Strategies for the design of minor groove binders: a re-evaluation based on the emergence of site-selective carbohydrate binders

Studies on minor groove binders provide new insights into DNA structure and recognition, information that may in the future serve as the basis for the design of synthetic binders targeted to particular minor groove sites. Carbohydrate-based minor groove binders are emerging as a particularly interesting and important class of ligands for DNA.

Chemistry & Biology January 1995, **2**:7–12

Small molecules that bind in the minor groove of DNA frequently have cytotoxic activity because they interfere with the binding of proteins necessary for DNA replication and transcription. Some of these small molecules have proven very useful as antitumor agents because they selectively kill rapidly-dividing cells. This has stimulated efforts to design molecules that bind at designated sites in the minor groove. It is believed that groove binders with increased selectivity will produce a greater biological response for a given dose (and hence cause fewer toxic side effects) than non-selective groove binders [1]. Molecules that target particular DNA sites also have the potential to be used for the selective suppression of transcription from particular gene sequences [2,3]. But to design sequence-selective minor groove binders, one needs to know how the structures of various minor groove sites differ and how molecules sense these differences. Studies on existing minor groove binders, both small molecules and proteins, are leading to a better understanding of possible strategies for sequence-selective groove binding. Below I evaluate current views of the minor groove and then discuss how carbohydrate-based DNA binders, which are emerging as an important class of minor groove binders, fit into the picture for DNA recognition and new ligand design.

The evolving view of the minor groove

For many years the major groove was the focus of most studies aimed at understanding sequence-specific DNA recognition. This emphasis grew largely out of the belief that complementary networks of hydrogen bonds provide the primary basis for specific DNA recognition. There are more hydrogen bond donors and acceptors on the major groove edge of each base pair than on the minor groove edge. There are, therefore, more opportunities for discriminating different base pairs using hydrogen bonds from the major groove. In the minor groove, the principal difference between base pairs is that G–C

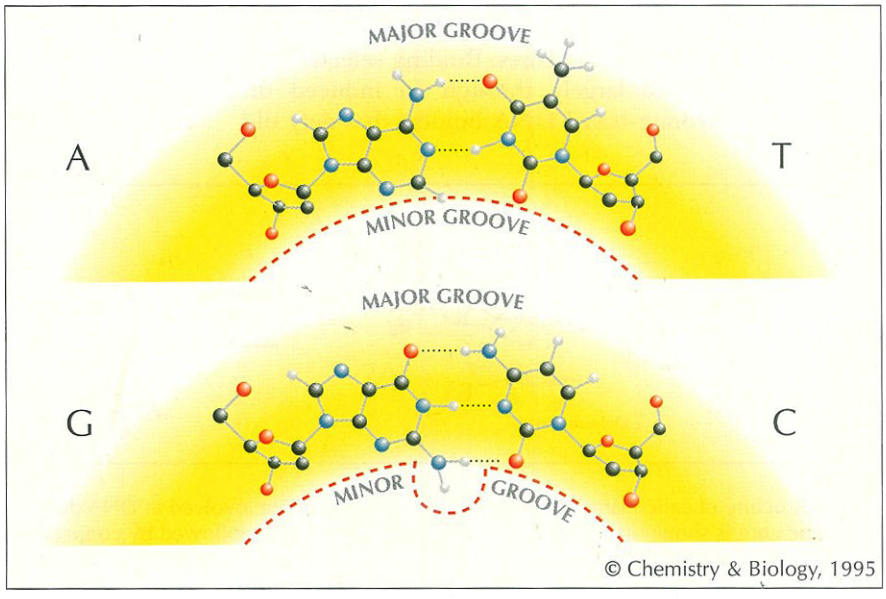

Fig. 1. Functionality in the minor groove at **(a)** A–T and **(b)** G–C base pairs. The exocyclic NH_2 group of the guanidinium protrudes into the minor groove (red outline, lower diagram).

base pairs contain an exocyclic amino group that protrudes into the groove. This amino group makes the steric and electronic environment of the minor groove at G–C base pairs profoundly different from that at A–T base pairs (Fig. 1). Thus, while designing ligands to discriminate A–T-rich and G–C-rich DNA sites in the minor groove seemed feasible, the potential for achieving greater sequence discrimination was regarded by many as limited [4].

