

**SEQUENCE SPECIFIC INTERACTIONS,  
INTRACELLULAR DISTRIBUTION AND THE EFFECTS  
ON TRANSCRIPTION OF NOVEL DNA MINOR  
GROOVE-BINDING POLYAMIDES.**

**BY**

**CELIA CAROLINE O'HARE**

**Cancer Research UK Drug-DNA Interactions Research Group,  
Department of Oncology, Royal Free and University College Medical  
School, University College London, London, W1W 7BS.**

Submitted to the University of London for PhD examination

June 2004.

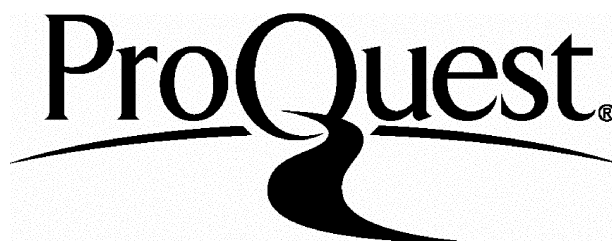
ProQuest Number: U643388

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest U643388

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.  
Microform Edition © ProQuest LLC.

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346



## ABSTRACT

Targeting DNA with small molecules has been a useful therapeutic approach in the treatment of human cancers. Interest in selective disruption of replication and transcription by DNA-interactive ligands has intensified with the wealth of new genetic targets provided by the discovery of key genes associated with genetic diseases such as cancer. This has resulted in the development not only of novel conventional chemotherapy agents but also of molecules with potential to modulate the expression of disease related genes. Minor groove-binding polyamides offer a promising means of disrupting gene expression at predetermined sites. Investigations are presented into the sequence specificity, intracellular uptake and effects on transcription of a number of novel polyamides. A structure-binding activity relationship was determined for a polyamide series comprised of thiazole (T), imidazole (I) or pyrrole (P) heterocycles examining the effects of linker length, heterocycle permutation and N-terminal head group. Linkage with a cross-linked motif increased binding affinity for N-terminal amino or formyl polyamides, with six or seven carbons linkers producing the strongest binding. Optimal selectivity was achieved for the TIP series with either TIP-C7-formyl or TIP-C7-amino, with specificity of the leading thiazole modulated by the adjacent head group. Use of an N-terminal acetyl or hydrogen abrogated binding irrespective of linkage or heterocycle composition. Examination of the binding of formyl-III showed a hundred-fold preference for a single TG mismatch site, relative to a CG match site demonstrating the utility of an I:I pair for TG mismatch recognition. Evaluation of the uptake and distribution of TIP-C7-amino in cells suggested that it was not taken up via passive uptake mechanisms, although it was able to bind DNA when directly introduced into the nucleus. This was irrespective of the cell line examined, although a cell-type dependent anti-proliferative effect was observed at high concentrations. TIP-C7-amino failed to inhibit proliferation in *S.cerevisiae* and caused consistent changes in the transcription of only 0.5% of yeast genome, suggesting inefficient cell uptake. This was in contrast to distamycin, which produced a detectable effect on both the proliferation and transcription of these cells. These studies have implications for the design of linked polyamides as sequence selective agents targeting DNA in cells.

# INDEX

Title page.....	1
Abstract.....	2
Index.....	3
Lists of Figures and Tables.....	7
Abbreviations.....	13
Acknowledgements.....	15

## CHAPTER 1: INTRODUCTION

<b>1.1</b>	<b>Cancer.....</b>	<b>16</b>
<b>1.2</b>	<b>Anticancer strategies.....</b>	<b>17</b>
<b>1.3</b>	<b>DNA as a molecular target for anticancer strategies.....</b>	<b>20</b>
<b>1.4</b>	<b>The structure of DNA.....</b>	<b>24</b>
<b>1.5</b>	<b>Strategies for targeting nucleic acids.....</b>	<b>28</b>
	1.5.1 Antimessage.....	28
	1.5.2 Antigene.....	35
<b>1.6</b>	<b>Targeting DNA using small ligands.....</b>	<b>41</b>
	1.6.1 Covalent modification.....	42
	1.6.2 Non-covalent binding.....	43
<b>1.7</b>	<b>DNA recognition on the minor groove by small ligands.....</b>	<b>44</b>
	1.7.1 Features of minor groove recognition.....	45
	1.7.2 Covalent interactions: chemically reactive minor groove binders.....	47
	1.7.3 Non-covalent interactions: reading the helix in the minor groove.....	55
<b>1.8</b>	<b>The development of polyamides or “lexitropsins”.....</b>	<b>69</b>
	1.8.1 Introduction of a GC recognition element.....	69
	1.8.2 Groove width and 2:1 binding.....	72
	1.8.3 AT degeneracy.....	75
	1.8.4 Stacking and linkage across the groove.....	76
	1.8.5 A hairpin versus cross-linked linkage motif.....	79
	1.8.6 Extending the recognition site.....	82
	1.8.7 Polyamides as molecular vectors.....	84
	1.8.8 Polyamides as molecular tools, antiretrovirals and modulators of gene expression.....	87
<b>1.9</b>	<b>Experimental aims.....</b>	<b>89</b>

## **CHAPTER 2: DNA SEQUENCE RECOGNITION OF WATSON-CRICK BASE PAIRS BY NOVEL THIAZOLE-CONTAINING CROSS-LINKED POLYAMIDES**

<b>2.1</b>	<b>Introduction.....</b>	<b>92</b>
<b>2.2</b>	<b>Materials.....</b>	<b>98</b>
<b>2.3</b>	<b>Methods.....</b>	<b>101</b>
2.3.1	Preparation of a genomic DNA template.....	101
2.3.2	Preparation of a 5'singly end labelled fragment.....	103
2.3.3	DNase I footprinting.....	105
2.3.4	Determination of equilibrium association constants.....	107
<b>2.4</b>	<b>Results.....</b>	<b>110</b>
2.4.1	Sequence specific binding of thiazole-containing polyamides.....	110
2.4.2	Binding affinity and selectivity of the TIP series.....	145
<b>2.5</b>	<b>Discussion.....</b>	<b>150</b>

## **CHAPTER 3: RECOGNITION OF TG MISMATCHES IN THE DNA MINOR GROOVE USING TRI-IMIDAZOLE POLYAMIDES**

<b>3.1</b>	<b>Introduction.....</b>	<b>157</b>
<b>3.2</b>	<b>Materials.....</b>	<b>162</b>
<b>3.3</b>	<b>Methods.....</b>	<b>163</b>
3.3.1	Construction of a TG mismatched oligonucleotide duplex.....	163
3.3.2	DNase I footprinting of TG mismatched oligonucleotides.....	164
<b>3.4</b>	<b>Results.....</b>	<b>165</b>
3.4.1	Recognition of a single TG mismatch using the tri-imidazole f-III.....	165
3.4.2	Recognition of a double TG mismatch using the heterodimer PII-III...	166
<b>3.5</b>	<b>Discussion.....</b>	<b>169</b>

## **CHAPTER 4: CELL UPTAKE AND LOCALISATION OF POLYAMIDES AND SELECTED MINOR GROOVE BINDERS**

<b>4.1</b>	<b>Introduction.....</b>	<b>173</b>
<b>4.2</b>	<b>Materials.....</b>	<b>176</b>
<b>4.3</b>	<b>Methods.....</b>	<b>177</b>
4.3.1	Cell culture.....	177
4.3.2	Image acquisition and processing.....	177

4.3.3	Visualisation of drug uptake in living cells.....	179
4.3.4	Visualisation of drug uptake in fixed cells.....	180
4.3.5	Microinjection of live cells.....	180
4.3.6	Visualisation of drug uptake in isolated nuclei.....	181
4.3.7	Anti-proliferative activity.....	181
<b>4.4</b>	<b>Results.....</b>	<b>184</b>
4.4.1	Relative fluorescence and nuclear binding of distamycin, TIP-amino, TIP-C7-amino, JH-37, GWL-6 and SJG-136 in fixed cells.....	184
4.4.2	Uptake and intracellular distribution of TIP-C7-amino and TIP-amino in live MCF7 cells.....	188
4.4.3	Uptake and intracellular distribution of TIP-C7-amino in the live Jurkat cells.....	199
4.4.4	Microinjection of live MCF-7 cells with TIP-C7-amino.....	203
4.4.5	Uptake and intracellular distribution of SJG-136 and GWL-6 in live MCF-7 cells.....	205
4.4.6	Uptake and intracellular distribution of the hairpin polyamide JH-37 in MCF-7 cells.....	211
4.4.7	Uptake of TIP-amino, TIP-C7-amino, GWL-6 and JH-37 in isolated nuclei.....	211
4.4.8	Anti-proliferative activity of TIP-C7-amino in Jurkat and MCF-7 cells	216
<b>4.5</b>	<b>Discussion.....</b>	<b>220</b>

## CHAPTER 5: THE EFFECTS OF DISTAMYCIN AND THE THIAZOLE-CONTAINING POLYAMIDE TIP-C7-AMINO ON GLOBAL GENE EXPRESSION IN THE MODEL EUKARYOTIC SYSTEM

### *S. CEREVISIAE*

<b>5.1</b>	<b>Introduction.....</b>	<b>230</b>
<b>5.2</b>	<b>Materials.....</b>	<b>232</b>
<b>5.3</b>	<b>Methods.....</b>	<b>234</b>
5.3.1	Anti-proliferation assays.....	234
5.3.2	Gene expression profiling.....	236
<b>5.4</b>	<b>Results.....</b>	<b>244</b>
5.4.1	Anti-proliferative activity of distamycin A and TIP-C7-amino.....	244
5.4.2	Response of mRNA expression to treatment with distamycin A and TIP-C7-amino.....	247
<b>5.5</b>	<b>Discussion.....</b>	<b>269</b>

**CHAPTER 6: OVERALL DISCUSSION AND FUTURE WORK..... 277**

**APPENDICES..... 288**

**REFERENCES..... 292**

## LIST OF FIGURES

### CHAPTER 1

<b>1.1</b>	Watson-Crick B DNA.....	21
<b>1.2</b>	Phosphodiester and glycosidic linkages.....	24
<b>1.3</b>	Hydrogen bond donors at the groove edges of AT and GC base pairs.....	25
<b>1.4</b>	Packaging of DNA into a chromosome.....	27
<b>1.5</b>	Antigene versus antimessage strategies for nucleic acid targeting.....	28
<b>1.6</b>	Backbone and sugar modified antisense oligonucleotides.....	30
<b>1.7</b>	RNA interference.....	33
<b>1.8</b>	PNA.....	37
<b>1.9</b>	Pyrimidine triplex motif.....	39
<b>1.10</b>	Purine triplex motif.....	41
<b>1.11</b>	Common mechanisms of DNA interaction.....	43
<b>1.12</b>	Frequent positions of covalent base modification.....	47
<b>1.13</b>	(+)-CC-1065.....	48
<b>1.14</b>	Formation of a CPI-DNA adduct.....	49
<b>1.15</b>	Analogues of (+)-CC-1065.....	50
<b>1.16</b>	Reactive CI and CBI moieties.....	50
<b>1.17</b>	Pyrrolo[1,4]benzodiazepines.....	50
<b>1.18</b>	Formation of a PBD-DNA adduct.....	51
<b>1.19</b>	Analogues of pyrrolo[1,4]benzodiazepines.....	52
<b>1.20</b>	Mitomycin C.....	53
<b>1.21</b>	Enediynes.....	54
<b>1.22</b>	Chromomycin and mithramycin.....	56
<b>1.23</b>	Small non-covalent minor groove binders.....	58
<b>1.24</b>	Head to head linked benzimidazole structures.....	60
<b>1.25</b>	Analogues of berenil.....	63
<b>1.26</b>	Netropsin and distamycin.....	64
<b>1.27</b>	Intrusion of the guanine-2-amino group into the minor groove.....	67
<b>1.28</b>	Bonding components of sequence specific netropsin binding.....	68
<b>1.29</b>	Heterocycles for GC recognition.....	70
<b>1.30</b>	Development of molecules for GC recognition.....	71
<b>1.31</b>	2-ImN and 2-PyN.....	74
<b>1.32</b>	Heterocycles for AT recognition.....	76
<b>1.33</b>	Modes of dimeric polaymide binding.....	77
<b>1.34</b>	Covalent linkage of 2-PyN.....	79
<b>1.35</b>	A hairpin linked polyamide.....	81

<b>1.36</b>	Hairpin, cyclic and tandem linkages.....	81
<b>1.37</b>	Hairpin versus cross-linked linkage motifs.....	82
<b>1.38</b>	Isolexins and vinyllexins.....	84
<b>1.39</b>	Netrospin, distamycin and polyamide conjugates.....	86
<b>1.40</b>	Experimental strategy.....	91

## CHAPTER 2

<b>2.1a</b>	Molecular structures of distamycin A, unlinked TII, and unlinked and cross-linked TPI polyamides.....	95
<b>2.1b</b>	Molecular structures of unlinked and cross-linked TPP and TPI polyamides.....	96
<b>2.2</b>	Binding motifs for (a) TPP, (b) TII, (c) TII and (d) TIP.....	97
<b>2.3</b>	DNase I footprint titration with TPP-amino.....	112
<b>2.4</b>	DNase I footprint titration with TPP-formyl.....	113
<b>2.5</b>	Box diagrams of (a) TPP-amino and (b) TPP-formyl.....	114
<b>2.6</b>	DNase I footprint titration with TPP-C7-amino.....	115
<b>2.7</b>	Box diagram of TPP-C7-amino.....	116
<b>2.8</b>	DNase I footprint titration with TPP-acetyl.....	117
<b>2.9</b>	DNase I footprint titration with TPP-C7-acetyl.....	118
<b>2.10</b>	DNase I footprint titration with TII-amino and TII-acetyl.....	120
<b>2.11</b>	Box diagram of TII-amino.....	121
<b>2.12</b>	DNase I footprint titration with TPI-formyl.....	122
<b>2.13</b>	DNase I footprint titration with TPI-C5-formyl.....	123
<b>2.14</b>	Box diagrams of (a) TPI-formyl and (b) TPI-C5-formyl.....	124
<b>2.15</b>	DNase I footprint titration with TPP-formyl and TPI formyl.....	125
<b>2.16</b>	DNase I footprint titration with TII-amino, TPP-amino and TIP-amino.....	127
<b>2.17</b>	Box diagram of TIP-amino.....	128
<b>2.18</b>	DNase I footprint titration with TIP-acetyl.....	129
<b>2.19</b>	DNase I footprint titration with TIP-amino, TIP-C5-amino, TIP-C6-amino.....	130
<b>2.20</b>	DNase I footprint titration with TIP-C7-amino and TIP-C8-amino.....	131
<b>2.21</b>	Box diagram for TIP-C7-amino.....	132
<b>2.22</b>	DNase I footprint titration with TPP-C7-amino and TIP-C7-amino.....	133
<b>2.23</b>	DNase I footprint titration with TIP-C7-amino.....	134
<b>2.24</b>	DNase I footprint titration with TIP-formyl.....	135
<b>2.25</b>	DNase I footprint titration with TIP-C6-formyl, TIP-C7-formyl and TIP-C8-formyl.....	137
<b>2.26</b>	DNase I footprint titration with TIP-C6-formyl and TIP-C6-amino.....	138

<b>2.27</b>	DNase I footprint titration with TIP-C7-formyl and TIP-C7-amino.....	139
<b>2.28</b>	DNase I footprint titration with TIP-C8-formyl and TIP-C8-amino.....	140
<b>2.29</b>	DNase I footprint titration with TIP-C5-hydrogen.....	141
<b>2.30</b>	DNase I footprint titration with TIP-C7-hydrogen.....	142
<b>2.31</b>	DNase I footprint titration with TIP-C7-acetyl.....	143
<b>2.32</b>	Quantitative DNase I footprint titration with TIP-C7-amino.....	144
<b>2.33</b>	Binding isotherm for TIP-C7-amino.....	147

## CHAPTER 3

<b>3.1</b>	Molecular basis of 5-methylcytosine hotspots.....	157
<b>3.2</b>	Molecular basis for Im/Im recognition of GC and GT base pairs.....	159
<b>3.3</b>	Molecular structures if the tri-imidazole formyl-III and PII-III.....	160
<b>3.4</b>	Possible binding motifs for formyl-III and PII-III.....	161
<b>3.5</b>	DNase I footprint experiments with formyl-III.....	166
<b>3.6</b>	DNase I footprint experiments with PII-III.....	167
<b>3.7</b>	Box diagram of oligonucleotide binding by formyl-III and PII-III.....	168
<b>3.8</b>	Modes of binding for formyl-III.....	170
<b>3.9</b>	Recognition of TG mismatches by an Im/HP pair.....	172

## CHAPTER 4

<b>4.1</b>	Molecular structures of TIP-amino, TIP-C7-amino, SJG-136, GWL-6 and JH-37.....	175
<b>4.2</b>	Fixed MCF-7 cells exposed for 1h to 10µM doses of Hoechst 33258, TIP-C8-amino, TIP-C7-amino, TIP-amino.....	185
<b>4.3</b>	Fixed MCF-7 cells exposed for 1h to 10µM doses of SJG-136 and GWL-6.....	186
<b>4.4</b>	Fixed MCF-7 cells either untreated or exposed for 1h to 10µM doses of distamycin and JH-37.....	187
<b>4.5</b>	Live MCF-7 cells exposed for 1h to 0, 3, 10, 30µM doses of TIP-C7-amino over 24, 48, 72h.....	191
<b>4.6</b>	Live MCF-7 cells exposed for 1h to 0, 3, 10, 30µM doses of TIP-C7-amino over 24, 48, 72h.....	192
<b>4.7</b>	Live mitotic MCF-7 cells treated for 72h with 30µM TIP-C7-amino .....	193
<b>4.8</b>	Live MCF-7 cells exposed to 0 & 30µM doses of TIP-C7-amino for 72h, counter-stained with propidium iodide.....	194
<b>4.9</b>	Live MCF-7 cells exposed to 30µM doses of TIP-C7-amino for 48h,	



	counter-stained with both propidium iodide and CDFA AM.....	195
4.10	Live MCF-7 cells incubated for 1h with 10µM Hoechst 33258.....	196
4.11	Live MCF-7 cells exposed to 30µM doses of TIP-C7-amino for 48h, counter-stained with rhodamine 123.....	197
4.12	Live MCF-7 cells either untreated or treated for 72h with 30µM TIP-amino.....	198
4.13	Live Jurkat cells exposed to 30µM TIP-C7-amino over 24, 48, 72h .....	200
4.14	Live Jurkat cells exposed to 30µM TIP-C7-amino for 72h, counter-stained with acridine orange and CDFA AM.....	201
4.15	Live Jurkat cells stained with 10µM Hoechst 33258, plus CDFA AM and PI....	202
4.16	Live MCF-7 cells microinjected with TIP-C7-amino TRITC-dextran.....	204
4.17	Unfixed MCF-7 cells exposed to 10µM SJG-136 for 1h.....	206
4.18	Unfixed MCF-7 cells incubated for 24h with 0, 1, 3, and 10µM SJG-136.....	207
4.19	Unfixed MCF-7 cells incubated with 1, 3, and 10µM GWL-6 for 24, 48, 72h...	208
4.20	Unfixed MCF-7 cells exposed for 24h to 0, 1, 10 and 30µM GWL-6.....	209
4.21	Unfixed MCF-7 cells exposed to 30µM GWL-6 exposed for 10mins, 2h, 24h...	210
4.22	Live MCF-7 cells exposed to 30µM JH-37 for 10mins, 2h, 24h, 48h and 72h...	212
4.23	Live Jurkat cells and isolated nuclei exposed to 10µM CDFA AM for 10min...	213
4.24	Isolated Jurkat nuclei exposed to 10µM Hoechst 33258 or PI.....	214
4.25	Isolated Jurkat nuclei exposed to 10µM TIP-C7-amino, TIP-amino, JH-37 and GWL-6.....	215
4.26	Growth inhibition in Jurkat cells exposed to TIP-C7-amino.....	217
4.27	Viability of Jurkat cells exposed to TIP-C7-amino.....	218
4.28	Growth inhibition in MCF-7 cells exposed to TIP-C7-amino.....	219

## CHAPTER 5

5.1	An example image of a pair of Yeast Genefilters.....	238
5.2	Illustration of the normalisation of the distribution of intensities for two filter images.....	241
5.3	Illustration of a scatter plot of the intensities of the treated sample vs control....	241
5.4	Representation of the distribution of the log <sub>2</sub> transformed normalised intensity ratios.....	242
5.5	Graph of the fraction of control growth versus concentration for a 2 hour incubation of distamycin or TIP-C7-amino in <i>S. cerevisiae</i> .....	244
5.6	Graphs to show the anti-proliferative activity of distamycin & TIP-C7-amino in <i>S. cerevisiae</i> on continuous treatment over 24 hours.....	245
5.7	Graph to show cell viability versus concentration for TIP-C7-amino and distamycin A.....	246

<b>5.8</b>	Typical scatter plot showing the linear relationship between the raw intensities of the control and treated samples, before and after normalisation.....	247
<b>5.9</b>	Scatter plots showing the standard deviation versus the mean ratio of all genes for the parallel and serial treatments with 100µM distamycin.....	267
<b>5.10</b>	Graph to show the minimum replicate number required for a p value of <0.05 for the parallel and serial data sets.....	268

## **LIST OF TABLES**

### **CHAPTER 1**

<b>1.1</b>	The major drug families used in chemotherapy.....	18
<b>1.2</b>	Heterocycle pairing rules.....	75

### **CHAPTER 2**

<b>2.1</b>	Summary of binding activity for thiazolated polyamides.....	145
<b>2.2</b>	Equilibrium association constants for the TIP polyamides.....	148
<b>2.3</b>	Binding site specificity of TIP polyamides.....	149

### **CHAPTER 4**

<b>4.1</b>	Anti-proliferative activity of TIP-C7-amino in Jurkat cells over time.....	218
------------	--	-----

### **CHAPTER 5**

<b>5.1</b>	Top 20 ORFs for three parallel treatments with 100µM distamycin.....	249
<b>5.2</b>	Functions of the protein products of characterised genes derived from top 20 ORFs from parallel treatments with 100µM distamycin.....	250
<b>5.3</b>	Top 20 ORFs for three serial treatments with 100µM distamycin.....	252
<b>5.4</b>	Functions of the protein products of characterised genes derived from top 20 ORFs from serial treatments with 100µM distamycin.....	253
<b>5.5</b>	Complete list of ORFs common to both parallel and serial treatments with 100µM distamycin.....	254
<b>5.6</b>	Protein products of characterised genes appearing in all treatments with 100µM distamycin associated with protein processing and metabolism.....	255
<b>5.7</b>	Protein products of characterised genes appearing in all treatments with 100µM distamycin associated with cell transport mechanisms.....	255
<b>5.8</b>	Protein products of characterised genes appearing in all treatments with	

	100µM distamycin associated with DNA/RNA processing.....	256
<b>5.9</b>	Protein products of characterised genes appearing in all treatments with 100µM distamycin associated with protein folding and stress responses.....	257
<b>5.10</b>	Protein products of characterised genes appearing in all treatments with 100µM distamycin associated with. the cytoskeleton.....	257
<b>5.11</b>	Protein products of characterised genes appearing in all treatments with 100µM distamycin associated with lipid metabolism.....	257
<b>5.12</b>	Top 20 ORFs for three parallel treatments with 100µM TIP-C7-amino.....	259
<b>5.13</b>	Functions of the protein products of characterised genes derived from top 20 ORFs from parallel treatments with.100µM TIP-C7-amino.....	260
<b>5.14</b>	Top 20 ORFs for three serial treatments with 100µM TIP-C7-amino.....	261
<b>5.15</b>	Functions of the protein products of characterised genes derived from top 20 ORFs from serial treatments with.100µM TIP-C7-amino.....	262
<b>5.16</b>	Complete list of ORFs common to both parallel and serial treatments with 100µM TIP-C7-amino.....	263
<b>5.17</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with cell transport mechanism.....	264
<b>5.18</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with DNA/RNA processing and metabolism.	264
<b>5.19</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with cytoskeletal organisation.....	265
<b>5.20</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with glucose metabolism.....	265
<b>5.21</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with lipid metabolism.....	266
<b>5.22</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with protein folding.....	266
<b>5.23</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with ribosome structure and assembly.....	266
<b>5.24</b>	Protein products of characterised genes appearing in all treatments with 100µM TIP-C7-amino associated with sexual conjugation.....	266

## **ABBREVIATIONS**

A	Adenine
ATP	Adenine triphosphate
APS	Ammonium persulfate
C	Cytosine
CBI	Chloromethyl-5-hydroxyl-1,2-dihydro-3H-benz[e]indole
CD	Circular dichroism
CDFA AM	Carboxyfluorescein diacetate
CPI	Cyclopropyl pyrroloindole
cps	Counts per second
DAPI	4', 6-diamidino-2-phenylindole
DSB	Double-strand break
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
EDTA	Ethylethylenediamine tetraacetic acid
FCS	Foetal calf serum
FIV	Feline immunodeficiency virus
G	Guanine
g	Gram
Hepes	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HIV-1	Human immunodeficiency virus type 1
Im	Imidazole-2-carboxamide
K <sub>a</sub>	Equilibrium association constant
kb	Kilobase
LTR	Long terminal repeat
M	Molar
mg	Milligram
ml	Millilitre
mM	Millimolar
μM	Micromolar
μl	Microlitre
MAR	Matrix associated regions
mRNA	Messenger ribonucleic acid

ng	Nanogram
nM	Nanomolar
NMR	Nuclear magnetic resonance
ORF	Open reading frame
pmol	Picomoles
PBD	Pyrrolo[2,1-c][1,4]benzodiazepine
PCR	Polymerase chain reaction
PNA	Peptide nucleic acid
Py	Pyrrole-2-carboxamide
RNA	Ribonucleic acid
RNAi	RNA interference
RNase H	Ribonuclease H
rpm	revolutions per minute
SDS	Sodium dodecyl sulfate
SiRNA	Small interfering RNA
SPR	Surface plasmon resonance
T	Thymine
TAE	Tris-acetic acid EDTA buffer
TBE	Tris-boric acid EDTA buffer
TBP	TATA binding protein
TEMED	<i>N, N, N, N</i> -Tetramethyl-ethylenediamine
TFO	Triplex-forming oligonucleotide
Th	Thiazole- 2-carboxamide
ThI	Thiazole-imidazole-imidazole
TIP	Thiazole-imidazole-pyrrole
TPI	Thiazole-pyrrole-imidazole
TPP	Thiazole-pyrrole-pyrrole
U	Enzyme unit
UV	Ultraviolet

## **ACKNOWLEDGEMENTS**

I would like to thank Professor John Hartley for his supervision and patience, Professor Alan Entwistle for his help and advice regarding the fluorescence microscopy work and Ritu Garg who carried out the microinjection. I would also like to thank Dr Aiden Flynn for his invaluable help with the mathematical analyses. Finally I wish to thank my partner, parents and friends without whose support this work would not have been possible.

# CHAPTER 1

## INTRODUCTION

### 1.1 Cancer

Cancer is a disease typified by progressive genomic change resulting from the accumulation of genetic defects (Vogelstein & Kinzler 1993). These can range from point mutations to changes in chromosome complement, affecting the expression of genes governing the normal regulatory mechanisms of cell proliferation and homeostasis. Irregular tissue growth in the form of a tumour is considered to arise as a consequence of an imbalance between normal cellular proliferation and programmed cell death. Whilst over 100 distinct types of cancer and tumour subtypes have been distinguished collectively within specific organs, the majority of malignant cancer genotypes identified may be characterised by six fundamental pathophysiological changes; growth signal autonomy, an insensitivity to antigrowth signals, evasion of programmed cell death (apoptosis), an ability to undergo indefinite proliferation (immortality), sustained angiogenesis, and tissue invasion and metastasis (as reviewed by Hanahan & Weinberg 2000). These traits may be acquired during the course of tumour development by a series of genomic alterations, each of which may confer a growth advantage leading to the uncontrolled expansion and eventually malignant transformation of a clonal cell population. This appears to occur as a multi-step process, with normal cells evolving via a succession of pre-cancerous lesions into an invasive or metastatic tumour (Foulds, 1954).

Frequently, abnormal growth is benign or not cancerous, resulting in the formation of discrete tissue masses such as papillomas (warts), breast fibroadenomas or leiomyomas (fibroids). These can be treated by surgical removal and in some cases, as for example with fibroadenomas, may resolve spontaneously for reasons that are poorly understood. Benign growth, however, can occasionally be dangerous if it occurs at a critical organ location, for example benign gliomas can produce lethal pressure in the brain if untreated. Furthermore, a number of abnormal growths, such as colonic polyps or cervical dysplasia, are common precursors to invasive or malignant growth. Indeed, the formation of some precursors, for example colonic polyps, may be associated with

heritable syndromes such as familial adenomatous polyposis (FAP) leading to a predisposition to malignant disease (Vogelstein *et al.*, 1988). Once transformed to a cancerous genotype, a malignant tumour becomes invasive, spreading into surrounding tissue. In some cases a tumour will then metastasise, with small clumps of cancerous cells migrating, often via the lymphatic system, from the primary growth site. These micro-metastases are asymptomatic and can attach to a new area of tissue, where initially supplied by the local vasculature they proliferate to form secondary tumours. It is these secondary lesions that often prove life-threatening, causing 90% of cancer deaths (Hanahan & Weinberg 2000), and for which better anticancer therapies are urgently needed.

## **1.2 Anticancer strategies**

A number of therapeutic strategies are currently available in the clinic for the treatment of malignant tumours. If the tumour mass is discrete and located at an accessible site, surgical removal is usually the first course of treatment. This is often followed by localised radiotherapy to destroy cells at the periphery of the excision site, some of which may be remnants of the excised tumour. However, discrimination between normal and cancerous tissue from patient biopsies can be problematic. This can be particularly difficult when distinguishing between non-malignant hyperplastic growth and low-grade lesions, where cells exhibit a malignant phenotype but are still localised to their tissue site of origin. Thus a successful long-term outcome is dependent both on the rapid detection and accurate classification of the primary tumour.

When cancer is not restricted to a discrete tissue mass, as for example in leukaemia, or when a tumour mass has metastasised, chemotherapy is a critical component of treatment. This strategy involves the use of single or multiple chemical agents, either in series or in combination, acting principally to control cellular proliferation by the inhibition of growth or the induction of apoptosis. Currently, approximately 80 chemotherapy drugs are approved by the FDA for use in direct cancer treatment in the USA (<http://fda.gov/cder/cancer/druglistframe.html>). Many of these are small molecules able to reduce tumour growth by the disruption of DNA processing events in dividing



cells. This has proved to be a useful therapeutic target in the treatment of human cancers, especially in tumours with raised rates of proliferation, which show an increased sensitivity to agents blocking DNA replication and transcription (Hurley, 2002). A summary of the major drug families frequently used in chemotherapy is shown below in Table 1.1.

<u>ALKYLATING AGENTS</u>	<u>ANTIMETABOLITES</u>	<u>ANTIMITOTIC ANTIBIOTICS</u>
Nitrogen mustards e.g. mechlorethamine, chlorambucil, melphalan, cyclophosphamide.	Antifolates e.g. methotrexate.	Anthracycline antibiotics e.g. doxorubicin (adriamycin) daunorubicin.
Nitrosoureas e.g. CCNU, methyl-CCNU, BCNU.	Pyrimidine analogues e.g. 5-fluorouracil, Cytosine arabinoside (ara C).	Vinca alkaloids e.g. vincristine, vinblastine, etoposide (VP-16).
Methanesulphonate esters e.g. busulphan.	Purine analogues e.g. 6-thioguanine, 6-mercaptopurine	Non-anthracycline antibiotics e.g. actinomycin D, mitomycin C, bleomycin.
Aziridines e.g. thio-tepa, AZQ, DZQ, triethylenemelamine.		
Triazenes e.g. Dacarbazine, procarbazine, temozolomide.		

**Table 1.1** Summary of the major drug families used in chemotherapy.

A high rate of patient morbidity is often related to chemotherapy, however, due to the limited selectivity of the agents used in this form of cancer treatment. These can produce an associated toxicity in other regions of normal tissue turnover, for example the bone marrow and stomach and follicular epithelia, generating unpleasant side effects to treatment. This toxicity can be dose limiting, restricting the therapeutic index of the anticancer agent used, so there is only a small difference between the effective treatment dose and the lethal dose. Furthermore, for some patients the efficacy of the chemotherapy drugs used can be undermined by either inherent or acquired drug resistance, whereby a sub population of tumour cells escape the effects of the cytotoxic drugs administered resulting in the selective regrowth of a tumour more resistant to therapy. This is a significant problem in cancer treatment, as acquired resistance is the

most frequent reason for the failure of chemotherapeutic regimes in patients with initially responsive tumours (Pratt *et al.*, 1994).

In recent years the aim of novel anticancer strategies has been to refine the selectivity and delivery of therapeutic agents whilst minimising associated systemic toxicity. To this end, a wide range of pre-clinical and clinical anticancer agents have been devised to target each of the common physiological properties displayed by tumours. The proliferation of hormone responsive tumours, such as some forms of breast and prostate cancers, can be significantly reduced by both pre and post-surgical treatment with oestrogen or androgen antagonists, for instance tamoxifen or flutamide. Furthermore, there is now evidence to suggest that hormone antagonists such as tamoxifen can be chemopreventative, deferring tumour onset when given prophylactically to patients with a heritable risk (e.g. BRCA 1 or BRCA 2 mutations) (Pritchard *et al.*, 2003). The problem of constitutive growth factor signalling is being tackled by the development of inhibitors of signal transduction. For instance the tyrosine kinase inhibitors, Iressa™ and Gleevec™ have been shown to inhibit EGFR and Bcr-Abl signalling, respectively (Woodburn *et al.*, 1997; Druker *et al.*, 1996), whilst the monoclonal antibody Herceptin™ can inhibit the action of the Her 2 cell surface receptor which is overexpressed in approximately 25% of human breast cancers (Sledge & Miller 2003). Advances are also being made in the suppression of tumour-associated neoangiogenesis using anti-vasculature agents such as angiostatin and endostatin (Tonini *et al.*, 2003), as well as antibodies directed against proangiogenic factors such as vascular endothelial growth factor, VEGF (Tonini *et al.*, 2003; Longo *et al.*, 2003).

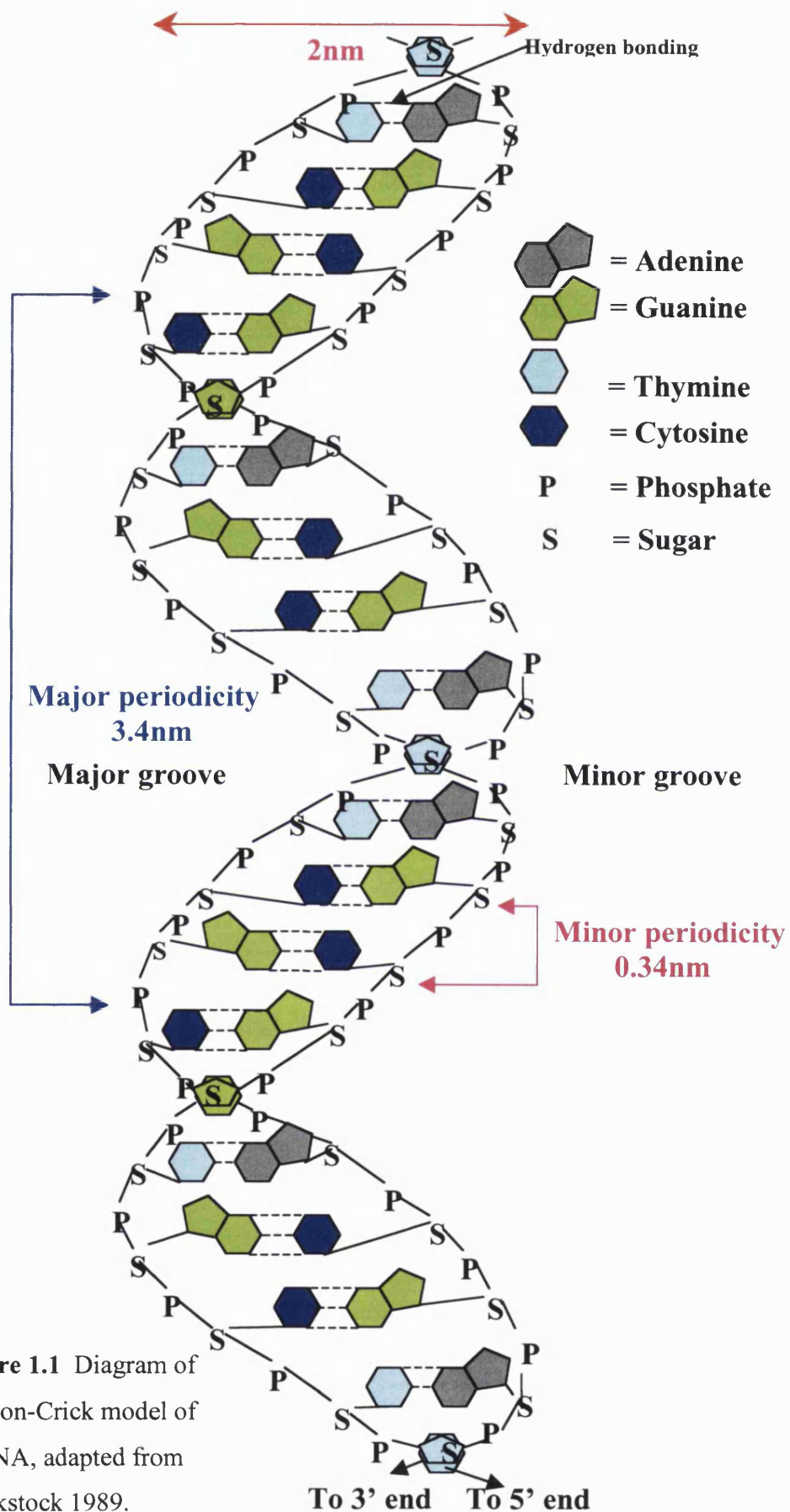
The rational design of new DNA-interactive agents, nevertheless, remains a crucial component of current anticancer research. This encompasses not only the evolution of novel chemotherapy agents with increased DNA specificity in an effort to reduce systemic toxicity, but also of a new generation of DNA interactive molecules able to modulate the expression of critical genes associated with genetic disease, including cancer. Thus, development of sequence selective agents, alongside progress in the fields of genomics and proteomics, provides a useful approach towards the individualisation of

patient therapy with multiple applications in pharmacogenetic screening and diagnosis of a wide spectrum of genetic diseases.

### **1.3 DNA as a molecular target for anticancer agents**

There are many advantages to the strategy of using DNA as a molecular target for cancer therapy. Although a relatively complex macromolecule, the structure of DNA (Figure 1.1) has been well characterised since its initial elucidation by Watson and Crick (Watson & Crick, 1953), facilitating the rational design of DNA-interactive drugs. As a sizeable structure, DNA also affords a large number of putative binding sites for both natural and synthetic ligands. Furthermore, due to its coding function, the maintenance of the molecular integrity of DNA is crucial. Therefore, a relatively small number of critically placed irreversible lesions can successfully disrupt the replication and transcription cycles of DNA, if they go undetected by intracellular repair mechanisms.

Traditional chemotherapy agents, consisting of a variety of structurally diverse molecules capable of interacting with DNA, have exploited these properties, producing critical genetic damage in cells to initiate their death. In general, these agents have been shown to produce an antitumour effect by irreversibly blocking DNA replication or transcription, either directly by the formation of covalent adducts with the DNA or indirectly by the inhibition of critical DNA processing enzymes (Hurley 2002). These mechanisms prevent further cell division, activating a cascade of molecular events leading to cell death. So far however, these drugs have predominantly relied on targeting the high levels of DNA replication and transcription associated with many tumours, rather than selecting for discrete regions of the DNA sequence associated with a transformed genotype. Although many traditional chemotherapy agents demonstrate modest sequence selectivity (as reviewed by Hartley, *et al.*, 1988; Hartley & Souhami 1993), this occurs at relatively small stretches of sequence present in many regions of the genome. Thus, sensitivity to these agents has been as a function of the level of general disruption incurred on genomic replication and transcription in rapidly dividing cells, rather than by the selective disruption of unique tumour related sequences.



**Figure 1.1** Diagram of Watson-Crick model of B-DNA, adapted from Blackstock 1989.

As a consequence traditional chemotherapy agents act indiscriminately on rapidly proliferating tissue, whether it is normal or cancerous.

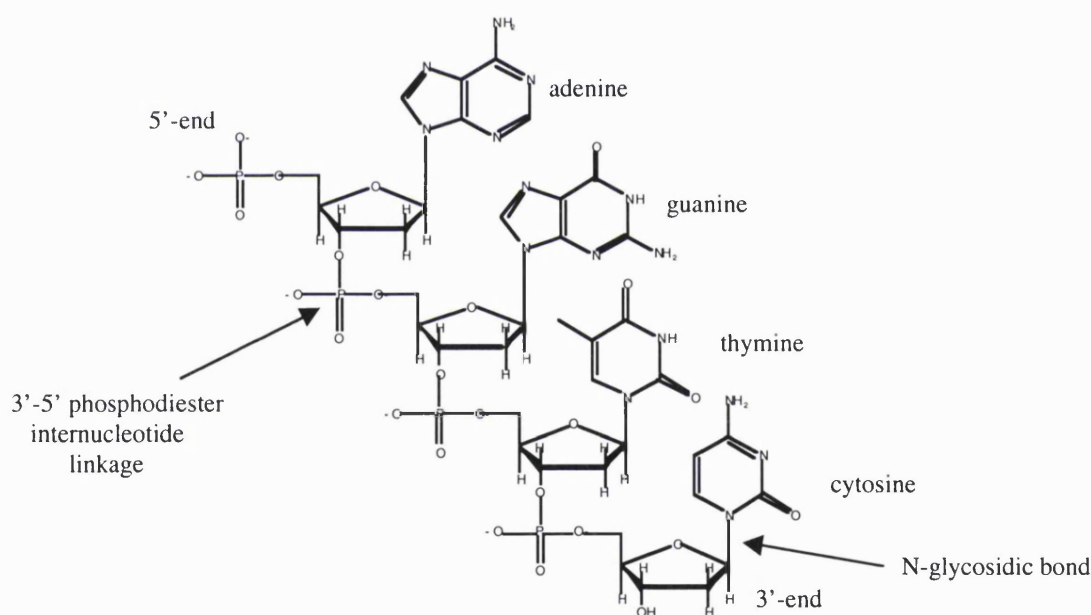
In addition to the benefits of increasing the selectivity of DNA damaging agents, a further advantage of targeting DNA lies in the potential for altering the levels of expression of individual genes and thereby the transcription of their associated protein products. Efforts to discriminate between normal and transformed cells more effectively have resulted in the detailed investigation of the genetic changes involved in the progression of normal tissue to malignancy. Examination of the genetic profiles of common tumours has revealed a correlation between the development of a malignant phenotype and the overexpression, mutation and / or deletion of key genes associated with the regulation of cell growth and tissue homeostasis. These genes and their associated protein products are predominantly involved in the control of the cell cycle, such as p53 (Lane & Crawford 1979; Linzer & Levine 1979; Levine 1997) and p21 (El-Deiry *et al.*, 1993; Harper *et al.*, 1993; Xiong *et al.*, 1993); programmed cell death (apoptosis), including bcl-2 (Tsujimoto *et al.*, 1984; Adams & Cory 1998); and differentiation. These genes, together with the 30,000 plus gene sequences recently identified by the Human Genome Project, provide many novel genetic targets. As a consequence, DNA targeting strategies are becoming increasingly important in an effort to generate selective agents able to modulate the expression of candidate genes. This provides a potentially useful approach for the development of therapeutic agents that could alter expression levels of individual genes, down regulating activated oncogenes such as ras (Midgley & Kerr 2002) and bcr-abl (Lugo *et al.*, 1990; Deininger *et al.*, 2000) and restoring the activity of tumour suppressor genes and their associated proteins products such as p53 and pRB (Levine 1997), by interacting with regions of unique sequence either in the target gene itself or its promoter elements. Accordingly, interest has been renewed in the rational design of a new generation of DNA binding agents with enhanced sequence selectivity, using both naturally occurring and existing synthetic DNA ligands as a structural basis.

There are also disadvantages to using DNA as a molecular target, however, including primarily its lack of accessibility both at a cellular and molecular level. At a cellular level mammalian nuclear DNA, which is generally regarded as the main target of DNA-reactive drugs, is encased within a number of membrane-bound compartments within the cell core. The plasma membrane can restrict the transport of molecules into the cell, preventing intracellular passage and access to the genomic material. This can be exacerbated by active transport mechanisms associated with drug resistance, with some cells exhibiting a multidrug resistance phenotype (MDR). In these cases, cells express high levels of a transmembrane protein P-glycoprotein (Juliano & Ling 1976), which acts as an efflux pump actively transporting drug localised in the cytoplasm back out of the cell (Varma *et al.*, 2003). In addition, internalised DNA targeting ligands can interact with numerous intracellular molecules other than nucleic acids. Indeed, another form of drug resistance involves the interaction of drugs with intracellular thiols, including glutathione (GSH). This can act as a non-specific molecular competitor, binding covalently to some cytotoxic drugs, thereby decreasing their availability for nucleic acid interactions (Fojo & Bates 2003).

Accessibility may also be a problem at a molecular level due to the physical compaction of the DNA macromolecule, which can be extensively supercoiled and associated with protein. Access of putative binding sites may therefore depend on DNA processing events such as replication or transcription, which are enabled by reorganization of the local macromolecular structure and associations. During these processes the DNA structure is affected by superhelical stresses when DNA unwinds, which may in turn influence the reactivity of these open regions to either DNA-binding proteins or drugs. Moreover, once access has been achieved not all drug induced DNA lesions persist, with some being successfully repaired by intracellular mechanisms thus restoring the overall integrity of the DNA. However, whilst these factors complicate the design of DNA-selective agents, DNA still remains one of the most attractive molecular targets for clinical drug development.

## 1.4 The Structure of DNA

The nucleic acids DNA and RNA play a critical role in the storage of genetic information and protein biosynthesis for all living organisms. DNA occurs in the nucleus of most living cells encoding the genetic information of an entire organism. It is a flexible macromolecule composed of nucleotides, which are linked into a polymeric chain by a series of phosphodiester linkages (Figure 1.2). Two polynucleotide chains are arranged as a double-stranded helix, which can be right or occasionally left-handed assuming different forms (A-DNA, B-DNA, C-DNA, D-DNA and Z-DNA) depending on the chemical microenvironment. However, B-DNA predominates in living eukaryotic cells. The constituent nucleotides are themselves comprised of a deoxyribose sugar, a phosphate group and either a double-ringed purine (adenine, A or guanine, G) or a single-ringed pyrimidine base (thymine, T or cytosine, C) (Figure 1.2).

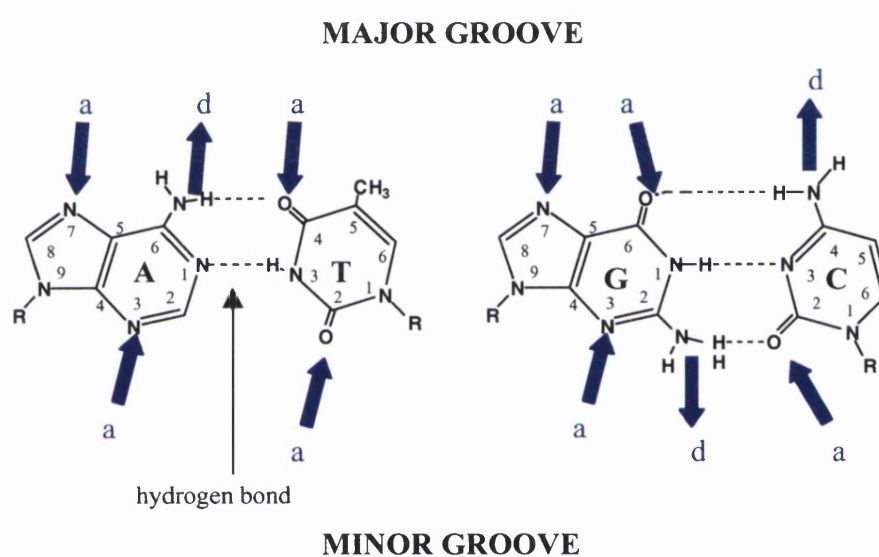


**Figure 1.2** A tetranucleotide sequence showing phosphodiester and glycosidic linkages.

In this structure, each polynucleotide chain is oriented with the phosphodiester linkages facing the outer edge of the molecule to create two antiparallel repeating sugar phosphate backbones. These are held together as a duplex by reciprocal hydrogen bonding between internally placed complementary purine-pyrimidine base pairs,



whereby the planar base pairs stack above each other approximately perpendicular to the helical axis, with thymine forming two hydrogen bonds with adenine and guanine forming three with cytosine (Figure 1.3). It is the hydrogen bonding, in combination with hydrophobic interactions acting vertically between adjacent base pairs causing them to stack, which contributes to the overall stability of the duplex. A stereochemical consequence of this is that the polynucleotide chains orient anti-parallel with respect to each other, such that the internucleotide phosphodiester linkages lie in opposing directions for each chain.



**Figure 1.3** A schematic diagram of Watson-Crick base pairing showing the potential hydrogen bond donors (d) and acceptors (a) at the edges of A: T and G: C base pairs in the major and minor grooves, where A is adenine, T is thymine, G is guanine and C is cytosine.

Each base pair is attached to the sugar phosphate backbone by two corresponding glycosidic bonds (Figure 1.2). Their relative alignment within the double helix results in the formation of external helical grooves or indentations of different sizes, known as the major and minor grooves. The floor of each groove is comprised of the edges of consecutive base pairs running along the helical axis, whilst the two sugar phosphate backbones form the walls either side. Dimensions of these grooves vary between DNA

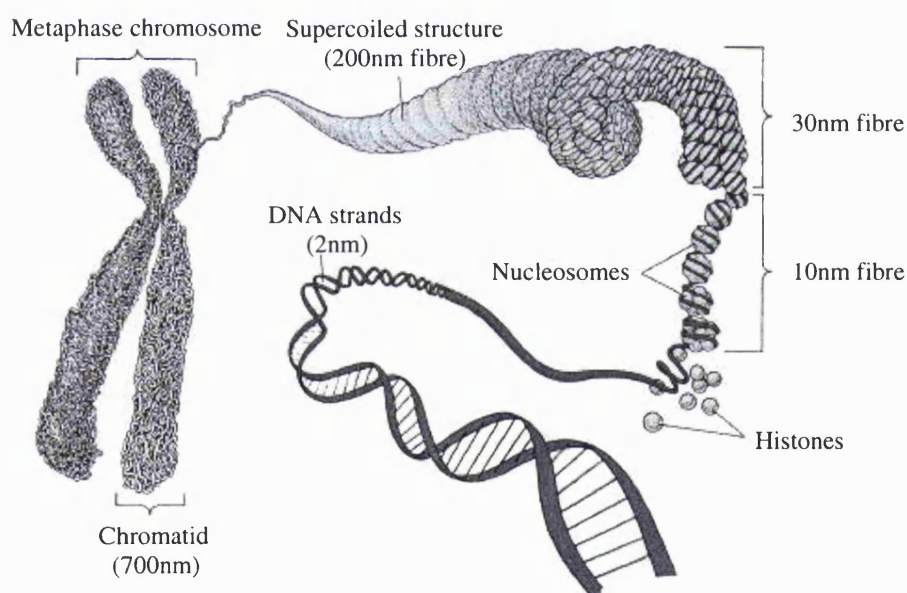


conformations, which can exhibit a range of periodicities, angles of base pairing and numbers of base pairs per helical turn; nevertheless, the major groove is generally wider than the minor groove. The predominant intracellular form of DNA and the most stable, however, is B-DNA, which has 10.5 base pairs per helical turn with an approximate rise of 0.34 nm per base pair. For this structure the average size of the major groove is 1.17 nm wide and 0.88 nm deep, whilst the minor groove is smaller measuring 0.57 nm in width and 0.75 nm in depth. Groove size may also vary according to the local sequence composition, with the minor groove width narrowing over poly dA poly dT runs to 0.3-0.4 nm (Fratini *et al.*, 1982; Nelson *et al.*, 1987) and widening over GC runs (Fagan & Wemmer 1992; Neidle 1992). Furthermore, TA steps have been shown to have a greater deformability than poly dA or dT stretches, due to a compression of the major groove resulting in an associated widening of the minor groove (Goodsell *et al.*, 1994; Dickerson 1998).

Both grooves contain numerous functional groups, which protrude from the edges of the base pairs running along the groove floor. These provide multiple hydrogen bond donors and acceptors, which act as sites of contact for DNA-binding proteins as well as for other smaller molecular ligands such as naturally occurring antibiotics. Local distribution of the electrostatic potential of each groove is sequence dependent, with greater negative potential along AT runs (Lavery & Pullman 1981). The wider major groove exposes numerous functional groups along the base edges that line its floor, offering a large proportion of available bond donors and acceptors. Thus the majority of regulatory proteins bind DNA selectively by forming hydrogen bonds predominantly to specific groups positioned along the major groove (Berg & von Hippel 1988). In contrast the floor of the minor groove is filled with fewer protein recognition sites, however, its lesser dimensions allow the preferential association of many small DNA interactive ligands, which may require more extensive electrostatic van der Waals contacts for binding (Zimmer *et al.*, 1986).

The DNA molecule is seldom free *in vivo* but occurs in most prokaryotic and eukaryotic cells as a supercoiled complex associated with DNA-binding proteins to form

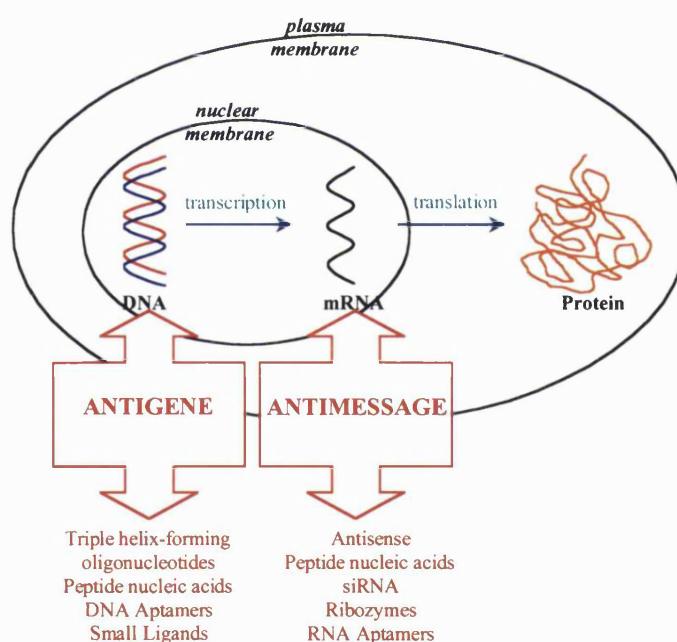
compacted molecules or chromatin as shown in Figure 1.4. This is built up from 200 base pair portions of DNA, negatively coiled twice around a protein core comprised of eight basic proteins or histones. A series of nucleosomes are then coiled into a helical array to produce a protein associated 30nm fibre, with 6 nucleosomes per turn. During replication and transcription regions of DNA are unwound by groups of enzymes known collectively as topoisomerases, allowing access to regulatory proteins, polymerases and possibly DNA-interactive molecules. The general accessibility of compacted nucleosomal DNA to small molecular ligands *in vivo* is not clearly understood. Indirect evidence of the efficacy of numerous cytotoxic agents together with a number of small ligands shown to directly interact with cellular DNA nevertheless suggests that small ligands are able either to penetrate the supercoiled DNA structure, or to access open regions during replication and transcription.



**Figure 1.4** Illustration of the packaging of DNA strands into a chromosome as adapted from <http://voh.chem.ucla.edu/vohtar>.

## 1.5 Strategies for targeting nucleic acids

Sequence selective DNA binding is considered to be a crucial component of the antitumour effect exerted by many traditional chemotherapy agents and naturally occurring antitumour antibiotics. Their limited therapeutic efficacy to date, however, has resulted in a continuing effort to produce novel compounds with enhanced affinities and specificities for DNA, using conventional agents as a structural basis. The recent characterisation of a number of molecular targets for cancer has renewed interest in the selective targeting of nucleic acids and thereby the rational design of sequence specific agents. This has led to the development of new types of anticancer agent able to target known DNA sequences in order to modulate the expression and transcription of selected gene sequences related to malignant phenotypes. Several strategies have been devised to achieve this, targeting nucleic acids either at the level of the gene or at the level of the message (Figure 1.5).



**Figure 1.5** Schematic diagram to illustrate the “antigene” versus “antimessage” strategies for nucleic acid targeting.

### 1.5.1 Antimessage

The disruption of gene expression by targeting nucleic acids at the level of the message was originally demonstrated by Paterson and coworkers, who showed that single-

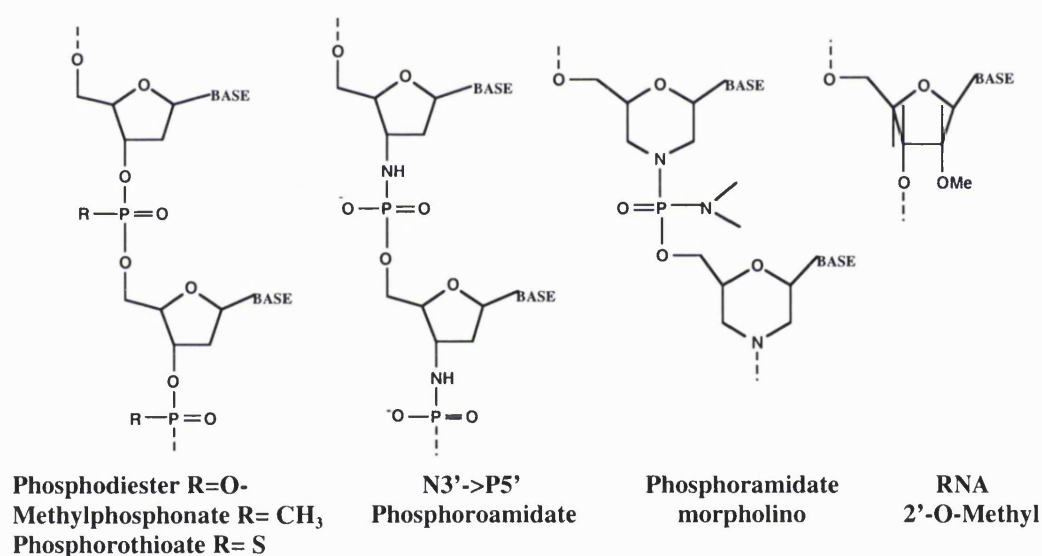
stranded nucleic acids could inhibit the translation of transcript RNA in a cell free model (Paterson *et al.*, 1977). It was later shown that oligonucleotides complementary to the 3' region of the Rous Sarcoma virus could inhibit its replication in cell culture (Zamecnik & Stephenson 1978) and that targeted antisense oligonucleotides could inhibit gene expression in cells (Izant & Weintraub 1984). Gene silencing by targeting transcript RNA has the advantage of the relative accessibility of mRNA during transcription or translation as compared to the genomic DNA, which is compacted and covered in associated proteins. However, many transcript copies must be targeted in contrast to a small number of gene copies and modulation of gene expression may be a transient effect due to the continuous cycle of production and degradation of mRNA. Nevertheless, a number of strategies to block expression at the level of the message are currently under development, including antisense oligonucleotides, small interfering RNAs (siRNA) and ribozymes.

#### *1.5.1.1 Antisense*

At present antisense is the most common oligonucleotide-based antimessage strategy for the inhibition of gene expression. Gene silencing is effected by the hybridisation of a single-stranded deoxyoligonucleotide, often 15-20 nucleotides long, by Watson:Crick base pairing to mRNA or pre-mRNA of complementary sequence transcribed from a selected gene target. Depending on the properties of the antisense oligonucleotide used, translation of the hybridised gene transcript into a protein product is prevented usually either by the degradation of the antisense:mRNA heteroduplex by endogenous ribonuclease H (RNase H) present in both the nucleus and cell cytoplasm or by steric hindrance of mRNA processing and translation (Crooke 1999; Dias & Stein 2002).

Although antisense oligonucleotides with phosphodiester backbones have been shown to inhibit translation of target mRNAs in both cell free systems and microinjected oocytes (as reviewed by Tidd 1993), they are prone, however, to degradation by intracellular nucleases (Wickstrom 1986; Akhtar *et al.*, 1991; Eder *et al.*, 1991). As a result, many analogue structures have been developed which are more resistant to degradation, including the use of modified methylphosphonate or phosphorothioate backbones

(Figure 1.6). These alternative backbones provide antisense molecules with different properties. For example, whilst uncharged methylphosphonate backbones are able to form stable nuclease resistant dimers with complementary polynucleotides, they are unable to elicit RNase H cleavage of hybridised transcripts. In contrast, negatively charged phosphorothioate oligonucleotides of correct stereochemical orientation retain the combined properties of nuclease resistance, a high affinity for RNA and the ability to induce RNase H cleavage (Crooke 1998). As they are easily synthesised and produce exceptional antisense activity they are currently the most widely used class of antisense compounds with several undergoing clinical trials (as reviewed in Jansen & Zangemeister-Wittke 2002). In 1998 the FDA approved the first commercially available antisense drug Vitravene, which is currently used to treat cytomegalovirus (CMV) retinitis in AIDS patients.



**Figure 1.6** Backbone and sugar-modified antisense oligonucleotides.

Despite many promising properties, a number of adverse sequence independent effects have been observed with phosphorothioate oligonucleotides due to their high affinity for various cellular proteins, causing associated toxicity on systemic application (Guvakova *et al.*, 1995; Fennewald & Rando 1995; Agrawal 1999). They are also prone to

irrelevant cleavage, inducing scission by RNase H of nontargeted mRNAs with as little as a 5 base pair complementarity with transcript RNA (Monia *et al.*, 1993; reviewed by Stein 2000). As a result, a new generation of antisense oligonucleotides have been designed in an effort to reduce non-specific binding interactions and nuclease degradation whilst improving the affinity for specific mRNAs (Figure 1.6). These are able to operate via both RNase-H dependent and independent mechanisms (Baker & Monia 1999). Modification of the ribose sugar unit by replacing the H at the 2' position of ribose with a methyl group (RNA 2'-O-methyl) has produced molecules that are nuclease resistant and able to exerting a potent antisense effect by interfering with translation initiation and ribosomal assembly (Baker *et al.*, 1997). Alternatively, substituting the 3' oxygen of ribose with an amine group has created N3'->P5' phosphoroamidate structures able to form stable complexes with both RNA and single and double-stranded DNA (Chen *et al.*, 1995; Gryaznov *et al.*, 1995; Gryzanov 1999). Other structural modifications have included replacing deoxyribose with a morpholine ring and the substitution of phosphodiester linkages with an uncharged phosphorodiamidate backbone (Summerton *et al.*, 1997a, 1999). These structures are both stable and nuclease resistant in biological models (Hudziak *et al.*, 1996), demonstrating targeted antisense activity in cell free systems and in some cell culture lines (Taylor *et al.*, 1998; Summerton *et al.*, 1997b). Chimeric oligonucleotides comprised of different combinations of these structural components have also been developed, taking advantage of the associated properties of each to optimise uptake, transcript hybridisation and sequence specificity (Agrawal *et al.*, 1997, Hamma & Miller 1999; Miller & Hamma 1999).

Despite the potential of antisense oligonucleotides, some serious drawbacks remain, many of which are shared with other methods of DNA targeting. Irrespective of the molecular charge of the antisense oligonucleotide, cell uptake occurs mainly by endocytosis whereby internalised molecules are retained within intracellular endosomal or lysosomal compartments preventing access to the nuclear envelope (Tonkinson *et al.*, 1994; Shoji *et al.*, 1991; Stein 1999). At low concentrations endocytosis may be receptor mediated (Loke *et al.*, 1989; Yakubov *et al.*, 1989) whilst at higher concentrations these

become saturated and uptake occurs by adsorptive endocytosis and pinocytosis (De Diesbach *et al.*, 2000). Although some leakage occurs, a large proportion of exogenously administered antisense oligonucleotides are retained in intracellular endosomal vesicles and are unable to interact with target mRNAs (Gray *et al.*, 1997; Stein *et al.*, 1993). As a consequence a number of delivery systems have been investigated to enhance uptake, including the use of neutral and cationic liposomes such as Lipofectin and cationic biodegradable polymers, such as poly-L-lysine (as reviewed extensively by Dias & Stein 2002 and references therein).

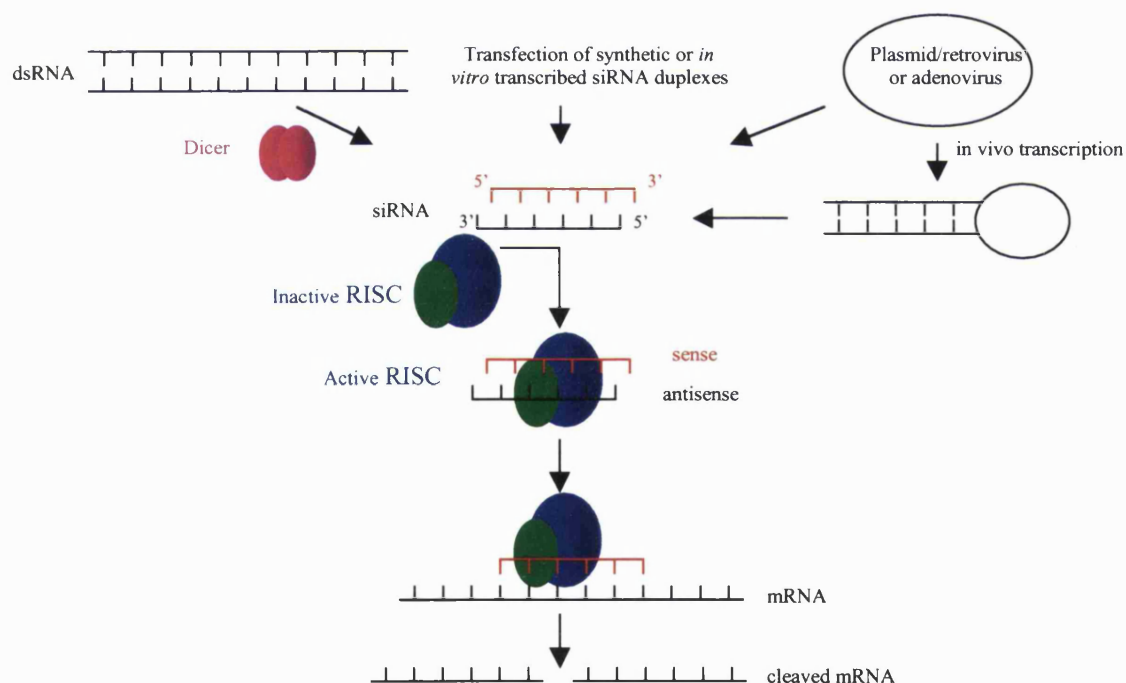
Another problem encountered has been poor accessibility of the target site due to intramolecular folding of the mRNA target, which can interfere with hybridisation. This may be overcome by optimisation of the sequence targeted. Various approaches have been used to identify optimal antisense sequence, including predictive computer RNA folding algorithms and 'walking along the mRNA', whereby the activity of a series of antisense oligonucleotides, directed at multiple sites along the target mRNA is evaluated in model cell culture systems (Bacon & Wickstrom 1991). However, due to the time-consuming nature of this process, mapping the RNase H accessibility of hybridisation sites has also been carried out using random oligonucleotide libraries (Ho *et al.*, 1996; Akhtar 1998) and more recently a combinatorial approach, scanning oligodeoxynucleotide arrays (Milner *et al.*, 1997). This technique uses DNA microarrays comprised of series of overlapping antisense oligonucleotides immobilised onto a nylon membrane or glass slide, designed to hybridise to each possible region of sequence within a given radiolabelled target mRNA (Southern *et al.*, 1994). This may be used to identify cooperatively acting antisense oligonucleotides as well as enable screening of mRNA from cell extracts, as mRNA structures are likely to differ *in vitro* and in biological systems (Scherr & Rossi 1998).

#### 1.5.1.2 RNA interference

A recent development in oligonucleotide based nucleic acid targeting strategies has been the selective post-transcriptional inhibition of gene expression by small interfering 21-23mer double stranded RNA molecules (siRNA or RNAi) (Fire *et al.*, 1998; Hammond



*et al.*, 2000; Sharp *et al.*, 2001). Like some forms of antisense oligonucleotides, gene silencing by RNA interference is affected by the targeted degradation of mRNA (Tuschl *et al.*, 1999). Considered to have evolved as an antiviral defence strategy in plants and lower invertebrates to prevent viral infection and genomic damage from insertable genetic elements, the mechanism for naturally occurring RNA interference has so far been partially elucidated in plants, worms and flies (as reviewed by Hannon 2002). SiRNA is produced endogenously in these organisms and is generated from the cleavage of longer stretches of RNA by an RNase III related enzyme, Dicer (Bernstein *et al.*, 2001; Knight & Bass 2001, Ketting *et al.*, 2001). The resultant siRNA molecule has an unusual double stranded structure with each strand of the 21-23 nucleotide long duplex overhanging at the 3' end (Zamore *et al.*, 2000; Elbashir *et al.*, 2001a). One of these strands is then incorporated into an RNA-induced silencing complex (RISC) forming a protein:RNA complex which acts as a nuclease, digesting mRNA of complementary sequence (Bernstein *et al.*, 2001; Lipardi *et al.*, 2001; Zamore *et al.*, 2000, Figure 1.7).



**Figure 1.7** Proposed mechanism of RNA interference (Lieberman *et al.*, 2003).

Although a recent development in nucleic acid targeting, RNA interference has been used successfully to selectively inhibit the expression a wide range of genes in several



mammalian cell types (Elbashir *et al.*, 2001b; McManus *et al.*, 2002). These have included genes related to cancerous phenotypes such as Bcr-Abl, mutated Ras and human papilloma virus E6 and E7 (Lieberman *et al.*, 2003). Nevertheless, despite the potential of this strategy a number of hurdles remain, precluding immediate therapeutic use. Although resistant to ribonucleases, like antisense oligonucleotides siRNA is prone to degradation by endogenous nucleases making intracellular delivery problematic. Thus chemical modifications such as 5' capping, modification of ribose sugars and substituting phosphorothioate linkages have been used to improve nuclease resistance. Furthermore, gene silencing has been effected by the introduction of synthetic siRNA duplexes into cells by transfection or by *in vivo* transduction by expression from stem-loop structure encoded by plasmids and viral vectors (reviewed extensively in Hannon, 2002; Lieberman *et al.*, 2003). Other possible methods for exogenous delivery include covalent linkage of siRNA to antibodies, cell surface receptor ligands, basic peptides or incorporation into liposomes.

#### 1.5.1.3 Ribozymes

A third anti-mRNA strategy involves the use of ribozymes. These are catalytic oligoribonucleotides able to bind RNA of complementary sequences via Watson-Crick base pairing, causing its degradation by the sequence selective cleavage of its component phosphodiester linkages. Naturally occurring ribozymes include self-splicing introns, RNase P and small catalytic RNAs such as hammerhead, hairpin, hepatitis delta virus and VS ribozyme motifs (reviewed extensively in Puerta-Fernandez *et al.*, 2003). Whilst the tertiary structure varies considerably between ribozyme classes, all types induce the site-specific cleavage of mRNA transcripts, a property with therapeutic potential for the inhibition of gene expression. Furthermore, group I intron ribozymes derived from self-splicing introns may be used for site-specific trans-splicing of therapeutic RNA sequences into a target transcript, a process known as ribozyme-mediated RNA repair (reviewed by Long *et al.*, 2003). Unlike other antimessage strategies, this has the potential for repair of disease-associated mutant RNAs including transcripts involved in carcinogenesis (as reviewed by Phylactou 2000). Indeed,

recently a trans-splicing ribozyme has been used to repair mutant p53 RNA in human cells lines (Watanabe *et al.*, 2000).

Most gene silencing studies using ribozymes, however, have used small catalytic hairpin or hammerhead RNAs which are also proving to be promising therapeutic tools for cancer treatment (Phylactou 2000; Kurrek 2003). As with small interfering RNAs, a variety of delivery systems have been used, including exogenous administration of synthetic ribozymes using cationic lipids and intracellular transcription from plasmid vectors transfected into target cells. The use of viral vectors encoding ribozyme expression cassettes have the advantage of enabling the modulation of the ribozymal concentration and intracellular localisation by using different promoters to drive ribozyme synthesis (Kato *et al.*, 2001). Problems encountered using ribozymes are common to other antimessage strategies, such as the nucleolytic degradation of presynthesised ribozymes. Nuclease stability may be increased by their structural modification, although this is further complicated by reduction in the catalytic activity produced by the resultant conformational changes. Nevertheless, modifications of the constituent RNA structure such as the use of fluorine, amine or 2'-O-alkyl groups at the 2' position of pyrimidine nucleosides have been used successfully with only a small loss of catalytic activity (Beigelman *et al.*, 1995; Zinnen *et al.*, 2002; Kurrek 2003). As a result modified hammerhead and hairpin ribozymes have been used to modulate the replication and transcription of viral RNAs associated with diseases such as hepatitis, HIV, mumps and influenza in addition to transcript targets involved in the induction or progression of tumours (Phylactou 2000). Currently two clinical trials are in progress examining the therapeutic effect of chemically modified ribozymes in breast and colorectal cancer (Sullenger & Gilboa 2002).

### 1.5.2 Antigene

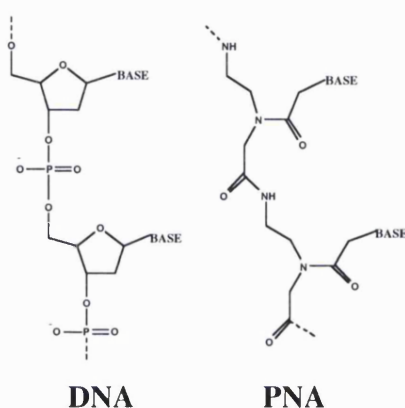
The antigene strategy relies on the rational design of synthetic molecules able to bind to double stranded DNA in a sequence specific manner in order to modulate the expression of a specific gene. This has the advantage of requiring the targeting of only 2 copies or alleles of a given gene as compared to thousands of gene transcript copies. Furthermore,

controlling expression at the level of the gene has the potential of providing a long term effect relative to the transient effect of strategies such as antisense or ribozymes, in which degraded mRNA is rapidly replaced due to the continued transcription of the target gene. However, possible drawbacks include the relative inaccessibility of target sequences within chromatinous DNA in the nucleus, as well as competition for binding sites with native DNA binding proteins. Synthetic molecules designed for this task have ranged from macromolecules, including triple helix forming oligodeoxynucleotides, peptide nucleic acids, zinc finger proteins, peptide libraries and DNA aptamers, to smaller ligands including minor groove binding ligands such as pyrrolobenzodiazepines and polyamides.

#### *1.5.2.1 Peptide-Nucleic Acid chimeras (PNAs)*

PNA is an achiral nucleic acid analogue constituted from DNA bases attached via methylene carbonyl linkages to an uncharged polyamide backbone comprised of N-(2-aminoethyl)glycine units as shown in Figure 1.8 (Egholm *et al.*, 1993; Nielsen *et al.*, 1991). These linkages differ from the peptide bonds in amino acids and as a result PNA is resistant to peptidases and proteases providing the molecule with intracellular stability. As it is uncharged, PNA is able to hybridise with nucleic acids with a high affinity, doing so in a sequence selective manner forming stable duplexes and triplexes with either single-stranded or double-stranded DNA or RNA (Nielsen *et al.*, 1991; Jensen *et al.*, 1997). This enables its use in either antimessage or antigene strategies for the modulation of gene expression. As it is not a substrate for RNase H or other RNases, PNA exerts an antisense effect by inhibiting mRNA splicing, translation initiation or elongation through steric hindrance (Hanvey *et al.*, 1992; Boffa *et al.*, 1996; Cutrona *et al.*, 2000). As a result PNA has been used successfully to down-regulate the expression of a variety of molecular targets (Phylactou 2000). Furthermore, PNA may be used to mediate an antigene effect as homopyrimidine PNAs cause DNA strand displacement, resulting in the formation of a PNA<sub>2</sub>/DNA triplex (or P-loop) at homopurine tracts (Betts *et al.*, 1995). These can inhibit replication or transcription by arresting DNA or RNA polymerases at transcription factor binding sites (Demidov *et al.*, 1995; Nielsen 1999; Ray & Nordén 2000). As with other nucleic acid targeting

molecules, PNA is not readily internalised by cells, although this can vary considerably between cell types (Larsen *et al.*, 1999). Thus initially delivery relied on using high concentrations of PNA or microinjection, streptolysin O, liposomes or cationic lipids. Uptake can be enhanced by conjugating PNA to DNA oligomers (Uhlmann 1998) or to receptor ligands or to anti-receptor antibodies (Basu & Wickstrom 1997). However, PNA is most readily internalised when covalently linked to a nuclear localisation signal peptide (Cutrona *et al.*, 2000; Pooga *et al.*, 2001).



**Figure 1.8** DNA versus PNA, showing the modified backbone structure of PNA.

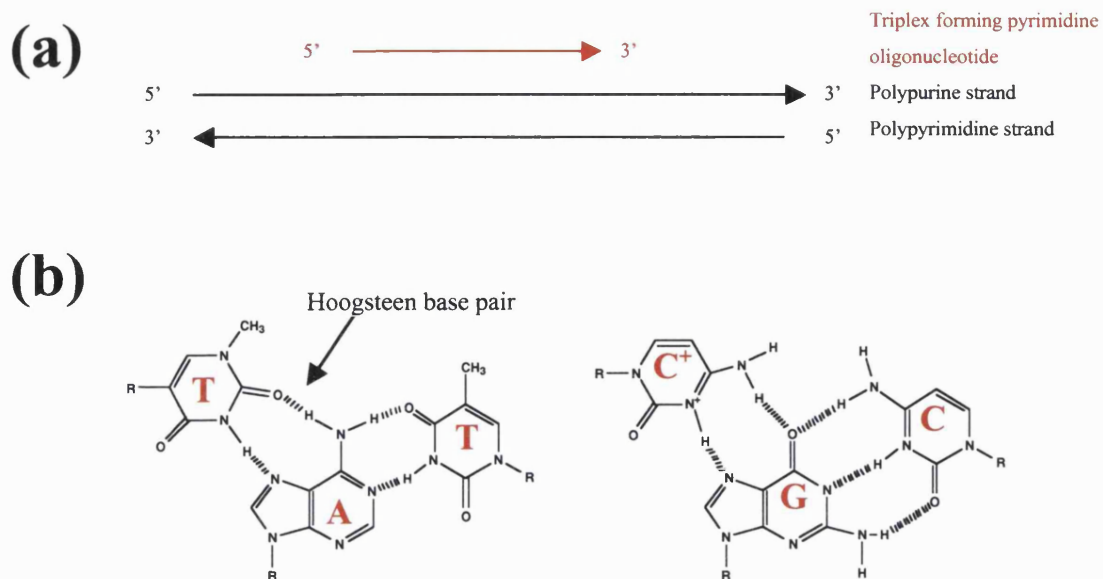
#### 1.5.2.2 Triplex-forming oligonucleotides (TFOs)

Oligonucleotides may bind in a sequence specific manner to the major groove of double stranded DNA at polypurine:polypyrimidine tracts with high specificity to form local triple helical structures (Felsenfeld *et al.*, 1957). This property enables the selective recognition of DNA sequence (Moser & Dervan 1987; Le Doan *et al.*, 1987), with potential for therapeutic use in antigene targeting strategies (Hélène 1991; Thuong & Hélène 1993; Vasquez & Wilson 1998; Chan & Glazer 1997; Praseuth *et al.*, 1999). Triplex formation occurs by either Hoogsteen or reverse Hoogsteen hydrogen bonding between a third oligonucleotide strand and the polypurine strand of a Watson: Crick base paired duplex (Radhakrishnan & Patel 1994; Frank-Kamenetskii & Mirkin 1995). The third strand may be composed of either homopyrimidine or homopurine bases, which bind within the major groove of the DNA duplex in a parallel or antiparallel orientation respectively (Figures 1.9a and 1.10a). Thus an additional pyrimidine strand

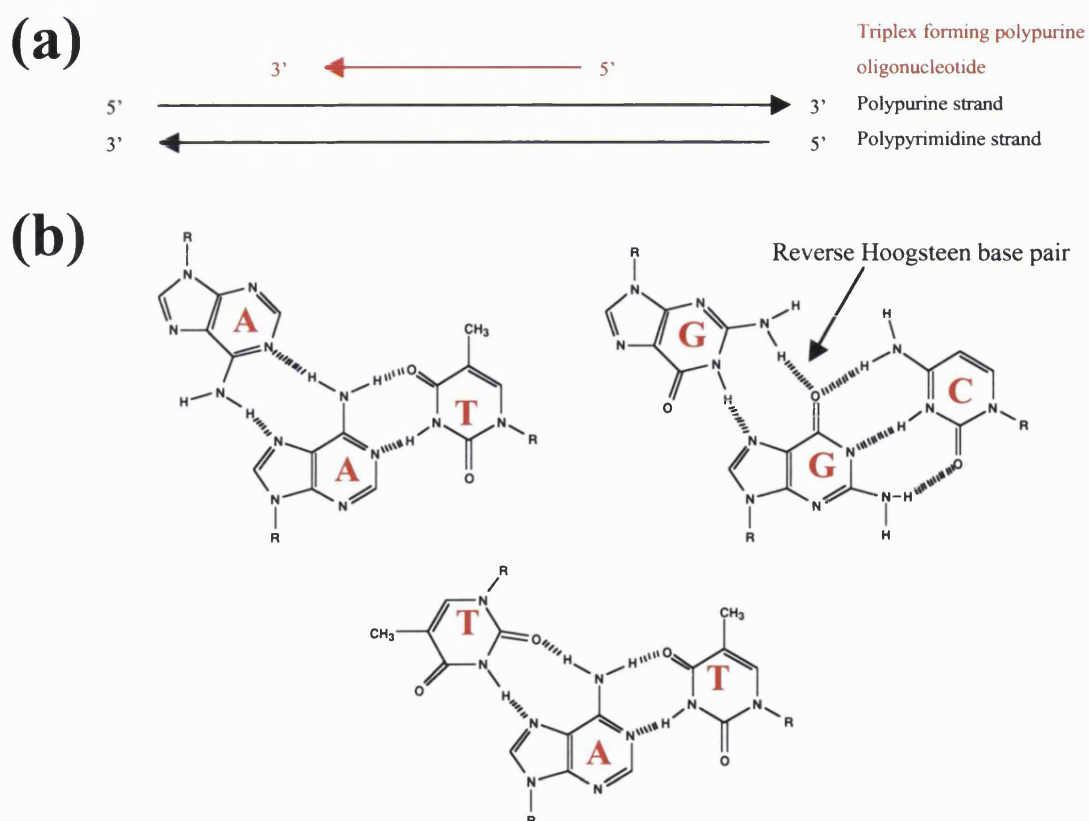
can bind to a complementary purine strand forming a pyrimidine-purine-pyrimidine triplex, so that thymine bonds with an AT base pair and cytosine with a GC base pair, to produce T:A:T and C:G:C triplets (Figure 1.9b). Similarly, synthetic purine stretches can constitute a pyrimidine-purine-purine triplex, with stable T:A:A, C:G:G and T:A:T triplets formed with the purine strand of the Watson:Crick duplex as shown in Figure 1.10 (Beal & Dervan 1991; Pilch *et al.*, 1991). Whilst formation of the purine motif is pH independent, the pyrimidine motif occurs at low pH due to the required protonation of the cytosine residues in the third strand. Furthermore, the pyrimidine motif triplets, T:A:T and C<sup>+</sup>:G:C triplets generate an isomorphous structure enabling the incorporation of the third strand into the major groove without distortion of the phosphodiester backbones (Giovannangeli *et al.*, 1992). In contrast, the base triplets of the purine motif produce some backbone distortion, which affects the stacking and thus stability of the three strands. Triplex formation is further stabilised *in vitro* by the presence of divalent metal cations, which reduce repulsion between the negatively charged phosphodiester backbones (Malkov *et al.*, 1993; Potaman & Soyfer 1994).

Despite successful use of TFOs to induce site-specific mutations, sequence selective strand cleavage and modulation of gene expression *in vitro* (reviewed in Chan & Glazer 1997), a number of problems have been encountered with their use under physiological conditions. Although polypurine:polypyrimidine sequences are common in mammalian genomes, numerous attempts have been made to extend the range of base recognition to include regions of mixed sequence (as reviewed by Gowers & Fox 1999). Furthermore, the requirement of pyrimidine motif structures for low pH and high concentrations of divalent cations such as Mg<sup>2+</sup> for the formation of a stable triplex is difficult to reproduce in cells at a physiological pH (Blume *et al.*, 1999). Moreover, homopurine third strands have a tendency to form self-associating G tetrads at physiological levels of K<sup>+</sup>, inhibiting triplex formation (Arimondo *et al.*, 2001).

A number of modifications have been made to the constituent bases and phosphodiester backbone of the third strand in an effort to improve triplex stability at physiological pH



**Figure 1.9** The pyrimidine triplex motif showing (a) the parallel orientation of the third polypyrimidine strand and (b) the base triplets T·A·T and C<sup>+</sup>·G·C formed by Hoogsteen hydrogen bonding, as adapted from Seidman & Glazer 2003.



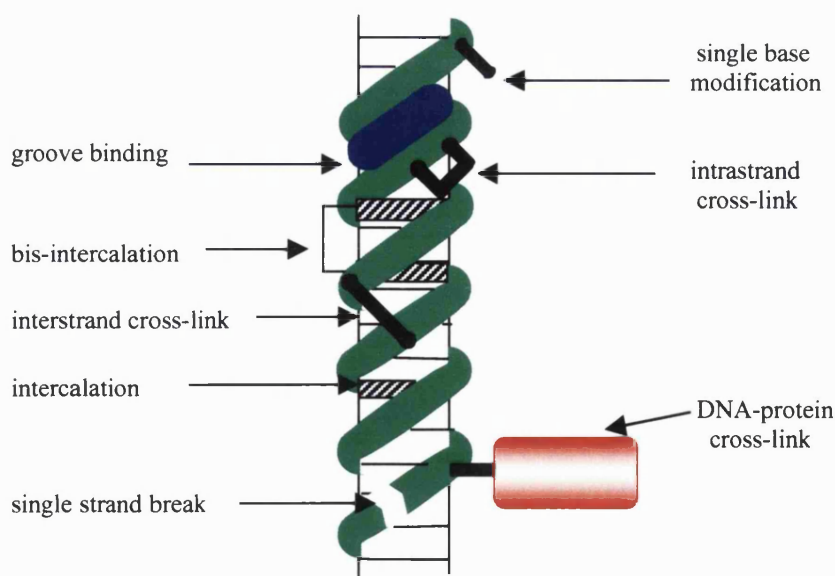
**Figure 1.10** The purine triplex motif, showing (a) the antiparallel orientation of the third polypurine strand and (b) the base triplets of A·A·T, G·G·C and T·A·T.

and resistance to intracellular nucleases (reviewed in detail by Agrawal 1999). The substitution of natural bases with synthetic analogues, for example replacing cytosine with 5-methylcytosine, reduces the pH restriction in the pyrimidine motif, enabling binding at neutral pH (Lee *et al.*, 1984). Furthermore, guanine analogues such as 2'-deoxy-6-thioguanosine (Gee *et al.*, 1995, Vasquez *et al.*, 1995) or 7-chloro-7-deazaguanines (Aubert *et al.*, 2001) can enhance stability of the purine motif by reducing tetrad formation. Further stabilisation has been achieved by covalent attachment, either at the 5' or 3' ends or at internal positions within the third strand, of minor groove binding and intercalating agents (Kukreti *et al.*, 1997; Praseuth *et al.*, 1999). Use of alkylating or photoreactive agents at these positions has provided novel sequence selective properties. Numerous modifications have also been made to triplex oligonucleotides to reduce susceptibility to intracellular nuclease degradation. These include altering  $\beta$ -glycosidic linkages to their  $\alpha$  anomers (Thuong & Hélène 1993; Noonberg *et al.*, 1995), substitution of component deoxyribose sugars with ribose (Thuong & Hélène 1993), as well as the use of methylphosphonate, phosphoramidate, morpholino and PNA backbones (reviewed by Praseuth *et al.*, 1999; Seidman & Glazer 2003).

Whilst selected TFOs can modulate the expression of endogenous genes and viral sequences *in vitro* and in cells (Cooney *et al.*, 1988; Postel *et al.*, 1991; Praseuth *et al.*, 1999; Guntaka *et al.*, 2002), in some cases it has been difficult to prove that the effect has been directly mediated by triplex formation. Nevertheless, modulation is thought to occur by steric interference of replication, transcription, transcription factor and topoisomerase/helicase binding. However, there are a number of limiting factors for therapeutic use of TFOs. These include problems with low cell permeability requiring the use of a variety of vectors such as cationic lipids and peptide conjugation (Praseuth *et al.*, 1999), the inhibition of triplex formation by nucleosomes (Brown & Fox 1996; Espinas *et al.*, 1996) and the restricted range of sequence recognition (Gowers & Fox 1999). As a result, although many antisense agents are undergoing clinical trials, there are currently no TFOs in pre-clinical trials.

## 1.6 Targeting DNA using small molecular ligands

Small ligands may interact with DNA via a number of molecular mechanisms as summarised in Figure 1.11, which may involve either irreversible or reversible binding. Irreversible binding of a chemical agent with DNA usually requires some form of covalent interaction. Such agents may be monofunctional, possessing a single reactive group, or bifunctional with two and can elicit a wide range of base modifications in both grooves of the DNA. These can occur at single bases forming monoadducts, as well as extending over more than one base to form cross-links in the case of bifunctional ligands. Strand breaks may also be introduced into the DNA by a number of means including the metabolism of chemical agents to produce highly reactive species such as oxygen radicals and the inhibition of DNA complexes with its processing and repair enzymes which break the DNA strands in order to reorganise the local structure. These lesions, together with spontaneous DNA damage by methylation, deamination, depurination and oxidation (Friedberg *et al.*, 1995) may be either mutagenic or cytotoxic depending on the site of damage, with critically placed lesions triggering a chain of molecular events leading to cell death unless they are repaired. Non-covalent ligand binding is mediated by reversible electrostatic interactions such as van der Waals forces or hydrogen bonding, as well as by non-polar interactions allowing intercalation and groove-binding which may occur either in the major or minor groove of the DNA.



**Figure 1.11** Illustration depicting the common mechanisms of DNA binding and damage produced by small DNA-interactive ligands.

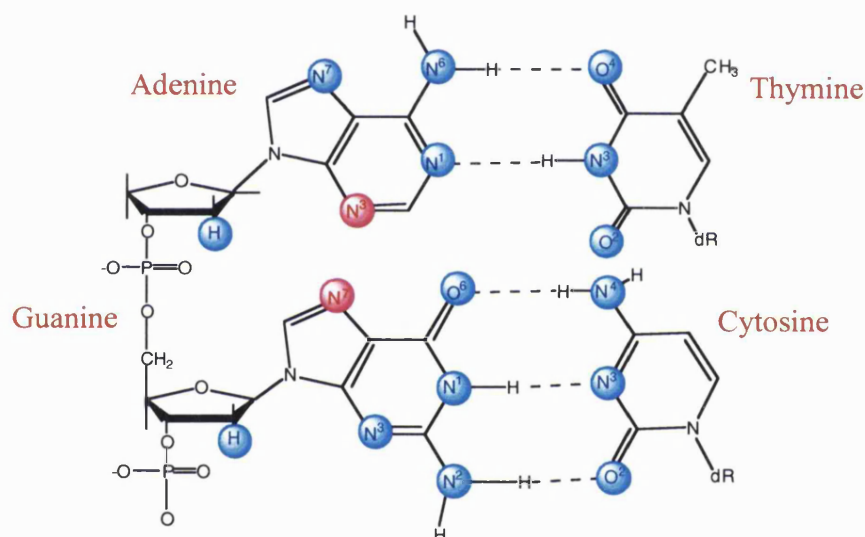


### 1.6.1 Covalent modification

DNA damage involves the modification of its structural components by the formation or breakage of either permanent or transient chemical bonds. Small ligands can produce such damage by the covalent modification of its constituent nucleotides. This can be in the form of base or phosphate alkylation, interstrand or intrastrand cross-links and single or double strand breaks, with most DNA damaging agents producing more than one type of lesion. Whilst much chemical damage is consistently repaired, failure to do so can result in the loss of the sequential integrity of the DNA leading to mutation. Alternatively irreversible damage at critical locations in the DNA can disrupt crucial DNA processing events such as replication and transcription, causing cell cycle arrest promoting cell death.

DNA bases are vulnerable to alkylation at physiological pH and the antitumour activity of alkylating agents is thought to occur as a result of such lesions. These agents are electrophilic and have an affinity for the numerous nucleophilic centres in the DNA structure. As a result they can produce a variety of covalent single base modifications (Figure 1.12), with the highly nucleophilic guanine-N7, adenine-N3 and adenine-N7 positions most readily alkylated (as reviewed extensively by Hartley 1993b). Such modifications weaken the glycosidic bonds between the bases and their associated sugars and therefore often lead to depurination/depyrimidination and the formation of alkali-labile abasic sites (Loeb & Preston 1986). When an alkylating agent is bifunctional it can also react with two nucleosides to form cross-links in the DNA either in the same strand (intrastrand) or in the opposite strand (interstrand). This occurs less frequently than single base modification possibly due to the additional stereochemical constraints of bifunctional alkylation. Nevertheless, although cross-links account for a relatively small proportion of alkylations, they are often considered to be critical lesions. This may be because if unrepaired, interstrand cross-links prevent strand separation and thereby interfere with DNA replication and transcription. Consequently, a number of cross-linking agents such as nitrogen mustard (Kohn *et al.*, 1966; Fox & Scott 1980) and mitomycin C (Iyer *et al.* 1963; Borowy-Borowski *et al.*, 1990) have been used as chemotherapeutic agents.

Strand breakage often accompanies other forms of DNA damage and thus may not always be the primary cytotoxic lesion. However, it is the primary lesion produced by bleomycin and enediyne antibiotics (Povrik 1983; Smith & Nicolaou 1996). The binding of these agents to DNA results in the release of reactive radical species, which induce the formation of both single and double strand breaks. Double strand breaks are usually formed when both DNA strands are broken and the opposing breaks are close enough to sever the DNA molecule. Although a small proportion of chemically induced strand breakage results in the formation of double strand breaks, like cross-links, they are more disruptive to DNA processing and are thus considered more mutagenic unless repaired.



**Figure 1.12** Frequent positions of covalent base modification in DNA, as adapted from Friedberg *et al.*, 1995.

### 1.6.2 Non-covalent binding

Non-covalent interactions with DNA can take the form of intercalation or groove binding. Intercalation occurs when planar aromatic ligand molecules are inserted between stacked base pairs, causing a local structural distortion of the DNA, which stiffens and unwinds to accommodate the intercalator. Due to this distortion intercalators can only bind between alternate base pairs, limiting their range of interactions (Wilson 1990) and thus their sequence specificity. Two intercalating units

may be linked to produce a bis-intercalator, extending the sequence specificity. Some potent antitumour agents interact with DNA primarily by intercalation between the DNA bases but nevertheless have a groove-binding component, for example positioning a bulky oligosaccharide or peptide group into the groove. Anthracycline antibiotics such as arugomycin and nogalamycin position bulky constituent oligosaccharide components in both grooves, whilst mono-intercalators such as daunomycin and adriamycin place their sugar groups into the minor groove only. Thus, in these cases although the intercalating unit of the ligand structure primarily determines binding, these minor groove-binding substituents influence flanking sequence requirement (Frederick, *et al.*, 1990).

Non-covalent groove binding is a form of ligand interaction with DNA, which occurs via electrostatic interactions such as hydrogen bonding, van der Waals forces and hydrophobic interactions. These attractive forces are sensitive to distance and thus binding can depend on the three-dimensional shape of the ligand, with groove-binding favoured by molecules which fit neatly into the DNA groove. Isohelicity of a molecule, whereby it follows the curvature of the DNA helix, facilitates insertion into either the major or minor groove of DNA parallel to those grooves. In some cases, non-covalent groove binding can act as a component of subsequent covalent interaction. An example of this is the binding of the CPI-containing antibiotic (+)-CC1065, which will be discussed in a later section.

### **1.7 DNA recognition in the minor groove by small ligands**

In recent years a large number of agents have been synthesised based structurally on naturally occurring minor groove binders including, distamycin, netropsin, (+)-CC-1065, anthramycin and Hoechst 33258. These have been developed in an effort to generate novel therapeutic compounds with sequence selective anticancer and antiviral properties, providing an alternative strategy to current oligonucleotide based techniques for nucleic acid targeting such as antisense. Classic minor groove binding ligands can be resolved into two groups based on their charge; cationic and neutral, however the

majority are positively charged. These may interact with DNA via non-covalent or covalent interactions or combinations of the two.

The majority of DNA-binding control proteins bind in the major groove, with some interactions extending into the minor groove, probably due to the relatively high numbers of available hydrogen bond donors and acceptors present there. Nevertheless some important regulatory proteins, including polymerases and the TATA binding protein (TBP), utilise the minor groove. Many small ligands including xenobiotics usually less than 1000Da in size, bind predominantly in the minor groove possibly as a result of better van der Waals contacts in this narrower groove. This phenomenon may have been as a result of the evolutionary adaptation of some prokaryotic species to produce antibiotics disrupting the protein-DNA interactions of competing organisms, by capitalising on the relative lack of occupancy of the minor groove (Lown, 1990; Lown 1997). Therefore, although at first instance the minor groove does not appear to contain as much structural information as the major groove, there are nevertheless a number of advantages in targeting this region.

#### 1.7.1 Features of minor groove recognition

Minor groove recognition relies on a series of structural features. These include the charge, shape, chirality, stereochemistry, flexibility and in the case of longer ligands phasing, as well as the ability of incoming hydrogen bond donors or acceptors to recognise specific base pairs. Minor groove binding ligands generally share a number of common structural features, which enable both covalent and non-covalent binding between the groove walls. A large number of minor groove binding ligands are cationic complementing the electrostatic potential of AT-rich regions. These are drawn initially towards AT-rich sequences by a non-specific electrostatic attraction for the negative charge associated with the phosphate backbones of the DNA. This is compounded by the deep negative charge potential near the bottom of the minor groove at the AT region (Lavery *et al.*, 1982). For GC-rich regions, the deepest negative potential well is associated with the major groove, whilst for AT-rich sequences this occurs in the minor groove.

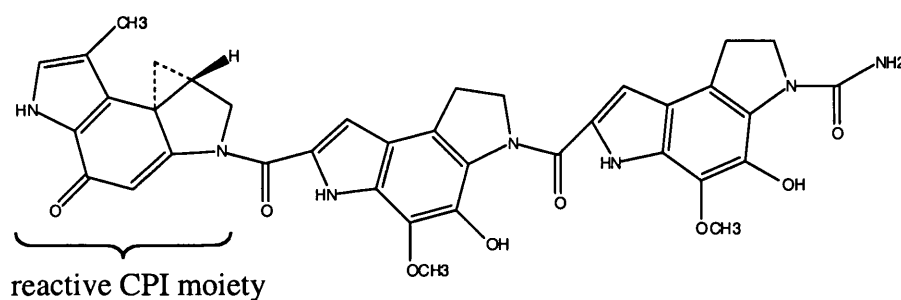
The minor groove floor is formed from the base edges providing a convex surface, which is favourably complemented by the concave shapes of many groove-binding ligands. This shape dependence is strongly affected by changing the chirality of portions of the ligand, with racemates often demonstrating poor binding. Once positioned within the groove, the stereochemistry of the chemical groups constituting the ligand, together with their ability to accept or donate hydrogen bonds, determines its alignment alongside specific base pairs. For example, the protrusion of the guanine amino group into the groove often produces steric hindrance to ligand binding, establishing an AT binding preference. Furthermore, hydrogen bonding at the base edges provides an ability to discriminate between AT and GC base pairs due to the different selection and spatial arrangement of bond donors and acceptors available for each base. Close association with the groove floor then allows other stabilising non-bonded interactions, such as hydrophobic interactions and van der Waals contacts with the groove walls although their influence on binding remains somewhat controversial (Sauers 1995; Chang & Cheng 1996). These are distance dependent and thus are optimised in the smaller dimensions associated with the minor groove, especially in the narrow environment provided by polyA.polyT stretches. As a consequence many natural non-covalent binding ligands show general AT selectivity, although the requirements for their detailed sequence recognition differs.

Small molecules such as Hoechst 33258 and distamycin, bind exclusively in the minor groove, whilst others require only partial groove binding, often placing bulky substituents such as a sugar, in the case of daunomycin or adriamycin, or a peptide in the case of actinomycin into the groove (Geierstanger & Wemmer 1995). DNA is relatively flexible, so some molecules are able to bind despite causing major distortion to the local DNA structure, as for example is the case for chromomycin and mithramycin, which target GC-rich regions as they require a wider groove. Others cause only minor distortion due to their small size and isohelicity with the DNA helix, for example Hoechst 33258, DAPI, netropsin and distamycin.

## 1.7.2 Covalent interactions: chemically reactive minor groove binders as possible agents of chemotherapy

### 1.7.2.1 CPI-containing antibiotics

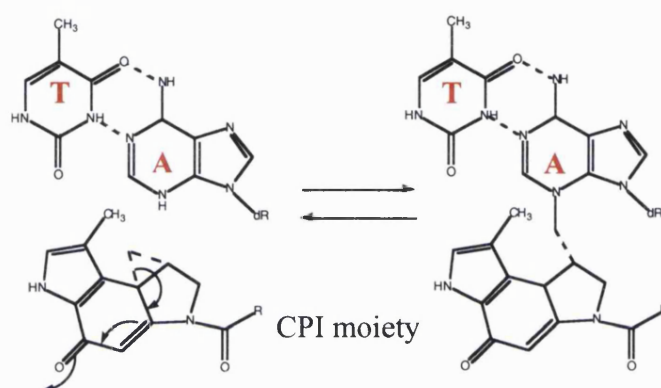
(+)-CC-1065 and the structurally related duocarmycins form a class of potent antitumour antibiotics originally isolated from various *Streptomyces* strains (Hanka *et al.*, 1978; Yasuzawa *et al.*, 1988; Takahashi *et al.*, 1988). As confirmed by single X-ray crystallography (Chidester *et al.*, 1981), (+)-CC-1065 is composed of three pyrroloindole units; a reactive cyclopropyl pyrroloindole moiety (CPI) and two further non-covalent binding subunits, as shown in Figure 1.13. It is a monofunctional alkylating agent which is thought to act via a shape-dependent catalysis mechanism, whereby on non-covalent binding in the minor groove, (+)-CC-1065 adopts a helical conformation producing helical twist in its linking N2 amide, disrupting the stabilising influence of the group.



**Figure 1.13** The molecular structure of (+)-CC-1065

Non-covalent binding activates nucleophilic attack by the CPI moiety causing alkylation at the N3 position of adenine in the minor groove (Boger *et al.*, 1991a, b, 1997a, 1999), occurring via a reversible adenine N3 addition to the least substituted carbon of the reactive CPI moiety within selected AT rich regions as shown in Figure 1.14 (Hurley & Needham-VanDevanter, 1986; Reynolds, *et al.*, 1985). As a result, (+)-CC-1065 selectively targets the consensus sequence 5'-(T/A)(T/A)A-3', with 5-TAA-3' being the preferred site of alkylation (Hurley *et al.*, 1984, 1988; Reynolds *et al.*, 1985; Boger *et al.*, 1991 a, b) as shown in Figure 1.14. Non-covalent binding is a prerequisite for DNA

alkylation by (+)-CC-1065 (Boger *et al.*, 1991c, 1996). Interestingly, the (-) enantiomers of CC-1065 and the structurally related duocarmycin drug family are less reactive than (+) (Warpehoski *et al.*, 1988), suggesting the importance of molecular shape in groove-binding. These form covalent adducts with the ring system pointing in opposite direction to (+) indicating that the geometry of non-covalent bound (-) racemate was unfavourable for the alkylation reaction.

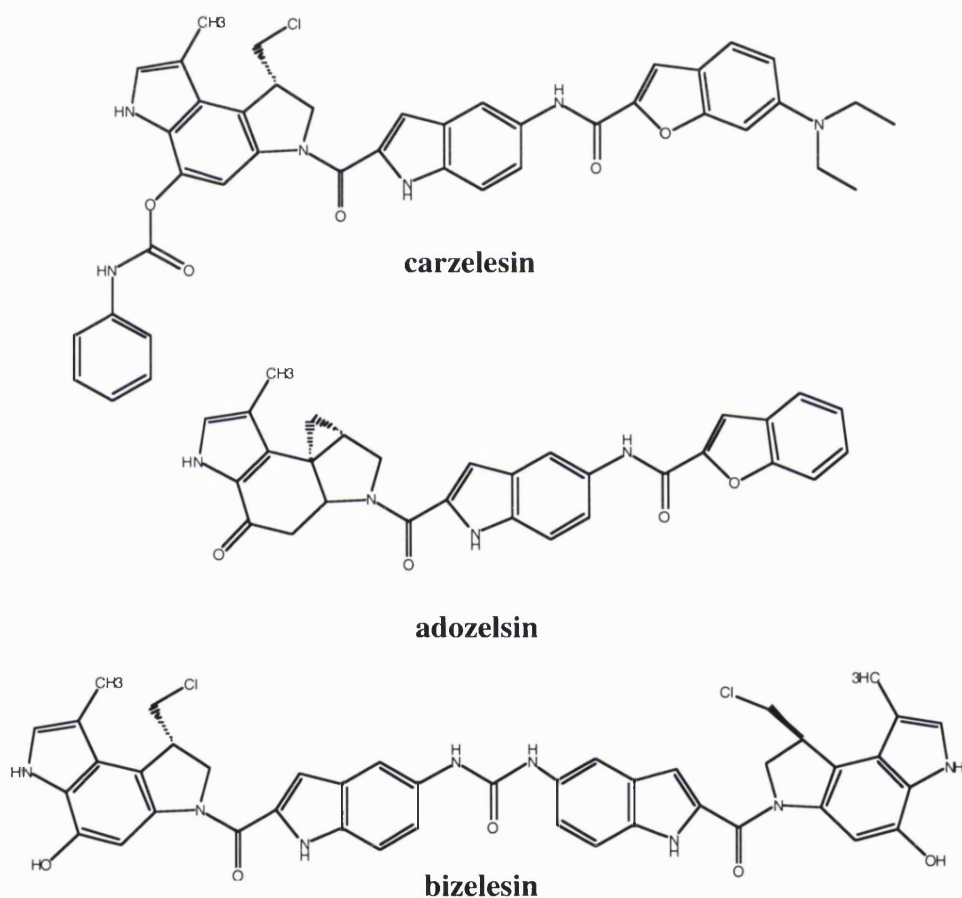


**Figure 1.14** Formation of a CPI-DNA adduct.

Whilst (+)-CC-1065 is a highly potent antitumour agent, exerting its cytotoxicity through covalent binding to DNA (Martin *et al.*, 1981, Li *et al.*, 1982), *in vivo* studies demonstrated that it produced a delayed concomitant hepatotoxicity at therapeutic doses prohibiting its use in the clinic (McGovren *et al.*, 1984). To circumvent this, structurally related analogues were synthesised which retained the reactive CPI portion, such as adozelesin and carzelesin (Figure 1.15). These proved to be highly effective antitumour agents *in vivo* (McGovren *et al.*, 1984; Bhuyan *et al.*, 1992 a, b; Li *et al.*, 1991). The subsequent development of a bifunctional alkylator based on (+)-CC-1065, bizelesin, resulted in an increased sequence selective capability and cytotoxicity than that observed with the monofunctional agents (Ding & Hurley 1991; Mitchell *et al.*, 1991). This formed interstrand cross-links across two adenine N3-positions, spanning a total of six to seven base pairs (Sun & Hurley 1993).

Despite the promising activity of these analogues in preclinical studies and preliminary clinical trials (Carter *et al.*, 1996; Pitot *et al.*, 2002), significant myelotoxicity has been

reported for both adozelesin and bizelesin in Phase II trials (Volpe *et al.*, 1996; Cristofanilli *et al.*, 1998). However importantly, recent studies of bizelesin indicate that it may preferentially target AT rich matrix associated regions (MAR), areas of genomic DNA vital to replication, providing evidence that small sequence-specific molecules can produce non-random region-specific damage to genomic DNA (Woynarowski *et al.*, 2000; Woynarowski 2002). This may ultimately prove to be a significant factor in the high antiproliferative and antireplicative potency of some minor groove binding drugs including CPI-containing molecules.

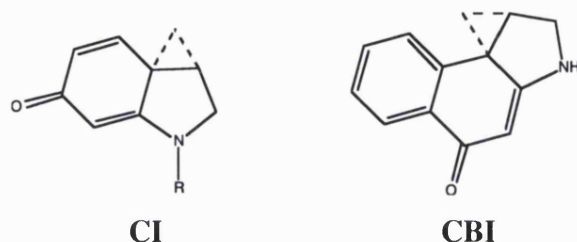


**Figure 1.15** The analogues of (+)-CC-1065, adozelesin, carzelesin and bizelesin.

A series of modifications have been made to the CPI DNA alkylation subunit to investigate its contribution to the affinity and selectivity of DNA association (reviewed in Boger *et al.*, 1997b). Reduction of the reactive CPI moiety found in (+)-CC-1065 to



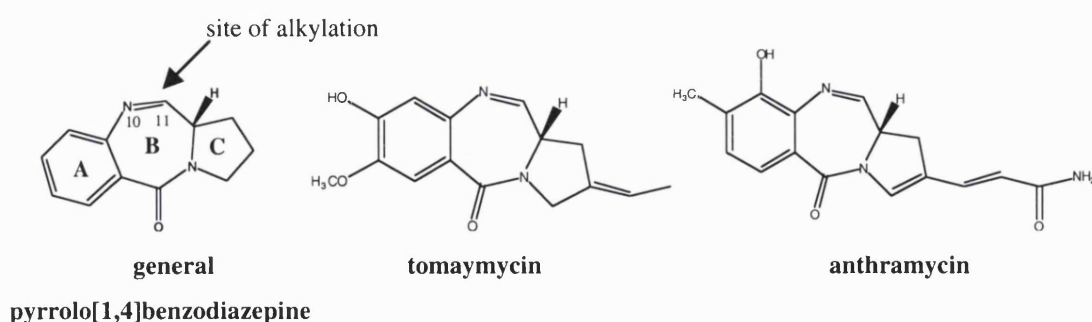
CI (Figure 1.16) was found to result in a more reactive unit of alkylation, with similarities in the DNA alkylation patterns of CPI and CI derivatives indicating that CI represents the minimum pharmacophore (Boger *et al.*, 1991c, 1996). A number of CBI based analogues of (+)-CC-1065 have also been synthesised and evaluated (Boger *et al.*, 1993, 1995). These were found to have enhanced stability and potency relative to corresponding CPI analogues, alkylating with unaltered sequence specificity but at a greater rate and efficiency.



**Figure 1.16** The reactive CI and CBI moieties.

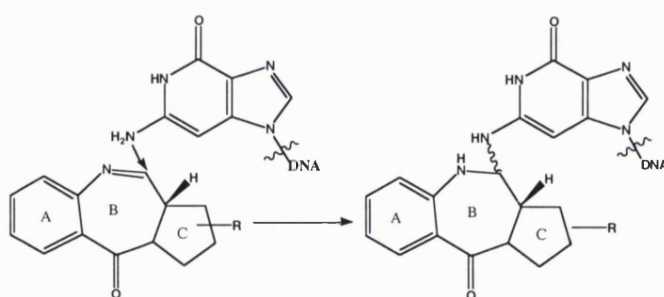
#### 1.7.2.2 The Pyrrolo(1,4)benzodiazepenes

Like (+)-CC-1065, the pyrrolo[2,1][1,4]benzodiazepines (PBDs) are another family of antitumour antibiotics produced by diverse species of *Streptomyces* and include a wide range of compounds exemplified by anthramycin (Leimgruber *et al.*, 1965 a, b) and tomaymycin (Arima *et al.*, 1972) as shown in Figure 1.17. The members of this molecular series share three component rings (A, B, or C), but may differ in the substituent groups attached to their A and C rings, as well as by the degree of saturation of the pyrrole moiety (C), which can be fully saturated or unsaturated at either C2-C3 (endocyclic) or C2 (exocyclic) positions.



**Figure 1.17** The pyrrolo[1,4]benzodiazepine antitumour antibiotics.

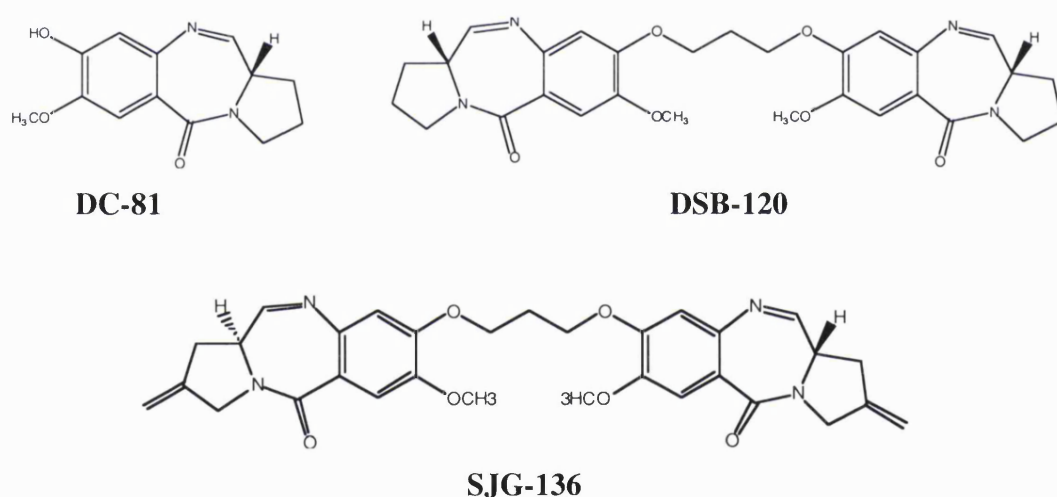
Much work has established that PBDs bind in the minor groove, forming an acid-labile aminal bond between the C11 position of its constituent B ring and the exocyclic N2 position of guanine (as reviewed by Hurley, 1977; Hurley & Thurston 1984; Hurley & Needham-VanDevanter 1986). The chirality of the C11 position produces a right-handed twist in the molecule allowing it to fit within the minor groove, initially non-covalently (Jones 1990). Like (+)-CC-1065, this chiral centre appears to be critical to binding, as racemization of the C11 position results in a significant reduction in the DNA-binding activity and cytotoxicity of PBD analogues (Miyamoto *et al.*, 1977). On insertion into the groove, the chemical group at N10-C11, usually either an imine or a carbinolamine methyl ether, provides an electrophilic centre for the nucleophilic attack of the guanine N2 (Figure 1.18). This results in the production of a covalent adduct at that site typically spanning 3 base pairs, which has been shown by footprinting to occur preferentially at 5'-PuG Pu-3' (Hertzberg *et al.*, 1986).



**Figure 1.18** Formation of a PBD-DNA adduct.

The antitumour activity of PBDs is considered to be a product of their sequence selective interactions with DNA, causing the inhibition of replication and transcription (Neidle *et al.*, 1994). PBDs elicit potent biological effects *in vitro*, inhibiting nucleic acid synthesis and transcription by RNA polymerase (Kohn *et al.*, 1968; Puvvada *et al.*, 1992, 1997). Furthermore, the *in vitro* cytotoxicity of a number of PBDs has been shown to correlate with their relative binding affinities, suggesting that this effect is a result of DNA binding (Puvvada *et al.*, 1993). Thus in an effort to exploit their potential selective anticancer properties, numerous synthetic analogues have been developed (Thurston *et al.*, 1990, 1999, Baraldi *et al.*, 1994), including PBD dimers with varying

linker lengths (Bose *et al.*, 1992a, Thurston *et al.*, 1996). These were synthesised so that a central alkyl linker joined two monomeric DC-81 PBD units, as for example in the PBD dimer DSB-120 (Figure 1.19). Extensive molecular modeling, NMR and kinetic studies established that this dimer produced sequence selective interstrand cross-linking, extending across six base pairs with a reactivity preference for 5'-PuGATCPy-3' or 5'-PuGATCPu-3' sites (Bose *et al.*, 1992b; Mountzouris *et al.*, 1994; Jenkins *et al.*, 1994, Smellie *et al.*, 2003). However, whilst exhibiting potent *in vitro* toxicity it possessed little *in vivo* antitumour activity (Bose *et al.*, 1992a, b; Walton *et al.*, 1996). Nevertheless, a C2/C2' exo-unsaturated analogue, SJG-136, has been developed, which has both *in vitro* and *in vivo* activity (Gregson *et al.*, 1999, 2001; Smellie *et al.*, 2003) and has currently been selected for clinical trials.

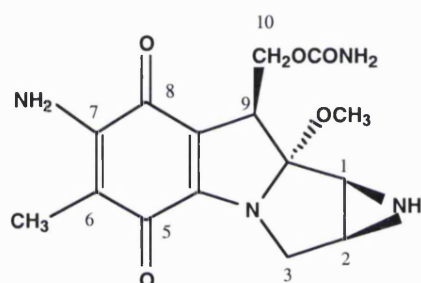


**Figure 1.19** PBD monomer, DC-81, and the derivative dimers, DSB-120 and SJG-136.

#### 1.7.2.3 Mitomycin C

The mitomycins are another class of potent antitumour antibiotics derived from *Streptomyces caespitosus* by Hata and coworkers. Mitomycin C (Figure 1.20) has demonstrated significant antitumour activity and is used extensively in cancer chemotherapy. Like the PBDs, mitomycin C binds in the minor groove, facilitating the

formation of covalent adducts with DNA with a high specificity for the N2-position of guanine (Tomasz *et al.*, 1987). For alkylation to occur, reductive activation of the quinone group to a semi or hydroquinone is required. This may be achieved in cells by the reductive metabolism of mitomycin C by cellular flavoenzymes or mimicked *in vitro* by chemical reducing agents. Alkylation of the DNA can be either monofunctional or bifunctional, producing monoadducts or cross-links, respectively. Furthermore, the cross-links can be either intrastrand or interstrand with alkylation occurring across two neighbouring guanosines at their N2 amino groups via the aziridine at C1 and the carbamate at C10 (Tomasz *et al.*, 1988). These occur with almost absolute selectivity for 5'-CG sequences (Teng *et al.*, 1989).

