UNIVERSITY OF SOUTHAMPTON

New strategies for assessing the sequence selective binding of small molecules to DNA

by

Manuel Lavesa Curto

A Thesis Submited for the Degree of
Doctor of Philosophy

Division of Biochemistry and Molecular Biology,
School of Biological Sciences,
Faculty of Science.

August 2001

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE
SCHOOL OF BIOLOGICAL SCIENCES
DIVISION OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Doctor of Philosophy

NEW STRATEGIES FOR ASSESSING THE SEQUENCE SELECTIVE BINDING OF SMALL MOLECULES TO DNA

By Manuel Lavesa Curto

Several small molecules are known to bind to DNA in a sequence specific fashion, recognizing between 2-4 base pairs. One powerful technique which has been widely used for determining their sequence recognition properties is that of footprinting, in which the ligand protects its binding region from digestion by agents such as DNase I. However, a severe limitation of this technique is that the (unknown) preferred binding site(s) for a ligand must be present within the footprinting substrate, which is typically less than 200 base pairs long . Since there are 10 possible dinucleotide steps, 64 different trinucleotides, 136 tetranucleotides and *256* pentanucleotides, it is clear that as the specificity of the ligand increases there is a reduced chance of finding the correct site within any given restriction fragment. Even if the preferred binding site is present, a proper analysis of the specificity should examine the binding to related sequences, which may also not be present. For these reasons we have developed two different approaches towards overcoming this limitation of footprinting.

REPSA: (Restriction Enzyme Protection Selection and Amplification) is a novel technique used to determine the binding sites of DNA binding ligands. This technique uses type IIs restriction enzymes (which cut several bases away from their recognition sites) to selectively cleave unbound DNA, while DNA fragments to which the ligand is bound are protected from digestion. The reaction is performed with oligonucleotides containing recognition sites for these enzymes which flank a central region of random bases around their cutting site. The uncut products of the reaction are amplified by PCR, and are subjected to further rounds of drug binding and enzyme cleavage, enriching for oligonucleotides containing the preferred ligand binding sites. After several rounds of selection the fragments are cloned and sequenced, revealing the preferred binding sites. We have successfully employed this technique for confirming the binding sites of TANDEM (TpA) and distamycin (AT), though it has been less successful for determining the binding sites of echinomycin (CpG).

MULTISITE: We have prepared a DNA footprinting substrate which contains all 136 possible tetranucleotide sequences, and have used DNase I footprinting to examine the binding of some ligands with known recognition properties. These ligands include actinomycin, which is known to bind to GpC; echinomycin (CpG), TANDEM (TpA), mithramycin (GpC, GpG) and the AT-selective minor groove binding ligands distamycin and Hoechst 33258. These ligands show large differences in their binding to tetranucleotides which contain the preferred dinucleotide steps. For example ACGT and ACGG are much better binding sites for echinomycin than GCGC, CCGG and GCGG.

PREFACE

This work was performed between September 1997 and July 2001 at the University of Southampton, in the Division of Biochemistry and Molecular Biology.

Some of the results presented in chapter 5 have been published as follows:

Lavesa, M and Fox, K.R. (2001). Preferred binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM determined using a universal footprinting substrate. *Analytical Biochemistry* **293**, 246-250.

I would like to thank Keith for all his help and guidance during this project. A very special thanks is also due to my wife, Jenny, for her encouragement and support during the hard times and for becoming an English woman. I couldn't have done it without you!. My appreciation and gratitude also go to my family for their support during all these years and to all the lab colleagues that put up with my temper, but especially to Darren and Richard for their friendship.

Table of Contents

CHAPTER 1

Introduction

CHAPTER 2

Materials and Methods

CHAPTER 3

Determination of the best conditions for REPSA

(Restriction Enzyme Protection Selection and Amplification)

CHAPTER 4

Drug binding sites determined by REPSA

CHAPTER 5

MULTISITE

CHAPTER 6

Work in progress

REFERENCES

Apendix A

Computer program

Figure index

Table index

Table number **In or after page**

ABBREVIATIONS

A, T, C, G	Adenine, Thymine, Cytosine, Guanine
AMPS	20% (w/v) Ammonium Persulphate
ATP	Adenosine triphosphate
cps	counts per second
cfu	colony formation units
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetracetic acid
IPTG	Isopropylthio-β-*D*-galactoside
N	any nucleotide(A, T, C, G)
NaOAc	Sodium Acetate
NMR	Nuclear Magnetic Resonance
OD	Optical density
R	Purine
RNA	Ribonucleic acid
TBE	Tris-HCl, boric acid, EDTA 5X buffer
TE	10 mM Tris-HCl, pH 7.5, containing 0.1 mM EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris-hydroxymethyl-aminomethane
UV	Ultra violet
XGal	5-Bromo–4-chloro–3-indolyl-β-D-galactoside
Y	Pyrimidine

CHAPTER 1

Introduction

The study of drug-DNA interactions is important for the design of novel chemotherapeutic agents, possibly leading to progress in cancer therapy. Small molecules which interact with DNA can be classified by their mode of binding, namely covalent or non-covalent. Drugs which bind covalently can be further divided into cross-linking agents such as the bifunctional alkylating agents mitomycin and melphalan and compounds like cis-platin, and strand-breaking antibiotics such as bleomycin, streptonigrin and phleomycin. The non-covalent binding group can be subdivided into intercalating agents, including the anthracyclines (e.g. daunomycin) and quinoxaline antibiotics, and those which do not intercalate but bind to the minor groove of DNA such as netropsin and distamycin (Neidle and Waring, 1993). Several of these agents bind with sequence selectivity. Consequently, the identity of the binding sites has been the focus of intense research for many years with the ultimate aim of designing gene specific DNA binding drugs.

DNA is one of the few pharmacological receptors whose structure is known in precise detail, permitting (in theory) the rational design of drugs which can bind to it. Thus, a basic knowledge of the DNA structure is important for understanding the mechanisms of action of these drugs.

1.1 DNA Structure

The three-dimensional structure of DNA was proposed in 1953 (Watson and Crick, 1953) and was suggested to be composed of two helical chains each coiled around a common axis, forming a right-handed thread and running in opposite directions. Each chain consists of deoxyribonucleotides with 3',5' linkages (Fig. 1.1). The purine and pyrimidine bases are positioned on the inside of the helix carrying the genetic information, and the phosphate groups and deoxyribose units are on the outside performing a structural

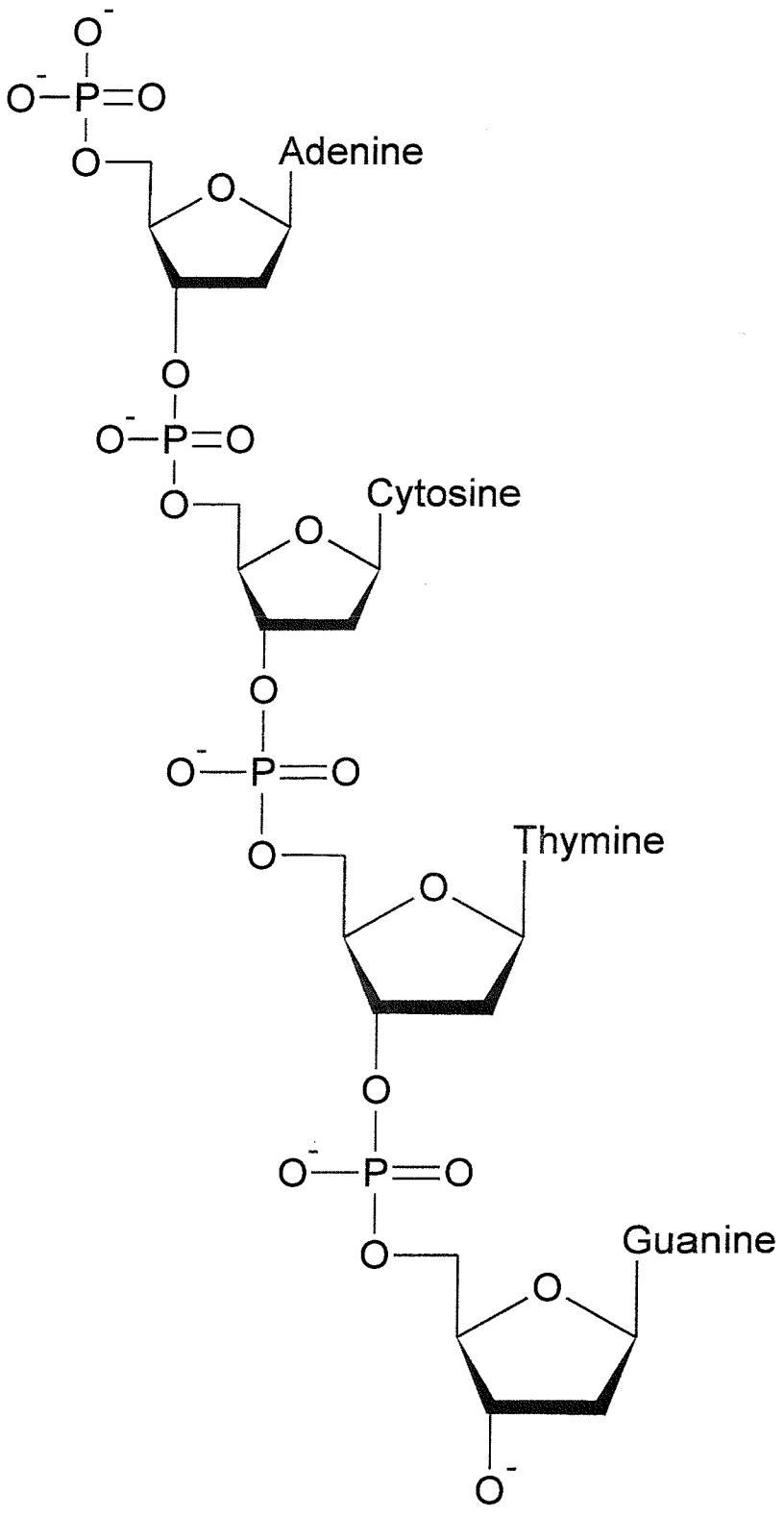


Figure 1.1 A selection of the polynucleotide chain in DNA

role. The planes of the sugars are nearly at right angles to those of the bases which are perpendicular to the helix axis. The diameter of the helix is 20 Å. Adjacent bases are separated by 3.4Å along the helix axis and related by a rotation of 36 degrees. Consequently, the helical structure repeats every 10 residues on each chain, that is, a spanning distance of 34 Å. The two chains are held together by hydrogen bonds between pairs of bases. Thus, one member of the pair has to be a purine and the other a pyrimidine for bonding to occur and adenine must pair with thymine, and guanine with cytosine (Fig1.2) due to steric and hydrogen bonding factors (Watson and Crick, 1953). This base-pairing architecture has been experimentally supported as the molar ratio of adenine to thymine, and the molar ratio of guanine to cytosine, are always very close to unity (Zamenhof *et al.*, 1952). Hydrogen bonding may occur in two or more different conformations. The most common one, found in all DNA in nature, is the Watson-Crick base pair and the other is known as the Hoogsteen base pair (Hoogsteen, 1959) (Fig1.3).

The right-handed regular structure of DNA is known as B-DNA. However, it is known that the structure of DNA is not uniform and that it can assume different conformations. The factors influencing the equilibrium between the structures include sequence, superhelical stress, ionic conditions, and protein binding. The most common structures are the B-DNA and A-DNA, and the most unusual include Z-DNA (Dickerson *et al.*, 1982), and triple helices. These structures will be considered in turn.

B-DNA has two types of groove (Timsit, 1999) that appear since the glycosidic bonds of a base pair are not diametrically opposite each other. One of the grooves is called the major groove (12 Å wide) and the other the minor groove (6 Å wide). The minor groove contains the pyrimidine O-2 and the purine N-3 of the base pair, and the major groove is on the opposite side of the pair (Fig 1.2) which is also slightly deeper than the minor groove (8.5 versus 7.5 Å). The bases are perpendicular to the helix axis, which runs through the centre of the helix.

A-DNA, like B-DNA, is a right-handed double helix formed by antiparallel chains

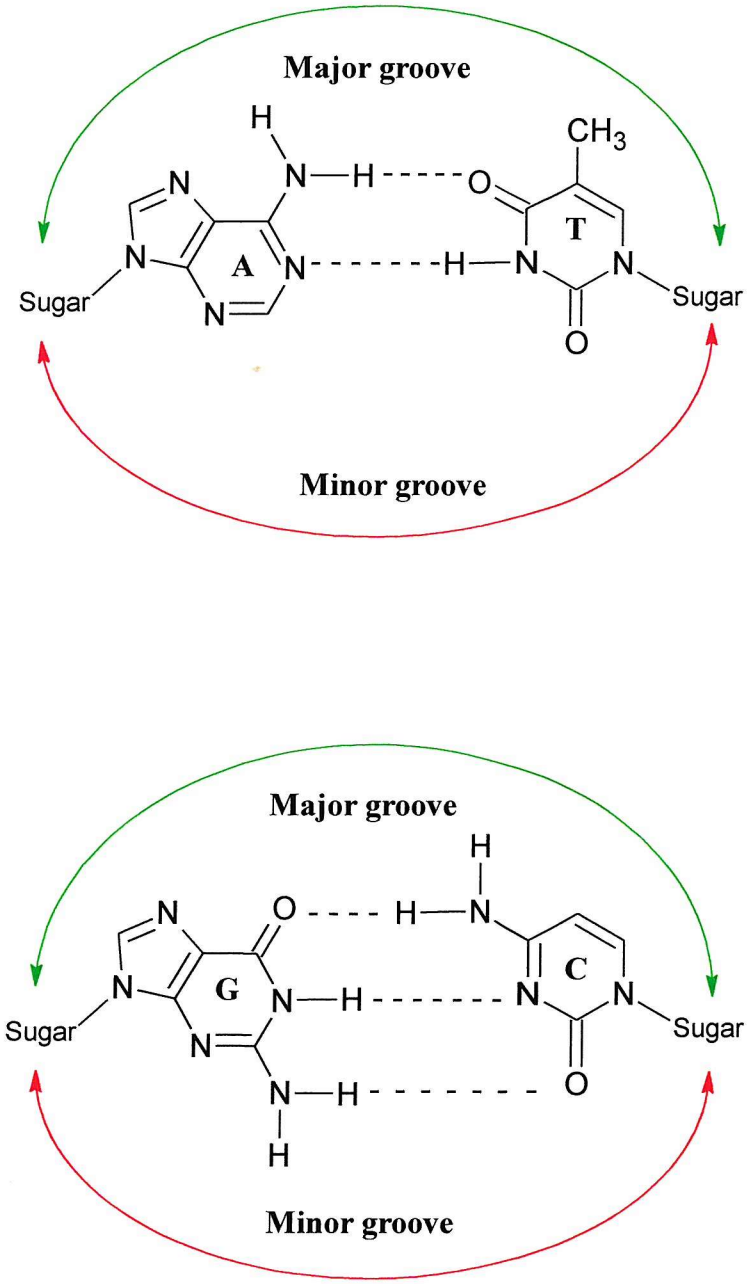
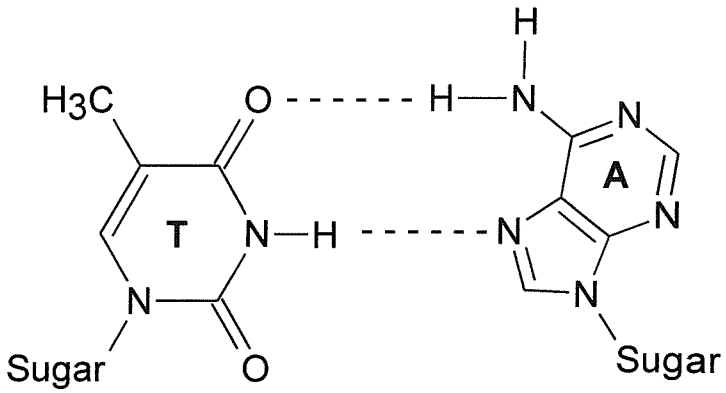
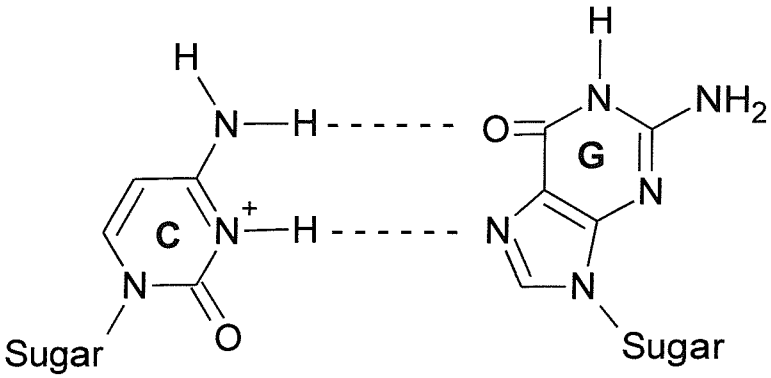


Figure 1.2 Watson and Crick GC and AT base pairs showing the Major and Minor groove.



Hoogsteen T-A



Hoogsteen C-G

Figure 1.3 Hoogsteen base pairs

held together by Watson-Crick base-pairing. This structure is favoured when dehydration occurs because its phosphate groups bind fewer water molecules than do those in B-DNA. The main difference between these two structures lies in the puckering of the sugars. In the case of A-DNA the sugar has a C3'-*endo* configuration and a C2'-*endo* configuration is found in B-DNA (Fig. 1.4a). In addition, the helix of A-DNA is wider and shorter than that in B-DNA possessing a very deep major groove, the minor groove albeit very shallow and can hardly be termed a groove at all (Dickerson *et al.*, 1982). The bases are inclined to the helical axis, which does not run through the centre of the base pairs, forming a structure similar to a spiral staircase.

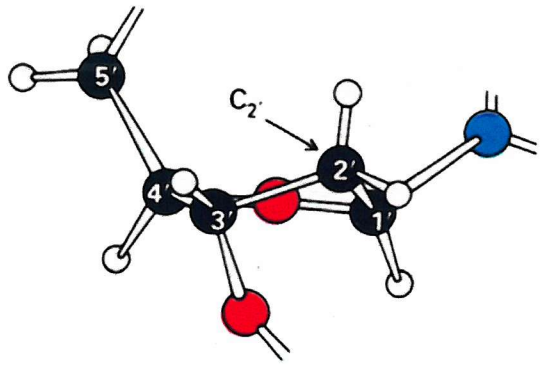
Z-DNA is a left-handed double helix again formed by Watson-Crick base pairing. It is thinner and more elongated than B-DNA. The minor groove is very deep and the major groove is completely flattened out on the surface of the molecule. This form of DNA was discovered by Rich and co-workers (Rich *et al.*, 1984) while studying the crystal structure of CGCGCG. They observed that the bases alternate between *syn* and *anti* conformations (Fig 1.4b). Since the *syn* conformation is more stable for purines than for pyrimidines, Z-DNA is favoured in nucleotide sequences that have alternating purines and pyrimidines (especially alternating GC). The formation of Z-DNA is favoured by conditions of high salt and can be stabilized by negative supercoiling.

Triple helices are formed in regions of DNA that contain runs of pyrimidines in one strand because these regions may fold back on themselves to form three stranded structures. These helices contain one polypurine and two polypyrimidine tracts (Lee *et al.*, 1984). One of the polypyrimidine strands runs parallel to the polypurine strand (i.e. in the same 5' to 3' orientation) and forms Hoogsteen base pair interactions (Fig.1.5).

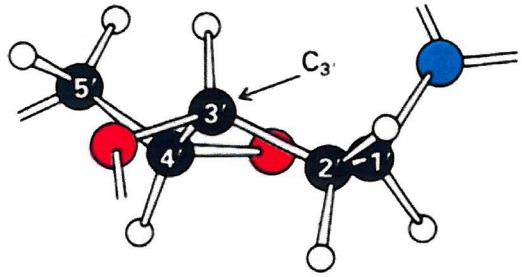
1.2 INTERCALATION

In 1961 Lerman was the first to describe the intercalation model explaining the binding of proflavine and ethidium to DNA. His original observations come from X-ray diffraction patterns and viscometric and sedimentation studies. In the presence of proflavine

a)

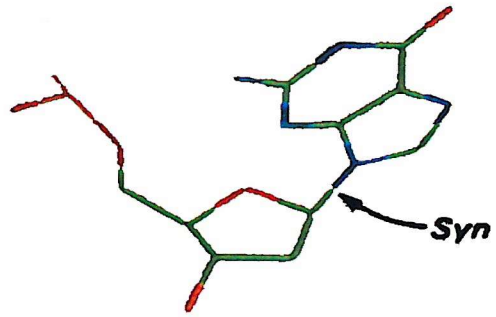


$C_{2'}$-*endo*



$C_{3'}$-*endo*

b)



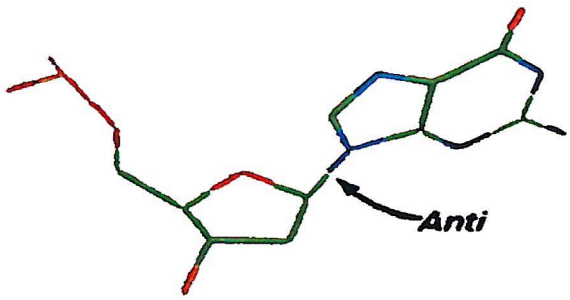


Figure 1.4 a) Sugar puckering markedly affects the orientation of the phosphodiester bridges and the glycosidic bond $C_{2'}$-*endo* (found in B-DNA) and $C_{3'}$-*endo* (found in A- DNA) denote which atom of the ribose ring lies above the plane in the orientation shown here. b) The *syn* and *anti* conformations of the glycosidic bond

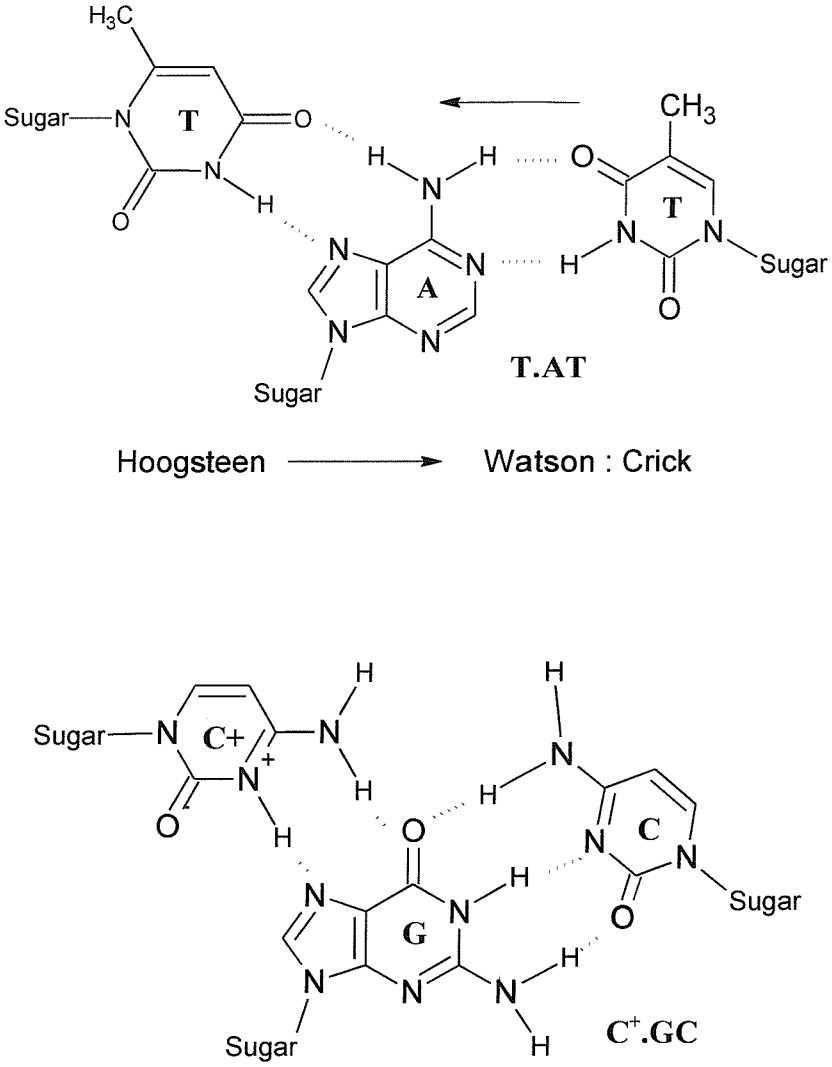


Figure 1.5 Structures of the parallel triplets showing the Hoogsteen and Watson and Crick base pairing.

a closer packing of the molecules was observed, which showed that proflavine binds internally to the helix, with a concomitant increase in viscosity and a decrease in the sedimentation coefficient of DNA. These changes in viscosity and sedimentation could be attributed to a lengthening of the DNA molecule by the drug. Thus, Lerman proposed that the flat polycyclic aromatic ring system of a drug molecule becomes inserted (i.e .intercalated) between the adjacent base-pairs of the double helix, causing it to unwind locally. At present, there are many techniques to determine if a drug interacts with DNA by intercalation, including changes in superhelical density, viscometry and electronic dichroism. In addition, there are three basic criteria that have to be satisfied for intercalation to occur; (1) the helix has to be extended, (2) it has to be demonstrated that there is local unwinding in the presence of the drug and (3) the plane of the polycyclic aromatic ring system of the bound drug has to be parallel to the base-pairs, therefore, perpendicular to the helical axis.

1.3 DNA MINOR GROOVE BINDING AGENTS

Most minor groove ligands bind to AT rich regions. This selection has been attributed to several factors:

1) There are two hydrogen bonding acceptors sites in all the dinucleotide pairs (TA, AT, CG and GC). However, guanine has a hydrogen bond donor at the N2 position, which is sticking out into the centre of the minor groove. This N2 hydrogen is thought to impede binding of ligands, thus making AT or TA a more favourable site (Fig. 1.2).

2) In the minor groove the AT or poly A regions have been observed to be narrower than GC regions (Laughton and Luisi, 1998), thus, increasing the van der Waals forces and the electrostatic interactions between the ligand and the bases within these regions.

3) There is a greater negative electrostatic potential within AT regions. (Pullman, 1983).

There are many AT selective compounds which are known to bind in the minor groove including distamycin, netropsin, Hoechst 33258, pentamidine, DAPI, berenil, etc... (Neidle and Waring, 1993), but for this section we will concentrate on distamycin and Hoechst 33258. Other minor grove ligands to be considered include the pyrrole polyamides which have been suggested as specific sequence reading agents, and mithramycin which is a GC-selective minor groove binding antibiotic.

1.3.1 Distamycin

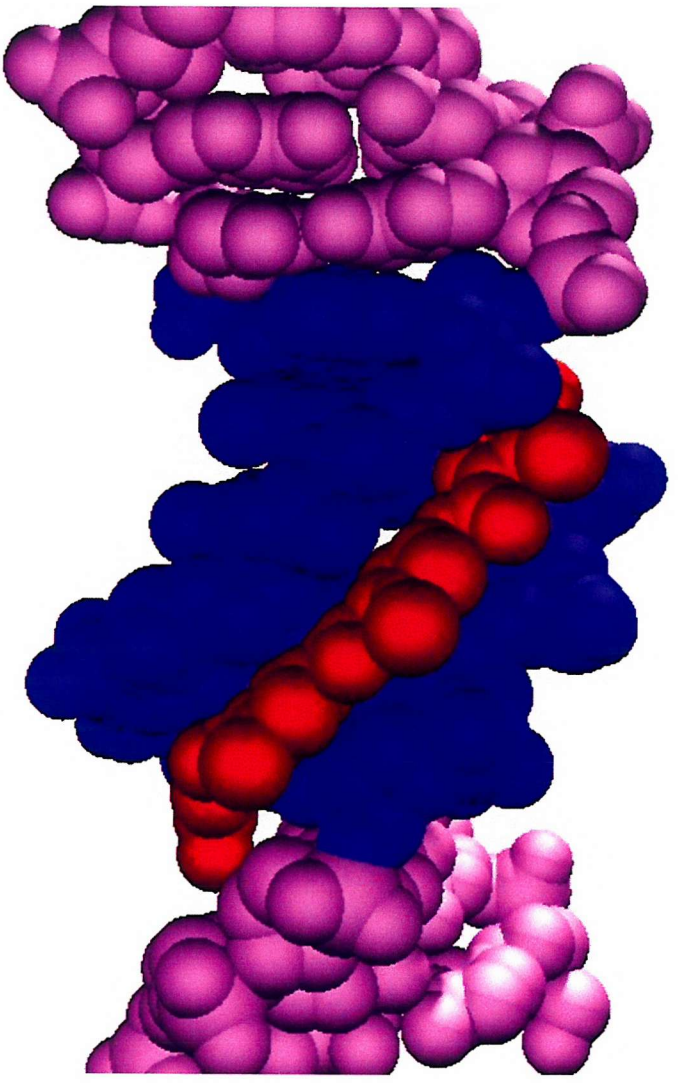
Distamycin is an antibiotic that binds to the minor groove of DNA. Its structure (Fig. 1.6) consists of three N-methyl- pyrrole rings, joined by amide groups, with formamide and propylamidinium end groups.

1.3.1.1 Sequence selectivity

Footprinting studies (section 1.6) using DNase I (Fox and Waring, 1984; Portugal and Waring, 1987a; and Fish *et al.*, 1988), hydroxyl radicals (Churchill *et al.*, 1990; Portugal and Waring., 1987b) and methidiumpropyl-EDTA-Fe(II) (Van Dyke *et al.*, 1982) as cleaving agents, showed that distamycin binds preferentially to regions of high A and T content. It protects 3 to 4 base pairs and the binding site appears to be 5 base pairs long. In addition, a DNase I footprinting study (Abu-Daya *et al.*, 1995) showed that there is a ranking order for distamycin binding AAAA = AATT > ATTA / TAAT = TATA > ATAT. Similar binding affinities were found using circular dichroism (Chen and Sha, 1998). Both methods show that AAAA is among the best binding sites for this agent.

Initially it was believed that distamycin could not bind to GC regions due to the interference of the amino group at the C2 position of guanosine. However, footprinting

A)



B)

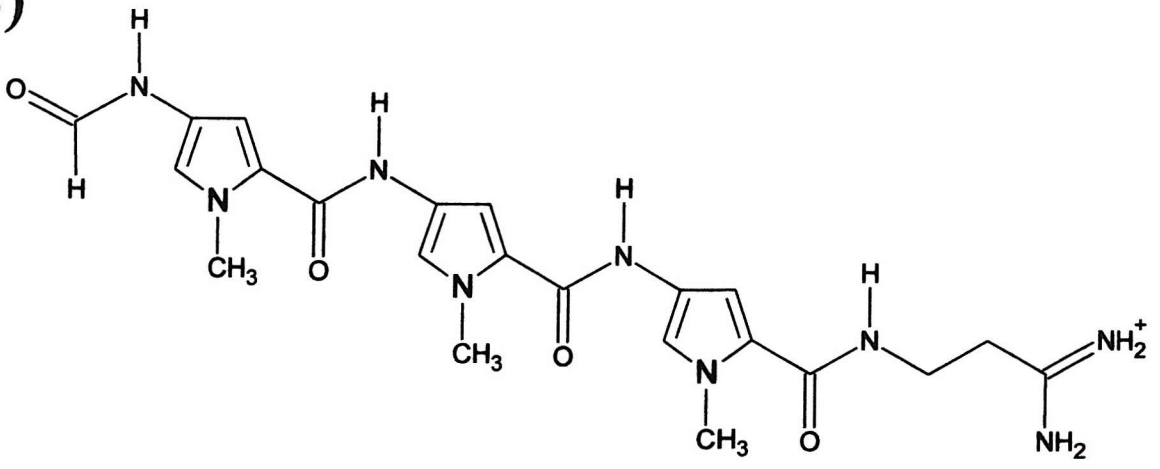


Figure 1.6 A) Crystal structure of distamycin (red) bound to AAATTT (blue), the view is through the minor groove. The image was obtained from the pdb file submitted by Coll et al. (1987), using the WebLab ViewerLite and rendered with Pov-Ray for windows. B) Molecular structure of distamycin.

studies showed that, in certain circumstances, distamycin can bind to sites which contain a GC base pair situated either in the centre of the binding site or at the flanking regions. (Churchill *et al.*, 1990; Van Dyke *et al.*, 1982). In addition, spectrophotometric studies showed that the ligand does not bind to poly(dDAP-dT) (DAP = 2,6 diamino purine) (Fig.1.7) (Dasgupta *et al.*, 1990), thereby, confirming the negative role of the guanine 2-amino group. In addition, there is a footprinting study which shows that distamycin binds to CpI but not to DAPpT; the conclusion from these results was that the 2-amino group of guanine acted as a negative signal for binding (Bailly and Waring, 1995).

1.3.1.2 Mode of binding

One of the first X-ray crystal structures of this antibiotic class was that of netropsin bound to CGCGAATT^{Br}CGCG. This compound is related to distamycin and binds to DNA by a similar mechanism (Kopka *et al.*, 1985). The analysis of the structure revealed that netropsin bound to the minor groove by displacing a spine of water molecules. In addition, the antibiotic was found to bind to DNA by forming hydrogen bonds between its amide groups and the adenine N3 and the thymine O2 (Fig 1.8). Netropsin caused a widening of the minor groove upon binding and it was found to conform to the DNA curvature. This malleability is due to the N-methyl-pyrrole rings, which can twist relative to each other and form different angles in order to conform to the shape of the minor groove (Kopka *et al.*, 1985). This feature has also been observed for distamycin (Ekambareswara *et al.*, 1988). Distamycin crystalized with the DNA dodecamer d(CGCAAATTTGCG) (Coll *et al.*, 1987), showed that its binding is very similar but not identical to that of netropsin. The differences are due to changes in the chemical structure, as the distamycin molecule contains a formamide-methyl-pyrrole group compared to a guanidinium group found in netropsin (Kopka *et al.*, 1985; Coll *et al.*, 1987). This difference in structure allows distamycin to make more hydrogen bonds than netropsin. The bonding occurs between the distamycin amino groups 1, 5, and 7 which form bifurcated hydrogen bonds simultaneously to both an adenine N3 and a thymine O2 atom, while the amino group 9 formed a single bond to adenine N3 (Fig1.9) (Coll *et al.*, 1987). In addition, there are strong van der Waals interactions between the pyrrole H3's of distamycin and the adenine C2H's, which were found in an NMR study of distamycin complexed with d(CGCGAATTCGCG)$_2$ (Fig

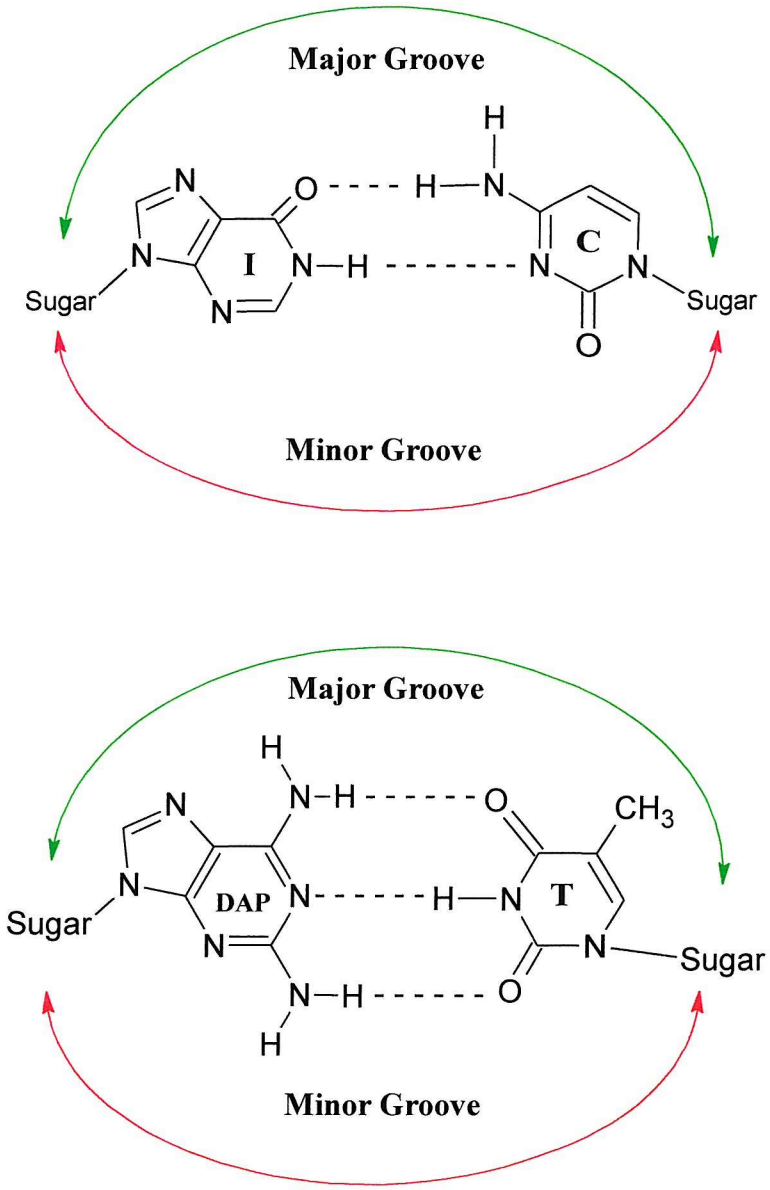


Figure 1.7 Structures of Inosine (I) and 2,6 diaminopurine (2-aminoadenine) (DAP) . The major and minor grooves of the helix are indicated. Compare these structures with Fig. 1.1.4)

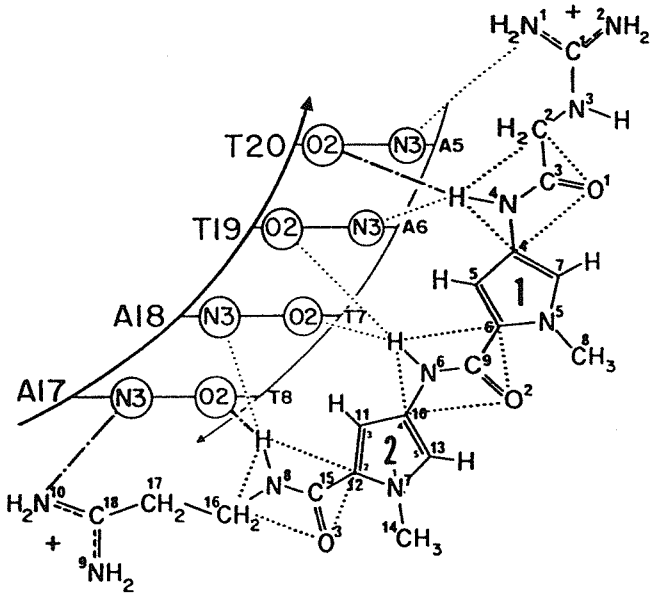


Figure 1.8 Diagramatic representarion of netropsin binding to DNA. DNA is shown only by a minor groove ladder, with adenine N3 and thymine O2 atoms indicated.. Dot-dashed lines indicate indicate N-to-N or N-to-O distances short enough to be standard hydrogen bonds, whereas dotted lines indicate distances of 3.2 A or more. Taken from Kopka *et al.* (1985)

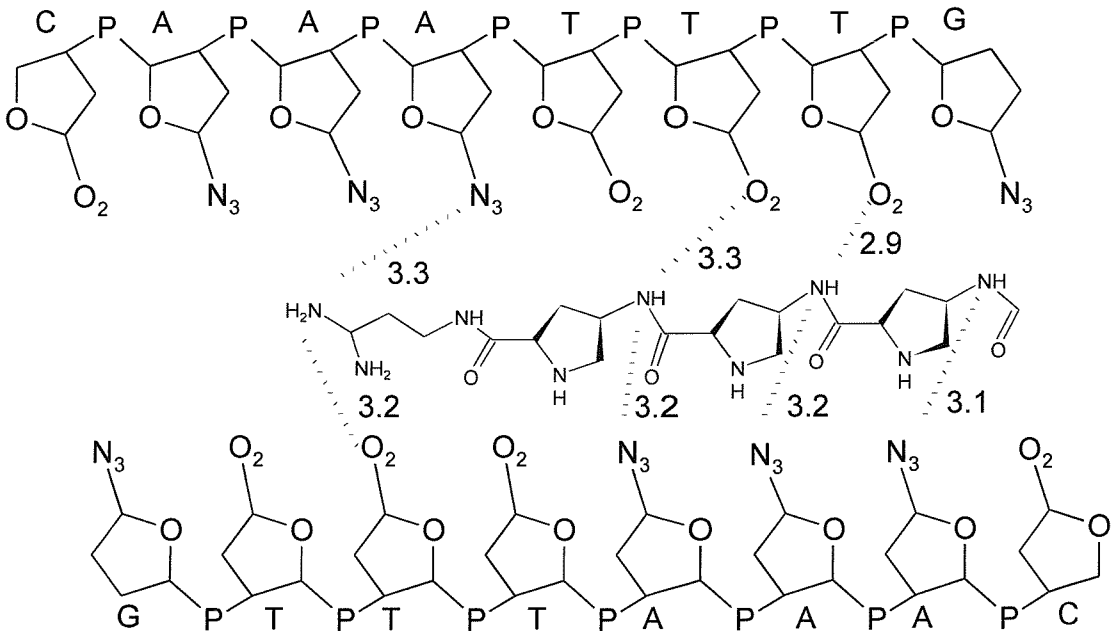


Figure 1.9 A schematic diagram illustrating the hydrogen bonding between distamycin and the electronegative nitrogen and oxygen atoms in the floor of the minor groove. Dashed lines indicate hydrogen bonding interactions and numbers indicate the length in A of the hydrogen bonds. The first amino group N1 is on the left of the distamycin molecule and N-9 is on the right. Based in Coll *et al.* (1987).

1.10)(Klevit *et al.*, 1986).

NMR (Fagan and Wemmer, 1992; Pelton and Wemmer, 1989, 1990) and circular dichroism spectroscopy (Rentzeperis *et al.*, 1995) demonstrated that distamycin is also able to bind to DNA in a 2:1 molar ratio. Some of the sequences that exhibited this 2:1 binding mode were ATATAT and CIIICC (Fagan and Wemer, 1992), AAATTT and AAATT (Pelton and Wemmer, 1989, 1990) and AAATT and AATTT (Renzeperis *et al.*, 1995). The first distamycin molecule binds to DNA with strong affinity and with favourable enthalpy. The second distamycin molecule has a lower binding affinity but has a much more favourable enthalpy, due to the van der Waals interactions, and the entropy is less favourable due to the interaction of water and ions (Rentzeperis *et al.*, 1995). The stability of this 2:1 complex is dependent on the hydrogen bonds, van der Waals forces and interaction due to the stacking of the drug-drug and drug-DNA. Moreover, a widening of the minor groove has been observed suggesting that this AT region must be distorted in order to accommodate the ligand (Pelton and Wemmer, 1989). Most of the NMR structures of the 2:1 complex have been shown in sequences containing AATTT or AAATTT with different flanking sites (Rentzeperis *et al.*, 1995; Pelton and Wemmer, 1990; and Fagan and Wemmer, 1992). Experiments with circular dichroism (Chen and Sha, 1998) showed that distamycin formed the 2:1 complex when the binding site contained at least five base pairs, these binding sites were ranked by their binding cooperativity as follows: AAGTT, ATATA ≥AAACT > AATAA, AAATA, AAAGT > AATAT . TAAAA ≥AAATT ≥AAAAA ≥ATAAA, AAAAT and by their dissociation rates (slowest at the left) AAGTT > ATATA > AAATA > AATAA ≥ AAATT > TAAAA > AAAGT, AAACT, ATAAA, AATAT, AAAAA, AAAAT. The ranking order is similar except for AAACT and AAAGT. In addition, there is an NMR spectroscopy study which uses an oligonucleotide that contains the sequences AAGTT and AACTT (Geierstanger *et al.*, 1993). In this study they used a distamycin analogue 2-ImD (imidazole distamycin) as well as distamycin and they found that a complex of 1:1:1 2-ImD, distamycin, DNA complex was formed (Fig.1.11) at a higher rate than 2:1 2-ImD, DNA complex and 2:1 distamycin, DNA complex. Also the binding affinities of these ligands to AAGTT and AACTT were found to be in the order 2-

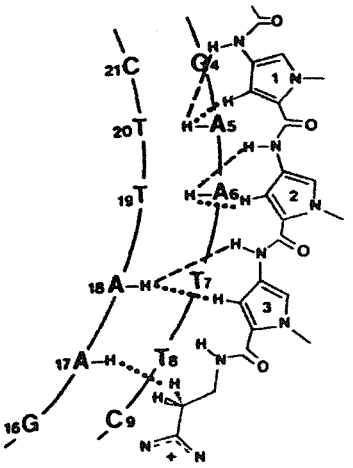


Figure 1.10 Diagramatic representation of the distamycin binding site in the central AATT region of d(CGCGAATTCGCG)$_2$. The observed NOE contacts from assigned adenine C2H resonances to distamycin amide protons (dashed lines) and the non-exchangeble distamycin protons (dotted lines are indicated. Taken from Klevit *et al.* (1986)

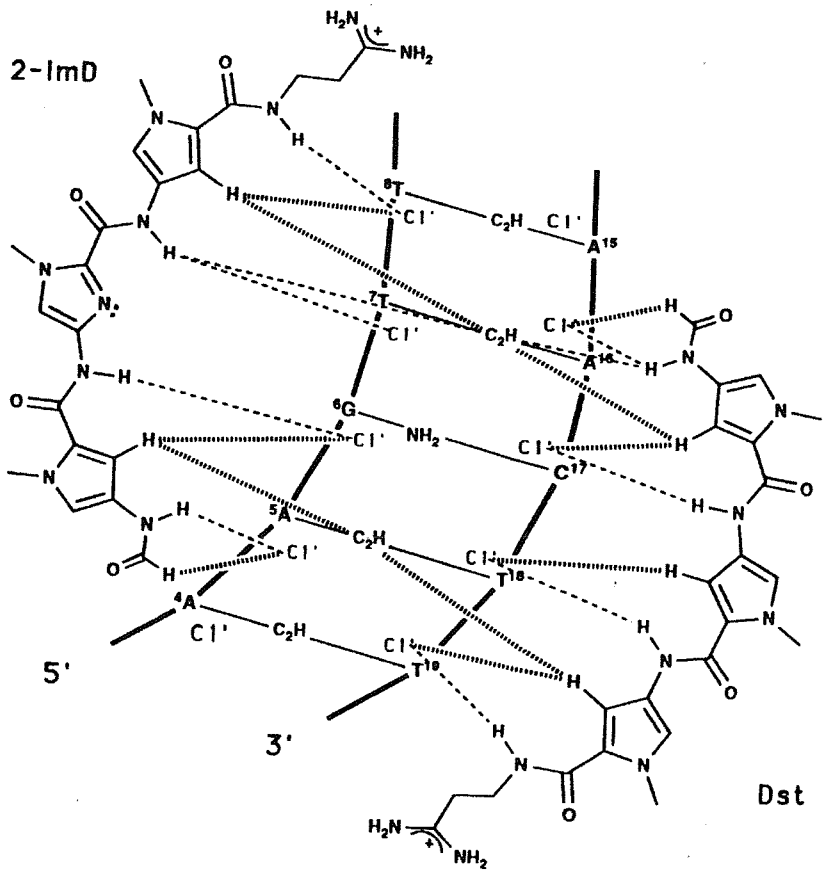


Figure 1.11 Intermolecular contacts between the 2-ImD and distamycin (Dst)ligand of the 1:1:1 2-ImD:Dst:DNA complex with d(CGCAAGTTGGC):d(GCCAAGTTGCG). NOE contacts between amide protons of the ligands and DNA protons are indicated as dashed lines, while contacts between other ligand protons and DNA are shown as stippled lines. Ligand-DNA contacts involving the amidinium and methylene -protons as well as ligand-ligand contacts are not included. This figure was taken from Geierstanger *et al.* 1993)

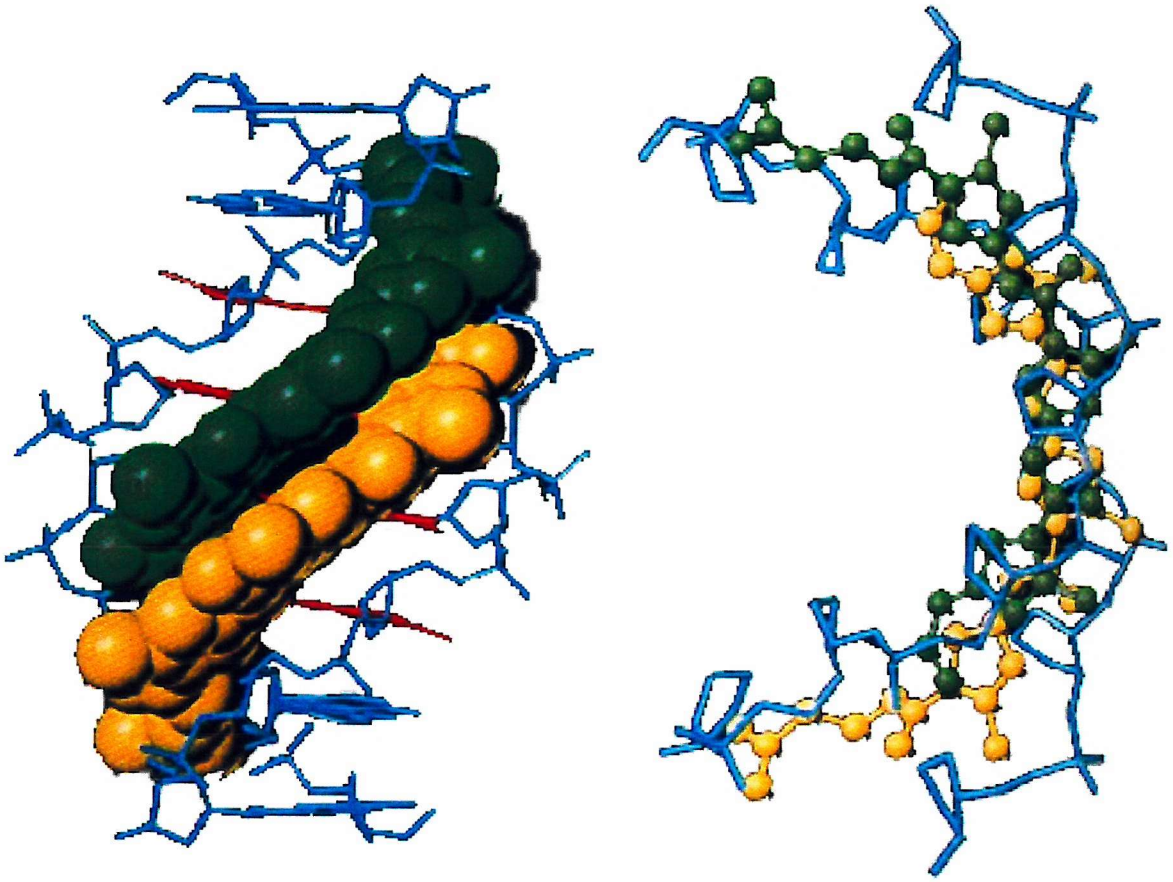


Figure 1.12 Three dimentional structures of the 2:1 distamycin-DNA complex with d(ICATATIC)$_2$ in the left one the ligands are shown in space-filled models in green and yellow, and the DNA is shown as a stick model with the central AT base-pairs in red. The right structure shows the sandwiching of the two drugs between the sugar-phosphate backbone of DNA. All the bases are omitted for clarity. The drugs are in ball and stick model, the DNA in stick. taken from Chen *et al.*(1997).

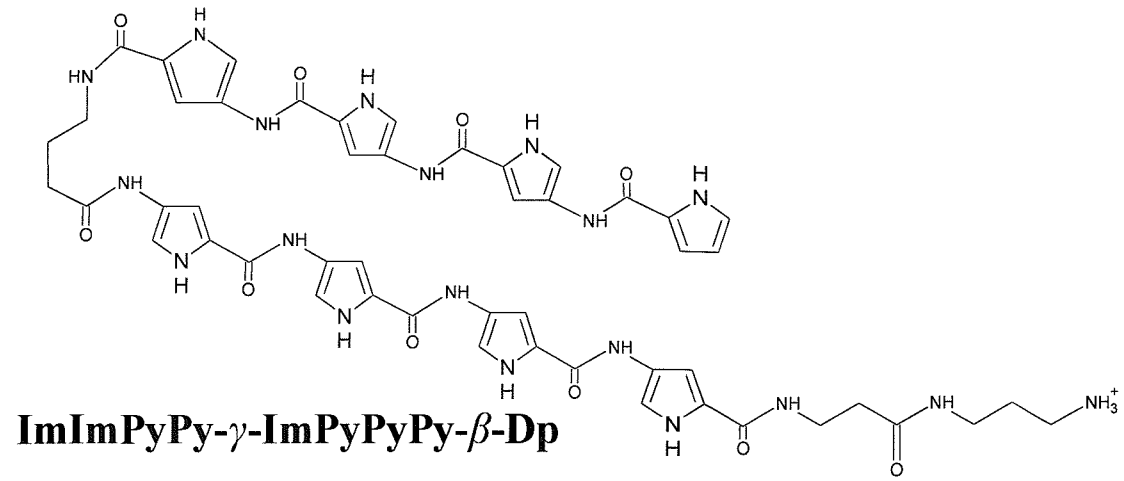
ImD.Dist.DNA > Dist.Dist.DNA > 2-ImD.2-ImD.DNA > Dist.2-ImD.DNA. (Singh *et al.*, 1994). The imidazole nitrogen of the 2-ImD ligand was bound to the guanine NH2, thus recognizing GC base pairs.

The crystal structures of the 2:1 distamycin, DNA complexed with d(GTATATAC) (Mitra *et al.* 1999), d(ICITACIC) and d(ICATATIC) (Chen *et al.* 1997) have been solved (Fig. 1.12). The IpC base pair resembles the minor groove of the ApT base pair due to the lack of the 2-amino group , however, the major groove resembles the GpC base-pair (Fig.1.7). Thus, distamycin binds to A-T/I-C regions but not to G-C base pairs due to the steric hindrance of the 2-amino group of guanine.

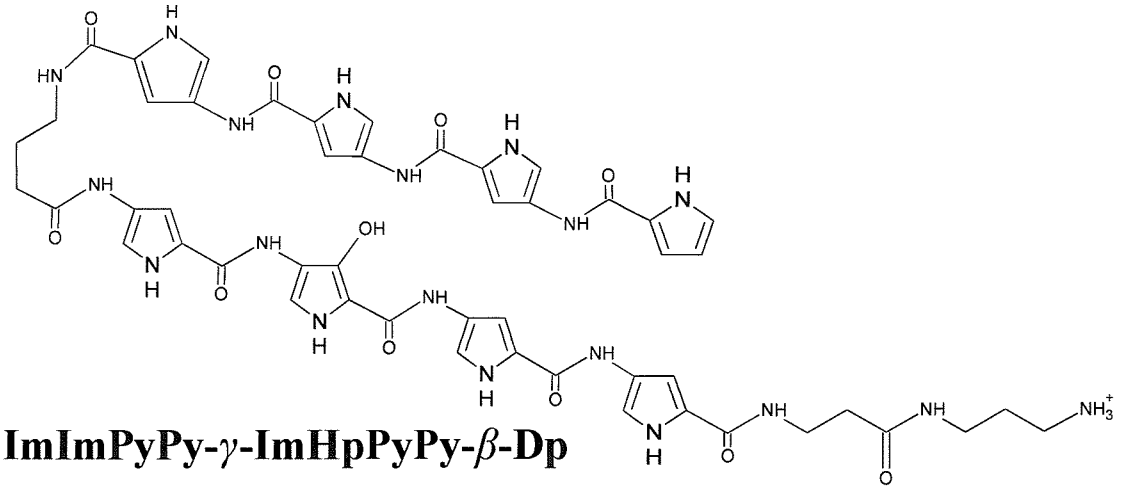
The 2:1 complex of distamycin with DNA led to the development of polyamides. These synthetic agents have a hairpin structure composed of N-methylimidazole (Im),N-methylpyrrole (Py) and N-methyl-3-hydroxypyrrole as building blocks (see Fig.1.13). Their composition is different depending for which target sequence they were developed. The Im/Py pair discriminates a GpC from a CpG, and both from ApT/TpA base-pairs, a Py/Py pair is selective for both ApT and TpA and a Hp/Py pair distinguishes TpA from ApT and GpC/CpG base-pairs (Kielkopf *et al.*, 2000; Urbach *et al.*, 1999). Also, to enhance the DNA binding affinity and specificity of these compounds a C-terminal beta-alanine residue is added preceding a dimethylpropylamine (Dp) group (Parks *et al.*, 1996 Kielkopf *et al.*, 2000).

1.3.2 Hoechst 33258

Hoechst 33258 is a synthetic drug which binds in the minor groove of DNA and consist of four rings linked by flexible bonds, the rings are in the order phenol-benzimidazole-benzimidazole-piperazine (Fig 1.14). Hoechst 33258 has been used clinically for the treatment of pancreatic carcinoma (Kraut *et al.*, 1991), and it serves as a precursor for the development of novel anti-cancer drugs. It is also employed in the detection of DNA and chromosomes in conjuction with the light microscope (Araki *et al.*, 1987) by virtue of its fluorescent properties when bound to B-DNA. However, recent evidence suggests that upon forming a very stable DNA-duplex-Hoechst complex, the fluorescent activity



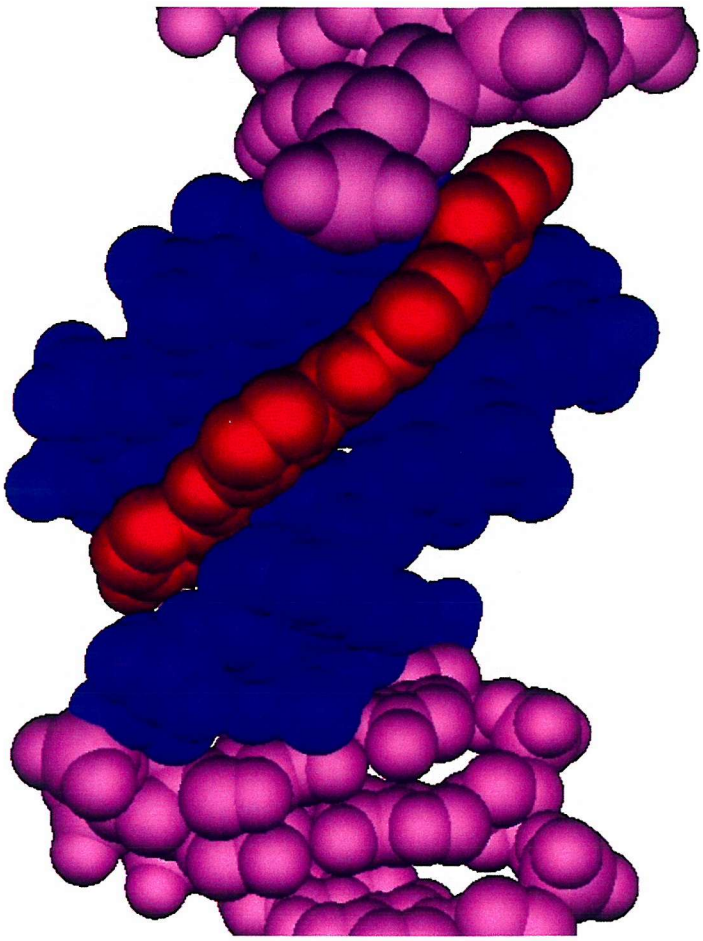
ImImPyPy-γ-ImPyPyPy-β-Dp



ImImPyPy-γ-ImHpPyPy-β-Dp

Figure 1.13 Molecular structure of two polyamides

A)



B)

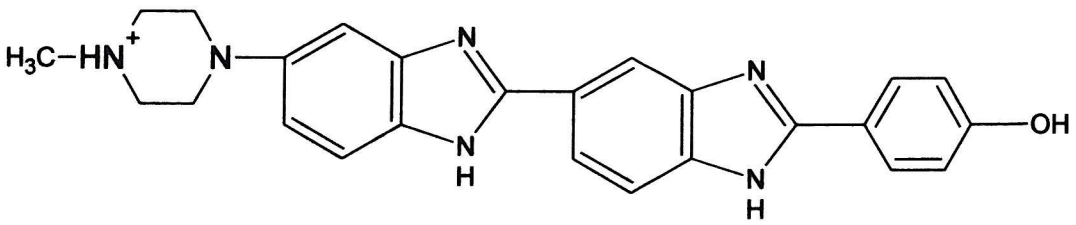


Figure 1.14 A) Crystal structure of Hoechst 33258 (red) bound to AAATTT (blue), the view is through the minor groove. The Image was obtained from the pdb file submitted by Vega, M. *et al.*(1994), using the WebLab ViewerLite and rendered with Pov-Ray for windows. B) Molecular structure of Hoechst 33258.

disappears and consequently the DNA concentration determined by this method is not accurate (Loontiens *et al.*, 1991).

Hoechst has also been found to inhibit the activity of both topoisomerase I (Cheng *et al.*, 1993) and topoisomerase II (Woynarowski *et al.*, 1989).

1.3.2.1 Sequence selectivity

Footprinting studies using methidiumpropyl-EDTA Fe(II) as a cleaving agent, demonstrated that Hoechst has a DNA binding region of 4 to 5 base pairs in length, and like distamycin and netropsin has a preference for binding to AT-rich regions which can be flanked by GC base pairs (Harshman and Dervan, 1985) . Another footprinting study with DNase I has shown that Hoechst does not bind every AT site equally, its sequence selectivity follows the ranking order: AATT > AAAA > ATTA / TAAT = TTAA = ATAT > TATA (Abu-Daya *et al.*, 1995). In addition, the best binding sites were also ranked by a combinatorial method which involves the use of a microchip. In this experiment the results were AATT > AAAT > AAAA > AATA > TAAT > ATAT > TAAA > ATAA > TTAA > TATA which are in agreement with the ranking before mentioned (Drobyshev *et al.*, 1999). Electric linear dichroism experiments show that Hoechst cannot bind to GC rich regions, as a minor groove binding agent, because the 2-amino group of the guanine bases impairs binding in this groove. Instead Hoechst changes its mode of binding and becomes an intercalating agent (Bailly *et al.*,1993).

Spectrophotometric studies show that Hoechst binds very tightly to AT rich sequences, with an association constant of the order of 6x 10^8 M^{-1} at 25 °C on calf thymus DNA (Loontiens *et al.*, 1991).

1.3.2.2 Mode of Binding

X-ray crystallography (Quintana *et al.*, 1991) shows that Hoechst binds DNA in a 1:1 ratio. NMR and crystal structure studies show that the stability of binding of Hoechst and AT rich DNA sequences is dependent on hydrogen bonds between the NH groups of the

benzimidazole rings and the O2 of thymine and the N3 of adenine (Searle and Embrey, 1990; Clark *et al.*, 1996). In addition there are van der Waals contacts between the adenine bases H2's and the deoxyribose O4's and the drug.

A recent NMR study (Gavathiotis *et al.* 2000) examines the formation of a 2:1 Hoechst/DNA complex on the sequence $d(CTTTTGCAAAAG)_2$ (Fig. 1.15). Hoechst, unlike distamycin, does not stack the two molecules side by side; instead they are accommodated adjacent to each other in a symmetrical orientation without disturbing the symmetry of the duplex; the binding of the two drug molecules is highly co-operative, and this species is the only one present in solution. However, if the sequence is changed to $d(GAAAAGCTTTTC)_2$, then the orientational specificity is lost and there is evidence of multiple complexes being formed. These results led to the conclusion that TpG junctions direct the orientational preference of Hoechst. However an ApG junction does not have distinct structural features within the minor groove to be recognised.

1.3.3 Mithramycin

The anti-tumour agent mithramycin and the related compounds chromomycin and olivomycin constitute the aureolic acid class of antibiotics. Mithramycin has been used in the therapy of disseminated testicular carcinomas (Ream *et al.*, 1968) Paget's disease (Elias and Evans, 1972) and hypercalcemia (Kiang *et al.*, 1979; Mundy *et al.*, 1983; Rosol and Capen, 1987), despite its toxicity. These agents have similar structural features, such as an aglycon moiety with five attached hexopyranoses (Fig 1.16)(Thiem and Meyer, 1979; Thiem and Meyer, 1981; Yoshimura *et al.*, 1988), and act by inhibiting DNA and RNA synthesis *in vivo* usually with some preference for the latter (Fok and Waring, 1972; Kaziro and Kamiyama, 1965). Also it has been shown that mithramycin inhibits RNA synthesis by binding to GC rich sequences in eukaryotic promoters such as the SV40 early promoter (Ray *et al.*, 1989) and c-myc P1 and P2 promoters (Snyder *et al.*, 1991). In addition, it has been shown to inhibit the binding of regulatory proteins, such as the transcription factor Sp1 (Snyder *et al.*,1991).

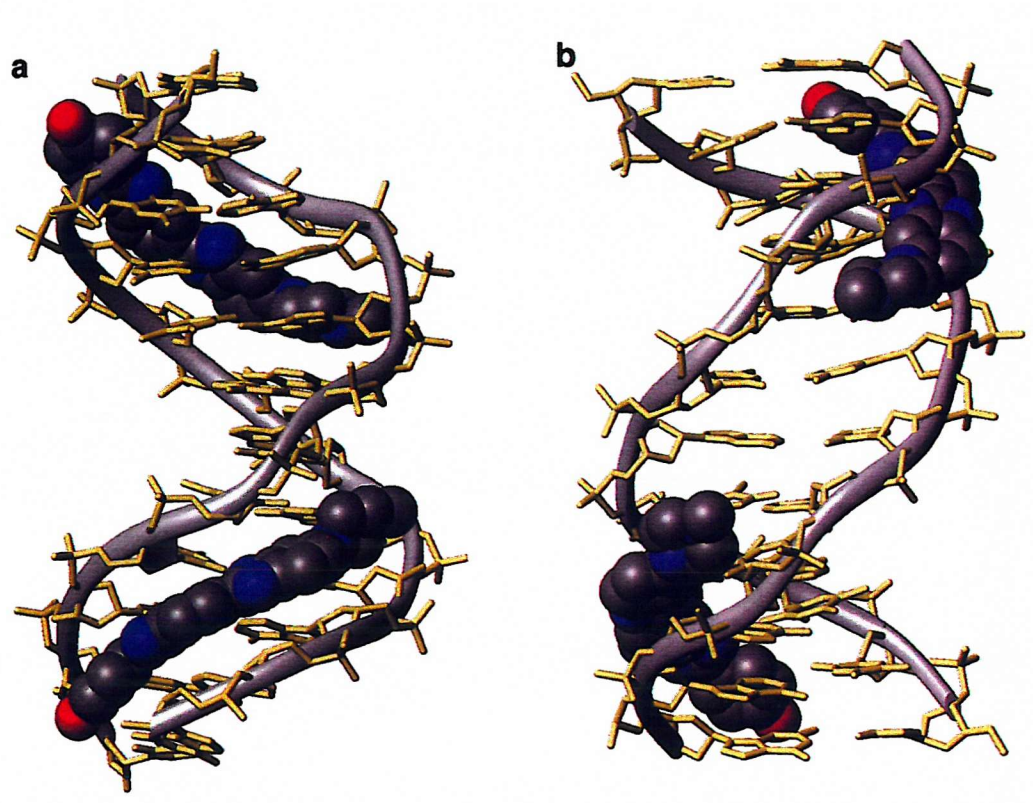


Figure 1.15 3 dimensional structures of the complex between Hoechst 33258 and DNA on d(CTTTTGCAAAAG)$_2$. a) illustrates the change in groove width along the sequence with widening at the junctions with the central GC steps b) View of the minor groove indicating the relative separation of the two N-methypiperazine rings. Taken from Gavathiotis *et al.* (2000)

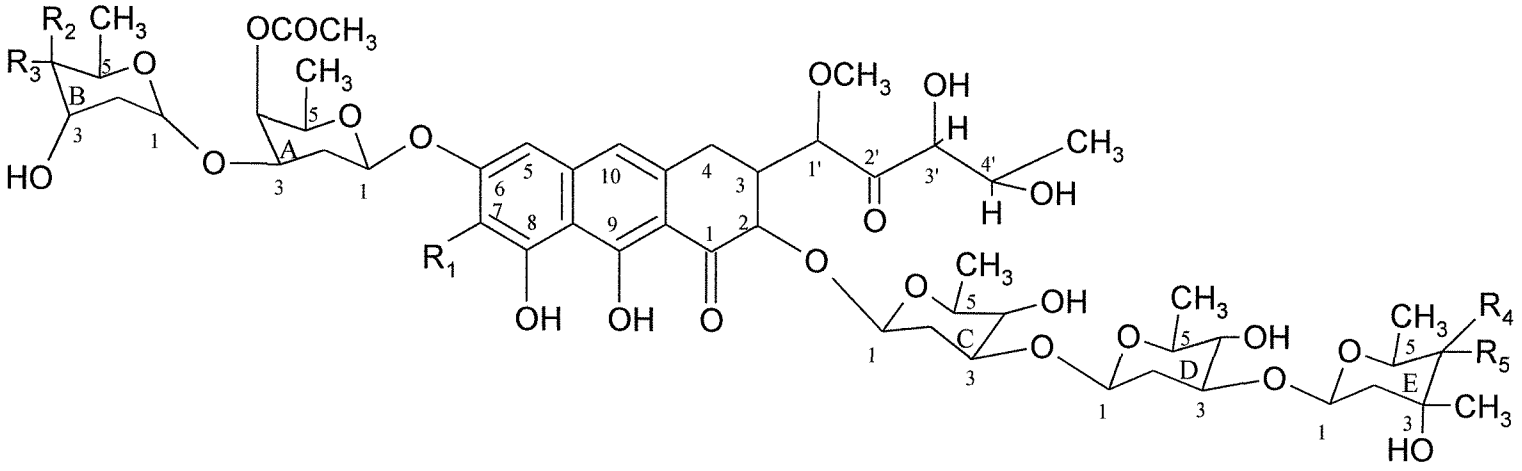


Figure 1.16 The molecular structure of mithramycin and chromomycin A_3
Mithramycin: $R_1 = CH_3$, $R_2 = R_4 = H$, $R_3 = R_5 = OH$.
Chromomycin A_3 : $R_1 = CH_3$, $R_2 = OCH_3$, $R_3 = R_5 = H$, $R_4 = COCH_3$

These agents *in vitro* inhibit RNA and DNA polymerase reactions (Ward *et al.*, 1965) in the presence of Mg^{2+} (Cons and Fox,1989; Nayak *et al.*, 1973). This reaction is specific for magnesium, and it can not be replaced by calcium. Therefore, in the treatment of hypercalcemia (Kiang *et al.* 1979; Mundy *et al.* 1983; Rosol and Capen 1987), the action of mithromycin cannot be explained by this agent binding directly to calcium ions (Cons and Fox, 1989).

1.3.3.2 Sequence selectivity

Footprinting studies using Methidiumpropyl-EDTA/Fe(II) (Van Dyke and Dervan, 1983) showed that the preferred binding sites for mithramycin, olivomycin and chromomycin in the presence of magnesium are at least 3 base pairs long and contain no less than 2 contiguous dG-dC base pairs. The preferred sites for chromomycin with magnesium were in decreasing affinity 3'-GGG, CGA > CCG, GCC > CGA, CCT > CTG-5'(Van Dyke and Dervan, 1983). Circular dichroism experiments using decamers of the form d(GTA-XGCY-TAC) and d(GTA-XCGY-TAC) found that the tetranucleotide preference for chromomycin A was in the order GGCC > CGCG > GCGC, CCGG > AGCT > ACGT, TGCA > TCGA (Liu and Chen, 1994). Other studies using DNase I and DNase II (Fox and Howarth, 1985) and hydroxyl radicals (Cons and Fox, 1989, Carpenter *et al.*, 1993) as probes found that mithramycin seems to recognise GC rich regions with a preference for those containing the GpG (or CpC) step. In addition, chromomycin and olivomycin were found to have similar but not identical binding sites to those of mithramycin, and that these antibiotics bound to DNA tighter than mithramycin (Cons and Fox, 1989). NMR experiments using the DNA sequences d(ATGCAT), d(ATCGAT), d(TATGCATA) and d(ATAGCTAT) showed that mithramycin bound strongly to sequences that contained the GC base pair but no binding was observed for the CG containing oligomer (Banville *et al.*, 1990). The conclusion of this results was that mithramycin and chromomycin bound better to 5'-GC than 5'-CG, and that the best binding sites are 5'-GG, 5'-CC and 5'-GC.

It was observed in some of these footprinting experiments that runs of A and T,

flanking the binding site of the aureolic antibiotics, had enhanced cleavage by DNase I (Fox and Howarth, 1985; Cons and Fox, 1990; Cons and Fox, 1991; Carpenter *et al.,* 1993; Carpenter *et al.,* 1994) and DNase II (Carpenter *et al.*, 1993), an effect which is probably caused by drug-induced DNA structural changes.

1.3.3.3 Mode of Binding

Initially there was a dispute as to whether these aureolic antibiotics bound to the major (Berman *et al.*, 1985; Keniry *et al.,* 1987; Kam *et al.*,1988; Berman and Kam, 1989) or minor groove (Gao and Patel, 1989a; Gao and Patel, 1989b, Banville *et al.*, 1990). However, it has now been established that the minor groove is the site of binding for these agents. Gao and Patel (1989a) showed that chromomycin bound as a dimer to the self complementary sequence d[TTGGCCAA] and that when it was bound to the DNA sequence GGCCGGCC the minor groove became wider and shallower. Later Banville *et al.* (1990) confirmed that the binding was to the minor groove with a stoichiometry of 2 drug molecules:1 hexamer duplex: 1 magnesium ion and that the DNA changed conformation from B-DNA to A-DNA (Fig 1.17), thus widening and deepening the minor groove. Therefore, these results show that chromomycin causes a deformation of the DNA and that it binds to DNA in the minor groove. The widening of the minor groove was also suggested for mithramycin (Cons and Fox, 1990) which was found to bind to $d(ACCCGGGT)_2$ at ratios of 4 drug molecules:1duplex forming dimer-dimer interactions (Keniry *et al.*, 1991; Keniry *et al.,* 1993). Mithramycin binds as a dimer to the DNA by forming symmetrical sequence specific hydrogen bonds to the 2-amino group of guanine with the hydroxyl group of the aglycon C^8 (Fig 1.16)(Sastry and Patel, 1993). The CDE trisaccharide of mithramycin changes its conformation in order to allow the two dimers to bind to the minor groove of DNA, demonstrating that the CDE-trisaccharide plays an important role in determining the binding selectivity of this agent (Keniry *et al.,* 1993). These trissacharides are located within the minor groove, thus interacting with both strands of the duplex. The CD disaccharide interacts with its hydrophobic edges while the E saccharide interacts with its hydroxyl bearing face and also forms hydrophobic molecular interactions. In addition, the AB dissacharide and the hydrophilic side chain form intermolecular contacts with the sugar-

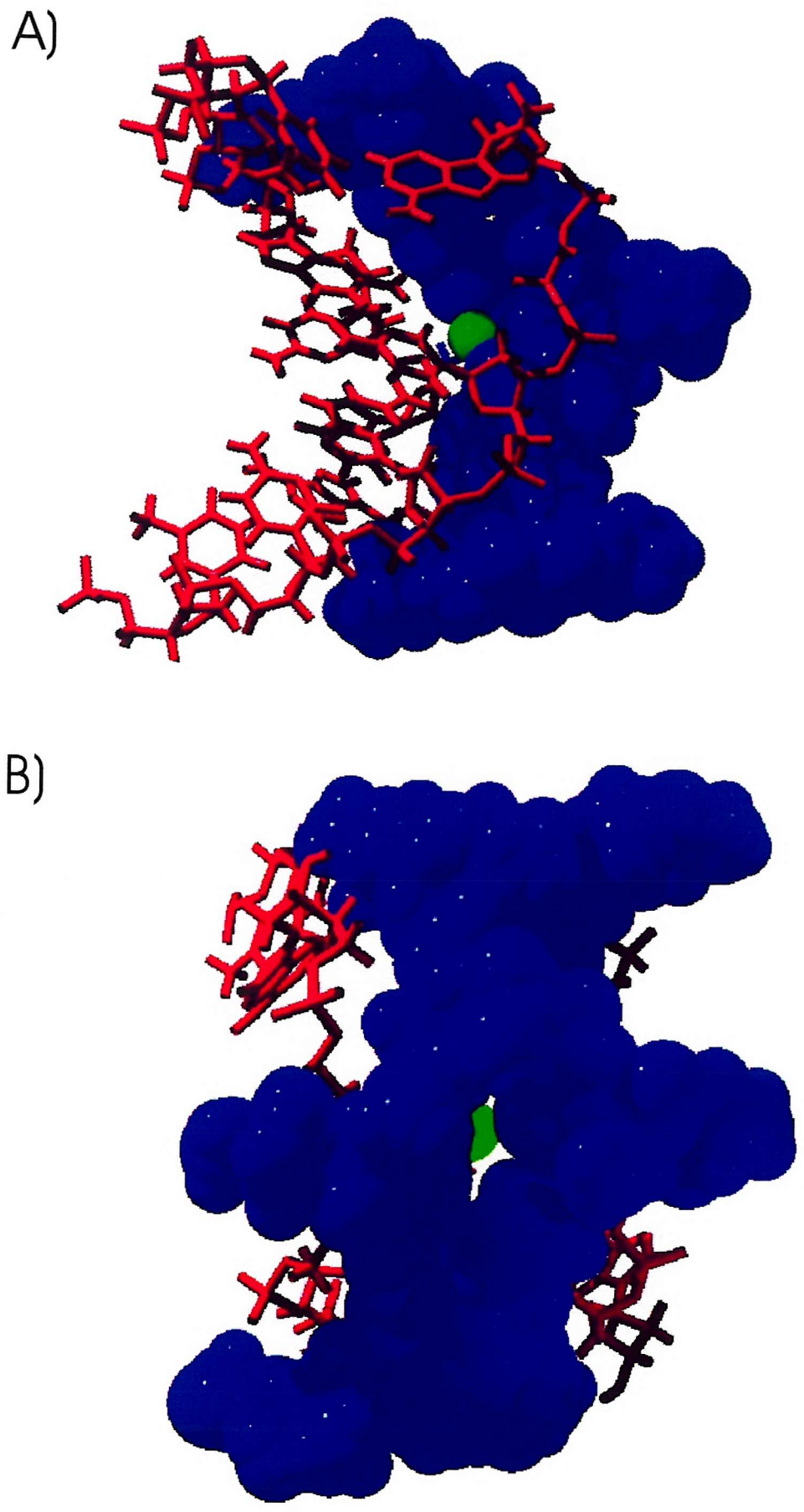


Figure 1.17 Crystal Struture of mithramycin (blue), bound to the DNA sequence (5'-TCGCGA-3')(red). A) view from along the phosphate backbone B) is a view from the minor groove. The Mg ion is a green sphere. Images were created in Swiss-Pdb viewer, from the pdb file 146D submitted by Sastry and Patel (1993), and rendered in Pov-ray for Windows.

