37a-d**) to a 11 base pair duplex $d(A_5XA_5)$-$d(T_5YT_5)$ (X-Y = T-A, G-C) in both A-DNA

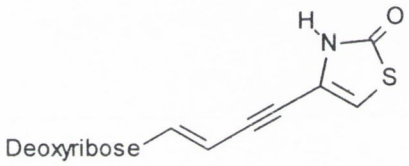
and B-DNA configurations was modelled using computer based molecular modelling. These studies predicted that **1.37d** would selective a T-A base pair over a G-C in B-DNA, with an affinity roughly halfway between that of the G*T-A and T*A-T base triplets. The nucleobase **1.37c** was predicted to show a preference for the T-A base pair in A-DNA, but with a relatively low affinity (similar magnitude as the binding of T-A by guanine). Synthesis and binding studies on these systems however have not been reported.



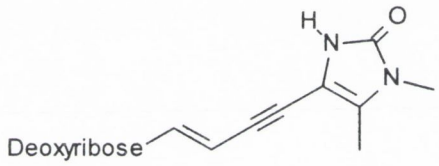
1.37a



1.37b



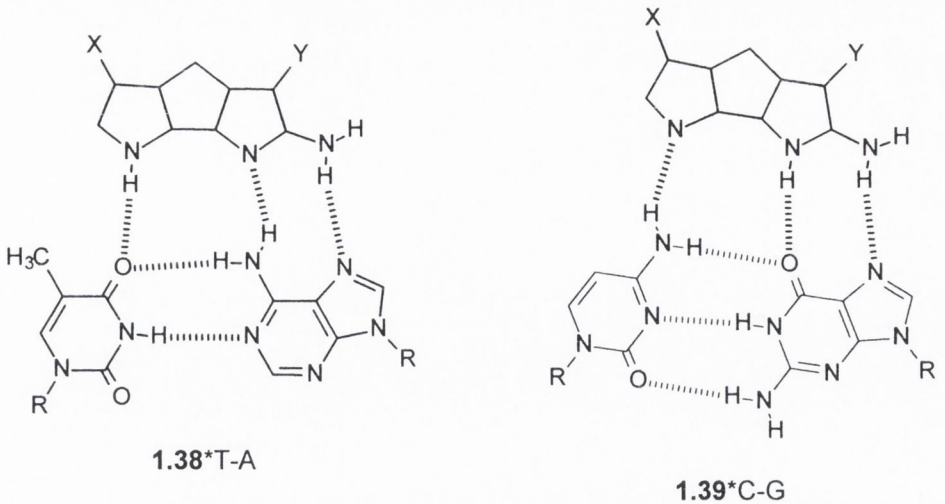
1.37c



1.37d

1.5 Aims of this Project

To date, the contribution of the Davis group to a general sequence selective DNA recognition scheme has been rather small. Preliminary studies carried out within the group have identified two (5,5,5)-tricyclic frameworks **1.38** and **1.39**, which display the required hydrogen bonding donor/acceptor patterns for recognition of the A-T/T-A and G-C/C-G base pairs respectively (see **Figure 1.21**). It was proposed that discrimination between A-T and T-A and between G-C and C-G base pairs would achieved by varying the point of attachment to the third strand. Attaching **1.38** or **1.39** at X would select for the T-A and C-G base pairs respectively, while attachment at Y would specify for A-t and G-C base pairs respectively.

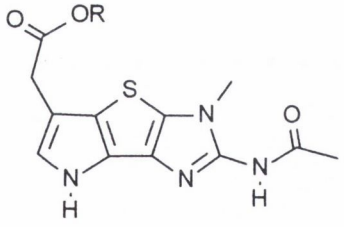


1.38*T-A

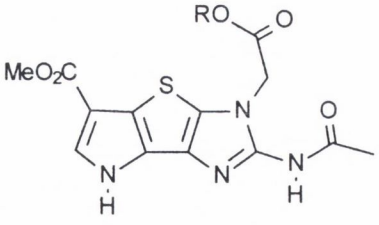
1.39*C-G

Figure 1.21 Proposed hydrogen bonding schemes for the **1.38***T-A and **1.39***C-G triplets

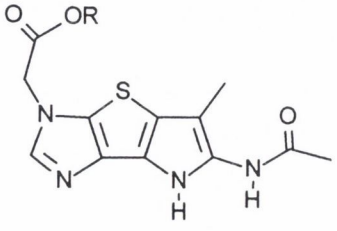
By a combination of computer-based literature searches and molecular modelling, four closely related molecules **1.40-1.43** were designed for base pair recognition.[1.48] The CO_2Me group of **1.41** and the *N*-acetyl groups of **1.40-1.43** were intended to moderate the electron-rich nature of the tricyclic system. In addition, the *N*-acetyl groups should help to fill the major groove and may provide triplex stability through π-stacking and hydrophobic interactions. The methoxycarbonylmethyl groups were intended to provide a convenient route for incorporating the ligands into a PNA or analogous backbone.[1.49] Molecular modelling indicated that such novel third strands, docked with complementary double helical DNA, would form stable triplexes.



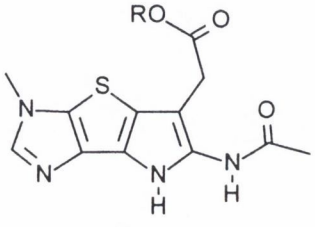
1.40



1.41



1.42

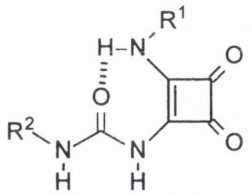


1.43

It was planned to initially focus our synthetic efforts on securing the proposed T-A ligand, **1.40**. It was hoped that on achieving **1.40**, the synthetic route could then be modified to afford the closely related structures **1.41**, **1.42** and **1.43**.

We intended that the binding affinity of **1.40** to a T-A base pair would initially be examined in organic solvents using NMR techniques. Then, having linked **1.40** to a suitable backbone, procedures such as UV thermal denaturation, affinity cleavage studies, *etc.* would be used to assess strength and specificity of DNA binding under conditions closer to a physiological environment. It was also considered not unreasonable to hope that a third strand containing **1.40** might cocrystallise with a complementary DNA duplex, allowing the triplex to be observed by x-ray crystallography.

In addition to the (5,5,5)-tricyclic system, a second potential G-C/C-G base pair ligand was identified in the *N*-carbamoyl squaramides of the type **1.44**. The squaramide unit had previously been shown by Ballester, Costa and co-workers to provide a distinctive array of hydrogen bonding sites for molecular recognition.[1.50] Compound **1.44** features a ureidyl substituent, the carbonyl of which would probably form an intramolecular hydrogen bond to the adjacent NH, resulting in a structure with a preorganised array of hydrogen bonding donor/acceptor sites.



1.44

It was thought that these novel compounds might bind C-G/G-C base pairs in organic solvents. The binding was expected to take place *via* interaction of the squaramide hydrogen bonding array with the cytosine exocyclic amine and the hydrogen bond acceptor sites on guanine, as shown in **Figure 1.22**.

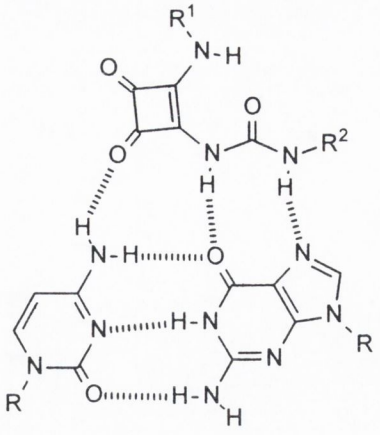


Figure 1.22 Proposed hydrogen bonding scheme for the **1.44***C-G triplet.

We were also aware that **1.44** bears an intriguing resemblance to guanine and envisaged the squaramide binding cytosine through the Watson-Crick hydrogen bonding sites (see **Figure 1.23**).

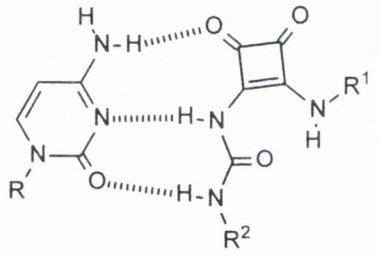


Figure 1.23 Proposed hydrogen bonding scheme of a C-**1.44** pair

The binding potency of these compounds was to be initially investigated by NMR techniques. It was hoped that if any were found to exhibit moderate to high binding of cytosine or cytosine-guanine base pairs, they would be incorporated into oligonucleotides and their ability to bind single and double stranded DNA examined.

Chapter Two

The Thieno[2,3-*d*]imidazole system

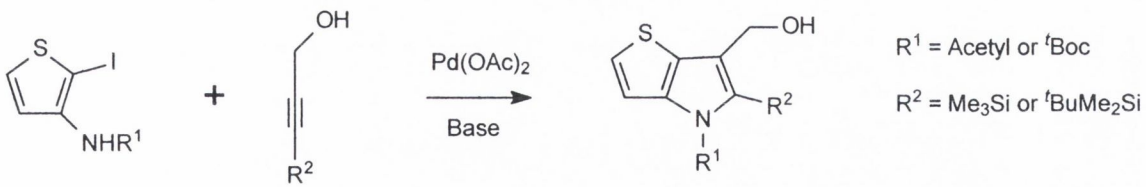
2.1 Preamble

The attempted synthesis of the T-A binding tricyclic compound **1.40** is described in this chapter. It was considered that 3,4 diaminothiophene, which had been described by Outurquin and Paulmier,[2.1] offered the best starting point for the synthesis of **1.40**, as it introduced two of the key nitrogen atoms necessary for the final product. From this diamine, we planned to explore the synthesis of both the thieno[3,2-*b*]pyrrolo system and thieno[2,3-*d*]imidazole. It was hoped that having achieved these bicycles, the methodologies could then be combined to allow the synthesis of **1.40**.

Section 2.2 deals with the synthesis of the key intermediate **2.1** and the first attempt to synthesise the thieno[3,2-*b*]pyrrole system. The attempted synthesis of 2,3-diaminothiophene derivatives and the conversion of these into the thieno[2,3-*d*]imidazole system is then covered in section 2.3.

2.2 The Attempted Synthesis of a 6-Substituted Thieno[3,2-*b*]pyrrole System

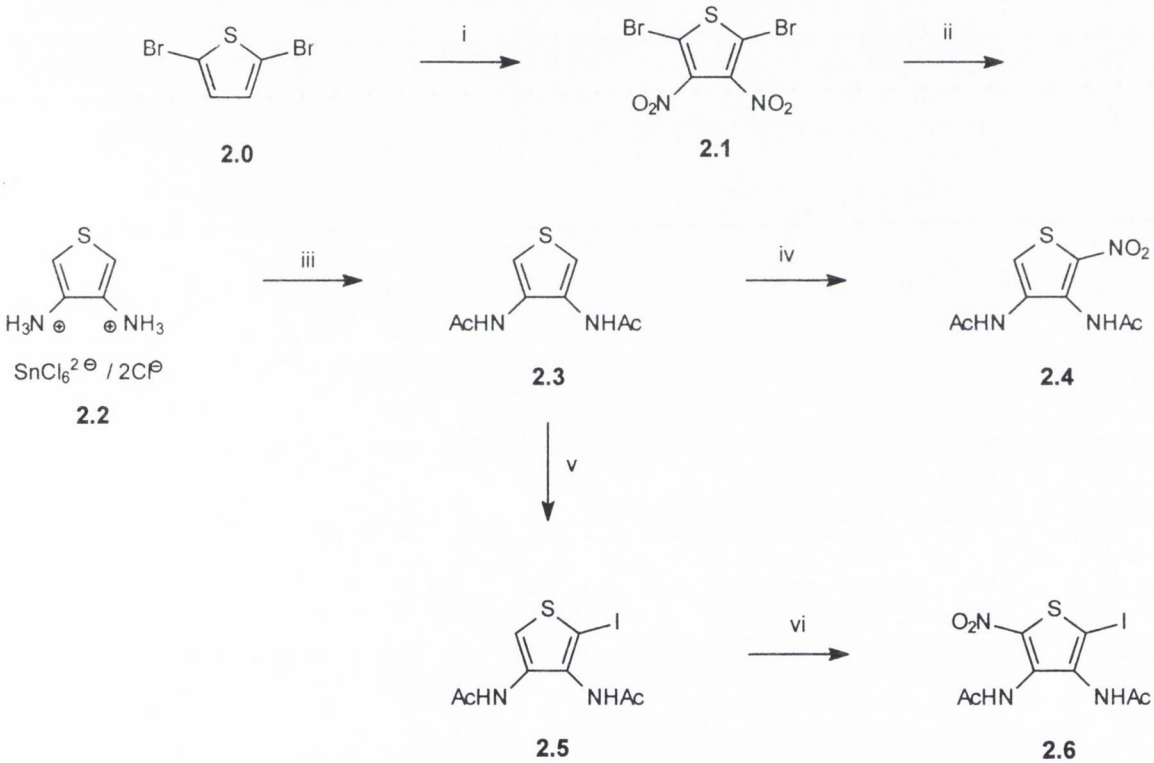
On searching the literature for thieno[3,2-*b*]pyrrole compounds, numerous procedures for the synthesis of 5-substituted[2.2] and 5,6-disubstituted[2.2, 2.3] derivatives were found. Of these procedures, the method of Gronowitz *et al*[2.3] for synthesising a 6-hydroxymethyl-5-silylthieno[3,2-*b*]pyrrole (see **Scheme 2.1**) seemed the most promising.



Scheme 2.1

It was felt that adapting this procedure to the synthesis of the tricyclic system **1.40** offered a number of advantages. Firstly, the silyl group might be a useful protecting group for the 5 position, which would be easily removed by protodesilylation at a later stage. Secondly, the alcohol, through oxidation to a carboxylic acid, would allow eventual linkage to a oligonucleotide backbone.

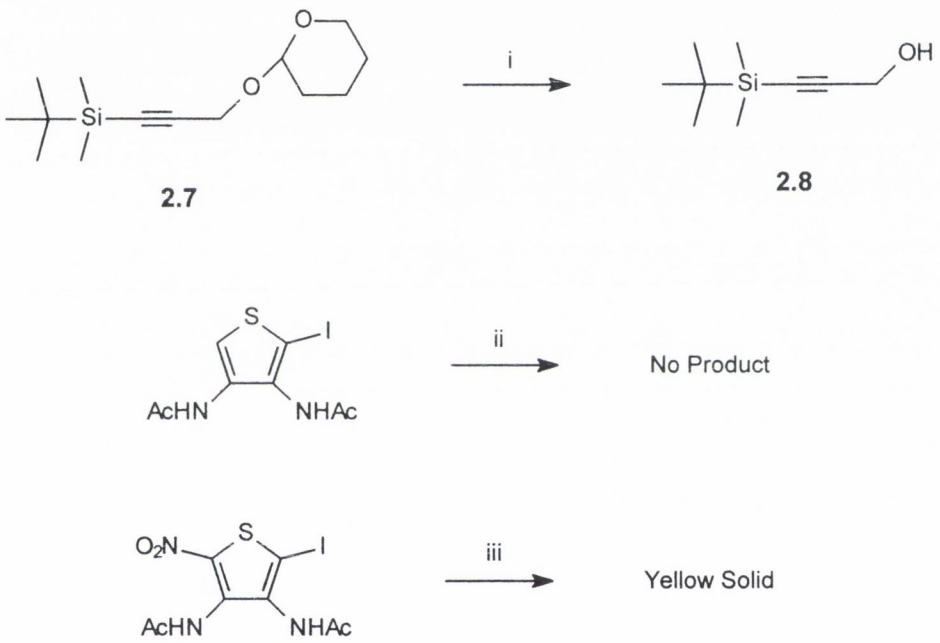
Scheme 2.2 shows the synthesis of the 3,4-diacetamido-2-iodothiophene **2.5** and 3,4-diacetamido-2-iodo-5-nitrothiophene **2.6**. First, the dibromothiophene was bisnitrated by the method of Easton.[2.4] The dibromo dinitrothiophene was reduced, bis(*N*-acetylated) and then mononitrated by a modified method of Outurquin and Paulmier to give **2.4**.[2.1] Attempted iodination of **2.4** by the use of *N*-iodosuccinimide or iodine monochloride failed to give any product. This, it is presumed, is due to the presence of the electron-withdrawing nitro group, which deactivates the aromatic ring towards electrophilic substitution. Iodination of the diacetamide **2.3**, by a method adapted from Brunett[2.5] afforded the iodide **2.5**. Nitration of **2.5** by a procedure adapted from Rinkes[2.6] gave the iodo nitro diacetamide **2.6**.



Scheme 2.2 *Reagents, conditions (and yields)*: i, conc. H_2SO_4, 65% SO_3 fuming H_2SO_4, fuming HNO_3, 20-30 °C, 5 minutes (38%); ii, conc. HCl, Sn pellets, <30 °C, 4 h; iii, saturated aqueous Na_2CO_3, Ac_2O, room temperature,17 h (13% from **2.1**); iv, fuming HNO_3, AcOH, Ac_2O, 20 °C, 2 h (81%); v, *N*-Iodosuccinimide, CH_2Cl_2, room temperature 2 days (77%); vi, fuming HNO_3, AcOH, Ac_2O, 20 °C, 3 h (38%) .

The deprotection of THP protected alcohol **2.7**[2.7] by the method of Mori[2.8] and the attempted addition of iodothiophenes **2.5** and **2.6** across the resulting alkyne **2.8** is shown in **Scheme 2.3**. The addition of **2.5** across alkyne **2.8** failed to give any product, and **2.5**

and **2.8** could be recovered quantitatively. The iodo nitrothiophene **2.6** under the same reaction conditions, gave a complex mixture after 2 days from which no heterocyclic material could be isolated. It was found however that if the reaction was stopped after only 2½ hours, the alkyne **2.8** was recovered quantitatively along with a yellow solid which appears to be either the 3- or 4-amino derivative of **2.6** as shown in **Figure 2.1**. The ^{1}H NMR supports this hypothesis, having three singlets at 2.06, 8.07 and 9.43 ppm which integrate for 3H, 2H and 1H respectively. The ^{13}C NMR is also supportive, having one CH_3 signal and five quaternary carbon signals, 22.84, 95.07, 122.24, 131.49, 146.67 and 169.33 ppm respectively.



Scheme 2.3 *Reagents, conditions (and yields)*: i, *p*-TsOH, MeOH, room temperature, 24 h (82%); ii, **2.8**, Bu_4NCl, Na_2CO_3, $Pd(OAc)_2$ (cat.), dry DMF, 95 °C, 2 days; iii, **2.8**, Bu_4NCl, Na_2CO_3, $Pd(OAc)_2$ (cat.), dry DMF, 80 °C, 150 minutes.

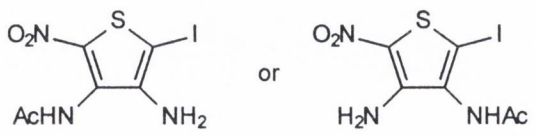
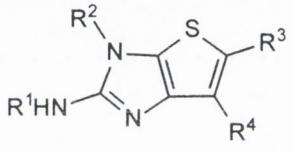


Figure 2.1 Possible structures of unknown yellow product from reaction of **2.5** with **2.8**.

2.3 The Attempts to Synthesise the 2-Aminothieno[2,3-*d*]imidazole System

Our attempts to secure the thieno[3,2-*b*]pyrrole system having been thwarted, we focused upon the of the thieno[2,3-*d*]imidazole system **2.9**. A search of the literature for 2-aminothieno[2,3-*d*]imidazole compounds failed to uncover any examples. There are however numerous procedures for the preparation of the related 2-aminobenzimidazole from 1,2-phenylenediamine using either cyanogen bromide[2.9] or cyanamide.[2.10] It was hoped that on obtaining a 2,3-diaminothiophene, these procedures could be adapted to give **2.9**. A second strategy envisaged, was based on the work of Schulze,[2.11] whereby reductive cyclisation of an *ortho*-cyanamido nitrothiophene would give **2.9**.



2.9

2.3.1 Investigation of the reactions of 2-nitro-3,4-diacetamido-thiophene 2.4

With these strategies in mind, the reduction of 2-nitro-3,4-diacetamidothiophene **2.4** was attempted. Pd/C catalysed hydrogenation of **2.4** in AcOH, or reduction with zinc powder, also in AcOH, gave tarry residues as product. The reduction of **2.4** by the action of tin powder in conc. HCl, followed by basification with aqueous NaOH and *in situ* acetylation with acetic anhydride, to yield a triacetamide, had been reported by Outurquin and Paulmier.[2.1] Attempts, however, to isolate the free amine by aqueous base washing, again yielded tarry residues. These results are not surprising as aminothiophenes are notoriously unstable and normally must be manipulated as ammonium salts or *N*-acyl derivatives. The presence of electron-withdrawing substituents however is often found to increase the stability of the amines.

In the case of **2.4**, the presence of the nitro group *ortho* to a amido substituent allows a 'push-pull' interaction to be set up, as represented by the canonical forms in **Figure 2.2**. This, in addition to stabilising **2.4**, results in the *ortho*-amide bearing a partial positive charge on the nitrogen. As a consequence of this, the 3-amido group should be more susceptible to base hydrolysis than the 4-amido.

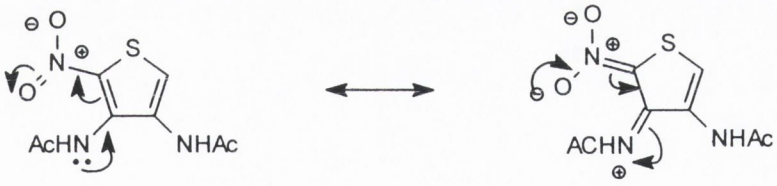


Figure 2.2

Scheme 2.4 shows the selective hydrolysis of **2.4** with KOH to afford the 4-acetamido-3-amino-2-nitrothiophene **2.10** in low yield after continuous extraction. The structure of **2.10** was confirmed by selective decoupling of the ^{1}H coupled ^{13}C NMR spectrum. Outurquin and Paulmier[2.1] reported a higher yield (36%) for this hydrolysis of **2.4** using conc. H_2SO_4, but this result could not be reproduced and instead an inseparable, 1:1 mixture by ^{1}H NMR ,of **2.10** and 3,4-diamino-2-nitrothiophene was isolated.

2.4 **2.10**

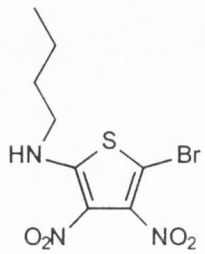
Scheme 2.4 *Reagents, conditions (and yields)*: i, 10% KOH, room temperature, 18 h (24%).

The synthesis of 4-acetamido-3-cyanamido-2-nitrothiophene was attempted by a method adapted from Anatol and Berecoechea;[2.12] however treating the amine **2.10** with cyanogen bromide in water gave the starting material unaffected. The suspicion that this result was due to low nucleophilicity of the amino group was supported by the failure of **2.10** to acetylate on treatment with acetic anhydride. This is presumably a result of the electronic interaction between the amino and nitro groups.

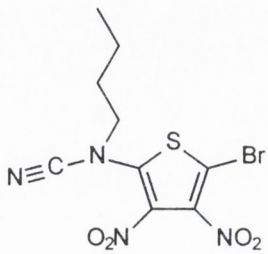
In an effort to reduce the nitro group and so remove this interaction, **2.10** was treated with tin powder in HCl to give a cream solid, presumed to be the 2-ammonium chloride salt. This solid however decomposed to a tar when treated with saturated aqueous $NaHCO_3$ in the presence of either cyanogen bromide, acetic anhydride or di-*tert*-butyl dicarbonate.

2.3.2 Investigation of the reactions of 5-amino-3,4-dinitrothiophenes 2.11 and 2.12

In addition to **2.4**, the amino dinitrothiophene **2.11** and (*N*-butylcyanamido) dinitrothiophene **2.12** were considered potential intermediates to a 2-aminothieno[2,3-*d*]imidazole. The butyl chain was included in order to prevent tautomerisation in the target imidazole ring.

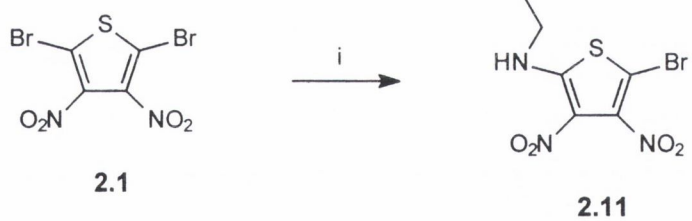


2.11



2.12

The synthesis of **2.11** is shown in **Scheme 2.5**. Butylamino S_NAr substitution of 2,5-dibromo-3,4-dinitrothiophene gave the butylamino thiophene **2.11**. Attempts to convert **2.11** to the cyanamido thiophene **2.12** by treatment with cyanogen bromide in water/THF left the amine unaffected.

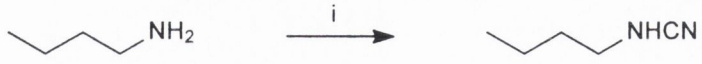


2.1 **2.11**

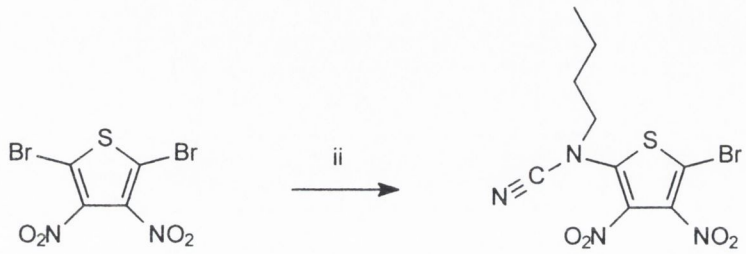
Scheme 2.5 *Reagents, conditions (and yields)*: i, Butylamine, CH_2Cl_2, 0 °C 1h then reflux for 18h (76%).

Having failed to introduce the cyano group into the aminothiophene, it was decided to attempt a *N*-butylcyanamido S_NAr substitution of **2.1**. The synthesis of butylcyanamide and its reaction with **2.1** is shown in **Scheme 2.6**. Butylcyanamide **2.12** was synthesised by the method of Ainley, Curd and Rose.[2.13] Treating **2.1** with **2.12** afforded the 5-butylcyanamido thiophene **2.13** in 17% yield after 2 hours. The low yield appears to be due to the product decomposing over the course of the reaction (the purified product decomposed considerably when placed *in vacuo* for 20 minutes). Regular analysis by TLC (CH_2Cl_2-hexane, 1:1) of the reaction also suggested this; the

amount of product was approximately constant over the 2 hours, while side products increased and the starting material was eventually completely consumed.



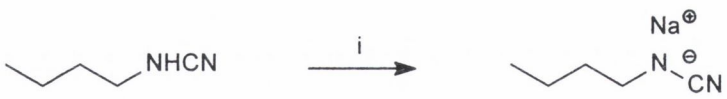
2.12



2.1 **2.13**

Scheme 2.6 *Reagents, conditions (and yields)*: i, CNBr, dry Et_2O, -5 °C then room temperature, 90 minutes (95%); ii, **2.12**, $^{i}Pr_2EtN$, dry CH_2Cl_2, 0 °C then reflux , 2 h (17%).

In the hope of shortening the reaction time and so perhaps increasing product yield, sodium butylcyanamide **2.14** was synthesised as shown in **Scheme 2.7** by modification of a method of Krommer.[2.14] Treatment of **2.1** with **2.14** in either dry DMF or dry DMSO appeared by TLC analysis to afford **2.13** but the reaction was slow, failing to go to completion within 2 hours and so offered no advantage over the previous method.



2.12 **2.14**

Scheme 2.7 *Reagents, conditions (and yields)*: i, NaOH powder, 2-propanol, 20 °C, 2 h (91%).

Due to the low yield and poor stability of **2.13,** it was decided to abandon this route to the 2-aminothieno[2,3-d]imidazole system. In a final bid to secure a 2,3-diaminothiophene, the reduction of **2.11** was attempted by a number of methods. Treatment of **2.11** with tin powder in conc. HCl, iron powder in AcOH, tin II chloride in conc. HCl and EtOH or zinc powder in 20% NaOH in EtOH left the starting material unaffected in each case.

2.11 was however affected by the action of tin II chloride in EtOH at 70 °C affording a pale yellow crystalline solid as product. The elemental analysis of this solid indicated that the bromine group had been lost, as is common in the reduction of *ortho*-bromo nitrothiophenes. However the remaining elements found, and the ^{1}H and ^{13}C NMR were not consistent with an aminothiophene product. The solid was found to be stable to the presence of base and was unaffected by treatment with Ac_2O. This would suggest that the reduction was incomplete, perhaps stopping at a nitroso product.

Chapter Three

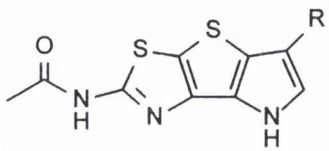
The *N*-Benzyl Protected Pyrrolo[2',3':4,5]thieno [3,2-*d*]thiazole 3.16

3.1 Preamble

Having failed in all attempts to synthesise a thieno[2,3-*d*]imidazole, due mainly it was felt, to the inaccessibility of 2,3-diaminothiophenes, our attention turned to the development of alternatives to the imidazole ring. Our choice of modifications was limited by the necessity to retain the base-pair specificity for which the original system had been designed. We wished to keep both the five membered ring and the imine-like nitrogen of the imidazole ring. Computer based molecular modelling[3.1] suggested that the pyrrolo[2',3':4,5]thieno[3,2-*d*]thiazole system **3.1**, produced by substituting a thiazole ring for the imidazole, would be a suitable replacement for pyrrolothieno[2,3-*d*]imidazole. Whilst replacing the nitrogen with a larger sulphur atom would obviously result in some distortion of the tricycle, modelling of **3.1** with a T-A base pair indicated that it would be able to form a stable triplet.

Besides the danger of losing specificity for the T-A base pair, the other major concern with the use of a thiazole ring was the loss of the backbone linkage site, which was to be used for an A-T base pair ligand. It was felt however that this loss was outweighed by the advantages of accessing a molecule, which could bind a T-A base pair. If the binding properties of **3.1** proved encouraging enough, efforts to synthesise the imidazole based system would be renewed.

The development of a synthetic route to the pyrrolo[2',3':4,5]thieno[3,2-*d*]thiazole system **3.1** is discussed in this chapter. It was originally planned to form the pyrrole ring, by an intramolecular Heck type reaction on an enamine such as **3.11**, to give a 5-ethoxycarbonyl-pyrrolo[2', 3':4, 5]thieno[3,2-*d*]thiazole. It was considered that the ester group would help to stabilise the potentially unstable, electron rich tricycle. The ethoxycarbonyl group would ultimately be replaced with a ethoxycarbonylmethyl group for incorporation of the tricycle into a PNA strand.

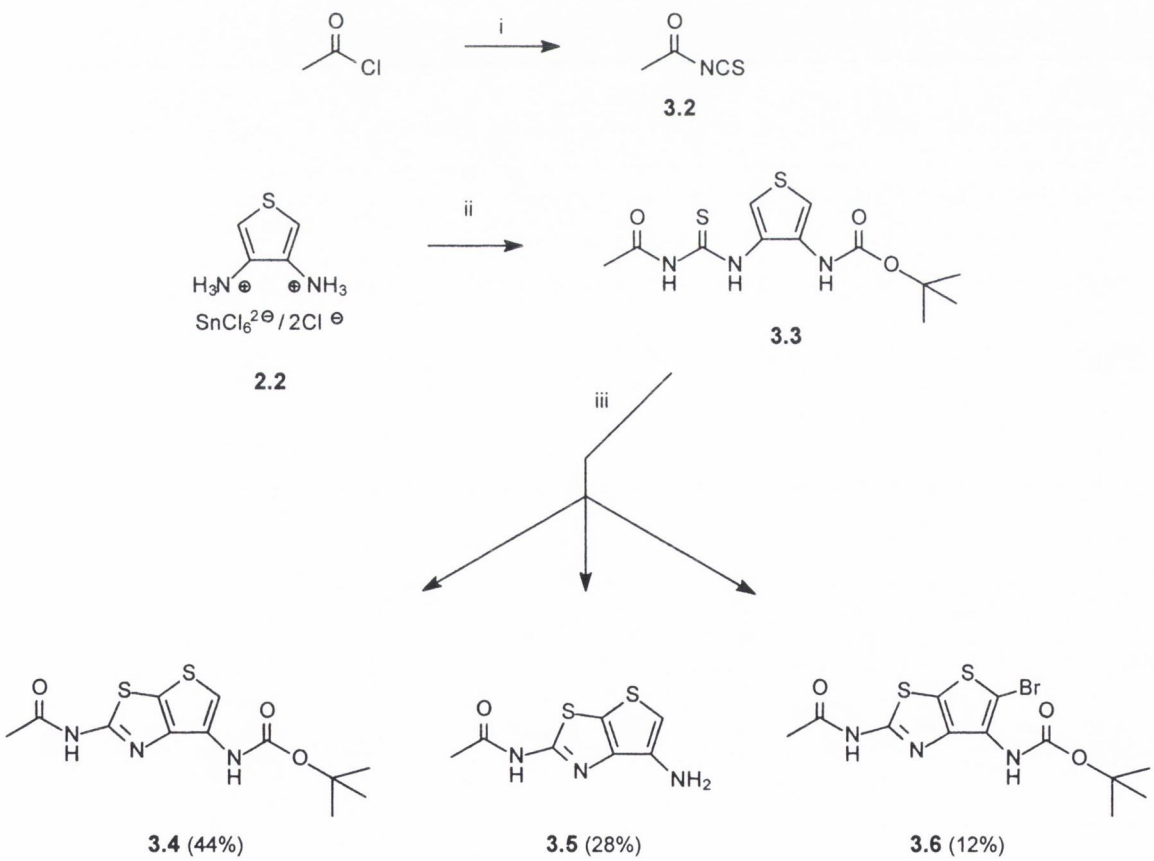


3.1

The synthesis of the thieno[3,2-*d*]thiazole bicycle is covered in section 3.2. Section 3.3 deals with the attempts to synthesise a variety of enamines as intermediates to the 5-ethoxycarbonyl tricycle. The synthesis of the *N*-benzyl protected 5-ethoxycarbonylmethyl tricycle **3.16** and attempts to remove the benzyl group are outlined in section 3.4.

3.2 The Synthesis of the 2-Acetamidothieno[3,2-*d*]thiazole System

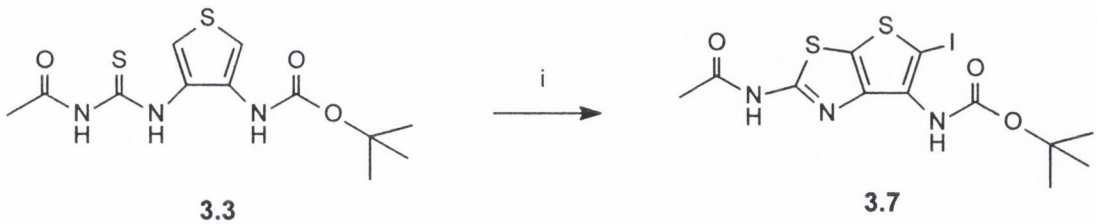
The steps involved in the formation of acetylisothiocyanate **3.2**, and the 2-acetamidothieno[3,2-*d*]thiazoles **3.4, 3.5** and **3.6** are outlined in **Scheme 3.1**. Acetylisothiocyanate was synthesised by a modification of a method of Takamizawa.[3.2] Initially, purification by vacuum distillation gave poor yields (<10%), but the yield was increased considerably (74%) by using flash chromatography eluting with diethyl ether instead.



Scheme 3.1 *Reagents, conditions (and yields)*: i, KSCN, dry acetone, room temperature, 20 minutes (74%); ii, Et_3N, dry CH_2Cl_2, <10 °C, 15 minutes, then **3.2**, room temperature, 18 h, then di-*tert*-butyl dicarbonate, reflux 18 h (47% from **3.2**); iii, Br_2, $CHCl_3$, room temperature, 10 minutes.

Grehn[3.3] reported the oxidative cyclisation of 1-acyl-3-(3-thienyl)-2-thioureas with bromine in AcOH to give 2-acylaminothieno[3,2-*d*]thiazoles. With the intention of employing this method, the ^{t}Boc-protected thiourea **3.3** was synthsised. Cyclisation of **3.3** with bromine in AcOH gave a 1:1 mixture (by TLC analysis) of the ^{t}Boc protected thienothiazole **3.4** and the 6-aminothienothiazole **3.5**. Suspecting that the ^{t}Boc was hydrolysed by a HBr/AcOH mixture (HBr is generated in the cyclisation), the solvent was changed to $CHCl_3$. Cyclising **3.3** in this solvent reduced the amount of free amine formed, but also resulted in a third product being generated. This product was confirmed as the 5-bromothienthiazole **3.6** by comparison with a sample synthesised by treating **3.4** with *N*-bromosuccinimide. Attempts to neutralise the HBr by carrying out the cyclisation in the presence of Et_3N merely resulted in a longer reaction time and the formation of other side products.

Prompted by the prospect of a one-pot conversion of **3.3** into a 5-iodothienothiazole, cyclisation using iodinating reagents was investigated. Treating **3.3** with iodine in $CHCl_3$ gave a mixture of products by TLC analysis. It was found however that using two equivalents of *N*-iodosuccinimide gave the thienothiazole **3.4** as a single product by TLC analysis and this was iodinated to the 5-iodothienothiazole **3.7** by a further equivalent. The one pot synthesis of **3.7** from **3.3** is outlined in **Scheme 3.2**. No signs of 6-aminothienothiazole products were detected using this method.



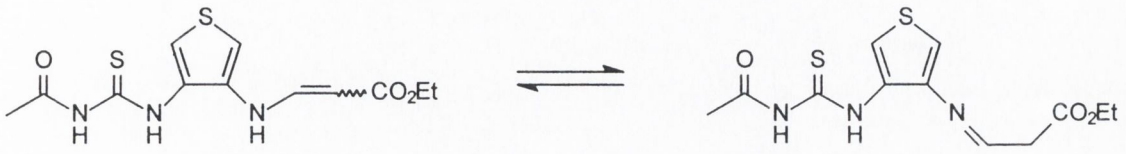
3.3 **3.7**

Scheme 3.2 *Reagents, conditions (and yields)*; *N*-Iodosuccinimide, $CHCl_3$, room temperature, 4 h (73%).

3.3 The Synthesis of Enamine Derivatives 3.10, 3.11 and 3.12

With the synthesis of the thieno[3,2-d]thiazole skeleton accomplished, our attention turned to the closure of the pyrrole ring. The first strategy considered, was to derivatise the 3-(4-aminothienyl)thiourea **3.9** as the enamine **3.10** and then cyclise this to give the corresponding thienothiazole. Halogenation followed by a Heck type reaction would give the tricycle. Although it was considered that there was a strong possibility of bromination

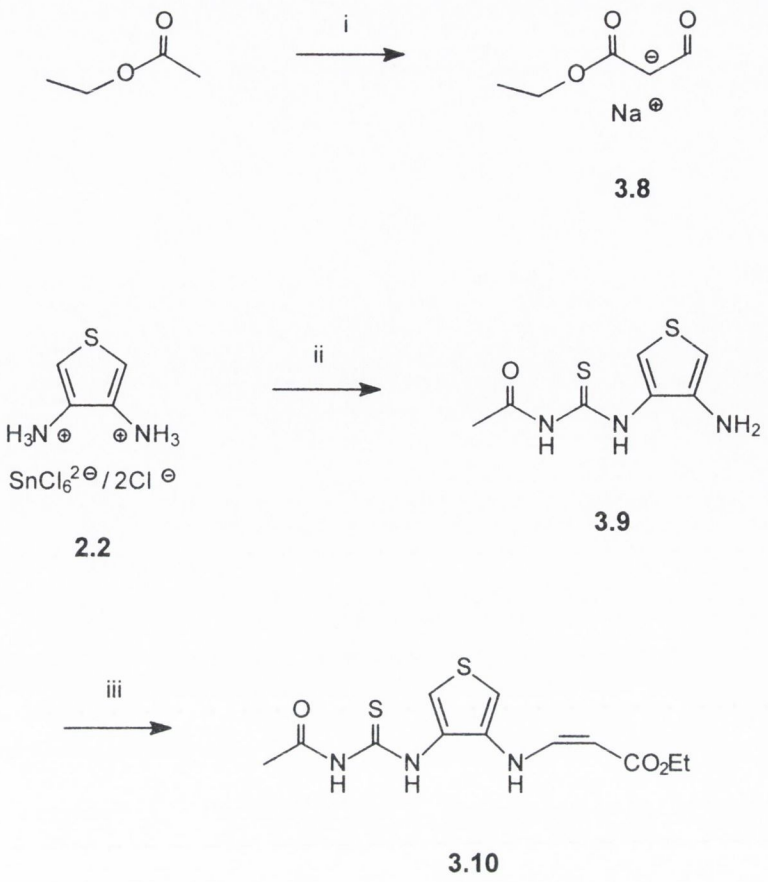
of the enamine double bond, it was thought that the enamine, being in equilibrium with the imine **3.10b** as in **Figure 3.1**, might not be as readily available for attack. The elegance of this strategy, giving the 5-ethoxycarbonylpyrrolothienothiazole in six steps from the dibromo dinitro thiophene **2.1** if it were to work, prompted its pursuit.



3.10 **3.10b**

Figure 3.1 Enamine-Imine tautomerisation of **3.10-3.10b**

The synthesis of sodium ethyl formylacetate **3.8** and the enamine **3.10** is outlined in **Scheme 3.3**. The ethyl formylacetate sodium salt **3.8** was synthesised by the method of Robins.[3.4] The well known instability of formylacetate esters precluded the purification and isolation of the ester and instead it was manipulated as the salt. Liberation of the ethyl formylacetate, by the action of AcOH, in the presence of the 3-amino 4-thiourea **3.9** gave the 3-enamino 4-thiourea **3.10**.



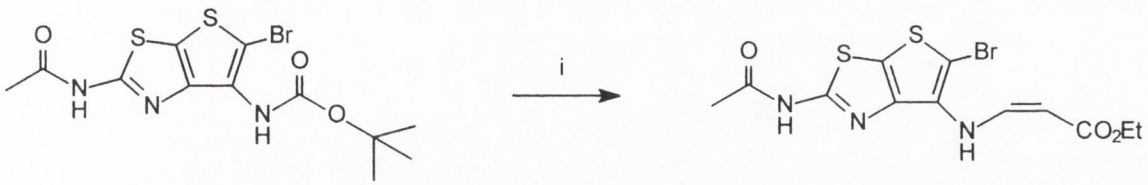
Scheme 3.3 R*eagents, conditions (and yields)*; i, Ethyl formate, sodium sand¶, toluene, room temperature, 24 h; ii, Et_3N, dry CH_2Cl_2, <10 °C, 15 minutes, then **3.2**, room temperature, 18 h (62% from **3.2**),; iii, **3.8**, AcOH, dry EtOH, reflux, 30 minutes, then room temperature, 15 h (45%).

¶ Sodium metal, toluene, vigorous stirring with reflux, then 0 °C with stirring.

The ^1H and ^{13}C NMR spectra indicated that **3.10** was present as the enamine tautomer only, while difference NOE experiments showed the double bond to be in the *cis* conformation. No signs of *cis/trans* isomerisation could be detected in a $CDCl_3$ solution of **3.10**, which was left stand at room temperature for 24 hours. This would indicate that enamine/imine interconversion was extremely slow, if occurring at all. It was therefore not surprising that attempts to cyclise **3.10** by treating with bromine in AcOH resulted in a complex precipitate, the structure of which could not be determined.

The failure of **3.10** to cyclise indicated that it would be necessary to form the thienothiazole bicycle before the enamine is introduced. This would require the inclusion of amino protection/deprotection steps in the synthetic strategy. As shown in **Scheme 3.4**, the 5-bromo 6-^{t}butylcarbamate **3.6** was deprotected with trifluoroacetic acid. Treating the ammonium trifluoroacetate salt with ethyl formylacetate gave a poor yield of

the *cis*-enamine **3.11**. The stereochemistry was assigned by comparision of the olefinic proton ^{1}H NMR coupling constants with those of **3.10**.