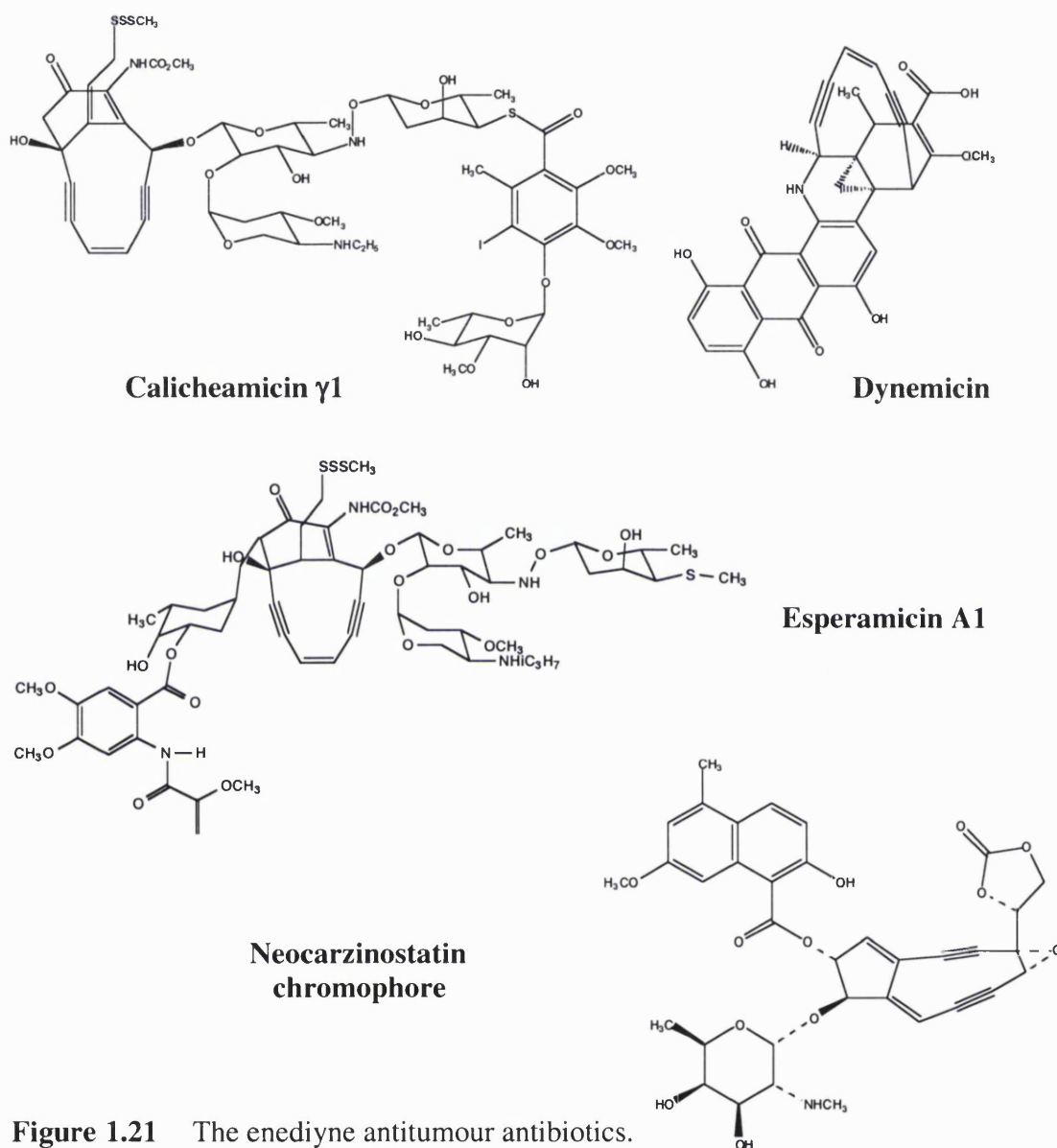


**Figure 1.20** Structure of mitomycin C.

#### 1.7.2.4 The Eneidyne antibiotics

The enediyne antibiotics are a class of extremely potent antitumor agents originally isolated from a variety of *Streptomyces* species. The enediynes include calicheamicin  $\gamma_1$  (Lee *et al.*, 1987a, b), dynemicin (Konishi *et al.*, 1989), esperamicin A1 (Golik *et al.*, 1987a, b) and the closely related neocarzinostatin (Edo *et al.*, 1985) (Figure 1.21). They contain an enediyne core, comprised of two acetylenic groups flanking a double bond within a nine or ten-membered ring chromophore. This may be associated with either a DNA intercalating group, such as the anthraquinone of dynemicin A, or a minor groove-binding function such as the oligosaccharide found in calicheamicin  $\gamma_1$ . Eneidiynes are not directly biologically active but can be considered natural prodrugs, which on

reductive activation undergo cycloaromatisation of the unsaturated enediyne core to produce cytotoxic diyl radicals (Smith & Nicolaou 1996).



**Figure 1.21** The enediyne antitumour antibiotics.

In the case of calicheamicin, its oligosaccharide portion binds tightly in the minor groove, where nucleophilic attack of the allylic trisulfide group by reducing thiols, such as intracellular glutathione triggers the relief of strain at the bridgehead position double bond. This changes the geometry of the molecule to favour Bergman cycloaromatisation, generating extremely reactive benzenoid diradicals in the process

(Bergman 1973). These cause the oxidative cleavage of DNA by removing hydrogen atoms from the deoxyribose sugar on both strands, generating DNA strand breaks. A high proportion of strand scission is as double strand breaks (DSBs), which have long been considered to be the reason for the potent antitumour activity of this class of agents. Indeed recent work has directly demonstrated that calicheamicin  $\gamma$ 1 induces double strand breaks with high efficiency in cellular DNA, producing a strong double strand break response and an associated cell cycle arrest (Elmroth *et al.*, 2003). Thus antitumour activity is mediated in this case by efficient DSB-inducing activity in cells.

The potent cytotoxicity and reactivity of the enediynes has generated much interest, in particular in calicheamicin  $\gamma$ 1 which has been shown to produce sequence-specific strand breaks at homopurine and homopyrimidine tracts such as 5'-TCCT-3' (Zein *et al.*, 1988; Walker *et al.*, 1992) and neocarzinostatin which can cause double strand breaks preferentially at the T residue of a GT step (Dedon & Goldberg 1990; Dedon *et al.*, 1992). Calicheamicin  $\gamma$ 1 and esperamicin A have both now entered clinical trials, however the lack of tumour specificity of these molecules in general has proved problematic in their use as putative anticancer agents. One successful approach has been their conjugation to a cancer cell recognising antibody, producing a new class of anticancer drug (gemtuzamab ozogamicin) recently approved for use in acute myelogenous leukaemia (Sievers 2001). Eneidyne cycloaromatisation can be triggered *in vitro* by introducing strain in the relatively stable molecule, causing production of radical species with a half-life of seconds (Grisson *et al.*, 1996). Thus another therapeutic approach has involved the design of novel triggering mechanisms, to promote diradical cyclisation in cancer cells (Dai *et al.*, 2002). A further strategy has involved redesign of the reactive enediyne core to provide both selective triggering and more discriminating diradical intermediates (Tuntiwechapikul *et al.*, 2002).

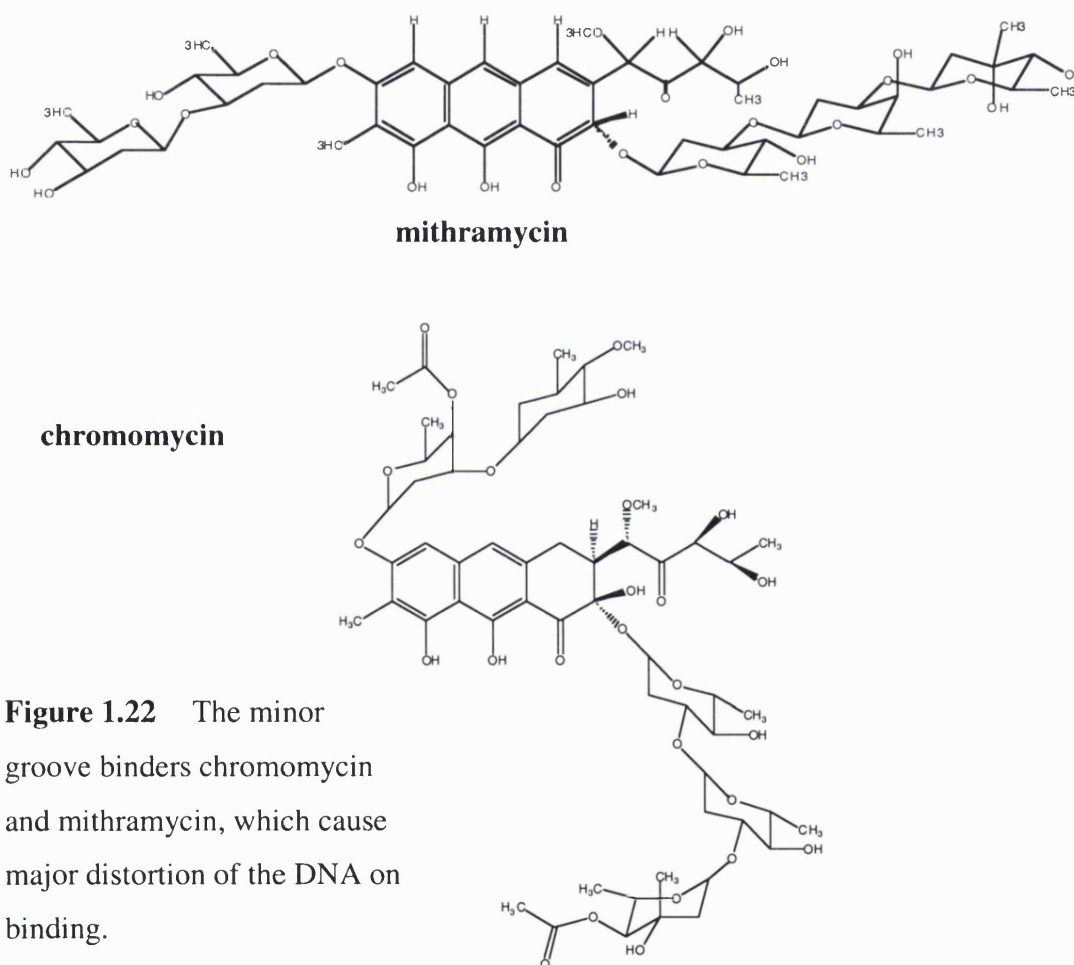
### 1.7.3 Non-covalent interactions: reading the helix via the minor groove.

#### 1.7.3.1 *Chromomycin & Mithramycin*

Chromomycin and mithramycin are members of the aureolic acid group of antibiotics and are composed of a chromophore with a trisaccharide attached to each end of the



molecule (Figure 1.22). Their cytotoxic effect is considered to be due to their interaction with DNA, which occurs via the minor groove. Binding as head-to-tail dimers, interaction in the minor groove causes major structural distortion (Gao *et al.*, 1992; Sastry & Patel 1993), with chromomycin binding at CG rich sequences of at least three base pairs, with a preference for 5'-GGG-3' and 5'-CGA-3' (Fox & Howarth 1985; Van Dyke & Dervan 1983). This may be facilitated by the wider groove associated with GC regions.



**Figure 1.22** The minor groove binders chromomycin and mithramycin, which cause major distortion of the DNA on binding.

### 1.7.3.2 Hoechst 33258 and benzimidazole-based ligands

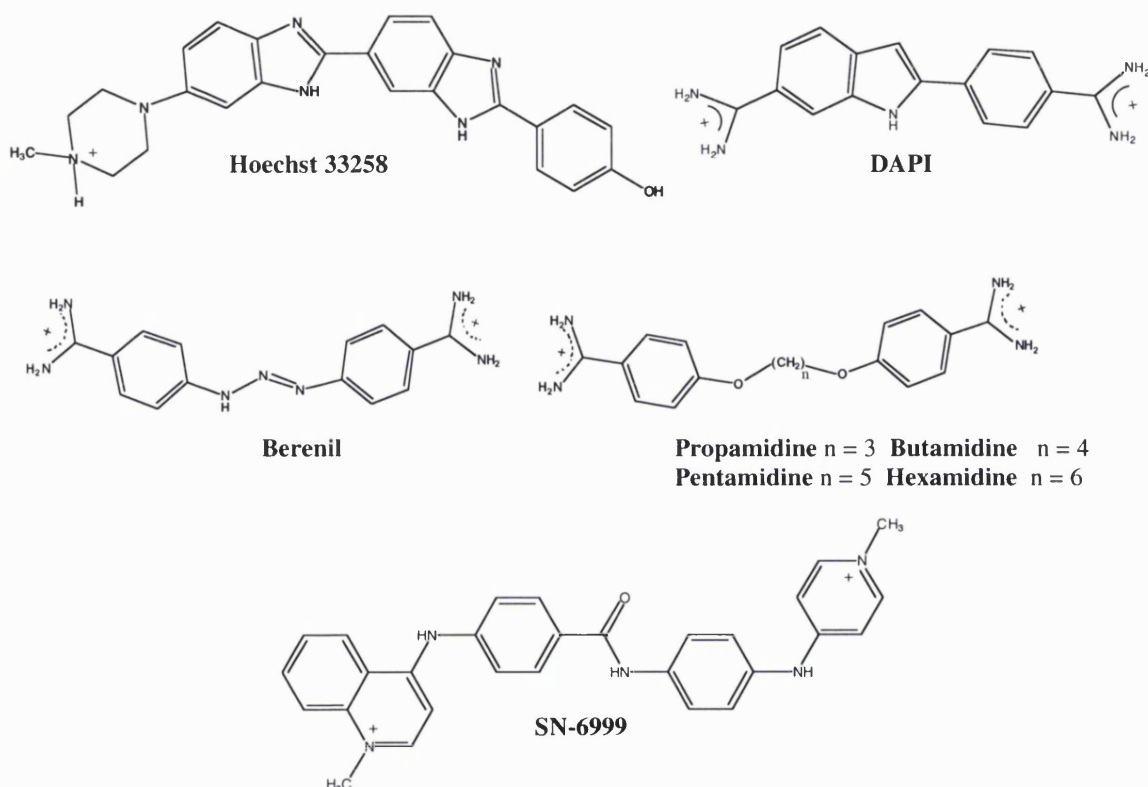
Hoechst 33258 is a crescent shaped molecule comprised of two bis-benzimidazole moieties linked head to tail (Figure 1.23). Originally synthesised as an antihelminthic agent, it is one of a series of compounds generated containing a bis-benzimidazole group

and is now used predominantly as a DNA specific stain. Hoechst 33258 exerts both antimicrobial and antitumour activity, with studies indicating that its cytotoxicity may be as a result of poisoning topoisomerase I by stabilising the cleavable complexes formed by topoisomerase I in a similar manner to etoposide with topoisomerase II (Kim *et al.*, 1996; Sun *et al.*, 1994). X-ray crystallography, NMR and footprinting studies have shown that it binds non-covalently in the minor groove with a preference for AT-rich regions extending over 4-5 base pairs (Teng *et al.*, 1988; Parkinson *et al.*, 1990; Harshman & Dervan 1985; Portugal & Waring 1988). Hoechst 33258 displays a sensitivity to variation of (A/T)<sub>4</sub> sequence, demonstrating a preference for AATT (Abu-Daya *et al.*, 1995). CD studies have indicated that it may also bind weakly to GC regions, partially intercalating at GCGC sequences (Bailly *et al.*, 1992). However, predominant DNA binding occurs via a combination of hydrogen bonding to the edges of AT base pairs facing into the minor groove, together with extensive van der Waals interactions with the non-polar DNA backbone. The molecule fits within the minor groove positioning the two NH groups of the bis-benzimidazoles towards the groove floor. This orientation facilitates hydrogen bonding, with the adjacent hydrogen bond donors, adenine N-3 and thymine O-2, of AT base pairs (Teng *et al.*, 1988; Carrondo *et al.*, 1989; Parkinson *et al.*, 1990; Fede *et al.*, 1991; Embrey *et al.*, 1993; Quintana *et al.*, 1991; Vega *et al.*, 1994; Spink *et al.*, 1994). The contribution of these molecular forces to the overall binding affinity remains controversial, however, with some evidence that the hydrophobic transfer of the ligand from solution to its binding site may provide the energy for complex formation (Haq *et al.*, 1997).

Groove width plays an important role in the binding of bis-benzimidazoles, with sequence specificity determined by the accessibility of binding sites, and the van der Waals forces and hydrophobic interactions associated with them rather than hydrogen bonding (Czarny *et al.*, 1995; Clark *et al.*, 1996a). This appears to be borne out by a recent NMR study which demonstrated that Hoechst 33258 could bind to DNA in a co-operative 2:1 complex, fitting two parallel molecules aligned head to head in the groove width (Gavathiotis *et al.*, 2000). This head to head alignment bound to the self complementary oligonucleotide duplex d(CTTTTGCAAAAG)<sub>2</sub>, with the planar phenyl



and benzimidazoles accommodated by the narrow poly dT tract, fitting tightly into the groove with strong van der Waals interactions. The bulky piperazine groups, which are adjacent to each other in this 2:1 complex, were positioned alongside the central 5'-GC-3' possibly as a consequence of the increased width at the GC base pairs.

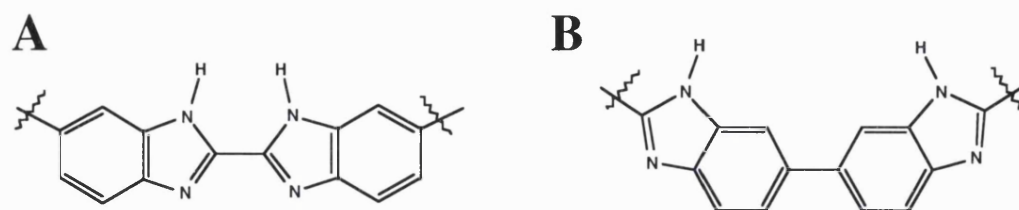


**Figure 1.23** Structures of several small non-covalent minor groove-binding agents.

A large number of analogues based on the bis-benzimidazole structure have been synthesised, in an effort to enhance DNA binding and targeted cytotoxicity. Many of these have been composed of two or three benzimidazole units linked head to tail as in Hoechst 33258, with a positively charged terminal group. These have, for example, modified the side chains of Hoechst, by replacing the terminal piperazine chain with an amidinium, an imidazoline or a tetrahydropyridinium group (Wood *et al.*, 1995; Clark *et al.*, 1997; Bostock-Smith *et al.*, 1998). Exchange of the bulky piperazine ring for these smaller moieties has resulted in changing the selectivity of the ligand along AT stretches, favouring narrower groove widths at the 3' end of the binding site. In

contrast, relocation of the hydroxyl group on the phenol ring of Hoechst 33258 from para to the meta position, was found not to improve ligand selectivity significantly (Ebrahimi *et al.*, 1992; Clark *et al.*, 1996a). Replacement of the benzimidazole units with other heterocycles including benzoxazole or pyridoimidazole has altered the selectivity to introduce some GC tolerance (Bathini *et al.*, 1990; Rao *et al.*, 1991; Kumar *et al.*, 1990). The tris-benzimidazole analogues introduced a further hydrogen bond acceptor (Kim *et al.*, 1996; Clark *et al.*, 1996b; Ji *et al.*, 2001). These analogues have an increased coverage, extending over 5 to 7.5 base pairs relative to the four to five spanned by bis-benzimidazoles. More recently a novel aryl-bis-benzimidazole amino acid analogue of Hoechst has been designed for incorporation into peptide combinatorial libraries (Bunkenborg *et al.*, 2002). This replaces the N-methyl piperazine group with a carboxyl group and the hydroxy group of the terminal phenol ring with an amino-methyl group to produce a molecule that may subsequently be extended using peptide linkages.

Head to head linkage of two bis-benzimidazoles via the imidazole ring (Figure 1.24 A) using a methylene chain up to four carbons in length has produced a number of symmetrical dicationic molecules that bind strongly to AT-rich tracts via the minor groove (Lombardy *et al.*, 1996). A number of these alkyl-linked benzimidazoles have demonstrated increased anti-*Pneumocystis carinii* activity, an organism associated with AIDS related infections, with decreased toxicity relative to the usual course of therapy, pentamidine. Recently an alternative series of head to head linked bis-benzimidazoles have been synthesised, linking two units via the benzene rings (Figure 1.24 B) (Mann *et al.*, 2001). This has produced a molecule that exhibits strong antitumour activity, especially against ovarian carcinomas, providing a structural basis for the rational development of tumour-active Hoechst analogues. Other strongly cytotoxic Hoechst analogues have included use of a covalent binding moiety via the introduction of a terminal carbamate with an attached chloroalkyl or bromoalkyl group (Bielawski *et al.*, 2001).



**Figure 1.24** Head to head linked benzimidazole structures.

#### 1.7.3.4 DAPI

DAPI (4', 6-diamidino-2-phenylindole) is a synthetic antibiotic extensively used as a fluorescent stain for chromosomal DNA (Figure 1.23). It has shown potent cytotoxic activity against a number of parasites; a property attributed to its high binding affinity for DNA disrupting its metabolism. Footprinting and structural studies have demonstrated that it binds in the minor groove of DNA at sites of three or more AT base pairs oriented with the phenyl-amidinine at the 5' end (Portugal & Waring 1988; Wilson *et al.*, 1990). Selective binding at a T: T mismatch has also been observed in the minor groove (Trotta & Paci 1998). The phenyl and indole rings are coplanar, fitting neatly into the narrow groove width associated with poly dAT runs. DAPI has been shown to inhibit the activity of RNA polymerase II, possibly as a function of its high binding affinity for A: T sequences in the minor groove. This may prevent the binding of TBP (TATA box binding protein) to its consensus sequence, inhibiting the recruitment of the transcription factors required by RNA polymerase II to initiate transcription (Chiang *et al.*, 1994). DAPI can also disrupt the activity of a number of other DNA processing enzymes associated with regulatory and structural functions, including DNA ligase, topoisomerases, exonuclease III and DNA polymerase (Parolin *et al.*, 1990). However, the wide range and varying levels of inhibition produced by DAPI cannot be attributed to AT-specific minor groove binding alone.

Although minor groove interaction at A: T sites is the preferred mode of binding of DAPI to double-stranded DNA, other modes of binding at GC or non-consecutive AT base pairs have been observed. For example, intercalation between A: U base pairs is favoured with double stranded RNA due to the decreased capability for stabilising van der Waals interactions in the wider and shallower minor groove width produced by these

base pairs (Tanious *et al.*, 1992), whilst major groove binding has been observed at poly [d(GC)<sub>2</sub>] sites (Kim *et al.*, 1993). DAPI may also intercalate with DNA at GC rich sequences (Kim *et al.*, 1993; Jansen *et al.*, 1993; Trotta *et al.*, 1995), with the guanine-N2 amino group preventing deep insertion of the phenyl ring of DAPI into the minor groove at GC sites (Waring & Bailly 1997). Furthermore, a 2:1 external ( $\pi,\pi$ ) stacking motif in the minor groove has been characterised using NMR, with the position and orientation of the ligand resembling those for intercalation of DAPI at G: C base pairs (Trotta *et al.*, 1996).

DAPI is a small molecule, which on binding to DNA produces little distortion in the local structure. Recent work, however, has indicated that its binding may effect the organisation of longer DNA segments, by inhibiting the assembly of inherently curved sections of the DNA into nucleosomes (Fitzgerald & Anderson 1999). Inhibition of nucleosome formation by DAPI was attributed to its selective binding at the numerous oligo A/T clusters that produce curved sections of DNA sequence. These structures are relatively uncommon in genomic DNA, but are found to occur in the control regions for DNA processing events such as transcription, replication and recombination. Whilst other small minor groove binding ligands such as Hoechst 33258 and distamycin have also been examined for this activity, DAPI was found to be a stronger inhibitor of nucleosomal assembly, possibly as it was more effective at stiffening the DNA, counteracting its intrinsic curvature and bendability (Albert *et al.*, 1999). Furthermore, DAPI has been shown to influence secondary structure, promoting the formation of a stable hairpin in the trinucleotide repeat sequence (ATT)<sub>6</sub> (Trotta *et al.*, 2000). NMR and molecular modelling studies indicated that the formation of DNA hairpins was induced by two DAPI molecules binding in the minor groove of this sequence. Trinucleotide repeat sequences are associated with a number of genetic disorders and therefore selective binding and identification of these regions may be of future therapeutic value.

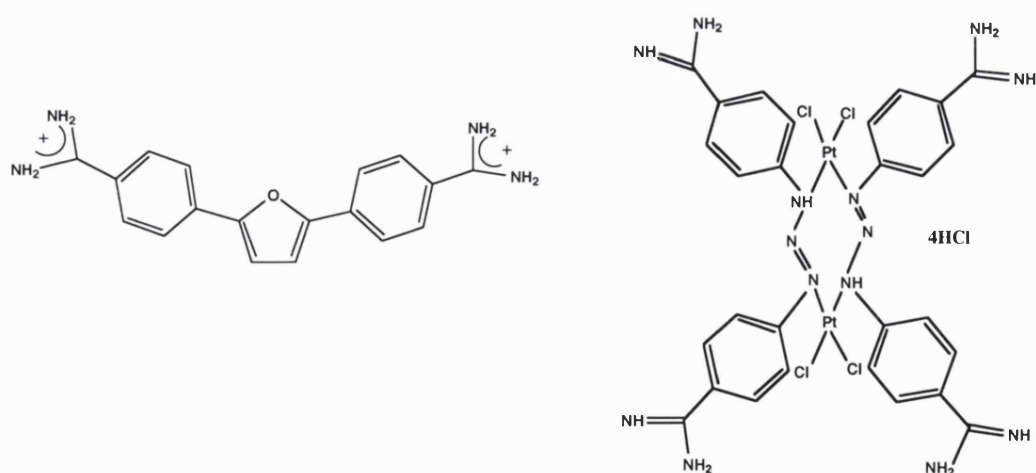
#### 1.7.3.5 Berenil

Berenil (1,3-bis(4-amidinophenyl)triazene) is an antitrypanosomal agent which exhibits weak cytotoxic and antiviral properties (Clercq & Dann 1980) (Figure 1.23). It is a potent inhibitor of viral, trypanosomal and mammalian topoisomerases (Fairfield *et al.*, 1979; Shaffer & Traktman 1987; Sekiguchi & Kmiec 1988; Portugal 1994) and has been shown to inhibit plasmid replication as a result of DNA binding (Coates *et al.*, 2002). It is a dicationic molecule comprised of two terminally substituted phenyl rings linked by a triazene spacer. Footprinting, NMR and crystallographic studies have shown that it binds predominantly in the minor groove of DNA in a 1:1 stoichiometry, with specificity for AT rich regions, binding both alternating AT and oligo(A) tracts containing at least three AT base pairs (Zimmer & Wahnert 1986; Portugal & Waring 1986, 1987; Yoshida *et al.*, 1990; Brown *et al.*, 1990, 1992; Lane *et al.*, 1991; Jenkins *et al.*, 1993). Its concave shape means that it is isohelical with the DNA, slotting along the minor groove floor forming hydrogen bonds from its charged amidinium groups to either the adenine N3 or thymine O2 positions at either end of the binding site (Brown *et al.*, 1990, 1992; Lane *et al.*, 1991). However, due to the small size of the molecule and thus its limited ability to follow the curvature of the helix, the interaction at the 5' end is indirect requiring a water molecule to bridge the gap between the base edge and the amidinium group (Brown *et al.*, 1990).

Thermodynamic and UV spectroscopic studies identified that at low ligand to DNA concentrations berenil also intercalated with the DNA (Schmitz & Hubner 1993). CD and NMR data later indicated that like DAPI, berenil intercalates with both DNA and RNA duplexes and triplexes, with selective binding following a hierarchy such that non-intercalative binding at poly(dA).poly(dT) is favoured over mixed binding modes at poly[d(A-T)]<sub>2</sub> and poly(rA).poly(rU), with minor groove binding requiring the presence of at least one 5'ApA-3' dinucleotide step (Pilch *et al.*, 1995). Also like DAPI, binding by berenil at AT clusters stiffens the DNA duplex at these regions, which mediate helix bending and possibly gene regulation by this process (Reinert 1999). Berenil binds AT-rich regions strongly but not uniquely, intercalating at GC-rich sites (Pilch *et al.*, 1995; Colson *et al.*, 1995; Waring & Bailly 1997; Barcelo *et al.*, 2001). However, replacement

of guanines with inosines, which lack the protruding exocyclic C2-amino of guanine, restores groove binding indicating that like ligands such as distamycin, DAPI and Hoechst 33258, this moiety is of critical importance to selectivity of berenil in the minor groove (Pilch *et al.*, 1995; Waring & Bailly 1997).

A number of berenil derivatives have been designed to enhance or alter the DNA binding characteristics of the parent molecule (Figure 1.25). For example, a furan based analogue, furamidine (2,5-Bis(4-guanyphenyl)furan), was synthesised replacing the central triazene linker with a furan moiety (Boykin *et al.*, 1995). It was found to bind more tightly than berenil to the oligonucleotide d(CGCGAATTCGCG)<sub>2</sub> due to additional non-bonded interactions between the furan group and the groove walls, making direct contact at both ends of the molecule via hydrogen bonds from the terminal amidinium groups and the O2 atoms of the thymine bases, (Laughton *et al.*, 1996; Trent *et al.*, 1996). Crystal structures have also been obtained for several 1,3-diaryltriazene analogues of berenil designed for the selective targeting of GC base pairs (Walton *et al.*, 1991). Furthermore, a platinated analogue of berenil (Pt-berenil), whereby two berenil molecules have been linked using cis-Pt(II) centres, has been shown to form interstrand crosslinks with DNA binding directed by the berenil moieties to AT rather than GC regions (Gonzalez *et al.*, 1996; 1999).



**Figure 1.25** Furamidine and Pt-berenil.

#### 1.7.3.6 Pentamidine and related bis(amindinophenoxy)-alkane diamidines

The bisbenzamidine compound series known as the bis(amindinophenoxy)-alkane diamidines (Figure 1.23) exhibit potent anti-protozoal activity and have been used to treat a wide range of fungal diseases including *Pneumocystis carinii* pneumonia. Whilst their mode of action has not been established, current evidence suggests that DNA binding may play a role in their biological activity although no direct correlation has been observed between their DNA binding properties and their activity against *P. carinii*. Footprinting, X-ray crystallography, NMR and molecular modelling studies have demonstrated that pentamidine and a number of structural derivatives bind to DNA in the minor groove in a sequence selective manner favouring AT rich sites (Waring 1970; Williamson 1979; Fox *et al.*, 1990; Cory *et al.*, 1992; Edwards *et al.*, 1992; Jenkins *et al.*, 1993). Molecular mechanics and footprinting studies show that for pentamidine, this occurs in the minor groove at sites of at least five consecutive AT with the molecule spanning over 4-5 base pairs such that the charged amidinium groups at either end may hydrogen bond with thymine O2 and adenine N3 positions in the DNA (Sansom *et al.*, 1990, Fox *et al.*, 1990). Sites composed of three or less AT base pairs were found not to be favoured by pentamidine (Fox *et al.*, 1990) or even by the shorter derivatives butamidine or propamidine (Bailly *et al.*, 1997). Also, little tolerance was observed for GC base pairs within putative binding sites. This was shown by footprinting studies examining binding to sequences comprised of the synthetic bases inosine and 2, 6-diaminopurine replacing guanine and adenine respectively which indicated that binding was prohibited due to steric hindrance caused by the guanine C2-amino group protruding into the minor groove (Bailly *et al.*, 1997).

The length of the alkyl chain linking benzamidine units has been shown to influence groove binding of pentamidine derivatives. Molecular mechanics calculations have indicated that analogues containing three and five linker methylenes run isohelical with the DNA curvature promoting strong DNA binding, whilst those with two, four or six produce more linear structures less favourable for groove-binding (Cory *et al.*, 1992). Later footprinting studies corroborated this, observing that bisbenzamidines with three and five carbon methylene linkers bound with greater selectivity than those with four or

seven to nine (Bailly *et al.*, 1997). Footprinting analysis of cis and trans-linked butamidine analogues demonstrated that the conformation of the linker was also an important factor in DNA binding, with cis-linked butamidine isomers able to bind more strongly than those linked in trans, despite comparable hydrogen bonding capacities (Bailly *et al.*, 1994). Thus in common with other minor groove binding compounds, the conformation of the ligand combined with the local DNA structure play important roles in DNA interactive binding.

#### 1.7.3.7 SN series

The SN series is a family of structurally related synthetic compounds comprised of bisquaternary ammonium heterocycles (bQAHs), which have been shown to have potent antitumour and antiviral activities *in vivo* and *in vitro* (Cain *et al.*, 1969; Robertson & Baguley 1982). These compounds bind in the minor groove, demonstrating a preference for four or more consecutive dA.dT base pairs (reviewed extensively by Reddy *et al.*, 1999) with many eliciting a strong antitumour effect in the murine leukaemia cell line L1210 (Denny *et al.*, 1979).

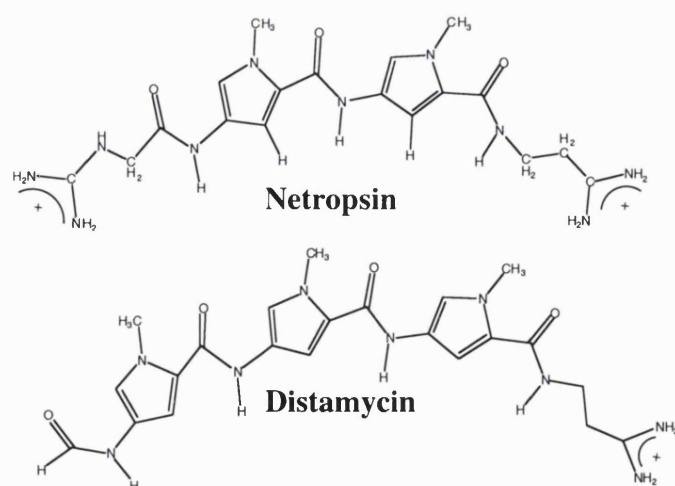
Whilst a great many bisquaternary heterocycles have been synthesised, the DNA binding properties of the bQAH SN-6999 in particular, an agent that has demonstrated strong inhibition of DNA polymerase activity *in vitro* (Kittler *et al.*, 1996, Figure 1.23), has been extensively characterised. CD, UV absorption, thermal melting and viscosity studies have indicated that SN-6999 is highly selective for AT base pairs with a specificity comparable to the naturally occurring antitumour antibiotic netropsin (Luck *et al.*, 1987). Furthermore, NMR and X-ray crystallography have shown that binding occurs in the minor groove of DNA over five consecutive dA.dT base pairs (Leupin *et al.*, 1986; Chen *et al.*, 1992; Gao *et al.*, 1993). Common to other minor groove-binding agents, this is stabilised by a combination of molecular forces, including the electrostatic attraction of the positively charged quaternary ammonium heterocycles for the negatively charged DNA, van der Waals interactions with the groove walls and the formation of a hydrogen bond between the amide NH of SN-6999 and the thymine-O2 position (Gao *et al.*, 1993). However, unlike comparable minor groove binders the SN-6999 molecule is



less curved, resulting in a straightening of the DNA helix to accommodate its binding (Gao *et al.*, 1993). NMR studies have also shown that SN-6999 can exhibit an orientation preference for binding with some DNA sequences, for example the duplex d(GGGAAAACGG).d(CCGTTTTTCCC), preferentially placing the pyridinium ring at the 5' end of the (+) strand (Rydzewski *et al.*, 1996). This may be due to increased van der Waals interactions with the groove walls as a result of the improved shape complementarity of the ligand in that orientation and the DNA, thereby indicating that the structural properties of the DNA sequence itself as well as those of the ligand can profoundly influence the characteristics of binding.

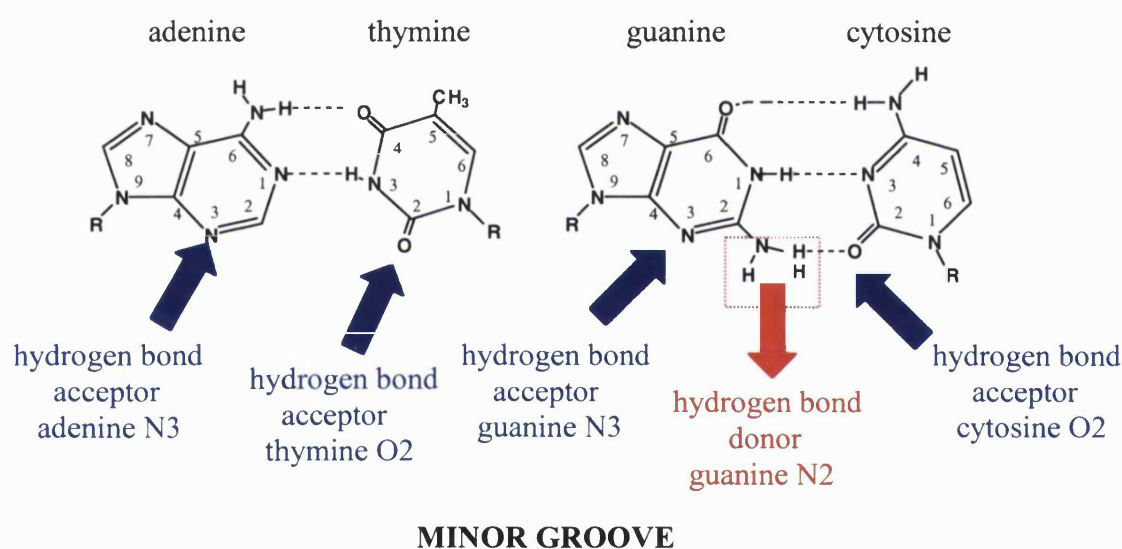
#### 1.7.3.8 Netropsin and distamycin

Netropsin and distamycin are naturally occurring antibiotics originally derived from the *Streptomyces* species *S. netropsis* (Finlay *et al.*, 1951) and *S. distamycetes* (as reviewed by Arcamone 1993), respectively. They are comprised of two and three pyrrole moieties with either an N-terminal guanidinium or formyl group and a C-terminal propylamide (Figure 1.26). These compounds have shown potent biological effects exhibiting antitumour and antiviral activities (Arcamone 1993). This may be as a consequence of an ability to interfere with the replication and transcription of DNA, as netropsin has been shown to inhibit template activity in *in vitro* DNA and RNA polymerase systems (Zimmer *et al.*, 1971; Chandra *et al.*, 1972).



**Figure 1.26** The minor groove binding antibiotics netropsin and distamycin.

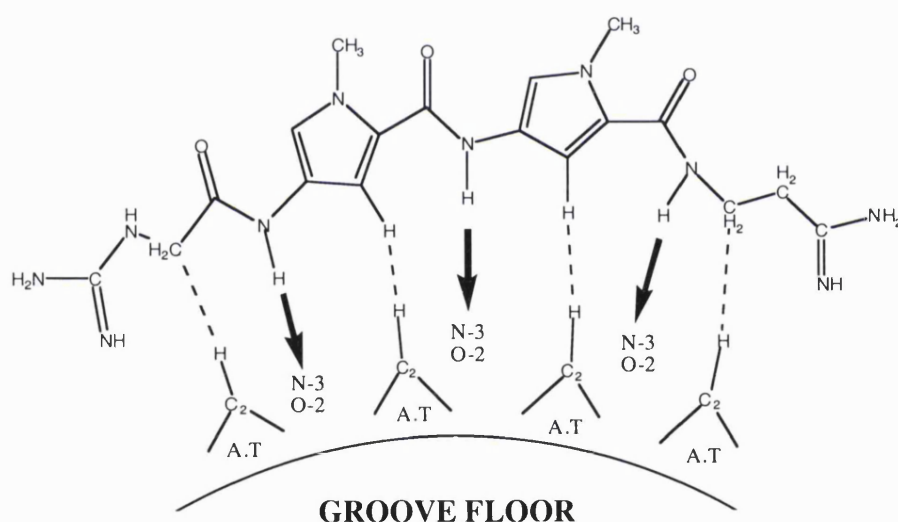
Initial studies of the interactions of netropsin with the simple polymers polydA.polydT and polyd(A.T) demonstrated a strong selectivity for AT base pairs (Zasedatelev *et al.*, 1974; Zimmer *et al.*, 1979; Breslauer *et al.*, 1987). Binding was thought to be limited to AT tracts due to the intrusion of the guanine 2-amino group in the minor groove causing a steric clash with the N-methylpyrrole C3 hydrogen, preventing binding to GC-rich regions (Wartell *et al.*, 1974) (Figure 1.27). Various footprinting methodologies, examining a wide variety of AT-rich sequences, confirmed the AT preference and further demonstrated that netropsin and distamycin targeted four and five base pairs, respectively (Van Dyke *et al.*, 1982; Lane *et al.*, 1983; Fox & Waring 1984; Portugal & Waring 1987) with preferential binding at 5'-AATT-3' and 5'-AAATT-3' over alternating AT sequences (Kopka *et al.*, 1985a, b; Coll *et al.*, 1987, 1989).



**Figure 1.27** A schematic diagram showing the intrusion of the guanine 2-amino group into the minor groove, as indicated by the dashed box. The available hydrogen bond donors and acceptors in the minor groove are also detailed.

A more detailed mechanism of binding was derived from crystallographic and NMR studies of netropsin (Coll *et al.*, 1989; Kopka *et al.*, 1985a, b, 1986) and distamycin (Coll *et al.*, 1987; Klevit *et al.*, 1986, Pelton & Wemmer 1988). As cationic molecules, binding was initiated by their electrostatic attraction to the negatively charged phosphate

backbone of DNA. Subsequently, their crescent-shaped geometry and planar molecular structure enabled them to follow the curvature of the DNA helix, fitting tightly into the narrow minor groove (Berman *et al.*, 1979; Patel 1982; Kopka *et al.*, 1983, 1985a, b). This placed the cationic ends of the molecules deep within the groove, an area of high negative electrostatic potential, orienting the constituent amide groups of the ligand towards the groove floor and allowing close contact between the pyrrole rings and the base edges. Here the amide hydrogens of N-methylpyrrolecarboxamides formed bifurcated hydrogen bonds with N3 of adenine and O2 of thymine of the AT base pairs in the minor groove, providing the correct orientation for strong van der Waals interactions between the ligand and the groove walls (Figure 1.28). Thus the pyrrole rings filled the groove with their hydrogens positioned too deeply to allow room for guanine NH<sub>2</sub> of a GC base pair.



**Figure 1.28** Bonding components contributing to the sequence-specific binding of netropsin to (AT)<sub>4</sub>, as adapted from Bailly & Chaires 1998. Heavy arrows indicate hydrogen bonding from donor to acceptor, whilst dashed lines indicate close van der Waals non-bonded contacts between the ligand and the DNA.

Dickerson and co-workers suggested that these non-bonded van der Waals interactions between the ligand methylenes and the chemical group on the C2 position of either the purine or pyrimidine base pairs, hydrogen in the case of adenine and amino with

guanine provided the structural information for sequence selective recognition, with the hydrogen bonding providing the appropriate reading frame for this to occur (Kopka *et al.*, 1985a). Thermodynamic studies later indicated that groove binding might be driven by a number of enthalpic factors; with a combination of van der Waals contacts, electrostatic interactions, hydrogen bonding and the displacement of water molecules from the minor groove providing the binding energy for this process (Marky *et al.*, 1987).

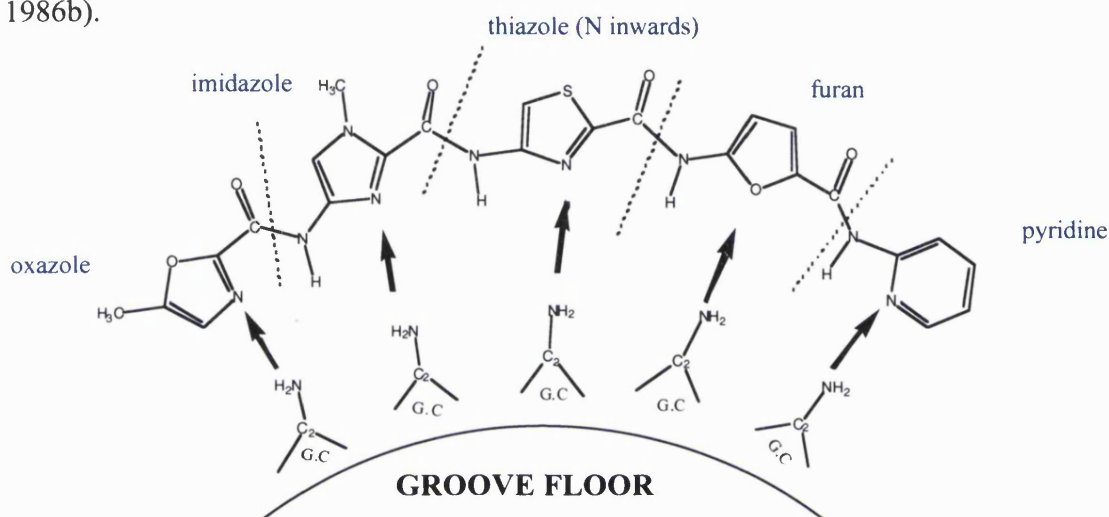
## **1.8 The development of polyamides or “lexitropsins”**

In recent years a number of structural derivatives of netropsin and distamycin have been synthesised in an effort to capitalise on their strongly selective minor groove binding properties. Modification of their inherent specificity for AT base pairs to GC was proposed to develop ligands capable of selective binding to DNA sequences with a mixed base pair content (Kopka *et al.*, 1985b; Lown *et al.*, 1986b). This has led to the development of information-reading oligopeptides or “lexitropsins”, currently described as polyamides due to their repeating peptide linkages. These molecules provide an alternative gene targeting approach, enabling DNA sequence discrimination with an equivalent binding affinity to native DNA-binding control proteins, raising the possibility of the artificial regulation of gene expression as a consequence (Trauger *et al.*, 1996a). Recent work has demonstrated that linked polyamides can modulate transcription of specific genes *in vitro* and in some cell culture situations (Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998, 1999a & b; Mapp *et al.*, 2000; Chiang *et al.*, 2000).

### **1.8.1 Introduction of a GC recognition element**

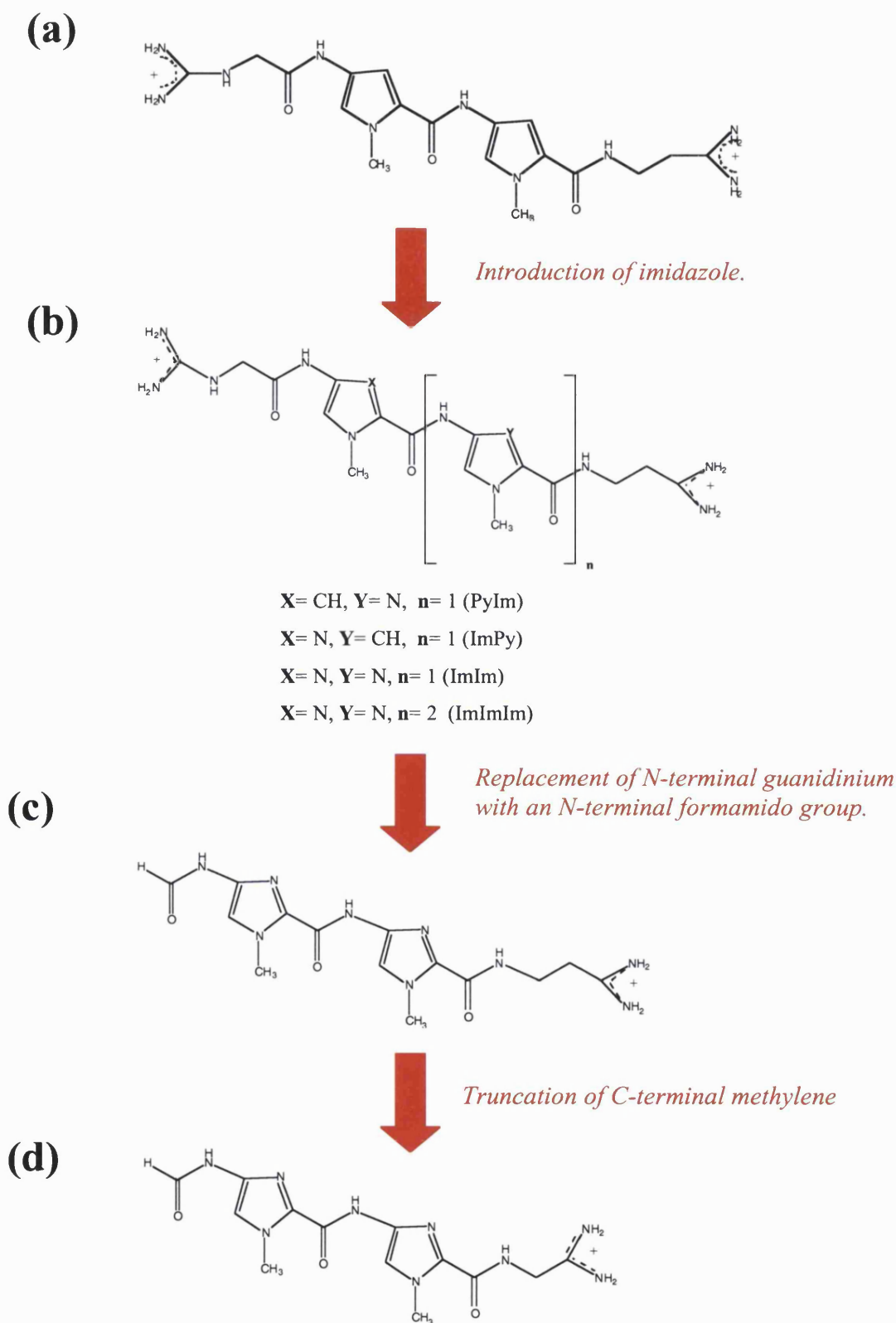
Structural analysis of netropsin:DNA complexes indicated that GC recognition might be achieved by the substitution of one or more of the pyrrole rings that constitute netropsin with an alternative heterocycle able to accommodate the amino group of guanine which protrudes into the minor groove (Kopka *et al.*, 1985b, Lown, *et al.*, 1986b). To this end, a number of aromatic heterocycles were synthesised, including imidazole, pyridine, furan, oxazole and thiazole with the N atom directed into the groove (Figure 1.29). It was considered that imidazole was a favourable candidate for GC recognition as it

would be able to fit neatly alongside guanine in the groove, favouring the formation of a hydrogen bond between the guanine N2-amino and the imidazole nitrogen (Kopka *et al.*, 1985b, Lown *et al.*, 1986b). Thus imidazole was used to replace the guanine-excluding pyrrole ring, in a series of dicationic netropsin-based analogues as shown in Figure 1.30 (b) (Lown 1986b). Whilst these analogues exhibited an increased acceptance of GC base pairs, footprinting studies indicated that they nevertheless retained some attraction for AT regions although this was reduced relative to distamycin and netropsin (Lown *et al.*, 1986b).



**Figure 1.29** Representation of putative GC recognition heterocycles adapted from Bailly & Chaires 1998. Heavy arrows indicate hydrogen bonding.

In order to decrease further the attraction for AT stretches, which have a strong negative electrostatic potential (Lavery & Pullman 1985), a series of two ringed imidazole-containing analogues were designed based more closely on the structure of distamycin (Krowicki & Lown 1987). These replaced the cationic N-terminal guanidinium head group with a non-polar N-terminal formyl, resulting in a monocationic molecule (Figure 1.30 c). Footprinting experiments with these analogues demonstrated that systematic replacement of the N-methylpyrrole groups with imidazole produced an increased selectivity for regions with significant GC content (Kissinger *et al.*, 1987). However, NMR studies later showed that a monocationic diimidazole polyamide, bound over the



**Figure 1.30** The development of GC recognition, showing the evolution of (a) netropsin into (b) a dicationic mono, di and triimidazole (c) a monocationic formamido diimidazole, (d) a truncated C-terminal formamido diimidazole.



5'-CCAT-3' of the oligonucleotide sequence d[CATGGCCATG]<sub>2</sub>, with the N-terminus oriented at the 5' end and the C-terminus located above the 3' thymine (Lee *et al.*, 1988a). Alignment at the C-terminus over the 3' thymine rather than over GC base pairs was attributed to a steric clash between the hydrogens atoms in the C-terminal methylene chain and the guanine2-NH<sub>2</sub>, resulting in an avoidance of a 3' GC base pair (Figure 1.30).

Alteration of the methylene chain enabled GC recognition at the C-terminus, suggesting that recognition of sequence at the 3' base pair was dictated by the van der Waals interactions between the C-terminal methylenes and the DNA (Lee *et al.*, 1988b). This demonstrated that whilst a level of GC tolerance could be provided by the introduction of an imidazole moiety, its use alone was insufficient to provide complete GC discrimination, with other structural factors also acting as important determinants of recognition.

### 1.8.2 Groove width and 2:1 binding

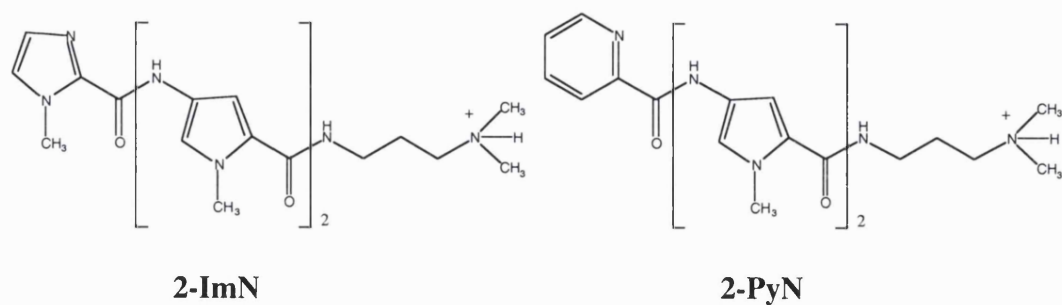
The discovery by NMR that distamycin could bind cooperatively as an antiparallel 2:1 complex within the minor groove opened up the possibility of enhancing the selectivity of polyamides by reading the structural information along both strands of the DNA simultaneously (Pelton & Wemmer 1989). This study indicated that whilst at low ligand concentrations 1:1 binding predominated, at higher concentrations two distamycin molecules were able to bind side-by-side within the minor groove. The two molecules stacked one on another, ring-on-amide so that they overlapped, with each molecule contacting one of the DNA strands. The 2:1 complex showed a strong orientational preference, with the N-terminal formyl head group aligned at the 5' end of the contacted strand. In contrast, netropsin did not form a 2:1 complex strongly favouring a 1:1 binding motif, possibly due to repulsion of the positive charges at both ends of the molecule.

Structural analysis of this dimeric binding mode demonstrated that the minor groove expanded to accommodate two parallel distamycin molecules (Pelton & Wemmer

1990). This concurred with similar studies on other minor groove binding ligands, which also exhibited a 2:1 binding mode, including actinomycin D and chromomycin (Scott *et al.*, 1988; Gao & Patel 1989; Wang *et al.*, 1989). Groove width was found to exert a profound influence on the binding characteristics of distamycin in both 1:1 and 2:1 binding modes, with preferred binding observed at polydA.polydT stretches over TA steps possibly due to the smaller amount of energy required to expand a narrow groove at these sites compared to narrowing the wider groove width found at TA steps (Fagan & Wemmer 1992). Therefore the groove width associated with particular base pair sequences strongly influenced the pattern of distamycin binding.

The potential for dimeric binding explained the unanticipated binding preferences observed for two previously synthesised netropsin-based analogues, pyridine-2-carboxamido-netropsin (2-PyN) and 1-methylimidazole-2-carboxamido-netropsin (2-ImN). These molecules consisted of either an N-terminal pyridine or imidazole ring followed by two pyrrole rings to target them to the sequence 5'-TGTTA-3' (Wade & Dervan 1987) (Figure 1.31). However, footprinting and affinity cleavage experiments, showed that rather than targeting this sequence, 2-PyN bound preferentially at two distinct sites 5'-TTTTT-3' and 5'-TGTC A-3', whilst 2-ImN strongly favoured 5'-TGTC A-3' (Wade & Dervan 1987, Wade *et al.*, 1992, 1993). Later structural characterisation of their binding revealed that like distamycin, these molecules could also bind cooperatively in an antiparallel 2:1 motif in the minor groove. The formation of a cooperative complex aligning the two GC recognition elements at each end of the dimer enabled recognition of two GC base pairs, explaining the preference for the 5'-TGTC A-3' over the anticipated 1:1 target 5'-TGTTA-3', with formation of a hydrogen bond between the imidazole N3 or pyridine N atom and the amino group of guanine (Mrksich *et al.*, 1992; Dwyer *et al.*, 1992). Thus in the 1:1 binding motif, 2-PyN had favoured the narrower 5'-TTTTT-3' whereas the wider mixed sequence 5'-TGTC A-3' facilitated 2:1 binding (Wade *et al.*, 1992).





**Figure 1.31** 1-methylimidazole-2-carboxamidonetropsin (2-ImN) and pyridine-2-carboxamidonetropsin (2-PyN).

Many studies have indicated that in a 1:1 complex the specificity of a polyamide is limited to differentiating between AT and GC base pairs, with indiscriminate recognition at base pair inversions such as AT versus TA. However, during 2:1 binding each polypeptide ligand is aligned to an adjacent DNA strand so that each of the constituent heterocycle rings contacts a single base, instead of the centre of the base pair, enabling potential sequence recognition along both strands. Unfortunately, as the pyrrole ring has a similar affinity for A and T, parallel pyrrole pairs (Py/Py) in a 2:1 binding mode are unable to break the degenerate discrimination of AT and TA base pairs (White *et al.*, 1996). In contrast, the imidazole heterocycle favours guanine and thus the pairing of an imidazole ring opposite pyrrole (Im/Py) provides specific GC recognition. This is attributed to the asymmetric arrangement of the guanine NH<sub>2</sub> hydrogens protruding into the groove, preferentially hydrogen bonding with an imidazole ring on the guanine side of the base pair rather than the cytosine side (Kopka *et al.*, 1997). Thus an Im/Py and Py/Im pair targets GC and CG, respectively (Wade *et al.*, 1992; Mrksich *et al.*, 1992; Wade *et al.*, 1993). Numerous footprinting and structural studies examining binding to a wide variety of sequences support the generality of these observations and have led to the establishment of a set of pairing rules defining the optimal side-by-side pairings of imidazole and pyrrole rings for binding at predetermined DNA sites as summarised in Table 1.2 (Mrksich *et al.*, 1993a; Geierstanger *et al.*, 1994 a & b; Mrksich *et al.*, 1995; Kielkopf *et al.*, 1998a, b).

Nevertheless, work continues in elucidating new heterocycle pairings with potentially improved recognition properties.

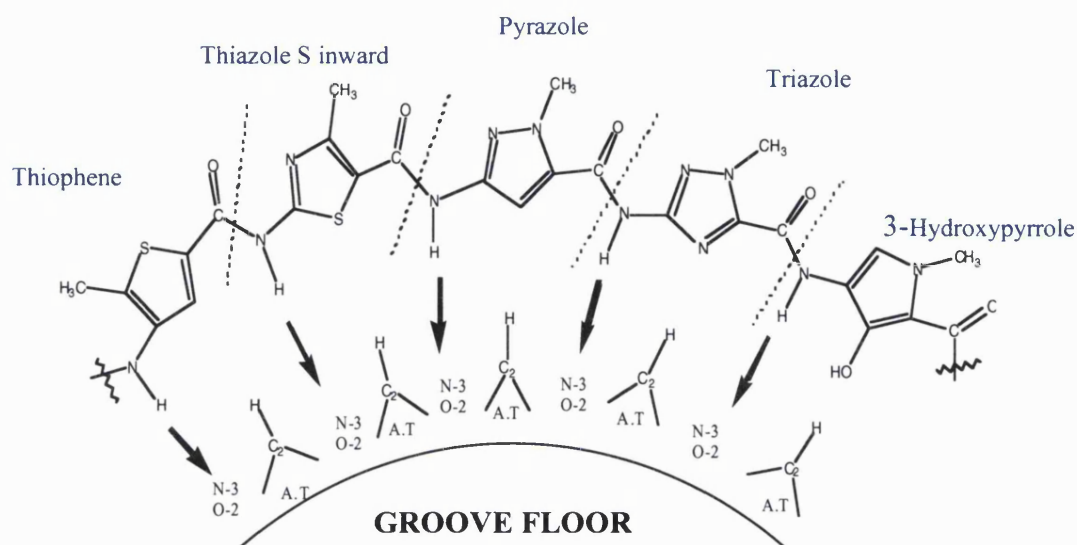
<u>heterocycle pair</u>	<u>base pair</u>
ImPy	GC
PyIm	CG
ImIm	GC or CG
PyPy	AT or TA

**Table 1.2** Summary of heterocyclic pairing rules, where Im is imidazole and Py is pyrrole.

### 1.8.3 AT degeneracy

Whilst strong specificity has been attained with imidazole/pyrrole pairings for GC recognition, AT discrimination with pyrrole pairs remains partially degenerate (White, *et al.*, 1996). As a consequence, a number of alternative heterocycles to pyrrole have been investigated to overcome this (Rao *et al.*, 1990a, b, c; Kopka *et al* 1997, 1998; Bremer *et al.*, 2000; Ellervik *et al.*, 2000; Nguyen *et al.*, 2001), however the most successful to date has been N-methyl-3-hydroxypyrrole (Hp) (Figure 1.32). High resolution X-ray crystallography has demonstrated that the hydroxyl group of hydroxypyrrole can fit tightly into the asymmetric cleft between thymine and adenine present in the minor groove, hydrogen bonding with the thymine O2 (Kielkopf *et al.*, 1998b). Its subsequent use has provided the first success in complete sequence recognition along both DNA strands in the minor groove (White *et al.*, 1998). However, although more selective than pyrrole for thymine, hydroxypyrrole exhibited a reduced binding affinity for this base relative to pyrrole (Kielkopf *et al.*, 1998b, 2000; White *et al.*, 1999). Furthermore, its selectivity was affected by its linear context in the polyamide, with Hp/Py pairs possibly limited to internal positions for TA discrimination (Ellervik *et al.*, 2000). As a result alternative heterocycles have been investigated in an attempt to improve both the affinity and specificity of AT versus TA discrimination, including hydroxybenzimidazole (Hz), which appears to mimic the interactions of

hydroxypyrrole with thymine (Renneberg & Dervan 2003), and the potential adenine discriminator thiazole (Rao *et al.*, 1990a, b, c; Kopka *et al.*, 1997, 1998).



**Figure 1.32** Representation of binding by putative AT recognition elements, adapted from Bailly & Chaires 1998.

To examine the effect on binding of including a thiazole heterocycle, a series of polyamides were synthesised containing thiazole present in two alternative orientations (Rao *et al.*, 1990a). These molecules were synthesised so that either the sulphur or the nitrogen atom of the attached thiazole would be directed into the minor groove. Those bearing nitrogen directed towards the DNA had a tolerance for GC base pairs in addition to binding to AT rich regions, possibly as a result of hydrogen bonding between thiazole nitrogen and guanine N2-NH<sub>2</sub> (Rao *et al.*, 1990b). In contrast, when thiazole was placed with the sulphur atom directed floorwards the compounds exhibited a strict preference for AT base pairs, providing an even greater selectivity for these regions than either netropsin or distamycin. This occurred as a result of a steric clash between the larger sulphur atom and the guanine N2-NH<sub>2</sub>, causing an avoidance of GC base pairs (Kumar *et al.*, 1990b; 1991). Avoidance of guanine N2-NH<sub>2</sub> appears to be a key component of sequence recognition in the minor groove, indicating that this group is an important carrier of biological information (Warpehsoki & Hurley 1988). For example, the strict discrimination of 3'terminal AT base pairs of a number of oligopeptide

antibiotics such as netropsin, distamycin, anthelvencin and kikumycin is attributed to such a steric clash between the guanine amino and their carboxyl terminii (Lown 1988). There is also evidence from NMR that in addition to groove binding, in the N-terminal position the thiazole ring may also avoid GC regions by intercalating between AT base pairs due to the steric demands of the bulky sulphur atom (Kumar *et al.*, 1990b). Thus as for other groove binding agents such as DAPI (Wilson *et al.*, 1989), a fine balance may exist between intercalation and groove binding for sequence recognition by this heterocycle.

#### 1.8.4 Stacking and linkage across the groove

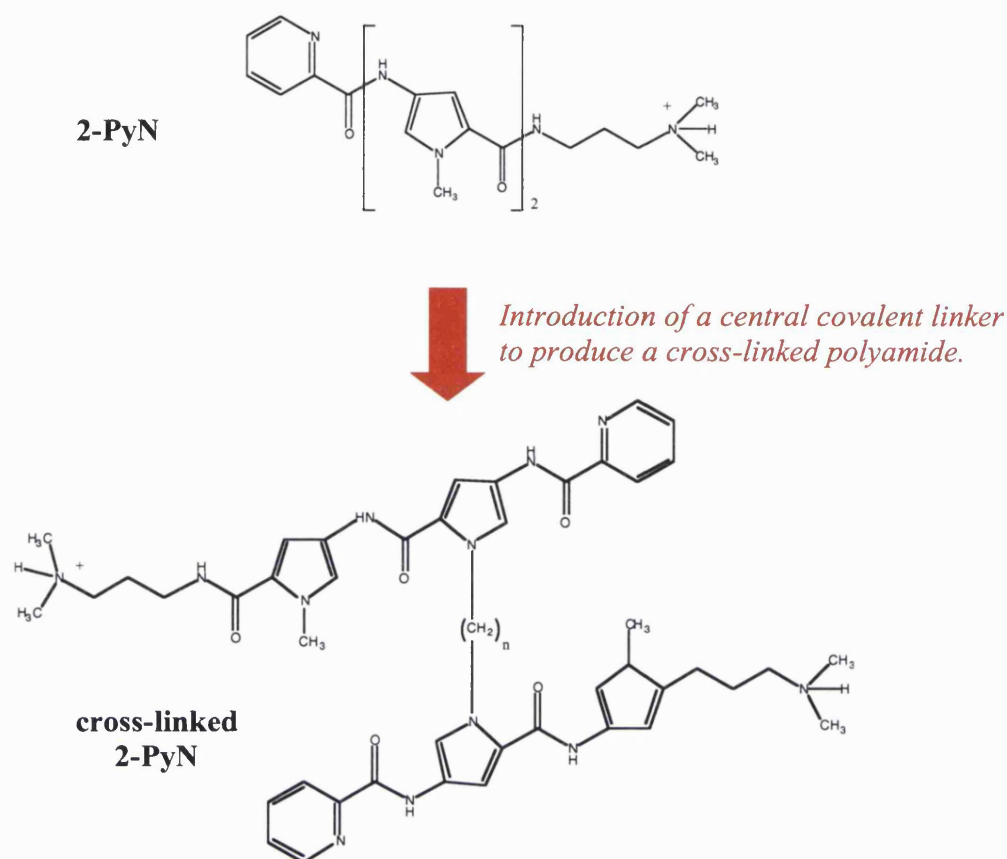
In the 2:1 motif, the relative stacking of two polyamide molecules within the groove becomes critical as slippage of one ligand over the other can pair up different heterocycles. In the absence of an amide group on the leading heterocycle ring, the two peptide chains completely overlap in a maximum overlap motif, whereby each peptide ring stacks on an amide of its neighbour. Addition of an N-terminal amide may cause the two peptides to slip in relation to each other by one amide, leading to a one-residue stagger motif. In this case the two ligands still stack ring on amide, however the first amide on each constituent peptide overhangs, resulting in an extended reading frame for a given heterocycle combination (Kopka *et al.*, 1997) (Lacy *et al.*, 2002a). Thus a single linear combination of heterocycles can have two different reading frames depending on the stacking arrangement of its constituent peptides as shown below in Figure 1.33.



**Figure 1.33** Two modes of binding enabling the polyamide ImPyPy to read two different DNA sequences, where Im is imidazole, Py is pyrrole, = indicates the carboxamide moiety and + represents the C-terminal cationic tail (Kopka *et al.*, 1997).

Tethering two oligopeptides across the minor groove with an appropriate linker can fix the stacking arrangement of the heterocycle pairs across the minor groove, as well as increasing the chances of dimeric binding. Although high drug: DNA ratios can induce cooperative binding in a 2:1 complex of unlinked polyamides within the groove, this may not always be energetically favourable. For example, monomeric polyamides containing thiazole with an inward directed sulphur atom have been shown to incur energetic penalties due to the steric interactions of this bulky group, making co-operative binding less likely as a consequence (Kumar *et al.*, 1991). However, optimal linkage can energetically favour dimeric binding, providing an overall binding affinity that exceeds the product of two stepwise binding constants resulting from the co-operative binding of two separate monomers (Chen *et al.*, 1996).

Enhanced binding was achieved with the aforementioned 2-PyN for its 2:1 binding site 5'-TGTCA-3', when two 2-PyN monomers were covalently linked via the central pyrrole-N1 position using a methylene linker ranging from three to six carbons in length as shown in Figure 1.34 (Mrksich & Dervan 1993b). Footprinting titrations revealed that the linked analogues of 2-PyN, bound to 5'-TGTCA-3' with a ten fold increase in affinity relative to the monomer. Furthermore, the ratio of binding affinities of 2-PyN for 5'-TGTCA-3' and 5'-TTTTT-3' was altered from 1:1 for the monomer to 25:1 for the dimeric form. Structural analysis of the binding of the 2-PyN covalent dimer revealed that linkage produced an almost equivalent series of DNA contacts as observed with the co-operative binding of two unlinked 2-PyN monomers (Dwyer *et al.*, 1993). Moreover, its covalent linkage to a tripyrrole ligand using a butyl linker to produce a heterodimer could expand the range of sequences that could be targeted (Mrksich & Dervan 1994a). These data, together with subsequent experiments by Lown and coworkers linking two alternative tripyrrole ligands (Chen & Lown 1994), indicated that covalent linkage of two polyamides could significantly improve both their binding affinity and specificity.



**Figure 1.34** Covalent linkage of 2-PyN monomer into a cross-linked dimer ( $n=3-6$ ).

#### 1.8.5 A hairpin versus cross-linked linkage motif

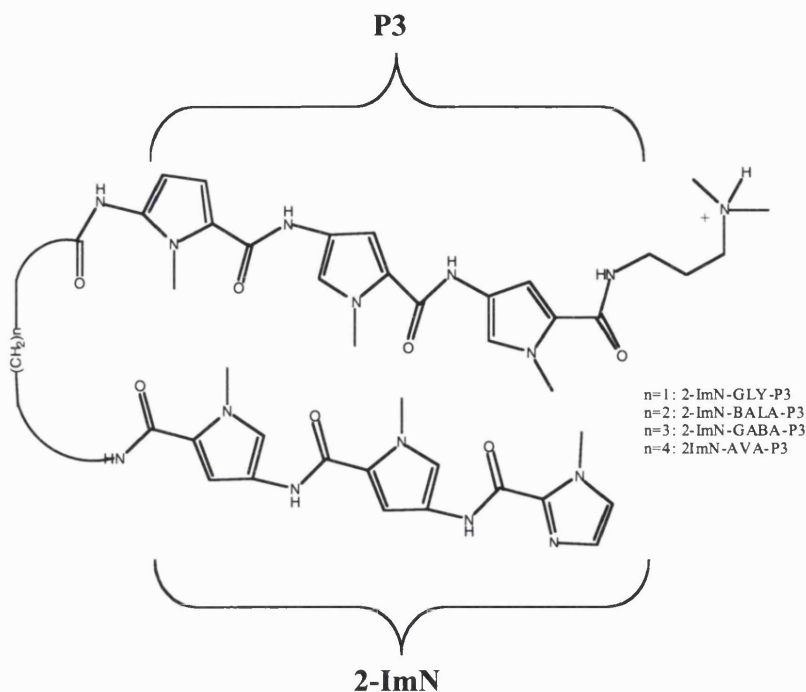
As the accurate alignment of the peptide rings alongside each base pair is crucial to providing both specificity and affinity for linked polyamides, a covalent linker must be of an optimal length and geometry to position both component ligands in close proximity to the base edges within the minor groove. To achieve this, two types of structures have been proposed; a hairpin motif, whereby the oligopeptides are linked head to tail using a suitable "turn peptide" (Mrksich *et al.*, 1994b), and a cross-linked or stapled structure where the central rings of two polyamide ligands are linked via a methylene bridge (Mrksich, & Dervan 1993b, 1994a; Chen & Lown 1994).

Hairpin linkages have the advantage of rapid linear synthesis using solid phase techniques. Different amino acid linkers have been tested for hairpins, varying from 1-4 methylene carbons as shown in Figure 1.35 (Mrksich *et al.*, 1994b). Glycine and  $\beta$ -

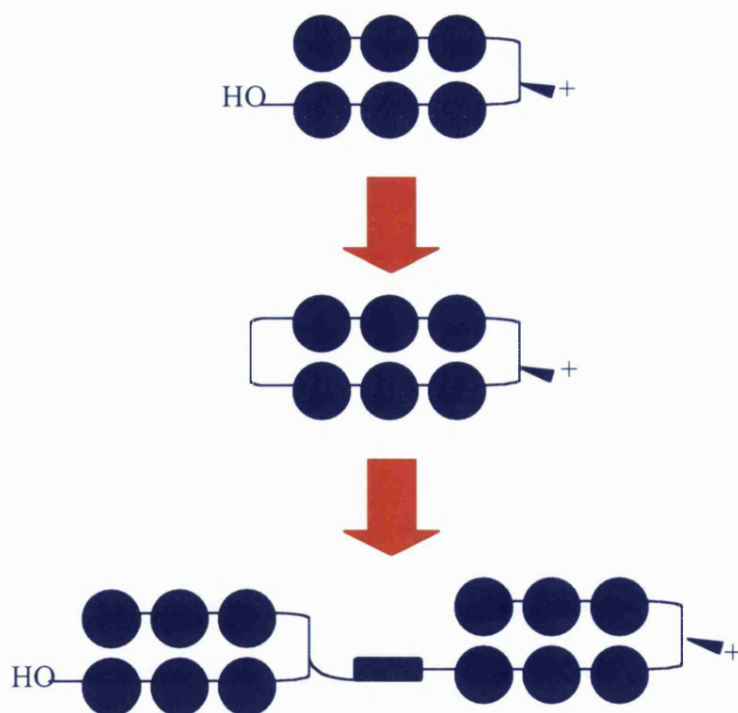
alanine linkers, one and two carbon atoms in length, respectively, were too short for optimal hairpin formation resulting in linear polyamides. However,  $\gamma$ -aminobutyric acid (three carbons) enabled the molecule to turn easily, with dimers linked using this “ $\gamma$ -turn” exhibiting up to a three hundred-fold increase in affinity relative to the corresponding monomers (Mrksich *et al.*, 1994b). The linker length was found to have a profound effect on the binding affinities of these molecules, with over a thousand-fold difference in the affinity observed between the best ( $\gamma$ -aminobutyric acid) and the worst linkers (5-aminovaleric acid). Addition of the  $\gamma$ -turn increased both the binding affinity and, through the inhibition of slipped binding motifs, the sequence selectivity (Trauger *et al.*, 1996b). Subsequent addition of a C-terminal  $\beta$ -alanine or “ $\beta$ -tail “ to facilitate the solid phase synthesis process was found to enhance binding affinity and specificity yet further (Parks *et al.*, 1996a, b), with the  $\gamma$ -turn and the  $\beta$ -tail demonstrating a selectivity for AT and TA base pairs (Swalley *et al.*, 1999). Recently a “U-pin” turn motif has been synthesised in an effort to provide a GC rather than AT tolerance at the linked end of the molecule (Heckel & Dervan 2003). Further variation on the basic hairpin structure has produced cyclic (Chen & Lown 1994; Herman *et al.*, 1999a) as well as tandem hairpin motifs where a number of hairpins may be linked in series (Herman *et al.*, 1999b) (Figure 1.36).

Whilst the structural components required for optimised DNA binding by hairpin polyamides have been investigated and reviewed in detail (Dervan & Burli 1999), they are less clearly defined for cross-linked molecules. Unlike hairpin molecules, which require a three carbon linkage or  $\gamma$ -turn to link two monomers ‘head to tail’, ethidium bromide displacement, CD, NMR and molecular modelling studies have indicated that cross-linkage or an “H-pin” requires a longer stretch of methylene groups to cross the groove width via the central heterocycle ring (Chen & Lown 1994, 1995; Chen *et al.*, 1996). These studies indicated that at least a five-carbon linker was required to span this distance, with linkers of six to eight carbons providing optimal linkage. A further increase in linker length resulted in a reduction in both affinity and specificity, with the ligand slipping across the binding site increasing the size of the site occupied (Chen *et al.*, 1996).





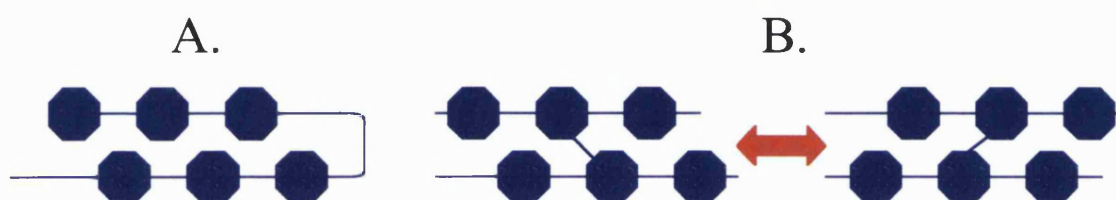
**Figure 1.35** A hexapeptide hairpin, where the terminal carboxylic acid of 2-ImN and terminal amine of the tripyrrole portion (P3) are connected with glycine (Gly),  $\beta$ -alanine ( $\beta$ Ala),  $\gamma$ -aminobutyric acid (GABA), 5-aminovaleric acid (Ava) (Mrksich *et al.*, 1994b).



**Figure 1.36** An illustration of the development of hairpin linkages showing the evolution of a flexible hairpin linker into cyclic and tandem linked polyamides.



As with cooperative dimers, two potential DNA binding modes are possible for linked polyamides, an overlap and a staggered orientation (Kopka *et al.*, 1997). The maximum overlap mode is favoured by the hairpin linkage due to the absence of an amide on the leading ring (Parks *et al.*, 1996a, b). Cross-linked polyamides, however, are synthesised with a leading N-terminal head group in general and thus are able to bind in the one-residue stagger motif, slipping along the peptide chain by one amide yielding an extended reading frame. Therefore the use of a central methylene linker with a degree of inherent flexibility enables binding in both modes for cross-linked dimers, although a one-residue staggered motif should be favoured (Kopka *et al.*, 1997; Lacy *et al.*, 2002a) (Figure 1.37).



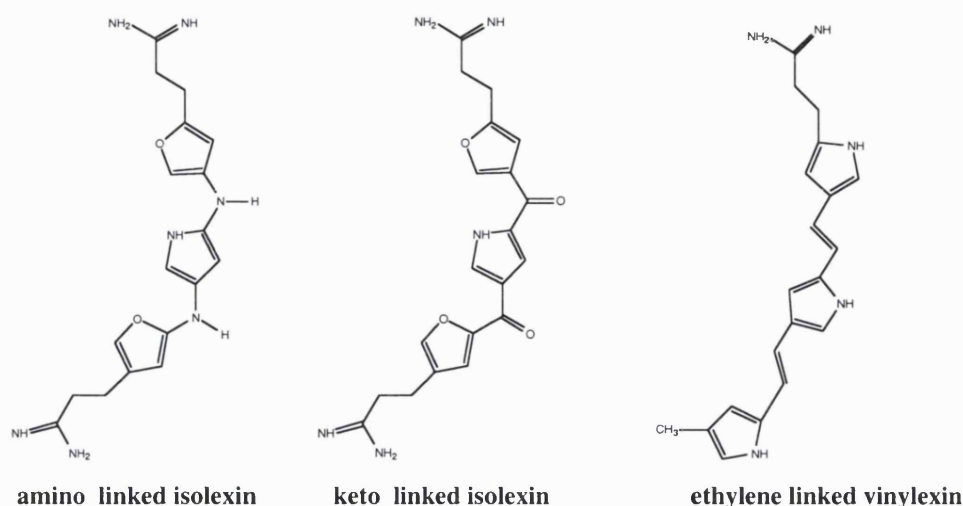
**Figure 1.37** Illustration of hairpin versus cross-linked linkages, showing binding in the maximum overlap motif for a hairpin structure (A) and in both modes for a cross-linked structure (B).

#### 1.8.6 Extending the recognition site

Unique sequence recognition in the genome requires an ability to target a stretch of sequence at least 16 base pairs in length. In an effort to extend the size of the binding site covered by netropsin and distamycin, a number of N-methyl-pyrrole carboxamide derivatives containing, three, four, five and six pyrrole rings were synthesised (Youngquist & Dervan 1985). Affinity cleavage experiments demonstrated that these homologues bound to sequences of five, six, seven and eight contiguous AT base pairs respectively, with the four and five ring analogues binding more tightly than the parent molecule, distamycin. However, further extension to incorporate a sixth pyrrole caused a decrease in binding affinity, as the molecule no longer matched the curvature of the minor groove (Youngquist & Dervan 1985). Theoretical modelling suggested that this

was due to a slight mismatch in length between the pyrrolecaboxamide unit and its target base pair, resulting in an inability to match the base pair rise in the minor groove over extended sequences (Goodsell & Dickerson 1986). Indeed repeat lengths can vary for different heterocycle moieties, with the repeat length of thiazole 0.9Å longer than pyrrole (Lown 1993). Thus as a polyamide increases in length, the hydrogen bonds and van der Waals contacts become out of phase with the spacing between nucleotide units of DNA, explaining the decrease in binding affinity observed with longer molecules. Phasing has also proved a problem for antiparallel polyamide dimers, with 2:1 binding of longer polyamides showing the same length dependence as for 1:1 binding whereby four to five rings increased affinity but further rings decreased both affinity and specificity of binding (Kelly & Dervan 1996).

Loss of phasing and isohelicity with increased numbers of heterocycles led to the design of short blocks of tri or tetra heterocyclic polyamides linked together, either “head to head” or “head to tail” with different amino acid linkers in order to rephase the molecule. In an effort to target longer AT sequences, netropsin molecules had been linked ‘head to head’, so that the two N-termini were linked using a hydrocarbon linker (Khorlin *et al.*, 1980). This strategy was subsequently re-employed to circumvent the phasing problem, covalently linking netropsin and distamycin analogues in a linear fashion using a variety of rigid and flexible linkers to enable effective bidentate 1:1 binding (Youngquist & Dervan 1987; Lown *et al.*, 1989; Wang & Lown 1992; Guo *et al.*, 1993; Burckhardt *et al.*, 1997). This provided successful bidentate binding along a site of sixteen base pairs by three tetrapyrrole units linked by a series of more flexible β-alanine linkers (Youngquist & Dervan 1987). Molecular modelling studies have suggested that improved isohelicity might also be attained by replacing the carboxamide bonds with shorter keto or amino linkages to produce isolexins (Goodsell & Dickerson 1986; Zakrzewska & Pullman 1988) or by introducing ethylene linkages between adjacent heterocycles (Zakrzewska *et al.*, 1988) as shown in Figure 1.38.



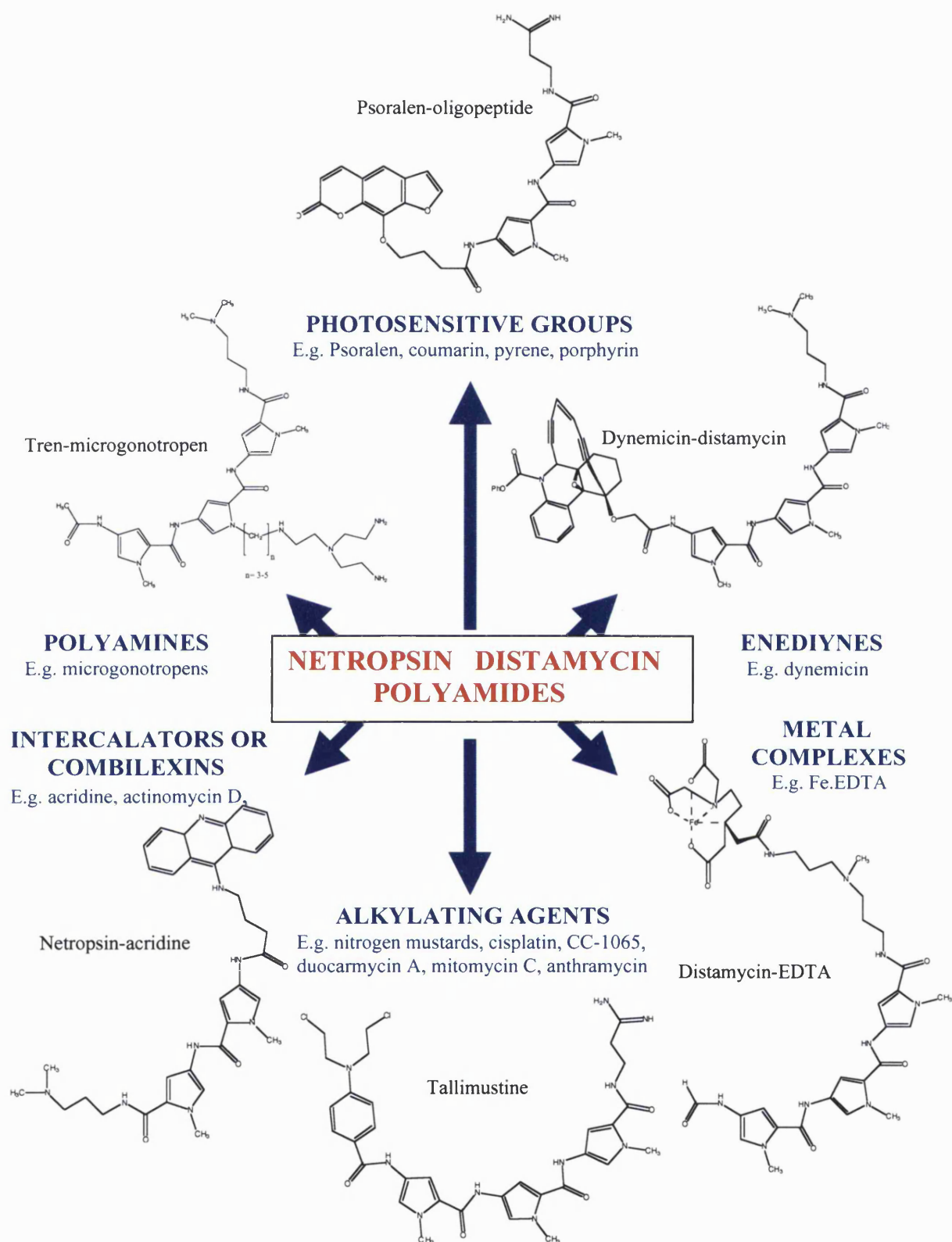
**Figure 1.38** Molecular structures of isolexins and vinylexins.

Recent work into the transcriptional modulation of gene sequences *in vitro* using hairpin ligands has generated new interest in extending the recognition site of polyamides which bind as antiparallel dimers (Turner *et al.*, 1997, Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998, 1999a, 1999b; Mapp *et al.*, 2000; Chiang *et al.*, 2000). Different strategies have been employed to approach the phasing problem of 2:1 binding, in an effort to target larger and therefore less common gene sequences. Hairpin polyamides have been extended by the insertion of  $\beta$ -alanine pairs between every two or three heterocycles acting as a molecular spring to loosen the rigid curvature of the ligand, with the binding affinity and specificity dictated by the parallel placement of  $\beta/\beta$ ,  $\text{Py}/\beta$  and  $\text{Im}/\beta$  pairs (Turner *et al.*, 1998). Cooperative binding of extended hairpins has also been used to target sites of 10 or 12 base pairs (Trauger *et al.*, 1998). An alternative approach has been the development of tandem motifs whereby a number of hairpin or cyclic units can be linked either “head to head” or “head to tail” in series (Herman *et al.*, 1999b). Work is ongoing investigating the optimal linker required for this, with use of a 5-aminovaleric acid or 2-(2-aminoethoxy)acetic acid linker joining two six ringed “head to tail” hairpins enabling the targeting of a ten base pair site (Kers *et al.*, 2002). Ten base pairs have also been successfully targeted using a “head to head” tandem hairpin dimer (Weyermann & Dervan 2002).

#### 1.8.7 Polyamides as molecular vectors for non-covalent and covalent binding

Netropsin, distamycin and subsequent polyamide derivatives have been used as molecular vectors for a wide range of DNA interactive agents (Figure 1.39). This has modified the groove-binding characteristics of a number of classical DNA binding molecules, as well as increasing the affinity and selectivity of both non-covalent and covalent binding by hybrid agents at defined regions of sequence. For example, conjugation of a linear or branching alkylamine chain at the central methyl-pyrrole heterocycle of distamycin has led to the development of a series of molecules called microgonotropens. These bind non-covalently in the minor groove, demonstrating a high affinity for DNA as the additional protonated alkylamine side-chains form electrostatic bonds with the negatively charged phosphodiester backbone (reviewed by Satz & Bruice 2002 and references therein). As a result, microgonotropens have been shown to inhibit transcription factor binding in a sequence selective manner *in vitro*, although little activity has been observed in whole cells possibly due to an inability to traverse the cell membrane (Bruice *et al.*, 1997; White *et al.*, 2002; Satz & Bruice 2002).

Photoreactive agents such as psoralen and coumarin, which intercalate with DNA forming covalent adducts at pyrimidine 5'-TpA steps on exposure to UV, have also been conjugated to polyamides to produce photosensitive oligopeptides (Hartley *et al.*, 1994; Rao *et al.*, 1994). These hybrid compounds demonstrate an increased selectivity for DNA sequence, producing light dependent lesions. Other intercalating moieties such as ellipticine and acridines have been attached to netropsin and distamycin structures to produce molecules with mixed binding modes known as combilexins (reviewed in Bailly & Hénichart 1994). More recently, polyamide-acridine conjugates have been used to disrupt major groove binding proteins by causing distortion of the local DNA structure (Fechter & Dervan 2003). Oligopeptides may also be linked with metal complexes such as Fe-EDTA, to produce molecules with a DNA cleaving moiety as observed with distamycin-EDTA (Schultz *et al.*, 1982, Figure 1.39). The Fe-EDTA substituent may be chemically reduced to generate hydroxyl radicals causing DNA strand breaks at the site of ligand binding, providing the basis for the binding site mapping technique, DNA affinity cleavage (Schultz *et al.*, 1983; Taylor *et al.*, 1984).



**Figure 1.39** Examples of netropsin, distamycin and polyamide conjugates.

High affinity sequence selective DNA cleavage has also been produced by the conjugation of netropsin and distamycin to enediyne antibiotics (reviewed extensively in Bailly & Chaires 1998, Figure 1.39).

Many classical major and minor groove-binding alkylating agents such as chlorambucil, cisplatin, mitomycin C, anthramycin, duocarmycin A, CC-1065 and their structural derivatives have also been attached to netropsin, distamycin and more recently polyamides. The alkylation patterns of these molecules can be significantly modified, with hybrid compounds exhibiting altered sequence selectivity, cytotoxicity and groove-binding properties. For example, several benzoic acid mustard (BAM) tethered oligopyrrole analogues of distamycin have been synthesised, including tallimustine (FCE 24517, Figure 1.39). These were demonstrated to be more cytotoxic than unconjugated BAM, with an increase in the cytotoxic potency observed as the number of constituent pyrrole rings increased (Wyatt *et al.*, 1994, 1995). Furthermore, whilst BAM alone formed lesions in the major groove at the N7 position of guanine residues, selective alkylation by tallimustine was found occur at the adenine N3 position in the minor groove (Broggini *et al.*, 1995). However, for minor groove binding to predominate, at least two pyrrole rings were required with both guanine N-7 and adenine N-3 lesions produced by a monopyrrole-amide BAM conjugate (Wyatt *et al.*, 1995). The nitrogen mustard chlorambucil has also been tethered to distamycin (Lee *et al.*, 1993) and more recently to hairpin polyamides with chlorambucil-polyamide conjugates shown to form sequence selective adducts *in vitro* and in cells (Wurtz & Dervan 2000; Wang 2003; Dudouet *et al.*, 2003). A number of pyrrolo[2,1-c][1,4]benzodiazepines-polyamide conjugates have also been synthesised, based on the covalent minor-groove binding agent anthramycin (Reddy *et al.*, 2000). These have been shown to be more cytotoxic than both natural and synthetic pyrrolo[2,1-c][1,4]benzodiazepines, with cytotoxicity related to both the length and heterocycles of the constituent polyamide (Kumar & Lown 2003). Likewise, polyamide structures have been conjugated to several CC-1065 derivatives CI, CBI and CPI, with CBI-polyamides recently shown to alkylate DNA, inhibiting its replication *in vitro* and in intracellular minichromosomal form (Wang *et al.*, 2002, 2003).

### 1.8.8 Polyamides as molecular tools, anti-retrovirals and modulators of gene expression

The current generation of linked polyamides bind as extended dimers within the minor groove, with covalent linkage of the component antiparallel oligopeptides producing sequence-reading ligands of improved binding affinity (Chen *et al.*, 1994, 1996; Mrksich *et al.*, 1993, 1994a, 1994b; Parks *et al.*, 1996a; Baliga *et al.*, 2001). As a consequence, polyamides able to bind and discriminate DNA sequence with an equivalent affinity to cellular DNA-binding proteins have become useful molecular tools in the elucidation of critical groove-binding contacts of protein-DNA complexes during transcription (Neely *et al.*, 1997, McBryant *et al.*, 1999; Winston *et al.*, 2000; Ehley *et al.*, 2002), as well as prospective candidate molecules for gene targeting strategies (Trauger *et al.*, 1996a; Gottesfeld *et al.*, 1997). These have included their use as components of larger artificial transcriptional activators (Mapp *et al.*, 2000; Ansari *et al.*, 2001) and the direct application of polyamides and polyamide conjugates to modulate transcription of specific genes *in vitro* and in some cell culture situations (Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998, 1999a, 1999b; Mapp *et al.*, 2000; Chiang *et al.*, 2000; Wang *et al.*, 2002; Oyoshi *et al.*, 2002, 2003). Retroviral DNA has provided many of the target sequences used in this work, with polyamides demonstrating potent activity as anti-retroviral agents (Wang & Lown 1992). This has been achieved by modulating viral integration into the host genome and subsequent replication (Clanton *et al.*, 1995; Neamati *et al.*, 1998; Ryabinin *et al.*, 2000; Sharma *et al.*, 2002) or by regulation of the transcription of the viral gene products (Dickinson *et al.*, 1998, 1999a, b; Schaal *et al.*, 2003).

Despite demonstrating promising DNA binding properties in cell free systems, the localisation of polyamides in cell nuclei and their subsequent binding to genomic sequence has been harder to verify directly. As with other nucleic acid targeting strategies, these factors are of critical importance for putative gene targeting in living cells using polyamides. Early cell uptake studies of a nitroxide spin-labeled analogue of netropsin using electron pair resonance spectroscopy indicated that it was readily taken

up and localised predominantly in the cell nucleus (Bailly *et al.*, 1989). More recently however, whilst fluorophore-labelled polyamides have been shown to stain DNA satellites and telomeric repeats in a sequence specific manner in isolated nuclei and chromosomes (Janssen *et al.*, 2000a; Maeshima *et al.*, 2001), their uptake and binding to genomic sequence has been inferred only indirectly by the gain- or loss-of-function phenotypes produced when developing *Drosophila melanogaster* are fed satellite-specific polyamides (Janssen *et al.*, 2000b) and more recently by the derepression of integrated latent HIV-1 in human lymphocytes (Coull *et al.*, 2002). In contrast, studies of the intracellular distribution of fluorescein-tagged distamycin analogues in human ovarian adenocarcinoma cell lines indicate little nuclear uptake, with localisation shown to be mainly cytoplasmic and mitochondrial (Sharma *et al.*, 2001), although several uptake studies carried out concurrently with the work in this thesis and discussed in detail in later chapters indicate that this may be cell-type dependent.

An additional consideration for the use of polyamides in gene targeting strategies has been the accessibility of target sequences when they are compacted as chromatin. Studies *in vitro* have demonstrated that effective polyamide binding occurs at fully or partially exposed regions of nucleosomal DNA, whereas binding sites fully obscured by associated histones are not accessible (Gottesfeld *et al.*, 2001). Moreover, polyamides binding with high affinity to sequences near the periphery or at the centre of a nucleosome can prevent translocation of the histone octamer along the DNA, a process required to enable access of DNA control proteins to the primary sequence, thereby inhibiting transcription from the nucleosomal template (Gottesfeld *et al.*, 2002). Indeed polyamides may be used as a bivalent clamp to lock together nucleosomal supergrooves (Suto *et al.*, 2003), adding to the complexity of potential eukaryotic gene regulation. Such studies suggest that the potential for shielding of genomic sequence by nucleosomes, together with the limited understanding of the dynamics of structural reorganisation during DNA processing events may significantly complicate the targeting of polyamides to cellular DNA.



## 1.9 Experimental aims

Much of the recent work investigating the interactions of polyamides with DNA has involved the extensive use of hairpin-linked molecules, as their linear structure may be rapidly synthesised by solid phase techniques. Whilst optimisation of these structures has been explored extensively, producing agents able to bind to predetermined sites with high affinity, the effective constituents for high affinity selective DNA binding by the alternative cross-linked polyamides remain to be fully characterised. Detailed examination of the optimal combinations of heterocycles, linker and head group, a moiety usually absent from hairpin structures, for this alternative linkage and the effects of these structural substituents on the ligand:DNA interactions could improve the understanding of selective binding by polyamides in general. Furthermore, the continued difficulty encountered in attaining consistent discrimination of A:T from T:A, with a binding affinity approaching that of the pyrrole heterocycle for both base pairs, means that novel heterocycle pairings are still needed to improve the complete sequence recognition already achieved with pyrrole:hydroxypyrrole pairs. More importantly, the absence of conclusive evidence for universal effective cell uptake, nuclear localisation and modulation of the transcription of genes present *in situ* in the genome, properties crucial to predetermined sequence-selective gene targeting in living cells, leaves the potential of polyamides as agents for gene therapy open to question.

The aims of the experimental work were to investigate polyamide:DNA interactions both *in vitro* and in cells, elucidating the binding selectivity and affinity of selected polyamides to Watson-Crick and non-canonical DNA sequences as well as examining their uptake, nuclear localisation and ability to modulate global gene expression profiles in cells (as illustrated in Figure 1.40). The structure-activity relationship of a series of novel unlinked and cross-linked thiazole-containing polyamides was investigated, exploring the effects on sequence-specific DNA binding of various permutations of constituent heterocycle, N-terminal head group and length of methylene linker used to cross-link the molecule. Comparison of the selectivity of a series of polyamides comprised of the heterocycle combination thiazole-imidazole-pyrrole at both maximum overlap and one-residue stagger binding sites enabled some assessment of the role of the

thiazole moiety in the discrimination of adenine when placed at the N-terminal position. Analysis of the relative binding affinity and selectivity of an unlinked N-terminal formyl tri-imidazole polyamide at match, single and double TG mismatch sites along a single DNA sequence allowed an evaluation of the use of polyamides in non-canonical sequence recognition. A comparison was made of the live intracellular uptake of selected polyamides used in the *in vitro* studies and covalent binding DNA ligands to establish directly the membrane permeability and nuclear localisation of polyamides without the use of conjugated fluorophores, to exclude the possibility of the attached moiety influencing the uptake properties of the polyamide. Finally, investigation of the global response of gene expression in a model organism *Saccharomyces cerevisiae* was evaluated on treatment with either distamycin or its structural derivative a cross-linked N-terminal thiazole polyamide in order to determine the effects of polyamides on the transcription of genomic DNA sequence.

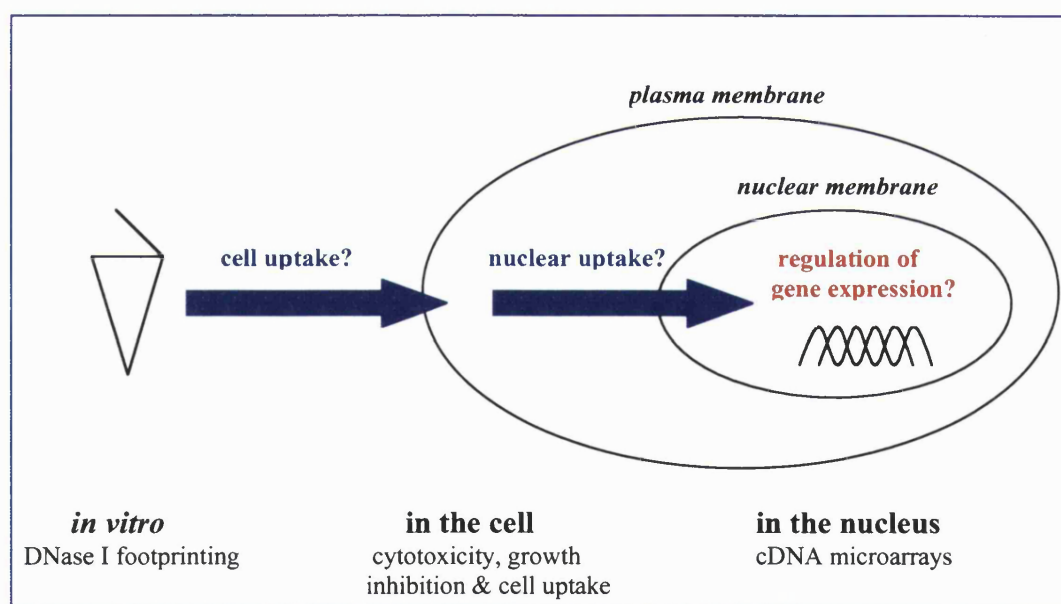


Figure 1.40 Illustration summarising experimental strategy.

## CHAPTER 2

### DNA SEQUENCE RECOGNITION OF WATSON-CRICK BASE PAIRS BY NOVEL THIAZOLE CONTAINING CROSS-LINKED POLYAMIDES.

#### 2.1 Introduction

A series of polyamide monomers and cross-linked homodimers comprised of different combinations of the heterocycles thiazole (Th), imidazole (Im) and pyrrole (Py), with alternative N-terminal head groups attached to the leading heterocycle, and cross-linked by differing lengths of alkanediyl linker, are presented in Figures 2.1 a-b. Constituent heterocycle combinations were chosen to target putative binding sites comprised of mixed sequence, with pyrrole favouring AT and imidazole favouring GC base pairs. The thiazole moiety, orientated so that the sulphur atom was directed towards the minor groove floor on binding, was included as a potential discriminator of adenine. In this position, the bulky sulphur atom produces a steric clash with guanine C2-NH<sub>2</sub> groups protruding into the groove resulting in an avoidance of GC sites and a strict preference for AT base pairs (Rao *et al.*, 1990b; Kopka *et al.*, 1997; Kopka *et al.*, 1998). This preference is lost if thiazole is inverted, directing instead the smaller nitrogen atom towards the groove floor (Rao *et al.*, 1990b; Kumar *et al.*, 1990; Kumar *et al.*, 1991).

At high concentrations unlinked polyamides may bind side-by-side within the minor groove with each peptide ligand contacting the adjacent strand of DNA (Pelton & Wemmer 1989). Cooperative 2:1 binding of unlinked ligands, however, may not always be energetically favourable (Kumar *et al.*, 1991). Covalent linkage of polyamides, either in a cross-linked or hairpin form, energetically favours binding as a 2:1 complex with DNA due to free energy additivity. In this alignment, each heterocyclic ring contacts a single base such that a pyrrole placed opposite a pyrrole in the corresponding ligand will bind AT or TA base pairs, an imidazole opposite a pyrrole will favour GC over CG and an imidazole: imidazole pairing will bind either GC or CG base pairs (Wade *et al.*, 1992; Mrksich *et al.*, 1992; Wade *et al.*, 1993; Mrksich & Dervan 1993; Geierstanger *et al.*, 1994; White *et al.*, 1996). This has led to the development of a set of

heterocycle pairing rules for sequence recognition (Trauger *et al.*, 1996a; White *et al.*, 1997).

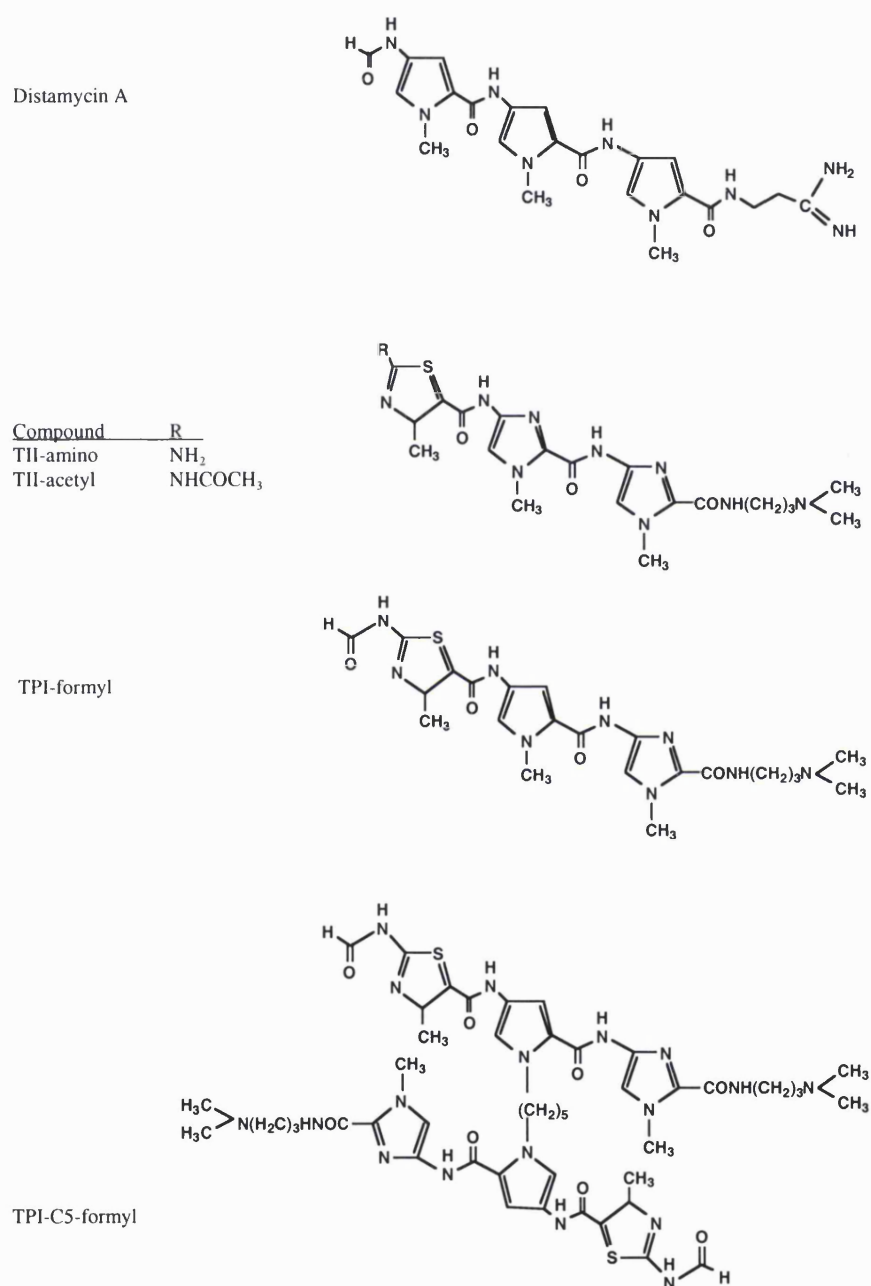
Furthermore, covalent linkage may dictate stacking in a 2:1 alignment (Kopka *et al.*, 1997). Choice of a central methylene linker with limited flexibility, enables two possible modes of binding as a 2:1 complex with DNA, an overlap or a stagger orientation as shown in Figure 2.2 for the heterocycle combinations examined here. In the absence of an amide group on the leading heterocyclic ring, the parallel peptide chains may completely overlap, with each peptide ring stacking on an amide of its neighbour. Addition of an N-terminal amide can cause the component peptides to slip in relation to each other by one amide, leading to a one-residue stagger motif. In this case the two ligands still stack ring on amide, however the first amide on each substituent peptide overhangs, resulting in an extended reading frame (Kopka *et al.*, 1997; Lacy *et al.*, 2002a). The potential for two modes of binding for any given heterocycle permutation of a cross-linked polyamide permits comparative study of the favoured modes for a range of heterocycle combinations.

The dimers investigated here were cross-linked with a flexible polymethylene chain  $-(CH_2)_n$  bridging the two central heterocycles via their N-methyl carbon atoms. Chain lengths ( $n=5-8$ ) were based on previous CD, NMR and ethidium fluorimetry studies of cross-linked tri-pyrrole polyamides (Chen & Lown 1994; Chen *et al.*, 1996). These identified that a minimum of five carbons was required to span the groove width, with comprehensive contact across the DNA duplex possible with hexanediyl (C6), heptanediyl (C7) and octanediyl (C8) linkages. Shorter linkers, such as propanediyl (C3) and butanediyl (C4), were unable to accommodate binding across the groove, whilst linkers of nine to twelve carbons also demonstrated poor binding strength relative to those of an intermediate length (C6-C8) (Chen *et al.*, 1995b; Chen *et al.*, 1996). Long alkanediyl linkers incurred large energetic penalties due to chain folding as well as ligand slippage across potential binding sites, thereby increasing the size of the site occupied (Chen *et al.*, 1996).

At the N-terminus of each of the presented compounds, one of four putative head groups was attached to the leading heterocycle; a hydrogen, an amino, a formamide (formyl) or an acetamide (acetyl). As proposed by Kopka and coworkers (Kopka *et al.*, 1997), unlinked N-formamido polyamides binding in a 2:1 complex with DNA have been shown to favour a one-residue stagger binding motif, whilst those lacking an N-terminal formamide, bind in a maximum overlap mode (Lacy *et al.*, 2002a). Preference for the overlapped motif is well documented for hairpin-linked polyamides, which also lack a head group, with only a hydrogen present at that position (Wemmer & Dervan 1997; Dervan & Burli 1999). At the N-terminus, a formamide group, as found in distamycin A, easily stacks in a 2:1 alignment and is well placed to interact with the DNA through hydrogen bonding and van der Waals interactions (Lacy *et al.*, 2002a). The alternative head groups attached here to the leading heterocycle were chosen to further investigate the influence on sequence recognition and bidentate binding of these active functionalities with different electrostatic properties. This enabled a comparative study of unnatural N-terminal head groups such as hydrogen or acetamide, which are unable to hydrogen bond with DNA, with hydrogen bond donors such as formamide and amino which are N-terminally placed in naturally occurring minor groove binding antibiotics (Sharma, *et al.*, 2000; Probst, *et al.*, 1965; Takaishi, *et al.*, 1972).

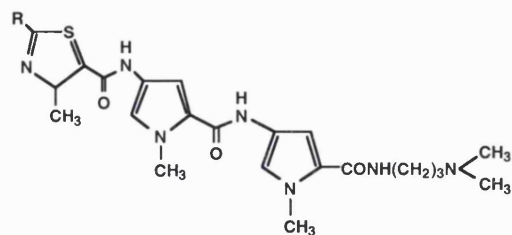
The characterisation of binding *in vitro* to a Watson-Crick DNA sequence by a series of novel unlinked and cross-linked thiazolated polyamides is presented in this chapter. The structure-activity relationship is investigated here using DNase I footprinting (Lane *et al.*, 1983) exploring the effect of heterocycle combination, methylene linker length and N-terminal head group on the sequence specificity and affinity of DNA binding. Quantitation of the binding affinity and selectivity of the heterocycle combination thiazole-imidazole-pyrrole (TIP) at putative DNA match and mismatch sites, for both the maximum overlap and one-residue stagger binding modes, is determined using a footprinting titration method (Brenowitz *et al.*, 1986a, b; Senear *et al.*, 1986). The comparison of a series of cross-linked ligands, able to bind in both binding motifs, enabled an evaluation of optimal linkage for these novel compounds, as well as an

investigation of the influence of the N-terminal head group together with the potential role of N-terminal thiazole in adenine discrimination.

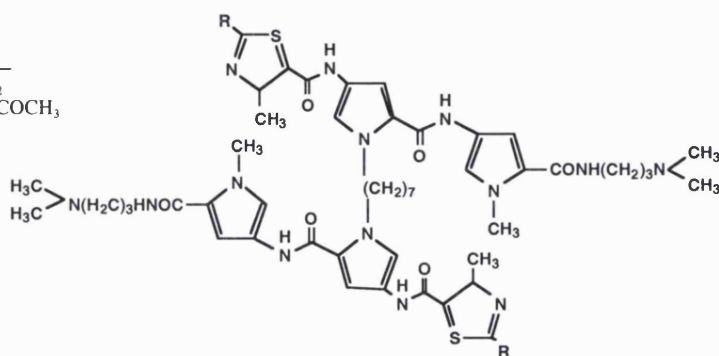


**Figure 2.1 (a)** Molecular structures of distamycin A, the unlinked (TII) thiazole-imidazole-imidazole polyamides, and the unlinked and cross-linked (TPI) thiazole-pyrrole-imidazole polyamides used in this study.

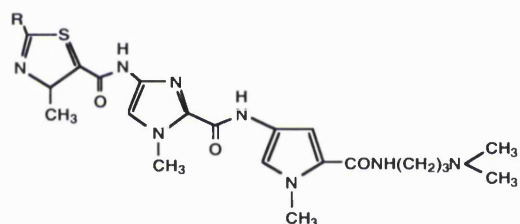
Compound	R
TPP-amino	NH <sub>2</sub>
TPP-formyl	NHCHO
TPP-acetyl	NHCOCH <sub>3</sub>



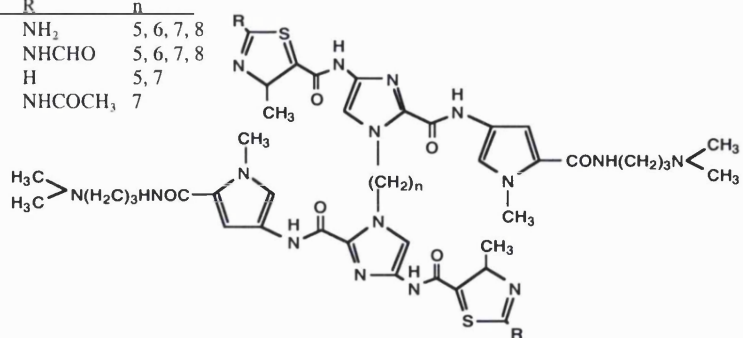
Compound	R
TPP-amino	NH <sub>2</sub>
TPP-acetyl	NHCOCH <sub>3</sub>



Compound	R
TIP-amino	NH <sub>2</sub>
TIP-acetyl	NHCOCH <sub>3</sub>

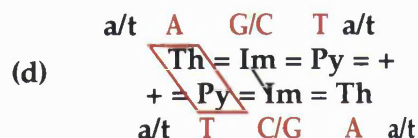
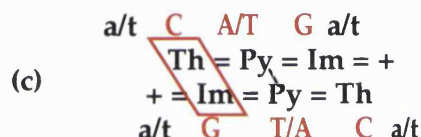
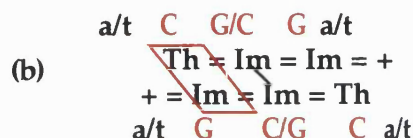
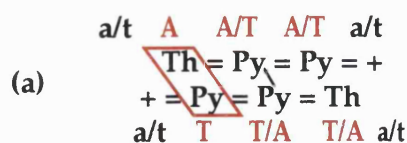


Compound	R	n
TIP-Cn-amino	NH <sub>2</sub>	5, 6, 7, 8
TIP-Cn-formyl	NHCHO	5, 6, 7, 8
TIP-Cn-H	H	5, 7
TIP-Cn-acetyl	NHCOCH <sub>3</sub>	7

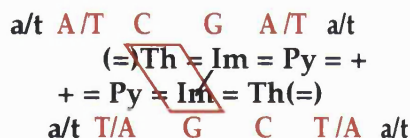
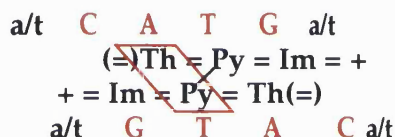
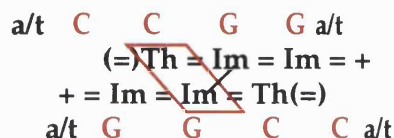
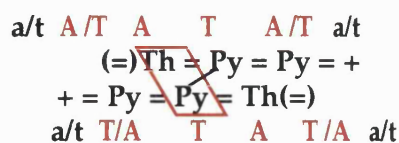


**Figure 2.1 (b)** Molecular structures of the unlinked and cross-linked (TPP) thiazole-pyrrole-pyrrole polyamides, and the unlinked and cross-linked (TIP) thiazole-imidazole-pyrrole polyamides used in this study.

(i) Maximum Overlap



(ii) One Residue Stagger



**Figure 2.2** Binding motifs for (a) TPP, (b) TII, (c) TPI, (d) TIP. The DNA target sequence is indicated in capital letters. The red box indicates heterocycle rings, which are side-by-side in the minor groove; Th is thiazole, Im is imidazole, Py is pyrrole. The equal sign, =, signifies an amide and the plus sign, +, signifies the charged tail. (i) Maximum Overlap Motif – In this motif the two ligands completely overlap stacking ring-on-amide. (ii) One Residue Stagger Motif - This motif requires a leading amide, shown in parentheses (=). Here the polyamides slip relative to the maximum overlap, still stacking ring-on-amide, but moving the charged tails further apart. The one residue stagger motif has a longer reading frame, but still preserves the same number of ring-on-amide interactions.



## 2.2 Materials

### 2.2.1 Investigational compounds

Compounds were kindly provided by Professor J. William Lown, Department of Chemistry, University of Alberta, Edmonton, AB, Canada. The synthesis and characterisation of the compounds have been reported (Sharma, *et al.*, 2000). A minimum of 1mg of each compound was dissolved in distilled and deionised water at a 10mM stock concentration and stored at  $-20^{\circ}\text{C}$  until use. Dilutions were freshly prepared for each experiment in distilled and deionised water.

### 2.2.2 Plasmid DNA

Plasmid pBSFIV34TF10 (4.8 mg/ml) was kindly provided by Dr Tom Phillips, Vaccine Research Institute, San Diego, CA, U.S.A. and was stored in 10mM Tris-HCl (pH 7.8), 1mM EDTA at  $-20^{\circ}\text{C}$ .

### 2.2.3 Radioisotope

$[\gamma\text{-}^{32}\text{P}]$  ATP (5000 Ci/mmol, 10 mCi/ml) was purchased from Amersham International.

### 2.2.4 Chemicals

Tris-hydroxymethyl-methylamine (Tris), ethylenediaminetetraacetic acid (EDTA), ethylenediaminetetraacetic acid disodium salt ( $\text{Na}_2\text{EDTA}$ ), sodium chloride (NaCl), anhydrous sodium acetate ( $\text{C}_2\text{H}_3\text{O}_2\text{Na}$ ), boric acid ( $\text{H}_3\text{BO}_3$ ), sucrose, bromophenol blue, xylene cyanole, formic acid, ethidium bromide, N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid] (Hepes), magnesium chloride ( $\text{MgCl}_2$ ), potassium permanganate ( $\text{KMnO}_4$ ), dithiothreitol (DTT), ammonium persulphate (APS), piperidine, tetramethylethylenediamine (TEMED), phosphate buffered saline (PBS), hydrogen peroxide ( $\text{H}_2\text{O}_2$ ), salmon sperm DNA, Hoechst 33258 and phenol:chloroform:isoamylalcohol (25:24:1) were purchased from Sigma.

Sodium dodecyl sulphate (SDS), glacial acetic acid, formamide, isopropanol, potassium chloride (KCl), calcium chloride ( $\text{CaCl}_2$ ), sodium azide ( $\text{NaN}_3$ ), disodium tetraborate ( $\text{Na}_2\text{B}_4\text{O}_7$ ), triethylamine ( $(\text{C}_2\text{H}_5)_3\text{N}$ ) and dimethyl sulfoxide (DMSO) were purchased from BDH.

Biogel-P6 gel (fine grade 45-90  $\mu\text{m}$ ) was purchased from Biorad.

Electrophoresis grade agarose was purchased from Invitrogen.

Glycogen (20mg/ml) was purchased from Roche Diagnostics.

Ultrapure deoxynucleoside triphosphates (dNTPs) solutions (100mM) were purchased from Pharmacia.

X-Omat Kodak film was purchased from Anachem.

### 2.2.5 Enzymes

T4 polynucleotide kinase (10 units/ $\mu\text{l}$ ) supplied in 50mM Tris-HCl (pH 7.6), 25mM KCl, 5mM DTT, 0.1 $\mu\text{M}$  ATP, 0.2 mg/ml BSA and 50% glycerol (v/v) was purchased from Invitrogen.