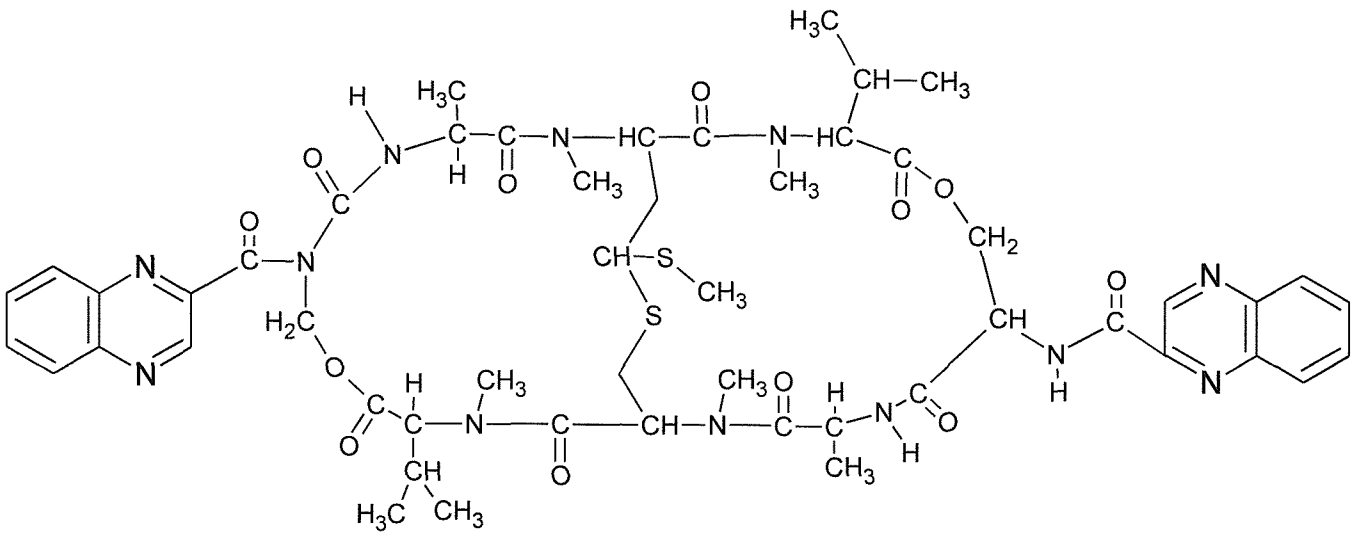
phosphate backbone. (Sastry and Patel, 1993)

1.4 QUINOXALINE ANTIBIOTICS

The quinoxaline antibiotics were first isolated in the 1950's, they are natural compounds produced by a wide variety of streptomycetes (Katagiri *et al.*, 1975). This class of antibiotics has a structure characterized by two moieties of quinoxaline-2-carboxylic acid attached to a cross-bridged octadepsipeptide ring containing L- and D- amino acids (Katagiri *et al.*, 1975; Lee and Waring, 1978a). These antibiotics are divided into two families depending on the structure of the cross-bridge. The quinomycins have a thioacetal cross-bridge while the triostins have a disulphide. The triostins may be the biosynthetic precursors of the quinomycins (Lee and Waring, 1978a). The quinoxaline antibiotics are active against Gram positive bacteria, viruses and diverse experimental tumours (Katagiri *et al.*, 1975). The best-known member of the quinomycins is echinomycin (Fig.1.18) and of the triostins is triostin A (Fig.1.19). In each of the natural antibiotics four of the peptide bonds are methylated. These methyl groups are absent in the related synthetic derivative des-N-tetramethyl-triostin A (designated TANDEM) (Fig.1.20a). (N-MeCys^3,N-MeCys^7) TANDEM is a derivative of TANDEM in which two of the N-methyl groups are retained and which has very similar binding properties(Fig 1.20b).

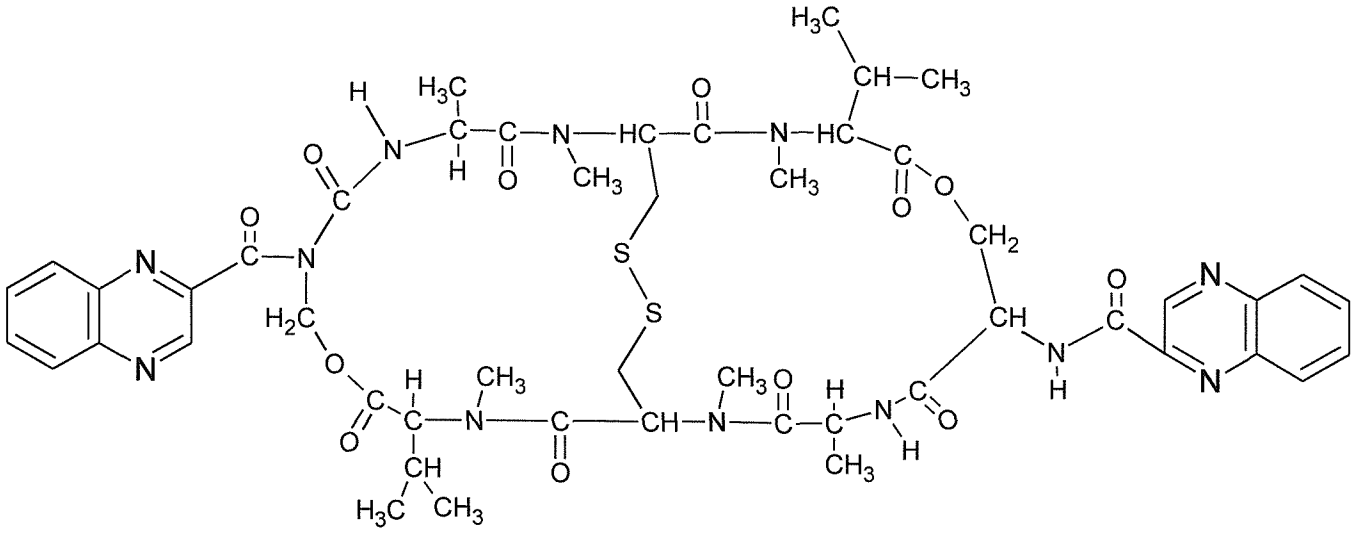
1.4.1 Bifunctional intercalation

Echinomycin, at low ionic strength, was found to remove and reverse the supercoiling of circular DNA with an unwinding angle almost twice that of the monofunctional intercalator ethidium, assessed by sedimentation studies. Viscosity measurements with sonicated rod-like DNA fragments showed an extension of the double helix of 6.3 Å per bound echinomycin molecule, which is again double that of ethidium. Thus, from this work it was concluded that echinomycin is a bifunctional intercalator which functions by inserting both quinoxaline chromophores into the DNA helix (Waring and Wakelin, 1974). All the other quinoxaline antibiotics have been shown to possess this mode of binding to DNA (Lee and Waring, 1978a).



Echinomycin

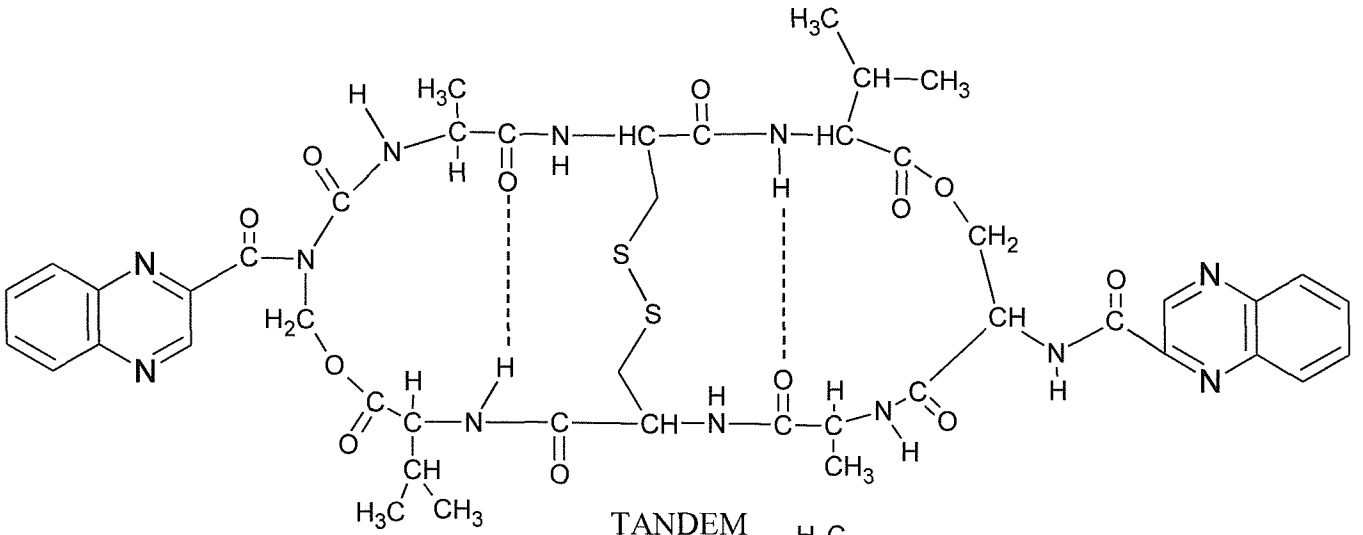
Figure 1.18 Molecular structure of echinomycin



Triostin A

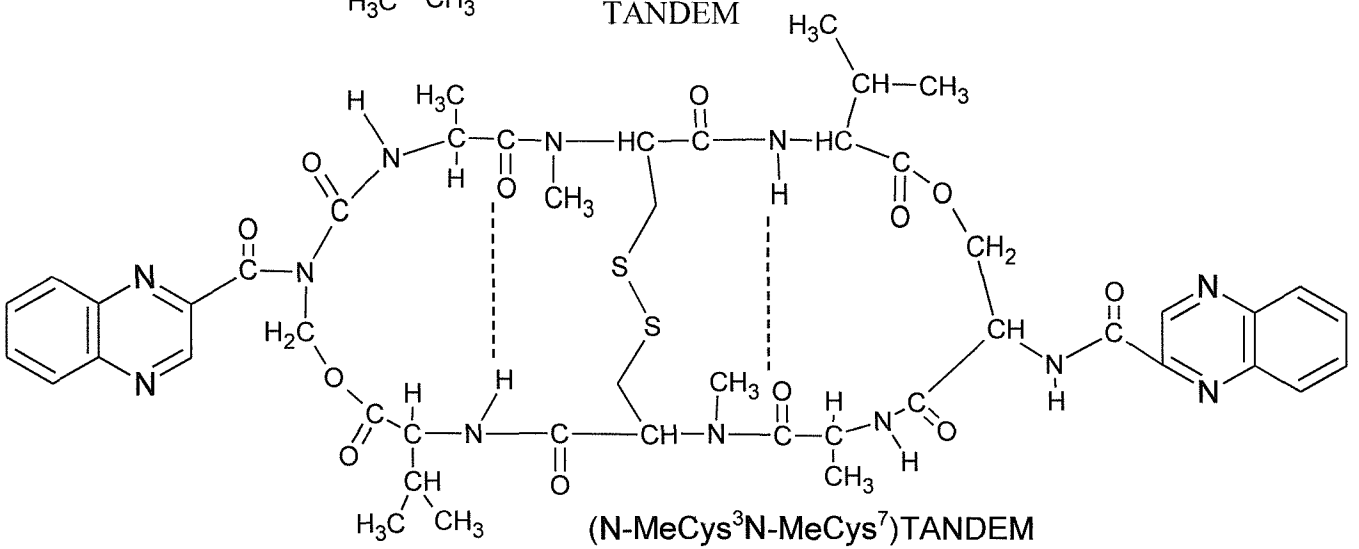
Figure 1.19 Molecular structure of Triostin A

A)



TANDEM

B)



(N-MeCys^{3}N-MeCys^{7})TANDEM

Figure 1.20 Molecular structures of A) TANDEM and B) its derivative (N-MeCys^{3} N-MeCys^{7})TANDEM

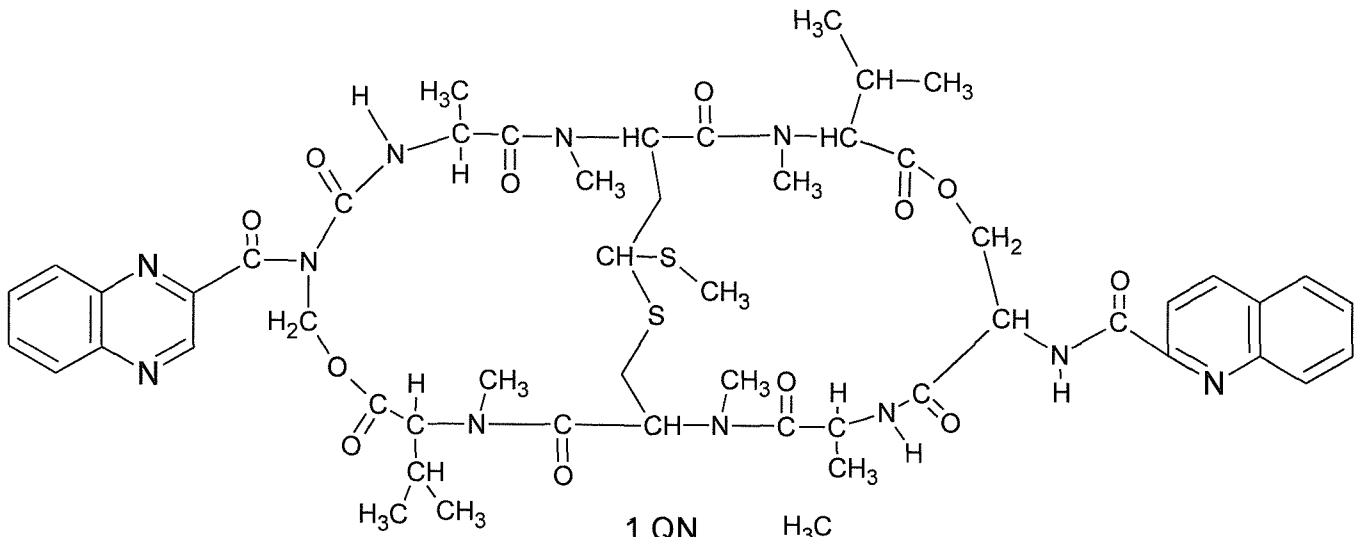
1.4.2 Structure-activity relationships

The availability of several natural and synthetic analogues of the antibiotics provides the possibility of studying their structure-activity relations for DNA binding. Lee and Waring (1978b) used a derivative of echinomycin with a broken thioacetal cross-bridge, but intact lactone linkages, and found that this cross-bridge was essential for echinomycin binding. Fox *et al.* (1980a) used an analogue of TANDEM that has the two Cys residues replaced by Ala so that the cross-bridge is totally lacking (designated [Ala^3,Ala^7]TANDEM), and demonstrated that although this compound could still bind to DNA with the same selectivity as the parent drug its binding constant was three orders of magnitude lower. These results suggested that the cross-bridge is not absolutely essential for binding to DNA , though it evidently increases the binding strength significantly by maintaining the integrity of the peptide ring (Lee and Waring, 1978b). In addition, if one or both of the D-amino acid centres bearing the quinoxaline chromophores is epimerized to the L-isomer the molecule is inactivated (Fox *et al.*, 1980a) as well as when the chromophores are removed. Several chromophore substitutions have been made such as 1QN and 2QN (Fox *et al.*, 1980b; Fig.1.21) to study the nature of the binding to the DNA helix. 1QN and 2QN (possessing one and two quinoline chromophores replacing the natural quinoxalines) showed a stronger binding to DNA; with an altered binding preference suggesting that the chromophores are not merely "inert plugs that serve only to help locate the peptide portion of the molecule within one of the grooves of the DNA helix after the fashion of a staple"(Fox *et al.*,1980b). Thus, the conclusion from all these results is that the binding energy of the quinoxaline antibiotics does not derive from any one identifiable structural feature but depends upon the disposition of substituents throughout the whole molecule, involving both chromophores as well as the peptide portion (Fox *et al.*, 1980a).

1.4.3 Conformation studies

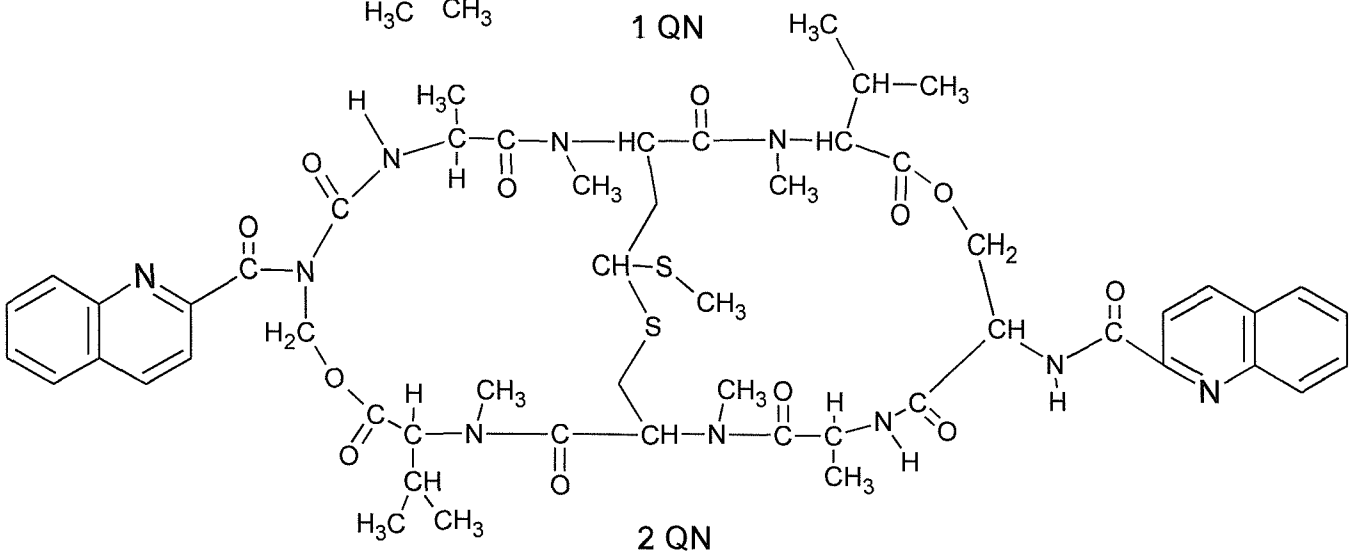
In order to explain the bifunctional intercalation of the quinoxaline antibiotics their structure should follow a set of requirements 1) the chromophores have to lie on the same side of the peptide ring; 2) their planes should be more or less parallel; 3) the vertical distance between those planes should be an integral multiple (X) of 3.4 Å, i.e. the

A)



1 QN

B)



2 QN

Figure 1.21 Molecular structures of A) 1 QN and B) 2 QN

theoretical spacing required to accommodate X base pairs; 4) the space between the chromophores should be essentially free from obstruction by any other substituents attached to the peptide ring (Sheldrick *et al.*, 1984).

1.4.3.1 Echinomycin

For echinomycin (Fig. 1.22) the results obtained by NMR spectroscopy (Cheung *et al.*, 1978) showed that the octapeptide ring behaves as a more or less rigid disk. The two quinoxaline chromophores are separated by 10 Å, which is almost exactly the correct distance to accommodate two sandwiched DNA base-pairs and satisfies the neighbour-exclusion hypothesis (Cairns, 1962), and they extend outwards almost perpendicular to the plane of the octapeptide disk, with their own planes nearly parallel, satisfying all the prerequisites of a bifunctional intercalator.

1.4.3.2 TANDEM

The crystal structure of TANDEM (Fig.1.23) has been determined (Viswamitra *et al.*, 1981; Hossain *et al.*, 1982) and the most significant features are as follows. The octapeptide ring is a relatively rigid structure and the chromophores are attached to it at opposite ends; the distance between them is about 6.5 Å even though the separation of the D-Ser units is 12.2 Å, which is not enough for bifunctional intercalation. However, if the molecule rotates by 20° about the D-Ser C-N bonds then the required separation of 10 Å is achieved to allow two DNA base-pairs to be accommodated (Fig1.24). The molecule contains two internal hydrogen bonds connecting the alanine carbonyls with the valine NH's. This bond is not possible in natural antibiotics since the valine NH's are methylated. This difference appears to be very important for determining the sequence selectivity of these ligands.

1.4.4 Sequence selectivity

1.4.4.1 Echinomycin

Early studies with natural DNAs, of varying base composition, demonstrated that

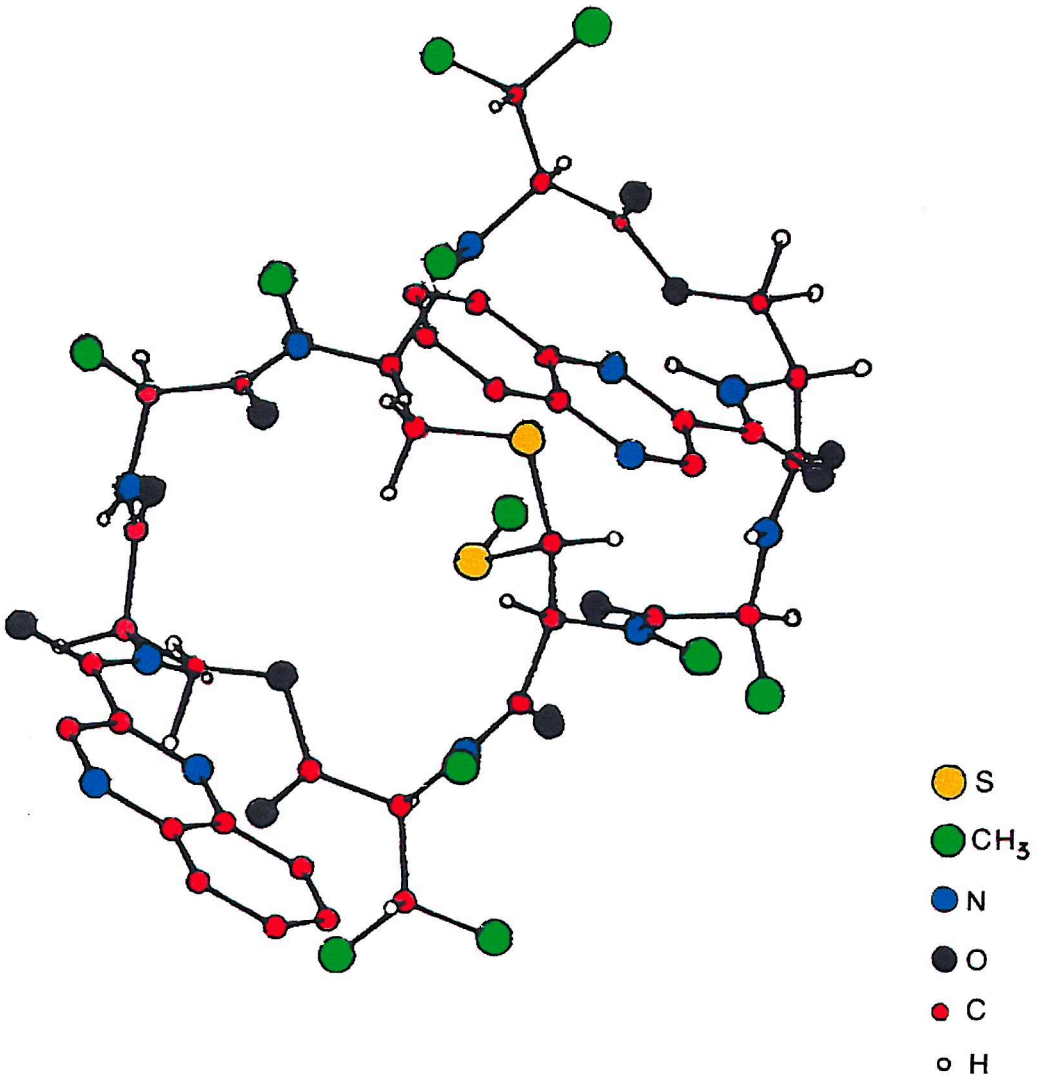


Figure 1.22 Crystal structure of echinomycin taken from Cheung *et al.* 1978.

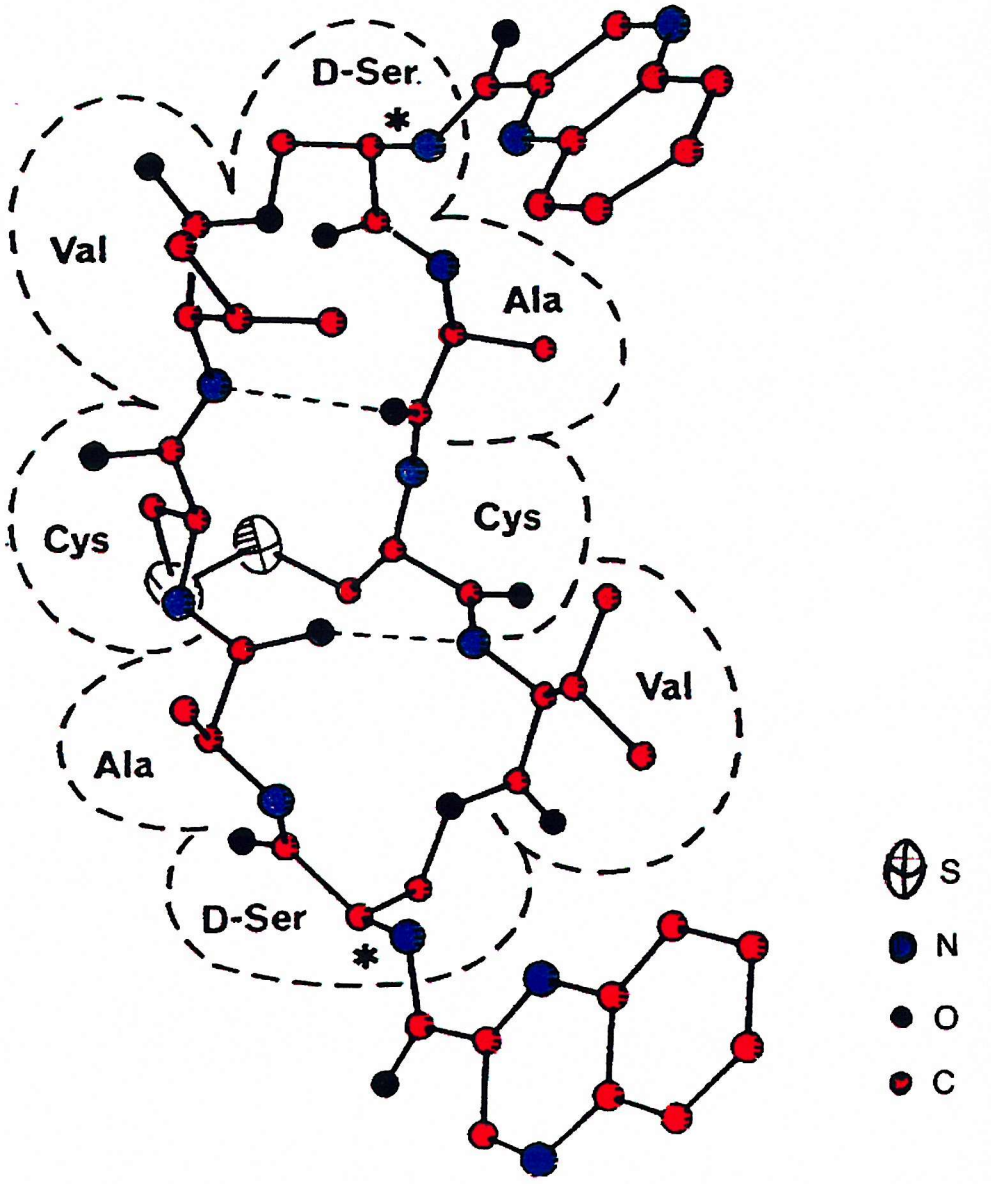
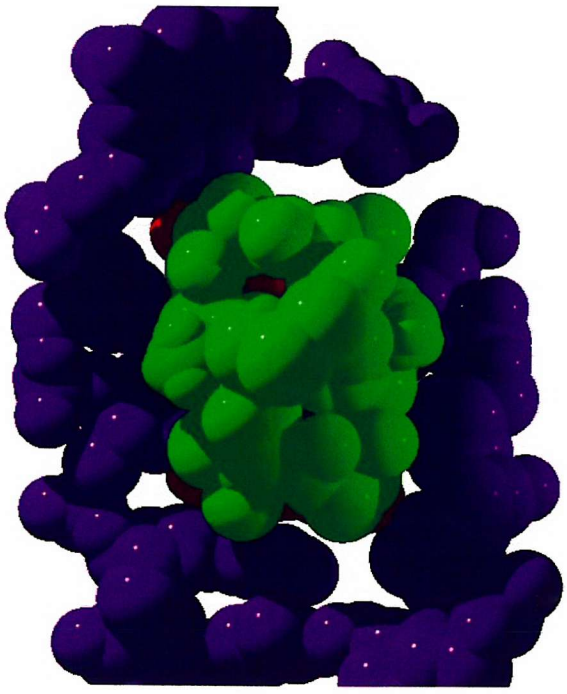


Figure 1.23 Crystal structure of TANDEM taken from Viswamitra *et al.*, 1981

a)



b)

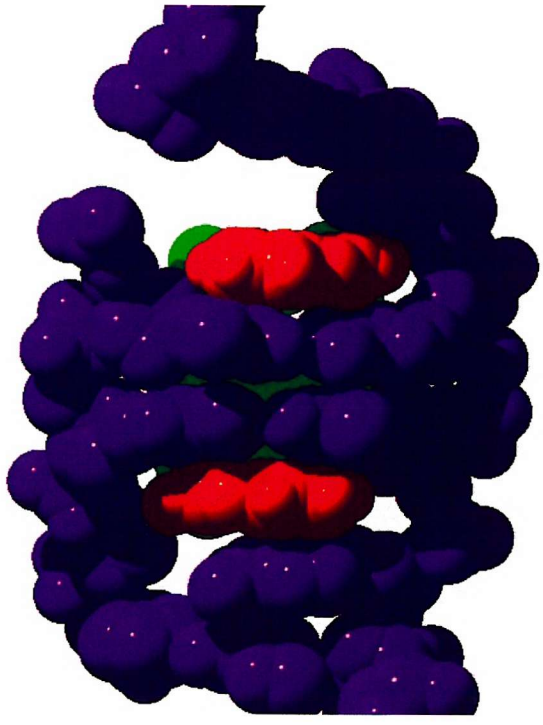


Figure 1.24 TANDEM bound to the DNA sequence [d(GATATC)]2 a) viewed through the minor groove. b) viewed through the major groove. Color coding DNA purple, TANDEM green and the chromophores red. Images were created in Swiss-Pdb viewer, from the pdb file 2DA8 submitted by Addess *et al.* (1993) , and rendered in Pov-ray for Windows.

echinomycin displayed a clear preference for binding to GC rich DNAs (Wakelin and Waring, 1976) with an association constant of 0.3 x 10^6 M^{-1} for *C. perfringens* DNA (30% G+C) compared to 3.1 x 10^6 M^{-1} for *M. lysodeikticus* (72% G+C). The preferred binding sites of echinomycin are centred around a CpG step as judge by footprinting studies using DNases I and II (Low *et al.*, 1984b) and methidiumpropyl-EDTA-Fe(II) (MPE-Fe(II)) (van Dyke and Dervan, 1984). The MPE-Fe(II) studies revealed that the binding site is composed of 4 base pairs and that the strongest binding sites are 5'-3' TCGT and ACGT. It was also noticed that some regions adjacent to the binding site are more sensitive to cleavage by the probes in the presence of the antibiotic than that of the free DNA, suggesting that these regions might have an altered conformation from B-DNA. The elucidation of the crystal structures of ligands bound to short oligonucleotides (Ugetto *et al.*, 1985) has improved our knowledge of the interaction between these antibiotics and their DNA targets . The first such structure was determined for echinomycin bound to CGTACG (Ugetto *et al.*, 1985). Two echinomycin molecules are bound per duplex, sandwiching the CpG steps. Specific hydrogen bonds are formed between the alanine carbonyls and the 2-amino group of guanines, as well as between the alanine NHs and N3s of guanine. These contacts were sufficient to explain the selectivity for the dinucleotide CpG. A surprising feature of this structure was that the central AT base pairs were found in the Hoogsteen conformation, rather than in the Watson-Crick base pairing. A similar effect was also noted for the extra GC pairs in the structure of the drug bound to GCGTACGC (Wang *et al.*, 1986).

The 2-amino group of guanine has been shown to be essential for echinomycin binding. This was found by footprinting experiments in which the guanine bases were replaced by inosine on either one or both strands. These results showed that echinomycin did not bind to CpI in either case, demonstrating that even when one 2-amino group is absent the binding of echinomycin to DNA is abolished (Marchand *et al.*, 1992). Similar experiments were conducted in which the adenine was substituted by 2,6 diaminopurine (DAP), and showed excellent echinomycin binding, confirming the important role of the 2-amino group. Indeed echinomycin bound tighter to TpDAP in the doubly substituted DNA

than to CpG sites in natural DNA (Bailly and Waring, 1995).

Echinomycin (and Triostin A) increase the susceptibility of purines on the 3'-side of CG to modification by diethylpyrocarbonate (DEPC)(Portugal *et al.* ,1988; Mendel and Dervan, 1987; Jeppesen and Nielsen, 1988) . Although this was first interpreted as arising from the formation of Hoogsteen pairs (Mendel and Dervan, 1987), this now seems unlikely since DEPC modifies N7 of purines, a portion which is involved in hydrogen bonding of Hoogsteen base pairs (Portugal *et al.*, 1988; Mclean and Waring, 1988). This is now interpreted as arising from a local unwinding of the DNA duplex (Jeppesen and Nielsen, 1988). This effect is specially noteworthy for CGA.

Conformational studies of echinomycin-DNA binding have been examined by NMR spectroscopy; Gao and Patel (1988) compared the interaction of echinomycin with d(ACGT) and d(TCGA) and showed that the A1*T4 base pairs adopted Hoogsteen pairing for the echinomycin-d(A1-C2-G3-T4) complex while the T1*A4 base pairs adopted Watson-Crick pairing for the echinomycin-d(T1-C2-G3-A4) complex in aqueous solution. However these results did not address the question of Hoogsteen base pair formation in the interior of a DNA duplex. This was studied by Gilbert and co-workers (Gilbert *et al.*, 1989) examining the complex between echinomycin and the DNA octamer $[d(ACGTACGT)]_2$. Their results showed that at low temperatures both the terminal and the interior A*T base pairs are Hoogsteen base paired in the drug-DNA complex in solution. However, as the temperature was raised the internal A*T Hoogsteen base pairs started to exchange between Hoogsteen and open or Watson-Crick base pair conformations, while the base pairs on the ends of the duplex remained in the Hoogsteen conformation up to at least 45°C. Therefore, their conclusion was that *in vivo* the conformation of the A*T base pairs flanking echinomycin binding sites may be in an open rather than a Hoogsteen base paired state. Another DNA octamer $[d(TCGATCGA)]_2$ (Gilbert and Feigon, 1991) showed that the complex with echinomycin did not form Hoogsteen base pairs and that the terminal and internal A*T base pairs were Watson-Crick base paired at all temperatures ranging from 0 to 45°C. One possible explanation for this problem is that the stacking interactions between

the bases and the quinoxaline ring must contribute to the sequence requirements for Hoogsteen base pair formation (Gao and Patel, 1988). The double ring of the adenine might stack better over the quinoxaline ring in the $[d(ACGTACGT)]_2$ complex when the A*T bases are Hoogsteens base paired, but Hoogsteen geometry does not increase the stacking between the T and the quinoxaline ring in complex $[d(TCGATCGA)]_2$ enough to overcome the increased electrostatic repulsion caused by bringing the phosphate groups closer together (Gilbert and Feigon, 1991). In addition the binding of echinomycin to $[d(ACGTACGT)]_2$ showed that echinomycin could bind in a 2:1 ratio and that the binding was highly cooperative (Gilbert and Feigon, 1991). This cooperative binding was also found in sequences that were spaced by four base pairs such as ACGTATACGT (Gilbert and Feigon, 1992). These results were also corroborated by footprinting studies (Bailly *et al.*, 1996). The extent of echinomycin cooperativity depends on the bases flanking the binding site and on the distance between the CpG steps. In addition, it is thought that the cooperativity is primarily due to the DNA's structural changes induced by echinomycin upon binding (Bailly *et al.*, 1996).

1.4.4.2 Triostin A

The cleavage inhibition patterns (footprints) with DNase I (Low *et al.*, 1984a) of triostin A are very similar to those obtained for echinomycin, and revealed that all its binding sites contained the dinucleotide step CpG. The preferred binding sequence is XCGY, where X and Y are either purine or pyrimidine. This close similarity in the footprints for both antibiotics is surprising because other experiments designed to measure their binding affinity to a variety of synthetic and naturally occurring DNAs (Lee and Waring, 1978a) demonstrated that triostin A, although possessing a higher binding constant for *M. lysodeikticus* DNA (72% G+C content) than for *Clostridium perfringens* DNA (30% G+C content) bound better to poly(dA-dT) than to poly(dG-dC). The reverse is true in the case of echinomycin. However, these polynucleotides may adopt different helical structures as compared to natural DNAs (Klug *et al.*, 1979; Cohen *et al.*, 1981). These results exaggerate small differences in the intrinsic sequence preferences of the antibiotics. The crystal structure of Triostin A (Fig. 1.25) takes a form of a cradle with the quinoxaline

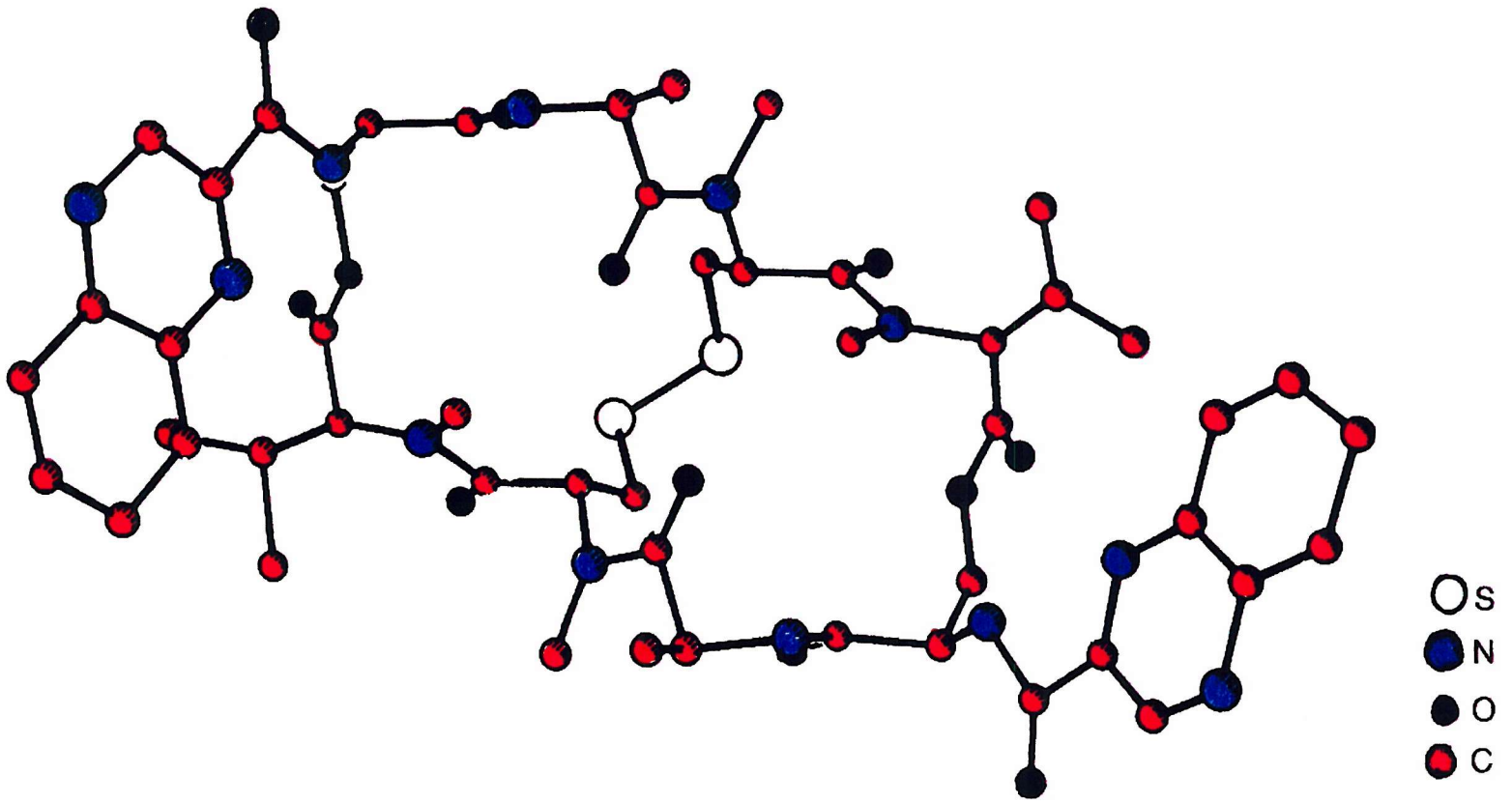


Figure 1.25 Crystal structure of Triostin A taken from Sheldrick *et al.*, 1984

chromophores, Ala methyls, and Val N-methyls projecting upwards from the edge. Assuming there is free rotation about the Ser N-C bonds, then the planes of the quinoxaline chromophores could become parallel and adjust their separation to 10.2 Å and accommodate two DNA base pairs as echinomycin (Sheldrick *et al.*, 1984).

1.4.4.3 TANDEM

Equilibrium binding studies with natural and synthetic DNAs have revealed that TANDEM (lacking the four N-methyl groups) binds selectively to DNAs which are rich in AT residues (Lee and Waring, 1978b), in clear contrast to the selectivity of triostin A or echinomycin. The AT-selectivity was confirmed by early footprinting studies, though these could not distinguish between the binding to ApT or TpA. On the basis of the crystal structure of TANDEM Viswamitra and co-workers (Viswamitra *et al.*, 1981) proposed that the drug might bind better to ApT. They noted the presence of the intramolecular hydrogen bonds between the alanine carbonyls and the NH of valine, preventing binding to CpG since no hydrogen bonds can be formed to the 2-amino group of guanine. The authors suggested that the NH atoms of the alanine residues might be positioned pointing towards the DNA helix, in a position ready to form hydrogen bonds to the 2-keto group of thymine, explaining the selectivity for ApT. In contrast it has been reported that poly d(TAC)* poly d(GTA) binds TANDEM while poly d(ATC)* poly d(GAT) does not, which suggest that the drug recognises TpA rather than ApT (Evans *et al.*, 1982). Footprinting experiments with other TANDEM analogues confirmed the possibility of binding to TpA. [N-MeCys^{3}, N-MeCys^{7}]TANDEM shows the same preference as TANDEM revealing that the AT or GC selectivity does not arise as a direct consequence of methylation of the Cys-amide bonds (Low *et al.*, 1986). In addition, another analogue containing lactone linkages instead of alanine NHs did not bind DNA, confirming the requirement for the alanine NH group (Low *et al.*, 1986). Nonetheless these studies could not rigorously distinguish between binding to ApT or TpA. TpA was tentatively suggested to be the binding site since *Rsa 1* (GTAC) is inhibited by TANDEM whereas *Sau* 3A1 (GATC) is not.(Low *et al.*, 1986)

Further footprinting (Waterloh *et al.*, 1992, Lavesa *et al.*, 1993) and NMR

spectroscopy (Addess *et al.*, 1993) studies using [N-MeCys^3, N-MeCys^7]TANDEM, were performed to establish the binding selectivity of this agent. All of these suggest TpA as the recognition site. More recent work (Fletcher *et al.*, 1995) has shown that the binding stability of [N-MeCys^3, N-MeCys^7]TANDEM with isolated TpA steps decreases in the order ATAT > TTAA > TATA, and the dissociation from longer runs of AT is much slower than those from the short sites. This indicated that for a strong drug/DNA interaction surrounding A and T base pairs around the TpA site are required.

In order to explain the difference in binding specificities of TANDEM (TpA) and its parent compound triostin A (CpG) several footprinting and NMR studies were performed. NMR experiments with [N-MeCys^3,N-MeCys^7]TANDEM (Addess and Feigon, 1994a) and X-ray crystallography with TANDEM (Viswamitra *et al.*, 1981) found that these ligands formed two intramolecular hydrogen bonds between the Val NH protons and the Ala carbonyls. However, these hydrogen bonds cannot be formed in triostin A because of the N-methyl substituents attached to both Val amides. These internal hydrogen bonds explain the reason why TANDEM and [N-MeCys^3, N-MeCys^7]TANDEM do not bind to CpG steps, as these bonds would be required to break in order to accommodate the guanine 2-amino group, and this would be energetically unfavourable (Addess and Feigon, 1994a). TANDEM and [N-MeCys^3,N-MeCys^7]TANDEM form hydrogen bonds between the alanine NH and adenine N3 (Addess *et al.*, 1993). These hydrogen bonds have been shown to be essential for the binding of these agents to DNA (Olsen *et al.*, 1986). Although this hydrogen bonding pattern explains why TANDEM can bind to TpA it does not explain why the ligand cannot bind to other YpR steps, interacting with N3 of guanine. This is shown by a footprinting study in which the adenines were substituted by 2,6 diaminopurine (DAP) (Fig 1.7), in which TANDEM bound to TpDAP as well as it bound to TpA. These data led to the idea that the selectivity of TANDEM for TpA sites was probably due to stacking and hydrophobic interactions and/or maybe due to an indirect recognition of DNA structure, and not to an obstructive effect of the guanine 2-amino group (Bailly and Waring, 1998) . On the other hand triostin A and echinomycin when bound to DNA form two intermolecular hydrogen bonds between the Ala CO and two guanine 2-amino groups which are critical for

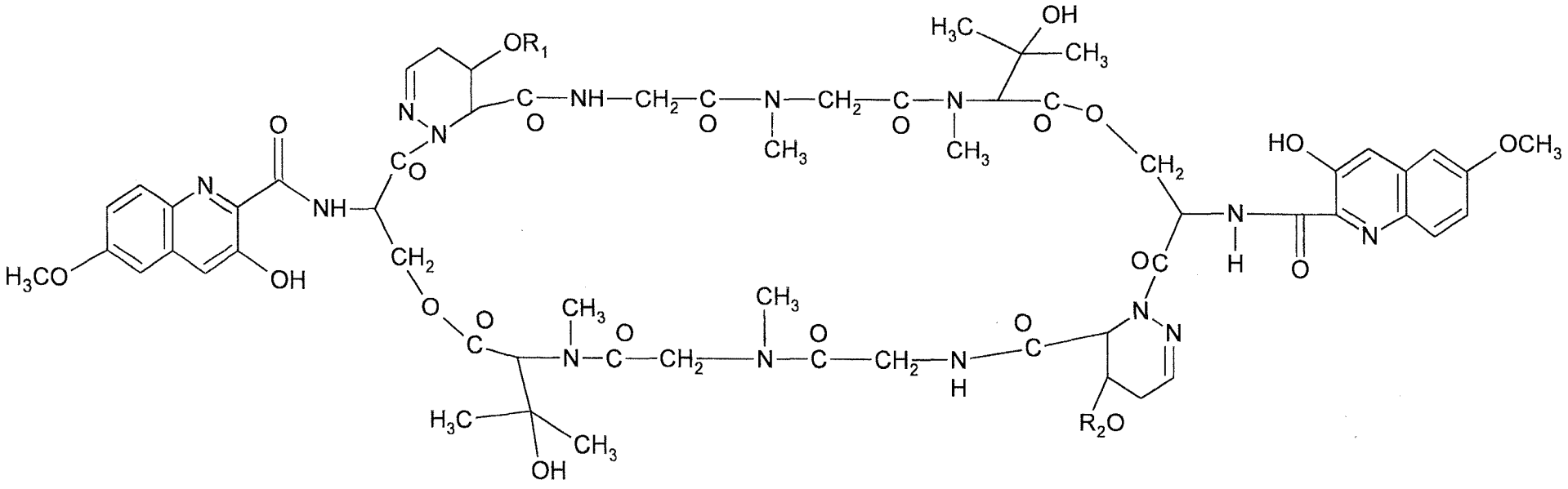
the sequence specific binding (Addess and Feigon, 1994a; Bailly and Waring, 1998; Marchand *et al.*, 1992) to CpG in solution. Triostin A was found to bind less tightly to CpG than [N-MeCys^3,N-MeCys^7]TANDEM does to a TpA step suggesting that the intercalation of the chromophores into the CpG step is not energetically favourable; this is shown by triostin A requiring four intermolecular bonds (two bonds between the Ala CO and the 2-amino group of guanine and two intermolecular hydrogen bonds between the Ala NH and the guanine N3) to stabilize the complex between triostin A and [d(GACGTC)]2, (Addess and Feigon, 1994a). In addition, footprinting data showed that triostin A bound better to TpDAP than to CpG suggesting that probably the intercalation of the quinoxaline rings into TpDAP is more energetically favourable than to CpG (Bailly and Waring, 1998).

NMR studies conducted with [N-MeCys^3,N-MeCys^7]TANDEM and d(GGACITCC)$_2$ showed that [N-MeCys^3, N-MeCys^7]TANDEM bound to CpI regions even though the binding was less tight than that to TpA sites (Addess and Feigon, 1994b). In contrast, a footprinting study using inosine (I) (Fig 1.7) found that TANDEM did not bind to TpI or CpI steps (Bailly and Waring, 1998). This difference was explained on the basis of drug concentration, due to NMR experiments requiring a higher concentration of DNA and ligand compared to footprinting experiments. However, still the question remains as to why TANDEM binds preferentially to TpA and not to CpG (Bailly and Waring, 1998).

1.4.5 Luzopeptins

The luzopeptins are a related class of antitumour antibiotics isolated from fermentation broths of an aerobic strain of actinomycete (Ohkuma *et al.*,1980; Tomita *et al.*, 1980). They contain two substituted quinoline chromophores linked by a decadepsipeptide (Huang *et al.*, 1980) (Fig.1.26).

Luzopeptin has been shown to bind to DNA by bisintercalation, assessed by viscometric and fluorometric studies (Arnold and Clardy, 1981), in a manner similar to the quinoxaline antibiotics. The two classes of antibiotics are related although there are some



Luzopeptin A $R_1 = R_2 = COCH_3$
Luzopeptin B $R_1 = H$ $R_2 = COCH_3$
Luzopeptin C $R_1 = R_2 = H$

Figure 1.26 Molecular structures of the Luzopeptins

structural differences, for example, the luzopeptins have ten amino acids in the cyclic peptide as opposed to eight in the quinoxaline family. In addition quinoline chromophores are substituted with both methoxy and hydroxyl groups, and they lack a sulphur containing cross-bridge (Fox *et al.*, 1988).

Footprinting studies with DNase I and micrococcal nuclease have demonstrated that the luzopeptins have a binding preference for alternating A and T nucleotides on duplex DNA (Fox *et al.*, 1988). NMR studies have suggested that it recognizes the sequence 5'-ATGC (Searle *et al.*, 1989). However, these luzopeptin antibiotics have very little selectivity and their binding interaction with DNA remains unclear (Bailly *et al.*, 2000).

1.5 ACTINOMYCIN D

Actinomycin D is a chromopeptide composed of two pentapeptide lactones and a phenoxazone chromophore (Fig.1.27). It is used clinically to treat tumours such as Wilms' tumour, gestational choriocarcinoma and rhabdomyosarcoma. This agent is very cytotoxic causing apoptosis in some cells. Therefore, the search for a less cytotoxic agent having the same anticancer properties of Actinomycin is in progress. Actinomycin's pharmacological action is due to its interactions with DNA and its inhibition of DNA-dependent RNA synthesis.

1.5.1 Sequence selectivity

Actinomycin was found to be selective for GpC sites as suggested by binding studies (Cerami *et al.*, 1967) and by enzymatic footprinting with DNase I (Fox and Waring, 1984) and MPE-Fe(II) (Van Dyke *et al.*, 1982). However, not all GpC sites bound with the same affinity because of the effects due to the flanking base pairs. Thus, the binding strength as measured by dissociation kinetics was found to decrease in the following order TGCA > CGCG > AGCT >> GGCC (Chen, 1989). In addition, footprinting experiments show that actinomycin D dissociates faster from sites flanked by $A_n{\cdot}T_n$ than $(AT)_n$. In dissociation kinetics within alternating AT regions, TGCA is a better binding site than AGCT, and CGCA is better than GGCA; GpC sites flanked by $(AC)_n{\cdot}$ $(GT)_n$ are good binding sites and

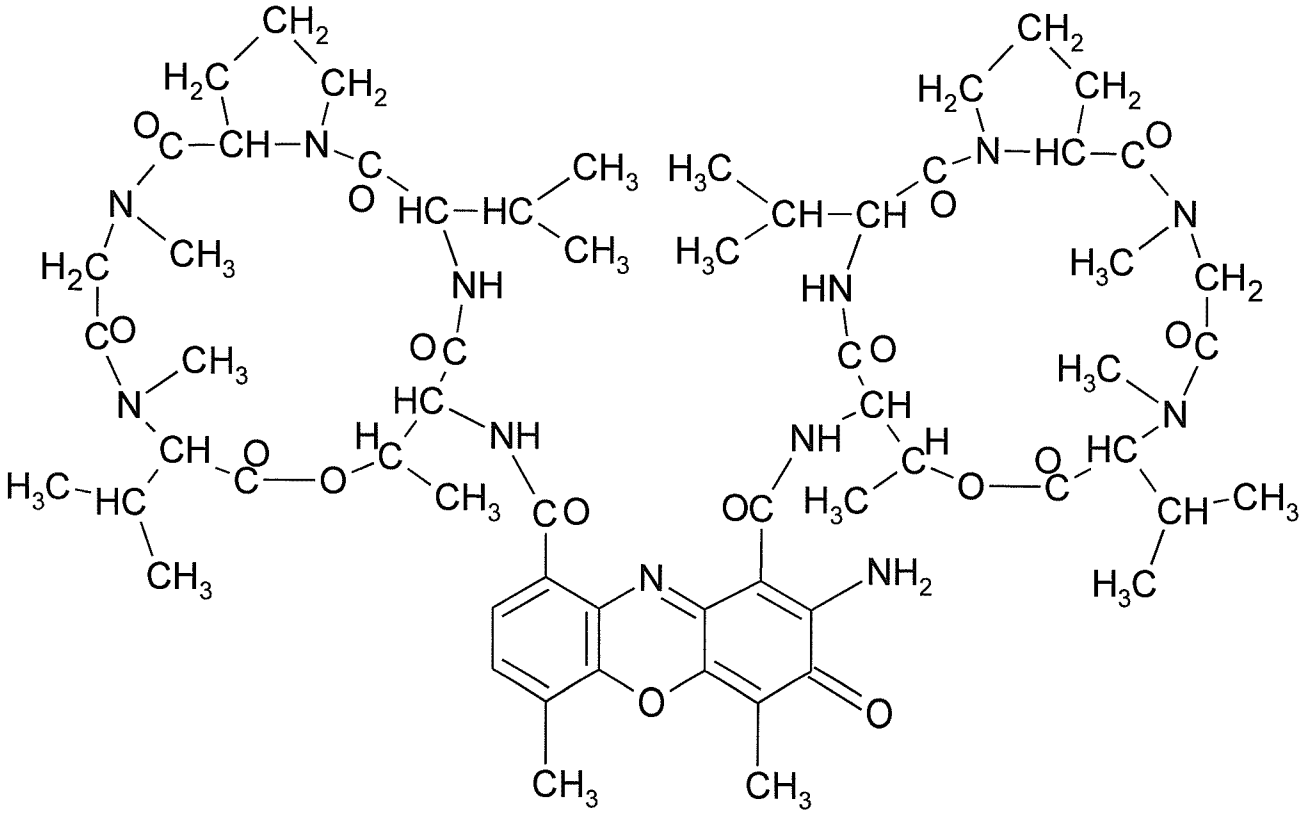


Figure 1.27 Molecular structure of actinomycin D

dissociation is faster from TGCA (Fletcher and Fox, 1996).

1.5.2 Mode of binding

Initially it was proposed that actinomycin D bound to DNA by a direct interaction of the guanine 2-amino proton with an acceptor group within the actinomycin chromophore. This was based upon chemical studies of deoxy nucleotides possessing different percentages or lacking the guanine residue (Goldberg *et al.*, 1962). Alternatively other experiments using spectroscopy suggested that actinomycin bound to DNA by a mechanism of intercalation (Muller and Crothers,1968). This conclusion was also reached by experiments showing that supercoiled DNA was unwound by actinomycin D (Waring, 1970). Moreover, these results were supported by X-ray crystallography (Sobell, 1974) showing that actinomycin bound to DNA by intercalating the phenoxazone ring between the dinucleotide sequence dG-dC and the peptide subunits were in the minor groove interacting, via hydrogen bonds, with deoxyguanosine residues on opposite chains. Other NMR evidence supports the intercalating mode of binding of actinomycin (Patel, 1974; Krugh, 1972; Arison and Hoogsteen, 1970). Hydrogen bonding does occur between the guanine 2-amino group of DNA and the threonine carbonyls of the drug (Chen *et al.*, 1996). The importance of the 2-amino group was assessed by footprinting experiments using 2,6-diaminopurine (DAP) and inosine (I) as substitutes for adenine and guanine, the results were that actinomycin did not bind to IpC, however, it bound to DAPpT. These results suggest that actinomycin requires the 2-amino group for binding (Bailly and Waring, 1995; Waring and Bailly, 1994). The cyclopeptide side-chains of actinomycin D interact with the minor groove, forming hydrophobic interactions between the methyl groups of each threonine and the sugar ring of guanine and between the five-membered ring of each proline and the sugar ring of cytosine (Chen *et al.*, 1996).

Actinomycin binds very tightly to DNA and dissociates very slowly (Muller and Crothers, 1968). Stopped flow experiments suggested that actinomycin bound to natural DNA with an association constant of 4×10^3 M^{-1} and then migrated to sites with higher affinity until it found the optimal binding site. (Fox and Waring, 1984; Fox and Waring,

1986). From this suggestion the shuffling hypothesis was raised to explain the patterns observed in a time dependent footprinting experiment. This hypothesis suggested that actinomycin bound first to non specific sites and then shuffled along the DNA helix to locate preferred binding sites. To test the validity of the hypothesis footprinting experiments with DNaseI (Bailly *et al.*, 1994) were conducted with 7-azidoactinomycin, a derivative of actinomycin that has the same binding properties (Graves and Wadkins, 1989), and the results suggested that actinomycin D might bind to DNA in a two step process. First the actinomycin molecule recognises the structure of the minor groove of DNA, and will bind only to those sites that contain a suitably wide minor groove, such as in GC rich sequences. then the molecule moves along the helix axis until a GpC site is located and hydrogen contacts are established (Bailly *et al.*, 1994).

The crystal structure of actinomycin bound to d(GAAGCTTC) (Kamitori and Takusagawa, 1992) showed that there are three different structures depending on the positions of the cyclic depsipeptides. Actinomycin seems to be able to change its conformation without breaking essential hydrogen bonds between the drug and the DNA. It forms four hydrogen bonds between its threonine moiety and guanine and two hydrogen bonds between the N2 amino group of phenoxazone and the DNA backbone. The four bonds between the threonine and guanine seem to be responsible for the specificity of this agent for GpC. In addition, their results suggest that actinomycin D can bind not only to B-form DNA but also to A-form DNA (Kamitori and Takusagawa, 1994).

NMR experiments conducted with short oligonucleotides containing at least two GC sites, such as d(ATGCGCAT) (Scott *et al.*, 1988a), dGCGC (Wilson *et al.*, 1986; Scott *et al.*, 1988b) and d(AAGCGCTT) (Chen *et al.*, 1996) indicated that actinomycin D could also bind in a 2:1 (drug/DNA) ratio (Fig. 1.28). The binding of the second actinomycin molecule to GCGC is highly anti-cooperative (Scott *et al.*, 1988b). However, when two actinomycin molecules were bound to $dA_1T_2G_3C_4G_4C_6A_7T_8$ unwinding of the duplex occurred, changing the orientation of the nucleotides with respect to each other. Also the binding caused a change in the conformation from C2'-endo to C3'-endo of the sugars of T_2, C_4 and C_6. These

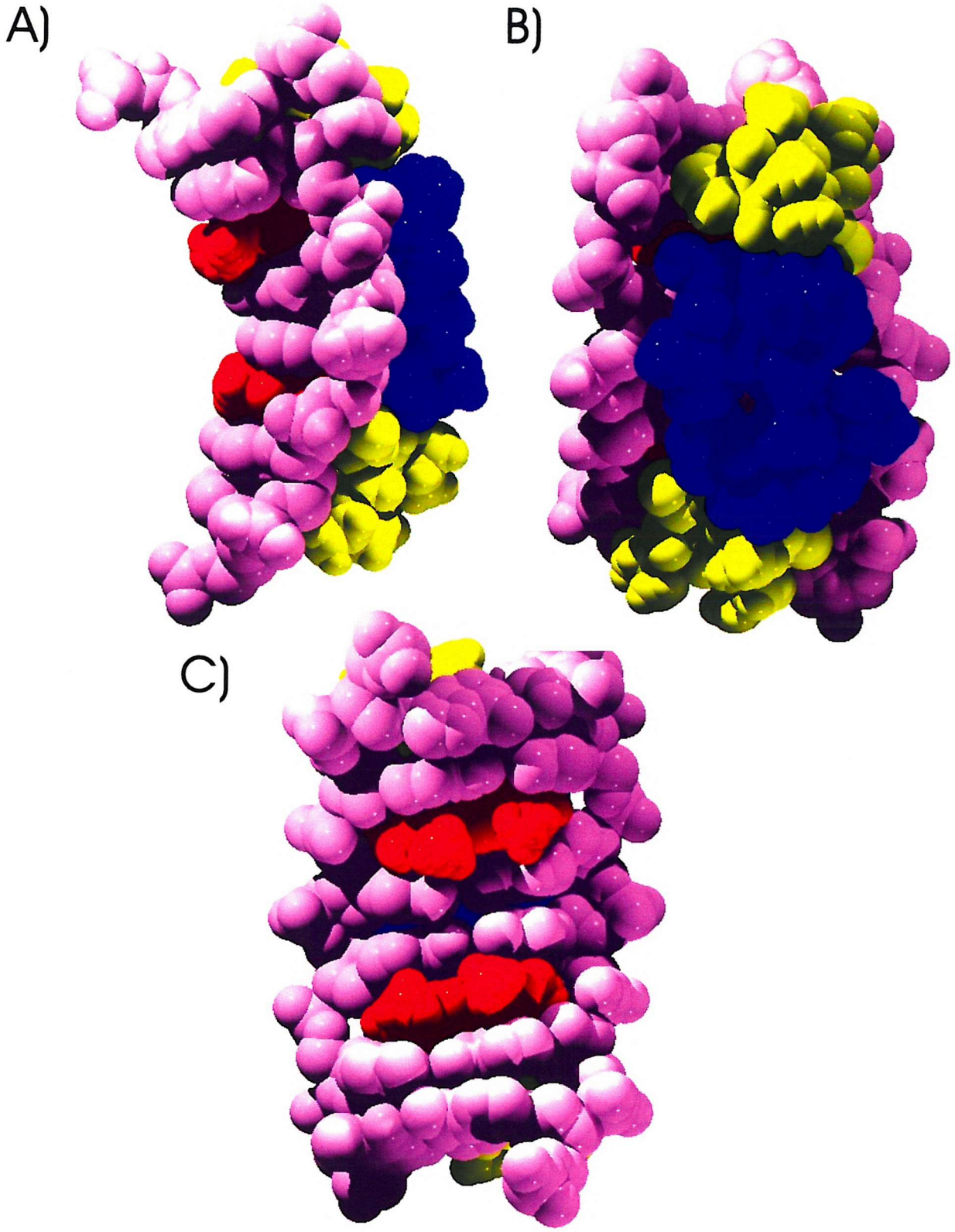


Figure 1.28 Structure of actinomycin-d(AAGCGCTT) complex in a 2:1 ratio. Color coding:phenoxazone ring, red; the two cyclic pentapeptide lactone rings are in yellow (Q ring) andin blue (B ring); DNA duplex, purple;A) this view shows the curvature of the DNA and the separation by two base pairs of the intercalated phenoxazone rings of adjacently bound actinomycin molecules looking along the phosphate backbone. B) Space filling view shows the minor groove to emphasize the complementary fit between the cyclic side-chains of actinomycin and the sugar-phosphate backbone of the DNA of the minor groove complex. C) A view from the major groove showing the intercalation of the phenoxazone ring. Images were created in Swiss-Pdb viewer, from the pdb file 1FJA submitted by Chen *et al.*, (1996),

data provide an answer to how do the peptide rings of actinomycin fit into the narrow minor groove of B-DNA and shows that the duplex is a very flexible molecule (Scott *et al.*, 1988a) even within a short distance (Wartell *et al.*, 1974). In addition, the 2:1 complex has been crystalized interacting with d(GAAGCTTC) and revealed that the complex is stabilized not only by intercalation, but also by forming hydrogen bonds between the DNA's guanine amino group N-2 and ring nitrogen N-3 and the actinomycin's carbonyl oxygen atom and amide groups of threonine residues (Kamitori and Takusagawa, 1992).

1.6 FOOTPRINTING

One of the most commonly used techniques for assessing sequence specific DNA binding is that of footprinting. This method was devised by Galas and Schmitz (1978) and was initially developed to study the interaction of proteins with DNA and to determine the specific sequences to which they bind. Subsequently, this technique has been used to great effect to investigate small ligands and antibiotics binding to DNA. In this technique, uniform DNA fragments of known sequence and length (typically 100-500 base-pairs) are tagged on one end of one strand (5' or 3') with the label ^{32} P. These DNA fragments are exposed to DNA-binding molecules (i.e. proteins or drugs) which will bind to their preferred binding sites among the 100 to 500 base pairs available. The fragments are subsequently digested with a suitable probe (typically DNase I). A bound protein or drug protects the base pairs that they cover from cleavage by the enzyme, and this is visualized on an autoradiogram of a sequencing gel as a gap or lighter region in the otherwise continuous ladder of bands (Fig 1.29). Control cleavage is run alongside to show all the steps of the continuous ladder so that precise identification of the protected regions is made possible. Several probes, both chemical and synthetic, have been used as footprinting agents and include DNase I (Laskowski, 1971; Suck, 1982); Micrococcal nuclease (Dingwall *et al.*, 1981); copper phenanthroline (Sigman *et al.*, 1979); methidiumpropyl-EDTA. Fe(II) (MPE) (Hertzberg and Dervan, 1982); and hydroxyl radicals (Haber and Weiss, 1934).

The ideal footprinting probe should be sequence neutral and generate an even ladder of DNA cleavage products in the absence of the ligand. This is almost true for certain

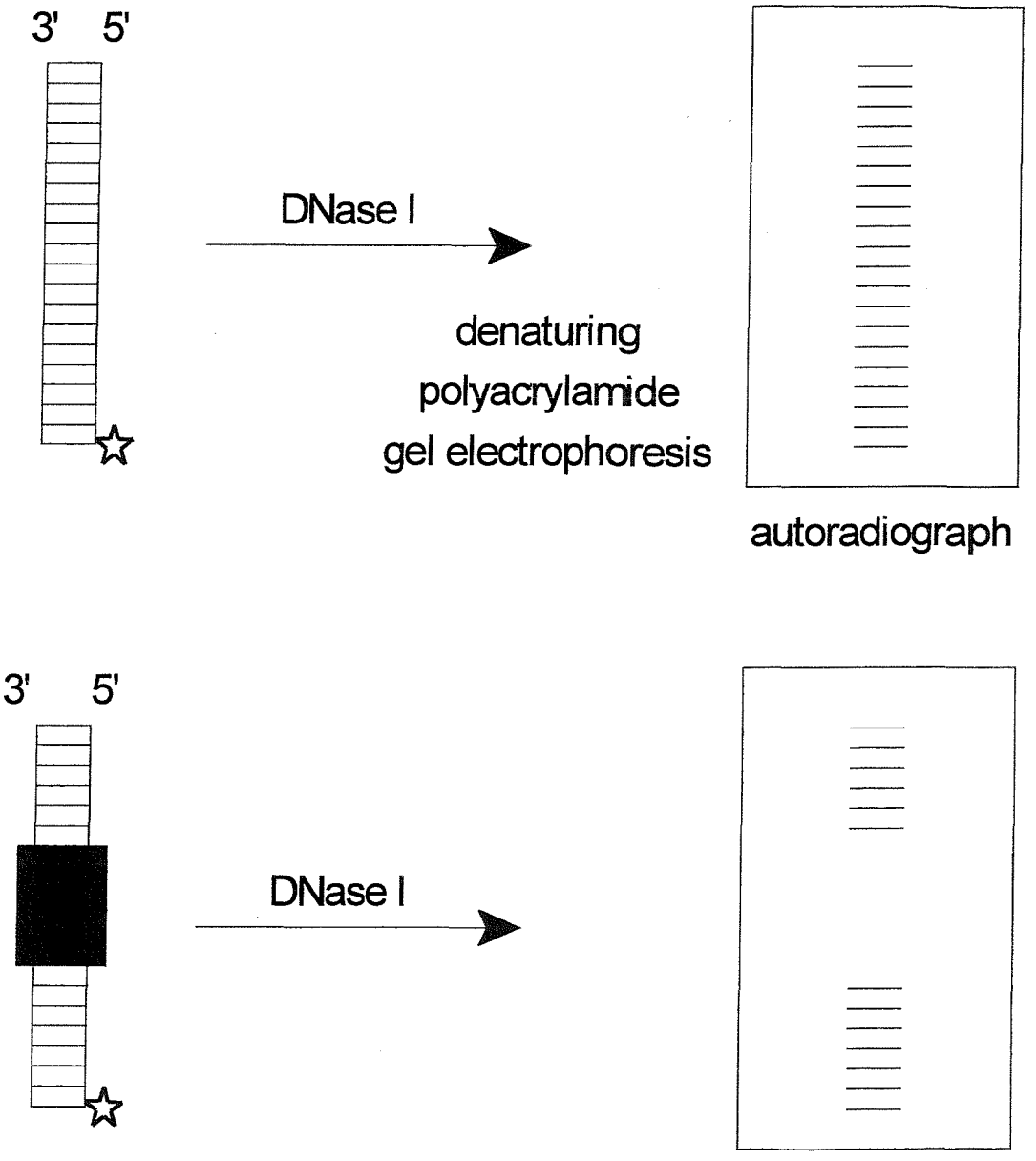


Figure 1.29 Schematic representation of the footprinting technique. The black rectangle represents the bound ligand, and the star represents the radioactive label.

chemical probes, such as methidiumpropyl-EDTA (MPE) (Van Dyke *et al.*, 1982) and hydroxyl radicals (Tullius and Dombroski, 1986). However, DNase I is the most commonly used probe because of its cost and ease of use, even though it produces a cleavage pattern that varies according to the DNA sequence and local structure (see section 1.7.1.1). Drugs that modify DNA conformation in sites adjacent to their binding sequence can also enhance enzyme activity. This is noticed in cases were the cleavage of these sites in drug-free controls is very poor. The enzymes and synthetic probes have their own characteristic sequence-structural recognition properties (see section 1.6) and they give different information about the nature of drug induced structural changes in the DNA (Fox, 1992).

