University of Southampton

Faculty of Medicine, Health and Life Sciences

School of Biological Sciences

Sequence-Specific Recognition of DNA from the Minor Groove

by

Andrew James Hampshire

Thesis for the degree of Doctor of Philosophy

June 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF BIOLOGICAL SCIENCES

Doctor of Philosophy

SEQUENCE-SPECIFIC RECOGNITION OF DNA FROM THE MINOR GROOVE

by Andrew James Hampshire

The human genome has long been a target for therapy for many diseases, including cancers. Intercalating agents and minor groove binding ligands such as polyamides have been developed with the intention of selectively targeting DNA with high binding affinity and sequence specificity. However, *in vivo* selectivity coupled with high affinity has proved hard to achieve, and many current chemotherapy drugs binding strongly but with low specificity.

Footprinting and fluorescence melting techniques were utilised in this thesis to study a range of novel DNA-binding ligands to increase the knowledge of this class of compounds. So as to assess the binding selectivity of DNA binding ligands, four minimal length novel footprinting sequences were designed, containing every di-, tri- and symmetrical hexa-nucleotide step (the latter split between two fragments). Initial studies on classic DNA-binding ligands, such as distamycin, showed the substrates to be useful for footprinting and the results elaborated on their precise binding preferences.

The novel universal DNA fragments were then combined with ligand-specific fragments to study the binding of novel analogues of the bis-intercalator TANDEM to the dinucleotide TpA, especially when flanked by AT-base pairs; binding was attenuated or abolished by removing the disulphide cross-bridge or replacing the quinoxaline groups with naphthylene rings. However, the binding affinity was increased by 30-40-fold by substituting the two valine residues in the octadepsipeptide ring for lysines.

Polyamides are a group of DNA minor groove ligands with established binding rules for selectivity recognising all four base pair combinations through the use of pyrrole, imidazole and hydroxypyrrole rings. However, the most selective hairpin polyamides are too large to be useful drugs and are generally limited to targeting a maximum of ten base pairs of DNA. The sequence selectivity of a range of polyamide ligands containing novel ring moieties was examined with a view to improving their selectivity and affinity. It was found that C-terminal pyridoimidazole or benzimidazole ring systems increased the affinity for A or T compared to a pyrrole group. The position of pyridoimidazoles and benzimidazoles in 2:1 dimers is also important, with pyridoimidazole creating a staggered dimer. An isopropyl-thiazole ring system was shown to target guanine selectively, while also creating a staggered 2:1 complex. The introduction of an imidazole to the polyamide in place of a pyrrole decreased the binding affinity significantly, as well as altered the sequence selectivity.

Contents

Chapter II: Materials and Methods

Chapter III: Novel Universal Footprinting Substrates

Chapter IV: Novel Analogues of the Bis-Intercalator TANDEM

Chapter V: Polyamide and Hoechst-Derived Minor Groove Binding Ligands

Chapter VI: Minor Groove Binding Ligands Containing Isopropyl-Thiazole Units

Figures, Equations and Table Contents

Acknowledgements

The research in this thesis was supported by a research studentship from the Biotechnology and Biological Sciences Research Council (BBSRC) and a CASE studentship from Eurogentec UK Ltd.

The minor groove ligands described in Chapter 5 were kindly provided by Dr Malvinder Singh (University of Saskatchewan); TANDEM and it analogues (Chapter 4) were provided by Dr Mark Searcey (School of Pharmacy, University of London – now School of Chemical Sciences and Pharmacy, University of East Anglia); and the Thiazotropsin ligands (Chapter 6) were provided by Prof. Colin Suckling and Prof. Roger Waigh (University of Strathclyde).

Most oligonucleotides were kindly provided by Prof. Tom Brown (University of Southampton, School of Chemistry).

I would like to thank my supervisor Keith Fox for his support and advice throughout my doctorate studies, with particular reference to his assistance during the writing of this thesis.

Thanks must also be made to the members of the Fox Lab with whom I rubbed shoulders on a daily basis. Foremost are: Dave Rusling, who was a great help in his mentoring role; Phil Rachwal, for the sports updates; Vicki Tear, for her sanity and tolerance in a male-dominated lab; Steve Harrison, my bench mate and hence empire-crusher; and Dave Norris, who was always a useful friend to have around! My appreciation is also owed to Peter James, Kerensa Jones, Loïc LeStrat, Nick Brown and Matt Smith, whose combined knowledge furthered mine as I was starting in the lab.

Finally, thank you to my family and friends who carried me on their shoulders when things were tough. You will never appreciate how grateful I am for your belief in me. Many thanks.

Abbreviations

2YT	Yeast tryptone media
bp(s)	Base pair(s)
Bzi	Benzimidazole
DMSO	Dimethyl sulphoxide
DNase I	Deoxyribonuclease I enzyme
Dp	Dimethylaminopropylamide
DTT	Dithiothreritol
EDTA	Ethylenediaminetetraacetic acid
HIV	Human Immunodeficiency Virus
Hp	N-methylhydroxypyrrole (Hydroxypyrrole)
I	Inosine
Im	N-methylimidazole (Imidazole)
IPTG	Isopropyl-β-D-thiogalactopyranoside
Mp	Methylpiperazine
Py	N-methylpyrrole (Pyrrole)
PCR	Polymerase Chain Reaction
Pzi	Pyridoimidazole
R	Purine
rpm	Revolutions per minute
S	Cytosine or Guanine base
ssDNA	Single-strand-DNA
T_m	Melting temperature
TANDEM	des-*N*-tetramethyl-triostin A
TBE	Tris-Borate-EDTA buffer
TE buffer	Tris-EDTA buffer
TEMED	N,N,N',N'-Tetramethylethylenediamine
Th	Isopropyl-thiazole
Tris	Tris(hydroxymethyl)aminomethane
W	Adenine or Thymine
X-gal	5-bromo-4-chloro-3-indoylyl-β-Galactosidase
Y	Pyridimidine

SI units are used throughout, as are standard three letter abbreviations for amino acids.

Chapter I
Introduction

Why Target DNA?

The central dogma of molecular biology states that the flow of genetic information travels from DNA to RNA and finally to the translation into proteins (Crick, 1970). Every biological process uses this cascade for gene expression, with cancers, viruses and microbial pathogens using it to elicit damaging effects in humans. It is therefore necessary to target the cascade selectively to block these unwanted processes.

Every molecule of DNA creates many RNA molecules, which in turn create scores of protein molecules. It is rational to target the smallest number of molecules at the start of the cascade to elicit an effect. Therefore DNA is the obvious target to affect molecular processes and alter gene expression. When the secondary structure of DNA was first described (Watson and Crick, 1953), at a time when the structure of most drug targets was not known, it presented one of the first macromolecular targets for drug design. It is now clear that many diseases, including cancers, viruses (including HIV) and genetic diseases such as cystic fibrosis might be treated with ligands that target unique DNA sequences to alter gene expression.

Selective Targeting of DNA

Deoxyribonucleic acid (DNA) is a polymer of sugar-phosphate bound bases: Thymine (T) and cytosine (C) are pyrimidines; adenine (A) and guanine (G) are purines. These bases, each attached to a deoxyribose sugar, are connected in chains by phosphodiester linkages. Nuclear double stranded DNA is composed of two antiparallel strands, which contain hydrogen bonds between complementary bases in the two strands, creating AT and GC base pairs. Three hydrogen bonds connect the bases in a GC pairs, with two in an AT pair. The opposing strands of DNA create a double helical structure, with the two strands twisted around each other in such a manner as to create major and minor grooves in the normal B-form of the helix.

The major and minor grooves of DNA generate possible drug target sites, with their unique patterns of hydrogen bonds. Many proteins interact with DNA via one or other of these grooves and thereby affect transcription, replication and many other cellular processes (Nordhoff and Lehrach, 2007). Figure 1.1 shows the possible hydrogen bonding sites on the exposed edges of the base pairs in the minor and major grooves. Interaction

with these provides a means of achieving sequence selectivity, though it should be remembered that electrostatic forces, van der Waal interactions and hydrophobic contacts also play important roles in determining binding affinity and selectivity (Wemmer and Dervan, 1997).

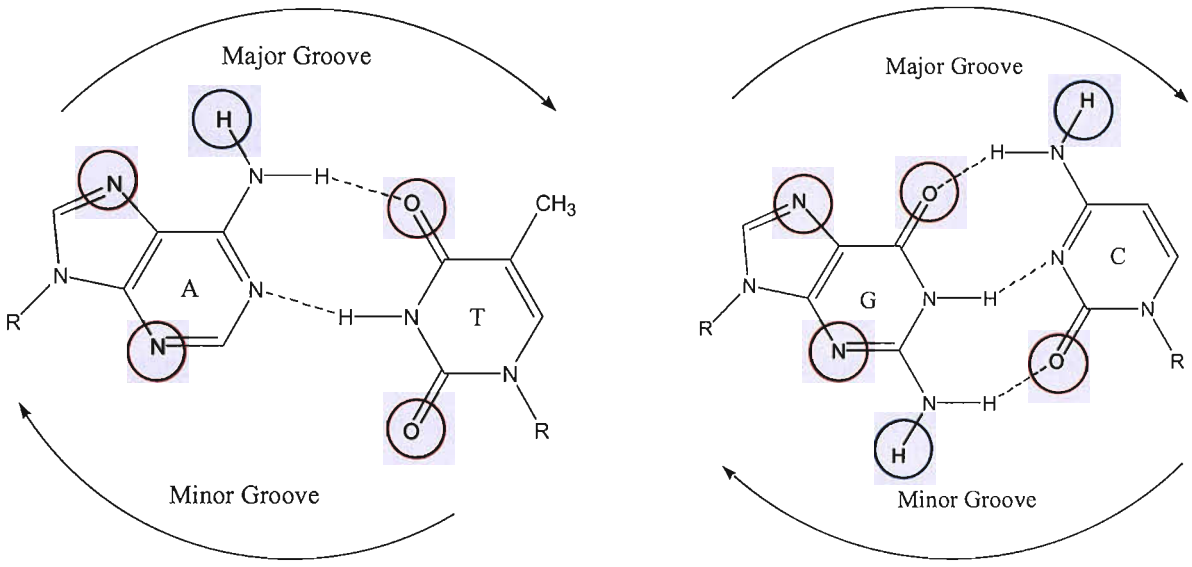


Figure 1.1: Possible hydrogen bonding sites on the DNA base pairs. Red circles indicate hydrogen bond acceptors, blue circles indicate hydrogen bond donors. The methyl group of thymine may also hinder interactions with the major groove or provide a hydrophobic contact point.

DNA Minor Groove Binding Ligands

Although the minor groove contains less information than the major groove, it provides an ideal location for small molecules to bind to DNA. Its narrow structure allows for close contact with ligands, which can then form specific interactions with the exposed faces of the base pairs.

Several minor groove binding ligands have been widely studied, the most important of which are the polyamides and their derivatives. Intercalating ligands may also make specific interactions with the minor groove, whilst threading intercalating agents such as nogalamycin place functional groups in both the major and minor grooves.

Polyamides

Ligands such as the polyamides have been developed to selectively target the minor groove and to make contacts with the exposed edges of the base pairs (Neidle, 2001). Most interactions with the bases involve the N3 on purines and O2 on pyrimidines. The exocyclic amino group of guanine also forms many water-mediated contacts in protein-DNA complexes, whilst the O4' of deoxyribose is involved in a significant number of hydrophobic interactions.

Many of the hydrophobic contacts have been shown to be secondary, accompanying energetically more favourable hydrogen bonding to N3 or O2 (Morávek *et al.*, 2002). However, for ligand binding to occur, existing hydrogen bonds between water and the ligand and DNA need to be broken to accommodate the ligand. The DNA minor groove has a spine of water running its length (Soler-López *et al.*, 1999) and this needs to be removed to allow ligand binding. This is energetically expensive, though entropically favourable, and plays a major role in ligand-DNA interactions (Marky and Breslauer, 1987; Chalikian *et al.*, 1994; Rentzeperis *et al.*, 1995; Guerri *et al.*, 1998).

Many small molecules that bind to specific sequences within the DNA minor groove have been discovered or synthesised. The first compounds had several common features: They were comprised of aromatic rings; had an overall curvature that matched that of the shape of the floor of the minor groove; had hydrogen bond donors and acceptors on the inside edge; had one or more positive charges; and they selectively bound to AT base pairs (Wemmer, 2000).

The first minor groove ligands to be studied were the natural products netropsin and distamycin (Figure 1.2). Early binding studies demonstrated that these ligands bound selectively to A/T-rich DNA, in a mechanism that did not involve intercalation (Zimmer *et al.*, 1971; Wartell *et al.*, 1974). NMR and crystallographic studies have since clearly defined their mode of binding within the minor groove (Patel, 1982; Coll *et al.*, 1987).

These ligands contain pyrrole rings linked by amide bonds with either a formyl or guanidine group at the N-terminus and a propylamidine at the C-terminus (Cozzi and Mongelli, 1998). Both of these ligands bind to A/T-rich sequences of four or more consecutive AT base pairs. The pyrrole rings cannot form contacts with the 2-amino group of guanine, which therefore sterically hinders the interaction with GC base pairs (Van Dyke *et al.*, 1982; Kopka *et al.*, 1985; Lown *et al.*, 1986).

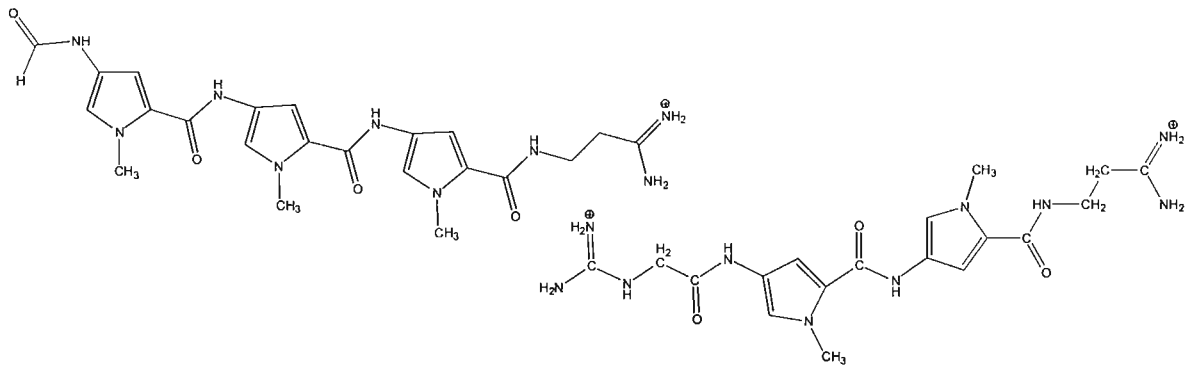


Figure 1.2: Structures of distamycin (left) and netropsin (right).

Crystallographic and NMR studies have shown that these ligands bind snugly in the minor groove in a 1:1 fashion (a single ligand molecule binding at a single DNA binding site) at low concentrations (Patel, 1982; Coll *et al.*, 1987; Gonzalez *et al.*, 2001), taking advantage of the relatively deep, narrow minor groove at the A/T-rich sequences, as shown in Figure 1.3.

The crescent shape of the ligands matches the curvature of the minor groove and the NH-groups of the amides point into the groove where they make hydrogen bonds to acceptors on the base pairs. The positively charged tails are positioned deep in the groove where there is the highest electrostatic potential. From all this information, it is apparent that van der Waals forces, electrostatic interactions and hydrogen bonding contribute to the binding energy (Guerri *et al.*, 1998; Morávek *et al.*, 2002).

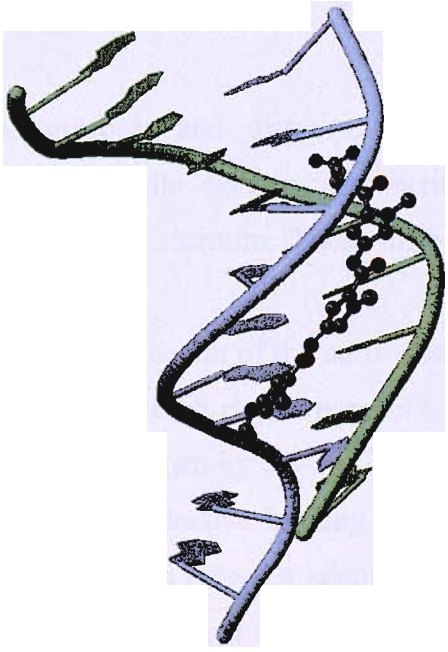


Figure 1.3: DNA double helix d(GGCCAATTGG) with a distamycin molecule bound in a 1:1 mode (taken from the RCSB Protein Data Bank Online [PDB ID – 1JTL]).

Footprinting and other studies have shown that distamycin and netropsin bind better to the sequences AAAA and AATT than to TTAA and TATA (Zimmer *et al.*, 1979; Abu-Daya *et al.*, 1995; Abu-Daya and Fox, 1997; Pelton and Wemmer, 1989) and this selectivity is thought to arise from sequence dependent differences in the shape of the minor groove. A_n sequences tend to have the highest affinity as $A_n.T_n$ tracts possess a narrow minor groove, with which the ligands make close contacts (Yoon *et al.*, 1988; Wemmer and Dervan, 1997). $A_n.T_n$ tracts also have a relatively rigid helix that is ideal for minor groove ligand binding. This occurs because the AT base pairs can adopt a high propeller twist, generating bifurcated hydrogen bonds between adjacent base pairs (Nelson *et al.*, 1987). An ApT step base stacks the thymine methyl group with the adjacent adenine and the sugar phosphate backbone in such a way as to again prevent excessive bending of the helix.

In contrast, TpA steps disrupt the structure of the DNA groove and produce a wider minor groove (Yoon *et al.*, 1988; Wemmer and Dervan, 1997). This is because the thymine methyl groups project into the major groove without any significant stacking interactions with either the adjacent adenine or sugar phosphate backbone, making this step more flexible than $A_n.T_n$ or ApT, widening the minor groove and preventing close ligand-DNA contacts (Suzuki *et al.*, 1996; Allemann and Egli, 1997). In the case of distamycin, TpA steps at the 5'-end or the centre of the binding site have the greatest effect on the affinity (Gonzalez *et al.*, 2001). This may be because the ligand binds with its uncharged N-terminus facing the 5'-end of any stretch of adenine residues, where the ligand sits shallower in the groove (Kopka *et al.*, 1985; White *et al.*, 1997b).

Structural differences between netropsin and distamycin affect binding selectivity: Netropsin contains two pyrrole rings while distamycin contains three; and they have different charged groups on their N and C-termini. These differences subtly affect how these two ligands bind.

Netropsin sits symmetrically in the centre of the narrow minor groove with the two pyrrole rings slightly non-coplanar so that each ring is parallel to the walls of the groove. This causes the minor groove to be forced open by 0.5-2.0 Å and the helix is bent back by 8 ° across the region of attachment. Its selective binding is not due to hydrogen bond formation, but close van der Waals contacts between adenine C2 hydrogens and the CH groups on the pyrrole rings of the ligand (Kopka *et al.*, 1985; Goodwin *et al.*, 2005).

Distamycin binds in a similar manner to netropsin, but has a larger binding site size and requires longer A/T-tracts. It also interacts with the base pairs immediately flanking

the high affinity AATT site (Abu-Daya and Fox, 1997). Whilst both netropsin and distamycin can both bind with low affinity to A_n tracts of three base pairs, netropsin shows strong binding to tracts of four AT base pairs. In contrast, distamycin will only bind well to an A_n tract of five base pairs or longer, and the sequence AAATT is an especially good binding site (Abu-Daya and Fox, 1997; Baraldi *et al.*, 2000). This extra selectivity made distamycin a more appealing research subject.

Distamycin has shown promise as an anti-tumour drug, selectively inhibiting binding of two important transcription factors in muscle-specific gene expression that are involved in cancer, SRF and MEF2, to their A/T-rich elements (Taylor *et al.*, 1997). Distamycin also possesses antiviral and antiprotozoal activity (Ginsburg *et al.*, 1993; Cozzi and Mongelli, 1998; Orfeo *et al.*, 1999); and has also been used as a vehicle for targeting alkylating agents to DNA in the form of tallimustine and brostallicin, thereby increasing their cytotoxicity (Baraldi *et al.*, 2000; Fedier *et al.*, 2003).

Over the years numerous synthetic curved aromatic ligands have been prepared in attempts to enhance the binding affinity or alter the sequence specificity.

Probably the most widely used synthetic A/T-specific DNA minor groove binding ligand is Hoechst 33258 (Figure 1.4). This compound is used as a cytological DNA staining agent due to an intense increase in its fluorescence upon binding to DNA and it is often used for microscope imaging (Latt and Wohlleb, 1975).

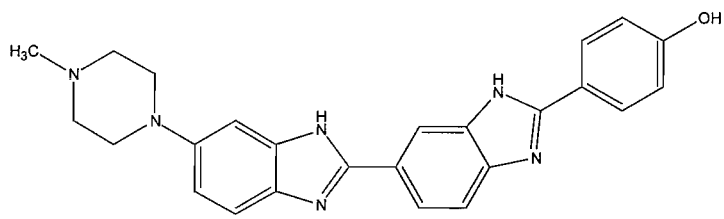


Figure 1.4: Structure of Hoechst 33258.

Hoechst 33258 binds in the minor groove to sequences of 4-5 AT base pairs, with the two benzimidazole NHs forming a bridge on the floor of the minor groove between adjacent adenine N3 and thymine O2 atoms on opposite DNA strands (Harshman and Dervan, 1985; Vega *et al.*, 1994; Aymami *et al.*, 1999; Gavathiotis *et al.*, 2000; Bostock-Smith *et al.*, 2001). The flanking phenol and N-methylpiperazine rings interact with at least one additional base pair on each site of the binding site (Harshman and Dervan, 1985) and it

has been suggested that the N-methylpiperazine may form a water-mediated interaction to guanines adjacent to the A/T-tract (Gavathiotis *et al.*, 2000). Close van der Waals contacts with the non-polar backbone are also important for strong binding.

Hoechst 33258 has a very low affinity for two consecutive AT base pairs, which increases significantly for three and reaching a plateau for four AT base pairs (preferentially binding to AATT (Breusegem *et al.*, 2002)). TpA steps are particularly weakly bound by Hoechst 33258 compared to the other minor groove binding ligands (Abu-Daya *et al.*, 1995); although NMR studies have shown that the ligand prefers a flexible DNA structure (Bostock-Smith *et al.*, 2001). The presence of GC base pairs inhibits binding due to hindrance from the 2-amino group of guanine (Drobyshev *et al.*, 1999), though some weaker binding may be permitted at these sequences (Bailly *et al.*, 1993).

Hoechst 33258 has been used as a therapeutic agent for solid tumours (Bielawski *et al.*, 2001), disrupting topoisomerases I and II and resulting in cytotoxic DNA strand cleavage. The ligand has also been reported to protect against radiation-induced DNA-strand breakage, but its high mutagenicity and cytotoxic effects limit its use as a protector of normal tissue during radiotherapy (Tawar *et al.*, 2003).

These minor groove binding ligands (together with others such as berenil, DAPI and pentamidine) are all A/T specific. They also display only limited sequence selectivity as they recognise fairly short (4-5) base pair targets.

2:1 Binding

During NMR experiments on the binding of distamycin to A/T-containing sequences, a new type of complex was discovered when the site contained more than four AT base pairs. Two distamycin molecules are bound side-by-side at the sequence AAATT (Pelton and Wemmer, 1989). In this complex the two ligand molecules are arranged antiparallel, with each partner in contact with one wall of the minor groove, making contacts to one of the DNA strands, as shown in Figure 1.5. The groove binding was similar to that seen with 1:1 binding, with the polyamides positioned such that each ring makes contact with a single base (Walker *et al.*, 1997).

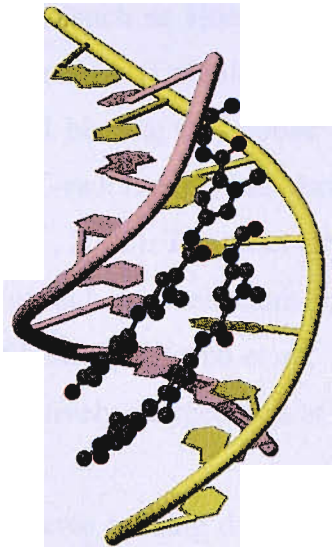


Figure 1.5: Distamycin bound to d(GTATATAC) DNA in a 2:1 mode (taken from the RCSB Protein Data Bank Online [PDB ID – 378D]).

A major difference between the 1:1 and 2:1 binding of distamycin is that the minor groove is substantially wider (by 3.5-4 Å) (Wemmer and Dervan, 1997), so that it is more like the size of a mixed B-DNA sequence than the narrow groove that is typically observed in A/T-rich regions.

In titrations of an AAATT-containing sequence with distamycin, it was found that 1:1 complexes formed at low ligand concentrations; these were replaced by the 2:1 complex as ligand concentration increased. It was estimated that the binding constant for the first ligand was about ten times greater than that of the second (Rentzeperis *et al.*, 1995). This indicates that the DNA minor groove is quite flexible, changing width to accommodate a second ligand once the first is bound. It therefore follows that there should be a stronger preference for the 2:1 complex in sequences with an intrinsically wide groove, while the converse should occur with sequences that adopt a very narrow minor groove. This has been demonstrated with AAAAA, for which the 2:1 binding mode is disfavoured relative to 1:1, while the 2:1 binding mode is favoured for the sequence TATAT (since the TpA steps increase the minor groove width) (Fagan and Wemmer, 1992). By substituting IC base pairs for AT, leaving the minor groove functional groups unchanged but increasing the width of the minor grove, the 2:1 binding mode is further enhanced (Wemmer and Dervan, 1997), indicating that DNA groove shape and flexibility strongly affect the binding behaviour.

Other minor groove binding ligands such as Hoechst 33258, DAPI and berenil can also form 2:1 complexes at high concentrations (Loontiens *et al.*, 1990; Gavathiotis *et al.*, 2000; Rosu *et al.*, 2002). However, this 2:1 binding may not be through minor groove binding, as intercalation of these drugs into G/C-rich regions has also been proposed (Loontiens *et al.*, 1990; Bailly *et al.*, 1993; Rosu *et al.*, 2002). The observation that Hoechst 33258 functions as a clamp for two DNA chains at high concentrations, causing unwinding of the duplex, also indicates a different mode of binding (Utsuno *et al.*, 1999). Netropsin does not engage in 2:1 binding, since it contains positively charged tails at both ends.

Research into new DNA minor groove binding drugs therefore shifted to concentrate on 2:1 binding ligands. This was due to the realisation that this provides increased sequence specificity and increased binding affinity than the conventional 1:1 binding. In addition, when two molecules of drug bind in the DNA minor groove, the groove is fully filled, so that processes such as replication and transcription are inhibited more efficiently.

G/C Selectivity: Introducing imidazoles

The major limitation with netropsin and distamycin-based ligands is that they are A/T selective in both the 1:1 and 2:1 binding modes. This is because pyrrole rings cannot form contacts with the 2-amino group of guanine, which therefore sterically hinders the interaction with GC base pairs. Dickerson and Lown suggested that by replacing the pyrrole (Py) rings with imidazoles (Im) (exchanging a carbon for nitrogen), the exocyclic NH_2 group of guanine might be read in the 1:1 binding mode (Kopka *et al.*, 1985; Lown *et al.*, 1986).

Peter Dervan's group at the California Institute of Technology synthesised the polyamide ImPyPy and similar ligands, which were expected to bind in the 1:1 mode to the sequence $(G/C)(A/T)_2$ (Wade and Dervan, 1987; Wade *et al.*, 1992). However, the molecules did not bind as expected, but recognised the five base pair sequence 5'-(A/T)G(A/T)C(A/T) (Wade and Dervan, 1987; Wade *et al.*, 1992). In addition, affinity cleavage experiments challenged the 1:1 binding mode (Wade *et al.*, 1992). NMR studies confirmed that the ligand bound cooperatively to this sequence in a 2:1 binding mode (Mrksich *et al.*, 1992; Buchmueller *et al.*, 2006).

Inspection of crystallographic and NMR data indicated that only A/T-rich DNA sequences contain the narrow grooves that are necessary for strong 1:1 binding of netropsin and distamycin (Coll *et al.*, 1987; Fagan and Wemmer, 1992). Since the introduction of G/C-residues produces a wider minor groove, it follows that in the 1:1 mode ligands will not be able to make good interactions with both walls of the groove.

Wemmer's group used distamycin analogues (using the ligand PyImPy), since distamycin (unlike netropsin) can bind in a 2:1 mode. NMR studies showed that this molecule bound cooperatively to AAGTT in preference to AAATT (Dwyer *et al.*, 1992). In this complex, one imidazole N3 formed a hydrogen bond with the 2-amino group of guanine, though the other imidazole, placed adjacent to the cytosine, did not make any hydrogen bond contacts with this base.

Using these simple principles it is possible to design sequence-specific minor groove binding ligands, which interact in the 2:1 mode, with an imidazole for recognizing guanine paired with pyrrole against the cytosine-containing strand. In this way, Im/Py targets GC, Py/Im recognises CG and Py/Py targets AT or TA (White *et al.*, 1997a), as shown in Figure 1.6. The C-terminal tails of the molecules are specific for AT base pairs (White *et al.*, 1997b). Interestingly, Im/Im has been shown to recognise a GT mismatch (Lacy *et al.*, 2004).

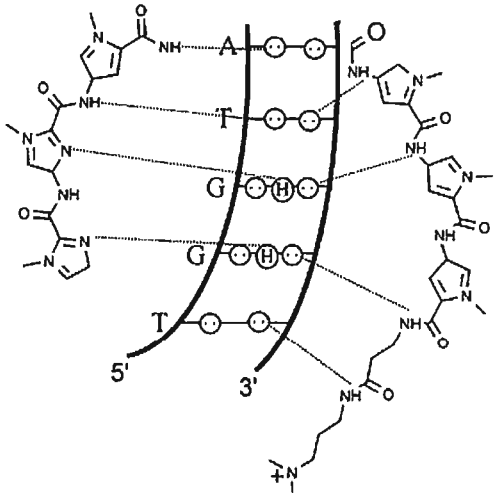


Figure 1.6: Imidazole and pyrrole binding to GC and AT sequences. Circles with dots represent the lone pairs of purine N(3) and pyrimidine O(2) and circles containing an H represent the 2-amino group of guanine (taken from Wemmer and Dervan (1997)).

Other studies have searched for novel guanine-targeting polyamide units. One recent example is imidazopyridine, which binds to guanine with comparable affinity to imidazole (Renneberg and Dervan, 2003; Marques *et al.*, 2004). 3-pyrazole (Zhan and Dervan, 2000)

and oxazole (Doss *et al.*, 2006) have also shown increased selectivity for guanine residues. Expanding the repertoire of aromatic rings for minor groove recognition ensures the strongest possible binding can be established for the varying microstructure of the DNA at different sequences.

Covalently linked dimers

The side-by-side binding of two polyamides in the 2:1 mode provides a means for recognising unique DNA sequences. However the use of such unlinked heterodimers presents a problem, as the two halves of the dimer can pair in other combinations which have different sequence specificities. For example, although the ImPyPy.PyImPy combination should recognise GC(A/T), the homodimers of ImPyPy and PyImPy will bind to G(A/T)G and (A/T)G(A/T) respectively. In addition, the two halves of longer molecules may slip relative to each other, altering the sequence specificity (Mrksich and Dervan, 1993a).

These problems have been overcome by covalently linking the two halves of the dimer. This has been achieved in three different ways, using hairpins, cyclic or central H-Pins (also known as crosslinked or "stapled") to join the two halves of the dimer (Figure 1.7) (Mrksich *et al.*, 1994; Cho *et al.*, 1995; Mrksich and Dervan, 1994). All these structures enhance the specificity and affinity of binding compared to unlinked ligands.

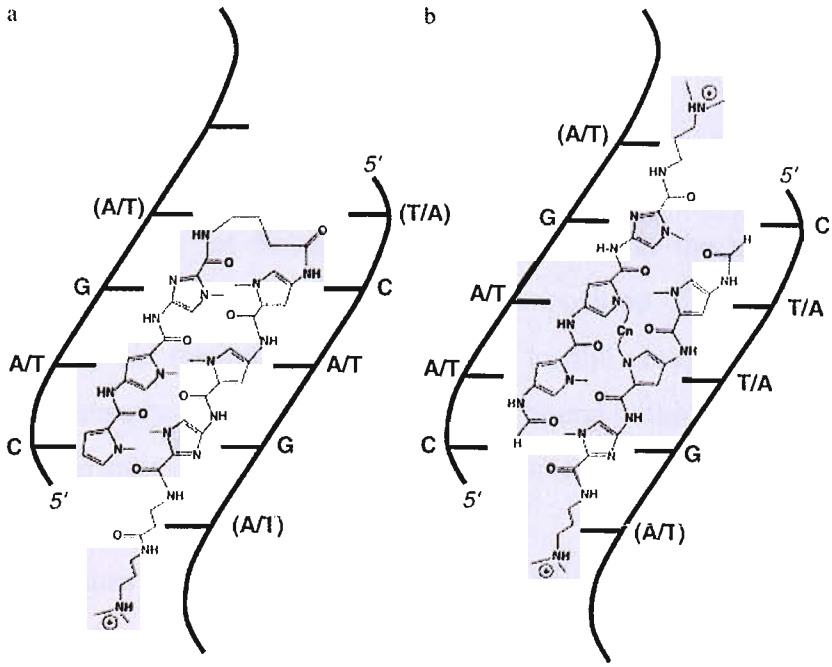


Figure 1.7: Examples of different types of dimer linkage: a) Hairpin b) H-Pin (taken from Walker *et al.* (1998a)).

The most widely used method for joining the two halves of the dimer is the hairpin linkage (Figure 1.7a). Dervan's group initially tested a series of hairpin linkers, with one to four methylene carbons attached to the ligand molecules via amines and carboxylates (Mrksich *et al.*, 1994). The three-carbon linker (gamma amino butyric acid (γ)) gave the highest binding affinity. Linkers with fewer carbons produced strained molecules, while longer linkers caused steric clashes. The γ-linker forms a tight turn, which fits well at the base of the minor groove, but also has a strong preference for binding at AT base pairs (de Clairac *et al.*, 1997; Swalley *et al.*, 1999). This is probably due to steric clash with the 2-amino group of guanine.

By connecting the carboxyl and amino terminals with a γ-linker in a head-to-tail manner, the binding affinity is increased by about 100 fold, high enough to compete successfully with transcription factors and other DNA binding proteins (Dervan and Bürli, 1999). These hairpin structures also have higher specificity than their monomeric counterparts. Hairpin polyamides containing eight rings therefore have an affinity and specificity that is comparable to endogenous DNA minor groove-binding proteins (Dervan and Bürli, 1999).

Cyclic polyamides containing the same γ-linkage have also been tested, though these compounds were more challenging to synthesise. These compounds showed high affinity for DNA, but their specificity was reduced significantly, suggesting that some flexibility is required to optimise local contacts (Cho *et al.*, 1995; Wemmer and Dervan, 1997; Herman *et al.*, 1999).

The H-Pin linkage (Figure 1.7b) uses the N-methyl groups of the pyrrole and imidazole rings as attachment points for linking the two molecules in the centre of the dimer. (Mrksich and Dervan, 1994). Studies varying the composition of the linkage have found that dimers linked by six methylene groups (with a formyl group on the N-terminus) or seven methylene groups (with an amino N-terminus), bound within the DNA minor groove with high binding affinity compared to longer or shorter linkages (Mrksich and Dervan, 1993b; O'Hare *et al.*, 2002).

The N-terminus of the ligand also influences the binding characteristics of crosslinked polyamides and affects the optimal linker length. This is probably due to the two hydrogen bonds that are created between the formyl group and the minor groove floor. Hydrogen bond interactions of the N-terminal head group with DNA may be necessary for

crosslinked side-by-side binding (O'Hare *et al.*, 2002), which are not required for hairpin linkage.

Linked polyamides have two potential DNA-binding modes, in which the two halves are fully overlapped or staggered. In the maximum overlap mode (favoured by hairpin linkage), 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 neighbour. 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 can also stack ring-on-amide, but 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 maximising the distance between the charges (O'Hare *et al.*, 2002).

Inclusion of β-alanine

At an early stage in the development of synthetic DNA minor groove binding ligands it was found that the binding affinity does not improve for 2:1 polyamides with greater than five units (Kelly *et al.*, 1996). This occurs because the unit repeat length and curvature of the polyamide does not match that of DNA and the two become out of register. A method to make minor groove binding ligands longer, so as to enhance their sequence specificity, but without compromising their affinity was therefore required.

Glycine and β-alanine have been used as short flexible inserts, in place of a pyrrole or imidazole rings; these introduce flexibility into the molecule and allow ligands to bind to long target sites (Trauger *et al.*, 1996; Swalley *et al.*, 1997). β-alanine (β) is the most effective linker, producing ligands with high binding affinity and specificity. The β/β pair is specific for AT base pairs, while β/Py or Py/β increase the binding affinity and specificity relative to Py/Py (Turner *et al.*, 1998). These β-containing polyamides therefore provide an optimal combination of size, flexibility and alignment of the polyamide-paired aromatic subunits in the DNA minor groove, allowing the targeting of 11 to 16 base pairs, and highly cooperative dimeric polyamides using this motif have been now documented (Dervan and Bürli, 1999).

Hydroxypyrroles: AT, TA specificity

One problem with using distamycin-derived ligands for sequence recognition is that the Py/Py pair does not distinguish between AT and TA. Rational design of an adenine or thymine-specific module was a difficult prospect due to the similar hydrogen bonding properties of the minor groove-accessible faces of adenine, thymine and cytosine (Walker *et al.*, 1998a). However, in the late-1990s Dervan's group introduced a third aromatic ring (N-methylhydroxypyrrole, Hp) to discriminate between AT and TA (White *et al.*, 1998). Hydroxypyrrole was tested on the assumption that a small substituent on one corner of the Py/Py pair would favour thymine over adenine by steric clash with adenine. An hydroxyl group was chosen as this substituent as it was small, but also because it provided the possibility of forming hydrogen bonds with the lone pair electrons at O2 of thymine (Kielkopf *et al.*, 2000; Dervan, 2001). Py/Hp recognises AT base pairs while Hp/Py targets TA. This now provides a means for selectively targeting each DNA base, as shown in Figure 1.8.

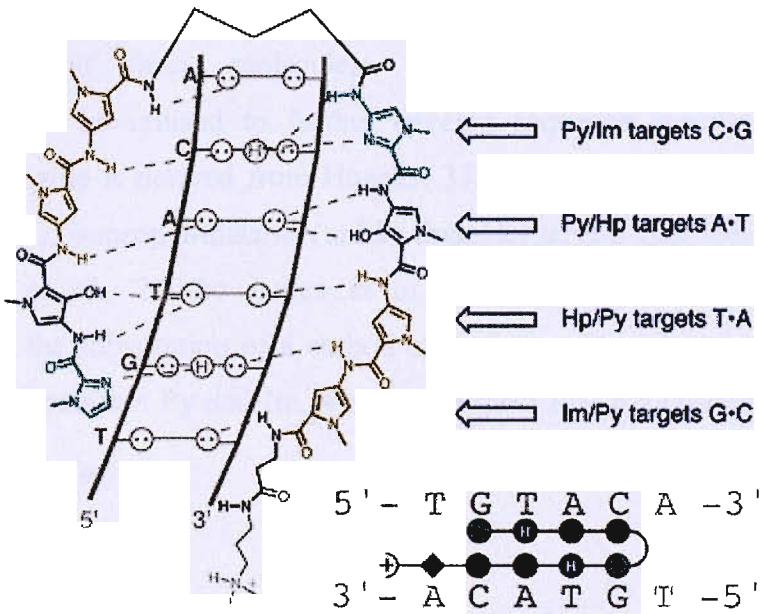


Figure 1.8: Molecular recognition of the minor groove of DNA by hairpin polyamides containing Py, Im and Hp. Circles with dots represent lone pairs of purine N3 and pyrimidine O2 and circles containing an H represent the 2-amino group of guanine (taken from Dervan (2001)).

However, other means for discriminating between TA and AT are still being explored (as hydroxypyrrole degrades over time in the presence of acid or free radicals, so is not very biologically robust (Renneberg and Dervan, 2003)), including the use of the thiazole ring

as an adenine discriminator paired opposite pyrrole (Kopka *et al.*, 1997). At present, hydroxypyrrole remains the unit of choice for distinguishing between AT and TA, although the related hydroxybenzimidazole has recently shown selective targeting of thymine (Renneberg and Dervan, 2003; Marques *et al.*, 2004). Thiophene or furan based ligands have also shown great potential (Mallena *et al.*, 2004; Chaires *et al.*, 2004), with chlorothiophene found to target thymine selectivity (Doss *et al.*, 2006).

A range of different polyamide subunits (for targeting each base) are required to target specific DNA tracts (and so genes) as strongly as possible, as the microstructure of the minor groove varies for different sequences and so a subunit may bind a region of DNA more strongly than another and vice versa at a different site.

Other polyamide subunits relevant to this thesis

The work on polyamides described in this thesis studies novel minor groove binding ligands that contain some alternative subunits. Figure 1.9 shows some of the subunits that have been used in these molecules: Isopropyl-thiazole; benzimidazole; and pyridoimidazole will be utilised to further develop sequence specificity and binding affinity. Benzimidazole is derived from Hoechst 33258 and should target AT base pairs (Briehn *et al.*, 2003); isopropyl-thiazole (unlike thiazole) targets guanines (Anthony *et al.*, 2004a; Anthony *et al.*, 2004b; James *et al.*, 2004). Pyridoimidazole differs from benzimidazole by the substitution of a carbon atom for a nitrogen atom. By comparison with the difference between Py and Im, we might expect that pyridoimidazole will target guanine.

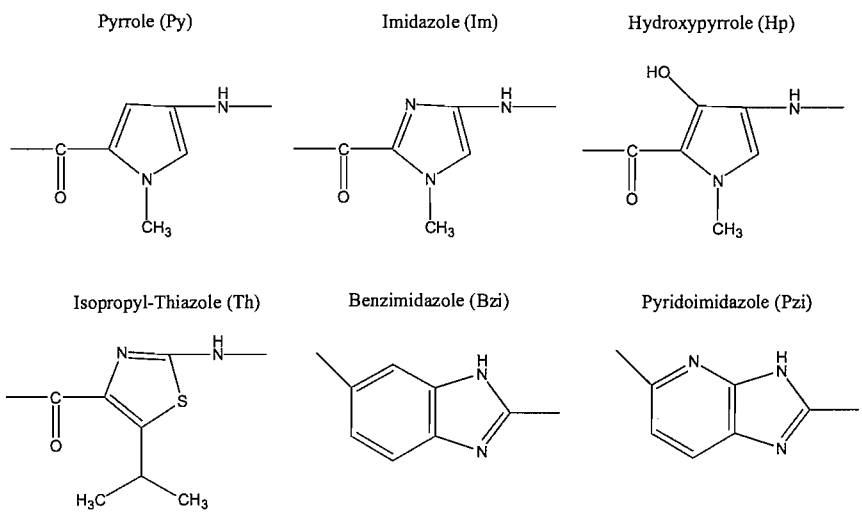


Figure 1.9: Structures of pyrrole, imidazole, hydroxypyrrole, isopropyl-thiazole, benzimidazole and pyridoimidazole.

Dimethylaminopropylamide (Dp) and methylpiperazine (Mp) have been used as C-terminal moieties that contain positive charges, which may aid binding and directional orientation (White *et al.*, 1997b; Brown *et al.*, 2007).

Potential use of polyamides

The potential for these DNA minor groove binding ligands is extraordinary and the first polyamides and their derivatives are already in clinical or experimental use (Dervan, 2001; Murty and Sugiyama, 2004).

Most DNA is found in chromatin in the form of nucleosomal DNA. It therefore follows that polyamides have to be able to bind DNA in this form to affect cellular processes before the genes are expressed. It has been discovered that nucleosomal DNA sequences facing away from, or even partially facing, the histone octamer are fully accessible by polyamides (Gottesfeld *et al.*, 2001; Gottesfeld *et al.*, 2002; Leslie and Fox, 2002; Suto *et al.*, 2003). Binding is only prevented by interactions between the DNA and the histone octamer (Gottesfeld *et al.*, 2001). Minor groove ligands such as polyamides are therefore viable pharmacological tools.

Anti-cancer drugs

Polyamides are of interest as potential anti-cancer drugs as they can affect a number of cancer-related processes.

The distamycin-alkylating agent conjugates brostacillin and tallimustine have been used to selectively target DNA mismatch repair-deficient human colon tumour cells to induce cytotoxicity (Fedier *et al.*, 2003), whilst a hairpin polyamide-alkylating agent conjugate has been used for gene-silencing of the same tumour cells (Shinohara *et al.*, 2006). Gene silencing has also been exhibited with polyamide-targeted copper-mediated DNA cleavage (Subramanian *et al.*, 2006).

Transcription factor binding sites on DNA are attractive potential targets for minor groove binding ligands, as their blockades should result in modulation of gene expression which is uncontrolled in cancer. Modulation of transcription factor binding using sequence-specific binding compounds has received intensive investigation in recent years and sequence selective compounds have been shown to selectively inhibit the binding of various transcription factors, including LEF-1 involved in colon cancer (Supekova *et al.*, 2002) and

Ets involved in breast cancer (Chiang *et al.*, 2000). The human multi-drug resistance 1 gene promoter region has also been targeted by a hairpin polyamide (Buchmueller *et al.*, 2005).

Minor groove binding agents based on distamycin and Hoescht 33258 have also been shown to inhibit DNA helicases (Soderlind *et al.*, 1999; Brosh Jr *et al.*, 2000), topoisomerases (Bell *et al.*, 1997; Flores *et al.*, 2006), DNA replication processes in general (Sun and Hurley, 1992) and DNA repair (Brooks *et al.*, 1999). All of which are important in cancer therapy.

Many of these target proteins bind in the major groove of DNA (or at least a component does). Minor groove binding ligands are therefore used due to their property of distorting the DNA helix on binding, thereby preventing binding in the major groove as well as the minor groove (Blattes *et al.*, 2006). However, modifications to polyamides to include a major groove binding element have been developed. For example, an eight-ring hairpin polyamide linked to an Arg–Pro–Arg tripeptide on the carboxyl terminus has been shown to clamp a positive patch to the DNA backbone and interfere with crucial protein–phosphate contacts (Bremer *et al.*, 1998). Inhibition of major groove binding protein–DNA interactions may also be achieved using polyamide–oligodeoxynucleotide (*i.e.* a triplex forming oligonucleotide) conjugates that bind simultaneously to the minor and major grooves of DNA (Dervan and Bürli, 1999).

Anti-infection drugs

One of the main focuses of polyamides research for treating human infections is the human immunodeficiency virus (HIV). The polymerase II promoter region of HIV-1 has been targeted selectively (Dickinson *et al.*, 1998; Tutter and Jones, 1998; Ehley *et al.*, 2002), as has the long terminal repeat involved in host-DNA integration and transcription (Coulli *et al.*, 2002; Mischiati *et al.*, 2004).

Other viruses have also been targeted with polyamides, mainly through inhibition of transcription factor binding, such as the human cytomegalovirus (Dickinson *et al.*, 1999) and human papilloma virus (Schaal *et al.*, 2003).

The protozoan parasite *Plasmodium falciparum* is responsible for the most lethal form of human malaria and has a genome comprised of 82 % AT base pairs, with large tracts of DNA comprised almost entirely of these bases. Several A/T-selective minor groove

binding compounds have been demonstrated to be effective against the blood stages of this intracellular parasite *in vitro*, including distamycin (Ginsburg *et al.*, 1993). Polyamides have also shown potential as anti-fungal (Marini *et al.*, 2003) and anti-bacterial agents (Simon *et al.*, 2000).

Other uses

The human transforming growth factor β1 involved in several cardiovascular diseases including stroke and ischemic heart disease, as well as progressive renal disease has also been selectively targeted by a polyamide (Lai *et al.*, 2005; Matsuda *et al.*, 2006).

DNA minor groove binding ligands have also been used as transcription activators to upregulate gene expression (Arora *et al.*, 2002; Schmitz and Schepers, 2004; Kwon *et al.*, 2004). This is possible by targeting a sequence upstream of a transcription binding site by a hairpin polyamide, which is linked by a flexible tether to short activating peptides (the activation domain). The complex transcription machinery is recruited to the transcription factor binding site and so the gene is expressed.

Polyamides have also found use as a biochemical tool, such as fluorescently labelling telomeres (Maeshima *et al.*, 2001). By fluorescently tagging a telomere-targeted polyamide, staining is more rapid than conventional hybridisation methods. Telomere length can therefore be quickly established to observe the effects of this property on the cell cycle of, for example, cancer cells.

***

It is estimated that 50 % of the DNA sites on any gene promoter can now be targeted with minor groove binding ligands (Dervan, 2001), which should be sufficient to target most important transcription factors. Despite this, recent investigations have shown that some inhibition *in vitro* is not matched by *in vivo* activity (Chiang *et al.*, 2000; Best *et al.*, 2003). However, the pharmacokinetics, bioavailability and toxicity of polyamides still need to be established, and it seems likely that one day DNA minor groove binding drugs will form the basis of new classes of gene targeted therapies.

Several other sequence specific DNA binding ligands have been used in the work described in this thesis. Their properties are described below.

Mithramycin

The aureolic acid anti-tumour antibiotic mithramycin (Figure 1.10) and the related compounds chromomycin and olivomycin contain an aglycone chromophore to which several sugar residues are attached; the A-B disaccharide and C-D-E trisaccharide segments projecting from opposite ends of the chromophore (Bakhaeva *et al.*, 1968).

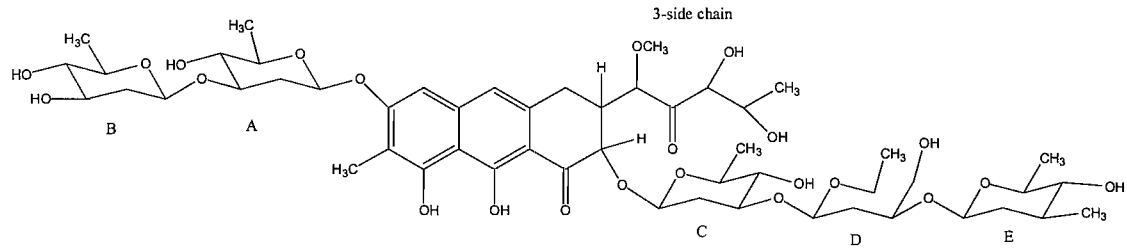


Figure 1.10: Structure of mithramycin.

Mithramycin is coordinated by Mg^{2+} ions to form an antiparallel dimer through interactions between the cation and the O^9 and O^1 oxygens of the aglycone chromophores (Sastry and Patel, 1993). The aglycone chromophores then bind in the DNA minor groove through the formation of hydrogen bonds between the 2-amino protons of guanine and acceptor phenolic hydroxyl O^8 oxygens, although van der Waals interactions are also important (Sastry and Patel, 1993; Sastry *et al.*, 1995). Mithramycin is therefore a G/C-specific minor groove binding ligand, with at least two contiguous GC base pairs necessary for binding (Van Dyke and Dervan, 1983).

Fox and Howarth (1985), Cons and Fox (1989) and Carpenter *et al.* (1993) have used DNase I footprint to show CpG steps to be weak mithramycin binding sites, whilst GpC are stronger and GpG/CpC is the preferred sequence.

The saccharide groups are also involved in binding, with the C-D segment interacting edgewise in the antiparallel alignment with G/C bases through hydrophobic contacts (Sastry and Patel, 1993). The E sugar lies in the minor groove across both duplex strands and creates hydrogen bonds with guanine or cytosine (it can accept or donate hydrogen bonds, so reducing selectivity by the ligand), as well as hydrophobic interactions, to stabilise ligand binding at the terminals (Sastry and Patel, 1993). The A-B disaccharide segment interacts in the minor groove with the sugar phosphate backbone and acts as a weak clamp for the mithramycin dimer to prevent it sliding along the helix groove (Sastry and Patel, 1993). Although Van Dyke and Dervan (1983) observed that at least two

contiguous GC base pairs were required for recognition, they also saw that a minimum of three base pairs were involved in binding. NMR studies have shown that binding of the saccharide groups of mithramycin means that a sequence of six base pairs is actually involved in ligand binding (Sastry and Patel, 1993).

The mithramycin dimer cannot be accommodated in a narrow minor groove (Sastry and Patel, 1993). In such regions, strong binding will occur if there is a neighbouring TpA step present, which will create a "kink" in the DNA so widening the minor groove and allowing the E sugar to be accommodated within the groove (Sastry *et al.*, 1995).

Mithramycin has been found to inhibit both cancer growth and bone resorption by binding G/C-rich DNA, which blocks binding of Sp-family transcription factors to regulatory elements. By stopping Sp factors binding to the c-src promotor region, for example, many human cancers can be treated (Remsing *et al.*, 2003a). Mithramycin also induces myeloid differentiation of HL-60 promyelocytic leukaemia cells, so is an effective agent in certain patients with chronic granulocytic leukaemia (Miller *et al.*, 1987). Mithramycin SK is a recently discovered derivative of mithramycin (Remsing *et al.*, 2003b), with a hydroxy-methoxy-oxo-butyl 3-side chain the only difference. This substitution elicits up to nine-fold higher activity against melanoma, leukaemia and central nervous system cancer cells.

DNA Intercalators

Another major group of ligands that bind to DNA are the intercalators (Baguley, 1991; Waring and Bailly, 1994; Braña *et al.*, 2001; Dawson *et al.*, 2007). These compounds all possess planar polyaromatic systems, which bind by insertion between DNA base pairs, often favouring pyrimidine/purine steps (Braña *et al.*, 2001). These chromophores can be linked to other groups (such as peptides), which play an important role in determining the affinity and selectivity of binding.

A selection of DNA intercalating agents that have been used in this investigation are described on the following pages.

Actinomycin D

Actinomycin D is a chromopeptide comprised of a phenoxazone chromophore connected to two depsipeptide ring systems of mainly hydrophobic residues, as shown in Figure 1.11.

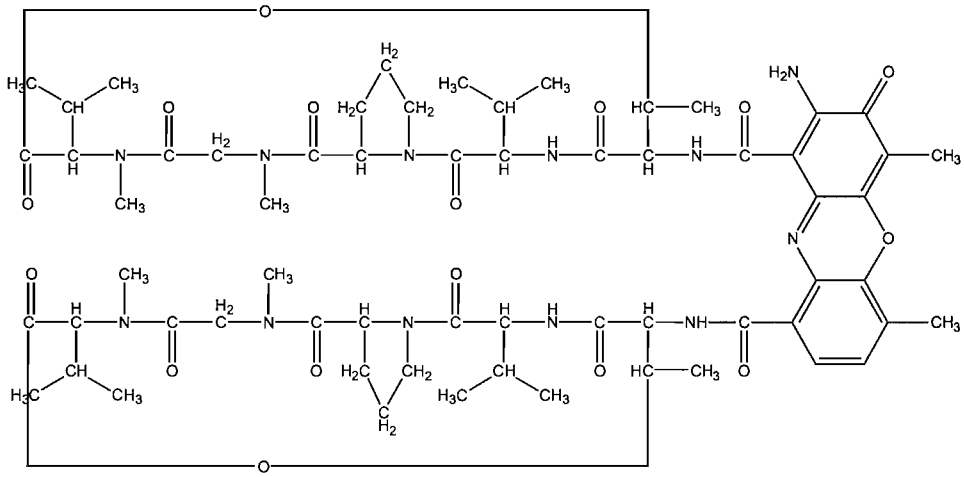


Figure 1.11: Structure of actinomycin D.

Actinomycin D binds to double stranded DNA by intercalation of its phenoxazone chromophore, positioning the cyclic depsipeptides in the minor groove. Its preferred binding site is GpC, as shown by footprinting (Van Dyke *et al.*, 1982; Scamrov and Beabealashvilli, 1983; Fox and Waring, 1984a) and structural studies (Muller and Crothers, 1968; Sobell and Jain, 1972; Sobell, 1973). It binds to this sequence by forming hydrogen bonds between its threonine carbonyls and the 2-amino groups of guanines positioned on either side of the intercalation site (Kamitori and Takusagawa, 1994; Chen *et al.*, 2003a). The importance of the 2-amino group has been shown by studies using inosine-containing DNA fragments (Jennewein and Waring, 1997), although weaker binding has been seen with ApC and GpN steps (Chen *et al.*, 2004).

Binding of actinomycin D to DNA is in several phases, as the ligand binds to an optimal site (Muller and Crothers, 1968; Fox and Waring, 1984b; Brown and Shafer, 1987) and consequently changes DNA conformation (Waterloh and Fox, 1991a). The phenoxazone chromophore causes the DNA helix to be unwound by rotating one base pair at the intercalation site (Kamitori and Takusagawa, 1994), so allowing the depsipeptide rings to bind strongly in the minor groove (Kamitori and Takusagawa, 1994). Overall intercalation of actinomycin unwinds the DNA helix by 26 ° (Waring, 1970; Muller and Crothers, 1968), creating an asymmetrically wound helix that extends for several base pairs beyond the actual binding site (Kamitori and Takusagawa, 1994). This produces a long-

range effect on DNA conformation that may be responsible for the anti-cooperative binding observed with a GCGC oligomer (Scott *et al.*, 1988) and explains the enhanced DNase I cleavage seen in footprinting studies that is propagated into neighbouring regions of A/T DNA (Waterloh and Fox, 1991a).

Actinomycin D can also significantly change DNA structure to produce its optimal GpC binding site, either by forming hairpins (Chen *et al.*, 2003b) or by looping out regions of single strand DNA (Robinson *et al.*, 2001; Chen *et al.*, 2004). By forming hairpins in single strand DNA, actinomycin can serve as a sequence-specific single strand DNA binding agent that inhibits HIV and other retroviruses replicating through single stranded DNA intermediates (Chin *et al.*, 2003). Actinomycin D has previously been used as an anti-tumour drug by inhibiting transcription (Horwitz and McGuire, 1978; Chen, 2002).

Quinoxaline antibiotics – Bis-intercalators

Echinomycin

Echinomycin (Figure 1.12) is a member of the quinoxaline group of antibiotics that also includes triostin A. These compounds possess a cyclic octadepsipeptide, which contains a thioacetal or disulphide cross-bridge, to which two quinoxaline chromophores are attached (Waring and Wakelin, 1974; Wakelin, 1986). They bind to DNA by the simultaneous bis-intercalation of the quinoxaline rings, positioning the cyclic peptide in the minor groove, where it makes specific contacts with the bases.

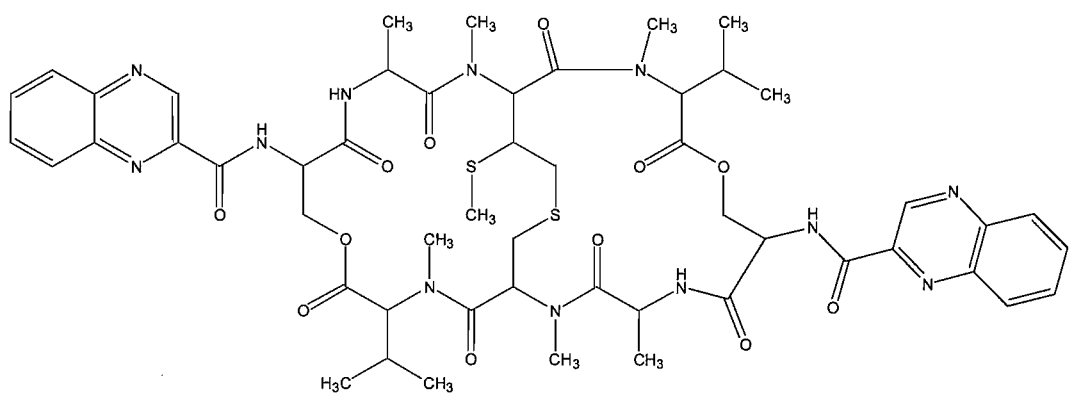


Figure 1.12: Structure of echinomycin.

Echinomycin binds selectively to the dinucleotide step CpG (Low *et al.*, 1984a; Van Dyke and Dervan, 1984), with these two base pairs sandwiched between the quinoxaline

chromophores. However, other weaker binding sites have been observed, such as TpG (Van Dyke and Dervan, 1984), CpC and CpA (Waterloh and Fox, 1991b). The 2-amino group of guanine is therefore the only essential feature of the binding site, as shown by experiments with DNA fragments containing inosine or diaminopurine (Jennewein and Waring, 1997).

A number of crystal and NMR structures have revealed the details of the interaction between echinomycin and DNA, to explain the selectivity for CpG (Wang *et al.*, 1984; Ughetto *et al.*, 1985; Gao and Patel, 1988; Gilbert and Feigon, 1992). The crucial interactions are hydrogen bonds between the carbonyls of the two alanines and the 2-amino groups of the guanines (Wang *et al.*, 1984; Ughetto *et al.*, 1985). The valine residues also aid binding by creating van der Waals interactions with the deoxyribose sugars in the DNA backbone (Ughetto *et al.*, 1985). The other residues in the depsipeptide ring (serine and cysteine) are important for the structural integrity of the backbone ring (Ughetto *et al.*, 1985).

The intercalation of the quinoxaline chromophores of echinomycin alters local DNA duplex conformation (Low *et al.*, 1984) to unwind the DNA helix by 48 ° in regions flanking the CpG binding site (Waring and Wakelin, 1974; Fox *et al.*, 1986), twice the unwinding angle of monofunctional intercalators such as ethidium. Elongation of the DNA also occurs (Tseng *et al.*, 2005), with these two effects altering six base pairs of DNA (Tseng *et al.*, 2005), even though only four base pairs are directly involved in binding (Van Dyke and Dervan, 1984). The distortion of the DNA destabilises the base pairing adjacent to the CpG site, enabling the formation of Hoogsteen pairing at flanking AT base pairs, which has been observed in NMR and crystal structures (Wang *et al.*, 1984; Gilbert and Feigon, 1992). However, this Hoogsteen base pairing is not essential for binding (Gilbert and Feigon, 1992; Sayers and Waring, 1993). Echinomycin does not bind to all CpG sites with equal affinity and the strongest is observed at ACGT and TCGT, with adenine selectivity most likely due to an interaction between the base and the valines of the ligand (Ughetto *et al.*, 1985; Gao and Patel, 1988). Cooperative binding has been observed when echinomycin binds to multiple adjacent sites (Gilbert and Feigon, 1992; Bailly *et al.*, 1996; Leslie and Fox, 2002), although alternating G/C-tracts create anti-cooperativity (Wakelin and Waring, 1976).

Echinomycin has recently been shown to have impressive activity against the malarial parasite P.*fulciparum* (LD_{50} = 7-10 ng/ml) (Castillo *et al.*, 2003) and so may have potential

use in the field of antimalarials. It has also shown potential as an anti-tumour agent (Dolma *et al.*, 2003), inhibiting transcription factor binding (Kong *et al.*, 2005), chromatin decondensation, nuclear assembly and DNA replication (May *et al.*, 2004). New echinomycin analogues have recently been developed to further the pharmacological use of the ligand (Kim *et al.*, 2004).

TANDEM

Triostin **A** ***N*-Dem**ethylated (TANDEM) is a synthetic derivative of the quinoxaline antibiotic triostin A (Viswamitra *et al.*, 1981; Ughetto *et al.*, 1985; Wakelin, 1986), which lacks the N-methyl groups on the valine and cysteine residues, as shown in Figure 1.13.

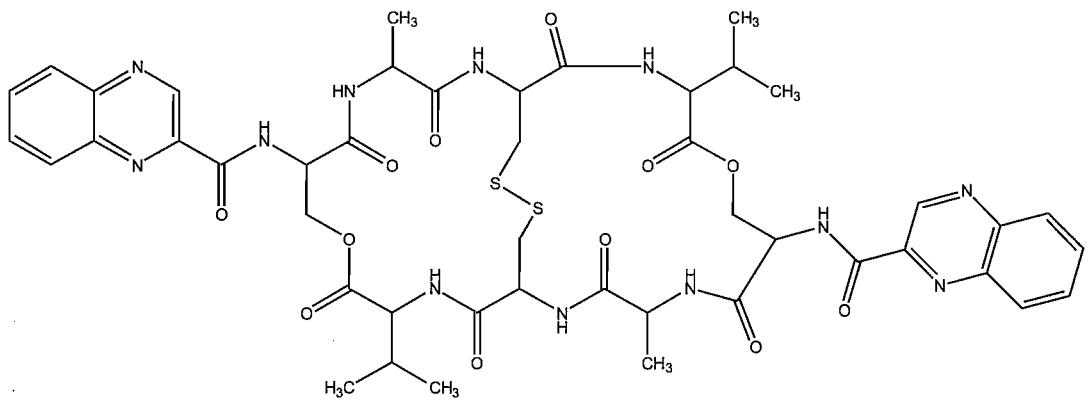


Figure 1.13: Structure of TANDEM.

This small alteration results in a dramatic change in its sequence selectivity and TANDEM binds selectively to TpA steps (Lee and Waring, 1978; Low *et al.*, 1984b; Waterloh *et al.*, 1992; Lavesa *et al.*, 1993). This binding is about 20-fold weaker than the parent compounds echinomycin and triostin A (Lee and Waring, 1978).

NMR studies (Addess *et al.*, 1993) have revealed that TANDEM binds in the same way as the parent compounds, with the two quinoxaline rings bis-intercalating around the TpA step and causing local unwinding of the DNA helix, while the peptide ring lies in the minor groove (Addess *et al.*, 1993). The electrostatic component of the stacking interactions between the chromophores and bases that make up intercalation sites plays a very important role in binding affinity (Marco *et al.*, 2005).

The change in sequence selectivity is due to the formation of intramolecular hydrogen bonds between the two valine NH groups and the alanine carbonyls (which are involved in hydrogen bonds to the 2-amino group of guanine in the parent compounds) (Addess *et al.*, 1993). This interaction rotates the side chains of the valines inwards, so

narrowing the width of the peptide ring. Adenine bases are then targeted by the formation of intermolecular hydrogen bonds between the N3 atom of the base and the alanine NH groups (Addess *et al.*, 1993). However, this hydrogen bonding is not thought to be the dominant factor in TANDEM binding, and the selectivity is mainly attributed to steric or hydrophobic interactions with a minor groove of suitable dimensions (Bailly and Waring, 1998). Strong binding requires the presence of an intact disulphide cross-bridge (Lee and Waring, 1978; Malkinson *et al.*, 2005), though an analogue in which alanines replace the cysteines still binds to A/T-rich DNA, albeit with a much lower affinity (Fox *et al.*, 1980a). It has long been known that TANDEM binds cooperatively to poly(dA-dT) (Lee and Waring, 1978) and more recent work has shown that the neighbouring bases to TpA affect the binding; ATAT shows the highest affinity while there is little interaction with TTAA (Fletcher *et al.*, 1995; Lavesa and Fox, 2001). Longer alternating A/T-tracts generate stronger binding sites (Fletcher *et al.*, 1995).

Nogalamycin

Nogalamycin is an anthracycline antibiotic that bears bulky sugar residues at both ends of an anthracycline chromophore (Figure 1.14). It is a member of a class of DNA binding agents known as "threading intercalators". These compounds intercalate into duplex DNA, but leave pendant groups in *both* DNA grooves. In the case of nogalamycin, the positively charged fused bicycle amino sugar binds in the major groove and the uncharged nogalose sugar (with methyl ester) in the minor groove (Searle *et al.*, 1988; Williams *et al.*, 1990). This mode of binding requires local denaturation of the DNA duplex, so as to insert the ligand.

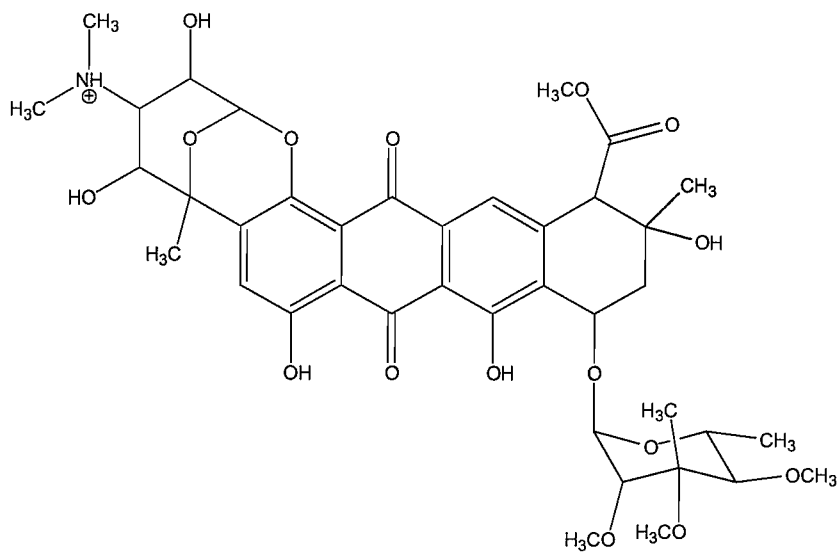


Figure 1.14: Structure of nogalamycin.

Since the local DNA structure has to be unpaired before it can bind, nogalamycin displays very slow association and dissociation kinetics (Fox and Waring, 1984c; Fox *et al.*, 1985), which are both faster for A/T- than G/C-rich DNAs. DNase I footprinting studies have shown that nogalamycin possesses some limited sequence binding preferences and appears to bind best to regions of alternating purines and pyrimidines that contain all four DNA bases (Fox and Waring, 1986; Fox and Alam, 1992). Several crystallographic and NMR structures have been determined for nogalamycin bound to short DNA fragments in which the ligand is intercalated between different base steps: CpG (Gao *et al.*, 1990; Robinson *et al.*, 1990); TpG (Searle *et al.*, 1988); and CpA (Zhang and Patel, 1990). This large array of binding sequences emphasises that nogalamycin does not have an absolute sequence binding preference, although a hydrogen bond may be formed between the 2-amino group of guanine and the methyl ester of nogalamycin (Williams *et al.*, 1990). The ligand appears to discriminate between different DNA sequences by virtue of their structure and stability. Intercalation is always at YpR steps; but the best binding sites contain combinations of all four bases.

Techniques for studying DNA binding specificity and affinity

Three main techniques are used in this thesis to study the selectivity and affinity of different DNA binding ligands. They are: DNase I footprinting; hydroxyl radical footprinting; and fluorescence melting.

Footprinting

DNA footprinting is a protection assay in which cleavage of DNA is inhibited at discrete locations by the sequence-specific binding of a ligand or protein. A DNA fragment of known sequence and length (typically a restriction fragment of 100-200 bp), which has been selectively radiolabelled at one end of one strand, is lightly digested (single hit kinetics) by a suitable endonucleolytic probe in the presence and absence of the drug under investigation, as shown in Figure 1.15. The cleavage agent is prevented from cutting around the drug-binding sites, so that when the products of reaction are separated on a denaturing polyacrylamide gel and exposed to autoradiography, the position of the ligand can be seen as a gap (footprint) in the otherwise continuous ladder of bands (Fox, 1997).

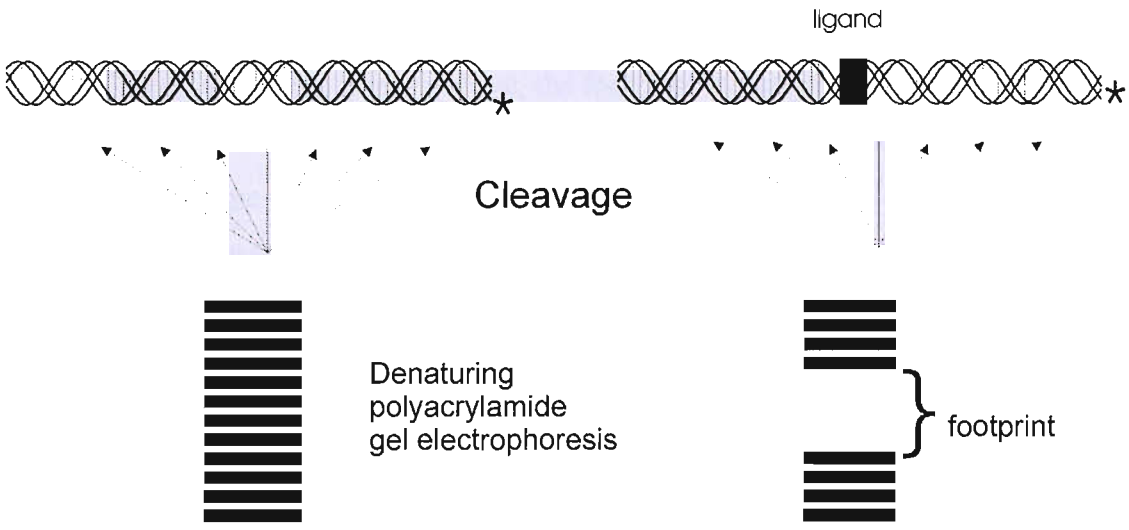


Figure 1.15: Footprinting in the absence (left) and presence (right) of a DNA binding ligand.

DNase I footprinting

A commonly used endonucleolytic probe for footprinting is the monomeric glycoprotein DNase I. This is a double-strand specific endonuclease that introduces single-strand nicks in the phosphodiester backbone of DNA, cleaving the O3'-P bond. The enzyme requires the presence of divalent cations (especially calcium and magnesium). It binds by inserting an exposed loop into the DNA minor groove, interacting with the phosphate backbone and groove walls (Suck *et al.*, 1986; Lahm and Suck, 1991; Weston *et al.*, 1992). DNase I does not have any simple sequence dependency but it produces an uneven cleavage pattern in which $A_n.T_n$ tracts and G/C-rich regions are poorly cleaved (Drew and Travers, 1984). ApT is also cut much more efficiently than TpA. Regions of duplex DNA with unusually narrow minor grooves (such as $A_n.T_n$ (Fox, 1992)) are poor substrates, as the enzyme cannot insert its exposed loop into the minor groove. Crystal structures of oligonucleotides-DNase I complexes show that the DNA is bent towards the major groove; this may be an essential part of the catalytic mechanism and explains why rigid DNA sequences (such as G/C-rich regions) are poor substrates for the enzyme.

DNase I has a molecular weight of 30.4 kD and covers about 10 bp of DNA (one complete turn of the DNA helix). The large size of this probe causes an overestimation of drug binding site sizes, as it cannot cleave close to the bound ligand. Another factor that

affects the interpretation of DNase I cleavage patterns is that although DNA bases lie perpendicular to the helical axis, they are inclined relative to the phosphodiester backbone. Since DNase I binds across the minor groove, the footprint is staggered by 2-3 bases in the 3'-direction across the two strands. When a DNA-binding ligand is added, the closest approach of the enzyme is not the same on each strand (Figure 1.16). DNase I can therefore approach closer to the ligand on the lower strand, while the region of the upper strand protected extends by a further 2-3 bases beyond the actual ligand-binding site. As a result, DNase I footprints are staggered by about 2-3 bases in the 3'-direction across the two strands.

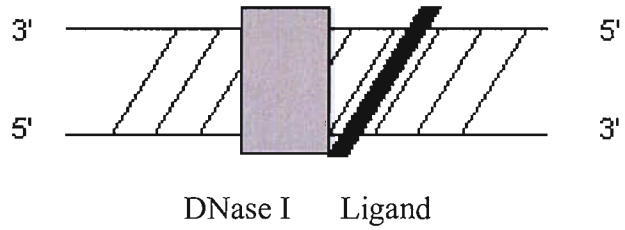


Figure 1.16: DNase I binding to and cleaving DNA in the presence of a ligand, showing the 3' stagger.

Hydroxyl radical footprinting

Hydroxyl radicals are another endonucleolytic probe that is used in DNA footprinting (Tulius, 1988). These are formed using the Fenton reaction (Equation 1.1) between ferrous ions and hydrogen peroxide. The hydroxyl radical is a highly reactive and freely diffusible species, which can approach close to the bound ligand as a result of its small size. The resolution problems suffered by DNase I are therefore not a factor in hydroxyl radical footprinting. In addition, hydroxyl radicals generate an even cleavage pattern in the drug-free lanes (though cleavage is attenuated in $A_n.T_n$ tracts).

Although the exact mechanism of hydroxyl radical cleavage is still unclear, it is accepted that the free radicals attack the DNA backbone at the C4' or C1' positions, leading to subsequent base removal and strand scission (Pogozelski *et al.*, 1995). Although greater resolution makes hydroxyl radical footprinting more accurate, it is a more complex and time-consuming technique so is generally used only after preliminary DNase I experiments. The reaction is quenched by free radical scavengers, so cannot be used for ligands such as echinomycin and TANDEM, for which stock solutions are prepared in

DMSO. It also cannot be used for ligands that bind in the DNA major groove, such as triplex-forming oligonucleotides.

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + \bullet OH$$

Equation 1.1: The Fenton reaction, showing the creation of hydroxyl radicals. The Fe^{2+} is usually complexed with EDTA to prevent a direct interaction of Fe^{2+} with DNA.

Fluorescence melting

Another way to study DNA binding ligands is through DNA melting experiments. Ligand molecules that bind to duplex DNA will stabilise it and so increase the melting temperature (T_m). This has typically been studied by measuring changes in the absorbance at 260 nm, determining the relative binding affinities of different ligands by comparing their effects on the T_m. A more sensitive and high throughput variation of this technique uses synthetic oligonucleotides containing a fluorophore attached to one strand and a quencher linked opposite this on the other strand (Darby *et al.*, 2002). When the fluorophore and quencher are in close proximity, as in the duplex, the fluorescence is quenched. When the structure melts, the fluorophore and quencher are separated and there is a large increase in fluorescence. These experiments are performed in the Roche LightCycler, which has an excitation source of 488 nm, with the fluorescence emissions measured at 520 nm.

By testing a series of different DNA duplexes with varying concentrations of ligand, the preferred binding site of the ligand may be established by observing which duplex is stabilised the most. The concentration dependence of the different melting profiles can be used to estimate the relative binding strength of different ligands.

Aims of this Thesis

This thesis describes studies on the interaction of several ligands with DNA. These studies aim to determine their preferred binding sites, and to understand the molecular basis of their selectivity. These studies have used DNase I and hydroxyl radical footprinting as well as fluorescence melting experiments.
These studies are divided into four main sections:

Chapter III: Novel universal footprint substrates
Although footprinting is a powerful technique for examining sequence selectivity, it is limited by the sequences that are available in any given DNA fragment. The work in this chapter develops novel DNA footprinting substrates that contain distinct combinations of two, three, four and six base pair sequences. These fragments are tested with several well characterised DNA binding agents.

Chapter IV: Novel analogues of the bis-intercalator TANDEM
DNase I footprinting studies, using several different DNA fragments, have been used to examine the binding and selectivity of several novel analogues of TANDEM.

Chapter V: Novel minor groove binding ligands
The sequence selectivity of derivatives of Hoechst 33258 and other novel polyamides are investigated using DNase I and hydroxyl radical footprinting. These new polyamides contain benzimidazole or pyridoimidazole rings in addition to the usual pyrrole and imidazole groups.

Chapter VI: Minor groove binding ligands containing isopropyl-thiazole units
The DNA binding properties of two novel derivatives of the synthetic ligand thiazotropsin, which contains isopropyl-thiazole ring systems, are examined by footprinting and fluorescence melting experiments.

Chapter II

Materials and Methods

Materials

Materials are grouped according to their usage. Many items were used for more than one protocol, so will be described in their first occurrence.

Cloning and dideoxy sequencing

Oligodeoxynucleotides with BamHI restriction sites for ligation into plasmids were either purchased from Eurogentec UK Ltd. (Romsey, UK) (DiN and TriN), or provided by Andrea Di Salvo (University College Dublin) (TANa/TANb) and Tom Brown (University of Southampton, School of Chemistry) (all other sequences) and stored at –20 °C in water. pUC18, pUC19 and Taq DNA Polymerase were bought from Sigma-Aldrich (Poole, UK). dNTPs (dGTP, dCTP, dATP, dTTP), T4 DNA ligase and BamHI were from Promega (Southampton, UK). Radioactive α-^{32}P-dATP was purchased from Amersham Biosciences (Little Chalfont, UK) with an initial activity of ~110 TBq/mmol (3000 Ci/mmol).

The T7-Dideoxy Sequencing Kit was bought from the USB Corporation (Cleveland, OH, U.S.A), whilst Sequagel (19:1 Acrylamide:Bisacrylamide solution containing 8 M urea) was purchased from National Diagnostics (Hull, UK). 3 MM paper and Saran wrap used when drying the gels were bought from Whatmann and GRI respectively.

Radiolabelling

HindIII, SacI, EcoRI, PstI, Acc65I and XbaI restriction enzymes were bought from Promega. α-^{32}P-dCTP and γ-^{32}P-ATP were purchased from Amersham Biosciences (Little Chalfont, UK) with an initial activity of ~110 TBq/mmol (3000 Ci/mmol). AMV Reverse Transcriptase was purchased from Sigma-Aldrich and Accugel (40 % (w/v) 19:1 Acrylamide:Bisacrylamide solution) was from National Diagnostics. T4 Polynucleotide kinase (PNK) was bought from New England Biolabs (Hitchin, UK) and X-ray film used to show the accurate location of the radiolabelled plasmid was from Kodak.

Ligands

Distamycin A, Hoechst 33258, mithramycin, actinomycin D and echinomycin were bought from Sigma-Aldrich, with nogalamycin being a gift from P.F. Wiley (Upjohn Company, Kalamazoo). TANDEM was a gift from Richard Olsen (Department of Chemistry, Utah State University) whilst solid-phase-TANDEM was a gift from Mark Searcey (School of

Pharmacy, University of London), along with Bis-acm-TANDEM, [Lys^4,Lys^8]-TANDEM, Benzylated-[Lys^4,Lys^8]-TANDEM, Naphthyl-TANDEM, Hemi-naphthyl-TANDEM, Biotinated-[Lys^4,Lys^8]-TANDEM, Bis-acetate-TANDEM and Mono-quinoxaline-TANDEM. Malvinder Singh (University of Saskatchewan) provided the Series A, B, C and D ligands, and Colin J. Suckling and Roger D. Waigh (University of Strathclyde) provided Thiazotropsin A, Thiazotropsin B and Thiazotropsin C.

DNase I and hydroxyl radical footprinting

Bovine DNase I was purchased from Sigma-Aldrich and stored at –20 °C at a concentration of 7200 U/ml in 150 mM NaCl and 1 mM $MgCl_2$.

Ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA), L-ascorbic acid and hydrogen peroxide (30 % (w/w) solution) were all bought from Sigma-Aldrich, whilst ammonium ferrous sulphate was purchased from BDH.

For both dideoxy sequencing and footprinting, a Phosphorimager Screen and the Storm 860 Phosphorimager from Molecular Dynamics (Amersham Pharmacia Biotech UK Limited, (Little Chalfont, UK)) were used.

Analysis of the footprinting gels was carried out using ImageQuant version 5.0 Build 050 software from Molecular Dynamics and SigmaPlot 9 for Windows.

Fluorescence melting

Oligodeoxynucleotides were provided by Prof. Tom Brown (University of Southampton, School of Chemistry) and stored at –20 °C in water. For each pair of oligodeoxynucleotides, one is 5'-fluorescein labelled, whilst the complimentary strand has a 3'-methyl red quencher moiety present. Fluorescence melting curves were determined in a Roche Lightcycler with an excitation source at 488 nm using Roche capillaries, taking fluorescence readings at 520 nm (Darby *et al.*, 2002).

Fluorescence analysis was carried out using the Roche Lightcycler software.

DNA Fragments

Various DNA fragments were designed and synthesised for this work, and some were used that were from previous work in the laboratory. These are described below.

MS1/MS2

MS1 and MS2 are fragments that contain every possible tetranucleotide sequence (Lavesa and Fox, 2001). Both fragments contain the same sequence, bar two point mutations shown in red below, cloned into the BamHI site of pUC18, but in opposite orientations. The sequences below show the strand that is seen when the fragments are labelled at the 3'-end of the HindIII site.

MS1

5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGC
GATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATG
TAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'

MS2

5'-GGATCCGCATTCGAGGCTGAGATGACAAACCAGACCCCACCGGACGTACTTTACA
TAACTCTTCACGCCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCTTGATCGC
GCTACGACTAGTGCAGTTGGAAATCGGCCATGTGTATTGCCGCATATGGATCC-3'

pAAD1

pAAD1 contains every self-complementary $(A/T)_4$ sequence separated by CGCG (Abu-Daya *et al.*, 1995). The fragment is cloned into the BamHI site of pUC18. The strand that is shown is the one that is labelled at the 3'-end of the HindIII site.

5'-GTACGCGTTAACGCGCGATATCGCGCGTAATCGCGCGTATACGCGCGAATTCGC-3'

P6

The P6 substrate is a sequence previously used in the laboratory that contains a series of long $(A/T)_{12}$ tracks. This is cloned into the BamHI site of pUC18.

5'-GGATCGGAAATAAATAAATCCGGTTTTTTAAAAAACCGGATATATATATATCC
GGAAAAAATTTTTTCCGGATCC-3'

_Tyr_T(43-59)

_Tyr_T(43-59) is a derivative of the _tyr_T promoter sequence of _E. coli_ (Brown and Fox, 1999). The _tyr_T sequence was used in the first footprinting experiments with TANDEM (Low _et al._, 1984a). The strand shown is the one that is labelled at the 3'-end of the EcoRI site.

```
5'-GGGAACCCCCACCACGGGGTAATGCTTTTTACTGGCCTGCTCCCTTCTCGGGAAG
CGGGGCGCTTCATATCAAATGACGCGCCGCTGTAAAGTGTTAGGAAGAGAAAAAAGAA
CTGGTTGCGTAATTTTCATCCGTAACGGATTAAAGGTAACCGGAATT-3'
```

DiN

DiN was prepared as a short synthetic sequence that contains three copies of each the 10 different dinucleotide steps. This was cloned into the BamHI site of pUC18.

```
5'-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3'
```

TriN

TriN was prepared as a short synthetic sequence that contains each of the 32 trinucleotide steps. This was cloned into the BamHI site of pUC18.

```
5'-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3'
```

HexA and HexB

HexA and HexB were designed as synthetic fragments that, between them, contain all 64 symmetrical hexanucleotide sequences. The sequences were designed so that HexA contains the sites for EcoRI and PstI, but not HindIII and SacI, while HexB contains the sites for HindIII and SacI, but not EcoRI and PstI. Each sequence was cloned in both orientations into the BamHI site of pUC19. During this work HexArev was only obtained as a dimer (HexArev2), while HexBrev could not be isolated. The opposite strand of HexBfor was therefore visualised by cleaving with a different combination of restriction enzymes (see Table 2.1). It was subsequently realised that these inserts are multiples of three base pairs and therefore still give blue colonies on Xgal/IPTG plates. HexAfor and HexBfor contain in frame stop codons and so yield white colonies. HexArev (monomer)

and HexBrev have now been successfully cloned but the footprinting results with these sequences are not included in this thesis.

HexAfor

5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAAT
TCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCT
AGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'

HexArev2

5'-GGATCCATGGCCACGTGCACTAGTACTTAAGCGCTAGCATGCGCAGCTGCA
GGCCTAGGTACCGGTTAACGTTTAAACCGCGGTCCGGAATTCTAGATCTATAGC
TAAATTTGGCGCCATATATCGATATCCCGGGATCCATGGCCACGTGCACTAGTA
CTTAAGCGCTAGCATGCGCAGCTGCAGGCCTAGGTACCGGTTAACGTTTAAACC
GCGGTCCGGAATTCTAGATCTATAGCTAAATTTGGCGCCATATATCGATATCCC
GGGATCC-3'

HexBfor

5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAG
CTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCG
ACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'

HexBrev

5'-GGATCCATGCATTAATTCGAATATTGATCATGACGTCGACATGTACATATG
TATATACGCGCGTACGCGTATACGTAGCGCGCTTATAAGCTTGCAATTGCCGGC
TAATTAGGGCCCTCGAGCTCGCGATCGGCCGGATCC-3'

STRATHA/STRATHB

STRATHA and STRATHB are complementary sequences containing several variations of the proposed target site for Thiazatropsin B [(A/T)CGCG(T/A)]. The original cloned sequence was STRATHB (cloned into the BamHI site of pUC19), with STRATHA being gained by labelling at the opposite end of the sequence. STRATHA is visualised when the DNA is labelled at the 3'-end of the Acc65I site, whilst STRATHB is visualised by labelling at the 3'-end of the HindIII site.

STRATHA

5'-GATGGATCCTTAATCCGCATAACCCCTTTACGCGAAATCGCCTAATCGGGATATCG
CGAATTCGCGCAAAACGCGTATAGCGTTAAGCGCAATACGCATACCGGATCCGAG-3'

STRATHB

5'-CACGGATCCGGTATGCGTATTGCGCTTAACGCTATACGCGTTTTGCGCGAATTCGC
GATATCCCGATTAGGCGATTTCGCGTAAAGGGGTTATGCGGATTAAGGATCCATC-3'

SASK1/SASK2

SASK1 and SASK2 (which is the same sequence cloned in the opposite orientation) were designed to contain variants of the proposed binding sites (four base pair G/C-tracts) for the Series B ligands (described in chapter 5). These were cloned into the BamHI site of pUC19. The strands shown are the ones visualised by labelling the 3'-end of the HindIII site.

SASK1

5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCC
ATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACC
GGTTGCCACGGATCC-3'

SASK2

5'-GGATCCGTGGCAACCGGTTGCAAGATCTTGCTTGGCCAAGCAAGGACTTGCAAGGC
CTTGCAAGGTCTTGCATGGCCATGCAAGGGCTTGCAACGCGTTGCATGGCCTAGCAAGCGC
TTGCTGGATCC-3'

SASK3/SASK4

SASK3 and SASK4 (which is the same sequence cloned in the opposite orientation) were designed to contain variants of the proposed binding sites (a long A/T-tract and $[CWWGWWC]_n$, where W is adenine or thymine) for the Series C ligands (described in chapter 5). These were cloned into the BamHI site of pUC19. The strands shown are the ones visualised by labelling the 3'-end of the HindIII site.

SASK 3

5′-GGTGGATCCAGCAAGCGCGCTTGCAACCCTATAGGGTTGCAAGTCTTGCAAATAT
ATATATTTGCAAGATCTTGCAAGGCTTGCGCCAAGCTTGCAAGGGTATACCCTTGCAA
GCTAGCTTGCAAGCGCTTGC-3′

SASK 4

5′-GCAAGCGCTTGCAAGCTAGCTTGCAAGGGTATACCCTTGCAAGCTTGGCGCAAGC
CTTGCAAGATCTTGCAAATATATATATTTGCAAGACTTGCAACCCTATAGGGTTGCAA
GCGCGCTTGCTGGATCCACC-3′

TANa/TANb

TANa and TANb are complementary sequences, which contain a single TATA site in the middle and were used for studying the binding of TANDEM and its analogues (chapter 4).

TANa

5′-CCACGTCGCTGACCACCTGCGCAGGTCCATATATGGCCAACTCGGTGCAT
CGCTCACTGGACACATCAGTCCATGAATGACTCGATGACTCAATGACTCG-3′

TANb

5′-TCATTCATGGACTGATGTGTCCAGTGAGCGATGCACCGAGTTGGCCATAT
ATGGACCTGCGCAGGTGGTCAGCGACGTGGGCATTGAGCGATGCAGGCAG-3′

Fluorescently labelled oligonucleotides

The oligonucleotides used in the fluorescence melting experiments are shown below. Targets 1-8 are described in James *et al.* (2004), whilst target 9 (containing the proposed binding site of Thiazotropsin B) was designed during this work.

	Name	Sequence
1	ACTAGT	5′-F-CCGACTAGTGC-3′ 3′-Q-GGCTGATCACG-5′
2	TGATCA	5′-F-CCGTGATCAGC-3′ 3′-Q-GGCACTAGTCG-5′
3	TGATCT	5'-F-CCGTGATCTGC-3' 3'-Q-GGCACTAGACG-5'
4	TGTACT	5'-F-CCGTGTACTGC-3' 3'-Q-GGCACATGACG-5'
5	TGAACT	5'-F-CCGTGAACTGC-3' 3'-Q-GGCACTTGACG-5'

	Name	Sequence
6	TCTAGT	5'-F-CCGTCTAGTGC-3' 3'-Q-GGCAGATCACG-5'
7	ATATAT	5'-F-CGCATATATGGC-3' 3'-Q-GCGTATATACCG-5'
8	AAAAAG	5'-F-CGCAAAAAGGC-3' 3'-Q-GCGTTTTTCCG-5'
9	ACGCGT	5'-F-CCGACGCGTGC-3' 3'-Q-GGCTGCGCACG-5'

Methods

The methods described are grouped according to the procedure in which they are used. Certain methods have been used in more than one section. In such cases, only the first use is described in detail.