This picture of the potential for sequence-selective binding in the minor groove has changed considerably in the past few years because of several developments. In 1989, Wemmer and colleagues [5] showed that distamycin, the prototypical A–T selective minor groove binder, can bind to DNA as an antiparallel dimer. This unexpected finding forced a re-evaluation of the nature of A–T selectivity and the role of a narrow minor groove in binding site selection. More importantly, the finding raised new possibilities for the design of minor groove binders that are selective for sequences containing mixed A–T and G–C base pairs, or for sequences containing only G–C base pairs, using a hydrogen bonding strategy (see below).

Attitudes towards minor groove recognition have also changed as more structural information about protein–DNA recognition has accumulated and ideas about how selectivity can be achieved have expanded to include mechanisms other than complementary arrays of hydrogen bonds. For example, sequence-dependent DNA flexibility, or the ability of particular runs of base pairs to distort to provide a complementary binding surface for proteins or other ligands, can provide a high degree of binding selectivity without the formation of specific arrays of hydrogen bonds. In a striking example of this, the crystal structure of a minor groove TATA box-binding protein (TBP) shows that it induces severe bending of the TATA box and widening of the minor groove [6]. There are only five protein–base hydrogen bonds in the complex, leaving unsatisfied thirteen of seventeen possible hydrogen bond acceptors on the minor groove edges of the bases. Binding selectivity is apparently achieved largely through an induced fit. Finally, carbohydrate-based DNA binders that defy old paradigms for minor groove recognition have recently been identified. These molecules are also helping to change perceptions about how much selectivity is possible in minor groove binding and how that selectivity can be achieved.

Carbohydrate-based minor groove binders

DNA binders that contain carbohydrates have been known for over thirty years. For most of that time, however, the carbohydrate portions of DNA-binding glycoconjugates were believed to contribute little to the selectivity that the molecules displayed. That view changed rapidly with the discovery of the enediyne antitumor antibiotic calicheamicin. Calicheamicin consists of a bicyclic enediyne moiety attached to a tetrasaccharide-aryl tail (Fig. 2). Under reducing conditions, the enediyne rearranges to produce a 1,4-aryl diradical, which abstracts hydrogen atoms from the DNA backbone initiating DNA strand scission. Studies on the cleavage selectivity of calicheamicin showed that it displays a new kind of specificity. It is neither A–T selective nor G–C selective [7]. Instead, it seems to bind to a variety of sites containing three or more pyrimidines in a row. The steric and electronic environments of many of the sites appear to be very different because some contain guanine amino groups and some do not [7,8]. Any mechanism for achieving binding selectivity must be able to explain calicheamicin's affinity for sites which present very different functionality. Investigations from several laboratories have shown that the carbohydrate portion of calicheamicin is the principal DNA-binding element and is largely responsible for the oligopyrimidine selectivity [9–11]. Based on an analysis of different pyrimidine-rich binding sites, we proposed that the binding selectivity involves an induced fit process [8].

The first structural evidence for an induced fit recognition process came from circular dichroism measurements by Sugiura and coworkers [12] and NMR experiments in our laboratories [13]. Our NMR studies showed that calicheamicin binds in the minor groove with the oligosaccharide tail contacting the bases in the recognition sequence (Fig. 3). The oligosaccharide curves gently to follow the winding path of the minor groove. The curvature in the oligosaccharide is largely due to the

