

**Some pages of this thesis may have been removed for copyright restrictions.**

If you have discovered material in AURA which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our [Takedown Policy](#) and [contact the service](#) immediately

**Base Modified Peptidic Nucleic Acids: Synthesis  
and Crystallographic Analysis of Intermediates**

**GEETA SOOD**

**DOCTOR OF PHILOSOPHY**

**Aston University**

**September 1998**

This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with the author, and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.

**Base Modified Peptidic Nucleic Acids:  
Synthesis and Crystallographic Analysis of Intermediates**

**A Thesis submitted by Geeta Sood B.Sc. Hons. for the degree of  
Doctor of Philosophy**

**1998**

**Abstract:** Peptidic Nucleic Acids (PNAs) are achiral, uncharged nucleic acid mimetics, with a novel backbone composed of *N*-(2-aminoethyl)glycine units attached to the DNA bases through carboxymethylene linkers. With the aim of extending and improving upon the molecular recognition properties of PNAs, the aim of this work was to synthesise PNA building block intermediates containing a series of substituted purine bases for subsequent use in automated PNA synthesis.

Four purine bases: 2,6-diaminopurine (D), isoGuanine (isoG), xanthine (X) and hypoxanthine (H) were identified for incorporation into PNAs targeted to DNA, with the promise of increased hybrid stability over extended pH ranges together with improvements over the use of adenine (A) in duplex formation, and cytosine (C) in triplex formation.

A reliable, high-yielding synthesis of the PNA backbone component *N*-(2-butyloxycarbonyl-aminoethyl)glycinate ethyl ester was established. The precursor *N*-(2-butyloxycarbonyl)amino acetonitrile was crystallised and analysed by X-ray crystallography for the first time. An excellent refinement ( $R = 0.0276$ ) was attained for this structure, allowing comparisons with known analogues. Although chemical synthesis of pure, fully-characterised PNA monomers was not achieved, chemical synthesis of PNA building blocks composed of diaminopurine, xanthine and hypoxanthine was completely successful.

In parallel, a second objective of this work was to characterise and evaluate novel crystalline intermediates, which formed a new series of substituted purine bases, generated by attaching alkyl substituents at the N9 or N7 sites of purine bases. Crystallographic analysis was undertaken to probe the regiochemistry of isomers, and to reveal interesting structural features of the new series of similarly-substituted purine bases.

The attainment of the versatile synthetic intermediate 2,6-dichloro-9-(carboxymethyl)purine ethyl ester, and its homologous regioisomers 6-chloro-9-(carboxymethyl)purine ethyl ester and 6-chloro-7-(carboxymethyl)purine ethyl ester, necessitated the use of X-ray crystallographic analysis for unambiguous structural assignment. Successful refinement of the disordered 2,6-diamino-9-(carboxymethyl)purine ethyl ester allowed comparison with the reported structure of the adenine analogue, ethyl adenin-9-yl acetate. Replacement of the chloro moieties with amino, azido and methoxy groups expanded the internal angles at their point of attachment to the purine ring. Crystallographic analysis played a pivotal role towards confirming the identity of the peralkylated hypoxanthine derivative diethyl 6-oxo-6,7-dihydro-3*H*-purine-3,7-diacetate, where two ethyl side chains were found to attach at N3 and N7.

**Keywords:** Peptide Nucleic Acids (PNAs), Antisense and Antigene Theory, Triplex-Forming Oligonucleotides (TFOs), X-Ray Crystallographic Analysis, 2,6-Diaminopurine (D), isoGuanine (isoG), Xanthine (X), Hypoxanthine (H).

## ACKNOWLEDGEMENTS

© Geeta Sood, 2007. All Rights Reserved.

**THIS THESIS IS DEDICATED TO  
MY MOTHER KAMLESH SOOD:**

**FOR ALL YOUR LOVE AND UNDERSTANDING, AND  
FOR ALL YOUR SUPPORT AND ENCOURAGEMENT**

*~ A mother's love can never be repaid, but I dedicate this Thesis to you;  
to repay you for your kindness, for your encouragement and  
for everything you have ever done for me .  
I hope it makes you proud of me ~*

**~ I also thank my sister Neeta and my brother Neil  
for their humour and thoughtfulness, and  
mostly for always being there for me ~**

**With All My Love Always**

**Geeta**

## ACKNOWLEDGEMENTS

I would very much like to thank both my supervisors Dr William Fraser and Dr Carl Schwalbe for their expert advice and knowledge, and for their reassuring guidance during my time at Aston. I will always regard their constant aid and assistance as invaluable.

I would also like to acknowledge my industrial sponsors Celltech Therapeutics in Slough for their financial support and interest in the project. Special thanks to Dr Mike A. W. Eaton and Mr Jim Turner for their regular contact and assistance.

I am grateful to EPSRC for their financial support towards this project, and to IUCr for the financial support for presenting my work at the IUCr Conference in St. Louis, Missouri, USA.

Special thanks go to all technical assistance - Mrs Karen Farrow, Mr Mike Davies and Mr Chris Bache, and to Mr Graham Smith for his design work for the crystallographic data. Also, many thanks to Mr Peter Ashton and Miss Claudine Clark for undertaking mass spectral analysis work.

Thanks to all those in the Medicinal Chemistry section I have had the pleasure of working with whilst undertaking this research, in particular Andrew Walsh, and to those in the Pharmaceutical Sciences division who have aided, assisted and guided me.

Finally and very importantly, special thanks go to my very dearest and closest friends: Deeps, Sonia, L.Sanj, Jas, Raj, Sat, Tony, Bob, and Tar. Thank you for your humour, your support, your patience, but mostly for your genuine friendship! Special thanks also go to Gurmit and Pushpa for your support and affection. It did not go unnoticed.

# ABBREVIATIONS

A	adenine
Ac	acetyl
AC	6-amino-2-chloropurine
Aeg	<i>N</i> -(2-aminoethyl)glycine
Ar	argon
$\beta$ -Ala	$\beta$ -alanine
Bn	benzyl
Boc	<i>tert</i> -butyloxycarbonyl
<i>N</i> -Boc	<i>N-tert</i> -butoxycarbonyl
BBB	blood brain barrier
BOP	1-benzotriazolyl- <i>N</i> -oxytris(dimethylamino)-phosphoniumhexafluorophosphate/benzotriazole-1-yl-oxy-tris(methylamino)phosphonium hexafluorophosphate
Bz	benzoyl
C	cytosine
$^{\circ}$ C	degrees celsius
Cbz	benzyloxycarbonyl/carbobenzoxy
CIF	crystallographic information file
CHCl <sub>3</sub>	chloroform
CH <sub>2</sub> Cl <sub>2</sub>	dichloromethane
CH <sub>3</sub> CN	acetonitrile
d	day
D	2,6-diaminopurine
D-Ala	D-alanine
DL-Ala	DL-alanine
DCC	<i>N,N'</i> -dicyclohexylcarbodiimide
DCM	dichloromethane
DhbtOH	3,4-dihydro-4-oxo-1,2,3-benzotriazine
DIEA	<i>N,N</i> -diethylamine
DIPEA	<i>N,N</i> -dipropylethylamine
DMAP	4-dimethylaminopyridine
DMF	<i>N,N</i> -dimethylformamide
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dsDNA	double-stranded deoxyribonucleic acid
Dz	2,6-diazidopurine
Et <sub>3</sub> N	triethylamine
EtOAc	ethyl acetate

EtOH	ethanol
ESMS	electrospray mass spectrometry
<i>f</i>	atomic scattering factor
FABMS	fast atom bombardment mass spectrometry
Fmoc	9-fluorenylmethoxycarbonyl
$F_c/F_{calc}$	calculated structure factor
$F_o/F_{obs}$	observed structure factor
g	gram
G	guanine
Gly	glycine
h	hour(s)
H	hypoxanthine/inosine
H-bonding	hydrogen bonding
HATU	<i>O</i> -(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate/2-(1 <i>H</i> -9-azabenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	<i>O</i> -(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HCl	hydrochloric acid
HMPA	hexamethylphosphoric triamide
HOAt	<i>N</i> -1-hydroxy-7-azabenzotriazole
HOBt	<i>N</i> -1-hydroxy-7-benzotriazole
HPLC	high pressure liquid chromatography
I	inosine/hypoxanthine
IsoG	isoguanine
IUCr	International Union of Crystallography
IR	infra-red
J	pseudoisocytosine
L	litre
L-Ala	L-alanine
LSIMS	liquid secondary ion mass spectrometry
M	molar
5 <sup>m</sup> C	5-methylcytosine
m5 <sup>ox</sup> C	5-methyl-6-oxo-cytidine
8 <sup>ox</sup> A	<i>N</i> <sup>6</sup> -methyl-8-oxo-adenine
mg	milligram
min	minute
mL	millilitre
mM	millimolar
mmol	millimole

## CONTENTS

mol	mole
mp	melting point
MP	methylphosphonate
MeOH	methanol
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
NaN <sub>3</sub>	sodium azide
NaOH	sodium hydroxide
NMR	nuclear magnetic resonance
ODN	oligodeoxynucleotide
Pet. Ether	petroleum ether fraction(s)
PNA	peptidic nucleic acid/peptide nucleic acid
ppm	parts per million
PS	phosphorothioate
PyBOP	benzotriazolyl-oxy-tris[pyrrolidino]-phosphonium hexafluorophosphate/ benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
PyBroP	bromo-tris-pyrrolidino-phosphonium hexafluorophosphate
R	discrepancy factor
RNA	ribose nucleic acid
RNase H	ribonuclease H
r.t.	room temperature
s	second
ssDNA	single-stranded DNA
T	thymine
Temp.	temperature [Kelvin]
TBTU	2-(1 <i>H</i> -benzotriazol-1-yl)-1,1,3,3,-tetramethyluronium
TFA	trifluoroacetic acid
TFO	triplex-forming oligonucleotide
THF	tetrahydrofuran
TLC	thin-layer chromatography
tRNA	transfer ribonucleic acid
UV	ultra-violet
%	percent
μg	microgram
μL	microlitre
μm	micrometre
μM	micromolar
μmol	micromole
X	xanthine
Z	benzyloxycarbonyl/carbobenzoxy



# CONTENTS

	Page
Title	1
Abstract	2
Dedication	3
Acknowledgements	4
Abbreviations	5
Contents	8
List of Figures	18
List of Tables	22

## CHAPTER 1

### **Introduction: The Structural and Molecular Recognition Properties of Peptidic Nucleic Acids (PNAs) Compared With DNA Oligonucleotides and Analogues**

<b>1.1</b>	<b>The Aims and Objectives of the Research Project</b>	<b>23</b>
<b>1.2</b>	<b>Peptidic Nucleic Acids - DNA Mimetics</b>	<b>23</b>
<b>1.2.1</b>	Why Modify the DNA Backbone?	23
<b>1.2.2</b>	Comparison of the PNA and DNA Backbone Structures	24
<b>1.2.3</b>	The Improved Properties of PNA Compared with DNA Oligomers	25
<b>1.3</b>	<b>Molecular Recognition Processes in Duplex and Triplex Formation</b>	<b>26</b>
<b>1.3.1</b>	The Nature of Hybridisation Interactions Between Purine and Pyrimidine Bases	26
<b>1.3.2</b>	The Antisense and Antigenic Approaches Towards the Regulation of Gene Expression	29
<b>1.3.3</b>	Gene Transcription and Translation Arrest in the Antisense Approach	31
<b>1.3.4</b>	Essential Pharmacokinetic Properties of Oligomers for Antisense Agents for Antisense Applications	32
<b>1.4</b>	<b>Molecular Recognition and Stability Properties of PNAs</b>	<b>33</b>
<b>1.4.1</b>	Requirements for Application of PNA as an Antisense Agent	33
<b>1.4.2</b>	H-Bonding and Base Stacking Interactions in Hybridisation of PNA and DNA Oligomers	34
<b>1.4.3</b>	Hybridisation of PNA with DNA	34
<b>1.4.4</b>	Strand Displacement Properties of PNAs	35

	<b>Page</b>
<b>1.4.5</b> P-Loops as Artificial Transcription Promoters	37
<b>1.4.6</b> Transcription Arrest via Use of PNA Oligomers	38
<b>1.4.7</b> Translation Arrest by PNA Oligomers	38
<b>1.4.8</b> Comparison of the Hybrid Stability Depending on PNA and DNA Backbone Structures	39
<b>1.4.9</b> The Targeting Ability of PNA Oligomers as a Consequence of the Neutral Backbone	39
<b>1.5</b> Evaluation of Studies on DNA Backbone Modifications	41
<b>1.5.1</b> The Scope for Additional Backbone Modifications in DNA	41
<b>1.5.2</b> Ionic and Non-Ionic Backbones	41
<b>1.5.3</b> Potential Modification of the PNA Backbone	45
<b>1.5.4</b> The Scope for PNA Backbone Modifications : Hybridisation, Stability and Targeting to DNA	48
<b>1.5.5</b> Chimeric PNAs	49
<b>1.6</b> The Future of PNAs	50
<b>1.6.1</b> Therapeutic and Diagnostic Applications of PNAs	50
<b>1.6.2</b> Mechanistic Applications of PNAs	51
<b>1.6.3</b> Evaluation of the Future for PNAs	52

## **CHAPTER 2**

### **The Design of Base-Modified PNA Oligomers**

<b>2.1</b> The Effects on Duplex and Triplex Stability via Base Modification	54
<b>2.1.1</b> Base-Modified Oligomers Incorporating D, and IsoG, X and H for Duplex and Triplex Studies	54
<b>2.1.2</b> Hybridisation Properties of 2,6-Diaminopurine (D) Compared to Adenine	54
<b>2.2</b> pH Independent Triplex Formation via Base Modification	55
<b>2.2.1</b> Why are Base-Modified PNA Oligomers for Triplex Formation of Interest?	55
<b>2.2.2</b> Design of Base-Modified PNA Oligomers for pH-Independent Triplex Formation	57
<b>2.2.3</b> Rationale for the Choice of IsoG, X, and H for Incorporation into PNA Oligomers	57

	<b>Page</b>
<b>2.3 The Known Effects of Modified Bases in DNA and PNA Oligomers</b>	<b>58</b>
2.3.1 5-Methylcytosine (5 <sup>m</sup> C)	58
2.3.2 Pseudoisocytosine (J)	59
2.3.3 Pyrazole Pyrimidines P1 and P2	60
2.3.4 5-Methyl-6-Oxo-Cytidine (m5 <sup>ox</sup> C)	62
2.3.5 N <sup>6</sup> -Methyl-8-Oxo-Adenine (8 <sup>oxa</sup> A)	63
<b>2.4 Use of Non-Natural Bases for pH Circumvention for Triplexation Studies for Potential pH Independent Triplex Formation</b>	<b>63</b>
2.4.1 IsoGuanine (isoG) as a Potential Mimic of C <sup>+</sup> for Stable Triplex-Formation	63
2.4.2 The Importance of N7 Regioisomeric DNA Bases and Their Derivatives	65
<b>2.5 The Effects of Xanthine and Hypoxanthine Upon Molecular Recognition of PNA Oligomers</b>	<b>66</b>
2.5.1 The Known Contributions of Xanthine and Hypoxanthine to the Hybrid Stability of DNA Oligonucleotides	66
2.5.2 The Potential Hybridisation Properties of Xanthine Compared with Hypoxanthine	67
<b>2.6 The Potential of Hypoxanthine as a Universal PNA Base</b>	<b>68</b>
<b>2.7 Conclusion</b>	<b>69</b>

## **CHAPTER 3**

### **The Synthesis of Base-Modified PNA Intermediates**

<b>3.1 Aims of the Synthetic Work</b>	<b>70</b>
<b>3.2 Potential Protecting Groups for the Primary Amino Group of PNA Monomers</b>	<b>70</b>
<b>3.3 Synthesis of The PNA Backbone Component</b>	<b>71</b>
3.3.1 The Synthesis of Ethyl N-(Boc-aminoethyl) glycinate 1	71
<b>3.4 The Synthesis of PNA Monomers Composed of DNA Bases: Literature Precedents</b>	<b>77</b>
<b>3.5 The Synthesis of PNA Monomers Composed of DNA Bases: Current Strategies</b>	<b>81</b>
3.5.1 PNA Monomers Composed of Novel Bases D, Dz, X, H and isoG	81

	<b>Page</b>
3.5.2 The Versatility of 2,6-Dichloropurine <b>5</b> as a Starting Material	82
<b>3.6 Synthesis of the PNA Monomer Incorporating the Novel Base 2,6-Diaminopurine (D) <b>82</b></b>	<b>86</b>
3.6.1 Synthesis of the PNA Building Block Incorporating 2,6-Diaminopurine <b>82</b> via its 2,6-Diazidopurine Precursor <b>84</b>	86
3.6.2 Synthesis of the PNA Building Block Incorporating 2,6-Diaminopurine <b>80</b> via Direct Alkylation of 2,6-Dichloropurine <b>75</b>	88
3.6.3 Protection of the Exocyclic Amino Groups of the 2,6-Diaminopurine Derivative <b>82</b>	91
3.6.4 Coupling of PNA Building Blocks to Ethyl <i>N</i> -(Boc-aminoethyl) glycinate <b>1</b> : Literature Precedents	92
3.6.5 Attempted Coupling of the 2,6-Diaminopurine Building Block <b>84</b> to Ethyl <i>N</i> -(Boc-aminoethyl) glycinate <b>1</b>	94
3.6.6 Recent Literature Describing the Synthesis of the target 2,6-Diaminopurine PNA Monomer <b>6</b>	97
<b>3.7 Attempted Synthesis of the Building Block Incorporating IsoG <b>9</b> Using the Versatile 2,6-Dichloropurine Base <b>2</b></b>	<b>98</b>
<b>3.8 Synthesis of the Building Block Incorporating Xanthine <b>21</b></b>	<b>101</b>
3.8.1 Direct Alkylation of Xanthine <b>3</b> Using Ethyl Bromoacetate <b>27</b>	101
3.8.2 Attempted Synthesis of the PNA Monomer Incorporating Xanthine <b>7</b> from 2,6-Dichloropurine <b>75</b>	101
3.8.3 Attempted Coupling of the Xanthine Building Block <b>87</b> to Ethyl <i>N</i> -(Boc-aminoethyl) glycinate <b>1</b>	104
<b>3.9 The Synthesis of the PNA Monomer Incorporating Hypoxanthine</b>	<b>104</b>
3.9.1 Direct Alkylation of Hypoxanthine <b>4</b> using Ethyl Bromoacetate <b>27</b>	104
3.9.2 6-Chloropurine <b>89</b> as a Versatile Starting Material for Synthesis of the PNA Monomer Incorporating Hypoxanthine <b>4</b>	105
<b>3.10 The Chemical Shift Values of Related Purine Analogues from <sup>1</sup>H and <sup>13</sup>C NMR</b>	<b>108</b>

## **CHAPTER 4**

### **X-Ray Crystallography for Structure Determination**

<b>4.1 The Uses of X-Ray Crystallography</b>	<b>112</b>
--	------------

	<b>Page</b>
4.1.1 The Value of X-ray Crystallography and Molecular Modelling Techniques for Structure Characterisation	112
<b>4.2 The Crystal Analysis Procedure</b>	<b>113</b>
4.2.1 The Principal Data Requirements	113
4.2.2 Summary of the Stages in Structure Determination	113
<b>4.3 Crystal Geometry, Symmetry and Space Groups</b>	<b>115</b>
4.3.1 The Symmetry of the Crystal - The Unit Cell	115
4.3.2 The Miller Indices	115
4.3.3 The Reciprocal Cell	117
4.3.4 Bragg's Law in Reciprocal Space	117
4.3.5 Crystal Systems and Their Symmetry	118
4.3.6 Point Groups and Space Groups	119
<b>4.4 Intensity Data Collection</b>	<b>122</b>
4.4.1 The Wavelength of X-Rays	122
4.4.2 Diffraction of X-rays and Use of a Monochromator	122
4.4.3 Geometrical Parameters from Diffraction Data	123
4.4.4 Crystal Selection, Mounting and Alignment	124
<b>4.5. Data Reduction</b>	<b>125</b>
4.5.1 Applied Corrections to Intensity Data	125
4.5.2 The Lorentz-Polarisation Correction	125
4.5.3 Absorption	126
4.5.4 Crystal Deterioration and Percentage (%) Decay	127
<b>4.6 Data Conversion</b>	<b>127</b>
4.6.1 Scaling	127
4.6.2 Absolute Scaling and Temperature Factors	127
4.6.3 Preliminary Stages to the Final Structure	128
<b>4.7 The Phase Problem</b>	<b>128</b>
<b>4.8 Determination of Atomic Parameters</b>	<b>129</b>
4.8.1 Direct Methods	129
<b>4.9 Structure Refinement</b>	<b>130</b>
4.9.1 Fourier Syntheses	130
4.9.2 Least Squares	132

	<b>Page</b>
<b>4.9.3</b> Weighting Scheme	132
<b>4.10.1</b> Determination of Structure Correctness	133
<b>4.10.2</b> The Discrepancy Factor ( <i>R</i> )	133
<b>4.11</b> Molecular Plotting	133

## **CHAPTER 5**

### **Structure Analysis of PNA Backbone Intermediates**

<b>5.1</b> X-Ray Structural Analysis and Evaluation	135
<b>5.2.1</b> The Value of X-Ray Crystallography for Structural Analysis of PNA Backbone Precursors	135
<b>5.2.2</b> The Focal Points of This Crystallographic Work	135
<b>5.3</b> General Methods	136
<b>5.4</b> Comparison of the N9 6-Chloropurine 125 with the Regioisomeric N7 6-Chloropurine 126 Ring System and the Effect of Chloro Substitution at C2	136
<b>5.4.1</b> Crystal Data for 6-Chloro-(9-carboxymethyl)purine ethyl ester 125	136
<b>5.4.2</b> Crystal Data for 6-Chloro-(7-carboxymethyl)purine ethyl ester 126	136
<b>5.4.3</b> Crystal Data for 2,6-Dichloro-(9-carboxymethyl)purine ethyl ester 76	136
<b>5.4.4</b> Geometrical Features of 2,6-Dichloro-(9-carboxymethyl)purine ethyl ester 76	141
<b>5.5</b> Replacement of the C2 and C6 Chloro Groups with Amino, Azido and Methoxy Substituents	144
<b>5.5.1</b> Crystal Data for 2,6-Diamino-(9-carboxymethyl)purine ethyl ester 80	144
<b>5.5.2</b> Crystal Data for Ethyl-Adenin-9-yl acetate 94	144
<b>5.5.3</b> Comparison between Ethyl-Adenin-9-yl Acetate 94 and 2,6-Diamino-(9-carboxymethyl)purine ethyl ester 80	144
<b>5.5.4</b> Evaluation of Structural Changes in 80, 84 and 92 on Replacement of the Chloro Moieties with Amino (80), Azido (84) and Methoxy (82) Substituents	149

	<b>Page</b>
5.5.5 Crystal Data for 6-Amino-2-Methoxy-(9-carboxymethyl)purine methyl ester <b>92</b>	149
5.5.6 Crystal Data for 2,6-Diazido-(9-carboxymethyl)purine methyl ester <b>84</b>	149
5.6.1 <b>Characterisation of Diethyl 6-Oxo-6,7-dihydro-3<i>H</i>-purine-3,7-diacetate <b>124</b></b>	<b>153</b>
5.6.2 Crystal Data for Diethyl 6-Oxo-6,7-dihydro-3 <i>H</i> -purine-3,7-diacetate <b>124</b>	153
<b>5.7 Evaluation of PNA Backbone Intermediates</b>	<b>156</b>
5.7.1 Structural Comparison of the Boc-Protected Intermediate <b>17</b> with Some Reported Analogues	156
5.7.2 Crystal Data for <i>N</i> -(Boc-amino)acetonitrile <b>17</b>	156
5.7.3 Comparison of 1-Hydroxyiminophenylcarbamate <b>20</b> at 213 K and 293 K	158
5.7.4 Crystal Data for <b>20</b>	158
<b>5.8 Evaluation of Packing Features of Related Heterocyclic Compounds</b>	<b>161</b>
5.8.1 Information Derived from Stacking Diagrams	161
5.8.2 Base Stacking and H-Bonding Interactions	161
5.8.3 Stacking Arrangement of 6-Chloro-(9-carboxymethyl)purine ethyl ester <b>125</b>	162
5.8.4 Stacking Arrangement of 6-Chloro-(7-carboxymethyl)purine ethyl ester <b>126</b>	162
5.8.5 Stacking Arrangement of 2,6-Dichloro-(9-carboxymethyl)purine ethyl ester <b>77</b>	165
5.8.6 Stacking Arrangement of 2,6-Diamino-(9-carboxymethyl)purine ethyl ester <b>80</b>	165
5.8.7 Stacking Arrangement of 6-Amino-2-methoxy-(9-carboxymethyl)purine methyl ester <b>92</b>	168
5.8.8 Stacking Arrangement of 2,6-Diazido-(9-carboxymethyl)purine methyl ester <b>84</b>	168

	<b>Page</b>
<b>5.8.9</b> Stacking Arrangement of Diethyl 6-Oxo-6,7-Dihydro-3 <i>H</i> -purine-3,7-diacetate <b>124</b>	168

## **CHAPTER 6**

### **Experimental**

<b>6.1</b> Experimental	172
<b>6.1.1</b> General Methods	172
<b>6.2</b> Chemical Synthesis	172
<b>6.2.1</b> ( <i>N</i> -Boc-amino)acetonitrile <b>17</b>	172
<b>6.2.2</b> <i>N</i> -Boc-ethylenediamine <b>14</b>	172
<b>6.2.3</b> <i>tert</i> -butyl-4-Nitrophenylcarbonate <b>24</b>	173
<b>6.2.4</b> <i>N</i> -Boc-ethylenebenzylamine <b>21</b> from <i>N</i> -(Boc)ethylenebenzylamine <b>14</b> , with the side product 1-Hydroxyiminophenylcarbamate <b>20</b>	174
<b>6.2.5</b> Ethyl <i>N</i> -(Boc-aminoethyl)glycinate <b>1</b>	174
<b>6.2.6</b> <i>N</i> -(2-aminoethyl)glycine <b>29</b>	175
<b>6.2.7</b> 2,6-Dichloro-9-(carboxymethyl)purine ethyl ester <b>76</b> and 2,6-dichloro-7-(carboxymethyl)purine ethyl ester <b>78</b>	175
<b>6.2.8</b> 2,6-Dichloro-9-(carboxymethyl)purine methyl ester <b>75</b> and 2,6-Dichloro-7-(carboxymethyl)purine methyl ester <b>79</b>	176
<b>6.2.9</b> 6-Chloro-9-(carboxymethyl)purine ethyl ester <b>125</b> and 6-chloro-7-(carboxymethyl)purine ethyl ester <b>126</b>	177
<b>6.2.10</b> 6-Chloro-9-(carboxymethyl)purine methyl ester <b>127</b> and 6-chloro-7-(carboxymethyl)purine methyl ester <b>128</b>	177
<b>6.2.11</b> 2,6-Diazido-9-(carboxymethyl)purine methyl ester <b>84</b>	178
<b>6.2.12</b> 2,6-Diamino-9-(carboxymethyl)purine methyl ester <b>81</b>	178
<b>6.2.13</b> 2,6-Diamino-9-(carboxymethyl)purine ethyl ester <b>80</b>	179
<b>6.2.14</b> 2,6-Diamino-9-(carboxymethyl)purine)acetic acid <b>82</b>	179
<b>6.2.15</b> 6-Amino-9-(carboxymethyl)-2-methoxypurine methyl ester <b>85</b>	180
<b>6.2.16</b> <i>N</i> -(Benzyloxycarbonyl)imidazole <b>100</b>	180
<b>6.2.17</b> 1-(Benzyloxycarbonyl)-3-ethylimidazolium tetrafluoroborate <b>98</b>	181



	<b>Page</b>
<b>6.2.18</b> 6-Amino-2-chloropurine <b>110</b>	181
<b>6.2.19</b> 6-Amino-2-chloro-9-(carboximido)purine <b>112</b>	181
<b>6.2.20</b> Attempted synthesis of 6-amino-2-benzyloxy-9(carboxymethyl)- purine ethyl ester <b>113</b>	181
<b>6.2.21</b> 2,6-Dibenzyloxy-xanthin-9-yl acetic acid <b>121</b>	182
<b>6.2.22</b> 2,6-Diallyloxy-xanthin-9-yl acetic acid <b>122</b>	183
<b>6.2.23</b> Xanthin-9-yl acetic acid <b>87</b>	183
<b>6.2.24</b> Diethyl 3,7-hypoxanthyl diacetate <b>124</b>	183
<b>6.2.25</b> 6-Benzyloxy-hypoxanthin-9-yl acetic acid <b>132</b>	184
<b>6.2.26</b> 6-Benzyloxycarbonyl-hypoxanthin-7-yl acetic acid <b>135</b>	184
<b>6.2.27</b> Attempted Synthesis of <i>N</i> -(2-Boc-aminoethyl)- <i>N</i> -(diazido-9-acetyl)- glycine <b>102</b>	185
<b>6.2.28</b> Attempted Synthesis of <i>N</i> -(2-Boc-aminoethyl)- <i>N</i> -(diaminopurin-9- acetyl)glycine <b>71</b>	185
<b>6.2.29</b> Attempted Synthesis of <i>N</i> -(2-Boc-aminoethyl)- <i>N</i> -(xanthin-9- acetyl)glycine <b>7</b>	186
<b>6.2.30</b> Attempted Synthesis of <i>N</i> -(2-Boc-aminoethyl)-2,6-dibenzyloxy- <i>N</i> - (xanthin-9-acetyl)glycine <b>73</b>	186
<b>6.2.31</b> Attempted Synthesis of <i>N</i> -(2-Boc-aminoethyl)-6-benzyloxy- <i>N</i> - (hypoxanthin-9-acetyl)glycine <b>72</b>	186
<b>6.3 Crystallography</b>	<b>187</b>
<b>6.3.1</b> Data Collection	187
<b>6.3.2</b> 6-Chloro-9-(carboxymethyl)purine ethyl ester <b>125</b>	188
<b>6.3.3</b> 6-Chloro-7-(carboxymethyl)purine ethyl ester <b>126</b>	188
<b>6.3.4</b> 2,6-Dichloro-9-(carboxymethyl)purine ethyl ester <b>76</b>	188
<b>6.3.5</b> 2,6-Diamino-9-(carboxymethyl)purine ethyl ester <b>80</b>	188
<b>6.3.6</b> 6-Amino-2-methoxy-(carboxymethyl)purine methyl ester <b>92</b>	188
<b>6.3.7</b> 2,6-Diazido-9-(carboxymethyl)purine methyl ester <b>84</b>	188
<b>6.3.8</b> Diethyl 3,7-hypoxanthyl diacetate <b>124</b>	189
<b>6.3.9</b> ( <i>N</i> -Boc-amino)acetonitrile <b>17</b>	189
<b>6.3.10</b> 1-Hydroxyiminophenylcarbamate <b>20</b>	189

	<b>Page</b>
<b>REFERENCES</b>	<b>190</b>
<b>APPENDIX</b>	<b>202</b>

## List of Figures

	<b>Page</b>
<b>Figure 1.1</b>	Comparison of the PNA and DNA backbone structures. 24
<b>Figure 1.2</b>	Structures of unsubstituted purine and pyrimidine bases. 26
<b>Figure 1.3</b>	Watson-Crick base pairing in (a) A.T and (b) G.C duplexes. 27
<b>Figure 1.4</b>	Complementary Watson-Crick and Hoogsteen pairing of (a) T*A.T and (b) C <sup>+</sup> *G.C triplexes. 28
<b>Figure 1.5</b>	Formation of (a) T*A.T, (b) A*A.T and (c) G*G.C homopurine-targeted triple helices. 29
<b>Figure 1.6</b>	(a) Normal gene expression inhibited by (b) Antisense and (c) Antigene compounds. 30
<b>Figure 1.7</b>	The current status of the multiple mechanisms involved in the action of antisense oligonucleotides. 32
<b>Figure 1.8</b>	Key hurdles to therapeutic applications of antisense oligonucleotides. 33
<b>Figure 1.9</b>	The three requirements for effective application of PNAs as antisense agents. 34
<b>Figure 1.10</b>	Strand displacement P-loop complex formation between dsDNA and homopyrimidine PNA. 36
<b>Figure 1.11</b>	The guanidinium (DNG) backbone. 43
<b>Figure 1.12</b>	Potential sites for modification on the PNA backbone. 45
<b>Figure 1.13</b>	Modifications to the repeating unit in the original PNA structure. 47
<b>Figure 1.14</b>	A chimeric PNA. 49
<b>Figure 2.1</b>	Comparison of Watson-Crick pairing between (a) A.T and (b) D.T (and D.U) base pairs in DNA oligomers. 55
<b>Figure 2.2</b>	Targeting of an A.T DNA base pair by (a) T and (b) Br <sup>5</sup> U in the third antigene strand. 56
<b>Figure 2.3</b>	Targeting of a G.C DNA base pair using C <sup>+</sup> or 5 <sup>m</sup> C in the third antigene strand. 56
<b>Figure 2.4</b>	The (a) N3 and (b) N5 tautomeric forms of pseudoisocytosine (J). 59
<b>Figure 2.5</b>	Representation of the two tautomeric forms of pseudoisocytosine (J) in a (a) G.J base pair and (b) a J*G.C base triplet. 59
<b>Figure 2.6</b>	Hydrogen bonding in (a) P1*G.C and (b) P2*G.C base triplets. 61
<b>Figure 2.7</b>	Hydrogen bonding in the G <sup>7N</sup> *G.C triplet. 62

<b>Figure 2.8</b>	Triplex formation of m <sup>5</sup> oxC and a G.C base pair.	62
<b>Figure 2.9</b>	Triplex formation of 8 <sup>oxo</sup> A with an (a) G.C base pair and (b) an A.U base pair.	63
<b>Figure 2.10</b>	IsoC in a Watson-Crick pair with (a) the N1 and (b) N3 tautomers of isoG.	64
<b>Figure 2.11</b>	The (a) N1 and (b) N3 tautomers of isoG.	64
<b>Figure 2.12</b>	Hybridisation of isoG with a G.C base pair.	65
<b>Figure 2.13</b>	The isomorphous (a) X <sup>7N*</sup> A.T and (b) G <sup>7N*</sup> G.C triplexes.	66
<b>Figure 2.14</b>	Isomorphism of (a) H <sup>*</sup> A.T and (b) X <sup>*</sup> A.T with isoG <sup>*</sup> G.C triplets.	67
<b>Figure 2.15</b>	Ionisation of (a) X compared with (b) H.	67
<b>Figure 2.16</b>	Targeting of the dzaX base to A.T base pairs.	68
<b>Figure 3.1</b>	Three alternative protecting groups to Boc for the amino moiety of <b>1</b> .	70
<b>Figure 3.2</b>	Coupling of protected PNA monomer units P <sub>1</sub> and P <sub>2</sub> .	71
<b>Figure 3.3</b>	Comparison of <sup>1</sup> H NMR spectra of (a) synthesised ethyl <i>N</i> -(Boc-aminoethyl)glycinate <b>1</b> and (b) commercially available <b>1</b> .	75
<b>Figure 3.4</b>	Protected PNA building block monomers, composed of the bases (a) T, (b) C, (c) G, (d) A, and (e) J.	77
<b>Figure 3.5</b>	Benzyloxycarbonyl, benzyl and allyl protecting groups.	78
<b>Figure 3.6</b>	Boc-protected PNA monomers composed of the novel isoG, D, Dz, X and H purine bases.	82
<b>Figure 3.7</b>	Mechanism of alkylation of 2,6-dichloropurine in the presence of the weak base K <sub>2</sub> CO <sub>3</sub> .	85
<b>Figure 3.8</b>	Proposed mechanism for the alkylation of 2,6-dichloropurine <b>75</b> in the presence of the strong base NaH.	90
<b>Figure 3.9</b>	Resonance forms of intermediates on substitution of <b>75</b> C2 and C6.	98
<b>Figure 3.10</b>	Likely mechanisms of ester to the acetic acid <b>121</b> conversion.	103
<b>Figure 3.11</b>	Relative positions of selected <sup>1</sup> H and <sup>13</sup> C shifts.	110
<b>Figure 4.1</b>	Flow diagram showing stages towards the determination of crystal structures of small molecules.	114
<b>Figure 4.2</b>	Representation of the unit cell dimensions of a crystal.	116
<b>Figure 4.3</b>	Example of a lattice plane and an interleaving plane.	116
<b>Figure 4.4</b>	Significance of the symbols in Bragg's law.	118

<b>Figure 4.5</b>	Representations of the seven crystal systems; a) triclinic, b) monoclinic, c) orthorhombic, d) trigonal, e) hexagonal, f) tetragonal, g) cubic.	118
<b>Figure 4.6</b>	A point symmetry operation; a centre of symmetry produced by 2-fold axis combined with reflection.	121
<b>Figure 5.1</b>	ORTEPII view <sup>263</sup> of 6-chloro-(9-carboxymethyl)purine ethyl ester <b>125</b> .	138
<b>Figure 5.2</b>	ORTEPII view <sup>263</sup> of 6-chloro-(7-carboxymethyl)purine ethyl ester <b>126</b> .	139
<b>Figure 5.3</b>	ORTEPII view <sup>263</sup> of 2,6-dichloro-(9-carboxymethyl)purine ethyl ester <b>75</b> .	142
<b>Figure 5.4</b>	Disposition of the methyl and ethyl ester side chains with respect to the heterocyclic ring.	143
<b>Figure 5.5</b>	ORTEPII view <sup>263</sup> of the independent molecule 2,6-diamino-(9-carboxymethyl)purine ethyl ester <b>80</b> .	145
<b>Figure 5.6</b>	ORTEPII view <sup>263</sup> of ethyl-adenin-9-yl acetate <b>94</b> .	146
<b>Figure 5.7</b>	Packing arrangement of ethyl adenin-yl acetate <b>94</b> .	148
<b>Figure 5.8</b>	ORTEPII view <sup>263</sup> of 6-amino-2-methoxy-(9-carboxymethyl)purine methyl ester <b>92</b> .	150
<b>Figure 5.9</b>	ORTEPII view <sup>263</sup> of 2,6-diazo-(9-carboxy-methyl)purine methyl ester <b>84</b> .	152
<b>Figure 5.10</b>	ORTEPII view <sup>263</sup> of diethyl 6-oxo-6,7-dihydro-3 <i>H</i> -purine-3,7-diacetate <b>124</b> .	154
<b>Figure 5.11</b>	ORTEPII view <sup>263</sup> of <i>N</i> -(Bocamino)acetonitrile <b>17</b> .	157
<b>Figure 5.12</b>	Formation of the three types of H-bonds exhibited by the oxime group.	158
<b>Figure 5.13</b>	ORTEPII view <sup>263</sup> of H-bonding 1-hydroxyimino phenyl carbamate.	159
<b>Figure 5.14</b>	H-Bonding between two molecules of <b>20</b> .	160
<b>Figure 5.15</b>	Stacking arrangement of 6-chloro-(9-carboxymethyl)purine ethyl ester <b>125</b> .	163
<b>Figure 5.16</b>	Stacking arrangement of 6-chloro-(7-carboxymethyl)purine ethyl ester <b>126</b> .	164
<b>Figure 5.17</b>	Stacking arrangement of 2,6-dichloro-(9-carboxymethyl)purine ethyl ester <b>76</b> .	166
<b>Figure 5.18</b>	Stacking arrangement of 2,6-diamino-(9-carboxymethyl)purine ethyl ester <b>80</b> .	167
<b>Figure 5.19</b>	Stacking arrangement of 6-amino-2-methoxy-(9-carboxy methyl)purine methyl ester <b>92</b> .	169

- Figure 5.20** Stacking arrangement of 2,6-diazido-(9-carboxymethyl)purine methyl ester **84**. 170
- Figure 5.21** Stacking arrangement of diethyl 6-oxo-6,7-dihydro-3*H*-purine-3,7-diacetate **124**. 171

## List of Tables

		Page
<b>Table 1.1</b>	Structure of phosphorus-containing backbone modifications.	42
<b>Table 1.2</b>	Structure of phosphorus-free backbone modifications.	42
<b>Table 1.3</b>	Amide, urea, carbamate and alkyl chain backbone modifications.	43
<b>Table 1.4</b>	Table to illustrate the effect of selected modified backbones on $T_m$ and nuclease resistance.	44
<b>Table 1.5</b>	The effect of PNA backbone modifications on $T_m$ .	48
<b>Table 3.1</b>	Possible combinations of coupling reagents for coupling <b>80</b> to protected building blocks.	94
<b>Table 3.2</b>	Different conditions employed to attain successful coupling of <b>80</b> to <b>1</b> .	95
<b>Table 3.3</b>	$^1\text{H}$ Chemical shifts of the series of similarly alkylated purine bases.	108
<b>Table 3.4</b>	$^{13}\text{C}$ Chemical shifts of the series of similarly alkylated purine bases.	109
<b>Table 4.1</b>	Characteristics of the seven crystal systems.	119
<b>Table 5.1</b>	Selected geometrical parameters: bond lengths, bond angles and bond torsion angles for <b>76</b> , <b>125A</b> and <b>125B</b> , and <b>126</b> .	140
<b>Table 5.2</b>	Selected geometrical parameters to enable comparison between <b>80</b> and <b>94</b> .	155
<b>Table 5.3</b>	Selected geometrical parameters for the hypoxanthine derivative <b>124</b> .	156
<b>Table 5.4</b>	Comparison of the cell axes of <b>20</b> obtained at 213 K and 293 K.	160

# CHAPTER 1

## **Introduction: The Structural and Molecular Recognition Properties of Peptidic Nucleic Acids (PNAs) Compared With DNA Oligonucleotides and Analogues**

### **1.1 The Aims and Objectives of the Research Project**

The future of gene therapy,<sup>1,2</sup> an exciting new technology for gene expression manipulation,<sup>3</sup> relies heavily upon the exploitation of the molecular recognition properties<sup>4-6</sup> of synthetic reagents such as Peptidic Nucleic Acids<sup>7,8</sup> (PNAs) (Figure 1.1), with the ability to hybridise sequence-specifically<sup>9,10</sup> to targeted single-stranded (ss) and double-stranded (ds) nucleic acids.<sup>11</sup> A principal objective of this thesis work was to design and synthesise novel PNA monomers composed of non-natural purine bases for subsequent incorporation into PNA oligomers. The use of novel bases was envisaged to potentially ameliorate those hybridisation<sup>12</sup> and stability properties<sup>13</sup> displayed by conventional DNA analogues, and to extend the molecular recognition properties for targeting to nucleic acid sequences<sup>6</sup>.

Crystallographic analysis of crystalline PNA intermediates was undertaken to provide unambiguous structural determination, to probe the regiochemistry of isomeric analogues, and finally to reveal interesting structural features of a new series of similarly substituted purine bases.

### **1.2 Peptidic Nucleic Acids - DNA Mimetics**

#### **1.2.1 Why Modify the DNA Backbone?**

The exigency for new strategies to overcome those problems currently associated with antisense drugs based upon DNA<sup>14</sup> or RNA has resulted in the advent of PNAs.<sup>15,16</sup> Though the double helix of DNA is able to store, retrieve and communicate the genetic information of a living organism, research has shown that use of DNA oligonucleotides as possible therapeutic drugs<sup>17</sup> is significantly limited by the following properties: inadequate stability, rapid degradation in the presence of nucleases, and a poor cellular uptake.<sup>18</sup> DNA oligonucleotides are found endogenously in eukaryotic cells. The development of automated, solid phase nucleic acid chemistry has allowed the rapid synthesis in the laboratory not only of DNA oligomers, but also of synthetic analogues that are currently providing novel gene-regulating molecules.

PNAs were investigated in 1991 for their improved properties by the Danish group led by Peter E. Nielsen.<sup>19-21</sup> Initially, molecular modelling studies identified those nucleic acid analogues with low energy conformations based on a novel peptidic backbone. The



synthesis of PNAs (Figure 1.1) containing the DNA bases A, T, C, and G was subsequently undertaken to improve upon those undesirable properties associated with the natural DNA sugar-phosphate backbone, whilst preserving sequence-specific hybridisation properties of both DNA and RNA. PNAs are a third generation of modified DNA analogues<sup>6,21,23</sup> anticipated as candidates for application as biological tools. These uses include inhibition of abnormal gene expression and potential diagnostic and biomolecular aids in drug therapy, including screening for genetic mutations with possible applications against diseases,<sup>25</sup> particularly those currently recalcitrant to existing therapies.<sup>26,27</sup>

### 1.2.2 Comparison of the PNA and DNA Backbone Structures

PNAs are achiral, uncharged DNA mimetic compounds<sup>15,19,26</sup> composed of covalently linked *N*-(2-aminoethyl)glycine (Aeg) units, to which the heterocyclic purine bases are attached via carboxymethylene linkers<sup>12</sup> (Figure 1.1). PNAs represent a dramatic deviation from the natural DNA structure as the entire deoxyribose sugar-phosphate DNA backbone is substituted with peptidic units, thus affording an improved DNA mimic<sup>13</sup> possessing remarkable properties, notably superior sequence-specificity<sup>28</sup> and exceptional binding affinity<sup>6</sup> compared with DNA oligomers.

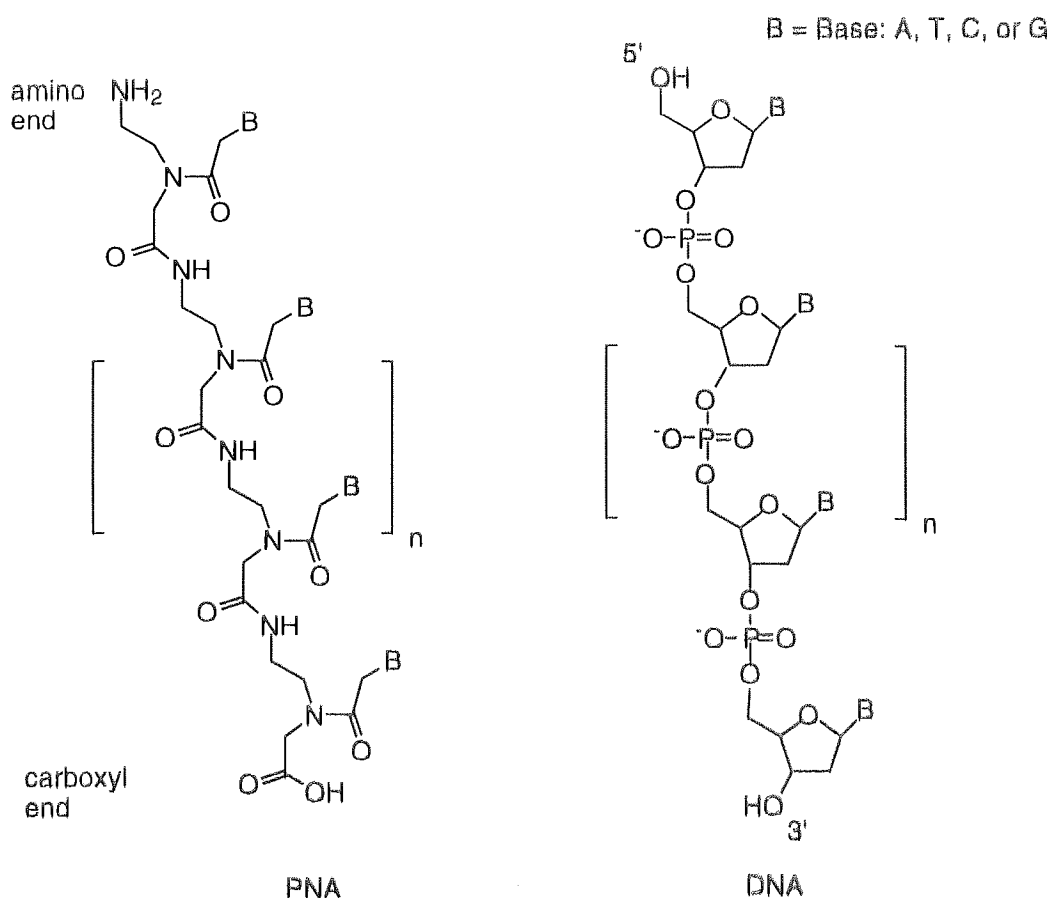


Figure 1.1. Comparison of the PNA and DNA backbone structures.

The PNA and DNA backbones are homomorphous with each other<sup>19</sup> where the PNA backbone monomer unit is composed of six bonds to maintain the exact interbase distances found in natural DNA.<sup>31</sup> Sufficient flexibility is retained permitting correct orientation and unconstrained base pairing with targeted DNA. Analogous to DNA, the PNA monomer unit contains three bonds from the backbone to the base.<sup>32</sup> However, whilst the DNA backbone is comprised of negatively charged phosphates and five-membered sugar rings, the PNA backbone, in contrast, is electrically neutral with the base linked to the backbone through a carboxymethylene linker.

### 1.2.3 The Improved Properties of PNA Compared with DNA Oligomers

The desirable and unique properties of PNAs include the ability to hybridise specifically<sup>26,33,34</sup> and with unprecedented affinity<sup>6,37,38</sup> to complementary ssDNA<sup>37</sup> and dsDNA<sup>22,39</sup> sequences, to form PNA.DNA duplexes<sup>37,40,41</sup> and PNA<sub>2</sub>.DNA (PNA\*PNA.DNA) triplexes<sup>38,39,42</sup> which display enhanced thermal stability. This eventuates in possibilities to control gene expression via triplex formation or strand displacement (section 1.4.4) of targeted genes.<sup>40</sup> The absence of a negative charge, a prominent characteristic of the DNA backbone, eliminates electrostatic repulsion during duplex and triplex formation between PNA and the DNA targets, thus affording complexes of enhanced stability.<sup>22</sup> The presence of achiral linkages<sup>21,43</sup> avoids diastereoisomeric mixtures which are found to be problematic in applications of phosphorothioates and other backbone-modified oligomers in DNA targeting. Regular interbase distances, rigid amide bonds arising from sp<sup>2</sup> hybridisation, the presence of tertiary amide linkages, highly flexible aminoethyl linkers and finally, the ability to hydrogen bond intramolecularly, are all essential factors that govern and contribute towards PNA's favourable properties.<sup>44,45</sup>

As modified glycines comprise part of the PNA backbone, PNA is termed a pseudopeptide. The amide bonds found in the PNA backbone are not identical to those found in natural peptides, and a peptidic backbone with purine and pyrimidine base side chains subsequently results in a combination that is not easily recognised by either nucleases or proteases.<sup>6</sup> Consequently, the half-life of these compounds is markedly extended both *in vitro* and *in vivo* compared with those oligonucleotides with a phosphodiester backbone.<sup>46</sup> The stability of PNAs<sup>47,48</sup> to proteolysis was investigated where PNA was subjected to incubation in the presence of a control peptide. Decathymidyl PNA was shown to remain virtually, structurally unchanged, and observed also to be resistant to aminopeptidases, dipeptidyl aminopeptidases, carboxypeptidases and nucleases. This confirmed that the *N*-(2-aminoethyl)glycine based peptide backbone could not be readily cleaved by proteolytic enzymes.

## 1.3 Molecular Recognition Processes in Duplex and Triplex Formation

### 1.3.1 The Nature of Hybridisation Interactions Between Purine and Pyrimidine Bases

PNAs containing the DNA purine and pyrimidine bases adenine (A), thymine (T), cytosine (C)<sup>49</sup> and guanine (G),<sup>4,5</sup> nature's chemical carriers of genetic information, in addition to the synthetic pyrimidine analogue isopseudocytosine (J) (Figure 1.2), exhibit proven abilities to hybridise with natural nucleic acids. This offers the possibility of regulating gene expression<sup>4,6</sup> in a controlled manner through duplex and triplex interactions with target nucleic acid sequences.<sup>6,50</sup> Specificity of the complementary bases of an oligonucleotide for its targeted nucleic acid sequence derives from the molecular recognition properties governed by Watson-Crick<sup>52,53</sup> and Hydrogen bonding.<sup>54</sup>

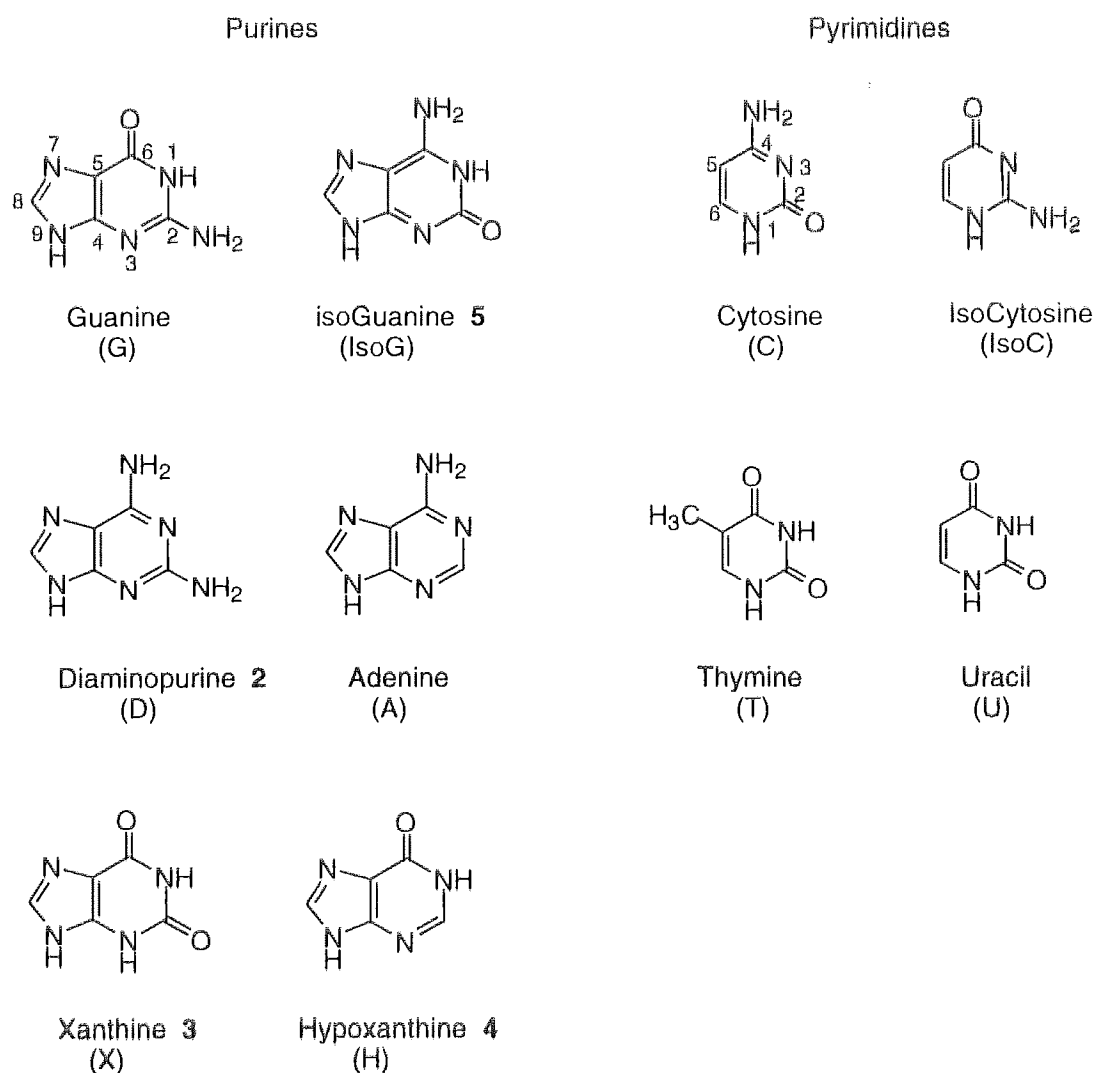
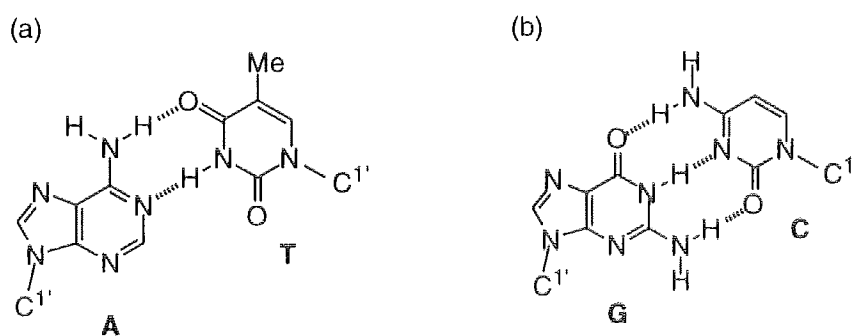


Figure 1.2. Structures of unsubstituted purine and pyrimidine bases.

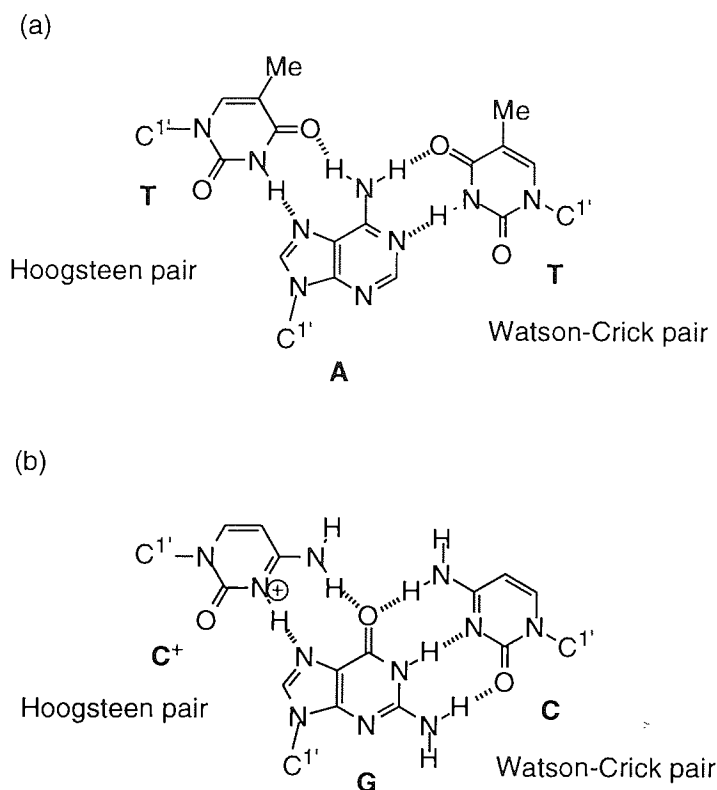
The reversible nature of H-bonding between base pairs allows double helical strands to pair and unpair in exactly the same conformation.<sup>55</sup> This pairing specificity in DNA is the underlying principle of transfer for genetic information, essential for replication, recombination, transcription and translation. Hybridisation between complementary oligomers is characterised by a large enthalpy gain<sup>56,57</sup> and a significant entropy loss, attributable to the formation of a highly ordered and fairly rigid duplex structure that forms from two flexible and much less ordered strands.

Nucleic acids exhibit base pairing properties according to the Watson-Crick bonding motifs,<sup>52</sup> where mutual recognition<sup>4</sup> between the purines A and G, for the pyrimidines T and C respectively (Figure 1.2), results in complementary H-bonding interactions, and thus the formation of A.T and G.C base pairs (Figure 1.3), via H-bond acceptor and donor interactions between the carbonyl and amino groups.<sup>53,58</sup> Triplex-Forming Oligonucleotides (TFOs) containing purine bases recognise polypurine targets of Watson-Crick-paired duplexes,<sup>59</sup> and align in a parallel<sup>60</sup> or antiparallel<sup>61</sup> orientation via Hoogsteen bonding,<sup>5</sup> depending upon the nature of the bases in the TFO; pyrimidine or purine. The two types of triple helices are classified according to their mode of binding. The first of these consists of the recognition of a Watson-Crick A.T base pair in the duplex strand by a Hoogsteen-paired T upon the third strand residing in a parallel orientation in the major groove,<sup>60</sup> and similarly between C<sup>+</sup> for G.C base pairs<sup>5</sup> as shown in Figure 1.4. The stability of the resulting complexes<sup>62-64</sup> stems from the complementary interactions between the bases in T<sup>\*</sup>A.T and C<sup>+</sup>\*G.C TFOs.<sup>65</sup>



**Figure 1.3.** Watson-Crick base pairing in (a) A.T and (b) G.C duplexes.

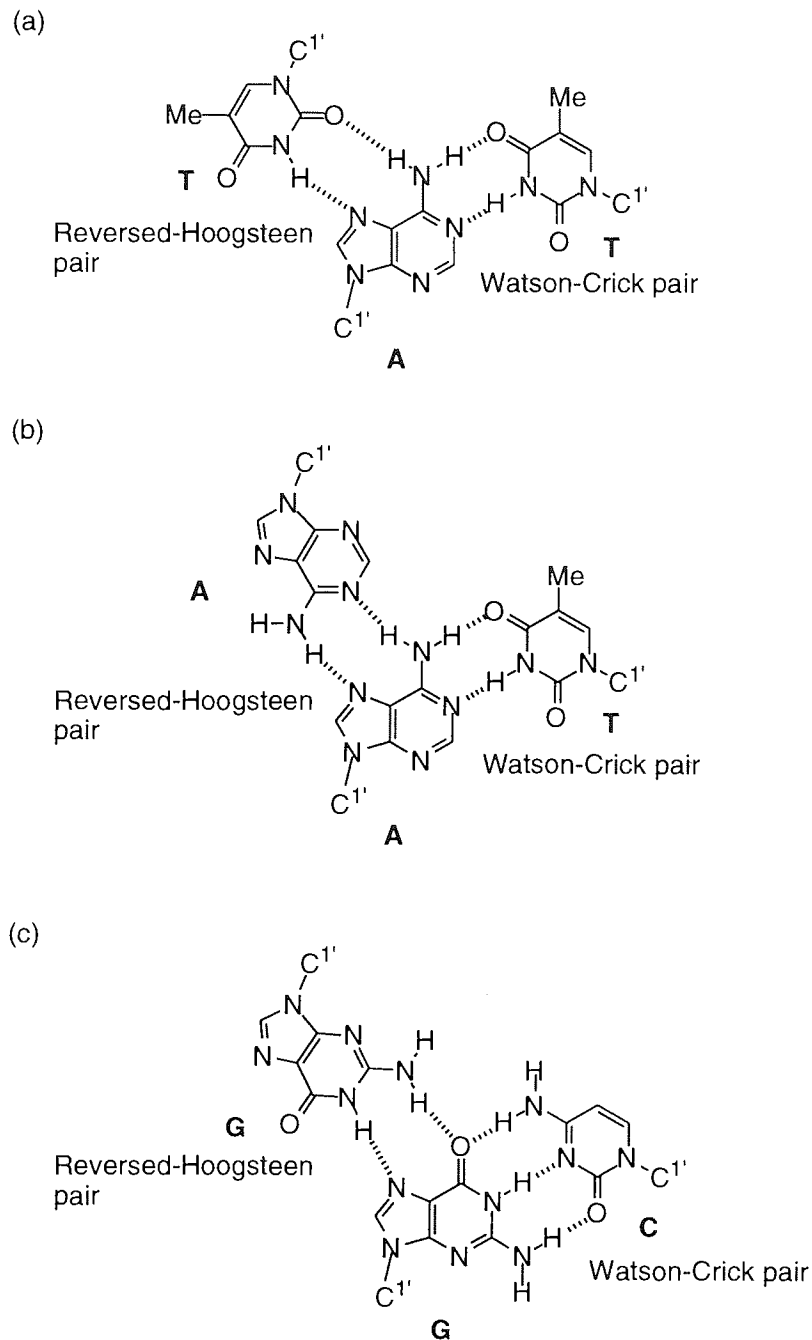
The second mode of binding involves a purine-rich TFO strand aligned in an antiparallel orientation<sup>66</sup> to the purine strand of the target duplex, but involving Reversed-Hoogsteen<sup>51</sup> T<sup>\*</sup>A.T,<sup>67</sup> A<sup>\*</sup>A.T and G<sup>\*</sup>G.C base triplets (Figure 1.5).<sup>66</sup> Whilst offering similar geometry and sites of interaction from T targeted to an A.T base pair, protonation of the N3 of cytosine in acidic media is deemed necessary to maintain the integrity of the Hoogsteen pair, thus permitting its hybridisation with the G.C base pairs.<sup>68</sup>



**Figure 1.4.** Complementary Watson-Crick and Hoogsteen pairing in (a) T\*A.T and (b) C<sup>+</sup>\*G.C triplexes.

However, the resulting isomorphous C<sup>+</sup>\*G.C triplet displays a significant degree of instability at physiological pH and accounts for the strong pH dependence in the binding affinity of C<sup>+</sup>-containing oligodeoxynucleotides.<sup>69,70</sup>

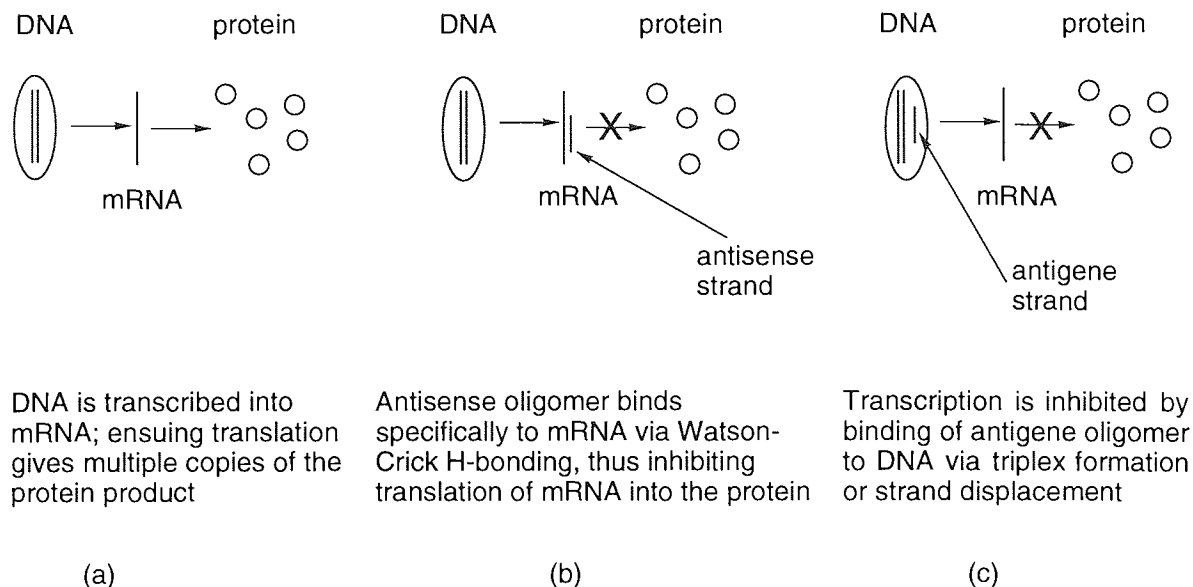
The international interest in synthesising non-natural bases has arisen from attempts to eliminate the exigency for C<sup>+</sup> through replacement with bases which display the necessary H-bond donor sites to allow pH-independent triplex formation.<sup>70</sup> This forms the basis of the design work discussed in Chapter 2. In contrast, A\*A.T and G\*G.C triplexes are non-isomorphous with each other, and are uncharged and stable over a wider pH range (Figure 1.5). The non-isomorphism exhibits destabilising distortions within the TFO backbone. The resulting neutral, rigid and pH-independent base triplets represent bidentate recognition of the base pairs embedded within a duplex structure. Simultaneous binding of pyrimidine oligonucleotides to adjacent purine tracts on alternate strands of the Watson-Crick duplex has also been investigated as a way of extending the structural motif to mixed sequences of the form (pyrimidine)<sub>m</sub>.(purine)<sub>n</sub>, but is not considered further in this work.<sup>55</sup>



**Figure 1.5.** Formation of (a) T\*A.T, (b) A\*A.T and (c) G\*G.C triple helices.

### 1.3.2 The Antisense and Antisense Approaches Towards the Regulation of Gene Expression

Zamackeon and Stephenson<sup>71</sup> proposed modified synthetic oligonucleotides as a class of potential therapeutic agents capable of exerting effects upon the body's cellular system and interacting in a rational way with mRNA.<sup>72</sup> This property was ascribed to their capacity to base pair with a disease-related DNA oligonucleotide, thus inhibiting further synthesis of mRNA.<sup>73,74</sup> Foreordained knowledge of the intended nucleic acid target, sequence, or 'receptor' allows the synthesis of an antisense oligonucleotide<sup>75,76</sup> complementary to the target, hence equating to "rational" drug design.<sup>77,78</sup>



**Figure 1.6.** (a) Normal gene expression inhibited by (b) Antisense and (c) Antigene compounds.

Hybridisation of an antisense oligonucleotide with its target nucleic acid (Figure 3.6) in aqueous solution is a dynamic process reaching a point of equilibrium within a specific set of environmental conditions. By their very nature, antisense oligonucleotides<sup>75</sup> cannot deliver total blockade of a nucleic acid target sequence.

The antigene<sup>79</sup> and antisense<sup>12,80,81</sup> strategies for potential treatment of diseases at the gene expression<sup>209</sup> level have attracted wide attention, in particular regarding PNAs.<sup>11</sup> The axiom of 'Antisense Technology'<sup>83</sup> emanates from the ability of oligonucleotides to impede individual gene expression, and hence inhibit translation processes in a sequence-specific manner. Two potentially useful strategies (Figure 1.6) for the regulation of gene expression are:

1) The antisense strategy:<sup>75</sup> a passive process focusing at the mRNA or pre-mRNA level (Figure 1.6(a)), where hybridisation of a viral RNA via Watson-Crick<sup>52</sup> or Reversed Watson-Crick base pairing is targeted to complementary RNA sequences of a pre-determined mRNA. Gene inhibition is attained via functional blockade, ribosomal reading, or activation of RNaseH; an enzyme that selectively degrades the message. This is achieved by blockage of the sites essential for effective binding of an mRNA molecule (Figure 1.6 (b)), including interaction of ribosomes in transport processes from the nucleus to the cytoplasm. Whereas translation inhibition may be blocked by either a duplex- or a triplex-forming PNA, translation elongation arrest of the ribosome requires a triplex-forming PNA.

2) The antigene strategy:<sup>84,85</sup> the use of TFOs as potential antiviral<sup>86</sup> and therapeutic agents<sup>87</sup> for the suppression of individual genes, for example, oncogenes implicated in

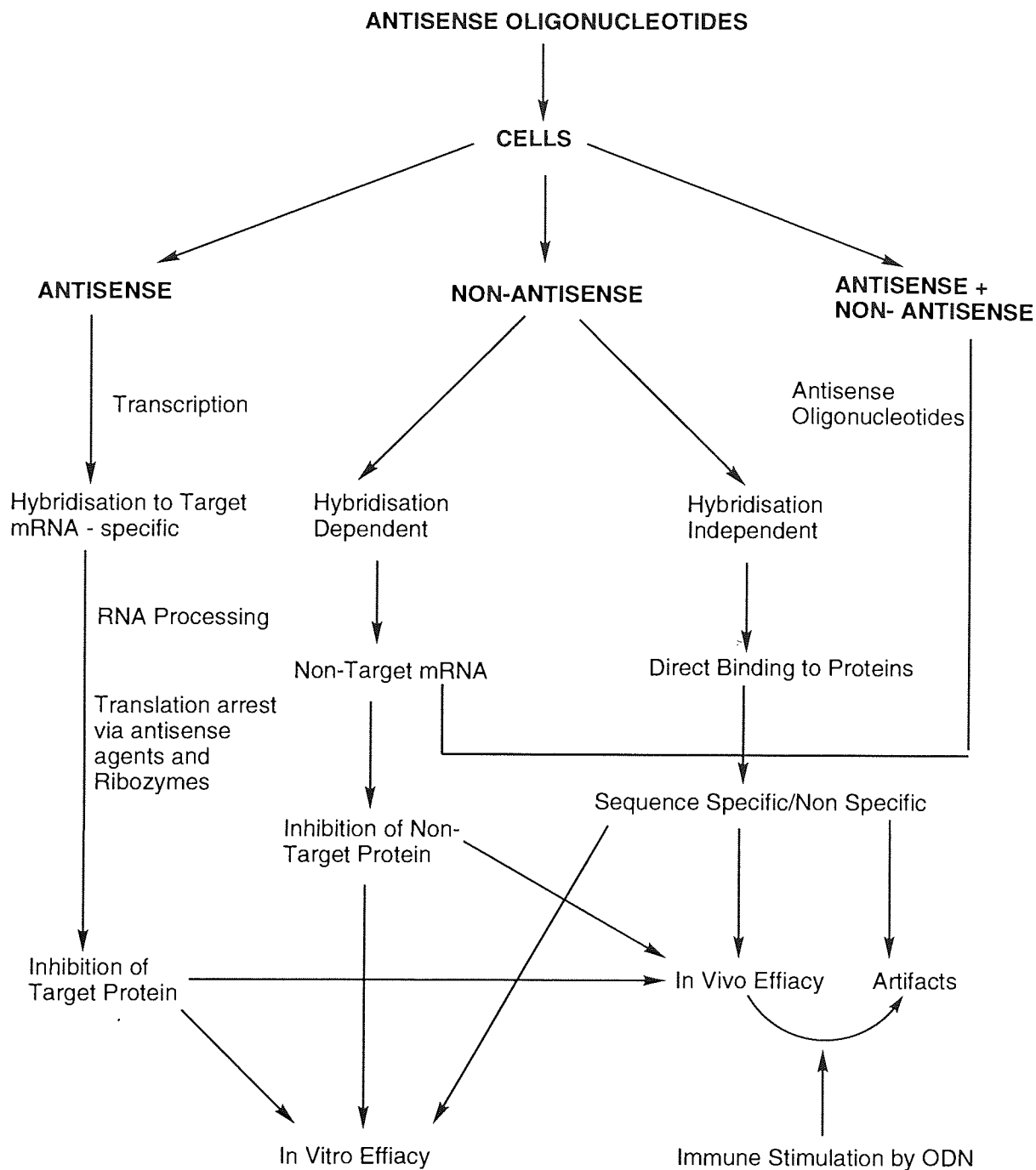
cancer. This strategy offers a new and highly selective chemotherapeutic strategy towards the treatment of human disease as a consequence of the ability of TFOs to inhibit transcription<sup>88,89</sup> of genetic information via site specific interaction of a TFO, such as an mRNA oligonucleotide, for complementary targeted duplex DNA or RNA, using Hoogsteen H-bonding donor and acceptor sites on a purine pairing strand. This affords triple helical molecules,<sup>34,90</sup> whilst preserving the specificity element, thus resulting in translocation arrest, and hence regulation of gene function (Figure 1.6 (c)).<sup>91</sup>

### **1.3.3 Gene Transcription and Translation Arrest in the Antisense Approach**

In analogy with the modified DNA backbones, the interaction of PNA with different targets may be mediated via multiple mechanisms. These include transit of the spliced mRNA from the nucleus to the cytoplasm, translation through inhibition of the assembly of initiation factors such as ribosomal subunits at the start codon, and inhibition of ribosome progression along the coding sequence of the mRNA. PNA oligomers may also modulate gene expression via hybridisation to mRNA<sup>92</sup> in the cytoplasm, thus impeding translation through at least two mechanisms; competition with the translation machinery or induction of mRNA cleavage.<sup>95</sup> Gene expression may also be modulated at the level of transcription<sup>94,95</sup> through triplex formation with dsDNA using TFOs containing contiguous purines on one strand and pyrimidines on another, or via inhibition through hybridisation of the antisense nucleic acid to nascent RNA, to pre-mRNA or to the splicing machinery at intron-exon junctions. These mechanisms are summarised in Figure 1.7, and have culminated in the development of specific tools in molecular biology, medicinal chemistry and pharmacology, providing a rational basis for new therapeutic applications.<sup>97-99</sup> This has consequently revolutionised the applications of PNAs within the diagnostic field due to its portentously modified backbone compared with other backbones based on DNA 2'-deoxyribose.

A major challenge to the antisense field has been the identification of potential therapeutic or diagnostic agents<sup>99</sup> offering effective inhibition over selected target sequences, with no adverse effect upon non-targeted genetic sequences.<sup>100</sup> Certain proteins are down-regulated by the intracellular release of antisense RNA that is complementary to the targeted RNA segment. It has been demonstrated that PNAs may specifically down-regulate gene activity by blocking interaction of the transcription factor,<sup>101</sup> NF- $\kappa$ B, with 5' regulatory DNA sequences to prevent transcriptional transactivation both *in vitro*, and when transfected into cells in culture (Figure 1.7).<sup>101</sup> This is attributable to the ability of PNAs to hybridise to the targeted DNA with pronounced specificity. Gel retardation, circular dichroism, nuclear magnetic resonance (NMR), electrospray mass spectroscopy (MS), and capillary gel electrophoresis techniques have been used to study the unique hybridisation properties exhibited by PNA.<sup>22,26,32</sup>

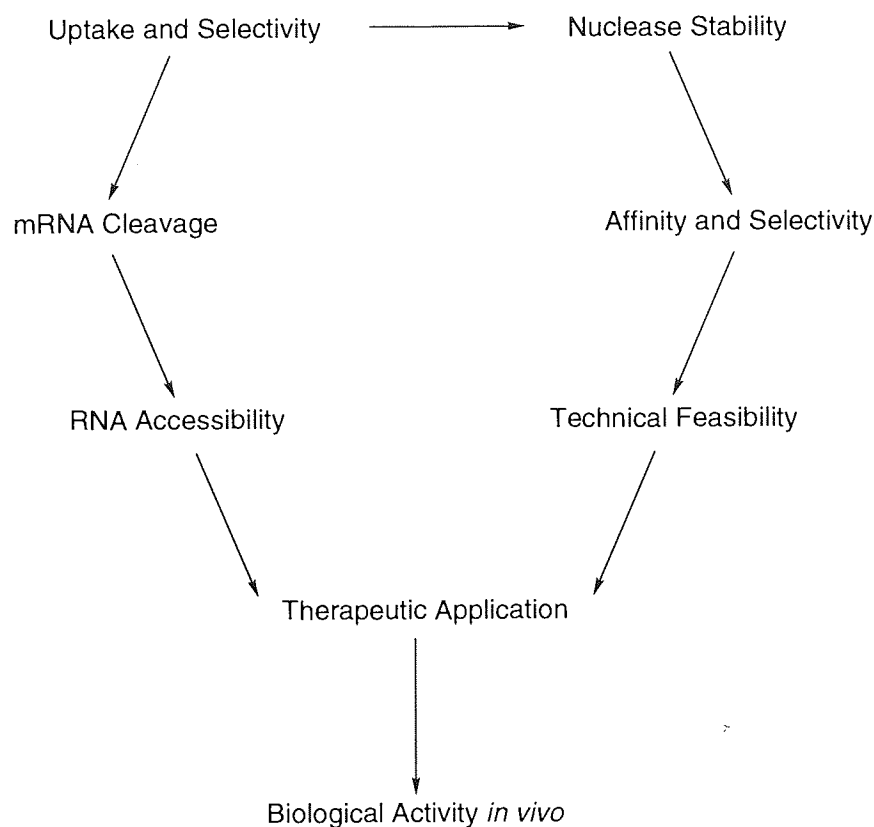




**Figure 1.7.** The current status of the multiple mechanisms involved in the action of antisense oligonucleotides.

### 1.3.4 Essential Pharmacokinetic Properties of Oligomers for Antisense Applications

Abundant evidence indicates substantial developments are required to improve the biological efficacy of PNA, its pharmacokinetic properties, and its bioavailability. Criteria which need to be optimised to make PNA therapeutically viable are shown in Figure 1.8. Oligonucleotide-dependent enzymatic functions are inhibited by PNA.DNA complexes, which also affects cleavage by restriction enzymes, translation and reverse transcription, thus increasing PNA's potential as an attractive chemotherapeutic agent.