3.6 **3.11**

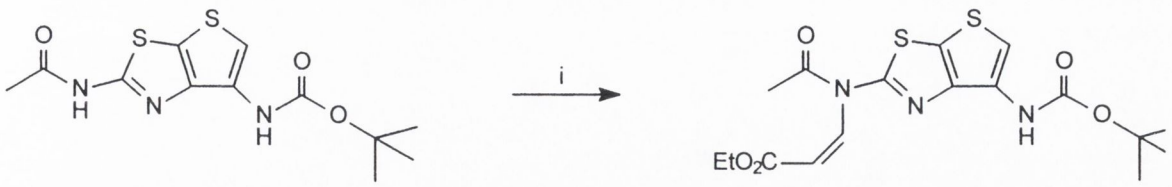
Scheme 3.4 *Reagents, conditions (and yields)*; i, Trifluoroacetic acid, room temperature, 30 minutes, then **3.8**, AcOH, dry EtOH, room temperature, 24 h (12%).

Besides the poor yields, this reaction also suffered from poor reproducibility. The reaction often failed on small scale and invariably gave no products when run on a larger scale (>50 mg). This unreliability prompted us to explore other means of forming enamines. The reaction of primary amines with alkynes to afford enamines is well documented. Heindel and co-workers[3.5] reported the reaction of anilines with methyl propiolate in methanol at room temperature. Employing this procedure, the 6-amine **3.5** failed to react with ethyl propiolate at either room temperature or at reflux. Similarly, deprotecting **3.6** with TFA, washing with aqueous base to liberate the amine and subsequent treatment with ethyl propiolate failed to give any product.

Nucleophilic substitution of **3.5** on the α,β-unsaturated ester, ethyl iodoacrylate, was tried next. There was no noticeable reaction in CH_3CN, DMF, MeOH or $CHCl_3$ with either ^{i}Pr$_2$EtN or DBU present. The use of powdered K_2CO_3 in refluxing acetone did however promote some reaction and analysis by TLC showed a large number of spots. Unfortunately, the number of products generated in this reaction suggested that while these conditions may promote alkylation of the 6-amino group, the 2-acetamido group was also being alkylated.

The suitability of these K_2CO_3/acetone conditions for amide alkylation was demonstrated by the reaction of the 2-acetamido 6-^{t}butylcarbamate **3.4** with ethyl iodoacrylate as shown in **Scheme 3.5**. The product of this reaction while being a single spot by TLC, looked by ^{1}H NMR to be a 2:1 mixture of two enamine compounds. Difference NOE experiments indicated that the major component of the mixture was the alkylated amide. The coupling constants of the olefinic protons (9.0 Hz) of the major component indicated that it was the *cis* isomer **3.12**. The minor component gave no

response in NOE experiments, suggesting that it was the product of alkylation at the carbamate nitrogen.

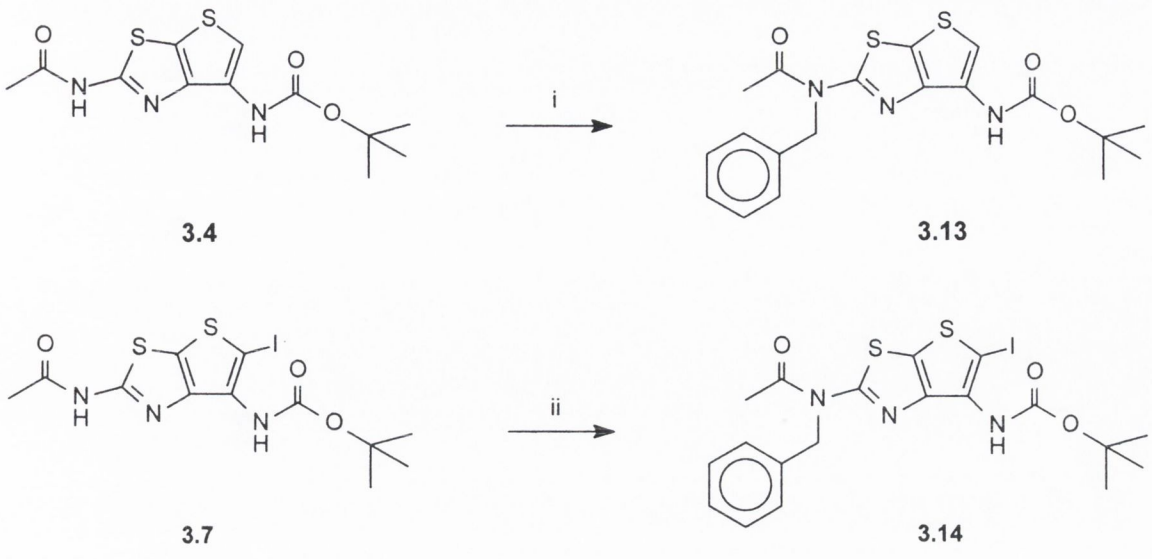


3.4 **3.12**

Scheme 3.5 *Reagents, conditions (and yields)*; i, Ethyl iodoacrylate, powdered K_2CO_3, dry acetone reflux, 48 h (38%).

The synthesis of the enamine **3.12** had demonstrated that alkylation of 2-acetamido 6-carbamates occurs preferentially at the acetamido nitrogen. Therefore, it would be necessary to protect the amide NH before the carbamate NH could be derivatised. The benzyl group was initially chosen for this purpose, owing to a number of reasons. Firstly, it was felt that the two general methods of removing the benzyl group would be compatible with our system. Acid hydrolysis could be used to remove the ^{t}Boc group at the same time as the benzyl group, while the alternative method, using dissolving metal reduction or hydrogenation should pose no threat to the integrity of the electron-rich tricyclic system. Secondly, the benzyl CH_2 would allow verification of the regiospecificity of the protection, by difference NOE experiments.

Scheme 3.6 shows the benzyl protection of 2-acetamido 6-^{t}butylcarbamate **3.4** and 2-acetamido 5-iodo 6-^{t}butylcarbamate **3.7**. Refluxing **3.4** with benzyl bromide and K_2CO_3 in acetone gave the protected amide **3.13** as a single product, in good yield. The configuration of **3.13** was confirmed by difference NOE ^{1}H NMR experiments. **3.14** was achieved by a similar procedure, this time using DMF as solvent due to the poor solubility of the starting material in acetone. As for **3.13,** the point of attachment of the benzyl group was confirmed by difference NOE experiments.



3.4 **3.13**

3.7 **3.14**

Scheme 3.6 *Reagents, conditions (and yields)*; i, Benzyl bromide, powdered K_2CO_3, dry acetone, 50 °C, 22 h (85%), ii, , Benzyl bromide, powdered K_2CO_3, dry DMF, 64 °C, 24 h (63%).

Attempts to produce the 6-enamino derivative of **3.13** by reaction with ethyl iodoacrylate and K_2CO_3 in acetone gave no reaction products. Using the more vigorous conditions of sodium hydride in dry DMF also failed to affect the starting material. Similarly, **3.14** failed to react with iodoacrylate and K_2CO_3 in acetone.

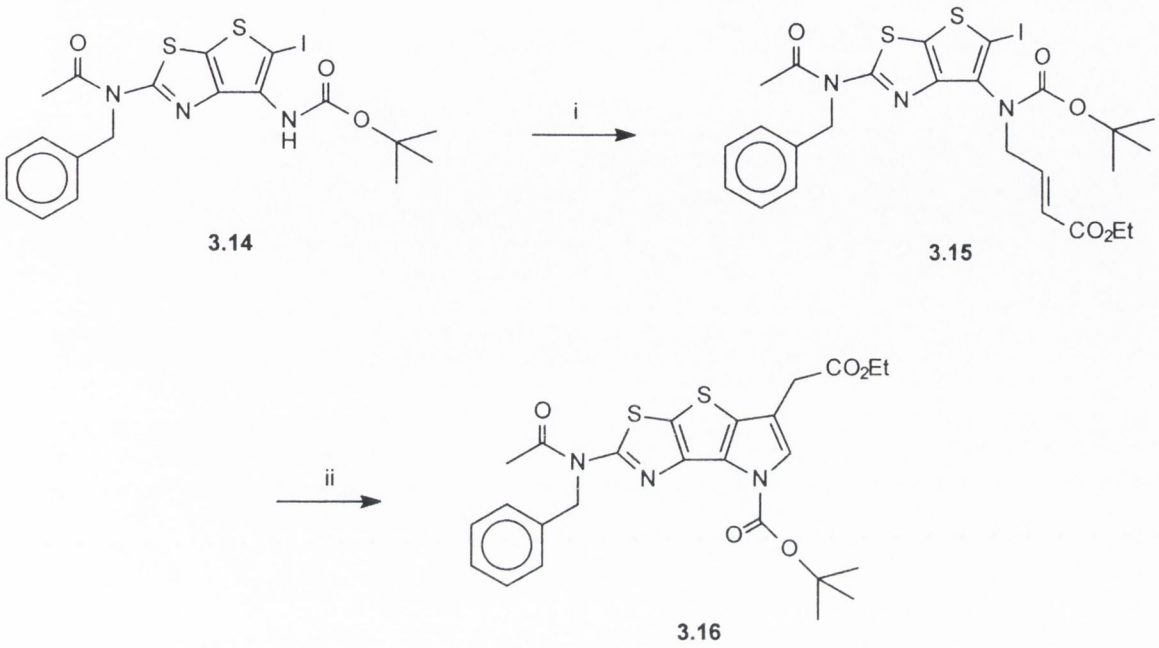
Due to the difficulties in converting **3.13** and **3.14** to their enamine derivatives it was decided to suspend the development of the 5-ethoxycarbonyl tricyclic system and instead focus on the 5-ethoxycarbonylmethyl tricycle.

3.4 The Synthesis of 5-Ethoxycarbonylmethyl-pyrrolo[2',3',:4,5]thieno [3,2-*d*]thiazole 3.16

Wensbo and co-workers[3.6] published a procedure for the synthesis of indole-3-acetic acids and heteroanalogues including thienopyrrole systems. Allylation of *N*-^{t}Boc protected *o*-iodothienyl amines with ethyl 4-bromocrotonate, followed by palladium catalysed ring closure in a one-pot reaction, yielded *N*-^{t}Boc protected thienopyrroles.

Scheme 3.7 shows the steps in the conversion of **3.14** to the tricycle **3.16** employing similar methodology. **3.14** was originally allylated by the method of Wensbo, using ethyl 4-bromocrotonate and Cs_2CO_3 in dry DMF at 50-60 °C. This method was time consuming, failing to go to completion in 2 days, and afforded the product in unremarkable yield of 55%. Switching to sodium hydride as the base gave complete

conversion to **3.15** after 2 hours at room temperature although it was necessary to make a further addition of 0.5 equivalents of allylating agent at 1 hour.



Scheme 3.7 *Reagents, conditions (and yields)*; i, Sodium hydride, dry DMF, room temperature, 15 minutes then ethyl 4-bromocrotonate, room temperature, 2 h (60%); ii, $Pd(OAc)_2$, $P(Ph)_3$, $^iPr_2EtN$, dry CH_3CN, 65 °C, 21 h (69%).

The palladium catalysed cyclisation of **3.15** was first carried out following the method of Wensbo. Treating **3.15** with $Pd(OAc)_2$, K_2CO_3 or Cs_2CO_3 and $P(Ph)_3$ in DMF at 65 °C gave poor yields (~20%) of **3.16**. A number of other experimental conditions were investigated and optimum conditions were found to be $Pd(OAc)_2$, $P(Ph)_3$ and $^iPr_2EtN$ in dry CH_3CN at 65 °C. In order to confirm its structure, **3.16** was recrystallised from CH_2Cl_2-petroleum ether and the x-ray structure solved as shown in **Figure 3.2**.

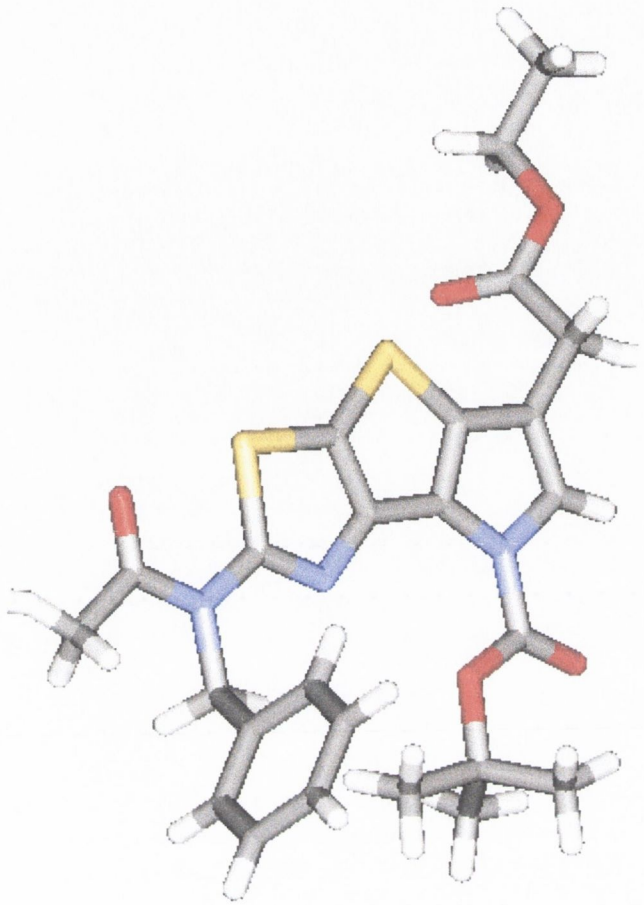


Figure 3.2 X-ray crystal structure of **3.16**.

3.4.1 Deprotection of 5-ethoxycarbonylmethyl-pyrrolothienothiazole 3.16

With the tricyclic skeleton formed, the only steps remaining were the removal of the ^{t}Boc and benzyl protecting groups. It was originally hoped that acid hydrolysis would remove both protecting groups simultaneously. Treating **3.16** with either trifluoroacetic acid/CH_2Cl_2 or formic acid[3.7] at room temperature afforded brown solids. The ^{1}H NMR spectra of these crude solids suggested that the ^{t}Boc group had been removed whilst the benzyl group was left intact. Heating the reaction mixtures to 60 °C merely gave unassignable (by ^{1}H NMR) decomposition products after several days.

Two other methods[3.8] of debenzylation by acid hydrolysis were also examined. Stirring **3.16** in 48% HBr at room temperature for 1 hour gave a single spot by TLC analysis. This product appears to be the result of the hydrolysis of the carbamate and ester, however the benzyl group was still intact, its signals being still observable by ^{1}H NMR. Treating **3.16** with phenol and orthophosphoric acid at room temperature for 18 hours gave an identical result.

The second standard method of debenzylation of amides is by reduction. The use of sodium metal in liquid ammonia[3.9] or lithium metal and dibenzyl ether in liquid ammonia[3.10] afforded no deprotection products in either case. TLC analysis of the resultant residues failed to reveal any UV or ninhydrin active products.

Debenzylation of **3.15** *via* palladium catalysed hydrogenation was also attempted. Transfer hydrogenation catalysed by either 5% Pd-C or $Pd(OH)_2$, and using cyclohexene as a hydrogen source in MeOH at 65 °C proved ineffective. Hydrogenation of **3.15** under a hydrogen atmosphere and catalysed by $Pd(OH)_2$ in AcOH also left the starting material unchanged.

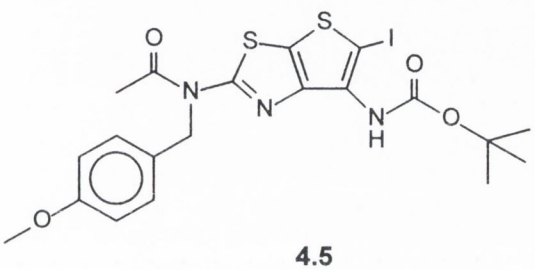
The unsuccessful attempts to remove the benzyl protecting group encouraged us to seek alternative means of amide protection. These efforts are recounted in the chapter 4.

Chapter Four

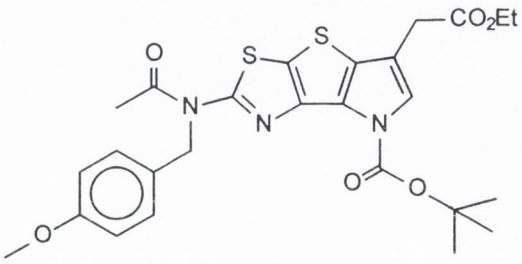
Alternative Protecting Groups for 3.7

4.1 Preamble

The introduction of a variety of protecting groups to **3.7** will be discussed in this chapter. Having been unsuccessful in removing the benzyl group from either **3.15** or **3.16**, it was desirable to find a protecting group, which would be more labile. These efforts culminated in the synthesis of **4.5**, which was then carried forward to afford the ^{t}Boc *p*-methoxybenzyl protected tricycle **4.7**.



4.5

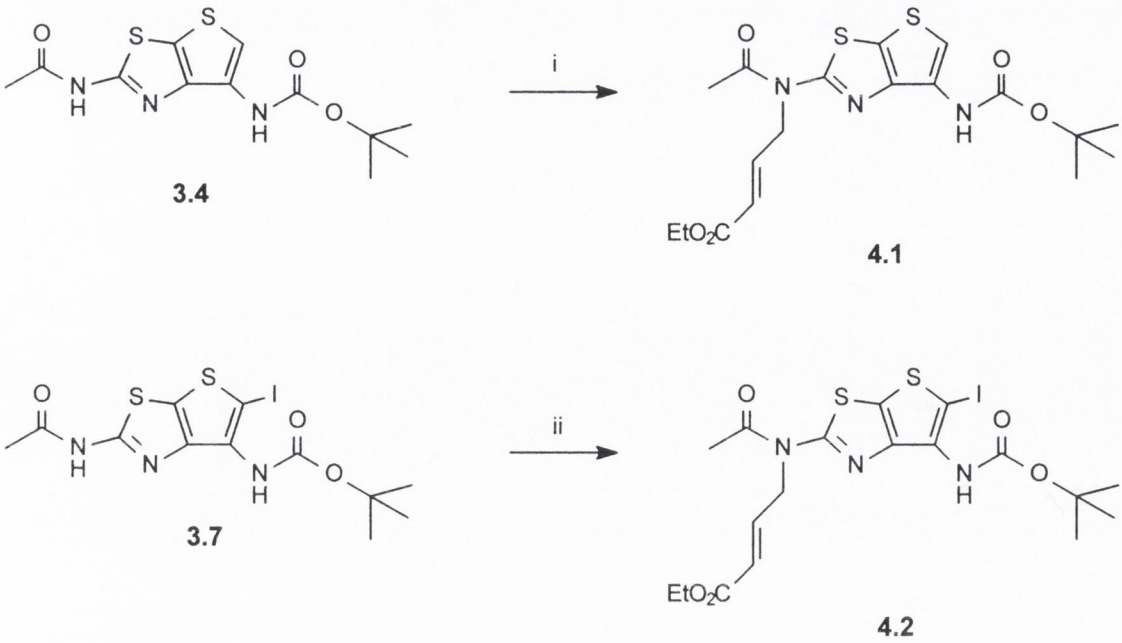


4.7

Section 4.2 describes the attempts to circumvent the need for amide protection by direct allylation at the carbamate nitrogen of **3.4** and **3.7** with ethyl 4-bromocrotonate. The efforts to introduce a variety of protecting groups in place of the benzyl group are outlined in section 4.3. The synthesis of **4.7** and the attempts to remove the *p*-methoxybenzyl from both **4.7** and **4.5** are covered in section 4.4

4.2 Allylation of 2-Acetamido 6-^{t}Butylcarbamates 3.4 and 3.7

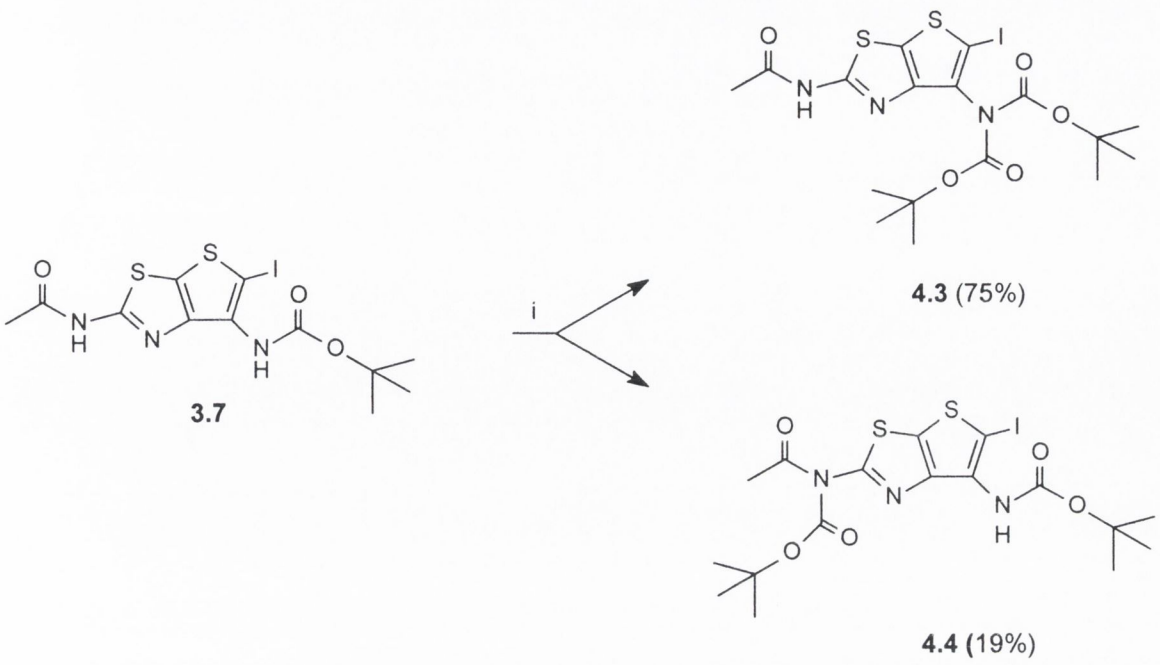
Despite having shown that alkylation of the 2-acetamido 6-carbamates **3.4** and **3.7** occurs preferentially at the acetamido nitrogen (see section 3.3), it was considered worthwhile to attempt regiospecific allylation of **3.4** and **3.7**. If allylation were found to occur at the carbamate nitrogen, the need for acetamide protection would no longer arise. As shown in **Scheme 4.1** treating **3.4** or **3.7** with ethyl 4-bromocrotonate and K_2CO_3 in CH_3CN gave the undesired regioisomers **4.1** and **4.2**. The allylation of **3.7** using the same conditions, only in dry DMF, again gave the unwanted isomer **4.2**, in 45% yield. As in previous cases, the position of the allyl group was determined by the use of difference NOE ^{1}H NMR experiments.



Scheme 4.1 *Reagents, conditions (and yields)*; i, Ethyl 4-bromocrotonate, K_2CO_3, dry CH_3CN, 60 °C, 17 h (26%); ii, Ethyl 4-bromocrotonate, K_2CO_3, dry CH_3CN, 60 °C, 17 h (64%).

4.3 Alternative Methods of Protecting 2-Acetamido 6-^{t}Butylcarbamate 3.7

The results of the allylation of **3.4** and **3.7** confirmed that an amide blocking/protecting group was necessary in the synthetic strategy. The reaction of **3.7** with a number of candidates was investigated for this purpose. Initial attempts to introduce a second ^{t}Boc group by using di-*tert*-butyl dicarbonate and K_2CO_3 in acetone, failed. It was later found however that using DMAP and $^i Pr_2EtN$ in CH_3CN at 60 °C gave a mixture of two products as shown in **Scheme 4.2**.



Scheme 4.2 *Reagents and conditions*; i, Di-*tert*-butyl dicarbonate, DMAP, $^{i}Pr_2EtN$, dry CH_3CN, 60 °C, 135 minutes.

NOE studies of these compounds were non-productive, as it was not possible to see any interaction between the methyl of the acetamido group and the ^{t}Boc group. The assignment of **4.3** and **4.4** was originally based solely on ^{1}H NMR spectra. In the ^{1}H spectrum of the major product there are just three singlets at 1.45, 2.19, and 9.65 ppm which integrate for 18, 3 and 1 protons respectively. This spectrum is consistent with two equivalent ^{t}Boc groups as would be expected with the 2-acetamido 6-bis(^{t}Boc)amine **4.3**. The spectrum of the minor product also supports these assignments. This spectrum also contains three singlets, at 1.43, 1.58 and 7.9 ppm, which integrate for 12, 9 and 1 protons respectively. This is consistent with the 2-(*N*-acetyl-*N*-^{t}Boc)amino 6-(*N*-^{t}Boc)amine **4.4**, for which one could expect the ^{t}Boc groups to be non-equivalent. The reason for the apparent upfield shift in the acetyl CH_3 from ~2.2 to 1.43 ppm (under one of the ^{t}Boc signals) is however unknown.

After considerable effort, x-ray quality crystals of **4.3** were obtained from $CHCl_3$-MeOH and the crystal structure solved (see **Figure 4.1**). This confirmed the presence of two ^{t}Boc groups on the 6-amino nitrogen. From this it would appear that ^{t}Boc protection of **3.7** results in preferential protection at the 6-carbamate nitrogen to give the undesired bis-protected compound **4.3**, with the desired **4.4** formed only in low yield.

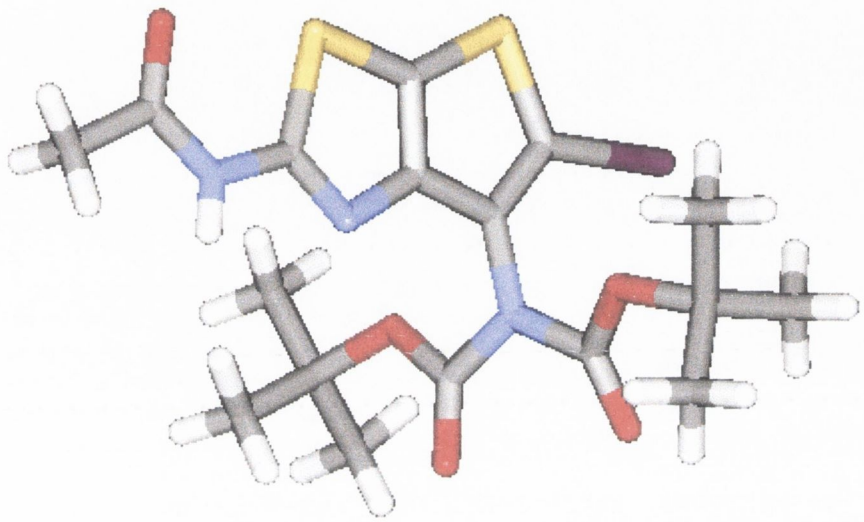


Figure 4.1 X-ray structure of bis(^{t}Boc) protected 4-amine **4.3**. MeOH solvent molecule removed for clarity

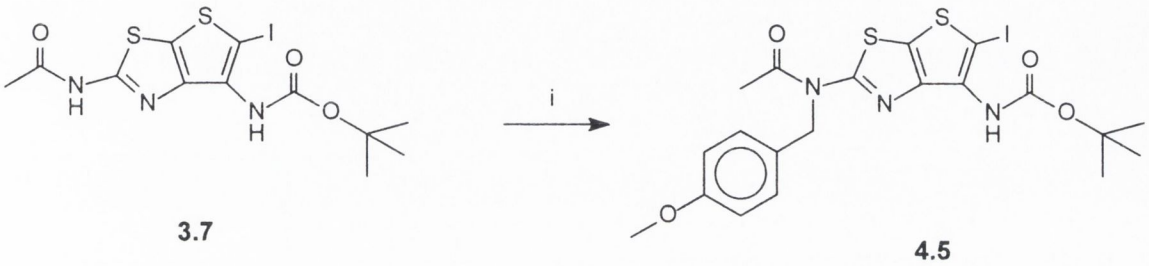
Protection with a ^{t}butyldimethylsilyl group was also considered. It was found that using ^{t}butyldimethylsilyl chloride and K_2CO_3 with or without a catalytic amount of imidazole in dry DMF at 50 °C left **3.7** unaffected.

The alkylation of the amide nitrogen with benzyl derivatives was examined also. Each of these groups, trityl, benzhydryl and *p*-methoxybenzyl, would be expected to be more susceptible to acid hydrolysis than the simple benzyl group.

Treating **3.7** with trityl bromide and K_2CO_3 in dry DMF or CH_3CN gave no reaction products. This is probably at least partly due to steric hindrance; the three bulky phenyl rings preventing the amide nitrogen attacking the substrate. The benzhydryl group was likewise unreactive. Heating **3.7** with diphenylmethyl bromide and K_2CO_3 in dry

DMF or Cs_2CO_3 in dry CH_3CN left the starting material unaffected. Similarly, using sodium hydride in dry DMF gave no reaction products.

The reaction of **3.7** with *p*-methoxybenzyl chloride did proceed successfully, as shown in **Scheme 4.3**. The location of the methoxybenzyl group was confirmed by ^1H NMR difference NOE experiments.



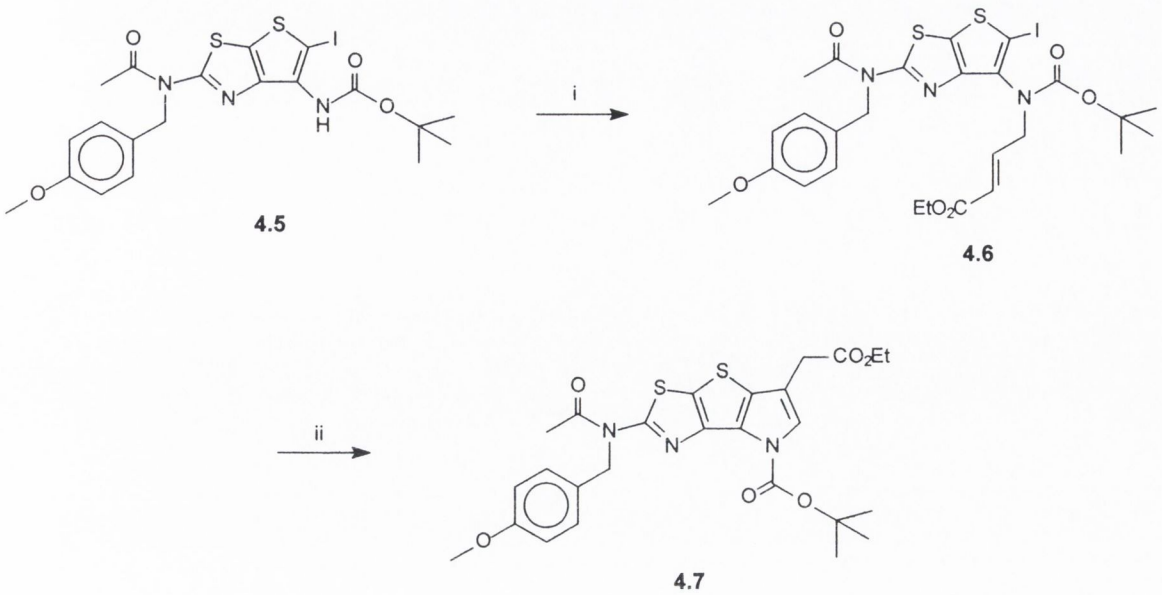
3.7 **4.5**

Scheme 4.3 *Reagents, conditions (and yields);* i, *p*-Methoxybenzyl chloride, powdered K_2CO_3, dry DMF, 65 °C, 3½ h (67%).

The derivatisation and subsequent cyclisation of **4.5** plus the removal of the protecting groups is covered in section 4.4.

4.4 The Synthesis of *p*-Methoxybenzyl Protected Tricycle 4.7 and Attempts to Remove the *p*-Methoxybenzyl Group from 4.5 and 4.7

The allylation and subsequent cyclisation of **4.5** is shown in **Scheme 4.4**. The yield for the allylation step is considerably lower than the analogous reaction for the benzyl protected amide **3.14**, however time constraints had prohibited the optimisation of the reaction. Palladium catalysed cyclisation of **4.6** gave the *p*-methoxybenzyl protected tricycle **4.7** in good yield.



4.5 4.6 4.7

Scheme 4.4 *Reagents, conditions (and yields)*; i, Sodium hydride, dry DMF, room temperature, 15 minutes then ethyl 4-bromocrotonate, room temperature, 100 minutes (28%); ii, $Pd(OAc)_2$, $P(Ph)_3$, $^iPr_2EtN$, dry CH_3CN, 72 °C, 29 h (78%).

With the ^{t}Boc methoxybenzyl protected tricycle **4.7** achieved, the removal of both protecting groups was our next objective. Smith and co-workers[4.1] reported the removal of the *p*-methoxybenzyl moiety to afford a cyclic amide in good yield by the action of neat trifluoroacetic acid at room temperature for 2½ hours. Subjecting **4.7** to the same reaction conditions resulted in the loss of the ^{t}Boc group within 90 minutes (by ^{1}H NMR), however prolonged exposure to the acid (18 hours) left the *p*-methoxybenzyl group unaffected.

Due to a lack of time it was decided to study the removal of the methoxybenzyl from the precursor **4.5**, hoping to then apply a successful method to **4.7**. Standing **4.5** in trifluoroacetic acid at room temperature for 1 day removed the ^{t}Boc group only, heating the same solution to 60 °C for 7 days resulted in decomposition products, which were unassignable by ^{1}H NMR. Parallel treatment of **4.5** in formic acid gave similar results.

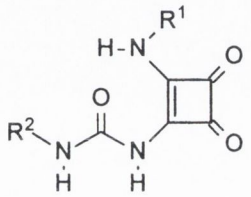
Finally removal of the protecting group was attempted by oxidation. Kronenthal and co-workers[4.2] reported the oxidative dearylation of *p*-methoxybenzyl protected Azetidinones, by the action of ceric ammonium nitrate. Employing this procedure however resulted in a vivid pink precipitate, which was indecipherable by ^{1}H NMR. Presumably the electron rich **4.5** was not stable under the oxidative conditions employed.

Chapter Five

Studies of the *N*-Carbamoyl Squaramide System

5.1 Preamble

This chapter is an account of the work carried out on the *N*-carbamoyl squaramides of the type **1.44**. We intended synthesising a selection of chloroform soluble *N*-carbamoyl squaramide compounds and then investigating the nucleoside recognition properties of these by using ^{1}H NMR techniques. In the event, the self-association of these molecules proved to be their most significant property.



1.44

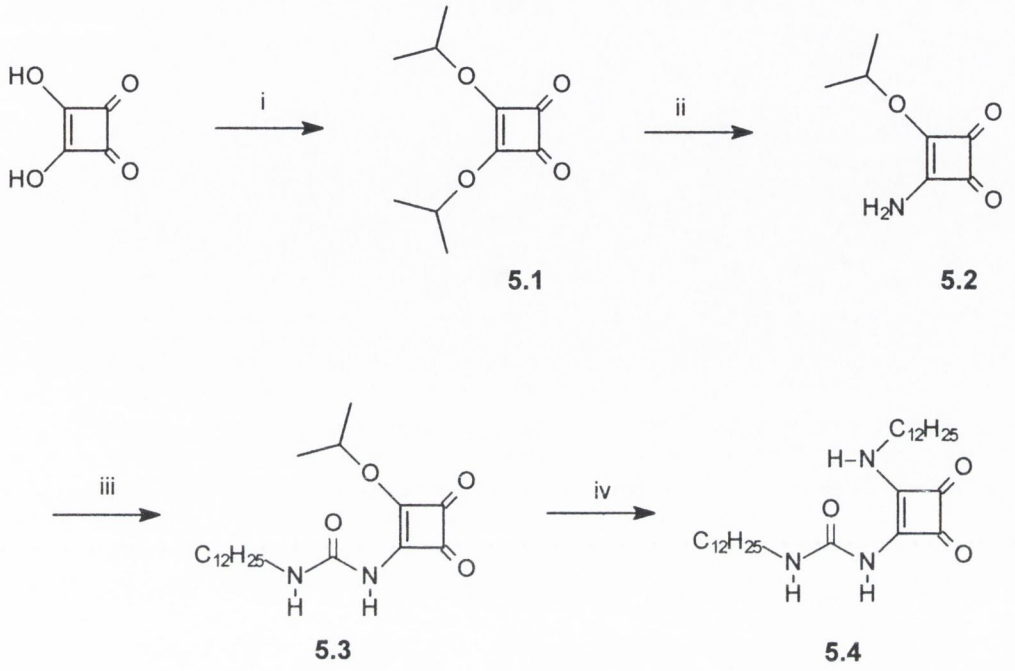
The synthesis of the squaramide **5.7** and the initial binding studies with triacetyl cytidine in $CDCl_3$, are covered in section 5.2. Section 5.3 deals with the synthesis of the squaramides **5.9**, **5.12** and **5.13**, and the study of the dimerisation of these, in the presence of deuteroacetonitrile. The binding of cytidine by squaramides is returned to in section 5.4, where the early binding studies are re-examined. Binding studies in 5% (v/v) $CD_3CN/CDCl_3$ are also covered in this section.

5.2 Initial Studies of Amino Urea 5.5 and 5.7

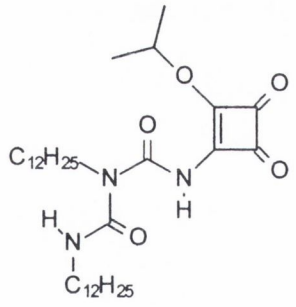
The synthesis of the dodecylcarbamoyl squaramide **5.5** is outlined in **Scheme 5.1**. The diisopropyl squarate **5.1** was obtained by a modification of a method of Liebeskind,[5.1] using toluene instead of benzene and catalysing the reaction with *p*-TsOH. The resultant diester was aminated based on a procedure of Liebeskind,[5.2] to afford the ester squaramide **5.2**.

The ester carbamoyl squaramide **5.3** was achieved by treatment of **5.2** with dodecylisocyanate and *N,N*-diisopropylethylamine. Initially, **5.3** was found to further react with the isocyanate to give a side product which by ^{1}H and ^{13}C NMR appears to contain two dodecyl chains. This side product is tentatively assigned the structure **5.5**. To avoid this side product and the resultant low product yield, an excess of **5.2** was used which was readily recovered at the workup stage. Treatment of **5.3** with an excess of

dodecylamine gave the carbamoyl squaramide **5.4**. Unfortunately **5.4** was found to be insoluble in chloroform and so was not further studied.



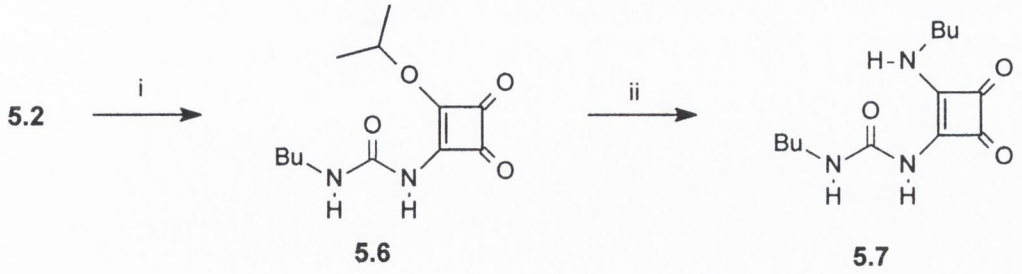
Scheme 5.1 *Reagents, conditions (and yields)*: i, 2-propanol, *p*-TsOH, toluene, reflux, 48 h; ii, ammonia gas, MeOH-CH_2Cl_2, room temperature, 20 minutes (75%); iii, dodecylisocyanate, $^iPr_2EtN$, dry CH_3CN, room temperature, 18 h (69%); iv, dodecylamine, CH_2Cl_2, room temperature, 5 days (18%)



5.5

Despite having found **5.3** to be quite soluble in $CHCl_3$, parallel work on the butylcarbamoyl squaramide **5.7** proved more fruitful. The synthesis of **5.7** is shown in **Scheme 5.2**. The synthesis proceeded in a similar fashion to that of the dodecyl derivatives. An excess of the ester squaramide **5.2** is again necessary in order to avoid the reaction of ester carbamoyl squaramide **5.6** with the isocyanate. Treatment of **5.6** with

butylamine afforded the butylcarbamoyl squaramide **5.7**, which was a $CHCl_3$-soluble, solid.

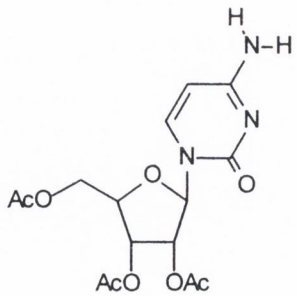


Scheme 5.2 *Reagents, conditions (and yields)*; i, butylisocyanate, Et_3N, dry CH_3CN, room temperature, 18 h (64%); ii, Butylamine, CH_2Cl_2, room temperature, 20 h (87%)

5.2.1 Initial investigation of nucleoside recognition properties of 5.7

It was found that when solutions of **5.7** in $CDCl_3$ were diluted from 243 mM to 9 mM and 0.4 mM to 0.1 mM the ^1H NMR spectra showed little change. The C-3 NH resonances were shifted slightly (δ 10.51 $\rightarrow$ δ 10.53 and δ 10.44 $\rightarrow$ δ 10.37 respectively), as were the carbamoyl NH resonances (δ 6.63 $\rightarrow$ δ 6.60 and δ 6.57 $\rightarrow$ δ 6.51 respectively) and C-4 NH resonances (δ 8.33 $\rightarrow$ δ 8.34 and δ 8.31 $\rightarrow$ δ 8.30 respectively). The slightness of these resonance shifts coupled with the fact that the NH peaks are poorly resolved at low concentrations led us to assume that **5.7** was not self-associating.

Encouraged by this result, we turned towards investigating the binding of triacetyl cytidine **5.8,** by **5.7**, using NMR titration techniques.



5.8

A concentration study of **5.8** in $CDCl_3$ covering the range of concentrations from 41.9 mM to 11.2 mM was carried out. The H-5 signal shifted downfield by 0.09 ppm while the H-6 shifted by 0.05 ppm over the course of the experiment. Both chemical shift

data set were fitted to dimerisation models, (see **Figure 5.1a** and **5.1b**) yielding a dimerisation constant of 2 (±2) M^{-1} (calculated using Hostest version 5.0[5.3]).

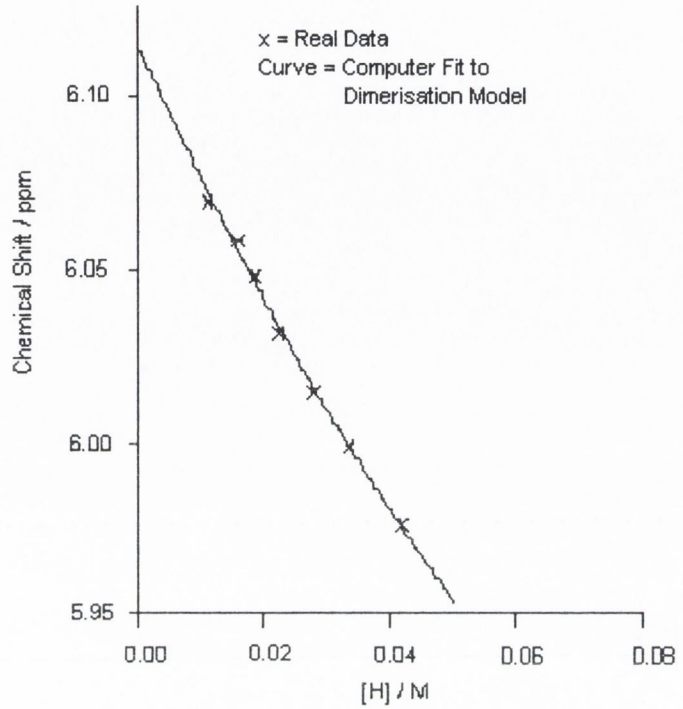


Figure 5.1a Chemical shift of the H-5 of **5.8** in $CDCl_3$ versus its concentration

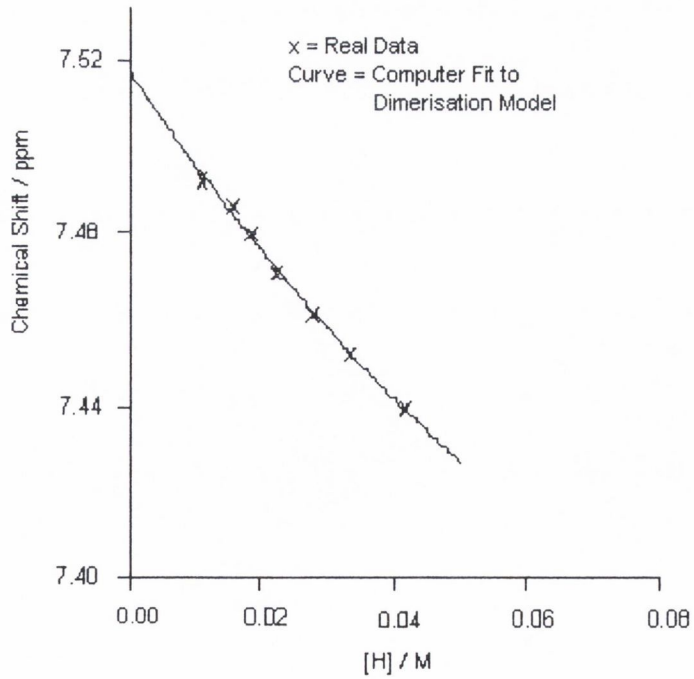


Figure 5.1b Chemical shift of the H-6 of **5.8** in $CDCl_3$ versus its concentration

A binding study by NMR titration experiment, of **5.8** as host and squaramide **5.7** as guest in $CDCl_3$ produced a disappointing result. A graph for the change in shift of the H-5 versus the concentration of guest, while being a smooth curve, failed to fit a 1:1 host-guest model. This behaviour was observed for host initial concentrations of 1.9 mM (see **Figure 5.2a**) and 4.1 mM (see **Figure 5.2b**). The H-6 signals shifted in a similar manner but to a smaller extent.