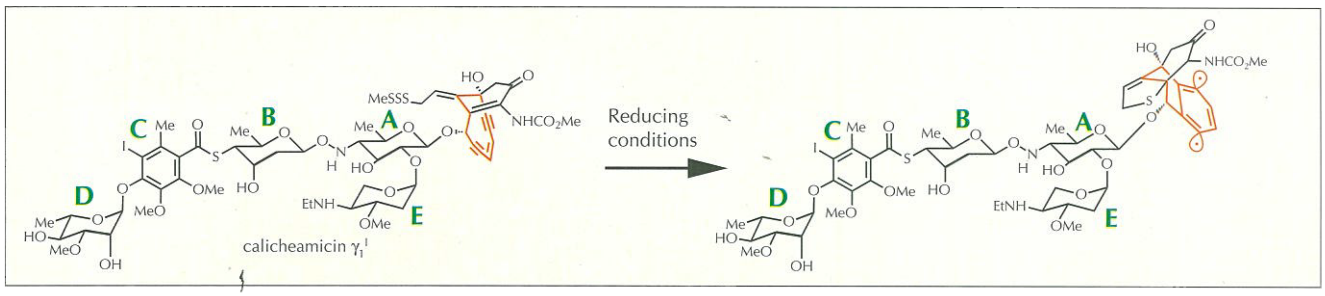

Fig. 2. Structure of calicheamicin γ_1^I and diradical intermediate involved in DNA damage. The ten-membered benzenoid system of the enediyne core is shown in red. Cleavage of the allylic trisulfide followed by conjugate addition of the resulting thiolate to the bridgehead double bond triggers the cyclization to form the diradical, which can attack DNA, initiating cleavage. Elements A–E of the oligosaccharide tail are labelled.

