Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
ELSEVIER
Bioorganic & Medicinal Chemistry Letters 14 (2004) 1353–1356

DNA binding of a short lexitropsin

Nahoum G. Anthony,[b,c] Keith R. Fox,[a] Blair F. Johnston,[c] Abedawn I. Khalaf,[b] Simon P. Mackay,[c] Iain S. McGroarty,[c] John A. Parkinson,[b] Graham G. Skellern,[c] Colin J. Suckling[b,*] and Roger D. Waigh[c]

[a]*Department of Physiology and Pharmacology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO9 3TU, UK*
[b]*Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, UK*
[c]*Department of Pharmaceutical Sciences, University of Strathclyde, 27 Taylor Street, Glasgow G4 0NR, UK*

Received 25 July 2003; revised 12 November 2003; accepted 18 November 2003

Abstract—Footprinting, capillary electrophoresis, molecular modelling and NMR studies have been used to examine the binding of a short polyamide to DNA. This molecule, which contains an isopropyl-substituted thiazole in place of one of the *N*-methylpyrroles, is selective for the sequence 5′-ACTAGT-3′ to which it binds with high affinity. Two molecules bind side-by-side in the minor groove, but their binding is staggered so that the molecule reads six base pairs, unlike the related natural products, which tend to bind to four-base-pair sequences. The result suggests that high affinity and selectivity may be gained without resort to very large molecules, which may be difficult to deliver to the site of action.

1. Introduction

Since the discovery of netropsin and distamycin and the observation that they could bind side-by-side in the minor groove of DNA, with selectivity for A/T sequences, there has been a great deal of work aimed at achieving specific binding to any given base-pair sequence.[1–3] The potential for such an approach in the therapy of diseases such as cancer scarcely needs to be elaborated, at a time when the human genome has been published.

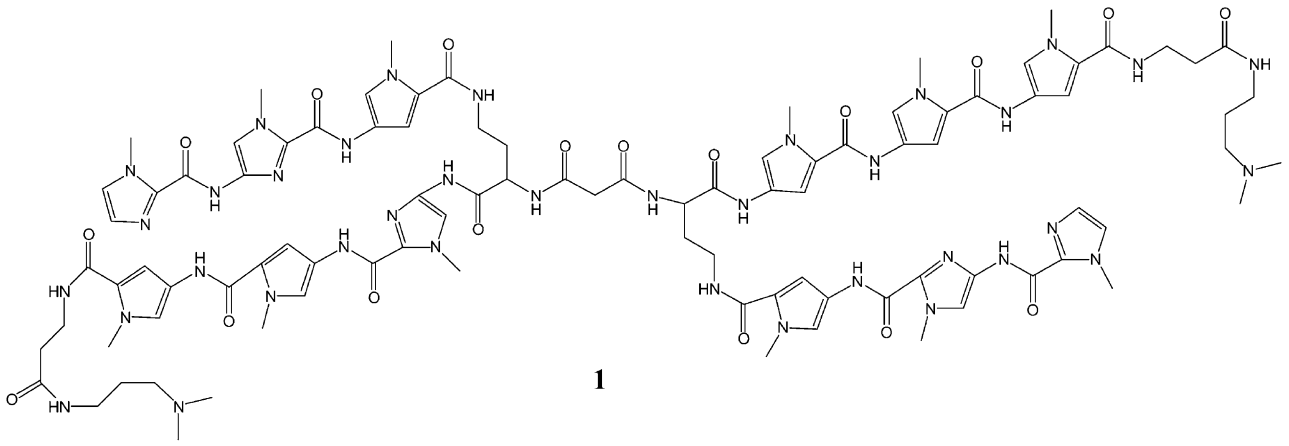
The excellent work that has been carried out on polyamide minor groove binders, for example by the groups of Dervan[1] and Lown[4] has resulted in compounds able to bind very selectively to target sequences, but at the price of molecular size. Typically, molecules which will read base pair sequences of the required 10 base pair length or more have been based[5] on hairpin structures such as **1**, with molecular weights over 2000 Da. While this has allowed convincing demonstrations of their ability to interrupt biochemical processes in vitro,[6] the 'rule of five' is violated in at least two respects; the molecular weight is too high and there are too many hydrogen bond donors.[7] Molecules such as **1** would be expected to have difficulty in passing through biological membranes and might not reach the desired site of action, even if given by injection. Despite this, there are some indications of biological activity in whole animals, notably in the fruit fly *Drosophila melanogaster*[8] but there is no doubt that smaller molecules would be more likely to become therapeutic entities.