Cloning

Annealing and amplification

For DiN and TriN both strands were synthesised and were mixed in equal amounts (1 µl of 100 µM of each) in a total volume of 15 µl (in sterile water containing 1 x 10 x Taq polymerase buffer). These were annealed by heating to ~95 °C in a water bath then cooling slowly.

The sequences for cloning SASK1, SASK3, STRATHA, HexA and HexB were made as single stranded templates, which were amplified by PCR using reverse and forward primers. The polymerase chain reaction comprised of 0.5 µM of each primer, 0.5 nM template, 1 x 10 x Taq polymerase buffer, 1 mM $MgCl_2$, 500 µM dNTP mix (dATP, dGTP, dCTP, dTTP) and 1 µl Taq polymerase (5 u/µl) in a final solution volume of 100 µl (made up with 80 µl sterile water), using the protocol in Figure 2.1.

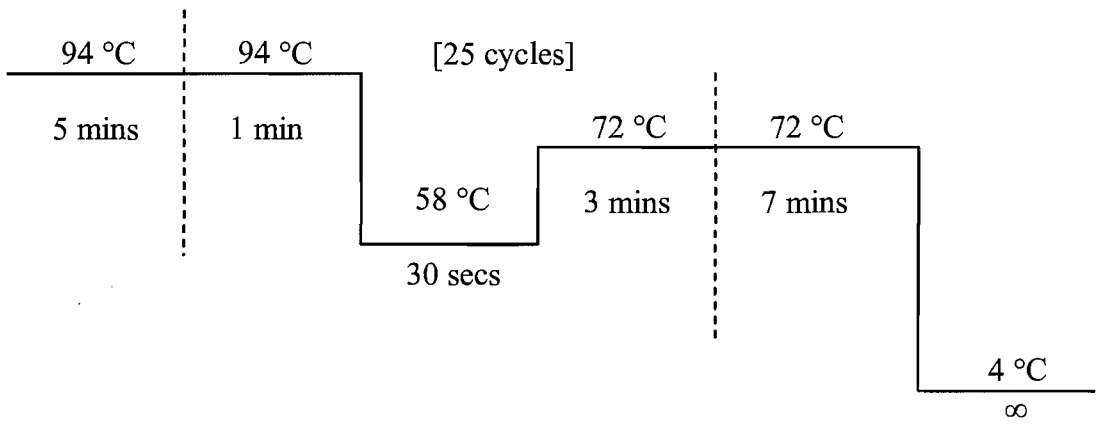


Figure 2.1: PCR protocol for amplification of DNA substrates for footprinting studies.

A 2 % agarose gel (containing 5 μg/ml ethidium bromide) was run to elute the amplified DNA from the reaction components and the desired PCR product was trapped in a well that had been cut into the agarose gel. The isolated DNA fragment was concentrated by ethanol precipitation. This procedure was found to be necessary in order to inactivate and remove the Taq polymerase, which would otherwise fill in the sticky ends at later stages in the cloning procedure. The dried DNA pellet was resuspended in 40 μl sterile water. 20 μl of this was then added to 20 μl sterile water, 4 μl Buffer E (6 mM Tris-HCl, 6 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, pH 7.5 at 37 °C) and 1 μl BamHI (10 u/μl) before incubating at 37 °C for 1 hour to generate "sticky ends" ready for ligation into the plasmid. The BamHI was removed by ethanol precipitation and the DNA pellet resuspended in 17 μl sterile water.

Ethanol precipitation

A volume of DNA solution was combined with an amount of 3 M sodium acetate equivalent to 1/9 of the volume of DNA solution, and 3 volumes of ethanol. The solution was then left on dry ice for 30 minutes before spinning for 10 minutes at 13000 rpm, discarding the supernatant. The DNA pellet was then washed with 100 μl 80 % ethanol and spun at 13000 rpm for a further 2 minutes, discarding the supernatant and drying in a SpeedVac for 5 minutes.

Ligation of insert into plasmid

5 μl of the annealed, amplified and BamHI-cut oligonucleotides were combined with 10 μl BamHI-cut pUC18 (or pUC19) plasmid (DiN and TriN were cloned into pUC18, the rest were cloned into pUC19), to which 2 μl 10x T4 DNA ligase buffer (300 mM Tris-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP) and 2 μl T4 DNA ligase were added. This was incubated for at least 3 hours at room temperature to ensure ligation of the insert into the pUC18 or pUC19.

BamHI cleavage of pUC18/pUC19

1 μl pUC18 or pUC19 (237 μg/ml) in a total volume of 18 μl (sterile water) was combined with 2 μl 10 x multicore buffer and 1 μl BamHI before incubation at 37 °C for 1 hour. The cleavage and was stopped by ethanol precipitation and the DNA pellet resuspended in 10 μl sterile water.

Transformation into competent cells

10 µl of ligation mixture was added to 200 µl competent TG2 cells and incubated on ice for 40 minutes. The cells were then heat shocked at 45 °C for 1 minute before returning to ice. The transformed cells were plated onto agar plates containing 100 µg/ml carbenicillin, 0.02 % X-gal, 1 mM IPTG and grown overnight at 37 °C.

Competent cell preparation ($CaCl_2$ method)

A 5 ml 2YT (16 g Tryptone, 10 g Yeast Extract, 5 g NaCl per litre) culture was innoculated with a single colony of TG2 (a strain of *E.*coli K12 for blue/white selection) from an agar plate. This was then incubated at 37 °C overnight, shaking at 350 rpm. 500 µl of this liquid culture was then added to 50 ml 2YT media and incubated as before until it reached an A_{600} of 0.5 - 0.8. The cells were then harvested by centrifugation at 5000 rpm for 5 minutes (at 4 °C) and the resultant pellet resuspended in 20 ml sterile, ice cold, transformation buffer (50 mM $CaCl_2$, 10 mM Tris-HCl at pH 7.4). After incubation on ice for 30 minutes, the cells were again centrifuged at 5000 rpm for 5 minutes before being resuspended in 5 ml transformation buffer and stored at 4 °C.

IPTG/X-gal agar plates

100 ml sterile 39.5 g/l blood-based agar was melted and cooled to ~50 °C. 100 µl 100 mg/ml carbenicillin, 1 ml 100 mM IPTG (in water) and 1 ml X-gal (20 mg/ml in dimethylformamide) were added before immediately pouring into petri-dishes and leaving to set. The set plates were stored at 4 °C.

Screening of colonies

The pUC18 and pUC19 plasmids contain a carbenicillin-resistance gene, so ensuring that only TG2 cells carrying the plasmid grow up on the agar plates. There is also a LacZ gene incorporated in the plasmid containing the polylinker multiple cloning site, which, if expressed, complements the partial gene product present in the TG2 cells to produce active β-galactosidase. In the presence of its inducer, IPTG, this enzyme will degrade the chromogenic substrate X-gal, producing a blue colouration. However, if an insert is present within the polylinker multiple cloning site, the LacZ gene will be disrupted, so producing colonies of TG2 cells without blue colouration. Recombinant TG2 cells were therefore

selected by aseptically picking off white colonies. A selection of white colonies were then grown up in 5 ml 2YT media, containing 10 µl 100 mg/ml carbenicillin, overnight at 37 °C, shaking at 350 rpm. 30 % glycerol stocks were made from the liquid cultures (600 µl culture to 400 µl 50 % glycerol), while the rest of the broth was divided between two 1.5 ml microcentrifuge tubes and centrifuged at 5000 rpm for 7 minutes. The pellets produced were purified using a QIAgen miniprep kit.

QIAprep spin miniprep (using a microcentrifuge)

Plasmid was prepared from 5 ml overnight culture according to the manufacturer's instructions. The pelleted cells were resuspended in 250 µl Buffer P1 and lysed by gently mixing with 250 µl Buffer P2. 350 µl Buffer N3 was then added (mixing gently immediately on addition) before the microcentrifuge tubes were centrifuged at 13000 rpm for 10 minutes. The supernatant was transferred into QIAprep Spin Columns and centrifuged at 13000 rpm for a further 60 seconds. The flow-through was discarded and 0.5 ml Buffer PB added to wash the column. After spinning for 60 seconds at 13000 rpm (discarding the flow-through on completion), 0.75 ml Buffer PE was added to the column (again to wash). The columns were again centrifuged for 60 seconds at 13000 rpm, and once the flow-through had been discarded, the spin was repeated to remove any residual wash buffer. The columns were then placed in clean microcentrifuge tubes and the DNA was eluted by adding 50 µl Buffer EB, first letting stand for 60 seconds, then centrifuging at 13000 rpm for 60 seconds.

Dideoxy Sequencing

Dideoxy sequencing was performed with a T7-dideoxy sequencing kit (Pharmacia) according to the manufacturer's instructions. 40 µl of the purified DNA were transferred to clean microcentrifuge tubes and 10 µl 2 M NaOH added. After leaving at room temperature for 10 minutes, the DNA was precipated with ethanol by adding 15 µl 3 M sodium acetate pH 4.8, 35 µl sterile water and 300 µl 100 % ethanol, followed by washing with 70 % ethanol. The pellet produced was resuspended in 10 µl sterile water.

To each tube was added 2 µl annealing buffer and 2 µl universal primer, both from the T7-Dideoxy Sequencing Kit. The samples were then incubated first at 37 °C for 20 minutes, then at room temperature for at least a further 10 minutes.

A polymerase mix (12 µl Label Mix A (from the T7-Dideoxy Sequencing Kit), 7 µl sterile water, 1 µl α-^{32}P-dATP) and enzyme / buffer mix (6.5 µl enzyme dilution Buffer, 1.5 µl T7 Polymerase (both from the T7-Dideoxy Sequencing Kit)) were made on ice before mixing together. 6 µl of this final mixture was added to each microcentrifuge tube and incubated at room temperature for 5 minutes. Meanwhile, 2.5 µl each of ddGTP-short mix, ddCTP-short mix, ddATP-short mix and ddTTP-short mix (a set for each sample, all coming from the T7-Dideoxy Sequencing Kit) were incubated at 37 °C for 5 minutes in clean tubes.

4.5 µl of the radioactive sample was then added to each ddNTP-short mix before incubating at 37 °C for a further 5 minutes to let polymerisation occur. The reaction was halted by adding 5 µl Stop solution (from the T7-Dideoxy Sequencing Kit).

The DNA was denatured by boiling for 2 minutes then crash-cooling on ice and run on a 9 % denaturing electrophoresis gel (18 ml Sequagel, 5 ml 10 x buffer (216 g Tris, 110 g Boric acid, 18.8 g EDTA, 1 kg urea in 2 l hot water), 27 ml diluent (1 kg urea in 2 l hot water), 200 µl 20 % (w / v) ammonium persulphate and 40 µl N,N,N',N'-tetramethylethylenediamine (TEMED)) at 1500 V, 42 W using 1 x TBE running buffer until the first (dark blue) band reached the bottom of the gel (about 1.5 hours).

The glass plates were then separated and the gel fixed by soaking in 10 % (v / v) acetic acid for 10 - 15 minutes. The gel was then transferred to Whatmann 3 MM paper and dried under vacuum at 90 °C for 1 hour. The dried gel was exposed to a phosphorimager screen for 20 minutes before being scanned at 100 nm resolution. The sequence for each sample was then read off and checked.

Changing the insert orientation in the plasmid

All the synthetic DNA inserts used in this work had BamHI sites at both ends and could therefore be cloned in either orientation. To obtain the opposite orientation (useful for visualising the other strand in footprinting studies), 1 µl of the DNA was added to 1 µl BamHI, 2 µl Buffer E and 16 µl sterile water and incubated at 37 °C for 1 hour (to excise the insert). Re-ligation was then carried out in the same way as normal ligation after the DNA had been precipated with ethanol. After transformation (which produced mainly white colonies with a few blue colonies resulting from religation without the insert), a selection of the white colonies were then grown and sequenced as before. All the longer DNA substrates were therefore available for footprinting in both orientations.

Radiolabelling

α-^{32}P-dATP / α-^{32}P-dCTP radiolabelling

To 3'-radiolabel the DNA fragment, the purified plasmid DNA from a 5 ml culture (in a volume of 50 µl) was cleaved with two restriction enzymes (1.5 µl each of 10 u/µl stocks) as shown in Table 2.1 in the presence of 5 µl 10 x multi-core buffer or other suitable buffer.

Substrate	Labelling Enzyme	2nd Enzyme
MS1 / MS2	HindIII	SacI
pAAD1	HindIII	SacI
TyrT(43-59)	EcoRI	AvaI
DiN	HindIII	SacI
TriN	HindIII	SacI
HexAfor / HexArev2	HindIII	SacI
HexBfor	EcoRI	PstI
HexBrev (using HexBfor clone)	XbaI	EcoRI
SASK1 / SASK2	HindIII	SacI
SASK3 / SASK4	EcoRI	PstI
STRATHB	HindIII	SacI
STRATHA (using STRATHB clone)	Acc65I	PstI

Table 2.1: Enzymes used to 3'-label each DNA substrate.

After incubation at 37 °C for 1.5 hours, 1.2 µl α-^{32}P-dATP was added to the DNA (using suitable safety precautions), except for HexBrev, for which α-^{32}P-dCTP was used to enable visualisation of the opposite strand. Unlabelled dTTP and dGTP were also added to the STRATHA mixture to fill in the 3'-sticky end of the DNA. The DNA was then incubated with 0.6 µl AMV reverse transcriptase for 1 hour at 37 °C.

20 µl 20 % Ficoll solution (including 10 mM EDTA and 0.1 % bromophenol blue in sterile water) was added to the radiolabelled DNA before separating from the remainder of the plasmid and unbound dATP / dCTP on a non-denaturing polyacrylamide gel

(typically 5.6% - 7 ml Accugel, 10 ml 5 x TBE (108 g Tris, 55 g boric acid, 9.4 g EDTA in 2 l sterile water), 33 ml sterile water, 200 µl 20 % (w / v) ammonium persulphate and 40 µl N,N,N',N'-tetramethylethylenediamine (TEMED)). The gel was run at 800 V, 20 W using 1 x TBE running buffer for about 1.5 hours, until the blue dye reached about three quarters down the gel.

On completion of electrophoresis, the position of the labelled DNA fragment was established by exposing an X-ray film to the gel for about 5 minutes. The relevant band was then cut from the gel and the DNA eluted by adding 300 µl Tris-EDTA (10 mM Tris-HCl at pH 7.5, 10 mM EDTA) and shaking at 350 rpm overnight at room temperature. The eluted DNA was then precipitated with ethanol and resuspended in a suitable volume of Tris-EDTA (10 mM Tris-HCl at pH 7.5, 0.1 mM EDTA) so as to give a reading of about 10 counts per second/µl on a hand-held Geiger counter.

γ-^{32}P-ATP radiolabelling

The TANa and TANb DNA substrates were 5'-radiolabelled using γ-^{32}P-ATP. 2 µl of 2 µM DNA was added to 14 µl sterile water, along with 1 µl T4 polynucleotide kinase (PNK) and 2 µl 10 x PNK buffer. 1 µl γ-^{32}P-ATP was also added before the mixture was incubated at 37 °C for over an hour. After incubation, 10 µl DNase I stop solution (10 mM EDTA, 1 mM NaOH, 0.1 % bromophenol blue, 80 % formamide) was added and the mixture heated at 100 °C for 3 minutes before crash cooling on ice. The labelled sample was then separated from unbound ATP by electrophoresis on a 8 % denaturing gel (16 ml Sequagel, 5 ml 10 x buffer (216 g Tris, 110 g boric acid, 18.8 g EDTA, 1 kg urea in 2 l hot water), 27 ml diluent (1 kg urea in 2 l hot water), 200 µl 20 % (w / v) ammonium persulphate and 40 µl N,N,N',N'-tetramethylethylenediamine (TEMED)) at 1500 V, 42 W using 1 x TBE running buffer until the dye band reached about three quarters down the gel. The radiolabelled DNA was located and eluted from the gel as with 3'-labelling. 2 µl of the unlabelled complementary strand (either TANb or TANa) was then added to the labelled DNA and annealed by heating to ~95 °C in a water bath and cooling slowly. This product can then be directly used as a footprinting substrate, but for clearer results the labelled duplex is separated from excess complementary strand by non-denaturing electrophoresis in the same way as 3'-labelling.

Footprinting

DNase I footprinting

1.5 µl radiolabelled DNA was mixed with 1.5 µl ligand solution (dissolved in 10 mM Tris-HCl at pH 7.5, 10 mM NaCl at differing concentrations depending on the experiment) and incubated at room temperature for 30 minutes. A GA track marker was also prepared. The mixture was then digested with 2 µl DNase I (typically 0.01 units/ml dissolved in 20 mM NaCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$) for exactly 1 minute before being quenched with 4 µl DNase I stop solution. The DNA was then denatured by incubating at 100 °C for 3 minutes and crash-cooling on ice before running on a denaturing polyacrylamide gel (typically 8% (w/v) composed of 16 ml Sequagel, 5 ml 10 x Buffer (216 g Tris, 110 g boric acid, 18.8 g EDTA, 1 kg urea in 2 l hot water), 27 ml Diluent (1 kg urea in 2 l hot water), 200 µl 20 % (w/v) ammonium persulphate and 40 µl N,N,N',N'-tetramethylethylenediamine (TEMED)) at 1500 V, 42 W using 1 x TBE running buffer until the dye band reached the bottom of the gel (about 1.5 hours).

The gel plates were then separated and the gel fixed and dried as for sequencing. The dried gel was then exposed to a phosphorimager screen overnight before scanning at 100 nm resolution.

The range of ligand concentrations tested was based on the results of broad-range preliminary experiments.

Charged ligands will often stick to the microcentrifuge tubes, so a lower concentration than wanted will be transferred to the DNA when mixed for incubation. When this is seen to be having a significant effect on apparent binding strength and is making selectivity of binding unclear, siliconised tubes are used throughout the above protocol. This obviously creates a marked increase in apparent strength of binding for the ligand, so results obtained using this method should only be compared with similarly sourced data. This adaptation was used when studying the binding of Thiazotropsin B.

GA track marker

A GA track marker shows the location of purines within the labelled DNA sequence, thereby allowing the identification footprinting sites in relation to the DNA sequence. The GA track marker was prepared by mixing 1.5 µl labelled DNA with 20 µl sterile water and

4 µl DNase I stop solution (10 mM EDTA, 1 mM NaOH, 0.1 % bromophenol blue, 80 % formamide). The sample was then incubated at 100 °C for about 40 minutes with the microcentrifuge cap open to allow evaporation. The marker was then run alongside the ligand lanes.

Hydroxyl radical footprinting

2 µl radiolabelled DNA was mixed with 10 µl ligand solution (dissolved in 10 mM Tris-HCl at pH 7.5, 10 mM NaCl at differing concentrations depending on the experiment) and incubated at room temperature for 30 minutes. A GA track marker was also prepared as before.

Once the 30 minute incubation was complete, the mixture was digested with 10 µl hydroxyl radical mix (4 µl 100 mM ammonium ferrous sulphate in 1 ml sterile water, 5 µl 0.5 M EDTA in 1 ml sterile water, 100 µl 100 mM ascorbic acid in 1 ml sterile water, 10 µl hydrogen peroxide in 1 ml sterile water mixed in a 1:1:2:2 ratio respectively) for 15 minutes before precipitating with ethanol as usual. The pellets produced were resuspended in 8 µl DNase I stop solution. The DNA was then denatured by incubating at 100 °C for 3 minutes and crash-cooling on ice before running in the same fashion as DNase I footprinting.

The range of ligand concentrations tested was based on the results of DNase I footprinting experiments.

Quantitative analysis

Once the phosphorimager screen had been scanned on a Storm 860 Phosphorimager, there are three ways to use ImageQuant to analyse the gels.

Footprinting plots – C_{50}

Quantitative footprinting analysis provides a means for estimating the dissociation constant of the ligand at each binding site. This is achieved by comparing the intensity of bands within the footprint at different ligand concentrations. Quantitative estimates of the binding affinities can be obtained by measuring the intensity of the bands in each footprint. These are normalised (to account for differences in gel loading or cleavage) either by dividing by

the intensity of a band which is not affected by the ligand or by the total intensity of all bands in each lane. The "intensity" refers to the radioactivity contained within the volume of each band, corrected for any background, although it is usually easier to combine the intensities of all the bands within a footprint to examine the effect of the ligand on the region of DNA. This can provide more accurate data, but assumes that the footprint only contains a single ligand binding site.

After calculating the relative intensity of the footprint at different ligand concentrations these are combined to produce a "footprinting plot" which is fitted by the simple binding equation $I/I_0=C_{50}/(L+C_{50})$, where I and I_0 are the relative band intensities in the presence and absence of the ligand respectively and L is the ligand concentration (Dabrowiak and Goodisman, 1989). If the DNA concentration is lower than the dissociation constant then C_{50}, then the ligand concentration, at which the intensity of the bands in the footprint is reduced by 50 %, approximates to the dissociation constant of the ligand from the DNA (K_d). Since the radiolabelled DNA concentration is usually less than 10 nM this condition is fulfilled for ligands with K_d values of 100 nM or weaker.

The analysis assumes that DNase I cleavage is zero when the ligand binding site is fully occupied (i.e. that I/I_0 tends to zero at high ligand concentrations). This is not always observed. The residual band intensity could be because the ligand-DNA complex is still cleaved (albeit less well than the control), although this is unlikely with minor groove binding ligands. The residual cleavage can also be caused by impurities in the DNA preparation, which yield bands that are not a direct result of DNase I cleavage. We therefore often allow for this residual intensity, by fitting the data to the equation $I/I_0=C_{50}/(L+C_{50}) + R$, where R is the residual cleavage (which must be positive).

Densitometer plots

A useful tool to determine where cleavage protection has occurred by a ligand, especially when using hydroxyl radicals (due to the usually more even cleavage pattern), is the densitometry plot. The plot is a simple graph of band intensity against DNA sequence, where a plot for the control lane can be compared with a ligand-treated lane, with footprints showing as clear dips in the plot. More precise binding sites can be established in this way compared to mere visual observation.

Differential cleavage plots

If the cleavage pattern of the control lane is uneven, then densitometry plots can be misleading or unclear. In these cases, a simple means to identify ligand binding sites along a DNA fragment is to construct a differential cleavage plot (Fox and Waring, 2001). These plots represent the cleavage of each bond in the ligand-treated lanes, compared with that in the ligand-free control lanes. The differential cleavage of each bond is then given by the log of the difference between the bands of the control and ligand lanes (log(control band intensity/ligand band intensity)). The analysis assumes single-hit kinetics as multiple cleavage events will cause an overrepresentation of bands that are close to the label (near the bottom of the gel).

Fluorescence Melting

These were prepared using the same conditions as described in James *et al.* (2004). Lightcycler capillaries contained a total volume 20 µl (5 µl fluorescently-labelled DNA strand at 1 µM, 5 µl quencher-labelled DNA strand at 1 µM, 5 µl ligand at 4 x final concentration (varies between experiments, replaced by buffer for controls), 5 µl buffer (10 mM Na Phosphate, 20 mM NaCl at pH 7.4)) were run in a Roche LightCycler. The samples were typically denatured by heating at 1 °C/min to 95 °C, held for 5 minutes, then annealled by cooling to 30 °C at 1 °C/min, holding for a further 5 minutes, before heating back to 95 °C as before. Fluorescence readings were taken at each heating / cooling step.

The range of ligand concentrations tested was based on the results of footprinting experiments.

Quantitative analysis of melting data

Melting temperatures (T_m) were determined from the maximum in the first derivative of the melting and annealing profiles using the Roche LightCycler programme. These data showed the relative change in stability of the duplex caused by ligand-DNA binding and so is an indicator of relative binding affinity of the ligand to the duplex.

All footprinting and fluorescence melting experiments were repeated to ensure the results were reproducible.

Chapter III

Novel Universal Footprinting Substrates

Introduction

As exemplified in the rest of this thesis, footprinting is a powerful technique for assessing the sequence specificity of DNA-binding. For some novel DNA binding ligands the investigator will have no prior knowledge of the sequence specificity, while for others the primary binding site might be predicted, but without knowledge of the likely secondary binding sites. The ligands examined in the other chapters in this thesis have been tested on a range of footprinting substrates but in some instances conclusive binding rules are still hard to establish. DNA footprinting substrates are therefore required that contain as many different binding sites as possible for any ligand (so called "universal sequences"). This will also allow identification of the preferred binding sites as well as providing a direct comparison of related (secondary) binding sites, that may differ by one or more base pairs, or that are flanked by different neighbouring bases.

For ligands that possess pronounced sequence selectivity, the problem of finding the exact sequence within any given fragment becomes greater. There are $4^n/2$ n-mer sequences if n is odd and $(4^n+4^{n/2})/2$ if n is even (Lavesa and Fox, 2001). There are therefore 10 different dinucleotides, 32 trinucleotides, 136 tetranucleotides, 512 pentanucleotides and 2080 hexanucleotides. Since a standard 35 cm electrophoresis gel used for footprinting studies can only resolve 100-180 base pairs, most footprinting substrates are 100-200 base pairs long. It is therefore clear that as the length of the preferred target site of a ligand is increased, there is a greater chance that it will not be represented in any given footprinting substrate. We therefore need a series of footprinting substrates that contain as many overlapping combinations of different sequences as possible. These sequences should be combined so as to generate the shortest possible fragment.

Researchers in the School of Mathematics at the University of Southampton have produced algorithms for calculating the minimal length DNA fragments that are required to contain all possible n-mer sequences (Anderson *et al.*, 2006). Two hypotheses have been proposed.

The n-1 hypothesis

The n-1 hypothesis states that the minimum length (L) of a universal fragment must be equal to the number of possible n-mers plus (n-1). This allows for the overlapping of adjacent n-mers, *e.g.* three optimally overlapped trinucleotides require a 5-mer length of

DNA to account for the non-overlap at the ends of the linear fragment (3 + (3-1) = 5). In this way, two equations are derived: One calculates the minimum length of a DNA fragment that contains all the possible n-mers when n is odd (*e.g.* trinucleotides) (Equation 3.1a); and one to calculate the minimum fragment length when n is even (*e.g.* dinucleotides) (Equation 3.1b).

$$\text{a) Odd } n = \frac{4^n}{2} + n - 1 \qquad \text{b) Even } n = \frac{4^n + 4^{n/2}}{2} + n - 1$$

Equation 3.1: The minimum number of base pairs in a universal sequence containing all possible n-mers for: a) n is odd; b) n is even.

The minimum length for a universal fragment that contains all possible dinucleotide sequence is therefore 11 base pairs, 34 base pairs for every possible trinucleotide and 139 base pairs for every possible tetranucleotide.

However, it was discovered that if n is greater than 2 and even, then the n-1 rule is not correct and longer fragments are necessary to contain all the possible n-mer sequences (Anderson *et al.*, 2006).

The n>2 and even hypothesis

The limitation with the n-1 rule is that with n-mers for which n is greater than 2 and even, self-complementary n-mers create what is referred to as "lollipop loops" (Anderson *et al.*, 2006), when combining overlapping sequences. These lollipop loops occur where the sequence path has come to a dead-end and can only be continued by repeating at least one of the sequences. This is analogous to the Koenigsberg Bridges problem (Figure 3.1a), a mathematical conundrum. If each bridge represents a point where the DNA sequence must pass to create a minimal length universal sequence (as the adjacent n-mers are overlapped), then at least one bridge must be passed over twice to involve every bridge in the final sequence. These "lollipop loops" only occur with even numbered n-mers and arise because some sequences are self-complementary (Anderson *et al.*, 2006). As a result of this, Anderson *et al.* (2006) showed that the shortest fragment containing every possible tetranucleotide must be 144 base pairs long. This explains why random computer searches failed to detect shorter sequences (Kwan *et al.*, 2003).

a) b)

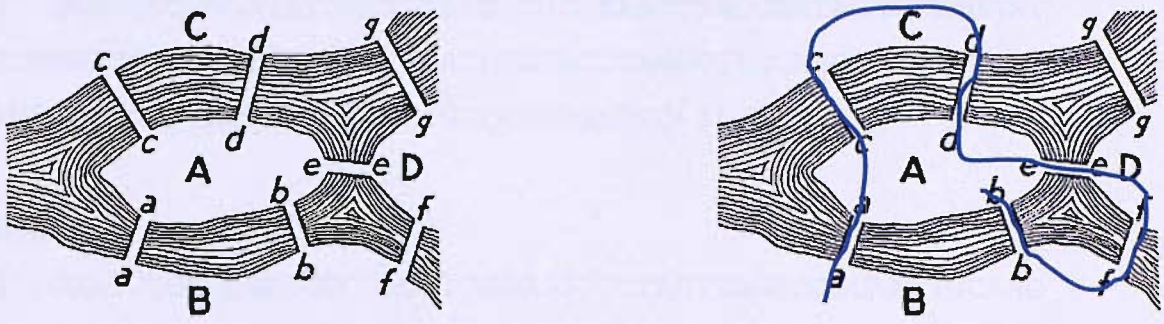


Figure 3.1: The Koenigsberg Bridges problem: a) The 7 bridges and 4 islands; b) A hypothetical route across the bridges without going back on ones self, showing that it is impossible to do so and go over all of the bridges.

Previous studies have generated a universal footprinting fragment (MS1/MS2) that contains every possible tetranucleotide sequence (Lavesa and Fox, 2001), which has been widely used in footprinting experiments (Malkinson *et al.*, 2005; Martin *et al.*, 2005).

The work presented in this chapter describes the preparation of footprinting substrates that contain dinucleotides, trinucleotides and hexanucleotides and their use in footprinting experiments, along with MS1/MS2. DiN is a fragment that contains three copies of every dinucleotide sequence; TriN contains every trinucleotide sequence. It is clearly not possible to design fragments that contain all 2080 hexanucleotide sequences (though this might be achieved using a panel of appropriate fragments), so, since many DNA binding ligands are (pseudo)symmetrical, two novel footprinting substrates (HexA and HexB) have been prepared that between them contain all 64 different symmetrical hexanucleotides. As far as possible the design of these fragments avoided stretches of DNA that are poor substrates for DNase I, such as G/C-rich regions and $A_n.T_n$ tracts. The four novel DNA fragment sequences are shown in Figure 3.2.

DiN

5′-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3′

TriN

5′-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3′

HexA

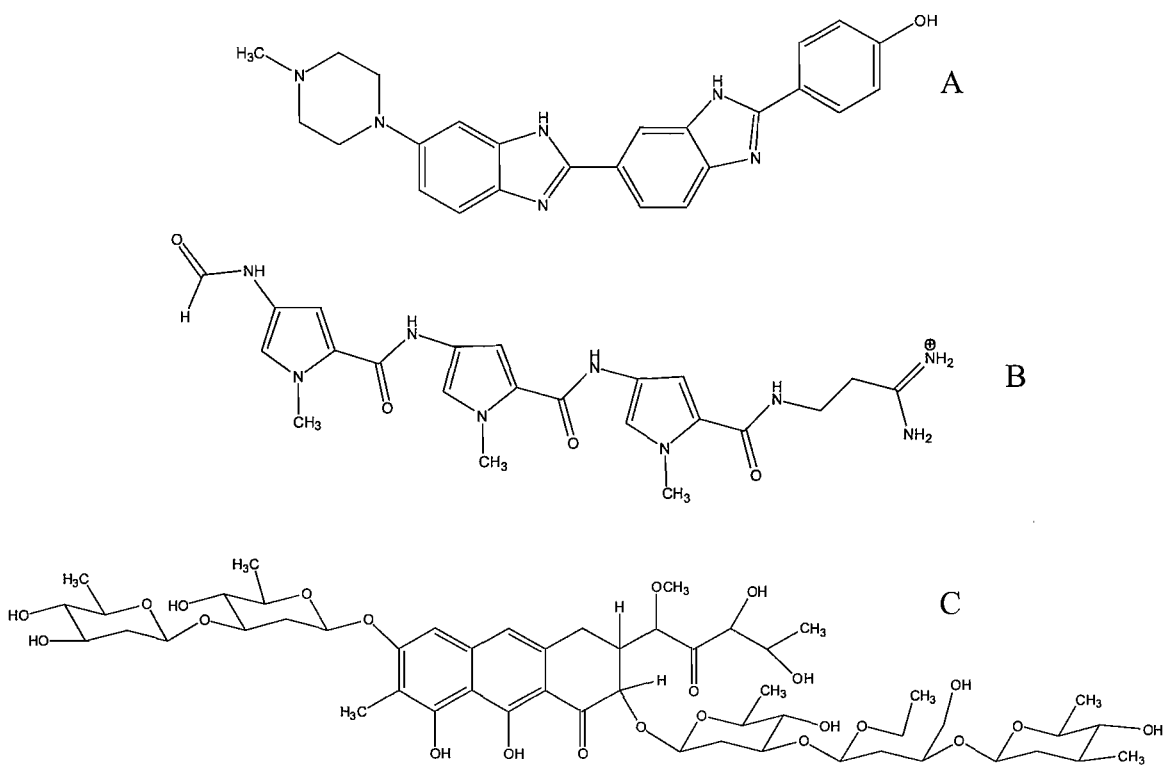
5′-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAAT
TCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCT
AGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3′

HexB

5′-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAG
CTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCG
ACGTCATGATCAATATTCGAATTAATGCATGGATCC-3′

Figure 3.2: Sequences of the four novel DNA fragments used in this chapter. These sequences were each cloned into the BamHI site of pUC18 or pUC19 and footprinting fragments containing these sequences were generated by cleaving with appropriate restriction enzymes as described in chapter 2 (Materials and Methods).

These DNA fragments were used in DNase I footprinting experiments with the well-characterised ligands: Distamycin; Hoechst 33258; mithramycin; nogalamycin; actinomycin D; and echinomycin (Figure 3.3). Studies with these ligands were particularly useful as their primary preferred binding targets have been well characterised.



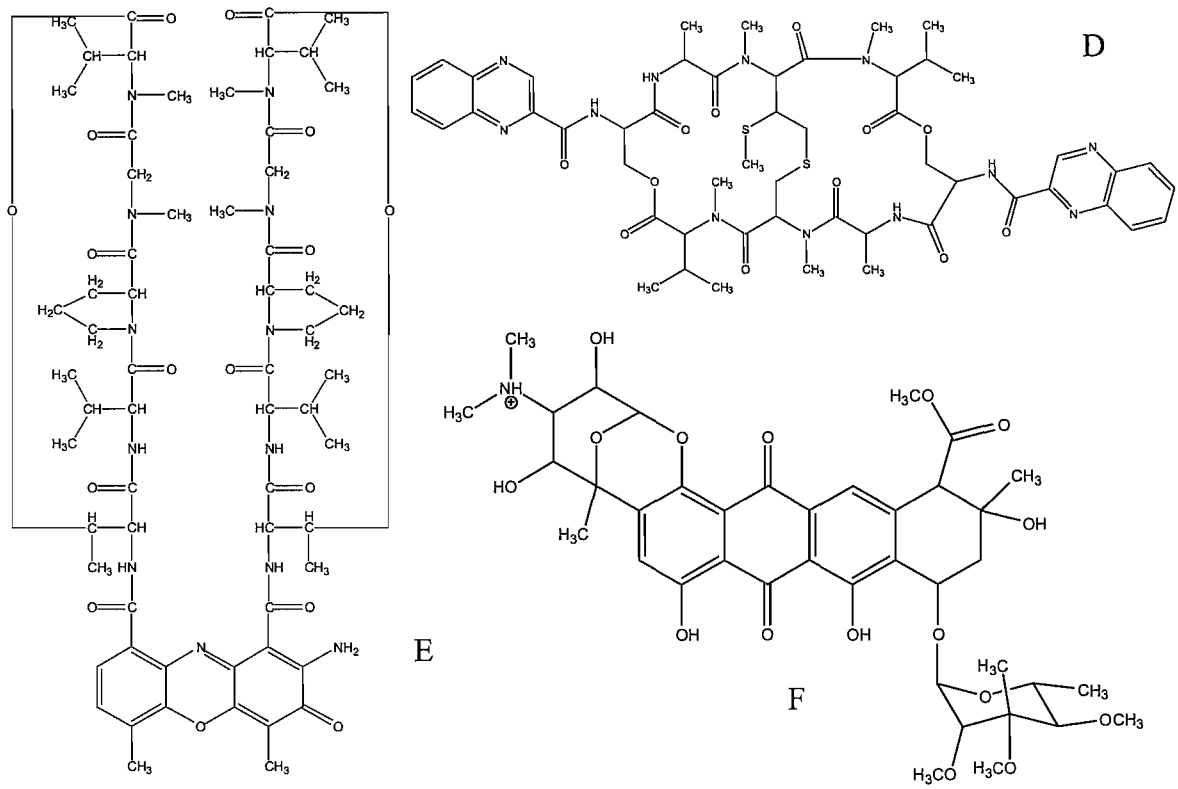


Figure 3.3: Structures of Hoechst 33258 (A); distamycin (B); mithramycin (C); echinomycin (D); actinomycin D (E); and nogalamycin (F).

Results

Design, cloning and sequencing of the universal substrates

The sequences of the oligonucleotides used to prepare the universal footprinting substrates are shown in Figure 3.2. They were each cloned into the BamHI site of pUC18 (or pUC19) as described in the Materials and Methods section (chapter 2). For DiN and TriN both strands were synthesised and annealed before cloning. For both HexA and HexB only one strand was synthesised; a primer was annealed to the 3'-end and extended with Taq polymerase to produce the full length duplex. Successful clones were picked as white colonies from agar plates containing carbenicillin, IPTG and X-gal. These were subjected to dideoxy sequencing and the sequencing gels are presented in Figure 3.4.

DiN and TriN were easily obtained as monomeric clones of these sequences. HexAfor and HexBfor were also obtained as monomers. Since these sequences are each longer than the fragments used in most footprinting experiments it was desirable to obtain clones in which these sequences had been inserted in each direction (as for MS1 and MS2) so as to be able to view each end of the sequence independently. I attempted to invert the

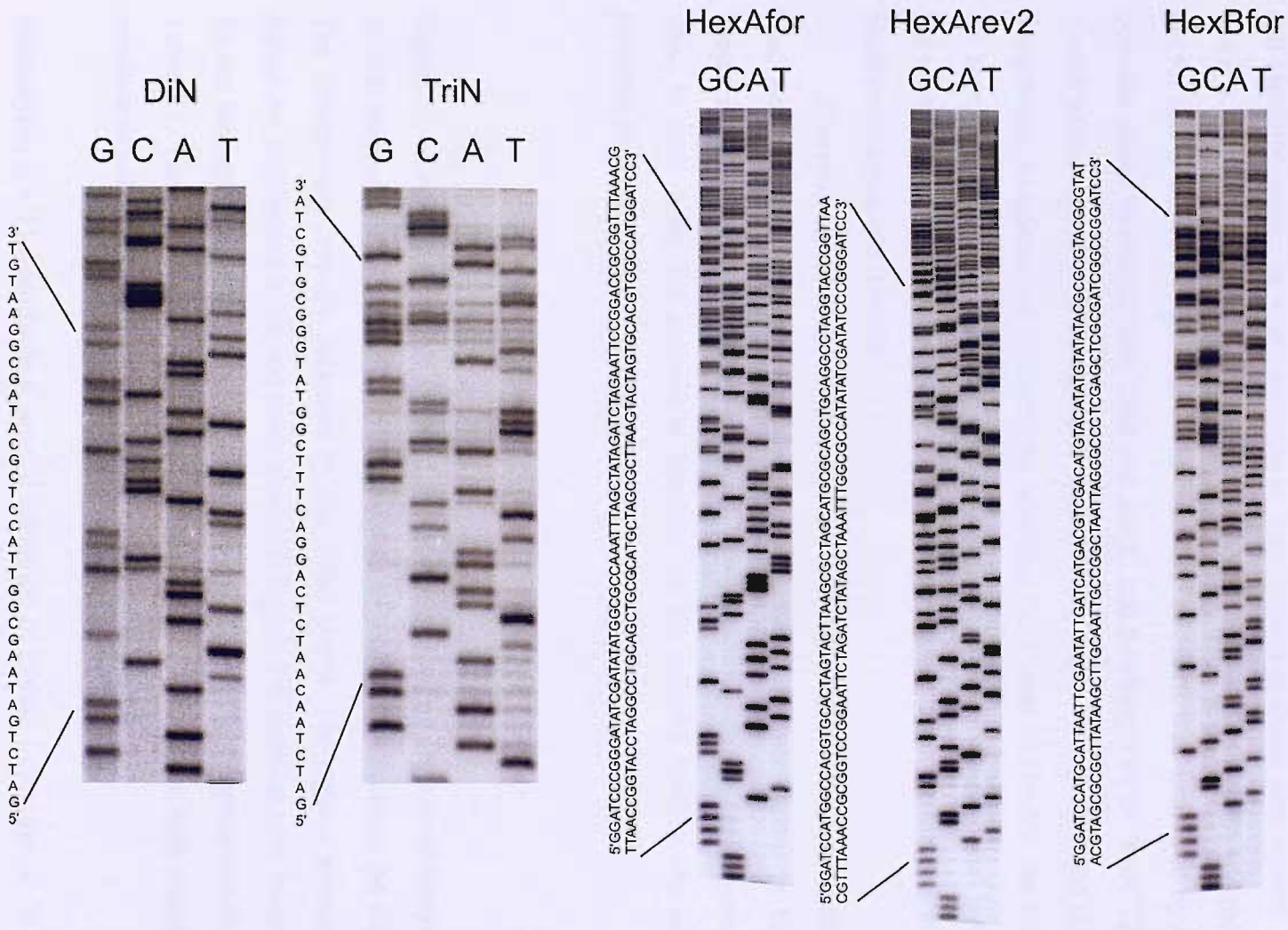


Figure 3.4: Dideoxy sequencing gels of DiN, TriN, HexAfor, HexArev2 and HexBfor. The sequence is shown alongside each gel. For HexArev2 only the first part of the dimeric insert is indicated.

insert by cutting HexAfor and HexBfor with BamHI and religating. This procedure consistently failed to produce inverted monomeric clones. After completion of this work it was realised these inserts are multiples of three base pairs and so did not completely inactivate the LacZ gene, and successful clones appeared as (light) blue colonies. HexAfor and HexBfor appear as white colonies as they contain in-frame stop codons. During this work, a dimeric version of HexA was obtained (designated HexArev2) and the sequencing gel for this is shown in Figure 3.4. The other end of HexB was visualised by labelling the opposite strand by cutting with XbaI and EcoRI and labelling with α-^{32}P-dCTP, though the 3'-end of the insert is not read as it runs off the end of the gels. Note that when the footprinting fragments are obtained by labelling the 3'-end of HindIII cut DNA (HexA), the same strand is visualised as on the sequencing gels. The opposite strand is visualised on the sequencing gels compared to the footprinting fragment when the DNA is labelled at the EcoRI restriction site (HexB).

Footprinting substrates were prepared from these clones and were used in DNase I (and hydroxyl radical) footprinting reactions with the ligands shown in Figure 3.3. It should be noted that the observed footprints often encompass two or more possible binding sites. In such cases, the analysis is focussed on the primary binding site and the other potential sites are indicated in parentheses.

Universal Dinucleotide Sequence

Figure 3.5 shows the results of DNase I footprinting experiments with distamycin; Hoechst 33258; mithramycin; nogalamycin; actinomycin D; and echinomycin on the DiN fragment. The footprinting sites are indicated by the filled boxes. The regions protected by each ligand are highlighted in the sequence shown in Figure 3.6, together with footprinting plots for the binding of distamycin to DiN. The C_{50} values for all the ligands are summarised in Table 3.1. The preferred binding sites for each of these ligands are well established so the results will only be briefly described.

Distamycin [$(A/T)_4$] produces a general inhibition of DNase I cleavage at 10 μM, though this can be divided into four regions that contain A/T-residues. The best binding site corresponds to the sequence TTAT (towards the bottom of the gel) the next best site is around ATT (the top footprint). TAA and TAT (the middle footprints on the gel) are bound more weakly.

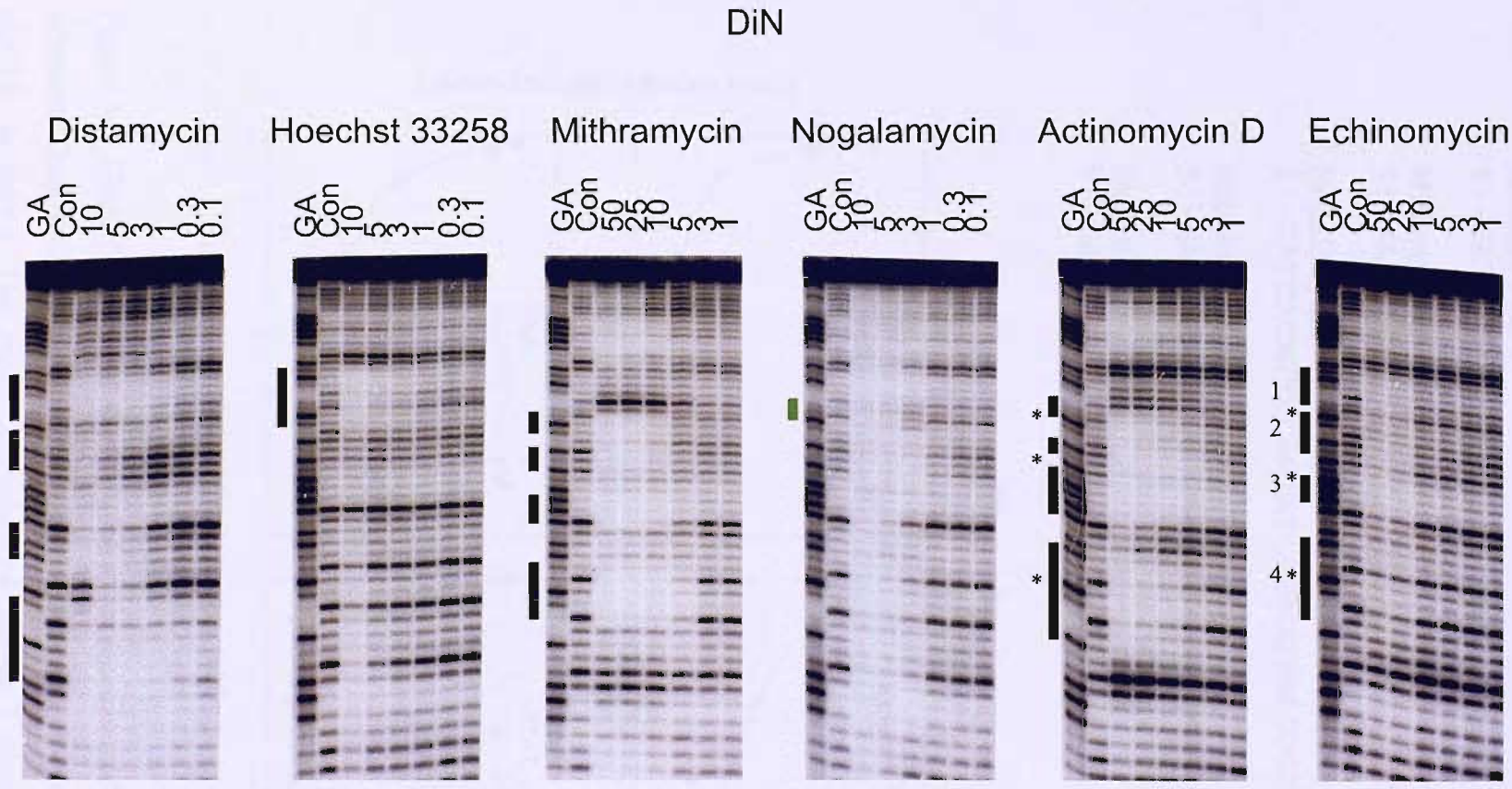


Figure 3.5: DNase I footprinting gels of seven ligands on the DiN substrate. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The boxes indicate the location of the footprints, which are also highlighted in the sequences shown in Figure 3.6. The site marked in green corresponds to a region of relatively enhanced cleavage with 3 μM nogalamycin. Sites marked "*" are GpC sites for actinomycin D and CpG sites for echinomycin. Sites are numbered as mentioned in the text. The experiments were performed in 10 mM Tris-HCl pH 7.4 containing 10 mM NaCl, except for those with mithramycin, which also contained 10 mM $MgCl_2$.

Distamycin
5'-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3'

Hoechst 33258
5'-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3'

Mithramycin
5'-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3'

Nogalamycin
5'-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3'

Actinomycin D
5'-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3'

Echinomycin
5'-GATCACATTCCGCTATGCGAGGTAACCGCTTATCA-3'

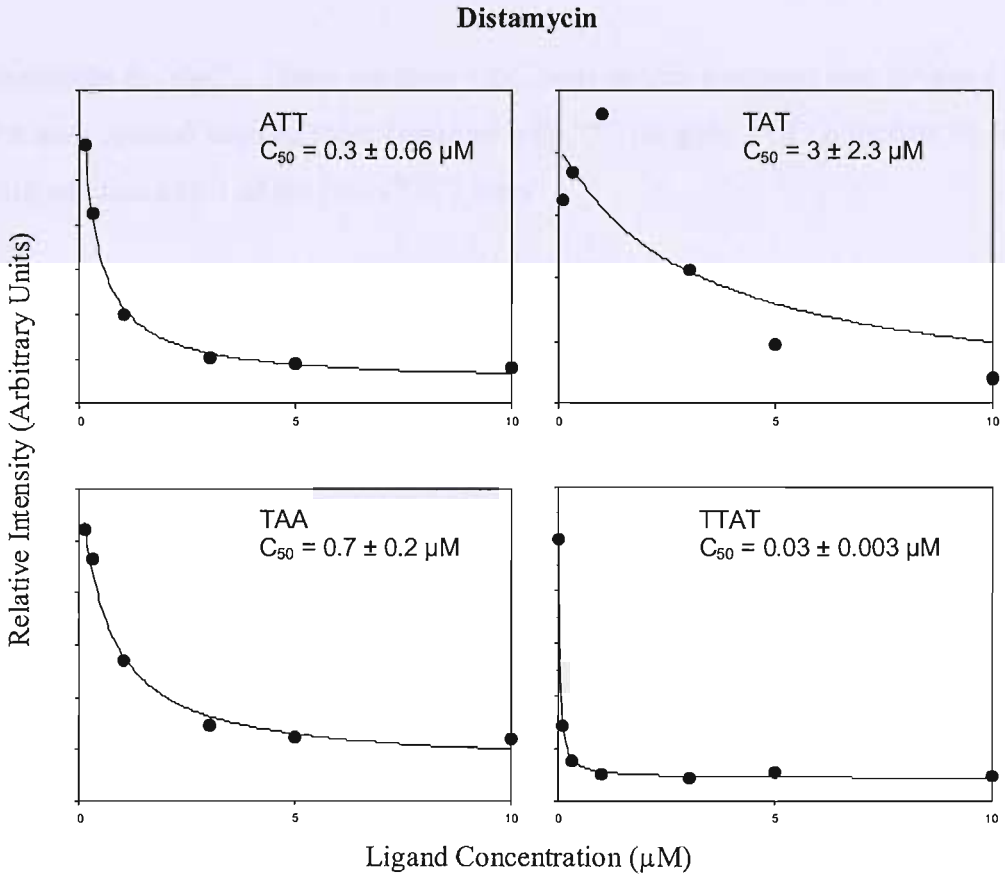


Figure 3.6: [Top] Footprinting sites for the seven ligands (red) on the DiN footprinting substrate. Only the sequence of the insert, which was cloned into the BamHI site of pUC18, is shown. The region of enhanced DNase I cleavage in the presence of nogalamycin is shown in green. [Above] Examples of footprinting plots determined for the interaction of distamycin with sites in DiN.

Hoechst 33258 [$(A/T)_4$] only generated one DNase I footprint on DiN, at the sequence ATT.

Mithramycin [$(G/C)_4$] produces a long region of protection covering several G/C-rich regions. This persists to lower concentrations around the sequence ACCGCT (towards the bottom of the gel). The ligand appears to bind to this site better than to TCCGCT.

Nogalamycin [$(RY)_n$] has produced a non-specific inhibition of cleavage throughout the fragment at concentrations of 3 μM and above. It is interesting to note the enhanced cleavage at CCG (indicated by the green box) revealing a region to which the ligand is not bound. However, another CCG site is also present in this fragment and this does not show enhanced cleavage.

Actinomycin D [GpC]. There are three GpC sites in this fragment and DNase I footprints can be seen around each of them (marked with "*" on gel). TGCG appears to be a better binding site that either of the two CGCT sites.

Echinomycin [CpG]. This fragment contains three CpG steps that would be expected to bind echinomycin (marked with "*" on gel). At concentrations of 25 μM and above there is a general attenuation of DNase I cleavage throughout the insert. However, within this attenuated cleavage pattern are four footprints (as marked). GCGA (site 3) and TCCGCT (site 2) are bound most strongly, with ACCGCT (site 4) bound more weakly, suggesting that base pairs further away from the core CpG motif affect the binding affinity. ACAT (site 1) is the weakest bound site, indicating that CpG is not necessarily required for binding.

The results with this short fragment, containing three copies of every dinucleotide step, are generally consistent with our previous knowledge of the preferred binding sites for these ligands. However, given the low complexity of the sequence these do not reveal any significant new features about the sequence preference for these ligands. It should also be noted that, in most cases it would not have been possible to determine the selectivity of these ligands based on this data alone. Clearly other universal footprinting fragments are required for general use.

	DiN			
Distamycin	ATT 0.3 ± 0.06	TAT 3 ± 2.3	TAA 0.7 ± 0.2	TTAT 0.03 ± 0.003
Hoechst 33258	ATT 4 ± 1.5			
Mithramycin	TCCGCT 9 ± 7	ATG 5.2 ± 3.6	AGG 3.5 ± 2	ACCGCT 3.3 ± 2.6
Nogalamycin	CCG (Stronger than control)			
Actinomycin D	CGCT 1.4 ± 0.5	TGCG 0.6 ± 0.1	CGCT 1.4 ± 0.4	
Echinomycin	Site 1 ACAT 25 ± 7	Site 2 TCCGCT 9 ± 4	Site 3 GCGA 8 ± 4	Site 4 ACCGCT 19 ± 5

Table 3.1: C_{50} values (μM) for the interaction of the seven ligands on the DiN fragment. The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). The site bound by nogalamycin is a site at which increased DNase I cleavage was observed.

Universal Trinucleotide Sequence

Figure 3.7 shows the results of DNase I footprinting experiments with distamycin; Hoechst 33258; mithramycin; nogalamycin; actinomycin D; and echinomycin on the TriN fragment. Footprinting sites are indicated by the filled boxes. The regions protected by each ligand are highlighted in the sequence shown in Figure 3.8 together with footprinting plots for the binding of distamycin to TriN. The C_{50} values for all the ligands are summarised in Table 3.2. As with the DiN fragment, the preferred binding sites for each of these ligands are well established so the results will only be briefly described.

Distamycin [$(A/T)_4$]. Like DiN, this fragment does not contain any $(A/T)_4$ sites that might be expected to bind distamycin. Nonetheless, the ligand has clearly affected the DNase I cleavage pattern at concentrations of 1 μM and above. TriN contains three $(A/T)_3$ sites, of which ATT (site 4) appears to be the best binding; AAA (site 3) requires higher concentrations to produce a footprint, while ATA (site 2) is weaker still. Surprisingly, one of the best footprints appears close to the top of this insert and is associated with the sequence ATCTAGCA (site 1), which does not contain a long A/T-tract, and demonstrates that distamycin can bind to some sequences that contain G/C residues.

Hoechst 33258 [$(A/T)_4$]. Hoechst 33258 is known to have similar selectivity to distamycin, yet the footprints produced are not the same. The ligand binds to each of the $(A/T)_3$ tracts and again ATT (site 4) is better than AAA (site 3), which is better than ATA (site 2). A footprint is again evident towards the top of the insert (site 1), though only at the highest concentrations of Hoechst 33258.

Mithramycin [$(G/C)_4$]. TriN contains one long G/C-tract (CGCCC; site 2), which binds mithramycin at similar concentrations to the CCC/GGGG sites that flank the insert. Clear protection from DNase I cleavage is also observed at CCG (site 3) and to a lesser extent GC (site 1), while a weaker footprint can also be seen around GTCC (site 4).

Nogalamycin [$(RY)_n$]. At concentrations of 3 μM and above nogalamycin produced a general non-specific inhibition of DNase I cleavage. However, a region in the centre of the oligopurine tract (GAAAG) shows a small amount of enhanced cleavage, which at least demonstrates a region to which nogalamycin is not bound, or binds less well.

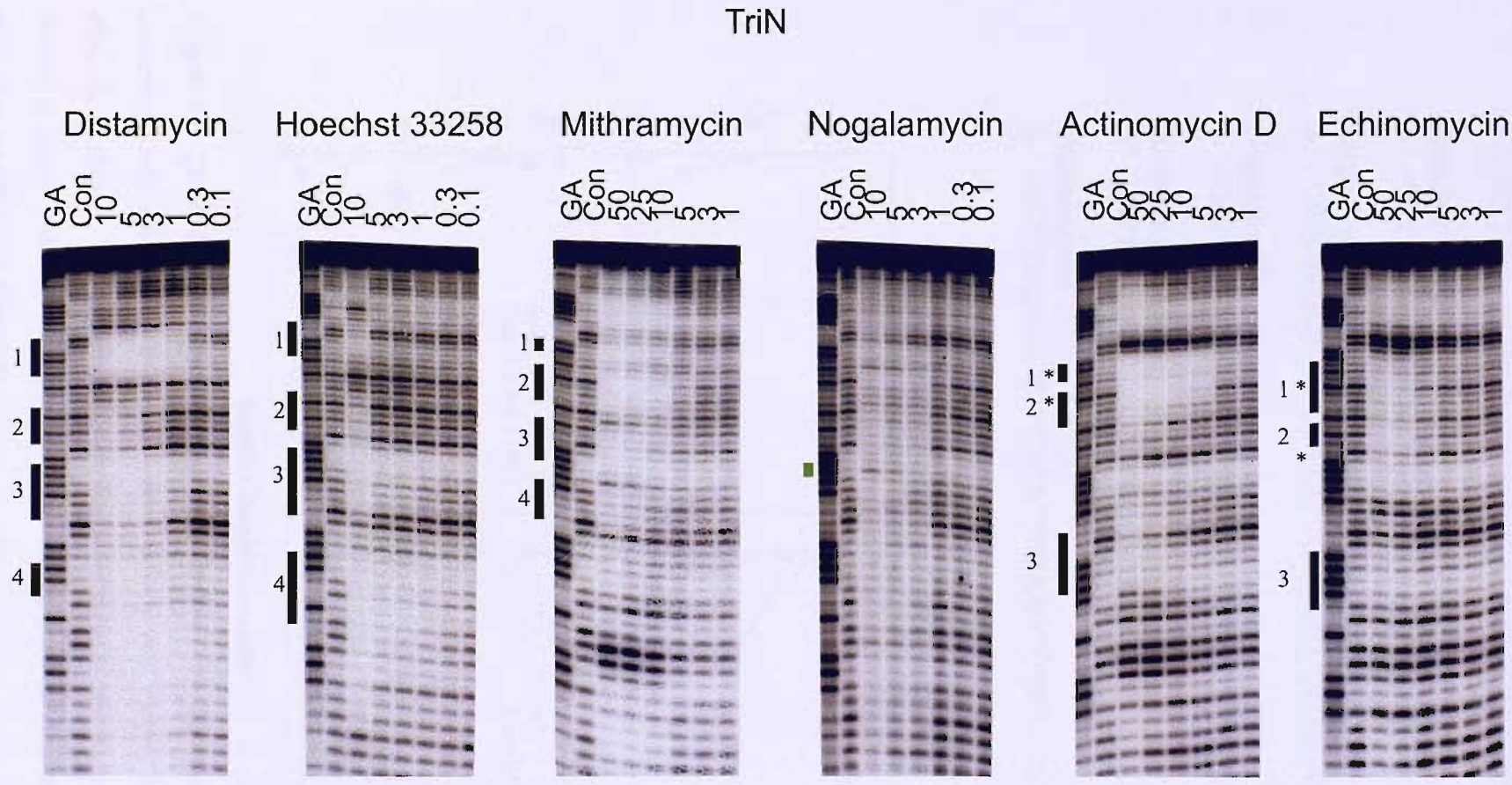


Figure 3.7: DNase I footprinting gels for the seven ligands with the TriN substrate. The footprints are indicated by the filled boxes and are also highlighted on the sequence in Figure 3.8. The site marked in green indicates a few bases that are cleaved better in the presence of nogalamycin. Sites marked "*" are GpC sites for actinomycin D and CpG sites for echinomycin. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

Distamycin
5'-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3'

Hoechst 33258
5'-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3'

Mithramycin
5'-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3'

Nogalamycin
5'-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3'

Actinomycin D
5'-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3'

Echinomycin
5'-GATCTAGCACGCCCATACCGAAAGTCCTGAGATTGTTA-3'

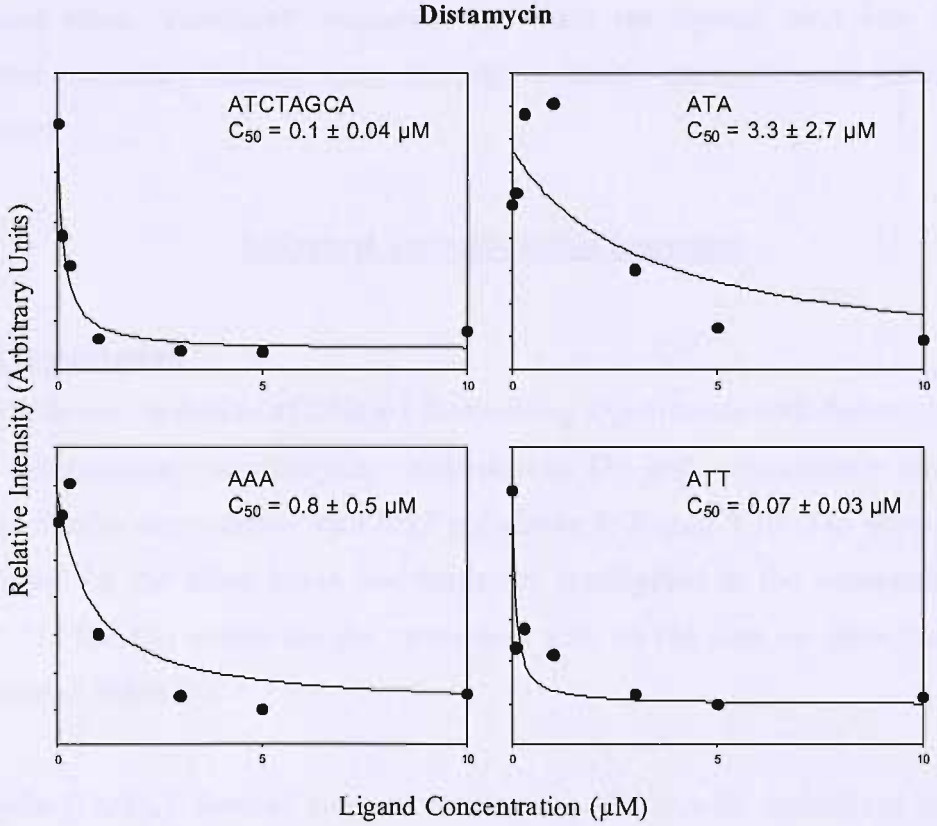


Figure 3.8: [Top] DNase I footprinting sites (in red) for the seven ligands on the TriN substrate. Only the sequence of the insert, which was cloned into the BamHI site of pUC18, is shown. The region of enhanced cleavage in the presence of nogalamycin is shown in green. [Above] Examples of footprinting plots for the interaction of distamycin with sites on this DNA fragment.

Actinomycin [GpC]. TriN contains two GpC steps, towards the top of the insert (marked with "*" on gel) and a clear DNase I footprint is evident in this region. Bands towards the bottom of this footprint reappear before those towards the top, suggesting that AGCA (site 1) is a better binding site than CGCC (site 2). An additional region of protection can be seen with higher actinomycin concentrations at GTCC (site 3).

Echinomycin [CpG]. TriN contains two CpG steps (marked with "*" on gel) and echinomycin produces two, almost adjoined footprints in the region that contains these sites: ACGC (site 1) and CCGA (site 2). There is also some attenuated cleavage further down the gel at the highest echinomycin concentrations at CTAG (site 3).

The results with this fragment, containing every trinucleotide step, are again consistent with our previous knowledge of the preferred binding sites for these ligands, though they also reveal some "canonical" sequences to which the ligands bind less well. Some unexpected secondary binding sites are also evident, especially with distamycin and actinomycin.

Universal Tetranucleotide Sequence

DNase I footprinting

Figure 3.9 shows the results of DNase I footprinting experiments with distamycin; Hoechst 33258; mithramycin; nogalamycin; actinomycin D; and echinomycin on the MS1 fragment. Similar experiments with MS2 are shown in Figure 3.10. The footprinting sites are indicated by the filled boxes and these are highlighted in the sequences shown in Figure 3.11. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 3.3.

Distamycin [$(A/T)_4$]. Several large footprints are evident with distamycin on MS1 and MS2 at concentrations of 3 μM and above. However, fewer clear footprints corresponding to the best sites are evident at 1 μM and below. It can be seen that distamycin produces clear footprints on MS1 and MS2 at AATA/TATT (site 2), ATTT/AAAT (site 3), AATTA/TAATT (site 8), TTAT/ATAA and TAAA/TTTA (site 9) and TTTT (site 10 on MS1; the complementary sequence on MS2 is AAAC and is bound weakly). Amongst the

	TriN			
Distamycin	Site 1	Site 2	Site 3	Site 4
	ATCTAGCA	ATA	AAA	ATT
	0.1 ± 0.04	3.3 ± 2.7	0.8 ± 0.5	0.07 ± 0.03
Hoechst 33258	Site 1	Site 2	Site 3	Site 4
	ATCTAGCA	ATA	AAA	ATT
	12 ± 8	13 ± 12	3.6 ± 1.6	0.1 ± 0.04
Mithramycin	Site 1	Site 2	Site 3	Site 4
	AG	CGCCC	ACCGA	GTCC
	9 ± 4	3.7 ± 2	3.8 ± 2	18 ± 6
Nogalamycin	GAAAG			
	(Stronger than control)			
Actinomycin D	Site 1	Site 2	Site 3	
	AGCA	CGCC	GTCC	
	1.4 ± 0.6	1.5 ± 0.6	3 ± 1	
Echinomycin	Site 1	Site 2	Site 3	
	ACGC	CCGA	CTGA	
	5 ± 1	5.6 ± 3.3	50 ± 24	

Table 3.2: C_{50} values (μM) for the interaction of the seven ligands on the TriN fragment. The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). The site bound by nogalamycin is a site at which increased DNase I cleavage was observed.

$(A/T)_4$ sites, ATAT (site 1 on MS2), TTAA (site 5) and TATA (site 7) are weaker sites. A surprisingly good footprint is evident on both strands within the sequence TTCTAT (site 6), suggesting that a single GC base pair can be tolerated within some of the binding sites (also seen more weakly with ACTAGT (site 4) and TCATCTCA (site 11).

Hoechst 33258 $[(A/T)_4]$. The footprints for Hoechst 33258 mainly correspond to the best sites that are seen with distamycin, with less evidence for additional regions of protection at higher concentrations. ATTT/AAAT (site 1), AATTA/TAATT (site 3), TAAA/TTTA (site 4) and TTTT (site 5 on MS1; the complementary sequence on MS2 is AAAC and is bound weakly) appear as good binding sites on both MS1 and MS2, of which AATTA and ATTT are the best. Of the secondary binding sites only observed at higher concentrations than distamycin, the DNase I cleavage is affected the least at TTAA (site 2).

Mithramycin $[(G/C)_4]$. This G/C-selective ligand has produced many footprints on these fragments. It appears that at concentrations of 25 μM and above mithramycin has bound to almost every G/C-tract that is two or more bases long (marked with filled boxes on the gels). However, at 10 μM fewer sites are occupied by the ligand (see the data with MS1) and these consist of longer $(G/C)_4$ tracts: GCGCG (site 1), CCCG (site 2), GGGCG (site 3) and GGGG (site 5). CCGG (site 4) exhibits weaker binding.

Nogalamycin $[(RY)_n]$. At concentrations of 3 μM and above nogalamycin produces a general inhibition of DNase I cleavage on MS1 and MS2, suggesting that it is binding without any significant sequence selectivity. However there are a few regions of protection that persist to lower concentrations, such as CATGTGTAT (site 1 on MS2), TGCAC/GTGCA (site 2), GTATA/TATAC (site 3), GCGTG/CACGC (site 4), TACATA (site 5 on MS2), GTACGT (site 6 on MS1), GTCTG/CAGAC (site 7) and ATGCG (site 8 on MS1), each of which consists of a region of alternating purines and pyrimidines with mixed base composition.

Actinomycin D [GpC]. Clear DNase I footprints are evident at every GpC step on MS1/MS2 in the presence of actinomycin D (marked with "*" on gels). TGCA (site 1) appears to be the best site and is stronger than AGCA/TCGT (site 5). It is interesting to see that two closely spaced GpC steps (*e.g.* AGCGCG/CGCGCT; site 2) appear to bind the ligand less well than isolated GpCs, though this may be because GCG/CGC is a poor

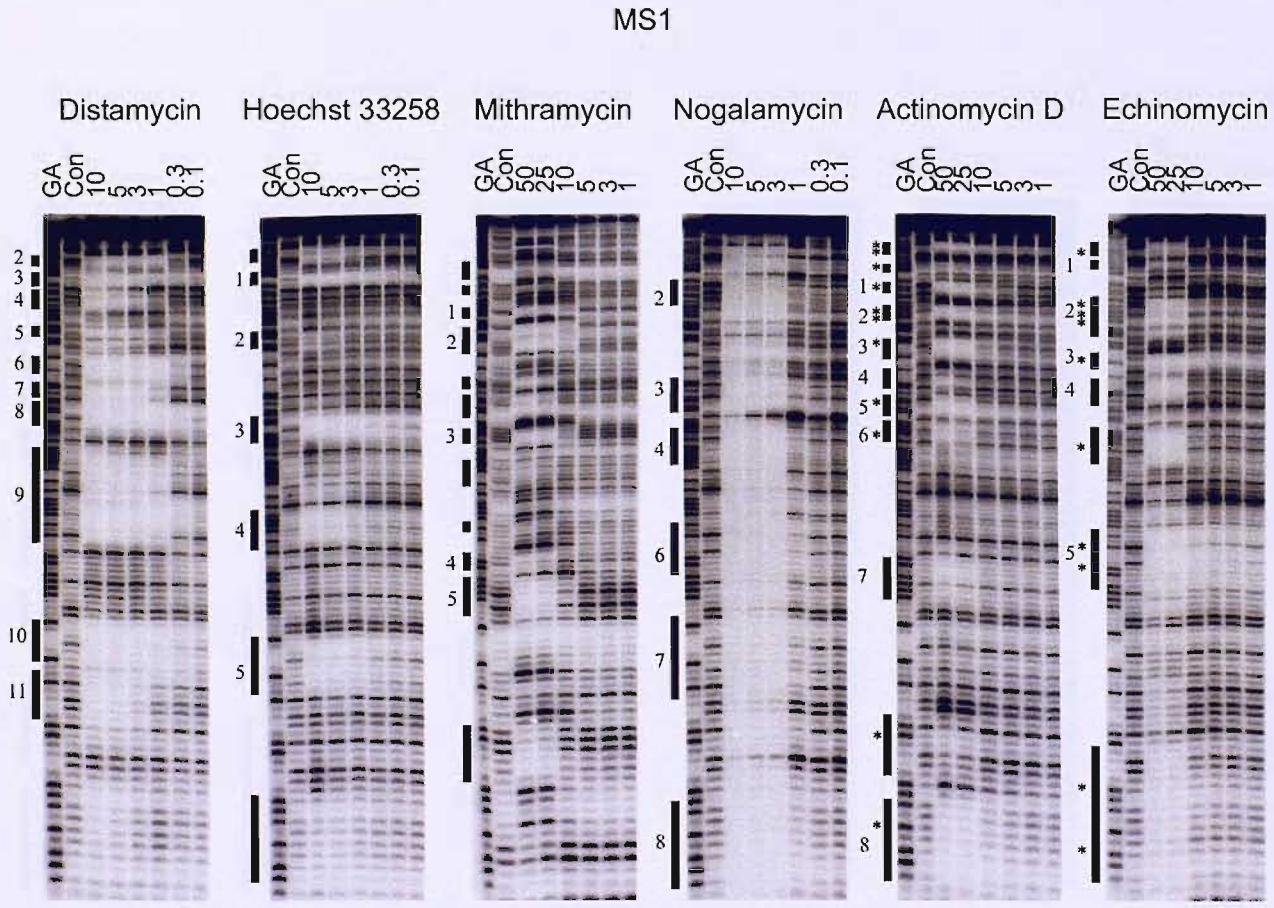


Figure 3.9: DNase I footprinting gels for the seven ligands with the MS1 fragment. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.11. Sites marked "*" are GpC sites for actinomycin D and CpG sites for echinomycin. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

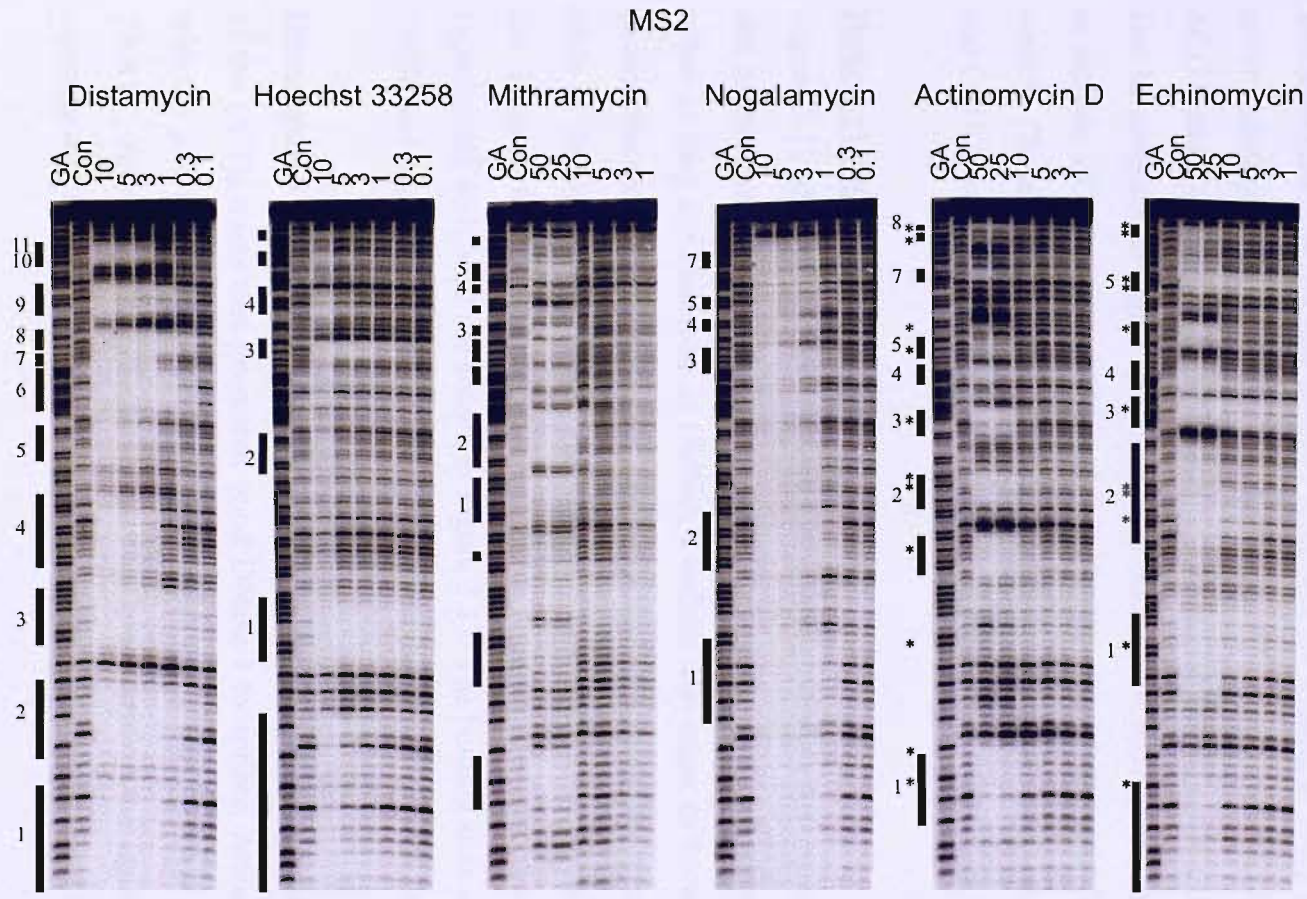


Figure 3.10: DNase I footprinting gels for the seven ligands with the MS2 fragment. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.11. Sites marked "*" are GpC sites for actinomycin D and CpG sites for echinomycin. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

binding site (*e.g.* TGCG/CGCA; site 8). Adjacent oligopurine or oligopyrimidine tracts also seem to inhibit binding (*e.g.* AGCTCCC/GGGAGCT; site 3) and AGGGCG (site 6 on MS1)). There may also be some interaction with non-GpC sites (*e.g.* GGTG/CACC; site 7) and TATC/GATA and GGTA/TACC (site 4).

Echinomycin [CpG]. Clear DNase I footprints are evident at every CpG step on MS1/MS2 in the presence of echinomycin (marked with "*" on gels), with CCGT/ACGG (site 3) and ACGT and CCGG (site 5) being the best sites. However, some CpA and TpG steps may also be protected at higher concentrations, such as GCAGA (site 1 on MS1; this is mutated to include a CpG step on MS2 (site 1 on MS2)). CCTGG/CCAGG (site 4) is also bound weakly. Closely spaced CpG steps seem to be bound better than isolated CpGs *(e.g.* ACGT and CCGG; site 5), though weaker binding is seen at CGCG (GCGCGA/TCGCGC; site 2).

Hydroxyl radical footprinting

Figure 3.12 shows hydroxyl radical footprinting patterns for distamycin, Hoechst 33258 and mithramycin on MS1 (the other ligands are unsuitable for hydroxyl radical footprinting either as they are dissolved in DMSO, a radical scavenger or because previous work has shown that they do not generate hydroxyl radical footprints (actinomycin D). Densitometer plots of the control lane and a ligand-treated lane are shown to the right of each gel, and the clearest footprinting sites are marked with the sequences of these sites. Similar experiments with MS2 are shown in Figure 3.13. The binding sites for distamycin, Hoechst 33258 and mithramycin are also highlighted on the sequence of MS1/MS2 in Figure 3.13.

Distamycin [$(A/T)_4$]. Clear hydroxyl radical footprints are produced by distamycin at most of the $(A/T)_4$ tracts that generated good DNase I footprints. Attenuated cleavage is seen with 0.1 μM ligand at the sites: TTCTA, AATTA, TTAT, TAAA and TTTT (MS1); and TAATT, AAAT and TATT (MS2). At this concentration the ligand has no effect on the cleavage at ATAT, TATA and TTAA, indicating TpA steps are not bound.

Hoechst 33258 [$(A/T)_4$]. As seen with DNase I, Hoechst 33258 produces similar footprinting patterns to distamycin. However Hoechst 33258 seems to be more selective for A/T-tracts and it has no effect on the sites containing single GC base pairs. It is interesting to note that a 3 bp A/T-tract is bound by Hoechst 33258, which does not show protection with distamycin (AAT).

Distamycin: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Hoechst 33258: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Mithramycin: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Nogalamycin: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Actinomycin D: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Echinomycin: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Figure 3.11: [Top] Sequences of the MS1 (top strand) and MS2 (bottom strand) fragments, indicating the footprints produced by the different ligands.

					Distamycin					
MS1	Site 2 AATA 0.2 ± 0.07	Site 3 ATTT 0.1 ± 0.05	Site 4 ACTAGT 1.2 ± 0.4	Site 5 TTAA 1 ± 0.4	Site 6 TTCTAT 0.1 ± 0.03	Site 7 TATA 0.6 ± 0.3	Site 8 AATTA 0.03 ± 0.004	Site 9 TTAT,TAAA 0.16 ± 0.03	Site 10 TTTT 0.05 ± 0.005	Site 11 TCATCTCA 0.7 ± 0.3
MS2	Sites 10 & 11 TGAGATGA 3 ± 1.5	Site 9 TTTA,ATAA 0.2 ± 0.1	Site 8 TAATT 0.1 ± 0.03	Site 7 TATA 0.8 ± 0.3	Site 6 ATAGAA 0.2 ± 0.1	Site 5 TTAA 1.3 ± 0.7	Site 4 ACTAGT 2 ± 1.6	Site 3 AAAT 0.6 ± 0.1	Site 2 TATT 1 ± 0.5	Site 1 ATAT 0.5 ± 0.3
					Hoechst 33258					
MS1	ATAT,AATA 1.5 ± 0.3	Site 1 ATTT 0.1 ± 0.01	Site 2 TTAA 11 ± 2	Site3 AATTA 0.05 ± 0.01	Site 4 TAAA 2 ± 0.6	Site 5 TTTT 0.5 ± 0.1	AAT 4 ± 1			
MS2	ATT 9 ± 3	AAAC 6 ± 2	Site 4 TTTA,ATAA 3 ± 1	Site 3 TAATT 0.1 ± 0.02	Site 2 TTAA 7 ± 3	Site 1 AAAT 0.2 ± 0.04	TATT,ATAT 2.3 ± 0.8			

Mithramycin												
MS1	GGCAG -	CTGCAC -	Site 1 GCGCG 10 ± 7	Site 2 CCCG 8 ± 3	CCTGG 22 ± 8	GC 41 ± 23	Site 3 GGGCG 10 ± 4	GAG 110 ± 83	GTACG -	Site 4 CCGG 34 ± 19	Site 5 GGGG 8 ± 5	GCCTCG 15 ± 7
MS2	CGAGGC 49 ± 31	Site 5 CCCC 20 ± 14	Site 4 CGTAC 33 ± 25	CTC -	Site 3 CGCCC 19 ± 5	GC 44 ± 19	CCAGG 22 ± 11	Site 2 CGGG 13 ± 10	Site 1 CGCGC 18 ± 13	GTGCAG 59 ± 40	CGGCC 62 ± 48	GCCGC 16 ± 14
Nogalamycin												
MS1	Site 2 TGCAC 0.2 ± 0.05	Site 3 GTATA 0.3 ± 0.04	Site 4 GCGTG 0.6 ± 0.25	Site 6 GTACGT 0.3 ± 0.05	Site 7 GTCTG 0.6 ± 0.3	Site 8 ATGCG 0.6 ± 0.2						
MS2	Site 7 CAGAC 1.4 ± 0.6	Site 5 TACATA 0.7 ± 0.2	Site 4 CACGC 1.5 ± 0.8	Site 3 TATAC 0.8 ± 0.3	Site 2 GTGCA 0.6 ± 0.2	Site 1 CATGTGTAT 0.5 ± 0.2						
Actinomycin D												
MS1	TGCGGCA 5 ± 1	GGCA 86 ± 29	Site1 TGCA 4 ± 1	Site 2 AGCGCG 8 ± 2	Site 3 AGCTCCC 9 ± 3	Site 4 TATC,GGTA 33 ± 8	Site 5 AGCA 15 ± 3	Site 6 AGGGCG 17 ± 4	Site 7 GGTG 24 ± 6	AGCC 54 ± 24	Site 8 TGCG 8 ± 3	
MS2	Site 8 CGCA 51 ± 17	GGCT 86 ± 34	Site 7 CACC 59 ± 31	Site 5 TGCT 24 ± 10	Site 4 TACC,GATA 86 ± 64	Site 3 GGGAGCT 16 ± 8	Site 2 CGCGCT 16 ± 7.5	TGCA 15 ± 9.5	Site 1 CGCA 35 ± 30			
Echinomycin												
MS1	GCGG 16 ± 6	Site 1 GCAGA 28 ± 10	Site 2 TCGT,GCGCGA 6 ± 3	Site 3 CCGT 0.5 ± 0.1	Site 4 CCTGG 20 ± 12	GCGT 18 ± 8	Site 5 ACGT,CCGG 0.3 ± 0.04	TCGA,GCGG 7 ± 3				
MS2	CCGC,TCGA 13 ± 3	Site 5 CCGG,ACGT 0.5 ± 0.1	ACGC 18 ± 7	Site 4 CCAGG 18 ± 6	Site 3 ACGG 0.2 ± 0.1	Site 2 TCGCGC,ACGA 7 ± 2	Site 1 TCGG 9 ± 2	CCGC 11 ± 5				

Table 3.3: C_{50} values (μM) derived from quantitative analysis of the MS1 and MS2 gels presented in this chapter (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). Sites marked "–" correspond to footprints that are evident on visual inspection, but for which quantitative analysis did not provide a clear value.

Mithramycin [$(G/C)_4$]. Mithramycin produces clear hydroxyl radical footprints in a number of locations, each of which corresponds to a G/C-rich region. These correspond to the regions protected from DNase I cleavage. In general these are located at $(G/C)_4$ tracts, though some shorter regions that contain G/C residues also appear to be bound when adjacent to a TpA step, whilst CCGG is again seen to be a weak site.

The results with these fragments, containing every tetranucleotide step, are consistent with our previous knowledge of the preferred binding sites for these ligands, though expanding the step length from trinucleotide has allowed more binding rules to be developed.

Universal Symmetrical Hexanucleotide Sequences

Distamycin [$(A/T)_4$]

DNase I footprinting

Figure 3.14 shows the results of DNase I footprinting experiments with distamycin on HexAfor, HexArev2, HexBfor and HexBrev fragments, thereby showing the interaction with every symmetrical hexanucleotide sequence. On these gels, the footprinting sites are indicated by the filled boxes and these are highlighted in the sequences shown in Figure 3.15. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 3.4.

HexA: Every $(A/T)_4$ site except (G/C)TTAA(G/C) produces a clear footprint on HexAfor at concentrations of 3 μM and below. The strongest sites consist of A/T-tracts of 6 bp, such as ATATAT (site 2), TTTAAA (site 6; indicating that a TpA step is tolerated) and AAATTT (site 3). The 4 bp A/T-tracts GAATTC (site 5) and CTATAG (site 4; both only clearly defined on HexArev2) are also strong sites, but GATATC (site 1) is weaker, suggesting that longer A/T-tracts are preferred, even though only 4 bp are required for distamycin binding (through comparison of ATATAT and GATATC). A very good footprint is evident at GTACTA (site 9), indicating that a single GC base pair can be tolerated within some of the binding sites. Sites with two GC base pairs in between A/T regions are bound more weakly, such as ATGCTA (site 8), TACCTA (site 7) and ATGGAT (site 10).

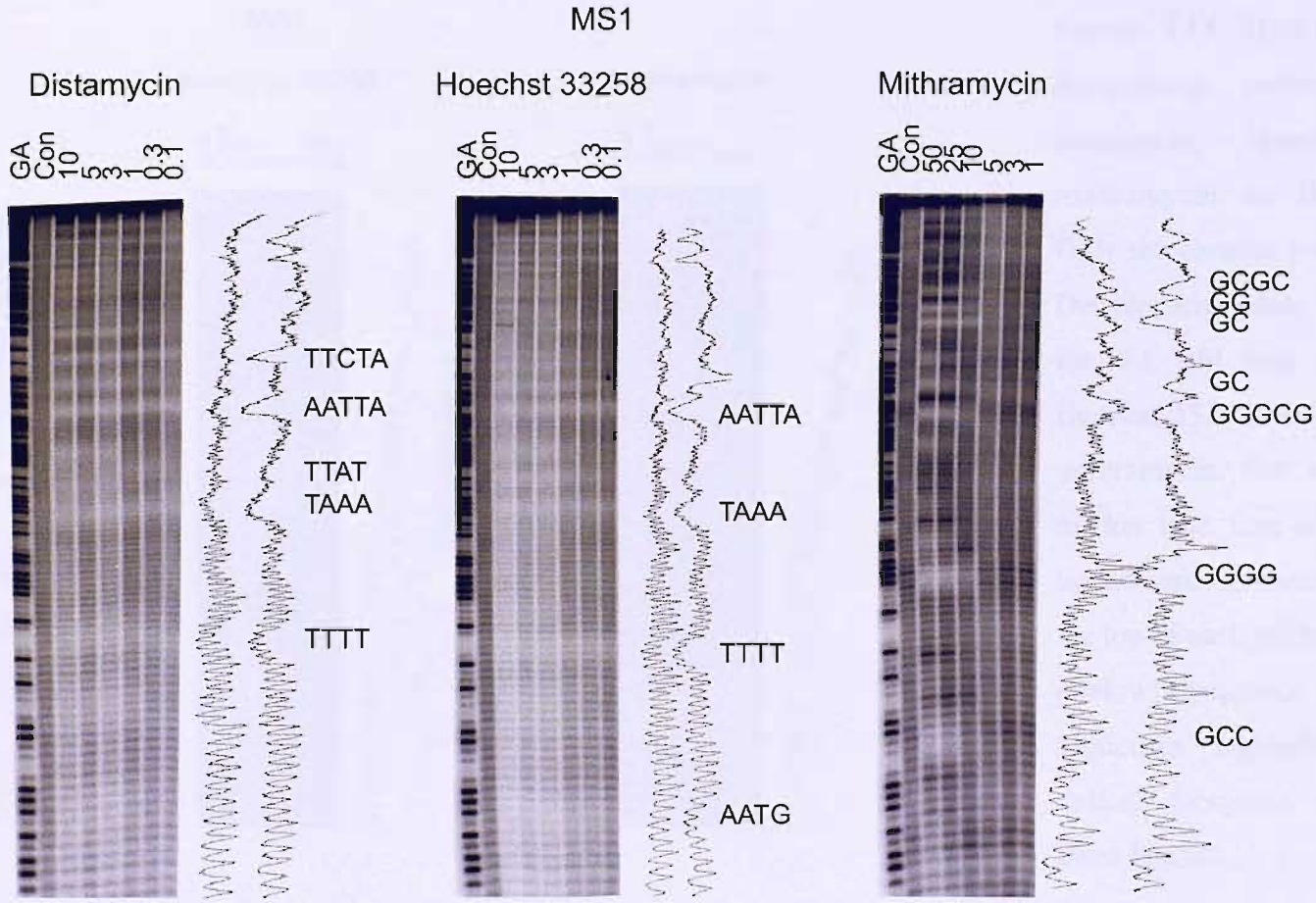


Figure 3.12: Hydroxyl radical footprinting patterns produced by distamycin, Hoechst 33258 and mithramycin on the MS1 substrate. Only the clearest footprints are marked. Densitometer plots were derived from the 0.1 μM lane for distamycin and Hoechst 33258, and from the 10 μM lane for mithramycin. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

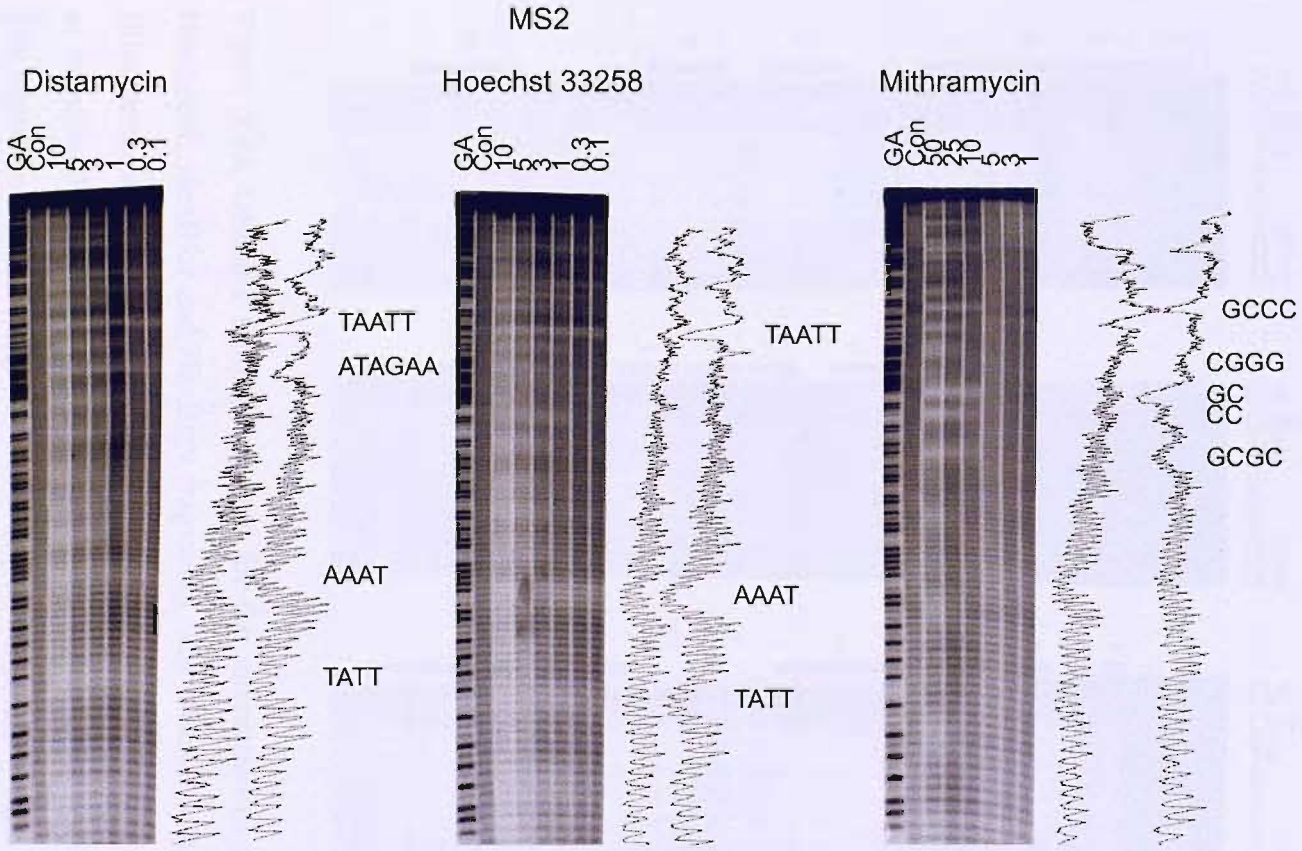


Figure 3.13: [Left] Hydroxyl radical footprinting patterns produced by distamycin, Hoechst 33258 and mithramycin on the MS2 fragment. Only the clearest footprints are marked. Densitometer plots were derived from the 0.1 μM lane of distamycin and Hoechst 33258, and the 10 μM lane of mithramycin. GA is a purine specific marker lane. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

[Below] Sequence of MS1 and MS2 sequences highlighting the hydroxyl radical footprints observed with the three ligands.