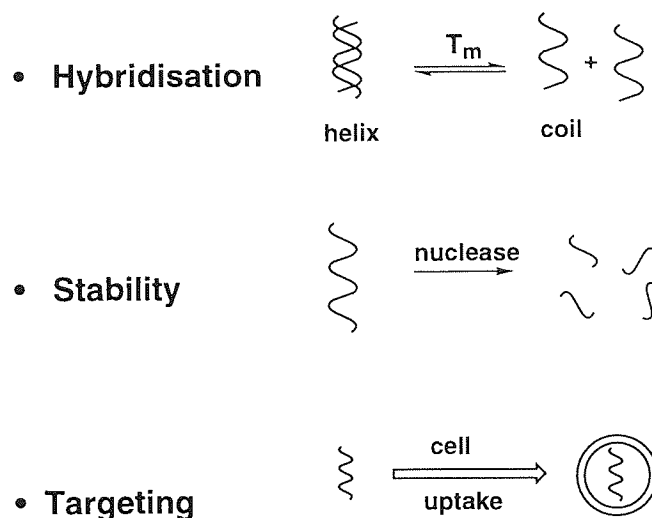
**Figure 1.8.** Key hurdles for therapeutic applications of antisense oligonucleotides.

However, its poor cell uptake, attributable to the absence of a charged backbone,<sup>39,102</sup> limits therapeutic application of PNA oligomers and is currently one of the most important factors warranting investigation in order to overcome those obstacles limiting the applications of PNA oligomers as viable therapeutic agents.

## 1.4 Molecular Recognition and Stability Properties of PNAs

### 1.4.1 Requirements for Application of PNA as an Antisense Agent

Important criteria<sup>75</sup> that need to be fulfilled by any drug for its cellular targets include accessibility to the drug, a high affinity for the cellular target of the drug, specificity for the drug, *in vivo* stability,<sup>103</sup> bioavailability, and favourable pharmacokinetics. The effectiveness of PNA as an antisense agent necessitates consideration of three essentially significant factors: hybridisation, stability and targeting (Figure 1.9), and these are discussed in the following sections.



**Figure 1.9.** The three requirements for effective application of PNAs as antisense agents.

#### 1.4.2 H-Bonding and Base Stacking Interactions in Hybridisation of PNA and DNA Oligomers

The affinity of PNA and other oligomers for their receptor sequences, results from two prominent sources of hybridisation interactions. The two major contributors to the associated free energy of duplex and triplex formation include hydrogen bonding (H-bonding)<sup>53,58</sup> via Watson-Crick base pairing which is base-sequence-independent, and vertical stacking of the co-planar bases (base stacking) which is base-sequence-dependent. Oligonucleotide affinity varies as a function of duplex sequence and length; lowered affinity is attributable to mismatched base pairing and as a function of adjacent bases. The affinity of a PNA antisense oligonucleotide for its target nucleic acid sequence may be increased by structural modification, with the aim of achieving a better blockade at equivalent intracellular concentration compared to the unmodified antisense oligonucleotide. Compared with DNA, it must be realised that the potential risk associated with the higher affinity exhibited by PNA oligomers may result in decreased specificity of the modified antisense oligonucleotide for its target nucleic acid sequence. Ensuing, indiscriminate hybridisation of PNAs to non-targeted sequences is thus likely to incapacitate the cell with adverse consequences upon molecular recognition properties for target sequences.

The properties of PNAs and their molecular recognition modes,<sup>104</sup> arising as a consequence of the neutral backbone,<sup>32</sup> are briefly surveyed.

#### 1.4.3 Hybridisation of PNA with DNA

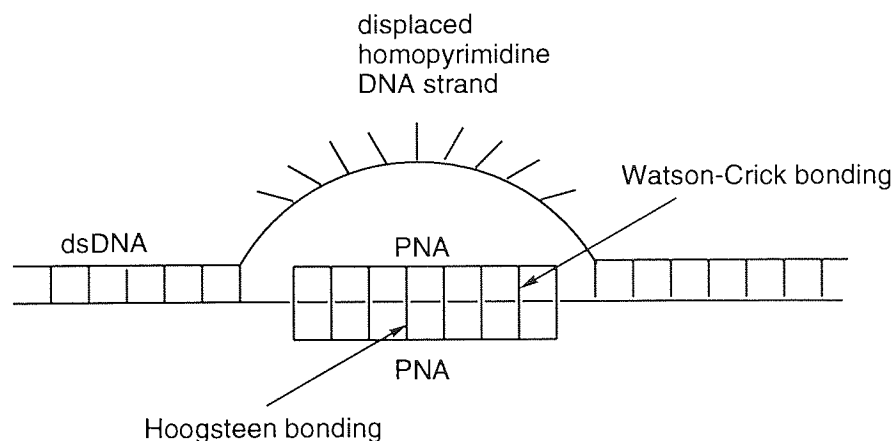
Hybridisation between ssPNA and PNA.DNA duplexes occurs in the parallel orientation (PNA *N*-terminal opposite to the equivalent 5' terminal of DNA) (Figure 1.1) with

lowered affinity, and in the anti-parallel orientation (PNA *N*-terminal opposite the 3' terminal of DNA) with increased affinity, but with negative consequences including undesirable and irreversible hybridisation, and lowered specificity. The preference for the antiparallel orientation appears to be a general feature of mixed sequence PNA.DNA duplexes.<sup>37,105,106</sup> However, adoption of the parallel orientation of a PNA oligomer, with the Hoogsteen strand in triplex formation, results in the significantly more stable (PNA)<sub>2</sub>.DNA triplex.

The specificity<sup>107</sup> of an oligonucleotide for its target sequence results from complementary H-bonding between bases of the antisense oligonucleotide for the target nucleic acids. Base-pairing interactions between sense and antisense nucleic acids generally follow the Watson-Crick H-bonding motifs. These two interactions are discussed for the crystalline compounds synthesised in this work, in the relevant sections of Chapter 5. The specificity of PNAs for complementary sequences was initially investigated by Nielsen *et al*<sup>108,109</sup> with use of the melting temperature ( $T_m$ ); an important parameter in assessing the physiochemical properties of oligonucleotides, and measured as the point at which 50% of the DNA double helix dissociates during helix-coil transition.  $T_m$  was used to characterise the affinity of the PNA oligomer for its targeted sequence, thus determining the duplex stability directly and the effect of mismatch introduction in PNA oligomers.  $T_m$  is altered by phosphodiester backbone modification, and increases with the substitution of modified bases with an ability to form increasingly stable Watson-Crick base pairs with complementary bases. This is exemplified by the presence of the additional H-bond in D, thus forming three H-bonds with T in DNA oligomers compared with its monoamino analogue adenine which can form only two (Figure 2.1). Base-pair mismatches in PNA oligonucleotides opposite GTAC<sup>110</sup> were observed to result in lowered  $T_m$ s, ranging from 8-20°C, thus providing compelling evidence ssDNA recognition by PNA occurs via Watson-Crick A.T and G.C pairing,<sup>4,6</sup> and confirming PNA possesses a higher sequence discrimination compared to DNA. Conclusive evidence regarding these factors was provided in a study by Nielsen *et al*,<sup>5,36</sup> where a single mismatch was observed to be increasingly more destabilising compared with the corresponding mismatch in a DNA.DNA duplex; a single mismatch in mixed PNA.DNA 15-mer was seen to lower  $T_m$  values by 15 °C on average, but only by 11 °C for the corresponding DNA.DNA interactions.

#### 1.4.4 Strand Displacement Properties of PNAs

Those molecular recognition properties exhibited by DNA oligonucleotides are also applicable to the duplexation and triplexation processes displayed by PNAs, but with one very interesting difference.<sup>111</sup> The stoichiometry<sup>112</sup> displayed for PNA oligomers is 2:1 arising from PNA<sub>2</sub>.DNA triplex formation.<sup>113</sup>



**Figure 1.10.** Strand displacement P-loop complex formation between dsDNA and homopyrimidine PNA.

This mode of recognition between PNA and its targets is reflected in PNA's unique hybridisation mode termed P-loop formation,<sup>114,115</sup> or alternatively, strand displacement<sup>116</sup> (Figure 1.1.0). In this hybridisation mode, a homopyrimidine PNA displaces the complementary homopyrimidine DNA strand<sup>12,23</sup> from dsDNA, thus producing a PNA-DNA Watson-Crick duplex which is subsequently trapped by a second PNA strand through Hoogsteen bond<sup>117</sup> formation within the homopurine DNA. This affords an internal PNA\*PNA·DNA triplex in a strand displacement complex (Figure 1.10). The non-complementary DNA strand is extruded as a single stranded loop. Strand displacement<sup>118</sup> results in the formation of a kinetically<sup>119</sup> and thermodynamically<sup>6</sup> stable product. The observed robustness of the strand displacement complex at high ionic strength is indicative of a kinetic trapping<sup>68</sup> of the labile PNA·DNA duplex intermediate, in accordance with sequence discrimination<sup>119</sup> exhibited by DNA towards a double-stranded (ds) PNA target. Hence, the strand displacement mechanism is essentially manipulation of gene expression via transcription initiation. This mechanism has been confirmed by investigations reported by Demidov *et al.*<sup>119,120</sup>

The binding motif is thus context-dependent as homopyrimidine PNAs pair with complementary polypurine targets to form 2:1 stoichiometric (PNA)<sub>2</sub>·DNA complexes as determined by UV-titration curves, and with mixed PNA strands to afford 1:1 PNA·DNA heteroduplexes with a mismatch sensitivity comparable to that found in dsDNA.<sup>115</sup> Though the thermal stability of PNA·ssDNA duplexes is lower than that of homopyrimidine (PNA)<sub>2</sub>·DNA triplexes, it is still higher than corresponding DNA oligomer duplexes, as confirmed by studies where PNA·DNA hybridisation produced monophasic, well-defined melting curves with higher  $T_m$  values than corresponding DNA·DNA duplexes. Furthermore, the observed difference of  $T_m \sim 10$  °C per base pair was characteristic of base stacking interactions. The discovery of a linear increase of  $T_m$  with increasing oligomer length, implied electrostatic contributions for hybridisation from

positive charges was of minor contribution.<sup>75,294</sup>

The pronounced hysteresis in the melting behaviour of PNAs is indicative of a very slow rate of triplex formation. This factor accounts for why a homopyrimidine PNA, when targeted to dsDNA oligomers, prefers the invasion of a DNA double helix, as opposed to the formation of a conforming PNA\*DNA.DNA triplex. Triplexes formed by strand invasion are characterised by both high thermodynamic stability and high specificity, which may appear contradictory. However, as the strand invasion reaction is kinetically controlled, the mechanism allows screening for the target sequence without trapping thermodynamically stable, but incorrectly matched, complexes. Strand invasion is doubly remarkable as PNAs must initiate recognition at target sequences that are already base-paired and then continue binding, overcoming entropic considerations which favour reformation of the parent duplex.

Strand displacement has resulted in the initiation of a new avenue for the development of sequence-specific dsDNA-binding ligands as gene therapeutic agents and as biomolecular tools within genome analyses. One such example entails PNA acting as a specific gene-activating reagent. In this approach, those sequences in dsDNA targeted by PNA may be considered to be artificial transcription promoters, positively controlled by the complementary PNA which acts as a transcription factor. RNA polymerase recognition for the P-loop in strand displacement recognition then initiates RNA transcription from the displaced strand. Interestingly, according to a contrasting observation, PNA will arrest transcription elongation of RNA bound to the template strand of the dsDNA. The high stability and sequence-specificity of PNA-DNA duplexes has been exploited in the analysis of single base pair mutations in PNA employing polymerase chain reaction (PCR),<sup>121-123</sup> where selective amplification or suppression of target sequences differing by one base pair, has been observed to occur. In another application, the use of labelled PNAs to directly map individual DNA molecules was attempted by Matsudaira *et al*,<sup>30</sup> by exploiting the strand invasion mode. The use of fluorescently labelled molecules bound to complementary sequences in the duplex DNA using 'strand invasion' was seen to enable direct observation of the position of the labelled section of DNA, thus allowing application for precise gene location as part of the mapping efforts of the Human Genome project.

#### **1.4.5 P-Loops as Artificial Transcription Promoters**

Resemblances between PNA strand displacement hybrids and transcription initiation complexes became apparent from investigations which demonstrated transcription initiation occurred at P-loops, specifically upon the looped-out strand, which acted as the template (Figure 1.10). The process was observed to be dependent upon the length of the resulting transcripts which correspond unerringly to run-off transcripts initiated at the bound PNA. Both transcription initiation and transcription arrest were seen to be distinguished by the

size of the P-loops; transcription arrest requires smaller P-loops compared with the promotion of transcription initiation, which has been observed to be more efficient with larger P-loops.

#### **1.4.6 Transcription Arrest Using of PNA Oligomers**

The principal theoretical advantage of PNA application for transcription inhibition, that's the synthesis of a protein on an mRNA template, is that transcription represents the initial step in the intermediary stage of RNA metabolism, thus providing substantial leverage for drug therapy. Transcription inhibition may proceed by targeting PNA oligomers to the coding region of the gene, for binding to the template DNA strand, thus arresting elongation with use of RNA polymerases. The potential importance of PNA oligomers lies particularly for targets that may be difficult to inhibit at the post-transcriptional level. However, sequence-specific binding is not yet possible as runs of homopurines are required. Several strategies to circumvent the requirement of purine-pyrimidine runs have been devised including purine-forming triplex structures at higher pH values.

Another obstacle associated with the application of PNAs as antisense agents *in vivo* is the associated slow rate of strand invasion under moderate salt concentrations. Salt concentrations above 50 mM are known to lower the binding rate of simple PNAs with dsDNA dramatically.<sup>30</sup> Though this results in a complex of remarkable stability, it also limits the cell uptake of the resulting complex. However, strand displacement has been observed to occur in concentrations of 500 mM NaCl in a low-salt buffer which significantly accelerates the formation of the (PNA)<sub>2</sub>.DNA complex. It is the high stability and sensitivity of triplexes to single base mismatches, and the broad range of potential DNA target sequences, that provide the basis for translation-elongation arrest using PNAs. Homopyrimidine PNA decamers have exhibited the most discriminative and efficient clamping, with formation of highly stable (PNA)<sub>2</sub>.DNA triplexes, thus causing polymerase inhibition. These properties are thus envisioned to enable PNAs to act as key components in a wide range of powerful new procedures for the study and manipulation of nucleic acids.

#### **1.4.7 Translation Arrest by PNA Oligomers**

A basic principle for the design of an antisense agent is to achieve a balance between conferring sufficient specificity for the target sequence, whilst maintaining useful physio-chemical properties to ensure cell penetration<sup>124</sup> and maintenance of adequate concentration for the desired biological activity. Minor structural differences can also have a significant effect upon the activity of DNA analogues. The majority of traditional biological targets<sup>68,125</sup> for medicinal chemists are proteins, receptors or enzymes. PNA antisense oligonucleotides have been shown to modify cellular function in studies where they exhibited control of protein production at the nucleic acid level for both transcription

and translation (Figure 1.7).<sup>126,127</sup> From the above, it may be surmised that PNA is a potent antisense mediator, where translation initiation is inhibited through the targeting of PNAs with mixed purine-pyrimidine sequences through duplex formation, whilst inhibition of translation elongation is attained via triplex-forming homopyrimidine and homopurine PNAs.

#### **1.4.8 Comparison of the Hybrid Stability Depending on PNA and DNA Backbone Structures**

One of the unique features of PNA is its ability to hybridise to an oligonucleotide irrespective of its orientation. The stability of double stranded nucleic acids, determined by both intrinsic and extrinsic factors, is a balance between ensuring sufficient integrity of the genetic code, whilst being flexible enough to allow manipulation and random exchanges of information. Bis-PNAs,<sup>114</sup> the linkage of two homopyrimidine PNA oligomers which distinguish themselves by forming (PNA)<sub>2</sub>.DNA triplexes of extremely high thermal stability, have been reported for use in hybridisation investigations. Their application was shown to demonstrate a substantial improvement in the hybridisation efficiency of PNA oligomers to ssDNA. The ability of PNA oligomers to invade the double helix of DNA was attributed in part to its neutral backbone, but predominantly to H-bonding interactions formed between the backbone of the Hoogsteen PNA strand with an oxygen of the phosphate resident upon the DNA backbone. The thermal stabilities of hybrids formed between PNA oligomers with complementary DNA, RNA and PNA oligomers of identical sequences, has been observed to display consistently higher thermal stabilities than DNA oligomers: PNA-PNA > PNA-RNA > PNA-DNA > (RNA-DNA > DNA-DNA).<sup>6,38,106</sup> The formation of mixed PNA.DNA strand displacement complexes<sup>3</sup> are observed to be strongly favoured at low ionic strength (< 50 mM NaCl) with a thermal stabilisation of 1 °C per base pair at physiological ionic strength compared with an oligomer duplex. This contention, attributed to the lack of electrostatic repulsion between the two strands, is supported by further studies where PNA.DNA and DNA.DNA duplexes were shown to possess equal thermal stability at ionic strengths above 1 M Na<sup>+</sup>; at low ionic strengths, repulsion between the polyionic DNA backbones facilitated displacement with the neutral PNA.<sup>38,106</sup> However, once formed at low ionic strengths the strand displacement complex is exceedingly stable at high ionic strengths (> 0.5 M NaCl) towards a dsDNA target.<sup>36,38</sup>

#### **1.4.9 The Targeting Ability of PNA Oligomers as a Consequence of the Neutral Backbone**

The charge residing upon the phosphodiester backbone enables successful movement of DNA through the cell membrane into the cytoplasm.<sup>129</sup> For efficient application as a powerful pharmaceutical, effective cell uptake of PNA oligomers is crucial.<sup>128,129</sup> To reach the cell interior, the cell membrane which often acts as a barrier to the entry of foreign particles and thus could prevent the entry of PNAs, must be passed. Though the



neutral backbone of PNAs improves affinity and specificity, the cell uptake<sup>124,130</sup> is adversely affected resulting in lowered mobility. This has resulted in the investigation of alternative possibilities to ascertain the delivery of PNA oligonucleotides into the cytoplasm.<sup>131-133</sup> The intrinsic structure of PNA oligomer strands may be modified to improve cell uptake.<sup>134</sup> Location of the oligonucleotide is determined by cell penetration behaviour which, in turn, is determined by the nature of the oligonucleotide modification. The accumulation of "plastic" oligonucleotides, those that distribute themselves unevenly, is hoped to exhibit beneficial results, as an accumulative affect in a localised region could prove to be potentially invaluable.

Alternative modes of entry, following systematic administration in an attempt to facilitate cell uptake for improving oligonucleotide passage through the membrane, include the use of conjugation of PNA oligonucleotides to transcellular drug delivery systems such as carrier molecules, for delivery of antisense therapeutics to the target tissue. Phagocytosis or endocytosis, an active transport method for oligonucleotides, transpiring from non-specific binding of the oligonucleotide to the membrane, involves gradual internalisation thus providing an encouraging approach towards improved penetration for PNA cell uptake. Lipophilicity may be related to the behaviour of PNAs in membranes, and is therefore of significant interest in devising drug delivery systems. PNA modification with groups specifically targeted to receptor proteins on cell membranes include lysosomes, liposomes<sup>135</sup> and protein conjugates of specific carriers.<sup>136</sup> Such groups allow cell-type-specific-directed delivery, where penetration of the membrane is followed by slow efflux of PNA; release of cell content upon cell fusion eventuates in their release within endosomal compartments, but consequates in poor bioavailability *in vivo*.

The problems associated with cell permeability need to be challenged and improved with subsequent examination of intracellular distribution. As such, an alternative strategy for circumventing the problems associated with the delivery mechanisms of PNAs was recently reported.<sup>132</sup> Studies regarding penetration of the blood brain barrier (BBB)<sup>132,133</sup> were observed to improve via use of vector-mediated delivery with a biotinylated oligomer coupled to a conjugate of avidin. Whereas the BBB transport of PNAs is negligible in the absence of a transcellular delivery system, as a possible solution for improving PNA uptake, biotinylated PNAs were targeted to the brain and successfully transported through the BBB *in vivo* via coupling to potential therapeutic drugs of a peptide drug delivery vector. The use of such delivery systems is now hoped to allow for *in vivo* testing of the pharmacological activity of PNAs in animal models. More recent studies have also shown technetium-99-labelled PNAs hybridise *in vivo* to beads conjugated with the complementary strand, indicating the stability of PNAs *in vivo*.<sup>131,137,138</sup> Furthermore, Tyler *et al*<sup>129</sup> have recently presented the first evidence of successful intracellular delivery of PNAs *in vivo*, with subsequent inhibition of protein expression. Hence, the key to

improving the cell uptake of PNAs may actually lie in the fact that when left unmodified, these small neutral molecules are able to better surpass the cell membrane intact.<sup>139</sup> Attachment of other molecules to the PNA oligomer may allow for visualisation but with an associated hindering of the transport of PNAs. Hence, these studies provide encouragement for improving the cell uptake of PNA, and thus making them more viable as therapeutic agents.

## **1.5 Evaluation of DNA Backbone Modifications**

### **1.5.1 The Scope for Other Backbone Modifications in DNA**

An increased appreciation for the characteristics and requirements of modified oligonucleotides<sup>72,76</sup> to display optimum biological activity with simultaneous elimination of current problems posed with use of phosphodiester-linked DNA oligomers, has resulted in numerous approaches for improving significant characteristics of rationally-modified oligonucleotide analogues.<sup>14,22,140-143</sup> Modified phosphodiester oligonucleotides were initially synthesised to improve nuclease resistance, to increase accessibility across cellular membranes, and to encompass force repulsion upon hybridisation with the complementary target, in order to attain stronger association. Minimising the number of chiral centres, preserving isostericity with the natural phosphodiester backbone, and retaining interatomic distances in the synthetic oligonucleotide to allow pre-determined hybridisation with the target nucleic acid, necessitate consideration. Consequently, a number of structural modifications to improve cell uptake, degradation, solubility, partition coefficients and intrinsic physiochemical properties have been introduced, adding to modified PNA backbones.<sup>144-146</sup>

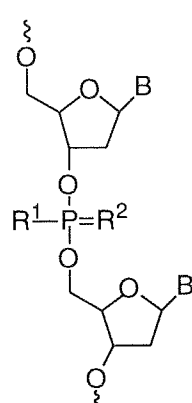
### **1.5.2 Ionic and Non-Ionic Backbones**

Much literature exists<sup>76</sup> reporting backbone modifications such as anionic S-oligonucleotides (phosphorothioate),<sup>147</sup> neutral methylphosphonates (MP)<sup>76</sup> (Table 1.1), and positively charged guanidinium oligomers.<sup>115</sup> Investigations for alternate DNA analogues<sup>14,50,148</sup> include the synthesis of oligonucleotides and polynucleotides containing unnatural bases or unnatural sugars, those consisting of nucleosides connected via linkages other than phosphodiester bonds, and those with alternative base-to-backbone linkers compared to conventional sugar phosphate moieties.<sup>140</sup>

Included in the first generation of backbone modified oligonucleotides are the methylphosphonates, phosphoramidates, phosphorothioates<sup>149</sup> and phosphotriesters, all of which retain the phosphorus atom (Table 1.1). Despite the observed increase in resistance toward nucleases, an associated lowered binding affinity was exhibited by this class of modified oligomers, in particular the S- and MP-oligonucleotides. This was attributed to the presence of a chiral centre, thus resulting in the production of several non-resolvable

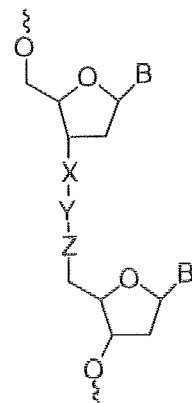
diastereoisomers; only a proportion possessed the correct geometry for hybrid formation with the complementary target. The recent synthesis of achiral phosphorous analogues has aided towards eliminating this problem, though clinically there appears to be no great benefit in the use of stereochemically pure phosphorothioates over random S/R analogues.<sup>140</sup>

**Table 1.1.** Structure of phosphorus-containing backbone modifications

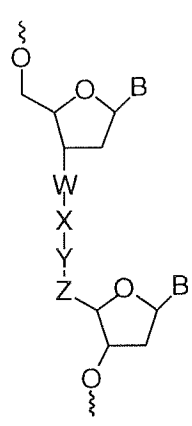
	<b>R<sup>1</sup></b>	<b>R<sup>2</sup></b>
	S <sup>-</sup>	O
	CH <sub>3</sub>	O
	NHR	O
	OR	O
	S <sup>-</sup>	S
	O <sup>-</sup>	O <sup>-</sup>

The second generation of modified backbones<sup>150,151</sup> involve introduction of a group with a non-phosphorus atom<sup>152</sup> (Tables 1.2 and 1.3). Nuclease resistance and the binding affinity of RNA for its complement were both observed to improve. This category is comprised of aminomethyl, hydroxymethyl phosphonates, phosphoramidates, methyl phosphorothioate, amides<sup>153-155</sup>, carbamates, sulfonate, thioformacetal, and ether amongst other backbones.<sup>156</sup> *N*-methylhydroxylamine and thioformate backbones however, have failed to show any improvement for binding affinity or nuclease resistance.

**Table 1.2.** Structure of phosphorous-free backbone modifications

	<b>X</b>	<b>Y</b>	<b>Z</b>
	O	SiR <sub>2</sub>	O
	CH <sub>2</sub>	S	CH <sub>2</sub>
	O	CH <sub>2</sub>	O
	CH <sub>2</sub>	NCH <sub>3</sub>	O
	S	CH <sub>2</sub>	O

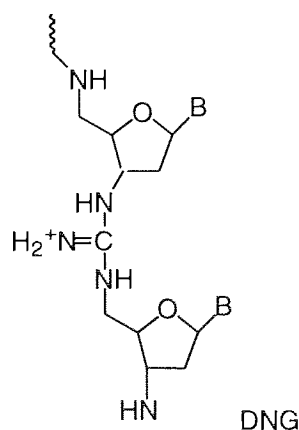
**Table 1.3.** Amide, urea, carbamate, and alkyl chain backbone modifications.



W	X	Y	Z
NR	CO	CH <sub>2</sub>	CH <sub>2</sub>
CH <sub>2</sub>	CH <sub>2</sub>	NH	CO
CH <sub>2</sub>	CO	NR	CH <sub>2</sub>
CH <sub>2</sub>	NH	CO	CH <sub>2</sub>
CO	NH	CH <sub>2</sub>	CH <sub>2</sub>
NR <sup>2</sup>	CO	NR <sup>1</sup>	CH <sub>2</sub>
O	CO	NR	CH <sub>2</sub>
NR	CO	O	CH <sub>2</sub>
CH <sub>2</sub>	CO	CH <sub>2</sub>	CH <sub>2</sub>
CH <sub>2</sub>	CH <sub>2</sub>	CH <sub>2</sub>	CH <sub>2</sub>

Backbones composed of amides are expected to display increased stability under physiological conditions and penetration through cell membranes attributable to overall charge reduction of the oligonucleotides, though this is in complete contradiction with studies reported with the use of PNAs. Those properties that have been observed to have improved include: biological properties, action upon RNaseH substrate, interaction with proteins, enzymes and cellular factors, cellular pharmacokinetic properties, clearance of metabolites, site-specific delivery, longer plasma half-life, and partial thromboplastin time ( $\alpha$ PTT).

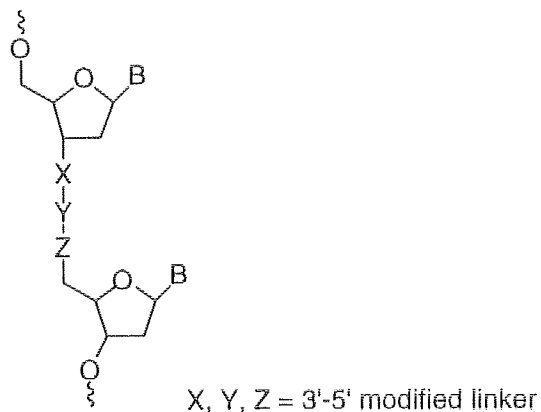
The effect of backbone charge on duplexation is illustrated in the binding of polycationpentamic thymidyl deoxyribonucleic guanidinium (DNG) (Figure 1.11), an alternative backbone developed by Dempcy *et al.*<sup>115</sup>



**Figure 1.11.** The guanidinium (DNG) backbone.

Replacement of guadinium linkages<sup>157</sup> '[-NHC(=NH<sub>2</sub><sup>+</sup>)NH]' for phosphodiester linkages were shown to improve the stability of both duplex and triplex structures upon hybridisation to negatively charged ssDNA, as a consequence of electrostatic attractions.

**Table 1.4.** The effect of selected modified backbones on T<sub>m</sub> and nuclease resistance.



Backbone Moieity	3'-5' Linker			T <sub>m</sub>	Nuclease Resistance
	X	Y	Z		
Carbamate	O	CO	NH	-/+	++
Carbonate	O	CO	O	+	+
Formacetal	O	CH <sub>2</sub>	O	x	++
Methyleneimino	CH <sub>2</sub>	NH	O	++	++
Dimethylhydrazo	CH <sub>2</sub>	NCH <sub>3</sub>	CH <sub>3</sub>	++	++
Methylimino	CH <sub>2</sub>	O	NCH <sub>3</sub>	+	++
Methylimino	CH <sub>2</sub>	NCH <sub>3</sub>	O	++	++
Oxime	CH <sub>2</sub>	=N	O	-/+	++
Phosphate	O	PO <sub>2</sub>	O	++	-
Silyl	O	SiR	O	+	x
Sulphonate	O	SO <sub>2</sub>	CH <sub>2</sub>	x	++
Sulphonamide	O	SO <sub>2</sub>	NH	++	++
Sulphur	CH <sub>2</sub>	S(O)N	CH <sub>2</sub>	+++	x
Thioformacetal	O	CH <sub>2</sub>	S	-	++

+ = increase; - = decrease; x = not measured or no significant effect.

Mesmaeker *et al* have recently reported the investigation of eighteen types of backbone replacements.<sup>158</sup> The thermodynamic stability of the duplexes formed between the modified backbone and RNA complements was investigated to improve the poor penetration of current phosphodiester backbones. It was found that an increased resistance to nucleases occurred concurrent with a decrease in affinity for the RNA complement, and that extension of the modified backbones destabilised resulting duplexes; the extent of destabilisation was found to be dependent upon the point of insertion of the extension.

Despite the numerous investigations regarding backbone modification, which have resulted in a whole array of modified backbones, no single backbone<sup>148</sup> has yet been shown to simultaneously increase stability, binding affinity, and specificity.

### 1.5.3 Potential Modifications to the PNA Backbone

Significant factors requiring increased consideration for potential backbone modification include conformational flexibility, charge,<sup>143</sup> hydration, chemical stability, and synthesis, which must be readily adaptable to modern solid phase synthetic methods. Furthermore, though negatively charged backbones may improve specific properties such as cell uptake,<sup>129</sup> they may also be inherently unfavourable for other properties<sup>159</sup> due to repulsive forces; consideration is thus given to neutral or positively charged alternatives, as reflected in the synthesis of the PNA,<sup>160</sup> aminoalkyl, and carbamate backbones, for example. The internucleotide linkage of the modified backbone also needs to be strongly hydrated to present a hydrophilic face to aqueous solutions as provided with the use of negatively charged backbones, and stable with respect to general acid and base catalysis; a property not exhibited by current phosphodiester linkages.

As with DNA, a variety of options exist for modifying the PNA unit<sup>85,161,162</sup> (Figure 1.12) in order to extend its future applications.

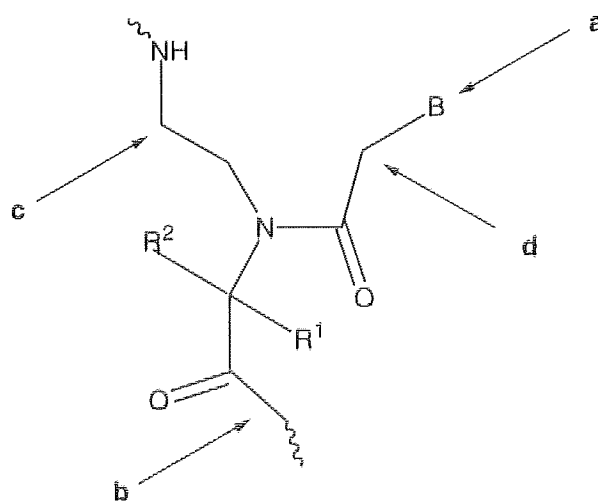


Figure 1.12. Potential sites for modification on the PNA backbone.

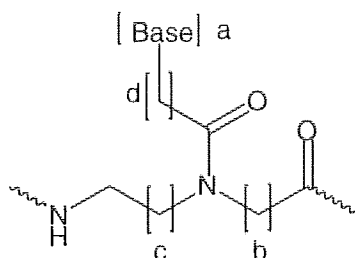
Four sections of the monomer backbone lend themselves to potential modification:

- a) The heterocyclic bases
- b) The backbone amino acid ( $R^1$  and  $R^2$ ) moieties
- c) The aminoalkyl section
- d) The backbone-to-base linker

Attempts to improve the backbone structure and the introduction of asymmetry into achiral PNA may influence its direction of binding with complementary DNA and RNA. The latter property may be impaired by i) covalent linking of a chiral acid, oligonucleotide or polypeptide at the PNA terminus;<sup>104</sup> ii) constructing a PNA backbone with an amino acid carrying a base on the side-chain; iii) incorporation of various amino acids within the backbone. Though certain alterations are permitted without severe penalties in DNA/RNA-PNA hybrid stability, extensive modification consequates in the introduction of severe implications upon PNA's structural,<sup>163</sup> chemical and physical properties. An observed lowering in the stability of backbone modified PNA-DNA hybrids, compared with unmodified hybrids, was ascribed to geometric constraints within PNA, and to a loss in entropy involving hybrid formation.<sup>161</sup> Differences between mismatched, and matched unmodified PNA<sub>2</sub>.DNA hybrids were observed to be comparable to those for corresponding modified hybrids.

Base modification,<sup>164</sup> with which this work is primarily concerned and is detailed in chapter 2, was considered to be the most significant aspect towards increasing the stability and hybridisation properties of the peptidic backbone, whilst keeping all other factors constant. Modification of a heterocyclic base should ensure H-bonding and the base stacking interactions remain undisrupted. Base modification is particularly attractive as it avoids interference with RNaseH activity of the targeted RNA. Purine derivatives, such as isoC<sup>165</sup> and D, introduce a third hydrogen bond thus contributing towards the stabilisation of duplex formation, as shown by alternating deaza-adenine and thymine residues. Inosine occurs naturally within the wobble position of various anticodons of tRNAs, and has been shown to hybridise non-specifically, with subsequent application in oligonucleotide mixtures for locating sequences that are ambiguous as a result of genetic code degeneracy.<sup>165</sup> Halide substitution and hydrophobic modifications, for example in 5-methyl- and 5-propyne pyrimidines, have been shown to increase  $T_m$ . The design aspects of base modification, and their anticipated applications towards extending molecular recognition properties in PNA oligomers, are evaluated in chapter 2.

Research conducted by Fuji *et al*,<sup>166</sup> and Hyrup *et al*, concerning elongation in the PNA unit (Figure 1.12) via incorporation of an additional methylene group at (b), (c) and (d), furnished the modified *N*-(thymine-1-ylacetyl)- $\beta$ -aminoalanine (b = 2, c = 1, d = 1) and *N*-(cytosine-1-ylacetyl)- $\beta$ -aminoalanine moieties (b = 2, c = 1, d = 1), *N*-(3-aminopropyl)glycine (b = 1, c = 2, d = 1), and the propanoic acid linker (b = 1, c = 1, d = 2), respectively (Figure 1.13).



**Figure 1.13.** Modifications to the repeating unit in the original PNA structure.

These modifications enabled the comparison of stability and specificity properties with the original PNA structures. Modification at (b), via incorporation of a  $\beta$ T (( $\beta$ -alanine)thymine) or an apgT ((aminopropyl)glycine thymine) monomer in a 10-mer homothymine PNA, was shown to reduce the  $T_m$  of hybrids formed with ssDNA, by 13 and 11°C respectively. Alternatively, the incorporation of ethylene and propylene carbonyl linkers (d), which results in increased distances between the base and the peptidic backbone, consequated in a significantly lowered affinity for complementary ssDNA targets, reflected in a lowering in  $T_m$  (~18°C). These observations implicated that extended monomers with modifications at (b) and (d) experienced a dramatic reduction in their affinity for ssDNA, but retained their specificity. However, analysis of  $T_m$  values also revealed that those hybrids with extended backbone showed greater instability compared to those with an extended linker.

The formation of unstable complexes emphasises the importance of retaining the correct interbase distances in the backbone to preserve base pairing and to allow formation of a regular backbone geometry. The extension of backbone-linker in DNA hybrids with PNA oligomers also imposes adverse steric and structural constraints, arising from structural incompatibilities between (aminoethyl)glycine and methylene extended backbones.<sup>323</sup> In summary, it is far more costly to accommodate an additional methylene group within the backbone, compared with the same modification for backbone-linker. Inspection of Table 1.5 allows one to observe that backbone-to-base extension is less detrimental to the stability of PNA:DNA complexes than  $\beta$ T or apgT unit linker extension. However, it should also be realised that the optimal backbone for a genetic material is not necessarily that which yields the most stable duplex; too great a stability would infer non-availability of genetic information, essential for transcription and replication processes.<sup>167</sup>



**Table 1.5.** The effect of PNA backbone modifications on  $T_m$ .

Modifying Unit	b	c	d	$T_m$ ( $^{\circ}\text{C}$ )
Ethylglycine	$\text{CH}_2$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	0
Propinoyl	$\text{CH}_2$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2\text{CH}_3$	21
Propyl	$\text{CH}_2$	$\text{CH}_2\text{CH}_2\text{CH}_2$	$\text{CH}_2$	8
$\beta$ -Alanine	$\text{CH}_2\text{CH}_2$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	10
Retro-inverse	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	$\text{CH}_2$	8.5
D-Alanine	$\text{CH}(\text{CH}_3)$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	1
L-Alanine	$\text{CH}(\text{CH}_3)$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	4.5
D-Glutamic acid	$\text{CH}(\text{CH}_2\text{CH}_2\text{CO}_2\text{H})$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	2.3
D-Serine	$\text{CH}(\text{CH}_2\text{OH})$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	0.6
L-Serine	$\text{CH}(\text{CH}_2\text{OH})$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	2.3
D-Lysine	$\text{CH}(\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2)$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	1
L-Lysine	$\text{CH}(\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2)$	$\text{CH}_2\text{CH}_2$	$\text{CH}_2$	0

#### 1.5.4 The Scope for PNA Backbone Modifications: Hybridisation, Stability and Targeting to DNA

The introduction of functionality and chirality<sup>96,168</sup> in the PNA backbone may be assessed via construction of modified PNA monomers<sup>169</sup> composed of alternative amino acids to glycine.<sup>170</sup> These modifications provide opportunities to tailor the hybridisation behaviour of PNA via conformational freedom for increasing solubility properties, for finely tuning the stability of hybrids with nucleic acids, and for providing convenient attachment points for various biologically active ligands. Such investigations have showed substitution of glycine units in the PNA backbone with chiral amino acids result in a slight loss in binding affinity of the modified PNA backbone to DNA and RNA. Modification of glycine to alanine amino acids has been shown to introduce loss of chiral integrity, ensuing in a large number of stereoisomers of the oligomer, where D-amino acids<sup>171</sup> were observed to be more easily accommodated in the backbone of a PNA-DNA duplex compared with L-amino acids.<sup>131</sup> The solubility of PNAs were observed to increase without compromising the stability and selectivity of hybrids with DNA. Finally, insertion of positively charged side-chains produced higher  $T_m$  values compared with neutral side chains (which are better than negatively charge side chains), particularly between two PNA strands which led to

rapid, efficient hybridisation with DNA compared with polyethylene glycol-based linkers. Effects imposed by the reversal of the orientation between two nucleobases to produce a reversed amide were investigated by Lagriffoul *et al.* An appreciable difference in the thermal stability between the preferred parallel and antiparallel triplexes was observed, where the thermal stability of the former was observed to decrease slightly, and the thermal stability of the latter to slightly increase. As the binding affinity to complementary DNA was preserved, these investigations confirmed the necessity to retain PNA recognition properties.

### 1.5.5 Chimeric PNAs

PNA-DNA chimeras<sup>172</sup> are potentially interesting molecules as they combine advantageous properties<sup>163</sup> of PNA and DNA (Figure 1.14). Whilst the PNA section provides high binding affinity and specificity, the DNA section is able to interact with DNA recognising enzymes, such as RNase H and DNA polymerases. Modelling studies were undertaken to investigate the need for a linker to connect the two oligomers in a mode that would allow stacking between the neighbouring bases in the PNA and DNA sections of the chimera.<sup>172</sup>

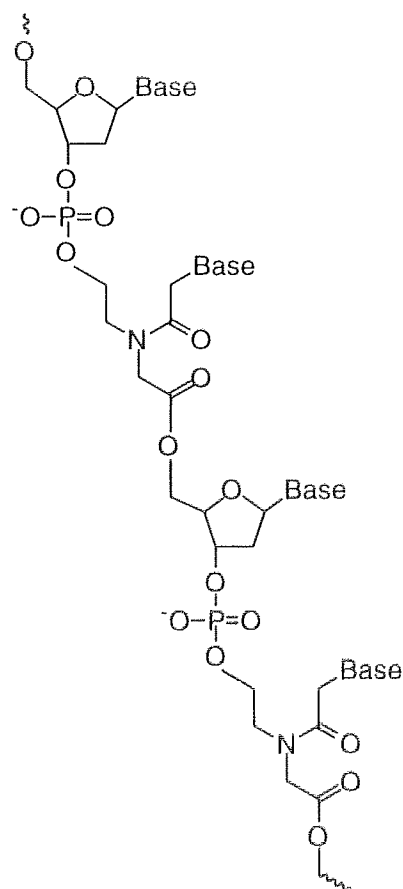


Figure 1.14. A chimeric PNA.

PNA oligomers, terminating in a hydroxyl group instead of the primary amine, were anticipated to constitute an appropriate linker in terms of base stacking and the number of bonds between the nucleobases and orientation, where both parts of the chimera would bind in the preferred antiparallel orientation to DNA.

From analysis of the aforementioned literature, it may be concluded only certain alterations or deviations from the original (aminoethyl)glycine backbone may be permitted without severe penalties to DNA/RNA-PNA hybrid stability. As an overall conclusion, at this stage of research regarding the modification of the original PNA backbone, the basic structure of the PNA backbone cannot be beneficially altered by extension from ethyl to propyl, glycine to  $\beta$ -alanine, or acetyl to propionyl, nor by reducing the nucleobase linker amide, nor reversing the amide linkage within the backbone, without adversely affecting stability, specificity or affinity. However, much freedom seems to arise in placing (functional) substituents on the backbone, as exemplified by exchanging the non-functional glycine for other non-natural amino acids;<sup>131</sup> the new type of substituent and also the stereochemistry at the newly created chiral centre however, exert different effects on the PNA hybridisation properties. Inclusion of cyclic substituents, as exemplified by the 'cyclohexyl substitution' at the amino ethyl linker is even possible provided the 'correct' stereoisomer is chosen. Modifications of the original *N*-acetyl-*N*-(2-aminoethyl)glycine PNA backbone has not yet yielded derivatives that hybridise with increased affinity to DNA or RNA. Therefore, this backbone seems to exhibit a balance between favourable constraints and flexibility required of an acyclic backbone.

## 1.6 The Future of PNAs

### 1.6.1 Therapeutic and Diagnostic Applications of PNAs

*In vitro* studies imply that PNAs possess the potential for development into efficient and specific gene-activating reagents<sup>173</sup> with clear advantages over conventional phosphodiester and phosphorothioate oligonucleotides. However, in terms of therapeutic applications *in vivo*,<sup>17</sup> several issues need to be addressed. In order for PNA to bind to dsDNA, a polypurine target<sup>39</sup> must be present in the gene of interest, thus limiting the number of possible targets within a gene. However, this is partly overcome with strand invasion, which is unaffected by this factor. Secondly, the formation of strand displacement complexes occurs very slowly at physiological salt concentrations. However, improvements of the binding rate have been undertaken via construction of bis-PNAs, but an approach to enable PNAs to bind at any physiological ionic strength still needs to be devised, to further the applications of PNA in this field.

### 1.6.2 Mechanistic Applications of PNAs

Numerous mechanistic applications<sup>174</sup> of PNA oligomers are likely. PNAs may be developed into biomolecular tools<sup>12,122,175</sup> and more importantly into antisense (targeting mRNA) or antigene (dsDNA) drugs.<sup>44</sup> A further interesting application of PNA stems from the information it may be able to provide regarding the origin of genetic information. More excitingly, PNA can function as a template for the chemical synthesis of RNA and vice versa, thus illustrating, in principle, information transfer between PNA and RNA is possible. The capability of PNAs to act as sequence-specific inhibitors against restriction enzyme cleavage,<sup>91</sup> as electron microscopy markers of homopurine tracts, as gene activating agents via artificial transcription factors, as transcription elongation inhibitors, and, in co-operation with single-strand-specific nucleases, as 'synthetic restriction enzymes', has resulted in wide application of PNAs for the isolation of transcriptionally active DNA,<sup>30</sup> nucleic acid purification,<sup>29</sup> human telomerase inhibition<sup>130</sup> and visualisation, PCR amplification enhancement,<sup>121,122,176</sup> antisense inhibition, and mutation screening,<sup>25</sup> amongst others.

PNAs, in conjunction with nuclease  $S_1$ , have been used as artificial restriction enzymes.<sup>91</sup> In situations where the PNA site of dsDNA is overlapping or proximal to a restriction enzyme cleavage site, binding of PNA to that site should prevent cleavage by the enzyme. Complementary to this technique is that where dsDNA is cleaved by the ss-specific nuclease  $S_1$ . If two adjacent PNA binding sites are present in the dsDNA, double strand cleavage may be accomplished with  $S_1$ , subsequent to hybridisation of PNA to these sites. It is hoped that this technique will allow the design of 'restriction enzymes' and serve for chromosome mapping and gene cloning.

The ability of PNAs to act as probes<sup>177</sup> labelled with biotin or fluorescein, or with reporter enzymes, permits application of PNAs as powerful tools in cloning and blotting procedures, arising from high thermal stability of hybrids, greater specificity of interaction reliant upon the reduced hybridisation times, increased range of labelling options and finally, from increased latitude in the selection of stringency conditions.

The development of oligonucleotides-based inhibition of telomerase, the enzyme responsible for maintaining the length of telomeres (chromosome ends) during replication, is especially attractive relative to other approaches for rational inhibitor design. In contrast to somatic cells, the majority of cancer cells have been found to possess telomerase activity.<sup>178</sup> Telomerase carries an RNA which function as a template for extension of the chromosome ends. PNAs complementary to a sequence in this RNA template have been shown to work as very efficient, sequence-specific inhibitors of telomerase activity both *in vitro* and in permeabilised cells. A comparison of the ability of PNA and phosphorothioate oligonucleotides respectively, to inhibit telomerase activity, showed that PNA possessed

clear advantages, including an essentially lowered concentration of PNA necessary for specific inhibition. This suggests that PNAs are useful lead compounds for the development of inhibitions that will function *in vivo* and hence, such compounds may be utilised to probe the intracellular function of telomerase and adapted to test the hypothesis that inhibition of telomerase can lead to suppression of tumour growth, thus acting as a potential target for cancer therapy.

### 1.6.3 Evaluation of the Future for PNAs

The reported literature on PNAs<sup>180</sup> has a bearing on many aspects of medicinal chemistry,<sup>181</sup> biology, biochemistry, and pharmacology, ranging from basal molecular recognition, self-assembly and chiral induction aspects, to molecular biology tools and gene therapeutic drugs, to our understanding of the structure and function of Nature's genetic material, DNA, its natural predecessors and origin. Hence, PNAs should also be regarded as structural and self-recognising systems with potential utility within both the traditional oligonucleotide field,<sup>17</sup> as well as in other areas of science.<sup>48</sup> By virtue of its ability to sequence-specifically inhibit transcription, the most exciting application for PNAs is its potential for development into an antigene drug,<sup>11,79</sup> though it must also be remembered that there is little advantage in gaining increased specificity with the new agents at the expense of poor efficacy.

In summary, the DNA and RNA binding properties of PNA, its chemical and biological stability, and its synthetic versatility, has furnished PNA as a very attractive molecule for several purposes. These include the development of antisense<sup>182,183</sup> and antigene therapeutic agents as well as biomolecular tools in diagnostics and molecular biology. For instance, protein synthesis (translation) can be sequence-selectively and efficiently blocked via hybridisation of PNA to mRNA. Analogously, PNA bound to the promoter, or within the coding region of a gene, can inhibit transcription by blocking the DNA access or the progression of the RNA polymerase. Continuing efforts in the field of PNA synthesis may yield novel base analogues, possibly expanding the recognition repertoire for triplex as well as duplex formation, which is the main aim of this work. PNAs are also envisioned for use in combinatorial libraries, and in processes analogous to the selection of nucleic acid aptamer molecules. Diagnostic assays for employing PNA for sequence-specific<sup>98</sup> recognition of DNA already exist, possibly leading to new, commercially available, diagnostic and analytical tools. Likewise, several molecular biology methods based on PNA technology have been developed. These include sequence-specific cleavage of plasmid DNA by PNA targeting of nuclease S1. This is a rare cleavage strategy for genome mapping using the restriction enzyme principle modulation of the PCR technique to detect single base mutations as well as a PNA-based nucleic acid capture technique.<sup>121,122</sup>

Perhaps the most exciting use of PNA, that within antisense and antigene therapy, is linked to the challenge of developing efficient and safe methods for PNA cell permeation. To date, attempts to deliver PNA into living cells have been predominantly unsuccessful, though recent reports provide encouraging news, and with many possible strategies to explore, *in vivo* use of PNA may one day become reality. To be more than just a laboratory curiosity, PNAs must offer the potential for significant improvement over existing protocols, or at least permit the development of new procedures that would not otherwise have been possible. The rapid strides that are being made in the PNA field augur well for the future of the antisense oligonucleotide-based therapeutics. Ongoing research should take PNAs far into the future; indeed, it must not be forgotten that PNAs were only first described in 1991.<sup>4</sup>

## CHAPTER 2

### The Design of Base-Modified PNA Oligomers

#### 2.1 Effects on Duplex and Triplex Stability via Base Modification

The present work focuses upon the development of base-modified PNA oligomers incorporating novel bases for targeting to complementary DNA bases, with potentially enhanced duplex<sup>24</sup> and triplex<sup>102,184</sup> stability, and specificity compared to natural bases.

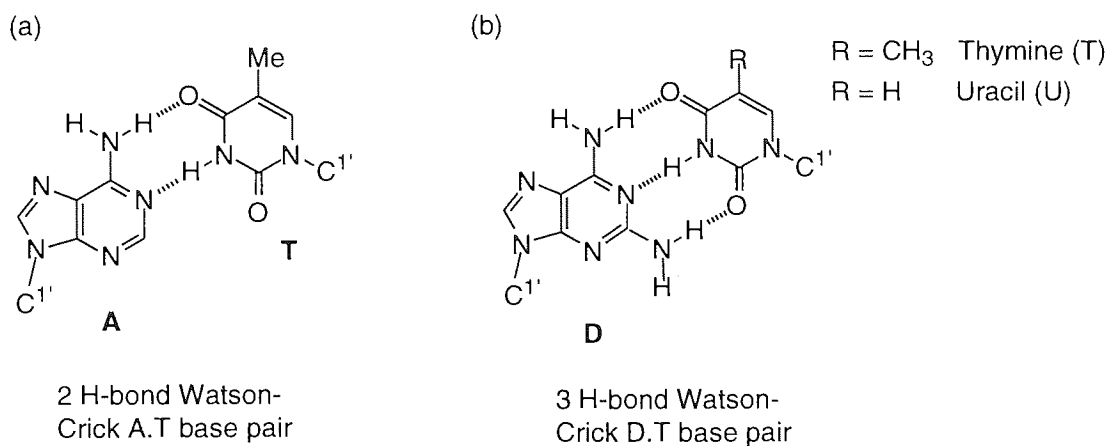
##### 2.1.1 Base-Modified Oligomers Incorporating Diaminopurine (D), and Isoguanine (isoG), Xanthine (X), and Hypoxanthine (H) for Duplex and Triplex Studies

The first objective of this work focused upon increasing the stability of duplex-forming oligomers via incorporation of the novel base 2,6-diaminopurine (D),<sup>185</sup> in place of the base adenine (A). This substitution was envisaged to increase the stability of PNA oligomers in duplexation studies<sup>186</sup> attributable to the additional C2 amino group in D.<sup>187</sup> The second focus of this research was to circumvent pH limitations arising from hybrid formation of PNA with G.C pairs by using isoGuanine (isoG); a pH-independent replacement for C<sup>+</sup>. Incorporation of xanthine (X) and hypoxanthine (H) would allow formation of isomorphous base triplexes with A.T pairs. These design aspects were anticipated to present new perspectives regarding improved molecular recognition properties of PNAs for DNA targets, whilst alleviating the limitations imposed by use of the conventional DNA bases A, T, C, and G.<sup>70,190-194</sup>

##### 2.1.2 Hybridisation Properties of 2,6-Diaminopurine (D) Compared to Adenine (A)

2,6-Diaminopurine is a naturally occurring nucleobase<sup>185</sup> structurally similar to adenine, but possessing an additional amino group at the C2 site.<sup>195</sup> The replacement of adenine (Figure 2.1 (a)) with 2,6-diaminopurine residues (Figure 2.1 (b)) has been shown to stabilise DNA duplexes; Chollet *et al*<sup>196</sup> reported an observed increase in thermal stability of the DNA duplex by 0 to 2 °C per D.T DNA base pair compared with that of A.T base pairs. Studies with similar replacements of A for D bases, have also shown sequence-specificity to improve in particular cases.<sup>197</sup>

It was envisaged an approach entailing the targeting of a polyD-containing PNA oligomer strand to an A.T-paired DNA duplex would displace A via strand displacement for subsequent formation of D.T (PNA.DNA) duplexes. The presence of the additional C2 NH donor on D was expected to confer increased stability to the D.T base pair compared with that exhibited by the A.T base pair.



**Figure 2.1.** Comparison of Watson-Crick pairing between (a) A.T and (b) D.T (and D.U) base pairs in DNA oligomers.

It is with regard to the formation of a three H-bond motif D.T (PNA.DNA), compared with the less stable two H-bond motif A.T (DNA.DNA), that the efficiency of strand displacement, the affinity, and the stability of D-containing PNA oligomers was expected to improve. Direct comparison of the stability of the D.T base pair in PNA oligomers with the known effects displayed by A.T base pairs incorporated in PNA oligomers, also pertain interest.

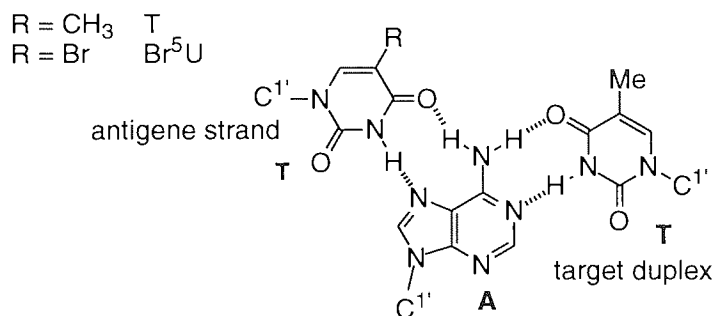
Following the completion of our experimental work, and during the course of this thesis write-up, Nielsen *et al* reported on PNA oligomers containing D.<sup>197</sup> The presence of D.T base pairs in PNA oligomers was shown to increase the  $T_m$  of decameric and pentameric DNA, RNA and PNA duplex strands. An observed increase in stability of 2.5 to 6.5 °C per D.T base pair compared with PNA A.T base pairs, was found; an increase slightly higher than that observed with comparable DNA oligomers.<sup>196</sup> The increased sequence-specificity following the introduction of a D residue was attributed to its characteristic "perfect three H-bond" property compared with the two H-bond property displayed by A. Both T and U possess the ability to mimic A, and thus base pair with D, as depicted in Figure 2.1. Hence, these studies confirmed the anticipated, increased contribution D makes towards increasing duplex stability.

## 2.2 pH Independent Triplex Formation via Base Modification

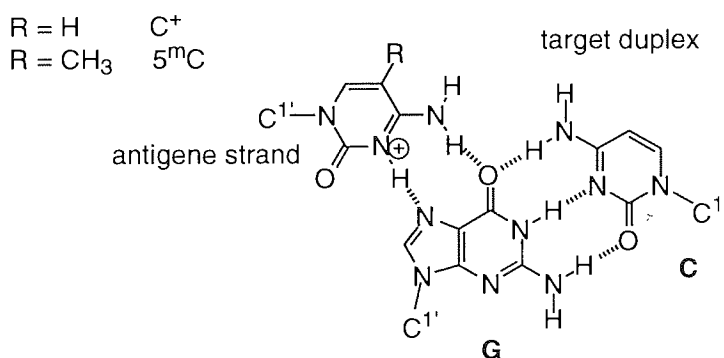
### 2.2.1 Why are Base-Modified PNA Oligomers for Triplex Formation of Interest?

Studies of sequence-specific triplex formation<sup>69,70,197</sup> by synthetic TFOs<sup>109</sup> are of cardinal interest, where sequence-selective recognition of targeted duplex DNA requires both affinity<sup>198</sup> for the target base and discrimination against the remaining three bases.





**Figure 2.2.** Targeting of an A.T DNA base pair by (a) T and (b)  $\text{Br}^5\text{U}$  in the third antigene strand.



**Figure 2.3.** Targeting of a G.C DNA base pair using  $\text{C}^+$  or  $5^m\text{C}$  in the third antigene strand.

Triplex stability may be rationalised by single specific H-bonding<sup>39</sup> contacts where  $\text{A}^*\text{A.T}^{199}$ ,  $\text{T}^*\text{A.T}$ ,<sup>51</sup> and  $\text{G}^*\text{G.C}$  each possess two H-bonding contacts (Section 1.3.1, Figure 1.5). However, the recognition of T by other base pairs in a TFO is not specific. This is as a consequence of its preferred recognition for A.T base pairs compared with C.G base pairs as observed in studies conducted by Faucon *et al.*<sup>61</sup> This non-degenerate recognition suggests a novel base with increased affinity that specifically recognises C.G base pairs is required, particularly if sequences containing more than one C.G base pair are to be targeted. The most promising way to reconcile affinity and selectivity is to utilise extended heterocyclic systems<sup>164</sup> with an ability to simultaneously utilise all major groove H-bonding sites of the Watson-Crick base pair for molecular recognition, thus increasing triplex stability.<sup>55</sup> The protonation of N3 sites on cytosine ( $\text{C}^+$ ) residues<sup>40,200</sup> in TFOs permits H-bonding with G.C base pairs in targeted DNA duplexes, enabling  $\text{C}^+*\text{G.C}$  triplet formation, as depicted in Figure 2.3,<sup>70,201</sup> where the  $\text{C1}'$  is the sugar anomeric carbon.

An absence of protonation results in a high degree of triplex destabilisation as the formation of only one, instead of two, viable Hoogsteen hydrogen bonds is possible. The

pH of an isolated C nucleoside is 4.3, but has been observed to increase to pH 6.7 in oligonucleotides when paired with complementary bases, presumably due to the presence of the polyanionic environment which encourages electrostatic attraction between positively charged C<sup>+</sup> residues on the TFO with the negatively charged backbone of the targeted duplex.<sup>200</sup> Despite the apparent rise in the pK<sub>a</sub> value, the necessity to protonate C culminates in a considerable pH dependence with adverse consequences upon the structural stability of the resulting triplets.<sup>201</sup> This imposes constraints upon the sequence composition and environment of potential triplexes, significantly limiting the application of C-containing TFOs *in vivo*, where intracellular pH is regulated at approximately 7.3.<sup>60,66,69</sup>

T<sup>\*</sup>A.T and C<sup>+</sup>\*G.C are classified as isomorphous triplets as their N-C1' bonds are superimposable. Non-isomorphism results in considerable triplex destabilisation for random purine sequences. Povsnic *et al*<sup>70</sup> investigated Br<sup>5</sup>U for targeting DNA A.T base pairs in substitution of T on the third, antigene strand (Figure 2.2). Though the substitution of T with Br<sup>5</sup>U was shown to increase binding affinity, and confer a large increase in duplex DNA cleavage efficiency over an extended pH range, it did not change the pH profile significantly. It was postulated the subsequent enhanced hybridisation was attributable to enhanced hydrophobicity, to base stacking,<sup>202</sup> and to the electronic complementarity of the Hoogsteen base pair.<sup>70</sup> These include the electron-withdrawing bromo substituent which increased the acidity at the superior N3-H donor site, and decreased the electron-donating properties of the poorer carbonyl H-bond acceptor.<sup>70</sup>

### **2.2.2 Design of Base-Modified PNA Oligomers for pH-Independent Triplex Formation**

As shown by Dervan and Povsnic,<sup>64,70,203</sup> TFOs containing C<sup>+</sup> dissociate in solutions of increasing pH,<sup>188</sup> attributable to the energy penalties associated with the loss of the Hoogsteen N3-H-O6 H-bond. Electrostatic repulsion arises between contiguous C<sup>+</sup> pyrimidine tracts in the TFO, thus causing distortion of the third antigene strand. This work, and recent literature reports,<sup>68</sup> emphasise the importance of studying the design aspects of sequence composition upon the stability of duplex and triplex structures.<sup>188,204</sup> Recent efforts directed towards the synthesis of a series of TFOs composed of non-natural bases,<sup>205-207</sup> have expanded the repertoire of sites which may be targeted by TFOs. Several approaches towards the design of pH-independent bases for stable triplet formation<sup>68</sup> have been undertaken, and are briefly surveyed.

### **2.2.3 Rationale for the Choice of IsoG, X and H for Incorporation into PNA Oligomers**

The purine bases X, H, and isoG (Figures 2.14, and 2.12 respectively) potentially mimic an N3-protonated C (Figure 2.3). Formation of two Hoogsteen H-bonds with G in the

target Watson-Crick paired strand with elimination for the need of protonation in acidic media should allow for successful triplex formation. pH-dependency limits the viability of triplexes formed from C<sup>+</sup> and 5<sup>m</sup>C;<sup>188</sup> the use of alternative purines should thus increase the scope of the molecular recognition properties of oligomers containing these modified nucleobases. Novel H-bonding schemes prompt incorporation of base pairs at specific positions in oligonucleotides. The rationale to use isoG in PNAs as a substitute for C<sup>+</sup> was based upon the ability of isoG to circumvent pH dependence<sup>68</sup> in a pyrimidine- and purine-motif triple-helical complex.<sup>204</sup> The use of isoG in particular, was based upon the premise to apply a molecular frame that would place H-bond donors and acceptors in the appropriate positions to allow the recognition, and hence formation of G.C base pairs with two specific H-bonds, whilst maintaining a *N*-(2-aminoethyl) glycine backbone geometry compatible with the pyrimidine triple-helix motif.

Chemical work regarding base modification has previously and invariably focused upon the development of non-natural bases<sup>205,206,209</sup> whose energetics of triple helix formation are less sensitive to pH, thus benefiting from applications where near physiologically relevant conditions are a pre-requisite. Reported, modified bases designed to comply with this requirement include, the pyrimidine analogues pseudoisocytosine<sup>183</sup> (J), 5-methylcytosine<sup>70</sup> (5<sup>m</sup>C), and the purine analogues<sup>68,206</sup> (P1 and P2), 8-oxo-adenine<sup>207,233</sup> (8-oxoA), and the 7N regioisomer of G<sup>205</sup> (G<sup>7N</sup>).

## 2.3 The Known Effects of Modified Bases in DNA and PNA Oligomers

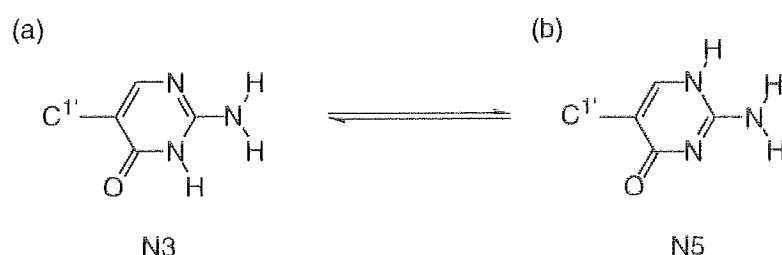
### 2.3.1 5-Methylcytosine (5<sup>m</sup>C)

A quantitative study regarding the energetics of triple helix formation for targeting different purine tracts confirmed 5<sup>m</sup>C provides complementary solutions towards the recognition of DNA G.C base pairs by mimicking C<sup>+</sup> in sequence-specific interactions with G.<sup>208</sup> In a study conducted by Povsnic *et al*,<sup>70</sup> the structural consequences of introducing 5<sup>m</sup>C at a single site opposite a G residue in an intramolecular triplex was investigated to achieve improved hybrid stability over the use of C<sup>+</sup>. From a study of triplexes containing 5<sup>m</sup>C, the pK<sub>a</sub> of N3 was postulated to be a major determinant of triplex stability. It was shown methylation at C extended the pH range (~ 0.4 units) for triple helix stability, attributable to the increase in the pK<sub>a</sub> of N3 in the triplex.<sup>70</sup> A higher pK<sub>a</sub> yields increasingly stable triplexes at physiological pH.<sup>201</sup> Compared with C<sup>+</sup>, an associated increase in binding affinity of the pyrimidine 5<sup>m</sup>C to its contiguous G.C target base pairs was observed by Povsnic *et al*, affording stable 5<sup>m</sup>C\*G.C triplets<sup>70</sup> (Figure 2.3). The ability of the electron-donating methyl substituent to stabilise protonation at N3 in complexed C<sup>+</sup> is significantly higher compared to its stabilising effect upon isolated C<sup>+</sup>. One possible interpretation implies substitution of hydrogen, with a methyl moiety at C5, promotes binding of the oligonucleotide via a hydrophobic effect. However, sequences of ascending

contiguous  $5^m\text{C}^+ \cdot \text{G} \cdot \text{C}$  tracts have been observed to display lowered stabilities at physiological pH as a consequence of inferior energetics.<sup>70</sup> This is attributed to positive charge repulsions between adjacent protonated N3 sites of contiguous  $5^m\text{C}$  bases, suggesting  $\text{pK}_a$  values are sequence-dependent,<sup>64</sup> and thus limiting the utility of  $5^m\text{C}$  for oligonucleotide directed triple-helix formation, particularly in G-rich sequences.<sup>208</sup>

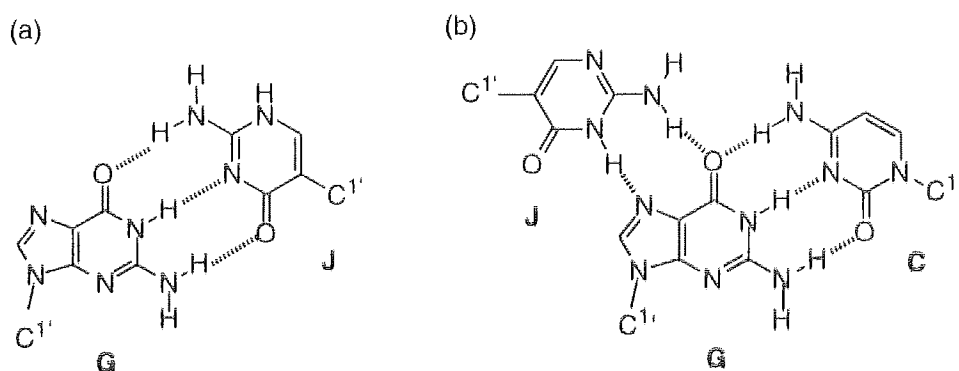
### 2.3.2 Pseudoisocytosine (J)

Pseudoisocytosine<sup>183</sup> (J) is a non-natural pyrimidine that is tautomerically ambiguous; the shift of a proton from N3 (Figure 2.4 (a)), where it presents the same H-bonding pattern form as its replacement  $\text{C}^+$ , to N5 (Figure 2.4 (b)) allows it to then adopt the same H-bonding pattern form as its replacement, C.



**Figure 2.4.** The (a) N3 and (b) N5 tautomeric forms of pseudoisocytosine (J).

Hence, the ability to tautomerise enables the formation of three H-bonds with G in the Hoogsteen motif (Figure 2.5 (a)), and two H-bonds with G.C base pairs in the Watson-Crick motif (Figure 2.5 (b)). This suggests that the use of synthetic bases expands the genetic alphabet to allow not only the formation of the full set of Watson-Crick base pairs, but also of a complete set of possible Hoogsteen base pairs. The H-bond donor-acceptor pattern of J is such that, dependent upon the tautomeric conformation it adopts, it may function either as a C analogue for Hoogsteen base pairing with G (Figure 2.5 (a)), or as a  $\text{C}^+$  analogue for Watson-Crick base pairing with G (Figure 2.5 (b)).



**Figure 2.5.** Representation of the two tautomeric forms of J in (a) a G.J base pair and a (b) a  $\text{J}^* \cdot \text{G} \cdot \text{C}$  base triplet.

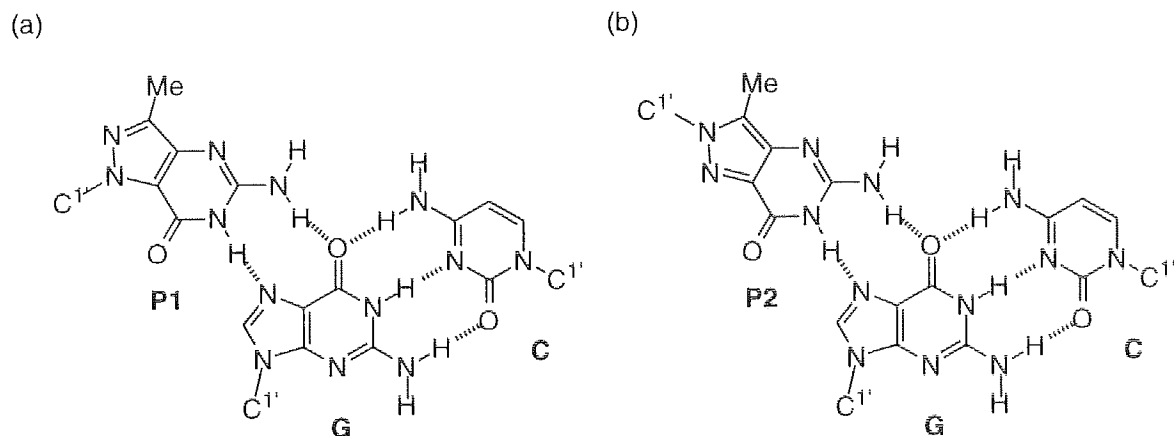
However, the G.J duplex has been shown to exhibit lowered stability compared with its replacement, the G.C base pair. The tautomeric ambiguity should not create serious problems regarding the hybridisation of J to duplex DNA in the major groove, though it does make it difficult to incorporate J at a specific sequences in an oligonucleotide synthesised by template-directed polymerisation.

In DNA oligomers, the inclusion of J in the antiparallel strand was found to be pH-independent indicating its preference to participate in Watson-Crick bonding as opposed to Hoogsteen bonding. This avoids the necessity for protonation as documented by Benner *et al.*,<sup>206</sup> who reported that J supported pH-independent triplex formation from 6.3 to 8.0 when incorporated in an appropriate oligonucleotide strand. Egholm *et al.*<sup>183</sup> synthesised PNA dimers, two PNA oligomers linked together, containing J in substitution of C<sup>+</sup>, with T and C in one strand, and T and J in the other PNA strand. Ensuing complex formation between the PNA oligomers with complementary oligonucleotides showed that thermal stabilities at acidic pH were comparable to those exhibited by the dimeric-PNA complex. The most efficient binding was observed with C in the anti-parallel (Watson-Crick) PNA strand, and with J in the parallel (Hoogsteen) PNA strand. This implies that dimeric PNAs could be designed to incorporate J with optimal Watson-Crick<sup>52</sup> and Hoogsteen recognition. The dimeric-PNAs were also found to display superior binding to monomeric PNAs for targeting duplex DNA by strand invasion.

### 2.3.3 Pyrazole Purines P1 and P2

Pyrazolo purines P1 and P2<sup>68,206</sup> (Figure 2.6 (a)) also possess a donor-acceptor H-bonding pattern on the pairing edge that mimics the Hoogsteen base-pairing-edge of C<sup>+</sup> (Figure 2.3). The use of increased sequence-specific interactions for pairing to contiguous G.C base pairs results in triplets with an ability to hybridise over an extended pH range, relative to both C<sup>+</sup> and 5<sup>m</sup>C. The proposed pairing alignment for the P1\*G.C triplex (Figure 2.6 (a)) involves installation of the imino proton to bind via Hoogsteen H-bonding with the N7 atom of G, and with the amino proton H-bonded to the O6 atom of the same G, in accordance with the C<sup>+</sup>\*G.C triplet (Figure 2.3). The pyrimidine of the third strand lies in the major groove of the Watson-Crick duplex,<sup>61</sup> parallel to the purine Watson-Crick strand and antiparallel to the pyrimidine Watson-Crick strand.

The P1 base has been shown to be accommodated comfortably in TFOs composed of G.C base pairs, and to participate in H-bonding interactions consistent with the triplexes shown in Figure 2.2 and Figure 2.3.<sup>206</sup> Oligonucleotides containing P1 bind to G-rich purine tracts in duplex DNA at pH 7.8. However, the issue remains whether triplexes containing multiple P2 residues are energetically disfavoured relative to 5<sup>m</sup>C at neutral pH. The possible structural perturbations induced by the introduction of this P2 ring system, compared with the P1 ring system, were assessed by Dervan *et al.*<sup>206</sup>



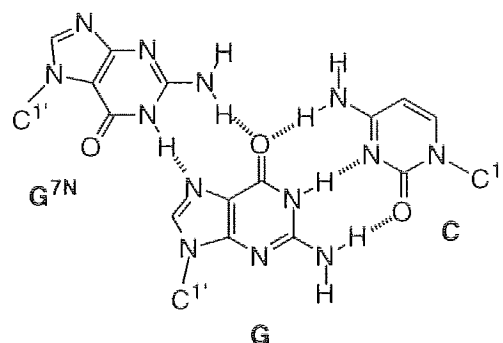
**Figure 2.6.** Hydrogen bonding in (a) P1\*G.C and (b) P2\*G.C base triplets.

Through a quantitative study it was shown that though both 5<sup>m</sup>C and P1 provide complementary solutions to the recognition of G.C base pairs, the orientation of both the phosphodiester backbone and the control of glycosidic conformation pertain increased importance. The structural implications of introducing P1 bases in an intramolecular pu-py-triplex were investigated by Radhakrishnam *et al.*,<sup>189</sup> where they were shown to exhibit sequence-specific binding of J in DNA oligomers with duplex DNA sites containing 6 contiguous G.C base pairs. Increased affinity compared with 5<sup>m</sup>C at physiological pH,<sup>68</sup> and facile accommodation in an otherwise pyrimidine third strand segment whilst participating in H-bonding interactions consistent with C<sup>+</sup>, was also observed.

It is proposed that the introduction of P1 increases the number of purine sequences available by oligonucleotide-directed triple helix formation for hybridisation to single sites in DNA. The marked difference and common drawback between isomeric P1 and its chemical isomer P2 is that P2 does not exhibit an acceptable degree of affinity to any Watson-Crick base pair. Figure 2.6 (a) shows that neither P1\*G.C nor P2\*G.C triplets are isomorphous with T\*A.T and C<sup>+</sup>\*G.C triplets (Figure 2.6 (b)). A weaker affinity cleavage pattern of the DNA duplex for P2\*G.C triplets, compared with P1\*G.C triplets, is attributed to the energetically unfavourable distortion of the third-strand backbone, arising as a consequence of the orientation of P2 which disfavors the *anti* conformation in the triple helix. The non-isomorphosity between P2\*G.C triplets and adjacent T\*A.T triplets results in an incompatible pu-pu-py bonding motif. These factors were addressed in the design aspects of this thesis work.

Purine-like non-natural nucleobases, described for the substitution of C<sup>+</sup>, include G<sup>7N</sup>; the N7 regioisomer of G.<sup>205</sup> Synthesised by Hunziker *et al.*,<sup>205</sup> G<sup>7N</sup> (Figure 2.7) closely resembles its P1 analogue (Figure 2.6 (a)). The attachment of the deoxyribose moiety at the N7-position of the glycosylated G analogue results in the reversal of the third strand

orientation, thus permitting binding to G.C residues (Figure 2.7) with exigency of backbone distortion.