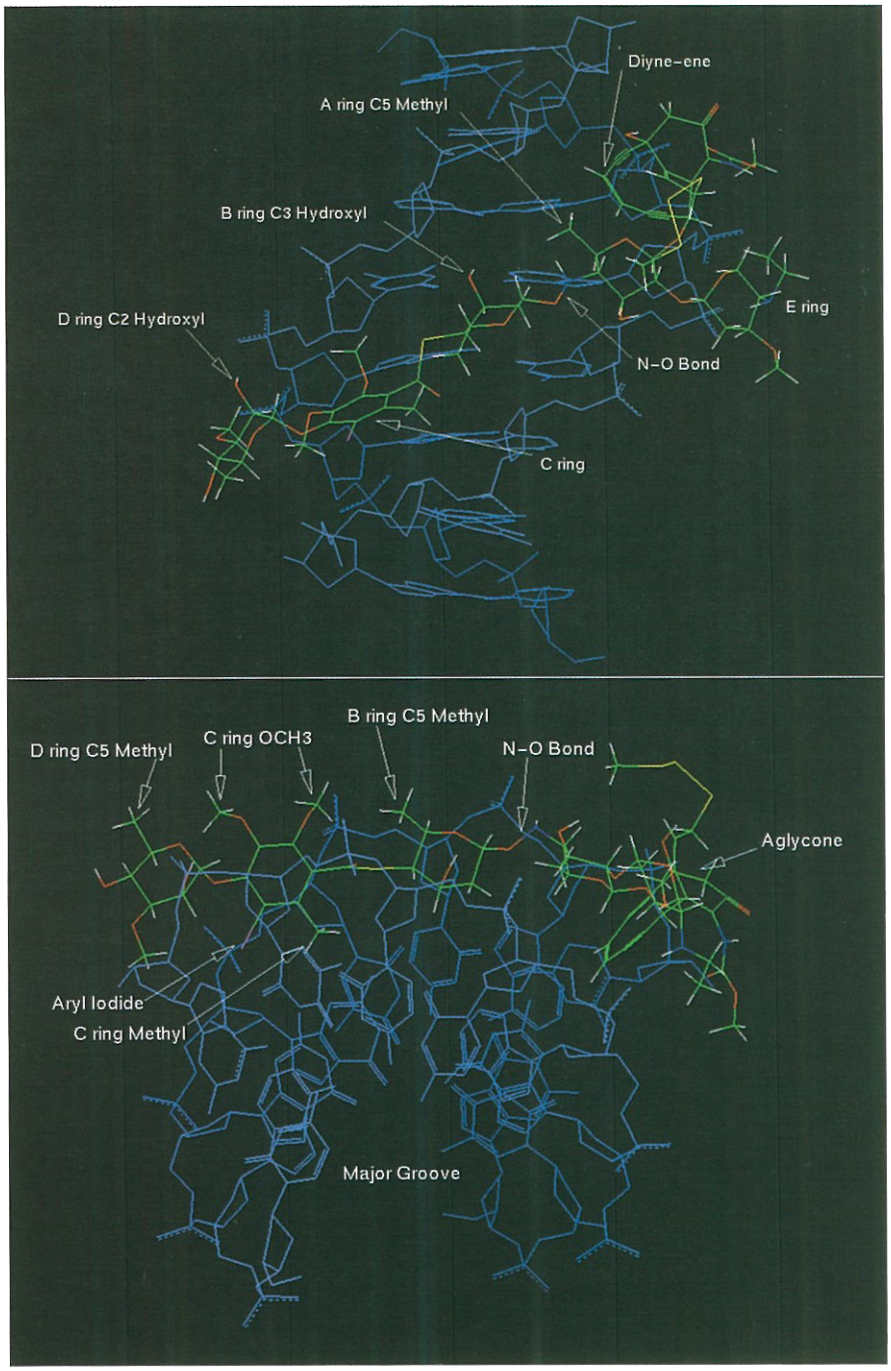

Fig. 3. Two views of calicheamicin bound in the minor groove of DNA. The front view (top) shows calicheamicin tracking along the minor groove. The side view (bottom) shows which functional groups on calicheamicin point towards the floor of the groove and which point out towards the solvent.

N–O bond in calicheamicin, which enforces an unusual conformation between the A and B sugars and allows the carbohydrate tail to track along the minor groove [13–15]. The position of the oligosaccharide with respect to the recognition sequence is essentially identical regardless of whether the sequence recognized is ACCT [13], TTTT [16], or TCCT [17]. The similar positioning is further evidence that oligopyrimidine sequences have some structural features in common that calicheamicin is able to sense. These structural features are not common hydrogen bonding arrays. Although one typically thinks of oligosaccharides as heavily hydroxylated molecules with a large number of potential hydrogen bond donors and acceptors, the calicheamicin oligosaccharide is unusually hydrophobic and contains only a small number of hydrogen-bonding partners. Of these, most do not contact the bases in the floor of the minor groove, but instead are directed towards the ribose–phosphate backbone. These hydrogen bonds appear to function primarily to anchor the molecule in the groove. In any event, there are no electrostatic contacts between calicheamicin and the floor of the groove which are common to all the different recognition sequences. Instead, changes observed in some of the DNA helix parameters upon binding calicheamicin (including an increase in groove width and changes in the conformation of some of the ribose sugars) indicate that pyrimidine sequences may share an ability to distort

readily to accommodate the drug. The distortion is much less dramatic than that produced by TATA binding proteins, but may be sufficient to explain the energetic differences that lead to specificity of binding.

There are a large number of other DNA binders that contain sugars as components, although in many cases the role of the sugars is still unclear. Calicheamicin is the first example of a DNA-binding molecule in which the carbohydrate portion itself binds site-selectively to DNA [9,10], but several examples are known in which the carbohydrates are essential for binding ([18], and references therein). A comparison of calicheamicin with these other DNA-binding glycoconjugates reveals similarities in the structures of the carbohydrates that may prove useful in design. For example, like the calicheamicin oligosaccharide, the carbohydrates in the aureolic acid antibiotics are quite hydrophobic [19]. Many of the sugars are 2,6-dideoxy sugars. Of the remaining hydroxyls, many are alkylated or acylated (Fig. 4). The hydrophobicity of the sugars is probably significant in helping to drive the binding [20]. The hydrophobicity makes DNA-binding sugars very different from the hydrophilic cell-surface carbohydrates involved in protein recognition. Other minor groove binding glycoconjugates such as chromomycin A_3 also distort DNA upon binding [21] (Fig. 4). Hence, sequence-dependent DNA flexibility may be significant in the site-selectivity displayed by many carbohydrate-containing minor groove binders. The relative rigidity of carbohydrates may be important in forcing the DNA to adapt to the shape of the carbohydrate molecule.

Strategies for the design of new minor groove binders

One of the most effective strategies for the design of site-selective minor groove binders is to construct derivatives of the netropsin/distamycin class of minor groove binders ,in which some of the individual pyrrole rings are replaced by other structures (such as imidazole or pyridine) intended to impart selectivity for G–C base pairs [22, 23]. The idea is to use a known minor groove binding scaffold but to modify it to include