Distamycin: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Hoechst 33258: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Mithramycin: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

HexB: Every $(A/T)_4$ site has been bound by distamycin on HexBfor (which does not contain any (G/C)TTAA(G/C) sites). As with HexA, $(A/T)_6$ sequences are preferred, with TAATTA (site 1) the strongest site. TTATAA (site 3), TATATA (site 5, with ATAT), AATATT (site 6) and AATTAAT (site 7) are slightly weaker binding sites. $(A/T)_4$ sites are bound less strongly by distamycin and CAATTG (site 2) and GTATAC (site 4) have similar affinity, suggesting that TpA steps can be tolerated within the binding sites.

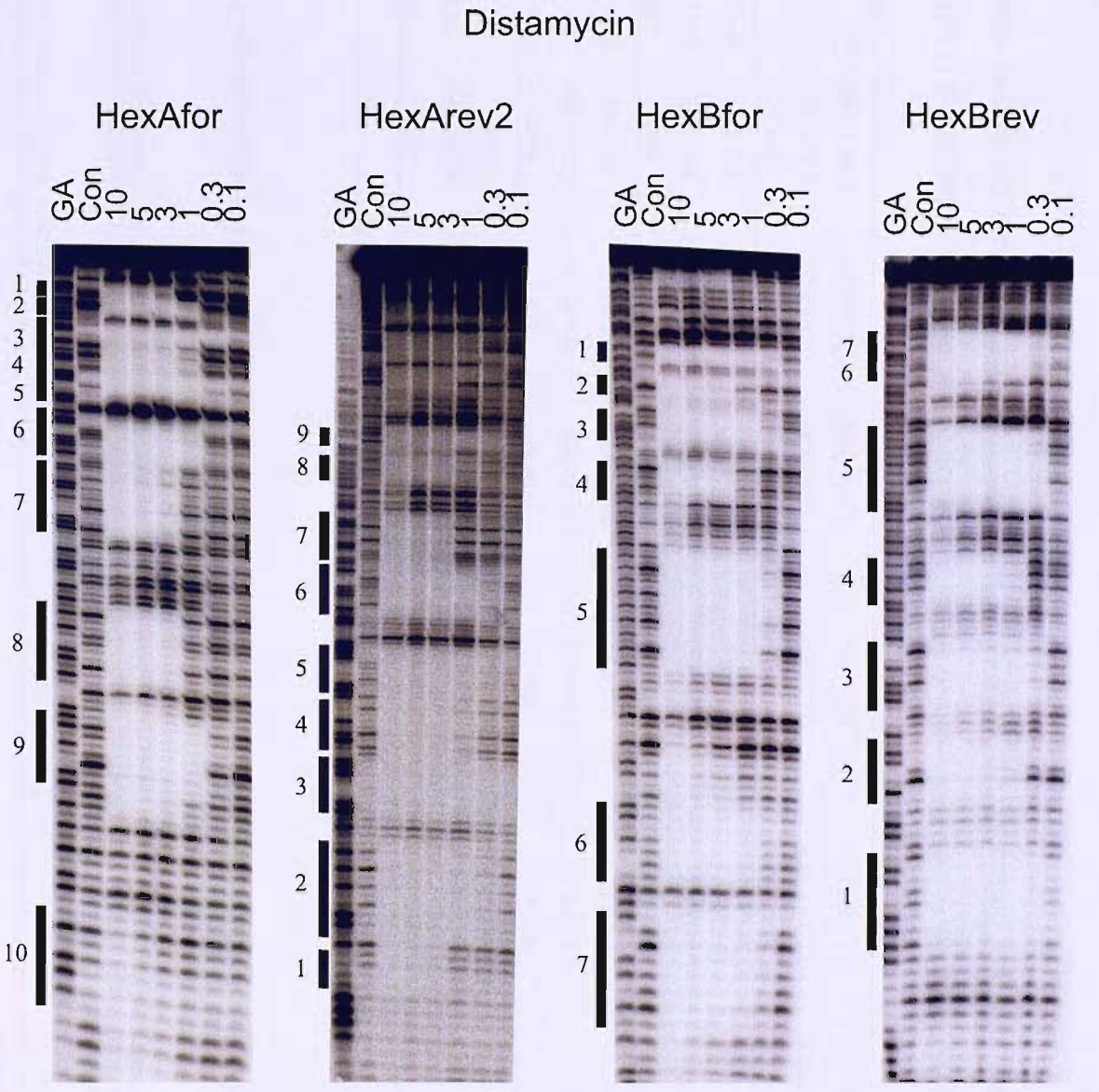


Figure 3.14: DNase I footprinting gels for the interaction of distamycin with HexAfor, HexArev2, HexBfor and HexBrev fragments. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.15. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

Distamycin: HexAfor (top); HexArev2 (bottom)

5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'

3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'

Distamycin: HexBfor (top); HexBrev (bottom)

5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'

3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'

Figure 3.15: Sequences of the HexA (top) and HexB (above) fragments, indicating the footprints (in red) produced by distamycin.

	Distamycin								
HexAfor	Site1	Site 2	Sites 3, 4 & 5	Site 6	Site 7	Site 8	Site 9	Site 10	
	GATATC	ATATAT	AAATTT,TATA,AATT	TTTAAA	TACCTA	ATGCTA	GTACTA	ATGGAT	
	2.2 ± 1.4	0.4 ± 0.2	0.6 ± 0.4	0.4 ± 0.2	2 ± 1.4	1 ± 0.7	0.2 ± 0.1	5.5 ± 3.6	
HexArev2	Site 9	Site 8	Site 7	Site 6	Site 5	Site 4	Site 3	Site 2	Site 1
	TAGTAC	TAGCAT	TAGGTA	TTTAAA	GAATTC	CTATAG	AAATTT	ATATAT	GATATC
	0.1 ± 0.02	1 ± 0.4	3.5 ± 2	0.2 ± 0.03	0.1 ± 0.01	0.4 ± 0.1	0.03 ± 0.01	0.1 ± 0.03	1.3 ± 0.5
HexBfor	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7		
	TAATTA	CAATTG	TTATAA	GTATAC	TATATA,ATAT	AATATT	AATTAAT		
	0.1 ± 0.07	0.7 ± 0.2	0.2 ± 0.1	1.1 ± 0.7	0.14 ± 0.08	0.3 ± 0.1	0.13 ± 0.06		
HexBrev	Sites 7 & 6	Site 5	Site 4	Site 3	Site 2	Site 1			
	ATTAATT,AATATT	ATAT,TATATA	GTATAC	TTATAA	CAATTG	TAATTA			
	0.1 ± 0.03	0.1 ± 0.04	0.7 ± 0.6	0.14 ± 0.08	0.4 ± 0.25	0.04 ± 0.01			

Table 3.4: C_{50} values (μM) derived from quantitative analysis of the interaction of distamycin with HexA and HexB fragments (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

Hydroxyl radical footprinting

Figure 3.16 shows hydroxyl radical footprinting patterns for distamycin on the HexAfor, HexArev2, HexBfor and HexBrev fragments. Densitometer plots of the control lane and a ligand-treated lane are shown on the right of each gel. The clearest footprinting sites are marked by the filled boxes on the left side of the gels and these are highlighted in the sequences shown in Figure 3.16 below the gels.

HexA: Clear hydroxyl radical footprints are produced by distamycin at all the sites that generated good DNase I footprints, including all $(A/T)_4$ sequences apart from (G/C)TTAA(G/C) (CTATAG is also not bound on HexArev2). The footprints observed at the lower concentrations appear smaller than those seen with DNase I. Within the $(A/T)_6$ sites, only the central (underlined) bases of AT<u>ATAT</u> (site 1), A<u>AATTT</u> (site 2) and TT<u>TAAA</u> (site 5) show attenuated cleavage at 0.1 μM distamycin. The $(A/T)_4$ sites CT<u>ATAG</u> (site 3; only on HexAfor, suggesting it is a weaker site) and G<u>AATTC</u> (site 4) are also bound. It is interesting to note that whilst TACCTA (site 6) and ATGCTA (site 7) show cleavage protection on HexAfor (as observed with DNase I), only TACTAG (site 8) is bound on both orientations of the fragment, supporting the observation with DNase I footprinting that two GC base pairs between A/T regions are less favoured than a single GC base pair.

HexB: Every $(A/T)_4$ sequence shows attenuated cleavage by distamycin on HexB except for GTATAC and CATATG, indicating TpA steps are not bound well. As with the previous fragment, footprinting sites do not cover the entire A/T-tract, as shown by T<u>AATTA</u> (site 1), C<u>AATTG</u> (site 2), TT<u>ATAA</u> (site 3) and T<u>ATATA</u> (site 4; only on HexBfor, suggesting it is a weaker site). <u>AATATT</u> (site 5) and ATT<u>AATT</u> (site 6) only show clear protection on HexBrev due to poor resolution of the HexBfor gel. The position of the footprints within these A/T-tracts seems to indicate TpA steps are not bound by distamycin. The underlined footprint sites correspond to the HexBfor results when the sequence is bound on both orientation of the fragment.

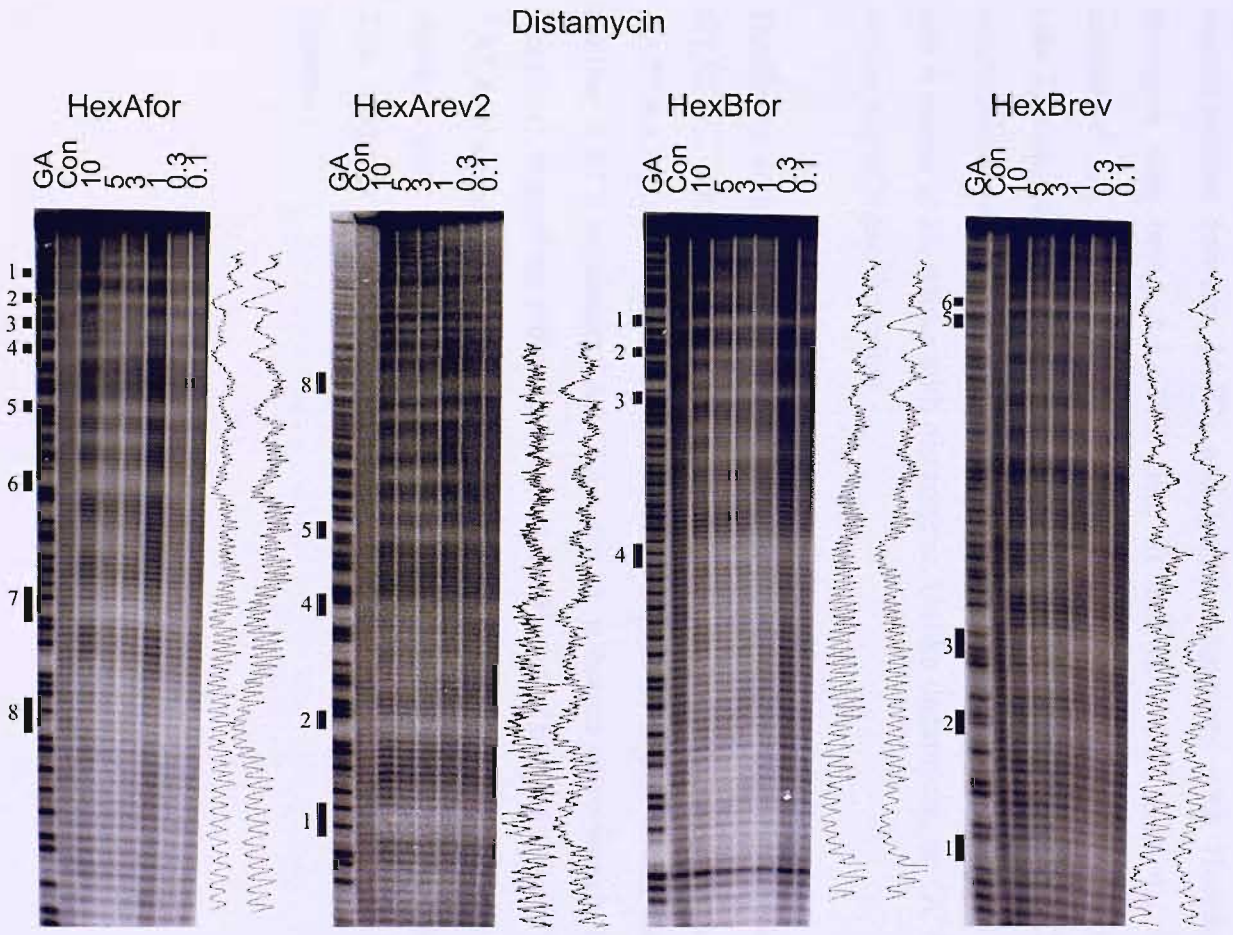


Figure 3.16: [Left] Hydroxyl radical footprinting patterns produced by distamycin on HexAfor, HexArev2, HexBfor and HexBrev fragments. Only the clearest footprints are marked. Densitometer plots were derived from the 0.1 μM lanes of distamycin. Sites are numbered as mentioned in the text. GA is a purine specific marker lane. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. [Below] Sequence of HexA and HexB fragments highlighting (in red) the hydroxyl radical footprints observed with distamycin.

Distamycin: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

Distamycin: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

Hoechst 33258 [$(A/T)_4$]

DNase I footprinting

Figure 3.17 shows the results of DNase I footprinting experiments with Hoechst 33258 on HexAfor, HexArev2, HexBfor and HexBrev fragments, thereby showing the interaction with every symmetrical hexanucleotide sequence. In these gels, the footprinting sites are indicated by the filled boxes and these are highlighted in the sequences shown in Figure 3.18. Examples of footprinting plots derived from the binding of Hoechst 33258 to HexBrev can be found in Figure 3.19. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 3.5.

HexA: Every $(A/T)_4$ site except CTATAG is bound by Hoechst 33258. There seems less discrimination between $(A/T)_4$ and $(A/T)_6$ sequences with Hoechst 33258, with the strongest sites being AAATTT (site 3) and GAATTC (site 4). It can therefore be concluded that the $(A/T)_4$ sequence AATT is the preferred binding site, with ATATAT (site 2) bound less strongly. GATATC (site 1), TTTAAA (site 5) and CTTAAG (site 6) are bound more weakly. Clearly TTAA is not bound well, although it is interesting that this site is bound at all, unlike with distamycin. Unlike distamycin, no GC base pairs are found within any of these footprints.

HexB: As with the HexA fragments, every $(A/T)_4$ site is bound by Hoechst 33258 except GTATAC. AATT is again the preferred $(A/T)_4$ sequences, with CAATTG (site 2) and AATTAAT (site 7) the best sites. AATATT (site 6) is also bound strongly. Interestingly another AATT sequence, TAATTA (site 1), is bound more weakly (most noticeably on HexBfor), suggesting that a TpA step at the 5'-end of the site reduces the binding affinity. TATA is a poor binding site; TATATA (site 5, with ATAT) and TTATAA (site 3) have similarly poor affinity, while GTATAC only shows weak protection on HexBrev (site 4). This suggests $(A/T)_6$ sequences are preferred over $(A/T)_4$-tracts when TpA steps are present.

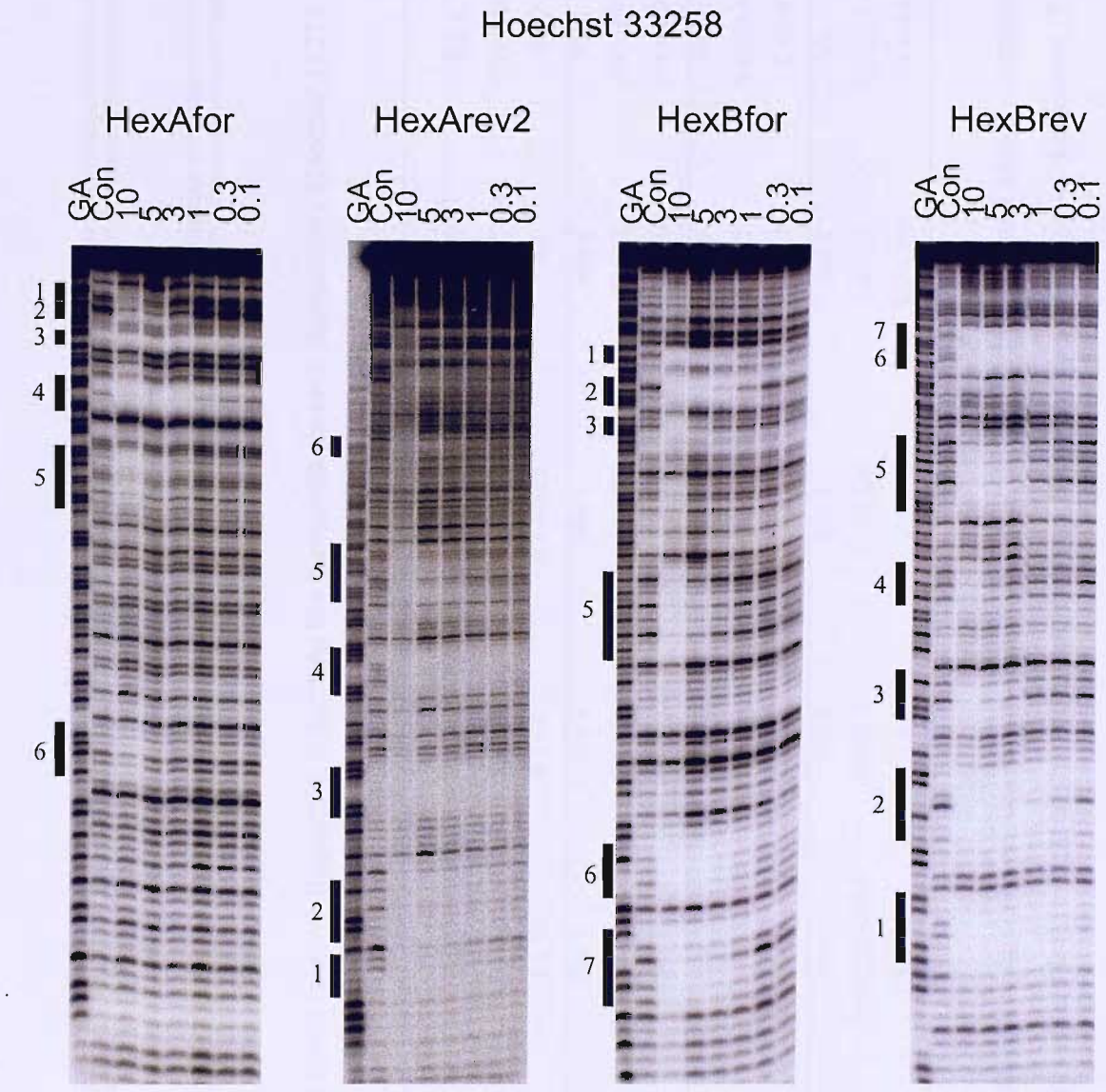


Figure 3.17: DNase I footprinting gels for the interaction of Hoechst 33258 with HexAfor, HexArev2, HexBfor and HexBrev fragments. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.18. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

Hoechst 33258: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

Hoechst 33258: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

Figure 3.18: Sequences of the HexA (top) and HexB (above) fragments, indicating the footprints (in red) produced by Hoechst 33258.

	Hoechst 33258					
HexAfor	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
	GATATC	ATATAT	AAATTT	GAATTC	TTTAAA,TTAA	CTTAAG
	8 ± 2	1.4 ± 0.5	0.1 ± 0.1	0.3 ± 0.15	8 ± 2	10 ± 2
HexArev2	Site 6	Site 5	Site 4	Site 3	Site 2	Site 1
	CTTAAG	TTAA,TTTAAA	GAATTC	AAATTT	ATATAT	GATATC
	9 ± 1	4 ± 1	0.03 ± 0.01	0.2 ± 0.1	0.06 ± 0.01	0.1 ± 0.04
HexBfor	Site 1	Site 2	Site 3	Site 5	Site 6	Site 7
	TAATTA	CAATTG	TTATAA	TATATA,ATAT	AATATT	AATTAAT
	2 ± 0.6	1 ± 0.7	7 ± 4	4 ± 2.5	1 ± 0.7	1 ± 0.6
HexBrev	Sites 7 & 6	Site 5	Site 4	Site 3	Site 2	Site 1
	ATTAATT,AATATT	ATAT,TATATA	GTATAC	TTATAA	CAATTG	TAATTA
	0.1 ± 0.01	1.5 ± 0.5	5 ± 2	2 ± 1	0.2 ± 0.01	0.1 ± 0.01

Table 3.5: C_{50} values (μM) derived from quantitative analysis of the interaction of Hoechst 33258 with HexA and HexB fragments (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

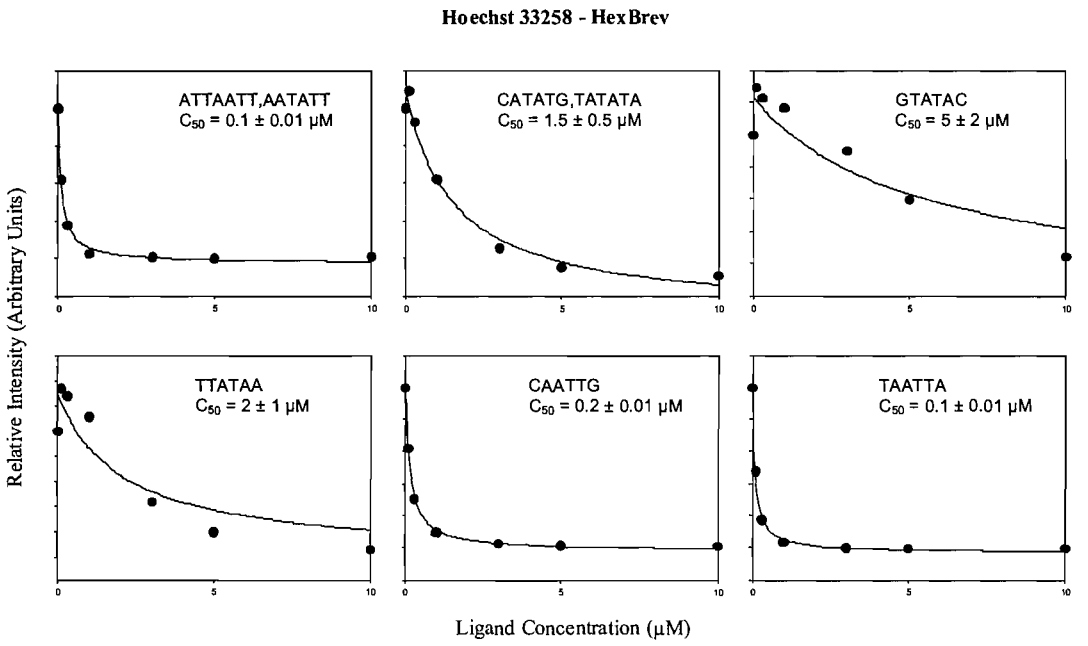


Figure 3.19: Examples of footprinting plots derived from the binding of Hoechst 33258 to HexBrev.

Hydroxyl radical footprinting

Figure 3.20 shows hydroxyl radical footprinting patterns for Hoechst 33258 on HexAfor, HexArev2, HexBfor and HexBrev fragments. Densitometer plots of the control lane and a ligand-treated lane are shown on the right of each gel. The clearest footprinting sites are marked by the filled boxes on the left side of the gels and these are highlighted in the sequences shown in Figure 3.20 below the gels.

HexA: Every $(A/T)_6$ site shows attenuated hydroxyl radical cleavage with 0.1 μM Hoechst 33258 on HexA, such as ATATAT (site 1), AAATTT (site 2) and TTTAAA (site 4). As with distamycin, the footprints at the lower concentrations (shown as underlined bases) cover only a part of the A/T-tract. GAATTC (site 3) is the only $(A/T)_4$ site that shows attenuated cleavage at 0.1 μM, with GATATC, CTATAG, GTTAAC and CTTAAG only producing footprints at higher concentrations. This supports the DNase I footprinting results that AATT is the preferred Hoechst 33258 binding site and that $(A/T)_6$ sequences are preferred over $(A/T)_4$ sites when TpA steps are present.

HexB: Every $(A/T)_4$ sequence shows attenuated cleavage by Hoechst 33258 on HexB at

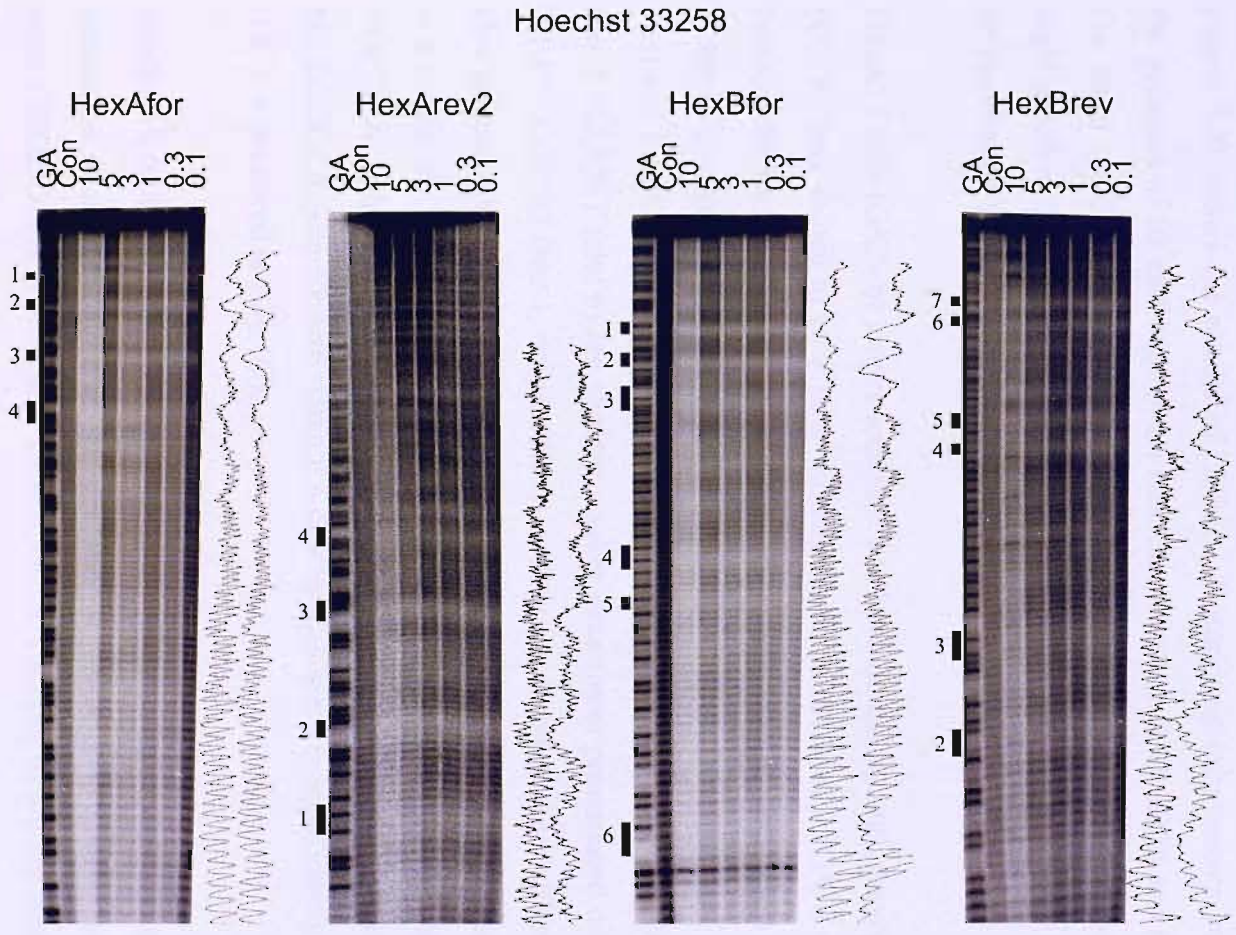


Figure 3.20: [Left] Hydroxyl radical footprinting patterns produced by Hoechst 33258 on HexAfor, HexArev2, HexBfor and HexBrev fragments. Only the clearest footprints are marked. Densitometer plots were derived from the 0.1 μM lanes. Sites are numbered as mentioned in the text. GA is a purine specific marker lane. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. [Below] Sequence of HexA and HexB fragments highlighting (in red) the hydroxyl radical footprints observed with Hoechst 33258.

Hoechst 33258: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

Hoechst 33258: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

0.1 μM except for GTATAC. Unclear cleavage patterns in the control lane prohibit resolution of good footprints at AATTAAT (at the bottom of the gel). As with the previous fragments, the footprints do not cover the entire A/T-tracts, as shown by TAATTA (site 1; only on HexBfor as it is unclear on HexBrev), CAATTG (site 2), TATATA (site 4), CATATG (site 5), AATATT (site 6) and ATTAATT (site 7; only on HexBrev as it is unclear on HexBfor). TTATAA (site 3) is also bound. The underlined footprint sites correspond to the HexBfor results when the sequence is bound on both orientation of the fragment.

Mithramycin [$(G/C)_4$]

DNase I footprinting

Figure 3.21 shows the results of DNase I footprinting experiments with mithramycin (in the presence of 10 mM $MgCl_2$) on HexAfor, HexArev2, HexBfor and HexBrev fragments. On these gels, the footprinting sites are indicated by the filled boxes and these are highlighted in the sequences shown in Figure 3.22. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 3.6.

HexA: Every $(G/C)_4$ or longer sequence has been bound by mithramycin except TCCGGA (CCGG was shown to be a poor binding site on MS1/MS2). $(G/C)_4$ and $(G/C)_6$ sites are bound with similar affinity and the best sequences have adjacent TpA steps on both sides of the G/C-tract, such as TTAACCGGTA (site 5; even though CCGG is normally a weak site) and TAGCGCTTAA (site 8). TATATGGCGCC (site 2), CCGCGGTTTAAA (site 4) and TAGGCCT (site 6) are slightly weaker sites (most pronounced on HexArev2), whilst ATATCCCCGG (site 1; only on HexArev2) is a poor binding site. Shorter G/C-regions are also protected when adjacent to TpA steps and TAGCTATA (site 3; flanked by TpA steps) is a strong site. TAGTGCACG (site 9) shows weaker affinity. It is interesting to note that TGCGCA with no adjacent TpA step (site 7) binds with similar affinity to the best sites, but TGGCCA (site 10; also without a nearby TpA step) is significantly weaker, suggesting $[GC]_n$ is preferred to $[G]_n.[C]_n$.

HexB: All $(G/C)_4$ and longer sequences show mithramycin footprints on HexB, with visual inspection of the gels most clearly revealing differences in binding strength between the sites. GGGCCCTA (site 3) and TAAGCGCGCTA (site 6) are the strongest binding sites,

while TATACGCGTA (site 8) and TACGCGCGTATATA (site 9) are weaker sites, indicating that binding is better when there are more GpC than CpG steps. This is supported by the observations that AAGCTTATA (site 5) is a better binding site than TACGTATA (site 7), and TAGCCGGC (site 4) is a poor site. TCGCGA (site 2, with no adjacent TpA step; only on HexBfor) binds better than CCGGCCG (site 1, also with no nearby TpA step; only on HexBfor), again suggesting CCGG is a poor site.

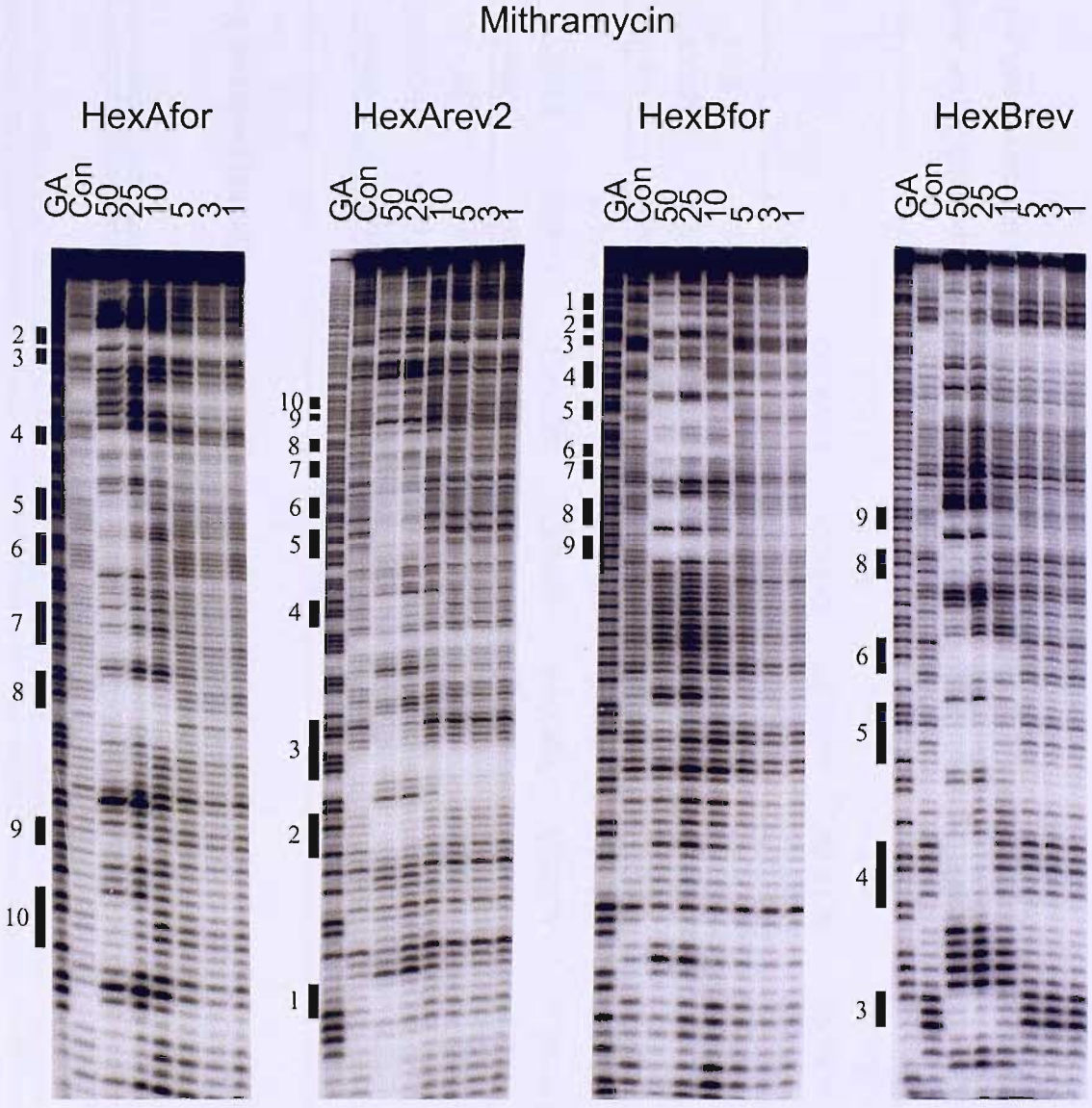


Figure 3.21: DNase I footprinting gels for the interaction of mithramycin with HexAfor, HexArev2, HexBfor and HexBrev fragments in the presence of 10 mM $MgCl_2$. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.22. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

Figure 3.22: Sequences of the HexA (top) and HexB (above) fragments, indicating the footprints (in red) produced by mithramycin.

	Mithramycin									
HexAfor	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	
	GGCGCC	TAGCTA	CCGCGG	ACCGGT	AGGCCT	TGCGCA	AGCGCT	GTGCACG	TGGCCA	
	15 ± 6	19 ± 6	11 ± 5	12 ± 5	10 ± 5	10 ± 4	6 ± 3	20 ± 5	22 ± 12	
HexArev2	Site 10	Site 9	Site 8	Site 7	Site 6	Site 5	Site 4	Site 3	Site 2	Site 1
	TGGCCA	CGTGCAC	AGCGCT	TGCGCA	AGGCCT	ACCGGT	CCGCGG	TAGCTA	GGCGCC	TCCCGG
	47 ± 27	65 ± 49	14 ± 6	30 ± 15	21 ± 13	16 ± 9	31 ± 16	25 ± 19	21 ± 12	32 ± 25
HexBfor	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	
	CCGGCCG	TCGCGA	GGGCCC	GCCGGC	AAGCTT	GCGCGC	TACGTA	ACGCGT	CGCGCG	
	9 ± 5	8 ± 7	6 ± 4	8 ± 4	7 ± 6	6 ± 3	9 ± 6	8 ± 4	7 ± 4	
HexBrev	Site 9	Site 8	Site 6	Site 5	Site 4	Site 3				
	CGCGCG	ACGCGT	GCGCGC	AAGCTT	GCCGGC	GGGCCC				
	109 ± 34	32 ± 18	8 ± 3	13 ± 5	16 ± 11	6 ± 3				

Table 3.6: C_{50} values (μM) derived from quantitative analysis of the interaction of mithramycin with HexA and HexB fragments (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

Hydroxyl radical footprinting

Figure 3.23 shows hydroxyl radical footprinting patterns for mithramycin on HexAfor, HexArev2, HexBfor and HexBrev fragments in the presence of 10 mM $MgCl_2$. Densitometer plots of the control lane and a ligand-treated lane are shown on the right of each gel. The clearest footprinting sites are marked by the filled boxes on the left side of the gels and these are highlighted in the sequences shown in Figure 3.23 below the gels.

HexA: Every $(G/C)_4$ site shows attenuated hydroxyl radical cleavage in the presence of high concentration of mithramycin on HexA, but at 10 μM only TTAACCGGTA (site 1), TAGGCCT (site 2) and TAGCGCTTAA (site 3) show clear footprints. As seen with the DNase I footprinting experiments, TpA steps are important (TGCGCA and TGGCCA do not bind at the lower concentrations). Interestingly, TATATGGCGCC and CCGCGGTTTAAA are not protected at the lower concentrations, suggesting that either $(G/C)_4$ sequences are preferred over $(G/C)_6$ sites, or that TpA steps have to be abutted against the G/C-tract to affect the binding affinity.

HexB: The hydroxyl radical footprinting pattern of mithramycin on HexB supports the DNase I experiments and adjacent TpA steps appear to improve the binding at $(G/C)_4$ and longer sequences, such as in GGGCCCTA (site 2), TAGCCGGC (site 3), TAAGCGCGTA (site 4), TATACGCGTA (site 6) and TACGCGCGTATATA (site 7). TCGCGA (site 1; only on HexBfor) is the only binding site without a nearby TpA step. AAGCTTATA (site 4) is the only footprint at an isolated GpC step, while there is no binding to TACGTA. This supports the previous evidence that GpC steps are preferred over CpG. The CCGGCCG G/C-tract at the extremes of the gels is unclear and so cannot be analysed. However, this would be expected to be a poor site as there is no TpA step nearby this CCGG sequence.

Nogalamycin $[(RY)_n]$

Figure 3.24 shows the results of DNase I footprinting experiments with nogalamycin on HexAfor, HexArev2, HexBfor and HexBrev. In these gels, the footprinting sites are indicated by the filled boxes and these are highlighted in the sequences shown in Figure 3.25. C_{50} values for the interaction with all the sites on these fragments are summarised in Table 3.7.

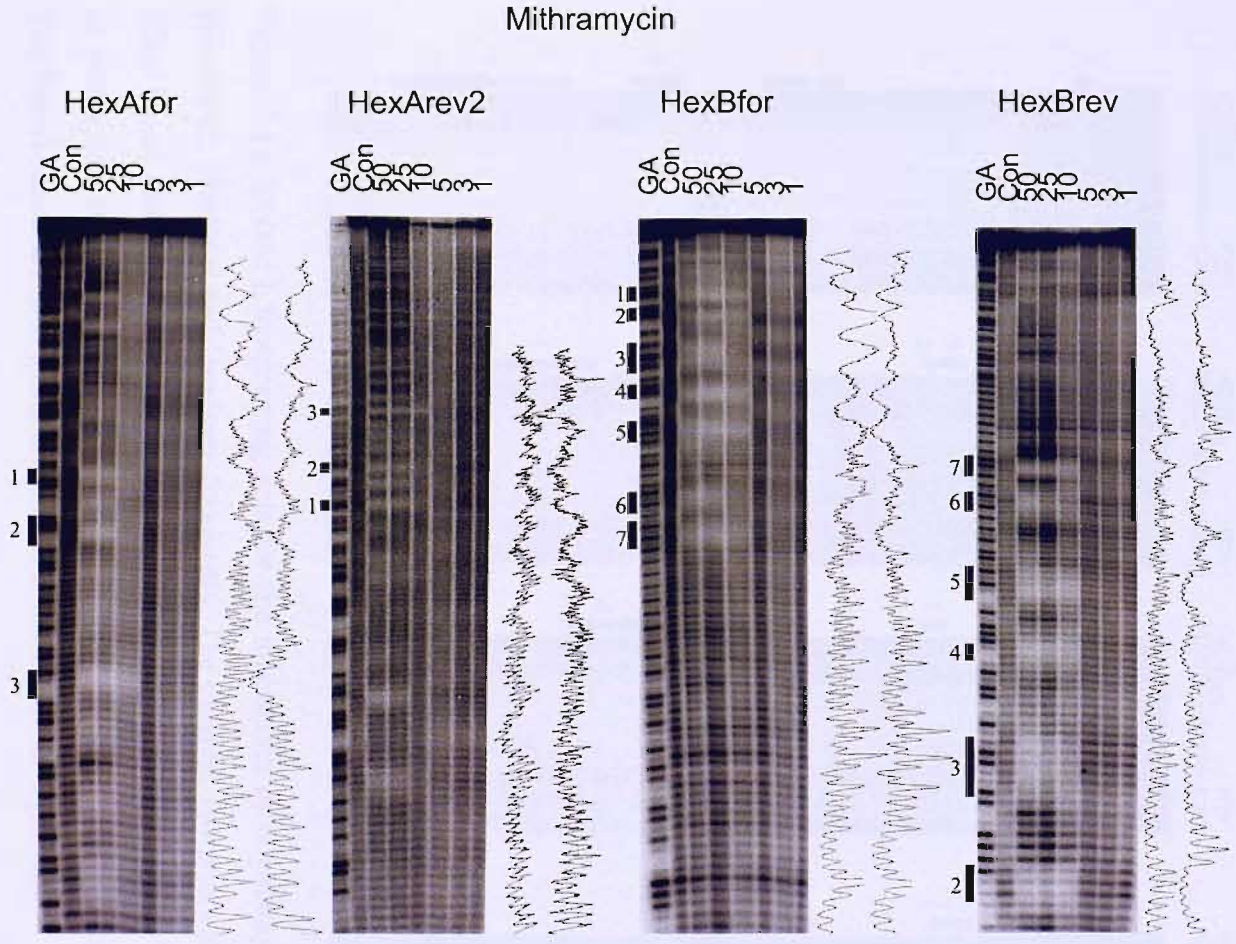


Figure 3.23: [Left] Hydroxyl radical footprinting patterns produced by mithramycin on HexAfor, HexArev2, HexBfor and HexBrev fragments in the presence of 10 mM $MgCl_2$. Only the clearest footprints are marked. Densitometer plots were derived from the 10 μM lanes. Sites are numbered as mentioned in the text. GA is a purine specific marker lane. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.
[Below] Sequence of HexA and HexB fragments highlighting (in red) the hydroxyl radical footprints observed with mithramycin.

Mithramycin: HexAfor (top); HexArev2 (bottom)
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'

Mithramycin: HexBfor (top); HexBrev (bottom)
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'

HexA: Nogalamycin produces non-specific inhibition of DNase I cleavage at concentrations of 3 μM and above. However, at lower concentrations every $(RY)_5$ sequence is protected, such as ATATATG (site 1), TGCAGC (site 2), TGCGCATGC (site 3) and ACGTG (site 5, with CTAG; only clear on HexAfor). Only one $(RY)_4$ site is protected at the lower concentrations; AGTACT (site 4; only clear on HexAfor). On comparing this site to some of those not bound (*e.g.* AGCGCT, CCGCGG and GATATC), it is apparent that binding is best to alternating RY sites that contain all four bases. It is also interesting to note that GGTACC is not bound (while AGTACT is), suggesting at least six base pairs are involved in binding.

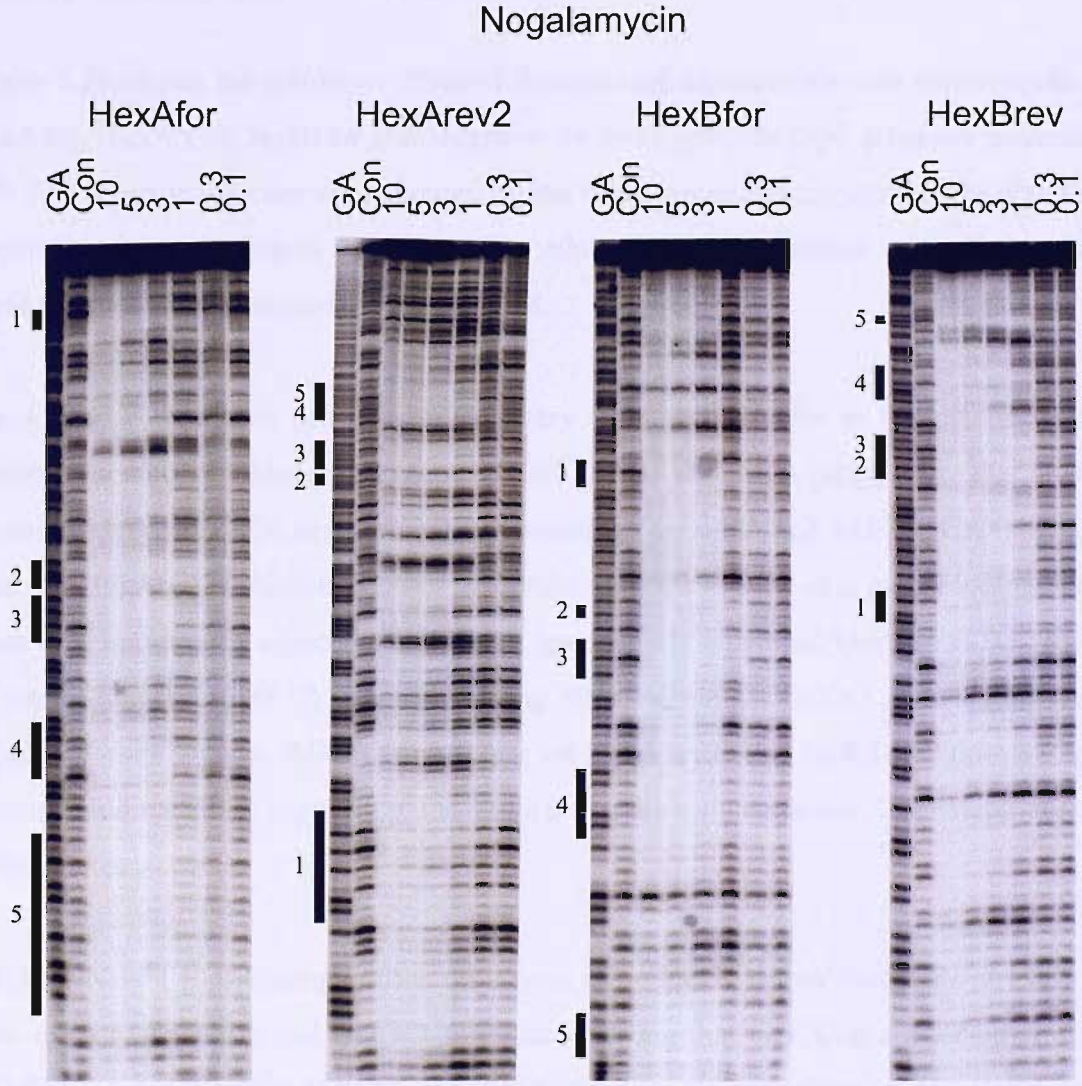


Figure 3.24: DNase I footprinting gels for the interaction of nogalamycin with HexAfor, HexArev2, HexBfor and HexBrev fragments. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.25. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

HexB: 1 μM nogalamycin protects almost every $(RY)_6$ sequence, including TACGTATA (site 1), ATACATAT (site 2; only clear on HexBfor), GTACATGT (site 3; only clear on HexBfor), CATGATCA (site 4) and ATGCATG (site 5). The exceptions are GCGCGC and ACGCGTACGCGCGTA. These two poor sites both contain long alternating G/C-tracts. No $(RY)_4$ sites are protected by 1 μM ligand on HexB (such as TCGCGA, TTGCAA, TTATAA, GACGTC and AATATT), suggesting that longer $(RY)_n$ sequences are required for strong nogalamycin binding.

Actinomycin D [GpC]

Figure 3.26 shows the results of DNase I footprinting experiments with actinomycin D on HexAfor, HexArev2, HexBfor and HexBrev. In these gels, the GpC steps are marked with "*". The footprinting sites are indicated by the filled boxes and these are highlighted in the sequences shown in Figure 3.27. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 3.8.

HexA: Clear footprints can be seen at every GpC on HexAfor in the presence of low concentrations of actinomycin, except AGGCCT and TGGCCA (suggesting that GGCC is a poor binding site). The best binding site contains TGCGCA and AGCTGCA (site 4; only clear on HexArev2). Although other footprints indicate GCGC is a poor binding site, the close (but not directly adjacent) GpC steps have clearly enhanced binding. GTGCAC (site 6) and TAGCTA (site 2) are also strong sites, while CCGCGG (site 3) is weaker. AGCGCT (site 5, with AGCA; only clear on HexArev2) and GGCGCC (site 1) are the poorest binding sites, supporting the theory that directly adjacent GpC steps are weak binding sites.

HexB: DNase I footprinting with actinomycin D on HexB shows footprints at every GpC step, except CGGCCG and GGGCCC (again showing that GGCC is a poor binding site). ATGCAT (site 7) and AAGCTT and TTGCAA (site 4; only clear on HexBrev) are the strongest binding sites, while TCGCGA and AGCT (site 1; only clear on HexBfor) is a much weaker site. It is surprising that AGCGCGCT (site 5) and CGCGCG and ACGCGT (site 6) are good binding sites, even though these GpC steps are located in G/C-rich regions. This suggests multiple adjacent GpC steps are permitted if in a long alternating G/C-tract.

Nogalamycin: HexAfor (top); HexArev2 (bottom)
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'

Nogalamycin: HexBfor (top); HexBrev (bottom)
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'

Figure 3.25: Sequences of the HexA (top) and HexB (above) fragments, indicating the footprints (in red) produced by nogalamycin.

	Nogalamycin				
HexAfor	Site 1	Site 2	Site 3	Site 4	Site 5
	ATATATG	TGCAGC	TGCGCATGC	AGTACT	ACGTG,CATG
	0.3 ± 0.2	0.3 ± 0.05	0.2 ± 0.05	0.2 ± 0.05	0.2 ± 0.05
HexArev2	Sites 5 & 4	Site 3	Site 2	Site 1	
	CATG,CACGT,GTAC	GCATGCGCA	GCTGCA	CATATAT	
	0.4 ± 0.1	0.3 ± 0.1	0.2 ± 0.05	0.3 ± 0.1	
HexBfor	Site 1	Site 2	Site 3	Site 4	Site 5
	TACGTATA	ATACATAT	GTACATGT	CATGATCA	ATGCATG
	0.6 ± 0.2	0.3 ± 0.1	0.2 ± 0.05	0.6 ± 0.3	0.3 ± 0.1
HexBrev	Site 5	Site 4	Sites 3 & 2	Site 1	
	CATGCAT	TGATCATG	ACATGTACATATGTA	TATACGTA	
	0.3 ± 0.1	0.3 ± 0.05	0.2 ± 0.05	0.2 ± 0.1	

Table 3.7: C_{50} values (μM) derived from quantitative analysis of the interaction of nogalamycin with HexA and HexB fragments (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

It is interesting that TGCCGGCT (site 3; only clear on HexBrev) is a weak binding site, while GGGCCC (site 2; only clear on HexBrev) is very poor (possibly explaining why it is not protected on HexBfor). These two sites suggest both $C_n.G_n$ and $G_n.C_n$ (respectively) create poor actinomycin D binding sites.

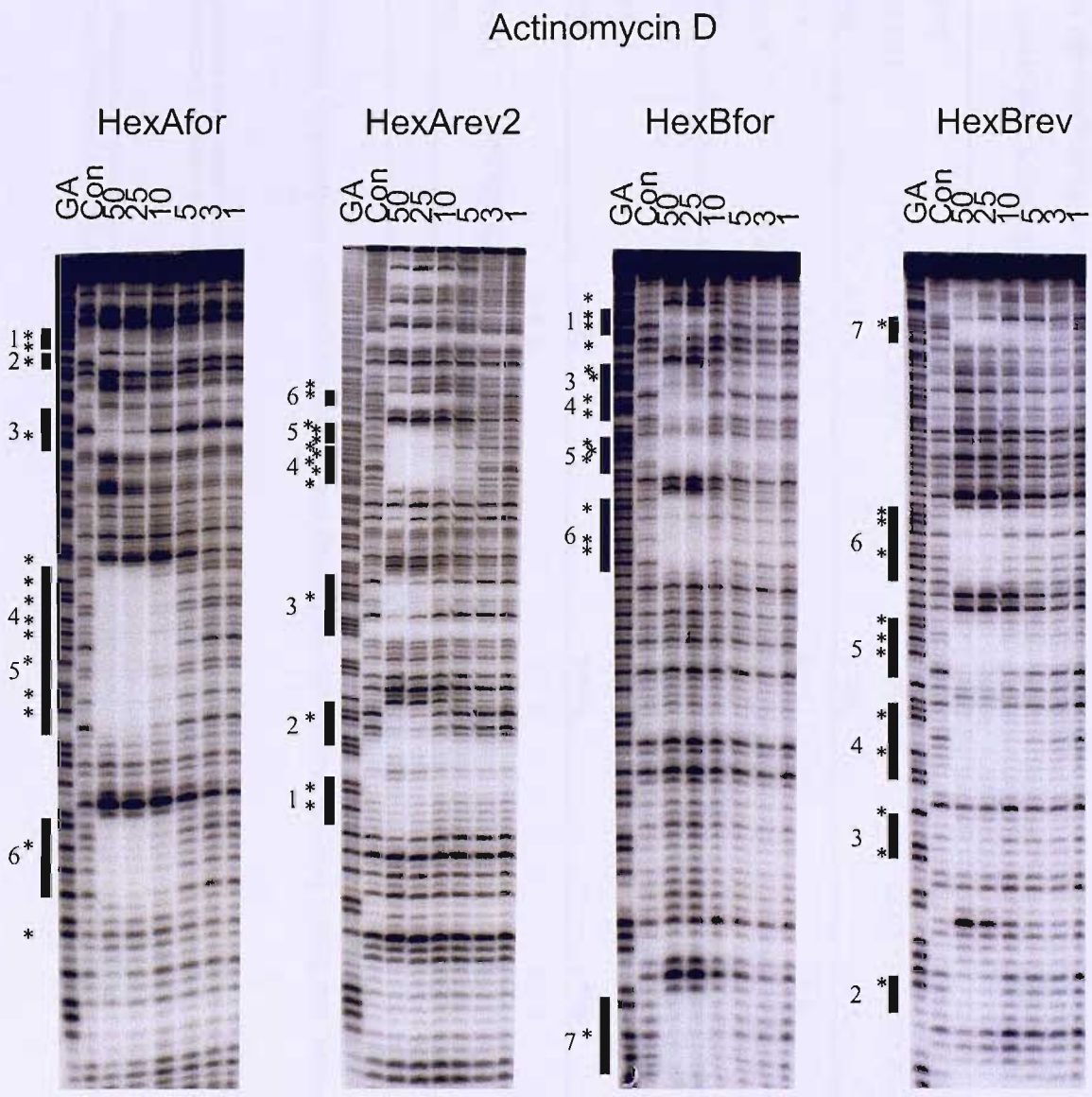


Figure 3.26: DNase I footprinting gels for the interaction of actinomycin D with HexAfor, HexArev2, HexBfor and HexBrev fragments. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.28. Sites are numbered as mentioned in the text and "*" indicates GpC steps. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

Actinomycin D: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

Actinomycin D: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

Figure 3.27: Sequences of the HexA (top) and HexB (above) fragments, indicating the footprints (in red) produced by actinomycin D.

	Actinomycin D					
HexAfor	Site 1	Site 2	Site 3	Sites 4 & 5	Site 6	
	GGCGCC	TAGCTA	CCGCGG	TGCAGCT,TGCGCA,TGCT,AGCGCT	GTGCAC	
	20 ± 6	9 ± 4	9 ± 5	5 ± 3	5 ± 3	
HexArev2	Site 6	Site 5	Site 4	Site 3	Site 2	Site 1
	GTGCAC	AGCGCT,AGCA	TGCGCA,AGCTGCA	CCGCGG	TAGCTA	GGCGCC
	6 ± 5	10 ± 8	4 ± 3	13 ± 10	7 ± 6	12 ± 7
HexBfor	Site 1	Sites 3 & 4	Site 5	Site 6	Site 7	
	TCGCGA,AGCT	AGCCGGCA,TGCA,AGCT	AGCGCGCT	ACGCGT,CGCGCG	ATGCAT	
	25 ± 9	14 ± 5	15 ± 9	17 ± 5	6 ± 3	
HexBrev	Site 7	Site 6	Site 5	Site 4	Site 3	Site 2
	ATGCAT	CGCGCG,ACGCGT	AGCGCGCT	AAGCTT,TTGCAA	TGCCGGCT	GGGCCC
	1.5 ± 0.4	5 ± 1	3.5 ± 1	2 ± 0.5	26 ± 12	51 ± 39

Table 3.8: C_{50} values (μM) derived from quantitative analysis of the interaction of actinomycin D with HexA and HexB fragments (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

Echinomycin [CpG]

Figure 3.28 shows the results of DNase I footprinting experiments with echinomycin on HexAfor, HexArev2, HexBfor and HexBrev. In these gels, all CpG steps are marked with "*". The footprinting sites are indicated by the filled boxes and these are highlighted in the sequences shown in Figure 3.29. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 3.9.

HexA: DNase I footprints can be seen at every CpG step on HexA in the presence of echinomycin. ACGT is the best binding sequence, with AACGTT (site 5; only clear on HexArev2) and CACGTG (site 9) the strongest sites, while WGCGCW is much weaker (TGCGCA (site 7) and AGCGCT (site 8); both only clear on HexAfor, emphasising they are weak sites). A weak footprint is also seen at the similar GGCGCC (site 3; only clear on HexArev2). ATCGAT (site 2; only clear on HexArev2) is a poor binding site, suggesting that TCGA is a weak binding sequence. CCCGGG (site 1; only clear on HexArev2) is a relatively strong binding site, with weaker binding seen at WCCGGW (ACCGGT (site 6; only clear on HexArev2) and TCCGGA (site 4; with CCGCGG)). Interestingly, a weak footprint is evident at CCATGG (site 10; only clear on HexAfor), even though this does not contain a CpG step (binding is probably to CpA).

HexB: Echinomycin generates DNase I footprints at most CpG steps on this fragment (TCGCGA and TTCGAA are the exceptions). TCGACGT (site 8) containing two CpG steps is the best binding site, generating a footprint at lower concentrations than TACGTA (site 5). The weaker site CTCGAG (site 2) is better than GCCGGC (site 3), while no footprint is evident at TTCGAA (suggesting at least six base pairs affect echinomycin binding, probably due to differences in the local DNA structure). Sites that contain two directly adjacent CpG steps such as GCGCGC (site 4) and TCGCGA (to which the ligand does not bind) produce weaker footprints than two nearby sites, such as CCGGCCGA (site 1; only clear on HexBfor) and TCGACGT (site 8). However, ACGCGCGT (site 6, with ACGCGT) is a very strong binding site, indicating that three directly adjacent CpG steps can produce a good footprint. It is interesting to note that sequences containing two or more CpA steps (and no CpG) are also bound well, and two nearby CpA/TpG steps (ACATATGT; site 7) are stronger than three contiguous CpA steps (CCATGCAT; site 9).

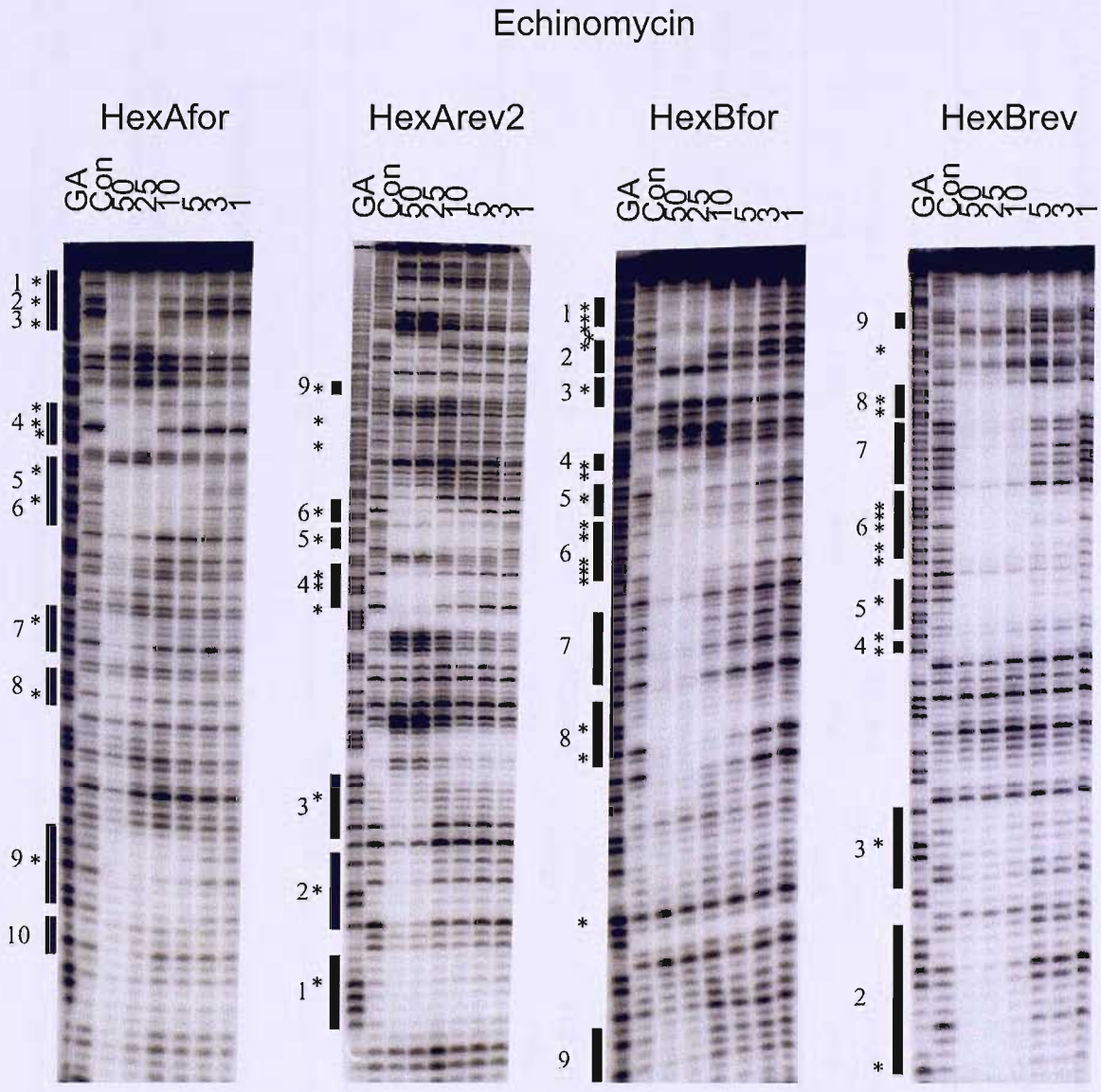


Figure 3.28: DNase I footprinting gels for the interaction of echinomycin with HexAfor, HexArev2, HexBfor and HexBrev fragments. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 3.29. Sites are numbered as mentioned in the text and "*" indicates CpG steps. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

Echinomycin: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

Echinomycin: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

Figure 3.29: Sequences of the HexA (top) and HexB (above) fragments, indicating the footprints (in red) produced by echinomycin.

	Echinomycin								
HexAfor	Sites 1, 2 & 3 CCCGGG,TCGA, GCGC 12 ± 4	Site 4 TCCGGA, CCGCGG 8 ± 4	Sites 5 & 6 AACGTT, ACCGGT 1.5 ± 0.8	Site 7 TGCGCA 28 ± 21	Site 8 AGCGCT 57 ± 47	Site 9 CACGTG 1 ± 0.4	Site 10 CCATGG 35 ± 29		
HexArev2	Site 9 CACGTG 0.7 ± 0.1	Site 6 ACCGGT 9 ± 3	Site 5 AACGTT 1 ± 0.2	Site 4 CCGCGG, TCCGGA 8 ± 3	Site 3 GGCGCC 27 ± 26	Site 2 ATCGAT 11 ± 6	Site 1 CCCGGG 1 ± 0.3		
HexBfor	Site 1 CCGGCCGA 3 ± 1	Site 2 CTCGAG 5 ± 2	Site 3 GCCGGC 11 ± 5	Site 4 GCGCGC 52 ± 50	Site 5 TACGTA 1 ± 0.5	Site 6 ACGCGT, ACGCGCGT 1 ± 0.8	Site 7 ACATATGT 3 ± 1.5	Site 8 TCGACGT 0.5 ± 0.4	Site 9 ATGCATGG 12 ± 9
HexBrev	Site 9 CCATGCAT 15 ± 8	Site 8 ACGTCGA 0.7 ± 0.07	Site 7 ACATATGT 4 ± 2	Site 6 ACGCGCGT, ACGCGT 0.5 ± 0.07	Site 5 TACGTA 0.2 ± 0.05	Site 4 GCGCGC 13 ± 6	Site 3 GCCGGC 5 ± 3	Site 2 CTCGAG 6 ± 4	

Table 3.9: C_{50} values (μM) derived from quantitative analysis of the interaction of echinomycin with HexA and HexB fragments (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

Discussion

The universal footprinting substrates developed in this chapter should be useful for studying novel ligands with unknown sequence selectivity, as these contain every combination of di-, tri-, tetra- and symmetrical hexanucleotides. In this chapter, seven well-characterised ligands have been footprinted against universal dinucleotide (DiN), trinucleotide (TriN), tetranucleotide (MS1/MS2) and symmetrical hexanucleotide (HexA and HexB) sequences (all of these, except MS1/MS2 have been developed during this work). These studies were performed in order to demonstrate the use of these fragments and to discover any novel insights into the interaction of these ligand with their preferred sites, when flanked by different combinations of base pairs.

Distamycin

Distamycin, which was one of the first natural polyamide ligands to be discovered, has long been known to bind to A/T-rich sequence and is selective for $(A/T)_4$ sites (Pelton and Wemmer, 1989; Van Dyke *et al.*, 1982).

The results with DiN confirm that a four base pair A/T-tract (TTAT) is a better binding site than shorter A/T-tracts. ApT steps (ATT) are also favoured over TpA steps (TAA; supporting Abu-Daya *et al.*, 1995), and the ligand binds to sites with two such steps (*e.g.* TAT) less well than those with one (TAA and ATT).

DNase I footprinting experiments with TriN support these findings. It is interesting to note that while two ApT or TpA steps are unfavourable, sites with one ApT step are better than simple A_n tracts (supporting Fagan and Wemmer, 1992). This is most likely due to the way in which the curved structure of the ligand matches the helical shape of these different sites.

The experiments with MS1/MS2 support these findings, with ApT steps preferred over TpA and $(A/T)_4$-tracts preferred over $(A/T)_3$ and $(A/T)_2$ sites. TTTT is one of the best sites along with AATTA, suggesting that although ATT is better than AAA (results with TriN) $A_n.T_n$ is bound well if n is greater than 3.

The presence of GC base pairs between A/T-tracts will affect the local DNA structure and the 2-amino group of guanine will also sterically inhibit binding (Patel, 1982). However, some GC base pairs have been tolerated in long A/T-tracts on MS1/MS2, probably by the ligand traversing the GC base pair (as observed by Kopka *et al.*, 1985),

although this has weakened binding affinity.

The novel symmetrical hexanucleotide fragments (HexA and HexB) have enabled further exploration of the preferred binding sites for distamycin. $(A/T)_6$-tracts are generally favoured over $(A/T)_4$ (Abu-Daya and Fox, 1997). AAATTT is the best binding site on this fragment, consistent with the suggestion that A_nT_n is a strong target (Coll *et al.*, 1987). The hydroxyl radical footprints suggest that distamycin does not cover the entire sequence and it appears to be located in the central portion, around the ApT step (Fagan and Wemmer, 1992), as opposed to the $A_n.T_n$ tracts at either end. The preference of distamycin for binding to long A/T-tracts (longer than the ligand itself) is most likely due to the local DNA structure.

It is interesting to note that alternating $(AT)_n$ base pairs are a better binding site than TTAA (though this can be bound if it is close to other A/T-tracts). TATA appear to be a better binding site than TTAA, even though it possesses two TpA steps (supporting work by Abu-Daya *et al.*, 1995). The inclusion of an ApT step seems to reduce the unfavourable DNA structure that is generated by the TpA step.

Hoechst 33258

Hoechst 33258 is a synthetic minor groove binding ligand, which has long been known to be selective for $(A/T)_4$ (Martin and Holmes, 1983; Harshman and Dervan, 1985).

DNase I footprinting experiments on DiN showed binding only at ATT, and that TpA steps inhibit ligand binding. In this sequence an $(A/T)_4$-tract containing a single TpA step is not protected, indicating that Hoechst 33258 has a more exact sequence binding requirement than distamycin. This has previously been suggested by Breusegem *et al.* (2002) and is most likely due to the more rigid structure of Hoechst 33258.

The results with TriN again demonstrated that the ligand does not bind well to sites that contain TpA steps even if a ApT step is present (ATA), and Abu-Daya *et al.* (1995) have previously shown that there is a 50-fold difference in binding between AATT and TATA. ATT is the preferred site on this fragment, with AAA weaker. It therefore appears that, unlike distamycin, Hoechst 33258 binds to ApT better than ApA, though this is much better than TpA.

The footprinting studies on MS1/MS2 support these results. It is interesting to note that $(AT)_4$-tracts are generally preferred over $(A/T)_3$-tracts (Abu-Daya and Fox, 1997), though TTAA is the worst binding site (Abu-Daya *et al.*, 1995; Breusegem *et al.*, 2002).

This supports the suggestion that the lack of ApT steps weakens Hoechst 33258 binding (Rosu *et al.*, 2002).

As observed with distamycin, lengthening the nucleotide step on the DNA target to include every symmetrical hexanucleotide (HexA and HexB) has shown Hoechst 33258 binds to six base pair A/T-tracts better than four base pair tracts (supporting Abu-Daya and Fox (1997)). TpA steps generally reduce the affinity for A/T-tracts, but these steps are tolerated better in longer A/T-tracts. Taking into account that only two hydrogen bonds are possible with Hoechst 33258 (Harshman and Dervan, 1985) and that hydroxyl radical footprinting shows sites shorter than the A/T-tracts to be bound, these TpA steps are probably not involved directly in binding on the longer sequences (seen by Bostock-Smith *et al.* (2001)). The weaker binding to sites with nearby TpA steps is most likely due to the effect of this dinucleotide on the local DNA structure (Abu-Daya *et al.*, 1995). Unlike distamycin, Hoechst 33258 does not bind to sites in which GC base pairs interrupt the A/T-tracts, as suggested by Bailly *et al.* (1993).

Mithramycin

The aureolic acid anti-tumour antibiotic mithramycin preferentially binds in the DNA minor groove to $(G/C)_4$ sequences (Sastry and Patel, 1993; Carpenter *et al.*, 1993).

The results with DiN confirm the G/C-selectivity. However, the observation that ACCGCT is favoured over TCCGCT indicates that an ApC step at the 5'-end of the target site is preferred over TpC. This is possibly because the purine/pyrimidine step distorts the local DNA structure by widening the minor groove, so allowing the mithramycin saccharide units to bind more tightly (as suggested by Cons and Fox (1989) and Sastry *et al.* (1995)).

DNase I footprinting experiments with TriN indicate that mithramycin binds better to longer G/C-tracts, with CGCCC and CCG favoured over GC, as suggested by Carpenter *et al.* (1993). GTCC is a weak binding site and interruption by the thymine base clearly disrupts mithramycin binding.

Footprinting studies with MS1/MS2 support these findings, confirming that longer G/C-tracts generally create stronger binding sites. All the $(G/C)_4$-tracts bind the ligand with similar affinities. The only exception is CCGG, which is a weak binding site, suggesting that CpG steps reduce the binding affinity, consistent with the results of Carpenter *et al.* (1993).

Mithramycin also generally does not bind to CCGG on HexA and HexB. The results with these fragments suggest that the ligand binds better when there are more GpC steps than CpG (Carpenter *et al.*, 1993; Sastry and Patel, 1993). Alternating G/C residues are also preferred over G_nC_n (even though they contain CpG steps). Unlike most other ligands studied in this chapter, mithramycin does not bind better to longer G/C-tracts and $(G/C)_6$ sites have comparable affinities to $(G/C)_4$. However, it is thought that the ligand prefers a distorted DNA structure around its $(G/C)_4$ binding site, allowing the saccharide groups to bind well in the groove (Sastry and Patel, 1993).

The structure of DNA near the primary binding site is clearly very important, and adjacent TpA steps appear to enhance binding, maybe by creating a "kink" in the DNA and widening the minor groove, thereby allowing better interaction with the saccharide groups (Sastry *et al.*, 1995). CCGG is only bound when it is adjacent to a TpA step, as too are shorter $(G/C)_2$ regions. Placing a TpA step on both sides of a G/C-tract increases binding strength even more, though TpA steps that are not contiguous with the G/C-tract do not enhance the binding affinity.

Nogalamycin

Nogalamycin is a threading intercalating agent that targets $[YR]_n$ sequences (Fox and Waring, 1986; Searle *et al.*, 1988; Gao *et al.*, 1990; Robinson *et al.*, 1990; Zhang and Patel, 1990). No specific footprints were observed on DiN, though the enhanced cleavage seen at CCG suggests that this is a site to which the ligand binds less well, even though it has previously been shown to bind to CpG (Gao *et al.*, 1990; Robinson *et al.*, 1990; Fox and Alam, 1992) and another CCG site present on the fragment is bound well. Fox and Alam (1992) suggest runs of pyrimidine or purine base pairs are weaker binding sites. We therefore suggest that the longer sequence context (TTCCG) at the enhanced CCG site is responsible for preventing binding.

Mainly non-specific binding is observed with TriN. However, the observation that a GAAAG sequence shows enhanced DNase I cleavage at high concentrations supports the suggestion that contiguous adjacent purines inhibit nogalamycin binding.

The footprinting experiments with MS1 and MS2 show some selective binding to regions of alternating purine and pyrimidine bases which contain a mixed base composition, as suggested by Fox and Waring (1986).

HexA and HexB give further insight into nogalamycin's selectivity and it is

apparent that the best binding sites contain $[RY]_n$, although the only $[RY]_4$ site bound on HexA and HexB at low nogalamycin concentrations is AGTACT, confirming the suggestion that regions of mixed base composition are preferred. Long $[GC]_n$ tracts are poor binding sites since nogalamycin requires DNA breathing before it can bind, and GpC is the hardest dinucleotide to unstack (Ornstein *et al.*, 1978). Surrounding G/C-tracts will stabilise the DNA duplex and will also hinder the interaction with regions that otherwise might have been good sites.

Actinomycin D

Actinomycin D binds to DNA by intercalating its phenoxazone chromophore between a GpC step (Kamitori and Takusagawa, 1994; Chen *et al.*, 2003a). The initial DNase I footprinting on DiN confirmed the interaction with GpC steps, with alternating R/Y steps preferred around the GpC binding site (Scamrov and Beabealashvilli, 1983).

The results with TriN showed that contiguous C or G around the GpC step produce weaker binding sites (CGCC compared to AGCA). The observation that GTCC is also bound (though weakly) is interesting. A similar result was obtained by Wadkins *et al.* (2000) and Chen *et al.* (2004) examining the binding of actinomycin to a single-stranded DNA sequence. It is possible that the central AT base pair in the GTC site is looped out, allowing binding to the remaining GpC. Conversely, it is also possible that the central T is merely tolerated by the ligand, which is binding to the guanine alone, though the presence of two 2-amino groups is thought to be mandatory for the interaction of actinomycin with double stranded DNA (Jennewein and Waring, 1997). Interestingly, the GTC sequence in MS1/MS2 is not bound by actinomycin, but some footprints with MS1/MS2 do not contain GpC steps, and the ligand appears to be binding (albeit weakly) to GpT/CpA and ApT sites. This suggests that the gross structure of the target site may be more important than hydrogen bond formation at these alternative motifs (Chen *et al.*, 2004).

The results with MS1/MS2 confirm that placing RpY steps adjacent to GpC produces better binding sites (TGCA is a better site than AGCA) and long purine or pyrimidine tracts reduce the binding. Two contiguous GpC steps are weak binding sites, either due to competition (exclusion) between adjacent actinomycin molecules (Scott *et al.*, 1988) or because the structure of a G/C-tract is not optimal for binding (Orstein *et al.*, 1978).

The footprinting experiments on HexA and HexB confirm that TGCA is a good

binding site, and that this is much better than AGCT. CCGCGG is a poor binding site, supporting the suggestion that G/C-tracts create a rigid structure that is not optimal for strong ligand binding, and two adjacent GpC binding sites again inhibit binding, although nearby but non-adjacent GpC sites show good binding (Scott *et al.*, 1988). The observation that GGCC and CCGG (adjacent to a GpC step) create poor actinomycin D binding reinforces the evidence that G/C-tracts generally inhibit binding (Scamrov and Beabealashvilli, 1983). It is interesting to note, however, that three contiguous GpC steps show good binding. This is most likely because exclusion is not a factor in this case (two of the three possible intercalation sites are available for binding) and so cooperative binding may be occurring to overcome any DNA structural limitations.

Echinomycin

The bis-intercalator echinomycin is known to bind selectively to CpG steps (Wakelin and Waring, 1976; Van Dyke and Dervan, 1984; Ughetto *et al.*, 1985; Low *et al*, 1984b). The results with DiN confirm that binding is to CpG, though a CpA/TpG step is also bound (albeit weakly). This suggests that the presence of a CpG step may not be an absolute requirement for binding and that other DNA structural features may generate potential binding sites. (Waterloh and Fox, 1991b). It is interesting to note that TCCGCT is a better binding site than ACCGCT, indicating that echinomycin binding is sensitive to the sequence and/or structure over at least six base pairs.

DNase I footprinting experiments with TriN show that although every CpG is targeted, CTAG is also weakly bound. The interaction with some CpA and TpA steps is most likely due to recognition of a feature of DNA structure, rather than base specific binding, in the same way that TANDEM recognises TpA steps (Bailly and Waring, 1998). This is consistent with the observation that TpD is a better binding site that CpG (D = diaminopurine) (Tseng *et al.* (2005) even though they present similar groups in the DNA minor groove.

The footprinting studies on MS1/MS2 again show that although all the CpG sites generate footprints, there is some weaker binding to CpA/TpG. It is clear that multiple nearby CpG steps enhance binding (this cooperative binding was also seen by Gilbert and Feigon (1992) and Bailly *et al.* (1996)), but directly adjacent CpG steps produce poor binding sites (Wakelin and Waring, 1976). As suggested for actinomycin D, G/C-tracts may generally be too rigid for strong binding. Alternatively two molecules of echinomycin

cannot bind side-by-side for steric reasons, as expected from the size of the echinomycin binding site (four base pairs).

The results with HexA and HexB show that ACGT is the best binding site and that TCGA is much weaker. This may be related to the observation that triostin A causes adenine bases around the CpG core to adopt a *syn* conformation, so generating Hoogsteen AT base pairs (Wang *et al.*, 1984). Gao and Patel (1988) showed that ACGT involves Hoogsteen base pairing whilst TCGA uses Watson-Crick pairing, and that ACGT generates the best binding sites. However, Sayers and Waring (1993) suggested that this difference in base conformation has a negligible effect on binding strength. It is interesting to note that ATCGAT is a stronger site than TTCGAA, suggesting that the flanking bases are also important in ligand binding, although it is unclear whether TACGTA binds differently to AACGTT.

Two contiguous CpG steps are again seen to weaken binding on the HexA and HexB fragments, but sequences containing three contiguous CpG steps show good binding. This is probably because the outermost CpG sites in CGCGCG are bound by two echinomycin molecules, which do not interfere with each other, in contrast to the two sites in CGCG (Gilbert and Feigon, 1992).

General Conclusions

Universal oligonucleotide sequences are a valuable tool for footprinting studies to establish the approximate binding selectivity and affinity of novel DNA binding ligands. The development of novel minimal length universal dinucleotide, trinucleotide and symmetrical hexanucleotide sequences has enabled binding studies on ligands containing various size binding sites and sequence selectivities.

The ligands tested in this chapter have already been widely studied on other natural and synthetic footprinting substrates and these studies support or expand on the selectivity described in the literature.

The fragments presented in this chapter generate good DNase I and hydroxyl radical cleavage patterns, with few regions of poor cleavage. Now that both HexArev and HexBrev have been cloned and sequenced correctly (see Appendix), a useful catalogue of DNA substrates is available for future use to test novel ligands with completely unknown sequence specificity

A complete universal hexanucleotide footprinting sequence is unlikely to be prepared due to the large number of steps (2080) making it difficult to design, and the fragment length would be much too long for conventional footprinting studies. However, with algorithms for the design of this fragment now discovered (Anderson *et al.*, 2006), the main factor inhibiting the use of such a long sequence is in the gel electrophoresis and footprinting reactions themselves. Single hit kinetics would also be very difficult to achieve, on average requiring only one cut per 2000 base pairs. Cutting such a designed fragment into several smaller sections might be one approach, but would be very time-consuming and labour-intensive, as many multiple footprinting reactions would need to be carried out in parallel. Another technique therefore needs to be developed. Fluorescence footprinting might be one such technique, where DNA is fluorescently labelled instead of radiolabelled. Running such a long fragment on a conventional capillary sequencing machine would allow fluorescence detection of the footprinting sites. The position of a footprint could easily be located by observing how far the DNA had flowed through the capillaries and binding affinity could be approximated by using an appropriate range of ligand concentrations. The main drawback from this method is that with radiolabelled DNA, only attomolar quantities of radiation are present in each band on a gel. Fluorophores are currently not able to be visualised at this kind of concentration, so the DNA would have to be concentrated substantially before running, so making calculations of binding affinity very difficult.

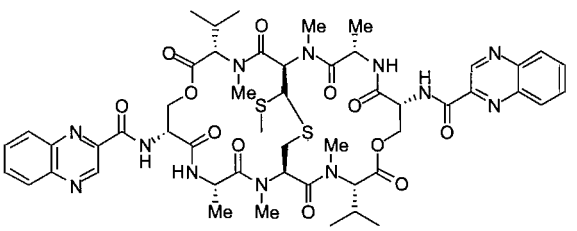
A more immediate use for these novel fragments footprinting substrates would be to investigate the wide range of so-called well-classified ligands to establish further details of their selectivity and to elaborate on any secondary binding sites.

Chapter IV

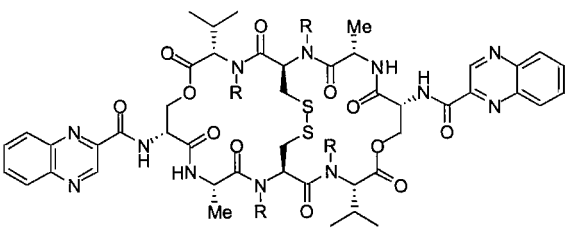
Novel Analogues of the Bis-Intercalator TANDEM

Introduction

Echinomycin (1) and triostin A (2) are members of the quinoxaline family of antitumour antibiotics that bind to DNA by bis-intercalation (Waring and Wakelin, 1974), inserting their chromophores on either side of CpG (Low *et al.*, 1984b). Sequence selectivity is achieved by the formation of hydrogen bonds between the carbonyl groups of the alanines and the 2-amino groups of guanines. It has also been demonstrated that the bases on either side of the CpG affect the binding, such that echinomycin prefers to bind ACGT (Van Dyke and Dervan, 1984). **T**riostin **A** ***N*-Dem**ethylated (TANDEM; 3) is a synthetic analogue of triostin A that lacks the N-methyl groups, which are present on four of the peptide bonds in the parent compound. This small alteration results in a dramatic change in sequence selectivity and TANDEM binds selectively to TpA steps, particularly ATAT (Lavesa *et al.*, 1993; Lavesa and Fox, 2001). This occurs because the alanine carbonyl groups form internal hydrogen bonds to the previously methylated amide groups, so removing the ability to bind to CpG.



1. Echinomycin



2. Triostin A; R = Me

3. TANDEM; R = H

TANDEM has previously been synthesised using relatively demanding solution-phase chemistry. However, solid-phase synthesis of TANDEM has recently been developed (Malkinson *et al.*, 2005), allowing the synthesis of a series of analogues with modified chromophores and amino-acid backbones. This chapter examines the sequence selectivity of these novel TANDEM analogues, as well as confirming the properties of the TANDEM that has been prepared by this new method. In the results described below, TANDEM refers to the newly synthesised compound.

TANDEM Analogues

The structures of the TANDEM analogues used in this work are shown in Figure 4.1.

Disulphide bridge modification

Bis-acm TANDEM

The synthetic precursor to TANDEM has a broken disulphide bridge, in which the two sulphurs are protected by acetamide groups.

Modification of intercalating groups

Hemi-naphthyl TANDEM

TANDEM has two quinoxaline ring systems for bis-intercalating into DNA. One of these rings has been substituted for naphthalene to form hemi-naphthyl TANDEM.

Naphthyl TANDEM

Both quinoxaline ring systems of TANDEM have been substituted for naphthalenes to form naphthyl TANDEM.

Mono-quinoxaline TANDEM

In this derivative, one of the quinoxaline intercalating groups has been removed and replaced with an acetyl group.

Bis-acetate TANDEM

In this derivative both the quinoxaline rings of TANDEM have been replaced with acetate.

Octadepsipeptide ring Val^4 and Val^8 modification

Biotinylated [Lys^4,Lys^8]TANDEM

This ligand was made with the intention of using it in nanotechnology to create nanostructures from DNA complexes. The ligand would be targeted to a synthetic piece of DNA, where the biotin molecules extending from the TANDEM octadepsipeptide ring could then be bound by streptavidin or avidin, allowing for more similar nanostructures to adhere, so building up a structure in whatever shape desired. One such nanostructure that could be bound by the biotinylated [Lys^4,Lys^8]TANDEM-DNA complex is a gold-DNA nanowire discussed in Ongaro *et al.* (2004), acting in this case as a scaffolding.

In this derivative, biotin molecules are attached to lysine residues that have been placed at positions 4 and 8 in the octadepsipeptide ring.

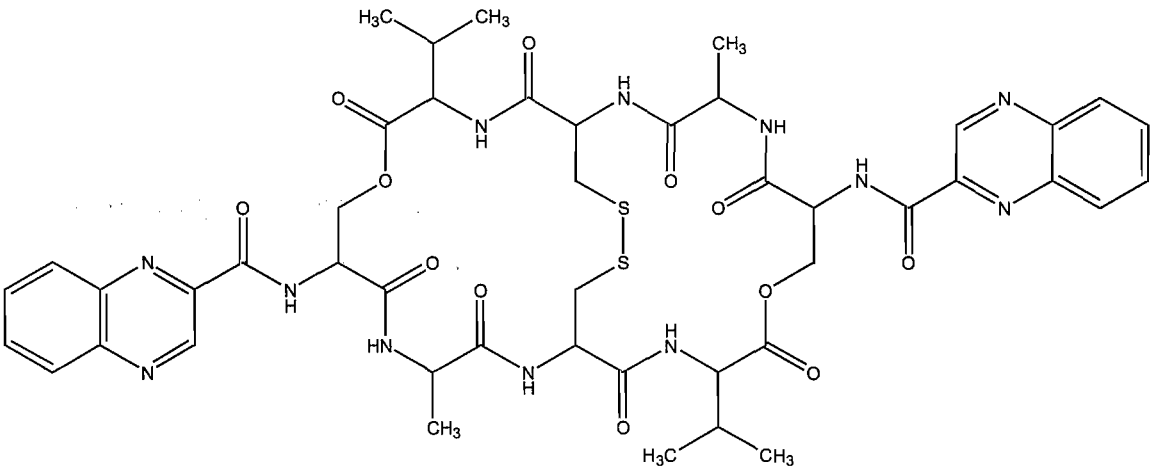
Benzylated [Lys^4,Lys^8]TANDEM

The synthetic precursor to biotinylated [Lys^4,Lys^8]TANDEM contains benzyl groups on the ends of the side chains of the lysines at positions 4 and 8.

