# Sequence Specificity of the Binding of 9-Aminoacridine- and Amsacrine-4-carboxamides to DNA Studied by DNase I Footprinting†

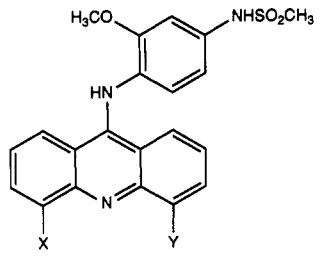
C. Bailly,‡,§ W. A. Denny,∥ L. E. Mellor,‡ L. P. G. Wakelin,⊥ and M. J. Waring*,‡

*Department of Pharmacology, University of Cambridge, Cambridge CB2 1QJ, U.K., St. Luke's Cancer Research Fund, Dublin 6, Ireland, and Cancer Research Laboratory, University of Auckland School of Medicine, Private Bag, Auckland, New Zealand*

*Received September 24, 1991; Revised Manuscript Received January 8, 1992*

ABSTRACT: DNase I footprinting has been used to probe the sequence selectivity of binding of a series of intercalating amsacrine-4-carboxamides and a related 9-aminoacridine-4-carboxamide to three DNA restriction fragments. These ligands have good experimental antileukemic activity, and for those members of the series that gave evaluable footprints, our principal finding is that they bind preferentially to GC-rich regions in agreement with the conclusion of equilibrium and kinetic measurements. The highest affinity sites generally occur in clusters of GC base pairs with runs of AT pairs being excluded from binding. It is important to appreciate that the 9-aminoacridine- and amsacrine-4-carboxamides exhibit a very high degree of selectivity for GC sites which, to our knowledge, has not been previously matched by acridine derivatives in footprinting experiments. The principal determinant of specificity appears to be the 4-carboxamide group itself since neither variations in the terminal funtionality of the 4-carboxamide sidechain nor the presence of the 9-anilino substituent modifies sequence preferences. The molecular origins of selectivity may be discerned in terms of potential hydrogen bonding interactions between the 4-carboxamide moiety and carbonyl oxygen and amino groups of GC base pairs in the DNA minor groove at CG dinucleotide sites. The related therapeutic agent amsacrine failed to inhibit cleavage by DNase I, so no conclusion can be drawn concerning its binding selectivity, save to note that amsacrine does not possess the 4-carboxamide group which appears to be the crucial determinant of GC specificity. Whether selectivity for binding to GC-rich sequences is an important element in the antitumor activity of both the 9-aminoacridine- and amsacrine-4-carboxamides remains to be determined.

The clinically-useful antileukemia drug amsacrine (**1**) (Arlin, 1983) is the most prominent member of a large class of 9-anilinoacridines (Denny et al., 1982) that have been widely




**1**: X = Y = H
**2**: X = $CH_3$, Y = $CONHCH_3$

studied for their antitumor effects, for their activity as topoisomerase II poisons (Nelson et al., 1984; Liu, 1989), and for their capacity to bind to DNA. Anilinoacridines bind moderately tightly to double-stranded DNA: the affinity constant of amsacrine for calf thymus DNA is $1.5 \times 10^5$ $M^{-1}$ at ionic strength 0.01 (Wilson et al., 1981) in an enthalpy-driven process (Wadkins & Graves, 1989; Graves & Wadkins, 1990). These agents share the well-known propensity of acridines to intercalate, the helix unwinding angle of amsacrine being 20.5° (Waring, 1976). Crystal structures of 9-anilinoacridines, including amsacrine, show that the plane of the phenyl group is oriented almost orthogonally (about 70°) to the plane of the acridine ring (Karle et al., 1980; Neidle et al., 1986). On the basis of the findings of crystallographic studies of 9-aminoacridine complexed to RNA dinucleotides (Sakore et al., 1979), it has been widely speculated (Wilson et al., 1981; Wadkins & Graves, 1989) that the anilino ring of amsacrine lies in the DNA minor groove when the acridine chromophore is fully intercalated, although the opposing view based on kinetic arguments has also been taken (Denny & Wakelin, 1986). Molecular and quantum mechanical calculations support location of the anilino ring in the minor groove (Chen et al., 1988; Graves & Wadkins, 1990). Simple 9-anilinoacridines dissociate very rapidly from DNA (Denny & Wakelin, 1986), and this kinetic lability has severely restricted information concerning the nucleotide sequence preferences of amsacrine binding. Hence, to date no footprinting studies of amsacrine–DNA complexes have been reported (see below). The limited sequence preference information that is available comes from equilibrium binding studies, several of which (Wilson et al., 1981; Wadkins & Graves, 1989) indicate little, if any, difference in affinity for amsacrine binding to the alternating copolymers poly(dA-dT)·poly(dA-dT) and poly(dG-dC)·poly(dG-dC). Complementing this lack of knowledge about the sequence selectivity of amsacrine, there is a paucity of detailed structural information about the amsacrine–DNA complex since the latter has not yet yielded to X-ray crystallographic analysis and it dissociates too rapidly for con-

† This work was supported by grants from the Cancer Research Campaign, the Medical Research Council, and the Royal Society to M.J.W. and from the Auckland Division of the Cancer Society of New Zealand and the Medical Research Council of New Zealand to W.A.D. C.B. acknowledges the support of INSERM.

‡ University of Cambridge.
§ Present address: INSERM U16, Place de Verdun, 59045 Lille, France.
∥ University of Auckland School of Medicine.
⊥ St. Luke's Cancer Research Fund.

0006-2960/92/0431-3514$03.00/0 

formational studies by two-dimensional NMR spectroscopy.

Many active antitumor agents of the DNA-intercalating class have one or more side chains appended to the intercalating chromophore, and it has been suggested (Feigon et al., 1984; Muller & Crothers, 1968; Wakelin et al. 1987) that an important effect of such side chains is to reduce the rate of dissociation of the ligand–DNA complex, thereby increasing the residence time of the ligand at particular binding sites. In the majority of cases these side chains appear to lie in the DNA minor groove rather than the major groove. Attaching additional side chains to the chromophore of amsacrine modifies the kinetic and thermodynamic characteristics of anilinoacridine–DNA complexes and alters biological behavior. For example, the 5-methyl-4-(methylcarbamoyl) derivative of amsacrine (compound **2**) has activity against experimental solid tumors and is currently in clinical trial (Hardy et al., 1988). A study of the DNA binding kinetics of this class of compound, the amsacrine-4-carboxamides (including compounds **4–8**; Table I), together with the "parent" compounds amsacrine (**1**) and the 9-aminoacridine-4-carboxamide **3** has recently been reported (Wakelin et al., 1990; Wakelin & Denny, 1990). This investigation led to the suggestion that, at equilibrium, the amsacrine-4-carboxamides bind with the acridine chromophore in a position of maximum overlap with the base pairs, and with the carboxamide side chain lying in the DNA minor groove and the bulky anilino chain in the major groove. The association kinetics for compounds with complex carboxamide side chains are slow, which is compatible with the need to thread the side chain through a transient opening in the double helix associated with the natural "breathing" motions of DNA. The dissociation kinetics of these agents are strongly side-chain structure-dependent, with several analogues [particularly **5** and **6**, having a C(O)NH-$(CH_2)_2$NRR side chain] dissociating much more slowly than amsacrine. Both equilibrium and kinetic measurements show that the amsacrine-4-carboxamides bind selectively to poly(dG-dC)·poly(dG-dC) compared to poly(dA-dT)·poly(dA-dT), and the kinetic data for binding to calf thymus DNA imply that the highest affinity sites in natural sequence DNA are GC-rich (Wakelin et al., 1990; Wakelin & Denny, 1990).

The enhanced kinetic stability of the DNA complexes of the amsacrine-4-carboxamides prompted us to explore their binding sequence preferences in more detail using the DNase I footprinting technique, and we report the results of this work here. Footprinting involves probing to single base pair resolution drug-induced inhibition of cleavage of defined-sequence DNA fragments by enzymatic or chemical means, and the method depends in part on the relative persistence of ligands at different binding sites. Since simple intercalators often dissociate rapidly from DNA, their optimal binding sites can be difficult to detect by standard footprinting methods. By contrast, the family of compounds studied here generally give good DNase I footprinting patterns. The DNA substrate chosen was the *tyr* T fragment from *Escherichia coli* which has been employed in many previous investigations of drug–DNA binding specificity (Low et al., 1984a,b; Fox & Waring, 1984, 1986). For comparative purposes, additional experiments were performed with two unrelated DNA fragments isolated from a different source, thereby providing an assessment of the sequence selectivity of the drugs with respect to a wide variety of potential binding sites. Since there is a large body of unusually detailed information concerning the kinetic and equilibrium DNA-binding properties of this closely-related series of agents, we measured the effects of the six compounds over a substantial concentration range, so that some impression of their relative site-specific affinities might be gained. We discuss the relevance of our findings both with respect to the mechanism and sequence selectivity of binding of DNA threading agents and with respect to their implications for the sequence selectivity of amsacrine.

