0026-895X/89/020232-10$02.00/0

MOLECULAR PHARMACOLOGY, 35:232-241

The Binding of Nonintercalative Drugs to Alternating DNA Sequences

FEDERICO GAGO,[1] CHRISTOPHER A. REYNOLDS, and W. GRAHAM RICHARDS
Physical Chemistry Department, Oxford University, Oxford OX1 3QZ, United Kingdom
Received June 24, 1988; Accepted November 14, 1988

SUMMARY

Molecular mechanics methods have been applied to suggest possible models for netropsin and related compounds binding to two different sequences of DNA, namely poly[d(AT)]·poly[d(AT)] and poly[d(GC)]·poly[d(GC)], and to evaluate the different contributions to the binding affinities of these compounds in the ethidium displacement assay. The geometries found after energy refinement suggest that one of the reasons for the selectivity of binding of these agents to A+T-rich DNA regions could be the different widths of the minor groove of the double strand of DNA found in the complexes of these drugs with both DNA sequences.

Netropsin and distamycin A (Fig. 1) are probably the best characterized of a group of minor groove-binding drugs that show a preferential binding to runs of A,T bases over G,C pairs in double-stranded DNA (1, 2). They can interfere with both replication and transcription by blocking the template function of DNA and they show interesting antiviral and antitumor activities (2).

Evidence for the specific recognition between small molecules such as these and a given sequence of DNA is generally obtained from "footprinting" and DNA cleavage inhibition experiments (3). Thus, it has been shown that netropsin and distamycin A preferably bind to runs of a single base (A or T) rather than to the alternating ones, whereas the opposite seems to be true for berenil (4). However, it is known that apparent site sizes and effects of flanking nucleotides not only reflect the extent of true specific binding sites but are also governed by steric disturbances to the DNA helix (4).

The thermodynamic nature of the interaction between these drugs and selected DNA host duplexes can be characterized experimentally by a combination of spectroscopic and calorimetric techniques. Unfortunately, this information is not readily available for many drugs. Netropsin is one compound of this group for which data exist showing that the free energies of binding to both ATAT and poly(dA)·poly(dT) are very similar (−12.7 and −12.2 kcal mol^{-1}, respectively), although with different enthalpic and entropic contributions in each case (5). The binding free energy for the complexation of this drug with GCGC is smaller (−7.1 kcal mol^{-1}) even though the binding entropy is qualitatively similar to that of the alternating A+T copolymer (5).

The ethidium displacement assay has been widely used to gain information about the binding affinity of drugs for a particular DNA and their discriminatory ability for different DNAs (1). This assay provides a C_{50} value for the drug-DNA interaction (defined as the micromolar drug concentration necessary to displace 50% of DNA-bound ethidium) that has been shown to be inversely related to the drug-DNA association constant (6).

Much interesting theoretical work has also been done on the binding of some nonintercalative drugs to homopolymeric AT sequences, notably by Pullman and colleagues (see Ref. 7 for a review), but comparatively little on the binding to the alternating sequences, for which some relevant experimental data exist (1). Some molecular mechanics calculations have shown (8) that, for netropsin, the structures of a complex with a six-base pair oligonucleotide having the sequence AAAAAA and of that with the alternating TATATA are indeed significantly different, providing perhaps an explanation for the different thermodynamics of the association of this antibiotic with poly(dA)·poly(dT) and ATAT. It must be noted, however, that no counterions or explicit solvent molecules were included in this model.

In view of the important differences between netropsin binding to alternating and homopolymeric runs of AT and GC, we have focused our calculations on the binding of a series of nonintercalating agents to the minor groove of ATAT and GCGC. Lacking all the necessary direct information about the binding affinity of these agents for the two sequences studied, we have relied on the C_{50} values from the ethidium displacement

F.G. is grateful to the Consejo Superior de Investigaciones Científicas (Spain) for a postdoctoral grant. The project is in part pursuant to a contract with the National Foundation for Cancer Research.
[1] Permanent address: Departamento de Farmacologia, Universidad de Alcalá de Henares, Madrid, Spain.

ABBREVIATIONS: ATAT, poly[d(AT)]·poly[d(AT)]; QSAR, quantitative structure-activity relationship; MEP, molecular electrostatic potential; RMS, root mean square; LXN, lex-netropsin, the imidazole analogue of netropsin; LXD, lex-distamycin A, the imidazole analogue of distamycin A.