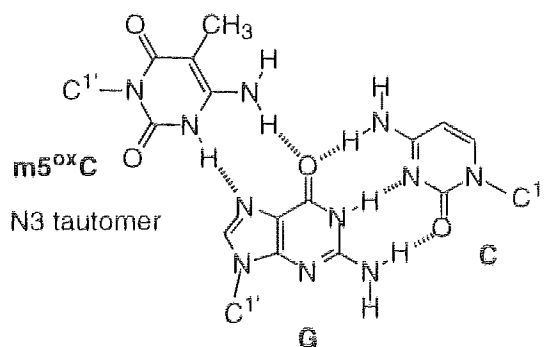


**Figure 2.7.** Hydrogen bonding in the  $G^{7N}^*G.C$  triplet.

The ensuing  $G^{7N}^*G.C$  base triplets, isomorphous with  $T^*A.T$  triplets, become parallel to the purine Watson-Crick strand.<sup>205</sup> The stabilities of DNA triple helical complexes containing  $G^{7N}$ , have been shown to decrease in the order:  $G^{7N}^*G.C > G^{7N}^*C.G > G^{7N}^*A.T > G^{7N}^*T.A$ <sup>68</sup> with affinities comparable to those of  $5^mC$ ; similar findings to those exhibited by P1. The high affinity and selectivity of  $G^{7N}$  are consistent with a model where  $G^{7N}$ , within the pyrimidine motif, forms two specific H-bonds to the purine strand of the Watson-Crick  $G.C$  base pair with remarkable specificity following incorporation in a pyrimidine oligonucleotide. On the basis of these results, it was envisioned a new parallel-stranded motif comprising of wholly N7 purines would be a possibility for DNA recognition through triplex formation.<sup>205</sup>

#### 2.3.4 5-Methyl-6-Oxo-Cytidine ( $m5^{ox}C$ )

5-Methyl-6-oxo-cytidine ( $m5^{ox}C$ ), a substituted pyrimidine, has been shown to bind to DNA  $G.C$  base pairs (Figure 2.8) with increased selectivity but decreased stability compared to  $5^mC$ .<sup>209</sup>

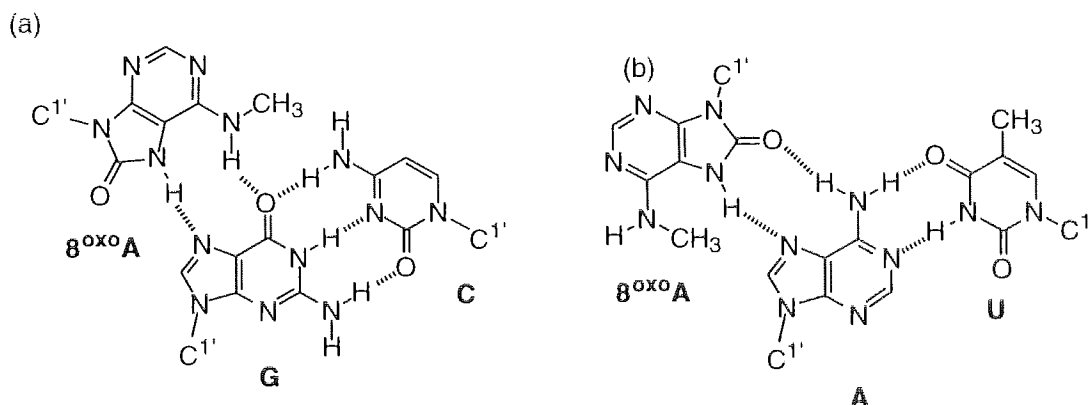


**Figure 2.8.** Triplex formation between  $m5^{ox}C$  and a  $G.C$  base pair.

The  $T_m$  of triple to double helix transitions, for triplexes containing  $m5^{ox}C$ , has been shown to be much lower at neutral pH than  $5^mC^+$ -containing triplexes. Indeed,  $m5^{ox}C$ -containing duplexes are less stable than their natural counterparts.  $m5^{ox}C$  possesses the ability to tautomerise, with preference conferred to the N1 tautomer, to bind to G.C base pairs. This tautomerism affords a two Hoogsteen bond with the G.C pair (Figure 2.8) compared to the unstable, one Hoogsteen bond that would result with the  $m5^{ox}C$  N3 tautomer.

### 2.3.5 $N^6$ -Methyl-8-Oxo-Adenine ( $8^{oxo}A$ )

$N^6$ -methyl-8-oxoadenine ( $8^{oxo}A$ )<sup>207</sup> which tends to exist in its *syn* conformation,<sup>203</sup> binds to C.G (Figure 2.9 (a)) and U.A (Figure 2.10 (b)) triads, the latter stabilised by a H-bond between the N6 exocyclic amino group and the O4 of uridine, to afford the triplets depicted in Figure 2.9. However, these mismatched triplets are less stable than  $8^{oxo}A^*G.C$  triplets. The lack of structural isomorphism between  $B^*G.C$  (where  $B = P1, G^{7N}, 8^{oxo}A$ ) and  $T^*A.T$  triplets, which also prevents binding for P2, was shown to result in lowered affinity for targeting an alternated  $(G.A)_5$  versus a contiguous  $G_6$  strand.<sup>68</sup> These studies confirmed the associated energy penalty for non-isomorphism of adjacent triplets.



**Figure 2.9.** Triplex formation of  $8^{oxo}A$  with an (a) G.C base pair and (b) a U.A base pair.

## 2.4 Use of Non-natural Bases for pH Circumvention for Triplexation Studies

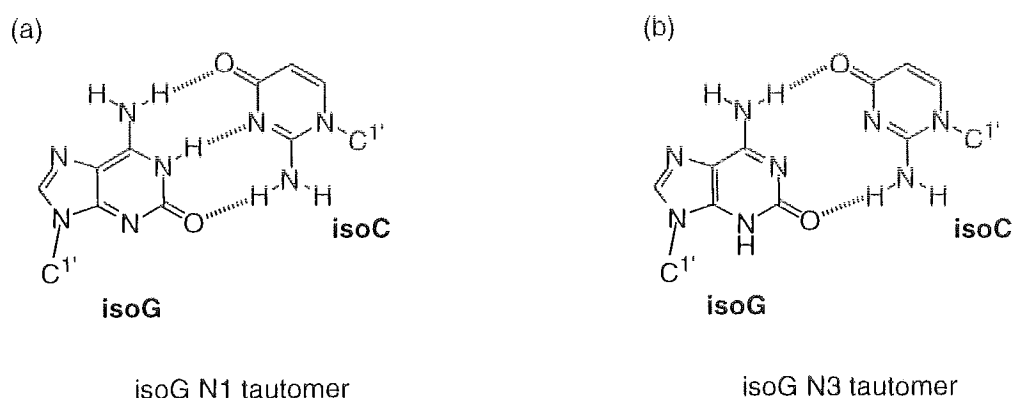
### 2.4.1 IsoGuanine (isoG) as a Potential Mimic of $C^+$ for Stable Triplex-Formation

Though  $C^+$  and  $5^mC$  extend the pH range, they still comprise a certain degree of pH dependency and backbone distortion, arising from electrostatic repulsion between contiguous base pairs. A neutral heterocyclic mimetic of  $C^+$  would enable the synthesis of



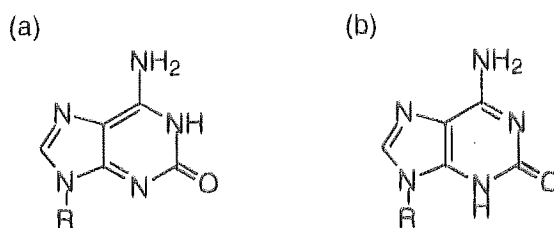
oligomers displaying enhanced specific affinity for duplex DNA, under conditions that mimic the intracellular environment. It was towards this aspect we pursued an interest in the design of isoG as a modified base with an ability to mimic C<sup>+</sup>, and hence form stable triple helices at physiological pH in its isomorphous orientation, thus circumventing pH dependency. The alleviation of the repulsion associated with 5<sup>m</sup>C and C<sup>+</sup>, with concurrent absence of the characteristic energetic penalties associated with the previously discussed bases, offers great potential for DNA targeting.

Studies have shown that when isoG tautomers are of comparable stabilities, complementary binding selects the tautomer that yields the most stable duplex structure.<sup>211</sup> Evidence of this was documented using isocytosine<sup>212</sup> (isoC) in the isoC-isoG scenario investigated by Switzer *et al.*,<sup>165</sup> who showed that both DNA and RNA polymerase direct the incorporation of isoG to pair with isoC in a DNA template. These findings show isoC pairs solely with the N1 tautomer of isoG (Figure 2.10).



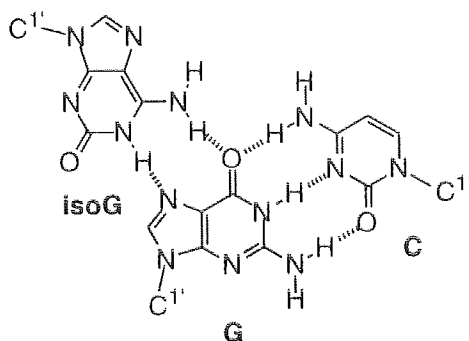
**Figure 2.10.** IsoC in a Watson-Crick pair with (a) the N1 and (b) N3 tautomers of isoG.

The utilisation of isoG PNA monomers, in particular the A-mimicking isoG N1 tautomer (Figure 2.11(a)) for targeting A.T base pairs, possesses the ability to successfully substitute C<sup>+</sup> and hybridise using Watson-Crick pairing, in an antiparallel orientation to the target duplex, with Reversed-Hoogsteen bonds.<sup>51</sup>



**Figure 2.11.** The (a) N1 and (b) N3 tautomers of isoG (where R = CH<sub>2</sub>OCH<sub>2</sub>CH<sub>3</sub>).

The N3 isoG tautomer (Figure 2.11(b)) does not possess the correct orientation for such binding to G.C base pairs. This results in an increasingly stable triple helix motif (isoG\*G.C) compared with G.C base pairs, as depicted in Figure 2.12.

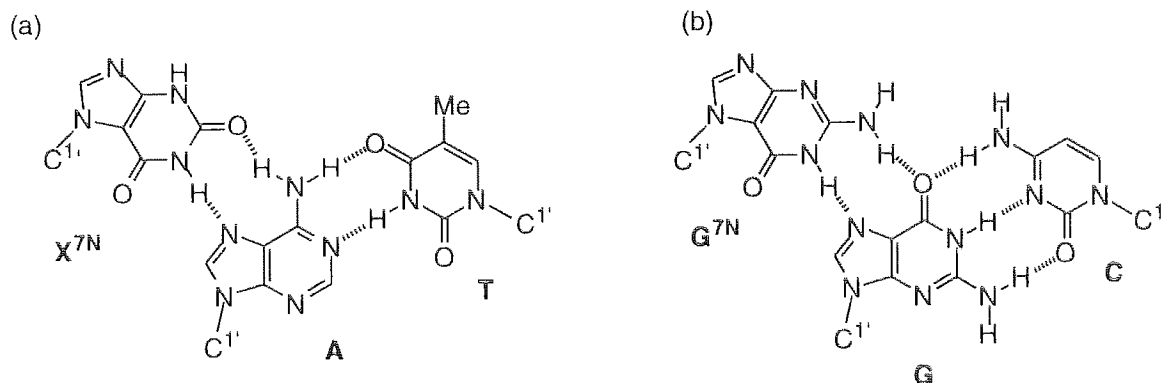


**Figure 2.12.** Hybridisation of isoG with a G.C base pair.

#### 2.4.2 The Importance of N7 Regioisomeric DNA Bases and Their Derivatives

Though we were primarily concerned with the design and synthesis of the N9 regioisomers of base-modified PNA oligomers, it was recognised that the design of new DNA binding motifs, in particular the stability of isolated planar triplets, was dependent upon the structural isomorphism of contiguous base triplets.<sup>51</sup> Isomorphism governs the energy penalty associated with backbone distortion for ascending contiguous triplets. This factor pertains consideration regarding the necessity to ensure pairing of G.C base pairs with a matching base that maintains integrity of both the backbone and the nucleobase distances. The design and synthesis of alternative combinations of isomorphous triplets minimise the effects of backbone distortion, hence reducing adverse effects imposed upon triple helix stability.<sup>213</sup> The mode that introduces the least perturbation to the Watson-Crick base pair is less important than the energy penalty associated with backbone distortion imposed by non-isomorphous triplets.<sup>51</sup> To date, only homopyrimidine oligonucleotides form a third strand without backbone distortion. Substitution at N7 may intrude perturbations that may be invoked to explain this discrimination, including adverse effects upon base orientation.

In this thesis work, the N7 regioisomers of modified nucleobases were identified to act as potential precursors, with an ability to maintain a regular distance between adjacent bases. This permits the investigation and assessment of the optimal orientation for the third strand relative to the duplex. The DNA regioisomers X<sup>7N</sup>, A<sup>7N</sup>, I<sup>7N</sup> and G<sup>7N</sup>, maintain a regular distance between adjacent bases,<sup>73</sup> where the pyrimidine or purine is attached through a glycosidic C1'-N9 or C1'-N7 bond to the sugar phosphate backbone of phosphodiester-linked oligomers. When the C1' atom of the Watson-Crick base pair is in a fixed orientation, the position of the C1' atom of the nucleoside in the third strand is dependent upon the base triplet.



**Figure 2.13.** The isomorphous (a)  $X^{7N^*}A.T$  and (b)  $G^{7N^*}G.C$  triplexes.

In order to circumvent the backbone distortions that result from non-isomorphism, the N7 regioisomer of the non-natural nucleobase X was recognised to form  $X^{7N^*}A.T$  base triplets, isomorphous with  $C^{+*}G.C$  base triplets at physiological pH, as shown in Figure 2.13 (a). The glycosidic bond ( $C1'-N7$ ) of the  $X^{7N}$  purine connects the N7 site, as opposed to N9, to the sugar. Pyrimidine-like bases isoC, and J, P1<sup>206</sup> and m5oxC<sup>209</sup> are all isomorphous with  $T^*A.T$  triplets, where the H atom at the N3 site permits binding to G in a pH-independent fashion.

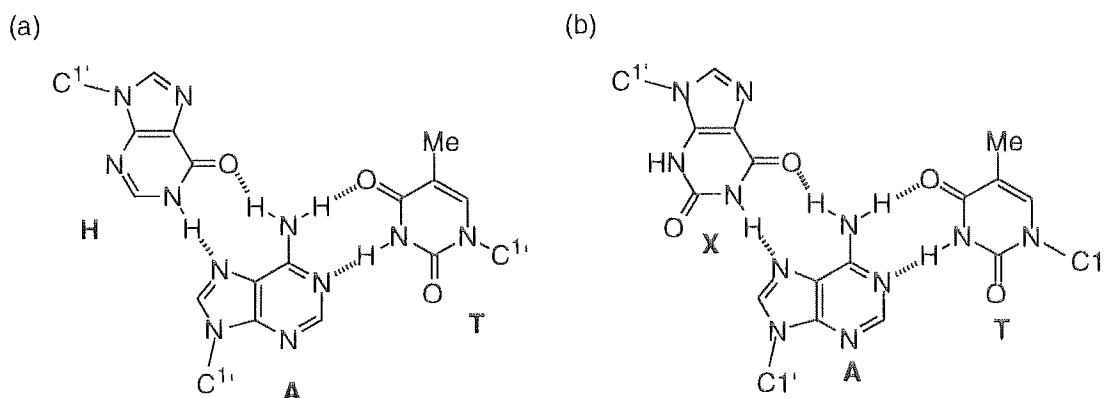
## 2.5 The Effects of Xanthine and Hypoxanthine Upon Molecular Recognition of PNA Oligomers

### 2.5.1 The Known Contributions of Xanthine and Hypoxanthine to the Hybrid Stability of DNA Oligonucleotides

It was recognised that the control of glycosidic conformation, compatible with the peptidic backbone within the py-pur-py motif,<sup>204</sup> is as important a design criterion as matching H-bond donors-acceptors of heterocycles for eliminating geometrical constraints for base-specific triple-helix formation. The substitution of  $C^+$  with 5<sup>m</sup>C (Figure 2.3) for targeting G.C base pairs, affords a base triplet isomorphous with  $T^*A.T$ . (Figure 2.3). Similarly, an isoG\*G.C in a PNA triplet is non-isomorphous with  $T^*A.T$  since intra-base distance is far greater. In order to retain the isomorphous property, and hence reduce an element of instability as a potential consequence of backbone distortion, conferred by the non-isomorphous isoG\*G.C and  $T^*A.T$  triplets, we focused upon the synthesis of alternate bases for binding to A.T base pairs, where X and H base-modified PNA oligomers were employed to address these requirements.

It was recognised that the substitution of T in the third strand of PNAs with X, and its structural analogue H, provided complementary solutions to the recognition of A.T base pairs. The design thus focused upon the non-standard X.T and H.T base pairs envisaged as potential pairings for incorporation in PNA oligomers to target G.C- and A.T-paired DNA oligonucleotide sequences. The resultant stable  $H^*A.T$  and  $X^*A.T$  triplets, as

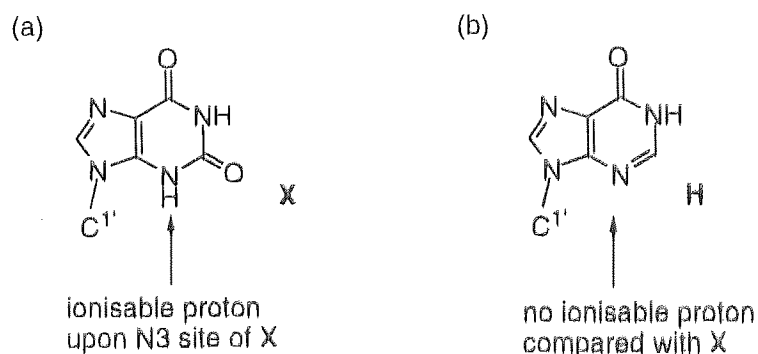
depicted in Figure 2.14 (a) and (b) respectively, are isomorphous with isoG\*G.C triplets (Figure 2.12).



**Figure 2.14.** Isomorphism of (a) H\*A.T and (b) X\*A.T with isoG\*G.C triplets (Figure 2.12).

### 2.5.2 The Potential Hybridisation Properties of Xanthine Compared with Hypoxanthine

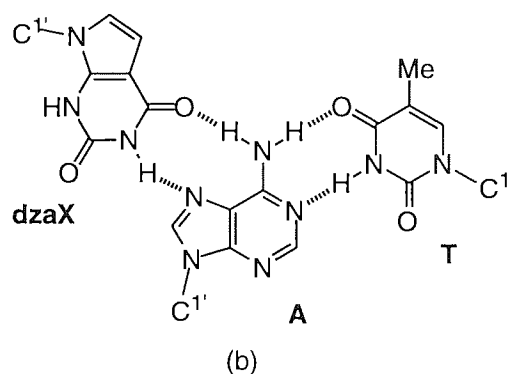
We opted to synthesise the X PNA building block due to the presence of the additional C2 carbonyl bond within the pyrimidine moiety of the heterocyclic ring, compared with H which possesses only two potential H-bond sites relative to its natural homologue. X was hypothesised to yield oligomers exhibiting greater binding affinity and increased stability, attributed to the additional resulting donor-acceptor base pair bonds. X presents a H-bond 'acceptor-donor-acceptor' pattern compatible to its complementary standard base, T (Figure 2.16). Though the N3 tautomer of X (Figure 2.15 (a)) allows stable triplexation formation with A.T base pairs (Figure 2.14 (a)), with a  $pK_a$  of 5.7 when free in solution, X is far more acidic as a heterocycle than standard nucleobases. Whilst the  $pK_a$  of this nucleobase increases when incorporated in DNA oligonucleotides, the acidity may also present problematic aspects. This is attributed to the ionisable proton upon the N3 site, which is highly susceptible to ionisation at physiological pH, thus conferring a potential element of pH dependence.



**Figure 2.15.** Ionisation of (a) X compared with (b) H.

In turn, this may result in repulsive forces with a subsequent lowered degree of stability. Hence, the possibility for replacement of T with X at physiological pH is unfeasible. In contrast, as H does not contain an ionisable proton on N3, the employment of H for targeting A.T base pairs ensures there is no element of pH dependency (Figure 2.15 (b)).

An alternative approach to eliminate the potential introduction of pH dependency by X, is to substitute the N3 site with a CH group. Milligan *et al*<sup>214</sup> identified a purine analogue with substitution at N7 of X, and a CH group in replacement of the ionisable N3, afforded 7-deaza-2'-deoxyxanthosine (dzaX)<sup>214</sup> (Figure 2.16).



**Figure 2.16.** Targeting of the dzaX base to A.T base pairs.

dzaX was shown to enhance the binding of oligodeoxynucleotides to duplex DNA in the anti-parallel orientation under intracellular conditions, increasing the  $pK_a$  from 5.3 (uncomplexed X) to 7.8 (bound to A.T base pairs). This potentially allows increased binding characteristics due to the increased regular geometry of the third strand. The observed enhancement of binding displayed by dzaX compared with T, could potentially reflect a decrease in the repulsive forces between the phosphodiester backbone in the third strand, with that in the duplex. Under physiological conditions, the substitution of dzaX for T significantly increases the utility of the anti-parallel motif and may prove useful for *in vivo* inhibition of gene expression.

## 2.6 The Potential of Hypoxanthine as a Universal PNA Base

A number of attempts have been made to use modified bases that behave in a non-discriminatory manner towards the natural bases for oligonucleotide primers and probers. A universal base which could substitute for any of the four natural bases would be of great utility for manipulating DNA *in vitro*. In previous work, investigation has focused upon the thermal stability studies of oligonucleotides containing a variety of potentially suitable analogues. However, towards this aspect except for H, only 3-nitropyrrole has been investigated in respect of enzymatic reactions, with primers containing a universal DNA

base.<sup>215</sup> In efforts to extend the molecular recognition properties of PNA, the purine base hypoxanthine (H) was identified as a possible candidate for exhibiting properties enabling it to behave as a non-discriminatory base for incorporation in PNA oligomers, without causing a significant destabilisation of the duplex, and with a potential to exhibit a higher  $T_m$ . The poor ability of H to bind with lower affinity, conferring decreased stability shown in DNA-containing oligonucleotides, is attributed to the presence of one less carbonyl bond at the N3 site relative to its X analogue, the very property that allows exploitation of H as a universal base. The associated lowered specificity enables targeting to a series of nucleobases, where a stringent complementary motif pattern is not a prerequisite. Potential applications of universal DNA bases include PCR, where the base is attached through N9 to 2'-deoxyribose in oligonucleotide primers.<sup>216</sup> The isomeric N7 nucleoside, where the N7 of H is connected to configured 2'-deoxyribose, displays interesting DNA recognition properties when incorporated into triplex-forming oligonucleotides. If H is able to display stabilisation properties with DNA duplexes by stacking interactions, rather than by H-bonding, and to pair with little variation when set opposite the 4 natural bases, its applications may be extended towards a possible universal base.<sup>217</sup>

## **2.7 Conclusion**

In conclusion, those problems posed with the use of current bases, both modified and unmodified, and their subsequent implications in duplex and triplex formation, have been addressed in this chapter. The use of isoG, X and H have been chosen to address, overcome and improve upon these aspects.

# CHAPTER 3

## The Synthesis of Base-Modified PNA Intermediates

### 3.1 Aims of the Synthetic Work

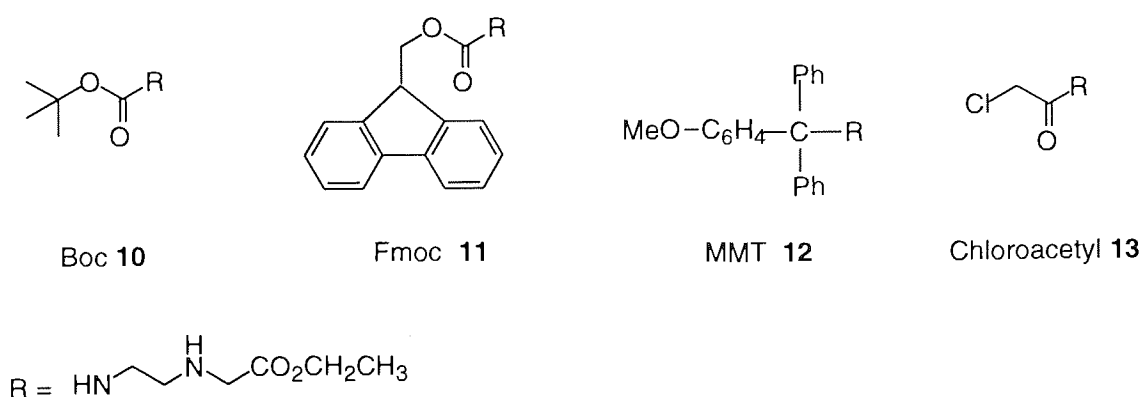
The three aims of this synthetic work were to:

- synthesise *N*-(2-butyloxycarbonyl-aminoethyl)glycinate ethyl ester **1**, for coupling to modified PNA bases
- synthesise novel PNA building blocks composed of the purine bases D **2**, X **3**, H **4**, and isoG **5**
- to examine crystalline derivatives by X-ray crystallography

### 3.2 Potential Protecting Groups for the Primary Amino Group of PNA Monomers

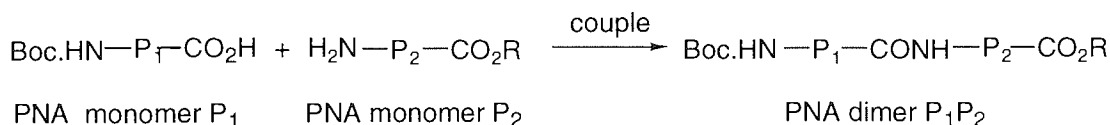
Automated solid-phase synthesis of PNA oligomers composed of the DNA and novel heterocyclic bases requires suitably protected building blocks.<sup>218-221</sup> This in turn necessitates the synthesis of the protected backbone component *N*-(Boc-aminoethyl)glycinate ethyl ester **1**, for coupling to the corresponding acetic acid derivatives of the appropriate base, to afford the new PNA monomers **6**, **7**, **8**, and **9**.

In addition to *tert*-butoxycarbonyl (Boc) **10**, the alternative protecting groups<sup>13</sup> such as fluorenylmethoxycarbonyl (Fmoc)<sup>222</sup> **11**, monomethoxytrityl (MMT)<sup>223</sup> **12**, and chloroacetyl (ClCH<sub>2</sub>CO-) **13**, have been employed in PNA chemistry for amino protection and are shown in Figure 3.1. The four amino protecting groups **10**, **11**, **12** and **13** require contrasting conditions for efficient and highly selective removal, without compromising either the stability of the remaining base protecting groups or affecting the linkage to the solid support.



**Figure 3.1.** Three alternative protecting groups, to Boc, for the amino moiety of **1**.

The use of the Boc group **10** in PNA synthesis is well-documented and provides a synthetic method, ideal for automation, to afford desired products in high yields.<sup>220,224</sup> Boc protection of the peptidic backbone **1** permits a specific mode of coupling between the amino group of one monomer for attachment to the carboxyl group of the proximate monomer, alleviating potential side reactions between primary amino groups of the same monomer. This process may be illustrated using general peptidic compounds, P<sub>1</sub> and P<sub>2</sub>, where R is a carbonyl protecting group such as an ethyl group for example, or an attachment to a solid-support as shown in Figure 3.2.



**Figure 3. 2.** Coupling of protected PNA monomer units P<sub>1</sub> and P<sub>2</sub>.

The Fmoc-mediated synthesis of PNA monomers devised by Thomson *et al*,<sup>222</sup> was reported two years after the onset of this work and has been shown to provide an alternative choice to the Boc moiety **10** for amino protection.<sup>152,223,225</sup> However, few other reports have since appeared, despite the offer of the several advantages over the Boc group including: milder synthesis conditions, improved monomer solubility and increased coupling efficiency.<sup>152,225</sup>

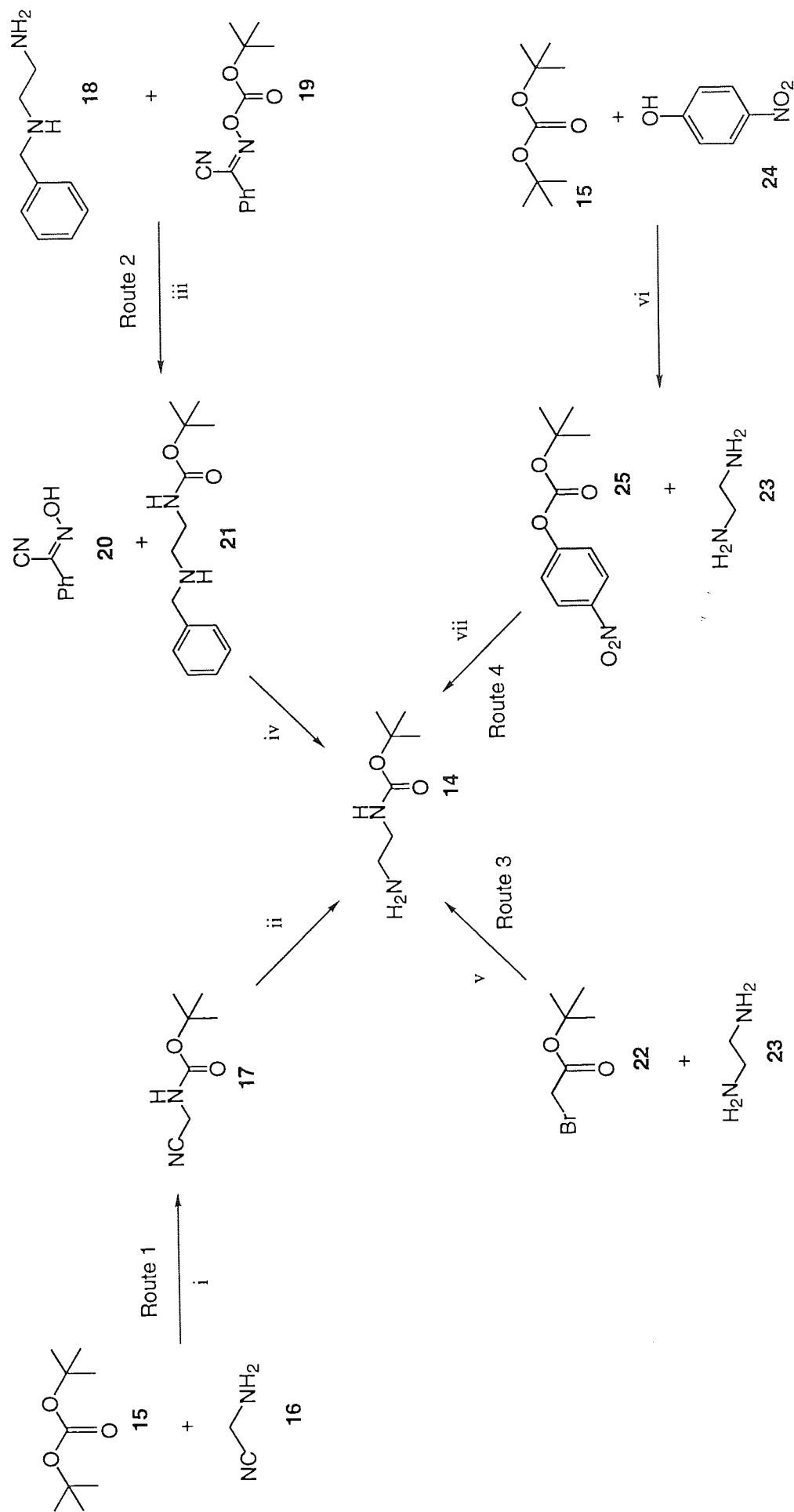
Breipohl *et al*<sup>223,226</sup> employed the MMT-protected peptidic backbone as an alternative choice to the Boc-protected backbone **1** for synthesis of PNA-DNA chimerae.<sup>172,226</sup> This was deemed necessary as the repeated treatment of PNA-DNA chimerae with trifluoroacetic acid (TFA), essential for Boc deprotection, and the harsh treatment with hydrofluoric acid (HF), essential for cleavage from the solid support, rendered the Boc/CBz strategy incompatible for PNA-DNA chimerae synthesis. This was attributed to the increased sensitivity of the DNA component of the dimers to strong acids. The procedure was also observed to be time-consuming, and involve the formation of several side-products which produced unstable intermediates and gave low overall yields.

### 3.3 Synthesis of the PNA Backbone Component **1**

#### 3.3.1 The Synthesis of Ethyl *N*-(Boc-aminoethyl) glycinate **1**

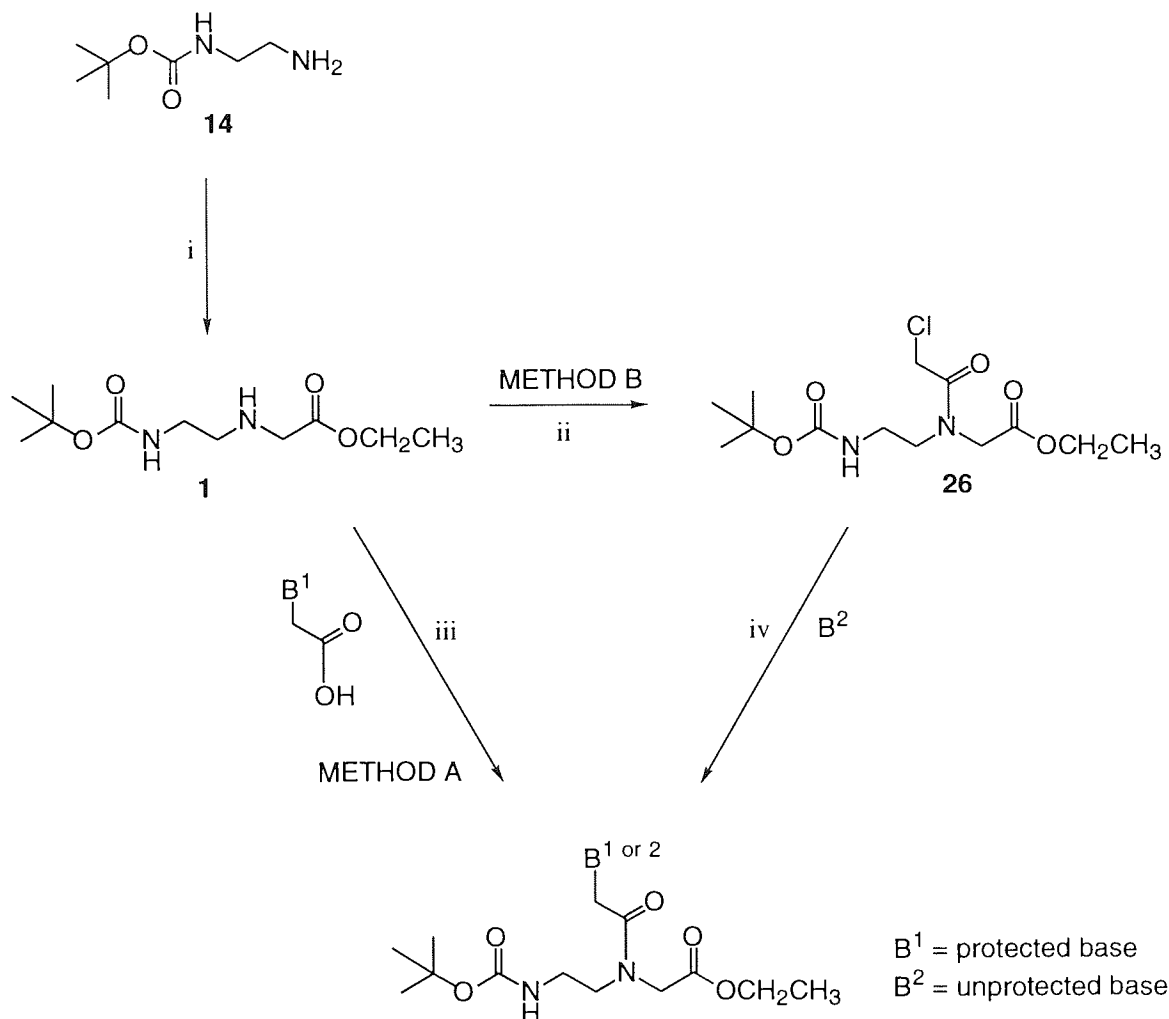
As reported by Nielsen *et al*,<sup>218</sup> the PNA backbone **1** was first synthesised in its glycinate form as ethyl *N*-(Boc-aminoethyl) glycinate, which has since been superseded with the use of the *N*-(2-aminoethyl)glycine side chain (Aeg).<sup>21,227,228</sup> This, and subsequent reported syntheses towards the common precursor *N*-(Boc-amino)ethylenediamine **14**, of the backbone component **1**,<sup>229</sup> are shown in Scheme 3.1.





**Scheme 3.1.** Reagents: i,  $\text{Et}_3\text{N}/\text{CH}_2\text{Cl}_2/\text{r.t.}$ ; ii, Raney Nickel/ $\text{NH}_3/\text{EtOH}/60 \text{ atm.}/24 \text{ h}$ ; iii, 1,4-dioxane/ $0 \text{ }^\circ\text{C}$ ; iv,  $\text{EtOH}/\text{H}_2/10\% \text{ Pd on charcoal}/5 \text{ d}$ ; v,  $\text{CH}_2\text{Cl}_2/\text{r.t.}/24 \text{ h}$ ; vi, DMF/ $24 \text{ h}/\text{r.t.}$ ; vii, 1,4-dioxane/ $\text{Na}_2\text{CO}_3/120 \text{ }^\circ\text{C}/2 \text{ h}/0 \text{ }^\circ\text{C}$ .

The final stages of the two routes employed for the synthesis of the PNA backbone component **1** from **14**, are shown in Scheme 3.2.<sup>218,219</sup>



**Scheme 3.2.** Reagents: i,  $\text{BrCH}_2\text{CO}_2\text{CH}_2\text{CH}_3/\text{KF}/\text{celite}/70\text{ }^\circ\text{C}/20\text{ min}/\text{CH}_3\text{CN}$ ; ii,  $\text{Et}_3\text{N}/\text{ClCH}_2\text{COCl}/\text{CH}_2\text{Cl}_2/0\text{ }^\circ\text{C}$ ; iii,  $\text{DCC}/\text{DhbtOH}/\text{DMF}/\text{DMAP}/70\text{ }^\circ\text{C}/48\text{ h}$ ; iv,  $\text{NaH}/\text{DMF}/75\text{ }^\circ\text{C}/1\text{ h}$ .

Following the evaluation of several alternative methods, (routes 2, 3 and 4), the procedure reported by Meltzer *et al* (route 1) was adopted to afford **1** in excellent purity and increased yields.<sup>219</sup> The preparation of *N*-(Boc-amino)ethylenediamine **14** in two steps from aminoacetonitrile hydrochloride **16** was undertaken according to Scheme 3.2. The precursor *N*-(Boc-amino)acetonitrile **17** was obtained as a yellow oil<sup>219</sup> on acylation of **16** with a slight excess of di-*tert*-butyl dicarbonate **15**, in the presence of excess triethylamine. Recrystallisation from hexane furnished **17** as white crystals in consistently improved yields of 80%, compared with the reported literature, where a yellow viscous oil was reported. Structural confirmation of **17** was obtained for the first time by X-ray crystallography, with the acquirement of a very good refinement value.<sup>217</sup> In chapter 5,

the crystal structure of **17** is compared with that of structurally similar Boc-protected amino acids.<sup>18,216</sup>

Hydrogenolysis of **17** using Raney nickel in ethanol saturated with ammonia (10% NH<sub>3</sub>/EtOH) at 50 psi afforded **14** in near quantitative yield as a dark green, viscous oil. Alkylation of *N*-(Boc)ethylenediamine **14** in anhydrous CH<sub>3</sub>CN, with ethyl bromoacetate BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub> **27** in the presence of excess KF/Celite, was carried out at elevated temperature to yield the desired backbone component **1** as a pale yellow oil (Scheme 3.2). A short work-up furnished the Boc-protected peptidic backbone **1** in purer form than the commercially available material, as shown by <sup>1</sup>H NMR (Figure 3.3).

Alternative, but less successful attempted methods for the synthesis of *N*-(Boc-amino)ethylenediamine **14**, are shown in Scheme 3.1 (routes 2, 3 and 4), and are briefly discussed. Brief reaction of 2-(*tert*-butoxycarbonyloximino)-2-phenylacetonitrile (Boc-ON) **19**, and the commercially available (*N*-benzyl)ethylenediamine **18** at r.t. (Scheme 3.1, route 2), afforded mono-*N*-Boc-benzylamine **21** as a yellow, viscous oil, concurrent with large quantities of the oximino-2-phenylacetonitrile by-product **20**. This necessitated repeated extraction for isolation of **14**. Flash column chromatography using EtOAc-MeOH (4:1), proceeded with repeated recrystallisation, eventually gave very small quantities of the pure material **14**. Structural determination of the **20** was confirmed by X-ray crystallography.

Hydrogenolysis of the small quantity of **21** attained from this route, the requirement for large quantities of 10% Pd on charcoal and lengthy reaction periods (5 d) for conversion of the benzyl-protected **21** to **14**, furnished low yields of impure **14**, where spectroscopic data was not able to provide sufficient proof regarding the acquirement of **14** (Scheme 3.1). Hence, this procedure was abandoned due to its inefficiency, which resulted in very poor yields of **14**.

Alternatively, reaction of *tert*-butyl bromoacetate **22** with an excess of ethylenediamine **23** (Scheme 3.1, route 3) in CH<sub>2</sub>Cl<sub>2</sub> at r.t., produced **14** in a crude form. Solvent extraction afforded the precursor **14** in 27 % yield. This approach was not pursued any further due to the poor yields of **14** isolable in pure form.<sup>222</sup>

In route 4, Boc-4-nitrophenylcarbonate **25** was prepared by refluxing a mixture of Boc anhydride **15**, 4-nitrophenol **24**, and sodium carbonate in 1,4-dioxane. A solution of **25** in DMF was added to a stirred solution of ethylenediamine **23** according to the literature method.<sup>4</sup> However, the extremely low yield (11%) of *N*-(Boc-amino)ethylenediamine **14**, obtained as a pale yellow oil following a lengthy solvent extraction procedure, promoted use of a higher yielding and more efficient method.

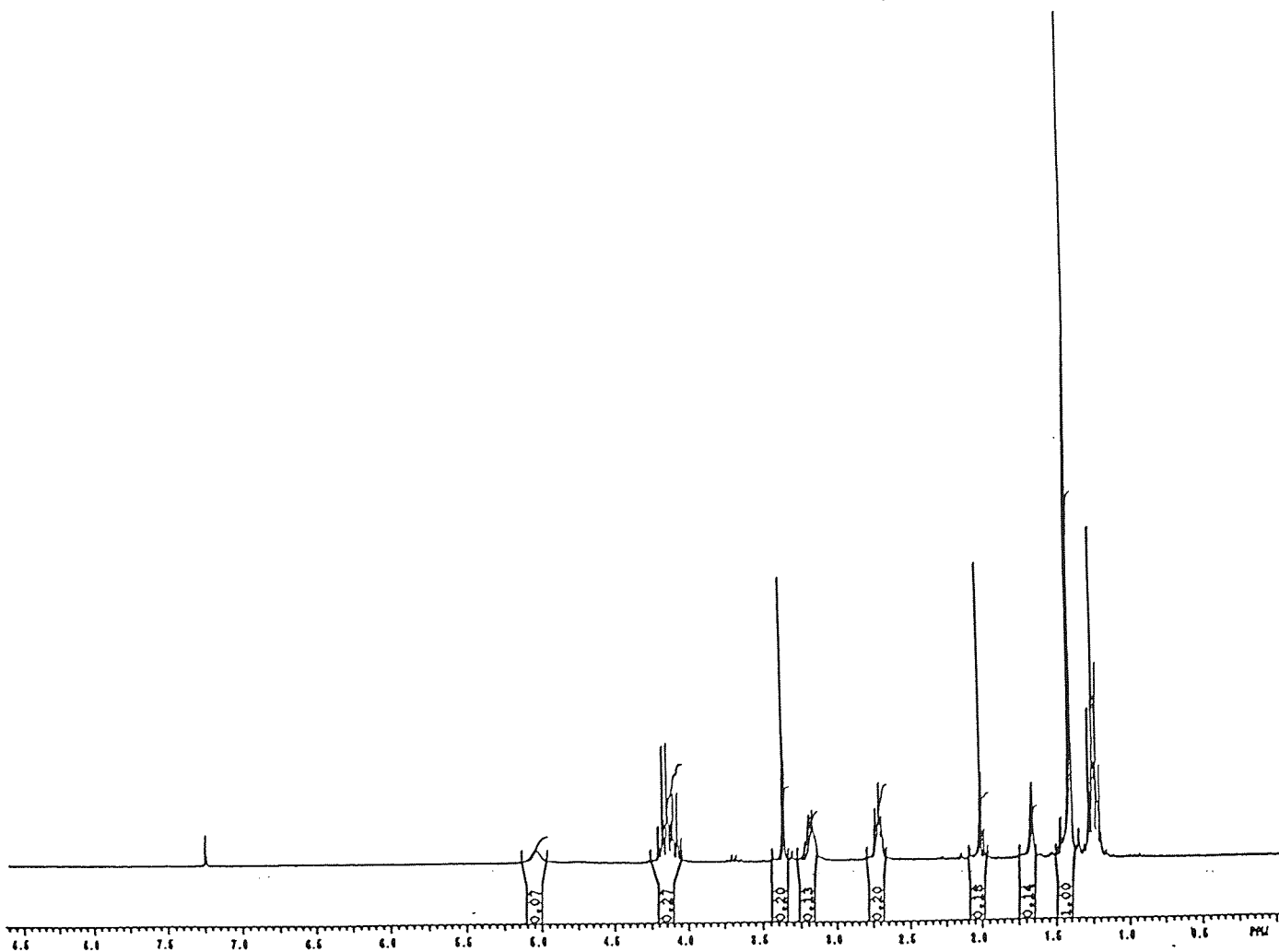
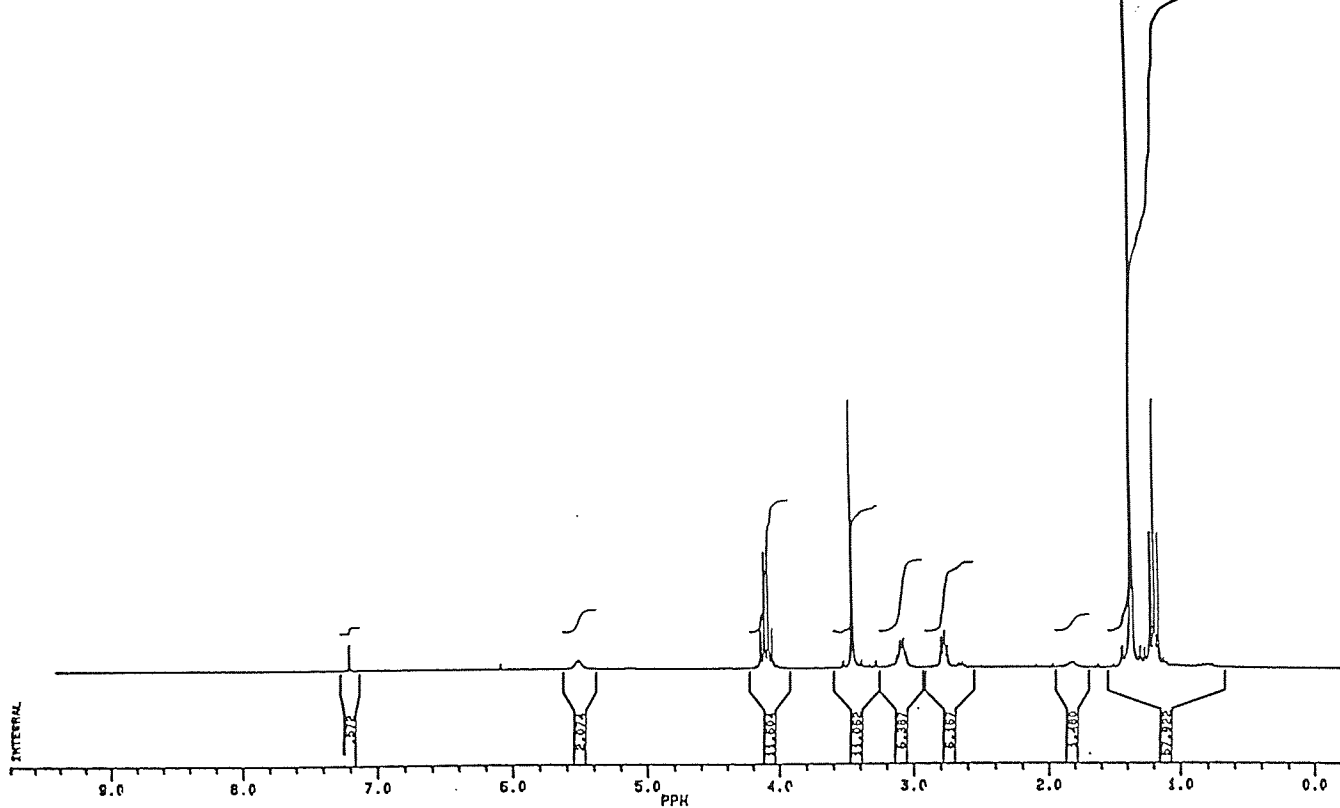
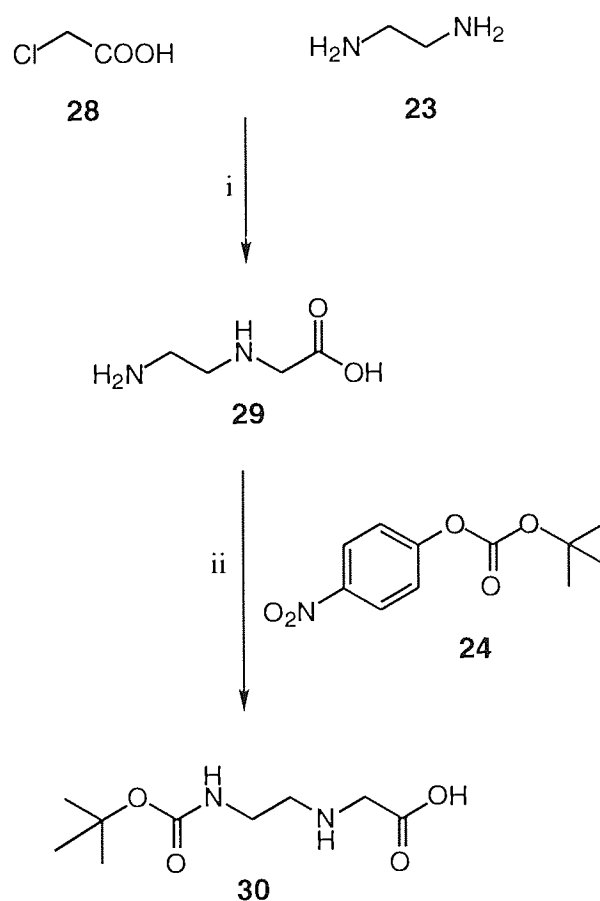


Figure 3.3. Comparison of  $^1\text{H}$  NMR spectra of (a) synthesised ethyl *N*-(Boc-aminoethyl) glycinate **1** and (b) commercially available **1**.

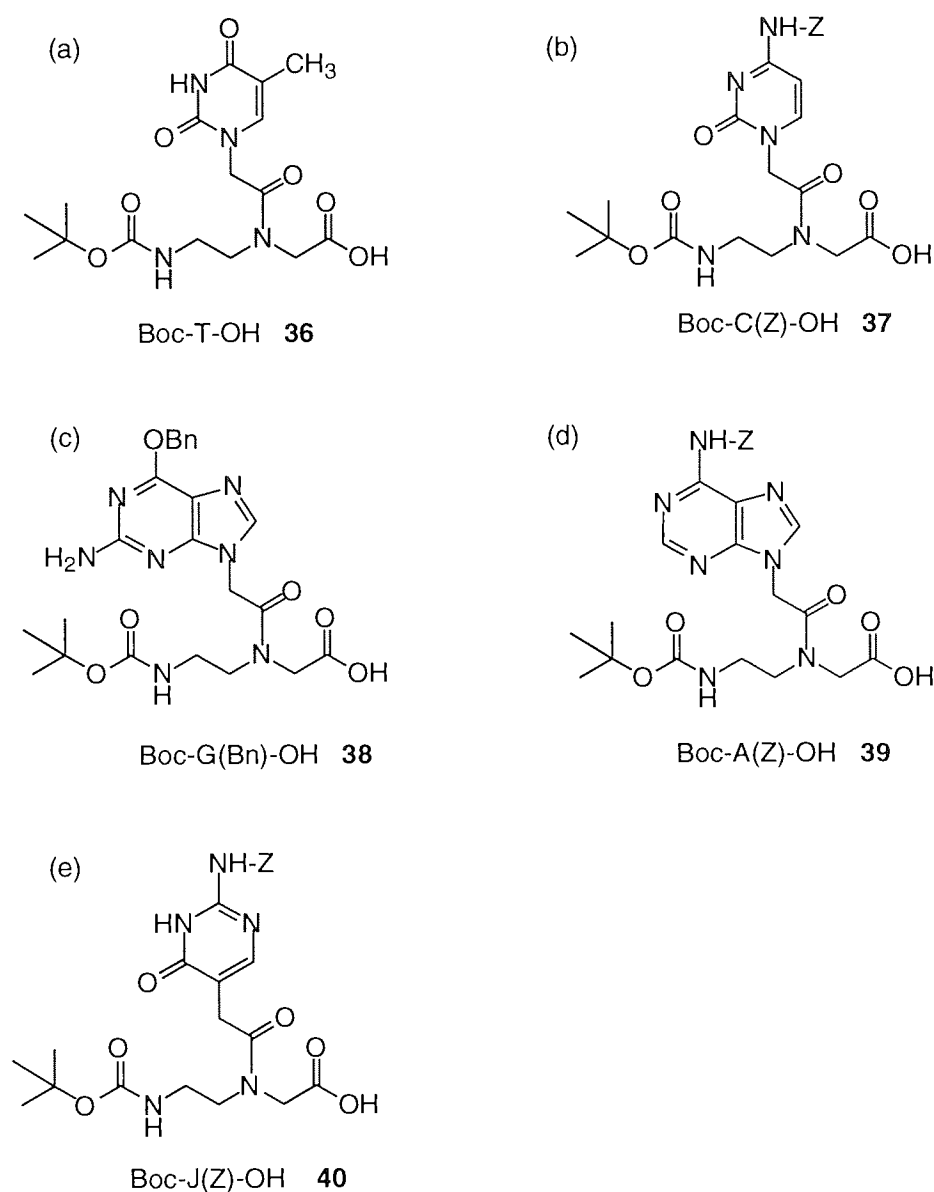
An alternative procedure towards the synthesis of the carboxylic acid *N*-(2-aminoethyl)glycine **30** was also undertaken.<sup>228</sup> Addition of chloroacetic acid at 0 °C to **23** gave *N*-(2-aminoethyl)glycine) **29**. Solvent extraction, followed by recrystallisation from EtOH-H<sub>2</sub>O, afforded **29** as a highly pure, white crystalline compound in 60% yield (Scheme 3.3). Dropwise addition of Boc *p*-nitrophenylcarbonate **24** in 1,4-dioxane, to a solution of **29** in 50% aqueous 1,4-dioxane, ensued in the release of H<sup>+</sup>, formed by attack of the terminal amino group upon the carbonyl centre of **24**. Subsequent pH fluctuations at 11.2 were controlled via dropwise addition of a buffer solution. Compound **29** was isolated in 21% yield following solvent extraction and a further follow-up procedure afforded the desired compound **30**. However, the need to manually maintain the pH at 11.2 for 16 h, which afforded **30** in low yields with incomplete analysis, prompted the investigation of an alternative method with increased efficiency and purity, and without the need for such labour intensive and time-consuming stages. Furthermore, in accordance with the literature, it was decided the glycinate, rather than the glycine form of **1**, would be synthesised.



**Scheme 3.3.** Reagents: i, ClCH<sub>2</sub>COOH; ii, 2 M NaOH/50% aq. 1,4-dioxane/r.t./24 h.

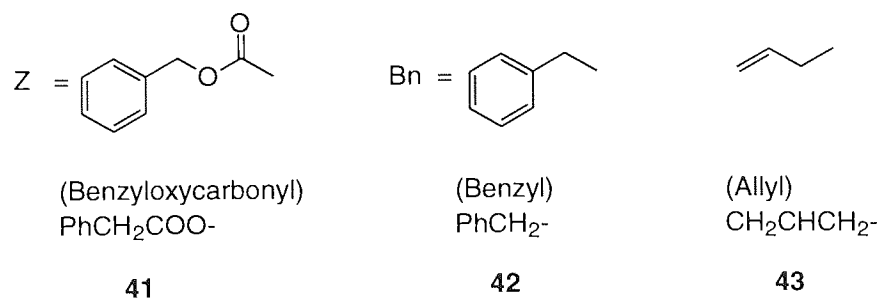
### 3.4 The Synthesis of PNA Monomers Composed of DNA Bases: Literature Precedents

At the outstart of this work, literature precedents reported the successful incorporation of the conventional DNA bases T **31**, C **32**, G **33** and A **34** into PNA monomers.<sup>4,5,231</sup> More recently, the incorporation of the non-natural, substituted pyrimidine J **35**, has been reported (Figure 3.4).<sup>183</sup>



**Figure 3.4.** Protected PNA building block monomers, composed of the bases (a) T, (b) C, (c) G, (d) A, and (e) J.

Figure 3.4 shows the three principal protecting groups used in DNA and PNA chemistry for the protection of the amino groups (Z), or for introducing the protected carbonyl functionalities, allyl or benzyl, at the appropriate sites of the heterocyclic bases.



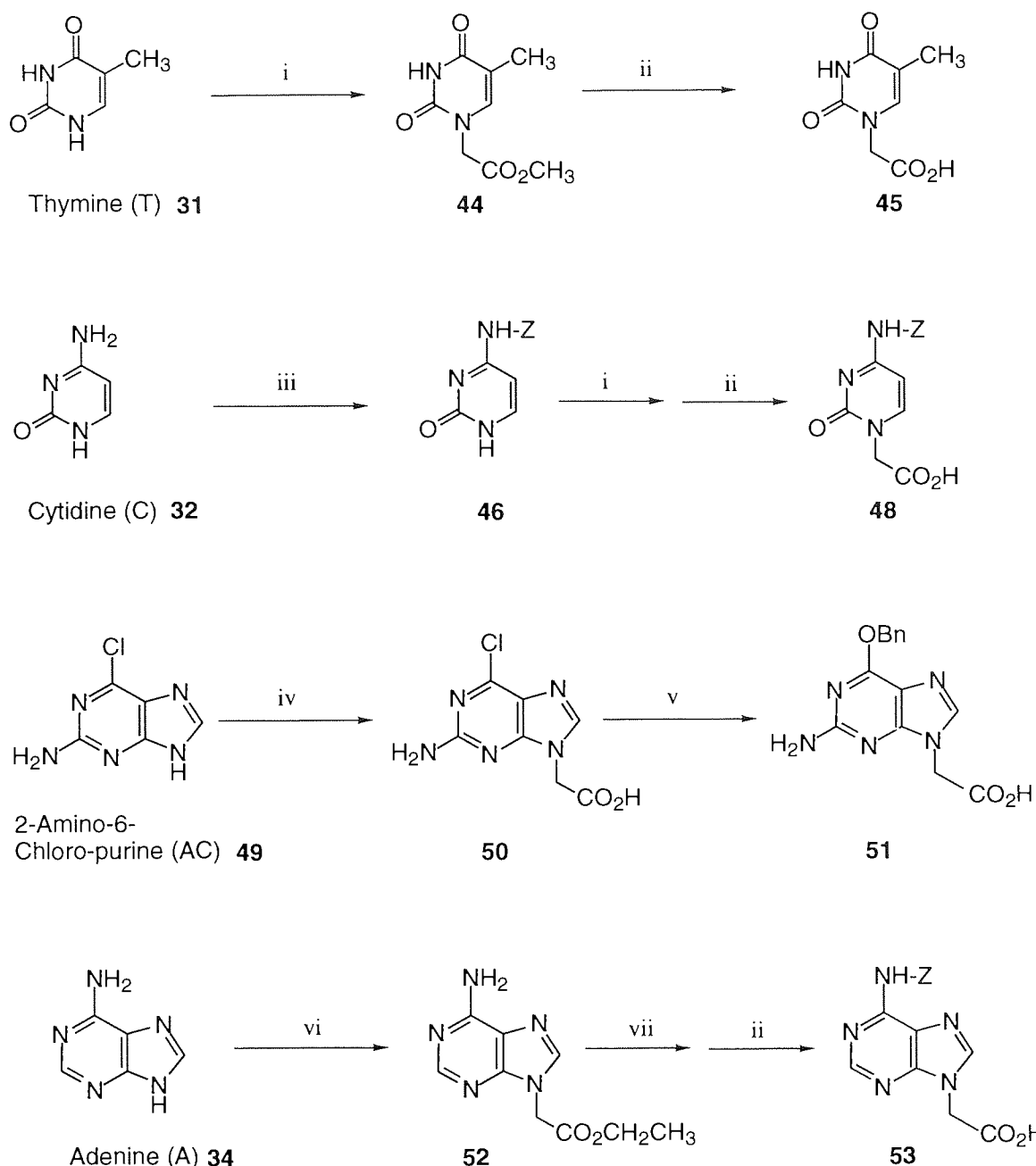
**Figure 3.5.** Benzyloxycarbonyl, benzyl and allyl protecting groups.

Two principal strategies have been employed in the synthesis of PNA monomers composed of DNA nucleobases (Scheme 3.2).<sup>218,219</sup> The first procedure (Method A), as reported by Dueholm *et al.*,<sup>218</sup> involves direct attachment of the protected *N*-carboxylated base to the *N*-(2-aminoethyl)glycine side-chain of the backbone component,<sup>227</sup> ethyl *N*-(Boc-aminoethyl)glycinate ester **1**.<sup>218</sup> The direct alkylation of T **31**, and Z-protected C **32** and A **34** using benzyloxycarbonyl chloride ( $\text{PhCH}_2\text{OCOCl}$ ), **54**, was followed by hydrolysis of the respective methyl, and ethyl esters with use of **1**, to afford the corresponding acids **45**, **48** and **53**.<sup>21,218,219</sup>

Acids **45**, **48**, and **53** were coupled to *N*-(2-Boc-aminoethyl)glycinate **1** using the activating agents dicyclohexylcarbodiimide (DCC) **55** and 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine (DhbtOH) **56**. Hydrolysis of **36**, **37**, and **39** using LiOH, yielded the respective PNA monomers, **57**, **58**, and **59** (Scheme 3.5). Condensation of G **60**, mediated by PyBroP **61** and diethylamine (DIEA) with the methyl **55**, rather than the ethyl ester of *N*-(2-Boc-aminoethyl)glycine **1**, was followed by hydrolysis with NaOH to afford **38** (Scheme 3.4).<sup>218</sup>

The second method (Method B, Scheme 3.2), reported more recently by Meltzer *et al.*,<sup>219</sup> is a potentially more general and direct route involving direct coupling of the common intermediate *N*-(chloroacetyl)-ethyl-*N*-(Boc-aminoethyl)glycinate ester **26**, to the unprotected purine and pyrimidine bases, T **31**, C **32**, A **33** and AC (2-amino-6-chloropurine) **49**, to yield the respective A, T, C, and AC PNA monomers, as depicted in Scheme 3.6.

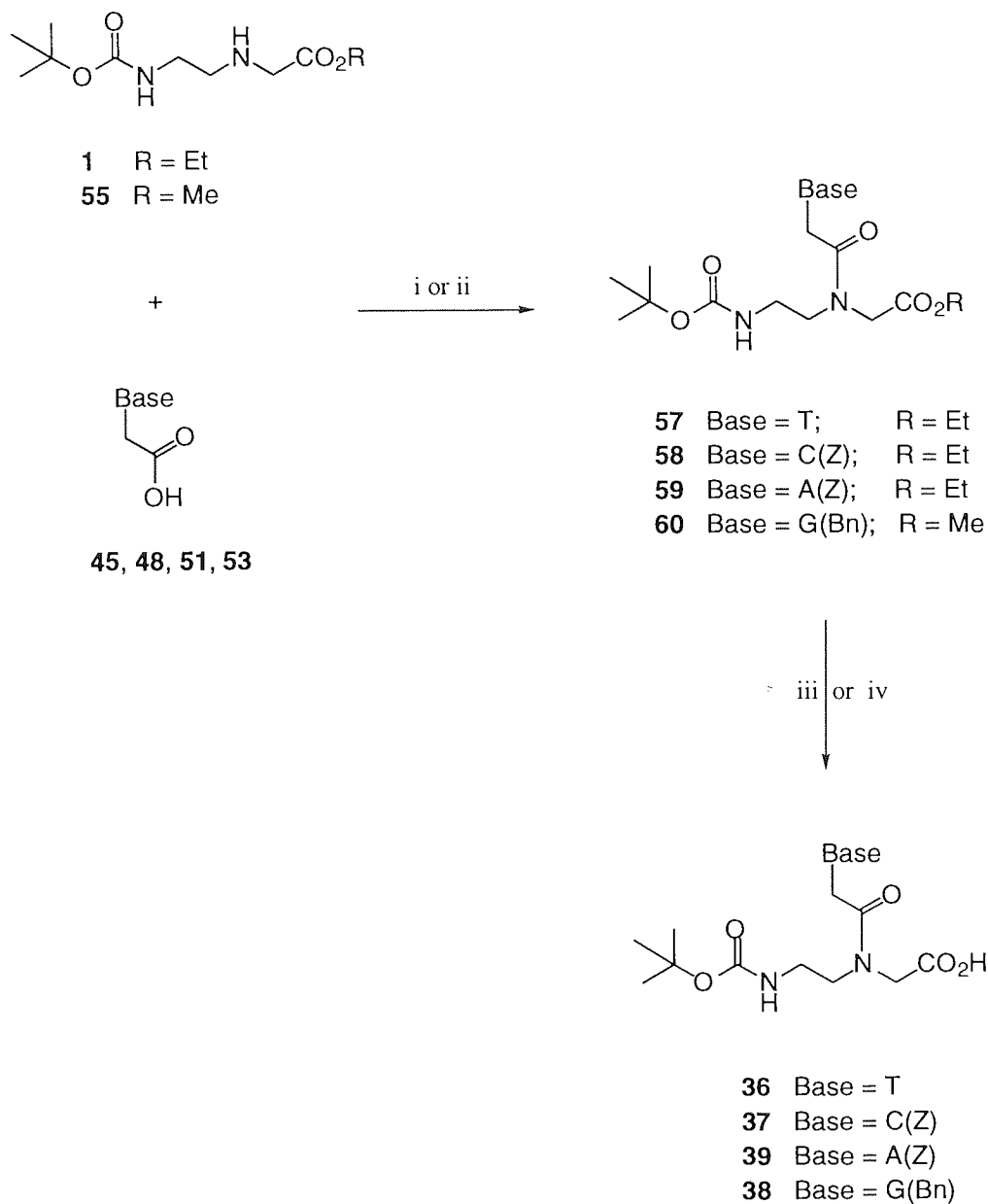
**26** was obtained via reaction of **1** in  $\text{CH}_2\text{Cl}_2$  with chloroacetyl chloride **28**, in the presence of triethylamine ( $\text{Et}_3\text{N}$ ), as shown in Scheme 3.2. Subsequent coupling of the purine bases **31-34** with **26** produced **57** and **62-64**. Hydrolysis of the **57**, **62** and **63** esters conducted with NaOH, gave the respective Boc-protected acids.



**Scheme 3.4.** Reagents: i,  $\text{BrCH}_2\text{CO}_2\text{CH}_2\text{CH}_3/\text{K}_2\text{CO}_3/\text{DMF}/\text{r.t.}/24 \text{ h}$ ; ii,  $4 \text{ M HCl (aq.)}/2 \text{ M NaOH (aq.)}/0 \text{ }^\circ\text{C (aq)}$ ; iii,  $\text{PhCH}_2\text{OCOCl}/\text{pyridine}/24 \text{ h}/\text{r.t.}/\text{NaOH (aq.)}$ ; iv)  $\text{BrCH}_2\text{CO}_2\text{H}/\text{K}_2\text{CO}_3/\text{DMF}/20 \text{ h}/\text{r.t.}$ ; v,  $\text{PhCH}_2\text{O-Na}^+/\text{130 }^\circ\text{C}/2 \text{ h}/\text{DMF}/1 \text{ M NaOH (aq.)}$ ; vi,  $\text{BrCH}_2\text{CO}_2\text{CH}_2\text{CH}_3/\text{NaH}/\text{DMF}/\text{r.t.}/24 \text{ h}$ ; vii,  $\text{PhCH}_2\text{OCOIm}^+\text{EtBF}_4^-/\text{DMF}/\text{MeOH}/2 \text{ M NaOH (aq.)}/\text{r.t.}$

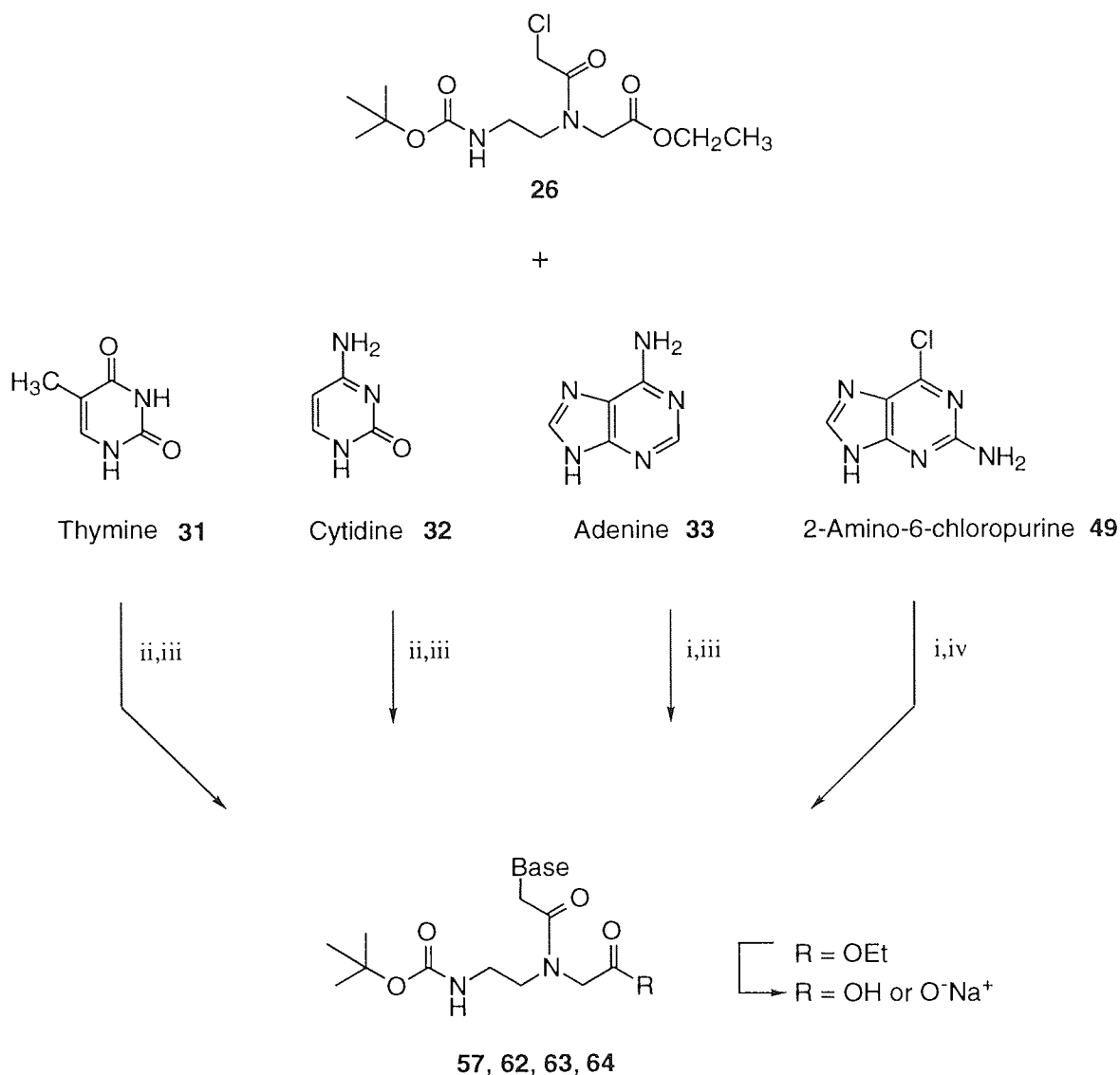
The alkylation of T **31**, and C **32** was shown to be regiospecific, whilst flash column chromatography was required for the successful isolation of the desired N9-regioisomer of A **33**.





**Scheme 3.5.** Reagents: i, DCC/DhbtOH; ii, PyBroP/DIEA; iii, LiOH/THF/H<sub>2</sub>O; iv, NaOH/EtOH/H<sub>2</sub>O.

Reaction of AC **49** with **26** in a solution of pulverised K<sub>2</sub>CO<sub>3</sub> and DMF using the procedure reported by Harden *et al*, furnished **64** (Figure 3.6).<sup>219</sup> The chloro substituent was transformed into the carbonyl group via reaction with 2.5 M NaOH followed by neutralisation using 4 M HCl to give **66** in 74% yield.

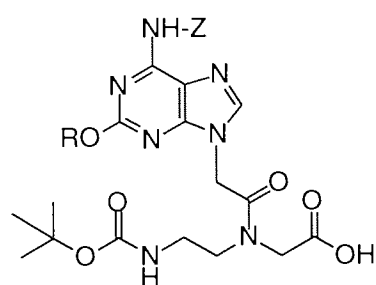


**Scheme 3.6.** Reagents : i, BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>/NaH/DMF/75 °C/1 h; ii, BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>/K<sub>2</sub>CO<sub>3</sub>/DMF/r.t./4 h; iii, 1 M NaOH/1 M HCl/4 °C; iv, 2.5 M NaOH/1 M HCl/4 °C.

### 3.5 The Synthesis of PNA Monomers Composed of DNA Bases: Current Strategies

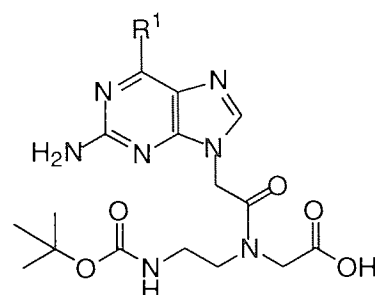
#### 3.5.1 PNA Monomers Composed of Novel Bases D, Dz, X, H and isoG

In this work, the desired PNA monomers of interest, composed of the novel bases D **2**, its precursor Dz **67**, X **3** and its analogue H **4**, and isoG **5**, as shown in Figure 3.6, were synthesised using the Boc/Z protecting group strategy.<sup>21,124</sup>



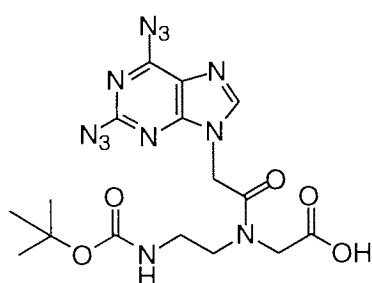
Boc-isoG(Z, R)-OH;

**68** R = Bn  
**69** R = Allyl

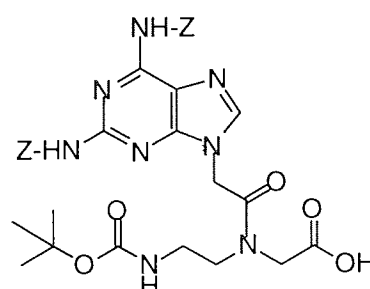


Boc-AC-OH

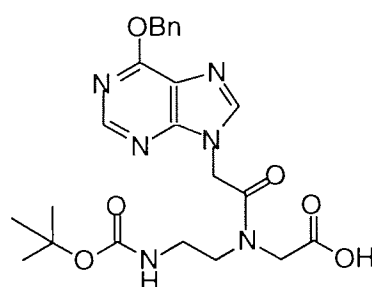
**64** R<sup>1</sup> = Cl; R<sup>2</sup> = CH<sub>2</sub>CH<sub>3</sub>  
**66** R<sup>1</sup> = OH; R<sup>2</sup> = H



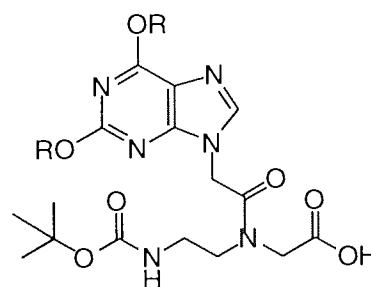
Boc-Dz-OH **70**



Boc-D(Z)-OH **71**



Boc-H(Bn)-OH **72**



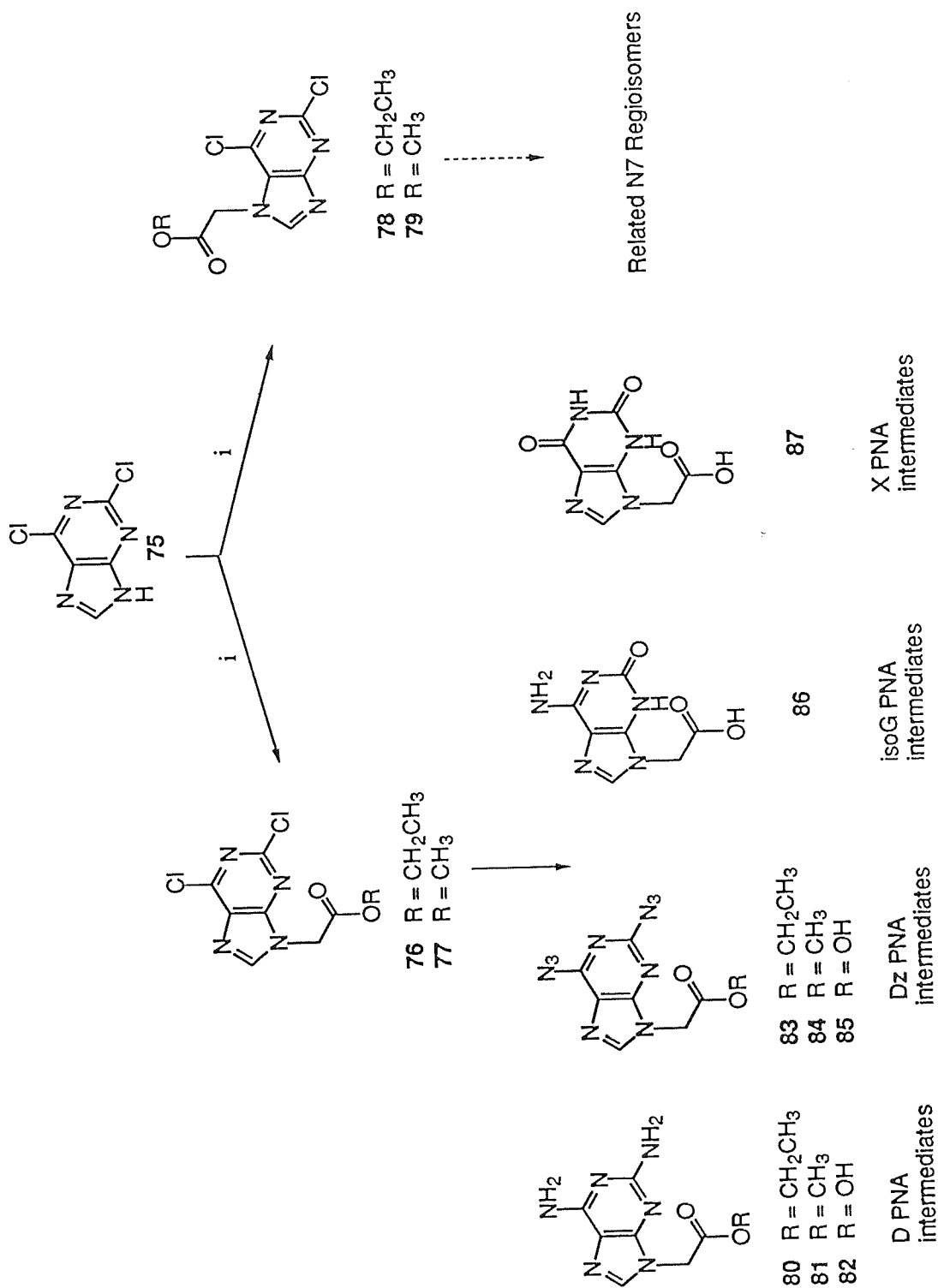
Boc-X(R)-OH

**73** R = Bn  
**74** R = Allyl

**Figure 3.6.** Boc-protected PNA monomers composed of the novel isoG, D, Dz, X and H purine bases.

### 3.5.2 The Versatility of 2,6-Dichloropurine 75 as a Starting Material

2,6-dichloropurine **75** was selected as a starting material for four of the five target bases, due to its established versatility.