1.7 FOOTPRINTING PROBES

1.7.1 Enzymatic probes

1.7.1.1 DNase I

DNase I is the most commonly used probe in footprinting studies. It is a double-strand specific endonuclease which needs a divalent metal ion (i.e. calcium, magnesium or manganese) as a cofactor to cleave the O-3'-phosphate bond. Thus, it introduces single strand nicks in the DNA (Laskowski, 1971; Suck, 1982). Although it cuts all phosphodiester bonds it shows a mild preference for cutting at the steps TpX, XpT, XpA and ApX, where X is any base (Laskowski, 1971). The crystal structure of the enzyme has been determined (Suck *et al.*, 1984)(Fig 1.30) and its mechanism for binding to and cutting DNA has been suggested (see Fig 1.31) (Suck and Oefner, 1986). The crystal structure shows that an exposed loop of the enzyme binds DNA from the minor groove catalysing nucleophilic attack of a water molecule at the phosphorus atom. Its specificity therefore arises from structural recognition of this groove which is a parameter that is influenced by the local helical twist, the base pair propeller twist and the precise orientation of the scissile phosphate bond (Lomonossoff *et al.*, 1981; Drew and Travers, 1984). The optimum width of the minor groove for efficient cleavage is 10-12 Å, if it is narrower than this value the efficiency of the enzyme is significantly reduced. This is confirmed by studies of homopolymeric runs of A and T which adopt a structure with a very narrow minor groove,

Figure 1.30 Crystal structure of Dnase I taken from Weston *et al.*, 1992)

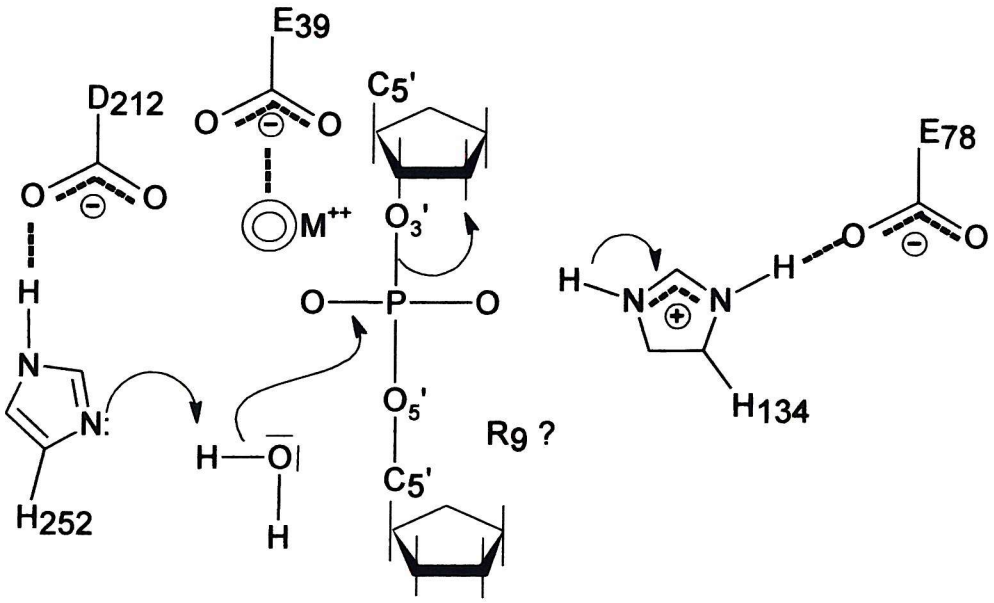


Figure 1.31 Proposed mechanism of action of DNase I (D212, E39, E78, H252, H134, and R9 are parts of the enzyme). From the disposition of the residues and assuming that the reaction involves an in-line attack at the phosphorous group by a water molecule, it seems most likely that H252 acts as a proton acceptor from a water molecule opposite the O3' atom. The subsequent attack at the phosphorous group with inversion of its configuration gives rise to a pentacovalent transition state that maybe stabilized by the metal ion and possibly by the side chain of R9. H134 is in an ideal position to protonate the leaving O3'. The metal ion (double circle denoted M^{++}) may also be important for the correct positioning of the phosphate group. Adapted from Weston et al., (1992) .

and are cut very poorly (Drew and Travers, 1984). Thus, the enhancements in DNase I activity caused by drugs are usually a consequence of changes in the minor groove. Low *et al.* (1984b) were the first to observe drug-induced enhancement in DNase I activity when footprinting echinomycin on the *Tyr* T DNA fragment. In addition they noticed that these enhancements were mainly confined to runs of A and T residues, consequently it was suggested that echinomycin might open up the grooves adjacent to its intercalation sites, converting the AT rich sequences into a form which more closely resembles mixed sequence DNA which is more susceptible to cleavage. Other researchers found that the enhancements of DNase I activity depended on the drugs used and not on the DNA sequences (Fox and Waring, 1984). Since these initial observations many ligands have been shown to induce enhancements of DNase I activity.

The DNase I crystal structure also revealed that the DNA bends towards the major groove (opening up the minor groove) in the presence of the enzyme (Lahm and Suck, 1991). Therefore, it has been suggested that DNA flexibility is an important parameter affecting its cleavage efficiency, explaining why GC regions (which are more rigid) are poor substrates for the enzyme. Also DNase I, when used as a footprinting enzyme, causes an overestimation of the bases which are protected by the agent bound to DNA. Presumably this overestimation of approximately three base pairs is caused by the size of the enzyme ($M_r \sim 31{,}000$; 40 Å diameter)(Drew and Travers, 1984). In addition DNase I footprints are usually staggered in the 3'-direction across the two strands of the duplex. This occurs because the enzyme binds across the width of the DNA minor groove, and closest phosphates are staggered by 2-3 base pairs across the strands (Fig 1.32).

1.7.1.2 DNase II

The structural requirements of this enzyme are less well defined than for DNase I and its DNA cleavage pattern is very uneven. It is therefore not used very often as a footprinting probe, instead it is used in the detection of DNA structural changes induced by ligand binding. DNase II does not need a metal ion cofactor and has a pH optimum of 4.8 (Bernardi, 1971). It cleaves the O-5' phosphate bond, consequently if the DNA is labelled at

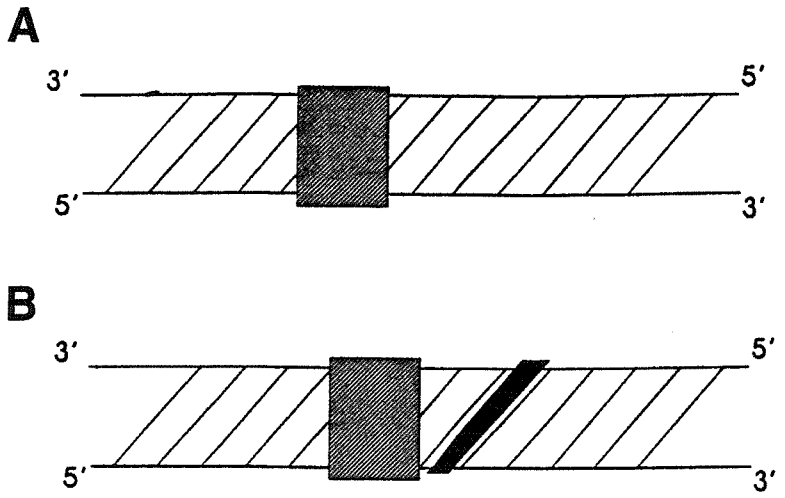


Figure 1.32 Diagram of the 3'-staggered cleavage produced by DNase I The DNA helix is viewed along the minor groove. The ligand is represented by the black box and Dnase I is represented by the hatched box This figure is taken from Fox 1997

the 3'-end the fragments produced will bear one less phosphate group than corresponding products from DNase I cleavage. This causes the DNase II cleavage products to migrate slower in polyacrylamide gels. DNase II prefers to cleave purine rich sequences, although it will also attack single-stranded substrates (Drew, 1984). Optimum cleaving sites for this enzyme are runs of A and G or G alone, the efficiency decreases for runs of A alone. Regions of alternating A and T bases are not digested (Suggs and Wagner, 1986). DNase II cleavage patterns do not correlate across the two DNA strands, and it is presumed that the enzyme recognizes a particular structure adopted by purine rich strands in double helical DNA (Drew, 1984). The asymmetrical cleavage could be due to either some asymmetry in the DNA structure or to the cleavage mechanism of the enzyme. DNase II has been found to cleave better when the distance between the same-strand phosphates is small (Drew and Travers, 1984).

DNase II was first used in drug-DNA studies to confirm whether the echinomycin induced changes in DNA structure seen with DNase I were due to the enzyme or to some unidentified artefact (Low *et al.*, 1984b). The enzyme has shown drug-induced enhancements and many of them coincide with regions that were enhanced with DNase I. Thus, these enhancements indicate that there is some drug-induced alteration in the DNA structure that is transmitted from the drug binding site through a series of A/T base pairs. DNase II has also been used to probe the interaction of nogalamycin and its derivatives with DNA (Fox, 1988).

1.7.1.3 Micrococcal nuclease

Micrococcal Nuclease (MNase), with Ca^{2+} as a cofactor, cleaves the O-5' phosphate bond (Deiters *et al.*, 1982). The cleavage occurs almost exclusively at bonds pA and pT (Dingwall *et al.*, 1981), a preference which has been suggested to arise from the greater conformational mobility of AT-rich sequences that present a more open structure (Flick *et al.*, 1986). However, the enzyme does not cleave all the pA and pT bonds with the same efficiency; homopolymeric runs of A and T are cleaved less well than regions of alternating A and T residues (Flick *et al.*, 1986;Fox and Waring, 1987). MNase has been found to be

sensitive to local DNA conformational flexibility and to the nature of the nucleotide at the cutting site. Consequently, MNase is useful in determining changes in the local stability (breathing) of the DNA duplex surrounding drug binding sites (Fox and Waring, 1987). However, its major use is in the preparation of nucleosome particles and in studies of chromatin structure (McGhee and Felsenfeld, 1980), since it cleaves linker DNA much more readily than nucleosome-bound core regions.

Several ligands have been found to cause enhancements in MNase digestion without altering its sequence specificity, as shown in areas of high AT content (Fox and Waring, 1987). Some of the drugs that enhance MNase activity include echinomycin (Fox and Waring, 1987), luzopeptin (Fox *et al.*, 1988), bleomycin (Fox *et al.*, 1987) and distamycin-like compounds (Portugal and Waring, 1987a; 1988).

1.7.2 Chemical Probes

1.7.2.1 EDTA/Fe(II) ('hydroxyl radical')

EDTA/Fe(II) uses the Fenton reaction which involves the decomposition of H_2O_2 with the concomitant oxidation of Fe^{2+} and the intermediate formation of hydroxyl radicals (Haber and Weiss, 1934).

$$\mathbf{Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + HO^- + HO^+}$$

The hydroxyl radicals oxidize the deoxyribose of the DNA backbone cleaving the DNA (Tullius, 1987; Tullius *et al.*, 1987). Hydroxyl radicals are especially useful because of their small size, the fact that they do not perturb the DNA structure and since EDTA/Fe(II) is negatively charged it does not bind to DNA (Tullius, 1987). Regions protected from cleavage are staggered by 2-3 bonds towards the 3' direction. suggesting that cleavage occurs from the minor groove as described for DNase I. Hydroxyl radicals are also useful for studying regions that posses unusual conformations, in particular regions which posses a narrow minor groove as found in runs of adenines in Kinetoplast bent DNA (Burkhoff and

(Burkhoff and Tullius, 1987). Because of its small size, and the even ladder of cleavage products it is useful for mapping drug DNA contacts at high resolution (Tullius, 1987). The hydroxyl radicals have been used successfully for determining the binding sites of distamycin (Churchill *et al.*, 1990) and mithramycin (Carpenter *et al.*, 1993). Unfortunately not all DNA binding ligands give footprints with hydroxyl radicals e.g. actinomycin (Churchill *et al.*, 1990), and the quinoxaline antibiotics echinomycin, TANDEM and Triostin A because these agents are dissolved in organic solvents which inhibit the free radical reaction.

1.7.2.2 Methidiumpropyl-EDTA/Fe(II)

Methidiumpropyl-EDTA/Fe(II) (MPE) has been developed by Dervan and coworkers (Hertzberg and Dervan, 1982) and was one of the first chemical cleavage agents to be used as a footprinting probe (Van Dyke *et al.*, 1982). This compound (Fig.1.33) consists of a methidium chromophore, which intercalates between the DNA bases, to which is attached a reactive tail containing Fe^{2+}, responsible for generating free radicals thereby cleaving DNA. MPE cleaves DNA with no base dependence and with only minor sequence dependence (Van Dyke and Dervan, 1984).

1.7.2.3 Other chemical footprinting probes

There are many chemical probes that have been used in footprinting experiments and these include: 1,10-Phenanthroline/CuI (OP-Cu) (Sigman *et al.*, 1979), uranyl(vi) ion (Nielsen *et al.*, 1988), Diethyl pyrocarbonate (DEPC) specific for A (Miles, 1977), Dimethyl sulfate (DMS) specific for G (Lawley and Brookes, 1963), Potassium permanganate specific for T (Rubin and Schmid, 1980), Osmium tetroxide (OsO_4) specific for T (Friedmann and Brown, 1978), Bromoacetaldehyde specific for A (McLean and Waring, 1988).

1.8 REPSA (Restriction Enzyme Protection Selection and Amplification)

Gene therapeutic applications require agents with high sequence specificity to enable

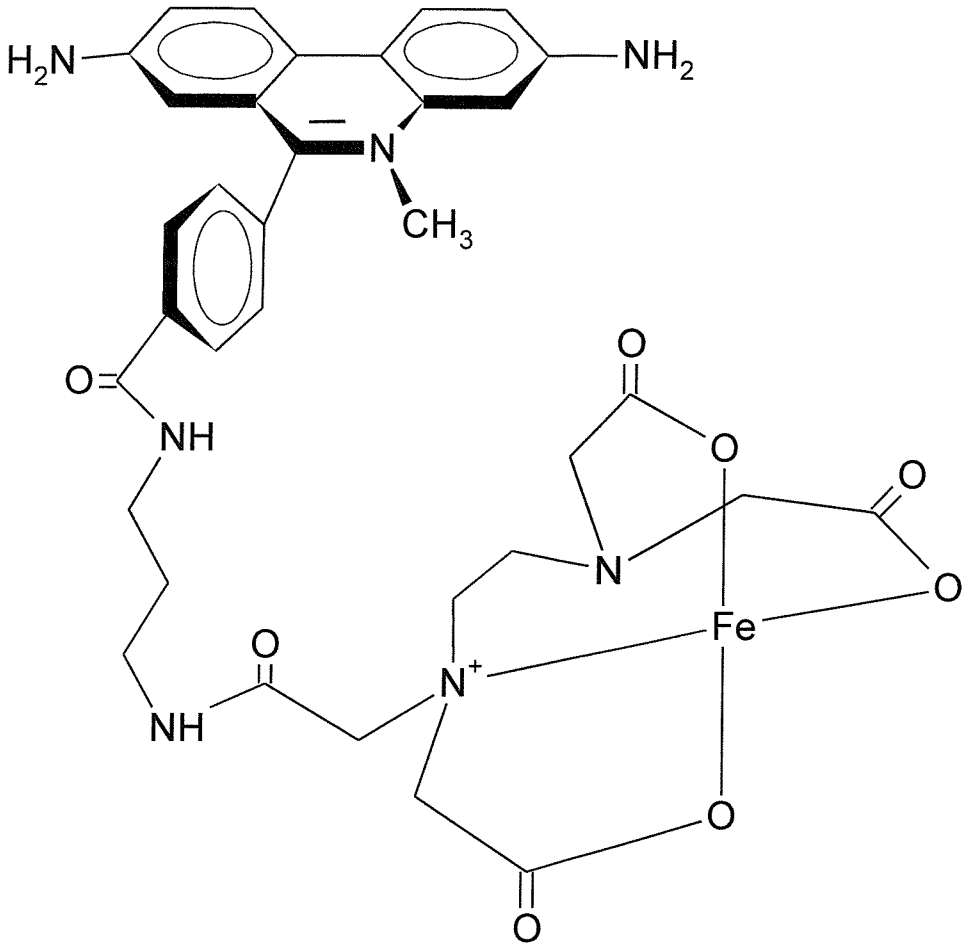


Figure 1.33 Molecular structure of Methidiumpropyl-EDTA/Fe(II)

recognition of unique sites within the human genome. To target a single site within the human genome of 3 $x10^9$ base pairs requires recognition of at least 17 base pairs (Thoung and Helene, 1993). Therefore, the interest in highly selective compounds has increased and the technique most frequently used for recognition of binding sites is that of footprinting. However, like all techniques it has its limitations. The most crucial of these is that the size of the DNA fragment is limited to 200 base pairs and cannot be increased without losing resolution. Therefore, the binding site of the ligand has to be present within these 200 base pairs. The probability of finding the recognition sequence of a ligand within the fragment depends on the selectivity of the ligand. If the ligand is selective for 1 base pair the number of combinations is 2 (i.e. AT or GC), for 2 base pairs the number is 10 (not 16 as a result of the self-complementarity of the bases and these are AA=TT, GG=CC, GT=AC, TG=CA, AG=CT, GA=TC, TA, GC, CG and AT). For 3 base pairs the number of combinations is 32, for 4 is 136 , for 5 is 512. The general equations are then:

For an odd number of bases $4^n/2$

For an even number of bases $(4^n + 4^{n/2})/2$.

and the probability is inversely proportional to the number of combinations. Thus the more selective the ligand the less likely it is to find the preferred binding site within a 200 base pair fragment. Footprinting reactions for highly selective ligands therefore require some prior knowledge of the exact ligand binding site. Thus, in order to survey the several million base pairs required to identify their hypothetical binding sites at physiological drug concentrations a new technique is needed.

Presently there are combinatorial methods which have been successful in identifying the preferred binding sites of proteins (Wright and Funk, 1993). These methods depend on the physical separation of ligand-DNA complexes from free DNA either by electrophoresis or by ligand specific antibodies. However, these techniques cannot be used to isolate small molecule-DNA complexes because these complexes are often not sufficiently stable to allow their partition, or there are no available means for their physical isolation. Therefore,

new combinatorial methods are required to study drug-DNA interactions. One such technique is REPSA (Hardenbol and Van Dyke, 1996).

1.8.1 Theory

REPSA (Restriction Enzyme Protection Selection and Amplification) consists of several steps which are represented in figure 1.34.

1) Protection step The ligand is exposed to a pool of oligonucleotides which contain a random mixture of bases in the central region, flanked by fixed sequences containing a recognition site for the selecting type IIs restriction enzyme. From these random sequences the ligand will bind to those for which it has the highest affinity, thus protecting them from cleavage by the enzyme.

2) Selection step The random pool of oligonucleotides is subjected to cleavage by the restriction enzyme. Sequences to which the drug is not bound will be cleaved, while drug binding sites will be protected.

3) Amplification step The protected sequences are subsequently amplified using the PCR, and in this way increase the ratio of selected versus unselected sequences.

This process is repeated for several rounds until we are sure that sufficient specificity has been obtained.

4) Cloning and Sequencing The product is then cloned into pUC 18 and sequenced in order to determine the sequences with the highest affinity for the ligand.

1.8.1.1 Choice of restriction enzyme

The restriction enzymes should be of the type IIs. This class of enzyme cuts at a

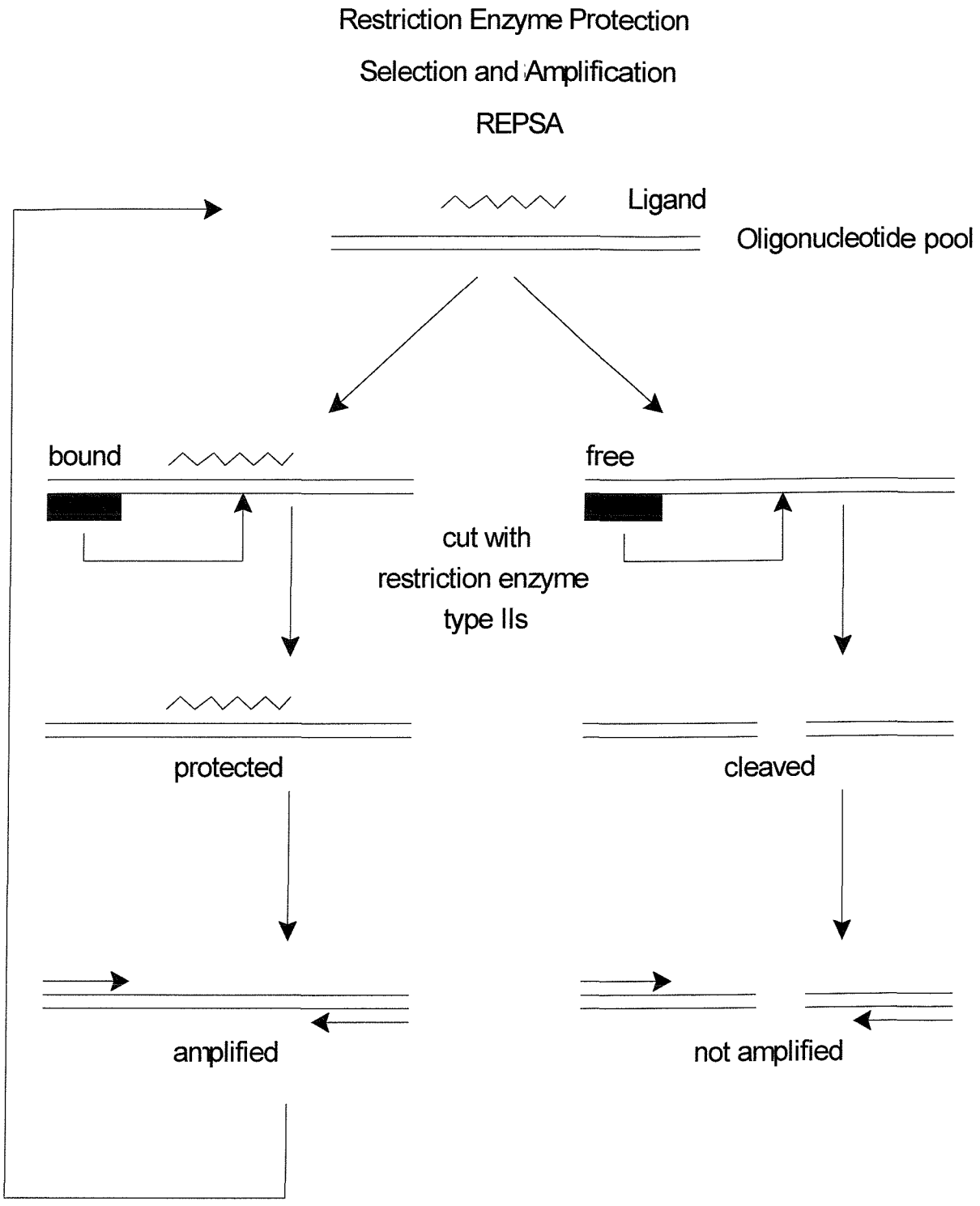


Figure 1.34 Schematic diagram showing the different stages of REPSA. The black box represents the enzyme binding site and the arrow is the enzymes cleavage site. The random region is located in the centre of the template.

specific distance away from its recognition site without regard to the sequence at the cleavage site. Also there are some type II enzymes which contain some random bases within their recognition sequence that could be used for REPSA. Examples of type IIs enzymes are *Fok* I which recognizes the sequence GGATG, cutting the DNA 9 and 13 bases away, and *Bsg* I which recognizes GTGCAG and cuts the DNA strands 16 and 14 bases away. An example of the type II enzymes is *Bgl* I which has a recognition sequence GCCNNNNNGGC and cuts within the N region. Another requirement is that the restriction enzyme must be inhibited by the ligand through DNA binding.

1.8.2 Applications of REPSA

To date REPSA has been successfully used in three separate studies. The first study was concerned with the determination of DNA sequences that were selected by a triplex-forming oligonucleotide, 21 base pairs long, of a purine-motif. The results of this study showed that REPSA was a powerful method for investigating ligand-DNA interactions. In addition, the data indicated the possible existence of a new base triplet, G.AT (Hardenbol and Van Dyke, 1996). The next study identified the best DNA binding sites for the TATA binding subunit (hTBP) of the human general transcription factor. The most significant sequence was 5'-TATAAATA-3 followed by 5'-TATAAAAG-3', this last sequence corresponds to the adenovirus-2 major late TATA box sequence (Hardenbol *et al.*, 1997a). The last study determined the preferred DNA binding sites of distamycin A (Hardenbol *et al.*, 1997b); however in this experiment there are a few points of concern which will be highlighted here. First the template was designed with a central region containing 21 random base pairs, this length is longer than that required for an individual distamycin binding site, which is about 5 base pairs long. Thus, the results obtained cannot be used quantitatively to determine the best binding site for distamycin. Also the number of sequences theoretically possible is larger than the number of sequences practically obtainable, and this is due to the number of molecules provided at the beginning of the experiment being too low (25 fmol), therefore, not all the sequences were represented within the sample. These are some of the reasons why in our experiments we will be using different conditions from those reported in this paper. (See chapter 3). However, their results show that distamycin binds to runs of A

and T and proves that REPSA is a valid technique for determining binding sites.

1.9 PURPOSE OF THIS WORK

The purpose of this work is to find new methods for assessing the preferred sites of sequence selective DNA binding molecules. Two techniques have been used to accomplish this goal. The first is REPSA which has been used successfully by Hardenbol and Van Dyke (1996). The other is a footprinting substrate that we have designed that contains all the possible tetranucleotides.

CHAPTER 2

Materials and methods

2.1 Chemicals and enzymes

Oligodeoxynucleotides were purchased from Oswel (Southampton, UK) and stored at -20°C. Bovine DNase I was bought from Sigma and stored at -20°C at a concentration of 7200 U/ml in 150 mM NaCl and 1mM $MgCl_2$. *Bsg* I was obtained from New England Biolabs. AMV reverse transcriptase, T4 DNA ligase and all other restriction endonucleases were purchased from Promega. Taq DNA Polymerase was purchased from Bioline. *Bam*H I-cut pUC 18 treated with alkaline phosphatase and the T7- dideoxy sequencing Kit , were obtained from Pharmacia Biotech UK.

Radioactive [α^{32}P]ATP was purchased from Amersham International at an initial activity of 3000 Ci/mmol.

Distamycin, echinomycin, actinomycin D, mithramycin and Hoechst 33258 were obtained from Sigma. [*N*-MeCys^3,*N*-MeCys^7]TANDEM was a gift from Professor Olsen.

Polyacrylamide concentrates (Sequagel and Accugel) were purchased from National diagnostics. X-ray film and Saran wrap were obtained from GRI; all other chemicals and materials were purchased from Sigma or Fisher Scientific.

2.2 REPSA non-radioactive method

The precise details of the REPSA protocol were varied during this work, so as to optimise the process. The full details of these changes are presented in chapter 3. The conditions described below correspond to those most commonly used.

2.2.1 Template formation

The double stranded template for REPSA, containing a central region of redundant

bases (see sequences below), was prepared by PCR using the single strand oligonucleotide and the right primer. This could not be prepared by simply annealing complementary strands, as the central region would contain mismatched base pairs. By using a small number of PCR cycles (4), each double strand is freshly made, rather than resulting from annealing of product molecules.

```
5'-GGTAGGATCCAGAGTGCAG-3'
5'-GGTAGGATCCAGAGTGCAGNNNNNNNNNNNNNNNNNGACCTGAGAGGATCCATGG-3'
                                          3'-CTGGACTCTCCTAGGTACC-5'
```

2 µl ssDNA template (30 ng/µl) was mixed with 200 nM primers, 4 µl 10 mM dNTPs, 6 µl 25 mM $MgCl_2$, 1 µl 10 x Taq buffer(160 mM $(NH_4)_2SO_4$, 670 mM Tris-HCl pH 8.8 @ 25 °C, 0.1% Tween-20) and made up to 100 µl with water and 1µl 5 u/µl Taq DNA polymerase. PCR was performed for four cycles with the following conditions 95 °C for 1 min denaturing, 60 °C for 3 min annealing and extension. This should yield a final concentration of dsDNA template of approximately 10 ng/µl.

2.2.2 First round of REPSA

2.2.2.1 Ligand binding (Protection)

5 µl of dsDNA template (approximately 50 ng), 2 µl of drug which was previously dissolved in Tris-Na (10 mM Tris-HCl pH 7.5 and 10 mM NaCl) at the desired concentration, was added to a mixture which was suitable for *Bsg* I cleavage containing 1 µl of 10x NEBuffer 4 , (500 mM potassium acetate, 200 mM Tris acetate, 100 mM magnesium acetate, 10 mM DTT pH 7.9 @ 25 °C) 0.8 µl 1 mM S-adenosylmethionine and 3 µl H_2O. This mixture was incubated at room temperature for 30 min.

2.2.2.2 Enzyme cleavage (Selection)

1 µl of *Bsg* I (3000 u/ml) was added to the previous sample and incubated at 37 °C for 30 min. The reaction was stopped by the addition of 2 µl of 0.5 M EDTA and diluted with 90 µl of water to yield a final dsDNA concentration of approximately 0.5 ng/µl. For more details of this step refer to Chapter 3.

2.2.2.3 Amplification

2 µl of the diluted sample (1 ng dsDNA) was added to a PCR tube containing the

same reaction mixture as used for preparing the dsDNA template (2.2.1). Hot start PCR was started with a cycle profile which changed throughout the project (for more details refer to Chapter 3). 30 µl samples were removed after 9, 12 and 15 cycles, and the PCR products were analysed using a 2% agarose gel electrophoresis (see section 2.3). The sample that yielded a visible product with the least number of cycles was selected for the next round of REPSA.

2.2.3 Second and Subsequent rounds of REPSA

5 µl of the amplified material from the previous round was used for the next round of REPSA. This was subjected to ligand binding, cleavage and amplification as described before (see Figure 1.8.1). This procedure was repeated several times to achieve sufficient selection, and then the sample was cloned and sequenced (section 2.16).

2.3 Agarose gel Electrophoresis

In order to determine the presence of PCR product after each round of REPSA, samples were run on a 2% agarose gel (1.2 g agarose in 60 ml of 1xTBE (11 g Tris, 5.5 g boric acid and 1 g EDTA per litter) and 6 µl of 10 mg/ml ethidium bromide). 8 µl of each sample (or the whole sample when DNA extraction was performed) was mixed with 4 µl of 20 % ficoll (20% ficoll (w/v) 10 mM EDTA and 0.1% of either bromophenol blue or xylene cyanol). Xylene cyanol was most often chosen, as bromophenol blue migrated at a similar speed to the PCR product. Alternatively bromophenol blue was used in a well by itself to mark the location of the product. The gel was run at 100 V (60-100 mA) until the bromophenol blue reached the middle of the gel. Bands in the gel were visualized under ultraviolet light and a picture was obtained showing the PCR product with the slower mobility and the primers which run faster (Fig.2.3.1). Note that the primers are still in a vast excess, as required for ensuring that no products anneal to each other, thereby avoiding the formation of mismatches in the central region.

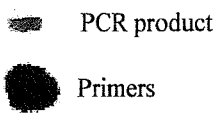


Figure 2.1 shows the primer and product as seen after exposing the agarose gel to ultraviolet radiation.

Agarose gel electrophoresis was also used to analyse plasmid DNA. In this case a 0.7% agarose gel was used (0.42 g agarose in 60 ml of 1xTBE (11 g Tris, 5.5 g boric acid and 1 g EDTA per litter) and 6 µl of 10 mg/ml ethidium bromide). The gels were run at the same voltage and current as above, and for the same duration. There are three different DNA species which can be observed under ultraviolet illumination, depending on their migration within the agarose: nicked DNA with the slowest mobility; linear DNA of intermediate mobility and supercoiled DNA with the greatest mobility.

2.4 Extraction of DNA from agarose gels

2.4.1 QIAEX II

For samples prepared by REPSA, the DNA was extracted from the agarose gels on alternate rounds, or when primer dimers were noticed in the agarose gel. The DNA extraction kit QIAEX II (from QIAGEN, West Sussex UK) was used according to the manufacturer's instructions. The band containing the DNA was excised from the gel (the slowest running band). Then 1.5 ml of Buffer QX1 and 30 µl of QIAEX II (silica gel) were added, the agarose was melted by incubation at 50°C for 10 min and the DNA was adsorbed to the silica pellet by gently mixing every 2 min, while it was incubating. The Eppendorf was then centrifuged at 13,000 rpm for 30 sec and the supernatant was discarded. To remove residual agarose contaminants the pellet was redissolved in 0.5 ml of QX1 and centrifuged at 13,000 rpm for 30 sec. After removing all traces of the supernatant the pellet was washed twice with buffer PE to eliminate residual salt contaminants. Then the pellet was allowed to dry for 10 to 15 min or until the pellet became white. The DNA was extracted by resuspending the pellets in 20 µl of water (pH between 7.0 and 8.5) and incubated for 5 min at room temperature, centrifuged for 30 sec and the supernatant placed into a clean tube. The resuspension, incubation and centrifugation steps were repeated once again to elute all the DNA. The DNA was concentrated by rotary evaporation under vacuum and the pellet resuspended in 15 µl of water and used for the next round of REPSA.

2.4.2 Crush and Soak Method

The procedure we followed was a modified method of Sambrook *et al.*, (1989). The DNA band was extracted from the agarose gel and chopped into small cubes which were placed in an Eppendorf tube (no more than 0.3 g of agarose per tube). Then 150 to 200 µl of

water was added to the Eppendorf together with 300 to 500 µl of buffered phenol (Tris-EDTA) pH 7.5. The sample was vortexed for one minute and dipped into liquid nitrogen for another 30 sec to 1 min, it was then centrifuged at 14,000 rpm for 6 to 8 min. The vortexing, freezing and centrifuging steps were repeated several times until no agarose was visible in the lower layer. The upper aqueous layer was extracted and placed into a clean tube, to this layer an equal volume of chloroform / isoamyl alcohol (24 : 1) was added, mixed, and centrifuged for 2 min. The upper phase was removed and placed into another tube to which 1/10 of the total volume of sodium acetate pH 5.2 was added together with 3 times the total volume of 96% ethanol. The sample was placed in dry ice for 15 to 30 min, and subsequently centrifuged at 14,000 rpm and 4°C. The supernatant was discarded carefully and the pellet was washed with 70% ethanol. After washing, the pellet was dried in a speedvac and resuspended in 10 µl of water.

2.5 REPSA Radioactive method

2.5.1 Labelling of Right Primer

The right primer was labelled with [γ] ^{32} P-ATP, and the procedure for performing this kind of labelling has been fully described in section 2.18.2.

2.5.2 Labelling the template

The following mixture was added to a tube: 2 µl of template, 5 µl of labelled right primer (prepared as described above), 1 µl of 100 mM dNTP (containing 25 mM of each dNTP), 5 µl of Taq buffer, 1 µl of 50 mM $MgCl_2$, 40 µl water and 1 µl Taq polymerase. This sample was extended on the PCR with the following conditions: 2 min at 94°C and 10 min at 64°C for only one cycle. Then the sample was loaded onto a 9% non-denaturing gel (11 ml Accugel in 50ml mixture) after adding 20 µl of 20% Ficoll. The band containing the labelled template was extracted from the gel and was eluted overnight with 200 to 300 µl of Tris-EDTA (10mM Tris-HCl, 0.1 mM EDTA, pH 8.0)

2.5.3 First Round of REPSA

2.5.3.1 Ligand binding (Protection)

The labelled template was ethanol precipitated (see section 2.7) and resuspended in

18 µl of Tris-EDTA (10mM Tris-HCl, 0.1 mM EDTA, pH 8.0). To this sample 2 µl of ligand (dissolved in Tris-Na to a concentration which was 10 x the final concentration) was added and left to equilibrate for 30 min.

2.5.3.2 Enzyme cleavage (Selection)

After equilibration 2.5 µl of 10x NEBuffer 4 , (500 mM potassium acetate, 200 mM Tris acetate, 100 mM magnesium acetate, 10 mM DTT pH 7.9 @ 25°C) 2.5 µl SAM (0.8 mM S-adenosylmethionine) and 1 µl *Bsg* I were added and incubated at 37°C for 30 min. The reaction was stopped by boiling the sample for 3 min.

2.5.3.3 Amplification

In a PCR tube 10 µl of the cleaved fragment was mixed with 5 µl of the labelled right primer (see section 2.5.1). 2 µl of left primer, 1 µl of 25 mM dNTP, 5 µl of 10 x Taq buffer(160 mM $(NH_4)_2SO_4$, 670 mM Tris-HCl pH 8.8 @ 25°C, 0.1% Tween-20), 1 µl of 50 mM $MgCl_2$, 25 µl of H_2O and 1 µl of 5u/µl Taq.DNA polymerase. The tube was placed in the PCR and cycled for 15 cycles with the following settings: 30 sec at 92°C and 3 min at 64°C. Two more cycles were then performed with an excess of unlabelled right primer so as to prevent self-annealing of products by adding 2.5 µl unlabeled primer, 1 µl 25 mM dNTP and 1 µl Taq DNA polymerase. This sample was mixed with 20 µl of 20% Ficoll and loaded onto an 8% non- denaturing polyacrylamide gel. The band was extracted from the gel and was eluted overnight with 200 to 300 µl of Tris-EDTA (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0).

2.5.4 Second and Subsequent rounds of REPSA

The DNA from above was ethanol precipitated (see section 2.7) and resuspended in 18 µl of Tris-EDTA (10mM Tris-HCl, 0.1 mM EDTA, pH 8.0), and then the ligand was added again as in section 2.5.3.1.This procedure was repeated several times to ensure that sufficient selection had occurred, and then the sample was cloned and sequenced (see section 2.16).

2.6 Phenol extraction

Phenol extractions were performed to clean the DNA from the last round of REPSA

before cloning and sequencing, so as to remove any Taq DNA polymerase which is present in the PCR sample. In addition, phenol extractions were also used to remove drugs or enzymes from the DNA at some stages of REPSA (for more details see Ch 3).

The phenol extraction was started by mixing 200 µl of phenol (Biophenol/$CHCl_3$/ isoamylalcohol 25/2/1 pH 8) with 200 µl of water containing the DNA. The sample was spun for 5 min and the top layer removed and added to another 200 µl of phenol, (Biophenol/ $CHCl_3$/ isoamylalcohol 25/2/1 pH 8). After centrifugation for 5 min the upper phase was removed and added to an Eppendorf containing 200 µl of ether (which removed any remaining phenol) and centrifuged for 5 min. The bottom layer was removed and added to another 200 µl of ether, and centrifuged for 5 min. Finally the bottom layer was removed and left at 37°C for about 30 min. to evaporate the remaining ether. This sample was then ethanol precipitated to recover the DNA.

2.7 Ethanol precipitation

A volume of a DNA solution was combined with an amount of 3 M sodium acetate equivalent to 1/9 to 1/10 of the original DNA volume. To this was added a quantity of 95% ethanol which was equal to 3 times the total volume. The sample was then placed in dry ice for 10 min. and spun for 10 min. at 14,000 rpm. The supernatant was discarded and the pellet was washed with 100 µl 70% ethanol. The DNA was dried in a speedvac and redissolved to the desired volume in either water or TE (10 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA).

2.8 Staggered cleavage of *Bam*H I generating sticky ends

After ethanol precipitation the sample was redissolved in 45 µl of water. This was cleaved with *Bam*H I, so as to generate "sticky ends " for subsequent cloning, by adding 5 µl of buffer E (6 mM Tris-HCl, 6 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, pH 7.5 at 37°C from Promega) and 3 µl of *Bam*H I (10 u/µl) and the sample was incubated at 37°C for at least 30 min.

2.9 Ligation of the insert to the plasmid

The sample from 2.8 was ethanol precipitated, to remove the *Bam*H I, and redissolved in 30 µl of water. 15 µl of this was stored at -20°C and the remaining 15 µl was mixed with 1 µl of pUC 18 which had been *Bam*H I cut and treated with alkaline

phosphatase, 2 µl of 10x ligase buffer (300 mM Tris-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP) and 2 µl of T4 DNA ligase. This reaction mixture was incubated for at least two hours at room temperature.

2.10 Competent cells

2.10.1 $CaCl_2$ method

2.10.1.1. Cell preparation

A 5 ml 2YT (16 g Tryptone, 10 g Yeast Extract, 5 g NaCl per liter) culture was inoculated either with 50 µl of TG2 cells from a frozen glycerol stock or with a TG2 colony from an agar plate and incubated overnight at 37°C with shaking at 350 rpm. 500 µl of this culture was added to 50 ml of 2YT media (16 g Tryptone, 10 g Yeast Extract, 5 g NaCl per liter) and grown to an A_{600} of 0.4 to 0.7. The cells were harvested by centrifugation at 5000 rpm for 5 min. and the pellet was resuspended in 50 ml sterile, ice cold, transformation buffer (50 mM $CaCl_2$, containing 10 mM Tris-HCl (pH 7.4)) and incubated for 30 min on ice. The cells were centrifuged at 5000 rpm for 5 min., resuspended in 5 ml transformation buffer and stored at 4°C. The cells remained viable for about 1 week as judged by controls of known concentration.

2.10.1.2 Transformation

200 µl of competent cells was added to each ligation mixture from section 2.9 and left on ice for 30 minutes. The cells were heat shocked at 45°C for up to 2 min and the tube was returned to the ice and left for about 10 min before plating out on agar plates containing 100 µg/ml ampicillin, 0.5 mM IPTG, 0.01% Xgal and incubated overnight at 37°C.

2.10.2 Electroporation method

2.10.2.1 Cell preparation

3 ml of an overnight LB culture was added to 1 l of LB media (10 g Tryptone, 5 g Yeast Extract, 10 g NaCl per litre) and incubated at 37°C until the A_{600} was between 0.5 to 0.7, at which time the flask was chilled for 10 min on ice, harvested by centrifugation at 3,000 rpm for 20 min. The cells were washed three times by resuspending in 1x, 0.5x, 0.5x volumes of ice cold water, and harvested between each step as above. The pellet was then resuspended in 20 ml of 10% glycerol (ice cold) and centrifuged at 3,000 rpm for 20 min. The cells were finally resuspended in 2 ml of 10% glycerol (ice cold) and aliquots of 50 µl

were placed into sterile Eppendorf tubes (pre-chilled in dry-ice) and store at -80°C until required for transformation. The transformation efficiency was tested by using controls of known concentration. Typical values were $5x10^6$ to $1x10^7$ cfu/µg DNA.

2.10.2.2 Transformation

The electrocompetent cells were thawed on ice for 2 to 5 min. Then 1-2 µl of ligation mixture (DNA) was added and gently mixed with the pipette tip. The cell/DNA mixture was pipetted into a 0.1 cm electro-cuvette and electroporated using a Biorad Gene Pulser set to a resistance of 200 Ω, capacitance 25 µFD and Voltage 1.25 Kv. Immediately after electroporation, 1 ml of LB (10 g tryptone, 5 g yeast extract and 5 g NaCl per liter) containing 20 mM glucose, was added and shaken at 200 rpm, 37°C for 45 min. The cells were spread-plated onto agar plates containing 100 µg/ml ampicillin , 0.5 mM IPTG, 0.01% Xgal and incubated overnight at 37°C.

2.11 Preparation of Agar plates

Agar (39.5 g/l Blood based agar) was autoclaved and cooled to 50°C, then 1 ml of 100 mM IPTG, 1 ml of 2% X-Gal (in DMF) and 20 mg ampicillin were added immediately before pouring into petri-dishes. The plates were allowed to set for a few hours before storage at 4°C.

2.12 Screening of colonies

Plasmid pUC18 contains the lacZ gene which can complement the partial gene product from the host *E. coli* TG2 to produce active ß-galactosidase. This enzyme, in the presence of the inducer IPTG, degrades the chromogenic substrate Xgal releasing a blue colour. Wild type colonies, containing the intact pUC18 are therefore blue. The pUC18 polylinker multiple cloning site is positioned within this gene so that inserted fragments inactivate the gene giving rise to white colonies. Some of the white colonies were aseptically picked off each of the plates and grown overnight at 37°C in 5ml 2 YT cultures. 30% glycerol stocks were made from these overnight cultures (600 µl of broth to 400 µl of 50% glycerol). The remainder of the culture was used for preparing plasmid DNA as described below (sec 2.13, 2.14, and 2.15) which was then sequenced by the dideoxy method (sec 2.16).

2.13 Wizard Miniprep

Plasmid DNA was recovered from 5 ml 2 YT (16 g Tryptone, 10 g Yeast Extract, 5 g NaCl per liter) cultures using the"Wizard Plus Miniprep DNA purification System" (Promega). The cells were harvested by spinning at 6,000 rpm for 5 min. Then they were resuspended in 200 µl of Cell resuspension solution (50 mM Tris-HCl, 10 mM EDTA, 0.1 mg/ml RNase A) lysed with 200 µl alkaline cell lysis solution (0.2 M NaOH, 1% SDS) and cell debris precipitated with 200 µl Neutralization solution (3 M potassium acetate, 2 M acetic acid at pH 4.8). The cell debris was pelleted by centrifugation at 14,000 rpm for 5-10 minutes and the supernatant was collected. The plasmid was extracted from the supernatant by adding 1 ml of Magicprep DNA purification resin solution. This solution was then filtered through a maxiprep column which was connected to a vacuum manifold. The resin was then washed with 2 ml of column wash solution (200 mM NaCl, 20 mM Tris-HCl pH 7.5, 5 mM EDTA mixed with an equal volume of ethanol) and the resin was allowed to dry under vacuum for a further 10 min. The plasmid was extracted from the resin by adding 55 µl of water (preheated to 37°C), leaving it at room temperature for a few minutes and then centrifuging for 40 sec at 8,000 rpm.

2.14 QIAprep Spin miniprep

The QIAprep miniprep was the preferred method for plasmid preparation as it does not involve the addition of a DNA binding resin, using instead a silica-gel membrane which binds the DNA under high-salt conditions. The plasmid is subsequently released from the filter by adding a low-salt buffer or water. The first step of this miniprep involves the recovery of the plasmid from overnight 5 ml 2YT (16 g Tryptone, 10 g Yeast Extract, 5 g NaCl per liter) cultures. This was performed by dividing each culture into three 1.5 ml Eppendorfs which were centrifuged at 5,000 to 7,000 rpm for 5 min. Each pellet was resuspended in 83 µl of buffer P1 (Qiagen). The three resuspended pellets were pooled into one Eppendorf and then the lysis was performed by adding 250 µl of buffer P2 (Qiagen). Neutralisation was achieved by adding 350 µl of buffer N3. Then the sample was gently mixed and centrifuged at 14,000 rpm for 10 min. The supernatant was decanted or pipetted into a QIAprep spin column and was centrifuged at 14,000 rpm for 1 min. The endonucleases were removed by washing the filter with 0.5 ml of Buffer PB (Quiagen) and centrifuging at 14,000 rpm for 1 min. Then the salts were cleared by adding 0.75 ml of

buffer PE (Quiagen) and centrifuged at 14,000 rpm for 1 min. After each of these centrifugation steps the flow-through was discarded. The column was placed in a clean Eppendorf tube and the plasmid was eluted by adding 50 µl of buffer EB (10 mM Tris-HCl, pH 8.5) and was left standing for 1 min after which the sample was centrifuged at 14,000 rpm for 1 min. Then the flow- through was collected, as it contained the plasmid, and the column was discarded..

2.15 One tube plasmid Miniprep

This method was occasionally employed to screen clones in order to test if the plasmid contained the full insert or only primer dimers, prior to sequencing. For this quick and inexpensive method, colonies were grown as before, then 1.5 ml of culture was centrifuged at 6,000 rpm for 5 min. The supernatant was discarded and the pellet resuspended in 200 µl of STET buffer (8%w/v sucrose,0.1%v/v TritonX-100, 50 mM EDTA, 50 mM Tris-HCl pH 8.0) 4 µl of lysozyme (50 mg/ml)was added and incubated at room temperature for 5 min. This yielded a viscous solution which was subsequently boiled for 45 sec and centrifuged at 14,000 rpm for 10 min. The supernatant was placed in a clean tube and 8 µl of Cetyltrimethylammoniumbromide (CTAB (5%w/v)) was added, which gave a visibly cloudy suspension, and the sample was centrifuged for 5 min at 14,000 rpm. The pellet was resuspended in 300 µl of 1.2 M NaCl, by vigorous vortexing, and precipitated by addition of 750 µl of ethanol and centrifugation at 14,000 rpm for 10 min. The pellet was rinsed with 100 µl of 70% ethanol/water, dried under vacuum and resuspended in 20 µl of water (Del Sal *et al.*, 1988). This DNA was then labelled (see section 2.18). and analysed by non-denaturing polyacrylamide gel electrophoresis.

2.16 Dideoxy sequencing

40 µl of the plasmid solution obtained from the Promega or Quiagen minipreps (as described above) was denatured by adding 10 µl 2 M NaOH and incubated at room temperature for 10 min. This was then precipitated by adding 15 µl 3 M sodium acetate (pH 4.8), 35 µl water and 300 µl ethanol and placed on dry ice - ethanol for 10 min. The sample was centrifuged for 10 min and the pellet washed with 70% ethanol. The pellet was then dried under vacuum and resuspended in 10 µl distilled water.

This DNA was then sequenced using a T7 sequencing kit (Pharmacia) according to the manufacturers instructions.

To the 10 µl of denatured DNA were added 2 µl annealing buffer and 2 µl universal primer (sequence: 5'-dGTAAAACGACGGCCAGT-3' concentration as provided by the manufacturer) and incubated at 37°C for 20 min. This was then stored at room temperature for at least 10 min before sequencing.

The principle of the dideoxy sequencing (or chain termination method) consists of annealing an oligonucleotide primer onto the single stranded DNA template that is to be sequenced. This primer/template duplex is then used as a substrate for chain extension from the 3' end of the primer by T7 DNA polymerase (other polymerases can be used such as the Klenow, Taq polymerase and reverse transcriptase). Four separate synthesis reactions are carried out, each containing a small amount of a different dideoxynucleotide triphosphate (ddNTP), that is ddATP, ddTTP, ddCTP, or ddGTP as well as the four deoxynucleotide triphosphates. Incorporation of ddNTPs causes chain termination because they lack a 3'-hydroxyl group. Random low level incorporation of a specific ddNTP, in competition with the normal dNTP analogue , will result in a mixture of different length chains. These four sets of products are then fractionated, alongside each other, on a denaturing polyacrylamide gel as explained below.

The labelled products of the sequencing reaction were resolved on 10% polyacrylamide gels containing 8 M urea. These were run at 1500 V, 41 Watts for about 2 hrs and treated as described under gel electrophoresis below (sec.2.23).

2.17 Multisite Strand

The top strand of the multisite 160 base fragment, whose sequence is shown below, was purchased from Oswell together with a 17 base right primer which is shown as a short sequence situated at the bottom of the illustration. The complementary strand was synthesized by polymerase extension using a PCR with the following settings: 94°C for 1 min and 60°C for 5 min. The sample was cycled for 4 cycles.

MS Sequence

5'-CGAC**GGATCC**GCATTCGAGGCTGAGATGACAAAACAGACCCCACCGGACGTACTTTACATAACTCTTCACG
CCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCTTGATCGCGCTACGACTAGTGCAGTTGGAAATCGG
CCATGTGTATTGCCGCATAT**GGATCC**GAGC-3'
3'-GCGTATA**CCTAGG**CTCG-5'

The DNA fragments were then cleaved with *Bam*H I (see section 2.8), whose cleavage sites are indicated as the bold and underlined bases in the sequence above, to provide 'sticky ends' for ligation into the *Bam*H I site of pUC18 (see section 2.9). Then cloning and sequencing were performed.

This sequence contains all 136 possible tetranucleotides and was designed to minimise the presence of long GC or AT-tracts, to avoid homopurine/homopyrimidine sequences and to generate an even distribution of individual dinucleotide steps. Two clones (MS1 and MS2) were obtained in which the insert was placed in opposite orientations. Labelling the 3'-end of the *Hind* III site (see section 2.18) allowed one to visualize the upper portion of the strand in MS2 and the lower portion of the strand in MS1.

2.18 Labelling of DNA fragments

2.18.1 Labelling with $[\alpha]^{32}$P-dATP

DNA fragments containing the inserts, either derived from REPSA or from the multisite or from previous work, were excised from the plasmid by cutting with restriction enzymes *Hind* III and *Sac* I. The DNA was subsequently labelled at the 3'-end of the *Hind* III site by incubating with $[\alpha]^{32}$P-dATP and reverse transcriptase. The labelled DNA fragments of interest were separated from the remainder of the plasmid on a non-denaturing 6% polyacrylamide gel. The position of the labelled DNA was determined by autoradiography (exposing the X-ray film for approximately 10 min). The band was then cut from the gel and eluted into 300 µl of TE (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA) by gentle shaking at 37°C overnight. Subsequently the DNA was ethanol precipitated (see section 2.7) and resuspended in a suitable volume of 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0 so as to give at least a reading of 10 c.p.s/µl as determined by a hand held Geiger counter.

2.18.2 Labelling with [γ]^{32} P-ATP

In an Eppendorf tube 2 µl of oligonucleotide was combined with 2 µl of [γ]^{32} P-ATP, 2 µl of 10 x T4 polynucleotide kinase buffer (700 mM Tris-HCl (pH 7.6), 100 mM $MgCl_2$, 50 mM dithiothreitol) and 1 µl T4 polynucleotide kinase (PNK) . The mixture was incubated at 37°C for 30 minutes. The reaction was stopped with 10 µl of DNase I Stop solution and after denaturing the sample for one minute it was immediately loaded onto a 10 % (w/v) denaturing polyacrylamide gel (i.e. 28 ml of Sequagel in 50 ml mixture). The band containing the labelled oligonucleotide was extracted from the gel and eluted into 300 µl of Tris-EDTA (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA) by gentle shaking at 37°C overnight. Subsequently the primer was ethanol precipitated (see section 2.7) and resuspended in a suitable volume of Tris-EDTA (10mM Tris-HCl, 0.1 mM EDTA, pH 8.0) so as to give at least a reading of 100 c.p.s/µl as determined by a hand held Geiger counter.

2.19 DNase I Footprinting

1.5 µl of drug solution (dissolved in 10 mM Tris-HCl, 10 mM NaCl, pH 8.0) was added to 1.5 µl labelled DNA. Digestion was started by adding 2 µl DNase I (about 0.01 units/ml, dissolved in 20 mM NaCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$) and quenched after one minute by adding 3.5 µl of DNase I stop solution (10 mM EDTA, 1 mM NaOH, 0.1% bromophenol blue, 80% formamide). Then the samples were loaded onto a denaturing gel (see section 2.23)..

2.20 EDTA/ FeII (Hydroxyl radical) Footprinting

For each ligand a control, high and low concentration reaction tubes were prepared. To each of these tubes 2 µl of DNA was added together with 10 µl of the ligand dissolved at the chosen concentrations with Tris-Na (10 mM Tris-HCl, 10 mM NaCl, pH 8.0). Then the mixtures were allowed to reach equilibrium for 30 min. During these 30 min the following mixtures were made:

A) 2 µl of 100 mM $Fe(NH_4)_2SO_4$ in 1000 µl of H_2O.

B) 4 µl of 0.5 M EDTA in 1000 µl of H_2O.

C) 10 µl of 100 mM Ascorbate in 90 µl of H_2O..

D) 10 µl of 30 % H_2O_2 in 1000 µl of H_2O.

The master mix was made rapidly by mixing 40 µl A + 40 µl B + 80 µl C + 80 µl D. 6 µl of this mixture was added to each reaction tube and was incubated for 20 min at room temperature. Then the reaction was stopped by ethanol precipitation (see section 2.7) in this case 2 µl of 3M NaOAc and 70 µl of ethanol were added. After the ethanol precipitation the samples were redisolved in 8 µl of DNase I Stop solution (10 mM EDTA, 1 mM NaOH, 0.1% bromophenol blue, 80% formamide) and were loaded onto a denaturing gel (see section 2.23)..

2.21 Diethyl pyrocarbonate (DEPC)

3 µl of drug solution (dissolved in 10 mM Tris-HCl, 10 mM NaCl, pH 8.0) was added to 3 µl of DNA and was left to equilibrate for 30 min. Then 5 µl of DEPC were added and was left to digest for 20 min at room temperature. The reaction was quenched by ethanol precipitation (sec 2.7). In this case 2 µl of 3 M NaOAc, 10 µl of H_2O and 70 µl of 100% EtOH were added. The samples were cleaved by the addition of 50 µl of 10% piperidine and incubation for 20 min at 100°C. Then the tubes were placed in a speedvac until dried. After the speedvac the samples were washed with 70 % ethanol and redissolved in 8 µl of DNase I stop solution (10 mM EDTA, 1 mM NaOH, 0.1% bromophenol blue, 80% formamide) and were loaded onto a denaturing gel (see section 2.23)..

2.22 GA Track marker

The GA track marker is used in order to identify the bands within a footprinting gel. This marker shows the location of the purines within the DNA sequence thus allowing the identification of the bases within the footprints. The GA track was made by mixing 1.5 µl of labelled DNA, 5 µl of DNase I stop solution (10 mM EDTA, 1 mM NaOH, 0.1% bromophenol blue 80% formamide) and 20 µl H_2O. The mixture was incubated at 100°C for 30 min with the cap open to allow evaporation and reduce the volume to about 5-6 µl, which is enough for loading onto the gel and produce a GA track.

2.23 Gel electrophoresis

Products of the footprinting reactions were resolved by 10% (w/v) polyacrylamide gels containing 8 M urea, made up in TBE (11 g Tris, 5.5 g boric acid and 1 g EDTA per

litre) buffer. Samples were boiled for at least 3 min and crash-cooled on ice before loading onto the gel. 40 cm x 20 cm x 0.3 mm gels were run at 1500 V, 41 W for about 2 hours. The glass plates were then separated and the gel fixed by soaking in 10% (v/v) acetic acid for 10 min. The gel was then transferred to Whatmann 3 MM paper, dried under vacuum at 80°C for 1 hour and subjected to autoradiography at -70°C with an intensifying screen.

2.24 Quantitative analysis of footprinting gels

Gels were exposed to a storage-phosphor screen and analysed using a Molecular Dynamics Storm 860 phosphorimager. The intensity of bands in each footprint was measured using the ImageQuant Software. The values were normalised with respect to the total band intensity in each lane to correct for differences in gel loading and DNase I digestion. Footprinting plots were analysed according to the equation $I/I_0 = C_{50}/(L + C_{50})$, were I and I_0 are the band intensities in the presence and absence of the ligand respectively, L is the ligand concentration and C_{50} is the drug concentration which reduces the intensity of bands in the footprint by 50%. Since the concentration of target DNA (typically about 10 nM) is well bellow the ligand dissociation constant, C_{50} approximates to the dissociation constant. Curves were fitted to the data points using Fig P for windows (ver.1) (Biosoft). This program calculates the standard error for each C_{50} value from the square of the variance found on the diagonal of the covariance matrix. Errors quoted for the C_{50} values therefore refer to the curve fitting procedure and not to the individual data points.

CHAPTER 3

Determination of the best conditions for REPSA

3.1 Introduction:

This chapter is mainly concerned with the development of REPSA as a tool for determining the preferred DNA binding sites of small molecules, so as to assess whether the interaction with their primary recognition sequences is affected by the flanking base pairs. As we have seen in the introduction (see section 1.8) REPSA has been successfully employed for determining the binding sites of distamycin. However, the sequences used in this study contained a very large random cassette (21 base pairs) yielding many overlapping and adjacent binding sites within each fragment (Hardenbol and van Dyke, 1997). Therefore, we modified REPSA in order to be useful for identifying single unique binding sites. One of the modifications was the synthesis of sequences that contained only five random base pairs with one restriction enzyme recognition site within the fixed region. The reasons for these changes will be addressed below. As the experiments progressed many technical problems were encountered and these are addressed in turn. A scheme showing the different parameters that were altered is presented in Fig 3.1. This chapter outlines the various modifications to the REPSA protocol which were employed during this work.

3.2 Which ligands?

The ligands that were chosen for this project were distamycin, Hoechst 33258, mithramycin, echinomycin, [*N*-MeCys^3,*N*-MeCys^7]TANDEM and actinomycin D (see sections 1.3, 1.4 and 1.5). The reason for choosing these ligands was because they have each been very well characterized by NMR spectroscopy, footprinting and X-ray crystallography and their primary binding sites are well known. In addition, these agents possess a high affinity for their binding sites which may be affected by the flanking bases. These agents also represent a diverse series of ligands which bind to DNA by different

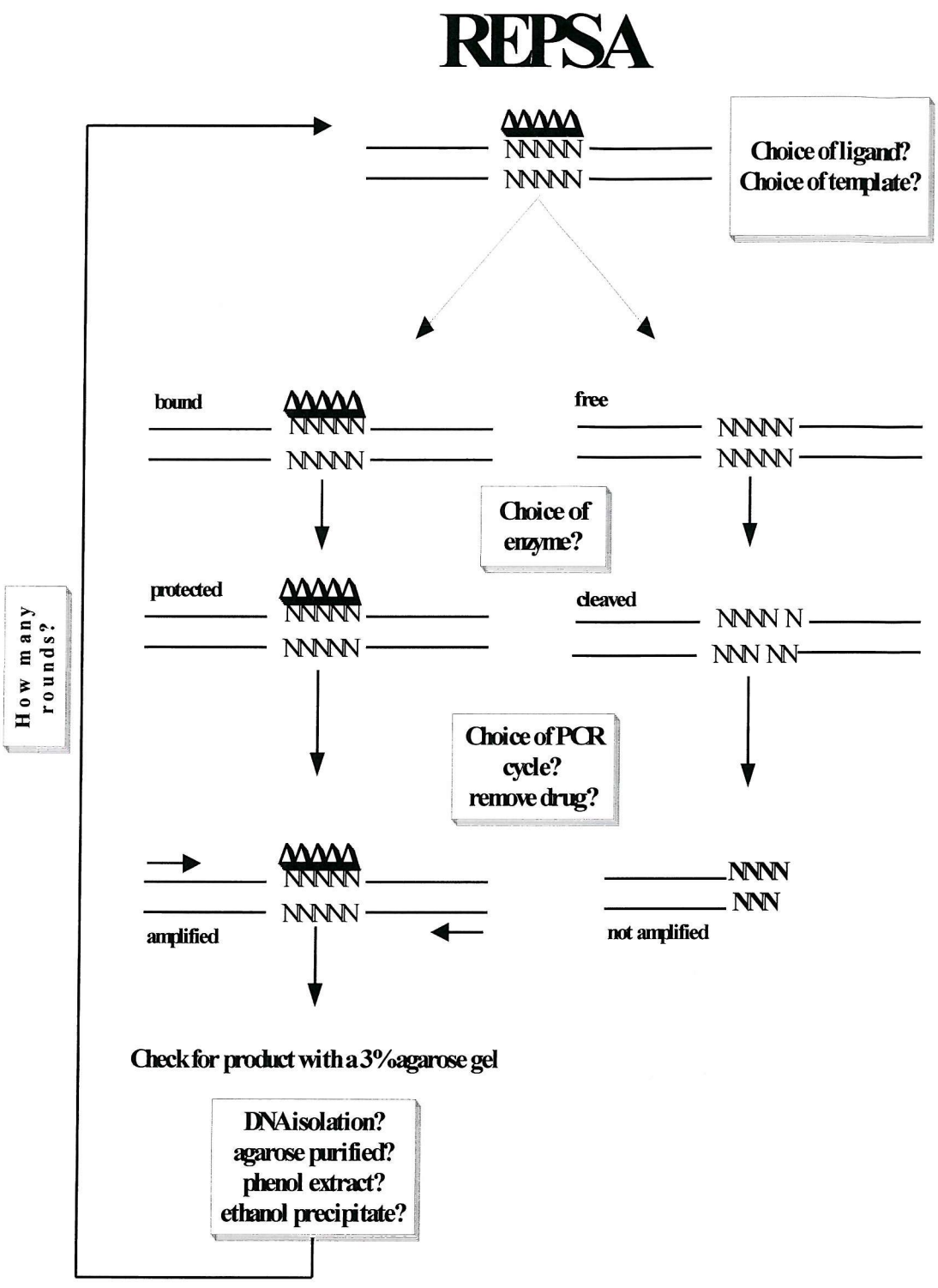


Figure 3.1 A diagram showing all the different steps of REPSA and the questions that were addressed during development

mechanisms including intercalation, bis-intercalation and groove binding. These ligands should therefore be excellent tools for evaluating the accuracy of REPSA, since we can already predict most of the primary sequences that should be targeted.

3.3 Which template?

The REPSA templates have to be designed in such a way that the fixed region contains the binding site for the restriction enzyme while the cleavage site is within the random region. Also to make it easier for the final cloning step the restriction site for *Bam*HI is included at both ends of the template. Most of the templates that are used in this work contain a region of five random bases because several of the known DNA binding agents have a recognition site between 2 to 4 base pairs long. This region should therefore only be able to accommodate one ligand molecule

3.3.1 Template design for *BglI*

BglI is a type II enzyme that recognizes a sequence which contains five random base pairs within its recognition site (GCCNNNNNGGC) with cleavage occurring in the middle of the random region.

The template that was designed for *BglI* was:

```
5'-GTGGAGCTTGGATCCACCAG-3'
5'-GTGGAGCTTGGATCCACCAGCCNNNNNGGCGTGGTAGTGGAGGATCCATG-3'
                                  3'-CACCATCACCTCCTAGGTAC-5'
```

The bold region is the cleavage site for *Bam*HI, which is included to permit subsequent cloning, and the bold and underlined region is the recognition sequence for *BglI*. The short sequences shown above and below the template are the PCR primers used for amplification.

3.3.2 Template design for *BsgI*

*Bsg*I is a type IIs (the s refers to staggered cleavage), and cleaves the DNA at 16 (5'-3') and 14 (3'-5') bases away from its binding site

Five different templates were designed for *Bsg*I experiments containing other 5-random bases at different locations , or one long 17-base random sequence. Their sequences are shown below:

```
     5'-GGTAGGATCCAGAGTGCAG-3'                    ↓
1)   5'-GGTAGGATCCAGAGTGCAGNNNNNNNNNNNNNNNNGACCTGAGAGGATCCATGG-3'
2)   5'-GGTAGGATCCAGAGTGCAGACAGGTTCACTCNNNNNGACCTGAGAGGATCCATGG-3'
3)   5'-GGTAGGATCCAGAGTGCAGGCACTCAGNNNNNGCTGGACCTGAGAGGATCCATGG-3'
4)   5'-GGTAGGATCCAGAGTGCAGGACAGNNNNNTCGACTGGACCTGAGAGGATCCATGG-3'
5)   5'-GGTAGGATCCAGAGTGCAGGNNNNNGCACTCGACTGGACCTGAGAGGATCCATGG-3'
                                                  3'-CTGGACTCTCCTAGGTACC-5'
```

The bold bases are the recognition site for *Bam*HI (included to permit subsequent cloning), the underlined and bold are the recognition site for *Bsg*I and the arrow shows the position at which cleavage takes place. These templates are called 1-17 template (1), 13-17 template (2), 9-13 template (3), 6-10 template (4) and 2-6 template (5), in descending order respectively. The short sequences are the PCR primers used for the amplification of these templates.

These different templates were examined as early results obtained with the 1-17 template (described below) suggested that many of the good binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM and distamycin were located closer to the recognition site for the enzyme than to its cleavage site. This indicated that the best results might be obtained when the random region was located away from the cleavage site. Once again the random region was only five base pairs long because most known binding agents recognise between 2 to 4 base pairs.

3.3.3 Template design for *Fok*I

*Fok*I is a type IIs restriction enzyme and has the recognition sequence 5'-GGATG-3'. It cleaves the DNA 9 and 13 nucleotides away from this binding site. The following template was designed for this enzyme in which the random region was positioned within the cleavage site.

```
5'-GTCCAAGCTTCTGGAGGGATG-3'              ↓
5'-GTCCAAGCTTCTGGAGGGATGGTCCTGAACNNNNNGTTCACCTCTGCACGATCCTAG-3'
                                         3'-AAGTGGAGACGTGCTAGGATC-5'
```

The bold bases on the left are the recognition site for *Hind*III and the ones on the right of the template are the recognition site for *Sau*3A1. These were included to permit cloning of the selected fragments. The underlined and bold are the recognition site for *Fok*I and the arrow indicates where the cleavage takes place.

3.4 Which enzyme?

Our initial experiments used type II enzymes that had random sequences within their recognition site, because of their simplicity of action. The first enzyme was *Bgl*I, which has a recognition sequence 5'-GCCNNNNNGGC-3'. However, experiments with this enzyme were stopped because Hoechst 33258 failed to protect the site from cleavage and distamycin needed very high concentrations to do so (see section 3.4.1). Similar experiments with *Psh*AI which has a recognition sequence 5'-GACNNNNGTC-3 were attempted. This enzyme was also not used for the same reasons as *Bgl*I. For these enzymes , if the enzyme cuts from the DNA major groove then the minor groove of the random region will face away from the enzyme surface. As a result ligands that bind in this region may not block the enzyme activity.

The next enzyme used was *Bsg*I. This type IIs enzyme has been used in previous REPSA studies by Hardenbol and Van Dyke (1997) and their results show that the cleavage of this enzyme is inhibited by distamycin when bound to DNA.. *Bsg*I cleaves the DNA at 16 (5'-3') and 14 (3'-5') bases away from its binding site. Another type IIs enzyme *Fok*I, was also employed in some experiments because we were concerned about the appearance of some mutations in experiments with *Bsg*I. Similar mutations were found in experiments with *Fok*I and *Bsg*I was the selected enzyme for all future experiments due to its lower cost.

3.4.1 Ligand protection from cleavage by *Bgl*I and *Psh*AI

Before doing REPSA it is essential to determine whether the enzyme is inhibited by ligand binding. This was only tested for *Bgl*I and *Psh*AI in this work since *Fok*I and *Bsg*I have previously been shown to be inhibited by distamycin (Hardenbol and van Dyke,1992). To this end a template was

designed containing the recognition sequence for *BglI* with a known good binding site for Hoechst 33258 (AAATT) in the random region. Thus, the template had the following sequence:

5'-GCTCCTGCC***AAATT***GGCTCCTCG-3'

This template was labelled with [γ-^{32}P] dATP (sec.2.18.2) and then exposed to 0, 0.1, 1, and 10 μM Hoechst 33258, and was digested with 1μl of *BglI* (10 u/μl) for 1, 5, and 30 minutes. The samples were loaded onto a denaturing polyacrylamide gel and the results are shown in Fig 3.2. This shows that even at a concentration of 10μM, Hoechst 33258 does not protect the site from cleavage by *BglI*. The experiment was repeated with higher concentrations (100, 50, 10 μM) of Hoechst 33258 and the same result was found (Fig 3.3). It therefore appears that *BglI* is not a suitable enzyme for use in REPSA, at least with minor groove binding agents.

A similar experiment was done with *Psh*AI.. A new template was designed containing *Psh* AI recognition site, and the best Hoechst binding site situated in the random region.

5'-TCAGTGCGAC***AATT***GTCGCGCACTG-3'

This template was also labelled with [γ-^{32}P] dATP (sec.2.18.2) and then exposed to Hoechst 33258 were (0, 0.01, 0.1, and 10 μM), and 0.4 u /μl *Psh* AI. The samples were digested for 1, 5, and 30 min. and stopped with DNase I stop solution and run on a denaturing polyacrylamide gel. The results are shown in Fig 3.4 and reveal that Hoechst 33258 does not protect this site from cleavage.

Since distamycin also binds to regions of high A and T content (Fox and Waring 1984), the *BglI* template was used to test if this agent could protect from *BglI* cleavage . The results are presented in Fig. 3.5 and show that distamycin only protected the template from cleavage at high concentrations and that significant cleavage still occurred in the presence of 10 μM ligand..

From these experiments it is clear that these enzymes are not potently inhibited by minor groove binding ligands. It is possible that these enzymes cleave DNA from the major groove, while the agents Hoechst 33258 and distamycin bind in the minor groove; therefore very little protection was observed. Distamycin may be successful at high concentrations possibly because it changes its stoichiometry from 1:1 to 2:1, thus, distorting the DNA structure more and thereby inhibiting

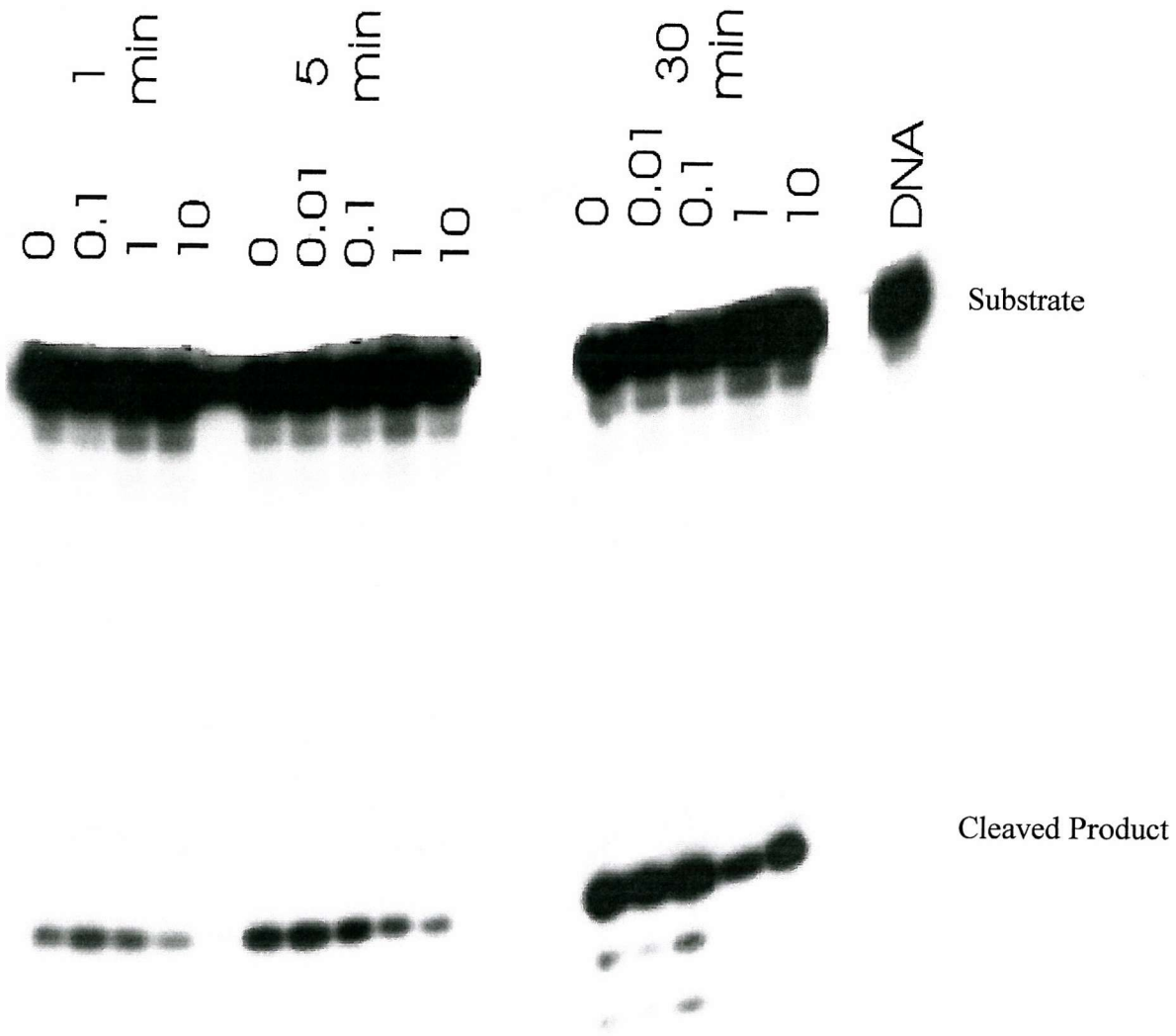


Figure 3.2 A denaturing polyacrylamide gel showing the cleavage pattern after digestion with *Bgl* I for 1, 5 and 30 minutes. The DNA was reacted with different concentrations of Hoechst 33258 which are indicated at the top of the gel lane in μM. The lane labeled DNA is the control for uncleaved DNA.

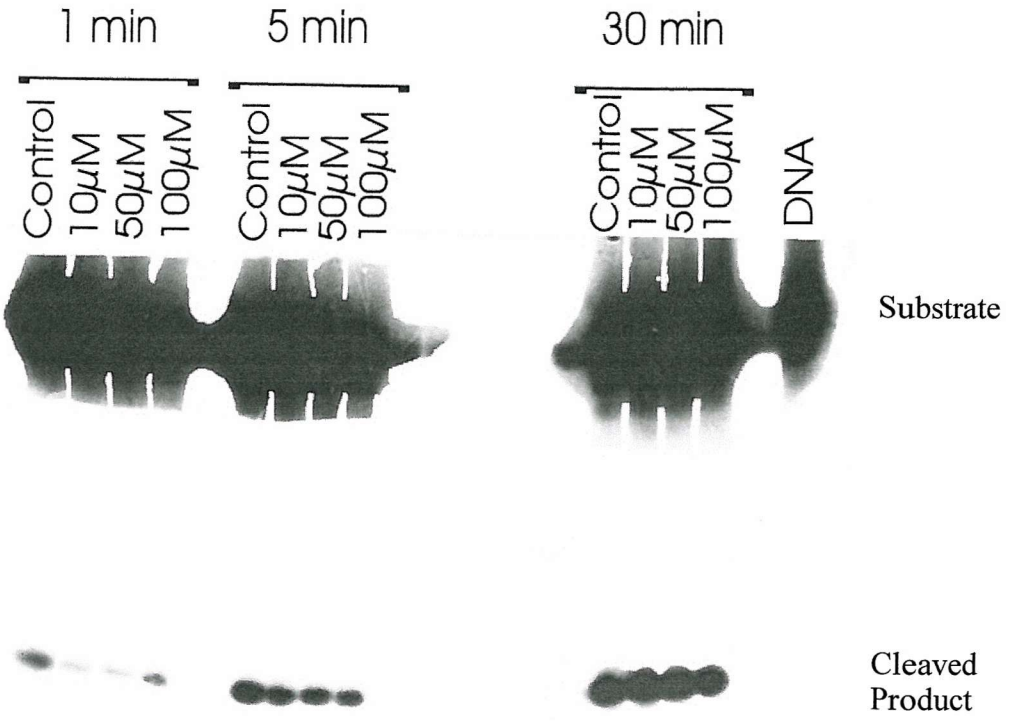


Figure 3.3 A denaturing polyacrylamide gel showing the cleavage pattern after digestion with *Bgl* I for 1, 5 and 30 minutes. The DNA was reacted with different concentrations of Hoechst 33258 which are indicated at the top of the gel lane in μM. The last lane on the right is the control for uncleaved DNA.