Vent DNA Polymerase (2 units/ $\mu\text{l}$ ) supplied in 100mM KCl, 0.1mM EDTA, 10mM Tris-HCl (pH 7.4), 1mM DTT, 0.1% Triton X-100 and 50% glycerol (v/v) was purchased from New England Biolabs.

RQ1 RNase-Free DNase (1 unit/ $\mu\text{l}$ ) supplied in 10mM Hepes (pH7.5), 50% glycerol (v/v), 10mM  $\text{CaCl}_2$  and 10mM  $\text{MgCl}_2$  and Not I (10 units/ $\mu\text{l}$ ) supplied in 10mM Tris-HCl (pH7.4), 0.1% Triton X-100, 500mM KCl, 0.1mM EDTA, 1mM DTT, 0.5mg/ml BSA, 50% glycerol (v/v) were purchased from Promega.

### 2.2.6 Buffers

Forward reaction buffer (5X) consisting of 350mM Tris-HCl (pH 7.6), 5mM 2-mercaptoethanol, 50mM  $\text{MgCl}_2$ , 500mM KCl was purchased from Invitrogen.

Vent ThermoPol reaction buffer (1X) consisting of 10mM KCl, 10mM  $(\text{NH}_4)_2\text{SO}_4$ , 20mM Tris-HCl (pH 8.8), 2mM  $\text{MgSO}_4$ , 0.1% Triton X-100 was purchased from New England Biolabs.

Biogel-P6 spin column solution: 8% P6-biogel (w/v), 0.02% sodium azide (w/v) in distilled and deionised water.

Sucrose loading buffer: 0.6% sucrose (w/v), 0.04% bromophenol blue (w/v) and 0.04% xylene cyanole (w/v) in distilled and deionised water.

TAE agarose electrophoresis buffer: 40mM Tris, 20mM acetic acid, 2mM  $\text{Na}_2\text{EDTA}$ , pH8.1.

Footprinting buffer (2X): 10mM Tris (pH 7.0), 1mM MgCl<sub>2</sub>, 50mM KCl, 1mM EDTA, 0.5mM DTT, 20mM Hepes.

DNase I Stop solution: 200mM NaCl, 30mM EDTA (pH8.0), 1% SDS (w/v).

Formamide loading dye: 0.04% bromophenol blue (w/v), 0.04% xylene cyanole (w/v), 98% deionised formamide (v/v).

Tris-boric acid electrophoresis buffer: 89mM Tris, 89mM boric acid, 2mM EDTA, pH8.3

1X SSC: 15mM sodium citrate pH7.0, 150mM NaCl.

#### 2.2.7 DNA size markers

1kb DNA ladder (1mg/ml) in 10mM Tris-HCl pH 7.5, 50mM NaCl and 1mM EDTA was purchased from Invitrogen.

φX174 DNA/Hinf I ladder (1mg/ml) in 10mM Tris-HCl pH 7.5, 1mM EDTA was purchased from Promega.

#### 2.2.8 Kits

A ready-to-use SequaGel-6 sequencing system, consisting of SequaGel-6 monomer solution (19:1 acrylamide: bisacrylamide, 8M urea) and SequaGel complete buffer solution (0.89M Tris borate, 0.89M boric acid, 20mM EDTA) was purchased from National Diagnostics.

A Bio101 GENECLAN II kit containing 6M sodium iodide, 'glassmilk' silica matrix and New Wash concentrate (NaCl, Tris, EDTA) was purchased from Anachem.

A Wizard<sup>®</sup> genomic DNA purification kit, containing solutions for nuclei lysis, protein precipitation, DNA rehydration and RNase A (4mg/ml) was purchased from Promega.

#### 2.2.9 Synthetic oligodeoxynucleotide primers

The synthetic forward primer 5'-CTC GAG CTC GAG TGG GAT GAG-3', identified as OBUSN1, binds to the sequence region 698 to 718 of the plasmid pBSFIV34TF10 and was used to extend the top strand of the non-coding 5' long terminal repeat region (LTR) of the FIV strain 34TF10. The reverse primer 5'-AGA TCT AGA TCT TGC GAA GTT CTC GG-3', identified as OBUSN3, binds to the complementary strand at

position 1076 to 1051 of pBSFIV34TF10 and was used to extend the bottom strand of the 5' LTR region. The primers, purchased from MWG Biotech, were designed such that the resultant double stranded PCR product spanned 355 base pairs of the FIV 5' LTR region contained within plasmid pBSFIV34TF10 from 698 to 1076 (for LTR region derived from pBSFIV34TF10 sequence see appendix I).

For generation of a double stranded fragment of the human HER2/neu promoter, the synthetic primers 5'-GAG GGA GAA AGT GAA GCT GGG AGT T-3', identified as NEU-F, and 5'-CCC TGG TTT CTC CGG TCC CAA-3' identified as NEU-R designed to span a region of 192 base pairs from 339 to 531 (for Neu sequence see appendix II) were purchased from MWG Biotech.

## **2.3 Methods**

### **2.3.1 Preparation of a genomic DNA template**

#### *Cell Culture:*

The K562 human myeloid leukaemia cell line (Andersson, *et al.*, 1979) was grown in suspension culture in RPMI 1640 medium supplemented with 5% foetal calf serum and 2mM glutamine. Cells were incubated at 37°C in an atmosphere of 5% (v/v) carbon dioxide (CO<sub>2</sub>) in air and maintained in an exponential growth phase within a concentration range of  $5 \times 10^4 - 1 \times 10^6$  cells ml<sup>-1</sup>. Cells were passaged using conventional procedures and screened regularly for mycoplasma.

#### *DNA isolation:*

DNA was isolated from  $3 \times 10^6$  K562 cells by means of a Wizard® Genomic DNA Purification Kit, using the reagents and enzymes provided and following the manufacturer's protocol.

#### *DNA Restriction:*

Following isolation of the DNA, supercoils in the resultant genomic prep were removed to facilitate access of DNA sequences for subsequent pcr by digestion with Not I, an enzyme with no restriction sites in the genomic sequence of interest, the neu promoter

region, and a low number of sites in the overall genome. Digestion was carried out using 10 units of Not I in its appropriate reaction buffer at 37°C overnight. Following restriction, the DNA was precipitated with 0.3M sodium acetate and 3 volumes of 95% ethanol, washed once in 80% ethanol and lyophilised dry.

#### *DNA Measurement:*

The concentration of the genomic DNA isolated was measured by fluorimetry with a Perkin-Elmer LS-2B fluorimeter. A standard curve was generated over a concentration range of 0 to 10 µg of salmon sperm DNA in 1X SSC (15mM sodium citrate pH7.0, 150mM NaCl) labelled for 10 minutes at room temperature with the DNA binding agent Hoechst 33258 (10µg/ml), in a final volume of 3ml. Aliquots of genomic DNA were resuspended in 1X SSC and labelled with Hoechst 33258 (10µg/ml). This enabled measurement of DNA concentration as a function of the fluorescence of the DNA binding agent Hoechst 33258. Fluorescence was measured at 360nm excitation and 450nm emission. Following measurement, the final concentration of the DNA isolated was adjusted to 1mgml<sup>-1</sup>.

#### *Generation of a 192 base pair template of the HER2/neu promoter:*

Exponential amplification of a 192 base pair fragment of the HER2/neu promoter was carried out using the oligodeoxynucleotide primers NEU-F and NEU-R, to generate a purified DNA template of the promoter region from the genomic prep for further rounds of amplification. Amplification proceeded in a total volume of 50µl containing 100ng of the genomic DNA prep, 25pmoles of the forward primer NEU-F, 25pmoles of reverse primer NEU-R, 800µM of a dNTP mix, 1 unit of Vent polymerase and 1X Vent Thermopol reaction buffer. PCR was carried out using a PTC-100 Programmable Thermal Cycler (MJ Research) and the cycling conditions consisted of an initial 4 minute denaturation step at 94°C, followed by a cycle of 1 minute at 94°C, 1 minute annealing at 62°C and 1minute extension at 72°C, for 35 cycles. After amplification, the 50µl PCR mixture was taken up in 5µl of 10X sucrose loading buffer and loaded in 2 x 25µl aliquots onto a 2.5% agarose gel (dimensions 8 x 10.5 x 1cm) containing 0.4µg/ml of ethidium bromide, alongside 1µg of a 1kb DNA size marker. Electrophoresis was

carried out for 2 hours at 50 volts in 1X TAE buffer. Following electrophoresis, the 192 base pair fragment was visualised by UV fluorescence and excised. The fragment was isolated from the agarose gel slices using the constituent reagents provided in a GENECLAN II kit following the manufacturer's standard protocol.

### 2.3.2 Preparation of a 5' singly end-labelled fragment

#### *5'-end labelling of primers with [ $\gamma$ - $^{32}$ P] ATP:*

For the 5'LTR of FIV34TF10, the oligodeoxynucleotide primers OBUSN1 or OBUSN3 (6 pmol) were 5' end-labelled with T4 polynucleotide kinase (30 units), [ $\gamma$ - $^{32}$ P] ATP (40 $\mu$ Ci) in 1X forward reaction buffer in a final volume of 20 $\mu$ l, at 37°C for 45 minutes. Following incubation, the labelled primer was placed at 65°C for 10 minutes to heat inactivate the kinase reaction and allowed to cool to room temperature. Once cool, 13 $\mu$ l of distilled water was added and the total volume of 33 $\mu$ l was loaded onto a Biospin-6 column. The column was prepared immediately prior to use by loading 1ml of Biospin-6 solution into an empty column, which was spun at 1200g to pack and remove residual storage buffer. The packed column was then washed through with 200 $\mu$ l of distilled water. The labelling reaction was loaded onto the dry column and spun for 5 minutes at 1200g to elute the end-labelled primer, which was then added to the primer extension reaction. In order to examine the top strand of the 5' LTR region of FIV34TF10, the OBUSN1 primer was 5' end-labelled prior to PCR, whilst the OBUSN3 primer was 5' end-labelled prior to PCR for examination of the bottom strand.

For the HER2/neu promoter, the oligodeoxynucleotide primers NEU-F or NEU-R (40 pmol) were 5'-end labelled and purified on a Biospin-6 column as previously described. The labelling reaction was then added to the primer extension reaction. To examine the top strand of a 192 base pair fragment of the HER2/neu promoter region, the NEU-F primer was 5' end-labelled prior to PCR, whilst the NEU-R primer was 5' end-labelled prior to PCR for examination of the bottom strand.

#### *Exponential Primer Extension:*

For the 5'LTR of FIV34TF10, exponential amplification of the LTR region was carried out with one of the two primers, either OBUSN1 or OBUSN3, labelled. Extension proceeded in a total volume of 50µl containing 10ng of template DNA pBSFIV34TF10, 6 pmoles of labelled primer, 6 pmoles of unlabelled primer for the complementary strand, 800µM of a dNTP mix, 1 unit of Vent polymerase and 1X Vent Thermopol reaction buffer (stored as a 10X stock at -20°C). PCR was carried out using cycling conditions consisting of an initial 4 minute denaturation step at 94°C, followed by a cycle of 1 minute at 94°C, 1 minute annealing at 63°C and 1 minute extension at 72°C, for a total of 35 cycles. An unlabelled extension reaction was carried out in parallel.

For the HER2/neu promoter, exponential amplification of the HER2/neu promoter region was carried out with one of the two primers, either NEU-F or NEU-R, labelled. Amplification proceeded in a total volume of 50µl containing 100ng of the purified 192 base pair template, 40pmoles of labelled primer, 40pmoles of unlabelled primer for the complementary strand, 800µM of a dNTP mix, 1 unit of Vent polymerase and 1X Vent Thermopol reaction buffer. PCR was carried out for 35 cycles using the conditions listed in section 2.3.1.

#### *Purification of the double stranded single-end labelled fragment:*

Following amplification of the 5'LTR of FIV34TF10, the 50µl labelled PCR mixture was loaded in 2 x 25µl aliquots onto a 2.5% agarose gel containing no ethidium bromide (dimensions 8 x 10.5 x 1cm) as described in section 2.3.1, alongside 1µg of a 1kb DNA marker and a 25µl aliquot of the unlabelled primer extension reaction. Samples were then electrophoresed for 2 hours at 50 volts in 1X TAE buffer. Following electrophoresis, the 2 lanes containing the DNA size marker and unlabelled extension reaction were excised and stained for 5 minutes in 100ml of 1X TAE buffer containing 0.4µg/ml of ethidium bromide. The bands were visualised by UV fluorescence and the position of the unlabelled 355 base pair PCR fragment was marked on a ruler. This was used as a reference point to ascertain the position of the labelled 355 base pair fragment aliquots on the remaining portion of unstained agarose gel, which were then excised.

Labelled DNA was isolated from the agarose gel slices a GENECLAN II kit as detailed previously.

For the HER2/neu promoter, after amplification the 50µl labelled PCR mixture was loaded as onto an ethidium bromide free 2.5% agarose gel, alongside 1µg of a φX174 DNA/Hinf I marker and a 25µl aliquot of the unlabelled primer extension reaction. Electrophoresis and isolation of the labelled DNA was carried out as above.

### 2.3.3 DNase I Footprinting

#### *Drug: DNA reactions:*

Polyamides (prepared as in Sharma, *et al.*, 2000) were incubated with 1000 cps of 5' singly end-labelled fragment of the 5'LTR of FIV34TF10LTR in 2x footprinting buffer (10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes), at room temperature for 30 minutes in a total volume of 50µl. Cleavage was initiated by the addition of the drug treated sample to 0.1 units of DNase I diluted in ice cold 10mM Tris pH 7.0 from a stock solution (1unit/µl) and 1µl of a 1:1 solution of 250mM MgCl<sub>2</sub>, 250mM CaCl<sub>2</sub>. The reactions were performed at room temperature and stopped after 3 minutes by the addition of 100µl of a stop mix containing 200mM NaCl, 30mM EDTA pH 8 and 1% SDS. The cleavage products were phenol: chloroform extracted by vortexing with an equal volume of phenol: chloroform: isoamylalcohol (25:24:1), spinning for 5 minutes at 13000rpm and removing the upper aqueous layer containing the labelled DNase I digested DNA into a fresh Eppendorf. This was then ethanol precipitated in the presence of 1µg glycogen, washed once in 80% ethanol and lyophilised dry. The samples were resuspended in 4 µl of formamide loading dye, denatured for 5 min at 90°C and cooled on ice prior to loading on a 6% denaturing polyacrylamide gel.

For quantitative experiments polyamide incubation was carried out over a wide concentration range at small dose increments. Each experiment was carried out in at least triplicate on each DNA strand. The conditions were optimised for cleaving an estimated amount of 0.25µg of DNA per lane by examining cleavage over a range of



both DNase I concentrations and incubation periods. This enabled the minimum concentration of DNase I required to produce a full size range of cleavage products to be determined, whilst ensuring that on average each DNA fragment was cleaved only once. This prevented a bias being introduced into the overall distribution of cleavage products, since over-digestion produces a greater proportion of short fragments as a result of multiple cleavage events. Therefore the DNase I concentration and incubation period used were adjusted so that at least 50% of the full-length fragment remained.

To examine binding at the HER2/neu promoter, the polyamide TIP-C7-amino was incubated with 1000 cps of the 5' single end-labelled HER2/neu fragment in 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes, at room temperature for 30 min, in a total volume of 50µl. Cleavage reactions and subsequent phenol/chloroform extraction were performed as detailed above.

#### *Maxam Gilbert Sequencing Lanes:*

A purine-specific marker lane was generated using 2000 cps of 5' single-end labelled DNA, prepared as described above. Labelled DNA was incubated at room temperature for 5 minutes in 70% formic acid. Following incubation, samples were snap frozen with a dry ice/ ethanol bath and dried by lyophilisation. The DNA pellet was resuspended in 65µl of a freshly diluted and chilled 10% piperidine solution and incubated at 90°C for 30 minutes. The DNA was then precipitated with 0.3M sodium acetate and 2 volumes of absolute isopropanol, washed twice with 80% ethanol and lyophilised dry. Samples were taken up in 4µl formamide loading dye, heated to 90°C for 5 minutes, then placed on ice prior to loading onto the polyacrylamide gel.

A thymine specific marker lane was generated using 2000 cps of 5' single-end labelled DNA, prepared as described previously. Labelled DNA was suspended in 5µl of distilled water and preheated to 90°C for 3 minutes. After 3 minutes, 10µl of 0.1mM potassium permanganate freshly diluted in 2mM sodium borate pH 9.5 was rapidly added and the sample incubated for a further 80 seconds at 90°C. The reaction was

terminated by ethanol precipitation of the DNA, which was subsequently lyophilised dry. Piperidine cleavage was effected as described above.

A cytosine specific marker lane was generated using 2000 cps of 5' single-end labelled DNA, prepared as above. Labelled DNA was suspended in 5 $\mu$ l of distilled water and incubated with 15 $\mu$ l of a freshly prepared hydrogen peroxide mix (245mM hydrogen peroxide, 10mM EDTA, 133mM triethylamine acetate pH7) for 20 minutes at room temperature. Following incubation, samples were snap frozen with a dry ice/ethanol bath and dried by lyophilisation. Piperidine cleavage was carried out as above.

#### *Polyacrylamide Gel Electrophoresis:*

Electrophoresis was carried out on a BioRad Sequi-Gen sequencing gel apparatus using a gel size of 21cm x 50cm x 0.4mm. Prior to pouring, the gel plates were tightly clamped in a 21cm casting tray containing a silicone gasket. The 6% polyacrylamide denaturing sequencing gels were prepared from a National Diagnostics SequaGel-6 kit comprised of a ready-to-use SequaGel-6 monomer solution and a SequaGel complete buffer solution. For each gel, 80ml of SequaGel-6 monomer and 20ml of complete buffer were freshly mixed with 250 $\mu$ l of APS solution (0.25mgml<sup>-1</sup>). Following polymerisation, each gel was pre-run with TBE buffer to warm the gel to approximately 50°C. Once loaded, samples were electrophoresed at 1500-2000V, maintaining the gel temperature at 50-55°C and the electrophoresis run was terminated when the bromophenol blue marker had migrated approximately 40cm. After the run, the gel plates were separated and the gel peeled away from the back plate using a sheet of Whatman 3MM chromatography paper. This was then placed on a further sheet of Whatman DE81 filter paper and onto a porous drying support. Gels were dried under vacuum for 2 hours at 80°C using a Biorad 5850 gel dryer and exposed to X-ray film for 24 hours (X-OMAT, Kodak). Exposure times varied, but generally footprinting gels were exposed overnight without a screen.

#### 2.3.4 Determination of equilibrium association constants for binding sites

##### *Integration of optical density:*

Digitised images of footprint titrations were acquired by scanning autoradiograms (resolution of 100 $\mu\text{m}^2$ ) of at least 3 distinct repeat experiments for each TIP polyamide on each DNA strand using a BioRad GS-670 imaging densitometer interfaced with a 7100/60 Power PC. Images were stored as 12 bit grayscale data sets as Molecular Analyst 2.1.1 files without further manipulation or processing. Subsequently, images were retrieved under Molecular Analyst 2.1.1 and binding sites were identified by eye from the digital gel image as clearly discernable discrete blocks of diminishing band intensity, which progressively decreased with increased ligand concentration across the lanes. At each allocated binding site, the integrated optical density of a selected region of bands placed centrally within the site were acquired for each concentration of ligand using the auto integrate facility of the analysis program. This command automatically corrects for variation in the local background intensity, by integrating the optical density of a 2 pixel width perimeter outside and adjacent to each binding site, averaging and then subtracting this value from the integrated optical density acquired for the binding site itself. In addition, to correcting for experimental variation of the total DNA loaded per lane, the integrated optical density of a standard band (reference site) outside the regions protected from DNase I cleavage by ligand binding was also acquired for each lane analysed. The corrected volume integration values of the binding sites together with the reference site allowed the apparent DNA target site saturation,  $\theta_{\text{app}}$ , to be calculated for each concentration of ligand using the following equation:

$$\theta_{\text{app}} = 1 - (I_{\text{tot}}/I_{\text{ref}})/(I_{0\text{tot}}/I_{0\text{ref}}) \quad (1)$$

where  $I_{\text{tot}}$  and  $I_{\text{ref}}$  correspond to the integrated volumes of the target and reference sites, respectively.  $I_{0\text{tot}}$  and  $I_{0\text{ref}}$  are the integrated volumes for those sites in the absence of polyamide.

*Calculation of  $K_a$  and fitting of binding site isotherms:*

The data were fitted to a Langmuir binding model (eq 2), where  $[L]_{\text{tot}}$  is the total polyamide concentration,  $K_a$  is the equilibrium association constant and  $\theta_{\text{min}}$  and  $\theta_{\text{max}}$  are the site saturation values when the site is unoccupied or saturated, respectively.

$$\theta_{\text{fit}} = \theta_{\text{min}} + (\theta_{\text{max}} - \theta_{\text{min}}) \frac{K_a [L]_{\text{tot}}}{1 + K_a [L]_{\text{tot}}} \quad (2)$$

Values for  $\theta_{\text{min}}$ ,  $\theta_{\text{max}}$  and  $K_a$  were obtained by minimising the difference between  $\theta_{\text{app}}$  and  $\theta_{\text{fit}}$ , using a non-linear optimisation procedure with the constraint  $\theta_{\text{max}} > \theta_{\text{min}}$ . Goodness of fit between  $\theta_{\text{fit}}$  and  $\theta_{\text{app}}$  was assessed by calculating a Pearson correlation coefficient between the actual and the model values. The confidence level was obtained using a test statistic for a t-distribution and any fit with  $p > 0.05$  was rejected. At least three sets of data were used in determining each association constant for five target match and mismatch sites.

The  $\theta_{\text{app}}$  values for each site were used to plot each individual-site isotherm. As the  $\theta_{\text{app}}$  values often do not span the range 0 to 1, these were normalised by subtracting  $\theta_{\text{min}}$  and dividing by  $\theta_{\text{max}} - \theta_{\text{min}}$  (eq 3). The normalised values ( $\theta_{\text{norm}}$ ) were then plotted against the ligand concentration. Isotherms were plotted to demonstrate the accuracy of the optimisation procedure.

$$\theta_{\text{norm}} = (\theta_{\text{app}} - \theta_{\text{min}}) / (\theta_{\text{max}} - \theta_{\text{min}}) \quad (3)$$

## 2.4 Results

### 2.4.1 Sequence-specific binding of novel thiazole-containing cross-linked polyamides.

DNase I footprinting was used to determine the sequence specific DNA binding of a series of novel polyamides consisting of the heterocycles thiazole, imidazole and pyrrole, either as monomers or cross-linked dimers (Figures 2.1(a) and 2.1(b)). Four heterocycle permutations were examined, thiazole-pyrrole-pyrrole (TPP), thiazole-imidazole-imidazole (TII), thiazole-pyrrole-imidazole (TPI) and thiazole-imidazole-pyrrole (TIP). Each combination had one of four candidate N-terminal head groups attached to the leading heterocycle, an amino, formyl, acetyl or hydrogen and were either monomers or cross-linked homodimers with alkanediyl linkers ranging from 5-8 carbons in length. Although the range of possible polyamide permutations, comprised of the assorted component heterocycles, head groups and linkers, was not complete, general binding trends were observed across the compound series.

Footprinting experiments were carried out to examine either an individual compound over a wide range of increasing concentration (0-100 $\mu$ M), or multiple compounds over a small concentration range to facilitate comparison of DNA binding of closely related ligands. Experiments used for the quantitative analysis of binding affinity and specificity of the TIP heterocycle series were carried out over small incremental increases in ligand concentration to provide accurate binding isotherms from which to derive the binding constants at each protected site. Footprints or areas of DNase I protection due to ligand binding were assigned by eye as discrete regions of diminishing band intensity clearly discernable across all the lanes, whereby the intensity at these sites progressively decreased with increasing ligand concentration. Where protection extended over a large number of base pairs at high concentrations, the primary site was assigned from the area protected at the onset of footprinting. Extended areas of protected DNA sequence seen at high ligand concentrations only were not assigned as footprints, but as a non-specific binding due to ligand saturation. Experiments were carried out on both strands of the 5'LTR fragment to confirm binding sites, and targeted sequences were identified by comparison with the corresponding region in sequencing lanes run alongside the footprinting reactions on each gel. Ligand binding sites were

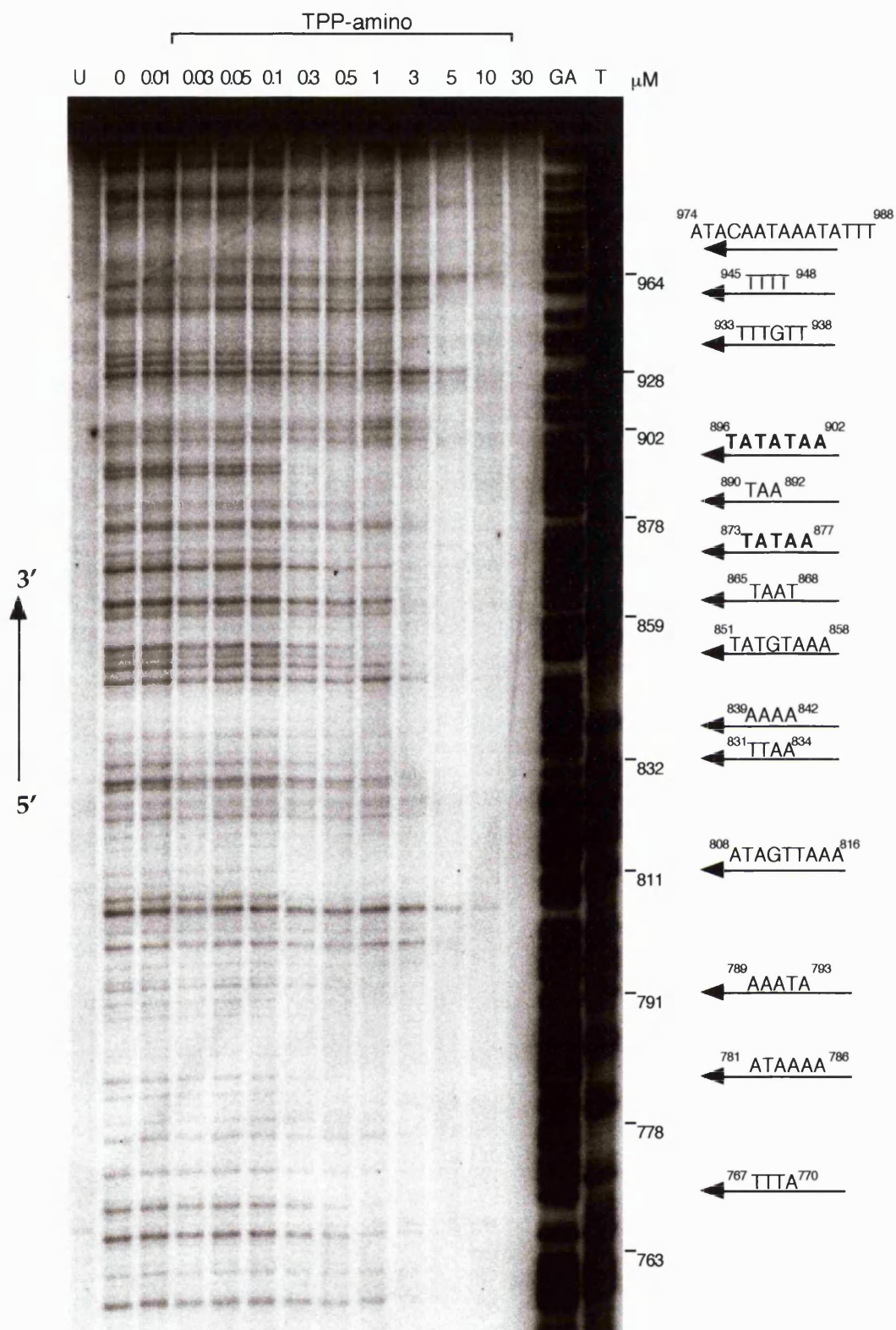
then summarised for each polyamide by diagrammatic representation as a composite box diagram, whereby the sites observed from multiple gels along both DNA strands were denoted as filled, hatched or open blocks alongside the sequence of the double stranded LTR fragment generated.

#### *Thiazole-pyrrole-pyrrole:*

The thiazole-pyrrole-pyrrole (TPP) series consisted of N-terminal amino, formyl and acetyl monomers as well as C7-linked N-terminal amino and acetyl homodimers (Figure 2.1(b)). Within this heterocycle series, the amino and formyl TPP monomers bound to identical sequence regions along both strands of the 5'LTR fragment, with binding sites centred at multiple runs of AT base pairs (Figures 2.3 and 2.4). TPP-amino, however, showed up to a ten-fold greater affinity for most sites compared to TPP-formyl. Nevertheless, the same order of site preference was observed for both ligands, with the strongest binding observed at the two TATA boxes (sites 873 & 896) present and at <sup>781</sup>5'-ATAAAA-3'<sup>786</sup>, <sup>789</sup>5'-AAATA-3'<sup>793</sup> and <sup>808</sup>5'-ATAGTTAAA-3'<sup>816</sup>. Diagrammatic representation of the binding produced by the amino and formyl TPP monomers is shown in Figure 2.5.

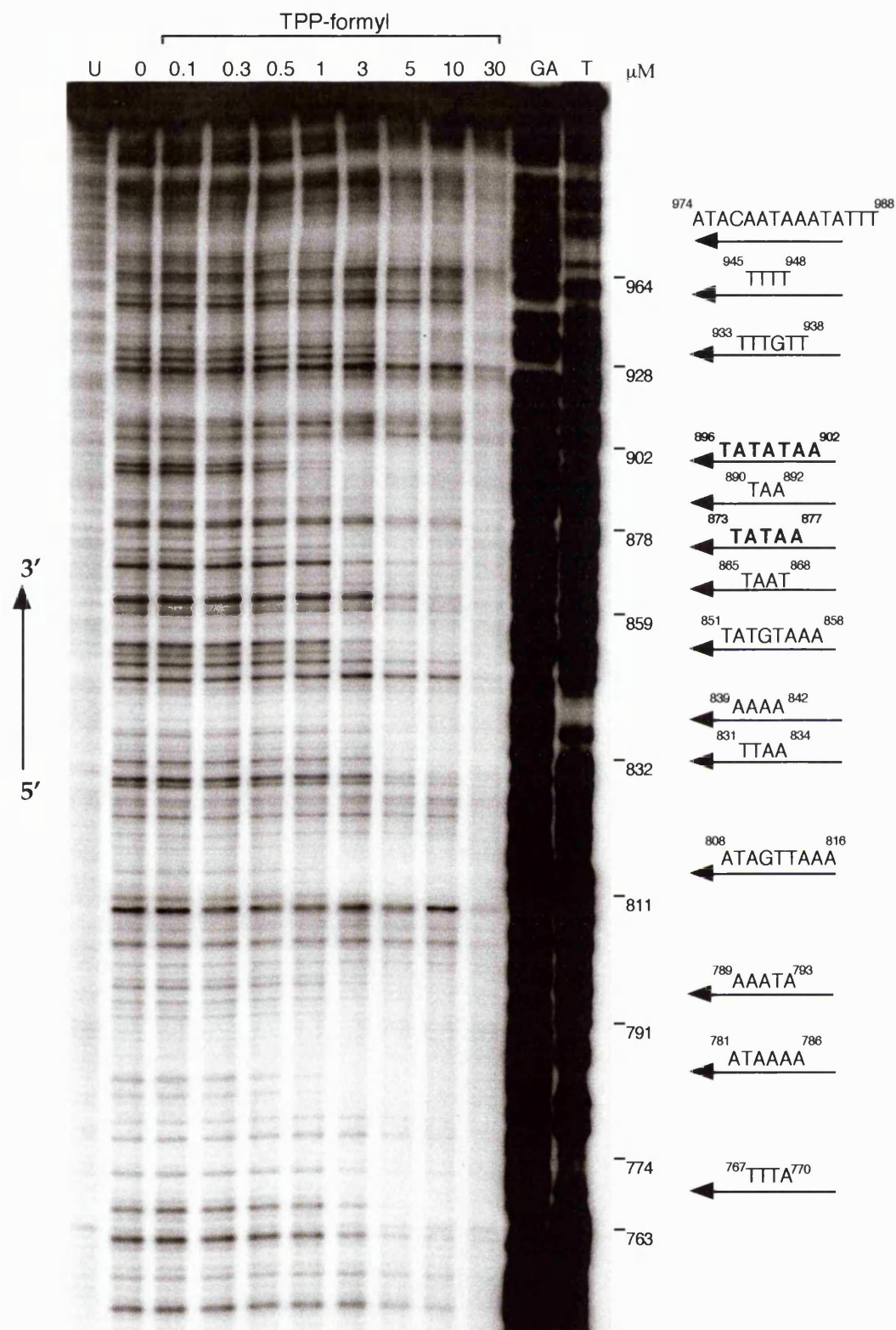
Linkage of N-terminal amino-TPP with a C7 methylene chain to produce a cross-linked dimer did not change the location of protected sites but did enhance binding at selected sites (Figure 2.6). A ten fold increase in binding was observed at <sup>839</sup>5'-AAAA-3'<sup>842</sup> and <sup>831</sup>5'-TTAA-3'<sup>834</sup>, as well as a six fold increase at <sup>933</sup>5'-TTTGTT-3'<sup>938</sup>, relative to the amino-TPP monomer. At the remaining majority of sites however, comparable binding strength was observed for the linked and unlinked TPP-amino. As observed for the two monomers, the strongest binding occurred at the two TATA boxes, <sup>781</sup>5'-ATAAAA-3'<sup>786</sup>, <sup>789</sup>5'-AAATA-3'<sup>793</sup> and <sup>808</sup>5'-ATAGTTAAA-3'<sup>816</sup> (Figure 2.7).

In contrast to both the amino and formyl TPP monomers, no binding was observed on either DNA strand with the TPP-acetyl monomer at concentrations of up to 30μM (Figure 2.8). Furthermore, linkage with a C7 linker did not improve binding, with no



**Figure 2.3** DNase I footprint titration with TPP-amino on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing multiple binding sites. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

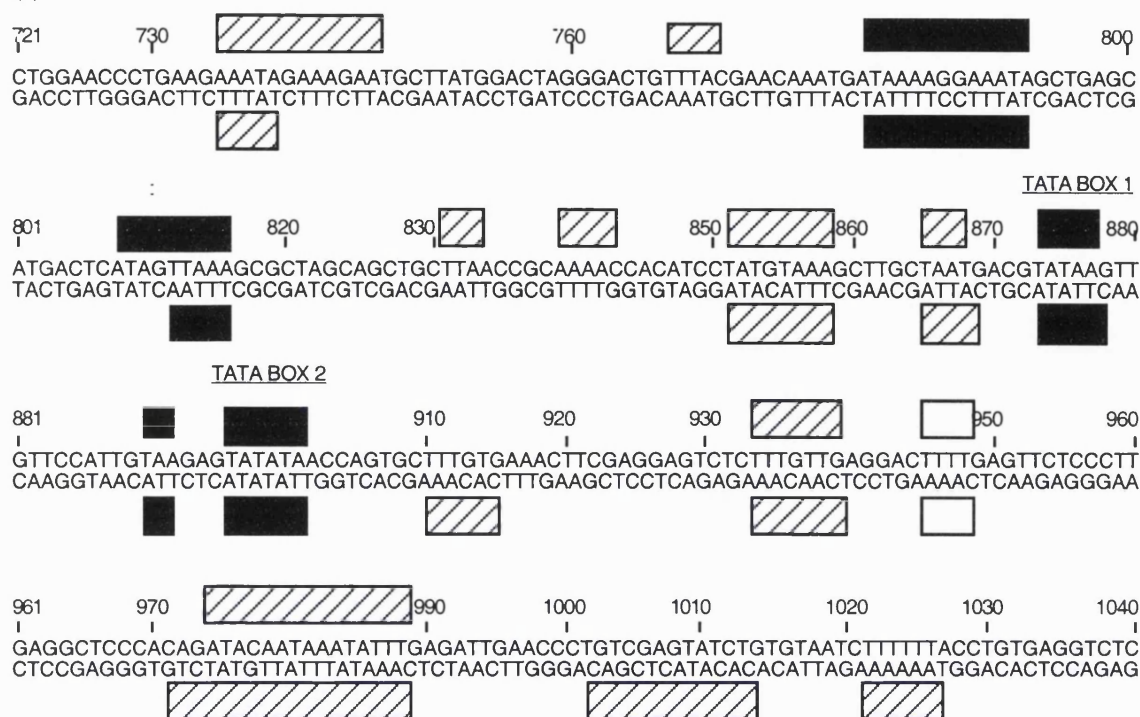




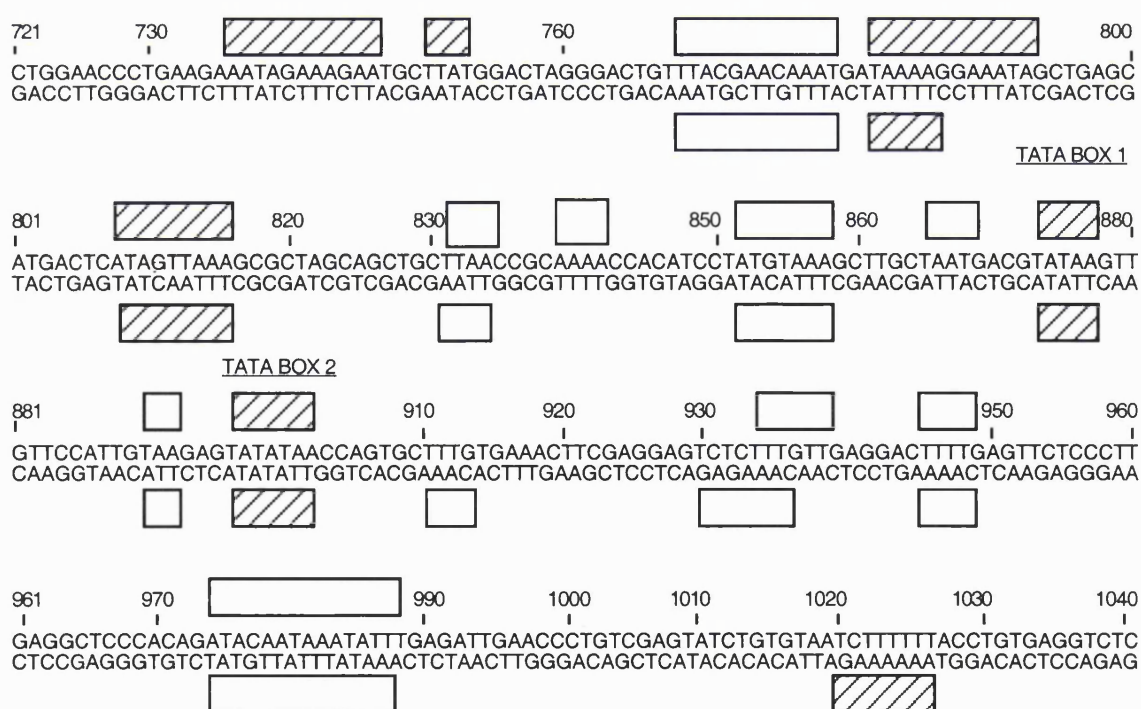
**Figure 2.4** DNase I footprint titration with TPP-formyl on the top strand of the 5'- $^{32}$ P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing multiple binding sites. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM  $MgCl_2$ , 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.



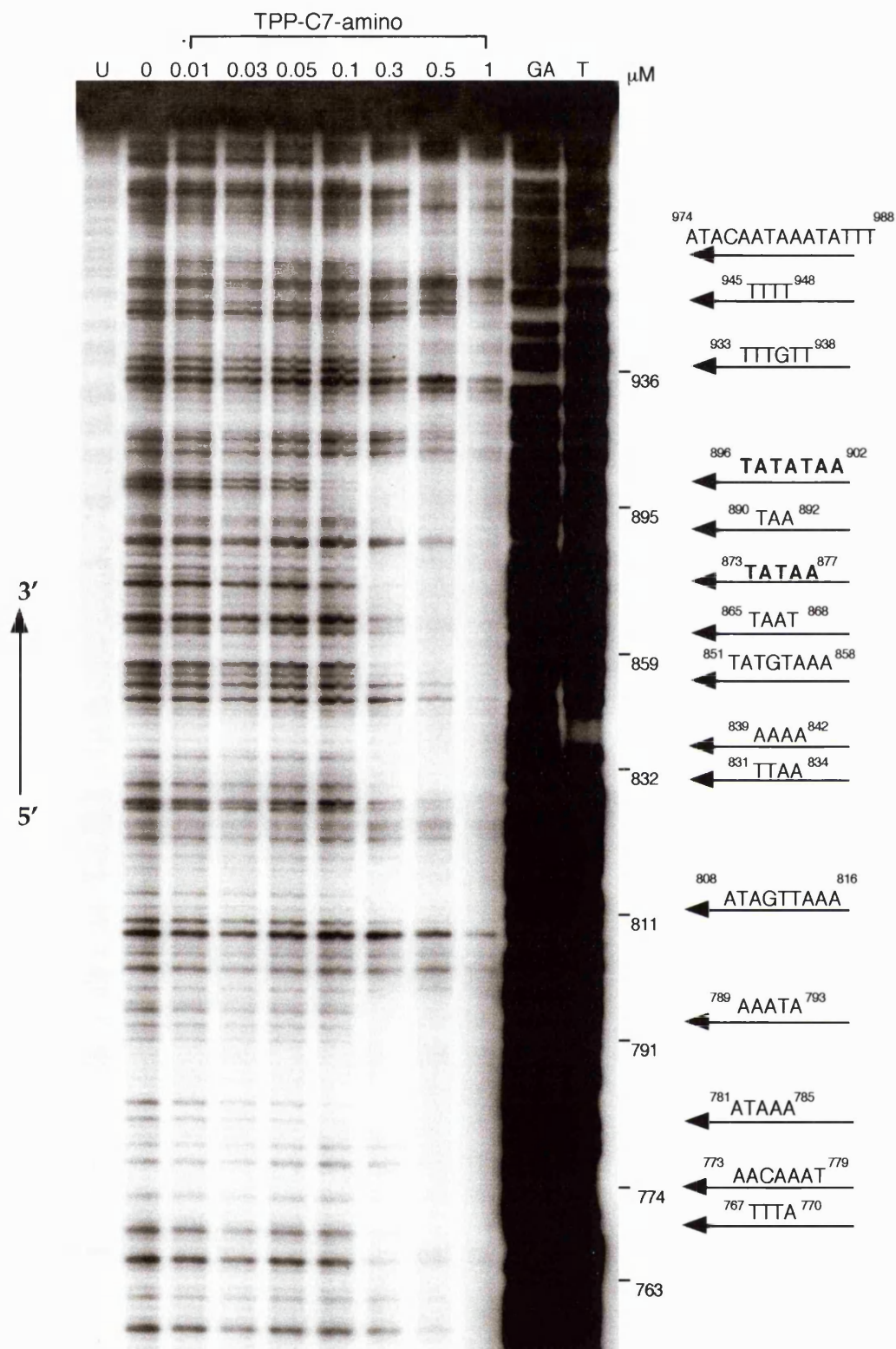
(i)



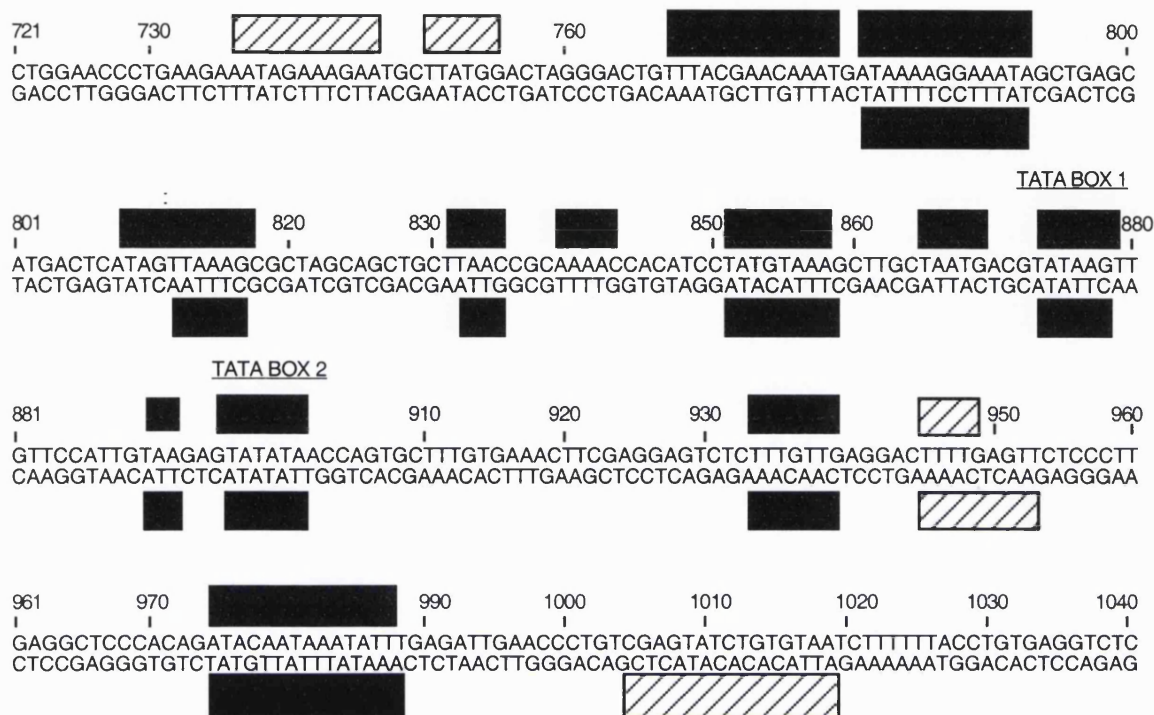
(ii)



**Figure 2.5** Box diagrams showing the multiple binding sites of (i) TPP-amino and (ii) TPP-formyl on the 355bp FIV34TF10 LTR. Open bars denote weak binding ( $\geq 5\mu\text{M}$ ), whilst hatched and filled bars denote moderate ( $\geq 1\mu\text{M}$ ) and strong ( $\geq 0.3\mu\text{M}$ ) binding.

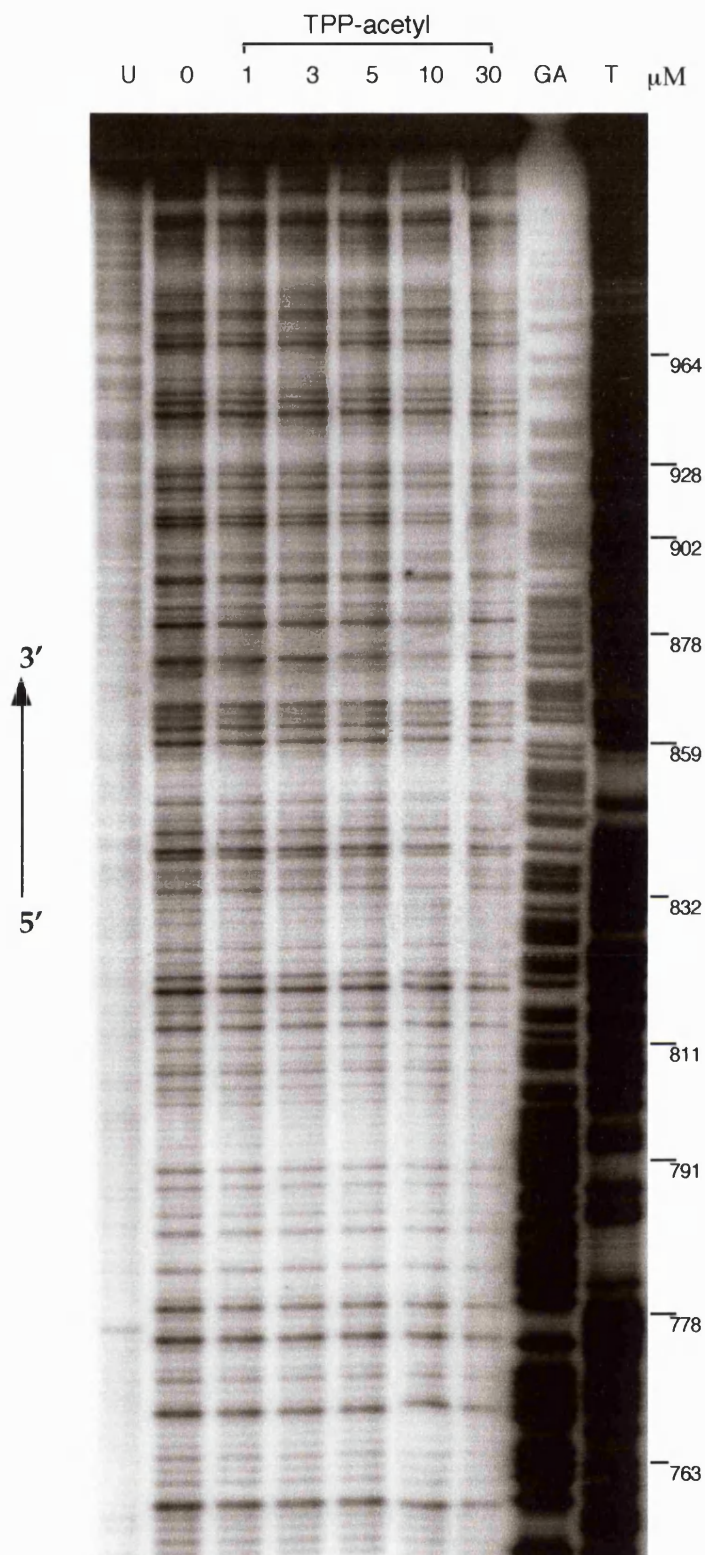


**Figure 2.6** DNase I footprint titration with TPP-C7-amino on the top strand of 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing multiple binding sites. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

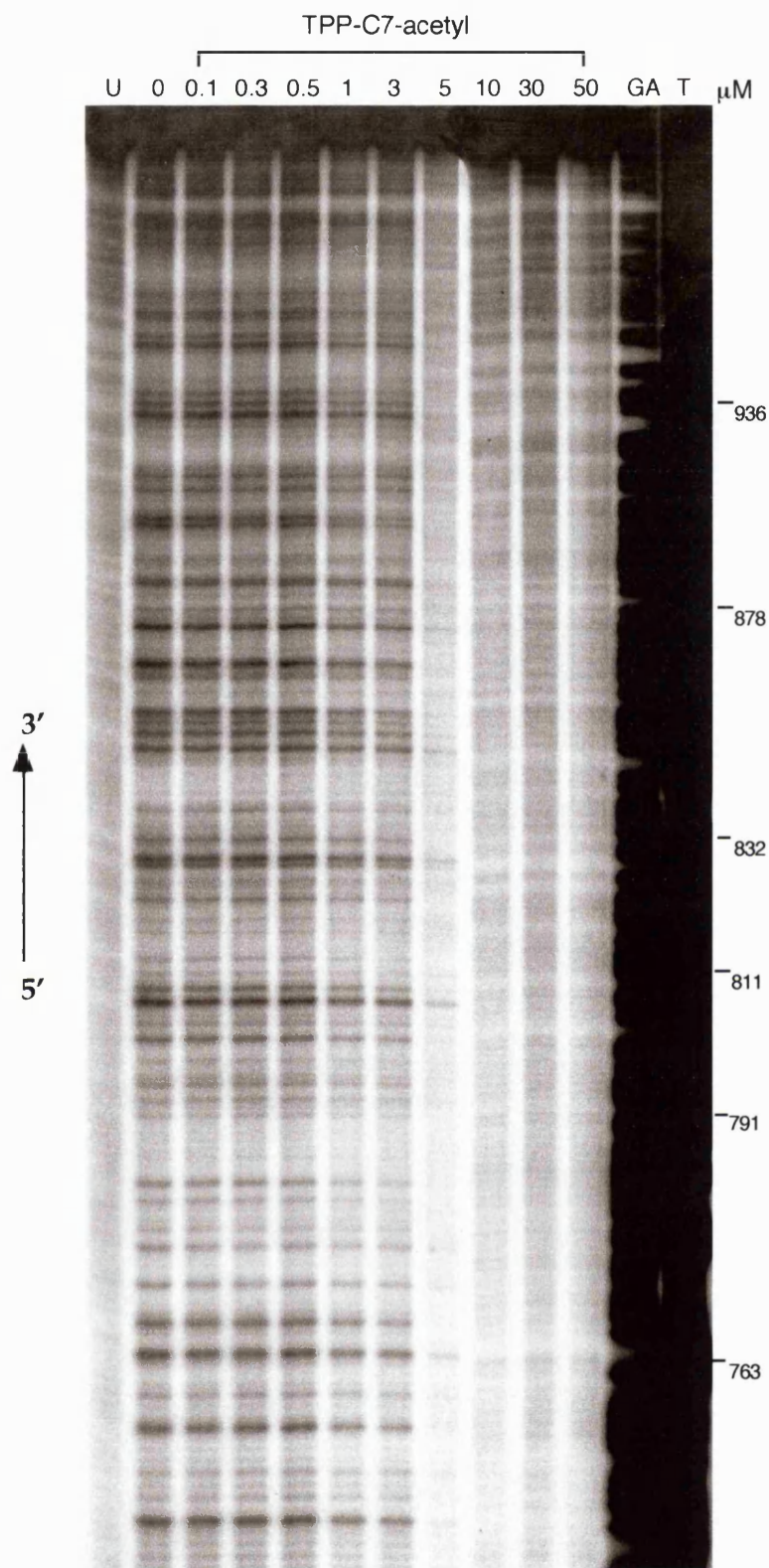


**Figure 2.7** Box diagram showing the multiple binding sites of TPP-C7-amino on the 355bp FIV34TF10 LTR fragment. Hatched bars denote moderate binding ( $\geq 1\mu\text{M}$ ) whilst filled bars denote strong binding ( $\geq 0.3\mu\text{M}$ ).





**Figure 2.8** DNase I footprint titration with TPP-acetyl on the top strand of the 5'- $^{32}\text{P}$  labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing no binding. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM  $\text{MgCl}_2$ , 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.



**Figure 2.9** DNase I footprint titration with TPP-C7-acetyl on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing no binding. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

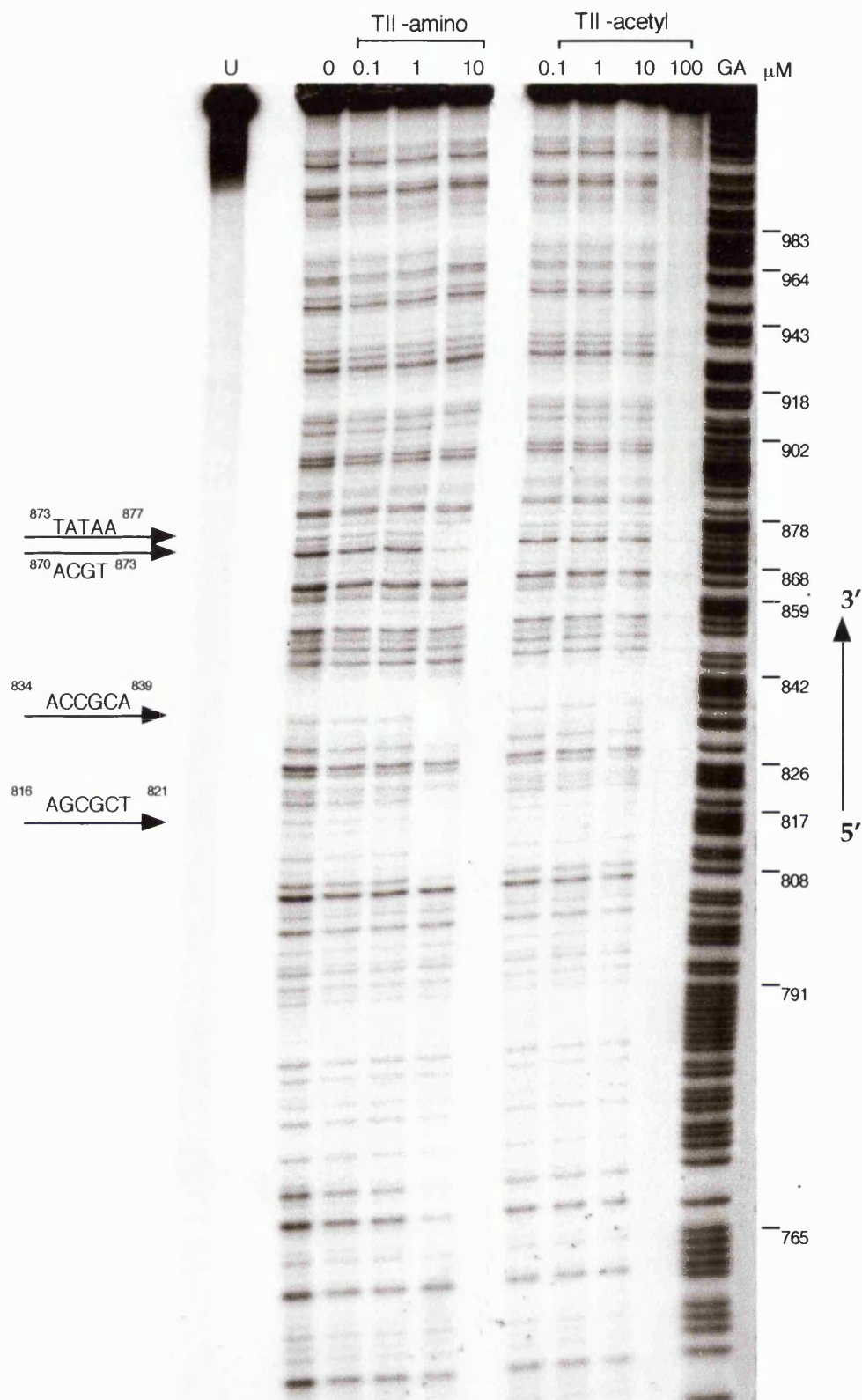
discrete footprints observed and complete loss of the DNase I cleavage pattern at concentrations above 5 $\mu$ M (Figure 2.9).

#### *Thiazole-imidazole-imidazole:*

The thiazole-imidazole-imidazole (TII) heterocycle series was composed of only two monomers, N-terminal amino and N-terminal acetyl-TII (Figure 2.1 (a)). TII-amino showed predominant binding at three putative sites at a concentration of 10 $\mu$ M, as shown in Figure 2.10 lanes 3-5. Two of these were one-residue stagger mismatch sites for this heterocycle combination centred at GC runs and consisted of four GC base pairs flanked either side by AT runs (<sup>816</sup>5'-AGCGCT-3'<sup>821</sup>, <sup>834</sup>5'-ACCGCA-3'<sup>839</sup>). Weak binding also appeared to occur at TATA box 1, with protection extending into the adjacent site (<sup>870</sup>5'-ACGT-3'<sup>873</sup>), however it was difficult to discriminate between these two sites. Binding sites along both strands of the LTR region are indicated in Figure 2.11. The TII-acetyl monomer failed to demonstrate binding within the FIVLTR (Figure 2.10 lanes 6-9).

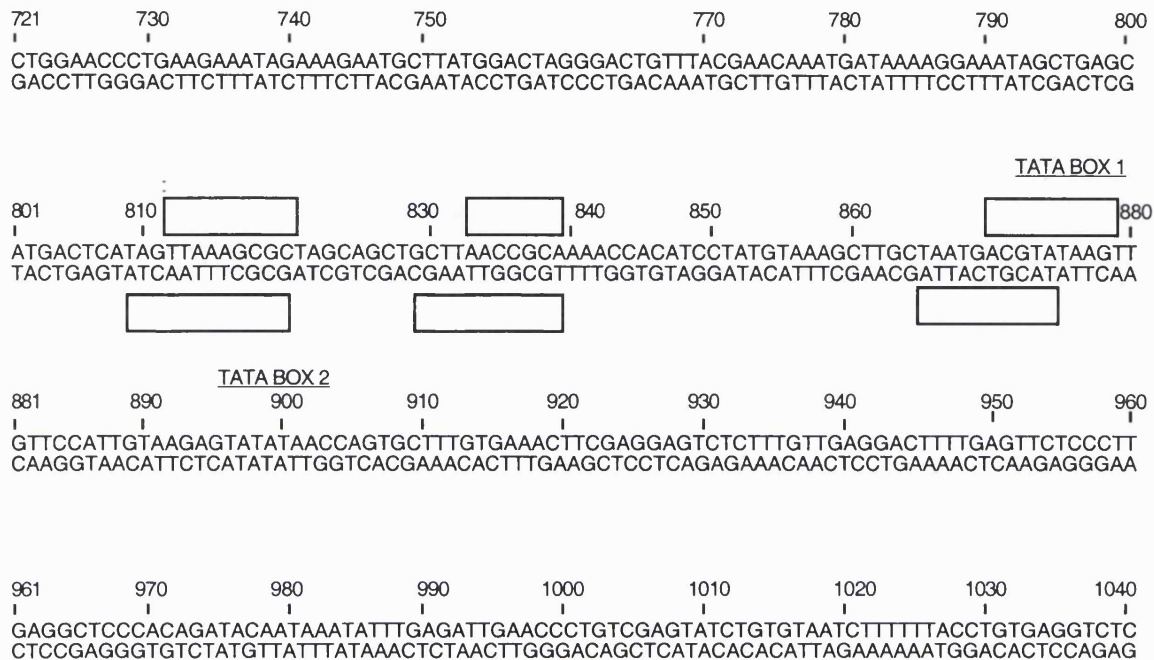
#### *Thiazole-pyrrole-imidazole:*

The thiazole-pyrrole-imidazole (TPI) series was comprised of an N-terminal formyl TPI monomer and C5 linked dimer (Figure 2.1 (a)). The TPI formyl monomer bound to sites along the LTR fragment centred at runs of AT base pairs at concentrations of 3 $\mu$ M and above, as shown in Figure 2.12. The strongest binding observed at the two TATA boxes in addition to a further site <sup>808</sup>5'-ATAGTTAAAGCGC-3'<sup>820</sup>. Linkage of two formyl-TPI monomers with a C5 methylene linker resulted in binding at the same sites favoured by the monomer, however binding strength of the ligand increased by a factor of ten at all sites (Figure 2.13). Detailed binding of the monomer and dimer along both strands of the FIV LTR region is represented as a box diagram in Figure 2.14. Interestingly, the pattern of binding observed with the TPI-formyl monomer was very similar to that seen with the corresponding TPP-formyl monomer. Subsequent comparative footprinting of these two compounds confirmed a shared site preference (Figure 2.15). Two to three-fold weaker binding was observed at favoured binding sites for TPI-formyl relative to TPP-formyl. Furthermore, around the sites



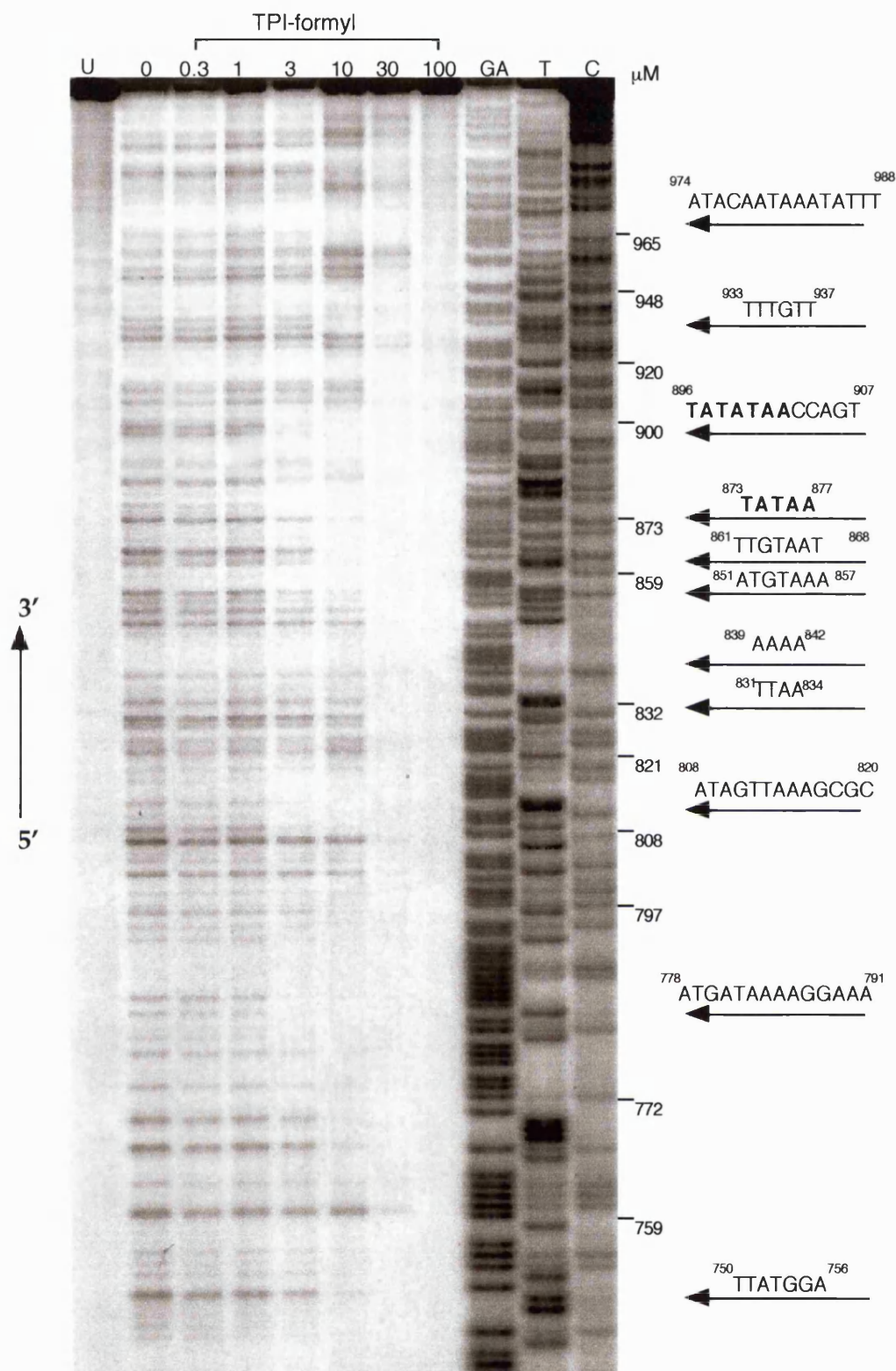
**Figure 2.10** DNase I footprint titration with TII-amino and TII-acetyl on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.



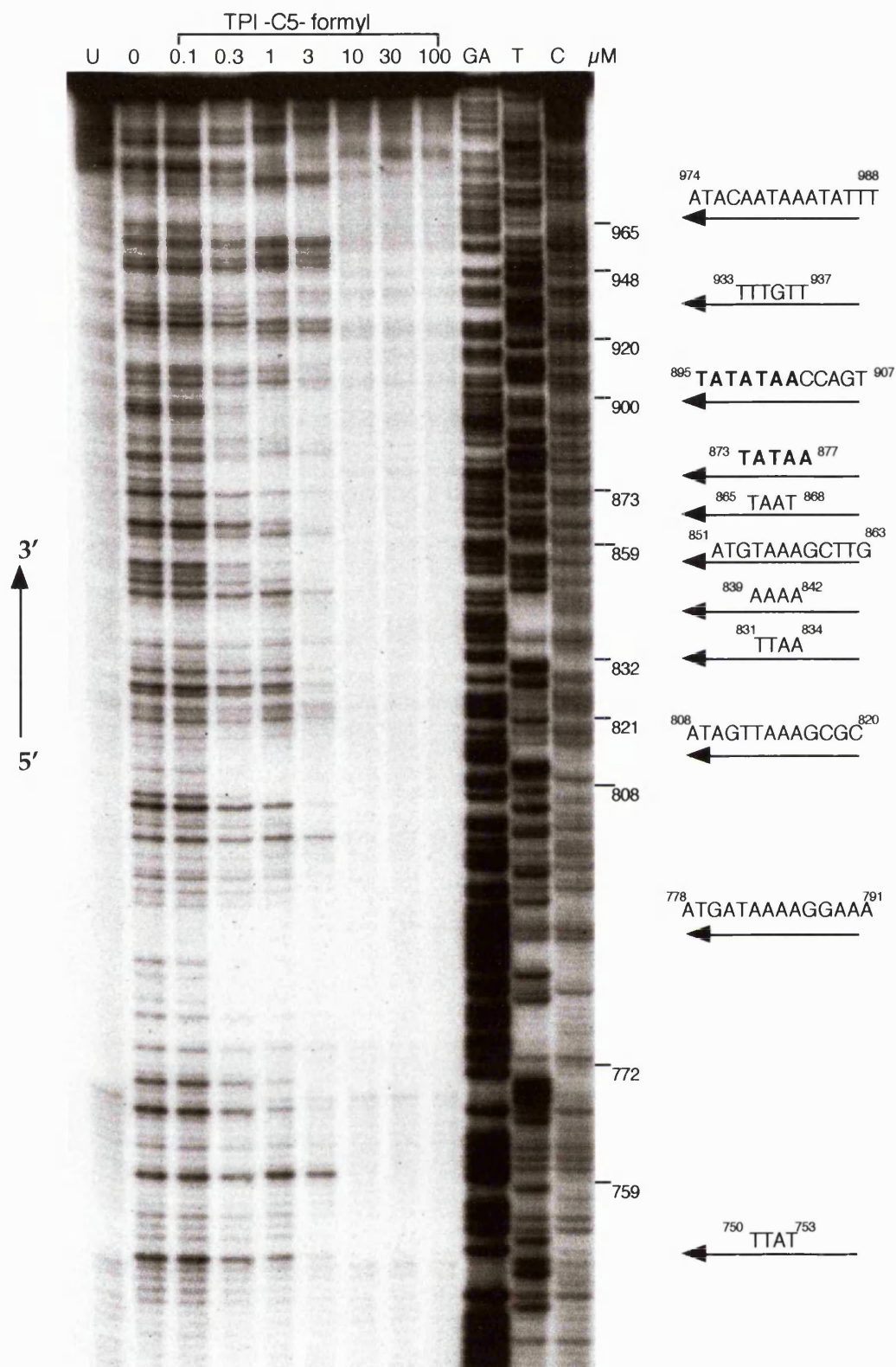


**Figure 2.11** Box diagram showing the binding sites of TII-amino on the 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10. Open bars denote weak binding ( $\geq 5\text{mM}$ ).



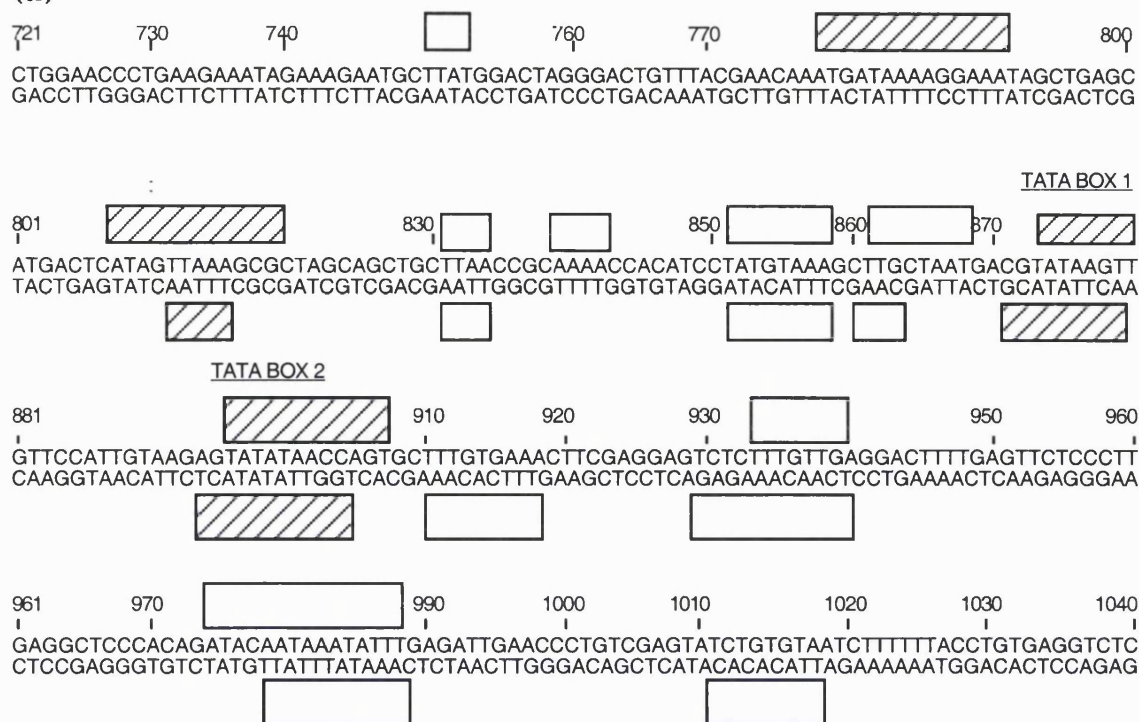


**Figure 2.12** DNase I footprint titration with TPI-formyl on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing multiple binding sites. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction, T is the thymine marker reaction and C is the cytosine marker reaction.

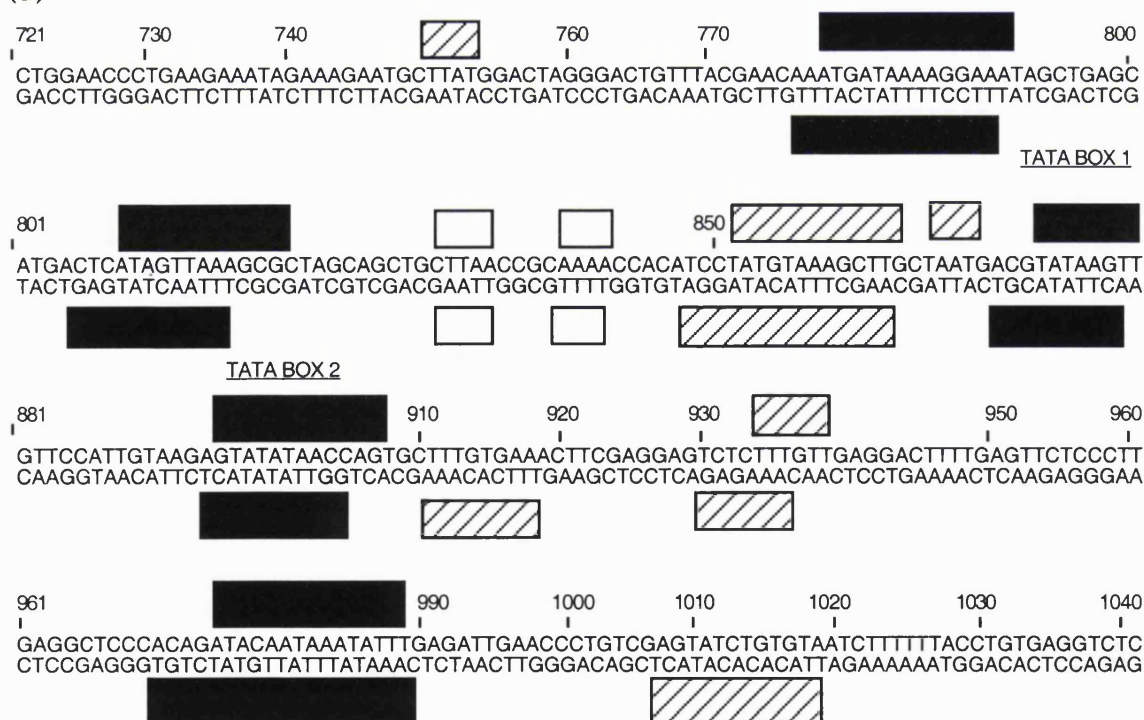


**Figure 2.13** DNase I footprint titration with TPI-C5-formyl on the top strand of the 5'- $^{32}\text{P}$  labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing multiple binding sites. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM  $\text{MgCl}_2$ , 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction, T is the thymine marker reaction and C is the cytosine marker reaction.

(a)

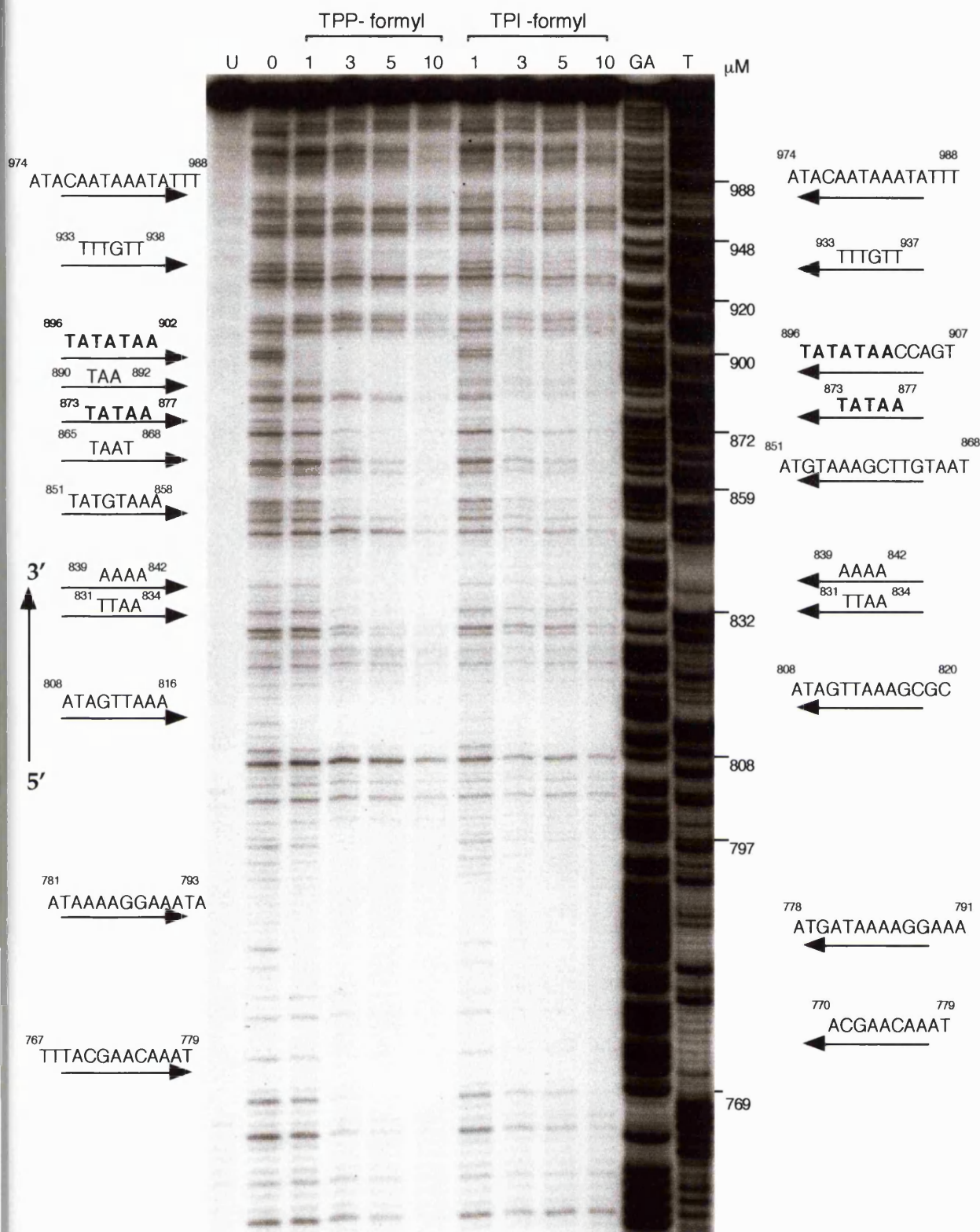


(b)



**Figure 2.14** Box diagrams showing the multiple binding sites of (a) TPI-formyl and (b) TPI-C5-formyl on the 355bp FIV34TF10 LTR. Open bars denote weak binding ( $\geq 5\mu\text{M}$ ), whilst hatched and filled bars indicate moderate ( $\geq 1\mu\text{M}$ ) and strong binding ( $\geq 0.3\mu\text{M}$ ).





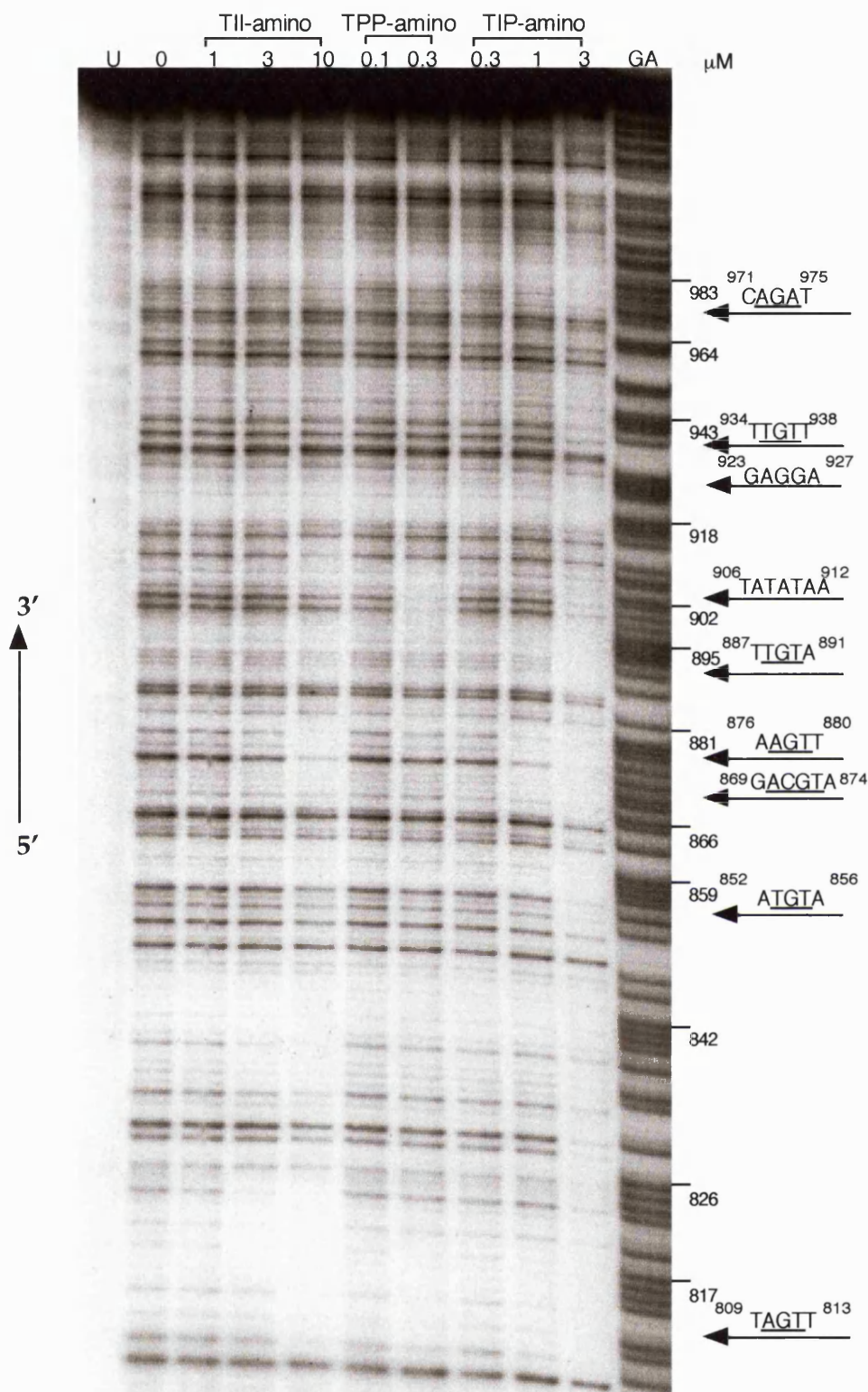
**Figure 2.15** DNase I footprint titration with TPP-formyl and TPI-formyl on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing multiple binding sites. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

<sup>781</sup>5'-ATAAAAGGAAATA-3'<sup>793</sup> and <sup>767</sup>5'-TTTACGAACAAAT-3'<sup>779</sup> as observed for TPP-formyl, TPI-formyl was less selective for the each AT run producing a larger footprint extending from 770 to 791. These two sites could be resolved separately for TPP-formyl.

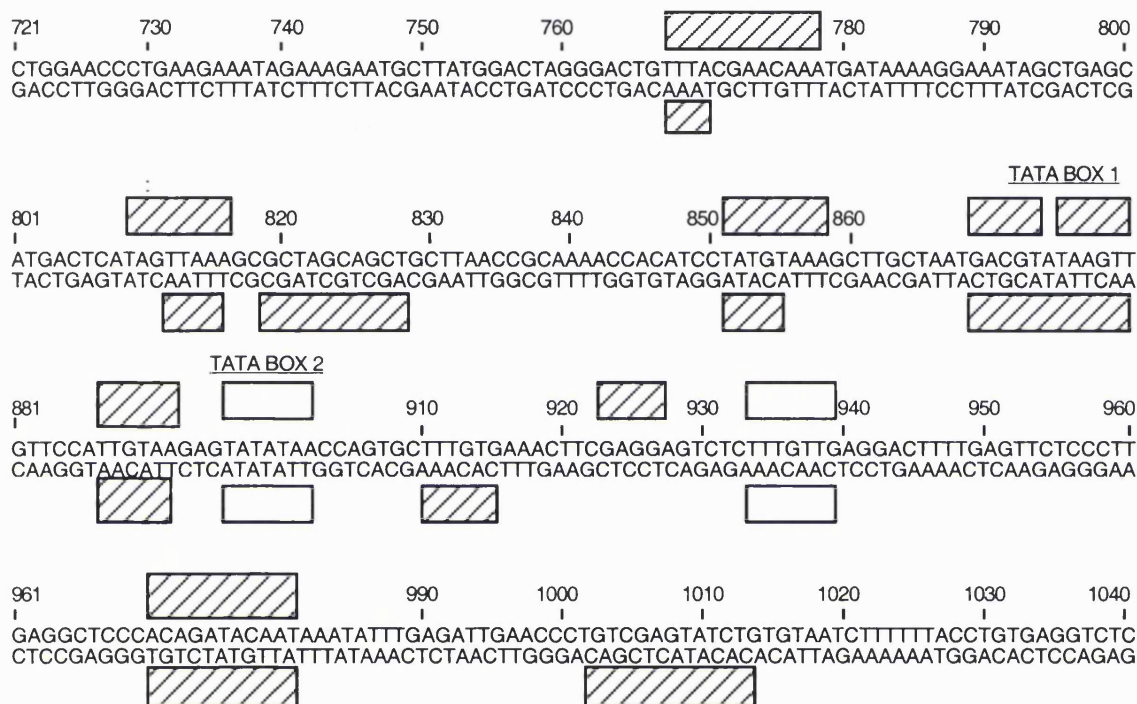
#### *Thiazole-imidazole-pyrrole:*

The near complete group of thiazole-imidazole-pyrroles (TIPs) was composed of unlinked N-terminal amino and acetyl monomers as well as N-terminal hydrogen, amino, formyl or acetyl dimers cross-linked with a five to eight carbon methylene linker (Figure 2.1 (b)). Within the N-terminal amino group, discrete binding was observed with the TIP monomer at a number of sites, at concentrations of 1-3 $\mu$ M (Figure 2.16). These included a maximum overlap match site, <sup>809</sup>5'-TAGTT-3'<sup>813</sup>, as well as four maximum overlap mismatches, <sup>852</sup>5'-ATGTA-3'<sup>856</sup>, <sup>887</sup>5'-TIGTA-3'<sup>891</sup>, <sup>934</sup>5'-TTGTT-3'<sup>938</sup> and <sup>971</sup>5'-CAGAT-3'<sup>975</sup>. Binding was also observed at a single one-residue stagger mismatch site, <sup>869</sup>5'-GACGTA-3'<sup>874</sup>, as well as at a further overlap match site <sup>876</sup>5'-AAGTT-3'<sup>880</sup>. Weak binding was also observed at TATA box 2 <sup>906</sup>5'-TATATAA-3'<sup>912</sup> and at a further site <sup>923</sup>5'-GAGGA-3'<sup>927</sup>. A diagrammatic summary of binding is shown in Figure 2.17. In contrast to the TPP and TPI polyamides, TIP-amino did not share a binding preference with the other heterocycle combinations examined as demonstrated by comparative footprinting of TIP-amino with the N-terminal amino-TII and TPP monomers, showing a distinct pattern of nuclease protection (Figure 2.16). In common with the other heterocycle combinations, however, replacement of the N-terminal amino head group with an acetyl resulted in an abrogation of discrete binding as shown in Figure 2.18.

Cross-linkage of TIP-amino with a five, six, seven or eight carbon chain retained the site preference observed with the monomer. TIP-C5-amino also produced weak protection at these sites, binding with comparable strength to that of the TIP-amino at most sites (Figure 2.19). Use of a six carbon linker resulted in a small increase in binding at most sites (Figure 2.19), however, lengthening the linker to seven or eight

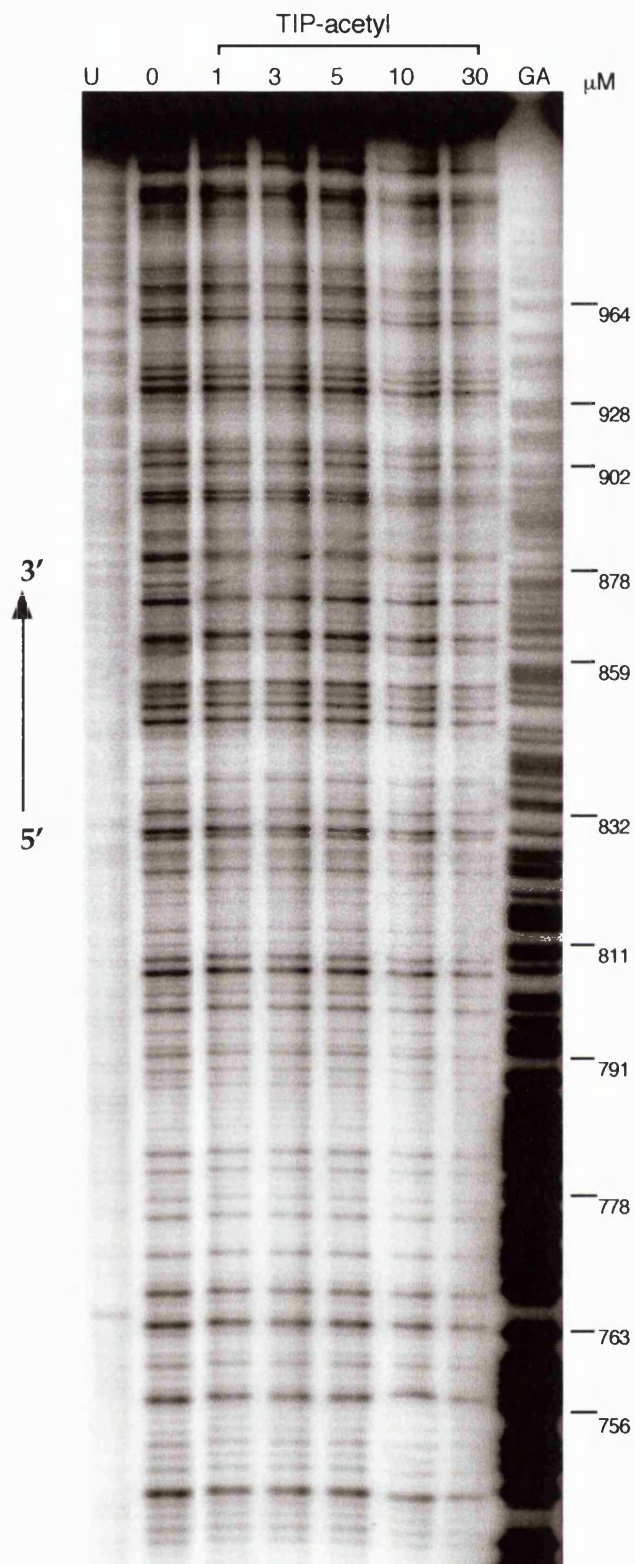


**Figure 2.16** DNase I footprint titration with TII-amino, TPP-amino and TIP-amino on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing binding sites for TIP-amino. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control and GA is the purine marker reaction.



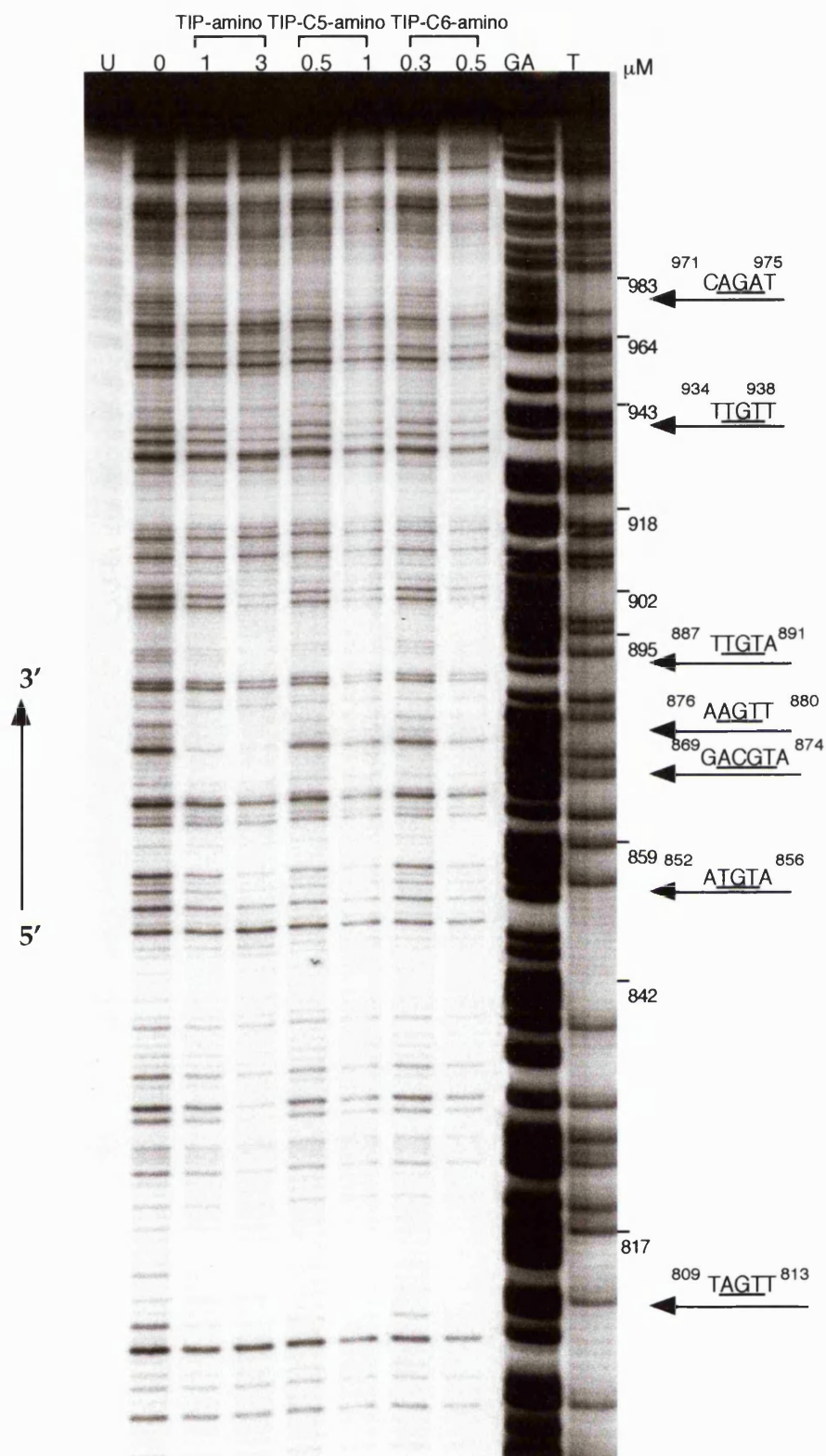
**Figure 2.17** Box diagram showing the multiple binding sites of TIP-amino on the 355bp FIV34TF10 LTR fragment. Open bars denote weak binding ( $\geq 5\mu\text{M}$ ) whilst hatched indicate moderate binding ( $\geq 1\mu\text{M}$ ).



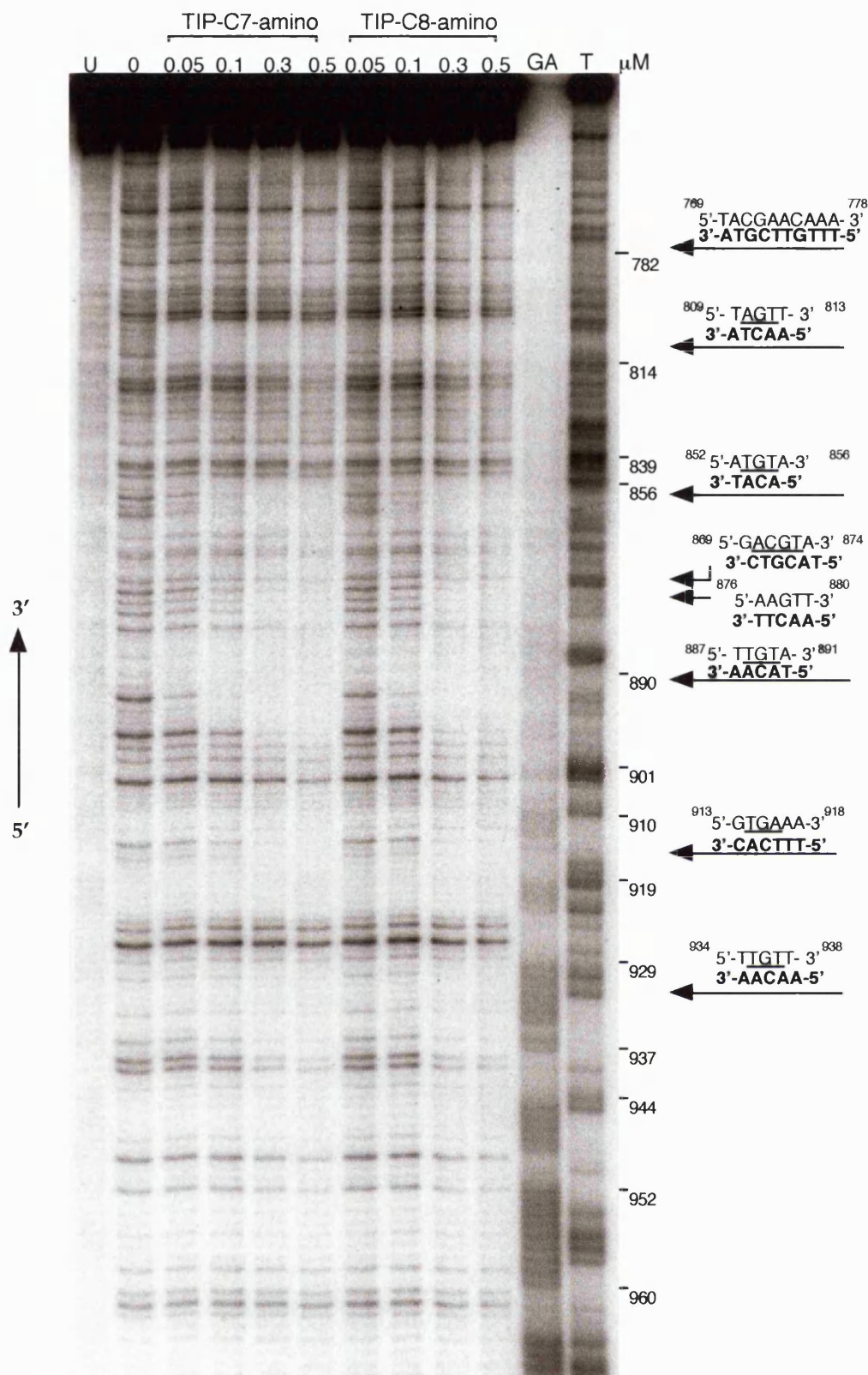


**Figure 2.18** DNase I footprint titration with TIP-acetyl on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing no binding. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

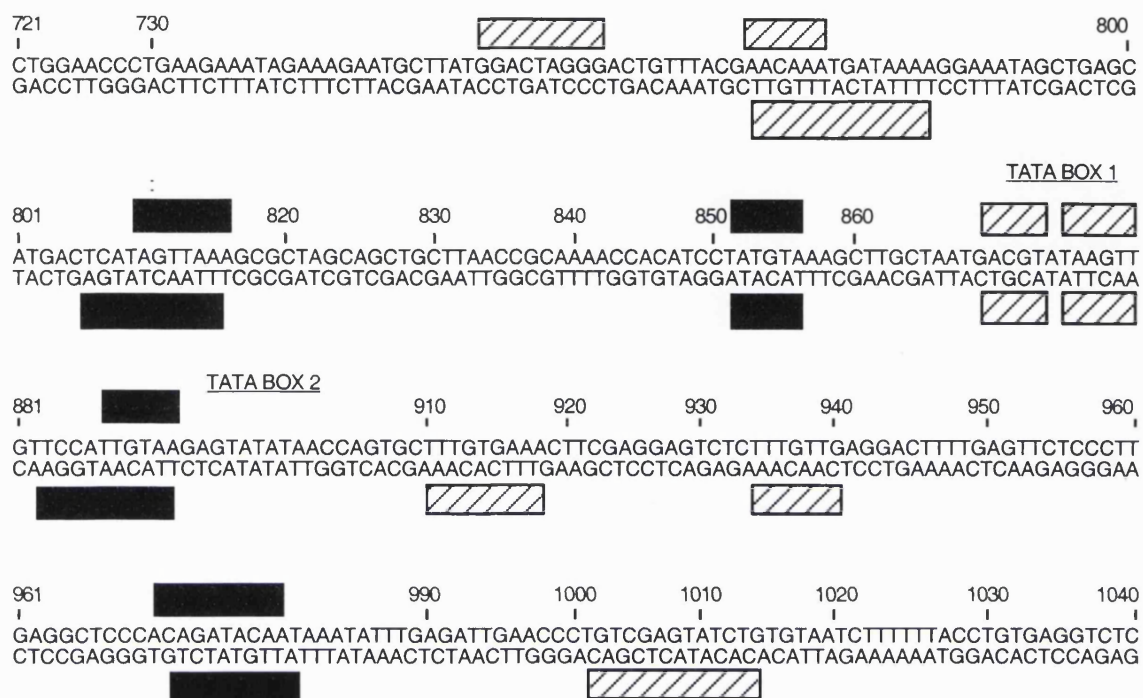




**Figure 2.19** DNase I footprint titration with TIP-amino, TIP-C5-amino and TIP-C6-amino on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing discrete binding. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

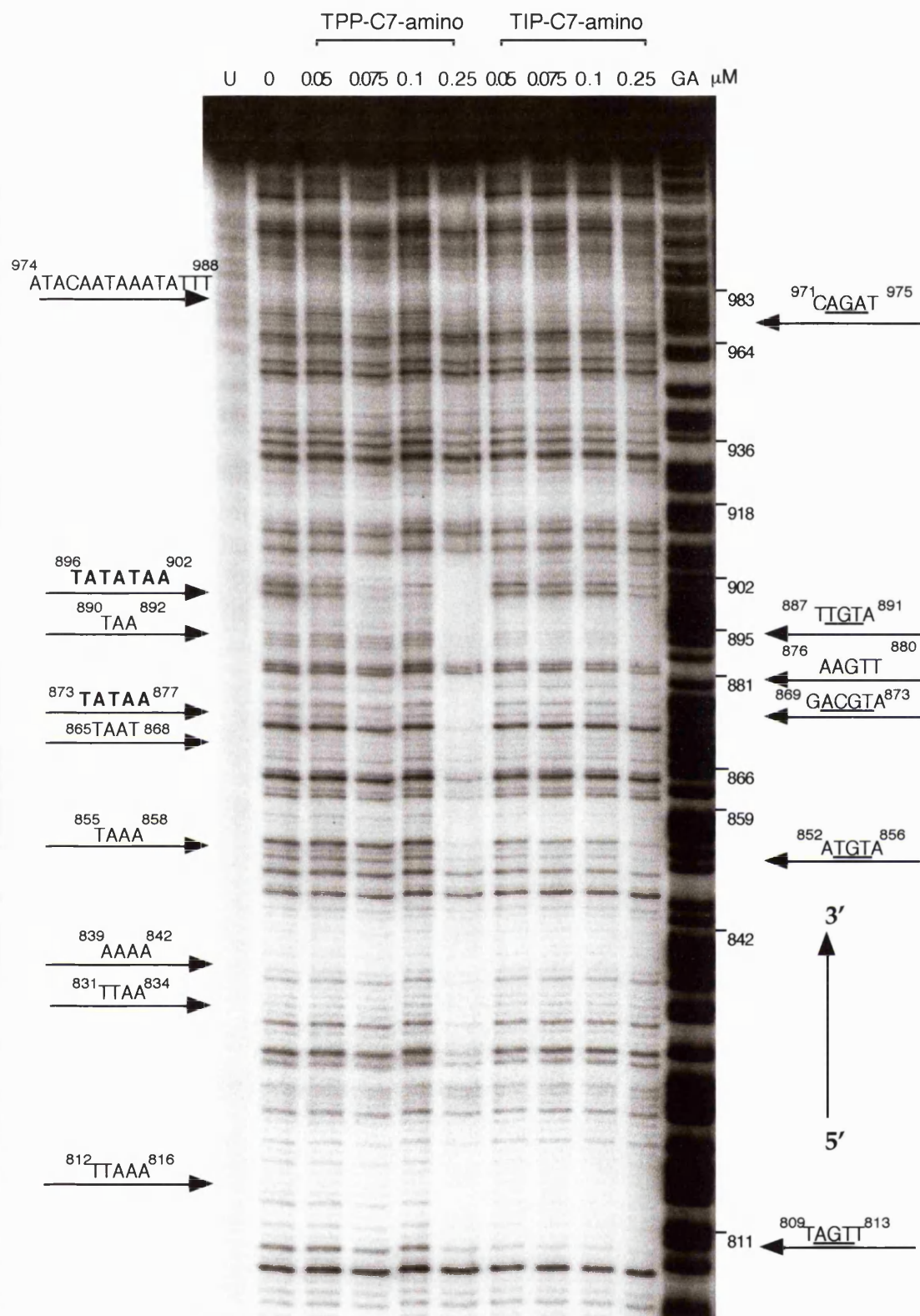


**Figure 2.20** DNase I footprint titration with TIP-C7-amino and TIP-C8-amino on the bottom strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, with binding sites for both compounds on this strand indicated in bold type alongside the corresponding top strand sequence. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

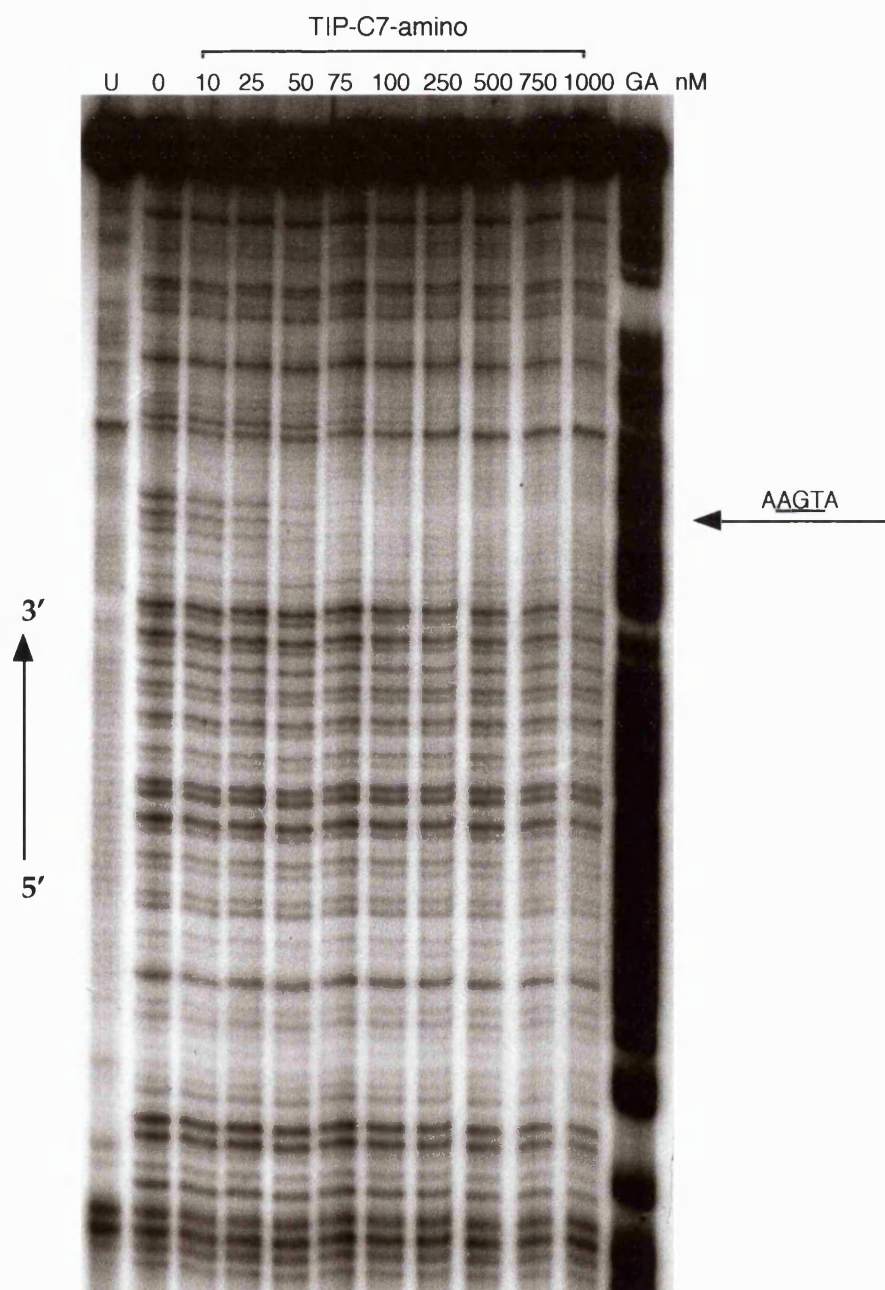


**Figure 2.21** Box diagram showing the multiple binding sites of TIP-C7-amino on the 355bp FIV34TF10 LTR. Hatched bars denote moderate binding ( $\geq 0.3\mu\text{M}$ ) whilst filled bars denote strong binding ( $\geq 0.1\mu\text{M}$ ).

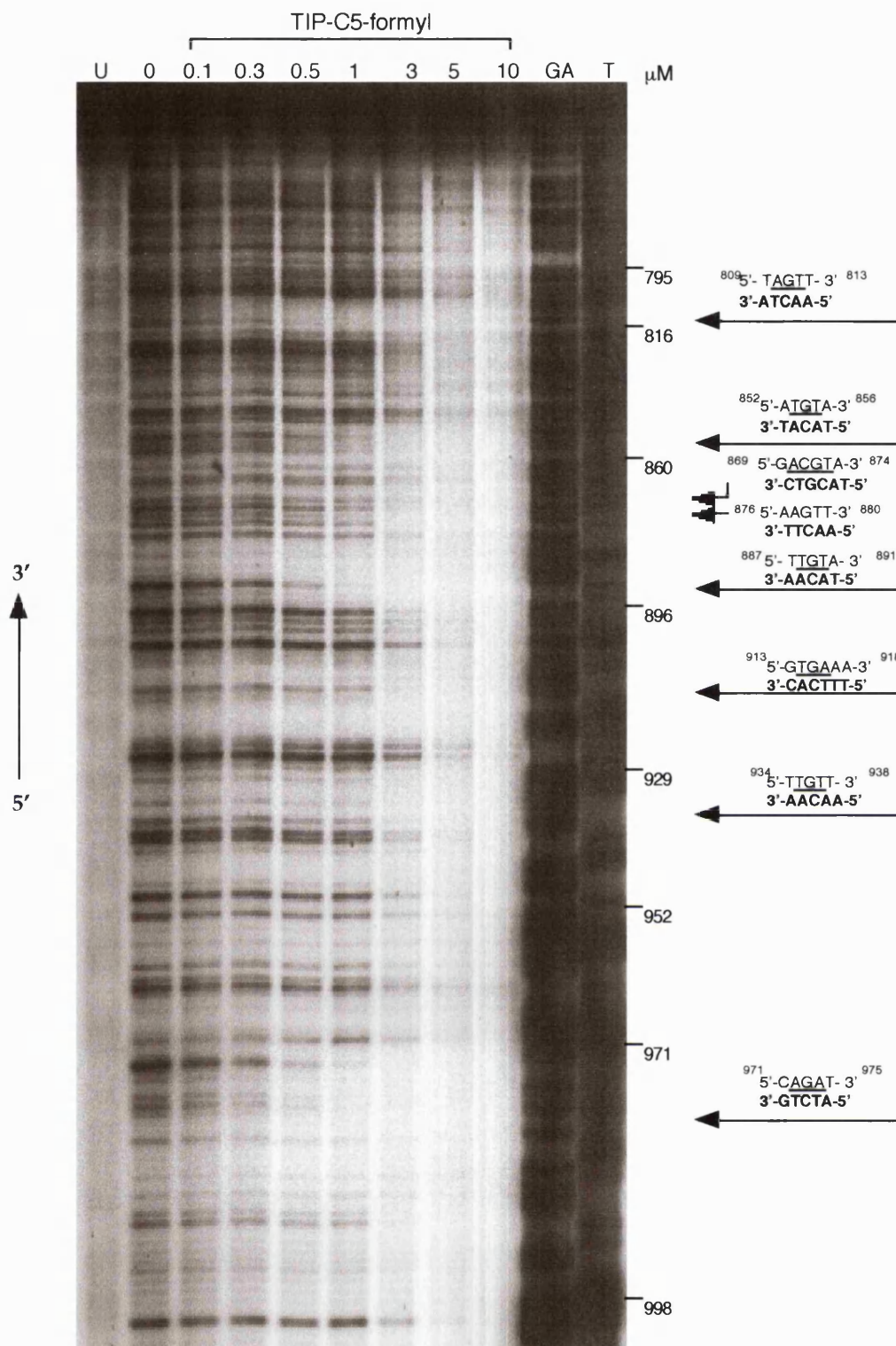




**Figure 2.22** DNase I footprint titration with TPP-C7-amino and TIP-C7-amino on the top strand of 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing binding sites for both compounds. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control and GA is the purine marker reaction.



**Figure 2.23** DNase I footprint titration with TIP-C7-amino on the top strand of 5'-<sup>32</sup>P labelled 192 bp fragment of the Her2/neu promoter region, showing a single binding site. All reactions contain 1000cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control and GA is the purine marker reaction.

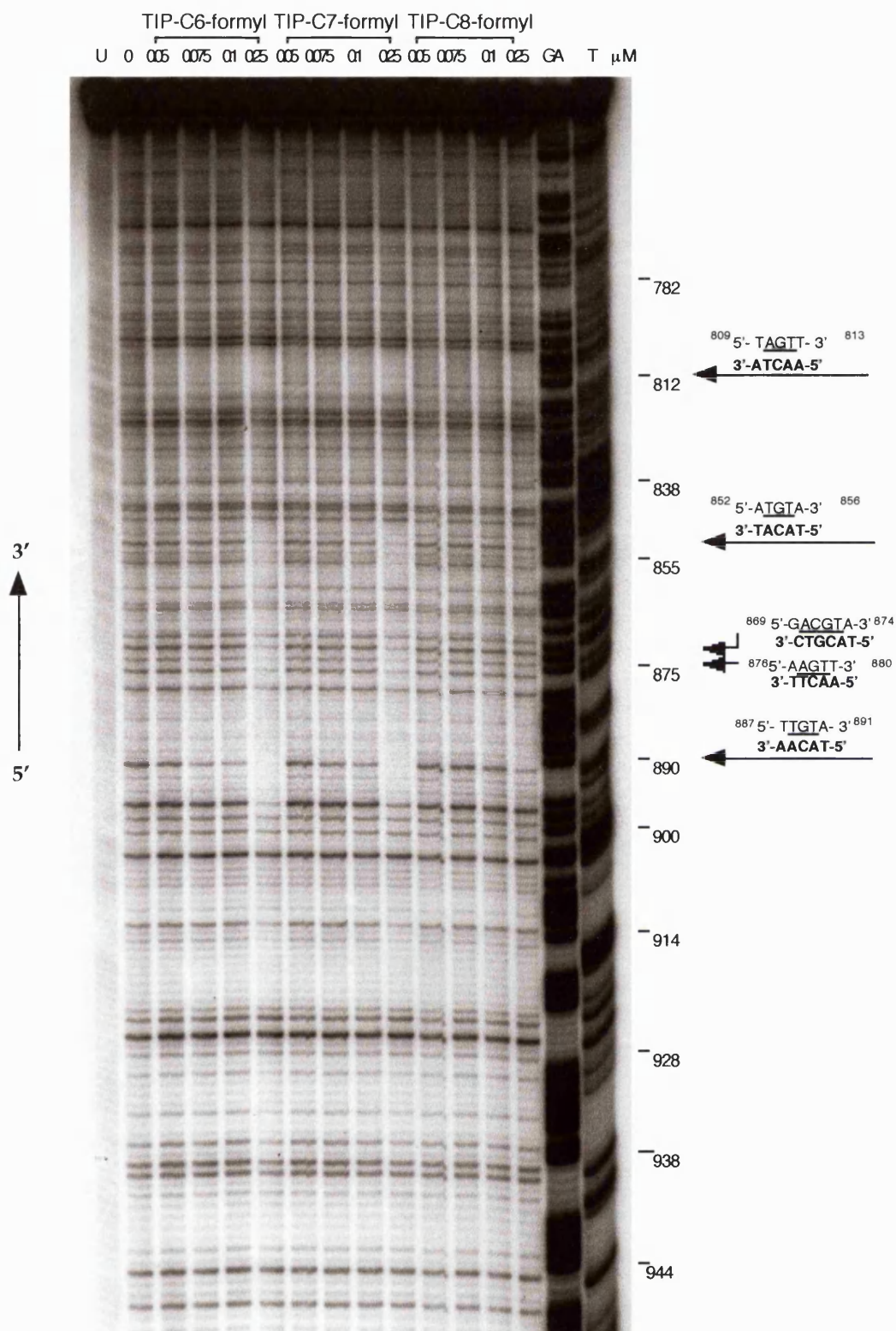


**Figure 2.24** DNase I footprint experiment with TIP-C5-formyl on the bottom strand of 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, with binding sites on this strand indicated in bold type alongside the corresponding top strand sequence. All reactions contain 1000cps of restriction fragment, 10mM Tris pH7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

carbons improved binding by up to ten fold, resulting in DNA protection at submicromolar concentrations (Figure 2.20). The C7 and C8 TIP-amino homodimers showed similar binding strength across most sites, although weaker binding was observed at <sup>869</sup>5'-GACGTA-3'<sup>874</sup> for TIP-C7-amino relative to TIP-C8-amino (Figure 2.20). Binding sites produced by TIP-C7-amino along both strands of the LTR region are shown in Figure 2.21. Comparison of the C7 linked amino-TIP and amino-TIP confirmed the difference in the protection sites for the two cross-linked heterocycle combinations, with TIP-C7-amino binding over extended AT runs and TIP-C7-amino binding at discrete sites (Figure 2.22). TIP-C7-amino also displayed exclusive binding at submicromolar concentrations, at a single maximum overlap match site, 5'-AAGTA-3', in the promoter region of human Her 2/neu (Figure 2.23).