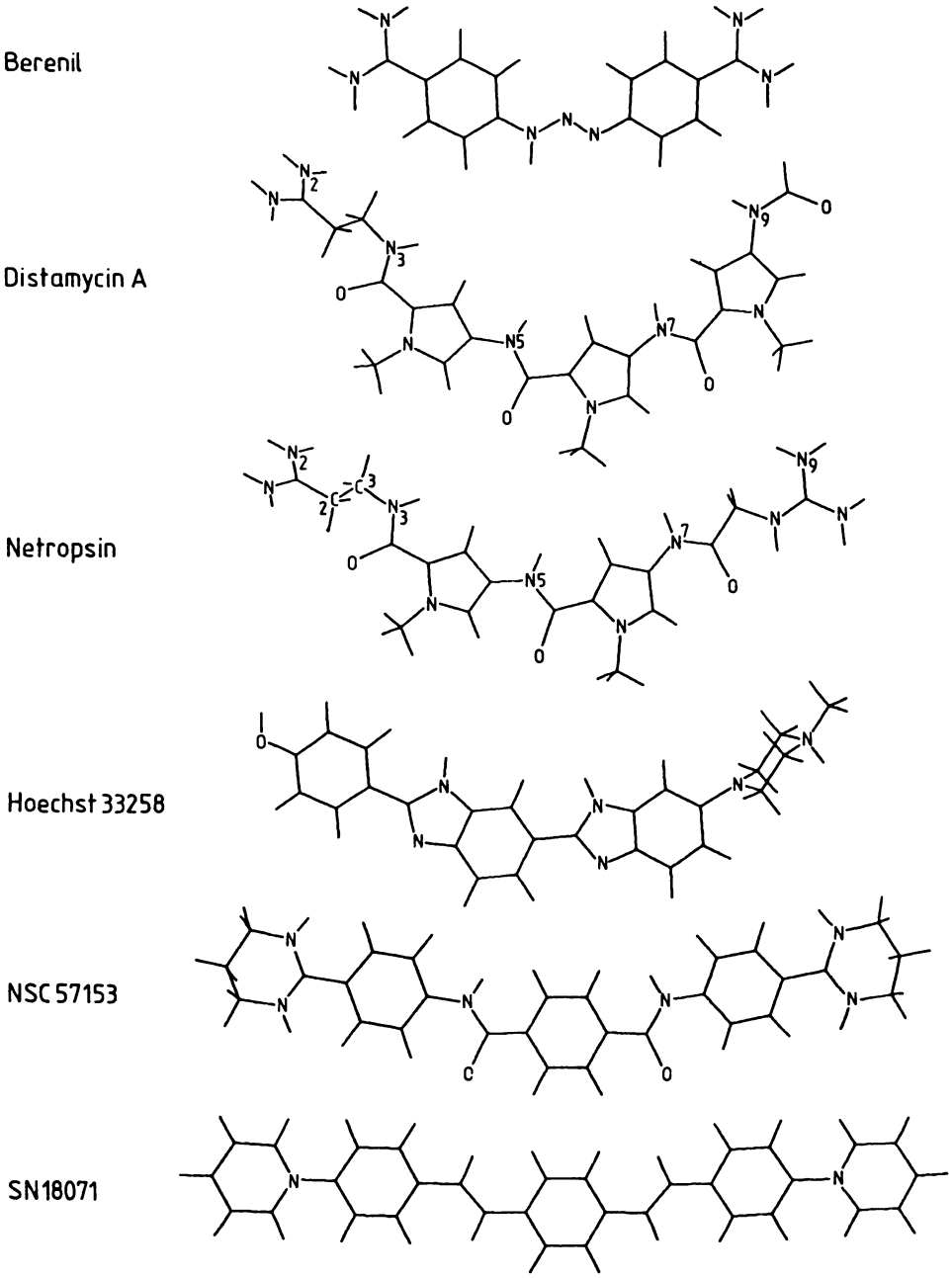

Fig. 1. Optimized structures of some nonintercalative DNA-binding drugs. The atom positions for netropsin and distamycin A relevant to the text have been numbered.

assay (1). A pioneering study on the QSARs for this class of agents had already revealed that the C_{50} values for drug interaction with ATAT incorporated significantly more of the variance in the biological data than those obtained from poly(dA)·poly(dT) or GCGC (9). For this reason, the experimental data published (1) refer to the alternating sequences and not to the homopolymeric ones, which were the subject of previous correlation work (Ref. 7 and references cited therein). Another interesting fact arising from this QSAR work was that the ability of these compounds to discriminate between different DNA polymers, as provided by the ratio of their C_{50} values, is independent of the absolute magnitude of these values (1). This quotient was the major variable in the regression equation that related antitumor activity to physicochemical parameters of a large series of bisquaternary ammonium heterocycles (9) and it proved to be of value in assessing the preference of some of these compounds for A+T-rich DNA. Instead of using these values as a direct measure of affinity (7), we have estimated the discriminatory ability of each agent in terms of the ratio of the C_{50} values obtained for ATAT and GCGC. From the differences in interaction energy calculated for the drugs and the corresponding DNA sequences, it is possible to get a theoretical measure of selectivity. Both sets of values are thus relative and can be compared in a qualitative way.

In this manner, we have attempted to further clarify the principles that govern drug-DNA interactions in the minor groove. A better understanding of these factors can help in the design of agents capable of recognizing a specific DNA sequence, thus providing a basis for selectivity.

Methods

Drugs

Six compounds belonging to the same category of nonintercalative agents (Fig. 1) and showing the same preference for alternating AT

sequences over GC pairs, as measured in the ethidium displacement assay (1), were initially included in this study. They were chosen because they present some distinct features; four of them (berenil, netropsin, NSC 57153, and SN 18071) carry two positive charges whereas distamycin A and Hoechst 33258 have only one positive charge and SN 18071 was specifically designed to contain no hydrogen bonding functions that could account for its selectivity of binding.

There is an interest in the development of G·C recognition elements. In this respect, substitution of imidazole for pyrrole in all or some of the rings of netropsin and distamycin A has been thought to provide the framework for drugs with an improved binding to the G·C-containing oligonucleotides. The term "lexitropsins" ("information-reading" molecules) has been proposed to name these structures (10). In order to assess the validity of this approach, we substituted imidazole for pyrrole in all of the rings of these two antibiotics, which yielded two more molecules that were included in our calculations.

The geometries for berenil (11), netropsin (10, 12), and Hoechst 33258 (13) were created by adding hydrogens to the crystal structures using a molecular modeling system (14) and standard bond lengths and angles. This system was also used to construct the rest of the drug molecules from fragments obtained from the Cambridge Structural Databank.

The crystal structure of netropsin (as the sulfate salt) has long been known (12) and that of a complex of one netropsin molecule with a dodecamer having the sequence CGCGAATT(Br)CGCG has been solved at a resolution of 2.2 Å (10). The main difference between the two structures is that the propylamidinium group of the molecule is rotated by 65° about the C2-C3 bond (Fig. 1) in the sulfate salt, causing it to be almost perpendicular to the plane of the rest of the molecule, whereas in the bound state the molecule is essentially planar. This effect probably arises from the attraction of both cationic terminal residues for the sulfate ions sandwiched between them in the concave part of the molecule (2). Distamycin A and the two lexitropsins were built up having in mind the part of their structures in common with netropsin.

Berenil. From saturation binding studies with calf thymus DNA, a ratio of one molecule of drug bound for every four or five nucleotides was estimated, although the amount of bound drug doubled for synthetic homopolymers (2). Recent footprinting experiments using DNase I and micrococcal nuclease have revealed that berenil recognizes and binds preferably to AT-rich regions that are at least four base pairs long (4).

Netropsin and distamycin A. From DNA cleavage inhibition experiments, it has been estimated that the minimal protected site for both netropsin and distamycin A is four base pairs (15).

Hoechst 33258. This dye has been used as an effective DNA-binding fluorochrome for chromosomal banding and exhibits a fluorescence enhancement upon interaction with DNA that increases with dA·dT content (2). It offers a protection pattern against DNA cleavage similar to that of netropsin and distamycin A, which suggests a binding site size of five ± one base pairs (16). However, in the complex of Hoechst 33258 with a dodecanucleotide of sequence $(CGCGAATTCGCG)_2$, solved by X-ray analysis at a resolution of 2.2 Å (13), there are only three base pairs involved in the central recognition site.

Bisquaternary ammonium heterocycles. Several geometrical isomers are possible for the two conjugated aromatic molecules studied. The *cis*-isomers of NSC 57153 and SN 18071 possess an overall curved shape that allows them to give a better fit within the minor groove than the corresponding *trans*-isomers (Fig. 1). In view of previous calculations that showed us that the *cis* and *trans* forms were energetically almost equivalent, we have considered only the former for the two compounds, which lie closer to a curve of 20 Å radius.

Oligonucleotides

The double-strand DNA structures were generated in the B-conformation with the aid of a program that builds a computerized model restricting all bond lengths and bond angles to those established experimentally for fiber DNA. Both ATAT and GCGC chains consisted of two strands, each with 12 5′-nucleotides. This length is sufficient to avoid end effects in the central region, where the drug is bound. The residues corresponding to A, T, G, and C were numbered according to their position along a given strand in the 5′-3′ direction.