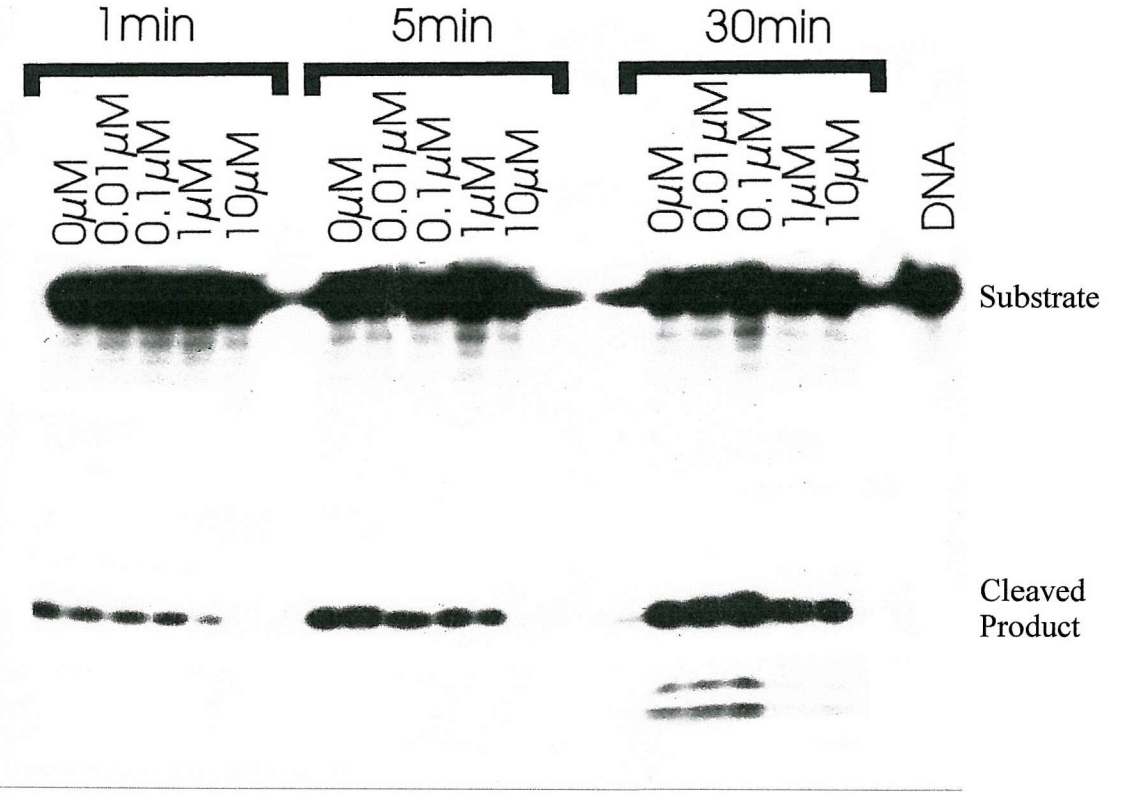


Figure 3.4 A denaturing polyacrylamide gel showing the cleavage pattern after digestion with *Psh* A I for 1, 5 and 30 minutes. The DNA was reacted with different concentrations of Hoechst 33258 which are indicated at the top of the gel lane in μM. The last lane on the right is the control for uncleaved DNA.

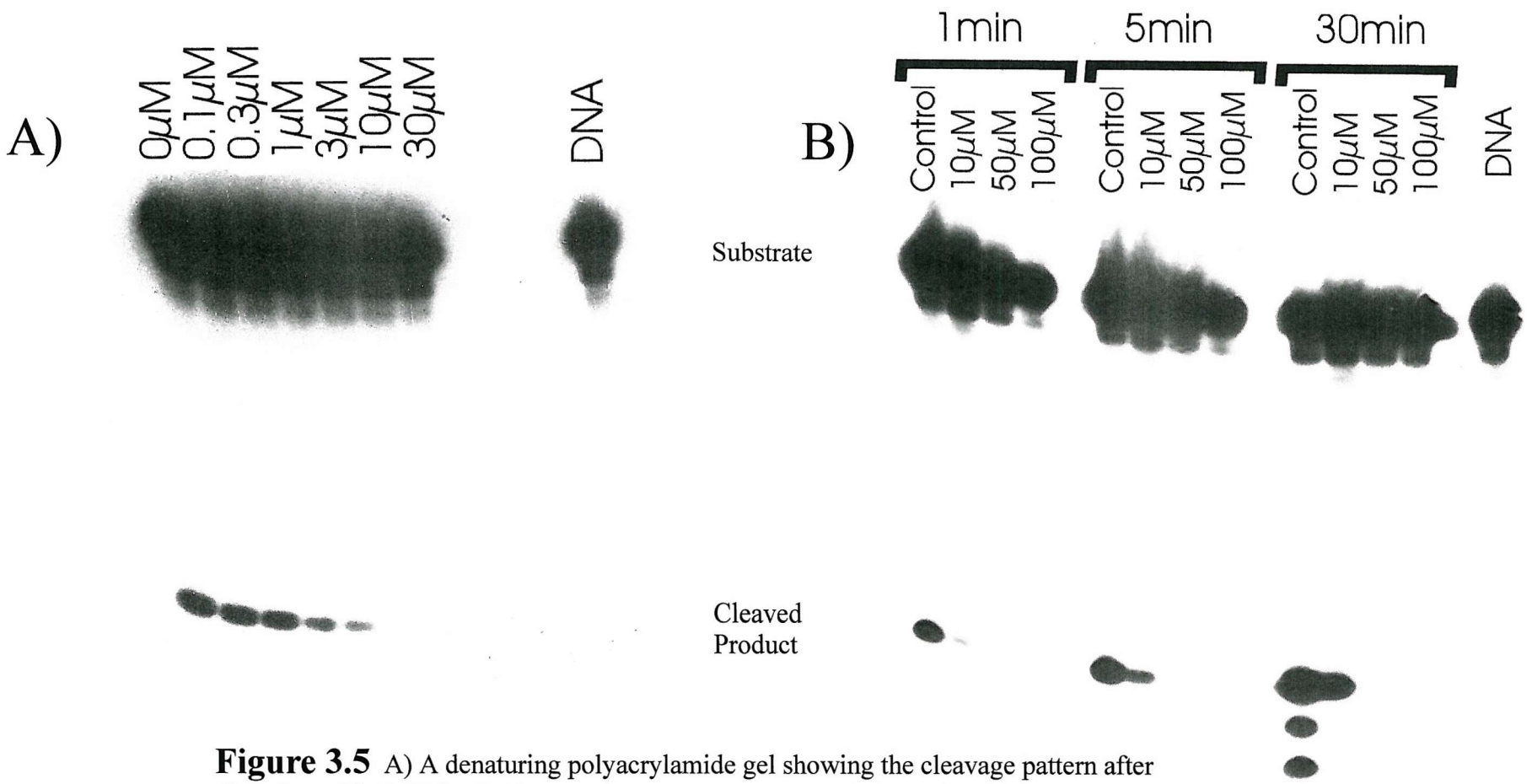


Figure 3.5 A) A denaturing polyacrylamide gel showing the cleavage pattern after digestion with *Bgl* I for 1 minute at low concentrations of distamycin which are indicated at the top of each lane in μM B) A denaturing polyacrylamide gel showing the cleavage pattern after digestion with *Bgl*I for 1, 5 and 30 minutes. The DNA was reacted with different concentrations of distamycin which are indicated at the top of the gel lane in μM.The lanes labeled DNA are the controls for uncleaved DNA.

cleavage by the enzyme.

3.4.2 Cleavage of template by *Fok*I

Efficient use of REPSA requires that all the unbound DNA is digested by the selecting restriction enzyme. It is therefore necessary to add sufficient enzyme to ensure good cleavage over an appropriate digestion time. Too much enzyme for too long will cause cleavage of some selected sites while too little will retain unbound sites. The time course of digestion by *Fok*I was therefore examined. First the template was made double stranded as previously described and radiolabeled by including [α-^{32}P]dATP in the PCR reaction. This template was exposed to different concentrations of the restriction enzyme (i.e. 1.25 units/ µl and 2 units /µl *Fok*I) for different periods of time (1, 5, 10, 30 and 60 minutes). Afterwards the samples were denatured and loaded onto a denaturing polyacrylamide gel (Fig 3.7). This was exposed to a phosphoimaging screen to quantify the amount of cleaved fragment. The results are shown in Fig. 3.6

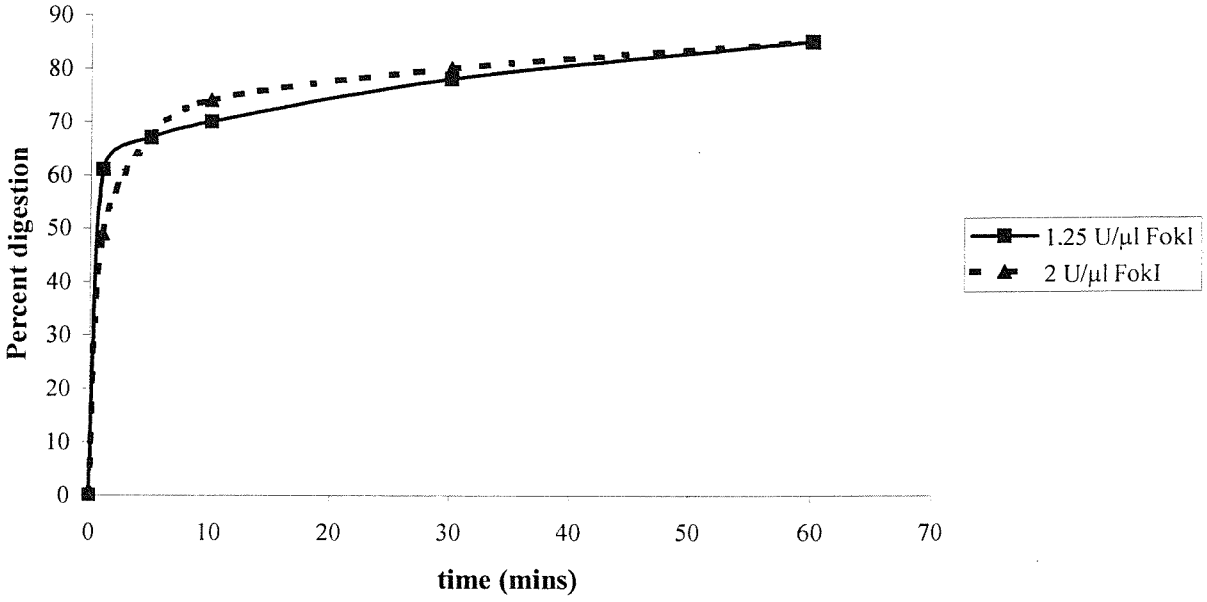


Figure 3.6 Graph showing the percentage cleavage of *Fok*I over time

It can be seen that after 60 minutes digestion both concentrations of the enzyme yielded 85% cleaved DNA. In REPSA experiments with this enzyme, the concentration chosen was 1.25 units/ µl with a digestion time of 60 minutes.

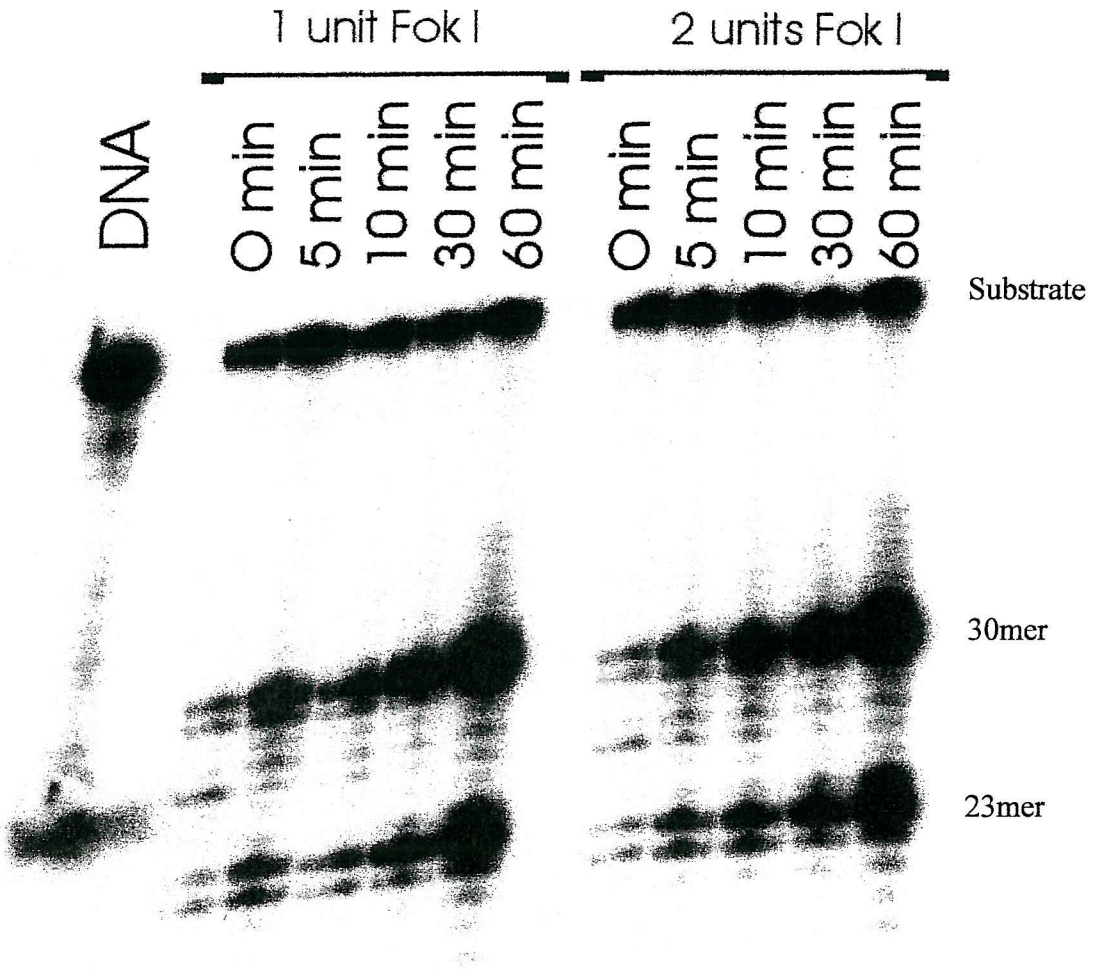


Figure 3.7 A denaturing polyacrylamide gel showing the cleavage pattern after digestion with *Fok*I for 0, 5, 10, 30 and 60 minutes at two different enzyme concentrations The lane marked DNA is the control for uncleaved DNA. Two cleavage products are evident because *Fok*I cuts the fragment assymetrically, yielding labelled products of length 30 and 23 bases.

3.4.2 Does *Bsg*I cleave all DNA sequences?

In some early REPSA experiments using *Bsg*I as the selecting enzyme, with 1 μM Hoechst 33258 and the 13-17 random template, some clones were obtained which were clearly not related to Hoechst binding sites. The sequences of these are shown below and contain GC base pairs (instead of AT) within the selected region.

```
C1) 5'-GATCCAGAGTGCAGACAGGTTCACTC TTCCG GACCTGAGAGGATC-3'
C2) 5'-GATCCAGAGTGCAGACAGGTTCACTC AGAGA TGGACCTGAGAGGATC-3'
C3) 5'-GATCC   GAG CTCGTACCCGAGTCTCTCAGTCCGTCGATATCTGTTGCACT
     CTGATCCAGAGTGCAGACAGGTTCACTC CAACA GACCTGAGAGGATC-3'
C4) 5'-GATCCAGAGTGCAGACAGGTTCACTC ACAAG GACCTGAGAGGATC-3'.
C5) 5'-GATCCAGAGTGCAGACAGGTTCACTC AGGGG GAC TGAGAGGATC-3'
```

We reasoned that these could arise from either, failure of REPSA to select bound DNA, or the inability of the enzyme to cut some sequences from the pool of random oligonucleotides. Plasmids containing these sequences were therefore subjected to cleavage with *Bsg*I. Since these plasmids should only contain one *Bsg*I site successful cleavage should convert closed circular to linear DNA. The results for sequence C1 are shown in Fig 3.8, while those for sequences C2 - C5 are shown in Fig 3.9.

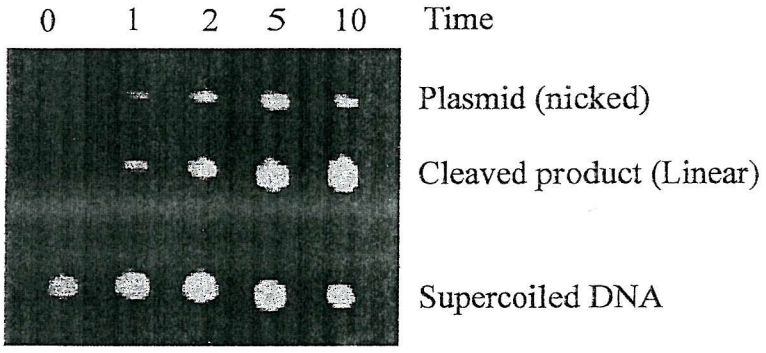


Figure 3.8 A 0.7 % agarose gel showing the cleavage pattern of 10μl of plasmid containing C1 with 1 μl of *Bsg*I obtained at different digestion times.

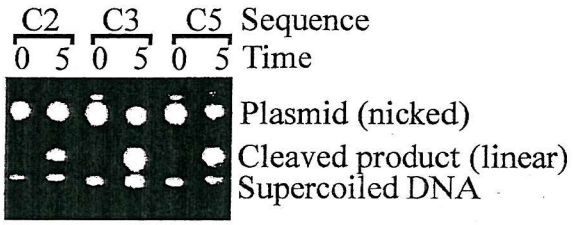


Figure 3.9 A 0.7% agarose gel showing the pattern of C2, C3 and C5 after digestion with *Bsg*I for 0 and 5 minutes.

These two figures show that *Bsg*I is able to cut all these clones, confirming that the failure to select Hoechst binding sites arises from some other aspects of the REPSA protocol. Another experiment using the 6-10 random template and 20 μM [*N*-MeCys^3, *N*-MeCys^7]TANDEM REPSA yielded sequences which were again GC-rich and were not related to the known ligand binding sites. These sequences are shown below.

```
1)    5'-GATCCAGAGTGCAGGACAG GCATA TCGACTGGACCTGAGAGGATC-3'
2)    5'-GATCCAGAGTGCAGGACAG GGGGG TCGACTGGACCTGAGAGGATC-3'
3)    5'-GATCCAGAGTGCAGGACAG GGGGG TCGACTGGACCTGAGAGGATC-3'
4)    5'-GATCCAGAGTGCAGGACAG TATGC TCGACTGGACCTGAGAGGATC-3'
5)    5'-GATCCAGAGTGCAGGACAG GGAAA TCGACTGGACCTGAGAGGATC-3'
6)    5'-GATCCAGAGTGCAGGACAG TAGTA TCGACTGGACCTGAGAGGATC-3'
7)    5'-GATCCAGAGTGCAGGACAG GTTGG TCGACTGGACCTGAGAGGATC-3'
8)    5'-GATCCAGAGTGCAGGACAG GGAGG TCGACTGGACCTGAGAGGATC-3'
9)    5'-GATCCAGAGTGCAGGACAG GGACG TCGACTGGACCTGAGAGGATC-3'
10)   5'-GATCCAGAGTGCAGGACAG GAAAA TCGACTGGACCTGAGAGGATC-3'
11)   5'-GATCCAGAGTGCAGGACAG GGAGC TCGACTGGACCTGAGAGGATC-3'
12)   5'-GATCCAGAGTGCAGGACAG TCGAT TCGACTGGACCTGAGAGGATC-3'
13)   5'-GATCCAGAGTGCAGGACAG GTTGC TCGACTGGACCTGAGAGGATC-3'
14)   5'-GATCCAGAGTGCAGGACAG GGGGG TCGACTGGACCTGAGAGGATC-3'
15)   5'-GATCCAGAGTGCAGGACAT AGATG TCGACTGGACCTGAGAGGATC-3'
16)   5'-GATCCAGAGTGCAGGACAT TAAGG TCGACTGGACCTGAGAGGATC-3'
```

We therefore again considered the possibility that some DNA sequences might be able to adopt unusual DNA structures which are resistant to digestion with *Bsg*I. These will then be selected in the REPSA technique by virtue of their unusual structures, rather than from ligand binding. The sequences of 4 of these clones obtained with [*N*-MeCys^3,*N*-MeCys^7]TANDEM were selected for testing *Bsg*I cleavage.

2) 5'-GATCCAGAGTGCAGGACAG ***GGGGG*** TCGACTGGACCTGAGAGGATC
7) 5'-GATCCAGAGTGCAGGACAG ***GTTGG*** TCGACTGGACCTGAGAGGATC
8) 5'-GATCCAGAGTGCAGGACAG ***GGAGG*** TCGACTGGACCTGAGAGGATC
9) 5'-GATCCAGAGTGCAGGACAG ***GGACG*** TCGACTGGACCTGAGAGGATC

Plasmids containing these sequences were prepared and digested with *Bsg*I for various times (0, 5, 10, 15 and 30 min.), the digestion was stopped with the addition of 2 µl of 0.5M EDTA. Then the products of digestion were resolved on a 0.7% agarose gel. Since each of these plasmids only contains one cutting site for *Bsg*I, the digestion of supercoiled DNA should produce a single linear DNA product. The results of these reactions are shown in Fig 3.10.

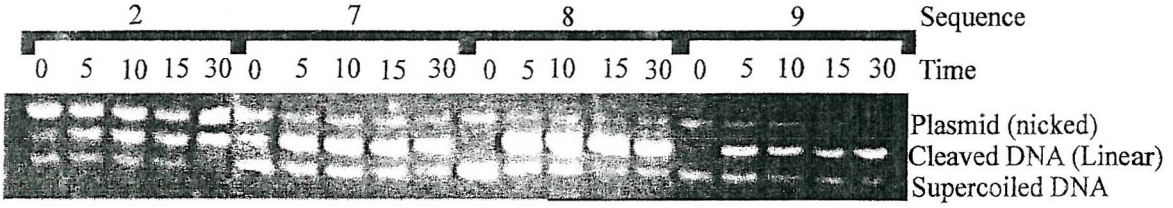


Figure 3.10 A 0.7 % agarose gel showing the cleavage patterns of the 4 templates shown above with *Bsg*I.

It can be seen that sequences 2 and 9 are cleaved at a slightly slower rate than 7 and 8. However, all of these sequences are cleaved by *Bsg*I, suggesting that their structure is not resistant to digestion by this enzyme. It is therefore likely that [*N*-MeCys^3,*N*-MeCys^7] TANDEM binding to this random region (located between the enzyme binding and cutting sites) is unable to protect from *Bsg*I cleavage. This is considered further in Chapter 4.

It is well known that G-rich sequences can form quadruplex structures involving Hoogsteen binding between guanines. Therefore, in order to reduce the incidence of these unusual structures dGTP was replaced with C^7-dGTP (7-deaza-2'-deoxyguanosine) in some of the PCR reactions. This base analogue is able to form normal G-C base pairs, but is unable to form Hoogsteen base pairs.

3.5 Chosing a PCR cycle

The initial PCR conditions were chosen by calculating the melting temperatures of the primers used for the 13-17 template, designed for *Bsg*I, with the following simple equation:

$$Tm = 4(G+C) + 2(A+T)$$

This gave a value of 60°C for both primers. The annealing temperature was chosen to be 5°C lower than the melting temperature of the primers (i.e. 55°C). The chosen extension temperature was 57°C. However the extension and annealing temperatures were changed during the course of this work due to appearance of primer dimers and/or non-selected binding sites and are specified in the individual sections. The denaturation temperature was always 94°C.

3.5.1 Testing different annealing and extension temperatures

The 13-17 template, which was designed for *Bsg*I, was made dsDNA under different annealing and extension temperatures, using the concentrations specified in the Materials and Methods.

1) 55°C annealing, 57°C extension.
2) 55°C annealing, 65°C extension.
3) 60°C annealing, 65°C extension.
4) 60°C annealing, 70°C extension.
5) 23°C annealing, 50°C extension.
6) 60°C annealing, 74°C extension.
7) 65°C annealing, 74°C extension.

All these conditions yielded a PCR product which was visible on ethidium stained agarose gels after 4 cycles. These templates were tested for deletions or mutations, before proceeding to REPSA, by cloning and sequencing. This was only performed for the most stringent conditions (6, and 7). The resulting sequences obtained from eight of these clones were:

```
    5'-GATCCAGAGTGCAGACAGGTTCACTC NNNNN GACCTGAGAGGATC-3'
6,1)5'-GATCCAGAGTGCAGACAGGTTCACTC CAGAT GACCTGAGAGGATC-3'
6,2)5'-GATCCAGAGTGCAGACAGGTTCACTC GCCGA GACCTGAGAGGATC-3'
6,3)5'-GATCCAGAGTGCAGACAGGTTCACTC GAGCG GACCTGAGAGGATC-3'
6,4)5'-GATCCAGAGTGCAGACAGGTTCACTC GGGGC GACCTGAGAGGATC-3'
7,1)5'-GATCCAGAGTGCAGACAGGTTCACTC ACAAA GACCTGAGAGGATC-3'
7,2)5'-GATCCAGAGTGCAGACAGGTTCACTC GAGTC GACCTGAGAGGATC-3'
7,3)5'-GATCCAGAGTGCAGACAGGTTCACTC TGGGG GACCTGAGAGGATC-3'
7,4)5'-GATCCAGAGTGCAGACAGGTTCACTC GCGAG GACCTGAGAGGATC-3'
```

These sequences are complete with no deletions or mutations and appear to be sufficient for forming intact DNA templates. Therefore, the template preparation under conditions for 7 were tested with 0.1 μM Hoechst 33258 in REPSA. However primer dimers, in which the central portion was deleted, were evident after five rounds of REPSA, their sequences are shown below.

```
     5'-GATCCAGAGTGCAG-3'
     5'-GATCCAGAGTGCAGACAGGTTCACTCNNNNNGACCTGAGAGGATCCATG-3'
                                      3'-CTGGACTCTCCTAGGTAC-5'

1H)  5'-GATCCAGAGTGCAGAC                    CTGAGAGGATCCATG-3'
2H)  5'-GATCCAGAGTGCAG                   GACCTGAGAGGATCCATG-3'
3H)  5'-GATCCAGAGTGCAG                   GACCTGAGAGGATCCATG-3'
4H)  5'-GATCCAGAGTGCAGAC                    CTGAGAGGATCCATG-3'
5H)  5'-GATCCAGAGTGCAGAC                    CTGAGAGGATCCATG-3'
6H)  5'-GATCCAGAGTGCAGAC                    CTGAGAGGATCCATG-3'
7H)  5'-GATCCAGAGTGCAGAC                    CTGAGAGGATCCATG-3'
8H)  5'-GATCCAGAGTGCAGAC                    CTGAGAGGATCCATG-3'
```

Several procedures were therefore tested for purifying the full length products, removing the shorter primer dimers. These are described in section 3.7.

3.5.2 More temperature tests during REPSA

In an attempt to prevent the formation of primer dimers several more PCR conditions were tested for the 13-17 template during the REPSA reaction with *Bsg*I and 1 μM Hoechst 33258

8) 60°C for annealing and extension.

9) 50°C for annealing and 72°C for extension.

10) 25°C for annealing and 50°C for extension.

The products obtained at each temperature were analysed by agarose gel electrophoresis. For condition 8, primer dimers were evident after REPSA round seven, for condition 9 there was primer dimer formation after REPSA round five, while condition 10 did not show formation of primer dimers. However, when that product was cloned and sequenced there was no evidence for selection of Hoechst 33258 binding sites. It was therefore decided to try another enzyme (*Fok*I) in the hope that this might yield better selectivity.

An experiment with *Fok*I and 0.1 μM Hoechst 33258, and with the PCR set at 50°C for annealing and extension for 9, 12 and 15 cycles , showed that primer dimers were still formed after 4 rounds of REPSA. Therefore it was decided that the best way to exclude primer dimers from accumulating in successive rounds of PCR was to separate them from the product on an agarose gel and then purify the DNA. Several methods were used for this and are described in section 3.7. The selected restriction enzyme for future experiments was *Bsg*I, due to its lower cost, and the PCR cycle which was chosen was 95°C for denaturing for 1 min and the annealing and extension temperature was set at 60°C for 4 minutes. Product was always collected after 9, 12 and 15 PCR cycles.

3.6 Drug removal before PCR

In early experiments the products from the restriction digestion were simply added to the PCR reaction without stopping the reaction (the restriction enzyme will be inactivated by the first denaturation cycle) and without any purification to remove the drug or enzyme. This was the procedure used by Hardenbol and van Dyke (1997). However we routinely stopped the restriction enzyme by addition of 2 μl of 0.5 M EDTA (final concentration of 10 μM). The total volume was then adjusted to 100 μl with water giving a DNA concentration of 10 ng / μl. 2 μl of this reaction were used in the next round of REPSA. With this protocol we obtained an unusually high number of sequences containing base mutations in the template when examining the binding of [*N*-MeCys^3,*N*-MeCys^7]TANDEM . We therefore explored the possibility that this might arise from the presence of small amounts of drug

which are carried over into the PCR mixture.

To find out if there is any interference of the drug, the following experiment was performed. The dsDNA template was formed under three different conditions a) in the presence of 10 μM [*N*-MeCys^3,*N*-MeCys^7]TANDEM, b) in the presence of 10 μM echinomycin and c) a control with no drug. These PCR products were subsequently cloned and sequenced and the following sequences were obtained:

a) For [*N*-MeCys^3,*N*-MeCys^7]TANDEM

```
     5'-GATCCAGAGTGCAGGCACTCAG NNNNN GCTGGACCTGAGAGGATC-3'
1)   5'-GATCCAGAGTGCAGGCACTCAG ACACA GCTGGACCTGAGAGGATC-3'
2)   5'-GATCCAGAGTGCAGGCACTCAG AATGT GCTGGACCTGAGAGGATC-3'
3)   5'-GATCCAGAGTGC□GGCACTCA  GAAGC GCTGGACCTGAGAGGATC-3'
4)   5'-GATCCAGAGTGCAGGCACTCAG AAAGG□GCTGGACCTGAGAGGATC-3'
5)   5'-□□□□□□□□□□□□□□□□GGCACTCAG TTAAG GCTGGACCTGAGAGGATC-3'
6)   5'-GATCCAGAGTGCAGGCACTCAG GGACG GCTGGACCTGAGAGGATC-3'
7)   5'-GATCCAGAGTGCAGGCACTCAG TGGCG GCTGGACCTGAGAGGATC-3'
8)   5'-GATCCAGAGTGCAGGCACTCAG GGCGG GCTGGACCTGAGAGGATC-3'
9)   5'-GATCCAGAGTGCAGGCACTCAG□CAACG GCTGGACCTGAGAGGATC-3'
```

b) For echinomycin

```
     5'-GATCCAGAGTGCAGGCACTCAG NNNNN GCTGGACCTGAGAGGATC-3'
1)   5'-GATCCAGAGTGCAGGCACTCAG AGTAC GCTGGACCTGAGAGGATC-3'
2)   5'-GATCCAGAGTGCAGGCACTCAG GGGGG GCTGGACCTGAGAGGATC-3'
3)   5'-GATCCAGAGTGCAGGCACTCAG AGAGG GCTGGACCTGAGAGGATC-3'
4)   5'-GATCCAGAGTGCAGGCACTCAG GGGAG GCTGGACCTGAGAGGATC-3'
5)   5'-GATCCAGAGTGCAGGCACTCAG AGGAC GCTGGACCTGAGAGGATC-3'
6)   5'-GATCCAGAGTGCAGGCACTCAG GGCGG GCTGGACCTGAGAGGATC-3'
```

c) For control

```
     5'-GATCCAGAGTGCAGGCACTCAG NNNNN GCTGGACCTGAGAGGATC-3'
1)   5'-GATCCAGAGTGCAGGCACTCAG AGGGG GCTGGACCTGAGAGGATC-3'
2)   5'-GATCCAGAGTGCAGGCACTCAG GCAGT GCTGGACCTGAGAGGATC-3'
3)   5'-GATCCAGAGTGCAGGCACTCAG TAGAC GCTGGACCTGAGAGGATC-3'
4)   5'-GATCCAGAGTGCAGGCACTCAG AAGTG GCTGGACCTGAGAGGATC-3'
5)   5'-GATCCAGAGTGCAGGCAGTCAG AACTC GCTGGACCTGAGAGGATC-3'
6)   5'-GATCCAGAGTGCAGGCAGTCAG ACAAA GCTGGACCTGAGAGGATC-3'
```

From these sequences it can be seen that 4/9 sequences obtained with [*N*-MeCys^3,*N*-MeCys^7]TANDEM contained mutational errors (marked with □) while 0/6 mutants were

obtained for both echinomycin and the control. It is not clear why [*N*-MeCys^3,*N*-MeCys^7] TANDEM, and not echinomycin, induces these mutations. However, in order to prevent this effect in subsequent experiments the [*N*-MeCys^3,*N*-MeCys^7]TANDEM was removed from the digestion mixture before PCR amplification. This was achieved by phenol extraction followed by ethanol precipitation.

3.7 Which DNA isolation method ?

In the paper by Hardenbol and van Dyke (1997), the primers were radio-labelled and PAGE gels were used to check for the presence of appropriate PCR product. However, in the experiments described in this thesis PCR product formation was assessed on 2% agarose gels which were stained with ethidium bromide. After successive rounds of REPSA, a faster migrating species was often observed as shown in Fig 3.11. This was characterized by sequencing and the results indicated that this product was due to primer dimer formation. Therefore, a primer dimer control was synthesized with the following sequence.

5'-GGTAGGATCCAGAGTGCAGGACCTGAGAGGATCCATGG-3'

This template was made double stranded using the usual right primer, and was used as a marker for monitoring the formation of the unwanted product.

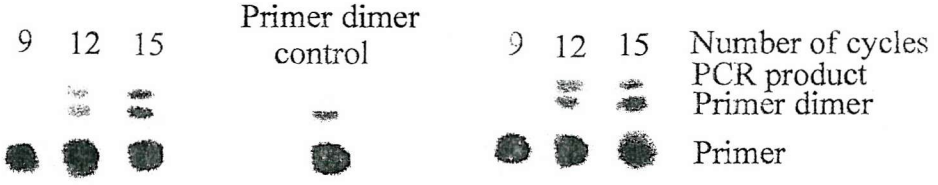


Figure 3.11 A 2% agarose gel showing the primer dimer and how the control served to detect this unwanted product.

It was also thought that the formation of primer dimers might be due to the accumulation of primers which are present when the DNA from one round of REPSA was utilized for the next round carrying some primer with it. Therefore, different techniques

were used to separate the product from the primers, and any primer dimers.

3.7.1 Agarose Trapping method

In this method a well was made in the middle of the agarose gel in the trajectory of the DNA sample and was filled with buffer. The migration of the DNA band was monitored and once the desired product reached the well it was extracted by removing the buffer with the DNA in it. This was then ethanol precipitated to remove impurities . This method was not very successful, and involved isolation of small quantities of DNA from relatively large volumes.

3.7.2 QIAEX II clean up kit

The details of the procedure are described in section 2.4.1. We first tested the efficiency of this clean up kit because the templates used in this work are only 54 base pairs long and we were concerned that the DNA might be lost during purification.

The PCR sample (100 µl) was concentrated in a speedvac until the volume was about 15 µl. It was then loaded onto a 2% agarose gel stained with ethidium bromide which was run for 30 min., to allow separation of PCR product from the primers or primer dimers. The desired PCR product was excised from the agarose gel and was purified following the QIAEX II procedure (see section 2.4.1). After purification the DNA was redissolved in half of the initial volume (50 µl), assuming that 50 % of the DNA was lost during clean up to yield approximately the same final DNA concentration as it was in the original PCR sample. This assumption was based on the manufacturer's claim that QIAEX II has a 50% efficiency for fragments which are less than 50 bases long. REPSA (section 2.2.2.1) was emulated by mixing 5 µl of the purified DNA template with 6 µl of water, which substitutes for the added drug and enzyme, and 1 µl of NEBuffer. Then the sample was further diluted with 2 µl of 0.5M EDTA and 90µl of water, which are added after the enzyme digestion in a normal REPSA experiment (section 2.5.3.2). 10 µl of this sample was added to a PCR tube containing the usual PCR mix (section 2.5.3.3) and was

placed in the PCR which was set for an annealing and extension temperature of 50°C. The sample was amplified for 9, 12, 15, 18, 21, and 24 cycles, 10 µl of sample were collected at each stage. Finally 5 µl of each sample were loaded in a 2% agarose gel stained with ethidium bromide. The results obtained are shown in Fig 3.12.

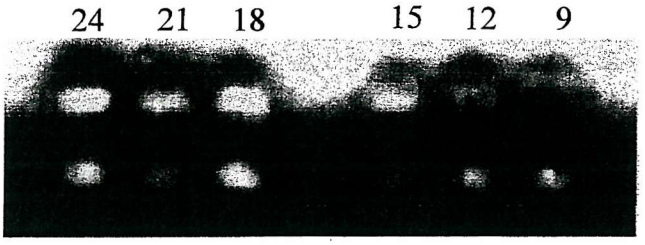


Figure 3.12 2% agarose gel showing the bands after using the QIAEX II clean up kit for 1 round of REPSA

It can be seen that the PCR product was detected after 9 cycles. Thus, QIAEX II was an excellent procedure and was used periodically to purify the DNA when primer dimer formation was noted.

3.7.3 Crush and Soak method

For the procedure see section 2.4.2 in Materials and Methods. This method was also tested and it was successful in extracting the DNA from the agarose gel. Also it is a cheaper method than QIAEX II and it was used at every round of REPSA.

3.8 How many rounds of REPSA?

This will depend on the number of random bases that the template contains, since the higher the number of bases the greater number of combinations that are present. If the template has a short random region then selection is accomplished within a short number of rounds. On the other hand, a long random region provides a higher number of different sequences and more rounds of REPSA will be needed. In this work for all the templates that contain 5 random bases 8 rounds of REPSA were performed. However, for the 1-17 random template the rounds of REPSA were increased to 12.

3.9 The perfected REPSA cycle

As a result of the investigations described above the following procedure was adopted for REPSA.

The template was made dsDNA as explained in section 2.2.1

3.9.1 First round of REPSA

1) Protection: 5 µl of dsDNA template (10 ng/µl) , 2 µl of drug dissolved in Tris-Na (10 mM Tris-HCl pH 7.5 and 10 mM NaCl) to the desired concentration, was added to a mixture which is suitable for *Bsg*I cleavage containing 1 µl of 10x NEBuffer 4 , (500 mM potassium acetate, 200 mM Tris acetate, 100 mM magnesium acetate, 10 mM DTT pH 7.9 @ 25°C) 0.8 µl SAM (final concentration 80 µM S-adenosylmethionine) and 3 µl H_2O, were mixed and incubated at room temperature for 30 min.

2) Selection: 1 µl of *Bsg*I (3000 u/ml) was added to the previous sample and incubated at 37°C for 30 min. The reaction was stopped by the addition of 2 µl of 0.5 M EDTA and diluted with 90 µl of water to yield a final dsDNA concentration of approximately 0.5 ng/µl. Then the sample was phenol extracted and ethanol precipitated to remove all the drug and enzyme and then redissolved in 100 µl of water before the amplification step.

3) Amplification: 10 µl of the previous sample was added to the standard PCR mix (section 2.5.3.3). Hot start PCR was favoured because it decreases the formation of primer dimers. The cycle profile was 94°C for 1 min. for denaturing and 60°C for 4 min. for annealing and extension. The product was collected after 9, 12, and 15 cycles.

4) DNA isolation: The selected product was loaded on a 3% agarose gel and was extracted and purified by the Crush and Soak method (section 2.4.2) before proceeding to the next round of REPSA.

3.10 Radioactive REPSA

In addition to the preferred REPSA cycle a radioactive method was also employed.

3.10.1 Template Formation

The right primer was radio-labelled at its 5'-end with [γ]^{32} P-ATP using polynucleotide kinase. This labelled primer was used for preparing the duplex template by polymerase extension with a PCR that had the following settings: 94 °C for 2 min and 64 °C for 10 min. Only one extension was performed. The sample was then mixed with 20 µl of 20 % ficoll loading dye and was loaded onto a 9 % non-denaturing polyacrylamide gel. The band containing the labelled template was extracted from the gel and was eluted overnight with 200 to 300 µl of Tris-EDTA (10mM Tris-HCl, 0.1 mM EDTA, pH 8.0)

3.10.2 First Round of radioactive REPSA

1) Protection: The radio-labelled template from above was ethanol precipitated and redissolved in 18 µl of Tris-EDTA (10 mM Tris-HCl , 0.1 mM EDTA, pH8.0). Then 2 µl of ligand was added.

2) Selection:: After equilibration for 30 min 2.5 µl of 10x NEBuffer 4 , (500 mM potassium acetate, 200 mM Tris acetate, 100 mM magnesium acetate, 10 mM DTT pH 7.9 @ 25 °C) 2.5 µl SAM (final concentration 80 µM S-adenosylmethionine) and 1 µl *Bsg*I were added and the cleavage reaction was incubated at 37 °C for 30 min.

3) Amplification: 10 µl of the sample from above was added to the radioactive PCR mix (section 2.5.3.3). Hot start PCR was still favoured because it decreases the formation of primer dimers. The cycle profile was 94 °C for 30 sec. for denaturing and 64 °C for 3 min. for annealing and extension. The number of cycles was increased to 15 in order to ensure that all the labelled right primer was incorporated thereby achieving maximum specific activity. At the end of these 15 cycles an excess of unlabelled right primer was

added and two more cycles were performed. This removes any duplexes which contain self-annealed PCR products generated during the reaction incorporating small quantities of radiolabelled primers.

4) DNA isolation: After the amplification 20 µl of 20 % ficoll was added to the sample and then it was loaded onto an 8 % non-denaturing polyacrylamide gel. The band containing the labelled sample was extracted from the gel and was eluted overnight with 200 to 300 µl of Tris-EDTA (10mM Tris-HCl, 0.1 mM EDTA, pH 8.0).

CHAPTER 4

Drug binding sites determined by REPSA

4.1 Introduction

The previous chapter explored several of the parameters which were varied during development of the REPSA technique. In this chapter I present the results that have been obtained under these conditions. In each case, since the primary sequence specificity of each of the ligands has been well documented, it is possible to assess which experiments have successfully selected for good ligand binding sites.

4.2 Distamycin with the 1-17 random template

Distamycin (1 μM) was tested with the 1-17 random template and the following sequences (Fig 4.1) were obtained. It can be seen that most of these sequences contain at least one $(AT)_4$ tract, and that several contain longer (AT) tracts.

1)	5'-GTTGTAGAAG**AAAAT**GT-3'	13)	5'-C**TAATT**GGGC**TTTATTAA**-3'
2)	5'-**TATTA**CTGGGGAAGTAG-3'	14)	5'-**TAAA**GACGGGCTG**ATA**GT-3'
3)	5'-GGG**AAAA**CAG**TTTATA**G-3'	15)	5'-G**AA**G**TAA**GGC**AA**G**TATA**-3'
4)	5'-**TTA**C**AA**GG**TA**G**TA**G**AA**G-3'	16)	5'-**AAAA**GAG**AA**GGGAGACGT-3'
5)	5'-**TATAAA**GTGGTGGAGTG-3'	17)	5'-**TA**G**TT**CGAGGTG**TAA**CA-3'
6)	5'-GTTAGTACGAAG**TAAA**C-3'	18)	5'-G**TAATAA**G**AATGTA**GTG-3'
7)	5'-GAAGTACGTAGG**AAAA**G-3'	19)	**TAT**GAG**AA**GTGGCAGGCG**TTAATA**-3
8)	5'-G**AAAA**GGG**ATAAAA**CAG-3'	20)	5'-**TA**GAGAGAGAGGG**TA**GT-3'
9)	5'-GTG**TAATAT**GGTAGTAG-3'	21)	5'-**AA**G**AAA**CC**ATTA**GTGCA-3'
10)	5'-**ATAAAT**GG**TA**GG**AT**C-3'	22)	5'-CAGGAGGGAGGG**AAA**GA-3'
11)	5'-**ATA**G**TA**GGGACG**AAATA**-3'	23)	5'-**TTTAAAA**CGG**ATGTA**GT-3'
12)	5'-GGG**AA**GGCTGGG**AATT**GTG-3'	24)	5'-G**TATA**GGGAG**ATGAA**GAG-3'

Figure 4.1 Data obtained after 12 rounds of REPSA for the 1-17 DNA template and 1 μM distamycin.. The bold and underlined bases represent potential sites for drug binding. The sequences represent the random region within the template.

From this figure it is clear that most of the sequences contain long runs of As and Ts

except for sequences 4, 17, 20 and 22 that are more like random sequences in which the As and Ts are separated by C and G residues. These data can be compared with the best distamycin binding sites which were determined by circular dichroism (Chen and Sha, 1998) and by footprinting (Abu-Daya *et al.*, 1995). The footprinting experiment ranking order was AAAA = AATT > ATTA / TAAT = TATA > ATAT, and similar binding affinities were found using circular dichroism (Chen and Sha, 1998). Both methods agree that AAAA is among the best binding sites and we see it present in sequences 1, 3, 7, 8, 16 and 23. On the other hand the worst binding sites are known to contain TpA steps (ATAT and AATA) and these are only present on sequences 9 and 19. It is also possible that some of these sequences could represent sites at which distamycin is binding in a 2:1 fashion (all sequences that contain runs of A's and T's which are of five bases or more in length) such as sequences 1, 2, 3, 5, 8, 9, 10, 11, 13, 18, 19, and 23. These data suggest that most of these sequences are genuinely selected by distamycin and that REPSA is a viable technique for determining binding sites of DNA ligands.

4.3 Hoechst 33258 with the 1-17 random template

Hoechst 33258 (1 μM) also yielded some positive clones with the 1-17 random template, which are shown in Fig 4.2.

1) 5'-TGGGGAGTAGGGGAG**AATAAT**-3'
2) 5'-AGAAAGGTGACAC**TATTA**-3'
3) 5'-GG**TA**CG**AA**GAGCTGTG**TA**-3'
4) 5'-GCCGG**TAA**CTGCGAGC**TA**-3'
5) 5'-GAG**AAATA**CAGGGTG-3'
6) 5'-GGGG**AAAAAAA**GAACGT-3'

Figure 4.2 Data obtained after 12 rounds of REPSA for the 1-17 DNA template and 1 μM Hoechst 33258. The bold and underlined bases represent potential sites for drug binding. The sequences represent the random region within the 1-17 random template.

The number of sequences is not sufficient to make any firm conclusions. However, we can still compare these results with those obtained by other methods. Footprinting

studies with DNase I have shown that the sequence selectivity of this agent is not equal for all AT- rich regions (Abu-Daya *et al.*, 1995). The ranking order followed is AATT > AAAA > ATTA/TAAT = ATAT > TATA. In addition the ranking was also obtained by a combinatorial method which involved the use of a microchip and showed that the best sites were AATT > AAAT > AAAA > AATA > TAAT > ATAT > TAAA > ATAA > TTAA > TATA. Our results do not contain an AATT site. However we have AAAT (sequence 5), AAAA (sequence 6), AATA (sequence 5), TAAT (sequence 1), ATAA (sequence 1). Within our sequences we do not have any TTAA or TATA represented. Two sequences (3 and 4) do not contain any good binding sites, indicating that probably these two sequences (3 and 4) were not selected by this agent.

4.4 [*N*-MeCys^3,*N*-MeCys^7]TANDEM with the 1-17 random template

[*N*-MeCys^3,*N*-MeCys^7]TANDEM (20 μM) also yielded positive results with this template, which are represented in Fig 4.3.

1) 5'-**ATA**GGGGGG**TAT**GGGAG-3'
2) 5'-G**TATATAT**GGTCCAGGG-3'
3) 5'-G**ATATATAAA**GTGGGAG-3'
4) 5'-G**ATA**G**ATATA**GAG**ATAA**-3'
5) 5'-ACAAGC**AATAAATA**GGT-3'
6) 5'-C**ATATA**CAAG**AATA**GAG-3'
7) 5'-AG**ATA**GG**ATA**GCAAGGG-3'
8) 5'-**ATA**GGGG**ATAAA**CAGTG-3'
9) 5'-GG**TAT**G**ATA**G**TTA**GGGG-3'
10) 5'-**ATAATA**GTGAGATCTGA-3'
11) 5'-GTAAG**TAAA**C**ATATAA**G-3'
12) 5'-G**TAT**G**AAT**G**ATA**G**TA**GC-3'
13) 5'-**ATA**GGGGCCG**TAA**GAGT-3'
14) 5'-G**ATA**G**AAT**G**ATA**G**TA**GC-3'
15) 5'-**AATATA**G**TA**GGG**AATA**G-3 '
16) 5'-**ATAAA**GCAG**TAA**GC**ATA**-3'
17) 5'-GAGTG**TA**G**TATATA**GGG-3'
18) 5'-ATGTG**TATA**GGTTGGAG-3
19) 5'-CCAGATAGCG**ATATT**GA-3'
20) 5'-**TATATA**CGGGGGGGGCA-3'
21) 5'-GATGTG**TATT**GTGGGGGG-3'
22) 5'-AGG**ATAT**GGGAG**ATAA**G-3'

Figure 4.3 Data obtained after 12 rounds of REPSA for the 1-17 DNA template and 20 μM [*N*-MeCys^3,*N*-MeCys^7]TANDEM. The bold and underlined bases represent potential sites for drug binding. The sequences represent the random region within the template.

It can be seen that, as expected, all these sequences contain TpA steps which have

been shown to be required for [*N*-MeCys^3,*N*-MeCys^7]TANDEM binding. In addition most of these TpA steps are surrounded by other A's or T's. Sixteen of the 22 sequences contain alternating A/T tracts of four bases and longer (2, 3, 4, 5,6,10,11, 15,16,17,18,19,20,21, and 22).

The large bias toward A/T binding sites can be seen in that these 22 clones contain 57 TpA steps and 50 ApT steps, but only 9 GpC and 3 CpG. It should also be noted that, even though the best binding site may be TpA, simply counting the abundance of each dinucleotide step may not necessarily give the expected answer. For example if the best tetranucleotide site is ATAT, this has twice as many ApT steps as TpA. The only sequence containing an isolated TATA is clone 18, though we might argue that this represents ligand binding to GTAT or ATAG . The tetranucleotide TTAA, which was identified as one of the weakest (A/T)TA(A/T) binding sites (Fletcher and Fox, 1995) is not present in any of these sequences. Although several studies have suggested that [*N*-MeCys^3,*N*-MeCys^7]TANDEM binds to TpA only when this is flanked by A and T residues, many of the clones contain (A/T)TA(G/C) sites. The binding of [*N*-MeCys^3,*N*-MeCys^7]TANDEM to TpA steps which are flanked by G and C residues is consistent with the footprinting data presented in chapter 5.

In addition it is worth noting that the A/T content of the clones obtained with distamycin is different from those with [*N*-MeCys^3,*N*-MeCys^7]TANDEM. None of the distamycin clones contain alternating AT-tracts while none of the [*N*-MeCys^3,*N*-MeCys^7]TANDEM clones contain AAAA.

These data demonstrate that REPSA can be successfully employed for investigating fragments which contain good binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM.

4.5 Echinomycin with the 1-17 random template

Similar experiments were performed with echinomycin (10 μM) on the 1-17 random template. The sequences of 12 clones obtained are shown in Fig 4.4

1) 5'-GGTAAGGGGACTGTGGT-3'
2) 5'-AATGAGGATAGAGAGGG-3'
3) 5'-TGGGAGTAAAGAGAGGG-3'
4) 5'-**CG**CAACAAACTAGAAGG-3'
5) 5'-C**CG**AGAGATAGCAGGGC-3'
6) 5'-ACAA**CG**G**CG**AGGAGGGA-3'
7) 5'-ATAAGGGGGGTATTGG-3'
8) 5'-AGTGCCAGGAG**CG**AGTC-3'
9) 5'-GGAGGGAATA**CG**AGACA-3'
10) 5'-GAGAG**CG**GGGGGAGGAG-3'
11) 5'-GAGAAG**CG**ATGGGGCAGC-3'
12) 5'-TAGACAGGGGGGAGGAG-3'

Figure 4.4 Data obtained after 12 rounds of REPSA for the 1-17 DNA template and10μM echinomycin.. The bold and underlined bases represent potential sites for drug binding. The sequences represent the random region within the template.

In contrast to the results with [*N*-MeCys^3,*N*-MeCys^7]TANDEM these clones appear to contain very few echinomycin binding sites (CpG). Only seven sequences out of twelve contain a CpG step. It is worth comparing these results with those obtained with the multisite fragment (chapter 5) in which the best binding sites for echinomycin are shown to be ACGT, ACGG and ACGA and the worst were CCGG and GCGG. Sequence 6 contains ACGG and sequence 9 contains ACGA, which are good binding sites. However, sequence 10 contains one of the worst binding sites GCGG. These results suggest that REPSA has not selected for good echinomycin binding sites under these conditions. The reasons for this result are not clear.

4.6 Other random templates

The previous sections have shown that REPSA can be used with a long random template to isolate binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM and distamycin (but not echinomycin). Since we are interested in determining the interactions of ligands with single binding sites we performed similar experiments with templates containing only five random bases.

4.6.1 Distamycin with the 13-17 random template

As there was a small number of colonies on the agar plates after transformation, only a few sequences were obtained for this template with 1 μM distamycin . Unfortunately none of these fragments showed any selectivity for this ligand. The random sequences obtained for this template are presented in Fig 4.5.

1) 5'-AGTAC-3'
2) 5'-GCAGG-3'
3) 5'-AAGGC-3'

Figure 4.5 Data obtained after 8 rounds of REPSA for the 13-17 DNA template with 1 μM distamycin. The sequences represent the random region within the template.

These data are not as good as they were for the 1-17 random template and do not contain any distamycin binding site. This may be because the random sequence is located within the cleavage site of the restriction enzyme. Examination of the sequences obtained with the 1-17 random template also reveals that most of the good binding sites for this agent are located towards the middle and to the 5' end of the sequence and not to the 3' end.

4.6.2 [*N*-MeCys^3,*N*-MeCys^7]TANDEM with the 13-17 random template

Similar experiments were performed with [*N*-MeCys^3,*N*-MeCys^7]TANDEM (20 μM) and the 13-17 random template. The results are shown in Figure 4.6

1) 5'-**TA**ACC-3' 6) 5'-GCAAA-3'
2) 5'-AGCGT-3' 7) 5'-CG**TA**C-3'
3) 5'-GATGC-3' 8) 5'-TGGGG-3'
4) 5'-AAGAA-3' 9) 5'-CAAGT-3'
5) 5'-GACTC-3'

Figure 4.6 Data obtained after 8 rounds of REPSA for the 13-17 DNA template with 20 μM [*N*-MeCys^3,*N*-MeCys^7]TANDEM. The bold and underlined bases represent potential sites for drug binding. The sequences represent the random region within the template.

As with distamycin it appears that this template has not yielded positive results for [*N*-MeCys^3,*N*-MeCys^7]TANDEM, since very few of these sequences contain TpA steps.

4.6.3 Echinomycin with the 13-17 random template

We also examined the selectivity of echinomycin (10 μM) using the 13-17 random

template, even though this ligand did not produce successful results with the 1-17 random template. The results for 14 clones are presented in Figure 4.7

1) 5'-TC A**CG**GG G-3'	8) 5'-T**C G**GGAG G-3'
2) 5'-T**C G**G**CG**G G-3'	9) 5'-T**C G**GC**CG** G-3'
3) 5'-T**C G**AGC**C G**-3'	10) 5'-TC **CG**AGG G-3'
4) 5'-T**C G**G**CG**A G-3'	11) 5'-TC **CG**GGG G-3'
5) 5'-TC **CG**AGG G-3'	12) 5'-T**C G**GGGG G-3'
6) 5'-T**C GCG**GA G-3'	13) 5'-T**C G**AGGG G-3'
7) 5'-T**C G**GGCA G-3'	14) 5'-T**C G**GGGG G-3'

Figure 4.7 Data obtained after 8 rounds of REPSA for the 13-17 DNA template with 10 μM echinomycin. The bold and underlined bases represent potential sites for drug binding. The sequences represent the random region within the template. The first TC and the last G belong to the fixed regions located at either side of the template. The five random bases are located in the middle.

In contrast to the results with the 1-17 random template it can be seen that all these sequences contain at least one CpG step. It should, however be noted that in most instances the CpG is located at the 5'-end of the random region using the terminal C of the constant region. One of the best echinomycin binding sites as determined by multisite (chapter 5) ACGG is found in sequence 1. Several of the clones contain the sequence TCGG which is an intermediate to high affinity binding site (2, 4, 7, 8, 9, 12 and 14). In addition the low affinity sites such as CCGG and GCGG are present in 9 and 11, and 2 and 6 respectively. It should also be noted that several of the sites contain G_n tracts (11- 14), possibly suggesting that these sequences may be refractory to *Bsg* I cleavage on account of the unusual structure adopted by the DNA.

4.6.4 [*N*-MeCys^3,*N*-MeCys^7]TANDEM with the 9-13 random template

This template was only used with [*N*-MeCys^3,*N*-MeCys^7]TANDEM and the clones obtained are shown in Fig 4.8.

These clones yielded sequences that contain many mutations and deletions within the fixed region of the template. In addition the random region does not contain many good

binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Only clone 1 contains a TA tract TATAA which has been shifted due to the deletions of three bases within the fixed region so that its position is 6 bases away from the enzyme binding site, thus simulating the 6-10 template.

```
                                                    ↓
     5'-GATCCAGAGTGCAGGCACTCAG  NNNNN    GCTGGACCTGAGAGGATC-3'
1)   5'-GATCCAGAGTGCAGG□A□T□AG  TATAA GGGAGGGACCTGAGAGGATC-3'
2)   5'-GATCCAGAGTGCAG□□A□T□AG  TCTAGAAGTGACGACCTGAGAGGATC-3'
3)   5'-GATCCAGAGTGCAGGCACTCAG  CGAGC    GCTGGACCTGAGAGGATC-3'
4)   5'-GATCCAGAGTGCAGG□ACAGAA  CGATC   GACTGGACCTGAGAGGATC-3'
5)   5'-GATCCAGAGTGCAGGCACTCAG  AGGGC    GCTGGACCTGAGAGGATC-3'
6)   5'-GATCCAGAGTGCAGGCACTCAG  CGGAA    GCTGGACCTGAGAGGATC-3'
```

Figure 4.8 Data obtained after 8 rounds of REPSA for the 9-13 DNA template with 20 μM [*N*-MeCys^3,*N*-MeCys^7]TANDEM. The bold and underlined bases represent potential sites for drug binding. The bases in bold and italics represent mutations that were found to have occurred in these clones. The bold bases are the enzyme recognition site and the down arrow indicates the enzyme cleavage site. The □ represents the bases that are deleted.

4.6.5 [*N*-MeCys^3,*N*-MeCys^7]TANDEM with the 6-10 random template

The clones obtained with this template are shown in Fig 4.9 Most of these sequences do not contain the expected binding site for [*N*-MeCys^3,*N*-MeCys^7]TANDEM.

```
1)   5'-G GCATA T-3'          9)   5'-G GGACG T-3'
2)   5'-G GGGGG T-3'          10)  5'-G GAAAA T-3'
3)   5'-G GGGGG T-3'          11)  5'-G GGAGC T-3'
4)   5'-G TATGC T-3'          12)  5'-G TCGAT T-3'
5)   5'-G GGAAA T-3'          13)  5'-G GTTGC T-3'
6)   5'-G TAGTA T-3'          14)  5'-G GGGGG T-3'
7)   5'-G GTTGG T-3'          15)  5'-G AGATG T-3'
8)   5'-G GGAGG T-3'          16)  5'-G TAAGG T-3'
```

Figure 4.9 Data obtained after 8 rounds of REPSA for the 6-10 DNA template with 20 μM [*N*-MeCys^3,*N*-MeCys^7]TANDEM The bold and underlined bases represent potential sites for drug binding. The sequences represent the random region within the template. The first G and the last T belong to the fixed regions located at either side of the template. The five random bases are located in the middle.

This template yielded many clones which have a high GC content and the sequence GGGGG appears in 3/16 clones. It therefore seems that [*N*-MeCys^3,*N*-MeCys^7]TANDEM was not selecting most of these sequences and that maybe they were selected because they were resistant to digestion by *Bsg*I (section 3.4.2), due to the formation of unusual structures within the random region. We considered this possibility and tested some of these sequences for *Bsg*I cleavage as described in section 3.4.2. The result was that *Bsg*I could cleave all the templates, though at different rates. These sequences also contain some good [*N*-MeCys^3,*N*-MeCys^7]TANDEM binding sites such as clone 1 with ATAT, and the intermediate binding sites such as clone 4 and 6 with GTAT . The worst binding sites were clone 6 GTAG and clone 16 GTAA.

The 6-10 template has yielded 4/16 clones that could be possible binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM. These results are more positive than those found for the13-17 templates and the 9-13 template. Nonetheless these sites are found with low frequency and are not enriched above the number that would be expected randomly.

4.6.6 [*N*-MeCys^3,*N*-MeCys^7]TANDEM with the 2-6 random template

This template is the last one containing a five random region, which is located two bases away from the enzyme recognition site. The results obtained with this template are presented in Fig 4.10.

This template gave many clones with mutations and deletions. However, 7/16 clones contain potential binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM within the random region. The most frequent binding site within these clones is site ATAG which appears in clones 1, 3, 10, 14 and 16. This site is shown to have an intermediate binding affinity (section 5.4), with a C_{50} of 17 μM (Table 5.2). The other potential [*N*-MeCys^3,*N*-MeCys^7]TANDEM sites present in these clones are site GTAA (clone 4) and CTAG (clone 13). The binding affinity of these sites decreases in the order CTAG > ATAG > GTAA . However, it should again be emphasized that these TpA sites do not appear to have been

	5'-GATCCAGAGTGCAGG	NNNNN	GCACTCGACTGGACCTGAGAGGATC-3'
1)	5'-GATCCAGAGTGCAGG	**ATA**GG	GCACTCGACTGGACCTGAGAGGATC-3'
2)	5'-GATCCAGAGTGCAGG	AGAAT	□C□□□CGACTGGACCTGAGAGGATC-3'
3)	5'-GATCCAGAGTGCAGG	C**ATA**G	GCACTCGACTGGACCTGAGAGGATC-3'
4)	5'-GATCCAGAGTGCAGG	**TAAA**G	GCACTCGACTGGACCTGAGAGGATC-3'
5)	5'-GATCCAGAGTGCAGC	TCAGG	□***CGGATAGTA***GGACCTGAGAGGATC-3'
6)	5'-GATCCAGAGTGCAGC	TCGAG	GCGGATAGTAGGACCTGAGAGGATC-3'
7)	5'-GATCCAGAGTGCAGG	ACAGT	***TGGG***TCGACTGGACCTGAGAGGATC-3'
8)	5'-GATCCAGAGTGCAGG	ACAGT	***ATGG***TCGACTGGACCTGAGAGGATC-3'
9)	5'-GATCCAGAGTGCAGG	ACAGG	***AATA***TCGACTGGACCTGAGAGGATC-3'
10)	5'-GATCCAGAGTGCAGG	**AATA**G	GCACTCGACTGGACCTGAGAGGATC-3'
11)	5'-GATCCAGAGTGCAGG	ACAGT	***AATG***TCGACTGGACCTGAGAGGATC-3'
12)	5'-GATCCAGAGTGCAGG	**A**CAGAA**G**	GCACTCGACTGGACCTGAGAGGATC-3'
13)	5'-GATCCAGAGTGCAGG	C**TA**GG	GCACTCGACTGGACCTGAGAGGATC-3'
14)	5'-GATCCAGAGTGCAGG	C**AATA**	GCACTCGACTGGACCTGAGAGGATC-3'
15)	5'-GATCCAGAGTGCAGA	CTGCT	***TGAAGGGGGAA***GACCTGAGAGGATC-3'
16)	5'-GATCCAGAGTGCAGG	GC**ATA**	GCACTCGACTGGACCTGAGAGGATC-3'

Figure 4.10 Data obtained after 8 rounds of REPSA for the2- 6 DNA template with 20 μM [*N*-MeCys^3,*N*-MeCys^7]TANDEM The bold and underlined bases represent potential sites for drug binding. The bold and italics bases represent mutations from the original sequence. The □ represents the bases that are deleted, the bold bases are insertions.

enriched beyond what would have been expected from a random distribution.

4.7 Footprinting of sequences obtained with [*N*-MeCys^3,*N*-MeCys^7]TANDEM and 1-17 random template

We tested the binding of [*N*-MeCys^3,*N*-MeCys^7]TANDEM to the sequences selected by REPSA from the 1-17 random template by performing footprinting experiments on DNA fragments containing these inserts. Fragments 1, 3, 5, 6, 19, 20, 21 and 22 were chosen for these experiments and their sequences are shown bellow:

```
1)   5'-GATCCTCTCAGGTC CTCCCATACCCCCCTAT CTGCACTCTGGATC-3'
3)   5'-GATCCTCTCAGGTC CTCCCACTTTATATATC CTGCACTCTGGATC-3'
5)   5'-GATCCAGAGTGCAG ACAAGCAATAAATAGGT GACCTGAGAG
6)      GATCCAGAGTGCAG CATATACAAGAATAGAG GACCTGAGAGGATC-3'
19)  5'-GATCCTCTCAGGTC TCAATATCGCTATCTGG CTGCACTCTGGATC
20)         CAGAGTGCAG TATATACGGGGGGGGCA GACCTGAGAGGATC-3
21)  5'-GATCCTCTCAGGTC CCCCCCACATACACATC CTGCACTCTGGATC-3'
22)  5'-GATCCTCTCAGGTC CTTATCTCCCATATCCT CTGCACTCTGGATC-3'
```

Since these sequences were all cloned into the *Bam*HI site of pUC18 plasmid polylinker, suitable footprinting substrates containing these inserts were prepared by cutting the plasmids with *Hin*dIII and *Sac*I and labelling at the 3'-end of the *Hin*dIII site with α-^{32}P-dATP. It should be noted that sequences 5 and 6, and 19 and 20 are each contained in the same fragments as part of multimeric inserts. DNase I footprinting experiments with these fragments are presented in Figures 4.11 - 4.13.

Sequence 1 contains two potential [*N*-MeCys^3,*N*-MeCys^7]TANDEM sites GTATG and GATAG. DNase I footprinting experiments with this fragment are presented in the left hand panel of Fig. 4.11. It can be seen that [*N*-MeCys^3,*N*-MeCys^7]TANDEM reduces the DNase I cleavage at both sites in a concentration dependent fashion, though the protection at GATAG (CTATC) appears to be greater than at GTATG (CATAC). The intensity of bands within these footprints was measured using a phosphorimager, and footprinting plots derived from these data are shown in Figure 4.14. C_{50} values were estimated from these plots by fitting the data to a

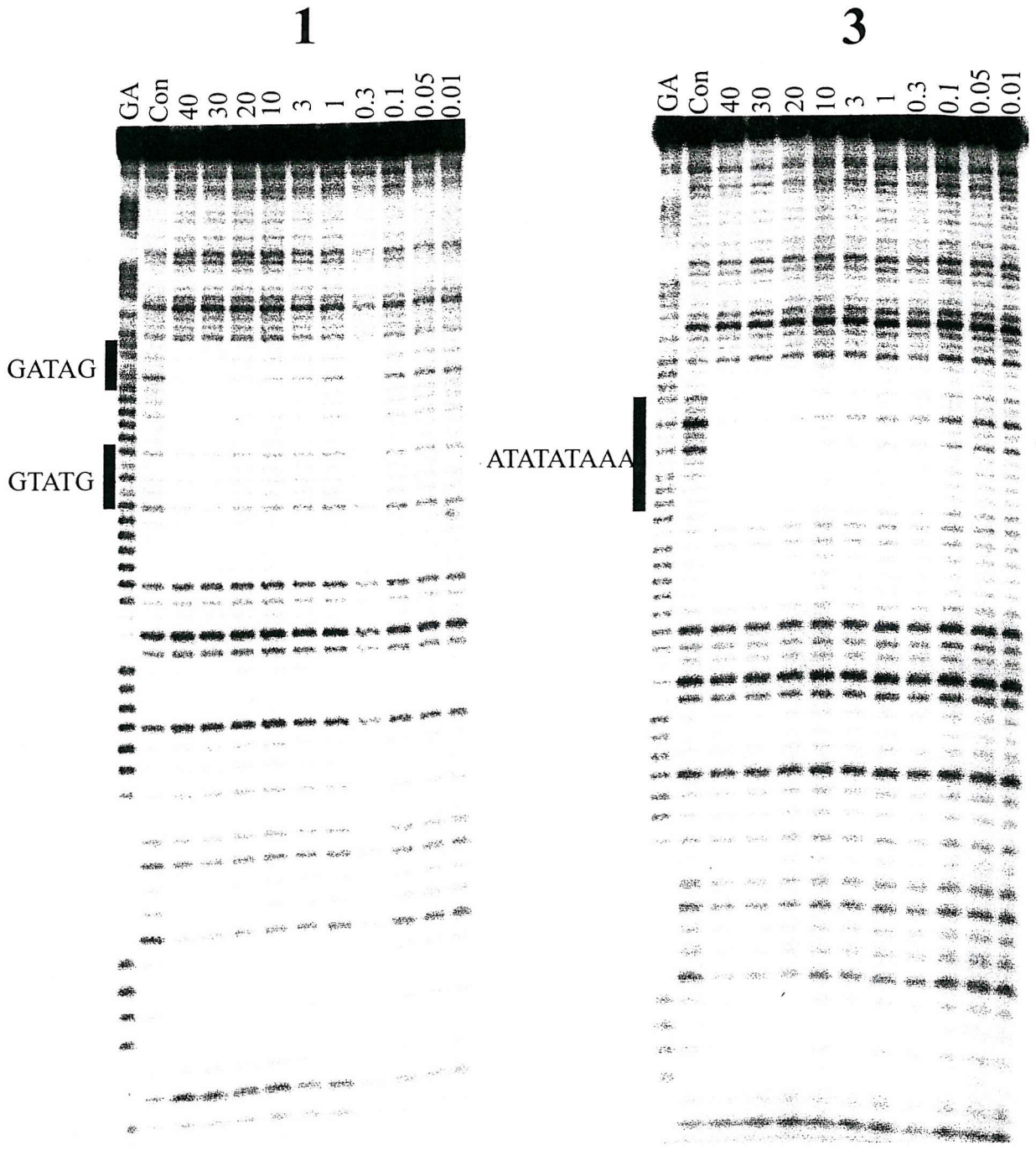


Figure 4.11 DNase I cleavage patterns of sequences 1 and 3 in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM . Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the potential target sites. The track labelled GA is a marker specific for guanine and adenine. The track labelled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

simple binding curve (section 2.24); these values are presented in Table 4.1. Although these footprints demonstrate that [*N*-MeCys^3,*N*-MeCys^7]TANDEM does bind to these sites, the C_{50} values are not as strong as might be expected for tight binding sites.

Sequence 3 has a longer alternating AT tract to which [*N*-MeCys^3,*N*-MeCys^7]TANDEM might bind, in the sequence ATATATAAA. The results of DNase I footprinting experiments with this fragment are shown in the right hand panel of Figure 4.11. It can be seen that [*N*-MeCys^3,*N*-MeCys^7]TANDEM produces a footprint around this AT-site, which is much clearer than those generated with sequence 1. The footprint is located in the $(AT)_n$ portion of this region, though it extends into the A-tract on account of the known 3'-stagger of the enzyme. A footprinting plot derived from these data is shown in Figure 4.14 and the estimated C_{50} value (0.017 ± 0.012) is presented in Table 4.1. This low value confirms that this is indeed a good binding site for [*N*-MeCys^3,*N*-MeCys^7]TANDEM. However, it can be seen that the binding curve does not represent a good fit to the data points. Indeed inspection of the autoradiograph reveals that, although the footprint decreases in a concentration dependent fashion and is much less pronounced at lower ligand concentrations, bands within this region have not returned to that in the control at the lowest concentration of [*N*-MeCys^3,*N*-MeCys^7]TANDEM (0.01 μM). Although the intensity of the bands in the footprint increases between 40 μM and 1 μM there is very little further change below 0.1 μM. This cannot be explained by suggesting variations in the extent of digestion since the intensity of other bands in the fragment, in regions outside the ligand binding site, is constant across all the lanes. It therefore seems that the interaction of [*N*-MeCys^3,*N*-MeCys^7]TANDEM with this site is not described by a simple binding curve. This effect, which may be the result of cooperative interactions between bound ligands, is considered further in the Discussion.

Sequences 5 and 6 are cloned into the same DNA fragment, which contains a multimeric DNA insert of REPSA sequences. DNase I footprinting experiments with this fragment are shown in the left hand panel of Fig. 4.12. The lower site (CTATTTATT) corresponds to sequence 5, while the upper two sites (CTATT and TATAT) are from sequence 6. It can be seen that [*N*-MeCys^3,*N*-MeCys^7]TANDEM produces footprints at each of these sites (as well as at

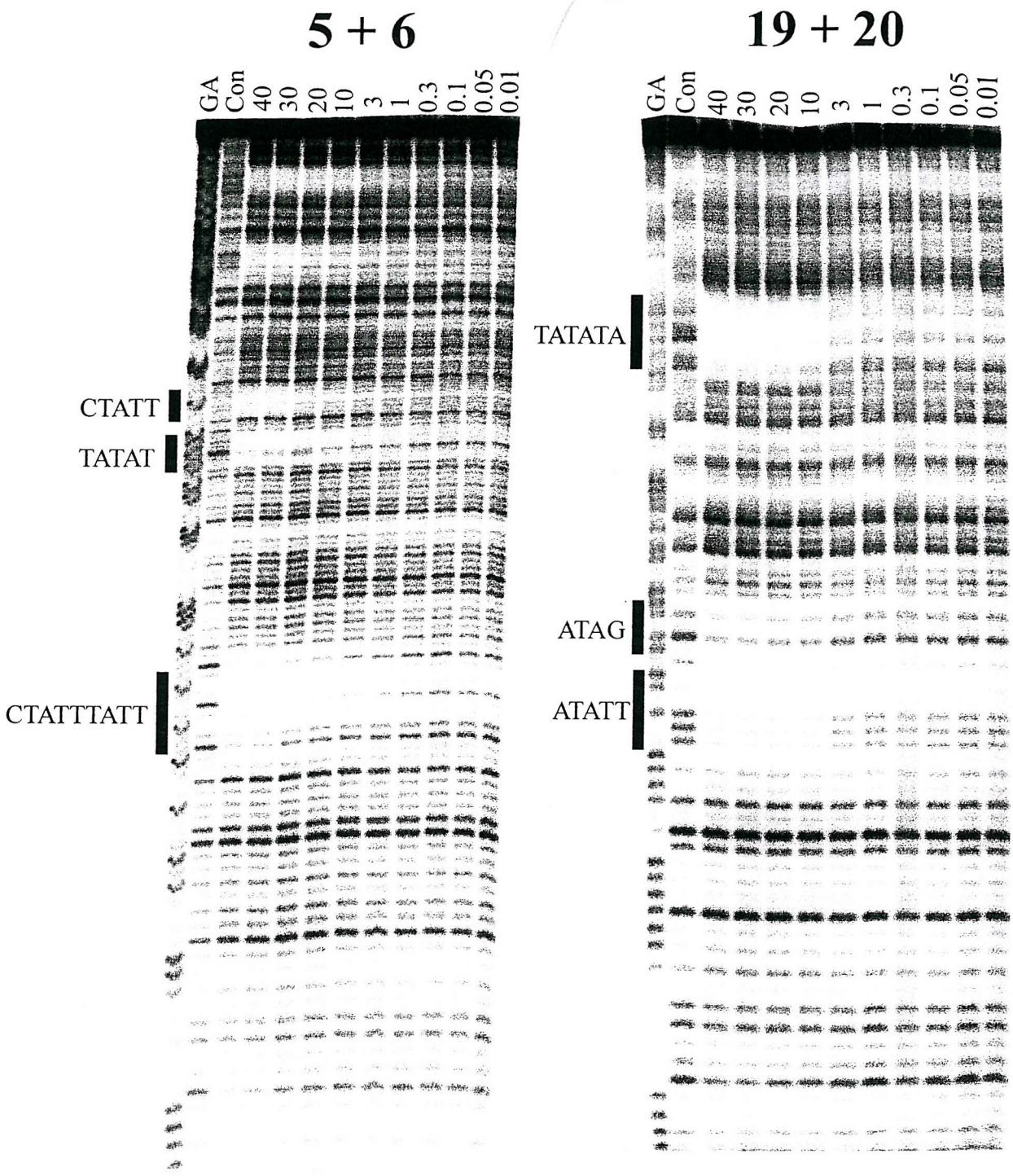


Figure 4.12 DNase I cleavage patterns of sequences 5 + 6 and 19 + 20 in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM . Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the potential target sites. The track labelled GA is a marker specific for guanine and adenine. The track labelled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end. There are other footprints towards the top of gel 5 + 6 because this fragment contained multiple REPSA sequences. However, only these two sequences were identified on the sequencing gel.

other unidentified sites higher up the gel in other parts of this multimeric insert). Visual inspection of these patterns suggests that the lowest target represents the best binding site. It should also be notice that the cleavage of bands in the sequences TATAT and CTATTTATT has not returned to that in the control by the lowest ligand concentration. Footprinting plots derived from these data are presented in Figure 4.14, and their C_{50} values are summarised in Table 4.1. Once again it can be seen that the curves are not good fits to the data points, especially for TATTTAT and TATAT.