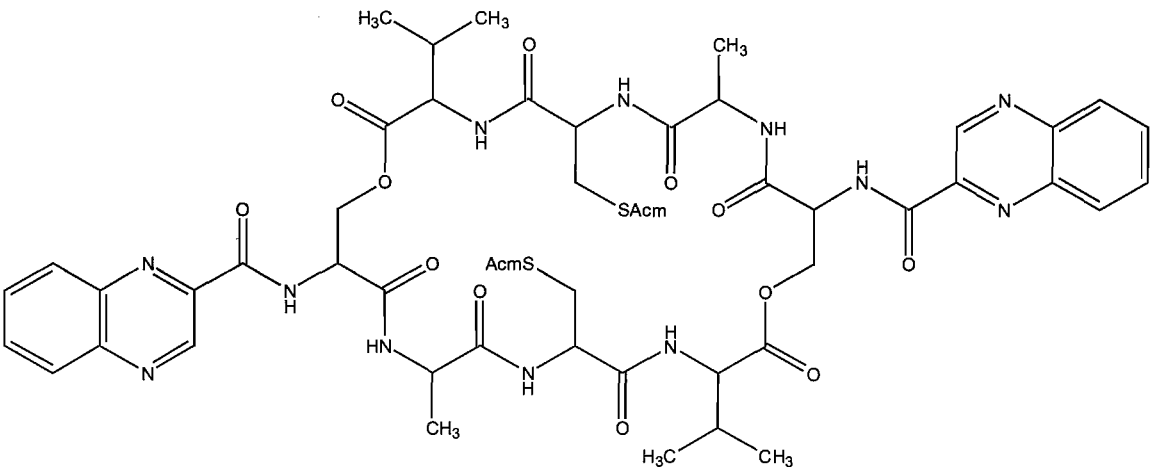
[Lys^4,Lys^8] TANDEM

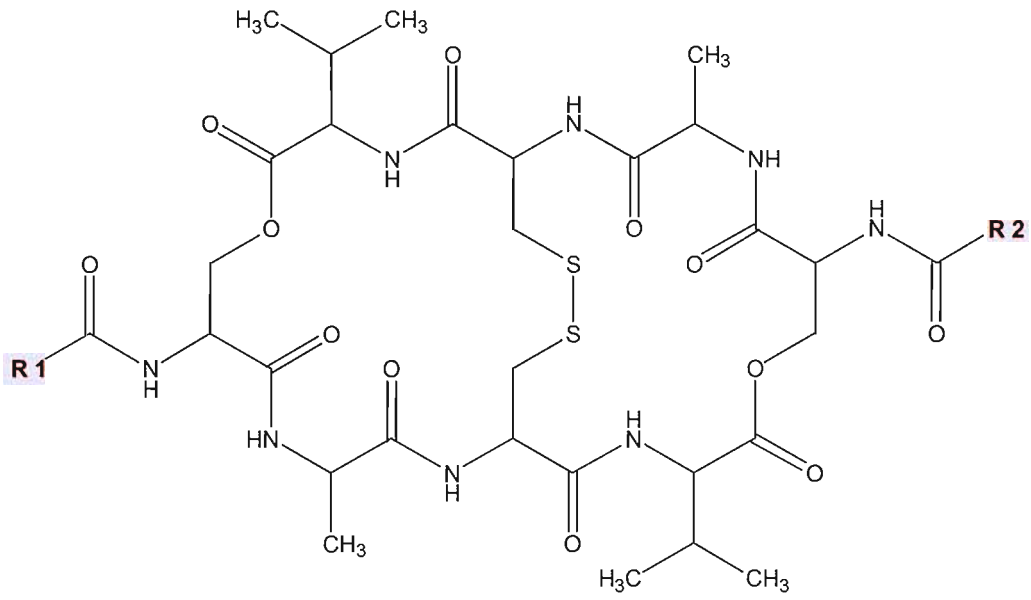
This synthetic precursor to benzylated [Lys^4,Lys^8]TANDEM, is similar to TANDEM, but with Val^4 and Val^8 substituted by lysine residues.

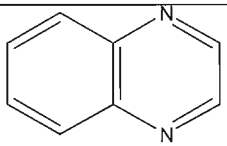
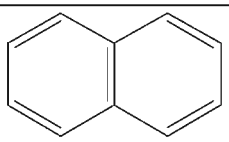
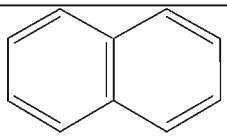
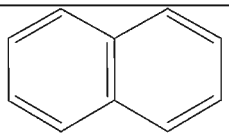
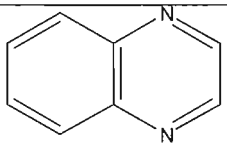
TANDEM

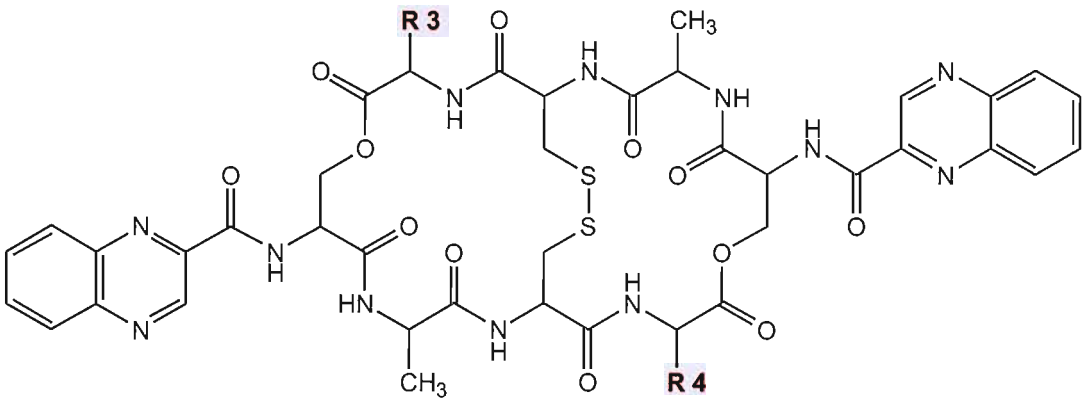


Bis-acm TANDEM





	R1	R2
Hemi-naphthyl TANDEM		
Naphthyl TANDEM		
Mono-quinoxaline TANDEM		**CH_3**
Bis-acetate TANDEM	**CH_3**	**CH_3**



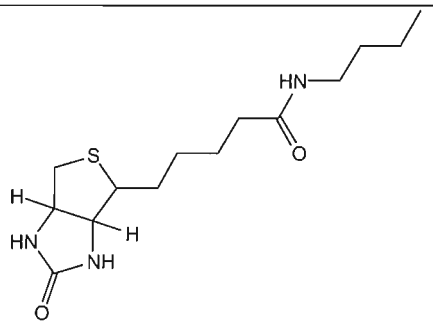
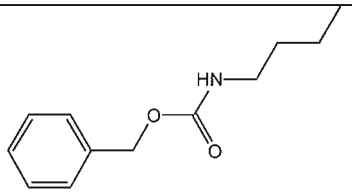
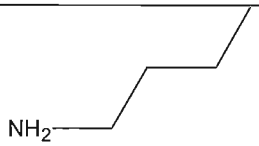
	R3 / R4
Biotinylated [Lys^4,Lys^8]TANDEM	
Benzylated [Lys^4,Lys^8]TANDEM	
[Lys^4,Lys^8]TANDEM	

Figure 4.1: Molecular structures of TANDEM; Bis-acm TANDEM; Hemi-naphthyl TANDEM; Naphthyl TANDEM; Mono-quinoxaline TANDEM; Bis-acetate TANDEM; Biotinylated [Lys^4,Lys^8]TANDEM; Benzylated [Lys^4,Lys^8]TANDEM; and [Lys^4,Lys^8]TANDEM.

Results

The observed footprints presented in this chapter often encompass two or more possible binding sites. In such cases, the analysis is focussed on the primary binding site and the other potential sites are indicated in parentheses.

TANDEM

Before examining the properties of the novel TANDEM derivatives, the sequence selectivity of the newly synthesised TANDEM was confirmed. The first footprinting studies with this compound using the *tyr*T DNA fragment (Low *et al.*, 1984a) showed binding to ATAT and ATAA. The results of DNase I footprinting experiments with the new TANDEM on this fragment (actually *tyr*T(43-59)) are shown in the first panel of Figure 4.2. It can be seen that there is a single clear footprint at the ATAT site (the ATAA site to which binding was observed in earlier work is not present in this shorter *tyr*T fragment). Weaker footprints are also observed at GTAA and TTAG. These results confirm the previous observations and demonstrate that the newly synthesised TANDEM binds to TpA. Similar experiments with fragment pAAD1 are shown in the central panel of Figure 4.2. This fragment contains several different $(A/T)_4$ sites and is similar to the fragments used by Lavesa *et al.* (1993). It can be seen that TANDEM produces a clear footprint at ATAT, with a weaker site at TATA. There is no interaction with the other $(A/T)_4$ sites, including TTAA. This is again consistent with the previous observations. SASK1 (designed for the studies with the Series B minor groove binding ligands in chapter 5) contains a single TpA step (CTAG) and DNase I footprinting experiments with this fragment (third panel of Figure 4.2) show a single footprint at this site. The sequences of these fragments can be found in the Materials and Methods chapter (chapter 2).

These results confirm that the newly synthesised TANDEM has the same properties as the original material and binds to TpA, especially when this is flanked by A/T residues. C_{50} values, derived from quantitative analysis of these footprints are shown in Table 4.1 and show that this ligand binds to the best of these sites with a C_{50} of about 10 μM, consistent with previous observations.

MS1/MS2

Figure 4.3 shows the results of DNase I footprinting experiments with TANDEM and the novel derivatives on the MS1 (left four panels) and MS2 (right four panels) DNA fragments, with TpA steps marked "+". These fragments, which contain the same sequence

in opposite orientations, contain every possible tetranucleotide. The location of the observed footprints is highlighted in Figure 4.4 and the C_{50} values derived for these sites are summarised in Table 4.2.

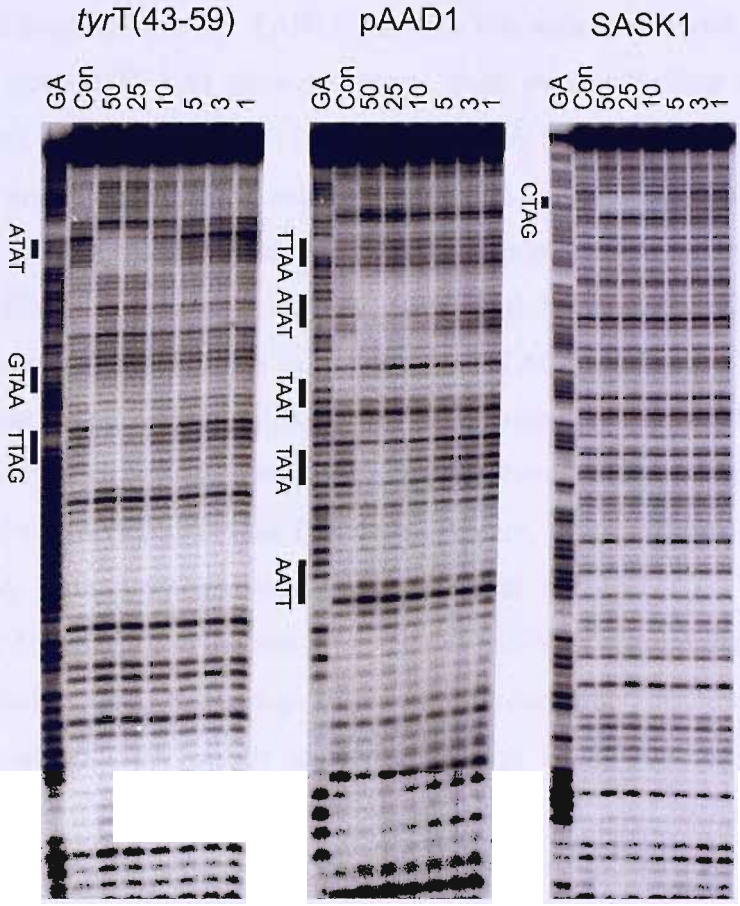


Figure 4.2: DNase I footprinting gels for TANDEM on *tyr*T(43-59), pAAD1 and SASK1. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

	TANDEM			
*tyr*T(43-59)	ATAT 12 ± 5	GTAA, TTAG 45 ± 27		
pAAD1	TTAA -	ATAT 4 ± 1	TAAT -	TATA 6 ± 1
SASK1	CTAG 12 ± 7			

Table 4.1: C_{50} values (μM) for the interaction of: TANDEM with TpA containing binding sites on *tyr*T, pAAD1 and SASK1 fragments. The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). Sites marked "–" correspond to TpA sites that showed no cleavage protection with TANDEM at the concentrations tested.

The interaction of [N-MeCys^3,N-MeCys^7]TANDEM with these sequences has previously been reported (Lavesa and Fox, 2001) and the results with TANDEM confirm these results. It can be seen that ATAT (site 1) and GTATAG (site 4, with CTAT and TTAG) are the preferred binding sites for TANDEM, with C_{50} values of about 8-10 μM. All the other TpA sites (except TTAA) show footprints, with weaker binding to ATAC (site 2), CTAG and GTAG (site 3) and TTAT, GTAA and GTAC (site 5).

It can be seen that the Hemi-naphthyl TANDEM produces much weaker footprints than TANDEM. Several of the weaker TANDEM footprints are no longer evident with this derivative and DNase I cleavage is only attenuated by this ligand at the strongest TANDEM binding sites (ATAT (site 1), CTAG and GTAG (site 2) and GTATAG (site 3)). It seems that substitution of one quinoxaline ring by naphthalene has drastically reduced the binding affinity and only the strongest sites are bound. Experiments with Naphthyl TANDEM failed to affect the DNase I cleavage pattern, even at ligand concentrations as high as 200 μM, with a weak attenuation evident at ATATGTATATA on MS1 at a concentration of 500 μM (not shown). The removal of both nitrogens in the quinoxaline ring has most likely altered the charge distribution around the ring, thereby affecting its ability to intercalate. Similar results to Hemi-naphthyl TANDEM can be seen with the Biotinylated [Lys^4,Lys^8]TANDEM. The footprints are much weaker than with TANDEM and only the best TANDEM sites are occupied. ATAT (site 1) and TATA (site 3) are still the preferred binding sites and the selectivity seems to be unaltered (CTAG and GTAG (site 2) show a weaker footprint). Benzylated [Lys^4,Lys^8]TANDEM produced a very similar effect to Naphthyl TANDEM (not shown); only very weak binding was evident at ATATGTATATA with 500 μM ligand on MS1.

In contrast to the other derivatives, [Lys^4,Lys^8]TANDEM alters the DNase I cleavage patterns at much lower concentrations than TANDEM. The footprints are located in similar regions, but are evident at much lower concentrations. As observed with the parent compound, ATAT (site 1) is the best binding site with good binding also observed at TATA (site 2). It appears that [Lys^4,Lys^8]TANDEM binds more strongly than TANDEM, but without sacrificing the selectivity for TpA (the same sites show footprints on both ligands (not labelled for [Lys^4,Lys^8]TANDEM)).

Bis-acm TANDEM, Mono-quinoxaline TANDEM and Bis-acetate TANDEM did not produce any changes in the DNase I cleavage patterns, even at concentrations as high as 500 μM. These results demonstrate the importance of the chromophores and the intact disulphide cross-bridge.

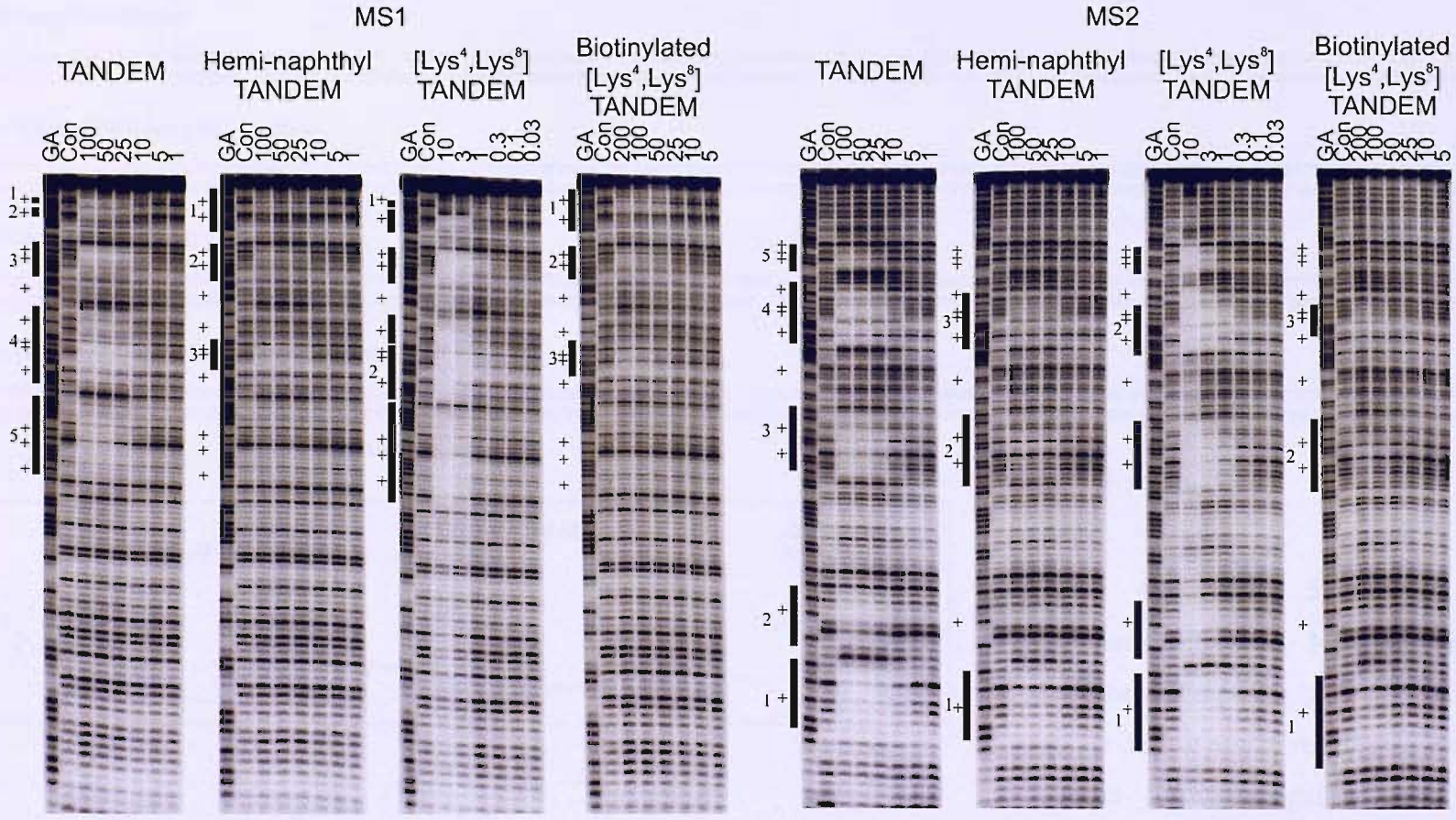


Figure 4.3: DNase I footprinting gels for TANDEM, Hemi-naphthyl TANDEM, [Lys^4,Lys^8]TANDEM and Biotinylated [Lys^4,Lys^8]TANDEM on the MS1/MS2 DNA fragments. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The footprints are indicated by the boxes and the sequence of the fragments is shown in Figure 4.4. Gels are numbered as mentioned in the text and TpA steps are marked "+".

TANDEM: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

Hemi-naphthyl TANDEM: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

[Lys^4,Lys^8] TANDEM: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

Biotinylated [Lys^4,Lys^8] TANDEM: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

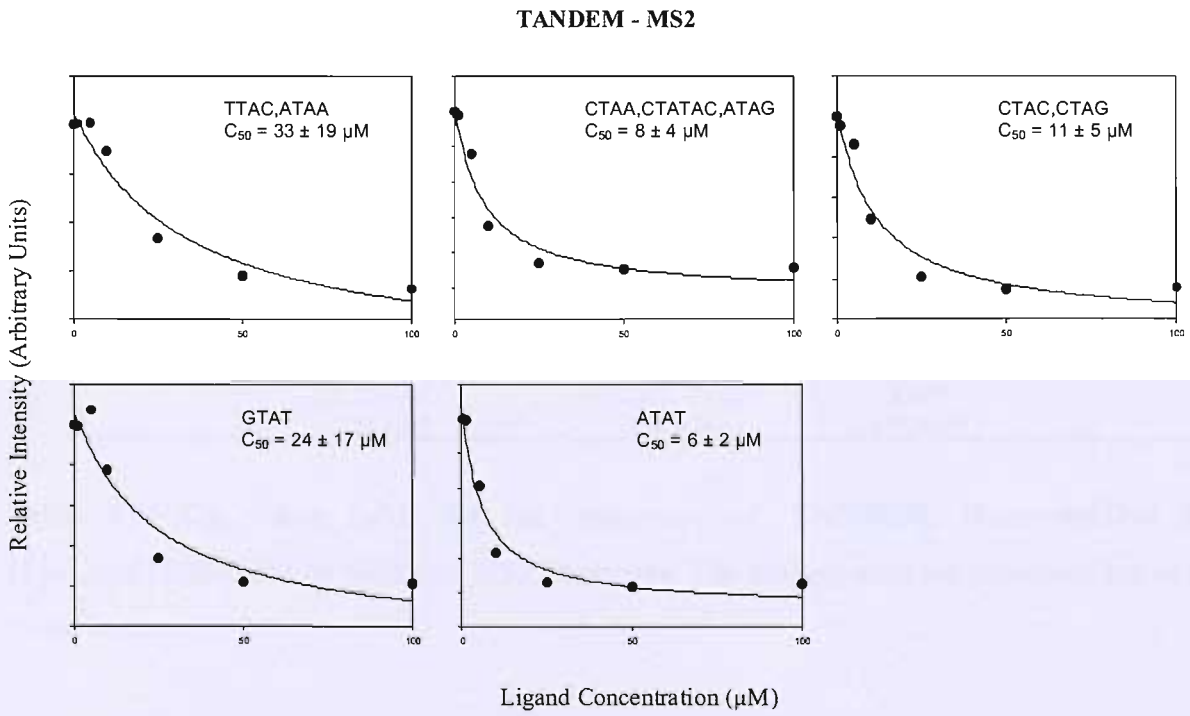


Figure 4.4: [Top] Sequences of fragments MS1 and MS2 showing the footprinting sites for TANDEM and its derivatives in red. These were derived from the gels shown in Figure 4.3. [Left] Example of footprinting plots for TANDEM binding to MS2.

			TANDEM			
MS1	Site 1 ATAT 10 ± 3	Site 2 ATAC 21 ± 11	Site 3 CTAG, GTAG 12 ± 4	Site 4 CTAT, GTATAG, TTAG 8 ± 3	Site 5 TTAT, GTAA, GTAC 19 ± 10	
MS2	Site 5 TTAC, ATAA 33 ± 19	Site 4 CTAA, CTATAC, ATAG 8 ± 4	Site 3 CTAC, CTAG 11 ± 5	Site 2 GTAT 24 ± 17	Site 1 ATAT 6 ± 2	
			Hemi-naphthyl TANDEM			
MS1	Site 1 ATAT, ATAC 7 ± 4	Site 2 CTAG, GTAG 13 ± 6	Site 3 GTATAG 18 ± 10			
MS2	Site 3 CTAA, CTATAC, ATAG 8 ± 2	Site 2 CTAC, CTAG 20 ± 7	Site 1 ATAT 14 ± 8			
			[Lys^4,Lys^8]TANDEM			
MS1	Site 1 ATAT 0.04 ± 0.02	ATAC 1 ± 0.4	CTAG, GTAG 0.6 ± 0.2	CTAT 2 ± 1	Site 2 GTATAG, TTAG 2.5 ± 1.5	TTAT, GTAA, GTAC 2 ± 1
MS2	TTAC, ATAA 3 ± 1	Site 2 CTATAC, ATAG 0.3 ± 0.05	CTAC, CTAG 0.5 ± 0.1	GTAT 1.4 ± 0.4	Site 1 ATAT 0.2 ± 0.04	
			Biotinylated [Lys^4,Lys^8]TANDEM			
MS1	Site 1 ATAT, ATAC 21 ± 12	Site 2 CTAG, GTAG 35 ± 14	Site 3 GTATAG 45 ± 17			
MS2	Site 3 CTATAC 156 ± 47	Site 2 CTAC, CTAG 284 ± 264	Site 1 ATAT 154 ± 117			

Table 4.2: C_{50} values (μM) for the interaction of: TANDEM; Hemi-naphthyl TANDEM; [Lys^4,Lys^8]TANDEM; and Biotinylated [Lys^4,Lys^8]TANDEM on MS1 and MS2 fragments. The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

HexA and HexB

Figure 4.5 shows the results of DNase I footprinting experiments with TANDEM and its derivatives on HexAfor (left four panels) and HexArev2 (right four panels), with TpA steps marked "+". These fragments (together with HexBfor and HexBrev, which are shown in Figure 4.7) contain every symmetrical hexanucleotide sequence. The binding sites for these ligands are highlighted on the sequence of HexA in Figure 4.6 along with some representative footprinting plots for [Lys^4,Lys^8]TANDEM. The C_{50} values for the interaction of these ligands with all their binding sites on these fragments are summarised in Table 4.3.

It can be seen that TANDEM has bound to every TpA step on these fragments, except TTAA (in both TTTAAA and GTTAAC). The best binding site is at ATATAT (site 1), while CTAG (site 2) seems to be especially weak.

As with MS1/MS2, it can be seen that Hemi-naphthyl produces weaker footprints than TANDEM, which appear at much higher ligand concentrations and generally only result in attenuated DNase I cleavage. However, the selectivity seems to be unaffected and ATATAT (site 1) is still the preferred binding site. A similar effect is seen with Biotinylated [Lys^4,Lys^8]TANDEM, which also shows weak binding, with ATATAT as the preferred site (site 1). Napthyl TANDEM only produced a weak attenuation of DNase I cleavage around the sequence GATATCGATATAT at 500 μM ligand (not shown). No binding was observed with Bis-acm TANDEM, Mono-quinoxaline TANDEM or Bis-acetate TANDEM.

[Lys^4,Lys^8]TANDEM produced the same footprinting pattern as TANDEM, but required much lower concentrations (often less than 1 μM). ATATAT (site 1) is still the preferred binding site and no interaction is seen at TTAA, except at 10 μM for which there appears to be non-specific binding.

Figure 4.7 shows the results of similar experiments with TANDEM and its derivatives on HexBfor (left four panels) and HexBrev (right four panels), with TpA steps marked "+". The binding sites for these ligands are highlighted on the sequence of HexB in Figure 4.8 along with some representative footprinting plots for [Lys^4,Lys^8]TANDEM. The C_{50} values for the interaction of these ligands with all their binding sites on these fragments are summarised in Table 4.4.

It is seen that, as expected, TANDEM produces a footprint at most TpA steps. The exceptions are TTAA and CGCGTACGCGCG, the latter having alternating G/C-tracts that

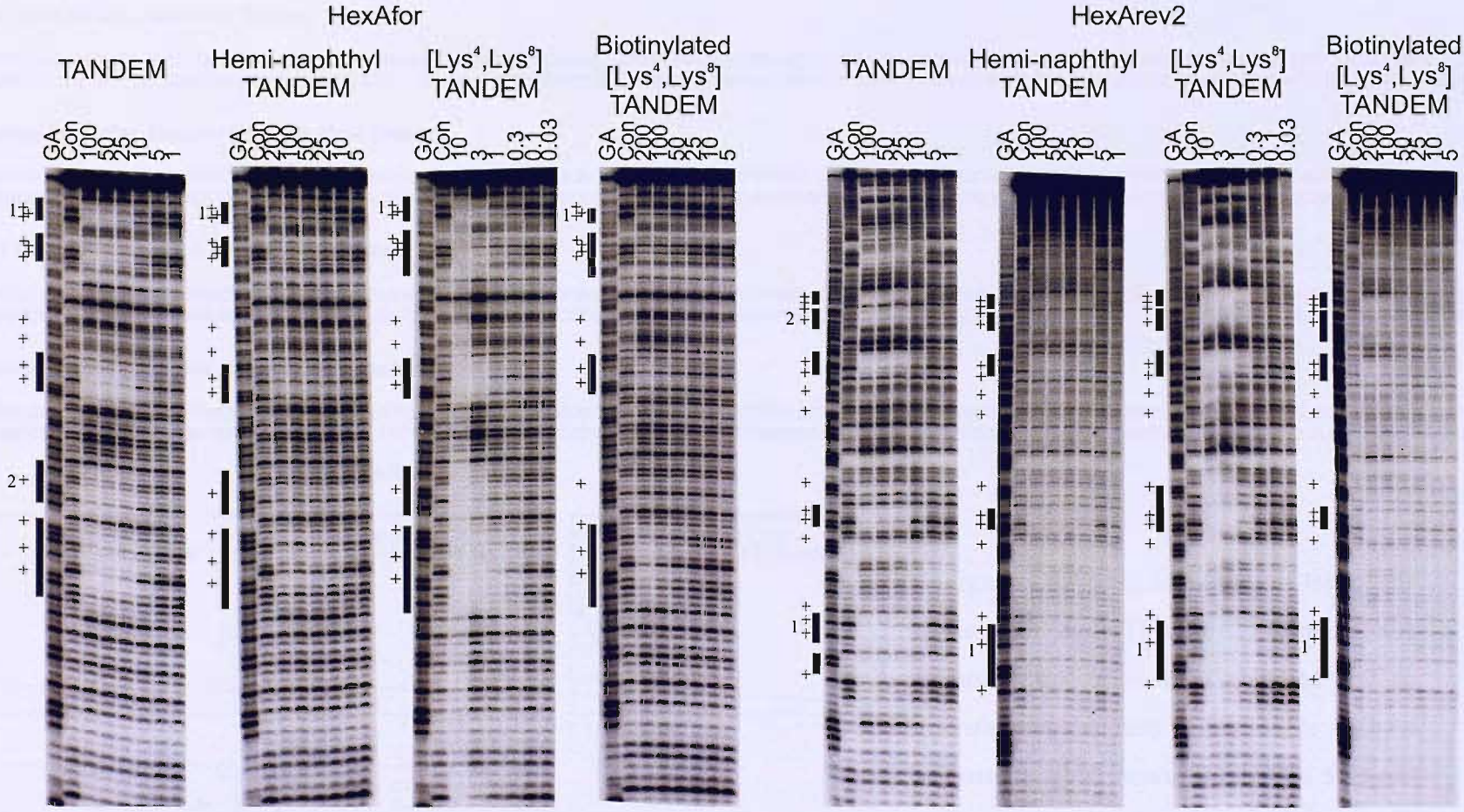


Figure 4.5: DNase I footprinting gels for TANDEM, Hemi-naphthyl TANDEM, [Lys^4,Lys^8]TANDEM and Biotinylated [Lys^4,Lys^8]TANDEM with the HexAfor and HexArev2 DNA fragments. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The footprints are highlighted by the filled boxes and TpA steps are marked "+". Gels are numbered as mentioned in the text.

TANDEM: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

Hemi-naphthyl TANDEM: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

[Lys^4,Lys^8] TANDEM: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

Biotinylated [Lys^4,Lys^8] TANDEM: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

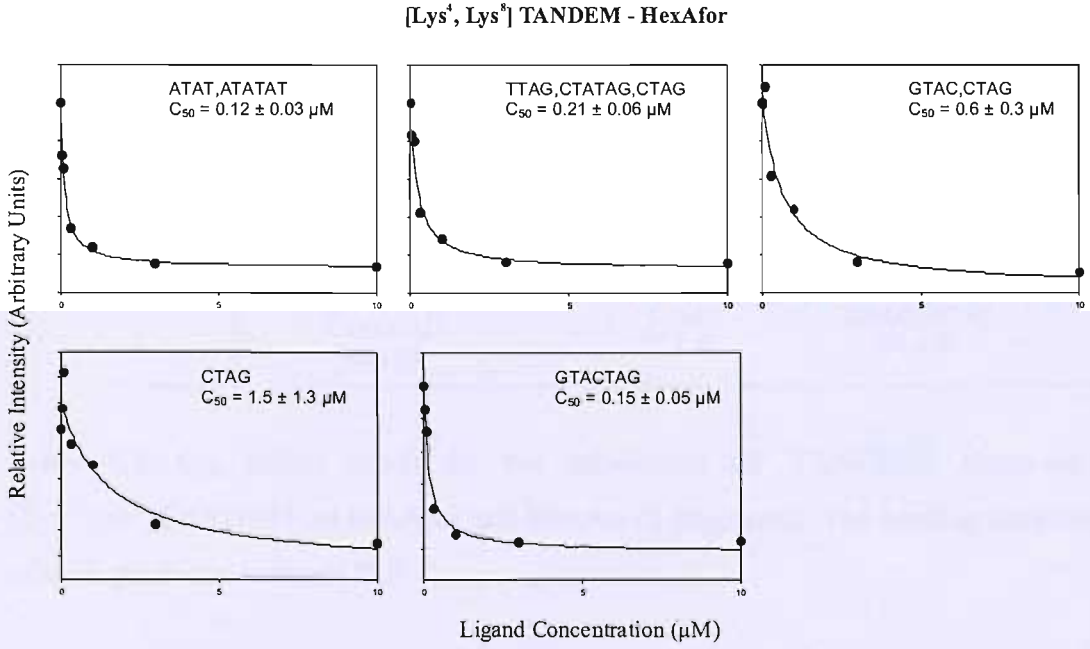


Figure 4.6: [Top] Sequences of fragments HexAfor and HexArev2 showing the footprinting sites for TANDEM and its analogues (in red). These were derived from the gels shown in Figure 4.5. [Left] Example of footprinting plots for [Lys^4,Lys^8]TANDEM binding to sites in this fragment.

TANDEM						
HexAfor	Site1 ATAT, ATATAT 6 ± 2	TTAG, CTATAG, CTAG 7 ± 4	GTAC, CTAG 11 ± 6	Site 2 CTAG 15 ± 9	GTACTAG 7 ± 4	
HexArev2	CTAGTAC 18 ± 8	Site 2 CTAG 22 ± 11	CTAG, GTAC 16 ± 8	CTATAG 13 ± 8	Site 1 ATATAT 7 ± 3	ATAT 17 ± 6
Hemi-naphthyl TANDEM						
HexAfor	Site 1 ATAT, ATATAT 8 ± 4	TTAG, CTATAG, CTAG 11 ± 5	GTAC, CTAG 17 ± 10	CTAG 24 ± 13	GTACTAG 10 ± 6	
HexArev2	CTAGTAC 3.3 ± 2.5	CTAG 2.6 ± 1.8	CTAG, GTAC 5.5 ± 3.8	CTATAG 4.3 ± 3	Site 1 ATATAT 2 ± 1.3	
[Lys^4,Lys^8]TANDEM						
HexAfor	Site 1 ATAT, ATATAT 0.1 ± 0.03	TTAG, CTATAG, CTAG 0.2 ± 0.06	GTAC, CTAG 0.6 ± 0.3	CTAG 1.5 ± 1.3	GTACTAG 0.15 ± 0.05	
HexArev2	CTAGTAC 0.17 ± 0.06	CTAG 0.4 ± 0.2	CTAG 0.5 ± 0.2	CTATAG 0.5 ± 0.2	Site 1 ATATAT 0.12 ± 0.02	
Biotinylated [Lys^4,Lys^8]TANDEM						
HexAfor	Site 1 ATATAT 39 ± 8	TTAG, CTATAG, CTAG 43 ± 8	GTAC, CTAG 118 ± 33	GTACTAG 75 ± 24		
HexArev2	CTAGTAC 42 ± 24	CTAG 71 ± 41	CTAG, GTAC 61 ± 30	CTATAG 43 ± 22	Site 1 ATATAT 29 ± 13	

Table 4.3: C_{50} values (μM) for the interaction of: TANDEM; Hemi-naphthyl TANDEM; [Lys^4,Lys^8]TANDEM; and Biotinylated [Lys^4,Lys^8]TANDEM on HexAfor and HexArev2 fragments. The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

will make the DNA too rigid for strong ligand binding. GTATATAC (site 1, with ATAT) and AATATT (site 2) are the best binding sites. Examination of the C_{50} values for these sequences and HexA suggests that YTAR sites are amongst the weakest sites. Once again much higher concentrations of Hemi-naphthyl TANDEM are required to produce a footprint, though in this case clear regions of protection can be seen at GTATATAC (site 3, with ATAT) and AATATT (site 4), with slightly weaker binding at TTATAA (site 2) and CTAA and TTAG (site 1). It is clear that Hemi-naphthyl TANDEM has the same sequence selectivity as the parent compound, though it binds less well.

As seen with the HexA substrates, Biotinylated [Lys^4,Lys^8]TANDEM binds less well than TANDEM, though again the binding sites are similar to those of Hemi-naphthyl TANDEM. [Lys^4,Lys^8]TANDEM again binds better than TANDEM, producing footprints at sub-micromolar concentration at similar sites. As with HexA, no binding was observed with Bis-acm TANDEM, Mono-quinoxaline TANDEM, Benzylated [Lys^4,Lys^8]TANDEM and Bis-acetate TANDEM.

TANa/TANb

These sequences correspond to the two labelled strands of a synthetic DNA fragment that was designed to contain a single binding site for TANDEM (ATATAT). The results of DNase I footprinting experiments with TANDEM and the best of its derivatives on this fragment are shown in Figure 4.9, with TpA steps marked "+". Note that, in contrast to all the other footprinting substrates, this fragment was labelled at the 5'-end of each strand. The footprinting site for these ligands is highlighted on the sequence of this fragment in Figure 4.10, along with the footprinting plots. Their C_{50} values are summarised in Table 4.5.

It can be seen that all these ligands produce a single footprint at the ATATAT site, which is evident on both strands, though with very different affinities. The C_{50} for TANDEM is about 3-4 μM; Hemi-naphthyl TANDEM is about 10-12 μM, Biotinylated [Lys^4,Lys^8]TANDEM is about 30 μM, while [Lys^4,Lys^8]TANDEM is about 0.1 μM, approximately 30-40 times stronger than the parent compound, TANDEM.

Addition of avidin to biotinylated [Lys^4,Lys^8]TANDEM

A range of concentrations of avidin (25, 50 and 100 μM) was added to the DNA-bound Biotinylated [Lys^4,Lys^8]TANDEM (200 and 100 μM) before DNase I cleavage to see if it was possible to form a biotin-avidin complex and so increase the footprinting site size.

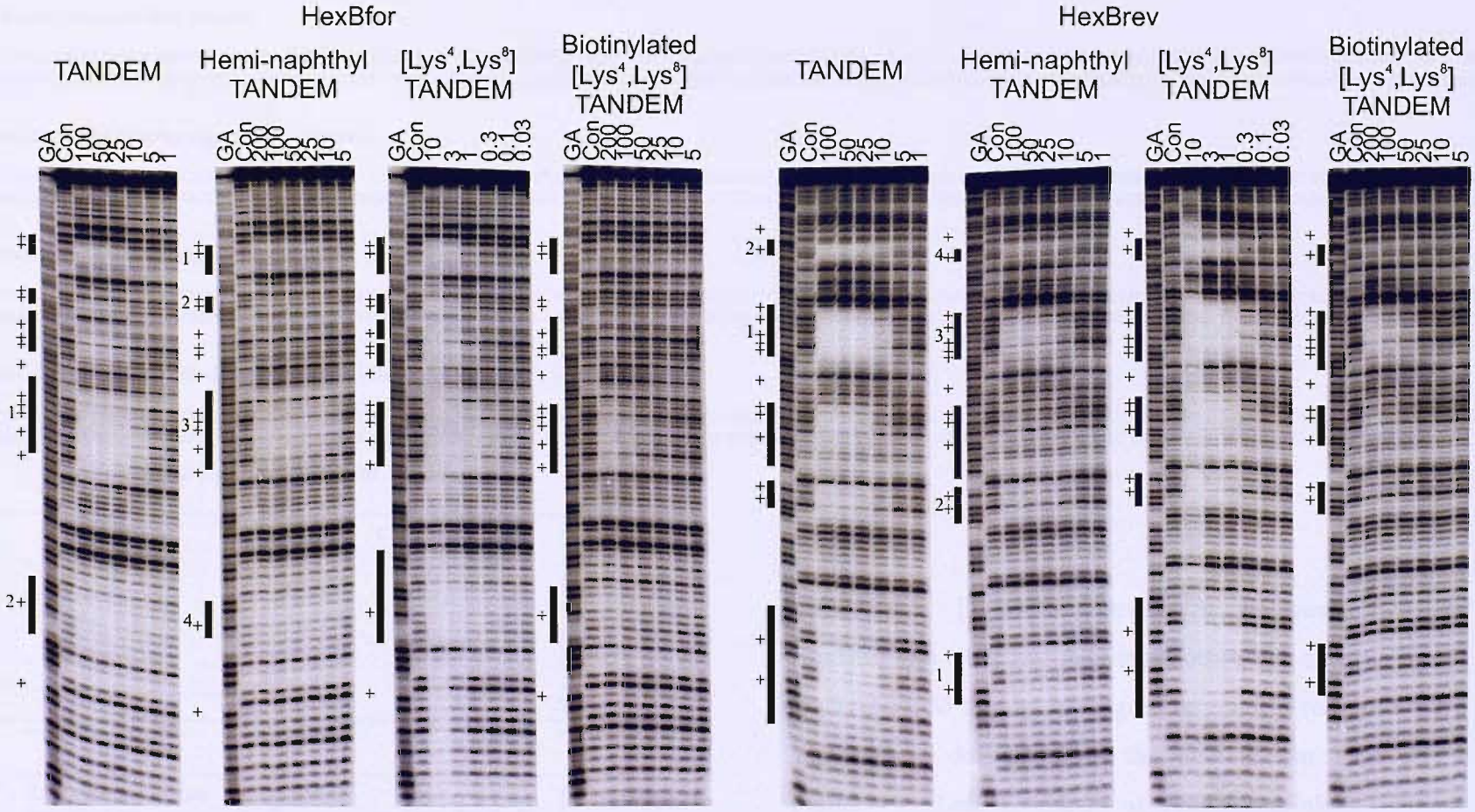


Figure 4.7: DNase I footprinting gels for TANDEM, Hemi-naphthyl TANDEM, [Lys^4,Lys^8]TANDEM and Biotinylated [Lys^4,Lys^8]TANDEM on the HexBfor and HexBrev DNA fragments. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The footprints are marked with filled boxes and TpA steps are marked "+". Gels are numbered as mentioned in the text.

TANDEM: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

Hemi-naphthyl TANDEM: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

[Lys^4,Lys^8] TANDEM: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

Biotinylated [Lys^4,Lys^8] TANDEM: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

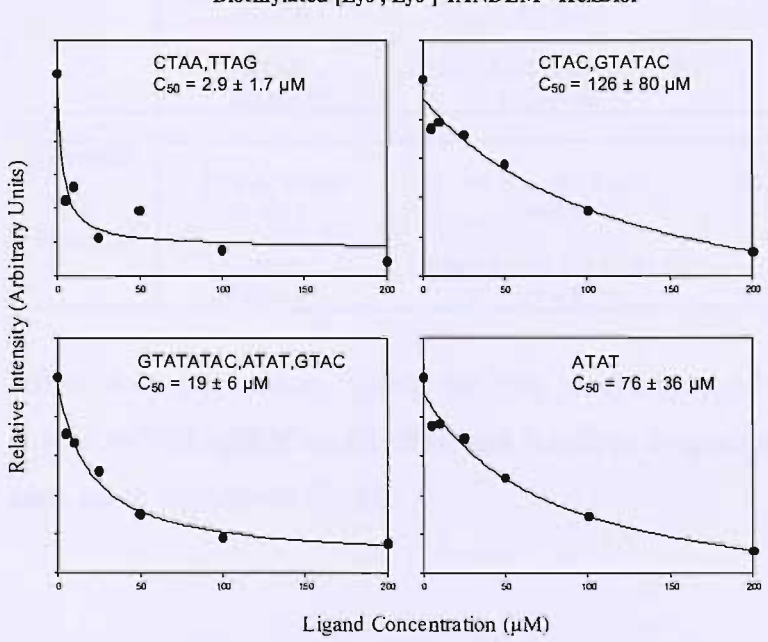


Figure 4.8: [Top] Sequences of fragments HexBfor and HexBrev showing footprinting sites for TANDEM and its analogues (shown in red). These were derived from the gels shown in Figure 4.7. [Left] Examples of footprinting plots for the interaction of biotinylated [Lys^4,Lys^8]TANDEM with sites on this fragment.

TANDEM						
HexBfor	CTAA, TTAG 15 ± 7	TTATAA 14 ± 6	CTAC, GTATAC 16 ± 7	Site 1 GTATATAC, ATAT 4 ± 1.5	Site 2 ATAT 4.4 ± 1	
HexBrev	Site 2 ATAT 4.2 ± 1	Site 1 GTAC, ATAT, GTATATAC 3.8 ± 1.5	GTATAC, GTAG 8 ± 3	TTATAA 7 ± 3	CTAA, TTAG 8 ± 4	
Hemi-naphthyl TANDEM						
HexBfor	Site 1 CTAA, TTAG 11 ± 3	Site 2 TTATAA 6 ± 3.5	Site 3 GTATATAC, ATAT 5 ± 2	Site 4 ATAT 4.4 ± 1.4		
HexBrev	Site 4 ATAT 11 ± 6	Site 3 GTAC, ATAT, GTATATAC 10 ± 7	GTATAC, GTAG 15 ± 9	Site 2 TTATAA 10 ± 7	Site 1 TTAG 16 ± 12	
[Lys^4,Lys^8]TANDEM						
HexBfor	CTAA, TTAG 1 ± 0.5	TTATAA 1 ± 0.4	CTAC 1.4 ± 0.6	GTATAC 0.3 ± 0.06	GTATATAC, ATAT, GTAC 0.06 ± 0.05	ATAT 0.5 ± 0.2
HexBrev	ATAT 0.4 ± 0.1	GTAC, ATAT, GTATATAC 0.1 ± 0.01	GTATAC, GTAG 0.2 ± 0.05	TTATAA 0.3 ± 0.06	CTAA, TTAG 0.4 ± 0.2	
Biotinylated [Lys^4,Lys^8]TANDEM						
HexBfor	CTAA, TTAG 3 ± 1.7	CTAC, GTATAC 126 ± 80	GTATATAC, ATAT, GTAC 19 ± 6	ATAT 76 ± 36		
HexBrev	ATAT 14 ± 4	GTAC, ATAT, GTATATAC 27 ± 6	GTATAC, GTAG 51 ± 6	TTATAA 52 ± 15	TTAG 60 ± 37	

Table 4.4: C_{50} values (μM) for the interaction of: TANDEM; Hemi-naphthyl TANDEM; [Lys^4,Lys^8]TANDEM; and Biotinylated [Lys^4,Lys^8]TANDEM on HexBfor and HexBrev fragments. The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

This was attempted with the two HexB fragments. The success of this interaction is important for the use of the biotinylated TANDEM in nanotechnology. In these experiments the products of the footprinting reaction were extracted with phenol and precipitated before loading onto the gel so as to remove the protein. These experiments (not shown) were unsuccessful; the protein showed a total inhibition of DNase I cleavage at all the concentrations tested, even with a large excess of Biotinylated [Lys^4,Lys^8]TANDEM.

TANDEM	
TANa	ATATAT
	3.4 ± 0.7
TANb	ATATAT
	4.2 ± 1.4
Hemi-naphthyl TANDEM	
TANa	ATATAT
	11 ± 8
TANb	ATATAT
	12 ± 7
[Lys^4,Lys^8]TANDEM	
TANa	ATATAT
	0.1 ± 0.01
TANb	ATATAT
	0.07 ± 0.01
Biotinylated [Lys^4,Lys^8]TANDEM	
TANa	ATATAT
	35 ± 10
TANb	ATATAT
	29 ± 8

Table 4.5: C_{50} values (μM) for the interaction of: TANDEM; Hemi-naphthyl TANDEM; [Lys^4,Lys^8]TANDEM; and Biotinylated [Lys^4,Lys^8]TANDEM on TANa and TANb fragments.

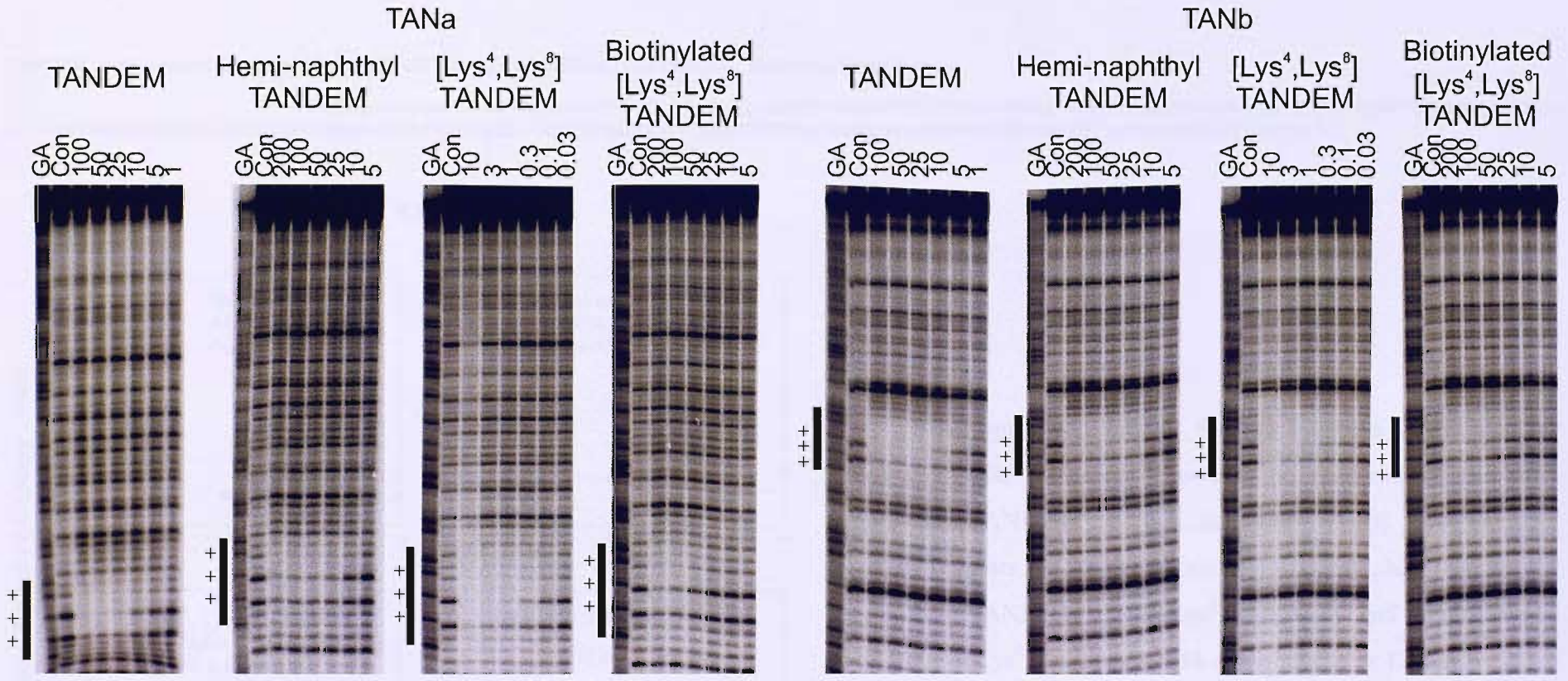


Figure 4.9: DNase I footprinting gels for TANDEM, hemi-naphthyl TANDEM, [Lys^4,Lys^8]TANDEM and biotinylated [Lys^4,Lys^8]TANDEM on the synthetic DNA fragment TANa/TANb. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The footprints are indicated by the filled boxes and TpA steps are marked "+". The DNA was labelled at the 5'-end of either strand.

TANDEM; Hemi-naphthyl TANDEM; [Lys^4,Lys^8] TANDEM; Biotinylated [Lys^4,Lys^8] TANDEM: TANa (top); TANb (bottom)

```
                    5'-CCACGTCGCTGACCACCTGCGCAGGTCCATATATGGCCAACTCGGTGCATCGCTCACTGGACACATCAGTCCATGAATGACTCGATGACTCAATGACTCG-3'
3'-GACGGACGTAGCGAGTTACGGGTGCAGCGACTGGTGGACGCGTCCAGGTATATACCGGTTGAGCCACGTAGCGAGTGACCTGTGTAGTCAGGTACTTACT-5'
```

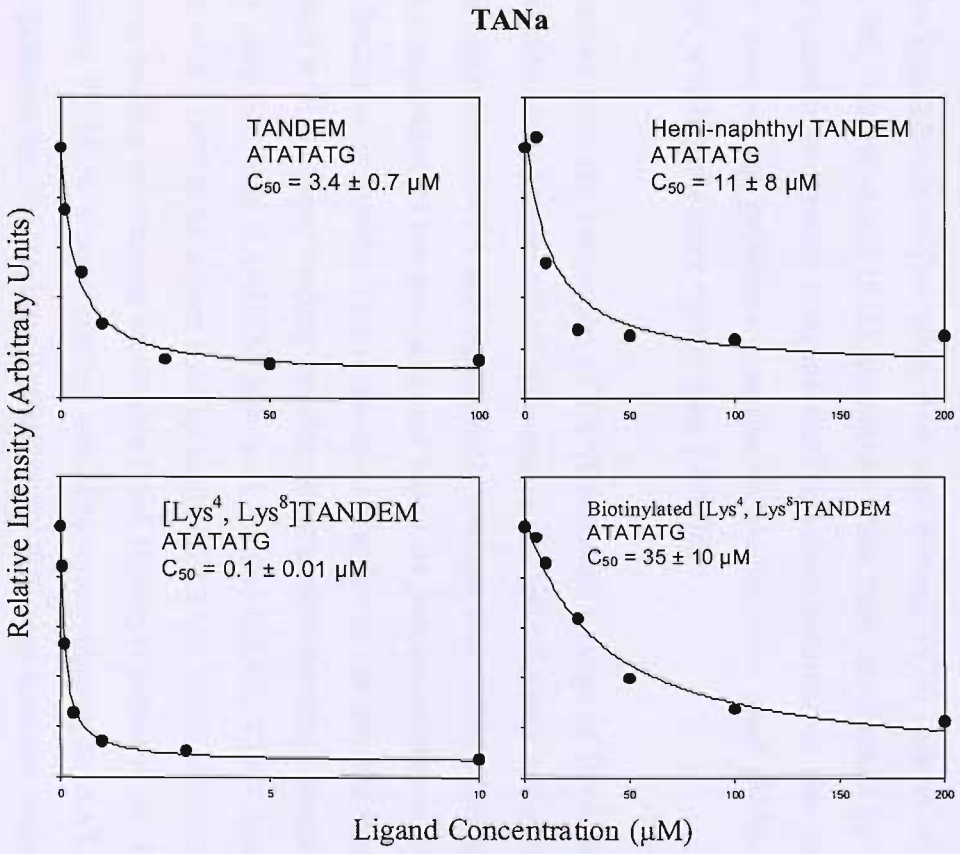


Figure 4.10: [Top] Sequence of the TANa/TANb DNA fragment showing the footprinting site for TANDEM and its analogues. [Left] Footprinting plots for the interaction of TANDEM, hemi-naphthyl TANDEM, [Lys^4,Lys^8]TANDEM and biotinylated [Lys^4,Lys^8]TANDEM with this site in TANa.

Antibacterial activity

Although the naturally occurring compounds echinomycin and triostin A are potent antibiotics, TANDEM has no antibacterial activity. It is not clear whether this is because it binds less well or because it is more rapidly degraded by peptidases as it lack the N-methyl groups. Since [Lys^4,Lys^8]TANDEM binds 30-40 times stronger than TANDEM, we tested whether this compound has any antibiotic activity against *Staphylococcus aureus*. However, in simple dilution experiments neither this compound nor TANDEM had any effect on the growth of *S. aureus* at concentrations up to 25 μM. In contrast, echinomycin displayed a minimum inhibitory concentration of 24 nM.

Discussion

TANDEM

The results presented in this chapter confirm that, as expected, the TANDEM that has been prepared by solid-phase synthesis has the same properties as that prepared by solution-phase method. The ligand binds to TpA steps (Lee and Waring, 1978; Low *et al.*, 1984b; Waterloh *et al.*, 1992; Addess *et al.*, 1992), especially when these are flanked by A and T residues. It is not possible to directly compare the C_{50} values obtained in this work with previous work as most of the published quantitative studies have used [N-$MeCys^3$,N-$MeCys^7$]TANDEM, which binds more tightly than TANDEM.

The C_{50} values derived from the interaction of TANDEM with a range of DNA fragments confirm that TANDEM binds best to ATAT ((Lavesa *et al.*, 1993; Lavesa and Fox, 2001); this site is slightly better than TATA and significantly stronger than CTAG. No protection is evident at TTAA regardless of the next adjacent bases, as previously observed (Lavesa *et al.*, 1993 and Fletcher *et al.*, 1995). This is thought to be due to the unusual structure of this sequence, which is also a poor binding site for minor groove binding ligands such as distamycin. Other sites that bind TANDEM less well include GTAT, TTAT (previously shown by Fletcher *et al.*, 1995 to be a poor binding site) and ATAC, while alternating G/C-tracts prevent strong binding by creating unsuitable DNA rigidity (Orstein *et al.*, 1978). In general, the sequence YTAR is a poor binding site. The strong binding to AATATT (in HexB substrates) indicates that the inhibitory effect of YY/RR steps does not extend to the

next flanking bases. The observation that ATATAT is preferred over other NATATN sites (also seen by Fletcher *et al.*, 1995) on HexA and HexB indicates either cooperative binding of TANDEM to adjacent sites (which seem unlikely as the TpA steps are too close together to permit the simultaneous bonding of two TANDEM molecules) or that the structure of long $(AT)_n$ tracts is better able to support the binding of this ligand, as suggested by Waterloh *et al.* (1992). This ATATAT site is bound on the TANa/TANb substrate producing a C_{50} value of between 3-4 μM.

The studies with the other derivatives provide information on important features of the bifunctional intercalators.

Role of the disulphide bridge

The disulphide bridge is broken in Bis-acm TANDEM, though the rest of the molecule is the same as TANDEM. This derivative failed to produce any footprints, confirming that the cross-bridge is an essential part of the structure. This has previously been suggested with the compounds [Ala^3,Ala^7]TANDEM (Fox *et al.*, 1980a) and a chemically broken derivative of echinomycin (Lee and Waring, 1978). Removing the disulphide bridge will make the compound less rigid; the chromophores may not be oriented in the correct position for intercalation and the functional groups on the peptide backbone will not be correctly positioned.

Modification of the intercalating groups

Hemi-naphthyl TANDEM and Naphthyl TANDEM

Substituting one of the intercalating quinoxaline groups for a naphthalene ring severely reduced the binding affinity, but had little effect on the sequence selectivity. The weaker binding of this ligand means that only the strongest TANDEM binding sites (TATATA and AATATT) showed cleavage protection. The results with TANa/TANb confirm that the ligand still binds to the sequence ATATAT but shows a three-fold decrease in affinity.

Replacing both quinoxaline rings with naphthalene appears to abolish binding altogether (up to 500 μM). This unexpected observation suggests that although the sequence selectivity is primarily determined by the cyclic peptide, the chromophores are more than inert wedges that intercalate between the bases. A role for the chromophores has

been suggested in previous studies with the analogues 1QN and 2QN, in which one or both of the quinoxaline chromophores of echinomycin were replaced with quinolines (Fox *et al.*, 1980b; Bailly *et al.*, 1999). This modification had little effect on the overall selectivity, but caused significant changes in relative affinities for different sequences. Removal of both nitrogens in the quinoxaline ring will alter the charge distribution around the ring and may thereby affect its ability to intercalate. It is also worth noting that quinoxaline rings alone are not DNA intercalators and that they only bind in this fashion when placed in the context of the complete antibiotic molecule. It may also be significant that most aromatic intercalating compounds possess either heteroatoms (such as nitrogen, as in the phenathridine chromophore of ethidium or acridine rings) or have other electronegative groups attached to the rings (as in the anthracyclines). It is possible that the naphthalene rings cannot stack efficiently between the DNA base pairs.

It would be interesting to investigate the properties of other ring-substituted analogues of TANDEM, possibly including larger three or four ring aromatic groups. We might expect these derivatives to bind with greater affinity, though the results with the simple chromophore derivatives presented in the chapter suggest that the effects may not be simple to predict.

Mono-quinoxaline TANDEM & Bis-acetate TANDEM

The results with these derivatives demonstrate that, although the selectivity may arise from interactions with the peptide backbone, this alone is not sufficient for DNA binding. The removal of only one of the intercalating groups of TANDEM abolishes binding.

Modifications of the octadepsipeptide ring - Val^4 and Val^8 modifications

Biotinylated [Lys^4,Lys^8]TANDEM and Benzylated [Lys^4,Lys^8]TANDEM

Biotinylated [Lys^4,Lys^8]TANDEM was synthesised with the intention of using it in nanotechnological applications. This derivative exhibited much weaker binding than the parent compound and only bound to the best TANDEM sites. The much weaker binding than [Lys^4,Lys^8]TANDEM could arise because the long side chains interfere with the binding. It could also be attributed to the lack of the positive charge or hydrogen bonding interactions that are possible with the lysine residues. A similar, though more pronounced effect is evident with the benzylated compound.

Adding avidin to the DNA-bound Biotinylated [Lys^4,Lys^8]TANDEM ligand resulted in complete protection of the DNA from cleavage, even at low concentrations. Although it could be suggested that the biotins are interfering with ligand binding, so are therefore not available to bind with the avidin, recent studies have discovered a high-affinity interaction between DNA and avidin that causes DNA condensation (Morpurgo *et al.*, 2004), possibly explaining the lack of specific binding.

[Lys^4,Lys^8] TANDEM

DNase I footprinting studies on [Lys^4,Lys^8]TANDEM (which was a synthetic precursor to the benzylated compound) showed that this compound binds more strongly than TANDEM, but with the same selectivity. This stronger binding has caused greater differentiation between the best binding sites of TANDEM and it can now be seen that ATAT is clearly preferred over TATA. The 30-40 times stronger binding suggests that there are favourable interactions between the lysine amino groups (which may be positively charged) and the negatively charged DNA phosphodiester backbone. This observation is especially interesting as the valine residues at this position in the parent compounds form close hydrophobic interaction with the walls of the DNA minor groove (Ughetto *et al.*, 1985). Introducing a basic amino acid at this position might therefore have been expected to diminish the binding (although the aliphatic $(CH_2)_4$ lysine chain may allow an interaction with the DNA backbone). This therefore raises the possibility of introducing other amino acids at this position to further enhance the affinity or modulate the sequence selectivity. Another (untested) favourable property of this compound is that it should be more soluble than the parent compound TANDEM (6 μM; Lee and Waring, 1978).

TANDEM does not possess any antibacterial activity, either as a result of its lower binding affinity than triostin A and echinomycin, or because it is degraded by peptidases as it lacks the N-methyl groups. It was therefore thought that [Lys^4,Lys^8]TANDEM, which binds more tightly, might show some antibacterial activity. However, this also showed no activity against *S. aureus* at concentrations up to 25 μM. This suggests that the low binding strength is not the reason for the lack of TANDEM's activity. It therefore seems likely that the methyl groups of the natural antibiotics protect them from degradation. It might therefore be interesting to examine the biological activity of related derivates that contain two of the N-methyl groups (N-$MeCys^{3,7}$ or N-$MeLys^{4,8}$).

Chapter V

Polyamide and Hoechst-Derived Minor Groove Binding Ligands

Introduction

The minor groove of DNA is an ideal target for small molecules that bind using a combination of hydrogen bonds, electrostatic forces, van der Waal interactions and hydrophobic contacts to achieve high affinity and sequence selectivity (Wemmer and Dervan, 1997). Distamycin and netropsin were the first natural compounds found to bind to the minor groove selectively, targeting A/T regions of DNA (Zimmer *et al.*, 1971; Wartell *et al.*, 1974). The first such synthetic compound was the dye Hoechst 33258 (containing benzimidazole rings). This compound was also A/T-selective and bound in a 1:1 configuration with the DNA at low concentrations (Harshman and Dervan, 1985).

Pelton and Wemmer (1989) observed distamycin can bind to longer A/T-tracts as a dimer forming a 2:1 antiparallel complex. This structure has allowed the possibility of targeting different bases selectively by exploiting differences between the atoms presented into the DNA minor groove. N-methylimidazole was the first ring unit to be used in this context, forming hydrogen bonds with the exocyclic 2-amino group of guanine (Kopka *et al.*, 1985; Lown *et al.*, 1986; Wade and Dervan, 1987; Wade *et al.*, 1992). N-methylhydroxypyrrole has since been used to discriminate between TA and AT, using steric differences between adenine and thymine (White *et al.*, 1998; Kielkopf *et al.*, 2000).

All four DNA bases can therefore be targeted using different ring systems in a 2:1 configuration. However, long polyamides containing these ring systems come out of phase with the DNA (Kelly *et al.*, 1996), so limiting the number of bases that can be targeted. In addition some polyamides do not bind as well as expected to their intended targets (Chiang *et al.*, 2000; Best *et al.*, 2003).

It is therefore important to develop novel sequence selective compounds by experimenting with different ring moieties to improve the binding. To this end, derivatives of Hoechst 33258 and standard polyamides have been developed to study the effects of the introduction of benzimidazole (targets A/T), pyridoimidazole (proposed to target A/T), various other ring systems and novel linkages between molecules. 39 different ligands (which have been divided into four series) are discussed in this chapter.

Series A

Ten Series A ligands are shown in Figure 5.1 along with Hoechst 33258. These are all derivatives of Hoechst 33258 and contain two benzimidazole rings. The ligands differ from

Hoechst 33258 by substituting the N-terminal phenol with different ring systems, some only varying from one another by the position of an atom in a ring. Data obtained from these ligands, when compared to Hoechst 33258, will indicate how different N-terminal substitutions affect binding selectivity and affinity.

Series B

The 13 Series B ligands shown in Figure 5.2 are conjugates of Hoechst 33258 and polyamide derivatives, with the exception of M2M-B084, which is a classic Dervan polyamide. Many of the ligands have a Hoechst 33258 C-terminus and a polyamide N-terminus. M2M-B074 has a pyridoimidazole moiety at the C-terminus instead of a benzimidazole, whilst M2M-B083 includes a gamma amino butyric acid linker in the middle of the polyamide, and may be able to form a hairpin dimer. M2M-B091 contains a C-terminal imidazole ring and an N-terminal benzimidazole, in the opposite configuration to Hoechst 33258. M2M-B095 has this inverse N-terminal benzimidazole moiety as well, whilst M2M-B097 and M2M-B099 have two consecutive C-terminal benzimidazole moieties. The different configurations of these conjugates will provide insight into the effect of position of a ring moiety within the polyamide, as well as show how different components affect the overall binding properties of a ligand.

Series C

The 8 Series C ligands shown in Figure 5.3 are comprised of two Hoechst 33258 and polyamide derivatives in a "Head-to-Head" or "Tail-to-Tail" configuration, joined together by a linker region. The design is such that the linker may be able to cross the minor groove, with each monomer binding to opposite strands. This binding mode may be preferred over normal antiparallel binding due to the orientation of the ligand. In this way, a potentially limitless length of DNA can be targeted, as the ligands overlap in the minor groove as heterodimers with the same binding rules as used for normal polyamides.

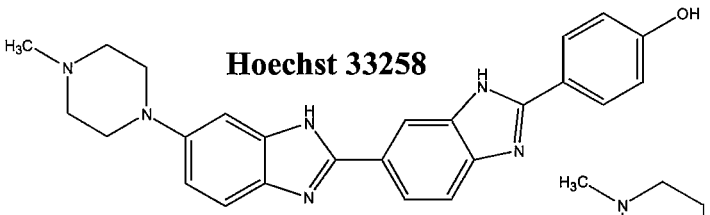
Series D

The 8 Series D ligands shown in Figure 5.4 are polyamide-Hoechst 33258 derivatives similar to the Series B ligands. However, instead of having benzimidazole moieties at the termini, Series D ligands have internal benzimidazole rings, often next to an N-terminal imidazole. Of particular interest is M2S-D88, which has alternating Bzi/Py moieties and so will provide insight into the effect of a central benzimidazole. Studies on these ligands will

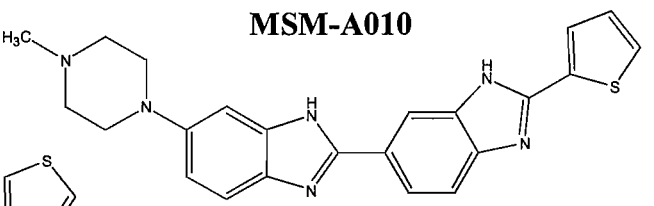
show the effect of benzimidazole position on binding.

These ligands were examined by DNase I and hydroxyl radical footprinting using various DNA substrates, some of which were designed specifically for certain ligands (*e.g.* SASK3/SASK4 for Series C ligands).

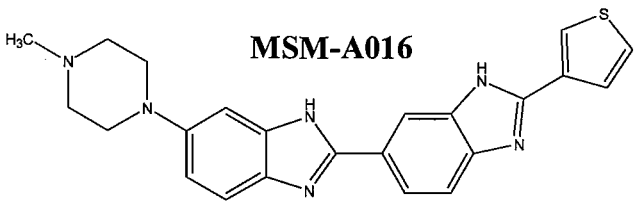
Hoechst 33258



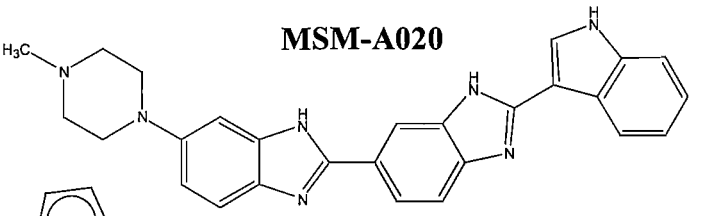
MSM-A010



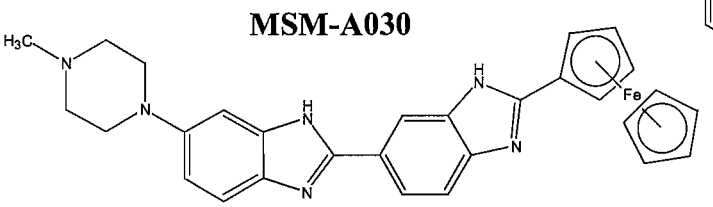
MSM-A016



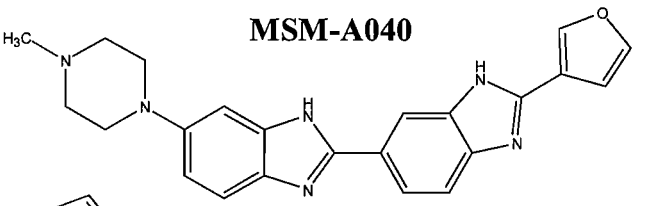
MSM-A020



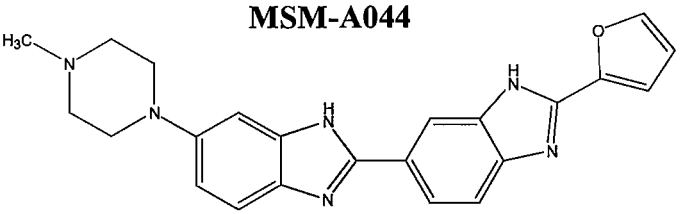
MSM-A030



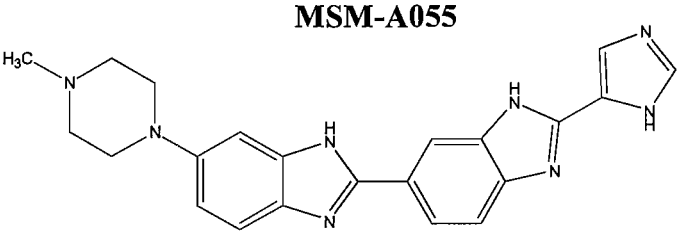
MSM-A040



MSM-A044



MSM-A055



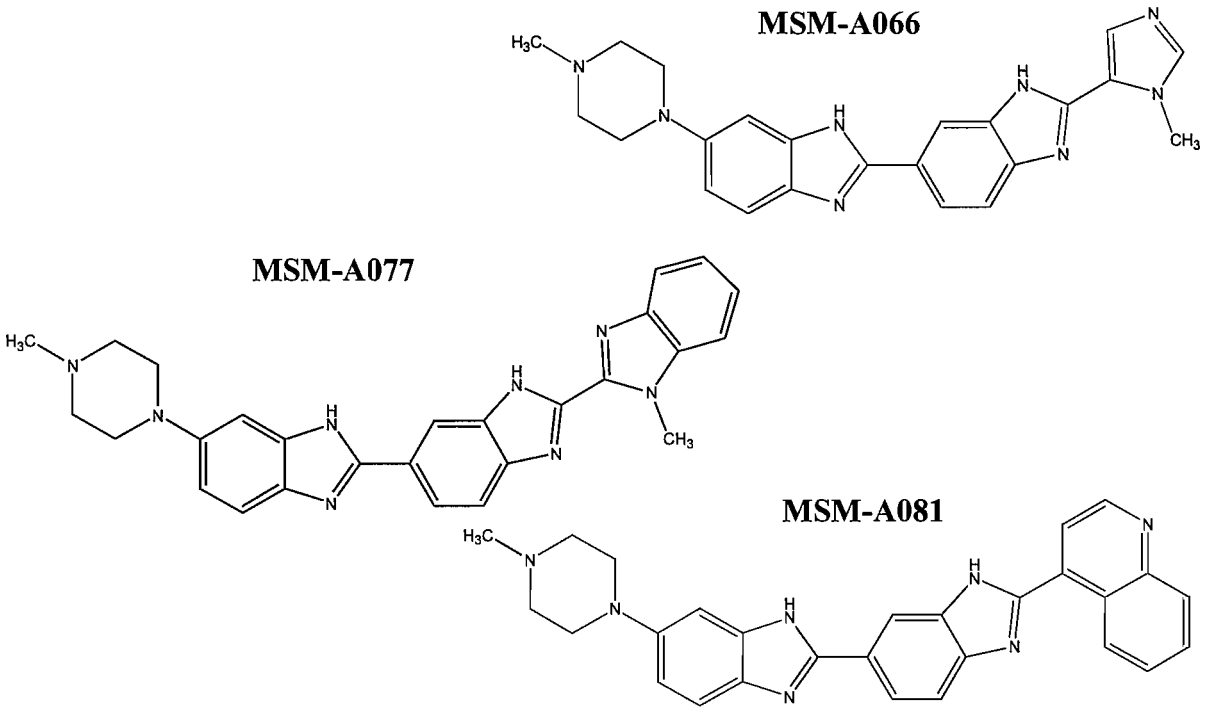
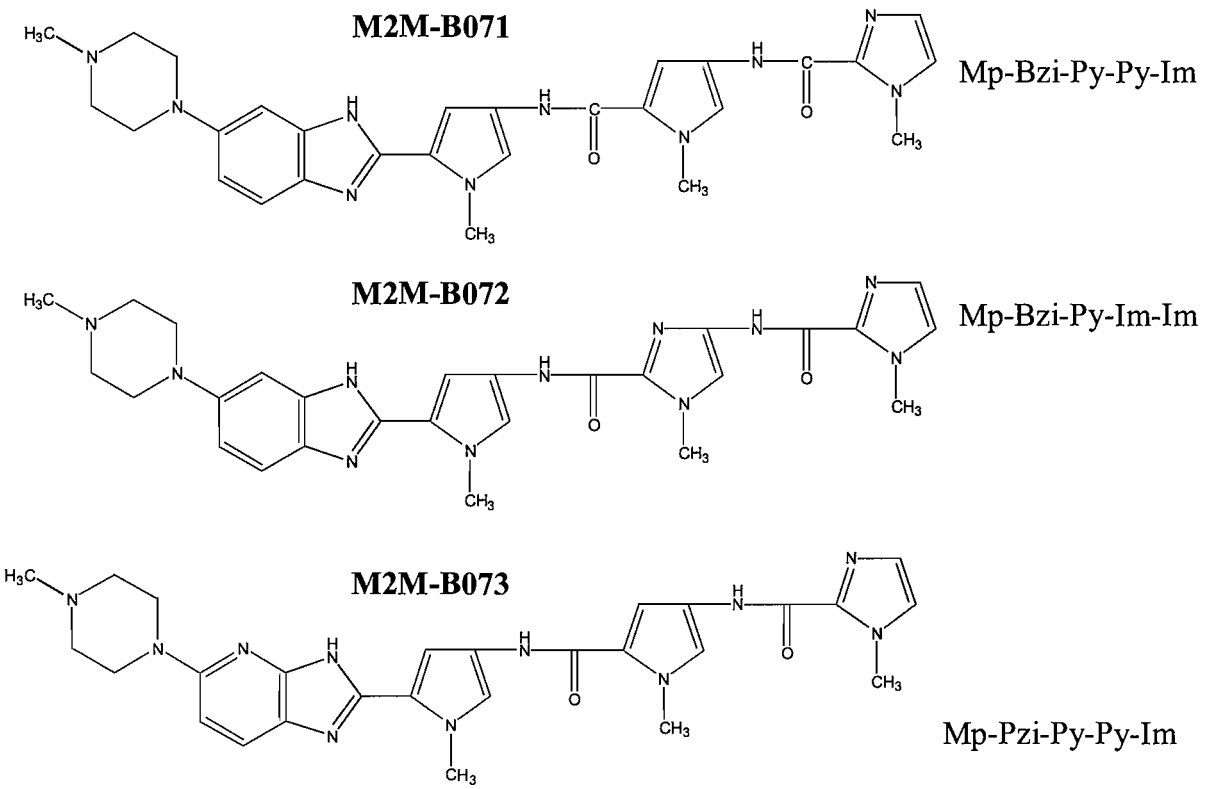
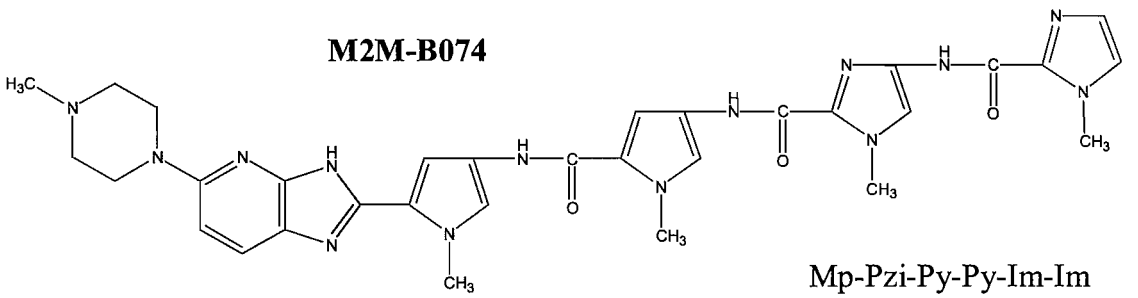


Figure 5.1: Structures of the Series A ligands. These are all derivatives of Hoechst 33258.

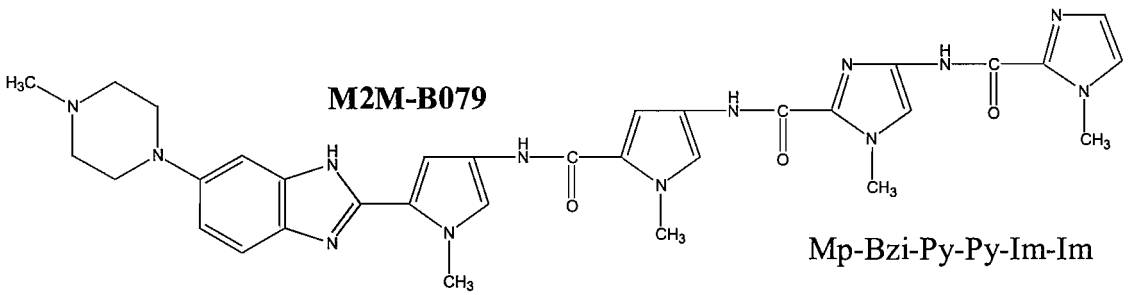


M2M-B074



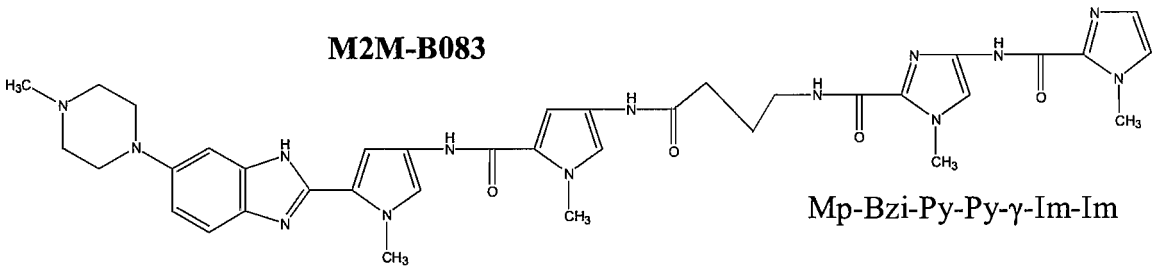
Mp-Pzi-Py-Py-Im-Im

M2M-B079



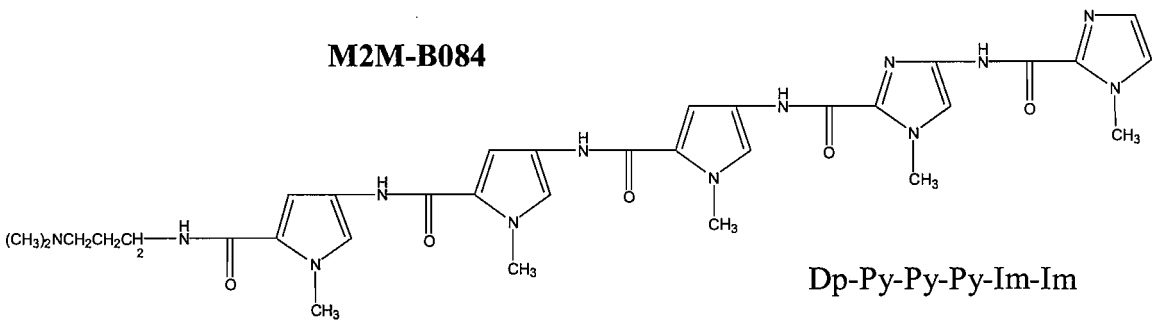
Mp-Bzi-Py-Py-Im-Im

M2M-B083



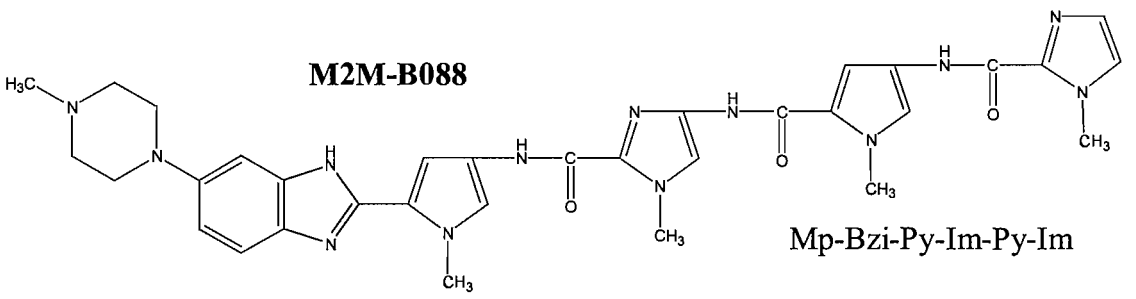
Mp-Bzi-Py-Py-γ-Im-Im

M2M-B084



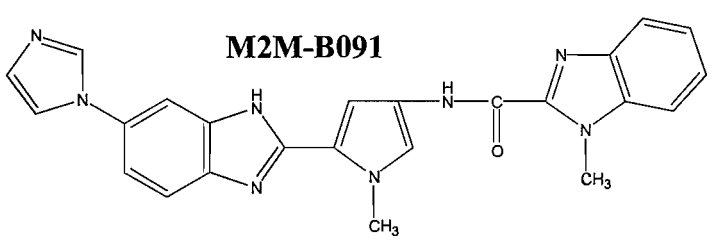
Dp-Py-Py-Py-Im-Im

M2M-B088



Mp-Bzi-Py-Im-Py-Im

M2M-B091



Im-Bzi-Py-Bzi

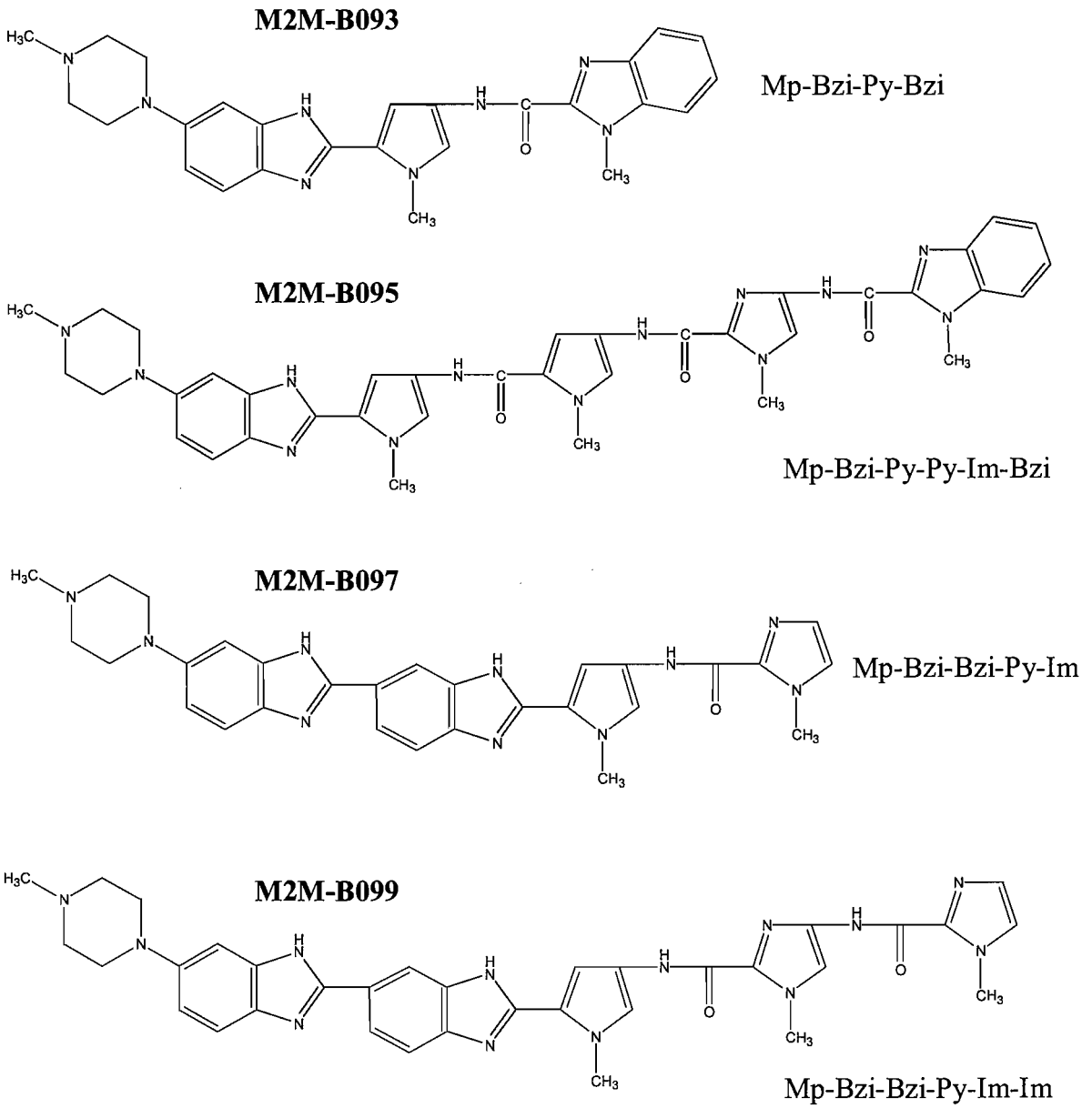


Figure 5.2: Structures of the Series B ligands. These are conjugates of Hoechst 33258 with pyrrole/imidazole polyamides. M2M-B084 is a standard pyrrole/imidazole polyamide.

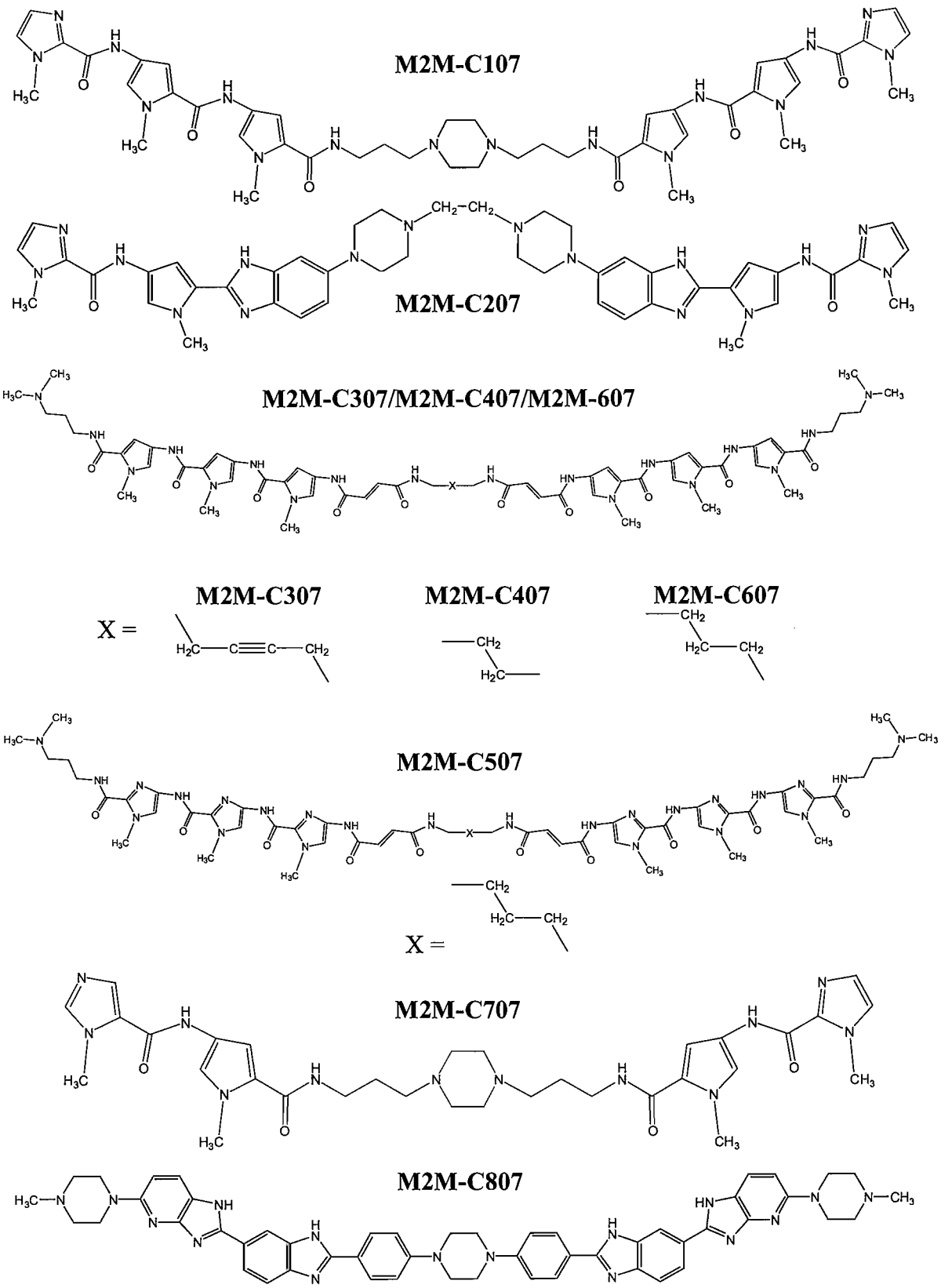


Figure 5.3: Structures of the Series C ligands, which are "Head-to-Head" or "Tail-to-Tail" Hoechst 33258 and polyamide derivatives, joined by various linkers.