Force Field

In our calculations we followed a molecular mechanics approach making use of the AMBER suite of programs (17). An "all atom" force field representation was used for the small molecules but in the case of the DNA the hydrogens bonded to carbon atoms were not explicitly included ("united atom" approximation) for computational efficiency (18). The drug atoms were assigned the van der Waals and hydrogen bonding parameters of corresponding AMBER atom types; the additional parameters necessary were obtained in accordance with the interpolation method presented by Weiner *et al.* (19). Atomic partial charges for the drug atoms were calculated using the semiempirical molecular orbital method AM1 within the AMPAC program (20). In order to be more consistent with the charges on the DNA atoms, it would have been preferable to fit the charges to *ab initio* MEPs of the drug molecules. But, because of their size, it would have been necessary to split them into fragments and such an approach could have led to considerable error, due to the highly conjugated nature of these systems. Therefore, the electronic distribution was obtained by a single AMPAC run. We have carried out a number of *ab initio* test calculations on small molecule-small molecule interactions in order to study the relative merits of MEP-derived charges and AM1 charges. The total interaction energy was evaluated using the supermolecule approach, with a counterpoise correction. Calculations were then performed in which one molecule was treated quantum mechanically, and the other molecule as a set of point charges that were included as a perturbation in the Hamiltonian operation (21). Generally, better agreement with the full *ab initio* results was obtained for the MEP charges. However, in some cases the AM1 charges gave better agreement. Moreover, the MEP charges tended to overestimate the interaction energy, whereas the AM1 charges tended to underestimate it. There may be, therefore, some advantage in mixing the two sets of charges. Some justification of the validity of the new parameters and charges is also provided by the good agreement between the conformational energy changes for the drugs calculated using molecular mechanics and AMPAC (see below).

Counterions

Although sodium ions could, in principle, enter into close contact with N-3 atoms of adenines and O-2 atoms of thymines, in the newly determined structure of poly(dA)·poly(dT) (22) the octahedral coordination of sodium is undistorted only if the ions are positioned on the side of the minor groove and not within it. Consequently, a monovalent counterion was placed in the plane of each phosphate group and given an enlarged van der Waals radius so as to mimic the effect of hydration. The nonbonded terms for the counterions were: charge = 1, van der Waals radius = 5.0 Å, and well depth = 0.1 kcal mol^{-1}. With these parameters, the ions resemble counterions surrounded by six water molecules (23). A distance-dependent dielectric function was used to reduce the strong electrostatic interactions between charged centers because the solvent was not explicitly included in this model.

Energy Minimization and Model Building

The B-DNA structure was energy refined as follows: firstly, only the counterions were allowed to move, which resulted in their positioning at an average distance of 6.3 Å from the phosphorus and nearly equidistant from two neighboring phosphate groups; the whole macromolecule was then minimized using the conjugate gradient method until convergence (as judged by a RMS gradient of less than 0.1 kcal mol^{-1} $Å^{-1}$). All the small molecules were also energy refined until the RMS gradient was less than 0.1 kcal mol^{-1} $Å^{-1}$.

Docking of the drugs was accomplished by means of the interactive

molecular graphics program HYDRA (24) implemented on a Silicon Graphics IRIS 3120 workstation. For model building, we have relied on the useful information provided by the X-ray structures of the complexes of B-DNA with netropsin (10) and Hoechst 33258 (13).

Experience suggests that the intermolecular space is rich with energy minima for ligands interacting with macromolecules. Alternative models were created by reversing the orientation of the drugs in the minor groove, always favoring as many hydrogen bonding contacts as possible and minimizing the overlap of the van der Waals surface of the drug with that of the oligonucleotide. In this respect, the energy function within HYDRA is especially useful because it allows one to explore a lot of conformational space interactively in a fraction of the time it takes to do an AMBER calculation. The best complexes from this preliminary search were then fully optimized as above.

Results and Discussion

The component analysis of the total energies of these complexes can be seen in Table 1, which shows the nonbonded, electrostatic, and hydrogen bond contributions to the interaction energy.

Conformational Changes upon Binding.

We have obtained a measure of the conformational energy change in the double strand of DNA as the difference in energy of the dodecanucleotides refined in the absence of the drug and the polynucleotide part of the complexes, both in the presence of counterions (ΔE_{DNA}; Table 2). The difference between the energy of the ligands in their free and bound forms (ΔE_{drug}; Table 2) was calculated both by molecular mechanics and by the AM1 molecular orbital method (20) and is a measure of the distortion the molecule has to undergo in order to make favorable interactions with the polynucleotides. The use of the semiempirical method is an additional means of checking that we have calibrated the parameters correctly.

Taking ΔE_{DNA} and ΔE_{drug} together, a measure of the distortion energy, or conformational energy change, is obtained for each complex, representing the increase in energy that the individual molecules undergo in order to form the complex. The binding energy is then obtained making use of the simple equation:

$$\text{Binding Energy} = E_{\text{drug-DNA}} + \Delta E_{DNA} + \Delta E_{drug}$$

where $E_{\text{drug-DNA}}$ is the interaction energy between the ligand and the DNA molecule (Table 1), ΔE_{DNA} is the difference in the energy of the DNA fragment before and after interacting with the drug (Table 2), and ΔE_{drug} is the distortion energy of the ligand calculated with respect to the optimized energy of the free molecule (Table 2, MM). These binding energies are given in Table 3.

It appears that the DNA molecules do not undergo profound conformational changes (ΔE_{DNA} is less than 0.2% of the total energy), in agreement with experimental evidence for nonintercalative agents (2) and previous calculations (7). Both the total energy of each strand and the interaction energy between

TABLE 1

Breakdown of the interaction energy/kcal mol^{-1} of the drugs in the different complexes

ELE is the electrostatic contribution, NB corresponds to the 6–12 nonbonded dispersion-repulsion term, and HB is the corresponding 10–12 term for hydrogen bonding interactions (18). Drug-DNA1 and drug-DNA2 list the values of the total interaction energy of each drug with either the first or the second strand of the dodecanucleotide, respectively; the total energy of interaction of the drugs with each nucleotide is denoted by drug-DNA.