Sequences 19 and 20 are cloned into the same DNA fragment, which contains a dimeric DNA insert of REPSA sequences. DNase I footprinting experiments with this fragment are shown in the right hand panel of Fig. 4.12. The lower sites (ATATT and ATAG) correspond to sequence 20, while the upper site (TATATA) is from sequence 19. It can be seen that [*N*-MeCys^3,*N*-MeCys^7]TANDEM produces footprints at all these sites. Visual inspection of the gel reveals that TATATA is the best of these three binding sites, while ATAG is much weaker; ATATT appears to have intermediate affinity for [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Footprinting plots derived from these data are shown in Figure 4.14 and the C_{50} values are summarised in Table 4.1. Once again it can be seen that the curves are not good fits to the data points, since the intensity of bands does not return to that in the control, even at the lowest ligand concentration.

Sequence 21 contains only one TpA step, which might represent a [*N*-MeCys^3,*N*-MeCys^7]TANDEM binding site, in the sequence GTATG. A DNase I footprinting experiment with this fragment is shown in the left hand panel of Fig. 4.13. It can be seen that, although the bands are attenuated at the highest ligand concentration, the ligand has not produced a clear footprint. A footprinting plot derived from this gel is shown in Fig. 4.14 and the estimated C_{50} value is shown in Table 4.1. In contrast to some of the results with other REPSA fragments, the intensity of bands has returned to that in the control at the lower ligand concentrations, and the simple binding curve seems to be a reasonable fit to the data points. This is clearly a weak [*N*-MeCys^3,*N*-MeCys^7]TANDEM binding site, suggesting that it may not have been selected in REPSA by virtue of its ability to bind [*N*-MeCys^3,*N*-MeCys^7]TANDEM.

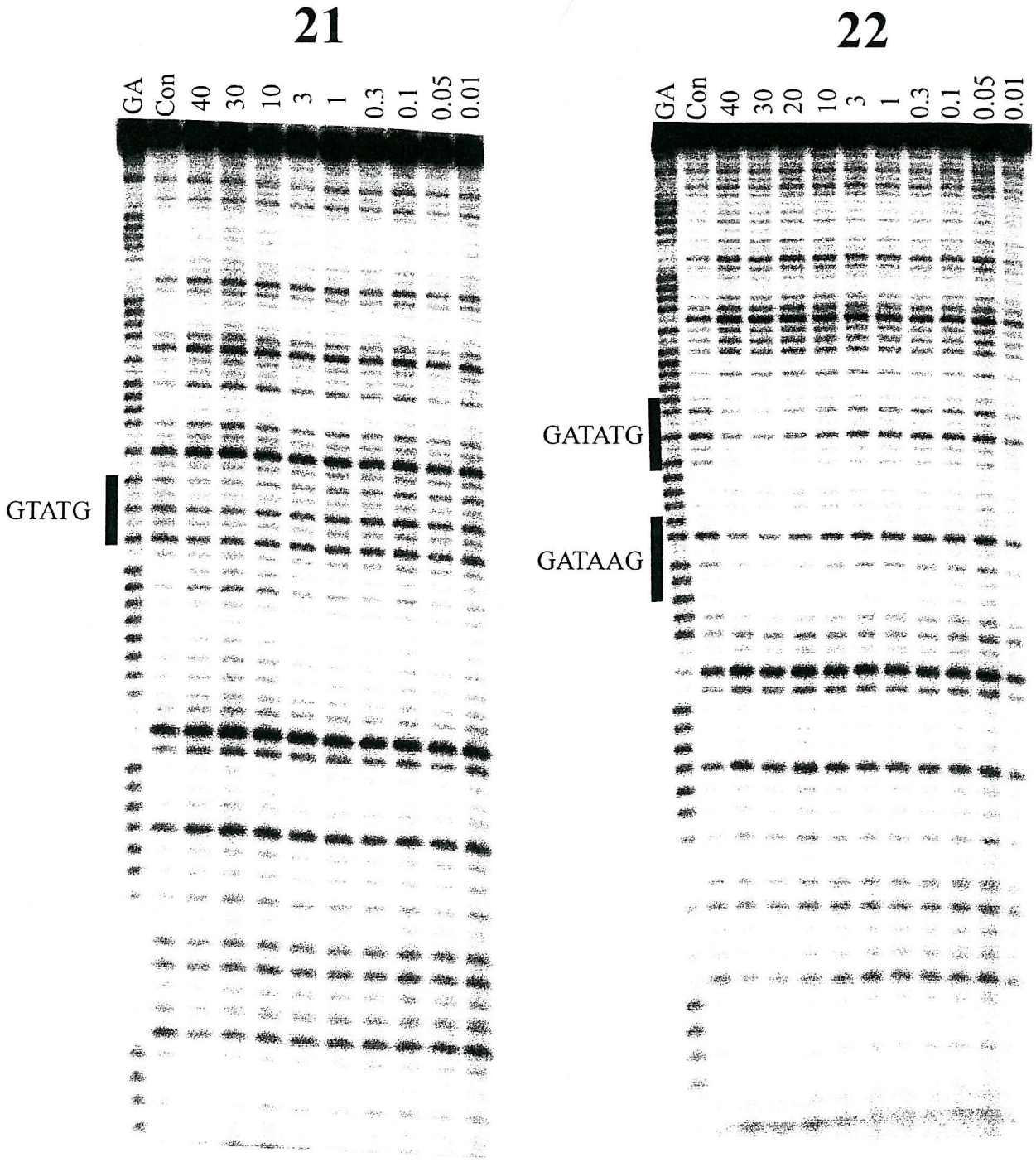


Figure 4.13 DNase I cleavage patterns of sequences 21 and 22 in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the potential target sites. The track labelled GA is a marker specific for guanine and adenine. The track labelled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

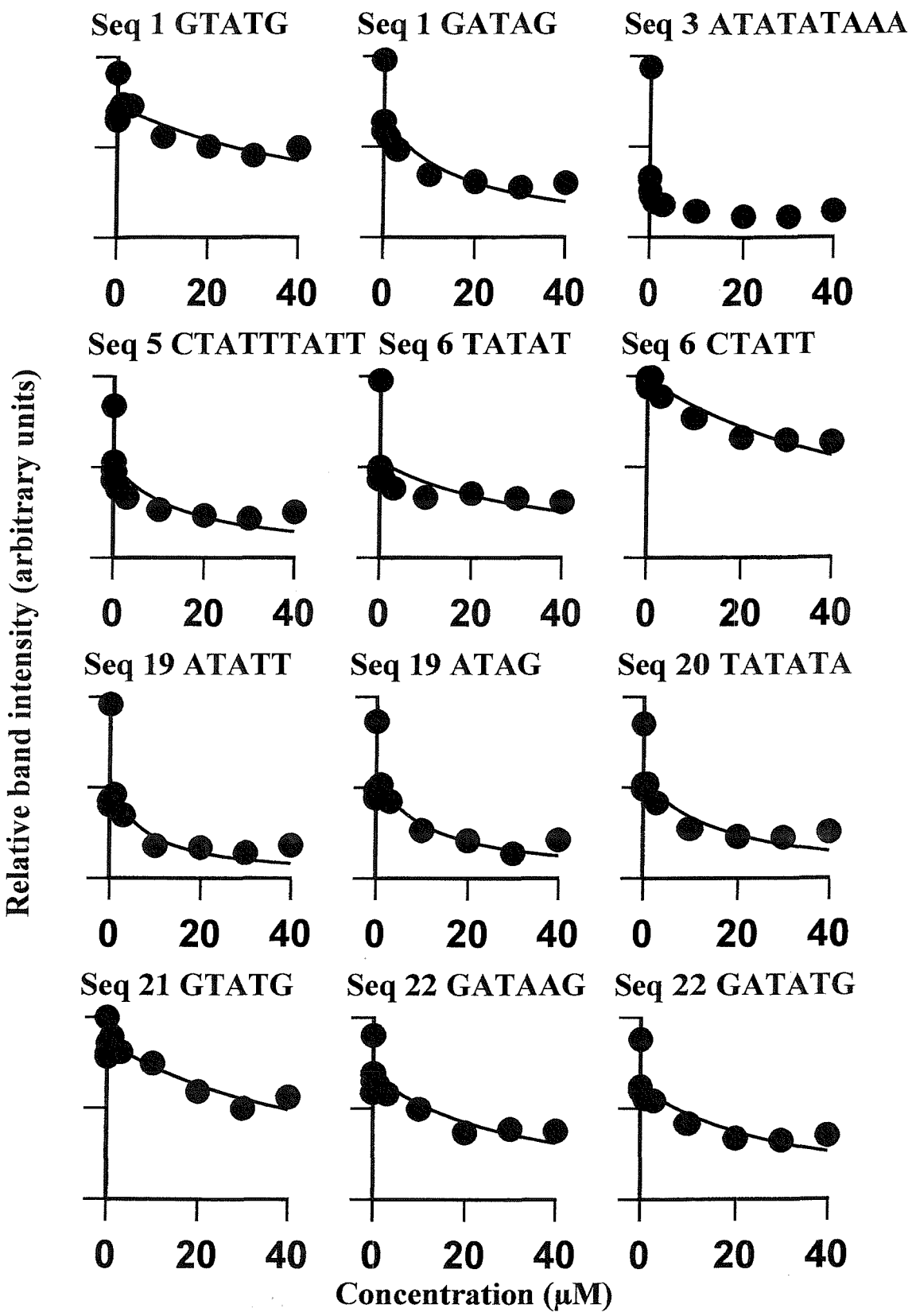


Figure 4.14 Footprinting plots showing the interaction of [*N*-MeCys^3,*N*-MeCys^7] TANDEM with the TpA sites of DNA fragments obtained with REPSA. Data was analysed as described in section 2.24.

Sequence 22 contains two potential [*N*-MeCys^3,*N*-MeCys^7]TANDEM binding sites (GATAAG and GATATG). The results of a DNase I footprinting experiment with this fragment are shown in the right hand panel of Fig. 4.13. It can be seen that [*N*-MeCys^3,*N*-MeCys^7]TANDEM causes attenuated cleavage at both sites at the highest concentrations, but neither site produces a clear footprint. Footprinting plots derived from these data are shown in Figure 4.14 and the C_{50} values are summarised in Table 4.1 In this instance the bands have returned to that in the control at low ligand concentrations.

Sequence #	Site	Sequence	C_{50}
1	1	GTATG	55 ± 18
	2	ATAG	16 ± 8
3	1	ATATATAAA	0.017 ± 0.012
5	1	CTATTTATT	15 ± 9
6	1	TATAT	36 ± 28
	2	CTATT	52 ± 8
19	1	ATATT	7 ± 5
	2	ATAG	11 ± 6
20	3	TATATA	14 ± 7
21	1	GTATG	52 ± 12
22	1	GATAAG	28 ± 10
	2	GATATG	32 ± 10

Table 4.1: C_{50} values (μM) determined for the interaction of [*N*-MeCys^3,*N*-MeCys^7]TANDEM with the various TpA sites on all the REPSA sequences. The data were derived from quantitative analysis (section 2.24) of the footprinting plots shown in Figure 4.14.

4.8 Discussion

4.8.1 REPSA

The results presented in the first part of this chapter suggest that REPSA can be successfully employed for isolating the preferred binding sites for the minor groove binding

ligand distamycin and the AT-selective bis-intercalator [*N*-MeCys^3,*N*-MeCys^7]TANDEM on the template with a 17 base pair random region. These results with distamycin confirm the results of Hardenbol & van Dyke (1997b), who also used a long random template to isolate long AT-tracts. Surprisingly we were unable to isolate good echinomycin binding sites using this method.

Evidence for the success of the technique is provided by some of the following observations. (i) Most of the sequences isolated with distamycin contained AT-tracts of 4 or more base pairs, in contrast to the (failed) selection with echinomycin, for which only one out of twelve clones contained such an AT-tract. (ii) Although both [*N*-MeCys^3,*N*-MeCys^7]TANDEM and distamycin produced AT-rich sequences, they have different arrangements. Thirteen out of 24 sequences with distamycin contained A_n tracts of three base pairs or longer. In contrast only 4 out of 22 sequences selected with [*N*-MeCys^3,*N*-MeCys^7]TANDEM contained such A_n tracts. (iii) Similarly 11 [*N*-MeCys^3,*N*-MeCys^7]TANDEM sequences contained the sequences ATAT or TATA while only five distamycin clones contained this sequence.

The failure of REPSA to isolate echinomycin binding sites from the 1-17 random template is difficult to explain. It may be significant that REPSA is most successful with AT-selective ligands.

The use of a long random sequence to identify binding sites for small ligands is, however, limited since it will inevitably isolate fragments that contain multiple, overlapping binding sites. We were therefore interested in using this technique with fragments containing shorter blocks of random bases in order to isolate unique, isolated drug binding sites. In the first series of experiments we placed a 5 base random region in the vicinity of the enzyme cleavage site, in the expectation that this would be the best region for achieving selection. The results for distamycin and [*N*-MeCys^3,*N*-MeCys^7]TANDEM with this fragment were disappointing, and we failed to enrich the sequences for canonical binding sites for these ligands. In contrast to the results with the random 17-mer, this template appeared to select some good echinomycin

binding sites, which were located at the 5'-edge of the random region. On a statistical basis we would expect only 25% of the echinomycin clones to contain a G as the first base, whereas 10/14 were found to begin with a G generating a 5'- CpG step. The significance of these results is even more apparent when these sequences are compared with those obtained with [*N*-MeCys^3,*N*-MeCys^7]TANDEM and distamycin on the 13-17 fragment, which produced a more random base distribution. It is not clear why REPSA appears to successfully isolate echinomycin binding sites using the short template, in contrast to the longer random region. However, it should be noted that, even though the long random template did not yield successful clones with echinomycin, the sequences generated were biased toward GC-residues (60%). This may reflect some feature of echinomycin binding, though it may also be due to the inability of the selecting enzyme to cleave some GC-rich sequences.

Examination of the selected sequences with distamycin (Fig. 4.1) suggested that more of the ligand binding sites were located at the 5'-end than the 3'-end of the random region. This, together with the observed location of the CpG echinomycin binding sites, prompted us to examine whether selection could be achieved when the binding sites were placed in different positions between the recognition and cutting sites for the enzyme. There is no *a priori* reason why binding to the cutting site alone should be sufficient or necessary to inhibit enzyme cleavage. Indeed, if the enzyme cuts DNA from the major groove, then the binding of a ligand in the opposing minor groove may have not inhibited the reaction. In this situation cleavage may be inhibited by drug binding to an intervening site, by either steric occlusion or by distorting the DNA structure. These other random templates were only tested with [*N*-MeCys^3,*N*-MeCys^7]TANDEM, since much less is known about its sequence recognition properties than the other ligands. These experiments were unsuccessful. This therefore suggests that REPSA is most successful for isolating multiple overlapping binding sites. It may be that, although these drugs reduce the efficiency of restriction enzyme cleavage, their kinetics are such that selection is only achieved at high levels of saturation in which many bound drug molecules may cooperate to inhibit the enzyme sufficiently. However, it may be that with the benefit of hindsight, [*N*-MeCys^3,*N*-MeCys^7]TANDEM may not have been the best ligand for these studies. Firstly, several studies together with those presented below, suggest that [*N*-

MeCys^3,*N*-MeCys^7]TANDEM binds to AT-containing DNA in a cooperative fashion { Lee & Waring, 1978b}, and that the binding to long alternating AT tracts is much stronger, with a slower dissociation rate, than to short AT-tracts. Secondly the dissociation rate for [*N*-MeCys^3,*N*-MeCys^7]TANDEM is relatively fast from all sequences (except poly(dA-dT)), which may obviate the selection process (Fletcher *et al.*, 1995). Further experiments with [*N*-MeCys^3,*N*-MeCys^7]TANDEM might employ templates with intermediate length random regions (6-8 bases).

Although REPSA appears to have successfully isolated fragments containing good binding sites for [*N*-MeCys^3,*N*-MeCys^7]TANDEM and distamycin, it has yet to be employed for determining the binding sites of a ligand with unknown sequence specificity.

4.8.2 Footprinting of the sequences selected with [*N*-MeCys^3,*N*-MeCys^7]TANDEM.

DNase I footprinting experiments were performed with several of the clones selected by REPSA, in order to confirm that they contained good [*N*-MeCys^3,*N*-MeCys^7]TANDEM binding sites. All these sites yielded DNase I footprints, though they displayed a range of binding affinities. These experiments clearly demonstrated that [*N*-MeCys^3,*N*-MeCys^7]TANDEM binds best to long AT-tracts, an observation which is pursued further in Chapter 6. In general the footprints were much clearer for the sites containing regions of alternating AT, than for clones containing other TpA steps. The lowest C_{50} value was obtained for clone 3, which contains the sequence ATATATAAA.

One interesting observation from these experiments is that in several cases the simple binding curve does not adequately describe the data points, and produces higher C_{50} values than expected with large errors. In general the weaker binding sites, which produce attenuated cleavage rather than full protection, are adequately described by the curves, while the stronger sites are not. It appears that something unusual is occurring with the strongest binding sites. An

unusual effect at the strong binding sites is also seen in the footprinting patterns themselves. In several instances the DNase I cleavage pattern does not return to that in the control even at very low ligand concentrations. This is not due to variations in the extent of digestion, since cleavage of surrounding bands is unaffected. In any case this would be removed since, for generating the footprinting plots, the intensity of bands is normalised with respect to bands outside the footprint which are not affected by ligand binding. Neither can it be explained by simply invoking strong ligand binding. This may be best illustrated by examining the cleavage patterns with sequence 3 (Fig. 4.11). This gives a clear footprint (>90% protection) with 40 μM ligand which begins to fade in a concentration dependent fashion with 75% protection at 1 μM. For a simple binding equation, in which the %bound = $L/(L + K_d)$ (where L is the free ligand concentration and K_d is the dissociation constant) then a value of 75% bound at 1 μM would equate to a dissociation constant of 0.3 μM. Using this parameter we can then estimate that only 3% should be bound at 0.01 μM ligand, in contrast to the 60% binding which is still observed.

If we perform the estimation beginning at the low ligand concentrations, then 60% bound at L = 0.01 μM yields a K_d of about 0.007 μM. On the basis of this very low dissociation constant we would predict that 98% should be bound at 0.3 μM, in contrast to the observed value of about 75%. It appears that the binding equation does not adequately describe the binding process. Examination of the footprinting plots in Fig.4.14 reveals that, for those instances where the fit is especially poor, the data are better described by a simple binding curve if the point for the control (zero drug concentration) is omitted. Revised footprinting plots, omitting the initial point, are presented in Fig. 4.14 and the calculated C_{50} values are listed in Table 4.2. Notice that the curves in Fig 4.15 are not fitted to zero. This analysis does not alter the conclusion that the best binding sites contain longer regions of alternating AT, but alters the apparent affinity of the binding sites.

Why then are the data points not adequately described by a simple binding curve? We can offer several explanations. Firstly it has previously been proposed that [*N*-MeCys^3,*N*-MeCys^7]TANDEM binds to alternating AT tracts with positive cooperativity (Lee & Waring, 1976). Although the origin of this effect is not clear, it may arise from drug-induced changes in DNA structure which potentiate the binding of further ligand molecules. Unusual kinetic

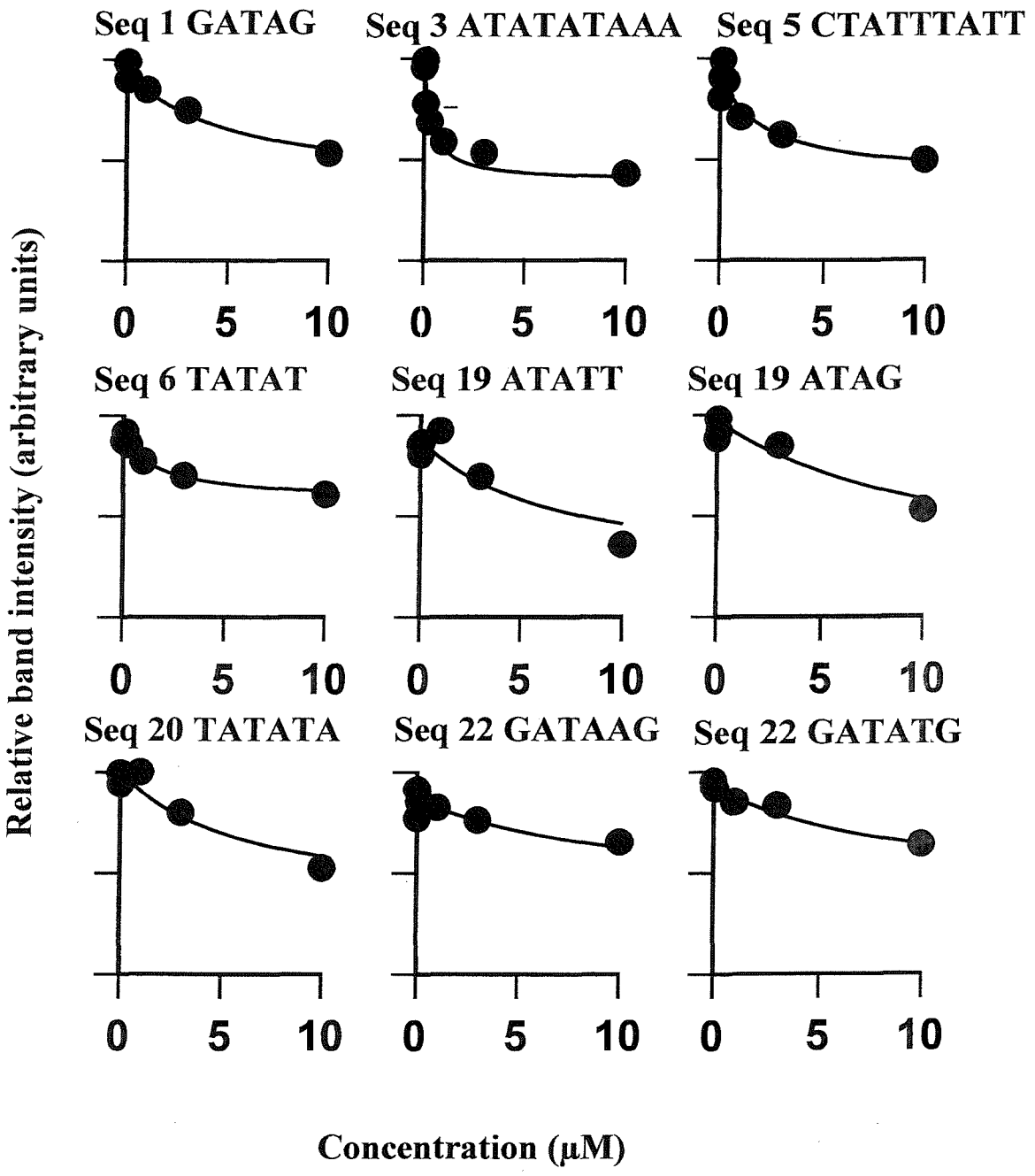


Figure 4.15 Footprinting plots showing the interaction of [*N*-MeCys^3,*N*-MeCys^7] TANDEM with the best TpA sites of DNA fragments obtained with REPSA, without the control values. Data was analysed as described in section 2.24.

Sequence #	Site	Sequence	C_{50} without control
1	1	GTATG	
	2	ATAG	5 ± 1
3	1	ATATATAAA	0.4 ± 0.2
5	1	CTATTTATT	2 ± 1
6	1	TATAT	1.6 ± 0.6
	2	CTATT	
19	1	ATATT	7 ± 4
	2	ATAG	11 ± 8
20	3	TATATA	6 ± 3
21	1	GTATG	
22	1	GATAAG	8 ± 3
	2	GATATG	11 ± 7

Table 4.2: C_{50} values (μM) determined for the interaction of [*N*-MeCys^3,*N*-MeCys^7]TANDEM with the various TpA sites on all the REPSA sequences without controls. The data were derived from quantitative analysis (section 2.24) of the footprinting plots shown in Figure 4.15.

effects for the binding of [*N*-MeCys^3,*N*-MeCys^7]TANDEM have also been observed in other studies. The dissociation from poly(dA-dT) is always described by a single exponential function, indicative of a single binding mode, yet the rate constant is strongly dependent on the saturation, becoming faster at higher levels of saturation (Fox *et al.*, 1982, Fletcher *et al.*, 1995). Secondly the observation may be a consequence of having overlapping binding sites. The simple binding curve describes the interaction of a ligand with isolated binding sites. The situation may be different for multiple overlapping sites. We can envision a situation in which the binding of one molecule, at any location along the repeat, is very strong, and in which the interaction with further molecules is much weaker. If there are two such identical sites, then at low concentrations they are equally likely to be occupied, and will therefore each be protected by 50%. This situation is more akin to negative cooperativity (in contrast to the previous suggestions), but may provide a qualitative explanation for the unusual binding curves observed with some sequences.

The persistence of a portion of the footprint at lower drug concentrations suggests that a fraction of the drug molecules are very tightly bound (maybe even by an unknown covalent mechanism). This is also consistent with the observation that the REPSA clones obtained with [*N*-MeCys^3,*N*-MeCys^7]TANDEM were more prone to contain mutations than with the other ligands.

CHAPTER 5

MULTISITE

5.1 Introduction

Several small molecules are known to bind to DNA in a sequence specific fashion, recognizing between 2-4 base pairs. One powerful technique which has been widely used for determining their sequence recognition is that of footprinting (section 1.8). However, as mentioned before, one limitation of this technique is that the unknown preferred binding site(s) for a ligand must be present within the footprinting substrate, which is typically less than 200 base pairs long. Since there are 10 possible dinucleotide steps, 32 different trinucleotides, 136 tetranucleotides and 512 pentanucleotides (Dervan, 1986), it is clear that, as the specificity of the ligand increases, there is a reduced chance of finding the correct site within any given restriction fragment. Even if the preferred binding site is present, a proper analysis of the specificity should examine the binding to related sequences, which may also not be present. We have therefore prepared a DNA footprinting substrate which contains all 136 tetranucleotide sequences.

5.2 Construction of the sequence

In order to construct the substrate a computer program was made in C++, the details of which are in figure (5.1) and in appendix A. The program was used to form sequences that contain all the tetranucleotides and in addition they had to comply with the following criteria:

1) The length of the strand had to be less than 160 bases in order to visualize most of the sequence on a single footprinting gel.

2) As far as possible the strand should be free of homopurine homopyrimidine tracts, and

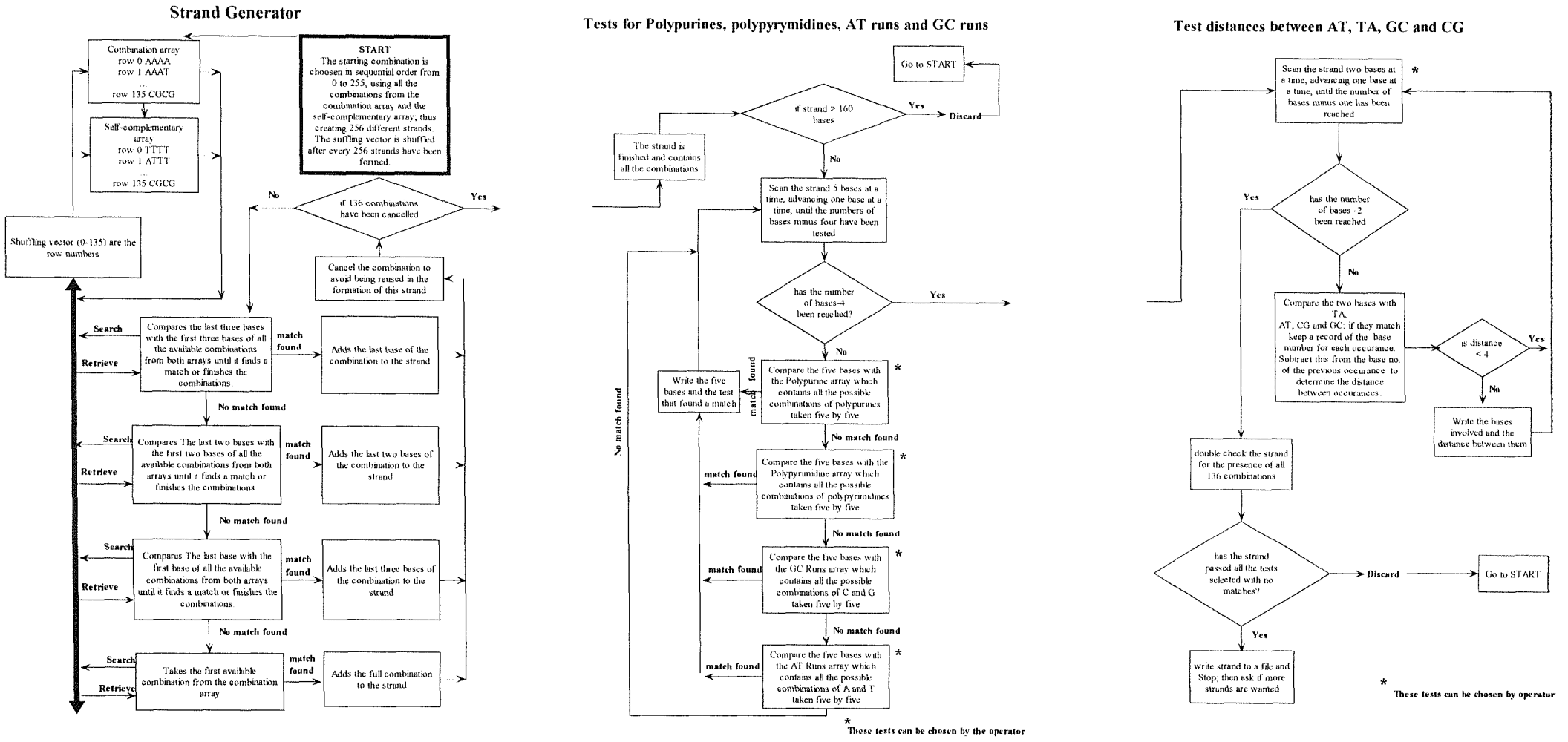


Figure 5.1 Flow chart of the computer program

regions of alternating AT or GC. These were avoided because they are known to adopt unusual structures and in some cases are poor substrates for DNase I (Laskowsky , 1979).

3) As far as possible each of the different GpC, CpG, ApT and TpA sites should be separated from each other preferably by more than five bases in order to unambiguously assign the binding sites. This will obviously not be possible for sequences such as TATA which contain two TpA steps.

The program was allowed to run for two weeks in a personal computer with a Pentium 200 MMX. At that time it made 10^8 sequences, out of the possible 10^{209}, without finding one that complied with all the criteria. Among the sequences generated by this program we chose one which satisfied as many of the criteria as possible. The chosen sequence and its characteristics are shown below.

5'-GCATTCGAGGCTGAGATGACAAAACAGACCCCACCGGACGTACTTTACATAA
CTCTTCACGCCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCTTGATCG
CGCTACGACTAGTGCAGTTGGAAATCGGCCATGTGTATTGCCGCATAT-3'
This strand has 156 characters

Undesirable features
This strand contains polypurines GGGAG, GGAAA, and polypyrimidines CTCTT, TCTTC.
This strand contains several CGs in a row CGCCC, CGCGC, CGGCC, GCCGC

CG, GC, AT, TA test:

This test scans the strand from 5' to 3' and measures the number of bases that separate each CpG, GpC, ApT, and TpA and reports only those which are closer than 5 bases. Also it counts the number of occurrences of each CpG, GpC, ApT and TpA step.

Num of CpG = 11
Num of GpC = 11
Num of ApT = 12
Num of TpA = 13
3 of the CpG steps are separated by only 4, 3 and 2 bases.
3 of the TpA steps are separated by only 4, 3, and 2 bases.
2 of the ApT steps are separated by only 4 bases.
1 of the GpC steps is separated by only 1 base from the next GpC.

The sequence GGATCC was added to both ends of this strand to allow cloning into the *Bam*HI site of pUC18. This oligonucleotide was synthesized by Oswell DNA service, Southampton UK. The complementary strand was generated by annealing a short primer (5'- GCTCGGATCCATATGCG-3') to the 3'-end, and the sequence was made fully double stranded by extension with *Taq* DNA polymerase. The resulting fragment was cut with *Bam*HI and cloned into the *Bam*HI site of pUC18. Two clones were obtained in which the sequence was in opposite orientations. The combination of these two clones allows easy analysis of both ends of the fragment in footprinting experiments. The sequences of these fragments are:

MS1

5'-GGATCCATATGCGGCAATACACATGGCCGATTTCCAACTGCACTAGTCGTAG
CGCGATCAAGGTTAAGCTCCCGTTCTATCCTGGTATAGCAATTAGGGCGTGAAG
AGTTATGTAAAGTACGTCCGGTGGGGTCTGTTTTGTCATCTCAGCCTCGAATGC
GGATCC -3'

MS2

5'-GGATCCGCATTCGAGGCTGAGATGACAAAACAGACCCCACCGGACGTACTTT
ACATAACTCTTCACGCCCTAATTGCTATACCAGGATAGAACGGGAGCTTAACCT
TGATCGCGCTACGACTAGTGCAGTTGGAAATCGGCCATGTGTATTGCCGCATAT
GGATCC-3'

5.3 Echinomycin

Figure 5.2 shows DNase I digestion of fragments MS1 and MS2 in the presence of various concentrations of echinomycin. It can be seen that, as expected, echinomycin produces footprints at most of the CpG sites (see section 1.4.5.1). Looking at the data for MS1 it can be seen that site 1 (GCGG) is a poor echinomycin binding site. On MS2 the footprint at site 7 (TCGC) is also weak, though this is not easily resolved from the adjacent site 8 (GCGC). Therefore, when two sites overlap we chose bands towards the top and

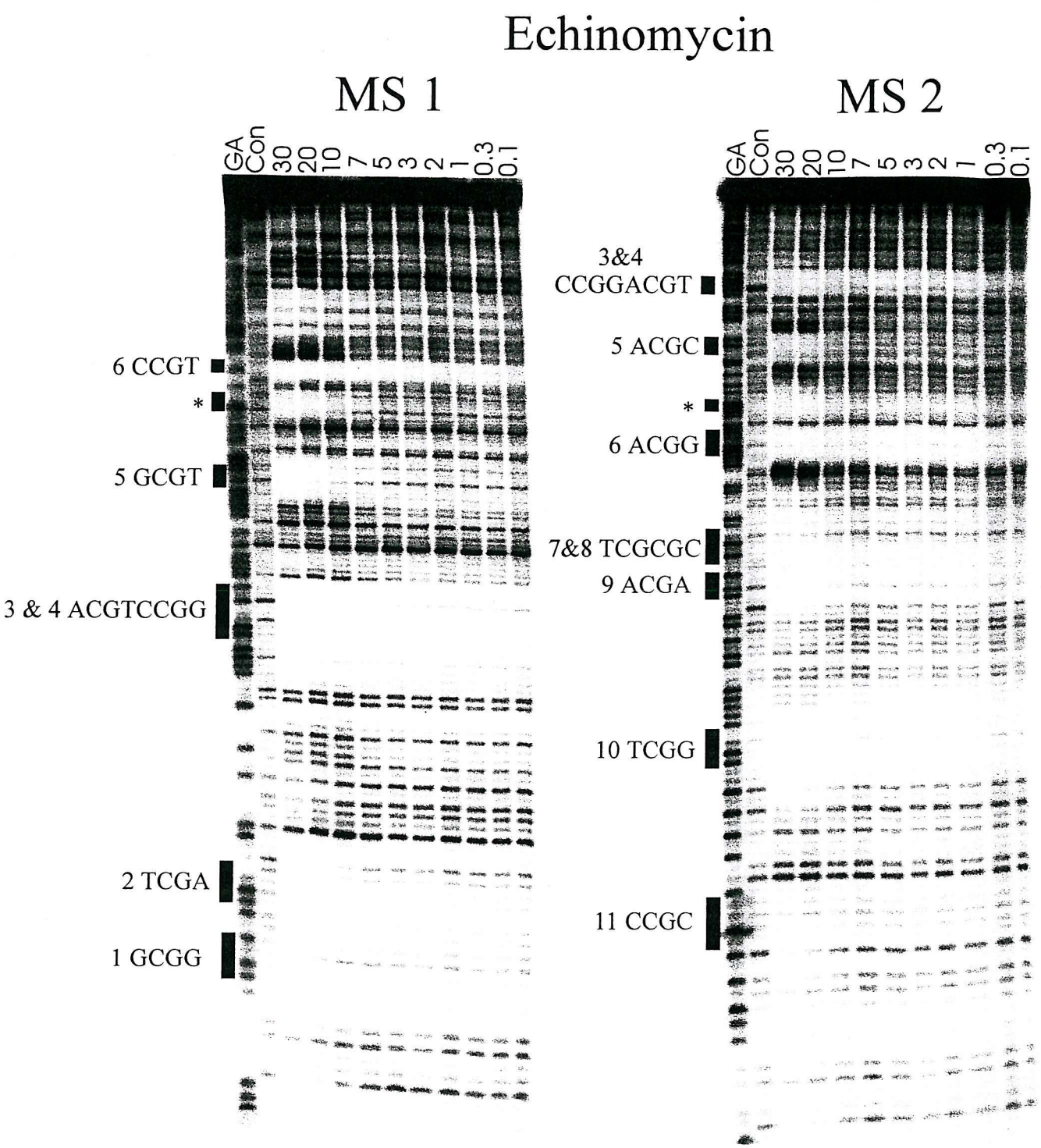


Figure 5.2 DNase I footprinting of MS1 and MS2 in the presence and absence of echinomycin. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each CpG step on the fragment. The track labeled GA is a marker specific for guanine and adenine. The site labeled * is an unexpected binding site. The fragments are labeled at the 3'-end.

bottom of the footprint for the separate analysis. It can be seen that the footprints at sites 4 (ACGT) and 6 (CCGT) persist to lower concentrations than the other sites. These two sites also produce best footprints on MS2. In addition it can be seen that sites 7 (TCGC) and 10 (TCGG) and 11(CCGC) require higher ligand concentrations to produce a footprint. Footprinting plots, derived from quantitative analysis of these data, are presented in Figure 5.3 and yield the C_{50} values shown in table 5.1.

		C_{50}	
Site	**Sequence**	**MS1**	**MS2**
1	GCGG/CCGC	16 ± 4	Not Resolved
2	TCGA	12 ± 5	Not Resolved
3	CCGG	No footprint	Not Resolved
4	ACGT	Too low to measure	Not Resolved
5	GCGT/ACGC	11 ± 3	21 ± 4
Site *		14 ± 3	25 ± 4
6	CCGT/ACGG	Too low to measure	Too low to measure
7	TCGC/GCGA	Not Resolved	32 ± 5
8	GCGCG/CGCGC	Not Resolved	18 ± 9
9	ACGA/TCGT	Not Resolved	11 ± 6
10	TCGG/CCGA	Not Resolved	22 ± 6
11	CCGC/GCGG	Not Resolved	20 ± 7

Table 5.1 C_{50} values (μM) determined for the interaction of echinomycin with the various CpG sites on fragments MS1 and MS2. The data were derived from quantitative analysis (section 2.24) of the footprinting plots shown in Figure 5.3.

From these data it appears that the best echinomycin binding sites are ACGT and ACGG, while GCGA and TCGG are poor ligand binding sites. The remainder of the sites have intermediate affinities. In general it appears that the flanking G and C bases produce weaker binding sites.

In addition to these footprints at CpG sites, echinomycin produced a footprint at an

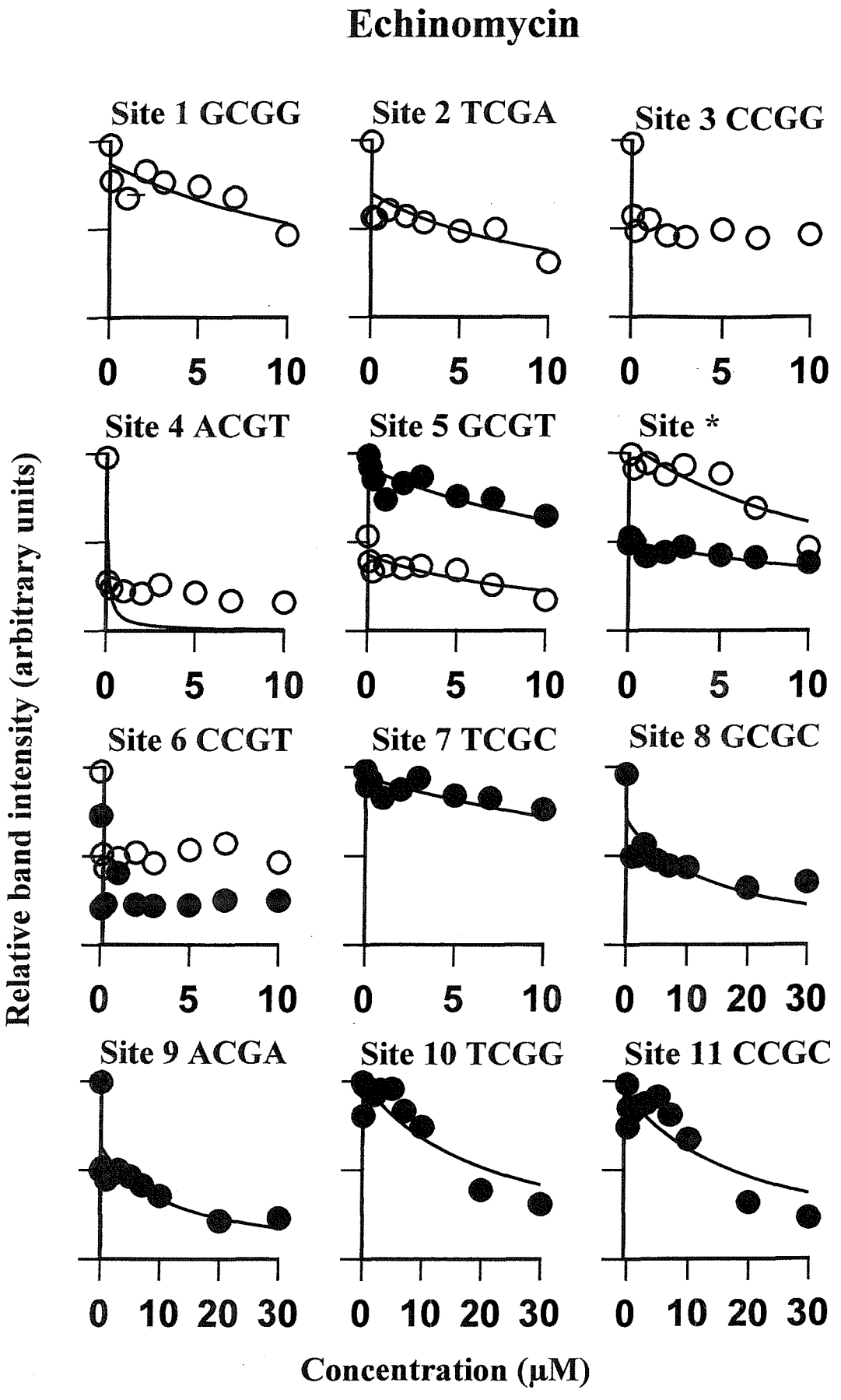


Figure 5.3 Footprinting plots showing the interaction of echinomycin with the CpG sites in DNA fragments MS1 and MS2, ○ corresponds to data for MS1 and ●are taken from MS2. Data was analysed as described in section 2.24.

additional site which does not contain a CpG step. This unknown site is marked with * on the gel (Fig.5.2) and appears within the sequence TATCCTGGTATG, though it is not possible to rigorously identify the binding site by DNase I footprinting . Although this is not a strong binding site it appears to be better than several of the other CpG sites and this will be explored further in section 5.3.2.

5.3.1 DEPC patterns for echinomycin with MS1 and MS2

Diethyl-pyrocarbonate (DEPC) is a chemical probe which reacts with purines at the N7 position (located in the major groove of B-DNA) resulting in products in which the imidazole ring has been opened (Vincze *et al.*, 1973). Subsequent treatment with alkali leads to DNA strand scission, since the N9 glycosidic bond is labilized due to the carbethoxylation and ring opening. Previous studies have shown that echinomycin renders adenines on the 3'-side of its binding site (i.e. CG<u>A</u>) hyperreactive to DEPC (Portugal *et al.*, 1988; Mendel and Dervan, 1987; Jeppesen and Nielsen, 1988). The results of DEPC modification of fragments MS1 and MS2 in the presence and absence of echinomycin are presented in Fig. 5.4. The weak cleavage in the control lane corresponds to modification of purines (A and G). The gel shows that there are enhancements on most of the CG<u>A</u> sites in the presence of echinomycin. However, no such enhancement is evident at GCGA and CCGA both of which are weak echinomycin sites. In addition other sites which do not contain a CpG step show enhancements ATC<u>AA</u>, TCC<u>AA</u>, ACC<u>A</u> and AG<u>AA</u> . These results may help in identifying the unknown echinomycin binding site.

5.3.2 A different binding site for Echinomycin

As mentioned previously, echinomycin appears to bind to a site that does not contain a CpG step. This is located in the sequence 5'- TATCCTGGTATA-3', though the exact binding site can not be determined by DNase I footprinting. In order to confirm the new echinomycin binding site a new DNA fragment was designed, which contained the estimated sequence plus another good echinomycin binding site 5'-ACGT-3'. This fragment contained the insert:

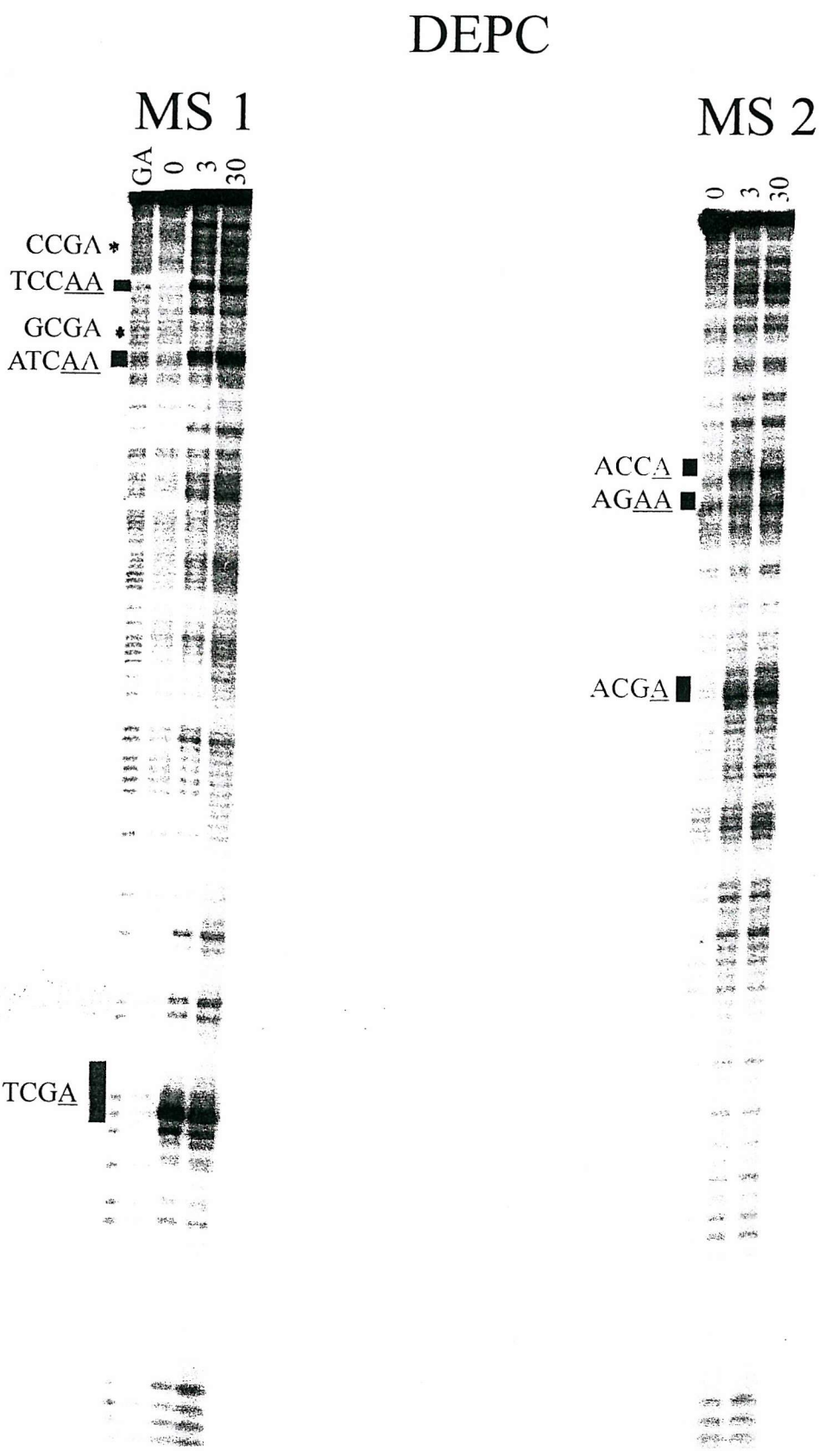


Figure 5.4 DEPC footprinting gel of MS 1 and MS 2 in the presence and absence of echinomycin. Drug concentrations (μM) are indicated at the top of each gel lane. The boxes indicate the regions of enhancement. The * indicates CGA sites that did not produced enhancements. The fragments are labeled at the 3'end. The track labeled GA is a marker specific for guanine and adenine.

E8

5'-GATCACGTTCTATCCTGGTATAGATC-3'

This oligonucleotide was annealed to its complementary sequence, and ligated into *Bam*HI cut pUC18. After transforming into *E. coli* the plasmid was isolated and its sequence was confirmed by dideoxy sequencing . The *Hin*dIII-*Sac*I fragment containing this insert was labelled in the usual way (section 2.18.1). The DNase I footprinting gel obtained with the DNA fragment E8 is presented in Fig 5.5, and the differential cleavage plot obtained from this gel is shown in Fig 5.6. It can be seen that a footprint is generated around the CpG step in the presence of echinomycin, though the apparent affinity of this site (ACGT) was lower than expected since the experiment with MS1 and MS2 had suggested that this was one of the best echinomycin binding sites. In addition a faint footprint can be seen in the vicinity of the unknown binding site. Although this confirms that echinomycin binds to this site, the exact location of the drug is still uncertain. For this reason we decided to examine the sensitivity of the DNA to diethyl pyrocarbonate (DEPC) instead of DNase I.

In section 5.3.1 we showed the DEPC modification patterns of echinomycin with MS1 and MS2. Now we will do a similar experiment with E8 to try to determine the binding site more precisely. The results of such an experiment are shown in Fig 5.4. It can be seen that echinomycin enhances the reactivity of several, but not all, adenines in this sequence (T<u>A</u>TACC<u>A</u>GG<u>A</u>TAG<u>AA</u>). Of all these sites ACC<u>A</u> has the largest enhancement. When we consider the results from the DNase I footprinting and the DEPC modification patterns of MS1, MS2 and E8 together (Fig 5.2, 5.5 and 5.6) it suggests that ACCA is a candidate for the identity of the unknown binding site found within the MS1 and MS2 fragment. Also, note that there is an enhancement for this site in MS2 (Figure 5.4 right hand panel).

5.3.3 Summary

The MS1and MS2 fragments show that echinomycin does not bind to all CpG sites with the same affinity. The binding affinity appears to be influenced by the identity of the

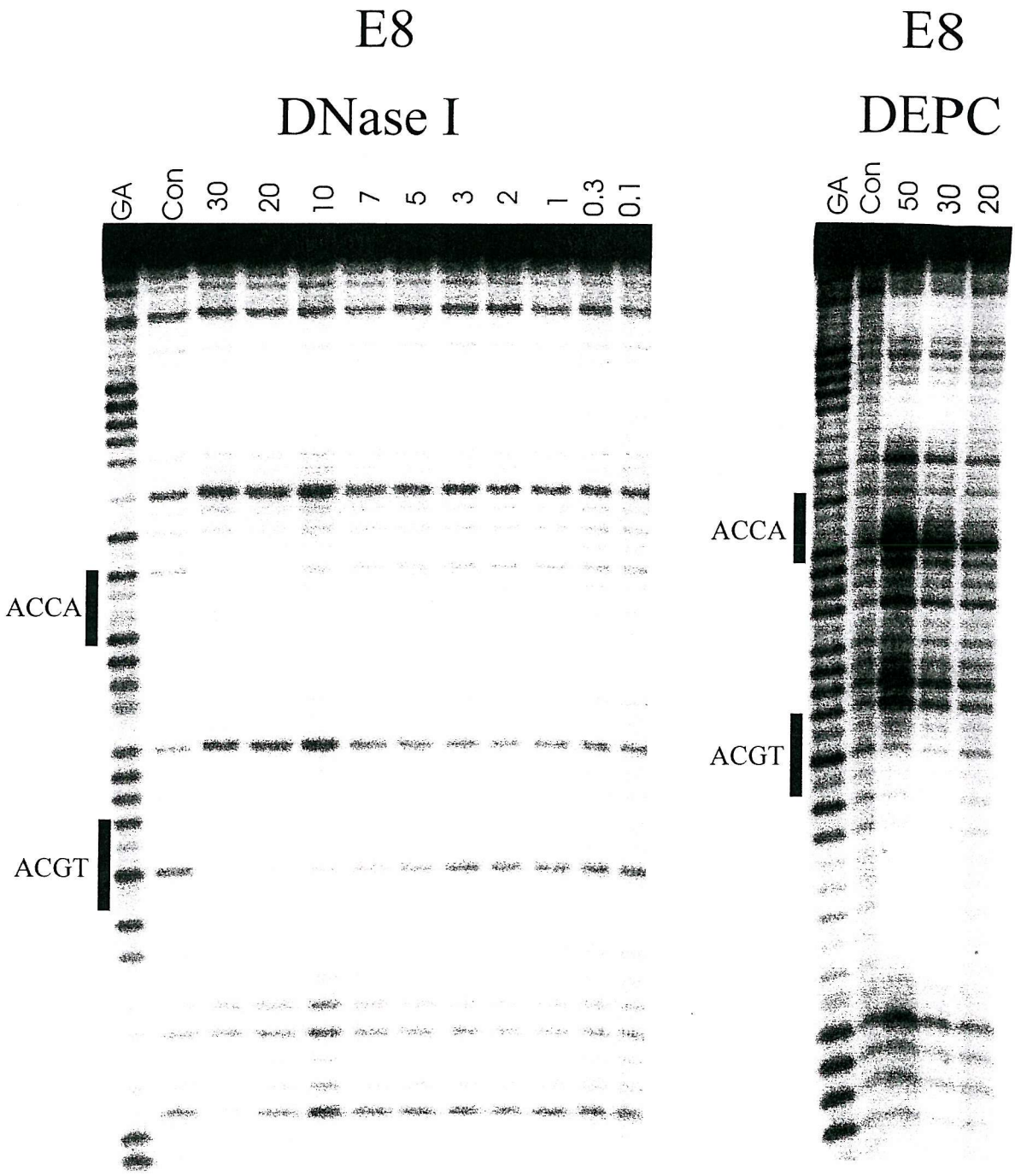


Figure 5.5 DNase I and DEPC footprinting of the E8 in the presence and absence of echinomycin. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of the known echinomycin binding site and of the suggested new binding site. The track labeled GA is a marker specific for guanine and adenine. The track labeled 'Con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

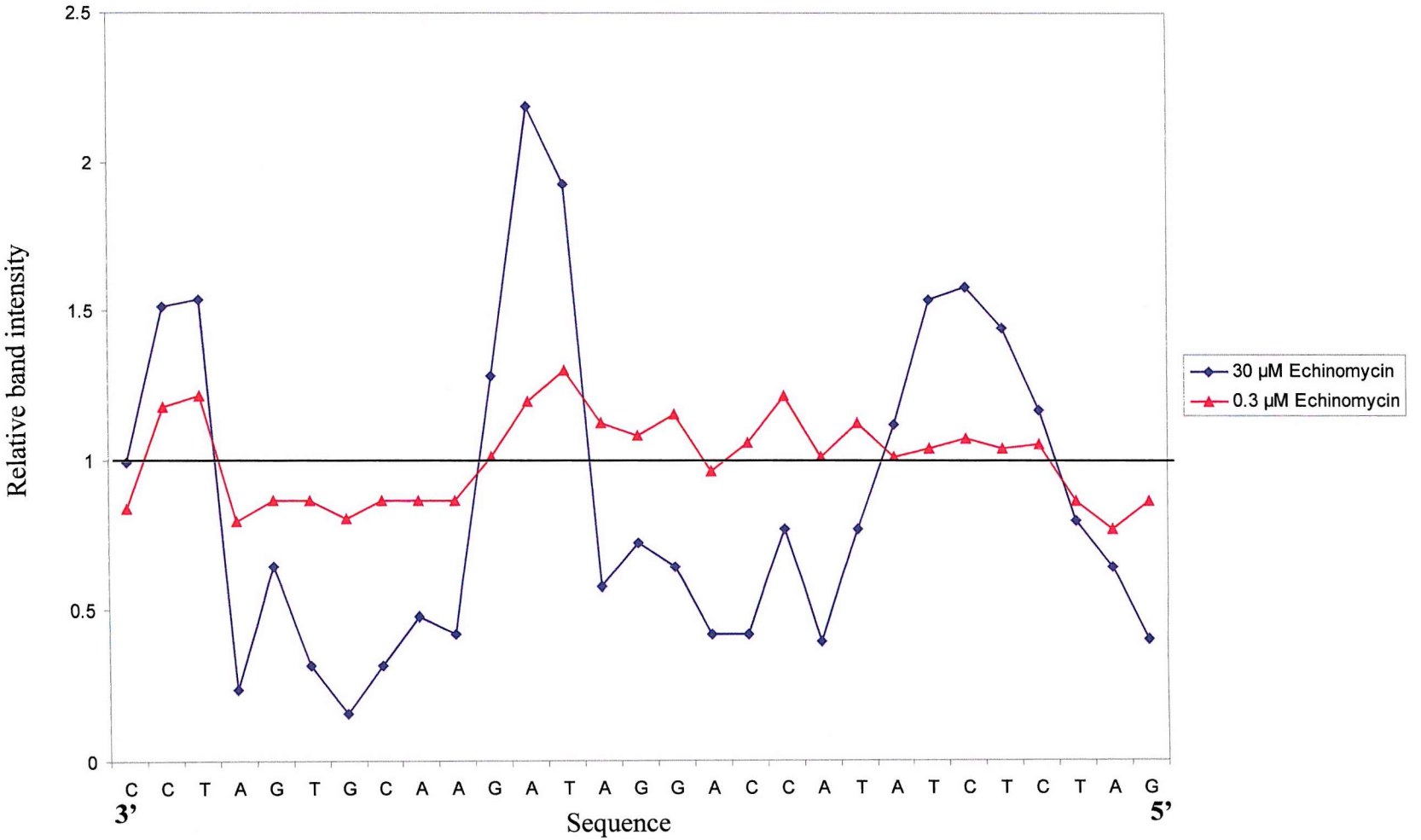


Figure 5.6 Differential cleavage plot showing the relative band intensities of the DNase I footprinting gel shown in Fig 5.5 for the highest and lowest concentrations (30 μM and 0.3 μM) of echinomycin. Values less than one show protection from cleavage while those greater than one show relative enhancement.

flanking bases. In general it appears that flanking by either G or C bases produces the weakest binding sites, while those sites which are flanked by A or T on both sides produces the strongest binding sites. The best sites appear to be ACGT (site 4) and TCGA (site 2), and the weakest sites are GCGA (site 7) and CCGG (site 3). Also a new binding site for echinomycin has been discovered with these fragments which is tentatively assigned to the sequence ACCA. This finding will be considered further in the Discussion.

5.4 [*N*-MeCys^3, *N*-MeCys^7]TANDEM

Figure 5.7 shows the interaction of [*N*-MeCys^3, *N*-MeCys^7]TANDEM with DNA fragments MS1 and MS2. This synthetic ligand has previously been shown to bind to the dinucleotide TpA, especially when this is surrounded by A and T residues (section 1.4.5.3). It can be seen that clear footprints are evident at most of the TpA sites, including some in which this step is flanked by GC residues. The various sites are less easily characterized with MS1 since they are located toward the centre and top of the gel, but are much better resolved on MS2. Visual inspection suggests that site ATAT (site 12) produces the best footprint, which persists to lower concentrations than the other sites; the bands within this site are still attenuated at 3 μM ligand concentration. In contrast cleavage of site TTAA (site 8) is barely affected, even at the highest ligand concentrations.

Footprinting plots showing the relative intensity of bands in each footprint at different ligand concentrations are shown in Fig 5.8. In some instances, even with this synthetic fragment, the binding sites are close together and it is difficult to resolve the interaction with isolated sites. Where two sites are close together we chose bands toward the top and bottom of the footprint for the analysis. Sites 5 (GTAT) and 6 (ATAG) overlap in the sequence GTATAG and binding to one site must exclude binding to the other. However, both of these sites are represented elsewhere on the fragment GTAT (site 11) and ATAG (site 7). In addition, ambiguities in individual drug binding sites were resolved by referring to the observation that DNase I footprints are staggered in the 3' direction relative to the drug binding site. C_{50} values corresponding to the ligand concentration which reduces the intensity of bands within each footprint by 50% are presented in table 5.2. These C_{50}

[*N*-MeCys^3, *N*-MeCys^7] TANDEM

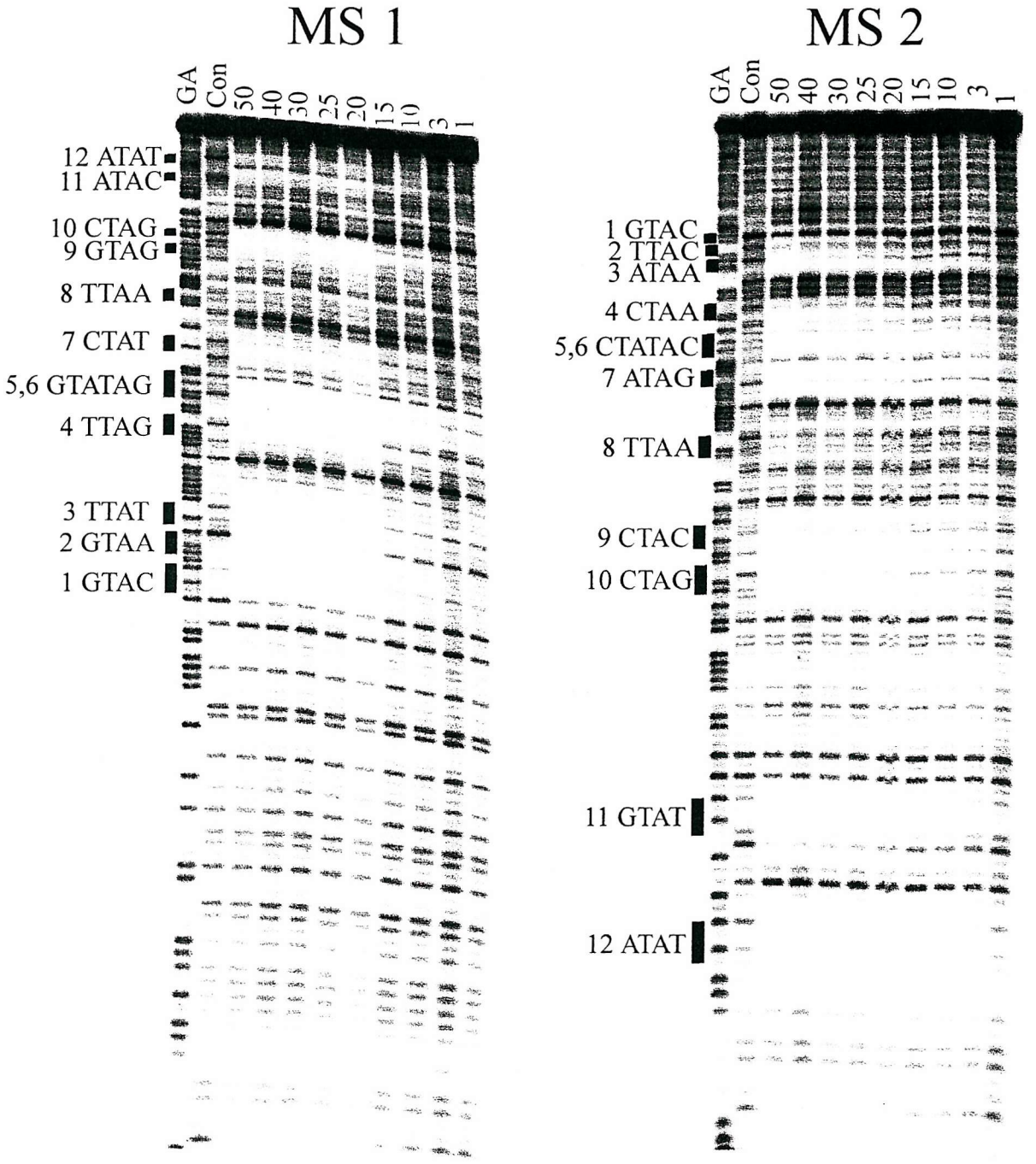


Figure 5.7 DNase I footprinting of MS1 and MS2 in the presence and absence of [*N*-MeCys , *N*-MeCys] TANDEM. Drug concentrations (μM) are indicated at the top of each gel lane.The black boxes indicate the position of each TpA step on the fragment. The track labeled GA is a marker specific for guanine and adenine. The track labeled 'Co is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

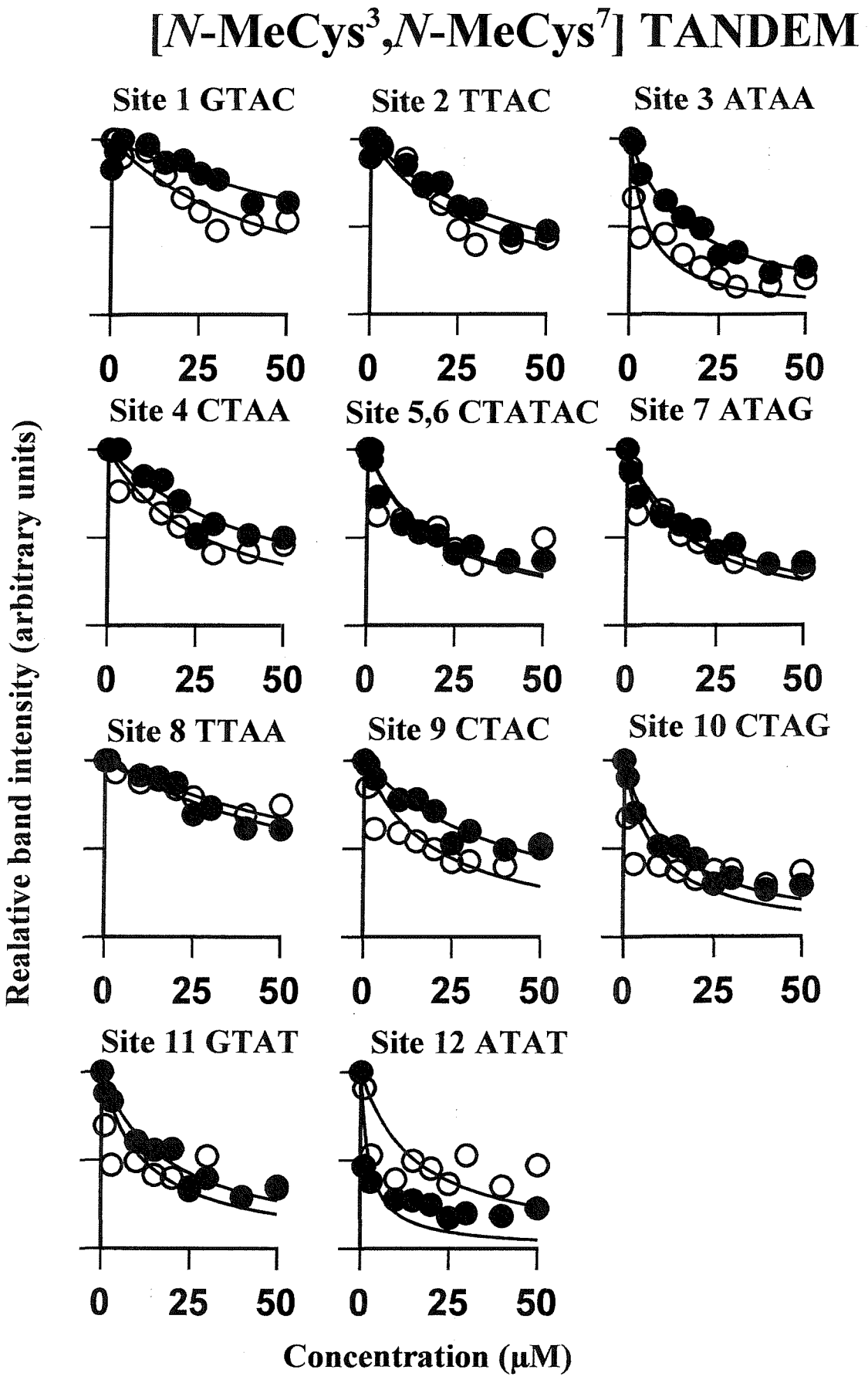


Figure 5.8 Footprinting plots showing the interaction of [*N*-MeCys^3,*N*-MeCys^7] TANDEM with the TpA sites in DNA fragments MS1 and MS2, ○ corresponds to data for MS1 and ●are taken from MS2. Data was analysed as described in section 2.24.

values confirm that ATAT and TTAA are the best and worst sites respectively with affinities that differ by more than 1 order of magnitude. It is worth noting that footprints toward the top of the gel are less easy to quantify than those toward the bottom and that as a result the C_{50} values with MS1 are more accurate for low site numbers whereas greater reliance can be placed on the high site numbers for MS2.

		C_{50}	
Site	**Sequence**	**MS1**	**MS2**
1	GTAC	43 ± 5	94 ± 15
2	TTAC/GTAA	31 ± 4	43 ± 4
3	ATAA/TTAT	5 ± 1	16 ± 1
4	CTAA/TTAG	27 ± 3	45 ± 6
5,6	CTATAC/GTATAG	20 ± 4	19 ± 2
7	ATAG/CTAT	17 ± 2	20 ± 2
8	TTAA	100 ± 10	80 ± 10
9	CTAC/GTAG	20 ± 4	42 ± 3
10	CTAG	9 ± 3	13 ± 1
11	GTAT	12 ± 3	18 ± 2
12	ATAT	16 ± 4	2.5 ± 0.9

Table 5.2 C_{50} values (μM) determined for the interaction of [*N*-MeCys^3, *N*-MeCys^7]TANDEM with the various TpA sites on fragments MS1 and MS2. The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 5.8.

5.4.1 Summary

The footprinting experiments with fragments MS1 and MS2 have demonstrated that the binding of [*N*-MeCys^3, *N*-MeCys^7]TANDEM is affected by the identity of the surrounding bases. Not all TpA steps are good binding sites. The data suggest that the best binding sites contain the sequences ATAX and XTAT. This is consistent with the original observation that the ligand binds very tightly to poly(dA-dT) (Lee and Waring, 1978). These binding preferences are considered further in the Discussion. The interaction of [*N*-MeCys^3, *N*-MeCys^7]TANDEM with long regions of poly(dA-dT) is further explored in chapter 6.

5.5 Actinomycin D

DNase I footprinting experiments showing the interaction of actinomycin D with the fragments MS1and MS2 are presented in Fig 5.9. This ligand has previously been shown to bind to the dinucleotide GpC (section 1.5.1). Inspection of the gels shows that actinomycin D produces clear footprints around each of the GpC sites. Looking at the data for MS1 it can be seen that sites 3 (GGCG) and 2 (AGCC) are only protected at the highest ligand concentrations, while footprints at the other sites persist to much lower ligand concentrations.

		C_{50}	
Site	**Sequence**	**MS1**	**MS2**
1	TGCG/CGCA	6 ± 2	Not Resolved
2	AGCC/GGCT	12 ± 4	Not Resolved
	CCGGTGGGG	10 ± 3	Not Resolved
3	GGCG/CGCC	10 ± 2	Not Resolved
4	AGCA/ TGCT	5 ± 2	2.6 ± 0.8
5	AGCT	5 ± 2	3 ± 0.6
	TCCTGGT/ACCAGGA	24 ± 12	17 ± 5
6 & 7	AGCGCG/CGCGCT	6± 2	3 ± 1
8	TGCA	4 ± 2	2 ± 1
9	GGCC	Not resolved	No Footprint
10	TGCC/GGCA	Not Resolved	1 ± 0.2
11	CGCA/TGCG	Not Resolved	1 ± 0.15

Table 5.3 C_{50} values (μM) determined for the interaction of actinomycin with the various GpC sites on the fragment MS1 and MS2 The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 5.10.