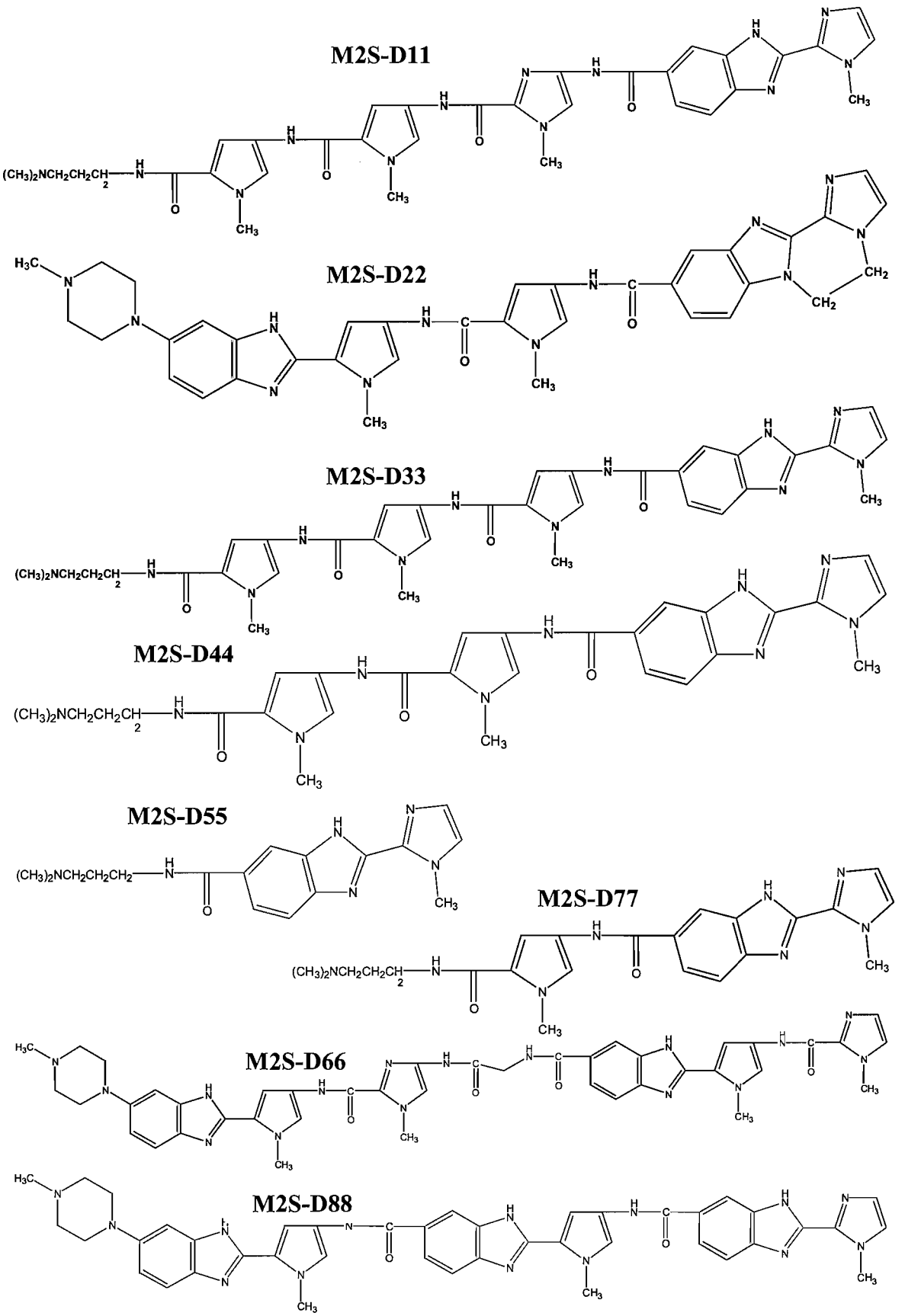


Figure 5.4: Structures of the Series D ligands, which examine the effect of internal benzimidazole rings.

Results

DNase I and hydroxyl radical footprinting were used to establish the sequence selectivity and binding affinity of the 40 ligands shown in Figures 5.1- 5.4. For each ligand, initial studies were carried out on the universal tetranucleotide sequence MS1/MS2, to determine the approximate binding selectivity. On the basis of these results, and the predicted selectivity, further specifically designed sequences (such as SASK1/SASK2) were then used in order to define the binding preferences more precisely.

The results for Series B and D are grouped together due to the similar nature of these ligands. The results for the Series C ligands therefore follow after those for Series D.

It should be noted that the observed footprints often encompass two or more possible binding sites. In such cases, the analysis is focussed on the primary binding site and the other potential sites are indicated in parentheses.

Hoechst 33258 Analogues: Series A

The Series A ligands were first footprinted on fragments MS1 or MS2 to establish their approximate binding selectivity. From the results it was clear that all these ligands are A/T selective and so they were tested against pAAD1, which contains different $(A/T)_4$ sites to better assess their binding selectivity and affinity.

M2M-A044, M2M-A077 and M2M-A081 were not tested on MS1 and MS2 as they were obtained after the other ligands.

MS1/MS2

Figure 5.5 shows the results of DNase I footprinting experiments with Hoechst 33258, M2M-A020, M2M-A030 and M2M-A055 on MS1, while Figure 5.6 shows the results of similar experiments with Hoechst 33258, M2M-A010, M2M-A016, M2M-A020, M2M-A040 and M2M-A066 on MS2. Figure 5.7 shows the sequence of MS1/MS2 with the observed footprinting binding sites highlighted, together with an example of footprinting plots derived from the binding of M2M-A030 to MS1. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.1.

It can be seen that Hoechst 33258 produces DNase I footprints at most of the $(A/T)_4$ sites on MS1 and MS2; as expected AATT (in AATTA/TAATT) and ATTT/AAAT

are the best 4 bp binding sites, producing footprints at the lowest ligand concentrations, whilst TTTT on MS1 is weaker. On MS2, the sequence corresponding to TTTT is AAAC (due to a point mutation at this site) at which no footprint is produced. Assuming Hoechst 33258 is binding in a 1:1 mode, $(A/T)_4$ sequences are therefore preferred over $(A/T)_3$ sites. This is confirmed by the weaker binding of AATG/CATT. No significant binding is observed for TTAA and TATA on MS1 or MS2 by Hoechst 33258, and TpA steps are generally bound weakly (such as TAAA/TTTA, TTAT/ATAA, ATAT and AATA/TATT) compared to sites containing only ApT or ApA/TpT steps.

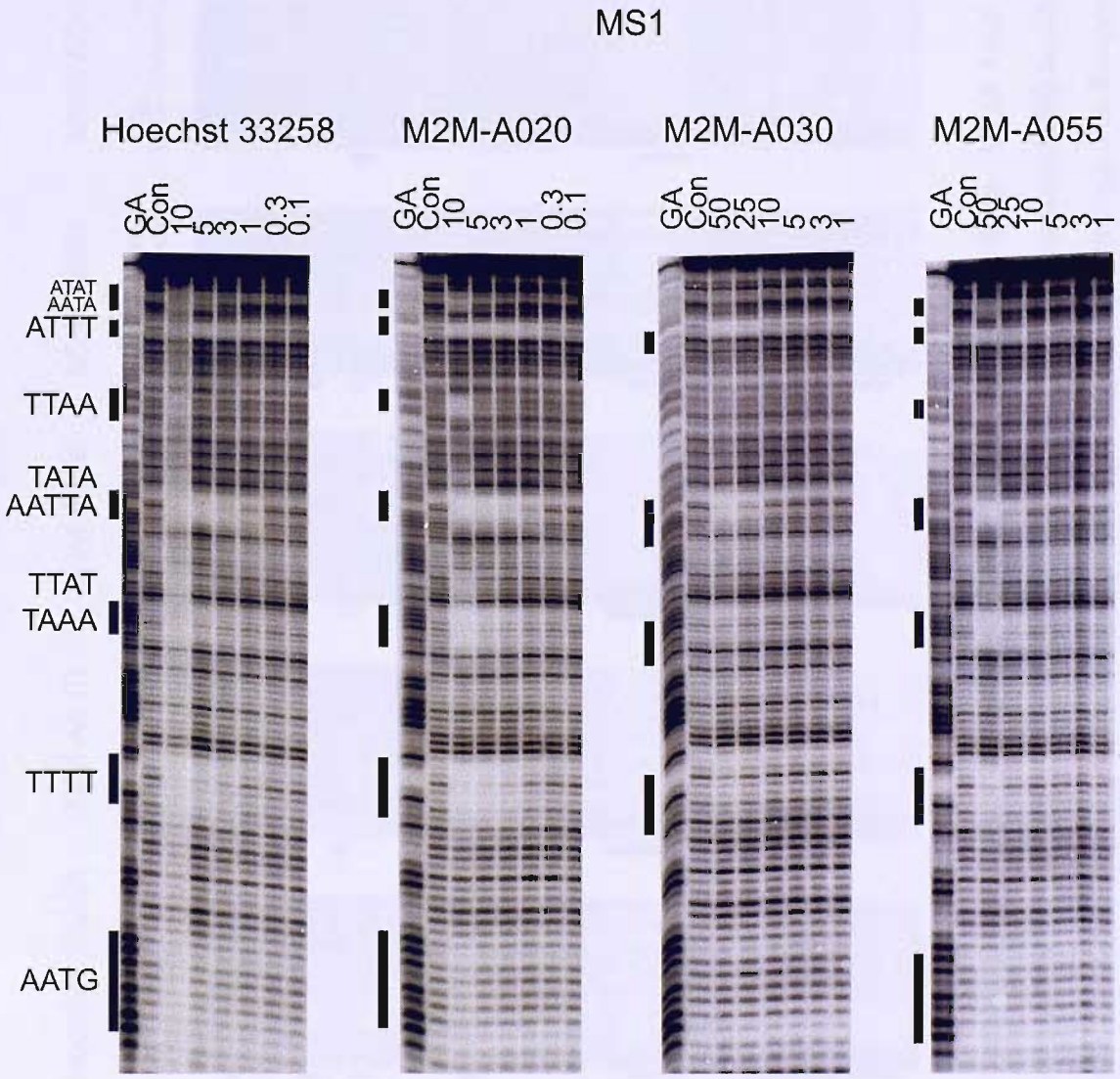


Figure 5.5: DNase I footprinting experiments for Hoechst 33258, M2M-A020, M2M-A030 and M2M-A055 with MS1. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The $(A/T)_4$ sites are indicated to the left of the gels and the boxes adjacent to each gel shows binding. The footprints are highlighted in Figure 5.7.

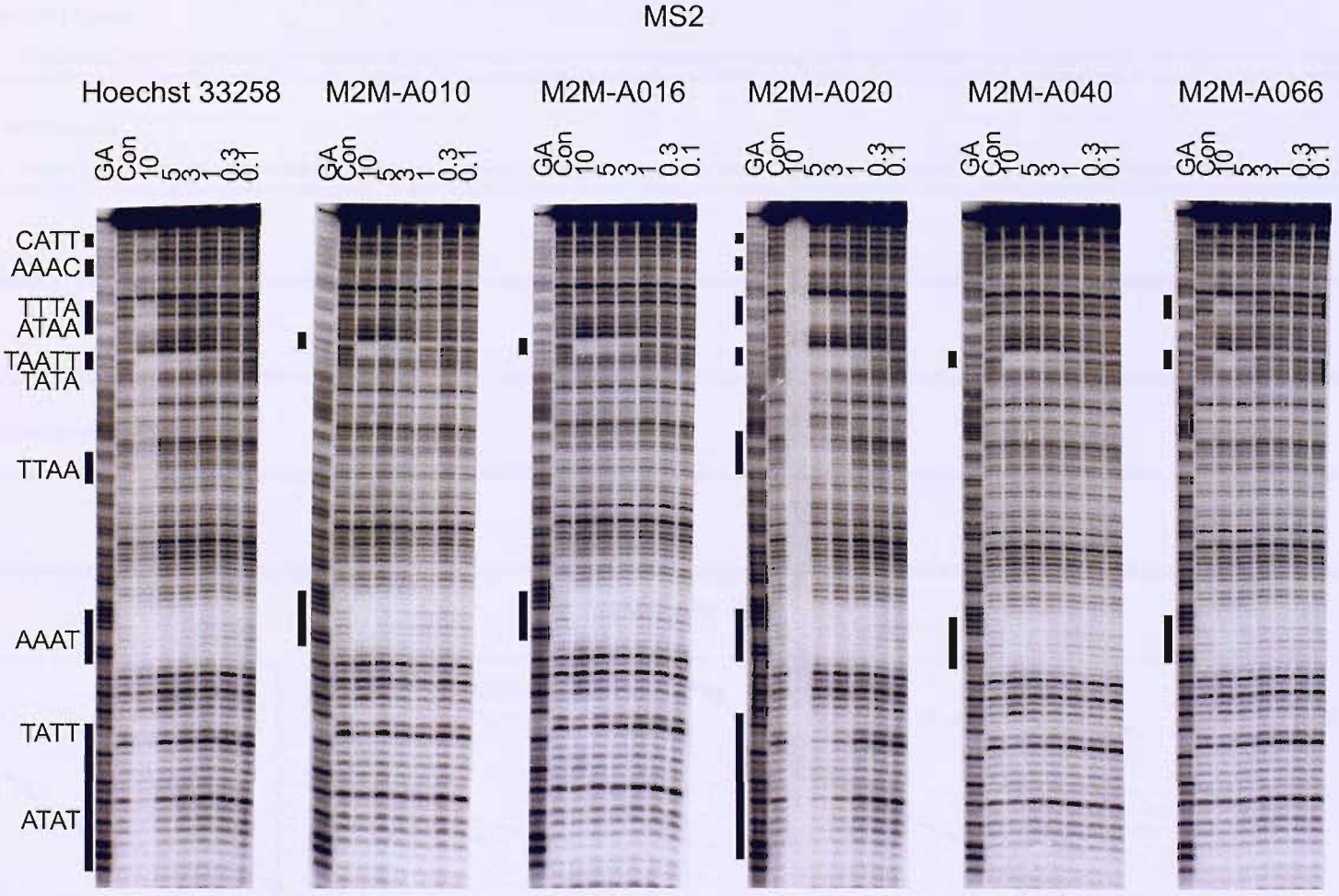


Figure 5.6: DNase I footprinting experiments for Hoechst 33258, M2M-A010, M2M-A016, M2M-A020, M2M-A040 and M2M-A066 with MS2. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The $(A/T)_4$ sites are indicated to the left of the gels and the boxes adjacent to each gel shows binding. The footprints are highlighted in Figure 5.7.

Hoechst 33258: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

M2M-A020: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

M2M-A030: MS1

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
```

M2M-A055: MS1

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
```

M2M-A010; M2M-A016; M2M-A040: MS2

```
5'-GGATCCGCATTCGAGGCTGAGATGACAAACCAGACCCCACCGGACGTACTTTACATAACTCTTCACGCCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCTTGATCGCGCTACGACTAGTGCAGTTGGAAATCGGCCATGTGTATTGCCGCATATGGATCC-3'
```

M2M-A066: MS2

```
5'-GGATCCGCATTCGAGGCTGAGATGACAAACCAGACCCCACCGGACGTACTTTACATAACTCTTCACGCCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCTTGATCGCGCTACGACTAGTGCAGTTGGAAATCGGCCATGTGTATTGCCGCATATGGATCC-3'
```

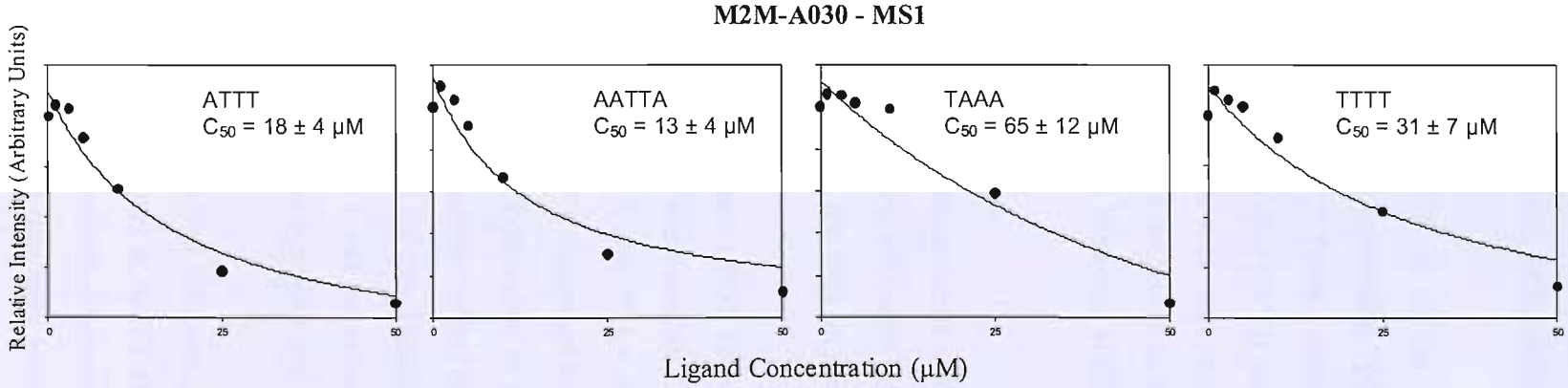


Figure 5.7: [Top] Sequences of fragments MS1 and MS2, showing footprinting sites for the Series A ligands in red. These were derived from the gels in Figures 5.5 and 5.6. [Above] Example of footprinting plots for M2M-A030 binding MS1.

It can be seen that M2M-A010, M2M-A016, M2M-A030 and M2M-A040 only produce footprints at the best Hoechst 33258 sites (AATT and AAAT/ATTT), although M2M-A030 also shows very weak binding to TAAA and TTTT on MS1. The rank order of their affinities to AATT and AAAT/ATTT is M2M-A040 > M2M-A016 ≥ M2M-A010 > M2M-A030, though they are all weaker than Hoechst 33258.

M2M-A020 produces similar footprints to Hoechst 33258, at all the A/T-tracts, though TTAA is bound more strongly relative to other sites (sites containing TpA steps are generally stronger for this ligand). M2M-A055 binds best to AATT (with a similar affinity to M2M-A030), but is less effective at ATTT, binding to the other $(A/T)_4$ sites weakly (TTAA is much weaker). M2M-A066 binds AATT and ATTT with approximately the same affinity as M2M-A010, but this ligand also produces a good, but weaker, footprint at TTTA and ATAA. TATA is not bound by any Series A ligand on MS1 and MS2.

pAAD1

Figures 5.8 and 5.9 show the results of DNase I footprinting experiments with Hoechst 33258 and the Series A ligands on pAAD1. The $(A/T)_4$ sites are indicated by the filled boxes. The C_{50} values for the interaction of these ligands with all the sites on this fragment are summarised in Table 5.1.

It can be seen that, as previously reported (Abu-Daya *et al.*, 1995), Hoechst 33258 produces a footprint at all the $(A/T)_4$ sites. AATT is the best site and Hoechst 33258 binds to this about 100 times stronger than to ATAT and TAAT; TATA and TTAA are weaker. M2M-A010, M2M-A016 and M2M-A040 also bind well to AATT (although 5-, 2- and 3-fold weaker, respectively, than Hoechst 33258); about 150 times better than to ATAT and TAAT, though no interaction is evident at TATA and TTAA. A similar effect is seen with M2M-A066 and M2M-A077 (both 50-fold weaker binding to AATT than Hoechst 33258), though there is only a 10-15-fold difference in C_{50} between AATT and the other two sites. However, M2M-A066 binds TAAT stronger than ATAT, with the opposite true for M2M-A077.

M2M-A020 and M2M-A044 produce similar, but slightly less selective, patterns of protection to Hoechst 33258, with about 40-fold higher C_{50} values at AATT than ATAT and TAAT. TATA and TTAA are weaker sites, though the discrimination between the best and worst sites is less than with Hoechst 33258, suggesting that they may have a greater tolerance for TpA steps. A similar effect is seen with M2M-A055 and M2M-A081, with protection similar to Hoechst 33258, with about 40-fold higher C_{50} values at AATT than

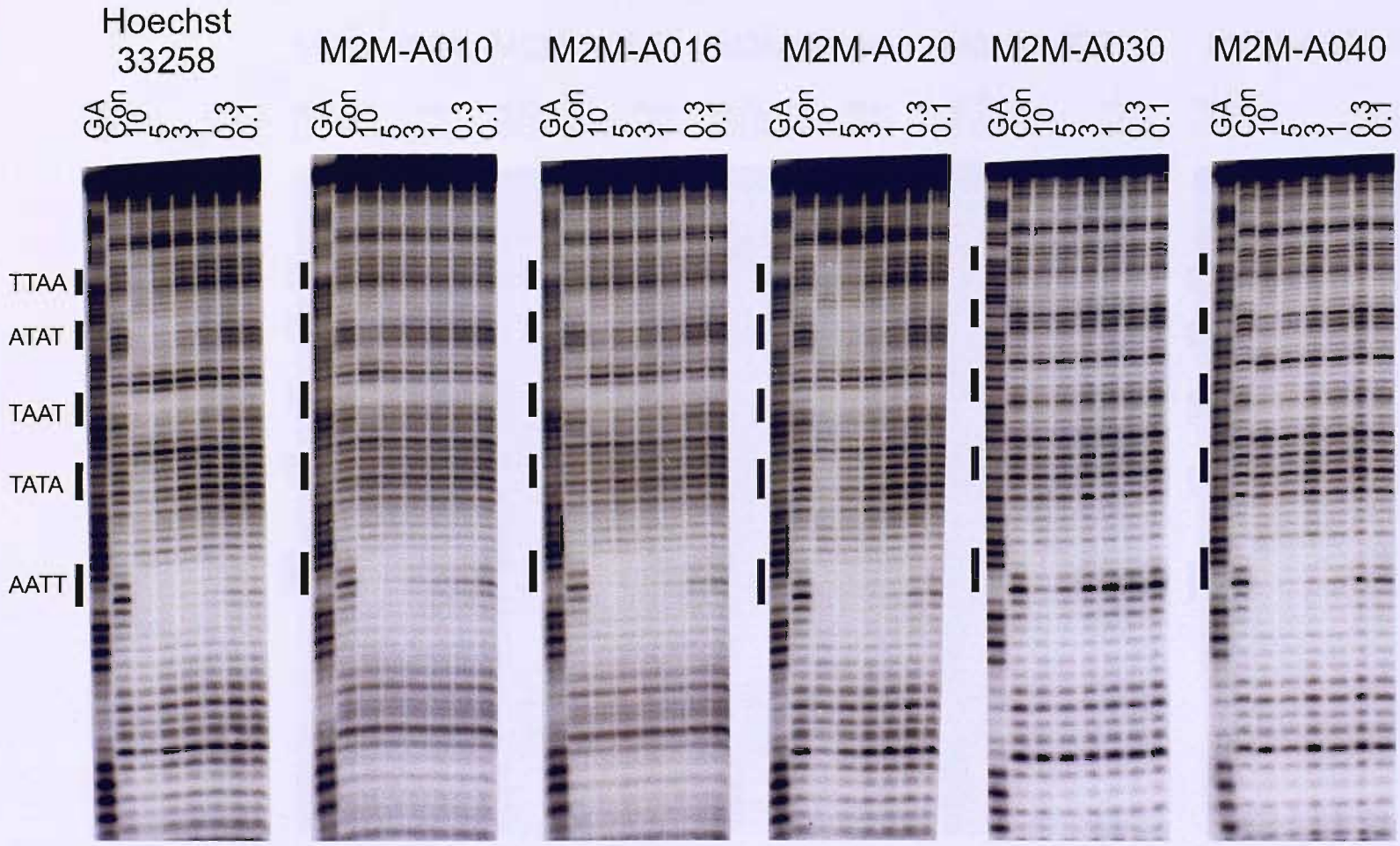


Figure 5.8: DNase I footprinting experiments for Hoechst 33258, M2M-A010, M2M-A016, M2M-A020, M2M-A030 and M2M-A040 with the fragment pAAD1. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane. The $(A/T)_4$ sites are marked.

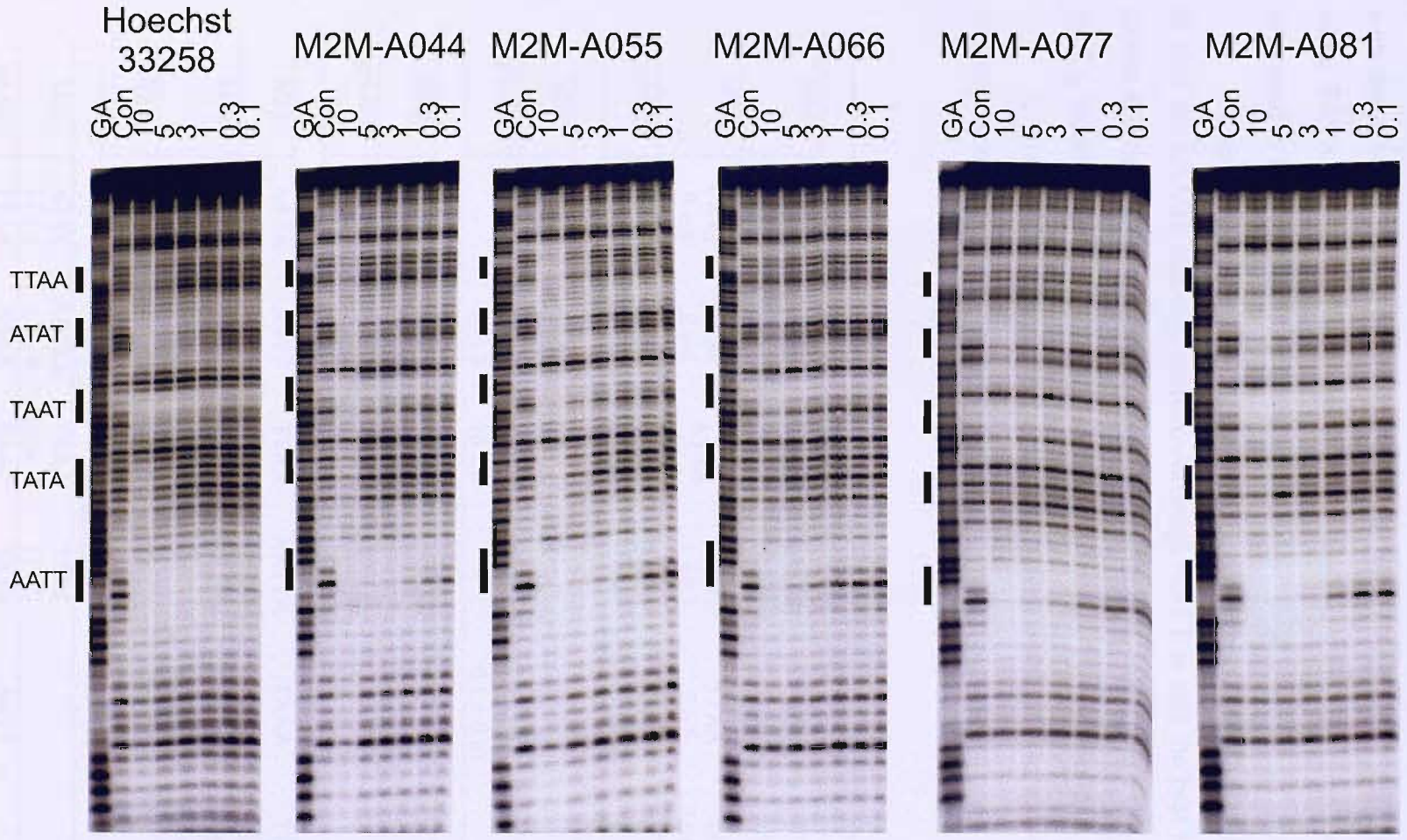


Figure 5.9: DNase I footprinting experiments for Hoechst 33258, M2M-A044, M2M-A055, M2M-A066, M2M-A077 and M2M-A081 on the pAAD1 fragment. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The $(A/T)_4$ sites are marked.

ATAT and TAAT. TATA and TTAA are weaker sites, though the discrimination between the best and worst sites is less than with Hoechst 33258, suggesting that they may have a greater tolerance for TpA steps. A similar effect is seen with M2M-A055 and M2M-A081, with about 5- and 10-fold higher C_{50} values at AATT than ATAT and TAAT respectively, while TATA and TTAA, although still the weakest sites, are bound relatively strongly compared to Hoechst 33258.

As noted with MS1/MS2, much higher concentrations of M2M-A030 are required to alter the pAAD1 cleavage pattern. Although AATT is still the best site for this ligand (about 300 times weaker than Hoechst 33258), there is only about a 2-4-fold difference in the C_{50} values at the best and worst sites, with every $(A/T)_4$ site bound.

In general the ligands that bind less well show less discrimination between the different arrangements of $(A/T)_4$ residues.

Hoechst 33258							
MS1	ATAT, AATA 7 ± 1	ATTT 0.6 ± 0.6	TTAA 15 ± 7	AATTA 0.6 ± 0.3	TAAA 6 ± 1	TTTT 2 ± 0.5	AATG 6 ± 2
MS2	CATT 9 ± 4	AAAC 8 ± 4	TTTA, ATAA 6 ± 3	TAATT 1 ± 0.1	TTAA 7 ± 5	AAAT 0.6 ± 0.2	TATT, ATAT 4 ± 1
pAAD1	TTAA 8 ± 5	ATAT 2 ± 1	TAAT 1.5 ± 1	TATA 6 ± 4	AATT 0.02 ± 0.01		
M2M-A010							
MS2	TAATT 4 ± 2	AAAT 3 ± 1.					
pAAD1	TTAA -	ATAT 16 ± 3	TAAT 16 ± 4	TATA -	AATT 0.1 ± 0.03		
M2M-A016							
MS2	TAATT 3 ± 1	AAAT 3 ± 1					
pAAD1	TTAA -	ATAT 11 ± 3	TAAT 10 ± 3	TATA -	AATT 0.04 ± 0.02		
M2M-A020							
MS1	AATA 11 ± 5	ATTT 0.2 ± 0.2	TTAA 8 ± 2	AATTA 0.4 ± 0.2	TAAA 4 ± 1	TTTT 1 ± 0.3	AATG 10 ± 4
MS2	CATT 5 ± 2	AAAC 3 ± 1	TTTA, ATAA 2.5 ± 1	TAATT 0.2 ± 0.05	TTAA 2 ± 1	AAAT 0.3 ± 0.1	TATT, ATAT 1 ± 0.5
pAAD1	TTAA 5.5 ± 3.5	ATAT 2 ± 0.5	TAAT 3 ± 1	TATA 5 ± 3	AATT 0.05 ± 0.01		
M2M-A030							
MS1	ATTT 18 ± 4	AATTA 13 ± 4	TAAA 65 ± 12	TTTT 31 ± 7			
pAAD1	TTAA 22 ± 2	ATAT 12 ± 2	TAAT 10 ± 1	TATA 21 ± 5	AATT 6 ± 2		

M2M-A040							
MS2	TAATT	AAAT					
	1 ± 0.2	1.5 ± 0.2					
pAAD1	TTAA	ATAT	TAAT	TATA	AATT		
	-	8 ± 2	8 ± 3	-	0.05 ± 0.01		
M2M-A044							
pAAD1	TTAA	ATAT	TAAT	TATA	AATT		
	19 ± 18	3.5 ± 1	5 ± 2	11 ± 8	0.1 ± 0.01		
M2M-A055							
MS1	AATA	ATTT	TTAA	AATTA	TAAA	TTTT	AATG
	42 ± 10	34 ± 5	163 ± 65	15 ± 4	25 ± 5	32 ± 9	37 ± 11
pAAD1	TTAA	ATAT	TAAT	TATA	AATT		
	8 ± 5	3.5 ± 1	3.5 ± 1	5 ± 3	0.5 ± 0.1		
M2M-A066							
MS2	TTTA,ATAA	AATT	AAAT				
	9 ± 3	4 ± 1	6 ± 2				
pAAD1	TTAA	ATAT	TAAT	TATA	AATT		
	-	15 ± 3	8 ± 1	-	1 ± 0.4		
M2M-A077							
pAAD1	TTAA	ATAT	TAAT	TATA	AATT		
	-	8 ± 1	14 ± 2	-	1 ± 0.1		
M2M-B081							
pAAD1	TTAA	ATAT	TAAT	TATA	AATT		
	31 ± 22	5 ± 1	7 ± 1	25 ± 18	0.5 ± 0.2		

Table 5.1: C_{50} values (μM) derived from quantitative analysis of the gels presented in this chapter (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). Sites marked "–" indicate A/T-tracts on the pAAD1 DNA at which no significant cleavage protection was observed.

Hoechst 33258 / Polyamide Conjugates: Series B

The Series B ligands were first footprinted with DNase I using the MS1 and MS2 substrates to establish their approximate sequence selectivities. Further DNase I experiments were performed using HexA and HexB, together with hydroxyl radical footprinting to confirm this selectivity and to more accurately define the binding sites. Only gels for one orientation of each of these fragments are presented; in each case the other strands gave similar results.

Once their general selectivity had been established, DNase I and hydroxyl radical footprinting experiments were carried out using the specifically-designed fragments SASK1/SASK2 to define the binding preferences more accurately.

M2M-B091, M2M-B093 and M2M-B095 showed no binding to the MS1/MS2 or SASK1/SASK2 fragments.

M2M-B071

Utilising known binding rules of polyamides, M2M-B071 with the sequence Mp-Bzi-Py-Py-Im can bind as a 2:1 homodimer in three ways, as shown in Figure 5.10.

```
  Mp-Bzi-Py-Py-Im                Mp-Bzi-Py-Py-Im
          Im-Py-Py-Bzi-Mp                Im-Py-Py-Bzi-Mp
5'-W   W  G  W  C  W   W     5'-W   W  W  G  C  W  W   W

                  Mp-Bzi-Py-Py-Im
                       Im-Py-Py-Bzi-Mp
                5'-S   G  W  W  C   W
```

Figure 5.10: Possible DNA binding modes for M2M-B071.

Figure 5.11 shows the results of DNase I footprinting experiments with M2M-B071 on MS1, HexAfor, HexBrev, SASK1 and SASK2; and hydroxyl radical footprinting experiments on HexAfor and HexBfor fragments. Figure 5.12 shows the sequences of these fragments with the DNase I footprinting sites highlighted, together with an example of footprinting plots derived from the binding of M2M-B071 to SASK1. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.2.

MS1/MS2: It can be seen that M2M-B071 has bound strongest to TGTCA (site 5), with AGTCG (site 1, with TGCA) a weaker site. These contain the central portion of the top left binding mode in Figure 5.10 (*i.e.* GWC); the interaction with AGTCG suggests that G or C can be accommodated around the GWC to produce weaker binding sites, though there is no interaction with GGTCT. CGTTCT (site 2, with AGCT), containing a cytosine adjacent to the proposed GWWC core, is also bound less well than site 1. Other GWWC sequences (such as CGATCA and AGTACG) are not protected, indicating that the presence of AGCT in the site 2 footprint is aiding binding affinity. Weak binding is observed at AGCA (site 3), although surrounding guanine/cytosine bases again prevent binding in the case of AGCC, AGCGCG and GGCG.

The relatively strong binding to AAGAGTTAT (site 4) suggests that the ligand may also be able to bind to long A/T-tracts (probably by 1:1 binding) with guanine bases targeted by the imidazole, as suggested by Lown *et al.* (1986).

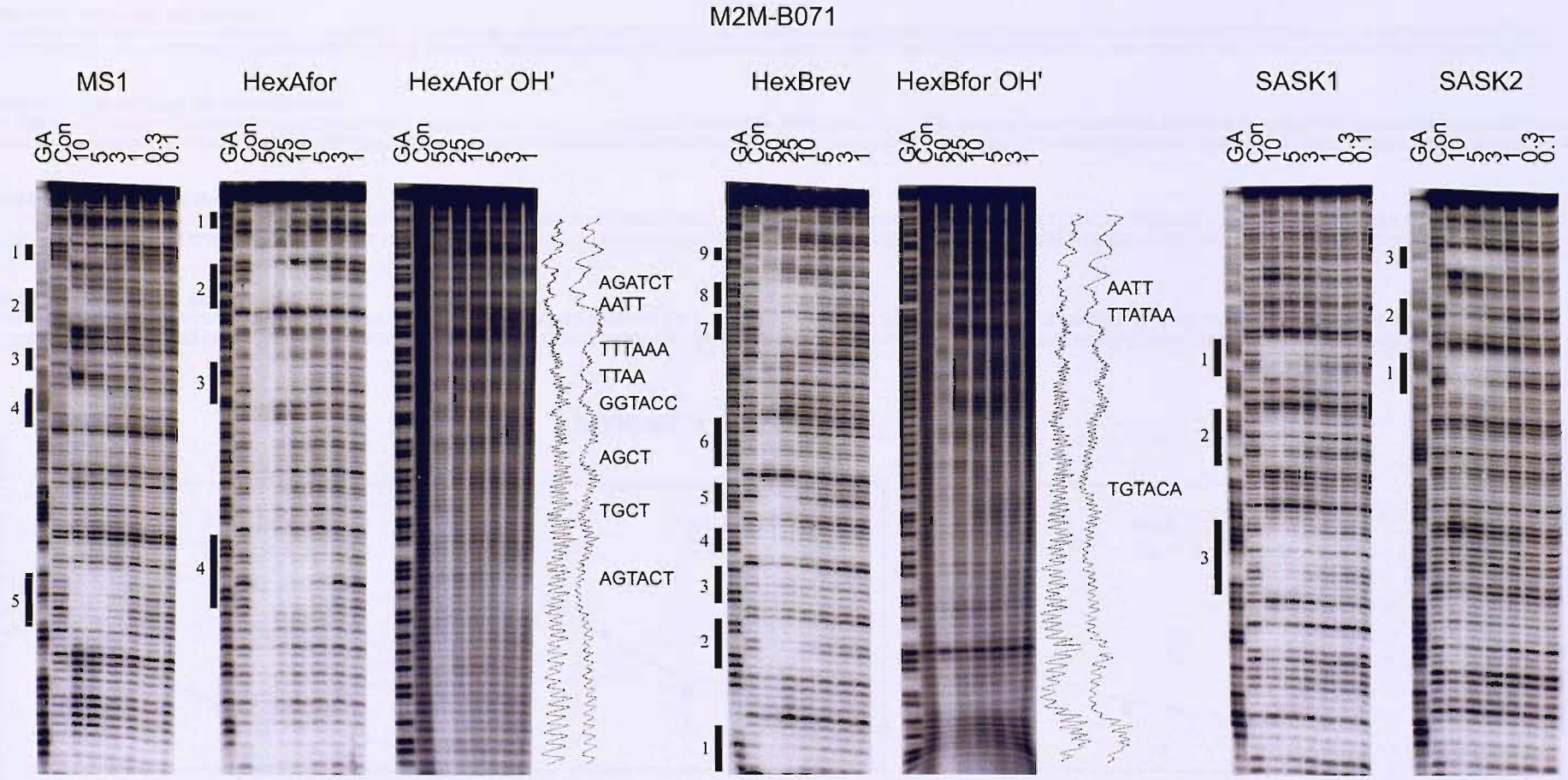


Figure 5.11: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B071 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.12; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 3 μM lane). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B071: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

M2M-B071: HexA for (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

M2M-B071: HexB for (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

M2M-B071: SASK1 (top); SASK2 (bottom)

```
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'
```

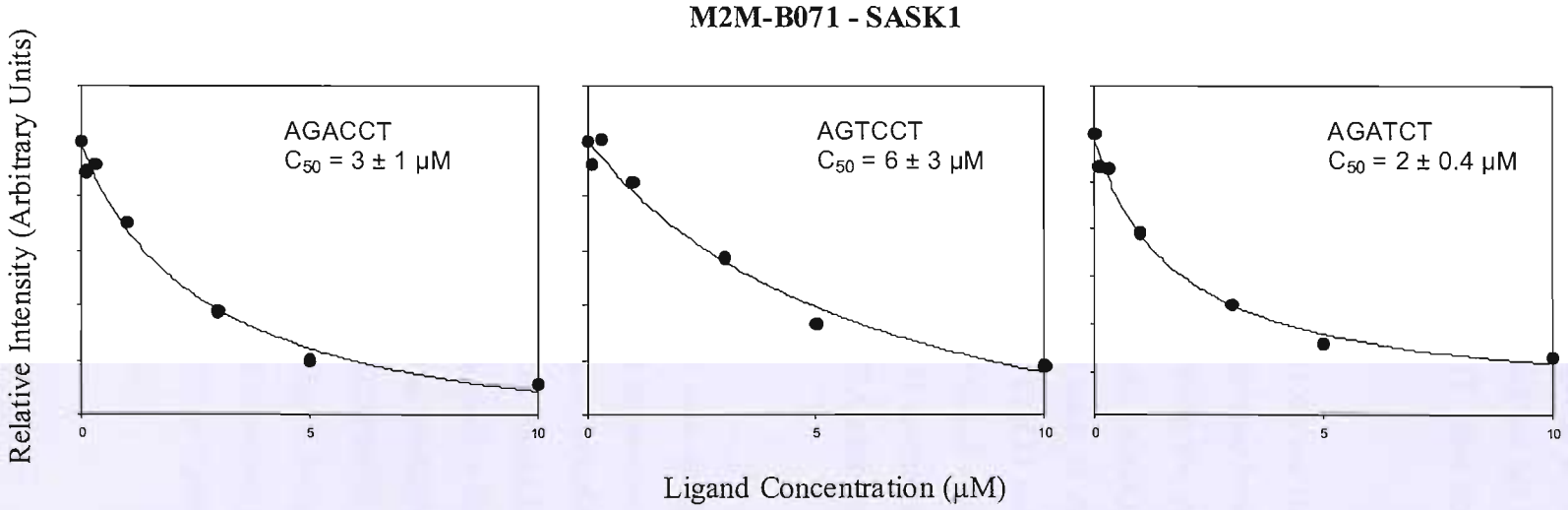


Figure 5.12: [Top] Binding sites for M2M-B071 on the tested fragments (from Figure 5.11). [Above] Example of footprinting plots for the interaction of M2M-B071 with binding sites in the SASK1 DNA fragment.

HexA: Only four footprints are evident on this fragment at concentrations of 25 μM or below, with AGATCT (site 2) and AGTACT (site 4) the best. This supports the proposal that WGWWCW is the binding site. Weaker footprints are evident at GGTACC (site 3) and GGATCC (site 1), suggesting that GC base pairs around the central GWWC weaken the binding, possibly through steric hindrance of the 2-amino group of guanine.

The hydroxyl radical footprinting gels show clear footprints at NGATCN sites (where N is any base) and some GC sites (particularly those which are not flanked by GC base pairs). A/T-tracts, except those containing alternating A/T, also show attenuated cleavage, suggesting that there is also some 1:1 binding.

HexB: TGTACA (site 7) and TGATCA (site 8, with TGACGTCG) are the best binding sites on the HexB fragment, supporting WGWWCW as the preference binding sequence. CGATCG is not protected, confirming that GC base pairs flanking the central GWWC weaken the binding. Long A/T-tracts such as TAATTA (site 2, with GGCT) and TTATAA (site 5, with AGCT) are bound more strongly than sequences such as AGCT (site 1), TGCC (site 3), TGCA (site 4) and AGCGCGCT (site 6), whilst CGCG and GGCC show no binding. This may indicate some 1:1 binding. ATTAATT (site 9) is a weak binding site.

Hydroxyl radical footprinting experiments with M2M-B071 on the HexB fragments show that A/T-tracts without alternating bases are bound. TGTACA also shows attenuated cleavage, although binding to TGATCA is less clear.

SASK1/2: It can be seen that M2M-B071 has bound only three sites on the specifically-designed SASK1/2 fragments. These fragments were designed to contain variations of $(G/C)_4$-tracts flanked by AT base pairs (*i.e.* sites such as AGCGCT, AGGCCA, AGCCCT etc); these hexanucleotide sites are separated by TGCT or TGCA. AGACCT (site 1) and AGATCT (site 3) are the strongest sites, with AGTCCT (site 2) a weaker site. The binding at site 3 is consistent with the suggestion that WGWWCW is the preferred binding site. Site 1 and 2 could either correspond to binding to GWC or to WGWWCW sites in which one of the central AT base pairs is replaced by cytosine. No binding is evident at GC sites that are flanked by GC base pairs (GGGCCC, CGCG and CGCGCG are not protected).

Hydoxyl radical footprinting experiments with M2M-B071 are presented later in the chapter, with the comparable ligand M2M-B097 (Figure 5.37).

	M2M-B071									
MS1	Site 1 TGCA, AGTCG 4 ± 2	Site 2 AGCT, CGTTCT 4 ± 1	Site 3 AGCA 6 ± 2	Site 4 AAGAGTTAT 4 ± 0.5	Site 5 TGTCA 1 ± 0.3					
MS2	TGACA 0.2 ± 0.1	ATAACTCTT 6 ± 2	TGCT 7 ± 2	AGAACT, AGCT 2 ± 0.4	CGACT, TGCA 2 ± 0.6					
HexAfor	Site 1 GGATCC 19 ± 7	Site 2 AGATCT 3 ± 1	Site 3 GGTACC 10 ± 3	Site 4 AGTACT 3 ± 1						
HexArev2	AGTACT 2 ± 1	GGTACC 12 ± 5	AGATCT 2 ± 0.5	GGATCC 19 ± 13						
HexBfor	GGATCC 67 ± 23	AGCT 31 ± 15	TAATTA, AGCC 20 ± 9	GGCA 19 ± 8	TGCA 13 ± 5	AGCT, TTATAA 13 ± 6	AGCGCGCT 22 ± 11	TGTACA 9 ± 5	CGACGTCA, TGATCA 6 ± 4	AATTAAT, TGCA 11 ± 5
HexBrev	Site 9 TGCA, ATTAATT 25 ± 9	Site 8 TGATCA, TGACGTCG 9 ± 4	Site 7 TGTACA 6 ± 3	Site 6 AGCGCGCT 23 ± 15	Site 5 TTATAA, AGCT 10 ± 6	Site 4 TGCA 17 ± 8	Site 3 TGCC 18 ± 6	Site 2 GGCT, TAATTA 8 ± 4	Site 1 AGCT 18 ± 7	
SASK1	Site 1 AGACCT 3 ± 1	Site 2 AGTCCT 6 ± 3	Site 3 AGATCT 2 ± 0.4							
SASK2	Site 3 AGATCT 2.5 ± 1	Site 2 AGGACT 4.5 ± 1	Site 1 AGGTCT 1.7 ± 0.4							

Table 5.2: C_{50} values (μM) derived from quantitative analysis of the M2M-B071 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

M2M-B072

Polyamide binding rules usually do not place two imidazole rings opposite each other (White *et al.*, 1997a; Walker *et al.*, 1998b), so the most likely binding configuration for the M2M-B072 ligand is shown in Figure 5.13.

```
Mp-Bzi-Py-Im-Im
     Im-Im-Py-Bzi-Mp
5'-W   G  G  C  C   W
```

Figure 5.13: Possible DNA binding modes for M2M-B072.

Figure 5.14 shows the results of DNase I footprinting experiments with M2M-B072 on MS1, HexAfor, HexBfor, SASK1 and SASK2 together with hydroxyl radical footprinting experiments on HexAfor and HexBrev. Figure 5.15 shows the sequences of these fragments with the observed footprinting binding sites highlighted. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.3.

MS1/MS2: It can be seen that M2M-B072 has bound every $(G/C)_4$ or longer site on the MS1/MS2 fragment, except CCCG (which bears no resemblance to the proposed binding site in Figure 5.13). The only GGCC sequence in these fragments has been removed by a point mutation in MS1 and so is only present on MS2 (CGGCCA). This change has removed the binding to this site on MS1 (the presented strand; although cleavage is protected on MS2). The preferred binding site for M2M-B072 on this fragment is GGGCG (site 3), with GCGGC (site 1), GCGCG (site 2), GGGG (site 4) and GCGG (site 5) evident as weaker sites.

HexA: Only four footprints are evident on this fragment. AGGCCT (site 2) is the best site, with TGGCCA (site 4) weaker, suggesting that all six base pairs are involved in defining the best ligand binding and that surrounding YpR steps alter the affinity for the central GGCC. GGCGCC (site 1) shows a weaker footprint and TGCGCA (site 3) is weaker still. This suggests a central GCGC can be bound by the ligand; either the central CG is tolerated within the proposed binding mode or there is secondary binding to GC alone (*i.e.* only the central Py-Im of the ligands binds as a 2:1 complex with the DNA). There is no interaction with the G/C-tracts CCGG; CCGCGG; and CCCGGG.

Hydroxyl radical footprinting experiments also show attenuated cleavage at these three sites (GGCGCC, AGGCCT and TGGCCA), supporting the DNase I results.

HexB: It can be seen that M2M-B072 has bound every $(G/C)_4$ on the HexB fragments, except CCGGCCG and GCCGGC (both of which contain CCGG, which showed no binding with HexA). GGGCCC (site 2) and CGCGCG (site 4, with CGCG) are the best binding sites, with GCGCGC (site 3) and TCGCGA (site 1) weaker. Site 4 probably appears stronger than sites 3 and 1 as the presence of two possible binding sites in the footprint of site 4 will affect the total binding affinity of the footprint. While GGGCCC contains the proposed GGCC binding site, the interaction with sites that contain alternating G/C-tracts suggests the presence of a secondary binding mode in which only GC is selectively bound. The observation that ATGCAT (site 5) is bound well (stronger than GCGCGC and TCGCGA) is consistent with this suggestion, though TTGCAA and AAGCTT show no cleavage protection. It therefore appears that Y/R steps are preferred around the secondary GC binding site.

Hydroxyl radical footprinting confirms the interaction with three of the strongest DNase I binding sites, with footprints evident at CGCGCG, ACGCGT and ATGCAT. It is not clear if there is attenuated cleavage at GGGCCC. It is interesting to note that ACGCGT shows cleavage protection whilst TCGCGA and GCGCGC do not, emphasising the importance of six base pairs in ligand binding.

SASK1/SASK2: Five clear footprints are evident on these fragments. AAGGCCTT (site 4) is the strongest site, with TAGGCCAT (site 1) slightly weaker, suggesting that AGGCCT is preferred over AGGCCA. ATGGCCAT (site 3) and TTGGCCAA (site 5) are weaker sites, suggesting that TGGCCA (sites 3 and 5) produces weaker binding than AGGCCT (sites 4 and 1). Many other $(G/C)_4$ sites are not bound by the ligand (*e.g.* ACCGGT).

A footprint is evident at ACGCGT (site 2) but there is no interaction with AGCGCT or AGCCCT, supporting the suggestion that Y/R steps are preferred around the secondary (GC) binding site.

Hydroxyl radical footprinting experiments with M2M-B072 are presented later in this chapter, along with the ligands M2M-B073 and M2M-B074 (Figure 5.23).

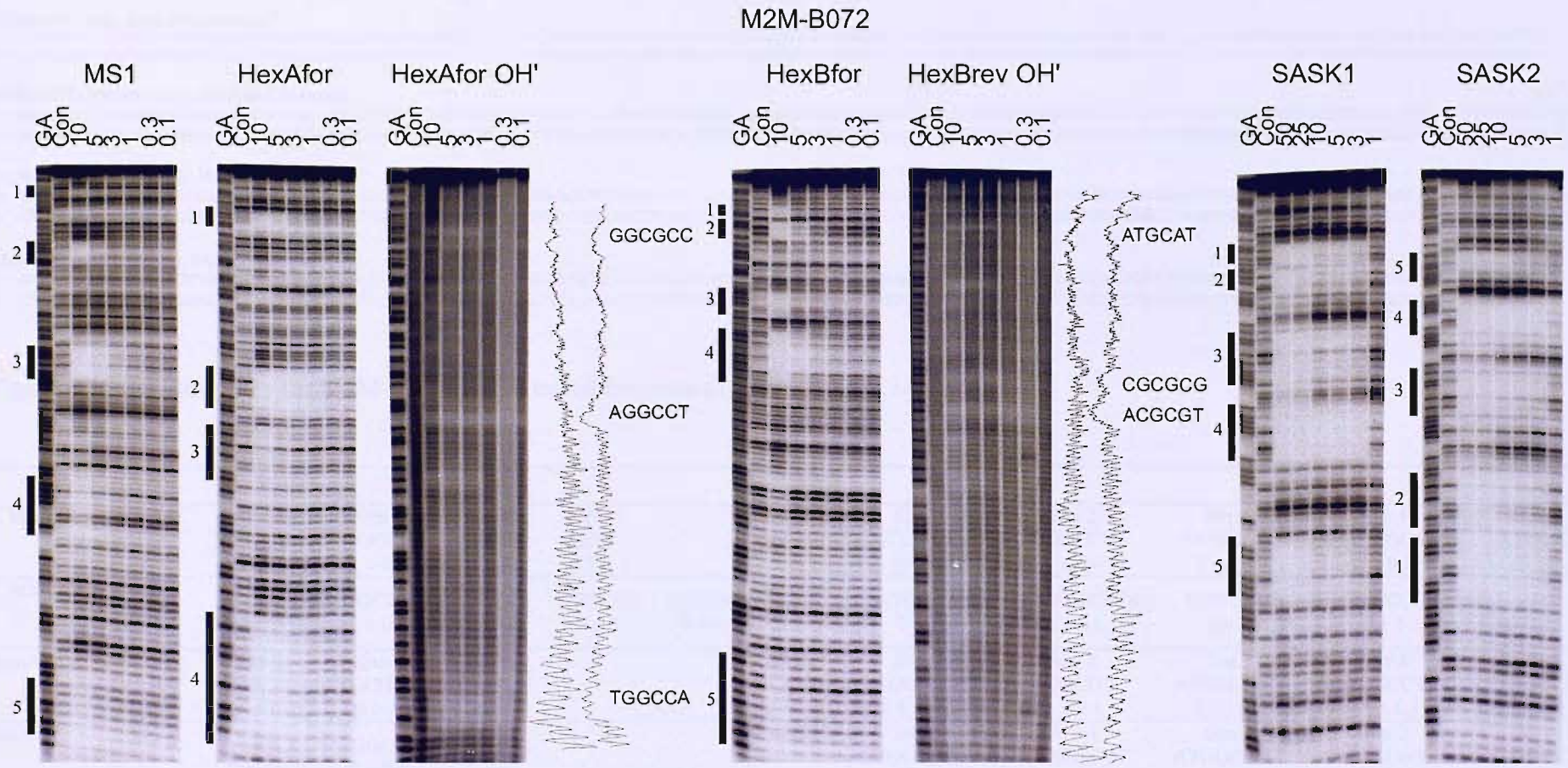


Figure 5.14: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B072 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.15; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 1 μM lane). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B072: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

M2M-B072: HexAfor (top); HexArev2 (bottom)
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'

M2M-B072: HexBfor (top); HexBrev (bottom)
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'

M2M-B072: SASK1 (top); SASK2 (bottom)
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'

Figure 5.15: Binding sites for M2M-B072 on the tested fragments (from Figure 5.14).

M2M-B072												
MS1	Site 1 GCGGC 7 ± 6.6	Site 2 GCGCG 8 ± 3	Site 3 GGGCG 1 ± 0.4	Site 4 GGGG 6 ± 2	Site 5 GCGG 6.5 ± 2		HexBfor	Site 1 TCGCGA 10 ± 5	Site 2 GGGCCC 2 ± 0.4	Site 3 GCGCGC 15 ± 10	Site 4 CGCG,CGCGCG 2 ± 0.3	Site 5 ATGCAT 6 ± 3
MS2	CCGC 10 ± 4	CCCC 7 ± 3	CGCCC 2 ± 0.6	CGCGC 7 ± 3	CGGCCA 6 ± 3	GCCGC 8 ± 4	HexBrev	ATGCAT 7 ± 3	CGCGCG,CGCG 2.5 ± 1	GCGCGC 6 ± 2	GGGCCC 2 ± 1	
HexAfor	Site 1 GGCGCC 6 ± 1	Site 2 AGGCCT 0.06 ± 0.01	Site 3 TGCGCA 26 ± 12	Site 4 TGGCCA 1.5 ± 0.3			SASK1	Site 1 TAGGCCAT 0.6 ± 0.2	Site 2 ACGCGT 1 ± 0.4	Site 3 ATGGCCAT 1.3 ± 0.7	Site 4 AAGGCCTT 0.4 ± 0.1	Site 5 TTGGCCAA 2.4 ± 1
HexArev2	TGGCCA 6 ± 3	TGCGCA 11 ± 6	AGGCCT 1 ± 0.7	GGCGCC 5 ± 3			SASK2	Site 5 TTGGCCAA 18 ± 8	Site 4 AAGGCCTT 0.2 ± 0.08	Site 3 ATGGCCAT 1.4 ± 0.8	Site 2 ACGCGT 3 ± 1.7	Site 1 ATGGCCTA 0.6 ± 0.2

Table 5.3: C_{50} values (μM) derived from quantitative analysis of the M2M-B072 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

M2M-B073

M2M-B073 has a very similar structure to M2M-B071; the only difference is the substitution of a pyridoimidazole for benzimidazole. Assuming that pyridoimidazole acts as a hydrogen bond donor, three possible binding sites can be proposed, as shown in Figure 5.16.

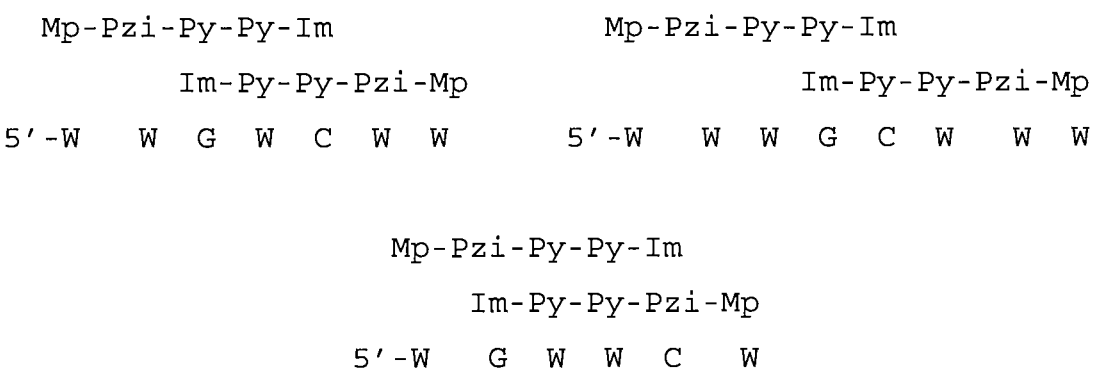


Figure 5.16: Possible DNA binding modes for M2M-B073.

Figure 5.17 shows the results of DNase I footprinting experiments with M2M-B073 on MS1, HexAfor, HexBfor, SASK1 and SASK2, together with hydroxyl radical footprinting experiments on HexAfor and HexBrev. Figure 5.18 shows the sequences of these fragments with the observed footprinting sites highlighted, together with an example of footprinting plots derived from the binding of M2M-B073 to HexAfor. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.4.

MS1/MS2: It can be seen that the best binding sites for M2M-B073 on MS1/MS2 are TGTCA (site 4) and AGTCG (site 1), with CGTTCT (site 2) a weaker site. This suggests that the ligand preferentially binds GWC over GWWC. No binding is observed at GGTCTG (as with M2M-B071), suggesting that G/C bases around the central binding site reduce the affinity. There is no evidence of binding to an isolated GC (top right in Figure 5.16), but the footprints at $(A/T)_4$ sites (TTAT and TAAA, site 3) indicate the possibility of 1:1 binding.

HexA: Only two clear footprints are evident on HexA, with AGATCT (site 1) and AGTACT (site 2) binding with similar affinity. The lack of binding to GGTACC reiterates the findings of MS1/MS2 that surrounding GC base pairs weaken binding. No isolated GC

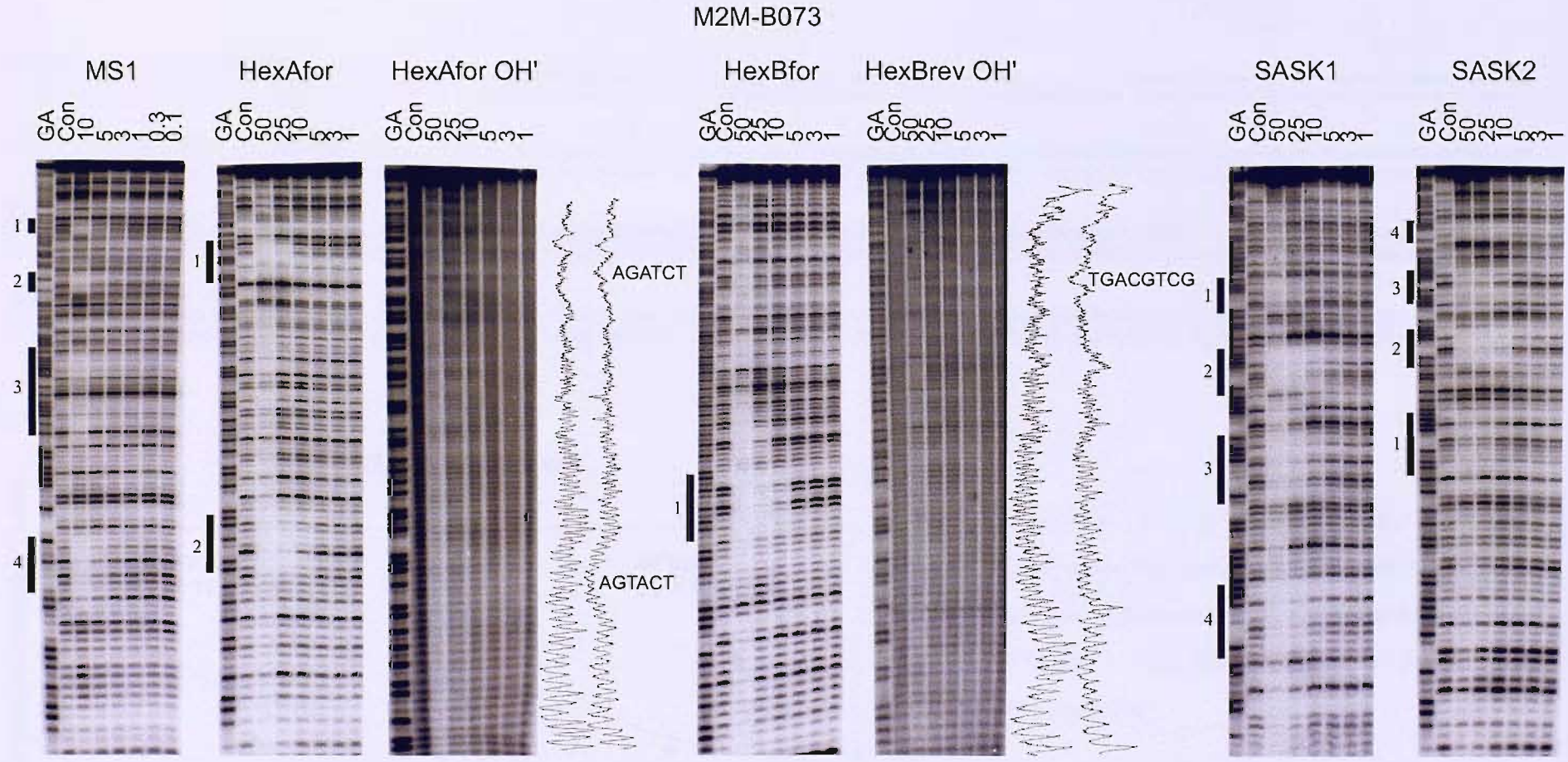


Figure 5.17: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B073 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.18; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 3 μM lane for HexAfor and 5 μM for HexBrev). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B073: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

M2M-B073: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

M2M-B073: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

M2M-B073: SASK1 (top); SASK2 (bottom)

```
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'
```

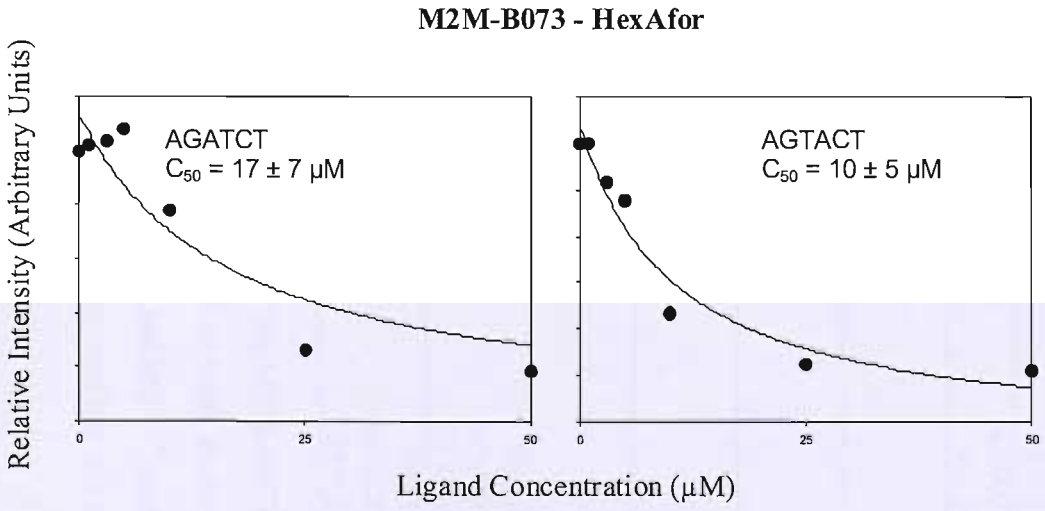


Figure 5.18: [Above] Binding sites for M2M-B073 on the tested fragments (from Figure 5.17). [Left] Example of footprinting plots for the interaction of M2M-B073 with binding sites in the HexAfor DNA fragment.

steps are bound by the ligand, while no GWC sites are present on the fragment.

Hydroxyl radical footprinting experiments support the DNase I findings, with just AGATCT and AGTACT showing cleavage protection.

HexB: CGACGTCA is the only sequence bound by M2M-B073 on the HexB fragment, suggesting the GWC binding mode of Figure 5.16 is preferred. No binding is evident at any of the GWWC (CGATCG, CGTACG and TGTACA) sites. This is contrary to the HexA results, indicating that GWC (not present on HexA) is the primary binding site, with GWWC a secondary target (therefore not bound on HexB). No binding is seen at GC sites.

Hydroxyl radical footprinting experiments also show that TGACGTCG to be selectively bound by the ligand, with the densitometer plots only differing significantly at this sequence.

	M2M-B073			
MS1	Site 1 AGTCG 5 ± 1	Site 2 CGTTCT 6 ± 1	Site 3 TTAT,TAAA 9 ± 2	Site 4 TGTCA 3 ± 1
MS2	TGACA 6 ± 2	ATAA 10 ± 2	AGAACG 8 ± 2	CGACT 4 ± 1
HexAfor	Site 1 AGATCT 17 ± 7	Site 2 AGTACT 10 ± 5		
HexArev2	AGTACT 14 ± 7	AGATCT 14 ± 9		
HexBfor	Site 1 CGACGTCA 6 ± 3			
HexBrev	TGACGTCG 5 ± 1			
SASK1	Site 1 AGCCCT 30 ± 9	Site 2 AGACCT 8 ± 4	Site 3 AGTCCT 15 ± 8	Site 4 AGATCT 17 ± 6
SASK2	Site 4 AGATCT -	Site 3 AGGACT 19 ± 6	Site 2 AGGTCT 4 ± 2	Site 1 AGGGCT 50 ± 31

Table 5.4: C_{50} values (μM) derived from quantitative analysis of the M2M-B073 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). Sites marked "–" correspond to footprints evident by visual inspection, but for which quantitative analysis did not provide a clear value.

SASK1/SASK2: It can be seen that AGACC (site 2) is the best binding site on the SASK1/SASK2 fragment, suggesting GWC is the preferred binding sequence. AGTCCT (site 3) is a weaker site, along with AGATCT (site 4). AGCCCT (site 1) is the weakest binding site and is the only GC site showing cleavage protection (possibly because the cytosine adjacent to the guanine is tolerated within the GWC binding conformation).

Hydoxyl radical footprinting experiments with M2M-B073 are presented later in this chapter, with the ligands M2M-B072 and M2M-B074 (Figure 5.23).

M2M-B074

Binding of two imidazoles opposite each other in a 2:1 complex is unlikely to occur (White *et al.*, 1997a; Walker *et al.*, 1998b), so M2M-B074 (Mp-Pzi-Py-Py-Im-Im) can bind as a homodimer in two ways, as illustrated in Figure 5.19.

```
   Mp-Pzi-Py-Py-Im-Im                Mp-Pzi-Py-Py-Im-Im
        Im-Im-Py-Py-Pzi-Mp                Im-Im-Py-Py-Pzi-Mp
 5'-W   G  G  W  C  C   W        5'-W   W  G  G  C  C  W   W
```

Figure 5.19: Possible DNA binding modes for M2M-B074.

Figure 5.20 shows the results of DNase I footprinting experiments with M2M-B074 on MS2, HexAfor, HexBfor, SASK1 and SASK2, together with hydroxyl radical footprinting experiments on HexAfor and HexBfor. Figure 5.21 shows the sequences of these fragments with the observed footprinting binding sites highlighted. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.5.

MS1/MS2: It can be seen that M2M-B074 has bound the MS1/MS2 fragment to produce five very clear footprints. The best footprints are located at CGCCC (site 3) and CGGCCA (site 5), which are closely related to the GGCC binding mode shown in Figure 5.19. A strong footprint at CCGGACG (site 2) could be due to binding at GGCC or GWC. No footprints are observed at other $(G/C)_4$ tracts (such as CCGC, CCCC, CGGG, CGCGC and GCCGC), in contrast to M2M-B072.

ACGACTA (site 4) and ATGACAA (site 1) are also protected, though only at higher concentrations. It is possible that this corresponds to binding to GWC.

HexA: DNase I footprinting experiments with M2M-B074 on the HexA fragment show strong binding to TGGCCA (site 5) and AGGCCT (site 4), consistent with the WGGCCW binding mode. Other G/C-tracts are bound more weakly, such as GGCGCC (site 1) and CCGCGG (site 2, with TCCGGA), though there is no protection at CCCGGG, ACCGGT and WGCGCW. The observation that WGCGCW does not bind, whilst GGCGCC does, indicates long G/C-tracts are preferred for good binding if the primary binding site (GGCC) is not present.

Interestingly, AACGTT (site 3) shows a weak footprint, but binding here may be in the 1:1 mode to the surrounding long A/T-tracts (TTTAAACGTTAA, the longest A/T-tract in the tested fragments).

Hydroxyl radical footprinting experiments support these DNase I experiments, with the strongest binding sites again see at GGCGCC, CCGCGG, AGGCCT and TGGCCA.

HexB: It can be seen that the sequences GGGCCC (site 2) and GTCGACGTC (site 4) are the preferred binding sites for M2M-B074, supporting both the GGCC and GWC binding modes in Figure 5.19. Other long G/C-tracts are also bound more weakly, with GCGCGC (site 3) a much stronger site than CCGGCCG (site 1), even though the latter contains the GGCC binding site. It is interesting to note that GCCGGC and CGCGCG are not bound, indicating selectivity beyond just binding to long G/C-tracts. The only $(G/C)_4$ sequence, ACGCGT, is also not bound.

Hydroxyl radical footprinting experiments support these DNase I results, although binding to GCGCGC is unclear (CGGCCG and GGGCCC still show cleavage protection). It is interesting to note that the GTCGACGTC site is clearly split into two footprints around GTCGAC and GTC using hydroxyl radicals, both of which support the GWC binding mode.

SASK1/SASK2: DNase I cleavage is protected at every GGCC site on the SASK1/SASK2 fragment, with TAGGCCAT (site 1) and AAGGCCTT (site 4) the strongest sites. ATGGCCAT (site 3) and TTGGCCAA (site 5) are slightly weaker sites, suggesting that YGGCCR is a weaker site (supported by the weaker binding to CGGCCG on HexB). The only other $(G/C)_4$ sequence bound is AAGCCCTT (site 2), though this requires higher concentrations than the other sites. No footprints are evident at AGCGCT, ACGCGT nor ACCGGT.

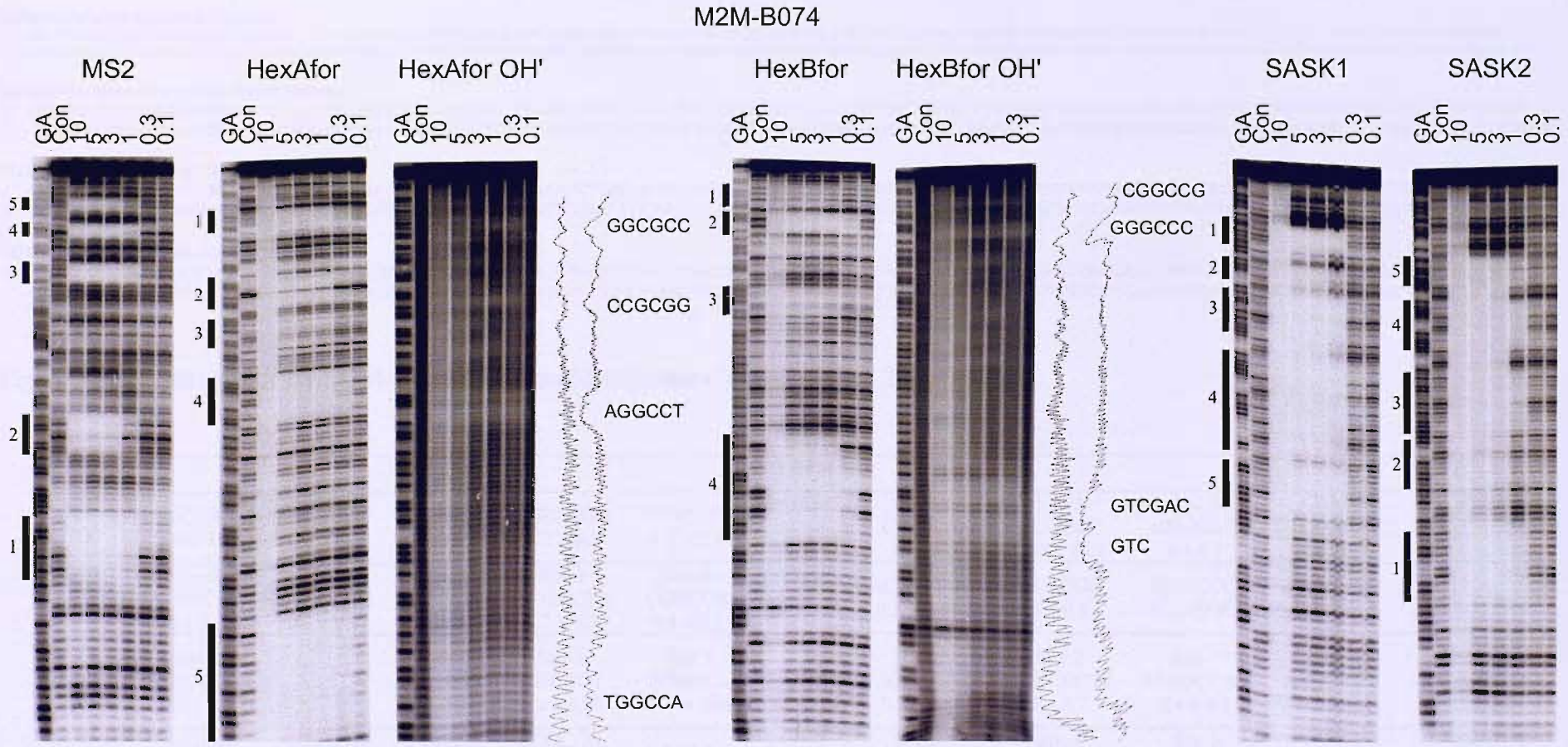


Figure 5.20: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B074 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.21; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 0.3 μM lanes). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B074: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

M2M-B074: HexAfor (top); HexArev2 (bottom)
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'

M2M-B074: HexBfor (top); HexBrev (bottom)
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'

M2M-B074: SASK1 (top); SASK2 (bottom)
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'

Figure 5.21: Binding sites for M2M-B074 on the tested fragments (from Figure 5.20).

M2M-B074											
MS1	TGGCAG 20 ± 16	TAGTCGT 0.5 ± 0.2	GGGCG 0.2 ± 0.1	CGTCCGG 0.2 ± 0.1	TTGTCAT 0.7 ± 0.3	HexBfor	Site 1 CCGGCCG 15 ± 6	Site 2 GGGCCC 0.06 ± 0.02	Site 3 GCGCGC 3 ± 0.7	Site 4 GTCGACGTC 0.06 ± 0.01	
MS2	Site 5 ATGACAA 2 ± 0.7	Site 4 CCGGACG 0.3 ± 0.1	Site 3 CGCCC 0.3 ± 0.1	Site 2 ACGACTA 0.7 ± 0.3	Site 1 CGGCCA 0.4 ± 0.1	HexBrev	GACGTCGAC 0.1 ± 0.04	GCGCGC 3 ± 0.9	GGGCCC 0.1 ± 0.03		
HexAfor	Site 1 GGCGCC 0.3 ± 0.2	Site 2 CCGG, CCGCGG 0.5 ± 0.2	Site 3 AACGTT 3 ± 1	Site 4 AGGCCT 0.1 ± 0.03	Site 5 TGGCCA 0.05 ± 0.01	SASK1	Site 1 TAGGCCAT 0.3 ± 0.1	Site 2 AAGCCCTT 1 ± 0.7	Site 3 ATGGCCAT 0.4 ± 0.2	Site 4 AAGGCCTT 0.3 ± 0.1	Site 5 TTGGCCAA 0.3 ± 0.1
HexArev2	TGGCCA 0.2 ± 0.05	AGGCCT 1 ± 0.4	AACGTT 11 ± 7	CCGCGG, CCGG 3 ± 1	GGCGCC 2 ± 1	SASK2	Site 5 TTGGCCAA 1 ± 0.7	Site 4 AAGGCCTT 0.3 ± 0.1	Site 3 ATGGCCAT 0.4 ± 0.2	Site 2 AAGGGCTT 2 ± 1	Site 1 ATGGCCTA 0.4 ± 0.2

Table 5.5: C_{50} values (μM) derived from quantitative analysis of the M2M-B074 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

SASK1/SASK2 Hydroxyl radical footprinting:

M2M-B072, M2M-B073 and M2M-B074

Figure 5.23 shows the results of hydroxyl radical footprinting experiments with M2M-B072, M2M-B073 and M2M-B074 on SASK1. These have been grouped together so as to emphasise the difference in sequence selectivity between these related ligands.

It can be seen that both M2M-B072 and M2M-B074 bind to sites containing GGCC sequences, with TAGGCCAT, ATGGCCAT, AAGGCCTT and TTGGCCAA all showing attenuated cleavage. Some variation in the binding sequence can be tolerated by M2M-B072, with AACGCGTT and AAGACCTT bound, whilst M2M-B074 has a more relaxed selectivity, binding AACGCGTT, AAGCCCTT, AAGACCTT and AAGTCCTT (the latter two sites possibly correspond to binding to GWC sites). M2M-B073 produces hydroxyl radical footprints at the same sites seen with DNase I footprinting; AAGACCTT, AAGTCCTT and AAGTACTT, supporting both the 3- and 4-overlap binding modes shown in Figure 5.16.

M2M-B079

The M2M-B079 only differs in structure from M2M-B074 by the substitution of the pyridoimidazole for a benzimidazole. The proposed configurations of binding are therefore the same as M2M-B074 and are shown in Figure 5.22.

```
   Mp-Bzi-Py-Py-Im-Im                Mp-Bzi-Py-Py-Im-Im
        Im-Im-Py-Py-Bzi-Mp                   Im-Im-Py-Py-Bzi-Mp
 5'-W   G  G  W  C  C   W        5'-W   W  G  G  C  C  W   W
```

Figure 5.22: Possible DNA binding modes for M2M-B079.

Figure 5.24 shows the results of DNase I footprinting experiments with M2M-B079 on MS2, HexAfor, HexBfor, SASK1 and SASK2, together with hydroxyl radical footprinting experiments on HexAfor and HexBfor. Figure 5.25 shows the sequences of these fragments with the observed footprinting binding sites highlighted. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.6.

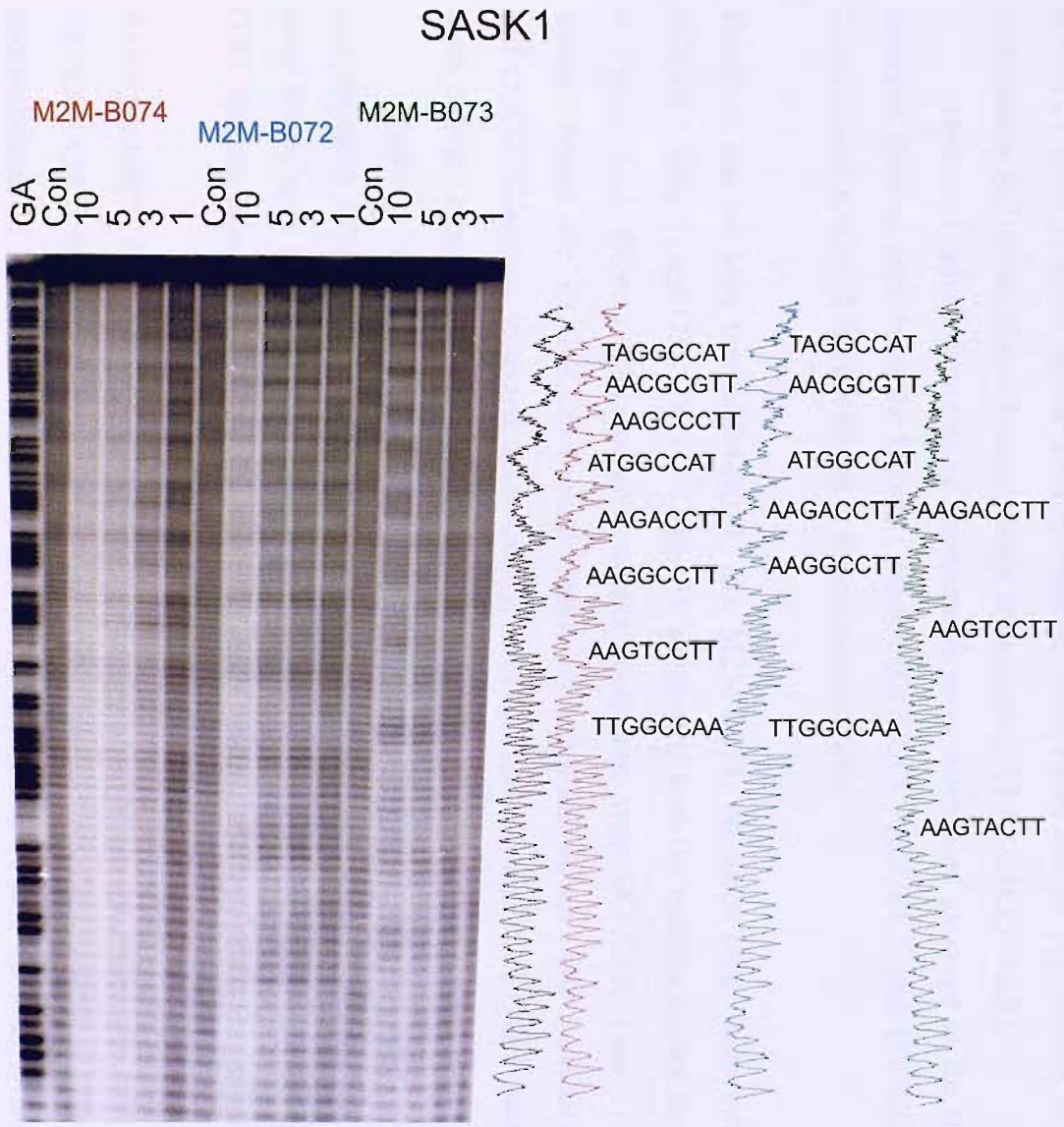


Figure 5.23: Hydroxyl radical footprinting gel of M2M-B074, M2M-B072 and M2M-B073 with the SASK1 fragment. Only the clearest footprints are marked. Densitometer plots were derived from the 1 μM lane for M2M-B074 and M2M-B072; and the 3 μM lane for M2M-B073. Densitometer plots are colour-coded with the ligand names above the gel (the control densitometer plot is black). GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane.

MS1/MS2: It can be seen that M2M-B079 has similar binding selectivity to the related M2M-B074 (Figure 5.20), but requires higher ligand concentrations (has weaker affinity). The strongest binding sites are at CCGGACG (site 2), CGCCC (site 3) and CGGCCA (site 6), with weaker binding at ACGACTA (site 5) and ATGACAA (site 1). These sequences support both of the binding modes suggested in Figure 5.22, but with mismatches widely tolerated. There is no protection of the G/C-tracts CCGC, CCCC, CGCGC and GCCGC, as with M2M-B074, but CGGGAGC is bound weakly, indicating M2M-B079 is less selective.

HexA: M2M-B079 has bound to HexA with similar selectivity to M2M-B074, except that TGGCCA (site 5) is a better site than AGGCCT (site 4). GGCGCC (site 1) has bound with comparable affinity to AGGCCT, but CCGCGG (site 2, with TCCGGA) is a weak binding sequence. The other G/C-tracts (CCCGGG, ACCGGT and WGCGCW) again show no cleavage protection, indicating a more specific binding mode than just long G/C-tracts. AACGTT (site 3) again shows weak protection, most likely due to 1:1 binding to the surrounding A/T-tracts (this is located within the sequence TTTAAACGTTAA).

Hydroxyl radical footprinting experiments show attenuated cleavage at the three strongest binding sites from the DNase I studies (GGCGCC, AGGCCT and TGGCCA); CGCGG and AACGTT are not bound at these concentrations.

HexB: It can be seen that M2M-B079, like M2M-B074, produces clear footprints at GGGCCC (site 1) and GTCGACGTC (site 3), supporting both the binding modes shown in Figure 5.22. Other G/C-tracts give weaker footprints, with GCGCGC (site 2) the weakest bound site. CCGGCCG (bound weakly by M2M-B074), GCCGGC, ACGCGT and CGCGCG are not protected. Interestingly, AATATT (site 4) and AATTAAT (site 5) show stronger binding than GCGCGC, maybe indicating a preference for 1:1 binding.

Hydroxyl radical footprinting experiments show attenuated cleavage only at the two strongest DNase I footprinting sites: GGGCCC and GTCGACGTC. However, the latter site is split into two distinct footprints (GTCGAC and GTC), suggesting multiple GWC binding sites and explaining the strong DNase I binding affinity for the longer site.

SASK1/SASK2: Every GGCC-containing site is bound universally by M2M-B079 on the SASK1/SASK2 fragments and a single very long region of protection is evident at concentrations of 3 μM and above. The selectivity appears to be less than M2M-B074;

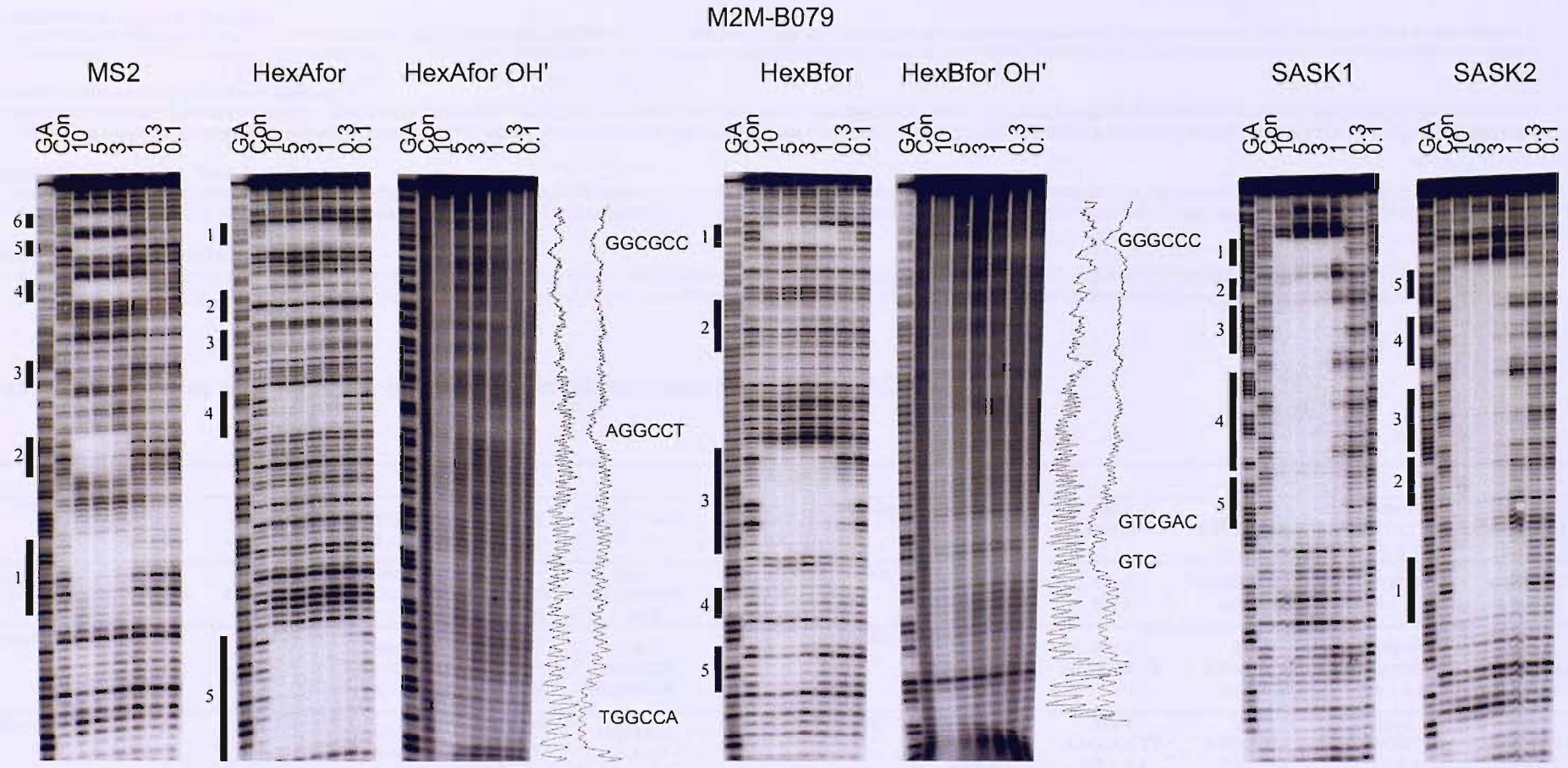


Figure 5.24: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B079 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.25; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 0.3 μM lanes). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B079: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTACGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

M2M-B079: HexAfor (top); HexArev2 (bottom)
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'

M2M-B079: HexBfor (top); HexBrev (bottom)
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'

M2M-B079: SASK1 (top); SASK2 (bottom)
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCCGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'

Figure 5.25: Binding sites for M2M-B079 on the tested fragments (from Figure 5.24).

M2M-B079												
MS1	TGGCAG 16 ± 9	TAGTCGT 0.5 ± 0.2	GCTCCCG 2.8 ± 0.5	GGGCG 0.2 ± 0.1	CGTCCGG 0.3 ± 0.1	TTGTCAT 0.5 ± 0.2	HexBfor	Site 1 GGGCCC 0.2 ± 0.06	Site 2 GCGCGC 12 ± 3	Site 3 GTCGACGTC 0.1 ± 0.02	Site 4 AATATT 2 ± 0.7	Site 5 AATTAAT 2 ± 0.4
MS2	Site 6 ATGACAA 2 ± 0.5	Site 5 CCGGACG 0.9 ± 0.3	Site 4 CGCCC 0.6 ± 0.2	Site 3 CGGGAGC 5.5 ± 1.4	Site 2 ACGACTA 1 ± 0.5	Site 1 CGGCCA 0.8 ± 0.4	HexBrev	ATTAATT 8 ± 6	AATATT 8 ± 5	GACGTCGAC 0.7 ± 0.3	GCGCGC 9 ± 5	GGGCCC 0.5 ± 0.3
HexAfor	Site 1 GGCGCC 0.2 ± 0.07	Site 2 CCGG, CCGCGG 1.4 ± 0.7	Site 3 AACGTT 6 ± 2	Site 4 AGGCCT 0.2 ± 0.05	Site 5 TGGCCA 0.05 ± 0.01		SASK1	Site 1 TAGGCCAT 0.5 ± 0.2	Site 2 AAGCCCTT 1 ± 0.8	Site 3 ATGGCCAT 0.5 ± 0.3	Site 4 AAGGCCTT 0.6 ± 0.4	Site 5 TTGGCCAA 0.6 ± 0.4
HexArev2	TGGCCA 0.3 ± 0.1	AGGCCT 1.4 ± 0.5	AACGTT 9 ± 2	CCGCGG. CCGG 5 ± 1	GGCGCC 1.5 ± 0.7		SASK2	Site 5 TTGGCCAA 0.8 ± 0.3	Site 4 AAGGCCTT 0.9 ± 0.4	Site 3 ATGGCCAT 0.6 ± 0.2	Site 2 AAGGGCTT 1 ± 0.5	Site 1 ATGGCCTA 0.5 ± 0.2

Table 5.6: C_{50} values (μM) derived from quantitative analysis of the M2M-B079 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

TAGGCCAT (site 1), ATGGCCAT (site 3), AAGGCCTT (site 4) and TTGGCCAA (site 5) bind with similar affinity. The only other $(G/C)_4$ sequence to be bound is AAGCCCTT (site 2), which contains one base different from the proposed best site (*i.e.* GCCC instead of GGCC) but this is not bound as strongly as with M2M-B074. No footprints are evident at AGCGCT, ACGCGT and ACCGGT.