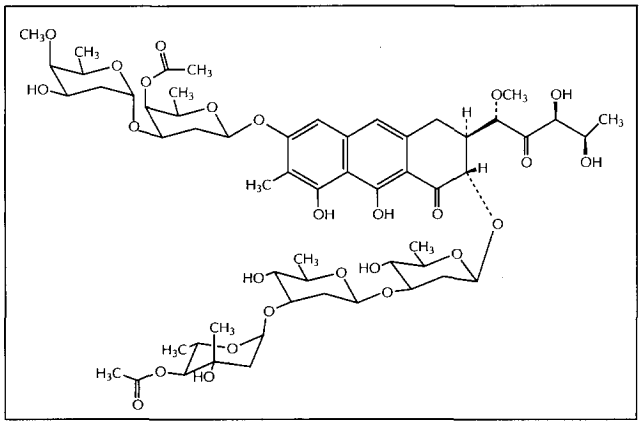

Fig. 4. Structure of chromomycin A_3, a member of the aureolic acid class of DNA-binding molecules. Many of the OH groups on the sugars on this class of molecules are either alkylated or acylated.

hydrogen-bond acceptors that can contact the amino groups in the minor groove at G–C base pairs [24]. The first efforts to implement this strategy were unsuccessful in the sense that they produced binders with unanticipated selectivity that could not easily be rationalized [23]. Wemmer's finding that distamycin can bind to DNA as an antiparallel dimer [5] provided the clue that allowed others to interpret the results of some of their efforts to make G–C-selective binders by modifying the distamycin/netropsin framework. For example, Dervan's efforts to design ligands that recognize DNA sequences including G–C base pairs suddenly made sense if the ligands were assumed to bind to DNA as dimers [25]. Since Wemmer's report in 1989 [5], the laboratories of both Dervan [26] and Lown [27] have made a number of modified ligands based on the distamycin/netropsin motif that bind as antiparallel dimers at designated sequences in the minor groove of DNA. In a particularly dramatic result, Dervan, Wemmer and coworkers [28] designed and characterized a molecule that binds as a dimer to the sequence 5'-(A,T)GCGC(A,T)-3'. The specificity is due largely to a network of hydrogen bonds between each ligand and each strand of the DNA. Thus, now that a scaffold has been identified that provides for enough hydrogen bonding contacts to each base pair to allow for better discrimination [4], it appears to be possible to design selective minor groove binders targeted to a wide range of different DNA sequences, including mixed sequences and G–C sequences, using patterns of hydrogen bonds.

The carbohydrate-containing minor groove binders suggest another strategy for design. An induced fit process has a large role in the site-selective binding of both calicheamicin and the aureolic acid antibiotics such as chromomycin, mithramycin and olivomycin to the minor groove of DNA, and one might envision a design strategy focused on exploiting sequence-dependent DNA flexibility. The difficulty in implementing such a strategy is two-fold: first, the relationship between DNA sequence and the ability to undergo particular deformations is not currently well understood, and, second, even if more were known about which types of sequences share an ability to deform in a similar manner, designing ligands *de novo* to exploit this flexibility would be difficult. But just as the netropsin/distamycin class of antibiotics provided the starting point for implementing a design strategy emphasizing hydrogen bonds, calicheamicin and the other carbohydrate-containing DNA binders provide blueprints for the design of ligands that sense DNA flexibility.

Studies on calicheamicin have suggested that pyrimidine-rich sequences share an ability to adapt their conformation in a particular way. Starting with scaffolds related to calicheamicin, it may be possible to design other oligopyrimidine-selective binders and thereby learn more about what is important in determining the selectivity for drug binding. The structural work on calicheamicin suggests that the shape of the molecule,

which is primarily determined by the conformation of the linkages between sugars, is critical in determining its selectivity of binding. It may be possible to design simplified ligands that have similar shapes to calicheamicin by preserving the glycosidic linkages while altering the individual sugars so that they are easier to produce. In a similar manner, it should be possible to use chromomycin and other aureolic acid antibiotics as blueprints for designing glycoconjugates that are G–C-selective binders, thereby clarifying the contributions of induced fit and hydrogen bonding to selectivity. Again, by determining which features are critical for binding and which are not, it may be possible to design radically simplified systems [29]. Ultimately, these simplified systems may provide a starting point for the design of ligands with altered or increased selectivity.

Conclusion

Site-selective minor groove binding ligands have a number of important uses. Above I have described two strategies for the design of new minor groove binding ligands based on existing scaffolds. One strategy emphasizes the role of hydrogen bonding in determining selectivity. By designing molecules with hydrogen bonding partners complementary to particular sequences, it is possible to achieve a high degree of selectivity in minor groove binding.