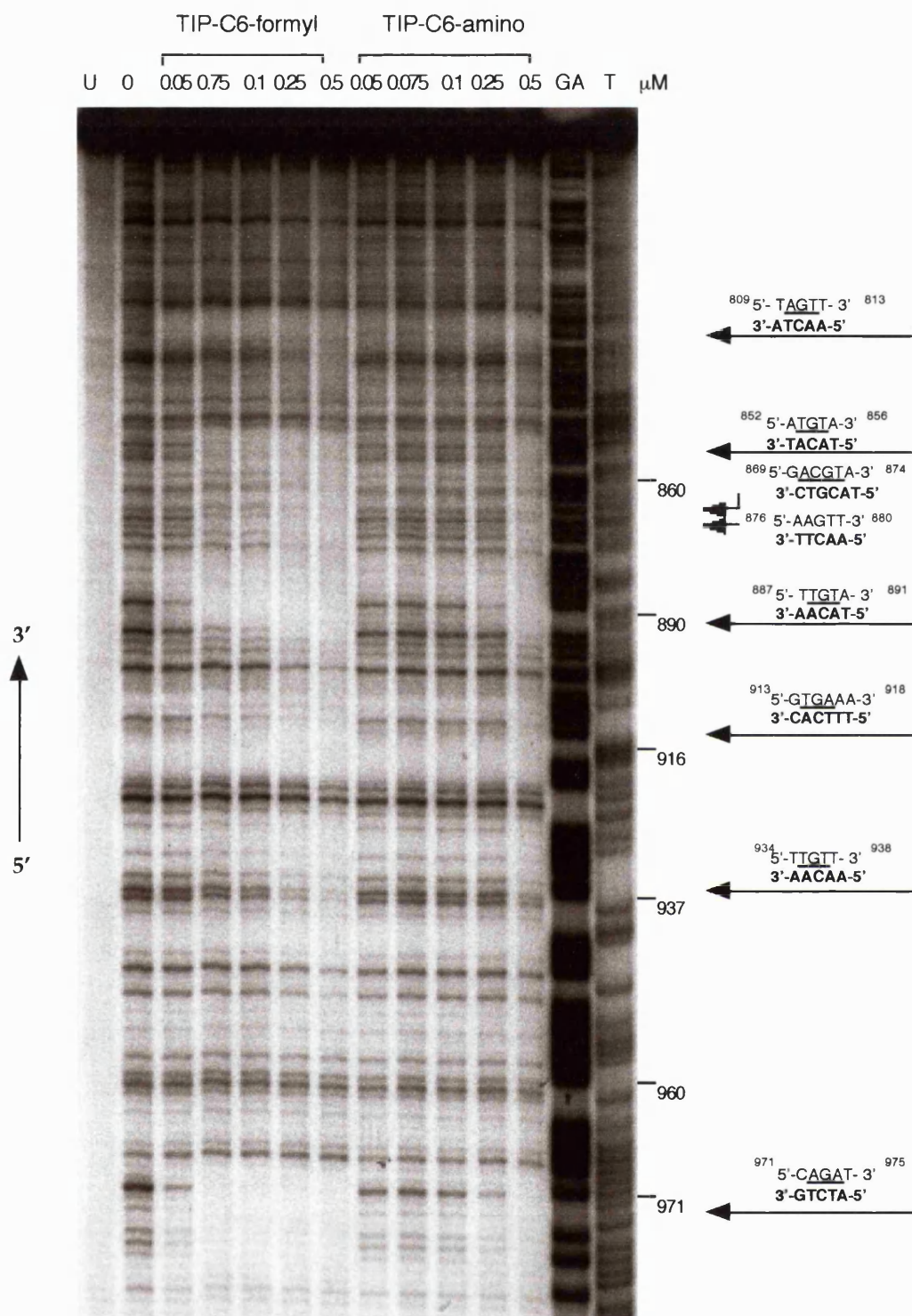
The N-terminal formyl-TIP series shared binding site preference with the amino-TIPs. As found in the amino-TIP series, the five carbon linked formyl-TIP was the weakest binder within the linked formyl-TIPs. TIP-C5-formyl bound most strongly at the maximum overlap match site <sup>809</sup>5'-TAGTT-3'<sup>813</sup> and the overlap mismatch <sup>971</sup>5'-CAGAT-3'<sup>975</sup> at concentrations of 0.5µM and above (Figure 2.24). Extension of the linker to six or seven carbons significantly increased the binding strength at the five principal protection sites, <sup>809</sup>5'-TAGTT-3'<sup>813</sup>, <sup>852</sup>5'-ATGTA-3'<sup>856</sup>, <sup>887</sup>5'-TTGTA-3'<sup>891</sup>, <sup>934</sup>5'-TTGTT-3'<sup>938</sup> and <sup>971</sup>5'-CAGAT-3'<sup>975</sup> (Figure 2.25). Binding strength progressively decreased on further extension of the linker to C8, relative to the C6 and C7 linked N-terminal formyl compounds (Figure 2.25). Comparative footprinting of the C6-C8 linked amino and formyl-TIPs established a binding order for the linked TIP compounds, whereby TIP-C6-formyl > TIP-C6-amino (Figure 2.26), TIP-C7-formyl = TIP-C7-amino (Figure 2.27) and TIP-C8-formyl < TIP-C8-amino (Figure 2.28). N-terminal hydrogen TIP, with either a five or seven carbon linker failed to show discrete binding in the region (Figures 2.29 and 2.30), as did C7 linked N-terminal acetyl TIP (Figure 2.31). At high ligand concentrations, however, non-specific binding due to ligand saturation resulted in the loss of cleavage products. A general binding summary for the complete range of heterocycle, linker and N-terminal head group combinations examined is shown in Table 2.1.



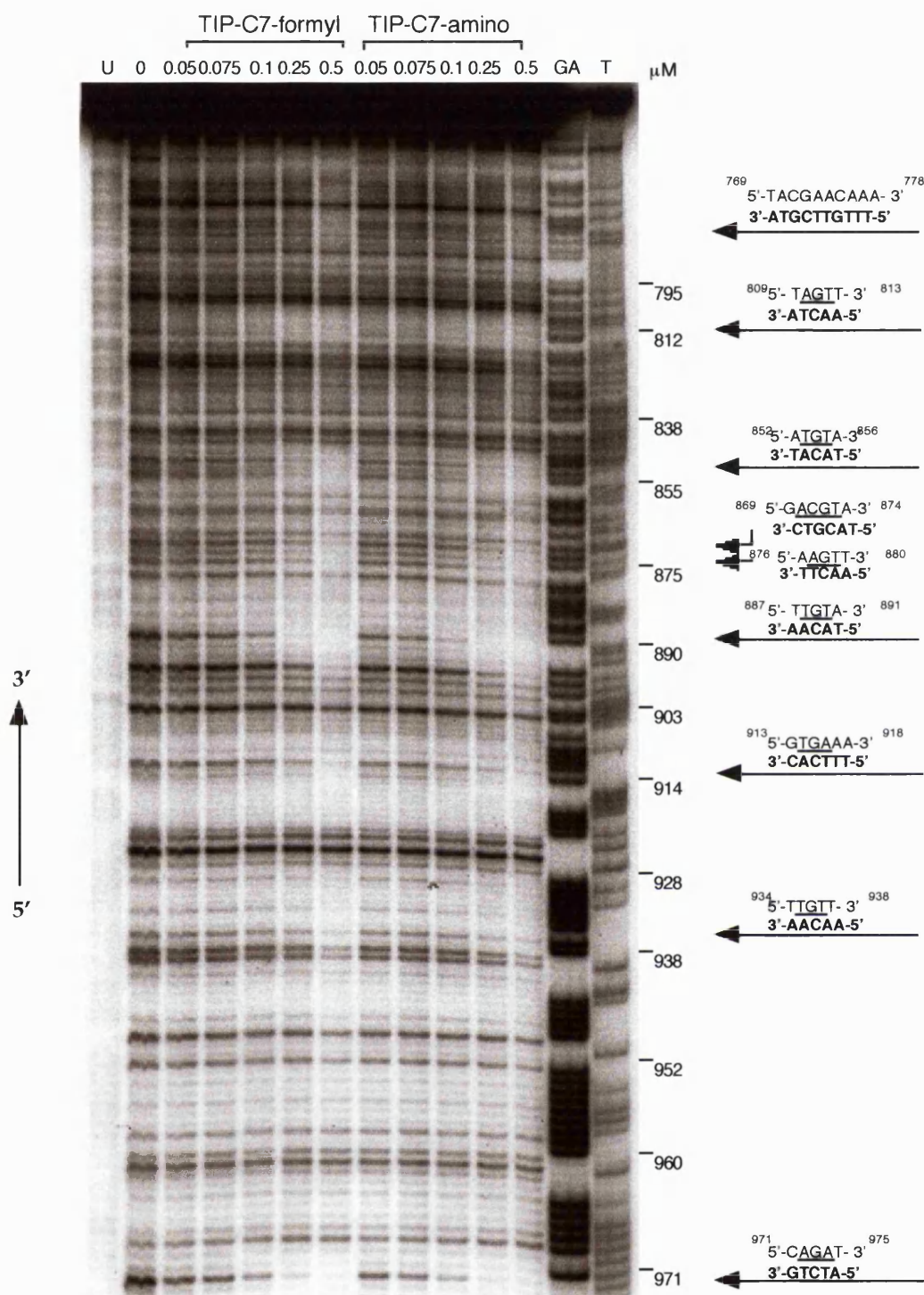


**Figure 2.25** DNase I footprint titration with TIP-C6-formyl, TIP-C7-formyl and TIP-C8-formyl on the bottom strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, with binding sites on this strand indicated in bold type alongside the corresponding top strand sequence. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.



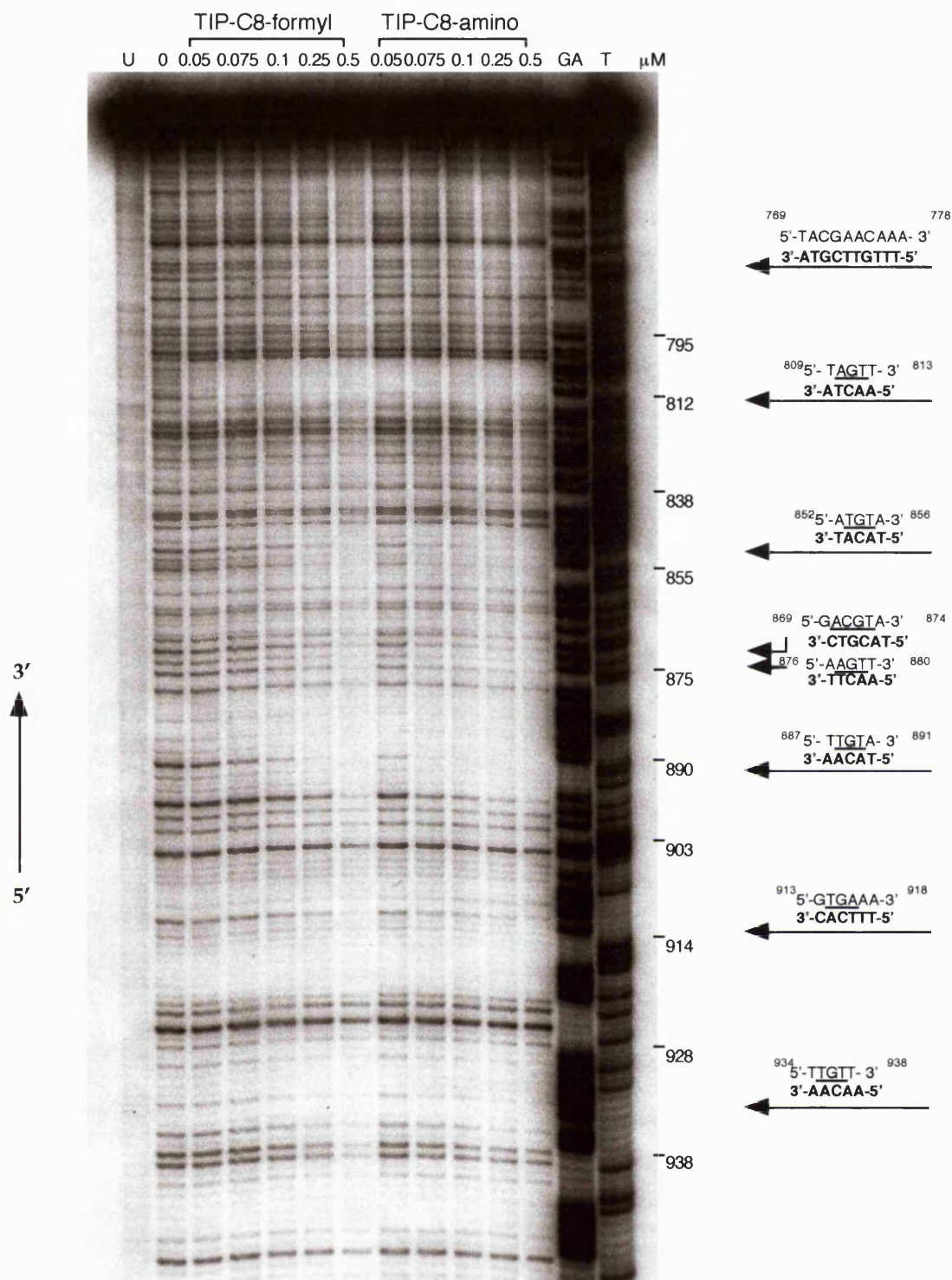


**Figure 2.26** DNase I footprint titration with TIP-C6-formyl and TIP-C6-amino on the bottom strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, with binding sites for both compounds on this strand indicated in bold type alongside the corresponding top strand sequence. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

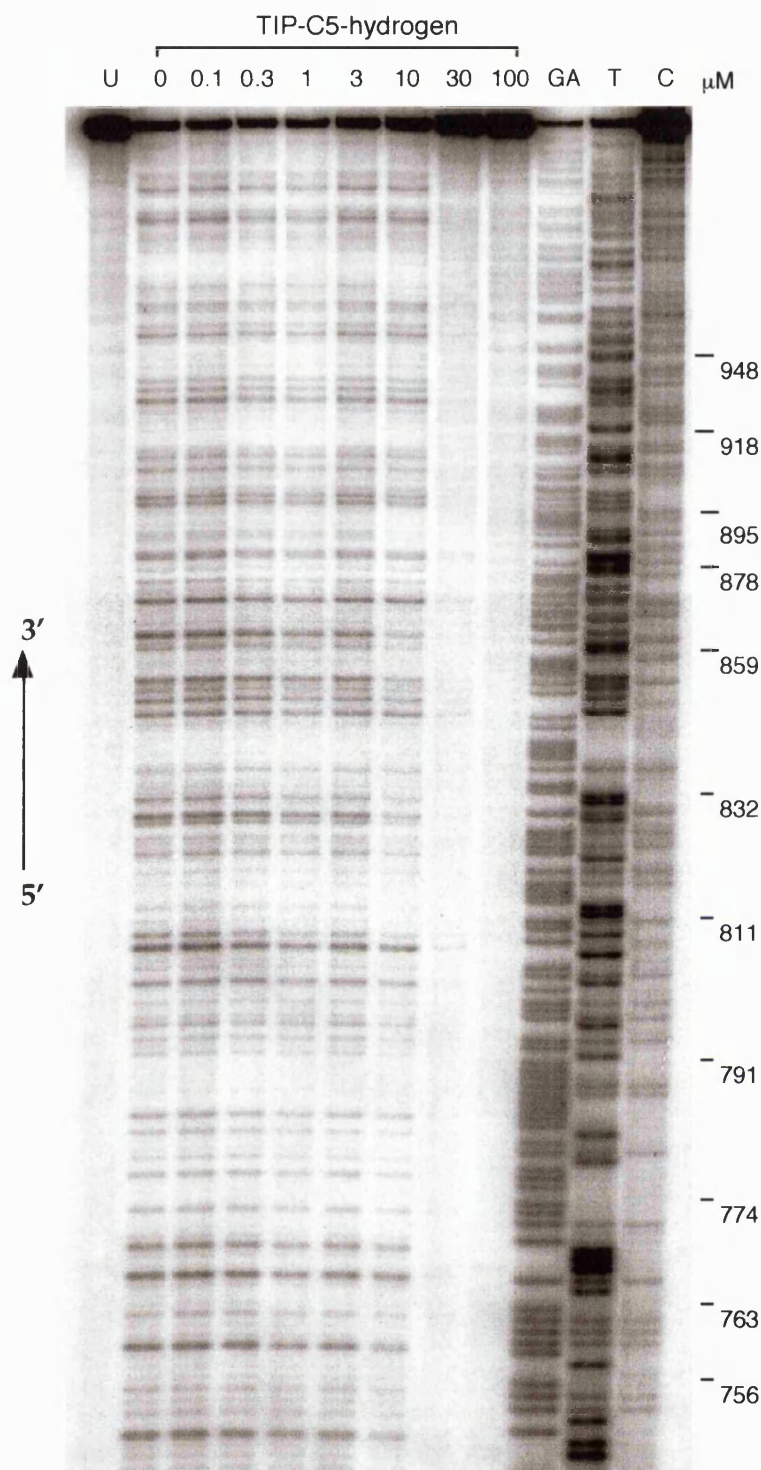


**Figure 2.27** DNase I footprint titration with TIP-C7-formyl and TIP-C7-amino on the bottom strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, with binding sites for both compounds on this strand indicated in bold type alongside the corresponding top strand sequence. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

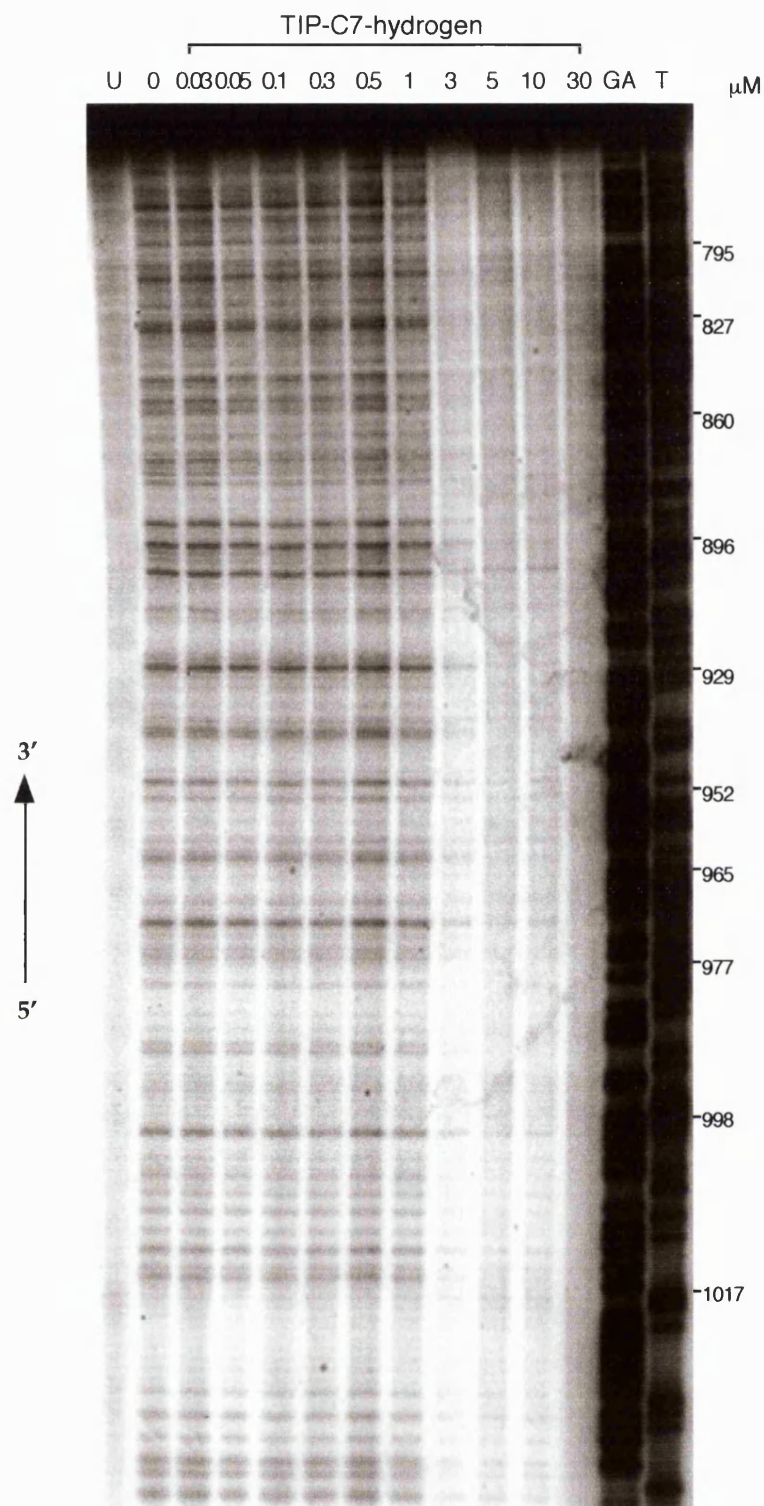




**Figure 2.28** DNase I footprint titration with TIP-C8-formyl and TIP-C8-amino on the bottom strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, with binding sites for both compounds on this strand indicated in bold type alongside the corresponding top strand sequence. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.

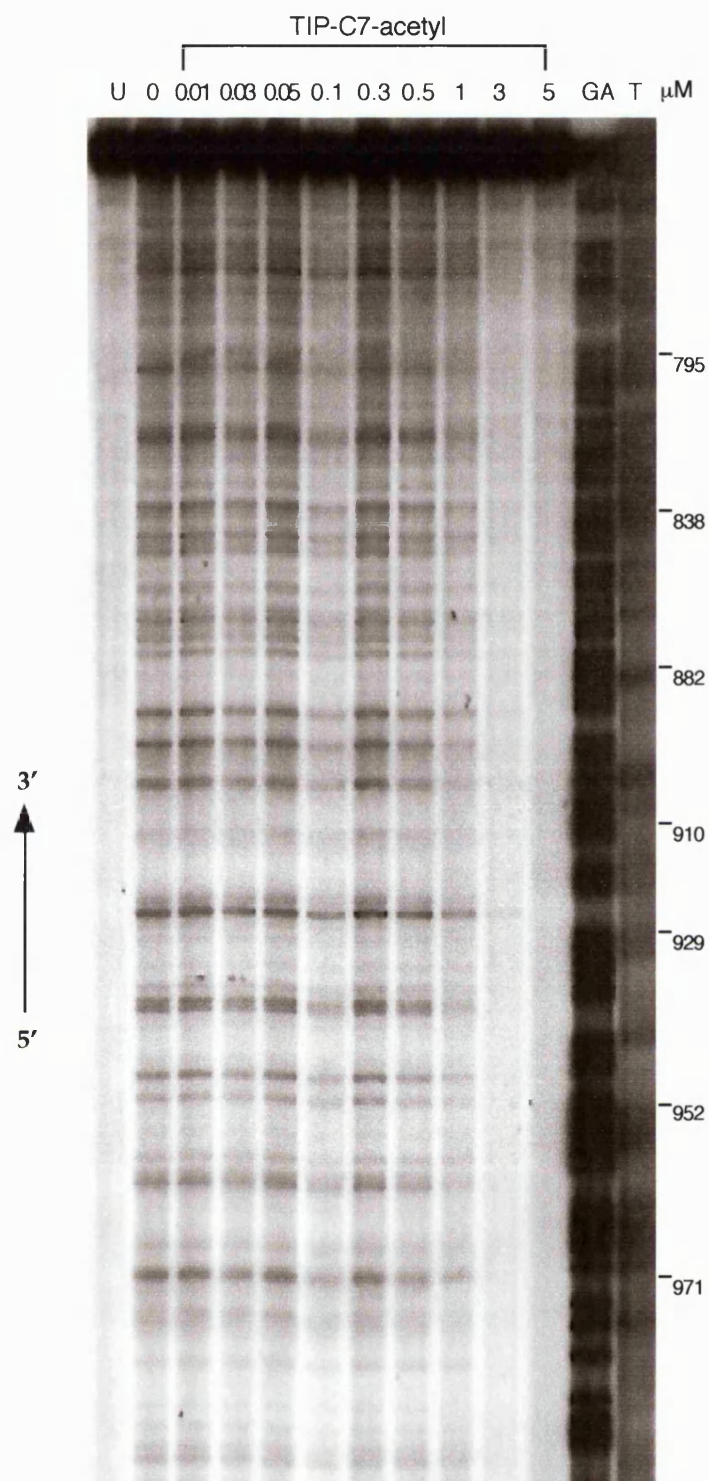


**Figure 2.29** DNase I footprint titration with TIP-C5-hydrogen on the top strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing no binding. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction, T is the thymine marker reaction and C is the cytosine marker reaction.

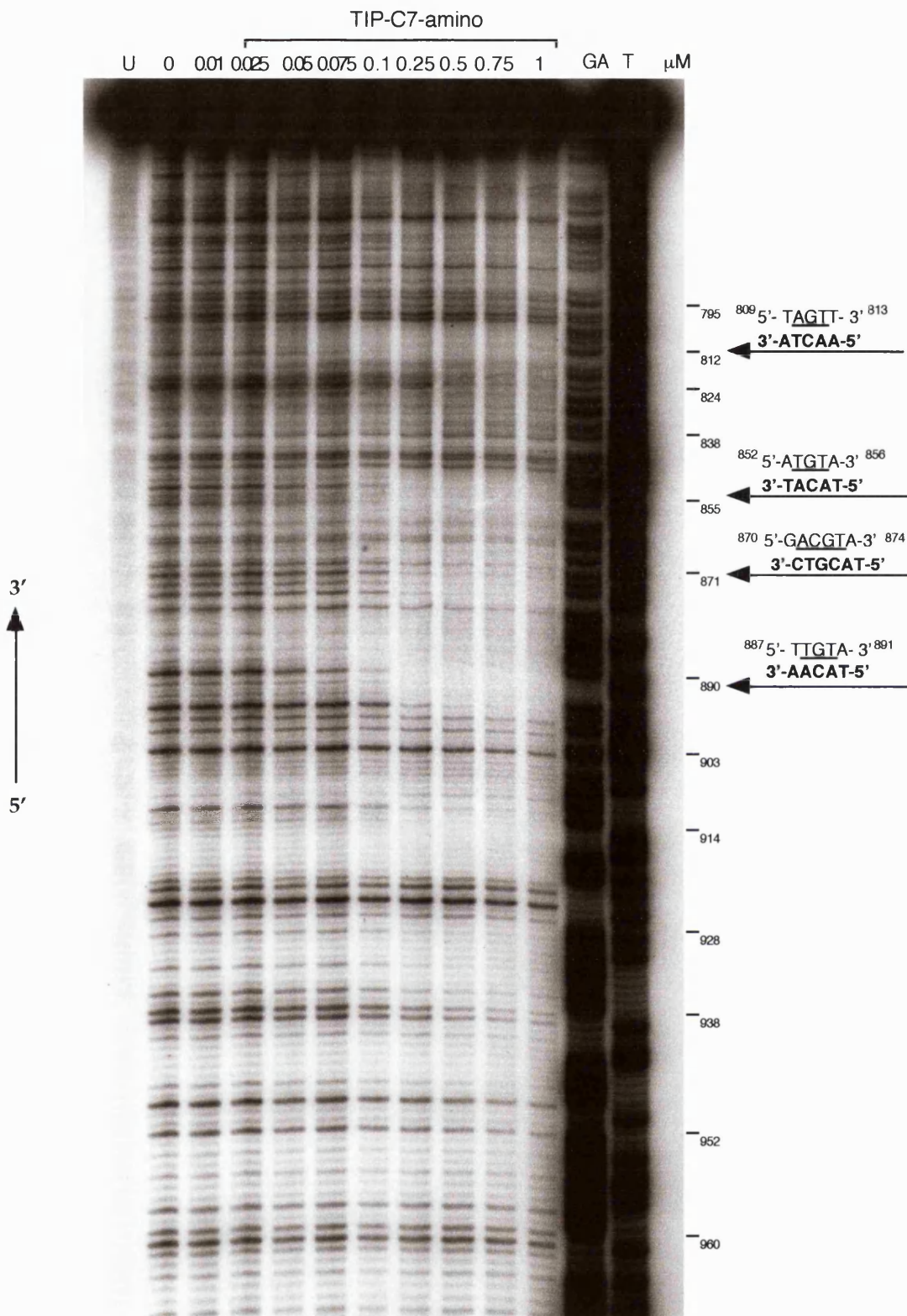


**Figure 2.30** DNase I footprint titration with TIP-C7-hydrogen on the bottom strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing no binding. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.





**Figure 2.31** DNase I footprint titration with TIP-C7-acetyl on the bottom strand of the 5'  $^{32}\text{P}$  labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing no binding. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM  $\text{MgCl}_2$ , 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.



**Figure 2.32** Quantitative DNase I footprint titration with TIP-C7-amino on bottom strand of the 5'-<sup>32</sup>P labelled 355bp FIV34TF10 LTR fragment from plasmid pBSFIV34TF10, showing four of the five binding sites quantitated. All reactions contain 1000 cps of restriction fragment, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. U is the untreated control, 0 is the DNase I treated control, GA is the purine marker reaction and T is the thymine marker reaction.



polyamide	linker	N-terminal head	binding activity
TII	none	NH <sub>2</sub>	+
TII	none	NHCOCH <sub>3</sub>	-
TPI	none	NHCHO	+
TPI	C5	NHCHO	++
TPP	none	NH <sub>2</sub>	++
TPP	none	NHCHO	++
TPP	none	NHCOCH <sub>3</sub>	-
TPP	C7	NH <sub>2</sub>	++++
TPP	C7	NHCOCH <sub>3</sub>	-
TIP	none	NH <sub>2</sub>	++
TIP	none	NHCOCH <sub>3</sub>	-
TIP	C5	H	-
TIP	C5	NH <sub>2</sub>	+
TIP	C5	NHCHO	+
TIP	C6	NH <sub>2</sub>	+++
TIP	C6	NHCHO	++++
TIP	C7	H	-
TIP	C7	NH <sub>2</sub>	++++
TIP	C7	NHCHO	+++
TIP	C7	NHCOCH <sub>3</sub>	-
TIP	C8	NH <sub>2</sub>	++++
TIP	C8	NHCHO	+++

**Table 2.1** Binding activity summary for the polyamides studied, where P is pyrrole, I is imidazole and T is thiazole. Polyamides were either unlinked monomers or linked via a C5, C6, C7 or C8 methylene chain. N-terminal head groups included hydrogen (H), amino (NH<sub>2</sub>), formyl (CHO) and acetyl (COCH<sub>3</sub>). The plus sign, +, signifies relative overall binding activity, with ++++ denoting the strongest binding and the weakest. The minus sign, -, signifies no discrete binding.

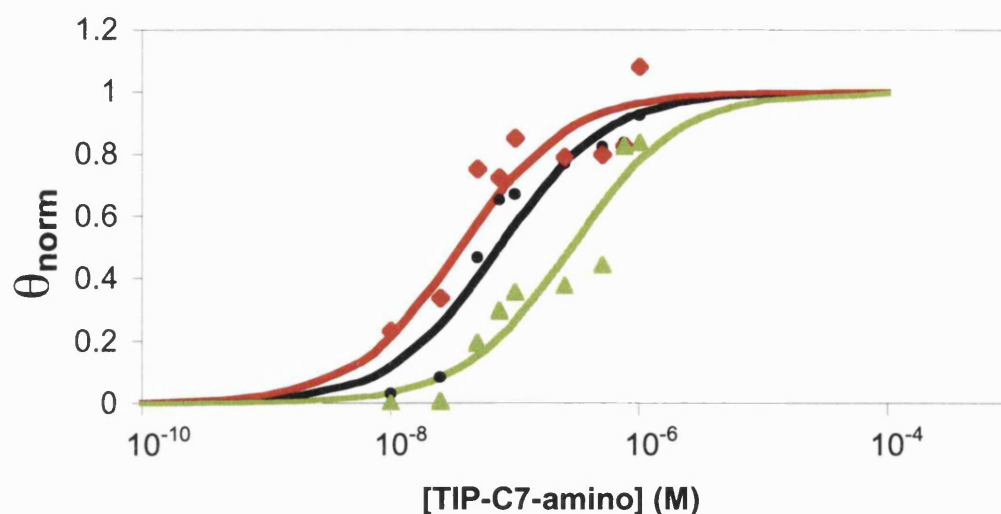
#### 2.4.2 Binding affinity and selectivity of the thiazole-imidazole-pyrrole series.

As shown by comparative footprint analyses in the previous section, the near complete thiazole-imidazole-pyrrole polyamides (TIP) series showed discrete footprints at low doses, exhibiting both affinity and selectivity for potential binding sites. In the 355 base pair fragment generated from 5' LTR region of FIV, a number of putative match and mismatch sites were present in both the maximum overlap and one-residue stagger binding modes for this heterocycle combination (Figure 2.2 (d)). To quantitate the affinity at these sites and thereby assess sequence selectivity for this series, association constants for five predominant binding sites were determined as described in section 2.3.4.

Footprinting experiments used for quantitation were generated over small incremental increases in ligand concentration to provide a large number of data points for curve fitting, with data from at least 3 repeat experiments used in this process. A typical titration used in the quantitative analysis of binding by the polyamide TIP-C7-amino is shown in Figure 2.32 and the binding isotherm derived for this compound is shown in Figure 2.33. The  $K_a$  values calculated at the binding sites  $^{809}5'$ -TAGTT -3'<sup>813</sup>,  $^{887}5'$ -TTGTA -3'<sup>891</sup>,  $^{852}5'$ -ATGTA -3'<sup>856</sup>,  $^{971}5'$ -CAGAT-3'<sup>975</sup>, and  $^{869}5'$ -GACGTA -3'<sup>874</sup> for the TIP polyamide series are given in Table 2.2.

##### *Binding Affinity:*

Within the N-terminal amino series the C7 and C8 linked compounds produced the strongest binding. Reduction in the methylene chain length to C6 produced a fifteen-fold decrease in binding affinity at the maximum overlap match site  $^{809}5'$ -TAGTT -3'<sup>813</sup>, relative to the C7 and C8 linked compounds as shown in Table 2.2. Further reduction to a chain length of 5 methylenes produced weak binding overall; therefore  $K_a$  values were not quantitated for this compound. The unlinked TIP-amino bound with a comparable affinity to the C6 linked compound at sites  $^{809}5'$ -TAGTT -3'<sup>813</sup> and  $^{869}5'$ -GACGTA -3'<sup>874</sup> but produced weaker binding at the remaining sites.



**Figure 2.33** A typical binding isotherm from the DNase I quantitative footprint titration experiments for TIP-C7-amino, where (♦) are the data points for the <sup>809</sup>5'-TAGTT-3'<sup>813</sup> site, (●) for <sup>852</sup>5'-ATGTA-3'<sup>856</sup> and (▲) for <sup>869</sup>5'-GACGTA-3'<sup>874</sup> and (-), (-) and (-) are the best fit binding titration isotherms for those sites obtained by a nonlinear least-squares algorithm.  $\theta_{\text{norm}}$  values were obtained by adding  $\theta_{\text{min}}$  and dividing by  $\theta_{\text{max}} - \theta_{\text{min}}$  and were calculated to fit the range of values between 0 and 1.

Polyamide	$K_a \times 10^7$ ( $M^{-1}$ )				
	<sup>809</sup> 5'-TAGTT-3' <sup>813</sup>	<sup>887</sup> 5'-TTGTA-3' <sup>891</sup>	<sup>852</sup> 5'-ATGTA-3' <sup>856</sup>	<sup>971</sup> 5'-CAGAT-3' <sup>975</sup>	<sup>869</sup> 5'-GACGTA-3' <sup>874</sup>
TIP-amino	0.2 ( $\pm$ 0.04)	0.1 ( $\pm$ 0.1)	0.07 ( $\pm$ 0.07)	0.06 ( $\pm$ 0.04)	0.2 ( $\pm$ 0.2)
TIP-C6-amino	0.2 ( $\pm$ 0.2)	0.5 ( $\pm$ 0.4)	0.3 ( $\pm$ 0.2)	0.7 ( $\pm$ 0.5)	0.3 ( $\pm$ 0.2)
TIP-C7-amino	2.7 ( $\pm$ 1.6)	2.3 ( $\pm$ 1.4)	1.0 ( $\pm$ 0.4)	2.8 ( $\pm$ 1.6)	0.4 ( $\pm$ 0.1)
TIP-C8-amino	3.1 ( $\pm$ 1.8)	3.3 ( $\pm$ 2.1)	1.6 ( $\pm$ 0.7)	2.4 ( $\pm$ 0.7)	1.1 ( $\pm$ 0.7)
TIP-C6-formyl	4.0 ( $\pm$ 1.8)	2.4 ( $\pm$ 1.5)	1.2 ( $\pm$ 0.85)	3.4 ( $\pm$ 1.8)	1.2 ( $\pm$ 0.8)
TIP-C7-formyl	2.4 ( $\pm$ 1.0)	1.2 ( $\pm$ 0.6)	0.7 ( $\pm$ 0.02)	1.2 ( $\pm$ 0.6)	0.4 ( $\pm$ 0.1)
TIP-C8-formyl	1.3 ( $\pm$ 0.4)	1.0 ( $\pm$ 0.4)	0.5 ( $\pm$ 0.2)	1.7 ( $\pm$ 0.6)	0.4 ( $\pm$ 0.1)

**Table 2.2** Equilibrium association constants ( $K_a \times 10^7 M^{-1}$ ) for the TIP polyamides. Values reported are the mean from at least three independent DNase I footprinting experiments, with the standard deviation indicated in parentheses.

In contrast, for the N-terminal formyl series, the C6 linker produced the strongest overall binding at putative sites, demonstrating the greatest binding affinity ( $K_a=4.0 \times 10^7 \text{ M}^{-1}$ ) at the maximum overlap match site  $^{809}5'\text{-TAGTT-}3'^{813}$  for any compound in either series. An increase in the carbon linker to C7 resulted in an approximately two-fold decreased binding affinity at most sites compared to the C6 compound. Increasing the linker length further to an eight carbon chain, produced a progressive decrease in binding affinity at all sites. A comparison of the C6-C8 compounds is shown previously in Figure 2.25. In common with the amino series, the C5 linked formyl compound was a weak binder and as such, was not quantitated. No discrete binding was observed when the TIP compounds contained an N-terminal hydrogen or acetyl head group in either the unlinked or cross-linked form.

#### *Binding Specificity:*

Comparison of the binding affinities at the individual binding sites allowed a determination of the relative binding of the TIP compounds between the maximum overlap (sites  $^{809}5'\text{-TAGTT-}3'^{813}$ ,  $^{887}5'\text{-TTGTA-}3'^{891}$ ,  $^{852}5'\text{-ATGTA-}3'^{856}$ ,  $^{971}5'\text{-CAGAT-}3'^{975}$ ) and the stagger (site  $^{869}5'\text{-GACGTA-}3'^{874}$ ) modes of binding. In addition, comparison of the four maximum overlap sites allowed an assessment of the discrimination of the compounds for A: T versus T: A sequences. Table 2.3 lists the ratios of binding at the stagger mode binding site ( $^{869}5'\text{-GACGTA-}3'^{874}$ ) and maximum overlap mismatch sites ( $^{887}5'\text{-TTGTA-}3'^{891}$ ,  $^{852}5'\text{-ATGTA-}3'^{856}$ ,  $^{971}5'\text{-CAGAT-}3'^{975}$ ) compared to the putative maximum overlap match site ( $^{809}5'\text{-TAGTT-}3'^{813}$ ). It should be noted, however, that only a very approximate assessment of specificity could be made using the binding affinity values calculated due to the large accompanying standard deviations.

Within the amino series only the C7 and C8 linked compounds showed specificity for the maximum overlap over the stagger mode of binding (Table 2.3). This difference was seven-fold in the case of TIP-C7-amino. The C6-C8 formyl compounds also showed some specificity for the maximum overlap mode, which was again maximal for the C7 compound (six-fold). The unlinked TIP-amino compound gave comparable binding at

the maximum overlap match site and the one-residue stagger site. This compound did however show some discrimination between the maximum overlap match site and the three mismatch sites indicating that the thiazole/pyrrole pair could discriminate between A: T and T: A. This specificity was also maintained with the formyl linked compounds and was over three-fold in some cases. The specificity for the match site was less with the linked amino compounds and absent in the case of the TIP-C6-amino.

Polyamide	OVERLAP	OVERLAP	OVERLAP	STAGGER
	AGT/TGT	AGT/TGT	AGT/AGA	AGT/ACGT
	(809-813/ 887-891)	(809-813/ 852-856)	(809-813/ 971-975)	(809-813/ 869-874)
TIP-amino	2.0	2.9	3.3	1.0
TIP-C6-amino	0.4	0.7	0.3	0.7
TIP-C7-amino	1.2	2.7	1.0	6.8
TIP-C8-amino	0.9	1.9	1.3	2.8
TIP-C6-formyl	1.7	3.3	1.2	3.3
TIP-C7-formyl	2.0	3.4	2.0	6.0
TIP-C8-formyl	1.3	2.6	0.8	3.3

**Table 2.3** Effects of linker length and head group on specificity of binding, where specificity is defined as  $K_a(\text{AGT})/K_a(\text{XGX})$ .

## 2.5 Discussion

The binding affinities of a series of polyamide monomers and cross-linked dimers comprised of the heterocycles thiazole, imidazole and pyrrole were determined to evaluate the effects on selectivity and binding of differing heterocycle combinations, the introduction of alternative leading N-terminal head groups and variation in the linking methylene chain length. Further quantitative DNase I footprint titrations with the TIP heterocycle series, of a DNA sequence containing putative match and mismatch sites for both the maximum overlap and one residue stagger binding modes, allowed the measurement of binding constants at each site. The presence of alternative mismatches for a given match site enabled specificity of binding to be determined and consequently the potential base recognition properties of the thiazole heterocycle to be assessed. Concurrent gel shift experiments by Dickerson and coworkers examined the binding of the TIP compound series to short oligomers containing match sites for both modes of binding (O'Hare *et al.*, 2002).

Only limited evaluation of the effects of different heterocycle combinations on sequence selectivity was possible due to the incomplete nature of the compound series. Furthermore, despite the presence of a leading terminal head group which should favour binding in the staggered motif (Kopka *et al.*, 1997, Lacy *et al.*, 2002a), many of the compounds examined here favoured binding in the maximum overlap motif. This may have been due in some cases to the limited number of putative match and mismatch sequences present in the DNA sequence examined for such a large number of compounds. N-terminal amino or formyl thiazole-pyrrole-pyrrole (TPP) showed binding at numerous AT sites, with the strongest binding in this series observed with the cross-linked TPP-C7-amino. The selectivity of the individual constituent thiazole and pyrrole heterocycles could not be accurately assigned as the sequence examined was AT rich and the compounds bound at multiple overlapping sites. Thiazole-pyrrole-imidazole (TPI) showed weak binding as an N-formyl monomer, with a ten-fold increase in binding observed with the C5 linked dimer. These shared site specificity with the TPP binders, targeting multiple AT runs, with TPI-formyl exhibiting a decreased affinity for these sites. N-terminal amino thiazole-imidazole-imidazole (TII) strongly favoured sites

centred on GC runs, showing predominant binding at two one-residue stagger mismatch sites. Although the LTR region contained few putative sites in either binding mode for this heterocycle combination, little unspecific binding was observed. Thus the presence of two contiguously placed imidazoles strongly dominates selectivity. N-terminal amino and formyl thiazole-imidazole-pyrroles (TIP) also demonstrated discrete binding at selected maximum overlap and one-residue stagger sites for this heterocycle combination.

Despite the limited range of compounds within each heterocycle combination studied, some general trends regarding structure activity were observed (Table 1). Certain linkage and head group combinations proved to be particularly weak binders. As seen previously by Chen and coworkers (Chen & Lown 1994; Chen *et al.*, 1996), linkage of polyamide monomers capable of DNA binding showed greatly improved affinity when cross-linked with a methylene linker of suitable length. This was demonstrated in both the TPP and TIP heterocycle series with the linkage of TPP-amino and TIP-amino with a seven carbon methylene chain. However, weak binding was observed with a five carbon linker for either the N-terminal formyl or amino TIPs, offering no increase in affinity compared to the monomers. Interestingly, this was not the case for TPI-C5-formyl, which showed a ten-fold increase in binding relative to its corresponding monomer.

The influence of the N-terminal head group attached to the leading aromatic ring was also evident across the heterocycle combinations. In all cases, no DNA binding was observed for compounds with an acetyl or hydrogen head group, irrespective of linkage. In contrast, those with either an N-terminal amino or formyl group proved to be moderate to good binders, depending on the suitability of the associated linker. These observations suggest that the presence of an N-terminal hydrogen or acetyl may abrogate binding in cross-linked polyamides, as observed previously with C7 linked compounds by Burckhardt and coworkers (Burckhardt *et al.*, 2000). This is not the case however, for hairpin polyamides, which bind efficiently with an N-terminal hydrogen (Parks *et al.*, 1996a, b).



A more detailed study of the near complete TIP series using quantitative DNase I footprinting demonstrated that within the N-terminal amino-TIP series, the C7 and C8-linked compounds bound most strongly to the putative binding sites. No significant binding was observed at sites other than these. The TIP-C7-amino proved the most selective, with up to a seven-fold selectivity for the maximum overlap match site, AGT. A significant decrease in both affinity and selectivity was seen in this series with reduced methylene chain lengths of five or six carbons. Alteration of the leading head group to a formyl resulted in an optimal methylene linker of six carbons. TIP-C6-formyl produced the strongest binding with a three-fold specificity for the overlap match site over the one-residue stagger site. Reduced binding was seen with the TIP-C7-formyl, which nevertheless showed up to a six-fold specificity for the overlap match site. The C8-linked formyl demonstrated a decrease in both binding affinity and specificity. Gel shift experiments by Dickerson and coworkers (O'Hare *et al.*, 2002) strongly supported the footprinting data, giving the same order of binding affinity in both the amino and formyl series. Overall, however, the formyl series demonstrated a lower affinity for the DNA in both binding modes, requiring an elevated drug to DNA ratio. This is consistent with the CD experiments of Burckhardt and coworkers (Burckhardt *et al.*, 2000) comparing C7-linked compounds with an amino or formyl head group.

Whilst polyamides linked using a hairpin motif have been extensively investigated (Dervan *et al.*, 1999), the optimised structure for cross-linked polyamides remains to be defined. Several studies of cross-linking in polyamides have demonstrated significant improvements in binding on linkage of the individual polypeptide ligands (Chen & Lown 1994; Chen *et al.*, 1996; Burckhardt *et al.*, 2000). The detailed assessment of linkage with a range of possible linkers presented here for the TIP heterocycle series, indicates a C6 linker with an N-terminal formyl and a C7 linker with an N-terminal amino provides suitable alignment of the polyamide dimer within the minor groove. This is lost when the linker is reduced to five carbons, possibly because comprehensive contact of the dimeric ligand with both strands of the DNA is restricted. In this case, 2:1

binding may require two C5 linked molecules, in a similar fashion to 2:1 binding by polyamide monomers.

This study reveals that the nature of the N-terminal moiety has a profound influence on the binding characteristics of cross-linked polyamides. Both the footprinting data presented here and gel shift experiments carried out at UCLA by Dickerson and coworkers indicate that the use of a hydrogen or acetyl head group may inhibit binding. Furthermore, replacement of an amino head group with a formyl alters the optimal linker required for binding. Therefore, there may be a combined effect on polyamide binding by the linkage and head group. Polyamide specificity may also be modulated in this way, resulting in an overall improved specificity for match site binding with an N-terminal formyl.

It is significant that both of the successful head groups, amino and formyl, have a NH group capable of hydrogen bonding with DNA, whereas the unsuccessful head groups, hydrogen and acetyl, do not. In a side-by-side unlinked di-imidazole complex with leading formyl groups (Kopka *et al.*, 1998), both of the formyl NH make hydrogen bonds with the floor of the minor groove: 2.94 Å from cytosine O2 or 2.80 Å from guanine N3 (table 1 of Kopka *et al.*, 1998). Hydrogen bonded interactions of the head group with DNA appear to be required for cross-linked side-by-side binding, although such an interaction is not demanded for a hairpin complex (Dervan *et al.*, 1999). It is interesting to note that amino and formyl are N-terminally placed in naturally occurring minor groove binding antibiotics such as anthelvencin, kikumycin B and distamycin, whereas hydrogen and acetyl are not (Sharma *et al.*, 2000; Probst *et al.*, 1965; Takaishi *et al.*, 1972).

This work also highlights the importance of the structural context of component heterocycles in DNA sequence recognition by polyamides. The thiazole ring, tested in the terminal position of the TIP polyamide series as a reading element for adenine, bound its AGT target in the maximum overlap motif with both the amino and formyl head groups. The overlap motif pairs the thiazole ring with pyrrole, and thus thiazole

reads adenine and pyrrole thymine as expected. The footprinting results with TIP-C6-formyl showed a preference for AGT over TGT by up to three-fold. Furthermore, the footprinting indicated that AGT is strongly favoured over ACGT, the one-residue stagger motif, in which thiazole is next to cytosine and imidazole next to guanine. Therefore, thiazole appears to favour adenine moderately in a Th/Py pair with a poor CG specificity when paired with imidazole.

These results suggest that the thiazole moiety provides very moderate binding and discrimination of adenine when in the polyamide N-terminal position. A previous study concluded that a thiazole/pyrrole pair binds poorly to all four Watson-Crick base pairs (Nguyen *et al.*, 2001). However, in that study the thiazole was placed internally in a hairpin. Positioning of the heterocyclic moiety within a polyamide may be crucial. Indeed, the successful T/A discrimination seen with hydroxypyrrole sited internally within the hairpin motif has not been found when hydroxypyrrole is placed at the terminal position (Ellervik *et al.*, 2000). Binding affinities of hairpins containing hydroxypyrrole vary considerably based on the position of the hydroxypyrrole ring within the hairpin motif (White *et al.*, 1998). Calculation of relative binding affinities of imidazole and hydroxypyrrole rings by Goodsell and co-workers (Walker *et al.*, 1997) indicate that the hydroxypyrrole ring binds with a lower affinity for AT base pairs compared to the affinity of the imidazole ring for GC pairs. In the present study the specificity for adenine observed with thiazole in the terminal position is comparable with the specificity for thymine provided by a hydroxybenzamide/pyrrole pair in eight ring hairpin polyamides (Ellervik *et al.*, 2000). The specificity of thiazole placed internally with a cross-linked polyamide remains to be determined.

A number of alternative heterocycle pairs examined recently to optimise the discrimination of A from T include hydroxybenzamide/pyrrole (Hb/Py) (Ellervik *et al.*, 2000), pyrazole/pyrrole (Pz/Py) (Nguyen *et al.*, 2001) and hydroxybenzimidazole/pyrrole pairs (Hz/Py) (Renneberg & Dervan 2003). Some of these heterocycles are six-membered aromatic rings, which it has been suggested may enhance contact with the minor groove due to their reduced curvature relative to the

commonly used five-membered rings (Renneberg & Dervan 2003; Briehn *et al.*, 2003). However, at present most of these heterocycles have been investigated in a central position only. Nevertheless, an N-terminal 3-methoxy or 3-chloro thiophene residue paired opposite a pyrrole has shown moderate selectivity for AT over TA when placed at the N-terminus of a hairpin motif (Foister *et al.*, 2003). It is interesting to note that this heterocycle is structurally very similar to thiazole, possessing a bulky sulphur atom which disfavours contact with the floor of the minor groove.

The influence of the relative positioning of the heterocycle within the polyamide is also demonstrated here. This is most clearly shown with the TPP and TPI compounds which share site preference in the LTR region examined. Despite the presence of putative match sites in both binding modes for the TPI combination (Figure 2.2), binding nevertheless predominates at AT runs. Whilst binding at these sites is reduced relative to TPP, this strongly suggests that the adjacent thiazole and pyrrole moieties determine the binding preference. Although, imidazole generally exhibits good GC specificity, this is overcome when placed at the C-terminus in this context. The AT richness of the examined sequence may also possibly contribute to this effect. Thus the lateral context of the heterocycle within the molecule must be taken into consideration in addition to its parallel position in 2:1 binding.

Recent work by Burckhardt and coworkers (Burckhardt *et al.*, 2002) investigating the binding of N-terminal amino TPP and TIP has suggested that sequence recognition may be favoured by the presence of T-A steps adjacent to the target sequence. In this concurrent study, CD and footprinting experiments confirmed that TPP-amino bound at pure AT tracts, whilst TIP-amino bound to both maximum overlap and one-residue stagger sites, with binding enhanced if the sites were flanked by T-A steps. It has been proposed that the presence of T-A steps may result in a compression of the major groove and thus an associated widening of the minor groove (Goodsell *et al.*, 1994; Dickerson 1998; Mack *et al.*, 2001). This widening may subsequently facilitate binding of a dimeric peptide within the minor groove, favouring 2:1 binding at that site. Thus the base composition of the flanking sequence providing the structural context of the target

site, as well as the structure of the polyamide itself should also be considered when interpreting the efficacy of polyamide binding.

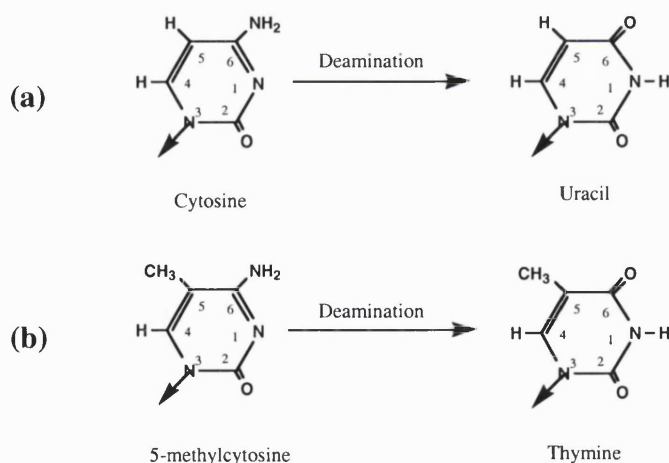
Differences in the two types of polyamides, hairpin versus cross-linked, complicate direct comparison of the thiazole ring in the two linkage motifs. In the hairpin motif there is effective binding in the absence of a head group. In contrast, in the cross-linked motif the head group attached to the leading ring may significantly modulate its binding characteristics, with formyl enhancing adenine selectivity and acetyl or hydrogen inhibiting binding for cross-linked polyamides with a leading thiazole ring. Additionally the number of constituent rings in the dimer, six in the case of the compounds used in this study and eight in the thiazolated hairpin study (Nguyen *et al.*, 2001), may significantly affect the affinity of binding. N-formyl-thiazole may provide sequence discrimination in cross-linked polyamides, with thiazole favouring adenine over thymine. Moreover, the leading head group in combination with the length of methylene linker have an important influence on the binding characteristics of cross-linked polyamides.

## CHAPTER 3

### RECOGNITION OF TG MISMATCHES IN THE DNA MINOR GROOVE USING TRI-IMIDAZOLE POLYAMIDES.

#### 3.1 Introduction

Although discrimination by polyamides of all four Watson-Crick base pairs has been achieved *in vitro* in the minor groove, their use in the selective recognition of non-canonical DNA sequence remains to be determined. Mutations resulting in mismatched base pairs have been implicated in tumorigenesis, with TG mismatches, arising from the spontaneous deamination of 5-methylcytosine or defective replication, responsible for most common mutations in human ras oncogenes (Land *et al.*, 1983; Almoguera *et al.*, 1988; Lall *et al.*, 1998; Watanabe *et al.*, 1999). Whilst methylation of cytosine provides a crucial mechanism for the discrimination of active genes in vertebrates, 5-methylcytosines are hotspots for spontaneous mutation as shown in Figure 3.1. This is due to the inherent instability of the cytosine residue and such events account for a third of point mutations observed in heritable human disease (Lall *et al.* 1998). Thus, despite the presence of specific cellular TG mismatch repair mechanisms, some TG mismatches may persist, eluding repair (Toft & Arends 1998).



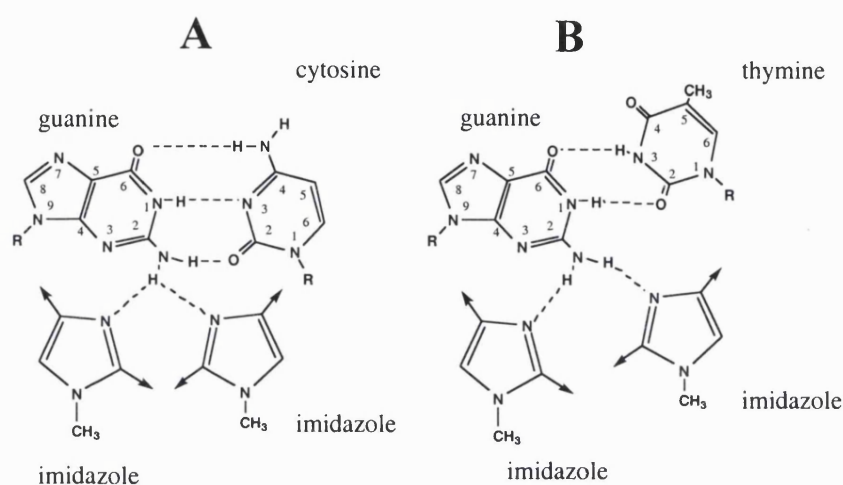
**Figure 3.1** Molecular basis of 5-methylcytosine hot spots showing the different deamination products of cytosine and 5-methylcytosine. **(a)** Deamination of cytosine to uracil: removed by uracil-DNA glycosylase and replaced with cytosine, preventing mutation. **(b)** Deamination of 5-methylcytosine to thymine is not removed, so that it pairs with adenine at the next round of replication giving a CG to TA transition.

Molecular ligands able to bind specifically to DNA sequences incorporating mismatched base pairs would have therapeutic benefit, promoting mismatch recognition and repair in cells, possibly preventing malignant transformation. Structural and thermodynamic analysis of mismatches, including the TG mismatch, has indicated they adopt a wobble conformation producing minor local structural distortion, with the stability of the mismatch dependent on the flanking sequence (Hare *et al.*, 1986; Hunter *et al.*, 1987; Allawi *et al.*, 1997). This may provide a basis for mismatch recognition and the development of small molecules able to discriminate T: T, A: G and G: G mismatches is already in progress (Trotta & Paci; Jackson *et al.*, 1999; Jackson & Barton 2000; Nakatani *et al.*, 2001a & b).

Recent NMR, CD and surface plasmon resonance (SPR) studies have demonstrated that a tri-imidazole polyamide analogue of distamycin, formyl-imidazole-imidazole-imidazole (formyl-III or AR-1-144), binds co-operatively in a 2:1 mode not only at GC base pairs but also at TG mismatches in short oligomer stretches (Yang *et al.*, 1999b; Lacy *et al.*, 2002b). This work suggests that the imidazole/imidazole (Im/Im) ring pair, previously shown to exhibit weak and degenerate recognition at GC base pairs (White *et al.*, 1997; Yang *et al.*, 1999a), may provide a suitable TG recognition motif further extending the current set of polyamide heterocycle pairing rules. When guanine is paired with thymine its N2 amino group is free to form two individual hydrogen bonds with a side-by-side Im/Im pair), as shown in Figure 3.2, in a similar fashion to TG mismatch recognition by Vsr mismatch endonuclease (Fox *et al.*, 2000. This was confirmed by thermodynamic and kinetic studies showing formyl-III bound cooperatively to a single TG base pair in a short DNA oligomer, as a homodimer and as a heterodimer alongside imidazole-pyrrole-imidazole, with selectivity due to the slow dissociation rate on binding of the 2:1 complex (Lacy *et al.*, 2002b). However, whilst formyl-III has shown preferential binding at a single TG mismatch in short DNA segments, its ability to discriminate TG mismatches in a longer sequence context remains to be demonstrated. More recently, polyamide-fluorophore conjugates, containing an Im/Im ring pair, have been used to distinguish T: G mismatches also in



short DNA segments providing further evidence of the potential utility of this heterocycle pairing in TG mismatch recognition (Rucker et al., 2003).

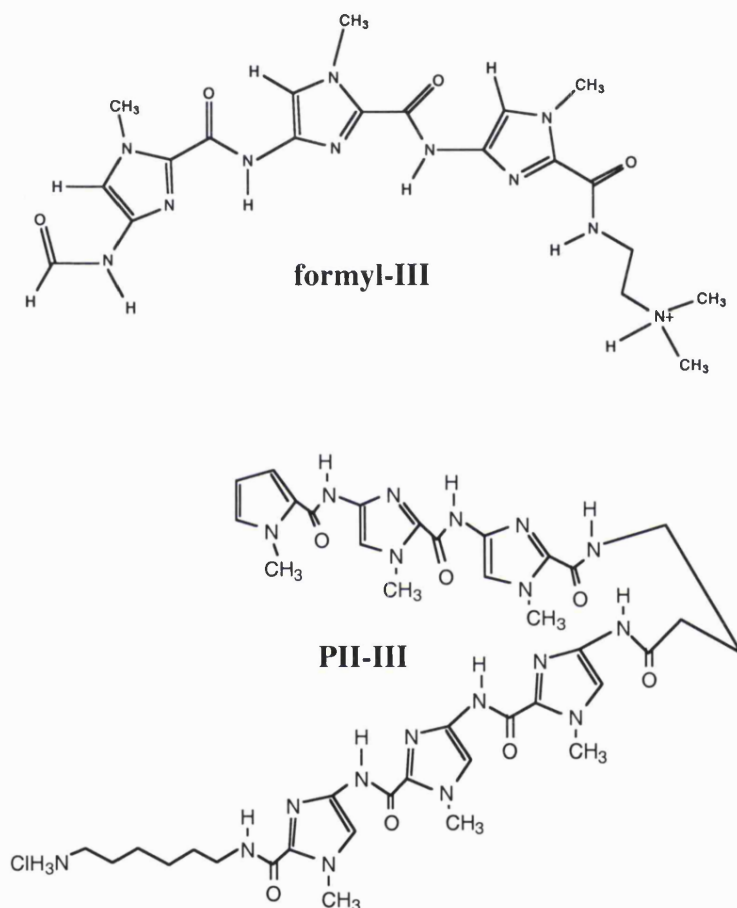


**Figure 3.2** Molecular basis for the recognition of GC and GT base pairs with the imidazole/imidazole ring pair. **A.** an Im/Im pair recognises a G:C base pair and **B.** an Im/Im pair recognises a G:T base pair, as adapted from Yang *et al.*, 1999 a & b.

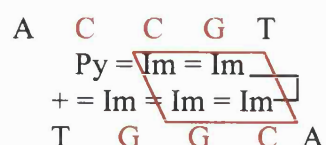
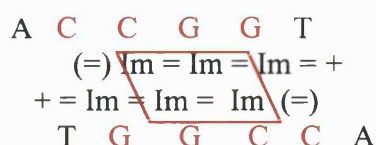
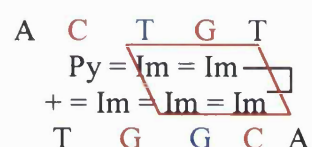
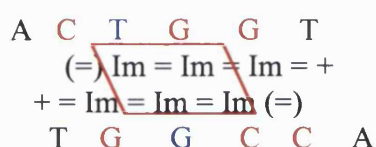
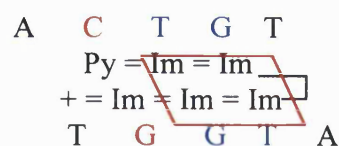
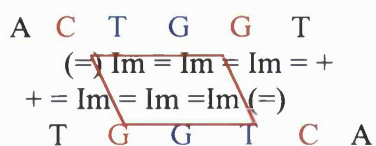
Further to TG mismatch recognition, an ability to differentiate between single and double mismatches is of biological importance as a tandem double T: G mismatch is rarely found *in vivo*. For this purpose, the hairpin-linked polyamide heterodimer PII-III was designed by Lee and coworkers to recognise two consecutive TG mismatches. Binding at a double tandem TG mismatch site by PII-III, containing two consecutive Im/Im ring pairs in a maximum overlap motif, should be favoured over a single mismatch due to the increased affinity of Im/Im for TG over CG base pairs observed previously (Yang *et al.*, 1999b). Moreover, as the binding affinity of Im/Im ring pairs for CG is also weaker than that observed with an PyIm pair, the introduction of an N-terminal pyrrole unit may dock the polyamide in a maximum overlap motif alongside an initial CG base pair, aligning the two Im/Im pairs next to the consecutive TG mismatches.

This chapter presents the *in vitro* binding of the polyamide monomer formyl-III (AR-1-144) and the heterodimer PII-III (EM-1-101c) as shown in Figure 3.3, to an 80 and a 78

base pair DNA sequence respectively containing TG mispairs. Their relative selectivity for three putative binding sites (Figure 3.4), a Watson-Crick match site, a single TG mismatch and a consecutive double TG mismatch placed in series along each oligonucleotide was investigated using DNase I footprinting (Lane *et al.*, 1983) in an evaluation of their use as agents for TG mismatch recognition.



**Figure 3.3** Molecular structures of monomer formyl-imidazole-imidazole-imidazole (formyl-III) and hairpin-linked heterodimer pyrrole-imidazole-imidazole –tri-imidazole (PII-III).

**A****B****(a)****(b)****(c)**

**Figure 3.4** Several possible binding motifs for **A.** formyl-III and **B.** PII-III at **(a)** the CG match, **(b)** the single TG mismatch and **(c)** the consecutive TG double mismatch sites along the annealed oligodeoxynucleotides IMG1/2 for formyl-III and IMG1/5/IMG1/6 for PII-III, where Im is imidazole and Py is pyrrole. When binding with DNA in a 2:1 ratio, formyl-III should adopt a one-residue stagger motif due to the presence of a leading N-terminal formamide head group. In contrast, the hairpin-linked heterodimer PII-III should adopt a maximum overlap motif, due to the absence of a leading head group. The equal sign, =, signifies an amide, with (=) denoting the leading formamide head group (NHCHO). The plus sign, +, signifies the charged tail ((CH<sub>2</sub>)<sub>2</sub>NMe<sub>2</sub>H<sup>+</sup>) and the square bracket, ], denotes the hairpin linker. The DNA target sequence is indicated in red capital letters, with TG mispairs indicated in blue.

## 3.2 Materials

### 3.2.1 Investigational compounds

Compounds were kindly provided by Professor Moses Lee, Department of Chemistry, Furman University, Greenville, North Carolina, USA. The synthesis and characterisation of the compounds have been reported (Lee *et al.*, 1993). A minimum of 1mg of each compound was dissolved in DMSO at a 10mM stock concentration and stored at –20°C until use. Compound dilutions were freshly prepared for each experiment in distilled and deionised water.

### 3.2.2 Oligodeoxynucleotide DNA

A double-stranded oligodeoxynucleotide was designed for each compound to contain three putative binding sites, a CG match, a single TG mismatch and a consecutive TG double mismatch. These were designed to occur in the favoured binding mode of each compound, a one-residue stagger motif for formyl-III and a maximum overlap motif for PII-III.

The 80mer oligodeoxynucleotide sequences 5'-AGG TGA GCA GGT CCA TAC TGG TTT GCA CCT CGA GGT TAC CGG TAT CTG CTC CAG CTC AAC TGG TAA CCT GCA CCT GGT CG-3' (IMGT1) and 5'-CGA CCA GGT GCA GGT TAC TGG TTG AGC TGG AGC AGA TAC CGG TAA CCT CGA GGT GCA AAC CGG TAT GGA CCT GCT CAC CT-3' (IMGT2) were purchased from MWG-Biotech and stored at a concentration of 1µg/µl in 10mM Tris-HCl (pH 7.8), 1mM EDTA at –20°C.

The 78mer oligodeoxynucleotide sequences 5'-AGG TGA GCA GGT CCA TAC TGT TTT GCA CCT CGA GGT TAC CGT ATC TGC TCC AGC TCA ACT GTA ACC TGC ACC TGG TCG-3' (IMGTO5) and 5'-CGA CCA GGT GCA GGT TAT GGT TGA GCT GGA GCA GAT ACG GTA ACC TCG AGG TGC AAA ACG GTA TGG ACC TGC TCA CCT -3' (IMGTO6) were purchased from Oswel and stored at a concentration of 1µg/µl in 10mM Tris-HCl (pH 7.8), 1mM EDTA at –20°C.

The strands for each duplex were designed to be non-self-complementary and unable to form hairpin secondary structures spontaneously. TG mismatches were positioned at either end of the oligodeoxynucleotides, which contained a central putative Watson-Crick CG match site. This ensured annealing of the end-labelled heteroduplex prior to drug treatment, despite a possible decrease in its stability due to the presence of mismatched regions. No alternative favourable binding sites were present in the duplexes.

### **3.2.3 Chemicals, Kits, Radioisotope, Enzymes and Buffers**

A Zymoclean Gel/DNA recovery kit was purchased from Zymo Research and a SequaGel sequencing system, comprised of concentrate (19:1 acrylamide/bisacrylamide), diluent (urea) and buffer (0.89M Tris-Borate, 20mM EDTA pH8.3, urea) was purchased from National Diagnostics. A  $\phi$ X174/Hinf I size marker was purchased from Promega.

The remaining chemicals, radioisotope, enzymes and buffers used were as described in materials and methods section 2.2.

## **3.3 Methods**

### **3.3.1 Construction of a TG mismatched oligonucleotide duplex**

5 $\mu$ g of the DNA oligodeoxynucleotides, IMGT1 and IMGTO5, were 5'-end labelled with  $^{32}$ P- $\gamma$  dATP using T4 polynucleotide kinase as described previously in section 2.2.3.1. Following radio-labelling, the oligodeoxynucleotides were annealed in a 1:1 ratio to 5 $\mu$ g of complementary oligodeoxynucleotide strand, IMGT2 in the case of labelled IMGT1 and IMGTO6 in the case of labelled IMGTO5, by heating at 95°C for 3 min, 65°C for 1 min and cooling slowly to 4°C. The labelled and annealed duplexes were separated from residual unannealed oligodeoxynucleotide strands by electrophoresing the annealing reaction alongside a  $\phi$ X174/Hinf I size marker on a 3% agarose gel containing 0.4 $\mu$ g/ml ethidium bromide. This was carried out for 20 min at high voltage (150V) to prevent diffusion into the surrounding 1X TAE running buffer. Following electrophoresis, the annealed product was excised and purified with a Zymoclean gel/DNA recovery kit following the manufacturer's protocol. In order to

examine ligand binding to the corresponding oligodeoxynucleotide strands, IMGT2 and IMGTO6 were end-labelled and annealed to unlabelled IMGT1 and IMGTO5, respectively, as described above.

### 3.3.2 DNase I footprinting of TG mismatched oligonucleotides

Formyl-III (prepared as in Lee *et al.*, 1993) or PII-III were incubated with 0.5µg of the double-stranded 5'-single end-labelled mismatched oligodeoxynucleotide in 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes, at room temperature for 30 min, in a total volume of 50µl. DNA cleavage was initiated by the addition of the drug treated sample to 2µl (1 U) DNase I diluted in ice cold 10mM Tris pH 7.0 from a stock solution (1U/µl) and 1µl of a solution of 250mM MgCl<sub>2</sub>, 250mM CaCl<sub>2</sub>. The reactions were performed at room temperature and stopped after 3 min by the addition of 100µl of a stop mix containing 200mM NaCl, 30mM EDTA pH 8 and 1% SDS. The cleavage products were phenol/chloroform extracted and ethanol precipitated in the presence of 1µl glycogen (20 mg/ml), washed once in 80% ethanol and lyophilised dry. A purine-specific (GA) marker lane was generated with 0.5µg of the appropriate 5' singly end-labelled and annealed oligodeoxynucleotide DNA duplex by incubation with 50µl of 98% (v/v) formic acid at room temperature for 8 minutes. Following incubation, the sequencing reaction was snap frozen by placing in a dry ice/ethanol bath and dried by lyophilisation. Piperidine cleavage was carried out as described previously in methods section 2.2.4.2. Reaction samples were then resuspended in 1X formamide loading dye, denatured for 5 min at 90°C, cooled on ice and electrophoresed at 2000V for 2 hours on a 10% denaturing polyacrylamide gel, prepared as described in section 2.2.4.3 using 40ml of SequaGel concentrate, 50ml of SequaGel diluent and 10ml of SequaGel buffer. Gels were dried under vacuum at 80°C for 2 hours and exposed to film for 24 hours (X-OMAT, Kodak).

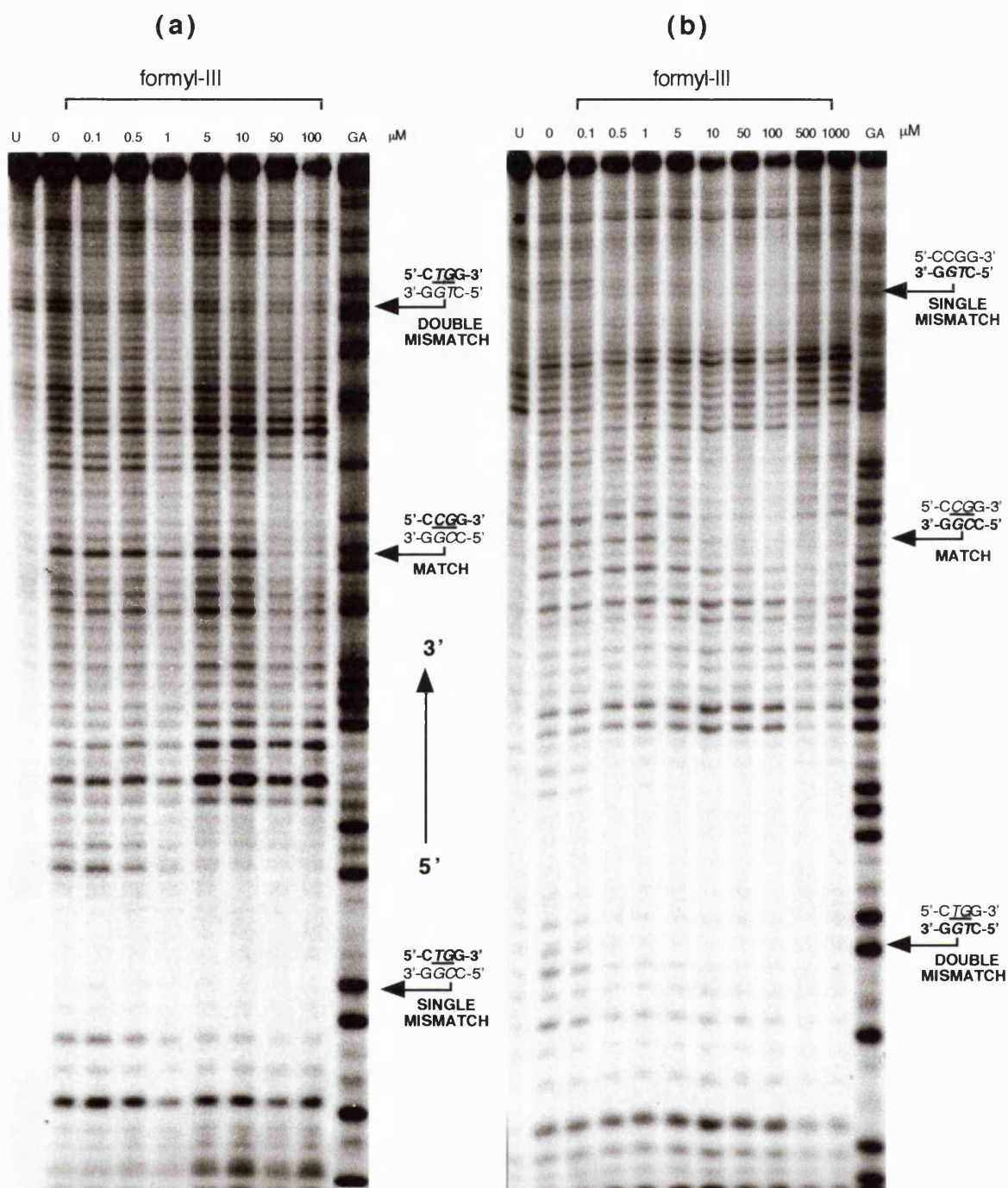
### 3.4 Results

#### 3.4.1 Recognition of a single TG mismatch using the tri-imidazole formyl-III

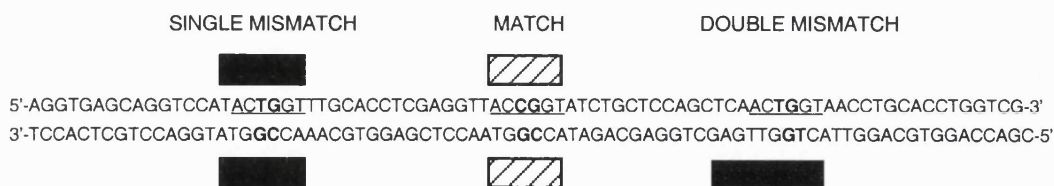
DNase I footprinting titrations on both strands of the oligodeoxynucleotide sequence IMGT1/2, containing a series of CG match and TG mismatch sites, provided an evaluation of the binding and relative specificity of the tri-imidazole, formyl-III (Figure 3.3) for TG mismatches along a single DNA segment. Comparison of the relative binding preference of formyl-III at a CG match site (5'-CCGG-3'/3'-GGCC-5'), a single TG mismatch site (5'-CTGG-3'/3'-GGCC-5') and a double TG mismatch site (5'-CTGG-3'/3'-GGTC-5') enabled an assessment of the affinity of the imidazole/imidazole ring pair in a 2:1 binding motif for TG mispair recognition (Figure 3.4 A). Footprinting was carried out on both strands to accurately assign binding sites, as the cleavage by DNase I of the DNA sequences examined was not sequence neutral.

Typical footprint titrations along both 80 base pair oligodeoxynucleotide strands are shown in Figure 3.5. Binding by formyl-III was observed at both the CG match and single TG mismatch sites on each strand. However, whilst strong binding was observed at the single TG mismatch site (5'-CTGG-3'/3'-GGCC-5') over a concentration range of 0.5µM and above, only weak binding was observed at the CG match site (5'-CCGG-3'), initiating at 50µM. Binding at the consecutive TG double mismatch site (5'-CTGG-3'/3'-GGTC-5') was harder to identify clearly. Multiple footprinting titrations of the top strand (IMGT1) appeared to indicate little binding at this site (Figure 3.5 (a)). However, further footprint titrations of formyl-III on the corresponding oligodeoxynucleotide strand (IMGT2) appeared to show strong binding (0.5µM and above) alongside the double TG mismatch site (Figure 3.5 (b)). No binding was observed at any other sites along the 80 base pair sequence. A diagrammatic representation of the binding of formyl-III is shown in Figure 3.6.





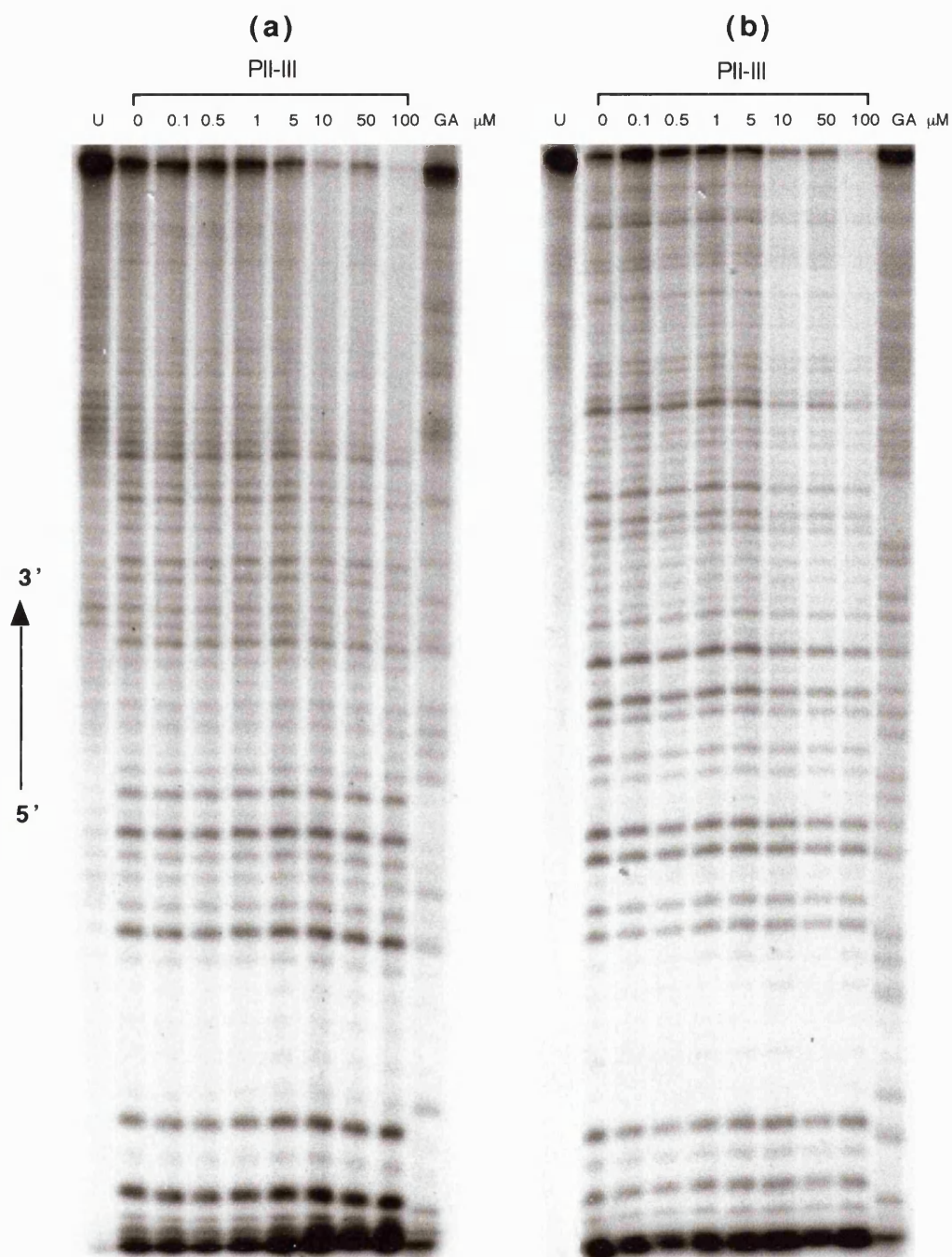
**Figure 3.5** DNase I footprint experiments with formyl-III on both strands of a 5'-<sup>32</sup>P labeled 80bp DNA sequence containing a core match site CCGG, a single TG mismatch site CTGG and a double TG mismatch site CTGG. All reactions contain 0.5μg of DNA, 10mM Tris pH 7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. (a) Top strand; U is the untreated control, 0 is the DNase I treated control and GA is the purine marker reaction. (b) Bottom strand; U is the untreated control, 0 is the DNase I treated control and GA reaction is the purine marker reaction.



**Figure 3.6** Box diagram depicting the binding of formyl-III to the annealed oligodeoxynucleotides IMGT1/IMGT2, whereby the solid box denotes strong binding (0.5μM and above) and the hatched box weaker binding (50μM and above). Binding at the double mismatch site was only resolved for the bottom strand. Putative binding sites are underlined, with the Watson-Crick 'match' site and the TG mismatch sites indicated in bold type.

#### 3.4.2 Recognition of a double TG mismatch using the hairpin heterodimer PII-III

DNase I footprinting titrations on both strands of the oligodeoxynucleotide sequence IMGTO5/6, containing a series of maximum overlap CG match and TG mismatch sites, enabled the binding and relative specificity of the hairpin-linked heterodimer PII-III (Figure 3.3) to be assessed at a single and a consecutive double TG mismatch along a single DNA segment. Typical footprint titrations along both of the oligonucleotide strands IMGTO5 and IMGTO6 are shown in Figure 3.7. No discrete footprints were observed at any sites along the 78 base pair sequence on either strand, although full length DNA was progressively lost with increasing ligand concentration.



**Figure 3.7** DNase I footprint experiments with PII-III on both strands of a 5'-<sup>32</sup>P labeled 78bp DNA sequence containing a core match site 5'-CCG-3'/3'-GGC-5', a single TG mismatch site 5'-CTG-3'/3'-GGC-5' and a double TG mismatch site 5'-CTG-3'/3'-GGT-5'. All reactions contain 0.5μg of DNA, 10mM Tris pH7.0, 1mM EDTA, 50mM KCl, 1mM MgCl<sub>2</sub>, 0.5mM DTT and 20mM Hepes. (a) Top strand; U is the untreated control, 0 is the DNase I treated control and GA is the purine marker reaction. (b) Bottom strand; U is the untreated control, 0 is the DNase I treated control and GA reaction is the purine marker reaction.

### 3.5 Discussion

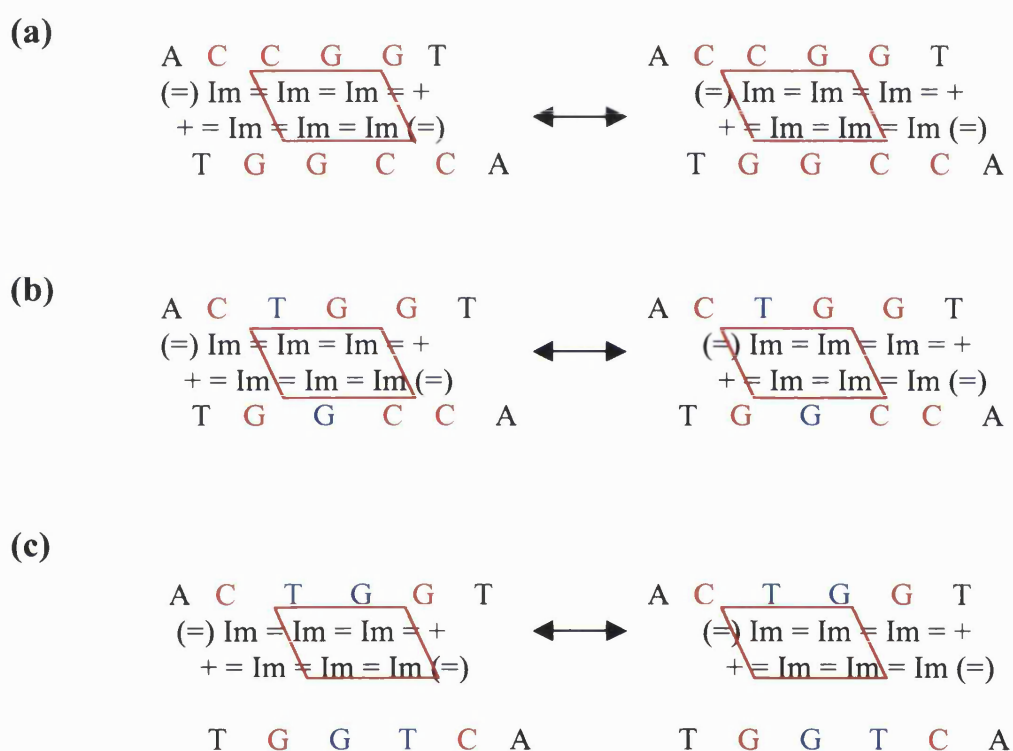
DNase I footprinting was used to examine the binding and relative selectivity of the polyamides formyl-III and PII-III for TG mismatches placed in series within long synthetic DNA segments. A double-stranded oligodeoxynucleotide was designed for each compound to contain three putative binding sites, a CG match, a single TG mismatch and a consecutive TG double mismatch. These were present in the favoured binding mode of each compound, a one-residue stagger motif for formyl-III and a maximum overlap motif for PII-III. The occurrence of all three sites within a single heteroduplex enabled evaluation of the relative site preference for both the tri-imidazole monomer formyl-III, binding in a 2:1 ratio with the DNA, and the hairpin-linked heterodimer PII-III. This provided a direct assessment of the suitability of the Im/Im ring pair for either single or consecutive double TG mismatch recognition in the context of the adjacent heterocycle rings present during side-by-side binding.

Formyl-III bound at both the Watson-Crick match, 5'-CCGG-3', and single TG mismatch sites, 5'-CTGG-3'. There was, however, a hundred-fold preference in binding at the single mismatch, relative to the Watson-Crick match site. This confirms previous work by Lee and coworkers using SPR, which demonstrated a thirty-fold increase in binding affinity for the single mismatch, CTGG, over the match site, CCGG, using short DNA hairpins (Lacy *et al.*, 2002b). In that study binding by formyl-III was also seen at a double mismatch site, CTGGdm, which whilst weaker than that observed at other sites, was of a similar affinity to that seen with distamycin at A<sub>3</sub>T<sub>3</sub> (Lacy *et al.*, 2002b). In the present study, however, binding at the consecutive TG double mismatch site was less clear, with no discernable binding observed along the top strand of the heteroduplex (IMGT1). On this strand it is possible that binding at the double mismatch site may have been masked by bands produced by the presence of shortened mismatched oligonucleotides present as impurities. This seems likely given the bands observed at that site in the uncleaved control lane. Further footprinting using sequence neutral cleavage agents, such as hydroxyl radicals or MPE, may ascertain binding by formyl-III at the double mismatch site more clearly. Nevertheless, the preferential binding observed here by formyl-III at the single mismatch site suggests that the Im/Im ring pair,



present during side-by-side binding of two formyl-III monomers, also favours TG rather than CG base pairs when they arise in a single DNA sequence.

Whilst the three putative target sites were designed for binding by formyl-III in the one-residue stagger motif, it cannot be excluded that formyl-III may also be binding here in a maximum overlap motif as shown in Figure 3.8. This should be a less favourable conformation, however, due to the presence of a leading formamide head group. Indeed, formyl-III has been shown to bind only in the one-residue stagger motif to the core sequence TGG in previous studies (Yang *et al.*, 1999b).

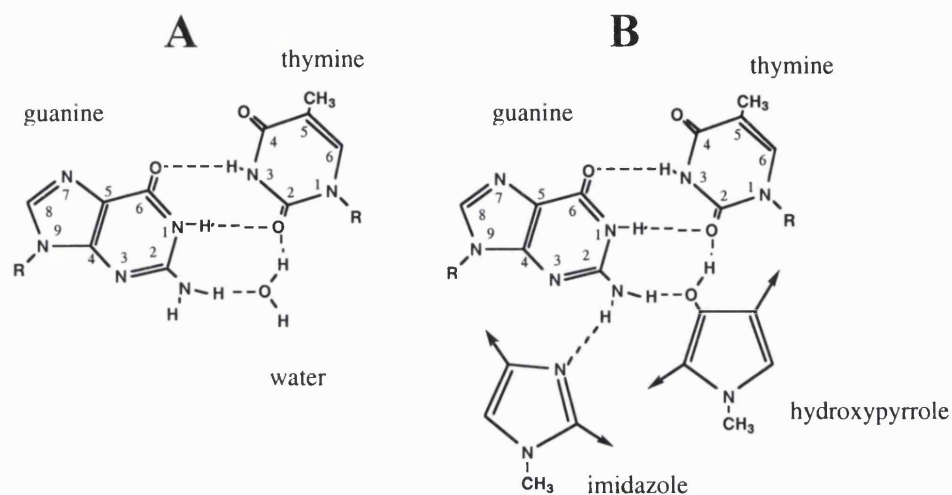


**Figure 3.8** Possible binding modes for formyl-III in the maximum overlap motif at (a) the CG match, (b) the single TG mismatch and (c) at the double TG mismatch sites along IMGT1/2, where Im is imidazole and Py is pyrrole. The equal sign, =, signifies an amide with, (=), denoting the leading formamide head group (NHCHO) and the plus sign, +, signifying the charged tail ((CH<sub>2</sub>)<sub>2</sub>NMe<sub>2</sub>H<sup>+</sup>). The DNA target sequence is indicated in red capital letters, with TG mismatched base pairs indicated in blue.

Discrimination of a single and double consecutive T: G mismatch was also assessed using the hairpin-linked polyamide heterodimer PII-III. Binding was examined at a putative Watson-Crick site and at a single and double consecutive TG mismatch, all designed for ligand binding in a maximum overlap motif. No discrete binding was seen, however, by PII-III at either the CG or TG binding sites present. The observed loss of the full-length band with increasing ligand concentration suggests that non-specific binding was occurring at high doses. The reason for the lack of binding specificity demonstrated by this polyamide is unclear. The complete absence of binding initially suggests that the structure of the polyamide, for example the linkage, is sub-optimal for discrete binding at the given sites. However, the sequences or structural conformation of the putative target or flanking sites may also restrict binding. Alternatively the constructed oligonucleotide duplex itself may be unstable. It is possible that the presence of the two mismatched sites in the double stranded annealed heteroduplex IMGTO5/IMGTO6 produce sufficient instability to result in the disintegration of the duplex into its two constituent single strands, thereby preventing discrete ligand binding.

The work presented in this chapter, examining the binding of formyl-III to a series of putative target sites present along a single oligonucleotide confirms previous NMR, CD, SPR and thermodynamic studies (Yang *et al.*, 1999b; Lacy *et al.*, 2002b), showed preferential binding at a single TG mismatch. Weaker binding was observed at a neighbouring Watson-Crick CG binding site, whilst strong binding at a consecutive TG double mismatch was resolved along only one strand of the heteroduplex examined. No discrete binding was observed with the hairpin-linked heterodimer PII-III. Nevertheless, the Im/Im ring pair has been shown here to be strongly selective for a single TG mismatch when placed in the same DNA sequence as an alternative Watson-Crick CG match site. This indicates that the Im/Im ring pair is a suitable recognition motif for TG mismatches, thereby extending the current set of heterocycle pairing rules. Lee and coworkers, however, have found that formyl-III is unable to discriminate TG from GT (Yang *et al.*, 1999b). Whilst this would not prevent the use of formyl-III in TG mismatch screening protocols, for example incorporated into a sensor chip for pharmacogenomic diagnosis, future gene targeting strategies may require differentiation

of degenerate non-canonical base pairs. This could be provided perhaps by further heterocycle pairings, such as imidazole-hydroxypyrrole (Im/Hp) as shown in Figure 3.9, or heterodimers, including ImImPyPy-PyPyImPy (Rucker *et al.*, 2003).



**Figure 3.9** Molecular basis for the possible recognition of TG mismatches by an imidazole/hydroxypyrrole ring pair, adapted from Yang *et al.*, 1999b. **A.** A conserved H<sub>2</sub>O bridging the G: T mismatch as observed in the crystal structure of Hunter *et al.* 1987 **B.** Potential discrimination of G: T from T: G by and Im/Hp ring pair as proposed by Yang and coworkers whereby the hydroxy group of Hp may adopt the role of the bridging water molecule shown in A.



## CHAPTER 4

### CELLULAR UPTAKE AND LOCALISATION OF THIAZOLE-CONTAINING POLYAMIDES AND POLYAMIDE-PBD CONJUGATES.

#### 4.1 Introduction

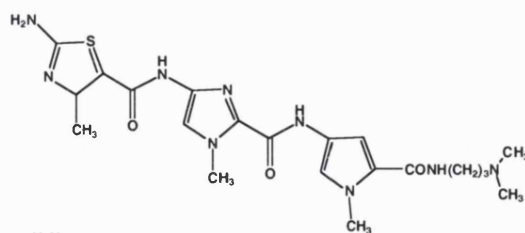
Whilst polyamides have repeatedly been shown to bind in a sequence-specific manner *in vitro*, capable of modulating transcription (Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998; Dickinson *et al.*, 1999a, b; Mapp *et al.*, 2000) and binding selectively within nucleosomal DNA (Gottesfeld *et al.*, 2001; Gottesfeld, *et al.*, 2002) and within isolated cell nuclei and chromosomes (Janssen *et al.*, 2000a; Maeshima *et al.*, 2001; Gygi *et al.*, 2002), substantiation of nuclear targeting in living cells has been less conclusive. Indirect evidence has been provided by the specific gain or loss of function phenotypes in *Drosophila melanogaster* when fed satellite-specific polyamides (Janssen *et al.*, 2000b), as well the inhibition of viral replication in human lymphocytes (Dickinson *et al.*, 1998) and targeted transcriptional derepression from the HIV-1 LTR (Coull *et al.*, 2002). Furthermore, polyamides conjugated to both major and minor groove binding alkylating agents such as chlorambucil and the CC-1065 analogue, CBI, have produced sequence-specific lesions in cells, causing G<sub>2</sub>/M arrest and DNA synthesis and growth inhibition (Wang *et al.*, 2002, 2003).

Recent work, however, examining the intracellular distribution of fluorophore-tagged polyamides by direct visualisation using fluorescence microscopy has provided only ambiguous evidence of live cell uptake. Whilst intracellular polyamide localisation has been confined to the cytosol or mitochondria in ovarian and breast cell lines (Sharma *et al.*, 2001; Belitsky *et al.*, 2002), nuclear accumulation has been observed in some T cell lines, suggesting uptake may be cell-type dependent (Belitsky *et al.*, 2002). Furthermore, *in vitro* binding studies of fluorophore-polyamide conjugates indicate that covalent linkage of a fluorophore to a polyamide may decrease its DNA binding affinity relative to untagged molecules with the equivalent number of amide units (Battacharya *et al.*, 2002). Thus, attachment of a fluorescent or an alkylating moiety to the polyamide, changing its size and electrostatic properties, could significantly transform

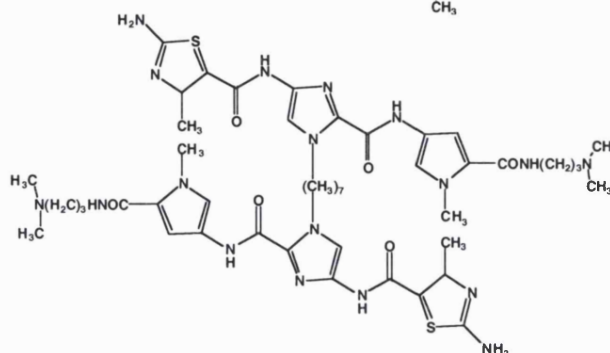
the binding and uptake characteristics of the conjugate and may not provide a representative model of unconjugated polyamide uptake in living cells.

Due to the aromatic content of polyamides, they may fluoresce upon excitation with UV (~345nm), emitting light in the blue range of the visible spectrum. This effect is potentiated when they are bound to DNA (Baliga & Crothers 2000) and may obviate the need for fluorophore conjugation. Exploiting this property, a study of cell uptake and intracellular localisation of an untagged cross-linked thiazolated polyamide dimer, TIP-C7-amino (Figure 4.1), was undertaken to evaluate the efficacy of intracellular uptake and distribution in two human cell lines, the breast line MCF-7 and leukaemic T cell line Jurkat, using fluorescence microscopy. Uptake was compared to a range of covalent and non-covalent synthetic minor groove binding ligands, including the TIP-amino monomer (Figure 4.1), a hairpin polyamide dimer, JH-37, a pyrrolobenzodiazepine (PBD) dimer, SJG-136, and a novel PBD-polyamide conjugate, GWL-6 (Figure 4.2). Previous studies of SJG-136 and GWL-6 have demonstrated potent cytotoxic activities, which correlate with an ability to produce covalent DNA adducts in cells (Gregson *et al.*, 2001; Thurston and coworkers, personal communication). This suggests that these molecules are able to traverse both the plasma and nuclear membranes, and therefore may be used to assess the inherent sensitivity of this method in the examination of drug uptake in living cells. Furthermore, the effect of TIP-C7-amino on growth was evaluated using standard anti-proliferative assays.

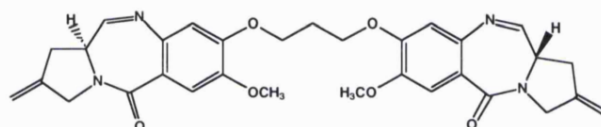
**TIP-amino**



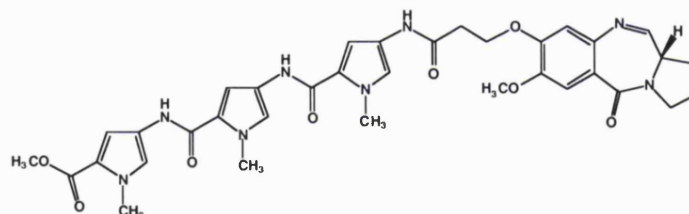
**TIP-C7-amino**



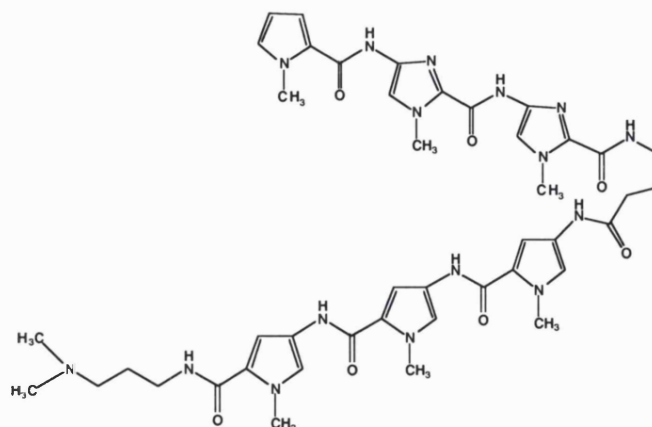
**SJG-136**



**GWL-6**



**JH-37**



**Figure 4.1** Molecular structures of the polyamide monomer TIP-amino, cross-linked dimer TIP-C7-amino, PBD dimer SJG-136, the polyamide-PBD conjugate GWL-6 and the hairpin linked polyamide dimer JH-37.

#### 4.2.1 Investigational compounds

Compounds were kindly provided by Professor J. William Lown, Department of Chemistry, University of Alberta, Edmonton, Canada, (TIP-amino and TIP-C7-amino), Professor David Thurston, Department of Pharmaceutical and Biological Chemistry, The School of Pharmacy, University of London, London, UK (SJG-136 and GWL-6) and Professor Moses Lee, Department of Chemistry, Furman University, Greenville, North Carolina, USA (JH-37). The synthesis and characterisation of the compounds have been reported (Sharma *et al.*, 2000; Gregson *et al.*, 2001). A minimum of 1mg of each compound was dissolved in the appropriate solvent at a 10mM stock concentration and stored at -20°C until use. Dilutions were freshly prepared for each experiment in distilled and deionised water. Compounds requiring organic solvents such as DMSO or DMA to make up the stock were serially diluted in aqueous media such that the final organic solvent concentration was less than 0.1% (v/v).

#### 4.2.2 Chemicals

Tris-hydroxymethyl-methylamine base (Tris base), Hoechst bisbenzimidazole 33258 (Hoechst 33258), 3,8-Diamino-5-(3-diethylaminopropyl)-6-phenylphenanthridinium iodide methiodide (propidium iodide), (+)-1-(2,5-dimethoxyphenyl)-2-aminopropane hydrochloride (DMA) tetramethylrhodamine isothiocyanate (TRITC)-conjugated dextran, sucrose and sulphorhodamine B were purchased from Sigma.

2-(6-Amino-3-imino-3H-xanthen-9-yl) benzoic acid methyl ester (Rhodamine 123), 5-carboxyfluorescein diacetate acetoxymethyl ester (CFDA AM) and 3,6-bis(Dimethylamino)acridinium chloride (acridine orange) were purchased from Molecular Probes. Methanol, acetic acid, dimethyl sulfoxide (DMSO) and trichloroacetic acid (TCA) were purchased from BDH.

#### 4.2.3 Buffers

Phosphate buffered saline (PBS) tablets were purchased from Sigma and dissolved in deionised water to yield 0.01M phosphate buffer, 0.0027M potassium chloride and 0.137M sodium chloride, pH7.4 at 25°C.

#### 4.2.4 Culture medium

Dulbecco's Minimum Essential Medium (DMEM), Hams F-12 and RPMI 1640 were supplemented with 10% (v/v) foetal calf serum and 2mM glutamine, which were all purchased from Autogen Bioclear.

#### 4.2.5 Kits

A Nuclei EZ Prep Nuclei Isolation kit, containing Nuclei EZ lysis buffer and Nuclei EZ storage buffer was purchased from Sigma.

### 4.3 **Methods**

#### 4.3.1 Cell culture

The MCF-7 human breast adenocarcinoma cell line (Soule, *et al.*, 1973) was grown as a monolayer culture in DMEM: Hams F-12 medium supplemented with 10% (v/v) foetal calf serum and 2mM glutamine. The Jurkat human leukaemic T cell lymphoblast cell line (Schneider, *et al.*, 1977) was grown in suspension culture in RPMI 1640 medium supplemented with 10% (v/v) foetal calf serum and 2mM glutamine. Cells were incubated at 37°C in an atmosphere of 5% (v/v) carbon dioxide (CO<sub>2</sub>) in air and maintained in an exponential growth phase. Cells were passaged using conventional procedures and screened for mycoplasma regularly.

#### 4.3.2 Image Acquisition and Processing

A Zeiss Axiophot epifluorescence microscope equipped with a mercury halide light source and narrow band blue, green and red filters, was used to view cells which were visualised with a 40x (NA 0.75) Plan Neofluar phase objective. Images were captured using Image Pro Plus software either as phase contrast or in the corresponding blue, green or red channels using a Coolview 12 CCD camera (Photonic Science). The filter set was custom made so that UV excitation was blue emitted, blue excitation was green emitted, green excitation was red emitted and red excitation was far red emitted. The mercury halide bulb was allowed to burn for at least 20 minutes prior to use to stabilise emission intensities and the bulb had been used for less than 50 hours for most experiments. Cells were found under phase contrast so there was no prior epifluorescent

illumination of cells to exclude photobleaching effects. On identification of a suitable cell area, trans-illumination was switched off and an epifluorescence exposure made. Exposure times and gain (electronic signal amplification) were optimised using Image Pro Plus image capture software so that pixel intensities were in the linear range to avoid pixel saturation. Exposure and gain were then fixed for a given experiment and images captured for at least six randomly selected areas in a typical sample. All relevant controls were contained in every individual experiment so that any changes in absolute fluorescence intensity, due to for example the number of hours that the bulb had been used, were controlled. Cross-talk, whereby a single fluorescent marker may be visualised in multiple emission channels, was eliminated not only by the use of a custom built filter set but also by optimisation of the marker concentrations used in singly stained samples which were viewed across the channels and concentrations and times of staining were chosen such that no signals were detected in inappropriate channels. Some dyes, for example acridine orange and rhodamine 123, have broad band emission and will fluoresce in two channels (red/green) even with narrow band filters. In these cases, however, checks were made to ensure no cross-talk into the blue channel (i.e. the drug channel) took place. In experiments using multiple fluorescence markers the images for the most photosensitive channel (blue) were recorded first, followed by the green and red images.

Fluorescence data sets were captured at constant exposures for every experiment to record quantitative differences in fluorescence images and stored as 12 bit grayscale data sets as Image Pro Plus files without any further manipulation or processing. Subsequently, images were retrieved under Image Pro Plus and as appropriate subjected to image enhancement by contrast stretching, whereby the absolute pixel intensities recorded are expanded to occupy the full 12 bit image depth, and were resaved as secondary images. Image Pro Plus files were then converted from 12 bit to 16 bit data sets to render them compatible with Adobe Photoshop 6.0. The images presented in the figures were derived from this tertiary data set. In the case of the phase contrast images these were subjected to adjustment of the contrast levels in order to normalise the range of image contrast and background for the different experiments. The fluorescence

images were not adjusted so as to preserve the original intensity differences captured microscopically.

#### 4.3.3 Visualisation of drug uptake in living cells using fluorescence microscopy

##### *Adherent cells:*

A total of  $5 \times 10^4$  MCF-7 human breast cancer cells were seeded onto 13mm round coverslips in flat-bottomed 24 well plates and allowed to grow for 48 hours at 37°C in a volume of 1ml of DMEM: Hams F-12 medium supplemented with 10% (v/v) foetal calf serum and 2mM glutamine. Drug treatments were carried out by replacement of the growth medium with 0.5ml of fresh medium, to which 10µl of drug appropriately diluted in sterile dH<sub>2</sub>O had been added. Cells were incubated with drug for 1, 2, 24, 48 or 72 hours and one hour prior to examination were counter-stained with individual or combinations of the fluorescent markers 5-carboxyfluorescein diacetate acetoxymethyl ester (CFDA AM) (1µM), rhodamine 123 (10µM), Hoechst 33258 (10µM) or acridine orange (10µM), which were added to the culture medium. Staining with propidium iodide (0.01mgml<sup>-1</sup>) was carried out 10 minutes prior to visualisation. Following this, the medium was aspirated and cells were washed once with phosphate buffered saline. The coverslips were removed from the wells, mounted onto glass microscope slides and viewed immediately as described above.

##### *Suspension cells:*

A total of  $1.8 \times 10^5$  Jurkat human T lymphoid cells were aliquotted in a volume of 100µl per well into 24 well plates. Drug treatments were carried out in a volume of 1ml of the growth medium RPMI 1640, supplemented with 10% (v/v) foetal calf serum and 2mM glutamine, to which 10µl of drug appropriately diluted in sterile dH<sub>2</sub>O had been added. Cells were then grown in suspension at 37°C for 24, 48 and 72 hours. Post-incubation, the medium was aspirated from the settled suspension cells, which were washed once with PBS. After allowing the cell suspension to resettle, the PBS wash was aspirated and a volume of 10µl of residual buffer containing the settled cells was spotted onto a glass microscope slide. A coverslip was applied and excess cell suspension was drawn



off using lens tissue. Phase and corresponding fluorescence images were captured as described in section 4.3.2.

#### 4.3.4 Visualisation of drug uptake in fixed cells using fluorescence microscopy

MCF-7 cells were aliquotted into 24 well plates and grown as described in section 4.3.1 for 48 hours at 37°C. Cells were then fixed for 3 minutes in methanol and washed once with PBS. Fixed cells were treated in PBS with 10µM concentrations of Hoechst 33258, TIP-C7-amino, TIP-amino, distamycin, SJG-136, GWL-6 and JH-37 for 1 hour. Post-incubation, cells were washed once with PBS. The coverslips were mounted onto a glass microscope slide and images captured as previously described.

#### 4.3.5 Microinjection of live cells

MCF-7 cells were seeded at  $5 \times 10^4$  cells ml<sup>-1</sup> in Petri dishes (25mm), each containing a single glass coverslip (13mm), allowed to adhere and grown in DMEM: Hams F-12 medium supplemented with 10% (v/v) foetal calf serum and 2mM glutamine for 48 hours in at 37°C. TIP-C7-amino was mixed with tetramethyl rhodamine isothiocyanate-conjugated dextran (TRITC-dextran) to yield final concentrations of 0.3mg/ml polyamide and 1mg/ml fluorochrome marker in phosphate buffered saline. MCF-7 cells were microinjected (50 cells per coverslip) with the polyamide/fluorochrome mix using an Eppendorf transjector 5246 and an Eppendorf hydraulic manual micromanipulator 5171, whilst viewed under a Zeiss IM35 microscope. Needle capillaries used to inject the polyamide/TRITC-dextran mix were pulled from filament-containing borosilicate glass to diameters of 0.2-0.5 mm using an Eppendorf micropipette puller. Cells were allowed to recover by incubating the cultures at 37°C for 1 hour, after which the coverslips were removed from the Petri dishes and mounted onto glass microscope slides. Cells were viewed and images captured as detailed in section 4.3.2.

#### 4.3.6 Visualisation of drug uptake in isolated nuclei using fluorescence microscopy

##### *Isolation of cell nuclei:*

A total of  $2 \times 10^7$  Jurkat cells were harvested from a suspension culture kept growing in log phase as described previously. The cells were spun down at 1500rpm for 5 minutes

at 4°C, the supernatant aspirated and the cell pellet set on ice. To ensure a single cell suspension and avoid clumping, the pellet was washed by vortexing briefly and resuspending in 1ml of ice-cold Dulbecco's Phosphate Buffered Saline (PBS). This was vortexed again and a further 9ml of cold PBS was added to the cell suspension, which was then set on ice. Cells were spun down at 1500rpm for 5 minutes at 4°C and the supernatant aspirated. The cell pellet was vortexed briefly and resuspended in 0.5ml of ice-cold Nuclei EZ lysis buffer. The suspension was re-vortexed and a further 3.5ml of lysis buffer was added. The cells were allowed to lyse on ice for 10 minutes and nuclei were collected by centrifugation at 1500rpm for 5 minutes at 4°C. The supernatant containing cytoplasmic components was then aspirated and the lysis step repeated. Following the second round of lysis, the suspension was homogenised using a Dounce homogeniser, layered in 500µl volumes over 2ml of 30% (w/v) sucrose solution made up in Nuclei EZ lysis buffer and spun at 3000rpm for 10 minutes. Following centrifugation, the supernatant was aspirated and the pelleted nuclei were resuspended in a total of 100µl of Nuclei EZ storage buffer.

#### *Drug treatments and visualisation of drug uptake in isolated cell nuclei:*

A 1:10 dilution of the stock nuclei was made in Nuclei EZ storage buffer for each drug treatment. Diluted nuclei were then treated on ice for 10 minutes with either 10µM TIP-amino, TIP-C7-amino, GWL-6, JH-37, propidium iodide, Hoechst 33258. Diluted nuclei and intact cells were also incubated with 10µM CDFA AM for 20min. These samples were incubated at 37°C due to the metabolic nature of the enzymatic cleavage of CDFA AM to its fluorescent state. Following treatment, a 5µl aliquot of the treated nuclei or cells was dropped onto a microscope slide, covered with a glass coverslip and visualised as described in section 4.3.2. Images were captured under phase contrast and in the appropriate channel, blue for TIP-amino, TIP-C7-amino, GWL-6, JH-37 and Hoechst 33258; red for propidium iodide and green for CDFA AM.

#### 4.3.7 Anti-proliferative activity

The anti-proliferative activity of the polyamide TIP-C7-amino was evaluated in both the adherent MCF-7 breast cell line and in the suspension Jurkat T cell line using standard

assays. The sulphorhodamine B (SRB) assay (Rubinstein, *et al.*, 1990) based on colorimetric measurement of SRB stain associated with intracellular protein in viable cells was used to determine growth inhibition in the adherent line. Cellular proliferation was measured in the suspension line by counting cells using a haemocytometer, as employed in the determination of a standard growth curve.

#### *Adherent cells:*

MCF-7 cells were plated out in triplicate at  $1.5 \times 10^3$  cells in a volume of 100 $\mu$ l per well in 96 well flat-bottomed microtitre plates and allowed to adhere overnight. Drug treatments were carried out the following day by the addition of an equal volume of media (100 $\mu$ l) containing double the final drug concentration. Cells were incubated with drug for 120 hours at 37°C, when cells in the untreated controls approached confluence. Post-incubation, 100 $\mu$ l of 30% (v/v) acetic acid was added to each well and the cells were left to fix for 20 minutes at 4°C. The acetic acid was removed by flicking the plates, which were then carefully washed 5x with tap water and air-dried. Following this, 100 $\mu$ l of the protein stain sulphorhodamine B (0.4% (w/v) in 1% (v/v) acetic acid) was added to each well and plates were incubated at room temperature for 20 minutes. Cells were washed 5x with 1% (v/v) acetic acid and air-dried for 1 hour. The SRB dye was solubilised by the addition of 100 $\mu$ l of 10mM Tris base to each well and the plates were left for a further 20 minutes at room temperature. The absorbance at 540nm ( $A_{540nm}$ ) was read using an automatic plate reader (Tecan Spectrofluor Plus) equipped with XFLUOR software. The fraction of control growth was calculated from the following equation:

$$\text{Fraction of control growth} = A_{540nm} \text{ of treated cells} / A_{540nm} \text{ untreated control}$$

The mean fraction of control growth and standard deviation for each concentration was calculated.

### *Suspension cells:*

A total of  $1.8 \times 10^5$  Jurkat cells were plated out in quadruplicate in a volume of 100 $\mu$ l per well in a 24 well plate. Drug treatments were carried out in a volume of 1ml of the growth medium RPMI 1640, supplemented with 10% (v/v) foetal calf serum and 2mM glutamine, to which 10 $\mu$ l of drug appropriately diluted in sterile dH<sub>2</sub>O had been added. Cells were incubated with drug for up to 120 hours at 37°C. At 24, 48, 72, 96 and 120 hours, the cells in each well were resuspended using a Pasteur pipette and 30 $\mu$ l aliquots were taken and counted using a haemocytometer. Five fields on both sides of the haemocytometer were counted for each well for each drug concentration using an Olympus CM2 phase contrast microscope. The mean cell count and standard deviation was calculated for each drug concentration and expressed as the fraction of control growth for each time point as follows:

$$\text{Fraction of control growth} = \text{Mean count of treated cells} / \text{Mean count of untreated control}$$

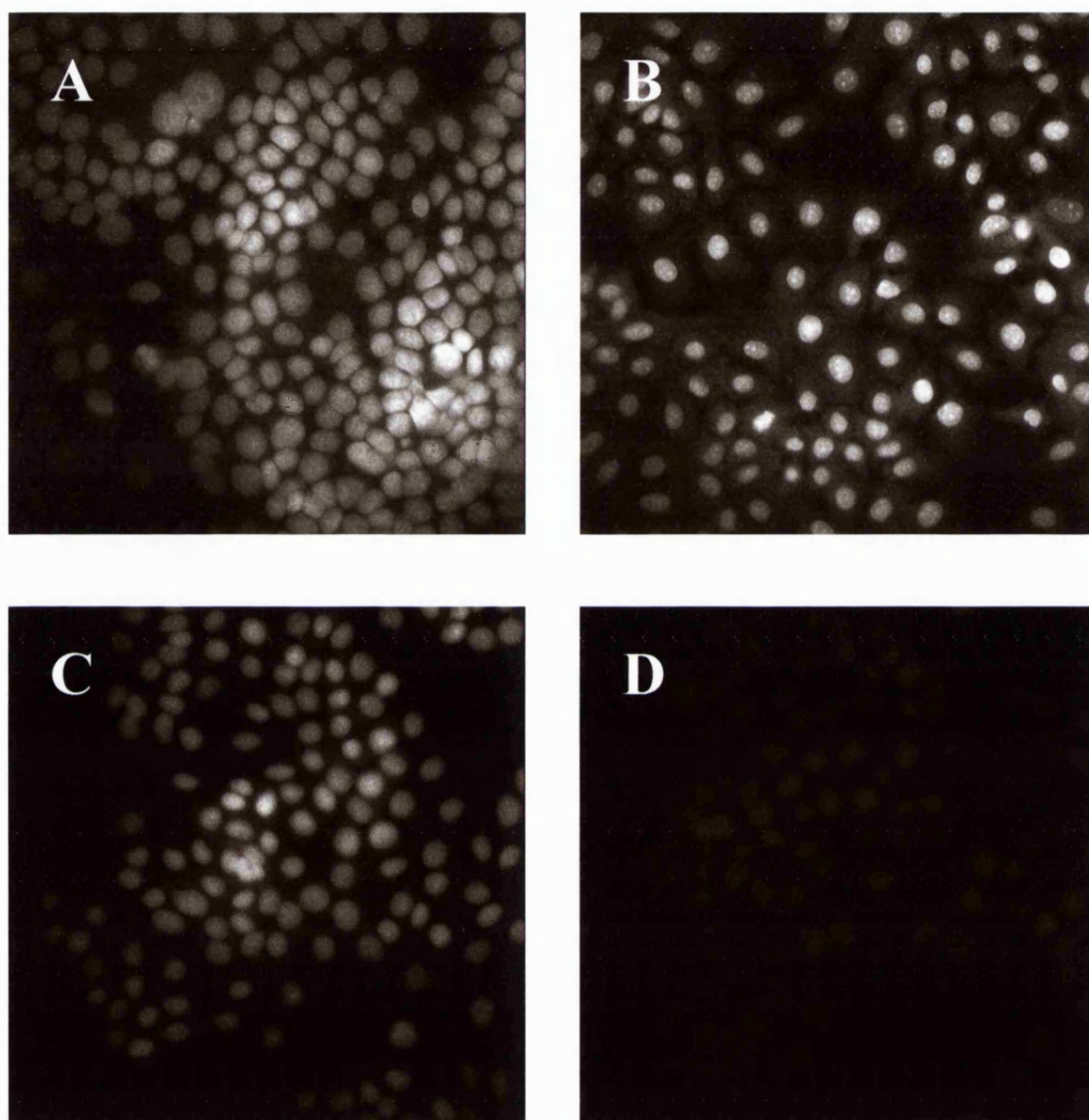
To assess the proportion of viable cells remaining after 72 and 120 hours, cells were double-stained with 1 $\mu$ M CFDA AM and 0.01mgml<sup>-1</sup> propidium iodide for 1 hour. A 5 $\mu$ l aliquot from each well was then dropped onto a glass microscope slide, covered with a coverslip and cells were viewed using a Zeiss epifluorescence microscope equipped with a mercury halide light source and narrow band blue, green and red filters with a 40x (NA 0.75) Neofluar objective. The percentage of viable cells was calculated by counting the number of cells stained with CFDA AM as a proportion of the total cell number in 3 fields of view for each well aliquot. At 72 and 120 hours, only one plate, containing four replicates for each given concentration, was counter-stained. Therefore, a mean and the standard deviation of the mean of three independent experiments could not be calculated in this case.