The reverse titration of **5.7** as host and **5.8** as guest in $CDCl_3$ also produced a discouraging result. The carbamoyl NH signal shifted downfield by over 2.5 ppm during the course of the experiment. Curiously, the C-3 NH signal initially shifted upfield from 10.52 to 10.21 ppm, but after adding approximately 4 equivalents of guest the signal started to shift downfield to finish at 10.24 ppm. As **Figure 5.3a** and **5.3b** show, the graphs of the NH chemical shifts versus guest concentration could not be fitted to a 1:1 binding model.

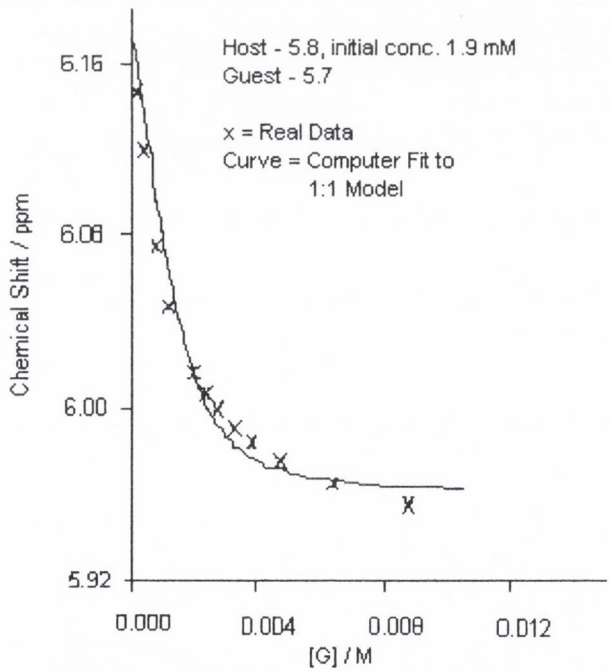


Figure 5.2a Chemical shift of the H-5 of **5.8** in $CDCl_3$ versus the concentration of **5.7**. Initial concentration of **5.8** was 1.9 mM.

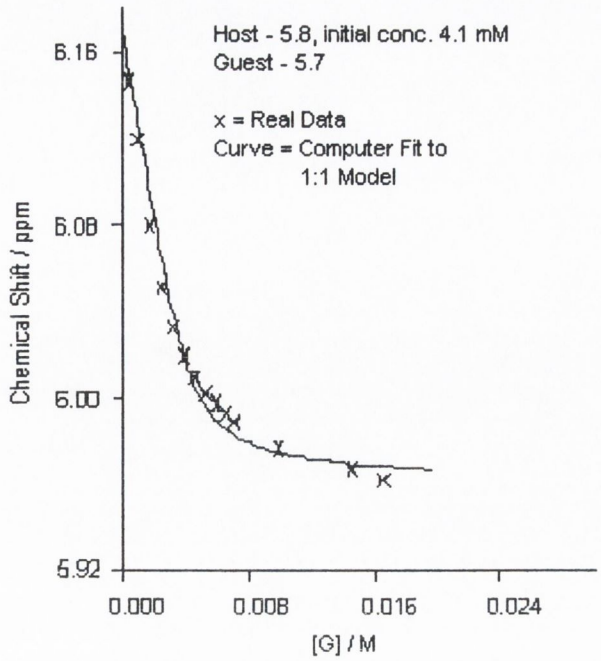


Figure 5.2b Chemical shift of the H-5 of **5.8** in $CDCl_3$ versus the concentration of **5.7**. Initial concentration of **5.8** was 4.1 mM.

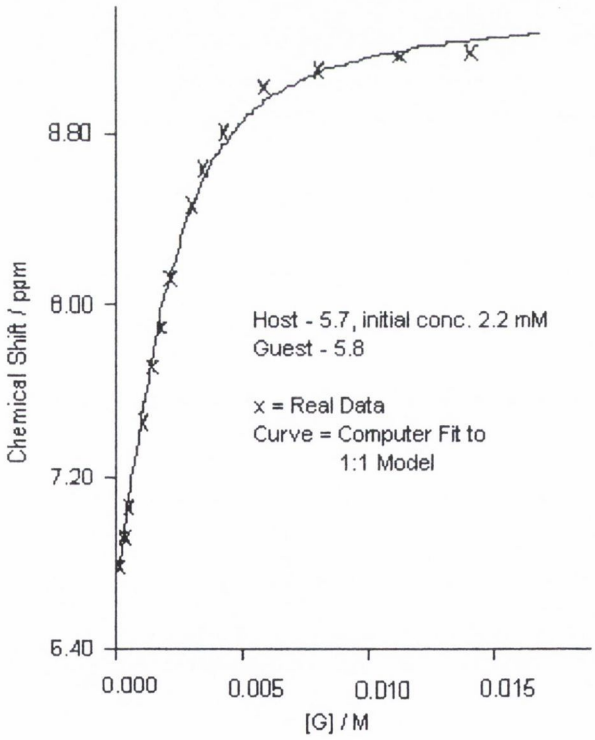


Figure 5.3a Chemical shift of the carbamoyl NH of **5.7** in $CDCl_3$ versus the concentration of **5.8**

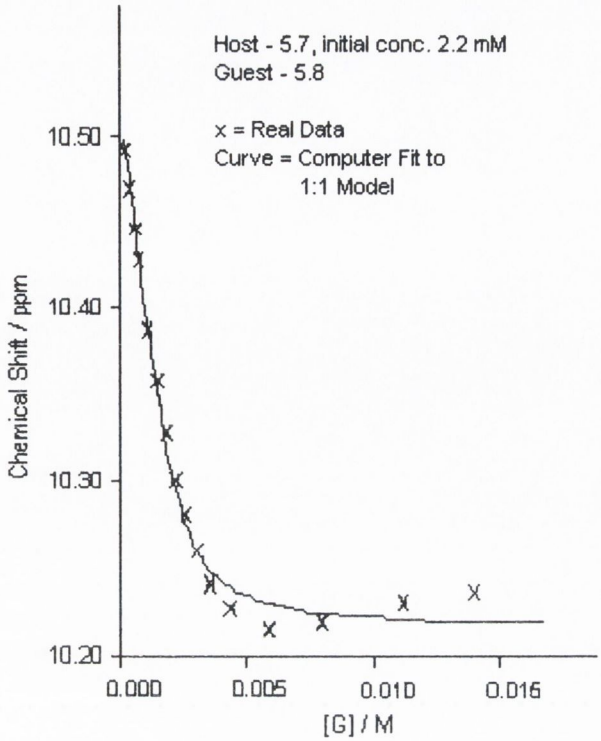
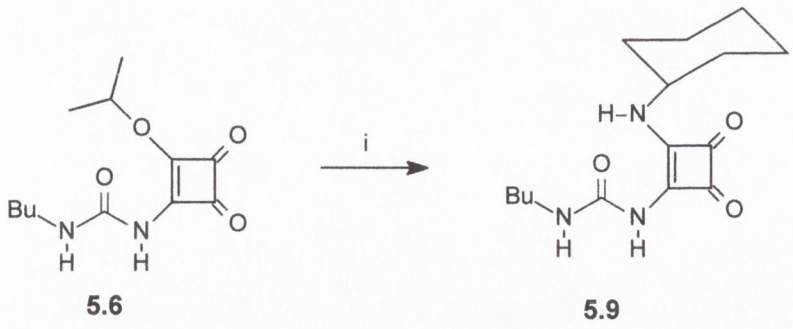


Figure 5.3b Chemical shift of the C-3 NH of **5.7** in $CDCl_3$ versus the concentration of **5.8**

5.3 Dimerisation Studies of *N*-Carbamoyl Squaramides

Whilst the results of the above binding studies could not be explained by a 1:1 binding model, it was apparent that there was some interaction between **5.7** and **5.8**. A possible reason for these unexpected results was that **5.7** forms dimers at the concentrations at which the experiments were carried out. Although no self-association had been detected in the earlier dilution experiments, the possibility of strong dimerisation could not be discounted. Adding weight to this argument, a FAB^+ mass spectrum of **5.8** revealed in addition to the monomer at m/z 268 (MH^+), a significant (~30%) signal for the dimer at m/z 535 (M_2H^+). Signals for higher aggregates M_nH^+ (n = 3-6) appeared at <3%.

Suspecting dimerisation, efforts were made to acquire a crystal structure of an *N*-carbamoyl squaramide. The cyclohexylamino squaramide **5.9** was synthesised as shown in **Scheme 5.3**. X-ray quality crystals of **5.9** were obtained and the crystal structure solved. As shown in **Figure 5.4**, **5.9** exists as a centrosymmetric dimer in the crystalline state, held together by 4 hydrogen bonds from the carbamoyl NH and the C-3 NH to the C-2 carbonyl oxygen. The carbamoyl squaramide units are essentially planar and exhibit the predicted intramolecular hydrogen bond



Scheme 5.3 *Reagents, conditions (and yields)*: i, Cyclohexylamine, CH_2Cl_2, room temperature, 20 h (82%)

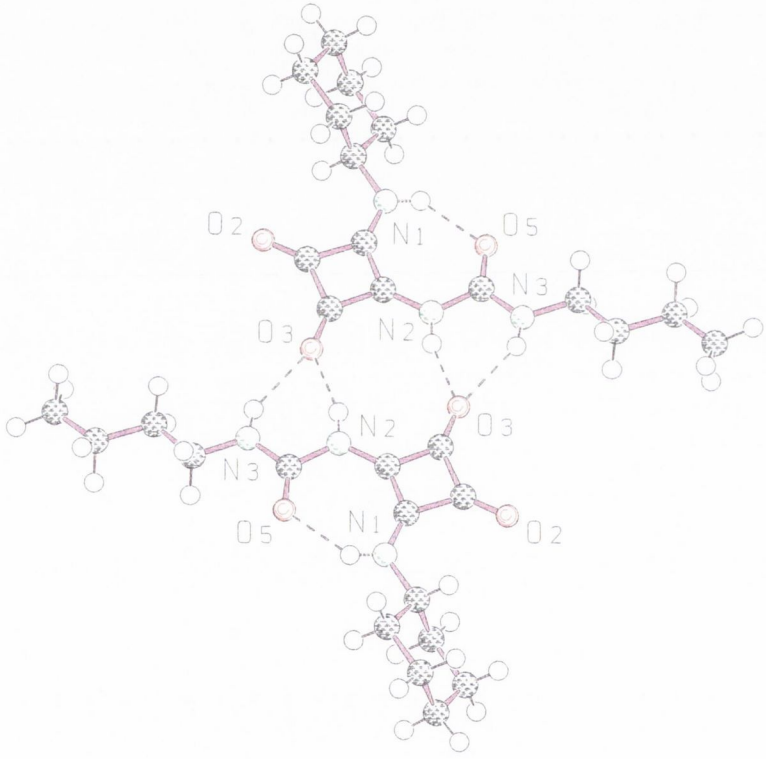


Figure 5.4 Crystal structure of **5.9**. Selected distances (Å) are N(2)H···O(3) 1.937, N(3)H···O(3) 2.108, N(1)H···O(5) 2.120.

Having shown that **5.9** dimerises in the solid state, we set about testing the hypothesis that strong dimerisation was taking place in $CDCl_3$. A concentration study of **5.9** in $CDCl_3$ covering the range of concentrations from 0.2 mM to 15.4 mM showed a small downfield shift in the C-3 NH signal (0.13 ppm), which on fitting to a dimerisation model (see **Figure 5.5**) yielded an association constant of 1.1×10^6 M^{-1}. The error analysis

by hostest, yielded an unrealistic standard deviation of ±53.3 x10^6 (this error would indicate a monomer chemical shift range of ≈ –5 to 27 ppm) and clearly indicates that the calculated association constant is unreliable. The carbamoyl NH signal shifted in a similar fashion but by only 0.05 ppm. A study over the concentration range from 0.6 mM to 30.1 mM produced qualitatively similar graphs, but it proved impossible to calculate an association constant from this data.

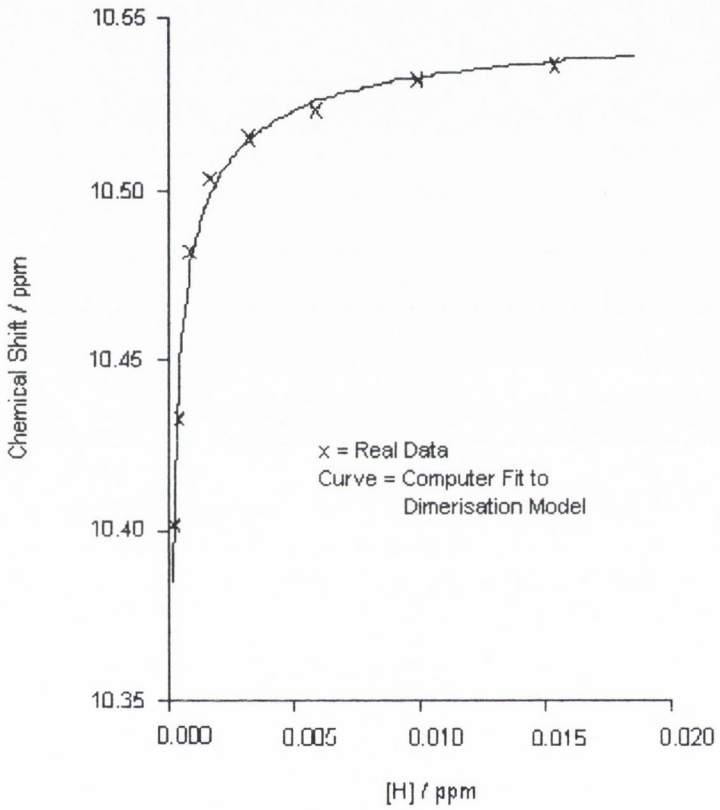


Figure 5.5 Chemical shift of the C-3 NH of **5.9** in $CDCl_3$ versus its concentration

While the above studies failed to produce association constants that are statistically reliable, they do suggest strong self-association of **5.9**. In order to probe this further, NMR studies in the presence of deuteroacetonitrile, a strong hydrogen bond acceptor solvent, were carried out. On adding 5% CD_3CN to a $CDCl_3$ solution of *N,N'*-dibutylurea, a non-dimerising system, the NH signal shifts downfield by 0.13 ppm from 4.21 to 4.34 ppm, presumably due to the formation of NH···N≡C hydrogen bonds. When a $CDCl_3$ solution of **5.9** (4.4 mM) was titrated with CD_3CN, the signals of all three NH protons moved upfield. The carbamoyl NH and C-4 NH moved by -0.67 and -0.70 ppm respectively, while the C-3 NH moved by over -2.1 ppm over the course of the

experiment (see **Figure 5.6**). The observed motions are conceivably the result of the disruption of the **5.9-5.9** dimer by competition for hydrogen bond donors by CD_3CN with resultant loss or weakening of the stronger NH··O=C hydrogen bonding.

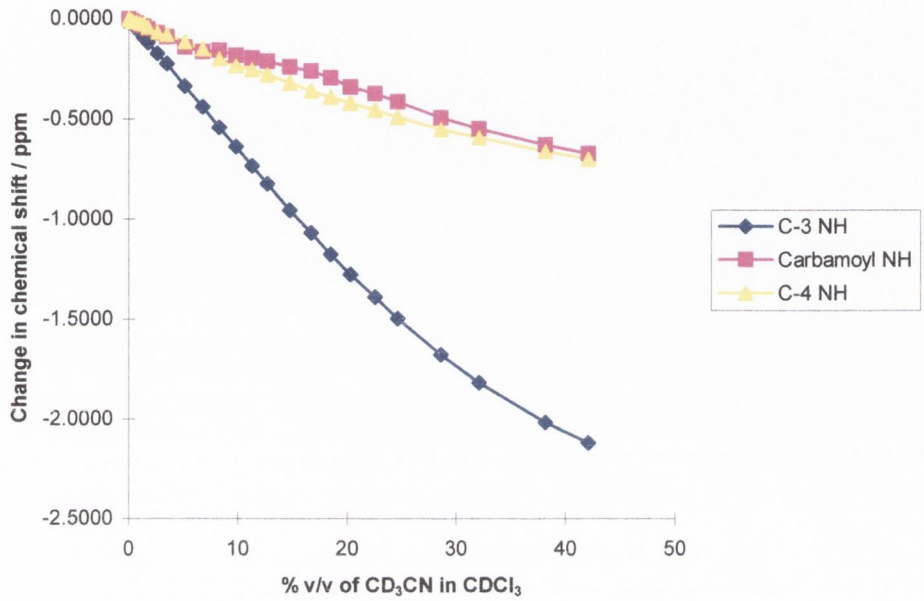


Figure 5.6 Change in chemical shift of the NH signals of **5.9** versus the percentage v/v of CD_3CN in $CDCl_3$

Quantitative ^1H NMR concentration studies of **5.7** and **5.9** were carried out in 5% (v/v) $CD_3CN/CDCl_3$ over the range of concentrations from 0.2 mM to 25 mM. The C-3 NH signals of both systems shifted downfield by 0.92 ppm over the course of the experiment and could be fitted to dimerisation models (see **Figure 5.7** and **5.8a**) to afford association constants of 6116 (±1207) M^{-1} and 5778 (±1057) M^{-1} respectively.

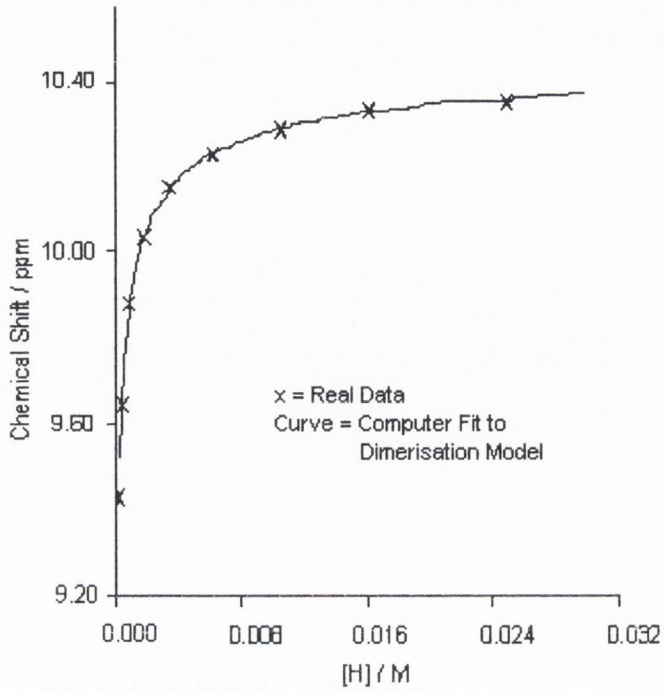


Figure 5.7 Chemical shift of the C-3 NH of **5.7** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

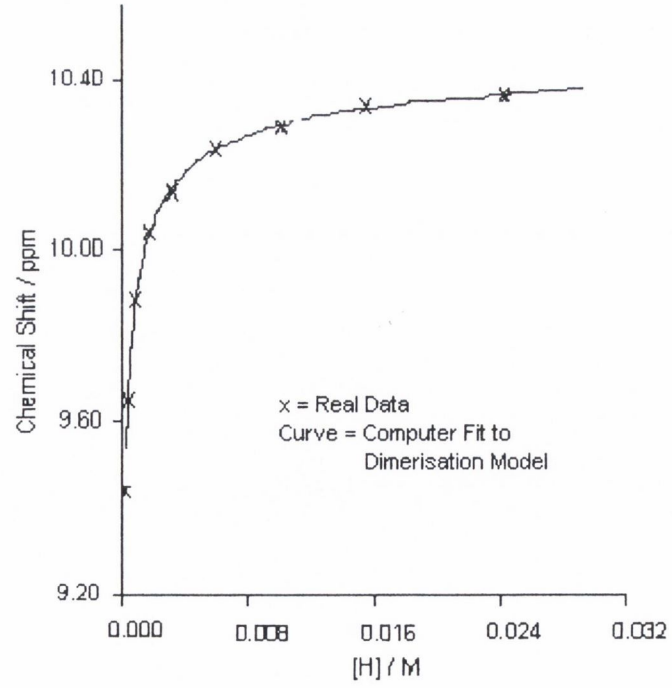


Figure 5.8a Chemical shift of the C-3 NH of **5.9** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

The carbamoyl NH signals also moved downfield, although to a lesser extent (~0.2 - 0.3 ppm). Due to signal broadening it was not possible to follow these peaks as accurately, but their general movement supported the analyses of the C-3 NH signals. Fitting the carbamoyl NH chemical shift data of **5.9** to a dimerisation model afforded an association constant of 4577 (±1814) M^{-1} (see **Figure 5.8b**).

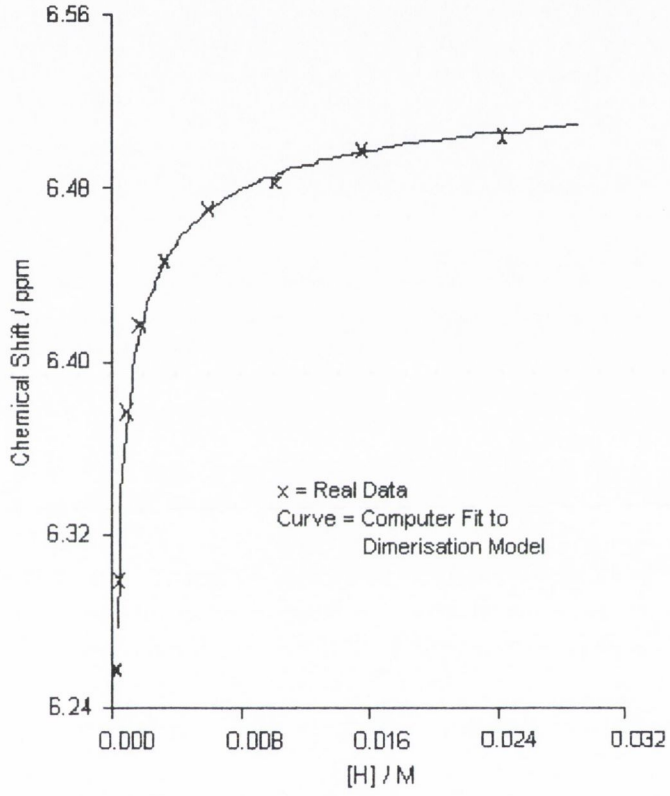


Figure 5.8b Chemical shift of the carbamoyl NH of **5.9** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

A concentration study of **5.9** was also carried out in 1% (v/v) DMSO-d_6/$CDCl_3$ over the concentration range 0.2 mM to 13.8 mM. Again the C-3 NH signal shifted downfield and a fit of the chemical shift data to a dimerisation model (see **Figure 5.9**) gave the lower association constant of 176 (±13) M^{-1}.

A re-examination of the chemical shift data from the concentration study of **5.9** in $CDCl_3$ gave association constants between 7.8 x 10^5 and 1.4 x 10^6 M^{-1}. This is based on the assumption that (δ_{dimer} - $\delta_{monomer}$) is likely to be between 3 and 4 ppm in this solvent.

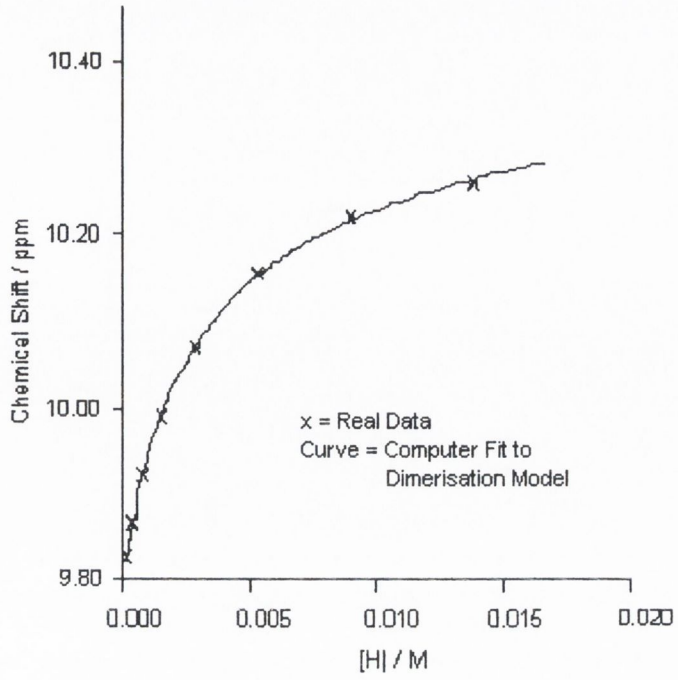
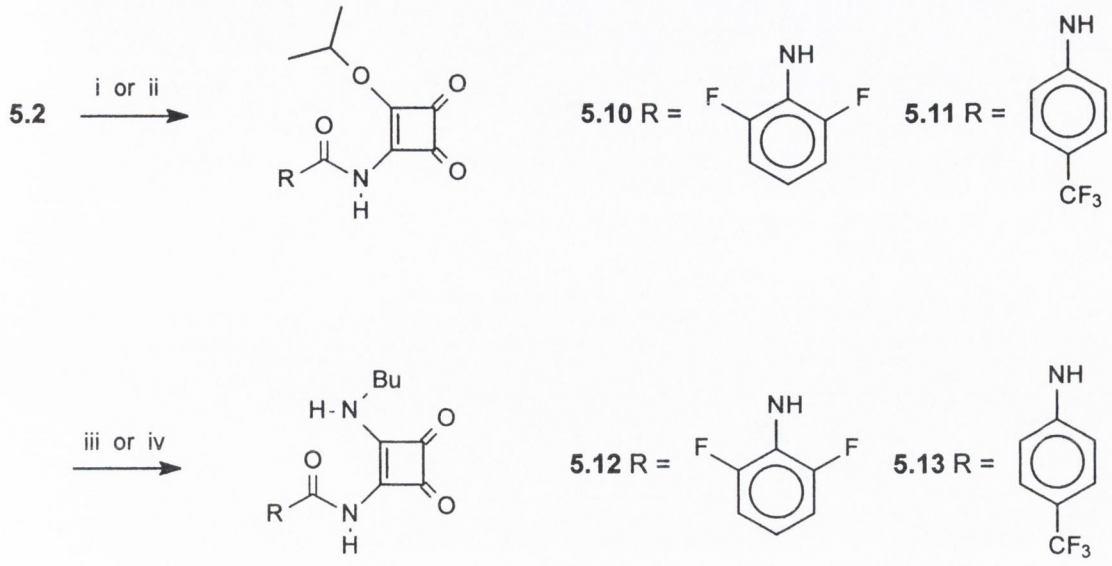
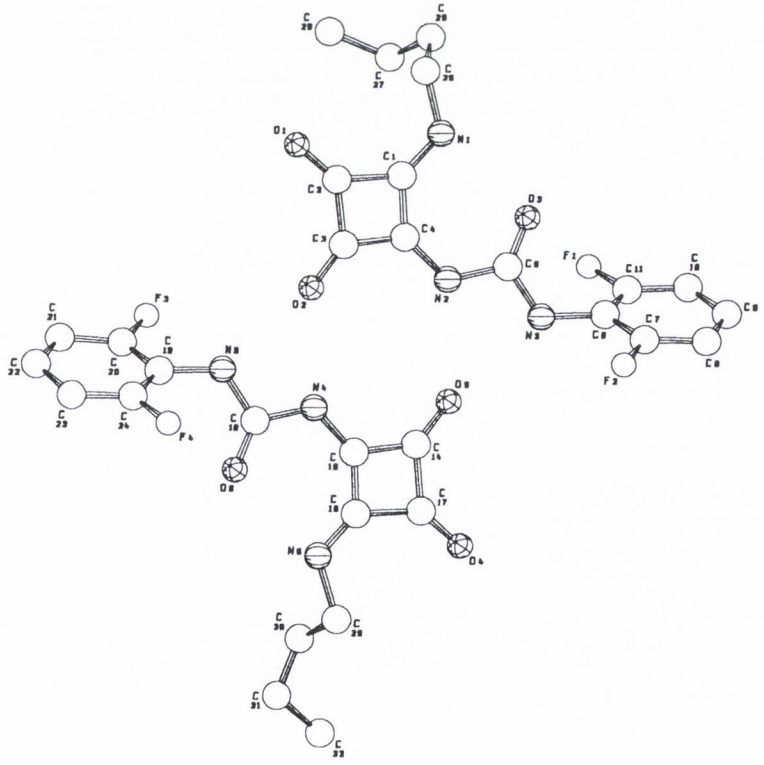


Figure 5.9 Chemical shift of the C-3 NH of **5.9** in 1% (v/v) DMSO-d_6/$CDCl_3$ versus its concentration

Having demonstrated strong dimerisation of the two *N*-butylcarbamoyl squaramides **5.7** and **5.9** in 5% (v/v) $CD_3CN/CDCl_3$, it seemed promising to investigate the self-association of other carbamoyl squaramides. In particular, we felt that by introducing electron-withdrawing aryl groups at the carbamoyl nitrogen, the acidity of the carbamoyl NH might be increased, thus enhancing dimerisation. Attempts to synthesise *p*-nitrophenylcarbamoyl squaramide and 2,4-difluorophenylcarbamoyl squaramide systems met with failure, yielding insoluble products. The 2,6-difluorophenylcarbamoyl squaramide **5.12** and α,α,α-trifluorotolylcarbamoyl squaramide **5.13** were synthesised as shown in **Scheme 5.4**. **5.12** was recrystallised from CH_2Cl_2-petroleum ether and the x-ray structure solved (see **Figure 5.10**). It can be clearly seen that **5.12** dimerises in a fashion similar to **5.9**. The crystals were not of sufficent quality to calculate hydrogen positions, however heavy atom distances between the nitrogens of the urea and the C-2 carbonyl oxygens could be measured and indicate hydrogen bond formation.



Scheme 5.4 *Reagents, conditions (and yields)*; i, 2,6-Difluorophenylisocyanate, Et_3N, dry CH_3CN, room temperature, 18 h (96% of **5.10**); or ii, α,α,α-Trifluorotolylisocyanate, Et_3N, dry CH_3CN, room temperature, 18 h, (**5.11**); iii, Butylamine, CH_2Cl_2, room temperature, 20 minutes, (67% of **5.12**); or iv, Butylamine, $CHCl_3$, room temperature, 18 h, (11% of **5.12** from the isocyanate)



SCHAKAL

Figure 5.10 Crystal structure of **5.12**. Selected distances (Å) are N(5)···O(2) 2.86, N(4)···O(2) 2.76, N(5)···O(6) 2.78,. N(2)···O(5) 2.80, N(3)···O(5) 2.95, N(1)···O(3) 2.77

During a concentration study of **5.12** in 5% (v/v) $CD_3CN/CDCl_3$ over the range of concentrations from 0.3 mM to 24.2 mM, the C-3 NH signal moved downfield by 0.94 ppm. The chemical shift data fitted to a dimerisation model to afford an association constant of 9818 (± 897) M^{-1} (see **Figure 5.11a)**. The carbamoyl NH peak was broad, partially hidden under the C-4 NH signal and poorly resolved at low concentrations and so could not be followed accurately, but it does support the analysis by its movement (see **Figure 5.11b**).

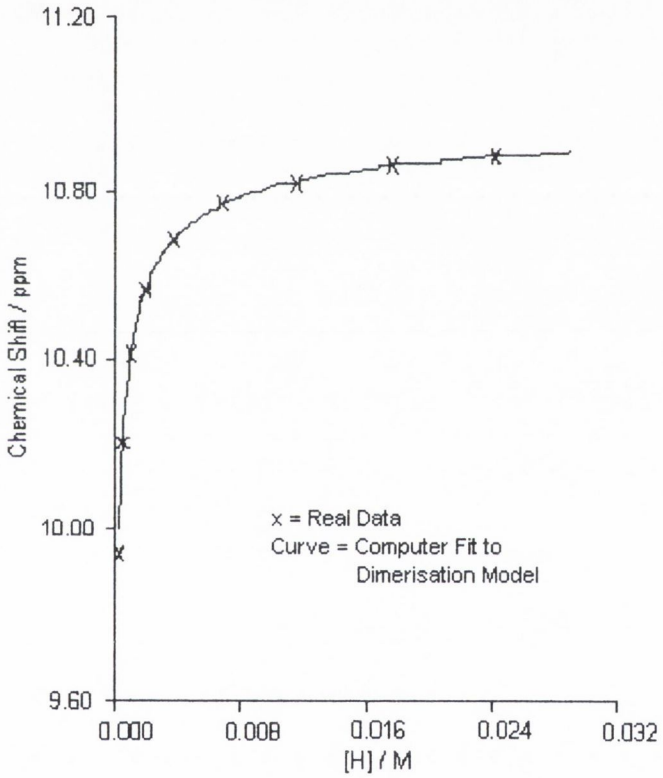


Figure 5.11a Chemical shift of the C-3 NH of **5.12** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

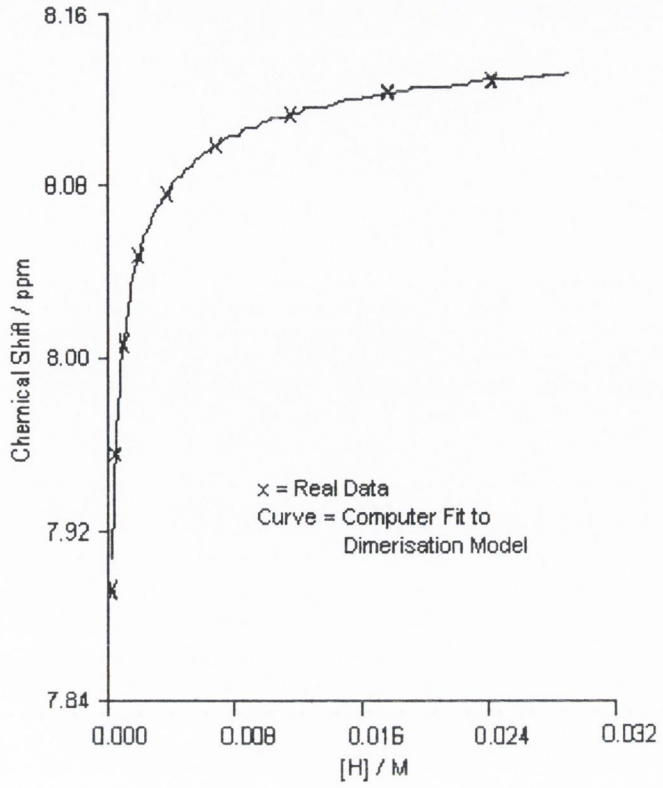


Figure 5.11b Chemical shift of the carbamoyl NH of **5.12** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

The concentration study of **5.13** covering the range of concentrations from 0.2 mM to 13.5 mM also showed a downfield shift in the C-3 NH and carbamoyl NH signals, by 0.77 and 0.24 ppm respectively. The chemical shift data however could not be fitted accurately to a dimerisation model for either NH signal (see **Figure 5.12a** and **5.12b).** Unusually the C-3 NH signal broadened considerably as the concentration increased, a feature indicative of an increase in the rate of exchange, possibly due to the emergence of alternative hydrogen bonding patterns.

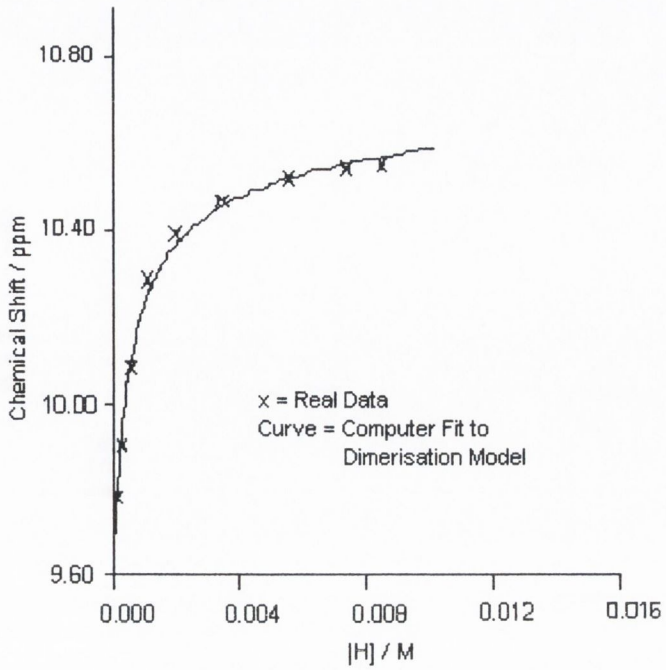


Figure 5.12a Chemical shift of the C-3 NH of **5.13** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

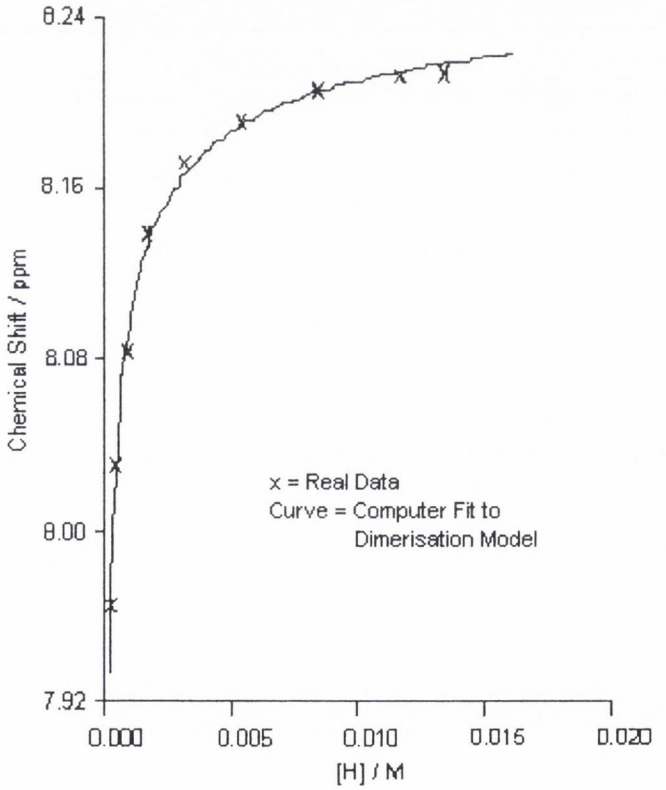
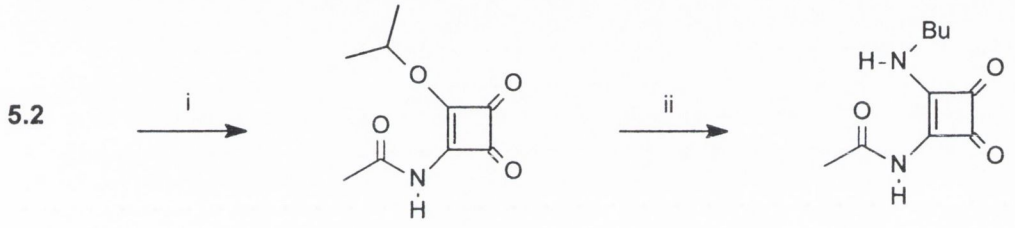


Figure 5.12b Chemical shift of the carbamoyl NH of **5.13** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

5.3.1 Dimerisation study of N-acetyl-N'-butyl squaramide 5.15

The importance of the C-3 NH in the dimerisation of carbamoyl squaramide was clearly shown in the studies carried out above. The extent to which the carbamoyl NH participated was questionable however, it being difficult to follow due to signal broadening. In order to quantify the co-operative effect of the carbamoyl hydrogen bonding, we planned to study the *N*-acetyl squaramide **5.15,** which incorporates an intramolecular hydrogen bond similar to the previous squaramides, while lacking a carbamoyl group. **5.15** was synthesised as shown in **Scheme 5.5**.



5.14 **5.15**

Scheme 5.5 *Reagents, conditions (and yields)*; i, Ac_2O, Et_3N, dry CH_3CN, room temperature, 19 h, (38%); ii, Butylamine, CH_2Cl_2, 20 h, (88%)

A concentration study of **5.15** in 5% (v/v) $CD_3CN/CDCl_3$ over the range of concentrations from 0.2 mM to 24.3 mM showed a weak dimerisation. The chemical shift data of the acetyl NH could be fitted to a dimerisation model (see **Figure 5.13**) which produced an association constant of 116 M^{-1}. This result would indicate that the carbamoyl NH is indeed co-operating in the self-association of the *N*-carbamoyl squaramides, resulting in an increase in the strength of the association by over an order of magnitude.

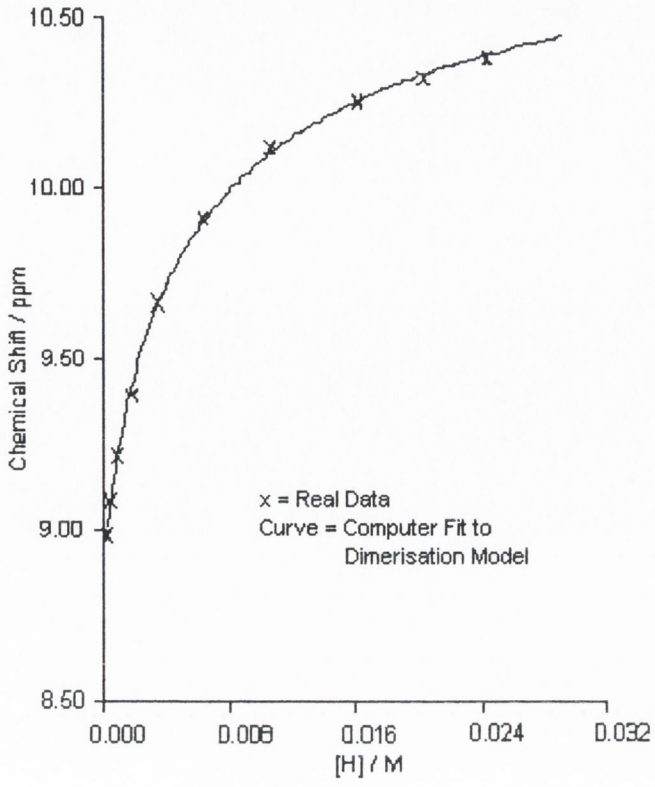


Figure 5.13 Chemical shift of the C-3 NH of **5.15** in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

5.4 Re-examination of Cytidine Binding by 5.7 and 5.9

With the self-associative properties of the carbamoyl squaramides thoroughly examined, the anomalous results of the earlier binding studies of **5.7** with triacetyl cytidine **5.8** in $CDCl_3$ could be at least partially explained. Re-examination of the binding studies of **5.8** as host and **5.7** as guest gave promising results. Incorporating a dimerisation constant of 1×10^6 M^{-1} for **5.7** in the analysis of the chemical shift data, yielded curves which fitted well to a 1:1 binding isotherm as shown in **Figure 5.14a** and **5.14b**. These curves afforded binding constants of >50,000 (±6200) M^{-1} (see **Figure 5.2a** and **5.2b** for original graphs).