The second strategy emphasizes induced fit as a mechanism for achieving selectivity. Relatively rigid ligands which force the DNA to mold around them can bind quite selectively as they exploit differences in DNA flexibility at different sequences. It is not yet known if the same degree of binding selectivity can be achieved using this strategy as using a hydrogen-bonding strategy. Nevertheless, this strategy may have some advantages that a 'pure' hydrogen bonding strategy does not have. For example, it has been shown that both mithramycin and the calicheamicin oligosaccharide can displace transcription factors that bind in the major groove [2,3]. The displacement is presumably related to the effects that these minor groove binders have on DNA structure. In contrast, binding of netropsin in the minor groove is compatible with the binding of at least some transcription factors in the major groove [30]. (Netropsin is very effective at preventing the binding of the general transcription factor TATA box binding protein, which binds in the minor groove [31].) These findings suggest that in order to disrupt transcription of a particular gene with a minor groove binder, one wants a molecule that is selective for the regulatory region in question and that affects the DNA conformation in a way that prevents binding of the necessary transcription factors. Ultimately, of course, this goal will probably be best met by developing hybrid approaches to the design of minor groove ligands in which both sequence-dependent DNA flexibility and hydrogen bonding have a role in determining the preferred sites (as they are thought to do in the case of the aureolic acid antibiotics). The work that is currently being done to understand existing minor groove binders will lay the groundwork for more sophisticated strategies to the design of new minor groove binding ligands.

References

1. Lown, J.W. (1990). Molecular mechanisms of DNA sequence recognition by groove binding ligands: biochemical and biological consequences. In *Molecular Basis of Specificity in Nucleic Acid–Drug Interactions.* (Pullman, B. & Jortner, J., eds), pp. 103–122, Kluwer Academic Publishers, Netherlands.
2. Snyder, R.C., Ray, R., Blume, S. & Miller, D.M. (1991). Mithramycin blocks transcriptional initiation of the c-myc P1 and P2 promoters. *Biochemistry* **30**, 4290–4297.
3. Ho, S.N., Boyer, S.H., Schreiber, S.L., Danishefsky, S.J. & Crabtree, G.R. (1994). Specific inhibition of formation of transcription complexes by a calicheamicin oligosaccharide: a paradigm for the development of transcriptional antagonists. *Proc. Natl. Acad. Sci. USA* **91**, 9203–9207.
4. Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804–808.
5. Pelton, J.G. & Wemmer, D.E. (1989). Structural characterization of a 2:1 distamycin A-d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723–5727.
6. Kim, J.L. & Burley, S.K. (1994). 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat. Struct. Biol.* **1**, 638–653.
7. Walker, S., Landovitz, R., Ding, W.-D., Ellestad, G.A. & Kahne, D. (1992). Cleavage behavior of calicheamicin γ^I and calicheamicin T. *Proc. Natl. Acad. Sci. USA* **89**, 4608–4612.
8. Lee, M.D., Ellestad, G.A. & Borders, D.B. (1991). Calicheamicins: discovery, structure, chemistry, and interaction with DNA. *Accounts Chem. Res.* **24**, 235–243.
9. Aiyar, J., Danishefsky, S.J. & Crothers, D.M. (1992). Interaction of the aryl tetrasaccharide domain of calicheamicin γ_1^I with DNA: influence on aglycon and methidiumpropyl-EDTA-Iron(II)-mediated DNA cleavage. *J. Am. Chem. Soc.* **114**, 7552–7554.
10. Nicolaou, K.C., Tsay, S.-C., Suzuki, T. & Joyce, G.F. (1992). DNA-carbohydrate interactions. Specific binding of the calicheamicin γ_1^I oligosaccharide with duplex DNA. *J. Am. Chem. Soc.* **114**, 7555–7557.
11. Li, T., Zeng, Z., Estevez, V.A., Baldenius, K.U., Nicolaou, K.C. & Joyce, G.F. (1994). Carbohydrate–minor groove interactions in the binding of calicheamicin γ^I to duplex DNA. *J. Am. Chem. Soc.* **116**, 3709–3715.
12. Uesugi, M. & Sugiura, Y. (1993). New insights into sequence recognition process of esperamicin A_1 and calicheamicin γ^I: origin of their selectivities and 'induced fit' mechanism. *Biochemistry* **32**, 4622–4627.
13. Walker, S., Murnick, J. & Kahne, D. (1993). Structural characterization of a calicheamicin–DNA complex by NMR. *J. Am. Chem. Soc.* **115**, 7954–7961.
14. Walker, S., Yang, D., Gange, D. & Kahne, D. (1991). Conformational analysis of the N–O bond in the calicheamicin oligosaccharide. *J. Am. Chem. Soc.* **113**, 4716–4717.
15. Walker, S., Gange, D., Gupta, V. & Kahne, D. (1994). Analysis of hydroxylamine glycosidic linkages: structural consequences of the N–O bond in calicheamicin. *J. Am. Chem. Soc.* **116**, 3197–3206.
16. Walker, S.L., Andreotti, A.H. & Kahne, D.E. (1994). NMR characterization of calicheamicin γ^I bound to DNA. *Tetrahedron* **50**, 1351–1360.
17. Paloma, L.G., Smith, J.A., Chazin, W.J. & Nicolaou, K.C. (1994). Interaction of calicheamicin with duplex DNA: role of the oligosaccharide domain and identification of multiple binding modes. *J. Am. Chem. Soc.* **116**, 3697–3708.
18. Silva, D.J. & Kahne, D.E. (1993). Studies of the 2:1 chromomycin A_3–Mg^{2+} complex in methanol: role of the carbohydrates in complex formation. *J. Am. Chem. Soc.* **115**, 7962–7970.
19. Walker, S., Valentine, K.G. & Kahne, D. (1990). Sugars as DNA binders: a comment on the calicheamicin oligosaccharide. *J. Am. Chem. Soc.* **112**, 6428–6429.
20. Ding, W.-D. & Ellestad, G.A. (1991). Evidence for hydrophobic interaction between calicheamicin and DNA. *J. Am. Chem. Soc.* **113**, 6617–6620.
21. Gao, X., Mirau, P. & Patel, D.J. (1992). Structure refinement of the chromomycin dimer–DNA oligomer complex in solution. *J. Mol. Biol.* **223**, 259–279.
22. Lown, J.W., Krowicki, K., Bhat, U.G., Skorobogaty, A., Ward, B. & Dabrowiak, J.C. (1986). Molecular recognition between oligopeptides