Most of the published work has reflected the assumption that selectivity and affinity of binding will depend on the ability of the ligand to hydrogen bond to the bottom of the minor groove, where the bases offer a sequence-dependent pattern of H-bond donors and acceptors. This assumption has given useful and interesting results, but calorimetric data suggest that the primary driving force is hydrophobic, the main energy benefit of binding coming from the loss of structured water in the groove and around the ligand.[9]

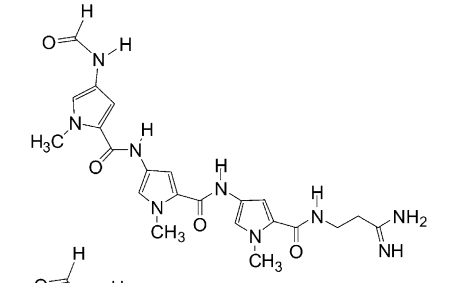
Computer modelling of DNA and of typical ligands, such as distamycin **2**, shows that there is a shape mis-

* Corresponding author. Tel.: +44-141-548-2271; fax: +44-141-548-4733; e-mail: c.j.suckling@strath.ac.uk

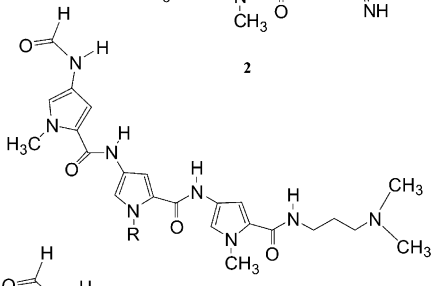
0960-894X/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2003.11.068



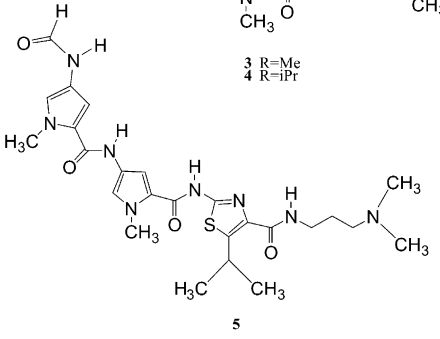
1



2



3 R=Me
4 R=iPr



5

match. Distamycin, although capable of adopting the required helical shape to fit the minor groove, is largely sp^2 and planar, whereas the surface of the minor groove is very uneven, reflecting the tetrahedral nature of the atoms constituting the DNA backbone (Fig. 1).

Our approach to the design of distamycin analogues has incorporated a strategy to raise the oil/water partition coefficient, which should aid binding as well as penetration through lipid membranes, and at the same time to change the shape of the molecule to potentially be capable of following the contours of the minor groove. Such an approach is of necessity partially empirical, since the minor groove is flexible and capable of adapting to fit a variety of ligands. Particularly significant is the ability of DNA to accommodate ligands either singly or side-by-side in the minor groove, depending on the base-pair sequence. It is obvious that the design of ligands that can bind singly, for at least part of their binding contact, with high affinity, will allow the base-pair code to be read with smaller molecules.