Hydoxyl radical footprinting experiments with M2M-B079 are presented later in this chapter, with the ligands M2M-B084, M2M-B088 and M2M-B099 (Figure 5.41).

M2M-B083

M2M-B083 is unique in the Series B set of compounds due to the inclusion of a gamma amino butyric acid linkage. This has been found to be the optimum linker length for a hairpin polyamide (Mrksich *et al.*, 1994) and so M2M-B083 may be able to bind as a hairpin dimer in the configuration shown in Figure 5.26.

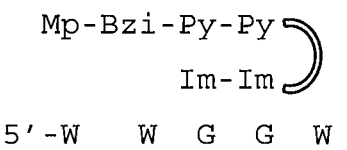
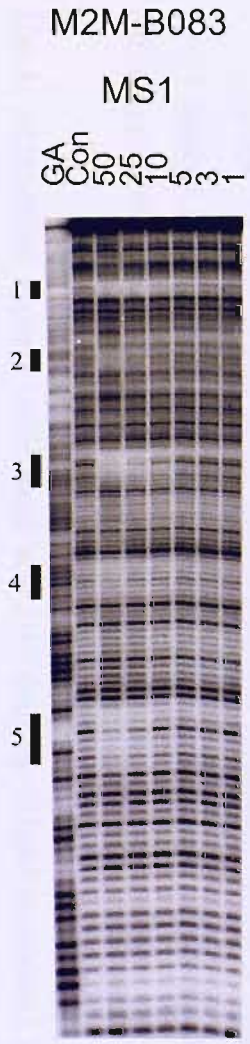


Figure 5.26: Possible DNA binding mode for M2M-B083. The curved structure represents the γ-linkage.

Figure 5.27 shows the results of DNase I footprinting experiments with M2M-B083 on MS1. The footprinting sites are highlighted against the sequence of this fragment and C_{50} values for the interaction with all the sites on this fragment are summarised in Table 5.7.

It can be seen that M2M-B083 has chiefly bound A/T-tracts (indicative of selective binding to just one strand), with AATTA (site 3) the preferred site. TTTT (site 5) and TAAA (site 4) are much weaker binding sites, suggesting that an ApT step is preferred by the ligand for strong binding. ATTTCCAA (site 1) is also a strong binding site and includes the proposed binding sequence in Figure 5.26. However, the weaker binding to AAGGTT (site 2) suggests the former sequence is being targeted at the adjacent A/T-tract.

Since the proposed binding site (WWGGW) is bound more weakly than simple A/T-tracts, this ligand was not investigated further.



M2M-B083					
	Site 1	Site 2	Site 3	Site 4	Site 5
MS1	ATTTCCAA	AAGGTT	AATTA	TAAA	TTTT
	13 ± 4	20 ± 6	10 ± 4	63 ± 11	42 ± 12
MS2	TAATT	AACCTT	TTGGAAAT		
	3 ± 0.8	11 ± 2	6 ± 1		

Figure 5.27: [Left] DNase I footprinting gel showing the interaction of M2M-B083 with the MS1 fragment. The footprints are indicated by the filled boxes and are also highlighted on the sequence below. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane. [Below] Binding sites for M2M-B083 on the MS1/MS2 fragments.

Table 5.7 (Above): C_{50} values (μM) derived from quantitative analysis of the M2M-B083 MS1/MS2 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

M2M-B083: MS1 (top); MS2 (bottom)

5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'

3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

M2M-B084

M2M-B084 is a classic Dervan polyamide (similar to that presented in Kielkopf *et al.* (1998)), which differs from the other ligands in Series B by containing only pyrrole and imidazole ring moieties and by having a C-terminal dimethylaminopropylamide moiety. There are two configurations in which M2M-B084 (Dp-Py-Py-Py-Im-Im) can potentially bind as a 2:1 complex within the DNA minor groove, as shown in Figure 5.28.

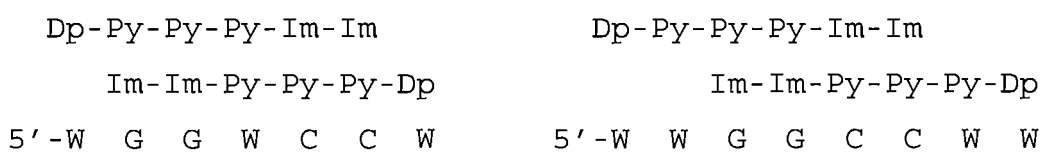


Figure 5.28: Possible binding sites for M2M-B084 on duplex DNA.

Figure 5.29 shows the results of DNase I footprinting experiments with M2M-B084 on MS2, HexAfor, HexBfor, SASK1 and SASK2, together with hydroxyl radical footprinting experiments on HexAfor and HexBfor. Figure 5.30 shows the sequences of these fragments with the observed footprinting binding sites highlighted, together with an example of footprinting plots derived from the binding of M2M-B084 to HexBfor. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.8.

MS1/MS2: It can be seen that M2M-B084 has bound CGGCCA (site 7) with the strongest affinity, so appears to be more selective for the GGCC binding mode than the similar compounds M2M-B074 and M2M-B079 (though M2M-B084 binds with a weaker affinity). Weaker binding sites such as ATGACAA (site 1), CCGGACG (site 2) and ACGACTA (site 6) suggest that binding to GWC is still occurring, though binding to CGCCC (site 3) and CGGGAGC (site 5) indicates that some other G/C-tracts are also bound. No footprints are observed at TCCGCA, ACCCCA, CGCGC and GCCGC.

TTGCTA (site 4) also shows weaker cleavage protection, which is most likely due to 1:1 binding in the surrounding A/T-tracts (TAATTGCTATA) as observed for M2M-B074 and M2M-B079.

HexA: M2M-B084 appears to be more selective than the related compound M2M-B079 (though it binds with weaker affinity) producing footprints at three sequences: TGGCCA (site 3), AGGCCT (site 2) and GGCGCC (site 1). The proposed binding sequence GGCC

is therefore represented, with GGCGCC only differing from this by one base. Combining the data for the HexAfor and HexArev2 fragments shows that these three sites have similar binding affinities, though GGCGCC is possibly weaker. Other G/C-tracts, such as CCCGGG, WCCGGW, CCGCGG and WGCGCW show no cleavage protection. Significantly, GGACC is not bound by M2M-B084, suggesting the five-unit 2:1 overlap in Figure 5.28 is not recognised (even though the central core GWC is bound at other sequences).

Hydroxyl radical footprinting experiments show attenuated cleavage at AGGCCT and TGGCCA.

HexB: It can be seen that M2M-B084 is again more selective than M2M-B079, binding preferentially to GGGCCC (site 1), which contains the proposed recognition sequence GGCC. CGCGCG (site 2, with ACGCGT) is bound more weakly, but GCGCGC and TCGCGA show no protection (the adjacent alternating G/C-tracts of site 2 are most likely creating cooperative binding of the ligand). Although there are no footprints in the other long G/C-tracts (CCGGCCG and GCCGGC), GTCGACGTC (site 3) is bound, although relatively weaker than with M2M-B079, suggesting that there may be some interaction with GWC. The A/T-selectivity seen with M2M-B079 is not apparent with M2M-B084.

Hydroxyl radical footprinting only shows binding at GGGCCC and CGCGCG, suggesting that the secondary binding to GWC is not as pronounced as with M2M-B079.

SASK1/SASK2: The binding selectivity and affinity of M2M-B084 to SASK1/SASK2 is similar to that of the related M2M-B079. TAGGCCAT (site 1), ATGGCCAT (site 3), AAGGCCTT (site 4) and TTGGCCAA (site 5) are bound with the same affinity, with AAGCCCTT (site 2; which has a single base different from GGCC) a weaker site. The other G/C-tracts (AGCGCT, ACGCGT and ACCGGT) show no cleavage protection.

Hydoxyl radical footprinting experiments with M2M-B084 are presented later in this chapter, with the ligands M2M-B079, M2M-B088 and M2M-B099 (Figure 5.41).

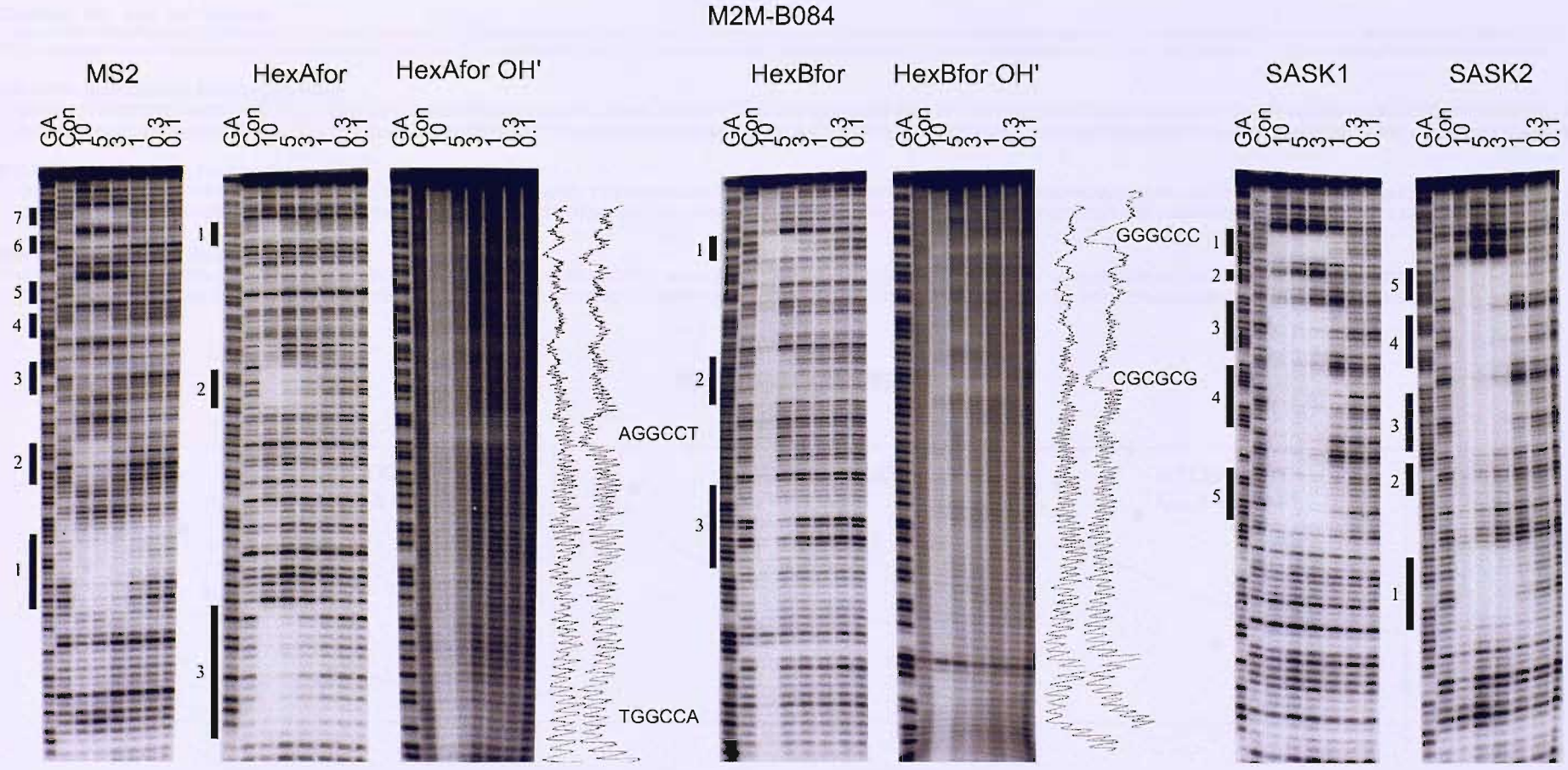


Figure 5.29: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B084 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.30; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 0.3 μM lane for HexAfor and 1 μM for HexBfor). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B084: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

M2M-B084: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

M2M-B084: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

M2M-B084: SASK1 (top); SASK2 (bottom)

```
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'
```

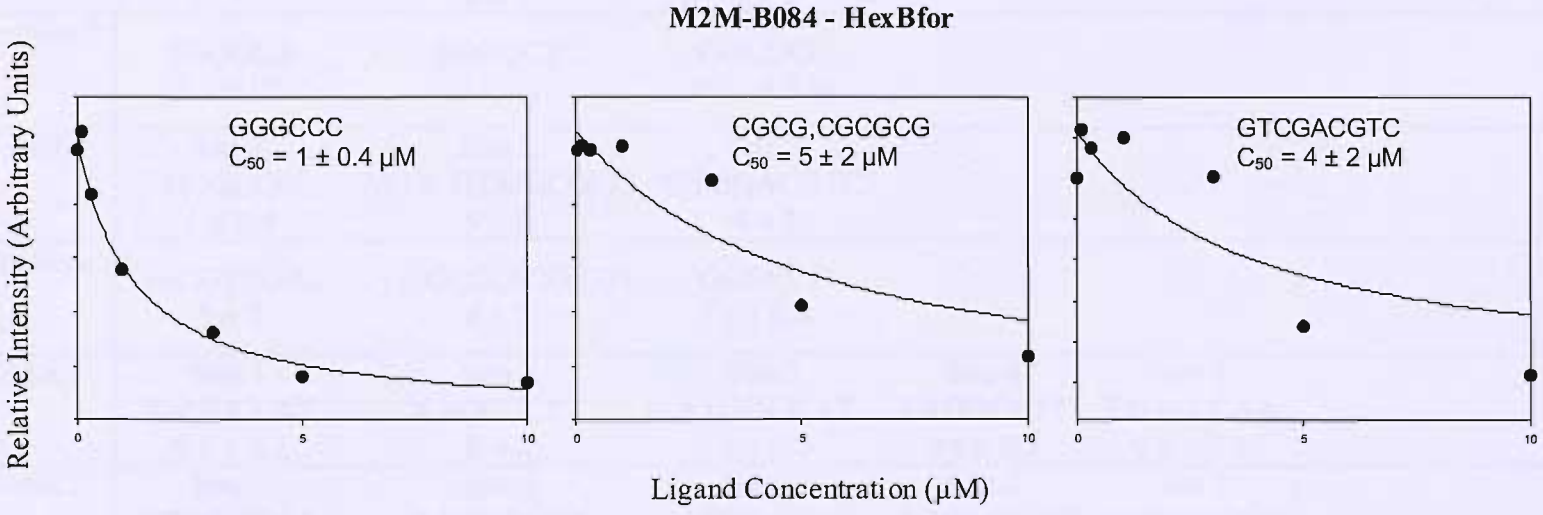


Figure 5.30: [Top] Binding sites for M2M-B084 on the tested fragments (from Figure 5.29). [Above] Example of footprinting plots for the interaction of M2M-B084 with binding sites in the HexBfor DNA fragment.

M2M-B084							
MS1	TGGCAG 31 ± 19	TAGTCGT 3 ± 0.4	GCTCCCG 4 ± 0.6	TAGCAA 5 ± 1	GGGCG 2 ± 0.5	CGTCCGG 2 ± 0.6	TTGTCAT 1.5 ± 0.7
MS2	Site 7 ATGACAA 2.4 ± 1	Site 6 CCGGACG 2 ± 0.6	Site 5 CGCCC 3 ± 1	Site 4 TTGCTA 4 ± 1	Site 3 CGGGAGC 3 ± 1	Site 2 ACGACTA 2 ± 1	Site 1 CGGCCA 0.7 ± 0.2
HexAfor	Site 1 GGCGCC 13 ± 7	Site 2 AGGCCT 4 ± 1	Site 3 TGGCCA 3 ± 1				
HexArev2	TGGCCA 2.3 ± 1.7	AGGCCT 2 ± 0.7	GGCGCC 0.07 ± 0.06				
HexBfor	Site 1 GGGCCC 1 ± 0.4	Site 2 ACGCGT,CGCGCG 5 ± 2	Site 3 GTCGACGTC 4 ± 2				
HexBrev	GACGTCGAC 5 ± 3	CGCGCG,ACGCGT 4 ± 2	GGGCCC 1 ± 0.4				
SASK1	Site 1 TAGGCCAT 0.5 ± 0.2	Site 2 AAGCCCTT 5 ± 2	Site 3 ATGGCCAT 1 ± 0.5	Site 4 AAGGCCTT 0.6 ± 0.2	Site 5 TTGGCCAA 0.6 ± 0.2		
SASK2	Site 5 TTGGCCAA 0.5 ± 0.2	Site 4 AAGGCCTT 0.5 ± 0.2	Site 3 ATGGCCAT 0.7 ± 0.3	Site 2 AAGGGCTT 3 ± 1	Site 1 ATGGCCTA 0.4 ± 0.1		

Table 5.8: C_{50} values (μM) derived from quantitative analysis of the M2M-B084 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

M2M-B088

Due to the arrangement of ring moieties in the polyamide M2M-B088, only two configurations of 2:1 binding are possible without clashing imidazoles (shown in Figure 5.31), although the second (shown on the bottom in Figure 5.31) exposes unpaired imidazole rings.

```
          Mp-Bzi-Py-Im-Py-Im
                  Im-Py-Im-Py-Bzi-Mp
    5'-W   W  G  C  G  C  W   W


        Mp-Bzi-Py-Im-Py-Im
                      Im-Py-Im-Py-Bzi-Mp
5'-W   W  W  C  G  C  G  W  W   W
```

Figure 5.31: Possible DNA binding modes for M2M-B088.

Figure 5.32 shows the results of DNase I footprinting experiments with M2M-B088 on MS1, HexAfor, HexBrev, SASK1 and SASK2, together with hydroxyl radical footprinting experiments on HexAfor and HexBfor. Figure 5.33 shows the sequences of these fragments with the observed footprinting binding sites highlighted, together with an example of footprinting plots derived from the binding of M2M-B088 to MS1. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.9.

MS1/MS2: It can be seen that the binding of M2M-B088 is generally weaker than that observed with the other Series B ligands. The only three sequences bound on MS1 are A/T-tracts, suggestive of 1:1 binding; AATTA (site 1), TAAA (site 2) and TTTT (site 3). However, several other $(A/T)_4$ sites are not bound, with no protection observed at ATAT, AATA, ATTT, TATA and TTAT, while TTAA is bound only weakly on MS2.

Weaker binding to GCCGC, GGCC and AGTGCA was also observed on the MS2 fragment (all of which differ from the proposed GCGC binding mode by one or two bases) suggesting G/C binding is secondary for M2M-B088. No other G/C-tracts, including GCGCG, showed cleavage protection on either fragment.

HexA: The binding appears to be more G/C-selective on the HexA fragments than observed with MS1/MS2 and only two long A/T-tracts show DNase I footprints:

AAATTTA (site 2) and TTTAAA (site 4, with TTAA). Binding to GCGC is seen at several places on this fragment, such as GGCGCC (site 1), TGCGCA (site 6) and AGCGCT (site 7), although up to two base permutations from this may be permitted: AGATCT (site 3), AGGCCT (site 5), AGTACT (site 8) and TGCACG (site 9) are bound. However, there is no protection at the sequences TGGCCA, CCGCGG, CCGG and CCCGGG.

Hydroxyl radical footprinting experiments show attenuated cleavage at all the DNase I footprinting sites at the top of the gel (GGCGCC, AAATTT, AGATCT, TTTAAA and AGGCCT). However, the inherently weak binding of this ligand makes it difficult to assess whether hydroxyl radical cleavage is also attenuated at the other DNase I sites.

HexB: It can be seen that M2M-B088 has mainly bound to A/T-tracts on the HexB fragments, with ATTAATT and AATATT (site 6), TATATA (site 4), TTATAA (site 3), AATT (site 2) and TAATTA (site 1) all exhibiting protection from cleavage. The binding of M2M-B088 is generally weak, and many of the G/C-tracts in this fragment (including CGCGCG, which contains the proposed GCGC binding site) do not show footprints. The only G/C-tract to show protection from cleavage is GACGTC (site 5), which contains RCGY (which is similar to the GCGC proposed site).

Hydroxyl radical footprinting experiments show attenuated cleavage at three A/T-tracts; TAATTA, AATT and AATATT.

SASK1/SASK2: Only one footprint at AGCGCT (site 1) is seen on the SASK1/SASK2 fragments. The binding to this site is relatively weak compared to the other Series B ligands, with other G/C-tracts (GGCC, CGCG, GCCC and CCGG) showing no cleavage protection.

Hydroxyl radical footprinting experiments with M2M-B088 are presented later in this chapter, with the ligands M2M-B079, M2M-B084 and M2M-B099 (Figure 5.41).

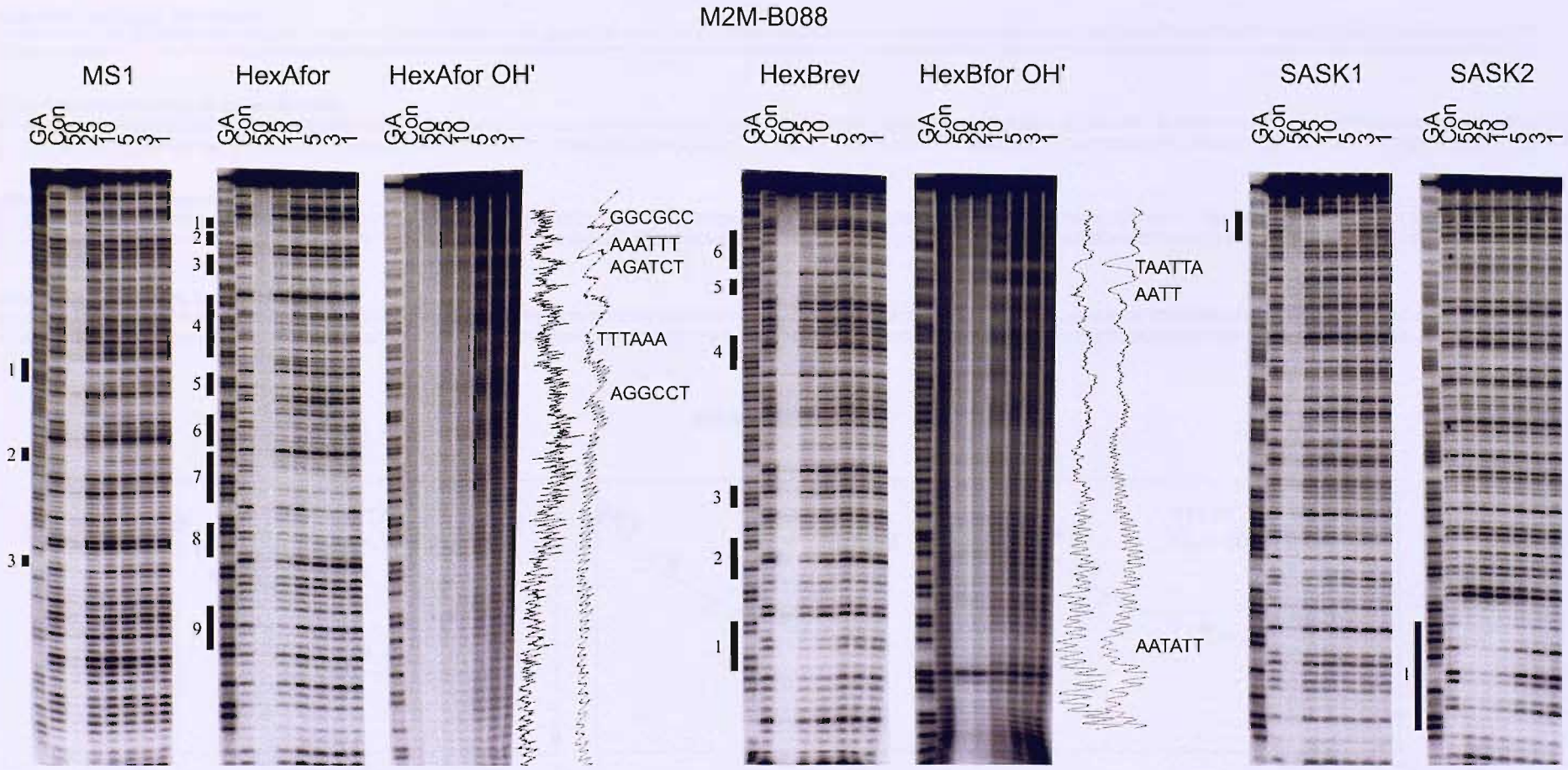


Figure 5.32: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B088 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.33; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 3 μM lanes). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B088: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

M2M-B088: HexA for (top); HexA rev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

M2M-B088: HexB for (top); HexB rev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

M2M-B088: SASK1 (top); SASK2 (bottom)

```
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'
```

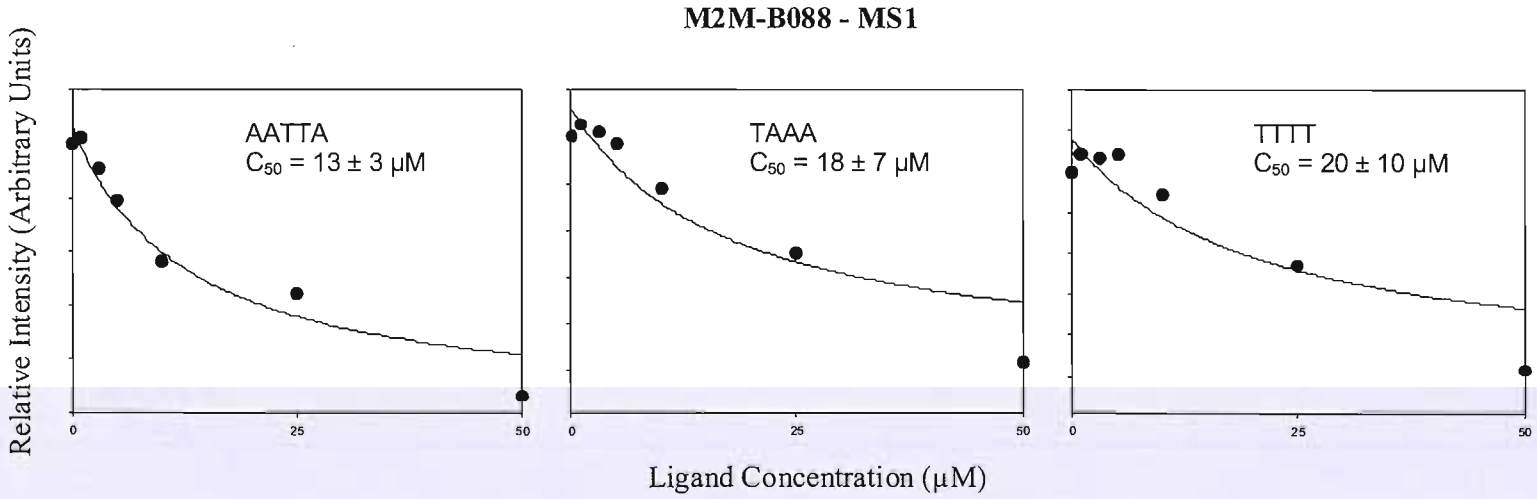


Figure 5.33: [Top] Binding sites for M2M-B088 on the tested fragments (from Figure 5.32). [Above] Example of footprinting plots for the interaction of M2M-B088 with binding sites in the MS1 DNA fragment.

	M2M-B088								
MS1	Site 1 AATTA 13 ± 3	Site 2 TAAA 18 ± 7	Site 3 TTTT 20 ± 10						
MS2	AAAC -	TAATT 11 ± 8	TTAA 30 ± 25	AGTGCA 26 ± 20	GGCC 23 ± 13	GCCGC 25 ± 18			
HexAfor	Site 1 GGCGCC 43 ± 21	Site 2 AAATTTA 73 ± 15	Site 3 AGATCT 27 ± 9	Site 4 TTTAAA,TTAA 26 ± 12	Site 5 AGGCCT 19 ± 12	Site 6 TGCGCA 12 ± 3	Site 7 AGCGCT 5 ± 2	Site 8 AGTACT 7 ± 2	Site 9 TGCACG 14 ± 8
HexArev2	CGTGCA 32 ± 19	AGTACT 30 ± 21	AGCGCT 21 ± 10	TGCGCA 26 ± 7	AGGCCT 19 ± 8	TTAA,TTTAAA 23 ± 10	AGATCT 20 ± 10	TAAATTT 13 ± 5	GGCGCC 19 ± 10
HexBfor	TAATTA 27 ± 10	AATT 16 ± 3	TTATAA 28 ± 17	TATATA 16 ± 8	GACGTC 19 ± 14	AATATT 12 ± 10	AATTAAT 9 ± 7		
HexBrev	Site 6 ATTAATT,AATATT 20 ± 5	Site 5 GACGTC 18 ± 9	Site 4 TATATA 22 ± 12	Site 3 TTATAA 15 ± 7	Site 2 AATT 15 ± 7	Site 1 TAATTA 6 ± 3			
SASK1	Site 1 AGCGCT 8 ± 4								
SASK2	Site 1 AGCGCT 6 ± 4								

Table 5.9: C_{50} values (μM) derived from quantitative analysis of the M2M-B088 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). Sites marked "–" correspond to footprints evident by visual inspection, but for which quantitative analysis did not provide a clear value.

M2M-B097

There are three possible configurations which might allow 2:1 binding for M2M-B097, two of which have benzimidazole paired with imidazole, as shown in Figure 5.34.

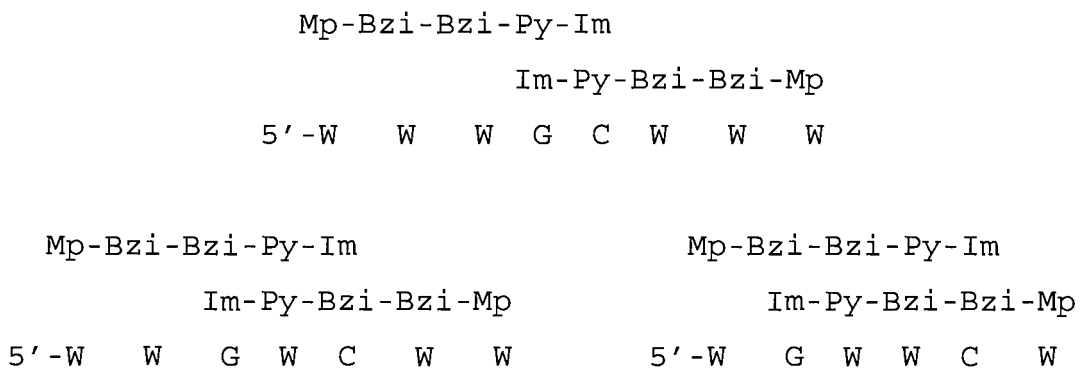


Figure 5.34: Possible DNA binding modes for M2M-B097.

Figure 5.35 shows the results of DNase I footprinting experiments with M2M-B097 on MS2, HexAfor, HexBfor, SASK1 and SASK2 fragments; and hydroxyl radical footprinting experiments on HexAfor and HexBfor fragments with densitometer plots. Figure 5.36 shows the sequences of the DNase I tested fragments with the observed footprinting binding sites highlighted. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.10.

MS1/MS2: It can be seen that M2M-B097 has bound to two GWC sites on MS1/MS2: CGACT (site 4, with TGCA) and TGACAAA (site 9). Some weaker binding is evident at GWWC (GGATCC (site 1, with ATAT) and AGAACG (site 6). However, footprints are not observed at all the GWWC and GWC sites, with GGTCT and TGTACG not showing cleavage protection.

Every $(A/T)_4$ tract is bound, indicative of 1:1 binding; such as TTAT (site 2), AAAT (site 3), TAATTGCT (site 7) and TTTA and ATAA (site 8). The binding to TTAA (site 5) is weaker.

HexA: Four M2M-B097 footprints are apparent on the HexA fragments, with the A/T-tracts AAATTTA (site 1) and AATT (site 2) the strongest bound. Other A/T-tracts such as ATATAT and TTTAAA are not bound, maybe because TpA steps hinder the binding. AGTACT (site 4) is a slightly weaker binding site, with GGTACC (site 3) weaker still.

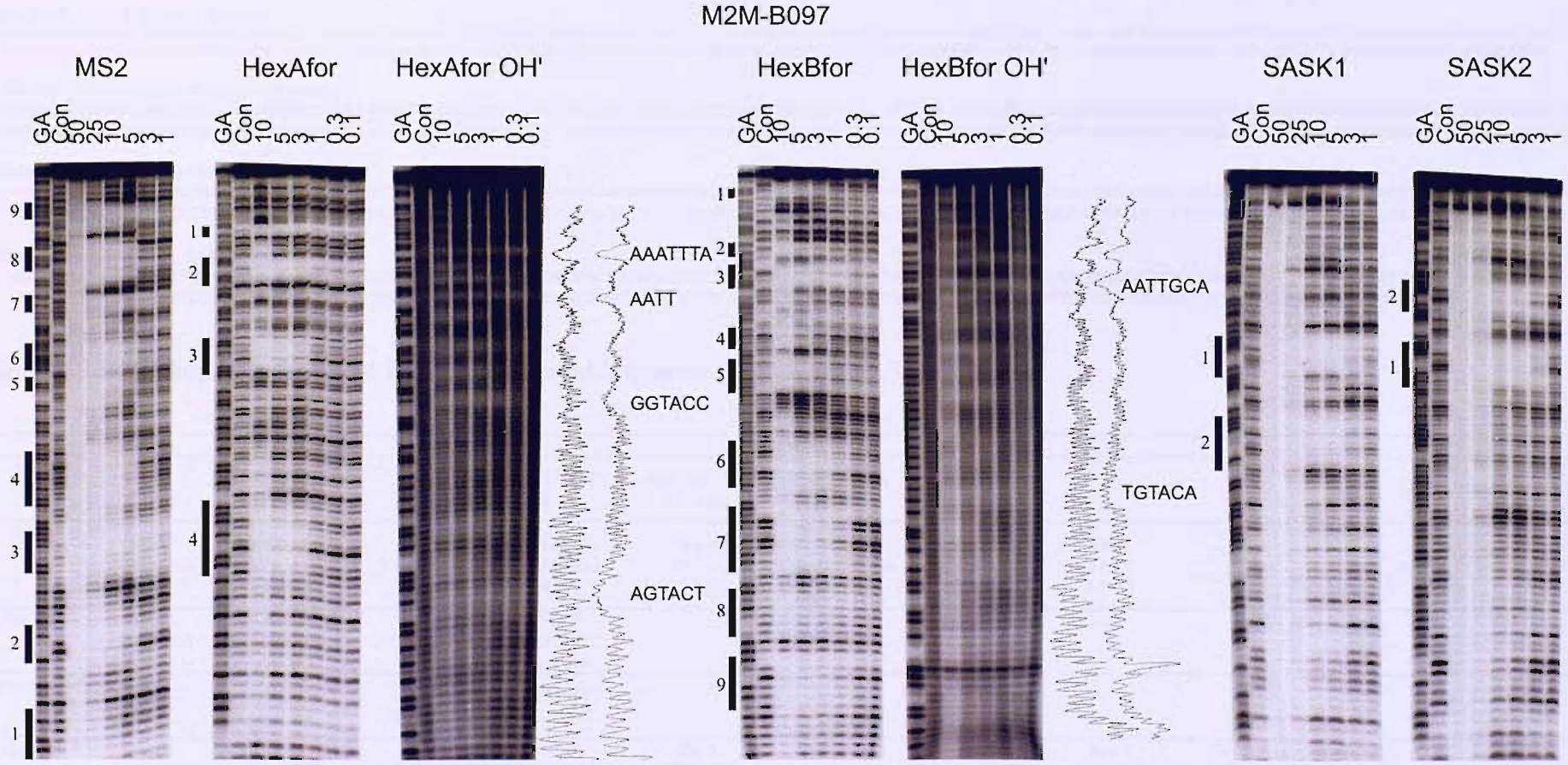


Figure 5.35: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B097 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.36; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 0.3 μM lanes). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B097: MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

M2M-B097: HexAfor (top); HexArev2 (bottom)

```
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'
```

M2M-B097: HexBfor (top); HexBrev (bottom)

```
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'
```

M2M-B097: SASK1 (top); SASK2 (bottom)

```
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'
```

Figure 5.36: Binding sites for M2M-B097 on the tested fragments (from Figure 5.35).

M2M-B097												
MS1	ATTT 72 ± 69	TGCA,AGTCG 5 ± 1	TTAA 23 ± 20	CGTTCT 6 ± 2	AGCAATTA 0.5 ± 0.2	TTAT,TAAA 3 ± 0.7	TTTTGTCA 1.3 ± 0.3			SASK1	Site 1 AGACCT 0.5 ± 0.1	Site 2 AGTCCT 0.9 ± 0.2
MS2	Site 9 TGACAAA 0.3 ± 0.15	Site 8 TTTA,ATAA 1 ± 0.5	Site 7 TAATTGCT 0.4 ± 0.1	Site 6 AGAACG 4 ± 2	Site 5 TTAA 5 ± 2	Site 4 CGACT,TGCA 2 ± 1	Site 3 AAAT 2 ± 0.5	Site 2 TATT 2.5 ± 1	Site 1 ATAT, GGATCC 11 ± 6	SASK2	Site 2 AGGACT 1 ± 0.3	Site 1 AGGTCT 0.7 ± 0.2
HexAfor	Site 1 AAATTTA 0.3 ± 0.15	Site 2 AATT 1 ± 0.3	Site 3 GGTACC 3 ± 2	Site 4 AGTACT 1 ± 0.4								
HexArev2	AGTACT 0.5 ± 0.2	GGTACC 2 ± 0.7	AATT 0.1 ± 0.07	TAAATTT 0.2 ± 0.2								
HexBfor	Site 1 GGATCC 8 ± 4	Site 2 TAATTA 4 ± 1.6	Site 3 AATTGCA 0.07 ± 0.04	Site 4 TACGTATA 4 ± 1.7	Site 5 CGTACG 6 ± 3.5	Site 6 TGTACA 1 ± 0.4	Site 7 CGACGTCA 0.5 ± 0.2	Site 8 AATATT 1.8 ± 0.5	Site 9 AATTAAT 0.4 ± 0.1			
HexBrev	ATTAATT 6 ± 2	TGACGTCG 5 ± 3	TGTACA 5 ± 3	CGTACG 7 ± 4	TATACGTA 6 ± 4	TGCAATT 2 ± 0.5	TAATTA 2 ± 1					

Table 5.10: C_{50} values (μM) derived from quantitative analysis of the M2M-B097 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

It can therefore be suggested that A/T bases surrounding the GWWC core are preferred over flanking G/C bases. This could explain the lack of binding to GGACC, even though it contains the preferred GWC sequence. There is no evidence of binding to GC sequences.

Hydroxyl radical footprinting experiments show attenuated cleavage at AAATTTA, AATT, GGTACC and AGTACT, which correspond to the best DNase I binding sites.

HexB: Many sequences on the HexB fragment appear to have been bound with similar affinity by M2M-B097 and there is one long footprint at 10 μM and above. However, protection at the following sites persists to lower concentrations: AATTGCA (site 3), for which binding probably occurs at the A/T-tract, as other GC sites are not bound; TGTACA (site 6), which contains the sequence GWWC; and CGACGTCA (site 7), containing the sequence GWC. Other strong sites are at A/T-tracts, including TAATTA (site 2), TACGTATA (site 4), AATATT (site 8) and AATTAAT (site 9). Weaker binding sites are GGATCC (site 1) and CGTACG (site 5), both containing the GWWC sequence, whilst CGATCG does not show protection.

Hydroxyl radical footprinting experiments show attenuated cleavage at AATTGCA and TGTACA. The former can be explained by suggesting 1:1 binding, whilst TGTACA contains the GWWC sequence proposed in Figure 5.34 for the preferred 2:1 binding mode.

SASK1/SASK2: Only two footprints are evident on the SASK1/SASK2 fragments: AGACCT (site 1) is marginally stronger than AGTCCT (site 2). It is interesting to note that AGATCT is not bound by M2M-B097 (contrary to the findings with the similar ligand M2M-B071). No GC sequences are bound and the affinity of binding for M2M-B097 is generally stronger than M2M-B071.

SASK1/SASK2 Hydroxyl radical footprinting:
M2M-B071 and M2M-B097

Figure 5.37 shows the results of hydroxyl radical footprinting experiments with M2M-B071 and M2M-B097 on the SASK1 fragment. It can be seen these two ligands show the same hydroxyl radical footprinting patterns with attenuated cleavage at AAGACCTT (site 1), AAGTCCTT (site 2) and AAGATCTT (site 3), suggestive of both GWC and GWWC binding. This selectivity is clearest on the differential cleavage plot and is also seen on SASK2 (not shown).

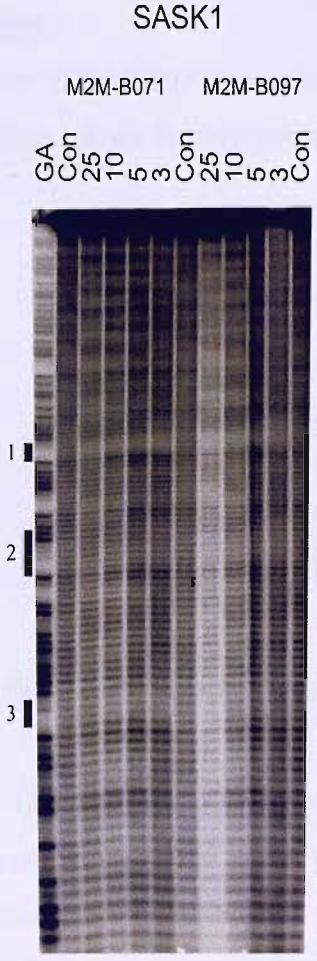
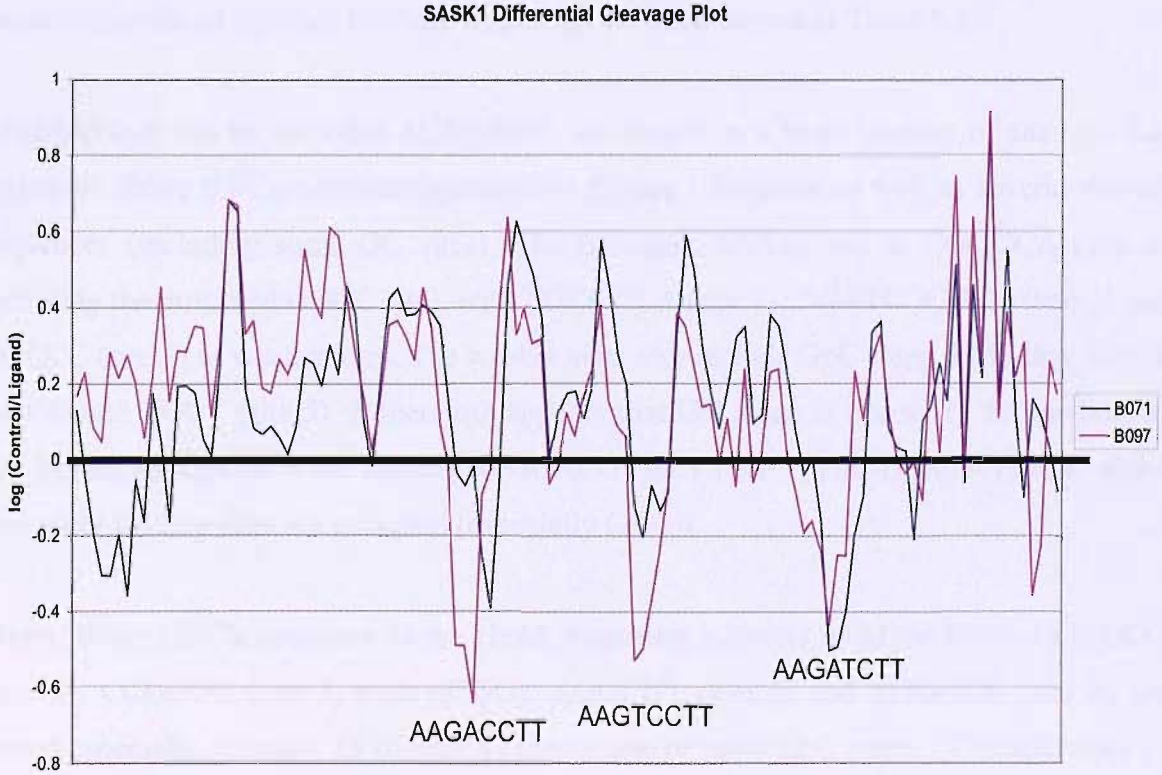


Figure 5.37: [Left] Hydroxyl radical footprinting gel of M2M-B071 and M2M-B097 with the SASK1 fragment. Only the clearest footprints are marked and are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane.

[Below] Differential cleavage plot of M2M-B071 and M2M-B097 binding with the SASK1 DNA fragment (5'-3') derived from the 3 μM lanes of the hydroxyl radical gel.



M2M-B099

Compound M2M-B099 is similar to M2M-B097, but with an extra N-terminal imidazole moiety. This allows for two potential binding conformations, as shown in Figure 5.38.

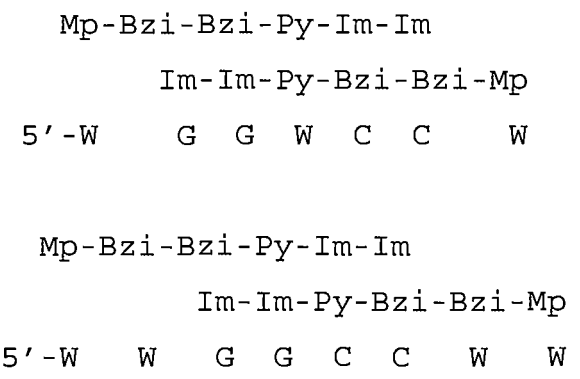


Figure 5.38: Possible DNA binding modes of M2M-B099.

Figure 5.39 shows the results of DNase I footprinting experiments with M2M-B099 on MS2, HexAfor, HexBfor, SASK1 and SASK2, together with hydroxyl radical footprinting experiments on HexAfor and HexBfor. Figure 5.40 shows the sequences of these fragments with the observed footprinting binding sites highlighted. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.11.

MS1/MS2: It can be seen that M2M-B099 has bound to a large number of sites on this fragment. Every $(G/C)_4$ sequence generates a DNase I footprint as well as several shorter sequences (including some GC sites). The strongest binding site is CGGCCA (site 2, including the proposed GGCC site), with TGCCGCA (site 1), TGATCGCGCT (site 3) and CGCCC (site 6) as weaker sites. The weaker sites also include GpC steps, including AGCT (site 4) and TGCT (site 5). It therefore appears that GC alone is necessary for binding of this ligand, though the weak binding to GGACGTACT (site 7) and TGACA (site 8) show that other binding sites are occupied (especially GWC).

HexA: Every $(G/C)_4$ sequence on the HexA fragments is bound by M2M-B099. GGCGCC (site 2), CCGCGG (site 3, with CCGG), AGGCCT (site 5) and TGGCCA (site 8) are bound especially strongly, all of which contain one or more GpC steps. CCCGGG (site 1) and ACCGGT (site 4) are weaker binding sites, indicating that G/C-tracts without a GpC step can also be bound. Weak binding is observed at TGCGCA (site 6) and AGCGCT (site

7), indicating that the central base sequence of site 2 (GCGC) is not sufficient to produce strong binding alone.

Hydroxyl radical footprinting experiments confirm these observations, with attenuated cleavage observed at CCCGGG, GGCGCC, TCCGGA, CCGCGG, ACCGGT, AGGCCT and TGCGCA. It is not clear whether AGCGCT and TGGCCA are protected or not.

HexB: M2M-B099 has bound every $(G/C)_4$ sequence on the HexB fragments. GGGCCC (site 2), GCCGGC (site 3) and GCGCGC (site 4) are bound strongly, supporting the theory that GC is necessary for strong binding. However, footprints are also observed at GTCGACGTC (site 6) and the A/T-tracts AATATT (site 7) and AATTAAT (site 8, with TGCA), suggesting that 1:1 binding may also be occurring.

It is interesting to note that CGGCCG (site 1) is a weaker binding site than GGGCCC (site 2), even though they contain the same central tetranucleotide sequence, indicating that all six base pairs are involved in binding. CGCGCG (site 5, with ACGCGT) is also a weaker binding site, suggesting that three adjacent GpC steps (in GCGCGC) are preferred over two steps.

Hydroxyl radical footprinting experiments of the HexB fragments support these DNase I studies, with attenuated cleavage observed at GGGCCC, GCCGGC, GCGCGC, ACGCGT, CGCGCG and AATATT. The DNase I footprint at GTCGACGTC (site 6) has been resolved into two separate binding sites, with two GTC sites binding. Binding to CGGCCG and AATTAAT is not clear as these sites are at the two extremes of the gel.

SASK1/SASK2: It can be seen that every GpC step has been bound by M2M-B099, except when these are located in alternating G/C-tracts (GCGC and CGCG). GGCC sites TAGGCCAT (site 1, with TGCT), ATGGCCAT (site 3, with TGCA), AAGGCCTT (site 5, with TGCA) and TTGGCCAA (site 7, with TGCT) are bound with similar affinity to sites with only a lone GC (TTGCAA (site 2), ATGCAA (site 4), TTGCAA (site 6) and TTGCAA (site 8)) or GCC (TTGCCA (site 9), indicating only a single GpC is necessary for strong binding. This suggests that in the GGCC binding mode in Figure 5.38, only the central Py-Im is affecting the sequence selectivity.

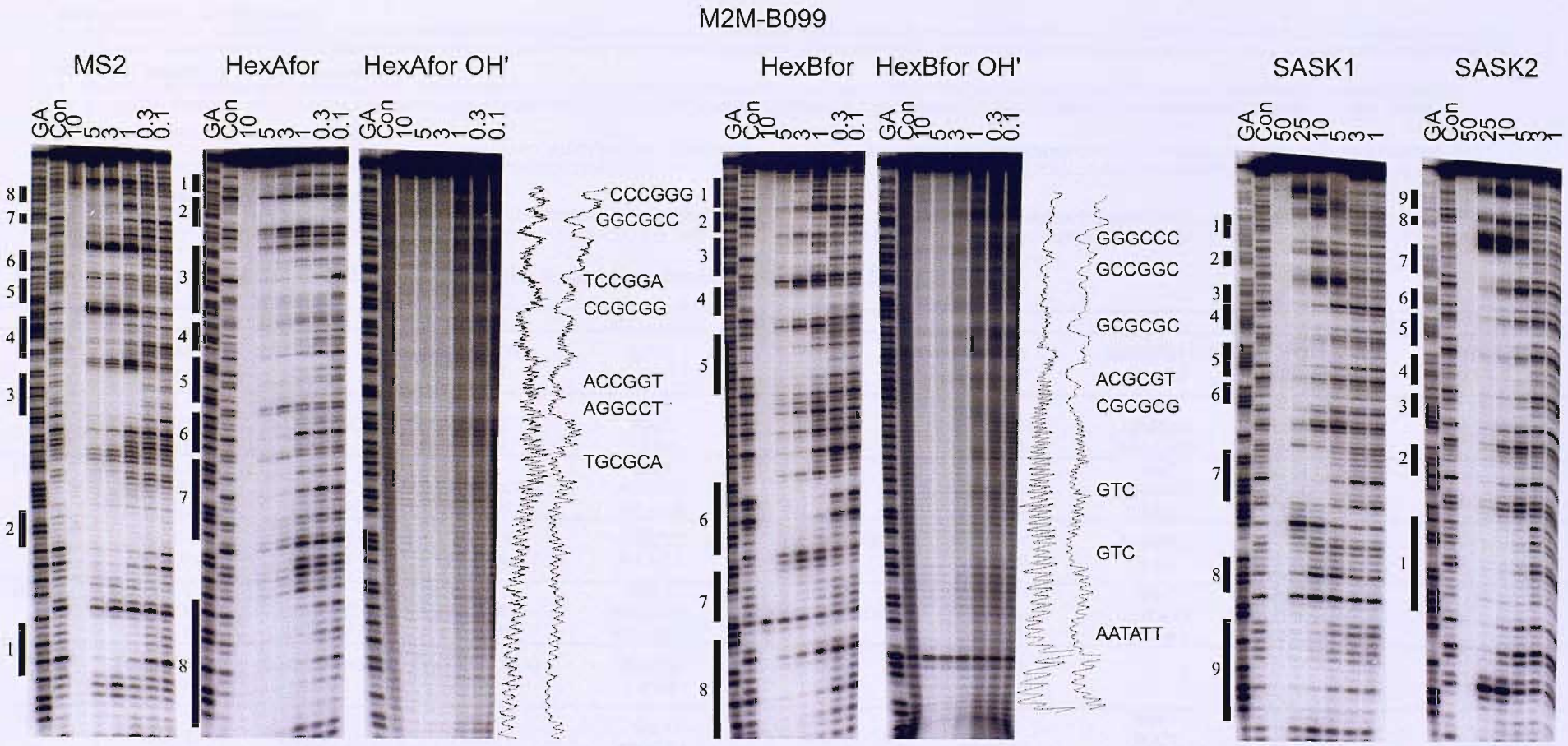


Figure 5.39: DNase I and hydroxyl radical footprinting gels showing the interaction of M2M-B099 with the tested fragments. The DNase I footprints are indicated by the filled boxes and are also highlighted on the sequences in Figure 5.40; while the hydroxyl radical footprint sequences are shown (densitometer plots are taken from the 0.3 μM lanes). Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-B099: MS1 (top); MS2 (bottom)
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

M2M-B099: HexAfor (top); HexArev2 (bottom)
5'-GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTTTAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGGCCATGGATCC-3'
3'-CCTAGGGCCCTATAGCTATATACCGCGGTTTAAATCGATATCTAGATCTTAAGGCCTGGCGCCAAATTTGCAATTGGCCATGGATCCGGACGTCGACGCGTACGATCGCGAATTCATGATCACGTGCACCGGTACCTAGG-5'

M2M-B099: HexBfor (top); HexBrev (bottom)
5'-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGCGCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAATTAATGCATGGATCC-3'
3'-CCTAGGCCGGCTAGCGCTCGAGCTCCCGGGATTAATCGGCCGTTAACGTTCGAATATTCGCGCGATGCATATGCGCATGCGCGCATATATGTATACATGTACAGCTGCAGTACTAGTTATAAGCTTAATTACGTACCTAGG-5'

M2M-B099: SASK1 (top); SASK2 (bottom)
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'

Figure 5.40: Binding sites for M2M-B099 on the tested fragments (from Figure 5.39).

	M2M-B099								
MS1	TGCGGCA 30 ± 29	GGCA 54 ± 51	AGCGCGATCA 0.2 ± 0.1	AGCT 1 ± 0.3	AGCA 1.4 ± 0.7	GGGCG 0.2 ± 0.07	AGTACGTCC 0.7 ± 0.3	TGTCA 0.4 ± 0.1	TGCGG 0.4 ± 0.1
MS2	Site 8 TGACA 7 ± 4	Site 7 GGACGTACT 1 ± 0.5	Site 6 CGCCC 0.5 ± 0.2	Site 5 TGCT 1 ± 0.5	Site 4 AGCT 2 ± 0.6	Site 3 TGATCGCGCT 0.3 ± 0.1	Site 2 CGGCCA 0.04 ± 0.01	Site 1 TGCCGCA 0.3 ± 0.07	
HexAfor	Site 1 CCCGGG 2 ± 0.4	Site 2 GGCGCC 0.06 ± 0.04	Site 3 CCGG,CCGCGG 0.1 ± 0.05	Site 4 ACCGGT 0.6 ± 0.3	Site 5 AGGCCT 0.1 ± 0.03	Site 6 TGCGCA 1.6 ± 0.8	Site 7 AGCGCT 1.3 ± 0.5	Site 8 TGGCCA 0.4 ± 0.1	
HexArev2	TGGCCA 1.7 ± 0.4	AGCGCT 7.2 ± 6.9	TGCGCA 5.2 ± 4.8	AGGCCT 0.5 ± 0.1	CCGCGG,CCGG 0.4 ± 0.1	GGCGCC 0.1 ± 0.01	GGATCC 4 ± 2		
HexBfor	Site 1 CGGCCG 6 ± 4	Site 2 GGGCCC 0.4 ± 0.07	Site 3 GCCGGC 0.3 ± 0.07	Site 4 GCGCGC 0.3 ± 0.05	Site 5 CGCG,CGCGCG 1 ± 0.4	Site 6 GTCGACGTC 0.2 ± 0.03	Site 7 AATATT 0.7 ± 0.1	Site 8 AATTAAT,TGCA 0.2 ± 0.04	
HexBrev	TGCATTAATT 3 ± 2	GACGTCGAC 1.4 ± 0.7	CGCGCG,CGCG 2 ± 1.5	GCGCGC 1 ± 0.8	GCCGGC 1 ± 0.9	GGGCCC 0.8 ± 0.5			
SASK1	Site 1 TGCT, TAGGCCAT 8 ± 2	Site 2 TTGCAA 14 ± 5	Site 3 TGCA, ATGGCCAT 12 ± 5	Site 4 ATGCAA 5 ± 2	Site 5 TGCA, AAGGCCTT 6 ± 2	Site 6 TTGCAA 6 ± 2	Site 7 TGCT, TTGGCCAA 5 ± 3	Site 8 TTGCAA 65 ± 51	Site 9 TTGCCA 7 ± 3
SASK2	Site 9 TGGCAA 5 ± 2	Site 8 TTGCAA 64 ± 60	Site 7 TTGGCCA,AGCA 7 ± 2	Site 6 TGCA,AAGGCCTT 6 ± 2	Site 5 TGCA,AGGTCT 3 ± 1	Site 4 TGCA,ATGGCCAT 3 ± 1	Site 3 ATGCAA 3 ± 1	Site 2 TTGCAA 7 ± 3	Site 1 ATGGCCTA,AGCA 3 ± 1.5

Table 5.11: C_{50} values (μM) derived from quantitative analysis of the M2M-B099 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

SASK1/SASK2 Hydroxyl radical footprinting:

M2M-B079, M2M-B084, M2M-B088 and M2M-B099

Figure 5.41 shows the results of hydroxyl radical footprinting experiments with M2M-B079, M2M-B084, M2M-B088 and M2M-B099 on the SASK1 fragment.

It can be seen that both M2M-B079 and M2M-B084 have bound to the GGCC sites (TAGGCCAT, ATGGCCAT, AAGGCCTT and TTGGCCAA) and are more selective than the related ligands M2M-B072 and M2M-B074 (shown in Figure 5.23). M2M-B099 has also bound these four GGCC sequences, but also produces attenuated cleavage at three sites that contain single GpC steps (TGCA, TGCA and TGCC), supporting previous findings that only GC is necessary for binding (and not GGCC).

M2M-B088 has only clearly bound the proposed binding site of AAGCGCTT selectively on the SASK1/SASK2 fragment, when observed by hydroxyl radical footprinting.

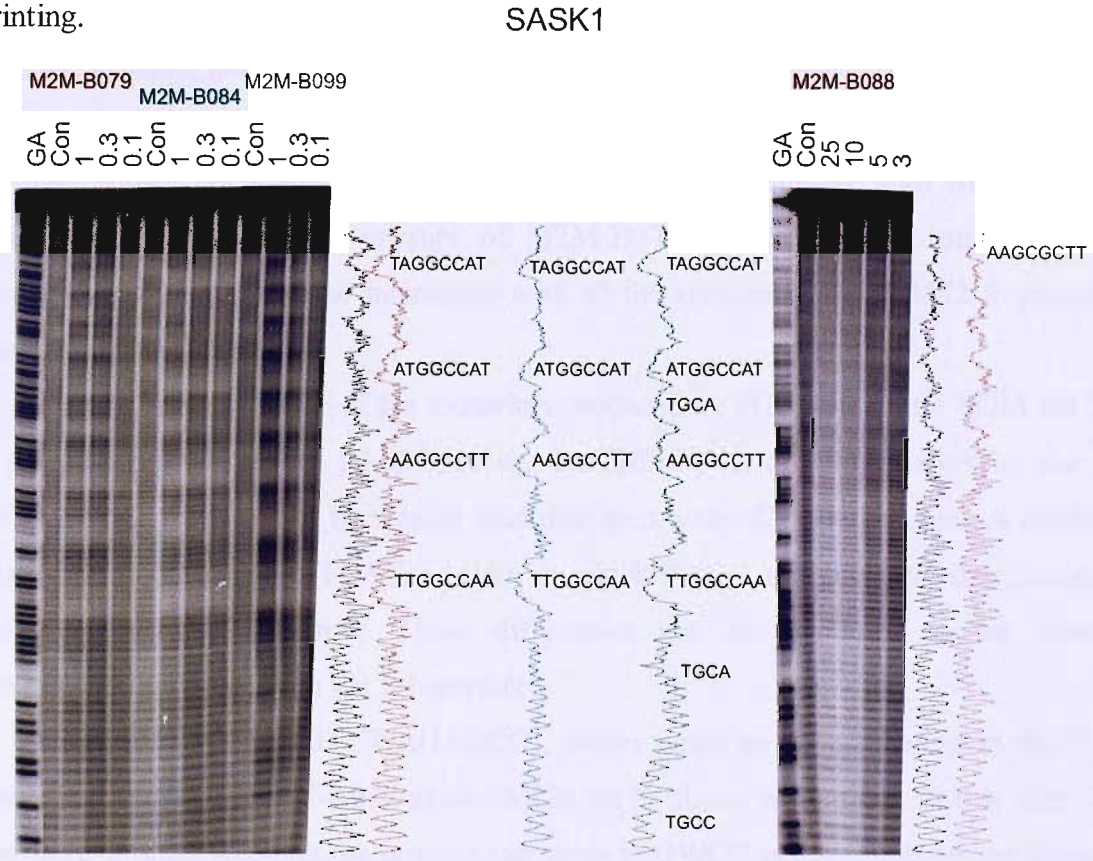


Figure 5.41: Hydroxyl radical footprinting gels of M2M-B079, M2M-B084 and M2M-B099; and M2M-B088 with SASK1. Only the clearest footprints are marked. Densitometer plots were derived from the 0.1 μM lane for M2M-B079, M2M-B084 and M2M-B099; and from the 3 μM lane for M2M-B088 and are colour-coded with the ligand names above the gel (the control densitometer plots are black). GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane.

1:1 mix heterodimer (M2M-B071 and M2M-B072)

The Series B ligands are designed to bind to DNA in a 2:1 conformation as homodimers. However, it is possible to mix some of the ligands to create viable 2:1 heterodimers to select for different sequences.

The mixing of M2M-B071 (Mp-Bzi-Py-Py-Im) and M2M-B072 (Mp-Bzi-Py-Im-Im) in a 1:1 fashion can give the potential for two different binding configurations to occur for a heterodimer binding to duplex DNA, assuming normal polyamide binding rules are applied. These are shown in Figure 5.42.

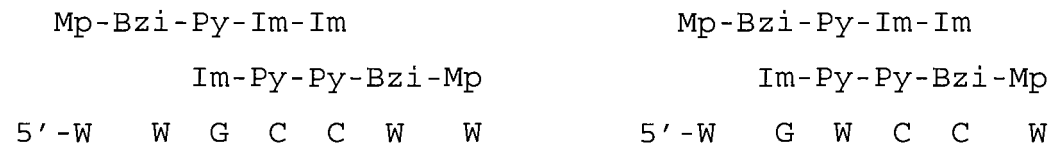


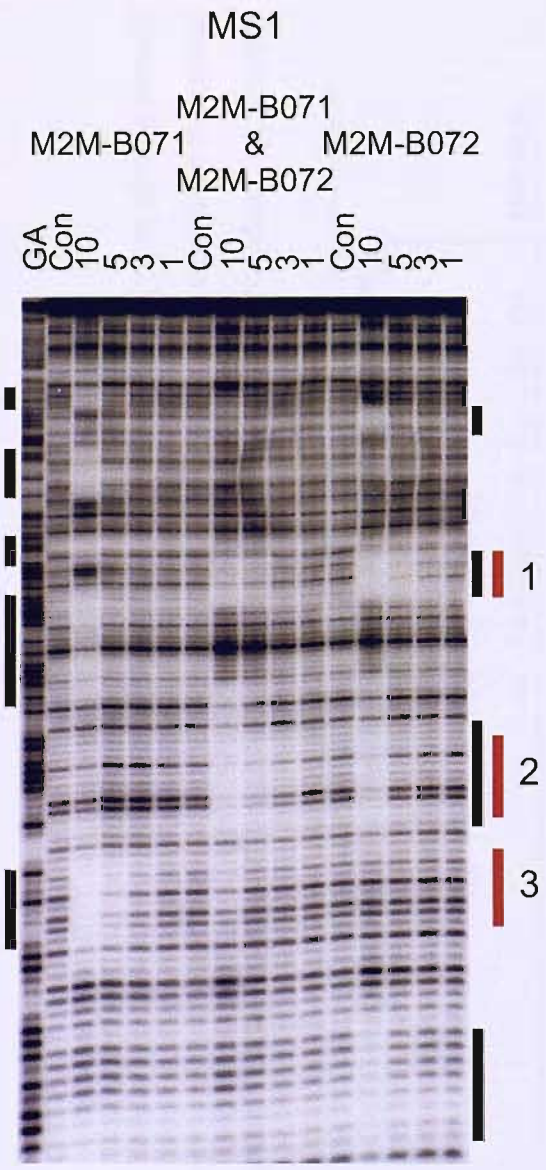
Figure 5.42: Possible DNA binding modes for a M2M-B071/M2M-B072 heterodimer.

Figure 5.43 shows the results of DNase I footprinting experiments with M2M-B071, M2M-B072 and an equimolar mixture of M2M-B071 and M2M-B072 on the MS1 fragment. The C_{50} values for the interaction with all the sites on the MS1/MS2 fragments are summarised in Table 5.12.

It can be seen that three of the footprints produced by M2M-B071 and M2M-B072 alone are also present in the 1:1 mix lanes. The affinity of the 1:1 mixture at site 1 (GGGCG/CGCCC) appears to be weaker than that seen with M2M-B072 alone. A similar situation is seen at site 3 (TGTCA/TGACA), in which the 1:1 mix has a weaker binding affinity than M2M-B071 alone. These differences can be attributed to the lower concentration of each ligand in the 1:1 mixture.

However, site 2 (GGGTCT/AGACCC), shows equal or better binding in the 1:1 mixture compared with M2M-B072 alone (M2M-B071 shows no binding at this site). It can therefore be suggested that the mixture can binds to GWCC as a heterodimer, as shown in Figure 5.42 above.

Other combinations of ligands: M2M-B072/M2M-B073; M2M-B072/M2M-B097; and M2M-B073/M2M-B097 exhibited no evidence of heterodimer formation on MS1 and MS2 fragments.



		Site 1	Site 2	Site 3
M2M-B071	MS1		GGGTCT	TGTCA
			-	2.5 ± 1
	MS2		AGACCC	Unclear
			-	
1:1 Mix	MS1	GGGCG	GGGTCT	TGTCA
		4 ± 1	1.7 ± 0.3	10 ± 6
	MS2	CGCCC	AGACCC	
		12 ± 5	8 ± 4	
M2M-B072	MS1	GGGCG	GGGTCT	
		1 ± 0.4	6 ± 3	
	MS2	CGCCC	AGACCC	
		1.6 ± 0.5	8 ± 2	

Figure 5.43 (Left): DNase I footprinting gel showing the interaction of M2M-B071, M2M-B072 and a 1:1 mixture with the MS1 fragment. The footprints are indicated by the filled boxes, with the red boxes representing binding of the 1:1 mixture. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

Table 5.12 (Above): C_{50} values (μM) derived from quantitative analysis of the gel in Figure 5.39 (with the standard errors shown).

Hoechst 33258 / Polyamide Conjugates: Series D

Using known pairing rules (and assuming that benzimidazole will not bind opposite another benzimidazole, as with Hoechst 33258), Table 5.13 shows the possible binding modes for the Series D ligands that showed cleavage protection on the tested fragments.

M2S-D11	Dp-Py-Py-Im-Bzi-Im Im-Bzi-Im-Py-Py-Dp 5′-W W G C G C W W
M2S-D22	Mp-Bzi-Py-Py-BzN-Im Im-BzN-Py-Py-Bzi-Mp 5′-W W G G C C W W
	Mp-Bzi-Py-Py-BzN-Im Im-BzN-Py-Py-Bzi-Mp 5′-W G G W C C W
	BzN-Im is linked by a six-membered ring and so is sterically unable to bind as a dimer with itself.
M2S-D33	Dp-Py-Py-Py-Bzi-Im Im-Bzi-Py-Py-Py-Dp 5′-W G W W W C W
	Dp-Py-Py-Py-Bzi-Im Im-Bzi-Py-Py-Py-Dp 5′-W W G W W C W W
M2S-D66	Mp-Bzi-Py-Im-Gly-Bzi-Py-Im Im-Py-Bzi-Gly-Im-Py-Bzi-Mp 5′-W G W C W G W C W
M2S-D88	Mp-Bzi-Py-Bzi-Py-Bzi-Im Im-Bzi-Py-Bzi-Py-Bzi-Mp 5′-W W W G W W C W W W
	The Series B ligand results show a six ring overlap is unlikely to occur.

Table 5.13: Proposed DNA binding sites for the Series D ligands. The benzimidazole, presenting an unbound nitrogen to the minor groove in M2S-D22 is termed "BzN". The glycine linkage in M2S-D66 is shown as "Gly".

The Series D set of compounds was initially examined by DNase I footprinting with MS1 and MS2 to establish approximate binding selectivity and affinity. DNase I footprinting experiments were then carried out on the SASK1/SASK2 fragments, since the initial observations suggested that the ligands were G/C selective, making SASK1/SASK2 an ideal substrate.

MS1/MS2

Figure 5.44 shows the results of DNase I footprinting experiments with M2S-D11, M2S-D22 and M2S-D66 on the MS1 fragment; and M2S-D33 and M2S-D88 on the MS2 fragment. Figure 5.45 shows the sequence of MS1/MS2 with the observed footprinting binding sites highlighted, together with an example of footprinting plots derived from the binding of M2S-D11 to MS1. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.14.

M2S-D11: It can be seen that the interaction between the MS1/MS2 fragment and M2S-D11 is generally weak, with only three A/T-tracts showing DNase I footprints (suggestive of 1:1 binding). The long site TATAGCAATTA (site 1) shows the strongest affinity, whilst the shorter sites 2 (TAAA) and 3 (TTTT) are weaker. There is no evidence for the interaction with any of the proposed G/C-containing sites.

M2S-D22: DNase I footprints are evident at every $(A/T)_4$ site on this fragment (suggestive of 1:1 binding); AATA (site 1), ATTT (site 2), TTAA (site 3), TATA (site 5), AATTA (site 6), TTAT (site 7), TAAA (site 8) and TTTT (site 9). However, the footprint at site 4 (CGTTCTAT) may indicate a different binding mode (though other GWWC sites are not bound).

M2S-D33: This is a generally weaker binder than the other Series D ligands and M2S-D33 produces DNase I footprints at every A/T-tract on the MS1/MS2 fragment; TTAT (site 1), TAAA (site 2), TTAA (site 3), TAATTGCTATA (site 5) and TTTA and ATAA (site 6) binding with similar affinity. Interestingly, ATAGAACG (site 4) is bound by the ligand, supporting the GWWC binding mode although no other GWWC sequences show protection. This binding is therefore most likely due to an interaction with the short A/T-tract (supported by weak binding to ATGACAAA (site 7).

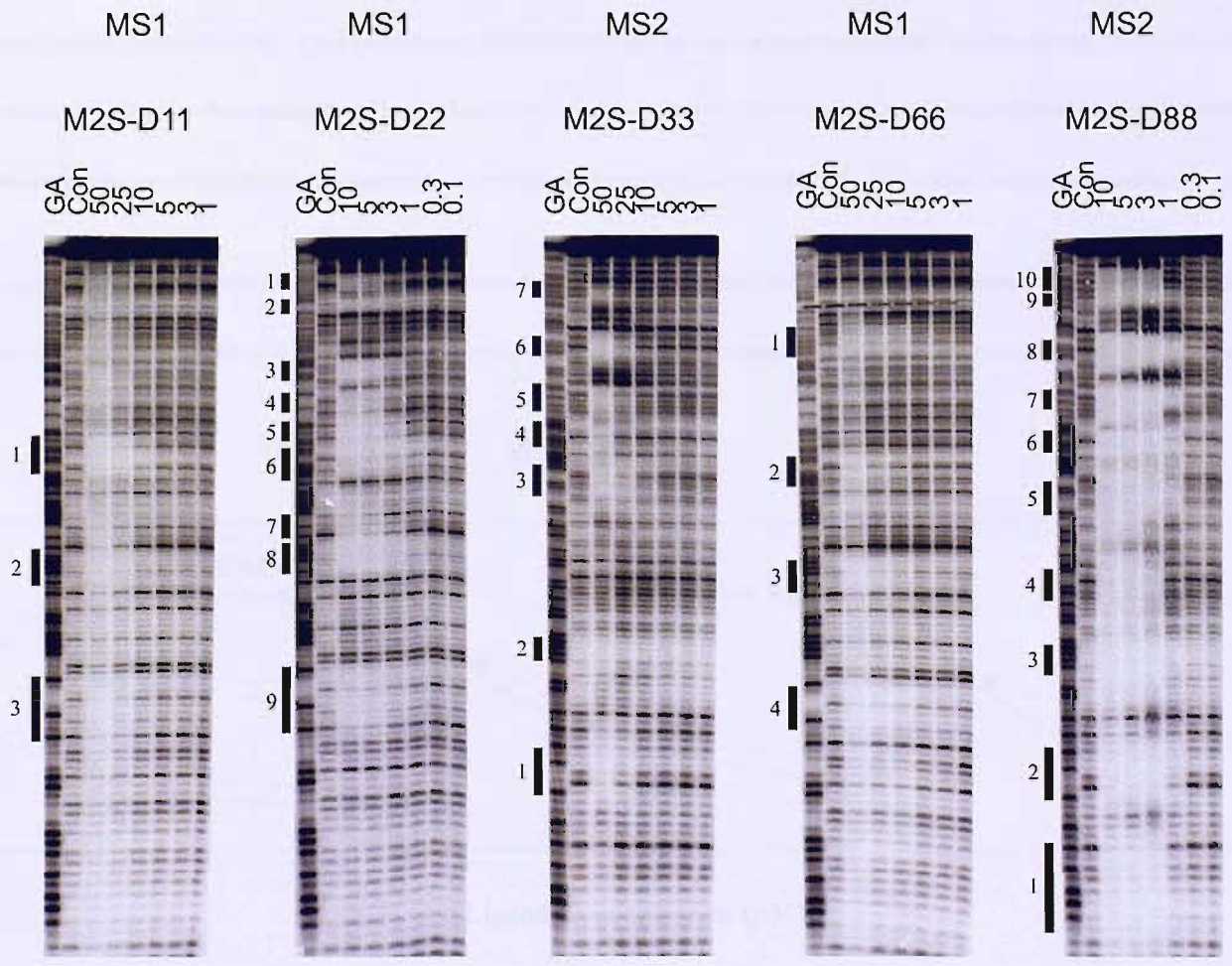


Figure 5.44: DNase I footprinting gels for the interaction of the Series D ligands with MS1 and MS2 fragments. The clearest footprints are indicated by the filled boxes and are highlighted in the sequence shown in Figure 5.45. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2S-D11: MS1
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'

M2S-D22: MS1
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'

M2S-D66: MS1
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'

M2S-D33: MS2
5'-GGATCCGCATTCGAGGCTGAGATGACAAACCAGACCCCACCGGACGTACTTTACATAACTCTTCACGCCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCTTGATCGCGCTACGACTAGTGCAGTTGGAAATCGGCCATGTGTATTGCCGCATATGGATCC-3'

M2S-D88: MS2
5'-GGATCCGCATTCGAGGCTGAGATGACAAACCAGACCCCACCGGACGTACTTTACATAACTCTTCACGCCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCTTGATCGCGCTACGACTAGTGCAGTTGGAAATCGGCCATGTGTATTGCCGCATATGGATCC-3'

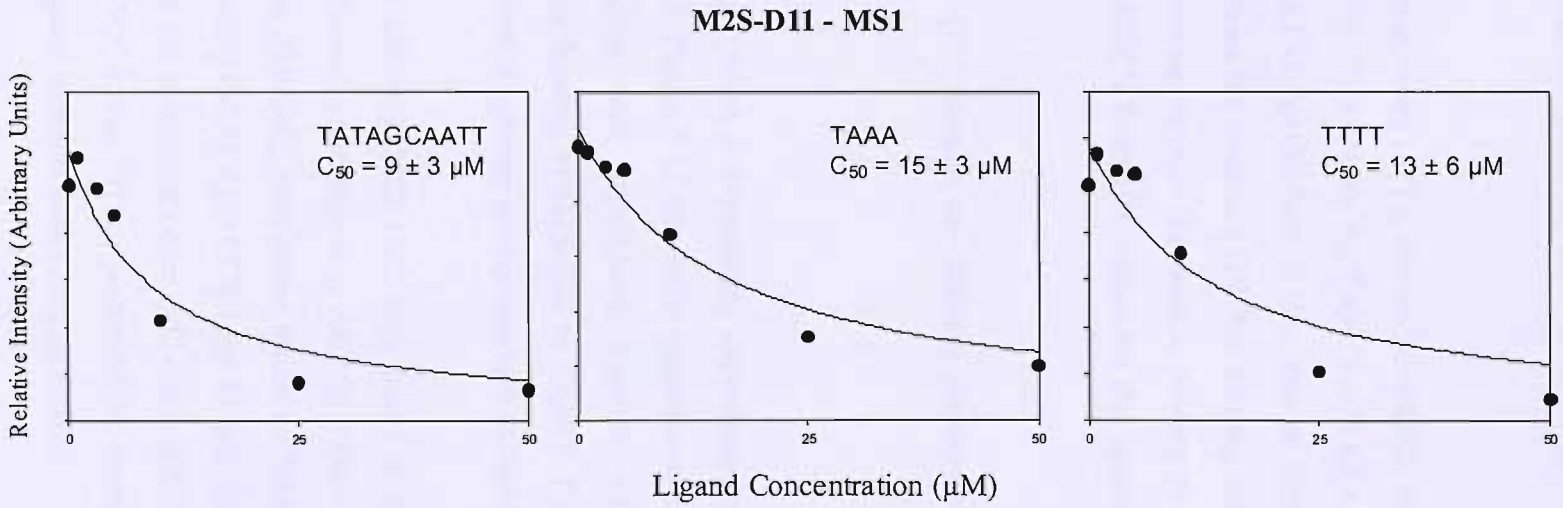


Figure 5.45: [Top] DNase I footprinting sites (in red) for the Series D ligands on the MS1/MS2 fragments. Only the sequence of the insert, which was cloned into the BamHI site of pUC18, is shown. [Above] Examples of footprinting plots for the interaction of M2S-D11 with the MS1 fragment.

M2S-D66: It can be seen that M2S-D66 produces the best footprints at AGCGCGA (site 1) and TTTT (site 4) on the MS1/MS2 fragment. AATTA (site 2) and TAAA (site 3) are the only weaker binding sites, with not every A/T-tract bound. It is also interesting that there are no footprints at G/C-tracts, suggesting that the proposed binding to GWCWGWC does not occur.

M2S-D88: Footprints are evident at every $(A/T)_4$ site on MS1/MS2, including ATAT (site 1), TAAT (site 2), AAAT (site 3), TTAA (site 5), TAATTGCTATA (site 7) and TTTA and ATAA (site 8). Binding to ATAGAACG (site 6) may also be attributable to 1:1 A/T-tract binding, though it also contains the proposed GWWC binding site (no other GWWC site is bound though). It is interesting to note the weaker binding to CGACTA (site 4), ATGACAAA (site 9) and CGAGGCT (site 10), indicating the ligand may bind to other sequences than just A/T-tracts.

M2S-D44, M2S-D55 and M2S-D77 showed no cleavage protection on the MS1/MS2 fragment.

SASK1/SASK2

Figure 5.46 shows the results of DNase I footprinting experiments with M2S-D22 and M2S-D88 on SASK1 and SASK2. Figure 5.47 shows the sequence of SASK1/SASK2 with the observed footprinting binding sites highlighted, together with an example of footprinting plots derived from the binding of M2S-D22 to SASK2. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.15.

M2S-D22: It can be seen that although M2S-D22 only binds at high concentrations, producing weak footprints, the clearest protection is at AGATCT (site 4), again indicating a different binding conformation (GWWC) than those stated in Table 5.13 (GGCC and GGWCC). AGGCCA (site 1, with TGCA) AGACCT (site 2) and AGTCCT (site 3) are even weaker sites. No footprints are evident at other G/C-tracts (GCGC, GCCC, CGCG, CCGG and GGCC (except AGGCCA; site 1)) nor isolated GpC steps. There is therefore only limited evidence that this ligand interacts with GWC or GGCC.

M2S-D88: Only four sequences on SASK1/SASK2 show specific cleavage protection by M2S-D88 (non-specific binding is seen down to 5 μM). These sites; AGCGCT and

				M2S-D11					
MS1	Site 1	Site 2	Site 3						
	TATAGCAATTA	TAAA	TTTT						
	9 ± 3	15 ± 3	13 ± 6						
				M2S-D22					
MS1	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
	AATA	ATTT	TTAA	CGTTCTAT	TATA	AATTA	TTAT	TAAA	TTTT
	4 ± 1	0.4 ± 0.1	2.6 ± 0.5	2.5 ± 0.8	1.3 ± 0.4	0.5 ± 0.1	1.3 ± 0.2	1 ± 0.3	3 ± 0.3
				M2S-D33					
MS2	Site 7	Site 6	Site 5	Site 4	Site 3	Site 2	Site 1		
	ATGACAAA	TTTA,ATAA	TAATTGCTATA	ATAGAACG	TTAA	AAAT	TATT		
	67 ± 52	20 ± 11	33 ± 18	37 ± 35	20 ± 9	35 ± 22	23 ± 15		
				M2S-D66					
MS1	Site 1	Site 2	Site 3	Site 4					
	AGCGCGA	AATTA	TAAA	TTTT					
	8 ± 4	24 ± 7	20 ± 5	6 ± 1					
				M2S-D88					
MS2	Site 10	Site 9	Site 8	Site 7	Site 6	Site 5	Site 4	Site 3	Site 2	Site 1
	ATT,CGAGGCT	ATGACAAA	TTTA,ATAA	TAATTGCTATA	ATAGAACG	TTAA	CGACTA	AAAT	TAAT	ATAT
	6 ± 2	1.2 ± 0.8	0.2 ± 0.07	0.3 ± 0.05	0.5 ± 0.1	0.7 ± 0.2	1.5 ± 0.9	1 ± 0.2	0.5 ± 0.2	0.9 ± 0.2

Table 5.14: C_{50} values (μM) derived from quantitative analysis of the Series D MS1/MS2 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

AGGCCA (site 1), AGACCT (site 2), AGTCCT (site 3) and AGATCT (site 4) have similar C_{50} values. The only link between these sequences is WGNNCW (N being any base), but other GGCC sites and GCCC are not bound. It can be assumed that AGGCCA is therefore not the ligand target in site 1, and that base steps in the GNNC binding sequence are consequently essential.

M2S-D11, M2S-D33, M2S-D44, M2S-D55, M2S-D66 and M2S-D77 showed no cleavage protection on the SASK1/SASK2 fragment.

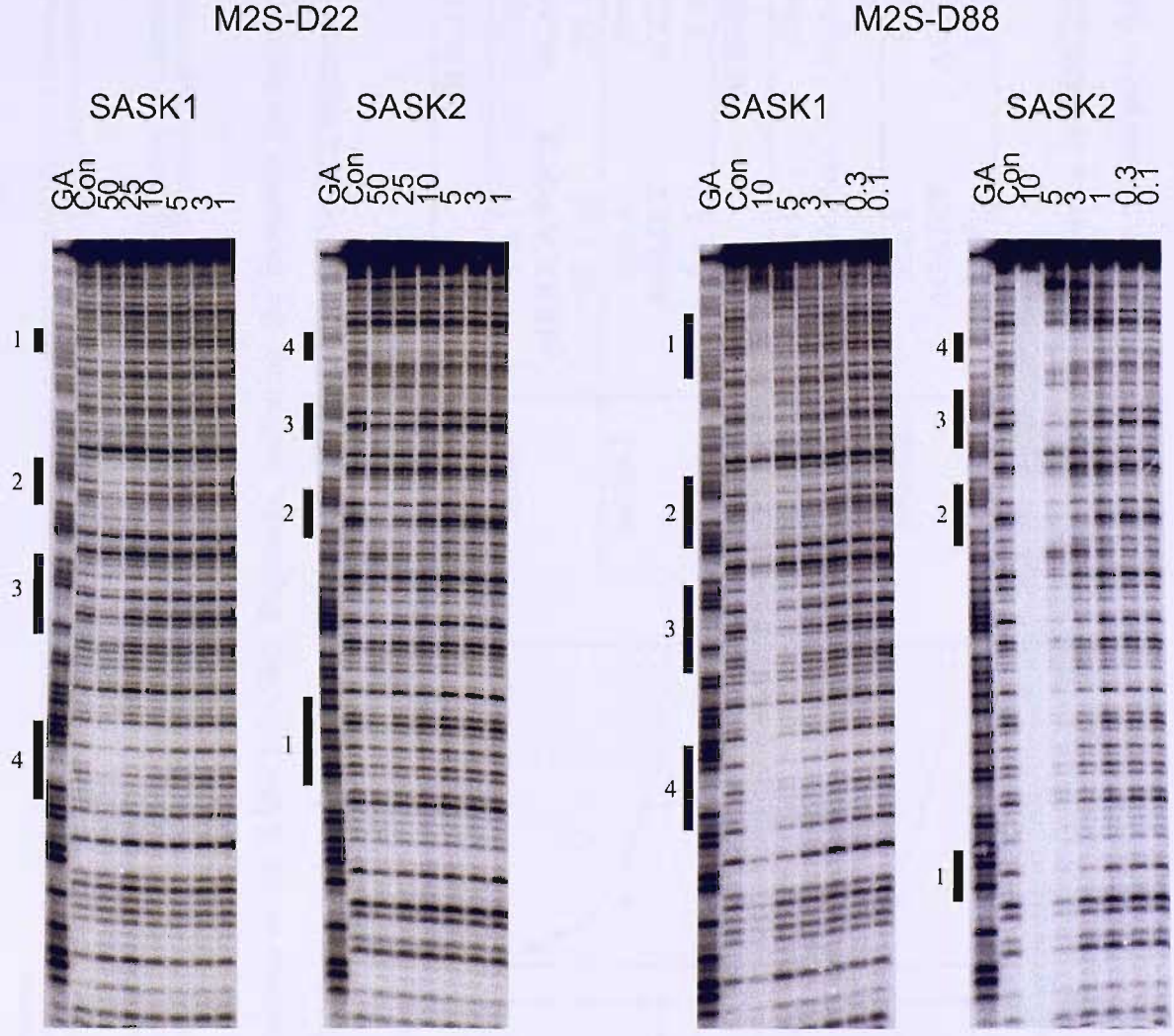


Figure 5.46: DNase I footprinting gels for the interaction of M2S-D22 and M2S-D88 with SASK1 and SASK2 DNA fragments. The clearest footprints are indicated by the filled boxes and are highlighted in the sequences shown in Figure 5.47. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

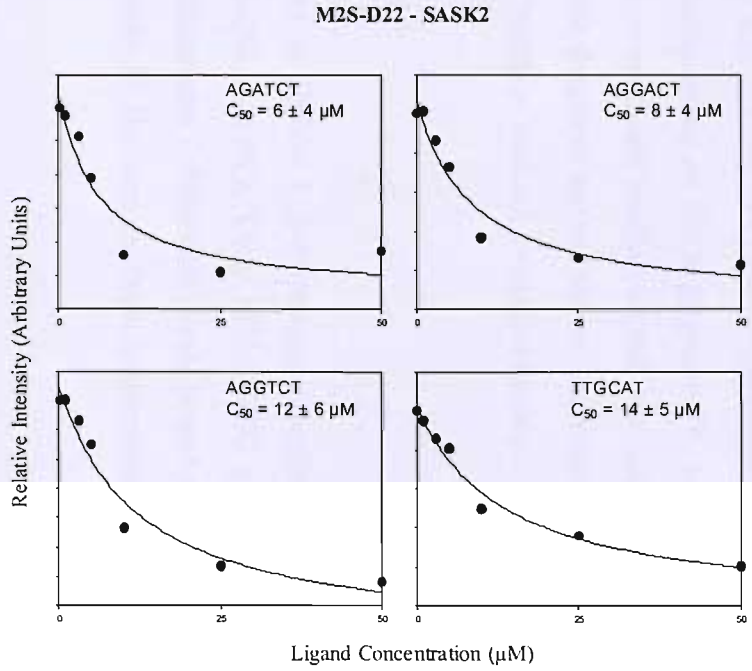
M2S-D22: SASK1 (top); SASK2 (bottom)

```
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'
```

M2S-D88: SASK1 (top); SASK2 (bottom)

```
5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'
```

Figure 5.47: [Above] Sequences of the SASK1/SASK2 fragments, indicating the footprints (in red) produced by M2S-D22 and M2S-D88. [Below left] Examples of footprinting plots derived from the binding of M2S-D22 to the SASK2 fragment.



	M2S-D22			
SASK1	Site 1 AGGCCA,TGCA 52 ± 10	Site 2 AGACCT 29 ± 5	Site 3 AGTCCT 43 ± 8	Site 4 AGATCT 18 ± 3
SASK2	Site 4 AGATCT 6 ± 4	Site 3 AGGACT 8 ± 4	Site 2 AGGTCT 12 ± 6	Site 1 TGCA,TGGCCT 14 ± 5
	M2S-D88			
SASK1	Site 1 AGCGCT,AGGCCA 6 ± 2	Site 2 AGACCT 4 ± 2	Site 3 AGTCCT 3 ± 1.5	Site 4 AGATCT 2 ± 1
SASK2	Site 4 AGATCT 2 ± 0.5	Site 3 AGGACT 2 ± 0.6	Site 2 AGGTCT 2 ± 0.8	Site 1 AGCGCT 1.5 ± 0.6

Table 5.15 (Above right): C_{50} values (µM) derived from quantitative analysis of the Series D SASK1/SASK2 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

Linked Hoechst 33258 / Polyamide Conjugates: Series C

Using known pairing rules, the binding modes in Table 5.16 are possible for the Series C ligands. The other ligands M2M-C107, M2M-C207, M2M-C507 and M2M-C707 showed no cleavage protection in preliminary experiments and so are not included.

M2M-C307 **M2M-C407** **M2M-C607**	Dp-Py-Py-Py Dp-Py-Py-Py Py-Py-Py-Dp Py-Py-Py-Dp 5′-W W W W W W W W W W W W W W W W W W
	The three ligands differ only by the linkages.
M2M-C807	Mp-Pzi-Bzi Mp-Pzi-Bzi Bzi-Pzi-Mp Bzi-Pzi-Mp 5′-W W W W W W W W W W W W W W W W

Table 5.16: Proposed DNA binding sites for the Series C ligands. The number of bases occupied by the linkages is unknown, so those shown are just representative of an undefined occupied space.

The Series C compounds were initially tested on the MS1/MS2 and SASK1/SASK2 fragments to establish approximate binding selectivity and affinity. DNase I footprinting experiments were then carried out on the specifically-designed fragment SASK3/SASK4. Hydroxyl radical footprinting was utilised to better establish binding selectivity on these fragments. Hydroxyl radical footprinting on DNA fragment P6, which contains long A/T-tracts, was carried out to better establish binding of the A/T-selective ligands. DNase I footprinting experiments for this fragment are not shown, as control DNase I cleavage was very poor in these long $A_n.T_n$ tracts, so making footprints unclear.