## Materials and Methods

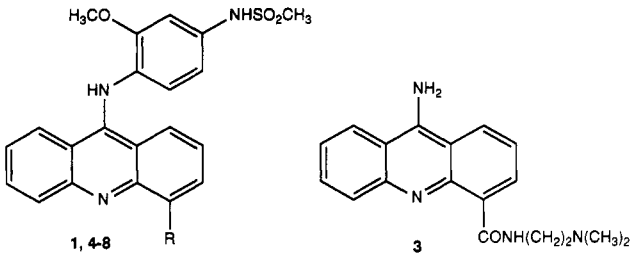
*Ligands.* Amsacrine (**1**), the 9-aminoacridine-4-carboxamide **3**, and the amsacrine-4-carboxamides **4–8** were prepared as previously described (Cain et al., 1977; Denny et al., 1982). The compounds were used as isethionate or hydrochloride salts and were dissolved in water (10 mM stock solution). Subsequent dilutions were made in 10 mM Tris–10 mM NaCl buffer, pH 7.0.

*DNA Fragments.* A 160 bp duplex DNA fragment containing the tyrosine tRNA promoter (*tyr* T fragment) was isolated from the plasmid pKMΔ-98 and 3′-labeled on the Watson strand (with [$\alpha$-$^{32}$P]dCTP) or on the Crick strand (with [$\alpha$-$^{32}$P]dATP) as previously described (Drew & Travers, 1984; Low et al., 1984a,b). Two other fragments of 133 and 253 bp were obtained from the plasmid pBS (Stratagene, La Jolla, CA) on digestion with the restriction enzymes *Ava*I and *Pvu*II and were 3′-end-labeled with [$\alpha$-$^{32}$P]dCTP (60 mCi) and AMV reverse transcriptase (Pharmacia) under standard conditions. Following electrophoresis on a 6% nondenaturing polyacrylamide gel (for 2.5 h at 200 V), the 133 and 253 bp bands were separated, identified by autoradiography, excised, and isolated by elution in 500 mM ammonium acetate–10 mM magnesium acetate buffer. The DNA was collected in the supernatant upon filtration and then precipitated twice with ethanol, centrifuged, lyophilized, and resuspended in 10 mM Tris–10 mM NaCl buffer pH 7.0.

*DNase I Footprinting.* The protocol for performing footprinting measurements was as previously described (Low et al., 1984a,b). Briefly, the reaction mixture (final volume 10 $\mu$L) contained a singly end-labeled DNA fragment, together with the tested drug at different concentrations in 10 mM Tris-HCl buffer (pH 7.0) containing 10 mM NaCl, 0.4 mM $MgCl_2$, and 0.4 mM $MnCl_2$. After preincubation at 37 °C for 30 min, the sample was digested with 0.006 unit/mL DNase I for 1 and 5 min at 20 °C. The reaction was stopped by adding $Na_2EDTA$ in a solution of formamide–electrophoresis dyes (0.1% xylene cyanol, 0.1% bromophenol blue). Samples were then loaded onto the gel after heating at 90 °C for 3 min. The cleavage products of the DNase I reactions were analyzed on 0.3 mm thick, 8% polyacrylamide gels containing 8 M urea and Tris–borate–EDTA buffer (pH 8.3). After 2-h electrophoresis at 1500 V, the gels were fixed in 10% acetic acid, transferred to Whatman 3MM paper, dried under vacuum at 80 °C, and subjected to autoradiography (Kodak, X-OMAT AR) at −70 °C with an intensifying screen. All experiments were performed at least in triplicate in order to ensure consistent behavior of the tested drugs and to verify that autoradiographs submitted to microdensitometry were properly representative.

*Densitometry and Numerical Analysis.* Autoradiographs were scanned using a multichannel computer-operated gel scanner in the Medical Research Council Laboratory of Molecular Biology (Cambridge, U.K.). Gel profiles were plotted and displayed on a raster graphics screen. Quantitative analysis of the gel electrophoresis profiles was performed by simple integration of the area under each peak, using a computer program developed specially for the purpose (Smith & Thomas, 1990). Band intensities were transformed into values for the fractional cleavage as previously described (Fox & Waring, 1986). Plots of ln ($f_a/f_c$) versus band number, where

Table I: Structures and DNA-Binding Parameters for Amsacrine-4-carboxamides and Related Compounds[a]




| compd no. | R | binding constant $K_{(0)} \times 10^{-6}$ ($M^{-1}$) AT | GC | $R$ | dissociation kinetics $1/\tau_{av}$ ($s^{-1}$) CT | AT | GC | IC |
|---|---|---|---|---|---|---|---|---|
| 1 | H | 0.30 | 0.30 | 1.0 | 600 | | | |
| 3 | - | 22 | 50 | 2.3 | 8.6 | 210 | 5.3 | 150 |
| 4 | $C(O)NHCH(OH)CH_2OH$ | 2.5 | 37 | 15 | 2.4 | | | |
| 5 | $C(O)NH(CH_2)_2N(CH_3)_2$ | 25 | 66 | 2.6 | 0.30 | 44 | 0.3 | 6.1 |
| 6 | $C(O)NH(CH_2)_2NH(CH_2)_2OH$ | 22 | 49 | 2.2 | 0.58 | | | |
| 7 | $C(O)NH(CH_2)_2N(CH_3)_2$ | 2.3 | 4.0 | 1.7 | 2.9 | | | |
| 8 | $C(O)NHCH_2C(O)NH(CH_2)_2NH(CH_2)_2OH$ | 28 | 66 | 2.4 | 0.62 | | | |

[a] The association constant $K_{(0)}$ expressed in base pair units for binding of each drug to poly(dA-dT) and poly(dG-dC), designated AT and GC, respectively, in buffer of ionic strength 0.01, pH 7.0, was measured using an ethidium displacement method as described by Baguley and Falkenhaug (1981). $R$ is the ratio of affinities for the two polynucleotides, i.e., $R = K_{(0)}^{GC}/K_{(0)}^{AT}$. The dissociation constants were measured in buffer of ionic strength 0.1, pH 7.0 at 20 °C, using the sodium dodecyl sulfate (SDS) sequestration method at final [DNA] = 200 μM, final [ligand] = 10 μM, and final [SDS] = 10 mM for calf thymus DNA, designated CT, poly(dA-dT), designated AT, poly(dG-dC), designated GC, and poly(dI-dC), designated IC. The average dissociation rate is given as the harmonic mean time constant defined as $1/\tau_{av} = 1/\sum[A_i/(100 \times \tau_i)]$, where $A_i$ represents the amplitude and $\tau_i$ the time constant of each individual kinetic transient observed. The data are taken from Wakelin and Denny (1990).

$f_a$ and $f_c$ are the fractional cleavage at a given bond in the presence and absence of the ligand, respectively, were used to represent the differential cleavage at each internucleotide bond. The results are displayed on a logarithmic scale for convenience so that positive values indicate enhanced cleavage whereas negative values indicate blockage.

## Results

*Equilibrium and Kinetic Binding Data.* It is useful to review the kinetic and equilibrium binding properties of the compounds studied as a prelude to describing their sequence selectivity since this will aid interpretation of the footprinting data. To this end, Table I gives the structures of the compounds and summarizes some kinetic and thermodynamic parameters of their binding to calf thymus DNA and synthetic polynucleotides [taken from Wakelin et al. (1990) and Wakelin and Denny (1990)]. The association constants as measured by the ethidium displacement assay show that amsacrine (**1**) has no discernible sequence preference of binding, which contrasts with the results of theoretical studies suggesting a degree of AT specificity for this ligand (Chen et al., 1988). It is evident, however, that the 4-carboxamide side chain confers a clear-cut selectivity for GC over AT base pairs on the acridine chromophore, ranging from a factor of 1.7 to 15 depending on the nature of the carboxamide substituent (Table I, compounds **3–8**). This preferential binding to GC sequences by compound **3** is also predicted by computational studies (Chen et al., 1987). The reciprocal harmonic mean time constants for dissociation of the calf thymus DNA complexes of compounds **3–8** (Table I) reveal the significant stabilizing effect of a bulky 4-carboxamide side chain. The most kinetically stable complexes are formed by the $C(O)NH(CH_2)_2NMe_2$ analogue **5** which dissociates 2000 times more slowly than amsacrine (**1**) whose complex has a lifetime of only a few milliseconds. Even when the side chain does not contain an additional positive charge, as in compound **4**, there is a large stabilization of the complex compared to that of amsacrine (**1**), and interestingly this ligand exhibits the greatest sequence preference. The DNA complex of compound **5** dissociates 29-fold more slowly than does that of the corresponding 9-aminoacridine-4-carboxamide (**3**), showing the importance the threading mechanism plays in kinetically stabilizing DNA–drug complexes [for discussion see Wakelin et al. (1990) and Wakelin and Denny (1990)]. Finally, Table I includes the data for the dissociation of compounds **3** and **5** from synthetic polynucleotides that reveal the similarity of their rates of dissociation from poly(dG-dC)·poly(dG-dC) and calf thymus DNA, a finding implying that the high-affinity sites for these compounds in natural sequence DNA are GC-rich (Wakelin & Denny, 1990).