On MS2 no footprint is evident at site 9 (GGCC). The footprinting plots derived from the gels are presented in figure 5.10 and the C_{50} values are shown in table 5.3. It is worth noting that, as [*N*-MeCys^3, *N*-MeCys^7]TANDEM above, the various sites are less

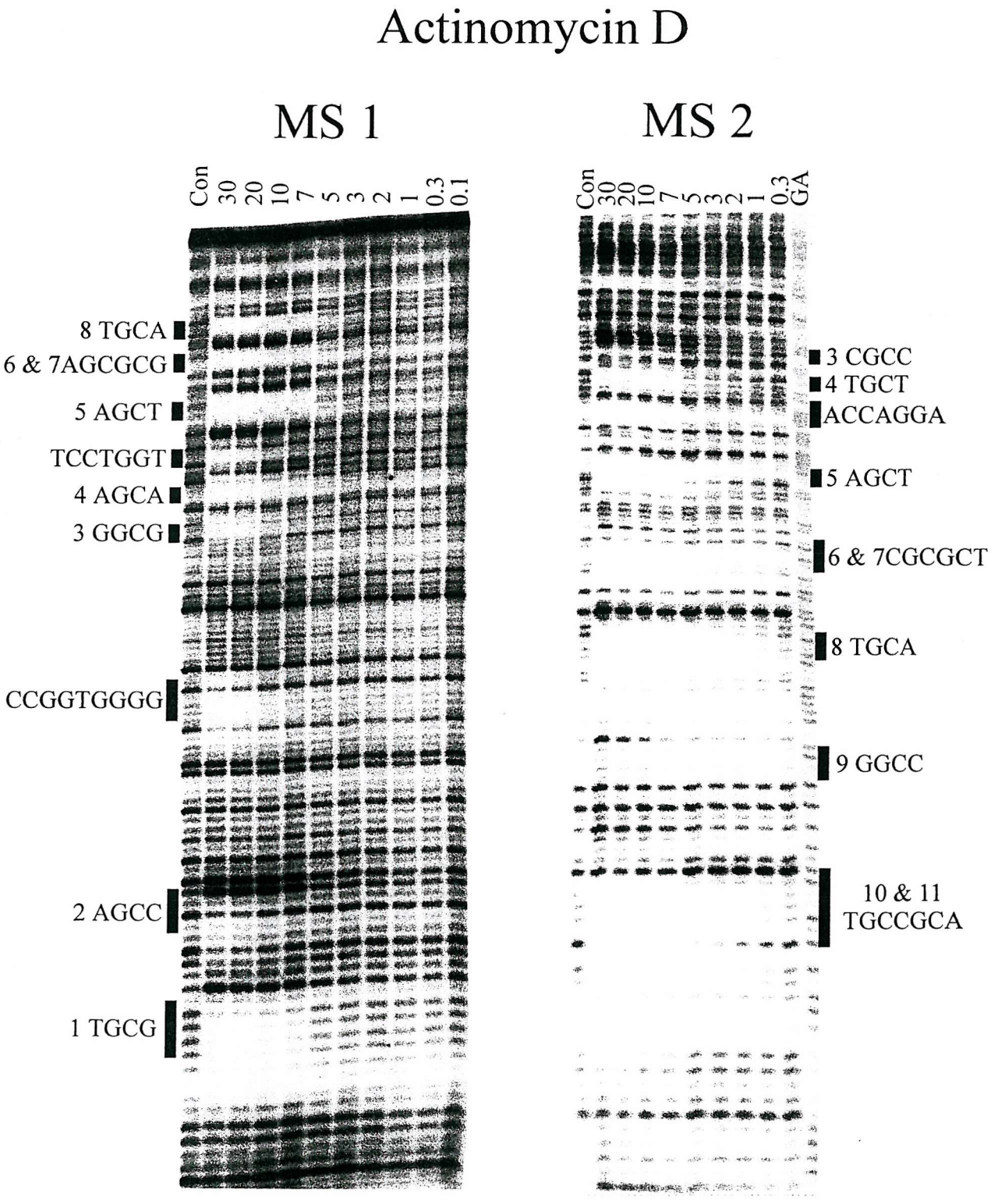


Figure 5.9 DNase I footprinting of MS1 and MS2 in the presence and absence of actinomycin D. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each GpC step on the fragment. The track labeled GA is a marker specific for guanine and adenine.The track labeled 'Con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

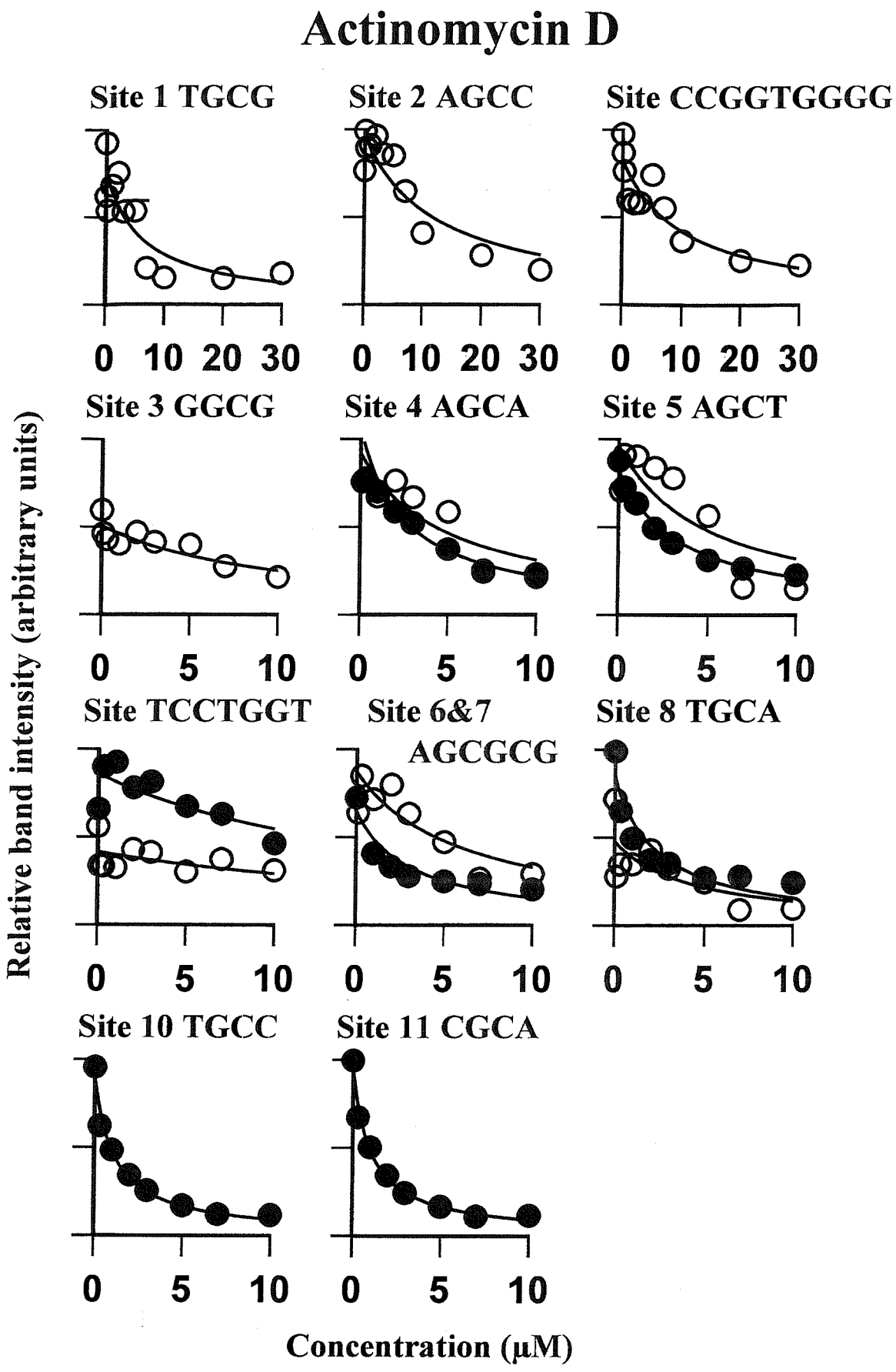


Figure 5.10 Footprinting plots showing the interaction of actinomycin D with the GpC sites in DNA fragments MS1 and MS2, ○ corresponds to data for MS1 and ●are taken from MS2. Data was analysed as described in section 2.24.

easily characterized with MS1 since they are located toward the centre and top of the gel, but are much better resolved on MS2. Therefore, when comparing binding sites between MS1 and MS2 more importance is given to the C_{50} values obtained with MS2.

These data confirm that GGCC (site 9), AGCC (site 2) and GGCG (site 3) are weak binding sites and suggest that GGCA (site 10) and TGCG (site 11) are amongst the best binding sites for this agent. In addition these results show that actinomycin D also binds to other sites that do not contain a GpC step. These sites are among the sequences 5'-CCGGTGGGG-3' (MS1) and 5'-TCCTGGT-3' (MS2). The affinity for these sites is low, though higher than that obtained for site 9 (GGCC), and only high concentrations of the ligand will provide a footprint. These sites have not been studied further in this thesis, but are considered further in the Discussion.

5.5.1 Summary

These results indicate that actinomycin D binding to GpC is also affected by the flanking regions and show that the best binding sites are GGCA (site 10) and TGCG (site 11), and the weakest site is GGCC (site 9). We also found that this ligand is able to bind to other sequences besides its preferred GpC sequence, and this subject will be considered further in the Discussion.

5.6 Distamycin

DNase I footprinting patterns of the fragments MS1 and MS2 in the presence and absence of distamycin are represented in Fig 5.11. This ligand has previously been shown to have a preference for regions of high A and T content (see section 1.3.1.1). Visual inspection of the footprinting gels confirms that all the sites that contain at least four A and/or T residues are protected from DNase I cleavage by the ligand. Of all these sites TTTT (site1) and AATT (site 4) seem to display the highest affinity for distamycin. However, this ligand seems to bind efficiently to all sites that contain a TA tract including one that contains a C residue in the middle (TTCTATC).

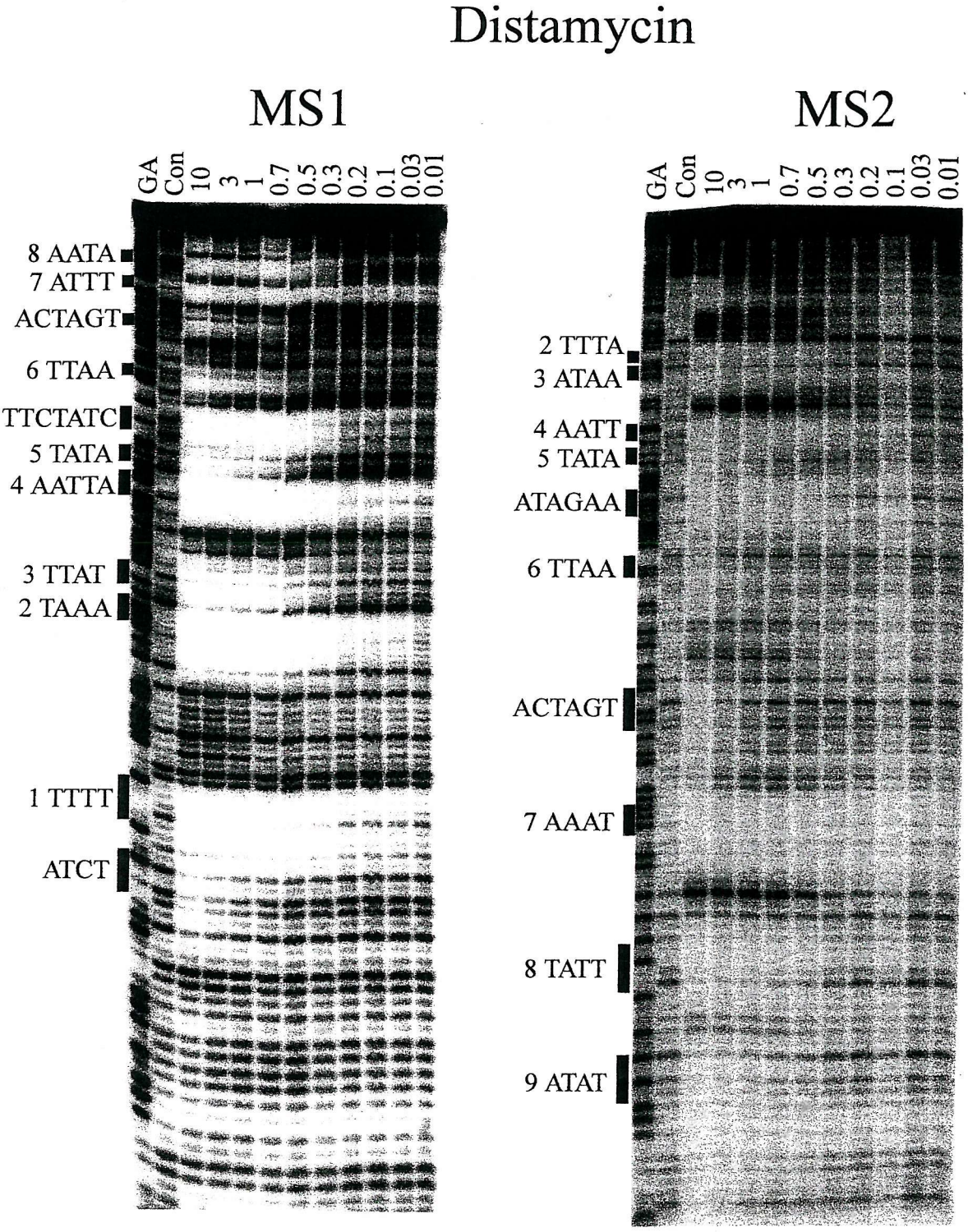


Figure 5.11 DNase I footprinting of MS1 and MS2 in the presence and absence of distamycin. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each AT track step on the fragment. The track labeled GA is a marker specific for guanine and adenine. The track labeled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labeled at the 3'-end.

The band intensities of the gels were analysed with the phosphoimager and the footprinting plots obtained are shown in Fig. 5.12. The C_{50} values are presented in table 5.4. It should be noted that the absolute values for C_{50} are higher for MS2 than MS1, we have no explanation for this difference, but the important feature is the ranking order of sites in each case is the same.

		C_{50}	
Site	**Sequence**	**MS1**	**MS2**
	ATCTCA/TGAGAT	1.8 ± 0.68	Not Resolved
1	TTTT/AAAA	0.01 ± 0.006	Not Resolved
2	TAAA/TTTA	0.2 ± 0.05	0.5 ± 0.2
3	TTAT/ATAA	0.4 ± 0.08	7 ± 2
4	AATTA/TAATT	0.01 ± 0.006	0.3 ± 0.1
5	TATA	0.4 ± 0.1	2 ± 0.5
	TTCTAT/ATAGAA	0.07 ± 0.02	0.6 ± 0.3
6	TTAA	3 ± 0.7	5 ± 1
	ACTAG/CTAGT	1 ± 0.5	2 ± 0.65
7	ATTT/AAAT	0.6 ± 0.3	3 ± 1
8	AATA/TATT	0.2 ± 0.04	0.5 ± 0.2
9	ATAT	Not resolved	0.6 ± 0.2

Table 5.4 C_{50} values (μM) determined for the interaction of distamycin with the various AT tracks on the fragments MS1 and MS2. The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 5.12.

These C_{50} values support our visual inspection and show that TTTT/AAAA is the best binding site together with AATTA /TAATT while TTAA is an especially poor site. This result can be compared with those obtained previously by footprinting and circular dichroism (Abu-Daya *et al.*, 1995; Chen and Sha, 1998). Our results show that AATTA= AAAA > TATA >TAAA = ATAT > TTAT > AATA >ATTT >TTAA. Those obtained by footprinting show AAAA= AATT > ATTA/ TAAT = TATA > ATAT (Abu-Daya *et al.*, 1995) and similar binding affinities were found by circular dichroism (Chen and Sha, 1998).

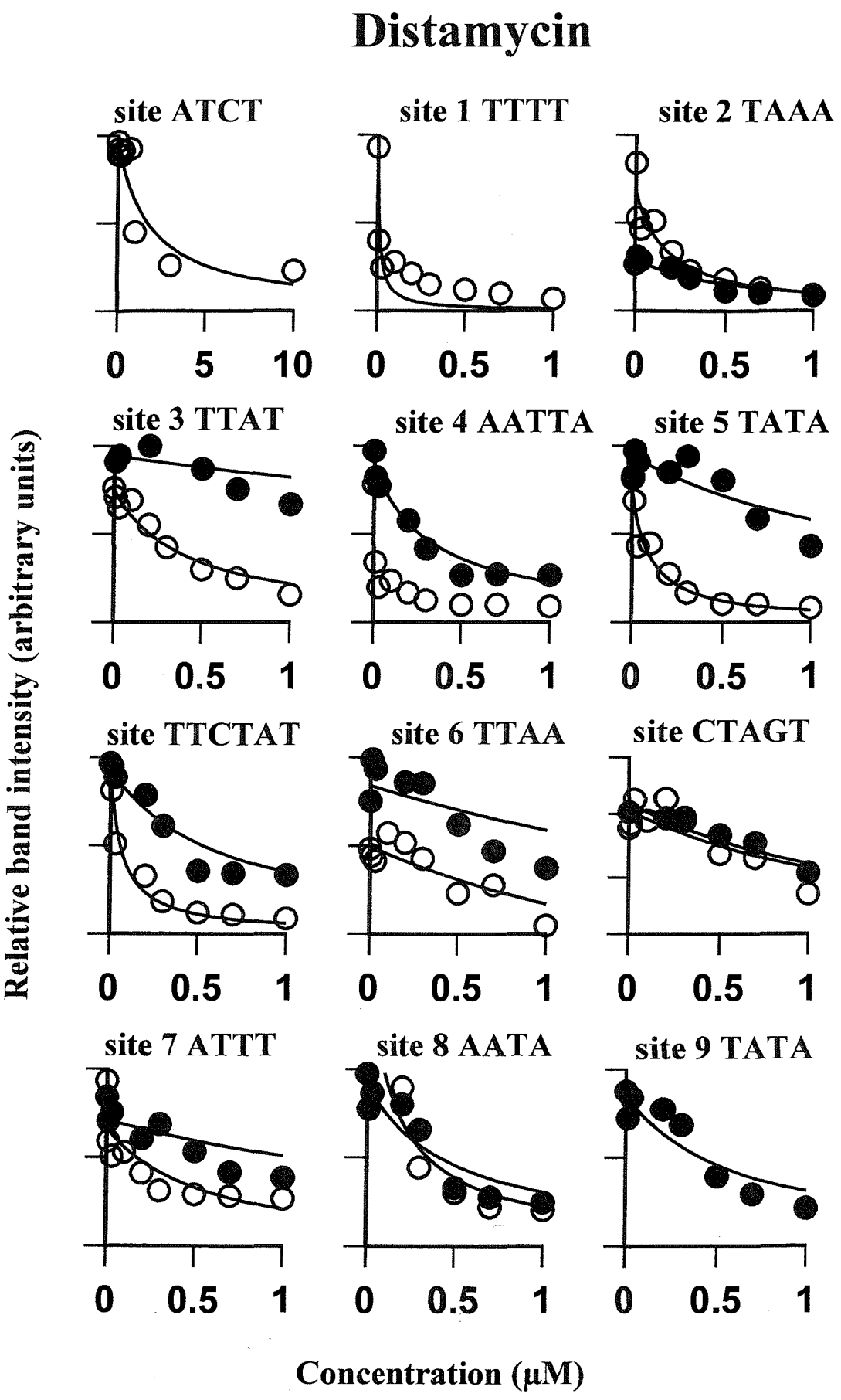


Figure 5.12 Footprinting plots showing the interaction of distamycin with the AT tracts in DNA fragments MS1 and MS2, ○ corresponds to data for MS1 and ●are taken from MS2. Data was analysed as described in section 2.24.

All these results confirm that AAAA and AATT are among the best binding sites for this ligand. In addition to these TA tracts distamycin bound to some sites that contain C in the middle such as ATCTCA and TTCTAT. This is consistent with some previous observations that this agent is able to bind to sites that contain a GC base pair situated either at the centre or at the sides of the binding site (Churchill *et al.*, 1990; Van Dyke *et al.*, 1982). These results will be considered further in the Discussion.

5.6.1 Hydroxyl radical footprinting of distamycin

Hydroxyl radical footprinting is not as good as DNase I footprinting for assessing ligand binding strengths, however it provides a better indication of the ligand's location. We therefore used hydroxyl radical footprints to identify the distamycin binding sites more accurately. Hydroxyl radical footprinting gels of MS1 and MS2 in the presence and absence of distamycin are shown in Fig. 5.13. It can be seen that this probe produces an even ladder of cleavage in the controls and that there are attenuated bands at each of the AT tracks. Densitometer traces obtained with the phosphoimager and the ImageQuant software are shown in Fig 5.14 for MS1 and Fig. 5.15 for MS2. On these densitometry traces each valley indicates a distamycin binding site. This analysis helps to detect binding sites that were not very evident in the DNase I footprinting gel (Fig 5.11). Several of these sites include GC residues, such as TTGA, CATC and CTAGTT. (Fig 5.15). This illustrates the power of hydroxyl radical cleavage over DNase I footprinting. If we compare the DNase I footprinting gels (Fig 5.11) with the intensity plots obtained for the hydroxyl radical gels (Fig 5.14 and 5.15) it can be seen that the DNase I footprints are broader and at times cover more than one binding site. For example Sites 4, 5 and TTCTAT on MS1 (Fig 5.11 left hand panel) almost appear as a single footprint at the higher concentrations, this effect is even more evident with MS2 (Fig 5.11 right hand panel) for the same sites. However, in the hydroxyl radical intensity plots (Fig 5.14 and 5.15) and the hydroxyl radical gel (Fig 5.13), we can see that the demarcation within footprints is much clearer and the separation between sites 4, 5 and CTAGTT is more evident for both MS1 and MS2. These differences in footprinting demarcation are due to the difference in size of these probes. DNase I is a big

Distamycin

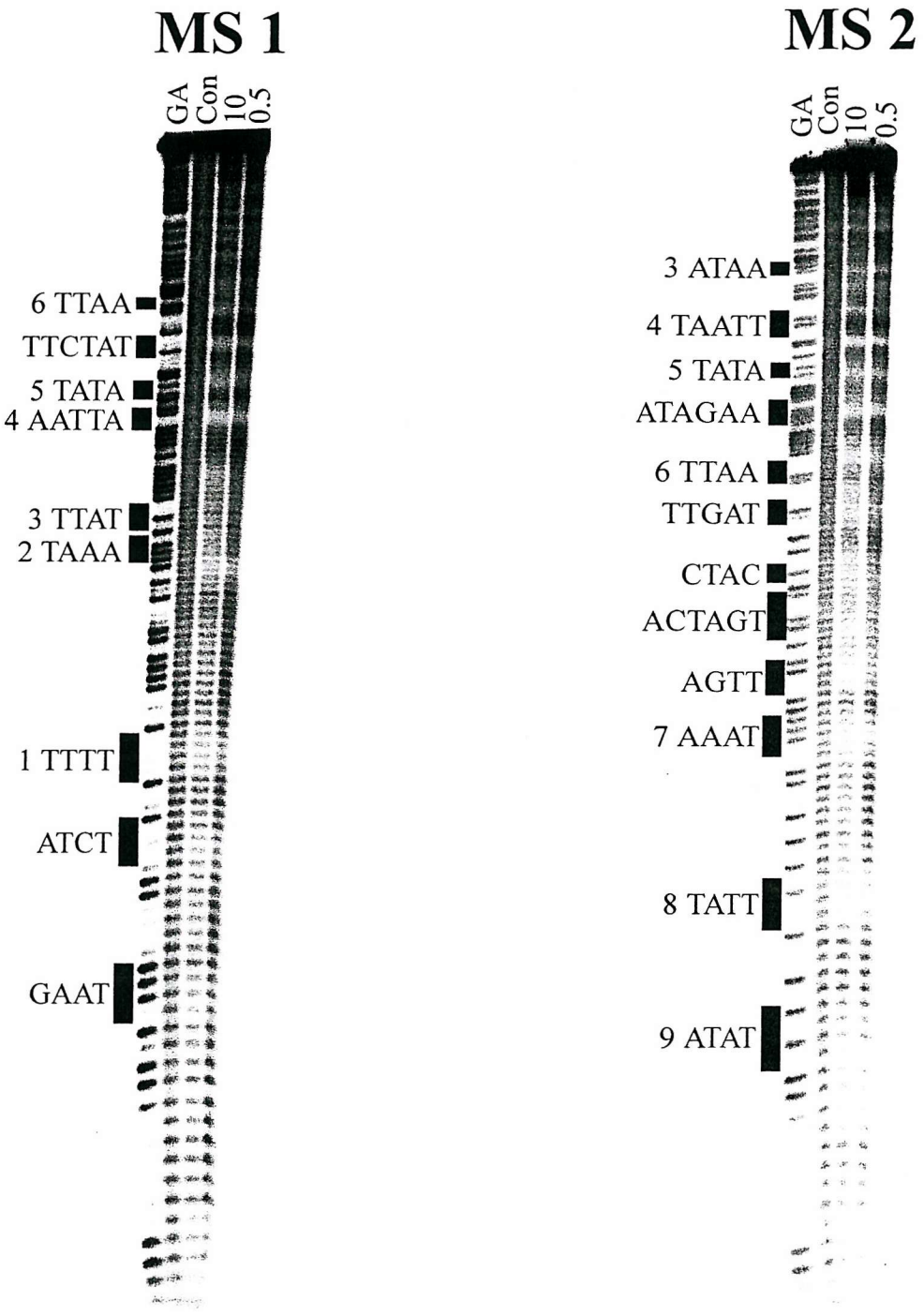


Figure 5.13 Hydroxyl radical footprinting patterns for MS 1 and MS 2 in the presence and absence of distamycin. Ligand concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each AT track step on the fragment. The track labeled GA is a marker specific for guanine and adenine. The track labeled 'Con' is the control which shows the digestion in the absence of ligand.

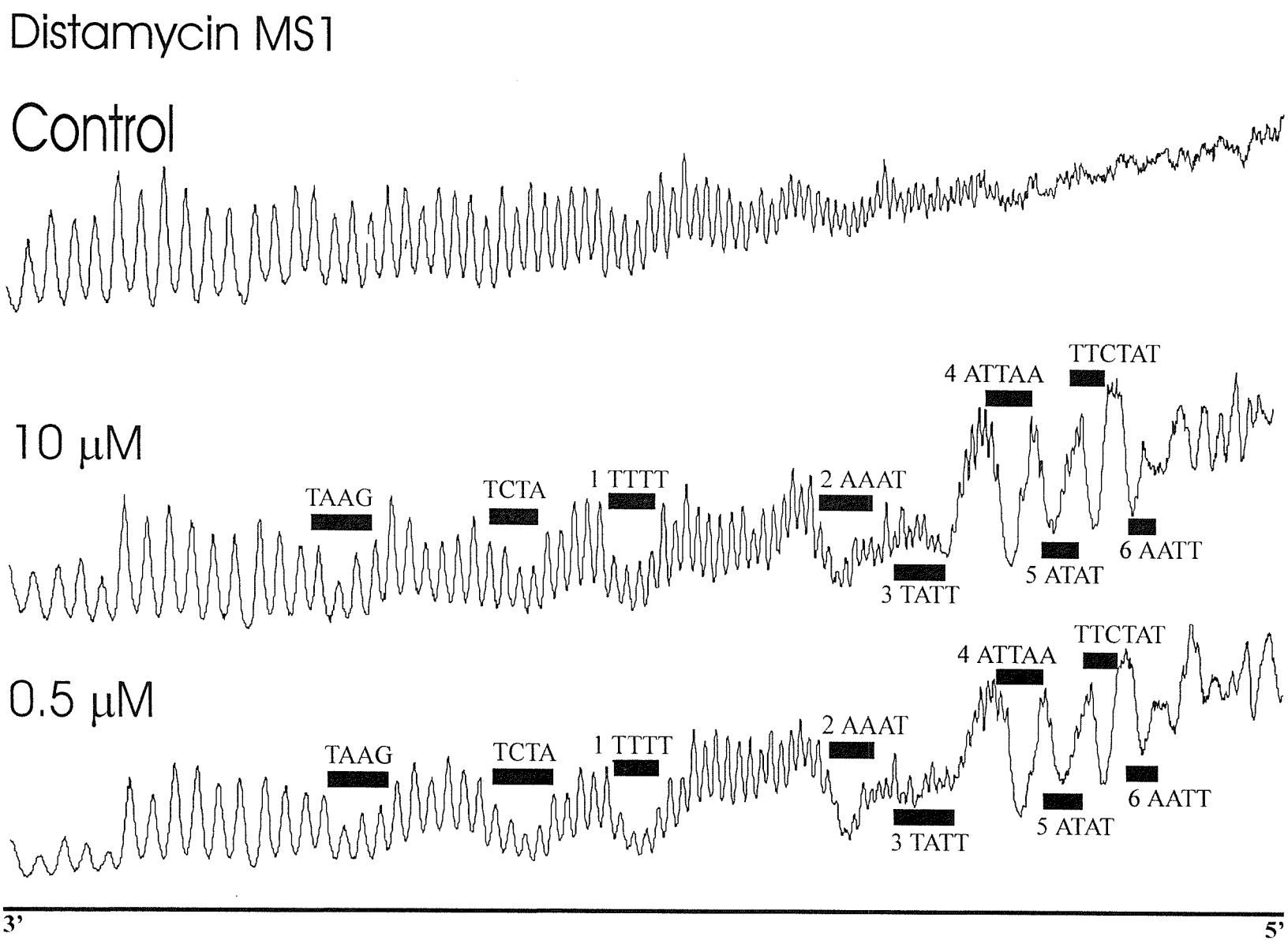


Figure 5.14 Intensity plots for hydroxyl radical cleavage of fragment MS 1 in the absence and presence of distamycin. The traces were obtained from analysis of the phosphorimages presented in figure 5.13, using ImageQuant software the trace and sequences run from 3'-5', reading from left to right, so that the right hand end corresponds to the top of the autoradiograph.

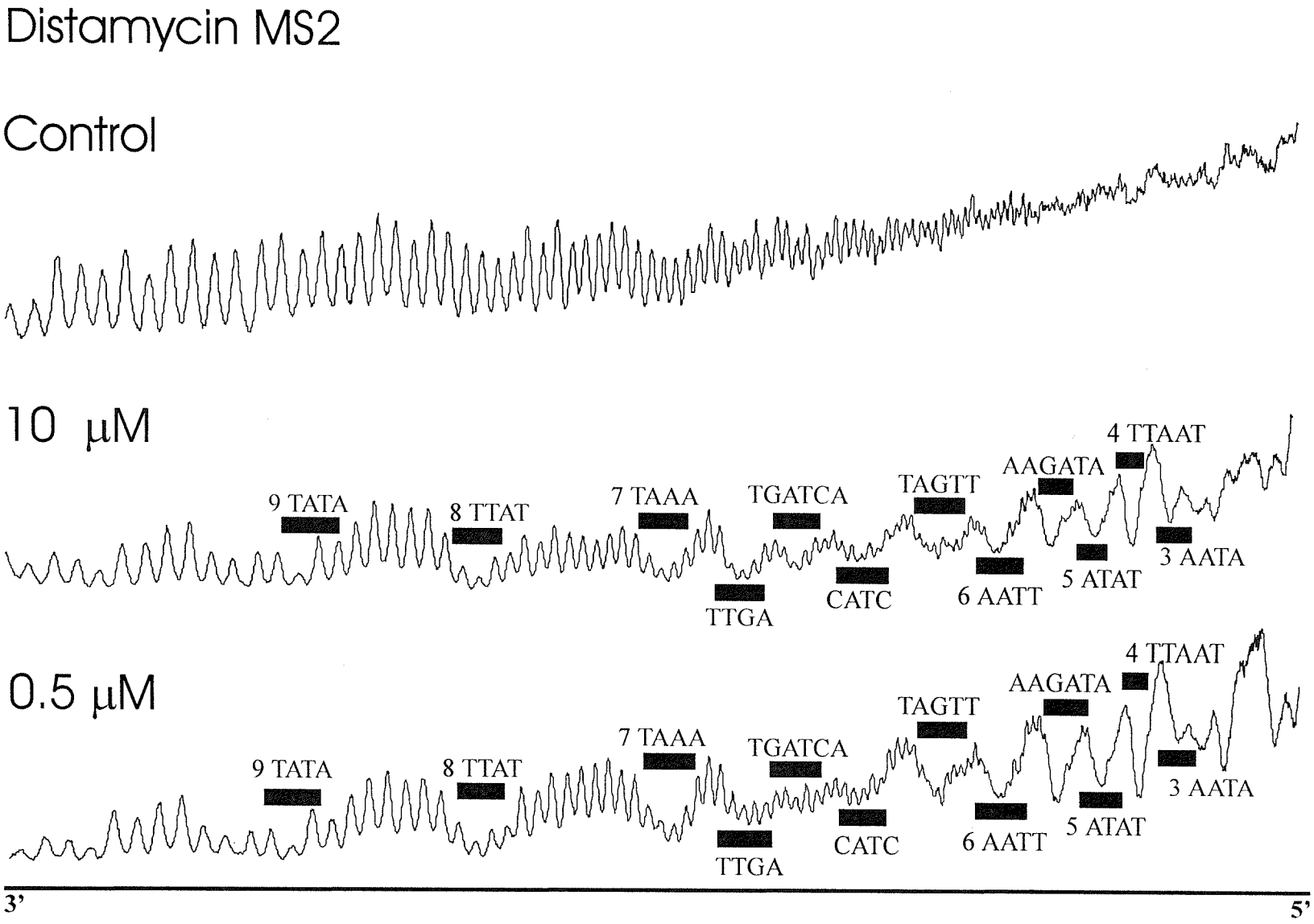


Figure 5.15 Intensity plots for hydroxyl radical cleavage of fragment MS2 in the absence and presence of distamycin. The traces were obtained from analysis of the phosphorimages presented in figure 5.13, using ImageQuant software the trace and sequences run from 3'-5', reading from left to right, so that the right hand end corresponds to the top of the autoradiograph

molecule (M_r ~ 31,000; 40 Å diameter)(Drew and Travers, 1984) while hydroxyl radicals are small molecules which do not perturb the DNA structure. Hydroxyl radicals are uncharged but heavily reactive and they extract the hydrogen atom from the deoxyribose in the backbone (Tullius, 1987), therefore yielding a more accurate size and position of the binding sites.

5.6.2 Summary

DNase I footprinting of the MS1 and MS2 fragments in the presence and absence of distamycin show that this agent prefers to bind to regions of high AT content. The site with the highest affinity is AAAA followed by AATT. In addition this agent binds to sites that contain a CG base pair in the middle or on the flanking sides of the site, thus, corroborating the results obtained by Churchill *et al*, (1990) and Van Dyke *et al.* (1982). These results are considered further in the Discussion

5.7 Hoechst 33258

This ligand also prefers regions which are high in A and T content (section 1.3.2.1). DNase I footprinting patterns for this ligand on MS1 and MS2 are shown in Fig 5.16. It can be seen that most of the AT regions are protected by drug binding. However it is worth noting that even though the sequence recognition properties of Hoechst 33258 are similar to distamycin, the footprinting gels are very different for these two ligands (compare Fig. 5.16 with 5.11). In general the footprints with Hoechst 33258 are more discrete, and are found in AT-containing sites. Footprinting plots derived from these data are shown in Fig.5.17 and the C_{50} values are shown below (table 5.5).

Many of the sites have very good affinity for Hoechst 33258 and their C_{50} values are too low to be measured accurately. These low C_{50} values were expected since the footprints at these sites persist even at the lowest concentrations. This problem can not be solved by using ever lower ligand concentrations as we estimate the concentration of DNA substrate to be about 1 nM.

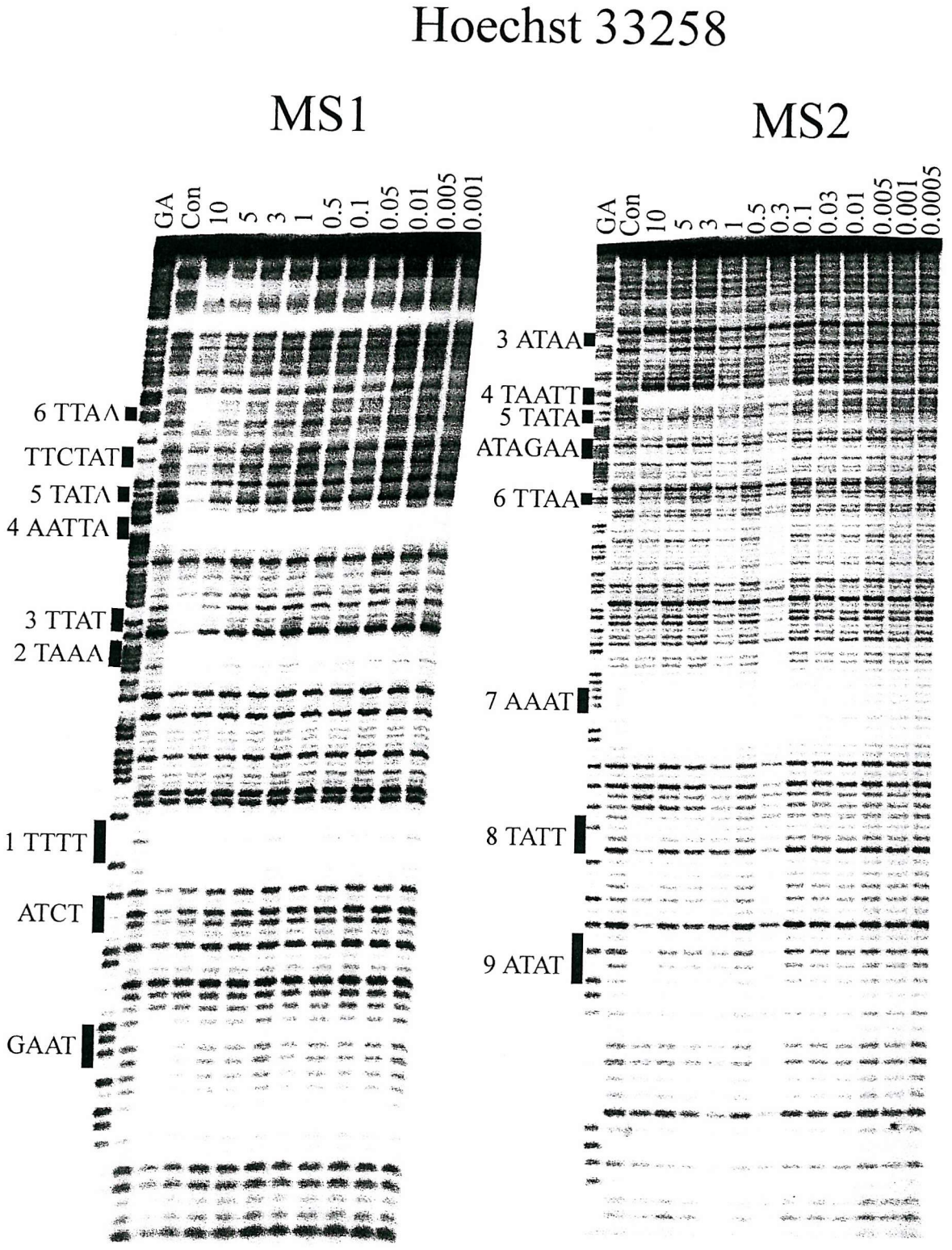


Figure 5.16 DNase I footprinting of MS1 and MS2 in the presence and absence of Hoechst 33258. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each AT track on the fragment. The track labeled GA is a marker specific for guanine and adenine. The track labeled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labeled at the 3'-end.

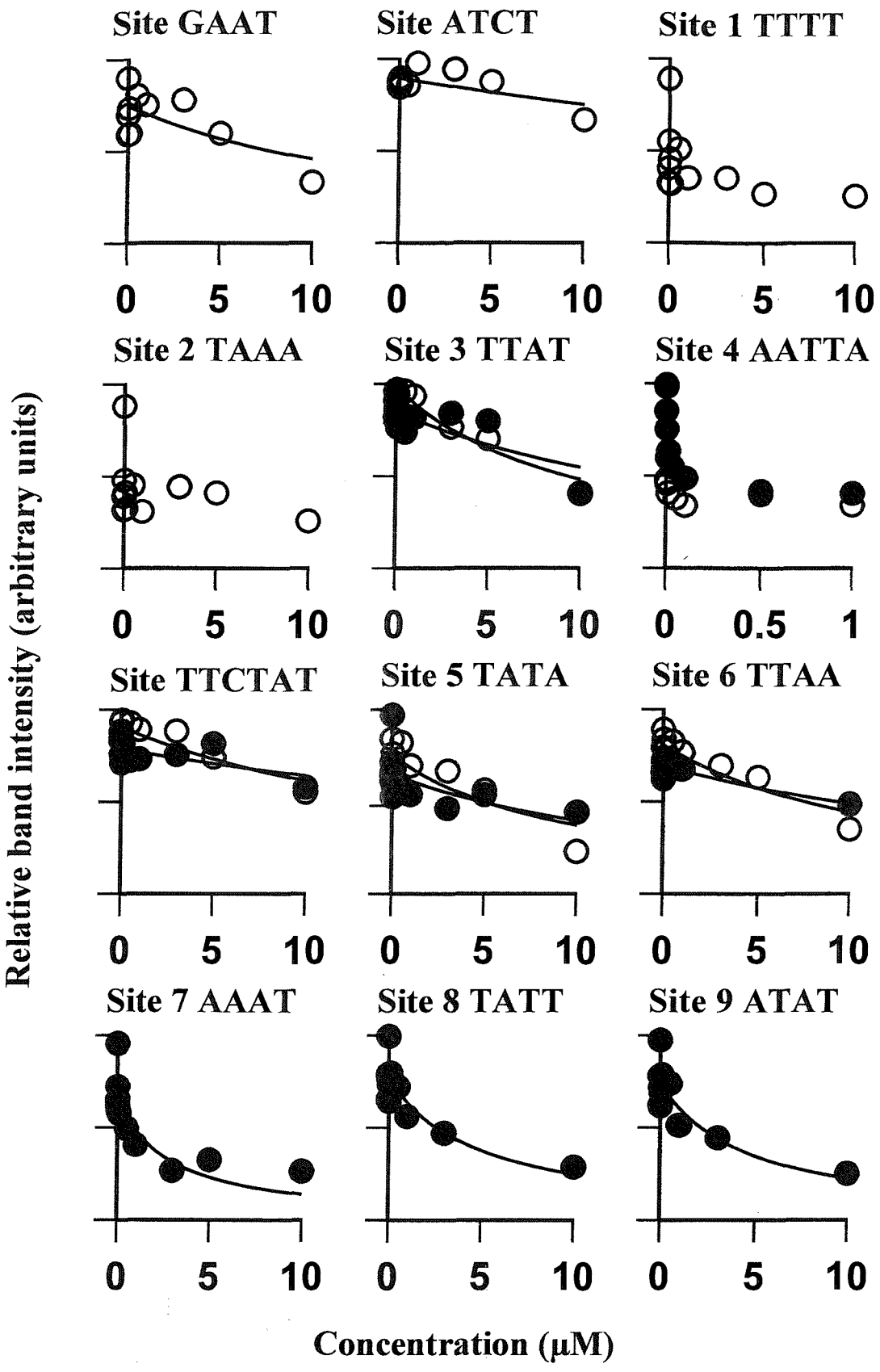


Figure 5.17 Footprinting plots showing the interaction of Hoechst 33258 with the AT tracts in DNA fragments MS1 and MS2, O corresponds to data for MS1 and ●are taken from MS2. Data was analysed as described in section 2.24.

Site	Sequence	C_{50} MS1	C_{50} MS2
	GAAT/ATTC	16 ± 9	Not resolved
	ATCT/AGAT	52 ± 25	Not resolved
1	TTTT/AAAA	Too low to measure	Not resolved
2	TAAA/TTTA	Too low to measure	Not resolved
3	TTAT/ATAA	9 ± 1	18 ± 7
4	AATTA/TAATT	Too low to measure	Too low to measure
5	TATA	10 ± 3	16 ± 10
	TTCTAT/ATAGAA	21 ± 6	45 ± 23
6	TTAA	12 ± 3	7 ± 3
7	ATTT/AAAT	Not resolved	3 ± 1
8	AATA/TATT	Not resolved	5 ± 2
9	ATAT	Not resolved	4 ± 2

Table 5.5 C_{50} values (μM) determined for the interaction of Hoechst 33258 with the various AT sites on fragments MS1 and MS2. The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 5.17.

The ranking order of these sites could be compared with those found by other groups. For example Abu-Daya *et al.* (1995) found the following ranking AATT > AAAA > ATTA / TAAT = TTAA = ATAT > TATA on DNase I footprinting experiments and Drobyshev *et al.* (1999) obtained AATT > AAAT > AAAA > AATA > TAAT > ATAT > TAAA > ATAA > TTAA > TATA using a combinatorial method with a microchip (section 1.3.2.1). Our results could be ranked as TTTT, TAAA, AATTA, TTAT > ATTT > ATAT > AATA > TTAA ≥TATA. All these experiments seem to suggest that AATT and TTTT are the best binding sites for this agent while TTAA and TATA are among the worst. Also, this agent seems to be able to bind to sites that contain a GC base pair situated in the middle or at the sides of a binding site, just like distamycin. These sites are GAAT, ATCT and TTCTAT. We will see in the next section that there are some more sites, with a CG base pair in the middle, to which this agent is binding.

5.7.1 Hydroxyl radical footprinting of Hoechst 33258

Hydroxyl radicals are a useful probe for determining the exact location of the ligand within the fragment. The results obtained with hydroxyl radical digestion of the fragments MS1 and MS2 in the presence and absence of Hoechst 33258 are presented in Fig 5.18 Direct inspection of these gels shows that most of the attenuations occur within the AT regions. Densitometer traces for these data are shown in Fig 5.19 for MS1 and 5.20 for MS2. These plots confirm that all the regions of high AT content produce a footprint. The MS2 results (Fig 5.20) reveal several sites that contain a CG base pair in the middle or at the sides of the binding site. Some of these sites are AGTT, ACTAGT, CTAC, TTGAT, and TTCA. It is possible that Hoechst binds to these sites like distamycin, as a minor groove binding agent, or by intercalation, as this agent is known to intercalate in areas which contain CG base pairs (Bailly *et al.*, 1993). However, only three of these sites (GAAT, ATCT and TTCTAT) seem to show footprints on the DNaseI footprinting gels (Fig 5.16), thus suggesting that these sites have very low affinity for this agent as can be seen in table 5.5. Of these sites GAAT seems to be the best with a C_{50} value of 16 ± 9 μM. Also if we compare the footprint size of the DNase I gels (Fig 5.16) with those of the hydroxyl radical gels (Fig 5.18) we can see that the footprints from the DNase I gels are longer than those of the hydroxyl radicals gels. The reasons for these differences were explained in section 5.6.1.

5.7.2 Summary

The DNase I footprinting gels for this agent (Fig 5.16) show that Hoechst has high affinity for AATT and TTTT and low affinity for TTAA and TATA. These gels also show that Hoechst is able to bind to sites that contain a CG base pair in the middle or at the sides of the binding region. These sites were recognise as being GAAT, ATCT and TTCTAT. These results will be considered further in the Discussion.

5.8 Mithramycin

DNase I footprinting gels for this agent are presented in Fig 5.21. This ligand has previously been shown to bind to GC-rich sequences with better binding to regions

Hoechst 33258

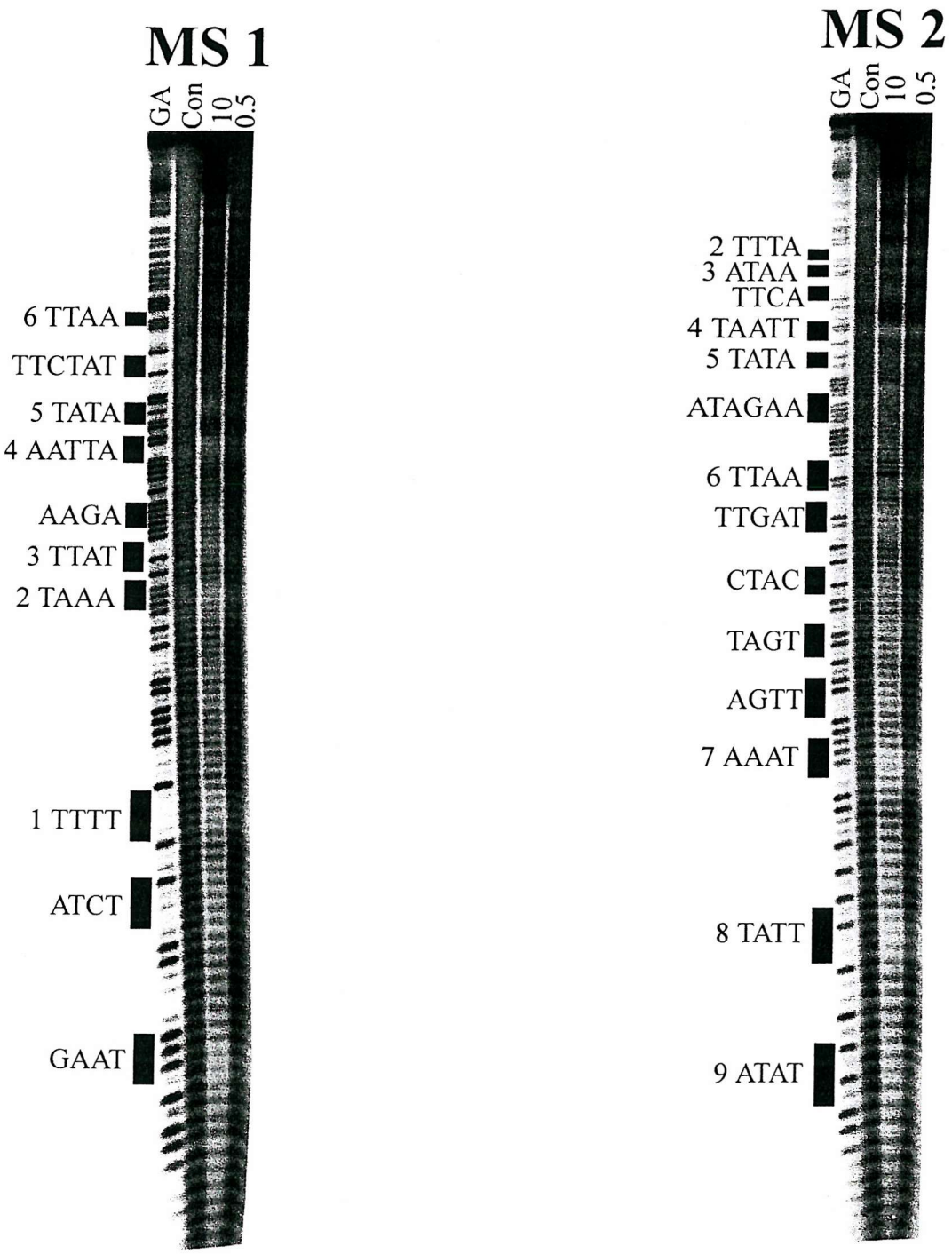


Figure 5.18 Hydroxyl radical footprinting patterns for MS 1 and MS 2 in the presence and absence of Hoechst 33258. Ligand concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each AT track step on the fragment. The track labeled GA is a marker specific for guanine and adenine, and the track labeled 'Con' is the control which shows the digestion in the absence of ligand.

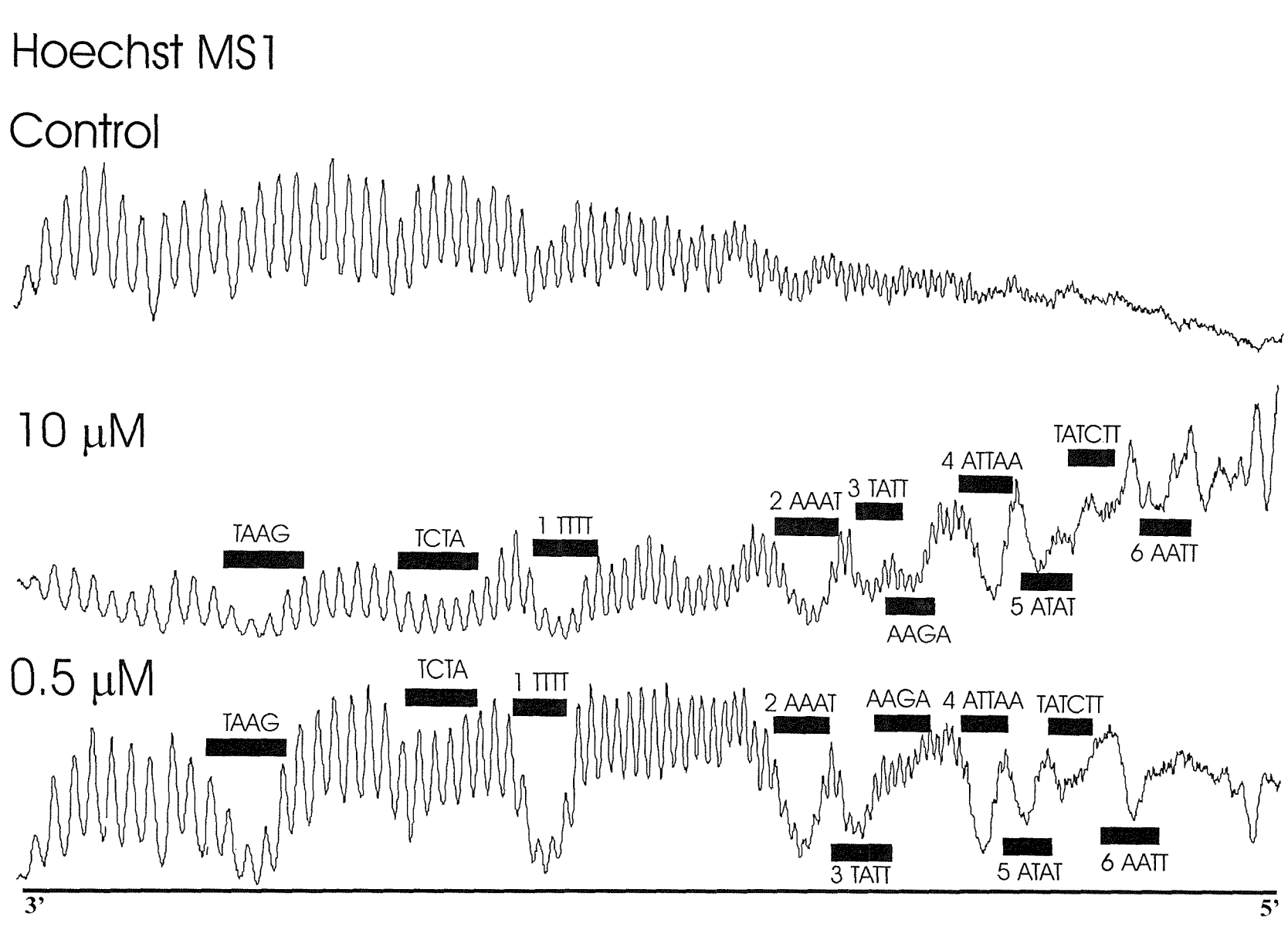


Figure 5.19 Intensity plots for hydroxyl radical cleavage of fragment MS 1 in the absence and presence of Hoechst 33258 The traces were obtained from analysis of the phosphorimages presented in figure 5.20, using ImageQuant software. The trace and sequences run from 3'-5', reading from left to right, so that the right hand end corresponds to the top of the autoradiograph.

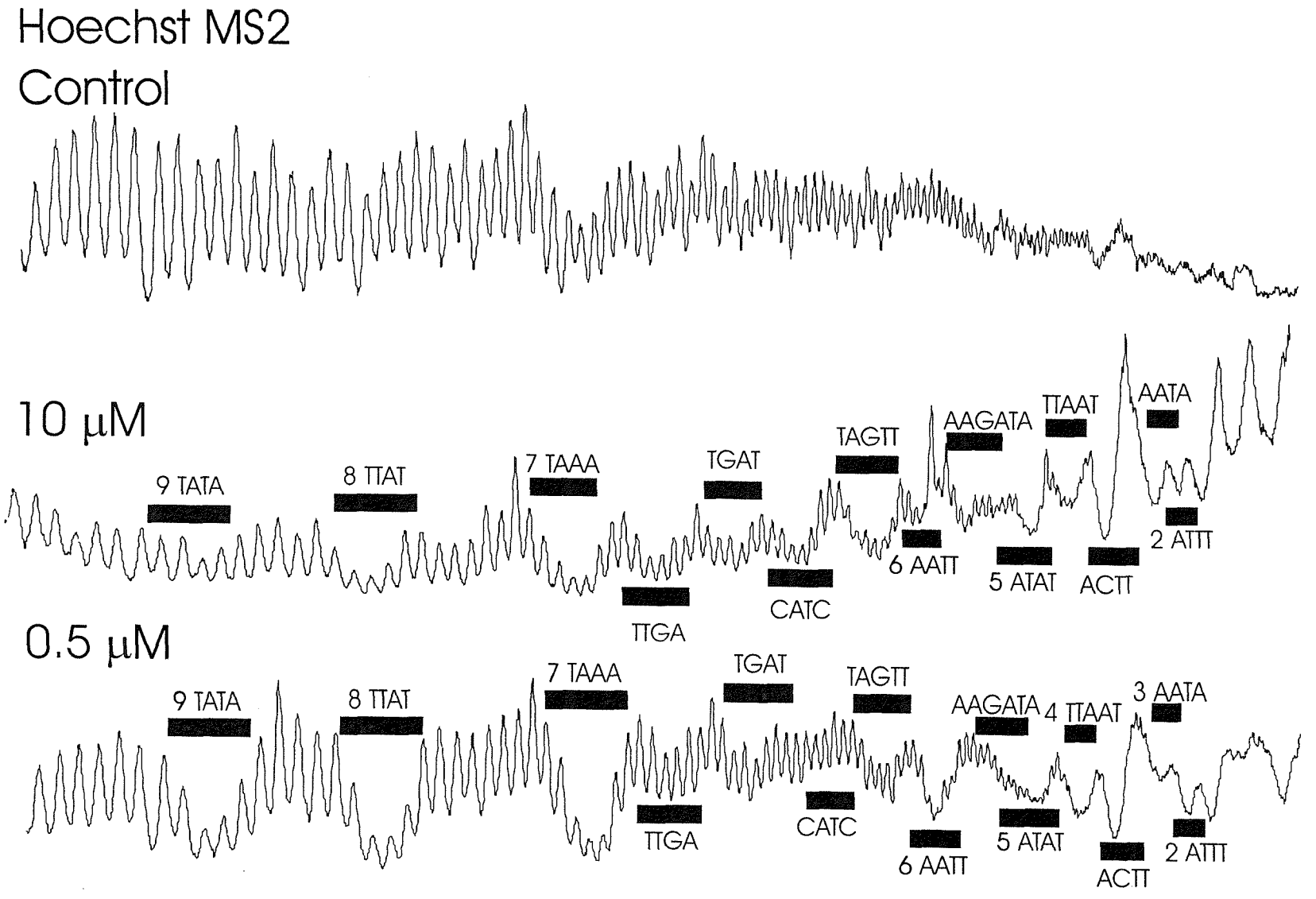


Figure 5.20 Intensity plots for hydroxyl radical cleavage of fragment MS 2 in the absence and presence of Hoechst 33258 The traces were obtained from analysis of the phosphorimages presented in figure 5.20, using ImageQuant software. The trace and sequences run from 3'-5', reading from left to right, so that the right hand end corresponds to the top of the autoradiograph.

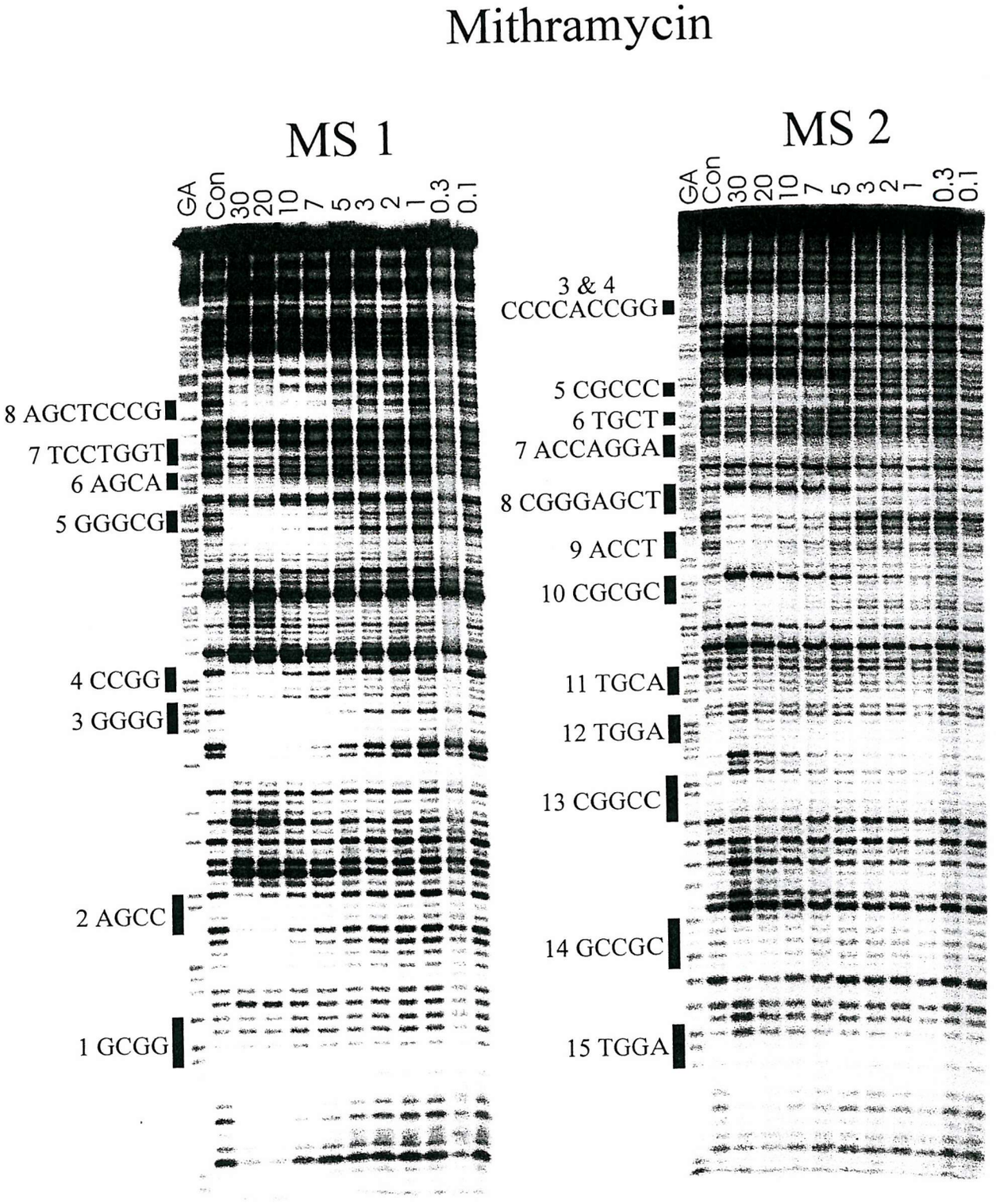


Figure 5.21 DNase I footprinting of MS1 and MS2 in the presence and absence of mythramycin. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each GC tracks on the fragment. The track labeled GA is a marker specific for guanine and adenine. The track labeled 'Con' is the control which shows the digestion in the absence of ligand.

containing 5'-GG, 5'-CC and 5'-GC than 5'-CG (section 1.3.3.2). Visual inspection of the footprinting gels reveals that the ligand produces clear footprints at each of the GC containing regions. The footprinting plots showing the binding strengths of the different sites are presented in Fig 5.22 and the C_{50} values obtained are shown in table 5.6

		C_{50}	
Site	**Sequence**	**MS1**	**MS2**
1	GCGG/CCGC	8 ± 4	Not resolved
2	AGCC/GGCT	7 ± 2	Not resolved
3	GGGG/CCCC	4 ± 1	Not resolved
4	CCGG	13 ± 4	Not resolved
5	GGGCG/CGCCC	5 ± 1	12 ± 4
6	AGCA/TGCT	17 ± 10	12 ± 3
7	TCCTGGT/ACCAGGA	13 ± 6	29 ± 8
8	AGCTCCCG/CGGGAGCT	3 ± 1	6 ± 1
9	AGGT/ACCT	Not resolved	7 ± 2
10	GCGCG/CGCGC	Not resolved	5 ± 2
11	TGCA	Not resolved	No footprint
12	TCCA/TGGA	Not resolved	No footprint
13	GGCCG/CGGCC	Not resolved	19 ± 10
14	GCGGC/GCCGC	Not resolved	21 ± 11
15	TCCA/TGGA	Not resolved	No footprint

Table 5.6 C_{50} values (μM) determined for the interaction of mithramycin with the various GC tracks on the fragments MS1 and MS2. The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 5.22

The best binding sites are site 8 AGCTCCCG followed by site 3 GGGG, site 5 GGGCG and site 10 CGCGC. The poor binding observed at site 13 (CGGCC) was unexpected as GGCC was found to be the best binding site in a study by Liu and Chen (1994). Inspection of the footprinting gel for MS2 (Fig 5.21) reveals a small enhancement at this site in contrast to the expected footprint. This may suggests some interaction of the

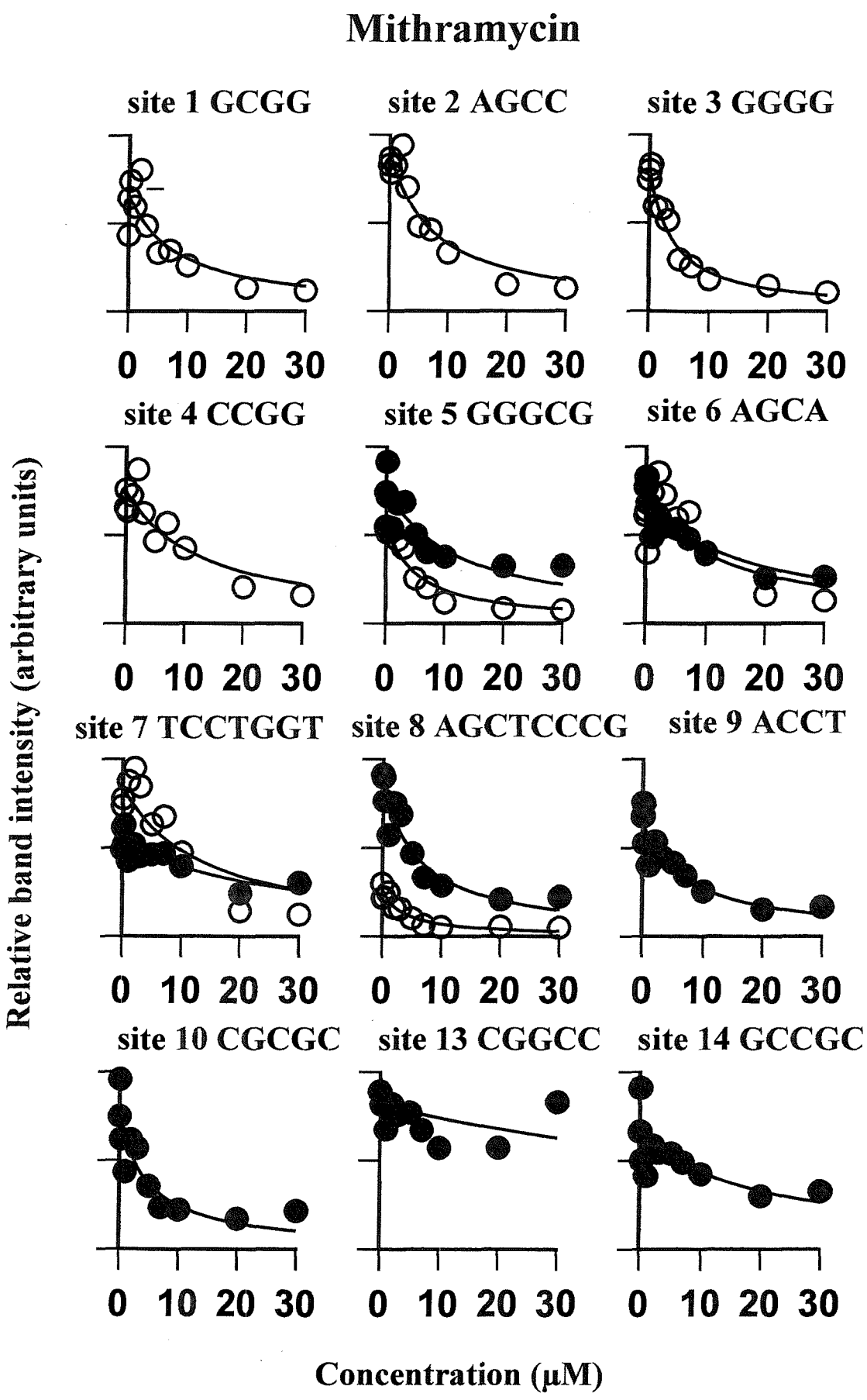


Figure 5.22 Footprinting plots showing the interaction of mithramycin with the regions containing 5'-GG, 5'-GC and 5'-CC in DNA fragments MS1 and MS2, ○ corresponds to data for MS1 and ●are taken from MS2. Data was analysed as described in section 2.24.

ligand with this site, though with low affinity. It is possible that the binding to this site is reduced by the presence of the CpG at the 5'-end.

5.8.1 Hydroxyl radical footprinting of mithramycin

This chemical reagent was used to better determine the binding positions of mithramycin. Figure 5.23 shows the hydroxyl radical footprinting results for MS1 and MS2 obtained in the presence and absence of mithramycin. It can be seen that all the regions containing high CG content yield an attenuated cleavage. Densitometer traces derived from these gels are shown in Fig 5.24 for MS1 and Fig.5.25 for MS2. The gel for MS2 (Fig 5.23) confirms that mithramycin is not binding to site13 (CGGCC). Also if we compare the footprint size of the DNase I gels (Fig 5.21) with those of the hydroxyl radical gels (Fig 5.23) we can see that the footprints from the DNase I gels are longer than those of the hydroxyl radicals gels. The reason for these differences are explained in section 5.6.1.

5.8.2 Summary

Footprinting of MS1 and MS2 gels with DNase I confirms that mithramycin binds to regions of high GC content preferring sequences that contain 5'-GG, 5'-CC, and 5'-GC. However it seems that the presence of a CpG step within the site causes a decrease in affinity, possibly explaining the very low affinity observed at CGGCC (site 13). These results are considered further in the Discussion.

5.9 Discussion

The MS1 and MS2 fragments were designed to contain all the 136 tetranucleotide sequences, and have been used successfully for identifying the binding sites for echinomycin, [*N*-MeCys^3, *N*-MeCys^7]TANDEM, actinomycin D, distamycin, Hoechst 33258 and mithramycin. These fragments demonstrate that in each case the identity of the bases surrounding the binding site influences the affinity of the ligand. Therefore, these fragments could be used in the identification of the preferred binding sites for novel ligands of unknown specificity. They can also provide a comparison between dinucleotide steps

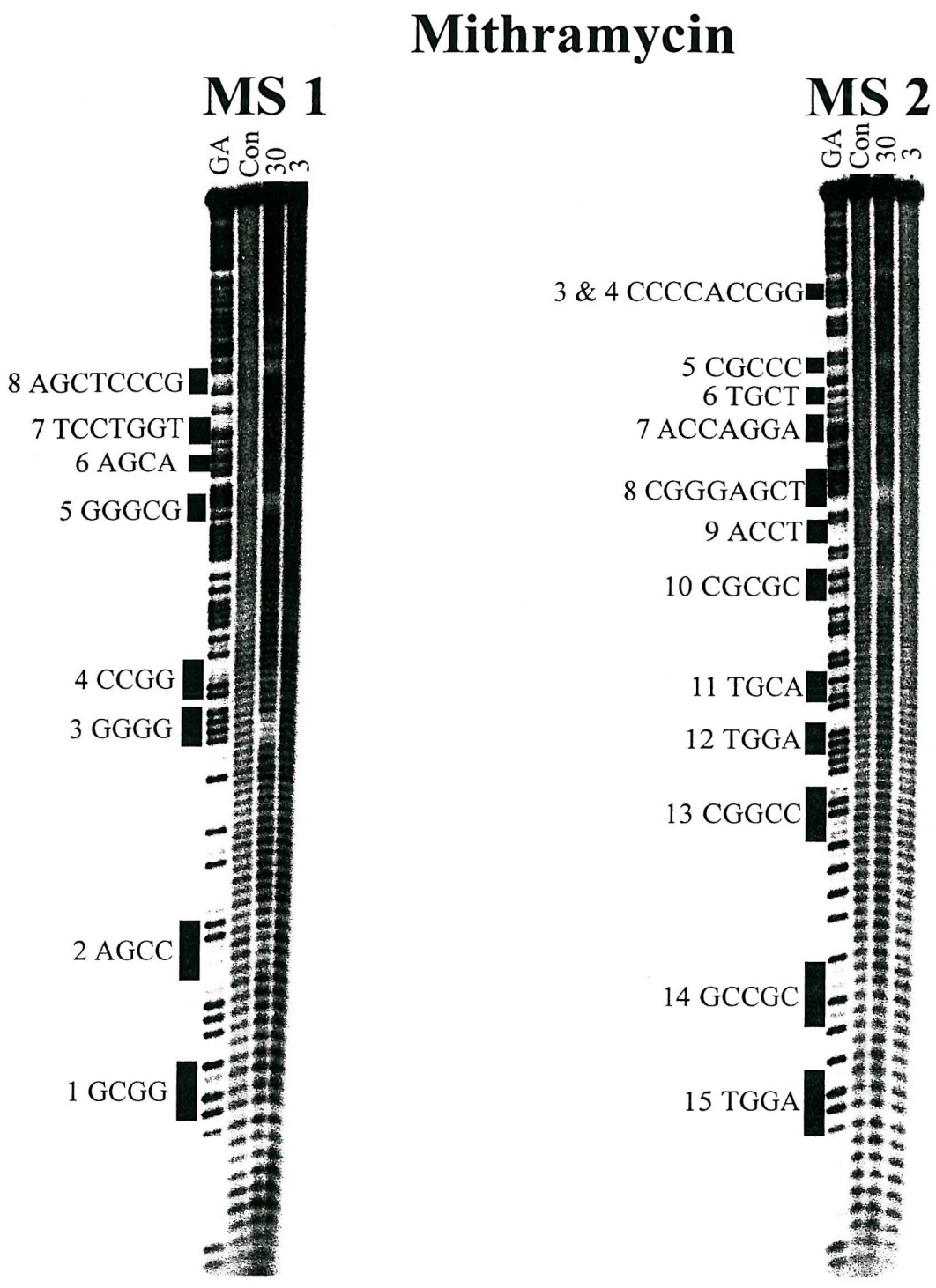


Figure 5.23 Hydroxyl radical footprinting patterns for MS 1 and MS 2 in the presence and absence of mithramycin. Ligand concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the position of each GC track step on the fragment. The track labeled GA is a marker specific for guanine and adenine. The track labeled 'Con' is the control which shows the digestion in the absence of ligand.

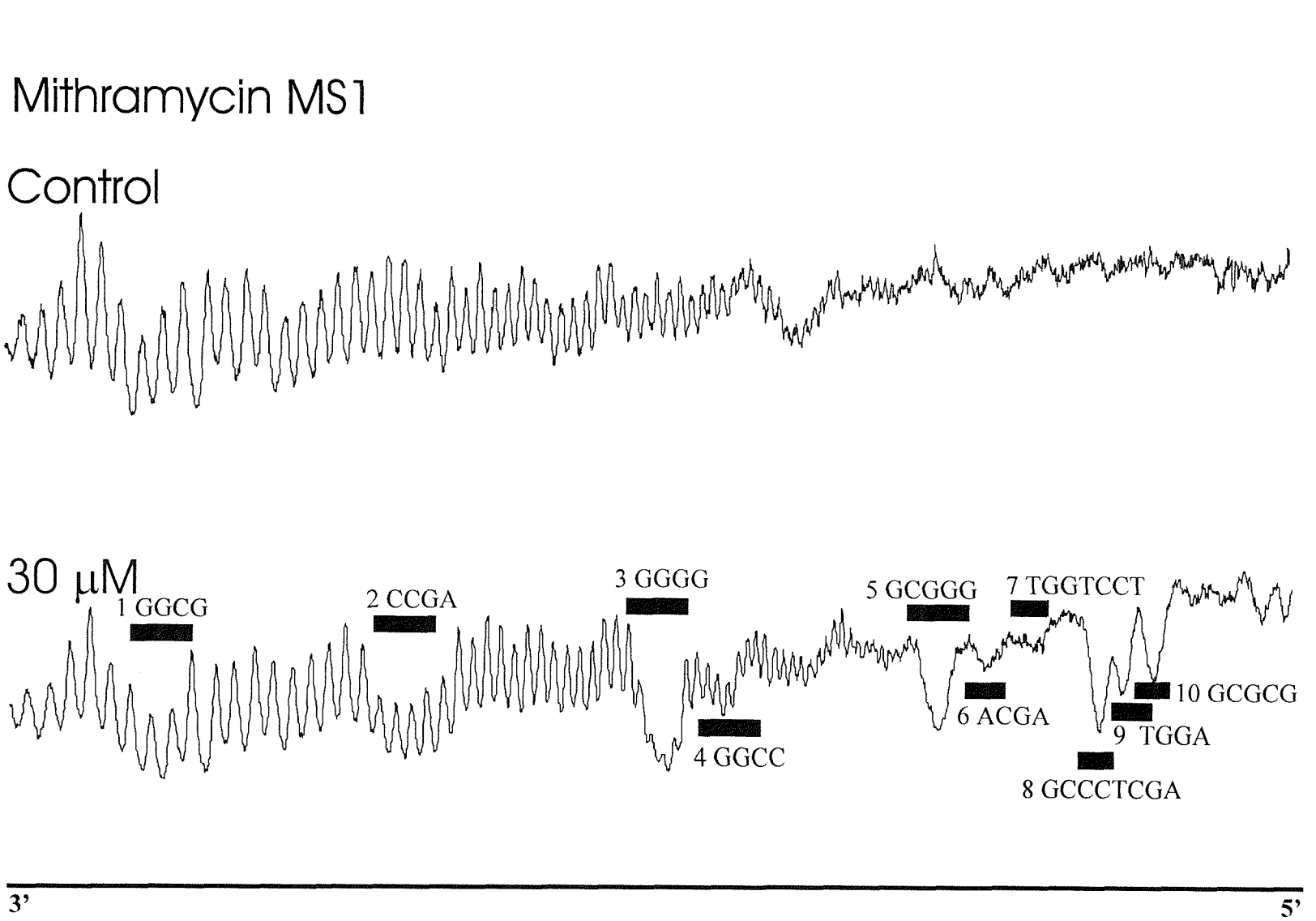


Figure 5.24 Intensity plots for hydroxyl radical cleavage of fragment MS 1 in the absence and presence of mythraycin The traces were obtained from analysis of the phosphorimages presented in figure 5.27, using ImageQuant software. The trace and sequences run from 3'-5', reading from left to right, so that the right hand end corresponds to the top of the autoradiograph.