**Scheme 3.7.** Reagents: *i*,  $\text{BrCH}_2\text{CO}_2\text{CH}_2\text{CH}_3$  or  $\text{BrCH}_2\text{CO}_2\text{CH}_3/\text{K}_2\text{CO}_3/\text{anhydrous CH}_3\text{CN}/\text{r.t.}/24 \text{ h}$ .

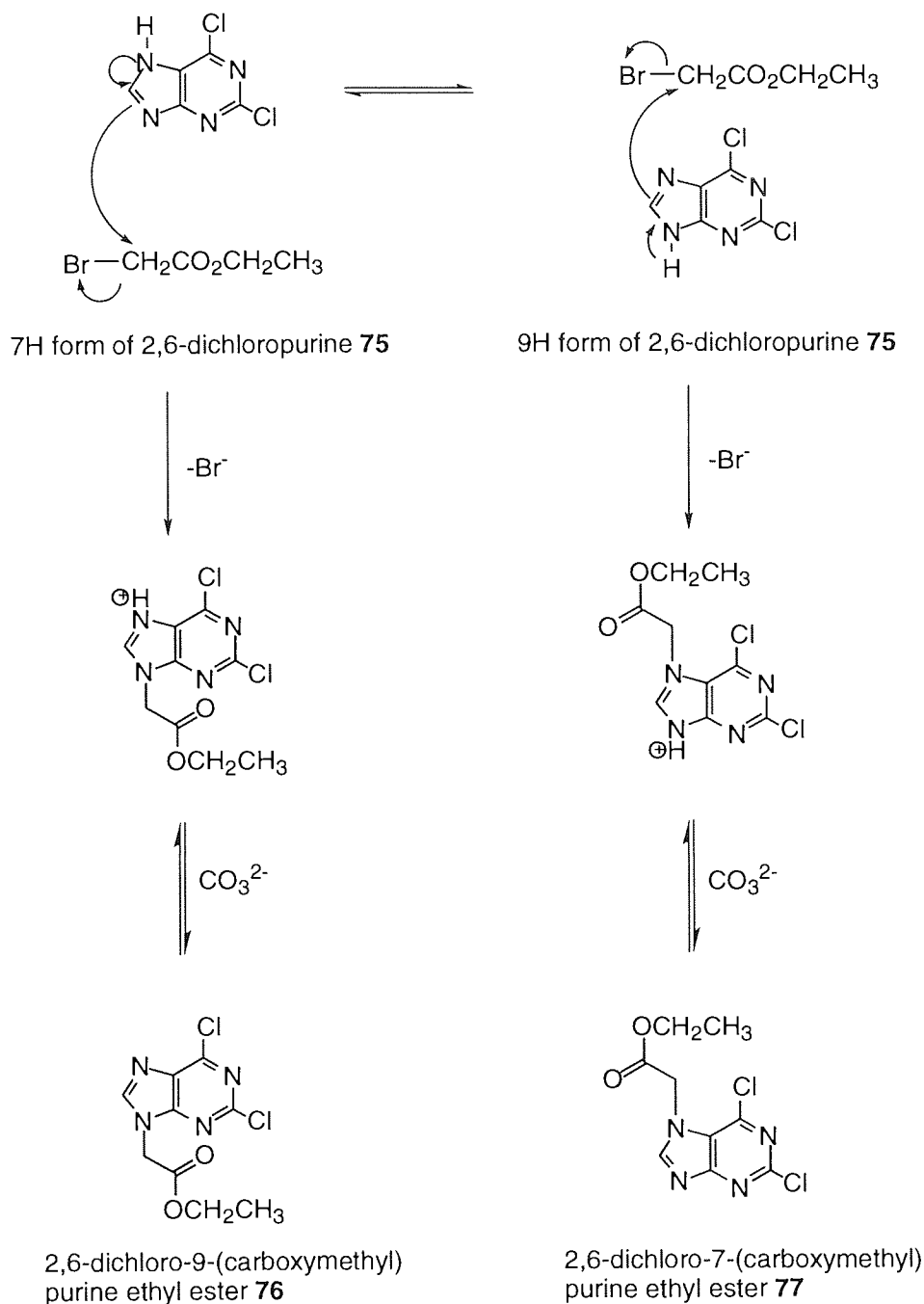
The proposed synthetic route (Scheme 3.7) focused upon alkylation of 2,6-dichloropurine **75** to yield the functionalised 2,6-dichloro-9-(carboxymethyl)purine ethyl and methyl esters, **76** and **77** respectively, as key intermediates towards the synthesis of each of the target PNA building block intermediates **82**, **85**, **86** and **87** (Scheme 3.7).

As base alkylation in the PNA series has been reported with use of the methyl **88** and also of the ethyl **27** ester of bromoacetate, the alkylation of **75** was undertaken with both **88** and **27**, respectively, to probe any differences in yield of product. Direct alkylation of **75** with  $\text{BrCH}_2\text{CO}_2\text{CH}_2\text{CH}_3$  **27** as the alkylating agent and  $\text{K}_2\text{CO}_3$  as the weak base, produced a mixture of two regioisomeric products 2,6-dichloro-9-(carboxymethyl)purine ethyl ester **76** and 2,6-dichloro-7-(carboxymethyl)purine ethyl ester **78**. Due to the similarity in  $R_f$  values of the N9 and N7 regioisomers **76** and **78**, their separation via flash column chromatography proved fairly inefficient. However, recrystallisation from MeOH enabled isolation of the N9 isomer as pale yellow crystals in usable yields of 51%. The N7 regioisomer was obtained from the mother liquor as a golden brown viscous oil (32%). X-ray crystal structure analysis of the N9 regioisomer **76**<sup>232</sup> with inspection of spectroscopic analyses, provided unambiguous proof regarding regioisomeric assignment of **76** and **78**. Alkylation of **75** with methyl bromoacetate **88** under similar experimental conditions furnished the N9 **78** and N7 **79** methyl esters in 39% and 28% yields, respectively. Subsequent reactions were undertaken with the crystalline ethyl ester of **76**.

Regioselective substitution for *N*-alkylation of purine bases depends upon the choice of heterocycle, the leaving group, the alkylating agent, the strength of the base and the reaction temperature.<sup>235</sup> The presence and the disposition of substituted alkyl groups already attached to the heterocycle also exert a prominent influence upon the site of substitution.<sup>234,235</sup> Selective alkylation of a broad series of purines has rarely been observed to be regioselective; alkylation of purine bases tends to produce mixtures of N9- and N7-substituted products with predominant yields of the N9 isomer.<sup>234-236</sup> Literature precedents regarding synthetic efforts towards the preparation of an N9 regioisomer in the absence of its N7 regioisomer, have been predominantly unsuccessful. One exception however, involves the transient protection of the N9 site of both 2,6-dichloropurine **75** and 6-chloropurine **89** with methylcobaloxime, to produce alkylated N7 adducts in excess of 50%, with smaller quantities of the N9 regioisomer.<sup>237</sup>

A feasible proposal regarding the mechanism of substitution in the presence of a weak base suggests direct attack occurs upon the *N* site which is more open to attack, that is where there is no proton present, to afford a cationic intermediate. Subsequent loss of the acidic proton at that *N* site is thought to furnish the respective N9- or N7-substituted regioisomer. Hence, from this mechanism, the final distribution ratio of N9:N7 substituted 2,6-dichloropurine **78** is inversely dependent upon the initial N9:N7 distribution ratio of

unsubstituted 2,6-dichloropurine **75** (Figure 3.7).



**Figure 3.7.** Proposed mechanism for substitution of 2,6-dichloropurine **75** in the presence of the weak base  $\text{K}_2\text{CO}_3$ .

Investigations by Kjelhberg and Johansson *et al*<sup>236</sup> regarding the reactivity and regioselectivity of modified purines towards alkyl halides showed the predominant formation of the N9 isomer compared to the N7 isomer. One such example was provided by reaction of 6-chloroguanine with 1-bromobutane to afford N9:N7 proportions in a 4:1 ratio, as judged by  $^1\text{H}$  NMR,<sup>236,238</sup> in accordance with our results. Also in parallel with our findings, a typical ratio for the isomers was observed, where the distribution was governed by the nature of the substituent at the C6 site of the heterocycle.

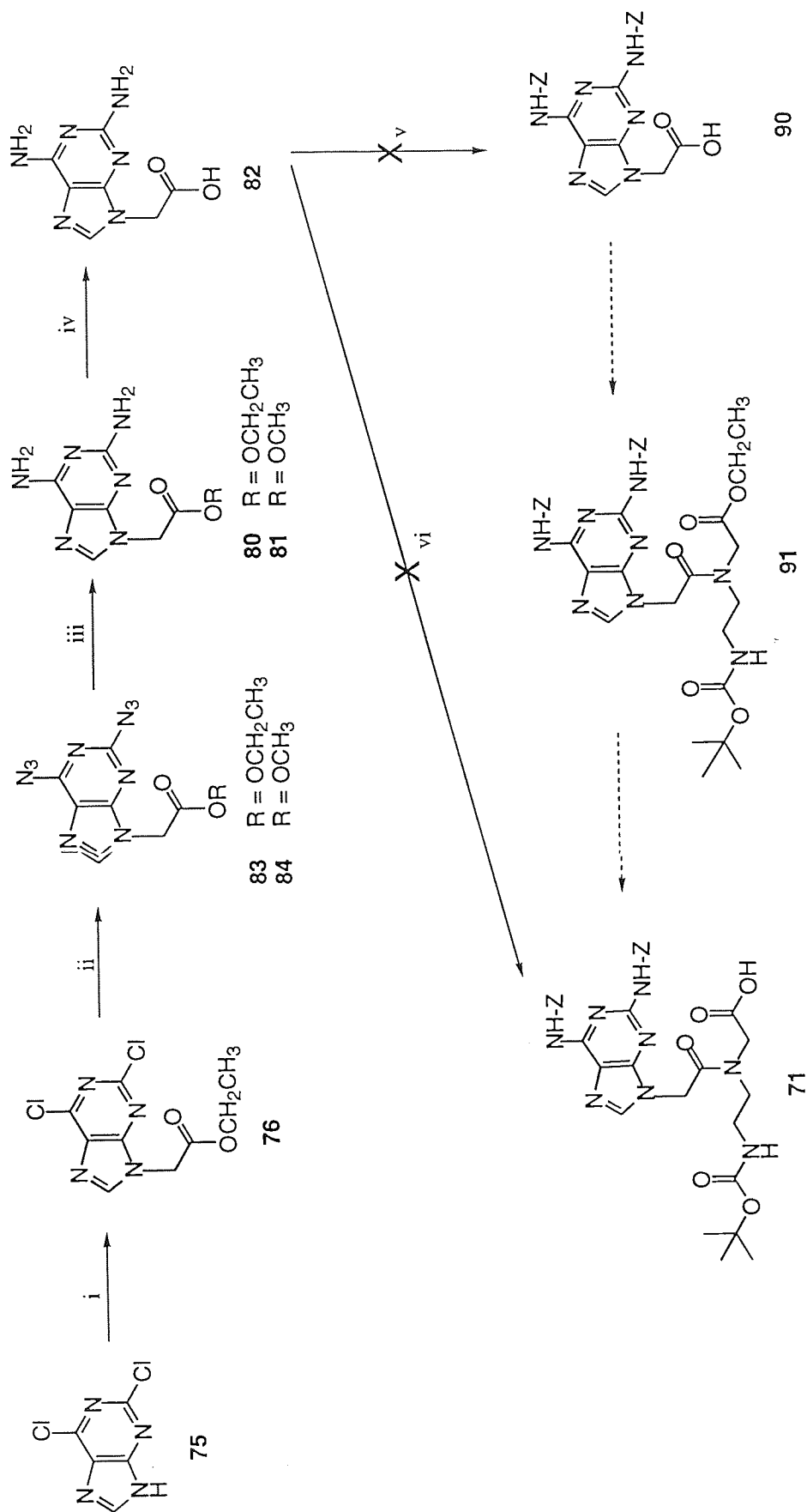
### 3.6 Synthesis of the PNA Monomer Incorporating the Novel Base 2,6-Diaminopurine (D) 82

#### 3.6.1 Synthesis of the PNA Building Block Incorporating Diaminopurine 2 via its Diazidopurine Precursor 84

The attempted synthesis of the desired 2,6-diaminopurine PNA monomer **71** made use of the versatile 2,6-dichloropurine **75**. Azidation of **75** to give the intermediate 2,6-diazido-9-(carboxymethyl)purine ethyl ester **83**<sup>239</sup> with ensuing hydrogenolysis furnished 2,6-diamino-9-(carboxymethyl)purine ethyl ester **80**<sup>240</sup> (Scheme 3.8).

The successful substitution of both chloro groups on **75** with azido moieties was executed via treatment of **75** with sodium azide (NaN<sub>3</sub>), in the presence of excess MeOH-acetone (4:1). TLC and NMR analysis showed the crude mixture to contain three compounds. Additionally, from mass spectral analysis it appeared likely that the methyl ester of the diazido homologue **84** had formed, presumably through transesterification of the ethyl ester homologue of diazidopurine **83**.<sup>241</sup> Slow crystallisation of the crude product mixture from anhydrous CH<sub>3</sub>CN over a period of several days, resulted in the successful isolation of a white crystalline product which, from spectroscopic analysis, indicated the acquirement of the methyl ester of the diazido analogue **84**. Structural confirmation of **84** was achieved by X-ray crystallographic analysis (chapter 5) providing unambiguous proof for the methyl, rather than the ethyl ester of **84**.<sup>241</sup> Foundering of repeated attempts, using flash column chromatography and X-ray crystallography, to isolate the ethyl ester of **83** was attributed to the close proximity of the *R<sub>f</sub>* values. This compelled us to focus upon **84**, successfully isolated in its pure crystalline form. However, hydrogenolysis of the azido mixture without separation of the components was undertaken, followed by recrystallisation from EtOAc-MeOH, to observe if the resulting amino products could be more easily separated than their respective diazido precursors. This reaction afforded 2,6-diamino-9-(carboxymethyl)purine ethyl ester as pale yellow crystals in 39% yield,<sup>239</sup> concurrent with the by-product 6-amino-9-(carboxymethyl)-2-methoxypurine methyl ester **92** as pale white crystals.<sup>242</sup> As spectroscopic characterisation of **92** was inconclusive, X-ray crystallographic analysis was undertaken, and provided proof for structural determination of **92**.

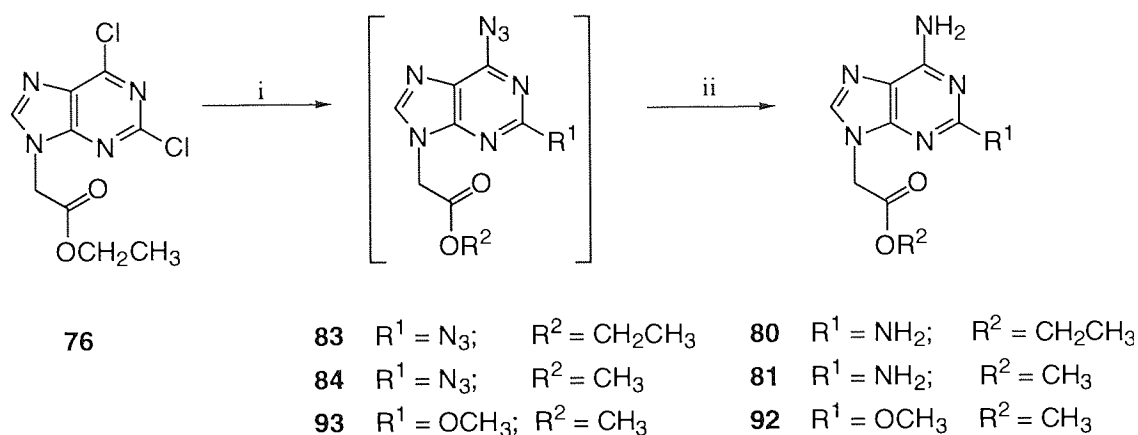
Both the methyl ester analogues of 2,6-diazido-9-(carboxymethyl)purine **84** and 6-azido-9-(carboxymethyl)-2-methoxypurine **93** and the ethyl ester analogue of 2,6-diazido-9-(carboxymethyl)purine **83** were attained during azidation of **76**.



**Scheme 3.8.** Reagents: i, BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>/K<sub>2</sub>CO<sub>3</sub>/anhydrous CH<sub>3</sub>CN/r.t.; ii, NaN<sub>3</sub>/acetone-MeOH/75 °C/72 h; iii, H<sub>2</sub>/10% Pd on charcoal/EtOH/72 h/r.t.; iv, 2 M NaOH/4 M HCl/20 min/reflux; v, *N*-(benzyloxycarbonyl)-*N'*-ethylimidazolium tetrafluoroborate/24 h/0 °C; vi, ethyl *N*-(Boc-aminoethyl)glycinate/DCC/Dhbt(OH)/DMF/CH<sub>2</sub>Cl<sub>2</sub>/24 h/r.t.; v, H<sub>2</sub>/10% Pd on charcoal/EtOH/r.t.



Hydrogenolysis of the azido product mixture furnished 2,6-diamino-9-(carboxymethyl)purine ethyl ester **80**, 2,6-diamino-9-(carboxymethyl)purine methyl ester **81** and 6-amino-9-(carboxymethyl)-2-methoxypurine methyl ester **92**<sup>242</sup> (Scheme 3.9).

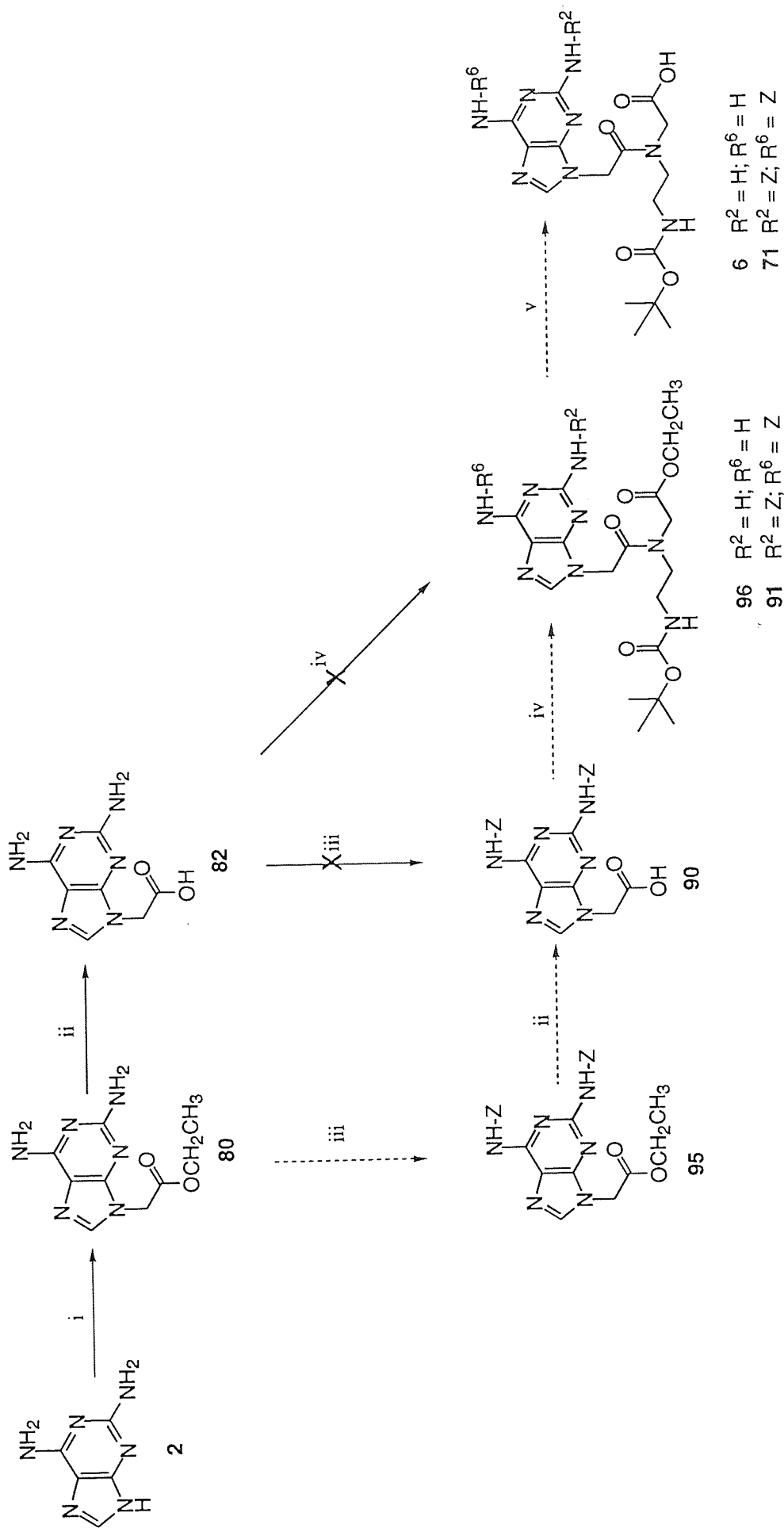


**Scheme 3.9.** Reagents: i,  $NaN_3/MeOH/-Me_2CO/75\text{ }^\circ C/72\text{ h}$ ; ii,  $H_2/10\% Pd\text{ on charcoal}/72\text{ h}$ .

Subsequent azidation reactions were thus conducted in EtOH to avoid the transesterification step which occurred with use of excess MeOH. Use of EtOH provided the ethyl ester **83** in 29% yield, exclusively. Ensuing reduction of **81**, via catalytic hydrogenolysis under an atmosphere of  $H_2$ , was conducted during four days at r.t. using 10% Pd on charcoal. The desired 2,6-diaminopurine derivative **80** was isolated in 43% yield.<sup>239</sup> The acetic acid derivative of 2,6-diamino-9-(carboxymethyl)purine **82** was obtained as a pale yellow powder in 36% yield from saponification of **80** using 2 M NaOH, and precipitation with 4 M HCl (Scheme 3.10).

### 3.6.2 Synthesis of the PNA Building Block **82** Incorporating 2,6-Diaminopurine via Direct Alkylation of 2,6-Dichloropurine **75**

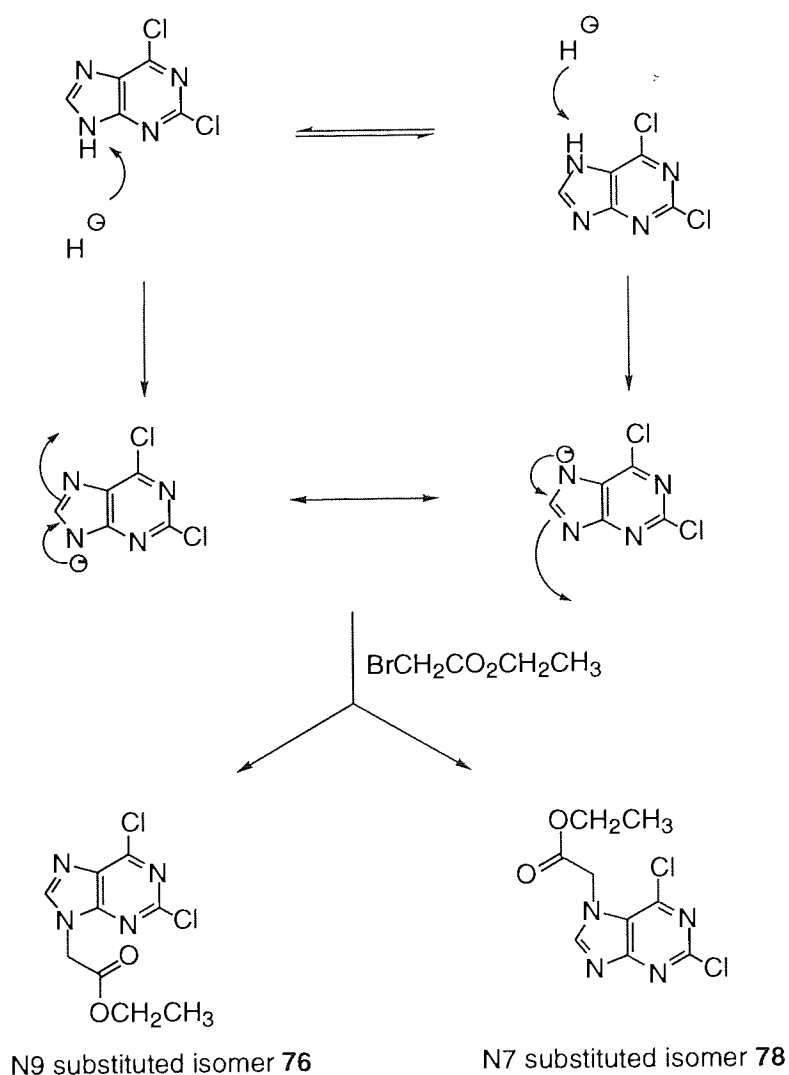
The above method generated side-products which resulted in reduced yields of the desired component **82** and proved problematic for chromatographic isolation of the desired intermediate **83**. Despite the versatility of **75**, an alternative, higher yielding method, with elimination of by-products, was deemed necessary (Scheme 3.10). Furthermore, the potential danger associated with this reaction attributed to the use of refluxing organic solvents containing  $NaN_3$ , necessitated an alternative route for the direct synthesis of D-containing PNA building block **82**. Inspection of PNA literature precedent revealed the direct alkylation of A **34** with use of a strong base, afforded the N9 alkylated adenine **94** regioisomer in excellent yield, but with little or none of the corresponding N7 regioisomer.<sup>218</sup> NaH was used during alkylation to deprotonate the nucleophilic centres on 2,6-diaminopurine **2** for adduct formation. Reaction of the anionic intermediate with  $BrCH_2CO_2CH_2CH_3$  **27** at r.t. in DMF produced the ethyl ester product **80**.



**Scheme 3.10.** Reagents: i,  $\text{BrCH}_2\text{CO}_2\text{CH}_3/\text{K}_2\text{CO}_3/\text{anhydrous CH}_3\text{CN}/\text{r.t.}$ ; ii, 2 M NaOH/4 M HCl/20 min/reflux; iii, *N*-(benzyloxycarbonyl)-*N'*-ethylimidazoliumtetrafluoroborate/24 h/0 °C; iv, ethyl *N*-(Boc-aminoethyl)glycinate/DCC/DhbtOH/DMF/ $\text{CH}_2\text{Cl}_2$ /24 h/r.t.; v,  $\text{H}_2$ /10% Pd on charcoal/EtOH/r.t.

Compound **80** was isolated as a pale orange powder in 79% yield with none of the corresponding N7 isomer. Subsequent saponification of **80** successfully produced the desired **82** as a pale yellow powder in 70% yield (Scheme 3.10).

The highly basic nature of alkoxide and hydroxide anions in the presence of a strong base implies the favoured reaction for heterocyclic substitution with purines, entails proton abstraction of the protonated N9 or N7 site (Figure 3.8). Increased susceptibility of the activated anion is proceeded by nucleophilic attack from the incoming electron-deficient ethyl ethanoate cation, in accordance with the  $S_N1$  mechanism. This affords the respective N9- or N7-substituted regioisomer. Resonance forms of the activated 2,6-diaminopurine anion may be drawn where the negative charge is located on either the N9 or the N7 site (Figure 3.8).



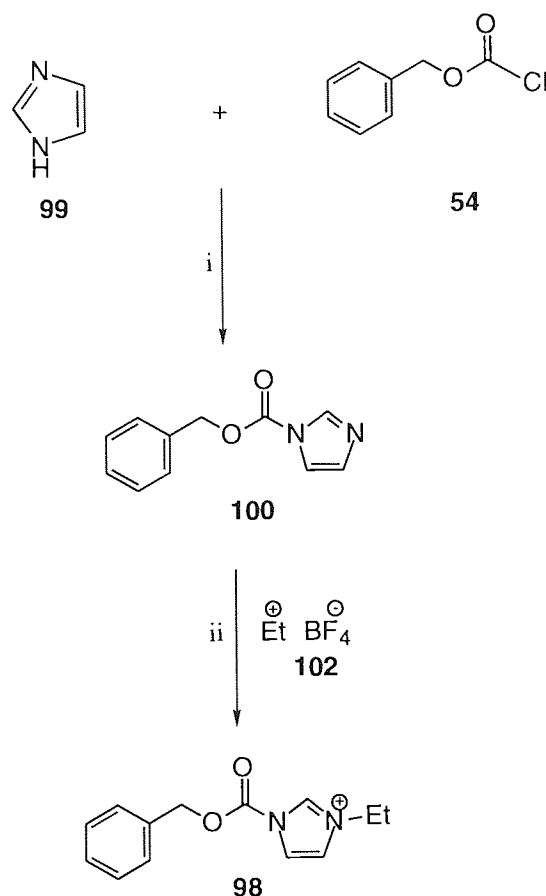
**Figure 3.8.** Proposed mechanism for the alkylation of 2,6-dichloropurine **75** in the presence of the strong base NaH.

### 3.6.3 Protection of the Exocyclic Amino Groups of the 2,6-Diaminopurine Derivative **82**

The heterocyclic bases of PNA require protection of exocyclic amino groups during automated synthesis to alleviate undesired reactions.<sup>218,219</sup> The primary amino group of the base to be coupled to the carbonyl group of the PNA backbone experiences competition during coupling from unprotected amino groups. This consequences in decreased yields, but more importantly in the formation of an array of unwanted side-reactions, including chain extension from this position to give complex product mixtures which are impossible to separate and purify.

In the DNA series, acetyl and benzoyl moieties remain favourable amino protecting groups for A and C, and isobutyryl for G. In the PNA series, the benzyloxycarbonyl (Z) **41** moiety prevails as the more appropriate base protecting group. During cleavage of Z-protected bases from the solid support following PNA oligomer synthesis, concomitant removal of Z groups, under those conditions normally applied in peptide chemistry, is accomplished using hydrogen fluoride;<sup>218</sup> mild conditions which allow selective, facile and efficient removal, with no adverse affect upon other bonds, in particular the peptide bonds. Though the use of the Z group is specifically used for amino group protection,<sup>243</sup> **41** is also used for those bases displaying poor solubility properties to increase the lipophilicity of PNA monomers, thus bestowing increased solubility in organic solvents compared with the corresponding unprotected analogues. Introduction of the Z moiety **41** is effected via reaction of the amino group with benzyl chloroformate **97** and sodium hydroxide, as in the case for the C PNA monomer **37**, whilst Rapoport's reagent; *N*-(benzyloxycarbonyl)-*N'*-ethylimidazoliumtetrafluoroborate **98**, is employed for protection of PNA monomers containing A. Rapoport's reagent **98** was synthesised according to Scheme 3.11.<sup>243</sup>

Attempted reaction of **82** with **98** failed to give **90** (Scheme 3.10) as analysed by spectroscopic data. The absence of a multiplet, characteristic of a benzyl group, was attributed to the low reactivity of the amino groups to undergo reaction with both Rapoport's reagent **97** and also the alternatively used benzyl chloroformate **98**. Furthermore, the significantly poor solubility exhibited by **82** was thought to be the major factor contributing to the resistance of the amino groups to react with these reagents. The failure to protect the amino groups of **82** raised suspicions that the C2 and C6 amino groups would not be sufficiently reactive during PNA synthesis to undergo undesired reactions. Hence, the alternative possibility for using **82** in its unprotected form was considered.

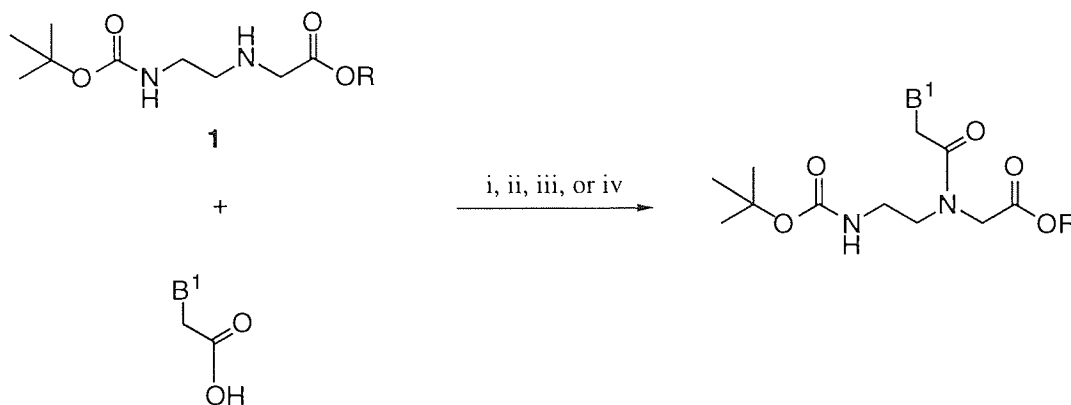


**Scheme 3.11.** Reagents: i, toluene/0 °C; ii, Et<sup>+</sup>BF<sub>4</sub><sup>-</sup>/CH<sub>2</sub>Cl<sub>2</sub>/0 °C/4 h.

### 3.6.4 Coupling of PNA Building Blocks to ethyl *N*-(Boc-aminoethyl) glycinate **1**: Literature Precedents

Synthesis of PNA monomers proceeds via coupling of the nucleobase acetic acid derivative with the protected backbone **1**. At the advent of PNA monomer synthesis, the first homothymine oligomers were assembled from activated pentafluorophenyl (pfp) ester monomers using the glycine of **1**.<sup>4</sup> However, recent literature precedents reports *in situ* activation with the glycinate of **1** using DCC **55** and DhbtOH **56** for improved synthesis of T, and C and A (*Z*)-protected building blocks **57**, **58** and **59**.<sup>218,219</sup> Hydrolysis with LiOH converts the glycinate functionality to the glycine, to furnish the desired PNA monomers **36**, **37** and **39**. The G monomer **38** is prepared in its protected form G-Bn, to facilitate solubility for coupling **60** with the backbone component **1** using PyBrop **61**. Hydrolysis using aqueous NaOH affords the desired G PNA monomer **38**.

For synthesis of PNA monomers **57-60**, documented literature reports the successful use of the coupling agents DCC **55** with DhbtOH **56** for T **31**, C **32** and A **34**, and PyBrop **61** for G **33**. *N,N'*-di-isopropylcarbodiimide (DIPCDI) has recently found use as a more potent coupling reagent. Carpino *et al*<sup>244</sup> have reported the inclusion of uronium salts to increase coupling yields and reaction kinetics. Coupling of the building block bases to **1** is shown in Scheme 3.12.

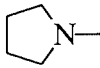
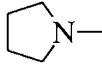
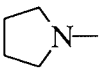
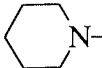


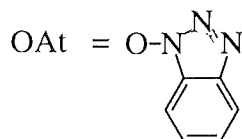
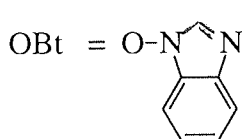
**Scheme 3.12.** Reagents. i, DCC/DhbtOH ( $B^1 = T, C\text{-Z}, \text{ or } A\text{-Z}$ ); ii, PyBroP/DIEA ( $B^1 = G\text{-O}^6\text{-Bn}$ ;  $R = \text{Me}$ ); iii, LiOH/THF/ $\text{H}_2\text{O}$  ( $B^1 = T, C\text{-Z}, \text{ or } A\text{-Z}$ ;  $R = \text{Et}$ ); iv, NaOH/EtOH/ $\text{H}_2\text{O}$  ( $B = G\text{-O}^6\text{-Bn}$ ;  $R = \text{Me}$ ).

Efficient amide bond formation between the base and the peptidic backbone **1** under mild conditions, principally to eliminate further side-reactions, necessitates chemical activation of the carboxyl component upon the *N*-protected nucleobase. Though dicyclohexylcarbodiimide (DCC) **55** provides one of the highest levels of activation, its principal drawback is the well-documented formation of the insoluble *N*-acylurea, dicyclohexylurea (DCU), during activation and acylation.

Generally, in amide bond formation of peptides, the coupling agent BOP which produces the toxic chemical hexamethylphosphoric triamide (HMPA) has predominantly been replaced by BroP and PyBroP (benzotriazolyl-oxo-tris[pyrrolidino]-phosphonium hexafluorophosphate derivatives (Table 3.1). PyBroP, an analogue of BOP, displays similar coupling rates but with the absence of HMPA formation and unlike DCC, is not hygroscopic nor moisture sensitive. Hence, where reactions failed with employment of DCC, PyBroP was chosen as the alternative coupling agent. Coupling additives, *in situ* activating agents, are predominantly uronium or phosphonium salts based on 1-hydroxy-7-azabenzotriazole (HOAt) or 1-hydroxybenzotriazole (HOBT).<sup>244</sup> Used in conjunction with an active ester and carbodiimide, such as HBTU and HATU,<sup>245</sup> (Table 3.1) their observed advantages for coupling of hindered amino acids include reduced racemisation and increased acylation rates. More importantly, *N*-acyl urea formation is suppressed via rapid reaction of the base with the highly reactive *O*-acyl urea to form the active ester, which then reacts with DhbtOH to yield the desired ester.

**Table 3.1.** Possible combinations of coupling reagents for coupling **1** to protected building blocks.

Phosphonium salts			Uronium salts		
	$\begin{array}{c} \text{R} \\ \text{R} \diagup \text{P}^{\oplus} \text{-X} \\ \text{R} \end{array}$	$\text{PF}_6^{\ominus}$		$\begin{array}{c} \text{R} \\ \text{R} \diagup \text{C}^{\oplus} \text{-X} \\ \text{R} \end{array}$	$\text{PF}_6^{\ominus}$
R	X	Name	R	X	Name
$(\text{CH}_3)_2\text{N}$	OBt	BOP	$(\text{CH}_3)_2\text{N}-$	OBt	HBTU
$(\text{CH}_3)_2\text{N}$	Br	BrOP	$(\text{CH}_3)_2\text{N}-$	OAt	HATU
	OBt	PyBOP		OBt	HBPyU
	OBt	PyBOP		OBt	HBPIP-U



### 3.6.5 Attempted Coupling of the Diaminopurine Building Block **82** to ethyl *N*-(Boc-aminoethyl) glycinate **1**

Coupling of the nucleobase **82** to the peptidic backbone **1** (Scheme 3.10) was attempted using various sets of experimental conditions, as shown in Table 3.2. Where DCC **55** was used, the general procedure involved addition of the backbone component **1** dissolved in DMF, to a stirred solution of DCC **55**, the acetic acid derivative of the nucleobase, and HOBT or DhbtOH, in  $\text{CH}_2\text{Cl}_2$  at  $0^\circ\text{C}$ . DMF was deemed necessary for the solubility of the building block, as **82** showed very poor dissolving properties with use of  $\text{CH}_3\text{CN}$ . Even with use of the DMF solvent, **82** was observed to precipitate and not fully dissolve. Variation to the duration, temperature, solvent, and concentration of the reaction, failed to increase the solubility of **82**. Continuous monitoring of the reaction via TLC analysis was observed to display only starting materials, with no newly formed spots. Overnight stirring at r.t. with filtering to remove the by-product DCU, was followed by solvent extraction which gave a yellow powdery residue. NMR analysis and spectroscopic data

failed to reveal the presence of the expected peaks, including the presence of the expected amide bond.

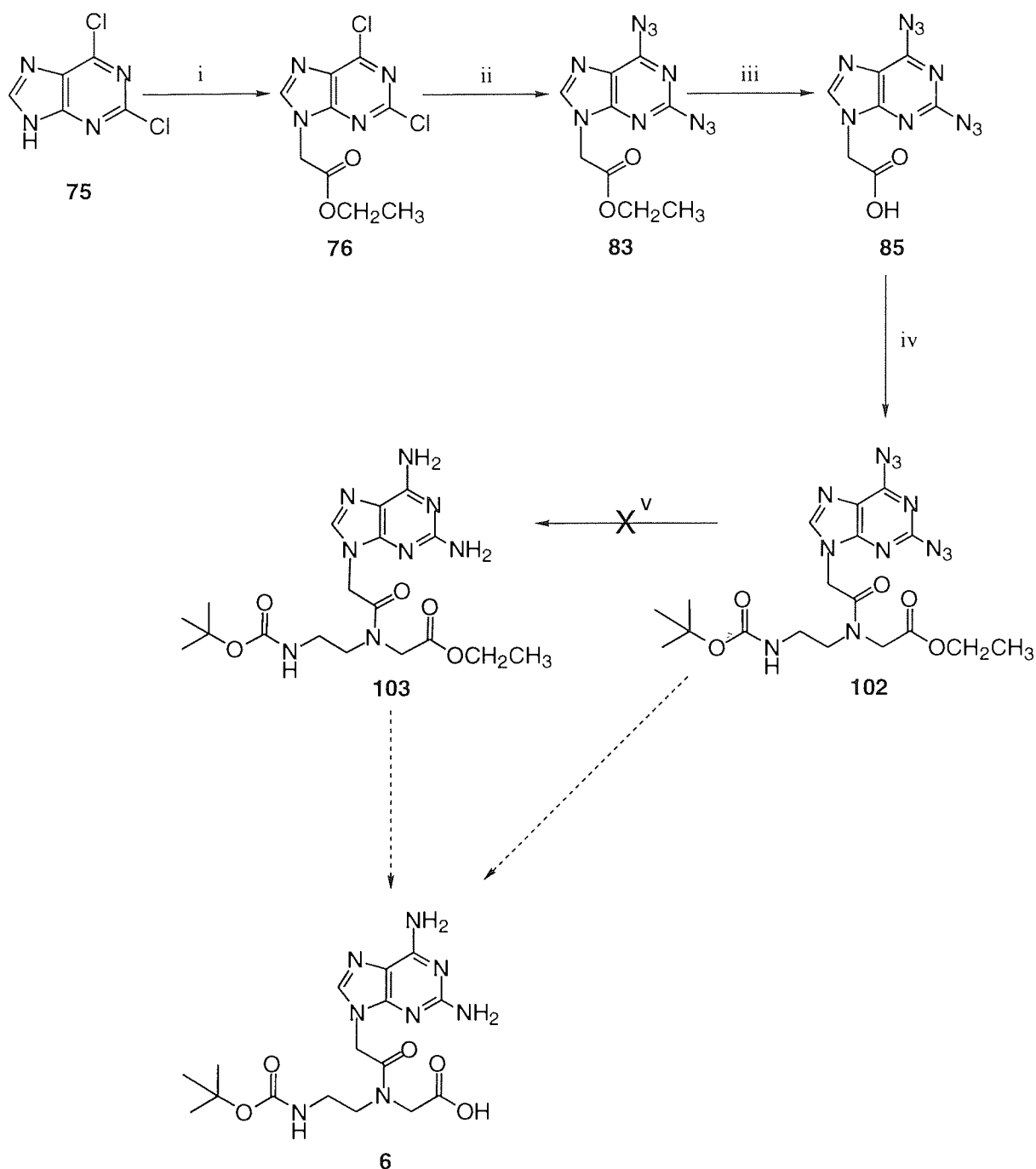
Alternative coupling and activation agents, including the application of PyBrop **61** and DMAP in place of DCC **55** and DhbtOH **56**, with similar experimental conditions using DMF as the solvent, also failed to couple **82** to **1**. Reluctance of the carboxyl moiety of **82** to react with the amino group on **1** to afford the amide product **96**, was attributed to the poor solubility of unprotected **82**, with subsequent failure to react with **1** to produce the desired PNA monomer **96**. Further inspection of NMR spectra showed the presence of the starting material **55**, but none of the peptidic backbone **1**. Thus, it was not possible to attempt the final step for conversion of **96** to **6**, as shown in Scheme 3.10. Acquisition of the D monomer **96** thus required either increased solubility of the D building block **82** for successful reaction with **1**, or use of an alternative route.

**Table 3.2.** Experimental conditions employed to attempt successful coupling of **1** to **82**.

Coupling Reagent	Coupling Agent	Temp. (°C)	Catalyst	Base
DCC	HOBt	r.t.	-	DIPEA
DCC	DhbtOH	0	DMAP	DIPEA
DCC	HOBt	0	-	DIPEA
DCC	DhbtOH	r.t.	DMAP	-
DCC	HOBt	0	DMAP	-
PyBrop	-	r.t.	DMAP	-
PyBrop	DhbtOH	r.t.	-	Et <sub>3</sub> N
PyBrop	DhbtOH	r.t.	DMAP	-
PyBrop	DhbtOH	r.t.	DMAP	Et <sub>3</sub> N
PyBrop	DhbtOH	r.t.	DMAP	DIPEA

Hydrogenolysis of **102**, following the coupling of **1** with **85** may provide an alternative approach towards the synthesis of **6** (Scheme 3.13). Therefore, coupling of the diazido building block **85** to **1** was attempted under conditions analogous to those used with DCC **55** for the attempted coupling of the 2,6-diaminopurine building block **82** and **1**. Hence, to the acetic acid of the diazidopurine compound **85** in DMF was added the peptide **1** with **55**, and the catalyst DMAP with the activating agent DhbtOH **56**. Compound **85** was observed to exhibit significantly superior solubility in DMF compared with **82**.





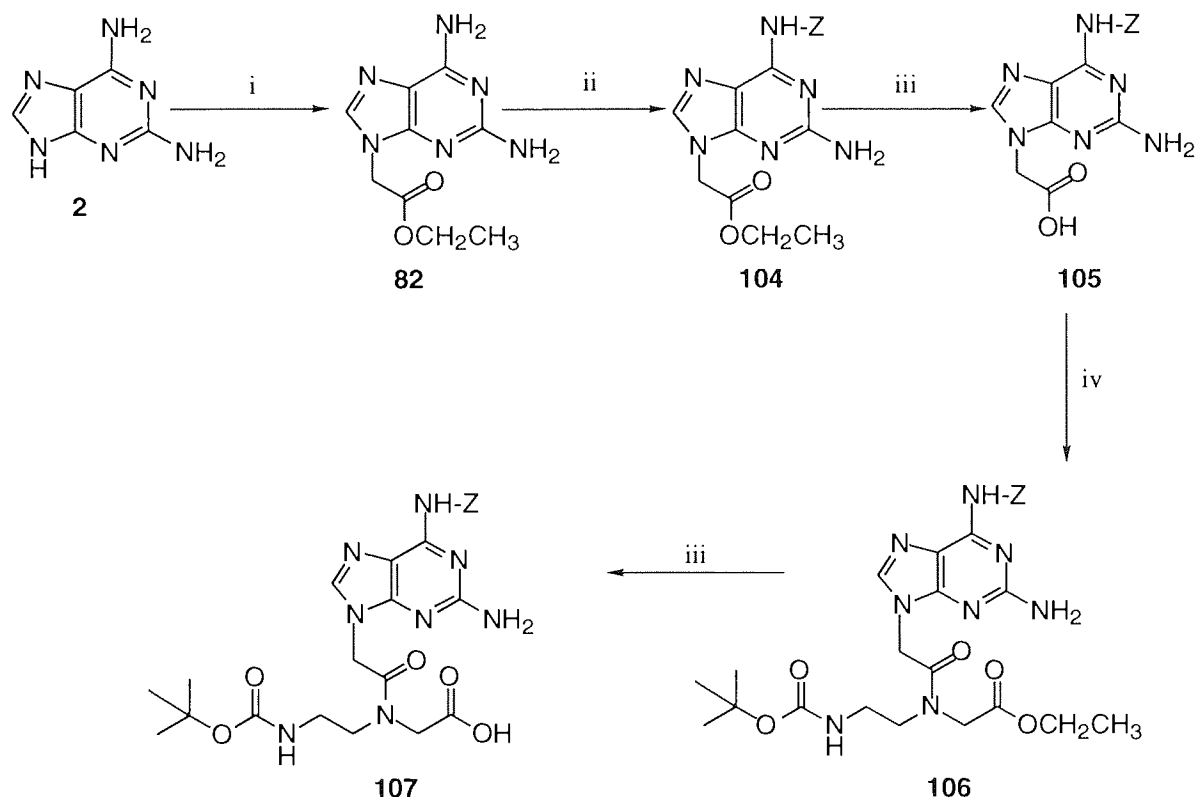
**Scheme 3.13.** Reagents: i,  $\text{BrCH}_2\text{CO}_2\text{CH}_2\text{CH}_3/\text{K}_2\text{CO}_3/\text{anhydrous CH}_3\text{CN}/\text{r.t.}$ ; ii, sodium azide/acetone-EtOH/ $75^\circ\text{C}/72\text{ h}$ ; iii,  $2\text{ M NaOH}/4\text{ M HCl}/20\text{ mins}/\text{reflux}$ ; iv, DCC/DhbtOH/DMF/r.t.; v,  $\text{H}_2/10\% \text{ Pd on charcoal}/\text{EtOH}/\text{r.t.}/48\text{ h}$ .

A simple work-up procedure involving solvent extraction afforded a mixture which displayed two spots from TLC analysis. The organic phase was evaporated to dryness *in vacuo* to yield a pale yellow powder. Inspection of NMR analysis showed a presence of the desired peaks corresponding to the acquisition of the desired product **102**. Additional peaks arising from the presence of some impurities were also detected. NMR analysis and accurate mass spectral analysis confirmed the synthesis of the desired diazidopurine building block **102**, where the  $\text{M}^+$  value was 460; in agreement with the calculated  $\text{M}^+$  of

460. However, repeated attempts using both solvent extraction and flash column chromatography failed to remove the impurities. Subsequent hydrogenolysis of the crude product **102** failed to yield **6**, despite numerous attempts where the quantity of the catalyst and duration of the reaction was altered. TLC analysis failed to show any newly-formed spots, and inspection of  $^1\text{H}$  NMR spectra showed an absence of the desired amino groups.

### 3.6.6 Recent Literature Precedent Describing the Synthesis of the 2,6-Diaminopurine PNA Monomer **6**

Very recent literature by Nielsen *et al.*,<sup>186</sup> which appeared during the drafting of this project, reports the successful synthesis of the D PNA monomer **6** (Scheme 3.14). Alkylation of **2** afforded **82** in lowered yields compared with this thesis work, although identical conditions were used. Coupling of **105** to **1** was conducted according to Scheme 3.14 to furnish **106** in yields of 70%. Attempts by Nielsen *et al.* to protect both amino groups also failed; only the amino group resident on C6 was successfully protected, confirming the low reactivity of the C2 amino group. The initial stages from the method reported by Nielsen *et al.* were identical to those undertaken in our studies.

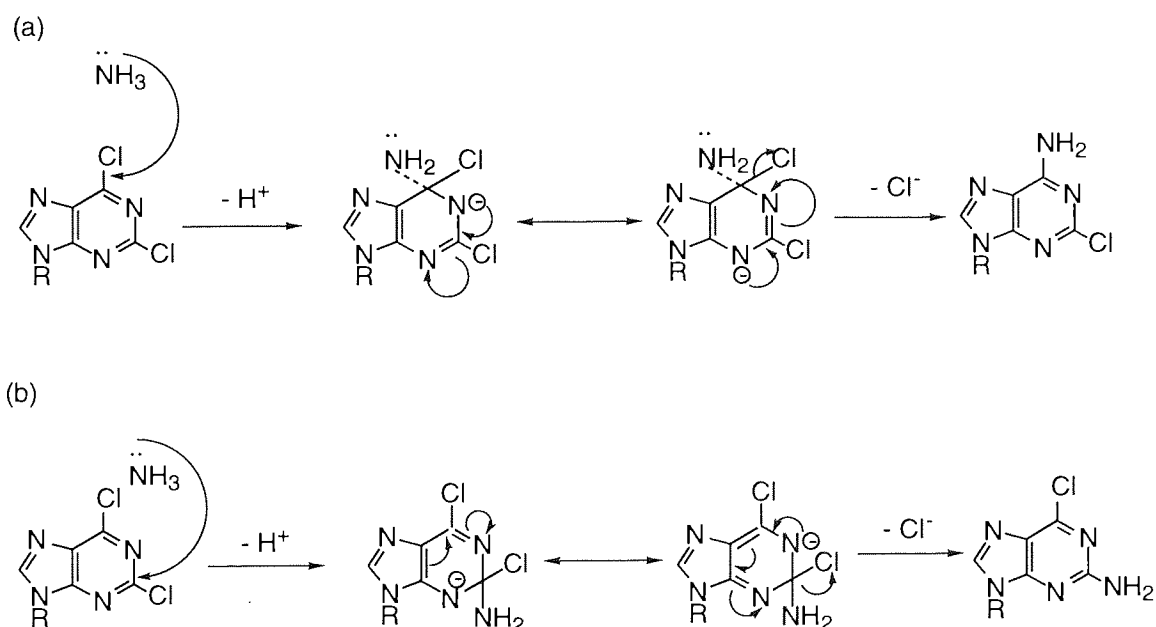


**Scheme 3.14.** Reagents. i, NaH/BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>; ii, *N*-benzyloxycarbonyl-*N'* methyl imidazolium triflate; iii, NaOH/HCl; iv, DCC/DhbtOH/ ethyl *N*-(2-Boc-aminoethyl)glycinate.

Amino protection was achieved using the triflate variant, rather than the ethyl imidazolium tetrafluoroborate, of Rapoport's reagent *N*-benzyloxycarbonyl-*N*'-methylimidazolium triflate **108**. **107** was prepared in 70% yield by brief reaction of methyl trifluoromethylsulfonate **109** with a cold solution of benzyloxycarbonylimidazole **100**. Saponification of **104** afforded **105** in 88% yield. Coupling of **1** with **105** using DCC **55** and DhbtOH **56**, followed by a solvent extraction follow-up procedure furnished **106** in 74% yield. The D monomer was finally acquired through hydrolysis of **107** with dilute NaOH as a white solid in 82% yield. PNA oligomers were synthesised using the standard protocols.

### 3.7 Attempted Synthesis of the Building Block Incorporating IsoG Using the Versatile 2,6-Dichloropurine Base **86**

As 2-amino-6-chloropurine (AC) **49** was used as a versatile starting material for the synthesis of the G PNA monomer **38**, it was recognised 6-amino-2-chloropurine **110** would be an obvious choice of base for the synthesis of the isoG building block **111**. The greater reactivity of the 6-chloro moiety of **49** compared to the C2-amino moiety for regioselective nucleophilic displacement, is well-documented from similar studies on hexose and pentose nucleosides.<sup>232</sup> Substitution with an amino moiety of one of two halogens on a heterocyclic ring occurs at r.t. under mild conditions, and leads to a progressive increase in the basic strength of the purine. However, ensuing nucleophilic displacement of the remaining halogen moiety necessitates increasingly vigorous conditions, normally executed via heating at high temperatures with increased reaction times.



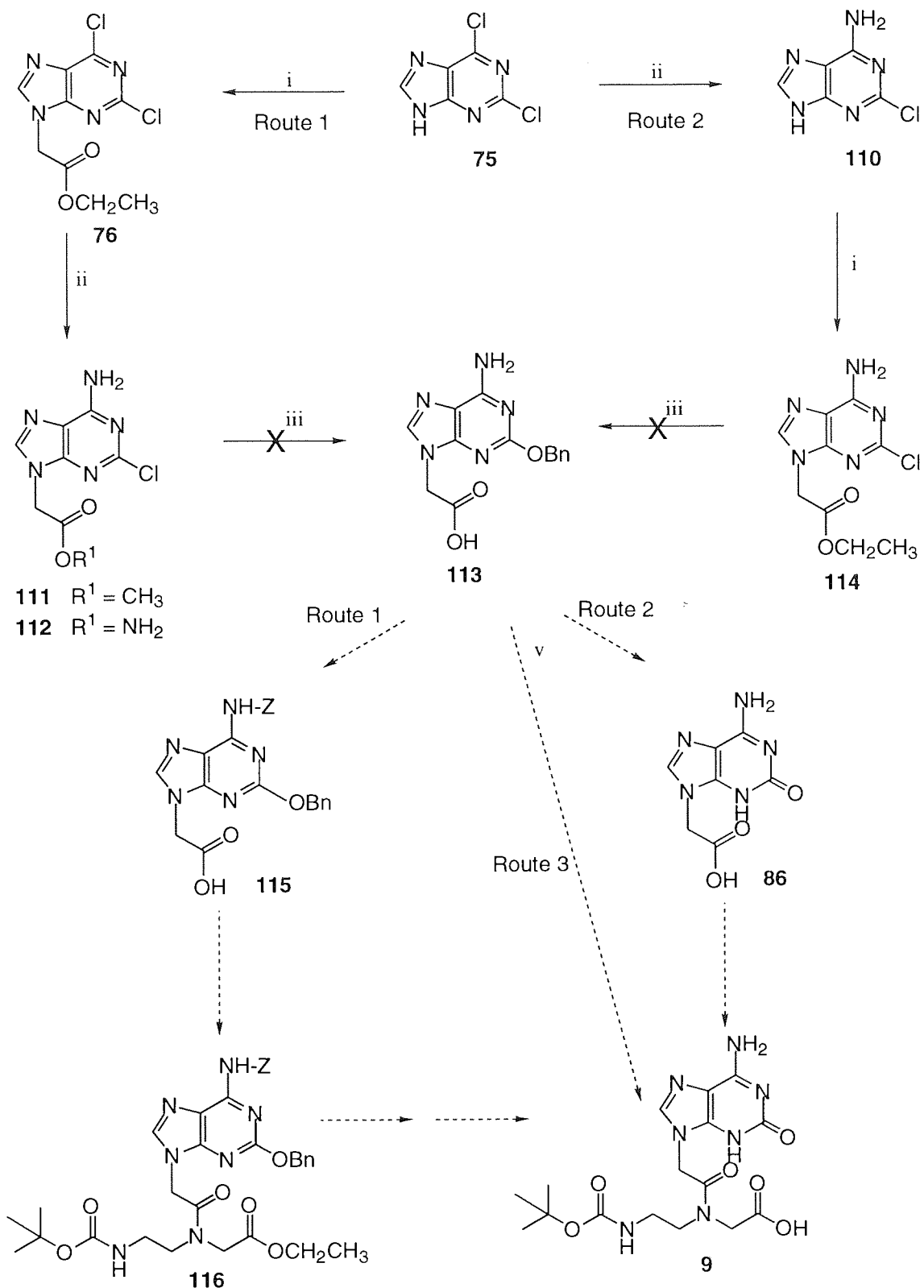
**Figure 3.9.** Resonance forms of intermediates formed by nucleophilic addition to the C6 and C2 sites of **76**.

The increased reactivity of C6 compared with C2 may be explained by examining the relative energies of the intermediates formed from the attack at C6 versus C2 (Figure 3.9). Once installed, the inductive effect of the C2 halogen group and the electron-releasing effect of C6-amino group deactivates further attack at C2 by NH<sub>3</sub>.

Amination of **75** via route 1, with a solution of ammonia saturated in methanol to selectively displace the C6 chloro group for an amino group, afforded **112**. Inspection of spectroscopic data indicated acquisition of the carboximide **112**. The process is thought to have occurred via concomitant amide formation, where an observed molecular ion peak M<sup>+</sup> = 228 corresponded to the calculated mass of the desired intermediate **112**. An alternative possibility for the acquisition of the acetic acid of **113** is thought to have occurred via transesterification of the ethyl ester followed by hydrolysis to yield the acetic acid **113**.

Installation of the O atom at C2 in its protected form occurred slowly and was attempted through a one step reaction using either the benzyl or the allyl ether via addition of **112** to a 1 M sodium benzyloxalate **117** or allyloxalate **118** mixture at 60 °C. Saponification of **111** to its acetic acid would furnish **113**. Completion of the synthetic procedure via route 1, as shown in Scheme 3.15, relied upon successful preparation of **115** via protection of the C6 amino group using benzyloxycarbonyl chloride **54** to increase the solubility of **115**. Coupling of **115** to **1**, followed by hydrolysis of the backbone ester of **116** to its acid, with a final hydrogenolysis stage was hoped to afford the isoG PNA building block **9**. Hence, a novel synthesis of the PNA monomer incorporating isoG **9**, from readily accessible materials, was established.

Preliminary investigations to install the benzyloxy or allyloxy group at C2 of **113** were unsuccessful as indicated from <sup>1</sup>H NMR analysis. Though <sup>1</sup>H NMR spectra showed an absence of starting materials, the concurrent absence of multiplets corresponding to a benzyl or allyl functionality reflected the failure of the C6 amino group to react with the benzyl or allyl oxalate. Repeated attempts with variation to reactions conditions failed to furnish **115** and this approach was not pursued further. Since initial attempts at functionalisation of the dichloropurine adduct proved more troublesome than expected, an alternative strategy was pursued (Scheme 3.15, route 2) which involved functionalisation of **110**, successfully attained via amination of **75**, with ensuing alkylation to afford **114**. Route 3 was recognised as a third pathway, where it was suspected reaction of **113** with **1** would proceed successfully due to the increased solubility of **113** attributable to the lipophilic benzyl groups. Hydrogenolysis of the intermediate would furnish the desired PNA monomer incorporating isoG. Scheme 3.15 summarises the subsequent stages of the three planned routes towards furnishing the isoG PNA monomer **9**.

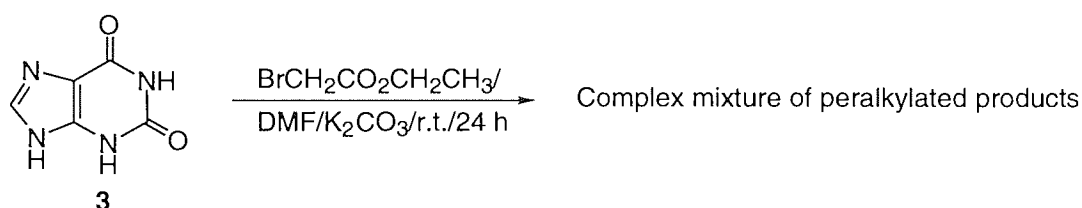


**Scheme 3.15.** Reagents: i, BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>, K<sub>2</sub>CO<sub>3</sub>, anhydrous CH<sub>3</sub>CN, r.t.; ii, NH<sub>3</sub>/MeOH/75 °C/28 h; iii, sodium benzyloxide/70 °C/24 h; v, H<sub>2</sub>/10% Pd on charcoal/r.t./24 h.

### 3.8 Synthesis of the Building Block 121 Incorporating Xanthine

#### 3.8.1 Direct Alkylation of Xanthine 3 Using Ethyl Bromoacetate 27

Since direct alkylation of the A<sup>239</sup> **34** and D<sup>240</sup> **2** bases gave the desired N9 regioisomers in high yields, the same method was adopted for the synthesis of the X building block **7**. Direct alkylation of X **3** was attempted with **27**. Use of DMF as the solvent was deemed essential as **3** was observed to be only sparingly soluble in anhydrous CH<sub>3</sub>CN. However, as observed by TLC and NMR analysis, a mixture of components was obtained. Inspection of <sup>1</sup>H NMR and mass analysis revealed the direct alkylation of **3** had resulted in a peralkylated mixture. Any one of the possible N1, N3, N7, N9, O2 and O6 sites were recognised as possible positions where alkylation may have occurred (Scheme 3.16).

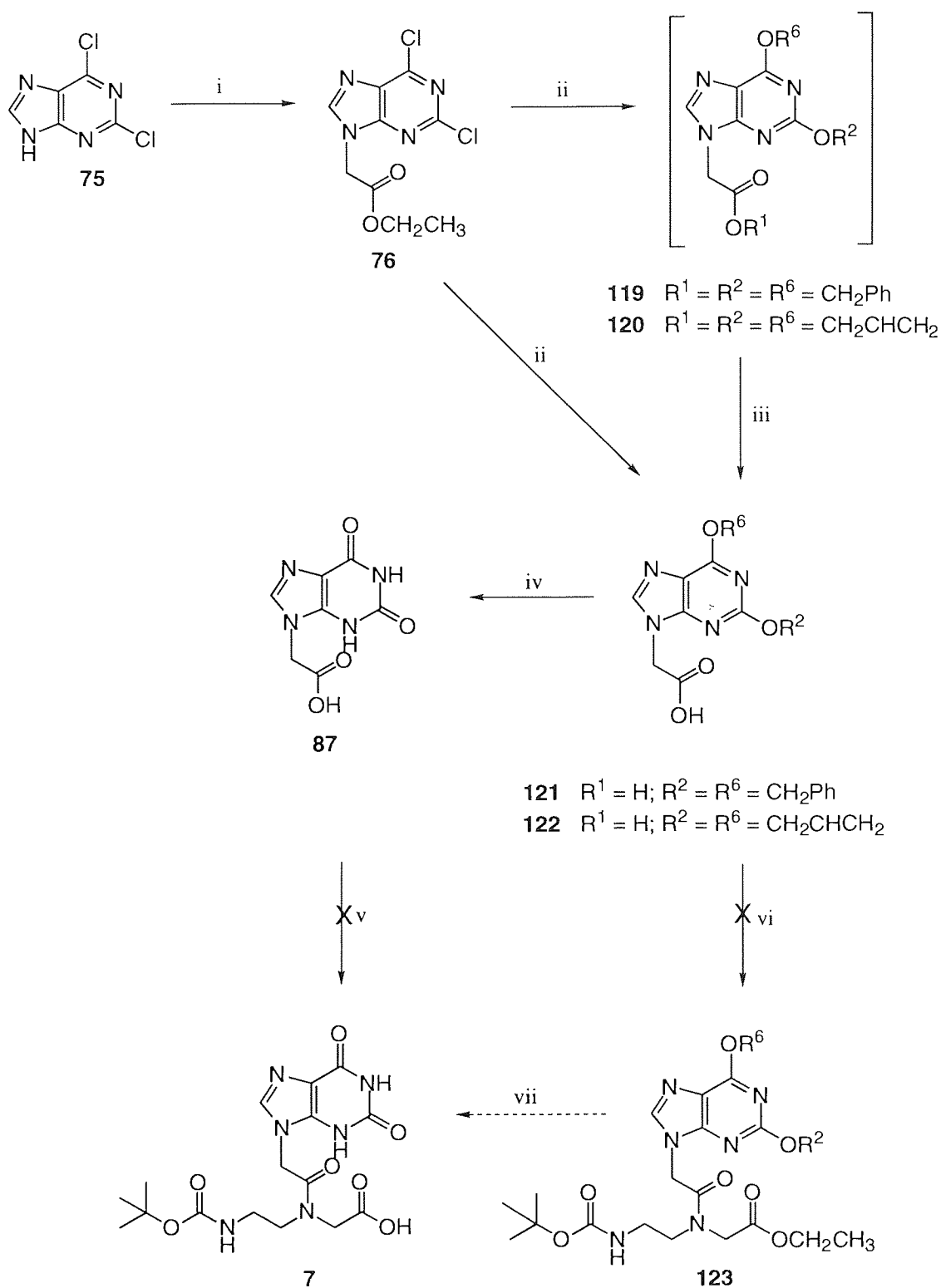


**Scheme 3.16.** Alkylation of X for representation of the potential sites for substitution from direct attack on X **3**.

Failure to isolate the regioisomers via flash column chromatography was attributed to the close proximity of the *R<sub>f</sub>* values observed by TLC analysis, thus preventing conclusive identification for each of the individual products present in the resultant complex mixture.

#### 3.8.2 Attempted Synthesis of the PNA Monomer 7 Incorporating Xanthine 3 from 2,6-Dichloropurine 75

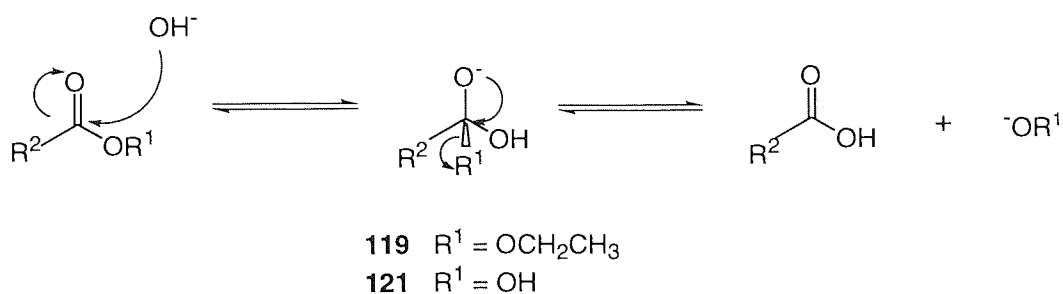
Failure of the direct alkylation approach to furnish **87** prompted us to revert to our initial strategy involving use of the versatile 2,6-dichloropurine **75**, for synthesis of the xanthine building block **87** (Scheme 3.17). Literature precedent<sup>64</sup> regarding substitution of nucleobases in the hexose DNA series at the N9 site with a sugar group, have reported the synthesis of the X PNA monomer via benzylation of 2,6-dichloropurine derivatives. Hence, the synthesis of the PNA X building block **87** was undertaken via a similar four stage procedure, as outlined in Scheme 3.17. Overnight reaction of **76** with the sodium salt of benzyl alcohol **117** or allyl alcohol at 60 °C, generated *in situ*, successfully resulted in substitution of both chloro groups to incorporate two oxygen centres at C2 and C6 in their benzylic or allylic protected forms. The product isolated from this reaction was characterised from spectroscopic analysis and analytical data which supported the acetic acid of **121** as the correct structure, and not the expected benzyl ester form of **119**.



**Scheme 3.17.** Reagents: i,  $\text{BrCH}_2\text{CO}_2\text{CH}_2\text{CH}_3/\text{K}_2\text{CO}_3/\text{anhydrous CH}_3\text{CN}/\text{r.t.}/24 \text{ h}$ ; ii, 1 M sodium benzyloxalate/ $70^\circ\text{C}/72 \text{ h}$ ; iii, 2 M NaOH/4 M HCl; v,  $\text{H}_2/\text{Pd}$ ; vi, 1 M sodium allyloxalate/ $70^\circ\text{C}/72 \text{ h}$ .

It was suspected formation of the acetic acid occurred either via direct hydrolysis of the ethyl ester as a consequence of the excess MeOH used, which was deemed essential for

elution of **119**, or from transesterification of the ethyl ester to the benzyl ester with subsequent hydrolysis to yield **121** (Figure 3.9).



**Figure 3.10.** Possible transesterification of the ester to the acetic acid **121**.

Conversion of the ethyl group during the benzylation or allylation stage was thus fortuitous as it circumvented the need for the additional saponification step to produce **121** and **122** (Scheme 3.17). Flash column chromatography with  $\text{CH}_2\text{Cl}_2$  permitted removal of excess benzyl alcohol.

Replacement of sodium metal via the alternative NaH method<sup>218</sup> to form the benzylate anion, enabled isolation of **121** via a simpler and shorter solvent extraction procedure to afford **121** in significantly higher yields and purity. Hydrogenolysis of **121** using 10% Pd on charcoal in ethanol for 4 days at r.t., yielded **87** as a pale white powder.

Allylic systems<sup>246</sup> behave analogously to benzylic systems. Homogenous Pd(0)-catalysed allyl cleavage has been shown to be very effective in hexose DNA synthesis, where allyl groups are cleaved under Noyori conditions<sup>246</sup> following DNA synthesis. The allyl protecting group was introduced in this work as an alternative choice to the benzyl protecting group. This was attributed to the need to circumvent heterogeneous catalytic hydrogenolysis which is required for the removal of the benzyl protecting group following PNA synthesis. Sodium allyloxalate **118** was generated using allyl alcohol instead of benzyl alcohol, for reaction with Na metal, under those conditions used for the synthesis of sodium benzyloxalate **117**. It was anticipated substitution of the chloro groups with allyl groups with hydrolysis of **119**, would yield the acetic acid derivative **121**. A final hydrogenolysis step to cleave the allyl moieties, thus ensuing in the incorporation of carbonyl functionalities at C2 and C6, would thus furnish the desired product **87**. However, from spectroscopic analysis allylation of **76** revealed the direct acquirement of the acetic acid of **121**, thus eliminating the need for saponification. Though flash column chromatography and attempted solvent extraction failed to furnish **121** in its pure form, NMR and mass analysis supported its successful formation. As **121** was acquired with decreased impurity and yields in the allyl reaction compared with the simpler and higher yielding benzyl reaction, the use of the allyl route was not pursued.



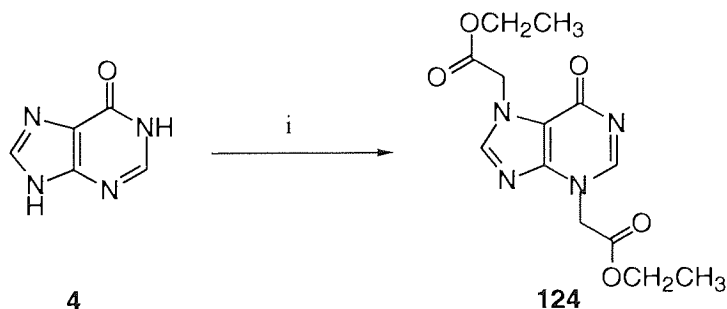
### 3.8.3 Attempted Coupling of the Xanthine Building Block **121** to ethyl *N*-(Boc-aminoethyl) glycinate **1**

Coupling reactions described in Table 3.2 for the synthesis of the D PNA monomer **82** were also applied for the attempted synthesis of the X PNA monomer **7**. The failure to couple **7** to **1** was attributed to the poor solubility of **7** where a change of solvent from CH<sub>3</sub>CN to DMF showed no improvement. The alternative coupling of **121** in its benzyl-protected form would impart increased solubility, and hence synthesis of **122** was attempted using **121**, with DCC and DhbtOH. The solubility was observed to significantly improve, and **122** completely dissolved after a short period, where TLC monitoring showed a newly-formed spot. Solvent extraction was undertaken in an attempt to remove impurities from the product mixture. The ambiguous analytical and spectroscopic data obtained did not support full characterisation of the xanthine building block structure. However, inspection of <sup>1</sup>H NMR analysis was encouraging, with peaks present for each of the expected features in the derived PNA monomer **123**. The pure form of **123** could not be isolated, but it is suspected that the novel X PNA monomer **123** can be acquired in its pure form with improved yields using this method.

## 3.9 The Synthesis of the PNA Monomer Incorporating Hypoxanthine **130**

### 3.9.1 Direct Alkylation of Hypoxanthine **4** Using Ethyl Bromoacetate **27**

In analogy with xanthine **3**, the direct alkylation of hypoxanthine **4** produced a mixture of complex by-products (Scheme 3.18) composed of two major components as observed by TLC analysis. Due to the close proximity of the *R<sub>f</sub>* values, use of flash column chromatography proved unsuccessful for isolation of the major components. Mass spectroscopic data indicated the acquisition of peralkylated derivatives. Recrystallisation of the crude mixture, containing **124** from EtOAc, furnished the major component **124** as white crystals in 61% yield.<sup>247</sup> X-ray structure analysis provided conclusive structural proof for **124**. Crystal analysis revealed that substitution had occurred unexpectedly at the N3 and N7 sites (Scheme 3.18).



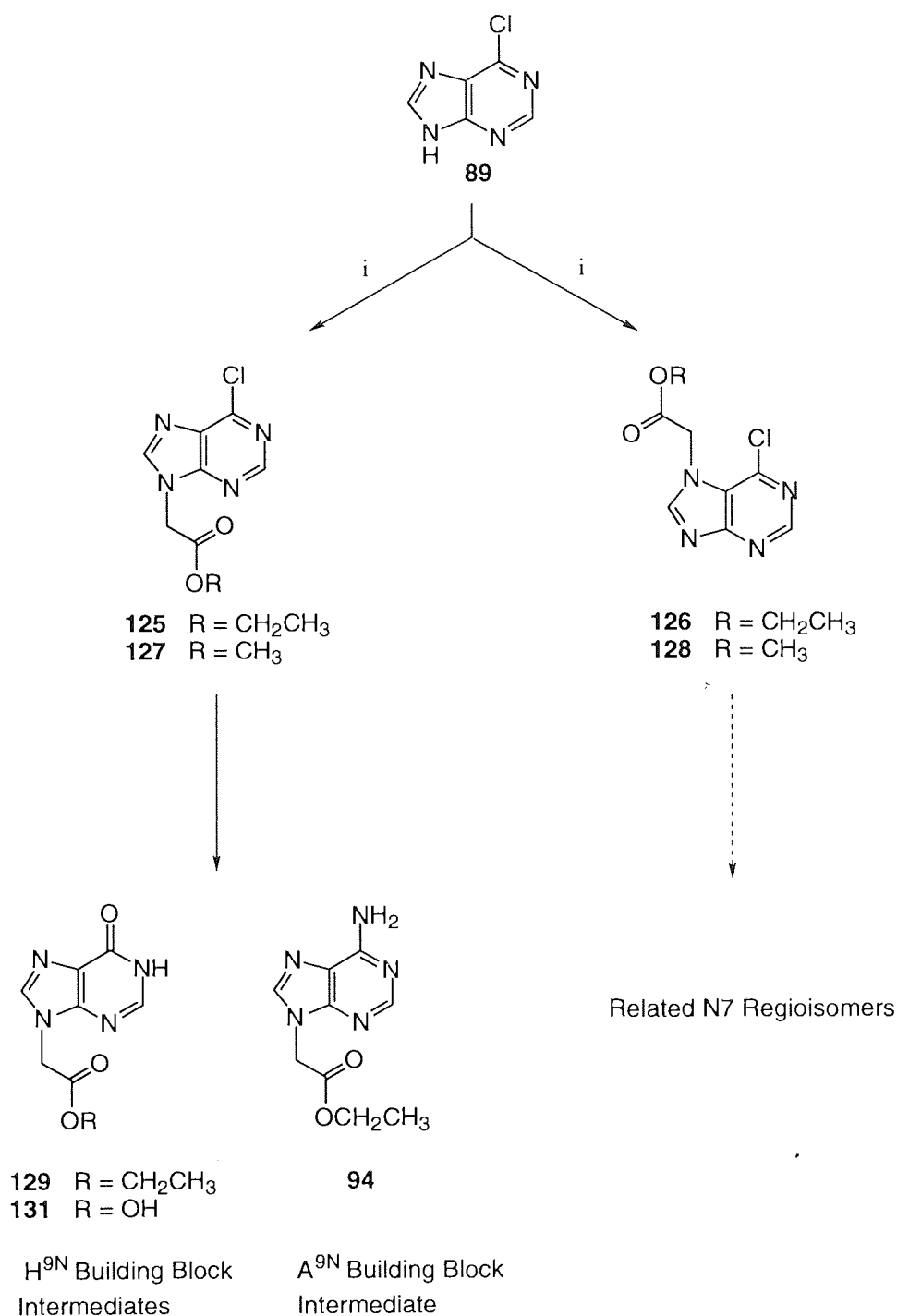
**Scheme 3.18.** Reagents: i, BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>/K<sub>2</sub>CO<sub>3</sub>/anhydrous CH<sub>3</sub>CN/r.t./24 h.