MS1/MS2 and SASK1/SASK2

Figure 5.48 shows the results of DNase I footprinting experiments with the Series C ligands on MS1/MS2 and SASK1/SASK2. Only M2M-C407 and M2M-C807 showed cleavage protection on these fragments (M2M-C807 only bound to SASK1/SASK2). The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.17.

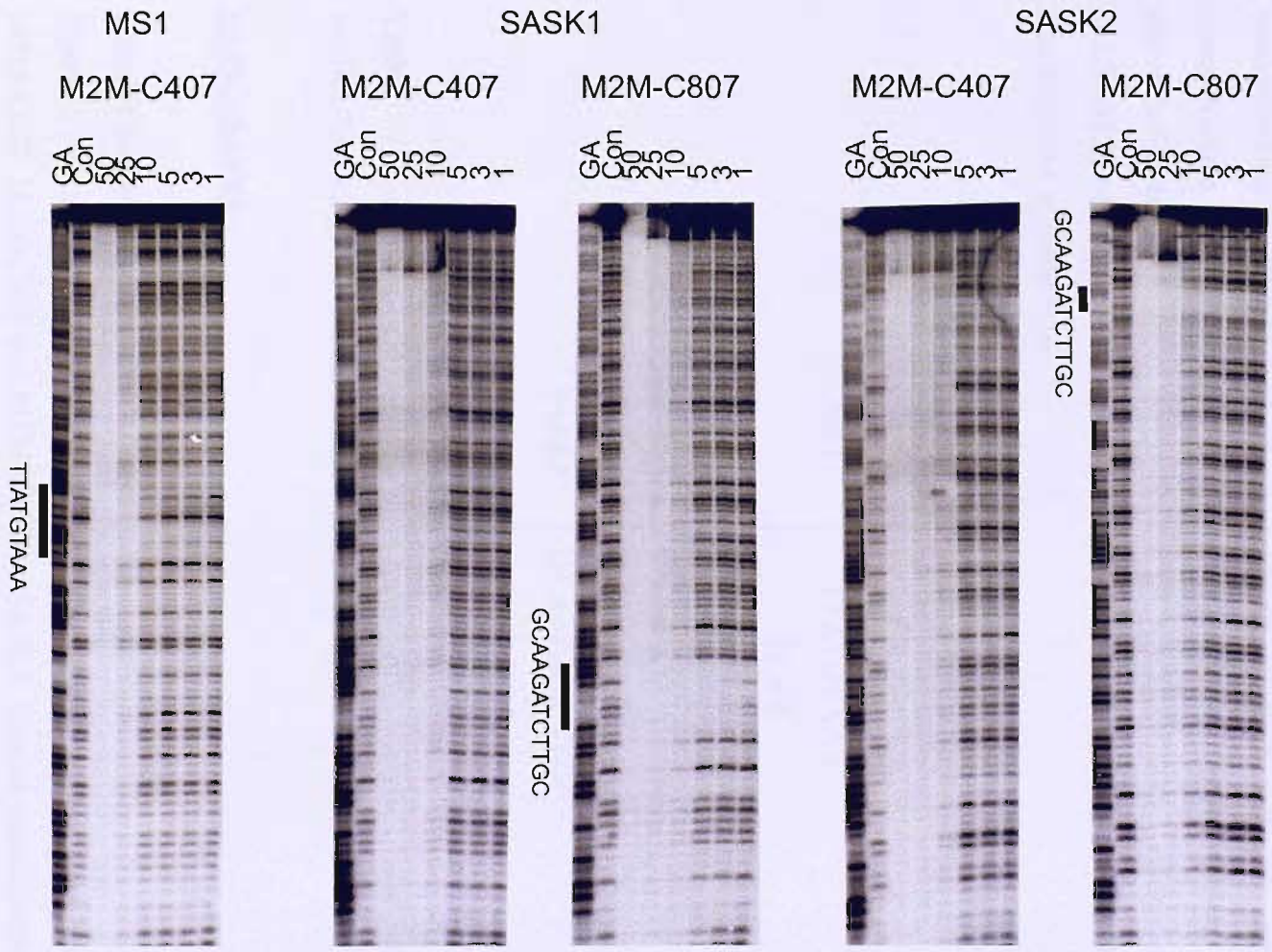


Figure 5.48: DNase I footprinting gels for the interaction of M2M-C407 with the MS1 fragment, and M2M-C407 and M2M-C807 with SASK1 and SASK2 fragments. The clearest footprints are indicated by the filled boxes and sequences. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

MS1/MS2: M2M-C407 shows non-selective binding at 50 μM, with one region of attenuated cleavage at lower concentrations around the sequence TTATGTAAA. This is the longest A/T-tract in the fragment, though it contains a single GC base pair in the centre.

SASK1/SASK2: M2M-C407 has bound the SASK1/SASK2 fragment non-selectively at concentrations of 10 μM and above, with no selective footprints apparent at lower concentrations. M2M-C807 also binds non-selectively at concentrations of 10 μM and above, but a distinct footprint is produced at GCAAGATCTTGC, evident on both SASK 1 and SASK2, which persists to low concentrations; this suggests a different binding mode to that proposed in Table 5.16.

M2M-C407	
MS1	TTATGTAAA
	15 ± 9
M2M-C807	
SASK1	GCAAGATCTTGC
	1 ± 0.02
SASK2	GCAAGATCTTGC
	2 ± 0.5

Table 5.17: C_{50} values (μM) derived from quantitative analysis of the Series C MS1 and SASK1/SASK2 gels (with the standard errors shown).

SASK3/SASK4

DNase I footprinting

Figure 5.49 shows the results of DNase I footprinting experiments with M2M-C307, M2M-C407, M2M-C607 and M2M-C807 on SASK3. Similar experiments on SASK4 are shown in Figure 5.50. Figure 5.51 shows the sequence of these fragments with the observed footprinting binding sites highlighted. An example of footprinting plots derived from the binding of M2M-C807 to SASK3 is in Figure 5.52. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 5.18.

It can be seen that M2M-C307, M2M-C407 and M2M-C607 have all produced a

single footprint on this fragment around the sequence AAATATATATTT (site 1). M2M-C407 binds to this the strongest, possibly explaining its observed binding to MS1/MS2 and SASK1/SASK2, whilst M2M-C307 is weaker and M2M-C607 the weakest ligand.

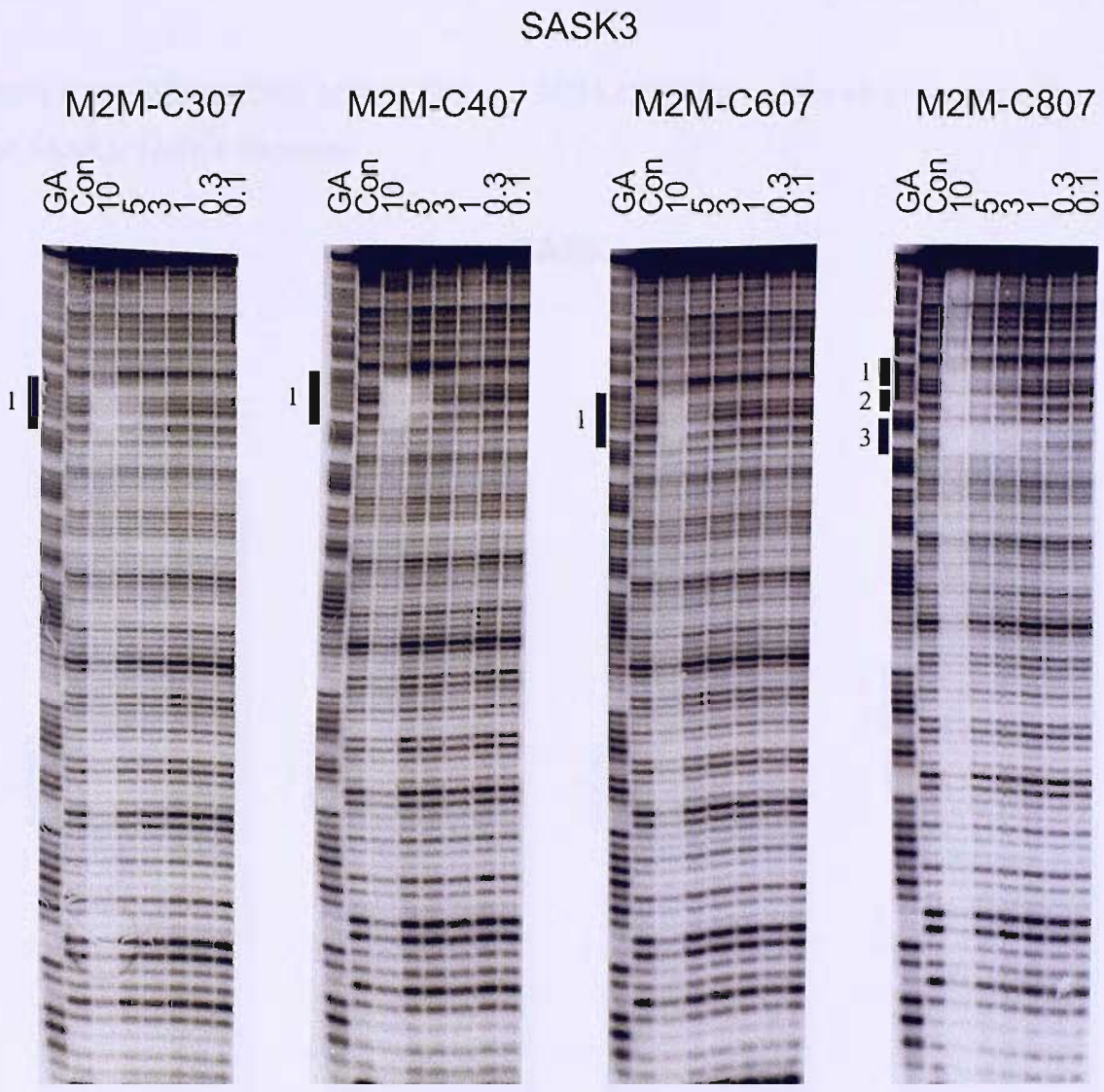


Figure 5.49: DNase I footprinting gels for the interaction of the Series C ligands with the SASK3 fragment. The clearest footprints are indicated by the filled boxes and are highlighted in the sequences shown in Figure 5.51. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-C807 is unique in that it binds to other sites as well; AGATCTTGC (site 3) is the best site, with AAATATA (site 1) and ATATTT (site 2) bound as weaker sites. The best

site supports the results with SASK1/SASK2, which suggested that GWWCWWG might be the preferred target, whilst the splitting of the AAATATATATTT sequence into two distinct footprints suggests the central alternating region ($[AT]_4$) is a weak site for M2M-C807.

M2M-C107, M2M-C207, M2M-C507 and M2M-C707 showed no cleavage protection on the SASK3/SASK4 fragment.

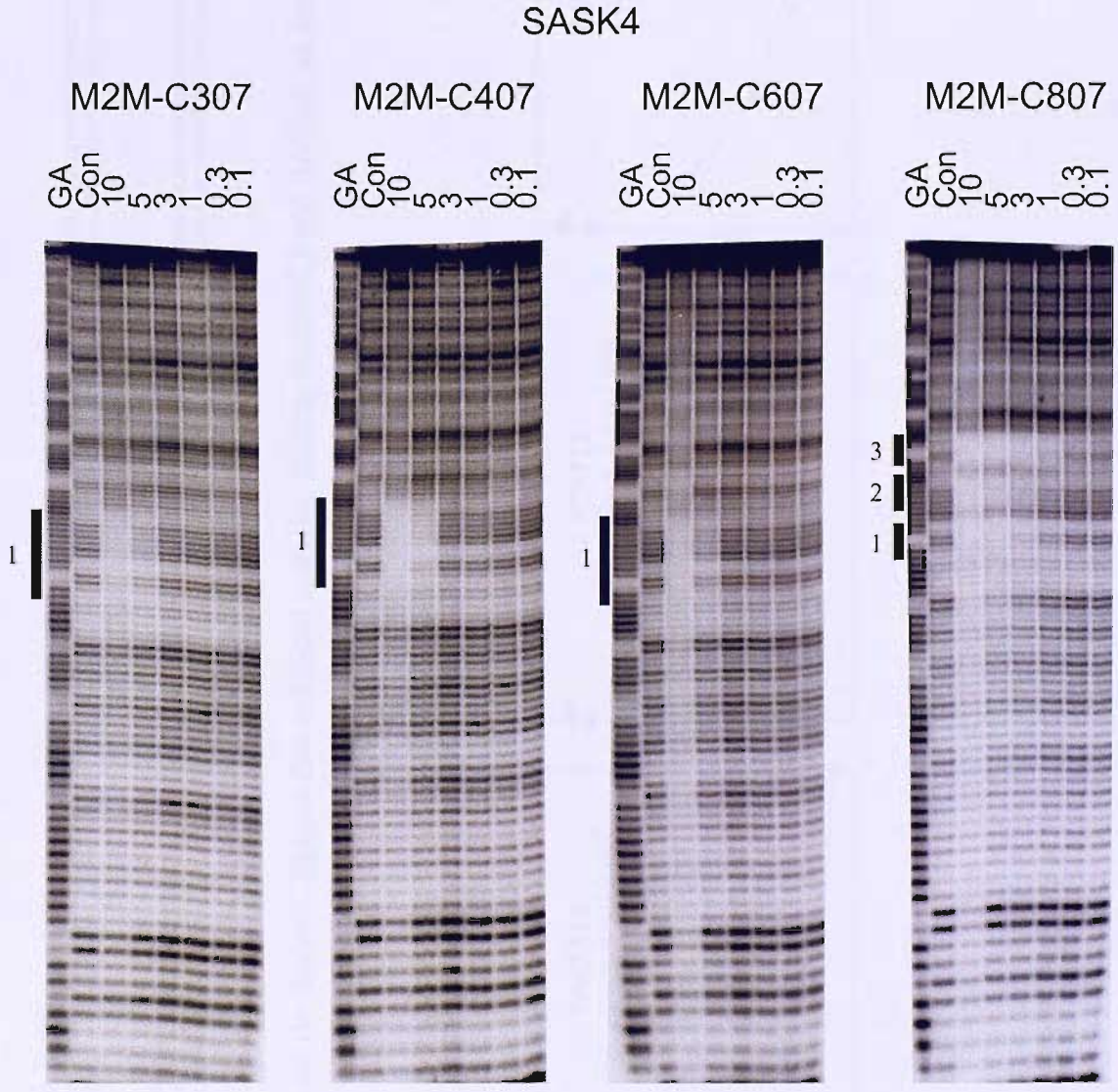


Figure 5.50: DNase I footprinting gels for the interaction of the Series C ligands with the SASK4 fragment. The clearest footprints are indicated by the filled boxes and are highlighted in the sequences shown in Figure 5.51. Sites are numbered as mentioned in the text. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane.

M2M-C307; M2M-C407; M2M-C607: SASK3 (top); SASK4 (bottom)

```
5'-GGATCCAGCAAGCGCGCTTGCAACCCTATAGGGTTGCAAGTCTTGCAAATATATATATTTGCAAGATCTTGCAAGGCTTGCGCCAAGCTTGCAAGGGTATACCCTTGCAAGCTAGCTTGCAAGCGCTTGCAAGATACTTGGCATGGATCC-3'
3'-CCTAGGTCGTTCGCGCGAACGTTGGGATATCCCAACGTTCAGAACGTTTATATATATAAACGTTCTAGAACGTTCCGAACGCGGTTCGAACGTTCCCATATGGGAACGTTCGATCGAACGTTCGCGAACGTTCTATGAACCGTACCTAGG-5'
```

M2M-C807: SASK3 (top); SASK4 (bottom)

```
5'-GGATCCAGCAAGCGCGCTTGCAACCCTATAGGGTTGCAAGTCTTGCAAATATATATATTTGCAAGATCTTGCAAGGCTTGCGCCAAGCTTGCAAGGGTATACCCTTGCAAGCTAGCTTGCAAGCGCTTGCAAGATACTTGGCATGGATCC-3'
3'-CCTAGGTCGTTCGCGCGAACGTTGGGATATCCCAACGTTCAGAACGTTTATATATATAAACGTTCTAGAACGTTCCGAACGCGGTTCGAACGTTCCCATATGGGAACGTTCGATCGAACGTTCGCGAACGTTCTATGAACCGTACCTAGG-5'
```

Figure 5.51: Binding sites for the Series C ligands that exhibited selective binding to SASK3 and SASK4, as derived from Figures 5.49 and 5.50.

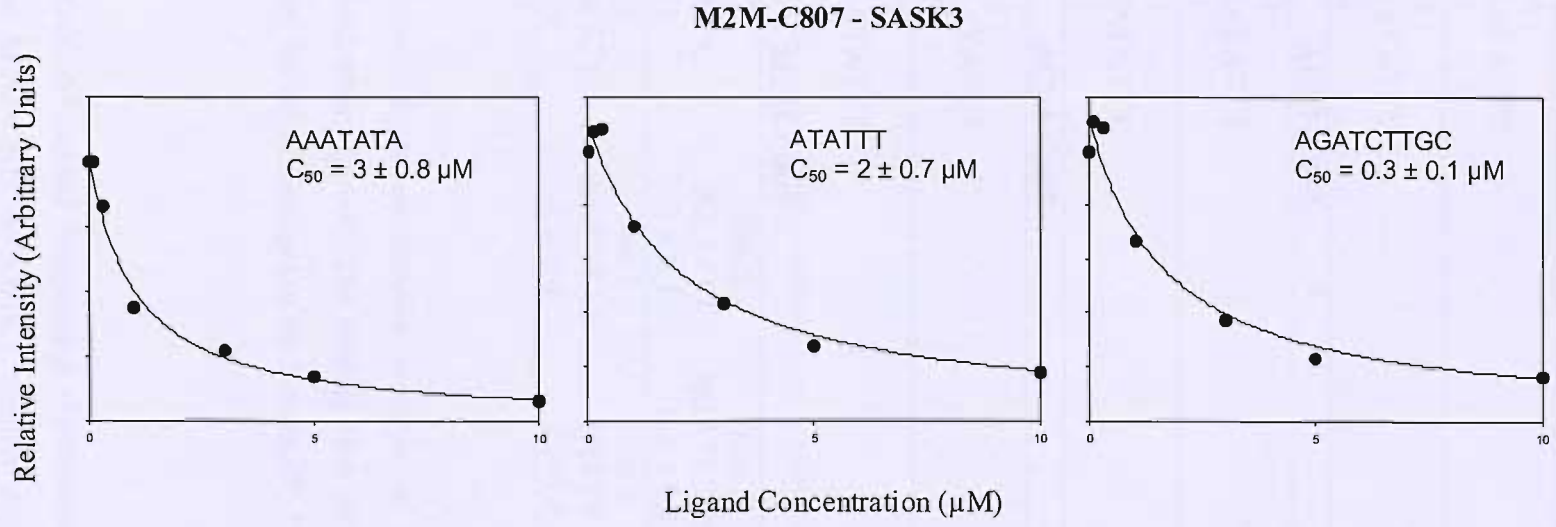


Figure 5.52: Examples of footprinting plots for the interaction of M2M-C807 with SASK3.

M2M-C307			
SASK3	Site 1 AAATATATATATTT 5 ± 2		
SASK4	Site 1 AAATATATATATTT 5 ± 3		
M2M-C407			
SASK3	Site 1 AAATATATATATTT 3 ± 1		
SASK4	Site 1 AAATATATATATTT 3 ± 1		
M2M-C607			
SASK3	Site 1 AAATATATATATTT 6 ± 3		
SASK4	Site 1 AAATATATATATTT 7 ± 5		
M2M-C807			
SASK3	Site 1 AAATATA 3 ± 0.8	Site 2 ATATTT 2 ± 0.7	Site 3 AGATCTTGC 0.3 ± 0.1
SASK4	Site 3 GCAAGATCT 1 ± 0.3	Site 2 AATATAT 2 ± 0.4	Site 1 TATTT 2 ± 0.4

Table 5.18: C_{50} values (μM) derived from quantitative analysis of the Series C SASK3/SASK4 gels (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3').

Hydroxyl radical footprinting

Figure 5.53 shows the results of hydroxyl radical footprinting experiments with M2M-C307, M2M-C407, M2M-C607 and M2M-C807 on SASK3. Similar experiments with SASK4 fragment are shown in Figure 5.54.

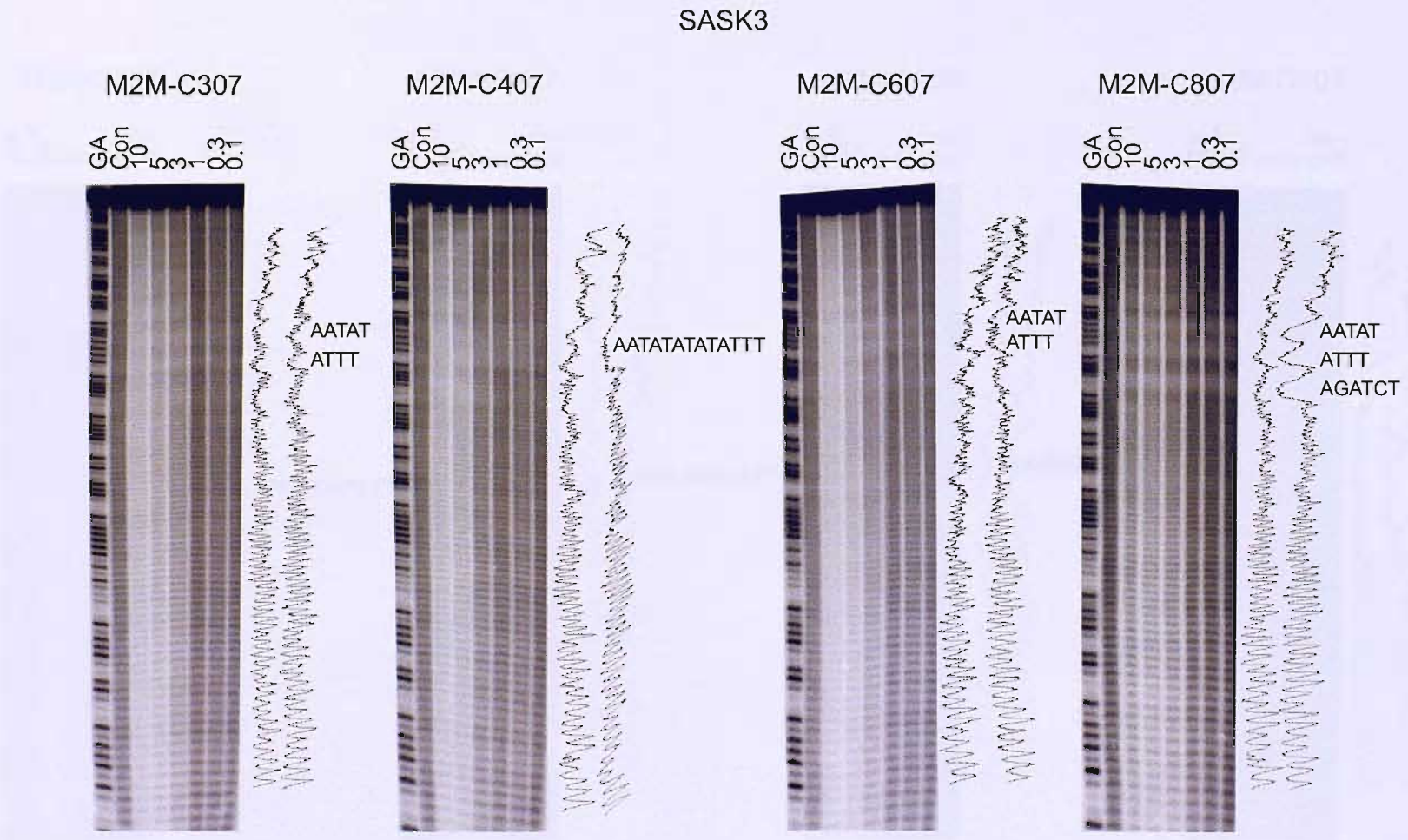


Figure 5.53: Hydroxyl radical footprinting gels of the Series C ligands with the SASK3 fragment. Only the clearest footprints are marked. Densitometer plots were derived from the 0.3 μM lane for all ligands. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane.

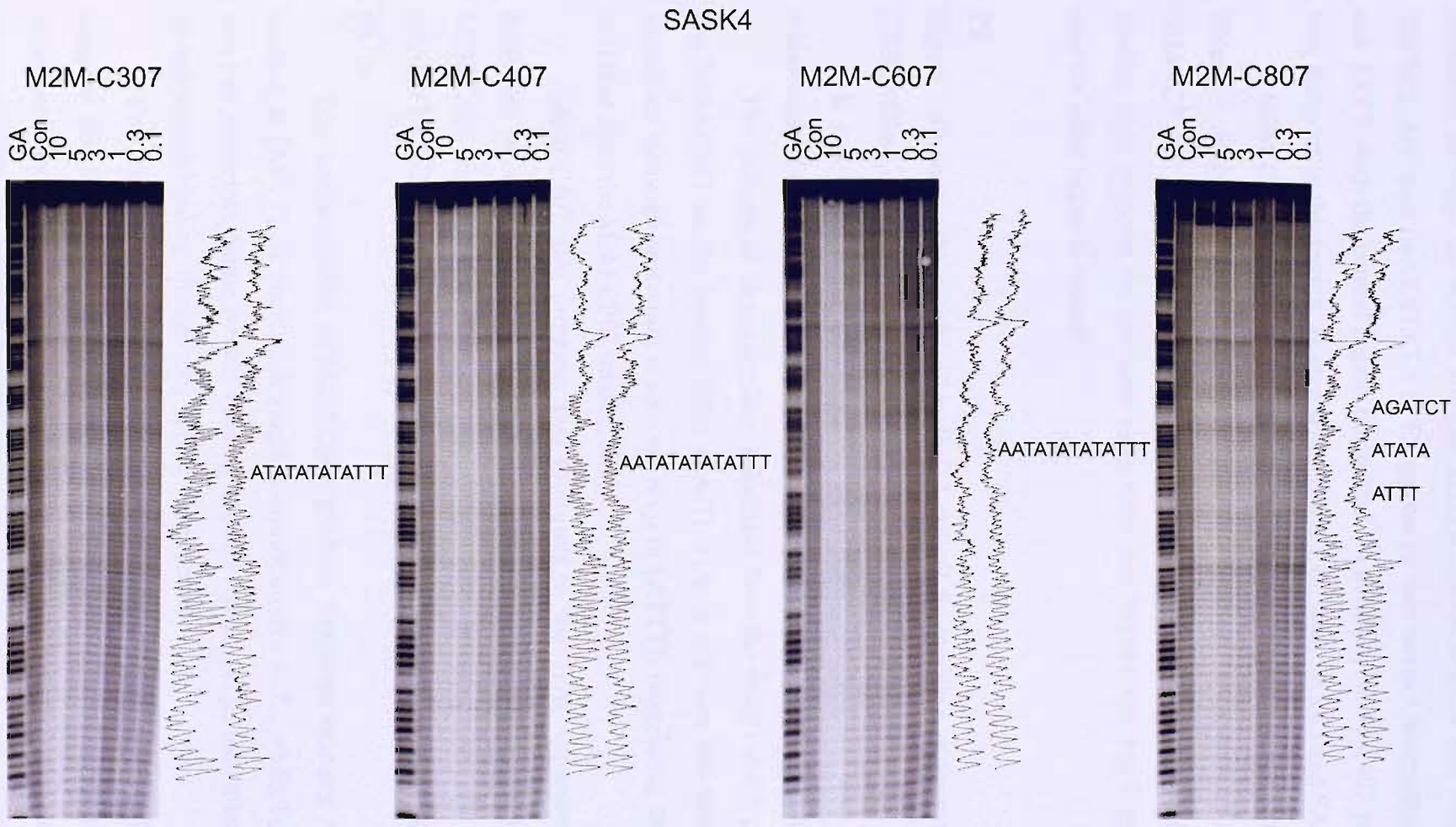


Figure 5.54: Hydroxyl radical footprinting gels of the Series C ligands with the SASK4 fragment. Only the clearest footprints are marked. Densitometer plots were derived from the 0.3 μM lane for all ligands. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane.

The hydroxyl radical footprinting experiments show more clearly defined regions of attenuated cleavage for M2M-C307 and M2M-C607 on SASK3 than seen with DNase I. The long A/T-tract (AATATATATTT) appears as two distinct footprints around AATAT and ATTT, with the central region not bound. In contrast, M2M-C407 produces a single long footprint in this region, as seen with DNase I, on both SASK3 and SASK4.

M2M-C807 attenuates cleavage on SASK3 and SASK4 in the same places as the DNase I footprints, with attenuated cleavage evident around AATAT, ATTT and AGATCT on SASK4. The observation that the long A/T-tract contains two discrete binding sites supports the previous suggestion that M2M-C807 has a greater selectivity than the other Series C ligands.

P6

Figure 5.55 shows the results of hydroxyl radical footprinting experiments with M2M-C307, M2M-C407, M2M-C607 and M2M-C807 on the P6 fragment.

It can be seen that M2M-C307, M2M-C407, M2M-C607 and M2M-807 bind within each of the A/T-tracts (A_6T_6, $[AT]_6$, T_6A_6 and $[ATTT]_3$) at high concentrations.

The differential cleavage plots (calculated from the results at 0.3 μM ligand) show that M2M-C307 has the greatest effect at A_6T_6. T_6A_6 is split into two footprints, whilst no significant cleavage protection is seen at $[AT]_6$ or $[ATTT]_3$ (suggesting that TpA steps are excluded from the M2M-C307 binding site).

M2M-C407 also produces two footprints in the T_6A_6 sequence, while multiple footprints are seen with $[ATTT]_3$, indicating that the target binding sequence is short for M2M-C407 (another possible explanation for the footprints on the MS1/MS2 and SASK1/SASK2 fragments). It is not clear whether cleavage is attenuated at A_6T_6 and $[AT]_6$.

Low concentrations of M2M-C607 produce footprints at every A/T-tract, though binding at $[ATTT]_3$ is unclear. A single footprint covers A_6T_6, while T_6A_6 is again split into two footprints. Some weaker binding is observed at $[AT]_6$, indicating that TpA steps do not abolish binding, though they appear to reduce it.

Two distinct footprints are seen at both A_6T_6 and T_6A_6 with M2M-C807, which suggests that the centre of these sites (the ApT and TpA steps) are bound less well. No significant binding is observed to $[AT]_6$ and $[ATTT]_3$, although cleavage attenuation at the terminals of the P6 fragment suggest strong binding to GATC sites, as seen with previous fragments.

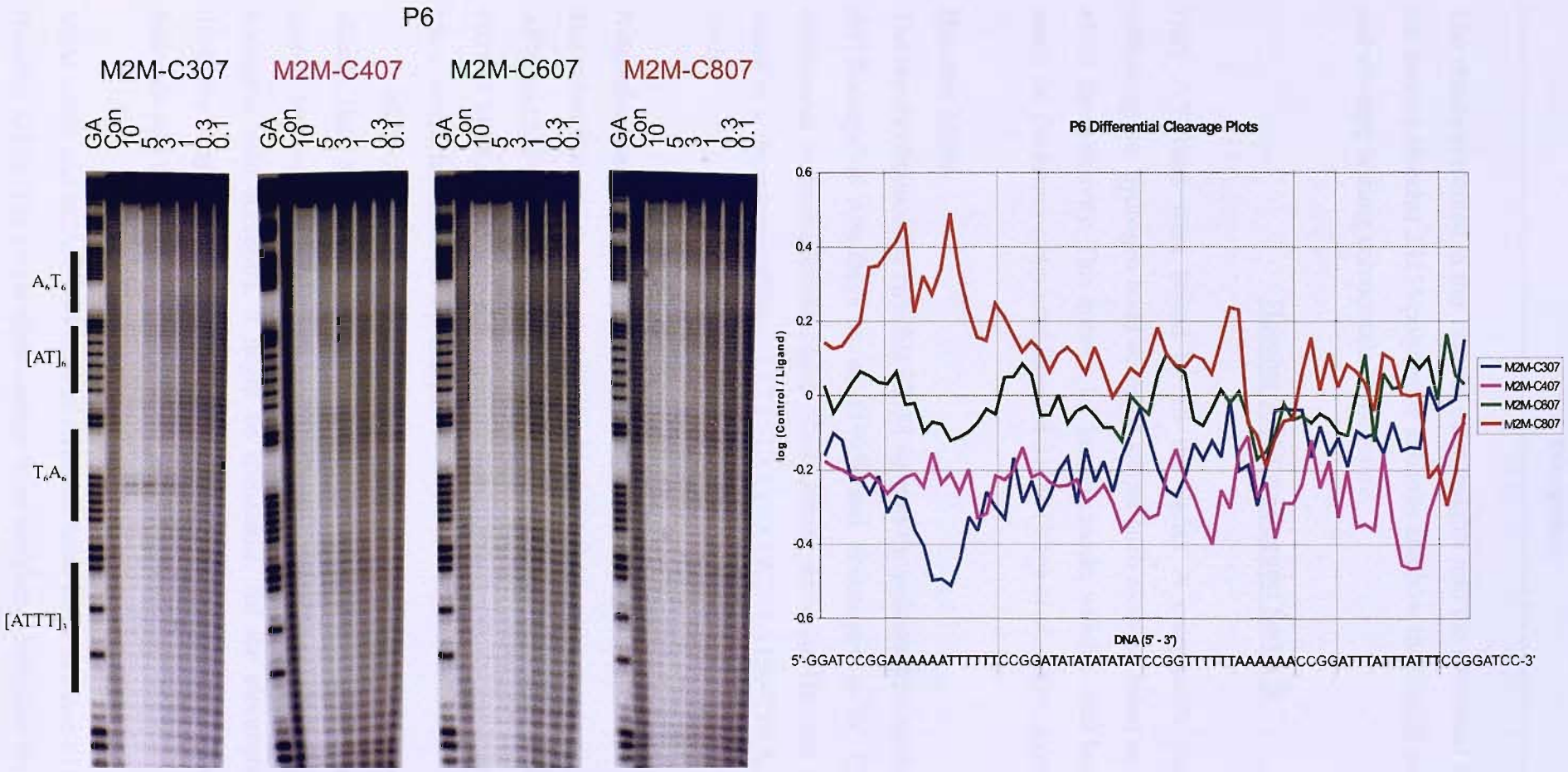


Figure 5.55: [Left] Hydroxyl radical footprinting gel of the Series C ligands with the P6 fragment. A/T-tracts are marked to the left. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane. [Right] Differential cleavage plot of the Series C ligands against the P6 DNA sequence (5'-3') derived from the 0.3 μM lanes of the hydroxyl radical gels.

Discussion

The results presented in this chapter offer insights into the individual binding properties of the various Hoechst 33258/polyamide subunits and how they might lead to more selective and stronger binding compounds in the future.

Hoechst 33258 Analogues: Series A

Only A/T-tracts were bound by all the Series A compounds (and Hoechst 33258), indicating any hydrogen bond acceptor atoms, such as those found on M2M-A066, do not affect the selectivity. This infers a 1:1 binding mode, which is well known as the preferred mode for the parent compound Hoechst 33258 (Vega *et al.*, 1994; Aymami *et al.*, 1999).

Hoechst 33258

The results obtained for Hoechst 33258 support the published research on this ligand, with the findings of Abu-Daya *et al.* (1995) and Breusegem *et al.* (2002) most directly comparable to those presented here. These papers show that Hoechst 33258 binds to A/T-tracts in a decreasing affinity AATT>TAAT=ATAT>TATA=TTAA, as is found in this study.

N-terminal substitution: Thiophene and furan

The substitution of the Hoechst 33258 N-terminal phenyl group with thiophene in M2M-A010 and M2M-A016 has reduced the binding affinity. TpA steps significantly reduce the affinity, more so than with Hoechst 33258, suggesting that the bulky terminal phenyl group has a limited tolerance for TpA steps.

M2M-A016 binds to the preferred AATT sequence with about 2-fold stronger affinity than M2M-A010, suggesting that the thiophene sulphur facing into the minor groove increases the affinity without affecting the selectivity (so it is not acting as a hydrogen bond acceptor). It might be speculated that the electrostatic potential of the thiophene ring is now positioned away from the DNA minor groove, so forcing the molecule into the groove, thereby enhancing the affinity.

M2M-A040 and M2M-A044 have N-terminal furan rings in place of the phenol group of Hoechst 33258. The single atom change from thiophene decreased the binding strength to

AATT on pAAD1, while also relaxing the selectivity so that it includes some TpA steps. This may explain the strong binding of M2M-A040 to A/T-tracts in the MS1/MS2 fragment. This relaxed selectivity is most pronounced with M2M-A044, for which the preference for AATT is reduced while binding to sites containing TpA steps is increased.

The apparent increase in the affinity of M2M-A044 for sites containing TpA steps may be related to the way the furan oxygen is positioned, so as to poke deeper into the minor groove of DNA than with M2M-A040. Since no extra hydrogen bonds are present, it can be suggested that the electrostatic potential of the furan ring, now facing more directly away from the groove, forces stronger binding of the ligand in the wider groove at TpA steps.

N-terminal substitution: Bulky ring moieties

M2M-A020, M2M-A030, M2M-A077 and M2M-A081 are related by the substitution of the Hoechst 33258 N-terminal phenol for bulky ring groups (benzopyrrole, ferrocene, an inverted N-methylbenzimidazole and quinoline respectively). Although the overall binding affinity for A/T-tracts has been weakened by these substitutions, the relative affinity for TpA steps has strengthened relative to Hoechst 33258.

Taking into account the findings for the thiophene and furan-substituted ligands, it can be suggested that the bulky ring groups force the ligand into the groove for strong binding. This effect is most noticeable with M2M-A020, M2M-A030 and M2M-A081, where the ring moieties are positioned pointing away from the DNA minor groove. However, the observation that affinity of these ligands is decreased compared to Hoechst 33258 (500-fold in the case of M2M-A030) indicates that this effect is limited to TpA steps (relaxing the specificity for ApA and ApT).

Although it might be suggested that the benzopyrrole of M2M-A020 forms an extra hydrogen bond at TpA steps, this cannot account for the increased TpA binding of the other ligands. It can therefore be proposed that TpA sites are bound relatively strongly as the wider groove at these steps can accommodate the bulky ring systems of these ligands. The bulky N-terminal groups (and methyl group in the case of M2M-A077) may also interact with the DNA backbone to strengthen binding at TpA steps.

N-terminal substitution: Imidazole

The substitution of the Hoechst 33258 N-terminal phenol for imidazole in M2M-A055 and N-methylimidazole in M2M-A066 has also generally decreased binding strength, though

with no overall change in selectivity. The imidazole nitrogen clearly does not form hydrogen bonds with the 2-amino group of guanine.

A relative increase in tolerance of M2M-A055 and M2M-A066 for TpA steps, relative to Hoechst 33258 and M2M-A040, suggests that their amide (M2M-A055) and N-methyl (M2M-A066) groups play a role in the sequence selectivity. It is possible that these groups increase the size of the N-terminal ring, allowing the ligands to more fully occupy the wide minor groove at TpA steps. It is also possible that these groups interact with the sugar-phosphate backbone (as with M2M-A077), due to the sequence dependence of DNA local structure. This could explain the variation in selectivity between M2M-A055, M2M-A066 and M2M-A077, with slightly different ring structures allowing different sequences of DNA to be preferentially bound.

General Conclusions

The results for the Series A compounds have shown that changing one ring system can have subtle effects on selectivity and affinity of binding.

It is clear that Hoechst 33258 is the strongest binding of the ligands tested, suggesting that an N-terminal phenyl group is best for strong binding with A/T-tracts. However, Hoechst 33258 is selective for ApA or ApT steps, discriminating against TpA. To overcome this selectivity, a ring system that more fully occupies the wider groove at TpA is required. It is possible that selectivity can also be altered through sequence-specific interactions between a ligand and the DNA sugar-phosphate backbone.

Hoechst 33258 / Polyamide Conjugates: Series B

The Series B ligands are conjugates of Hoechst 33258 and polyamides, containing benzimidazole, pyridoimidazole, N-methylpyrrole and N-methylimidazole ring systems. The changes affected the sequence selectivity and affinity, leading to a better understanding of the role each ring moiety plays in binding to the DNA minor groove.

N-terminal Py-Im

M2M-B071 (Mp-Bzi-Py-Py-Im), M2M-B073 (Mp-Pzi-Py-Py-Im) and M2M-B097 (Mp-Bzi-Bzi-Py-Im) all have an N-terminal Py-Im, which, in the 2:1 mode, might target two GC base pairs that are separated by different numbers of AT base pairs (a two-unit overlap might bind GC; a three-unit overlap will bind GWC; and a four-unit overlap will bind

GWWC).

M2M-B071 has a clear preference for the three-unit overlap GWC, with weaker binding to GWWC and GC. This optimal overlap is probably because the benzimidazole/imidazole pairing, which would be generated by the four-unit overlap, is sterically unfavourable (contrary to work on benzimidazole-containing hairpin polyamides (Briehn *et al.* (2003)); whilst the two-unit overlap does not create enough ligand-ligand and ligand-DNA contacts for strong binding.

When the benzimidazole of M2M-B071 is substituted for a pyridoimidazole in M2M-B073, GWC (*i.e.* the three-unit overlap) is still the preferred binding site, but the affinity for GWWC is reduced significantly (GC binding is negligible). This indicates that the pyridoimidazole dislikes binding opposite imidazole more than benzimidazole. A pyridoimidazole ring system is therefore a useful tool for creating staggered 2:1 ligand binding, although increased A/T binding (compared with M2M-B071) suggests that some 1:1 binding also results from this substitution.

M2M-B097 contains two contiguous benzimidazole ring systems (instead of Bzi-Py in M2M-B071). This ligand largely targets A/T-tracts, suggesting that it binds in a classic 1:1 mode. However, binding to GWC sequences (especially when flanked by AT base pairs) is stronger than with M2M-B071. It has a clear preference for this site over GWWC and GC (which are both relatively weaker than with M2M-B071). The two contiguous benzimidazoles of M2M-B097 may be too rigid (and therefore less annular) for both rings to bind in a 2:1 conformation, hence preferring the three-unit overlap. It appears that some Bzi/Im pairing is permitted in this case; possibly indicating that it is the Mp-Bzi terminal that inhibits binding of the other Series B ligands.

N-terminal Im-Im

M2M-B072 (Mp-Bzi-Py-Im-Im), M2M-B074 (Mp-Pzi-Py-Py-Im-Im), M2M-B079 (Mp-Bzi-Py-Py-Im-Im), M2M-B084 (Dp-Py-Py-Py-Im-Im) and M2M-B099 (Mp-Bzi-Bzi-Py-Im-Im) all have an N-terminal Im-Im ring system, which might target two consecutive GC base pairs.

The results showed that M2M-B072 binds to GpC steps, and that the proposed target site (GGCC) is bound with similar affinity to GpC steps that are surrounded by AT base pairs. This suggests that the central GpC alone is sufficient for ligand binding. However, the different affinities of the various GGCC sites (*e.g.* AGGCCT and TGGCCA) suggest that six base pairs are involved in defining the complete binding site. It can

therefore be proposed that only the central Py-Im is directly responsible for determining the selectivity in the 2:1 mode, and that the terminal Bzi/Im pairs are redundant (terminal Mp-Bzi/Im pairs were shown to be non-permissive with M2M-B071) as shown in Figure 5.56. With only two units targeting the DNA selectively, binding is generally weaker than that observed for the other N-terminal Im-Im ligands, although the affinity for this sequence is clearly modulated by the surrounding base pairs.

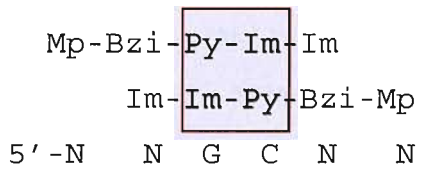


Figure 5.56: Proposed binding mode of M2M-B072, in which only the central Py-Im pair is responsible for the selectivity.

Lengthening the polyamide by adding an extra pyrrole ring as in M2M-B074 (while also substituting benzimidazole with pyridoimidazole) allows a four-unit 2:1 overlap, with no Pzi/Im pairing. This has resulted in GGCC being substantially preferred over isolated GpC steps; the binding affinity is also strengthened, presumably due to the four-unit overlap. The surrounding base pairs appear to affect the selectivity, and YGGCCR is a relatively poor binding site. The ligand also binds weakly to GWC, suggesting that it may also bind with a five-unit overlap (with pyridoimidazole paired with imidazole). However, the Pzi/Im pairing does not impart any selectivity, so GGWCC is not preferentially targeted, in a similar fashion to M2M-B072. The stronger binding affinity of M2M-B074 compared to M2M-B072 has relaxed the sequence selectivity; some single base changes weaken but do not abolishing the binding, while some G/C-tracts that do not contain GpC steps are also bound weakly. This may suggest that van der Waals forces and electrostatic interactions play a more important role than hydrogen bonding. It is also interesting to note that M2M-B074 also binds to some long A/T-tracts, presumably in the 1:1 mode, maybe because longer compounds are more likely to use this binding conformation. This is consistent with the observation that sequence selectivity is decreased with longer ligands (Kelly *et al.*; 1996). However, there is no indication to suggest footprint size is significantly larger for the longer ligand (as could be expected).

M2M-B079 differs from M2M-B074 by the substitution of the pyridoimidazole with benzimidazole. This single atom change has generally weakened the binding affinity

and relaxed the selectivity for GGCC, as well as resulting in more binding to long A/T-tracts (in the 1:1 mode). It is clear that pyridoimidazole, as well as creating a staggered 2:1 complex (shown with M2M-B073), binds more strongly than benzimidazole. This stronger affinity is coupled with reduced binding to related but different sequences and to A/T-tracts (as a 1:1 complex).

M2M-B084 is a Dervan-type polyamide (similar to that in Kielkopf *et al.*, 1998), with a dimethylaminopropylamide C-terminal (instead of methylpiperazine) and a pyrrole ring in place of the benzimidazole of M2M-B079. These changes resulted in increased selectivity for GGCC compared with M2M-B074 and M2M-B079, although the surrounding base pairs do not affect the selectivity, unlike with M2M-B074. However, the overall binding affinity of M2M-B084 is weaker, suggesting that the benzimidazole and pyridoimidazole rings produce stronger binding than pyrrole (possibly explaining the decreased selectivity). The proposed secondary binding site at GGWCC is not bound by M2M-B084, though there is a weak interaction with GWC (weaker relative to M2M-B074 and M2M-B079). This indicates that an unlinked five-unit 2:1 overlap is unstable for polyamides with a terminal Im-Im, and only the core Py-Py-Im is involved in selective recognition (in a manner similar to M2M-B072).

M2M-B099 is a similar ligand to M2M-B097, with two contiguous benzimidazole rings at its C-terminus. Binding would therefore be expected at GGCC, as the results with M2M-B097 showed that two contiguous benzimidazoles are too rigid for 2:1 binding. Surprisingly, this ligand only seems to recognise the central GC, as GGCC is bound with the same affinity as isolated GpC steps. This is similar to the results with M2M-B072, where two contiguous N-terminal imidazoles bound opposite Bzi-Py, with only the two core Im/Py pairs responsible for the selectivity (Figure 5.57). However, the long unpaired terminal regions clearly affect the binding, as long G/C-tracts are preferred (as with M2M-B074); directly adjacent GpC steps are poor sites; and affinity is relatively strong.

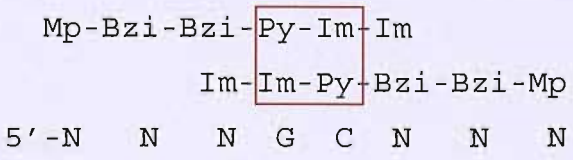


Figure 5.57: Preferred binding mode of M2M-B099, in which only the central Py-Im is responsible for the sequence selectivity in the 2:1 mode.

It is also interesting that M2M-B099 binds weakly to GWC, supporting a five-unit overlap in the 2:1 mode, though there is no binding to GGWCC. This supports the results with M2M-B097 that Bzi/Im pairs are permitted if the benzimidazole is not the terminal ring; and with M2M-B072 that a terminal Bzi-Im pair does not impart any sequence selectivity.

N-terminal Im-Py-Im

M2M-B088 differs from the previous ligands by having an alternating Py/Im structure (Mp-Bzi-Py-Im-Py-Im), which might target alternating GC base pairs. This ligand has a weaker binding affinity than the other Series B ligands (opposite to that seen with a shorter ligand f-ImPyIm (Buchmueller *et al.*, 2006)). Some 1:1 binding to A/T-tracts is also seen (especially when the preferred binding site is not present in the fragment). The proposed GCGC binding site is strongly bound (although some cleavage protection is observed at derivatives such as GTGC and GGCC on MS1/MS2), supporting a four-unit overlap of the ligand in the 2:1 mode. There is no evidence for binding as a two-unit overlap, probably because not enough 2:1 ligand-DNA contacts can be established. The sequence around the primary binding site is also important, and GC base pairs considerably reduce the binding affinity (probably due to steric hindrance from the 2-amino group of guanine). This makes WGCGCW the preferred binding sequence.

Hairpin polyamide

M2M-B083 is unique in the Series B compounds as it includes a gamma amino butyric acid linkage, which has been found to be the optimum linker length for hairpin formation (Mrksich *et al.*, 1994). However, in this case the hairpin configuration is less favourable than binding to A/T-sequences, suggesting that the C-terminal half binds in a 1:1 complex with the DNA. This is probably because the N-terminal Im-Im is too short to establish strong ligand-DNA contacts. The presence of some weaker binding to the predicted hairpin sequence suggests that a longer benzimidazole-containing hairpin polyamide could still be viable.

N-terminal reversed-Bzi

M2M-B091, M2M-B093 and M2M-B095 all possess an N-terminal benzimidazole ring, which is orientated so that it points towards the C-terminus (reversed compared to the normal ligands). This orientation abolishes binding.

1:1 heterodimers

Since M2M-B071 and M2M-B072 bind to completely different sites as homodimers, it was interesting to examine the behaviour of a 1:1 mixture of these compounds, to see whether they might form an unlinked heterodimer (similar to Mrksich and Dervan, 1993a and Geierstanger *et al.*, 1993). The combination of these two ligands produces an additional footprint at AGACCC, which is consistent with the two monomers fully overlapping in a 2:1 conformation, with benzimidazole binding opposite imidazole. This conformation is not favoured by the previously discussed ligand homodimers.

It is interesting to note that none of the other ligand pairs gave evidence of heterodimer formation. The pyridoimidazole of M2M-B073 did not permit heterodimer formation (compared with the benzimidazole of M2M-B071), as did the two contiguous benzimidazoles of M2M-B097.

General Conclusions

It appears that for 2:1 unlinked polyamide binding, the optimum overlap of the two monomers is four units. Two units proved too weak (M2M-B071), even for a linked dimer (M2M-B083), whilst five units seems to be too rigid for good binding (M2M-B074, M2M-B079, M2M-B084 and M2M-B099). However, two units are adequate for strong and selective 2:1 binding if other regions of the oligopeptides aid binding, even though they do not affect the selectivity (M2M-B072 and M2M-B099).

Benzimidazole generally does not bind opposite imidazole in the 2:1 mode, though an exception is when benzimidazole is not the terminal ring system (methylpiperazine may therefore be important in this regard too). Two contiguous benzimidazole rings also inhibit fully overlapped 2:1 binding, probably due to their rigidity (M2M-B097 and M2M-B099). A reversed benzimidazole ("pointing" towards the C-terminal) at the N-terminus of a Hoechst 33258/polyamide conjugate abolishes binding (M2M-B095 compared to M2M-B079).

Pyridoimidazole shows interesting selectivity when compared with benzimidazole, producing stronger binding affinity and increased selectivity (M2M-B074 compared to M2M-B079). Pyridoimidazole is also useful for staggering a 2:1 complex, so targeting a longer sequence of DNA (M2M-B073), which is not seen with benzimidazole (M2M-B071). It is important to note also that pyridoimidazole and benzimidazole seem to bind stronger than pyrrole (M2M-B074 and M2M-B079 compared to M2M-B084), as suggested by Briehn *et al.* (2003), although the selectivity is generally decreased.

The general binding specificities of the Series B ligands that exhibited selective DNA recognition are presented in Table 5.19.

Ligand	Structure	Binding site preference
M2M-B071	Mp-Bzi-Py-Py-Im	GWC>GWWC=GC
M2M-B072	Mp-Bzi-Py-Im-Im	GC=GGCC
M2M-B073	Mp-Pzi-Py-Py-Im	GWC>>GWWC>GC
M2M-B074	Mp-Pzi-Py-Py-Im-Im	GGCC>GWC>>A/T
M2M-B079	Mp-Bzi-Py-Py-Im-Im	GGCC>GWC>A/T
M2M-B083	Mp-Bzi-Py-γ-Py-Im-Im	A/T>WWGG
M2M-B084	Dp-Py-Py-Py-Im-Im	GGCC>>GWC
M2M-B088	Mp-Bzi-Py-Im-Py-Im	WGCGCW
M2M-B097	Mp-Bzi-Bzi-Py-Im	GWC>>GWWC, A/T
M2M-B099	Mp-Bzi-Bzi-Py-Im-Im	GC=GGCC>GWC

Table 5.19: Summary of the preferential binding sites of the Series B ligands.

Hoechst 33258 / Polyamide Conjugates: Series D

The Series D ligands are again based on the Dervan series of polyamides (Dervan, 2001). However, benzimidazole has been incorporated at either the N-terminus of the ligands (linked to an imidazole ring), internally, or at the C-terminus (linked to Mp) to alter binding selectivity and affinity.

N-terminal Bzi-Im

M2S-D11, M2S-D33, M2S-44, M2S-D55 and M2S-D77 are all standard polyamides, but with a benzimidazole-imidazole at their N-termini. The results suggest that the benzimidazole and imidazole are too close for selective binding in the DNA minor groove, though an internal hydrogen bond may be formed between the imidazole and the benzimidazole, so abolishing binding. This could explain the complete lack of binding of M2S-D55, M2S-D77 and M2S-D44, which should therefore only be able to make one (M2S-D77) or two (M2S-D44) selective contacts with the DNA (none in the case of M2S-D55), which may not be sufficient.

M2S-D33 (Dp-Py-Py-Py-Bzi-Im) displays some selective binding. This is

predominantly to A/T-tracts, though the observation that short sequences are bound (3 bp in some cases) suggests a 1:1 binding mode (staggering of a 2:1 complex would bind longer sequences). This is supported by the weaker binding of M2S-D11 (Dp-Py-Py-Im-Bzi-Im), which only binds to A/T-tracts even though another imidazole ring is present. The weaker binding is indicative of only two pyrrole rings forming DNA interactions.

N-terminal linked Bzi-Im

M2S-D22 is similar to other Series D ligands, but the N-terminal Bzi-Im is linked by a six-membered ring. The benzimidazole ring is also in a different orientation to normal, with a hydrogen bond-accepting nitrogen facing into the groove. It also has a C-terminal Mp-Bzi, which may increase the binding affinity without affecting selectivity.

With the possibility of an internal hydrogen bond removed for this ligand, weak binding is observed at both the proposed sequences GGCC and GGWCC (though only GWC for the latter, probably due to the terminal Bzi/Im pair not creating recognition, as observed for the Series B ligands). However, another binding mode must be responsible for the observed stronger binding to GWWC. It is possible that the terminal BzN-Im complex is too tight to bind two contiguous bases well. If the imidazole binds guanine, then BzN will occupy space which will stagger the neighbouring pyrrole, leaving an unrecognised base in the centre of the complex, as shown in Figure 5.58.

```
      Mp-Bzi-Py-Py---Im
               Im---Py-Py-Bzi-Mp
5'-W   W  G  W W  C  W   W
```

Figure 5.58: Proposed binding mode for M2S-D22.

Internal benzimidazole-containing ligands

M2S-D66 has a glycine linker connecting two Bzi-Py-Im oligopeptides to allow more flexibility in the ligand. Surprisingly, this resulted in largely A/T-selective binding, suggesting that the glycine linker is too flexible to allow stable 2:1 binding at the proposed site (GWCWGWC) (the linker might also cause the second half of the ligand to be out of phase with the first half).

However, a $[GC]_n$ tract is also bound by M2S-D66, indicating an alternative 2:1 binding configuration. The glycine linkers may not bind opposite each other, but instead

the ligand is staggered, with the benzimidazole rings opposite the linkers as shown in Figure 5.59. The flexibility of the linker is compensated for by the opposing bulky benzimidazole group, while the inherent flexibility of the central Gly-Bzi pairs might permit binding opposite the 2-amino group of guanine, allowing binding to $[GC]_n$.

```
Mp-Bzi-Py-Im-Gly-Bzi-Py-Im
        Im-Py-Bzi-Gly-Im-Py-Bzi-Mp
5'-W   W  G  C  N   N   G  C  W   W
```

Figure 5.59: Proposed binding mode for M2S-D66. The glycine linkage is shown as "Gly".

M2S-D88 has an alternating Bzi-(Py/Im) sequence (Mp-Bzi-Py-Bzi-Py-Bzi-Im) and so can only bind in a specific 2:1 configuration, as benzimidazole does not bind opposite another benzimidazole, as seen with Hoechst 33258 (Bostock-Smith *et al.*, 2001; Rosu *et al.*, 2002). It appears that the proposed GWWC binding site is not entirely correct, as central GC base pairs are also permitted. However, GGCC is not bound, suggesting that different-base steps are required for binding to GNNC. This ambiguity in the centre suggests that the central Py-Bzi pairs have loose contacts with the DNA that can overcome the steric hindrance with the 2-amino group of guanine. This is most likely due to the three benzimidazole moieties on each monomer creating a decreased annular shape in the whole ligand compared with conventional polyamides, so weakening ring contacts with the DNA.

General Conclusions

An N-terminal Bzi-Im does not bind in the DNA minor groove, possibly due to an internal hydrogen bond being formed between the two ring systems. However, binding is enabled by inverting the benzimidazole (so removing the possible hydrogen bond donor). The close proximity of these ring systems allows only one base to be selectively recognised, with the benzimidazole producing a stagger relative to the DNA, while imidazole forms a hydrogen bond. Moving the benzimidazole to an internal position restricts the possible binding sites (as it cannot be placed opposite another benzimidazole), but also relaxes the selectivity when multiple benzimidazoles are present. This is possibly due to changes in the shape of the ligand, which may be too rigid for selective binding. Adding a glycine residue within the compound removes this rigidity problem, but creates a phasing problem between the ligand and DNA.

The general binding selectivities of the Series D ligands that exhibited selective DNA recognition are presented in Table 5.20.

Ligand	Structure	Binding site preference
M2S-D11	Dp-Py-Py-Im-Bzi-Im	A/T
M2S-D22	Mp-Bzi-Py-Py-BzN-Im	GWWC, A/T > GGCC, GWC
M2S-D33	Dp-Py-Py-Py-Bzi-Im	A/T
M2S-D66	Mp-Bzi-Py-Im-Gly-Bzi-Py-Im	A/T, $[GC]_n$
M2S-D88	Mp-Bzi-Py-Bzi-Py-Bzi-Im	A/T, GNNC

Table 5.20: Summary of the preferential binding sites of the Series D ligands.

Linked Hoechst 33258 / Polyamide Conjugates: Series C

The Series C ligands are "Head-to-Head" or "Tail-to-Tail" linked Hoechst 33258/polyamide conjugates, which were designed so as to enable binding of a single compound to both DNA strands within the minor groove.

Tail-to-tail linked dimers

M2M-C107, M2M-C207 and M2M-C707 are all tail-to-tail linked dimers, which showed no binding to the tested fragments. This lack of binding is probably due to the absence of positively charged C-terminal tails which normally aid binding and directional orientation (White *et al.*, 1997b).

Pyrrole-containing head-to-head linked dimers

M2M-C307, M2M-C407 and M2M-C607 are pyrrole-containing head-to-head linked dimers; the only difference between the three ligands is the linker region ($-CH_2-C\equiv C-CH_2-$, $-[CH_2]_2-$ and $-[CH_2]_3-$ respectively). They are therefore all A/T-selective ligands. However, M2M-C407 binds shorter tracts than the other ligands (as seen in the multiple binding sites at $[ATTT]_3$ on the P6 fragment). The shorter binding site is probably because this short linker occupied the least number of base pairs crossing the minor groove (a longer linker would occupy more space in crossing the DNA). The observation that M2M-C407 is the strongest binding ligand suggests that its linker is the most favourable for in-phase binding to DNA. The binding to TTATGTAAA on MS1/MS2 also suggests cooperative binding

for M2M-C407, as lone $(A/T)_4$ sites are not bound (there is no evidence of this with the other ligands).

M2M-C307 has a slightly longer linker than M2M-C407, which weakens the overall binding affinity, probably because the ligand comes out of phase with the DNA. The weaker binding has removed the binding to TpA seen with the shorter ligand, while the longer linker also appears to have increased the length of its DNA binding sequence. M2M-C607 has the longest linker in this series and produces the weakest binding of the pyrrole-containing head-to-head linked dimers. However, M2M-C607 tolerates TpA steps better than the others, most likely because the long-linker is able to bridge the DNA at the wider minor groove.

Imidazole-containing head-to-head linked dimer

M2M-C507 is comprised entirely of imidazole rings, linked by $-[CH_2]_3-$. No binding was observed for this compound on the tested DNA fragments, which was not surprising as imidazole prefers not to bind opposite another imidazole (Walker *et al.*, 1998b; White *et al.*, 1997a). M2M-C507 also does not bind the long A/T-tract of SASK3/SASK4 and so does not appear to bind as an A/T-specific 1:1 monomer.

Pyrido- and benzimidazole-containing head-to-head linked dimer

M2M-C807 is comprised of pyridoimidazole and benzimidazole rings. Conventional binding rules suggest that it will show selective binding to A/T-tracts. This is indeed seen but there is no cleavage protection at TpA steps (probably because the short, rigid benzene-piperazine-benzene linkage is not able to traverse the wider minor groove at these sequences).

The interesting discovery about this head-to-head linked dimer is that it does not bind simply to A/T-tracts, and GWWCWWG is bound more strongly than the A/T-tracts on both SASK1/SASK2 and SASK3/SASK4 fragments, with P6 exhibiting protection at similar GATC binding sites. It seems likely that this is due to the pyridoimidazole moiety rotating about the bond between it and the benzimidazole group, exposing two hydrogen bond accepting nitrogens to face the minor groove for selective binding to guanine. This reorientation of the ligand will also create a more annular shape, which may assist binding. Assuming that pyridoimidazole does not bind well opposite another ring moiety (as seen with the Series B ligands), the binding mode shown in Figure 5.60 explains the observed binding sites. The linker is apparently too short and rigid to occupy space along the DNA,

and so it binds directly across the groove.

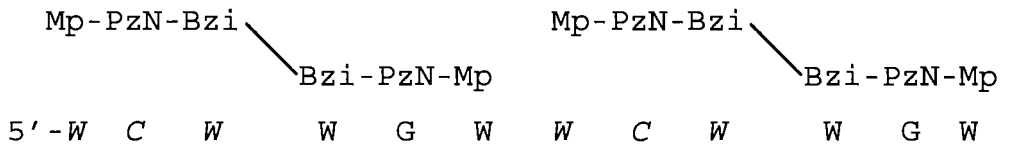


Figure 5.60: Proposed mode for M2M-C807. The reorientated pyridoimidazole is termed PzN.

General Conclusions

The Series C set of ligands are unique in that they are designed to bind across the minor groove of DNA, so targeting both strands with one ligand. If these ligands bind in a 2:1 fashion, then an indefinite length of DNA could be targeted, thereby overcoming the usual length restriction as long ligands comes out of phase with the DNA.

Apart from confirming that a positive terminus is required for polyamide binding, the results indicate that the linker is crucial for binding. A -$[CH_2]_2$- linker creates stronger binding than -CH_2-C≡C-CH_2- and -$[CH_2]_3$-, probably because differences in length alter the phasing between the ligand and DNA. Target site length is also related to linker length (a shorter linkage binds shorter sequences), whilst a longer linker is useful for binding to the wider minor groove at TpA steps. G/C-selectivity is also possible with the cross-DNA-binding ligands, as shown by the guanine selectively of pyridoimidazole in M2M-C807.

Chapter VI

Minor Groove Binding Ligands Containing Isopropyl-Thiazole Units

Introduction

A large number of small molecules are known to bind to the minor groove of duplex DNA (Neidle, 2001; Dervan, 2001). These compounds, of which the best known are the natural products distamycin and netropsin or the synthetic bisbenzimidazole Hoechst 33258, are generally selective for A/T-rich sequences, to which they bind in a 1:1 mode (Kopka *et al.*, 1985; Abu-Daya *et al.*, 1995). The molecules make close contacts with the narrow minor groove that is typically found in A/T-rich sequences; binding to G/C-containing sequences is hindered by their wider groove and by steric clash with the 2-amino group of guanine.

A large number of modifications to these simple structures have been attempted in order to change the sequence specificity, but in general these have resulted in a relaxed specificity, with the inclusion of other sites as well as A/T-tracts (Lown *et al.*, 1986). However, the observation that distamycin can bind to some sequences in an antiparallel side-by-side 2:1 mode (Pelton *et al.*, 1989) led to the development of the hairpin pyrrole/imidazole polyamides as specific sequence-reading agents. These compounds recognise sequences by the side-by-side pairing of N-methylpyrrole and N-methylimidazole groups in the DNA minor groove. The simple code recognizes CG with a pyrrole/imidazole (Py/Im) pair, Im/Py targets GC and the Py/Py pair binds to both AT and TA (Dervan *et al.*, 2005; Wemmer *et al.*, 2000; Wemmer *et al.*, 1997). Although these compounds are able to recognise specific DNA sequences, they are large molecules that often exhibit difficulty in trafficking to the nucleus of cells and are currently limited to targeting a maximum of ten base pairs of DNA due to the ligand becoming out of phase with the DNA (Dervan, 2001).

Novel distamycin analogues have therefore been developed with heterocyclic rings of different shapes and sizes so as to produce molecules with improved potential for following the contours of the minor groove (Anthony *et al.*, 2004a; James *et al.*, 2004; Khalaf *et al.*, 2004; Anthony *et al.*, 2004b). One method that has been attempted is to use smaller molecules that can dimerise in the minor groove, but to control the ways in which they can overlap by including bulky substituents (such as isopropyl-thiazole). The size of this group prevents side-by-side stacking of another molecule within the minor groove, thereby forcing the two half halves of the dimer to be staggered. This staggered overlap also increases the ligand binding site size. The first such ligand to be investigated was Thiazotropsin A (Figure 6.2A), in which one of the N-methylpyrroles is substituted by an

isopropyl-substituted thiazole (*i.e.* Py-Py-Th). The thiazole group would be expected to be selective for GC-base pairs (like imidazole), but the bulky isopropyl groups prevent side-by-side stacking with N-methylpyrrole. On the basis of this model it was predicted that this compound should be selective for the sequence ACTAGT (shown in Figure 6.1), a prediction that was confirmed by DNase I and hydroxyl radical footprinting (Anthony *et al.*, 2004b; James *et al.*, 2004).

In the present study, the ligands Thiazotropsin B (Figure 6.2B) and Thiazotropsin C (Figure 6.2C) (derivatives of the parent compound) have been investigated and compared with Thiazotropsin A.

Thiazotropsin B has one of the N-methylpyrrole groups of Thiazotropsin A replaced with N-methylimidazole, as well as a slightly altered N-terminus, from a formyl to a acetyl group. By comparison with the Dervan rules for sequence recognition and allowing for the staggered side-by-side binding of these compounds, it can be predicted that this compound should bind to the sequence (A/T)CGCG(A/T) (shown in Figure 6.1).

Thiazotropsin C has the N-methylimidazole of Thiazotropsin B replaced by another isopropyl-substituted thiazole. Using conventional Dervan binding rules, treating the isopropyl-thiazole as binding like an imidazole moiety, this ligand should also bind to the sequence (A/T)CGCG(A/T). However, the inability of both isopropyl-thiazoles to stack opposite pyrrole rings will create a further problem and no selective binding is therefore predicted.

```
   Dp-Th-Py-Py-f                    Dp-Th-Py-Im-a
       f-Py-Py-Th-Dp                    a-Im-Py-Th-Dp
5'- W  C  W  W  G  W             5'- W  C  G  C  G  W
```

Figure 6.1: Proposed binding sites of Thiazotropsin A (left) and Thiazotropsin B (right). W is adenine or thymine; f is formyl; a is acetyl.

DNase I and hydroxyl radical footprinting and fluorescence melting experiments have been used to explore the sequence specificity of these compounds, with the results compared to those for Thiazotropsin A.

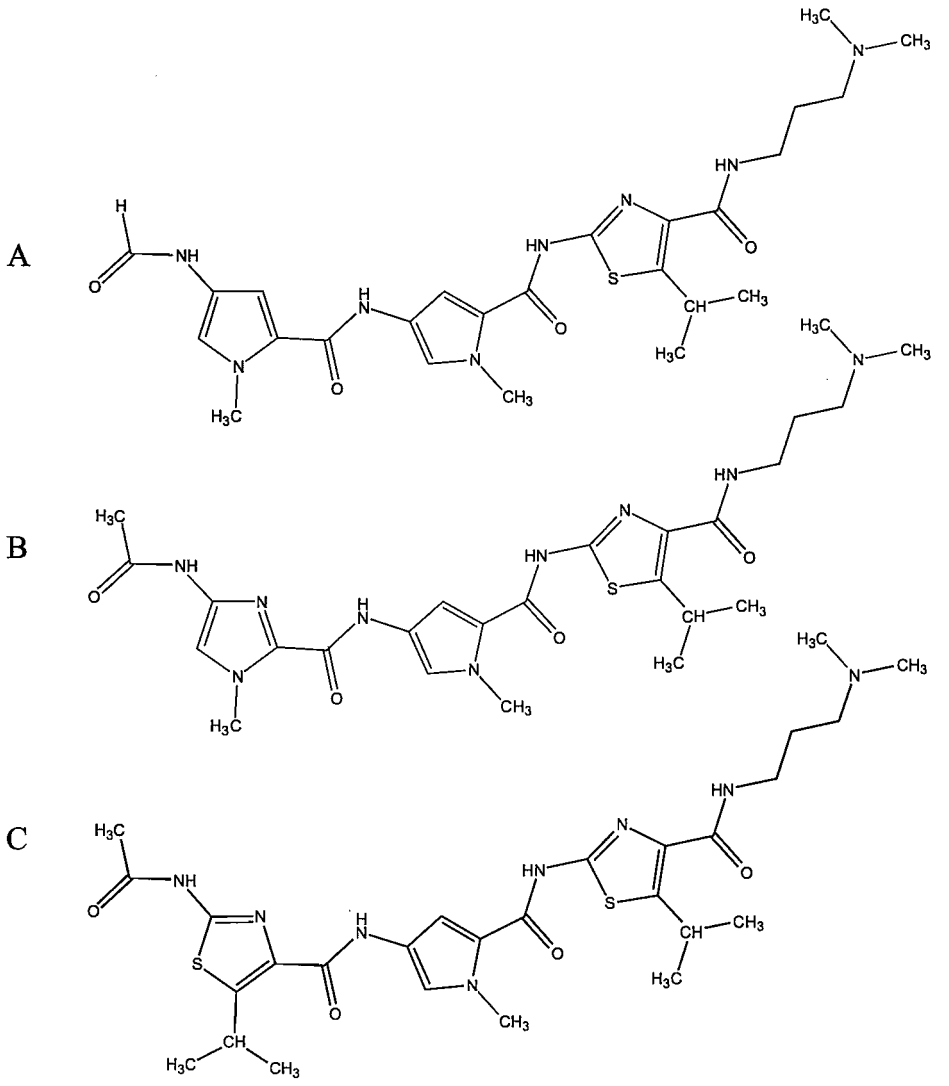


Figure 6.2: Structure of Thiazotropsin A (A), Thiazotropsin B (B) and Thiazotropsin C (C).

Results

The three ligands investigated in this chapter were initially studied with DNase I footprinting using universal tetranucleotide sequences MS1 and MS2 to establish their approximate binding selectivity and to compare them with the previously published results for Thiazotropsin A (James *et al.*, 2004). Hydroxyl radical footprinting was also utilised to better establish binding on these sequences. Thiazotropsin C showed no selective binding at this stage and so was not investigated further.

Thiazotropsin B was next tested with the footprinting substrates SASK1/SASK2 and STRATHA/STRATHB, which were designed to contain variants of its predicted

binding site. The results were confirmed by hydroxyl radical footprinting experiments and compared to the findings of experiments targeting the Thiazotropsin A ligand to STRATHA/STRATHB.

The selectivities of Thiazotropsins A and B were also examined using fluorescence melting experiments, using nine synthetic DNA duplexes.

MS1/MS2

Figure 6.3 shows the results of DNase I footprinting experiments with Thiazotropsins A, B and C on fragments MS1 and MS2. The left panel shows a direct comparison between Thiazotropsins A and B at a single concentration of each ligand, from which it can be seen that the two ligands produce different footprinting patterns. The right hand panels show the interaction of Thiazotropsins B and C with MS1 and MS2 over a range of concentrations. Figure 6.4 shows the sequence of the MS1/MS2 DNA fragment with observed binding sites highlighted, together with an example of footprinting plots derived from the binding of Thiazotropsin B to MS2. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 6.1.

It can be seen that 10 μM Thiazotropsin A produces a clear footprint at ACTAGT (marked with "*" on gels) as well as at all the A/T-tracts that are greater than 3 bp in length (shown as "+" on gels). The footprints produced by Thiazotropsin B are different and are at three sites comprising tracts of guanine and cytosine bases of variable lengths and composition. Some binding to A/T-tracts is also evident (marked "+" on gels). It is interesting to note that there is a footprint at TTTT on MS1 (site 2), while there is no protection at AAAC in the equivalent position on MS2 (the sequences are different at this position due to a single point mutation). Although MS1 and MS2 do not contain the proposed binding site for Thiazotropsin B (ACGCGT), one of the best footprints is located at GCGCGA (site 1; see Table 6.1), which differs from this by only one base pair at one end of the site.

In contrast, Thiazotropsin C does not produce any clear footprints, but causes a general non-specific inhibition of DNase I cleavage at concentrations of 25 μM and above. This ligand was not investigated further.

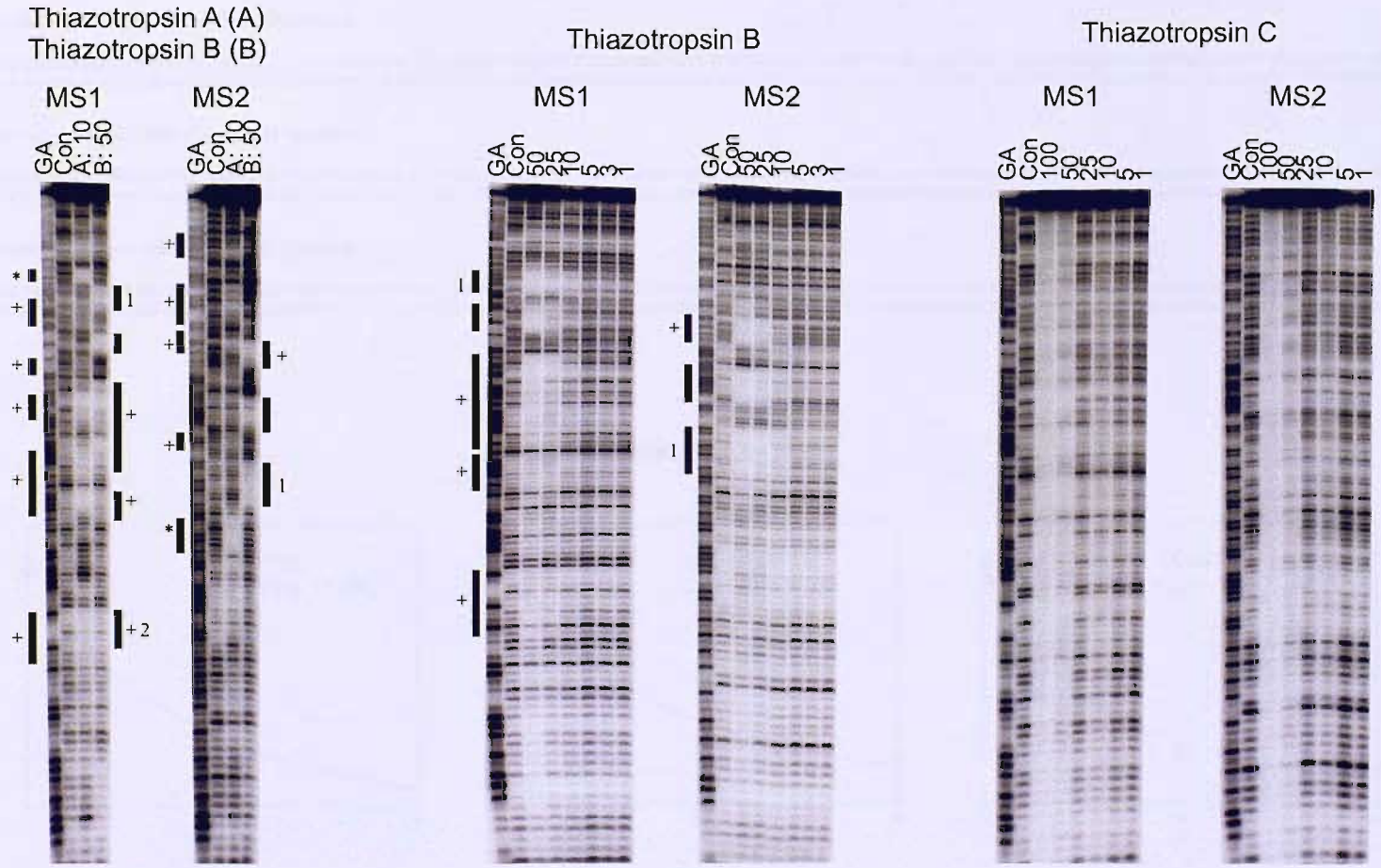


Figure 6.3: DNase I footprinting gels for Thiazotropsins A and B with the MS1 and MS2 DNA fragments. GA is a marker lane specific for purines. Con is a control lane. The ligand concentrations (μM) are shown at the top of each gel lane. The footprints marked relate to the sequences in Figure 6.4. The footprint marked with "*" is ACTAGT, A/T-tracts are marked "+", whilst other sites are numbered as mentioned in the text.

Thiazotropsin A (single concentration): MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGTCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

Thiazotropsin B (single concentration): MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGTCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

Thiazotropsin B (concentration range): MS1 (top); MS2 (bottom)

```
5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGTCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'
```

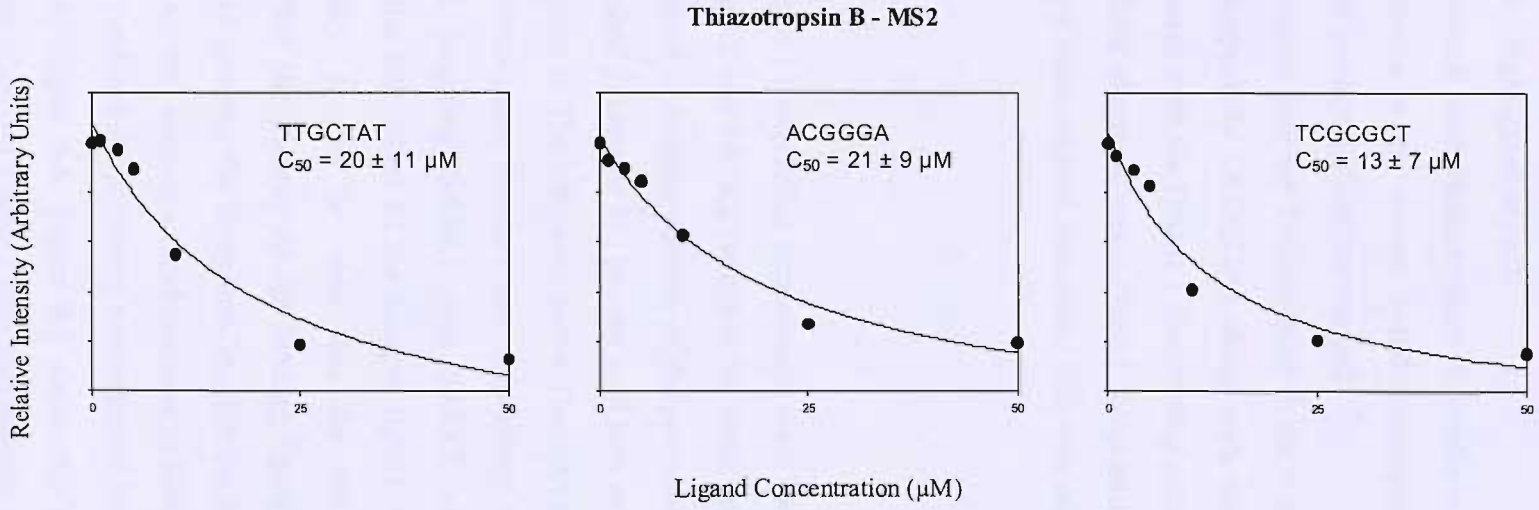


Figure 6.4: [Top] Sequences of fragments MS1 and MS2 showing the footprints produced by Thiazotropsins A and B, derived from Figure 6.3. [Bottom] Examples of footprinting plots derived from analysis of the gels shown in Figure 6.3.

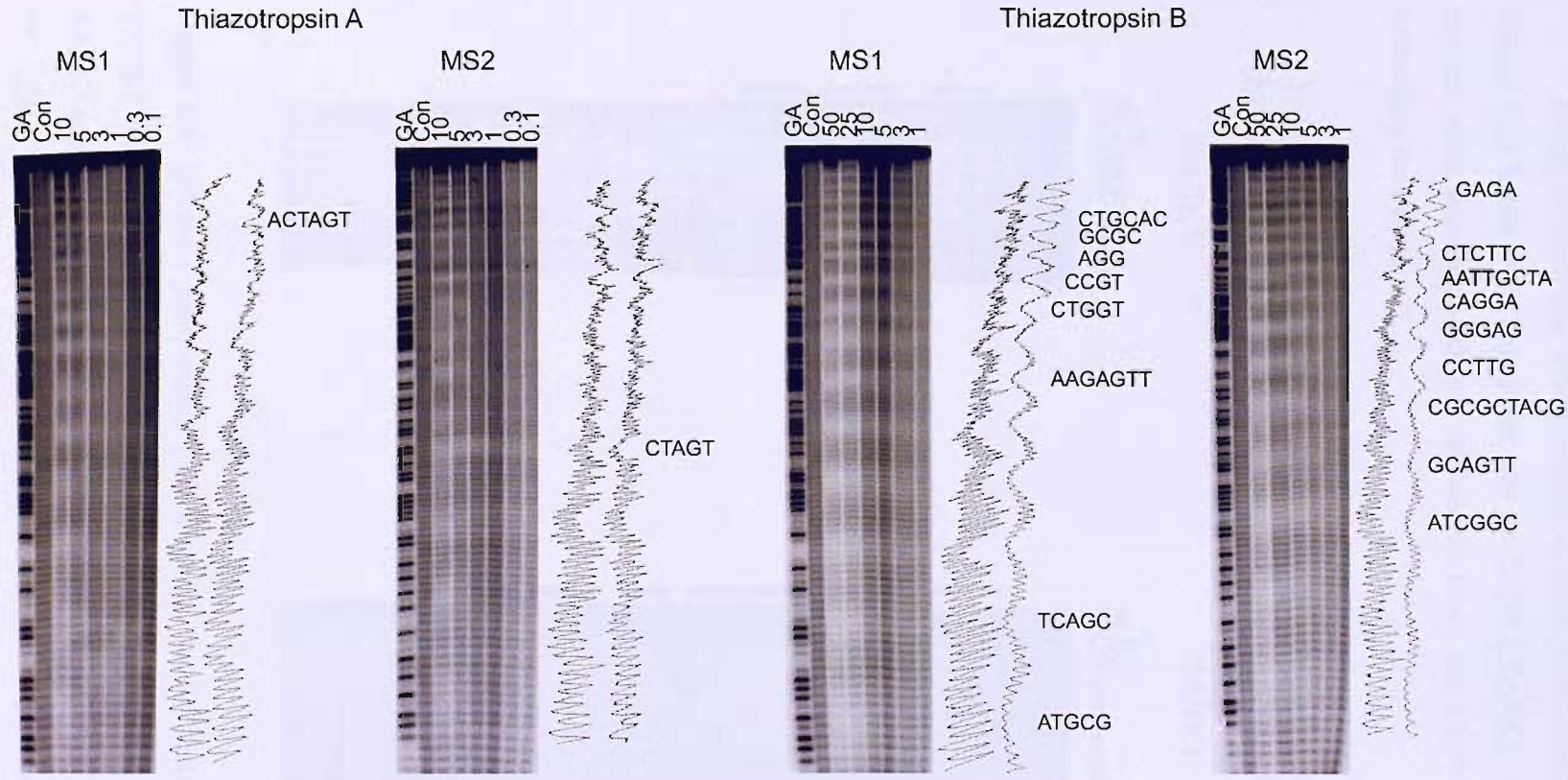
Figure 6.5 shows hydroxyl radical footprinting patterns for Thiazotropsins A and B on MS1 and MS2. Densitometer plots of the control lane and a ligand-treated lane are shown on the right of each gel, and the clearest footprinting sites are marked by the sequences of these sites. The binding sites for Thiazotropsin B are also summarised on the sequence of MS1/MS2 shown beneath these gels (highlighted in red).

Inspection of Figure 6.5 reveals that Thiazotropsin A produces only one clear footprint, at ACTAGT, which persists to the lowest ligand concentration tested. The footprints in the A/T-tracts are much weaker and are not marked.

The hydroxyl radical footprinting gels for Thiazotropsin B show a clear protection at the sites seen with DNase I (especially GCGCGA) along with several regions of protection within G/C-tracts, consistent with the DNase I footprinting experiments (Figure 6.3). Weak footprints other than those observed with DNase I footprinting are also seen, mainly in the A/T-tracts, indicating a more relaxed selectivity than was observed in Figure 6.3.

SASK1/SASK2

Figure 6.6 shows the results of DNase I footprinting experiments with Thiazotropsin B on SASK1 and SASK2 fragments. SASK1 and SASK2 (which is the same sequence cloned in the opposite orientation) were designed to contain variants of the proposed $(G/C)_4$ binding sites for the Series B ligands (described in chapter 5) , but are used here as they contain the proposed binding site for Thiazotropsin B. The left panel shows Thiazotropsin B binding to SASK1 over a range of concentrations using standard microcentrifuge tubes, whilst the right panels are of the ligand targeting SASK1 and SASK2 using siliconised microcentrifuge tubes, to prevent the interaction of the charged ligand with the tube, so affecting apparent binding affinity. It can be seen that the use of siliconised microcentrifuge tubes does not affect the identity of the binding ligand sites, but does reduce the concentrations required to generate the footprints. In addition, footprinting plots with the non-siliconised tubes (not shown) showed a rapid decrease in footprint intensity at lower drug concentrations, which is evident in the abrupt concentration dependence of the footprints in the left hand panel of Figure 6.6. Figure 6.7 shows the sequence of the SASK1/SASK2 DNA fragment with observed binding sites highlighted, together with an example of footprinting plots derived from the binding of Thiazotropsin B to SASK1.



Thiazotropsin B: MS1 (top); MS2 (bottom)

5'-GGATCCATATGCGGCAATACACATGGCAGATTTCCAACTGCACTAGTCGTAGCGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAGAGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGCGGATCC-3'
3'-CCTAGGTATACGCCGTTATGTGTACCGGCTAAAGGTTGACGTGATCAGCATCGCGCTAGTTCCAATTCGAGGGCAAGATAGGACCATATCGTTAATCCCGCACTTCTCAATACATTTCATGCAGGCCACCCCAGACCAAACAGTAGAGTCGGAGCTTACGCCTAGG-5'

Figure 6.5: Hydroxyl radical footprinting experiments showing the binding of Thiazotropsins A and B to MS1 and MS2. Densitometer plots are shown alongside each panel for the control and ligand treated lanes (0.1 μM for Thiazotropsin A and 25 μM for Thiazotropsin B) along with the bound sequences. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane. The sequence shown below indicates the binding sites for Thiazotropsin B.

The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 6.1.

It can be seen that Thiazotropsin B binds best to ACGCGT (marked with "*" in Figure 6.6) better than any other site on SASK1/SASK2. This difference is especially evident when working with the siliconised tubes for which this site is about 100-times better than the others.

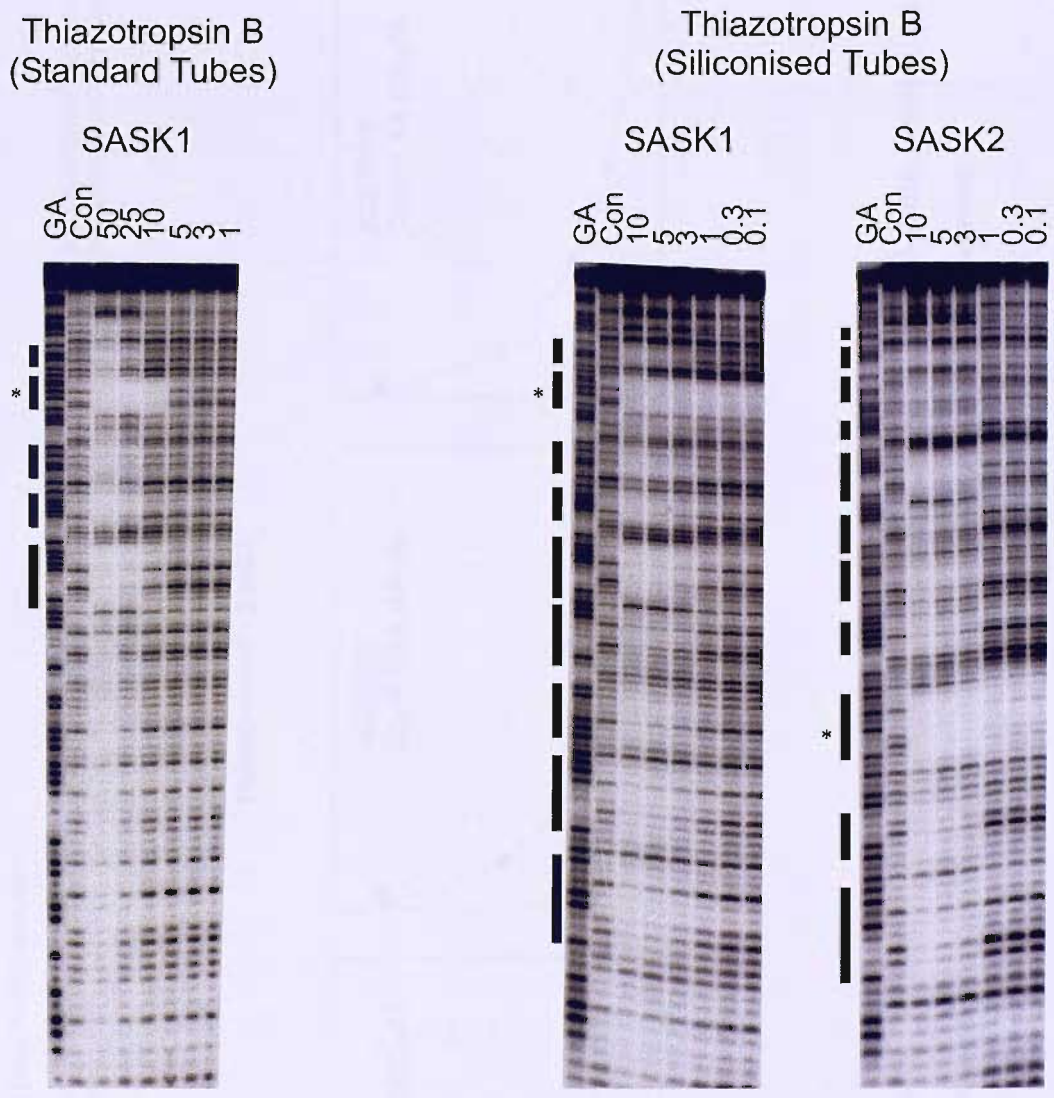


Figure 6.6: DNase I footprinting gels showing the interaction of Thiazotropsin B with SASK1 and SASK2. [Left] SASK1 tested using standard tubes; [Right] SASK1 and SASK2 tested using siliconised tubes. The preferred ACGCGT binding site is marked with "*". GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane. The footprints marked relate to the sequences in Figure 6.7.