*DNase I Footprinting Data.* The ability of all the compounds listed in Table I to inhibit the nuclease activity of DNase I was examined using the 160 bp *tyr* T DNA fragment (cut between the *Eco*RI and *Ava*I restriction sites). While amsacrine itself consistently fails to give a footprinting signal even at concentrations as high as 1 mM, the amsacrine-4-carboxamide derivatives **5**, **7**, and **8** and the 9-aminoacridine-4-carboxamide **3** give clear patterns of inhibition of cutting, showing that these agents bind preferentially to certain nucleotide sequences. In contrast, the amsacrine derivative **4** only poorly inhibits DNase I activity and together with compound **6** gives rather undistinguished footprints from which it is difficult to extract any clear evidence for sequence-selective binding.

As an example of the quality of the primary data, Figure 1 shows the effects of increasing concentrations of compound **7** on the DNase I cleavage pattern of *tyr* T DNA 3′-end-labeled on the top (Watson) strand and, in a separate experiment, 3′-end-labeled in the bottom (Crick) strand. Selective blockage of cutting at several sites can be seen on the Watson strand with low band intensities evident in mixed sequences of equal A+T and G+C composition or in G+C-rich regions around positions 20–25, 35–45, 57, 70–75, 78–80, 95–110, 115, and 117–120. This pattern of inhibition is mirrored almost perfectly on the Crick strand, the sole difference being an additional area of restricted cutting around positions 87–94. The loss of intensity in all these regions is strictly dependent on the concentration of the ligand, indicating that the sites

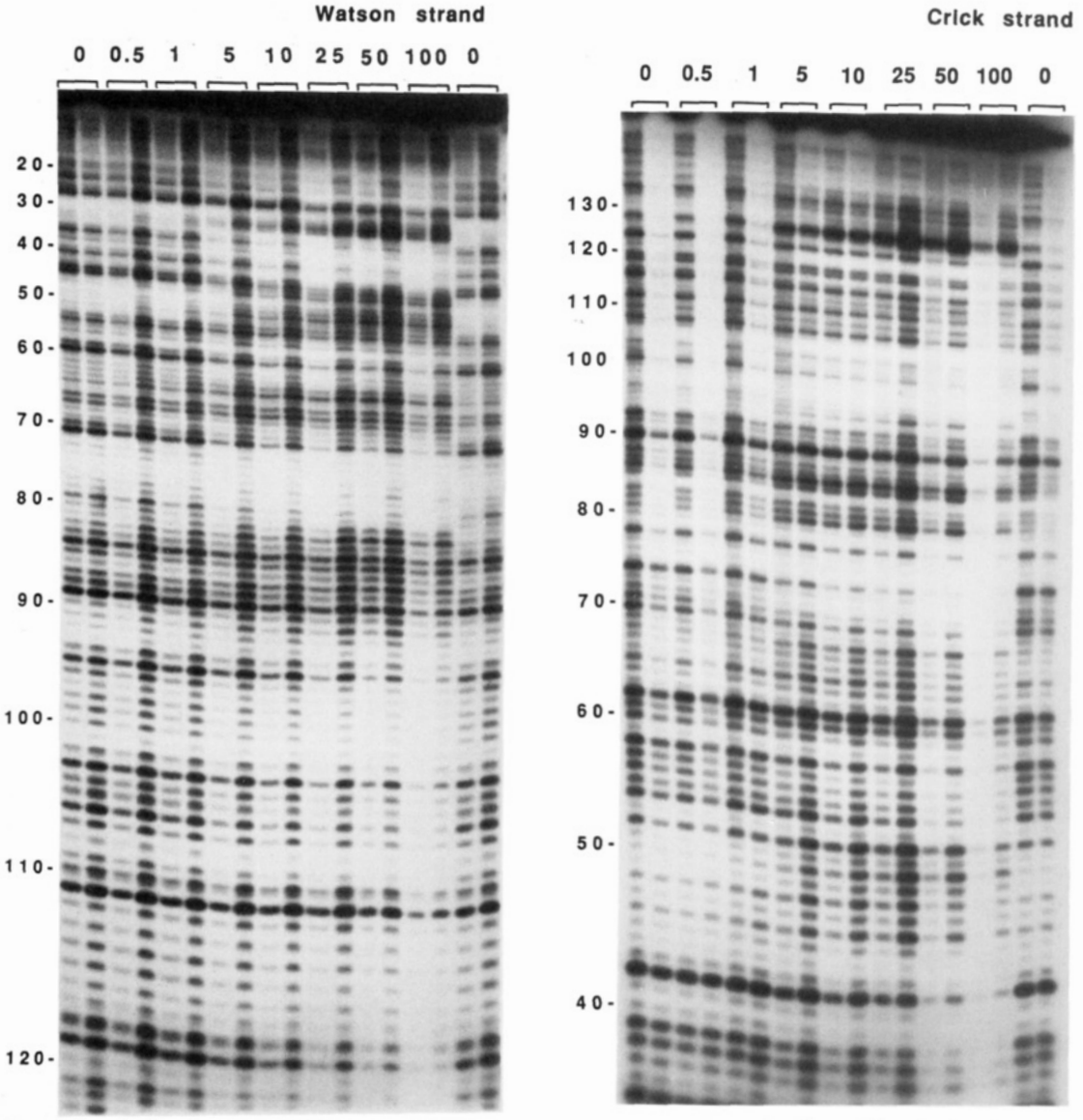



FIGURE 1: DNase I footprinting of compound **7** bound to the 160 bp duplex DNA fragment from *E. coli* containing the tyrosine tRNA promoter region. The DNA was labeled with [$\alpha$-$^{32}$P]dCTP for the Watson strand (left-hand panel) or with [$\alpha$-$^{32}$P]dATP for the Crick strand (right-hand panel) in the presence of AMV reverse transcriptase. Numbers on the left side of the gels refer to the sequence shown in Figure 2. The concentrations of the drug tested vary from 0 (no drug, control) to 100 $\mu$M as indicated at the top of each set of two lanes which represent digestion by DNase I for 1 and 5 min at room temperature.

are protected from cleavage as a direct result of competitive drug binding. Dose-dependent enhancements in cleavage rates relative to that in drug-free controls occur at flanking sequences surrounding certain binding sites. The domains of enhanced cutting, positions 28–33, 46–53, 62–68, and 82–93 in the Watson strand, are markedly AT-rich, as are practically all such tracts which become sensitized to DNase I cleavage in the presence of DNA-binding ligands.

Figure 2 shows differential cleavage plots for compounds **3**–**8** evaluated by analyzing autoradiograms like those illustrated in Figure 1 with a scanning microdensitometer. These plots were generated at a common drug concentration of 50 $\mu$M except in the case of the Watson strand for compound **4** whose poor efficacy at inhibiting DNase I activity necessitated the use of 10-fold more ligand. Such plots often display a symmetrical skewing of the footprint pattern to the 3′ side of the binding site for intercalating agents, but this feature is not evident with the agents studied here. Figure 2 delineates more clearly the inhibition and enhancement patterns described above for compound **7**. The agent with the most comparable footprinting pattern to **7** is the homologue **5**, whose cutting motif is practically identical except that its amplitudes of inhibition and enhancement are somewhat larger at positions 92–118 and 46–53, respectively, on the Watson strand. Compound **8**, the ligand with a complex glycinamide side chain, yields a footprint that is very similar to that of **7** on the Crick strand. On the Watson strand, however, inhibition of cutting is substantially reduced at positions 35–45 and 70–80, and the cleavage of some phosphate bonds within these regions is actually enhanced. Indeed, the most prominent characteristic of the Watson strand footprint of **8** is the widely-distributed increase in cutting rates. The hydroxyethyl derivative 6 shares this peculiarity with compound **8** on the Watson strand, the footprint pattern showing little inhibition and a widespread enhancement of cutting. On the Crick strand, the motif is again reminiscent of that of the glycinamide derivative, but now the differential cleavage ratios are very diminished. Overall, it is apparent that compound **6** only weakly inhibits DNase I activity, and it is therefore difficult to discern any sequence-dependent variations in the cutting patterns. Similar comments also pertain to the behavior of the diol derivative **4**, where the footprint is compressed in an