The aberrant upfield shift of the C-3 NH in the earlier reverse titration study as depicted in **Figure 5.3b**, can perhaps also be explained. As cytidine is added to the solution of **5.7**, a **5.7-5.8** host-guest complex is formed, accompanied by the dissociation of the **5.7-5.7** dimer. If the chemical shift of the host-guest complex is at a lower frequency than that of the dimer then an upfield movement can be expected on adding **5.8**.

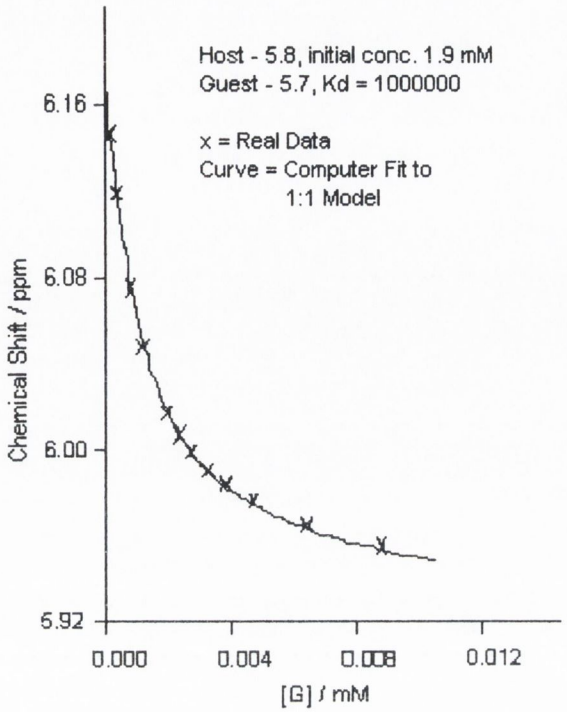


Figure 5.14a 1:1 Binding curve of **5.7** with **5.8** in $CDCl_3$ incorporating an estimated dimerisation constant of **5.7** in the analysis. Initial concentration of **5.8** was 1.9 mM

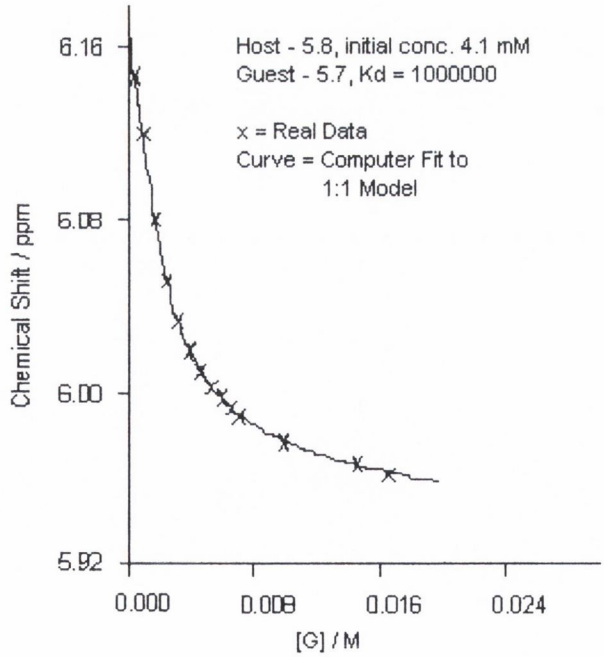


Figure 5.14b 1:1 Binding curve of **5.7** with **5.8** in $CDCl_3$ incorporating an estimated dimerisation constant of **5.7** in the analysis. Initial concentration of **5.8** was 4.1 mM

These results suggested that a more comprehensive study of cytidine binding in 5% (v/v) $CD_3CN/CDCl_3$ might be fruitful. A concentration study of tris(^{t}Butyldimethylsilyl) cytidine[5.4] (TBDMS cytidine) in 5% (v/v) $CD_3CN/CDCl_3$ covering the range of concentrations from 1.7 mM to 77.6 mM showed an upfield movement of the H-6 chemical shift, and the curvature of the data set indicated a dimerisation constant of 2 M^{-1} (see **Figure 5.15**). The NH_2 chemical shift exhibited a similar curvature but due to signal broadening and its temporarily concealment under the H-5 peak, the movement could not be accurately analysed.

A binding study by NMR titration experiment of squaramide **5.9** as host and TBDMS cytidine as guest, showed a large downfield shift of the carbamoyl NH signal. The chemical shift data could be fitted to a 1:1 binding model incorporating the dimerisation constant of 5778 M^{-1} for **5.9**, to afford a binding constant of 10449 (±540) M^{-1} (see **Figure 5.16a**). The C-3 NH also exhibited a downfield motion, although to a much smaller extent (0.15 ppm). A graph of this signal shift, while not fitting a 1:1 binding model as accurately as above, does support this analysis and afforded a binding constant of 6877(±1351) M^{-1} (see **Figure 5.16b**). These results demonstrate that the **5.9**-cytidine complex forms with a potency comparable to that of the guanosine-cytidine base-pair (10^4-10^5 M^{-1} in $CDCl_3$),[5.5] and in fact the binding would be considerably higher were it not for the strong dimerisation of **5.9** which must be first overcome.

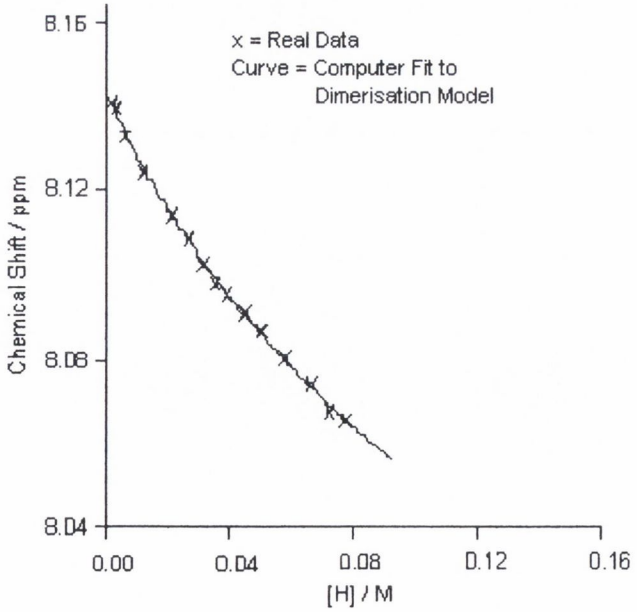


Figure 5.15 Chemical shift of the H-6 of TBDMS cytidine in 5% (v/v) $CD_3CN/CDCl_3$ versus its concentration

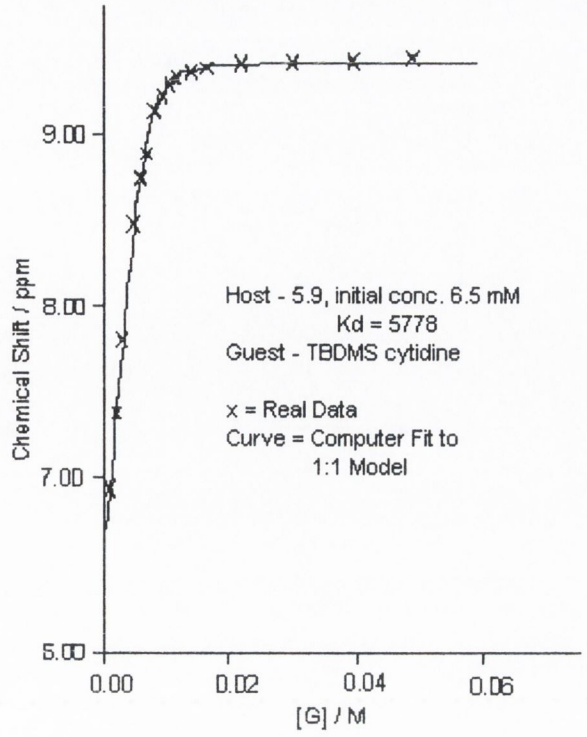


Figure 5.16a Chemical shift of the carbamoyl NH of **5.9** in 5% (v/v) $CD_3CN/CDCl_3$ versus the concentration of TBDMS cytidine

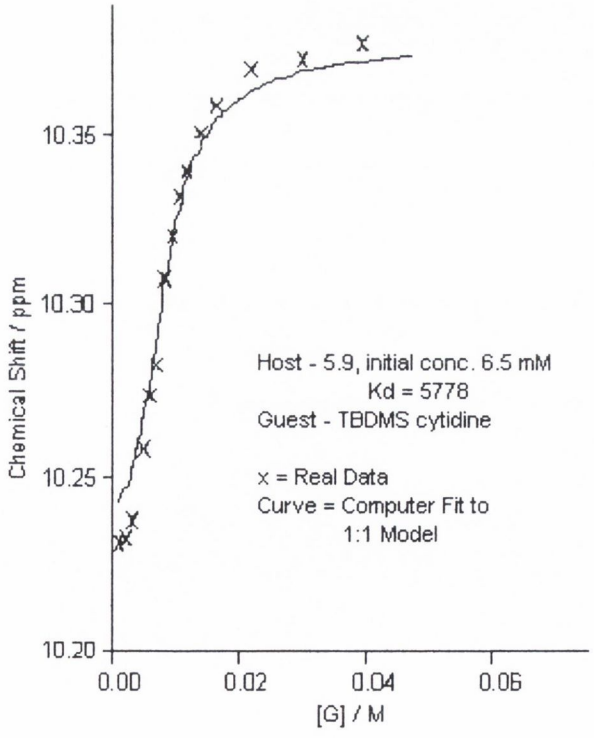


Figure 5.16b Chemical shift of the C-3 NH of **5.9** in 5% (v/v) $CD_3CN/CDCl_3$ versus the concentration of TBDMS cytidine

The plan to investigate binding of guanosine-cytidine base-pairs by carbamoyl squaramides using ^{1}H NMR techniques was abandoned, owing to a number of factors which arose from the previous studies. Firstly, the use of ^{1}H NMR titration experiments techniques would be dependent on the assumption that the guanosine-cytidine base-pair could be treated as a single, associated species[1.41] under the study conditions. Owing to the necessity to carry out binding studies in $CD_3CN/CDCl_3$ mixtures, a solvent mixture in which the base-pair would be expected to be considerable dissociated, it was felt that this assumption would not hold. Secondly, the large **5.9**-cytidine association constant implies that there would be effective competition between **5.9** and guanosine, for cytidine binding sites which would also disrupt the base-pair.

Chapter Six

Experimental

GENERAL

^{1}H and ^{13}C NMR spectra were recorded on Bruker MSL-300 or Bruker DPX-400 spectrometers and were referenced internally to the residual non-deuterated solvent signals (δ_H $CHCl_3$, 7.26; δ_H DMSO, 2.52; δ_H acetone, 2.05) and (δ_C $CDCl_3$, 77.0; δ_C DMSO, 39.5 central peak) signals. *J*-Values are given in Hz. The DEPT 135 ° technique was used to assign (CH_2) signals. The DEPT 90 °, TOCSY and difference NOE techniques were used to further assign NMR signals of the more important intermediates and final receptors. Chemical shifts are reported: value (number of protons, description of absorption, coupling constant(s) where applicable, and assignment). Abbreviations used: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad, vbr = very broad, dd = double doublet dt = double triplet and dq = double quartet. Melting points were determined using open capillary tubes in Electrothermal 9100 melting point block and are uncorrected. IR spectra were recorded using a Mattson Genesis II FT-IR spectrometer or a Perkin-Elmer FT-IR Paragon 1000 spectrometer. Sample films were supported on sodium chloride plates. Analytical TLC was carried out on DC-Alufolien Kieselgel $60F_{254}$ 0.2 mm plates, and compounds were visualised by UV fluorescence, Ninhydrin dip and phosphomolybdic acid dip.

Elemental analyses were carried out in the microanalytical laboratory, Department of Chemistry, University College Dublin. Mass spectra were recorded by Peter Ashton at the University of Birmingham. X-ray diffraction studies were carried out by Maija Nissinen, Department of Chemistry, University of Jyväskylä or by Sylvia Draper, Department of Chemistry, Trinity College Dublin.

All solvents were distilled before use. Dry solvents and common reagents were purified according to the methods of Perrin, Armarego and Perrin.[6.1]

Chromatography of reaction products was carried out using Kieselgel 60 (Merck) 400-230 mesh, by the method of Still *et al.*[6.2]

2,5-Dibromo-3,4-dinitrothiophene 2.1[2.4]

Concentrated H_2SO_4 (40 ml), fuming H_2SO_4 (65% SO_3, 30 ml) and fuming HNO_3 (d. 1.42, 32.4 ml) were cooled to 5 °C. 2,5-Dibromothiophene (36.9 g, 17.2 ml 0.152 mol) was added dropwise with stirring, maintaining the temperature between 20-30 °C. The mixture was allowed stir for 5 minutes and then poured on ice (600 g). The solid was collected by filtration, washed with water (100 ml) and sucked dry in air. The solid was stirred in MeOH (40 ml) for 30 minutes and then let stand overnight at 4 °C. The solid was collected by filtration, washed with cold MeOH and dried *in vacuo* to afford the dibromo dinitrothiophene **2.1** (18.36 g, 36%) as a brown solid; mp. 135-136°C (literature[2.4] 134-135 °C); δ_C(75.5 MHz, $CDCl_3$) 113.41 (CBr), 140.64 (CNO_2); TLC R_f 0.29 developed in CH_2Cl_2-hexane (1:1).

***N*-(4-Acetylamino-thiophen-3-yl)-acetamide 2.3**

This procedure is modified from that of Outurquin and Paulmier.[2.1] A suspension of 2,5-dibromo-3,4-dinitrothiophene **2.1** (15.55 g, 46.8 mmol) in concentrated HCl (325 ml) was cooled to 5 °C, and tin granules (33.68 g, 0.284 mol) added portionwise maintaining the temperature below 30 °C. The mixture was allowed stir for 4 hours and then let stand overnight at 4 °C. The solid was collected by filtration, washed with Et_2O (5x30 ml) and dried in air to yield the bis(ammonium) *salt* **2.2** as a brown solid (5.3 g). The salt **2.2** (5.3 g) was stirred in saturated aqueous Na_2CO_3 (100 ml) for 5 minutes, acetic anhydride (7.5 ml, 80 mmol) was added and the resultant mixture stirred at room temperature for 17 hours. The solid was collected by filtration, washed with water (2x25 ml) and recrystallised from acetone to afford the diacetamide **2.3** (1.18 g, 13%) as grey crystals; mp. 209-211°C (literature[2.1] 207-208 °C); δ_H(300 MHz, DMSO-d_6) 2.10 (3H, s, CH_3), 7.58 (1H, s, CH), 9.53 (1H, br s, NHCO); δ_C(75.5 MHz, DMSO-d_6) 23.50 (CH_3), 110.70 (ArCH), 128.90 (ArC), 168.04 (NHCO); TLC R_f 0.30 developed in $CHCl_3$-MeOH (9:1).

***N*-(4-Acetylamino-2-nitro-thiophen-3-yl)-acetamide 2.4**[2.1]

A solution of the diacetamide **2.3** (0.446 g, 1.8 mmol) in Ac_2O (11.2 ml) and AcOH (11.2 ml) was cooled to 0 °C. A mixture of HNO_3 (d. 1.42, 0.47 ml) in AcOH (5 ml) was then added dropwise with stirring. The temperature was maintained at 20 °C for 2 hours and the mixture then poured on ice (20 g). The mixture was adjusted to pH 7 with saturated aqueous Na_2CO_3, extracted with Et_2O (3x30 ml), the organic phases combined

and washed with brine (30 ml). The organic layer was dried over $MgSO_4$, the solvent evaporated under reduced pressure and the residue purified by recrystallisation from hexane-EtOAc (3:2) to afford the title compound **2.4** (0.4435 g, 81%) as orange crystals; mp. 156-158 °C (literature[2.1] 153-155 °C); δ_H(300 MHz, acetone-d_6) 2.07 (3H, s, CH_3) 2.28 (3H, s, CH_3) 8.17 (1H, s, ArCH) 9.45 (1H, br s, NH) 9.78 (1H, br s, NH); δ_C(75.5 MHz, DMSO-d_6) 23.34 (CH_3) 23.42 (CH_3) 117.26 (ArCH) 128.45 (CNH) 131.77 (CNH) 139.10 (CNO_2) 168.56 (NHCO) 168.76 (NHCO); TLC R_f 0.42 developed in EtOAc.

_N_-(4-Acetylamino-2-iodo-thiophen-3-yl)-acetamide 2.5

This procedure is modified from that of Brunett.[2.5] _N_-Iodosuccinimide (0.755 g, 3.35 mmol) was added to a suspension of the diacetamide **2.3** (0.669 g, 3.39 mmol) in dry CH_2Cl_2 (17 ml). The mixture was stirred under argon at room temperature for 48 hours. The solid was collected by filtration washed with CH_2Cl_2 (2x10 ml) to afford the *title compound* **2.5** (0.95 g, 77%) as a cream solid; mp.194-196 °C; (Found C 29.61, H 2.81, N 8.41, S 10.08, I 39.24.$C_8H_9IN_2O_2S$ requires C 29.64, H 2.80, N 8.64, 9.89, I 39.15%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3359, 3264, 1681, 1644, 1545, 1499, 1424, 1365, 1261, 1020, 955, 861, 744; δ_H(300 MHz, DMSO-d_6) 2.07 (3H, s, CH_3), 2.08 (3H, s, CH_3), 7.89 (1H, s, CH), 9.24 (1H, br s, NHCO), 9.38 (1H, br s, NHCO); δ_C(75.5 MHz, DMSO-d_6) 23.25 (CH_3), 23.36 (CH_3), 76.58 (CI), 114.65 (ArCH), 132.79 (ArCNH), 133.50 (ArCNH), 168.09 (NHCO), 169.18 (NHCO); TLC R_f 0.36 developed in $CHCl_3$-MeOH (9:1).

_N_-(4-Acetylamino-2-iodo-5-nitro-thiophen-3-yl)-acetamide 2.6

This procedure is modified from that of Rinkes.[2.6] A solution of the iodide **2.5** (0.85 g, 2.6 mmol) in AcOH (28.2 ml) and Ac_2O (28.2 ml) was cooled to 0 °C. Concentrated HNO_3 (d. 1.42, 0.56 ml) in AcOH (5.7 ml) was added dropwise with stirring and the mixture stirred at 20 °C for 3 hours. The resultant solid was collected by filtration, washed with water (2x10 ml) and dried *in vacuo* to afford a cream solid. This was purified by flash chromatography eluting with $CHCl_3$-MeOH (98:2) to afford the *title compound* **2.6** (0.368 g, 38%) as white crystals; mp. 244-246 °C; (Found C 25.87, H 2.12, N 11.15, S 8.65, I 33.91. $C_8H_8IN_3O_4S$ requires C 26.03, H 2.18, N 11.38, S 8.69, I 34.38%); δ_H(300 MHz, DMSO-d_6) 2.06 (3H, s, CH_3), 2.10 (3H, s, CH_3), 9.55 (1H, br s, NHCO), 10.18 (1H, br s, NHCO); δ_C(75.5 MHz, DMSO-d_6) 22.84 (CH_3), 23.00 (CH_3),

88.38 (CI), 132.12 (CNH), 136,64 (CNH), 141.18 (CNO_2), 167.73 (CONH), 168.81 (CONH); TLC R_f 0.45 developed in $CHCl_3$-MeOH (9:1).

3-(^{t}Butyldimethylsilyl)-2-propyn-1-ol 2.8

This procedure is modified from that of Mori and Funaki.[2.8] *p*-Toluenesulphonic acid (0.285 g) was added to a stirred solution of the tetrahydropyran protected alkyne **2.7** (1.294 g, 5.12 mmol) in MeOH (26 ml) and stirred at room temperature for 24 hours. The mixture was partitioned between water (30 ml) and EtOAc (30 ml), the organic layer obtained and the aqueous layer washed with EtOAc (2x30 ml). The organic phases were combined and washed with saturated aqueous Na_2CO_3 solution (2x30 ml), brine (30 ml) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure to yield a yellow oil. This was further purified by flash chromatography eluting with hexane-EtOAc (19:1) to afford the unprotected alcohol **2.8** (0.714 g, 82%) as a white crystalline solid; mp. 34-35 °C (literature[6.3] 36-38 °C); δ_H(300 MHz, $CDCl_3$) 0.11 (6H, s, $2CH_3$), 0.93 (9H, s, $C(CH_3)_3$), 1.74 (1H, t, OH), 4.27 (2H, d, CH_2); δ_C(75.5MHz, $CDCl_3$) -4.70 ($Si(CH_3)_2$), 16.41 (Si-C), 26.02 (CH_3), 51.65 (CH_2), 88.90 (C≡C), 104.42 (C≡C); TLC R_f 0.27 developed in hexane-EtOAc (9:1).

Attempted addition of *N*-(4-acetylamino-2-iodo-5-nitro-thiophen-3-yl)-acetamide 2.6 across 2.8

A mixture of the iodo nitro diacetamide **2.6** (0.092 g, 0.25 mmol), the alkyne **2.8** (0.085 g, 0.5 mmol), Bu_4NCl (0.075 g, 0.25 mmol), Na_2CO_3 (0.135 g, 1.25 mmol) and $Pd(OAc)_2$ (0.003 g, 0.0135 mmol) in dry DMF was heated to 80 °C while stirring under argon for 150 minutes. The DMF was evaporated under reduced pressure and the residual DMF removed by repetitive addition and evaporation under reduced pressure of toluene. The resultant brown mixture underwent flash chromatography (3% MeOH in $CHCl_3$) to give an almost quantitative yield of the alcohol **2.8** (0.076 g) and a yellow solid (0.063 g); δ_H(300 MHz, DMSO-d_6) 2.06 (3H, s), 8.07 (2H, br s), 9.43 (1H, br s); δ_C(75.5 MHz, DMSO-d_6) 22.84 (CH_3), 95.07, 122.24, 131.00, 146.67, 169.33; TLC R_f 0.41 developed in $CHCl_3$-MeOH (9:1).

_N_-(3-Amino-2-nitro-thiophen-4-yl)-acetamide 2.10

A solution of the 2-nitro diacetamide **2.4** (0.50 g, 2.1 mmol) in 10% (w/v) aqueous KOH (40 ml) was stirred at room temperature for 18 hours. The solution was continuously extracted with EtOAc for 4 hours, the organic phase dried over $MgSO_4$ and the solvent evaporated under reduced pressure to yield the 3-aminothiophene **2.10** (0.10 g, 24%) as a yellow oil (literature[2.1] δ_H(DMSO-d_6) 2.07 (CH_3), 7.85 (Ar-H), 4.5-5.5 (NH_2), 9.65 (NH); δ_H(300 MHz, DMSO-d_6) 2.08 (3H, s, CH_3), 7.88 (1H, s, ArH), 7.99 (2H, br s, NH_2), 9.69 (1H, br s, NH); δ_C(75.5 MHz, DMSO-d_6) 23.23 (CH_3), 120.17 (ArCNH), 120.58 (ArCH), 125.35 ($ArCNH_2$), 144.45 ($ArCNO_2$); TLC R_f 0.34 developed in $CHCl_3$-MeOH (9:1).

(5-Bromo-3,4-dinitro-thiophen-2-yl)-butyl-amine 2.11

Butylamine (1.31 g, 18 mmol) in CH_2Cl_2 (36 ml) was added over 1 hour to a solution of the dibromide **2.1** (3.01 g, 9 mmol) in CH_2Cl_2 (9ml) at 0 °C with stirring. The solution was refluxed overnight and the solvent evaporated under reduced pressure. The resultant residue was purified by flash chromatography eluting with hexane-EtOAc (11:9) to afford the *thienyl butylamino* **2.11** (2.23 g, 76 %) as a yellow solid; mp. 98-99 °C; (Found C 29.73, H 3.11, N 12.91, S 9.91, Br 24.40. $C_8H_{10}BrN_3O_4S$ requires C 29.64, H 3.11, N 12.96, S 9.89, Br 24.65%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3293, 2948, 2928, 2869, 1592, 1548, 1498, 1447, 1420, 1350, 1253, 1113, 1081, 930, 908, 750, 668; δ_H(300 MHz, $CDCl_3$) 0.98 (3H, t, CH_3) 1.46 (2H, m, CH_2) 1.75 (2H, m, CH_2) 3.34 (2H, m, NCH_2) 8.56 (1H, br s, NH); δ_C(75.5 MHz, $CDCl_3$) 13.45 (CH_3) 19.86 (CH_2) 30.43 (CH_2) 47.75 (CH_2) 92.90 (CBr) 117.10 (CNO_2) 140.98 (CNO_2) 159.60 (CNH); TLC R_f 0.33 developed in CH_2Cl_2-hexane (6:4).

Butylcyanamide 2.12

This procedure is modified from that of Ainley and co-workers.[2.13] A solution of butylamine (0.61 ml, 6.22 mmol) in Et_2O (3.5 ml) was stirred at -10 °C. Cyanogen bromide (0.33 g, 3.11 mmol) was added in portions such that the temperature was maintained below -5 °C. The solution was allowed to warm to room temperature and stirred for 90 min. After this time, analysis by TLC developed in CH_2Cl_2 showed one product component which stained lilac with ninhydrin, R_f 0.40. The mixture was then purified by passing through a silica plug eluting with Et_2O to afford the butylcyanamide

2.12 (0.288 g, 95%) as a yellow oil; δ_H(300 MHz, $CDCl_3$) 0.88 (3H, t, CH_3), 1.33 (2H, m, CH_2), 1.53 (2H, m, CH_2), 2.99 (2H, m, CH_2), 4.57 (1H, s br, NH); δ_C(75.5 MHz, $CDCl_3$) 13.39 (CH_3) 19.27 (CH_2) 31.49 (CH_2) 45.50 (CH_2) 117.00 (CN).

(5-Bromo-3,4-dinitro-thiophen-2-yl)-butyl-cyanamide 2.13

A solution of butylcyanamide (75 mg, 0.77 mmol) in dry CH_2Cl_2 (2 ml), followed by *N, N*-diisopropylethylamine (133µl, 0.77 mmol) in dry CH_2Cl_2 (1 ml), was added dropwise to a stirred solution of **2.1** (0.26 g, 0.77 mmol) in dry CH_2Cl_2 (6 ml) at 0 °C, maintaining the temperature at 0 °C. The solution was allowed warm to room temperature and then refluxed for 2 hours. After this time, analysis by TLC developed in CH_2Cl_2-hexane (1:1) showed virtually complete consumption of the starting material. The solvent was evaporated under reduced pressure and the residue partially purified by flash chromatography eluting with CH_2Cl_2-hexane (1:1) to afford impure thienyl cyanamide (0.165 g) as an orange oil. This oil was purified by flash chromatography eluting with hexane-EtOAc (3:1) to yield the cyanamide **2.13** (0.045 g, 17%) as a yellow oil which rapidly decomposed at room temperature; δ_H(300 MHz, $CDCl_3$) 0.99 (3H, t, CH_3), 1.46 (2H, m, CH_2), 1.82 (2H, m, CH_2), 3.60 (2H, t, CH_2); δ_C(75.5 MHz, $CDCl_3$) 13.44 (CH_3), 19.35 (CH_2), 29.54 (CH_2), 56.89 (CH_2), 108.12 (C), 109.75 (C), 118.81(C), 129.68(C), 142.38 (C); TLC R_f 0.36 developed in hexane-EtOAc (75:25).

Sodium butylcyanamide 2.14

This procedure is modified from that of Weiss and Krommer.[2.14] A solution of butylcyanamide (0.29 g, 2.9 mmol) in isopropyl alcohol (1 ml) was stirred at 20 °C and finely powdered NaOH (0.12 g, 3.1 mmol) added. The slurry was stirred at 20 °C for 2 hours, during which time the NaOH slowly disappeared. The mixture was diluted with Et_2O (2 ml) and filtered to remove any solids. The solvent was evaporated under reduced pressure and the residue dried *in vacuo* to afford the sodium salt **2.14** (0.32 g, 91%) as a white solid; δ_H(300 MHz, DMSO-d_6) 0.84 (3H, t, CH_3), 1.29 (4H, m, $2CH_2$), 2.75 (2H, t, CH_2); δ_C(75.5 MHz, DMSO-d_6) 14.23 (CH_3), 20.06 (CH_2), 35.78 (CH_2), 48.66 (CH_2), 134.42 (CN).

Attempted reduction of dinitro amine 2.11

Tin(II) chloride (0.95 g, 3.6 mmol) and the dinitro amine **2.11** (0.32 g, 1.0 mmol) were dissolved in dry EtOH (2 ml). The mixture was protected from the atmosphere with a calcium chloride tube, stirred at 70 °C for 90 minutes and poured on ice (10 g). The mixture was adjusted to pH 9 with 2 M NaOH and extracted with EtOAc (3x10 ml). The organic phases were combined, washed with brine (10 ml) and dried over $MgSO_4$ to afford a yellow/brown solid (0.123 g) which was ~80% one major component by TLC R_f 0.49 developed in MeOH-$CHCl_3$ (9:1). The aqueous phase containing a cream suspension was further treated with 2 M NaOH until the suspension dissolved. The aqueous phase was washed with EtOAc (3x20 ml), the organic phases combined, washed with brine (20 ml) and dried over $MgSO_4$ to afford a yellow solid (0.123 g) as single component by TLC R_f 0.49 developed in MeOH-$CHCl_3$ (9:1) which could be crystallised in $CDCl_3$-hexane; (Found C 41.67, H 5.69, N 18.18, S 14.12.); δ_H(300 MHz, DMSO-d_6) 0.93 (3H, t), 1.36 (2H, m), 1.66 (2H, m), 3.46 (m, partially obscured by H_2O signal), 4.12 (2H, s), 10.48 (1H, t), 10.95 (1H, s); δ_C(75.5 MHz, DMSO-d_6) 13.54 (CH_3), 19.36 (CH_2), 29.24 (CH_2), 30.49 (CH_2), 47.00 (CH_2) 113.17 (C), 150.36 (C), 172.15 (C).

Acetyl isothiocyanate 3.2

This procedure is modified from that of Takamizawa.[3.2] Acetyl chloride (2.55 ml, 35.9 mmol) was added dropwise to a stirred solution of potassium thiocyanate (3.5 g, 36.1 mmol) in dry acetone (35 ml) and the solution stirred for 20 minutes. The solids were removed by filtration and washed with ether (2 x 20ml). The filtrate was concentrated under reduced pressure to ~20 ml and the residue purified by passing through a silica plug eluting with Et_2O. The solvent was evaporated under reduced pressure, maintaining the temperature below 20 °C, to yield the isothiocyanate **3.2** (2.7 g, 74%) as a pale yellow liquid (literature[3.2] ν_{max}/cm^{-1} 1950-1980, 1725); ν_{max}(film from CH_2Cl_2)/cm^{-1} 1961, 1732; δ_H(400 MHz, $CDCl_3$) 2.30 (s, CH_3); TLC R_f 0.43 developed in hexane-EtOAc (97:3).

[4-(3-Acetyl-thioureido)-thiophen-3-yl]-carbamic acid ^{t}butyl ester 3.3

Triethylamine (4.0 ml, 29 mmol) was added dropwise to a stirred suspension of bis(ammonium) salt **2.2** (3.0 g) in dry CH_2Cl_2 (100 ml) while maintaining the temperature below 10 °C. The mixture was protected from the atmosphere with a calcium chloride tube and stirred during 15 minutes while allowed warm to room temperature. Acetyl isothiocyanate **3.2** (1.43 g, 14.2 mmol) in dry CH_2Cl_2 (220 ml) was added dropwise during 90 minutes and the mixture stirred for 18 hours. TLC analysis (EtOAc-MeOH, 95:5) of the reaction mixture revealed complete consumption of the starting material to give one major UV active product. The suspension was evaporated under reduced pressure and the resultant residue suspended in dry CH_2Cl_2 (60 ml). Di-*tert*-butyl dicarbonate (3.6 g, 16.5 mmol) was added and the mixture refluxed under an atmosphere of argon for 18 hours. The solvent was removed under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (95:5) to afford the *title compound* **3.3** (2.10 g, 47% from **3.2**) as a pale brown solid; mp. 154-156 °C; (Found C 45.51, H 5.34, N 13.28. $C_{12}H_{17}N_3O_3S_2$ requires C 45.70, H 5.43, N 13.32%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3259, 3191, 3009, 2980, 2932, 1692, 1534, 1455, 1370, 1337, 1245, 1160, 1049, 1023, 861, 757; δ_H(400 MHz, $CDCl_3$) 1.51 (9H, s, $C(CH_3)_3$), 2.17 (3H, s, CH_3), 6.67 (1H, br s, NHCOO, exchanges with D_2O), 7.33 (1H, br s, ArH), 7.59 (1H, d, *J* 3.5, ArH), 9.54 (1H, br s, ArCNHCS, exchanges with D_2O), 12.06 (1H, br s, OCNHCS, exchanges with D_2O); δ_C(100.6 MHz, $CDCl_3$) 24.20 (CH_3), 28.27 ($C(CH_3)_3$), 81.15 ($C(CH_3)_3$), 112.63 (ArCH), 117.34 (ArCH), 128.81 (Ar*C*NHCS), 130.68

(ArCNHCOO), 153.70 (NHCOO), 171.42 (NHCO), 179.24 (NHCS); TLC R_f 0.65 developed in CH_2Cl_2-EtOAc (9:1).

(2-Acetylamino-thieno[3,2-*d*]thiazol-6-yl)-carbamic acid ^{t}butyl ester 3.4

Bromine (0.62 g, 3.86 mmol) in $CHCl_3$ (10 ml) was added dropwise over 30 minutes to a stirred solution of ^{t}Boc protected thiourea **3.3** (1.22 g, 3.86 mmol) in $CHCl_3$ (50 ml) at room temperature. The mixture was allowed stir for 10 minutes after which time, analysis by TLC developed in CH_2Cl_2-EtOAc (9:1) showed ~80% of the starting material consumed. A further addition of bromine (0.10 g, 0.6 mmol) in $CHCl_3$ (2 ml) was made and the mixture allowed stir for 10 minutes. The solvent was removed under reduced pressure and the resulting residue purified by flash chromatography sequentially eluting with CH_2Cl_2-EtOAc (9:1) followed by CH_2Cl_2-EtOAc (8:2) and finally with CH_2Cl_2-EtOAc (1:1) to afford *^{t}Boc protected thienothiazole* **3.4** (0.53 g, 44%) as a white solid; mp. 180-181 °C, ν_{max}(film from CH_2Cl_2)/cm^{-1} 3259, 3067, 2978, 2933, 1692, 1548, 1432, 1368, 1274, 1159, 1056, 1002, 888, 844, 757; δ_H(400 MHz, $CDCl_3$) 1.46 (9H, br s, $C(CH_3)_3$), 2.25 (3H, s, CH_3), 7.39 (1H, br s, ArH), 7.66 (1H, br s, NH), 10.54 (1H, br s, NH); δ_C(100.6 MHz, $CDCl_3$) 22.85 (CH_3), 28.17 ($C(CH_3)_3$), 81.03 ($C(CH_3)_3$), 108.99 (ArCH), 123.93 (ArC), 126.07 (ArC), 146.40 (ArC), 152.95 (NHCOO), 160.59 (thiazolo C-2), 168.35 (NHCO); TLC R_f 0.31 developed in CH_2Cl_2-EtOAc (9:1).

***N*-(6-Amino-thieno[3,2-*d*]thiazol-2-yl)-acetamide 3.5**

The title compound was also isolated from the flash chromatographic purification of the (2-Acetylamino-thieno[3,2-*d*]thiazol-6-yl)-carbamic acid ^{t}butyl ester **3.4** to afford the *4-amino-thienothiazole* **3.5** (0.23 g, 28%) as a cream solid; δ_H(400 MHz, DMSO-d_6) 2.17 (3H, s, CH_3), 4.86 (2H, br s, NH_2), 6.12 (1H, s, ArH), 12.22 (1H, br s, NH); δ_C(100.6 MHz, DMSO-d_6) 22.36 (CH_3), 97.75 (ArCH), 122.81 (ArC), 136.70 (ArC), 147.36 (ArC), 159.46 (thiazolo C-2), 168.66 (NHCO); TLC R_f 0.20 developed in CH_2Cl_2-MeOH (95:5).

(2-Acetylamino-5-bromo-thieno[3,2-*d*]thiazol-6-yl)-carbamic acid ^{t}butyl ester 3.6

The title compound was also isolated from the flash chromatographic purification of the (2-Acetylamino-thieno[3,2-*d*]thiazol-6-yl)-carbamic acid ^{t}butyl ester **3.4** to afford the *5-bromo-thienothiazole* **3.6** (0.185 g, 12%) as a cream solid; mp. 186-190 °C (decomp.);

(Found C 36.61, H 3.55, N 10.71, S 15.96, Br 20.73. $C_{12}H_{14}BrN_3O_3S_2$ requires C 36.74, H 3.60, N 10.71, S 16.35, Br 20.37%); δ_H(400 MHz, DMSO-d_6) 1.25 (9H, s, $C(CH_3)_3$), 2.19 (3H, s, CH_3), 8.73 (1H, br s, NH), 12.33 (1H, br s, NH); δ_C(100.6 MHz, DMSO-d_6) 22.32 (CH_3), 28.00 ($C(CH_3)_3$), 78.90 ($C(CH_3)_3$), 108.99 (ArCBr), 121.86 (ArC), 127.43 (ArC), 149.03 (ArC), 153.15 (NHCOO), 160.46 (thiazolo C-2), 168.89 (NHCO); TLC R_f 0.24 developed in CH_2Cl_2-EtOAc (8:2).

(2-Acetylamino-5-iodo-thieno[3,2-*d*]thiazol-6-yl)-carbamic acid ^{t}butyl ester 3.7

N-iodosuccinimide (1.34 g, 6.0 mmol) and thiourea **3.3** (0.53 g, 1.7 mmol) were stirred in $CHCl_3$ (25 ml) at room temperature for 4 hours. The solution was washed with water (3x15 ml) and the combined aqueous layers back-extracted with $CHCl_3$ (20 ml). The organic layers were combined and washed with brine (20 ml), dried over $MgSO_4$, and evaporated under reduced pressure . The resultant brown residue was purified by flash chromatography eluting with CH_2Cl_2-EtOAc (8:2) to afford *(2-Acetylamino-thieno[3,2-d]thiazol-6-yl)-carbamic acid ^{t}butyl ester* **3.4** (77 mg, 15%) spectroscopically identical to that obtained previously and *(2-Acetylamino-5-iodo-thieno[3,2-d]thiazol-6-yl)-carbamic acid ^{t}butyl ester* **3.7** (0.54 g, 73%) as a brown solid; mp. 180-190 °C (decomp.), ν_{max}(film from CH_2Cl_2)/cm^{-1} 3359, 3226, 3164, 3065, 2960, 2925, 2853, 1697, 1632, 1555, 1450, 1346, 1304, 1244, 1163, 1058, 995, 705; δ_H(400 MHz, DMSO-d_6) 1.44 (9H, s $C(CH_3)_3$), 2.18 (3H, s, CH_3), 8.63 (1H, br s, NH), 12.30 (1H, br s, NH); δ_C(100.6 MHz, DMSO-d_6) 22.34 (CH_3), 28.04 ($C(CH_3)_3$), 66.92 (ArCI), 78.71 ($C(CH_3)_3$), 125.99 (ArC), 131.63 (ArC), 149.56 (ArC), 153.22 (NHCOO), 160.76 (thiazolo C-2), 168.84 (NHCO); TLC R_f 0.20 developed in CH_2Cl_2-EtOAc (8:2).

Sodium ethyl formylacetate 3.8[3.4]

Freshly cut sodium metal (0.48 g, 21 mmol) was added to dry toluene (15 ml) and heated with rapid stirring until the sodium melted. The continuously stirred suspension of molten sodium was cooled quickly by lowering the reaction flask into an ice bath to produce sodium sand. The suspension was then allowed come to room temperature and a mixture of ethyl formate (.85 ml, 10.5 mmol) and ethyl acetate (1 ml, 10.3 mmol) added with stirring. The mixture was stirred at room temperature over 24 hours to afford the ethyl formylacetate sodium salt **3.8** as a cream suspension. This was used without further purification.

1-Acetyl-3-(4-amino-thiophen-3-yl) thiourea 3.9

Triethylamine (6.5 ml, 47 mmol) was added dropwise to a stirred suspension of crude bis(ammonium) salt **2.2** (5.0 g) in dry CH_2Cl_2 (100 ml) while maintaining the temperature below 10 °C. The mixture was protected from the atmosphere with a calcium chloride tube and stirred during 15 minutes while allowed warm to room temperature. Acetyl isothiocyanate **3.2** (1.10 g, 10.1 mmol) in dry CH_2Cl_2 (500 ml) was added dropwise during 2 hours and the mixture stirred for 18 hours. The suspension was evaporated under reduced pressure, the residue taken up in CH_2Cl_2-EtOAc (8:2, 50 ml), the solids removed by filtration and washed with methanol (2x50 ml). The filtrate was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (85:15) to afford the *thiourea* **3.9** (1.50 g, 62% from **3.2**) as a pale brown solid, mp. 154-155 °C; (Found C 38.90, H 4.13, N 19.36, S 28.20. $C_7H_9N_3OS_2$ requires C 39.05, H 4.21, N 19.52, S 29.78%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3385, 3233, 3187, 3046, 1687, 1569, 1542, 1419, 1371, 1344, 1246, 1154, 1037, 853, 766; δ_H(400 MHz, $CDCl_3$) 2.21 (3H, s, CH_3), 3.53 (2H, br s, NH_2), 6.39 (1H, d, *J* 3.5, CH), 7.84 (1H, d, *J* 3.5, CH), 8.92 (1H, br s, ArCNHCO), 12.25 (1H, br s, OCNHCS); δ_C(100.6 MHz, $CDCl_3$) 24.36 (CH_3), 103.82 (ArCH), 116.12 (ArCH), 128.50 (ArCNH), 138.50 ($ArCNH_2$), 170.70 (NHCO), 177.00 (NHCS); TLC R_f 0.23 developed in CH_2Cl_2-EtOAc (9:1).

***(Z)*-3-{[4-(3-Acetyl-thioureido)-thiophen-3-yl]-amino}-propenoic acid ethyl ester 3.10**

The previously prepared crude suspension of sodium ethyl formylacetate **3.8** in toluene (3 ml) was added to dry EtOH (5 ml). Acetic acid (2 ml) was added to dissolve the mixture. The resultant solution was then added to thiourea **3.9** (0.15 g, 0.7 mmol) in dry EtOH (10 ml) and refluxed for 30 minutes. After this time, analysis by TLC developed in CH_2Cl_2-EtOAc (95:5) showed >95% of the starting material consumed and a single product spot R_f 0.46. The mixture was allowed cool to room temperature and stand for 15 hours. The solvent was evaporated under reduced pressure, and residual acetic acid was removed by repetitive addition and evaporation under reduced pressure of toluene (2x5ml). The residue was suspended in CH_2Cl_2 and the solids removed by filtration. The filtrate was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (95:5) to afford the *title compound* **3.10**

(104 mg, 45%) as a yellow solid; δ_H(400 MHz, $CDCl_3$) 1.30 (3H, t, *J* 7.0, CH_2CH_3), 2.24 (3H, s, $COCH_3$), 4.19 (2H, q, *J* 7.0, CH_2), 4.88 (1H, d, *J* 8.5, CHCOOEt), 6.73 (1H, d, *J* 3.5, ArH), 6.95 (1H, dd, *J* 12.0 and 8.5, C*H*NH), 7.82 (1H, d, *J* 3.5, ArH), 9.31 (1H, br s, ArCNHCS), 9.53 (1H, d, *J* 12.0, CHN*H*), 12.22 (1H, br s, CONHCS); δ_C(100.6 MHz, $CDCl_3$) 14.40 (CH_2CH_3), 24.36 ($COCH_3$), 59.35 (CH_2CH_3), 89.19 (*C*HCOOEt), 106.75 (ArCH), 118.09 (ArCH), 128.17 (ArC), 135.93 (ArC), 145.65 (CHNH), 169.94 (NHCO), 171.37 (NHCO), 178.77 (NHCSNH); TLC R_f 0.65 developed in CH_2Cl_2-EtOAc (8:2).