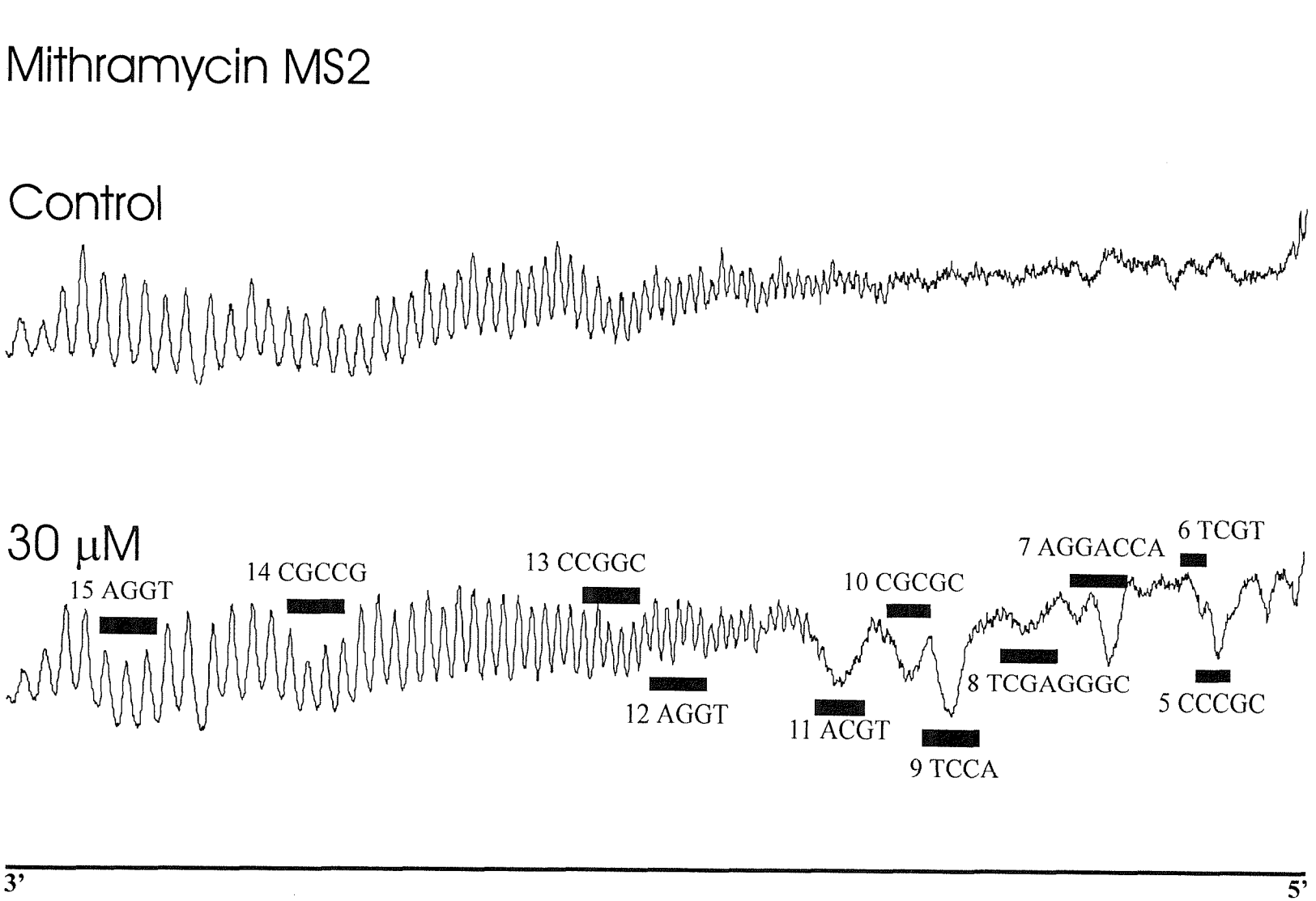


Figure 5.25 Intensity plots for hydroxyl radical cleavage of fragment MS 2 in the absence and presence of mythramycin The traces were obtained from analysis of the phosphorimages presented in figure 5.27, using ImageQuant software. The trace and sequences run from 3'-5', reading from left to right, so that the right hand end corresponds to the top of the autoradiograph

flanked by different sequences as in the present study. However, MS1 and MS2 still have some limitations, in particular it is sometimes not possible to resolve all the binding sites. Also there are binding regions that contain several tetranucleotides within the same site, for instance AATTA which is site 4 for Hoechst 33258 and for distamycin (table 5.4 and table 5.5) . This site contains two tetranucleotides (AATT and ATTA) to which these ligands bind, and is impossible to determine which one of them has the higher affinity for the ligand. In addition there is also the possibility that there are cooperative interactions between ligands which are bound to adjacent sites such as AGCGCG which are sites 6 and 7 for actinomycin D (table 5.3), thus affecting the value of the binding strength. In some of these cases the upper bands were used for analysis of the top binding site and the lower ones for the bottom site or the whole site was considered as a single site.

5.9.1 Echinomycin

The results presented in this chapter demonstrate that echinomycin does bind to all CpG sites and that the affinity of the binding is affected by the flanking regions. However, examination of table 5.1 reveals that the binding is not determined by the order of the pyrimidines or purines around the binding site It is known that intercalation into RpY steps is poorer than into YpR (Krugh and Reinhardt, 1975). Therefore, one would predict that sites containing RCGY would be worse than sites containing YCGR. However, the results show that ACGT (site 4) is the best binding site and ACGC (site 5) and GCGT (site 5) are intermediate, while CCGG (site 3) and TCGA (site 2) are good binding sites and CCGA (site 10) is among the worst sites. The only tendency that we notice is that in general it appears that the weakest sites are flanked by G or C bases while the strongest sites are flanked by A or T on both sides. In addition, some of our results are corroborated by a study which measured the dissociation kinetics of echinomycin from CpG binding sites (Fletcher and Fox, 1996). They found that echinomycin dissociated slower from ACGT (site 4) than from TCGA (site 2). One possible explanation for this difference is shown by an NMR study which demonstrated that the AT bases surrounding ACGT can adopt a Hoogsteen configuration while those flanking TCGA do not (Gao and Patel, 1988). Although the formation of Hoogsteen base pairs is thought not to occur on longer DNA molecules in

solution, these unusual conformations demonstrate the intrinsic moulding capability of each sequence, which may help or prevent echinomycin binding.

The footprinting gels for MS1 and MS2 also showed a new binding site that did not contain a CpG. The identity of this site was determined to be ACCA, after considering all the results obtained by DNase I footprinting and DEPC modification patterns of MS1, MS2 and E8.(Fig 5.2, Fig 5.5, Fig 5.6). There are previous studies which support our findings that echinomycin can bind to CpC (GpG) as a secondary site (Fox *et al.*, 1991). This interaction may occur through the formation of a specific hydrogen bond to the second guanine with the other symmetrically disposed bonding site left unoccupied (Fox *et al.*, 1991). This result leaded us to investigate other tetranucleotide sequences that contained CpC (GpG) in the middle. Unfortunately, most of these sites contained an adjacent CpG step and the only sites that could be examined were sites GGGG, ACCT and AGGT which did not produce a footprint.

The precise location of the binding sites could not be further investigated using hydroxyl radicals, because this ligand is dissolved in DMSO which quenches the free radical reaction.

5.9.2 [*N*-MeCys^3, *N*-MeCys^7]TANDEM

The results presented in this chapter demonstrate that, although [*N*-MeCys^3, *N*-MeCys^7]TANDEM can bind to all sites containing TpA steps, the affinity is affected by the flanking sequences. We have shown that the best tetranucleotide binding site for [*N*-MeCys^3, *N*-MeCys^7]TANDEM contains the sequence ATAT (site 12), consistent with the previous observation that it binds best to TpA steps which are flanked by A and T residues (Low *et al.*, 1984; Low *et al.*, 1986; Fletcher *et al.*, 1995). However, the binding is strongly influenced by the identity of the surrounding bases and all TpA steps do not constitute good binding sites. This is not unexpected since other studies have shown that TANDEM does not bind to IpC, for which the hydrogen bonding face exposed to the minor groove is identical to TpA (Bailly and Waring, 1998). It has therefore been suggested that the

selectivity is not determined by hydrogen bond formation but may originate from steric and/or hydrophobic interactions with a minor groove of suitable dimensions (Bailly and Waring, 1998).

Several of the TpA steps do not generate good binding sites. This is specially noteworthy for TTAA (site 8) which is the poorest binding site, demonstrating that TpA flanked by A and T does not necessarily constitute a good binding site. Examination of the C_{50} values in table 5.2 reveals that the binding is also not determined by the order of purines and pyrimidines around the binding site. One might predict that RTAY would be poor binding sites as each quinoxaline chromophore would be required to intercalate into an RpY step and these are known to be poorer intercalation sites than YpR (Krugh and Reinhardt, 1975). However of these sequences GTAC (site1) is one of the worst binding sites, ATAT (site 12) is the best and GTAT (site 11) has intermediate affinity. Similarly YTAR sites have variable affinities with TTAA (site 8) as a poor site and CTAG (site 10) among the best. The only trend that we note in these binding data is that sequences of the type ATAX/XTAT are always among the best binding sites, consistent with the original observation that the ligand binds very tightly to poly(dA-dT) (Lee and Waring, 1978).

Hydroxyl radical footprinting was not performed for this agent as it is dissolved in DMSO, which inhibits the free radical reaction.

5.9.3 Actinomycin D

This agent binds to all GpC containing sites (Fox and Waring 1984). However, the results show that the affinity for this dinucleotide is affected by the nature of the flanking regions. A close examination of table 5.3 shows that the binding strength of this ligand could be determined by the positions of purines and pyrimidines around the binding site. For example sites containing YGCR such as CGCA, TGCA and CGCG have intermediate to strong affinities for this ligand, while sites containing RGCY like AGCT, AGCC have intermediate to weak affinity. In addition GGCC, which is also RGCY, shows no footprint.

A previous publication has shown that the binding affinity of actinomycin D is weak at sites containing the sequence GGCX/XGCC with poor interaction at GGCC (Chen, 1989). Our data confirm that the interaction with GGCC (site 9) is extremely weak and that most sites containing the sequence GGCX are indeed weak, the only exception is site GGCA (site 10) which shows a very high affinity for this ligand (table 5.3). In addition another study suggested that site TGCA has stronger affinity for actinomycin binding than AGCT (Fletcher and Fox, 1995). In contrasts our results (table 5.3) reveal that the difference in affinity between these two sites is not very large.

DNase I footprinting gels (Fig 5.9) show two binding sites that do not contain a GpC step. These sites seem to be within the sequences CCGGTGGGG and TCCTGGT, and show a weak affinity for actinomycin D, although higher than GGCC (table 5.3). There are several studies that suggests that actinomycin D is able to bind to other dinucleotides besides GpC, (Watrerloh and Fox, 1992; Krugh , 1972). Some footprinting studies have shown that actinomycin binds better to GG (CC) than to GT (AC) and GA(TC) (Waterloh and Fox, 1992; Scamov and Beabealashvilli, 1983). In additon, there is an NMR study that suggest that actinomycin D binds to CpG by the same mechanism that it binds to GpC, i.e. by intercalation. However, the binding to CpG is not as strong as with GpC (Zhou *et al.*, 1989). Another group has studied the binding of actinomycin to $T(G)_nT$ and found that the affinity of this ligand decreased as follows TGGGT > TGGT > TGGGGT > TGT. (Bailey *et al.*, 1994). Therefore, considering all these information we could predict that actinomycin D is probably binding to CpG, GpG or TpG on the sequence CCGGTGGGG, and to TpG or GpG on the sequence TCCTGGT. To determine the exact locations of the binding sites would require further investigations. However, hydroxyl radical cleavage cannot be used for the determination of these sites, as this ligand quenches the free radical reaction (Churchill *et al.*, 1990).

5.9.4 Distamycin

The data obtained for distamycin shows that this ligand binding is also affected by the surrounding bases. This ligand binds preferentially to regions of high A and T content

(Fox and Waring, 1984;Churchill *et al.*, 1990). The best tetranucleotide sites for distamycin binding are AAAA and AATTA,. Sites with intermediate affinity are ATAT and TATA and the worst site is TTAA (table 5.4). These findings are in agreement with previous observations obtained by footprinting studies (Abu-Daya *et al.*, 1995) and by circular dichroism (Chen and Sha, 1998). These preferences could be explained by the structure of the DNA at these sites. For example the minor groove of site AAAA tends to be narrower towards the 3' end (Burkhoff and Tullius, 1997). This narrowing of the minor groove may increase the affinity of distamycin because this antibiotic, like netropsin, narrows the minor groove on binding, thereby improving non-bonded interactions with the walls of the groove (Coll *et al.*, 1989). If the groove is structurally narrower before binding then less energy is required for this distortion. The next best site was AATTA which, if we follow the same principles, should be narrower on the middle of the site thus facilitating distamycin binding. The intermediate affinity binding sites ATAT and TATA contain a normal B-DNA minor groove, thus requiring more energy for distamycin binding. In addition, these sites may have a low propeller twist angle which could impede the formation of a narrow minor groove. This low propeller twist angle seems to be due to the presence of TpA (Coll *et al.*, 1989). That is possibly why ATAT is a better binding site than TATA and TTAA is the worst binding site.

Distamycin also bound to some sites that contained a CG base pair in the middle of the site or in the flanking regions. These sites were identified as TTCTAT, ACTAG and ATCTCA. The binding affinity for these sites is quite low but is higher than that of the TTAA site (table 5.4). Previous results have shown that distamycin is able to accommodate a CG base pair within its binding site (Churchill *et al.*, 1990; van Dyke *et al.*, 1982). Hydroxyl radical experiments have shown even more sites that contained a CG base pair in the middle (Fig 5.13), these sites were GAAT, TTGA, CATC and CTAGTT and were not very evident in the DNase I footprinting gels, indicating that their affinities are probably lower than that of the TTAA site.

5.9.5 Hoechst 33258

Examination of the gels (Fig. 5.18) and table 5.5 shows that the affinity of this ligand is also affected by the flanking regions. Hoechst also prefers to bind to regions of high A and T content which can be flanked by GC base pairs (Harshman and Dervan, 1985), just like distamycin. However, if we compare the gels obtained for distamycin (Fig 5.11) with those obtained for Hoechst 33258 (Fig 5.18) it can be seen that the pattern is different; this means that these agents differ in their binding preferences. Hoechst 33258, like distamycin, does not bind with high affinity to every region that is high in A and T content. The preferred tetranucleotide (by visual inspection of the gel Fig 5.18) sequences seem to be AATT followed by TTTT and TAAA and the worst sites are TTAA and TATA. Similar results were obtained by other studies (Abu-Daya et al. 1995; Drobyshev et al., 1999). In general it seems that Hoechst discriminates against TpA steps. This could be due to the low propeller twist angle within this step which obstructs the formation of a narrow minor groove (Coll *et al.*, 1989)

Hoechst 33258, like distamycin, seems to be able to bind to sites that contain a CG base pair within the middle of the site. This antibiotic binds to sites such as GAAT, ATCT and TTCTAT although with a very low affinity (table 5.5), much lower than that found for distamycin (table 5.4).

Hydroxyl radical experiments revealed some more sites that contain a CG base pair within the middle. These sites were identified as AGTT, ACTAGT, CTAC, TTGAT and TTCA. However, this sites did not produce a footprint in the DNase I footprinting gels (Fig 5.18), indicating that these sites have very low affinity for Hoechst 33258. This ligand is known to be able to change its mode of binding, from a minor groove binding agent to an intercalating agent, in regions containing CG base pairs (Bailly *et al.*, 1993). Therefore, maybe Hoechst is binding to these sites as an intercalating agent.

5.9.6 Mithramycin

This agent binds preferentially to regions that contain the sequences 5'-GG, 5;-CC

and 5'-GC, and with a much lower affinity to sites containing 5'-CG (Banville *et al.*, 1990). Our results from DNase I footprinting experiments (Fig 5.25 and table 5.6) show that its preferred binding site is AGCTCCCG (site 8) followed by GGGG (site 3) and GGGCG (site 5). Site 8 may have a slightly lower C_{50} value than site 3 because there is some cooperativity between site AGCT and site CCCG (CGGG). However, the C_{50} values for all these sites are very similar (table 5.6) and it seems probable that mithramycin is binding to the GGG sequence which is common to all these sites. The results obtained show that the flanking regions affect the binding of mithramycin. For example site AGCT (part of site 8), which is adjacent to site CCCG, yields a footprint while site TGCA (site 11) does not. This result is consistent with a footprinting study (Carpenter *et al.*, 1994) which suggest that the reason for this preference stems from the fact that mithramycin tends to avoid YpR steps. This avoidance is primarily due to the narrowness of the minor groove which is generally associated with this YpR step. In addition, our results indicate that a similar effect occurs at sites AGGT (site 9) and TGGA (sites 12 and 15), in this case site 9 yields a C_{50} value of 7 ± 2 μM while sites 12 and 15 yield no footprint. These results demonstrate again that mithramycin is avoiding YpR steps. However, there are some exceptions to this rule because site CGCGC (site 10) shows some affinity for mithramycin and contains several YpR steps.

These data also demonstrate that mithramyin binds to GpG and GpC better than to CpG. This could be seen by comparing the affinities for the sites that contain these dinucleotides (table 5.6). Almost all the sites that contain a CpG step have lower affinity for mithramycin than those that do not have it, except again for site 10 (CGCGC).

The most surprising result was the absence of a footprint at site 13 (CGGCC). This result was unexpected as the binding to a similar site (GGCC) was found to be the best binding site in a study by Liu and Chen (1994). However, the affinity to this site may be reduced because of the presence of a CpG step (YpR), which we have shown to decrease the affinity for mithramycin binding.

5.10 Conclusion

All these results have shown that the surrounding base pairs affect the affinity of echinomycin, [*N*-MeCys^3, *N*-MeCys^7]TANDEM, actinomycin D, distamycin, Hoechst 33258 and mithramycin for their preferred binding sites.

We have demonstrated that fragments MS1 and MS2 are useful for determining the binding sites of ligands with unknown specificities, as all the results determined for echinomycin, [*N*-MeCys^3, *N*-MeCys^7]TANDEM, actinomycin D, distamycin, Hoechst 33258 and mithramycin are in agreement with previously published results.

CHAPTER 6

Interaction of [*N*-MeCys^3, *N*-MeCys^7]TANDEM with different length $(AT)_n$ tracts

6.1 Introduction

The results presented in Chapter 4 demonstrated that [*N*-MeCys^3, *N*-MeCys^7] TANDEM binds to long TA regions such as ATATATAAA better than short regions like GTAT. This is consistent with several older binding and footprinting studies which showed that [*N*-MeCys^3, *N*-MeCys^7]TANDEM bound best to TpA sites which are surrounded by A and T and that the interaction with poly(dA-dT) is very strong and highly cooperative (Lee and Waring, 1978). In order to better characterize [*N*-MeCys^3, *N*-MeCys^7]TANDEM binding to long TA regions footprinting experiments were performed on a range of DNA fragments containing different length $(AT)_n$ tracts.

6.2 Fragments used

The following fragments which contain different length $(AT)_n$ tracts were chosen.

AA2

5'-GATCGGCCATATGGCCGATC-3'.(Abu-Daya, 1995)

SmaD1G

5'-CGCGATATCGCG-3'.(Lavesa *et al.*, 1993)

p$(AT)_6$

5'-AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCGCGAATTCGCGTTTAAAC GCGTAATTACGCGTATATACGCGAAATTTCGCGATC-3'. (Abu-Daya and Fox, 1997)

p$(AT)_{10}$

5'-AATTCGAGCTCGGTACCCGGGGATCGCGTTTTTAAAAACGCGTATATATATAC

GCGAAAAATTTTTCGCGAATTCGCGATCCTCTAGAGTCGACCTGCAGGCATGCAA GCTT-3'. (Abu-Daya and Fox, 1997)

pAAD1

5'-GATCGCGAATTGCGCGCTATACGCGATTACGCGCGATATCGCGCTTAACGCG ATC-3'. (Abu-Daya *et. al.*, 1995)

A_nTAT_n

5'-GATCCGGAAAAATATTTTTCCGGAAAATATTTTCCGGAAATATTTCCGGAAT ATTCCGGATC-3'. (KRF gift)

$T(AT)_8CG(AT)_{15}$

5'-TATATATATATATATATCGATATATATATATATATATATATATATAT-3' (Waterloh and Fox, 1991)

$((TAA)_4CG(TTA)_4)_2$

5'-TAATAATAATAACGTTATTATTATTATAATAATAATAACGTTATTATTATT A-3'. (Waterloh and Fox, 1991)

$((TA)_5GC(TA)_5)_2$

5'-TATATATATAGCTATATATATATATATATATAGCTATATATATA-3'. (Waterloh and Fox, 1991)

The regions underlined and bold are the $(AT)_n$ tracts of interest, the other AT tracts present within these fragments do not contain a TpA step and were therefore of less interest.

AA2, $p(AT)_6$, $p(AT)_{10}$, pAAD1, $(A)_nTA(T)_n$, were cloned into the *Bam*HI site of pUC18 and $((TA)_5GC(TA)_5)_2$, $((TAA)_4CG(TTA)_4)_2$, $T(AT)_8CG(AT)_{15}$, and SmaD1G were cloned into the *Sma*I site of pUC19. These plasmids were transformed into *E. coli* TG2 as described in section 2.10.1.2. The sequences were confirmed by dideoxy sequencing with a T7 polymerase kit (Pharmacia)(section 2.16). The plasmids that contained the synthetic inserts were radiolabelled by digesting with *Hin*dIII and *Sac*I. The 3'-end of the *Hin*dIII site was labelled with $[\alpha^{32}P]$dATP using reverse transcriptase (section 2.18.1). The labelled fragments were separated on 6% polyacrylamide gels (section 2.18.1).

6.3 Footprinting experiments

We will start by exploring those sequences that contain short $(AT)_n$ tracts. The first of this is AA2 which contains a potential [*N*-MeCys³, *N*-MeCys⁷]TANDEM binding site namely CATATG The results of DNase I footprinting experiments with this fragment are shown in Fig. 6.1. It can be seen that bands in this target site are attenuated in the presence of [*N*-MeCys³, *N*-MeCys⁷]TANDEM, though surprisingly the footprint does not seem to be complete even at the highest ligand concentrations . The footprinting plot derived from these data is shown in Fig 6.4 yielding a C_{50} value of 15 ± 8 μM (Table 6.1).

The next site to be explored was GATATC containing the same central tetranucleotide (ATAT) but flanked by GC instead of CG. This site is found within the fragment SmaD1G. DNase I footprints with this fragment are shown in the right panel of Fig 6.1. In this case bands are again attenuated only at the highest concentration of [*N*-MeCys³, *N*-MeCys⁷]TANDEM. A footprinting plot derived from these data is shown in Fig. 6.4 yielding a C_{50} value of 15 ± 4 μM (Table 6.1). This binding site (GATATC) is also found in fragment pAAD1 (Fig 6.2) described below, in which it yields a comparable binding constant.

Fragment	Site	Sequence	C_{50} (μM)
AA2	1	CATATG	15 ± 8
SmaD1G	1	GATATC	15 ± 4
pAAD1	1	GTATAC	14 ± 5
	2	GTAATC	No Footprint
	3	GATATC	18 ± 7
	4	GTTAAT	No Footprint

Table 6.1 C_{50} values (μM) determined for the interaction of [*N*-MeCys³, *N*-MeCys⁷]TANDEM with fragments containing short $(AT)_n$ tracts. The data were derived from quantitative analysis (section 2.24) of the gels shown in Fig.6.1 and 6.2, fitting a simple binding equation to the footprinting plots shown in Figure 6.4.

The sequence of pAAD1 contains four potential [*N*-MeCys³, *N*-MeCys⁷]TANDEM

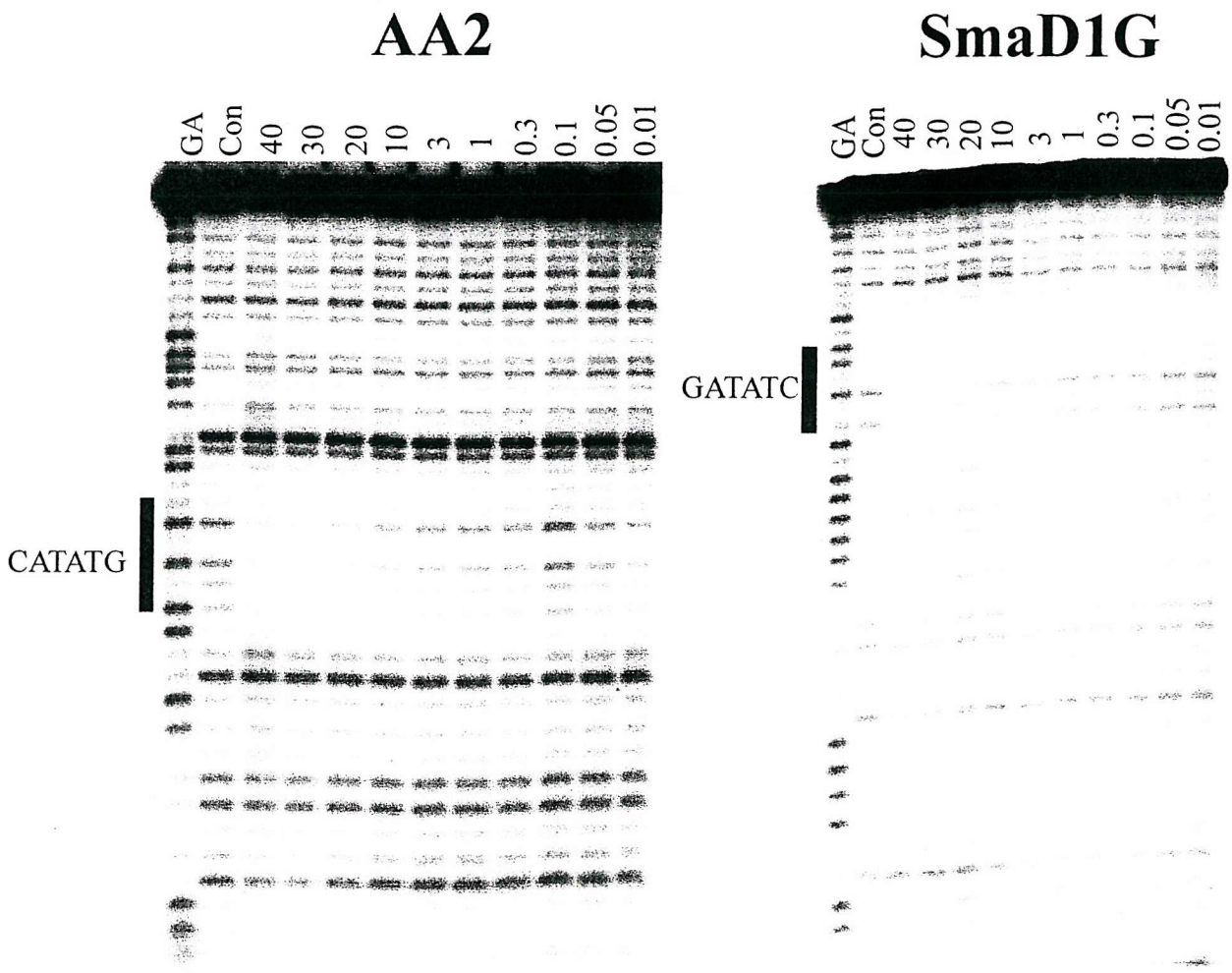


Figure 6.1 DNase I cleavage patterns of sequences AA2 and SmaD1G in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the sites of the TpA tracts. The track labelled GA is a marker specific for guanine and adenine. The track labelled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

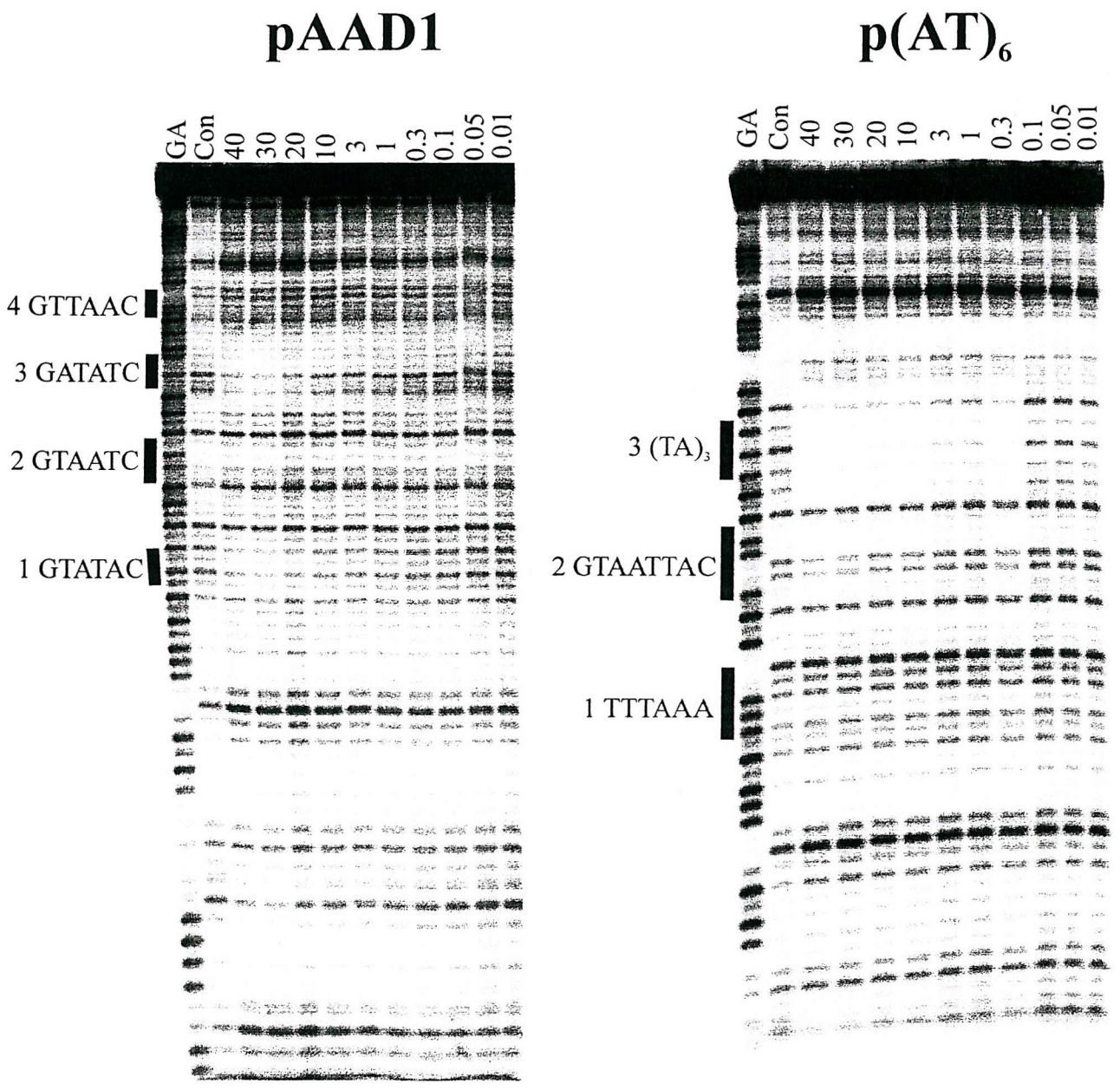


Figure 6.2 DNase I cleavage patterns of sequences pAAD1 and $p(AT)_6$ in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the sites of the TpA tracts. The lane labelled GA is a marker specific for guanine and adenine. The lane labelled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

binding sites (GTATAC, GTAA, GATATC, and TTAA). DNase I footprinting experiments with this fragment are shown in Fig 6.2 (left hand panel). Inspection of this gel reveals that sites TTAA and GTAA do not produce footprints with [*N*-MeCys^3, *N*-MeCys^7]TANDEM. This result is consistent with the results presented in Chapter 5 in which these sites also displayed very high C_{50} values (Table 5.2). However, the other sites GTATAC and GATATC yielded footprints which were analysed and the resulting footprinting plots are shown in Fig 6.4, yielding the C_{50} values which are presented in Table 6.1.

The next series of experiments explores [*N*-MeCys^3, *N*-MeCys^7]TANDEM binding to longer $(AT)_n$ regions with different sequences of A's and T's . The first fragment to be considered is $p(AT)_6$, which contains three potential [*N*-MeCys^3, *N*-MeCys^7]TANDEM binding sites which are TTTAAA, GTAATTAC and TATATA. Footprinting experiments with this fragment are shown in the right hand panel of Fig 6.2 and it can be seen that sites TTTAAA and GTAATTAC do not yield footprints. This again is not surprising as TTTAAA contains a central TTAA, which was a poor binding site in both AAD1(above) and MS2 (section 5.4). The symmetrical site GTAATTAC contains two potential binding sites GTAA/TTAC, which were also found to have high C_{50} values in the MS2 fragment (section 5.4 and Table 5.2). On the last site, TATATA, the bands are protected in the [*N*-MeCys^3, *N*-MeCys^7]TANDEM treated lanes and the footprint persists to concentrations below 0.3 μM. The footprinting plot for this site is shown in Fig 6.4 and the C_{50} value (0.05 ± 0.03) is listed in Table 6.2.

The sequence from $p(AT)_{10}$ contains two potential [*N*-MeCys^3, *N*-MeCys^7]TANDEM binding regions one is TATATATATA ($(TA)_5$) and the other TTTTTAAAAA. DNase I footprinting experiments with this fragment are presented in Fig 6.3. It can be seen that site TTTTTAAAAA does not yield any footprint. This result was expected because, as mentioned before, this site contains a central TTAA which has a very low affinity for [*N*-MeCys^3, *N*-MeCys^7]TANDEM. However, a clear footprint is evident at $(AT)_{10}$ which persists to the lowest [*N*-MeCys^3, *N*-MeCys^7]TANDEM concentration (0.01 μM). This site can be compared to the one found in $(AT)_6$, TATATA, and visual inspection of the gels (Fig 6.2 and 6.3) reveals that $(TA)_5$ is a better binding site. A footprinting plot showing the interaction

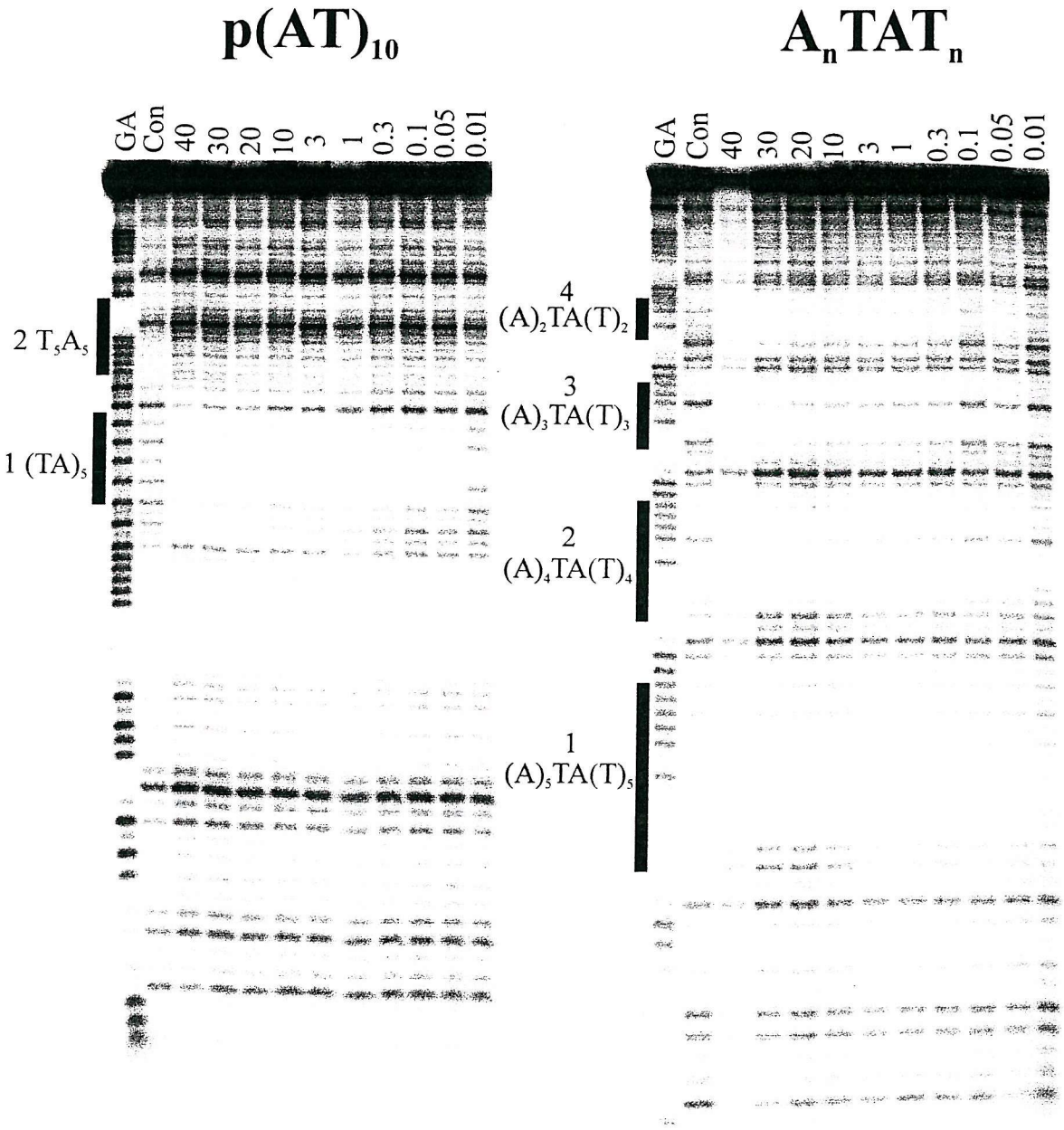


Figure 6.3 DNase I cleavage patterns of sequences $p(AT)_{10}$ and A_nTAT_n in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the sites of the footprints. The track labelled GA is a marker specific for guanine and adenine. The track labelled 'con' is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

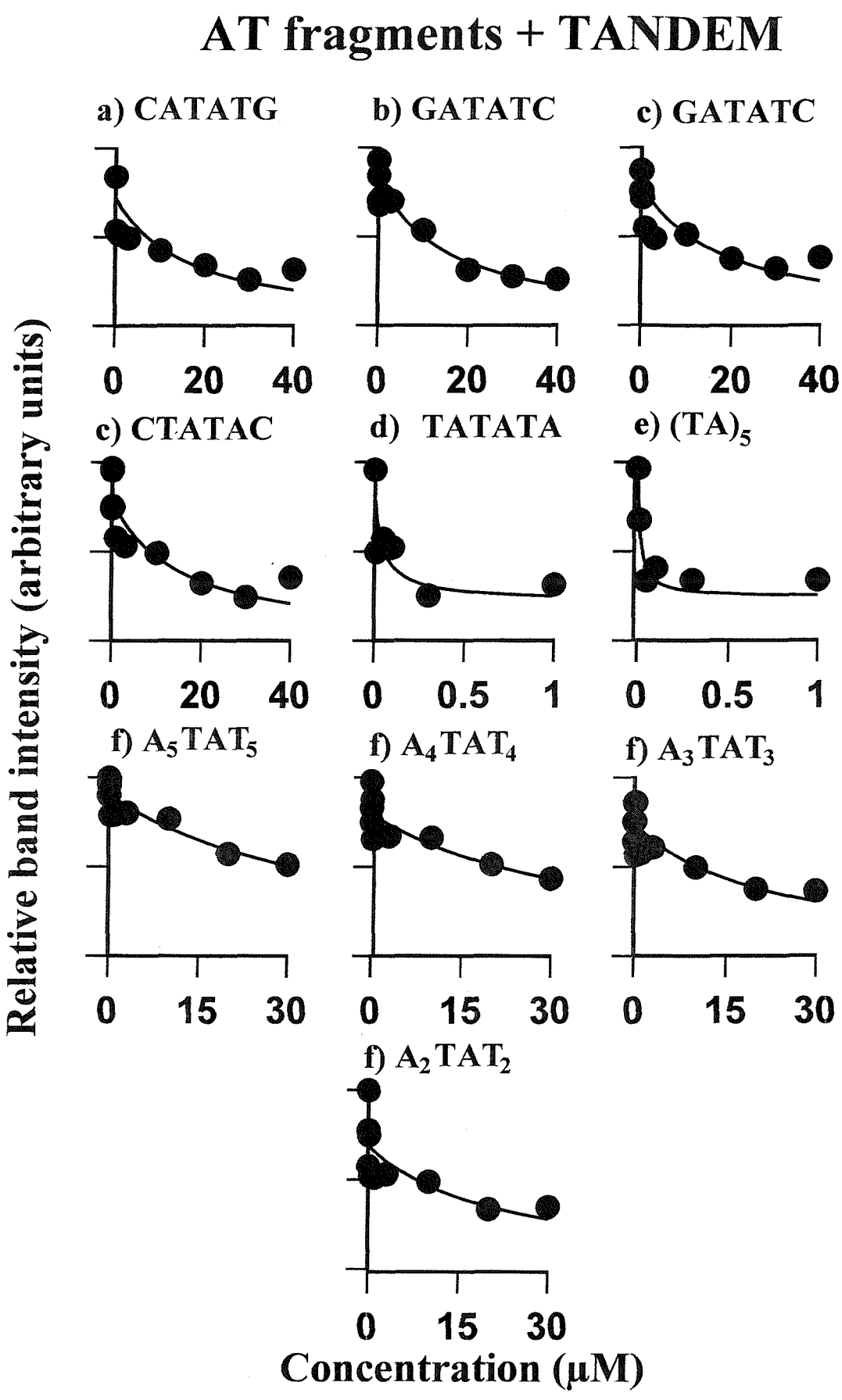


Figure 6.4 Footprinting plots showing the interaction of [*N*-MeCys^3, *N*-MeCys^7] TANDEM with a) AA2, b) SmaD1G, c) pAAD1, d) $p(AT)_6$, e) $p(AT)_{10}$ and f) A_nTAT_n. Data was analysed as described in section 2.24.

with this site is shown in Fig 6.4 and the C_{50} value (0.015 ± 0.006) is listed in Table 6.2. These results suggests that [*N*-MeCys³, *N*-MeCys⁷]TANDEM increases its affinity in proportion to the length of the alternating TA tract. Several fragments containing different lengths of alternating TA are described below.

The next sequence to be studied is A_nTAT_n. This fragment explores sites that contain TpA steps embedded in the sequence A_nTAT_n. We have already seen that TpA sites in T_nTAA_n (TTAA, TTTAAA and $(T)_4TA(A)_4$) have very low affinity for [*N*-MeCys³, *N*-MeCys⁷] TANDEM. The sites within this fragment are $(A)_5TA(T)_5$, $(A)_4TA(T)_4$, $(A)_3TA(T)_3$ and $(A)_2TA(T)_2$. A DNase I footprinting experiment using this fragment is shown in Fig. 6.3 (left panel). It can be seen that bands near each TpA step are attenuated at the highest [*N*-MeCys³, *N*-MeCys⁷]TANDEM concentrations. The C_{50} values which are shown in Table 6.2 show that the affinity of [*N*-MeCys³, *N*-MeCys⁷]TANDEM decreases as the number of As and Ts, flanking the central TpA site increases. Even though these differences are not large they demonstrate that [*N*-MeCys³, *N*-MeCys⁷]TANDEM binding is affected by the flanking regions around its preferred binding site TpA. The footprinting plots for these sites are presented in Fig 6.4.

Fragment	Site	Sequence	C_{50} (μM)
$p(AT)_6$	1	TTTAAA	No Footprint
	2	GTAATTAC	No Footprint
	3	TATATA	0.05 ± 0.03
$p(AT)_{10}$	1	$(TA)_5$	0.015 ± 0.006
	2	$(T)_4TA(A)_4$	No Footprint
A_nTAT_n	1	$(A)_5TA(T)_5$	38 ± 9
	2	$(A)_4TA(T)_4$	35 ± 13
	3	$(A)_3TA(T)_3$	21 ± 10
	4	$(A)_2TA(T)_2$	20 ± 11

Table 6.2 C_{50} values (μM) determined for the interaction of [*N*-MeCys³, *N*-MeCys⁷]TANDEM with long $(AT)_n$ tracts. The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 6.4.

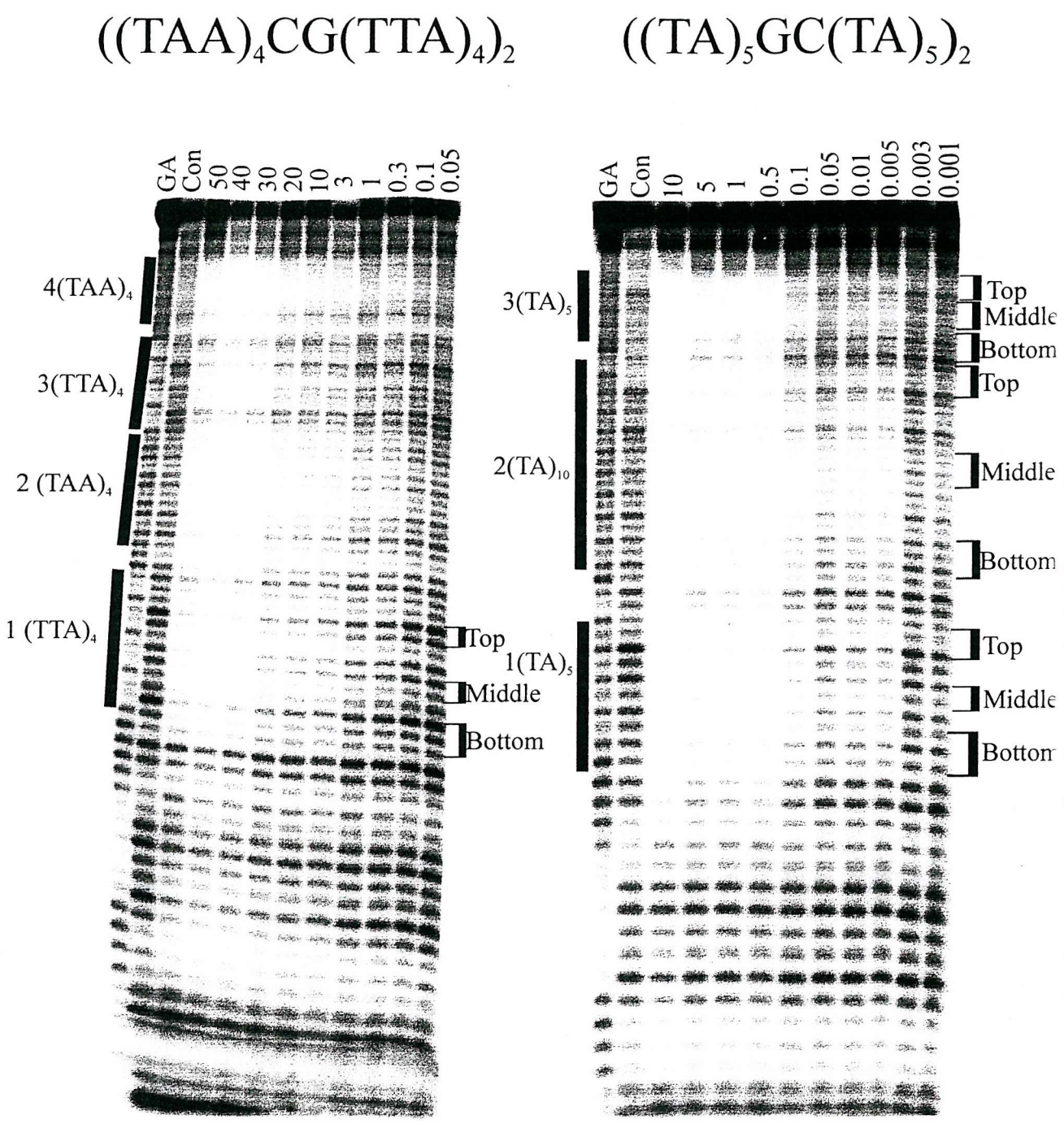


Figure 6.5 DNase I cleavage patterns of sequences $((TAA)_4GC(TTA)_4)_2$ and $((TA)_5GC(TA)_5)_2$ in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the sites of the TpA tracts. The track labelled GA is a marker specific for guanine and adenine. The track labelled 'con'is the control which shows the digestion in the absence of ligand. The fragments are labelled at the 3'-end.

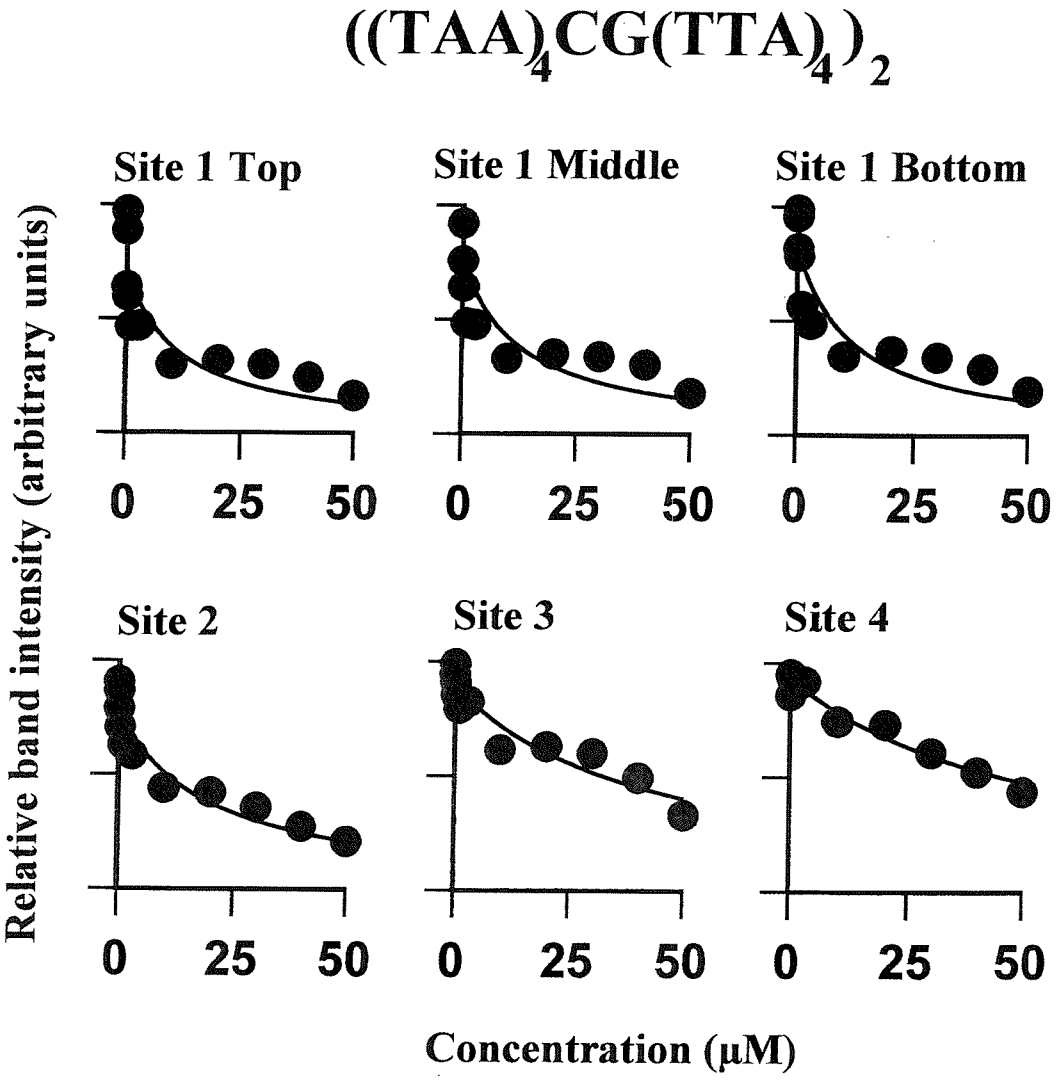


Figure 6.6 Footprinting plots showing the interaction of [*N*-MeCys^3, *N*-MeCys^7] TANDEM with sites in $((TAA)_4CG(TTA)_4)_2$. Data was analysed as described in section 2.24.

The next fragment investigates regions that contain four repeats of TTA(TAA) in which the TpA steps are located in the centre TTAT(ATAA). The results with this fragment are shown in Fig 6.5. It can be seen that bands in the AT-rich regions are attenuated at the highest concentrations of [*N*-MeCys^3, *N*-MeCys^7]TANDEM. Since each TTA(TAA) tract contains several equivalent TpA steps we decided to estimate wether [*N*-MeCys^3, *N*-MeCys^7]TANDEM bound best to the centre than to the edges of the tract. This analysis was performed on the lowest (TTA) tract while the footprints at the other sites were analysed over the entire region. Footprinting plots derived from these data are presented in Fig. 6.6 and the C_{50} values are summarized in Table 6.3. It can be seen that for site 1 there is no difference in apparent binding affinity to the various parts of this region. Sites 3 and 4 have large C_{50} values probably due to loss of resolution, as they are located towards the top of the gel. Some of these C_{50} values are similar to those obtained for the interaction with ATAA/TTAT presented in chapter 5.

Fragment	**Site**	**Sequence**	**C_{50} (μM)**
$((TAA)_4GC(TTA)_4)_2$	1 Top	$(TTA)_4$	11 ± 5
	1 middle		12 ± 6
	1 bottom		11 ± 4
	2	$(TAA)_4$	18 ± 4
	3	$(TTA)_4$	40 ± 7
	4	$(TAA)_4$	52 ± 5

Table 6.3 C_{50} values (μM) determined for the interaction of [*N*-MeCys^3, *N*-MeCys^7]TANDEM with $(TAA)_4GC(TTA)_4$. The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 6.6.

The last two fragments explore the binding of [*N*-MeCys^3, *N*-MeCys^7]TANDEM to different length alternating AT tracts. The first such fragment is $((TA)_5GC(TA)_5)_2$. This sequence contains two $(TA)_5$ at the edges, each of which has three ATAT sites at the centre with two CTAT/ATAG at the edges. The central region has the sequence $C(TA)_{10}G$. A DNase I footprinting gel with this fragment is shown in Fig.6.5 (right hand panel). It can be seen that [*N*-MeCys^3, *N*-MeCys^7]TANDEM has reduced the cleavage of all the bands in this sequence and that some regions of attenuation persist to concentrations below 0.01 μM. In

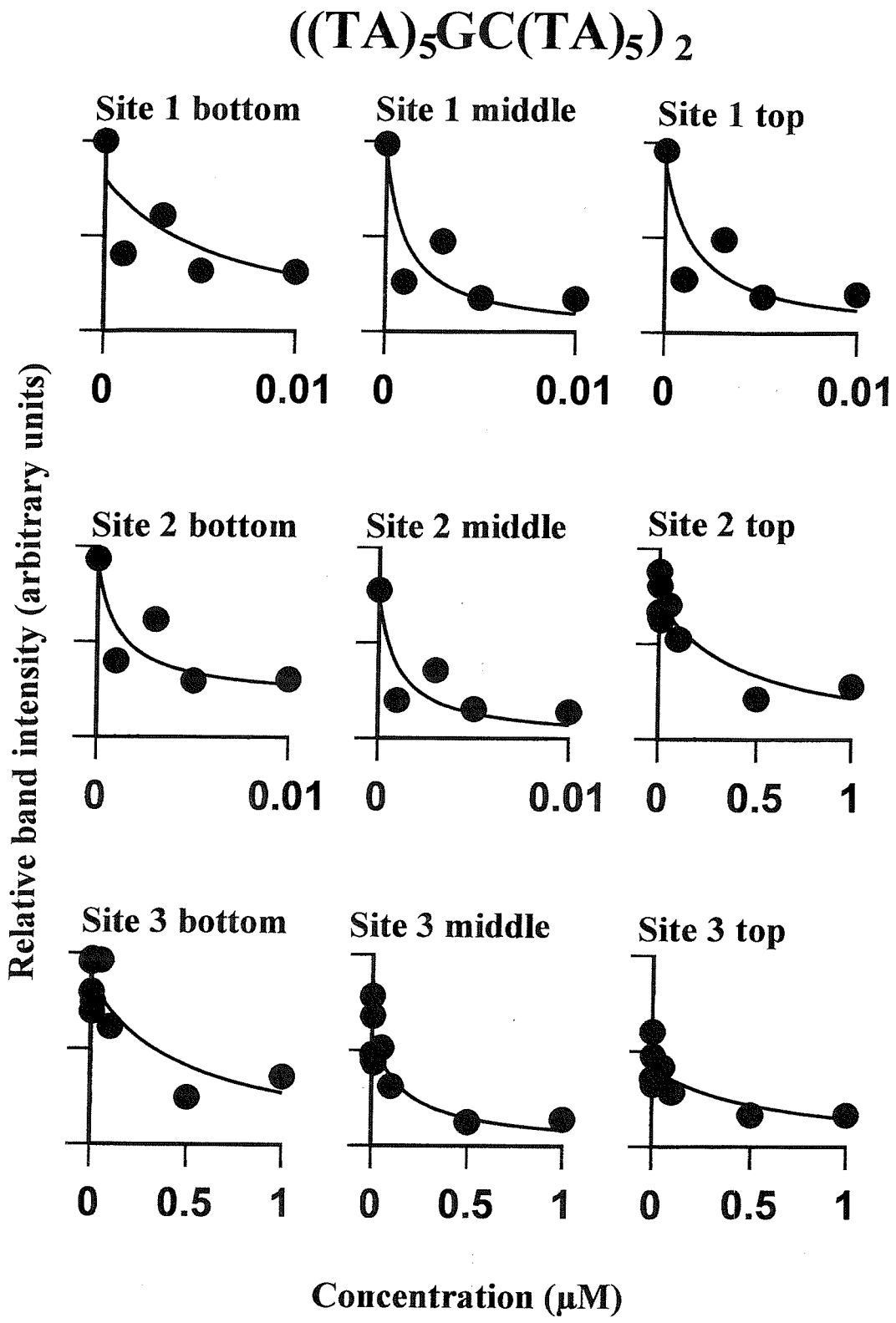


Figure 6.7 Footprinting plots showing the interaction of [*N*-MeCys^3, *N*-MeCys^7] TANDEM with sites in $((TA)_5GC(TA)_5)_2$. Data was analysed as described in section 2.24.

these regions although the band intensity changes in a concentration dependent fashion, it does not seem to return to that at the control at the lowest concentrations. It can also be seen that the footprints appear to be clearer at the centre than at the edges of each $(TA)_n$ tract. This observation suggests that [*N*-MeCys³, *N*-MeCys⁷]TANDEM has higher affinity for the centre of the site than for the outer edges. However, it should be emphasized that there is binding to all the TpA steps. Footprinting plots derived from these data are shown in Fig.6.7 and the C_{50} values are summarized in Table 6.4. These C_{50} values show that [*N*-MeCys³, *N*-MeCys⁷] TANDEM has very high affinity for $(AT)_n$ tracts. However, it should be noted that the ligand appears to have a higher affinity for the central regions of $(AT)_n$ tracts compared to the edges of $(AT)_n$ tracts. This shall be considered in greater detail in the Discussion

Fragment	Site	Sequence	C_{50} (μM)
$((TA)_5GC(TA)_5)_2$	1 bottom	$(TA)_5$	0.006 ± 0.005
	1 middle		0.001 ± 0.0006
	1 top		0.0015 ± 0.0009
	2 bottom	$(TA)_{10}$	0.001 ± 0.0008
	2 middle		0.001 ± 0.0005
	2 top		0.4 ± 0.2
	3 bottom	$(TA)_5$	0.5 ± 0.2
	3 middle		0.16 ± 0.1
	3 top		0.54± 0.37
$T(AT)_8CG(AT)_{15}$	1 bottom	$(TA)_{15}$	0.002 ± 0.0008
	1 middle		0.001 ± 0.0004
	1 top		0.03 ± 0.02
	2 bottom	$(TA)_8$	0.09 ± 0.06
	2 middle		0.005 ± 0.003
	2 top		0.005 ± 0.004

Table 6.4 C_{50} values (μM) determined for the interaction of [*N*-MeCys³, *N*-MeCys⁷]TANDEM with different lengths of alternating TpA steps. The data were derived from quantitative analysis (section 2.24) of the plots shown in Figure 6.7 and 6.9.

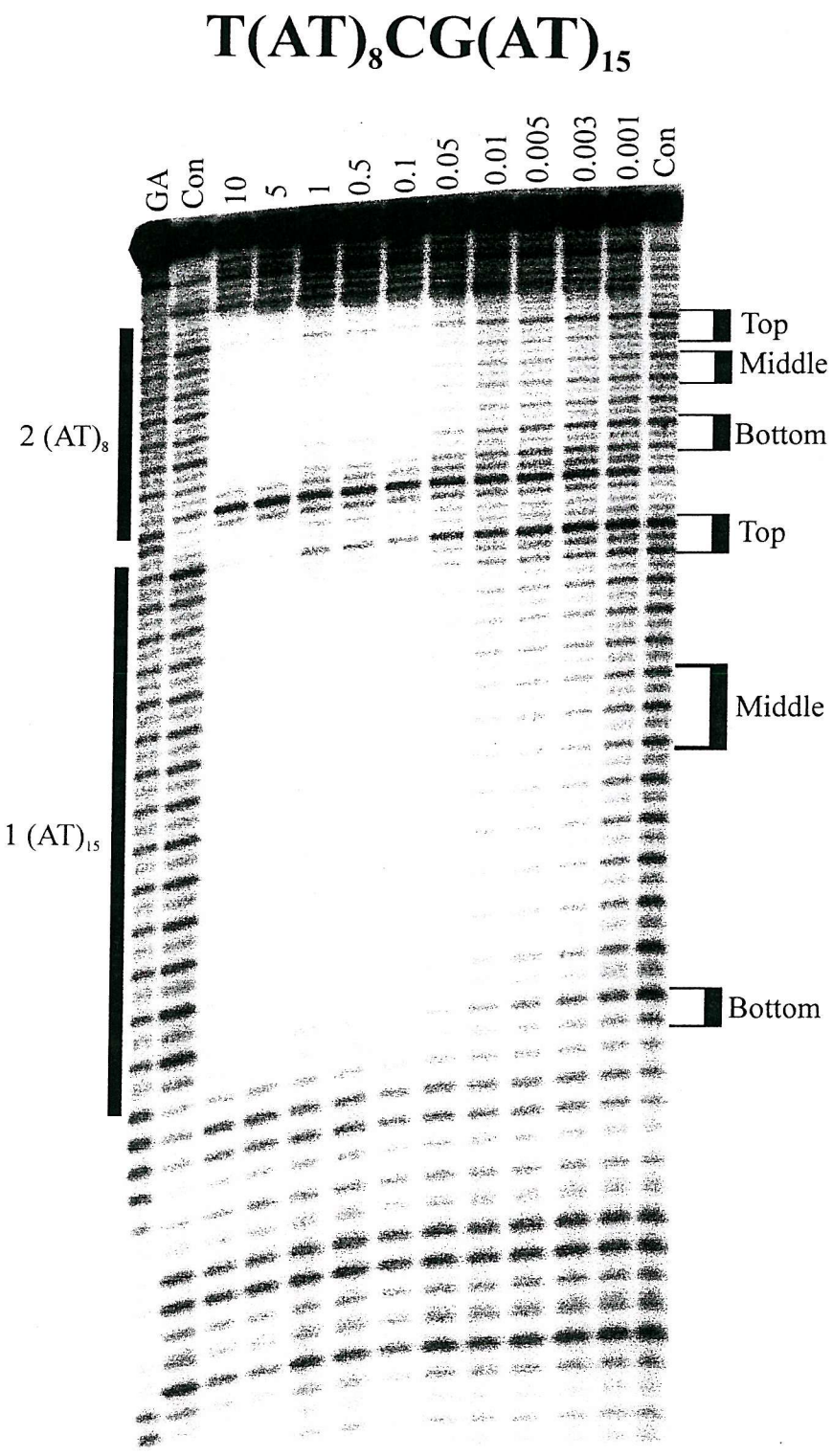


Figure 6.8 DNase I cleavage patterns of sequence $T(AT)_8CG(AT)_{15}$ in the presence and absence of [*N*-MeCys^3,*N*-MeCys^7]TANDEM. Drug concentrations (μM) are indicated at the top of each gel lane. The black boxes indicate the sites of the footprints. The track labelled GA is a marker specific for guanine and adenine. The track labelled 'con' is the control which shows the digestion in the absence of ligand. The fragment is labelled at the 3'-end. The brackets indicate the bands that have been analysed

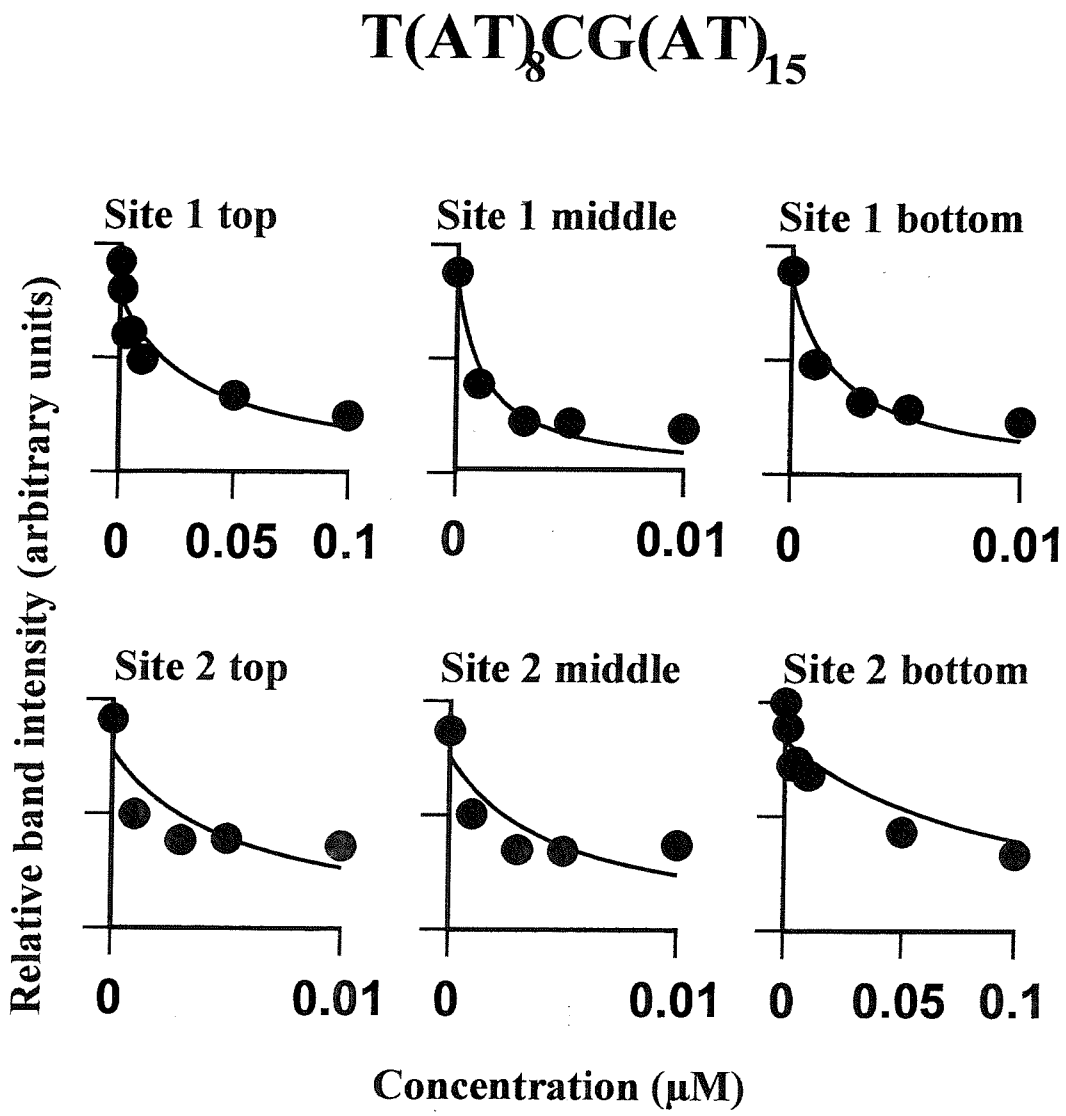


Figure 6.9 Footprinting plots showing the interaction of [*N*-MeCys^3, *N*-MeCys^7] TANDEM with sites in $T(AT)_8CG(AT)_{15}$. Data was analysed as described in section 2.24.

The next fragment examined contained the sequence $T(AT)_8CG(AT)_{15}$, and DNase I footprinting results with this are shown in Fig 6.8. This sequence contains two long TA tracts, site 1 has 14 and site 2 has 8 TpA steps. It can be seen that [*N*-MeCys³, *N*-MeCys⁷]TANDEM decreases the intensity of bands in all the AT-tracts in a concentration dependent fashion. The bands around the central TCGA site are not protected, as expected.

The protection at these sites also appears to vary with their location, so that bands in the centre of each AT-tract are attenuated to lower [*N*-MeCys³, *N*-MeCys⁷]TANDEM concentrations than those at the edges. This is most pronounced at the first ATAT below the central CG for which bands reappear at much higher [*N*-MeCys³, *N*-MeCys⁷]TANDEM concentrations than in the remainder of the fragment. Footprinting plots derived from these data are shown in Fig. 6.9 and the C_{50} values are summarized in Table 6.4. However, it should be noted that the intensity of the bands has not returned to that in the control, even at the lowest ligand concentration, except for the weakest site near the central CpG. This simple binding analysis may therefore not adequately describe the interaction, and maybe there is some cooperative process occurring which will be considered further in the Discussion.

These data shows that [*N*-MeCys³, *N*-MeCys⁷]TANDEM has a higher affinity for long alternating AT tracts, than for other TA tracts.

6.4 Discussion

The results presented throughout this thesis show that the affinity of [*N*-MeCys³, *N*-MeCys⁷]TANDEM for TpA sites changes depending on the flanking regions. Here we have explored different sequences containing $(TA)_n$ tracts and we have arranged them according to their length and composition.

6.4.1 Short $(TA)_n$ tracts

The sequences containing short isolated tetranucleotide sequences were AA2, pAAD1, and SmaD1G. The results obtained with these fragments show that even within tetranucleotide sites [*N*-MeCys³, *N*-MeCys⁷]TANDEM has a higher affinity for regions

containing alternating A and T residues, since ATAT has a higher affinity for [*N*-MeCys³, *N*-MeCys⁷]TANDEM than TATA and TTAA. In addition, all the ATAT show comparable C_{50} values (see Table 6.1 for CATATG and GATATC and Table 6.2 for A_2TAT_2). Therefore the ranking order of these short tetranucleotide sites in decreasing affinity is ATAT > TATA >>TTAA. These results are in agreement with previously published data (Fletcher *et al.*, 1995). However the sequence specificity of [*N*-MeCys³, *N*-MeCys⁷]TANDEM is still not fully understood. At first it was postulated that [*N*-MeCys³, *N*-MeCys⁷]TANDEM did not bind to CpG due to the formation of internal hydrogen bonds between the alanine CO groups and valine NH groups which prevented the interaction with the 2-amino group of guanine. In addition, an NMR study using the fragment GATATC proposed that [*N*-MeCys³, *N*-MeCys⁷]TANDEM made a hydrogen bond to the N-3 of adenine (Addess *et al.*, 1993) which should be equally feasible with the N-3 of guanine. However it has been shown that TANDEM does not bind to IpC even though this has a minor groove which is identical to TpA, with the same hydrogen bonding capacities (Bailly and Waring, 1998). Therefore, it appears that hydrogen bonding alone can not explain the binding specificity of TANDEM and that some other interactions such as steric and/or hydrophobic may play a role, probably involving the recognition of a minor groove with suitable flexibility and/or structure. In conclusion there is clearly something special about regions of alternating AT which generates good TANDEM binding sites.

6.4.2 non-alternating regions

6.4.2.1 T_nTAA_n and A_nTAT_n tracts

The sequences that contain these kind of sites are $p(AT)_6$, $p(AT)_{10}$ and A_nTAT_n. In general A_nTAT_n sites have very low affinity for [*N*-MeCys³, *N*-MeCys⁷]TANDEM while T_nTAA_n does not show any footprints. The results obtained for sites T_nTAA_n. were not surprising as these sites contain a central TTAA which has very low affinity for [*N*-MeCys³, *N*-MeCys⁷]TANDEM. In contrast the results obtained for A_nTAT_n show that the affinity of [*N*-MeCys³, *N*-MeCys⁷]TANDEM decreases as the number of As and Ts flanking the central TpA site increases. These results obtained for A_nTAT_n and T_nTAA_n could be due to changes within the structure of the minor groove. It is known that $(A)_n$ sequences tend to get narrower

towards the 3' end (Burkhoff and Tullius, 1997). Therefore these sequences will tend to be narrower towards the centre of the sequence, within the TpA step, probably preventing [*N*-MeCys^3, *N*-MeCys^7]TANDEM binding. This narrowing effect should increase with longer tracts of As and Ts yielding a lower affinity for [*N*-MeCys^3, *N*-MeCys^7]TANDEM. In addition it is possible that DNA flexibility plays a role in affecting the affinity since A_nT_n tracts are likely to be less flexible than other AT-containing sequences.

6.4.2.2 fragment $((TAA)_4GC(TTA)_4)_2$

This fragment has several potential [*N*-MeCys^3, *N*-MeCys^7]TANDEM binding sites with the sequence ATAA/ TTAT. These sites were found to have moderate C_{50} values in the MS1 and MS2 fragments (Table 5.2). Only sites 1 and 2 of this fragment showed comparable C_{50} values to those found for MS1 and MS2. The other sites (3 and 4) have higher C_{50} values (Table 6.3). It should be considered however that this result may be an artifact of the footprinting gel. This is due to the sites being located towards the top of the gel where the resolution is lower, and accurate binding data are harder to obtain.

6.4.3 Long alternating $(TA)_n$ tracts

The fragments that contained long alternating $(TA)_n$ tracts were $p(AT)_6$, $p(AT)_{10}$, $((TA)_5GC(TA)_5)_2$ and $T(AT)_8CG(AT)_{15}$. The results (Tables 6.2 and 6.4) seem to show that as the number of alternating TpA steps increase the affinity for [*N*-MeCys^3, *N*-MeCys^7] TANDEM increases as well. Examination of the gels (Figs 6.2, 6.3, 6.5 and 6.8) reveals that [*N*-MeCys^3, *N*-MeCys^7]TANDEM has a higher affinity for the central regions of each $(AT)_n$ tract than for the edges. This observation was previously made by Fletcher *et al.* (1995). There are several explanations for this phenomenon: firstly the cooperativity could be originated by interactions between drug molecules bound to adjacent sites. Secondly, the DNA structure could be disrupted by the binding of the ligand in such a way that further ligand binding is facilitated. Thirdly the structure of long $(TA)_n$ tracts is known to vary along its length (Klug *et al.*, 1979; McClellan *et al.*, 1986), even in the absence of ligand. Therefore, the binding of [*N*-MeCys^3, *N*-MeCys^7]TANDEM could simply be showing these

differences.

As we saw in chapter 4, a simple binding equation does not adequately describe the binding process of [*N*-MeCys^3, *N*-MeCys^7]TANDEM to long alternating $(TA)_n$ tracts. For example site 1 middle from fragment $((TA)_5GC(TA)_5)_2$ (Fig 6.5) gives an evident footprint at 40 μM ligand concentration, which fades in a concentration dependent fashion, with about 90% protection at 1 μM ligand concentration. For a simple binding equation in which the %bound = $L/(L + K_d)$ (where L is the free ligand concentration and K_d is the dissociation constant) would predict a Kd of about 0.11 μM. Now if we use this Kd for L = 0.01 μM it gives 8% bound in contrast to the about 80% that is still observed (Fig 6.5 and 6.7). The details explaining this observation are found in section 4.8.2. This observation could be due to several reasons: First it could be due to the positive cooperativity (Lee and Waring, 1976) and to the unusual kinetics found when TANDEM binds to poly(dA-dT) (Fox *et al.*, 1982). Second this observation could be a consequence of having overlapping binding sites.