Complex	Strand	NB	ELE	HB	drug-DNA1	drug-DNA2	drug-DNA
				kcal mol^{-1}			
ATAT-Berenil	1	−19.6	−79.7	−1.1	−100.4		−205.6
	2	−22.0	−82.8	−1.3		−105.2	
GCGC-Berenil	1	−19.0	−89.6	−2.1	−110.7		−199.4
	2	−21.3	−66.3	−1.1		−88.7	
ATAT-Distamycin A	1	−42.3	−51.4	−1.0	−94.7		−184.6
	2	−37.1	−51.2	−1.6		−89.9	
GCGC-Distamycin A	1	−36.9	−42.8	−1.2	−80.9		−146.3
	2	−22.3	−42.0	−1.1		−65.4	
ATAT-Hoechst 33258	1	−31.1	−44.4	−0.7	−76.2		−145.6
	2	−36.1	−33.1	−0.2		−69.4	
GCGC-Hoechst 33258	1	−28.8	−34.7	−0.6	−64.1		−128.9
	2	−28.9	−35.6	−0.3		−64.8	
ATAT-LXD	1	−40.0	−58.6	−2.0	−100.6		−179.0
	2	−38.3	−39.0	−1.1		−78.4	
GCGC-LXD	1	−38.9	−49.8	−1.3	−90.0		−163.8
	2	−30.4	−41.3	−2.1		−73.8	
ATAT-LXN	1	−29.1	−93.5	−2.6	−125.2		−244.8
	2	−36.5	−81.5	−1.6		−119.6	
GCGC-LXN	1	−35.4	−88.8	−2.2	−126.4		−234.6
	2	−29.9	−75.8	−2.5		−108.2	
ATAT-Netropsin	1	−33.2	−81.7	−1.4	−116.3		−244.2
	2	−30.2	−95.0	−2.7		−127.9	
GCGC-Netropsin	1	−29.2	−75.3	−2.3	−106.8		−220.3
	2	−26.7	−84.7	−2.1		−113.5	
ATAT-NSC 57153	1	−39.6	−74.9	−0.9	−115.4		−216.5
	2	−35.7	−65.1	−0.3		−101.1	
GCGC-NSC 57153	1	−31.5	−92.8	−0.4	−124.7		−211.3
	2	−29.9	−56.6	−0.1		−86.6	
ATAT-SN 18071	1	−32.1	−63.6	0.0	−95.7		−187.3
	2	−31.2	−60.4	0.0		−91.6	
GCGC-SN 18071	1	−25.1	−57.6	0.0	−82.7		−164.4
	2	−26.7	−55.0	0.0		−81.7	

TABLE 2

Conformational energy change (kcal mol^{-1}) of DNA upon binding of the drug (ΔE_{DNA}) and difference in energy of the different drugs in the optimal complexes with respect to their free state (ΔE_{drug}), calculated by means of molecular mechanics (MM; kcal mol^{-1}) and the AM1 molecular orbital method (AM1; heat of formation, kcal)

	Complexes with ATAT			Complexes with GCGC		
	ΔE_{DNA}	ΔE_{drug}		ΔE_{DNA}	ΔE_{drug}	
		MM	AM1		MM	AM1
	kcal mol^{-1}	kcal mol^{-1}	kcal	kcal mol^{-1}	kcal mol^{-1}	kcal
Berenil	13.4	8.2	12.4	27.5	8.8	13.0
Distamycin A	19.0	3.0	7.1	13.1	6.3	9.8
Hoechst 33258	16.5	1.4	2.7	12.2	1.7	2.9
LXD	14.5	8.2	1.6	19.3	8.8	3.8
LXN	22.9	9.1	4.6	20.4	7.5	4.9
Netropsin	20.8	7.2	13.0	20.8	9.5	10.1
NSC 57153	19.4	2.4	6.0	19.7	3.4	5.9
SN 18071	17.8	3.6	2.1	13.1	3.5	5.0

TABLE 3

Calculated binding energies/kcal mol^{-1} for the drugs in the two complexes

$\Delta E_{ATAT\text{-}GCGC}$ is the difference between the binding energies of the drugs to ATAT and GCGC. The ratio GC/AT expresses the number of times the drug concentration has to be increased with respect to that in the assay on ATAT in order to displace the same amount of ethidium from GCGC (1).

	ATAT Complexes	GCGC Complexes	$\Delta E_{ATAT\text{-}GCGC}$	GC/AT
	kcal mol^{-1}			
Berenil	−184.0	−163.1	−20.9	6.3
Distamycin A	−162.6	−126.9	−35.7	514.3
Hoechst 33258	−127.7	−115.0	−12.7	7.0
LXD	−156.3	−135.7	−20.6	
LXN	−212.8	−206.7	−6.1	
Netropsin	−217.2	−190.0	−27.2	27.3
NSC 57153	−194.7	−188.2	−6.5	5.0
SN 18071	−165.9	−147.8	−18.1	10.5

them become less negative upon binding of the drugs (data not shown). It must be kept in mind that the initial DNA model lacked a water spine of hydration that would undoubtedly stabilize the structure. In the crystal structure of the dodecanucleotide, co-crystallized with netropsin, studied by Kopka *et al.* (10), the minor groove was found to be 0.5–2.0 Å wider than in the nucleotide alone and several water molecules had been displaced by the drug. The same effect was observed in the complex of distamycin A with a different DNA dodecamer (25). In the classical B form the width of the minor groove is 12 Å, but in both the calcium and sodium salts of poly(dA)·poly(dT) it has been reported to narrow down to 9.2 Å (22). In our models, the initial width is also 12 Å, but after the minimization of the complexes we find significant differences between ATAT and GCGC sequences (see below).

Binding Energies

From the data in Table 3 it can be seen that the difference ($\Delta E_{ATAT\text{-}GCGC}$) between the binding energies of these drugs to ATAT and GCGC is always negative, which means that in every case the drug shows preferential binding to the sequence ATAT, which is in agreement with the experimental data. Moreover, according to our calculations, the discriminating ability decreases in the order distamycin A (ΔE = −35.7) > netropsin (ΔE = −27.2) > berenil (ΔE = −20.9) > SN 18071 (ΔE = −18.1) > Hoechst 33258 (ΔE = −12.7) > NSC 57153 (ΔE = −6.5), in good accord with the selectivity these compounds show in the ethidium displacement assay (1). The only discrepancy is found for berenil; its calculated selectivity is much greater than expected from the experimental data. The high value we find for the distortion of DNA in the case of the GCGC complex (Table 2) indicates that this molecule is inducing a considerable change in the DNA molecule. In spite of using counterions to neutralize the charge on the phosphates, the amidine groups of berenil not involved in hydrogen bonding to the base pairs interact strongly with the oxygen atoms of the phosphate backbone. This interaction is particularly strong in the case of GCGC; thence, the deformation inflicted on the DNA. However, the difference in the interaction energy of the drug with the two sequences, as taken from Table 1, is much smaller and of the same order of magnitude as that for NSC 57153, which gives a similar concentration ratio in the ethidium assay (cf. Table 3).

The two lexitropsins studied show a diminished selectivity with respect to the parent antibiotics (ΔE = −20.6 and −6.1 versus −35.7 and −27.2, respectively). Nonetheless, they bind better to ATAT than to GCGC, proving that they can still tolerate the original (5′-ATAT-3′) binding site. In footprinting experiments, LXN appeared to be less specific for AT-rich DNA than was netropsin, the former being able to recognize guanine in the binding sequence (26). In comparison, LXD seems more discriminatory, probably as a consequence of presenting a reduced total cationic charge, which makes the molecule less attracted toward the minor groove. In support of this view, another monocationic imidazole lexitropsin studied by Lown and co-workers (27) has been shown to be highly specific in binding to the sequence 5′-CCGT-3′ (or 5′-ACGG-3′). It is noteworthy that, when LXN was synthesized and tested in the ethidium displacement assay using calf thymus as the source of DNA (26), it showed a relative binding constant nearly identical to that of netropsin (≃ $1.8 \cdot 10^6$ M^{-1}). Indeed, we find that the binding energy is almost the same for both molecules (≃ −215 kcal mol^{-1}). A theoretical study of the specificity of binding of several lexitropsins to poly(dA)·poly(dT) and poly(dG)·poly(dC) has also shown that LXN does not exhibit a preference for this latter nucleotide either (28).