(*Z*)-3-[(2-Acetylamino-5-bromo-thieno[3,2-*d*]thiazol-6-yl)-amino]-propenoic acid ethyl ester 3.11

Trifluoroacetic acid (0.3 ml, 3.9 mmol) was added to ^{t}boc protected thienothiazole **3.6** (50 mg, 130 μmol) and stirred for 30 minutes at room temperature. The solution was evaporated under reduced pressure. Residual trifluoroacetic acid was removed by repetitive addition and evaporation under reduced pressure of toluene (2x1 ml). The residue was dissolved in dry EtOH (5ml) to which was added the previously prepared crude suspension of sodium ethyl formylacetate **3.8** in toluene (3 ml) and acetic acid (2 ml). The solution was stirred at room temperature over 24 hours. The solvent was evaporated under reduced pressure and the residue taken up in CH_2Cl_2 (10 ml). The solids were removed by filtration and washed with CH_2Cl_2 (20 ml). The solvent was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (9:1) to afford the *enamine* **3.11** (6 mg, 12%) as a cream solid; δ_H(400 MHz, DMSO-d_6) 1.26 (3H, t, *J* 7.0, CH_2CH_3), 2.20 (3H, s, $COCH_3$), 4.16 (2H, q, *J* 7, CH_2CH_3), 4.95 (1H, d, *J* 8, C*H*COOEt), 8.41 (1H, dd, *J* 12 and 8, C*H*NH), 10.28 (1H, d, *J* 12.5, CHN*H*), 12.49 (1H, br s, NHCO); δ_C(100.6 MHz, DMSO-d_6) 14.24 (CH_2CH_3), 22.29 ($COCH_3$), 58.96 (CH_2CH_3), 87.08 (*C*HCOOEt), 91.91 (ArCBr), 124.32 (ArC), 128.66 (ArC), 142.92 (ArC), 144.19 (CHNH), 161.00 (thiazolo C-2), 169.07 (CO), 169.36 (CO); TLC R_f 0.76 developed in CH_2Cl_2-EtOAc (9:1).

(*Z*)-3-{Acetyl-[6-(^{t}butoxycarbonylamino)-thieno[3,2-*d*]thiazol-2-yl]-amino}-propenoic acid ethyl ester 3.12

Ethyl iodoacrylate (12 μl, 93 μmol) was added to a mixture of the thienothiazole **3.4** (30 mg, 95 μmol) and dry powdered K_2CO_3 (13 mg, 94 μmol) in dry acetone (500μl) under an argon atmosphere. The mixture was stirred at 50 °C for 24 hours. Another 0.5

equivalents of acrylate was added and the mixture left stir for 6 hours. A further 0.5 equivalents of acrylate was then added and stirring continued for 18 hours. The solvent was removed under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (97.5:2.5) to afford the partially purified *title compound* **3.12** (15 mg, 38%) as a white film; δ_H(400 MHz, $CDCl_3$) 1.20 (3H, m, CH_2CH_3), 1.55 (9 H, s, $C(CH_3)_3$), 2.36 (3H, s, $COCH_3$), 4.13 (2H, m, CH_2), 6.17 (1H, d, CH), 7.14 (1H, d, CH), 7.34 (1H, br s, ArH), 7.45 (1H, br s, NH).

[2-(*N*-Acetyl-*N*-benzyl-amino)-thieno[3,2-*d*]thiazol-6-yl]-carbamic acid ^{t}butyl ester 3.13

Benzyl bromide (12 µl, 100 µmol) was added to a mixture of the thienothiazole **3.4** (30 mg, 95 µmol) and dry powdered K_2CO_3 (13 mg, 94 µmol) in dry acetone (1 ml) under an argon atmosphere. The mixture was stirred at 50 °C for 22 hours. The solvent was removed under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-hexane (9:1) to afford the *benzyl protected amide* **3.13** (33 mg, 85%) as a white powder; δ_H(400 MHz, $CDCl_3$) 1.54 (9H, s, $C(CH_3)_3$), 2.33 (3H, s, CH_3CO), 5.55 (2H, br s, CH_2), 7.17 (1H, br s, thieno CH or NH), 7.29 (5H, m partially obscured under $CHCl_3$ peak, ArH), 7.47 (1H, br s, thieno CH or NH); δ_C(100.6 MHz, $CDCl_3$) 22.90 (CH_3CO), 28.28 ($C(CH_3)_3$), 51.50 (CH_2), 81.06 ($C(CH_3)_3$), 107.86 (thieno CH), 125.37 (ArC), 126.09 (ArCH), 126.26 (ArC), 127.71 (ArCH), 128.99 (ArCH), 136.14 (ArC), 145.96 (ArC), 152.53 (NHCOO), 160.10 (thiazolo C-2), 170.63 (CH_3CO);); TLC R_f 0.51 developed in CH_2Cl_2-hexane (8:2).

[2-(*N*-Acetyl-*N*-benzyl-amino)-5-iodo-thieno[3,2-*d*]thiazol-6-yl]-carbamic acid ^{t}butyl ester 3.14

Benzyl bromide (215 µl, 1.87 mmol) was added to a mixture of 5-iodo thienothiazole **3.7** (0.82 g, 1.87 mmol) and dry powdered K_2CO_3 (0.26g, 1.88 mmol) in dry DMF (20 ml) under an argon atmosphere. The mixture was stirred at 64 °C for 24 hours. The mixture was partitioned between water (50 ml) and Et_2O (50 ml), the organic layer obtained and the aqueous layer washed with Et_2O (3x25 ml) and $CHCl_3$ (3x25 ml). The organic phases were combined and washed with water (25 ml), brine (25 ml) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2 to afford the *title compound* **3.14** (0.623 g, 63%) as

a white powder; mp. 163-164 °C (decomp.); (Found C 42.85, H 3.82, N 7.65. $C_{19}H_{20}IN_3O_3S_2$ requires C 43.11, H 3.81, N 7.94%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3310, 3065, 2977, 2930, 1715, 1651, 1533, 1496, 1453, 1393,1366, 1247, 1165, 1062, 1023, 983, 873, 755, 683; δ_H(400 MHz, $CDCl_3$) 1.46 (9H, s, $C(CH_3)_3$), 2.35 (3H, s, CH_3CO), 5.55 (2H, br s, CH_2), 6.16 (1H, br s, NH), 7.30 (6H, m, ArH); δ_C(100.6 MHz, $CDCl_3$) 22.51 (CH_3CO), 27.71 ($C(CH_3)_3$), 51.17 (CH_2), 71.76 (ArCI), 80.47 ($C(CH_3)_3$), 125.99 (ArCH), 127.26 (ArCH), 128.45 (ArCH), 128.99 (ArC), 130.79 (ArC), 135.68 (ArC), 147.68 (ArC), 152.27 (NHCOO), 161.71 (thiazolo C-2), 170.32 (CH_3CO); TLC R_f 0.43 developed in CH_2Cl_2-EtOAc (99:1).

(*E*)-4-{[2-(*N*-Acetyl-*N*-benzyl-amino)-5-iodo-thieno[3,2-*d*]thiazol-6-yl]-(^{t}butoxycarbonyl)-amino}-but-2-enoic acid ethyl ester 3.15

Sodium hydride (60% oil dispersion, 96 mg, 2.4 mmol) was added to a stirred solution of benzyl protected 5-iodo thienothiazole **3.14** (0.623 g, 1.2 mmol) in dry DMF (10 ml) under an atmosphere of argon. The mixture was stirred at room temperature for 5 minutes. Ethyl 4-bromocrotonate (75%, 216µl, 1.2 mmol) was added *via* a syringe over 1 minute and the mixture stirred for 1 hour. Analysis by TLC indicated that ~80% of the starting material and all of the crotonate had been consumed. A further addition of ethyl 4-bromocrotonate (75%, 108µl, 0.6 mmol) was made and the mixture stirred for 1 hour. The reaction mixture was partitioned between water (20 ml) and Et_2O (20 ml), the organic layer obtained and the aqueous layer washed with Et_2O (3x10 ml) and $CHCl_3$ (3x10 ml). The organic phases were combined, washed with water (20 ml), brine (20 ml) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2 to afford the *title compound* **3.15** (0.454 g, 60%) as a colourless glass; mp. 128-129 °C; (Found C 46.94, H 4.41, N 6.32. $C_{25}H_{28}IN_3O_5S_2$ requires C 46.81, H 4.40, N 6.32%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 2978, 2932, 1709, 1670, 1494, 1452, 1391, 1367, 1343, 1307, 1279, 1225, 1160, 1085, 1037, 978, 859, 757; δ_H(400 MHz, $CDCl_3$) 1.26 (9H, s and partially obscured t, *J* 7.0, CH_2CH_3 and $CH^a{}_3C(CH^b{}_3)_2$), 1.46 (3H, br s, $CH^a{}_3C(CH^b{}_3)_2$), 2.38 (3H, s, CH_3CO), 4.13 (2H, dq, *J* 7.0 and 2.5, CH_2CH_3), 4.26 (1H, dd, *J* 16.5 and 3.5, CH^aH^bNCOO), 4.44 (1H, dd, *J* 16.0 and 4.5, CH^aH^bNCOO), 5.39 (1H, d, *J* 16.6, CH^aH^bphenyl), 5.74 (1H, d, *J* 17.6, CH^aH^bphenyl), 6.07 (1H, dt, *J* 16.0 and 2.0, CHCOOEt), 6.93 (1H, br m, CHCHCOOEt), 7.27 (6H, m partially obscured under

$CHCl_3$ peak, ArH); δ_C(100.6 MHz, $CDCl_3$) 13.81 (CH_2CH_3), 22.46 ($COCH_3$), 27.51, ($C(CH_3)_3$), 49.04 (CH_2CH_3), 51.25 (CH_2phenyl), 52.94 (ArCI), 59.71 (CH_2CH), 80.50 ($C(CH_3)_3$), 122.07 (*C*HCOOEt), 125.68 (ArCH), 127.20 (ArCH), 128.43 (ArCH), 128.74 (ArC), 135.08 (ArC), 135.65 (ArC), 143.15 ($CHCH_2$), 148.01 (ArC), 153.30 (NCOO), 161.65 (thiazolo C-2), 165.60 (COOEt), 170.33 (CH_3CO); TLC R_f 0.19 developed in CH_2Cl_2-EtOAc (99:1).

2-(*N*-Acetyl-*N*-benzyl-amino)-5-(ethoxycarbonylmethyl)-pyrrolo[2',3':4,5]thieno[3,2-d]thiazole-7-carboxylic acid ^{t}butyl ester 3.16

N, *N*-Diisopropylethylamine (80 μl, 459 μmol) was added to a solution of the crotonylamino iodo thienothiazole **3.15** (100 mg, 156 μmol), $Pd(OAc)_2$ (4 mg, 18 μmol) and PPh_3 (10 mg, 38 μmol) in dry CH_3CN (1 ml) under an atmosphere of argon. The solution was let stand at 65 °C for 21 hours by which time, analysis by TLC developed in CH_2Cl_2-EtOAc (95:5) showed complete consumption of starting material and formation of one major product component (R_f 0.68). The solvent was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (99:1) to afford the *bis protected tricycle* **3.16** (55 mg, 69%) as a pale brown solid; mp. 161-162 °C; (Found C 58.40, H 5.31, N 8.09, S 12.27. $C_{25}H_{27}N_3O_5S_2$ requires C 58.46, H 5.30, N 8.18, S 12.48%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3162, 3126, 2359, 1737, 1669, 1480, 1444, 1372, 1342, 1254, 1203, 1160, 1107, 1023, 974, 847, 764; δ_H(400 MHz, $CDCl_3$) 1.32 (3H, t, *J* 7.0, CH_2CH_3), 1.58 (9H, s, $C(CH_3)_3$), 2.37 (3H, s, CH_3CO), 3.61 (2H, d, *J* 1.0, CH_2COOEt), 4.23 (2H, q, *J* 7.0, CH_2CH_3), 5.63 (2H, br s, CH_2phenyl), 7.32 (5H,m partially obscured under $CHCl_3$ peak, ArH), 7.36 (1H, s, pyrrole C-5 H); δ_C(100.6 MHz, $CDCl_3$) 14.20 (CH_3CH_2), 23.11 (CH_3CO), 28.11 ($C(CH_3)_3$), 32.34 (CH_2COOEt), 51.69 (CH_3CH_2OO), 61.15 (CH_2phenyl), 84.11 ($C(CH_3)_3$), 113.33 (ArC), 122.33 (ArCH), 126.83 (ArCH), 126.95 (ArC), 127.67 (ArCH), 128.77 (ArCH), 130.24 (ArC), 136.43 (ArC), 137.03 (ArC), 141.98 (ArC), 148.52 (ArC), 159.88 (thiazolo C-2), 170.41 (CO), 170.62 (CO); TLC R_f 0.50 developed in CH_2Cl_2-EtOAc (95:5).

Attempted deprotection of 3.16 by treatment with trifluoroacetic acid

A solution of **3.16** (10 mg, 19 μmol) in CH_2Cl_2:TFA (1:1, 500 μl) was stirred at room temperature for 1 hour. The solvent was evaporated under reduced pressure and the residual TFA was removed by repetitive addition and evaporation under reduced pressure

of toluene (2x2 ml) to yield a brown solid (10 mg); δ_H(400 MHz, $CDCl_3$) 1.33 (3H, t, CH_3), 2.38 (3H, s, CH_3CO), 3.65 (2H, s, CH_2COOEt), 4.24 (2H, q, $COOCH_2$), 5.56 (2H, m, CH_2phenyl), 7.25, (m, ArH), 8.54 (1H, br s, pyrrole C-5 H).

(*E*)-4-{Acetyl-[6-(ᵗbutoxycarbonylamino)-thieno[3,2-*d*]thiazol-2-yl]-amino}-but-2-enoic acid ethyl ester 4.1

Ethyl 4-bromocrotonate(75%, 12 μl, 65 μmol) was added to a mixture of thienothiazole **3.4** (20 mg, 64 μmol) and dry powdered K_2CO_3 (26 mg, 188 μmol) in dry CH_3CN (500 μl) under an argon atmosphere. The mixture was stirred at 60 °C for 17 hours. After this time, analysis by TLC developed in CH_2Cl_2-EtOAc (9:1) showed complete consumption of starting material and a single major product. The mixture was allowed cool to room temperature and the resultant white crystals isolated by filtration. The crystals were washed with hexane (1 ml), Et_2O (1ml), water (2 ml) and Et_2O (1 ml), sucked dry in air and then further dried *in vacuo* for 2 hours to afford the *title compound* **4.1** (7 mg, 26%) as white needles; (Found C 50.63, H 5.42, N 9.75, S 15.40. $C_{18}H_{22}N_3O_5S_2$ requires C 50.93, H 5.22, N 9.90, S 15.10%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3310, 2978, 2288, 1716, 1663, 1536, 1492, 1438, 1390, 1368, 1340, 1274, 1250, 1229, 1158, 1096, 1059, 1037, 958, 892, 842, 755, 690; δ_H(400 MHz, $CDCl_3$) 1.27 (3H, t, *J* 7.0, CH_2CH_3), 1.55 (9H, s, $C(CH_3)_3$), 2.38 (3H, s, CH_3CO), 4.19 (2H, q, *J* 7.0, CH_2CH_3), 5.03 (2H, br s, $CHCH_2$), 5.85 (1H, d, *J* 16.0, CHCOOEt), 7.05 (1H, dt, *J* 16.0 and 4.5, $CHCH_2$), 7.13 (1H, br s, NHCOO), 7.45 (1H,br s, ArH); TLC R_f 0.55 developed in CH_2Cl_2-EtOAc (9:1).

(*E*)-4-{Acetyl-[6-(ᵗbutoxycarbonylamino)-5-iodo-thieno[3,2-*d*]thiazol-2-yl]-amino}-but-2-enoic acid ethyl ester 4.2

Ethyl 4-bromocrotonate (75%, 13 μl, 71 μmol) was added to a mixture of 5-iodo thienothiazole **3.7** (30 mg, 68 μmol) and dry powdered K_2CO_3 (26 mg, 188 μmol) in dry CH_3CN (500 μl) under an argon atmosphere. The mixture was stirred at 60 °C for 17 hours. After this time, analysis by TLC developed in CH_2Cl_2-EtOAc (9:1) showed ~95% consumption of starting material and a single major product. The solvent was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (9:1) to afford the *title compound* **4.2** (24 mg, 64%) as a yellow solid; mp. 183-185 °C; ν_{max}(film from CH_2Cl_2)/cm^{-1} 2980, 1787, 1744, 1650, 1555, 1455, 1369, 1274, 1152, 1120, 990, 850, 778, 752; δ_H(400 MHz, $CDCl_3$) 1.29 (3H, t, *J* 7.0, CH_2CH_3), 1.53 (9H, s, $C(CH_3)_3$), 2.41 (3H, s, CH_3CO), 4.20 (2H, q, *J* 7.0, CH_2CH_3), 5.07 (2H, br d, *J* 3.0, NCH_2), 5.91 (1H, d, *J* 15.5, CHCOOEt), 6.15 (1H, br s, NHCOO), 7.04 (1H, dt, *J* 15.5 and 5.0, $CHCH_2$); δ_C(100.6 MHz, $CDCl_3$) 14.15 (CH_2CH_3), 22.55

($COCH_3$), 28.22, ($C(CH_3)_3$), 49.04 (CH_2CH_3), 60.76 (CH_2CH), 81.04 ($C(CH_3)_3$), 123.11 (*C*HCOOEt), 141.03 (*C*HCH_2); TLC R_f 0.21 developed in CH_2Cl_2-EtOAc (95:5).

^{t}Boc protection of (2-acetylamino-5-iodo-thieno[3,2-*d*]thiazol-6-yl)-carbamic acid ^{t}butyl ester 3.7

Di-*tert*-butyl dicarbonate (25 mg, 115 μmol), 5-iodo thienothiazole **3.7** (50 mg, 114 μmol), *N*,*N*-diisopropylethylamine (20 μl, 115 μmol) and DMAP (14 mg, 120 μmol) were dissolved in dry CH_3CN (1 ml) and stirred at 60 °C for 135 minutes. After this time, analysis by TLC developed in CH_2Cl_2-EtOAc (8:2) showed ~60% of the starting material had been consumed, to give a major product component (R_f 0.31) and a minor product (R_f 0.62). The solvent was evaporated under reduced pressure and the residue purified by flash chromatography, initially eluting with CH_2Cl_2-EtOAc (9:1) and then switching mid-way to CH_2Cl_2-EtOAc (8:2) to afford the starting material **3.7** (24 mg), the *bis(N-^{t}Boc) protected thienothiazole* **4.3** (24 mg, 75%); mp. 180-190 °C (decomp.); (Found C 37.87, H 3.91, N 7.56. $C_{17}H_{22}IN_3O_5S_2$ requires C 37.87, H 4.11,N 7.79%.); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3317, 3141, 2979, 1721, 1671, 1535, 1492, 1436, 1390, 1274, 1250, 1229, 1159, 1038, 959, 893, 843, 756; δ_H(400 MHz, $CDCl_3$) 1.45 (18H, s), 2.19 (3H, s), 9.65 (1H, br s); TLC R_f 0.28 developed in CH_2Cl_2-EtOAc (8:1) and the *N-^{t}Boc-N'-acetyl-N'-^{t}Boc protected thienothiazole* **4.4** white solid (6 mg, 19%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3233, 2979, 2928, 2854, 1794, 1752, 1707, 1555, 1455, 1393, 1368, 1272, 1249, 1156, 1123, 855, 760; δ_H(400 MHz, $CDCl_3$) 1.43 (12H, s), 1.58 (9H, s), 7.93 (0.6H, br s), TLC R_f 0.60 developed in CH_2Cl_2-EtOAc (8:2).

{2-[*N*-Acetyl-*N*-(4-methoxybenzyl)-amino]-5-iodo-thieno[3,2-*d*]thiazol-6-yl}-carbamic acid ^{t}butyl ester 4.5

4-Methoxybenzyl chloride (31 μl, 229 μmol) was added to a mixture of the 5-iodothienothiazole **3.7** (100 mg, 228 μmol) and dry powdered K_2CO_3 (94 mg, 681 μmol) in dry DMF (2.5 ml) under an argon atmosphere. The mixture was stirred at 65 °C for 3½ hours. Analysis by TLC developed in CH_2Cl_2-EtOAc (7:3) showed ~90% of the starting material was consumed and one major product (R_f 0.77) formed. The mixture was allowed cool to room temperature and stirred for 3 days. The solvent was evaporated under reduced pressure and the residual DMF was removed by repetitive addition and evaporation under reduced pressure of toluene (2x5 ml). The residue was purified by

flash chromatography eluting with CH_2Cl_2-EtOAc (98:2) to afford the *4-methoxybenzyl protected amide* **4.5** (85 mg, 67%) as a white powder; mp. 167-170 °C; (Found C 43.36, H 3.99, N 7.31, S 11.82, I 20.73. $C_{20}H_{22}IN_3O_4S_2$ requires C 42.94, H 3.96, N 7.51, S 11.46, I 22.68%); ν_{max}(film from CH_2Cl_2)/cm^{-1} 2972, 2929, 1703, 1657, 1513, 1493, 1391, 1248, 1162, 1063, 1033, 753; δ_H(400 MHz, $CDCl_3$) 1.48 (9H, s, $C(CH_3)_3$), 2.37 (3H, s, CH_3CO), 3.80 (3H, s, CH_3O), 5.47 (2H, s, CH_2), 6.17 (1H, br s, NH), 6.87(2H, d, *J* 8.5, C-2 ArH), 7.21 (2H, d, *J* 8.5, C-3 ArH); δ_C(100.6 MHz, $CDCl_3$) 22.99 (CH_3), 28.17 ($C(CH_3)_3$), 51.16 (CH_2), 55.28 (CH_3O), 80.92 ($C(CH_3)_3$), 114.27 (ArCH), 128.00 (ArC), 128.13 (ArCH), 131.20 (ArC), 146.20 (ArC), 152.74 (NHCOO), 159.19 (thiazolo C-2), 170.75 ($NCOCH_3$); TLC R_f 0.31 developed in CH_2Cl_2-EtOAc (95:5).

(*E*)-4-({2-[*N*-Acetyl-*N*-(4-methoxybenzyl)-amino]-5-iodo-thieno[3,2-*d*]thiazol-6-yl}-(^{t}butoxycarbonyl)-amino)-but-2-enoic acid ethyl ester 4.6

Sodium hydride (60% oil dispersion, 8 mg, 20 μmol) was added to a stirred solution of the 4-methoxybenzyl protected 5-iodo thienothiazole **4.5** (57 mg, 10μmol) in dry DMF (1 ml) under an atmosphere of argon. The mixture was stirred at room temperature for 15 minutes. Ethyl 4-bromocrotonate (75%, 19μl, 10 μmol) was added *via* a syringe and the mixture stirred for 40 minutes. Analysis by TLC indicated that ~80% of the starting material had been consumed. A further addition of sodium hydride (60% oil dispersion, 8 mg, 20 μmol) was made and the mixture stirred for 1 hour. The solvent was evaporated under reduced pressure and the residual DMF was removed by repetitive addition and evaporation under reduced pressure of toluene (2x3 ml). The residue was purified by flash chromatography eluting with CH_2Cl_2-EtOAc (98:2) to afford the *title compound* **4.6** (19 mg, 28%) as a white powder; mp. 133-135 °C; ν_{max}(film from CH_2Cl_2)/cm^{-1} 3303, 2971, 2931, 2836, 1710, 1659, 1611, 1514, 1435, 1391, 1249, 1164, 1064, 1032, 964, 871, 817, 757; δ_H(400 MHz, $CDCl_3$) 1.26 (11H, partially obscured br s and t, *J* 7.0, CH_2CH_3 and $CH^a{}_3C(CH^b{}_3)_2$), 1.48 (3H, br s, $CH^a{}_3C(CH^b{}_3)_2$), 2.41 (3H, s, CH_3CO), 3.79 (3H, s, CH_3O), 4.14 (2H, dq, *J* 7.0 and 2.5, CH_2CH_3), 4.29 (1H, d, *J* 17.0, CH^aH^bNCOO), 4.49 (1H, dd, *J* 17.0 and 5.5, CH^aH^bNCOO), 5.36 (1H, d, *J* 16.5, CH^aH^bphenyl), 5.64 (1H, d, *J* 16.0, CH^aH^bphenyl), 6.10 (1H, d, *J* 15.5, CHCOOEt), 6.85 (2H, d, *J* 8.5, C-2 ArH), 6.96 (1H, br m, CHCHCOOEt), 7.18 (2H, d, *J* 8.0, C-3 ArH); δ_C(100.6 MHz, $CDCl_3$) 13.79 (CH_2CH_3), 22.51 ($COCH_3$), 27.55, ($C(CH_3)_3$), 49.06 (CH_2CH_3), 50.82 (CH_2phenyl), 59.73 (CH_2CH), 80.52 ($C(CH_3)_3$), 113.82 (ArCH),

122.12 (*C*HCOOEt), 127.12 (ArCH), 127.63 (ArC), 143.18 (*C*HCH$_2$), 153.25 (NCOO), 158.69 (thiazolo C-2), 165.60 (COOEt), 170.31 (CH$_3$*C*O); TLC R_f 0.20 developed in CH_2Cl_2-EtOAc (95:5).

2-[*N*-Acetyl-*N*-(4-methoxybenzyl)-amino]-5-(ethoxycarbonylmethyl)-pyrrolo[2',3':4,5]thieno[3,2-d]thiazole-7-carboxylic acid ^{t}butyl ester 4.7

N, *N*-Diisopropylethylamine (15 µl, 86 µmol) was added to a solution of the crotonylamino iodo thienothiazole **4.6** (19 mg, 28 µmol), $Pd(OAc)_2$ (1 mg, 4 µmol) and PPh_3 (2 mg, 8 µmol) in dry CH_3CN (200 µl) under an atmosphere of argon. The solution was stirred at 72 °C for 20 hours, after which time, analysis by TLC showed the reaction to ~50% complete. Further additions of dry CH_3CN (300 µl), $Pd(OAc)_2$ (1 mg, 4 µmol) and triphenylphosphine (2 mg, 8 µmol) were made and stirred at 65 °C for an additional 9½ hours. The solvent was evaporated under reduced pressure and the residue purified by flash chromatography eluting with CH_2Cl_2-EtOAc (98.5:1.5) to afford the *p*-methoxybenzyl protected tricycle **4.7** (12 mg, 78%) as a clear glass; δ_H(400 MHz, $CDCl_3$) 1.32 (3H, t, *J* 7.0, CH_2CH_3), 1.61 (10H, s, $C(CH_3)_3$), 2.40 (3H, s, CH_3CO), 3.62 (2H, s, C*H*$_2$COOEt), 3.80 (3H, s, CH_3O), 4.23 (2H, q, *J* 7.0, C*H*$_2$CH$_3$), 5.56 (2H, br s, CH_2phenyl), 6.86 (2H, d, *J* 8.5, C-2 ArH), 7.32 (d partially obscured under $CHCl_3$ peak, *J* 8.5, C-3 ArH), 7.34 (1H, s, C-5 pyrrolo H);); TLC R_f 0.54 developed in CH_2Cl_2-EtOAc (95:5).

Attempted deprotection of 4.7 by treatment with trifluoroacetic acid

A solution of **4.7** (6 mg, 11 µmol) in TFA (500 µl) was stirred at room temperature for 90 minutes. The solvent was evaporated under reduced pressure and the residual TFA was removed by repetitive addition and evaporation under reduced pressure of toluene (2x2 ml) to yield a brown solid (5 mg); δ_H(400 MHz, $CDCl_3$) 1.30 (3H, t, CH_3), 2.42 (3H, s, CH_3CO), 3.62 (2H, s, C*H*$_2$COOEt), 3.79 (3H, s, OCH_3), 4.22 (2H, q, $COOCH_2$), 5.43 (1H, vbr s, CH_2phenyl), 6.84, (2H, d, ArH), 7.11 (2H, d, ArH), 8.86 (2H, vbr s).

3-Amino-4-isopropoxy-cyclobut-3-ene-1,2-dione 5.2.

This procedure is modified from that of Liebeskind.[5.1, 5.2] 3,4-Dihydroxycyclobut-3-ene-1,2-dione (2.0 g, 17.5 mmol) and *p*-toluenesulphonic acid (33 mg, 0.2 mmol) in isopropyl alcohol (15 ml) and toluene (20 ml) was refluxed on a Dean Stark apparatus for 48 hours. The solvent was evaporated under reduced pressure to afford the diisopropyl squarate **5.1** as a colourless oil (literature[5.1]. δ_H($CDCl_3$) 1.46 (CH_3), 5.35 (CH)); δ_H(400 MHz, $CDCl_3$) 1.46 (6H, d, *J* 6.0, $CH(CH_3)_2$), 5.34 (1H, m, $CH(CH_3)_2$); δ_C(100.6 MHz, $CDCl_3$) 22.26 ($CH(CH_3)_2$), 78.43 ($CH(CH_3)_2$), 183.58 (C), 188.79 (C); TLC R_f 0.25 hexane-EtOAc (9:1). The oil was taken up in CH_2Cl_2-MeOH (1:1, 10 ml) stirred at room temperature and ammonia gas slowly bubbled through the mixture for 20 minutes until the diester was completely consumed, as judged by TLC. The solvent was evaporated under reduced pressure and the residue purified by passing through a silica plug, eluting with EtOAc to afford the *title compound* **5.2** (2.05 g, 75%) as a white waxy solid; mp. 174-175 °C; (Found C 54.42, H 5.76, N 8.87. $C_7H_9NO_3$ requires C 54.19, H 5.85, N 9.03%.); δ_H(400 MHz, DMSO-d_6) 1.38 (6H, d, *J* 6.5, $CH(CH_3)_2$), 5.24 (1H, m, $CH(CH_3)_2$), 8.27 (2H, s br, NH_2); δ_C(100.6 MHz, DMSO-d_6) 22.61 ($CH(CH_3)_2$), 76.37 ($CH(CH_3)_2$), 174.59 (C), 177.35 (C), 182.96 (C), 189.88 (C), TLC R_f 0.31 developed in EtOAc-hexane (3:1).

1-Dodecyl-3-(2-isopropoxy-3,4-dioxo-cyclobut-1-enyl)-urea 5.3

Dodecylisocyanate (80 μl, 0.32 mmol) was added to a solution of **5.2** (0.10 g, 0.65 mmol) and *N,N*-diisopropylethylamine (60 μl, 0.32 mmol) stirred at room temperature in acetonitrile (2 ml). The resultant solution was stirred for 18 hours and the solvent evaporated under reduced pressure. The residue was purified by flash chromatography, eluting with CH_2Cl_2-EtOAc (14:1 graduated to 4:1) to afford the *dodecylurea* **5.3** (0.082 g, 72% from consumed starting material) as a white waxy solid; mp. 87-89 °C; (Found C 65.29, H 9.20, N 7.38. $C_{20}H_{34}N_2O_4$ requires C 65.54, H 9.35, N 7.64%.); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3266, 3200, 3101, 3049, 2921, 2852, 1807, 1716, 1687, 1580, 1467, 1432, 1376, 1298, 1145, 1095, 1038, 903, 826, 721, 630; δ_H(400 MHz, $CDCl_3$) 0.89 (3H, t, *J* 7.0, CH_2CH_3), 1.29 (18H, s and partially obscured m, $CH_3(CH_2)_9$), 1.51 (6H, d, *J* 6.0,

$CH(CH_3)_2$), 1.60 (2H, m, $NHCH_2CH_2$), 3.30 (2H, m, $NHCH_2$), 5.47 (1H, m, $CH(CH_3)_2$), 7.90 (1H vbr s, CH_2NH), 9.45 (1H, br s, $CH_2NHCONH$); δ_C(100.6 MHz, $CDCl_3$) 13.68 (CH_2CH_3), 22.24 (CH_2), 22.38 ($CH(CH_3)_2$), 26.42 (CH_2), 28.83 (CH_2), 28.89 (CH_2), 29.11 (CH_2), 29.18 (CH_2), 31.46 (CH_2), 40.56 ($NHCH_2$), 79.51 ($CH(CH_3)_2$), 152.32 (C), 167.40 (C), 181.11 (C), 183.25 (C), 188.08(C); TLC R_f 0.48 developed in CH_2Cl_2-EtOAc (9:1). The flash chromatographic purification of **5.3** also afforded starting material (52 mg) and a minor product (8 mg); δ_H(400 MHz, $CDCl_3$) 0.90 (6H, t, 2 CH_2CH_3), 1.28 (36H, s, 2 $CH_3(CH_2)_9$), 1.54 (9H, d and partially obscured m, $CH(CH_3)_2$ and 2 CH_2), 3.15 (2H, m, CH_2), 3.29 (2H, m, CH_2), 5.49 (1H, m, $CH(CH_3)_2$), 7.68 (1H vbr s, CH_2NH), 9.33 (1H, br s, $CH_2NHCONH$); δ_C(100.6 MHz, $CDCl_3$) 13.62 (CH_2CH_3), 22.20 (CH_2), 22.36 ($CH(CH_3)_2$), 26.41 (CH_2), 28.87 (CH_2), 29.16 (CH_2), 29.49 (CH_2), 31.44 (CH_2), 39.46 ($NHCH_2$), 40.56 ($NHCH_2$), 79.33 ($CH(CH_3)_2$), 151.89 (C), 158.80 (C), 167.44 (C), 180.98 (C), 183.135 (C), 188.18 (C); TLC R_f 0.66 developed in hexane-EtOAc (1:1).

1-Dodecyl-3-(2-dodecylamino-3,4-dioxo-cyclobut-1-enyl)-urea 5.4

Dodecylamine (38 mg, 0.2 mmol) was added with stirring to a solution of dodecylurea **5.3** (17 mg, 0.05 mmol) in CH_2Cl_2 (1 ml) The mixture was stirred at room temperature for 5 days. The resultant precipitate was isolated by filtration, washed with CH_2Cl_2 (2x2 ml) and dried in air to afford the title compound **5.4** (4 mg, 18%) as a white solid; δ_H(400 MHz, DMSO-d_6) 0.86 (6H, m, $2CH_3$), 1.24 (36H, m, $(CH_2)_9CH_3$), 1.44 (2H, m, CH_2), 1.53 (2H, m, CH_2), 3.12 (2H, m, CH_2), 3.57 (2H, m, CH_2), 6.95 (1H, br s, NH), 7.94 (1H, vbr s, NH).

1-Butyl-3-(2-isopropoxy-3,4-dioxo-cyclobut-1-enyl)-urea 5.6

Butylisocyanate (108 µl, 0.96 mmol) was added to a stirred solution of **5.2** (0.30 g, 1.9 mmol) and triethylamine (136 µl, 1.0 mmol) in acetonitrile (8 ml) at room temperature. The resultant solution was stirred for 18 hours and the solvent evaporated under reduced pressure. The residue was purified by flash chromatography, eluting with CH_2Cl_2-EtOAc

(14:1) to afford the *butylurea* **5.6** (0.15 g, 64% from consumed starting material) as a white crystalline solid; mp. 152-155 °C; (Found C 57.08, H 7.33, N 10.42. $C_{12}H_{18}N_2O_4$ requires C 56.68, H 7.14, N 11.02%.); δ_H(400 MHz, $CDCl_3$) 0.96 (3H, t, *J* 7.4, CH_2CH_3), 1.41 (2H, m, CH_3CH_2), 1.51 (6H, d, *J* 6.0, $CH(CH_3)_2$), 1.59 (2H, m, $CH_2CH_2CH_2$), 3.32 (2H, m, $NHCH_2$), 5.47 (1H, m, $CH(CH_3)_2$), 7.92 (1H vbr s, CH_2NH), 9.34 (1H, br s, $CH_2NHCONH$); δ_C(100.6 MHz, $CDCl_3$) 13.25 (CH_2CH_3), 19.55 (CH_2CH_3), 22.38 ($CH(CH_3)_2$), 30.90 ($CH_2CH_2CH_2$), 40.23 ($NHCH_2$), 79.57 ($CH(CH_3)_2$), 152.27 ($CH_2NHCONHC$), 167.37 (NHCONH), 181.08 (C), 182.75 (C), 188.03(C); TLC R_f 0.30 developed in CH_2Cl_2-EtOAc (9:1), and starting material (156 mg).

1-Butyl-3-(2-butylamino-3,4-dioxo-cyclobut-1-enyl)-urea 5.7

Butylamine (170 µl, 1.74 mmol) was added with stirring to a solution of **5.6** (0.15 g, 0.59 mmol) in CH_2Cl_2 (4 ml) The resultant solution was stirred at room temperature for 20 hours and the solvent evaporated under reduced pressure. The residue was purified by passing through a silica plug, eluting with CH_2Cl_2-EtOAc (9:1) to afford the *title compound* **5.7** (0.134 g, 87%) as a white waxy solid; mp. 180-182 °C; (Found C 58.29, H 7.89, N 15.64. $C_{13}H_{21}N_3O_3$ requires C 58.41, H 7.92, N 15.72%.); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3373, 3261, 3171, 3004, 2959, 2934, 2874, 1808, 1703, 1605, 1551, 1478, 1355, 1312, 1234, 1182, 1131, 1077, 1021, 714; FAB-MS (*m/z*, %) 268 ($M+H^+$, 100), 290 ($M+Na^+$, 54), 535 ($2M+H^+$, 31), 557 ($2M+Na^+$, 21), 802 ($3M+H^+$, 1), 824 ($3M+Na^+$, 2), 1069 ($4M+H^+$, 1), 1092 ($4M+Na^+$, 3), 1626 ($6M+Na^+$, 2); δ_H(400 MHz, $CDCl_3$) 0.97 (6H, m, $2CH_3$), 1.42 (4H, m, $2CH_2CH_3$), 1.57 (2H, m, $CONHCH_2CH_2$), 1.66 (2H, m, $NHCH_2CH_2$), 3.27 (2H, q, *J* 7.0, $CONHCH_2$), 3.76 (2H, q, *J* 7.0, $NHCH_2$), 6.62 (1H, br s, $NHCONHCH_2$), 8.36 (1H, br s, $CNHCH_2$), 10.54 (1H, br s, $NHCONHCH_2$); δ_C(100.6 MHz, $CDCl_3$) 13.51 (CH_3), 13.58 (CH_3), 19.49 ($CH_3CH_2(CH_2)_2NHC$), 19.85 ($CH_3CH_2(CH_2)_2NHCO$), 31.61 ($CH_2CH_2CH_2NHCO$), 32.90 ($CH_2CH_2CH_2NHC$), 40.12 (CH_2CH_2NHCO), 44.21 (CH_2CH_2NHC), 153.61 ($NHCONHC$), 160.30 (NHCONH), 172.48 (CH_2NHC), 183.11(CO), 183.42 (CO); TLC R_f 0.49 developed in CH_2Cl_2-EtOAc (8:2).

1-Butyl-3-(2-cyclohexylamino-3,4-dioxo-cyclobut-1-enyl)-urea 5.9

Cyclohexylamine (200 μl, 1.75 mmol) was added with stirring to a solution of **5.6** (0.14 g, 0.55 mmol) in CH_2Cl_2 (4 ml). The resulting solution was stirred at room temperature for 20 hours and the solvent evaporated under reduced pressure. The residue was purified by passing through a silica plug eluting with CH_2Cl_2-EtOAc (8:2) to afford the *title compound* **5.9** (0.132 g, 82%) as a white waxy solid; mp. 142-143 °C; (Found C 61.55, H 7.99, N 13.73. $C_{15}H_{23}N_3O_3$ requires C 61.41, H 7.90, N 14.32%.); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3373, 3262, 3172, 3018, 2932, 2856, 1808, 1700, 1601, 1552, 1471, 1370, 1348, 1243, 1181, 1083, 1050, 680; δ_H(400 MHz, $CDCl_3$) 0.99 (3H, t, *J* 7.5, CH3), 1.25 (1H, br m, cyclohexyl C*H*^{a}H^b), 1.44 (6H, m, CH_3CH_2 and 2 cyclohexyl CH_2), 1.57 (2H, m, $CH_3CH_2CH_2$), 1.66 (1H, br m, cyclohexyl CH^a*H*^{b}), 1.79 (2H, m, cyclohexyl CH_2), 2.04 (2H, m, cyclohexyl CH_2), 3.28 (2H, q, *J* 7.0, $NHCH_2$), 4.13 (1H, m, NHC*H*), 6.60 (1H, br s, CH_2NH), 8.38 (1H, d, *J* 7.5, N*H*CH), 10.55 (1H, br s, $CH_2NHCONH$); δ_C(100.6 MHz, $CDCl_3$)13.21 (CH_3), 19.46 (CH_3CH_2), 23.91 (2 cyclohexyl CH_2), 24.63 (cyclohexyl CH^{a}H^b), 31.20 ($CH_3CH_2CH_2$), 33.67 (2 cyclohexyl CH_2), 39.67 ($NHCH_2$), 52.82 (NHCH), 153.18 (NHCONHC), 159.97 (NHCONH), 171.32 (CH_2NHC), 182.65 (CO) TLC R_f 0.45 developed in CH_2Cl_2-EtOAc (8:2).

1-(2-6-Difluorophenyl)-3-(2-butylamino-3,4-dioxo-cyclobut-1-enyl)-urea 5.12

2,6,-Difluorophenylisocyanate (40 mg, 0.26 mmol) was added to a solution of **5.2** (40 mg, 0.26 mmol) and triethylamine (75 μl, 0.54 mmol) stirred at room temperature in acetonitrile (3 ml). The resultant solution was stirred for 18 hours and the solvent evaporated under reduced pressure. The residue was purified by flash chromatography, eluting with CH_2Cl_2-EtOAc (9:1) to afford the 2,6-difluorophenylurea **5.10** (46 mg, 96% from consumed starting material) as a white crystalline solid; δ_H(400 MHz,); 1.49 (6H, d, *J* 6.0, $CH(CH_3)_2$), 5.49 (1H, m, $CH(CH_3)_2$), 7.00 (2H, m, C-3 and C-5 ArH), 7.27 (m partially obscured under $CHCl_3$ peak, C-4 ArH), 8.55 (1H, br s, ArN*H*CONH), 9.47 (1H, br s, ArNHCON*H*); TLC R_f 0.23 developed in CH_2Cl_2-EtOAc (9:1), and starting material

(16 mg). Butylamine (100 µl, 1.0 mmol) was added with stirring to a solution of **5.10** (46 mg, 0.15 mmol) in CH_2Cl_2 (5 ml) The resultant solution was stirred at room temperature for 20 minutes and the solvent evaporated under reduced pressure. The residue was purified by passing through a silica plug, eluting with CH_2Cl_2-EtOAc (8:2) to afford the *title compound* **5.12** (32 mg, 38% from **5.2**) as a white solid; mp. 212-204 °C; (Found C 55.46, H 4.66, N 12.81. $C_{15}H_{15}F_2N_3O_3$ requires C 55.73, H 4.68, N 13.00%.); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3283, 3193, 3034, 2960, 2934, 2875, 1806, 1731, 1698, 1601, 1553, 1467, 1301, 1231, 1176, 1093, 1029, 876, 786, 736; δ_H(400 MHz, $CDCl_3$) 0.94 (3H, t, *J* 7.5, CH_3), 1.41 (2H, m, CH_3CH_2), 1.61 (2H, m, $CH_3CH_2CH_2$), 3.75 (2H, m,$NHCH_2$), 7.02 (2H, m, C-3 and C-5 ArH), 7.28 (m partially obscured under $CHCl_3$ peak), 8.04 (1H, br s, ArN*H*CONH), 8.21 (1H, br s, CH_2NH), 11.08 (1H, br s, ArNHCON*H*); TLC R_f 0.20 developed in CH_2Cl_2-EtOAc (8:2).