and nucleic acids: novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity. *Biochemistry* **25**, 7408–7416.
23. Wade, W.S. & Dervan, P.B. (1987). Alteration of the sequence specificity of distamycin on DNA by replacement of an N-methylpyrrole-carboxamide with pyridine-2-carboxamide. *J. Am. Chem. Soc.* **109**, 1574–1575.
24. Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R.E. (1985). The molecular origin of DNA–drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. USA* **82**, 1376–1380.
25. Wade, W.S., Mrksich, M. & Dervan, P.B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' Sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783–8794.
26. Mrksich, M. & Dervan, P.B. (1994). Design of a covalent peptide heterodimer for sequence-specific recognition in the minor groove of double-helical DNA. *J. Am. Chem. Soc.* **116**, 3663–3664.
27. Chen, Y.-H. & Lown, J.W. (1994). A new DNA minor groove binding motif: cross-linked lexitropsins. *J. Am. Chem. Soc.* **116**, 6995–7005.
28. Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1994). Design of a GC-specific DNA minor groove-binding peptide. *Science* **266**, 646–650.
29. Silva, D.J., Kraml, C.M. & Kahne, D. (1994). Chromomycin A_3 as a blueprint for designed metal complexes. *J. Am. Chem. Soc.* **116**, 2641–2642.
30. Oakley, M.G., Mrksich, M. & Dervan, P.B. (1992). Evidence that a minor groove-binding peptide and a major groove-binding protein can simultaneously occupy a common site on DNA. *Biochemistry* **31**, 10969–10975.
31. Chiang, S.-Y., Welch, J., Rauscher III, F.J., & Berman, T.A. (1994). Effects of minor groove binding drugs on the interaction of TATA box binding protein and TFIIA with DNA. *Biochemistry* **33**, 7033–7040.

Daniel Kahne, Department of Chemistry, Princeton University, Princeton, NJ 08544-1009, USA