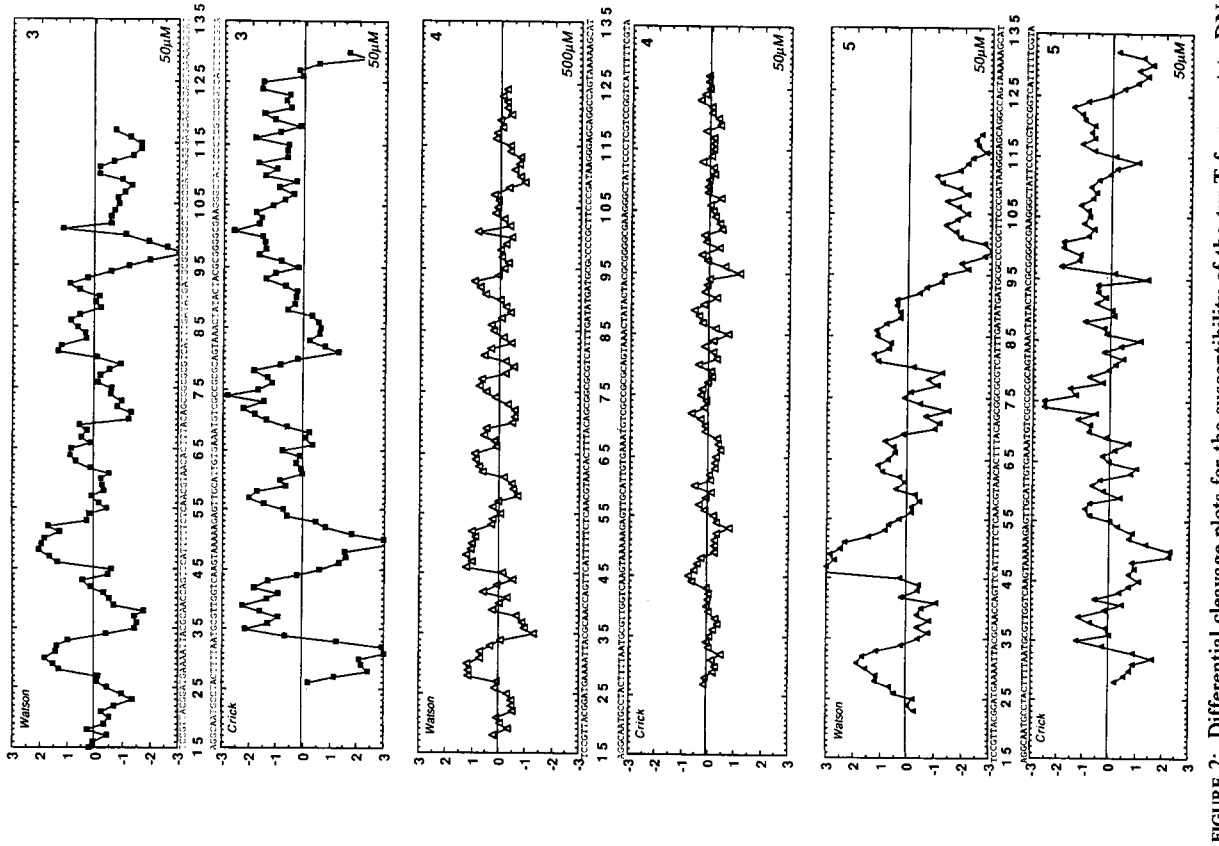



FIGURE 2: Differential cleavage plots for the susceptibility of the *tyr* T fragment to DNase I in the presence of 50 $\mu$M of **3**, **4**, **5**, **6**, **7**, and **8** (from top to bottom, left panel, and from top to bottom, right panel, respectively). For compound **4**, 500 $\mu$M was used for the top strand. The ordinate scales are the dimensionless difference $(\ln f_a - \ln f_c)$, where $f_a$ and $f_c$ represent values of fractional cleavage of a given bond, calculated as previously described (Fox & Waring, 1984). The upper part shows differential cleavage of the upper "Watson" strand, and the lower, of the complementary "Crick" strand. Negative values on the ordinate indicate protection by the bound ligand, while positive values indicate enhanced cleavage. The ordinate scales for the two strands are inverted, so that deviation of points *toward* the lettered sequence corresponds to a ligand-protected site and deviation *away* represents enhanced cleavage.

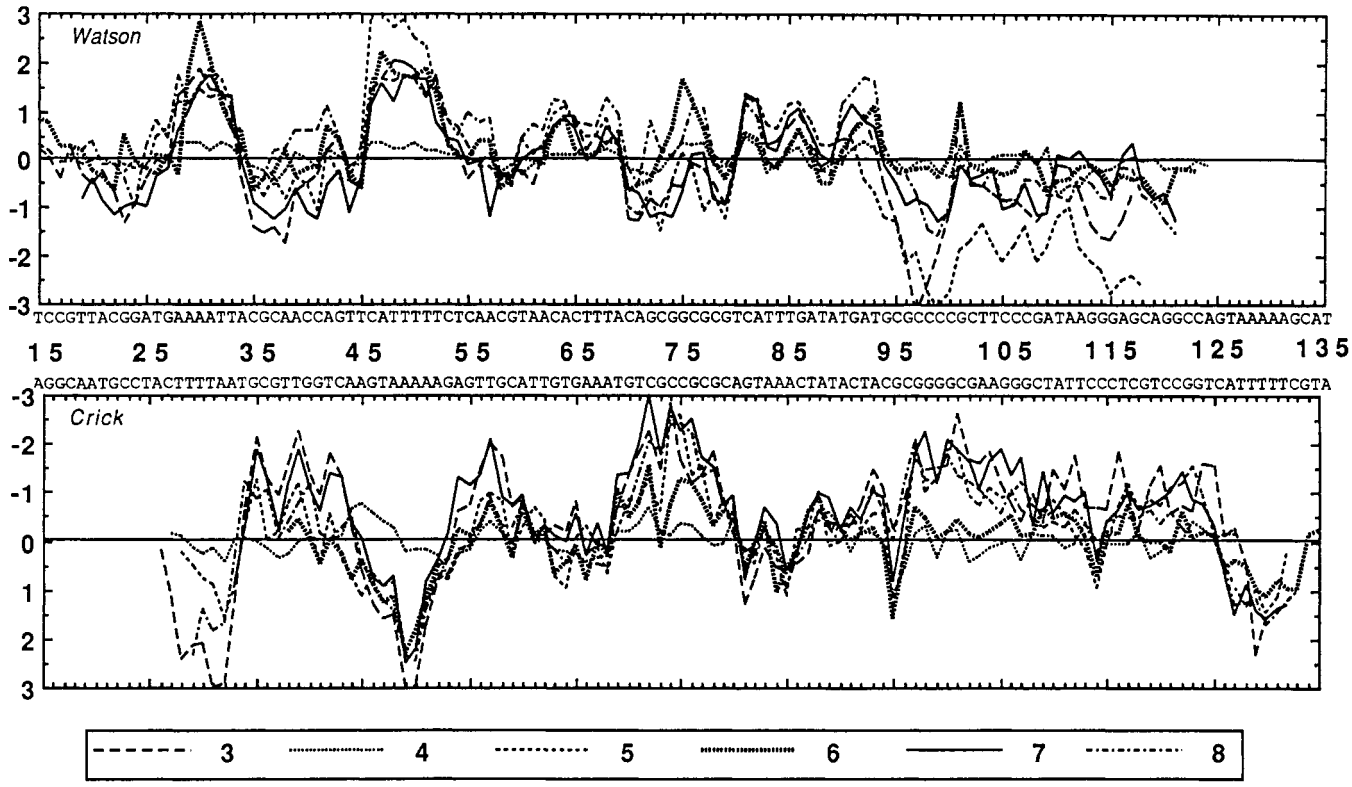



FIGURE 3: Superimposed differential cleavage plots for the susceptibility of the *tyr* T fragment to DNase I in the presence of 50 $\mu$M compounds **3**–**8**. All conditions identical to those in Figure 2.

even more extreme manner, so that as for compound **6** it is difficult to judge whether the ligand binds to any sequences preferentially. Lastly in Figure 2, we present the differential cleavage plot for the 9-aminoacridine-4-carboxamide **3** which possesses the same side chain as compound **5** but lacks the anilino substituent. Its footprinting pattern is practically indistinguishable from those of analogues **5** and **7**, which unequivocally demonstrates the dominance of the carboxamide side chain as the main determinant of sequence selectivity among these compounds. Figure 3 provides a summary overlay of all the footprinting motifs measured at 50 $\mu$M which makes clear that those agents which convincingly show selectivity (**3**, **5**, **7**, and **8**) share the same sequence preferences: namely, a strong tendency to block cutting on both DNA strands at positions 19–26, 34–45, 53–60, 69–80, and 92–125. This pattern is maintained over a ligand concentration range of 5–50 $\mu$M (data not shown).