The site of substitution of incoming substituents has been shown by Montgomery *et al*<sup>235</sup> to be dependent upon those substituent groups already present upon the ring. H possesses five potential sites where substitution may occur. The peralkylated H product **124** presumably forms due to the necessity to alleviate steric hindrance between two adjacently located ethyl acetate substituents, thus discouraging attachment at both the N3 and N9, or at both the O6 and N7 sites in the same hypoxanthyl derivative **124**. Hence, the regiochemistry of the major dialkylated product, is determined by the site of attachment of the first ethyl acetate fragment. The alkylation of N9- or N7- alkylated hypoxanthine derivatives has been shown to occur at the N3, or N1 site, respectively.<sup>236,247</sup>

### 3.9.2 6-Chloropurine **92** as a Versatile Starting Material for Synthesis of the PNA Monomer Incorporating Hypoxanthine **4**

As with the direct alkylation of X **3**, the direct alkylation of H **4** also produced peralkylated products. Hence, an alternative synthetic method for the synthesis of the H building block **128** was necessary. Montgomery *et al* have reported that alkylation of 6-chloropurine **89** in basic solution using **27** produces both the N7- and N9- substituted regioisomeric forms **125** and **126**, respectively.<sup>235</sup>

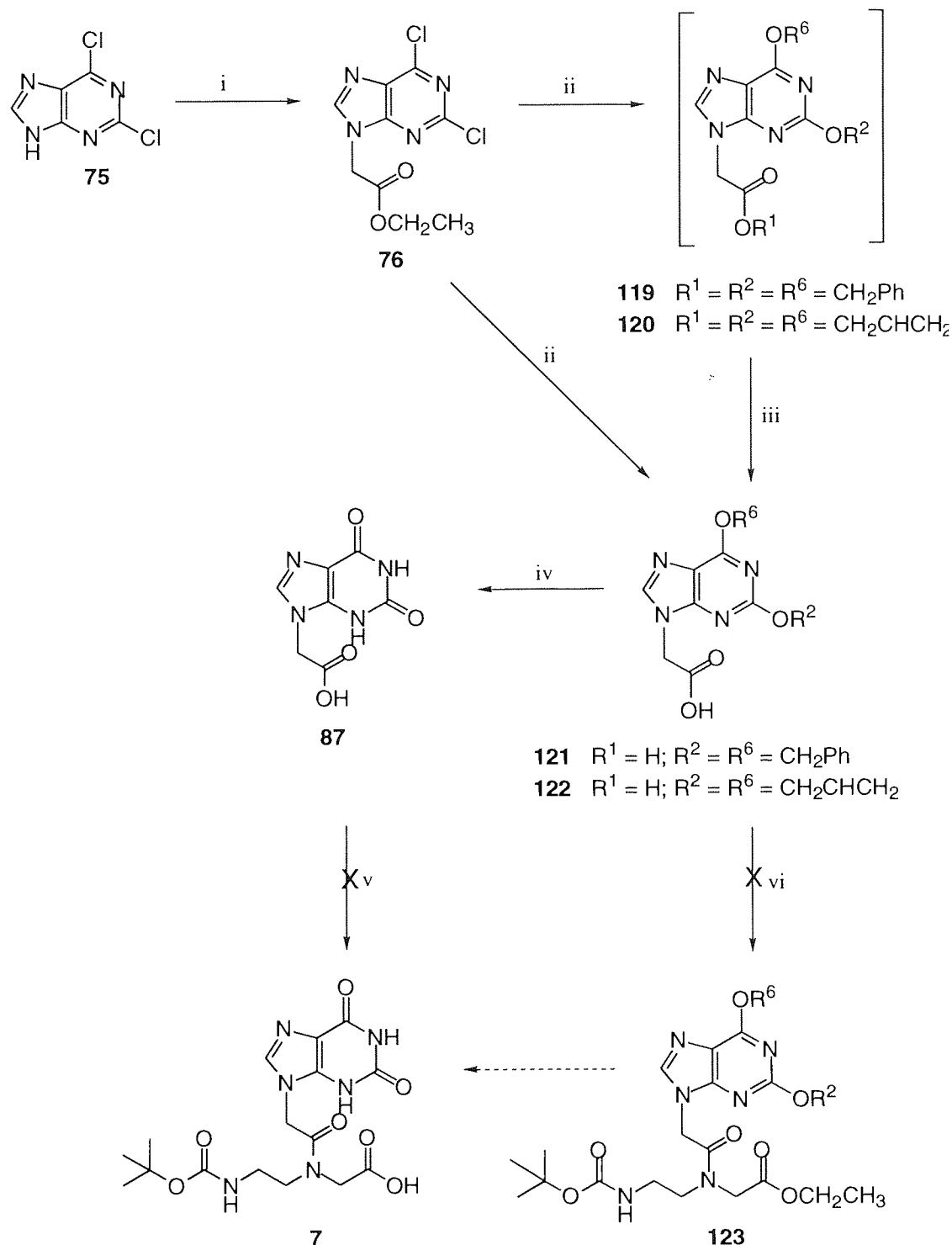
In this thesis work, the alkylation of 6-chloropurine **89**, under those conditions imposed upon the analogous procedure for **75** using **127** and **128**, provided a regioisomeric mixture of products.<sup>248,249</sup> NMR analysis showed these to be 6-chloro-9-(carboxymethyl)purine ethyl ester **125** and 6-chloro-7-(carboxymethyl)purine ethyl ester **126**, and 6-chloro-9-(carboxymethyl)purine methyl ester<sup>249</sup> **127** and 6-chloro-7-(carboxymethyl)purine methyl ester<sup>248</sup> **128** respectively (Scheme 3.19). Separation of the positional isomers, **125** from **126**, was observed to be easier than from the similar type of reaction involving the 2,6-dichloropurine regioisomers **76** and **78**, as a consequence of the increased  $R_f$  difference. Flash column chromatography elution with EtOAc, followed by recrystallisation from MeOH yielded the N9 regioisomer **125** as pale yellow crystals in 65% yield.<sup>249</sup> Further elution with EtOAc-MeOH (8:1) with recrystallisation from MeOH furnished the N7 regioisomer **126** as dark yellow crystals in 27% yield<sup>248</sup> (Scheme 3.19). X-ray crystallography analysis was undertaken to confirm the regiochemistry of **125** and **126**. Flash column chromatography elution with EtOAc, followed by recrystallisation from MeOH also yielded the N9 methyl ester **127** as pale yellow crystals in 52% yield. Further elution with EtOAc-MeOH (8:1) with recrystallisation from MeOH furnished the N7 methyl ester **128** as yellow crystals in 30% yield. X-ray crystallography analysis could not be taken to confirm the regiochemistry of the homologues attributable to the polycrystalline structures of **127** and **128**.



**Scheme 3.19.** Reagents: i, BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>/K<sub>2</sub>CO<sub>3</sub>/anhydrous CH<sub>3</sub>CN/r.t./24 h.

Displacement of the chloro group at C6 in **125** using **117** gave **132**. The desired compound **132** was obtained as a pale yellow powder in 39% yield from flash column chromatography using CH<sub>2</sub>Cl<sub>2</sub>. Hydrogenolysis to remove the benzyl moiety under H<sub>2</sub> during 3 d using 10% Pd on charcoal, successfully furnished the desired compound in 27% yield. Transformation of the N7 regioisomer of **76** into the N7 isomer of **135** was also performed according to Scheme 3.20. Both N9 and N7 isomers **132** and **135** were found to crystallise to afford yields of 66% and 37% respectively. However, due to the

polycrystalline structure of **132** and **137**, X-ray crystallographic analysis could not be applied to **132** or the N7 regioisomer of the xanthine acetic acid **135**, although these compounds were fully characterised by spectroscopic and analytical methods. Hydrogenolysis of **132** was successfully yielded *N*-alkyl H **130**. Scheme 3.20 shows the remaining step for synthesis of the PNA building block **8** which was not carried out successfully due to time constraints.



**Scheme 3.20.** Reagents: i,  $\text{BrCH}_2\text{CO}_2\text{CH}_3/\text{K}_2\text{CO}_3/\text{anhydrous CH}_3\text{CN}/\text{r.t.}/24 \text{ h}$ ; ii,  $1 \text{ M}$  sodium benzyloxalate/ $70 \text{ }^\circ\text{C}/72 \text{ h}$ ; iii,  $2 \text{ M}$   $\text{NaOH}/4 \text{ M}$   $\text{HCl}$ ; iv,  $\text{H}_2/\text{Pd}$ ; vi,  $1 \text{ M}$  sodium allyloxalate/ $70 \text{ }^\circ\text{C}/72 \text{ h}$ ; v,  $\text{DCC}/\text{DhbtOH}/\text{DMF}/\text{r.t.}$

Coupling of **131** using DCC **55** and DhbtOH **56**, appeared to successfully furnish **133**.  $^1\text{H}$  NMR analysis exhibited expected peaks corresponding to the groups in the H PNA monomer **133**, and were successfully assigned. However, due to time limitations, this reaction was repeated twice but usable quantities of **8** were never isolated.

### 3.10 Comparison of Chemical Shift Values for Related Purine Bases

Non-regioselective alkylation of various bases<sup>252</sup> was investigated by Osterman *et al.*,<sup>253</sup> who showed regioisomeric assignment of the N9- versus N7 substituted derivatives could principally be established through comparison of  $^1\text{H}$  and  $^{13}\text{C}$  NMR. The signals for the C8, C5 and C1 carbons for regioisomers in  $^{13}\text{C}$  spectra, and for the protons on the N1, N3, N7, C8 and N9 sites in  $^1\text{H}$  spectra should be somewhat different to assist in the regioisomeric assignment of these sites in the N9 and N7 regioisomers.

**Table 3.3.**  $^1\text{H}$  Chemical shifts of the series of similarly alkylated purine bases.

	<b>CH<sub>3</sub></b>	<b>CH<sub>2</sub></b>	<b>CH<sub>2</sub></b>	<b>C8-H</b>
<b>76</b>	1.21	4.20	5.24	8.70
<b>78</b>	1.21	4.20	5.44	8.33
<b>77</b>	3.73	-	5.26	8.70
<b>79</b>	4.03	-	5.45	8.82
<b>125</b>	1.20	4.19	5.27	8.68
<b>126</b>	1.20	4.22	5.45	8.77
<b>127</b>	3.71	-	5.28	8.67
<b>128</b>	3.74	-	5.47	8.76
<b>80</b>	1.20	4.12	4.84	7.67
<b>81</b>	1.21	-	4.84	7.68
<b>82</b>	-	4.79	-	7.66
<b>84</b>	3.71	-	5.61	8.41
<b>121</b>	-	4.57	5.64	8.25
<b>132</b>	-	4.52	5.62	8.36
<b>124</b>	1.21	4.13	4.83	8.11

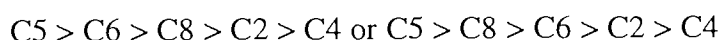
**Table 3.4.** <sup>13</sup>C Chemical shifts of the series of similarly alkylated purine bases.

	CH <sub>3</sub>	CH <sub>2</sub>	CH <sub>2</sub>	C-5	C-6	C-8	C-2	C-4	CO
76	14.0	44.5	62.8	130.2	151.8	146.2	153.1	153.2	166.1
78	14.1	48.1	62.1	122.6	143.8	150.8	153.4	153.9	167.3
77	45.9	-	54.3	130.3	151.5	150.3	156.2	162.2	168.0
79	48.0	61.4	-	138.0	151.4	148.9	157.6	163.3	168.5
125	14.1	44.9	61.9	130.7	149.4	148.1	152.0	152.2	167.6
126	14.2	48.0	62.0	122.6	142.6	151.6	152.2	152.7	168.2
127	44.8	-	52.9	130.7	149.4	148.1	152.1	152.2	168.1
128	44.1	51.8	54.0	-	151.3	152.7			
80	14.2	43.6	61.4	138.0	156.3	152.2	160.6	162.2	168.5
81	-	43.8	61.2	137.8	156.2	152.8	156.2	162.6	168.3
82	14.2	43.6	61.4	138.0	156.3	152.2	160.6	162.2	168.5
84	40.7	-	52.9	120.8	152.9	154.0	154.3	155.1	168.2
121	47.4	68.2	68.9	120.2	136.5	145.5	151.1	152.8	160.3
132	44.5	-	67.8	120.3	136.4	144.3	151.8	152.8	161.9

Tables 3.3 and 3.4 show the <sup>1</sup>H and <sup>13</sup>C chemical shift values for the series of newly substituted purines synthesised in this work. Table 3.4 shows chemical shifts for <sup>13</sup>C spectra. The change in chemical shift values for C8 and C6 was particularly affected, and observed to be dependent upon the regioisomer; N9 or N7. Studies by Jones, Pugmire<sup>116</sup> and Fuji<sup>166</sup> *et al* have shown that for unsubstituted purine, the chemical shift trend follows the order:



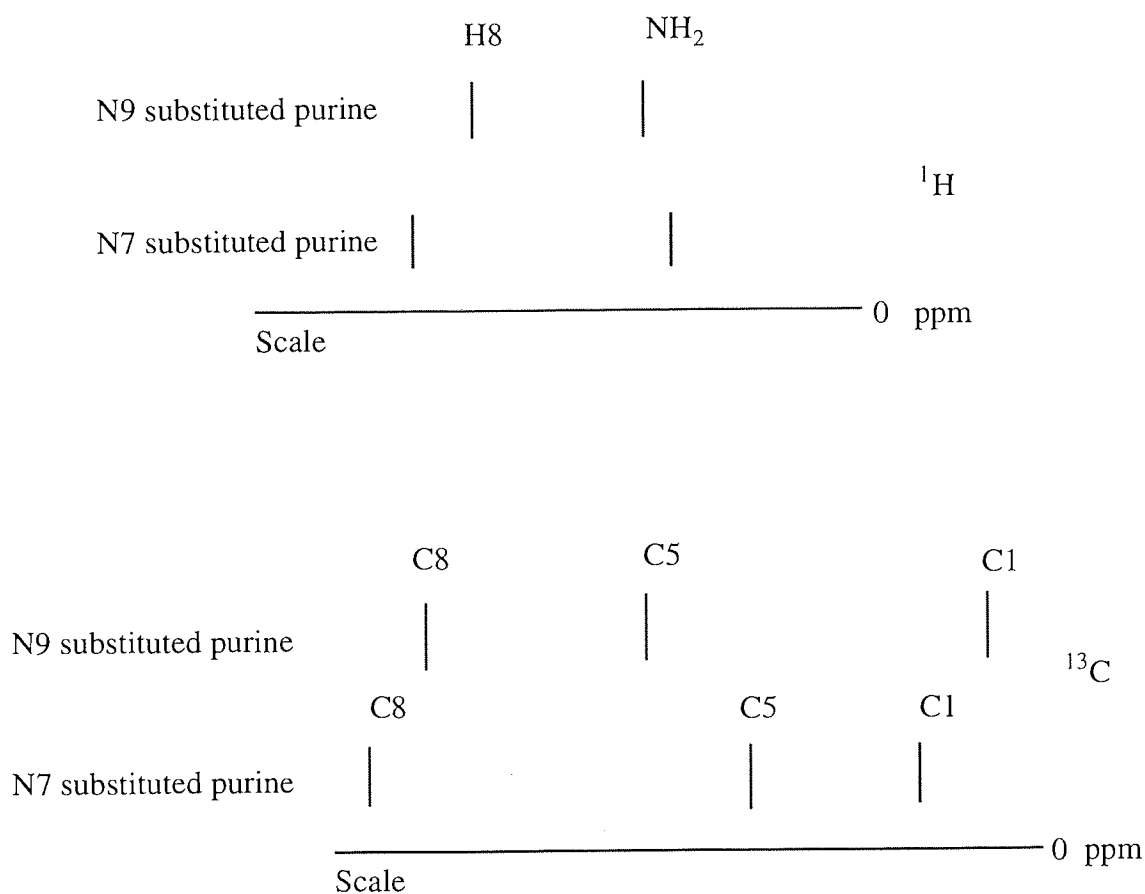
From inspection of <sup>13</sup>C spectra, it was observed that in general the trend for the purines in the alkylated series followed :



Both <sup>1</sup>H and <sup>13</sup>C spectroscopic techniques provide an opportunity to briefly study the chemical shift values for each of the substituted purine bases. The few protons present on pyrimidine and purine rings generally lead to well-resolved and easily interpretable <sup>1</sup>H

NMR spectra. Chemical shifts of heterocyclic protons are mostly dependent upon the electron density at the respective specific position. Chemical shift values of those protons directly attached to the heterocycle ring are observed at high field values, characteristic of electron-deficient sites. Substitution of the purine base results in slight modification of the chemical shifts in the substituted purine base. Variation in the ordering of  $^1\text{H}$  and  $^{13}\text{C}$  spectra for substituted purines is attributable to a ring current effect, arising from longer  $T_1$  relaxation times. The C4 signal in  $^{13}\text{C}$  spectra is observed to be smaller than the C6 signal due to a saturation effect arising from the longer  $T_1$  relaxation times. Following alkylation, proton assignment is based upon the reported relative ordering of the protons, where H6 is observed upfield to H2 which in turn is observed upfield to H8, as reported for substituted purines and analogues.<sup>116</sup> Similar trends were observed with *N*-alkylpurines in this thesis work, where alkylation at N9 and N7 moves C2 chemical shifts upfield, and alkylation at N1 and N3 moves C8 chemical shifts upfield.

The spectra of purine exhibits the ordering of the chemical shifts as C5, C3, C8, C6 and C2. The chemical shift trend established within the simple molecular spectrum forms the basis for interpreting the changes in chemical shifts of  $^{13}\text{C}$  spectra<sup>116,250</sup> for N7- and N9-alkylpurines<sup>109,235</sup> and are taken as a vindication of the validity of the theoretical estimates of the electronic structure.



**Figure 3.13.** Relative positions of selected  $^1\text{H}$  and  $^{13}\text{C}$  shifts.

$^{13}\text{C}$  NMR trends are depicted in Figure 3.13 where the shifts of the N9 isomer are shown relative to the corresponding shifts for the N7 regioisomer.<sup>236</sup> Exhibition of these characteristic shift differences assisted in the assignment of protons and carbons of well-resolved  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra,<sup>116,250</sup> and for distinguishing and characterising the N9 and N7 regioisomers<sup>33</sup> of each analogue. Rao *et al*<sup>251</sup> investigated the trends exhibited by the N9 and N7 regioisomers acquired from alkylation, and observed each carbon adjacent to an alkylated *N* atom moved upfield, in concert with our experimental results, and that the long-range electronic effects were preserved at C2 and C8.

From the aforementioned literature precedents and other numerous studies regarding alkylation of heterocyclic bases, trends for chemical shifts can be derived, and only ambiguous regioisomeric assignment can be made. Conclusive identity of the site of substitution can only be attained through X-ray crystallography which provides unambiguous information regarding the structural determination of the series of alkylated heterocycles in this work (chapter 5).



# CHAPTER 4

## X-Ray Crystallography for Structure Determination

### 4.1 The Uses of X-Ray Crystallography

The aim of X-ray crystallographic analysis is to independently confirm the structural formulae of crystalline compounds, to determine the crystal structure, and to evaluate the molecular conformation and configuration.

#### 4.1.1 The Value of X-ray Crystallography and Molecular Modelling Techniques for Structure Characterisation

The determination of molecular geometry is of vital importance in understanding the chemical structure and bonding of novel structures. Scientists have used computerised models of new chemical entities to help define activity profiles, geometries and reactivities. The Cambridge Crystallographic Database (CCD)<sup>254</sup> is a systematised class of an atlas-style compendium of all organic, organometallic and organic metal-complex crystal structures, in particular bibliographic, chemical and numerical data together with search routines, statistical software and molecular modelling programs. In pharmaceutical research involving new chemical entities, X-ray crystallography can provide patterns of molecular geometry and intermolecular interactions of analogous compounds. The initial steps are taken to attain crystallographic data and atomic co-ordinates from experimentation, a literature search, or searches of chemical databases such as the CCD.<sup>254</sup> Conformational analysis may be applied to the crystal structure to provide an ideal initial geometry. The use of these crystallographic geometries as a template with additional modelling techniques can then lead to the design of analogues and the determination of parameters for QSAR studies to complement biological data.

The majority of experimental data originates from X-ray, electron and neutron diffraction, and from microwave spectroscopy. Crystal structure analysis<sup>255</sup> from X-ray diffraction, provides data at a molecular level unattainable by conventional analytical procedures, to afford a precise 3-D structural determination of crystalline organic molecules. The development of new, and the enhancement of existing procedures, has allowed for solving, correcting and producing well-refined structures with high precision and confidence. High powered computing facilities, sophisticated software, and precision instrumentation confer data pertaining to structural images from which the exact determination and location of the individual atoms can be made. Subsequently, molecular dimensions and geometrical features including interatomic distances, bond and torsion angles, planarity, and angles between successive planes may be derived. Occasionally, the resulting 3-D representation of the atomic contents establishes the connectivity, molecular formula, and geometrical details, hitherto completely unknown, thus revealing the identity

of the compound if the structure is hitherto unknown. In this work the technique is used for both structural analysis of unknowns and for confirmation of proposed structures.

Three prominent limitations to the application of X-ray crystallography prevail. The first is the degree to which the computed model image is correct in terms of connectivity and geometrical dimensions. This arises from imperfections in the specimen crystal, errors in measurement of diffracted beams and from misinterpretation of models from maps. These errors are reflected in the difference between the observed ( $F_o$ ) and the calculated structure amplitudes ( $F_c$ ), conventionally reported as the  $R$  (discrepancy) factor. Secondly, nitrogen atoms may not be unambiguously distinguishable from carbon atoms; an important feature, particularly in the study of heterocyclic systems. However, this limitation may be partly overcome by the use of known bond lengths which assist in interpreting ambiguous data. A third and significant weakness, which often serves to provide confusion regarding the identity of the compound, is the difficulty in locating all H-atom positions. The latter two limitations arise as a consequence of the ability of electrons in the atoms contained within the crystal to scatter X-rays. Hence, atoms with similar electron density are difficult to distinguish and atoms with low electron density may not be located at all. This results in the acquisition of ambiguous data and consequently incorrect deduction of atom location and hence, structural identity.

## **4.2 The Crystal Analysis Procedure**

### **4.2.1 The Principal Data Requirements**

Crystal structure analysis comprises three principal stages:

- a) The experimental measurement of the unit cell dimensions and the intensities of a large fraction of the available diffracted beams of the crystal, dependent upon both the nature of the atoms and their relative positions within the unit cell.
- b) The derivation of an inferred atomic arrangement from the observed intensities of the diffraction maxima.
- c) Refinement of the proposed atom arrangement of scattering matter until calculated and observed intensities lie between the expected limits of error.

### **4.2.2 Summary of the Stages in Structure Determination**

An overview of the stages leading to crystal structure determination is outlined in Figure 4.1. Using a diffraction pattern, the first objective is to derive the unit-cell dimensions from planar spacing measurements which are directly related in a reciprocal mode to the dimensions of the crystal lattice. The space group may be deduced from the symmetry and systematic absences observed from the diffraction pattern. The formula weight of the unit cell contents can be precisely determined once the density of the crystal has been measured.

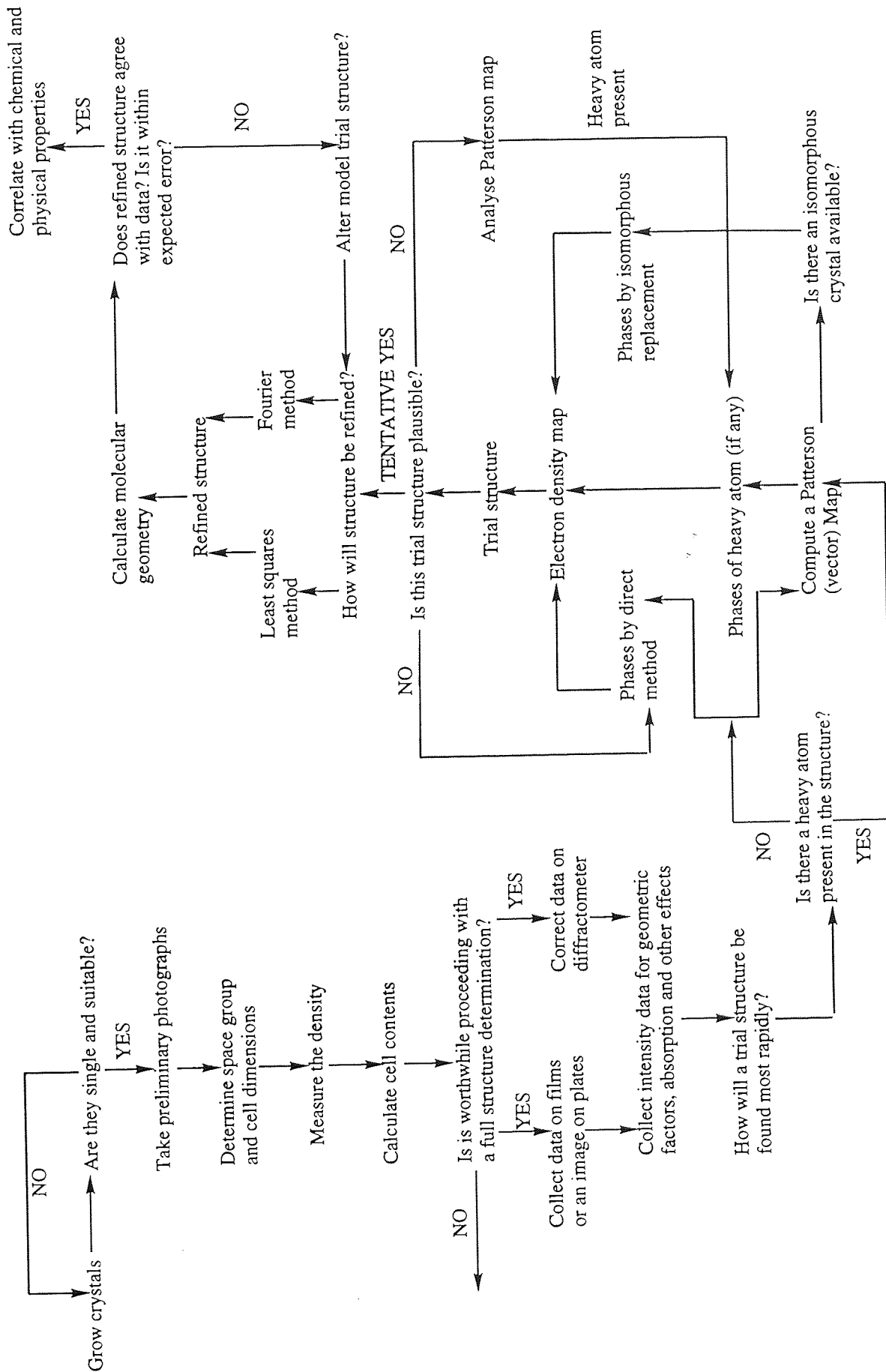


Figure 4.1. Flow diagram showing stages towards the determination of crystal structures of small molecules.

Ensuing data collection is followed by absorption correction which is sometimes delayed until after structure solution and partial refinement have been undertaken to yield the empirical formula. Data collection parameters are determined following investigation of higher metric symmetry in cell parameters. The presence of symmetry in the unit cell causes certain classes of diffraction maxima to possess equal intensity to others (Laue symmetry). Friedel's Law causes half of the diffraction maxima redundant if a centre of symmetry is present. In non-symmetric cases, these data will differ; the differences may or may not be measurable. The collection of the Laue-unique and Friedel-related data should preferably be undertaken for non-centrosymmetric space groups, though Laue unique data can also be undertaken for centrosymmetric space groups. Data reduction is undertaken with appropriate application of the background, profile, Lorentz, polarisation and scaling corrections.

Structural determination is undertaken using intensity measurements of most of the accessible 'reflections' (diffraction maxima) in the diffraction pattern, which are correlated, averaged and manipulated using geometrical factors to yield the relative values of  $|F|^2$ , together with the respective indices. Phase approximation to attain a trial structure is undertaken using either the direct methods approach or the Patterson method, where application of a computer program yields the 3-D structure. Subsequent refinement improves and furnishes the best possible fitting structure.

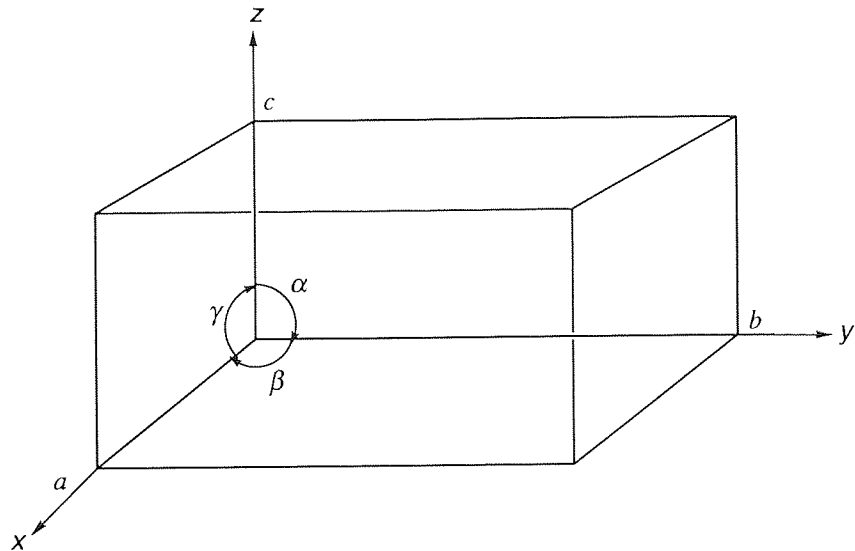
### **4.3 Crystal Geometry, Symmetry and Space Groups**

#### **4.3.1 The Symmetry of the Crystal - The Unit Cell**

Crystallisation is the process of atom or molecule arrangement from a fluid or solution to an ordered solid state. The eminently high degree of order characteristic of the crystalline state, allows exploitation via X-ray analysis of the internal periodic structure for crystal structure determination. In 3-D, the imaginary parallelepiped containing one unit of a translationally repeating pattern which displays the full symmetry of the lattice is termed the unit cell. The distances  $a, b, c$ , are its primitive translations;  $\alpha$  the angle between the  $b$  and  $c$  axes,  $\beta$  the angle between  $a$  and  $c$ , and,  $\gamma$  the angle between  $a$  and  $b$  (Figure 4.2).

#### **4.3.2 The Miller Indices**

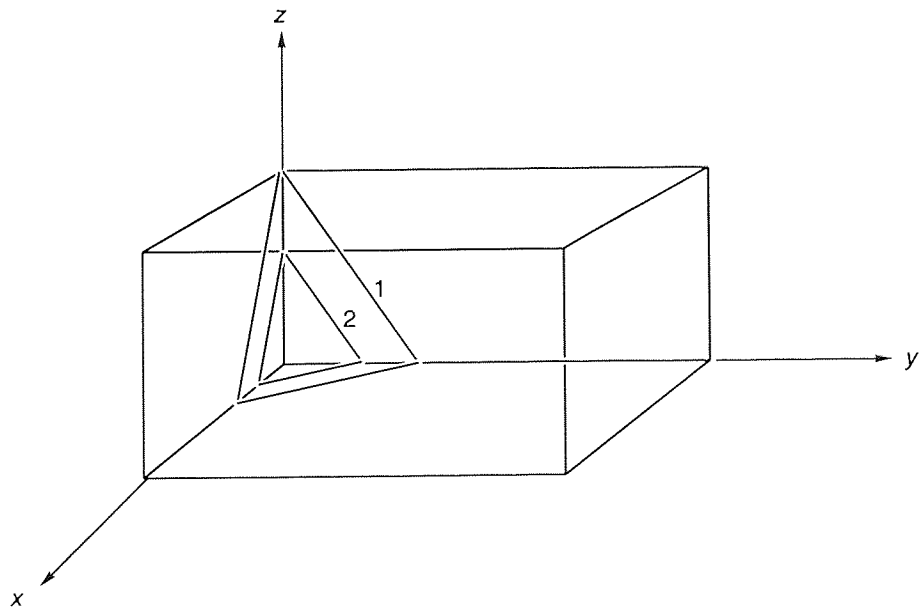
The resulting assembly of regularly stacked unit cells, each possessing identical contents, enables reduction of crystal analysis to spatial arrangement determination of those atoms within a single unit cell, or those within the asymmetric unit, if the unit cell possesses a degree of internal symmetry. In turn, the lattice of a crystal is termed a basic network of points upon which the repeating unit may be imagined to be arranged, to reproduce the regularly repeating crystal structure.



**Figure 4.2.** Representation of the unit cell dimensions of a crystal.

This network of points constitutes the formation of a stack of planes and the orientation of such planes may be defined by the ratio of the intercepts of its reference plane;  $a/h$ ,  $b/k$ ,  $c/l$ , where the  $h$ ,  $k$  and  $l$  integers used to define the plane are the designated Miller Indices, to which a symbol  $(hkl)$  may be given.

Also known as the parametral plane of the crystal, the reference plane can be any plane which has an intercept on the three crystal axes  $a$ ,  $b$  and  $c$  (Figure 4.3).



**Figure 4.3.** Example of a lattice plane and an interleaving plane.

For all planes parallel to a given direction, the ratio of the intercepts on the axes is the same. For a given value of the ratio  $a:b:c$ , the Miller indices of a plane are inversely proportional to its intercept on the axes, and thus fundamental to modern-day mathematical resolution of the crystal structure. The geometrical reason for their use arises from the fact that  $h/a, k/b, l/c$  are proportional to the direction cosines of the normal to the plane ( $hkl$ ), which is used for crystallographic definition. Miller indices may be used in three ways: to designate a set of lattice planes, to designate a particular member of the set, or to designate the face of a macroscopic crystal parallel to the set.

### 4.3.3 The Reciprocal Cell

There also exists a second related lattice, termed the reciprocal lattice whose unit cell lengths are denoted as  $a^*, b^*$  and  $c^*$ , and the interaxial angles as  $\alpha^*, \beta^*$  and  $\gamma^*$ . The relationship between the two correlates the fundamental translations of one lattice in a perpendicular orientation to the other. For orthogonal axes, the fundamental translations and dimensions of the reciprocal lattice are respectively parallel and inversely proportional to those of the crystal lattice. Thus, the geometry of the diffraction pattern attained from crystal analysis is related to the lattice and unit cell geometry of the crystal structure. Reciprocal cell parameters are used in diffractometer control programs to manipulate the diffraction geometry in order to derive the cell and crystal orientation parameters from selected observed reflections, with subsequent prediction of the positioning of all reflections for intensity measurements.

### 4.3.4 Bragg's Law in Reciprocal Space

One of the most significant features for solving crystal structures using X-ray crystallography derives from the work conducted by William Bragg in 1913. Bragg showed that the angular distribution of individual diffracted beams, produced when a beam of X-rays passed through a specific crystal orientation with deviations of  $2\theta$  from the direct X-ray beam, could be likened to a reflection of the incident beam from one of the sets of lattice planes present in the crystal; such a reflection could only occur if the parameters  $\lambda, d$  and  $\sin\theta$  upheld Bragg's Law:

$$2d\sin\theta = n\lambda$$

where  $\lambda$  = radiation wavelength,  $d$  = lattice plane spacing,  $\theta$  = angle of incidence of the X-ray beam, and  $n$  = integer (analogous to the order of the diffraction grating). From Bragg's equation, it may be derived that the angles of incidence and reflection are equal, and that the incoming and outgoing beams, and the normal to the reflecting planes, lie in one plane (Figure 4.4).

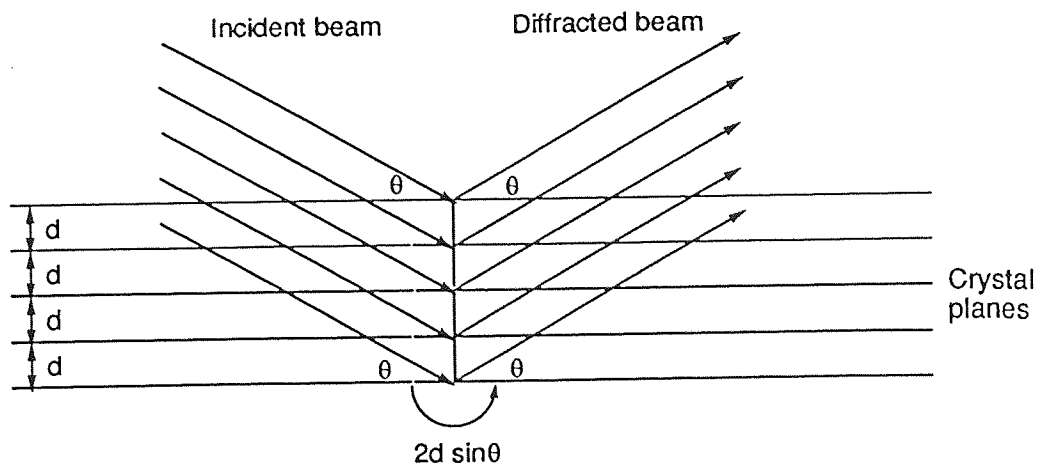


Figure 4.4. The significance of the symbols in Bragg's law.

Hence, the path difference ( $d$ ) between waves scattered from adjacent parallel lattice planes must be an integral number of wavelengths ( $n$ ), thus enabling the derivation of the position of diffraction maxima for values of the angle of incidence. The distance from each spot to the centre of an X-ray diffraction pattern is proportional to  $\sin\theta$ , demonstrating the reciprocal nature of the geometrical relationship between the crystal lattice and the angular spread of the scattering, and hence providing a convenient representation of the diffraction pattern geometry, where each reciprocal lattice point represents a Bragg reflection. However, it does not say anything about the intensities of diffraction maxima that are observed when Bragg's equation is satisfied.

#### 4.3.5 Crystal Systems and Their Symmetry

The existing 14 types of Bravais lattices may be divided into seven main systems:

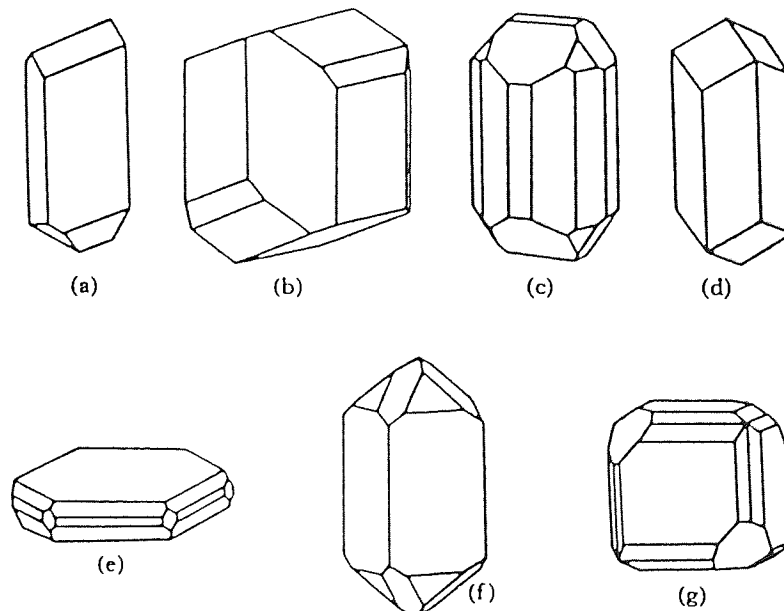


Figure 4.5. Representations of the seven crystal systems; a) triclinic, b) monoclinic, c) orthorhombic, d) trigonal, e) hexagonal, f) tetragonal, g) cubic.

Each crystal system is characterised by a specific minimum number of symmetry elements, and referable to certain characteristic axes (Table 4.1).

**Table 4.1.** Characteristics of the seven crystal systems.

Crystal System	Axial Length	Axial Angle	Symmetry Characteristics
Cubic	$a=b=c$	$\alpha=\beta=\gamma=90^\circ$	Three-fold rotation axes along all four body-diagonals. Four-fold axes parallel to each crystal axis. Two-fold axes present.
Tetragonal	$a=b\neq c$	$\alpha=\beta=\gamma=90^\circ$	Four-fold rotation axis parallel to $c$ . Two-fold rotation axes perpendicular to $c$ .
Trigonal	$a=b=c$	$\alpha=\beta=\gamma\neq 90^\circ$	—
Hexagonal	$a=b\neq c$	$\alpha=\beta=90^\circ$ ; $\gamma=120^\circ$	Six-fold axis parallel to $c$ . Two-fold rotation axes perpendicular to $c$ .
Orthorhombic	$a\neq b\neq c$	$\alpha=\beta=\gamma=90^\circ$	Three mutually perpendicular two-fold rotation axes
Monoclinic	$a\neq b\neq c$	$\beta\neq 90^\circ$ ; $\alpha=\gamma=90^\circ$	Two-fold rotation axis parallel to $b$
Triclinic	$a\neq b\neq c$	$\alpha\neq\beta\neq\gamma$	—

Other lattices possess additional centring; monoclinic with centring on one face, orthorhombic with centring on one face or on all faces, or with body centring and tetragonal with body centring. The symmetry of the lattice shows as metric symmetry (special values of unit cell axial ratios and angles). Conversely, special values of unit cell dimensions are a guide to lattice symmetry but sometimes a misleading value accidentally occurs resulting in incorrect determination.

#### 4.3.6 Point Groups and Space Groups

Molecular materials only occasionally pack so that molecular symmetry coincides with lattice symmetry; the analysis of intermolecular interactions such as H-bonding and base stacking shows that they frequently make use of the unit cell symmetry for efficient packing (section 5.8). The symmetry of unit cell contents may be derived from the symmetry of the diffraction pattern and from systematically absent reflections, with subsequent derivation of the Bravais lattice and the probable space group which may also reveal information on unit cell content and molecular packing prior to structure



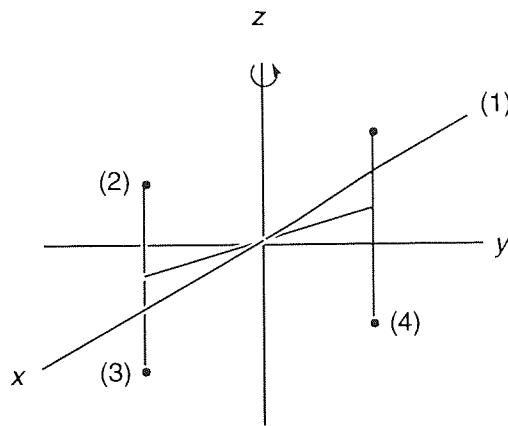
determination. Space group determination also vastly simplifies the analysis of the diffraction pattern as different regions of the pattern are known to be identical. The presence of a lattice restricts the possibilities of rotational symmetry to a total of five types. Point symmetry operations which have at least one point of an object unmoved, comprise:

- a) ***n*-fold rotation axes (1,2,3,4,6)**: a rotation of  $(360/n)^\circ$  results in self-coincidence of the structure.
- b) ***n*-fold rotatory-inversion axes (1, 2 or *m*, 3, 4, 6)**: application of this inversion operation with the origin of co-ordinate as the 'centre of inversion' implies that every point  $x,y,z$  becomes  $-x,-y,-z$ . An *n*-fold rotation-inversion axis implies a rotation of  $(360/n)^\circ$  followed by inversion through a point on the axis produces no apparent change in the structure.
- c) **mirror planes (*m*)**: application of this operation produces a reflection of the structure into self-coincidence.

It is characteristic of point groups that there is a point in space about which or through the symmetry elements may be considered to operate and which remains unmoved by the operations. These point symmetry operations can be combined in 32 distinct ways with the 14 Bravais Lattices to give the 3-D crystallographic point groups. Translation symmetry becomes possible as it is no longer necessary for successive applications of a symmetry operation to bring a point back to its original position. Subsequently, combinations of point-symmetry operations with their translations enables the derivation of the following space-symmetry operations:

- a) ***n*-fold screw axes  $n_r$** : the corresponding symmetry operation is a rotation about the axis by  $360^\circ/n$  followed by a translation parallel to the axis by  $r/n$  of the identity period along that axis ( $r =$  an integer less than  $n$ , where  $n = 1, 2, 3, 4, 6$  is the fold of the axis) of the unit cell length in that direction. For example, a 2-fold screw axis translates by  $1/2$  the repeat distance to the axis.
- b) **glide planes**: the combination of reflection across a plane with a translation parallel to the reflecting plane. The translation in such a plane occurs along an edge or face diagonal of the unit cell. A point equivalent by a simple translational symmetry operation, a lattice vector, must be reached after two glide translations; these translations may be half of the repeat distance along a unit cell edge, where the glide plane is referred to as an a-, b-, or c-glide depending upon the edge parallel to the translation.

From the combination of pure rotations, rotatory inversions, screw axes and glide planes, 230 combinations of symmetry elements are acquired for crystal symmetry description; each combination termed a space group may be found in the International Tables.<sup>256</sup>



**Figure 4.6.** A point symmetry operation; a centre of symmetry produced by 2-fold axis combined with reflection.

For the first steps, the lowest symmetry should be considered first, which is guided by inspection of the morphology or physical properties of the crystals, or from knowledge that the crystals must be chiral. This is followed by size, shape, and density determination of the unit cell to determine how many formula units per unit cell exist. Diffraction symmetry and Bravais lattice (P, A, B, C, I, F or R) determination is followed by data measurement. Diffraction or Laue groups provide the minimum symmetry information which may be determined from the diffraction pattern. Assuming anomalous data are not obtained, the symmetry of the diffraction point group, determined from the diffraction pattern, is used to determine the necessary data required. Comparison between the symmetry equivalent intensity data and the possible crystal symmetries yields Laue symmetry. Each ray, reflected from a given set of planes in a crystal, possesses an inherent intensity dependent upon the electron distribution. As each reciprocal lattice point corresponds to one such ray, each has its own associated intensity. A reciprocal lattice in which each point has been assigned a weight equal to the intensity of the diffracted ray is termed the intensity-weighted reciprocal lattice; the possible symmetries of such lattices, the Laue symmetries, are just those of the 11 centrosymmetric point groups that result when a centre is added to each of the 32 point groups.

Once the lattice has been assigned, the space group remains to be determined.<sup>255</sup> A method that allows the reconstruction of the reciprocal lattice, point-by-point, enables observation of the systematic presence or absence of specific groups of reflections and thus imply the presence or absence of certain symmetry elements. The systematic absence conditions, appropriate for the given Laue group, are tested to determine cell centring, glide planes and screw axes, with performance of a statistical test to detect a centre of symmetry. For many of the more common space groups, systematic absences uniquely determine the space group. Unfortunately, the information provided by systematic absences is not complete; only translational symmetry elements can be detected and there

are many space groups not fully defined by these elements. As such, for this large set of less common space groups, two or more possible space group choices exist, and hence each of these is tested in turn to give the most probable solution. Initially, the highest possible symmetry space group is considered towards symmetry reduction, where the cell parameters are tested for possible higher metric symmetry.

Symmetry operations involving translation are indicated by special conditions for certain classes of data, which are listed in Volume A of the International Tables<sup>256</sup> to provide confirmation for the successful determination of a space group. If ambiguity still exists and symmetry operations not involving translations are indicated, statistical tests on intensity distribution are applied to indicate the presence of such symmetry. Finally, if this fails and a guess must be made, it is helpful to take account of the relative frequency of the possible space groups. Up to this point the process is geometrically reliant. However, it is the task of X-ray crystallography to locate the exact atom positions and this depends upon a knowledge of the intensities of the X-ray diffraction patterns, and the measurement and subsequent interpretation of these.

## **4.4 Intensity Data Collection**

### **4.4.1 The Wavelength of X-Rays**

X-rays are a good energy source in crystallographic studies as the X-ray radiation wavelength (typically 0.5 - 1.8 Å) is of the same order as the interatomic distances between the scattering centres (0.8-3.0 Å). The extent to which X-rays are scattered by electrons, thus interacting to give rise to constructive or destructive interference between two or more waves for producing a peak of intensity where reflected waves are in phase, is dependent upon the atomic number and the angle of scattering. This enables determination of the crystal geometry. Most X-ray tubes used for diffraction have targets (anodes) composed of copper or molybdenum metal. Copper radiation (1.541 Å) is preferred for small crystals or large unit cells, whilst molybdenum radiation (0.7107 Å) is preferred for larger crystals of strongly absorbing materials and for very high resolution data. The internal radius of collimators, objects used to shape the X-ray beam hitting the crystal, is typically chosen to be somewhat larger than the size of the sample so it may be bathed in the incident beam at all times.

### **4.4.2 Diffraction of X-Rays and Use of a Monochromator**

The energy of the characteristic X-radiation emitted by the atoms of the target is supplied by a beam of electrons. Deceleration of these electrons may cause white radiation X-ray emission at a variety of wavelengths, which may be eliminated using a monochromator. Crystal units must be located at the points of a space-lattice to enable successful diffraction of X-rays by a crystal. As the regularly periodic X-ray beams with divergent wavelengths

sweep through the crystal, they are diffracted from a given set of planes in a crystal at different scattering angles. Each atom scatters a small fraction of the radiation and the crystal units thus become the centres of scattered wavelets, between which a definitive phase relationship holds owing to the regular spacing of the crystal lattice, and the fact that all the wavelets are derived from the same incident wave, in accordance with Bragg's Law. Therefore, a narrow band of wavelengths can be chosen by selecting a particular scattering angle of the monochromator crystal. An additional effect of the monochromator is to systematically alter the polarisation of the incident beam. However, this necessitates geometric corrections to subsequent intensity data.

#### 4.4.3 Geometrical Parameters from Diffraction Data

Any formula for the intensity of a spectrum must contain two types of geometrical factors, of which one is dependent upon the nature of the atoms in the unit cell and their arrangement, and the other upon the condition of the crystal and the method used to observe the spectrum. Hence, the derivation of the following two types of experimental data from an X-ray diffraction pattern enables molecular geometry determination of the cell components.

**a) Geometry:** the axial lengths ( $a$ ,  $b$  and  $c$ ) and the angles between these axes ( $\alpha$ ,  $\beta$  and  $\gamma$ ) of the unit cell may be found from the observed angular deviation (angle of scattering) of individual diffracted beams from the incoming, undeviated X-ray. The direction of the emergent beam at each value of  $\theta$  at which a diffraction maximum is observed, is a function only of the cell dimensions and of the wavelength of the radiation used. Dependent upon the spacings of the reciprocal lattice, and hence of the crystal lattice, the angle of scattering thus provides information regarding the size and shape of the unit cell.

**b) Location:** the atomic scattering factor. X-rays are able to cause electrons in atoms to vibrate thus causing them to act as sources of secondary radiation. The scattering power of a single atom, symbolised as  $f$ , is a measure of its ability to act as a secondary source of radiation, and is dependent upon its electronic structure and the angle of scattering. The scattering factor can be modified if the X-rays used are appreciably absorbed by the atom. The scattering of an assemblage of atoms can be closely approximated by summing the contributions to each scattered wave from each atom independently. Thus, the intensities of the diffracted beams, dependent upon the nature and arrangement of atoms within each cell, can be calculated at any angle from the sum of the waves scattered for different atoms. The precise atom locations are expressed with respect to the lattice planes in the crystal, as fractions of the unit cell.

#### 4.4.4 Crystal Selection, Mounting and Alignment

For successful crystal analysis, the crystal itself needs to be stable, display a strong diffracted intensity at scattering angles for the incoming radiation, which are accessible on the instrument being used, and to possess a small crystal mosaicity which determines the divergence of the diffracted beam and the resolution of the crystal. It is thus worthwhile ensuring a good, high quality crystal is selected, as the effects of a poor crystal propagate through the data collection, structure solution and refinement stages. This adversely affects the quality of the final structure resulting in highly unsatisfactory uncertainties which limit useful comparisons.

All crystals, except those of particularly high symmetry, rotate the plane of polarised light. Where the field between crossed polarisers is observed to be dark, the crystal itself shows as alternately dark and bright on rotation. A perfect single crystal is expected to give a sharp extinction providing the best optical indication of crystal quality. The crystals chosen for analysis should be large enough to produce an adequate diffraction pattern, yet small enough to minimise those absorption problems characteristic with the use of X-rays. The ultimate test of a crystal is its behaviour once mounted upon the diffractometer. Reflections possessing sufficient intensity to furnish both a good peak shape and to yield an efficient index produce suitable unit cells. Those which are curved or otherwise deformed, with attached significant passengers, or which show re-entrant angles, produce poor quality data.

The selection, mounting and positioning of the crystal upon the diffractometer requires exact alignment to attain high-precision data. Positioning occurs upon the centre of the crosswires on the  $\chi$  circle or its equivalent, where the first position is that where the crystal is directed vertically upwards ( $\chi = 0^\circ$ ) and the second where the crystal axis is aligned horizontally ( $\chi = 90^\circ$ ). Unit cell dimensions are obtained from least squares analysis of setting angles, collected for a set of approximately 25 accurately centred reflections. Once the orientation matrix is known, embodying both the unit cell dimensions and the angular relationship between unit cell axes and diffractometer axes, intensity measurements can be made.

For each reflection the expected diffractometer angle settings are calculated, in particular  $\omega$ , the rotation of the base of the diffractometer, and  $2\theta$ , the angle between the detector beam path and the direct beam. Since these calculated angles may not be completely correct and the crystal consists of slightly misaligned blocks,  $\omega$  and  $2\theta$  are scanned. The peak  $\omega$  scan angle is calculated and increased on each side for background determination, and the  $\omega$  scan rate varied depending upon the intensity. Reflection scanning, the movement of the crystal and the detector angle  $2\theta$ , should be narrow enough to reduce the inefficient use of diffractometer time and the background radiation thus increasing precision, but wide

enough to avoid reflection truncation and the introduction of systematic errors in the data. As a correlation with the collimator diameter, scan type and scan width exists, the aperture is set using those conditions that prevail during data collection. Diffraction symmetry must be initially determined upon a standard diffractometer to enable efficient intensity data collection. The speed at which diffraction data are collected is dependent upon the intensity of the X-ray source. Although the reflections from the crystal need to be reasonably intense in order to achieve high  $I/\sigma$  values, strong reflections are avoided at low angles due to their increased susceptibility to an irradiation-induced increase in mosaic spread.

## 4.5 Data Reduction

### 4.5.1 Applied Corrections to Intensity Data

The intensity data collected from the diffraction process constitutes the raw material from which crystal structures are derived. The data usually consist of a set of indices and six associated measurements; indices, diffractometer setting angles, background count before the scan, peak count accumulated during the scan, and background count after the scan. These physical measurements on the crystals provide the information required for construction of the structural model. The preliminary manipulation of these intensities, that is the conversion of the raw material intensities to a corrected, more generally usable form, is referred to as data reduction. This process is usually undertaken by the data reduction program DATREDXL,<sup>257</sup> or CADABS,<sup>258</sup> to correct for background, geometrical and possibly absorption factors which may manifest as physically meaningless values of important factors such as anisotropic atomic vibration parameters, in the ensuing structure refinement process. DATREDXL<sup>257</sup> is used for the purpose of converting the intensities obtained from the diffractometer to a form that can be read by MULTAN84,<sup>259</sup> an automatic crystal structure solution program, where the output is given in the form of  $h$ ,  $k$ ,  $l$ ,  $F_o$  and  $\sigma F_o$ .

The steps applied to the raw data for the purpose of data correction include total background subtraction, Lorentz-polarisation and intensity correction.

### 4.5.2 The Lorentz-Polarisation Correction

The intensities of the diffraction maxima show a variation in different directions and also vary significantly with the angle of scattering. The X-rays scattered by a group of atoms, specifically within one unit cell of a structure in any direction in which there is a diffraction maximum, possess a specific combination of amplitude and phase, measured relative to the scattering by a single electron. This is the structure factor, symbolised by  $F$  or  $|F_{hkl}|$  and is the most important quantity derived from the intensities.

The intensity of the scattered radiation is proportional to the square of the amplitude and hence the intensity of the diffracted beam corresponding to the diffraction maximum, where  $I_{hkl}$  is proportional to  $|F|^2$  divided by geometric factors.

$$|F| \propto \sqrt{I}$$

where  $|F|$  = structure factor;  $I$  = background corrected intensity.

The structure factor amplitude is calculated theoretically once the positions of the atoms in the cell are known. Furthermore, structure factors are used in the calculation of electron density maps from which the positions can be determined, thus it is customary to convert the intensities into 'observed' structure amplitudes  $|F_o|$ . The relationship between  $|F_o|$  and  $I$  depends upon a number of factors, primarily geometric, that are related to the individual reflection and to the apparatus used to measure its intensity.

$$|F_{hkl}| = (KI_{hkl}/Lp)$$

where  $L$  = the Lorentz factor, and  $p$ , the polarisation factor, is given as  $p = (1 + \cos^2 2\theta)/2$ .

The minimal input to the data reduction program consists of raw intensity data, each reflection identified by its indices  $h$ ,  $k$  and  $l$  and the cell parameters. Coded information with the method of data measurement is often able to provide the correct form for application of the Lorentz expression. The Lorentz factor ( $L$ ) which depends on the precise measurement technique used and upon the Bragg angle is a correction factor accounting for the time and place a rotating crystal spends in the reflection position to diffract the beam, which in turn depends upon the diffraction geometry. The polarisation term arises as a consequence of the nature of the X-ray beam and the manner in which its reflection efficiency varies with the reflection angle. The partial polarisation of an X-ray beam, when scattered by electrons from a crystal monochromator, must be taken into account as it affects the subsequent reflections from the crystal under study.

#### 4.5.3 Absorption

It is necessary to calculate the absorption for the actual path length travelled within the crystal by the beam reflecting from each infinitesimal portion of the crystal and to integrate these results over the volume of the crystal. The severity of absorption is dependent upon both the number of heaviest atoms and the crystal morphology. Two crystal-based methods may be used for absorption correction. The first is based upon accurate indexing of crystal faces and measurement of their distances from a common reference point. The second is based upon azimuthal scans where the crystal is rotated about the scattering

vectors of a number of reflections. The CADABS program<sup>258</sup> can derive absorption corrections from such azimuthal scan data. Two other corrections applied to the data include extinction, an expression for attenuation of the primary beam passing through the rotating crystal, and is dependent upon the distance of the reciprocal lattice point from the origin. Extinction predominantly affects strong, low-angle reflections and is corrected by a single correction factor as a variable of structure refinement. Thermal-diffuse scattering artificially enhances high-angle reflection intensity.

#### **4.5.4 Crystal Deterioration and Percentage Decay**

Data collection also involves the periodic measurement with ensuing merger of a set of six standard reflections, three intensity reflections and three orientation reflections to monitor any decomposition or slippage of the crystal. To ensure that reliable, relative intensities are attained, repeated checks for calibration of standard reflections are made. As data collection proceeds, the integrated intensities of those reflections measured may significantly change with time, where a monotonic decrease in diffraction intensity indicates possible crystal deterioration. Crystal decomposition, triggered by exposure to X-rays, is not an uncommon occurrence. Radiation damage may be detected via periodic intensity monitoring of a set of three reference reflections, as a function of data collection time. This enables a comparison between intensity values with subsequent calculation of the extent of radiation damage. Once data collection has proceeded without any overall trend in the intensities of the check reflections, the spread of measurements may be used to estimate the parameter if used in the beam correction. The purpose of decay correction is to place intensity measurements at different stages of the data collection on the same scale. Application of a simple linear intensity corrector to the measured data, at the measured periods using programs called DATREDXL<sup>257</sup> or CADABS,<sup>258</sup> adjusts for this observation.

### **4.6 Data Conversion**

#### **4.6.1 Scaling**

Once the necessary corrections have been applied to the individual reflections, it is often necessary to bring various sets of reflections to a common basis in order to attain a uniform data set, where the output from the data reduction program consists of a set of records, one to a reflection, containing the necessary information for subsequent calculations.

#### **4.6.2 Absolute Scaling and Temperature Factors**

Useful information may be obtained via statistical comparison of the observed intensity data with the theoretical predictions for a crystal composed of a random assemblage of atoms. Following from Bragg's Law, if a particular set of lattice planes coincide in



orientation and position with a densely populated planar set of atoms in a crystal, an intense diffraction maximum is detected as scattering from corresponding atoms occurring in phase. The amplitude for scattering electromagnetic radiation from a single atom, as a function of angle, is given by the scattering factor ( $f$ ) expressed in terms of the power of an equivalent number of electrons located at the position of the atomic nucleus. The variation of the scattering factor is a consequence of the finite size of the atom, which is regarded as the scattering source. Hence, the amplitude measured from the entire crystal reveals the type of atoms in the crystal and the phase angle between path lengths. Thermal motion also has an effect on the X-ray intensities which causes the scattering of the real atom to fall off more rapidly than that of the ideal, stationary model.

#### 4.6.3 Preliminary Stages to the Final Structure

The practical approach for advancing from the observed structure magnitudes to the final structure hereon is comprised of three stages. The first and most critical of these involves the development of a phase set and associated structural model, sufficiently close to correct for the application of Fourier methods. This phasing model is a poor representation of the entire chemical structure due to its incompleteness. In the second stage, successive cycles of structure factor and Fourier calculation are continually elaborated such that all the atoms in the molecule are located and placed in reasonably correct positions; only distinctions regarding atom types and bond orders should necessitate additional detail at this stage. The phases should be fairly well fixed, with little variation in detail required during the final stage. Subsequent refinement is undertaken to attain the best fit between  $F_o$  and  $F_c$ , and the best approximation to the true structure where atomic positions and temperature factors are adjusted.

#### 4.7 The Phase Problem

In determining an unknown structure, a Fourier summation needs to be performed to attain a 3-D image of the scattering matter. If the structure factors  $|F|$  and phase angles  $\alpha$  (for each  $h, k, l$ ) are known, the electron density distribution  $\rho$  of the unit cell can be calculated and plotted for all values of  $X, Y$  and  $Z$ , to give a 3-D electron density map. The electron density at a point  $X, Y, Z$  in a unit cell of volume  $V_c$ , is

$$\rho(XYZ) = 1/V_c \sum_h \sum_k \sum_l |F| \cos [2\pi (hX + kY + lZ) - \alpha]$$

If the electron density is known, assuming atoms to be at the centres of peaks the entire structure may be derived. Structure solutions would thus be superfluous in computation except for the principal difficulty that the available data consist only of the structure factor magnitudes and not of the phases. The necessity to supply these missing data: the problem of attaining estimates of the phase angles so that an image of the scattering matter can be calculated, constitutes the phase problem.

## 4.8 Determination of Atomic Parameters

Approximations to true electron density maps for solving the phase problem can be calculated by two approaches; the first by a direct calculation of initial phases which allow the deduction of an atomic model from a Fourier map; the second by the direct location of enough atoms in the cell to yield approximate but adequate phases for  $F_c$  calculation. The first uses direct methods, such as George Sheldrick's SHELXS program,<sup>260</sup> to find values of  $\alpha$  directly. The second method derives a trial structure via use of the Patterson, heavy atom, or isomorphous replacement methods, and calculates approximate values of  $\alpha$  for each reflection.

### 4.8.1 Direct Methods

The implementation of the direct methods program<sup>259,261</sup> for solving a crystal structure depends upon the program itself. However, all direct methods programs fail if a severe error is made in the choice of space group or if an incomplete set of data is collected. Of the available methods for solving the crystal structure of a small molecule, the most commonly applied is the direct method. Crystal structure may be solved using SHELXS,<sup>262</sup> which is particularly effective if a large proportion of the data is observed in the 1.0 Å region of reciprocal space and if no significant problems regarding the choice of space group exist.

Direct methods require that the structure factors be placed on an absolute scale. The magnitude of the structure factors depends upon the extent to which atoms co-operatively scatter and also upon the scattering angles themselves, where scattering occurs with decreased strength at high angles. Contribution from reflections with high structure factor amplitude tend to dominate any map calculated, including those with phases corresponding to the correct structure. Hence, only these high-value terms need to be initially considered in an attempt to obtain an approximation to the correct map.

Direct methods for both centrosymmetric and non-centrosymmetric structures are most successful when care is taken, initially, to find the first few phases, and to observe that reflections with indices related to those of many other strong reflections are used in the early stages. It is possible to derive relations among the phases of different reflections as there are limits to the possible phase angles for individual reflections even in non-centrosymmetric structures, but especially in centrosymmetric structures, where the phase angle of a structure factor  $F(hkl)$  is either 0° or 180°. These relations arise as a consequence that the electron density can never be negative and that it is near zero except for isolated resolved peaks at atomic positions. Hence, from their relationships it is often possible to derive phases for almost all strong reflections and so determine the structure from the resulting approximate electron density map. For centrosymmetric structures, relations can be found amongst these structure factors which involve the magnitudes of the

larger structure factors normalised in a certain way. The calculation of normalised structure factors from  $F_o$  values, affords the distribution of intensities. For a non-centrosymmetric structure this should follow the theoretical values for an acentric distribution, and for a centrosymmetric structure it should be centric. However, the presence of other symmetry elements may shift the distribution to centric and hypercentric respectively.

If the atoms can be located, the phase angle may approximate closely to the real value. Subsequently, an approximate electron density map can be calculated and interpreted using observed structure amplitudes and approximate phase angles, to yield a blend of the true structure with the trial structure.

#### 4.9 Structural Refinement

Repetitive refinement is necessary as the initially estimated phases may give a poor image of the scattering matter, and because the least squares equations are not linear. Refinement has thus been extended to the problem of fitting the observed diffraction intensities with those calculated. Refinement of the structure can be initiated once approximate positions have been determined. Atomic parameters are systematically varied to yield the best possible agreement of the observed structure factor amplitudes with those calculated for the proposed structure, to afford a 3-D structural model with significantly improved bond angles and lengths. Many successive iterative refinement cycles are usually needed before the structure reaches convergence as shifts in the atomic parameters, from cycle to cycle with respect to the expected experimental errors, are negligible. There are two common refinement techniques used: Fourier syntheses and the least squares method.

##### 4.9.1 Fourier Syntheses

The purpose of the Fourier Series is to calculate electron density over the whole unit cell starting from previously unscaled data in the form of  $h, k, l, F$ , and  $\alpha$ . However, the input amplitudes and phases of the structure factors possess a certain degree of inaccuracy, rendering the process as only an approximation. The solution of the crystal structure results from successive approximations, and is reliant upon the relation of the repeating structure and diffraction pattern, via Fourier summations.

Though Fourier syntheses are not as suitable for refinement as other techniques and cannot be used to refine scale and thermal parameters, if atoms are resolved, then fairly accurate positional parameters for the atoms can be determined. Subsequently, once the *approximate* positions and identities of all the atoms in the asymmetric unit are known, the amplitudes and phases of the structure factors can be readily calculated. The amplitudes and phases of the structure factors *calculated* from an approximation method will be to some degree incorrect, but at least a crude approximation is obtained, where the amplitudes can be replaced with measured values. Hence, one can calculate an

approximation to the true electron density by a 3-D Fourier summation of the experimentally *observed* structure factor amplitudes with the *calculated* phases which hold increased importance.

If the positions of the atoms in the structure are known, the calculated phase angle may be nearly correct. Subsequently, an approximate electron density map calculated with observed structure amplitudes and computed phase angles will contain a blend of the true structure with the trial structure used to compute the phase angles. If the trial structure contains most of the atoms of the true structure, at or near their true positions, the resulting electron density map will contain peaks near sites representing atoms which were omitted from the trial structure, in addition to those that were previously located.

In centrosymmetric structures, as the phase angles are either  $0^\circ$  or  $180^\circ$ , a slight error in the structure may not have any effect upon most of the phase angles. A map computed with observed  $|F|$  and computed phase angles may be correct even if the original model possessed some slight error. However, with non-centrosymmetric structures, for which the phase angles may have any values from  $0^\circ$  to  $360^\circ$ , there will be at least some errors in most of the phases, and consequently the calculated electron density map will be weighted more in the direction of the model used to calculate the phases than with a centrosymmetric structure. Once most of the known structure has been derived, difference maps rather than electron density maps are computed, where the coefficients for the calculation are the differences between  $F_o$  and  $F_c$  with the computed phase angle. This affords a map in which the positive region is representative of an area where too few electrons were included; a negative region represents the converse. If possible, H atoms are located from electron difference maps; any remaining H atoms are assumed to ride on attached atoms and are placed in calculated positions.

The main uses of Fourier syntheses in crystal structure determination are :

- 1) Patterson functions
- 2) Initial maps for a few data phased by direct methods
- 3) Difference maps to expand partially solved structures
- 4) Location of hydrogen atoms
- 5) Final checking of a structure

The major problem with Fourier sums is that the phases, which are vital to the summations, are at best only approximately known. As a Fourier synthesis is the best way of knowing whether a crystal structure has been solved, variations try to maximise useful data. During the refinement of the structure, if the magnitude of the observed structure factor,  $F_o$ , is markedly lower compared to the calculated structure factor  $F_c$ , the correction for extinction and refinement is carried out by SHELXL.<sup>262</sup> An excellent determination of

attainment of a good structure is a flat difference map at the end of refinement.

#### 4.9.2 Least Squares

The least squares method originally proposed by L egendre<sup>263</sup> depends, for its success, on the availability at the start of a reasonably good set of phases, that is a good trial structure where it makes the sum of the squares of the errors,  $\Delta F = |F_o| - |F_c|$ , a minimum. A cyclic process ensures an improved value for the individual parameters, until no further improvement is observed. Any anomalies in molecular geometry or packing should be scrutinised with great care and regarded with some scepticism. Since there are many more observations than parameters to be determined, statistical methods are used to fit an appropriate equation expressing the errors in  $|F|$  that result from errors in the trial structure.

At the conclusion of any least squares refinement process, it is always wise to calculate a difference Fourier synthesis. After the structure has been fully refined, parameters are obtained from the least-squares refinement as a set of co-ordinates and displacement parameters for each atom from which geometrical parameters, including bond lengths and angles, torsion angles, intermolecular and other non-bonded distances and least-squares planes with angles between can be calculated. Movement of atoms are calculated and corrections of apparent geometrical values applied to every derived result, an estimated standard deviation as a measure of its precision or reliability applied. As the technique is so suitable for high-speed computation, least squares refinement is the most common method used in structure improvement.

SHELXL93<sup>262</sup> refinement is based on full-matrix least-squares which calculates structure factors and accumulated least square totals solved later for parameter changes. Further refinement of parameters includes atomic co-ordinates, anisotropic thermal parameters (non-hydrogen atoms) and atomic isotropic vibration parameters (hydrogen atoms), scale factor for  $|F_o|$ , and an extinction parameter. Small variations in these parameters are made on each cycle to produce a test agreement between  $F_o$  and  $F_c$ .

#### 4.9.3 Weighting Scheme

The functions minimised in the least squares method carry a weighting factor for each observation. The weighting should be a measure of the reliability of each observation. If chosen correctly, the weighting factor is able to adjust the contribution of each observation to the normal equations such that the most reliable results are produced.

#### 4.10.1 Determination of Structure Correctness

The general criteria for assessing the likely correctness of a structure, that is, if the experimentally observed data are reasonably precise, and the model is appropriate, depend upon:

- a) The agreement of the individual observed structure factor amplitudes  $|F_o|$  with those calculated for the refined model  $|F_c|$ , which should be comparable to the estimated precision of the experimental measurements of the structure factors.
- b) The electron density difference map, phased with the final parameters of the refined model which should reveal no fluctuations in electron density greater than those expected on the basis of the estimated precision of the electron density.
- c) Any anomalies in the molecular geometry and packing or other derived quantities which should be scrutinised.

#### 4.10.2 The Discrepancy Factor ( $R$ )

The agreement index known as the  $R$  factor depends on the type and complexity of the structure, and on the quality of the experimental data. It is a measure of the precision of the derived structure, denoting how well the calculated model fits the observed data. The lower the  $R$  factor, the greater the confidence placed in the quality of the refined model.

$$R = \sum |\Delta F| / \sum |F_o| = (\sum ||F_o| - |F_c||) / \sum |F_o|$$

If initial phases are poor, the first approximations to the electron density will contain significant incorrect detail with peaks near, but isolated from the correct location, and the  $R$  factor will initially be large. The reliability is assessed by checking the estimated standard deviations of the derived results and inspection of those measures regarding the agreement for the values of  $|F_o|$  with the values of  $|F_c|$ , as indicated by the absence of any unexplained peaks in a final difference map, together with chemical integrity of the resulting structure.

#### 4.11 Molecular Plotting

In an effort to gain an increasingly detailed understanding regarding the conformation of molecules, application of molecular modelling acts as a complementary technique to X-ray crystallography, to provide a representation of the molecules in their lowest energy state, when isolated, in solution or within the solid crystal. The proposed model attained can be displayed using the ORTEP<sup>264</sup> program. Atoms in a molecule can be represented as ellipsoids whose size is related to the displacement parameters. If any ellipsoids are large or elongated, they may indicate atoms undergoing large thermal vibration. Alternatively, they may warn of undetected disorder or errors in the data. Use of these ORTEP drawings<sup>263</sup> also involves symmetry application for model assignment of the unit cell, thus enabling determination of molecular packing and intermolecular bonding.

Chem-X is a suite of programs produced by Chemical Design, to allow plotting from the monomer level, of either the unit cell contents or of individual molecules, to the macromolecular level, for projection and perspective or stereo representation. Software tools are provided for the construction and manipulation of molecules, upon which a variety of chemical and simple molecular mechanic calculations can be performed, affording energy contour maps, and allowing a basic, conformational analysis of individual molecules. QUANTA,<sup>266</sup> supplied by Molecular Simulations enables a dynamic display of up to 4500 atoms simultaneously, as well as docking of several molecules, to allow the visualisation of substrate-receptor interactions. Within QUANTA itself, intermolecular distances between two molecules may be displayed allowing the identification of possible H-bonds that are likely to be formed between adjacent structures.

# CHAPTER 5

## Structural Analysis of PNA Building Block Intermediates

### 5.1 X-Ray Structural Analysis and Evaluation

This chapter deals principally with the application of X-ray crystallographic analysis for evaluating the structural features and aspects of the crystalline compounds synthesised in this thesis work. All of the structures discussed here with the exception of the known oxime **20**, have now been published.<sup>217,232,239,241,242,247-249</sup> Crystal Information Files (CIF), supplementary data and the complete structure factor tables for **76**, **125**, **126**, **80**, **92**, **84**, **124** and **17** are all available on-line from the International Union of Crystallography (IUCr) and directly from the Cambridge Structural Database (CSD), which is also accessible via the Chemical Database system at Daresbury, Chester.

#### 5.2.1 The Value of X-Ray Crystallography for Structural Analysis of PNA Building Block Precursors

The primary intention of each crystallographic analysis undertaken in this work was to confirm and provide unequivocal proof of the crystal structure. Where alkylation reactions provided separable mixtures of regioisomers, comparison of analytical data derived for each member of the analogously substituted purine series with those of published literature allowed tentative assignment of regiochemistry. Spectroscopic techniques alone were not able to conclusively differentiate between regioisomers from all of the product mixtures. X-ray crystallographic analysis was thus particularly useful in providing unambiguous proof of regiochemical differentiation between 6-chloro-9-(carboxymethyl)purine ethyl ester **125** and 6-chloro-7-(carboxymethyl)purine ethyl ester **126**.<sup>248,249</sup> Crystallographic analysis was also necessary to confirm the procurement of the required N9 regioisomer of 2,6-dichloro-(carboxymethyl)purine ethyl ester analogue **76** from a mixture of two regioisomers.<sup>232</sup> Finally, crystallographic analyses were essential for characterising those products whose identity could not be determined using analytical and spectroscopic techniques alone.

#### 5.2.2 The Focal Points of This Crystallographic Work

The novel and interesting features arising from crystallographic data in this work were:

- Regioisomeric assignment and comparison of the N9 **80** and **125** with the N7 **126** chloro homologues, and assessment of the effects of regioisomerism on molecular structure.
- Evaluation of changes in purine ring geometry arising from replacement of chloro with amino, azido and methoxy substituents.



- Determination of the substitution pattern the peralkylated hypoxanthyl derivative **124**.<sup>247</sup>
- Comparison of the molecular interactions in the new series of substituted purines synthesised in this work, including H-bonding interactions and base-stacking arrangements.

### 5.3 Methods

Samples for all compounds under study in this work were recrystallised from mixtures of EtOAc-MeOH. X-ray crystallographic data were collected at 293 K. General methods for data collection are given in the experimental section (Chapter 6).

### 5.4 Comparison of N9 6-Chloropurine **125** with the Regioisomeric N7 6-Chloropurine **126** and the Effect of Chloro Substitution at C2

#### 5.4.1 Crystal Data for 6-Chloro-(9-carboxymethyl)purine ethyl ester<sup>248</sup> **125**

A single crystal (0.35 x 0.15 x 0.1 mm) of **125** (C<sub>9</sub>H<sub>9</sub>ClN<sub>4</sub>O<sub>2</sub>) crystallised in the triclinic space group  $P\bar{1}$ ,  $Z = 4$ . Geometrical parameters were:  $a = 5.872$  (2),  $b = 10.529$  (2),  $c = 19.096$  (5) Å,  $\alpha = 105.01$  (2),  $\beta = 93.85$  (2),  $\gamma = 105.00$  (2) °, and  $V = 1090.1$  (5) Å<sup>3</sup>.  $D_x = 1.466$  g/cm<sup>3</sup>;  $\Delta\rho_{\max} = 0.597$  e Å<sup>-3</sup>;  $\Delta\rho_{\min} = -0.256$  e Å<sup>-3</sup>;  $R(F) = 0.0780$ ;  $wR(F^2) = 0.2343$ ;  $\mu = 3.066$  mm<sup>-1</sup> and  $R_{\text{int}} = 0.027$ .

#### 5.4.2 Crystal Data for 6-Chloro-(7-carboxymethyl)purine ethyl ester<sup>249</sup> **126**

A single crystal (0.2 x 0.2 x 0.35 mm) of **126** (C<sub>9</sub>H<sub>9</sub>ClN<sub>4</sub>O<sub>2</sub>) crystallised in the monoclinic space group  $P2_1/c$ ,  $Z = 4$ . Geometrical parameters were:  $a = 7.724$  (6),  $b = 9.991$  (14),  $c = 13.820$  (3) Å,  $\alpha = 90$ ,  $\beta = 91.006$  (11),  $\gamma = 90$  °, and  $V = 1066.4$  (3) Å<sup>3</sup>.  $D_x = 1.499$  g/cm<sup>3</sup>;  $\Delta\rho_{\max} = 0.225$  e Å<sup>-3</sup>;  $\Delta\rho_{\min} = -0.256$  e Å<sup>-3</sup>;  $R(F) = 0.0339$ ;  $wR(F^2) = 0.0824$ ;  $\mu = 3.313$  mm<sup>-1</sup> and  $R_{\text{int}} = 0.077$ .