## 4.4 Results

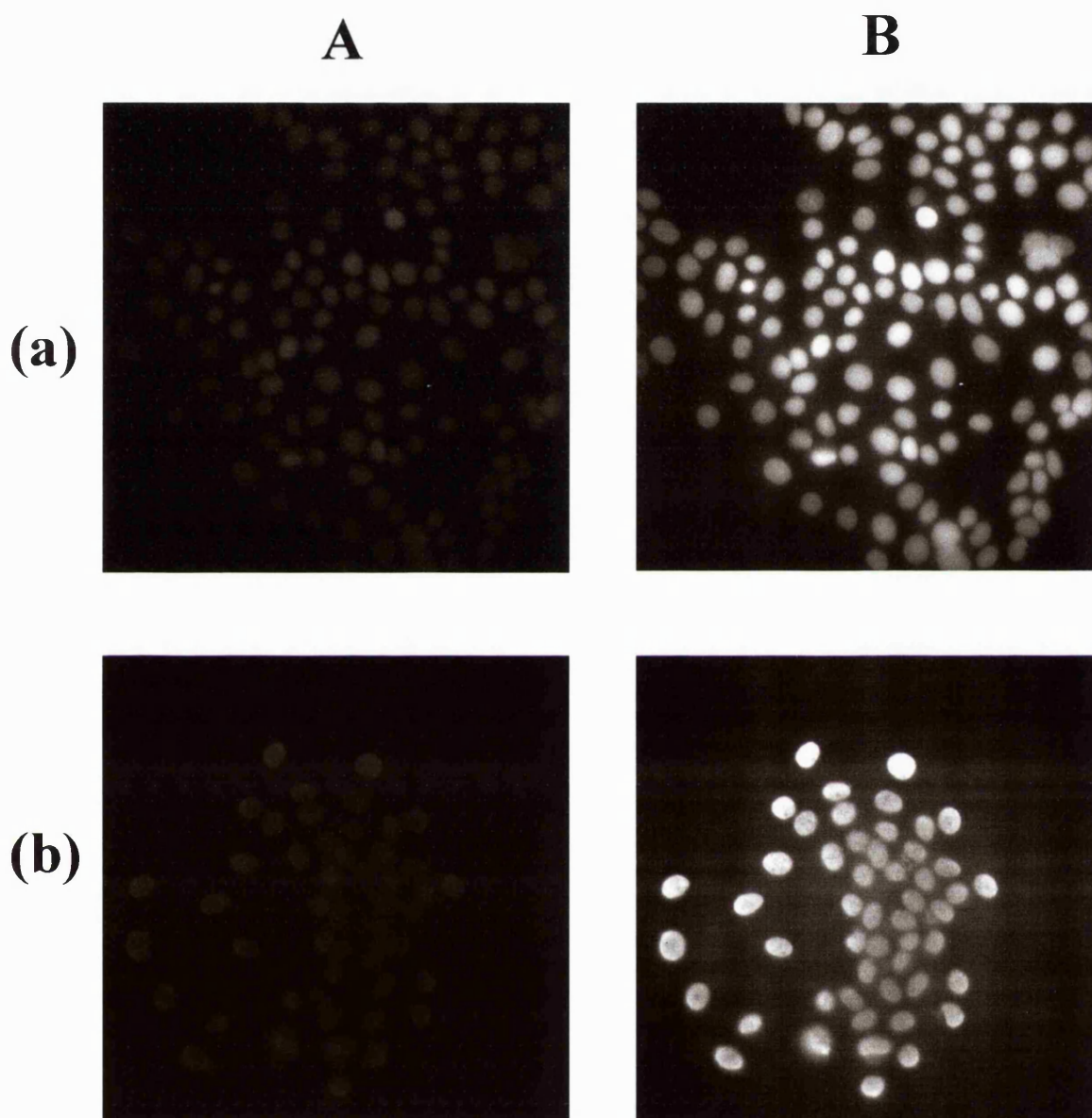
### 4.4.1 Relative fluorescence and nuclear binding of the minor groove-binding agents distamycin, TIP-amino, TIP-C7-amino, TIP-C8-amino, JH-37, GW6 and SJG-136 in fixed cells.

To establish nuclear binding and the relative fluorescence of the non-covalent and covalent minor groove binders distamycin, TIP-amino, TIP-C7-amino, TIP-C8-amino, JH-37, GWL-6 and SJG-136 in permeabilised cells, MCF-7 cells were fixed and treated for 1 hour with equimolar doses of these agents and visualised in the blue emission channel. As a positive control, their fluorescence and distribution was compared to an equimolar dose of the cell membrane permeant DNA selective dye Hoechst 33258, which also requires UV excitation and emits at 461nm in the blue range of the visible spectrum. Images were captured at optimised exposures, to enable comparison of the fluorescent intensity produced by these compounds on excitation with UV. Under these conditions, Hoechst 33258 produced substantially stronger fluorescence in the blue emission range than the other minor groove binders studied, showing discrete nuclear staining (Figure 4.2 A). This image required a four-fold decreased exposure compared to images captured for the other agents, due to the brightness of the fluorescence.

After Hoechst 33258, the polyamide dimer, TIP-C8-amino, produced the next strongest fluorescence on excitation with UV followed by TIP-C7-amino, with discrete nuclear binding observed in permeabilised cells (Figure 4.2 B & C). In contrast, the polyamide monomer, TIP-amino, was only weakly fluorescent, with nuclear binding barely visible without contrast enhancement at this exposure (Figure 4.2 D). The PBD dimer, SJG-136, also showed nuclear staining producing a weak but distinct fluorescence signal whilst the PBD-polyamide conjugate, GWL-6, showed trace fluorescence and nuclear binding (Figure 4.3 B (a) & (b)). Distamycin and the hairpin polyamide dimer, JH-37 did not produce detectable fluorescence when raw data images were examined (Figure 4.4 A (b) and A (c)). However, contrast enhancement of these images, amplifying the fluorescent signal collected, showed that these compounds also localised to the nucleus in permeabilised cells but were inherently weakly fluorescent (Figure 4.4 B (b) & B (c)). Thus, the relative inherent fluorescence on UV excitation was established for the

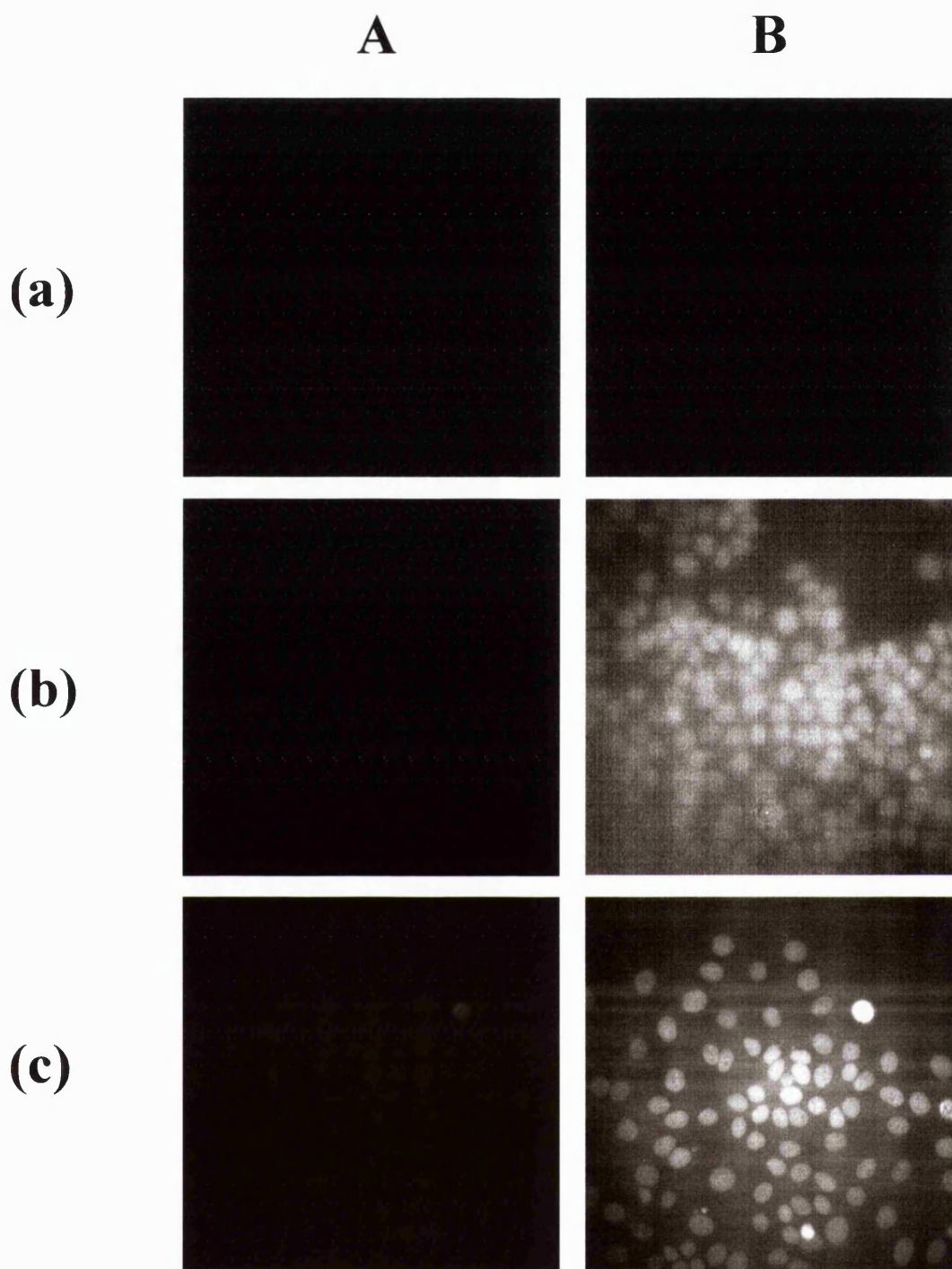


**Figure 4.2** Micrographs of fixed MCF-7 cells exposed for 1 hour to 10 $\mu$ M doses of A. Hoechst 33258 (1/4 exposure), B. TIP-C8-amino, C. TIP-C7-amino and D. TIP-amino, as viewed in the blue emission channel at a magnification of 125, showing the relative fluorescence and nuclear localisation of the compounds.



**Figure 4.3** Micrographs of fixed MCF-7 cells exposed for 1 hour to 10 $\mu$ M doses of (a) SJG-136 and (b) GWL-6, as viewed A. in the blue emission channel and B. in the blue emission channel with contrast enhancement at a magnification of 125, showing the relative fluorescence and nuclear localisation of the compounds.





**Figure 4.4** Micrographs of fixed MCF-7 cells, either (a) untreated or exposed for 1 hour to 10 $\mu$ M doses of (b) distamycin and (c) JH-37, as viewed A. in the blue emission channel and B. in the blue emission channel with contrast enhancement, at a magnification of 100 showing relative fluorescence and nuclear localisation of the compounds.

compounds examined, whereby Hoechst 33258 > > TIP-C8-amino > TIP-C7-amino > SJG-136 > GW-6 = TIP-amino > JH-37 > distamycin. No fluorescence was observed in fixed cells that had not been exposed to minor groove-binding agents, even after contrast enhancement (Figure 4.4 (a)).

#### 4.4.2 Live cell uptake and intracellular distribution of the polyamide dimer TIP-C7-amino and monomer TIP-amino in the human breast cell line MCF-7.

To examine uptake and intracellular distribution of the polyamide dimer, TIP-C7-amino, in living cells, unfixed MCF-7 cells were exposed to concentrations of up to 30 $\mu$ M of the polyamide for time periods ranging from 30 minutes to 72 hours, and were viewed in the blue channel on excitation with UV emitted from the mercury halide light source (Figure 4.5). Corresponding phase images are shown in Figure 4.6. No polyamide uptake was visible in either the cytoplasm or nuclei of living cells over 72 hours of continuous incubation, at doses of up to 3 $\mu$ M (Figure 4.5 A-C (a) & (b)). At 24 hours, however, a fine micro-precipitate was observed in the culture medium at the highest doses of 10 and 30 $\mu$ M after which a limited cytoplasmic uptake was visualised, with progressive intracellular accumulation at these doses over 72 hours (Figure 4.5 A-C (c) & (d)). Over this duration polyamide uptake was of a granular appearance and distribution was confined to the cytosol. Accumulation was intracellular as the nuclei of stained cells were clearly visible as dark shadows and were not obscured by any superficial binding to the cell surface. Nuclear localisation of TIP-C7-amino was detected by eye only in cells with compromised membrane morphology as seen under phase contrast. Furthermore, cells undergoing mitosis did not show chromosomal staining with this compound (Figure 4.7).

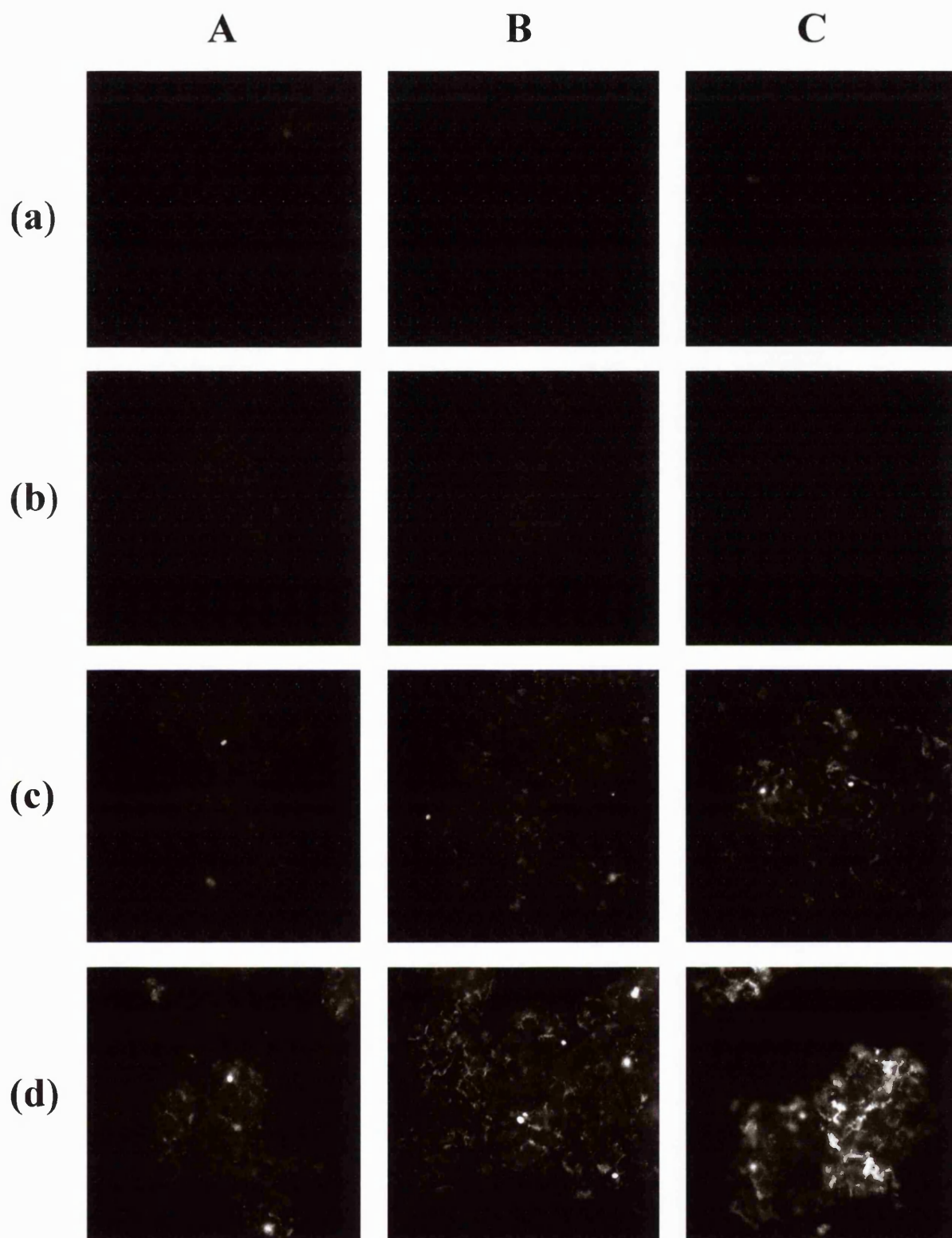
To assess the integrity of the plasma membrane of treated cells, MCF-7 cells exposed to TIP-C7-amino were counter-stained with the membrane impermeant nucleic acid dye propidium iodide, which emits at 617nm in the red portion of the visible spectrum, prior to visualisation and then viewed in both the blue and red emission channels (Figure 4.8). Under these conditions the intense nuclear staining observed in the blue channel in selected cells directly correlated with propidium iodide staining visualised in the red,

indicating that these cells were dead (Figure 4.8 B (b) & (c)). Further, this corresponded with a compromised morphology as seen under phase contrast (Figure 4.8 B (a)). Propidium iodide also stained membrane compromised cells in the untreated sample, however these did not show concomitant staining when viewed in the blue channel (Figure 4.8 A (b) & (c)). To confirm the viability of cells showing cytosolic accumulation, those treated with polyamide were counter-stained with both propidium iodide and CFDA AM, a green emitting membrane permeant dye rendered impermeant on cleavage by non-specific intracellular esterases yielding carboxyfluorescein, prior to visualisation and viewed in the red, green and blue emission channels as well as under phase contrast (Figure 4.9). As previously observed, cells showing nuclear localisation of TIP-C7-amino in the blue channel also stained with propidium iodide, as indicated by arrows in Figure 4.9 B & D. Furthermore, these cells did not stain with CFDA AM when observed in the green emission range (Figure 4.9 C). This was in contrast to cells showing cytosolic uptake of TIP-C7-amino alone, which also stained with CFDA AM. To provide a positive control for nuclear staining, live MCF-7 cells were stained with Hoechst 33258 alone for 1 hour, thereby enabling a comparison of intracellular localisation of the freely membrane permeant dye Hoechst 33258, with that of the polyamide TIP-C7-amino. These cells showed uniform nuclear uptake with no residual staining of the cytosol (Figure 4.10 B). The nuclei were clearly distinguishable from the surrounding cytosol when compared to the phase contrast image (Figure 4.10 A).

To examine whether cytosolic polyamide distribution co-localised with mitochondria in cells, MCF-7 cells were treated with TIP-C7-amino over 72 hours and counter-stained with rhodamine 123, a freely membrane permeant mitochondrial dye rendered impermeant on exposure to the intramitochondrial electrochemical gradient, prior to visualisation in the green and blue emission channels, as well as under phase contrast (Figure 4.11). Images captured indicated that the intracellular distribution of TIP-C7-amino visualised in the blue channel differed appreciably from that of rhodamine 123 detected here in the green channel (Figure 4.11 B & C). Rhodamine 123 localised to the mitochondria uniformly across cell sheets, whilst TIP-C7-amino accumulated

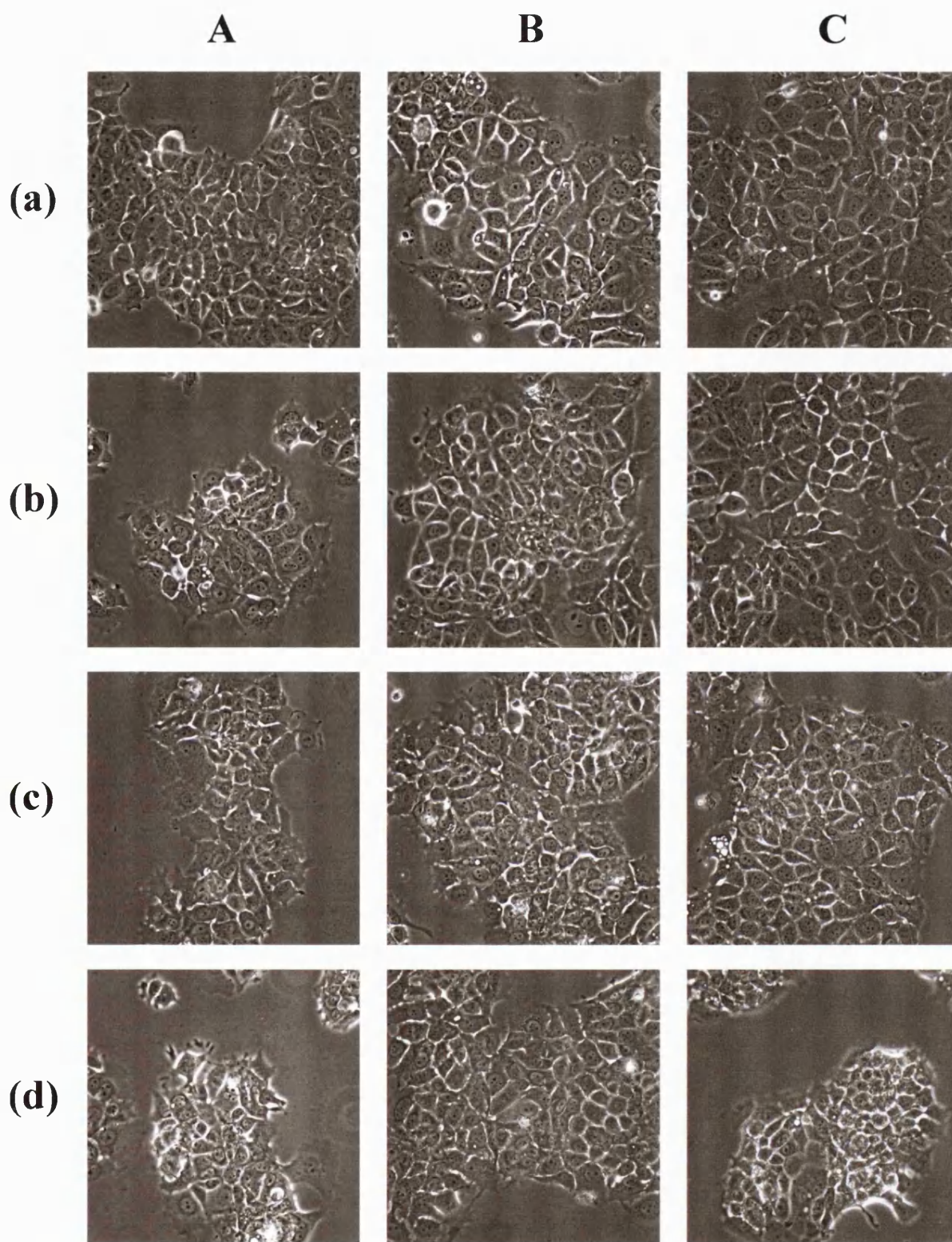
heterogeneously across the monolayer with discrete cytosolic aggregation in cells situated at the periphery of the groups of cells.

For comparison, the intracellular uptake and distribution of the polyamide monomer, TIP-amino, was also examined in living cells. MCF-7 cells were exposed to concentrations of up to 30 $\mu$ M of the polyamide monomer over 72 hours, and were viewed in the blue channel (Figure 4.12). No polyamide uptake was visible in either the cytoplasm or nuclei of living cells after 72 hours of continuous incubation, at doses of up to 30 $\mu$ M (Figure 4.12 B a-c). In contrast to the dimer, TIP-C7-amino, no micro-precipitate was observed in the culture medium. Further, no cytoplasmic uptake could be detected above the levels of inherent auto-fluorescence seen in the untreated controls (Figure 4.12 A a-c).

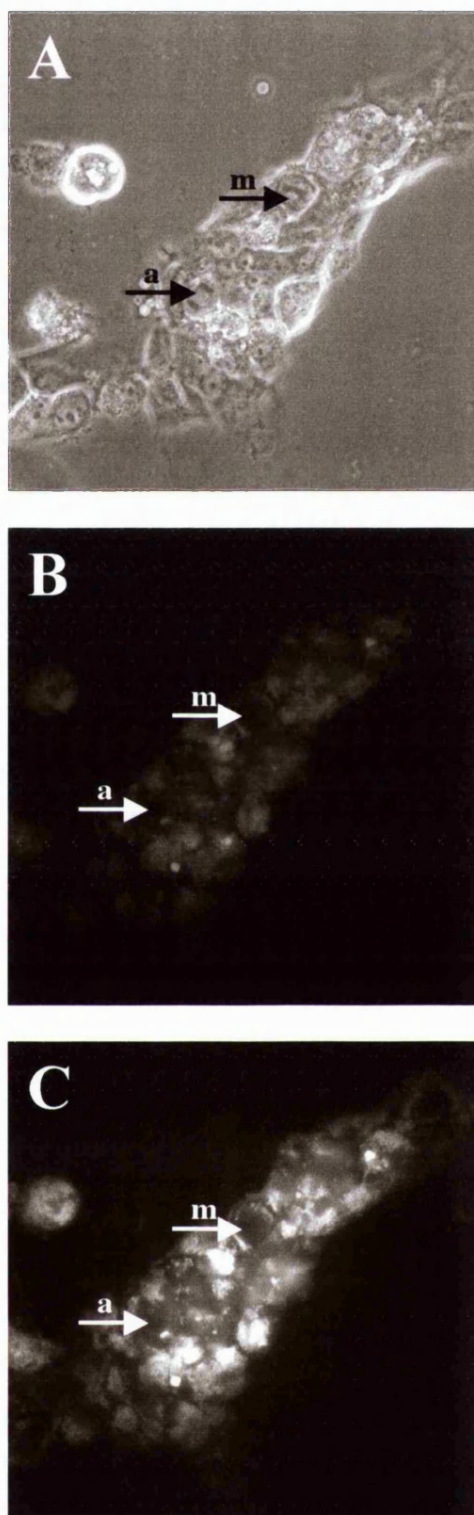


**Figure 4.5** Micrographs of live MCF-7 cells exposed to doses of (a) 0 $\mu$ M, (b) 3 $\mu$ M, (c) 10 $\mu$ M and (d) 30 $\mu$ M TIP-C7-amino over A. 24 hours, B. 48 hours and C. 72 hours as viewed in the blue emission channel at a magnification of 75, showing dose-dependent cytosolic uptake of TIP-C7-amino over time.



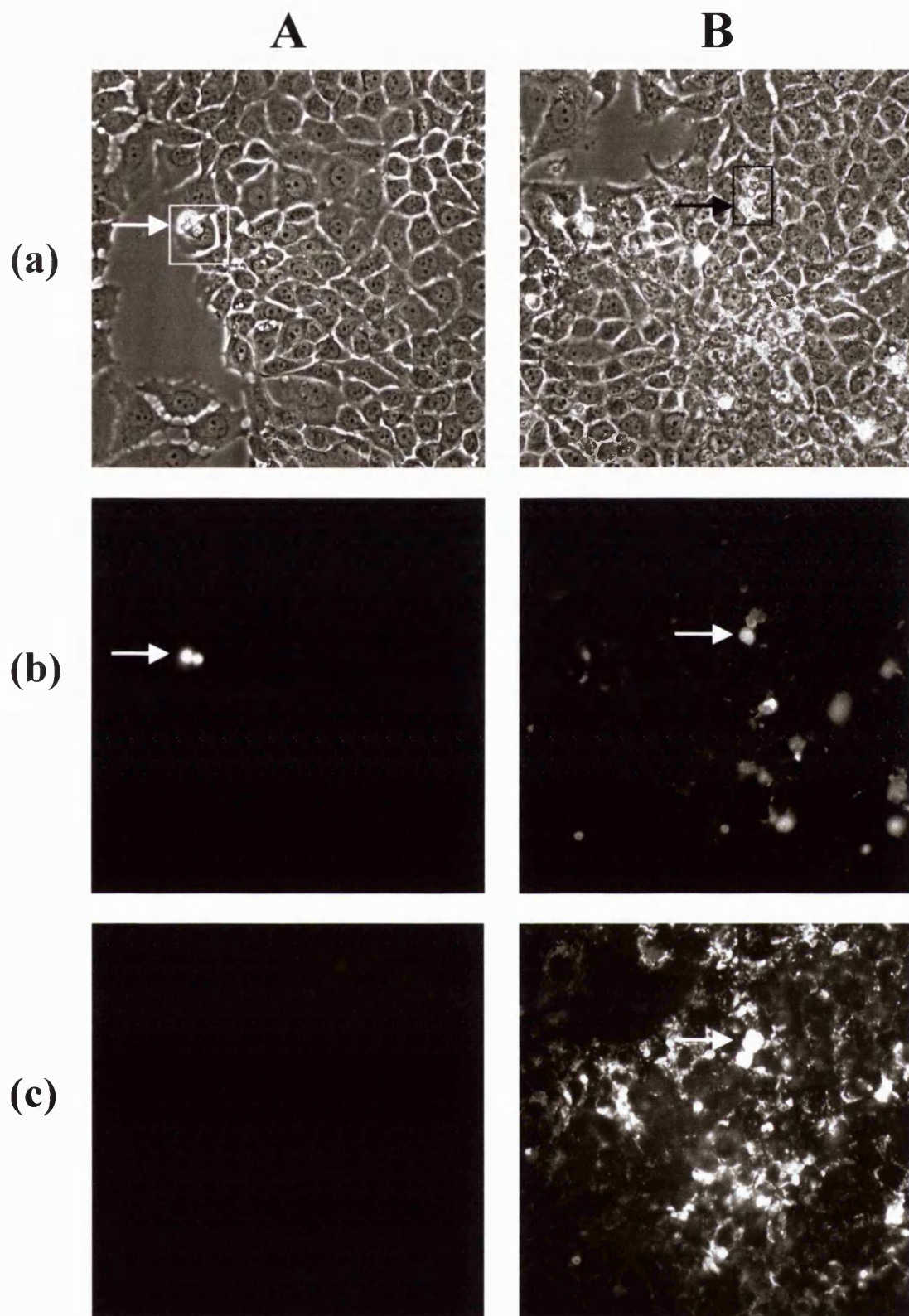


**Figure 4.6** Corresponding micrographs of live MCF-7 cells exposed to doses of (a) 0 $\mu$ M, (b) 3 $\mu$ M, (c) 10 $\mu$ M and (d) 30 $\mu$ M TIP-C7-amino over A. 24 hours, B. 48 hours and C. 72 hours as viewed under phase contrast at a magnification of 75, showing an absence of any effect on overall cell morphology.

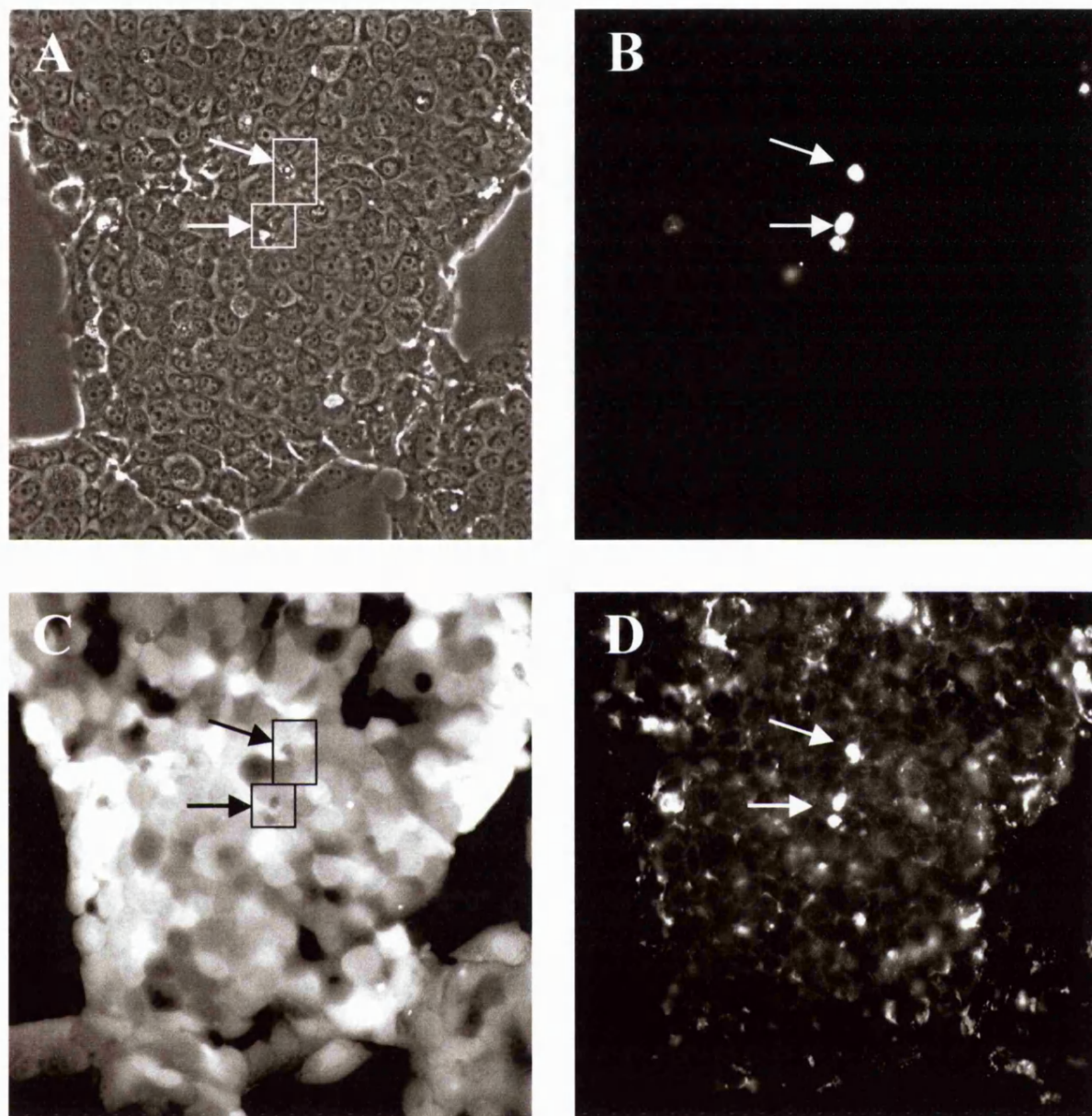


**Figure 4.7** Micrographs of live MCF-7 cells treated for 72 hours with 30 $\mu$ M TIP-C7-amino as viewed A. under phase contrast and B. in the blue emission channel and C. in the blue channel with contrast enhancement at a magnification of 150, showing an absence of chromosomal staining by TIP-C7-amino in mitotic cells. Cells at anaphase (a) and metaphase (m) are indicated with arrows.



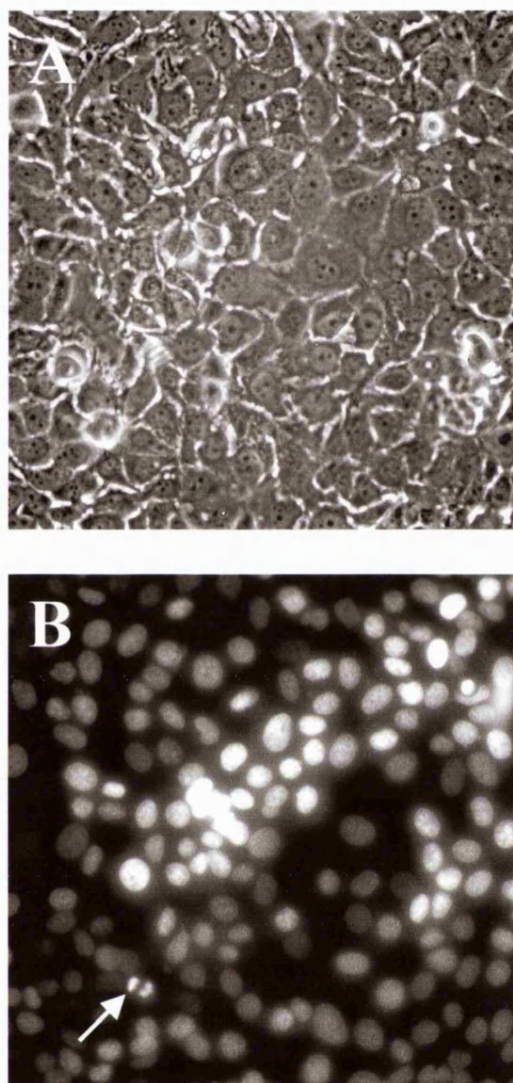


**Figure 4.8** Micrographs of live MCF-7 cells exposed to A. 0 $\mu$ M and B. 30 $\mu$ M TIP-C7-amino for 72 hours, counter-stained with 0.01 mgml<sup>-1</sup> of propidium iodide, viewed (a) under phase contrast, (b) in the red emission channel and (c) in the blue emission channel at a magnification of 110, showing that uptake in live cells is confined to the cytoplasm. Arrows indicate dead cells.

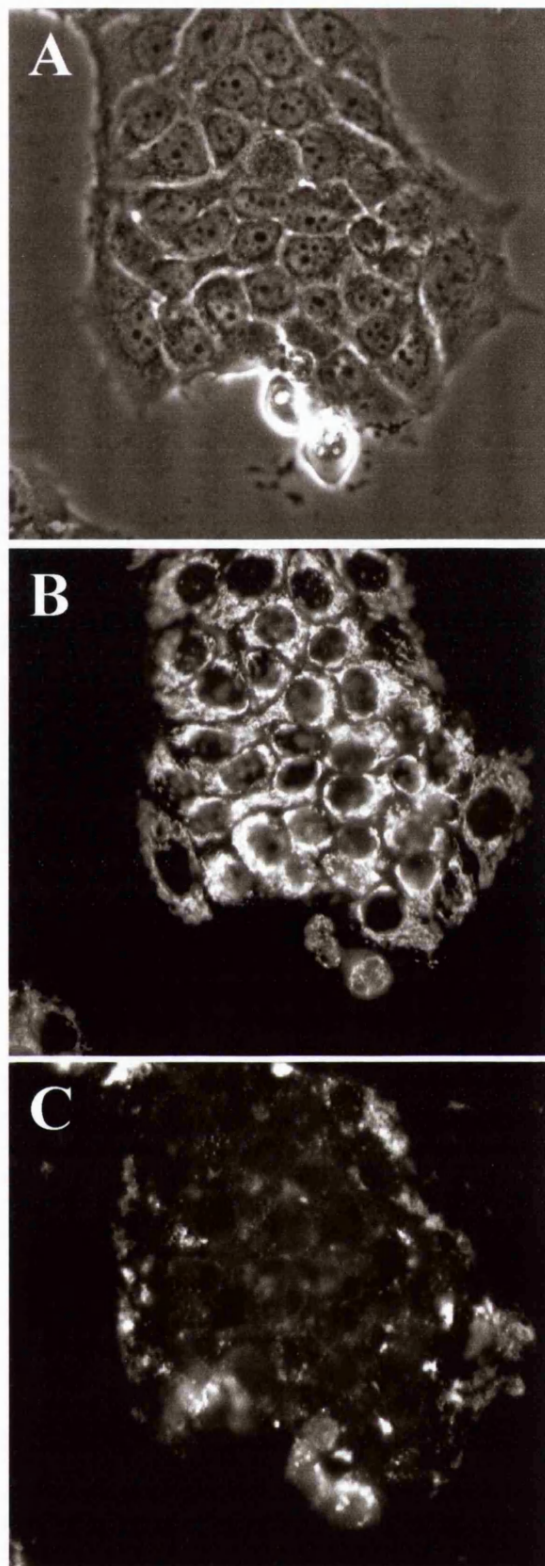


**Figure 4.9** Micrographs of live MCF-7 cells exposed to 30 $\mu$ M TIP-C7-amino for 48h, counter-stained with 0.01mgml<sup>-1</sup> propidium iodide and 1 $\mu$ M CFDA AM, viewed A. under phase contrast, B. in the red emission channel, C. in the green emission channel and D. in the blue emission channel at a magnification of 110, showing that nuclear uptake is limited to non-viable cells. Dead cells are indicated with arrows.

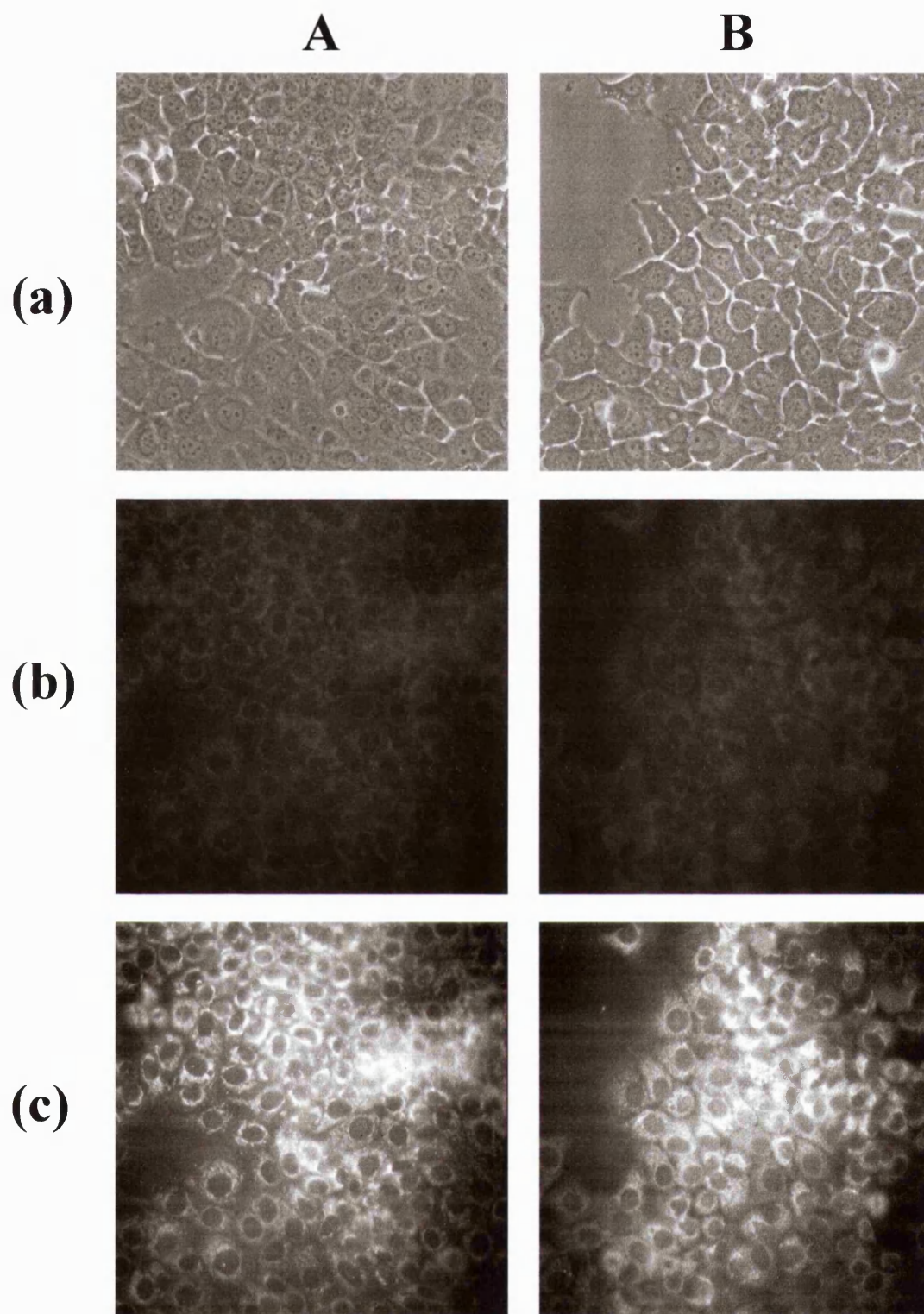




**Figure 4.10** Micrographs of live MCF-7 cells incubated for 1 hour with 10 $\mu$ M Hoechst 33258 as viewed A. under phase contrast and B. in the blue emission channel at a magnification of 125, showing the nuclear localisation of freely membrane permeant Hoechst 33258 in living cells. A mitotic cell at anaphase is indicated with an arrow.



**Figure 4.11** Micrographs of live MCF-7 cells exposed to 30 $\mu$ M TIP-C7-amino for 48 hours, counter-stained with 1 $\mu$ M rhodamine 123, viewed A. under phase contrast, B. in the green emission channel and C. in the blue emission channel at a magnification of 250, showing that uptake of TIP-C7-amino does not co-localise with mitochondria.



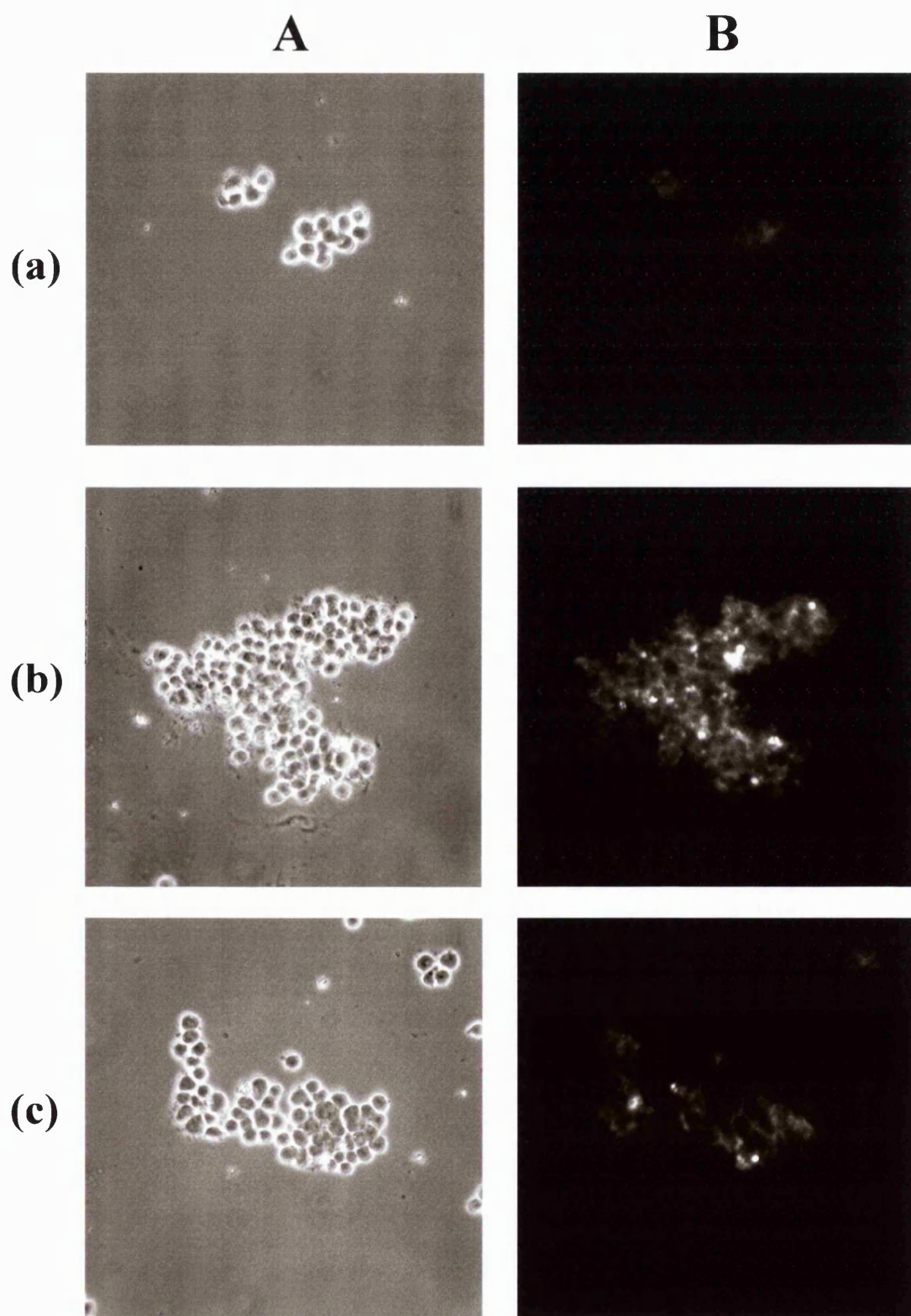
**Figure 4.12** Micrographs of live MCF-7 cells A. untreated and B. treated for 72 hours with 30 $\mu$ M TIP-amino monomer as viewed (a) under phase contrast (b) in the blue emission channel and (c) in the blue emission channel with contrast enhancement at a magnification of 125, showing that no nuclear localisation is visible in living cells, although detectable nuclear staining was seen with TIP-C7-amino in fixed cells (Fig. 4.2).

#### 4.4.3 Live cell uptake and intracellular distribution of the polyamide dimer TIP-C7-amino in the human T cell line Jurkat.

To assess polyamide uptake in a T cell line, live Jurkat cells were exposed to concentrations of up to 30 $\mu$ M of TIP-C7-amino over 72 hours and were viewed in the blue channel as well as under phase contrast. As with the MCF-7 cells, polyamide uptake was not detected either in the cytoplasm or nuclei of living cells for up to 24 hours of continuous incubation, at this dose range. Once more, a fine micro-precipitate began to appear in the culture medium between 24 and 48 hours at the highest doses of 10 and 30 $\mu$ M, after which progressive cell associated accumulation was observed at these doses between 48 to 72 hours (Figure 4.13). This was of a granular appearance and did not localise to the nuclei of living cells. Similar to the polyamide uptake observed in MCF-7 cells, intense staining was seen in cells exhibiting compromised membrane structure under phase contrast. Counterstaining with viable cell specific dyes, including CFDA AM and acridine orange, a dual-fluorescence nucleic acid stain emitting in the green (525nm) when DNA-associated and in the red (650nm) when RNA-associated, confirmed that these cells were non-viable (Figure 4.14). Furthermore, a corresponding anti-proliferative effect was observed qualitatively in this cell type at 10 and 30 $\mu$ M over 48-72 hours.

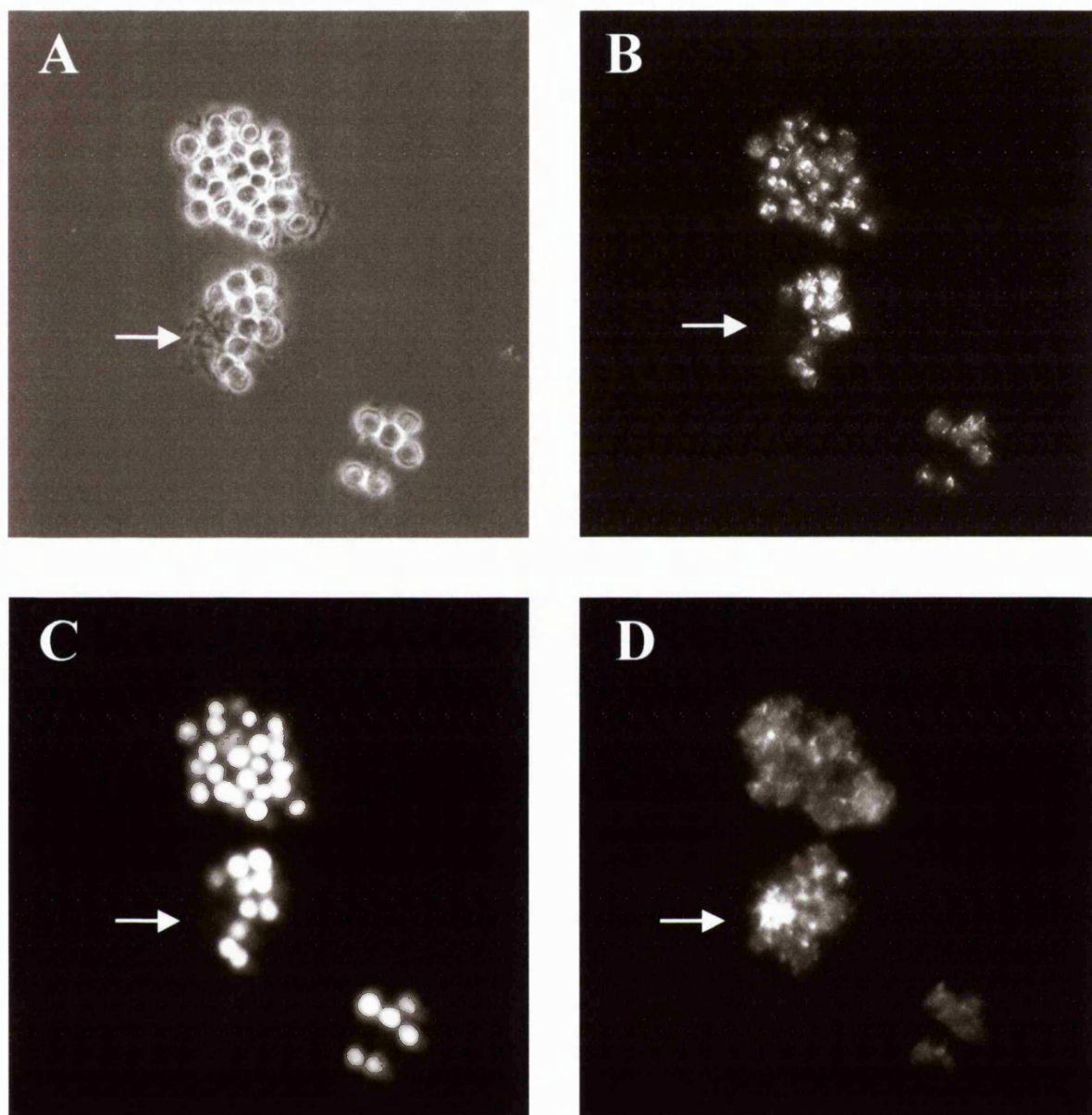
Live Jurkat cells were also stained with Hoechst 33258, plus CFDA AM and propidium iodide for 1 hour and viewed under phase contrast and in the appropriate emission channels (Figure 4.15). As with the MCF-7 cells, this provided a positive control for nuclear staining, allowing comparison of the localisation of the freely membrane permeant dye Hoechst 33258 in living cells, with that of TIP-C7-amino. Hoechst staining of Jurkat cells, however, demonstrated that discrimination of the nucleus from the cytosol was less straightforward at this magnification, as the nucleus encompassed a much larger cell volume compared to that in the MCF-7 cells (Figure 4.10). Corresponding images of Jurkat cells captured in the red and green channels indicated that whilst the freely permeable Hoechst indiscriminately stained both live and dead cells, cell viability could be clearly distinguished using CFDA AM and propidium iodide as previous (Figure 4.15 (b) C & D).



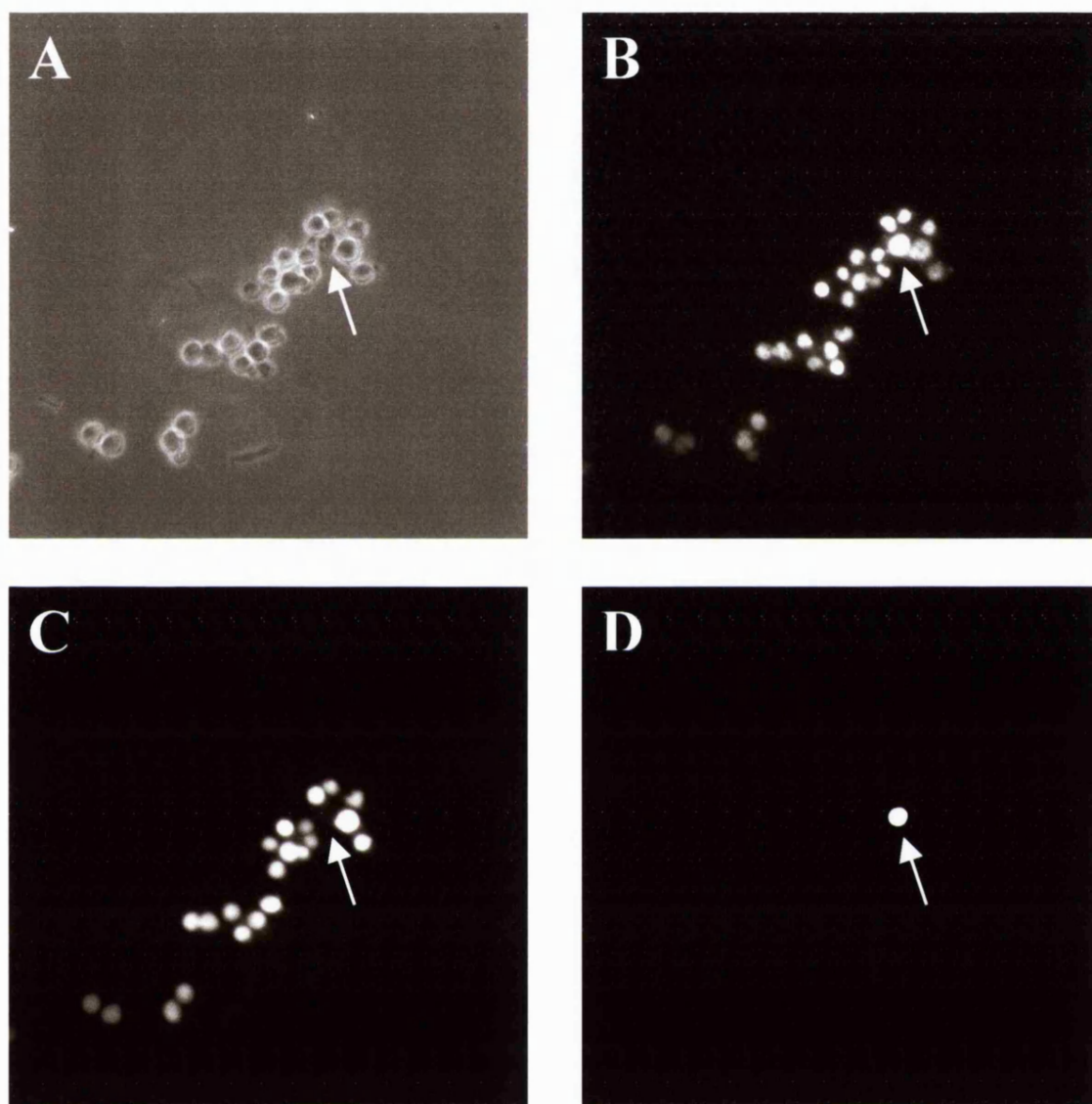


**Figure 4.13** Micrographs of live Jurkat cells exposed to 30 $\mu$ M TIP-C7-amino over (a) 24 hours, (b) 48 hours and (c) 72 hours, as viewed A. under phase contrast and B. in the blue emission channel at a magnification of 180, showing cell associated TIP-C7-amino uptake over time.





**Figure 4.14** Micrographs of live Jurkat cells exposed to 30 $\mu$ M TIP-C7-amino for 72 hours, counter-stained with 1 $\mu$ M acridine orange and 1 $\mu$ M CFDA AM, viewed A. under phase contrast, B. in the red emission channel, C. in the green emission channel and D. in the blue emission channel at a magnification of 220, showing that uptake does not co-localise with cellular nucleic acids and that nuclear staining is confined to non-viable cells. Dead cells are indicated with arrows.



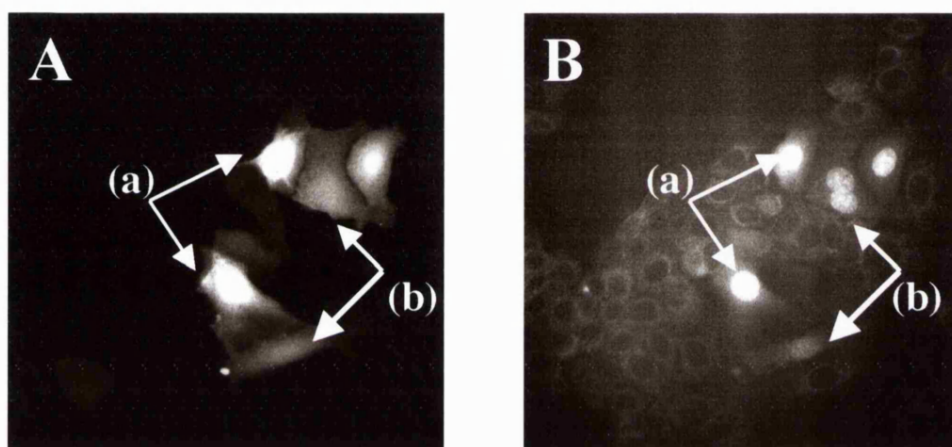
**Figure 4.15** Micrographs of live Jurkat cells stained with 10 $\mu$ M Hoechst 33258, plus 1 $\mu$ M CFDA AM and 0.01mgml<sup>-1</sup> propidium iodide, viewed A. under phase contrast, B. in the blue emission channel, C. in the green emission channel and D. in the red emission channel at a magnification of 180, showing nuclear localisation of Hoechst 33258 in both viable and non-viable cells. A dead cell is indicated with an arrow.

#### 4.4.4 Microinjection of live MCF-7 cells with TIP-C7-amino

TIP-C7-amino was directly injected into the cytoplasm of live MCF-7 cells to ascertain whether a polyamide once present in the cell cytosol at high concentrations could traverse the nuclear membrane and localise in the nuclei of living cells (Figure 4.16). The green-yellow excited (557nm) /orange-red emitting (576nm) fluorochrome marker TRITC- dextran was co-injected to identify injected cells, both under phase contrast where it appeared pink and in the red emission channel. Injection of TIP-C7-amino was not straightforward as at high concentration on contact with the culture medium at the cell surface it began to precipitate, progressively blocking the bore of the needle capillary.

Microinjected cells were viewed under phase contrast (data not shown) and in the red and blue emission channels (Figure 4.16 A and B). Successfully injected cells exhibited pink cytoplasmic staining when viewed under phase contrast due to the presence of the TRITC-dextran (data not shown). Injected cells were more clearly identified using TRITC-dextran in the red channel, as shown in Figure 4.16 A. In some cases, however, the TRITC-dextran appeared to be localised in the nucleus as well as in the cytoplasm of some cells, as indicated by arrows (a) in Figure 4.16 A. This demonstrated that in these cells the nucleus as well as the cytoplasm had been directly microinjected, as TRITC-dextran is normally excluded from the nucleus when the cytoplasm only is injected. Furthermore, when viewed in the blue channel with contrast enhancement these cells also showed strongly fluorescent nuclei (indicated by arrows (a) in Figure 4.16 B), which directly correlated with the nuclear localisation observed in the red channel (Figure 4.16 A). These data show that in a proportion of the injected cells TIP-C7-amino has been directly introduced into the nucleus. It cannot be discounted that in such cells at least part of the observed signal observed in the blue may be due to the presence of a high nuclear concentration of TRITC-dextran. However, in some cells cytoplasmic staining only was predominantly observed in the red channel, demonstrating that cytoplasmic microinjection only had occurred, as indicated by white arrows in Figure 4.16 A. In these cells a weak but detectable nuclear staining was seen in the blue channel, indicating that TIP-C7-amino may have crossed into the nucleus. The

possibility that nuclear uptake was as a result of the trauma of injection cannot be ruled out. However, whatever the mechanism for nuclear uptake, this is an important observation as it demonstrates that freely soluble TIP-C7-amino will stain the nuclei of living cells. Weak cytoplasmic auto-fluorescence was observed in cells that had not been microinjected with the polyamide/TRITC-dextran mix, when viewed in the blue channel with contrast enhancement, as shown in Figure 4.16 B.



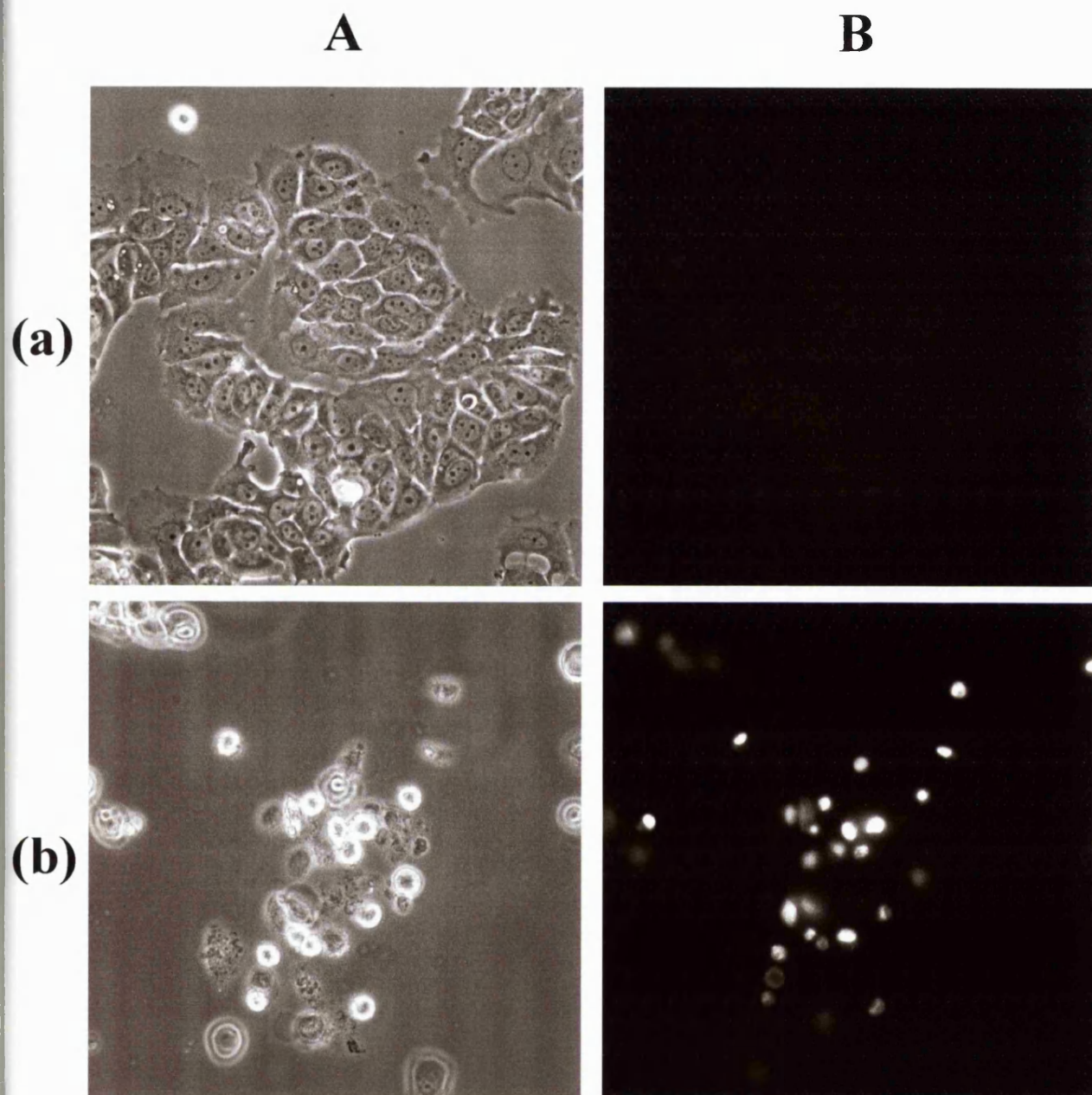
**Figure 4.16** Micrographs of live MCF-7 cells microinjected with a TIP-C7-amino/TRITC-dextran mix to yield final concentrations of 0.3mg/ml polyamide and 1mg/ml fluorochrome marker in phosphate buffered saline, as viewed A. in the red emission channel and B. in the blue emission channel with contrast enhancement at a magnification of 200, showing nuclear staining by TIP-C7-amino. Nuclear staining, due to direct injection of the TIP-C7-amino/TRITC dextran mix into the nucleus is indicated by arrows (a), whilst nuclear uptake following cytoplasmic staining only is indicated by

#### 4.4.5 Live cell uptake and intracellular distribution of the PBD dimer SJG-136 and the PBD-polyamide conjugate GWL-6 in the breast cell line MCF-7.

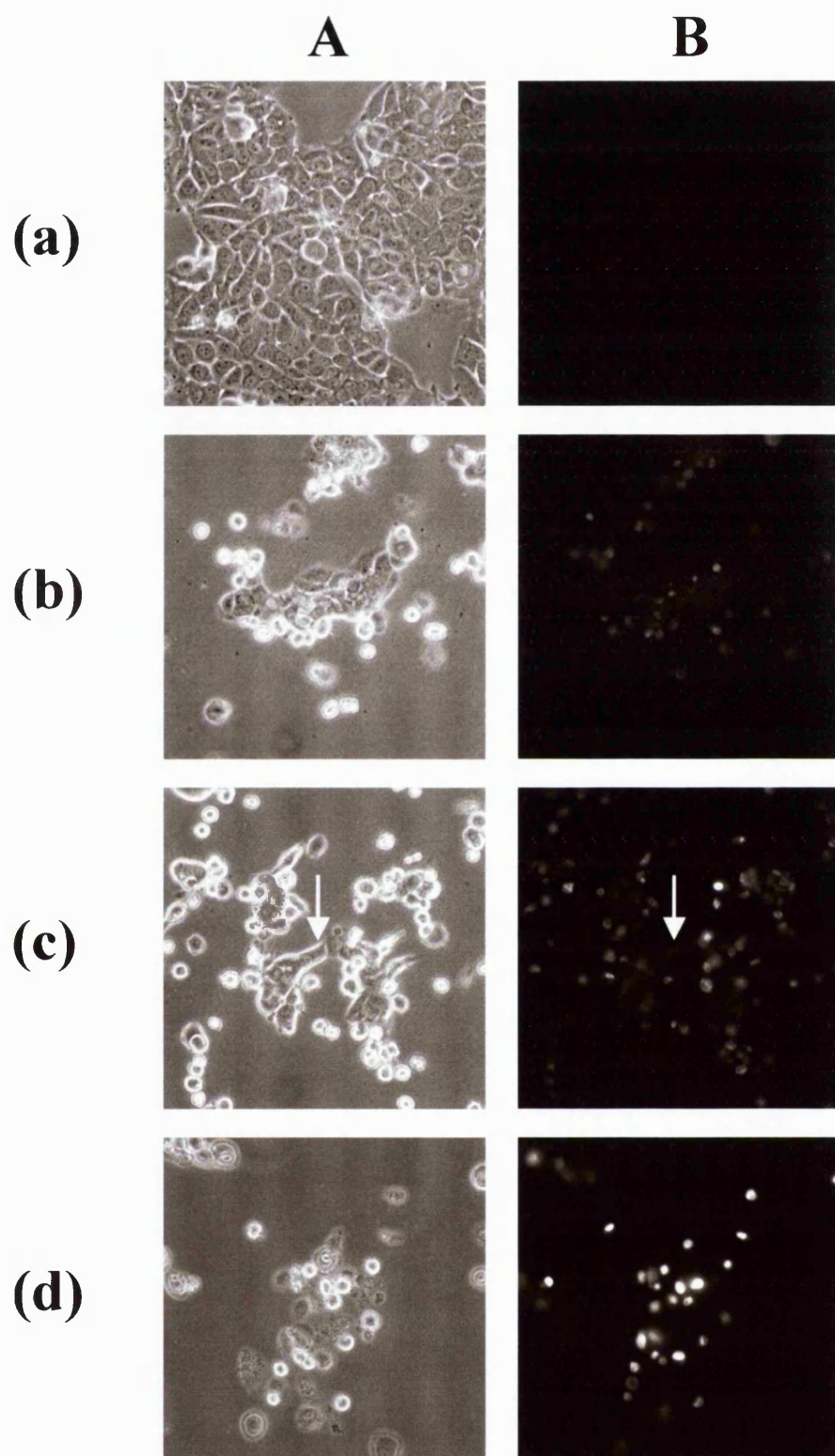
Intracellular uptake of TIP-C7-amino in unfixed cells was then compared to that of the PBD dimer, SJG-136, a known cytotoxic agent causing cross-links in cells (Gregson, *et al.*, 2001) and the novel PBD-polyamide conjugate, GWL-6, also cytotoxic and shown to cause DNA monoadducts *in vitro* (personal communication from Professor D. Thurston). MCF-7 cells were treated with a range of concentrations of SJG-136 and GWL-6 over time periods ranging from 10 minutes to 24 hours and then viewed in the blue channel as well as under phase contrast. Cells incubated with 10 $\mu$ M SJG-136 for 1 hour showed no detectable cellular uptake (Figure 4.17). Prolonged incubation with SJG-136 over 24 hours produced a potent cytotoxic effect at that dose and a concomitant staining of the dying cells when viewed in the blue channel. Corresponding phase contrast images for both time points clearly establish the differing cell morphologies after treatment. Cells incubated for 24 hours with a range of doses of SJG-136 showed dose dependent intracellular compound accumulation as well as an associated cytotoxicity (Figure 4.18). Localisation of this uptake was not possible due to the gross changes in cellular morphology and membrane damage incurred on treatment with SJG-136 by the time accumulation could be detected with this method.

Similarly, prolonged incubation of MCF-7 cells for 24 to 72 hours with a range of concentrations of the PBD-polyamide conjugate, GWL-6, produced an appreciable dose dependent cytotoxicity, shown clearly under phase contrast (Figure 4.19). Once more this was associated with weak staining when viewed in the blue channel, which was barely visible at the highest doses of 10 and 30 $\mu$ M in dying cells (Figure 4.20). Shorter exposures to a 30 $\mu$ M dose, ranging from 10 minutes to 2 hours, failed to detect intracellular compound accumulation in living cells (Figure 4.21). Contrast enhancement, amplifying the intensity of the fluorescence signal captured, did not demonstrate a noticeable increase in emitted signal above the inherent auto-fluorescence seen in untreated cells (Figures 4.20 and 4.21).



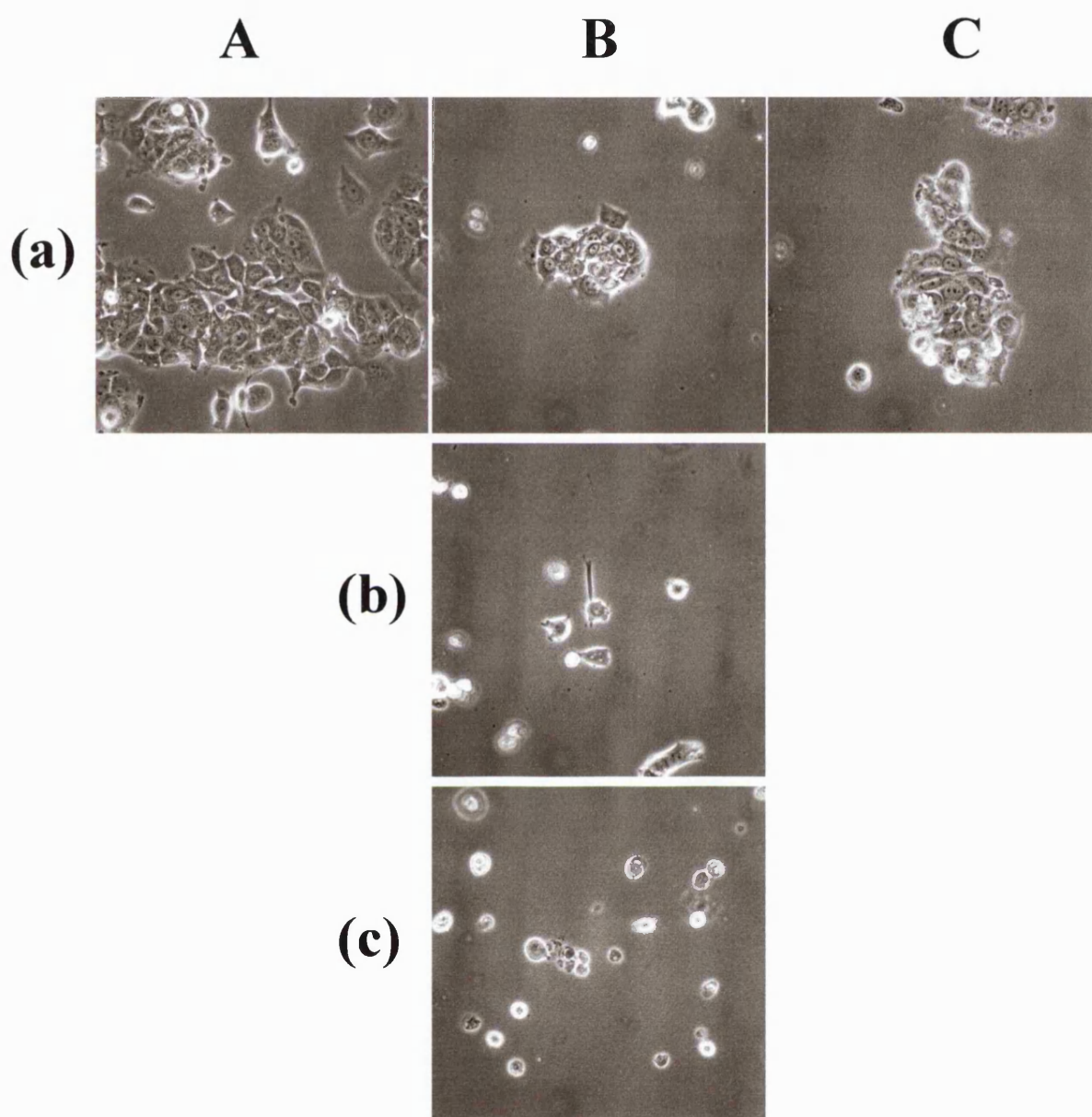


**Figure 4.17** Micrographs of unfixed MCF-7 cells exposed to 10 $\mu$ M SJG-136 for (a) 1 hour and (b) 24 hours, as viewed A. under phase contrast and B. in the blue emission channel at a magnification of 150, showing strong localised staining in dead cells only.

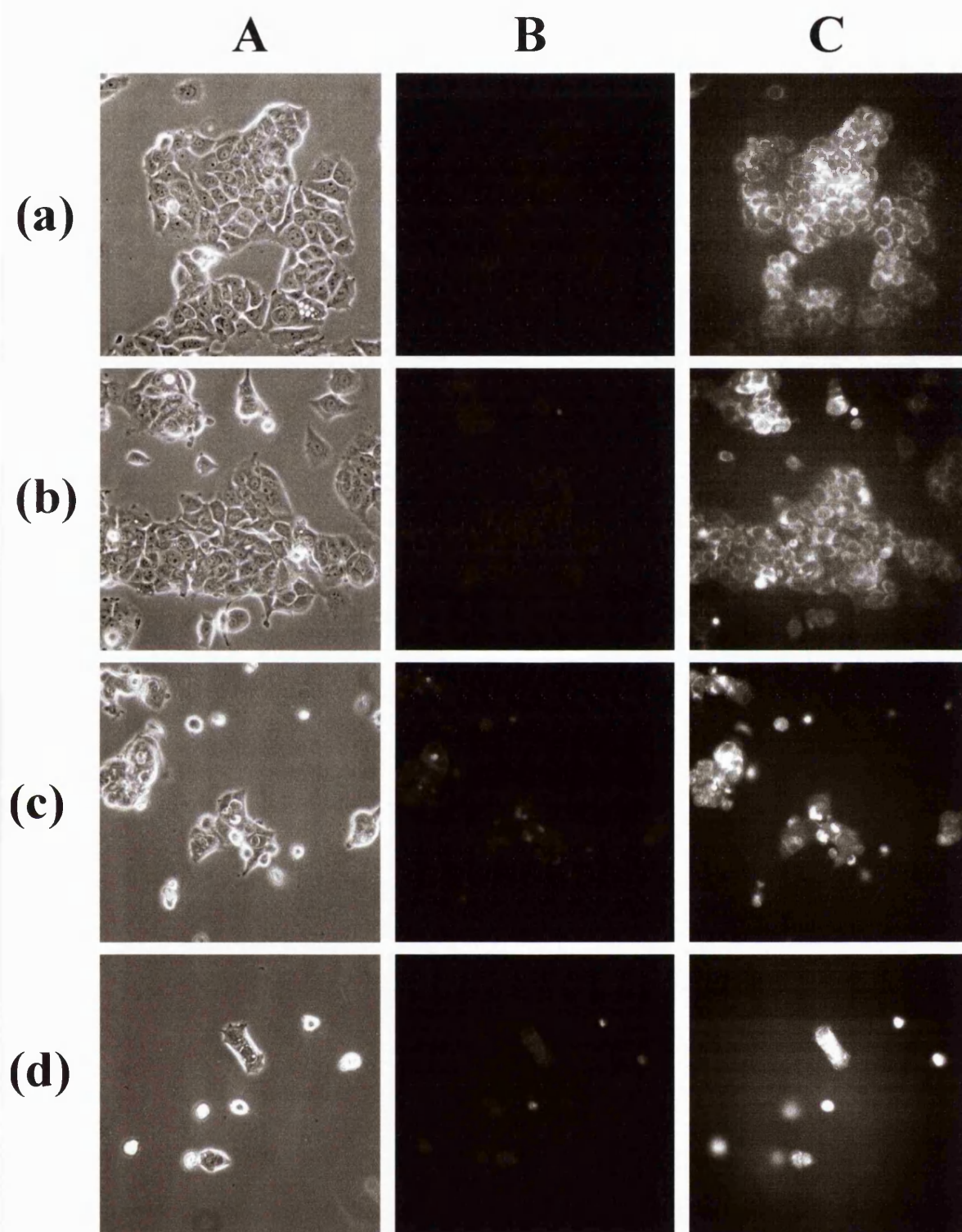


**Figure 4.18** Micrographs of unfixed MCF-7 cells incubated for 24 hours with (a) 0 $\mu$ M, (b) 1 $\mu$ M, (c) 3 $\mu$ M and (d) 10 $\mu$ M doses of the PBD dimer SJG-136, as viewed A. under phase contrast and B. in the blue emission channel at a magnification of 75, showing that dose-dependent localised staining is confined to dead or dying cells. Viable regions remain unstained as indicated by arrows.

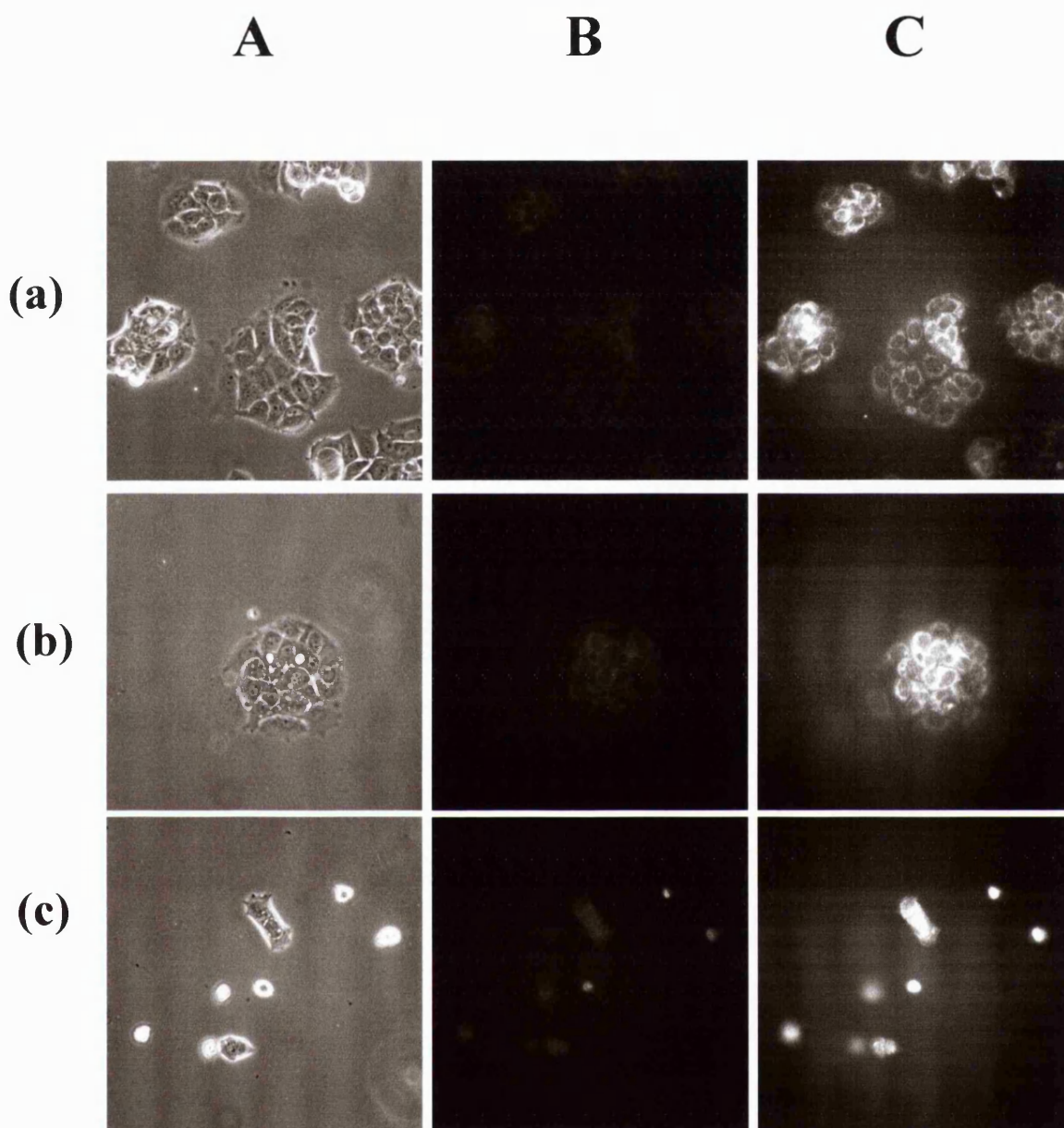




**Figure 4.19** Micrographs of unfixed MCF-7 cells incubated with (a) 1 $\mu$ M, (b) 3 $\mu$ M and (c) 10 $\mu$ M doses of the PBD-polyamide conjugate GWL-6 for A. 24 hours, B. 48 hours and C. 72 hours, as viewed under phase contrast at a magnification of 75, showing the gross changes in cellular morphology incurred on treatment.



**Figure 4.20** Micrographs of unfixed MCF-7 cells exposed for 24 hours to (a) 0 $\mu$ M, (b) 1 $\mu$ M, (c) 10 $\mu$ M and (d) 30 $\mu$ M doses of the PBD-polyamide conjugate GWL-6, as viewed A. under phase contrast, B. in the blue emission channel and C. in the blue emission channel with contrast enhancement at a magnification of 75, showing dose-dependent trace staining in dead or dying cells only.



**Figure 4.21** Micrographs of unfixed MCF-7 cells exposed to 30 $\mu$ M GWL-6 for (a) 10mins, (b) 2 hours and (c) 24 hours, as viewed A. under phase contrast, B. in the blue emission channel, C. in the blue emission channel with contrast enhancement at a magnification of 75, showing time-dependent trace staining in dead or dying cells only.

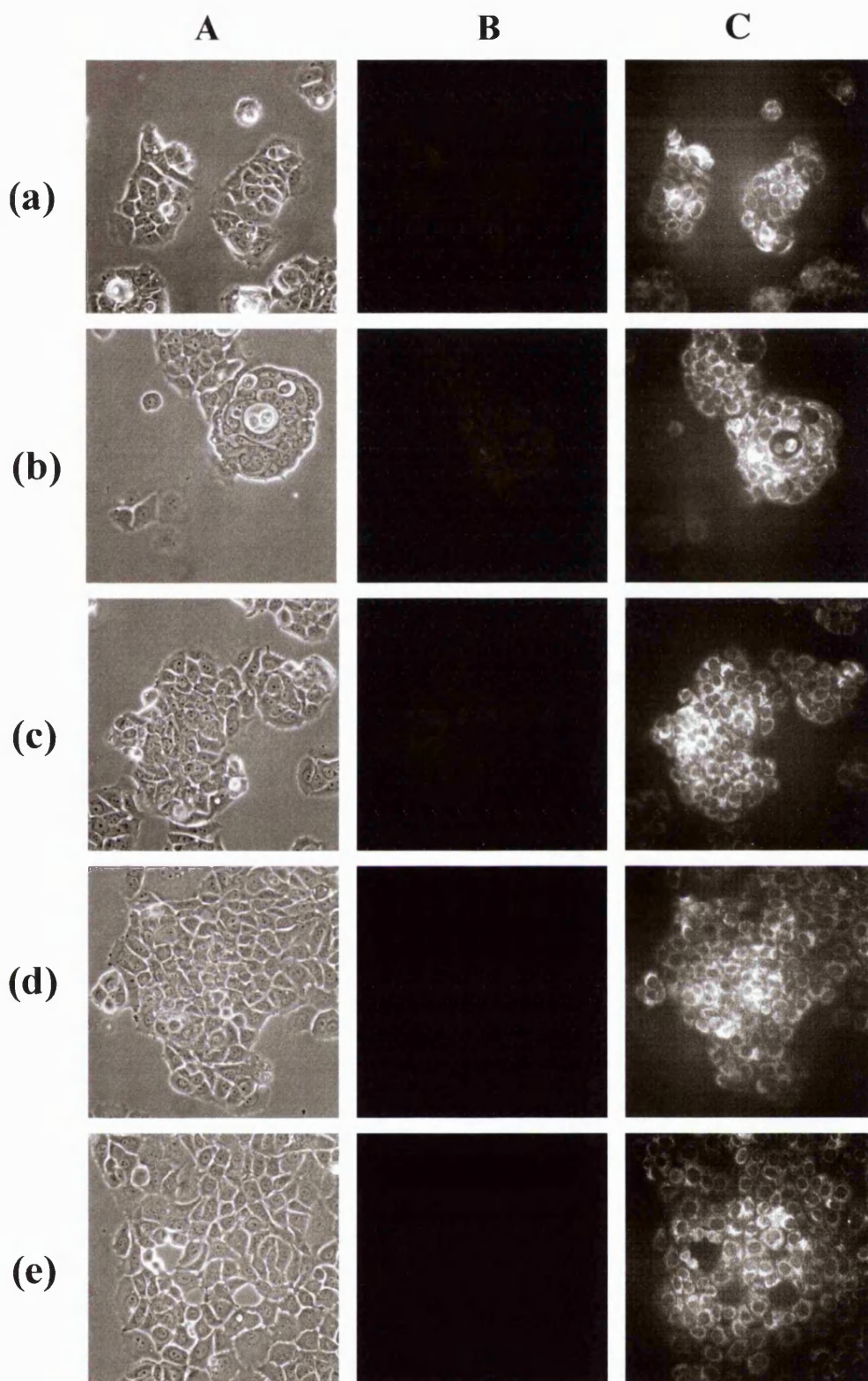
#### 4.4.6 Live cell uptake and intracellular distribution of the hairpin polyamide dimer JH-37 in the breast line MCF-7.

Live cellular uptake of the hairpin polyamide dimer, JH-37, was also evaluated in MCF-7 cells using fluorescence microscopy as above. Cells incubated with a dose of 30 $\mu$ M JH-37 for varying time durations ranging from 10 minutes to 72 hours, did not show any detectable intracellular accumulation of compound, when viewed in the blue emission range (Figure 4.22). Contrast enhancement of these images did not show an appreciable increase in the signal above the inherent cellular auto-fluorescence. Unlike the PBD dimer and PBD-polyamide conjugate, JH-37 did not produce a cytotoxic or anti-proliferative effect in these cells at doses of up to 30 $\mu$ M.

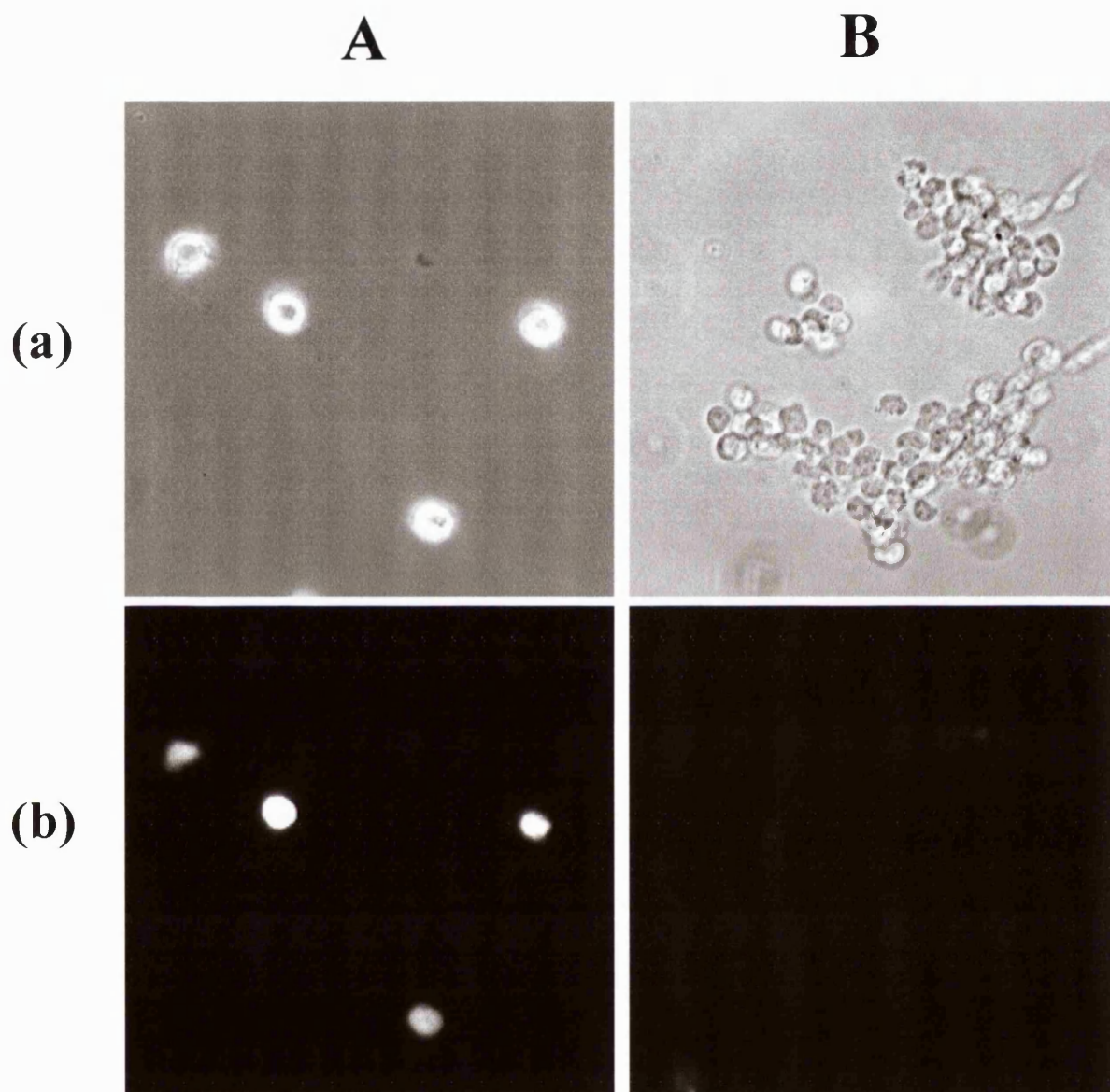
#### 4.4.7 Uptake in isolated cell nuclei.

To determine whether the nuclear envelope was acting as a barrier to nuclear localisation, uptake of the polyamides TIP-amino, TIP-C7-amino and JH-37 as well as the PBD-polyamide conjugate GWL-6 was evaluated in isolated Jurkat nuclei using fluorescence microscopy. Successful isolation of nuclei from whole cells was established by incubation of both Jurkat cells and the nuclei prep with 10 $\mu$ M CDFA AM for 20 minutes at 37°C (Figure 4.23). Although the cells and nuclei were difficult to differentiate morphologically under phase contrast or bright field (Figure 4.23 (a)), when viewed in the green channel the whole cells stained brightly whereas the nuclei did not (Figure 4.23 (b)). As CDFA AM is converted to its fluorescent state by intracellular esterases located in the cell cytoplasm, this indicated that the nuclei had lost both the cytoplasm and the surrounding plasma membrane. However, the isolated nuclei stained brightly with Hoechst 33258 and propidium iodide when viewed in the blue and red channels respectively (Figure 4.24), thereby demonstrating that they were intact and retained their DNA content. Incubation of the nuclei for 10 minutes with equimolar doses of TIP-C7-amino, TIP-amino, GWL-6 and JH-37 showed that the nuclei stained rapidly with all of the compounds examined when viewed in the blue emission channel (Figure 4.25). Contrast enhancement was required to visualise this



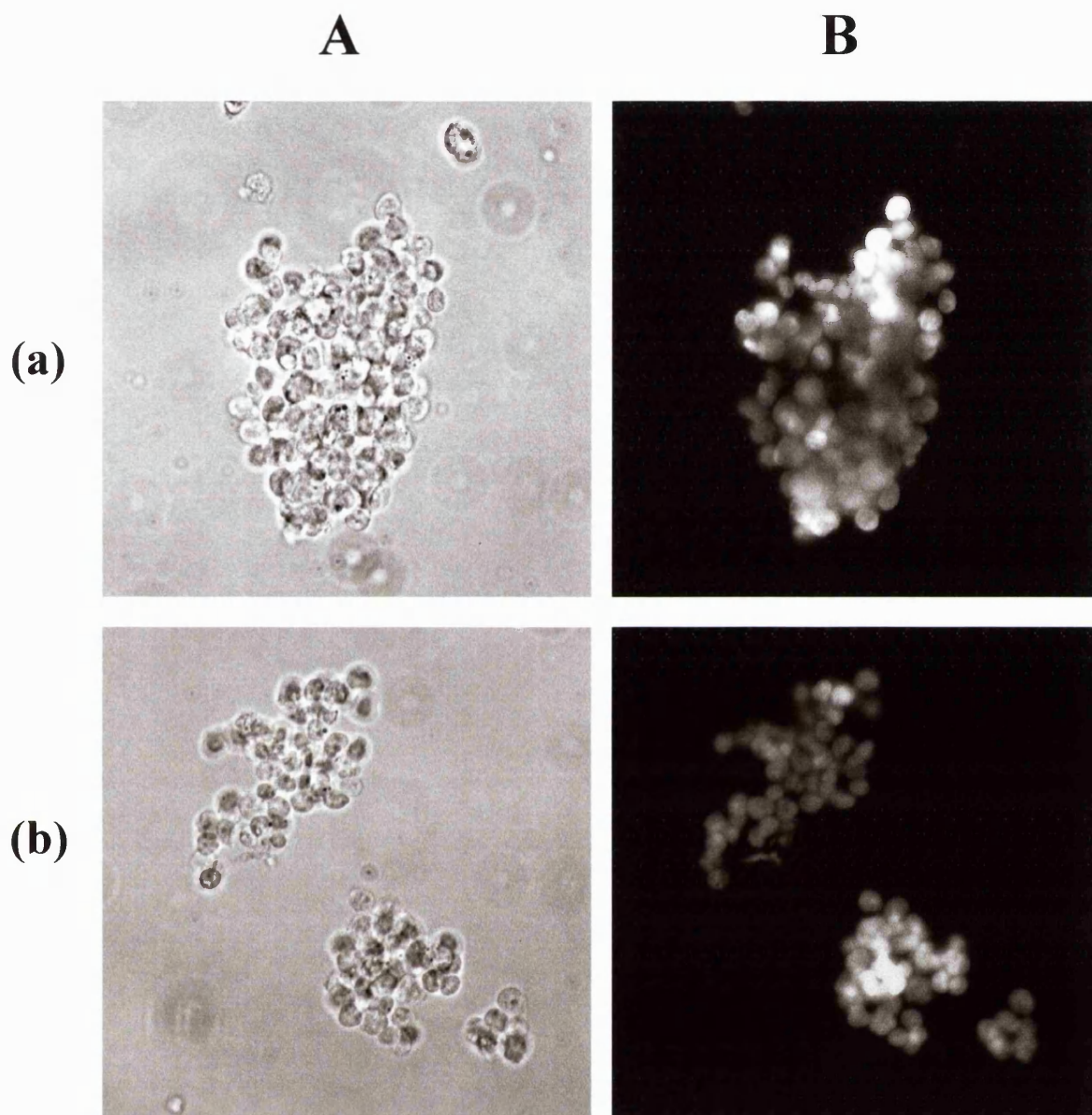


**Figure 4.22** Micrographs of live MCF-7 cells exposed to 30 $\mu$ M JH-37 for (a) 10min, (b) 2 hours, (c) 24 hours, (d) 48 hours and (e) 72 hours, as viewed A. under phase contrast, B. in the blue channel and C. in the blue channel with contrast enhancement at a magnification of 66, showing no localisation visible in living cells, although detectable nuclear staining was seen with JH-37 in fixed cells (Fig. 4.5).



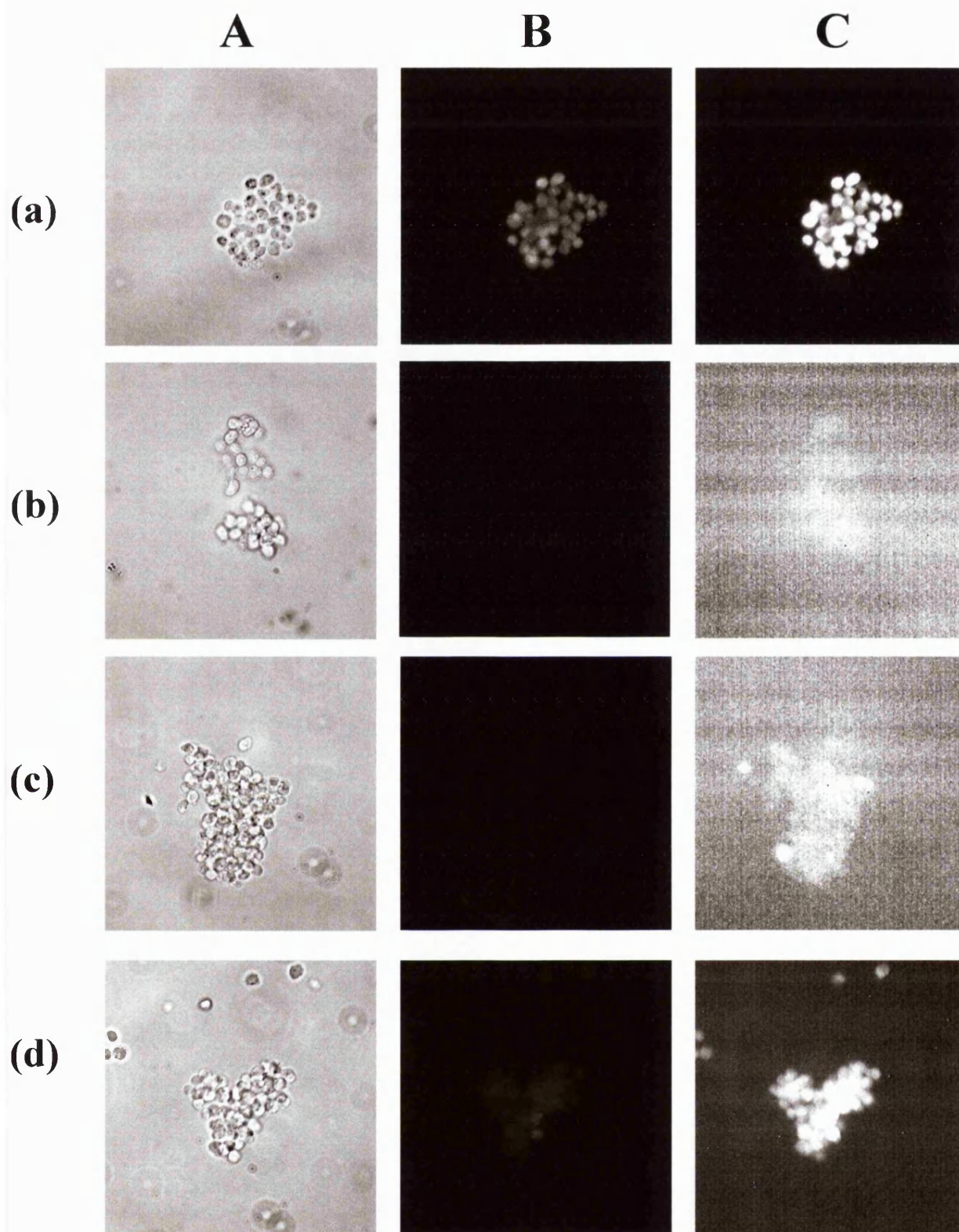
**Figure 4.23** Micrographs of A. Jurkat cells and B. Jurkat nuclei incubated with 10 $\mu$ M CDFA AM for 20mins as viewed under (a) phase contrast, (b) in the green channel at a magnification of 330, showing that the esterase activity in the cytoplasm of intact cells is not present in isolated nuclei.





**Figure 4.24** Micrographs of isolated Jurkat nuclei stained for 10 min with 10 $\mu$ M (a) Hoechst 33258 and (b) propidium iodide as viewed under A. bright field and B. in the blue channel for Hoechst or the red channel for propidium iodide at a magnification of 330, showing staining of the nuclear DNA.





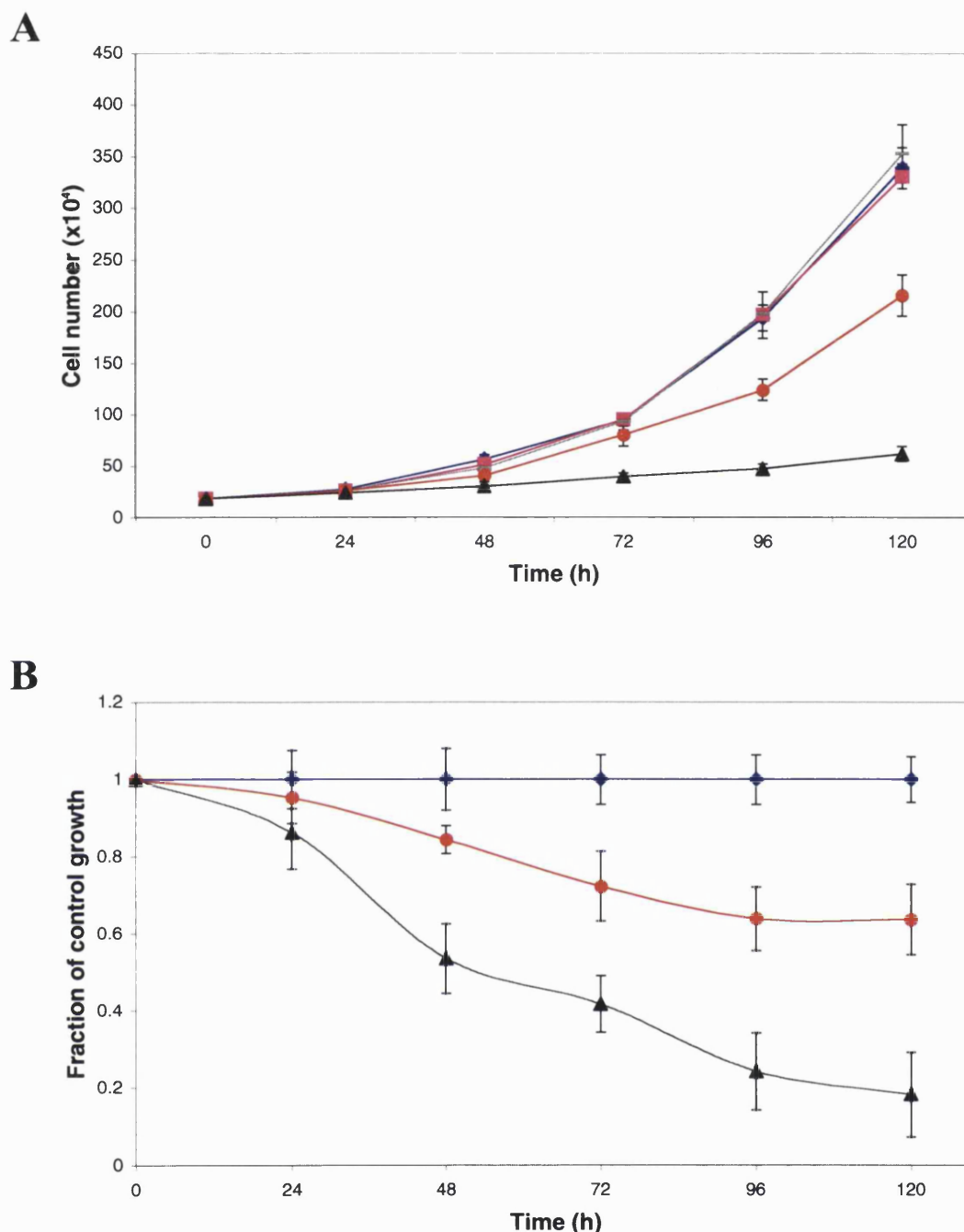
**Figure 4.25** Micrographs of isolated Jurkat nuclei stained for 10min with 10 $\mu$ M (a) TIP-C7-amino, (b) TIP-amino, (c) GWL-6 and (d) JH-37 as viewed A. under bright field, B. in the blue channel and C. in the blue channel with contrast enhancement at a magnification of 180, showing free nuclear uptake.

staining for TIP-amino, GWL-6 and JH-37 due to the reduced fluorescence of these compounds (Figure 4.25 C (b)-(d)).

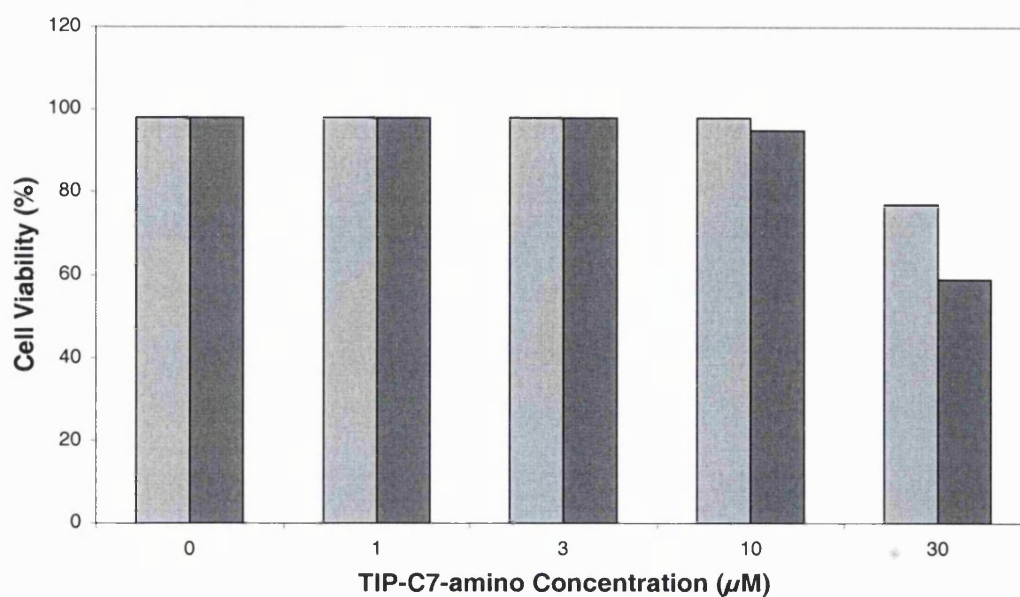
#### 4.4.8 Anti-proliferative activity of TIP-C7-amino in Jurkat and MCF-7 cells.

The qualitative growth inhibition observed in Jurkat cells on incubation with TIP-C7-amino in previous uptake experiments was investigated further using standard anti-proliferative assays over a prolonged incubation period. Whilst little growth inhibition was evident at up to 3 $\mu$ M TIP-C7-amino in Jurkat cells, doses of 10 and 30 $\mu$ M produced inhibition in growth of 36% and 82% respectively at 120 hours (Figure 4.26A and B). Polyamide treated Jurkat cells were counter-stained at 72 and 120 hours with propidium iodide and CDFA AM to differentiate between a decreased rate of proliferation and a loss of cell viability (Figure 4.27). This showed that for doses of up to 10 $\mu$ M, cells remained viable even after 120 hours of incubation with TIP-C7-amino. At 30 $\mu$ M, 77% of cells were viable at 72 hours, whilst at 120 hours only 59% of cultured cells were alive. The doses of TIP-C7-amino required to cause 50% growth inhibition relative to control growth ( $I_{50}$ ) over time were calculated and are shown in Table 1. These show that growth inhibition has reached a plateau by 120 hours. For comparison, proliferation was also assessed in MCF-7 cells incubated with TIP-C7-amino over 120 hours using the SRB assay. In contrast to the Jurkat cells, no decreased growth was observed in MCF-7 cells at up to 30 $\mu$ M over 120 hours (Figure 4.28), although at higher concentrations (100 $\mu$ M) some inhibition was observed (data not shown).

The growth of Jurkat cells was also examined after incubation with the TIP-amino monomer over 120 hours, during which no inhibitory effect was observed. Following repeated fixed cell staining experiments, the monomer was shown subsequently to have lost its ability to fluoresce and was therefore concluded to have degraded (data not shown).



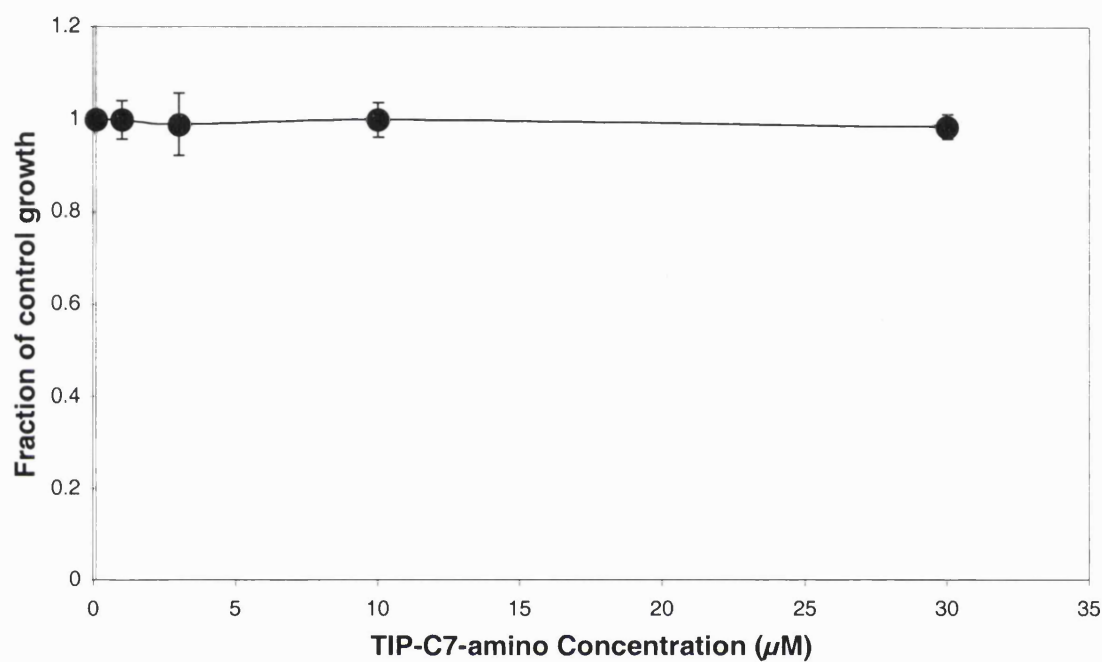
**Figure 4.26** **A.** Growth curves of Jurkat cells continuously incubated with 0μM(◆), 1μM (■), 3μM (-), 10μM (●) and 30μM (▲) TIP-C7-amino over 120 hours. Cells were treated with increasing concentrations of the polyamide dimer TIP-C7-amino and counted as described in section 4.1. **B.** Growth inhibition in Jurkat cells on continuous incubation with TIP-C7-amino over 120 hours. The fraction of control growth for each concentration was calculated as described and results are the mean of at least three independent experiments with error bars showing the standard deviation of the mean.



**Figure 4.27** Viability of Jurkat cells at 72h (■) and 120h (■) at 0μM, 1μM, 3μM, 10μM and 30μM TIP-C7-amino. Cells were treated with increasing concentrations of TIP-C7-amino, and at 72 hours and 120 hours were counter-stained for 1 hour with fluorescein triacetate and propidium iodide, harvested and viewed using an epifluorescence microscope under phase contrast and in the green and red channels as described in the methods section. Percentage cell viability was determined by counting the number of cells visible in the green channel in three fields of view and expressing it as a percentage of the total cell number in that field.

Time (hours)	I <sub>50</sub> (μM)
24	>30
48	>30
72	26
96	16.5
120	15.8

**Table 4.1** Anti-proliferative activity of TIP-C7-amino in Jurkat cells over time. I<sub>50</sub> is the dose required to cause 50% inhibition relative to control growth.



**Figure 4.28** Growth inhibition in MCF-7 cells on continuous incubation with TIP-C7-amino over 120 hours. Cells were treated with increasing concentrations of polyamide and stained with SRB. The fraction of control growth was calculated as described in section 4.1. Results are the mean of at least three independent experiments, with error bars showing the standard deviation of the mean value.

## 4.5 Discussion

Exploiting the property of inherent fluorescence on UV excitation, the cellular uptake of an untagged cross-linked polyamide, TIP-C7-amino, was examined in the two human cell lines MCF-7 and Jurkat using fluorescence microscopy. This enabled a direct assessment of its permeability, uptake and subsequent intracellular distribution in living cells. Polyamide uptake was compared in the two cell types, to address the possibility of a cell-type dependency of uptake, and to a range of structurally related minor groove binders, including distamycin, a TIP-amino monomer, a hairpin linked polyamide, JH-37, a PBD dimer, SJG-136, and a PBD-polyamide conjugate, GWL-6. The PBD dimer SJG-136 has been shown previously to produce covalent DNA adducts in cells traversing both the plasma and nuclear membranes (Gregson *et al.*, 2001), therefore its inclusion in the study enabled an evaluation of the sensitivity of fluorescence microscopy as a method for examining the uptake of small molecular ligands in cells. Finally, the effect of prolonged exposure to the polyamide, TIP-C7-amino, on the proliferation of both cell lines used in the uptake studies was assessed using standard anti-proliferative assays.

Initial studies of the relative fluorescence of these compounds in fixed cells demonstrated a wide range of fluorescent activity on excitation with UV. Distamycin, comprised of three pyrrole rings, and the hairpin linked polyamide JH-37, composed of four pyrrole and two imidazole units, produced the weakest fluorescence which was only detectable in the blue channel following contrast enhancement. This indicated that whilst neither the pyrrole nor imidazole heterocycles are very fluorescent on UV excitation, the emitted signal could still be detected with this methodology. The PBD-tripyrrole polyamide conjugate, GWL-6, and the TIP-amino monomer were also weakly fluorescent, but nevertheless produced a signal just visible in the blue channel without enhancement. In contrast, the PBD dimer, SJG-136 and the cross-linked polyamide dimers TIP-C8-amino and TIP-C7-amino all produced fluorescence that could be visualised clearly in the blue. As the PBD moiety is known to fluoresce on UV excitation, and the pyrrole heterocycle has been shown here not to be very fluorescent, the difference in the relative fluorescent activity of SJG-136, containing two PBD



moieties, and GWL-6 containing only one, may be accounted for by the relative number of component PBD units. Furthermore, the difference in fluorescence observed here between the thiazolated polyamides and the pyrrole and imidazole-containing polyamides, JH-37 and distamycin, suggests that the thiazole heterocycle itself may be strongly fluorescent. The increased fluorescence of the TIP dimers over the monomer may be due, therefore, to the increased number of constituent thiazole units. Furthermore, it is interesting to note that a small increase in the chain length of the linker used, from C7 to C8, also appears to affect the relative fluorescence of the two dimers examined, with C8 linked polyamide producing the greatest fluorescent signal observed for any of the novel compounds examined. Importantly, this work established that untagged polyamides could be visualised using fluorescence microscopy following excitation with a UV source.