Given the close correspondence between the differential cleavage patterns of the amsacrine-4-carboxamide **5** and the analogous 9-aminoacridine derivative **3** on the *tyr* T DNA fragment (Figures 2 and 3), we investigated further the similarity of their binding specificity by using two additional DNAs of length 133 and 253 base pairs obtained by digestion of the pBS plasmid with the restriction enzymes *Ava*I and *Pvu*II as described in the Materials and Methods section. As an example of the data, Figure 4 shows an autoradiogram obtained from DNase I digestion of these two pBS fragments in the presence of various concentrations of compounds **3** (left-hand panel, 133 bp fragment) and **5** (right-hand panel, 253 bp fragment). Here again, sequence-specific protection and enhancement of cleavage rates can be clearly observed in a ligand concentration-dependent manner. Inspection of the corresponding differential cleavage plots of the Watson strands at a concentration of 50 $\mu$M (Figure 5) confirms both the similar sequence specificity of these two compounds and endorses the finding that their preferred binding sites are located in G+C-rich regions of DNA (e.g, sites 19–31, 46–55, 65–75, 85–95, and 106–114 on the 133 bp fragment and 22–34, 44–46, 50–55, 66–69, 71–76, and 97–103 on the 253 bp fragment).

Table II summarizes the sequences that are most strongly protected from cleavage ($\ln f_a - \ln f_c < -1$ at 50 $\mu$M), evaluated considering data from both DNA strands at each site where possible, by compounds **3** and **5** in all three DNA fragments studied. At this level of protection the binding sites span a wide range in size from a single base pair up to a run of 32 base pairs. All of the highest affinity sites are very G+C-rich. These findings imply that, in general, a single GC base pair is a sufficient determinant for binding and unequivocally show that sites with the greatest affinity are found in clusters of GC base pairs where, having regard to the likelihood of neighbor-exclusion effects, individual GC sites either abut or overlap. Only in one case, at the sequence ATAT at positions 24–27 on the 133 base pair pBS fragment, is there any compelling evidence that high-affinity intercalation occurs between two adjacent AT base pairs (the other putative case, ATAA at positions 109–112 on *tyr* T DNA, appears somewhat problematical).

## Discussion

The compounds studied fall into two broad categories with respect to their ability to inhibit DNase I activity. Compounds **3**, **5**, **7**, and **8** are good inhibitors and show pronounced preferences for binding to GC-rich sequences whereas amsacrine and ligands **4** and **6** only inhibit cutting very weakly, if at all. In the case of the latter agents it is therefore difficult to draw any firm conclusions about their sequence preferences using DNase I footprinting methods. Although amsacrine binds by far the least tightly and dissociates much more rapidly than the other agents, compounds **4** and **6** share similar equilibrium and kinetic parameters with their congeners. It is reasonable to speculate that amsacrine fails to footprint

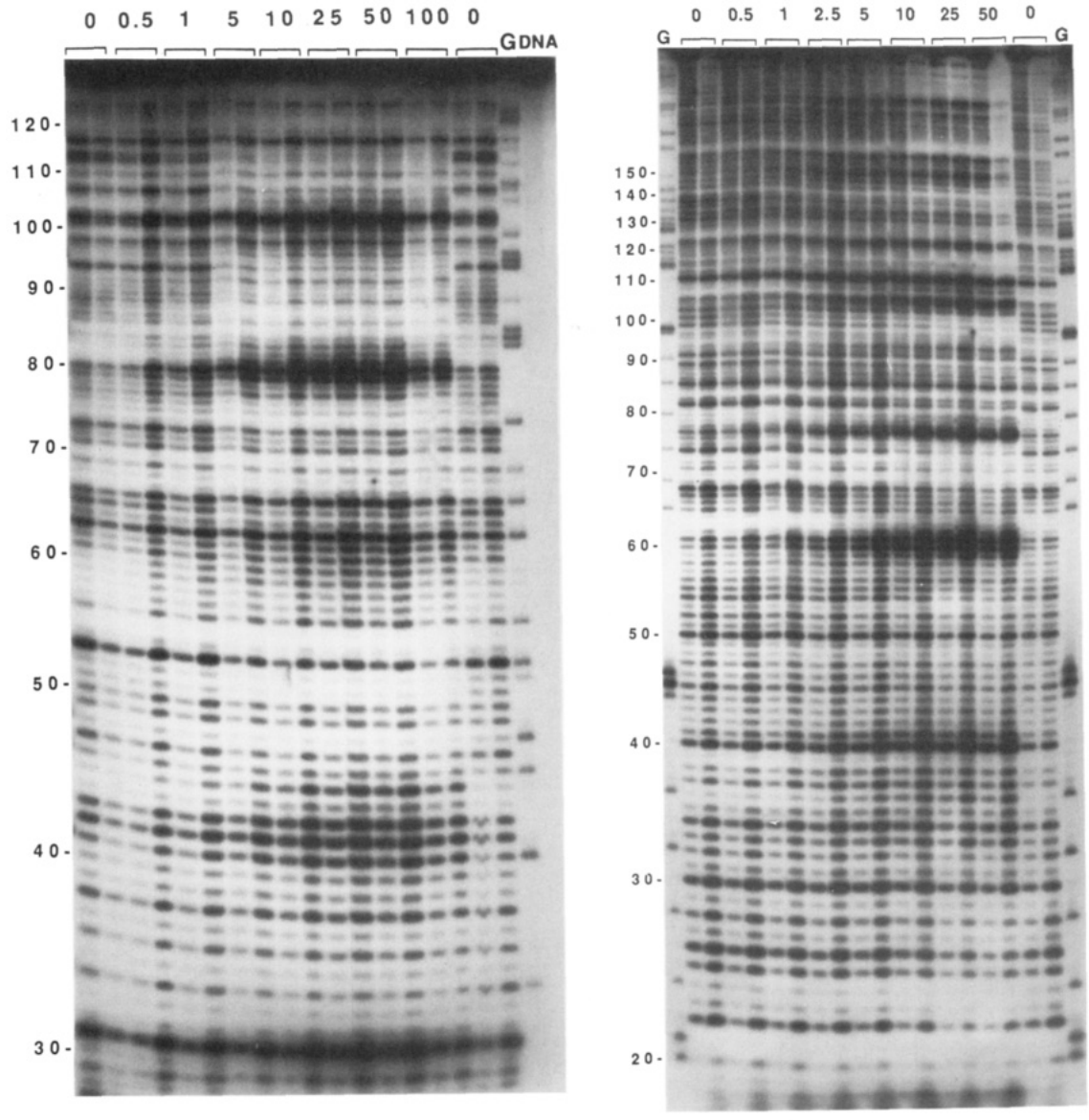



FIGURE 4: DNase I footprinting of compound **3** bound to the 133 bp (left-hand panel) and compound **5** bound to the 253 bp (right-hand panel) DNA fragments from the plasmid pBS. Labeling and purification of these fragments are described in the Materials and Methods section. Numbers on the left side of the gels refer to the sequences shown in Figure 5. Track G is a dimethyl sulfate–piperidine marker specific for guanine. Track labeled DNA represents the untreated control DNA. Other details as for Figure 1.

because of its exceptionally weak binding and short residence time given the difficulties previously experienced in footprinting simple intercalators like ethidium and proflavin that have fast binding kinetics (Fox & Waring, 1987). By contrast, on the face of it, the reasons for the failure of **4** and **6** to inhibit the enzyme must lie elsewhere than in inappropriate kinetic and equilibrium properties and would thus appear to be related to the structures of their DNA–ligand complexes. However, it is difficult to fathom what structural features distinguish the complexes of **4** and **6** in this regard given that these two ligands, in common with the other amsacrine-4-carboxamide derivatives, are believed to bind with the anilino ring positioned in one DNA groove and the bulky carboxamide side chain in the other (Wakelin et al., 1990; Wakelin & Denny, 1990). As a first guess one might suggest that the groove locations of their side chains are reversed compared to their stable-mates, but since either side chain would present a substantial blockage to the binding of DNase I in the minor groove (Wakelin et al., 1990; Wakelin & Denny, 1990), this proposal is not compelling.