REFERENCES

Abu-Daya, A. (1995). DNA structure and its recognition by minor groove binding ligands. Ph.D thesis

Abu-Daya, A.and Fox, K.R. (1997). Interaction of minor groove ligands with long AT tracts. *Nucl. Acids Res.* **25**, 4962-4969.

Abu-Daya,A.,Brown,P.M.,and Fox, K.R. (1995) DNA sequence preferences of several AT-selective minor groove binding ligands. *Nucl. Acid Res.* **23**,No 17, 3385-3392.

Addess, K.J. and Feigon, J. (1994a). Sequence specificity of quinoxaline antibiotics. 2. NMR studies of the binding of [N-MeCys^3,N-MeCys^7]TANDEM and Triostin A containing a CpI step. *Biochem.* **33**, 12397-12404.

Addess, K.J. and Feigon, J. (1994b). Sequence specificity of quinoxaline antibiotics. 1.Solution structure of a 1:1 complex between Triostin A and $[d(GACGTC)]_2$ and comparison with the solution structure of the [N-MeCys^3,N-MeCys^7]TANDEM-$[d(GACGTC)]_2$ complex. *Biochem.* **33**, 12386-12396.

Addess, K.J. Sinsheimer, J.S. and Feigon, J. (1993). Solution structure of a complex between [N-MeCys^3,N-MeCys^7]TANDEM and $[d(GATATC)]_2$. *Biochem.* **32**, 2498-2508.

Araki, T., Yamamoto, A. And Yamada, M (1987). Accurate determination of DNA content in single cell nuclei stained with Hoechst 33258 fluorochrome at high salt concentration. *Histochem.* **87:** 331-338.

Arison, B.H. and Hoogsteen, K. (1970). Nuclear magnetic resonance spectral studies on actinomycin D . Preliminary observations on the effect of complex formation with 5'-deoxyguanylic acid. *Biochem.* **9**, 3976-3983.

Arnold, E. and Clardy, J. (1981). Crystal and Molecular structure of BBM-928A, a novel antitumour antibiotic from actinomadura luzonensis. *J.Amer. Chem. Soc.* **103**, 1243-1244.

Bailey, S.A., Graves , D.E. and Rill, R. (1994). Binding of actinomycin to the $T(G)_nT$ motif of double-stranded DNA: Determination of the guanine requirem,ent in non-classical, non-GpC binding sites. *Biochemistry* **33**, 11493-11500.

Bailly, C. and Waring, M.J. (1995). Transferring the purine 2-amino group from guanines to adenines in DNA changes the sequence-specific binding of antibiotics. *Nucl. Acids Res.* **23**, 885-889.

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., Colson, P., Henichart, J.P. and Houssier, C.,(1993). The different binding modes of Hoechst 33258 to DNA studied by electric linear dichroism. *Nucl. Acids Res.*, **21**, 3705-3709

Bailly, C., Graves, D.E., Ridge, G. and Waring, M.J. (1994). Use of a photoactive derivative of actinomycin to investigate shuffling between binding sites on DNA. *Biochem.* **33**, 8736-8745.

Banville, D.L., Keniry, M.A., Kam, M. and Shafer, R.H. (1990). NMR studies of the interaction of chromomycin A_3 with small DNA duplexes. Binding to GC-containing sequences. *Biochem.* **29**, 6521-6534.

Berman, E. and Kam, M.(1989). The solution conformation of the antibiotic anticancer chromomycin A_3 by two-dimensional NMR spectroscopy. Computer-assisted modelling of receptor-ligand interactions: *theoretical aspects and applications to drug design* 217-227.

Berman, E., Brown, S.C., James, T.L. and Shafer, R.H. (1985). NMR studies of chromomycin A_3 interaction with DNA. *Biochem.* **24**, 6887-6893.

Bernardi, G. (1971). In: The Enzymes, Boyer, P.D. (Ed.), Academic Press, London, pp 271-287.

Burkhoff, A.M. and Tullius, T.D. (1987). The unusual conformation adopted by the Adenine tracts in Kinetoplast DNA. *Cell* **48**, 935-943.

Cairns, J. (1962). The Application of Autoradiography to the Study of DNA Viruses. *Cold Spring Harbor Symp. quant. Biol.*, **27**, 311-318.

Carpenter, M.L., Marks, J.N. and Fox, K.R.(1993). DNA-sequence binding preference of the GC-selective ligand mithramycin. *Eur. J. Biochem.* **215**, 561-566.

Carpenter, M.L., Cassidy, S.A.and Fox, K.R.(1994). Interaction of mithramycin with isolated GC and CG sites. *J. Mol. Recognition* **7**, 189-197.

Cerami, A., Reich, E., Ward, D.C. and Goldberg I.H. (1967). The interaction of actinomycin with : Requirement for 2 -amino group of purines. *Proc. Natl. Acad. Sci. U.S.A.*, **57**, 1036-1042.

Chen, F.-M. (1989). Observation of an Anomalously Slow Association Kinetics in the Binding of Actinomycin D to d(CATGGCCATG). *Biochem.* **29**, 7684-7690.

Chen, F. and Sha, F. (1998). Circular dichroic and kinetic differentiation of DNA binding modes of distamycin *Biochem.* **37**, 11143-11151.

Chen, H., Liu, X. and Patel, D.J. (1996). DNA bending and unwinding associated with actinomycin D antibiotics bound to partially overlapping sites on DNA. J. Mol. Biol. **258**, 457-479.

Chen, X., Ramakrishnan, B. and Sundaralingam, M. (1997). Crystal structures of the side-by-side binding of distamycin to AT-containing DNA octamers d(ICITACIC) and d(ICATATIC). *J. Mol. Bio.* **267**, 1157-1170.

Cheng,A.Y., Yu, C., Gatto, B., Liu, L.F. (1993) DNA minor groove-binding ligands: A different class of mammalian DNA topoisomerase I inhibitors. *Proc. Natl. Acad. Sci. USA* **90**, 8131-8135.

Cheung, H.T., Feeney, J., Roberts, G.C.K., Williams, D.H., Ughetto, G. and Waring, M.J. (1978). The Conformation of Echinomycin in Solution, *J. Am. Chem. Soc.* **100**, 46-54.

Churchill, M.E.A., Hayes, J.J. and Tullius, T.D.(1990) Detection of Drug Binding to DNA by Hydroxyl Radical Footprinting. Relationship of Distamycin Binding Sites to DNA Structure and Positioned Nucleosomes on 5S RNA Genes of *Xenopus. Biochem.* **29**, 6043-6050.

Clark, R.G., Gray, E.J., Neidle, S., Li Y.H. and Leupin, W.(1996). Isohelicity and phasing in Drug-DNA Sequence Recognition: Crystal Structure of a Tris(benzimidazole)-Oligonucleotide Complex. *Biochem.* **35,** 13745-13752.

Cohen, J.S., Wooten, J.B. and Chatterjee, C.L. (1981). Characterization of Alternating Deoxyribonucleic Acid Conformations in Solution by Phosphorous-31 Nuclear Magnetic Resonance Spectroscopy. *Biochem* **20**, 3049-3055.

Coll, M., Frederick, C.A., Wang, A.H.J., and Rich, A. A(1987). 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. USA* **84**, 8385-8389.

Cons, B.M.G. and Fox, K.R. (1989). High Resolution Hydroxyl Radical Footprinting of the Binding of Mythramycin and Related Antibiotics to DNA. *Nucl. Acids Res.* **17**, 5447-5459.

Cons, B.M.G. and Fox, K.R. (1990). The GC-selective ligand mithramycin alters the structure of $(AT)_n$ sequences flanking its binding sites. *FEBS Lett.* **264**, 100-104.

Cons, B.M.G. and Fox, K.R. (1991). Effects of the antitumor antibiotic mithramycin on the

structure of repetitive DNA regions adjacent to its GC-rich binding site. *Biochem.* **30**, 6314-6321.

Dasgupta, D., Howard, F.B., Sasisekharan, V., and Miles, H.T..(1990). Drug-DNA Binding Specificity: Binding of Netropsin and Distamycin to Poly(d2NH_2A-dT). *Biopolymers,* **30**, 223-227.

Deiters,, J.A., Galluci, J.C., and Holmes, R.R. (1982). Computer simulation of staphylococcal nuclease action on Thymidine 3',5'-bis (phosphate)(pdTp). *J. Amer. Chem. Soc.* **104**, 5457-5465.

Del Sal, G., Manfioletti, G. and Schneider, C. (1988). A one-tube plasmid DNA mini-preparation suitable for sequencing. *Nucl. Acids. Res.* **16**, 9878.

Dervan, P.B. (1986). Design of sequence-specific DNA-binding molecules. *Science* 232, 464-471

Dickerson, R.E., Drew, H.R., Conner,B.N., Wing, R.M., Fratini, A.V., and Kopka, M.L. (1982). The Anatomy of A-, B-, and Z-DNA. *Science* **216**, 475-485.

Dingwall, C., Lomonossoff, G.P. and Laskey, R.A. (1981). High sequence specificity of micrococcal nuclease. *Nucl. Acids Res.* **9**, 2659-2673

Drew, H.R. (1984). Structural specificities of five commonly used DNA nucleases. *J. Mol. Biol.* **176**, 535-557.

Drew, H.R. and Travers, A.A. (1984). DNA structural variation in the E.coly TyrT 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. *Nucl. Acid Res.* **27**, 4100-4105.

Ekambareswara Rao, K., Dasgupta, D. and Sasisekharan, V. (1988) Interaction of synthetic analogues of distamycin and netropsin with nucleic acids. Does curvature of ligand play a role in distamycin-DNA interactions?. *Biochem.* **27**, 3019-3024.

Elias, E.G. and Evans, J.T. (1972) Mithramycin in the treatment of Paget's disease of bone. *J. Bone Joint Surg.* **54**(8), 1730-1736.

Evans, D.H., Lee, J.S., Morgan, A.R. and Olsen, R.K. (1982). A method for the specific inhibition of poly [d(A-T)] synthesis using the A-T specific quinoxaline antibiotic TANDEM. *Can. J. Biochem.* **60**, 131-136.

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.

Fish, E.L., Lane, M.J., and Vournakis, J.N. (1988). Determination of equilibrium Binding Affinity of Distamycin and Netropsin to the Synthetic Deoxyoligonucleotide Sequence d(GGTATACC)$_2$ by Quantitative DNase I footprinting. *Biochem.* **27**, 6026-6032.

Fletcher, M.C. and Fox, K.R. (1996). Dissociation kinetics of actinomycin D from individual GpC sites in DNA. *Eur. J. Biochem.* **237**, 164-170.

Fletcher, M.C., Olsen, K.R. 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.

Flick, J.T., Eissenberg, J.C. and Elgin, S.C.R. (1986). Micrococcal nucleaseas a DNA structural probe: Its recognition sequences, their genomic distribution and correlation with DNA determinants. *J. Mol. Biol.* **190**, 619- 633.

Fok, J. and Waring, M.J. (1972) Breakdown of pulse-labeled ribonucleic acid in Bacillus megaterium, revealed by exposure to the antibiotics mithramycin, chromomycin, and nogalamycin. *Mol.Pharmacol.* **8**(1), 65-74.

Fox, K.R. (1988). Footprinting studies in the interactions of nogalomycin, arugomycin, decilorubicin and viriplanin with DNA. *Ancan. Drug Des.* **3**,157-168.

Fox, K.R. (1992). Use of Enzymatic and Chemical Probes to Determine the Effect of Drug Binding on Local DNA Structure. In advances in DNA sequence specific agents Hurley,L.H. Editor, Jai Press Inc. Vol 1, 167-214.

Fox, K.R. and Howarth, N.R. (1985). Investigations into the sequence-selective binding of mithramycin and related ligands to DNA. *Nucl. Acid Res.* **13**, 8695-8714.

Fox, K.R. and Waring, M.J. (1984). Structural variations produced by actinomycin and distamycin as revealed by DNAase I footprinting. *Nucl. Acid Res.* **12**, 9271-9285.

Fox, K.R.,and Waring, M.J. (1986). Footprinting reveals that nogalamycin and actinomycin shuffle between DNA binding sites. *Nucl. Acids Res.* **14**, 2001-2014.

Fox , K.R. and Waring, M.J. (1987). Footprinting at low temperatures: evidence that ethidium and other simple intercalators can discriminate between nucleotide sequences. *Nucl. Acids Res.* **15**, 491-507.

Fox, K.R., Marks, J.N. and Waterloh, K. (1991). Echinomycin binding to alternating AT. *Nucl. Acids Res.* **19**, 6725-6730.

Fox, K.R., Davies, H., Adams, G.R., Portugal, J. and Waring, M.J. (1988). Sequence-specific binding of luzopeptin to DNA. *Nucl. Acid Res.* **16**, 2489-2507.

Fox, K.R., Grigg, G.W. and Waring, M.J. (1987). Sequence selective binding of phleomycin to DNA. *Biochem. J.* **243**, 847-851.

Fox, K.R., Olsen, R.K. and Waring, M.J. (1982). Equilibrium and kinetic studies on the binding of des-*N*-tetrmethyltriostin A to DNA. *Biochim. Et Biophys. Acta*, **696**, 315-322.

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. Pharmac.* **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.

Friedmann, T. and Brown, D.M. (1978). Base specific reactions useful for DNA sequencing: methylene blue sensitized photo oxidation of guanine and osmium tetraoxide modification of thymine. *Nucl. Acids Res.* **5**, 615-622

Galas, D.J. and Schmitz, A. (1978). DNAase footprinting-simple method for detection of protein-DNA binding specificity. *Nucl. Acid Res.* **5**, 3157-3170.

Gao, X. and Patel, D.J. (1988). NMR studies of Echinomycin Bisintercalation Complexes with d(A1-C2-G3-T4) and d(T1-C2-C3-A4) duplexes in aqueous solution: Sequence Dependent Formation of Hoogsteen A1.T4 and Watson-Crick T1.A4. Base pairs Flanking the Bisintercalation site. *Biochem.* **27,** 1744-1751.

Gao, X. and Patel, D.J. (1989a). Solution structure of the chromomycin-DNA complex. *Biochem.* **28**, 751-762.

Gao, X. and Patel, D.J. (1989b). Antitumor drug-DNA interactions: NMR studies of echinomycin and chromomycin complexes. *Quater. Rev. Biophys.* **22**, 93-138.

Gavathiotis, E., Sharman, G.J., Searle, M.S. (2000). Sequence-dependent variation in DNA minor groove width dictates orientational preference of Hoechst 33258 in A-tract recognition: solution NMR structure of the 2:1 complex with d(CTTTTGCAAAAG)$_2$. *Nucl. Acids Res.* **28**, No.3, 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 analogue 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**, 4472-4482.

Gilbert, D.A., Van der Marel, G.A., Van Boom, J.H. and Feigon, J., (1989). Unstable Hoogsteen base pairs adjacent to echinomycin binding sites within a DNA duplex. *Proc. Natl. Acad. Sci. U.S.A.,* **86**, 3006-3010.

Gilbert, D.E. and Feigon, J. (1991). The DNA Sequence at Echinomycin Binding Sites Determines the Structural Changes Induced by Drug Binding: NMR Studies of Echinomycin Binding to $[d(ACGTACGT)]_2$ and $[d(TCGATCGA)]_2$.*Biochem.* **30**, 2483-2494.

Gilbert, D.E. and Feigon, J. (1992). Proton NMR study of the $[d(ACGTATACGT)]_2$ - 2 echinomycin complex: conformational changes between echinomycin binding sites. *Nucl. Acids. Res.* **20**, 2411-2420.

Goldberg, I.H., Rabinowitz, M.and Reich, E.(1962). Basis of actinomycin action, I. DNA binding and inhibition of RNA-polymerase synthetic reactions by actinomycin. *Proc. Natl. Acad. Sci. USA* **48**, 2094-2101.

Graves, D.E. and Wadkins, R.M. (1989). 7-acidoactinomycin D a novel probe for examining actinomycin D-DNA interactions. *J. Biol. Chem.* **264**, No 13, 7262-7266.

Haber, F. and Weiss, J. (1934). The catalytic decomposition of hydrogen peroxide by iron salts. *Proc. Roy. Soc. London Ser. A* **147**, 332

Hardenbol, P., Wang, J.C.and Van Dike, M.W. (1997b). Identification of preferred distamycin- DNA binding sites by the combinatorial method REPSA. *Bioconj. Chem.* **8**, 617-620.

Hardenbol, P., Wang, J.C.and Van Dike, M.W. (1997a). Identification of preferred hTBP DNA binding sites by the combinatorial method REPSA. *Nucl. Acids Res.* **25**, 3339-3344.

Hardenbol, P. and Van Dike, M.W. (1996). Sequence specifity of triplex DNA formation: analysis by a combinatorial approach, restriction endonuclease protection selection and amplification. *Proc. Natl. Acad. Sci. USA* **93**, 2811-2816.

Harshman, K.D. and Dervan, P.B.(1985). Molecular recognition of B-DNA by Hoechst 33258. *Nucl. Acids Res.* **13**, No13, 4825-4835.

Hertzberg, R.P., and Dervan, P.B. (1982). Cleavage of double helical DNA by methidiumpropyl-EDTA/Fe(II). *J. Amer. Chem. Soc.* **104**,313-315.

Hoogsteen, K. (1959). The structure of crystals containing a hydrogen-bonded complex of

1-methylthymine and 9-methyladenine *Acta Cristallogra.*, **12**, 822-823.

Hossain, M.B., van der Helm, D., Olsen, R.K., Jones, P.G., Sheldrick,G.M., Egert, E., Kennard, O., Waring, M.J. and Viswamitra, M.A. (1982). Crystal and Molecular Structure of the Quinoxaline Antibiotic Analogue TANDEM (Des-N-tetramethyltriostin A), *J. Am. Chem. Soc.* **104**, 3401-3408.

Huang, C.H., Mong, S. and Crooke, S.T. (1980). Interactions of a new Antitumour antibiotic BBM-928A with deoxyribonucleic acid. Bifunctional intercalative binding studies by fluorimetry and viscometry. *Biochem.* **19**, 5537-5542.

Jeppesen, C. and Nielsen, P.E. (1988). Detection of intercalation-induced changes in DNA structure by reaction with diethyl pyrocarbonate or potassium permanganate. Evidence against the induction of Hoogsteen base pairing by echinomycin. *FEBS Lett.* **231**, 172-176.

Kam, M., Shafer, R.H. and Berman, E. (1988). Solution conformation of the antitumor antibiotic chromomycin A_3 determined by the two-dimensional NMR spectroscopy. *Biochem.* **27**, 3581-3588.

Kamitori, S. and Takusagawa, F. (1992). Crystal structure of the 2:1 complex between d(GAAGCTTC) and the anticancer drug actinomycin D. *J. Mol. Biol.* **225**, 445-456.

Kamitori, S.and Takusagawa, F. (1994). Multiple binding modes of anticancer drug actinomycin-D-X-ray molecular modelling and spectroscopic studies of $d(GAAGCTTC)_2$ - actinomycin D complexes and its host DNA. *J. Am. Chem. Soc.* **116,** No10, 4154-4165.

Katagiri, K., Yoshida, T. and Sato, K. (1975). Mechanism of action and anti-tumour agents. In: Cororan, J.M. and Hahn, F.E. (Eds), Antibiotics III, Springer Verlag, Berlin, 234-251.

Kaziro, Y. and Kamiyama, M. (1965). Inhibition of RNA polymerase reaction by chromomycin A_3. *Biochem. and Biophys. Res. Comm.* **19**, 433-437.

Keniry, M.A., Banville, D.L., Levenson, C. and Shafer, R.H. (1991). NMR investigation of the interaction of mithramycin A with $d(ACCCGGGT)_2$. *FEBS Lett.* **289**, 210-212.

Keniry, M.A., Brown, S.C., Berman, E. and Shafer, R.H. (1987). NMR studies of the interaction of chromomycin A_3 with small DNA duplexes I. *Biochem.* **26**, 1058-1067.

Keniry, M.A., Banville, D.L., Simmonds, P.M. and Shafer, R.H. (1993). Nuclear Magnetic Resonance comparison of the binding sites of mithramycin and chromomycin on the self-complementary oligonucleotide $d(ACCCGGGT)_2$. *J. Mol. Biol.* **231**, 753-767.

Kiang, D.T., Loken, M.K. and Kennedy, B.J. (1979). Mechanism of the hypocalcemic effect of mithramycin. *J.C.E.&M.* **48**, 341-344.

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.

Klevit, R.E., Wemmer, D.E. and Reid, B.R.(1986). ^{1}H NMR studies on the interaction between distamycin A and a symmetrical DNA dodecamer. *Biochem.* **25**, 3296-3303.

Klug, A., Jack, A., Viswamitra, M.A., Kennard, O., Shakked, Z. and Steitz, T.A. (1979). A Hypothesis on a Specific Sequence-dependent Conformation of DNA and its Relation to the Binding of the lac-represor Protein. *J. Mol. Biol.*, **131**, 669-680.

Koo, H.-S, Wu, H.-M and Crothers, D.M. (1986). DNA bending at adenine thymine tracts. *Nature* **320**, 501-506.

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. USA* **82**, 1376-1380.

Kraut, E.H., Fleming, T, Segal, M., Neidhart, J.A., Behrens, B.C., Macdonald, J. (1991). Phase II study of Pibenzimol in pancreatic cancer- a southwest oncology group-study. *Investigational new drugs* **9**, No 1, 95-96.

Krugh, T.R. (1972). Association of actinomycin D and deoxyribonucleotides as a model for binding of the drug to DNA. *Proc. Natl. Acad. Sci. USA* **69**, 1911-1914.

Krugh, T.R. and Reinhardt, C.G. (1975). Evidence for sequence preferences in the intercalative binding of ethidium bromide to dinucleoside monophosphates. *J. Mol. Biol.* **97**, 133-162

Lahm, A. and Suck, D. (1991). DNase I-induced DNA Conformation 2Å Structure of a DNase I-Octamer Complex. *J. Mol. Biol.* **221**, 645-667.

Laskowsky, M. (1971). Deoxyribonuclease I. In: Boyer, P.D. (Ed), The enzymes, Academic Press, London, pp 289-316.

Laughton, C. and Luisi B.(1998). The Mechanics of Minor Groove Width Variation in DNA, and its implications for the Accommodation of Ligands. *J. Mol. Biol.* **288**, 953-963

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.

Lawley, P.D., and Brookes, P. (1963). Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem. J.* **89**, 127-128.

Lee, J.S. and Waring, M.J. (1978a). Bifunctional Intercalation and Sequence Specificity in the Binding of Quinomycin and Triostin Antibiotics to Deoxyribonucleic Acid. *Biochem. J.* **173**, 115-128.

Lee, J.S. and Waring, M.J. (1978b). Interaction between Synthetic Analogues of Quinoxaline Antibiotics and Nucleic Acids. *Biochem J.* **173**, 129-144.

Lee, J.S., Woodsworth, M.L., Latimer, L.J.P. and Morgan, A.R. (1984). Poly(pyrimidine). Poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucl. Acid Res.* **12**, 6603-6614.

Lerman, S. (1961). Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.*, **3**, 18-30.

Liu, C and Chen, F.M.(1994). Oligonucleotide studies of sequence-specific binding of chromomycin A_3 to DNA. *Biochem.* **33**, 1419-1424.

Lomonossoff, G.P., Butler, P.J.G. and Klung, A. (1981). Sequence-dependent variation in the conformation of DNA. *J. Mol. Biol.* **149**, 749-760.

Loontiens, F.G., McLaughlin, L.W., Diekmann,S. And Clegg, R.M.,(1991). Binding of Hoechst 33258 and 4',6-Diamidino-2-phenylindole to Self-Complementary Decadeoxynucleotides with Modified Exocyclic Base Substituents. *Biochem.* **30**, 182-189.

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 DNase I footprinting. *FEBS Lett.* **176**, 414-419.

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 sequence. *Nucl. Acid Res.* **12**, 4865-4879.

Low, C.M.L., Fox, K.R., Olsen, R.K., and Waring, M.J. (1986). DNA sequence recognition by under-methylated analogues of triostin A. *Nucl. Acid Res.* **14**, 2015-2033.

Marchand, C., Bailly, C., McLean, M.J., Moroney, S.E. and Waring, M.J. (1992). The 2-amino group of guanine is absolutely required for specific binding of the anti-cancer antibiotic echinomycin to DNA. *Nucl. Acids Res.* **20**, 5601-5606.

Marini, J.C., Levene, S.D., Crothers, D.M. and Englund, P.T. (1982). Bent helical structure In kinetoplast DNA *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7664-7668. Corrections

(1983) **80**, 7678.

McGhee, J.D. and Felsenfeld, G. (1980). Nucleosome structure. *Annu. Rev. Biochem.* **49**, 1115-1156.

McLean, M.J. and Waring, M.J. (1988). Chemical probes reveal no evidence of Hoogsteen base pairing in complexes formed between echinomycin and DNA in solution. *J. Mol. Recog.* **1**, 138-151.

McClellan, J.A., Palecek, E., Lilley, D.M.J. (1986). $(AT)_n$ tracts embedded in random sequence DNA -formation of a structure which is chemically reactive and torsionally deformable. *Nucl. Acids Res.* **14**, 9291-9309.

Mendel, D. and Dervan, P.B. (1987). Hoogsteen base pairs proximal and distal to echinomycin binding sites on DNA. *Proc. Natl. Acad. Sci. USA* **84**, 910-914.

Miles, E.W. (1977). Modification of histidyl residues in proteins by diethylpyrocarbonate. *Methods Enzymol.* **47**, 431-442.

Mitra. S.N., Wahl, M. C., and Sundaralingam, M. (1999). Structure of the side-by-side binding of distamycin to $d(GTATATAC)_2$. *Acta Cryst.* D**55**, 602-609.

Muller, W. and Crothers, D.M. (1968). Studies of the Binding of Actinomycin and Related Compounds to DNA. *J. Mol. Biol.* **35**, 251-290.

Mundy, G.R., Wilkinson, R. and Heath, D.A. (1983). Comparative study of available medical therapy for hypercalcemia of malignancy. *Am. J. of Medicine* **74**, 421-432.

Nayak, R., Sirsi, M. and Podder, S.K. (1973). Role of Magnesium Ion on the Interaction between Chromomycin A3 and Deoxyribonucleic Acid. *FEBS Lett.* **30**, 157-162.

Neidle S. and Waring, M.J. (1993). Molecular aspects of anticancer drug-DNA interactions, Macmillan, London.

Nielsen, P.E., Jeppesen, C., and Buchardt, O. (1988). Uranyl salts as photochemical agents for cleavage of DNA and probing of protein/DNA contacts. *FEBS Lett.* **235**, 122-124.

Ohkuma, H., Sakai, F., Nishiyama, Y., Ohbayashi, M., Imanishi, H., Konishi, M., Koshiyama, H. and Kawaguchi, H. (1980). BBM-928, A new antitumor antibiotic complex. I. Production, Isolation, Characterization and antitumor activity. *J. Antibiot.* **33**, 1087-1097.

Olsen, R.K., Ramasamy, K., Bhat, K.L., Low, C.M.L. and Waring, M.J. (1986).Synthesis and DNA-Binding studies of [Lac 2, Lac6]TANDEM, an analog of des-N-

tetramethyltriostin A (TANDEM) having a lactic acid substituted for each L-alanine residue. *J. Am. Chem. Soc.* **108**, No19, 6032-6036.

Parks, M.E., Baird, E.E. and Dervan, P.B.(1996). Optimization of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem Soc.* **118**, 6147-6152.

Patel, D.J. (1974). Peptide antibiotic-oligonucleotide interactions. Nuclear magnetic resonance investigations of complex formation between actinomycin D and d-ApTpGpCpApT in aqueous solution. *Biochem.* **13**, 2396-2402.

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. USA* **86**, No15, 5723-5727.

Pelton, J.G., and Wemmer, D.E. (1990) Binding modes of Distamycin A with $d(CGCAAATTTGCG)_2$ Determined by Two-Dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393-1399.

Portugal, J. And Waring, M.J. (1987a). Comparison of binding sites in DNA for berenil, netropsin and dystamycin A footprinting study. *Eur. J. Biochem* **167**, 281-289.

Portugal, J. And Waring, M.J.(1987b). Hydroxyl radical footprinting of the sequence-selective binding of netropsin and dystamicin to DNA. *FEBS Lett.* **225**, No 1,2, 195-200.

Portugal, J. and Waring, M.J. (1988). Assignment of DNA binding sites for 4',6-diamine-2-phenylindole and bisbenzimide (Hoechst 33285). A comparative footprinting study. *Biochim. Biophys. Acta* **949**, 158-168.

Portugal, J. Fox, K.R., Mclean, M.J., Richenberg, J.L. and Waring M.J. (1988). Diethyl pyrocarbonate can detect a modified DNA structure induced by the binding of quinoxaline antibiotics. *Nucl. Acids Res.* **16**, 3655-3670.

Pullman, B. (1983). Electrostatics of Polymorphic DNA. *J. Biomol. Struct. Dynamics* **1**(3), 773-794.

Quintana, J.R., Lipanov, A.A., and Dickerson, R.E.(1991). Low-Temperature Crystallographic Analyses of the Binding of Hoechst 33258 to the Double-Helical DNA Dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *Biochem.* **30**, 10294-10306.

Ray, R., Snyder, R.C., Thomas, S., Koller, C.A. and Miller, D.M.(1989). Mithramycin Blocks binding and Function of the SV40 Early Promoter. *J.Clin. Invest.*, **83** , 2003-2007.

Ream,N.W., Perlia, C.P., Wolter, J. and Taylor, S.G. (1968). Mithramycin therapy in disseminated germinal testicular cancer. *J.A.M.A.* **204**(12), 1030-1036.

Reich, E., Goldberg, I.H., and Rabinowitz, M. (1962). Structure-activity correlations of actinomycins and their derivatives. *Nature* **196**, 743-748.

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. *Biochem.* **34**, 2937-2945.

Rich, A., Nordheim, A. and Wang A.H.-J. (1984). The chemistry and biology of left-handed Z-DNA. *Ann. Rev. Biochem.* **53**, 791-846.

Rosol, T.J. and Capen, C.C. (1987). The effect of low calcium diet, mithromycin, and dichlorodimethylene bisphosphonate on humoral hypercalcimia of malignancy in nude mice transplanted with the canine adenocarcinoma tumor line (CAC-8) *J. Bone and Min. Res.* **2**, 395-405.

Rubin, C.M. and Schmid C.W. (1980). Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucl. Acids Res.* **8**, 4613-4619.

Sambrook, J., Fritsch, E.F. and Maniatis, T.(1989) Molecular cloning a laboratory manual 2nd Edition. Cold Spring Harbor Laboratory Press.

Sastry, M., Patel, D. J.(1993). Solution structure of the mithramycin dimer-DNA complex. *Biochem.* **32,** 6588-

Scamrov, A.V. and Beabealashvilli, R.Sh. (1983). Binding of actinomycin D to DNA revealed by DNase I footprinting. *FEBS Lett.* **164**, 97-101.

Scott. E.V., Zon, G., Marzilli. L.G. and Wilson, W.D. (1988a). 2D NMR investigation of the binding of the anticancer drug actinomycin D to duplex dATGCGCAT: conformational features of the unique 2:1 adduct. *Biochem.* **27**, 7940-7951.

Scott, E.V., Jones, R.L., Banville, D.L., Zon, G., Marzilli, L.G. and Wilson, W.D. (1988b). ^{1}H and ^{31}P NMR investigations of actinomycin D binding selectively with oligodeoxyribonucleotides containing multiple adjacent d(GC) sites. *Biochem.* **27**, 915-923.

Searle, M.S. and Embrey, J.K. (1990) Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ^{1}H NMR spectroscopy. *Nucl. Acids Res.*, **18,** 3753-3762.

Sheldrick, M.G., Guy, J.J., Kennard, O., Rivera, V. and Waring, J.M.(1984). Crystal and

Molecular Structure of the DNA-binding AntitumourAntibiotic Triostin A. *J. Chem. Soc. Perkin trans. II* 1601-1605.

Sigman, D.S., Graham, D.R., D'Aurora, V.,and Stern, A.M. (1979).Oxygen dependent cleavage of DNA by the 1,10-Phenanthroline cuprous complex. *J. Biol. Chem.* **254**, 12269-12271.

Singh, S.B., Ajay, Wemmer, D.E. and Kollman, P.A. (1994). Relative binding affinities of distamycin and its analogue to d(CGCAAGTTGGC)·d(GCCAACTTGCG): comparison of simulation results with experiment. *Proc. Natl. Acad. Sci. USA* **91**, 7673-7677.

Snyder, J.G., Hartman, N.G., D'Estantoit, B.L., Kennard, O., Remeta, D.P. and Breslauer, K.J. (1989). Binding of actinomycin D to DNA: Evidence for a non-classical high-affinity binding mode that does not require GpC sites. *Proc.Natl.Acad.Sci.U.S.A.* **86**, 3968-3972.

Snyder, R.C., Ray, R., Blume, S. and Miller, D.M. (1991). Mithromycin blocks transcriptional initiation of the c-myc P1 and P2 promoters. *Biochem.* **30**, 4290-4297.

Sobell H.M. (1974). The stereochemistry of actinomycin binding to DNA *Cancer Chemotherapy Reports. Part 1* **58**, 101-116.

Suck, D, Oefner, C., and Kabsch, W. (1984). Three-dimentional structure of bovine pancreatic DNAase I at 2.5 A resolution. *EMBO J.* **3**, 2423-2430.

Suck, D. and Oefner, C. (1986). Structure of DNase I at 2.0 A resolution suggest a mechanism for binding to and cutting DNA. *Nature* **321**, 620-625.

Suck, D. (1982). Crystallization and preliminary crystallographic data of bovine pancreatic deoxynbonuclease I. *J. Mol. Biol.* **162**, 511-513.

Suggs, J.W. and Wagner, R.W. (1986). Nuclease recognition of an alternating structure in a $d(AT)_{14}$ plasmid insert. *Nucl. Acid Res.* **14**,3703-3716.

Thiem, J. and Meyer, B. (1979). Studies on the structure of chromomycin A_3 by ^{1}H and ^{13}C nuclear magnetic resonance spectroscopy. *J. Chem. Soc., Perkin Trans. 2,* 1331-1336.

Thiem, J. and Meyer, B. (1981). Studies on the structure of olivomycin A and mithramycin by ^{1}H and ^{13}C nuclear magnetic resonance spectroscopy. *Tetrahedron* **37**, 551-558.

Thuong, N.T. and Helene, C. (1993). Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.* **32**, 666-690.

Timsit, Y. (1999). DNA structure and polymerase fidelity. *J. Mol. Biol.* **293**, No4, 835-853.

Tomita, K., Hoshimo, Y., Sasihara, T. and Kawaguchi, H. (1980). BBM-928 a new antitumour antibiotic complex. II taxonomic studies of the producing organism. *J. Antibiot.* **33**, 1098-1102.

Tullius, T.D. and Dombroski, B.A. (1986). Hydroxyl radical footprinting: high resolution information about DNA-protein contracts and application to repressor and proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 5469-5500.

Tullius, T.D., Dombroski, B.A., Churchill, M.E. and Kam, L. (1987). Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts. *Methods. Enzymol.* **155**, 537-558.

Tullius, T.D. (1987). Chemical snapshots of DNA: using the hydroxyl radical to study the structure of DNA and DNA protein complexes. *TIBS* **12**, 297-300.

Ughetto, G., Wang, A.H.J., Quigley, G.J., van der Marel, G., van Boom, J.H. and Rich, A. (1985). A comparison of the structure of echinomycin and triostin complexes to a DNA fragment. *Nucl. Acids Res.* **13**, 2305-2323.

Urbach, A.R., Szewczyk, J.W., White, S., Turner, J.M., Baird, E.E., and Dervan, P.B.(1999). Sequence selectivity of 3-hydroxypyrrole/pyrrole ring pairings in the DNA minor groove. *J. Am. Chem. Soc.* **121**, 11621-11629.

Van Dyke, M.W. and Dervan, P.B. (1984). Echinomycin Binding Sites on DNA. *Science* **225**, 1122-1127.

Van Dyke, M.W., Hertzberg, R., 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. USA* **79,** 5470-5474.

Van Dyke, M.W. and Dervan, P.B. (1983). Chromomycin, mithramycin, and olivomycin binding sites on heterogeneous deoxyribonucleic acid. Footprinting with (Methidiumpropyl-EDTA) iron(II). *Biochem.* **22**, 2373-2377.

Vega, M. C., Garcia Saez, I., Aymami, J., Eritja, R., Van der Marel, G. A., Van Boom, J. H., Rich, A., 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-

Viswamitra, M.A., Kennard, O., Cruse, W.B.T., Egert, E., Sheldrick, G.M., Jones, P.G., Waring, M.J., Wakelin, L.P.G. and Olsen, R.K. (1981). Structure of TANDEM and its implication for bifunctional intercalation into DNA. *Nature* **289**, 817-819.

Wakelin, L.P.G. and Waring, M.J. (1976). The Binding of Echinomycin to Deoxyribonucleic Acid. *Biochem. J.* **157**, 721-740.

Wang AH, Ughetto G, Quigley GJ, Rich A (1986) : Interactions of quinoxaline antibiotic and DNA: the molecular structure of a triostin A-d(GCGTACGC) complex. *J. Biomol Struct Dyn* **4**(3):319-42.

Ward, D.C., Reich, E. and Goldberg, I.H. (1965). Base specificity in the interaction of polynucleotides with antibiotic drugs. *Science* **149**, 1259-1263.

Waring, M.J. (1970). Variation of the supercoils in close circular DNA by binding of antibiotics and drugs: evidence for molecular models involving intercalation. *J. Mol. Biol.* **54**, 247-279.

Waring, M.J. and Bailly, C. (1994). The purine 2-amino group as a critical recognition element for binding of small molecules to DNA. *Gene* **149**, 69-79.

Waring, M.J. and Wakelin, L.P.G. (1974). Echinomycin : a bifunctional intercalating antibiotic. *Nature* **252**, 653-657.

Wartell, R.M., Larson, J.E. and Wells, R.D. (1974). The compatibility of netropsin and actinomycin binding to natural deoxyribonucleic acid. *J. Biol. Chem.* **250**, 2698-2702.

Waterloh , K. and Fox, K.R. (1991). The effects of actinomycin on the structure of $dA_n \cdot dT_n$ and (dA-dT)n regions surrounding its GC binding site. *J. Biological Chem.* **266**, 6381-6388.

Waterloh , K. and Fox, K.R. (1992). Secondary (non-GpC) binding sites for actinomycin on DNA. *Biochim.Biophys.Acta* **1131**, 300-306.

Waterloh, K., Olsen, R.K. and Fox, K.R. (1992). The bifunctional intercalator [N-MeCys^3, N-MeCys^7]TANDEM binds to the dinucleotide TpA. *Biochem.* **31**, 6246-6253.

Watson, J.D. and Crick, F.H.C. (1953). Molecular Structure of nucleic acids. *Nature* **171**,737-738.

Wells, R.D. and Larson, J.E. (1970). Studies on the binding of actinomycin D to DNA and DNA model polymers. *J.Mol.Biol.* **49**, 319-342.

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.

Wilson, W.D., Jones, R.L., Zon, G., Scott, E.V., Banville, B.L. and Marzilli, L.G. (1986). Actinomycin D binding to oligonucleotides with 5'd(GCGC)3' sequences. Definitive

^{1}H and ^{31}P NMR evidence for two distinct d(GC) 1:1 adducts and for adjacent site binding in a unique 2:1 adduct. *J. Am. Chem. Soc.* **108**, 7113-7114.

Woynarowski, J.M., Sigmund, R.D. and Beerman, T.A. (1989). DNA Minor Groove Binding Agents Interfere with Topoisomerase II Mediated Lesions Induced by Epipodophyllotoxin Derivative VM-26 and Acridine Derivative M-Amsa in Nuclei from L1210 Cells. *Biochem.* **28**, 3850-3855.

Wright, W.E. and Funk, W.D. (1993). CASTing for multicomponent DNA-binding complexes. *Trends Biochem. Sci.* **18**, 77-80.

Wu, M.H. and Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509-513.

Yoshimura, Y., Koenuma, M., Matsumoto, K., Tori, K. and Terui, Y. (1988). NMR studies of chromomycins, olivomycins, and their derivatives. *J. Antibiotics* **41**, 53-67.

Zamenhof, S., Brawerman, G., Chargaff, E. (1952). On the desoxypentose nucleic acids from several microorganisms. *Biochim. Biophys. Acta* **9**, 402-405.

Zhou, N., James, T.L. and Shafer, R.H. (1989). Binding of actinomycin D to $[d(ATCGAT)]_2$: NMR evidence of multiple complexes. *Biochemistry* **28**, 5231-5239.

Appendix A

Computer program

```
// DNA.cpp : Defines the entry point for the console application.
//

#include "stdafx.h"
#include <iostream.h>
#include <fstream.h>
#include <stdlib.h>
char* Ch1;
char* Ch2;
char* Ch3;
char* Ch4;
char* E;
char* NF;
char* Strand[160];
char* CALC[6];



char* QUART[136][4]={
{"C","T","A","C"},
{"T","C","T","T"},
{"T","G","A","T"},
{"C","G","A","A"},
{"T","G","C","A"},
{"G","C","T","T"},
{"C","C","G","T"},
{"G","C","G","C"},
{"C","G","A","C"},
{"G","G","A","C"},
{"A","A","T","T"},
{"C","C","A","T"},
{"C","C","T","G"},
{"T","T","C","T"},
{"C","G","T","G"},
{"C","G","A","G"},
{"T","C","T","A"},
{"C","T","T","G"},
{"T","C","A","T"},
{"C","G","G","G"},
{"C","G","T","C"},
```

```
{"C","T","C","A"},
{"G","A","T","C"},
{"A","T","A","T"},
{"T","G","T","T"},
{"C","A","A","C"},
{"G","C","T","C"},
{"T","G","C","T"},
{"C","T","T","T"},
{"G","T","T","T"},
{"C","A","A","T"},
{"T","G","G","T"},
{"T","A","G","T"},
{"T","T","A","A"},
{"G","A","C","A"},
{"G","G","T","C"},
{"T","C","G","A"},
{"G","G","A","A"},
{"A","G","C","T"},
{"G","T","G","T"},
{"C","A","G","C"},
{"G","G","A","T"},
{"G","C","G","T"},
{"T","G","A","A"},
{"A","G","G","T"},
{"C","G","T","A"},
{"C","C","A","A"},
{"C","T","A","T"},
{"C","C","C","C"},
{"C","T","G","T"},
{"G","G","C","A"},
{"A","T","T","T"},
{"G","C","A","T"},
{"C","T","A","G"},
{"T","G","T","A"},
{"G","G","G","T"},
{"G","T","G","A"},
{"C","A","C","T"},
{"G","T","C","A"},
{"G","A","T","T"},
{"C","C","T","T"},
{"G","T","A","A"},
{"C","C","G","C"},
{"C","G","T","T"},
{"G","C","C","T"},
{"A","C","G","T"},
```

```
{"G","G","C","T"},
{"G","G","G","C"},
{"G","C","A","C"},
{"T","A","T","T"},
{"A","G","A","T"},
{"G","C","A","A"},
{"G","C","T","A"},
{"G","T","T","A"},
{"C","C","T","A"},
{"T","C","G","T"},
{"T","T","G","T"},
{"G","C","C","A"},
{"C","A","G","A"},
{"T","G","G","A"},
{"C","G","G","T"},
{"G","G","T","A"},
{"G","G","G","A"},
{"C","G","G","A"},
{"G","T","A","T"},
{"C","G","C","T"},
{"C","G","C","G"},
{"G","G","T","T"},
{"C","A","T","G"},
{"C","T","T","A"},
{"C","G","C","A"},
{"C","A","A","A"},
{"G","G","C","C"},
{"T","C","C","T"},
{"C","G","C","C"},
{"C","C","A","C"},
{"G","A","G","A"},
{"C","T","A","A"},
{"T","A","T","A"},
{"C","G","A","T"},
{"C","A","T","C"},
{"T","A","A","T"},
{"G","T","T","C"},
{"G","T","C","T"},
{"G","T","A","C"},
{"C","T","T","C"},
{"C","C","C","T"},
{"C","C","A","G"},
{"G","A","A","T"},
{"T","A","C","T"},
{"T","T","T","A"},
```

```
{"G","A","T","A"},
{"C","C","G","A"},
{"A","G","T","T"},
{"A","C","A","T"},
{"C","C","G","G"},
{"C","A","G","T"},
{"C","T","C","C"},
{"C","G","G","C"},
{"C","A","T","T"},
{"C","T","G","A"},
{"G","C","G","A"},
{"G","A","A","A"},
{"C","C","T","C"},
{"A","C","T","T"},
{"C","A","C","C"},
{"G","A","C","T"},
{"T","C","A","A"},
{"C","C","C","A"},
{"C","T","G","C"},
{"C","A","C","A"},
{"C","A","T","A"},
{"T","T","A","T"},
{"G","A","G","T"},
{"T","T","T","T"},
{"C","T","C","T"}
};

char* MIRROR[136][4]={
{"G","T","A","G"},
{"A","A","G","A"},
{"A","T","C","A"},
{"T","T","C","G"},
{"T","G","C","A"},
{"A","A","G","C"},
{"A","C","G","G"},
{"G","C","G","C"},
{"G","T","C","G"},
{"G","T","C","C"},
{"A","A","T","T"},
{"A","T","G","G"},
{"C","A","G","G"},
{"A","G","A","A"},
{"C","A","C","G"},
{"C","T","C","G"},
{"T","A","G","A"},
```

```
{"C","A","A","G"},
{"A","T","G","A"},
{"C","C","C","G"},
{"G","A","C","G"},
{"T","G","A","G"},
{"G","A","T","C"},
{"A","T","A","T"},
{"A","A","C","A"},
{"G","T","T","G"},
{"G","A","G","C"},
{"A","G","C","A"},
{"A","A","A","G"},
{"A","A","A","C"},
{"A","T","T","G"},
{"A","C","C","A"},
{"A","C","T","A"},
{"T","T","A","A"},
{"T","G","T","C"},
{"G","A","C","C"},
{"T","C","G","A"},
{"T","T","C","C"},
{"A","G","C","T"},
{"A","C","A","C"},
{"G","C","T","G"},
{"A","T","C","C"},
{"A","C","G","C"},
{"T","T","C","A"},
{"A","C","C","T"},
{"T","A","C","G"},
{"T","T","G","G"},
{"A","T","A","G"},
{"G","G","G","G"},
{"A","C","A","G"},
{"T","G","C","C"},
{"A","A","A","T"},
{"A","T","G","C"},
{"C","T","A","G"},
{"T","A","C","A"},
{"A","C","C","C"},
{"T","C","A","C"},
{"A","G","T","G"},
{"T","G","A","C"},
{"A","A","T","C"},
{"A","A","G","G"},
{"T","T","A","C"},
```

```
{"G","C","G","G"},
{"A","A","C","G"},
{"A","G","G","C"},
{"A","C","G","T"},
{"A","G","C","C"},
{"G","C","C","C"},
{"G","T","G","C"},
{"A","A","T","A"},
{"A","T","C","T"},
{"T","T","G","C"},
{"T","A","G","C"},
{"T","A","A","C"},
{"T","A","G","G"},
{"A","C","G","A"},
{"A","C","A","A"},
{"T","G","G","C"},
{"T","C","T","G"},
{"T","C","C","A"},
{"A","C","C","G"},
{"T","A","C","C"},
{"T","C","C","C"},
{"T","C","C","G"},
{"A","T","A","C"},
{"A","G","C","G"},
{"C","G","C","G"},
{"A","A","C","C"},
{"C","A","T","G"},
{"T","A","A","G"},
{"T","G","C","G"},
{"T","T","T","G"},
{"G","G","C","C"},
{"A","G","G","A"},
{"G","G","C","G"},
{"G","T","G","G"},
{"T","C","T","C"},
{"T","T","A","G"},
{"T","A","T","A"},
{"A","T","C","G"},
{"G","A","T","G"},
{"A","T","T","A"},
{"G","A","A","C"},
{"A","G","A","C"},
{"G","T","A","C"},
{"G","A","A","G"},
{"A","G","G","G"},
```

```
{"C","T","G","G"},
{"A","T","T","C"},
{"A","G","T","A"},
{"T","A","A","A"},
{"T","A","T","C"},
{"T","C","G","G"},
{"A","A","C","T"},
{"A","T","G","T"},
{"C","C","G","G"},
{"A","C","T","G"},
{"G","G","A","G"},
{"G","C","C","G"},
{"A","A","T","G"},
{"T","C","A","G"},
{"T","C","G","C"},
{"T","T","T","C"},
{"G","A","G","G"},
{"A","A","G","T"},
{"G","G","T","G"},
{"A","G","T","C"},
{"T","T","G","A"},
{"T","G","G","G"},
{"G","C","A","G"},
{"T","G","T","G"},
{"T","A","T","G"},
{"A","T","A","A"},
{"A","C","T","C"},
{"A","A","A","A"},
{"A","G","A","G"}
};

char* PENTGC[32][5] = {

{"C","C","C","C","C"},
{"C","C","C","C","G"},
{"C","C","C","G","C"},
{"C","C","G","C","C"},
{"C","G","C","C","C"},
{"G","G","G","C","G"},
{"G","G","G","G","C"},
{"G","C","G","C","G"},
{"C","G","C","G","C"},
{"G","G","C","G","C"},
{"C","C","G","C","G"},
{"C","G","G","C","G"},
```

```
{"C","C","C","G","G"},
{"C","C","G","G","C"},
{"C","G","G","C","C"},
{"G","C","C","C","C"},
{"G","C","C","C","G"},
{"G","C","C","G","C"},
{"G","C","G","C","C"},
{"G","G","C","C","C"},
{"C","C","G","G","G"},
{"C","G","G","G","C"},
{"G","C","G","G","C"},
{"C","G","C","C","G"},
{"G","G","G","G","G"},
{"C","G","G","G","G"},
{"G","C","G","G","G"},
{"G","G","C","G","G"},
{"C","G","C","G","G"},
{"G","G","C","C","G"},
{"G","G","C","G","C"},
{"G","G","G","C","C"}

};

char* PENTAT[32][5] = {

{"A","A","A","A","A"},
{"A","A","A","A","T"},
{"A","A","A","T","A"},
{"A","A","T","A","A"},
{"A","T","A","A","A"},
{"T","T","T","A","T"},
{"T","T","T","T","A"},
{"T","A","T","A","T"},
{"A","T","A","T","A"},
{"T","T","A","T","A"},
{"A","A","T","A","T"},
{"A","T","T","A","T"},
{"A","A","A","T","T"},
{"A","A","T","T","A"},
{"A","T","T","A","A"},
{"T","A","A","A","A"},
{"T","A","A","A","T"},
{"T","A","A","T","A"},
{"T","A","T","A","A"},
{"T","T","A","A","A"},
```

```
{"A","A","T","T","T"},
{"A","T","T","T","A"},
{"T","A","T","T","A"},
{"A","T","A","A","T"},
{"T","T","T","T","T"},
{"A","T","T","T","T"},
{"T","A","T","T","T"},
{"T","T","A","T","T"},
{"A","T","A","T","T"},
{"T","T","A","A","T"},
{"T","T","A","T","A"},
{"T","T","T","A","A"}

};

char* PYRIMIDINE [32][5] = {

{"C","C","C","C","C"},
{"C","C","C","C","T"},
{"C","C","C","T","C"},
{"C","C","T","C","C"},
{"C","T","C","C","C"},
{"T","T","T","C","T"},
{"T","T","T","T","C"},
{"T","C","T","C","T"},
{"C","T","C","T","C"},
{"T","T","C","T","C"},
{"C","C","T","C","T"},
{"C","T","T","C","T"},
{"C","C","C","T","T"},
{"C","C","T","T","C"},
{"C","T","T","C","C"},
{"T","C","C","C","C"},
{"T","C","C","C","T"},
{"T","C","C","T","C"},
{"T","C","T","C","C"},
{"T","T","C","C","C"},
{"C","C","T","T","T"},
{"C","T","T","T","C"},
{"T","C","T","T","C"},
{"C","T","C","C","T"},
{"T","T","T","T","T"},
{"C","T","T","T","T"},
{"T","C","T","T","T"},
{"T","T","C","T","T"},
```

```
{"C","T","C","T","T"},
{"T","T","C","C","T"},
{"T","T","C","T","C"},
{"T","T","T","C","C"}

};

char * PURINE [32][5] = {

{"A","A","A","A","A"},
{"A","A","A","A","G"},
{"A","A","A","G","A"},
{"A","A","G","A","A"},
{"A","G","A","A","A"},
{"G","G","G","A","G"},
{"G","G","G","G","A"},
{"G","A","G","A","G"},
{"A","G","A","G","A"},
{"G","G","A","G","A"},
{"A","A","G","A","G"},
{"A","G","G","A","G"},
{"A","A","A","G","G"},
{"A","A","G","G","A"},
{"A","G","G","A","A"},
{"G","A","A","A","A"},
{"G","A","A","A","G"},
{"G","A","A","G","A"},
{"G","A","G","A","A"},
{"G","G","A","A","A"},
{"A","A","G","G","G"},
{"A","G","G","G","A"},
{"G","A","G","G","A"},
{"A","G","A","A","G"},
{"G","G","G","G","G"},
{"A","G","G","G","G"},
{"G","A","G","G","G"},
{"G","G","A","G","G"},
{"A","G","A","G","G"},
{"G","G","A","A","G"},
{"G","G","A","G","A"},
{"G","G","G","A","A"}

};

int
```

```
USED [136] ;

// This function will read the first combination from QUART and
// print the first four characters of the strand
        char*
        FirstQUART (int P){
        int N;
                N=P;

                Ch1 = QUART[N][0];
                Ch2 = QUART[N][1];
                Ch3 = QUART[N][2];
                Ch4 = QUART[N][3];
                USED[N] = 1;
                return Ch1, Ch2, Ch3, Ch4;
        }
// This function will read the first combination from MIRROR and
// print the first four characters of the strand
        char*
        FirstMIRROR (int P){
int N;
                N = P - 136  ;

                Ch1 = MIRROR[N][0];
                Ch2 = MIRROR[N][1];
                Ch3 = MIRROR[N][2];
                Ch4 = MIRROR[N][3];
                USED[N] = 1;
                return Ch1, Ch2, Ch3, Ch4;
        }

//This function compares the last three bases of
// the strand to the first three bases of the combination.

        char*
        ThreeBases (char* Ch1, char* Ch2, char* Ch3, char* Ch4){// start ThreeBases
char* Ch5;
char* Ch6;
char* Ch7;
char* Ch8;
int N = 0; // this variable will stop the program once all the combinations have been tried
int J = 0; // this variable will stop the loop if a match is found

                E = "O";
```

```
do {// start do
        // first we check that the combination has not been used
        // during strand formation.


        while (USED[N]==1)
                        N++;



                Ch5 = QUART[N][0];
                Ch6 = QUART[N][1];
                Ch7 = QUART[N][2];
                Ch8 = QUART[N][3];



                // This if statement compares the characters.
        if ((Ch2==Ch5)&&(Ch3==Ch6)&&(Ch4==Ch7)) {// start if

                Ch1=Ch5;
                Ch2=Ch6;
                Ch3=Ch7;
                Ch4=Ch8;
                USED[N] = 1;
                NF="A";
                J=1;
                E="I";


        }// end if

        else {// start else

                Ch5 = MIRROR[N][0];
                Ch6 = MIRROR[N][1];
                Ch7 = MIRROR[N][2];
                Ch8 = MIRROR[N][3];

                        if ((Ch2==Ch5)&&(Ch3==Ch6)&&(Ch4==Ch7)) {//
start if

                                Ch1=Ch5;
                                Ch2=Ch6;
```

```
                                        Ch3=Ch7;
                                        Ch4=Ch8;
                                        USED[N]= 1;
                                        NF="A";
                                        J=1;
                                        E="I";

                                    }// end if

                                    else {// start else

                                            N++;

                                    }// end else
                            }// end else
                           if (N > 135) {// start if
                            NF = "B";
                                    J=1;

                    }// end if

                    }// end do

                    while (J!= 1);

                    CALC[0]=Ch1;
                    CALC[1]=Ch2;
                    CALC[2]=Ch3;
                    CALC[3]=Ch4;
                    CALC[4]=NF;
                    CALC[5]=E;

                    return CALC[6];

}// end ThreeBases
// This function compares the last two characters of the strand
// to the first two characters of the combination.

        char*
                TwoBases (char* Ch1, char* Ch2, char* Ch3, char* Ch4){ // start TwoBases
char* Ch5;
char* Ch6;
char* Ch7;
char* Ch8;
int N = 0; // this variable will stop the program once all the combinations have been tried
```

```
int J = 0; // this variable will stop the loop if a match is found

        E = "O";

        do {// start do
            // first we check that the combination has not been used
            // during strand formation.

            while (USED[N]==1)
                        N++;

                Ch5 = QUART[N][0];
                Ch6 = QUART[N][1];
                Ch7 = QUART[N][2];
                Ch8 = QUART[N][3];

            // This if statement compares the characters.
                if ((Ch3==Ch5)&&(Ch4==Ch6)) { // start if

                 Ch1=Ch5;
                 Ch2=Ch6;
                 Ch3=Ch7;
                 Ch4=Ch8;
                 USED[N]= 1;
                 NF="A";
                 J=1;
                 E="I";

            } // end if

            else {// start else
                Ch5 = MIRROR[N][0];
                Ch6 = MIRROR[N][1];
                Ch7 = MIRROR[N][2];
                Ch8 = MIRROR[N][3];

                        if ((Ch3==Ch5)&&(Ch4==Ch6)) {// start if

                            Ch1=Ch5;
                            Ch2=Ch6;
                            Ch3=Ch7;
                            Ch4=Ch8;
                            USED[N]= 1;
                            NF="A";
                            J=1;
```

```
                                        E="I";

                                    }// end if

                                    else {//start else
                                            N++;
                                    }// end else

                            }// end else

                            if (N > 135) {// start if
                                    NF = "C";
                                    J = 1;

                            }// end if

                    }// end do

                    while (J!= 1);


                            CALC[0]=Ch1;
                            CALC[1]=Ch2;
                            CALC[2]=Ch3;
                            CALC[3]=Ch4;
                            CALC[4]=NF;
                            CALC[5]=E;



                            return CALC[6];

            }// end TwoBases

// This function compares the last character of the strand
// to the first character of the combination.

            char*
            OneBase (char* Ch1, char* Ch2, char* Ch3, char* Ch4){// start OneBase
char* Ch5;
char* Ch6;
char* Ch7;
char* Ch8;
int N = 0; // this variable will stop the program once all the combinations have been tried
int J = 0; // this variable will stop the loop if a match is found
```

```
E="O";
        do { // start do
            // first we check that the combination has not been used
            // during strand formation.

             while (USED[N]==1)
                        N++;


                    Ch5 = QUART[N][0];
                    Ch6 = QUART[N][1];
                    Ch7 = QUART[N][2];
                    Ch8 = QUART[N][3];

                // This if statement compares the characters.
                    if (Ch4==Ch5) { // start if

                     Ch1=Ch5;
                     Ch2=Ch6;
                     Ch3=Ch7;
                     Ch4=Ch8;
                     USED[N]= 1;
                     NF="A";
                     J=1;
                     E="I";

                 }// end if

                else {// start else
                    Ch5 = MIRROR[N][0];
                    Ch6 = MIRROR[N][1];
                    Ch7 = MIRROR[N][2];
                    Ch8 = MIRROR[N][3];

                    if (Ch4==Ch5) { // start if

                                Ch1=Ch5;
                                Ch2=Ch6;
                                Ch3=Ch7;
                                Ch4=Ch8;
                                USED[N]= 1;
                                NF="A";
                                J=1;
                                E="I";
```

```
                    } // end if

                    else { // start else
                                N++;
                          }// end else

               }// end else

               if (N > 135) {// start if
                NF = "D";
                     J=1;

          }// end if

          }// end do

          while (J!= 1);

          CALC[0]=Ch1;
          CALC[1]=Ch2;
          CALC[2]=Ch3;
          CALC[3]=Ch4;
          CALC[4]=NF;
          CALC[5]=E;



               return CALC[6];

     }// end OneBase

// This function takes the first combination available

     char*
          ZeroBase(char* Ch1, char* Ch2, char* Ch3, char* Ch4) {  // Start ZeroBase
char* Ch5;
char* Ch6;
char* Ch7;
char* Ch8;
int N = 0;
int J = 0;

          E="O";

          do {//start do
```

```
                    // first we check that the combination has not been used
                    // during strand formation.

                    while (USED[N]==1)
                                N++;


                            Ch5 = QUART[N][0];
                            Ch6 = QUART[N][1];
                            Ch7 = QUART[N][2];
                            Ch8 = QUART[N][3];



                             Ch1=Ch5;
                             Ch2=Ch6;
                             Ch3=Ch7;
                             Ch4=Ch8;
                             USED[N]=1;
                             NF="A";
                             E="I";
                             J=1;
            } //end do

            while (J!= 1);

            if (N > 136) {// start if
                    NF = "F";
             }// end if

            CALC[0]=Ch1;
            CALC[1]=Ch2;
            CALC[2]=Ch3;
            CALC[3]=Ch4;
            CALC[4]=NF;
            CALC[5]=E;

                    return CALC[6];

    } // end ZeroBase

// This function will analyse the final Strand
//to check that all the combinations are there

void Analysis (int P, int ChrtNum)
```

```
{ //start Analysis

        int N = 0;
        int J = 0;
        int C = 0;
        int K = ChrtNum - 3;
        char* Ch1;
        char* Ch2;
        char* Ch3;
        char* Ch4;
        char* Ch5;
        char* Ch6;
        char* Ch7;
        char* Ch8;


        // Open the file DNAtest.txt

        ofstream OutFile("DNAtest.txt", ios::app);

        OutFile << endl << " strand Num" << " " << P << endl;


                do {// start do

                        Ch1 = Strand[C];
                        Ch2 = Strand[C+1];
                        Ch3 = Strand[C+2];
                        Ch4 = Strand[C+3];


                        do {//start second do

                                Ch5 = QUART[N][0];
                                Ch6 = QUART[N][1];
                                Ch7 = QUART[N][2];
                                Ch8 = QUART[N][3];

                                if
((Ch1==Ch5)&&(Ch2==Ch6)&&(Ch3==Ch7)&&(Ch4==Ch8)){// start if

                                        OutFile << N <<" "<< " "<< " " << "QUART" <<"
"<< " " <<Ch1 << Ch2 << Ch3 << Ch4 << endl;
```

```
                    J=1;

                }// end if

                else {// start else

                    Ch5 = MIRROR[N][0];
                    Ch6 = MIRROR[N][1];
                    Ch7 = MIRROR[N][2];
                    Ch8 = MIRROR[N][3];

                                        if
((Ch1==Ch5)&&(Ch2==Ch6)&&(Ch3==Ch7)&&(Ch4==Ch8)){// start if

                                            OutFile << N <<" "<< " "
<<"MIRROR" <<" "<< " " << Ch1 << Ch2 << Ch3 << Ch4 << endl;
                                            J=1;

                                        }// end if

                                        else {// start else

                                            N++;

                                        }// end else

                }// end else

                if (N > 135) {// start if

                    J=1;

            }// end if
                }// end do

                while (J!= 1);

                C++ ;
                N = 0 ;
                J = 0 ;

            }// end do

            while (C <= K);
```

```
                    OutFile.close();


}// end Analysis

// This function will Analyse the strand for the presence
// of Polypurines and Polypyrimidines
// and all the GC and TA runs present on the strand.

void PolyGATC (int P, int ChrtNum){// start PolyGATC
char* Ch1;
char* Ch2;
char* Ch3;
char* Ch4;
char* Ch5;
char* Ch6;
char* Ch7;
char* Ch8;
char* Ch9;
char* Ch10;
int I = 0;
int N = 0;
int J = 0;
int C = 0;
int K = ChrtNum - 4;

// Open file "Analyser.txt"

ofstream OutFile ("Analyser.txt", ios::app);

        OutFile << "Polypurine and Polypyrimidine test" << endl
                << "strand Num" << " " << P << endl;


        do { // Start first do

                Ch1 = Strand[C];
                Ch2 = Strand[C+1];
                Ch3 = Strand[C+2];
                Ch4 = Strand[C+3];
                Ch5 = Strand[C+4];
```

```
do { // start second do

        // first we check if it is a Polypurine

        Ch6 = PURINE[N][0];
        Ch7 = PURINE[N][1];
        Ch8 = PURINE[N][2];
        Ch9 = PURINE[N][3];
        Ch10= PURINE[N][4];

if
((Ch1==Ch6)&&(Ch2==Ch7)&&(Ch3==Ch8)&&(Ch4==Ch9)&&(Ch5==Ch10))
{// start first if

        OutFile << "This strand contains a Polypurine" << " "
                << Ch1 << Ch2 << Ch3 << Ch4 << Ch5 << endl;
                                        I++;
                                        J=1;

}// end first if

// then we check if it is a Polypyrimidine if is not a
// Polypurine

else {// start first else

                Ch6 = PYRIMIDINE[N][0];
                Ch7 = PYRIMIDINE[N][1];
                Ch8 = PYRIMIDINE[N][2];
                Ch9 = PYRIMIDINE[N][3];
                Ch10= PYRIMIDINE[N][4];

        if
((Ch1==Ch6)&&(Ch2==Ch7)&&(Ch3==Ch8)&&(Ch4==Ch9)&&(Ch5==Ch10))
        {//start second if
                OutFile << "This strand contains a Polypyrimidine" << " "
                << Ch1 << Ch2 << Ch3 << Ch4 << Ch5 << endl;
                                        I++;
                                        J=1;

        }// end second if

        // Now will check for the presence of GC/CG runs
        // only if is not a Polypurine or a Polypyrimidine
```

```
                    else {// start second else

                            Ch6 = PENTGC[N][0];
                            Ch7 = PENTGC[N][1];
                            Ch8 = PENTGC[N][2];
                            Ch9 = PENTGC[N][3];
                            Ch10= PENTGC[N][4];

                            if
((Ch1==Ch6)&&(Ch2==Ch7)&&(Ch3==Ch8)&&(Ch4==Ch9)&&(Ch5==Ch10))
                    {//start third if
                                    OutFile << "This strand contains several CG in a row"
<< " "
                                            << Ch1 << Ch2 << Ch3 << Ch4 << Ch5 <<
endl;
                                                    I++;
                                                    J=1;
                            }// end third if

                            // Now if is none of the above will check for AT/TA runs

                            else {//start third else

                            Ch6 = PENTAT[N][0];
                            Ch7 = PENTAT[N][1];
                            Ch8 = PENTAT[N][2];
                            Ch9 = PENTAT[N][3];
                            Ch10= PENTAT[N][4];

                            if
((Ch1==Ch6)&&(Ch2==Ch7)&&(Ch3==Ch8)&&(Ch4==Ch9)&&(Ch5==Ch10))
                    {//start fourth if
                                    OutFile << "This strand contains several AT in a row"
<< " "
                                            << Ch1 << Ch2 << Ch3 << Ch4 << Ch5 <<
endl;
                                                    I++;
                                                    J=1;
                            }// end fourth if

                            else {// start fourth else

                                    N++;

                            }// close fourth else
```

```
                    }// close third else

                }// close second else

        }// close first else

                if (N>31) {// start fith if

                        J=1;

                }// end fith if

        }// end second do

                while (J != 1);

                        C++ ;
                        N = 0 ;
                        J = 0 ;

                }// end do

                while (C <= K);

                if (I < 2) {// start if

                OutFile.close();

                ofstream OutPut ("FinalResult.txt", ios::app);
                OutPut << P <<" "<< I <<endl;

                OutPut.close();
                }//end if

                } // end PolyGATC

//This function will test the distance between CG/GC. AT/TA

void CGATtest (int P, int ChrtNum) {// start CGAT

char* CH1;
char* CH2;
int Match = 0;
int C = 0; // scans the strand
```

```
int I = 0; // counts the number of outputs to choose
                        //the best result among the strands
int K = ChrtNum - 1; // stops the scaning loop when it
                // reaches the end of the strand
int CG = 0; //Number of CG occurences
int GC = 0; //Number of GC occurences
int AT = 0; //Number of AT occurences
int TA = 0; //Number of TA occurences

int W = 0; //Number of last CG occurence
int X = 0; //Number of last GC occurence
int Y = 0; //Number of last AT occurence
int Z = 0; //Number of last TA occurence

int Difference = 0; // distance between occurences

        // open the file "Analyser.txt"

//      ofstream OutFile ("Analyser.txt", ios::app);

//              OutFile << endl << "CG, GC, AT, TA test" << endl
//                      <<"strand Num" << " " << P << endl;


                do {// start first do

                        CH1 = Strand[C];
                        CH2 = Strand[C+1];



                        if ((CH1=="C")&&(CH2=="G")) { // start first if

                                W++;
                                C++;

                                if (W==1){ // start if

                                        CG = C ;

                                }//end if


                                else {// start else
```

```
                    Difference = (C - 2) - CG;
                    CG = C ;

                    }// end else


                if ((Difference < 5)&&(Difference != 0)){// start if

//                      OutFile << "CG's are only" <<" "<< Difference <<" "
//                              << "bases apart" << endl;
                                I++;

                }// end if

            }// end first if

            else {// start first else

            if ((CH1=="G")&&(CH2=="C")){// start second if

                X++;
                C++;

                if (X==1) {// start if

                    GC = C ;

                }//end if

                else {// start else

                    Difference = (C - 2) - GC;
                    GC = C ;

                }// end else

                if ((Difference < 5)&&(Difference != 0)){ // start if

//                      OutFile << "GC's are only" <<" "<< Difference
<<" "
//                            << "bases apart" << endl;
                              I++;

                }// end if
```

```
}// end second if

else { // start second else

if ((CH1=="A")&&(CH2=="T")) {// start third if

        Y++;
        C++;

        if (Y==1) {// strat if

                AT = C;

        }// end if

        else {// start else

                Difference = (C - 2) - AT;
                AT = C ;

        }// end else

        if ((Difference < 5)&&(Difference != 0)){// start if

//                              OutFile << "AT's are only" <<" "<< Difference
<<" "
//                              << "bases apart"<< endl;
                                I++;

        }// end if

} // end third if

        else {// start third else

                if ((CH1=="T")&&(CH2=="A")) {// start fourth if

                        Z++;

                        C++;

                        if (Z==1){// start if

                        TA = C ;
```

```
                                                    }// end if

                                                    else {// start else

                                                    Difference = (C - 2) - TA;
                                                    TA = C ;

                                                    }// end else

                                                    if ((Difference < 5)&&(Difference !=
0)){// start if

//                                                  OutFile << "TA's are only" <<" " <<
Difference <<" "
//                                                  << "bases apart" << endl;
                                                    I++;

                                        }// end if

                                }// end fourth if
                                        else {// start fourth else

                                                C++;

                                        }// end fourth else

                                }// end third else

                        }// end secon else

                }// end first else

        }// end do

                while (C<=K);

//              OutFile << "Num of CG =" <<" "<< W << endl;
//              OutFile << "Num of GC =" <<" "<< X << endl;
//              OutFile << "Num of AT =" <<" "<< Y << endl;
//              OutFile << "Num of TA =" <<" "<< Z << endl << endl;

//              OutFile.close();


                if (I < 2) {// start if
```

```
                ofstream OutPut ("FinalResult.txt", ios::app);
                OutPut << P <<" "<< I <<endl;

                OutPut.close();

                }// end if

        }// end of CGATtest

void main(){ // start main

int P = 0;
int I = 0;
int ChrtNum;
int N = 0;
int K = 0;

  // Open the file strand.txt

        ofstream OutFile("strand.txt");


        do {//start do
        if (P<= 135) {// start if
                FirstQUART (P);
        }// end if

        else {//start
                FirstMIRROR (P);
        }// end else


// write the first combination to the file


        OutFile << P << endl;
        OutFile <<  Ch1 << Ch2 << Ch3 << Ch4;
        Strand[0] = Ch1;
        Strand[1] = Ch2;
        Strand[2] = Ch3;
        Strand[3] = Ch4;

        ChrtNum = 4;

// Start the comparison of combinations
```

```
E = "O";
NF = "A";
I=0; // I counts the number of combinations that have been used
do {// start do

        if (NF == "A")
                K=1;
        if (NF == "B")
                K=2;
        if (NF == "C")
                K=3;
        if (NF == "D")
                K=4;
        if (NF == "F")
                K=5;

        switch (K)
                {// start switch
        case 1:

                        CALC [6] = ThreeBases(Ch1, Ch2, Ch3, Ch4);
                        Ch1=CALC[0];
                        Ch2=CALC[1];
                        Ch3=CALC[2];
                        Ch4=CALC[3];
                        NF=CALC[4];
                        E=CALC[5];

                        if ((E =="I")&&(ChrtNum<161)){//Start if

                        OutFile <<  Ch4 ;
                        I++;
                        Strand[ChrtNum] = Ch4;
                        ChrtNum = ChrtNum + 1;

                        }// end if

                        if (ChrtNum > 161 ) {// start if
                                OutFile << endl;
                                OutFile <<  "the strand is longrer than 160" << endl;
                                K = 6;
                        }// end if


                        break;
```

```
case 2:

                CALC[6] = TwoBases(Ch1, Ch2, Ch3, Ch4);
                Ch1=CALC[0];
                Ch2=CALC[1];
                Ch3=CALC[2];
                Ch4=CALC[3];
                NF=CALC[4];
                E=CALC[5];

                if ((E =="I")&&(ChrtNum < 160)) {// start if

                OutFile << Ch3 << Ch4;
                I++;
                Strand[ChrtNum] = Ch3;
                Strand[ChrtNum + 1] = Ch4;
                ChrtNum = ChrtNum + 2;

                }// end if

                if (ChrtNum > 161) { //start if
                        OutFile << endl;
                        OutFile << "the strand is longrer than 160" << endl;
                        K = 6;

                }// end if

                break;

case 3:

                CALC[6] = OneBase(Ch1, Ch2, Ch3, Ch4);
                Ch1=CALC[0];
                Ch2=CALC[1];
                Ch3=CALC[2];
                Ch4=CALC[3];
                NF=CALC[4];
                E=CALC[5];

                if ((E =="I")&&(ChrtNum < 159)){// start if

                OutFile << Ch2 << Ch3 << Ch4 ;
        I++;
                Strand[ChrtNum] = Ch2;
                Strand[ChrtNum + 1] = Ch3;
```

```
                Strand[ChrtNum + 2] = Ch4;
                ChrtNum = ChrtNum + 3;

                }// end if

                if (ChrtNum > 161) {// start if
                        OutFile << endl;
                        OutFile << "the strand is longrer than 160" << endl;
                        K = 6;
                }// end if

                break;

        case 4:

                CALC[6] = ZeroBase(Ch1, Ch2, Ch3, Ch4);
                Ch1=CALC[0];
                Ch2=CALC[1];
                Ch3=CALC[2];
                Ch4=CALC[3];
                NF=CALC[4];
                E=CALC[5];

                if ((E =="I")&&(ChrtNum < 158)) {// start if

                OutFile << Ch1 << Ch2 << Ch3 << Ch4;
                I++;
                Strand[ChrtNum] = Ch1;
                Strand[ChrtNum + 1] = Ch2;
                Strand[ChrtNum + 2] = Ch3;
                Strand[ChrtNum + 3] = Ch4;
                ChrtNum = ChrtNum + 4;

                }// end if

                if (ChrtNum > 161 ) { // start if
                   OutFile << endl;
                        OutFile << "the strand is longrer than 160" << endl;
                        K = 6;
                }// end if

                break;

        default:
                OutFile << "No combination found" << endl ;
```

```
                    K=6;


        }// end switch

        if (I > 134)
                K=6;

}// end do

        while (K!=6) ;

        if (ChrtNum < 159){

                for (N = ChrtNum + 1; N < 160; N++) {
                        Strand[N] = "O";

                }// end for

                        //Analysis (P, ChrtNum);
                        // PolyGATC (P, ChrtNum);
                        CGATtest (P, ChrtNum);

        }// end if

        // First we set USED to 0 again before the next strand is formed.

        for (N=0; N<136 ; N++) {// start for

                        USED[N]=0;

                }// end for


       //set I to 0 also

                I=0;

       // we print the character number
                OutFile << endl;
                OutFile << "the number of characters is "<< " "<< ChrtNum << endl;
               P++;
```

```
        } // end do
                while (P<273);
                OutFile.close();
}// end main
```