All of the compounds examined were shown to localise in the nucleus of fixed cells. Under these conditions the plasma and nuclear membranes have been permeabilised and the cytosol lost. The genomic DNA has partially lost its chromatinous superstructure and is thereby freely accessible to DNA binding agents. Therefore fixed cell studies can only demonstrate an ability to bind nuclear DNA in a membrane free system and do not reflect the transport and uptake processes present in living cells. Auto-fluorescence is not observed after fixation, as it is the product of the UV excitation of component soluble cytosolic molecules, which have been lost on membrane permeabilisation. Thus direct demonstration of the permeability, transport mechanisms and subsequent intracellular localisation of non-covalent and non-cytotoxic DNA binding ligands, such as polyamides, requires measurement in living cells.

Examination of the live uptake and distribution of TIP-C7-amino in the breast cell line MCF-7 demonstrated that at low micromolar doses neither cytosolic nor nuclear uptake was observed in viable cells using fluorescence microscopy. Uptake was seen, however, in dead cells at all concentrations as shown by propidium iodide and CFDA AM counter-staining. At 24 hours, the higher concentrations of polyamide formed a micro-precipitate visible in the culture medium under phase contrast. Subsequently a dose

dependent cytoplasmic accumulation of TIP-C7-amino occurred between 24 and 72 hours. In contrast, the TIP-amino monomer did not precipitate, showing no visible uptake in either the cytoplasm or nuclei of MCF-7 cells over 72 hours. Chromosomal staining during mitosis, in which the nuclear membrane has dissolved, leaving the chromatin freely accessible, was not observed for either compound. These results indicate that even allowing for the increased fluorescence of the TIP-amino dimer relative to that of the monomer, the strongly fluorescent cytosolic uptake of the thiazolated dimer visualised was predominantly intracellular aggregation of the micro-precipitate, possibly ingested by cellular transport processes such as phagocytosis or pinocytosis. These involve vesicular uptake and subsequently internalised molecules would be sequestered within the cytoplasm encased within a vesicular membrane. This in turn would prevent further nuclear uptake from the cytosol and would explain the absence of nuclear staining during mitosis. However, the TIP-amino monomer, which did not form a micro-precipitate, also failed to stain the nuclear material in dividing cells, suggesting that this compound was also not present in the cytosol in a freely soluble form at a sufficient concentration to be visualised.

Previous studies have demonstrated mitochondrial as well as cytosolic accumulation of polyamide in ovarian and breast cell lines using fluorescent tagging (Sharma *et al.*, 2001; Belitsky *et al.*, 2002). In the present study of the untagged polyamide TIP-C7-amino, however, there was no evidence of mitochondrial uptake, with the distribution of the internalised polyamide clearly differing from that of the mitochondria stained with the fluorescent MitoTracker rhodamine 123. This difference may be accounted for if the observed accumulation of TIP-C7-amino in live cells were mediated by a vesicular uptake mechanism, as sequestered polyamide would not be freely available to traverse the mitochondrial membrane. However, the absence of punctate cytoplasmic staining in cells microinjected with drug would argue against this. Alternatively, TIP-C7-amino may not be present in an adequate concentration within the mitochondria to be detected. Furthermore, covalent attachment of a fluorophore moiety to the polyamides used in prior studies may alter the electrostatic properties of the conjugates, thereby influencing

membrane permeability relative to untagged polyamides and accounting for observed differences in distribution.

As recent studies indicate that live cell uptake of polyamides may be cell-type dependent, with nuclear localisation observed in T cell lines (Belitsky *et al.*, 2002), the uptake and intracellular distribution of TIP-C7-amino was also examined in the T cell line Jurkat. Intracellular localisation was more difficult to establish in the Jurkat suspension, compared to the thinner cross-section of the MCF-7 monolayer, as the cells were smaller and their nuclei encompassed a larger relative cell volume. As observed on incubation with the MCF-7 cells, a micro-precipitate formed at high concentrations after which TIP-C7-amino appeared to be cell associated, with cytosolic uptake visible in viable cells. Internalised polyamide did not co-localise with nucleic acid staining by acridine orange, demonstrating that it was not visibly accumulating in the nucleus. CFDA AM counter-staining confirmed that areas of strong polyamide uptake were comprised of non-viable cells. Overall, no difference was observed in the visible intracellular uptake and distribution of TIP-C7-amino in MCF-7 and Jurkat cells in this study. However, as this was the only compound examined in both cell lines and its cytosolic uptake appeared to be predominantly as a solid aggregate, this might not be a general phenomenon.

To examine whether freely soluble internalised polyamide could cross the nuclear membrane and bind the genomic DNA of a living cell, TIP-C7-amino was directly injected into the cytosol of MCF-7 cells together with the fluorescent marker TRITC-dextran. A number of injected cells showed evidence of nuclear as well as cytoplasmic staining by TRITC-dextran, suggesting that the force of flow from injection might have disrupted the nuclear membrane in these cases, as TRITC-dextran is normally excluded from the nucleus. Some cells, however, did show cytoplasmic staining only with TRITC-dextran, with weak nuclear staining observed in the blue. Whilst nuclear localisation here may also be due to membrane trauma rather than native cell transport mechanisms, this result nonetheless indicates that unsequestered TIP-C7-amino can stain the nuclei of living cells and at sufficient concentration be visualised using

fluorescence microscopy. Thus, had internalised polyamide in the uptake experiments been free to traverse the nuclear membrane in sufficient concentrations, it would have been detected. As this was not observed in either MCF-7 or Jurkat cells, this strongly suggests that the polyamides examined are not freely membrane permeant and as such are either being at least partially excluded from the cell or being ingested, at undetectable levels in the case of TIP-amino monomer, in a sequestered form.

Examination of the uptake and distribution of the PBD dimer SJG-136 in MCF-7 cells provided a positive control for live cell uptake and nuclear localisation as SJG-136 forms covalent DNA adducts in cells following 1 hour of incubation at nanomolar concentrations (Gregson *et al.*, 2001). However, no intracellular uptake could be visualised after 1 hour using fluorescence microscopy, with uptake detected only after 24 hours at micromolar concentrations. Furthermore, intracellular distribution could not be ascertained as the cells stained were dead or dying. Interestingly, viable regions remained unstained despite exhibiting gross morphological abnormalities. This indicates that although these cells were beginning to die, whilst they retained intact membranes the PBD analogues, SJG-136 and GWL-6, were not present in sufficient internal concentrations to be detected. Only after the loss of membrane integrity did these compounds become freely permeant, entering the cells at sufficient levels for detection. This suggests that although PBD analogues must traverse both the plasma and nuclear membranes to produce DNA adducts, they are nonetheless not freely permeant. It therefore seems possible that cell death may be initiated by relatively few DNA lesions, assuming that this is the cytotoxic mechanism, which on progression cause membrane degeneration and subsequently a further influx of PBD molecules. Therefore, the PBD uptake visualised here is in membrane-compromised cells that are already dead or dying as a result of preliminary lesions. This initial nuclear PBD uptake, possibly taken up over the course of multiple cell divisions, is not visible in intact and viable cells even after contrast enhancement. Additionally, it is interesting to note that the presence of the polyamide conjugated onto the PBD moiety in GWL-6 does not prevent this compound from entering cells and eliciting a cytotoxic effect.

Uptake and distribution of the hairpin linked polyamide, JH-37, was also evaluated in live cells, although it had been shown to be only weakly fluorescent in fixed cell studies. It did not form a precipitate and neither nuclear nor cytoplasmic uptake was visualised after contrast enhancement. This suggested that JH-37 was also not freely permeable to the plasma or nuclear membrane. Were JH-37 able to accumulate in large concentrations within the nucleus it would have been visible on contrast stretching. As this was not observed, JH-37 was being at least partially excluded from cells or internalised either by passive or active cell processes at concentrations below the limits of detection for the current technique.

To determine whether the nuclear envelope constituted a significant barrier for the nuclear localisation of small molecular DNA ligands, uptake studies of TIP-C7-amino, TIP-amino, GWL-6 and JH-37 were also carried out in isolated Jurkat nuclei. Purified nuclei were clearly differentiated from whole Jurkat cells, which were morphologically similar, using CDFA AM, which identified intact cells only. Subsequent removal of the plasma membrane and surrounding cytoplasm enabled all of the compounds studied to traverse into the nucleus and stain the intact chromatin within 10 minutes, suggesting that the nuclear envelope was readily permeable. These data correlate with work by Janssen and co-workers, which demonstrated that fluorophore-tagged polyamides could selectively stain DNA satellites in isolated Kc nuclei (Janssen *et al.*, 2000a). Such results are unsurprising as the nuclear envelope is a double membrane perforated by numerous pores. These are nearly 90Å wide and are of sufficient size to allow the passage of all but the largest molecular assemblies, such as chromosomes and mature ribosomes, suggesting that small ligand uptake is more likely to be impeded either at the plasma membrane itself or sequestered within the cytoplasm.

In contrast to the live cell uptake studies, which failed to show a visible cell-type dependency of TIP-C7-amino uptake in MCF-7 and Jurkat cells over 72 hours, investigation of its anti-proliferative activity demonstrated a difference in growth inhibition in these cell lines over 120 hours. Whilst, no significant inhibition or cytotoxicity was observed at doses of 10 and 30µM over 120 hours of continuous

incubation in the MCF-7 cell line, a marked decrease in proliferation was seen in the T cell line Jurkat at 10 $\mu$ M over 72 to 120 hours. Indeed by 120 hours of incubation at this dose, growth had been suppressed by 37% whilst cell viability was approximately 95%. At 30 $\mu$ M, evidence of strong growth inhibition was observed from 48 hours of continuous incubation. By 72 hours, 60% inhibition of proliferation was observed with 75% cell viability, whilst at 120 hours there was 82% growth inhibition relative to the untreated controls. This was, however, associated with only 60% cell viability, and thus an ostensible cytotoxic effect was evident at this dose. Nevertheless, growth inhibition was the predominant effect elicited on treatment rather than cytotoxicity.

These results indicate that the T cell line Jurkat is more sensitive to prolonged exposure to TIP-C7-amino than the MCF-7 cells, although growth in the latter is also inhibited at higher doses (100 $\mu$ M). The reason for this difference in growth between the two cell lines is unclear but may be due to a number of factors. Although uptake of TIP-C7-amino over 72 hours was only visualised in both cell lines after the formation of a micro-precipitate, a number of parallel cell transport mechanisms may be ingesting the polyamide, possibly including low levels of passive diffusion across the cell membranes. As growth inhibition is most evident between 72 and 120 hours and incubation periods were terminated after 72 hours for the uptake studies it cannot be excluded that some nuclear uptake may have been detected over that time period had the experiments been extended. Leeching from intracellular vesicles containing sequestered polyamide could also allow low levels of soluble polyamide into the cytosol, available for subsequent nuclear uptake. However, it cannot be excluded that the growth inhibition observed in the Jurkat T cell line may not be dependent on interactions of the polyamide with nuclear DNA and may be the result of another molecular mechanism.

The work presented in this chapter demonstrates that fluorescence microscopy may be used to visualise directly the uptake and intracellular distribution of untagged molecular ligands in cells. With the ability to amplify the collected signal in the form of contrast enhancement, the microscopy employed here is as sensitive as conventional confocal microscopy. The fixed cell studies clearly indicate that were any of the untagged



compounds examined freely permeant and able to localise to the nuclear compartment at a significant concentration they would have been visible using this technique, either without or with contrast enhancement depending of their relative fluorescent activity. Thus under freely permeable conditions this method is sufficiently sensitive to visualise untagged as well as fluorophore-tagged ligands. The ability to visualise untagged molecules importantly excludes the possibility that the size and charge of a conjugated fluorophore may affect the dynamics of ligand uptake in cells. As nuclear localisation of polyamides in living cells was not observed, it appears within the detectable limits of this technique that the compounds examined are not reaching the nuclear compartment of MCF-7 cells at high concentrations in a freely soluble state. This strongly suggests that they are either at least partially excluded from the cell, in the cases where no intracellular uptake was observed, or are being sequestered intracellularly, as seen with TIP-C7-amino, either at the point of uptake or subsequent to internalisation. Importantly, the nuclear envelope does not appear to be a barrier to nuclear localisation.

As passive diffusion of small molecules into cells is strongly charge dependent, the electrostatic properties of the polyamides examined, for example the presence of the cationic amino head group in the case of the two TIP compounds, may not be compatible with membrane permeability. This has wider implications as future therapeutic applications for gene targeting using polyamides will rely extensively on their effective cellular uptake and nuclear localisation. Factors influencing delivery in living cells, such as molecular size, electric charge, non-specific protein binding and hydrophobicity/lipophilicity at physiological pH (Horobin *et al.*, 1990; Larsen *et al.*, 2000), must be considered in the design of future generations of polyamide molecules for use as gene regulators. Other gene targeting strategies including anti-sense have also encountered problems with delivery, as oligonucleotides are polyanions and will not readily pass the hydrophobic lipid bilayer constituting the plasma membrane in adequate amounts to show activity. This has resulted in the development of different types of carrier systems to enhance their absorption and distribution (Jaaskelainen *et al.*, 2002).

The conjugation of polyamides to alkylating moieties such as chlorambucil, CBI, CPI and PBDs nevertheless does not appear to prevent entry into cells, with the optimal resultant conjugates demonstrating an increased cytotoxicity and sequence specificity *in vitro* relative to the alkylating agents alone (Wang *et al.*, 1996; Kumar & Lown 2003; Wang *et al.*, 2002, 2003). Assuming that DNA alkylation is responsible for the observed cytotoxicity, this strongly suggests that alkylating conjugates are able to cross the cell membrane at sufficient levels to interact with the genomic DNA producing sequence-specific lesions and are not excluded by the attachment of a polyamide moiety. Thus the electrostatic properties of the alkylating moiety itself may alter the overall charge of the conjugate enabling its passage across the cell membranes. As observed here with SJG-136 and the PBD-polyamide conjugate GWL-6, it may be that relatively low concentrations are able enter the cell initially, but that these produce sufficient lesions to initiate cell death. Other structural factors may also increase the lipophilicity of these molecules, for example the hydrophobicity of their linkers (Lown *et al.*, 1989).

Subsequent to these studies, Crowley and coworkers have shown that the absence of nuclear localisation by polyamides in some cell lines may be due to lysosomal sequestration rather than impermeability of the nuclear envelope (Crowley *et al.*, 2003). They have demonstrated that by changing the fluorophore used for polyamide tagging from BODIPY to fluorescein, neutralising the overall charge of the fluorophore-polyamide conjugate from a weakly basic cation by the addition of an anionic moiety, that lysosomal sequestration is prevented and nuclear accumulation occurs as a result. Furthermore, in the human fibroblast cell line (RSF), use of the p-glycoprotein inhibitor verapamil can block lysosomal polyamide sequestration, possibly by the disruption of vesicular trafficking. This, however, did not occur in an epithelial cell line (HCT116) only and cannot be generally induced by other p-glycoprotein inhibitors such as bepridil, cyclosporin A and ketoconazole.

The results reported here demonstrate that the uptake and distribution of untagged molecular ligands, including polyamides, may be visualised in living cells by

fluorescence microscopy within the limits of detection. This has shown that untagged polyamides are not freely cell permeable and under some circumstances may be sequestered intracellularly, in turn preventing their nuclear accumulation. Polyamide present in a freely soluble form in the cell cytoplasm does appear to be able to accumulate in the nucleus as shown by microinjection, suggesting that the nuclear membrane itself may not be the primary barrier to nuclear accumulation. This is supported by the rapid and unimpeded uptake of the compounds in isolated nuclei. The inability of polyamides to localise in the nuclei of living cells does not appear to be cell type dependent, although there is some evidence of increased sensitivity in the T cell line Jurkat to growth inhibitory effects of prolonged polyamide exposure. However, this may be as a result of predominant intracellular accumulation of the polyamide examined as a solid aggregate. Moreover, the cytotoxic ligands SJG-136 and GWL-6, a PBD dimer and PBD-polyamide conjugate respectively, are also not freely permeant although they may nevertheless induce sufficient covalent lesions to initiate cell death, suggesting that relatively few DNA adducts may be required to produce a cytotoxic effect. Importantly, the use of untagged ligands provides an accurate assessment of their uptake and distribution in living cells. As Crowley and coworkers have clearly demonstrated recently, the physicochemical properties of molecules used for conjugation in fluorophore-tagged uptake studies may significantly alter the uptake characteristics of the resultant conjugate, providing a misleading model of live cell uptake (Crowley *et al.*, 2003).

## CHAPTER 5

### THE EFFECTS OF DISTAMYCIN AND THE THIAZOLE-CONTAINING POLYAMIDE TIP-C7-AMINO ON GLOBAL GENE EXPRESSION IN THE MODEL EUKARYOTIC SYSTEM *SACCHAROMYCES CEREVISIAE*.

#### 5.1 Introduction

The development of multiple strategies for the selective targeting of nucleic acids at either the level of the message or the gene provides potential for the artificial modulation of the transcription of disease-related genes including those implicated in oncogenesis. One recent antigene strategy has involved the synthesis of distamycin derivatives, or polyamides, able to bind to DNA sequences of increasing length with an affinity and selectivity resembling that of transcriptional regulation molecules in cells (Dervan 2001). Whilst difficulties persist in generating polyamides capable of targeting of sixteen base pairs, the minimum size required for identifying a unique region of sequence, whilst retaining optimal affinity and sequence selectivity, a number of polyamides have nevertheless been shown to be able to up or down-regulate the expression of selected genes *in vitro* on a consistent basis (Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998, 1999a, b; Chiang *et al.*, 2000; Mapp *et al.*, 2000). Furthermore, although the cell uptake characteristics of polyamides have not been fully characterised and current evidence suggests that they may not be freely permeable, the uptake of polyamide-dye conjugates has recently been demonstrated directly in some human T cell lines (Belitsky *et al.*, 2002; Dudouet *et al.*, 2003). As a consequence, few studies have been published examining the effect of polyamides on the expression of genes as well as predetermined site-specific binding of DNA by polyamides in living cells.

Recent development in the field of genomics involving the evolution of spotted cDNA microarrays (Brown & Botstein 1999) has enabled the simultaneous measurement of genome-wide mRNA expression. In this technique cDNA clones of known sequence are spotted and immobilised onto a glass slide or nylon membrane to produce a probe in the form of a cDNA grid known as cDNA microarray or 'DNA chip'. Pools of purified

mRNA from the cell populations under investigation are reverse transcribed into cDNA and then labelled prior to application to the cDNA microarray chip. The pools of cDNA may be labelled either using pairs of fluorescent dyes, where one cDNA pool is labelled with a red dye and the other with a green dye, or using radioisotopes such as  $^{33}\text{P}$  where both cDNA pools are labelled with the same isotope. Use of pairs of fluorescent dyes for labelling enables simultaneous application of the two pools of cDNA to the microarray chip, whereas radiolabelled cDNA pools require separate rounds of hybridisation. Comparison of the intensity of the signal produced from a given spot after hybridisation of both cDNA pools indicates the relative abundance of the corresponding mRNA in the two cell populations.

Using this technique, work carried out concurrently with experiments in this thesis, has investigated the levels of mRNA expression of polyamide-treated human DLD-1 colon cells (Supekova *et al.*, 2002) and lymphoid MT2 cells (Dudouet *et al.*, 2003). Both of these studies found that relatively few genes were affected by treatment, with different transcription profiles generated by polyamides targeted to different sequences (Dudouet *et al.*, 2003). Many of the gene transcripts consistently affected by treatment were identified as cell stress and various signal transduction genes. Expression levels of target genes were unaffected by polyamide treatment directly in DLD-1 cells. Analysis of the promoter regions of affected genes in order to identify potential polyamide binding sites linked to transcriptional regulation revealed numerous possible sites, however, no consensus binding motif was found (Supekova *et al.*, 2002).

The aim of the work presented in this chapter was to carry out a preliminary investigation into the effect of distamycin A and the polyamide TIP-C7-amino on global gene expression profiles in the eukaryotic model system *Saccharomyces cerevisiae* using DNA microarray hybridisation, in an effort to examine the ability of polyamides to modulate gene expression in cells. DNA hybridisation in *S. cerevisiae* provides a useful model in which to investigate gene expression profiling, as the relatively small genome (approximately 6000 genes) has been established and subjected to extensive mutational analysis. As the biological effects and cell uptake properties of polyamides

in yeast strains have not been defined, *S. cerevisiae* was treated with distamycin and TIP-C7-amino to evaluate potential anti-proliferative effects. These experiments provided a guide in the design of subsequent gene expression profiling experiments using nylon filter cDNA microarrays, which were carried out to evaluate the biological consequences of these treatments at the level of the gene.

## **5.2 Materials**

### **5.2.1 Investigational compound**

TIP-C7-amino provided and diluted as in materials section 2.2.

### **5.2.2 Radioisotope**

[ $\alpha$ -<sup>33</sup>P] CTP (10mCi/ml, ~2500Ci/mmol) was purchased from Amersham International.

### **5.2.3 Chemicals**

D-glucose, distamycin A, diethyl pyrocarbonate (DEPC),  $\beta$ -mercaptoethanol, yeast extract, sodium dodecyl sulphate (SDS), distamycin A and sodium chloride were purchased from Sigma

Dithiothreitol (DTT) 0.1M and mycological peptone were purchased from Invitrogen and Lab-M, respectively.

A 20mM dNTP mixture was made up from 100 $\mu$ l aliquots of the 100mM stocks of dGTP, dATP and dTTP purchased from Invitrogen in 200 $\mu$ l of DEPC treated water.

### **5.2.4 Enzymes**

Lyticase (5000U/ml) and Superscript™ II RNase H- Reverse transcriptase (200U/ $\mu$ l) were purchased from Biorad and Anachem, respectively.

### **5.2.5 Buffers**

5X First strand buffer containing 250mM Tris-HCl (pH 8.3), 375 mM KCl, 15mM MgCl<sub>2</sub> was purchased from Invitrogen

Y1 Buffer: 1M sorbitol, 0.1M EDTA pH7.4.

20X SSC: 17.5% sodium chloride, 8.8% trisodium citrate pH7.0.



DEPC water: 0.1% DEPC treated distilled deionised water.

#### 5.2.6 Growth media

YEPD agar: 1% mycological peptone, 1% yeast extract, 2% D-glucose, 1.8% bacteriological agar.

YEPD: 1% mycological peptone, 1% yeast extract, 2% D-glucose.

#### 5.2.7 Kits

An RNeasy kit, containing mini spin columns and buffers RW1 and RPE, was purchased from Qiagen.

A LIVE/DEAD yeast viability kit containing the yeast cell stains 10mM FUN1 and 5mM Calcofluor White M2R was purchased from Molecular Probes.

#### 5.2.8 Microarray Filters

Yeast Index Genefilters Microarrays, MicroHyb hybridisation solution, Poly dA (1µg/µl) and Oligo dT 1µg/µl were purchased from Invitrogen.

#### 5.2.9 Data analysis programs

OptiQuant imaging software and Pathways image analysis software were purchased from Research Genetics.

Statistical analysis of the data was carried out using Microsoft Excel.

#### 5.2.10 Identification of genes derived from expression profiles

Genes were identified using the *Saccharomyces* Genome database (SGD) at <http://www.yeastgenome.org/>.

## 5.3 Methods

### 5.3.1 Anti-proliferative activity of polyamides in *S. cerevisiae*

The anti-proliferative activity of distamycin A and the polyamide TIP-C7-amino were evaluated in the wild type *S. cerevisiae* strain DBY747 using two different assays. A standard colony counting assay was used for a preliminary investigation of cell proliferation following treatment for 2 hours with each investigational compound over a wide dose range. Cellular proliferation was then examined in the culture suspensions over time following treatment by measuring the optical density of the culture at 600nm. This enabled determination of the viability of aliquots of cells from the treated cultures, in an effort to discriminate between growth inhibition and cytotoxicity using cell staining techniques and subsequent visualisation by fluorescence microscopy.

#### *Measurement of growth using colony counting:*

A culture of yeast *S. cerevisiae* strain DBY747 was grown overnight in YEPD growth media in a shaking incubator at 30°C. The following day, 1 ml aliquots of  $5 \times 10^7$  cells were treated either with TIP-C7-amino or distamycin over a range of concentrations (0.1-100µM). Control samples were mock treated with an equivalent volume of drug vehicle, deionised H<sub>2</sub>O. Treated samples were incubated in an orbital shaker at 30°C for 2 hours. Following incubation, samples were spun down at 13000rpm for 30 seconds, the supernatant discarded and the pellets washed twice with 1ml of deionised H<sub>2</sub>O. These were resuspended in 1ml YEPD, diluted by 1/10000 in YEPD and 100µl aliquots of the dilution were plated out on two YEPD agar plates. Plates were incubated for 3 nights at 30°C and the colonies on each plate were counted manually to determine growth. Experiments were carried out in triplicate and the fraction of control growth was calculated by manually counting the number of colonies on drug treated plates as a proportion of the number growing on control plates. The mean fraction of control growth and standard deviation was calculated for each concentration.

#### *Measurement of growth using optical density:*

A culture of the *S. cerevisiae* wild type strain DBY747 was grown overnight in YEPD in an orbital shaker at 30°C. The following day, the cell culture was counted, diluted in

fresh pre-warmed YEPD and aliquotted into 2ml volumes at a concentration of  $2 \times 10^6$  cells per ml. Cells were treated in triplicate with a concentration range of 0-100 $\mu$ M with either TIP-C7-amino or distamycin A, whilst control samples were mock treated with an equivalent volume of drug vehicle, deionised H<sub>2</sub>O. Samples were incubated at 30°C over 24 hours during which culture growth was measured over time by taking 100 $\mu$ l aliquots from the treated cultures at 0, 2, 4, 6, 8, 10 and 24h intervals and determining the optical density of the samples at 600nm. This was done by diluting the time point aliquots 1:10 in YEPD growth medium and measuring the optical density of the samples at 600nm against a sample of YEPD alone using a an LKB Biochrom Ultraspec II 4050 UV/visible spectrophotometer.

To assess the proportion of viable cells remaining after 24 hours, cells were double-stained with the viability probes FUN 1 and Calcofluor White M2R. Staining was effected by washing a 50 $\mu$ l aliquot from each treated culture in 1ml of filter sterilised ddH<sub>2</sub>O containing 2% D (+) glucose and 10mM Na Hepes (pH 7.2). Cells were concentrated by centrifugation for 5 min at 13000 rpm, after which the supernatant was removed and the cells were resuspended in the 2% D (+) glucose, 10mM Hepes solution (pH 7.2). The stains FUN1 and Calcofluor White M2R were then added to each cell suspension to a final concentration of 10 $\mu$ M and 25 $\mu$ M, respectively, and samples were mixed thoroughly and incubated in a shaking incubator at 30°C for 30 min. Following staining, a 5 $\mu$ l aliquot of each stained cell suspension was trapped between a glass microscope slide and a 18mm x 18mm coverslip and viewed using a Zeiss Axiophot epifluorescence microscope equipped with a mercury halide light source and narrow band blue, green and red filters. Cells were viewed with a 40x (NA 0.75) Plan Neofluar phase objective under phase contrast and in the red and blue channels.

Calcofluor White 2MR is a UV-light excitable counterstain, which labels yeast cell walls a fluorescent blue regardless of the cell's metabolic state. The viability probe FUN 1 is excitable at 480nm and is used to stain metabolically active yeast. It is initially taken up into the cytoplasm by passive diffusion staining the cytosol fluorescent green, however, once internalised in live cells FUN1 is metabolised to an alternative

form which stains intravacuolar structures present in live cells a fluorescent red. Thus metabolically active cells with an intact plasma membrane exhibit strong punctate red staining of their intravacuolar structures together with a weak and diffuse green cytoplasmic fluorescence. Metabolically inactive cells with intact plasma membranes exhibit diffuse cytosolic staining but lack fluorescent intravacuolar bodies. Using these properties, the percentage of viable cells was calculated by counting the number of cells with red intravacuolar bodies as stained by FUN 1 as a proportion of the total cell number as stained by Calcofluor White M2R in 3 fields of view for each culture aliquot. A mean and the standard deviation of the mean were calculated for three independent experiments.

### 5.3.2 Profiling the response in gene expression on treatment of *S. cerevisiae* with Distamycin A or TIP-C7-amino

#### *Drug treatment:*

A 10ml culture of *S. cerevisiae* strain DBY747 was grown overnight in YEPD growth media in a shaking incubator at 30°C. The following day, the culture was diluted in fresh pre-warmed YEPD to provide an appropriate number of 1ml aliquots of  $5 \times 10^7$  cells. These were then treated with either 100µM TIP-C7-amino, 100µM distamycin A, or mock treated with an equivalent volume of the drug vehicle, deionised H<sub>2</sub>O, and incubated for a further 2 hours in the orbital shaker at 30°C. Drug treatments were carried out either in parallel or in series, in an effort to explore and then limit the number of experimental variables. Treatments carried out in parallel involved the simultaneous drug or mock treatment of a number of 1ml aliquots from the same cell culture batch, so that on comparison of these experiments the only biological variable was the drug treatment itself. Treatments carried out in series involved the drug or mock treatment of cells taken from a different batch of cell culture. This meant that for these experiments there were two biological variables, the drug treatment and the cell culture used.

#### *RNA extraction and quantification:*

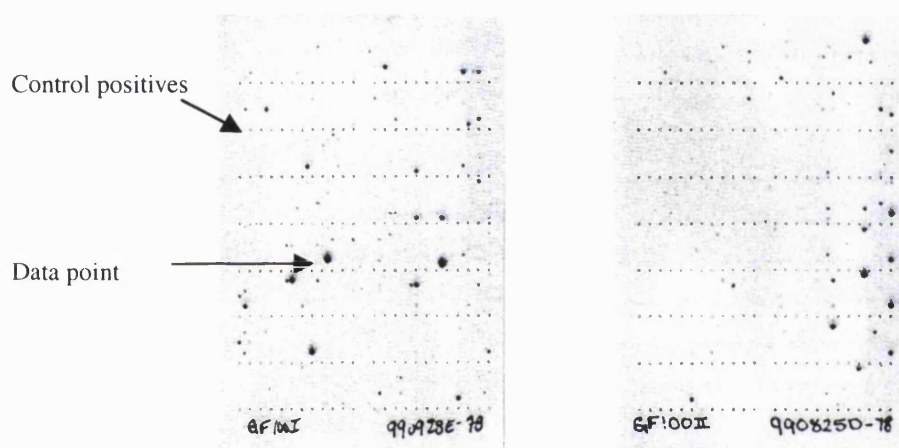
Following incubation with the investigational compounds, the treated samples were centrifuged at 2000rpm at 4°C for 30 seconds and the supernatant discarded. Each

sample was washed twice in 1 ml ddH<sub>2</sub>O, resuspended in 2ml Y1 buffer (1M sorbitol, 0.1M EDTA pH 7.4) with 2µl B-mercaptoethanol and 30µl (150U) of lyticase (5000U/ml, Biorad) to remove the cell wall and incubated at 30°C in an orbital shaker for 15 minutes. Samples were centrifuged at 1000rpm at 4°C for 5 min and the supernatant discarded. RNA was then extracted from the remaining spheroplasts using a Qiagen RNeasy kit following the manufacturers protocol and quantified by measurement of absorbance at 260nm using an LKB Biochrom Ultraspec II 4050 UV/visible spectrophotometer. The absorbance of each sample at 280nm was also measured to ascertain the level of contaminant protein concentration and an absorbance ratio of protein: RNA ( $A_{280}/A_{260}$ ) was calculated. Samples with a  $A_{280}/A_{260}$  ratio of over 2 were discarded, as this indicated high levels of contaminant protein, whilst the quality of extracted RNA with an absorbance ratio of less than 2 was further assessed by running a 1µg aliquot alongside a 1kb DNA marker on a 1% agarose gel. High quality RNA showed two distinct bands at 1 and 0.7 kbp corresponding to 28S and 18S ribosomal RNA, respectively.

#### *Preparation And Hybridisation Of Yeast Genefilter™ Microarrays:*

Prior to use, the two yeast microarray filters were stripped for 1 hour with 0.5% boiling SDS to remove any residuals from the nylon membranes. The two filters were placed DNA side up in a hybridisation roller tube so that they did not overlap and saturated with 5ml MicroHyb hybridisation solution and 5µl of a blocking agent poly dA (1 µg/µl), and pre-hybridised at 42°C for 2 hours in a Techne Hybridiser HB-1D. 1µg aliquots of treated RNA was then primed with 2µl oligo dT (1µg/µl) in a total volume made up to 10µl by diethyl pyrocarbonate treated ddH<sub>2</sub>O for 10 minutes at 70°C and chilled on ice for 2 minutes. The primed RNA samples were then used to generate <sup>33</sup>P-dCTP labelled cDNA probe by incubation with 300U of the reverse transcriptase Superscript II (200U/µl) at 37°C for 90 min, following the manufacturer's protocol. Unincorporated <sup>33</sup>P-dCTP was removed by spinning the labelled probes with 70µl ddH<sub>2</sub>O through a biospin-6 column at 1500 rpm for 5 min at room temperature. The counts per minute of the radiolabelled cDNA probe were then measured using a Wallac 1409 liquid scintillation counter. Probes were discarded if there was more than a 20% difference in

the measured counts. A labelled probe was denatured for 3 min at 100°C, added to the pre-hybridised filters and hybridised overnight for 15 hours at 42°C. Following hybridisation, the hybridisation mixture was discarded and the filters washed twice for 25 min at 50°C with a wash solution (2X SSC, 1% SDS). A final wash was carried out at room temperature for 15 min with a second wash solution (0.5XSSC, 1% SDS). Filters were placed on 3MM chromatography paper (Whatman) moistened with deionised H<sub>2</sub>O, wrapped in Saran wrap and exposed to a phosphor screen for 24 hours. The filters were scanned using a Packard Cyclone phosphor imager and visualised using OptiQuant software, as shown in Figure 5.1.



**Figure 5.1** An example image of a pair of Yeast Genefilters. Each filter has individual DNA clones from the yeast species *S. cerevisiae*, spotted onto and cross-linked to its surface. Each of these clones corresponds to the coding region of one of the 6144 genes or yeast “ORFs” (open reading frames) which can be identified from the complete yeast genome as listed in the *Saccharomyces* Genome Database (SGD). The 6144 ORFs are divided between the two filters so that each filter has 3072 clones, with each ORF spotted at a known position on that filter. Thus the two filters contain between them the entire complement of coding sequence from this organism. A system of positive controls (as indicated by an arrow) comprised of spots of total genomic DNA served to monitor the homogeneity of hybridisation runs, as well as facilitating the alignment of the filter images prior to analysis of the filters using the Pathways software.



### *Image Quantitation:*

Due to the degenerative effect of repeated cycles of stripping, each pair of Yeast Genefilters was used for no more than four hybridisation runs. Furthermore, in an effort to reduce variation within a parallel treatment data set, where possible the cDNA derived from such a treatment was hybridised to the same filter set. Thus at any one time cDNA derived from one of these drug treatments was hybridised to a single pair of filters (GF-100-a and GF100-b). Samples from the same treatment batch were hybridised in series to the same pair of filters. Phosphor images of the filter pair used for each hybridisation run were scanned using OptiQuant software and imported into the Pathways filter analysis software, for quantification of the raw intensity values of each ORF present on the filter pair. Each pair of filters was stripped for 1 hour in boiling 0.5% SDS between hybridisation runs.

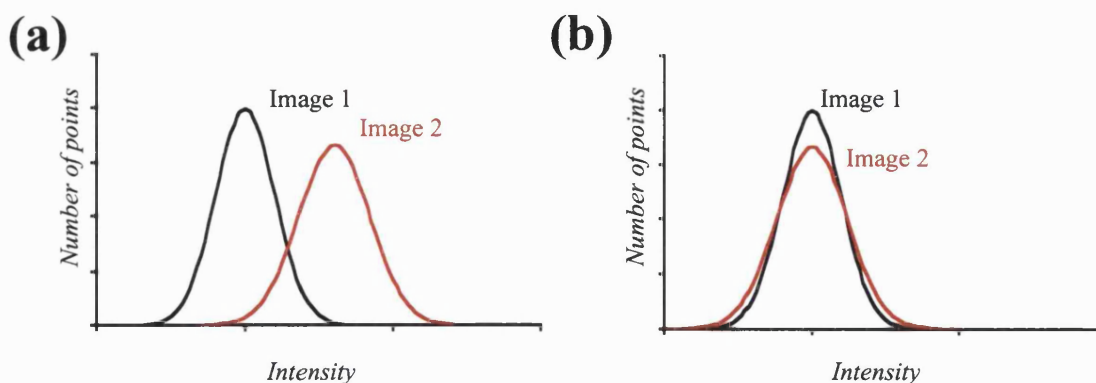
To enable the mapping of the ORFs spotted onto both filter membranes, the position and identity of the ORF DNA clones spotted are stored in the software in the form of two grids. On alignment of the filter images with the appropriate grid, the ORFs may be identified on the basis of their original grid location provided that the correct grid information has been selected. Each image was aligned individually to the appropriate filter grid (GF100-a or GF100-b) by eye using a crosshair to locate the four main control positives, one at each corner of the filter, so that the image skew was no greater than  $\pm 10^\circ$ . Once the image was aligned, the position of the grid was fine-tuned using the crosshairs to centre twelve further control points or spots by eye. This then enabled the appropriate clone grid to be accurately aligned and applied to the correct filter image, automatically providing a reference position for every spot on the image with each spot corresponding to a known ORF. With each ORF located, the raw spot intensities on that filter were measured, whereby the intensity was the numerical value assigned to the level of expression of a given gene or ORF. These values were then automatically stored as a single data set. This process was carried out separately for each filter in the pair and the data measured for each filter was stored as a separate dataset. This was repeated for each hybridisation run.

### *Comparison of filter datasets:*

To assess the effect of drug treatment on the expression levels of the entire complement of yeast genes, datasets of the raw intensities derived from each drug treatment were compared to those from the corresponding mock treated control. To enable direct comparison of two datasets, the intensities of spots need to be adjusted so that they represent the equivalent level of gene expression on both the drug treated filter and the control filter. This involves correcting for background and normalising the data.

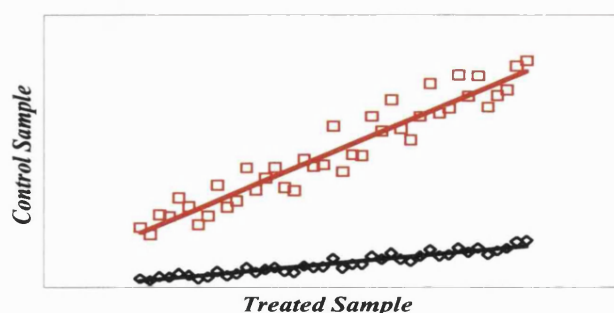
Background correction removes the contribution of noise caused by background radiation, non-specific hybridisation and bleed from adjacent spots. Correcting for background noise involves subtracting a value from all spot intensities so that those intensities that represent no change in gene expression have their values aligned to zero. This is of particular importance when comparing low intensities. Most correction techniques involve the subtraction from all spots of a global background value, obtained from the intensity of an area remote from all spots, or the subtraction from each spot of a local value, taken from an area adjacent to the spot. However, when the spot intensities are low, background values are often greater than the spot intensity resulting in a negative value. For these values subtraction of background from each of the corresponding filters introduces a large additive error which distorts the degree of variation at low intensities.

To enable the direct comparison of datasets across the entire range of data values, the intensities need to be normalised. This involves the multiplication of the intensity values by a factor calculated by dividing the average intensity value for all the data points on the drug treatment filter by those from the equivalent control filter. The normalisation process (as illustrated in Figure 5.2) accounts for differences in the efficiency of hybridisation and the radioactivity levels between different filter datasets, thereby enabling a direct comparison of the two filter images. Normalisation is often applied after background correction and results in the magnification of errors introduced by background correction.



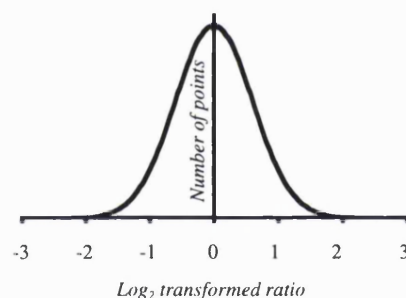
**Figure 5.2** Illustration of the normalisation of the distribution of intensities for two filter images, showing distribution (a) before and (b) after normalisation.

In order to overcome the problems of error associated with a multistep approach to background correction and normalisation, a model was employed to correct and normalise the data in a single step thereby reducing the accumulation of error as a result of serial transformations. The raw intensities from the treated sample were plotted against those from the control to produce a scatter plot (Figure 5.3). A linear model was fitted to the plot providing values for the gradient and intercept. The intercept represents the relative deviation from zero of the intensities that represent no change in gene expression within each dataset. The intercept was subtracted from all spot intensities in the treated sample. The gradient represents the relative efficiency of radioactive labelling and hybridisation of the two samples. All intensities in the control dataset were then multiplied by this gradient so that the overall efficiency ratio equalled one.



**Figure 5.3** Illustration of a scatter plot showing the intensities of the treated sample against the control before (◊) and after (◻) normalisation, with solid lines representing the regression model before (-) and after (-) normalisation.

Ratios of gene expression were calculated by dividing the corrected intensity of each ORF from the treated sample by the corresponding corrected intensity in the control. The resultant ratios were all greater than zero regardless of the difference in expression levels. For example, a ratio of 2:1 would give a value of 2.0 and a ratio of 1:2 would give 0.5. To clarify that these values are the same magnitude of change but in opposite directions, the raw ratios were imported into an Excel spreadsheet where they were  $\log_2$  transformed so that the values would be spread symmetrically around zero. Thus a ratio value of 0.5 is  $\log_2$  transformed to  $-1$  and 2 to  $+1$  (as illustrated in Figure 5.4). Furthermore, transformation of the raw expression data to a log scale eliminates the proportional relationship between the signal intensity and random error. The  $\log_2$  transformed ratios provided the basis for a gene expression matrix, whereby the relative expression levels of different genes could be examined by comparing rows in the expression matrix whilst the relative expression profiles of different samples (or treatments) could be assessed by comparing columns.



**Figure 5.4** Representation of the distribution of the  $\log_2$  transformed normalised intensity ratios.

*Statistical Analysis and Significance Inference Of The  $\log_2$  Transformed Ratios:*

As it is statistically inefficient to infer a fixed threshold cut off method to filter the  $\log_2$  transformed ratios, for example  $\pm 2$ , the mean values ( $\mu$ ) and standard deviation ( $\sigma$ ) of the  $\log_2$  transformed ratios were calculated for each gene across repeat drug treatments for all 6144 ORFs. A two-tailed-t-test was then used to statistically assess the likelihood of differential gene expression. The t-test statistic is a measure of the probability that two mean values obtained from independent measurements are the same. In this case, the two means to be compared are the mean value from the repeated measures of gene

expression and the mean value representing no change in gene expression. Using the  $\log_2$  transformed ratios no change in gene expression is represented by zero. The t-test statistic was calculated using the equation below, where n was the number of repeats:

$$t = \frac{\mu\sqrt{n}}{\sigma}$$

This measure was converted to a probability or p-value using tabulated values for a t-distribution where low probabilities indicate that it is unlikely there has been no change in expression of a particular gene.

Different experimental designs were used in order to assess and minimise the biological variation introduced by effects other than those due to treatment (e.g. those due to different cell populations). Experiments were carried out for parallel (i.e. same culture batches) and serial (i.e. different culture batches) designs. Comparisons were made between the significant genes across either parallel or serial experiments for a given treatment dose. Genes common to both designs (i.e. both with  $p < 0.05$ ) were listed separately. Genes in all the parallel design experiments were ranked in ascending order according to their probability and the top twenty genes were selected for further investigation. The biological functions of all selected genes were obtained using a genome database.

Variation was compared for the serial and parallel design experiments. The standard deviation was calculated for all genes in the serial and parallel experiments and was plotted against the mean differential expression ratio. A non-linear model was fitted to this plot. This model was used to estimate the expected variation for all levels of differential expression. Using these values for variation, the minimum number of replicates required to achieve significant differences in gene expression ( $p < 0.05$ ) was calculated for both serial and parallel designs. The number of replicates required was then plotted against the level of differential expression for each source of variation.

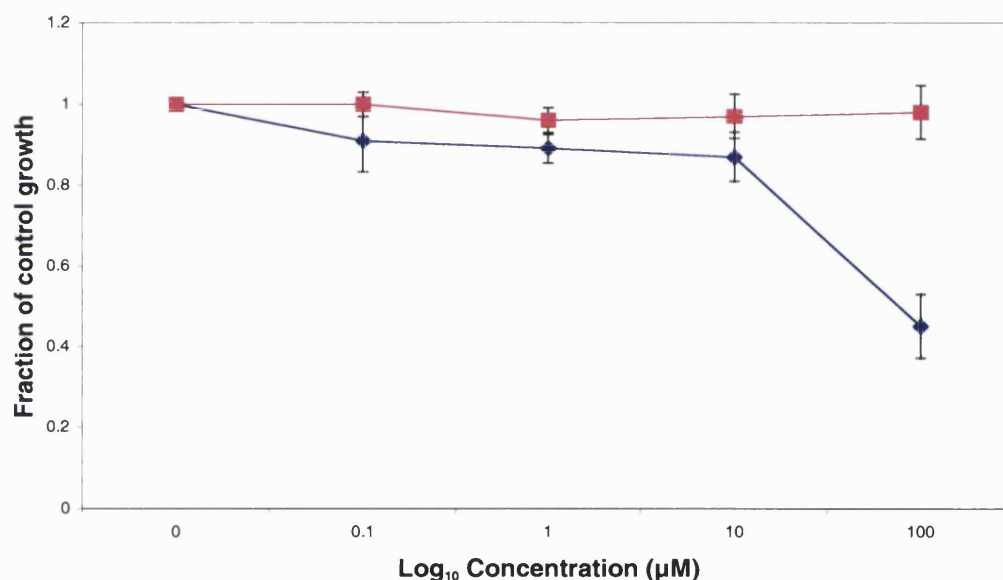
#### *Identification of the protein products of genes affected on treatment:*

The name, identity and possible biological function of the protein products of each gene/ORF were identified using the *Saccharomyces* Genome Database (SGD).

## 5.4 Results

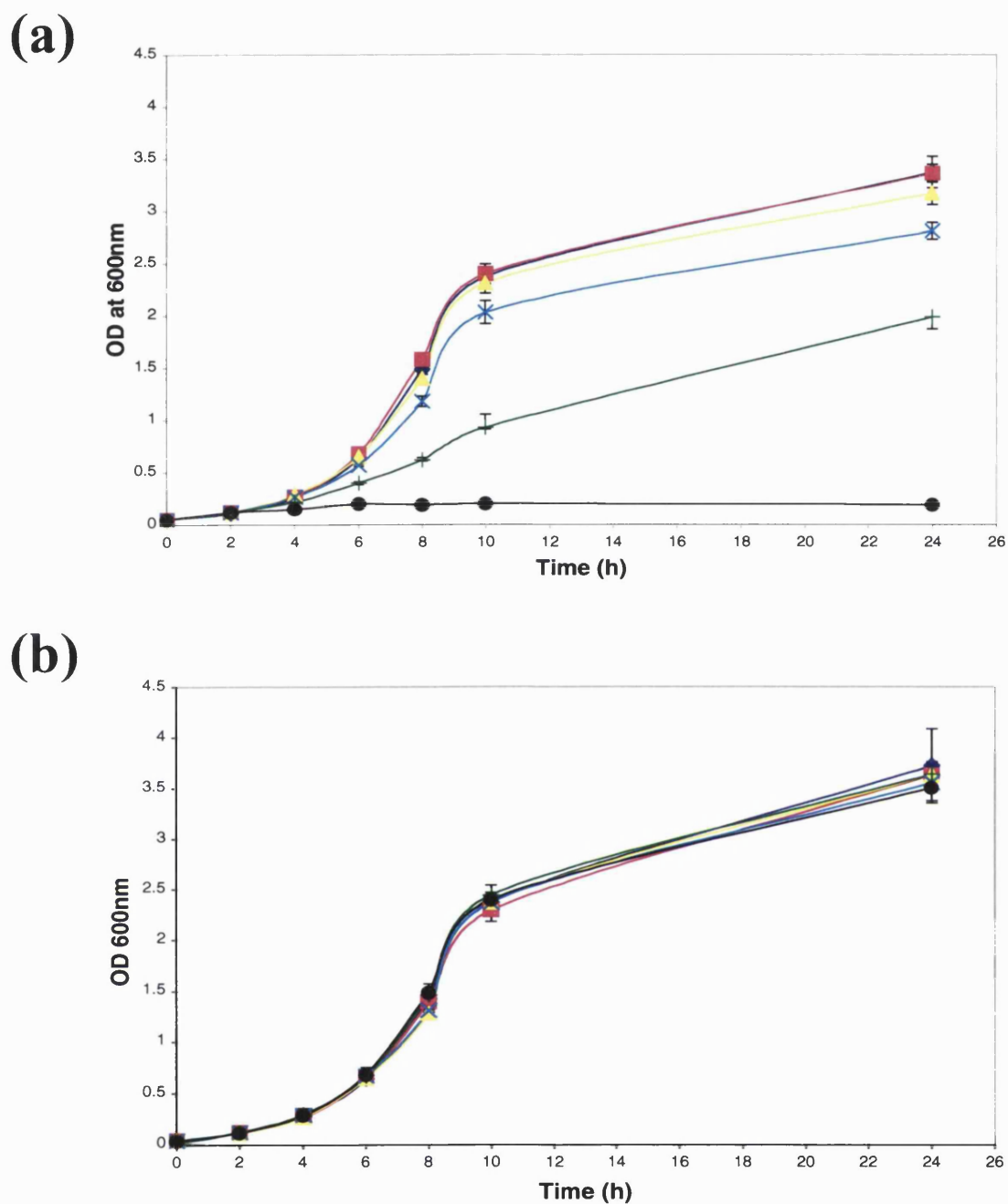
### 5.4.1 Anti-proliferative activity of distamycin A or TIP-C7-amino in *S. cerevisiae*.

To establish whether distamycin A or the polyamide TIP-C7-amino caused an anti-proliferative effect in the *S. cerevisiae* strain DBY747 cultures were treated with these agents over a wide dose range and over a prolonged time period. Colony-counting assays examining the growth response after incubation with these agents for two hours demonstrated differing responses to each agent. Whilst little growth inhibition was observed with TIP-C7-amino, which produced 98% of control growth at 100 $\mu$ M, distamycin A had a strong anti-proliferative effect over the dose range examined, produced up to 55% inhibition in colony growth (Figure 5.5).



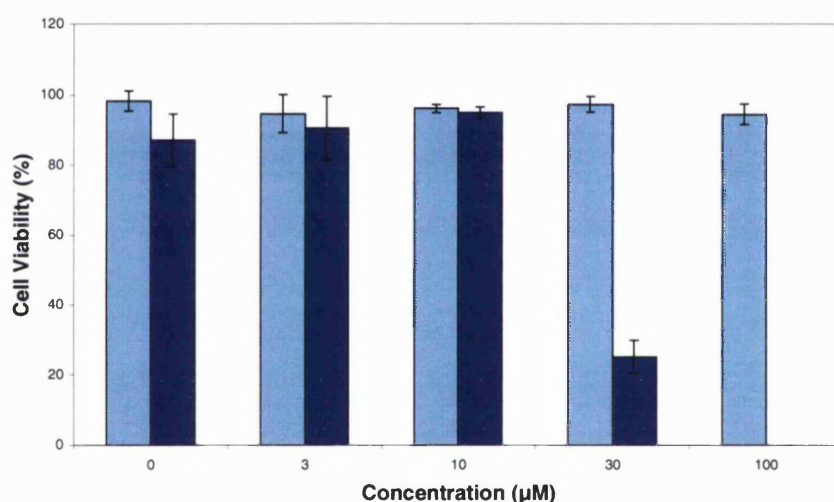
**Figure 5.5** Graph of the fraction of control growth versus dose for *S.cerevisiae* strain DBY747 after incubation with either distamycin A ( $\blacklozenge$ ) or TIP-C7-amino ( $\blacksquare$ ) for 2 hours. The mean % survival was calculated as described and results are the mean of at least three independent experiments with error bars showing the standard deviation of the mean.





**Figure 5.6** Graphs to show optical density at 600nm versus time for *S. cerevisiae* treated continuously with 0 $\mu$ M (◆), 1 $\mu$ M (■), 3 $\mu$ M (▲), 10 $\mu$ M (×), 30 $\mu$ M (+) and 100 $\mu$ M (●) (a) distamycin and (b) TIP-C7-amino over 24 hours. Cell cultures were treated with a range of concentrations of distamycin and the optical densities of the treated cultures were measured as described in section 5.3. Results are the mean of at least three independent experiments with error bars showing the standard deviation of the mean over 24 hours.

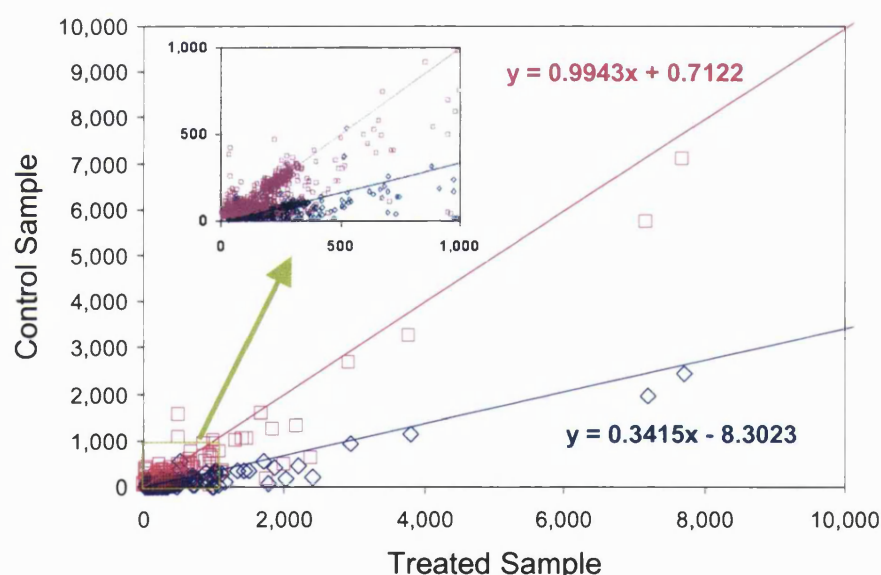
This difference in anti-proliferative activity between the two agents became more pronounced when viewed over time. Cells treated continuously with doses of 30 $\mu$ M and 100 $\mu$ M distamycin over 24 hours exhibited only 60% and 6%, respectively, of control growth (Figure 5.6a). In contrast, no significant anti-proliferative effect was observed with TIP-C7-amino even after 24 hours of continuous exposure to a dose of 100 $\mu$ M (Figure 5.6b). *S. cerevisiae* cells treated with either distamycin or TIP-C7-amino were counterstained at 24 hours with FUN1 and Calcofluor White M2R to differentiate between an inhibition in cell growth and a loss of cell viability (Figure 5.7). This showed that for doses of up to 10 $\mu$ M distamycin, the cells remained viable. At 30 $\mu$ M and 100 $\mu$ M, however, there was a sharp drop in viability to 25% and 0%, respectively. In contrast, cells remained viable at all doses of TIP-C7-amino.



**Figure 5.7** Graph to show cell viability versus concentration for TIP-C7-amino (■) and distamycin (■) at 0 $\mu$ M, 3 $\mu$ M, 10 $\mu$ M, 30 $\mu$ M and 100 $\mu$ M after 24 hours. Cells were treated with increasing concentrations of either TIP-C7-amino or distamycin, and at 24 hours were counterstained for 30 minutes with the yeast cell wall stain Calcofluor White M2R and the vacuolar stain FUN1, harvested and viewed using an epifluorescence microscope under phase contrast and in the blue and red channels as described in the methods section. Percentage cell viability was determined by counting the number of cells with stained vacuoles in the red channel in 3 fields of view and expressing it as a percentage of the total cell number in that field as determined in the blue channel using by counting the number of cells with stained cell walls.

#### 5.4.2 Response in mRNA expression levels on treatment of *S. cerevisiae* with distamycin or TIP-C7-amino

The effect of treating *S. cerevisiae* with either distamycin or TIP-C7-amino was evaluated by determining the expression levels of all 6144 yeast genes relative to an untreated population of cells for either three parallel (same cell population) or three serial (different cell populations) data sets. For each experiment to be directly comparable, with both the control and further biological repeats, each set of intensities measured had to be background corrected and normalised. This was carried out as a single step by linear regression, whereby the raw intensities of the treated sample were plotted against those of the control. The values derived for the intercept and gradient were then used to correct and normalise the data, as described in section 5.3.2. A typical scatter plot showing background correction and normalisation of the data for a single comparison is shown below in Figure 5.8.



**Figure 5.8** A typical scatter plot showing the linear relationship between the raw intensities of the control and treated samples before and after normalisation, where (◇) represents the raw data and (□) represents the normalised data. Solid lines represent the regression model before (-) and after (-) normalisation. The insert indicated by the green box shows a magnified section of the graph in the range of 0-1000.

The equations shown in this figure describe the linear relationship between the intensities of the control and treated samples before and after background correction and normalisation. These processes ensure that equivalent intensities in either the treated or control sample are the same. In this case the equation describing this linear relationship is  $y=x$  and on normalisation has a gradient equal to 1 and an intercept of 0. Resetting the intercept to 0 ensures that an intensity of 0 in the control sample corresponds to an intensity of 0 in the treated sample. A gradient equal to 1 sets the average ratio of intensities in the treated and control spots to be equal to 1 and signifies no change in gene expression. Consequently, ratios greater than 1 represent the overexpression of a given gene and ratios less than 1 indicate underexpression.

Two different experiment designs were compared in an effort to assess the variation introduced into the data as a result of the use of different cell populations. Thus samples of *S. cerevisiae* were treated with either 100 $\mu$ M distamycin or 100 $\mu$ M TIP-C7-amino either in parallel, using the same cell population, or in series, using different cell populations. For 100 $\mu$ M distamycin, 1233 and 244 genes were shown to have significant changes in expression at the  $p=0.05$  level for the parallel and serial treatments, respectively. For 100 $\mu$ M TIP-C7-amino, 848 and 283 genes were shown to have significant changes in their levels of expression ( $p=0.05$ ) for the parallel and serial treatments, respectively.

The top twenty ORFs of a total of 1233 genes derived from the three of 100 $\mu$ M distamycin parallel treatments are listed in ascending order of p value in Table 5.1. Of these ORFs, only fifteen were found to be characterised genes and these were predominantly upregulated, with only three genes showing downregulation. The changes in the ORF expression levels relative to the control are observed to range from -2.64 to +3.10. The molecular and biological functions of the protein products of these characterised genes are listed in Table 5.2. This shows that most of these include protein products involved in a wide range of unrelated biological processes. Five of these, however, appear to be involved in DNA or RNA processing events.

ORF NAME	GENE NAME	Stdev	Mean (Ratio)	Mean (Log <sub>2</sub> )	Fold Change	t	p
YDR070C	FMP16	0.01	0.48	-1.05	-2.08	128.00	<0.005
YDL096C	-	0.02	0.42	-1.24	-2.37	116.57	<0.005
YDR085C	AFR1	0.02	0.38	-1.40	-2.64	100.23	<0.005
YPR093C	-	0.01	0.71	-0.50	-1.41	93.82	<0.005
YHL024W	RIM4	0.03	2.90	1.54	2.90	81.98	<0.005
YML018C	-	0.01	0.64	-0.64	-1.56	74.55	<0.005
YDL225W	SHS1	0.02	1.70	0.76	1.70	70.89	<0.005
YEL071W	DLD3	0.04	3.10	1.63	3.10	66.94	<0.005
YDR233C	RTN1	0.03	1.96	0.97	1.96	56.01	<0.005
YGR123C	PPT1	0.02	1.65	0.72	1.65	55.38	<0.005
YIL036W	CST6	0.05	0.40	-1.33	-2.51	50.83	<0.005
YJR140C	HIR3	0.02	1.48	0.57	1.48	50.63	<0.005
YCL059C	KRR1	0.04	2.19	1.13	2.19	48.25	<0.005
YPL089C	RLM1	0.04	2.01	1.01	2.01	47.47	<0.005
YDL226C	GCS1	0.03	1.71	0.77	1.71	46.42	<0.005
YBR112C	CYC8	0.03	1.75	0.81	1.75	46.25	<0.005
YGR126W	-	0.04	0.52	-0.95	-1.93	45.95	<0.005
YER137C	-	0.03	0.60	-0.75	-1.68	44.88	<0.005
YLR043C	TRX1	0.02	1.55	0.63	1.55	44.23	<0.005
YIL142W	CCT2	0.05	2.43	1.28	2.43	44.04	<0.005

**Table 5.1** Top twenty ORFs listed in ascending order of p value derived from three samples treated in parallel with 100μM distamycin, where stdev is the standard deviation between the log<sub>2</sub> transformed intensity ratios for each ORF across three samples, the mean ratio is the intensity of the treated sample over the control, the mean log<sub>2</sub> value is the log<sub>2</sub> transformed intensity ratios, fold change is the change in expression of the treated sample over the control, t is the t-test statistic and p is the probability that there is no difference in gene expression between the control and treated samples. ORFs that do not represent characterised genes are denoted with a hyphen (-).

<b>GENE NAME</b>	<b>MOLECULAR FUNCTION</b>	<b>BIOLOGICAL FUNCTION</b>	<b>CELLULAR COMPONENT</b>	<b>FOLD CHANGE</b>
FMP16	-	-	Mitochondrion	-2.08
AFR1	Receptor signalling protein activity.	Signal transduction during conjugation.	Shmoo tip	-2.64
RIM4	RNA binding.	Premeiotic DNA synthesis.	Cytoplasm	2.90
SHS1	Structural constituent of cytoskeleton.	Cell morphogenesis.	Septin ring	1.70
DLD3	D-lactate dehydrogenase (cytochrome) activity.	Lactate metabolism.	Cytoplasm	3.10
RTN1	-	-	ER	1.96
PPT1	Protein serine/threonine phosphatase activity.	Protein amino acid phosphorylation.	Cytoplasm Nucleus	1.65
CST6	RNA polymerase II transcription factor activity.	DNA metabolism, transcription initiation from Pol II promoter.	Nucleus	-2.51
HIR3	Transcription corepressor activity.	G1/S specific transcription in mitotic cell cycle.	Nucleus	1.48
KRR1	-	rRNA processing, ribosome biogenesis.	Nucleolus	2.19
RLM1	DNA bending activity, DNA binding, transcriptional activator activity.	Cell wall organisation and biogenesis, positive regulation of transcription from Pol II promoter, signal transduction.	Nucleus	2.01
GCS1	ARF GTPase activator activity, actin binding.	ER to Golgi transport.	ER-Golgi intermediate compartment.	1.71
CYC8	General transcriptional repressor activity, transcription coactivator activity.	Negative regulation of transcription.	Nucleus	1.75
TRX1	Thiol-disulphide exchange intermediate activity.	DNA dependent DNA replication, response to oxidative stress.	Cytosol	1.55
CCT2	Chaperone activity	Protein folding, cytoskeleton organisation and biogenesis.	Cytoplasm Cytoskeleton	2.43

**Table 5.2** Molecular and biological functions of the protein products of characterised genes derived from the top twenty ORF list from samples treated in parallel with 100µM distamycin. Unknown function is denoted with a hyphen (-).



The top twenty ORFs of a total of 244 genes from the three serial treatments of 100 $\mu$ M distamycin are listed in ascending order of p value in Table 5.3. Thirteen of these ORFs were found to represent characterised genes. However, the majority of these were not shared with the top twenty list of ORFs for the corresponding parallel treatment. Only GCS1 and SHS1 appeared in both gene sets. Remarkably these showed very similar levels of expression, with levels in GCS1 changing by +1.71 and +1.72 in the parallel and serial repeats respectively, whilst SHS1 changed by +1.70 and +1.61, respectively. The molecular and biological functions of the characterised genes in the serial treatment are listed in Table 5.4. Once more, the majority of protein products are seen to be associated with a wide range of biological functions. Only four of the eighteen genes, RPB11, NUP42, RCL1 and SMP1, were found to be related to DNA or RNA processing events in this case.

Only the changes in expression of 66 ORFs were found to be common to both serial and parallel designs. These are listed in Table 5.5 and have p values ranging from 0.01 to 0.05. Of these ORFs, only 41 were found to be characterised genes and approximately 90% of these were upregulated. The protein products of these genes were then grouped into common biological functions and are listed as Tables 5.6-5.11. The largest group was that of protein products related to DNA and RNA processing events, with 27% of the protein products related to this function (Table 5.8). The next largest group was protein products associated with protein processing events, with almost 20% related to this function (Table 5.6). Remaining protein products were found to be associated with transport (7%, Table 5.7), protein folding and stress (4.5%, Table 5.9), the structure and organisation of the cytoskeleton (7%, Table 5.10) and lipid metabolism (7%, Table 5.11).

ORF NAME	GENE NAME	Stdev	Mean (Ratio)	Mean (Log <sub>2</sub> )	Fold Change	t	p
YDL226C	GCS1	0.02	1.72	0.78	1.72	64.10	<0.005
YAR027W	UIP3	0.01	0.83	-0.27	-1.20	56.11	<0.005
YOL005C	RPB11	0.01	1.22	0.29	1.22	55.24	<0.005
YCLX08C	FRM2	0.03	1.78	0.83	1.78	49.49	<0.005
YKR068C	BET3	0.05	2.46	1.30	2.46	41.23	<0.005
YPL088W	-	0.06	0.46	-1.12	-2.18	31.62	<0.005
YJL146W	IDS2	0.02	1.34	0.42	1.34	31.54	<0.005
YDR192C	NUP42	0.03	1.49	0.58	1.49	31.49	<0.005
YDR166C	SEC5	0.01	1.18	0.24	1.18	30.94	<0.005
YOL010W	RCL1	0.04	1.53	0.61	1.53	28.83	<0.005
YGR221C	TOS2	0.04	1.47	0.55	1.47	26.73	<0.005
YKL091C	-	0.03	1.31	0.39	1.31	26.35	<0.005
YBL102W	SFT2	0.02	1.24	0.31	1.24	25.98	<0.005
YGL083W	SCY1	0.04	1.47	0.55	1.47	24.42	<0.005
YDL225W	SHS1	0.07	1.61	0.69	1.61	18.23	<0.005
YBR182C	SMP1	0.05	1.43	0.51	1.43	17.59	<0.005
YBR161W	CSH1	0.09	1.84	0.88	1.84	17.09	<0.005
YDR188W	CCT6	0.05	1.41	0.49	1.41	16.54	<0.005
YML070W	DAK1	0.08	1.69	0.76	1.69	16.42	<0.005
YGR245C	SDA1	0.03	1.20	0.26	1.20	15.41	<0.005

**Table 5.3** Top twenty ORFs listed in ascending order of p value derived from three samples treated in series with 100μM distamycin, where stdev denotes the standard deviation between the log<sub>2</sub> transformed intensity ratios for each ORF for the three samples, the mean ratio is the intensity of the treated sample over the control, the mean log<sub>2</sub> value is the log<sub>2</sub> transformed intensity ratios, fold change is the change in expression of the treated sample over the control, t is the t-test statistic and p the probability that there is no difference in gene expression between the control and treated samples. ORFs that do not represent characterised genes are denoted with a hyphen (-).

<b>GENE NAME</b>	<b>MOLECULAR FUNCTION</b>	<b>BIOLOGICAL FUNCTION</b>	<b>CELLULAR COMPONENT</b>	<b>FOLD CHANGE</b>
GCS1	ARF GTPase activator activity, actin binding.	ER to Golgi transport.	ER-Golgi	1.72
UIP3	-	-	Nuclear membrane	-1.20
RPB11	DNA-directed RNA polymerase activity.	Transcription from the Pol II promoter.	DNA-directed RNA polymerase II, core complex.	1.22
FRM2	-	Negative regulation of fatty acid metabolism.	Cytoplasm Nucleus	1.78
BET3	-	ER to Golgi transport	TRAPP	2.46
IDS2	-	Meiosis	Cytoplasm Nucleus	1.34
NUP42	Structural molecule activity.	mRNA nucleus import, protein nucleus import, rRNA nucleus import, ribosomal nucleus import, tRNA nucleus import.	Nuclear pore	1.49
SEC5	-	Golgi to plasma membrane transport, cytokinesis, vesicle docking and fusion, polar budding.	Actin cap Exocist	1.18
RCL1	-	rRNA processing.	Nucleolus	1.53
TOS2	-	-	-	1.47
SFT2	-	Golgi to endosome transport.	Golgi membrane	1.24
SCY1	-	-	Clathrin-coated vesicle	1.47
SHS1	Structural constituent of cytoskeleton.	Cellular morphogenesis, cytokinesis, cell polarity.	Septin ring	1.61
SMP1	DNA bending activity, transcription factor activity.	Positive regulation of transcription form Pol II promoter.	Cytoplasm Nucleus	1.43
CSH1	Mannosyltransferase activity.	Sphingolipid biosynthesis.	Vacuole	1.84
CCT6	Chaperone activity.	Cytoskeleton organisation and biogenesis, proetin folding.	Chaperone-containing T complex Cytoplasm Cytoskeleton	1.41
DAK1	Glycerone kinase activity	Glycerol catabolism, response to stress.	Cytoplasm	1.69
SDA1	-	Actin cytoskeleton organisation and biosynthesis, start control point of mitotic cell cycle.	Nucleus	1.20

**Table 5.4** Molecular and biological functions of the protein products of characterised genes derived from the top twenty ORF list from samples treated in series with 100µM distamycin. Unknown function is denoted with a hyphen (-).

ORF NAME	GENE NAME	PARALLEL		SERIAL	
		Fold Change	p	Fold Change	p
YDL147W	RPN5	2.33	<0.01	2.81	0.02
YDR099W	BMH2	1.61	<0.01	1.50	0.03
YBR106W	PHO88	1.34	0.01	1.49	0.04
YBR155W	CNS1	1.40	0.05	1.66	0.01
YDR172W	SUP35	2.12	0.02	1.69	0.02
YAL003W	EFB1	1.40	0.01	1.30	0.01
YDR165W	TRM82	1.68	0.04	1.66	0.03
YDR228C	PCF11	1.48	0.03	1.40	0.05
YBR015C	TTP1/MNN2	1.63	0.04	1.48	0.01
YBR165W	UBS1	1.43	0.03	1.69	0.01
YAL022C	FUN26	1.46	<0.01	1.61	0.01
YBR159W	-	1.64	0.03	1.77	0.01
YBR166C	TYR1	1.86	0.04	1.90	0.03
YDR144C	MKC7	1.50	0.01	1.51	0.04
YBL077W	-	2.14	0.02	1.77	0.02
YBR160W	CDC28	1.59	0.05	1.99	0.02
YDL225W	SHS1	1.70	<0.01	1.61	<0.01
YCR013C	-	1.99	0.02	1.98	0.01
YBR026C	MRP1/ETR1	1.61	0.01	1.77	0.01
YDL161W	ENT1	1.80	0.01	1.69	0.03
YDL226C	GCS1	1.71	<0.01	1.72	<0.01
YHR156C	LIN1	1.32	0.03	1.33	0.02
YHR188C	GPI16	1.55	0.02	1.51	0.04
YJL026W	RNR2	1.60	<0.01	1.46	0.04
YGL173C	KEM1	1.57	0.01	1.35	0.05
YGL013C	PDR1	1.56	0.01	1.32	0.05
YEL036C	ANP1	1.84	<0.01	1.53	0.02
YFL040W	-	-1.32	<0.01	-1.40	0.05
YER071C	-	-1.60	0.01	-1.59	0.01
YIL075C	SEN3/RPN2	2.86	0.01	2.33	0.01
YHR084W	STE12	-1.58	0.01	-1.48	0.04
YER103W	SSA4	1.73	0.02	1.58	0.01
YFR051C	RET2	1.98	0.03	1.78	0.04
YGR123C	PPT1	1.65	<0.01	1.86	0.05
YGL202W	ARO8	1.71	0.01	1.49	0.01
YEL047C	-	1.52	0.03	1.91	0.04
YHR039C	MSC7	1.42	0.04	1.75	0.01
YHR054C	-	2.31	<0.01	1.92	0.01
YGL122C	NAB2	1.79	0.01	1.73	0.03
YIL053W	RHR2	2.06	0.02	1.99	0.04
YFR005C	SAD1	1.56	0.02	1.72	0.01
YKR068C	BET3	1.97	0.03	2.46	<0.01
YLL026W	HSP104	1.80	0.02	2.63	0.01
YML085C	TUB1	1.39	0.01	1.60	0.04
YLR056W	ERG3	2.06	<0.01	1.89	0.02
YJL123C	-	1.83	0.03	1.61	0.04
YML070W	DAK1	1.66	<0.01	1.69	<0.01
YML094W	GIM5	1.36	0.01	1.37	0.04
YKL214C	YRA2	1.25	0.01	1.31	0.01
YLR209C	PNP1	1.38	0.03	1.40	0.03
YML072C	-	2.38	0.05	1.61	0.05
YMR039C	SUB1	1.31	0.04	1.41	<0.01
YMR108W	ILV2	1.73	0.01	1.65	0.03
YKL212W	SAC1	1.83	0.02	1.87	0.02
YOL015W	-	1.49	0.01	1.82	0.02
YOR281C	PLP2	1.20	0.03	1.36	0.05
YPL014W	-	-1.69	0.03	-1.34	0.04
YPL063W	TIM50	1.52	0.03	1.64	0.04
YMR290C	HAS2	1.32	0.03	1.40	0.01
YOR202W	HIS3	-1.28	0.01	-1.28	0.02
YPR118W	-	1.36	0.04	1.32	0.05
YOL019W	-	-1.17	0.04	-1.30	0.03
YPL050C	MNN9	2.01	0.03	1.49	0.02
YOL155C	-	-1.81	0.03	-2.55	0.01
YPR127W	-	2.08	<0.01	2.42	0.02
YOL005C	RPB11	1.27	0.01	1.22	<0.01

**Table 5.5**  
Complete list of ORFs that are common to both parallel and serial treatments of *S.cerevisiae*. with 100µM distamycin. ORFs that do not represent characterised genes are denoted with a hyphen (-).

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
RPN5	Endopeptidase activity	Ubiquitin dependent protein catabolism.	Proteasome regulatory particle.	2.33	2.81
MNN2	Alpha-1,2-mannosyltransferase activity.	Protein amino acid glycosylation.	Golgi apparatus	1.63	1.48
UBS1	-	Protein ubiquitination, protein-nucleus export.	Nucleus	1.43	1.69
TYR1	Prephenate dehydrogenase activity.	Tyrosine metabolism.	Cytoplasm	1.86	1.90
MKC7	Aspartic-type signal peptidase activity.	Proteolysis and peptidolysis.	Cell wall	1.50	1.51
GPI16	GPI-anchor transamidase activity.	Attachment of GPI anchor to protein.	Integral to endoplasmic reticulum membrane.	1.55	1.51
ANP1 /MNN8	Mannosyltransferase activity.	N-linked glycosylation.	Golgi cis cisterna	1.84	1.53
SEN3 /RPN2	Endopeptidase activity, receptor activity.	Ubiquitin dependent protein catabolism.	Proteasome regulatory particle.	2.86	2.33
PPT1	Protein serine/threonine phosphatase activity.	Protein amino acid phosphorylation.	Cytoplasm Nucleus	1.65	1.86
ARO8	Aromatic amino acid transaminase activity.	Aromatic amino acid family metabolism.	Cytoplasm	1.71	1.49
ILV2	Acetolactate synthase activity.	Branched chain family amino acid biosynthesis.	Mitochondrion	1.73	1.65
HIS3	Imidazoleglycerol-phosphate dehydratase activity.	Histidine biosynthesis.	Cytoplasm	-1.28	-1.28
MNN9	Mannosyltransferase activity.	N-linked glycosylation.	Golgi cis face, Mannosyltransferase complex, Membrane	2.01	1.49

**Table 5.6** The protein products of characterised genes appearing in both parallel and serial 100µM distamycin treatments which are associated with protein processing and metabolism. Unknown function is denoted with a hyphen (-).

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
PHO88	Phosphate transporter activity	Phosphate transport.	Membrane	1.34	1.49
FUN26	Nucleoside transporter activity.	Nucleoside transport.	Membrane	1.46	1.61
RET2	Protein binding	ER to Golgi transport	COPI vesicle coat	1.98	1.78
BET3	-	ER to Golgi transport	TRAPP	1.97	2.46
TIM50	-	Mitochondrial matrix protein import.	Mitochondrial inner membrane .	1.52	1.64

**Table 5.7** The protein products of characterised genes appearing in both parallel and serial 100µM distamycin treatments which are associated with intracellular transport. Unknown function is denoted with a hyphen (-).

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
BMH2	DNA binding, Protein binding	RAS protein signal transduction, glycogen metabolism, signal transduction during filamentous growth.	Nucleus	1.61	1.50
SUP35	Translation release factor activity.	mRNA catabolism, translational termination.	Cytosol	2.12	1.69
EFB1	Translation elongation factor activity.	Translation elongation.	Ribosome	1.40	1.30
TRM82	Protein binding, tRNA (guanine-N7-)methyl transferase activity.	tRNA methylation.	Nucleus	1.68	1.66
PCF11	Cleavage/polyadenylation specificity factor activity, protein binding.	mRNA cleavage, mRNA polyadenylation, transcription termination from Pol II promoter.	mRNA cleavage factor complex.	1.48	1.40
CDC28	Cyclin-dependent protein kinase activity.	Regulation of cell cycle.	Cytoplasm Nucleus	1.59	1.99
LIN1	Protein binding.	-	Chromatin Nucleus	1.32	1.33
RNR2	Ribonucleoside-diphosphate reductase activity.	DNA replication.	Cytoplasm Nucleus	1.60	1.46
KEM1	5'-3' exoribonuclease activity, deoxyribonuclease activity, recombinase activity.	35S primary transcript processing, filamentous growth, mRNA catabolism.	Cytoplasm	1.57	1.35
PDR1	DNA binding, transcriptional activator activity.	Regulation of transcription from Pol II promoter, response to drug.	Nucleus	1.56	1.32
STE12	Transcription factor activity.	Conjugation with cellular fusion, positive regulation of transcription from Pol II promoter by pheromones.	Nucleus	-1.58	-1.48
NAB2	Poly(A) binding	mRNA polyadenylation, poly A and mRNA-nucleus export.	Cytoplasm Nucleus	1.79	1.73
SAD1	Pre-mRNA splicing factor activity.	Nuclear mRNA splicing via spliceosome.	Nucleus	1.56	1.72
YRA2	RNA binding	Poly(A)+mRNA-nucleus export.	Nucleus	1.25	1.31
PNP1	Purine-nucleoside phosphorylase activity	Purine nucleoside catabolism.	-	1.38	1.40
SUB1	Transcriptional coactivator activity.	Positive regulation of transcription from Pol II promoter, transcription termination from Pol II promoter.	Nucleus	1.31	1.41
HAS1	Putative ATP-dependent RNA helicase.	-	Nuclear membrane Nucleolus	1.32	1.40
RPB11	DNA-directed RNA polymerase activity.	Transcription from Pol II promoter.	DNA-directed RNA pol II complex.	1.27	1.22

**Table 5.8** The protein products of characterised genes appearing in both parallel and serial 100µM distamycin treatments which are associated with DNA/RNA processing.

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
CNS1	Chaperone activity	Protein folding	Cytoplasm	1.40	1.66
SSA4	Chaperone activity, heat shock protein activity.	Protein folding, response to stress.	Cytoplasm Nucleus	1.73	1.53
HSP104	ATPase activity, co-chaperone activity, heat shock protein activity.	Protein folding, response to stress.	Cytoplasm Nucleus	1.80	2.63

**Table 5.9** The protein products of characterised genes appearing in both parallel and serial 100µM distamycin treatments which are associated with protein folding and stress responses.

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
SHS1	Structural constituent of cytoskeleton.	Cellular morphogenesis, cytokinesis.	Septin ring	1.70	1.61
ENT1	Clathrin binding	Actin cortical patch assembly, actin filament organisation, endocytosis.	Actin cortical patch	1.80	1.69
GCS1	ARF GTPase activator activity, actin binding.	ER to Golgi transport, cell cycle dependent actin filament reorganisation.	ER-Golgi intermediate compartment.	1.71	1.72
TUB1	Structural constituent of the cytoskeleton.	Chromosome segregation.	Cytoplasmic, kinetochore, nuclear, polar microtubules and spindle pole body.	1.39	1.60
GIM5	Tubulin binding	Tubulin folding	Cytoplasm, prefoldin complex.	1.36	1.37

**Table 5.10** The protein products of characterised genes appearing in both parallel and serial 100µM distamycin treatments which are associated with the structure and organisation of the cytoskeleton.

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
ETR1	Enoyl-[acyl]-carrier protein] reductase activity.	Aerobic respiration, Fatty acid biosynthesis.	Mitochondrion	1.61	1.77
RHR2	Glycerol-1-phosphate activity.	Glycerol biosynthesis, response to osmotic stress.	Cytoplasm Nucleus	2.06	1.99
ERG3	C-5 sterol desaturase activity.	Ergosterol biosynthesis.	Endoplasmic reticulum	2.06	1.89
DAK1	Glycerone kinase activity	Glycerol catabolism, response to stress.	Cytoplasm	1.66	1.69
SAC1	Inositol/phosphatidylinositol phosphatase activity.	Dephosphorylation, phosphoinositide metabolism.	Integral to Golgi membrane.	1.83	1.87

**Table 5.11** The protein products of characterised genes appearing in both parallel and serial 100µM distamycin treatments which are associated with lipid metabolism and stress responses.



The top twenty ORFs of a total number of 848 derived from three samples treated in parallel with the polyamide TIP-C7-amino are listed in ascending order of p value in Table 5.12. In contrast to distamycin, these were predominantly downregulated. Fourteen of the identified ORFs corresponded to characterised genes and approximately 70% of these were downregulated. The molecular and biological functions of the protein products of the genes identified are listed in Table 5.13. This shows that they belong to a wide range of apparently unrelated functions.

The top twenty ORFs of a total ORF number of 283 derived from the corresponding set of serial treatments are listed in Table 5.14. Here, fifteen of the identified top twenty corresponded to characterised genes but in contrast to those derived from the parallel treatment, the majority of these were upregulated. Only two genes, PRE2 and KAR1 were downregulated. None of the genes identified in the top twenty of either the parallel or serial data sets were common to both lists. As with the parallel treatment, identification of the molecular and biological functions of the products of the genes included in the top twenty derived from the serial treatments demonstrated that they belonged to a wide range of unrelated functions. However four of these, MIG2, MIP16, RPB4 and MED7, did appear to be associated with DNA or RNA processing.

When the complete data sets for parallel and serial treatments were compared, only 38 ORFs were found to be common to both and are listed in Table 5.16. The majority of these, 65%, were downregulated and their p values ranged from <0.01 to 0.05. 25 of the ORFs represented characterised genes, with just over half downregulated. When the molecular and biological functions of gene products were identified, once more they were found to belong to a wide range of functions (Tables 5.17-5.24). As with distamycin, the largest group were related to DNA/RNA processing events, Table 5.18. These constituted 36% of the genes common to the serial and parallel data sets. The next largest groups were related to transport (16%, Table 5.17) and glucose metabolism (16%, Table 5.19). Remaining groups were cytoskeletal organisation (12%, Table 5.20), lipid metabolism (4%, Table 5.21), protein folding (4%, Table 5.22), ribosome structure and assembly (8%, Table 5.23) and sexual conjugation (4%, Table 5.24).

ORF NAME	GENE NAME	Stdev	Mean (Ratio)	Mean (Log <sub>2</sub> )	Fold Change	t	p
YER049W	-	0.00	1.63	0.70	1.63	1505.08	<0.005
YPL224C	MMT2	0.00	0.63	-0.66	-1.58	304.05	<0.005
YBR009C	HHF1	0.01	0.70	-0.51	-1.42	102.01	<0.005
YPL193W	RSA1	0.01	0.78	-0.36	-1.29	101.55	<0.005
YOL042W	NGL1	0.01	0.74	-0.43	-1.34	85.22	<0.005
YJL060W	BNA3	0.05	0.24	-2.04	-4.1	76.35	<0.005
YML133C	-	0.00	0.96	-0.07	-1.05	72.33	<0.005
YJL101C	GSH1	0.04	3.11	1.64	3.11	72.31	<0.005
YKR024C	DBP7	0.02	1.90	0.92	1.90	69.94	<0.005
YOR381W	FRE3	0.01	0.76	-0.40	-1.32	68.58	<0.005
YAL019W	FUN30	0.01	0.82	-0.28	-1.22	66.86	<0.005
YMR089C	YTA12	0.06	4.45	2.15	4.45	66.51	<0.005
YLR463C	-	0.01	0.84	-0.25	-1.19	60.16	<0.005
YLR250W	SSP120	0.01	0.77	-0.37	-1.29	48.49	<0.005
YMR116C	ASC1	0.01	1.31	0.39	1.31	47.31	<0.005
YLR042C	-	0.02	0.66	-0.60	-1.52	46.33	<0.005
YKL080W	VMA5	0.02	0.67	-0.58	-1.50	45.14	<0.005
YJR037W	-	0.03	0.64	-0.63	-1.55	43.87	<0.005
YMR007W	-	0.02	0.66	-0.61	-1.53	42.85	<0.005
YNL212W	VID27	0.02	0.69	-0.53	-1.44	41.85	<0.005

**Table 5.12** Top twenty ORFs listed in ascending order of p value derived from three samples treated in parallel with 100 $\mu$ M TIP-C7-amino, where stdev denotes the standard deviation between the log<sub>2</sub> transformed intensity ratios for each ORF for the three samples, the mean ratio is the intensity of the treated sample over the control, the mean log<sub>2</sub> value is the log<sub>2</sub> transformed intensity ratios, fold change is the change in expression ratio of the treated sample over the control, t is the t-test statistic and p the probability that there is no difference in gene expression between the control and treated samples. ORFs that do not represent characterised genes are denoted with a hyphen (-).

<b>GENE NAME</b>	<b>MOLECULAR FUNCTION</b>	<b>BIOLOGICAL FUNCTION</b>	<b>CELLULAR COMPONENT</b>	<b>FOLD CHANGE</b>
MMT2	-	iron ion homeostasis	Integral to membrane Mitochondrion	-1.58
HHF1	DNA binding.	Chromatin assembly/disassembly.	Nuclear nucleosome	-1.42
RSA1	Ribosome assembly.	Ribosomal large subunit assembly and maintenance.	Nucleoplasm	-1.29
NGL1	Endonuclease activity.	-	Mitochondrion	-1.34
BNA3	Arylformamidase activity	Nicotinamide adeninedinucleotide biosynthesis.	Cytoplasm Mitochondrion	-4.1
GSH1	Glutamate-cysteine ligase activity.	Glutathione biosynthesis, response to cadmium.	Intracellular	3.1
DBP7	ATP dependent RNA helicase activity.	35S primary transcript processing, ribosomal large subunit assembly and maintenance.	Nucleolus	1.90
FRE3	Ferric-chelate reductase activity.	Iron ion homeostasis, iron-siderochrome transport.	Integral to membrane Plasma membrane	-1.32
FUN30	-	Chromosome organisation and biogenesis.	Nucleus	-1.22
YTA12	ATPase activity, metallopeptidase activity.	Mitochondrial intermembrane space protein import, protein complex assembly, proteolysis and peptidolysis.	Cytoplasm m-AAA complex Mitochondrial inner membrane Mitochondrion	4.45
SSP120	-	Protein secretion.	Cytoplasm	-1.29
ASC1	-	-	Cytoplasm	1.31
VMA5	Hydrogen-transporting ATPase.	Vacuolar acidification.	Hydrogen transporting ATPase VI domain Vacuolar membrane	-1.50
VID27	-	Vacuole import and degradation.	Cytoplasm	-1.44

**Table 5.13** Molecular and biological functions of the protein products of characterised genes derived from the top twenty ORF list from samples treated in parallel with 100µM TIP-C7-amino. Unknown function is denoted with a hyphen (-).

ORF NAME	GENE NAME	Stdev	Mean (Ratio)	Mean (Log <sub>2</sub> )	Fold Change	t	p
YKL117W	SBA1	0.02	1.69	0.76	1.69	76.79	<0.005
YGL209W	MIG2	0.01	1.13	0.18	1.13	36.86	<0.005
YPL159C	PET20	0.03	1.40	0.49	1.40	33.42	<0.005
YPR012W	-	0.02	1.31	0.39	1.31	30.26	<0.005
YFR054C	-	0.02	0.81	-0.31	-1.24	27.85	<0.005
YOL075C	-	0.03	0.72	-0.48	-1.40	24.95	<0.005
YOR306C	MCH5	0.05	1.56	0.64	1.56	23.88	<0.005
YPR103W	PRE2	0.03	0.74	-0.44	-1.35	23.72	<0.005
YHL013C	-	0.04	0.72	-0.48	-1.40	22.24	<0.005
YHR015W	MIP6	0.03	1.30	0.38	1.30	21.98	<0.005
YKL156W	RPS27A	0.04	1.43	0.52	1.43	20.54	<0.005
YJL140W	RPB4	0.01	1.08	0.11	1.08	20.37	<0.005
YOR354C	MSC6	0.03	1.27	0.34	1.27	18.96	<0.005
YNL272C	SEC2	0.05	1.39	0.47	1.39	17.64	<0.005
YOR029W	-	0.02	0.87	-0.20	-1.15	17.36	<0.005
YNL279W	PRM1	0.06	1.50	0.58	1.50	17.03	<0.005
YLR175W	CBF5	0.03	1.24	0.31	1.24	16.36	<0.005
YER182W	FMP10	0.04	1.26	0.33	1.26	16.35	<0.005
YOL135C	MED7	0.05	1.36	0.45	1.36	16.07	<0.005
YNL188W	KAR1	0.05	0.70	-0.50	-1.42	16.06	<0.005

**Table 5.14** Top twenty ORFs listed in ascending order of p value derived from three samples treated in series with 100μM TIP-C7-amino, where stdev denotes the standard deviation between the log<sub>2</sub> transformed intensity ratios for each ORF for the three samples, the mean ratio is the intensity of the treated sample over the control, the mean log<sub>2</sub> value is the log<sub>2</sub> transformed intensity ratios, fold change is the change in expression of the treated sample over the control, t is the t-test statistic and p the probability that there is no difference in gene expression between the control and treated samples. ORFs that do not represent characterised genes are denoted with a hyphen (-).

<b>GENE NAME</b>	<b>MOLECULAR FUNCTION</b>	<b>BIOLOGICAL FUNCTION</b>	<b>CELLULAR COMPONENT</b>	<b>FOLD CHANGE</b>
SBA1	Co-chaperone activity	Protein folding	Cytoplasm Nucleus	1.69
MIG2	Specific RNA polymerase II transcription factor activity.	Glucose metabolism, regulation of transcription from Pol II promoter.	Nucleus	1.13
PET20	-	-	-	1.40
MCH5	Transporter activity.	Transport.	Membrane	1.56
PRE2	Endopeptidase activity.	Ubiquitin dependent protein catabolism.	Proteasome core complex.	-1.35
MIP6	RNA binding.	mRNA-nucleus transport.	Nuclear pore	1.30
RPS27A	Structural constituent of ribosome.	Protein biosynthesis.	Cytoplasm Cytosolic small ribosomal subunit	1.43
RPB4	DNA-directed RNA polymerase activity.	Transcription from the Pol II promoter.	DNA-directed RNA polymerase II core complex	1.08
MSC6	-	Meiotic recombination.	Mitochondrion	1.27
SEC2	Guanyl-nucleotide exchange factor activity.	Exocytosis.	Cytoplasmic vesicle	1.39
PRM1	-	Plasma membrane fusion.	Integral to membrane Shmoo tip	1.50
CBF5	Pseudouridylylate synthase activity.	35S primary transcript processing, rRNA modification.	Small nucleolar ribonucleoprotein complex.	1.24
FMP10	-	-	Mitochondrion	1.26
MED7	RNA polymerase II transcription mediator activity.	Transcription form Pol II promoter.	Mediator complex	1.36
KAR1	Protein binding.	Karyogamy during conjugation with cellular fusion.	Half bridge spindle body.	-1.42

**Table 5.15** Molecular and biological functions of the protein products of characterised genes derived from the top twenty ORF list from samples treated in series with 100µM TIP-C7-amino. Unknown function is denoted with a hyphen (-).

ORF NAME	GENE NAME	PARALLEL		SERIAL	
		Fold Change	p	Fold Change	p
YDR164C	SEC1	-1.45	0.02	-1.26	0.04
YDR387C	-	-1.52	0.01	-1.48	0.04
YBR195C	MSI1	-1.80	0.04	-1.39	0.03
YDR389W	SAC7	-1.42	0.01	-1.44	0.05
YDR191W	HST4	-1.87	0.02	-1.52	0.05
YBL094C	-	-1.25	0.03	-1.37	0.05
YDR297W	SUR2	1.24	0.01	1.33	0.02
YEL002C	WBP1	-1.22	0.01	-1.19	0.03
YIL072W	HOP1	-1.41	0.05	-1.38	0.04
YFR016C	-	-1.31	0.02	-1.26	0.04
YGR256W	GND2	-1.17	0.04	-1.24	0.03
YKL117W	SBA1	1.58	0.01	1.69	<0.01
YJL138C	TIF2	1.37	0.05	1.71	0.05
YML100W	TSL1	1.20	0.03	1.31	0.02
YMR031W-A	-	-1.49	0.01	-1.49	0.03
YJL063C	MRPL8	-1.19	0.02	-1.20	0.01
YJR047C	ANB1	-1.23	0.04	-1.17	<0.01
YJR009C	TDH2	1.46	0.02	1.40	0.03
YNL279W	PRM1	1.29	0.05	1.50	0.00
YNL135C	FPR1	1.43	<0.01	1.91	0.05
YOL112W	MSB4	1.59	0.04	1.34	0.04
YMR290C	HAS1	1.32	0.03	1.46	0.03
YOR170W	ESC8	1.35	0.05	1.43	0.02
YOL017W	-	-1.33	0.02	-1.33	0.03
YNL321W	-	1.54	<0.01	1.47	0.02
YPR077C	TUF1	1.43	0.02	1.42	0.04
YOR187W	-	1.20	0.01	1.21	0.03
YPR014C	-	-1.36	0.01	-1.32	0.02
YPL017C	-	-1.21	0.03	-1.18	0.03
YNL275W	-	-1.37	0.02	-1.40	0.04
YNL283C	WSC2	-1.63	0.04	-1.35	0.05
YOL075C	-	-1.57	0.02	-1.40	0.00
YOL115W	TRF4	-1.41	0.05	-1.25	0.02
YNR056C	BIO5	-1.60	0.03	-1.36	0.02
YOR029W	-	-1.12	0.03	-1.15	<0.01
YOR070C	GYP1	-1.52	0.03	-1.44	0.00
YOR094W	ARF3	-1.34	0.05	-1.15	0.04
YOR365C	-	-1.96	0.03	-1.85	0.01

**Table 5.16** Complete list of ORFs that are common to both parallel and serial treatments of *S.cerevisiae* with 100μM TIP-C7-amino. ORFs that do not represent characterised genes are denoted with a hyphen (-).

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
SEC1	SNARE binding	Exocytosis, non-selective vesicle docking /fusion.	Actin cap	-1.45	-1.26
BIO5	Permease activity	Biotin biosynthesis, vitamin cofactor transport.	Plasma membrane.	-1.60	-1.36
GYP1	Rab GTPase activator activity.	Vesicle mediated transport.	Golgi apparatus.	-1.52	-1.44
ARF3	ARF small monomeric GTPase activity.	Intracellular protein transport.	-	-1.34	-1.15

**Table 5.17** Protein products of characterised genes appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with transport. Unknown function or localisation is denoted with a hyphen (-).

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
MSI1	-	DNA repair, RAS protein signal transduction, chromatin silencing, nucleosome assembly.	Chromatin assembly complex Chromatin silencing complex	-1.80	-1.39
HST4	DNA binding.	Chromatin silencing at telomere, short chain fatty acid metabolism.	Cytoplasm Nucleus	-1.87	-1.52
HOP1	DNA binding.	Meiosis, synaptonemal complex formation.	Condensed nuclear chromosome Synaptonemal complex.	-1.41	-1.38
TIF2	RNA helicase activity, translation initiation factor activity.	Regulation of translational initiation.	Cytoplasm Ribosome	1.37	1.71
ANB1	Translation initiation factor activity.	Translation initiation.	Ribosome.	-1.23	-1.17
HAS1	Putative ATP-dependent RNA helicase.	-	Nuclear membrane Nucleolus	1.32	1.46
ESC8	-	Chromatin silencing.	Nucleus	1.35	1.43
TUF1	Translational elongation factor activity.	Translational elongation.	Mitochondrial matrix.	1.43	1.42
TRF4	DNA-directed DNA polymerase activity.	DNA topological change, mitotic chromosome condensation, sister chromatid cohesion.	Nucleus	-1.41	-1.25

**Table 5.18** Protein products of characterised genes appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with DNA/RNA processing and metabolism. . Unknown function is denoted with a hyphen (-).



GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
SAC7	Rho GTPase activator activity, signal transducer activity.	Cell cycle dependent actin filament reorganisation, small GTPase mediated signal transduction.	Intracellular	-1.42	-1.44
MSB4	Rab GTPase activator activity.	Actin filament organisation.	Bud tip Incipient bud site	1.59	1.34
WSC2	Transmembrane receptor activity.	Rho protein signal transduction, actin cytoskeleton organisation and biogenesis, response to heat.	Cytoplasm Membrane fraction.	-1.63	-1.35

**Table 5.19** The protein products of characterised genes appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with cytoskeletal organisation.

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
WBP1	Dolichyl-diphosphooligosaccharide-protein glycotransferase activity.	N-linked glycosylation, Cell cycle.	ER Nuclear membrane Oligosaccharyl transferase complex.	-1.22	-1.19
GND2	Phosphogluconate dehydrogenase activity.	Glucose metabolism.	Superpathway for gluconate utilisation, pentose phosphate pathway, RuMP cycle and formaldehyde assimilation.	-1.17	-1.24
TSL1	Enzyme regulator activity.	Response to stress, trehalose biosynthesis.	Alpha, alpha-trehalose-phosphate sythase complex (UDP-forming).	1.20	1.31
TDH2	Glyceraldehyde-3-phosphate dehydrogenase activity.	Gluconeogenesis, glycolysis.	Cell Wall Cytoplasm Cytosol Lipid particle	1.46	1.40

**Table 5.20** The protein products of characterised genes appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with glucose metabolism.

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
SUR2	Sphingosine hydroxylase activity.	Sphingolipid biosynthesis and metabolism.	ER	1.24	1.33

**Table 5.21** The protein product of a characterised gene appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with lipid metabolism.

GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
SBA1	Co-chaperone activity.	Protein folding	Cytoplasm Nucleus	1.58	1.69

**Table 5.22** The protein product of a characterised gene appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with protein folding.

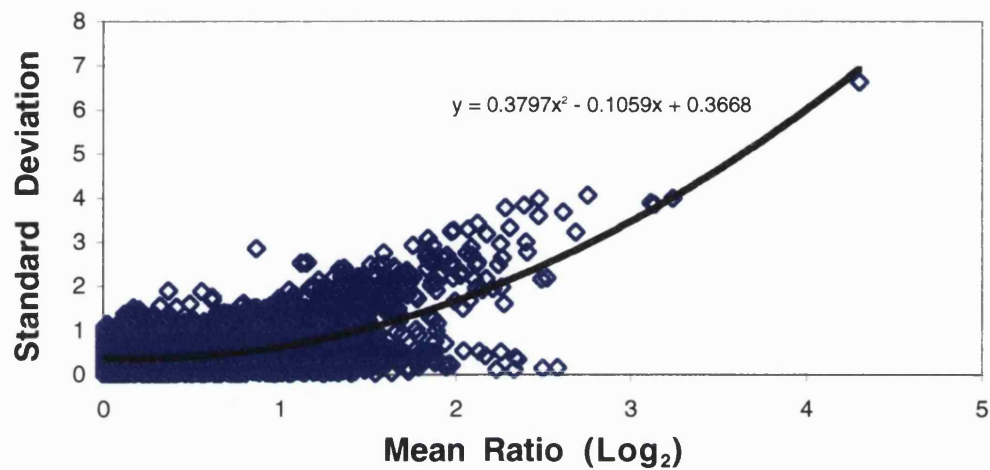
GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
MRPL8	Structural constituent of ribosome.	Mitochondrial genome maintenance, protein biosynthesis.	Mitochondrial large ribosomal subunit.	-1.19	-1.20
FPR1	Peptidyl-prolyl cis-trans isomerase activity.	Ribosome assembly.	Cytoplasm Nucleus	1.43	1.91

**Table 5.23** The protein product of characterised genes appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with ribosome structure and assembly.

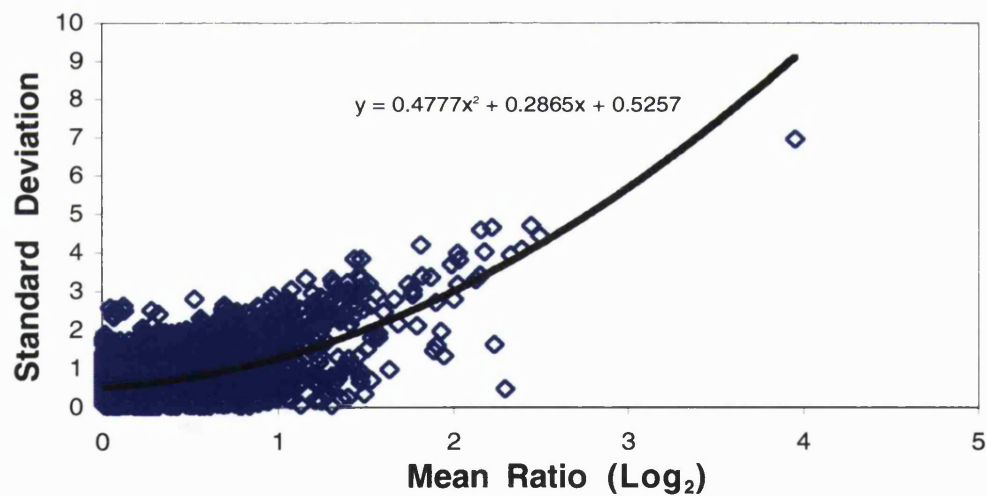
GENE NAME	MOLECULAR FUNCTION	BIOLOGICAL FUNCTION	CELLULAR COMPONENT	FOLD CHANGE	
				Serial	Parallel
PRM1	-	Plasma membrane fusion.	Integral to membrane Schmoo tip	1.29	1.50

**Table 5.24** The protein product of characterised genes appearing in both parallel and serial 100µM TIP-C7-amino treatments which are associated with sexual conjugation.

**(a)**

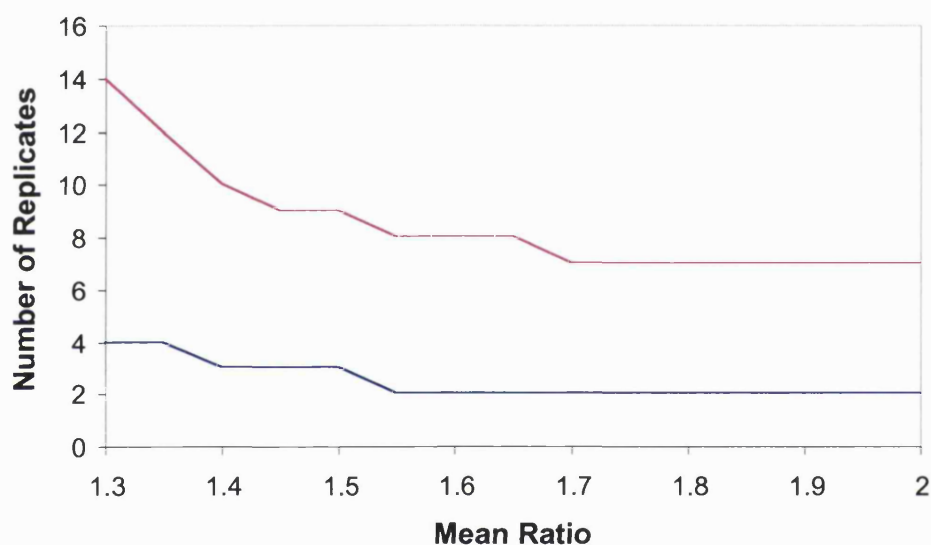


**(b)**



**Figure 5.9** Scatter plots showing the standard deviation versus the mean ratio of all genes for 100 $\mu$ M distamycin experiments where **(a)** shows the parallel treatments and **(b)** shows the serial treatments.

To evaluate the number of replicates required for each type of experimental design, scatter plots of the standard deviation of the intensities versus the mean  $\log_2$  ratio were plotted for the parallel and serial treatments of the 100 $\mu$ M distamycin treatments (Figure 5.9 a and b). These showed that for both the parallel and serial treatments that as the mean  $\log_2$  ratio increased the standard deviation also increased. When the number of replicates was plotted for each type of treatment against the mean ratio (Figure 5.10), this showed that as the mean ratio increased the number of replicates required decreased. Furthermore, as the standard deviation was higher for the serial treatments, the number of replicates required to achieve significance was greater for the serial data sets.



**Figure 5.10** Graph to show the minimum replicate number required for a p value of  $<0.05$  for a given level of differential expression, where (-) denotes the parallel data sets and (-) the serial data sets.

## 5.5 Discussion

A preliminary investigation was carried out into the biological consequences of the treatment of *Saccharomyces cerevisiae* with either distamycin A or the polyamide TIP-C7-amino. This was examined at a cellular level by determining the anti-proliferative effects of these compounds and at the level of the gene by examining the global gene expression profiles using DNA microarray hybridisation. The eukaryotic organism *S. cerevisiae* was chosen for this purpose as its small genome, comprised of approximately 6000 genes, has been established and subjected to extensive mutational analysis.

As the biological effects and uptake properties of polyamides have not been defined in yeast strains, *S. cerevisiae* was treated with distamycin and TIP-C7-amino to evaluate their anti-proliferative properties. Distamycin was found to have moderate anti-proliferative properties, with an  $IC_{50}$  value of  $90\mu M$  after two hours of treatment followed by growth over three days post-incubation. Continuous treatment of *S. cerevisiae* with distamycin produced 50% control growth with  $30\mu M$  at approximately 7 hours and at just over 4 hours with  $100\mu M$  distamycin. Cell viability studies demonstrated that the decreases in growth observed correlated with a loss in viability and thus distamycin was killing the cells over time rather than just inhibiting their proliferation. In contrast, the polyamide TIP-C7-amino did not show any anti-proliferative activity even after 24 hours of continuous incubation over the dose range examined, which ranged from 0- $100\mu M$ . It is possible, however, that a higher concentration of TIP-C7-amino might have produced an anti-proliferative effect.

The biological consequences of these compounds were further examined at the level of the gene in an effort to identify their effects on global gene expression. Two different experimental designs were used for this purpose, enabling an evaluation of the variation introduced into the data as a result of using different cell populations. Values obtained for changes in the levels of expression for all 6144 ORFs were ranked according to the probability that the differences observed over the three experiments were significant. This method was used to rank the ORFs for a number of reasons. The imposition of an arbitrary threshold cut off, for example  $\pm 2$ , results in the loss of the majority of the

information derived, as most of the changes in expression observed are usually small. This use of the threshold method introduces a bias into the analysis due to outlying data points, which may be less reliable. This method also fails to take into account that the importance of the change in expression is related to the importance of the gene itself. Thus it is possible that a small change in expression of a critical gene may have a more profound effect than a larger change in a gene which is involved in a less crucial mechanism. Statistical analysis accounts for the extent of differential gene expression and the variation between replicates, and can be used to obtain a measure of confidence in the results. In this study, the ranking method employed ensured that the genes showing consistent changes in their expression levels across repeat experiments were prioritised and enabled the reduction of the number of genes down to a manageable number for subsequent database searches.

In general, the parallel design, using the same population of cells for each treatment, provided less variation across the experiments. As a result, a large number of genes showed significant changes in their levels of expression. Serial treatments showed only 20-30% of the number genes derived from parallel treatments. Thus the cell population used appears to have a strong influence on the number of consistent changes observed between repeat experiments. However, whilst parallel treatments succeeded in reducing the variation between repeats, the biological significance of the changes observed under these conditions may be diminished relative to those observed across different cell populations. Therefore, the statistical threshold applied to results from the two designs may have to be adjusted to give the same level of confidence.

Another important consideration in the design of gene expression experiments is the number of replicates. In general, it is better to have many replicates as it increases the likelihood that the sample adequately represents the underlying biological variation. Due to the reduced variation, parallel experiments required fewer replicates to achieve the same level of statistical significance as the serial experiments. This suggests that repeat measurements performed in parallel do not fully capture the range of biological variation. When performing differential gene expression analysis, the experimental

design and number of replicates need to be optimised in order to achieve statistical significance and to account for the biological variation. Clearly, it is not possible to acquire an unlimited number of replicates, especially given the present prohibitive cost of microarrays. However this study provides estimates of the expected variation, which can be used to determine the optimal number of replicates in future experiments.

When the data for parallel and serial treatments were analysed individually, treatment with 100 $\mu$ M distamycin produced significant changes in the level of expression of 1233 ORFs for the parallel treatment and 244 ORFs across the serial treatments. Similarly, for 100 $\mu$ M TIP-C7-amino 848 and 283 ORFs showed significant expression changes in the parallel and serial treatments, respectively. As a significant change was determined by the p-value, a measure of the likelihood that the differences have occurred at random, it would be expected that 307 genes ( $6144 \times 0.05$ ) would come up at random for a p value of 0.05. Thus although these genes would fit the statistical criteria for significance, they would actually be false positives. As only 244 and 283 ORFs had a p value of  $<0.05$  for the serial treatments of 100 $\mu$ M distamycin and TIP-C7-amino, respectively, there is a very high chance that a large proportion of these may be false positives. This low number was probably due to an insufficient number of replicates.

In contrast, the higher numbers of ORFs derived from the parallel treatments suggests that only a relatively small proportion of these, 25% for 100 $\mu$ M distamycin and 36% for 100 $\mu$ M TIP-C7-amino, would be false positives. However in these cases, the high numbers of ORFs obtained made it prohibitive to examine all of their associated molecular and biological functions. For this reason, genes that came up consistently in both the parallel and serial experiments with a p value of  $<0.05$  were used for more extensive database searches. When the serial and parallel results are combined, the number of false positives should be reduced, with only 15 ORFs coming up at random ( $307 \times 0.05$ ). As 66 ORFs came up for the combined list for 100 $\mu$ M distamycin treatments this means that only 23% of these should be false positives. Similarly for 100 $\mu$ M TIP-C7-amino, 38 ORFs came up and thus up to 40% may be false positives in this case. Nevertheless, these are still relatively high numbers and thus differential



expression should be confirmed using either Northern Blotting or RT-PCR techniques prior to examination of promoter regions for any consensus binding sites, in order to exclude the potentially high number of false positives. The identification of consensus binding sites may enable differentiation between changes produced by direct interaction and those that occur as a result of downstream effects.

Analysis of the combined data sets showed that treatment with 100 $\mu$ M distamycin produced consistent changes in only 66 ORFs. 51 of these ORFs were found to represent characterised genes and the protein products of these genes were crudely sorted into six general biological functions. These included protein processing and metabolism, intracellular transport, DNA/RNA processing, protein folding and stress responses, structure and organisation of the cytoskeleton and lipid metabolism. Almost all the changes in expression measured involved upregulation. Indeed, only two genes were downregulated, HIS3 and STE12, which are involved in histidine biosynthesis and transcription factor activity, respectively.

The largest biological function group was associated with DNA/RNA processing and the protein product functions ranged from DNA binding (STE12 and PDR1), mRNA transcript cleavage (PCF11, KEM1 and SAD1) and polyadenylation (PCF11, NAB2 and YRA2). Indeed the largest change observed for this group was for the expression of NAB2 (+ 1.79 and 1.73). A number of the protein products are associated with the regulation of transcription from the Pol II promoter, PCF11, PDR1, SUB1 and RPB11. These changes appear to correlate with current opinion that the biological consequences of distamycin involve the disruption of DNA processing events such as transcription factor binding during transcription and the inhibition of helicase activity unwinding DNA during processing events such as replication.

The second largest functional group were the gene products associated with protein processing or metabolism. Relatively large changes in expression were observed for these genes, with up to 3 fold changes in the levels of expression (SEN3/RPN2 +2.86 and +2.33). A number of these, RPN5, UBS1, SEN3/RPN2 are associated with

ubiquitin dependent protein catabolism. It is conceivable that the increase in this catabolic activity may be associated with general protein breakdown due to the onset of cell death, although after 2 hours of continuous exposure, the incubation period for the microarray experiments, distamycin does not appear to be killing the cells. However by 4 hours, cell growth in the presence of 100 $\mu$ M distamycin is reduced to 60% of control growth. Thus it appears that the onset of death occurs between 2 to 4 hours. This may explain the upregulation of protein catabolism together with the relatively small amount of gene expression associated with cell stress responses at 2 hours (CNS1, SSA4, HSP104, RHR2 and DAK1). The timing of the onset of cell death may also explain why DNA/RNA processing events are upregulated rather than downregulated on disruption by distamycin. It is possible that at 2 hours cells may still be trying to overcome the disruptive effects incurred by distamycin binding and attempting to compensate for these effects by upregulating DNA replication and transcription. This may account for the slight stimulation in cell growth observed at 2 hours relative to control. Between 2 and 4 hours the cells become committed to death and these processes may actually be downregulated at this stage.

Analysis of the combined data sets showed that treatment with 100 $\mu$ M TIP-C7-amino produced consistent changes in only 38 ORFs. Twenty five of these ORFs represented characterised genes and the protein products of these genes were crudely sorted into seven general biological functions. These included transport, DNA/RNA processing and metabolism, cytoskeletal organisation, glucose metabolism, lipid metabolism, protein folding and ribosome structure and assembly. In contrast to the effects observed with 100 $\mu$ M distamycin, the ORFs in the combined data set were predominantly downregulated. However, characterised genes were found to be upregulated and downregulated in even measure.

Once more the largest functional group was associated with DNA/RNA processing. In contrast to treatment with 100 $\mu$ M distamycin, patterns in gene expression in this group were harder to elucidate as 4 of the genes were upregulated and 5 downregulated. This was particularly evident with genes associated with the initiation of translation, TIF2

and ANB1, which were up and downregulated respectively. Another example of conflicting patterns in gene expression was the downregulation of HOP1, a gene associated with meiosis. However, PRM1 was upregulated and is a gene required for shmoo formation, an extension of the plasma membrane required for the sexual conjugation of adjoining yeast cells. It might therefore be expected that these two genes would be co-regulated.

The second largest functional groups were gene products associated with vesicular transport and glucose metabolism. All of the genes associated with vesicular transport were downregulated. Genes associated with glucose metabolism were both up and downregulated. The upregulation of TDH2, associated with gluconeogenesis, and TSL1, associated with trehalose biosynthesis, suggested that cells are requiring increased levels of glucose, possibly as a response to stress induced by a glucose deficiency. This seems to be supported by the concomitant decrease observed in the expression levels of GND2, associated with the oxidative phase of the pentose phosphate pathway, whereby glucose is broken down to form pentose sugars for the synthesis of nucleic acids. However it is unclear why the cells exposed to the polyamide may require more glucose than untreated cells. If the increased requirement for glucose were only due to overgrowth of the cell cultures this would also have been seen in the control samples and thus would not appear as a differential effect. Were vesicular transport shown to be upregulated, in light of the sequestration observed in the previous chapter, the increased glucose requirement may have been due to active transport mechanisms associated with endocytosis and subsequent exocytosis of sequestered polyamide. However, as vesicular transport was observed to be downregulated this does not appear to be the case.

The remaining functional groups are composed of only one or two genes, making it difficult to define patterns of expression. The difficulty in interpreting the TIP-C7-amino data may be due to a number of factors. Firstly, this polyamide may not be having any significant biological effect on the yeast cells as seems to be suggested by the cytotoxicity data. This also seems to be confirmed by the relatively small number of

genes that show changes in expression levels on treatment and the high level of possible false positives among those genes. This is in contrast to the much higher numbers of genes that were effected by treatment with 100 $\mu$ M distamycin. The lack of a co-ordinated response among the functional groups which has made it difficult to define a pattern of gene expression, also seems to suggest that there is a significant possibility that many of these genes have come up at random. Once more, this was not the case for 100 $\mu$ M distamycin, which showed a predominant upregulation in the genes affected by treatment.

Comparison of the gene lists for distamycin A and TIP-C7-amino demonstrated few common features. No genes appeared to be shared between either the parallel, serial or complete gene lists for the two compounds. The only similarity observed was that the largest functional gene group was found to be associated with DNA/RNA processing for both compounds. However, within this functional group there were no genes shared between the distamycin and TIP-C7-amino treatments. Furthermore, whilst almost all of these genes were upregulated on treatment with distamycin, only 40% of the affected genes associated with DNA/RNA processing were upregulated when treated with TIP-C7-amino.

The absence of any biological effect on treatment with the polyamide TIP-C7-amino either at the level of the cell or at the level of the gene may suggest that this polyamide did not get taken up into cells readily over the incubation period used. This would correlate with results obtained in the previous chapter examining polyamide uptake in human cells. Furthermore, the absence of upregulation in genes associated with vesicular transport mechanisms suggests that after two hours the polyamide was not being internalised and sequestered. It is possible that such a process would require a much longer incubation period as was observed for human cells in chapter 4. There is always the possibility that the nylon microarray platform used here was not sensitive enough to measure biological effects of a small magnitude above the noise produced by the inherent variations in the technique. Indeed, the dose and incubation period chosen for the microarray experiments may not have been optimal for the measurement of

biological effects at the level of the message. Examination of the levels of gene expression following longer incubation periods may provide more defined patterns of gene expression. The transcript levels of genes of interest, for example those associated with DNA or RNA processing events, which are found to be consistently up or down regulated following treatment using microarrays might then be verified using alternative techniques measuring mRNA levels such as Northern Blotting or RT-PCR. This would enable the elimination of false positives.

## CHAPTER 6

### OVERALL DISCUSSION AND FUTURE WORK

The limited success of many current chemotherapy regimes highlights the need for a new generation of anticancer strategies, able to discriminate more effectively between tumour and normal tissue. As recent genomic studies indicate that tumours may exhibit different patterns of transcription to normal cells, novel approaches to combating cancer now include the exploration of transcriptional therapy (Pandolfi 2001; Darnell 2002). This aims to regulate the transcription of key disease-related genes either indirectly by targeting gene transcripts or alternatively by direct interaction with genomic DNA. Whilst numerous molecules are under development for this purpose, the current generation of minor groove-binding polyamides already demonstrate the stringent sequence recognition within double-stranded DNA that would be required for targeting genomic DNA. Moreover, they have been shown to achieve selective binding with a comparable affinity to native regulatory proteins, usually binding at nanomolar concentrations. To date this work has been carried out predominantly in *in vitro* models. Nevertheless, the selective binding observed with polyamides within nucleosomal DNA (Gottesfeld *et al.*, 2001; 2002), as well as their ability to up and down regulate transcription *in vitro* and in some cell culture situations (Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998, 1999a, b; Mapp *et al.*, 2000; Chiang *et al.*, 2000) demonstrates their potential as artificial transcriptional regulators and suggests that polyamides may provide some advantages over alternative strategies such as siRNA.

This thesis describes an investigation into the interactions of polyamides with DNA both *in vitro* and in cells. It examines the structural requirements for the *in vitro* binding of unlinked and linked polyamides to Watson-Crick and non-canonical DNA sequences. This includes the effects of linker length, heterocycle permutation and N-terminal head group on the affinity and selectivity of binding, together with the discriminatory properties of an imidazole:imidazole side-by-side pairing for selective TG mismatch recognition. In cells, the uptake and distribution of a thiazolated polyamide was investigated, in addition to its effects on global gene expression.