#### 5.4.3 2,6-Dichloro-(9-carboxymethyl)purine ethyl ester<sup>232</sup> **76**

A single crystal (0.5 x 0.5 x 0.25 mm) of **76** (C<sub>9</sub>H<sub>8</sub>Cl<sub>2</sub>N<sub>4</sub>O<sub>2</sub>) crystallised in the monoclinic space group  $C2/c$ ,  $Z = 8$ . Geometric parameters were:  $a = 25.585$  (14),  $b = 11.841$  (2),  $c = 8.012$  (3) Å,  $\alpha = 90$ ,  $\beta = 104.03$  (2),  $\gamma = 90$  ° and  $V = 2354.8$  (10) Å<sup>3</sup>.  $D_x = 1.552$  g/cm<sup>3</sup>;  $\Delta\rho_{\max} = 0.267$  e Å<sup>-3</sup>;  $\Delta\rho_{\min} = -0.207$  e Å<sup>-3</sup>;  $R(F) = 0.0353$ ;  $wR(F^2) = 0.1064$ ;  $\mu = 0.548$  mm<sup>-1</sup> and  $R_{\text{int}} = 0.0271$ .

Although the direct alkylation of 2,6-dichloropurine<sup>232</sup> **75** and 6-chloropurine<sup>248,249</sup> **89** afforded separable mixtures of regioisomeric products, crystal structure analysis was undertaken to provide proof for regioisomeric assignment of the respective regioisomers

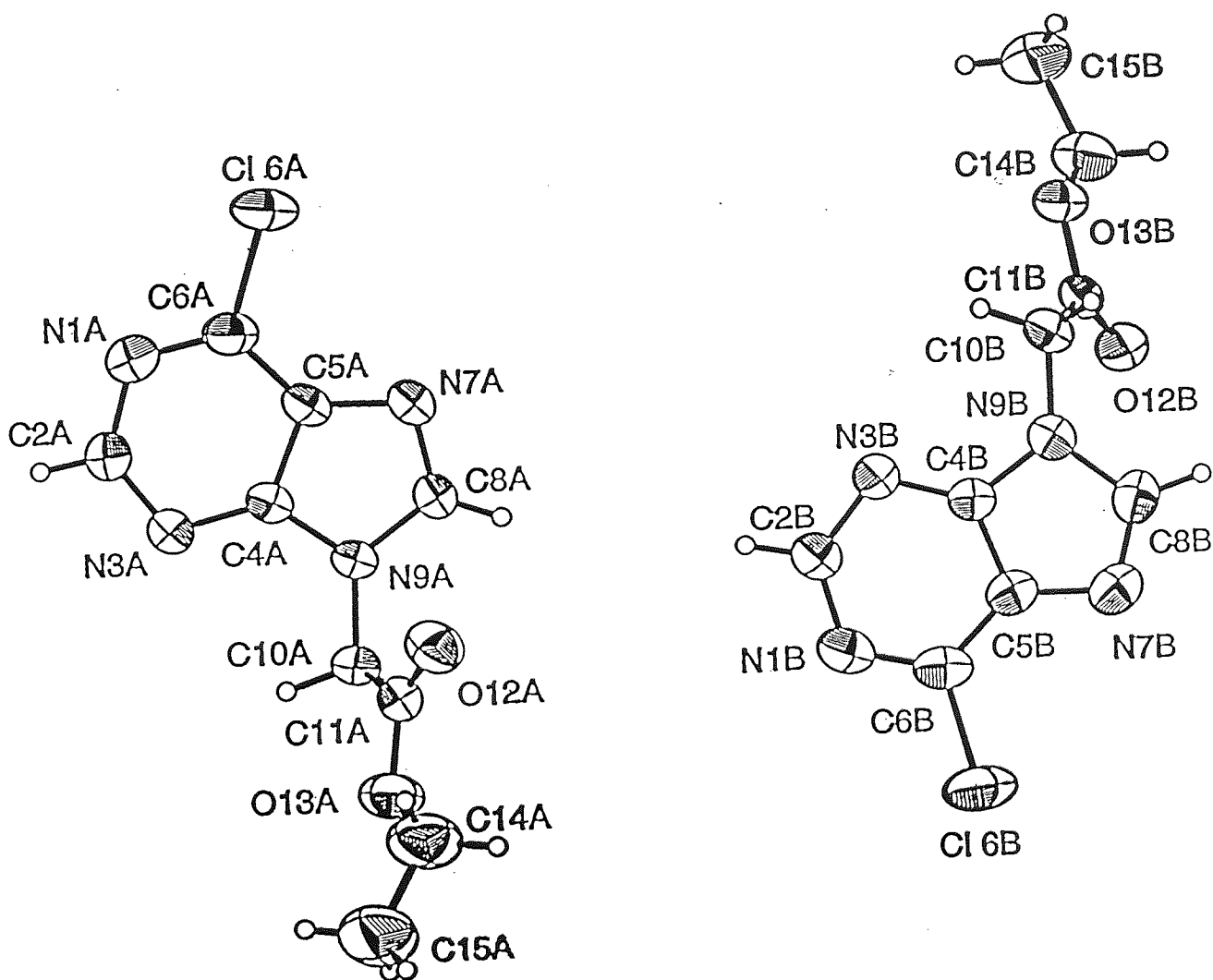
**76**, and **125** and **126**. The effects of chloro substitution were assessed by comparing the structural features of disubstituted chloro N9 product **75** (Figure 5.3) with its monosubstituted chloro analogues **125** (Figure 5.1) and **126** (Figure 5.2).

Two molecules (**125A**) and (**125B**), constitute the asymmetric unit of the N9 regioisomer (Figure 5.1) with just one molecule in **126** (Figure 5.2). The ring atoms in **125A** and **125B** are co-planar within  $\pm 0.006$  Å r.m.s. and  $\pm 0.009$  Å r.m.s. respectively; those in **126** are co-planar within  $\pm 0.006$  Å r.m.s. Bond distances in **125A** and **125B**, and **126** are mostly similar except for a distinct variation observed in the N7-C8 and C8-N9 bonds attributable to the substitution at these sites (Table 5.1). As a consequence, neighbouring bonds C4-N9 and C5-N7 exhibit variations in their respective bond lengths. From these data it is apparent that the respective bond length increases at the site of substitution; where N9 substitution occurs in **125A** and **125B**, and **76**, C8-N9 is 1.367 (4) and 1.377 (5) Å, and 1.376 (3) Å respectively, whereas C8-N9 is considerably shorter at 1.303 (3) Å in **126**. Similarly, substitution at N7 for **126** exhibits an N7-C8 bond distance of 1.363 (2) Å compared with a much shorter N7-C8 bond distances of 1.311 (5) and 1.301 (6), and 1.303 (3) Å for **125A** and **125B**, and **126**, respectively.

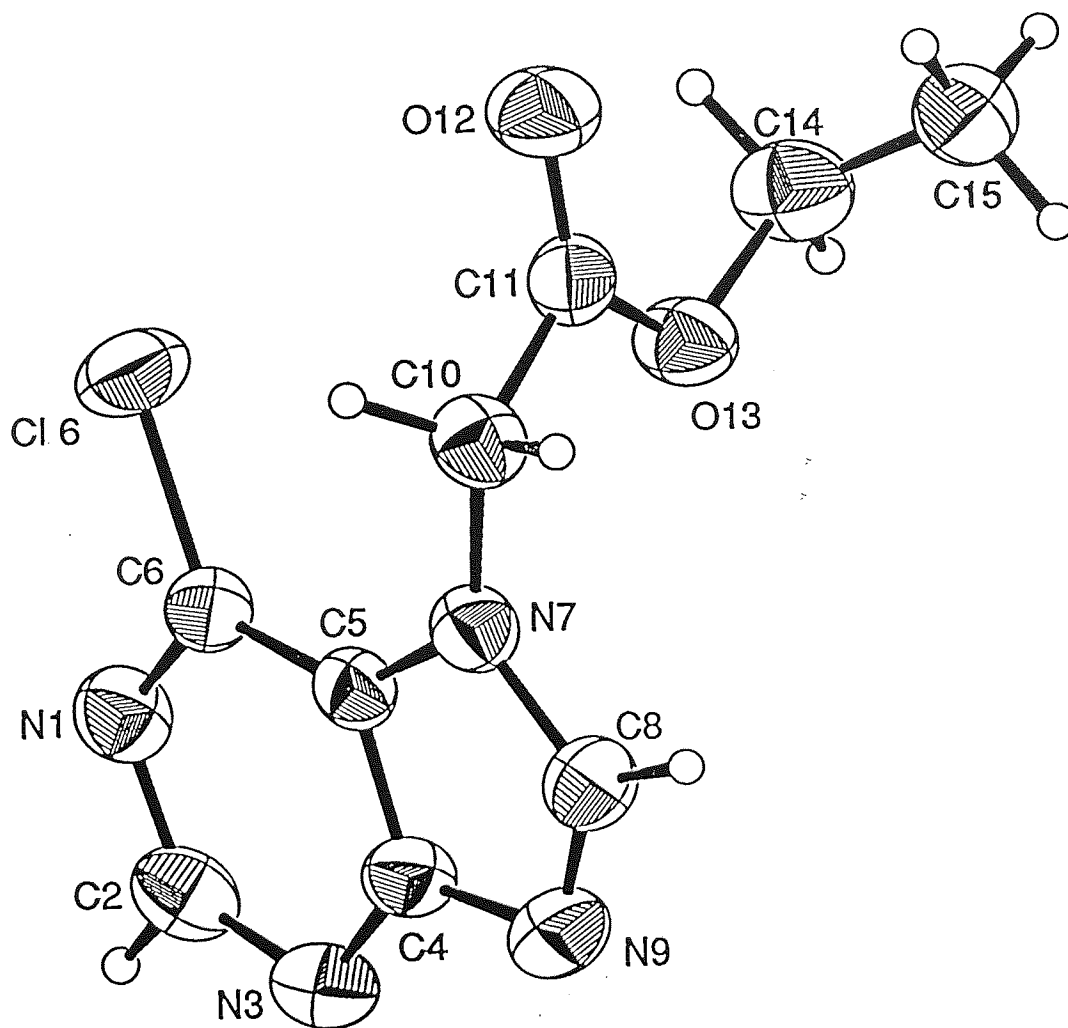
Comparison of the internal angles at N7, N9, C4 and C5 of the N9 isomer **125** with the N7 isomer **126**, illustrates how regioisomerism affects ring geometry. The corresponding C5-C4-N9, N7-C5-C4, C5-N7-C8, and C4-N9-C8 bond angles for the chloro analogues **76**, **125** and **126**, are summarised in Table 5.1. Attachment of the ethyl ester in **125** (Figure 5.1) at the N9 site, contracts C5-N7-C8 by 1.5 (4) ° and expands C4-N9-C8 by 1.45 (14) ° (Table 5.1) compared to **126**, where the side chain is attached at N7 (Figure 5.2).<sup>248,249</sup> From the observed increase in the internal angles at the respective substituted sites, it appears substitution at the respective N7 or N9 site relieves a certain degree of ring strain compared with the unsubstituted heterocyclic rings. Compared with **126** the internal angles within the five-membered ring of **125A** and **125B** are expanded at C5 by 4.3 (3) and 4.7 (3) °, and at N9 by 1.8 (13) and 1.3 (13) °, whereas bond angles at C4 and N7 are contracted by -4.2 (15) °, and -1.4 (14) and -1.3 (14) °, respectively.

In both independent molecules of **125** the side chain adopts an extended conformation; its carboxymethyl group plane intersects the very planar heterocycle at 60.1 (1) and 65.8 (1) ° respectively. The ethyl acetate side chain of **125A** and **125B** both avoid steric hindrance with the heterocycle by emerging from N9 with torsion angle C8-N9-C10-C11 equal to 58.2 (5) ° in **125A** and 65.1 (5) ° in **125B**. The **125A** side chain is almost linear with respect to the heterocyclic ring with torsion angles C9-C10-C11-O13 174.6 (3), C10-C11-O13-C14 -176.6 (4), C11-O13-C14-C15 159.4 (5) and C11-O13-C14-C15 159.4 (5) °; corresponding values of 168.9 (3), -177.9 (3), 169.1 (3) and 169.1 (3) ° are observed with the same substituent in **125B** (Figure 5.1).

The asymmetric unit of the N9 regioisomer **125** is comprised of two molecules (**125A**) and (**125B**)<sup>248</sup>.



**Figure 5.1.** ORTEPII view<sup>263</sup> of 6-chloro-(9-carboxymethyl)purine ethyl ester<sup>248</sup> **125** showing the labelling scheme for non-H atoms with thermal ellipsoids drawn at the 50% probability level and H atoms shown as small spheres with arbitrary radii.



**Figure 5.2.** ORTEP view<sup>263</sup> of 6-chloro-(7-carboxymethyl)purine ethyl ester<sup>249</sup> **126** showing the labelling scheme for non-H atoms with thermal ellipsoids drawn at the 50% probability level and H atoms shown as small spheres with arbitrary radii.

The ethyl acetate fragment attached at N7 in **126** avoids steric hindrance with both the heterocycle and the chlorine atom, by emerging almost orthogonally and continuing in an extended conformation with C8-N7-C10-C11 106.1 (2) ° and with C11-O13-C14-C15 169.3 (1) ° (Figure 5.2). Hence, all ester chains can be classified as essentially 180 ° *trans*.

**Table 5.1.** Selected geometrical parameters: bond lengths, bond angles and bond torsion angles for **76**, **125A**, **125B**, and **126**.

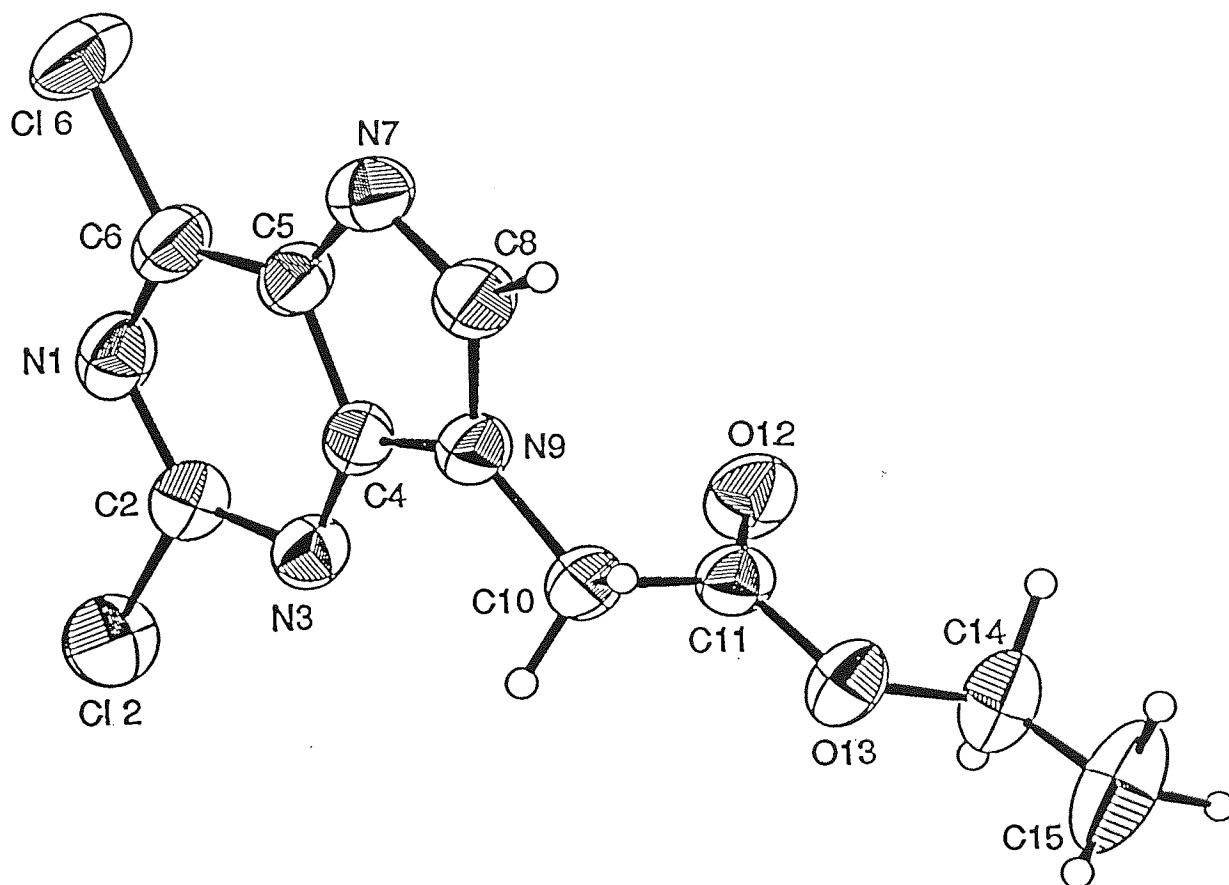
Geometrical Parameter	<b>76</b>	<b>125A</b>	<b>125B</b>	<b>126</b>
C4-N9	1.361 (3)	1.359 (4)	1.366 (4)	1.376 (2)
C5-N7	1.381 (3)	1.381 (4)	1.380 (5)	1.374 (4)
C8-N9	1.376 (3)	1.367 (4)	1.377 (5)	1.303 (3)
N1-C2	1.331 (3)	1.349 (5)	1.353 (5)	1.344 (2)
N1-C6	1.326 (4)	1.308 (5)	1.313 (5)	1.308 (2)
N7-C8	1.303 (3)	1.311 (5)	1.301 (6)	1.363 (2)
N1-C2-N3	129.8 (2)	128.0 (3)	128.2 (3)	127.9 (2)
N3-C4-C5	126.5 (2)	125.9 (3)	126.6 (3)	123.84 (14)
C5-C4-N9	105.6 (2)	105.6 (3)	105.6 (3)	109.8 1 (15)
C4-C5-N7	110.8 (2)	110.3 (3)	110.7 (3)	106.03 (12)
N1-C6-C5	121.5 (2)	122.0 (3)	121.4 (3)	121.06 (14)
C2-N1-C6	116.3 (2)	117.3 (3)	117.1 (3)	117.38 (14)
C2-N3-C4	111.0 (2)	112.2 (3)	111.7 (3)	113.44 (14)
C5-N7-C8	103.6 (2)	103.5 (3)	103.8 (3)	105.05 (14)
C5-N7-C10	-	-	-	128.59 (13)
C4-N9-C8	105.8 (2)	106.0 (3)	105.5 (3)	104.24 (13)
C4-N9-C10	127.0 (2)	126.5 (3)	126.2 (3)	-
C8-N9-C10	126.9 (2)	127.2 (3)	128.0 (3)	-
C8-N7-C10	-	-	-	126.37 (13)
C4-N9-C10-C11	81.0 (3)	-115.7 (4)	-107.7 (4)	-
C5-N7-C10-C11	-	-	-	-73.9 (2)
C8-N9-C10-C11	-92.9 (3)	58.2 (5)	65.1 (5)	-
C8-N7-C10-C11	-	-	-	106.1 (2)
N7-C10-C11-O13	-	-	-	161.3 (2)
N9-C10-C11-O13	166.0 (2)	174.6 (3)	168.9 (3)	-
C10-C11-O13-C14	178.9 (3)	-176.6 (4)	-177.9 (3)	-177.6 (4)
C11-O13-C14-C15	166.8 (4)	159.4 (5)	169.1 (3)	169.3 (1)

Comparison of dichloro- with monochloro substitution of the heterocyclic ring was evaluated via comparison of endocyclic bond angles and bond lengths of the heterocyclic ring of **76**, with **125** and **126**.

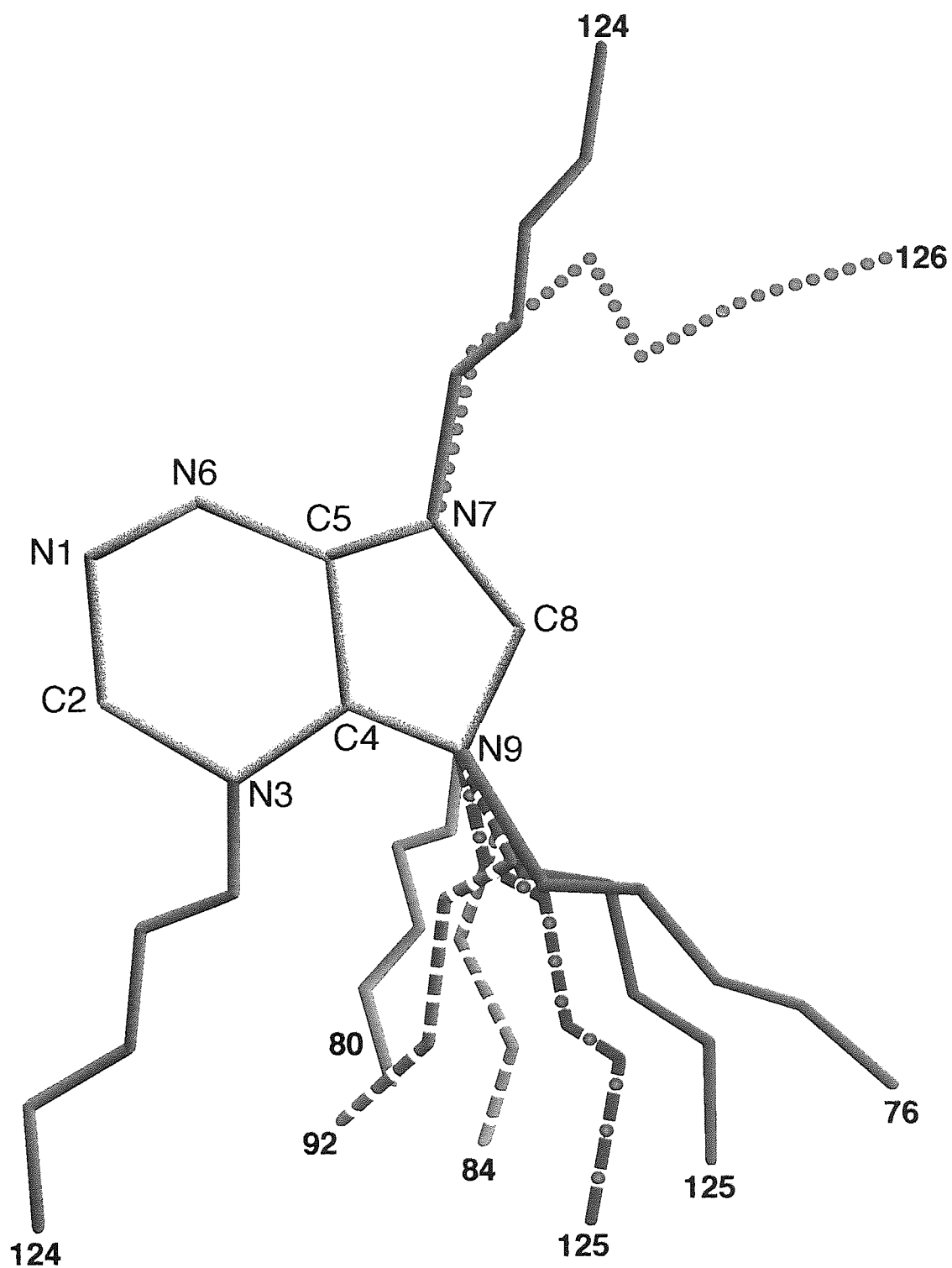
#### 5.4.4 Geometrical Features of 2,6-Dichloro-(9-carboxymethyl)purine ethyl ester **76**

Though the C2-Cl2 bond at 1.730 (3) Å is marginally longer than the bond which connects C6 to the more reactive Cl6 at 1.721 (2) Å, the C-Cl bond distances (Table 5.1) do not support a simplistic correlation of ground state bond length and strength with reactivity towards nucleophiles. An explanation for the less reactive C2-Cl2 bond is based on the stability of the resonance forms as previously discussed (Section 3.7). Chloro substitution upon most heterocycles is expected to expand the endocyclic bond angle at the substituted site, presumably via withdrawing electrons and diminishing repulsion of adjacent bonds. Whereas the average N1-C2-N3 and N1-C6-C5 angles are 127.1 and 119.2 ° as calculated in three accurately determined *9H* tautomer purine structures in reported literature,<sup>267,268,269</sup> and 127.3 ° and 122.8 ° respectively in three derivatives of 6-chloropurine,<sup>270,271,272</sup> corresponding angles are 129.8 (2) and 121.5 (2) ° in **76** (Figure 5.3). Analysis of these data revealed chloro substitution on the heterocycle significantly expands the endocyclic bond angle at the attached site. Thus, N1-C2-N3 is 1.6-1.9 ° greater in **76** compared with analogues **125** and **126**. In chlorobenzenes, a corresponding increment of 1.4 ° was observed at the atom acquiring the chloro substituent.<sup>200</sup>

Many of the geometrical features of **76** resemble those of its monochloro analogue **125**, where the internal N9, N7, C5 and C6 bond angles are essentially similar, as shown in Table 5.1.<sup>21</sup> In the five-membered ring the internal angles at N7 and N9 remain largely unchanged implying chloro substitution predominantly affects only the pyrimidine moiety of the heterocyclic ring. The internal bond angles C2-N3-C4 and C2-N1-C6 in **76** (sites adjacent to that substituted with the additional chloro moiety) are slightly contracted by 1.0 (4) and 1.2 (4) ° compared with **125A** due to the additional electron-withdrawing Cl atom at C2. Corresponding values are shown in Table 5.1. A comparison between **126** and **76** involves both the addition of the 2-chloro substituent and a change of regioisomer. This was observed to cause sizeable alternating changes in the internal angles of the six-membered ring, where decreases of 1.1 (2), 2.4 (2) and 1.5 (2) ° at N1, N3 and C5 respectively, and increases of 1.9 (3), 2.7 (2) and 0.4 (2) ° at C2, C4 and C6 were observed. In the five-membered ring the internal angle at N7 decreased by 1.5 (2) °, whilst that at N9 increased by 1.6 (2) °, in agreement with the trend observed between **125** and **126**. The ethyl ester chain of the N9 monochloro analogue **125B** (Figure 5.1) emerges in a fairly similar orientation with respect to the ring as the same chain upon the N9 dichloro analogue **76** (Table 5.1).



**Figure 5.3.** ORTEP view<sup>263</sup> of 2,6-dichloro-(9-carboxymethyl)purine ethyl ester<sup>232</sup> **76** showing the labelling scheme for non-H atoms with thermal ellipsoids drawn at the 50% probability level and H atoms shown as small spheres with arbitrary radii.



**Figure 5.4.** Disposition of the methyl and ethyl ester side chains with respect to the heterocyclic ring for crystal structures 76, 80, 84, 92, 124, 125 and 126.



The variability in the orientation of the purine substituent is compared in Figure 5.4. For **76**, **125A**, **125B**, and **126**, the ester chains emerge in an almost linear orientation to avoid steric hindrance with the ring. The ester group is arranged so that the purine ring and the continuation of the chain are in a *trans* orientation with the ring, the carbonyl oxygen atom thereby becoming *cis*.

## 5.5 Replacement of the C2 and C6 Chloro Groups with Amino, Azido, and Methoxy Substituents.

### 5.5.1 Crystal Data for 2,6-Diamino-(9-carboxymethyl)purine ethyl ester<sup>239</sup> **80**

A single crystal (0.6 x 0.35 x 0.35 mm) of **80** (C<sub>9</sub>H<sub>12</sub>N<sub>6</sub>O<sub>2</sub>) crystallised in the triclinic space group  $P\bar{1}$ ,  $Z = 4$ . Geometric parameters were:  $a = 8.612$  (2),  $b = 11.658$  (3),  $c = 11.914$  (2) Å,  $\alpha = 79.72$  (2),  $\beta = 85.97$  (2),  $\gamma = 76.73$  (2) °, and  $V = 1145.0$  (4) Å<sup>3</sup>.  $D_x = 1.432$  g/cm<sup>3</sup>;  $\Delta\rho_{\max} = 0.165$  e Å<sup>-3</sup>;  $\Delta\rho_{\min} = -0.208$  e Å<sup>-3</sup>;  $R(F) = 0.0422$ ;  $wR(F^2) = 0.1413$ ;  $\mu = 0.914$  mm<sup>-1</sup> and  $R_{\text{int}} = 0.042$ .

### 5.5.2 Crystal Data for Ethyl-adenin-9-yl acetate<sup>240</sup> **94**

**92** (C<sub>9</sub>H<sub>11</sub>N<sub>5</sub>O<sub>2</sub>) crystallised in the orthorhombic space group  $P_{bca}$ ,  $Z = 8$ . Geometric parameters were:  $a = 19.4945$  (14),  $b = 12.2427$  (7),  $c = 8.5428$  (5) Å, and  $V = 2038.8$  (2) Å<sup>3</sup>.  $D_x = 1.441$  g/cm<sup>3</sup>;  $R(F) = 0.0318$ ;  $wR(F^2) = 0.0856$ ;  $\mu = 0.898$  mm<sup>-1</sup> and  $R_{\text{int}} = 0.0145$ .

### 5.5.3 Comparison of the Known Ethyl-Adenin-9-yl Acetate **94** with 2,6-Diamino-(9-carboxymethyl)purine ethyl ester<sup>239</sup> **80**

The effects imposed by the additional C2 amino group upon the conformation of the molecule, the orientation of the ring, and the ordering within the molecule, were evaluated via comparison of **80** with its mono-amino analogue, the reported adenine ethyl ester **94**. Figure 5.5 shows that **80**<sup>239</sup> displays significant disorder which is not apparent in its adenine analogue **94**.<sup>240</sup> The structure of molecule **80A** was revealed easily using direct methods, but only ambiguous indications of the second independent molecule in the asymmetric unit were observed. Successive electron density maps gave a consensus position for most of the remaining heterocycle atoms with clear indications of disorder in N9 and C8, and the N9 side chain of **80**. As the side chains in sites B and C (Figure 5.5) are well separated they were refined in the normal manner. However, as the ring atoms are so close together, ensuing unstable refinements led to distorted thermal ellipsoids. Molecules **80B** and **80C** were thus constructed and refined subject to 50% occupancy and the 'SAME' restraint in SHELXL,<sup>265</sup> which restrained bond and 1,3 distances within the heterocycles to similar values as those in molecule **80A**.

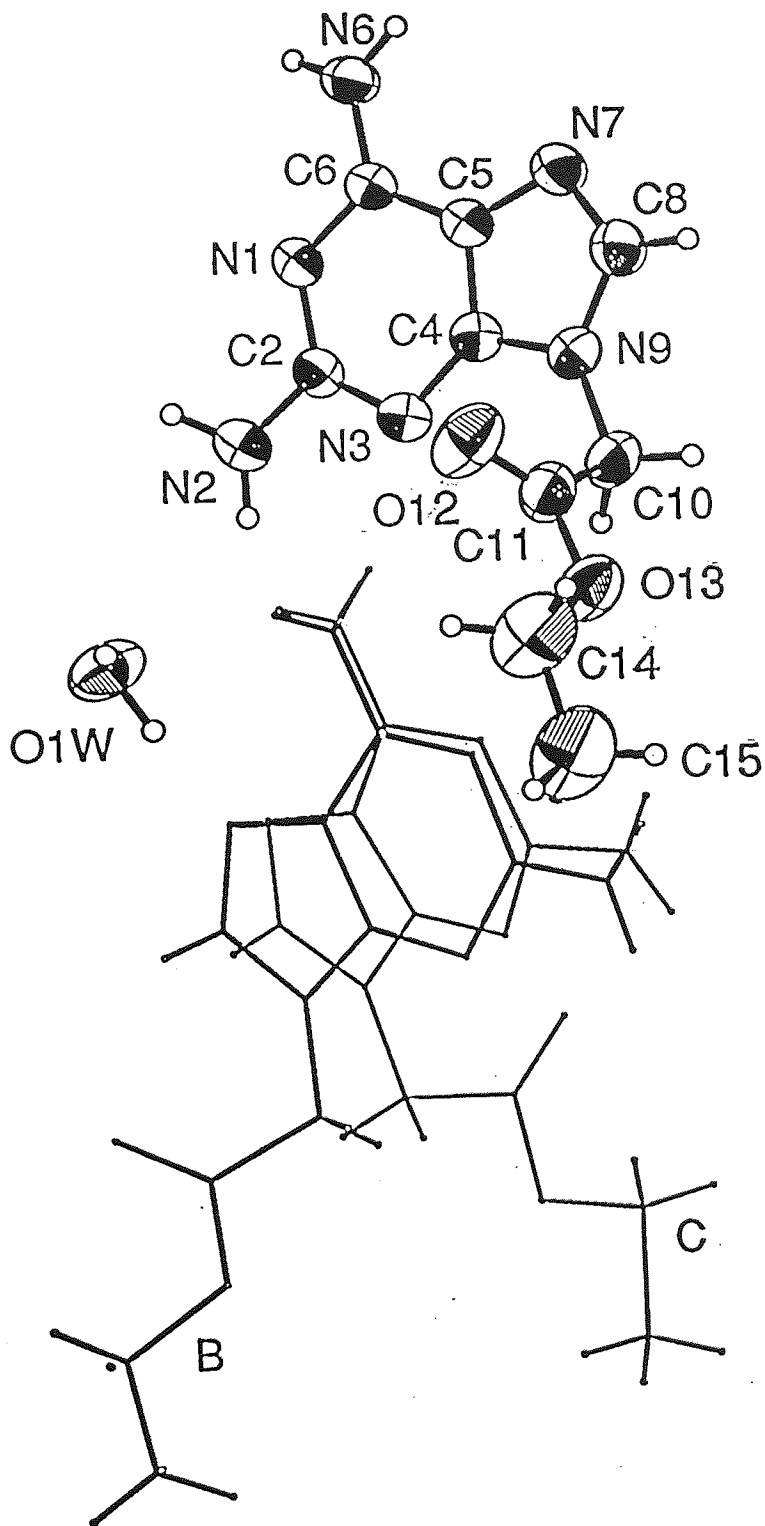


Figure 5.5. ORTEP view<sup>263</sup> of the independent molecule 2,6-diamino-(9-carboxymethyl)purine ethyl ester **80**. Two molecules of **80**, along with a molecule of water are present in the asymmetric unit; a well-ordered molecule (**80A**) with an accompanying disordered molecule **80B** or **80C**. For molecule **80A**, the displacement ellipsoids are drawn at the 50% probability level and the atom numbering scheme labelling scheme for non-H atoms shown as small spheres with arbitrary radii. The disordered molecule is depicted as a line drawing; of the two alternative sites, **80B** is drawn with thicker lines than **80C**.

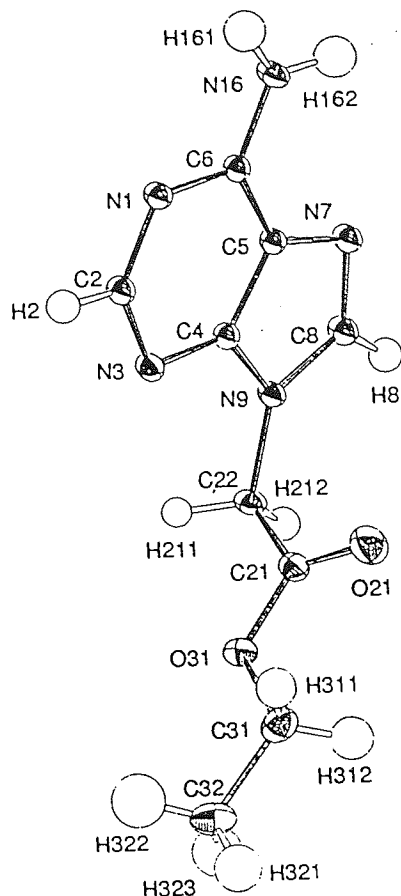


Figure 5.6. ORTEPII view<sup>263</sup> of ethyl-adenin-9-yl acetate<sup>240</sup> **94**.

In the crystal structure of **80** in its hemihydrate form  $C_9H_{12}N_6O_2 \cdot 0.5H_2O$ , the asymmetric unit therefore contains two molecules, one well-ordered molecule (**80A**) with an accompanying disordered molecule, either (**80B**) or (**80C**), with twofold positional disorder about a pivot point near the 6-amino group<sup>239</sup>. The H-bonded 6-amino group is almost invariant in molecules (**80B**) and (**80C**). Despite the initial difficulty in refining **80**, many geometric features of **80** resemble those previously found in its well-ordered and anhydrous adenine analogue **94** ethyl-adenin-9-yl acetate<sup>240</sup> (Figure 5.6).

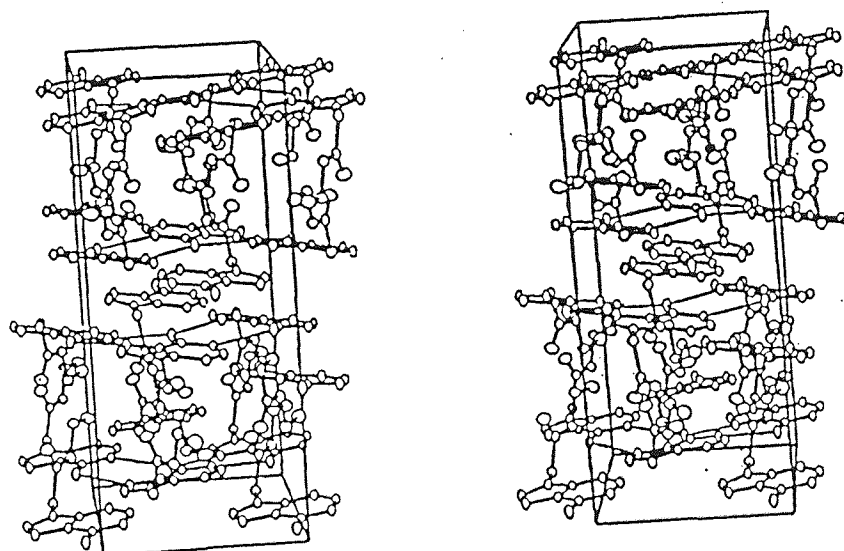
The ethyl acetate moiety of **94** which adopts an extended conformation, is virtually planar with a maximum deviation from the least squares plane of 0.04 (1) Å for O31, and almost perpendicular to the least squares plane defined by the adenine ring system<sup>240</sup>, where the angle between the two planes is 84.46 (4) ° and the torsion angle C8-N9-C10-C11 is 104.65 (13) °. This conformation differs significantly from the one found for other substituents in this position; in adenosine the equivalent torsion angle C8-C9-C22-C21 is 9.9 °. In molecule **80A**, atoms of the heterocycle are coplanar within 0.008 Å, similar to

its adenine analogue **94**. The ethyl ester chain at N9 in **80** emerges almost orthogonally from the heterocycle, where the torsion angle C8A-N9A-C10A-C11A is 93.0 (2) ° with the six atoms of the side chain coplanar within an r.m.s. deviation of  $\pm 0.092$  Å. Similarly, the side-chain atoms of (**80B**) and (**80C**) are almost coplanar, but with C8B-N9B-C10B-C11B and C8C-N9C-C10C-C11 torsion angles of 50.8 (6) and -97.0 (6) °, respectively. The orientation of the ethyl and methyl ester moieties for each of the substituted purines is revealed in Figure 5.4, where it can be seen that in order to avoid steric hindrance all chains emerge in an antiperiplanar disposition with respect to the heterocyclic ring.

**Table 5.2.** Selected geometrical parameters for **80**, **84**, **92** and **94**.

Compound	<b>80A</b>	<b>94</b>	<b>92</b>	<b>84</b>
N1-C2	1.353 (2)	1.3420 (14)	1.338 (2)	1.345 (4)
N1-C6	1.348 (2)	1.3567 (14)	1.347 (2)	1.324 (4)
C2-N3	1.340 (2)	1.315 (15)	1.326 (2)	1.313 (4)
C4-N9	1.373 (2)	1.3711 (14)	1.369 (2)	1.367 (4)
C5-N7	1.390 (2)	1.394 (14)	1.394 (2)	1.379 (4)
C2-R2	1.361 (2)	-	1.346 (2)	1.402 (4)
C6-R6	1.331 (2)	1.3402 (14)	1.330 (2)	1.389 (4)
N1-C2-N3	127.57 (12)	129.22 (10)	129.70 (12)	128.4 (3)
N3-C4-C5	127.63 (12)	127.73 (10)	128.07 (14)	127.3 (3)
C4-C5-C6	116.43 (12)	116.44 (10)	116.38 (12)	114.8 (3)
C5-C6-N1	118.12 (11)	117.48 (10)	118.44 (14)	120.1 (3)
C5-C4-N9	105.49 (11)	106.02 (9)	106.33 (11)	105.7 (2)
N7-C5-C4	110.83 (12)	110.33 (10)	110.37 (13)	111.1 (3)
N9-C8-N7	113.68 (13)	113.77 (10)	113.90 (12)	114.1 (3)
C6-N1-C2	119.15 (11)	118.93 (10)	118.71 (13)	117.8 (3)
C2-N3-C4	111.06 (11)	110.18 (9)	109.67 (13)	111.6 (3)
C4-N9-C8	106.05 (11)	106.15 (9)	105.70 (12)	105.6 (3)
C8-N7-C5	103.94 (12)	103.72 (9)	103.70 (12)	103.5 (3)
C4-N9-C10-C11	-78.0 (2)	-70.96 (14)	69.3 (2)	-76.8 (3)
C8-N9-C10-C11	93.0 (2)	104.65 (13)	-104.6 (2)	103.6 (3)
N9-C10-C11-O13	-178.25 (13)	152.62 (9)	-148.47 (13)	-177.0 (2)

The two H atoms of the amino group in **94** are potential H-bond donors; in addition to the N1, N3 and N7 atoms of the adenine ring, the O21 of the ethyl acetate fragment also has the ability to act as an H-bond acceptor. However, the H-bonding pattern is similar to that found in 9-methyl adenine with N16-H162...N1 and N16-H161...N7 H-bonds connecting molecules (equivalent to bonds in structures **80** and **94**), related by translational symmetry along the *c* axis. The amino group at N6 is involved in the donation of two intermolecular weak H-bonds to the ring N1 and ester carbonyl O11 atoms that created centrosymmetric dimers, which are further linked along the screw axis. The plane through the ring of **94** is almost parallel to the *ab* plane and the side chains orientated so the space between the adenine ring planes of neighbouring molecules, is filled (Figure 5.7)



**Figure 5.7.** Packing arrangement of ethyl-adenin-9-yl acetate **94**.

Compared with its adenine analogue<sup>240</sup> **94**, the N1-C2-N3 bond angle is compressed by 165 (16) ° in **80A** to accommodate the additional electron donating amino group at C2, while C5-C6-N1 is subsequently slightly expanded by 0.64 (15) °. Strong intermolecular H-bonds exist between the amino groups and the heterocyclic N atoms with an additional contact and weaker interaction to the water molecule which donates H-bonds to the N3 and N7 sites of molecule (**80B/C**) (Figure 5.5). The full use of donor and acceptor sites in molecule **80A** resembles the characteristic Watson-Crick three H-bond in the base pair that D is able to form with T, compared with the two H-bond base pair made by its adenine analogue **94**.

#### 5.5.4 Evaluation of the Structural Changes in **80**, **84** and **92** on Replacement of Chloro Substituents with Amino (**80**), Azido (**84**) and Methoxy (**92**) Substituents

##### 5.5.5 Crystal Data for 6-Amino-2-methoxy-(9-carboxymethyl)purine methyl ester<sup>241</sup> **92**

A single crystal (0.4 x 0.4 x 0.7 mm) of **92** (C<sub>9</sub>H<sub>11</sub>N<sub>5</sub>O<sub>3</sub>) crystallised in the monoclinic space group *P2<sub>1</sub>/n*, *Z* = 4. Geometrical parameters were: *a* = 11.085 (3), *b* = 7.7168 (6), *c* = 13.006 (2) Å,  $\alpha$  = 90,  $\beta$  = 99.97 (2),  $\gamma$  = 90 ° and *V* = 1095.8 (3) Å<sup>3</sup>. *D<sub>x</sub>* = 1.438 g/cm<sup>3</sup>;  $\Delta\rho_{\max}$  = 0.175 e Å<sup>-3</sup>;  $\Delta\rho_{\min}$  = -0.163 e Å<sup>-3</sup>; *R*(*F*) = 0.0380; *wR*(*F*<sup>2</sup>) = 0.1013;  $\mu$  = 0.947 mm<sup>-1</sup> and *R*<sub>int</sub> = 0.062.

##### 5.5.6 Crystal Data for 2,6-Diazido-(9-carboxymethyl)purine methyl ester<sup>242</sup> **84**

A single crystal (0.5 x 0.15 x 0.05 mm) of **84** (C<sub>8</sub>H<sub>10</sub>N<sub>10</sub>O<sub>2</sub>) crystallised in the monoclinic space group *P2<sub>1</sub>/c*, *Z* = 4. Geometrical parameters were: *a* = 12.761 (2), *b* = 12.128 (13), *c* = 7.997 (9) Å,  $\alpha$  = 90,  $\beta$  = 102.907 (10),  $\gamma$  = 90 ° and *V* = 1206.3 (3) Å<sup>3</sup>. *D<sub>x</sub>* = 1.510 g/cm<sup>3</sup>;  $\Delta\rho_{\max}$  = 0.266 e Å<sup>-3</sup>;  $\Delta\rho_{\min}$  = -0.270 e Å<sup>-3</sup>; *R*(*F*) = 0.0629; *wR*(*F*<sup>2</sup>) = 0.1961;  $\mu$  = 1.015 mm<sup>-1</sup> and *R*<sub>int</sub> = 0.0580.

Conventional analytical and spectroscopic techniques failed to disclose the identity of the unexpected by-product **92**, acquired as a consequence of using excess MeOH in azidation of the dichloro ethyl ester **76** followed by hydrogenolysis of the resulting diazido methyl ester **84** (section 3.6.1). X-ray crystallographic analysis served to prove that new groups had been introduced at the C2 and N9 sites.<sup>242</sup> From inspection of the ORTEP (Figure 5.8) plot, the C2 chloro group had been replaced by a methoxy group, and the methyl ester rather than the ethyl ester was present on N9.

The effects imposed upon the endocyclic bond angles and lengths as a consequence of chloro substitution for electron-donating amino **80**, azido **84**, or mesomeric electron donating methoxy groups **92**, are briefly surveyed. In these structures, the most pronounced effects were observed with changes in both the nature and position of the substituents. From Table 5.2, the bond angles and bond lengths at the key atoms of each of the purine rings are reflected in the electron donating or electron withdrawing nature of the ring substituents. Electron withdrawing chloro substituents in **76**, **125** and **126** have already been shown to expand the angle within the ring, whilst the electron donating NH<sub>2</sub> substituents, as for **80**, contract the angle.

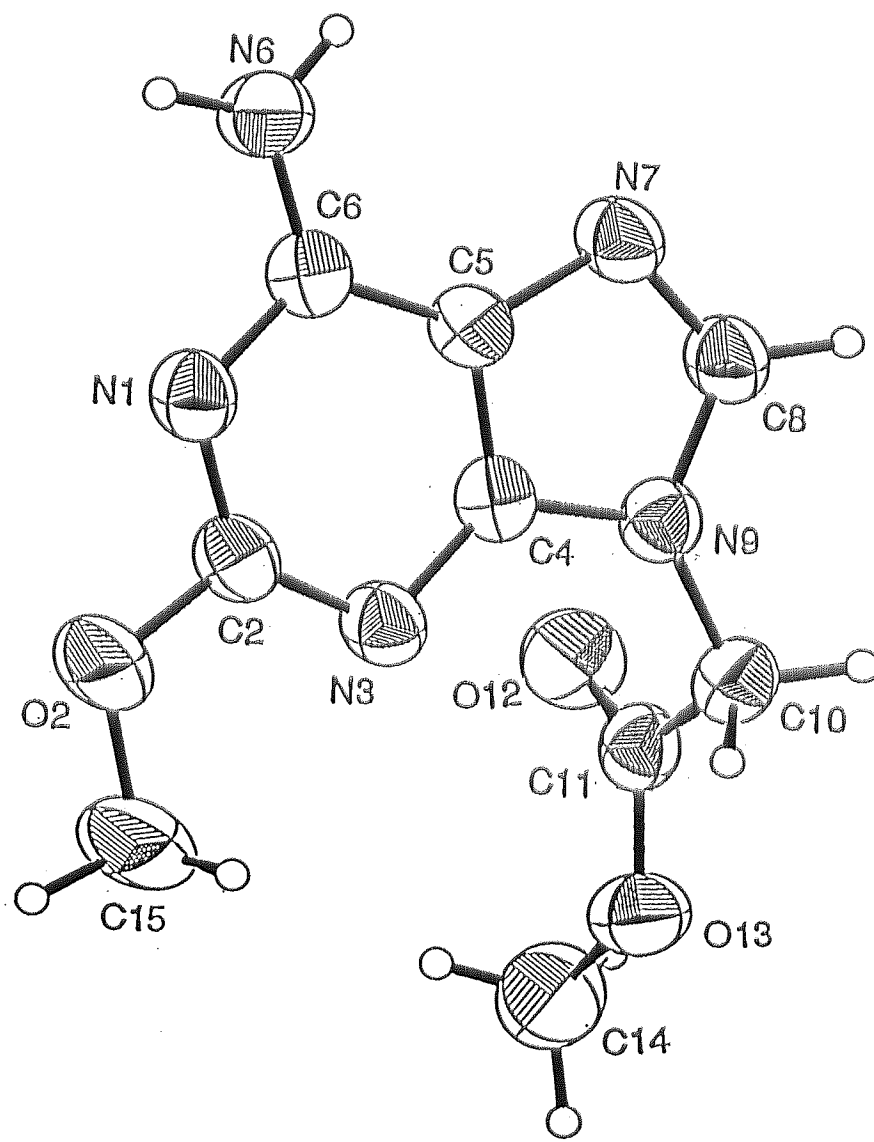


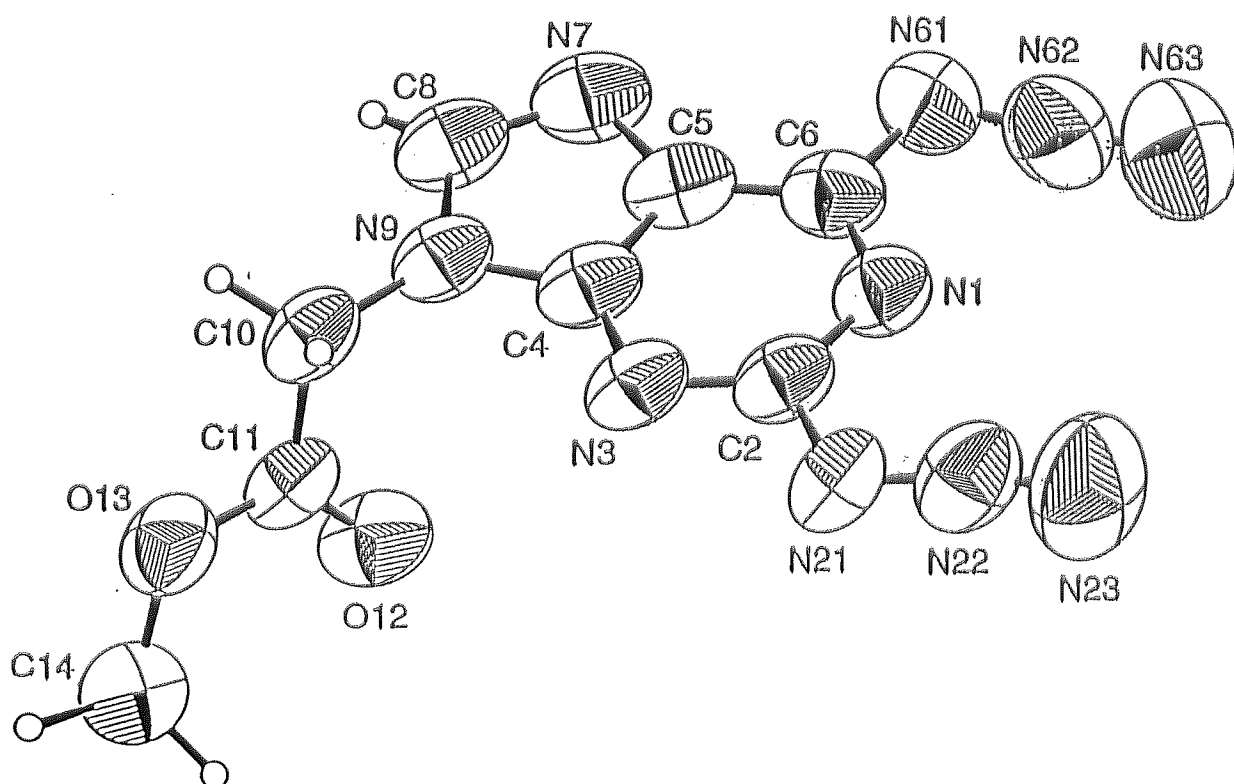
Figure 5.8 ORTEPII drawing<sup>263</sup> of 6-amino-2-methoxy-(9-carboxymethyl)purine methyl ester<sup>241</sup> 92, showing the labelling scheme for non-H atoms and the site and orientation of the side chain. Thermal ellipsoids are drawn at the 50% probability level. H atoms are shown as small spheres with arbitrary radii.

Compared to the known alkyl adenine analogue **94**, attachment of a methoxyl group at C2 in the homologue **92** expands N3-C2-N1 marginally by 0.48 (16) °. The presence of the 2-amino moiety in **80** contracts N3-C2-N1 by 2.13 (17) ° compared to amino-methoxy **92** and by 1.65 (16) ° compared to its ethyl adenin-9-yl analogue **94**. This phenomenon is caused by the presence of the additional lone pair of electrons on the nitrogen of the C2-amino group which becomes coupled with the electron rich environment of this moiety, and decreases these bond angles from normal trigonal geometry to minimise the repulsive intramolecular electrostatic forces. The internal angle N1-C6-C5 in amino-methoxy **92** is slightly compressed by 0.67 (17) ° and the C6-N6 bond length only marginally shorter by 0.005 (2) Å in **92** compared to its 2-amino-ethyl ester homologue **80**.<sup>239,242</sup>

As with all substituted purine analogues synthesised in this work, steric hindrance of the N9 side chain in **92** with the heterocycle was avoided by almost orthogonal emergence of the N9 side-chain with C8-N9-C10-C11 -104.6 (2) °. The atoms of the purine ring system are essentially coplanar within  $\pm 0.011$  (1) Å; the methyl group at O2 eclipsing N3 with N3-C2-O2-C14 1.7 (2) °.<sup>242</sup> Many geometrical features of the diazido compound **84** resemble those previously found in the dichloro **76** and diamino **80** ethyl ester homologues (Table 5.2).<sup>239,241</sup> The atoms of the heterocycle are coplanar within  $\pm 0.013$  (2) Å and the side chain avoids steric interference with the heterocycle by means of a large twist about the N9-C10 bond (Figure 5.9). The two azido groups of **84** emerge in a parallel orientation to each other, resembling a pair of jaws as evidenced by their near 0 ° torsion angles N1-C2-N21-N22 and N1-C6-N61-N62.<sup>241</sup> A similar disposition of azido groups was found in the crystal structure of 2,4-diazido-6-diazoacetylpyrimidine.<sup>274</sup>

However, replacement of both chloro groups in **76** with azido or amino groups in **84** and **80** respectively, alters internal and external angles at their attachment positions fairly significantly. Compared to the chloro ethyl ester homologue **76**, substitution of the chloro groups for azido groups in **84** results in decreased internal N1-C2-N3 and N1-C6-C5 angles, both by 1.4 (4) °, and increased external N1-C2-N21 and N1-C6-N61 angles by 3.1 (4) and 3.9 (4) °, respectively.<sup>232</sup> Although most bond distances are similar within the required 3 $\sigma$ , C5-C6 exhibits longer bond lengths by 0.025 (6) Å in **84**, whilst N9-C10 was shown to display shorter bond lengths by 0.017 (5) Å. Replacement of the chloro groups by electron-donating amino substituents contracts C5-C6-N1 by 3.38 (11) ° and N3-C2-N1 by 2.25 (12) ° respectively (Table 5.2), and expands the respective N3-C4-C5 and C6-N1-C2 internal angles by 1.1 (3) and 2.85 (11) °. Finally, substitution of azido groups with amino groups was seen to contract the internal C2 and C6 angles by 0.8 (12) and 2.0 (11) °, and expand the internal C5 and N1 angles by 1.6 (12) and 1.35 (11) °, respectively.





**Figure 5.9** ORTEP drawing [Johnson, 1976] of 2,6-diazo-(9-carboxymethyl)purine methyl ester<sup>242</sup> 84, showing the labelling scheme for non-H atoms and the site and orientation of the side ring with thermal ellipsoids drawn at the 50% probability level and H atoms shown as small spheres with arbitrary radii.

In **92**, the N1-C2-N3 internal angle, as expected, remains virtually unchanged, but C5-C6-N1 was observed to contract by 3.1 (14) ° due to substitution of the electron withdrawing chloro group with the mesomeric electron donating methoxy group at C2, whilst C6-N1-C2 and C2-N3-C4 are observed to expand by 2.4 (13) ° and contract by 1.3 (13) ° respectively.

In summary, from these observations, the C2 and C6 internal angles decrease upon progressive change of substituent from chloro to azido to amino moieties. This is attributed to the nature of the substituents, where the electron density of the substituent groups at C2 and C6 may alter the amount of ring strain; the intramolecular repulsion between neighbouring bonds. This is accomplished via contraction of the appropriate internal bond angles where the electron donating groups are attached, and expansion arising from attachment of electron withdrawing groups.

### 5.6.1 Characterisation of Diethyl 6-Oxo-6,7-dihydro-3H-purine-3,7-diacetate<sup>247</sup> **124**

### 5.6.2 Crystal Data for Diethyl 6-Oxo-6,7-dihydro-3H-purine-3,7-diacetate **124**

A single crystal (0.55 x 0.45 x 0.25 mm) of **124** (C<sub>13</sub>H<sub>16</sub>N<sub>4</sub>O<sub>5</sub>) crystallised in the orthorhombic space group *P*<sub>bca</sub>, *Z* = 8. Geometric parameters were: *a* = 11.356 (2), *b* = 13.243 (2), *c* = 20.002 (3) Å,  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 90^\circ$  and *V* = 3008.3 (8) Å<sup>3</sup>. *D<sub>x</sub>* = 1.361 g/cm<sup>3</sup>;  $\Delta\rho_{\max} = 0.213 \text{ e \AA}^{-3}$ ;  $\Delta\rho_{\min} = -0.270 \text{ e \AA}^{-3}$ ; *R*(*F*) = 0.0424; *wR*(*F*<sup>2</sup>) = 0.1246;  $\mu = 0.902 \text{ mm}^{-1}$  and *R*<sub>int</sub> = 0.0255.

The structural analysis of peralkylated **124** was undertaken to elucidate the substitution pattern of the two ethyl ester chains on the heterocyclic ring, judged to be present from spectroscopic analyses (Section 3.9.1). Crystal structure analysis revealed the two ethyl acetate fragments were attached to **124** at the N3 and N7 sites rather than the alternative N1, O6, and N9 sites.<sup>247</sup> The molecular geometry of **124** is in good agreement with documented literature, where substitution of hypoxanthine derivatives commonly occurs upon the N7 or N3 sites. In its isomeric form, the N7 of H has been observed to be connected to 2'-deoxyribose.<sup>275</sup> The substitution pattern of the two ethyl ester substituents of **124** (Figure 5.10) alleviates steric hindrance that may otherwise ensue from the alternative substitution patterns. Substitution at N7 and N3 places the chains such that steric hindrance is minimised.

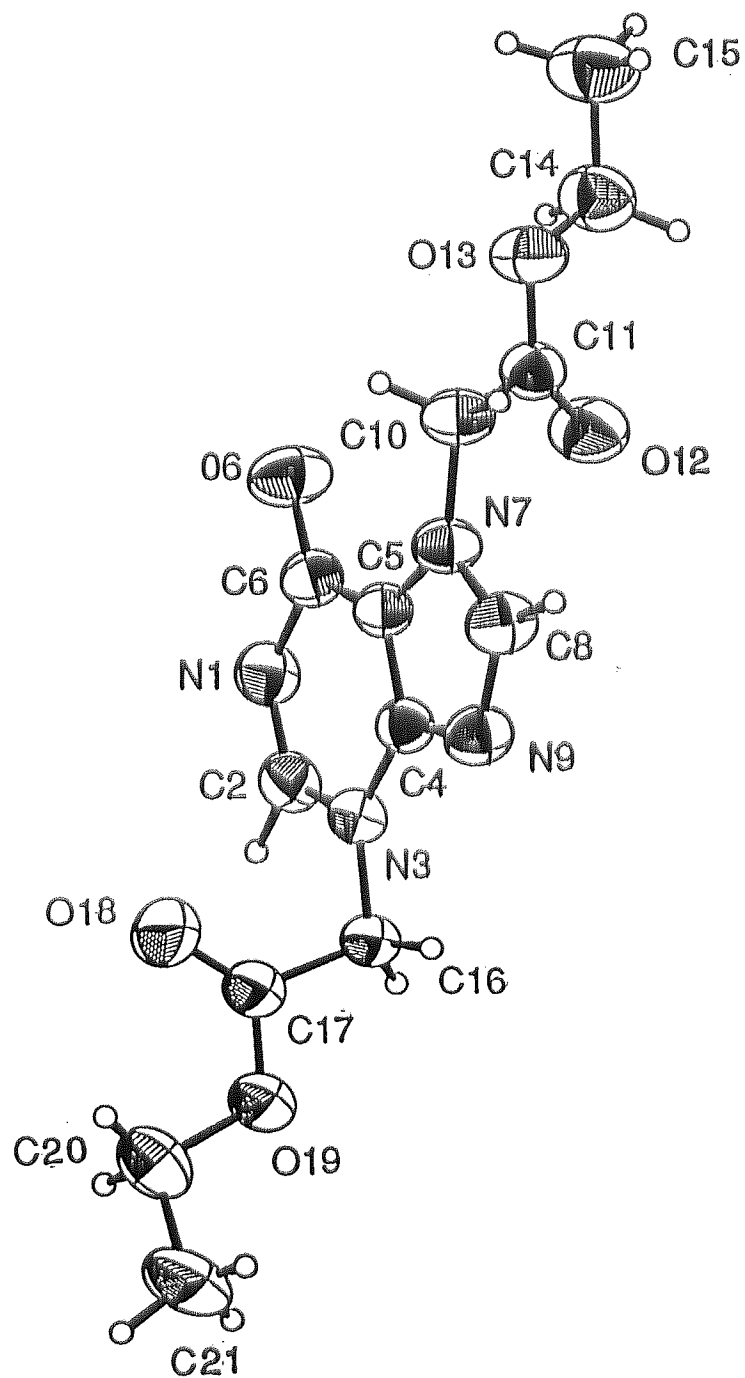


Figure 5.10. ORTEP drawing<sup>263</sup> of diethyl 6-oxo-6,7-dihydro-3*H*-purine-3,7-diacetate<sup>247</sup> 124 showing the labelling scheme for non-H atoms and the site and orientation of the side chain with thermal ellipsoids drawn at the 50% probability level.

As a consequence of the differences in the purine ring geometry of different tautomeric forms of H, anisotropic thermal ellipsoids characteristic of a slightly disordered molecule would be expected, if statistical distribution of several tautomeric forms within the structure is assumed. However, as observed from Figure 5.10, no obviously disordered molecules resulted from structure analysis.<sup>276,277</sup>

Evident in the crystal structure of the dialkyl compound is the near coplanar arrangement (Figure 5.10) of the heterocyclic ring atoms within  $\pm 0.022$  (1) Å, similar to all substituted purines in this series (Figure 5.4). Attachment of the side chains at N3 and N7 avoids steric hindrance from the purine ring, where the torsion angle C2-N3-C15-C16 -68.9 (2) ° is further from orthogonal than C8-N7-C10-O11 106.8 (2) °. The C16-O16 bond is marginally shorter than the other ester carbonyl group at C11-O11. Their observed orientation minimises steric hindrance with the ring, whilst the site of substitution minimises steric hindrance with each other.<sup>247</sup> As with the N9 76 and N7 77 dichloropurine isomers, the N9 bond angle at the unsubstituted N9 site is smaller compared with the corresponding angle of the substituted N7 site by a significant 3.47 (15) °.

**Table 5.3.** Selected geometrical parameters for **124**.

Bond	Length (Å)	Bond Angle	Angle (°)
N1-C2	1.294 (2)	N1-C2-N3	126.74 (14)
C6-O6	1.234 (2)	C2-N3-C4	115.75 (11)
C11-O11	1.197 (2)	N1-C6-C5	114.06 (11)
C16-O16	1.191 (2)	C8-N7-C5	106.27 (11)
		C8-N9-C4	102.80 (11)

Compared with values averaged between two independent molecules of the 1H lactam form of hypoxanthine,<sup>275</sup> the N1-C2 and C6-O6 carbonyl bond lengths are both shorter in **124** by 0.072 (3) Å and 0.012 (3) Å respectively, whilst the internal angles C8-C7-C5, N1-C6-C5 and C2-N3-C4 are expanded by 2.0 (6), 2.4 (2) and 3.7 (2) ° respectively. Unlike hypoxanthine and its respective hydrate,<sup>275,276</sup> nitrate<sup>277</sup> and hydrochloride<sup>279</sup> forms, **124** lacks N-H H-bond donor groups and therefore displays only an intermolecular contact from C2-H2 to O11: H2...O11<sup>i</sup> 2.24 (2), C2...O11<sup>i</sup> 3.11 (2) Å and C2-H2...O11<sup>i</sup> 153 (1) °; symmetry code: (i) -x, -y, 1-z, with no intermolecular H-bond contacts in the crystal structure.

**Table 5.4.** Selected bond torsion ( $^{\circ}$ ) values for the N3 and N7 substituents on **124**.

N7 Side Chain	Angle ( $^{\circ}$ )	N3 Side Chain	Angle ( $^{\circ}$ )
C4-C5-N7-C10	177.46 (12)	N1-C2-N3-C15	176.14 (13)
C5-N7-C10-C11	-68.9 (2)	C4-N3-C15-C16	106.86 (15)
N7-C10-C11-O11	-9.0 (2)	N3-C15-C16-O16	-6.5 (3)
N7-C10-C11-O12	172.02 (13)	N3-C15-C16-O17	172.84 (14)
C10-C11-O12-C13	-179.9 (2)	C15-C16-O17-C18	-179.5 (2)
O11-C11-O12-C13	1.8 (2)	O16-C16-O17-C18	-0.2 (3)
C11-O12-C13-C14	-172.9 (2)	C16-O17-C18-C19	-168.2 (2)

## 5.7 Evaluation of PNA Backbone Intermediates

### 5.7.1 Structural Comparison of the Boc-Protected Intermediate **17** with Some Reported Analogues

The synthesis of **17** (Figure 5.11) was first reported by Ravikumar *et al* as a pale yellow oil,<sup>229</sup> However, replication of the same method furnished **17** as pale yellow crystals which were recrystallised from MeOH to produce colourless crystals with a very good refinement factor (0.043),<sup>217</sup> which was observed to be superior to related Boc-protected amino acids<sup>230</sup>.

### 5.7.2 Crystal Data for *N*-(Boc-amino)acetonitrile<sup>229</sup> **17**

A single crystal (0.6 x 0.6 x 0.2 mm) of **17** (C<sub>7</sub>H<sub>12</sub>N<sub>2</sub>O<sub>2</sub>) crystallised in the orthorhombic space group *P*2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, *Z* = 4. Geometric parameters were: *a* = 8.5261 (7), *b* = 10.012 (7), *c* = 10.222 (14) Å,  $\alpha = 90$ ,  $\beta = 90$ ,  $\gamma = 90$  ° and *V* = 872.5 (2) Å<sup>3</sup>. *D<sub>x</sub>* = 1.189 g/cm<sup>3</sup>;  $\Delta\rho_{\max} = 0.0285$  e Å<sup>-3</sup>;  $\Delta\rho_{\min} = 0.276$  e Å<sup>-3</sup>; *R*(*F*) = 0.0276; *wR*(*F*<sup>2</sup>) = 0.0817;  $\mu = 0.729$  mm<sup>-1</sup> and *R*<sub>int</sub> = 0.043.

Conjugation within the urethane system is facilitated by near planarity about both the C-O and C-N bonds. In common with similar Boc-protected glycine<sup>230</sup> and alanine<sup>280</sup> amino acids, a *syn* periplanar relationship is observed between the O11 atom of the carbonyl group and the tertiary C7 atom, concurrent with an antiperiplanar displacement of the C10 methyl group to the carbonyl C5 atom. In common with other urethanes, C6-N5-C4 is ~ 10 ° lower than the trigonal value. However, the locations of the multiple bonded carbon atom, C2, and the carboxyl C atom in the amino acids, show more variability where C2-C3-N4-C5 is (-)-synclinal in **124**, but ranges from 62.9 to 110.7 ° in other Boc-protected amino acids. The only proton donor group, N4, donates a fairly weak intermolecular H-bond to a nearby N1 atom, related by (2-*x*, -0.5 + *y*, 0.5-*z*) with H4...N1 2.37 (2) and N4...N1 3.132 (2) Å.

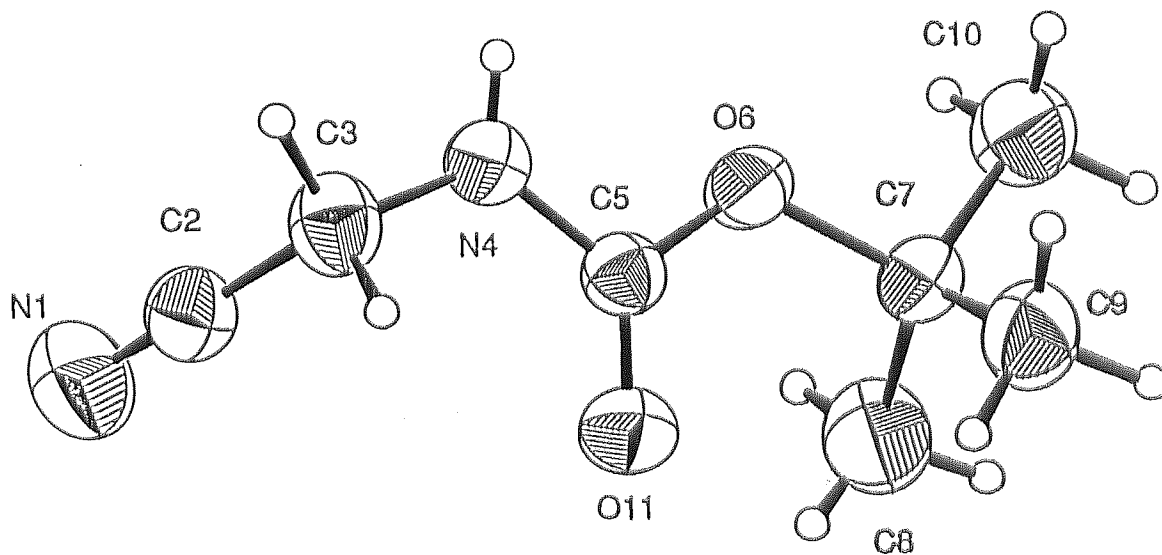


Figure 5.11. ORTEP view<sup>263</sup> of *N*-(Bocamino)acetonitrile **17**, showing the labelling scheme for non-H atoms and the site and orientation of the side chain with thermal ellipsoids drawn at the 50% probability level.

### 5.7.3 Comparison of 1-Hydroxyiminophenylcarbamate **20** at 213 K and 293 K

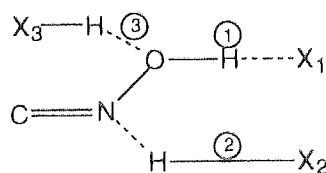
#### 5.7.4 Crystal Data for **20**

A single crystal (0.4 x 0.32 x 0.16 mm) of **20** (C<sub>8</sub>H<sub>6</sub>N<sub>2</sub>O) crystallised in the monoclinic space group  $P2_1/c$ ,  $Z = 4$ . Geometric parameters were:  $a = 8.8022$  (14),  $b = 11.061$  (3),  $c = 8.447$  (2) Å,  $\alpha = 90$ ,  $\beta = 116.607$  (5),  $\gamma = 90^\circ$  and  $V = 735.3$  (3) Å<sup>3</sup>.  $D_x = 1.320$  g/cm<sup>3</sup>;  $\Delta\rho_{\max} = 0.373$  e Å<sup>-3</sup>;  $\Delta\rho_{\min} = -0.449$  e Å<sup>-3</sup>;  $R(F) = 0.0897$ ;  $wR(F^2) = 0.2366$ ;  $\mu = 0.747$  mm<sup>-1</sup> and  $R_{\text{int}} = 0.028$ .

Crystal structural analysis of **20** at -60 °C has previously been documented by the Russian group led by Espenbetov<sup>231</sup>. In this thesis work, X-ray crystallography was undertaken at 293 K to confirm the identity of **20** (Figure 5.12).

Though a small degree of thermal expansion was expected with the rise in temperature from 213 K to 293 K, it was interesting to note the difference in increase for each of the respective axes; the largest increase was observed for axis  $c$  and the smallest for axis  $b$ . Comparison of bond lengths obtained from this work with those reported by Espenbetov did not show any unexpected differences.

Connecting molecules of **20** are linked around the  $2_1$  screw axis using H-bonds. One such bond of **20** exists between O11 of one molecule where OH acts as the donor group, with N9 of the connecting molecule, where the nitrile group acts as an acceptor to assist in H-bond formation. The oxime group has been shown<sup>281</sup> to participate in three types of H-bonds, as shown in Figure 5.12.



**Figure 5.12.** Formation of the three types of H-bonds exhibited by the oxime group of **20**.

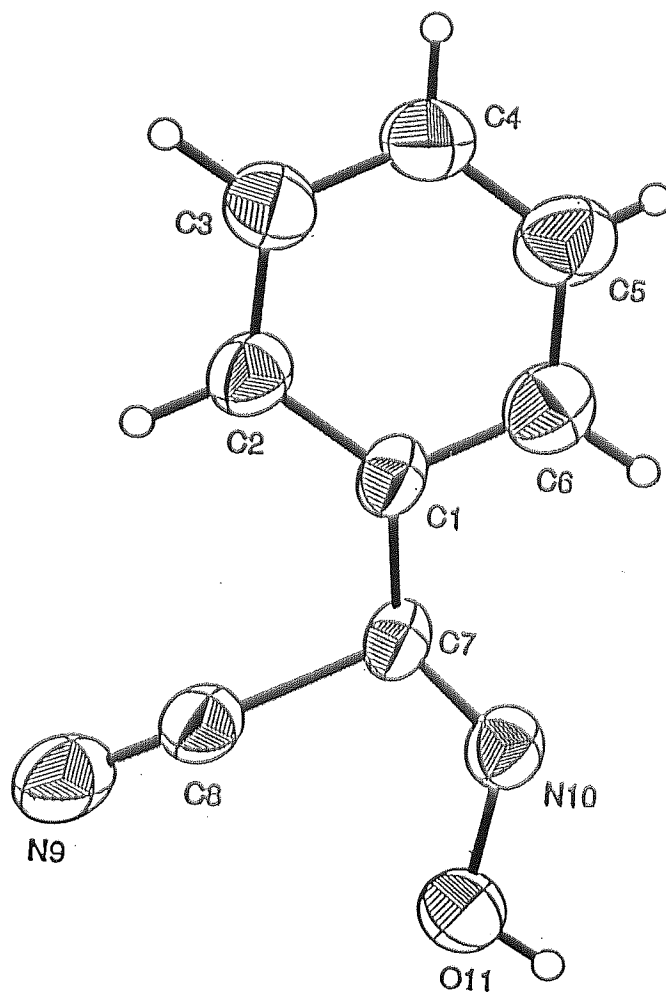


Figure 5.13. ORTEPII drawing<sup>263</sup> of 1-hydroxyiminophenylcarbamate **20**. Thermal ellipsoids are drawn at the 50% probability level.



Table 5.5. Comparison of the cell axes of **20** obtained at 213 K and 293 K

Cell Axes	Temp. (213 K)	Temp. (293 K)
<i>a</i>	8.714 (5)	8.802 (14)
<i>b</i>	11.022 (5)	11.061 (3)
<i>c</i>	8.297 (3)	8.447 (2)

Bertolasi *et al*<sup>281</sup> have showed the total distribution of N-O distances with a mean of 1.408 Å may be assigned to three classes. Grouping of the N-O distances in these classes gives rise to three partially overlapping sub-distributions as shown:

A (type 1 H-bond) = 1.388 (3) Å

B (type 1 + 2 H-bonds) = 1.416 (2) Å

C (type 1 + 2 + 3 H-bonds) = 1.425 (4) Å

The structure **20** has a type 1 H-bond and is therefore classified as a class A oxime. The symmetry operation  $2-x, -0.5+y, 1.5-z$  maps the two molecules. In Figure 5.14, it can be observed that when the second molecule of **20** is projected onto the same plane as the first molecule, the H-bond extends mainly along the *y*-axis with some component down the *x*-axis. The extensions of H-bonds along these axes is a valid explanation for why the *b* cell dimension displayed the smallest increase with respect to temperature increase, and why the *a* cell dimension exhibited a lower increase compared with the *c* cell dimension.

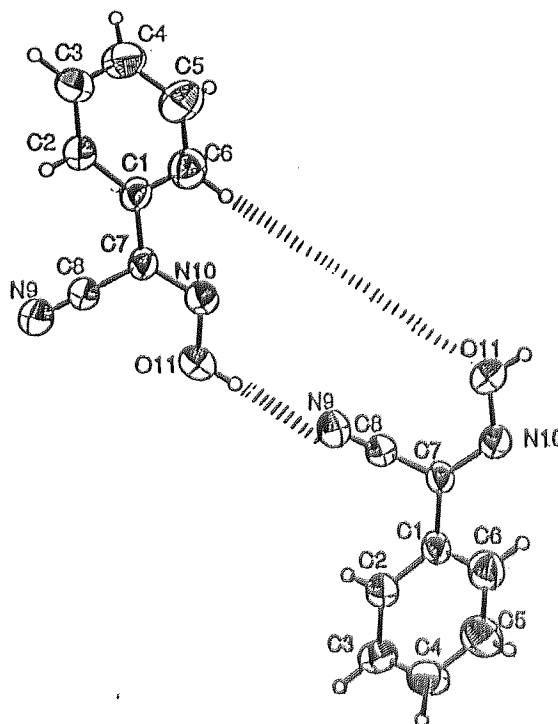


Figure 5.14. H-Bonding between two molecules of **20**

## 5.8 Evaluation of Packing Features of the Previous PNA Building Block Intermediates

### 5.8.1 Information Derived from Packing Diagrams

Application of computer graphics and molecular modelling programs, allows manipulation of the molecular orientation. This enables visual comparison of the adopted conformations, the stacking patterns within the unit cell and additionally, the types of inter- or intra-molecular H-bonding interactions which may influence the packing orientation.