The ability of the imidazole nitrogens of these molecules to form a good hydrogen bond to the exocyclic amino groups of guanine is reflected in Fig. 2.

The structural effects of water were neglected in the energy calculations because of the expensive computations required when dealing with so many water molecules. A distance-dependent dielectric function was used as a means of damping the electrostatic interactions. The desolvation effect upon binding certainly should contribute to the absolute values of the binding energy of each drug to DNA. The drugs studied should interact with water to different extents because they vary in surface area, charge, and hydrogen bonding pattern and, thus the orientation of the water molecules around them is also bound to be different. The final energy values, taking the solvent into account, could alter the relative affinity of these ligands for DNA. This is probably the reason why netropsin seems to bind better than distamycin A to both DNA sequences, in spite of some indication to the contrary (1). This happened to be the case when a supermolecule procedure was used to model the interaction of some of these drugs with water (7); the absolute values of their interaction energy with poly(dA)·poly(dT) were greatly reduced, bringing that of netropsin very

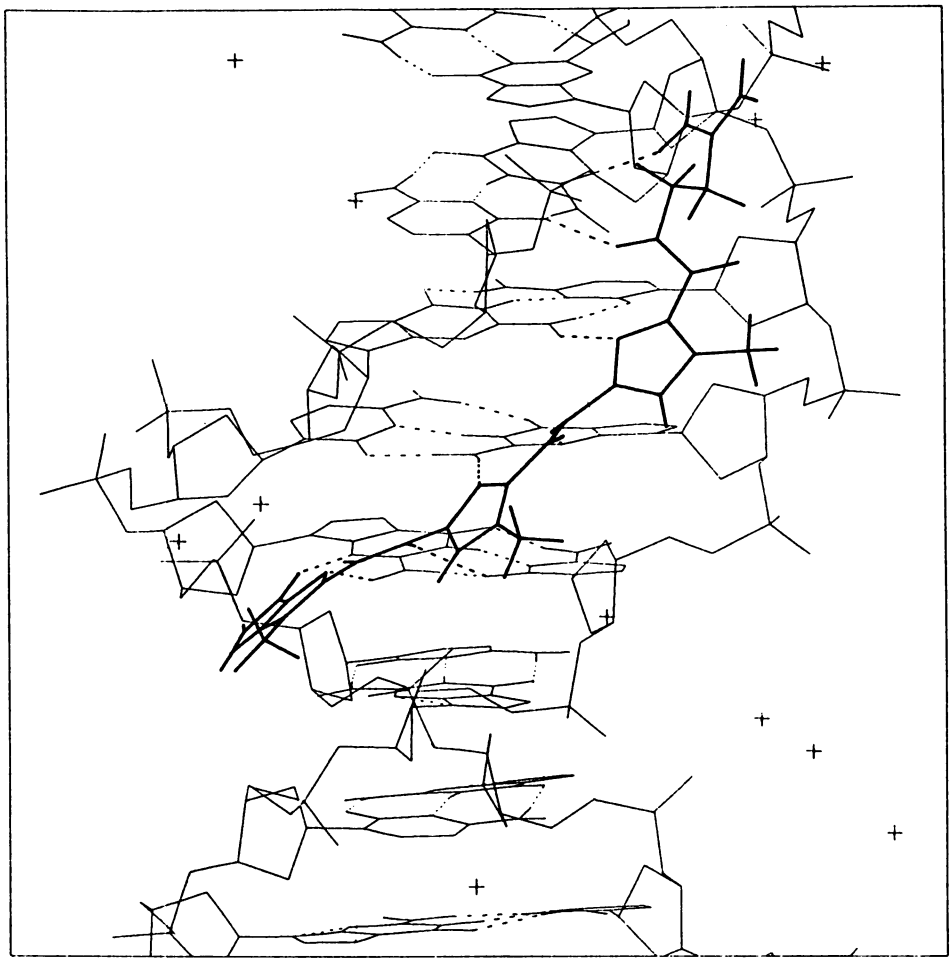

Fig. 2. Detail of the minor groove of DNA in the optimized complex between LXD and poly[d(GC)]·poly[d(GC)]. The drug molecule and the hydrogen bonds to the DNA have been *highlighted*. The counterions are represented as +.

close to the experimental value (5) and below that calculated for distamycin A. But, for a given molecule, this effect is likely to be of the same order of magnitude for both DNA sequences, and the fact that these drugs bind to AT·AT regions in preference to GC·GC pairs will remain unaltered.

Drug Binding to the Oligomeric Duplexes

Berenil. This drug hydrogen bonds to two adjacent thymine O-2 atoms in ATAT. Although in principle it could span three base pairs, the best interaction energy and the minimum energy complex is obtained when it covers only two, in agreement with previous work (11, 29). In the complex with GCGC, in contrast, this positioning is prevented by the amino group at the 2-position in guanine, and the best complex is found to be that in which the drug spans three base pairs. Footprinting experiments show that, in common with the antibiotics discussed below, this drug protects a stretch of at least four base pairs, although it is less tolerant to the presence of bases other than A or T close to the binding site (4). The differences in groove width between the two berenil complexes of lowest energy are not conclusive, although there are appreciable differences when the drug covers three base pairs in both cases.

Distamycin A and netropsin. Comparing the total energy of interaction between the drugs and the polynucleotides, the highest binding energy is found for netropsin and ATAT. This drug sits squarely in the center of the groove, with its pyrrole rings twisted in such a fashion that the shape of the molecule is complementary to the natural curvature of the B-DNA minor groove. The same noncoplanarity can be seen for the antibiotic in the ATAT-distamycin A complex (Fig. 3). The amide nitrogens can form hydrogen bonds to either an adenine N-3 or a thymine O-2 or indeed to both (bifurcated or "three-center" hydrogen bonds). The final geometry shows that the molecules cover a stretch of four base pairs. However, in the crystal structure of the d(CGCAAATTTGCG)$_2$-distamycin A complex (25), this drug covers five of the six A·T base pairs. DNA cleavage of DNA restriction fragments with distamycin-EDTA·Fe(II) and EDTA-distamycin·Fe(II) also indicates that distamycin A recognizes five base pairs (3), the major binding site being the sequence (5′-AAATT-3′). On the other hand, ^{1}HNMR studies (30) have shown that the four-base pair sequence 5′-AATT-3′ is already a perfectly good binding site for distamycin A.

These differences in binding site size might arise from structural modifications in the DNA depending on the sequence. In fact, the oligo(dA)·oligo(dT) stretch in two dodecamers studied containing homopolymeric runs of three (25) or six (31) A·T base pairs has a distinct structural modification, due to a larger than usual propeller twist associated with a system of bifurcated hydrogen bonds.