A comprehensive investigation into the effects of linker length and head group across a large number of novel cross-linked polyamides was carried out in Chapter 2. This enabled characterisation of the structural requirements for high affinity and selective DNA binding using the cross-linked motif. In accord with previous and concurrent studies, the binding trends observed showed the combined influence of these structural elements on both the affinity and selectivity of DNA binding for this linkage motif (Chen *et al.*, 1996; Burckhardt *et al.*, 2000; Lacy *et al.*, 2002a). These demonstrated that despite the slight variance in linker length required for optimal binding affinity, the best selectivity was achieved with a seven-carbon linker, in agreement with previous work by Chen and co-workers (Chen *et al.*, 1996). Furthermore, for the cross-linked motif the N-terminal head group profoundly influenced binding activity, a feature not observed with the hairpin linkage. In cross-linked molecules the uncharged hydrogen and acetamide head groups were unable to promote groove binding, whilst charged groups such as amino and formamide facilitated binding. This suggests that favourable interactions at the N-terminus, possibly via hydrogen bonding, are critical for optimal groove binding with the cross-linked motif.

The use of thiazole for the discrimination of A from T was also investigated in Chapter 2. Placed at the N-terminus of a cross-linked motif, thiazole demonstrated a moderate preference for adenine residues. This was not in agreement with previous work, which had been unable to detect such discrimination when thiazole was placed centrally within a hairpin (Nguyen *et al.*, 2001). Cross-linked and hairpin polyamides, however, appear to differ in their bonding requirements at the N-terminus, as exemplified by the abrogation of groove-binding with an N-terminal hydrogen in the cross-linked motif, a structural element commonly used in this position within hairpins. Furthermore, it is now evident that the sequence selectivity of heterocycles can be affected by their positional context within the polyamide, which might also account for the differences in selectivity observed with the thiazole moiety. Indeed, context dependency has been found to affect the selectivity of the 3-hydroxypyrrole moiety, which exhibits effective T from A discrimination only when centrally placed (White *et al.*, 1998; Ellervik *et al.*,



2000; Foister *et al.*, 2003). This is not observed with Im/Py pairings, which can differentiate G from C at both central and terminal positions.

To eliminate the possibility of context dependency on the moderate selectivity of thiazole for adenine, both the number and relative position of thiazole units would need to be examined more fully in either the hairpin or cross-linked motif. Improved AT discrimination may be provided by alternative side-by-side pairing to the thiazole/pyrrole pair examined here. For example 3-hydroxypyrrole, which has some preference for thymine, could be paired with thiazole to determine whether the combined affinity and selectivity for either adenine or thymine of each heterocycle in Th/Hp is greater than that of the individual Th/Py and Hp/Py pairs. However, the modest selectivity and affinity of both of these heterocycles relative to pyrrole and imidazole may ultimately restrict their use in the differentiation of AT base pairs.

Despite continuing investigation into heterocycle design there has been limited success to date in reliably discriminating TA from AT base pairs. Indeed, N-terminal pairings capable of binding with an affinity and specificity comparable to those of Im/Py for GC base pairs have yet to be elucidated. Thus, although the current set of pairing rules provides a useful basis for polyamide design, variation in the secondary structure of different DNA sequences means that predicting the optimal heterocycle combination required for binding at a predetermined site with the heterocycles available at present can still be unreliable. Faithful differentiation of these bases may require alternative strategies such as the use of heterodimers, as found with TG: GT discrimination (Rucker *et al.*, 2003), or even additional DNA contacts perhaps within the major groove.

The discrimination of non-canonical TG mismatches using an Im/Im heterocycle pairing was examined in Chapter 3. Previous and concurrent NMR, CD, SPR and thermodynamic studies investigating the ability of the Im/Im pair to discriminate TG from CG have used short oligomers containing a single putative binding site (Yang *et al.*, 1999b; Lacy *et al.*, 2002b). Here, however, the selectivity of an Im/Im pair was examined using a longer region of DNA sequence containing all three putative binding

sites, thereby enabling a direct comparison of binding preference. This demonstrated that Im/Im was strongly selective for a single TG mismatch when present alongside a putative CG match site, indicating that the Im/Im ring pair is a robust recognition motif for the TG mismatch. Furthermore, subsequent thermodynamic analysis has shown that the selectivity of Im/Im occurs as a result of favourable entropic contributions at the TG site (Lacy *et al.*, 2004).

The work presented here, together with that from Lee and coworkers, clearly demonstrates the potential of polyamides for mismatch recognition. However, although Im/Im pairs are highly selective for single TG mismatches they are unable to discriminate between TG and GT (Yang *et al.*, 1999b). It is possible that incorporation of Im/Im pairs into polyamide heterodimers such as ImPyIm-ImImIm (Yang *et al.*, 1999b) and ImImPyPy-PyPyImPy, which has been found to distinguish T: G from G: T (Rucker *et al.*, 2003), may provide more stable complexes with TG mismatches. Moreover, alternative heterocycle pairings might be used to discriminate TG from GT more effectively. Indeed Lee and coworkers have proposed that an imidazole-hydroxypyrrole (ImHp) ring pair may produce suitable discrimination (Yang *et al.*, 1999b). In this pair imidazole may interact with guanine whilst hydroxypyrrole might bind preferentially to the corresponding mismatched thymine. Indeed modelling studies suggest that in this position the Hp hydroxyl group might hydrogen bond in a manner similar to conserved water molecules as observed in the crystal structures of G: T wobble base pairs (Hunter *et al.*, 1987). This would enable bridging of both the guanine and thymine base pairs (Yang *et al.*, 1999b).

Notwithstanding the selectivity for both Watson-Crick and non-canonical sequences achieved to date with polyamides in *in vitro* systems, sequence selective binding using these molecules has been more difficult to translate into cellular models; a problem shared with alternative proposed transcriptional control strategies. Until quite recently studies examining cellular uptake of fluorophore-tagged polyamides have demonstrated limited internalisation and nuclear localisation in a narrow range of human cell lines (Sharma *et al.*, 2001; Belitsky *et al.*, 2002; Dudouet *et al.*, 2003). To circumvent the

need for a conjugated fluorophore, the uptake and intracellular distribution of an untagged polyamide in living cells was investigated using fluorescence microscopy in Chapter 4, exploiting instead its inherent fluorescence on UV excitation. This was significant as it enabled direct assessment of polyamide uptake and distribution in living cells. It thereby negated the need for an attached fluorophore, which can significantly alter both the uptake characteristics (Crowley *et al.*, 2003) and the DNA binding properties of the conjugate relative to the corresponding polyamide alone (Battacharya & Thomas 2002).

The slow intracellular accumulation and the absence of detectable nuclear localisation of the thiazolated polyamide examined here correlated with concurrent observations of polyamide uptake by other research groups (Sharma *et al.*, 2001; Crowley *et al.*, 2003). The cytosolic staining together with the absence of nuclear localisation, which persisted even after multiple rounds of mitosis, strongly indicated that the polyamide was not free to enter the nucleus. This suggested that the internalisation seen was via a vesicular uptake process sequestering the polyamide within cytosolic compartments, a phenomenon observed directly by Crowley and coworkers (Crowley *et al.*, 2003). Microinjection of the polyamide into cell nuclei confirmed that it was possible to visualise nuclear localisation using the fluorescence microscopy technique employed here and indicated that the absence of nuclear staining was as a result of the failure of the polyamide to accumulate at significant concentration within this organelle. The possibility of polyamide sequestration was further strengthened by evidence that the nuclear envelope did not provide a barrier to the distribution of freely soluble polyamide. This was clearly evident from the rapid staining seen here with isolated nuclei, in addition to the staining of isolated nuclei with fluorophore tagged polyamides as observed by Laemmli and co-workers (Janssen *et al.*, 2000a; Maeshima *et al.*, 2001).

Whilst some success was achieved in visualising the uptake of polyamides using fluorescence microscopy, it is possible that very low concentrations of nuclear polyamide may not have been detected. A more sensitive approach may be the labelling polyamides with, for example, tritiated hydrogen and tracking them using quantitative

electron autoradiography. This would also circumvent the use of large conjugated fluorophores for visualisation with all their associated effects on the charge and size of the molecule. Use of radiolabelling might then facilitate examination of the uptake of polyamides in different cell types.

Crowley *et al.*, (2003) have shown that charge is a critical factor in the effective cell uptake of polyamides and it is also known to influence the uptake of other putative transcriptional control molecules. However, it may not be the only factor influencing their uptake, as the cationic molecule netropsin readily accumulates in the nucleus of living cells (Bailly *et al.*, 1989). This may be as a result of its small size, as recent work has shown that nuclear localisation of polyamides can also be affected by their molecular weight, heterocycle content and linker composition (Best *et al.*, 2003; Edelson *et al.*, 2004). Furthermore uncharged molecules such as PNA are not readily internalised by cells (Larsen *et al.*, 1999). Thus, a number of other factors, including lipophilicity, may also play a part in the permeability of such molecules to the cell membrane (Lown *et al.*, 1989).

Moreover, the molecular basis for the differential uptake of polyamides observed in a number of cell lines is unclear. Whilst in some cases it may be explained by structural differences alone, the varying uptake and nuclear localisation of a single polyamide structure across different cell lines remains to be addressed (Belitsky *et al.*, 2002; Best *et al.*, 2003; Edelson *et al.*, 2004). It is possible that the type of cell targeted may also influence the efficiency of polyamide internalisation. In general, however, the range of factors recently shown to affect the cellular uptake of polyamides are shared with other transcriptional control strategies and highlight the difficulties encountered transporting complex synthetic molecules into cells. This suggests that in common with alternative antimessage or antigene approaches it may be difficult to deliver therapeutically relevant concentrations of polyamides required for selective transcriptional control.

As any future therapeutic applications for gene targeting using polyamides will rely extensively on their effective cell uptake and nuclear localisation, it will be crucial to

define clearly the structural limitations for their optimal delivery to nuclear DNA. This is especially important when considering the uptake of the much larger polyamides, which will be required for targeting longer sequences to prevent binding across the genome. Alternative delivery strategies are already being considered for larger polyamides, with for example investigations into the design of small self-assembling polyamide fragments (Poulin-Kerstien & Dervan 2003). However, other forms of delivery may be required, for example the use of liposomes or cationic lipids, as with other putative transcriptional control strategies. Once inside the cell, nuclear import might be enhanced by the addition of a nuclear localisation signal (NLS). These often occur in nature as clusters of basic amino acids, which are recognised by importins or karyopherins that facilitate transport into the nucleus. NLS-mediated nuclear import is utilised by numerous cellular proteins that are commonly shuttled between cytoplasm and nucleus. It is becoming evident, however, that this represents only one form of nuclear import, with increasing numbers of signalling molecules found to enter the nucleus via NLS and importin-independent pathways (Xu & Massague 2004). Nevertheless, as our understanding of the nucleocytoplasmic shuttling of cellular signal transducers and transcription factors improves it may be possible to mimic these uptake pathways with artificial transcriptional control molecules such as polyamides.

Although recent work has demonstrated the difficulties of delivering polyamides into cells, the antitumour and antiviral activities of polyamides observed over a number of years suggest that it is possible under certain conditions for polyamides to be internalised and for them to elicit a biological effect. Studies investigating their antiviral properties in lymphoid cell lines have shown that they can produce a significant reduction in the proliferation of viruses such as HIV-1 and FIV in infected cells (Lown *et al.*, 1989; Wang & Lown 1992; Dickinson *et al.*, 1998; Sharma *et al.*, 2002; Yang *et al.*, 2003). This is thought to occur as a consequence of targeted binding at predetermined sites in the viral genome, although not demonstrated directly. However the inhibition of retroviral replication observed does not necessarily prove that polyamides are interacting with the nuclear DNA and could be as a result of numerous modes of action (Wang & Lown 1992; Sharma *et al.*, 2002). This could explain the

reduced cytotoxicity of distamycin and related groove binders relative to their associated antiviral activity, as the two processes may not be occurring via the same molecular mechanism (Lown *et al.*, 1989). A recent study, however, shows the targeted derepression of HIV-1 replication in ACH2 cells, a chronically and latently infected T cell line where HIV-1 is fully integrated into the host genome. This strongly suggests that in this instance disruption of viral replication by polyamides is occurring at the level of the nucleus (Coull *et al.*, 2002).

The effect on global gene expression of the exposure of *S. cerevisiae* to a polyamide was evaluated in Chapter 5 and compared to the effect exerted by distamycin A, a weakly cytotoxic agent of similar structure. The uptake of distamycin and polyamides in yeast, an organism possessing a cell wall in addition to a cell membrane or spheroplast, has not been demonstrated directly. Nonetheless, the cytotoxic activity exerted by distamycin in the anti-proliferative assays carried out here suggested that it was being taken up by yeast cells. This appears to correlate with cytotoxicity studies in other eukaryotic cell lines, which have demonstrated distamycin to be moderately or weakly cytotoxic (Chandra *et al.*, 1972; Lown *et al.*, 1989). Given the biological effect exerted by distamycin on yeast cells, it is not unreasonable to assume that related molecules might also be taken up within this biological system. Anti-proliferative assays with the polyamide TIP-C7-amino, however, demonstrated that it had no cytotoxic or cytostatic activity in this organism, which alongside evidence of sequestration in Chapter 4 suggested that this compound may not be taken up by these cells.

Analysis of the combined data sets of differential gene expression in cells treated with 100 $\mu$ M distamycin or the polyamide TIP-C7-amino, revealed consistent changes in the transcription of approximately 1% and 0.5% of the yeast genome, respectively. This remarkably small number of affected genes is consistent with other recent microarray studies which have also observed that the expression levels of relatively few genes are affected by treatment with polyamides (Supekova *et al.*, 2002; Dudouet *et al.*, 2003). However, although treatment with distamycin appeared to produce a pattern of gene expression that could be loosely correlated with the anti-proliferative effect observed,

the same was not true for the polyamide TIP-C7-amino. The absence of any cytotoxic or cytostatic activity by the polyamide, together with the extremely small percentage of consistent changes in the levels of gene expression could be accounted for simply by statistical variation and strongly suggested that it was not being taken up by the cells.

The limited gene number consistently affected by distamycin treatment may have been due to the single incubation period examined of 2 hours, which was either just prior or coincided with the onset of cell death. The onset of death could be more clearly determined using flow cytometry examining the effects of distamycin on the cell cycle over time. Furthermore, examination of the transcription profiles produced over longer time periods, for example 4 and 6 hours, at which cell death was observed during continuous incubation in the anti-proliferative assays, may enable clustering of the genes affected over time. This may produce clusters of genes associated with a given metabolic pathway and could aid the differentiation between the direct interaction of distamycin and downstream effects. It would also provide more robust candidate genes for confirmation of up or down regulation by northern blotting or RT-PCR techniques, thereby enabling elucidation of some of the biological consequences of distamycin treatment in cells.

As lymphoid cell lines such as Jurkat and CEM appear to readily take up polyamides, it may also be profitable to examine the biological effects of increasing concentrations of polyamide on genomic transcription in T cells. Incubation of CEM cells with polyamide concentrations of 1-2 $\mu$ M have been shown to have no anti-proliferative effect (Dudouet *et al.*, 2003), which is consistent with the effects observed with low concentrations of polyamide in Jurkat cells this thesis. However, the anti-proliferative effect observed here over prolonged exposure and high concentrations indicate that there may be a concentration threshold at which non-specific binding to genomic DNA may occur. The effects on genomic transcription could be examined using microarrays, in combination with flow cytometry to determine possible mechanisms for the anti-proliferative effects observed. Once more, the examination of a number of incubation periods would enable clustering of the affected genes, providing more gene candidates for both RT-PCR and



the search for consensus binding motifs to determine whether the genes affected by polyamide exposure are the result of sequence specific binding or downstream events. Sequence specific binding could also be verified using *in vivo* footprinting techniques or by chromatin immunoprecipitation (ChIP), whereby non-covalent ligand-DNA interactions are captured by *in vivo* formaldehyde cross-linking of whole cells and the ligand-DNA complexes identified by immunoprecipitation.

Whilst investigation of the promoter regions from the genes identified was not possible in the present study due to the potentially high rate of false positives, analysis of potential polyamide binding sites linked to transcriptional regulation has been carried out by Supekova and co-workers (Supekova *et al.*, 2002). Promoter regions of the genes affected by polyamide treatment revealed numerous possible binding sites for the polyamides examined in that study, but failed to identify a common consensus motif for binding suggesting that the effects exerted on transcription were not sequence directed. It is possible that the differential expression observed occurred as a downstream consequence of polyamide binding, rather than by direct interaction with the polyamide itself. More recently however, different transcription profiles have been produced by polyamides targeted to different sequences and importantly these compounds have been shown to localise within the cell nucleus (Dudouet *et al.*, 2003). This seems to indicate that in this case each polyamide was having an individual effect possibly as a result of the different sequences to which they were targeted, although this was not shown to be as a direct consequence of targeted polyamide binding. Thus despite these preliminary studies, differential gene expression as a result of the direct interaction of polyamides binding at targeted sites within the genome in living cells remains to be demonstrated.

In conclusion the *in vitro* study of the selectivity and affinity of the interactions of DNA binding ligands such as polyamides is a crucial approach to understanding their possible mechanisms of action in cells. This has provided useful molecular vectors for the direction of conventional DNA damaging moieties to specific sequences, as observed with the numerous polyamide conjugates currently under development. Nevertheless, for polyamides to be employed successfully in gene targeting strategies elucidation of

the structural elements required for their efficient cell uptake and nuclear distribution is vital. These issues have been emphasised by numerous alternative antigene and antimessage strategies, which have already encountered difficulties in targeting adequate levels of ligand to show activity. Furthermore, for the evaluation and prediction of the biological consequences of polyamide binding in cells to be possible, a greater understanding is required of the accessibility of sequences in genomic DNA, the dynamics of nucleosomal organisation and the three-dimensional mechanisms of transcriptional activation. As a consequence, the design of the next generation of polyamides will require a more thorough understanding of these factors to achieve realistic therapeutic utility.

## APPENDIX I

DEFINITION: pBSLTR34TF10 comprised of the viral vector pBSSK and the 5' LTR of FIV as donated by Tom Phillips from Scripps Research Institute.

```
1  CTAAATTGTA AGCGTTAATA TTTTGTTAAA ATTCGCGTTA AATTTTTGTT
51  AAATCAGCTC ATTTTTTAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT
101 AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTGTGTC CAGTTTGGA
151 CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAAA GGGCGAAAAA
201 CCGTCTATCA GGGCGATGGC CCACTACGTG AACCATCACC CTAATCAAGT
251 TTTTGGGGT CGAGGTGCCG TAAAGCACTA AATCGGAACC CTAAAGGGAG
301 CCCCCGATTT AGAGCTTGAC GGGGAAAGCC GGCGAACGTG GCGAGAAAGG
351 AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA GGGCGCTGGC AAGTG TAGCG
401 GTCACGCTGC GCGTAACCAC CACACCCGCC GCGCTTAATG CGCCGCTACA
451 GGGCGCGTCC CATTCGCCAT TCAGGCTGCG CAACTGTTGG GAAGGGCGAT
501 CGGTGCGGGC CTCTTCGCTA TTACGCCAGC TGGCGAAAGG GGGATGTGCT
551 GCAAGGCGAT TAAGTTGGGT AACGCCAGGG TTTTCCCAGT CACGACGTTG
601 TAAAACGACG GCCAGTGAGC GCGCGTAATA CGACTCACTA TAGGGCGAAT
651 TGGGTACCGG GCCCCCCCTC GAGGTGACG GTATCGATAA GCTTGATCTC
    OBUSN1 →
701 GAGCTCGAGT GGGATGAGTA TTGGAACCT GAAGAAATAG AAAGAATGCT
751 TATGGACTAG GGA CTGTTTA CGAACAAATG ATAAAAGGAA ATAGCTGAGC
801 ATGACTCATA GTTAAAGCGC TAGCAGC TGC CTAACCGCAA AACCACATCC
851 TATGGAAAGC TTGCTAATGA CGTATAAGTT GTTCCATTGT AAGAGTATAT
901 AACCAGTGCT TTGTGAAACT TCGAGGAGTC TCTTTGTTGA GGACTTTTGA
951 GTTCTCCCTT GAGGCTCCCA CAGATACAAT AAATATTTGA GATTGAACCC
1001 TGTCGAGTAT CTGTGTAATC TTTTTTACCT GTGAGGTCTC GGAATCCGGG
    ← OBUSN3
1051 CCGAGAACTT CGCAAGATCT AGATCTAATC GAATTCCTGC AGCCCGGGGG
1101 ATCCACTAGT TCTAGAGCGG CCGCCACCGC GGTGGAGCTC CAGCTTTTGT
1151 TCCCTTTAGT GAGGGTTAAT TGCGCGCTTG GCGTAATCAT GGTCATAGCT
```

1201 GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAGC  
 1251 CCGGAAGCAT AAAGTGTAAG GCCTGGGGTG CCTAATGAGT GAGCTAACTC  
 1301 ACATTAATTG CGTTGCGCTC ACTGCCCCGCT TTCCAGTCGG GAAACCTGTC  
 1351 GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC  
 1401 GTATTGGGCG CTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC  
 1451 GTTCGGCTGC GCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT  
 1501 ATCCACAGAA TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC  
 1551 AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTCCAT  
 1601 AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG  
 1651 GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCTGGAA  
 1701 GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG  
 1751 TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA GCTCACGCTG  
 1801 TAGGTATCTC AGTTCGGTGT AGGTCGTTTC CTCCAAGCTG GGCTGTGTGC  
 1851 ACGAACCCCC CGTTCAGCCC GACCGCTGCG CTTATCCGG TAACTATCGT  
 1901 CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC  
 1951 TGGTAACAGG ATTAGCAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT  
 2001 TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC  
 2051 TCGGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG  
 2101 ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC  
 2151 AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT  
 2201 TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT  
 2251 GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA  
 2301 AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC TTGGTCTGAC  
 2351 AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT  
 2401 TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC  
 2451 GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA  
 2501 CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGG  
 2551 CCGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA

2601 ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC  
 2651 AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG  
 2701 TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT  
 2751 CCCCCATGTT GTGCAAAAAA GCGGTAGCTC CTTCGGTCCT CCGATCGTT  
 2801 GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT  
 2851 GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG  
 2901 GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT  
 2951 TGCTGTTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC  
 3001 TTTAAAAGTG CTCATCATTG GAAAACGTTT TCGGGGCGA AAACCTCTCA  
 3051 GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC  
 3101 AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG GGTGACAAAA  
 3151 ACAGGAAGGC AAAATGCCGC AAAAAAGGGA ATAAGGGCGA CACGGAAAT  
 3201 GTTGAATACT CATACTCTTC CTTTTTCAAT ATTATTGAAG CATTTATCAG  
 3251 GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA  
 3301 ACAAATAGGG GTTCCGCGCA CATTCCCCG AAAAGTGCCA C

## APPENDIX II

LOCUS: HUMHER201 757 bp DNA linear PRI 31-DEC-1994

DEFINITION: Human HER2 gene, promoter region and exon 1.

ACCESSION: M16789

```
1  CCCGGGGGTC CTGGAAGCCA CAAGGTAAAC ACAACACATC CCCCTCCTTG
51  ACTATGCAAT TTTACTAGAG GATGTGGTGG GAAAACCATT ATTTGATATT
101 AAAACAAATA GGCTTGGGAT GGAGTAGGAT GCAAGCTCCC CAGGAAAGTT
151 TAAGATAAAA CCTGAGACTT AAAAGGGTGT TAAGAGTGGC AGCCTAGGGA
201 ATTTATCCCG GACTCCGGGG GAGGGGGCAG AGTCACCAGC CTCTGCATTT
251 AGGGATTCTC CGAGGAAAAG TGTGAGAACG GCTGCAGGCA ACCCAGGCGT
301 CCCGGCGCTA GGAGGGACGA CCCAGGCCTG CGCGAAGAGA GGGAGAAAGT
                                     NEU-F  ──────────>
351 GAAGCTGGGA GTTGCCGACT CCCAGACTTC GTTGGAATGC AGTTGGAGGG
401 GGCGAGCTGG GAGCGCGCTT GCTCCCAATC ACAGGAGAAG GAGGAGGTGG
451 AGGAGGAGGG CTGCTTGAGG AAGTATAAGA ATGAAGTTGT GAAGCTGAGA
501 TTCCCCTCCA TTGGGACCGG AGAAACCAGG GGAGCCCCC GGGCAGCCGC
                                     <──────── NEU-R
551 GCGCCCCTTC CCACGGGGCC CTTTACTGCG CCGCGCGCCC GGCCCCACC
601 CCTCGCAGCA CCCC GCGCCC CGCGCCCTCC CAGCCGGGTC CAGCCGGAGC
651 CATGGGGCCG GAGCCGCAGT GAGCACCATG GAGCTGGCGG CCTTGTGCCG
701 CTGGGGGCTC CTCCTCGCCC TCTTGCCCC CGGAGCCGCG AGCACCCAAG
751 GTGGGTC
```

## REFERENCES

- Abu-Daya, A., Brown, P. M. & Fox, K. R. (1995) DNA sequence preferences of several AT selective minor groove binding ligands. *Nucleic Acids Res.* **23**, 3385-3392.
- Adams, J. M. & Cory, S. (1998) The bcl-2 protein family: arbiters of cell survival. *Science* **281**, 1322-1326.
- Agrawal, S. (1999) Importance of nucleotide sequence and chemical modifications of antisense oligonucleotides. *Biochim. Biophys. Acta* **1489**, 53-68.
- Agrawal, S., Jiang, Z., Zhao, Q., Shaw, D., Cai, Q., Roskey, A., Channavajjala, L., Saxinger, C. & Zhang, R. (1997) Mixed backbone oligonucleotides as second generation antisense oligonucleotides: *in vitro* and *in vivo* studies. *Proc. Natl., Acad. Sci. USA* **94**, 2620-2625.
- Akhtar, S., Kole, R. & Juliano, R. L. (1991) Stability of antisense DNA oligodeoxynucleotide analogs in cellular extracts and sera. *Life Sci.* **49**, 1793-1801.
- Akhtar, S. (1998) Antisense technology: selection and delivery of optimally acting antisense oligonucleotides. *J. Drug Targeting* **5**, 225-234.
- Albert, F. G., Eckdahl, T. T., Fitzgerald, D. J., & Anderson, J. N. (1999) Heterogeneity in the actions of drugs that bind in the DNA minor groove. *Biochemistry* **38**, 10135-10146.
- Allawi, H. T. & SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry* **36**, 10581-10594.
- Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N. & Perucho, M. (1988) Most human carcinomas of the exocrine pancreas contain mutant c-k-ras genes. *Cell* **53**, 549-554.
- Andersson, L. C., Nilsson, K., & Gahmberg, C. G. (1979) K562 - A human erythroleukaemic cell line. *Intl. J. Cancer* **23**, 143-147.



- Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B. & Ptashne, M. (2001) Towards a minimal motif for artificial transcriptional activators. *Chem. & Biol.* **102**, 1-10.
- Arcamone, F. M., Animati, F., Barbieri, B. Configliacchi, E., D'Alessio, R., Geroni, C., Giuliani, F. C., Lazzari, E., Menozzi, M. & Mongelli, N. (1989) Synthesis, DNA binding properties, and antitumour activity of novel distamycin derivatives. *J. Med. Chem.* **32**, 774-778.
- Arcamone, F. M. (1993) Synthesis and DNA binding selectivity of pyrrole-amidine oligopeptides. *Farmaco* **48**, 143-150.
- Arima, K., Kohsaka, M., Tamura, K. G., Imanaka, H. & Sakai, H. (1972) Studies on tomaymycin, a new antibiotic. Isolation and properties of tomaymycin. *J. Antibiot* **25**, 437-444.
- Arimondo, P. B., Garestier, T., Hélène, C. & Sun, J. S. (2001) Detection of competing DNA structures by thermal gradient gel electrophoresis: from self-association to triple helix formation by (G,A) containing oligonucleotides. *Nucleic Acids Res.* **29**, E15.
- Aubert, Y., Perrouault, L., Hélène, C., Giovannangeli, C. & Asseline, U. (2001) Synthesis and properties of triple helix-forming oligodeoxyribonucleotides containing 7-chloro-7-deaza-2'-deoxyguanosine. *Bioorg. Med. Chem.* **9**, 1617-1624.
- Bacon, T. A. & Wickstrom, E. (1991) Walking along human c-myc mRNA with antisense oligonucleotides; maximum efficiency at the 5' cap region. *Oncogene Res.* **6**, 13-19.
- Bailly, C., Henichart, J. P., Colson, P. & Houssier, C. (1992) Drug DNA sequence-dependent interactions analysed by electric linear dichroism. *J Mol Recognit.* **5**, 155-171.
- Bailly, C., Perrine, D., Lancelot, J-C., Saturnino, C., Robba, M. & Waring, M. J. (1997) Sequence selective binding to DNA of bis(amidinophenoxy)alkanes related to propamidine and pentamidine. *Biochem. J.* **323**, 23-31.

- Bailly, C., Donkor, I. O., Gentle, D., Thornalley, M. & Waring, M. (1994) Sequence selective binding to DNA of cis and trans-butamidine analogues of the anti-*Pneumocystis carinii* pneumonia drug pentamidine. *Mol. Pharmacol.* **46**, 313-322.
- Bailly, C., & Chaires, J. B. (1998) Sequence specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. *Bioconjugate Chem.* **9**, 513-536.
- Bailly C. & Hénichart, J. P. (1994) Molecular pharmacology of intercalator-minor groove binder hybrid molecules. In *Molecular Aspects of Anticancer Drug-DNA Interactions* (S. Neidle & M. J. Waring, Eds.) Vol.2, Macmillan, London, p162-196.
- Bailly, C., Catteau, J. P., Henichart, J. P., Reszka, K., Shea, R. G., Krowicki, K. & Lown, J. W. (1989) Subcellular distribution of a nitroxide spin-labeled netropsin in living KB cells. Electron paramagnetic resonance and sequence specificity studies. *Biochem. Pharmacol.* **38**, 1625-1630.
- Baker, B. F. & Monia, B. P. (1999) Novel mechanisms for antisense-mediated regulation of gene expression. *Biochim. Biophys Acta* **1489**, 3-18.
- Baker, B. F., Lot, S. S., Condon, T. P., Cheng-Flournoy, S., Lesnik, E. A., Sasmor, H. M. & Bennett, C. F. (1997) 2'-O-(2-methoxy)ethyl-modified anti-intercellular adhesion molecule 1 (ICAM-1) oligonucleotides selectively increase the ICAM-1 mRNA level and inhibit formation of the ICAM-1 translation initiation complex in human umbilical vein endothelial cells. *J. Biol. Chem.* **272**, 11994-12000.
- Baliga, R. & Crothers, D. M. (2000) On the kinetics of distamycin binding to its target sites on duplex DNA. *Proc. Natl. Acad. Sci. USA* **97**, 7814-7818.
- Baliga, R., Baird, E. E., Herman, D. M., Melander, C., Dervan, P. B. & Crothers, D. M (2001) Kinetic consequences of covalent linkage of DNA binding polyamides. *Biochemistry* **40**, 3-8.
- Baraldi, P. G., Leoni, A., Cacciari, B., Manfredini, S., Simoni, D., Bergomi, M., Menta, E. & Spinelli, S. (1994) Synthesis and antitumour activity of a new class of

pyrazolo[4,3-e]pyrrolo(1,2-a)(1,4)benzodiazepinone analogues of pyrrolo(2,1-c)(1,4)benzodiazepines. *J. Med. Chem.* **37**, 4329-4337.

Barcelo, F., Ortiz-Lombardia, M. & Portugal, J. (2001) Heterogeneous DNA binding modes of berenil. *Biochim. Biophys. Acta* **1519**, 175-184.

Basu, S. & Wickstrom, E. (1997) Synthesis and characterisation of a peptide nucleic acid conjugated to a d-peptide analog of insulin-like growth factor 1 for increased cellular uptake. *Bioconj. Chem.* **8**, 481-488.

Bathini, Y., Rao, K. E., Shea, R. G. & Lown, J. W. (1990) Molecular recognition between ligands and nucleic acids: novel pyridine and benzoxazole-containing agents related to Hoechst 33258 that exhibit altered DNA sequence specificity deduced from footprinting analysis and spectroscopic studies. *Chem. Res. Toxicol.* **3**, 268-280.

Battacharya, S. & Thomas, M. (2002) DNA binding properties of novel dansylated distamycin A analogues in which the fluorophore is directly conjugated to the N-methyl-pyrrole carboxamide backbone. *J. Biomol. Struct. Dyn.* **19**, 935-945.

Beale, P. A. & Dervan, P. B (1991) Second structural motif for recognition of DNA by oligonucleotide-directed triple helix formation. *Science* **251**, 1360-1363.

Beigelman, L., Mcswiggen, J. A., Draper, K. G., Gonzalez, C., Jensen, K., Karpeisky, A. M., Modak, A. S., Matulic-Adamic, J., Direnzo, A. B., Haeberli, P., Sweedler, D., Tracz, D., Grimm, S., Wincott, F. E., Thackaray, V. G. & Usman, N. (1995) Chemical modification of hammerhead ribozymes. *J. Biol. Chem.* **270**, 25702-25708.

Belitsky, J. M., Leslie, S. J., Paramjit, S. A., Beerman, T. A. & Dervan, P. B. (2002) Cellular uptake of N-methylpyrrole/N-methylimidazole polyamide-dye conjugates. *Bioorg. Med. Chem.* **10**, 3313-3318.

Berg, O. G. & Von Hippel, P. H. (1988) Selection of DNA binding sites by regulatory proteins. *Trends In Biochemical Sciences* **13**, 207-211.

Bergman, R. G. (1973) Reactive 1,4-dehydroaromatics. *Acc. Chem.Res.* **6**, 25-31.

- Berman, H. M., Neidle, S., Zimmer, C. & Thrum, H. (1979) netropsin, a DNA binding oligopeptide. Structural and binding studies. *Biochim. Biophys. Acta* **561**, 124-131.
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J., (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366.
- Best, T. P., Edelson, B. S., Nickols, N. G. & Dervan, P. B. (2003) Nuclear localisation of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. USA* **100**, 12063-12068.
- Betts, L., Josey, J. A., Veal, J. M. & Jordan, S. R. (1995) A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex. *Science* **270**, 1838-1841.
- Bhuyan, B. K., Smith, K. S., Adams, E. G., Petzold, G. L. & McGovren, J. P., (1992a) Lethality, DNA alkylation and cell cycle effects of adozelesin (U-73, 975) on rodent and human cells. *Cancer Res.* **52**, 5687-5692.
- Bhuyan, B. K., Smith, K. S., Adams, E. G., Wallace, T. L., Vonhoff, D. D., & Li, L. H. (1992b) Adozelesin, a potent new alkylating agent, cell killing kinetics and cell-cycle effects. *Cancer Chemother. Pharmacol.* **30**, 348-354.
- Bielawski, K., Wolczynski, S. & Bielawska, A. (2001) DNA-binding activity and cytotoxicity of the extended diphenylfuran bisamidines in breast cancer MCF-7 cells. *Biol Pharm Bull.* **24**, 704-706.
- Blackburn, G. M., (1990) DNA and RNA structure. In *Nucleic Acids In Chemistry And Biology*. G. M. Blackburn And M. J. Gaits, (Eds.), Oxford, Oxford University Press, p17-70.
- Blackstock, J. C. (1989) Nucleic acids. In *Guide To Biochemistry*. Wright, p78-90.
- Blume, S. W., Lebowitz, J., Zacharias, W., Guarcello, V., Mayfield, C. A., Ebbinghaus, S. W., Bates, P., Jones, D. E. Jr, Trent, J., Vigneswaran, N. & Miller, D. M. (1999) The integral divalent cation within the intermolecular purine\*purine.

pyrimidine structure: a variable determinant of the potential for and characteristics of the triple helical association. *Nucleic Acids Res.* **27**, 695-702.

Boffa, L. C., Morris, P. L., Carpaneto, E. M., Louissaint, M. & Allfrey, V. G. (1996) Invasion of the CAG triplet repeats by a complementary peptide nucleic acid inhibits transcription of the androgen receptor and tata binding protein genes and correlates with refolding of an active nucleosome containing a unique AR gene sequence. *J. Biol. Chem.* **271**, 13228-13233.

Boger, D. L., Munk, S. A. & Zarrinmayeh, H. (1991a) (+)-CC-1065 DNA alkylation: key studies demonstrating a noncovalent binding selectivity contribution to the alkylation selectivity. *J. Am. Chem. Soc.* **113**, 3980-3983.

Boger, D. L., Munk, S. A., Zarrinmayeh, H., Ishizaki, T., Haught, J., & Bina, M. (1991b) An alternative and convenient strategy for generation of substantial quantities of singly 5'-end labelled double stranded DNA for binding studies. Development of a protocol for examination of functional features of CC-1065 and duocarmycins that contribute to their sequence selective DNA alkylation properties. *Tetrahedron* **47**, 2661-26682.

Boger, D. L., Zarrinmayeh, H., Munk, S. A., Kitos, P. A. & Suntornwat, O. (1991c) Demonstration of a pronounced effect of noncovalent binding selectivity on the (+) CC-1065 DNA alkylation and identification of the pharmacophore of the alkylation subunit. *Proc. Natl. Acad. Sci. USA* **88**, 1431-1435.

Boger, D. L., Johnson, D. S., Palanki, M. S., Kitos, P. A., Chang, J., & Dowell, P. (1993) Evaluation of functional analogues of CC-1065 and the duocarmycins incorporating the cross-linking 9a-chloromethyl-1,2,9,9a-tetrahydrocyclopropa[c]benz[e]indole-4-one (C<sub>2</sub>BI) alkylation subunit. *Bioorg. Med. Chem.* **1**, 27-38.

Boger, D. L., Yun, W. & Han, N. (1995) 1,2,9,9a-tetrahydrocyclopropa[c]benz[e]indole-4-one (CBI) analogues of CC-1065 and the duocarmycins: synthesis and evaluation. *Bioorg. Med. Chem.* **3**, 1429-1453.

- Boger, D. L., Zhou, J. & Cai, H. (1996) Demonstration and definition of the noncovalent binding selectivity of agents related to CC-1065 by an affinity cleaving agent: noncovalent binding coincidental with alkylation. *Bioorg. Med. Chem.* **4**, 859-867.
- Boger, D. L. & Garbaccio, R. M. (1997a) catalysis of the CC-1065 and duocarmycin dna alkylation reaction: DNA binding induced conformational change in the agent results in activation. *Bioorg. Med. Chem.* **5**, 263-276.
- Boger, D. L., Boyce, C. W., Garbaccio, R. M. & Goldberg, J. A. (1997b) CC-1065 and the duocarmycins: synthetic studies. *Chem. Rev.* **97**, 787-828.
- Boger, D. L., & Garbaccio, R. M. (1999) Shape-dependent catalysis: insights into the source of catalysis for the CC-1065 and Duocarmycin DNA alkylation reaction. *Acc. Chem. Res.* **12**, 1043-1052.
- Borowy-Borowski, H., Lipman, R., Tomasz, M. (1990) Recognition between mitomycin C and specific DNA sequences for cross-link formation. *Biochemistry* **29**, 2999-3004.
- Bose, D. S., Thompson, A. S., Ching, J. S., Hartley, J. A., Berardini, M. D., Jenkins, T. C., Neidle, S., Hurley, L. H. & Thurston, D. E. (1992a) Rational design of a highly efficient irreversible DNA interstrand cross-linking agent based on the pyrrolobenzodiazepine ring system. *J. Am. Chem. Soc.* **114**, 4939-4941.
- Bose, D. S., Thompson, A. S., Smellie, M., Barardini, M. D., Hartley, J. A., Jenkins, T. C., Neidle, S. & Thurston D. E. (1992b) Effect of linker length on DNA binding affinity, cross-linking efficiency and cytotoxicity of C-8 linked pyrrolobenzodiazepine dimers. *J. Am. Chem. Soc. Chem. Commun.* **20**, 1518-1520.
- Bostock-Smith, C. E., Laughton, C. A. & Searle, M.S. (1998) DNA minor groove recognition by a tetrahydropyrimidinium analogue of Hoechst 33258: NMR and molecular dynamics studies of the complex with d(GGTAATTACC)<sub>2</sub>. *Nucleic Acids Res.* **26**, 1660-1667.
- Boykin, D. W., Kumar, A., Spychala, J., Zhou, M., Lombardy, R. J., Wilson, W. D., Dykstra, C. C., Jones, S. K., Hall, J. E., Tidwell, R. R., Laughton, C. A., Nunn,

- C. M. & Neidle, S. (1995) Dicationic diarylfurans as anti-*Pneumocystis carinii* agents. *J. Med. Chem.* **38**, 912-916.
- Bremer, R. E., Szewczyk, J. W., Baird, E. E. & Dervan, P. B. (2000) Recognition of the DNA minor groove by pyrrole-imidazole polyamides; comparison of desmethyl and N-methylpyrrole. *Bioorg. Med. Chem.* **8**, 1947-1955.
- Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986a) Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Methods Enzymol.* **130**, 132-181.
- Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986b) Footprint titrations yield valid thermodynamic isotherms. *Proc. Natl. Acad. Sci. USA* **83**, 8462-8466.
- Breslauer, K. J., Remeta, D. P., Chou, W. Y., Ferrante, R., Curry, J., Zaunczkowski, D., Snyder, J. G. & Marky, L. A. (1987) Ethalpy-entropy compensations in drug-DNA binding studies. *Proc. Natl. Acad. Sci. USA* **84**, 8922-8926.
- Briehn, C. A., Weyermann, P. & Dervan, P. B. (2003) Alternative heterocycles for DNA recognition: the benzimidazole/imidazole pair. *Chemistry* **9**, 2110-22.
- Broggini, M., Coley, H., Mongelli, N., Grandi, M., Wyatt, M. D., Hartley, J. A. & D'Incalci, M. (1995) DNA sequence specific adenine alkylation by the novel antitumour drug tallimustine (FCE 24517), a benzoyl nitrogen mustard derivative of distamycin. *Nucleic Acids Res.* **23**, 81-87.
- Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I. & Neidle, S. (1990) Crystal structure of a berenil-dodecanucleotide complex: the role of water in sequence-specific ligand binding. *EMBO J.* **9**, 1329-1334.
- Brown, D. G., Sanderson, M. R., Garman, E. & Neidle, S. (1992) Crystal structure of a berenil-d(CGCAAATTTGCG)<sub>2</sub> complex. An example of drug-DNA recognition based on sequence-dependent structural features. *J. Mol. Biol.* **226**, 481-490.



- Brown, P. M. & Fox, K. R. (1996) Nucleosome core particles inhibit DNA triple helix formation. *Biochem J.* **319**, 607-611.
- Brown, P. O. & Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21** (1 Suppl), 33-37.
- Bruice, T. C., Sengupta, D., Blasko, A., Chiang, S-Y. & Beerman, T. A. (1997) A microgonotropen branched decaaza decabutylamine and its DNA and DNA transcription factor interactions. *Bioorg. Med. Chem.* **5**, 685-692.
- Bunkenborg, J., Behrens, C. & Jacobsen, J. P. (2002) NMR characterization of the DNA binding properties of a novel Hoechst 33258 analogue peptide building block. *Bioconjug Chem.* **13**, 927-936.
- Burckhardt, G., Simon, H., Storl, K., Triebel, H., Walter, A., Lown, J. W. & Zimmer, C. (1997) DNA binding studies and influence on the activity of DNA topoisomerases of bis-netropsins: different effects of analogues containing cis and trans ethylene linkers. *J. Biomol. Struct. Dyn.* **15**, 81-95.
- Burckhardt, G., Fortsch, I., Simon, H., Birch-Hirschfeld, E., Kittler, L., Schutz, H., Sharma, S. K., Lown, J. W. & Zimmer, C. (2000) DNA sequence recognition of a cross-linked polyamide: CD studies, footprinting and effects on the activity of DNA gyrase. *J. Biomol. Struct. Dyn. Conv.* **11**, **2**, 355-363.
- Burckhardt, G., Simon, H., Birch-Hirschfeld, E., Kittler, L., Sharma, S. K., Lown, J. W. & Zimmer, C. (2002) DNA sequence recognition of thiazole-containing cross-linked polyamides can be favoured by T-A steps. *J. Biomol. Struct. Dyn.* **19**, 1101-1109.
- Cain, B. F., Atwell, G. J. & Seelye, R. N. (1969) Potential antitumour agents. X. Bisquaternary salts. *J. Med. Chem.* **12**, 199-206.
- Carrondo, M. A., Coll, M., Aymami, J., Wang, A. H., van der Marel, G. A., van Boom, J. H. & Rich, A. (1989) Binding of a Hoechst Dye to d(CGCGATATCGCG)<sub>2</sub> and its influence on the conformation of the DNA fragment. *Biochemistry* **28**, 7849-7859.

Carter, C. A., Waud, W. R., Li, L. H., De Koning, T. F., McGovren, J. P. & Plowman, J. (1996) Preclinical antitumour activity of bizelezin in mice. *Clin. Cancer. Res.* **2**, 1143-1149.

Chan, P. P. & Glazer, P. M. (1997) Triplex DNA: fundamentals, advances, and potential applications for gene therapy. *J. Mol. Med.* **75**, 267-82.

Chandra, P., Zunino, F., Götz, A., Wacker, A., Gericke, D., Di Marco, A., Casazza, A. M. & Giuliani, F. (1972) Template specific inhibition of DNA polymerases from RNA tumour viruses by distamycin and its structural analogues. *FEBS Lett.* **21**, 154-158.

Chang, D. K. & Cheng, S. F. (1996) On the importance of van der Waals interactions in the groove binding of DNA with ligands: restrained molecular dynamics study. *Int. J. Biol. Macromol.* **19**, 279-85.

Chen, S-M., Leupin, W., Rance, M. & Chazin, W. J. (1992) Two-dimensional NMR studies of d(GGTTAATGCGGT).d(ACCGCATTAACC) complexed with the minor groove binding drug SN-6999. *Biochemistry* **31**, 4406-4413.

Chen, Y. H. & Lown, J. W. (1994) A new DNA minor groove-binding motif: cross-linked lexitropsins. *J. Am. Chem. Soc.* **116**, 6995-7005.

Chen, Y-H, & Lown, J. W. (1995a) Design and synthesis of sequence-specific DNA minor groove recognising ligands of the cross-linked lexitropsins class. *Heterocycles* **41**, 1691-1707.

Chen Y. H. & Lown, J. W. (1995b) DNA minor groove binding of cross-linked lexitropsins: experimental conditions required to observed the covalently linked WPPW (groove wall-peptide-peptide groove wall) motif. *Biophys. J.* **68**, 2041-2048.

Chen, Y. H., Yang, Y., Lown, J. W. (1996) Optimisation of cross-linked lexitropsins. *J. Biomol. Struct. Dyn.* **14**, 341-355.

- Chen, J. K., Schultz, R. G., Lloyd, D. H. & Gryaznov, S. M. (1995) Synthesis of oligodeoxyribonucleotide N3'->P5' phosphoroamidates *Nucleic Acids Res.* **23**, 2661-2668.
- Chiang SY, Welch J, Rauscher FJ 3rd, Beerman TA. (1994) Effects of minor groove binding drugs on the interaction of TATA box binding protein and TFIIA with DNA. *Biochemistry* **33**, 7033-7040.
- Chiang, S. Y., Burli, R. W., Benz, C. C., Gawron, L., Scott, G. K., Dervan, P. B., Beerman, T. A. (2000) Targeting the Ets binding site of the her2/neu promoter with pyrrole-imidazole polyamides. *J. Biol. Chem.* **275**, 24246-24254.
- Chidester, C. G., Krueger, W. C., Mizasak, S. A., Duchamp, D. J., And Martin, D. B. (1981) The structure of CC-1065, a potent antitumour agent and its binding to DNA. *J. Am. Chem. Soc.* **103**, 7629-7635.
- Clanton, D. J., Buckheit, R. W., Terpening, S. J., Kiser, R., Mongelli, N., Lombardi Borgia, A., Schultz, R., Narayanan, V., Bader, J. P. & Rice, W. G. (1995) Novel sulfonated and phosphonated analogs of distamycin which inhibit the replication of HIV. *Antiviral Res.* **27**, 335-354.
- Clark, G. R., Squire, C. J., Gray, E. J., Leupin, W. & Neidle, S. (1996a) Designer DNA-binding drugs: the crystal structure of a meta-hydroxy analogue of Hoechst 33258 bound to d(CGCGAATTCGCG)<sub>2</sub>. *Nucleic Acids Res.* **24**, 4882-4889.
- Clark, G. R., Gray, E. J., Neidle, S., Li, Y. H. & Leupin W. (1996b) Isohelicity and phasing in drug-DNA sequence recognition: crystal structure of a tris (benzimidazole)-oligonucleotide complex. *Biochemistry* **35**, 13745-13752.
- Clark, G. R., Boykin, D. W., Czarny, A. & Neidle, S. (1997) Structure of a bis-amidinium derivative of Hoechst 33258 complexed to dodecanucleotide d(CGCGAATTCGCG)<sub>2</sub>: the role of hydrogen bonding in minor groove drug-DNA recognition. *Nucleic Acids Res.* **25**, 1501-1507.
- Clercq, E. D. & Dann, O. (1980) Diaryl amidine derivatives as oncornaviral DNA polymerase inhibitors. *J. Med. Chem.* **23**, 787-795.

- Coates, L., Ikpeazu, E. V., Chen, Y. & Valenzuela, M. S. (2002) Inhibition of DNA replication by berenil in plasmides containing poly(dA)poly(dT) sequences. *Plasmid* **47**, 120-128.
- Coll, M., Frederick, C. A., Wang, A. H., & Rich, A. (1987) A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTTCGC) and its complex with distamycin. *Proc. Natl. Acad. Sci. USA* **84**, 8385-8389.
- Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A., & Wang, A. H., (1989) Molecular structure of the netropsin d(CGCGATATCGCG) complex: DNA conformation in an alternating AT segment. *Biochemistry* **28**, 310-320.
- Colson, P., Houssier, C. & Bailly, C. (1995) Use of electric linear dichroism and competition experiments with intercalating drugs to investigate the mode of binding of Hoechst 33258, berenil, and DAPI to GC sequences. *J. Biomol. Struct. Dyn.* **13**, 351-366.
- Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) Site-specific oligonucleotide binding represses transcription of the human c-myc gene *in vitro*. *Science* **241**, 456-459.
- Cory, M., Tidwell, R. R. & Fairley, T. A. (1992) Structure and DNA binding activity of analogues of 1,5-bis(4-amidinophenoxy)pentane (pentamidine). *J. Med. Chem.* **35**, 431-438.
- Coull, J. J., He, G., Melander, C., Rucker, V. C., Dervan, P. B. & Margolis, D. M. (2002) Targeted derepression of the human immunodeficiency virus type 1 long terminal repeat by pyrrole-imidazole polyamides. *J. Virol.* **76**, 12349-12354.
- Cristofanilli, M., Bryan, W. J., Miller, L. L., Chang, A. Y., Gradishar, W. J., Kufe, D. W., & Hortobagyi, G. N. (1998) Phase II study of adozelesin in untreated metastatic breast cancer. *Anticancer Drugs* **9**, 779-782.
- Crooke, S. T. (1998) Molecular mechanisms of antisense drugs: RNase H. *Antisense Nucleic Acid Drug Dev.* **8**, 133-134.

Crooke, S. T. (1999) Molecular mechanisms of action of antisense drugs. *Biochim. Biophys. Acta* **1489**, 31-44.

Crowley, K.S., Phillion, D. P., Woodard, S. S., Schweitzer, B. A., Singh, M., Shabany, H., Burnette, B., Hippenmeyer, P., Heitmeyer, M. & Bashkin, J. K. (2003) Controlling the intracellular localisation of fluorescent polyamide analogues in cultured cells. *Bioorg. Med. Chem.* **13**, 1565-1570.

Cutrona, G., Carpaneto, E. M., Ulivi, M., Roncella, S., Landt, O., Farrarini, M. & Boffa, L. C. (2000) Effects in live cells of a c-myc anti-gene PNA linked to a nuclear localisation signal. *Nat. Biotechnol.* **18**, 300-303.

Czarny, A., Boykin, D. W., Wood, A. A., Nunn, C. M., Neidle, S., Zhao, M. & Wilson, W. D. (1995) Analysis of van der Waals and electrostatic contributions in the interactions of minor groove binding benzimidazoles with DNA. *J. Am. Chem. Soc.* **117**, 4716-4717.

Dai, W. M., Lai, K. W., Wu, A. X., Hamaguchi, W., Lee, M. Y. H., Zhou, L., Ishii, A., & Nishimoto, S. (2002) DNA cleavage potency, cytotoxicity and mechanism of action of a novel class of enediyne prodrugs. *J. Med. Chem.* **45**, 758-761.

Darnell, J. E. (2002) Transcriptional factors as targets for cancer therapy. *Nat. Rev. Cancer* **2**, 740-749.

Dedon, P. C. & Goldberg, I. H. (1990) Sequence-specific double-strand breakage of DNA by neocarzinostatin within a staggered cleavage site. *J. Biol. Chem.* **265**, 14713-14716.

Dedon, P. C., Jiang, Z. W. & Goldberg, I. H. (1992) Neocarzinostatin-mediated DNA damage in a model AGT.ACT site: mechanistic studies of thiol-sensitive partitioning of C-4' DNA damage products. *Biochemistry* **31**, 1917-1927.

Deininger, M. W. N., Goldman, J. M. & Melo, J. (2000) The molecular biology of chronic myeloid leukaemia. *Blood* **96**, 3343-3356.

- Demidov, V. V., Yavnilovich, M. V., Belotserkovskii, B. P., Frank-Kamenetskii, M. D. & Nielsen, P. E. (1995) Kinetics and mechanism of polyamide ('peptide') nucleic acid binding to duplex DNA. *Proc. Natl. Acad. Sci. USA* **92**, 2637-2641.
- Denny, W. A., Atwell, G. J., Baguley, B. C. & Cain, B. F. (1979) Potential antitumour agents. 29. Quantitative structure-activity relationships for the antileukaemic bisquaternary ammonium heterocycles. *J. Med. Chem.* **22**, 134-150.
- Dervan, P.B. & Burli, R. W. (1999) Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.* **3**, 688-693.
- Dervan, P. B. (2001) Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **9**, 2215-2235.
- Dias, N. & Stein, C. A. (2002) Antisense oligonucleotides: basic concepts and mechanisms. *Mol. Cancer. Ther.* **1**, 347-355.
- Dickerson, R. E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **26**, 1906-1926.
- Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M. & Dervan, P. B. (1998) Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. USA* **95**, 12890-12895.
- Dickinson, L. A., Trauger, J. W., Baird, E. E., Ghazal, P., Dervan, P. B. & Gottesfeld, J. M. (1999a) Anti-repression of RNA polymerase II transcription by pyrrole-imidazole polyamides. *Biochemistry* **38**, 10801-10807.
- Dickinson, L. A., Trauger, J. W., Baird, E. E., Dervan, P. B., Graves, B. J. & Gottesfeld, J. M (1999b) Inhibition of Ets-1 DNA binding and ternary complex formation between Ets-1, NF- $\kappa$ B, and DNA by a designed DNA-binding ligand. *J. Biol. Chem.* **274**, 12765-12773.
- Ding, Z. M., & Hurley, L. H., (1991) DNA interstrand cross-linking, DNA sequence specificity, and induced conformational changes produced by dimeric analogue of (+)-CC-1065. *Anticancer Drug Des.* **6**, 427-452.

- Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J., Lydon, N. B. (1996) Effects of a selective inhibitor of the abl tyrosine kinase on the growth of bcr-abl positive cells. *Nat. Med.* **2**, 561-566.
- Dudouet, B., Burnett, R., Dickinson, L. A., Wood, M. R., Melander, C., Belitsky, J. M., Edelson, B., Wurtz, N. R., Briehn, C., Dervan, P. B. & Gottesfeld, J. M. (2003) Accessibility of nuclear chromatin by DNA binding polyamides. *Chem. Biol.* **10**, 859-867.
- Dwyer, T. J., Geierstanger, B. H., Bathini, Y., Lown, J. W., Wemmer, D. E. (1992) Design and binding of a distamycin A analogue to d[CGCAAGTT]:d[GCCAACTTGCG]: synthesis, NMR studies, and implications for the design of sequence-specific minor groove binding oligopeptides. *J. Am. Chem. Soc.* **114**, 5911-5919.
- Dwyer, T. J., Geierstanger, B. H., Mrksich, M., Dervan P. B., Wemmer, D. E. (1993) Structural analysis of covalent peptide dimers, bis(pyridine-2-carboxamidonetropsin)(CH<sub>2</sub>)<sub>3-6</sub>, in complex with 5-TGACT-3' sites by two-dimensional NMR. *J. Am. Chem. Soc.* **115**, 9900-9906.
- Ebrahimi, S. E. S., Parkinson, J. A., Fox, K. R., McKie, J. H., Barber, J. & Douglas, K. T. (1992) Studies of the interaction of a meta-hydroxy analogue of Hoechst 33258 with DNA by melting temperature, footprinting and high-resolution <sup>1</sup>H NMR spectroscopy. *J. Chem. Soc. Chem. Commun.*, 1398-1399.
- Edelson, B. S., Best, T. P., Olenvuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A. & Dervan, P. B. (2004) Influence of structural variation on nuclear localisation of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res.* **32**, 2802-2818.
- Eder, P. S., DeVine, R. J., Dagle, J. M. & Walder, J. A. (1991) Substrate specificity and kinetics of degradation of antisense oligonucleotides by a 3' exonuclease in plasma. *Antisense Res. Dev.* **1**, 141-151.



- Edo, K., Mizugake, M., Koide, Y., Seto, H., Furihata, K., Otake, N. & Ishida, N. (1985) The structure of neocarzinostatin chromophore possessing a novel bicyclo-[7,3,0]dodecadiyne system. *Tetrahedron Lett.* **26**, 331-334.
- Edwards, K. J., Jenkins, T. C. & Neidle, S. (1992) Crystal structure of a pentamidine-oligonucleotide complex: implications for DNA-binding properties. *Biochemistry* **31**, 7104-7109.
- Egholm, M., Buckhardt, O., Christensen, L., Behrens, C., Frier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B., & Nielsen, P. E. (1993) PNA hybridises to complementary oligonucleotides obeying the Watson-Crick hydrogen bonding rules. *Nature* **365**, 566-568.
- Ehley, J. A., Melander, C., Herman, D., Baird, E. E., Ferguson, H. A., Goodrich, J. A., Dervan, P. B., & Gottesfeld, J. M. (2002) Promoter scanning for transcriptional inhibition with DNA-binding polyamides. *Mol. Cell. Biol.* **22**, 1723-1733.
- El-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, Lin D, Mercer WE, Kinzler KW, Vogelstein B. (1993) WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817-825.
- Elbashir, S. M., Lendeckel, W. & Tuschl, T. (2001a) RNA interference is mediated by 21 and 22 nucleotide RNAs. *Genes Dev.* **15**, 188-200.
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. & Tuschl, T. (2001b) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-498.
- Ellervik, U., Wang, C. C. C. & Dervan, P. B. (2000) Hydroxybenzamide/pyrrole pair distinguishes T.A from A.T base pairs in the minor groove of DNA. *J. Am. Chem. Soc.* **122**, 9354-9360.
- Elmroth, K., Nygren, J., Martensson, S., Ismail, I. H., & Hammarsten, O. (2003) Cleavage of cellular DNA by calicheamicin  $\gamma$ 1. *DNA Repair* **2**, 363-374.

- Embrey, K. J., Searle, M. S. & Craik, D. J. (1993) Interaction of Hoechst 33258 with the minor groove of the A + T-rich DNA duplex d(GGTAATTACC)<sub>2</sub> studied in solution by NMR spectroscopy. *Eur. J. Biochem.* **211**, 437-447.
- Espinas, M. L., Jimenez-Garcia, E., Martinez-Balbas, A. & Azorin, F. (1996) Formation of triple-stranded DNA at d(GA.TC)<sub>n</sub> sequences prevents nucleosome assembly and is hindered by nucleosomes. *J. Biol. Chem.* **271**, 31807-31812.
- Evans, W. E. & Relling, M. V. (1999) Pharmacogenomics: translating functional genomics into rational therapies. *Science* **286**, 487-491.
- Fagan, P. & Wemmer, D. E. (1992) Cooperative binding of distamycin A to DNA in the 2:1 mode. *J. Am. Chem. Soc.* **114**, 1080-1081.
- Fairfield, F. R., Bauer, W. R. & Simpson, M. V. (1979) Mitochondria contain a distinct DNA topoisomerase. *J. Biol. Chem.* **254**, 9352-9354.
- Fechter, E. J. & Dervan, P. B. (2003) Allosteric inhibition of protein-DNA complexes by polyamide-intercalator conjugates. *J. Am. Chem. Soc.* **125**, 8476-8485.
- Fede, A., Labhardt, A., Bannwarth, W. & Leupin, W. (1991) Dynamics and binding mode of Hoechst 33258 to d(GTGGAATTCCAC)<sub>2</sub> in the 1:1 solution complex as determined by two-dimensional <sup>1</sup>H NMR. *Biochemistry* **30**, 11377-11388.
- Felsenfeld, G., Davies, D. R. & Rich, A. (1957) Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **79**, 2023-2024.
- Fennewald, S. M. & Rando, R. F. (1995) Inhibition of high affinity basic fibroblast growth factor binding by oligonucleotides. *J. Biol. Chem.* **270**, 21718-21721.
- Finlay, A. C., Hochstein, F. A., Sobin, B. A. & Murphy, F. X. (1951) Netropsin, a new antibiotic produced by a *streptomyces*. *J. Am. Chem. Soc.* **73**, 314-343.
- Finlay, C. A., Hinds, P. W. & Levine, A. J., (1989) The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57**, 1083-1093.

- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.
- Fitzgerald, D. J. & Anderson, J. N. (1999) Selective nucleosome disruption by drugs that bind in the minor groove of DNA. *J. Biol. Chem.* **274**, 27128-27138.
- Foister, S., Marques, M. A., Doss, R. M. & Dervan, P. B. (2003) Shape selective recognition of TA base pairs by hairpin polyamides containing N-terminal 3-methoxy (and 3-chloro) thiophene residues. *Bioorg. Med. Chem.* **11**, 4333-4340.
- Fojo, T. & Bates, S. (2003) Strategies for reversing drug resistance. *Oncogene* **22**, 7512-7523.
- Foulds, L. (1954) The experimental study of tumour progression. Volumes I-II, London: Academic Press.
- Fox, K. R., & Waring, M. J. (1984) DNA structural variations produced by actinomycin and distamycin as revealed by DNase I footprinting. *Nucleic Acids Res.* **12**, 9271-9285.
- Fox, K. R. & Howarth, N. R. (1985) Investigations into the sequence-selective binding of mithramycin and related ligands to DNA. *Nucleic Acids Res.* **13**, 8695-8714.
- Fox, K. R., Sansom, C. E. & Stevens, M. F. (1990) Footprinting studies on the sequence-selective binding of pentamidine to DNA. *FEBS Lett.* **266**, 150-154.
- Fox, K. R., Allinson, S. L., Sahagun-Krause, H. & Brown, T. (2000) Recognition of GT mismatches by vsr mismatch endonuclease. *Nucleic Acids Res.* **28**, 2535-2540.
- Fox, M. & Scott, G. (1980) The genetic toxicology of nitrogen and sulphur mustard. *Mutat. Res.* **75**, 131-168.
- Frank-Kamenetskii, M. D. & Mirkin, S. M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65-95.

- Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATT<sup>B</sup>CGCG. *J. Biol. Chem.* **257**, 14686-14707.
- Frederick, C. A., Williams, L. D., Ughetto, G., Van Der Marel, G. A. & Van Boom, J. H. (1990) Structural comparison of anticancer-DNA complexes: adriamycin and daunomycin. *Biochemistry* **29**, 2538-2549.
- Friedberg, E. C., Walker, G. C. & Siede, W. (1995) DNA damage. In *DNA Repair And Mutagenesis*, ASM Press Washington DC, p1-47.
- Gao, X. L. & Patel, D. (1989) Solution structure of the chromomycin-DNA complex. *Biochemistry* **28**, 751-762.
- Gao, X. L., Mirau, P. & Patel, D. J. (1992) Structure refinement of the chromomycin dimer-DNA oligomer complex in solution. *J. Mol. Biol.* **223**, 259-79.
- Gao, Y. G., Sriram, M., Denny, W. A. & Wang, AH-J. (1993) Minor groove binding of SN-6999 to an alkylated DNA: molecular structure of d(CGC[<sup>ε</sup>G]AATTCGCG)-SN6999 complex. *Biochemistry* **32**, 9639-9648.
- Gavathiotis, E., Sharman, G. J. & Searle, M. S. (2000) Sequence-dependent variation in DNA minor groove width dictates orientational preference of Hoechst 33258 in A-tract recognition: solution NMR structure of the 2:1 complex with d(CTTTTGCAAAAG)<sub>2</sub>. *Nucleic Acids Res.* **28**, 728-735.
- Gee, J. E., Revankar, G. R., Rao, T. S. & Hogan, M. E. (1995) Triplex formation at the rat neu gene utilizing imidazole and 2'-deoxy-6-thioguanosine base substitutions. *Biochemistry* **34**, 2042-2048.
- Geierstanger, B. H., Dwyer, T. J., Bathini, Y., Lown, J. W. & Wemmer, D. E. (1993) NMR characterisation of a heterocomplex formed by distamycin and its analogue 2-ImD With d(CGCAAGTTGGC):d(GCCAAGTTGCG): preference for 1: 1: 1 2-Imd:Dst:DNA complex over the 2:1 2-Imd:DNA and the 2:1 Dst:DNA complexes. *J. Am. Chem. Soc.* **115**, 4474-4482.

- Geierstanger, B. H., Jacobsen, J. P., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994a) Structural and dynamic characterisation of the heterodimeric and homodimeric complexes of distamycin and 1-methylimidazole-2-carboxide-netropsin bound to the minor groove of DNA. *Biochemistry* **33**, 3055-3062.
- Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994b) Design of a GC-specific DNA minor groove-binding peptide. *Science* **266**, 646-650.
- Geierstanger, B. H. & Wemmer, D. E. (1995) Complexes of the minor groove of DNA. *Ann. Rev. Biophys. Biomol. Struct.* **24**, 463-493.
- Giovannangeli, C., Rougee, M., Garestier, T., Thuong, N. T. & Hélène C. (1992) Triple-helix formation by oligonucleotides containing the three bases thymine, cytosine, and guanine. *Proc. Natl. Acad. Sci. U S A* **89**, 8631-8365.
- Golik, J., Clardy, J., Dubay, G., Groenewold, G., Kawaguchi, H., Konishi, M., Krishnan, B., Ohkuma, H., Saitoh, K. & Doyle, T. W. (1987a) Esperamicins, a novel class of potent antitumor antibiotics. 2. Structure of esperamicin X. *J. Am. Chem. Soc.* **109**, 3461-3462.
- Golik, J., Dubay, G., Groenewold, G., Kawaguchi, H., Konishi, M., Krishnan, B., Ohkuma, H., Saitoh, K. & Doyle, T. W. (1987b) Esperamicins, a novel class of potent antitumor antibiotics. 3. Structures of esperamicins A1, A2, and A1b. *J. Am. Chem. Soc.* **109**, 3462-3464.
- Gonzalez, V. M., Amo-Ochoa, P., Perez, J. M., Fuertes, M. A., Masaguer, J. R., Navarro-Ranninger, C. & Alonso C. (1996) Synthesis, characterization and DNA modification induced by a novel Pt-berenil compound with cytotoxic activity. *J. Inorg. Biochem.* **63**, 57-68.
- Gonzalez, V. M., Fuertes, M. A., Jimenez-Ruiz, A., Alonso, C. & Perez, J. M. (1999) The formation of DNA interstrand cross-links by a novel bis-[Pt<sub>2</sub>Cl<sub>4</sub>(diminazene aceturate)<sub>2</sub>]Cl<sub>4</sub>·4H<sub>2</sub>O complex inhibits the B to Z transition. *Mol Pharmacol.* **55**, 770-777.
- Goodsell, D. S. & Dickerson, R. E. (1986) Isohelical analysis of DNA groove-binding drugs. *J. Med. Chem.* **29**, 727-733.

- Goodsell, D. S., Kaczor-Grzeskowiak, M. & Dickerson, R. E. (1994) The crystal structure of C-C-A-T-T-A-A-T-G-G: implications for bending of B-DNA at T-A steps. *J. Mol. Biol.* **239**, 79-96.
- Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. (1997) Regulation of gene expression by small molecules. *Nature* **387**, 202-205.
- Gottesfeld, J. M., Melander, C., Suto, R. K., Raviol, H., Luger, K. & Dervan P. B. (2001) Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides. *J. Mol. Biol.* **309**, 615-629.
- Gottesfeld, J. M., Belitsky, J. M., Melander, C., Dervan, P. B. & Luger, K. (2002) Blocking transcription through a nucleosome with synthetic DNA ligands. *J. Mol. Biol.* **321**, 249-263.
- Gowers, D. M. & Fox, K. R. (1999) Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res.* **27**, 1569-1577.
- Gray, G. D., Basu, S. & Wickstrom, E. (1997) Transformed and immortalised cellular uptake of oligodeoxynucleoside phosphorothioates, 3'-alkylamino oligodeoxynucleotides, 2'-o-methyl oligoribonucleotides, oligodeoxynucleoside methylphosphonates and peptide nucleic acids. *Biochem. Pharmacol.* **53**, 1465-1476.
- Gregson, S. J., Howard, P. W., Jenkins, T. C., Kelland, L., R. & Thurston, D. E. (1999) Synthesis of a novel C2/C2'-*exo* unsaturated pyrrolbenzodiazepine cross-linking agent with remarkable DNA binding affinity and cytotoxicity. *Chem. Commun.*, 797-798.
- Gregson, S. J., Howard, P. W., Hartley, J. A., Brooks, N. A., Adams, L. J., Jenkins, T. C., Kelland, L. R. & Thurston, D. E. (2001) Design, synthesis, and evaluation of a novel pyrrolbenzodiazepine DNA-interactive agent with highly efficient cross-linking ability and potent cytotoxicity. *J. Med. Chem.* **44**, 737-748.
- Grisson, J. W., Gunawardena, G. U., Klingberg, D. & Huang, D. (1996) The chemistry of enediynes, enyne allenes and related compounds. *Tetrahedron* **52**, 6453-6518.

- Gryaznov, S. M., Lloyd, D. H., Chen, J. K., Schultz, R. G., Dedionisio, L. A. Ratmeyer, L., & Wilson, W. D. (1995) Oligonucleotide N3'->P5' phosphoroamidates. *Proc. Natl. Acad. Sci. USA* **92**, 5798-5802.
- Gryzanov, S. M. (1999) Oligonucleotide N3'-> P5' phosphoroamidates as potential therapeutic agents. *Biochim. Biophys. Acta* **1489**, 131-140.
- Guntaka, R. V., Varma, B. R. & Weber, K. T. (2003) Triplex-forming oligonucleotides as modulators of gene expression. *Int. J. Biochem. Cell Biol.* **35**, 22-31.
- Guo, D., Gupta, R. & Lown, J. W. (1993) DNA sequence-selective binding of head-to-tail-linked bis-lexitropsins: relation of phasing to cytotoxic potency. *Anticancer Drug Des.* **8**, 369-397.
- Guvakova, M. A., Yakubov, L. A., Vlodavsky, I., Tonkinson, J. L. & Stein C. A. (1995) Phosphorothiate oligodeoxynucleotides bind to basic fibroblast growth factor, inhibit its binding to cell surface receptors, and remove it from low affinity binding sites on extracellular matrix. *J. Biol. Chem.* **270**, 2620-2627.
- Gygi, M. P., Ferguson, M. D., Mefford, H. C., Lund, K. P., O'Day, C., Zhou, P., Friedman, C., Van Den Engh, G., Stolowitz, M. L. & Trask, B. (2002) Use of fluorescent sequence-specific polyamides to discriminate human chromosomes by microscopy and flow cytometry. *Nucleic Acids Res.* **30**, 2790-2799.
- Hamma, T. & Miller, P. S. (1999) Syntheses of alternating oligo-2'-O-methylribonucleoside methylphosphonates and their interactions with HIV TAR RNA. *Biochemistry* **38**, 15333-15342.
- Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293-296.
- Hanahan, D. & Weinberg, R. A. (2000) The hallmarks of cancer. *Cell* **100**, 57-70.
- Hanka, L. J., Dietz, A., Gerpheide, S. A., Kuentzel, S. L. & Martin, D. G. (1978) CC-1065 (NSC-298223), a new antitumour antibiotic: production, *in vitro*



biological activity, microbiological assays and taxonomy of the producing microorganisms. *J. Antibiot.* **31**, 1211-1217.

Hannon, G. J. (2002) RNA interference. *Nature* **418**, 244-251.

Hanvey, J. C., Peffer, N. J., Bisi, J. E., Thomson, S. A. Cadilla, R. & Josey, J. A. (1992) Antisense and antigene properties of peptide nucleic acids. *Science* **258**, 1481-1485.

Haq, I., Ladbury, J. E., Chowdhry, B. Z., Jenkins, T. C. & Chaires, J. B. (1997) Specific binding of Hoechst 33258 to the d(CGCAAATTTGCG)<sub>2</sub> duplex: calorimetric and spectroscopic studies. *J. Mol. Biol.* **271**, 244-257.

Hare, D., Shapiro, L. & Patel, D. J. (1986) Wobble dG.dT pairing in the right-handed DNA: solution conformation of the d(C-G-T-G-A-A-T-T-C-G-C-G) duplex deduced from distance geometry analysis of nuclear overhauser effect spectra. *Biochemistry* **25**, 7445-7456.

Harper, J. W., Adami, G., Wei, N., Keyomarsi, K. & Elledge, S. (1993) The p21 cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805-816.

Harshman, K. D., & Dervan, P. B., (1985) Molecular recognition of B-DNA by Hoechst 33258. *Nucleic Acids Res.* **13**, 4825-4835.

Hartley, J. A., Lown, J. W., Mattes, W. B., & Kohn, K. W. (1988) DNA sequence specificity of antitumour agents. Oncogenes as possible targets for cancer therapy. *Acta Oncol.* **27**, 503-510.

Hartley, J. A. & Souhami, R. L. (1993) DNA-sequence selectivity of anticancer agents. In *Cancer Chemotherapy*. Hickman, J. A. & Tritton, T. R. (Eds), London, Blackwell Scientific Publications, p251-280.

Hartley, J. A. (1993b) Selectivity in alkylating agent-DNA interactions. In *Molecular Aspects of Anticancer Drug-DNA Interactions*. Volume 1, Neidle, S. & Waring, M. (Eds), London, Macmillan Press, p1-22.

- Hartley, J. A., McAdam, S. R., Das, S., Roldan, M. C., Haskell, M. K. & Lee, M. (1994) Molecular and cellular pharmacology of novel photoreactive psoralen and coumarin conjugates of pyrrole- and imidazole-containing analogues of netropsin. *Anticancer Drug Des.* **9**, 181-197.
- Heckel, A. & Dervan P. B (2003) U-pin polyamide motif for recognition of the DNA minor groove. *Chem. Eur. J.* **9**, 3353-3366.
- Hélène, C. (1991) The anti-gene strategy: control of gene expression by triplex-forming-oligonucleotides. *Anticancer Drug Des.* **6**, 569-584.
- Herman, D. M., Turner, J. M., Baird, E. E., & Dervan, P. B. (1999a) Cycle polyamide motif for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **121**, 1121-1129.
- Herman, D. M., Baird, E. E., & Dervan, P. B. (1999b) Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Eur. J.* **5**, 975-983.
- Hertzberg, R. P., Hecht, S. M., Reynolds, V. L., Molineux, I. J., & Hurley, L. H., (1986) DNA sequence specificity of the pyrrolo[1,4]benzodiazepine antitumour antibiotics, methidium propyl-EDTA-iron(II) footprinting analysis of the DNA binding sites for anthramycin related drugs. *Biochemistry* **25**, 1249-1258.
- Ho, S. P., Britton, D. H. O., Stone, B. A., Behrens, D. L., Leffet, L. M., Hobbs, F. W., Miller, J. A. & Trainor, G. L. (1996) Potent antisense oligonucleotides to the human multidrug resistance-1 mRNA rationally selected by mapping RNA-accessible sites with oligonucleotide libraries. *Nucleic Acids Res.* **24**, 1901-1907.
- Horobin, R. W. & Rashid, F. (1990) Interactions of molecular probes with living cells and tissues. Part 1. Some general mechanistic proposals, making use of a simplistic chinese box model. *Histochemistry* **94**, 205-209.
- Hudziak, R. M., Barofsky, E., Barofsky, D. F., Weller, D. L., Huang, S. B. & Weller D. D., (1996) Resistance of morpholino phosphorodiamidate oligomers to enzymatic degradation. *Antisense Nucleic Acid Drug Dev.* **6**, 267-272.

- Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D. & Kennard, O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5Å resolution. *J. Biol. Chem.*, **262**, 9962-9970.
- Hurley, L. H. (1977) Pyrrolo(1,4)benzodiazepine antitumour antibiotics. Comparative aspects of anthramycin, tomayamycin and sibiromycin. *J. Antibiot.* **30**, 349-370.
- Hurley, L. H., Reynolds, V. L., Swenson, D. H., Petzold, G. L., & Scahill, T. (1984) Reaction of the antitumour antibiotic CC-1065 with DNA: the structure of a DNA adduct with sequence specificity. *Science* **226**, 843-844.
- Hurley, L. H. & Thurston, D. E. (1984) Pyrrolo[1,4]benzodiazepine antitumour antibiotics: chemistry, interaction with DNA and biological implications. *Pharm. Res.* **2**, 52-59.
- Hurley, L. H., & Needham-Vandevanter, D. R., (1986) Covalent binding of antitumour antibiotics in the minor groove of DNA, mechanism of action of CC-1065 and pyrrolo(1,4)benzodiazepenes. *Acc. Chem. Res.* **19**, 230-237.
- Hurley, L. H., Lee, C. S., McGovren, J. P., Mitchell, M. A., Warpehoski, M. A., Kelly, R. C. & Aristoff, P. A (1988). Molecular basis for sequence-specific alkylation by CC-1065. *Biochemistry* **27**, 3886-3892.
- Hurley, L. H. (2002) DNA and its associated processes as targets for cancer therapy. *Nature Reviews Cancer* **2**, 188-200.
- Iyer, V. N. & Szybalski, W. (1963) A molecular mechanism of mitomycin action: linking of complementary DNA strands. *Proc. Natl. Acad. Sci. USA* **50**, 355- 362.
- Izant, J. G. & Weintraub, H. (1984) Inhibition of thymidine kinase gene expression by antisense RNA: a molecular approach to genetic analysis. *Cell* **36**, 1007-1015.
- Jaaskelainen, I. & Urtti, A. (2002) Cell membranes as barriers for the use of antisense therapeutic agents. *Mini Rev. Med. Chem.* **2**, 307-318.

- Jackson, B. A., Alekseyev, V. Y. & Barton, J. K. (1999) A versatile mismatch recognition agent: specific cleavage of a plasmid DNA at a single base mispair. *Biochemistry* **38**, 4655-4662.
- Jackson, B. A. & Barton, J. K. (2000) Recognition of base mismatches in DNA by 5, 6-chrysenequinone diimine complexes of rhodium (II): a proposed mechanism for preferential binding in destabilised regions of the double helix. *Biochemistry* **39**, 6176-6182.
- Jansen, K., Lincoln, P. & Nordén, B. (1993) Binding of DAPI analogue 2,5-bis(4-amidinophenyl)furan to DNA. *Biochemistry* **32**, 6605-6012.
- Jansen, B. & Zangemeister-Wittke, U. (2002) Antisense therapy for cancer- the time of truth. *Lancet Oncol.* **3**, 672-683.
- Janssen, S., Durussel, T. & Laemmli, U. K. (2000a) Chromatin opening of DNA satellites by targeted sequence-specific drugs. *Mol. Cell* **6**, 999-1011.
- Janssen, S., Cuvier, O., Muller, M., & Laemmli, U. K., (2000b) Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to *Drosophila melanogaster*. *Mol. Cell* **6**, 1013-1024.
- Jenkins, T. C., Lane, A. N., Neidle, S. & Brown, D. G. (1993) NMR and molecular modelling studies of the interaction of berenil and pentamidine with d(CGCAAATTTGCG)<sub>2</sub>. *Eur. J. Biochem.* **213**, 1175-1184.
- Jenkins, T. C., Hurley, L. H., Neidle, S. & Thurston, D. E. (1994) Structure of the covalent DNA minor groove adduct with a pyrrolobenzodiazepine dimer: evidence for sequence-specific interstrand cross-linking. *J. Med. Chem.* **37**, 4529-4537.
- Jensen, K. K., Orum, H., Nielsen, P. E. & Nordén, B. (1997) Kinetics for hybridisation of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry* **36**, 5072-5077.
- Ji, Y-H., Bur, D., Häsler, W., Runtz Schmitt, V., Dorn, A., Bailly, C., Waring, M. J., Hochstrasser, R. & Leupin, W. (2001) Tris-benzimidazole derivatives: design, synthesis and DNA sequence recognition. *Bioorg. Med.Chem.* **9**, 2905-2919.

- Jones, G. B., Davey, C. L., Jenkins, T. C., Kamal, A., Kneale, G. G., Neidle, S., Webster, G. D. & Thurston, D. E. (1990) The non-covalent interaction of pyrrolo[2,1-c][1,4]benzodiazepine-5,11-diones with DNA. *Anticancer Drug Des.* **5**, 249-264.
- Juliano, R. L. & Ling, V. (1976) A surface glycoprotein modulating drug permeability in chinese hamster ovary cell mutants. *Biochim. Biophys. Acta.* **455**, 152-62.
- Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikarov, I., Berk, A. J. & Dickerson, R. E. (1996) How proteins recognise the TATA box. *J. Mol. Biol.* **261**, 239-254.
- Kato, Y., Kuwabara, T., Warashina, M., Toda, H. & Taira, K. (2001) Relationships between the activities *in vitro* and *in vivo* of various kinds of ribozyme and their intracellular localisation. *J. Biol. Chem.* **276**, 15378-15385.
- Kelly, J. J. & Dervan, P. B. (1996) Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl. Acad. Sci. USA* **93**, 6981-6985.
- Kers, I. & Dervan, P. B. (2002) Search for the optimal linker in tandem hairpin polyamides. *Bioorg. Med. Chem.* **10**, 3339-3349.
- Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J. & Plasterk, R. H., (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654-2659.
- Khorlin, A. A., Krylov, A. S., Grokhovsky, S. L., Zhuze, A. L., Zasedatelev, A. S., Gursky, G. V., & Gottikh, B. P. (1980) A new type of AT-specific ligand constructed of 2 netropsin-like molecules. *FEBS Lett.* **118**, 311-314.
- Kielkopf, C. L., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998a) Structural basis for GC recognition in the DNA minor groove. *Nat. Struct. Biol.* **5**, 104-109.
- Kielkopf, C. L., White, S., Szewczyk, J. M., Turner, J. M., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998b) A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **282**, 111-115.

- Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. M., Turner, J. M., Baird, E. E., Dervan, P. B. & Rees, D. C. (2000) Structural effects of DNA sequence on T.A recognition by hydroxypyrrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **295**, 557-67.
- Kim, S. K., Eriksson, S., Kubista, M., & Norden, B. (1993) Interaction of 4', 6-diamidino-2-phenylindole (DAPI) with poly[d(G-C)<sub>2</sub>] and poly[d(G-m<sup>5</sup>C)<sub>2</sub>]: evidence for major groove binding of a DNA probe. *J. Am. Chem. Soc.* **115**, 3441-3447.
- Kim, J. S., Gatto, B., Yu, C., Liu, A., Liu, L. F. & LaVoie, E. J., (1996) Substituted 2,5'-bi-1H-benzimidazoles; topoisomerase I inhibition and cytotoxicity. *J. Med. Chem.* **39**, 992-998.
- Kissinger, K. L., Krowicki, K., Dabrowiak, J. C., & Lown, J. W. (1987) Molecular recognition between oligopeptides and nucleic acids: monocationic lexitropsins that display enhanced GC sequence dependent DNA binding. *Biochemistry* **26**, 5590-5595.
- Kissinger, K. L., Dabrowiak, J. C., & Lown, J. W. (1990) Molecular recognition between oligopeptides and nucleic acids: DNA binding specificity of a series of bis-netropsin analogues deduced from footprinting analysis. *Chem. Res. Toxicol.* **3**, 162-168.
- Kittler, L., Wähnert, U., Baguley, B. C., Bailly, C., Waring, M. J. & Löber, G. (1996) Sequence-specific binding of antitumour bisquaternary ammonium heterocycles to DNA and inhibition of polymerase activity *in vitro*. *Anticancer Drug Des.* **11**, 101-115.
- Klevit, R. E., Wemmer, D. E. & Reid, B. R. (1986) <sup>1</sup>H-NMR studies on the interaction between distamycin a and a symmetrical DNA dodecamer. *Biochemistry* **25**, 3296-3303.
- Knight, S. W. & Bass, B. L. (2001) A role for the RNase II enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269-2271.

- Kohn, K. W., Spears, C. L. & Doty, P. (1966) Interstrand cross-linking of DNA by nitrogen mustard. *J. Mol. Biol.* **19**, 266-288.
- Kohn, K. W., Bono, V. H. & Kann, H. E., (1968) Anthramycin, a new type of DNA-inhibiting antibiotic. Reaction with DNA and the effect on nucleic acid synthesis in mouse leukaemia cells. *Biochim. Biophys. Acta* **155**, 121-129.
- Konishi, M., Ohkuma, H., Matsumoto, K., Tsuno, T., Kamei, H., Miyaki, T., Oki, T., Kawaguchi, H., VanDuyne, G. D. & Clardy, J. (1989) Dynemicin A, a novel antibiotic with the anthraquinone and 1,5-diyne-3-ene subunit. *J. Antibiot.* **42**, 1449-1452.
- Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. (1983) Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **163**, 129-146.
- Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E., (1985a) Binding of an antitumour drug to DNA, netropsin and C-G-C-G-A-A-T-T-<sup>Br</sup>C-G-C-G. *J. Mol. Biol.* **183**, 553-563.
- Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E., (1985b) The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. USA* **82**, 1376-80.
- Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W., Dickerson, R. E. (1997) Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure* **5**, 1033-1046.
- Kopka, M. L., Han, G. W., Chiu, T. K., Walker, W. L., Lown, J. W. & Dickerson, R. E. (1998) DNA sequence recognition in the minor groove by polyamides, using a GC-Specific reading element: a perspective from crystallography. In *Structure, Motion, Interaction & Expression of Biological Macromolecules (Proceedings of the Tenth Conversation in Biomolecular Sterodynamics)* R. H. Sarma and M. H. Sarma (Eds.), Adenine Press, Schenectady, NY, p177-196.



Koziolkiewicz, M., Gendaszewska, E., Maszewska, M., Stein, C. A. & Stec, W. J., (2001) The mononucleotide-dependent, nonantisense mechanism of action of phosphodiester and phosphorothioate oligonucleotides depends on the activity of an ect-5'-nucleotidase. *Blood* **98**, 995-1002.

Krowicki, K. & Lown, J. W. (1987) Synthesis of novel imidazole-containing DNA minor groove binding oligopeptides related to the antiviral antibiotic netropsin. *J. Org. Chem.* **52**, 3493-3501.

Kukreti, S., Sun, J. S., Garestier, T. & Hélène, C. (1997) Extension of the range of DNA sequences available for triple helix formation: stabilization of mismatched triplexes by acridine-containing oligonucleotides. *Nucleic Acids Res.* **25**, 4264-4270.

Kumar, R. & Lown, J. W (2003) Synthesis and antitumour cytotoxicity evaluation of novel pyrrole[2,1-c][1,4]benzodiazepine imidazole containing polyamide conjugates. *Oncol. Res.* **13**, 221-233.

Kumar, S., Yadagiri, B., Zimmermann, J., Pon, R. T. & Lown, J. W. (1990a) Sequence specific molecular recognition and binding by a GC recognising Hoechst 33258 analogue to the decadeoxyribonucleotide d-[CATGGCCATG]<sub>2</sub>: structural and dynamic aspects deduced from high field <sup>1</sup>H-NMR studies. *J. Biomol. Struct. Dyn.* **8**, 331-357.

Kumar, S., Jaseja, M., Zimmermann, J., Yadagiri, B., Pon, R. T., Sapse, A. M. & Lown, J. W. (1990b) Molecular recognition and binding of a GC site-avoiding thiazole-lexitropsin to the decadeoxynucleotide d-[CGCAATTGCG]<sub>2</sub>: <sup>1</sup>H-NMR evidence for thiazole intercalation. *J. Biomol. Struct. Dyn.* **8**, 99-121.

Kumar, S., Bathini, Y., Joseph, T., Pon, R. & Lown, J. W. (1991) Structural and dynamic aspects of non-intercalative (1:1) binding of a thiazole-lexitropsin to the decadeoxyribonucleotide d-[CGCAATTGCG]<sub>2</sub>: an <sup>1</sup>H-NMR and molecular modelling study. *J. Biomol. Struct. Dyn.* **9**, 1-21.

Kurrek, J. (2003) Antisense technologies: improvement through novel chemical modifications. *Eur. J. Biochem.* **270**, 1628-1644.

- Lacy, E. R., Le, N. M., Price, C. A., Lee, M. & Wilson, W. D. (2002a) Influence of a terminal formamido group on the sequence recognition of DNA by polyamides. *J. Am. Chem. Soc.* **124**, 2153-2163.
- Lacy, E. R., Cox, K. K., Wilson, W. D. & Lee, M. (2002b) Recognition of T\*G mismatched base pairs in DNA by stacked imidazole-containing polyamides: surface plasmon resonance and circular dichroism studies. *Nucleic Acids Res.* **30**, 1834-1841.
- Lacy, E. R., Nguyen, B., Le, M., Cox, K. K., O'Hare, C. C., Hartley, J. A., Lee, M. & Wilson WD. (2004) Energetic basis for selective recognition of T\*G mismatched base pairs in DNA by imidazole-rich polyamides. *Nucleic Acids Res.* **32**, 2000-2007.
- Lall, L. & Davidson, R. L. (1998) Sequence directed base mispairing in human oncogenes. *Mol. Cell. Biol.* **18**, 4659-4669
- Land, H., Parada, L. F. & Weinberg, R. A. (1983) Cellular oncogenes and multi-step carcinogenesis. *Science* **222**, 771-777.
- Lane, A. N., Jenkins, T. C., Brown, T. & Neidle, S. (1991) Interaction of berenil with the Eco RI dodecamer d(CGCGAATTCGCG)<sub>2</sub> in solution studied by NMR. *Biochemistry* **30**, 1372-1385.
- Lane, D. P., & Crawford, L. V., (1979) T antigen is bound to a host protein in SV40 transformed cells. *Nature* **278**, 261-63.
- Lane, M. J., Dabrowiak, J. C., & Vournakis, J. N. (1983) Sequence specificity of actinomycin d and netropsin binding to pBR322 DNA analysed by protection from DNase I. *Proc. Natl. Acad. Sci. USA* **80**, 3260-3264.
- Larsen, A., Escargueil, A. E. & Skladanowski, A. (2000) Resistance mechanisms associated with altered intracellular distribution of anticancer agents. *Pharm. Ther.* **85**, 217-229.
- Larsen, H. J., Bentin, T. & Nielsen, P. E. (1999) Antisense properties of peptide nucleic acid. *Biochim. Biophys. Acta* **1489**, 159-166.

- Laughton, C. A., Tanious, F., Nunn, C. M., Boykin, D. W., Wilson, W. D. & Neidle, S. (1996) A crystallographic and spectroscopic study of the complex between d(CGCGAATTCGCG)<sub>2</sub> and 2, 5-bis(4-guanyphenyl)furan, an analogue of berenil. Structural origins of enhanced DNA binding affinity. *Biochemistry* **35**, 5655-5661.
- Lavery, R., & Pullman, B. (1981) The molecular electrostatic potential, steric accessibility and hydration of Dickerson's B-DNA dodecamer d(CpGpCpGpApApTpTpCpGpCpGp). *Nucleic Acids Res.* **9**, 3765-3777.
- Lavery, R., Pullman, B. & Zakrzewska, K. (1982) Intrinsic electrostatic properties and base sequence effects in the structure of oligonucleotides. *Biophys. Chem.* **15**, 343-351.
- Lavery, R. & Pullman, B. (1985) The dependence of the surface electrostatic potential of B-DNA on environmental factors. *J. Biomol. Struct. Dyn.* **2**, 1021-1032.
- Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J-L., Thuong, N. T., Lhomme, J., Hélène, C. (1987) Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-[alpha]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.* **15**, 7749-7760.
- Lee, J. S., Woodsworth, M. L., Latimer, L. J. & Morgan, A. R. (1984) Poly (pyrimidine) poly (purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res.* **12**, 6603-6614.
- Lee, M. D., Dunne, T. S., Siegel, M. M., Chang, C. C., Morton, G. O. & Borders, D. B. (1987a) Calicheimicins, a novel family of antitumor antibiotics. 1. Chemistry and partial structure of calicheimicin .gamma.II *J. Am. Chem. Soc.* **109**, 3464-3466.
- Lee, M. D., Dunne, T. S., Chang, C. C., Ellestad, G. A., Siegel, M. M., Morton, G. O., McGahren, W. J. & Borders, D. B. (1987b) Calicheimicins, a novel family of antitumor antibiotics. 2. Chemistry and structure of calicheimicin .gamma.II. *J. Am. Chem. Soc.* **109**, 3466-3468.

- Lee, M., Hartley, J. A., Pon, R. T., Krowicki, K., & Lown, J. W. (1988a) Sequence specific molecular recognition by a monocationic lexitropsin of the decadeoxyribonucleotide d-[CATGGCCATG]<sub>2</sub>: structural and dynamic aspects deduced from high field <sup>1</sup>H-NMR studies. *Nucleic Acids Res.* **16**, 665-684.
- Lee, M., Krowick, K., Hartley, J. A., Pon, R. T. & Lown, J. W. (1988b) Molecular recognition between oligopeptides and nucleic acids. Influence of van der Waals contacts in determining the 3' terminus of DNA sequences read by monocationic lexitropsins. *J. Am. Chem. Soc.* **110**, 3641-3649.
- Lee, M., Rhodes, A. L., Wyatt, M. D., Forrow, S. & Hartley, J. A. (1993) Design, synthesis, and biological evaluation of DNA sequence and minor groove selective alkylating agents. *Anticancer Drug Des.* **8**, 173-192.
- Leimgruber, W., Batcho, A. D., & Schenker, F. (1965a) The structure of anthramycin. *J. Am. Chem. Soc.* **87**, 5793-5795.
- Leimgruber, W., Stefanovic, V., Schenker, F., Karr, A., & Berger, J. (1965b) Isolation and characterisation of anthramycin a new tumour antibiotic. anthramycin. *J. Am. Chem. Soc.* **87**, 5791-5793.
- Leupin, W., Chazin, W J., Hyberts, S., Denny, W. & Wüthrich, K. (1986) NMR studies of the complex between the decadeoxynucleotide d-(GCATTAATGC)<sub>2</sub> and a minor groove-binding drug. *Biochemistry* **25**, 5902-5910.
- Levine, A. J. (1997) p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323-331.
- Li, L. H., Swenson, D. H., Schpok, S. L., Kuentzel, S. L., Dayton, B. D., & Krueger, W. C. (1982) CC-1065 (NSC298223), a novel antitumour agent that interacts with DNA. *Cancer Res.* **42**, 999-1004.
- Li, L. H., Kelly, R. C., Warpehoski, M. A., McGovren, J. P., Gebhard, I., & Dekoning, T. F. (1991) Adozelezin, a selected lead among cyclopropyl pyrroleindole analogues of the DNA binding antibiotic, CC-1065. *Invest. New Drugs* **9**, 137-148.

Lieberman, J., Song, E., Lee, S-K. & Shankar, P. (2003) Interfering with disease: opportunities and roadblocks to harnessing RNA interference. *Trends In Mol. Med.* **9**, 397-403.

Linzer, D. I. & Levine, A. J., (1979) Characterisation of a 54 kdalton cellular SV40 tumour antigen present in SV40 transformed cells and uninfected embryonal carcinoma cells. *Cell* **17**, 43-52.

Lipardi, C., Wei, Q. & Paterson, B. M. (2001) RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* **107**, 297-307.

Loeb, L. A. & Preston, B. D. (1986) Mutagenesis by apurinic/apyrimidinic sites. *Ann. Rev. Genet.* **20**, 201-30.

Loke, S. L., Stein, C. A., Zhang, X. H., Mori, K., Nakanishi, M., Subasinghe, C., Cohen J. S. & Neckers, L. M. (1989) Characterisation of oligonucleotide transport into living cells. *Proc. Natl. Acad. Sci. USA* **86**, 3474-3478.

Lombardy, R. L., Tanious, F. A., Ramachandran, K., Tidwell, R. R. & Wilson WD. (1996) Synthesis and DNA interactions of benzimidazole dications which have activity against opportunistic infections. *J. Med. Chem.* **39**, 1452-1462.

Long, M., Jones, J. P., Sullenger, B. A., Byun, J. (2003) Ribozyme-mediated revision of RNA and DNA. *J. Clinical Invest.* **112**, 312-318.

Longo, R., Sarmiento, R., Fanelli, M., Capacetti, B., Gattuso, D. & Gaspari, G. (2003) Anti-angiogenic therapy: rationale, challenges and clinical studies. *Angiogenesis* **5**, 237-256.

Lown, J. W., Krowicki, K., Balzarini, J. & De Clercq, E., (1986a) Structure-activity relationship of novel oligopeptides antiviral and antitumour agents related to netropsin and distamycin. *J. Med. Chem.* **29**, 1210-1214.

Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B. & Dabrowiak, J. C. (1986b) Molecular recognition between oligopeptides and nucleic acids: novel

imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity. *Biochemistry* **25**, 7408-7416.

Lown, J. W., (1988) Lexitropsins: rational design of DNA sequence reading agents as novel anticancer agents and potential cellular probes. *Anticancer Drug Des.* **3**, 25-40.

Lown, J. W., Krowicki, K., Balzarini, J., Newman, R. A., & De Clercq, E. (1989) Novel linked antiviral and antitumour agents related to netropsin and distamycin: synthesis and biological evaluation. *J. Med. Chem.* **32**, 2368-2375.

Lown, J. W. (1990) Molecular mechanisms of DNA sequence recognition by groove binding ligands: biochemical and biological consequences. In *Molecular Basis Of Specificity In Nucleic Acid-Drug Interactions*. B. Pullman And J. Jortner, (Eds.) Dordrecht, Kluwer Academic Publishers, p106-122.

Lown, J. W. (1993) Design of sequence-specific agents: lexitropsins. In *Molecular Aspects Of Anticancer Drug-DNA Interactions*. Vol. 1, Macmillan, London, p322-355.

Lown, J. W. (1997) DNA sequence recognition altered bis-benzimidazole minor groove binders. In *Advances In DNA Sequence Specific Agents*, Vol. 3, B. J. Graham (Eds.), Greenwich: JAI Press, p67-95.

Luck, G., Reinert, K. E., Baguley, B. & Zimmer, C. (1987) Interaction of the nonintercalative antitumour drugs SN-6999 and SN-18071 with DNA: influence of ligand structure on the binding specificity. *J. Biomol. Struct. Dyn.* **4**, 1079-1094.

Lugo, T. G., Pendergast, A. M., Muller, A. J. & Witte, O. N. (1990) Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science* **247**, 1079-1082.

Mack, D., Chiu, T. K. & Dickerson, R. E. (2001) Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts. *J. Mol. Biol.* **312**, 1037-1049.

- Maeshima, K., Janssen, S. & Laemmli, U. K. (2001) Specific targeting of insect and vertebrate telomeres with pyrrole and imidazole polyamides. *EMBO J.* **20**, 3218-3228.
- Malkov, V. A., Voloshin, O. N., Soyfer, V. N. & Frank-Kamenetskii, M. D. (1993) Cation and sequence effects on stability of intermolecular pyrimidine-purine-purine triplex. *Nucleic Acids Res.* **21**, 585-591.
- Mann, J., Baron, A., Opoku-Boahen, Y., Johansson, E., Parkinson, G., Kelland, L. R. & Neidle, S. (2001) A new class of symmetric bisbenzimidazole-based DNA minor groove-binding agents showing antitumor activity. *J. Med. Chem.* **44**, 138-144.
- Mapp, A. K., Ansari, A. Z., Ptashne, M. & Dervan, P. B. (2000) Activation of gene expression by small molecule transcription factors. *Proc. Natl. Acad. Sci. USA* **97**, 3930-3935.
- Marky, L. A. & Breslauer, K. J., (1987) Origins of netropsin binding affinity and specificity: correlations of thermodynamic and structural data. *Proc. Natl. Acad. Sci. USA* **18**, 4359-4363.
- Martin, D. G., Biles, C., Gerpheide, S. A., Hanka, L. J., Krueger, W. C., McGovren, J. P., Mizesak, S. A., Neil, G. L., Stewart, J. C. & Visser, J. (1981). CC-1065 (NSC 298223), a potent new antitumour agent improved production and isolation, characterisation and antitumour activity. *J. Antibiot.* **34**, 1119-1125.
- McBryant, S. J., Baird, E. E., Trauger, J. W., Dervan, P. B. & Gottesfeld, J. M. (1999) Minor groove dna-protein contacts upstream of a tRNA gene detected with a synthetic dna binding ligand. *J. Mol. Biol.* **286**, 973-981.
- McGovren, J. P., Clarke, G. L., Pratti, E. A. & Dekoning, T. F. (1984) Preliminary toxicity studies with the DNA-binding antibiotic, CC-1065. *J. Antibiot.* **37**, 63-70.
- McManus, M. T. & Sharp, P. A. (2002) Gene silencing in mammals by small interfering RNAs. *Nature Reviews Genetics* **3**, 737-747.

- Midgley, R. S. & Kerr, D. J. (2002) Ras as a target in cancer therapy. *Crit. Rev. Oncol. Hematology* **44**, 109-120.
- Miller, P. S. & Hamma, T. (1999) Studies on anti-HIV oligonucleotides that contain alternating methylphosphonate/phosphodiester linkages. *Antisense Nucleic Acids Drug Dev.* **9**, 367-370.
- Milner, N., Mir, K. U. & Southern, E. M. (1997) Selecting the effective antisense reagents on combinatorial oligonucleotide arrays. *Nature Biotechnol.* **15**, 537-541.
- Mitchell, M. A., Kelly, R. C., Winnieski, N. A., Hatzenbuehler, N. T. & Williams, M. G. (1991) Synthesis and DNA cross-linking by a rigid CPI dimer. *J. Am. Chem. Soc.* **113**, 8994-8995.
- Miyamoto, M., Kondo, S., Naganawa, H., Maeda, K., Ohno, M. & Umezawa, H. (1977) Structure and synthesis of neothramycin. *J. Antibiot.* **30**, 340-343.
- Monia, B. P., Lesnik, E. A., Gonzalez, C. Lima, W. F., Mcgee, D., Guinosso, C. J., Kawasaki, A. M., Cook, P. D. & Freier, S. M. (1993) Evaluation of 2'-modified oligonucleotides containing 2'-deoxy gaps as antisense inhibitors of gene expression. *J. Biol. Chem.* **268**, 14514-14522.
- Moser, H. E. & Dervan, P. B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **238**, 645-650.
- Mountzouris, J. A., Wang, J.-J., Thurston, D. E. & Hurley, L. H., (1994) Comparison of a DSB-120 DNA interstrand cross-linked adduct with the corresponding bis-tomaymycin adduct: an example of a successful template-directed approach to drug design based upon the monoalkylating compound tomaymycin. *J. Med. Chem.* **37**, 3132-3140.
- Mrksich, M., Wade, W. S., Dwyer, T. W., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992) Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. USA* **89**, 7586-7590.



- Mrksich, M. & Dervan, P. B. (1993a) Antiparallel side-by-side heterodimer for sequence-specific recognition in the minor groove of DNA by a distamycin /1-methylimidazole-2-carboxamide-netropsin pair. *J. Am. Chem. Soc.* **115**, 2572-2576.
- Mrksich, M. & Dervan P. B. (1993b) Enhanced sequence specific recognition in the minor groove of DNA by covalent peptide dimers: bis(pyridine- 2-carboxamidonetropsin) (CH<sub>2</sub>)<sub>3,6</sub>. *J. Am. Chem. Soc.* **115**, 9892-9899.
- Mrksich, M. & Dervan, P. B. (1994a) Design of a covalent peptide heterodimer for sequence-specific recognition in the minor groove of double-helical DNA. *J. Am. Chem. Soc.* **116**, 3663-3664.
- Mrksich, M., Parks, M. E. & Dervan, P. B. (1994b) Hairpin peptide motif. A new class of oligopeptides for sequence specific recognition in the minor groove of double-helical DNA. *J. Am. Chem. Soc.* **116**, 7983-7988.
- Mrksich, M. & Dervan, P. B. (1995) Recognition in the minor groove of DNA at 5'-(A,T)GCGC(A,T)-3' by a four ring tripeptide dimer. Reversal of the specificity of the natural product distamycin. *J. Am. Chem. Soc.* **117**, 3325.
- Nakatani, K., Sando, S. & Saito, I. (2001a) Improved selectivity for the binding of naphthyridine dimer to a guanine-guanine mismatch. *Bioorg. Med. Chem.* **9**, 2381-2385.
- Nakatani, K., Sando, S., Kumasawa, H., Kikuchi, J. & Saito, I. (2001b) Recognition of guanine-guanine mismatches by the dimeric form of 2-amino-1, 8-naphthyridine. *J. Am. Chem. Soc.* **123**, 12650-12657.
- Nakatani, K., Sando, S. & Saito, I. (2001c) Scanning of guanine-guanine mismatches in DNA by synthetic ligands using surface plasmon resonance. *Nat. Biotechnol.* **19**, 51 –55.
- Neamati, N., Mazumder, A., Sunder, S., Owen, J. M., Tandon, M., Lown, J. W. & Pommier, Y. (1998) Highly potent synthetic polyamides, bisdistamycins, and lexitropsins as inhibitors of human immunodeficiency virus type 1 integrase. *Mol Pharmacol.* **54**, 280-290.

- Neely, L., Trauger, J. W., Baird, E. E., Dervan, P. B. & Gottesfeld, J. M. (1997) Importance of minor groove binding zinc fingers within the transcription factor IIIA-DNA complex. *J. Mol. Biol.* **274**, 439-445.
- Neidle, S. (1992) Minor groove width and accessibility in B-DNA drug and protein complexes. *FEBS Lett.* **298**, 97-99.
- Neidle, S., Puvvada, M. S., & Thurston, D. E. (1994) The relevance of drug DNA sequence specificity to anti-tumour activity. *Eur. J. Cancer* **30A**, 567-568.
- Neidle, S. (2001) DNA minor groove recognition by small molecules. *Nat. Prod. Rep.* **18**, 291-309.
- Nelson, H. C. M., Finch, J. T., Luisis, B. F. & Klug, A. (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**, 221-226.
- Nguyen, D. H., Szewczyk, J. W., Baird, E. E., & Dervan, P. B. (2001) Alternative heterocycles for DNA recognition: an n-methylpyrazole/n-methylpyrrole pair specifies for A.T/T.A base pairs. *Bioorg. Med. Chem.* **9**, 7-17.
- Nielsen, P. E., Egholm, M., Berg, R. H. & Burchardt, O. (1991) Sequence selective recognition of DNA by strand displacement with a thymine substituted polyamide. *Science* **254**, 1497-1500.
- Nielsen, P. E. (1999) Peptide nucleic acids as therapeutic agents. *Curr. Opin. Struct. Biol.* **9**, 353-357.
- Noonberg, S. B., Francois, J. C., Praseuth, D., Guieysse-Peugeot, A. L., Lacoste, J., Garestier, T. & Hélène, C. (1995) Triplex formation with alpha anomers of purine-rich and pyrimidine-rich oligodeoxynucleotides. *Nucleic Acids Res.* **23**, 4042-4049.
- O'Hare, C. C., Mack, D., Tandon, M., Sharma, S. K., Lown, J. W., Kopka, M. L., Dickerson, R. E. & Hartley, J. H. (2002) DNA sequence recognition in the minor groove by cross-linked polyamides: the effect of N-terminal head group and linker length on binding affinity and specificity. *Proc. Natl. Acad. Sci. USA* **99**, 72-77.

- Oyoshi, T., Kawakami, W., Bando, T., Narita, A. & Sugiyama, H. (2002) Regulation of gene expression by a sequence-specific alkylating polyamide. *Nucleic Acids Res. Suppl.* **2**, 259-260.
- Oyoshi, T., Kawakami, W., Bando, T., Narita, A. & Sugiyama, H. (2003) Inhibition of transcription at a coding sequence by an alkylating polyamide. *J. Am. Chem. Soc.* **125**, 4752-4754.
- Pandolfi, P. P. (2001) Transcription Therapy For Cancer. *Oncogene* **20**, 3116-3127.
- Parkinson, J. A. Barber, J., Douglas, K. T., Rosamond, J. & Sharples, D. (1990) minor groove recognition of the self-complementary duplex d(CGCGAATTCGCG)<sub>2</sub> by Hoechst 33258: a high-field NMR study. *Biochemistry* **29**, 10181-10190.
- Parks, M. E., Baird, E. E. & Dervan, P. B. (1996a) Optimisation of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **118**, 6147-6152.
- Parks, M. E., Baird, E. E. & Dervan, P. B. (1996b) Recognition of 5'(A,T)GG(A,T)<sub>2</sub>-3' sequences in the minor groove of DNA by hairpin polyamides. *J. Am. Chem. Soc.* **118**, 6153-6159.
- Parolin, C., Montecucco, A., Chiarocchi, G., Pedrali-Noy, G., Valisena, S., Palumbo, M., & Palu, G. (1990) The effect of the minor groove-binding agent DAPI (4,6-diamidino-2-phenyl-indole) on DNA-directed enzymes: an attempt to explain inhibition of plasmid expression in *Escherichia coli*. *FEMS Microbiol. Lett.* **68**, 341-346.
- Patel, D. J. (1982) Antibiotic-DNA interactions: intermolecular nuclear overhauser effects in the netropsin-d(C-G-C-G-A-A-T-T-C-G-C-G) complex in solution. *Proc. Natl. Acad. Sci. USA* **79**, 6424-6428.
- Paterson, B. M., Roberts, B. E. & Kuff, E. L. (1977) Structural gene identification and mapping by DNA-mRNA hybrid arrested cell-free translation. *Proc. Natl. Acad. Sci. USA* **74**, 4370-4374.

- Pelton, J. G. & Wemmer, D. E. (1988) structural modelling of the distamycin A-d(CGCGAATTCGCG)<sub>2</sub> complex using 2D NMR and molecular mechanics. *Biochemistry* **27**, 8088-8096.
- Pelton, J. G. & Wemmer, D. E. (1989) Structural characterisation of a 2:1 distamycin A.d(CGCAAATTGGCC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723-5727.
- Pelton, J. G. & Wemmer, D. E. (1990) Binding modes of distamycin A with d(CGCAAATTTGCG)<sub>2</sub> determined by two-dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393-1399.
- Phylactou, L. (2000) Ribozyme and peptide-nucleic acid based gene therapy. *Adv. Drug Deliv. Rev.* **44**, 97-108.
- Pilch, D. S., Levenson, C. & Shafer, R. H. (1991) Structure, stability, and thermodynamics of a short intermolecular purine-purine-pyrimidine triple helix. *Biochemistry* **30**, 6081-6088.
- Pilch, D. S., Kirolos, M. A., Liu, X., Plum, G. E. & Breslauer, K. (1995) Berenil [1,3-bis(4'-amidinophenyl)triazene] binding to DNA duplexes and to a RNA duplex: evidence for both intercalative and minor groove binding properties. *Biochemistry* **34**, 9962-9976.
- Pitot, H. C., Reid, J. M., Sloan, J. A., Ames, M. M., Adjei, A. A., Rubin, J., Bagniewski, P. G., Atherton, P., Rayson, D., Goldberg, R. M., Erlichman, C. (2002) Phase I study of bizelesin (NSC 615291) in patients with advanced solid tumours. *Clin Cancer Res.* **8**, 712-717.
- Pooga, M., Land, T., Bartfai, T. & Langel, U. (2001) PNA oligomers as tools for specific modulation of gene expression. *Biomol. Eng.* **17**, 183-192.
- Portugal, J. (1994) Berenil acts as a poison of eukaryotic topoisomerase II. *FEBS Lett.* **344**, 136-138.
- Portugal, J. & Waring, M. J. (1986) Antibiotics which can alter the rotational orientation of nucleosome core DNA. *Nucleic Acids Res.* **14**, 8735-8754.

- Portugal, J. & Waring, M. J. (1987) Comparison of binding sites in DNA for berenil, netropsin and distamycin. A footprinting study. *Eur. J. Biochem.* **167**, 281-289.
- Portugal, J. & Waring, M. J., (1988) Assignment of binding sites in DNA for 4', 6-diamidine-2-phenylindole and bisbenzimidazole (Hoechst 33258): a comparative footprinting study. *Biochim. Biophys. Acta* **949**, 158-168.
- Postel, E. H., Flint, S. J., Kessler, D. J. & Hogan, M. E. (1991) Evidence that a triplex-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc mRNA levels. *Proc Natl Acad Sci U S A* **88**, 8227-8231.
- Potaman, V. N. & Soyfer, V. N. (1994) Divalent metal cations upon coordination to the N7 of purines differentially stabilize the PyPuPu DNA triplex due to unequal Hoogsteen-type hydrogen bond enhancement. *J. Biomol. Struct. Dyn.* **11**, 1035-1040.
- Poulin-Kerstien, A. T. & Dervan P. B. (2003) DNA-templated dimerisation of hairpin polyamides. *J. Am. Chem. Soc.* **125**, 15811-15821.
- Povrik, L. F. (1983) Bleomycin. In *Molecular Aspects Of Anticancer Drug Action*. S. Neidle & M. Waring (Eds), Macmillan Press London, p157-181.
- Praseuth, D., Guieyette, A. L. & Hélène, C. (1999) Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta* **1489**, 181-206.
- Pratt, W. B., Rudon, R. W., Ensminger, W. D. & Maybaum, J. (1994) In *The Anticancer Drugs*. (Eds) Oxford, Oxford University Press.
- Pritchard, R. S., Hill, A. K. D., Dijkstra, B., McDermott, E. W. & O'Higgins, N. J. (2003) The prevention of breast cancer. *British Journal Of Surgery* **90**, 772-283.
- Probst, G. W., Hoehn, M. M. & Woods, B. L. (1965) Anthelvincins, new antibiotics with anthelmintic properties. *Antimicrob. Agents Chemother.* **5**, 789-795.

- Puerta-Fernandez, E., Romero-Lopez, C., Barroso-delJesus, A., Berzal-Herranz, A. (2003) Ribozymes: recent advances in the development of RNA tools. *FEMS Microbiology Rev.* **27**, 75-97.
- Puvvada, M. S., Stephenson, P., Gibson, I. & Thurston, D. E., (1992) Inhibition of *in vitro* transcription of the n-ras oncogen by the pyrrolobenzodiazepines. *J. Pharm. Pharmacol.* **44**, 1055.
- Puvvada, M. S., Hartley, J. A., Jenkins, T. C. & Thurston, D. E. (1993) A quantitative assay to measure the relative DNA-binding affinity of pyrrolo[2,1-c][1,4]benzodiazepine (PBD) antitumour antibiotics based on the inhibition of restriction endonuclease BamHI. *Nucleic Acids Res.* **21**, 3671-3675.
- Puvvada, M. S., Forrow, S. A., Hartley, J. A., Stephenson, P., Gibson, I., Jenkins, T. C. & Thurston, D. E. (1997) Inhibition of bacteriophage T7 RNA polymerase *in vitro* transcription by DNA-binding pyrrolo[2,1-c][1,4]benzodiazepines. *Biochemistry* **36**, 2478-84.
- Quintana, J. R., Lipanov, A. A. & Dickerson RE. (1991) Low-temperature crystallographic analyses of the binding of Hoechst 33258 to the double-helical DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *Biochemistry* **30**, 10294-10306.
- Rao, K. E., Bathini, Y. & Lown, J. W. (1990a) Synthesis of novel thiazole containing DNA minor groove binding oligopeptides related to the antibiotic distamycin. *J. Org. Chem.* **55**, 728-737.
- Rao, K. E., Shea, R. G., Yadagiri, B., Lown, J. W. (1990b) Molecular recognition between oligopeptides and nucleic acids: DNA sequence specificity and binding properties of thiazole-lexitropsins incorporating the concepts of base site acceptance and avoidance. *Anticancer Drug Design* **5**, 3-20.
- Rao, K. E., Shea, R. G., Yadagiri, B. & Lown, J. W (1990). Influence of the methyl substituents of a thiazole-containing lexitropsins on the mode of binding to DNA. *Anti-Cancer Drug Design*, **5**, 3-20.

Rao, K.E. & Lown, J. W. (1991) Molecular recognition between ligands and nucleic acids: DNA binding characteristics of analogues of Hoechst 33258 designed to exhibit altered base and sequence recognition. *Chem. Res. Toxicol.* **4**, 661-669.

Rao, K. E., Gosselin, G., Mrani, D., Périgaud, C., Imbach, J. L., Bailly, C., Hénichart, J. P., Colson, P., Houssier, C. & Lown, J. W. (1994) Psoralen-lexitropsin hybrids: DNA sequence selectivity of photoinduced cross-linking from MPE footprinting and exonuclease ii stop assay, and mode of binding from electric linear dichroism. *Anticancer Drug Des.* **9**, 221-237.

Ray, A. & Nordén, B. (2000) Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *FASEB J.* **14**, 1041-1060.

Reddy, B. S., Damayanthi, Y. & Lown, J. W. (2000) Design, synthesis and *in vitro* cytotoxicity studies of novel pyrrolo[2,1-c][1,4]benzodiazepine (PBD)-polyamide conjugates and 2, 2'-PBD dimers. *Anticancer Drug Des.* **15**, 225-238.

Reddy, B. S., Sondhi, S. M. & Lown, J. W. (1999) Synthetic DNA minor groove-binding drugs. *Pharmacology & Therapeutics* **84**, 1-111.

Reinert, K. (1999) DNA Multimode interaction of berenil and pentamidine; double helix stiffening, unbending and bending. *J. Biomol. Struct. Dyn.* **17**, 311-331.

Renneberg, D. & Dervan, P. B. (2003) Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.* **125**, 5707-5716.

Reynolds, V. L., Molineux, I. J., Kaplan, D. J., Swenson, D. H., & Hurley, L. H. (1985) Reaction of the antitumour antibiotic CC-1065 with DNA location of the site of thermally induced strand breakage analysis of DNA sequence specificity. *Biochemistry* **24**, 6228-6237.

Robertson, I. G. C. & Baguley, B. C. (1982) A comparison of the requirements for antitumour activity and antibacteriophage lambda activity for a series of non-intercalative DNA-binding agents. *Eur. J. Cancer Clin. Oncol.* **18**, 271-279.