### 5.8.2 Base Stacking and H-Bonding Interactions

The double-helical structure of DNA is maintained via two prominent types of molecular interactions, H-bonding and base stacking (Section 1.3). As the molecules under study possess similar purine rings, both types of interactions were expected to occur. Horizontal H-bonding interactions are observed to lie in the plane of the bases and are mainly electrostatic in character, with an essential role in the stabilisation of nucleic acid secondary structures.<sup>184,280</sup> H-bonds have the general formula X-H...Y where X and Y are electronegative molecules or atoms. The strength of a H-bond, reflected in its length, depends on charges located on X, H and Y. H-bonding between purine and pyrimidine heterocycles is of the type N-H...N and N-H...O where the donor N-H group is either of the amino or the imino type. Under the influence of a polarising H-bond, the charge on X, H and Y is observed to adjust; H becomes more electropositive whilst X, Y become more electronegative. This effect leads to increased affinity of X, Y for accepting further H-bonds. Only **80**, **92** and **20** (diamino, amino methoxy and oxime) possess NH<sub>2</sub> or NH groups which can supply polar H atoms for conventional H-bonds. The remaining molecules synthesised in this work can only participate in weak C-H...X interactions where the carbon atom is adjacent to an electron-withdrawing atom, such as an heterocyclic nitrogen, as in the cases for the substituted heterocyclics, for example. H-bonds have been discussed previously for the individual structures in the appropriate sections of this chapter.

Base stacking,<sup>282</sup> which is composed of those interactions that lie perpendicular to the base planes, is important for the stabilisation of nucleic acid helices, and is analysed in the following section. Stacked arrangements are dominant in crystal structures of bases where the stacking patterns are observed to be specific with the polar substituents -NH<sub>2</sub>, =N, =O or halogen of one base superimposed over the aromatic system of the adjacent base. Since the overlap of  $\pi$ -electron clouds is a method for sharing the available electron density with the sigma framework, rings of greater electron deficiency are expected to possess a greater requirement to adopt a stacking mode. However, this propensity may be outweighed by H-bonding and other requirements for efficient packing. The presence of polarising hetero

atoms in, or attached to, an aromatic system has a major effect on electrostatic interactions, particularly for purine and pyrimidine heterocycles.<sup>281</sup>

The following representations display a specimen molecule rotated to a common orientation, in which the N1-C5 bond is horizontal and the C4-C5 bond is approximately vertical in the plane of the page. A transformed molecule is projected in the same way to bestow an optimum representation of the stacking pattern within the unit cell. In many cases, molecules are stacked pairwise about an inversion centre. When the inversion operation acts on a planar ring, the transformed ring is strictly parallel to the original ring to maximise overlap. Particularly important, regarding the series of homologous purines in this thesis work, is the orientation adopted by substituents relative to the heterocyclic ring in the stacked assembly. The subsequent effects imposed as a consequence of the emerging stacked pattern, are briefly discussed.

### 5.8.3 Stacking Arrangement of 6-Chloro-(9-carboxymethyl)purine ethyl ester 125

In Figure 5.15 (a) the independent molecule **125A** is represented by thermal ellipsoids; molecule **125B** is represented as a line drawing. The transformation of N3 of the independent molecule **125A** by the inversion (1-x, 1-y, 2-z) positions molecule A 3.226 (4) Å above the ring plane. There is only a small degree of overlap, but the adopted packing orientation allows adjacent molecules to be placed with the side chains located at opposite ends and directed in opposite directions for efficient packing.

In Figure 5.15 (b) molecule **125A** is represented as a line drawing and **125B** is represented by ellipsoids. The N3 site in molecule **125B** transformed by (1-x, -y, 1-z) lies 3.203 (11) Å above the plane of the ring. In both molecules, **125A** and **125B**, only the transformed N3 site projects within the outline of the original ring. N3 is situated at a perpendicular distance slightly greater than 3.2 Å. This is shorter than the 3.4 Å stacking distance of bases, characteristic in the DNA helix. The short spacing is made possible as only the compact N atoms are in contact.

### 5.8.4 Stacking Arrangement of 6-Chloro-(7-carboxymethyl)purine ethyl ester 126

Inversion of the ring by (1-x, -y, -z) allows greater overlap than occurs in the N9 regioisomer, but the perpendicular distance between the rings is greater thus reducing the base stacking interactions somewhat. The transformed C6 site of the ring lies 3.425 (2) Å above the ring plane (Figure 5.16).

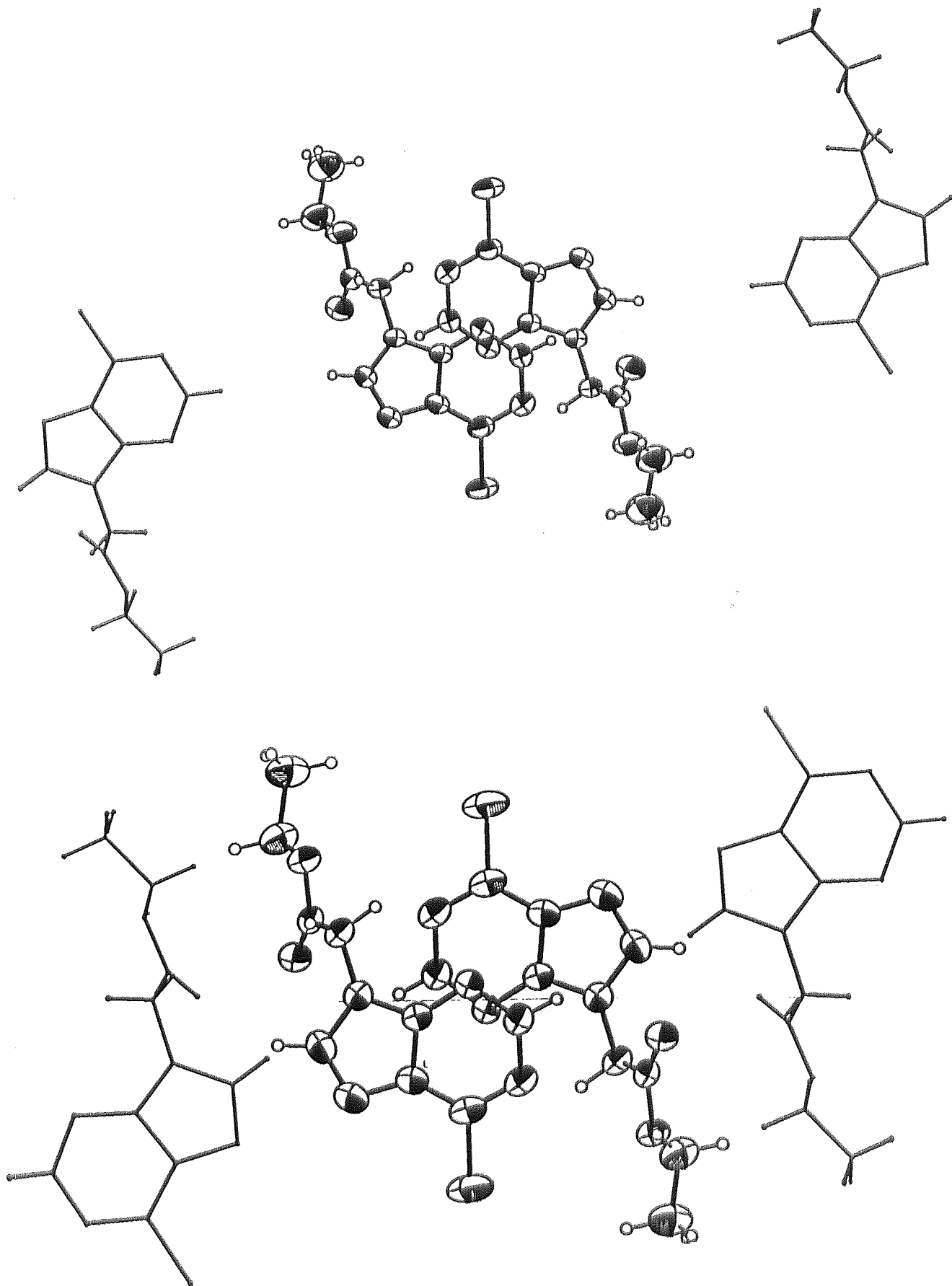


Figure 5.15. Stacking arrangement of 6-chloro-(9-carboxymethyl)purine ethyl ester<sup>248</sup> 125A and 125B.

(a) with a molecule transformed by  $(1-x, 1-y, 2-z)$  projected above the purine ring  
 (b) with a molecule transformed by  $(1-x, -y, 1-z)$  projected above the purine ring.

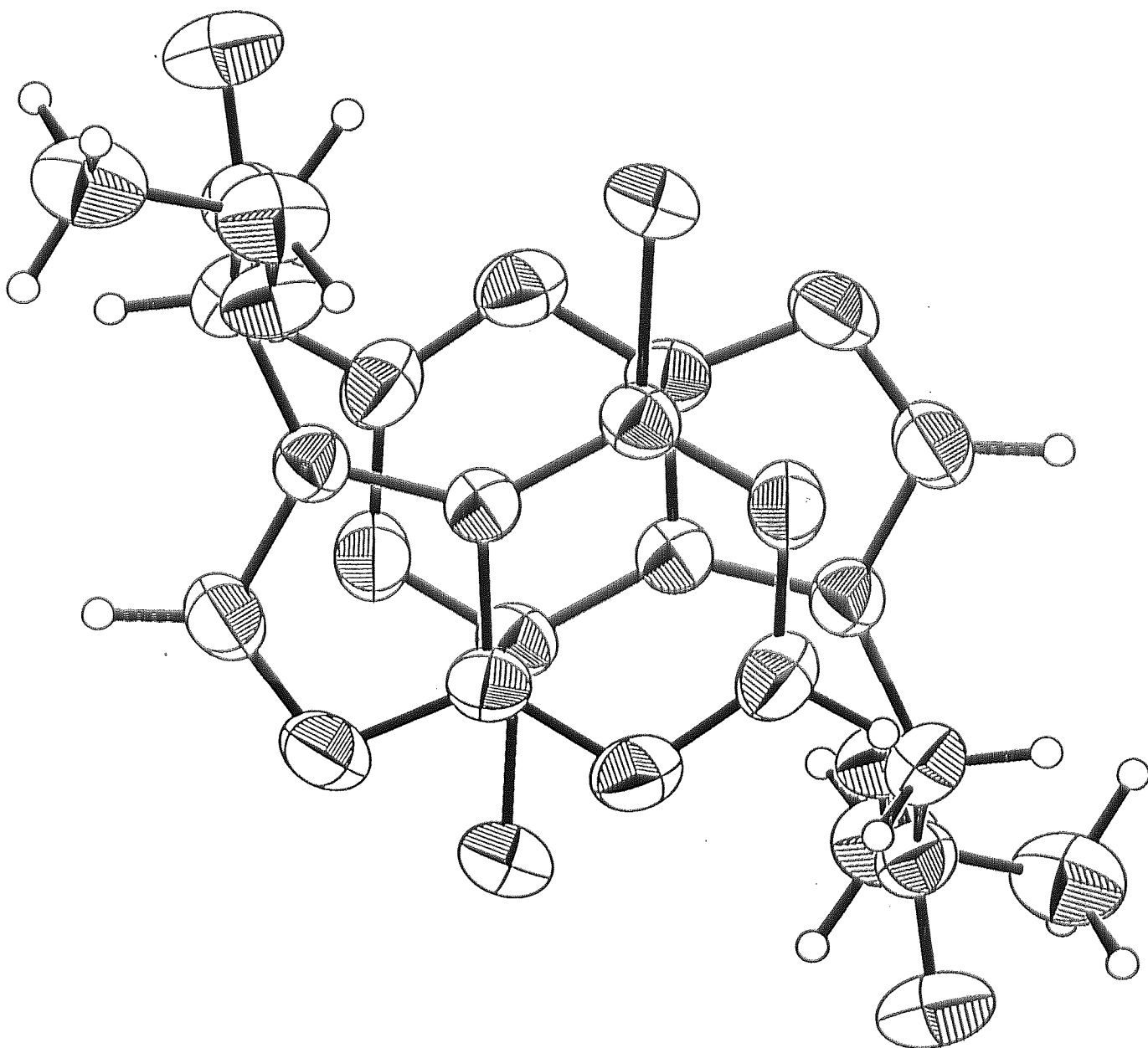


Figure 5.16. Stacking arrangement of 6-chloro-(7-carboxymethyl)purine ethyl ester<sup>249</sup>  
126.

### 5.8.5 Stacking Arrangement of 2,6-Dichloro-(9-carboxymethyl)purine ethyl ester **76**

In Figure 5.17 (a) the symmetry operation  $(x, 1-y, 0.5+z)$  positions the transformed atoms C5, C6, Cl6 3.648 (3), 3.804 (3), 3.485 (2) Å, respectively, above the plane of the original ring of molecule **76**. The plane of the transformed ring intersects the original plane at an angle of 16.94 °. It is proposed that this transformation, a glide plane operation, occurs in preference to a totally parallel alignment of the transformed ring with the original ring, to avoid steric clashes involving the bulky N9 side-chains. A straightforward translation may bring successive chains into conflict and an inversion could forced the chain against Cl6 of the stacked molecule. The Cl6 atom, which is expected to possess a partial negative charge, lies over the presumed positive C6 site.

A different molecule appears in Figure 5.17 (b), adjacent to the reference molecule. This transformed molecule of **76** is related by  $(x, 1-y, -0.5+z)$ , a translation of one unit cell length along the *c* axis, compared with the previous representation (Figure 5.17). N7, C8, N9 are positioned approximately in alignment with the edge of the original ring, and at distances of 3.858 (3), 3.786 (3), 3.416 (3) Å below the original ring.

### 5.8.6 Stacking Arrangement of 2,6-Diamino-(9-carboxymethyl)purine ethyl ester **80**

As before, molecule **80A** is represented by thermal ellipsoids, **80B** by heavy lines and **80C** by thin lines. In Figure 5.17 (a) the sites transformed by  $(1-x, 1-y, 2-z)$  and  $(1-x, 2-y, 2-z)$  are also shown. Every proton donor and acceptor site upon molecule **80A** is involved in H-bonding aided by the water of crystallisation. Hence, this accounts for the perfect ordering of molecule **80A**. For the alternative sites, **80B** and **80C**, N1 is completely uninvolved in H-bonding; the 2-amino group is able to make only one weak H-bond. The 6-amino group and N3 site act as the principal anchor points. Thus, the ring is able to revolve slightly in-plane, enabling the side-chain to adopt a completely different conformation compared with the orientation of the original ring.

In Figure 5.17 (b) molecule **80B** or **80C**, transformed by  $(-x, 1-y, 1-z)$ , is stacked with C2B and C4B of the transformed **80B** ring 3.28 (3) Å 3.20 (1) Å, below the original **80B** ring, and C2C and C4C 3.75 (3), 3.75 (1) Å below the original **80C** ring. It is expected the adoption of this specific mode of stacking will not be affected to any significant extent by the in-plane rotation, implicated by the disorder.

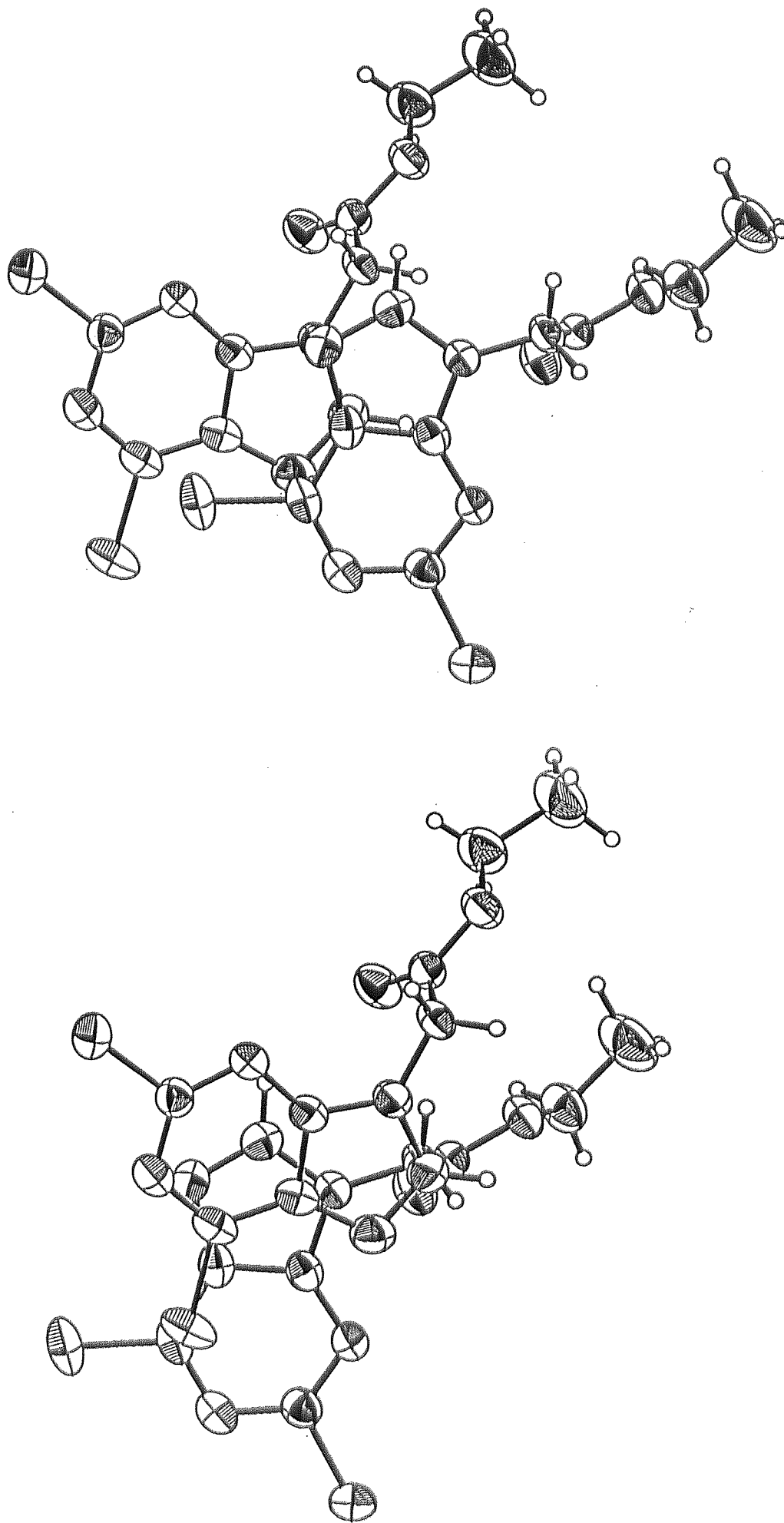
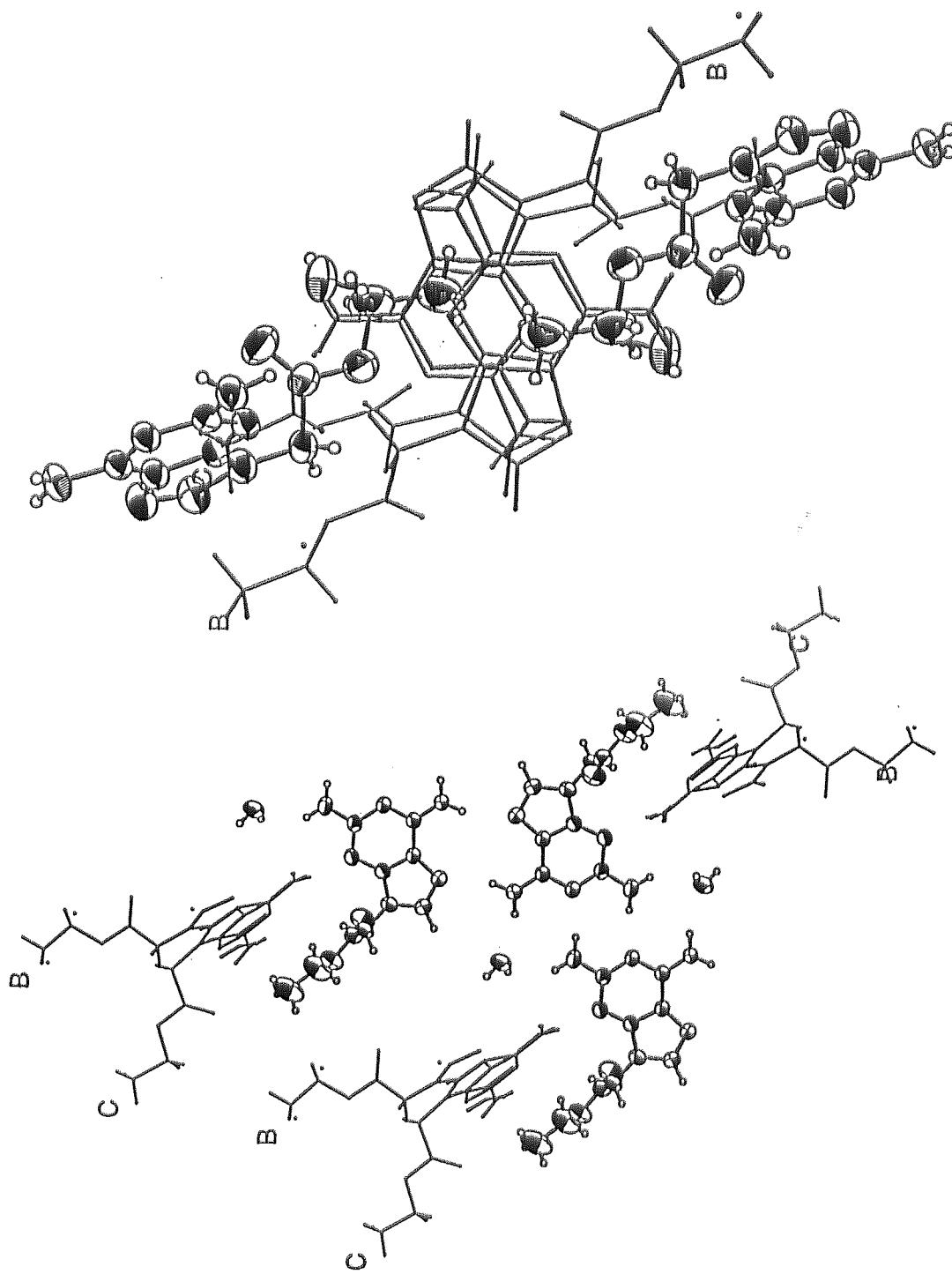


Figure 5.17. Stacking arrangement of 2,6-dichloro-(9-carboxymethyl)purine ethyl ester<sup>232 76</sup>.  
 (a) with a molecule transformed by (1-x, 1-y, 2-z) projected above the purine ring and (b).



**Figure 5.18.** Stacking arrangement of 2,6-diamino-9-(9-carboxymethyl)purine ethyl ester<sup>239 80</sup>.  
 a) with a molecule transformed by (1-x, 1-y, 2-z) projected above the purine ring  
 b) with a molecule transformed by (-x, 1-y, 1-z) projected above the purine ring.



### 5.8.7 Stacking Arrangement of 6-Amino-2-methoxy-(9-carboxymethyl)-purine methyl ester 92

The molecule shown stacked above the original molecule in Figure 5.19 (a) is generated by an inversion centre (1-x, -y, 1-z) where extensive overlap is observed. Transformed atoms N1, C2, C4, C5 are 3.3792 (7), 3.392 (2), 3.392 (2), 3.401 (2) Å respectively above the original ring plane.

In Figure 5.19 (b) the base-paired molecule is generated by an inversion centre giving 1-x, 1-y, 1-z. These paired H-bonds are identical in their geometry because of the symmetry:

N---H distance = 0.91 (2)

N---N distance = 3.055 (2)

H...N distance = 2.15 (2)

N--H...N angle = 171 (2)°

### 5.8.8 Stacking Arrangement of 2,6-Diazido-(9-carboxymethyl)purine methyl ester 84

Beneath the molecule represented in the standard orientation, lies a molecule related by a glide plane (x, 0.5-y, 0.5+z) with its C5, C6, N61 3.534 (2), 3.576 (3), 3.473 (4) Å below the original plane. The two ring planes are not parallel, but inclined at 3.87 (9)° (Figure 5.20).

### 5.8.9 Stacked Arrangement of Diethyl 6-oxo-6,7-dihydro-3H-purine-3,7-diacetate 124

The transformation (-x, -y, 1-z) appears to promote exceptionally efficient packing. There is considerable overlap of the pyrimidine rings in the stack. The transformed N1 and C2 sites lie 3.299 (2) and 3.308 (2) Å above the plane of the substituted hypoxanthine ring. This arrangement facilitates a C-H...O hydrogen bond from the original C8 to the O11 sites of the transformed molecule, which has a H...O distance 2.24 (2) Å, a C...O distance 3.177 (2) Å, and a C-H...O angle 153 (1)°. The parallel arrangement of the original and transformed side chains should facilitate tight packing whilst avoiding steric hindrance, from the bulky substituents at N3 and N7.

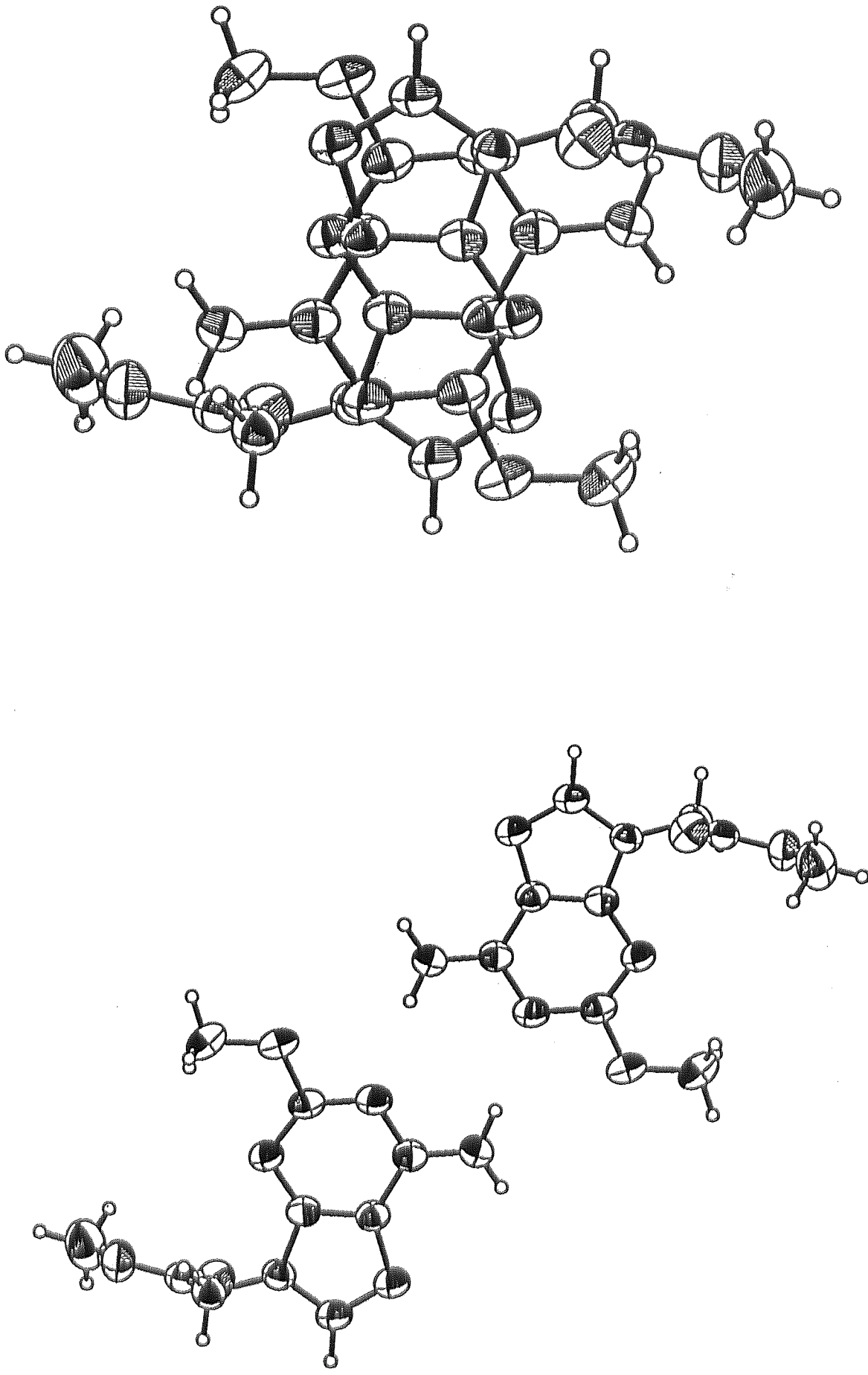


Figure 5.19. Stacking and base-pairing arrangement of 6-amino-2-methoxy-(9-carboxymethyl)purine methyl ester<sup>241</sup> 92.  
 (a) with a molecule transformed by (1-x, 1-y, 1-z) projected above the purine ring and (b) -x, 1-y, 1-z.

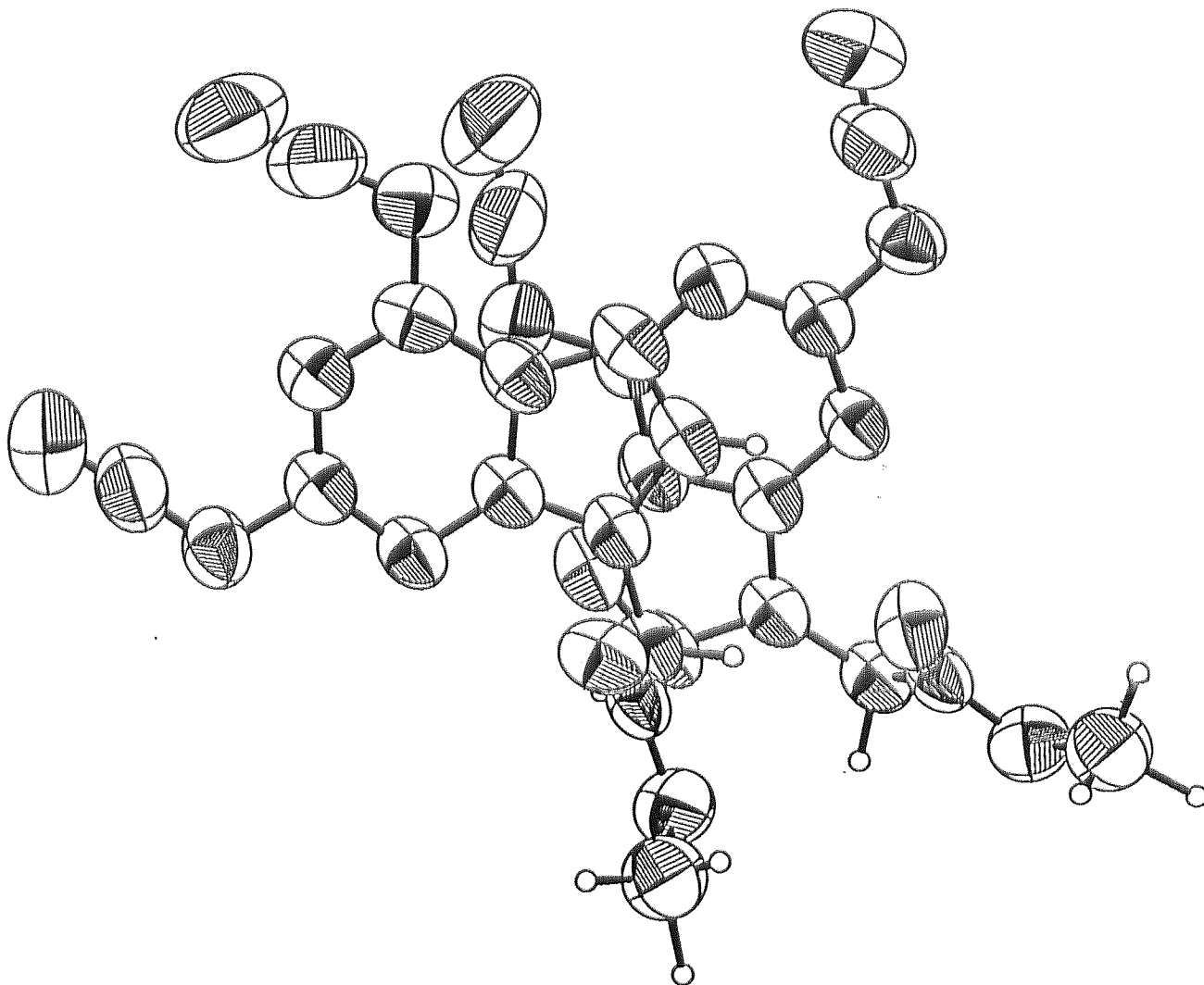


Figure 5.20. Stacking arrangement of 2,6-diaza-(9-carboxymethyl)purine methyl ester 84.

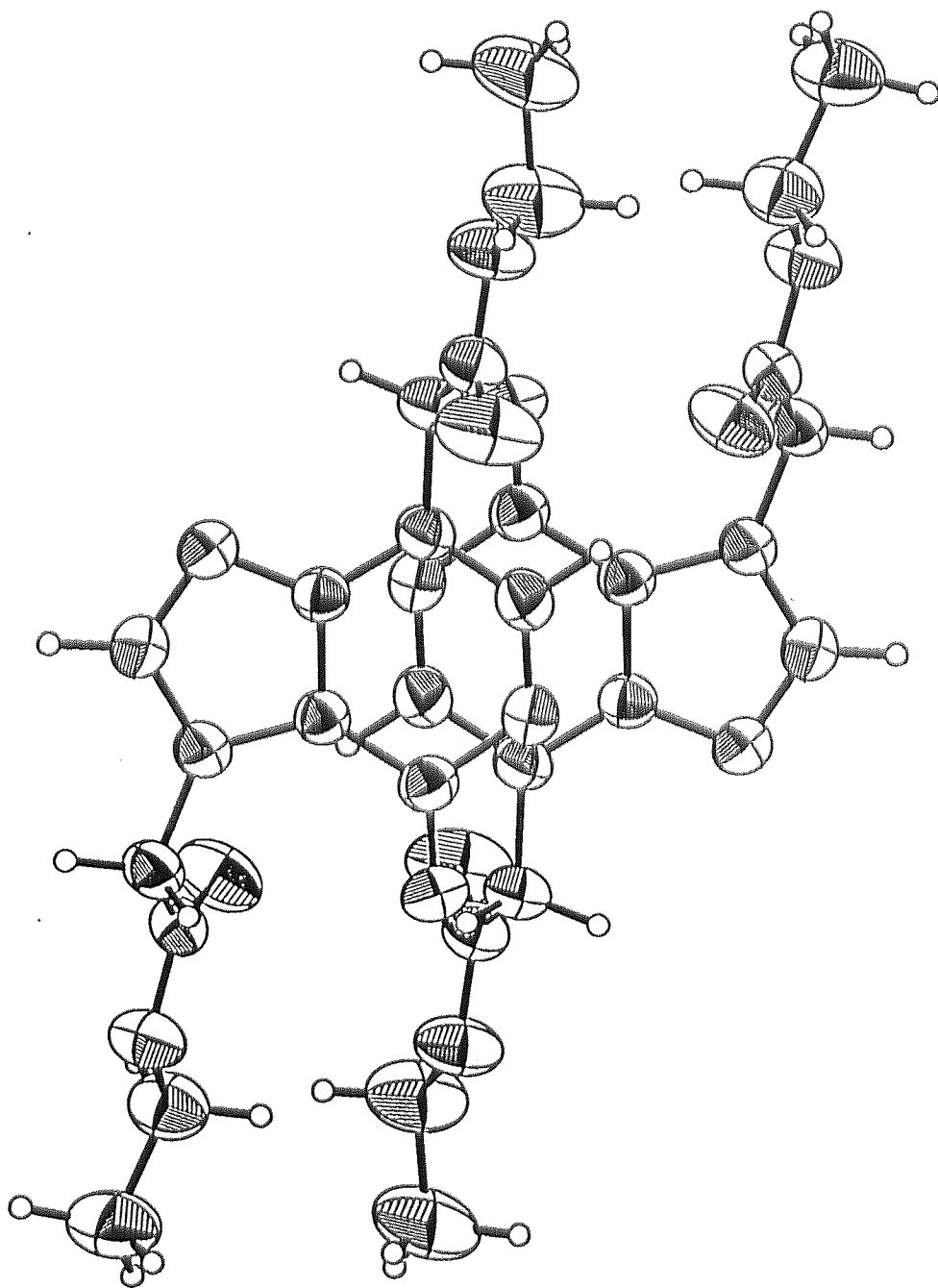


Figure 5.21. Stacking arrangement of diethyl 6-oxo-6,7-dihydro-3*H*-purine-3,7-diacetate 124.

# CHAPTER 6

## 6.1 Experimental

### 6.1.1 General Methods

NMR spectra were recorded on a Bruker AC-250 spectrometer at  $^1\text{H}$  (250.1 MHz) and  $^{13}\text{C}$  (62.9 MHz) and referenced to tetramethylsilane (TMS),  $\text{CDCl}_3$ , or  $d_6$ -DMSO. Mass spectrometry analyses were performed using VG Quatra I for  $\text{EI}^+$  or  $\text{CI}^+$  mode, or VG AutoSpec instruments for  $\text{FAB}^+$  mode, at EPSRC, Dept. of Chemistry, Birmingham University using a VG ProSpec instrument. Infra-red spectra were recorded on an FT-IR Mattson Galaxy 3000 Series spectrophotometer. Elemental Analyses were performed by Butterworths Laboratories, Middlesex. Flash column chromatography<sup>281</sup> was performed using Sorbsil C60 40/60H silica gel. TLC was performed using plastic-backed Kieselgel 60 silica gel plates containing a fluorescent indicator. Spots were visualised under UV light (254 nm) or with the aid of iodine. Melting points were measured uncorrected on a Kofler Reichert-Jung hot stage with the aid of a microscope (Cambridge Instruments).

## 6.2 Chemical Synthesis

### 6.2.1 (*N*-Boc-amino)acetonitrile<sup>229</sup> **17**

To a mixture of aminoacetonitrile.HCl **16** (30.0 g, 0.325 mol) and di-Boc anhydride **15** (70.9 g, 0.325 mol) in  $\text{CH}_2\text{Cl}_2$  (1 L) was added  $\text{Et}_3\text{N}$  (125 mL, 0.9 mol) dropwise, and the mixture stirred overnight at r.t. After rotary evaporation of the product mixture, the residue was dissolved in ether, washed with water, followed by brine and then dried ( $\text{MgSO}_4$ ). Evaporation of the solvent gave **17** (43.3 g, 85%). Recrystallisation from EtOAc gave the analytical sample; mp 155-158 °C; TLC (EtOAc)  $R_f$  0.58; IR (KBr disc):  $\nu_{\text{max}}$  3010, 2961, 2854, 2757, 2252, 1720, 1534, 935, 839, 783  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  1.46 (s, 9 H,  $(\text{CH}_3)_3$ ), 4.05 (d, 2 H,  $J$  6.0 Hz, CH-2), 5.30 p.p.m. (m, 1 H, NH,  $\text{D}_2\text{O}$  exchangeable);  $^{13}\text{C}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  28.1 ( $(\text{CH}_3)_3$ ), 29.0 ( $\text{CH}_2$ ), 81.1 (C), 115.6 (CN), 155.0 p.p.m. (CO); MS ( $\text{CI}^+$ ):  $m/z$  ( $I_r$ ) 157 (M+H, 93%), 118 (100%), 74 (26%), 72 (6%), 56 (8%); Anal. calcd. for  $\text{C}_7\text{H}_{12}\text{N}_2\text{O}_2$ : C, 53.8; H, 7.7; N, 17.9. Found: C, 53.8; H, 7.6; N, 17.8.

### 6.2.2 *N*-Boc-ethylenediamine **14**:

#### Method A

(*N*-Boc-amino)acetonitrile **17** (9.97 g, 0.064 mol) was dissolved in a slurry of 10% Pd/C (7.5 g) in anhydrous ethanol (55 mL) and hydrogenated at r.t. overnight in a Parr pressure vessel at 60 psi. The mixture was filtered through Celite and concentrated to dryness to afford 10.1 g (98%) **14** as a yellow oil; TLC (MeOH-EtOAc; 1:1 + 2%  $\text{NH}_4\text{OH}$ ):  $R_f$  0.32; Lit.<sup>219</sup>:  $R_f$  0.32. IR (KBr disc):  $\nu_{\text{max}}$  3018, 2966, 1662, 1534, 1392, 1215, 1099, 665,

482  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  1.35 (s, 2 H,  $\text{CH}_2$ ), 1.46 (s, 9 H,  $(\text{CH}_3)_3$ ), 2.9 (t, 2 H,  $J$  6.0 Hz,  $\text{CH}_2$ ), 3.4 (t, 2 H,  $J$  6.0 Hz,  $\text{CH}_2$ ), 5.2 p.p.m. (br s, 1 H, NH,  $\text{D}_2\text{O}$  exchangeable);  $^{13}\text{C}$  NMR [ $(\text{CDCl}_3)$ ]:  $\delta$  28.5 (3 x  $\text{CH}_3$ ), 70.0 ( $\text{CH}_2$ ), 77.7 ( $\text{CH}_2$ ), 155.9 ( $\text{C}(\text{CH}_3)_3$ ), 160.4 p.p.m. (CO); MS (electrospray):  $m/z$  ( $I_r$ ) 161 (M+H, 53%), 147 (18%), 133 (23%), 117 (12%), 105 (100%), 87 (18%), 43 (43%).

#### Method B

A solution of **14** (2.1 g, 0.130 mmol) in Raney nickel (5 g) in 10%  $\text{NH}_3/\text{EtOH}$  (350 mL) was hydrogenated at r.t. overnight in a Parr pressure vessel at 60 psi. The mixture was filtered through Celite and concentrated to dryness *in vacuo* to afford **14** (18.7 g, 87%), as a green oil TLC ( $\text{MeOH-EtOAc}$ ; 1:1) + 2%  $\text{NH}_4\text{OH}$ ):  $R_f$  0.34;  $^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  1.35 (s, 2 H,  $\text{CH}_2$ ), 1.46 (s, 9 H,  $(\text{CH}_3)_3$ ), 2.9 (t, 2 H,  $J$  6.0 Hz,  $\text{CH}_2$ ), 3.4 (t, 2 H,  $J$  6.0 Hz,  $\text{CH}_2$ ), 5.2 p.p.m. (br s, 1 H, NH,  $\text{D}_2\text{O}$  exchangeable). Remaining experimental data identical to analyses given in Method A.

#### Method C

To a vigorously stirred solution of ethylenediamine **23** (100 mL, 1.5 mol) in  $\text{CH}_2\text{Cl}_2$  (800 mL) at  $0^\circ\text{C}$  was added *tert*-butyl bromoacetate **22** (27.6 mL, 0.17 mol) in  $\text{CH}_2\text{Cl}_2$  (135 mL) during 5 h. The resulting mixture was allowed to warm slowly (3 h) to r.t. and left to stir overnight. The resulting reaction mixture was washed with water (3 x 200 mL), and the combined aqueous wash was back-extracted with  $\text{CH}_2\text{Cl}_2$  (175 mL). The combined organic layers were dried ( $\text{Na}_2\text{SO}_4$ ) and filtered. Analyses as for Method A.

#### Method D

##### 6.2.3 *tert*-butyl-4-Nitrophenylcarbonate **24**

4-Nitrophenylcarbonate **25** (25.5 g; 183.2 mmol) and sodium carbonate (58.28 g; 550 mmol) were stirred with anhydrous 1,4-dioxane (380 mL). Boc anhydride **15** (40.0 g, 183.2 mmol) in 1,4-dioxane (75 mL) was transferred to the mixture and the solution refluxed for 2 h. The resulting suspension was cooled to  $0^\circ\text{C}$  and filtered through Celite. The resultant filtrate was concentrated to a third, poured into ice and stirred for 1 h. Compound **24** (24.7 g, 78%) was collected by filtration, washed with water, dried over sicapent *in vacuo* and stored in a freezer; mp  $70\text{--}73^\circ\text{C}$ ; Lit.<sup>4</sup>  $71\text{--}73^\circ\text{C}$ .

*tert*-butyl 4-Nitrophenylcarbonate **24** (20 g, 83.6 mmol) was dissolved in anhydrous DMF (75 mL) and added dropwise during a period of 1 h to a solution of ethylenediamine **23** (56 mL) and DMF (75 mL). The solution was stirred for 24 h, evaporated to dryness and the resulting oil dissolved in 175 mL  $\text{H}_2\text{O}$ . The pH was adjusted to 3.5 at  $0^\circ\text{C}$  with 4 M HCl acid, filtered and extracted with  $\text{CHCl}_3$  (3 x 300 mL). The pH was adjusted to 12 with 2 M sodium hydroxide at  $0^\circ\text{C}$ , and the aqueous solution extracted with  $\text{CH}_2\text{Cl}_2$  (3 x 350 mL). Following treatment with saturated aqueous sodium chloride (350 mL), the

organic layer was dried over MgSO<sub>4</sub>, and the solution filtered and evaporated to dryness *in vacuo*, to produce **24** as a yellow, viscous oil (5.73 g, 27%); TLC (EtOAc): *R*<sub>f</sub> 0.38.

#### Method E

##### 6.2.4 *N*-Boc-ethylenebenzylamine **21** from *N*-(Boc)ethylenebenzylamine **14**, with the side product 1-Hydroxyiminophenylcarbamate **20**

To an ice-cooled solution of 2-(Boc-oximino)-2-phenylacetonitrile (Boc-ON) **19** (26.3g, 0.11 mol) and anhydrous 1,4-dioxane (20 mL) was added *N*-benzylethylenediamine **18** (16.3 g, 0.11 mol). The mixture was allowed to stir for 6 h at r.t. under Ar. The solvent was evaporated under high vacuum to yield the product mixture as a yellow oil. Flash chromatography, eluting with hexane-EtOAc-CH<sub>3</sub>CN 7:2:1 afforded the oxime by-product, 1-hydroxyiminophenylcarbamate **20**. Recrystallisation from EtOAc-Hexane afforded **20** (2 g, 6%), (6.4 g, 39%) as large colourless hexagonal crystals; mp 128-130 °C; Lit. mp<sup>280</sup>. 129-130 °C; TLC (C<sub>6</sub>H<sub>14</sub>-EtOAc-CH<sub>3</sub>CN 7:2:1): *R*<sub>f</sub> 0.29; IR (KBr disc):  $\nu_{\max}$  3383, 2239, 1452, 1284, 963, 870, 686 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  4.2 (s, 2 H, CH<sub>2</sub>), 7.70 (m, 2 H, CH<sub>2</sub>), 13.76 p.p.m. (b, 1 H, OH, D<sub>2</sub>O exchangeable); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  110.3, 125.7, (s, 2 H, CH<sub>2</sub>) 131.1 (s, 2 H, CH<sub>2</sub>), 131.6 p.p.m. (s, 2 H, CH<sub>2</sub>); MS (electrospray): *m/z* (*I*<sub>r</sub>) 164 (M+H, 67%), 146 (27%), 131 (48%), 121 (39%), 116 (100%) 106 (41%), 89 (96%). Further elution with EtOAc-MeOH 5:2 furnished **14** as a pale green oil (21.1 g, 79%); TLC (EtOAc:MeOH 9:1): *R*<sub>f</sub> 0.86; IR (KBr disc):  $\nu_{\max}$  2973, 1703, 1521, 1456, 1363, 1249, 1174, 740, 701, 453 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  1.35 (s, 9 H, *t*-Bu), 2.17 (t, 2 H, *J* 6.08 Hz, CH<sub>2</sub>), 2.46-2.52 (t, 2 H, *J* 7.5 Hz, CH<sub>2</sub>NH<sub>2</sub>), 2.75 (bs, 1 H, NHBoc, D<sub>2</sub>O exchangeable), 2.98 (q, 2 H, CH<sub>2</sub>NH, *J* 6.1 Hz), 3.67 (s, 2 H, CH<sub>2</sub>Ph), 6.75 (bs, 1 H, NH, D<sub>2</sub>O exchangeable), 7.21-7.24 p.p.m. (m, 5H, C<sub>6</sub>H<sub>5</sub>); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  28.4 (CH<sub>3</sub>), 40.1 (CH<sub>2</sub>), 48.7 (CH<sub>2</sub>), 52.91 (CH<sub>2</sub>), 77.6 (CO), 126.7 (CH, C<sub>6</sub>H<sub>5</sub>), 128.2 (CH, C<sub>6</sub>H<sub>5</sub>), 141.1 (CH), 155.9 p.p.m. (CH); MS (electrospray): *m/z* (*I*<sub>r</sub>) 251 (M+H, 30%), 195 (100%), 193 (4%), 177 (12%), 151 (31%) 120 (61%), 91 (100%), 57 (49%).

*N*-(Boc)-ethylenebenzylamine (1.075 g) **21**, was added to a slurry of anhydrous EtOH (20 mL) in 10% palladium in charcoal (0.40 g) and hydrogenolysed at r.t. for five d using a H<sub>2</sub>-filled balloon. The catalyst was removed via filtering of the product mixture and the solvent evaporated. The crude product was subjected to flash chromatography, eluting with EtOAc and recrystallised from MeOH. Analyses are as for Method A.

##### 6.2.5 Ethyl *N*-(Boc-aminoethyl)glycinate **1**

BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub> **27** (7.4 g, 0.044 mol) was added in one portion to a mixture of *N*-Boc-ethylenediamine **14** (9.97 g, 0.062 mol) and KF/Celite (26.04 g, 0.433 mol) in anhydrous CH<sub>3</sub>CN (dried over mol. sieves 3 Å; 200 mL) at 60 °C. The reaction mixture was cooled to r.t. after 20 min, filtered through Celite and concentrated to dryness *in*

*vacuo*. The residue was purified by flash chromatography using EtOAc-Hexane 1:1, followed by EtOAc, and finally by 10% MeOH-EtOAc to afford **1** (3.94 g, 26%) as a pale yellow oil; TLC (EtOAc-MeOH 9:1):  $R_f$  0.4; Lit.<sup>219</sup>. (EtOAc-MeOH 9:1):  $R_f$  0.4;  $^1\text{H}$  NMR [ $\text{CDCl}_3$ ]:  $\delta$  1.30 (t, 3 H,  $\text{CH}_3\text{CH}_2\text{O}$ ,  $J$  7.0 Hz), 1.46 (s, 9 H,  $(\text{CH}_3)_3\text{C}$ ), 1.70 (s, 1 H, NH), 2.7 (t, 2 H,  $\text{NHCH}_2$ ,  $J$  5.0 Hz), 3.2 (t, 2 H,  $\text{NHCH}_2$ ,  $J$  5.0 Hz), 3.4 (s, 2 H,  $\text{CH}_2\text{COO}$ ), 4.2 (q, 2 H,  $\text{CH}_3\text{CH}_2\text{O}$ ,  $J$  7.0 Hz), 5.2 p.p.m. (br s, 1 H, carbamate NH);  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  13.8 ( $\text{CH}_3\text{CH}_2\text{O}$ ), 28.0 ( $(\text{CH}_3)_3\text{C}$ ), 39.8, 48.4 ( $\text{NHCH}_2$ ), 50.0 ( $\text{CH}_2\text{CO}$ ), 60.4 ( $\text{CH}_3\text{CH}_2\text{O}$ ), 78.6 ( $(\text{CH}_3)_3\text{C}$ ), 155.8 (carbamate CO), 172.0 p.p.m. (ester CO); Anal. Calcd. for  $\text{C}_{11}\text{H}_{22}\text{N}_2\text{O}_4$ : C, 53.63; H, 9.00; N, 11.38.

#### 6.2.6 *N*-(2-aminoethyl)glycine **29**

Ethylenediamine (105 mL, 11.58 mol) was rapidly stirred at 0 °C with portion-wise addition of chloroacetic acid **28** (15 g, 0.1158 mol) under an atmosphere of Ar, ensuring each portion had completely dissolved before adding the next. The reaction was allowed to stir at 25 °C for 12 h and evaporated in *vacuo*. The resultant white paste was triturated with DMSO (250 mL), filtered and the resulting cake washed with DMSO (50 mL) and diethyl ether (300 mL). The ensuing fine, white needles were recrystallised by dissolving in a minimum of hot water and adding 3 equivalents of EtOH. The mixture was re-warmed and placed overnight in a fridge to afford the title compound as a white, crystalline product (58%); mp. 140-142 °C; Lit.<sup>228</sup> 141-143 °C. MS (electrospray):  $m/z$  ( $I_r$ ) 120 (M+H, 100%), 102 (10%). Mass calcd. for  $\text{C}_4\text{H}_{10}\text{N}_2\text{O}_2$ : 119.082. Found: 119.082. Anal. calcd. for  $\text{C}_4\text{H}_{10}\text{N}_2\text{O}_2 \cdot \text{H}_2\text{O}$ : C, 31.2; H, 9.2; N, 18.2. Found: C, 31.1; H, 9.2; N, 18.0.

#### 6.2.7 2,6-Dichloro-9-(carboxymethyl)purine ethyl ester **76** and 2,6-dichloro-7-(carboxymethyl)purine ethyl ester **78**

To a stirred solution of 2,6-dichloropurine **75** (5.36 g, 28.3 mmol) in anhydrous  $\text{CH}_3\text{CN}$  (40 mL) was added  $\text{K}_2\text{CO}_3$  (4.71 g, 34.1 mmol) followed by  $\text{BrCH}_2\text{COCH}_2\text{CH}_3$  **27** (5.57 g, 33.0 mmol). The product solution was stirred at r.t. for 48 h, under an atmosphere of Ar, filtered through Celite, and the solvent evaporated under vacuum. The residue was purified using flash column chromatography on silica and **76** eluted first with EtOAc. Recrystallisation from MeOH gave the title compound **76** (4.00 g, 52%) as colourless crystals; mp 121-123 °C; TLC (EtOAc):  $R_f$  0.55; IR (KBr disc):  $\nu_{\text{max}}$  3109, 3001, 1731, 1586, 1554, 1401, 1344, 1236, 1152, 1018  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  1.21 (t, 3 H,  $J$  7.1 Hz,  $\text{CH}_3$ ), 4.20 (q, 2 H,  $J$  7.1 Hz,  $\text{CH}_2$ ), 5.24 (s, 2 H,  $\text{CH}_2$ ), 8.70 p.p.m. (s, 1 H, H-8);  $^{13}\text{C}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  14.0 ( $\text{CH}_3$ ), 44.5 ( $\text{CH}_2$ ), 62.8 ( $\text{CH}_2$ ), 130.3 (C-5), 146.2 (C-8), 151.8 (C-6), 153.1 (C-2), 153.2 (C-4), 166.1 p.p.m. (CO); MS (CI):  $m/z$  ( $I_r$ ) 279 (M+H, 14%), 277 (M+H, 30%), 275 (M+H, 100%), 243 (4%), 241 (12%); Anal. calcd. for  $\text{C}_9\text{H}_8\text{Cl}_2\text{N}_4\text{O}_2$ : C, 39.3; H, 2.9; Cl, 25.8; N, 20.4. Found: C, 39.4; H, 2.7; Cl, 25.6; N, 20.2.



Further elution with EtOAc-MeOH (3:1) gave **78** (2.53 g, 32%) as a viscous, dark yellow crystalline residue; mp 78-79 °C; TLC (EtOAc):  $R_f$  0.45; IR (KBr disc):  $\nu_{\max}$  3108, 2897, 1752, 1602, 1535, 1477, 1408, 129, 1228, 1176, 1095, 994, 873  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  1.21 (t, 3 H,  $J$  7.1 Hz,  $\text{CH}_3$ ), 4.20 (q, 2 H,  $J$  7.1 Hz,  $\text{CH}_2$ ), 5.44 (s, 2 H,  $\text{CH}_2$ ), 8.83 p.p.m. (s, 1 H, H-8);  $^{13}\text{C}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  14.1 ( $\text{CH}_3$ ), 48.1 ( $\text{CH}_2$ ), 62.1 ( $\text{CH}_2$ ), 122.6 (C-5), 143.8 (C-6), 150.8 (C-8), 153.4 (C-2), 153.9 (C-4), 167.3 p.p.m. (CO); MS (electrospray):  $m/z$  ( $I_r$ ) 279 (M+H, 8%), 277 (M+H, 28%), 275 (M+H, 11%), 274 (M, 40%), 203 (66%), 201 (100%), 86 (42%), 77 (40%); Anal. calcd. for  $\text{C}_9\text{H}_8\text{Cl}_2\text{N}_4\text{O}_2$ : C, 39.3; H, 2.9; Cl, 25.8; N, 20.4. Found: C, 39.5; H, 2.8; Cl, 25.5; N, 20.2.

Similarly prepared were:

**6.2.8 2,6-Dichloro-9-(carboxymethyl)purine methyl ester 75 and 2,6-dichloro-7-(carboxymethyl)purine methyl ester 79**

From 2,6-dichloropurine **75** (1.00 g, 5.31 mmol),  $\text{CH}_3\text{CN}$  (20 mL),  $\text{K}_2\text{CO}_3$  (0.88 g, 6.38 mmol) and  $\text{BrCH}_2\text{COCH}_3$  **88** (1.00 g, 6.38 mmol). Flash column chromatography on silica eluting with EtOAc, followed by recrystallisation of the concentrated product-containing fractions from MeOH furnished the first title compound **79** (538 mg, 39%) as colourless crystals; mp 151-153 °C; TLC (EtOAc):  $R_f$  0.25. IR (KBr disc):  $\nu_{\max}$  3108, 2998, 2958, 1739, 1596, 1556, 1379, 1340, 1234, 1157, 951, 879  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  3.73 (s, 3 H,  $\text{CH}_3$ ), 5.26 (s, 2 H,  $\text{CH}_2$ ), 8.70 p.p.m. (s, 1 H, H-8);  $^{13}\text{C}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  45.9 ( $\text{CH}_2$ ), 54.3 ( $\text{CH}_2$ ), 130.3 (C-5), 150.3 (C-8), 151.5 (C-6), 156.2, (C-2) 162.2 (C-4), 168.0 p.p.m. (CO); MS ( $\text{CI}^+$ ):  $m/z$  ( $I_r$ ) 265 (M+H, 4%), 263 (M+H, 7%), 261 (M+H, 10%), 260 ( $\text{M}^+$ , 13%), 204 (13%), 203 (23%), 201 (33%), 77 (40%), 59 (100%); Anal. calcd. for  $\text{C}_8\text{H}_6\text{Cl}_2\text{N}_4\text{O}_2$ : C, 36.7; H, 2.7; Cl, 27.1; N, 21.4. Found: C, 36.8; H, 2.3; Cl, 27.2; N, 21.1.

Further elution with EtOAc with concentration of product-containing fractions and recrystallisation from MeOH furnished the second title compound **79** (392 mg, 28%) as a yellow crystalline product. mp 147-149 °C; TLC (EtOAc):  $R_f$  0.21; IR (KBr disc):  $\nu_{\max}$  3109, 2957, 1747, 1610, 1387, 1257, 1174, 987, 883  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  4.03 (s, 3 H,  $\text{CH}_3$ ), 5.45 (s, 2 H,  $\text{CH}_2$ ), 8.82 p.p.m. (s, 1 H, H-8);  $^{13}\text{C}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  48.0 ( $\text{CH}_3$ ), 61.4 ( $\text{CH}_2$ ), 138.0 (C-5), 148.9 (C-8), 151.4 (C-6), 157.6 (C-2), 162.0 (C-4), 168.5 p.p.m. (CO); MS ( $\text{CI}^+$ ):  $m/z$  ( $I_r$ ) 265 (M+H, 4%), 263 (M+H, 5%), 261 (M+H, 10%), 260 ( $\text{M}^+$ , 61%), 201 (92%), 165 (32%), 113 (31%), 104 (60%), 77 (39%), 59 (100%), 45 (27%), 41 (28%); Anal. calcd. for  $\text{C}_8\text{H}_6\text{Cl}_2\text{N}_4\text{O}_2$ : C, 36.8; H, 2.3; Cl, 27.2; N, 21.5. Found: C, 37.0; H, 2.3; Cl, 27.1; N, 21.2.

### 6.2.9 6-Chloro-9-(carboxymethyl)purine ethyl ester 125 and 6-Chloro-7-(carboxymethyl)purine ethyl ester 126

To a solution of 6-chloropurine **89** (3.52 g, 22.8 mmol) in dry CH<sub>3</sub>CN (25 mL) was added K<sub>2</sub>CO<sub>3</sub> (3.78 g, 27.4 mmol) and BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub> **27** (4.59 g, 22.8 mmol). After stirring at r.t. under Ar for 48 h, the product solution was filtered through Celite and the solvent evaporated under vacuum. The pale yellow residue was purified by flash column chromatography on silica. **125** was eluted first with EtOAc, followed by concentration of the filtrate and recrystallisation from MeOH to furnish **125** (3.58 g, 65%) as colourless crystals; mp 95 °C; TLC (EtOAc): *R<sub>f</sub>* 0.43; IR (KBr disc):  $\nu_{\max}$  3107, 2938, 1733, 1566, 1500, 1438, 1344, 1195, 939, 779 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  1.20 (t, 3 H, *J* 7.13 Hz, CH<sub>3</sub>), 4.19 (q, 2 H, *J* 7.13 Hz, OCH<sub>2</sub>), 5.27 (s, 2 H, CH<sub>2</sub>), 8.68 (s, 1 H, H-2), 8.79 p.p.m. (s, 1 H, H-8); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  14.1 (CH<sub>3</sub>), 44.9 (OCH<sub>2</sub>), 61.9 (CH<sub>2</sub>), 130.7 (C-5), 148.1 (C-8), 149.4 (C-4), 152.0 (C-2), 152.2 (C-6), 167.6 p.p.m. (CO); MS (EI): *m/z* (*I<sub>r</sub>*) 242 (M, 13%), 240 (M, 28%), 195 (8%), 167 (100%), 104 (18%), 86 (20%), 77 (36%); Anal. calcd. for C<sub>9</sub>H<sub>9</sub>ClN<sub>4</sub>O<sub>2</sub>: C, 44.9; H, 3.8; Cl, 14.7; N, 23.3. Found: C, 45.0; H, 3.7; Cl, 14.7; N, 23.3.

Further elution with EtOAc-MeOH 4:1, with concentration of the fractions and recrystallisation from MeOH, gave the second title compound **126** (1.48 g, 27%) as pale yellow, hexagonal crystals; mp 107-110 °C; TLC (EtOAc): *R<sub>f</sub>* 0.32. IR (KBr disc):  $\nu_{\max}$  3120, 2933, 1733, 1562, 1440, 779 cm<sup>-1</sup>. <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  1.20 (t, 3 H, *J* 7.1 Hz, CH<sub>3</sub>), 4.22 (q, 2 H, *J* 7.1 Hz, CH<sub>2</sub>), 5.45 (s, 2 H, CH<sub>2</sub>), 8.77 (s, 1 H, H-2), 8.84 p.p.m. (s, 1 H, H-8); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  14.2 (CH<sub>3</sub>), 48.0 (CH<sub>2</sub>), 62.0 (CH<sub>2</sub>), 122.6 (C-5), 142.6 (C-4), 151.6 (C-8), 152.2 (C-2), 152.7 (C-6), 168.2 p.p.m. (CO); MS (EI): *m/z* (*I<sub>r</sub>*) 242 (M, 15%), 240 (M, 40%), 167 (100%), 140 (16%), 86 (25%), 77 (28%). Anal. calcd. for C<sub>9</sub>H<sub>9</sub>ClN<sub>4</sub>O<sub>2</sub>: C, 44.9; H, 3.8; Cl, 14.7; N, 23.3. Found: C, 45.0; H, 3.7; Cl, 14.5; N, 23.1.

### 6.2.10 6-Chloro-9-(carboxymethyl)purine methyl ester 127 and 6-Chloro-7-(carboxymethyl)purine methyl ester 128

From 6-chloropurine **84** (2.0 g, 12.93 mmol), CH<sub>3</sub>CN (40 mL), K<sub>2</sub>CO<sub>3</sub> (2.14 g, 15.52 mmol) and BrCH<sub>2</sub>COCH<sub>3</sub> **88** (2.38 g, 15.56 mmol). Flash column chromatography eluting with EtOAc followed by recrystallisation of the concentrated product-containing fractions from MeOH furnished the title compound **127** (1.53 g, 52%) as pale yellow crystals; mp 113-116 °C; TLC (EtOAc): *R<sub>f</sub>* 0.48; IR (KBr disc):  $\nu_{\max}$  3103, 2971, 2935, 1733, 1601, 1423, 1332, 1178, 964, 792 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  3.71 (s, 3 H, CH<sub>3</sub>), 5.28 (s, 2 H, CH<sub>2</sub>), 8.67 (s, 1 H, H-8), 8.79 p.p.m. (s, 1 H, H-2); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  44.8 (CH<sub>2</sub>), 52.9 (CH<sub>2</sub>), 130.7 (C-5), 148.1 (C-8), 149.4 (C-6), 152.1 (C-2), 152.2 (C-4), 168.1 p.p.m. (CO); MS (EI): *m/z* (*I<sub>r</sub>*) 228 (M., 13%), 226 (M, 33%), 170 (18%), 169 (34%), 168 (56%), 167 (100%), 140 (26%), 104 (64%), 86 (41%), 79

(44%), 77 (71%), 59 (64%); Anal. calcd. for  $C_8H_7ClN_4O_2$ : C, 42.4; H, 3.1; Cl, 15.6; N, 24.7. Found: C, 42.6; H, 3.1; Cl, 15.5; N, 24.6.

Further elution with EtOAc-MeOH (3:1) with concentration of the product-containing fractions, followed by recrystallisation from MeOH, furnished **128** (0.85 g, 30%) as a pale yellow powder. mp 107-110 °C; TLC (EtOAc):  $R_f$  0.24; IR (KBr disc):  $\nu_{max}$  3102, 2971, 2935, 1751, 1595, 1567, 1498, 1336, 1336, 1220, 970  $cm^{-1}$ ;  $^1H$  NMR [ $(CD_3)_2SO$ ]:  $\delta$  3.74 (s, 3 H,  $CH_3$ ), 5.47 (s, 2 H,  $CH_2$ ), 8.76 (s, 1 H, H-8), 8.84 p.p.m. (s, 1 H, H-2);  $^{13}C$  NMR [ $(CD_3)_2SO$ ]:  $\delta$  14.2 ( $CH_3$ ), 44.1 ( $CH_3$ ), 51.8 ( $CH_2$ ), 129.2 (C-5), 151.3 (C-8), 149.4 (C-6), 152.1 (C-4), 152.2 (C-2), 168.1 p.p.m. (CO); MS (EI):  $m/z$  ( $I_r$ ) 228 (M, 25%), 226 (M, 82%), 181 (12%), 170.2 (12%), 169 (41%), 168 (67%), 140 (29%), 104 (47%), 86 (27%), 79 (28%), 77 (40%), 59 (48%); Anal. calcd. for  $C_8H_7ClN_4O_2$ : C, 42.4; H, 3.1; Cl, 15.6; N, 24.7. Found: C, 42.5; H, 3.1; Cl, 15.4; N, 24.6.

#### 6.2.11 2,6-Diazido-9-(carboxymethyl)purine methyl ester **84**

A mixture of **76** (351 mg, 1.34 mmol), sodium azide (95 mg, 6.15 mmol),  $(CH_3)_2CO$  (15 mL) and MeOH (55 mL) was refluxed for 4 d at 80 °C. The solvent was then evaporated to produce a brown residue which was purified using flash column chromatography, eluting with EtOAc. Recrystallisation from MeOH produced the title compound **84** (74 mg, 20%) as pale yellow-brown crystals. mp 126-128 °C; TLC (EtOAc):  $R_f$  0.34; IR (KBr disc):  $\nu_{max}$  3109, 2949, 2842, 2131, 1737, 1616, 1577, 1348, 1234, 995, 788, 626  $cm^{-1}$ ;  $^1H$  NMR [ $(CD_3)_2SO$ ]:  $\delta$  3.71 (s, 3 H,  $CH_3$ ), 5.16 (s, 2 H,  $CH_2$ ), 8.41 p.p.m. (s, 1 H, H-8);  $^{13}C$  NMR [ $(CD_3)_2SO$ ]:  $\delta$  40.7 ( $CH_2$ ), 52.9 ( $CH_3$ ), 120.8 (C-5), 146.2 (C-8), 152.9 (C-6), 154.3, (C-2), 155.1 (C-4), 168.2 p.p.m. (CO); MS (electrospray):  $m/z$  ( $I_r$ ) 275 (M+H 28%), 246 (18%), 160 (44%), 107 (28%), 95 (46%), 80 (74%), 59 (100%), 41 (36%), 28 (64%), 15 (76%); Acc. Mass for  $C_8H_6N_{10}O_2$ : Calcd 274.068. Found 274.068.

#### 6.2.12 2,6-Diamino-9-(carboxymethyl)purine methyl ester **81**

A mixture of **84** (4.04 g, 14.01 mmol), suspended in anhydrous EtOH (50 mL) with 10% palladium on charcoal (300 mg), was hydrogenolised at r.t. under  $H_2$  over 4 d. The product mixture was filtered through Celite and the solvent evaporated under vacuum to produce a crude, pale lilac product mixture. The title compound **81** was purified by flash column chromatography eluting with  $CHCl_2$ -MeOH; 8:1, followed by recrystallisation from EtOAc-MeOH (8:1) to furnish **81** as colourless crystals (538 mg, 39%); mp 144-145 °C; TLC (EtOAc-MeOH):  $R_f$  0.23; IR (KBr disc):  $\nu_{max}$  3425, 3327, 3215, 3164, 1753, 1603, 1472, 1417, 1346, 1282, 1221, 775, 650,  $cm^{-1}$ ;  $^1H$  NMR [ $(CD_3)_2SO$ ]:  $\delta$  4.12 (3 H,  $OCH_3$ ), 4.84 (s, 2 H,  $CH_2$ ), 5.84 (s, 2 H,  $NH_2$ ,  $D_2O$  exchangeable), 6.70 (s, 2 H,  $NH_2$ ,  $D_2O$  exchangeable), 7.67 p.p.m. (s, 1 H, H-8);  $^{13}C$  NMR [ $(CD_3)_2SO$ ]:  $\delta$

40.7 (CH<sub>3</sub>), 61.3 (OCH<sub>2</sub>), 112.8 (C-5), 137.8 (C-8), 156.2 (C-6), 160.5 (C-2), 162.6 (C-4), 168.3 p.p.m. (CO); MS (electrospray): *m/z* (*I<sub>r</sub>*) 223 (M+H, 100%), 205 (46%), 175 (40%); Acc. Mass for C<sub>8</sub>H<sub>10</sub>N<sub>6</sub>O<sub>2</sub>: Calcd 222.094. Found 222.094; Anal. calcd. for C<sub>8</sub>H<sub>10</sub>N<sub>6</sub>O<sub>2</sub>: C, 43.2; H, 4.5; N, 30.7 Found: C, 43.3; H, 4.8; N, 30.6.

### 6.2.13 2,6-Diamino-9-(carboxymethyl)purine ethyl ester **80**

#### Method A

2,6-Diazido-9-(carboxymethyl)purine methyl ester **84** (4.04 g, 14.1 mmol) was suspended in anhydrous EtOH (25 mL) and hydrogenolysed using 10% palladium on charcoal (300 mg) under H<sub>2</sub> for 4 d at r.t. The product mixture was filtered through Celite and the solvent evaporated under vacuum to give **80** in its crude form. The residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub>-MeOH and adsorbed onto silica. Flash column chromatography at medium pressure eluting with CH<sub>2</sub>Cl<sub>2</sub>-MeOH 8:1 followed by recrystallisation from EtOAc and MeOH afforded **80** (538 mg, 39%) as colourless crystals; mp 173-174 °C; TLC (CH<sub>2</sub>Cl<sub>2</sub>-MeOH): *R<sub>f</sub>* 0.23. IR (KBr disc):  $\nu_{\max}$  3388, 3306, 3168, 2981, 1747, 1603, 1429, 1348, 1259, 1024, 802, 749 cm<sup>-1</sup>. <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  1.20 (t, 3 H, *J* 7.1 Hz, CH<sub>3</sub>), 4.12 (q, 2 H, *J* 7.1 Hz, CH<sub>2</sub>), 4.84 (s, 2 H, CH<sub>2</sub>), 5.85 (s, 2 H, NH<sub>2</sub>, D<sub>2</sub>O exchangeable), 6.73 (s, 2 H, NH<sub>2</sub>, D<sub>2</sub>O exchangeable), 7.67 (s, 1 H, H-8). <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  14.2 (CH<sub>3</sub>), 43.6 (CH<sub>2</sub>), 61.4 (CH<sub>2</sub>), 112.8 (NH<sub>2</sub>), 138.0 (C-5), 152.2 (C-6), 156.3 (C-8), 160.6 (C-2), 162.2 (C-4), 168.5 (CO); MS (EI): *m/z* (*I<sub>r</sub>*) 236 (M, 50%), 163 (100%), 146 (30%), 94 (20%), 67 (22%); Anal. calcd. for C<sub>9</sub>H<sub>12</sub>N<sub>6</sub>O<sub>2</sub>·0.5 H<sub>2</sub>O: C, 44.1; H, 5.3; N, 34.3 Found: C, 43.9; H, 5.0; N, 34.0.

#### Method B

To a solution of 2,6-diaminopurine **2** (2.21 g, 14.7 mol) in anhydrous DMF (35 mL) was added NaH (60% dispersion in mineral oil) (0.77 g, 19.3 mmol) in small portions with stirring under Ar. The reaction mixture was stirred for a further 4 h at r.t. followed by addition of BrCH<sub>2</sub>CO<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub> **27** during 4 h. After stirring overnight, the solvent was evaporated and the residual orange oil was shaken with H<sub>2</sub>O (30 mL), causing the precipitation of **80** which was recrystallised from EtOAc to afford the title compound as a pale orange powder (2.39 g, 79%). All analytical data were identical to those obtained using Method A.

### 6.2.14 2,6-Diamino-9-(carboxymethylpurine)acetic acid **82**

To solution of 2,6-diamino-9-(carboxymethyl)purine ethyl ester **80** (1.0 g, 4.2 mmol) in H<sub>2</sub>O (20 mL) was added 2 M NaOH (12 mL). The solution was boiled for a period of 15 min and the resultant product mixture cooled to 0 °C and filtered through Celite. The title compound **80** was precipitated via addition of 4 M HCl acid and the yellow precipitate collected by filtration. Drying *in vacuo* afforded (0.62 g, 70%) **82** as a pale yellow powder; mp >300 °C; TLC (EtOAc-MeOH 4:1): *R<sub>f</sub>* 0.30; IR (KBr disc):  $\nu_{\max}$  3506,

3388, 3168, 2981, 1746, 1651, 1589, 1471, 1415, 1240, 1149, 864, 740, 609  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR  $[(\text{CD}_3)_2\text{SO}]$ :  $\delta$  4.79 (s, 2 H,  $\text{CH}_2$ ), 5.83 (b, 2 H,  $\text{NH}_2$ ,  $\text{D}_2\text{O}$  exchangeable), 6.69 (b, 2 H,  $\text{NH}_2$ ,  $\text{D}_2\text{O}$  exchangeable), 7.66 (s, 1 H, H-8), 12.23 p.p.m. (b, 1 H,  $\text{CO}_2\text{H}$ ,  $\text{D}_2\text{O}$  exchangeable);  $^{13}\text{C}$  NMR  $[(\text{CD}_3)_2\text{SO}]$ :  $\delta$  43.8 ( $\text{CH}_2$ ), 112.8 (C-5), 138.2 (C-8), 152.4 (C-6), 156.2 (C-2), 160.4 (C-4), 169.9 p.p.m. (CO); MS (CI<sup>+</sup>):  $m/z$  ( $I_r$ ) 209 (M+H, 13%), 193 (3%), 165 (34%), 151 (100%), 136 (10%), 44 (9%); Anal. calcd. for  $\text{C}_7\text{H}_8\text{N}_6\text{O}_2 \cdot \text{H}_2\text{O}$ : C, 36.6; H, 4.4; N, 37.2. Found: C, 36.9; H, 4.4; N, 36.9.

#### 6.2.15 6-Amino-9-(carboxymethyl)-2-methoxypurine methyl ester **85**

A mixture of 2,6-dichloro-9-(carboxymethyl)purine ethyl ester **76** (351 mg, 1.34 mmol), sodium azide (95 mg, 6.15 mmol),  $(\text{CH}_3)_2\text{CO}$  (15 mL) and MeOH (55 mL) was refluxed for 4 d at 80 °C. The solvent was then evaporated to produce a brown residue which was purified using flash column chromatography, eluting with EtOAc. 2,6-diazido-9-(carboxymethyl)purine methyl ester **84** was suspended in anhydrous EtOH (25 mL) and hydrogenolysed using 10% palladium on charcoal (300 mg) under  $\text{H}_2$  for 4 d at r.t. The product mixture was filtered through Celite and the solvent evaporated under vacuum to give **80** in its crude form, followed by recrystallisation from EtOAc-MeOH, furnished **85**. Recrystallisation from MeOH produced the title compound **85** as colourless crystals. Only minute quantities (~ 30 mg) were successfully isolated from the product mixture and were analysed by X-ray crystallography; mp 116-119 °C; TLC ( $\text{CH}_2\text{Cl}_2$ -MeOH):  $R_f$  0.27. IR (KBr disc):  $\nu_{\text{max}}$  3360, 3210 3168, 2981, 1757, 1603, 1448, 1259, 1115, 802, 749  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR  $[(\text{CD}_3)_2\text{SO}]$ :  $\delta$  1.22 (s, 3 H,  $\text{CH}_3$ ), 3.17 (s, 3 H,  $\text{OCH}_3$ ), 4.17 (s, 2 H,  $\text{CH}_2$ ), 6.13 (s, 2 H,  $\text{NH}_2$ ,  $\text{D}_2\text{O}$  exchangeable), 7.69 (s, 1 H, H-8).  $^{13}\text{C}$  NMR  $[(\text{CD}_3)_2\text{SO}]$ :  $\delta$  14.7 ( $\text{CH}_3$ ), 42.9 ( $\text{CH}_2$ ), 61.7 ( $\text{CH}_2$ ), 114.8 ( $\text{NH}_2$ ), 137.3 ( $\text{CH}_2$ ), 152.6 (C-4), 156.3 (C-8), 161.5 (C-2), 162.4 (C-6), 167.2 (CO).

#### 6.2.16 *N*-(Benzyloxycarbonyl)imidazole **100**

Benzyl chloroformate **97** (18.75 g, 0.11 mol) was added dropwise to a solution of imidazole **99** (15 g, 0.22 mol) in 150 mL at 0 °C, during a period of 1 h. The mixture was allowed to stir overnight under Ar, at r.t. The white viscous solution was filtered, concentrated *in vacuo*, and recrystallised from petroleum ether to yield *N*-(benzyloxycarbonyl)imidazole **100** (13.74 g, 68 %); mp 35-38 °C. Lit. mp<sup>243</sup> 37-38.5 °C TLC ( $\text{CH}_2\text{Cl}_2$ -MeOH 4:1):  $R_f$  0.30; IR (KBr disc):  $\nu_{\text{max}}$  3110, 3035, 2929, 1724, 1596, 1471, 1400, 1351, 1109, 725, 690  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR  $[(\text{CD}_3)_2\text{SO}]$ :  $\delta$  5.3 (s, 2 H,  $\text{CH}_2$ ), 7.0 (m, 1 H), 7.3, (m, 6 H,  $\text{C}_6\text{H}_5$ ), 8.0 (m, 1 H, C-5);  $^{13}\text{C}$  NMR  $[(\text{CD}_3)_2\text{SO}]$ :  $\delta$  65.0 ( $\text{CH}_2$ ), 116.8 (CH), 128.1 (m,  $\text{C}_6\text{H}_5$ ), 136.8 (s, CH), 148.3 p.p.m. (CO); MS (electrospray):  $m/z$  ( $I_r$ ) 413 (M+H, 32%), 391 (100%), 367 (10%), 176 (66%), 157 (49%), 136 (52%). Anal. Calcd. for  $\text{C}_{11}\text{H}_{10}\text{N}_2\text{O}_2$ : C, 65.3; H, 5.0; N, 13.9. Found: C, 65.0; H, 4.9; N, 13.7.