Thiazotropsin B (siliconised tubes): SASK1 (top); SASK2 (bottom)

5'-GGATCCAGCAAGCGCTTGCTAGGCCATGCAACGCGTTGCAAGCCCTTGCATGGCCATGCAAGACCTTGCAAGGCCTTGCAAGTCCTTGCTTGGCCAAGCAAGATCTTGCAACCGGTTGCCACGGATCC-3'
3'-CCTAGGTCGTTCGCGAACGATCCGGTACGTTGCGCAACGTTCGGGAACGTACCGGTACGTTCTGGAACGTTCCGGAACGTTCAGGAACGAACCGGTTCGTTCTAGAACGTTGGCCAACGGTGCCTAGG-5'

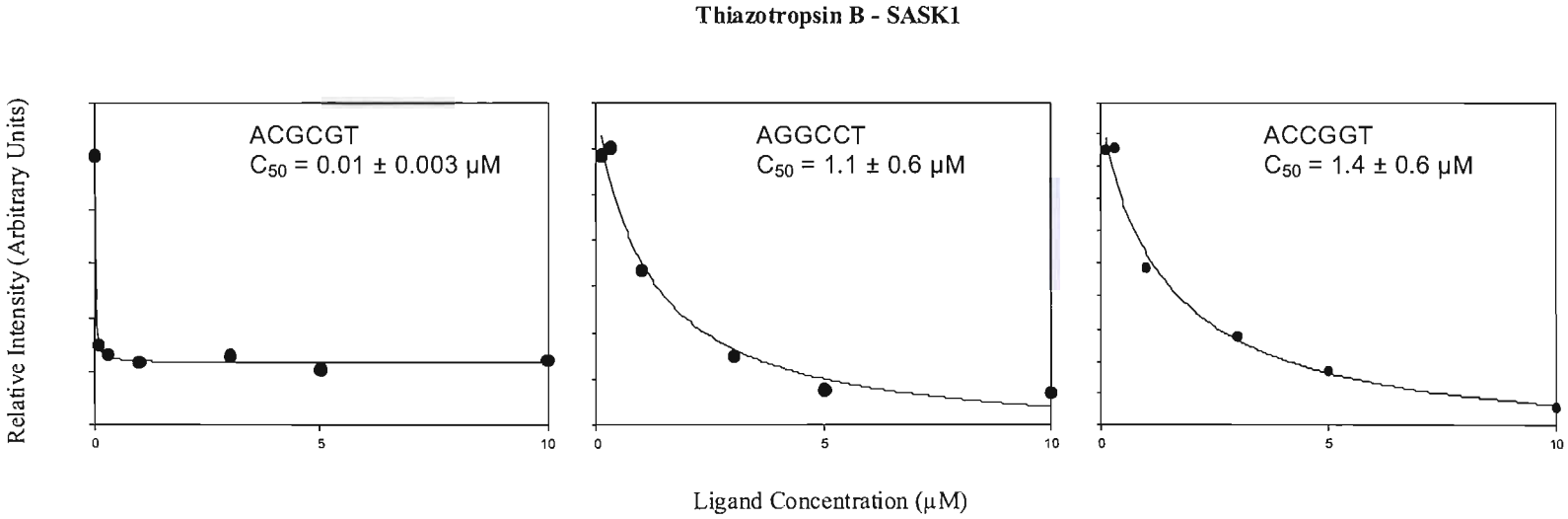


Figure 6.7: [Top] Binding sites for Thiazotropsin B on SASK1 and SASK2 using siliconised tubes (from Figure 6.6). [Above] Example of footprinting plots for the interaction of Thiazotropsin B with binding sites in the SASK1 DNA fragment.

STRATHA/STRATHB

Figure 6.8 shows the interaction of Thiazotropsins A and B with the STRATHA and STRATHB DNA fragments over a range of ligand concentrations. STRATHA and STRATHB are complementary sequences containing several variations of the proposed target site for Thiazatropsin B [(A/T)CGCG(T/A)]. The left two panels show DNase I footprints for Thiazotropsin A on STRATHA and STRATHB, and the right two panels show the effect of Thiazotropsin B on STRATHA and STRATHB. Figure 6.9 shows the sequence of the STRATHA/STRATHB DNA fragment with the binding sites of Thiazotropsins A and B highlighted, together with examples of footprinting plots derived from the binding of Thiazotropsin B to STRATHB. The C_{50} values for the interaction with all the sites on these fragments are summarised in Table 6.1.

Thiazotropsin A appears to bind to G/C-tracts with surrounding adenine and thymine bases (as shown in Table 6.1 and the unlabelled footprints of the gels). The C_{50} of binding to these tracts is about 1 μM, with the exception of the site at the 5'-end of STRATHA (site 1; GGATCC in Table 6.1), which displays a C_{50} of about 0.01 μM. The footprint here actually encroaches into the pUC19 polylinker, where a TCTAGA sequence is directly adjacent to the BamH1 site (*i.e.* TCTAGAGGATCC). This is the XbaI restriction site in pUC19 and is the preferred binding motif, so explaining the very strong binding affinity.

The footprinting pattern for Thiazotropsin B is very different and ACGCGT (site 4) is the preferred binding site for this ligand. Thiazotropsin B also binds well to ACGCGA (site 2) but the C_{50} value is about 3 times higher than to ACGCGT. TCGCGA (site 3) is the weakest site for Thiazotropsin B and the C_{50} value is 10 times greater than at ACGCGA. This may indicate that flanking purine/pyrimidine steps are important for binding, as the affinity is weaker with decreasing numbers of alternating purine/pyrimidine steps.

Figure 6.10 shows hydroxyl radical footprinting patterns for Thiazotropsin B on STRATHA and STRATHB. Densitometer plots of the control lane and a ligand-treated lane are shown on the right of each gel, and the clearest footprinting sites are marked by the sequences of these sites. The binding sites are also summarised on the sequence of STRATHA/STRATHB shown beneath these gels (highlighted in red).

It can be seen that hydroxyl radical cleavage is attenuated by Thiazotropsin B at every G/C-tract on STRATHA/STRATHB, producing more footprints than were evident

with DNase I. This is similar to the effects seen with both Thiazotropsins A and B with MS1 and MS2, in which hydroxyl radical footprinting showed more binding sites than DNase I. Visual inspection of the gels shows stronger binding to alternating G/C motifs than to, for example, GGG.

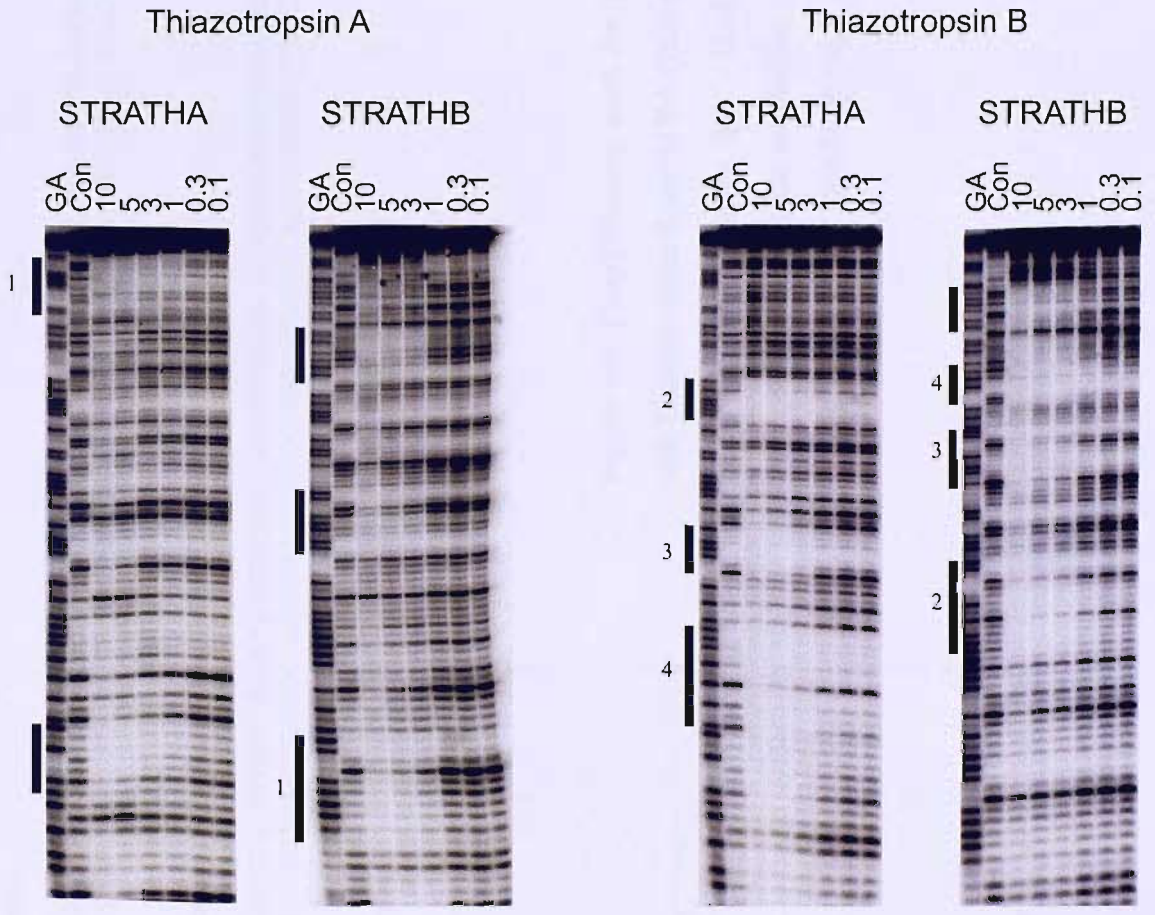


Figure 6.8: DNase I footprinting gels for Thiazotropsin A and Thiazotropsin B on the STRATHA and STRATHB DNA fragments. Only the clearest footprints are marked. The 3'-end of the sequence is missing with STRATHA as the fragment is labelled closer to the insert. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane. The footprints marked relate to the sequences in Figure 6.9 and are numbered as mentioned in the text.

Thiazotropsin A: STRATHA (top); STRATHB (bottom)

```
5'-GGATCCTTAATCCGCATAACCCCTTTACGCGAAATCGCCTAATCGGGATATCGCGAATTCGCGCAAAACGCGTATAGCGTTAAGCGCAATACGCATACCGGATCC-3'
3'-CCTAGGAATTAGGCGTATTGGGGAAATGCGCTTTAGCGGATTAGCCCTATAGCGCTTAAGCGCGTTTTGCGCATATCGCAATTCGCGTTATGCGTATGGCCTAGG-5'
```

Thiazotropsin B: STRATHA (top); STRATHB (bottom)

```
5'-GGATCCTTAATCCGCATAACCCCTTTACGCGAAATCGCCTAATCGGGATATCGCGAATTCGCGCAAAACGCGTATAGCGTTAAGCGCAATACGCATACCGGATCC-3'
3'-CCTAGGAATTAGGCGTATTGGGGAAATGCGCTTTAGCGGATTAGCCCTATAGCGCTTAAGCGCGTTTTGCGCATATCGCAATTCGCGTTATGCGTATGGCCTAGG-5'
```

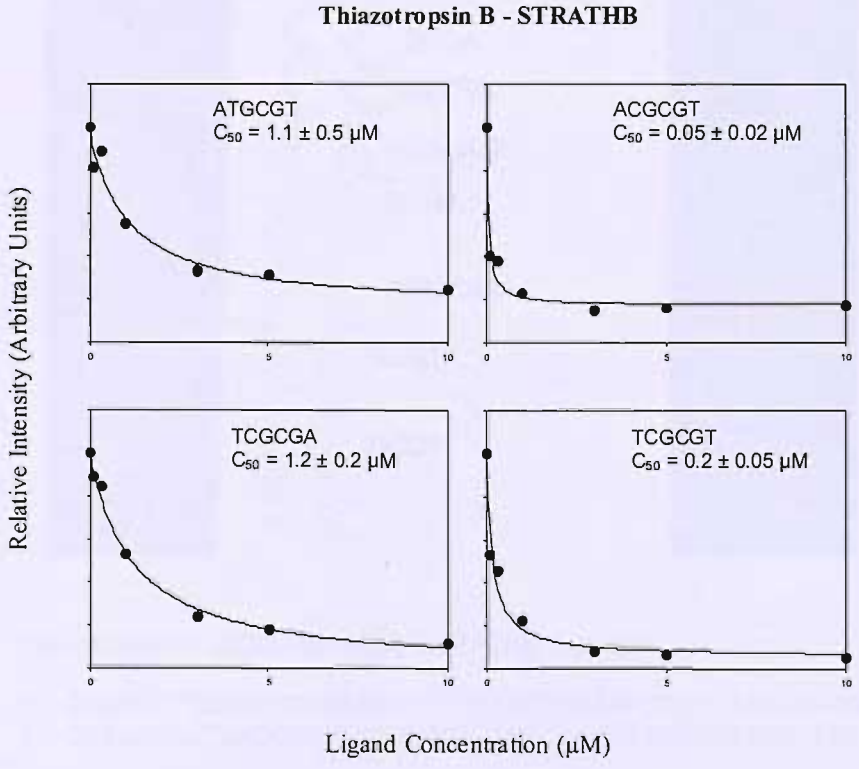


Figure 6.9: [Top] Binding sites for Thiazotropsin A and Thiazotropsin B on STRATHA and STRATHB, as derived from Figure 6.8. [Left] Examples of footprinting plots for the interaction of Thiazotropsin B with sites on the STRATHB DNA fragment.

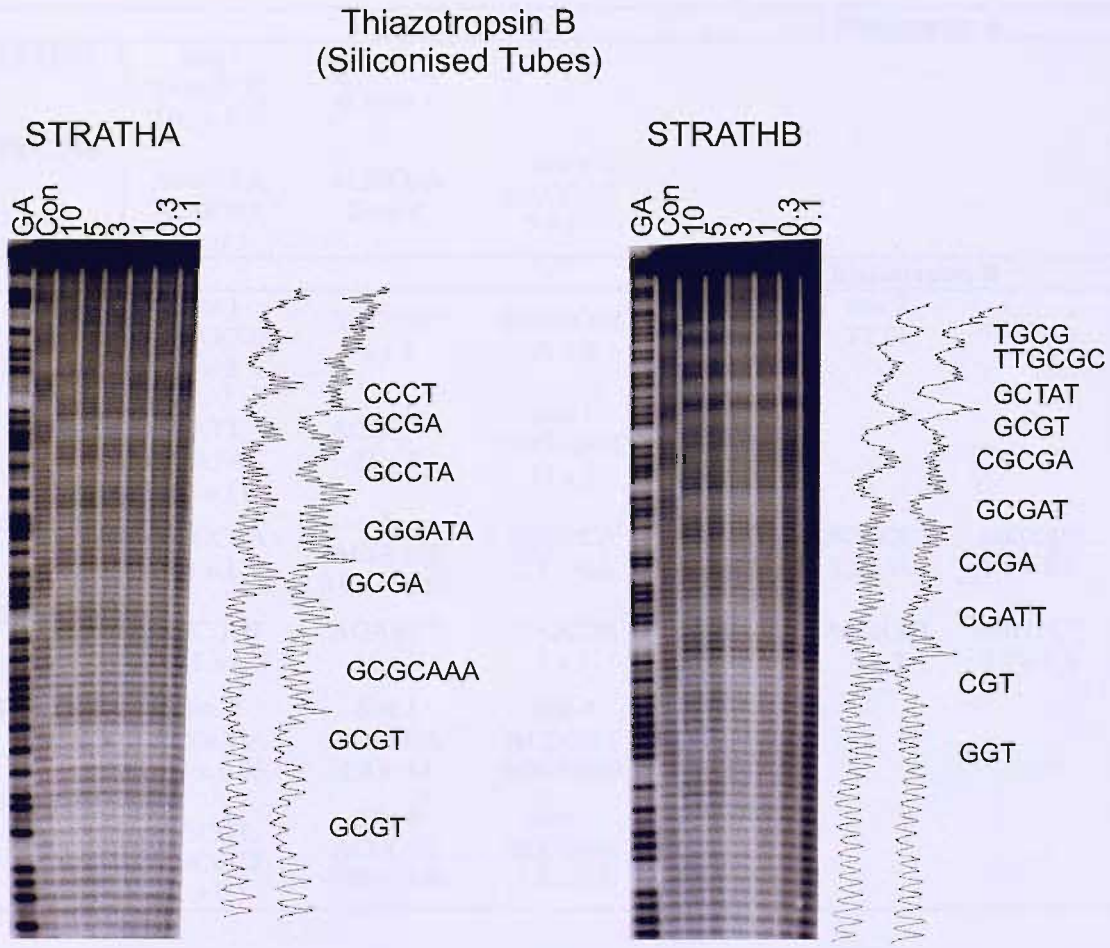


Figure 6.10: [Left] Hydroxyl radical footprinting gels for the interaction of Thiazotropsin B with STRATHA and STRATHB. The clearest footprints are indicated by the sequences. Densitometer plots are shown alongside the gels and were derived from the 3 μM lanes. GA is a marker lane specific for purines. Con is a control lane. Ligand concentrations (μM) are shown at the top of each gel lane.

[Below] Binding sites for Thiazotropsin B on STRATHA and STRATHB, derived from the hydroxyl radical gels.

Thiazotropsin B: STRATHA (top); STRATHB (bottom)

5'-GGATCCTTAATCCGCATAACCCCTTTACGCGAAATCGCCTAATCGGGATATCGCGAATTCGCGCAAAACGCGTATAGCGTTAAGCGCAATACGCATACCGGATCC-3'
3'-CCTAGGAATTAGGCGTATTGGGGAAATGCGCTTTAGCGGATTAGCCCTATAGCGCTTAAGCGCGTTTTGCGCATATCGCAATTCGCGTTATGCGTATGGCCTAGG-5'

Thiazotropsin A											
STRATHA	Site 1 GGATCC 0.01 ± 0.01	GCGTTA 0.7 ± 0.3									
STRATHB	ACGCTA, ACGCGT 1 ± 0.5	AGGCGA 2 ± 0.8	Site 1 AAGGAT 2 ± 1.7								
Thiazotropsin B											
MS1	Site 1 AGCGCGA 4 ± 2	TCCCGT 6 ± 3	AGGGCGT 13 ± 6	TAAA -	Site 2 TTTT -						
MS2	AATT, TATA 20 ± 11	ACGGGA 21 ± 9	Site 1 TCGCGCT 13 ± 7								
SASK1	AGGCCA 4 ± 1	* ACGCGT 0.01 ± 0.003	TGGCCA 2.5 ± 0.6	AGACCT 2 ± 0.5	AGGCCT 1.2 ± 0.6	AGTCCT 1.5 ± 0.7	TGGCCA 1.6 ± 0.6	AGATCT 1.4 ± 0.5	ACCGGT 1.4 ± 0.6		
SASK2	ACCGGT 21 ± 6	AGATCT 11 ± 2	TGGCCA 5 ± 1	AGGACT 2.7 ± 0.8	AGGCCT 4 ± 1	AGGTCT 2.7 ± 0.5	TGGCCA 1.8 ± 0.5	AGGGC 1.4 ± 0.5	* ACGCGT 0.1 ± 0.05	TGGCCT 1.5 ± 0.7	AGCGCT 1.4 ± 0.6
STRATHA	Site 2 ACGCGA 0.1 ± 0.06	Site 3 TCGCGA 1.4 ± 0.4	Site 4 ACGCGT 0.04 ± 0.02								
STRATHB	TGCGT, TGCGCT 1 ± 0.5	Site 4 ACGCGT 0.05 ± 0.02	Site 3 TCGCGA 1.2 ± 0.2	Site 2 TCGCGT 0.2 ± 0.05							

Table 6.1: C_{50} values (μM) derived from quantitative analysis of the gels presented in this chapter (with the standard errors shown). The binding sites are presented left to right in the order that they run from the top of each gel to the bottom (5'-3'). Sites marked "–" correspond to footprints that are evident on visual inspection, but for which quantitative analysis did not provide a clear value.

Fluorescence Melting

Figure 6.11 shows fluorescence melting curves for the effects of Thiazotropsins A and B on duplexes containing the sequences ACTAGT and ACGCGT (containing the suggested binding sites for Thiazotropsins A and B respectively). The changes in melting temperature produced by these ligands are summarised in Table 6.2 [Left]. It can be seen that Thiazotropsin A stabilises the ACTAGT duplex, producing ΔT_ms of 19 °C and 9.8 °C at 50 μM and 5 μM respectively. The other duplexes shown in Table 6.2 [Right] showed no stabilisation with Thiazotropsin A (James *et al.*, 2004). In contrast, Table 6.2 [Left] shows Thiazotropsin B stabilised ACGCGT more that ACTAGT though the ΔT_m values are lower than those produced by Thiazotropsin A with ACTAGT. For example, a ΔT_m of 9.8 °C is produced by 25 μM Thiazotropsin B against ACGCGT, whilst only 5 μM Thiazotropsin A stabilises its preferred sequence (ACTAGT) by the same amount. Table 6.2 [Right] shows the other duplexes tested showed no stabilisation with Thiazotropsin B.

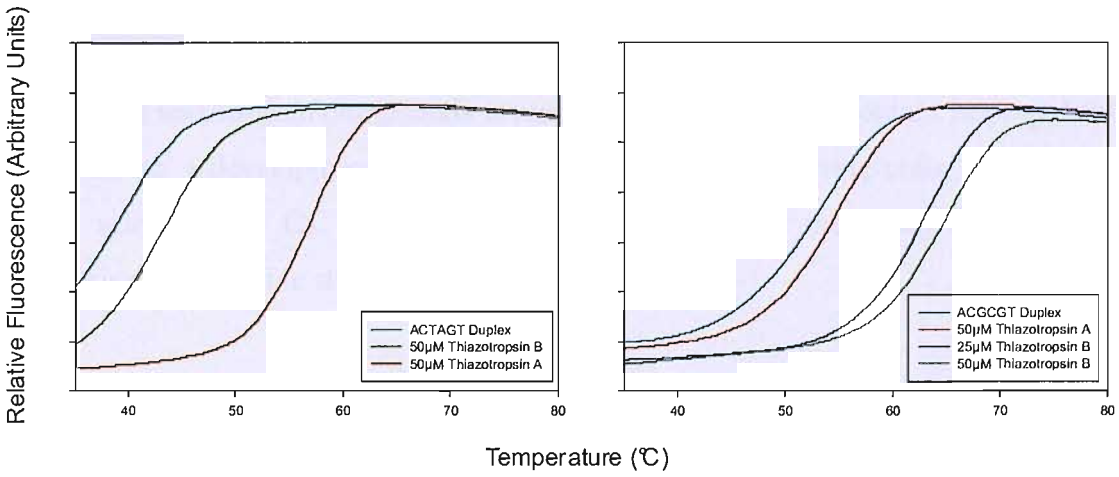


Figure 6.11: Fluorescence melting curves for sequences ACTAGT (left) and ACGCGT (right) in the presence of Thiazotropsin A and Thiazotropsin B. The free duplex is in cyan; 50 μM Thiazotropsin A is red; 25 μM Thiazotropsin B is navy; 50μM Thiazotropsin B is green.

	T_m (°C)	ΔT_m (°C)				
		50 μM	25 μM	10 μM	5 μM	1 μM
Thiazotropsin A						
ACTAGT	39.3	19.3	16.6	13.2	9.8	0.6
ACGCGT	53.7	3.1	1.2	0.3	0.1	0.0
Thiazotropsin B						
ACTAGT	39.3	4.0	1.1	0.5	0.2	0.2
ACGCGT	53.7	12.0	9.8	5.4	1.6	0.1

	T_m (°C)	ΔT_m (°C)
		10 μM
TGATCA	39.5	-0.2
TGATCT	40.1	-0.4
TGTACT	38.8	0.5
TGAACT	39.5	-0.1
TGTAGT	37.8	-0.3
ATATAT	38.4	0.1
AAAAAG	38.2	-0.2

Table 6.2: [Left] ΔT_m values produced by different concentrations of Thiazotropsin A and Thiazotropsin B on the sequences ACTAGT and CGCGT. [Right] ΔT_m values produced by a single concentration of Thiazotropsin B on the sequences shown. The duplex concentration was 0.25 μM. The melting temperatures of the free DNA duplexes are also shown. Each value is the average of at least 2 experiments which typically differed by less than 0.5 °C.

Discussion

The three Thiazotropsin compounds investigated in this chapter contain isopropyl-thiazole ring systems, a development on the well-classified polyamide ring units, with the thiazole group selective for GC base pairs (like imidazole) and the bulky isopropyl groups preventing side-by-side stacking with a pyrrole ring.

Thiazotropsin A

Thiazotropsin A (Dp-Th-Py-Py-f) has previously been shown to bind to ACTAGT (James *et al.*, 2004). This is thought to occur in the 2:1 binding mode as shown over page in Figure 6.12 [left], with the isopropyl-thiazole selective for a guanine and the two monomers staggered so as to avoid overlap between the N-methylpyrrole and isopropyl-thiazole units. This site is protected in footprinting experiments at concentrations as low as 0.3 μM. However, if the two halves of the dimer slip further, with even less overlap, then the configuration shown on the right in Figure 6.12 would be produced. On the basis of

conventional binding rules, we would predict that this might bind to the sequence WCWWWGW (W = A or T).

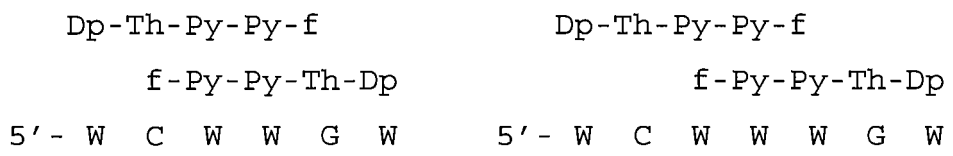


Figure 6.12: Possible binding sites for Thiazotropsin A.

The results with MS1/MS2 support the findings of previous work with this ligand and confirm that ACTAGT is the preferred site. A/T-tracts also seem to bind the ligand at 10 μM concentrations, as was noted in the previous study, possibly indicative of 1:1 binding. None of the footprints showed any evidence for the interaction WCWWWGW, suggesting that the 4 bp overlap shown on the left in Figure 6.12 is preferred. This could be because of the greater stabilisation afforded by the increased overlap or because the isopropyl-thiazole needs something to bind opposite (the formyl group in this case).

The footprinting experiments with STRATHA/STRATHB support the findings with MS1 and MS2, and the use of siliconised tubes increases the apparent binding affinity by 10-fold. The preferred site on these sequences appears to be TCTAGA. This suggests that the dimethylaminopropylamide at the C-terminus of the ligand is indiscriminate between adenine and thymine.

Fluorescence melting experiments support the binding preference for ACTAGT, which is clearly a much better binding site than ACGCGT (the preferred site for Thiazotropsin B). No stabilisation was seen with any of the other duplexes.

Thiazotropsin B

The only significant difference between Thiazotropsin A and Thiazotropsin B is the substitution of a carbon atom for a nitrogen atom in the third ring system to create an imidazole ring in place of a pyrrole ring. This change creates a different proposed 2:1 binding mode (Figure 6.13 over page).

```
        Dp-Th-Py-Im-a
             a-Im-Py-Th-Dp
5'- W  C  G  C  G  W
```

Figure 6.13: Proposed binding site for Thiazotropsin B.

The small change in Thiazotropsin B created a different binding selectivity, with the strongest site on MS1 being GCGCGA, which is similar to the proposed binding site shown in Figure 6.13. The change in selectivity is clearest when directly compared with Thiazotropsin A in Figure 6.3. The weaker binding to A/T-tracts is also interesting, with TTTT on MS1 exhibiting binding, whilst the complementary AAAC (due to the point mutation present in this sequence) is not clearly bound. This suggests that for the 1:1 binding three AT base pairs are not sufficient, whilst the ligand can bind to four successive AT pairs.

Although GCGCGA is the preferred site for Thiazotropsin B on MS1/MS2 (differing from the proposed binding site by just one base pair), the studies on SASK1/SASK2 show that AGGCCT (which differs from ACGCGT by two base pairs) also binds the ligand, but with twice the C_{50} value.

The use of the Thiazotropsin B specific sequence STRATHA/STRATHB has revealed some interesting features concerning the importance of the bases flanking the CGCG core. Binding to ACGCGT is about three times stronger than to ACGCGA, which in turn is about ten times stronger than to TCGCGA. This may indicate that purine/pyrimidine steps are important in binding, as affinity is weakened with decreasing numbers of alternating purine/pyrimidine steps. This is probably related to the local DNA structure which is thought to possess a wider minor groove, allowing stronger 2:1 binding.

The best binding site for Thiazotropsin A on STRATHA/STRATHB (ACTAGT) is about five times stronger than the best site for Thiazotropsin B. This is consistent with all the observations that Thiazotropsin B binds more weakly, as has been seen with other imidazole-containing polyamides targeting GCGC motifs (Dervan, 2001).

The fluorescence melting curves shown in Figure 6.11 and tabulated in Table 6.2 confirm that ACGCGT is the preferred target for Thiazotropsin B, though it binds to this less well than Thiazotropsin A does to its preferred sequence (ACTAGT). 5 μM Thiazotropsin A produced a ΔT_m value of 9.8 °C with ACTAGT, whilst 25 μM Thiazotropsin B was

required to produce a similar ΔT_m value with ACGCGT. This five fold difference is similar to that seen in the DNase I footprinting studies with STRATHA/STRATHB.

Interestingly, Thiazotropsin B seems also to stabilise the ACTAGT better than Thiazotropsin A stabilises ACGCGT. Although none of the other fluorescence melting duplexes showed stabilisation with either ligand, it is clear that Thiazotropsin B is less selective than Thiazotropsin A; a fact supported by footprinting studies and can only be attributable to the substitution of a pyrrole ring for an imidazole moiety.

Thiazotropsin C

Thiazotropsin C differs from Thiazotropsin B by the substitution of the imidazole ring for another isopropyl-thiazole. The following 2:1 binding mode (Figure 6.14) would therefore result.

```
        Dp-Th-Py-Th-a
             a-Th-Py-Th-Dp
5'- W  C  G  C  G  W
```

Figure 6.14: Proposed binding mode for Thiazotropsin C.

DNase I footprinting studies on the MS1/MS2 substrate only showed non-selective binding at 25 μM. The preferred sites for Thiazotropsins A and B are clearly not targeted by Thiazotropsin C. This non-selective binding suggests that the 2:1 binding mode is not permitted and that only 1:1 binding is evident at high ligand concentrations. A 2:1 binding mode is probably not favoured as the internal isopropyl-thiazole groups are too bulky to bind opposite other ring systems, though the terminal formyl or acetyl groups are small enough for 2:1 binding to occur opposite an isopropyl-thiazole in Thiazotropsins A and B.

General Conclusions

The results presented in this chapter support the findings of James *et al.*, (2004), concerning the sequence selectivity of Thiazotropsin A. Substitution of a further carbon atom by a nitrogen atom on just one of the ring moieties to create Thiazotropsin B has clearly had a large effect on the binding selectivity and affinity. By targeting guanine with

the new imidazole ring, the preferred binding site is changed from ACTAGT to ACGCGT, with the affinity for this new site about five times weaker than the parent compound.

Selective binding is abolished altogether when the imidazole in Thiazotropsin B is substituted for another isopropyl-thiazole, consistent with the suggestion that the large group is too bulky for 2:1 binding opposite another ring moiety. The Thiazotropsin C ligand can therefore only bind in the 1:1 mode at high concentrations.

Isopropyl-thiazole appears to be a useful monomer for increasing the binding site size of small molecule minor groove binding ligands. This bulky group forces staggering of the two partners in the 2:1 homodimer, thereby making the three ring system target four base pairs of DNA.

As with all the ligands presented in this thesis, NMR and X-ray crystallography studies with appropriate oligonucleotides will be necessary to confirm their proposed selectivity and binding mode. Although NMR studies have been carried out on Thiazotropsin A (Anthony *et al.*, 2004b), structural investigations on the other ligands would be useful, establishing if the bulky isopropyl-thiazole inhibits 2:1 binding against ring systems, as proposed in this chapter, yet still remaining permissive against the N-terminal domains of the ligands (formyl or acetyl).

A range of potential new ligands could be synthesised on the basis of these results. Of particular interest would be a series of different length ligands containing isopropyl-thiazole ring systems. This would give an insight into how isopropyl-thiazole affects the dimerisation and binding of long polyamides. As with all polyamides, the potential for new ring systems is large and variations on the isopropyl-thiazole moiety may also lead to stronger binding derivatives, with possibly different selectivity depending on the nature of the alteration.

The results in this chapter indicate that although an isopropyl-thiazole creates a stagger in a 2:1 homodimer, a small moiety is permitted to bind opposite it in the minor groove. Variations on this group may reveal an effect of the moiety (formyl and acetyl tested in this work) on binding strength and affinity.

Chapter VII
General Conclusions

General Conclusions

The aim of the work described in this thesis was to use DNA footprinting and fluorescence melting techniques to assess the selectivity and affinity of DNA-binding small molecules. Novel DNA footprinting substrates were prepared and these, along with other established fragments, were used to assess the sequence selectivity of novel minor groove ligands and to further probe the preferred binding site(s) of some well-known and novel ligands.

Chapter III: Novel universal footprinting substrates

Universal oligonucleotide substrates are useful for systematic footprinting studies. Their use in investigations on novel DNA-binding ligands gives an unbiased method for approximating sequence selectivity and affinity.

In this research, DNA fragments containing every dinucleotide, trinucleotide and symmetrical hexanucleotide were designed and prepared for footprinting studies. These substrates were intended to be of minimal length, while avoiding inappropriate regions (*e.g.* long G/C-tracts or long homopurine.homopyrimidine tracts). The usefulness of these fragments was demonstrated by performing DNase I and hydroxyl footprinting experiments with the well-characterised ligands: Distamycin; Hoechst 33258; mithramycin; nogalamycin; actinomycin D; and echinomycin. The results of these studies are consistent with previous knowledge, but expand our understanding of the fine details of their sequence selectivity.

Distamycin: The results confirm that this AT-selective ligand prefers ApT steps over TpA, with AATT as the best binding site. It is interesting to note that $A_n.T_n$ was only bound well in AT-tracts that were longer than three base pairs. A/T-tracts of six base pairs length were preferred over shorter four base pairs tracts, though hydroxyl radical footprinting showed that the ligand does not protect the whole of these long tracts. This indicates that distamycin binds best when a narrow minor groove surrounds its primary binding site.

Hoechst 33258: AATT is the preferred binding site for Hoechst 33258, and TpA steps are particularly bad for binding, more so than with distamycin. Interestingly, $A_n.T_n$ tracts are bound strongly at any length (unlike with distamycin). This might be simply because Hoechst 33258 has a smaller binding site, but may suggest that it prefers a more regular DNA structure than distamycin.

Mithramycin: As expected this ligand binds especially well to long G/C-tracts. However, the affinity for shorter tracts is enhanced if these are flanked by a neighbouring TpA step (Sastry *et al.*, 1995), while ApT steps seem to inhibit binding. This suggests that an uneven DNA structure is favoured for binding by mithramycin.

Nogalamycin: Purine/pyrimidine steps around a G/C-tract seem to produce the best nogalamycin binding sites, but the inclusion of all four bases also seems to be important. A mixed DNA sequence, with a dynamic DNA structure, is preferred for the threading intercalation binding of this ligand.

Actinomycin D: As expected, GpC is the preferred binding motif for this ligand, but adjacent sites attenuated the binding up to two or three base pairs away. This may be due to direct interaction with the cyclic peptides of actinomycin, but may also reflect differences in local DNA structure. Purine/pyrimidine steps produce the strongest binding sites.

Echinomycin: As observed by Wang *et al.* (1984) and Gao and Patel (1988), ACGT is the preferred binding site for echinomycin (in the NMR structures, adenines flanking the CpG core are arranged in the *syn* conformation). As with actinomycin, abutting binding sites show weak binding and the DNA sequence two base pairs from the central CpG affects the affinity.

The novel fragments, especially those with hexanucleotide repeats, are clearly useful DNA footprinting fragments and are ready for use to test the sequence selectivity of novel ligands.

Chapter IV: Novel analogues of the bis-intercalator TANDEM

TANDEM is one of a group of ligands, both natural and synthetic, that bind to DNA by bis-intercalation (Dawson *et al.*, 2007). TANDEM is a synthetic analogue of triostin A that lacks the N-methyl groups, which are present on four of the peptide bonds in the parent compound. This small alteration results in a dramatic change in sequence selectivity and TANDEM binds selectively to TpA steps, particularly ATAT (Lavesa *et al.*, 1993; Lavesa

and Fox, 2001). This occurs because the alanine carbonyl groups form internal hydrogen bonds to the previously methylated amide groups, so removing the ability to bind to CpG.

This study used TANDEM that had been prepared by novel solid-phase synthesis, which was shown to have identical properties (both affinity and selectivity) to solution-phase TANDEM (Waterloh *et al.*, 1992; Lavesa *et al.*, 1993; Lavesa and Fox, 2001). Analogues were also tested to study the effect minor changes have on binding selectivity and affinity.

Studies with novel TANDEM derivatives give new information on the important features of this ligand. As previously suggested (Fox *et al.*, 1980a; Lee and Waring, 1978), an intact cross-bridge is essential for binding, probably because this constrains the conformation. The chromophores are also essential for binding and the peptide backbone alone does not bind to DNA. Surprisingly, replacement of the quinoxaline chromophores with other similar aromatic rings, such as naphthylene has a pronounced effect on the affinity.

The most interesting observation is that replacement of the hydrophobic valine chains on residues 4 and 8 with lysines strengthens binding affinity 30-40 fold while not sacrificing selectivity. This is probably due to the favourable electrostatic interaction between this basic amino acid and the DNA phosphodiester backbone environment created between the positive charge of the lysines and the negative phosphodiester backbone of the DNA. Although the compound still does not possess any antibacterial activity, it is possible that a similar alteration to echinomycin and triostin A will greatly improve their biological activity.

Chapter V: Polyamide and Hoechst-derived minor groove binding ligands

Polyamides are a group of ligands mainly comprising pyrrole and imidazole rings, which form selective interactions with individual DNA bases in the minor groove. Long polymers of these rings have been designed (Dervan, 2001), but in general it has been found that these come out of phase with the DNA over long repeats (Kelly *et al.*, 1996), so limiting the number of bases that can be targeted selectively. The work in this thesis was therefore important for developing novel compounds by examining different ring moieties so as to improve binding and selectivity. To this end, derivatives of Hoechst 33258 (a minor groove binding ligand using benzimidazole rings to target the DNA) and standard polyamides have been developed to study the effects of the introduction of: Benzimidazole

(targets A/T); pyridoimidazole (proposed to target A/T); various other ring systems; and novel cross-DNA linkages between molecules.

Series A: Analogues of Hoechst 33258: The Series A ligands differ from Hoechst 33258 by substituting the N-terminal phenol with different ring systems. These studies suggest that changing the terminal phenyl group of Hoechst 33258 (Series A) decreases the binding affinity, but also subtly affects the selectivity. It appears that a large ring group pointing away from the minor groove of DNA creates stronger binding at TpA steps. It also appears that a methyl group facing away from the DNA groove increases the discrimination between similar sites (*e.g.* TAAT and ATAT), perhaps due to an interaction with the DNA sugar-phosphate backbone at particular local DNA structures.

Series B: Hoechst 33258/Polyamide Conjugates: The Series B ligands are conjugates of Hoechst 33258 and polyamide derivatives, with the exception of M2M-B084, which is a classic Dervan polyamide. Various ring systems are used in these ligands: Benzimidazole; pyridoimidazole; pyrrole; and imidazole.

It appears that for 2:1 unlinked polyamide binding, the optimum length of overlap of the two monomers is four units. A two unit overlap proves too weak, unless the rest of the ligand in some way interacts with the DNA to increase binding affinity without being selective (in cases of longer ligands). The introduction of benzimidazole and pyridoimidazole in these ligands has provided an insight into the binding of these groups. Two contiguous benzimidazole moieties are not able to bind effectively in a 2:1 manner, probably due to their rigidity. An inverse benzimidazole at the N-terminus of a Hoechst 33258/polyamide conjugate also abolishes binding. Pyridoimidazole shows interesting selectivity when compared to benzimidazole, seemingly creating a staggered arrangement of the polyamide monomers, so targeting a longer DNA sequence. It is important to note that pyridoimidazole and benzimidazole seem to bind stronger than pyrrole, as suggested by Briehn *et al.* (2003).

Series D: Hoechst 33258/Polyamide Conjugates: The Series D ligands are polyamide and Hoechst 33258 derivatives similar to the Series B ligands. However, instead of having benzoimidazole moieties at the terminals, these ligands have internal benzimidazole rings, often next to an N-terminal imidazole.

These ligands have shown that an N-terminal Bzi-Im does not exhibit strong binding unless linked as a complex, which is permissive to binding. However, it only appears to bind one base, due to the proximity of the hydrogen bond acceptors. It is apparent that the space occupied by the complex creates a stagger of the 2:1 ligand.

A glycine residue situated in the middle of a Hoechst 33258/polyamide conjugate weakens the binding, most likely due to the increased flexibility in the linker, so that the two halves of the compound come out of phase with the DNA and the opposing ligand monomer.

Series C: Linked Hoechst 33258/Polyamide Conjugates: The Series C ligands are comprised of two Hoechst 33258 and polyamide derivatives in a "Head-to-Head" or "Tail-to-Tail" configuration, joined together by a linker region. The design is such that the linker will traverse the DNA minor groove, with each monomer binding to opposite strands. These results confirm the literature that a positive terminus is required for polyamide binding (White *et al.*, 1997b) and indicate that a -$[CH_2]_2$- cross-DNA linker creates better binding than -CH_2-C≡C-CH_2- and -$[CH_2]_3$-. This is probably due to the rigidity incurred by the alkyne bond and the length of the propane linker being too long for good, in phase, binding. Linker length also seems to be important when considering the length of A/T DNA targeted, with a ligand with a shorter linker binding fewer bases. Only a propyl linkage shows relatively good tolerance of TpA steps in DNA, possibly due to the length of the linker.

It is unclear if 2:1 binding is occurring, with steric hindrance caused by the terminal dimethylaminopropylamide found in some of these ligands possibly inhibiting binding in this manner. Although pyridoimidazole rings are able to target guanine selectively in these ligands, the results indicate that more than one ligand is binding in close proximity, rather than in a 2:1 configuration (even though dimethylaminopropylamide is replaced by methylpiperazine in this case).

Chapter VI: Minor groove binding ligands containing isopropyl-thiazole units

The three Thiazotropsin compounds studied in this thesis contain isopropyl-thiazole ring systems. The thiazole group is selective for GC-base pairs (like imidazole) and the bulky isopropyl groups prevent side-by-side stacking with a pyrrole ring. Thiazotropsin A, containing two pyrroles and an isopropyl-thiazole moiety (*i.e.* Py-Py-Th) binds

preferentially to ACTAGT (Anthony *et al.*, 2004b; James *et al.*, 2004). The terminal bases can be either adenine or thymine, suggesting that the dimethylaminopropylamide at the C-terminus of the ligand is indiscriminate between these two bases.

Thiazotropsin B has one of the N-methylpyrrole groups of Thiazotropsin A replaced with N-methylimidazole (*i.e.* Im-Py-Th), as well as a slightly altered N-terminus, from a formyl to an acetyl group. The substitution of a pyrrole group by an imidazole dramatically altered the binding and the preferred binding site is now ACGCGT, with the imidazole targeting guanine. It is interesting to note that the binding of Thiazotropsin B to ACGCGT is about five fold weaker than Thiazotropsin A to its favoured binding site. A lower affinity has also been seen with other imidazole-containing polyamides that target GCGC motifs (Dervan, 2001).

Thiazotropsin C has the N-methylimidazole of Thiazotropsin B replaced by another isopropyl-substituted thiazole (*i.e.* Th-Py-Th). Although binding rules suggest this ligand could target the same site as Thiazotropsin B, only non-selective binding was observed. This non-selective binding suggests that the 2:1 binding mode is not permitted and that only 1:1 binding is evident at high ligand concentrations. A 2:1 binding mode is probably not favoured as the internal isopropyl-thiazole groups are too bulky to bind opposite other ring systems. Nevertheless, the terminal formyl or acetyl groups are small enough for 2:1 binding to occur opposite an isopropyl-thiazole in Thiazotropsins A and B.

The Thiazotropsin ligands have shown that isopropyl-thiazole is a useful monomer for increasing the binding site size of small molecule minor groove binding ligands. This bulky group forces staggering of the two partners in the 2:1 homodimer, thereby making the three ring system target five base pairs of DNA. The results also shown that imidazole weakens the binding affinity.

In conclusion, the results presented within this thesis have provided an insight into the selectivity and affinity of binding of a range of DNA binding ligand; especially intercalators and minor groove binding polyamides. In addition, the development of novel minimal length universal footprinting substrates will enable the systematic investigation of the sequence selectivity of novel ligands.

References

Abu-Daya, A., Brown, P.M. and Fox, K.R. (1995) DNA sequence preferences of several AT-selective minor groove binding ligands. *Nucleic Acids Res.* 23, 3385-3392.

Abu-Daya, A. and Fox, K.R. (1997) Interactions of minor groove binding ligands with long AT tracts. *Nucleic Acids Res.* 25, 4962-4969.

Addess, K.J., Gilbert, D.E., Olsen, R.K. and Feigon, J. (1992) Proton NMR studies of [N-MeCys^3,N-MeCys^7]TANDEM binding to DNA oligonucleotides: Sequence-specific binding at the TpA site? *Biochemistry* 31, 339-350.

Addess, K.J., Sinsheimer, J.S. and Feigon, J. (1993) Solution structure of a complex between [*N*-MeCys^3,*N*-MeCys^7]TANDEM and [d(GATATC)]$_2$. *Biochemistry* 32, 2498-2508.

Allemann, R.K. and Egli, M. (1997) DNA recognition and bending. *Chem. Biol.* 4, 643-650.

Anderson, J.W., Fox, K.R. and Niblo, G.A. (2006) A fast algorithm for the construction of universal footprinting templates in DNA. *J. Math. Biol.* 52, 307-342.

Anthony, N.G., Fox, K.R., Johnston, B.F., Khalaf, A.I., Mackay, S.P., McGroarty, I.S., Parkinson, J.A., Skellern, G.G., Suckling, C.J. and Waigh, R.D. (2004a) DNA binding of a short lexitropsin. *Bioorg. Med. Chem. Letts.* 14, 1353-1356.

Anthony, N.G., Johnson, B.F., Khalaf, A.I. Mackay, S.P., Parkinson, J.A., Suckling, C.J. and Waigh, R.D. (2004b) Short lexitropsin that recognizes the DNA minor groove at 5'-ACTAGT-3': Understanding the role of isopropyl-thiazole. *J. Am. Chem. Soc.* 126, 11338-11349.

Arora, P.S., Ansari, A.Z., Best, T.P., Ptashne, M. and Dervan, P.B. (2002) Design of artificial transcriptional activators with rigid poly-L-proline linkers. *J. Am. Chem. Soc.* 124, 13067-13071.

Aymami, J., Nunn, C.M. and Neidle, S. (1999) DNA minor groove recognition of a non-self-complementary AT-rich sequence by a tris-benzimidazole ligand. *Nucleic Acids Res.* 27, 2691-2698.

Baguley, B.C. (1991) DNA intercalating anti-tumour agents. *Anticancer Drug Des.* 6, 1-35.

Bailly, C., Colson, P., Hénichart, J-P. and Houssier, C. (1993) The different binding modes of Hoechst 33258 to DNA studied by electric linear dichroism. *Nucleic Acids Res.* 21, 3705-3709.

Bailly, C., Hamy, F. and Waring, M.J. (1996) Cooperativity in the binding of echinomycin to DNA fragments containing closely spaced CpG sites. *Biochemistry* 35, 1150-1161.

Bailly, C. and Waring, M.J. (1998) DNA recognition by quinoxaline antibiotics: Use of base-modified DNA molecules to investigate determinants of sequence-specific binding of triostin A and TANDEM. *Biochem J.* 330, 81-87.

Bailly, C., Echepare, S., Gago, F. and Waring, M.J. (1999) Recognition elements that determine affinity and sequence-specific binding to DNA of 2QN, a biosynthetic bis-quinoline analogue of echinomycin. *Anticancer Drug Des.* 14, 291-303.

Bakhaeva, G.P., Berlin, Yu.A., Boldyreva, E.F., Chuprunova, O.A., Kolosov, M.N., Soifer, V.S., Vasiljeva, T.E. and Yartseva, I.V. (1968) The structure of aureolic acid (mithramycin). *Tetrahedron Letts.* 32, 3595-3598.

Baraldi, P.G., Spalluto, G., Cacciari, B. and Romagnoli, R. (2000) DNA minor groove alkylating agents structurally related to distamycin A. *Expert Opin. Ther. Pat.* 10, 891-904.

Bell, A., Kittler, L., Lober, G. and Zimmer, C. (1997) DNA binding properties of minor groove binders and their influence on the topoisomerase II cleavage reaction. *J. Mol. Recognit.* 10, 245-255.

Best, T.P., Edelson, B.S., Nickols, N.G. and Dervan, P.B. (2003) Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12063-12068.

Bielawski, K., Wolczynski, S. and Bielawska, A. (2001) DNA-binding properties and cytotoxicity of extended aromatic bisamidines in breast cancer MCF-7 cells. *Pol. J. Pharmacol.* 53, 143-147.

Blattes, R., Monod, C., Susbielle, G., Cuvier, O., Wu, J-H., Hsieh, T-S., Laemmlie, U.K. and Käs, E. (2006) Displacement of D1, HP1 and topoisomerase II from satellite heterochromatin by a specific polyamide. *EMBO J.* 25, 2397-2408.

Bostock-Smith, C.E., Harris, S.A., Laughton, C.A. and Searle, M.S. (2001) Induced fit DNA recognition by a minor groove binding analogue of Hoechst 33258: Fluctuations in DNA A tract structure investigated by NMR and molecular dynamics simulations. *Nucleic Acids Res.* 29, 693-702.

Braña, M.F., Cacho, M., Gradillas, A., de Pascual-Teresa, B. and Ramos, A. (2001) Intercalators as anticancer drugs. *Curr. Pharm. Des.* 7, 1745-1780.

Bremer, R.E., Baird, E.E. and Dervan, P.B. (1998) Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch. *Chem. Biol.* 5, 119-133.

Breusegem, S.Y., Clegg, R.M. and Loontiens, F.G. (2002) Base-sequence specificity of Hoechst 33258 and DAPI binding to five $(A/T)_4$ DNA sites with kinetic evidence for more than one high-affinity Hoechst 33258-AATT complex. *J. Mol. Biol.* 315, 1049-1061.

Briehn, C.A., Weyermann, P. and Dervan, P.B. (2003) Alternative heterocycles for DNA recognition: The benzimidazole/imidazole pair. *Chem. Eur. J.* 9, 2110-2122.

Brooks, N., McHugh, P.J., Lee, M. and Hartley, J.A. (1999) The role of base excision repair in the repair of DNA adducts formed by a series of nitrogen mustard-containing analogues of distamycin of increasing binding site size. *Anticancer Drug Des.* 14, 11-18.

Brosh Jr, R.M., Karow, J.K., White, E.J., Shaw, N.D., Hickson, I.D. and Bohr, V.A. (2000) Potent inhibition of Werner and Bloom helicases by DNA minor groove binding drugs. *Nucleic Acids Res.* 28, 2420-2430.

Brown, P.M. and Fox, K.R. (1999) DNA triple-helix formation on nucleosome core particles. Effect of length of the oligopurine tract. *Eur. J. Biochem.* 261, 301-310.

Brown, S.C. and Shafer, R.H. (1987) Kinetic studies of actinomycin D binding to mono- oligo- and polynucleotides. *Biochemistry* 26, 277-282.

Brown, T., Taherbhai, Z., Sexton, J., Sutterfield, A., Turlington, M., Jones, J., Stallings, L., Stewart, M., Buchmueller, K., Mackay, H., O'Hare, C., Kluza, J., Nguyen, B., Wilson, D., Lee, M. and Hartley, J.A. (2007) Synthesis and biophysical evaluation of minor-groove binding c-terminus modified pyrrole and imidazole triamide analogs of distamycin. *Bioorg. Med. Chem.* 15, 474-483.

Buchmueller, K.L., Taherbhai, Z., Howard, C.M., Bailey, S.L., Nguyen, B., O'Hare, C., Hochhauser, D., Hartley, J.A., Wilson, W.D. and Lee, M. (2005) Design of a hairpin polyamide, ZT65B, for targeting the inverted CCAAT box (ICB) site in the multidrug resistant (MDR1) gene. *ChemBioChem* 6, 2305-2311.

Buchmueller, K.L., Bailey, S.L., Matthews, D.A., Taherbhai, Z.T., Register, J.K., Davis, Z.S., Bruce, C.D., O'Hare, C., Hartley, J.A. and Lee, M. (2006) Physical and structural basis for the strong interactions of the –ImPy- central pairing motif in the polyamide f-ImPyIm. *Biochemistry* 45, 13551-13565.

Carpenter, M.L., Marks, J.N. and Fox, K.R. (1993) DNA-sequence binding preference of the GC-selective ligand mithramycin. Deoxyribonuclease-I/deoxyribonuclease-II and hydroxy-radical footprinting at CCCG, CCGC, CGGC, GCCC and GGGG flanked by $(AT)_n$ and $A_n.T_n$. *Eur. J. Biochem.* 215, 561-566.

Castillo, U., Harper, J.K., Strobel, G.A., Sears, J., Alesi, K., Ford, E., Lin, J., Hunter, M., Maranta, M., Ge, H.Y., Yaver, D., Jensen, J.B., Porter, H., Robison, R., Millar, D., Hess, W.M., Condron, M. and Teplow, D. (2003) Kakadumycins, novel antibiotics from streptomyces sp NRRL 30566, an endophyte of grevillea pteridifolia. *FEMS Microbiol. Lett.* 224, 183-190.

Chaires, J.B., Ren, J., Hamelberg, D., Kumar, A., Pandya, V., Boykin, D.W. and Wilson, W.D. (2004) Structural selectivity of aromatic diamidines. *J. Med. Chem.* 47, 5729-5742.

Chalikian, T.V., Plum, G.E., Sarvazyan, A.P. and Breslauer, K.J. (1994) Influence of drug binding on DNA hydration: Acoustic and densimetric characterizations of netropsin binding to the poly(dAdT).poly(dAdT) and poly(dA).poly(dT) duplexes and the poly(dT).poly(dA).poly(dT) triplex at 25 °C. *Biochemistry* 33, 8629-8640.

Chen, C. (2002) The effect of two-day treatment of primary cultured ovine somatotropes with GHRP-2 on membrane voltage-gated K+ currents. *Endocrinology* 143, 2659-2663.

Chen, F.M., Sha, F. Chin, K.H. and Chou S.H. (2003a) Binding of actinomycin D to single-stranded DNA of sequence motifs d(TGTCT(n)G) and d(TGT(n)GTCT). *Biophys. J.* 84, 432-439.

Chen, F.M., Sha, F., Chin, K.H. and Chou, S.H. (2004) The nature of actinomycin D binding to d(AACCAXYG) sequence motifs. *Nucleic Acids Res.* 32, 271-277.

Chiang, S-Y., Bürli, R.W., Benz, C.C., Gawron, L., Scott, G.K., Dervan, P.B. and Beerman, T.A. (2000) Targeting the Ets binding site of the HER2/*neu* promoter with pyrrole-imidazole polyamides. *J. Biol. Chem.* 275, 24246-24254.

Chin, K.H., Chen. F.M. and Chou, S.H. (2003) Solution structure of the ActD-5'-CCGTT(3)GTGG-3' complex: Drug interaction with tandem G.T mismatches and hairpin loop backbone. *Nucleic Acids Res.* 31, 2622-2629.

Cho, J., Parks, M.E. and Dervan, P.B. (1995) Cyclic polyamides for recognition in the minor groove of DNA. *Proc. Natl. Acad. Sci. U.S.A.* 92, 10389-10392.

Ciardelli, T.L. and Olsen, R.K. (1977) Synthesis of des-*N*-tetramethyltriostin A, a bicyclic octadepsipeptide related to the quinoxaline antibiotics. *J. Am. Chem. Soc.* 99, 2806-2807.

Coll M., Frederick, C.A., Wang, A.H-J. and Rich, A. (1987) A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8385-8389.

Cons, B.M.G. and Fox, K.R. (1989) High resolution hydroxyl radical footprinting of the binding of mithramycin and related antibiotics to DNA. *Nucleic Acids Res.* 17, 5447-5459.

Coulli, J.J., He, G., Melander, C., Rucker, V.C., Dervan, P.B. and 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.

Cozzi, P. and Mongelli, N. (1998) Cytotoxics derived from distamycin A and congeners. *Curr. Pharm. Des.* 4, 181-201.

Crick, F.H. (1970) Central dogma of molecular biology. *Nature* 227, 561-563.
Dabrowiak, J.C. and Goodisman, J. (1989) in: *Chemistry and physics of DNA-ligand interactions* (N.R. Kallenbach, ed.) Adenine Press, New York, 143-174.

Darby, R.A.J., Sollogoub, M., McKeen, C., Brown, L., Risitano, A., Brown, N., Barton, C., Brown, T. and Fox, K.R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.* 30, e39.

Dawson, S., Malkinson, J.P., Paumier, D. and Searcey, M. (2007) Bisintercalator natural products with potential therapeutic applications: Isolation, structure determination, synthetic and biological studies. *Nat. Prod. Rep.* 24, 109-126.

de Clairac, R.P.L., Geierstanger, B.H., Mrksich, M., Dervan, P.B. and Wemmer, D.E. (1997) NMR characterization of hairpin polyamide complexes with the minor groove of DNA. *J. Am. Chem. Soc.* 119, 7909-7916.

Dervan, P.B. and Bürli, 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.

Dervan, P.B., Poulin-Kerstien, A.T., Fechter, E.J. and Edelson, B.S. (2005) Regulation of gene expression by synthetic DNA-binding ligands. *Top. Curr. Chem.* 253, 1-31.

Dickinson, L.A., Gulizia, R.J., Trauger, J.W., Baird, E.E., Mosier, D.E., Gottesfeld, J.M., and Dervan, P.B. (1998) Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12890-12895.

Dickinson, L.A., Trauger, J.W., Baird, E.E., Ghazel, P., Dervan, P.B. and Gottesfeld, J.M. (1999) Anti-repression of RNA polymerase II transcription by pyrrole-imidazole polyamides. *Biochemistry* 38, 10801-10807.

Dolma, S., Lessnick, S.L., Hahn, W.C. and Stockwell, B.R. (2003) Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. *Cancer Cell* 3, 285-296.

Doss, R.M., Marques, M.M., Foster, S., Chenoweth, D.M. and Dervan, P.B. (2006) Programmable oligomers for minor groove DNA recognition. *J. Am. Chem. Soc.* 123, 9074-9079.

Drew, H.R. and Travers, A.A. (1984) DNA structural variations in the E.coli *tyr*T promoter. *Cell* 37, 491-502.

Drobyshev, A.L., Zasedatelev, A.S., Yershov, G.M. and Mirzabekov, A.D. (1999) Massive parallel analysis of DNA-Hoechst 33258 binding specificity with a generic oligodeoxyribonucleotide microchip. *Nucleic Acids Res.* 27, 4100-4105

Dywer, T.J., Geierstanger, B.H., Bathini, Y., Lown, J.W. and Wemmer, D.E. (1992) Design and binding of a distamycin A analog to d(CGCAAGTTGGC). (GCCAACTTGCG): Synthesis, NMR studies, and implications for the design of sequence-specific minor groove binding oligopeptides. *J. Am. Chem. Soc.* 114, 5911-5919.

Ehley, J.A., Melander, C., Herman, D., Baird, E.E., Ferguson, H.A., Goodrich, J.A., Dervan P.B. and Gottesfeld, J.M. (2002) Promoter scanning for transcription inhibition with DNA-binding polyamides. *Mol. Cell. Biol.* 22, 1723-1733.

Fagan, P. and Wemmer, D.E. (1992) Cooperative binding of distamycin-A to DNA in the 2:1 mode. *J. Am. Chem. Soc.* 114, 1080-1081.

Fedier, A., Fowst, C., Tursi, J., Geroni, C., Haller, U., Marchini, S. and Fink, D. (2003) Brostallicin (PNU-166196) – a new DNA minor groove binder that retains sensitivity in DNA mismatch repair-deficient tumour cells. *Br. J. Cancer* 89, 1559-1565.

Fletcher, M.C., Olsen, R.K. and Fox, K.R. (1995) Dissociation of the AT-specific bifunctional intercalator [N-MeCys^3,N-MeCys^7]TANDEM from TpA sites in DNA. *Biochem J.* 306, 15-19.

Flores, L.V., Staples, A.M., Mackay, H., Howard, C.M., Uthe, P.B., Sexton III, J.S., Buchmueller, K.L., Wilson, W.D., O'Hare, C., Kluza, J., Hochhauser, D., Hartley, J.A. and Lee, M. (2006) Synthesis and evaluation of an intercalator-polyamide hairpin designed to target the inverted CCAAT box 2 in the topoisomerase IIα promoter. *ChemBioChem* 7, 1722-1729.

Fox, K.R., Olsen, R.K. and Waring, M.J. (1980a) Interaction between synthetic analogues of quinoxaline antibiotics and nucleic acids: Role of the disulphide cross-bridge and D-amino acid centres in des-N-tetramethyl-triostin A. *Br. J. Pharmacol.* 70, 25-40.

Fox, K.R., Gauvreau, D., Goodwin, D.C. and Waring, M.J. (1980b) Binding of quinoline analogues of echinomycin to deoxyribonucleic acid. Role of the chromophores. *Biochem J.* 191, 729-742.

Fox, K.R. and Waring, M.J. (1984a) DNA structural variations produced by actinomycin and distamycin as revealed by DNAase I footprinting. *Nucleic Acids Res.* 12, 9271-9285.

Fox, K.R. and Waring, M.J. (1984b) Kinetic evidence for redistribution of actinomycin molecules between potential DNA-binding sites. *Eur. J. Biochem.* 145, 579-586.

Fox, K.R. and Waring, M.J. (1984c) Evidence of different binding sites for nogalamycin in DNA revealed by association kinetics. *Biochim. Biophys. Acta* 802, 162-168.

Fox, K.R., Brassett, C. and Waring, M.J. (1985) Kinetics of dissociation of nogalamycin from DNA: Comparison with other anthracycline antibiotics. *Biochim. Biophys. Acta* 840, 383-392.

Fox, K.R. and 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. and Waring, M.J. (1986) Nucleotide-sequence binding preferences of nogalamycin investigated by DNase I footprinting. *Biochemistry* 25, 4349-4356.

Fox, K.R. (1992) Probing the conformations of eight cloned DNA dodecamers: CGCGAATTCGCG, CGCGTTAACGCG, CGCGTATACGCG, CGCGATATCGCG, CGCAAATTTGCG, CGCTTTAAAGCG, CGCGGATCCGCG and CGCGGTACCGCG. *Nucleic Acids Res.* 20, 6487-6493.

Fox, K.R. and Alam, Z. (1992) Footprinting studies of DNA-sequence recognition of nogalamycin. *Eur. J. Biochem.* 209, 31-36.

Fox, K.R. [edited] (1997) DNase I footprinting. *Methods in Molecular Biology: Drug-DNA Interaction Protocols* 90, 1-22.

Fox, K.R. and Waring, M.J. (2001) High-resolution footprinting studies of drug-DNA complexes using chemical and enzymatic probes. *Methods Enzymol.* 340, 412-430.

Gao, Y.G. and Patel, D.J. (1988) NMR studies of echinomycin bisintercalation complexes with d(A1-C2-G3-T4) and d(T1-C2-G3-A4) duplexes in aqueous solution: Sequence-dependent formation of Hoogsteen Al.T4 and Watson-Crick Tl.A4 base pairs flanking the bisintercalation site. *Biochemistry* 27, 1744-1751.

Gao. Y.G., Liaw, Y.C., Robinson, H. and Wang, A.H.J. (1990) Binding of the antitumor drug nogalamycin and its derivatives to DNA - structural comparison. *Biochemistry* 29, 10307-10316.

Gavathiotis, E., Sharman, G.J. and 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 $d(CTTTTGCAAAAG)_2$. *Nucleic Acids Res.* 28, 728-735.

Geierstanger, B.H., Dwyer, T.J., Bathini, Y., Lown, J.W. and Wemmer, D.E. (1993) NMR characterization of a heterocomplex formed by distamycin and its analog 2-ImD with d(CGCAAGTTGGC):d(GCCAACTTGCG): Preference for the 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.

Gilbert, D.E. and Feigon, J. (1992) Proton NMR study of the $[d(ACGTATACGT)]_2$-2 echinomycin complex: Conformational changes between echinomycin binding sites. *Nucleic Acids Res.* 20, 2411-2420.

Ginsburg, H., Nissani, E., Krugliak, M. and Williamson, D.H. (1993) Selective toxicity to malaria parasites by non-intercalating DNA-binding ligands. *Mol. Biochem. Parasitol.* 58, 7-15.

Gonzalez, C., Moore, M., Ribeiro, S., Schmitz, U., Schroth, G.P., Turin, L. and Bruice, T.W. (2001) The hybridization-stabilization assay: A solution-based isothermal method for rapid screening and determination of sequence preference of ligands that bind to duplexed nucleic acids. *Nucleic Acids Res.* 29, e85.

Goodwin, K.D., Long, E.C. and Georgiadis, M.M. (2005) A host-guest approach for determining drug-DNA interactions: An example using netropsin. *Nucleic Acids Res.* 33, 4106-4116.

Gottesfeld, J.M., Melander, C., Suto, R.K., Raviol, H., Luger, K. and 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. and Luger, K. (2002) Blocking transcription through a nucleosome with synthetic DNA ligands. *J. Mol. Biol.* 321, 249-263.

Guerri, A., Simpson, I.J. and Neidle, S. (1998) Visualisation of extensive water ribbons and networks in a DNA minor-groove drug complex. *Nucleic Acids Res.* 26, 2873-2878.

Harshman, K.D. and Dervan, P.B. (1985) Molecular recognition of B-DNA by Hoechst 33258. *Nucleic Acids Res.* 13, 4825-4835.

Herman, D.M., Turner, J.M., Baird, E.E. and Dervan, P.B. (1999) Cycle polyamide motif for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* 121, 1121-1129.

Horwitz, K.B. and McGuire, W.L. (1978) Actinomycin D prevents nuclear processing of estrogen receptor. *J. Biol. Chem.* 253, 6319-6322.

James, P.L., Merkina, E.E., Khalaf, A.I., Suckling, C.J., Waigh, R.D., Brown, T. and Fox, K.R. (2004) DNA sequence recognition by an isopropyl substituted thiazole polyamide. *Nucleic Acids Res.* 32, 3410-3417.

Jennewein, S. and Waring, M.J. (1997) Footprinting of echinomycin and actinomycin D on DNA molecules asymmetrically substituted with inosine and/or 2,6-diaminopurine. *Nucleic Acids Res.* 25, 1502-1509.

Kamitori, S. and Takusagawa, F. (1994) Multiple binding modes of anticancer drug actinomycin-D - X-ray, molecular modeling, and spectroscopic studies of d(GAAGCTTC)(2)-actinomycin-D complexes and its host DNA. *J. Am. Chem. Soc.* 116, 4154-4165.

Kelly, J.L., Baird, E.E. and Dervan, P.B. (1996) Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6981-6985.

Khalaf, A.I., Ebrahimabadi, A.H., Drummond, A.J., Anthony, N.G., Mackay, S.P., Suckling, C.J. and Waigh, R.D. (2004) Synthesis and antimicrobial activity of some netropsin analogues. *Org. Biomol. Chem.* 2, 3119-3127.

Kielkopf, C.L., Baird, E.E., Dervan, P.B. and Rees, D.C. (1998) Structural basis for G.C recognition in the DNA minor groove. *Nature Struct. Biol.* 5, 104-109.

Kielkopf, C.L., Bremer, R.E., White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B. and Rees, D.C. (2000) Structural effects of DNA sequence on T.A recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* 295, 557-567.

Kim, Y.B., Kim, Y.H., Park, J.Y. and Kim, S.K. (2004) Synthesis and biological activity of new quinoxaline antibiotics of echinomycin analogues. *Bioorg. Med. Chem. Letts.* 14, 541-544.

Kong, D., Park, E.J., Stephen, A.G., Calvani, M., Cardellina, J.H., Monks, A., Fisher, R.J., Shoemaker, R.H. and Melillo, G. (2005) Echinomycin, a small-molecule inhibitor of hypoxia-inducible factor-1 DNA-binding activity. *Cancer Res.* 65, 9047-9055.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E., (1985) The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. U.S.A.* 82, 1376-1380.

Kopka, M.L., Goodsell, D.S., Han, G.W., Chiu, T.K., Lown, J.W. and 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.

Kwan, A.H., Czolij, R., Mackay, J.P. and Crossley, M. (2003) Pentaprobe: A comprehensive sequence for the one-step detection of DNA-binding activities. *Nucleic Acids Res.* 31, e124.

Kwon, Y., Arndt, H-D., Choi, Y., Kawazoe, Y., Dervan, P.B. and Uesugi, M. (2004) Small molecule transcription factor mimic. *J. Am. Chem. Soc.* 126, 15940-15941.

Lacy, E.R., Nguyen, B., Le, M., Cox, K.K., O'Hare, C., Hartley, J.A., Lee, M. and Wilson, W.D. (2004) Energetic basis for selective recognition of TpG mismatched base pairs in DNA by imidazole-rich polyamides. *Nucleic Acids Res.* 32, 2000-2007.

Lahm, A. and Suck, D. (1991) DNase I-induced DNA conformation. 2 Å structure of a DNase I-octamer complex. *J. Mol. Biol.* 221, 645-667.

Lai, Y-M., Fukuda, N., Ueno, T., Matsuda, H., Saito, S., Matsumoto, K., Ayame, H., Bando, T., Sugiyama, H., Mugishima, H. and Serie, K. (2005) Synthetic pyrrole-imidazole polyamide inhibits expression of the human transforming growth factor-β1 gene. *J. Pharmacol. Exp. Ther.* 315, 571-575.

Latt, S.A. and Wohlleb, J.J. (1975) Optimal studies of the interaction of Hoechst 33258 with DNA, chromatin, and metaphase chromosomes. *Chromosoma* 52, 297-316.

Lavesa, M., Olsen, R.K. and Fox, K.R. (1993) Sequence-specific binding of [*N*-MeCys^3, *N*-MeCys^7] TANDEM to TpA. *Biochem J.* 289, 605-607.

Lavesa, M. and Fox, K.R. (2001) Preferred binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM determined using a universal footprinting substrate. *Anal. Biochem.* 293, 246-250.

Lee, J.S. and Waring, M.J. (1978) Interaction between synthetic analogues of quinoxaline antibiotics and nucleic acids. *Biochem J.* 173, 129-144.

Leslie, K.D. and Fox, K.R. (2002) Interaction of Hoechst 33258 and echinomycin with nucleosomal DNA fragments containing isolated ligand binding sites. *Biochemistry* 41, 3484-3497.

Loontiens, F.G., Regenfuss, P., Zechel, A., Dumortier, L. and Clegg, R.M. (1990) Binding characteristics of Hoechst 33258 with calf thymus DNA, poly[d(A-T)], and

d(CCGGAATTCCGG): Multiple stoichiometries and determination of tight binding with a wide spectrum of site affinities. *Biochemistry* 29, 9029-9039.

Low, C.M.L., Olsen, R.K. and Waring, M.J. (1984a) Sequence preferences in the binding to DNA of triostin A and TANDEM as reported by DNaseI footprinting. *FEBS Lett.* 176, 414-420.

Low, C.M.L., Drew, H.R. and Waring, M.J. (1984b) Sequence-specific binding of echinomycin to DNA: Evidence for conformational changes affecting flanking sequences. *Nucleic Acids Res.* 12, 4865-4879.

Lown, J.W., Krowicki, K., Bhat, U.G., Skorobogaty, A., Ward, B. and Dabrowiak, J.C. (1986) Molecular recognition between oligopeptides and nucleic acids: Novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity. *Biochemistry* 25, 7408-7416.

Maeshima, K., Janssen, S. and Laemmli, U.K. (2001) Specific targeting of insect and vertebrate telomeres with pyrrole and imidazole polyamides. *EMBO J.* 20, 3218-3228.

Malkinson, J.P., Anim, M.K., Zloh, M., Searcey, M., Hampshire, A.J. and Fox, K.R. (2005) Efficient solid-phase-based total synthesis of the bisintercalator TANDEM. *J. Organic Chem.* 70, 7654-7661.

Mallena, S., Lee, M.P.H., Bailly, C., Neidle, S., Kumar, A., Boykin, D.W. and Wilson, W.D. (2004) Thiophene-based diamidine forms a "super" AT binding minor groove agent. *J. Am. Chem. Soc.* 126, 13659-13669.

Marco, E., Negri, A., Luque, F.J. and Gago, F. (2005) Role of stacking interactions in the binding sequence preferences of DNA bis-intercalators: Insight from thermodynamic integration free energy simulations. *Nucleic Acids Res.* 33, 621-6224.

Marini, N.J., Baliga, R., Taylor, M.J., White, S., Simpson, P., Tsai, L. and Baird, E.E. (2003) DNA binding hairpin polyamides with antifungal activity. *Chem. Biol.* 10, 635-644.

Marky, L.A. and Breslauer, K.J. (1987) Origins of netropsin binding affinity and specificity: Correlations of thermodynamic and structural data. *Proc. Natl. Acad. Sci. U.S.A.* 84, 4359-4363.

Marques, M.A., Doss, R.M., Foister, S. and Dervan, P.B. (2004) Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *J. Am. Chem. Soc.* 126, 10339-10349.

Martin, R.F. and Holmes, N. (1983) Use of an 125I-labelled DNA ligand to probe DNA structure. *Nature* 302, 452-454.

Martin, C., Ellis, T., McGurk, C.J., Jenkins, T.C., Hartley, J.A., Waring, M.J. and Thurston, D.E. (2005) Sequence-selective interaction of the minor-groove interstrand cross-linking agent SJG-136 with naked and cellular DNA: Footprinting and enzyme inhibition studies. *Biochemistry* 44, 4135-4147.

Matsuda, H., Fukuda, N., Ueno, T., Tahira, Y., Ayame, H., Zhang, W., Bando, T., Sugiyama, H., Saito, S., Matsumoto, K., Mugishima, H. and Serie, K. (2006) Development of gene silencing pyrrole-imidazole polyamide targeting the TGF-β1 promoter for treatment of progressive renal diseases. *J. Am. Soc. Nephrol.* 17, 422-432.

May, L.G., Madine, M.A. and Waring, M.J. (2004) Echinomycin inhibits chromosomal DNA replication and embryonic development in vertebrates. *Nucleic Acids Res.* 32, 65-72.

Miller, D.M., Polansky, D.A., Thomas, S.D., Ray, R., Campbell, V.W., Sanchez, J. and Koller, C.A. (1987) Mithramycin selectively inhibits transcription of G-C containing DNA. *Am. J. Med. Sci.* 294, 388-394.

Mischiati, C., Finotti, A., Sereni, A., Boschetti, S., Baraldi, P.G., Romagnoli, R., Feriotto, G., Jeang, K-T., Bianchi, N., Borgatti, M. and Gambari, R. (2004) Binding of hybrid molecules containing pyrrolo[2,1-c][1,4]benzodiazepine (PBD) and oligopyrrole carriers to the human immunodeficiency type I virus TAR-RNA. *Biochem. Pharmacol.* 67, 401-410.

Morávek, Z., Neidle, S. and Schneider, B. (2002) Protein and drug interactions in the minor groove of DNA. *Nucleic Acids Res.* 30, 1182-1191.

Morpurgo, M., Radu, A., Bayer, E.A. and Wilchek, M. (2004) DNA condensation by high-affinity interaction with avidin. *J. Mol. Recognit.* 17, 558-566.

Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E. and 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. U.S.A.* 89, 7586-7590.

Mrksich, M. and 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. and Dervan, P.B. (1993b) Enhanced sequence specific recognition in the minor groove of DNA by covalent peptide dimers: Bis(pyridine-2-carboxamidonetropsin)$(CH_2)_{3-6}$. *J. Am. Chem. Soc.* 115, 9892-9899.

Mrksich, M. and Dervan, P.B. (1994) 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. and Dervan, P.B. (1994) 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.

Muller, W. and Crothers, D.M. (1968) On the structure of the actinomycin-DNA complex. *Hoppe Seylers Z. Physiol. Chem.* 349, 954.

Murty, M.S.R.C. and Sugiyama, H. (2004) Biology of N-methylpyrrole-N-methylimidazole hairpin polyamide. *Biol. Pharm. Bull.* 27, 468-474.

Neidle, S. (2001) DNA minor-groove recognition by small molecules. *Nat. Prod. Rep.* 18, 291-309.

Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* 330, 221-226.

Nordhoff, E. and Lehrach, H. (2007) Identification and characterization of DNA-binding proteins by mass spectrometry. *Adv. Biochem. Eng. Biotechnol.* 104, 111-195.

O'Hare, C.C., Mack, D., Tandon, M., Sharma, S.K., Lown, J.W., Kopka, M.L., Dickerson, R.E. and Hartley, J.A. (2002) 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. *Proc. Natl. Acad. Sci. U.S.A.* 99, 72-77.

Ongaro, A., Griffin, F., Nagle, L., Iacopino, D., Eritja, R. and Fitzmaurice, D. (2004) DNA-templated assembly of a protein-functionalized nanogap electrode. *Adv. Mater.* 16, 1799-1803.

Orfeo T., Chen L., Huang W., Ward G. and Bateman E. (1999) Distamycin A selectively inhibits acanthamoeba RNA synthesis and differentiation. *Biochim. Biophys. Acta* 1446, 273-285.

Ornstein, R.L., Rein, R., Breen, D.L. and Macelroy, R.D. (1978) Optimized potential function for calculation of nucleic-acid interaction energies .1. Base stacking. *Biopolymers* 17, 2341-2360.

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. U.S.A.* 79, 6424-6428.

Pelton, J.G. and Wemmer, D.E. (1989) Structural characterization of a 2:1 distamycin A·d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U.S.A.* 86, 5723-5727.

Pogozelski, W.K., McNeese, T.J. and Tullius, T.D. (1995) What species is responsible for strand scission in the reaction of [Fe(II)EDTA]($^{2-}$) and H_2O_2 with DNA. *J. Am. Chem. Soc.* 117, 6428-6433.

Remsing, L.L., Bahadori, H.R., Carbone, G.M., McGuffie, E.M., Catapano, C.V. and Rohr, J. (2003a) Inhibition of c-src transcription by mithramycin: Structure-activity relationships of biosynthetically produced mithramycin analogues using the c-src promoter as target. *Biochemistry* 42, 8313-8324.

Remsing, L.L., Gonzalez, A.M., Nur-e-Alam, M. (2003b) Mithramycin SK, a novel antitumor drug with improved therapeutic index, mithramycin SA, and demycarosyl-mithramycin SK: Three new products generated in the mithramycin producer streptomyces argillaceus through combinatorial biosynthesis. *J. Am. Chem. Soc.* 125, 5745-5753.

Renneberg, D. and Dervan, P.B. (2003) Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.* 125, 5707-5716.

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.

Robinson, H., Liaw, Y.C., Vandermarel, G.A., Vanboom, J.H. and Wang, A.H.J. (1990) NMR-studies on the binding of antitumor drug nogalamycin to DNA hexamer d(CGTACG). *Nucleic Acids Res.* 18, 4851-4858.

Robinson, H., Gao, Y-G., Yang, X-L., Sanishvili, R., Joachimiak, A. and Wang, A.H.J. (2001) Crystallographic analysis of a novel complex of actinomycin D bound to the DNA decamer CGATCGATCG. *Biochemistry* 40, 5587-5592.

Rosu, F., Gabelica, V., Houssier, C and De Pauw, E (2002) Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with double-stranded oligodeoxynucleotides by electrospray ionization mass spectrometry. *Nucleic Acids Res.* 30, e82.

Sastry, M. and Patel, D.J. (1993) Solution structure of the mithramycin dimer-DNA complex. *Biochemistry* 32, 6588-6604.

Sastry, M., Fiala, R. and Patel, D.J. (1995) Solution structure of mithramycin dimers bound to partially overlapping sites on DNA. *J. Mol. Biol.* 251, 674-689.

Sayers, E.W. and Waring, M.J. (1993) Footprinting titration studies on the binding of echinomycin to DNA incapable of forming Hoogsteen base pairs. *Biochemistry* 32, 9094-9107.

Scamrov, A.V. and Beabealashvilli, R.Sh. (1983) Binding of actinomycin D to DNA revealed by DNase I footprinting. *FEBS Lett.* 164, 97-101.

Schaal, T.D., Mallet, W.G., McMinn, D.L., Nguyen, N.V., Sopko, M.M., John, S. and Parekh, B.S. (2003) Inhibition of human papilloma virus E2 DNA binding protein by covalently linked polyamides. *Nucleic Acids Res.* 31, 1282-1291.

Schmitz, K. and Schepers, U. (2004) Polyamides as artificial transcription factors: Novel tools for molecular medicine? *Angew. Chem. Int. Ed.* 43, 2472-2475.

Scott, E.V., Jones, R.L., Banville, D.L., Zon, G., Marzilli, L.G. and Wilson, W.D. (1988) ^{1}H and ^{31}P NMR Investigations of actinomycin D binding selectivity with oligodeoxyribonucleotides containing multiple adjacent d(GC) sites. *Biochemistry* 27, 915-923.

Searle, M.S., Hall, J.G., Denny, W.A. and Wakelin, L.P.G. (1988) NMR-studies of the interaction of the antibiotic nogalamycin with the hexadeoxyribonucleotide duplex d(5'-GCATGC)2. *Biochemistry* 27, 4340-4349.

Shinohara, K-I., Sasaki, S., Minoshima, M., Bando, T. and Sugiyama, H. (2006) Alkylation of template strand of coding region causes effective gene silencing. *Nucleic Acids Res.* 34, 1189-1195.

Simon, H., Kittler, L., Baird, E., Dervan, P. and Zimmer, C. (2000) Selective inhibition of DNA gyrase *in vitro* by a GC specific eight-ring hairpin polyamide at nanomolar concentration. *FEBS Lett.* 471, 173-176.

Sobell, H.M. and Jain, S.C. (1972) Stereochemistry of actinomycin binding to DNA. II. Detailed molecular model of actinomycin-DNA complex and its implications. *J. Mol. Biol.* 68, 21-34.

Sobell, H.M. (1973) The stereochemistry of actinomycin binding to DNA and its implications in molecular biology. *Prog. Nucleic Acid Res. Mol. Biol.* 13, 135-190.

Soderlind, K.J., Gorodetsky, B., Singh, A.K., Bachur, N.R., Miller, G.G. and Lown, J.W. (1999) Bis-benzimidazole anticancer agents: Targeting human tumour helicases. *Anticancer Drug Des.* 14, 19-36.

Soler-López, M., Malinina, L., Liu, J., Huynh-Dinh, T. and Subirana, J.A. (1999) Water and ions in a high resolution structure of B-DNA. *J. Biol. Chem.* 274, 23683-23686.

Subramanian, M., Chander, R. and Chattopadhyay, S. (2006) A novel naturally occurring tripyrrole with potential nuclease and anti-tumour properties. *Bioorg. Med. Chem.* 14, 2480-2486.

Suck, D. and Oefner, C. (1986) Structure of DNase I at 2 Å resolution suggests a mechanism for binding to and cutting DNA. *Nature* 321, 620-625.

Sun, D. and Hurley, L.H. (1992) Effect of the (+)-CC-1065-(N3-adenine) DNA adduct on *in vitro* DNA-synthesis mediated by escherichia-coli DNA-polymerase. *Biochemistry* 31, 2822-2829.

Supekova, L., Pezacki, J.P., Su, A.I., Loweth, C.J., Riedl, R., Geierstanger, B., Schultz, P.G. and 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. and Luger, K. (2003) Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* 326, 371-380.

Suzuki, M., Yagi, N. and Finch, J.T. (1996) Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett.* 379, 148-152.

Swalley, S.E., Baird, E.E. and Dervan, P.B. (1997) A pyrrole-imidazole polyamide motif for recognition of eleven base pair sequences in the minor groove of DNA. *Chem. Eur. J.* 3, 1600-1607.

Swalley, S.E., Baird, E.E. and 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.

Tawar, U., Jain, A.K., Dwarakanath, B.S., Chandra, R., Singh, Y., Chaudhury, N.K., Khaitan, D., Tandon, V. (2003) Influence of phenyl ring disubstitution on

bisbenzimidazole and terbenzimidazole cytotoxicity: Synthesis and biological evaluation as radioprotectors. *J. Med. Chem.* 46, 3785-3792.

Taylor, A., Webster, K.A., Gustafson, T.A. and Kedes, L. (1997) The anti cancer agent distamycin A displaces essential transcription factors and selectively inhibits myogenic differentiation. *Mol. Cell. Biochem.* 169, 61-72.

Trauger, J.W., Baird, E.E., Mrksich, M. and Dervan P.B. (1996) 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.

Tseng, Y.D., Ge, H., Wang, X., Edwardson, J.M., Waring, M.J., Fitzgerald, W.J. and Henderson, R.M. (2005) Atomic force microscopy study of the structural effects induced by echinomycin binding to DNA. *J. Mol. Biol.* 345, 745-758.

Tullius, T.D. (1988) DNA footprinting with hydroxyl radicals. *Nature* 332, 663-664.

Turner, J.M., Swalley, S.E., Baird, E.E. and 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.

Tutter, A. and Jones, K.A. (1998) Chemicals that Footprint DNA: Hitting HIV-1 in the minor groove. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12739-12741.

Ughetto, G., Wang, A.H.-J., Quigley, G.J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1985) A comparison of echinomycin and triostin A complexed to a DNA fragment. *Nucleic Acids Res.* 13, 2305-2323.

Utsuno, K., Maeda, Y. and Tsuboi, M. (1999) How and how much can Hoechst 33258 cause unwinding in a DNA duplex? *Chem. Pharm. Bull.* 47, 1363-1368.

Van Dyke, M.W., Hertzberg, R.P. and 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. U.S.A.* 79, 5470-5474.

Van Dyke, M.M. and Dervan, P.B. (1983) Chromomycin, mithramycin, and olivomycin binding sites on heterogeneous deoxyribonucleic acid. Footprinting with (methidiumpropyl-EDTA)iron(II). *Biochemistry* 22, 2373-2377.

Van Dyke, M.M. and Dervan, P.B. (1984) Echinomycin binding sites on DNA. *Science* 225, 1122-1127

Vega, M.C., Garcia Saez, I., Aymami, J., Eritja, R., Van der Marel, G.A., Van Boom, J.H., Rich, A. and Coll, M. (1994) Three-dimensional crystal structure of the A-tract DNA dodecamer d(CGCAAATTTGCG) complexed with the minor-groove-binding drug Hoechst 33258. *Eur. J. Biochem.* 222, 721-726.

Viswamitra, M.A., Kennard, O., Cruse, W.B., Egert, E., Sheldrick, G.M., Jones, P.G., Waring, M.J., Wakelin, L.P. and Olsen, R.K. (1981) Structure of TANDEM and its implications for bifunctional intercalation into DNA. *Nature* 289, 817-819.

Wade, W.S. and Dervan, P.B. (1987) Alteration of the sequence specificity of distamycin on DNA by replacement of an N-methylpyrrolecarboxamide with pyridine-2-carboxamide. *J. Am. Chem. Soc.* 109, 1574-1575.

Wade, W.S., Mrksich, M. and 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.

Wadkins, R.M., Tung, C-S., Vallone, P.M. and Benight, A.S. (2000) The role of the loop in binding of an actinomycin D analog to hairpins formed by single-stranded DNA. *Arch. Biochem. Biophys.* 384, 199-203.

Wakelin, L.P.G. and Waring, M.J. (1976) The binding of echinomycin to deoxyribonucleic acid. *Biochem J.* 157, 721-740.

Wakelin, L.P.G. (1986) Polyfunctional DNA intercalating agents. *Med. Res. Rev.* 6, 275-340.

Walker, W.L., Landaw, E.M., Dickerson, R.E. and Goodsell, D.S. (1997) Estimation of the DNA sequence discriminatory ability of hairpin-linked lexitropsins. *Proc. Natl. Acad. Sci. U.S.A.* 94, 5634-5639.

Walker, W.L., Landaw, E.M., Dickerson, R.E. and Goodsell, D.S. (1998a) The theoretical limits of DNA sequence discrimination by linked polyamides. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4315-4320.

Walker, W.L., Kopka, M.L. and Goodwell, D.S. (1998b) Progress in the design of DNA sequence-specific lexitropsins. *Biopolymers* 44, 323-334.

Wang, A.H., Ughetto, G., Quigley, G.J., Hakoshima, T., van der Marel, G.A., van Boom, J.H. and Rich, A. (1984) The molecular structure of a DNA-triostin A complex. *Science* 225, 1115-1121.

Waring, M.J. (1970) Variations of the supercoils in closed circular DNA by binding of antibiotics and drugs: Evidence for molecular models involving intercalation. *J. Mol. Biol.* 54, 247-279.

Waring, M.J. and Wakelin, L.P.G. (1974) Echinomycin: A bifunctional intercalating antibiotic. *Nature* 252, 653-657.

Waring, M.J. and Bailly, C. (1994) DNA recognition by intercalators and hybrid molecules. *J. Mol. Recognit.* 7, 109-122.

Wartell, R.M., Larson, J.E. and Wells, R.D. (1974) Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid. *J. Biol. Chem.* 249, 6719-6731.

Waterloh, K. and Fox, K.R. (1991a) The effects of actinomycin on the structure of dA.dT and $(dA\text{-}dT)_n$ regions surrounding its GC binding site. *J. Biol. Chem.* 266, 6381-6388.

Waterloh, K. and Fox, K.R. (1991b) Interaction of echinomycin with $A_n.T_n$. and $(AT)_n$ regions flanking its CG binding site. *Nucleic Acids Res.* 19, 6719-6724.

Waterloh, K., Olsen, R.K. and Fox, K.R. (1992) Bifunctional intercalator [*N*-MeCys^3,*N*-MeCys^7]TANDEM binds to the dinucleotide TpA. *Biochemistry* 31, 6246-6253.

Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737-738.

Wemmer, D.E. and Dervan, P.B. (1997) Targeting the minor groove of DNA. *Curr. Opin. Struct. Biol.* 7, 355-361.

Wemmer, D.E. (2000) Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.* 29, 439-461.

Weston, S.A., Lahm, A. and Suck, D. (1992) X-ray structure of the DNase I-d(GGTATACC)$_2$ complex at 2.3 Å resolution. *J. Mol. Biol.* 226, 1237-1256.

White, S., Baird, E.E. and Dervan, P.B. (1997a) On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* 4, 569-578.

White, S., Baird, E.E. and Dervan, P.B. (1997b) Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA. *J. Am. Chem. Soc.* 119, 8756-8765.

White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E. and Dervan P.B. (1998) Recognition of the four Watson–Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* 391, 468-471.

Williams, L.D., Egli, M., Qi, G., Bash, P., Vandermarel, G.A., Vanboom, J.H., Rich, A. and Frederick, C.A. (1990) Structure of noglamycin bound to a DNA hexamer. *Proc. Natl. Acad. Sci. U.S.A.* 87, 2225-2229.

Yoon, C., Privé, 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. U.S.A.* 85, 6332-6336.

Zhan, Z-Y.J. and Devan, P.B. (2000) Alternative heterocycles for DNA recognition: A 3-pyrazole/pyrrole pair specifies for G.C base pairs. *Bioorg. Med. Chem. Letts.* 8, 2467-2474.

Zhang, X.L., Patel, D.J. (1990) Solution structure of the nogalamycin-DNA complex. *Biochemistry* 29, 9451-9466.

Zimmer, C., Reinert, K.E., Luck, G., Wahnert, U., Lober, G. and Thrum, H. (1971) Interaction of the oligopeptide antibiotics netropsin and distamycin A with nucleic acids. *J. Mol. Biol.* 58, 329-348.

Zimmer, C., Marck, C., Schneider, C. and Guschlbauer, W. (1979) Influence of nucleotide sequence on dA.dT-specific binding of netropsin to double stranded DNA. *Nucleic Acids Res.* 6, 2831-2837.