1-(α,α,α-Trifluorotolyl)-3-(2-butylamino-3,4-dioxo-cyclobut-1-enyl)-urea 5.13

α,α,α-Trifluorotolylisocyanate (69 µl, 0.48 mmol) was added to a solution of **5.2** (125 mg, 0.81 mmol) and triethylamine (67 µl, 0.48 mmol) stirred at room temperature in acetonitrile (5 ml). The resulting solution was stirred for 18 hours and the solvent evaporated under reduced pressure. The residue was partially purified by flash chromatography, eluting with CH_2Cl_2-EtOAc (9:1) to afford impure urea **5.11** as a white solid; δ_H(400 MHz, $CDCl_3$) 1.54 (6H, d, *J* 6.0, $CH(CH_3)_2$), 5.63 (1H, m, $CH(CH_3)_2$), 7.52 (2H, d, *J* 8.5, ArH), 7.64 (3H, q, *J* 8.5, ArH), 8.91 (1H, br s, NH). Butylamine (100 µl, 1.0 mmol) was added with stirring to a suspension of the above crude urea in $CHCl_3$ (1 ml). The solution was stirred at room temperature for 18 hours. The precipitate was removed by filtration, washed with $CHCl_3$ (1 ml) and the filtrate evaporated under reduced pressure. The residue was purified by flash chromatography, eluting with CH_2Cl_2-EtOAc (8:2) to afford the *title compound* **5.13** (19 mg, 11% from isocyanate) as a white solid; mp. 223-224 °C; (Found C 53.57, H 4.42, N 11.53. $C_{16}H_{16}F_3N_3O_3$ requires C 54.09, H 4.54, N 11.83%.); δ_H(400 MHz, $CDCl_3$) 1.01 (3H, t, *J* 7.3, CH_3), 1.48 (2H, m, CH_3CH_2), 1.71 (2H, m, $CH_3CH_2CH_2$), 3.83 (2H, m, CH_2NH), 7.63 (4H, dd, *J* 8.5 and

12.0, ArH), 8.31 (1H, br s, CH_2NH), 8.92 (1H, br s, ArNHCONH), 10.80 (1H, br s, ArNHCONH); δ_C(100.6 MHz, $CDCl_3$) 13.59 (CH_3), 19.58 (CH_2), 32.92 (CH_2), 44.64 (CH_2), 119.41 (ArCH), 126.35 (ArCH), 158.80 (C), 173.43 (C); δ_F(376.5 MHz, $CDCl_3$) -62.67 (s, CF_3).

_N_-(2-Butylamino-3,4-dioxo-cyclobut-1-enyl)-acetamide 5.15

Acetic anhydride (180 µl, 1.91 mmol) was added to a solution of **5.2** (205 mg, 1.32 mmol) and triethylamine (500µl, 3.6 mmol) stirred at room temperature in acetonitrile (2 ml). The resulting solution was stirred for 19 hours after which time, analysis by TLC developed in CH_2Cl_2-EtOAc (7:3) showed that ~80% of the starting material had been consumed and one major component produced. The solvent was evaporated under reduced pressure and the resultant residue purified by flash chromatography, eluting with CH_2Cl_2-EtOAc (7:3) to afford the *acetamide* **5.14** (100 mg, 38%) as a white crystalline solid; mp. 133-134 °C; (Found C 54.55, H 5.56, N 7.08. $C_9H_{11}NO_4$ requires C 54.82, H 5.62, N 7.10%.); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3274, 3209, 3100, 3004, 1804, 1737, 1721, 1593, 1526, 1469, 1403, 1330, 1262, 1194, 1094, 1035, 995, 900, 771; δ_H(400 MHz, $CDCl_3$); 1.52 (6H, d, *J* 6.0, $CH(CH_3)_2$), 2.37 (3H, s, CH_3CO), 5.63 (1H, m, $CH(CH_3)_2$), 9.50 (1H, br s, NHCO); δ_C(100.6 MHz, $CDCl_3$) 22.39 ($CH(CH_3)_2$), 22.81 (CH_3CO), 79.24 ($CH(CH_3)_2$), 166.35 (C), 167.09 (C), 184.15 (C), 187.38 (C); TLC R_f 0.33 developed in CH_2Cl_2-EtOAc (7:3). Butylamine (43 µl, 0.44 mmol) was added with stirring to a solution of **5.14** (86 mg, 0.44 mmol) in CH_2Cl_2 (5 ml) The resultant solution was stirred at room temperature for 20 hours and the solvent evaporated under reduced pressure. The residue was purified by passing through a silica plug, eluting with CH_2Cl_2-EtOAc (7:3) to afford the *title compound* **5.15** (81 mg, 29% from **5.2**) as a white crystalline solid; mp. 157-158 °C; (Found C 57.11, H 6.55, N 13.22. $C_{10}H_{14}N_2O_3$ requires C 57.13, H 6.71, N 13.32%.); ν_{max}(film from CH_2Cl_2)/cm^{-1} 3273, 3186, 3025, 2955, 2936, 2873, 1804, 1706, 1606, 1546, 1459, 1370, 1248, 1114, 1065, 1002, 972, 739; δ_H(400 MHz, $CDCl_3$) 0.98 (3H, t, *J* 7.0, CH_3), 1.43 (2H, m, CH_3CH_2), 1.64 (2H, m, $CH_3CH_2CH_2$), 2.33 (3H, s, CH_3CO), 3.76 (2H, m, $NHCH_2$), 7.87 (1H, br s, $NHCH_2$),

10.95 (1H, br s, NHCO); δ_C(100.6 MHz, $CDCl_3$) 13.57 (CH_3), 19.51 (CH_2), 23.04 (CH_3), 32.88 (CH_2), 44.27 ($NHCH_2$), 158.99 (C), 170.91 (C), 173.52 (C), 182.60 (C), 186.80 (C); TLC R_f 0.30 developed in CH_2Cl_2-EtOAc (7:3).

Typical Procedure for NMR Dimerisation Studies:

$CDCl_3$-CD_3CN (95:5, 600 μl) was taken in a NMR tube and 4 μl of a stock solution of **5.12** (38.7 mM) in $CDCl_3$-CD_3CN (95:5) was added. The tube was recapped and the solutions thoroughly mixed, ensuring that the tube walls were completely wetted. The ^1H NMR spectrum was then acquired (>1000 scans) at 30°C. Eight further additions of the **5.12** stock solution were made, each approximately double the previous addition, up to a total added volume of 1006 μl. The resultant solution was mixed thoroughly and the ^1H NMR spectrum acquired after each addition. The chemical shift data was then analysed using HOSTEST version 5.0[5.3] to calculate dimerisation constants K_d, monomer and dimer chemical shifts and error limits (standard deviations).

Typical Procedure for NMR Binding Studies:

A 600 μl $CDCl_3$-CD_3CN (95:5) solution of the squaramide **5.9** (6.5 mM) was taken in a NMR tube, and the ^1H NMR spectrum acquired. To this was added 10 μl of a $CDCl_3$-CD_3CN (95:5) solution of TBDMS cytidine (65.7 mM) and the tube recapped. The resultant solution were thoroughly mixed, ensuring that the tube walls were completely wetted, and the ^1H NMR spectrum acquired at 30 °C. Sixteen further additions of TBDMS cytidine stock solution were made, up to a total added volume of 1750 μl. The resultant solution was mixed thoroughly and the ^1H NMR spectrum acquired after each addition. The chemical shift data was then analysed using HOSTEST to calculate association constants K_a and error limits (standard deviations).

Figure 5.1a: NMR Dimerisation Study of 5.8 in $CDCl_3$

following the NMR signal of the H-5 proton

Total concentration of Host / M	Chemical shift of H-5 proton / ppm
0.0419	5.9760
0.0335	5.9986
0.0280	6.0143
0.0224	6.0312
0.0186	6.0481
0.0160	6.0582
0.0112	6.0695

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer}$ = 6.1152 ppm, δ_{dimer} = -4.7181 ppm

$K_d = 2\ (\pm 2)\ M^{-1}$

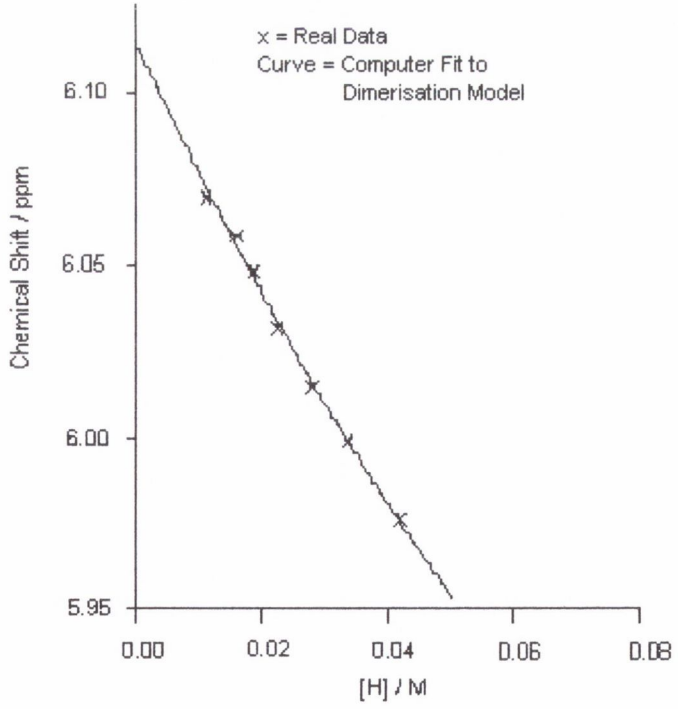


Figure 5.1b: NMR Dimerisation Study of 5.8 in $CDCl_3$
following the NMR signal of the H-6 proton

Total concentration of Host / M	Chemical shift of H-6 proton / ppm
0.0419	7.4393
0.0335	7.4518
0.0280	7.4606
0.0224	7.4706
0.0186	7.4794
0.0160	7.4857
0.0112	7.4919

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer} = 7.5180$ ppm, $\delta_{dimer} = 6.7734$ ppm

$K_d = 2\ (\pm 2)\ M^{-1}$

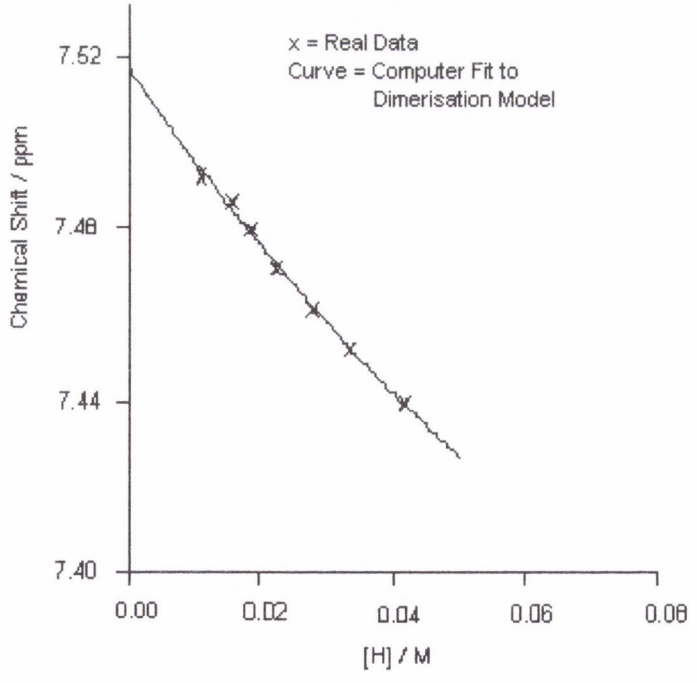


Figure 5.2a: NMR Binding Study of the Association of 5.8 with 5.7 in $CDCl_3$

following the NMR signal of the H-5 proton

Stock solutions: [H] = 0.0019 M, [G] = 0.0614 M

initial volume of host = 600 µl, δ_{Host} = 6.1806 ppm

K_d of Host = 2 M^{-1}

Total concentration of Guest / M	**Chemical shift of H-5 proton / ppm**
0.0002	6.1473
0.0004	6.1191
0.0008	6.0752
0.0012	6.0476
0.0020	6.0169
0.0024	6.0068
0.0027	5.9993
0.0033	5.9905
0.0038	5.9843
0.0047	5.9755
0.0064	5.9648
0.0088	5.9554

This graph shows real data from the table above plotted with a curve for fit to 1:1 model using the following (best fit) constants

K_a = 4831 (±3016) M^{-1}, $\delta_{Host\text{-}Guest}$ = 5.957 ppm

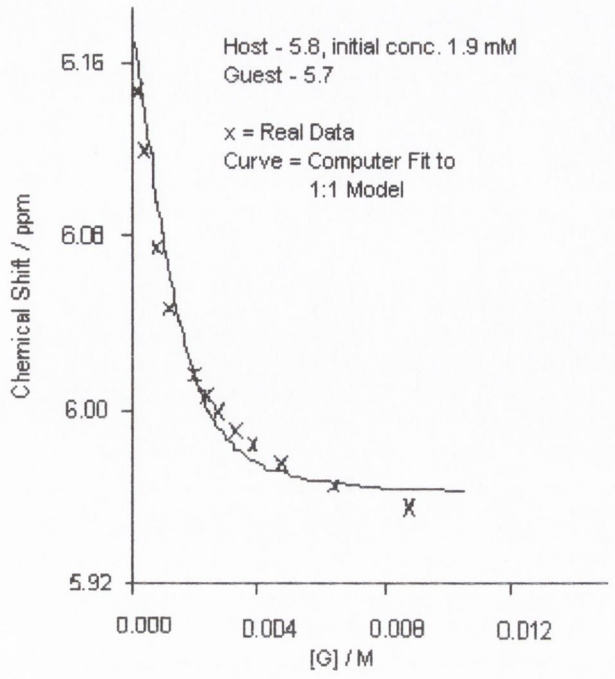


Figure 5.2b: NMR Binding Study of the Association of 5.8 with 5.7 in $CDCl_3$

following the NMR signal of the H-5 proton

Stock solutions: [H] = 0.0041 M, [G] = 0.0494 M

initial volume of host = 600 µl, δ_{Host} = 6.1774 ppm

K_d of Host = 2 M^{-1}

Total concentration of Guest / M	Chemical shift of H-5 proton / ppm
0.0005	6.1473
0.0010	6.1197
0.0017	6.0801
0.0025	6.0513
0.0032	6.0325
0.0039	6.0193
0.0046	6.0099
0.0053	6.0023
0.0059	5.9973
0.0066	5.9923
0.0071	5.9885
0.0099	5.9766
0.0145	5.9660
0.0165	5.9616

This graph shows real data from the table above plotted with a curve for fit to 1:1 model using the following (best fit) constants

$K_a = 2269\ (\pm 730)\ M^{-1}$, $\delta_{\text{Host-Guest}} = 5.9611$ ppm

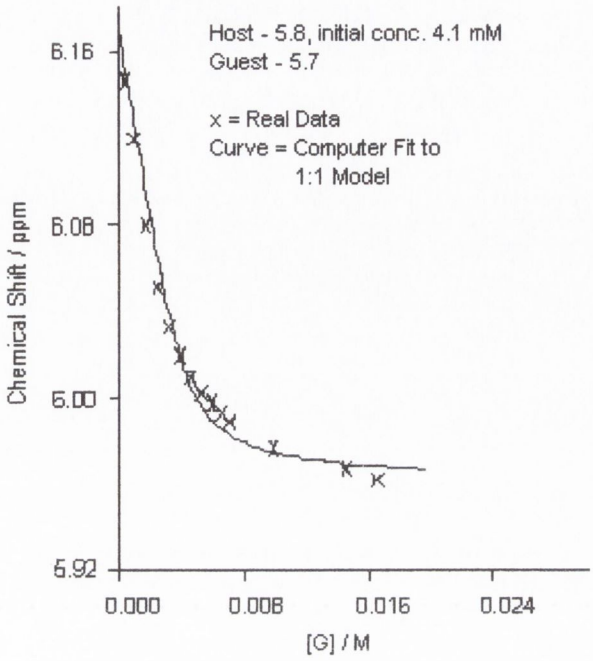


Figure 5.3a: NMR Binding Study of the Association of 5.7 with 5.8 in $CDCl_3$

following the NMR signal of the carbamoyl NH proton

Stock solutions: [H] = 0.0022 M, [G] = 0.0559 M

initial volume of host = 600 μl, δ_{Host} = 6.6132 ppm

K_d of guest = 2 M^{-1}

Total concentration of Guest / M	**Chemical shift of carbamoyl NH proton / ppm**
0.0002	6.7788
0.0004	6.9181
0.0006	7.0535
0.0011	7.4549
0.0015	7.7071
0.0018	7.8902
0.0022	8.1235
0.0030	8.4710
0.0035	8.6404
0.0043	8.8085
0.0058	9.0242
0.0080	9.1039
0.0112	9.1654
0.0140	9.1873

This graph shows real data from the table above plotted with a curve for fit to 1:1 model using the following (best fit) constants

K_a = 1192 (±93) M^{-1}, $\delta_{Host\text{-}Guest}$ = 9.4273 ppm

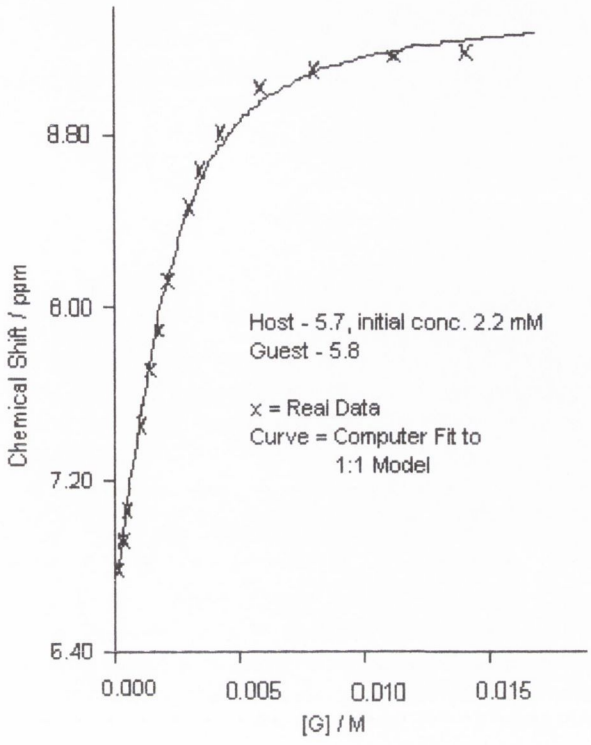

8.80
8.00
7.20
6.40
Chemical Shift / ppm
0.000
0.005
0.010
0.015
[G] / M
Host - 5.7, initial conc. 2.2 mM
Guest - 5.8
x = Real Data
Curve = Computer Fit to
1:1 Model

Figure 5.3b: NMR Binding Study of the Association of 5.7 with 5.8 in $CDCl_3$

following the NMR signal of the C-3 NH proton

Stock solutions: [H] = 0.0022 M, [G] = 0.0559 M

initial volume of host = 600 μl, δ_{Host} = 10.517 ppm

K_d of guest = 2 M^{-1}

Total concentration of Guest / M	Chemical shift of C-3 NH proton / ppm
0.0002	10.4919
0.0004	10.4693
0.0006	10.4455
0.0007	10.4279
0.0011	10.3878
0.0015	10.3577
0.0018	10.3288
0.0022	10.3012
0.0025	10.2812
0.0030	10.2599
0.0035	10.2410
0.0043	10.2272
0.0058	10.2147
0.0080	10.2185
0.0112	10.2297
0.0140	10.2360

This graph shows real data from the table above plotted with a curve for fit to 1:1 model using the following (best fit) constants

K_a = 3818 M^{-1}, $\delta_{Host\text{-}Guest}$ = 10.2059 ppm

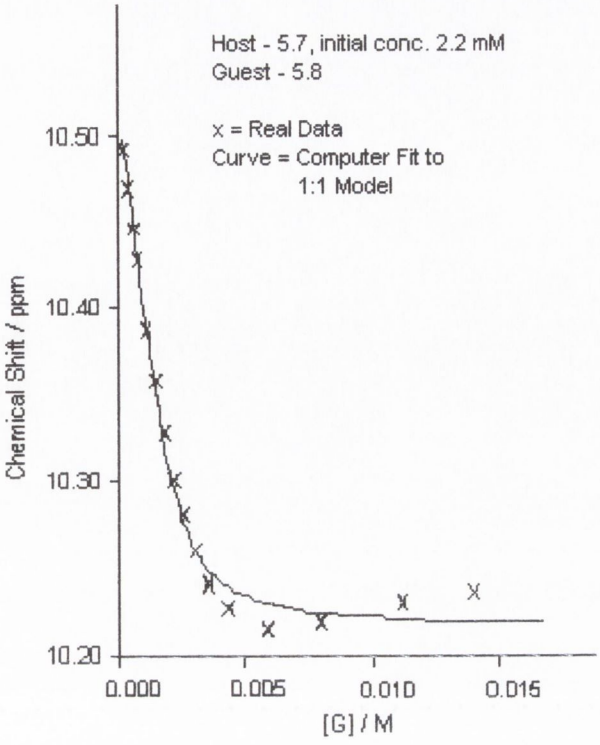

Host - 5.7, initial conc. 2.2 mM
Guest - 5.8
x = Real Data
Curve = Computer Fit to
1:1 Model
Chemical Shift / ppm
10.50
10.40
10.30
10.20
0.000
0.005
0.010
0.015
[G] / M

Figure 5.5: NMR Dimerisation Study of 5.9 in $CDCl_3$
following the NMR signal of the C-3 NH proton

Total concentration of Host / M	Chemical shift of C-3 NH proton / ppm
0.0002	10.4016
0.0004	10.4329
0.0009	10.4819
0.0017	10.5032
0.0032	10.5152
0.0059	10.5233
0.0100	10.5320
0.0154	10.5358

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer}$ = 6.9465 ppm, δ_{dimer} = 10.5567 ppm

K_d = 1125258 (±54420311) M^{-1}

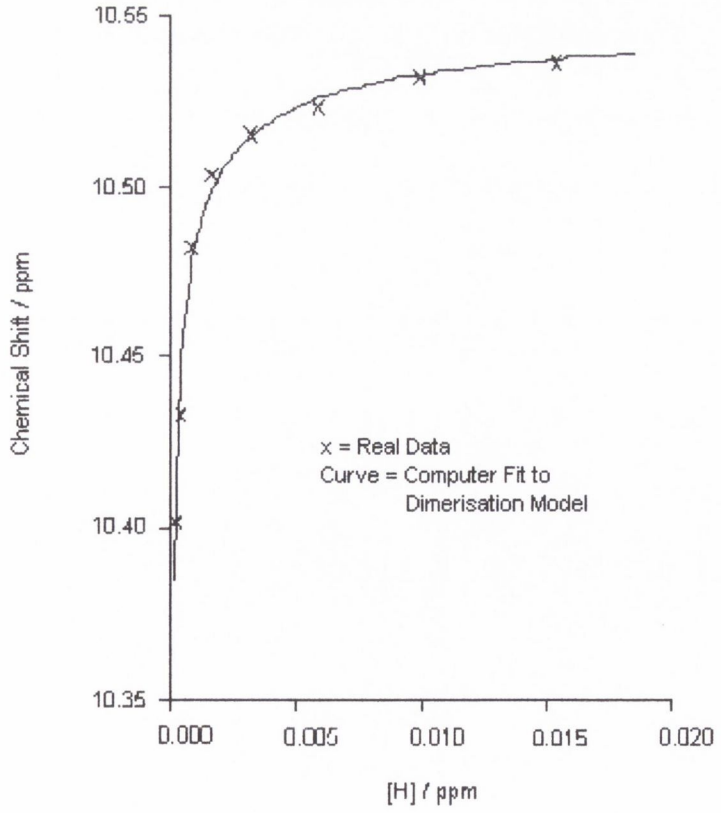


Figure 5.6: NMR Titration Study of 5.9 in $CDCl_3$ with CD_3CN

Initial concentration of **5.9** = 0.0044 M, Final concentration of **5.9** = 0.0026 M

Percentage (v/v) $CD_3CN/CDCl_3$	Change in chemical shift of C-3 NH proton / ppm	Change in chemical shift of Carbamoyl NH proton / ppm	Change in chemical shift of C-4 NH proton / ppm
0.1815	-0.0113	-0.0088	-0.0062
0.3623	-0.0251	-0.0138	-0.0113
0.7220	-0.0477	-0.0226	-0.0150
1.0791	-0.0690	-0.0364	-0.0213
1.4337	-0.0966	-0.0376	-0.0251
1.7857	-0.1179	-0.0527	-0.0401
2.6549	-0.1719	-0.0702	-0.0677
3.5088	-0.2233	-0.0891	-0.0815
5.1724	-0.3350	-0.1405	-0.1166
6.7797	-0.4378	-0.1643	-0.1518
8.3333	-0.5407	-0.1580	-0.1982
9.8361	-0.6385	-0.1819	-0.2333
11.2903	-0.7351	-0.1944	-0.2534
12.6984	-0.8242	-0.2107	-0.2797
14.7287	-0.9546	-0.2408	-0.3211
16.6667	-1.0675	-0.2597	-0.3575
18.5185	-1.1754	-0.2960	-0.3901
20.2899	-1.2745	-0.3387	-0.4214
22.5352	-1.3874	-0.3713	-0.4566
24.6575	-1.4953	-0.4139	-0.4892
28.5714	-1.6772	-0.4930	-0.5506
32.0988	-1.8164	-0.5494	-0.5945
38.2022	-2.0171	-0.6285	-0.6611
42.1053	-2.1200	-0.6736	-0.6974

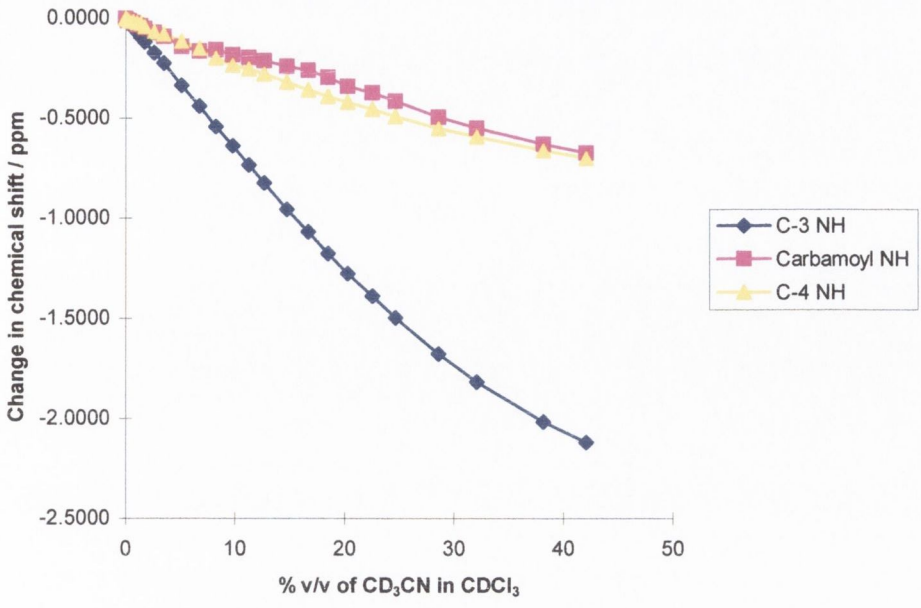

0.0000
-0.5000
-1.0000
-1.5000
-2.0000
-2.5000
Change in chemical shift / ppm
0
10
20
30
40
50
% v/v of CD3CN in CDCl3
C-3 NH
Carbamoyl NH
C-4 NH

Figure 5.7: NMR Dimerisation Study of 5.7 in 5% (v/v) $CD_3CN/CDCl_3$ following the NMR signal of the C-3 NH proton

Total concentration of Host / M	Chemical shift of C-3 NH proton / ppm
0.0002	9.4294
0.0005	9.6401
0.0009	9.8785
0.0018	10.0315
0.0034	10.1507
0.0062	10.2285
0.0105	10.2849
0.0162	10.3276
0.0249	10.3501

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer}$ = 8.0687 ppm, δ_{dimer} = 10.4919 ppm

K_d = 6116 (±1207) M^{-1}

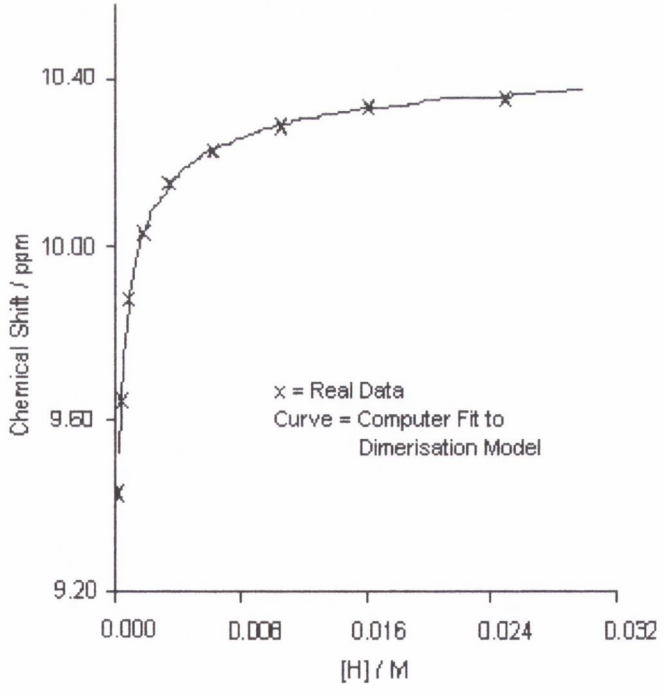


Figure 5.8a: NMR Dimerisation Study of 5.9 in 5% (v/v) $CD_3CN/CDCl_3$ following the NMR signal of the C-3 NH proton

Total concentration of Host / M	Chemical shift of C-3 NH proton / ppm
0.0002	9.4407
0.0004	9.6527
0.0009	9.8835
0.0017	10.0428
0.0033	10.1356
0.0059	10.2335
0.0101	10.2862
0.0155	10.3338
0.0243	10.3577

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer} = 8.1766$ ppm, $\delta_{dimer} = 10.4953$ ppm

$K_d = 5778\ (\pm 1057)\ M^{-1}$

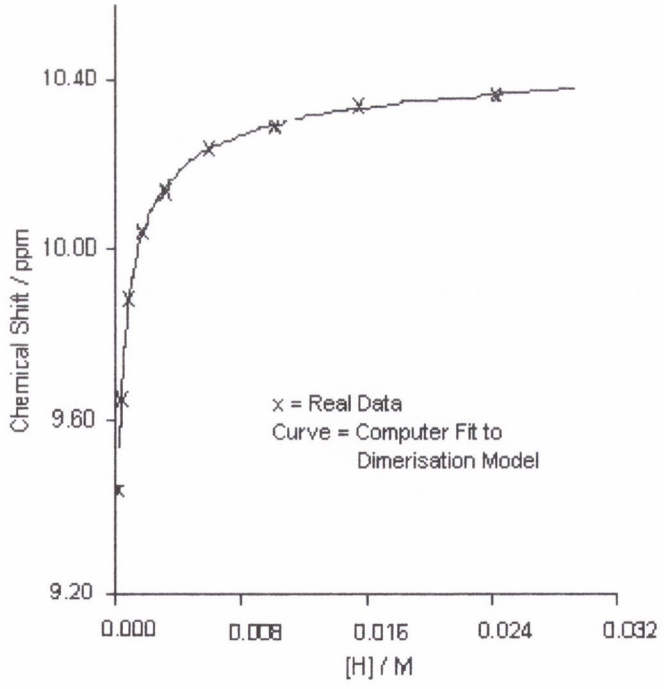


Figure 5.8b: NMR Dimerisation Study of 5.9 in 5% (v/v) $CD_3CN/CDCl_3$ following the NMR signal of the carbamoyl NH proton

Total concentration of Host / M	Chemical shift of Carbamoyl NH proton / ppm
0.0002	6.2570
0.0004	6.2996
0.0009	6.3774
0.0017	6.4175
0.0033	6.4464
0.0059	6.4702
0.0101	6.4827
0.0155	6.4978
0.0243	6.5041

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer} = 5.9536$ ppm, $\delta_{dimer} = 6.5443$ ppm

$K_d = 4577\ (\pm 1814)\ M^{-1}$

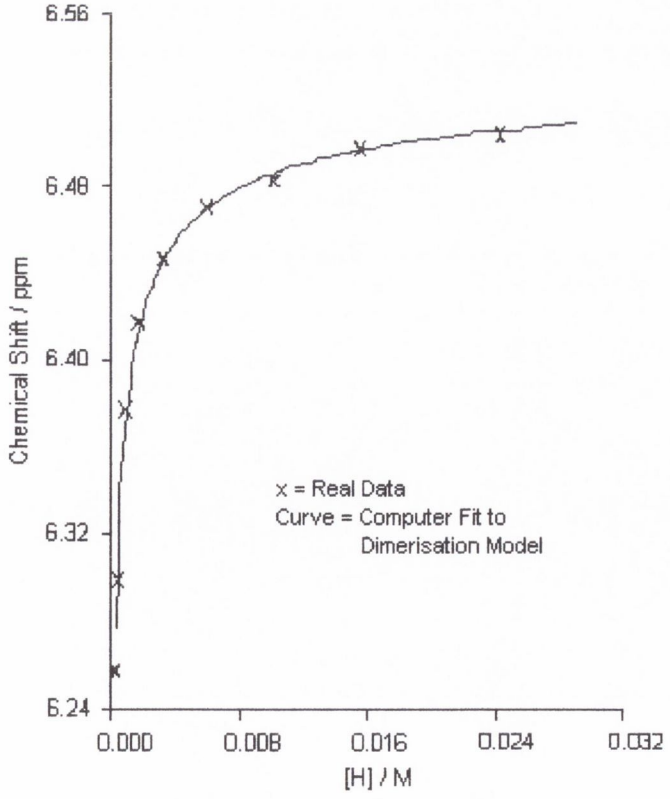


Figure 5.9: NMR Dimerisation Study of 5.9 in 1% (v/v) DMSO-d_6/$CDCl_3$ following the NMR signal of the C-3 NH proton

Total concentration of Host / M	Chemical shift of C-3 NH proton / ppm
0.0002	9.8271
0.0004	9.8659
0.0008	9.9224
0.0015	9.9889
0.0029	10.0679
0.0053	10.1545
0.0090	10.2197
0.0138	10.2586

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer} = 9.7815$ ppm, $\delta_{dimer} = 10.5367$ ppm

$K_d = 176\ (\pm 13)\ M^{-1}$

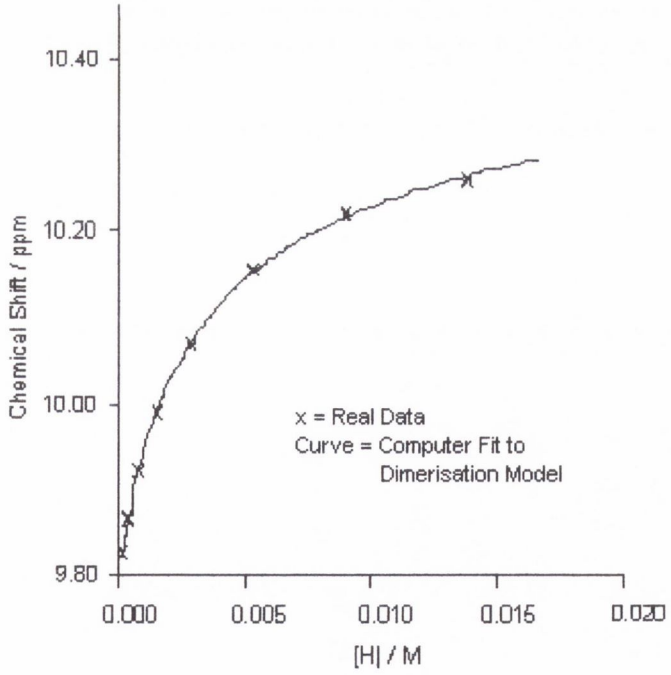


Figure 5.11a: NMR Dimerisation Study of 5.12 in 5% (v/v) $CD_3CN/CDCl_3$ following the NMR signal of the C-3 NH proton

Total concentration of Host / M	Chemical shift of C-3 NH proton / ppm
0.0003	9.9425
0.0005	10.2046
0.0010	10.4129
0.0020	10.5634
0.0037	10.6788
0.0068	10.7691
0.0116	10.8155
0.0177	10.8569
0.0242	10.8783

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer} = 8.0230$ ppm, $\delta_{dimer} = 11.0115$ ppm

$K_d = 9818\ (\pm 897)\ M^{-1}$

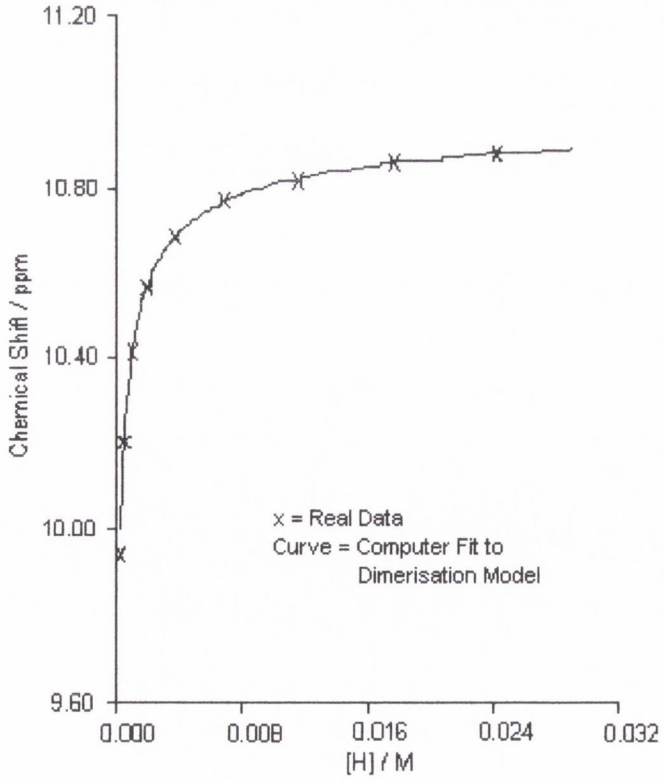


Figure 5.11b: NMR Dimerisation Study of 5.12 in 5% (v/v) $CD_3CN/CDCl_3$ following the NMR signal of the carbamoyl NH proton

Total concentration of Host / M	Chemical shift of Carbamoyl NH proton / ppm
0.0003	7.8927
0.0005	7.9554
0.0010	8.0069
0.0020	8.0470
0.0037	8.0759
0.0068	8.0985
0.0116	8.1122
0.0177	8.1235
0.0242	8.1286

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer}$ = 7.5423 ppm, δ_{dimer} = 8.1647 ppm

K_d = 5761 (±326) M^{-1}

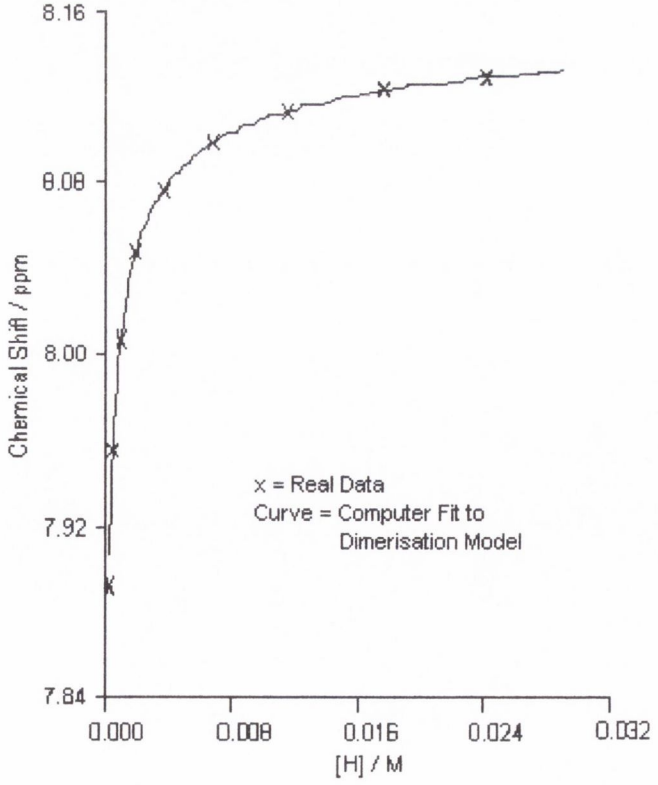


Figure 5.12a: NMR Dimerisation Study of 5.13 in 5% (v/v) $CD_3CN/CDCl_3$ following the NMR signal of the C-3 NH proton

Total concentration of Host / M	Chemical shift of C-3 NH proton / ppm
0.0002	9.7844
0.0005	9.9023
0.0009	10.0804
0.0017	10.2812
0.0032	10.3903
0.0054	10.4668
0.0084	10.5170
0.0117	10.5421
0.0135	10.5458

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer} = 9.3641$ ppm, $\delta_{dimer} = 10.7981$ ppm

$K_d = 1214\ (\pm 369)\ M^{-1}$

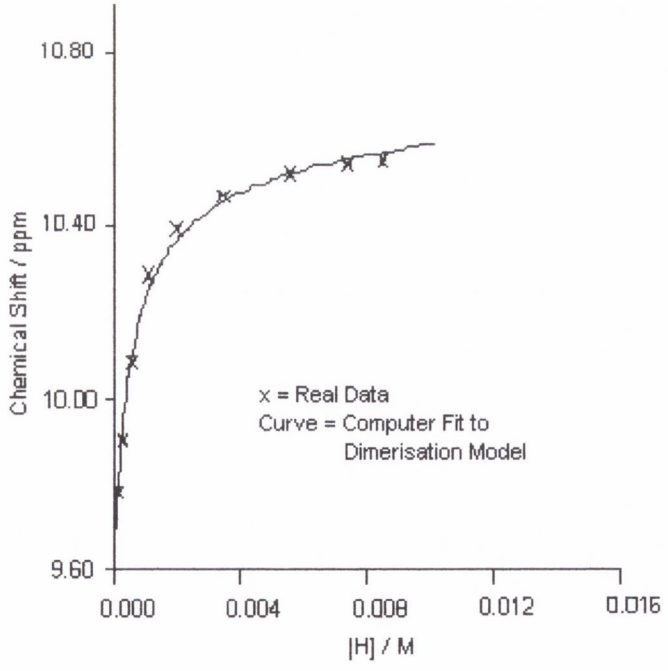


Figure 5.12b: NMR Dimerisation Study of 5.13 in 5% (v/v) $CD_3CN/CDCl_3$ following the NMR signal of the carbamoyl NH proton

Total concentration of Host / M	Chemical shift of Carbamoyl NH proton / ppm
0.0002	7.9655
0.0005	8.0307
0.0009	8.0834
0.0017	8.1386
0.0032	8.1712
0.0054	8.1913
0.0084	8.2051
0.0117	8.2126
0.0135	8.2139

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer}$ = 7.6867 ppm, δ_{dimer} = 8.2731 ppm

K_d = 3856 (±1198) M^{-1}

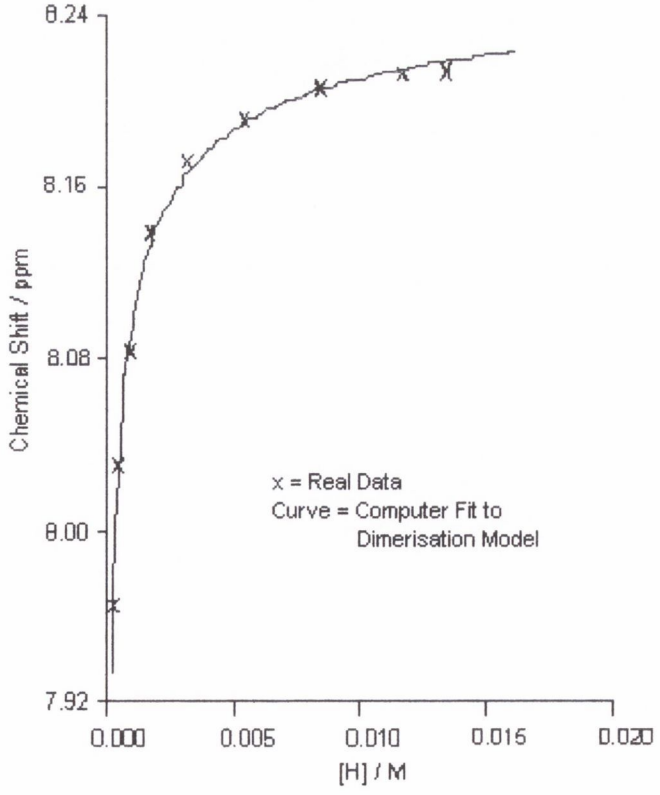


Figure 5.13: NMR Dimerisation Study of 5.15 in 5% (v/v) $CD_3CN/CDCl_3$
following the NMR signal of the C-3 NH proton