2. Results and discussion

Our first attempts to design analogues which would bind in a 1:1 ratio with DNA double helices centred on the introduction of branched chain alkyl groups; these would be expected to sterically obstruct side-by-side binding and would raise the oil/water partition coefficient, with potential benefits for membrane penetration. DNase I footprinting studies with the tyrT DNA fragment,[10] which has been widely used in other footprinting studies showed that the compounds **3** and **4** bound at AT-rich sites, producing similar patterns to those seen with distamycin (Fig. 2). Compound **4** bound more tightly (0.3 μM) than compound **3** (3 μM) as predicted from its more hydrophobic nature.

The results with **4** led us to synthesize analogues with branched alkyl side chains at different positions. The use of thiazole instead of pyrrole was particularly attractive, since the sulfur is large and will itself have a major effect on partitioning into biological membranes. Lexitropsin **5** has both a thiazole and an isopropyl side chain, which renders the sub-unit nearest to the positively charged tail very bulky. According to Dervan's rules,[1] compound **5** should be a selective binder for A/T base pairs where two pyrroles overlap in a side-by-side manner, but the thiazole should be tolerant of both AT and GC base pairs. In theory, **5** should bind to any sequence with four or more A/T base pairs and to many more which have a C/G base pair to either side of an A/T sequence, providing there at least three A/T pairs.

DNase I footprinting experiments with this compound showed a very different cleavage pattern from that expected, with a single region of protection around position 40 of tyrT DNA in the sequence 5′-AACTGGT (Fig. 2). This ligand clearly possesses a very different

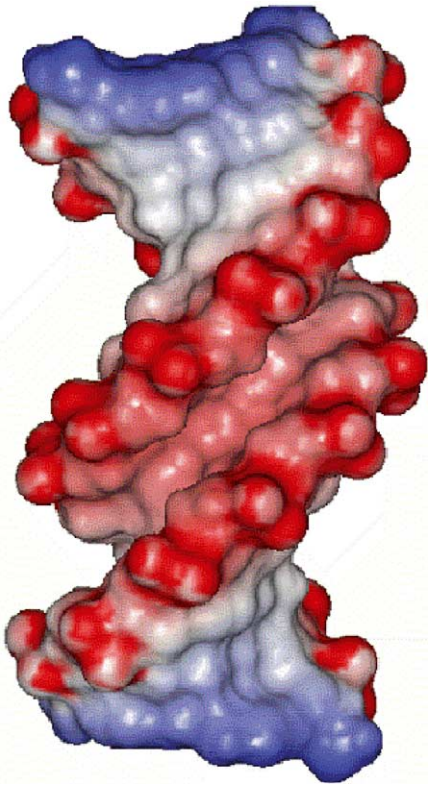


Figure 1. Surface of DNA minor groove in a region composed of A/T base pairs only. Red coloration indicates areas of negative charge.

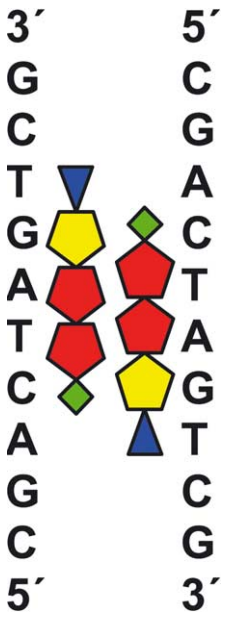


Figure 3. Schematic representation of the minor-groove binding mode between **5** and d(CGACTAGTCG) showing the six base-pair overlap which results from the staggered nature of the relationship between the two occupying ligands. Green, formyl 'head'; red, *N*-methyl pyrrole; yellow, thiazole; blue, DMAP 'tail'.