- Rubinstein, L., Shoemaker, R., Paull, K., Simon, R., Tosini, S. & Skehen, P., (1990) Comparison of *in vitro* anticancer-drug-screening data generated with a tetrazolium assay versus a protein assay against a diverse panel of human tumour cell lines. *J. Natl. Cancer Inst.* **82**, 1113-1117.
- Rucker, V. C., Foister, S., Melander, C. & Dervan, P. B. (2003) Sequence specific fluorescence detection of double stranded DNA. *J. Am. Chem. Soc.* **125**, 1195-1202.
- Ryabinin, V. A., Sinyakov, A. N., Richard De Soultrait, V., Caumont, A., Parissi, V., Zakharova, O. D., Vasyutina, E. L., Yurchenko, E., Bayandin, R., Litvak, S., Tarrago-Litvak, L. & Nevinsky, G. A. (2000) Inhibition of HIV-1 integrase-catalysed reaction by new DNA minor groove ligands: the oligo-1,3-thiazolecarboxamide derivatives. *Eur. J. Med. Chem.* **35**, 989-1000.
- Rydzewski, J. M., Leupin, W. & Chazin, W. (1996) The width of the minor groove affects the binding of the bisquaternary heterocycle SN-6999 to duplex DNA. *Nucleic Acids Res.* **24**, 1287-1293.
- Sastry, M. & Patel, D.J. (1993) Solution structure of the mithramycin dimer-DNA complex. *Biochemistry* **32**, 6588-6604.
- Sansom, C. E., Laughton, C. A., Neidle, S., Schwalbe, C. H. & Stevens, M. F. (1990) Structural studies on bio-active compounds. Part xiv. Molecular modelling of the interactions between pentamidine and DNA. *Anticancer Drug Des.* **5**, 243-248.
- Satz, A. L. & Bruice, T. C. (2002) Recognition in the minor groove of double-stranded DNA by microgonotropens. *Acc. Chem. Res.* **35**, 86-95.
- Sauers, R. R. (1995) An analysis of van der Waals attractive forces in DNA-minor groove binding. *Bioorg. Med. Chem. Lett.* **5**, 2573-2576.
- Schaal, T. D., Mallet, W. G., McMinn, D. L., Nguyen, N. V., Sopko, M. M., John, S. & Parekh, B. S. (2003) Inhibition of human papilloma virus E2 DNA binding protein by covalently linked polyamides. *Nucleic Acids Res.* **31**, 1282-1291.



Scherr, M. & Rossi, J. J. (1998) Rapid determination and quantitation of the accessibility to native RNAs by antisense oligodeoxynucleotides in murine cell extracts. *Nucleic Acids Res.* **26**, 5079-5085.

Schmitz, H. & Hubner, W. (1993) A thermodynamic and spectroscopic study on the binding of berenil to poly d(AT) and to poly(dA) x poly(dT). *Biophys. Chem.* **48**, 61-74.

Schneider, U., Schwenk, H. U. & Bornkamm, G. (1977) Characterisation of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukaemic transformed non-hodgkin lymphoma. *Int. J. Cancer* **19**, 621-626.

Schultz, P. G., Taylor, J. S. & Dervan, P. B. (1982) Design and synthesis of a sequence-specific DNA cleaving molecule (distamycin-EDTA) iron(II). *J. Am. Chem. Soc.* **104**, 6861.

Schultz, P. G. & Dervan, P. B. (1983) Sequence-specific double strand cleavage of DNA by bis(EDTA-distamycin-FeII) and EDTA-bis(distamycin)FeII. *J. Am. Chem. Soc.* **105**, 7748.

Scott, E. V., Zon, G., Marzilli, L. G., & Wilson, W. D. 2D NMR investigation of the binding of the anticancer drug actinomycin d to duplexed dATGCGCAT: conformational features of the unique 2:1 adduct. (1988) *Biochemistry* **27**, 7940-7951.

Seidman & Glazer, P. (2003) The potential for gene repair via triple helix formation. *J. Clin. Invest.* **112**, 487-494.

Sekiguchi, J. A. & Kmiec, E. B. (1988) Studies on DNA topoisomerase activity during *in vitro* chromatin assembly. *Mol. Cell. Biochem.* **83**, 195-205.

Senear, D. F., Brenowitz, M., Shea, M. A. & Ackers, G. K., (1986) Energetics of cooperative protein-DNA interactions: comparison between quantitative deoxyribonuclease footprint titration and filter binding. *Biochemistry* **25**, 7344-7354.

Shaffer, R. & Traktman, P. (1987) Vaccinia virus encapsidates a novel DNA topoisomerase with the properties of a eukaryotic type I enzyme. *J. Biol. Chem.* **19**, 9309-9315.

Sharma, S. K., Tandon, M. & Lown, J. W. (2000) Design and synthesis of novel thiazole-containing cross-linked polyamides related to the antiviral antibiotic distamycin. *J. Org. Chem.* **65**, 1102-1107.

Sharma, S. K., Morrissey, A. T., Miller, G. G., Gmeiner, W. H. & Lown, J. W. (2001) Design, synthesis, and intracellular localisation of a fluorescently labelled DNA binding polyamide related to the antibiotic distamycin. *Bioorg. Med. Chem. Lett.* **11**, 769-772.

Sharma, S. K., Billaud, J-N., Tandon, M., Billet, O., Choi, S., Kopka, M. L., Phillips, T. R. & Lown, J. W. (2002) inhibition of feline immunodeficiency virus (FIV) replication by DNA binding polyamides. *Bioorg. Med. Chem. Lett.* **12**, 2007-2010.

Sharp, P. A. (2001) RNA interference. *Genes Dev.* **15**, 485-490.

Shoji, Y., Akhtar, S., Periasamy, A., Herman, B., & Juliano, R. L., (1991) Mechanism of cellular uptake of modified oligodeoxynucleotides containing methylphosphonate linkages. *Nucleic Acids Res.* **19**, 5543-5550.

Sievers, E. L., (2001) Efficacy and safety of gemtuzumab ozogamicin in patients with CD33-positive acute myeloid leukaemia in first relapse. *Expert Opin Biol Ther.* **1**, 893-901.

Sledge, G. W. & Miller, K. D. (2003) Exploiting the hallmarks of cancer: the future conquest of breast cancer. *Eur. J. Cancer* **39**, 1668-1675.

Smellie, M., Bose, D., Thompson, A. S., Jenkins, T. C., Hartley, J. A. & Thurston, D. E. (2003) Sequence-selective recognition of duplex DNA through covalent interstrand cross-linking: kinetic and molecular modelling studies with pyrrolobenzodiazepine dimers. *Biochemistry* **42**, 8232-8239.

- Smith, A. L. & Nicolaou, K. C. (1996) The enediyne antibiotics. *J. Med. Chem.* **39**, 2103-2117.
- Soule, H. D., Vasquez, J., Long, A., Albert, S. & Brennan, M. (1973) A human breast cancer cell line from a pleural effusion derived from a breast carcinoma. *JNCI* **51**, 1409-1416.
- Southern, E. M., Casegreen, S. C., Elder, J. K., Johnson, M. Mir, K. U., Wang, L. & Williams J. C. (1994) Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. *Nucleic Acids Res.* **22**, 1368-1373.
- Spink, N., Brown, D. G., Skelly, J. V. & Neidle, S. (1994) Sequence-dependent effects in drug-dna interaction: the crystal structure of Hoechst 33258 bound in the d(CGCAAATTTGCG)<sub>2</sub> duplex. *Nucleic Acids Res.* **22**, 1607-1612.
- Stein, C. A. (2000) Is irrelevant cleavage the price of antisense efficacy? *Pharmacology & Therapeutics* **85**, 231-236.
- Stein, C. A., Tonkinson, J. L., Zhang, L. M., Yakubov, L., Gervasoni, J., Taub, R. & Rotenberg, S. A. (1993) Dynamics of the internalisation of phosphodiester oligodeoxynucleotides in HL60 cells. *Biochemistry* **32**, 4855-4861.
- Stein, C. A. (1999) Two problems in antisense biotechnology: *in vitro* delivery and the design of antisense experiments. *Biochim. Biophys. Acta* **1489**, 45-52.
- Sullenger, B. A. & Gilboa, E. (2002) Emerging clinical applications of RNA. *Nature* **418**, 252-8.
- Summerton, J., & Weller, D. (1997a) Morpholino antisense oligomers: design, preparation and properties. *Antisense Nucleic Acid Drug Dev.* **7**, 187-195.
- Summerton, J., Stein, D., Huang, S. B., Matthews, P., Weller, D. & Partridge, M. (1997b) Morpholino and phosphorothioate antisense oligomers compared in cell-free and in-cell systems. *Antisense Nucleic Acid Drug Dev.* **7**, 63-70.
- Summerton, J. (1999) Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim. Biophys. Acta* **1489**, 141-158.

- Sun, D. & Hurley, L. H. (1993) Analysis of the monoalkylation and cross-linking sequence specificity of bizelesin, a bifunctional alkylating agent related to (+)-CC-1065. *J. Am. Chem. Soc.* **115**, 5925-5933.
- Sun, Q., Gatto, B., Yu, C., Liu, A., Liu, L. F. & Edmond J. Lavoie (1994) Structure activity of topoisomerase I poisons related to Hoechst 33342. *Bioorg. Med. Chem. Lett.* **4**, 2871-2876.
- Supekova, L., Pezacki, J. P., Su, A. I., Loweth, C. J., Riedl, R., Geierstanger, B., Schultz, P. G. & Wemmer, D. E. (2002) Genomic effects of polyamide/DNA interactions on mRNA expression. *Chem. Biol.* **9**, 821-827.
- Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B. & Luger, K. (2003) Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **326**, 371-380.
- Swalley, S. E., Baird, E. E. & Dervan, P. B. (1999) Effects of  $\gamma$ -turn and  $\beta$ -tail amino acids on sequence specific recognition of DNA by hairpin polyamides. *J. Am. Chem. Soc.* **121**, 1113-1120.
- Takahashi, I., Takahashi, K., Ichimura, M., Morimoto, M., Asano, K., Kawamoto, I., Tomita, F., & Nakano, H. (1988) Duocarmycin A, a new antitumour antibiotic from *Streptomyces*. *J. Antibiot.* **41**, 1915-1917.
- Takaishi, T., Sugawara, Y. & Suzuki, M. (1972) Structure of kikumycin A and B. *Tetrahedron Lett.* **13**, 1873-1876.
- Tanious, F. A., Veal, J. M., Buczak, L., Ratmeyer, L. S., Wilson, W. D., (1992) DAPI (4',6-diamidino-2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and intercalation at AU sites. *Biochemistry* **31**, 3103-3112.
- Taylor, M. F., Weller, D. D., & Kobzik, L., (1998) Effect of TNF- $\alpha$  antisense oligomers on cytokine production by primary murine alveolar macrophages. *Antisense Nucleic Acid Drug Dev.* **8**, 199-205.

- Taylor, J. S., Schultz, P. G. & Dervan, P. B. (1984) DNA affinity cleaving. sequence specific cleavage of DNA by distamycin-EDTA.Fe(II) and EDTA-distamycin.Fe(II). *Tetrahedron* **40**, 457.
- Teng, M., Usman, C. A., Frederick, C. A. & Wang, A. H. J. (1988) The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG). *Nucleic Acids Res.* **16**, 2671-2790.
- Teng, S. P., Woodson, S. A. & Crothers, D. M. (1989) DNA sequence specificity of mitomycin cross-linking. *Biochemistry* **28**, 3901-3907.
- Thuong, N. T. & Hélène, C. (1993) Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angewandte Chem.* **32**, 666-690.
- Thurston, D. E., Jones, G. B. & Davis, M. E., (1990) Synthesis and reactivity of a novel oxazolo[2,3-*c*][1,4]benzodiazepine ring system with DNA recognition potential: a new class of anthramycins. *J. Chem. Soc. Chem. Commun.* **12**, 874-876.
- Thurston, D. E., (1993) Advances in the study of pyrrolo(2,1-*c*)(1,4)benzodiazepine (PBD) antitumour antibiotics. In *Molecular Aspects Of Anticancer Drug-DNA Interactions*. S. Neidle & M. Waring (Eds) London Macmillan, p54-88.
- Thurston, D. E., Bose, D. S., Thompson, A. S., Howard, P. W., Leoni, A., Croker, S. J., Jenkins, T. C., Neidle, Hartley, J. A. & Hurley, L. H., (1996) synthesis of sequence-selective C8-linked pyrrolo[2,1-*c*][1,4]benzodiazepine DNA interstrand cross-linking agents. *J. Org. Chem.* **61**, 8141-8147.
- Thurston, D. E., Bose, D. S., Howard P. W., Jenkins, T. C., Leoni, A., Baraldi, P. G., Guiotto, A., Cacciari, B., Kelland, L. R., Folpe, M. P. & Rault, S. (1999) Effect of A-ring modifications on the DNA binding behaviour and cytotoxicity of pyrrolo(2,1-*c*)(1,4)benzodiazepines. *J. Med. Chem.* **42**, 1951-1964.
- Tidd, D. M. (1993) Antisense oligonucleotides as potential inhibitors of gene expression. In *Molecular Aspects Of Anticancer Drug-DNA Interactions*. Volume I, S. Neidle & M. Waring (Eds) Macmillan, London, p274-299.

- Toft, N. J. & Arends, M. J. (1998) DNA mismatch repair and colorectal cancer. *J. Pathol.* **185**, 123-129.
- Tonini, T., Rossi, F. & Claudio, P. P. (2003) Molecular basis of angiogenesis and cancer. *Oncogene* **22**, 6549-6556.
- Tomasz, M., Lipman, R., Chowdary, D., Pawlak, J., Verdine, G. L. & Nakanishi, K., (1987) Isolation and structure of a covalent cross-link adduct between mitomycin C and DNA. *Science* **235**, 1204-1208.
- Tomasz, M., Chawla, A. K. & Lipman, R. (1988) Mechanism of monofunctional and bifunctional alkylation of DNA by mitomycin C. *Biochemistry* **27**, 3182-3187.
- Tonkinson, J. L. & Stein, C. A. (1994) Patterns of intracellular compartmentalisation, trafficking and acidification of 5'-fluorescein labelled phosphodiester and phosphorothioate oligodeoxynucleotides in HL60 cells. *Nucleic Acids Res.* **22**, 4268-4275.
- Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996a) Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559-561.
- Trauger, J. W., Baird, E. E., Mrksich, M. & Dervan, P. B. (1996b) Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9-13 base pairs. *J. Am. Chem. Soc.* **118**, 6160-6166.
- Trauger, J. W., Baird, E. E. & Dervan, P. B. (1998) Cooperative hairpin dimers for recognition of DNA by pyrrole-imidazole polyamides. *Angew. Chem. Int. Edit. Engl.* **37**, 1421-1423.
- Trent, J. O., Clark, G. R., Kumar, A., Wilson, W. D., Boyki, D. W., Hall, J. E., Tidwell, R. R., Blagburn, B. L. & Neidle, S. (1996) Targeting the minor groove of DNA: crystal structures of two complexes between furan derivatives of berenil and the DNA dodecamer d(CGCGAATTCGCG)<sub>2</sub>. *J. Med. Chem.* **39**, 4554-4562.
- Trotta, E., D'Ambrosio, E., Ravagnan, G., & Paci, M. (1995) Evidence for DAPI intercalation in CG sites of DNA oligomer [d(CGACGTCG)]<sub>2</sub>: A <sup>1</sup>H NMR study. *Nucleic Acids Res.* **23**, 1333-1340.

- Trotta, E., D'Ambrosio, E., Ravagnan, G., & Paci, M. (1996) Simultaneous and different binding mechanisms of 4', 6-diamidino-2-phenylindole to DNA hexamer (d(CGATCG))<sub>2</sub>. *J. Biol. Chem.* **271**, 27608-27614.
- Trotta, E. & Paci, M. (1998) Solution structure of DAPI selectively bound in the minor groove of a DNA TT mismatch-containing site: NMR and molecular dynamics studies. *Nucleic Acids Res.* **26**, 4706-4713.
- Trotta, E., Del Grosso, N., Erba, M. & Paci, M. (2000) The ATT strand of AAT.ATT trinucleotide repeats adopts stable hairpin structures induced by minor groove binding ligands. *Biochemistry* **39**, 6799-6808.
- Tsujimoto, Y., Finger, L. R., Yunis, J., Nowell, P. C. & Croce, C. M. (1984) cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* **226**, 1097-1099.
- Turner, J. M., Baird, E. E. & Dervan, P. B. (1997) Recognition of seven base pair sequences in the minor groove of DNA by ten ring pyrrole-imidazole polyamide hairpins. *J. Am. Chem. Soc.* **119**, 7636-7644.
- Turner, J. M., Swalley, S. E., Baird, E. E. & Dervan, P. B. (1998) Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **120**, 6219-6226.
- Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. & Sharp, P. A. (1999) Targeted mRNA degradation by double-stranded RNA *in vitro*. *Genes Dev.* **13**, 3191-3197.
- Tuntiwechapikul, W., David, W. D., Kumar, D., Salazaar, M. & Kerwin, S. M. (2002) DNA modification by 4-aza-3-ene-1,6-diynes: DNA cleavage, pH-dependent cytosine-specific interactions, and cancer cell cytotoxicity. *Biochemistry* **41**, 5283-5290.
- Uhlmann, E. (1998) Peptide nucleic acids (PNA) and PNA-DNA chimeras: from high binding affinity towards biological function. *Biol. Chem.* **379**, 1045-1052.

- Van Dyke, M. W., Hertzberg, R. P., & Dervan, P. B., (1982) Map Of distamycin , netropsin and actinomycin binding sites on heterogeneous DNA: DNA cleavage-inhibition patterns with methidiumpropyl-EDTA.Fe(II). *Proc. Natl. Acad. Sci. USA* **79**, 5470-5474.
- Van Dyke, M. W. & Dervan, P. B. (1983) Chromomycin, mithramycin, and olivomycin binding sites on heterogeneous deoxyribonucleic acid. Footprinting with (methidiumpropyl-EDTA)iron(II). *Biochemistry* **22**, 2373-2377.
- Varma, M. V., Ashokraj, Y., Dey, C. S. & Panchagnula R. (2003) P-glycoprotein inhibitors and their screening: a perspective from bioavailability enhancement. *Pharmacol Res.* **48**, 347-59.
- Vasquez, K. M., Wensel, T. G., Hogan, M. E. & Wilson, J. H.(1995) High-affinity triple helix formation by synthetic oligonucleotides at a site within a selectable mammalian gene. *Biochemistry* **34**, 7243-7251.
- Vasquez, K. M. & Wilson, J. H. (1998) Triplex-directed modification of genes and gene activity. *Trends Biochem. Sci.* **23**, 4-9.
- Vogelstein, B. & Kinzler, K. W. (1993) The multi-step nature of cancer. *Trends Genet.* **9**, 138-141.
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. & Bos, J. L. (1988) Genetic alterations during colorectal-tumour development. *N. Engl. J. Med.* **319**, 525-532.
- Volpe, D. A., Tomaszewski, J. E., Parchment, R. E., Garg, A., Flora, K. P., Murphy, M. J., Grieshaber, C. K. (1996) Myelotoxic effects of the bifunctional alkylating agent bizelesin on human, canine and murine myeloid progenitor cells. *Cancer Chemother Pharmacol.* **39**, 143-149.
- Wade, W. S. & Dervan, P. B. (1987) Alteration of the sequence specificity of distamycin on DNA by the replacement of an N-methylpyrrolicarboxamide with pyridine-2-carboxamide. *J. Am. Chem. Soc.* **109**, 1574-1575.



Wade, W. S., Mrksich, M. & Dervan, P. B. (1992) Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.

Wade, W. S., Mrksich, M. & Dervan, P. B. (1993) Binding affinities of synthetic peptides, pyridine-2-carboxamidenetropsin and 1-methylimidazole-2-carboxamidonetropsin, that form 2:1 complexes in the minor groove of double helical DNA. *Biochemistry* **32**, 11385-11389.

Walker, S., Landovitz, R., Ding, W.D., Ellestad, G. A. & Kahne, D. (1992) Cleavage behaviour of calicheamicin  $\gamma$ 1 and calicheamicin t. *Proc. Natl. Acad. Sci. USA* **89**, 4608-4612.

Walker, W. L., Kopka, M. L. & Goodsell, D. S. (1997) Progress in the design of DNA sequence specific lexitropsins. *Biopolymers* **44**, 323-334.

Walton A. R., Jenkins, T. C. & Neidle, S. (1991) Studies on 1,3-diaryltriazene analogues of berenil: molecules with potential GC base pair selectivity. *Acta Crystallogr B*. **47**, 771-775.

Walton, M. I., Goddard, P., Kelland, L. R., Thurston, D. E. & Harrap, K. R. (1996) Preclinical pharmacology and antitumour activity of the novel sequence-selective DNA minor groove cross-linking agent DSB-120. *Cancer Chemother. Pharmacol.* **38**, 431-438.

Wang, A. H., Cottens, S., Dervan, P. B., Yesinowski, J. P., Van Der Marel, G. A., & Boom, J. H. (1989) Interactions between a symmetrical minor groove binding compound and DNA oligonucleotides:  $^1\text{H}$  and  $^{19}\text{F}$  NMR studies. *J. Biomol. Struct. Dyn.* **7**, 101-117.

Wang, W. & Lown, J. W. (1992) Anti-HIV-1 activity of linked lexitropsins. *J. Med. Chem.* **35**, 2890-2897

Wang, Y., Gupta, R., Huang, L., Luo, W., & Lown, J. W. (1996) Design, synthesis cytotoxic properties and preliminary DNA sequencing evaluation of CPI-N-methylpyrrole hybrids. Enhancing effect of a trans double bond linker and role of the terminal amide functionality on cytotoxic potency. *Anticancer Drug Des.* **11**, 15-34.

- Wang, Y-D., Dziegielewski, J., Chang, A. Y., Dervan, P. B. & Beerman, T. A. (2002) Cell-free and cellular activities of a DNA sequence selective hairpin polyamide-CBI conjugate. *J. Biol. Chem.* **277**, 42431-42437.
- Wang, Y-D., Dziegielewski, J., Wurtz, N. R., Dziegielewska, B., Dervan, P. B. & Beerman, T. A. (2003) DNA crosslinking and biological activity of a hairpin polyamide-chlorambucil conjugate. *Nucleic Acids Res.* **31**, 1208-1215.
- Waring, M. (1970) Variations in the supercoils in closed circular DNA by binding of antibiotics and drugs; evidence for molecular models involving intercalation. *J. Mol. Biol.* **54**, 247-277.
- Waring, M. J. & Bailly, C. (1997) The influence of the exocyclic amino group characteristic of GC base pairs on molecular recognition of specific nucleotide sequences in DNA by berenil and DAPI. *J. Mol. Recogn.* **10**, 121-127.
- Warpehoski, M. A., Gebhard, I., Kelly, R. C., Krueger, W. C., Li, L. H., McGovren, J. P., Prairie, M. D., Wicniewski, N. & Wierenga, W. J., (1988) Stereoelectronic factors influencing the biological activity and DNA interaction of synthetic antitumour agents modelled on CC-1065. *J. Med. Chem.* **31**, 590-603.
- Warpehoski, M. A. & Hurley, L. H. (1988b) Sequence selectivity of covalent modification. *Chem. Res. Toxicol.* **1**, 315-333.
- Wartell, R. M., Larson, J. E. & Wells, R. D. (1974) Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid. *J. Biol.Chem.* **249**, 6719-6731.
- Watanabe, T. & Sullenger, B. A. (2000) Induction of wild type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proc. Natl. Acad. Sci. USA.* **97**, 8490-8494.
- Watanabe, H., Ha, A., Hu, Y. X., Ohtsubo, K., Yamaguchi, Y., Motoo, Y., Okai, T., Toya, D., Tanaka, N. & Sawabu, N. (1999) K-ras mutations in duodenal aspirate without secretin stimulation for screening of pancreatic and biliary tract carcinoma. *Cancer* **86**, 1441-1448.

- Watson, J. D. & Crick, F. H. C. (1953) Molecular structure of nucleic acids: a structure for deoxynucleic acids. *Nature* **171**, 737-738.
- Wemmer, D. E. & Dervan, P. B. (1997) Targeting the minor groove of DNA. *Curr. Opin. Struct. Biol.* **7**, 355-361.
- Weyermann, P. & Dervan, P. B. (2002) Recognition of ten base pairs of DNA by head to head hairpin dimers. *J. Am. Chem. Soc.* **124**, 6872-6878.
- White, S., Baird, E. E. & Dervan, P. B. (1996) Effects of AT/TA degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry* **35**, 12532-12537.
- White, S., Szewczyk, J. M., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998) Recognition of the four watson-crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-71 & Supplement.
- White, S., Turner, J. M., Szewczyk, J. M., Baird, E. E. & Dervan, P. B. (1999) Affinity and specificity of multiple hydroxypyrrole/pyrrole ring pairings for coded recognition of DNA. *J. Am. Chem. Soc.* **121**, 260-261.
- White, S., Baird, E. E. & Dervan, P. B. (1997) On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol* **4**:569-578.
- White, C. M., Satz, A. L., Gawron, L. S., Bruice, T. C. & Beerman, T. A. (2002) inhibiting transcription factor/DNA complexes using fluorescent microgonotropens (fmgs). *Biochim. Biophys. Acta* **1574**, 100-108.
- Wickstrom, E. (1986) Oligodeoxynucleotide stability in subcellular extracts and culture media. *J. Biochem. Biophys. Methods* **13**, 97-102.
- Winston, R. L., Ehley, J. A., Baird, E. E., Dervan, P. B. & Gottesfeld, J. M. (2000) Asymmetric DNA binding by a homodimeric bHLH protein. *Biochemistry* **39**, 9092-9098.
- Williamson, J. (1979) Effects of trypanocides on the fine structure of target organisms. *Pharm. Ther.* **7**, 445-512.

Wilson, W. D., Tanious, F. A., Barton, H. J., Streckowski, L. & Boykin, D. W. (1989) Binding of 4',6-diamidino-2-phenylindole (DAPI) to GC and mixed sequences in DNA: intercalation of a classical groove binding agent. *J. Am. Chem. Soc.* **111**, 5008-5010.

Wilson, W. D., Tanious, F. A., Barton, H. J., Jones, R. L., Fox, K., Wydra, R. L. & Streckowski, L. (1990) DNA sequence dependent binding modes of 4', 6-diamidino-2-phenylindole (DAPI). *Biochemistry* **29**, 8452-8460.

Wood, A. A., Nunn, C. M., Czarny, A., Boykin, D.W. & Neidle, S. (1995) Variability in DNA minor groove width recognised by ligand binding: the crystal structure of a bis-benzimidazole compound bound to the DNA duplex d(CGCGAATTCGCG)2. *Nucleic Acids. Res.* **23**, 3678-3684.

Woodburn, J. R., Barker, A. J., Gibson, K. H., Ashton, S. E., Wakeling, A. E., Carry, B. J., Scarlett, L. & Henthorn, L. R. (1997) ZD1839, an epidermal growth factor tyrosine kinase inhibitor selected for clinical development. *Proc. Am. Assoc. Cancer Res.* **38**, 633.

Wojnarowski, J. M., Napier, C., Trevino, A. V., & Arnett, B. (2000) Region-specific DNA damage by AT-specific DNA-reactive drugs is predicted by drug binding specificity. *Biochemistry* **39**, 9917-9927.

Wojnarowski, J. M., (2002) Targeting critical regions in genomic DNA with AT-specific anticancer drugs. *Biochim. Biophys. Acta* **1587**, 300-308.

Wurtz, N. R. & Dervan, P. B. (2000) Sequence specific alkylation of DNA by hairpin pyrrole-imidazole polyamide conjugates. *Chem. Biol.* **7**, 153-161.

Wyatt, M. D., Garbiras, B. J., Haskell, M. K., Lee, M., Souhami, R. L. & Hartley, J. A. (1994) Structure-activity relationship of a series of nitrogen mustard- and pyrrole-containing minor groove-binding agents related to distamycin. *Anticancer Drug Des.* **9**, 511-525.

Wyatt, M. D., Lee, M., Garbiras, B. J., Souhami, R. L. & Hartley, J. A. (1995) Sequence specificity of alkylation for a series of nitrogen mustard-containing

analogues of distamycin of increasing binding site size: evidence for increased cytotoxicity with enhanced sequence specificity. *Biochemistry* **34**, 13034-13041.

Xiong, Y., Hannon, G., Zhang, H., Casso, D., Kobayashi, R. & Beach, D. (1993) p21 is a universal inhibitor of cyclin kinases. *Nature* **366**, 701-704.

Xu L, & Massague J. (2004) Nucleocytoplasmic shuttling of signal transducers. *Nat Rev Mol. Cell Biol.* **5**, 209-219.

Yakubov, L. A., Deeva, E. A., Zarytova, V. F., Ivanova, E. M., Ryte, A. S., Yurchenko, L. V., Vlassov, V. V. (1989) Mechanism of oligonucleotide uptake by cells: involvement of specific receptors? *Proc. Natl. Acad. Sci. USA* **86**, 6454-6458.

Yang, X-L., Kaenzig, C., Lee, M. & Wang, A. H-J. (1999a) Binding of AR-1-144, a tri-imidazole DNA minor groove binder, to CCGG sequence analysed by NMR spectroscopy. *Eur. J. Biochem.* **263**, 646-655.

Yang, X-L, Hubbard, R. B., Lee, M., Tao, Z-F, Sugiyama, H. & Wang, A. H-J. (1999b) Imidazole-imidazole pair as a minor groove recognition motif for T: G mismatched base pairs. *Nucleic Acids Res.* **27**, 4183-4190.

Yasuzawa, T., Iida, T., Muroi, K., Ichimura, M., Takahashi, K., & Sano, H. (1988) Structures of the duocarmycins, novel antitumour antibiotics produced by *Streptomyces* Sp. *Chem. Pharm. Bull.* **36**, 3728-3731.

Yoshida, M., Banville, D. L. & Shafer, R. H. (1990) Structural analysis of d(GCAATTGC)<sub>2</sub> and its complex with berenil by nuclear magnetic resonance spectroscopy. *Biochemistry* **29**, 6585-6592.

Youngquist, R. S. & Dervan, P. B., (1985) Sequence-specific recognition of B-DNA by oligo(N-methylpyrrolocarboxamide). *Proc. Natl. Acad. Sci. USA* **82**, 2565-2569.

Youngquist, R. S. & Dervan, P. B., (1987) A synthetic peptide binds 16 base pairs of A,T double helical DNA. *J. Am. Chem. Soc.* **109**, 7564-7566.

- Zakrzewska, K. & Pullman, B. (1988) Theoretical study of the sequence selectivity of isolexins, isohelical DNA groove binding ligands. Proposal for GC minor groove specific compounds. *J. Biomol. Struct. Dyn.* **5**, 1043-1058.
- Zakrzewska, K., Randrianarivelino, M. & Pullman, B. (1988) Drug recognition of DNA. Proposal for GC minor groove specific ligands: vinylexins. *J. Biomol. Struct. Dyn.* **6**, 331-344.
- Zamecnik, P. C. & Stephenson, M. L. (1978) Inhibition of rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc. Natl. Acad. Sci. USA* **75**, 280-284.
- Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. (2000) RNAi. Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25-33.
- Zasedatelev, A. S., Gursky, G. V., Zimmer, C. & Thrum, H. (1974) Binding of netropsin to DNA and synthetic polynucleotides. *Mol. Biol. Rep.* **1**, 337-342.
- Zein, N., Sinha, A. M., McGahren, W. J. & Ellestad, G. A., (1988) Calicheamicin  $\gamma_1$ , an antitumour antibiotic that cleaves double stranded DNA site specifically. *Science* **240**, 1198-1201.
- Zimmer, C., Puschendorf, B., Grunicke, H., Chandra, P. & Venner, H. (1971) Influence of netropsin and distamycin A on the secondary structure and template activity of DNA. *Eur. J. Biochem.* **21**, 269-278.
- Zimmer, C., Marck, C., Schneider, C. & Guschlbauer, W. (1979) Influence of nucleotide sequence on dA.dT-specific binding of netropsin to double stranded DNA. *Nucleic Acids Res.* **6**, 2831-2837.
- Zimmer, C. & Wahnert, U. (1986) Non-intercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog. Biophys. Mol. Biol.* **47**, 31-112.

Zinnen, S. P., Domenico, K., Wilson, M., Dickinson, B. A., Beaudry, A., Molker, V., Daniher, A. T., Burgin, A. & Beigelman, L. (2002) Selection, design and characterisation of a new potentially therapeutic ribozyme. *RNA* **8**, 214-228.

## **PUBLICATIONS FROM THIS THESIS**

O'Hare, C. C., Mack, D., Tandon, M., Sharma, S. K., Lown, J. W., Kopka, M. L., Dickerson, R. E. & Hartley, J. H. (2002) DNA sequence recognition in the minor groove by cross-linked polyamides: the effect of N-terminal head group and linker length on binding affinity and specificity. *Proc. Natl. Acad. Sci. USA* **99**, 72-77.

Lacy, E. R., Nguyen, B., Le, M., Cox, K. K., O'Hare, C. C., Hartley, J. A., Lee, M. & Wilson WD. (2004) Energetic basis for selective recognition of T\*G mismatched base pairs in DNA by imidazole-rich polyamides. *Nucleic Acids Res.* **32**, 2000-2007.

# DNA sequence recognition in the minor groove by crosslinked polyamides: The effect of N-terminal head group and linker length on binding affinity and specificity

C. Caroline O'Hare\*, Darcy Mack†, Manju Tandon‡, Sanjay K. Sharma‡, J. William Lown‡, Mary L. Kopka†, Richard E. Dickerson†, and John A. Hartley\*§

\*Cancer Research Campaign Drug-DNA Interactions Research Group, Department of Oncology, Royal Free and University College Medical School, UCL, 91 Riding House Street, London, W1W 7BS, United Kingdom; †Molecular Biology Institute, University of California, Los Angeles, CA 90095; and ‡Department of Chemistry, University of Alberta, Edmonton, AB, Canada T6G 2G2

Contributed by Richard E. Dickerson, November 2, 2001

**Development of sequence-reading polyamides or "lexitropsins" with comparable DNA-binding affinities to cellular proteins raises the possibility of artificially regulated gene expression. Covalent linkage of polyamide ligands, with either a hairpin motif or crosslinking methylene bridge, has greatly improved binding affinity by ensuring their side-by-side register. Whereas hairpin polyamides have been investigated extensively, the optimized structure of crosslinked polyamides remains to be determined. This study examines a series of thiazole-imidazole-pyrrole (TIP) monomers and crosslinked dimers to evaluate the effects on selectivity and binding affinity of different N-terminal head groups attached to the leading thiazole ring and differing methylene linker lengths. Quantitative footprinting of a DNA sequence, containing potential match and mismatch sites for both maximum overlap and one-residue stagger binding modes, allowed measurement of binding constants at each putative site. Within an N-terminal amino TIP series, C7 and C8-linked compounds bound most strongly to these sites, whereas maximum binding affinity was observed for a C6 linker with a formyl head group. A C5 linker gave weak binding with either head group. A hydrogen or acetyl head group abrogated binding. Binding was confirmed by gel shift analyses. The highest specificity for the maximum overlap site over the one-residue stagger was observed with TIP-C7-amino. Selectivity of the leading thiazole was modulated by the head group, with N-terminal formyl TIP exhibiting up to 3-fold specificity for AGT over TGT, suggesting that N-formyl-thiazole may provide sequence discrimination of adenine over thymine. Moreover, the leading head group and methylene linker length significantly influences the binding characteristics of crosslinked polyamides.**

**M**odification of the sequence specificity of DNA minor groove binders netropsin and distamycin from AT to GC (1, 2) has led to the development of sequence-reading polyamides or lexitropsins. Strict AT sequence recognition inherent in the naturally occurring tripyrrole carboxamides has been expanded by replacement of the guanine-excluding pyrrole ring with a number of structurally related heterocycles including imidazole (3). Following the demonstration that distamycin may bind in a 2:1 binding mode (4), oligopeptides containing imidazole were found also to bind cooperatively in this motif to GC tracts (5, 6).

The current generation of polyamides binds as antiparallel dimers to opposite strands of the minor groove with each constituent peptide ring contacting an individual nucleotide base. Covalent linkage of the two component oligopeptides has produced sequence-reading ligands of improved binding affinity (7–13). As a consequence, synthesis of small molecules capable of DNA sequence discrimination with equivalent binding affinities to those of DNA-binding proteins raises the possibility of artificial regulation of gene expression (14). Indeed, recent work

has demonstrated that linked polyamides can modulate transcription of specific genes *in vitro* and in some cell culture situations (15–20).

Detailed studies of polyamide binding led to the development of pairing rules defining optimal combinations of imidazoles and pyrroles for binding at predetermined DNA sites (21–23). These rules were extended to address the problem of AT/TA discrimination through use of a 3-hydroxypyrrole unit for preferential thymine recognition, providing the first success in complete sequence recognition in the minor groove of DNA (24–26). Nevertheless, alternative solutions continue to be pursued in an effort to improve AT/TA discrimination, which include the potential use of thiazole as an adenine discriminator (27–29).

Accurate alignment of the peptide rings alongside each base pair is crucial to providing both sequence specificity and binding affinity for linked polyamides (6, 30, 31). Optimal length and geometry of the linkage is required to position both component ligands in close proximity to the base edges within the minor groove. To achieve this, two types of structures have been proposed: a hairpin, whereby the oligopeptides are linked head to tail by using a suitable "turn peptide" (8), and a crosslinked or stapled structure, where the central rings of two polyamide ligands are linked via a methylene bridge (7, 12, 32, 33). Whereas the binding affinity and selectivity of hairpin polyamides are well documented (34), the optimum conditions for methylene chain crosslinked polyamides remain to be defined.

For linked polyamides, two potential DNA-binding modes, an overlap and a staggered orientation, are possible (Fig. 1) (29). In the maximum overlap mode, favored by the hairpin linkage (9, 35), the absence of an amide on the leading ring holds the parallel peptide chains in specific register, whereby each ring stacks on an amide of its neighbor. In the one-residue stagger motif, polyamides with a leading amide head group can slip along the peptide chain by one amide yielding an extended reading frame. In this binding mode, polyamides also stack ring on amide; however, the first amide on each ligand overhangs. This sliding of the ligands in relation to each other allows further separation of the charged tails, reducing electrostatic repulsion by maximizing the distance between the charges.

In this report, we present a detailed study on the effect of the N-terminal head group and methylene linker length on the binding affinity and specificity of a series of thiazole-imidazole-

Abbreviation: TIP, thiazole-imidazole-pyrrole.

§To whom reprint requests should be addressed. E-mail: john.hartley@ucl.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



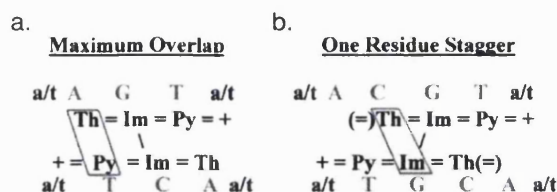


Fig. 1. Binding motifs of TIP. The box indicates which rings are side by side in the minor groove. (a) Maximum overlap motif. Th, thiazole; Im, imidazole; Py, pyrrole. The equal sign signifies an amide, and the plus sign signifies the charged tail. In the maximum overlap motif, the rings stack ring-on-amide. The DNA target sequence is indicated in capital letters, and the charged tails prefer adenine or thymine, a/t. (b) One residue stagger motif. This motif requires a leading amide shown in parentheses. Here, the polyamides slip relative to the maximum overlap, still stacking ring on amide, but moving the charged tails further apart. The one residue stagger motif has a longer reading frame but still preserves the same number of ring-on-amide interactions.

pyrrole (TIP) crosslinked polyamides (Fig. 2) by using quantitative DNase I footprinting and gel shift analyses.

#### Materials and Methods

**Quantitative DNase I Footprinting.** A 355-bp 5' <sup>32</sup>P-labeled PCR fragment was generated from the long terminal repeat region of plasmid pBSFIV34TF10, kindly provided by Tom Phillips (Vaccine Research Institute, San Diego, CA), using standard protocols. The labeled fragment was purified by agarose gel electrophoresis and isolated by using a Bio101 kit according to the manufacturer's instruction. Polyamides (prepared as in ref. 33) were incubated with 1,000 cps of 5' single end-labeled fragment

in 10 mM Tris, pH 7.0/1 mM EDTA/50 mM KCl/1 mM MgCl<sub>2</sub>/0.5 mM DTT/20 mM Hepes at room temperature for 30 min in a total volume of 50  $\mu$ l. Cleavage was initiated by the addition of the drug-treated sample to 2  $\mu$ l (0.1 unit) DNase I diluted in ice-cold 10 mM Tris, pH 7.0, from a stock solution (1 unit/ $\mu$ l, Promega) and 1  $\mu$ l of a solution of 250 mM MgCl<sub>2</sub>/250 mM CaCl<sub>2</sub>. The reactions were performed at room temperature and stopped after 3 min by the addition of 100  $\mu$ l of a stop mix containing 200 mM NaCl/30 mM EDTA, pH 8/1% SDS. The cleavage products were phenol/chloroform extracted and ethanol precipitated in the presence of 1  $\mu$ l of glycogen (20 mg/ml, Roche Diagnostics), washed once in 80% ethanol and lyophilized dry. The samples were resuspended in formamide loading dye, denatured for 5 min at 90°C, cooled on ice, and electrophoresed at 2,000 V for 2 h on a 6% acrylamide denaturing gel (Sequagel, National Diagnostics). The gels were dried under vacuum at 80°C and exposed to film for 24 h (X-Omat, Kodak). Densitometry was carried out on a Bio-Rad GS-670 imaging densitometer.

Equilibrium association constants were determined for target-binding sites for each polyamide (36–39). Volume integrations of five target sites and a reference site allowed the apparent DNA target site saturation,  $\theta_{app}$ , to be calculated for each concentration of ligand using the following equation:

$$\theta_{app} = 1 - (I_{tot}/I_{ref})/(I_{0 tot}/I_{0 ref}), \quad [1]$$

where  $I_{tot}$  and  $I_{ref}$  correspond to the integrated volumes of the target and reference sites, respectively.  $I_{0 tot}$  and  $I_{0 ref}$  are the integrated volumes for those sites in the absence of polyamide. The data were fitted to a Langmuir binding model (Eq. 2), where  $[L]_{tot}$  is the total polyamide concentration,  $K_a$  is the equilibrium association constant, and  $\theta_{min}$  and  $\theta_{max}$  are the site-saturation values when the site is unoccupied or saturated, respectively.

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min}) \frac{K_a [L]_{tot}}{1 + K_a [L]_{tot}} \quad [2]$$

Values for  $\theta_{min}$ ,  $\theta_{max}$ , and  $K_a$  were obtained by minimizing the difference between  $\theta_{app}$  and  $\theta_{fit}$ , using a nonlinear optimization procedure with the constraint  $\theta_{max} > \theta_{min}$ . Goodness of fit was assessed by calculating a Pearson correlation coefficient between the actual and the model values. The confidence level was obtained by using a test statistic for a  $t$  distribution, and any fit with  $P > 0.05$  was rejected. At least three sets of data were used in determining each association constant for five target match and mismatch sites.

**Gel Shift Assays.** A DNA decamer and dodecamer oligomer were prepared by solid phase synthesis at a 5  $\mu$ M scale on an Eppendorf ECOSYN D300 synthesizer. They were purified on a DE-52 ion exchange column using a KCl gradient, lyophilized, and resuspended in water to achieve the proper concentration. Polyamides were suspended in distilled water at the appropriate concentration.

Native gels (20%) using a 29:1 acrylamide:bis ratio were run at 4°C. The polyamides and the various DNA oligomers were incubated overnight at room temperature while shaking with 66 mM Mg(OAc)<sub>2</sub> and 3.33 mM sodium cacodylate, pH 7.0. Dye and Ficoll were then added so that the final concentration/lane was 250 ng DNA/50 mM Mg(OAc)<sub>2</sub>/2.5 mM sodium cacodylate, pH 7.0/3.375% Ficoll/0.0125% xylene cyanol/0.0125% bromophenol blue in a volume of 20  $\mu$ l. Gels were silver stained (40) by incubating the gel for 15 min with 1% HNO<sub>3</sub> followed by 15 min with 0.2% AgNO<sub>3</sub>. The gels were then rinsed quickly with water followed by developer for 5–10 min. The gels were fixed with 10% acetic acid. After fixing, the acetic acid was removed by rinsing with distilled water before drying.

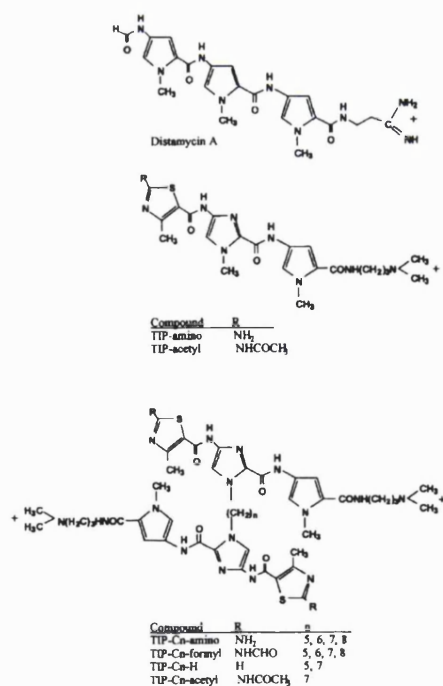
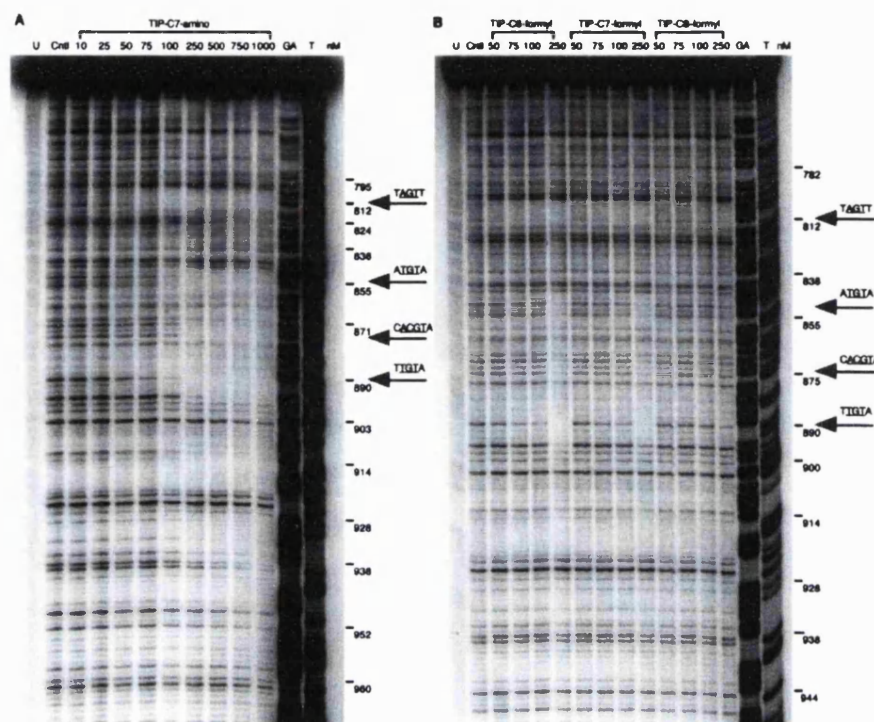


Fig. 2. Molecular structures of distamycin A, the unlinked thiazole-imidazole-pyrrole polyamides, and the crosslinked thiazole-imidazole-pyrrole polyamides used in this study.



**Fig. 3.** (A) Quantitative DNase I footprint titration experiment with TIP-C7-amino on the 5' <sup>32</sup>P-labeled 355-bp FIV34TF10 long terminal repeat fragment from plasmid pBSFIV34TF10, showing four binding sites. All reactions contain 1,000 cps of restriction fragment, 10 mM Tris, pH 7.0/1 mM EDTA/50 mM KCl/1 mM MgCl<sub>2</sub>/0.5 mM DTT/20 mM Hepes. Lane 1, untreated control; lane 2, DNase I-treated control; lanes 3–11, 10 nM, 25 nM, 50 nM, 75 nM, 100 nM, 250 nM, 500 nM, 1 μM TIP-C7-amino; lane 12, GA reaction; lane 13, T reaction. (B) DNase I footprint experiment with TIP-C6-formyl, TIP-C7-formyl, and TIP-C8-formyl on the 5' <sup>32</sup>P-labeled 355-bp FIV34TF10 long terminal repeat fragment from plasmid pBSFIV34TF10, showing four binding sites. All reactions contain 1,000 cps of restriction fragment, 10 mM Tris, pH 7.0/1 mM EDTA/50 mM KCl/1 mM MgCl<sub>2</sub>/0.5 mM DTT/20 mM Hepes. Lane 1, untreated control; lane 2, DNase I-treated control; lanes 3–6, 50 nM, 75 nM, 100 nM, 250 nM TIP-C6-formyl; lanes 7–10, 50 nM, 75 nM, 100 nM, 250 nM TIP-C7-formyl; lanes 11–14, 50 nM, 75 nM, 100 nM, 250 nM TIP-C8-formyl; lane 15, GA reaction; lane 16, T reaction.

## Results

**Quantitative DNase I Footprinting.** DNase I footprinting allowed the determination of binding affinity and binding specificity for a series of monomer and crosslinked dimer polyamides (Fig. 2). The compounds contained one of four different N-terminal head groups, and the crosslinked compounds contained methylene linkers ranging from 5 to 8. A typical footprint titration for compound TIP-C7-amino is shown in Fig. 3A. Putative match and mismatch sites were present in the DNA sequence studied for both the maximum overlap and one-residue stagger binding modes (Fig. 1). Four of these sites are indicated in Fig. 3A. Association constants for these sites, and a further site at sequence position 971, were determined for all compounds and are given in Table 1. No significant binding was observed at any other DNA sequences within the DNA fragment tested, including at a stagger mode binding site with a single base mismatch.

**Binding Affinity.** Within the N-terminal amino series, the C7- and C8-linked compounds produced the strongest binding. Reduction in the methylene chain length to C6 produced a 15-fold decrease in binding affinity at the maximum overlap match site relative to the C7- and C8-linked compounds. Further reduction to a chain length of 5 methylenes produced weak overall binding. The unlinked TIP-amino bound with an affinity similar to that of the C6-linked compound.

In contrast, for the N-terminal formyl series, the C6 linker produced the strongest overall binding at its target site, having the greatest binding affinity ( $K_a = 4.0 \times 10^7 \text{ M}^{-1}$ ) for any compound in either series. The C7-linked formyl compound had an approximately 2-fold decreased binding affinity compared with the C6 compound at most binding sites. The C8-linked compound gave a further decrease in binding affinity. A comparison of the C6–C8 compounds is shown in Fig. 3B. Similarly to the amino series, the C5-linked formyl compound was a weak binder. No significant binding was observed when the TIP compounds contained an N-terminal hydrogen or acetyl head group in either the unlinked or crosslinked form.

**Binding Specificity.** Comparison of the binding affinities at the individual binding sites allowed a determination of the relative binding of the TIP compounds between the maximum overlap (sites 812, 890, 851, and 971) and stagger (site 871) modes of binding. In addition, comparison of the four maximum overlap sites allowed an assessment of the discrimination of the compounds for A:T versus T:A sequences. Table 2 lists the ratios of binding at the stagger mode binding site (5'-ACGT-3') and maximum overlap mismatch sites (5'-TGT-3' and 5'-AGA-3') compared with the potential maximum overlap match site (5'-AGT-3').

Within the amino series, only the C7 and C8-linked com-



Table 1. Equilibrium association constants ( $K_a \times 10^7 \text{ M}^{-1}$ ) for the TIP polyamides

Polyamide	$K_a \times 10^7 (\text{M}^{-1})$				
	TAGTT (812)	TGTGA (890)	ATGTA (851)	CAGAT (971)	CACGTA (871)
TIP-amino	0.2 ( $\pm 0.04$ )	0.1 ( $\pm 0.1$ )	0.07 ( $\pm 0.07$ )	0.06 ( $\pm 0.04$ )	0.2 ( $\pm 0.2$ )
TIP-C5-amino	<0.1	<0.1	<0.1	<0.01	<0.01
TIP-C6-amino	0.2 ( $\pm 0.2$ )	0.5 ( $\pm 0.4$ )	0.3 ( $\pm 0.2$ )	0.7 ( $\pm 0.5$ )	0.3 ( $\pm 0.2$ )
TIP-C7-amino	2.7 ( $\pm 1.6$ )	2.3 ( $\pm 1.4$ )	1.0 ( $\pm 0.4$ )	2.8 ( $\pm 1.6$ )	0.4 ( $\pm 0.1$ )
TIP-C8-amino	3.1 ( $\pm 1.8$ )	3.3 ( $\pm 2.1$ )	1.6 ( $\pm 0.7$ )	2.4 ( $\pm 0.7$ )	1.1 ( $\pm 0.7$ )
TIP-C5-formyl	<0.1	<0.1	<0.1	<0.1	<0.1
TIP-C6-formyl	4.0 ( $\pm 1.8$ )	2.4 ( $\pm 1.5$ )	1.2 ( $\pm 0.85$ )	3.4 ( $\pm 1.8$ )	1.2 ( $\pm 0.8$ )
TIP-C7-formyl	2.4 ( $\pm 1.0$ )	1.2 ( $\pm 0.6$ )	0.7 ( $\pm 0.02$ )	1.2 ( $\pm 0.6$ )	0.4 ( $\pm 0.1$ )
TIP-C8-formyl	1.3 ( $\pm 0.4$ )	1.0 ( $\pm 0.4$ )	0.5 ( $\pm 0.2$ )	1.7 ( $\pm 0.6$ )	0.4 ( $\pm 0.1$ )
TIP-C5-H	<0.01	<0.01	<0.01	<0.01	<0.01
TIP-C7-H	<0.01	<0.01	<0.01	<0.01	<0.01
TIP-acetyl	<0.01	<0.01	<0.01	<0.01	<0.01
TIP-C7-acetyl	<0.01	<0.01	<0.01	<0.01	<0.01

Values reported are the mean from at least three independent DNase I footprinting experiments, with the standard deviation indicated in parentheses.

pounds showed specificity for the maximum overlap over the stagger mode of binding (Table 1). This was 7-fold in the case of TIP-C7-amino. The C6–C8 formyl compounds showed specificity for the maximum overlap mode, which was again maximal for the C7 compound (6-fold). The unlinked TIP-amino compound gave comparable binding at the maximum overlap match site and the one-residue stagger site. This compound, however, did show some discrimination between the maximum overlap match site and the three mismatch sites, indicating that the thiazole/pyrrole pair could discriminate between A:T and T:A. This specificity was also maintained with the formyl-linked compounds and was over 3-fold in some cases. The specificity for the match site was less with the linked amino compounds and absent in the case of the TIP-C6-amino.

**Polyamide–DNA Gel Shift Assays.** Concurrent with the footprinting experiments, PAGE native gel shift experiments were conducted on two oligomers, which were eventually to be used in crystallization trials. The gel shift results gave an additional assessment of the binding ability of the various polyamides with different linkers and head group. A decamer, CCAACGTTGG, contained a single site 5'-ACGT-3' for the one-residue stagger binding mode. A dodecamer, CGACTGCAGTCG, contained three binding sites: two maximum overlap sites, 5'-ACT-3' and 5'-AGT-3', and a further conceivable site, 3'-ACGT-5', in the one residue stagger mode. But, reading in the reverse direction, 3' to 5' has been reported to occur only under special conditions, e.g., N-terminal acetylation or addition of a glycine into the tail region (41). Side-by-side polyamides tend to always read in a 5' to 3' direction.

Complete binding of the TIP-C7-amino at the one-residue stagger match site in the decamer, CCAACGTTGG, at a

polyamide to DNA ratio of 2:1 and above was demonstrated (Fig. 4A). The unlinked TIP-amino also displayed complete binding at this site, but a higher polyamide to DNA ratio of 4:1 was necessary. In contrast, the TIP-C5-amino was unable to bind,

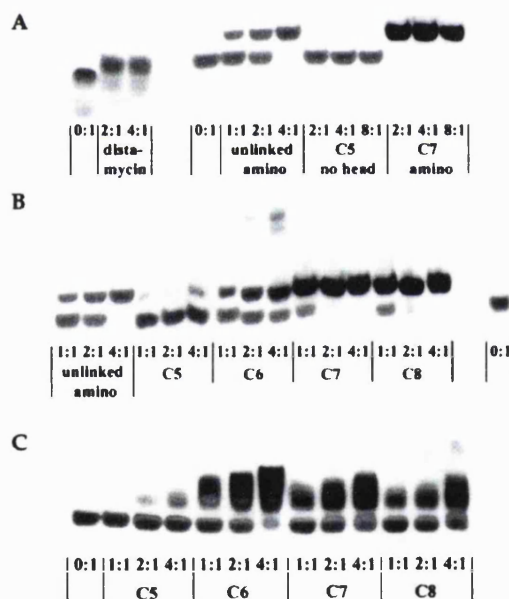


Fig. 4. Gel shift where 0:1 indicates no polyamide, only DNA oligomer. (A) TIP with various head groups. Distamycin bound to AAAATCTCTA is a control, with the drug binding to AAAA. The remaining lanes show polyamide:DNA ratios for TIP:CCAACGTTGG, with a DNA concentration of 2  $\mu\text{M}$ . The oligonucleotide contains a single ACGT one-residue stagger site. The head group and linker length is indicated. (B) TIP with amino head group. The two controls are unlinked TIP and 0:1, where 1 is the DNA sequence, CCAACGTTGG without polyamide. C5–C8 indicate the polyamide methylene linkers. (C) TIP with formyl head group. C5–C8 indicate the polyamide methylene linkers. The DNA sequence is CGACTGCAGTCG (1.7  $\mu\text{M}$ ), which contains three potential binding sites; a maximum overlap match site, AGT, a maximum overlap mismatch site, ACT, and a one-residue stagger site ACGT placed in the less favorable 3' to 5' direction. Note the diffuse bands in the shifted region, indicating multiple binding sites.

Table 2. Effects of linker length and head group on specificity of binding, where specificity is defined as  $K_a(\text{AGT})/K_a(\text{XGX})$

Polyamide	AGT/TGT (812/890)	AGT/TGT (812/851)	AGT/AGA (812/971)	AGT/ACGT (812/871)
TIP-amino	2.0	2.9	3.3	1.0
TIP-C6-amino	0.4	0.7	0.3	0.7
TIP-C7-amino	1.2	2.7	1.0	6.8
TIP-C8-amino	0.9	1.9	1.3	2.8
TIP-C6-formyl	1.7	3.3	1.2	3.3
TIP-C7-formyl	2.0	3.4	2.0	6.0
TIP-C8-formyl	1.3	2.6	0.8	3.3

even at an elevated 8:1 ratio of polyamide to DNA. Further gels using this oligomer were performed to establish the optimal length of methylene chain required for binding by the amino and formyl compound series. These indicated that within the N-terminal amino series, TIP-C7-amino and TIP-C8-amino produced the best binding for this site (Fig. 4B). The binding order was ascertained to be C7,C8 > unlinked > C6 > C5, respectively. In contrast, the N-terminal formyl polyamides showed optimal binding with a linker of six carbons (data not shown), and the binding order was C6 > C7 > C8 > C5.

Gel shifts with the second oligomer, CGACTGCAGTCG, determined optimal binding in the maximum overlap mode. The preferred sequence for this dodecamer in the overlap motif is AGT (Fig. 1), as the alignment of drug rings places an Im/Im pair next to a GC base pair. The Im/Im pair is redundant and can read either CG or GC. Assays with the N-terminal amino compounds using this oligomer resulted in the same order of overall binding as seen with the decamer above. With the formyl head group (Fig. 4C), the C6 linker proved to be optimal, and once more the compounds followed the binding order observed with the one-residue stagger motif.

With this DNA sequence, which has two binding sites in the maximum overlap motif, the gel shifts showed a smear indicating a number of conformers. This smearing was present on gels with either head group, amino or formyl. As the Im/Im pair in the TIP dimer is redundant and may read either GC or CG, this may account for two of the bands, ACT and AGT, in the diffused shifted region. Some of the lanes indicate a third band (Fig. 4C), which could be explained by having two polyamides bind concurrently.

The formyl head group tested in the gel shifts had a lower affinity for the DNA in both binding motifs. This was indicated by an increase in the drug to DNA ratio needed for effective binding. Optimal binding for the amino head group occurred at a 2:1 drug:DNA ratio for the ACGT sequence, whereas the best binding with the formyl occurred at a 4:1 ratio. However, the C6-formyl had a high affinity for AGT in the footprint results.

## Discussion

The binding affinities of a series of thiazole-imidazole-pyrrole polyamide monomers and crosslinked dimers were determined to evaluate the effects on selectivity and binding of the introduction of different N-terminal head groups attached to the leading thiazole ring, together with variation in the linking methylene chain length. Quantitative DNase I footprint titrations of a DNA sequence, containing putative match and mismatch sites for both the maximum overlap and one residue stagger binding modes, allowed the measurement of binding constants at each site. The presence of alternative mismatches for a given match site enabled specificity of binding to be determined and, consequently, the potential base recognition properties of the thiazole to be assessed. Concurrent gel shift experiments examined binding to short oligomers containing match sites for both modes of binding.

Quantitative DNase I footprinting demonstrated that within the N-terminal amino-TIP series, the C7 and C8-linked compounds bound most strongly to the putative binding sites. No significant binding was observed at sites other than these. The TIP-C7-amino proved the most selective, with up to a 7-fold selectivity for the maximum overlap match site, AGT. A significant decrease in both affinity and selectivity was seen in this series with reduced methylene chain lengths of five or six carbons. Alteration of the leading head group to a formyl resulted in an optimal methylene linker of six carbons. TIP-C6-formyl produced the strongest binding with a 3-fold specificity for the overlap match site over the one-residue stagger site. Reduced binding was seen with the TIP-C7-formyl, which nevertheless showed up to a 6-fold specificity for the overlap match

site. The C8-linked formyl demonstrated a decrease in both binding affinity and specificity. Gel shift experiments strongly supported the footprinting data, giving the same order of binding affinity in both the amino and formyl series. Overall, however, the formyl series demonstrated a lower affinity for the DNA in both binding modes, requiring an elevated drug to DNA ratio. This is consistent with the CD experiments of Burckhardt and coworkers (42) comparing C7-linked compounds with an amino or formyl head group.

Certain linkage and head group combinations proved particularly weak binders in both footprinting and gel shift experiments. For example, weak binding was observed with a five-carbon linker for either the N-terminal formyl or amino compound series. No binding was observed for C5 and C7-linked compounds lacking an N-terminal head group (hydrogen only) or with unlinked and C7-linked polyamides with an acetyl head group. These observations suggest that the presence of an N-terminal hydrogen or acetyl may abrogate binding in crosslinked polyamides, as was observed previously with C7-linked compounds by Burckhardt and coworkers (42). This is not the case, however, for hairpin polyamides, which bind efficiently with an N-terminal hydrogen (34).

Whereas polyamides linked using a hairpin motif have been extensively investigated (34), optimized binding conditions for crosslinked polyamides remain to be defined. Several studies of crosslinking in polyamides have demonstrated significant improvements in binding on linkage of the individual polypeptide ligands (7, 13, 42). The detailed assessment of linkage with a range of possible linkers presented here indicates a C6 linker with an N-terminal formyl, and a C7 linker with an N-terminal amino provides suitable alignment of the polyamide dimer within the minor groove. This is lost when the linker is reduced to five carbons, possibly because comprehensive contact of the dimeric ligand with both strands of the DNA is restricted. Furthermore, a C5 linker provides unfavorable binding compared with an unlinked monomer, suggesting that although suitable linkage enhances binding, suboptimal linkage may diminish effective binding.

This study reveals that the nature of the N-terminal moiety has a profound influence on the binding characteristics of crosslinked polyamides. Both the footprinting and gel shift data indicate that the use of a hydrogen or acetyl head group may inhibit binding. Furthermore, replacement of an amino head group with a formyl alters the optimal linker required for binding. Therefore, there may be a combined effect on polyamide binding by the linkage and head group. Polyamide specificity may also be modulated in this way, resulting in an overall improved specificity for match site binding with an N-terminal formyl.

It is significant that both of the successful head groups, amino and formyl, have an NH group capable of hydrogen bonding with DNA, whereas the unsuccessful head groups, -H and acetyl, do not. In a side-by-side unlinked di-imidazole complex with leading formyl groups (28), both of the formyl NH make hydrogen bonds with the floor of the minor groove: 2.94 Å from cytosine O<sub>2</sub> or 2.80 Å from guanine N<sub>3</sub> (see Table 1 of ref. 28). Hydrogen-bonded interactions of the head group with DNA seem to be required for crosslinked side-by-side binding, although such an interaction is not demanded for a hairpin complex (34).

The thiazole ring, tested here in the terminal position of the polyamide as a reading element for adenine, bound its AGT target in the maximum overlap motif with both the amino and formyl head groups. The overlap motif pairs the thiazole ring with pyrrole; thus, thiazole reads adenine and pyrrole thymine as expected. The footprinting results with TIP-C6-formyl showed a preference for AGT over TGT by up to 3-fold. Furthermore, the footprinting indicated that AGT is strongly favored over ACGT, the one-residue stagger motif, in which thiazole is next

to cytosine and imidazole is next to guanine. Therefore, thiazole seems to favor adenine moderately in a Th/Py pair, with poor CG specificity when paired with imidazole.

Our results therefore indicate that the thiazole moiety provides moderate binding and discrimination of adenine when in the polyamide N-terminal position. A previous study concluded that a thiazole/pyrrole pair binds poorly to all four Watson-Crick base pairs (43). However, in that study the thiazole was placed internally in a hairpin. Positioning of the heterocyclic moiety within a polyamide may be crucial. Indeed, the successful T/A discrimination seen with hydroxypyrrole sited internally within the hairpin motif has not been found when hydroxypyrrole is placed at the terminal position (44). Binding affinities of hairpins containing hydroxypyrrole vary considerably based on the position of the hydroxypyrrole ring within the hairpin motif (24). Calculation of relative binding affinities of imidazole and hydroxypyrrole rings by Goodsell and coworkers (45) indicates that the hydroxypyrrole ring binds with a lower affinity for AT base pairs compared with the affinity of the imidazole ring for GC pairs. In the present study, the specificity for adenine observed with thiazole in the terminal position is comparable with the specificity for thymine provided by a hydroxybenzamide/pyrrole pair in eight ring hairpin polyamides (44). The

specificity of thiazole placed internally with a crosslinked polyamide remains to be determined.

Differences in the two types of polyamides, hairpin versus crosslinked, complicate direct comparison of the thiazole ring in the two linkage motifs. In the hairpin motif, there is effective binding in the absence of a head group. In contrast, in the crosslinked motif the head group attached to the leading ring may significantly modulate its binding characteristics, with formyl enhancing adenine selectivity and acetyl or hydrogen inhibiting binding for crosslinked polyamides with a leading thiazole ring. Additionally, the number of constituent rings in the dimer, six in this case and eight in the hairpin, may significantly affect the affinity of binding. We suggest that *N*-formyl-thiazole may provide sequence discrimination in crosslinked polyamides, with thiazole favoring adenine over thymine. Moreover, we have demonstrated the important influence of the leading head group in combination with the length of methylene linker on the binding characteristics of crosslinked polyamides.

We thank Dr. Aiden Flynn for assistance with the numerical analysis. This work was supported by National Institutes of Health Grant GM-31299 (to R.E.D.) and Cancer Research Campaign Program Grant SP2000/0402 (to J.A.H.).

- Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1376–1380.
- Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B. & Dabrowiak, J. C. (1986) *Biochemistry* **25**, 7408–7416.
- Krowicki, K. & Lown, J. W. (1987) *J. Org. Chem.* **52**, 3493.
- Pelton, J. G. & Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5723–5727.
- Wade, W. S., Mrksich, M. & Dervan, P. B. (1992) *J. Am. Chem. Soc.* **114**, 8783–8794.
- Dwyer, T. J., Geierstanger, B. H., Bathini, Y., Lown, J. W. & Wemmer, D. E. (1992) *J. Am. Chem. Soc.* **114**, 5911–5919.
- Chen, Y. H. & Lown, J. W. (1994) *J. Am. Chem. Soc.* **116**, 6995–7005.
- Mrksich, M., Parks, M. E. & Dervan, P. B. (1994) *J. Am. Chem. Soc.* **116**, 7983–7988.
- Parks, M. E., Baird, E. E. & Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118**, 6147–6152.
- Baliga, R., Baird, E. E., Herman, D. M., Melander, C., Dervan, P. B. & Crothers, D. M. (2001) *Biochemistry* **40**, 3–8.
- Mrksich, M. & Dervan, P. B. (1993) *J. Am. Chem. Soc.* **115**, 9892–9899.
- Mrksich, M. & Dervan, P. B. (1994) *J. Am. Chem. Soc.* **116**, 3663–3664.
- Chen, Y. H., Yang, Y. & Lown, J. W. (1996) *J. Biomol. Struct. Dyn.* **14**, 341–355.
- Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996) *Nature (London)* **382**, 559–561.
- Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. (1997) *Nature (London)* **387**, 202–205.
- Mapp, A. K., Ansari, A. Z., Ptashne, M. & Dervan, P. B. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 3930–3935.
- Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M. & Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12890–12895.
- Dickinson, L. A., Trauger, J. W., Baird, E. E., Ghazal, P., Dervan, P. B. & Gottesfeld, J. M. (1999) *Biochemistry* **38**, 10801–10807.
- Dickinson, L. A., Trauger, J. W., Baird, E. E., Dervan, P. B., Graves, B. J. & Gottesfeld, J. M. (1999) *J. Biol. Chem.* **274**, 12765–12773.
- Chiang, S. Y., Burli, R. W., Benz, C. C., Gawron, L., Scott, G. K., Dervan, P. B. & Beerman, T. A. (2000) *J. Biol. Chem.* **275**, 24246–24254.
- Mrksich, M. & Dervan, P. B. (1993) *J. Am. Chem. Soc.* **115**, 2572–2576.
- Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994) *Science* **266**, 646–650.
- Mrksich, M. & Dervan, P. B. (1995) *J. Am. Chem. Soc.* **117**, 3325–3332.
- White, S., Szcwzyk, J. M., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998) *Nature (London)* **391**, Suppl., 468–471.
- Kielkopf, C. L., White, S., Szcwzyk, J. M., Turner, J. M., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998) *Science* **282**, 111–115.
- Kielkopf, C. L., Bremer, R. E., White, S., Szcwzyk, J. M., Turner, J. M., Baird, E. E., Dervan, P. B. & Rees, D. C. (2000) *J. Mol. Biol.* **295**, 557–567.
- Rao, K. E., Shea, R. G., Yadagiri, B. & Lown, J. W. (1990) *Anticancer Drug Design* **5**, 3–20.
- Kopka, M. L., Han, G. W., Goodsell, D. S., Chiu, T. K., Walker, W. L., Lown, J. W. & Dickerson, R. E. (1998) *Structure, Motion, Interaction and Expression of Biological Macromolecules, Proc. 10th Conversation* (Adenine Press, New York), pp. 177–191.
- Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W. & Dickerson, R. E. (1997) *Structure (Cambridge, UK)* **5**, 1033–1046.
- Mrksich, M., Wade, W. S., Dwyer, T. W., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7586–7590.
- Dwyer, T. J., Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1993) *J. Am. Chem. Soc.* **115**, 9900–9906.
- Chen, Y. H. & Lown, J. W. (1995) *Heterocycles* **8**, 1691–1707.
- Sharma, S. K., Tandon, M. & Lown, J. W. (2000) *J. Org. Chem.* **65**, 1102–1107.
- Dervan, P. B. & Burli, R. W. (1999) *Curr. Opin. Chem. Biol.* **3**, 688–693.
- Parks, M. E., Baird, E. E. & Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118**, 6153–6159.
- Brenowitz, M., Sencar, D. F., Shea, M. A. & Ackers, G. K. (1986) *Methods Enzymol.* **130**, 132–181.
- Brenowitz, M., Sencar, D. F., Shea, M. A. & Ackers, G. K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8462–8466.
- Sencar, D. F., Brenowitz, M., Shea, M. A. & Ackers, G. K. (1986) *Biochemistry* **25**, 7344–7354.
- Zhan, Z.-Y. J. & Dervan, P. B. (2000) *Bioorg. Med. Chem.* **8**, 2467–2474.
- Palfner, K., Kneba, M., Hiddemann, W. & Bertram, J. (1995) *BioTechniques* **19**, 926–929.
- Hawkins, C. A., Pelaez de Clairac, R., Dominey, R. N., Baird, E. E., White, S., Dervan, P. B. & Wemmer, D. E. (2000) *J. Am. Chem. Soc.* **122**, 5235–5236.
- Burckhardt, G., Fortsch, I., Simon, H., Birch-Hirschfeld, E., Kittle, L., Schutz, H., Sharma, S. K., Lown, J. W. & Zimmer, C. (2000) *J. Biomol. Struct. Dyn.* **11**, 355–363.
- Nguyen, D. H., Szcwzyk, J. W., Baird, E. E. & Dervan, P. B. (2001) *Bioorg. Med. Chem.* **9**, 7–17.
- Ellervik, U., Wang, C. C. C. & Dervan, P. B. (2000) *J. Am. Chem. Soc.* **122**, 9354–9360.
- Walker, W. L., Kopka, M. L. & Goodsell, D. S. (1997) *Biopolymers* **44**, 323–334.

# Energetic basis for selective recognition of T·G mismatched base pairs in DNA by imidazole-rich polyamides

Ellyn R. Lacy<sup>1</sup>, Binh Nguyen<sup>1</sup>, Minh Le<sup>1,2</sup>, Kari K. Cox<sup>2</sup>, Caroline O'Hare<sup>3</sup>, John A. Hartley<sup>3</sup>, Moses Lee<sup>2</sup> and W. David Wilson<sup>1,\*</sup>

<sup>1</sup>Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA, <sup>2</sup>Department of Chemistry, Furman University, Greenville, SC 29613, USA and <sup>3</sup>Department of Oncology, Royal Free and University College Medical School, University College London, London W1W 7BS, UK

Received January 8, 2004; Revised and Accepted March 9, 2004

## ABSTRACT

To complement available structure and binding results and to develop a detailed understanding of the basis for selective molecular recognition of T·G mismatches in DNA by imidazole containing polyamides, a full thermodynamic profile for formation of the T·G–polyamide complex has been determined. The amide-linked heterocycles f-ImImIm and f-PyImIm (where f is formamido group, Im is imidazole and Py is pyrrole) were studied by using biosensor-surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) with a T·G mismatch containing DNA hairpin duplex and a similar DNA with only Watson–Crick base pairs. Large negative binding enthalpies for all of the polyamide–DNA complexes indicate that the interactions are enthalpically driven. SPR results show slower complex formation and stronger binding of f-ImImIm to the T·G than to the match site. The thermodynamic analysis indicates that the enhanced binding to the T·G site is the result of better entropic contributions. Negative heat capacity changes for the complex are correlated with calculated solvent accessible surface area changes and indicate hydrophobic contributions to complex formation. DNase I footprinting analysis in a long DNA sequence provided supporting evidence that f-ImImIm binds selectively to T·G mismatch sites.

## INTRODUCTION

Polyamide DNA binding agents are of interest due to their sequence-specific binding characteristics (1–3), their ability to recognize mismatched base pairs, and their potential for use in modulation of gene expression (4). Through the combination of pyrrole (Py), imidazole (Im), and other heterocycles in a

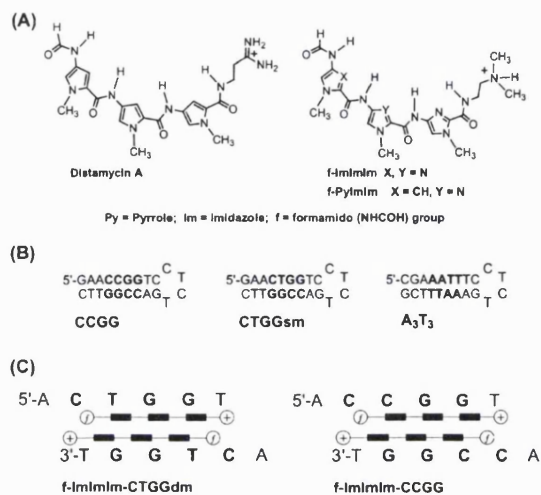
side-by-side pairing, polyamides capable of specifically recognizing Watson–Crick DNA sequences have been synthesized (1–3). In addition, structural, binding affinity and kinetics studies have shown the potential use of the side-by-side pairing of Im heterocycles for recognition of DNA sequences containing non-Watson–Crick base pairs (5,6). The amide-linked triimidazole, f-ImImIm where f is formamido (Fig. 1A), can discriminate T·G from Watson–Crick C·G base pairs as well as from other mismatches (5,6).

Possible uses of the Im/Im pair include rapid detection of single nucleotide polymorphisms (5,6). Rapid and accurate detection of T·G mismatch base pairs in DNA would be of significant benefit since this mismatch is responsible for most of the common mutations leading to formation of tumors in humans. In human bladder carcinoma, for example, a G·C to A·T transition at the 3'G of the GG doublet in codon 12 of the Ha-ras and Ki-ras proto-oncogene converts them to oncogenes (7–9). A T·G mismatch is often introduced by spontaneous deamination of 5-methylcytosine and can arise from errors in replication (10). Structural analysis of DNAs containing T·G mismatches, indicates that the mismatches adopt a wobble conformation and structural perturbations are mainly in the vicinity of the mismatch (11,12). NMR studies show that f-ImImIm forms an anti-parallel dimer in the DNA minor groove with a T·G mismatch site (5,6,13,14). In the mismatch complex the guanine-NH<sub>2</sub> group of the wobble T·G base pair forms two specific hydrogen bonds to the side-by-side Im/Im pair (5). Biosensor-surface plasmon resonance (SPR) analysis has shown that f-ImImIm is capable of strongly and specifically binding to a DNA sequence with only a single T·G mismatch (6,13). This novel T·G mismatch recognition motif has recently been corroborated by results from the Dervan laboratory, in which a hairpin polyamide containing a stacked Im/Im pair was found to bind at a T·G mismatch site (14).

There is a wealth of information on structure, binding affinity and specificity for the interactions of polyamides with DNA (1–3), however, the specific energetic factors that govern the binding events have rarely been characterized.

\*To whom correspondence should be addressed. Tel: +1 404 651 3903; Fax: +1 404 651 1416; Email: chewdw@panther.gsu.edu  
Correspondence may also be addressed to Moses Lee. Tel: +1 864 294 3368; Fax: +1 864 294 3559; Email: moses.lee@furman.edu





**Figure 1.** (A) Chemical structures of the polyamides: distamycin A, f-ImImIm and f-PyImIm. (B) Sequences of the DNA hairpins used in this study: CCGG and A<sub>3</sub>T<sub>3</sub> are Watson-Crick sequences while CTGGsm contains one T-G mismatch. (C) Schematic models of the complexes formed between f-ImImIm and the DNA sequences containing a double T-G mismatch (left) and CCGG (right) as inferred from NMR data.

Important calorimetric results have been reported for a very limited number of polyamides in complex with their target DNA sequences (15–17). Thermodynamic studies can provide valuable complementary information to the structural perspective and can improve our understanding of the molecular forces that govern the affinity and specificity of binding. Specifically, they can give additional insights into the energetic basis for the discrimination of the T-G mismatch from matched sites by f-ImImIm. We report here biosensor-SPR and calorimetric studies as a function of temperature as well as footprinting studies in a large sequence context for the interaction of f-ImImIm and f-PyImIm (Fig. 1A) with mismatched and matched sequence DNAs. By conducting the studies as a function of temperature, we were able to determine the heat capacity changes on binding that are critical for a complete thermodynamic understanding of biomolecular interactions (18–20).

## MATERIALS AND METHODS

### Reagents and biochemicals

**Compounds and buffers.** Distamycin A was purchased from Sigma Chemical Co. and used without further purification. *N*-[2-(Dimethylamino)ethyl]-1-methyl-4-[1-methyl-4-[4-formamido-1-methylimidazole-2-carboxamide]imidazole-2-carboxamide]imidazole-2-carboxamide (f-ImImIm) and *N*-[2-(dimethylamino)ethyl]-1-methyl-4-[1-methyl-4-[4-formamido-1-methylpyrrole-2-carboxamide]imidazole-2-carboxamide]imidazole-2-carboxamide (f-PyImIm) were prepared as described previously (6,21). Solutions of the compounds for calorimetric studies were prepared in MES20 buffer, 0.01 M MES [2-(*N*-morpholino)ethanesulfonic acid] adjusted to pH 6.25, 0.001 M EDTA and 0.2 M NaCl. Concentrations of the

compounds were determined spectrophotometrically by absorbance at 304 nm for distamycin and 303 nm for f-ImImIm and f-PyImIm. The extinction coefficients used were 34 000 M<sup>-1</sup> cm<sup>-1</sup> for distamycin (17) and 30 700 M<sup>-1</sup> cm<sup>-1</sup> for f-ImImIm and f-PyImIm at 25°C.

**DNA sequences.** DNAs were obtained as anionic exchange, HPLC purified products (Midland Certified Reagent Co). Biosensor-SPR binding studies were conducted with 5'-biotinated DNAs, and calorimetric studies were performed with non-biotinated DNAs. Solutions for the ITC experiments were prepared in degassed MES20 buffer. To ensure that the DNA was in the hairpin form, the solutions were heated to ~90°C and cooled rapidly in ice (22). The concentration of the DNA solutions was determined spectrophotometrically at 260 nm and 85°C using extinction coefficients per nucleotide of 9264, 9059 and 9400 M<sup>-1</sup> cm<sup>-1</sup> for CCGG, CTGGsm and A<sub>3</sub>T<sub>3</sub>, respectively. The extinction coefficients were calculated on a per strand basis by the nearest-neighbor method and divided by the number of nucleotides per strand (23–25). The extinction coefficients extrapolated to 25°C were 7824, 8107 and 7978 M<sup>-1</sup> cm<sup>-1</sup> for CCGG, CTGGsm and A<sub>3</sub>T<sub>3</sub>, respectively.

### Methods

**DNase I footprinting of TG mismatched oligonucleotides.** Five micrograms of an 80 base DNA 5'-AGG TGA GCA GGT CCA TAC TGG TTT GCA CCT CGA GGT TAC CCG TAT CTG CTC CAG CTC AAC TGG TAA CCT GCA CCT GGT CG-3' (MWG-Biotech AG), were 5'-end-labeled with [ $\gamma$ -<sup>32</sup>P]ATP using standard protocols. The labeled oligonucleotide was heated at 95°C for 3 min and annealed to 5  $\mu$ g of a complementary strand at 65°C for 1 min then cooled slowly to 4°C. The annealed product, containing a core match site (5'-CCGG-3') and a single T-G mismatch site (5'-CTGG-3') that are underlined above, was separated on a 3% agarose gel, excised and purified with a Zymoclean gel/DNA recovery kit (Anachem) using standard procedures. The polyamide f-ImImIm was incubated with 0.5  $\mu$ g of the double-stranded DNA in 10 mM Tris pH 7.0, 1 mM EDTA, 50 mM KCl, 1 mM MgCl<sub>2</sub>, 0.5 mM DTT and 20 mM HEPES, at room temperature for 30 min, in a total volume of 50  $\mu$ l. DNA cleavage was initiated by the addition of the polyamide treated sample to 2  $\mu$ l (1 U) DNase I diluted in cold 10 mM Tris pH 7.0 from a stock solution (1 U/ $\mu$ l, Promega) and 1  $\mu$ l of a solution of 250 mM MgCl<sub>2</sub>, 250 mM CaCl<sub>2</sub>. The reactions were performed at room temperature and stopped after 3 min by the addition of 100  $\mu$ l of a stop mix containing 200 mM NaCl, 30 mM EDTA pH 8 and 1% SDS. The cleavage products were phenol/chloroform extracted and ethanol precipitated in the presence of 1  $\mu$ l of glycogen (20 mg/ml; Roche Diagnostics), washed once in 80% ethanol and lyophilized. The samples were resuspended in formamide loading dye, denatured for 5 min at 90°C, cooled on ice and electrophoresed at 2000 V for 2 h on a 10% denaturing acrylamide gel (Sequagel, National Diagnostics). The gels were dried under vacuum at 80°C and exposed to film for 24 h (X-OMAT, Kodak).

**Isothermal titration calorimetry and biosensor-SPR.** Calorimetric titrations were performed with a VP-ITC (Microcal, Inc., Northampton, MA) or a CSC 4200 ITC

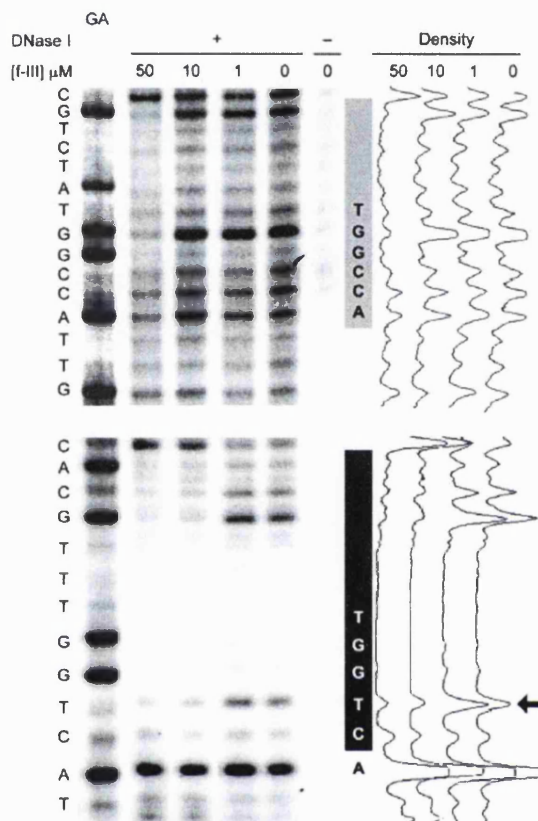
(Calorimetry Science Corp., Spanish Fork, UT). Software provided with the calorimeters is used for control and data collection. Electrical calibration of the instruments was performed by applying heat pulses using built-in modules. Chemical calibrations were performed by using: HCl/THAM, RNase A/2'CMP, and BaCl<sub>2</sub>/18-Crown-6 (Aldrich Chem. Co., Milwaukee, WI). The enthalpy values obtained in the calibrations differed by <2% from literature values (26–28).

ITC experiments were conducted by injecting 6–15  $\mu$ l (depending on the compounds concentrations and heat generated) of the polyamides every 300 s for a total of 16–34 injections into DNA solution. The concentrations were 0.0585, 0.2228 and 0.200 mM for f-ImImIm, f-PyImIm and distamycin, respectively. DNA hairpin concentrations were from 0.0035 to 0.010 mM. Similar experiments were performed to determine the heats of dilution of the polyamides with the DNA solution substituted by buffer. An experimental upper limit of 35°C was selected based on the  $T_m$  of the DNA, which was determined to be over 40°C. The heat produced for each injection of compound into DNA or buffer was obtained by integration of the area under each peak of the titration plots with respect to time. The heats of reaction were obtained by subtraction of the integrated heats of dilution of the compounds from the heats corresponding to the injection of compound into DNA. The heat of titration of the ligand into buffer varies slightly over the ratio range. Because of the variation, point-to-point rather than averaged subtraction was applied to ITC titration data. Data corresponding to the first injection were discarded. The binding enthalpy ( $\Delta H^\circ$ ) for each titration was obtained by fitting the results of heat per mole as a function of total molar ratio (polyamide/DNA).

SPR experiments were conducted using a BIACORE 2000 with an SA chip. The DNA hairpins were purchased with 5'-biotinated. Immobilizations, experimental details and data processing procedures have been previously described (6,21).

**Calculation of solvent accessible surface area.** Changes in solvent accessible surface area (SASA) upon complex formation were calculated using the GRASP program (29) with a probe radius of 1.4. Computations were performed for all atom NMR structures for f-ImImIm bound to the 5'-GAAGTGGTTC-3' duplex (CTGGdm) containing a core having two adjacent T·G mismatches (5). Carbon, carbon-bound hydrogen and phosphorus were defined as nonpolar atoms, while all other atoms were defined as polar. The change in SASA ( $\Delta$ SASA) on complex formation was calculated as the difference between the SASA of the complex and the sum of the SASAs of the free DNA duplex and the free compound in the bound conformation ( $\Delta$ SASA = SASA<sub>complex</sub> – [SASA<sub>DNA</sub> + SASA<sub>f-ImImIm</sub>]). Total contributions of the polar and nonpolar atoms to the surface areas were also obtained to calculate the predicted heat capacity changes ( $\Delta C_p^\circ$ ). The  $\Delta C_p^\circ$  values were calculated using the equation derived from protein folding (30) and recently modified for small molecule–DNA complexes (31):

$$\Delta C_p^\circ = (0.382 \cdot \Delta A_{np} - 0.121 \cdot \Delta A_p) \text{ cal/mol K}$$



**Figure 2.** DNase I footprinting results for the f-ImImIm complex with match (top) and single T·G mismatch (bottom) DNA sequences. Concentrations of f-ImImIm increase from 1 to 50  $\mu$ M and the f-ImImIm gel lanes are compared to a G/A sequencing ladder and to uncleaved DNA (–). The protected regions are denoted by gray and black rectangles, respectively. Density plots (far right) illustrate a decrease in band intensity upon increasing concentrations of f-ImImIm with protection of DNA cleavage at the single T·G mismatch indicated with a bold arrow.

## RESULTS

### Selective recognition of T·G mismatches by f-ImImIm

Although binding of f-ImImIm to single and double mismatch sites has been demonstrated in short DNA oligomers, selective binding has not been confirmed in the context of a long duplex sequence. For direct comparison of the binding of f-ImImIm to match and mismatch sites, DNase I footprinting was used with an 80 base DNA containing Watson–Crick CCGG and single T·G mismatch (5'-CTGG-3') sites (Fig. 2). Footprints were observed at both the T·G and match sites but footprints at the mismatch site appear at concentrations of 10  $\mu$ M while binding at the match site required more than 5-fold higher concentrations. No binding was observed at other sites in the DNA duplex under the footprinting conditions. At the highest concentration of 50  $\mu$ M, the footprints are larger than expected for the polyamide size. This has been observed with other



**Table 1.** Thermodynamic parameters for the interaction of polyamides with matched and mismatched DNA oligomers

Compound	DNA	$\Delta C_p^\circ$	15°C $\Delta H^\circ/T\Delta S^\circ$ $\Delta G^\circ$	20°C $\Delta H^\circ/T\Delta S^\circ$ $\Delta G^\circ$	25°C $\Delta H^\circ/T\Delta S^\circ$ $\Delta G^\circ$	35°C $\Delta H^\circ/T\Delta S^\circ$ $\Delta G^\circ$
f-ImImIm	CTGGsm	$-258 \pm 34$	-11.1/-2.1	-12.2/-3.3	-12.5/-3.3	-15.7/-6.9
	CCGG		-9.0	-8.9	-9.3, -9.2* -12.8/-5.5 -7.3, -7.3*	-8.8
f-PylImIm	CCGG	$-146 \pm 40$	-8.2	-12.9/-4.7 -8.1	-14.2/-6.1 -8.1	-15.2/-7.3 -7.9

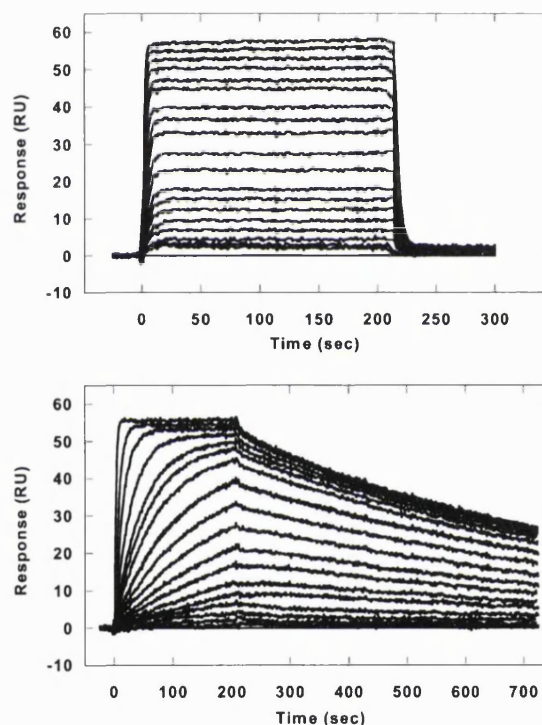
Values of  $\Delta H^\circ$  are obtained from calorimetry;  $\Delta G^\circ$  are from SPR or calorimetry (\*);  $T\Delta S^\circ$  are calculated from the relationship  $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$ ; the units for  $\Delta H^\circ$ ,  $\Delta G^\circ$  and  $T\Delta S^\circ$  are kcal/mol and for  $\Delta C_p^\circ$  are cal/mol K.

polyamides and probably arises from nonspecific interactions at the high concentrations.

#### Binding constants ( $K$ ), cooperativity and stoichiometry

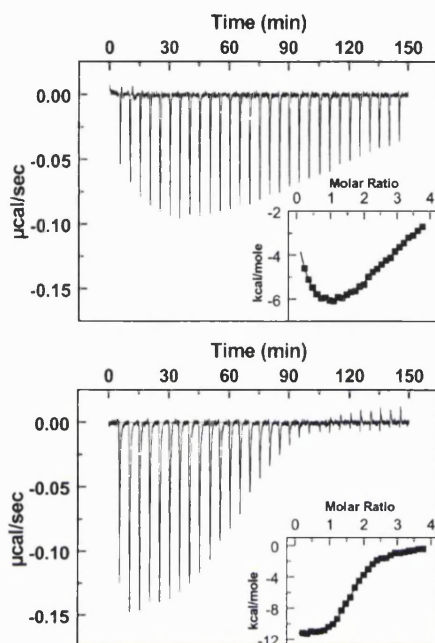
Biosensor-SPR sensorgrams for the interaction of f-ImImIm with DNA hairpin duplexes containing CCGG and CTGGsm binding sites (Fig. 1) are in Figure 3. The maximum instrument response ( $RU_{\max}$ ) obtained in the steady-state region corresponds to approximately twice the predicted response for binding of one molecule of the compound and indicates a 2:1 stoichiometry (2 mol of compound/mol of DNA hairpin) in agreement with previous studies (5,6). As can be seen in the figure, the kinetics for binding of the polyamide differ markedly for the different DNA sequences: f-ImImIm dissociates much more slowly from the mismatch complex (Fig. 3). Binding constants can be determined from fitting the steady state region of the sensorgrams, as previously described (6,21), or from fitting kinetics in the association and dissociation regions. As expected, all results were best fit with a model that has a 2:1 polyamide to DNA stoichiometry. For binding of f-ImImIm to both matched and mismatched sequences,  $K_1$  is significantly less than  $K_2$  indicating positive cooperativity in binding of the two molecules of f-ImImIm. Because of correlation between  $K_1$  and  $K_2$ , the error in fitting individual values of  $K_1$  and  $K_2$  is larger than that of the average ( $K = [K_1 \cdot K_2]^{1/2}$ ), for dimer complex formation. For comparison of the binding of polyamides to DNA, the average  $K$  values determined from the steady-state region are converted to binding free energies and reported in Table 1. Reporting the square root values places all results on a per-bound molecule basis and allows direct comparison between binding of monomers and dimers as well as with literature results. The rate constants obtained from fitting the kinetics curves allow calculation of  $K$  and with the polyamides these values agree with those from the steady-state analysis. Biosensor-SPR methods were also used to determine  $K$  as a function of temperature and the constants changed very little with temperature (Supplementary Material, Table S1). In summary, the presence of a T-G base pair in a DNA sequence has a pronounced influence on the binding of f-ImImIm and yields a much higher affinity than for a similar perfectly matched sequence.

**Calorimetric analysis of binding.** Large exothermic enthalpies are observed in ITC titrations of f-ImImIm into CCGG and CTGGsm (examples in Fig. 4). The titration heats for the



**Figure 3.** Typical SPR sensorgrams for the interaction of f-ImImIm with the CCGG hairpin (top) and CTGGsm hairpin (bottom) at 25°C. The maximum responses, with the same amount of DNA immobilized on each surface of the biosensor, are the same in both cases at high concentration indicating the same stoichiometry. Remarkable differences in binding kinetics are observed: fast on/off kinetics for f-ImImIm-CCGG and slower kinetics for the f-ImImIm-CTGGsm complex. The compound concentrations are from 0.75 to 26  $\mu$ M with CCGG and 0.01 to 7  $\mu$ M with CTGGsm.

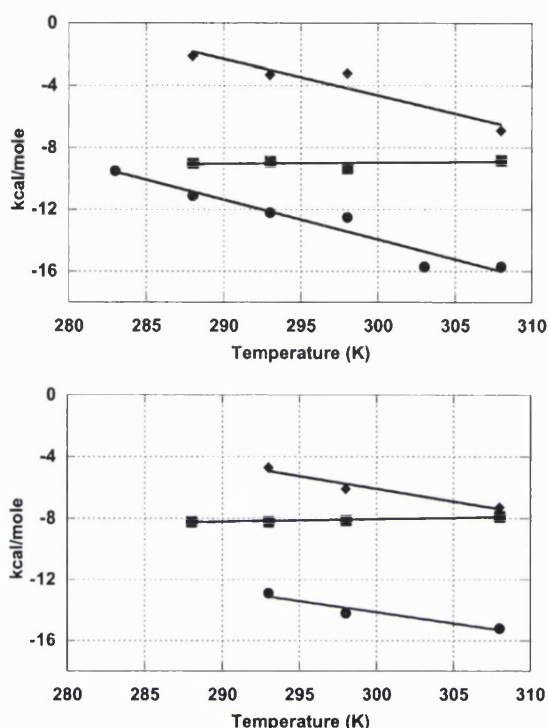
interaction of f-ImImIm with the CCGG hairpin are smaller than for CTGGsm and are spread over a broad ratio range as expected for weaker binding (6). The titration heat was converted to heat per mole as a function of total molar ratio, and point-to-point subtraction of dilution heat was applied. The data were fitted with a 2:1 model to give molar binding enthalpies ( $\Delta H^\circ$ ). Due to the large correlation among the  $\Delta H^\circ$



**Figure 4.** Baseline corrected ITC results, before heat of dilution subtraction, for titration of f-ImImIm into the CCGG hairpin in MES 20 buffer 25°C (top). Every peak represents the heat released from a 10 µl injection of 0.0585 mM f-ImImIm into 3.5 µM of DNA hairpin. The inset is the corrected molar heat/addition, after subtraction heat of dilution, with the best-fit curve for a 2:1 model. Similar titrations with the CTGGsm hairpin are shown (bottom). The relative small heat of dilution of f-ImImIm can be seen in the last two to three injections when all of DNA sites are saturated. It is interesting that distinct differences in the heat profiles are observed but that the averaged total binding heats are comparable (Table 1).

values for binding of the two f-ImImIm molecules in a cooperative complex, the average binding enthalpy, rather than the individual enthalpies ( $\Delta H_1^\circ$  and  $\Delta H_2^\circ$ ), is reported in Table 1. Titrations of f-PyImIm into the CCGG hairpin (Fig. 1) provide an excellent model for comparison of strong matched site binding with the mismatched complex of f-ImImIm. The enthalpic contributions are highly favorable for binding of both polyamides to their DNA recognition sequences. The  $\Delta H^\circ$  values along with  $\Delta G^\circ$  values from SPR allow calculation of  $\Delta S^\circ$  for binding (Table 1). The  $\Delta S^\circ$  values are negative, an unfavorable term in the  $\Delta G^\circ$  of binding. Although the average enthalpic contribution for the formation of f-ImImIm/CTGGsm and f-PyImIm/CCGG complexes is comparable, the entropic contribution is less favorable for f-PyImIm binding to CCGG at 298 K. The enthalpy is strongly temperature dependent and decreases (becomes more exothermic) as the temperature is increased (Fig. 5, top). Similar results were obtained for the interaction of f-PyImIm with its cognate CCGG sequence (Fig. 5, bottom). The experimental heat capacity changes for formation of the complexes were obtained from linear fitting of  $\Delta H^\circ$  as a function of temperature (Table 1) and all  $\Delta C_p^\circ$  values are negative.

The interactions of polyamides vary significantly with DNA sequence as shown in previous studies of distamycin binding



**Figure 5.** The dependence of the thermodynamic parameters on temperature for binding of f-ImImIm to CTGGsm (top) and f-PyImIm to CCGG (bottom). The entropy  $T\Delta S^\circ$  (diamonds) was calculated from the experimental  $\Delta G^\circ$  (squares) obtained from SPR, and  $\Delta H^\circ$  (filled circles) obtained from ITC. The slopes of  $\Delta H^\circ$  as a function of temperature provide the  $\Delta C_p^\circ$  values for complex formation (Table 1).

to AT-rich sequences (2,6,17,21). Unlike the imidazole containing polyamides, formation of the distamycin 2:1 complex with an  $A_3T_3$  sequence is governed by negative cooperativity with strong affinity in the first interaction ( $K_1 = 3.1 \times 10^8 \text{ M}^{-1}$ ,  $\Delta G_1^\circ = -11.6 \text{ kcal/mol}$ ) followed by weaker binding of a second molecule ( $K_2 = 8.0 \times 10^5 \text{ M}^{-1}$ ,  $\Delta G_2^\circ = -8.1 \text{ kcal/mol}$ ) (6). ITC titrations of distamycin were conducted as a model system for comparison of positive and negative cooperativity in dimer interactions (Supplementary Material, Fig. S1). Binding of the two molecules of distamycin to DNA is clearly distinguished in the plot, which has a quite different shape than with the Im polyamide titrations. Binding of the first distamycin has a lower enthalpy ( $\Delta H_1^\circ = -12.2 \text{ kcal/mol}$ ) than that of the second ( $\Delta H_2^\circ = -23.9 \text{ kcal/mol}$ ). The entropic contributions of  $-0.6$  and  $-15.9 \text{ kcal/mol}$  for distamycin binding to the two sites indicate that the entropic contribution is near zero for binding of the first molecule but is very unfavorable for the second. These results indicate that, as with the imidazole polyamides, distamycin binding is enthalpy driven.

**Calculated heat capacity changes.** Changes in SASA on complex formation have been correlated with  $\Delta C_p^\circ$  values for protein-protein, protein-DNA (32,33) and more recently for

drug–DNA interactions (34,35). The  $\Delta C_p^\circ$  values for the f-ImImIm–CTGGdm complex were calculated based on buried surface areas determined by using the reported NMR structure (5) with previously described methods (31,36). Although the structure for the complex with the CTGGsm sequence is not available, the values obtained with the CTGGdm sequence could be extrapolated to the complex formed with the single mismatch. The calculated heat capacity changes using two different sets of parameters are negative and in general agreement with the experimental values (Supplementary Material, Table S2).

## DISCUSSION

Previous biosensor binding studies have shown that f-ImImIm and other polyamides bind to short oligonucleotides (matched or mismatched sequences) with significantly different affinities (6,21). Fox and coworkers have also recently shown that some, but not all, chemical and enzymatic cleavage agents can recognize T·G (and other) mismatched base pairs in a long DNA sequence (37). The first question that we wished to answer in this study was whether f-ImImIm could selectively recognize a single T·G mismatch site in competition with excess Watson–Crick sites. Footprints from DNase I cleavage of an 80 base DNA sequence with a single T·G mismatch (Fig. 2) show strong protection from cleavage by f-ImImIm at the T·G site. These results are in agreement with the oligonucleotide studies by SPR that indicate stronger binding of f-ImImIm to the mismatched CTGGsm over the matched CCGG site.

A full thermodynamic data set is essential to complement the structural information available for T·G recognition by f-ImImIm and to provide a complete understanding the energetic basis for the T·G recognition. The values of  $\Delta H^\circ$  and  $\Delta C_p^\circ$  for binding reactions are most reliably obtained by calorimetry rather than indirectly from van't Hoff determinations (38–41). Calorimetry, on the other hand, may provide a less accurate  $K$  for strong binding reactions due to the relatively high concentrations that may have to be used if reaction heats are low. We have found biosensor-SPR methods to be particularly valuable in studies of DNA oligomer complexes with small molecules such as the polyamides (6,21) since a wide range of concentrations can be used and both kinetic and equilibrium results are obtained. The SPR studies are particularly useful when the  $K$  values are high, as with the f-ImImIm/T·G complex.

Biosensor-SPR methods clearly show that the polyamides form cooperative 2:1 complexes with their DNA recognition sites. The  $\Delta G^\circ$  values for binding are essentially constant with temperature suggesting strong enthalpy–entropy compensation in this temperature range (Fig. 5). Such compensation is observed for many biomolecular processes and suggests a significant hydrophobic component to the binding energetics (31,42–44). The SPR results quantitatively show that f-ImImIm binds much more strongly to the mismatched CTGGsm sequence than to the equivalent DNA with Watson–Crick base pairs. The  $\Delta G^\circ$  value for binding to T·G sites is  $\sim 2$  kcal/mol more favorable than for binding to equivalent CG sites. The improved binding of f-ImImIm to the T·G mismatch is correlated with much slower dissociation kinetics relative to the matched CCGG complex (6,21). The

slow dissociation kinetics for the mismatch complex could arise from at least two factors: the breakage of strong H-bonds (from the amino protons of G to f-ImImIm) and the rearrangement/re-uptake of water molecules in the DNA minor groove as described below.

Calorimetric studies are also consistent with formation of cooperative 2:1 complexes of f-ImImIm with both matched CG and mismatched T·G DNA sequences and with stronger affinity for the mismatch. The difference in the heat profiles for f-ImImIm binding to CTGGsm and CCGG (Fig. 4) arises both from differences in individual enthalpies ( $\Delta H_1^\circ$  and  $\Delta H_2^\circ$ ) and binding constants ( $K_1$  and  $K_2$ ) in the 2:1 complexes. Somewhat surprisingly given all of these differences, the calorimetric titrations suggest that the averaged total heats for dimer formation are comparable for formation of f-ImImIm–CTGGsm, f-ImImIm–CCGG, and f-PyImIm–CCGG complexes at 298 K. This is particularly surprising since comparison of the  $\Delta H^\circ$  and  $\Delta G^\circ$  values for binding (Table 1) indicates that the enthalpic contributions play the primary role in driving the interactions. Given that the calorimetric  $\Delta H^\circ$  values for binding are more negative than the total free energies of binding, there must be unfavorable compensating entropic contributions in each binding process. Since the  $\Delta H^\circ$  values for these interactions are similar at 298 K, it is the difference in the unfavorable entropic contribution that dictates the different binding affinities. The higher binding affinity for the f-ImImIm–CTGGsm complex means the  $\Delta S^\circ$  for formation of this complex must be less unfavorable than that of the complex with CCGG.

It is important at this point to combine results from the structural and thermodynamic studies to suggest possible explanations the similar  $\Delta H^\circ$  but quite different  $\Delta G^\circ$  and  $\Delta S^\circ$  for binding of f-ImImIm to match and mismatch sites. NMR and CD results suggest that significant conformational changes do not occur in the DNA binding sites on formation of these complexes (5,6). These observations imply that the entropic effects do not come from overall structural changes but from factors such as the loss of translational and rotational entropy of the polyamides and the opposing gain of entropy due to release of bound water molecules on complex formation (although the net change in bound water is unknown). Since the polyamides would be expected to be similarly hydrated and to release similar amounts of hydration water on complex formation, any differences in water released, among the complexes, would have to be due to differences in DNA hydration. Structural studies indicate that the T·G base pair adopts a wobble conformation with an unpaired free amino group of G that is strongly hydrated as part of the minor groove hydration matrix (12,45). On complex formation with f-ImImIm some of these tightly bound and highly ordered water molecules would be released from the T·G site to provide a favorable entropic contribution to complex formation. From this model the entropic contributions are predicted to be less unfavorable for formation of the f-ImImIm–CTGGsm complex than for the CCGG complex and this results in the enhanced affinity for the T·G complex. The difference in  $\Delta G^\circ$  between the f-ImImIm–CTGGsm and –CCGG complexes ( $\sim 2$  kcal/mol and mainly from entropy) is about the same order of magnitude as the entropy gain by releasing a single-bound water molecule from its site (46). Binding of a netropsin analog to a mixed sequence DNA

suggested that water is actually taken up upon complex formation (47), and it is difficult to see how this could enhance binding entropy. Initial molecular dynamics studies in our laboratory have suggested that minor groove hydration is quite varied among different AT DNAs, and it is possible that water release or uptake is quite dependent on different DNA sequences and ligands.

The negative  $\Delta C_p^\circ$  values for complex formation can provide useful insights about interactions and hydration (30–36). There are significant differences in the heat capacity changes for the polyamides and DNA sequences used in these experiments. The  $\Delta C_p^\circ$  for the f-ImImIm–CTGGsm complex, for example, is more negative than that for f-PyImIm–CCGG (Table 1). These  $\Delta C_p^\circ$  values indicate important differences in the formation of complexes with the CCGG DNA versus the T-G mismatch CTGGsm sequence. Structured water, such as the water of hydrophobic hydration, can have a large heat capacity and release of such water, for example on transfer of nonpolar groups into the interior of proteins or to the minor groove of DNA, contributes a negative term to the binding  $\Delta C_p^\circ$ . Since all of the polyamide complexes examined in this study have negative  $\Delta C_p^\circ$ , the binding enthalpy becomes more favorable (more negative) while the binding entropy is less favorable (also more negative) as the temperature increases. Such effects are larger in systems where bound and ordered water molecules are more strongly associated with the DNA groove as with CTGGsm. Calculation of buried surface areas from the NMR structures of f-ImImIm–CTGGdm and f-ImImIm–CCGG indicate that the observed  $\Delta C_p^\circ$  are correlated with water release from buried nonpolar and polar molecular surface.

Binding of polyamides to GC containing sequences has quite different characteristics than binding to AT-rich sequences. Binding of many imidazole-polyamides to sequences with GC or T-G mismatched base pairs displays positive cooperativity, while distamycin binds to  $A_3T_3$  with negative cooperativity (6). This difference may be explained in terms of minor groove width. The groove of AT is unusually narrow (0.3–0.4 nm) (48), but it is wider in mixed DNA sequences (0.5–0.6 nm) (49) as well as at T-G base pairs (12). Since the thickness of one polyamide ring is ~0.34 nm (50), the first distamycin molecule can fit snugly and make favorable van der Waals contacts with the walls of the minor groove in  $A_3T_3$ . For binding of the second molecule, however, the minor groove must widen sufficiently to accommodate the two stacked distamycin molecules in a 2:1 complex. On the other hand, the groove width of GC-rich or T-G sequences is wider and can bind two polyamides in a positively cooperative manner even at low ratios of compound to DNA. Interestingly, the experimental  $K$  and  $\Delta H^\circ$  for distamycin binding are clearly resolved (17,51 and this work) in ITC while those for polyamide binding with positive cooperativity are overlapped and difficult to separate. In summary, the results presented here provide a thermodynamic rationale for the specificity in molecular recognition of T-G mismatch base pairs by Im-polyamides.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

## ACKNOWLEDGEMENTS

This work was supported by NIH Grant GM61587, the Georgia Research Alliance (W.D.W.), and by the NSF-REU and Research Corporation (M.L.).

## REFERENCES

- Bailly, C. and Chaires, J.B. (1998) Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. *Bioconjug. Chem.*, **9**, 513–538.
- Wemmer, D.E. (2000) Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 439–461.
- Wemmer, D.E. (2001) Ligands recognizing the minor groove of DNA: Development and applications. *Biopolymers*, **52**, 197–211.
- Dervan, P.B. (2001) Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.*, **9**, 2215–2235.
- Yang, X.L., Hubbard, R.B., Lee, M., Tao, Z.F., Sugiyama, H. and Wang, A.H. (1999) Imidazole-imidazole pair as a minor groove recognition motif for T:G mismatched base pairs. *Nucleic Acids Res.*, **27**, 4183–4190.
- Lacy, E.R., Cox, K.K., Wilson, W.D. and Lee, M. (2002) Recognition of T\*G mismatched base pairs in DNA by stacked imidazole-containing polyamides: surface plasmon resonance and circular dichroism studies. *Nucleic Acids Res.*, **30**, 1834–1841.
- Land, H., Parada, L.F. and Weinberg, R.A. (1983) Cellular oncogenes and multistep carcinogenesis. *Science*, **222**, 771–778.
- Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N. and Perucho, M. (1988) Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell*, **53**, 549–554.
- Watanabe, H., Ha, A., Hu, Y.X., Ohtsubo, K., Yamaguchi, Y., Motoo, Y., Okai, T., Toya, D., Tanaka, N. and Sawabu, N. (1999) K-ras mutations in duodenal aspirate without secretin stimulation for screening of pancreatic and biliary tract carcinoma. *Cancer*, **86**, 1441–1448.
- Lall, L. and Davidson, R.L. (1998) Sequence-directed base mispairing in human oncogenes. *Mol. Cell. Biol.*, **18**, 4659–4669.
- Hare, D., Shapiro, L. and Patel, D.J. (1986) Wobble dG X dT pairing in right-handed DNA: solution conformation of the d(C-G-T-G-A-A-T-T-C-G-C-G) duplex deduced from distance geometry analysis of nuclear Overhauser effect spectra. *Biochemistry*, **25**, 7445–7456.
- Hunter, W.N., Brown, T., Kneale, G., Anand, N.N., Rabinovich, D. and Kennard, O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.*, **262**, 9962–9970.
- Lacy, E.R., Madsen, E.M., Lee, M. and Wilson, W.D. (2003) Polyamide dimer stacking in the DNA minor groove and recognition of T:G mismatched base pairs in DNA. In Demeunynck, M., Bailly, C. and Wilson, W.D. (eds), *DNA and RNA Binders: From Small Molecules to Drugs*. WILEY-VCH, Weinheim, Vol. 2, pp. 384–413.
- Rucker, V.C., Foister, S., Melander, C. and Dervan, P.B. (2003) Sequence specific fluorescence detection of double strand DNA. *J. Am. Chem. Soc.*, **125**, 1195–1202.
- Pilch, D.S., Poklar, N., Gelfand, C.A., Law, S.M., Breslauer, K.J., Baird, E.E. and Dervan, P.B. (1996) Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. *Proc. Natl Acad. Sci. USA*, **93**, 8306–8311.
- Pilch, D.S., Poklar, N., Baird, E.E., Dervan, P.B. and Breslauer, K.J. (1999) The thermodynamics of polyamide-DNA recognition: hairpin polyamide binding in the minor groove of duplex DNA. *Biochemistry*, **38**, 2143–2151.
- Rentzeperis, D., Marky, L.A., Dwyer, T.J., Geierstanger, B.H., Pelton, J.G. and Wemmer, D.E. (1995) Interaction of minor groove ligands to an AAATT/AATTT site: correlation of thermodynamic characterization and solution structure. *Biochemistry*, **34**, 2937–2945.
- Cooper, A., Johnson, C.M., Lakey, J.H. and Nollmann, M. (2001) Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.*, **93**, 215–230.
- Morton, C.J. and Ladbury, J.E. (1996) Water-mediated protein-DNA interactions: the relationship of thermodynamics to structural detail. *Protein Sci.*, **5**, 2115–2118.

20. Privalov, G.P. and Privalov, P.L. (2000) Problems and prospects in microcalorimetry of biological macromolecules. *Methods Enzymol.*, **323**, 31–62.
21. Lacy, E.R., Le, N.M., Price, C.A., Lee, M. and Wilson, W.D. (2002) Influence of a terminal formamido group on the sequence recognition of DNA by polyamides. *J. Am. Chem. Soc.*, **124**, 2153–2163.
22. Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A. and van Boom, J.H. (1988) Oligodeoxynucleotide folding in solution: loop size and stability of B- hairpins. *Biochemistry*, **27**, 6321–6326.
23. Cantor, C.R. and Tinoco, I., Jr (1965) Absorption and optical rotatory dispersion of seven trinucleoside diphosphates. *J. Mol. Biol.*, **13**, 65–77.
24. Cantor, C.R., Warshaw, M.M. and Shapiro, H. (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
25. Fasman, G.D. (1975) *Handbook of Biochemistry and Molecular Biology, Nucleic Acids*, 3rd Edn. CRC Press, Cleveland, OH.
26. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, **179**, 131–137.
27. Horn, J.R., Russell, D., Lewis, E.A. and Murphy, K.P. (2001) Van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: are there significant discrepancies? *Biochemistry*, **40**, 1774–1778.
28. Briggner, L.E. and Wadso, I. (1991) Test and calibration processes for microcalorimeters, with special reference to heat conduction instruments used with aqueous systems. *J. Biochem. Biophys. Methods*, **22**, 101–118.
29. Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
30. Spolar, R.S., Livingstone, J.R. and Record, M.T., Jr (1992) Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry*, **31**, 3947–3955.
31. Ren, J., Jenkins, T.C. and Chaires, J.B. (2000) Energetics of DNA intercalation reactions. *Biochemistry*, **39**, 8439–8447.
32. Makhatadze, G.I. and Privalov, P.L. (1990) Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. *J. Mol. Biol.*, **213**, 375–384.
33. Privalov, P.L. and Makhatadze, G.I. (1990) Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects. *J. Mol. Biol.*, **213**, 385–391.
34. Mazur, S., Tanious, F.A., Ding, D., Kumar, A., Boykin, D.W., Simpson, I.J., Neidle, S. and Wilson, W.D. (2000) A thermodynamic and structural analysis of DNA minor-groove complex formation. *J. Mol. Biol.*, **300**, 321–337.
35. Haq, I., Ladbury, J.E., Chowdhry, B.Z., Jenkins, T.C. and Chaires, J.B. (1997) Specific binding of hoechst 33258 to the d(CGCAAATTTGCG)<sub>2</sub> duplex: calorimetric and spectroscopic studies. *J. Mol. Biol.*, **271**, 244–257.
36. Spolar, R.S. and Record, M.T., Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
37. Brown, J., Brown, T. and Fox, K.R. (2003) Cleavage of fragments containing DNA mismatches by enzymic and chemical probes. *Biochem. J.*, **371**, 697–708.
38. Janin, J. (1995) Elusive affinities. *Proteins*, **21**, 30–39.
39. Ladbury, J.E. (1995) Counting the calories to stay in the groove. *Structure*, **3**, 635–639.
40. Naghibi, H., Tamura, A. and Sturtevant, J.M. (1995) Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proc. Natl Acad. Sci. USA*, **92**, 5597–5599.
41. Chaires, J.B. (1997) Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. *Biophys. Chem.*, **64**, 15–23.
42. Chaires, J.B. (1997) Energetics of Drug-DNA Interactions. *Biopolymers*, **44**, 201–215.
43. Gallagher, K. and Sharp, K. (1998) Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.*, **75**, 769–776.
44. Matulis, D., Rouzina, I. and Bloomfield, V.A. (2002) Thermodynamics of cationic lipid binding to DNA and DNA condensation: roles of electrostatics and hydrophobicity. *J. Am. Chem. Soc.*, **124**, 7331–7342.
45. Yang, X.L., Kaenzig, C., Lee, M. and Wang, A.H. (1999) Binding of AR-1–144, a tri-imidazole DNA minor groove binder, to CCGG sequence analyzed by NMR spectroscopy. *Eur. J. Biochem.*, **263**, 646–655.
46. Dunitz, J.D. (1994) The entropic cost of bound water in crystals and biomolecules. *Science*, **264**, 670.
47. Sidorova, N. and Rau, D.C. (1995) The osmotic sensitivity of netropsin analogue binding to DNA. *Biopolymers*, **35**, 377–384.
48. Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R.E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.*, **257**, 14686–14707.
49. Yoon, C., Prive, G.G., Goodsell, D.S. and Dickerson, R.E. (1988) Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc. Natl Acad. Sci. USA*, **85**, 6332–6336.
50. Pelton, J.G. and Wemmer, D.E. (1990) Binding modes of distamycin A with d(CGCAAATTTGCG)<sub>2</sub> determined by two-dimensional NMR. *J. Am. Chem. Soc.*, **112**, 1393–1399.
51. Wilson, W.D., Tanious, F.A., Barton, H.J., Wydra, R.L., Jones, R.L., Boykin, D.W. and Strekowski, L. (1990) The interaction of unfused polyaromatic heterocycles with DNA: intercalation, groove-binding and bleomycin amplification. *Anticancer Drug Des.*, **5**, 31–42.