### 6.2.17 1-(Benzyloxycarbonyl)-3-ethylimidazolium tetrafluoroborate **98**

Triethyloxonium tetrafluoroborate **101** (7.6 g, 40 mmol) was added portionwise at 0°C, under Ar, to *N*-Cbz-imidazole **100** (7.88 g, 38 mmol) in 110 mL CH<sub>2</sub>Cl<sub>2</sub>. The reaction mixture was allowed to approach r.t. and stirred for a further 3 h to furnish **98** which was isolated as a white oily product. <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]: δ 1.47 (t, 3 H, *J* = 7.1); 4.20 (q, 2 H, *J* = 7.1), 5.42 (s, 2 H), 7.3 (m, 6 H), 7.65 (m, 1 H), 8.98 (m, 1 H). mp 44.5-47 °C. Lit.<sup>243</sup> mp 45°C.

### 6.2.18 6-Amino-2-chloropurine **110**

A solution of 2,6-dichloropurine **75** (0.55 g, 2.9 mmol) in saturated NH<sub>3</sub>/MeOH solution (80 mL) was added to a pressure bomb and left to stir at 75 °C for 28 h. The solvent was evaporated to give the crude product **110** which was recrystallised from MeOH to yield **110** as a white powder; mp 226 °C; TLC (EtOAc): *R*<sub>f</sub> 0.29; IR (KBr disc):  $\nu_{\max}$  defective something above 3395, 2412, 1707, 1581, 933, 875, 634 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]: 4.08 (s, 2 H, CH<sub>2</sub>), 7.65 (s, 2 H, 2 NH<sub>2</sub>), 8.12 (s, 2 H, 2 -ONH<sub>2</sub>), 8.39 p.p.m. (s, 1 H, H-8); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]: δ 128.8 (C-5), 147.8 (C-8), 151.0 (C-6), 153.0 (C-2), 156.4 (C-5) p.p.m. (C-4); MS (electrospray): *m/z* (*I*<sub>r</sub>) 188 (100%), 170 (M+H, 4%), 153 (68%), 92 (38%), 73 (27%), 65 (34%), 53 (32%), 38 (35%). Acc. Mass for C<sub>5</sub>H<sub>4</sub>Cl: Calcd. 99.641. Found 99.641.

### 6.2.19 6-Amino-2-Chloro-9-(carboximido)purine **112**

A solution of **110** (50 mg, 0.18 mmol) in saturated NH<sub>3</sub>/MeOH (80 mL) was added to a bomb and left to stir at 75 °C for 30 h. The solvent was evaporated to give the crude title product which was purified by flash chromatography on silica, eluting with CH<sub>2</sub>Cl<sub>2</sub>-MeOH (2:1) to yield the **112** as a white powder; mp 228 °C; TLC (CH<sub>2</sub>Cl<sub>2</sub>-MeOH 3:1): *R*<sub>f</sub> 0.44; IR (KBr disc):  $\nu_{\max}$  3286, 2358, 1706, 1639, 1597, 1351, 1313, 1097, 937, 463 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]: δ 4.76 (q, 2 H, *J* 2.17 Hz, CH<sub>2</sub>), 7.35 (s, 2 H, 2 NH<sub>2</sub>), 7.76 (s, 2 H, 2 CONH<sub>2</sub>), 8.06 p.p.m. (s, 1 H, H-8); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]: δ 53.5 (CH<sub>2</sub>), 117.6 (C-6), 142.7 (C-2), 151.1 (C-4), 153.1 (C-5), 156.9 (C-8), 168.2 (CO); MS (EI): *m/z* (*I*<sub>r</sub>) 228 (M, 14%), 226 (29%), 184 (30%), 183 (58%), 182 (100%), 146 (18%), 94 (22%), 91(23%), 66 (24%), 44 (29%). Anal. calcd. for C<sub>7</sub>H<sub>7</sub>ClN<sub>6</sub>O: C, 37.1; H, 3.1; Cl, 15.6; N, 37.1. Found: C, 37.4; H, 2.9; Cl, 15.4; N, 37.2.

### 6.2.20 Attempted synthesis of 6-amino-2-benzyloxy-9-(carboxymethyl)purine ethyl ester **113**

Under a stream of Ar, sodium metal (0.8 g, \$ mmol) was dissolved in benzyl alcohol (20 mL) during 30 h with stirring at 70 °C to form **117**, to which was added 6-amino-2-chloropurine ethyl ester **111** (223 mg). After stirring for 72 h, the resulting yellow, thick solution was allowed to cool and added to a column of silica. The benzyl alcohol was

eluted with CH<sub>2</sub>Cl<sub>2</sub> and **113** was eluted with CH<sub>2</sub>Cl<sub>2</sub>-MeOH (10:3). Evaporation of the solvent and filtering to remove any remaining impurities yielded a crude residue. Attempted purification to yield crude **113**, produced a white powder. Spectroscopic data did not reveal a multiplet, characteristic of an expected benzyl group, around 57.3.

#### 6.2.21 2,6-Dibenzoyloxy-xanthin-9-yl acetic acid **121**

##### Method A

A 1 M solution of sodium benzyloxalate **117** was prepared under Ar by addition of sodium metal (1.35 g, 58.7 mmol) to benzyl alcohol (60 mL, 11.64 mmol), with stirring at 65 °C. Compound **76** (1.97 mg, 71.6 mmol) was added to the resultant gelatinous mixture and left to stir under Ar at 60 °C for 60 h to produce a yellow product mixture. On cooling the highly viscous solution was added to a column of silica and eluted at high pressure, via flash column chromatography with CH<sub>2</sub>Cl<sub>2</sub> to remove the benzyl alcohol. The crude form of the title product **121** was eluted using gradual increasing polarity of the eluent; CH<sub>2</sub>Cl<sub>2</sub>-MeOH (9:1; 7:1; 4:1). Concentration of the product solution to remove the solvent, followed by recrystallisation gave **121** (0.98 g, 35%) as a pale crystalline powder. mp 222-225 °C. TLC (CH<sub>2</sub>Cl<sub>2</sub>-MeOH 4:1): *R<sub>f</sub>* 0.30; IR (KBr disc):  $\nu_{\max}$  3110, 3035, 2929, 1724, 1596, 1471, 1400, 1351, 1109, 725, 690 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  4.57 (s, 2 H, CH<sub>2</sub>), 5.64 (s, 2 H, CH<sub>2</sub>), 7.34-7.73 (m, 2 x C<sub>6</sub>H<sub>5</sub>), 8.25 (s, 1 H, C-8), 8.63 p.p.m. (s, 1 H, C-2), 12.17 (bs, 1 H, OH, D<sub>2</sub>O exchangeable); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  47.4 (OCH<sub>2</sub>), 68.19 (CH<sub>2</sub>), 68.9 (CH<sub>2</sub>), 120.2 (C-5), 128.7 (C<sub>6</sub>H<sub>5</sub>), 136.5 (C-6), 145.5 (C-8), 151.1 (C-2), 152.8 (C-4), 160.3 (C-O); 160.7 p.p.m. (C-O); MS (electrospray): *m/z* (*I<sub>r</sub>*) 391 (M+H, 100%), 367 (10%), 176 (66%), 157 (49%), 136 (52%); Anal. Calcd. for C<sub>21</sub>H<sub>18</sub>N<sub>4</sub>O<sub>4</sub>: C, 64.4; H, 4.9; N, 14.3. Found: C, 64.2; H, 5.2; N, 14.6. Mass Calcd. 391.140. Mass Found 391.140.

##### Method B

Sodium metal (1.18 g, 51 mmol) was dissolved in benzyl alcohol (20 mL) and heated to 80 °C for 18 h to produce **117**. A solution of **76** (1.81 g, 6.58 mmol) dissolved in DMF (10 mL) was added to **117**, and the resulting orange suspension stirred under Ar at 80 °C for 30 h. After cooling to r.t. 1 M NaOH (aq) (100 mL) was added and the clear yellow solution washed with EtOAc (3 x 100 mL). The aqueous phase was acidified to pH 3 using 4 M HCl acid. The resultant precipitate was taken up in EtOAc (2 x 100 mL) and the aqueous phase extracted with EtOAc (200 mL). The combined organic phases were washed with brine (2 x 75 mL), dried using Na<sub>2</sub>SO<sub>4</sub>, filtered and evaporated *in vacuo*. Recrystallisation from MeOH furnished **121** (1.78 g, 69%) as pale white crystals. mp 223-224 °C. All analytical data were identical to those attained in method A.

### 6.2.22 2,6-Diallyloxy-xanthin-9-yl acetic acid 122

Sodium metal (0.4 g, 17.4 mmol) was dissolved in allyl alcohol (15 mL, 0.22 mol) under Ar at 0 °C during 1 h to produce **118**. **76** (0.344 g, 1.25 mmol) was dissolved in toluene (3 x 5 mL), added to **118** and left to stir for 6 h at 120 °C. The allyl alcohol was removed *in vacuo* and the crude product mixture purified by flash chromatography eluting with CH<sub>2</sub>Cl<sub>2</sub>-MeOH 2:1 to yield **122** as a feint-white powder. mp 276 °C; TLC (CH<sub>2</sub>Cl<sub>2</sub>-MeOH 2:1): *R<sub>f</sub>* 0.26; IR (KBr disc):  $\nu_{\max}$  3411, 3141, 1616, 1404, 1095, 675, 474 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  4.29 (s, 1 H, CH), 4.82 (d, 2 H, *J* 7.2 Hz, CH<sub>2</sub>), 5.28 (d, 2 H, *J* 7.2 Hz, CH<sub>2</sub>), 5.46 (d, 2 H, *J* 7.2 Hz, CH<sub>2</sub>), 5.96-6.03 (m, 2 x 5 H, 2 x CH<sub>2</sub>CHCH<sub>2</sub>), 7.71 p.p.m. (s, 1 H, H-8); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  45.7 (CH<sub>2</sub>), 67.2 (CH<sub>2</sub>CHCH<sub>2</sub>), 67.9 (CH<sub>2</sub>CHCH<sub>2</sub>), 117.5 (C-5), 118.1 (CH<sub>2</sub>CHCH<sub>2</sub>), 118.7 (CH<sub>2</sub>CHCH<sub>2</sub>), 133.2 (CH<sub>2</sub>CHCH<sub>2</sub>) 133.6, (CH<sub>2</sub>CHCH<sub>2</sub>), 143.6 (C-8), 154.1 (C-6), 161.5 (C-2), 169.9 (C-4), 191.3 (CO), (191.5 p.p.m. (CO). Acc. Mass for C<sub>7</sub>H<sub>6</sub>N<sub>4</sub>O<sub>4</sub>: Calcd. 290.281. Found 290.281.

### 6.2.23 Xanthin-9-yl acetic acid 87

**121** in anhydrous EtOH (20 mL) was hydrogenolysed with 10% palladium on carbon (185 mg). The reaction mixture was then left to stir for 72 h at r.t. under H<sub>2</sub>. The resulting product mixture was filtered through Celite, washed with ethanol and evaporated to yield the title compound **87**; mp >300 °C; TLC (CH<sub>2</sub>Cl<sub>2</sub>-MeOH 4:1): *R<sub>f</sub>* 0.30; IR (KBr disc):  $\nu_{\max}$  3054, 2973, 1963, 1604, 1568, 1425, 1309, 1066, 921, 840 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  3.36 (b, 1 H, NH), 3.71 (s, 1 H, NH), 4.62 (s, 2 H, CH<sub>2</sub>), 7.55 (s, C-8), 10.74 p.p.m. (bs, OH); <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  68.2 (CH<sub>2</sub>), 136.5 (8-C), 136.9 p.p.m. (OH); MS (electrospray): *m/z* (*I<sub>r</sub>*) 211 (M+H, 22%), 176 (17%), 154 (100%), 120 (17%), 91 (35%).  $\delta$  47.4 (CH<sub>2</sub>), 120.4 (C-5), 138.1 (C-6), 146.2 (C-8), 153.2 (C-2), 152.4 (C-4), 161.9 p.p.m. (CO). Acc. Mass for C<sub>7</sub>H<sub>6</sub>N<sub>4</sub>O<sub>4</sub>: Calcd. 210.150. Found 210.150.

### 6.2.24 Diethyl 3,7-hypoxanthyl diacetate 124

To a solution of hypoxanthine **4** (2.10 g, 15.4 mmol) in dry CH<sub>3</sub>CN (30 mL), was added K<sub>2</sub>CO<sub>3</sub> (2.13 g, 47.6 mmol) and BrCH<sub>2</sub>COCH<sub>2</sub>CH<sub>3</sub> (2.57 g, 48.4 mmol). After stirring for 48 h at r.t., the product solution was filtered and the solvent evaporated under vacuum. The residue was purified by flash column chromatography on silica, eluting with EtOAc-MeOH (8:1). Recrystallisation from MeOH gave **124** (2.88 g, 61%). mp 139-141 °C; TLC (ethyl acetate-MeOH 8:1): *R<sub>f</sub>* 0.40; IR (KBr disc):  $\nu_{\max}$  3120, 3060, 2994, 2940, 1750, 1731, 1700, 1630, 1605, 1605, 1550, 1345, 1230, 1020 cm<sup>-1</sup>; <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  1.21 (t, 6 H, *J* 7.1 Hz, 2 x CH<sub>3</sub>), 4.13 (q, 2 H, *J* 7.1 Hz, OCH<sub>2</sub>), 4.18 (q, 2 H, *J* 7.1 Hz, OCH<sub>2</sub>), 4.83 (s, 2 H, CH<sub>2</sub>N), 5.11 (s, 2 H, CH<sub>2</sub>N), 8.11 (s, 1 H, C-8), 8.38 p.p.m. (s, 1 H, CH<sub>2</sub>). <sup>13</sup>C NMR [(CD<sub>3</sub>)<sub>2</sub>SO]:  $\delta$  14.2 (2 x CH<sub>3</sub>), 44.6, (CH<sub>2</sub>), 47.3, (CH<sub>2</sub>), 61.5 (OCH<sub>2</sub>), 61.7 (OCH<sub>2</sub>), 122.9 (CO), 141.8 (C-8), 148.3 (CO), 149.1



(C-2), 156.0 (C-6), 167.9 (C-5), 168.3 p.p.m. (C-4); MS (electrospray):  $m/z$  ( $I_r$ ) 308 (M, 100%), 263 (25%), 236 (98%), 207 (25%), 177(18), 163(10%), 149(15%), 134(14%); Anal. calcd. for  $C_{13}H_{16}N_4O_5$ : C, 50.6; H, 5.2; N, 18.2 Found: C, 50.4; H, 5.2; N, 17.9.

#### 6.2.25 6-Benzyloxy-hypoxanthin-9-yl acetic acid 132

##### Method A

**125** (1.03 g, 363 mmol) dissolved in DMF (10 mL) was added to **117** (20 mL), and the resulting suspension stirred at 80 °C for 24 h. 1 M NaOH (aqueous, 100 mL) was added and the clear yellow solution washed using EtOAc (3 x 100 mL). The aqueous phase was acidified to pH 3 with 4 M HCl acid and the resulting precipitate taken up in EtOAc (2 x 100 mL); the aqueous phase was extracted with EtOAc (200 mL). The combined organic phases were washed with brine (2 x 75 mL), dried using  $Na_2SO_4$ , filtered and evaporated *in vacuo*. The residue was recrystallised from EtOAc to afford **132** as pale yellow crystals. mp 148-151°C; TLC (EtOAc-MeOH; 8:1):  $R_f$  0.42 IR (KBr disc):  $\nu_{max}$  3417, 2964, 1903, 1732, 1626, 1473, 1423, 1356, 1236, 1143, 887, 802, 748  $cm^{-1}$ ;  $^1H$  NMR [( $CD_3$ ) $_2$ SO]:  $\delta$  4.54 (s, 2 H,  $CH_2$ ), 5.61 (s, 2 H,  $CH_2$ ), 7.37-7.51 (m, 5 H,  $C_6H_5$ ), 8.36 (s, 1 H, H-8), 8.53 (s, 1 H, H-2), 12.32 p.p.m. (bs, 1 H, OH,  $D_2O$  exchangeable);  $^{13}C$  NMR [( $CD_3$ ) $_2$ SO]:  $\delta$  44.5 ( $CH_2$ ), 67.8 ( $CH_2$ ), 120.3 (C-5), 136.4 (C-6), 144.3 (C-8), 151.8 (C-2), 152.8 (C-4), 161.9 p.p.m. (CO); MS ( $CI^+$ ):  $m/z$  ( $I_r$ ) 285 (M+H, 100%), 193 (17%), 160 (54%), 146 (48%); Anal. calcd. for  $C_{14}H_{12}N_4O_3 \cdot H_2O$ : C, 55.6; H, 4.7; N, 18.5. Found: C, 55.5; H, 4.6; N, 18.3.

##### Method B

The above compound **132** was also prepared from sodium benzyloxalate **117** (1.2 g, 4.4 mmol) using the method described for 2,6-dibenzyloxy-xanthin-9-yl acetic acid **121** (Method A), to give **132** (0.82 g, 66 %). Identical analytical data to as the above method for the synthesis of **132** was attained.

#### 6.2.26 6-Benzyloxycarbonyl-hypoxanthin-7-yl acetic acid 135

Under a stream of Ar, a 1 M sodium benzyloxalate solution **117** was prepared by adding sodium metal (1.35 g, 58.7 mmol) to benzyl alcohol (60 mL, 11.64 mmol), with stirring at 65 °C. A solution of **125** (0.89 g, 290 mmol) dissolved in DMF (10 mL) was added to **117** (20 mL). The resulting suspension was stirred at 80°C for 30 h under Ar and left to cool to r.t. A 1 M NaOH solution (aq., 100 mL) was added and the clear yellow solution washed with EtOAc (3 x 100 mL). The aqueous phase was acidified to pH 3 with 4 M HCl acid. The resulting precipitate was taken up in EtOAc (2 x 100 mL) and the aqueous phase was extracted with EtOAc (200 mL). The combined organic phases were washed with brine (2 x 75 mL), and dried with  $Na_2SO_4$ . The solution was filtered and concentrated *in vacuo*, to produce a residue which was recrystallised from EtOAc to

furnish **135** (0.46 g, 56%) as dark yellow, needle-shaped polycrystals; mp 177-179 °C; TLC (EtOAc):  $R_f$  0.63; IR (KBr disc):  $\nu_{\max}$  3112, 2964, 1730, 1621, 1477, 1420, 1238, 1114, 798  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  5.17 (s, 2 H,  $\text{CH}_2$ ), 5.56 (s, 2 H,  $\text{CH}_2$ ), 7.35-7.45 (m, 5 H,  $\text{C}_6\text{H}_5$ ), 8.46 (s, 1 H, H-8), 8.55 (s, 1 H, H-2), 12.34 p.p.m. (b, 1 H, OH);  $^{13}\text{C}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  24.4 ( $\text{CH}_3$ ), 47.1 ( $\text{CH}_2$ ), 68.8 ( $\text{CH}_2$ ), 120.2 (C-5), 128.5 ( $\text{C}_6\text{H}_5$ ), 136.7 (C-6), 145.5 (C-8), 151.1 (C-2), 152.8 (C-4), 159.7 p.p.m. (CO); MS (electrospray):  $m/z$  ( $I_r$ ) 285 (M+H, 97%), 249 (43%), 259 (13%), 227 (100%), 217 (40%), 151 (11%), 81 (40%); Anal. calcd. for  $\text{C}_{14}\text{H}_{12}\text{N}_4\text{O}_3 \cdot \text{H}_2\text{O}$ : C, 55.6; H, 4.7; N, 18.5. Found: C, 55.9; H, 4.4; N, 18.5.

### Attempted Syntheses of Novel PNA Monomers Composed of the D, Dz, X, H, and IsoG Purine Bases

#### 6.2.27 Attempted Synthesis of *N*-(2-Boc-aminoethyl)-*N*-(diazido-9-acetyl)glycine **102**

Ethyl *N*-(2-Boc-aminoethyl)glycinate **1** (52 mg, 0.21 mmol), DhbtOH (35 mg, 0.21 mmol), DMAP (26 mg, 0.21 mmol), and **70** (50 mg, 0.19 mmol) were dissolved in DMF and DCC (44 mg, 0.21 mmol) at 0 °C under Ar. The ice-bath was removed after 4 h and the mixture left to stir overnight. The precipitated DCU was removed by filtration and washed with  $\text{CH}_2\text{Cl}_2$  (2 x 5 mL). To the combined filtrates was added more  $\text{CH}_2\text{Cl}_2$  (5 mL) and the solution washed successively with dilute aqueous  $\text{NaHCO}_3$  (3 x 10 mL), dilute aqueous  $\text{KHSO}_4$  (2 x 12 mL), and finally brine (1 x 15 mL). The precipitate in the organic phase was removed by filtration and **70** was precipitated once more with petroleum ether at 0 °C to afford crude **70**. Subsequent purification was undertaken via redissolving **70** in  $\text{CH}_2\text{Cl}_2$  and precipitating once more with petroleum ether. The organic phase was evaporated to dryness *in vacuo*, leaving a pale yellow powder (12 mg, 8%) which showed two spots on TLC (EtOAc-MeOH; 8:1):  $R_f$  0.42. Solvent extraction failed to isolate the desired pure product from the reaction mixture. NMR spectra showed a presence of the desired peaks corresponding to the acquisition of the title compound. Additional peaks arising from the presence of slight impurities were also detected. From accurate mass analysis which gave the mass of the desired compound, and from NMR analysis, it was concluded that the title compound had been successfully synthesised.  $^1\text{H}$  NMR [ $(\text{CD}_3)_2\text{SO}$ ]:  $\delta$  1.28 (t, 3 H,  $J$  7.1 Hz,  $\text{CH}_3$ ), (s, 9 H,  $(\text{CH}_3)_3$ ), 3.20 (q, 2 H,  $J$  6.1 Hz,  $\text{CH}_2$ ), 3.56 (t, 2 H,  $J$  6.3 Hz,  $\text{CH}_2$ ), 4.16 (s, 2 H,  $\text{CH}_2$ ), 4.22 (2 H, q,  $J$  7.1 Hz,  $\text{CH}_2$ ), 6.18 (s, 2 H,  $\text{CH}_2$ ), 8.05 p.p.m (s, 1 H, H-8).

#### 6.2.28 Attempted Synthesis of *N*-(2-Boc-aminoethyl)-*N*-(diaminopurin-9-acetyl)glycine **71**

Ethyl *N*-(2-Boc-aminoethyl)glycinate (130 mg, 0.53 mmol), DhbtOH (86 mg, 0.53 mmol), DMAP (64.5 mg, 0.53 mmol), and **82** (100 mg, 0.48 mmol) and DCC (109 mg,

0.53 mmol) were added to DMF (6 mL) at 0 °C under Ar. Regular TLC analyses to monitor the reaction were undertaken; spots corresponding to starting materials failed to disappear and no new spots were observed. Increasing the reaction temperature to 30 °C, and changing the coupling agents and coupling salts failed to couple **1** with **71**; failure of the reaction was attributed to the insolubility of **82**.

#### 6.2.29 Attempted Synthesis of *N*-(2-Boc-aminoethyl)-*N*-(xanthin-9-acetyl)glycine **7**

Ethyl *N*-(2-Boc-aminoethyl)glycinate **1** (82 mg, 0.33 mmol), DhbtOH (60 mg, 0.37 mmol), and **87** (77 mg, 0.37 mmol) and DCC (92 mg, 0.45 mmol) added to DMF (7 mL) at 0 °C under Ar. The ice-bath was removed after 4 h and the mixture left to stir overnight. From TLC analysis (EtOAc-MeOH; 8:1), the reaction mixture consisted of several components, all starting materials without any new spots corresponding to products. Increasing the temperature to 30 °C and use of PyBroP as an alternative coupling agent to DCC failed to couple **1** to **87**. Failure of the reaction could be attributed to the poor solubility of the base in its unprotected form.

#### 6.2.30 Attempted Synthesis of *N*-(2-Boc-aminoethyl)-2,6-dibenzyloxy-*N*-(xanthin-9-acetyl)glycine **73**

Ethyl *N*-(2-Boc-aminoethyl)glycinate **1** (180 mg, 0.73 mmol), DhbtOH (131.3 mg, 0.081 mmol), and **121** (229 mg, 0.81 mmol) and DCC (181 mg, 0.88 mmol) were added to DMF (8 mL) at 0 °C under Ar. The ice-bath was removed after 4 h and the mixture left to stir overnight. The precipitated DCU was removed by filtration and washed with CH<sub>2</sub>Cl<sub>2</sub> (2 x 5 mL). To the combined filtrates was added more CH<sub>2</sub>Cl<sub>2</sub> (5 mL) and the solution washed successively with dilute aqueous NaHCO<sub>3</sub> (3 x 10 mL), dilute aqueous KHSO<sub>4</sub> (2 x 12 mL), and finally brine (1 x 15 mL). The precipitate from the organic phase was removed by filtration and **73** was precipitated once more with petroleum ether at 0 °C to afford crude **73**. Subsequent, attempted purification was undertaken via redissolving **73** in CH<sub>2</sub>Cl<sub>2</sub> and precipitating once more with petroleum ether. The organic phase was evaporated to dryness *in vacuo*. Repeated attempts to remove impurities using solvent extraction, which produced brown oily residues, and thus isolate **73** in a pure form were unsuccessful. <sup>1</sup>H NMR [(CD<sub>3</sub>)<sub>2</sub>SO]: δ 1.23 (t, 3 H, *J* 7.1 Hz, CH<sub>2</sub>CH<sub>3</sub>), 3.24 (s, 9 H, (CH<sub>3</sub>)<sub>3</sub>), 3.24 (2 H, CH<sub>2</sub>CO), 3.82 (2 H, *J* 6.1 Hz, CH<sub>3</sub>CH<sub>2</sub>), 4.01 (2 H, t, *J* 7.1 Hz, CH<sub>2</sub>), 4.18 (q, 2 H, *J* 7.1 Hz, CH<sub>2</sub>), 4.99 (t, 2 H, *J* 7.1 Hz, CH<sub>2</sub>), 5.33 (2 H, CH<sub>2</sub>), 5.62 (1 H, q, NH), 7.24-7.45 (2 x 5 H, m, 2 x C<sub>6</sub>H<sub>5</sub>), 7.70 (1 H, H<sub>8</sub>).

#### 6.2.31 Attempted Synthesis of *N*-(2-Boc-aminoethyl)-6-benzyloxy-*N*-(hypoxanthin-9-acetyl)glycine **72**

Ethyl *N*-(2-Boc-aminoethyl)glycinate **1** (48 mg, 0.19 mmol), DhbtOH (32 mg, 0.19

mmol), and **132** (50 mg, 0.18 mmol) and DCC (40 mg, 0.19 mmol), DMAP (24 mg, 0.19 mmol) were added to DMF at 0 °C under Ar. The ice-bath was removed after 4 h and the mixture left to stir overnight. The precipitated DCU was removed by filtration and washed with CH<sub>2</sub>Cl<sub>2</sub> (2 x 5 mL). To the combined filtrates was added more CH<sub>2</sub>Cl<sub>2</sub> (5 mL) and the solution washed successively with dilute aqueous NaHCO<sub>3</sub> (3 x 10 mL) dilute aqueous KHSO<sub>4</sub> (2 x 12 mL), and finally brine (1 x 15 mL). The precipitate in the organic phase was removed by filtration and **72** was precipitated once more with petroleum ether at 0 °C to afford **72** in crude yield. Subsequent purification was undertaken via redissolving **72** in CH<sub>2</sub>Cl<sub>2</sub> and precipitating once more with petroleum ether. The organic phase was evaporated to dryness *in vacuo*. TLC analyses with varying solvent systems only showed streaks. However, NMR analysis on the sample proved to be promising and very encouraging, where peaks corresponding to all the major moieties of the H monomer were observed, but with the presence of impurities. Due to time constraints, the isolation of **72** in its pure form was not attempted.

### 6.3 Crystallography

#### 6.3.1 Data Collection

All intensity data sets were measured on an Enraf-Nonias CAD4 diffractometer by the  $\omega$ -2 $\theta$  scan technique. Data for **76** were collected with monochromatic Mo-K $\alpha$  radiation to a maximum of  $\theta$  of 25 ° and processed without absorption correction. All remaining data sets for **76**, **125**, **126**, **80**, **84**, **92**, **124** and **17** were collected with Cu radiation to  $\theta = 67$  °, or higher if crystal quality justified it. Empirical absorption corrections based on  $\psi$  (psi) scans were applied to these data with CADABS.<sup>258</sup> Trial structures were obtained by direct methods with MULTAN,<sup>259</sup> or with SHELXS<sup>264</sup> if MULTAN failed. All non-hydrogen atoms not located via direct method were found in Fourier maps using the SHELXL<sup>262</sup> program (Sheldrick), followed by full-matrix least squares refinement of coordinates and anisotropic displacement parameters. Where possible, hydrogen atoms were located from difference electron density maps, and both positions and isotropic temperature factors were refined. The remaining hydrogen atoms were placed in calculated positions and assumed to ride on the attached atoms. When advised by SHELXL,<sup>262</sup> further extinction corrections were introduced with least squares refinement of all the determined atomic co-ordinates, the anisotropic thermal parameters for non-hydrogen atoms and the isotropic temperature factors for the hydrogen atoms was conducted. Thermal ellipsoids have been drawn with ORTEP<sup>263</sup> and represent 50% probability.

Every two hours, three intensity monitor reflections were measured and for every 200 experimental reflections, three orientation monitor reflections were located. Estimated standard deviations were based on counting statistics and an allowance of 0.02 x (net intensity)<sup>2</sup> made for the minimum expected instability.

### **6.3.2 6-Chloro-9-(carboxymethyl)purine ethyl ester 125**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $22.8 \leq \theta \leq 29.1^\circ$ . Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $1.00 + 0.14 \tan \theta$  and a  $\omega$  scan speed of 0.4 to 2.1 deg min<sup>-1</sup>. A total of 3896 reflections were collected in the range of  $0 \leq h \leq 7$ ,  $-12 \leq k \leq 12$ ,  $-22 \leq l \leq 22$  and were merged to give 3826 unique reflections.

### **6.3.3 6-Chloro-7-(carboxymethyl)purine ethyl ester 126**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $22.1 \leq \theta \leq 40.7^\circ$ . Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $0.90 + 0.14 \tan \theta$  and a  $\omega$  scan speed of 0.9 to 4.1 deg min<sup>-1</sup>. A total of 3928 reflections were collected in the range of  $0 \leq h \leq 7$ ,  $-11 \leq k \leq 11$ ,  $-16 \leq l \leq 16$  and were merged to give 1898 unique reflections.

### **6.3.4 2,6-Dichloro-9-(carboxymethyl)purine ethyl ester 76**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $9.7 \leq \theta \leq 14.2^\circ$ . Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $1.00 + 0.35 \tan \theta$  and a  $\omega$  scan speed of 0.5 to 2.7 deg min<sup>-1</sup>. A total of 2524 reflections were collected in the range of  $-1 \leq h \leq 9$ ,  $0 \leq k \leq 14$ ,  $-30 \leq l \leq 30$  and were merged to give 1462 unique reflections.

### **6.3.5 2,6-Diamino-9-(carboxymethyl)purine ethyl ester 80**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $22.8 \leq \theta \leq 29.1^\circ$ . Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $1.20 + 0.14 \tan \theta$  and a  $\omega$  scan speed of 0.8 to 3.3 deg min<sup>-1</sup>. A total of 4966 reflections were collected in the range of  $-1 \leq h \leq 10$ ,  $-13 \leq k \leq 13$ ,  $-14 \leq l \leq 14$  and were merged to give 4126 unique reflections.

### **6.3.6 6-Amino-2-methoxy-(carboxymethyl)purine methyl ester 92**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $23.5 \leq \theta \leq 37.9^\circ$ . Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $0.90 + 0.14 \tan \theta$  and a  $\omega$  scan speed of 0.30 to 2.7 deg min<sup>-1</sup>. A total of 3786 reflections were collected in the range of  $-2 \leq h \leq 13$ ,  $0 \leq k \leq 9$ ,  $-15 \leq l \leq 15$  and were merged to give 1947 unique reflections.

### **6.3.7 2,6-Diazido-9-(carboxymethyl)purine methyl ester 84**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $22.4 \leq \theta \leq 44.3^\circ$ . Intensity data were collected by with  $\omega$  scan range of  $1.50 + 0.14 \tan \theta$  and a  $\omega$  scan speed of 0.6 to 1.8 deg min<sup>-1</sup>. A total of 4316 reflections were collected in the range of  $-15 \leq h \leq 15$ ,  $0 \leq k \leq 14$ ,  $-9 \leq l \leq 9$  and were

merged to give 2173 unique reflections.

#### **6.3.8 Diethyl 3,7-hypoxanthyl diacetate 124**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $22.7 \leq \theta \leq 29.1$  °. Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $0.90 + 0.35 \tan \theta$  and a  $\omega$  scan speed of 0.90 to 4.1 deg min<sup>-1</sup>. A total of 7043 reflections were collected in the range of  $0 \leq h \leq 10$ ,  $0 \leq k \leq 12$ ,  $-12 \leq l \leq 12$  and were merged to give 3093 unique reflections.

#### **6.3.9 (N-Boc-amino)acetonitrile 17**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $24.0 \leq \theta \leq 36.7$  °. Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $0.80 + 0.14 \tan \theta$  and a  $\omega$  scan speed of 0.9 to 4.1 deg min<sup>-1</sup>. A total of 1862 reflections were collected in the range of  $-1 \leq h \leq 14$ ,  $-16 \leq k \leq 16$ ,  $-1 \leq l \leq 25$  and were merged to give 1658 unique reflections.

#### **6.3.10 1-Hydroxyiminophenylcarbamate 20**

Unit cell dimensions were obtained from least squares analysis of setting angles of 25 unique reflections with  $5.62 \leq \theta \leq 67.8$  °. Intensity data were collected by  $\omega$ - $2\theta$  scans with  $\omega$  scan range of  $0.90 + 0.14 \tan \theta$  and a  $\omega$  scan speed of 0.7 to 3.3 deg min<sup>-1</sup>. A total of 2778 reflections were collected in the range of  $-9 \leq h \leq 10$ ,  $-13 \leq k \leq 13$ ,  $-10 \leq l \leq 0$  and were merged to give 1333 unique reflections.

## REFERENCES

1. Haig, D.; Kreig, A. M.; Stein, C. A. *Antisense and Nucleic Acid Drug Development* **1997**, *7*, 1-18.
2. Hélène, C. *European Journal of Cancer Research* **1991**, 1464-1471.
3. Diasio, R. B.; Zhang, R. *Antisense and Nucleic Acid Drug Development* **1997**, *7*, 239-243.
4. Egholm, M.; Nielsen, P. E.; Buchardt, O.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *114*, 9677-9678.
5. Egholm, M.; Behrens, C. *J. Chem. Soc. Chem. Commun.* **1993**, 800-801.
6. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Frier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993**, *365*, 566-568.
7. Egholm, M.; Nielsen, P. E.; Buchardt, O.; Berg, R. H. *Nucleic Acids Research* **1994** *19*, 145-148.
8. Dueholm, K. L. *New J. Chem.* **1997**, *21*, 19-31.
9. Lagriffoule, P.; Wittung, P.; Eriksson, M.; Jensen, K. K.; Norden, B.; Buchardt, O.; Nielsen, P. E. *Chem. Eur. J.* **1997**, *3*, 912-919.
10. Lomakin, A.; Frank-Kamenskii, M. D. *J. Mol. Biol.* **1998**, *276*, 57-70.
11. Noble, S. A.; Bonham, M. A.; Bisi, J. E.; Bruckenstein, D. A.; Brown, P. H.; Brown, S. C.; Cadilla, R. G. *Drug Development Research* **1995**, *34*, 184-195.
12. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, F.; Bonham, M. A.; Au, K. G.; Carter, S. G.; Bruckenstein, D. A.; Boyd, A. L.; Noble, S. A.; Babiss, L. E. *Science* **1991**, *254*, 1497-1499.
13. Hyrup, B.; Nielsen, P. E. *Bioorg. Med. Chem.* **1996**, *4*, 5-23.
14. Caruthers, M. H. *Acc. of Chem. Res.* **1991**, *24*, 278-284.
15. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. In 'Antisense Research and Applications' eds. Crooke, S. T.; Lebleu, B., CRC Press, **1992**, 363-372.
16. Jones, A. S. *Int. J. Biol. Macromolecules* **1979**, *1*, 194-207.
17. Putnam, D. A. *Am. J. Health-Syst. Pharm.* **1996**, *53*, 151-161.
18. Inouye, M. *Gene* **1988**, *72*, 25-34.
19. Nielsen, P.E.; Haaima, G. *Chem. Soc. Rev.* **1997**, 73-78.
20. Mrkisch, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.
21. Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *114*, 1895-1897.
22. Patel, D. J. *Nature* **1993**, *365*, 490-492.
23. Cherny, D. Y.; Belotserkovskii, B. P.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchardt, O.; Berg, R.; Nielsen, P. E. *Proc. Natl. Acad. Sci.* **1993**, *90*, 1667-1670.

24. Wittung, P.; Nielsen, P. E. *Nature* **1994**, *368*, 561-562.
25. Carlsson, C.; Jonsson, M.; Norden, B. *Nature* **1996**, *380*, 207.
26. Nielsen, P. E.; Egholm, M.; Buchardt, O. *Bioconjugate Chem.* **1994**, *5*, 3-7.
27. Nielsen, P. E. *Origins of Life and Evolution of the Biosphere* **1993**, *23*, 323-327.
28. Nielsen, P. E. *Bioconjugate Chem.* **1991**, *2*, 1-13.
29. Orum, H.; Nielsen, P. E.; Jorgensen, M.; Larsson, C.; Stanley, C.; Koch, T. *Biotechniques* **1995**, *19*, 472-480.
30. Boffa, L. C.; Carpaneto, E. M.; Allfrey, V. G. *Proc. Natl. Acad. Sci.* **1995**, *92*, 1901-1905.
31. Srinivasan, A. R.; Olson, W. K. *J. Am. Chem. Soc.* **1998**, *120*, 492-499.
32. Sen, S.; Nilsson, L. *J. Am. Chem. Soc.* **1998**, *120*, 4, 619-631.
33. Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.
34. Hyrup, B.; Egholm, M.; Nielsen, P. E.; Wittung, P.; Norden, B.; Buchardt, O. *J. Am. Chem. Soc.* **1994**, *116*, 7964-7970.
35. Griffith, M.; Risen, L. *J. Am. Chem. Soc.* **1995**, *117*, 5016-5022.
36. Srinivasan, A. R.; Olson, W. K. *J. Am. Chem. Soc.* **1998**, *120*, 1192-1197.
37. Leijon, M.; Graslund, A.; Nielsen, P. E.; Buchardt, O.; Norden, B.; Kristensen, S. M.; Eriksson, M. *Biochemistry* **1994**, *33*, 9820-9825.
38. Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Norden, B. *Nature* **1994**, *368*, 561-563.
39. Wittung, P.; Nielsen, P. E.; Norden, B. *Nucleosides and Nucleotides* **1997**, *16*, 599-602.
40. Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. *Abstract from Ninth Conversation in Biomolecular Stereodynamics* **1995**, A020, 20-24.
41. Mollegard, N. E.; Buchardt, O.; Egholm, M.; Nielsen, P. E. *Proc. Natl. Acad. Sci.* **1994**, *91*, 3892-3895.
42. Corey, D. R.; Iyer, M.; Norton, J.; Varnum, E.; Waggenspack, J. H. *FASEB* **1995**, A1391.
43. Urata, H.; Ueda, Y.; Suhara, H.; Nishioka, E.; Akagi, M. *J. Am. Chem. Soc.* **1993**, *115*, 9852-9853.
44. Noble, S. A.; Bonham, M. A.; Bisi, J. E.; Bruckenstein, D. A.; Brown, P. H.; Brown, S. C.; Cadilla, R.; Gaul, M. D.; Hanvey, J. C.; Hassman, C. F.; Josey, J. A.; Luzzio, M. J.; Myers, P. M.; Pipe, A. J.; Ricca, D. J.; Su, C. W.; Stevenson, C. L.; Thomson, S. A.; Wiethe, R. W.; Babiss, L. E. *Drug Development Research* **1995**, *34*, 184-195.
45. Sen, S.; Nilsson, L. *J. Am. Chem. Soc.* **1998**, *120*, 619-631.
46. Demidov, V.; Potaman, V. N.; Frank-Kamanetskii, M. D.; Egholm, M.; Buchardt, O.; Sonichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994**, *48*, 1310-1313.
47. Knudsen, H.; Nielsen, P. E. *Anti-Cancer Drug Design* **1997**, *8*, 113-118.



48. Danheiser, S. L. *Gen. Eng. News*, **1991**, *113*, 21.
49. Wittung, P.; Nielsen, P. E.; Norden, B. *Biochemistry* **1997**, *36*, 7973-7979.
50. Peymann, A.; Uhlmann, E.; Wagner, K.; Augustin, S.; Breipohl, G.; Will, D. W.; Shafer, A.; Wallmeier, H. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2636-2638.
51. Cheng, Y. K.; Pettitt, B. M.; Singleton, S.; Dervan, P. *J. Am. Chem. Soc.* **1992**, *114*, 4465-4474.
52. Kagawa, T. F.; Howell, M. L.; Tseng, K.; Ho, P. S. *Nucleic Acids Research* **1993**, *21*, 5978-5986.
53. Watson, J.; Crick, F. H. *Nature* **1953**, *171*, 737-738.
54. Gee, J. E.; Miller, D. M. *Am. J. Med. Sci.* **1992**, *304*, 366-373.
55. Horne, D.; Dervan, P. *J. Am. Chem. Soc.* **1990**, *112*, 2435-2437.
56. Rentzeperis, D.; Holder, J.; Marky, L. A.; Garcia, A. E. *Biophysical Journal* **1996**, *70*, MPOS12.
57. Searle, M.S.; Williams, D. H. *Nucleic Acids Research* **1993**, *21*, 2051-2056.
58. Brown, T.; Hunter, W. N.; Leonard, G. A. *Chem. Br.* **1993**, *29*, 484-488.
59. Trapene, T.; Op Ts'O, P. *J. Am. Chem. Soc.* **1994**, *116*, 10437-10449.
60. Moser, H. E.; Dervan, P. *Science* **1987**, 645-647.
61. Faucon, B.; Mergny, J. L.; Hélène, C. *Nucleic Acids Research* **1996**, *24*, 3181-3188.
62. Maher, L. J.; Dervan, P.; Wold, B. J. *Biochemistry* **1990**, *29*, 8820-8826.
63. Singleton, S.; Dervan, P. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965.
64. Singleton, S.; Dervan, P. *Biochemistry* **1992**, *31*, 10995-11003.
65. Gaffney, B. L.; Kung, P. P.; Wang, C.; Jones, R. A. *J. Am. Chem. Soc.* **1995**, *117*, 12281-12283.
66. Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360-1363.
67. Hudson, R. H. E.; Uddin, A. H.; Damha, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 12470-12477.
68. Priestley, E.; Dervan, P. *J. Am. Chem. Soc.* **1995**, *117*, 4761-476
69. Strobel, S.; Moser, H.; Dervan, P. *J. Am. Chem. Soc.* **1988**, *110*, 7927-7929.
70. Povsic, T. J.; Dervan, P. *J. Am. Chem. Soc.* **1989**, *111*, 3059-3061.
71. Zamecknik, P. *Antisense and Nucleic Acid Drug Development* **1997**, *7*, 199-202.
72. Julianao, R. L.; Akhtar, S. *Antisense Research and Development* **1992**, *2*, 165-176.
73. Goodchild, J. *Bioconjugate Chem.* **1990**, *1*, 165-187.
74. Cook, P. *Anti-Cancer Drug Design* **1993**, *6*, 585-607.
75. Carpino, L. A. *Accounts of Chemical Research* **1987**, *20*, 401-407.
76. Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 543-583.
77. Gigg, J.; Gigg, R. *J. Chem. Soc., Perkin Trans. 1* **1987**, 2411-2415.
78. Crooke, S. T. *Curr. Opin. Struc. Biol.* **1992**, *3*, 656-661.

79. Hanvey, J. C.; Peffer, N. J. *Science* **1992**, *258*, 1481-1485.
80. Schwartz, A. W. *Curr. Opin. Struc. Biol.* **1994**, *4*, 758-760.
81. Hanvey, J. C.; Peffer, N. J. *Science* **1992**, *258*, 1497-1515.
82. Hélène, C.; Toulme, J. J. *Biochimica et Biophysica Acta* **1990**, *1049*, 99-125.
83. Rose, D. J. *Anal. Chem.* **1993**, *65*, 3545-3549.
84. Hélène, C. *Anti-Cancer Drug Design* **1991**, *6*, 569-584.
85. Nielsen, P. E. *Perspectives in Drug Discovery and Design* **1996**, *4*, 76-84.
86. Nielsen, P. E. *Int. Antiviral News* **1993**, *1*, 37-39.
87. Maher, L. K. *Cancer Investigation* **1996**, *14*, 66-82.
88. Hamilton, S. E.; Iyer, M.; Norton, J. C.; Corey, D. R. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2897-2900.
89. Nielsen, P. E.; Egholm, M.; Buchardt, O. *Gene* **1994**, *149*, 139-145.
90. Kamenetskii, M. D. F.; Mirkin, S. M. *Int. J. Peptide Protein Res.* **1995**, *64*, 64-95.
91. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Nucleic Acids Research* **1993**, *21*, 197-200.
92. Good, L.; Nielsen, P. E. *Nature Biotech.* **1998**, *16*, 355-358.
93. Gambacorti-Passerini, C.; Mologni, L.; Bertazzoli, C.; Coutre, P.; Marchesi, E.; Grignani, F.; Nielsen, P. E. *Blood* **1996**, *88*, 1411-1417.
94. Larsen, H. J.; Nielsen, P. E. *Nucleic Acids Research* **1996**, *24*, 458-463.
95. Gambacorti-Passerini, C.; Mologni, L.; Grignani, F.; Morelli, D.; Marchesi, E.; Dermime, S.; Parmiani, G. *Blood* **1994**, *84*, A2406.
96. Sharma H. S.; Narayanan, R. *Bioessays* **1995**, *17*, 1055-1063.
97. Sanghvi, Y. S.; Hoke, G. D. *Nucleic Acids Research* **1993**, *21*, 3197-3203.
98. Wang, J.; Paleck, E.; Nielsen, P. E.; Rivas, G.; Cai, X.; Shiraishi, H.; Dontha, N.; Luo, D.; Farias, P. A. M. *J. Am. Chem. Soc.* **1996**, *118*, 7667-7670.
99. Faucon, B.; Mergny, J. L.; Hélène, C. *Nucleic Acids Research* **1996**, *24*, 3181-3188.
100. Milligan, J. F.; Matteucci, M. D.; Martin, J. C. *J. Med. Chem.* **1993**, *36*, 1923-1937.
101. Vickers, T. A.; Griffith, M. C.; Ramasamy, K.; Risen, L. M.; Freier, S. M. *Nucleic Acids Research* **1995**, *23*, 3003-3008.
102. Kim, S. K.; Nielsen, P. E. Egholm, M.; Buchardt, O.; Berg, R. H.; Norden, B. *J. Am. Chem. Soc.* **1993**, *115*, 6478-6481.
103. Sonichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994**, *48*, 1310-1313.
104. Corey, D. R.; Iyer, M.; Norton, J.; Waggenpack, J. J. *Biol. Chem.* **1995**, *2*, 14712-14717.
105. Miller, S. L. *Nature Struct. Biol.* **1996**, *3*, 167-169.
106. Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P.E.; Norden, B.; Graslund, A. *J. Am. Chem. Soc.* **1996**, *118*, 5544-5552.

107. Krasil'nikova, M. M.; Veselkov, A. G. *J. Mol. Biol.* **1996**, *30*, 226-230.
108. Sletten, J.; Jensen, L. H. *Acta Cryst.* **1969**, *B25*, 1608-1614.
109. Ressler, C. *J. Am. Chem. Soc.* **1956**, *78*, 5056-5059.
110. Wittung, P.; Nielsen, P.; Norden, B. *J. Am. Chem. Soc.* **1996**, *118*, 7049-7054.
111. Srinivason, A. R. *Biophysical Journal* **1996**, *70*, MPOS21.
112. Griffith, M.; Risen, L. M. *J. Am. Chem. Soc.* **1995**, *117*, 5016-5022.
113. Griffith, M. C.; Risen, L. M.; Grieg, M. J.; Leenik, C. A.; Sprankle, K. G.; Griffey, R. H.; Kiely, J. S.; Frier, S. M. *J. Am. Chem. Soc.* **1995**, *117*, 831-832.
114. Nielsen, P. E.; Egholm, M.; Buchardt, O. *J. Mol. Recog.* **1994**, *7*, 165-170.
115. Nielsen, P. E.; Christensen, L. *J. Am. Chem. Soc.* **1996**, *118*, 2287-2288.
116. Pugmire, R. J.; Grant, D. M.; Townsend, L. B.; Robins, R. K. *J. Am. Chem. Soc.* **1973**, *117*, 2791-2796.
117. Gee, J. E.; Robbins, I. Van der Laan, A. C.; van Boom, J. H.; Colombier, C.; Leng, M.; Raible, A. M.; Nelson, J. S.; Lebleu, B. *Antisense and Nucleic Acid Drug Development* **1998**, *8*, 103-111.
118. Peffer, N. J.; Hanvey, J. C.; Bisi, J. E.; Thomson, S. A.; Hassman, C. F.; Noble, S. A.; Babiss, L. E. *Proc. Natl. Acad. Sci.* **1993**, *90*, 10648-10652.
119. Demidov, V.; Yavnilovich, M. V.; Frank-Kamenetskii, M. D. *Biophysical Journal* **1997**, *72*, 2763-2769.
120. Demidov, V.; Potaman, V. N.; Frank-Kamanetskii, M. D.; Egholm, M.; Buchardt, O.; Stetsenko, D. A.; Veselovskaya, S. V.; Lubyako, E. N.; Potapov, V. K.; Azhikina, T. L.; Petersen, K. H.; Jensen, D. K.; Egholm, M.; Nielsen, P. E.; Buchardt, O. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 1119-1124.
121. Demers, D. B.; Curry, E. T.; Sozer, A. C. *Am. J. Hum. Gen.* **1995**, *57*, 120-124.
122. Orum, H.; Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O.; Stanley, C. *Nucleic Acids Research* **1993**, *21*, 5332-5336.
123. Thiede, C.; Bayerdorffer, E.; Blasczyk, R.; Wittig, B.; Neubauer, A. *Nucleic Acids Research* **1996**, *224*, 983-984.
124. Basu, S.; Wickstrom, E. *FASEB* **1996**, *10*, 2574.
125. Ariza, X.; Bou, V.; Vilarrasa, J. *J. Am. Chem. Soc.* **1995**, *117*, 3665-3673.
126. Wittung, P.; J. Kajanus, J.; Edwards, K.; Nielsen, P.; Norden, B.; Malmstrom, B. G. *FEBS Letters* **1995**, *375*, 317-320.
127. Zon, G. *Pharmaceut. Res.* **1988**, *5*, 539-548.
128. Eldrup, A. B.; Dal, Otto.; Nielsen, P. E. *J. Am. Chem. Soc.* **1997**, *119*, 11116-11117.
129. Gray, G. D.; Basu, S.; Wickstrom, E. *Biochem. Pharmacol.* **1997**, *53*, 1465-1476.

130. Simmons, C. G.; Mayfield, L. D.; Shay, J. W.; Corey, D. R. *FASEB Journal* **1997**, *11*, A1154, 1732.
131. Tyler, B. M.; Richelson, E.; McCormick, D. J.; Fauq, A.; Hoshall, C. V.; Douglas, C. L.; Groshan, K.; Cusack, B.; Lacy, B. *Biol. Psychiatry* **1998**, *43*, 121.
132. Pardridge, W. M.; Boado, R. J.; Kang, Y. S. *Proc. Natl. Acad. Sci.* **1995**, *92*, 5592-5596.
133. Boado, R. *Adv. Drug. Deliv. Rev.* **1995**, *15*, 73-107.
134. Norton, J.; Waggenspack, J. H.; Varnum, E.; Corey, D. R. *Bioorg. Med. Chem. Lett.* **1995**, *3*, 437-445.
135. Julianao, R. L.; Akhtar, S. *Nucleic Acids Research* **1991**, *19*, 5551-5559.
136. Simmons, C. G.; Pitts, A. E.; Mayfield, L. D.; Shay, J. W.; Core, D. R. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 3001-3006.
137. Mardiriossian, G.; Lei, K.; Rusckowski, M.; Chang, F.; Qu, T.; Egholm, M.; Hnatowich, D. J. *Journal of Nuclear Medicine* **1997**, *38*, 907-913.
138. Weiler, J.; Gausepohl, H.; Hauser, N.; Jensen, O. N.; Hoheisel, J. D. *Nucleic Acids Research* **1997**, *25*, 2792-2799.
139. Basu, S.; Wickstrom, E. *Bioconjugate Chem.* **1997**, *8*, 481-488.
140. Tarkoy, M.; Bolli, M.; Leumann, C. *Helvetica Chimica Acta* **1994**, *77*, 716-744.
141. Varma, R. S. *Synlett* **1993**, 621-637.
142. Kamenetskii, M. *Nature* **1991**, *354*, 505-506.
143. Smulevitch, S. V.; Simmons, C. G.; Norton, J. C.; Wise, T. W.; Corey, D. R. *Nature Biotech.* **1996**, *14*, 1700-1704.
144. Howarth, N. M.; Wakelin, L. P. G. *J. Org. Chem.* **1997**, *62*, 5441-5450.
145. Wittung, P.; Nielsen, P. E.; Norden, B. *Nucleosides and Nucleotides*, **1997**, *16*, 599-602.
146. Efimov, V. A.; Choob, M. V.; Buryakova, A. A.; Kalinkina, A. L.; Chakmakhcheva, O. G. *Nucleic Acids Research* **1998**, *26*, 566-575.
147. Stec, W. J.; Zon, G.; Egan, W.; Stec, B. *J. Am. Chem. Soc.* **1984**, *106*, 6077-6079.
148. Mesmaeker, A. D.; Almann, K. H.; Waldner, A.; Wendenborn, S. *Curr. Opin. Struc. Biol.* **1995**, *5*, 343-355.
149. Law, S. M.; Eritja, R.; Goodman, M. F.; Breslauer, K. J. *Biochemistry* **1996**, *35*, 12329-12337.
150. Dean, N. M.; Griffey, R. H. *Antisense and Nucleic Acid Drug Development* **1997**, *7*, 229-223.
151. Saxinger, C.; Ziang, R. *Proc. Natl. Acad. Sci.* **1997**, *94*, 2620-2625.
152. Plu, G. E.; Pilch, D. S. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 319-350.

153. Mesmaeker, A. D.; Waldner, A.; Lebrer-ton, J.; Hoffmann, P.; Fritsch, V.; Wolf, R. M.; Frier, S. M. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 227-231.
154. van der Laan, A. C. *Tetrahedron Lett.* **1996**, *37*, 7857-7860.
155. Goodnow, R. A.; Tam, D.; Pruess, D. L.; McComas, W. W. *Tetrahedron Lett.* **1997**, *38*, 3199-3202.
156. Schimdt, J. G.; Nielsen P. E.; Orgel, L. E. *Anal. Biochem.* **1996**, *235*, 239-241.
157. Dempcy, R. O.; Browne, K. A.; Bruice, T. *J. Am. Chem. Soc.* **1995**, *117*, 6140-6141.
158. De Mesmaeker, A.; Waldner, A.; Wendenborn, S.; Wolf, R. M. *Pure and Appl. Chem.* **1997**, *69*, 437-440.
159. Egli, M. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1894-1909.
160. Meier, C.; Engels, J. W. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1008-1010.
161. Hyrup, B.; Egholm, M.; Rolland, M.; Nielsen, P. E.; Berg, R. H.; Buchardt, O. *J. Chem. Soc. Chem. Commun.* **1993**, 518-519.
162. Juricova, K.; Smirckova, S.; Holy, A. *Coll. Czech. Chem. Commun.* **1995**, *60*, 237-250.
163. Gangamani, B. P.; Kumar, V. A.; Ganesh, K. N. *Tetrahedron* **1996**, *52*, 47, 15017-15030.
164. Sanghvi, Y. S. *Antisense Research and Applications*, **1993**, 273-288.
165. Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322-8323.
166. Fujii, M.; Yoshida, K.; Hidaka, J. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 637-640.
167. Nielsen, P. E. *Origins of Life and Evolution of the Biosphere* **1993**, *23*, 323-327.
168. Wittung, P.; Eriksson, M.; Lyng, R.; Nielsen, P. E.; Norden, B. *J. Am. Chem. Soc.* **1995**, *117*, 10167-10173.
169. Koch, T.; Naesby, M.; Wittung, P.; Jorgensen, M.; Larsson, C.; Buchardt, O.; Stanley, C. J.; Norden, B.; Nielsen, P. E.; Orum, H. *Tetrahedron Lett.* **1995**, *36*, 6933-6936.
170. Diederichsen, U. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 445-448.
171. Jones, A. S. *Int. J. Biolog. Macromolecules* **1979**, *1*, 194-207.
172. Uhlmann, E.; Will, D. W.; Breipohl, G.; Langner, D.; Rytte, A. *Angew. Chem. Int. Ed. Eng.* **1996**, *335*, 2632-2635.
173. Good, L.; Nielsen, P. E. *Antisense and Nucleic Acid Drug Development* **1997**, *7*, 431-437.
174. Nielsen, P. E. *Methods in Enzymology* **1996**, *267*, 426-433.
175. Demidov, V.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchardt, O.; Nielsen, P. E. *Nucleic Acids Research* **1993**, *21*, 2103-2107.
176. Demers, D. B.; Curry, E. T.; Egholm, M.; Sozer, A. C. *Nucleic Acids Research* **1995**, *23*, 3050-3055.
177. Behn, M.; Schuermann, M. *Thromb. Haemost.* **1998**, *79*, 773-777.

178. Nielsen, P. E. *Nucleic Acids Research* **1996**, *14*, 580-584.
179. Hamilton, S. E.; Pitts, A. E.; Katipally, R. R.; Jia, X.; Rutter, J. P. Davies, B. A.; Shay, J. W.; Wright, W. E.; Corey, D. R. *Biochemistry* **1997**, *36*, 11873-11880.
180. Zhang, L. G.; Min, J. M.; Zhang, L. B. *Progress in Biochemistry and Biophysics* **1996**, *23*, 209-213.
181. Nielsen, P. E. *Pure and Appl. Chem.* **1998**, *70*, 105-110.
182. Good, L.; Seeger, V. C. *Biotechniques* **1997**, *23*, 512.
183. Egholm, M.; Christensen, L.; Dueholm, K. L.; Buchardt, O.; Coull, J.; Nielsen, P. E. *Nucleic Acids Research* **1995**, *23*, 217-222.
184. Cheng, Y. K.; Pettit, B. M. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 225-231.
185. Strobel, S. A.; Cech, T. R.; Usman, N.; Beigelman, L. *Biochemistry* **1993**, *33*, 13824-13835.
186. Ullah, S.; Xu, Y. Z.; Day, R. S. *Biochemistry* **1995**, *34*, 7438-7442.
187. Orgel, L. E. *Nature* **1990**, *343*, 18-19.
188. Koh, J. S.; Dervan, P. *J. Am. Chem. Soc.* **1992**, *114*, 1470-1478.
189. Radhakrishnan, I.; Patel, D. J.; Priestley, E. S.; Nash, H. M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11228-11134.
190. Cherny, D. I.; Fourcade, A.; Svinarchuk, F.; Nielsen, P. E.; Malvy, C.; Delain, E. *Biophysical Journal* **1998**, *74*, 1015-1023.
191. Maher, L. J.; Dervan, P.; Wold, B. J. *Biochemistry* **1990**, *29*, 8820-8826.
192. Jones, R. J.; Lin, K. Y.; Milligan, J. F.; Wadwani, S.; Matteucci, M. D. *Tetrahedron*, **1993**, *58*, 2983-2991.
193. Yamagata, Y.; Tomita, K. *Acta Cryst.* **1988**, *C48*, 318-320.
194. Giovannangeli, C.; Rouges, M.; Thoung, N. T.; Hélène, C.; Gavestier, T. *Proc. Natl. Acad. Sci.* **1992**, *89*, 8631-8635.
195. Bailly, C.; Waring, M. J. *Nucleic Acids Research* **1998**, *26*, 4309-4314.
196. Chollet, A.; Kawashima, E. *Nucleic Acids Research* **1988**, *16*, 305-317.
197. Haaima, G.; Hansen, H. F.; Christensen, L.; Dahl, O.; Nielsen, P. E. *Nucleic Acids Research* **1997**, *25*, 4639-4643.
198. Jayarman, K.; Durland, R. H.; Rao, T. S.; Revankar, G. R.; Bodepudi, V.; Chaudary, N.; Guy-Caffey, J. *Nucleosides and Nucleotides* **1995**, *14*, 951-955.
199. Dagneaux, C.; Gousset, A. K.; Shchyolkina, A. K.; Ouali, M.; Letellier, R.; Liquier, J.; Florentiev, V. L.; Taillandier, E. *Nucleic Acids Research* **1996**, *24*, 4506-4512.
200. Colapietro, M.; Domenicano, A. *Acta Cryst.* **1977**, *B33*, 2240-2243.
201. Plum, G. E.; Pork, Y. W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. *Proc. Natl. Acad. Sci.* **1990**, *87*, 9436-9440.
202. Krasil'nikova, M. M.; Veselkov, A.G. *J. Mol. Biol.* **1996**, *30*, 226-230.

203. Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapene, T. L. *Biochemistry* **1992**, *31*, 6788-6792.
204. Zimmerman, S. C.; Schmitt, P. J. *Am. Chem. Soc.* **1995**, *117*, 10769-10770.
205. Hunziker, J.; Priestley, E. S.; Dervan, P.; Brunar, H. *J. Am. Chem. Soc.* **1995**, *117*, 2661-2662.
206. Krosigh, U. V.; Benner, S. A. *J. Am. Chem. Soc.* **1995**, *117*, 5361-5362.
207. Vkraczyk, S. H.; Milligan, J. F.; Wadini, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci.* **1992**, *89*, 3761-3764.
208. Lee, J. S. Woodsworth, M. L.; Latimer, L. J. P.; Morgan, A. R. *Nucleic Acids Research* **1984**, *12*, 60603-60607.
209. Xiang, G.; Bogacki, R.; McLaughlin, L. W. *Nucleic Acids Research* **1996**, *24*, 1963-1971.
210. Ono, A.; Ts'O, P. O.; Kan, L. *J. Am. Chem. Soc.* **1991**, *113*, 4032-4033.
211. Staubli, A. B.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *22*, 2637-2643.
212. Roberts C.; Brandru, R. *Tetrahedron Lett.* **1995**, *36*, 3601-3604.
213. Han, H.; Dervan, P. *Nucleic Acids Research* **1994**, *22*, 2837-2844.
214. Milligan, J. F.; Krawczyk, S. H.; Wadwani, S.; Matteuci, M. D. *Nucleic Acids Research* **1993**, *21*, 327-333.
215. Loakes, D.; Hill, F.; Linde, S. *Nucleosides and Nucleotides* **1995**, *14*, 1001-1003.
216. Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y., Matsubara, K. *Gene* **1985**, *38*, 271-274.
217. Sood, G.; Schwalbe, C. H.; Fraser, W. *Acta Cryst.* **1997**, *C53*, 959-961.
218. Dueholm, K.; Egholm, M.; Behrens, C.; Christensen, L.; Hansen, H. F.; Vulpius, T.; Petersen, K. H.; Berg, R. H.; Nielsen, P. E.; Buchardt, O. *J. Org. Chem.* **1994**, *59*, 5767-5773.
219. Meltzer, P. C.; Liang, A. Y.; Maatsudaira, P. *J. Org. Chem.* **1995**, *60*, 4305-4308.
220. Egholm, M.; Koch, T.; Hansen, H. F.; Andersen, P.; Larsen, T.; Batz, H. G.; Otteson, K.; Orum, H. *J. Peptide Res.* **1997**, *49*, 80-88.
221. Breipohl, G.; Will, D. W.; Peyman, A.; Uhlmann, E. *Tetrahedron* **1997**, *53*, 14671-14686.
222. Thomson, S. A.; Josey, J. A.; Cadilla, R.; Gaul, M. D.; Hasman, C. F.; Luzzio, M. J.; Pipe, A. J.; Reed, K. L.; Ricca, D. J.; Wiethe, R. W.; Noble, S. A. *Tetrahedron* **1995**, *51*, 6179-6194.
223. Breipohl, G.; Langner, K. D.; O'Malley, G.; Uhlmann, E. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 665-670.
224. Christensen, L.; Fitzpatrick, R.; Gildea, B.; Petersen, K. H.; Hansen, H. F.; Koch, T.; Egholm, M.; Buchardt, O.; Nielsen, P. E.; Coull, J.; Berg, R. H. *J. Peptide Res.* **1995**, *3*, 175-183.

225. Goodnow, R. A.; Richou, A. R.; Tam, S. *Tetrahedron Lett.* **1997**, *38*, 3195-3198.
226. Will, D. W.; Breipohl, G.; Lagner, D.; Knolle, J.; Uhlmann, E. *Tetrahedron* **1995**, *51*, 12069-12082.
227. Dueholm, K. L.; Egholm, M.; Buchardt, O. *Organic Preparations and Procedures* **1993**, *25*, 457-461.
228. Heimer, E. P. Felix, A. M. Lambor, T. J. Scheidi, F. Meinenhofer, J. *Int. J. Peptide Protein Res.* **1984**, *23*, 203-211.
229. Ravikumar, V. T. *Synth. Comm.* **1994**, *24*, 1767-1772.
230. Semertzidis, M.; Matsoukas, J.; Nastopoulos, V.; Hondrelis, J.; Voliotis, S.; Leban, I. *Acta Cryst.* **1989**, *C45*, 1474-1475.
231. Espenbetov, A. A.; Struchkov, Yu, T.; Rybakova, L. V. *J. Struc. Chem.* **1986**, *27*, 160-170.
232. Chan, D. M. C.; Schwalbe, C. H.; Sood, G.; Fraser, W. *Acta Cryst.* **1995**, *C51*, 2383-2386.
233. Green, G. R.; Grinter, T. J.; Kincey, P. M.; Jarvest, *Tetrahedron* **1990**, *46*, 6903-6914.
234. Montgomery, J. A.; Hewson, K.; Clayton, J. C.; Thomas, H. J. *Acta Cryst.* **1994**, *31*, 2202-2212.
235. Geen, G. R.; Grinter, T. J.; Kincey, P. M.; Jarvest, R. L. *Tetrahedron Lett.* **1990**, *46*, 6903-6014.
236. Kjelhberg, J.; Johansson, N. G. *Tetrahedron* **1986**, *42*, 6541-6544.
237. Dalby, C.; Bleasdale, C.; Clegg, W. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1696-1697.
238. Robins, M. J.; Zou, R.; Guo, Z.; Wnuk, S. *J. Org. Chem.* **1996**, *61*, 9207-9212.
239. Sood, G.; Schwalbe, C. H.; Fraser, W. *Acta Cryst.* **1997**, *C53*, 1624-1626.
240. Flensburg, C. *Acta Cryst.* **1994**, *C50*, 1480-1482.
241. Sood, G.; Schwalbe, C. H.; Fraser, W. *Acta Cryst.* **1997**, *C53*, 608-610.
242. Sood, G.; Schwalbe, C. H.; Fraser, W. *Acta Cryst.* **1998**, *C54*, 659-661.
243. Watkins, B. E.; Rapoport, H. *J. Org. Chem.* **1982**, *47*, 4471-4477.
244. Carpino, L.; El-Faham, A. *J. Org. Chem.* **1995**, *60*, 3561-3564.
245. Pon, R. T.; Yu, S. *Tetrahedron Lett.* **1997**, *38*, 3331-3334.
246. Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, *J. Am. Chem. Soc.* **1990**, *C54*, 1691-1696.
247. Sood, G.; Schwalbe, C. H.; Fraser, W. *Acta Cryst.* **1998**, *C54*, 114-116.
248. Sood, G.; Schwalbe, C. H.; Fraser, W. *Acta Cryst.* **1998**, *C54*, 1316-1318.
249. Sood, G.; Schwalbe, C. H.; Fraser, W. *Acta Cryst.* **1998**, *C54*, 1907-1909.
250. Baranay, G.; Merrifield, R. *J. Am. Chem. Soc.* **1977**, 7363-7365.
251. Rao, T. S.; Revankar, G. R. *J. Heterocyclic Chem.* **1995**, *32*, 1043-1049.



252. Fujii, T.; Saito, T.; Suzuki, T.; Kunugi, M. *Chemical and Pharmaceutical Bulletin* **1994**, *42*, 151-153.
253. Osterman, R. M.; McKittrick, B. A.; Chan, T. M. *Tetrahedron Lett.* **1992**, *33*, 4867-4870.
254. Allen, D. W.; Buckland, D.J.; Nowell, I. W. *J. Chem. Soc. Perkin Trans.*, **1976**, 1610-1612.
255. K. N. Trueblood & J. P. Glusker, 'Crystal Structure Analysis: A Primer', Oxford University Press, London, 1972.
256. International Tables for Crystallography. Volume A. Space Group Symmetry. Theo Hahn Pub. for IUCr by Kluwe Academic Publishers, London, 1996.
257. Brookhaven National Laboratory; Univ. of Birmingham 1986, *DATREDXL*. Univ. of Birmingham, England.
258. Gold, R. O.; Smith, I. University of Edinburgh. Enraf-Nonius 1989, CAD-4 Software. Version 5.0. Enraf-Nonius, Delft, The Netherlands.
259. Main, P.; Germain, G.; Woolfson, M. M. 1984. *MULTAN84*. A Computer Program for the Automatic Solution of Crystal Structures from X-ray Diffraction Data. Univs. of York, England, and Louvain, Belgium.
260. G. H. Stout; G. H. Jensen. 'X-Ray Crystallography; A Practical Guide', Wiley Press, New York, 1989.
261. Main, P.; Fiske, S. J.; Hull, S. E.; Lessinger, L.; Germain, G.; Declercq, J. P.; Woolfson, M. M. 1980, *MULTAN80*. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data. Univs. of York, England, and Louvain, Belgium.
262. Sheldrick, G. M. *SHELXL93*. Programme for the Refinement of Crystal Structures. University of Göttingen, Germany.
263. Johnson, C. K. 1976, *ORTEP II*. Report ORNL-5138. Oak Ridge National Laboratory, Tennessee, USA.
264. Chamers and Crawshaw; A Course in Statistics; Wiley Press, Oxford, 1991.
265. Sheldrick, G. *SHELX*. Programme for Crystal Structure Determination, 1976. University of Göttingen, Germany.
266. Motherwell, W. D. S. A Programme for Plotting Molecular and Crystal Structures, **1972**, University of Cambridge, England.
267. Takeda, T.; Ohashi, Y.; Sasada, Y. *Acta Cryst.* **1974**, *B30*, 825-827.
268. Itai, A.; Yamada, H.; Okamoto, T.; Iitaka, Y. *Acta Cryst.* **1977**, *B33*, 1816-1820.
269. Valle, G.; Piazzogna, G.; Ettore, R. *J. Chem. Soc. Dalton Trans.* **1985**, 1271-1273.
270. Mishnev, A. F.; Bleidelis, J.; Liepin'sh, E. E.; Ramzaeva, N. P.; Goncharova, I. N. *Khim. Geterotsykl. Soedin, SSSR* **1979**, 976-983.

271. Mishnev, A. F.; Bleidelis, J.; Goncharova, I. N.; Ramzaeva, N. P. *Latv. PSR Zinat. Akad. Vestis, Khim. Ser.* **1982**, 241-245.
272. Sternglanz, H.; Bugg, C. E. *Acta Cryst.* **1975**, B31, 2888-2891.
273. Domenicano, A.; Vaciago, A.; Coulson, C. A. *Acta Cryst.* **1975**, B14, 1630-1641.
274. Kartsev, V. G.; Aliev, Z. G.; Voronina, G. N.; Atovmyan, L. O. *Khim. Get. Soedin., SSSR*, **1990**, 575.
275. Schmalte, H. W.; Hanggi, G.; Dubler, E. *Acta Cryst.* **1988**, C44, 732-736.
276. Munns, A. R. I.; Tollin, P. *Acta Cryst.* **1970**, B26, 1101-1113.
277. Thewalt, U.; Bugg, C. E.; Marsh, R. E. *Acta Cryst.* **1970**, B26, 1089-1101.
278. Rosenstein, R. D.; Oberding, M.; Hyde, J. R.; Zubieta, J.; Karlin, K. D.; Seeman, N. C. *Cryst. Struct. Commun.* **1982**, 11, 1507-1513.
279. Sletten, J.; Jensen, L. H. *Acta Cryst.* **1969**, B25, 1608-1614.
280. Benedetti, E.; Blasio, B. di.; Pavone, V.; Pedone, C.; Toniolo, C.; Bonora, G. *M. Biopolymers* **1981**, 20, 1635-1649.
281. Bertolasi, V.; Gilli, G.; Veronese, A. C. *Acta Cryst.* **1982**, B38, 502-506.
282. Derek T. Hurt, In 'Introduction to The Chemistry and Biochemistry of` Pyrimidines, Purines and Pteridines', ed. John Wiley and Sons, 1979, pp 64-86.
283. Still, W. C.; Khan, M.; Mitra, A. *J. Org. Chem.* **1978**, 43, 2393-2925.

## APPENDIX

Crystal data and structure refinement for 1-Hydroxyiminophenylcarbamate **20**

F(000)	304
Data/restraints/parameters	1333 / 0 / 122
Goodness-of-fit on F <sup>2</sup>	1.038
Final R indices [I > 2σ(I)]	R1 = 0.0897, wR2 = 0.2300
R indices (all data)	R1 = 0.0955, wR2 = 0.2366
Largest diff. peak and hole	0.373 and -0.449 e.Å <sup>-3</sup>

A.1. Atomic coordinates ( $\times 10^4$ ) and equivalent isotropic displacement parameters ( $\text{Å}^2 \times 10^3$ ) for **20**. U(eq) is defined as one third of the trace of the orthogonalized U<sub>ij</sub> tensor.

	x	y	z	U(eq)
C(1)	5526 (2)	1133 (2)	807 (2)	40 (1)
C(2)	4481 (2)	2078 (2)	3777 (3)	49 (1)
C(3)	2833 (3)	2161 (2)	2395 (3)	58 (1)
C(4)	2252 (3)	1312 (2)	1079 (3)	61 (1)
C(5)	3292 (3)	368 (2)	1117 (3)	64 (1)
C(6)	4930 (3)	271 (2)	2473 (3)	55 (1)
C(7)	7270 (2)	1034 (1)	5263 (2)	38 (1)
C(8)	7978 (2)	2048 (2)	6449 (2)	45 (1)
N(9)	8542 (2)	2857 (2)	7363 (3)	63 (1)
N(10)	8152 (2)	73 (1)	5446 (2)	46 (1)
O(11)	9758 (2)	70 (1)	6856 (2)	54 (1)

A.2. Bond lengths [ $\text{\AA}$ ] and angles [ $^\circ$ ] for **20**

C(1)-C(2)	1.385 (3)
C(1)-C(6)	1.387 (3)
C(1)-C(7)	1.480 (2)
C(2)-C(3)	1.399 (3)
C(3)-C(4)	1.368 (3)
C(4)-C(5)	1.379 (4)
C(5)-C(6)	1.387 (3)
C(7)-N(10)	1.284 (2)
C(7)-C(8)	1.446 (2)
C(8)-N(9)	1.140 (3)
N(10)-O(11)	1.385 (2)
C(2)-C(1)-C(6)	119.8 (2)
C(2)-C(1)-C(7)	120.3 (2)
C(6)-C(1)-C(7)	119.9 (2)
C(1)-C(2)-C(3)	119.8 (2)
C(4)-C(3)-C(2)	120.2 (2)
C(3)-C(4)-C(5)	119.8 (2)
C(4)-C(5)-C(6)	120.8 (2)
C(5)-C(6)-C(1)	119.5 (2)
N(10)-C(7)-C(8)	120.6 (2)
N(10)-C(7)-C(1)	120.5 (2)
C(8)-C(7)-C(1)	118.9 (2)
N(9)-C(8)-C(7)	178.9 (2)
C(7)-N(10)-O(11)	111.6 (2)

Symmetry transformations used to generate equivalent atoms:

A.3. Anisotropic displacement parameters ( $\text{\AA}^2 \times 10^3$ ) for **20**. The anisotropic displacement factor exponent takes the form:  $-2 \pi^2 [h^2 a^{*2} U_{11} + \dots + 2 h k a^* b^* U_{12}]$

	U11	U22	U33	U23	U13	U12
C(1)	42 (1)	33 (1)	50 (1)	2 (1)	27 (1)	-4 (1)
C(2)	45 (1)	40 (1)	64 (1)	-2 (1)	26 (1)	-1 (1)
C(3)	44 (1)	55 (1)	73 (1)	9 (1)	26 (1)	7 (1)
C(4)	47 (1)	71 (2)	57 (1)	10 (1)	17 (1)	-5 (1)
C(5)	67 (2)	67 (2)	53 (1)	-11 (1)	21 (1)	-11 (1)
C(6)	9 (1)	48 (1)	59 (1)	-9 (1)	28 (1)	-2 (1)
C(7)	40 (1)	30 (1)	52 (1)	1 (1)	28 (1)	-2 (1)
C(8)	39 (1)	37 (1)	60 (1)	-2 (1)	22 (1)	-1 (1)
N(9)	53 (1)	45 (1)	83 (1)	-16 (1)	22 (1)	-3 (1)
N(10)	44 (1)	37 (1)	59 (1)	1 (1)	26 (1)	2 (1)
O(11)	45 (1)	46 (1)	70 (1)	3 (1)	23 (1)	8 (1)

A.4. Hydrogen coordinates ( $\times 10^4$ ) and isotropic displacement parameters ( $\text{\AA}^2 \times 10^3$ ) for **20**

	x	y	z	U(eq)
H(2)	4880 (33)	2650 (26)	4762 (30)	57 (6)
H(3)	2075 (51)	2824 (37)	2454 (43)	105 (11)
H(4)	1201 (46)	1433 (36)	93 (43)	92 (9)
H(5)	2957 (39)	-93 (30)	258 (44)	90 (10)
H(6)	5610 (33)	-354 (31)	2417 (31)	64 (7)
H(11)	10292 (16)	-453 (12)	6937 (28)	105 (12)