Total concentration of Host / M	Chemical shift of C-3 NH proton / ppm
0.0002	8.9816
0.0005	9.0844
0.0009	9.2174
0.0018	9.3930
0.0034	9.6640
0.0062	9.9111
0.0106	10.1156
0.0162	10.2523
0.0203	10.3225
0.0243	10.3840

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer}$ = 8.8667 ppm, δ_{dimer} = 11.1804 ppm

$K_d = 116\ M^{-1}$

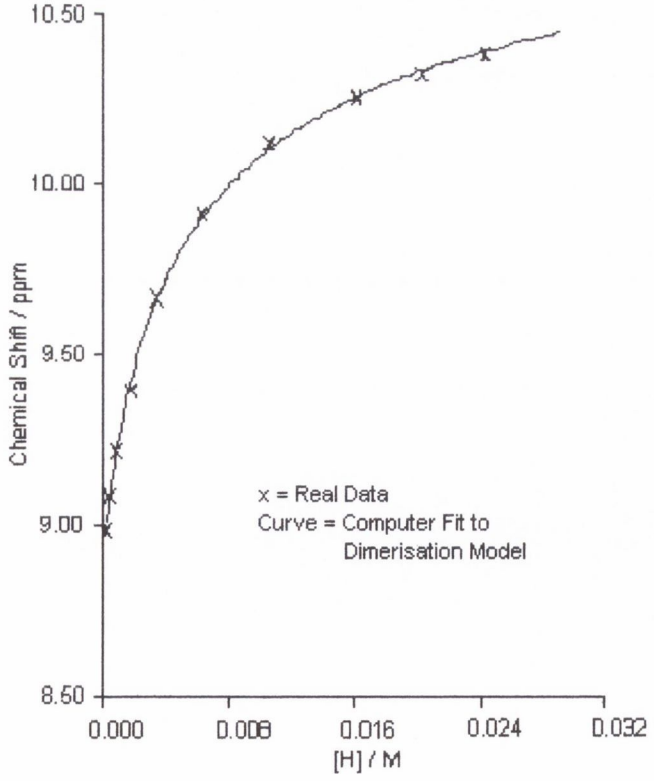


Figure 5.15: NMR Dimerisation Study of TBDMS Cytidine in 5% (v/v) $CD_3CN/CDCl_3$
following the NMR signal of the H-6 proton

Total concentration of Host / M	Chemical shift of H-6 proton / ppm
0.0017	8.1405
0.0034	8.1392
0.0065	8.1330
0.0122	8.1242
0.0219	8.1141
0.0270	8.1085
0.0314	8.1022
0.0354	8.0978
0.0390	8.0953
0.0451	8.0909
0.0502	8.0865
0.0581	8.0803
0.0664	8.0740
0.0722	8.0677
0.0776	8.0652

This graph shows real data from the table above plotted with a curve for ideal dimerisation using the following (best fit) constants

$\delta_{monomer}$ = 8.1433 ppm, δ_{dimer} = 7.7766 ppm

$K_d = 2\ M^{-1}$

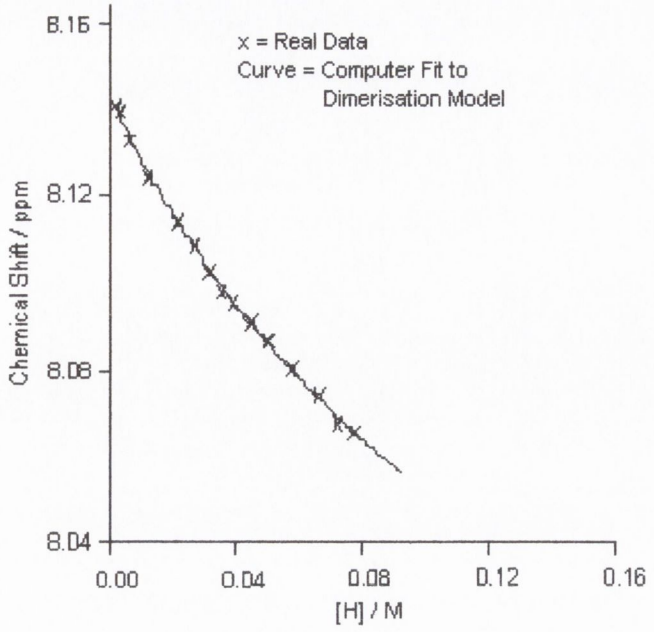


Figure 5.16a: NMR Binding Study of the Association of 5.9 with TBDMS Cytidine in 5% (v/v) $CD_3CN/CDCl_3$

following the NMR signal of the carbamoyl NH proton

Stock solutions: [H] = 0.0065 M, [G] = 0.0657 M

initial volume of host = 600 µl, δ_{Host} = 5.9038 ppm

K_d of host = 5778 M^{-1}, K_d of guest = 2 M^{-1}

Total concentration of Guest / M	Chemical shift of C-3 NH proton / ppm
0.0011	6.9494
0.0021	7.3746
0.0031	7.8037
0.0051	8.4811
0.0060	8.7432
0.0069	8.8875
0.0082	9.1308
0.0094	9.2199
0.0106	9.2839
0.0117	9.3203
0.0138	9.3604
0.0164	9.3867
0.0219	9.4118
0.0299	9.4181
0.0394	9.4244
0.0489	9.4419

This graph shows real data from the table above plotted with a curve for fit to 1:1 model using the following (best fit) constants

$K_a = 10449\ (\pm 540)\ M^{-1}$, $\delta_{Host\text{-}Guest} = 9.4260$ ppm

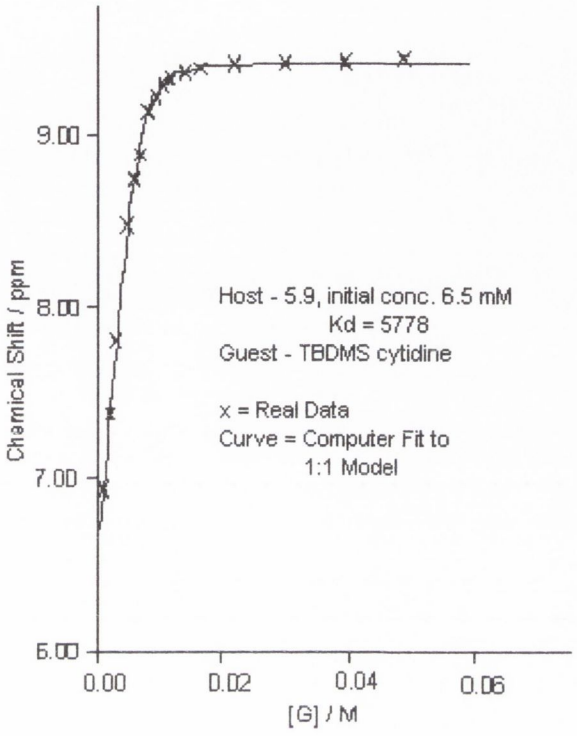


Figure 5.16b: NMR Binding Study of the Association of 5.9 with TBDMS Cytidine in 5% (v/v) $CD_3CN/CDCl_3$

following the NMR signal of the C-3 NH proton

Stock solutions: [H] = 0.0065 M, [G] = 0.0657 M

initial volume of host = 600 µl, δ_{Host} = 8.1766 ppm

K_d of host = 5778 M^{-1}, K_d of guest = 2 M^{-1}

Total concentration of Guest / M	Chemical shift of C-3 NH proton / ppm
0.0011	10.2310
0.0021	10.2322
0.0031	10.2372
0.0051	10.2586
0.0060	10.2736
0.0069	10.2824
0.0082	10.3075
0.0094	10.3200
0.0106	10.3313
0.0117	10.3389
0.0138	10.3501
0.0164	0.3577
0.0219	10.3690
0.0299	10.3715
0.0394	10.3765

This graph shows real data from the table above plotted with a curve for fit to 1:1 model using the following (best fit) constants

K_a = 6877±1351 M^{-1}, $\delta_{Host\text{-}Guest}$ = 10.3808 ppm

Chapter Seven

References

REFERENCES

1.1 Eugeny A. Lukhtanov, Alan G. Mills, Igor V. Kutyavin, Vladimir V. Gorn, Michael W. Reed and Rich B. Meyer, *Nucleic Acids Res.*, 1997, **25**, 5077.

1.2 Mark Matteucci, Kuei-Ying Lin, Teresa Huang, Richard Wagner, Daniel D. Sternbach, Mukund Mehrotra and Jeffrey M. Besterman, *J. Am. Chem. Soc.*, 1997, **119**, 6939; Evgeniy S. Belousov, Irina A. Afonina, Mikhail A. Podyminogin, Howard B. Gamper, Michael W. Reed, Pobert M. Wydro and Rich B. Meyer, *Nucleic Acids Res.*, 1997, **25**, 3440.

1.3 C. Hélène and J. J. Toulmé, *Oligodeoxynucleotides: Antisense inhibitors of gene expression*, Mc Millian Press, London, 1989, p 137-172.

1.4 Neidle S., Laurence H. Pearl and Jane V. Skelly, *Biochem. J.*, 1987, **243**, 1.

1.5 Gonzalo Colmenarejo, Francisco Montero and Guillermo Orellana, *Anti-Cancer Drug Design*, 1997, **12**, 239.

1.6 Achim H. Krotz, Louis Y. Kuo, Thomas P. Shields and Jacqueline K. Barton, *J. Am. Chem. Soc.*, 1993, **115**, 3877.

1.7 Achim H. Krotz, Brian P. Hudson and Jacqueline K. Barton, *J. Am. Chem. Soc.*, 1993, **115**, 12577.

1.8 Brian P. Hudson, Cynthia M. Dupureur and Jacqueline K. Barton, *J. Am. Chem. Soc.*, 1995, **117**, 9379.

1.9 Robert H. Terbrueggen and Jacqueline K. Barton, *Biochemistry*, 1995, **34**, 8227.

1.10 Robert H. Terbrueggen, Timothy W. Johann and Jacqueline K. Barton, *Inorg. Chem.*, 1998, **37**, 6874.

1.11 Sonya J. Franklin and Jacqueline K. Barton, *Biochemistry*, 1998, **37**, 16093.

1.12 Miquel Coll, Juan Aymami, Gijs A. van der Marel, Jaques H van Boom, Alexander Rich and Andrew H. -J. Wang, *Biochemistry*, 1989, **28**, 310; Miquel Coll, Christin A. Frederick, Andrew H. -J. Wang and Alexander Rich, *Proc. Natl. Acad. Sci. U.S.A.*, 1987, **84**, 8385; Mai-kun Teng, Nassim Usman, Christin A. Frederick and Andrew H. -J. Wang, *Nucl. Acids Res.*, 1988, **16**, 2671.

1.13 J. William Lown, Krzysztof Krowicki, U. Ganapathi Bhat, Andrew Skorobogaty, Brian Ward and James C. Dabrowiak, *Biochemistry*, 1986, **25**, 7408.

1.14 Warren S. Wade, Milan Mrksich and Peter B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 8783.

1.15 Milan Mrksich, Michelle E. Parks and Peter B. Dervan, *J. Am. Chem. Soc.*, 1994, **116**, 7983.

1.16 Michelle E. Parks, Eldon E. Baird and Peter B. Dervan, *J. Am. Chem. Soc.*, 1996, **118**, 6153.

1.17 James J. Kelly, Eldon E. Baird and Peter B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.*, 1996, **93**, 6981.

1.18 Clara L. Kielkopf, Eldon E. Baird, Peter B. Dervan and Douglas C. Rees, *Nature, Struct. Biol.*, 1998, **5**, 104.

1.19 John W. Trauger, Eldon E. Baird Milan Mrksich and Peter B. Dervan, *J. Am. Chem. Soc.*, 1996, **118**, 6160.

1.20 Rafael Peláez Lamamie de Clairac, Christian J. Seel, Bernhard H. Geierstanger, Milan Mrksich, Eldon E. Baird Peter B. Dervan and David E. Wemmer, *J. Am. Chem. Soc.*, 1999, **121**, 2956.

1.21 John W. Trauger, Eldon E. Baird and Peter B. Dervan, *Angew. Chem. Internat. Edit.*, 1998, **37**, 1421.

1.22 John W. Trauger, Eldon E. Baird and Peter B. Dervan, *J. Am. Chem. Soc.*, 1998, **120**, 3534.

1.23 Sarah White, Jason W. Szewczyk, James M. Turner Eldon E. Baird and Peter B. Dervan, *Nature*, 1998, **391**, 468.

1.24 Clara L. Kielkopf, Sarah White, Jason W. Szewczyk, James M. Turner, Eldon E. Baird, Peter B. Dervan and Douglas C. Rees, *Science*, 1998, **282**, 111.

1.25 G. Felsenfeld, David R. Davies and A. Rich, *J. Am. Chem. Soc.*, 1957, **79**, 2023.

1.26 Heinze E. Moser and Peter B. Dervan, *Science*, 1987, **238**, 645; Trung Le Doan, Loic Perrouault, Danièle Praseuth, Noureddine Habhoub, Jean-Luc Decout, Nguyen T. Thuong, Jean Lhomme and Claude Hélène, *Nucleic Acids Res.*, 1987, **15**, 7749.

1.27 Wendy M. Olivas and L. James Maher, *Biochemistry*, 1995, **3**, 278.

1.28 Jeremy S. Lee, Mary L. Woodsworth, Laura J. P. Latimer and A. Richard Morgan, *Nucleic Acids Res.*, 1984, **12**, 6603

1.29 Thomas Povsic and Peter B. Dervan, *J. Am. Chem. Soc.*, 1989, **111**, 3059; Luigi E. Xodo, Giorgio Manzini, Franco Quadrifoglio, Gijs, A. van der Marel and Jacques H. van Boom, *Nucleic Acids Res.*, 1991, **19**, 5625.

1.30 Jong Sung Koh and Peter B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 1470.

1.31 Paul S. Miller, Purshotam Bahn, Cynthia D. Cushman and Tina L. Trapane, *Biochemistry*, 1992, **31**, 6788; Michele C. Jetter and Frank W. Hobbs, *Biochemistry*, 1993, **32**, 3249.

1.32 Paul S. Miller, Guixia Bi, Sarah A. Kipp, Victor Fok and Robert K. DeLong, *Nucleic Acids Res.*, 1996, **24**, 730.

1.33 Guobing Xiang, Robert Bogacki and Larry W. McLaughlin, *Nucleic Acids Res.*, 1996, **24**, 1963.

1.34 Darren M. Gowers and Keith R. Fox, *Nucleic Acids Res.*, 1997, **25**, 3787.

1.35 Diego A. Gianolio and Larry W. McLaughlin, *J. Am. Chem. Soc.*, 1999, **121**, 6334.

1.36 T. Sudhakar Rao, Ross H. Durland, Dale M. Seth, Melissa A. Myrick, Veeraiah Bodepudi and Ganapathi R. Revankar, *Biochemistry*, 1995, **34**, 765.

1.37 Wendy M. Olivas and L. James Maher, *Nucleic Acids Res.*, 1995, **23**, 1936.

1.38 William A Greenberg and Peter B. Dervan, *J. Am. Chem. Soc.*, 1995, **117**, 5016.

1.39 Chin-Yi Huang, Cynthia D. Cushman and Paul S. Miller, *J. Org. Chem.*, 1993, **58**, 5048.

1.40 Chin-Yi Huang, Guixia Bi and Paul S. Miller, *Nucleic Acids Res.*, 1996, **24**, 2606.

1.41 Steven C. Zimmerman and Philippe Schmitt, *J. Am. Chem. Soc.*, 1995, **117**, 10769.

1.42 Shigeki Sasaki, Shouji Nakashima, Fumi Nagatsugi, Yoshitsugu Tanaka, Masao Hisatome and Minoru Maeda, *Tet. Lett.*, 1995, **36**, 9521.

1.43 Linda C. Griffin, Laura L. Kiesling, Peter A. Beal, Paul Gillespie and Peter B. Dervan, *J. Am Chem. Soc.*, 1992, **114**, 7976.

1.44 Karl M. Koshlap, Paul Gillespie, Peter B. Dervan and Juli Feigon, *J. Am. Chem. Soc.*, 1993, **115**, 7908.

1.45 Edmond Wang, Karl M. Koshlap, Paul Gillespie, Peter B. Dervan and Juli Feigon, *J. Mol. Biol.*, 1996, **257**, 1052.

1.46 Thomas E. Lehmann, William A. Greenberg, David A. Liberles, Carol K. Wada and Peter P Dervan, *Helv. Chim. Acta*, 1997, **80**, 2003.

1.47 Jeffrey H. Rothman and W. Graham Richards, *Biopolymers*, 1996, **39**, 795.

1.48 Xavier Doisy, Postdoctoral Report, University of Dublin, 1996.

1.49 Michael Egholm, Ole Buchardt, Peter E. Nielsen and Rolf H. Berg, *J. Am. Chem. Soc.*, 1992, **114**, 1895; Anusch Peyman, Eugen Uhlmann, Konrad Wagner, Sascha Augustin, Gerhard Breipohl, David W. Will, Andrea Schäfer and Holger Wallmeier, *Angew. Chem. Internat. Edit.*, 1996, **35**, 2636.

1.50 R. Prohens, S. Tomàs, J. Morey, P. M. Deyà, P. Ballester and A. Costa, *Tet. Lett.*, 1998, **39**, 1063; S. Tomàs, R. Prohens, M. Vega, M. C. Rotger, P. M. Deyà, P. Ballester and A. Costa, *J. Org. Chem.*, 1996, **61**, 9394.

2.1 Francis Outurquin and Claude Paulmier, *Bull. Soc. Chim. Fr.*, 1983, 153.

2.2 Francisco Garcia and Carmen Gálvez, *Synthesis*, 1985, 143.

2.3 David Wensbo, Ann Eriksson, Torsten Jeschke, Ulf Annby and Salo Gronowitz, *Tet. Lett.*,1993, **34**, 2823.

2.4 Ralph Mozingo, Stanton A. Harris, Donald E. Wolf, Charles E. Hoffhine, Nelson R. Easton and Karl Folkers, *J. Am. Chem. Soc.*, 1945, **67**, 2092.

2.5 Emery W. Brunett and Walter C. McCarthy, *J. Pharm. Sci.*, 1968, **57**, 2003.

2.6 I. J. Rinkes, *Rec. Trav. Chim.*, 1934, **54**, 643.

2.7 Synthesised by Xavier Doisy by a method adapted from Kenji Mori and Yuji Funaki, *Tetrahedron*, 1985, **41**, 2369.

2.8 Kenji Mori and Yuji Funaki, *Tetrahedron*, 1985, **41**, 2369.

2.9 P. Biddle, E. S. Lane and J. L. Willans, *J. Chem. Soc.*, 1960, 2369; R. Rastogi, S. Sharma and R. N. Iyer, *Eur. J. Med. Chem.*, 1979, **14**, 489.

2.10 B. Adcock, Alexander Lawson and D. H. Miles, *J. Chem. Soc.*, 1961, 5120; Stefan Weiss, Horst Michaud, Horst Prietzel and Helmut Krommer, *Angew. Chem. Internat. Edit.*, 1973, **12**, 841.

2.11 Jürgen Schulze, Hartmut Tanneberg and Hermann Matschiner, *Z. Chem.*, 1980, **20**, 436.

2.12 Jésus Anatol and Jean Berecoechea, *Bull. Soc. Chim. Fr.*, 1975, 395.

2.13 A. D. Ainley, F. H. S. Curd and F. L. Rose, *J. Chem. Soc.*, 1949, 98.

2.14 Stefan Weiss and Helmut Krommer, *Angew. Chem. Internat. Edit.*,1974, **13**, 546.

3.1 Molecular modelling employed MacroModel ver.5.5, run on a silicon graphics workstation. Energy minimisation carried out using AMBER* force field with PRCG (Polak-Ribiere Conjugate Gradient) method.

3.2 Akira Takamizawa, Kentaro Hirai and Keiko Matsui, *Bull. Chem. Soc. Japan*, 1963, **36**, 1214.

3.3 Leif Grehn, *J. Heterocyclic Chem.*, 1978, **15**, 81.

3.4 Roland K. Robins, Ganapathi R. Revankar, Darrell E. O'Brien, Robert H. Springer, Thomas Novinson, Anthony Albert, Keitaro Senga, Jon P. Miller and David G. Streeter, *J. Heterocyclic Chem.*, 1985, **22**, 601.

3.5 Ned D. Heindel, Peter D. Kennewell and Velmer B. Fish, *J. Heterocyclic Chem.*, 1969, **6**, 77.

3.6 David Wensbo, Ulf Annby and Salo Gronowitz, *Tetrahedron*, 1995, **51**, 10323.

3.7 J. Edward Semple, Pen C. Wang, Zenon Lysenko and Madeleine M. Joullié, *J. Am. Chem. Soc.*, 1980, **102**, 7505.

3.8 George F. Field, William J. Zally, Leo H. Sternbach and John F. Blount, *J. Org. Chem.*, 1976, **41**, 3853.

3.9 George F. Field, *J. Org. Chem.*, 1978, **43**, 1084.

3.10 Moohi Yoo Kim, John E. Starrett and Steven M. Weinreb, *J. Org. Chem.*, 1981, **46**, 5383.

4.1 Amos B. Smith, Thomas A. Rano, Noritaka Chida, Gary A. Sulikowski and John L. Wood, *J. Am. Chem. Soc.*, 1992, **114**, 8008.

4.2 David R. Kronenthal, C. Y. Han and Martha K. Taylor, *J. Org. Chem.*, 1982, **47**, 2765.

5.1 Lanny S. Liebeskind, Richard W. Fengl, K. R. Wirtz, and T. T. Shawe, *J. Org. Chem.*, 1988, **53**, 2482.

5.2 Lanny S. Liebeskind, Marvin S. Yu, and Richard W. Fengl, *J. Org. Chem.*, 1993, **58**, 3543.

5.3 C. S. Wilcox, in *Frontiers in Supramolecular Chemistry and Photochemistry*, ed. H.-J. Schneider and H. Durr, VCH, Weinheim, 1990, p. 123.

5.4 Kelvin K. Ogilvie, Aria L. Schifman and Christopher L. Penney, *Can. J. Chem.*, 1979, **57**, 2230.

5.5 Y. Kyogoku, R. C. Lord and A. Rich, *Biochim. Biophys. Acta.*, 1969, **179**, 10.

6.1 D. D. Perrin, W. L. F. Armarego and D. R. Perrin, in *Purification of Laboratory Chemicals*, Pergamon Press, Oxford 1980.

6.2 W. C Still, M. Kahn, A. Mitra, *J. Org. Chem.*, 1978, **43**, 2923.

6.3 Rick L. Danheiser, Eric J. Stoner, Hiroo Koyama, Dennis S. Yamashita, Carrie A. Klade, *J.Am.Chem.Soc.*, 1989; **111**; 4407.

Appendix A

X-Ray Crystallography Tables

APPENDIX A: X-RAY CRYSTALLOGRAPHY TABLES

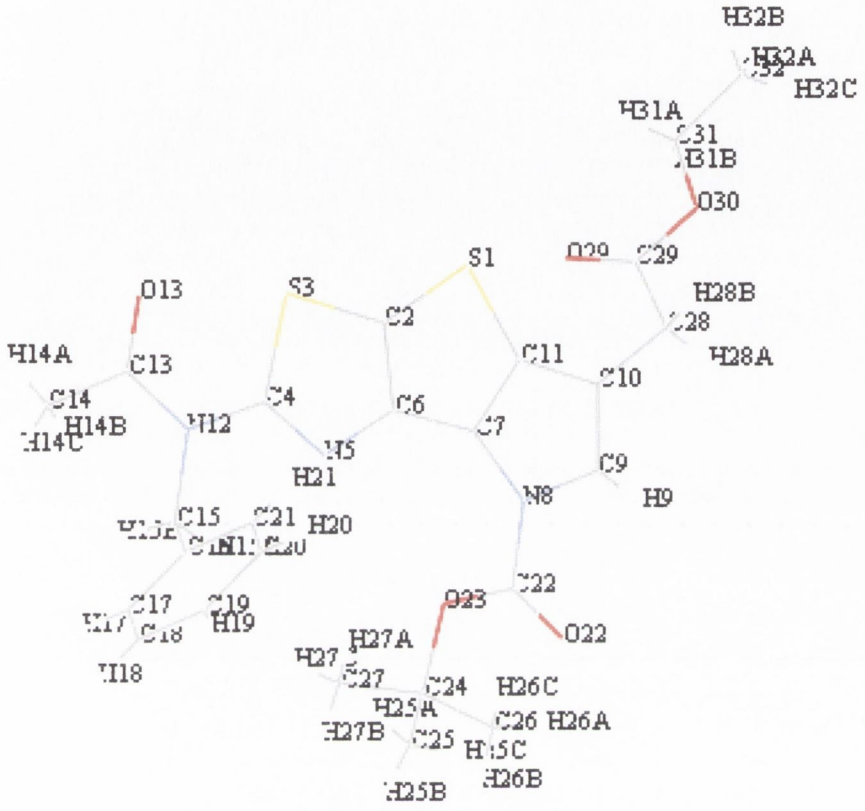


Table 1. Crystal data and structure refinement for $C_{25}H_{27}N_3O_5S_2$ - **3.16**.

Identification code	ad001	
Empirical formula	$C_{25}H_{27}N_3O_5S_2$	
Formula weight	513.62	
Temperature	173(2) K	
Wavelength	0.71073 Å	
Crystal system	triclinic	
Space group	P –1 No. 2	
Unit cell dimensions	a = 10.4958(3) Å	α= 104.765(2)°.
	b = 11.5214(4) Å	β= 110.424(2)°.
	c = 11.6097(4) Å	γ = 90.230(2)°.
Volume	1265.51(7) Å^3	
Z	2	

Density (calculated)	1.348 Mg/m^3
Absorption coefficient	0.251 mm^{-1}
F(000)	540
Crystal size	0.20 x 0.10 x 0.10 mm^3
Theta range for data collection	2.92 to 27.89°.
Index ranges	$0 \leq h \leq 13, -15 \leq k \leq 15, -15 \leq l \leq 13$
Reflections collected	5908
Independent reflections	5908 [R(int) = 0.0000]
Completeness to theta = 27.89°	97.7 %
Max. and min. transmission	0.9753 and 0.9515
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	5908 / 0 / 321
Goodness-of-fit on F^2	1.198
Final R indices [I>2sigma(I)]	R1 = 0.0616, wR2 = 0.1397
R indices (all data)	R1 = 0.0837, wR2 = 0.1478
Largest diff. peak and hole	0.366 and -0.391 $e.Å^{-3}$

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3) for ad001. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
C(2)	6888(3)	8524(3)	4194(3)	27(1)
C(4)	4939(3)	9518(3)	3287(3)	25(1)
C(6)	7057(3)	9768(2)	4634(3)	24(1)
C(7)	8402(3)	10147(2)	5581(3)	25(1)
C(9)	10492(3)	10832(3)	7100(3)	31(1)
C(10)	10467(3)	9614(3)	6761(3)	29(1)
C(11)	9147(3)	9189(3)	5804(3)	27(1)
C(13)	2645(3)	8941(3)	1735(3)	33(1)
C(14)	1334(3)	9338(3)	974(3)	39(1)
C(15)	3472(3)	11119(3)	2782(3)	30(1)
C(16)	3517(3)	11602(3)	1702(3)	28(1)
C(17)	2568(3)	12373(3)	1252(3)	33(1)
C(18)	2616(4)	12847(3)	278(3)	42(1)
C(19)	3598(4)	12539(3)	-257(3)	41(1)
C(20)	4528(3)	11771(3)	175(3)	40(1)
C(21)	4495(3)	11303(3)	1150(3)	36(1)

C(22)	9081(3)	12424(3)	6582(3)	31(1)
C(24)	7573(3)	13832(3)	5722(3)	37(1)
C(25)	7492(4)	14508(3)	6986(4)	51(1)
C(26)	8621(4)	14438(3)	5373(4)	48(1)
C(27)	6173(4)	13611(3)	4662(4)	51(1)
C(28)	11603(3)	8890(3)	7265(3)	34(1)
C(29)	11187(3)	7892(3)	7718(3)	33(1)
C(31)	11914(4)	6191(3)	8492(4)	53(1)
C(32)	12950(6)	5347(5)	8383(8)	121(3)
N(5)	5925(2)	10332(2)	4089(2)	26(1)
N(8)	9251(2)	11189(2)	6407(2)	27(1)
N(12)	3667(2)	9826(2)	2583(2)	27(1)
O(13)	2830(2)	7875(2)	1600(2)	43(1)
O(22)	9931(2)	13180(2)	7400(2)	42(1)
O(23)	7916(2)	12574(2)	5727(2)	32(1)
O(29)	10091(3)	7731(3)	7790(3)	57(1)
O(30)	12187(2)	7205(2)	8039(3)	48(1)
S(1)	8264(1)	7793(1)	4874(1)	30(1)
S(3)	5269(1)	8000(1)	3096(1)	28(1)

Table 3. Bond lengths [Å] and angles [°] for ad001.

C(2)-C(6)	1.381(4)
C(2)-S(3)	1.722(3)
C(2)-S(1)	1.724(3)
C(4)-N(5)	1.301(4)
C(4)-N(12)	1.402(4)
C(4)-S(3)	1.757(3)
C(6)-N(5)	1.387(4)
C(6)-C(7)	1.435(4)
C(7)-C(11)	1.381(4)
C(7)-N(8)	1.401(4)
C(9)-C(10)	1.355(4)
C(9)-N(8)	1.398(4)
C(10)-C(11)	1.427(4)
C(10)-C(28)	1.493(4)
C(11)-S(1)	1.737(3)
C(13)-O(13)	1.222(4)

C(13)-N(12)	1.379(4)
C(13)-C(14)	1.499(4)
C(15)-N(12)	1.475(4)
C(15)-C(16)	1.510(4)
C(16)-C(21)	1.391(4)
C(16)-C(17)	1.391(4)
C(17)-C(18)	1.391(4)
C(18)-C(19)	1.382(5)
C(19)-C(20)	1.371(5)
C(20)-C(21)	1.382(5)
C(22)-O(22)	1.199(4)
C(22)-O(23)	1.324(4)
C(22)-N(8)	1.406(4)
C(24)-O(23)	1.495(3)
C(24)-C(25)	1.508(5)
C(24)-C(26)	1.519(5)
C(24)-C(27)	1.520(5)
C(28)-C(29)	1.501(4)
C(29)-O(29)	1.199(4)
C(29)-O(30)	1.323(4)
C(31)-O(30)	1.462(4)
C(31)-C(32)	1.478(6)

C(6)-C(2)-S(3)	111.8(2)
C(6)-C(2)-S(1)	115.9(2)
S(3)-C(2)-S(1)	132.23(18)
N(5)-C(4)-N(12)	122.0(3)
N(5)-C(4)-S(3)	117.2(2)
N(12)-C(4)-S(3)	120.7(2)
C(2)-C(6)-N(5)	114.7(2)
C(2)-C(6)-C(7)	109.2(2)
N(5)-C(6)-C(7)	136.1(3)
C(11)-C(7)-N(8)	105.9(2)
C(11)-C(7)-C(6)	112.7(2)
N(8)-C(7)-C(6)	141.4(3)
C(10)-C(9)-N(8)	110.6(3)
C(9)-C(10)-C(11)	105.1(3)
C(9)-C(10)-C(28)	126.7(3)
C(11)-C(10)-C(28)	128.1(3)

C(7)-C(11)-C(10)	110.4(3)
C(7)-C(11)-S(1)	113.4(2)
C(10)-C(11)-S(1)	136.2(2)
O(13)-C(13)-N(12)	120.4(3)
O(13)-C(13)-C(14)	122.0(3)
N(12)-C(13)-C(14)	117.6(3)
N(12)-C(15)-C(16)	113.3(2)
C(21)-C(16)-C(17)	118.9(3)
C(21)-C(16)-C(15)	121.5(3)
C(17)-C(16)-C(15)	119.6(3)
C(16)-C(17)-C(18)	120.2(3)
C(19)-C(18)-C(17)	120.0(3)
C(20)-C(19)-C(18)	120.1(3)
C(19)-C(20)-C(21)	120.3(3)
C(20)-C(21)-C(16)	120.5(3)
O(22)-C(22)-O(23)	128.5(3)
O(22)-C(22)-N(8)	121.2(3)
O(23)-C(22)-N(8)	110.3(2)
O(23)-C(24)-C(25)	110.7(3)
O(23)-C(24)-C(26)	108.3(3)
C(25)-C(24)-C(26)	113.2(3)
O(23)-C(24)-C(27)	102.0(2)
C(25)-C(24)-C(27)	111.0(3)
C(26)-C(24)-C(27)	111.2(3)
C(10)-C(28)-C(29)	113.8(3)
O(29)-C(29)-O(30)	123.6(3)
O(29)-C(29)-C(28)	124.8(3)
O(30)-C(29)-C(28)	111.6(3)
O(30)-C(31)-C(32)	107.3(4)
C(4)-N(5)-C(6)	109.3(2)
C(9)-N(8)-C(7)	107.9(2)
C(9)-N(8)-C(22)	119.7(2)
C(7)-N(8)-C(22)	132.3(3)
C(13)-N(12)-C(4)	120.6(2)
C(13)-N(12)-C(15)	121.8(2)
C(4)-N(12)-C(15)	117.6(2)
C(22)-O(23)-C(24)	118.5(2)
C(29)-O(30)-C(31)	117.2(3)
C(2)-S(1)-C(11)	88.81(14)

C(2)-S(3)-C(4)	86.97(14)

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($Å^2 x\ 10^3$)for ad001. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2a^{*2}U^{11} + ... + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
C(2)	28(1)	26(2)	28(2)	8(1)	10(1)	3(1)
C(4)	27(1)	29(2)	23(1)	9(1)	11(1)	3(1)
C(6)	26(1)	23(1)	23(1)	7(1)	10(1)	2(1)
C(7)	27(1)	23(1)	27(2)	6(1)	12(1)	2(1)
C(9)	26(1)	34(2)	29(2)	9(1)	7(1)	2(1)
C(10)	26(1)	33(2)	30(2)	10(1)	11(1)	5(1)
C(11)	28(1)	24(1)	29(2)	7(1)	11(1)	2(1)
C(13)	29(2)	39(2)	30(2)	10(1)	9(1)	0(1)
C(14)	29(2)	47(2)	34(2)	9(2)	5(1)	1(1)
C(15)	30(2)	32(2)	29(2)	9(1)	12(1)	8(1)
C(16)	25(1)	30(2)	26(2)	7(1)	6(1)	2(1)
C(17)	32(2)	36(2)	33(2)	12(1)	14(1)	12(1)
C(18)	49(2)	39(2)	41(2)	19(2)	15(2)	15(2)
C(19)	51(2)	41(2)	36(2)	17(2)	19(2)	6(2)
C(20)	41(2)	46(2)	43(2)	16(2)	24(2)	8(2)
C(21)	32(2)	40(2)	40(2)	16(2)	14(1)	11(1)
C(22)	33(2)	25(2)	36(2)	7(1)	14(1)	2(1)
C(24)	42(2)	20(2)	49(2)	10(1)	15(2)	5(1)
C(25)	60(2)	29(2)	65(3)	4(2)	29(2)	10(2)
C(26)	56(2)	31(2)	61(2)	18(2)	21(2)	1(2)
C(27)	48(2)	31(2)	67(3)	16(2)	10(2)	11(2)
C(28)	27(2)	37(2)	38(2)	13(1)	10(1)	7(1)
C(29)	29(2)	38(2)	28(2)	10(1)	4(1)	6(1)
C(31)	61(2)	39(2)	51(2)	21(2)	6(2)	1(2)
C(32)	89(4)	71(4)	226(8)	95(5)	48(5)	34(3)
N(5)	26(1)	28(1)	25(1)	9(1)	9(1)	3(1)
N(8)	27(1)	23(1)	29(1)	6(1)	8(1)	1(1)
N(12)	24(1)	30(1)	26(1)	9(1)	7(1)	3(1)
O(13)	38(1)	36(1)	43(1)	9(1)	2(1)	-3(1)

O(22)	40(1)	30(1)	42(1)	2(1)	4(1)	-3(1)
O(23)	33(1)	22(1)	37(1)	8(1)	8(1)	4(1)
O(29)	44(2)	77(2)	72(2)	49(2)	29(1)	17(1)
O(30)	39(1)	39(1)	66(2)	27(1)	11(1)	9(1)
S(1)	32(1)	23(1)	34(1)	8(1)	11(1)	6(1)
S(3)	30(1)	25(1)	27(1)	6(1)	8(1)	-1(1)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for ad001.

	x	y	z	U(eq)
H(9)	11247	11369	7722	37
H(14A)	669	8631	471	59
H(14B)	1504	9759	399	59
H(14C)	972	9885	1555	59
H(15A)	4194	11583	3587	36
H(15B)	2579	11246	2881	36
H(17)	1884	12576	1611	39
H(18)	1975	13382	-19	50
H(19)	3628	12860	-925	49
H(20)	5199	11559	-197	48
H(21)	5145	10774	1445	43
H(25A)	6960	13996	7249	76
H(25B)	7048	15244	6900	76
H(25C)	8416	14723	7632	76
H(26A)	9533	14468	6014	73
H(26B)	8403	15260	5351	73
H(26C)	8603	13976	4532	73
H(27A)	6226	13075	3879	76
H(27B)	5878	14381	4508	76
H(27C)	5515	13234	4917	76
H(28A)	12351	9436	7984	41
H(28B)	11965	8537	6585	41
H(31A)	10984	5784	7966	63
H(31B)	11979	6480	9392	63

H(32A)	13024	5215	7542	181
H(32B)	12676	4576	8484	181
H(32C)	13836	5690	9051	181

Table 6. Torsion angles [°] for ad001.

S(3)-C(2)-C(6)-N(5)	-1.5(3)
S(1)-C(2)-C(6)-N(5)	-178.8(2)
S(3)-C(2)-C(6)-C(7)	177.67(19)
S(1)-C(2)-C(6)-C(7)	0.4(3)
C(2)-C(6)-C(7)-C(11)	-0.5(3)
N(5)-C(6)-C(7)-C(11)	178.4(3)
C(2)-C(6)-C(7)-N(8)	-179.1(3)
N(5)-C(6)-C(7)-N(8)	-0.2(6)
N(8)-C(9)-C(10)-C(11)	0.3(3)
N(8)-C(9)-C(10)-C(28)	178.7(3)
N(8)-C(7)-C(11)-C(10)	-0.3(3)
C(6)-C(7)-C(11)-C(10)	-179.3(2)
N(8)-C(7)-C(11)-S(1)	179.51(19)
C(6)-C(7)-C(11)-S(1)	0.4(3)
C(9)-C(10)-C(11)-C(7)	0.0(3)
C(28)-C(10)-C(11)-C(7)	-178.4(3)
C(9)-C(10)-C(11)-S(1)	-179.7(3)
C(28)-C(10)-C(11)-S(1)	1.9(5)
N(12)-C(15)-C(16)-C(21)	-43.4(4)
N(12)-C(15)-C(16)-C(17)	137.2(3)
C(21)-C(16)-C(17)-C(18)	-0.9(5)
C(15)-C(16)-C(17)-C(18)	178.5(3)
C(16)-C(17)-C(18)-C(19)	0.9(5)
C(17)-C(18)-C(19)-C(20)	-0.3(5)
C(18)-C(19)-C(20)-C(21)	-0.2(5)
C(19)-C(20)-C(21)-C(16)	0.2(5)
C(17)-C(16)-C(21)-C(20)	0.4(5)
C(15)-C(16)-C(21)-C(20)	-179.0(3)
C(9)-C(10)-C(28)-C(29)	129.2(3)
C(11)-C(10)-C(28)-C(29)	-52.8(4)
C(10)-C(28)-C(29)-O(29)	-6.5(5)

C(10)-C(28)-C(29)-O(30)	173.9(3)
N(12)-C(4)-N(5)-C(6)	-179.5(2)
S(3)-C(4)-N(5)-C(6)	-0.4(3)
C(2)-C(6)-N(5)-C(4)	1.2(3)
C(7)-C(6)-N(5)-C(4)	-177.7(3)
C(10)-C(9)-N(8)-C(7)	-0.5(3)
C(10)-C(9)-N(8)-C(22)	-177.8(3)
C(11)-C(7)-N(8)-C(9)	0.4(3)
C(6)-C(7)-N(8)-C(9)	179.1(4)
C(11)-C(7)-N(8)-C(22)	177.3(3)
C(6)-C(7)-N(8)-C(22)	-4.1(6)
O(22)-C(22)-N(8)-C(9)	-7.0(4)
O(23)-C(22)-N(8)-C(9)	171.8(2)
O(22)-C(22)-N(8)-C(7)	176.5(3)
O(23)-C(22)-N(8)-C(7)	-4.7(4)
O(13)-C(13)-N(12)-C(4)	1.1(4)
C(14)-C(13)-N(12)-C(4)	-177.4(3)
O(13)-C(13)-N(12)-C(15)	-179.2(3)
C(14)-C(13)-N(12)-C(15)	2.3(4)
N(5)-C(4)-N(12)-C(13)	179.6(3)
S(3)-C(4)-N(12)-C(13)	0.5(4)
N(5)-C(4)-N(12)-C(15)	-0.1(4)
S(3)-C(4)-N(12)-C(15)	-179.2(2)
C(16)-C(15)-N(12)-C(13)	-75.2(3)
C(16)-C(15)-N(12)-C(4)	104.5(3)
O(22)-C(22)-O(23)-C(24)	1.1(5)
N(8)-C(22)-O(23)-C(24)	-177.6(2)
C(25)-C(24)-O(23)-C(22)	-59.9(4)
C(26)-C(24)-O(23)-C(22)	64.7(4)
C(27)-C(24)-O(23)-C(22)	-178.0(3)
O(29)-C(29)-O(30)-C(31)	0.6(5)
C(28)-C(29)-O(30)-C(31)	-179.7(3)
C(32)-C(31)-O(30)-C(29)	161.7(4)
C(6)-C(2)-S(1)-C(11)	-0.1(2)
S(3)-C(2)-S(1)-C(11)	-176.7(2)
C(7)-C(11)-S(1)-C(2)	-0.2(2)
C(10)-C(11)-S(1)-C(2)	179.5(3)
C(6)-C(2)-S(3)-C(4)	1.0(2)
S(1)-C(2)-S(3)-C(4)	177.7(2)

N(5)-C(4)-S(3)-C(2)	-0.4(2)
N(12)-C(4)-S(3)-C(2)	178.7(2)

Symmetry transformations used to generate equivalent atoms:

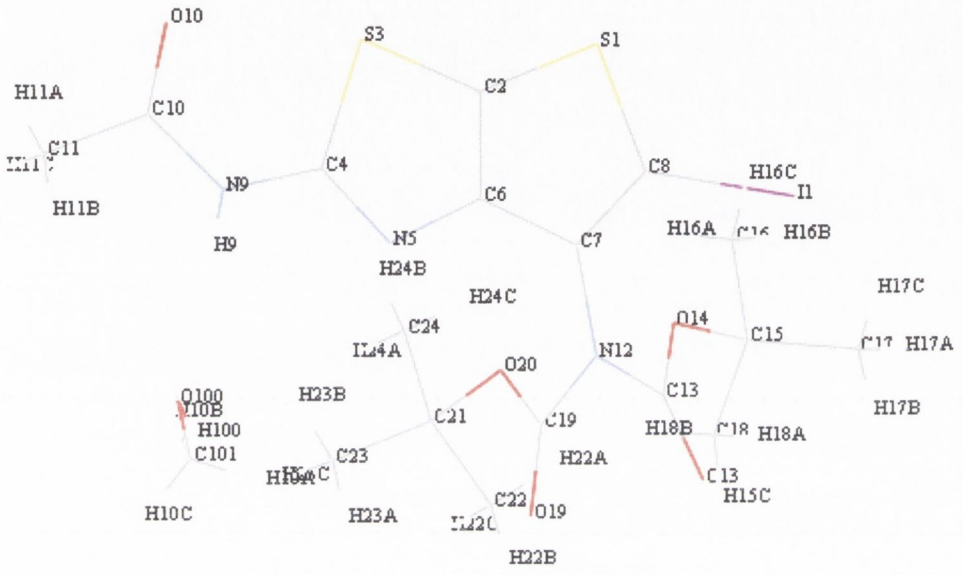


Table 1. Crystal data and structure refinement for $C_{18}H_{26}IN_3O_6S_2$ - **4.3**.