The hydrogen bond pattern between these two drugs and the minor groove of ATAT found in our models is illustrated diagrammatically in Fig. 4. Although it is generally accepted that these hydrogen bonds help to position the antibiotics properly along the minor groove, it has been proposed that the

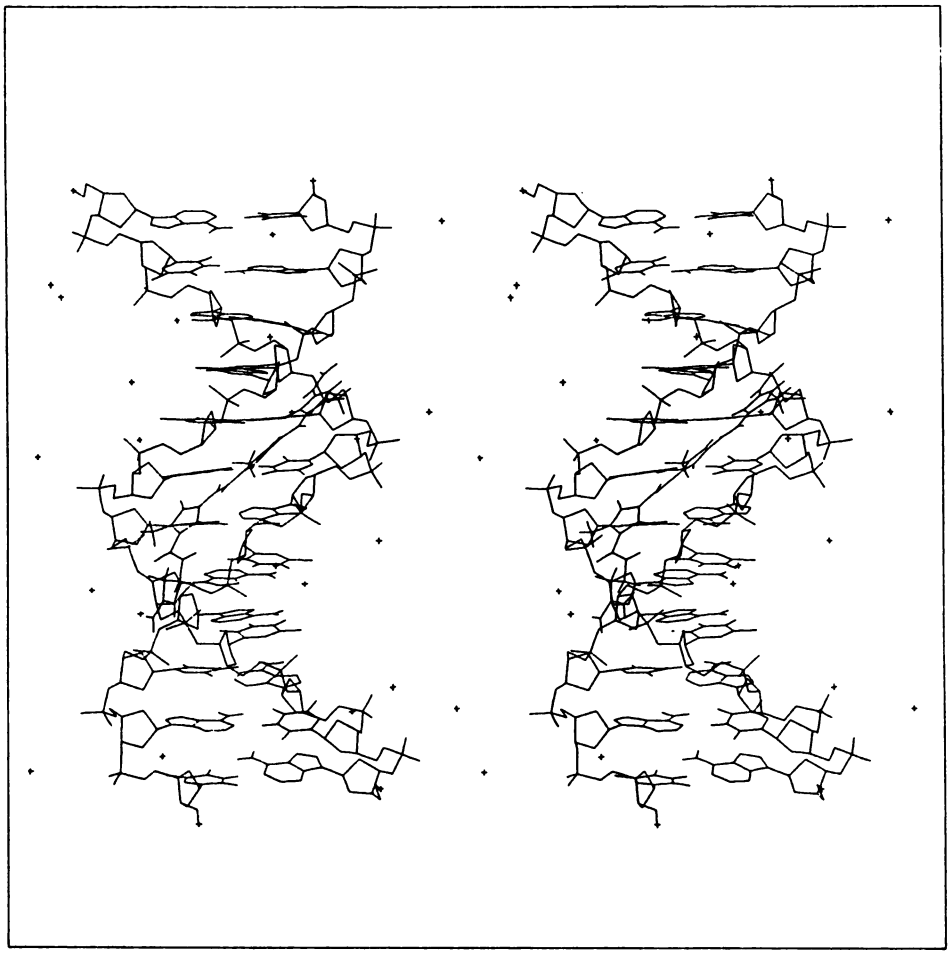

Fig. 3. Stereo view of the optimal complex between distamycin A and poly[d(AT)]·poli[d(AT)]. The counterions are represented as +.

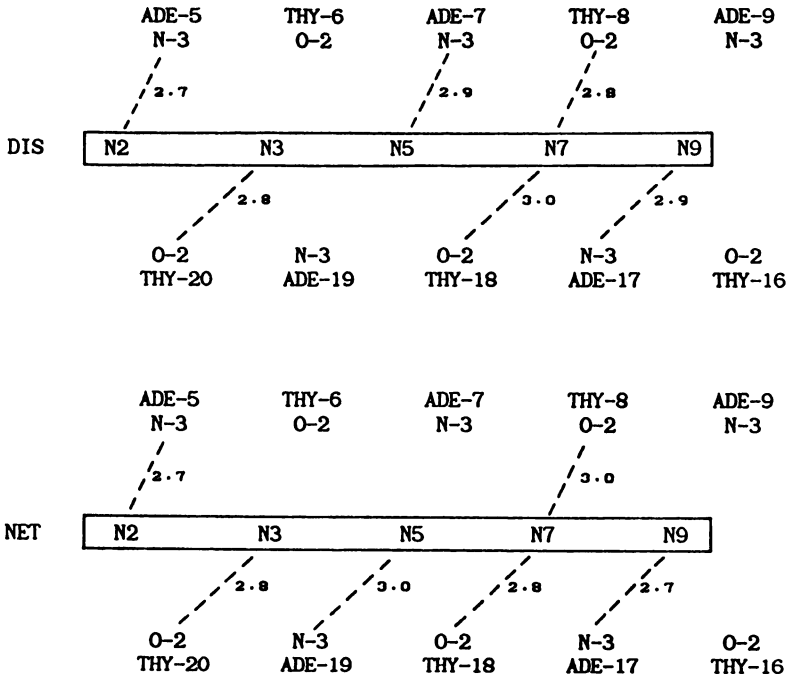

Fig. 4. A schematic representation of the hydrogen bonding between the electronegative atoms in the floor of the minor groove of poly[d(AT)]·poly[d(AT)] and distamycin A (*DIS*) and netropsin (*NET*). The interactions are indicated by *dashed lines* and *numbers* indicate the length/Å of the hydrogen bonds.

actual recognition (or "reading") of base sequences results from the nonbonded van der Walls packing contacts (10). In fact, for distamycin A (30) and netropsin (32), the strongest nuclear Overhauser effects in NMR come from the interactions between pyrrole CH hydrogens and adenine C-2 hydrogens, identifying these protons as the points of nearest contact between the drug and the DNA molecule. These NMR studies have also revealed that this type of drug can interact with the nucleotides in more than one manner.

The minor groove in the central part of the region covered

by the drugs in both ATAT complexes is narrower than in their GCGC counterparts. In the case of the lexitropsin complexes, this width is approximately the same in both sequences.

Hoechst 33258. The piperazine ring of this molecule can be found in two different locations in the crystal structure, related by 180° rotation about the central benzimidazole-benzimidazole bond (13), due to the existence of a conformational equilibrium in the free drug molecule in solution. Initial calculations on our complexes showed that a better interaction energy was obtained when this part of the molecule was buried within the minor groove (location A) and not sticking out of it (location B). In its complex with ATAT, the first benzimidazole NH group of this molecule (that bonded to the hydroxyphenyl ring) makes a bridging three-center hydrogen bond between the Thy-6 and Thy-20 O-2 atoms (2.8 and 2.9 Å, respectively) whereas the second NH hydrogen bonds to Ade-7 N-3 only (2.9 Å). In the GCGC complex, however, the two hydrogen bonds appear on the same strand, between the first NH and the O-2 atom of Cyt-20 (2.8 Å) and the second NH and Gua-19 N-3 (2.9 Å). In spite of these similarities, we find a significant difference between the binding energies of this compound to the two oligonucleotides. It has been argued (13) that the piperazine ring demands the wider groove characteristic of G·C regions, suggesting itself as a possible G·C-reading element for synthetic DNA sequence-reading drug analogues. We notice that this fragment does not preclude the molecule from binding to ATAT in preference to GCGC, although it does open up the minor groove in that region. The main difference between the two complexes seems to arise from the inability of the molecule to bring both strands of the GCGC oligonucleotide together to the same extent as it does with ATAT, which in turn does not give such favorable nonbonded and electrostatic interactions (Table 1). So, even in the presence of hydrogen bonds and a favorable positioning of the flat molecule in the center of the minor groove, the slightly greater twist that the molecule has to undergo to make those hydrogen bonds to GCGC possible, together with the local geometry of the DNA sequence in that region, results in a wider minor groove. The longer distances between the drug atoms and the groups on the floor and walls of this groove give rise to a poorer interaction energy.