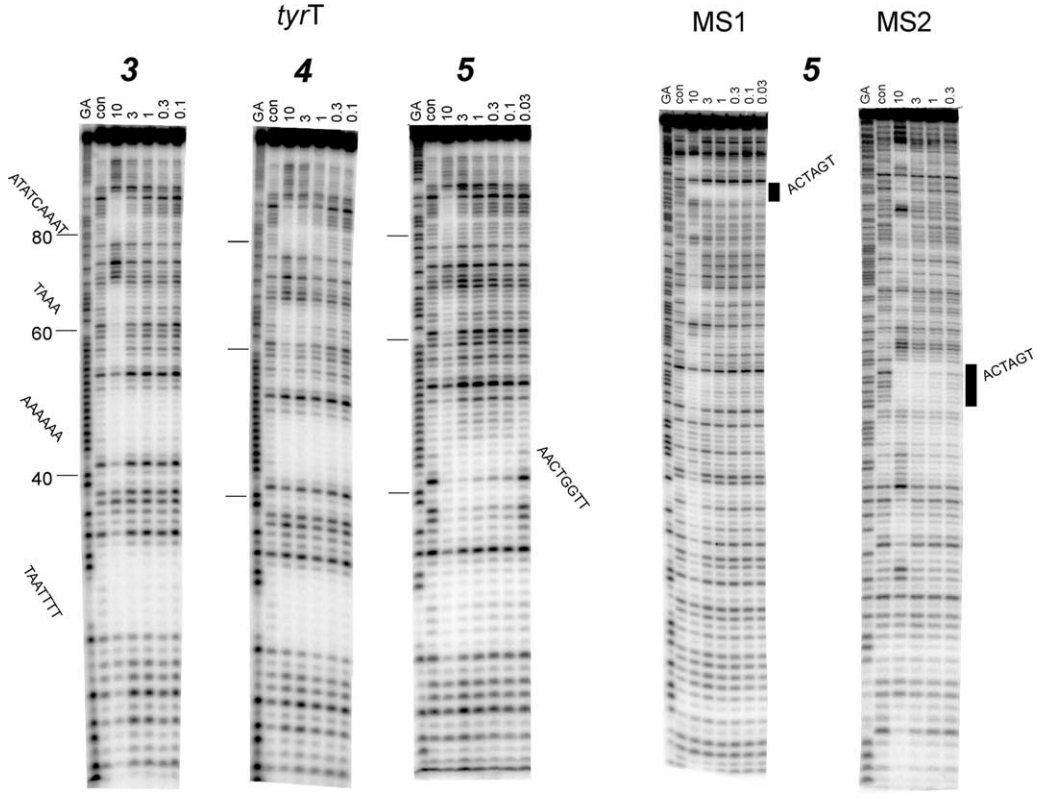


Figure 2. DNase I footprinting data for lexitropsins **3**, **4** and **5**. The first three panels used *tyr*T DNA as the footprinting substrate; the numbering scheme is the same as in previous publications.[10] The last two panels show the interaction of **5** with fragments MS1 and MS.[11] These fragments contain the same sequence in opposite orientations and are designed to contain all possible four-base-pair sequences. The ligand concentration (μM) is shown at the top of each gel lane; con indicates control while GA is a sequence marker specific for purines. The location of the footprints is indicated.

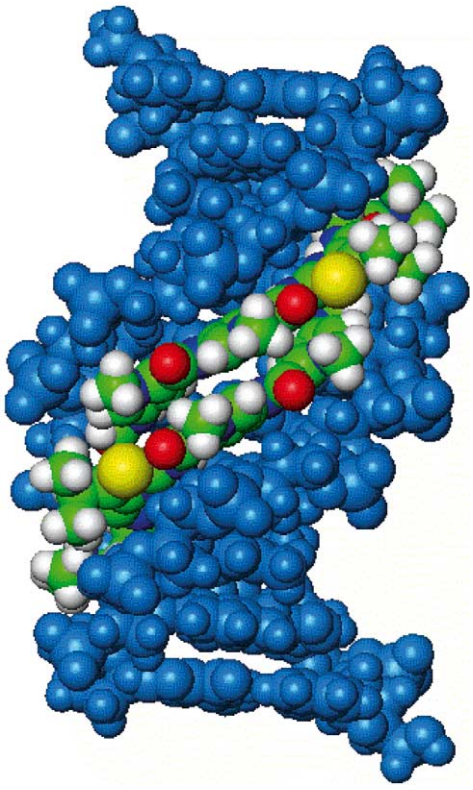


Figure 4. NMR-derived structure of the complex between **5** and 5′-CGACTAGTCG.