Identification code	ad002	
Empirical formula	$C_{18}H_{26}IN_3O_6S_2$	
Formula weight	571.44	
Temperature	173(2) K	
Wavelength	0.71073 Å	
Crystal system	monoclinic	
Space group	P 2_1/c No. 14	
Unit cell dimensions	a = 15.3515(3) Å	α= 90°.
	b = 8.8464(2) Å	β= 112.716(2)°.
	c = 19.3224(4) Å	γ = 90°.
Volume	2420.54(9) Å^3	
Z	4	

Density (calculated)	1.568 Mg/m^3
Absorption coefficient	1.532 mm^{-1}
F(000)	1152
Crystal size	0.30 x 0.20 x 0.20 mm^3
Theta range for data collection	3.15 to 25.02°.
Index ranges	$-18 \leq h \leq 18$, $-9 \leq k \leq 10$, $-22 \leq l \leq 22$
Reflections collected	7087
Independent reflections	4249 [R(int) = 0.0159]
Completeness to theta = 25.02°	99.3 %
Max. and min. transmission	0.7492 and 0.6564
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	4249 / 0 / 364
Goodness-of-fit on F^2	1.081
Final R indices [I>2sigma(I)]	R1 = 0.0220, wR2 = 0.0518
R indices (all data)	R1 = 0.0253, wR2 = 0.0533
Largest diff. peak and hole	0.604 and -0.498 $e.Å^{-3}$

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3) for ad002. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
C(2)	2519(1)	5766(2)	4312(1)	20(1)
C(4)	3977(2)	6682(2)	4270(1)	21(1)
C(6)	2693(2)	7279(2)	4440(1)	19(1)
C(7)	1956(2)	8041(2)	4577(1)	20(1)
C(8)	1263(2)	7073(2)	4552(1)	22(1)
C(10)	5318(2)	5732(3)	4040(1)	24(1)
C(11)	6216(2)	6185(3)	3966(2)	34(1)
C(13)	2180(2)	10110(2)	5481(1)	21(1)
C(15)	2769(2)	9118(3)	6760(1)	31(1)
C(16)	3193(3)	7578(3)	7047(2)	43(1)
C(17)	1856(2)	9368(3)	6867(2)	39(1)
C(18)	3478(2)	10366(4)	7094(2)	40(1)
C(19)	1911(2)	10690(2)	4175(1)	23(1)
C(21)	1451(2)	10815(3)	2813(1)	30(1)
C(22)	788(3)	12139(4)	2692(2)	54(1)
C(23)	2420(2)	11260(4)	2849(2)	53(1)

C(24)	1034(2)	9612(3)	2225(2)	39(1)
C(101)	4932(2)	10793(4)	3943(2)	61(1)
I(1)	20(1)	7595(1)	4686(1)	29(1)
N(5)	3528(1)	7812(2)	4419(1)	21(1)
N(9)	4830(1)	6876(2)	4208(1)	23(1)
N(12)	1980(1)	9627(2)	4736(1)	20(1)
O(10)	5011(1)	4446(2)	3946(1)	30(1)
O(13)	2006(1)	11357(2)	5637(1)	26(1)
O(14)	2561(1)	8968(2)	5943(1)	26(1)
O(19)	2168(1)	11980(2)	4294(1)	34(1)
O(20)	1546(1)	10002(2)	3512(1)	26(1)
O(100)	5587(1)	9811(2)	4456(1)	37(1)
S(1)	1476(1)	5217(1)	4376(1)	24(1)
S(3)	3427(1)	4893(1)	4144(1)	21(1)

Table 3. Bond lengths [Å] and angles [°] for ad002.

C(2)-C(6)	1.369(3)
C(2)-S(1)	1.722(2)
C(2)-S(3)	1.730(2)
C(4)-N(5)	1.307(3)
C(4)-N(9)	1.371(3)
C(4)-S(3)	1.766(2)
C(6)-N(5)	1.382(3)
C(6)-C(7)	1.425(3)
C(7)-C(8)	1.352(3)
C(7)-N(12)	1.434(3)
C(8)-S(1)	1.735(2)
C(8)-I(1)	2.074(2)
C(10)-O(10)	1.218(3)
C(10)-N(9)	1.370(3)
C(10)-C(11)	1.495(3)
C(13)-O(13)	1.200(3)
C(13)-O(14)	1.326(3)
C(13)-N(12)	1.417(3)
C(15)-O(14)	1.490(3)
C(15)-C(18)	1.509(4)

C(15)-C(17)	1.511(4)
C(15)-C(16)	1.520(4)
C(19)-O(19)	1.200(3)
C(19)-O(20)	1.331(3)
C(19)-N(12)	1.409(3)
C(21)-O(20)	1.486(3)
C(21)-C(24)	1.507(3)
C(21)-C(22)	1.509(4)
C(21)-C(23)	1.514(4)
C(101)-O(100)	1.407(4)

C(6)-C(2)-S(1)	112.60(16)
C(6)-C(2)-S(3)	110.84(16)
S(1)-C(2)-S(3)	136.55(13)
N(5)-C(4)-N(9)	121.73(19)
N(5)-C(4)-S(3)	116.77(16)
N(9)-C(4)-S(3)	121.49(16)
C(2)-C(6)-N(5)	116.48(19)
C(2)-C(6)-C(7)	112.51(19)
N(5)-C(6)-C(7)	131.01(19)
C(8)-C(7)-C(6)	111.38(19)
C(8)-C(7)-N(12)	125.51(19)
C(6)-C(7)-N(12)	123.09(19)
C(7)-C(8)-S(1)	113.57(16)
C(7)-C(8)-I(1)	127.20(16)
S(1)-C(8)-I(1)	119.22(12)
O(10)-C(10)-N(9)	120.7(2)
O(10)-C(10)-C(11)	123.6(2)
N(9)-C(10)-C(11)	115.7(2)
O(13)-C(13)-O(14)	127.9(2)
O(13)-C(13)-N(12)	123.45(19)
O(14)-C(13)-N(12)	108.64(17)
O(14)-C(15)-C(18)	110.0(2)
O(14)-C(15)-C(17)	109.2(2)
C(18)-C(15)-C(17)	112.8(2)
O(14)-C(15)-C(16)	101.17(19)
C(18)-C(15)-C(16)	111.3(2)
C(17)-C(15)-C(16)	111.7(2)
O(19)-C(19)-O(20)	127.2(2)

O(19)-C(19)-N(12)	124.5(2)
O(20)-C(19)-N(12)	108.34(17)
O(20)-C(21)-C(24)	102.03(18)
O(20)-C(21)-C(22)	109.9(2)
C(24)-C(21)-C(22)	111.5(2)
O(20)-C(21)-C(23)	109.5(2)
C(24)-C(21)-C(23)	110.0(2)
C(22)-C(21)-C(23)	113.3(3)
C(4)-N(5)-C(6)	108.67(17)
C(10)-N(9)-C(4)	123.77(19)
C(19)-N(12)-C(13)	120.20(17)
C(19)-N(12)-C(7)	119.97(17)
C(13)-N(12)-C(7)	119.44(17)
C(13)-O(14)-C(15)	120.23(17)
C(19)-O(20)-C(21)	120.71(16)
C(2)-S(1)-C(8)	89.91(10)
C(2)-S(3)-C(4)	87.23(10)

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($Å^2$x 10^3)for ad002. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U^{11} + ... + 2 h k a^* b^* U^{12}]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
C(2)	19(1)	18(1)	23(1)	-1(1)	8(1)	1(1)
C(4)	23(1)	20(1)	17(1)	-1(1)	6(1)	-2(1)
C(6)	20(1)	17(1)	18(1)	0(1)	6(1)	-1(1)
C(7)	23(1)	16(1)	18(1)	-1(1)	5(1)	2(1)
C(8)	22(1)	20(1)	24(1)	-2(1)	9(1)	2(1)
C(10)	25(1)	26(1)	19(1)	0(1)	6(1)	2(1)
C(11)	32(1)	35(2)	40(2)	-1(1)	19(1)	1(1)
C(13)	19(1)	21(1)	25(1)	-2(1)	10(1)	-1(1)
C(15)	40(1)	32(1)	20(1)	1(1)	12(1)	1(1)
C(16)	54(2)	43(2)	30(2)	10(1)	14(1)	11(2)
C(17)	48(2)	41(2)	34(2)	1(1)	23(1)	-1(1)
C(18)	45(2)	50(2)	25(2)	-7(1)	12(1)	-8(1)
C(19)	25(1)	18(1)	23(1)	1(1)	8(1)	5(1)

C(21)	42(1)	23(1)	22(1)	2(1)	11(1)	1(1)
C(22)	78(3)	41(2)	38(2)	8(2)	15(2)	24(2)
C(23)	60(2)	58(2)	49(2)	-15(2)	32(2)	-26(2)
C(24)	55(2)	35(2)	25(1)	-3(1)	12(1)	-7(1)
C(101)	73(2)	48(2)	62(2)	-2(2)	28(2)	2(2)
I(1)	24(1)	28(1)	37(1)	-6(1)	14(1)	1(1)
N(5)	23(1)	17(1)	22(1)	-1(1)	8(1)	-1(1)
N(9)	24(1)	18(1)	29(1)	-2(1)	11(1)	-3(1)
N(12)	25(1)	15(1)	20(1)	-2(1)	9(1)	1(1)
O(10)	29(1)	24(1)	37(1)	-5(1)	13(1)	1(1)
O(13)	31(1)	20(1)	30(1)	-5(1)	15(1)	2(1)
O(14)	35(1)	23(1)	18(1)	-1(1)	9(1)	4(1)
O(19)	52(1)	17(1)	31(1)	-1(1)	15(1)	-3(1)
O(20)	36(1)	19(1)	21(1)	1(1)	8(1)	-2(1)
O(100)	40(1)	28(1)	44(1)	-8(1)	15(1)	-8(1)
S(1)	21(1)	17(1)	33(1)	-4(1)	11(1)	-2(1)
S(3)	22(1)	16(1)	25(1)	-3(1)	9(1)	0(1)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for ad002.

	x	y	z	U(eq)
H(10A)	4496	11180	4162	91
H(10B)	4574	10244	3480	91
H(10C)	5267	11640	3831	91
H(9)	5069(19)	7700(30)	4316(15)	35(8)
H(11A)	6580(20)	5300(40)	4040(16)	57(9)
H(11B)	6520(20)	6990(40)	4287(17)	52(9)
H(11C)	6080(20)	6460(40)	3488(19)	56(9)
H(15C)	3230(20)	11380(40)	6912(17)	56(9)
H(16A)	3740(20)	7410(30)	6980(18)	53(10)
H(16B)	3330(20)	7470(30)	7600(20)	58(10)
H(16C)	2760(20)	6750(30)	6801(16)	46(8)
H(17A)	1990(19)	9300(30)	7370(17)	47(8)
H(17B)	1577(18)	10350(30)	6711(14)	38(7)

H(17C)	1370(20)	8620(40)	6587(16)	51(8)
H(18A)	3675(19)	10350(30)	7625(17)	46(8)
H(18B)	4010(20)	10280(30)	6983(16)	55(9)
H(22A)	190(20)	11620(30)	2722(16)	49(8)
H(22B)	1080(20)	12920(40)	3076(19)	62(10)
H(22C)	640(20)	12550(30)	2184(19)	57(9)
H(23A)	2720(20)	12070(40)	3222(19)	65(10)
H(23B)	2900(30)	10290(50)	3000(20)	98(13)
H(23C)	2360(20)	11590(40)	2340(20)	81(11)
H(24A)	1010(20)	10010(30)	1744(17)	49(8)
H(24B)	1420(20)	8710(30)	2367(15)	45(8)
H(24C)	400(20)	9280(40)	2198(16)	59(9)
H(100)	5870(20)	10230(40)	4847(18)	63(11)

Table 6. Torsion angles [°] for ad002.

S(1)-C(2)-C(6)-N(5)	178.44(15)
S(3)-C(2)-C(6)-N(5)	-0.3(2)
S(1)-C(2)-C(6)-C(7)	-1.5(2)
S(3)-C(2)-C(6)-C(7)	179.67(14)
C(2)-C(6)-C(7)-C(8)	0.6(3)
N(5)-C(6)-C(7)-C(8)	-179.4(2)
C(2)-C(6)-C(7)-N(12)	179.12(19)
N(5)-C(6)-C(7)-N(12)	-0.9(3)
C(6)-C(7)-C(8)-S(1)	0.7(2)
N(12)-C(7)-C(8)-S(1)	-177.85(16)
C(6)-C(7)-C(8)-I(1)	-178.44(15)
N(12)-C(7)-C(8)-I(1)	3.0(3)
N(9)-C(4)-N(5)-C(6)	179.50(19)
S(3)-C(4)-N(5)-C(6)	0.0(2)
C(2)-C(6)-N(5)-C(4)	0.2(3)
C(7)-C(6)-N(5)-C(4)	-179.8(2)
O(10)-C(10)-N(9)-C(4)	-1.0(3)
C(11)-C(10)-N(9)-C(4)	177.5(2)
N(5)-C(4)-N(9)-C(10)	-178.9(2)
S(3)-C(4)-N(9)-C(10)	0.6(3)
O(19)-C(19)-N(12)-C(13)	13.6(3)

O(20)-C(19)-N(12)-C(13)	-168.19(17)
O(19)-C(19)-N(12)-C(7)	-159.3(2)
O(20)-C(19)-N(12)-C(7)	19.0(3)
O(13)-C(13)-N(12)-C(19)	24.7(3)
O(14)-C(13)-N(12)-C(19)	-156.12(18)
O(13)-C(13)-N(12)-C(7)	-162.4(2)
O(14)-C(13)-N(12)-C(7)	16.7(3)
C(8)-C(7)-N(12)-C(19)	-113.6(2)
C(6)-C(7)-N(12)-C(19)	68.1(3)
C(8)-C(7)-N(12)-C(13)	73.6(3)
C(6)-C(7)-N(12)-C(13)	-104.8(2)
O(13)-C(13)-O(14)-C(15)	4.2(3)
N(12)-C(13)-O(14)-C(15)	-174.85(17)
C(18)-C(15)-O(14)-C(13)	-61.8(3)
C(17)-C(15)-O(14)-C(13)	62.5(3)
C(16)-C(15)-O(14)-C(13)	-179.6(2)
O(19)-C(19)-O(20)-C(21)	2.5(3)
N(12)-C(19)-O(20)-C(21)	-175.66(17)
C(24)-C(21)-O(20)-C(19)	178.9(2)
C(22)-C(21)-O(20)-C(19)	-62.7(3)
C(23)-C(21)-O(20)-C(19)	62.4(3)
C(6)-C(2)-S(1)-C(8)	1.60(17)
S(3)-C(2)-S(1)-C(8)	179.94(19)
C(7)-C(8)-S(1)-C(2)	-1.30(18)
I(1)-C(8)-S(1)-C(2)	177.89(13)
C(6)-C(2)-S(3)-C(4)	0.28(17)
S(1)-C(2)-S(3)-C(4)	-178.08(19)
N(5)-C(4)-S(3)-C(2)	-0.19(17)
N(9)-C(4)-S(3)-C(2)	-179.66(18)

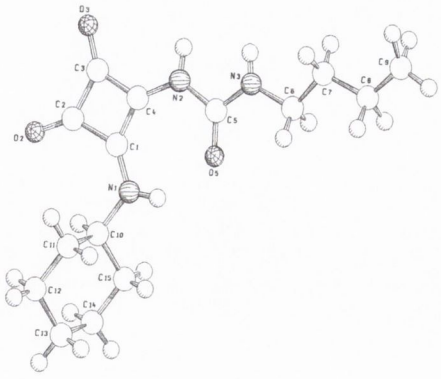
Symmetry transformations used to generate equivalent atoms:

Table 7. Hydrogen bonds for ad002 [Å and °].

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
N(9)-H(9)...O(100)	0.81(3)	2.01(3)	2.810(3)	173(3)
O(100)-H(100)...N(5)#1	0.80(3)	2.20(3)	2.952(3)	155(3)

Symmetry transformations used to generate equivalent atoms:

#1 -x+1,-y+2,-z+1



SCHAKAL

Table 1. Crystal data and structure refinement for $C_{15}H_{23}N_3O_3$ - **5.9**

Identification code	gd792s
Empirical formula	$C_{15}H_{23}N_3O_3$
Formula weight	293.36
Temperature	293(2) K
Wavelength	0.71073 Å
Space group	P1
Unit cell dimensions	a = 5.9800(4) Å α = 89.694(6) °
	b = 9.5090(5) Å β = 81.471(5) °
	c = 14.1543(8) Å γ = 87.322(5) °
Volume	795.10(8) $Å^3$
Z	2
Density (calculated)	1.225
Absorption coefficient	0.086 mm^{-1}
F(000)	316
Theta range for data collection	1.45 to 24.97 °
Index range	$0 \leq h \leq 6, -10 \leq k \leq 10, -15 \leq l \leq 15$
Reflections collected	3009
Independent reflections	2726
Refinement method	Full-matrix least-squares on F^2
Data /restraints / parameters	2723 / 0 / 282
Goodness-of-fit on F^2	1.070
Final R indices [$I>2\sigma(I)$]	R1 = 0.0510, wR2 = 0.1250
R indices (all data)	R1 = 0.1081, wR2 = 0.1575
Largest diff. peak and hole	0.315 and -0.179 $e.Å^{-3}$

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$ x 10^3) for $C_{15}H_{23}N_3O_3$. U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

Atom label	x	y	z	U(eq)
O(2)	1.1109(5)	0.9916(3)	0.8368(2)	0.0657(8)
O(3)	1.1628(4)	1.0531(3)	0.6069(2)	0.0642(8)
O(5)	0.4513(4)	0.7506(3)	0.6646(2)	0.0639(8)
N(1)	0.6539(5)	0.8246(3)	0.8235(2)	0.0478(8)
N(2)	0.7290(5)	0.8858(3)	0.5880(2)	0.0514(8)
N(3)	0.4912(5)	0.7765(4)	0.5039(2)	0.0573(9)
C(1)	0.8014(5)	0.8885(4)	0.7628(2)	0.0421(8)
C(2)	1.0097(6)	0.9649(4)	0.7719(2)	0.0470(9)
C(3)	1.0317(6)	0.9921(4)	0.6668(2)	0.0489(9)
C(4)	0.8330(5)	0.9154(4)	0.6640(2)	0.0436(8)
C(5)	0.5456(6)	0.7997(4)	0.5904(2)	0.0479(9)
C(6)	0.3021(8)	0.6907(7)	0.4916(3)	0.0703(14)
C(7)	0.2970(9)	0.6561(6)	0.3909(3)	0.0692(13)
C(8)	0.0974(11)	0.5711(7)	0.3764(4)	0.089(2)
C(9)	0.0685(15)	0.5364(10)	0.2822(5)	0.136(3)
C(10)	0.6721(6)	0.8110(4)	0.9256(2)	0.0421(8)
C(11)	0.8314(7)	0.6898(4)	0.9440(3)	0.0507(9)
C(12)	0.8493(7)	0.6801(5)	1.0500(3)	0.0576(10)
C(13)	0.6187(7)	0.6661(5)	1.1093(3)	0.0617(11)
C(14)	0.4567(7)	0.7856(5)	1.0889(3)	0.0564(10)
C(15)	0.4406(6)	0.7966(5)	0.9828(3)	0.0493(9)
H(1)	0.5412(77)	0.7890(45)	0.8012(31)	0.078(13)
H(2)	0.7908(63)	0.9146(38)	0.5297(29)	0.053(11)
H(3)	0.5641(73)	0.8118(45)	0.4557(32)	0.060(13)
H(61)	0.3056(59)	0.6054(39)	0.5335(27)	0.102(10)
H(62)	0.1605(63)	0.7499(36)	0.5188(25)	0.091(10)
H(71)	0.4436(66)	0.6144(37)	0.3619(26)	0.069(10)
H(72)	0.3002(58)	0.7415(38)	0.3516(25)	0.074(10)
H(81)	0.1083(57)	0.4777(37)	0.4195(25)	0.098(10)
H(82)	-0.0579(63)	0.6007(35)	0.4127(25)	0.103(10)
H(91)	-0.0667(60)	0.4735(36)	0.2793(24)	0.103(9)
H(92)	0.0598(65)	0.6262(39)	0.2464(29)	0.107(11)
H(93)	0.1993(61)	0.5381(38)	0.2271(27)	0.125(10)
H(10)	0.7293(53)	0.8943(35)	0.9442(23)	0.046(9)
H(112)	0.9745(67)	0.7005(37)	0.9082(27)	0.057(11)

H(113)	0.7715(58)	0.5977(38)	0.9206(25)	0.049(10)
H(121)	0.9157(64)	0.7617(44)	1.0685(28)	0.066(12)
H(122)	0.9586(85)	0.6018(53)	1.0642(33)	0.087(16)
H(131)	0.5494(83)	0.5827(53)	1.0971(34)	0.091(15)
H(132)	0.6307(63)	0.6640(39)	1.1796(30)	0.076(11)
H(141)	0.5199(68)	0.8744(44)	1.1122(29)	0.077(12)
H(142)	0.3051(75)	0.7752(43)	1.1210(30)	0.070(13)
H(151)	0.3774(62)	0.7135(41)	0.9622(26)	0.045(11)
H(152)	0.3365(65)	0.8708(41)	0.9696(26)	0.056(11)

Table 3. Bond lengths [Å] and angles [°] for $C_{15}H_{23}N_3O_3$

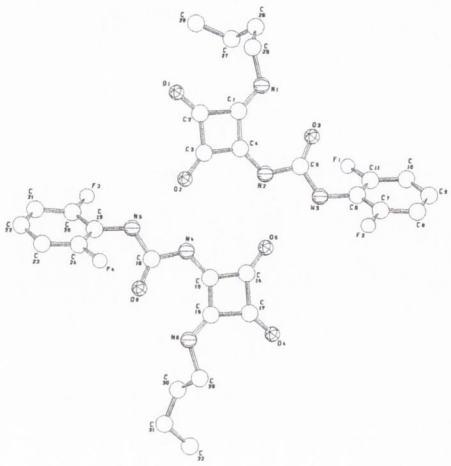
Atom labels	Bond length
O(2)-C(2)	1.207(4)
O(3)-C(3)	1.229(4)
O(5)-C(5)	1.218(4)
N(1)-C(1)	1.306(4)
N(1)-C(10)	1.469(4)
N(2-C(4)	1.356(4)
N(2)-C(5)	1.395(4)
N(3)-C(5)	1.335(4)
N(3)-C(6)	1.456(5)
C(1)-C(4)	1.407(4)
C(1)-C(2)	1.493(5)
C(2)-C(3)	1.496(5)
C(3)-C(4)	1.428(5)
C(6)-C(7)	1.470(6)
C(7)-C(8)	1.511(6)
C(8)-C(9)	1.412(7)
C(10)-C(11)	1.508(5)
C(10)-C(15)	1.508(5)
C(11)-C(12)	1.522(5)
C(12)-C(13)	1.515(6)
C(13)-C(14)	1.513(6)
C(14)-C(15)	1.523(5)
C(1)-N(1)-C(10)	123.0(3)
C(4)-N(2)-C(5)	125.2(3)

C(5)-N(3)-C(6)	121.1(3)
N(1)-C(1)-C(4)	136.2(3)
N(1)-C(1)-C(2)	133.8(3)
C(4)-C(1)-C(2)	89.9(3)
O(2)-C(2)-C(1)	135.1(3)
O(2)-C(2)-C(3)	137.4(3)
C(1)-C(2)-C(3)	87.5(3)
O(3)-C(3)-C(4)	134.6(3)
O(3)-C(3)-C(2)	136.4(3)
C(4)-C(3)-C(2)	89.0(3)
N(2)-C(4)-C(1)	137.4(3)
N(2)-C(4)-C(3)	129.0(3)
C(1)-C(4)-C(3)	93.6(3)
O(5)-C(5)-N(3)	125.0(3)
O(5)-C(5)-N(2	122.1(3)
N(3)-C(5)-N(2)	112.8(3)
N(3)-C(6)-C(7)	112.4(4)
C(6)-C(7)-C(8)	113.2(4)
C(9)-C(8)-C(7)	118.5(5)
N(1)-C(10)-C(1)1	111.7(3)
N(1)-C(10)-C(15)	110.0(3)
C(11)-C(10)-C(15)	111.3(3)
C(10)-C(11)-C(12)	110.3(3)
C(13)-C(12)-C(11)	111.2(3)
C(12)-C(13)-C(14)	110.9(3)
C(15)-C(14)-C(13)	111.5(4)
C(10)-C(15)-C(14)	110.5(3)

Table 4. Anisotropic displacement parameters ($Å^2$ x 10^3) for $C_{15}H_{23}N_3O_3$. The anisotropic displacement factor exponent takes the form $-2\pi^2[h^2a^{*2}U^{11} + ... + 2 h k a^* b^* U^{12}]$

Atom label	U11	U22	U33	U23	U13	U12
O(2)	0.067(2)	0.089(2)	0.047(2)	0.0022(13)	-0.0205(13)	-0.0297(15)
O(3)	0.058(2)	0.093(2)	0.0440(15)	0.0087(13)	-0.0073(12)	-0.0360(15)
O(5)	0.062(2)	0.090(2)	0.0420(15)	0.0082(13)	-0.0074(12)	-0.0370(15)
N(1)	0.045(2)	0.063(2)	0.038(2)	0.0060(13)	-0.0114(13)	-0.0145(15)
N(2)	0.048(2)	0.074(2)	0.036(2)	0.0066(14)	-0.0100(13)	-0.026(2)

N(3)	0.054(2)	0.085(2)	0.036(2)	0.001(2)	-0.0073(14)	-0.033(2)
C(1)	0.040(2)	0.049(2)	0.038(2)	0.0003(14)	-0.0068(14)	-0.005(2)
C(2)	0.046(2)	0.055(2)	0.042(2)	0.000(2)	-0.011(2)	-0.009(2)
C(3)	0.043(2)	0.065(2)	0.041(2)	0.002(2)	-0.008(2)	-0.015(2)
C(4	0.042(2)	0.053(2)	0.037(2)	0.0012(14)	-0.0076(14)	-0.007(2)
C(5)	0.043(2)	0.064(2)	0.039(2)	0.001(2)	-0.0068(15)	-0.015(2)
C(6)	0.060(3)	0.101(4)	0.054(3)	-0.010(3)	-0.009(2)	-0.039(3)
C(7)	0.072(3)	0.088(4)	0.052(2)	-0.002(2)	-0.016(2)	-0.029(3)
C(8)	0.096(4)	0.115(4)	0.063(3)	-0.003(3)	-0.022(3)	-0.052(4)
C(9)	0.151(6)	0.193(8)	0.076(4)	0.004(5)	-0.035(4)	-0.098(6)
C(10)	0.045(2)	0.048(2)	0.035(2)	0.0013(14)	-0.0082(14)	-0.010(2)
C(11)	0.048(2)	0.061(3)	0.044(2)	-0.005(2)	-0.008(2)	0.003(2)
C(12)	0.059(2)	0.066(3)	0.050(2)	0.001(2)	-0.017(2)	0.008(2)
C(13)	0.074(3)	0.068(3)	0.045(2)	0.014(2)	-0.015(2)	-0.012(2)
C(14)	0.044(2)	0.081(3)	0.042(2)	0.004(2)	-0.001(2)	-0.008(2)
C(15)	0.040(2)	0.064(3)	0.045(2)	0.005(2)	-0.009(2)	-0.003(2)



SCHAKAL

Table 1. Crystal data and structure refinement for $C_{60}F_8N_{12}O_{12}$ - **5.12**

Identification code	gavins	
Empirical formula	$C_{60}F_8N_{12}O_{12}$	
Formula weight	1232.72	
Temperature	293(2) K	
Wavelength	0.71073 Å	
Space group	P1	
Unit cell dimensions	a = 12.1339(19) Å	α = 90.00°
	b = 11.7176(13) Å	β = 104.09(3) °
	c = 23.246(7) Å	γ = 90.00 °
Volume	3205.6(11) $Å^3$	
Z	2	
Density (calculated)	1.277	
Absorption coefficient	0.108 mm^{-1}	
F(000)	1224	
Theta range for data collection	1.75 to 22.05 °	
Index range	$0 \le h \le 12$, $0 \le k \le 12$, $-24 \le l \le 23$	
Reflections collected	4168	
Independent reflections	3944 [R(int) = 0.1284]	
Completeness to theta = 22.05 °	99.5 %	
Refinement method	Full-matrix least-squares on F^2	
Data /restraints / parameters	3944 / 13 / 417	
Goodness-of-fit on F^2	1.130	
Final R indices [I>2σ(I)]	R1 = 0.0949, wR2 = 0.2594	

R indices (all data)	R1 = 0.1981, wR2 = 0.3504
Largest diff. peak and hole	0.466 and -0.304 e.Å^{-3}

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^2 x 10^3) for $C_{60}F_8N_{12}O_{12}$. U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

Atom label	x	y	z	U(eq)
O(2)	0.8688(5)	0.5841(7)	0.2050(3)	0.077(2)
C(4)	1.0063(8)	0.6589(7)	0.1534(4)	0.053(2)
O(1)	0.7457(5)	0.6969(7)	0.0761(3)	0.082(2)
N(2)	1.1131(5)	0.6375(6)	0.1887(3)	0.057(2)
O(3)	1.2145(5)	0.7215(6)	0.1290(3)	0.074(2)
F(1)	1.4028(5)	0.5723(6)	0.1334(3)	0.0933(19)
N(3)	1.3033(6)	0.6532(7)	0.2203(3)	0.063(2)
N(1)	0.9985(6)	0.7686(7)	0.0601(3)	0.070(2)
C(1)	0.9583(7)	0.7114(8)	0.0983(4)	0.056(2)
C(3)	0.8992(7)	0.6313(8)	0.1639(4)	0.060(2)
F(2)	1.4361(6)	0.7944(7)	0.3004(3)	0.115(2)
C(6)	1.4161(7)	0.6851(9)	0.2163(4)	0.061(3)
C(5)	1.2112(7)	0.6755(8)	0.1752(5)	0.059(2)
C(7)	1.4801(8)	0.7530(10)	0.2574(5)	0.073(3)
C(11)	1.4634(8)	0.6432(8)	0.1733(4)	0.060(2)
C(2)	0.8422(8)	0.6843(8)	0.1056(4)	0.062(3)
C(10)	1.5718(9)	0.6708(10)	0.1692(5)	0.077(3)
C(8)	1.5920(9)	0.7825(9)	0.2580(5)	0.085(3)
C(9)	1.6350(8)	0.7392(9)	0.2124(6)	0.080(3)
C(25)	0.9268(8)	0.8199(10)	0.0052(4)	0.085(3)
C(26)	0.9425(15)	0.7630(14)	-0.0495(5)	0.134(6)
C(27)	0.910(3)	0.646(2)	-0.0528(9)	0.277(17)
C(28)	0.788(4)	0.628(4)	-0.0775(16)	0.75(9)
F(3)	0.6251(5)	0.3773(6)	0.2057(3)	0.105(2)
F(4)	0.6417(5)	0.5813(6)	0.3800(3)	0.102(2)
O(4)	1.3052(5)	0.4502(6)	0.4347(3)	0.0710(19)
O(5)	1.1864(5)	0.5686(7)	0.3072(3)	0.087(2)
O(6)	0.8368(5)	0.4439(6)	0.3834(3)	0.073(2)
N(4)	0.9414(6)	0.5247(7)	0.3233(3)	0.061(2)
N(5)	0.7502(5)	0.5113(7)	0.2911(3)	0.061(2)
N(6)	1.0524(7)	0.3921(8)	0.4524(4)	0.086(3)

C(14)	1.1554(7)	0.5231(9)	0.3478(4)	0.065(3)
C(15)	1.0463(7)	0.4970(8)	0.3591(4)	0.056(2)
C(16)	1.0922(8)	0.4426(8)	0.4129(4)	0.058(2)
C(17)	1.2102(7)	0.4667(8)	0.4061(4)	0.059(2)
C(18)	0.8421(7)	0.4895(8)	0.3376(4)	0.057(2)
C(19)	0.6384(7)	0.4767(9)	0.2931(4)	0.064(3)
C(20)	0.5767(8)	0.4113(10)	0.2492(5)	0.076(3)
C(21)	0.4642(9)	0.3778(10)	0.2447(6)	0.095(4)
C(22)	0.4154(10)	0.4165(11)	0.2887(7)	0.091(4)
C(23)	0.4750(9)	0.4807(11)	0.3358(6)	0.087(4)
C(24)	0.5851(8)	0.5123(10)	0.3361(5)	0.071(3)
C(29)	1.1286(12)	0.3467(16)	0.5082(6)	0.128(6)
C(30)	1.104(3)	0.225(2)	0.5018(9)	0.285(19)
C(31)	1.104(3)	0.158(3)	0.5511(18)	0.49(4)
C(32)	1.219(4)	0.115(4)	0.5742(17)	0.37(2)

Table 3. Bond lengths [Å] and angles [°] for $C_{60}F_8N_{12}O_{12}$

Atom labels	Bond length
O(2)-C(3)	1.236(10)
C(4)-N(2)	1.377(11)
C(4)-C(1)	1.413(12)
C(4)-C(3)	1.417(12)
O(1)-C(2)	1.214(10)
N(2)-C(5)	1.377(11)
O(3)-C(5)	1.212(10)
F(1)-C(11)	1.327(10)
N(3)-C(5)	1.358(11)
N(3)-C(6)	1.444(11)
N(1)-C(1)	1.300(10)
N(1)-C(25)	1.484(12)
C(1)-C(2)	1.494(12)
C(3)-C(2)	1.499(13)
F(2)-C(7)	1.334(11)
C(6)-C(7)	1.337(13)
C(6)-C(11)	1.358(12)
C(7)-C(8)	1.399(14)

C(11)-C(10)	1.380(13)
C(10)-C(9)	1.365(15)
C(8)-C(9)	1.385(15)
C(25)-C(26)	1.490(15)
C(26)-C(27)	1.43(3)
C(27)-C(28)	1.46(2)
F(3)-C(20)	1.347(12)
F(4)-C(24)	1.352(12)
O(4)-C(17)	1.198(9)
O(5)-C(14)	1.221(10)
O(6)-C(18)	1.206(10)
N(4)-C(15)	1.379(10)
N(4)-C(18)	1.389(10)
N(5)-C(18)	1.376(11)
N(5)-C(19)	1.427(11)
N(6)-C(16)	1.283(11)
N(6)-C(29)	1.494(14)
C(14)-C(15)	1.443(12)
C(14)-C(17)	1.508(13)
C(15)-C(16)	1.393(12)
C(16)-C(17)	1.506(13)
C(19)-C(20)	1.349(14)
C(19)-C(24)	1.380(13)
C(20)-C(21)	1.400(15)
C(21)-C(22)	1.377(17)
C(22)-C(23)	1.380(17)
C(23)-C(24)	1.385(14)
C(29)-C(30)	1.45(2)
C(30)-C(31)	1.39(3)
C(31)-C(32)	1.46(2)

N(2)-C(4)-C(1)	137.7(7)
N(2)-C(4)-C(3)	128.7(8)
C(1)-C(4)-C(3)	93.6(8)
C(4)-N(2)-C(5)	123.0(7)
C(5)-N(3)-C(6)	121.0(7)
C(1)-N(1)-C(25)	123.9(8)
N(1)-C(1)-C(4)	134.8(8)
N(1)-C(1)-C(2)	135.1(8)
C(4)-C(1)-C(2)	89.8(7)

O(2)-C(3)-C(4)	134.0(9)
O(2)-C(3)-C(2)	136.5(8)
C(4)-C(3)-C(2)	89.5(7)
C(7)-C(6)-C(11)	117.7(9)
C(7)-C(6)-N(3)	120.1(9)
C(11)-C(6)-N(3)	122.1(9)
O(3)-C(5)-N(3)	124.9(8)
O(3)-C(5)-N(2)	124.3(9)
N(3)-C(5)-N(2)	110.8(8)
C(6)-C(7)-F(2)	118.9(9)
C(6)-C(7)-C(8)	123.2(10)
F(2)-C(7)-C(8)	117.9(10)
F(1)-C(11)-C(6)	118.6(9)
F(1)-C(11)-C(10)	118.2(9)
C(6)-C(11)-C(10)	123.1(10)
O(1)-C(2)-C(1)	135.9(9)
O(1)-C(2)-C(3)	137.0(8)
C(1)-C(2)-C(3)	87.1(7)
C(9)-C(10)-C(11)	117.4(10)
C(9)-C(8)-C(7)	116.5(11)
C(10)-C(9)-C(8)	122.0(9)
N(1)-C(25)-C(26)	112.4(10)
C(27)-C(26)-C(25)	112.7(13)
C(26)-C(27)-C(28)	114(2)
C(15)-N(4)-C(18)	120.9(8)
C(18)-N(5)-C(19)	121.1(8)
C(16)-N(6)-C(29)	121.6(10)
O(5)-C(14)-C(15)	134.5(9)
O(5)-C(14)-C(17)	137.3(8)
C(15)-C(14)-C(17)	88.2(8)
N(4)-C(15)-C(16)	139.2(8)
N(4)-C(15)-C(14)	126.4(9)
C(16)-C(15)-C(14)	94.4(7)
N(6)-C(16)-C(15)	135.8(9)
N(6)-C(16)-C(17)	134.1(9)
C(15)-C(16)-C(17)	90.1(7)
O(4)-C(17)-C(16)	136.3(9)
O(4)-C(17)-C(14)	136.3(9)
C(16)-C(17)-C(14)	87.3(7)
O(6)-C(18)-N(5)	124.7(8)

O(6)-C(18)-N(4)	125.4(8)
N(5)-C(18)-N(4)	109.9(8)
C(20)-C(19)-C(24)	116.6(9)
C(20)-C(19)-N(5)	119.6(9)
C(24)-C(19)-N(5)	123.7(10)
F(3)-C(20)-C(19)	118.3(9)
F(3)-C(20)-C(21)	117.2(11)
C(19)-C(20)-C(21)	124.5(11)
C(22)-C(21)-C(20)	116.1(12)
C(21)-C(22)-C(23)	122.2(10)
C(22)-C(23)-C(24)	117.8(11)
F(4)-C(24)-C(19)	119.1(9)
F(4)-C(24)-C(23)	118.3(10)
C(19)-C(24)-C(23)	122.6(12)
C(30)-C(29)-N(6)	101.1(14)
C(31)-C(30)-C(29)	121(2)
C(30)-C(31)-C(32)	108(2)

Table 4. Anisotropic displacement parameters ($Å^2$ x 10^3) for $C_{60}F_8N_{12}O_{12}$. The anisotropic displacement factor exponent takes the form $-2\pi^2[h^2a^{*2}U^{11} + ... + 2 h k a^* b^* U^{12}]$

Atom label	U11	U22	U33	U23	U13	U12
O(2)	0.046(3)	0.124(6)	0.065(4)	0.028(4)	0.020(3)	-0.006(4)
C(4)	0.056(6)	0.053(6)	0.058(6)	-0.002(5)	0.027(5)	-0.013(5)
O(1)	0.046(4)	0.132(7)	0.063(4)	0.014(4)	0.004(3)	0.002(4)
N(2)	0.024(4)	0.074(5)	0.074(5)	0.014(4)	0.012(3)	0.007(3)
O(3)	0.049(4)	0.096(5)	0.080(4)	0.032(4)	0.021(3)	0.004(3)
F(1)	0.067(4)	0.118(5)	0.098(4)	-0.015(4)	0.026(3)	-0.006(4)
N(3)	0.038(4)	0.087(6)	0.066(5)	0.011(4)	0.016(4)	-0.001(4)
N(1)	0.063(5)	0.089(6)	0.055(5)	0.019(5)	0.011(4)	-0.005(4)
C(1)	0.047(5)	0.072(6)	0.048(5)	0.008(5)	0.010(4)	-0.010(4)
C(3)	0.041(5)	0.082(7)	0.059(6)	0.010(5)	0.018(5)	-0.004(5)
F(2)	0.104(5)	0.148(7)	0.099(5)	-0.037(5)	0.039(4)	-0.011(5)
C(6	0.043(5)	0.068(7)	0.076(6)	0.011(6)	0.024(5)	0.000(5)
C(5)	0.042(6)	0.066(6)	0.075(7)	0.015(5)	0.026(5)	0.016(5)
C(7)	0.057(6)	0.085(8)	0.079(7)	-0.002(6)	0.022(6)	-0.009(6)
C(11)	0.051(6)	0.066(7)	0.065(6)	0.003(5)	0.014(5)	0.001(5)

C(2)	0.055(7)	0.067(7)	0.066(6)	-0.001(5)	0.017(5)	-0.005(5)
C(10)	0.053(7)	0.088(9)	0.089(8)	0.015(7)	0.017(6)	0.002(6)
C(8)	0.063(7)	0.076(8)	0.115(9)	0.014(7)	0.019(6)	-0.005(6)
C(9)	0.051(6)	0.077(7)	0.116(9)	0.025(7)	0.029(7)	0.004(5)
C(25)	0.073(7)	0.110(9)	0.063(6)	0.022(6)	0.000(5)	-0.007(6)
C(26)	0.221(18)	0.096(11)	0.078(8)	0.004(8)	0.021(9)	-0.035(11)
C(27)	0.50(5)	0.18(2)	0.091(15)	0.005(15)	-0.04(2)	0.03(3)
C(28)	1.11(16)	0.65(11)	0.26(4)	0.19(5)	-0.28(6)	-0.64(12)
F(3)	0.072(4)	0.148(6)	0.093(4)	-0.028(4)	0.018(3)	0.013(4)
F(4)	0.080(4)	0.131(6)	0.105(5)	-0.015(4)	0.042(4)	0.010(4)
O(4)	0.051(4)	0.097(5)	0.057(4)	-0.009(4)	-0.003(3)	0.009(3)
O(5)	0.044(4)	0.143(7)	0.077(5)	0.024(5)	0.020(3)	-0.005(4)
O(6)	0.055(4)	0.097(5)	0.074(5)	0.030(4)	0.025(3)	0.007(3)
N(4	0.036(4)	0.085(6)	0.062(5)	0.007(4)	0.012(4)	0.002(4)
N(5)	0.034(4)	0.082(6)	0.064(5)	0.014(4)	0.008(4)	0.002(4)
N(6)	0.070(6)	0.108(7)	0.080(6)	0.037(5)	0.018(5)	0.011(5)
C(14)	0.044(5)	0.087(8)	0.067(7)	0.011(6)	0.021(5)	0.005(5)
C(15)	0.042(5)	0.064(6)	0.064(6)	-0.007(5)	0.015(5)	0.004(4)
C(16)	0.060(6)	0.064(7)	0.051(6)	0.014(5)	0.014(5)	0.004(5)
C(17)	0.043(6)	0.076(7)	0.054(6)	-0.011(5)	0.007(4)	-0.001(5)
C(18)	0.044(5)	0.065(7)	0.067(6)	0.009(5)	0.026(5)	0.002(5)
C(19)	0.040(5)	0.080(7)	0.074(7)	0.013(6)	0.020(5)	0.002(5)
C(20)	0.048(6)	0.084(8)	0.096(8)	0.003(6)	0.014(6)	0.010(6)
C(21)	0.050(7)	0.089(9)	0.140(11)	0.012(8)	0.009(7)	0.010(6)
C(22)	0.058(7)	0.085(9)	0.132(11)	0.029(8)	0.027(8)	0.004(7)
C(23)	0.059(7)	0.092(9)	0.121(10)	0.038(8)	0.045(7)	0.016(7)
C(24)	0.056(6)	0.082(9)	0.080(7)	0.019(7)	0.026(6)	0.010(6)
C(29)	0.125(11)	0.157(15)	0.095(10)	0.060(11)	0.015(8)	0.045(11)
C(30)	0.46(5)	0.24(3)	0.108(14)	0.048(17)	-0.02(2)	0.15(3)
C(31)	0.54(7)	0.67(12)	0.31(5)	0.00(6)	0.22(5)	-0.16(8)
C(32)	0.37(4)	0.46(6)	0.24(4)	-0.05(4)	-0.03(3)	0.02(4)