NSC 57153. This compound shows strong DNA binding and a certain selectivity for ATAT over GCGC (cf. Table 3). The amide-linked aromatic skeleton has a curved shape that allows it to fit neatly into the minor groove of ATAT and make favorable interactions; it covers at least four base pairs, with one of the central amide nitrogens forming a three-center hydrogen bond to two adjacent O-2 atoms of Thy-8 and Thy-18, and the other one hydrogen bonding to Ade-7 N-3. The ring nitrogens are not sufficiently close to any electronegative base pair atom to be involved in hydrogen bonding. This pattern differs for the complex with GCGC, in which the central NH groups are too far off to make hydrogen bonding possible, but in which one of the ring NH groups is attracted by the sugar O-3′ atom and the other by a phosphate oxygen. Thus, stronger binding to the first strand of GCGC, relative to ATAT, compensate for the weaker binding to the second strand, resulting in a less pronounced difference between the total interaction energies than in the case of the above-mentioned antibiotics (cf. Table 1). The difference in width of the minor grooves in the two complexes is not very pronounced and is greater in the central part of the stretch to which the drug binds. This fact correlates with the lower selectivity of this molecule.

SN 18071. The conjugated system present in the whole length of this molecule makes it the most coplanar of all the drugs studied. It has been shown to be discriminatory toward the two nucleotide sequences considered in this study (1), despite the fact that it lacks any hydrogen-bonding functions. Theoretical calculations have already shown that it binds to the minor groove of poly(dA)·poly(dT) in preference to the major groove or to any of the grooves of poly(dG)·poly(dC) (7). Penetration within the minor groove is very similar for both alternating oligonucleotides and the main difference we find is, again, that the minor groove is somewhat wider in the GCGC complex. This width, measured as the shortest distance between the phosphorus atoms of opposite chains (22), is 9.1, 11.1, 10.7, and 9.3 Å in the central part of the ATAT complex, whereas in its GCGC counterpart these values are 10.3, 11.8, 11.1, and 10.7 Å, respectively.

Effect of the Counterions

The release of counterions must be a substantial contribution to the driving force for the binding of these cationic agents to the polynucleotides (33). We find that the positions of the counterions vary slightly upon binding of the charged drugs to the DNA and substantial differences emerge for the interaction energies between the "hydrated" ions and the drugs in the minor groove. The repulsion term for this interaction is lower for those drugs that bear only one positive charge (distamycin A, Hoechst 33258, and LXTD) and is always greater in the complexes with GCGC than in those with ATAT (data not presented). One possible explanation for this observation is that in the complexes with GCGC the drugs are somewhat more exposed to the surrounding medium, because of the wider minor groove, than in the complexes with ATAT.

Conclusions

Previous theoretical work had already shown that some nonintercalating agents bind better to homopolymeric AT sequences than to GC homopolymeric runs (Ref. 7 and references cited therein). The present results show the preference of another series of these nonintercalative compounds to bind to alternating AT sequences over GCGC sequences. The differences in the interaction energy values between the drugs and the two oligonucleotides satisfactorily correlate with the selectivity ratio provided by experiment.

In the lexitropsins related to netropsin and distamycin A that possess two or three imidazole rings (LXN and LXD, respectively), a decreased preference for the alternating runs of A and T is observed, not because of loss of affinity for ATAT regions but rather as a consequence of their increased binding to the GCGC polynucleotide. The experimentally determined binding affinity for LXN is comparable to that of the natural compound, as can be deduced from the thermal denaturation profiles and the measurements in the ethidium displacement assay (26), in agreement with our calculations and those of others using the homopolymeric AT sequence (28).

It has been reported that the electrostatic potential presents its most negative value in the minor groove of poly(dA)·poly(dT), whereas in poly(dG)·poly(dC) it is most negative in the major groove (34). This difference in the electrostatic potential could be thought of as the origin of the different binding energies of these nonintercalative drugs to these sequences, on the basis that they are all positively charged. But the recent finding that the electrostatic contribution is essen-

tially equal for netropsin binding to both ATAT and GCGC seems to suggest that deep penetration into the minor groove is not required to give rise to a large electrostatic contribution to the binding free energy (5). In fact, we find that netropsin (and the other drugs) are capable of interacting with this latter duplex, the difference in the electrostatic energy term being sometimes small with respect to the same term in the ATAT complexes (cf. Table 1). On the other hand, the van der Waals contacts between adenine C-2 and pyrrole ring CH groups of netropsin, suggested as the origin of the selectivity of binding to A-T nucleotides (10), disappear in the case of the lexitropsins but still these compounds show a high binding energy to ATAT (this study) and to poly(dA)·poly(dT) (28). However, as new hydrogen bonds are now possible between the imidazole nitrogens and the guanine NH_2 groups in the minor groove (Fig. 2), the complexes of these structures with the GCGC sequences are stabilized, and the differences between the binding energies to ATAT and GCGC become smaller (cf. Table 1).

The differences in the width of the minor groove seem to be relevant to the specificity of the interaction of these drugs with the two nucleotides studied. The longer interatomic distances in the case of the GCGC complexes result in a lower value not only in the electrostatic energy term but also in the dispersion-repulsion term, because the van der Waals interactions are calculated as a function of the inverse 6th and 12th powers of the distances (18). Thus, small changes in conformation have a large effect on the interaction energy term between drug and DNA, as well as on the stability of the complex. The different widths permissible in the two types of complexes might provide yet another reason why these drugs bind better to ATAT sequences than to alternating G·C pairs. This difference will probably be more pronounced in the case of the complexes of these drugs with poly(dA)·poly(dT) and poly(dG)·poly(dC) and is suggested as one of the reasons why this class of nonintercalative compounds bind better to A+T-rich DNA.

Our models must ultimately be confirmed by more accurate physical measurements, but even in the absence of confirmation they can be used in drug development. The fact that the differences in the values of interaction energy calculated for the complexes of a series of nonintercalative agents with two DNA sequences correlate with their relative binding affinities determined experimentally is encouraging. More importantly, this study provides some evidence that the conformation of the DNA in the region where the drugs bind, and more specifically the width of the minor groove, can be different depending on the sequence. Although we cannot guarantee that we have found the global minimum for any of these complexes, this difference is a constant trend found in all the minimized geometries investigated.