An alternative possibility is that compounds **4** and **6** simply fail to bind to the *tyr* T DNA fragment under the conditions of the footprinting assay, or maybe they bind in a different, perhaps nonintercalating, manner. This is a plausible explanation given that the footprinting buffer uniquely contains, compared to our previous DNA-binding studies with these agents, 400 $\mu$M $Mg^{2+}$ and 400 $\mu$M $Mn^{2+}$ and that compounds **4** and **6** have side chains with potential oxygen- and oxygen/nitrogen-containing chelating groups for these two metal ions, i.e., $CH(OH)CH_2OH$ in **4** and $NH(CH_2)_2OH$ in **6**. By contrast, compounds **3**, **5**, and **7** conspicuously lack metal-chelating ability of this type. Since $Mg^{2+}$ and $Mn^{2+}$ have reasonably high affinity for such chelates, generally forming octahedral complexes (Cotton & Wilkinson, 1972), and given that their combined $M^{2+}$ concentration is 800 $\mu$M compared to 50–500 $\mu$M for the ligands, it is possible that **4** and **6** might be sequestered as metal complexes in the cleavage assay. Such complexes may fail to bind to DNA altogether or might bind in some other mode not involving intercalation. We note that compound **8**, which gives reasonably good footprints, also has the potential to chelate divalent metal ions, but its side chain is complex and has the added capacity to engage in internal hydrogen bonding schemes. Lastly, on this point, it does not seem likely to us that the presence of $Mg^{2+}$ and $Mn^{2+}$ per se

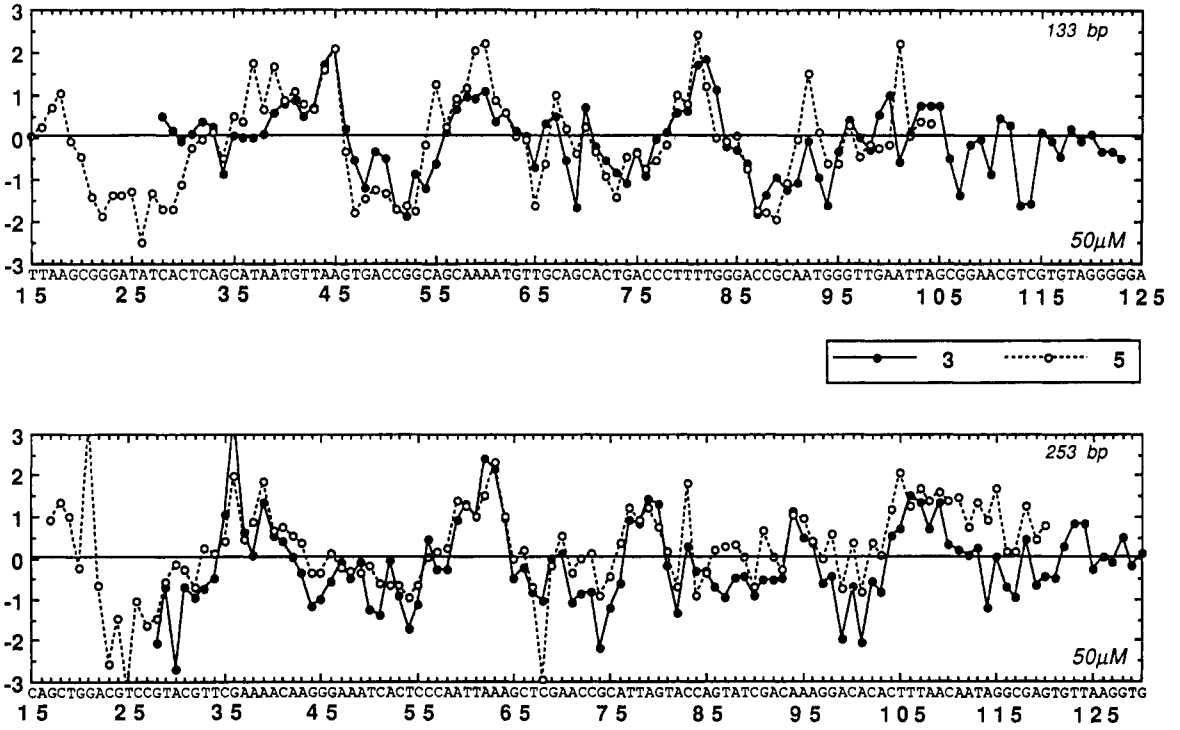



FIGURE 5: Superimposed differential cleavage plots for the susceptibility of the two pBS DNA fragments to DNase I in the presence of 50 μM compounds **3** (●) and **5** (○). Top, a 133 bp fragment from *Ava*I and *Pvu*II digestion of plasmid pBS; bottom, a 253 bp fragment obtained from the same digest.

Table II: Sequences of Highest Affinity Binding Sites for Compounds **3** and **5** as Inferred from Differential Cleavage Plots

| | compound **3** | | compound **5** | |
|---|---|---|---|---|
| DNA | position | sequence | position | sequence |
| *tyr* T | 23–24 | GG | | |
| | 35–43 | CGCAACCAG | 33–41 | CGCAACC |
| | 55–58 | CAAC | | |
| | 70–79 | CAGCGGCGCG | 70–79 | CAGCGGCGCG |
| | 93–125 | TGCGCCCCGCTTCCCGATAAGGGAGCAGGCAG | 94–123 | GCGCCCCGCTTCCCGATAAGGGAGCAGGC |
| 133 pBS | | | 21–30 | GGGATATCAC |
| | 51–54 | CGGC | 47–53 | TGACCGGC |
| | 69 | G | 65 | T |
| | | | 73 | TG |
| | 86–91 | ACCGCA | 87–90 | CCGC |
| | 94 | G | | |
| | 107 | G | | |
| | 113–114 | TC | | |
| 253 pBS | | | 23–28 | CGTCCG |
| | 28–31 | GTAC | | |
| | 44–45 | GG | | |
| | 50–51 | TC | | |
| | 53–55 | CTC | | |
| | | | 68 | C |
| | 72–75 | CCGC | 74 | G |
| | | | 84 | C |
| | 98–102 | GACAC | | |

selectively affects the binding properties of **4** and **6** as a result of metal ion condensation or specific site binding to DNA since this buffer system has been widely used to footprint DNA–ligand complexes with DNase I.

The most significant results of this study relate to the sequence preferences revealed for binding of the amsacrine-4-carboxamide derivatives **5** and **7** and for the 9-amino-acridine-4-carboxamide **3** which has the same carboxamide side chain as **5** but lacks its anilino substituent. All three compounds exhibit practically identical footprinting patterns with a strong preference for binding to GC-rich sequences. Assuming that all occupied binding sites have the same efficacy in inhibiting nuclease activity, the highest affinity sites generally occur in clusters of GC base pairs, the minimal high-affinity site being a single GC pair. These findings concur fully with the results of equilibrium and kinetic measurements, which suggest that **3** and **5** bind most tightly to GC-rich sequences in naturally-occurring eukaryotic DNA (Wakelin et al., 1987, 1990; Wakelin & Denny, 1990). The same kinetic and equilibrium studies led to a model of the DNA complexes of **3**, **5**, and **7** in which the acridine chromophore overlaps fully with the base pairs and the NH group of the 4-carboxamide moiety makes hydrogen bonding interactions with a cytosine $O_2$ carbonyl oxygen in the minor groove. This configuration places the 9-amino group or the 9-anilino substituent in the major groove. Compounds **3** and **5** have an ethyl link between their carboxamide NH and protonated dimethylamino terminal nitrogens, which, in the proposed model, permits them

to make an additional hydrogen bond from the latter to the cytosine oxygen [see Wakelin et al. (1987, 1990) and Wakelin and Denny (1990) for details]. By contrast, **7** has a propyl linkage which is too long to accommodate the protonated dimethylamino-to-cytosine oxygen hydrogen bond. It was found possible to correlate the kinetic and biological properties of the 9-aminoacridine- and amsacrine-4-carboxamides with their capacity to form this proposed bifurcated hydrogen bond (Wakelin et al., 1987, 1990; Wakelin & Denny, 1990). However, the model was unable to delineate in a precise manner which interactions between drug and DNA were responsible for selective binding to GC base pairs per se since the proposed hydrogen bonding scheme was equally applicable for binding to the thymine carbonyl oxygen in AT base pairs.

The footprinting data unequivocally provide answers to some of these questions, irrespective of whether the results are expressed within the context of the proposed binding model or not. First, it is clear that the 9-anilino substituent plays no part in determining GC selectivity among the amsacrine-4-carboxamide ligands since the sequence specificities of **3** and **5** are practically indistinguishable. Therefore, the origins of the selectivity must reside either in stacking interactions between the acridine chromophore and DNA or in the interactions between DNA and the carboxamide side chain. Since aminoacridines have no strong preference for binding to GC sequences (Fox & Waring, 1987), it appears that the carboxamide substituent must be the principal determinant of specificity for both the 9-aminoacridine- and amsacrine-4-carboxamides. Moreover, by observing that compounds **5** and **7** also have the same sequence selectivity, we can further conclude that interactions between the terminal protonated dimethylamino group of the carboxamide side chain and the DNA contribute little or nothing to determining that specificity. This conclusion is reinforced by the footprinting data for compound **8** which has yet a different 4-carboxamide side chain and whose sequence selectivity is again very similar to that of **3**, **5**, and **7**. Thus, it appears that GC specificity is conferred by possession of the 4-carboxamide group per se and that it is its interaction with the DNA base pairs which is the predominant feature. This finding can be readily rationalized within the terms of the Wakelin and Denny model (Wakelin et al., 1987, 1990; Wakelin & Denny, 1990) since in the proposed hydrogen bonding scheme the carbonyl oxygen of the 4-carboxamide group is well placed to hydrogen bond simultaneously to the guanine 2-amino group of a GC base pair, forming the other wall of the intercalation cavity. An AT base pair in this position could not participate in this second bonding interaction. Interestingly, such a model can also account for high-affinity binding adjacent to a single GC pair, or to alternating GC/AT sequences as is occasionally found experimentally (see Table II), since the NH group of the 4-carboxamide moiety can hydrogen bond to the pyrimidine carbonyl oxygen of either an AT or a GC base pair in a complex of this type, as mentioned above.