binding preference to the parent compounds; in order to properly characterize the preferred binding site(s) we performed further footprinting experiments with a fragment that contains every tetranucleotide sequence.[11] This revealed a single binding site at ligand concentrations below 3 μM, centred on the sequence 5′-ACTAGT (Fig. 2). This persisted to concentrations below 0.1 μM. Comparative binding studies using capillary electrophoresis with the DNA 12-mer 5′-AAATTATATTAT previously described[12] and the 10-mer 5′-CGACTAGTCG-3′ confirmed the selectivity of binding to the latter sequence.

NMR studies of binding between **5** and the oligonucleotide CGACTAGTCG were carried out using various protocols. Titration of **5** against an excess of the double helical oligonucleotide confirmed its high affinity for the minor groove. The data revealed 2:1 binding and confirmed the contribution of the isopropyl groups to ligand–ligand and ligand–groove binding, with part of the ligand overlapped, as shown schematically in Figure 3. Addition of excess **5** up to a ratio of 4:1 resulted in NMR signal broadening but with no additional changes in the appearance of the NMR data. Together with a reverse titration experiment in which the oligonucleotide was added to an excess of **5**, the data indicate aggregation and implicate a second, non-specific DNA binding mode for **5** once the primary binding site is fully occupied.

Partial overlap (Fig. 3), involving only the pyrrole bases of **5**, has the useful feature of extending the binding site, compared to simple analogues which tend to show overlap of all three rings, as expected for **3**. It is apparent that the isopropyl thiazole in **5** is too bulky to fit side-by-side, as shown in the computer graphic (Fig. 4), but may offer a useful 'step-down' function, allowing a smooth transition between the doubly occupied and totally unoccupied parts of the groove. NMR studies are planned for **4**, to determine whether the *N*-isopropyl function in **4** will prevent side-by-side binding.

The results with **5** are significant, since they demonstrate that it is possible to achieve high affinity and selectivity of DNA binding with relatively small molecules. It is hoped that analogues of **5** will show similar discrimination for other sequences and that extension of **5** will give structures that will be capable of reading the 10–12 base pair sequences which will increase the scope for biological activity.

Acknowledgements

We thank the Glasgow/Strathclyde Synergy fund for their support of this work, MRC for an allocation of NMR instrument time and the NMR staff at the National Institute for Medical Research, Mill Hill, UK for their assistance and cooperation. We thank the staff of the University of Edinburgh NMR facility for enabling NMR pilot studies on this research programme to be carried out.

References and notes

1. Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.
2. Wemmer, D. E. *Biopolymers (Nucl. Acid Sci.)* **2001**, *52*, 197.
3. Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291.
4. Sharma, S. K.; Reddy, B. S. N.; Lown, J. W. *Drugs Fut.* **2001**, *26*, 39.
5. Weyermann, P.; Dervan, P. B. *J. Am. Chem. Soc.* **2002**, *124*, 6872.
6. For example: Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218.
7. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Del. Rev.* **1997**, *23*, 3.
8. Janssen, S.; Durussel, T.; Laemmli, U. K. *Mol. Cell* **2000**, *6*, 999.
9. Haq, I.; Ladbury, J. E.; Chowdrhy, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997**, *271*, 244.
10. Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271.
11. Lavesa, M.; Fox, K. R. *Anal. Biochem.* **2001**, *293*, 246.
12. Hamdan, I. I.; Skellern, G. G.; Waigh, R. D. *Nucleic Acids Res.* **1998**, *26*, 3053.