The ultimate goal of designing a small molecule that will bind to a predetermined sequence in a highly specific manner will be accomplished only if all these factors are taken into account and a better knowledge of the local structure of a given stretch of the DNA helix is gained. In this respect, the idea of designing "A·T-rejecting elements" (13) as well as "G·C-reading" elements (10) seems to us particularly important.

References

1. Baguley, B. C. Nonintercalative DNA-binding antitumour compounds. *Mol. Cell. Biochem.* **43**:167–181 (1982).
2. Zimmer, C., and U. Wahnert. Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog. Biophys. Mol. Biol.* **47**:31–112 (1986).
3. Dervan, P. B. Design of sequence-specific DNA-binding molecules. *Science (Wash. D. C.)* **232**:464–471 (1986).
4. Portugal, J., and M. J. Waring. Comparison of binding sites for berenil, netropsin and distamycin: a footprinting study. *Eur. J. Biochem.* **167**:281–289 (1987).
5. Marky, L. A., and K. J. Breslauer. Origins of netropsin binding affinity and specificity: correlation of thermodynamic and structural data. *Proc. Natl. Acad. Sci. USA* **84**:4359–4363 (1987).
6. Baguley, B. C., and E.-M. Falkenhang. The interaction of ethidium with synthetic double-stranded polynucleotides at low ionic strength. *Nucleic Acid Res.* **5**:161–171 (1978).
7. Pullman, B. Carcinogens, antitumor agents and DNA. *Pontif. Acad. Sci. Scr. Varia* **70**:3–31 (1986).
8. Caldwell, J., and P. A. Kollman. A molecular mechanical study of netropsin-DNA interactions. *Biopolymers* **25**:249–266 (1986).
9. Denny, W. A., G. J. Atwell, B. C. Baguley, and B. F. Cain. Potential antitumour agents. 29. Quantitative structure-activity relationship for the antileukemic bisquaternary ammonium heterocycles. *J. Med. Chem.* **22**:134–150 (1979).
10. Kopka, M. L., C. Yoon, D. Goodsell, P. Pjura, and R. E. Dickerson. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. USA* **82**:1376–1380 (1985).
11. Pearl, L. H., J. V. Skelly, B. D. Hudson, and S. Neidle. The crystal structure of the DNA-binding drug berenil: molecular modelling studies of berenil-DNA complexes. *Nucleic Acid Res.* **15**:3469–3478 (1987).
12. Berman, H. M., S. Neidle, C. Zimmer, and H. Thrum. Netropsin, a DNA-binding oligopeptide: structural and binding studies. *Biochim. Biophys. Acta* **561**:124–131 (1979).
13. Pjura, P. E., K. Grzeskoniak, and R. E. Dickerson. Binding of Hoechst 33258 to the minor groove of B-DNA. *J. Mol. Biol.* **197**:257–271 (1987).
14. Ricketts, D. M. PERQ interactive molecular modelling system. *J. Mol. Graphics* **5**:63–70 (1987).
15. Van Dycke, M. W., R. P. Hertzberg, and P. B. Dervan. 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 (1982).
16. Harshman, K. D., and P. B. Dervan. Molecular recognition of B-DNA by Hoechst 33258$^+$. *Nucleic Acid Res.* **13**:4825–4835 (1985).
17. Singh, U. C., S. J. Weiner, J. W. Caldwell, and P. A. Kollman. *AMBER (UCSF): Assisted Model Building with Energy Refinement*, version 3.0. Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco (1986).
18. Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**:765–784 (1984).
19. Weiner, S. J., P. A. Kollman, D. T. Nguyen, and D. A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**:230–252 (1986).
20. Dewar, M. J. S., and J. J. P. Stewart. *Quantum Chem. Program Exc. (QCPE) Bull.* **6**:24a–24b (1986).
21. Reynolds, C. A., W. G. Richards, and P. J. Goodford. Prediction of selective bioreductive anti-tumour, anti-folate activity using a modified *ab initio* method for calculating enzyme-inhibitor interaction energies. *J. Chem. Soc. Perkin Trans. II* 551–556 (1988).
22. Alexeev, D. G., A. A. Lipanov, and I. Y. Skuratovskii. Poly(dA)·poly(dT) is a B-type double helix with a distinctively narrow minor groove. *Nature (Lond.)* **325**:821–823 (1987).
23. Singh, U. C., S. J. Weiner, and P. A. Kollman. Molecular dynamics simulations of d(C-G-C-G-A)·d(T-C-G-C-G) with and without "hydrated" counterions. *Proc. Natl. Acad. Sci. USA* **82**:755–759 (1985).
24. Hubbard, R. *HYDRA (Harvard York Drawing Program).* Chemistry Department, University of York, Heslington, York, England (1987).
25. Coll, M., C. A. Frederick, A. H.-J. Wang, and A. Rich. A bifurcated hydrogen-bonded conformation in the d(A·T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. USA* **84**:8385–8389 (1987).
26. Lown, J. W., K. Krowicki, U. G. Bhat, A. Skorobogaty, B. Ward, and J. C. Dabrowiak. Molecular recognition between oligopeptides and nucleic acids: novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity. *Biochemistry* **25**:7408–7416 (1986).
27. Kissinger, K., K. Krowicki, J. C. Dabrowiak, and J. W. Lown. Molecular recognition between oligopeptides and nucleic acids: monocationic imidazole lexitropsins that display enhanced GC sequence-dependent DNA binding. *Biochemistry* **26**:5590–5595 (1987).
28. Zakrzewska, K., R. Lavery, and B. Pullman. A theoretical study of the sequence specificity in binding of lexitropsins to B-DNA. *J. Biomol. Struct. Dyn.* **4**:833–843 (1987).
29. Gresh, N., and B. Pullman. A theoretical study of the non-intercalative binding of berenil and stilbamidine to double-stranded (dA-dT)$_n$ oligomers. *Mol. Pharmacol.* **25**:452–455 (1984).
30. Klevit, R. E., D. E. Wemmer, and B. R. Reid. ^{1}H NMR studies on the

interaction between distamycin A and a symmetrical DNA dodecamer. *Biochemistry* **25:** 3296–3303 (1986).
31. Nelson, H. C. M., J. T. Finch, B. F. Luisi, and A. Klug. The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature (Lond.)* **330:**221–226 (1987).
32. Patel, D. J. Antibiotic-DNA interactions: intermolecular nuclear Overhauser effects in the netropsin-d(CGCGAATTCGCG) complex in solution. *Proc. Natl. Acad. Sci. USA* **79:**6424–6428 (1982).
33. Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophysics* **11:**179–246 (1978).
34. Lavery, R., and B. Pullman. The dependence of the surface electrostatic potential of B-DNA on environmental factors. *J. Biomol. Struct. Dyn.* **2:**1021–1032 (1985).

Send reprint requests to: Dr. Federico Gago, Physical Chemistry Laboratory, Oxford University, South Parks Road, Oxford, OX1 3QZ, UK.