The DNase I cleavage inhibition data also provide additional, albeit circumstantial, insights into the structure of the DNA complexes of **3** and **5**. The Wakelin and Denny model (Wakelin et al., 1987, 1990; Wakelin & Denny, 1990) proposes that the 4-carboxamide side chains of both compounds make identical interactions with DNA principally because of the very close similarities in the equilibrium and kinetic behavior of their complexes. It is a significant finding, therefore, that the footprinting data should also be essentially identical for these two compounds, thereby reinforcing the notion that their side chains are involved in comparable interactions and are located in the same DNA groove. Indeed, this result is all the more signal since, taking into consideration all the footprinting data, it appears to highlight further the dominant importance of the 4-carboxamide group in determining both sequence specificity and the groove disposition of the side chains. In their theoretical modeling of the binding of compound **3** to DNA Chen et al. (1987) concluded that the ligand ought to have GC selectivity but that its carboxamide side chain is located in the wide groove with the carboxamide moiety making hydrogen bonds to the $O_6$ of guanine and the 4-amino group of cytosine in a CG dinucleotide site. This is, in fact, an analogous hydrogen bonding scheme to that described above for the minor groove interactions of the carboxamide group at this sequence. However, it is a scheme that would not obviously distinguish between GC and AT sequences given the similarity of the disposition of the hydrogen bonding groups of AT and GC base pairs in the major groove. In the Chen et al. (1987) model the protonated dimethylamino group makes hydrogen bonding interactions with the $N_7$ of guanine and the phosphate backbone and the chromophore is severely twisted in the intercalation site so that the 9-amino group can form additional hydrogen bonds with sugar and phosphate oxygens in the minor groove. Thus, there are currently two views about which groove the carboxamide side chains of these compounds lie in. Obviously, X-ray crystallographic and/or NMR structural studies are required to resolve this dilemma properly. Nevertheless, the footprinting data may contribute to the debate since it is known that DNase I cleaves DNA by binding and cutting in the minor groove (Drew & Travers, 1984; Suck & Oefner, 1986; Suck et al., 1988) and it is well appreciated that agents that compete for binding in this groove are good inhibitors of cleavage [e.g., Low et al. (1984a), Fox and Waring (1984), and Bailly et al. (1990a)] but that those that bind in the major groove such as cisplatin are not. Compound **3** is an excellent inhibitor of DNase I activity, as good as if not better than compound **5**, which is easily rationalized if the carboxamide side chain is bound in the minor groove. However, if the side chain is positioned in the major groove in the manner of the Chen et al. (1987) proposal, very little of the ligand is present in the minor groove to antagonize binding of the enzyme. Whatever the actual side-chain dispositions, it is becoming increasingly clear that the 9-aminoacridine- and amsacrine-4-carboxamides do take up a common, highly-preferred orientation on binding to DNA.

*Implications for Amsacrine Binding.* The selectivity of amsacrine binding to natural sequence DNA remains unknown, principally because the ligand will not give footprints in nuclease inhibition assays. Studies of binding to synthetic polynucleotides indicate that it binds preferentially to alternating pyrimidine–purine sequences but with no significant discrimination between AT or GC base pairs (Wilson et al., 1981). Theoretical calculations by Chen et al. (1988) discern a preference for alternating AT over GC sequences and place the anilino substituent of amsacrine in the DNA minor groove. The question arises as to whether the studies reported here for the amsacrine-4-carboxamides shed any light on the selectivity of amsacrine binding itself. Since our findings strongly implicate the 4-carboxamide group as the principal determinant of GC specificity, we would be forced to conclude that they do not and there are no grounds for extrapolating the behavior of the amsacrine-4-carboxamides to that of amsacrine. However, given the extent of the dominance of the 4-carboxamide group in determining GC selectivity, it does seem probable that the 5-methyl-4-(methylcarbamoyl) derivative of amsacrine, compound **2**, would bind preferentially

to GC-rich clusters in natural sequence DNA. This is all the more likely since **2** shows a preference for binding to poly(dG-dC)·poly(dG-dC) over poly(dA-dT)·poly(dA-dT) in common with all the amsacrine-4-carboxamides (Denny et al., 1984). Whether a propensity for binding to GC sequences would contribute to the activity of **2** against solid tumors, a biological property not shared by amsacrine, remains to be seen.

*Comparison with Other Intercalating Ligands.* Prior to the advent of footprinting techniques it was commonly believed that intercalating agents bound preferentially to GC-rich DNA and that minor groove binding ligands had high specificity for AT sequences. To some extent this view was reinforced by the finding that the ligand DAPI binds in the minor groove at AT-rich sequences and intercalates at GC-rich sites (Wilson et al., 1990). However, in the wake of the refinements to our knowledge given by the footprinting methods, it is now clear that GC selectivity is not an obligatory consequence of intercalation and that among intercalating agents a diversity of sequence recognition features exists. At one extreme, anthracene derivatives were reported to have AT specificity (Wilson et al., 1985), which contrasts with the findings for ethidium, propidium, and proflavin which show little, if any, preferential binding. The latter compounds bind best to mixed sequences of alternating pyrimidines and purines and are excluded from runs of adenines and thymines (Fox & Waring, 1987). Footprinting studies with benzophenanthridine (Bajaj et al., 1990) and anthraquinone derivatives (Fox et al., 1986) again revealed similar behavior without any marked GC selectivity. A more pronounced GC preference was reported with natural intercalators of the anthracycline family such as daunomycin (Chaires et al., 1990) and nogalamycin (Fox & Waring, 1986) and also with actinomycin (Fox & Waring, 1984) and echinomycin (Low et al., 1984a). In these cases, the glycan or peptide moiety appended to the chromophore actively contributes to the recognition process through the formation of key hydrogen bonds with DNA. This led to the discovery of the triplet sequences 5′-A/TCG and 5′A/TGC as the most favored binding sites for daunomycin (Gresh et al., 1985; Chaires et al., 1990). A clear GC specificity was also found for substituted phenazines related to phenyl neutral red (Muller et al., 1975), with a variety of platinum-containing intercalators (Wakelin et al., 1984), and for a derivative of ellipticine bearing an alkylamino side chain (Bailly et al., 1990b). These various examples of intercalating ligands and their specificity help to illustrate the fact that the findings presented here are neither unique nor particularly common. Nevertheless, it is important to appreciate that the 9-aminoacridine- and amsacrine-4-carboxamides exhibit a very high degree of selectivity for GC sites which, to our knowledge, has not been previously matched by other acridine derivatives.

A thorough understanding of both the sequence selectivity of DNA-binding anticancer drugs and the three-dimensional structure of their DNA complexes is an important prerequisite for the rational design of novel therapeutic agents of this class. Compounds such as the amsacrine-4-carboxamides that are able to place distinguishable side chains selectively in each DNA groove in a well-defined fashion may have important potential as templates for the specific delivery of other, reactive, functional groups to defined regions of DNA [see, for example, Prakash et al. (1990) and Gourdie et al. (1991)]. In a similar vein AT-specific minor groove binding ligands bearing such reactive groups also hold out the prospect of more selectively toxic antitumor agents (Gravatt et al., 1991). Lastly, we note that the combination of strongly GC-selective intercalators and AT-specific groove binders may provide a route to the design of ligands targeted to defined sequences [see, for example, Bailly et al., (1990a)].

## ACKNOWLEDGMENTS

We thank Dean Gentle for his excellent technical assistance. We are especially grateful to Dr. J. M. Smith of the Medical Research Council Laboratory of Molecular Biology for instruction in the use of the microdensitometer and to Sir Aaron Klug and Dr. Smith for providing access to computer facilities.

## REFERENCES

Abraham, Z. H. L., Cutbush, S. D., Kuroda, R., Neidle, S., Acheson, R. M., & Taylor, G. N. (1985) *J. Chem. Soc., Perkin Trans. 2*, 461–465.

Arlin, Z. (1983) *Cancer Treat. Rep. 67*, 967–970.

Baguley, B. C., & Falkenhaug, E.-M. (1978) *Nucleic Acids Res. 5*, 161–171.

Bailly, C., Helbecque, N., Henichart, J. P., Colson, P., Houssier, C., Rao, K. E., Shea, R. G., & Lown, J. L. (1990a) *J. Mol. Recognit. 3*, 26–35.

Bailly, C., OhUigin, C., Rivalle, C., Bisagni, E., Henichart, J. P., & Waring, M. J. (1990b) *Nucleic Acids Res. 18*, 6283–6291.

Bajaj, N. P. S., McLean, M. J., Waring, M. J., & Smekal, E. (1990) *J. Mol. Recognit. 3*, 48–54.

Cain, B. F., Atwell, G. J., & Denny, W. A. (1977) *J. Med. Chem. 20*, 987–996.

Chaires, J. B., Herrera, J. E., & Waring, M. J. (1990) *Biochemistry 29*, 6145–6153.

Chen, K.-X., Gresh, N., & Pullman, B. (1987) *FEBS Lett. 224*, 361–364.

Chen, K.-X., Gresh, N., & Pullman, B. (1988) *Nucleic Acids Res. 16*, 3061–3074.

Cotton, F. A., & Wilkinson, G. (1972) *Advanced Inorganic Chemistry*, 3rd ed., John Wiley & Sons, New York.

Denny, W. A., & Wakelin, L. P. G. (1986) *Cancer Res. 46*, 1717–1721.

Denny, W. A., Cain, B. F., Atwell, G. J., Hansch, C., Panthananickal, A., & Leo, A. J. (1982) *J. Med. Chem. 25*, 276–315.

Denny, W. A., Atwell, G. J., & Baguley, B. C. (1984) *J. Med. Chem. 27*, 363–367.

Drew, H. R., & Travers, A. A. (1984) *Cell 37*, 391–501.

Feigon, J., Denny, W. A., Leupin, W., Kearns, D. R. (1984) *J. Med. Chem. 27*, 450–465.

Fox, K. R., & Waring, M. J. (1984) *Nucleic Acids Res. 12*, 9271–9285.

Fox, K. R., & Waring, M. J. (1986) *Biochemistry 25*, 4349–4356.

Fox, K. R., & Waring, M. J. (1987) *Nucleic Acids Res. 15*, 491–507.

Fox, K. R., Waring, M. J., Brown, J. R., & Neidle, S. (1986) *FEBS Lett. 202*, 289–294.

Gao, Y.-G., Liaw, Y.-C., & Wang, A. H. J. (1990) *Biochemistry 29*, 9913–9918.

Gourdie, T. A., Prakash, A. S., Wakelin, L. P. G., Woodgate, P. D., & Denny, W. A. (1991) *J. Med. Chem. 34*, 240–248.

Gravatt, G. L., Baguley, B. C., Wilson, W. R., & Denny, W. A. (1991) *J. Med. Chem. 34*, 1552–1560.

Graves, D. E., & Wadkins, R. M. (1990) *Molecular Basis of Specificity in Nucleic Acid–Drug Interactions* (Pullman, B., Jortner, J., Eds.) pp 177–189, Kluwer Academic Press, Dordrecht, The Netherlands.

Gresh, N., Pullman, B., & Arcamone, F. (1985) *J. Biomol. Struct. Dyn. 3*, 445–466.

Hardy, J. R., Harvey, V. J., Paxton, J. W., Evans, P., Smith, S., Grove, W., Grillo-Lopez, A. J., & Baguley, B. C. (1988) *Cancer Res. 48*, 6593-6596.
Karle, J. M., Cysyk, R. L., & Karle, I. L. (1980) *Acta. Crystallogr., Sect. B B36*, 3012-3016.
Liu, L. F. (1989) *Annu. Rev. Biochem. 58*, 351-375.
Low, C. M. L., Drew, H. R., & Waring, M. J. (1984a) *Nucleic Acids Res. 12*, 4865-4879.
Low, C. M. L., Olsen, R. K., & Waring, M. J. (1984b) *FEBS Lett. 176*, 414-420.
Muller, W., & Crothers, D. M. (1968) *J. Mol. Biol. 35*, 251-290.
Muller, W., Bunemann, H., & Dattagupta, N. (1975) *Eur. J. Biochem. 54*, 279-285.
Neidle, S., Webster, G. D., Baguley, B. C., & Denny, W. A. (1986) *Biochem. Pharmacol. 35*, 3915-3921.
Nelson, E. M., Tewey, K. M., & Liu, L. F. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1361-1364.
Prakash, A. S., Denny, W. A., Gourdie, T. A., Valu, K. K., Woodgate, P. D., & Wakelin, L. P. G. (1990) *Biochemistry 29*, 9799-9807.
Sakore, T. D., Reddy, B. S., & Sobell, H. M. (1979) *J. Mol. Biol. 135*, 763-785.
Searle, M. S., Hall, J. G., Denny, W. A., & Wakelin, L. P. G. (1988) *Biochemistry 27*, 4340-4349.
Smith, J. M., & Thomas, D. J. (1990) *CABIOS 6*, 93-99.
Suck, D., & Oefner, C. (1986) *Nature 321*, 620-625.
Suck, D., Lahm, A., & Oefner, C. (1988) *Nature 332*, 465-468.
Wadkins, R. M., & Graves, D. E. (1989) *Nucleic Acids Res. 16*, 9933-9946.
Wakelin, L. P. G., & Denny, W. A. (1990) *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions* (Pullman, B., & Jortner, J., Eds.) pp 191-206, Kluwer Academic Press, Dordrecht, The Netherlands.
Wakelin, L. P. G., McFadyen, W. D., Walpole, A., & Roos, I. A. G. (1984) *Biochem. J. 222*, 203-215.
Wakelin, L. P. G., Atwell, G. J., Rewcastle, G. W., & Denny, W. A. (1987) *J. Med. Chem. 30*, 855-861.
Wakelin, L. P. G., Chetcuti, P., & Denny, W. A. (1990) *J. Med. Chem. 33*, 2039-2044.
Waring, M. J. (1976) *Eur. J. Cancer 12*, 995-1001.
Williams, L. D., Egli, M., Ughetto, G., van der Marel, G. A., van Boom, J. H., Rich, A., Wang, A. H.-J., & Frederick, C. A. (1990) *J. Mol. Biol. 215*, 313-320.
Wilson, W. D., Wang, Y. H., Kusuma, S., Chandrasekaran, S., Yang, N. C., & Boykin, D. W. (1985) *J. Am. Chem. Soc. 107*, 4989-4995.
Wilson, W. D., Tanious, F. A., Barton, H. J., Jones, R. L., Fox, K. R., Wydra, R. L., & Strekowski, L. (1990) *Biochemistry 29*, 8452-8461.
Wilson, W. R., Baguley, B. C., Wakelin, L. P. G., & Waring, M. J. (1981) *Mol. Pharmacol. 20*, 404-414.

# A Simple Spectral-Driven Procedure for the Refinement of DNA Structures by NMR Spectroscopy†

Howard Robinson and Andrew H.-J. Wang*
*Department of Physiology and Biophysics, University of Illinois at Champaign-Urbana, Urbana, Illinois 61801*
*Received October 28, 1991; Revised Manuscript Received January 30, 1992*

ABSTRACT: We have developed a simple and quantitative procedure (SPEDREF) for the refinement of DNA structures using experimental two-dimensional nuclear Overhauser effect (2D NOE) data. The procedure calculates the simulated 2D NOE spectrum using the full matrix relaxation method on the basis of a molecular model. The volume of all NOE peaks is measured and compared between the experimental and the calculated spectra. The difference of the experimental and simulated volumes is minimized by a conjugated gradient procedure to adjust the interproton distances in the model. An agreement factor (analogous to the crystallographic *R*-factor) is used to monitor the progress of the refinement. The procedure is an iterative one. The agreement is considered to be complete when several parameters, including the *R*-factor, the energy associated with the molecule, the local conformation (as judged by the sugar pseudorotation), and the global conformation (as judged by the helical *x*-displacement), are refined to their respective convergence. With the B-DNA structure of d(CGATCG) as an example, we show that DNA structure may be refined to produce calculated NOE spectra that are in excellent agreement with the experimental 2D NOE spectra. This is judged to be effective by the low *R*-factor of ~15%. Moreover, we demonstrate that not only are NOE data very powerful in providing details of the local structure but, with appropriate weighting of the NOE constraints, the global structure of the DNA double helix can also be determined, even when starting with a grossly different model. The reliability and limitations of a DNA structure as determined by NMR spectroscopy are discussed:

There has been a growing awareness that DNA molecules are highly polymorphic. It has been unequivocally shown that in addition to B-DNA a number of stable alternative DNA conformations also exist (Rich et al., 1984; Palecek, 1991). They may play important roles in biological systems. A fuller understanding of the biological function requires the detailed knowledge of the three-dimensional structure of DNA and its potential in adopting various conformations, some of which are yet to be uncovered.

In the past, X-ray crystallography has been the principal tool for determining the three-dimensional structure of DNA molecules. Its greatest virtue is that a carefully refined crystal

† This work was supported by grants from the NIH (GM-41612 and CA-52506) to A.H.-J.W.
* Corresponding author.

0006-2960/92/0431-3524$03.00/0 