Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 15–19

Targeting the DNA minor groove with fused ring dicationic compounds: Comparison of in silico screening and a high-resolution crystal structure

Nancy H. Campbell,[a] David A. Evans,[a,b] Michael P. H. Lee,[a] Gary N. Parkinson[a] and Stephen Neidle[a,*]

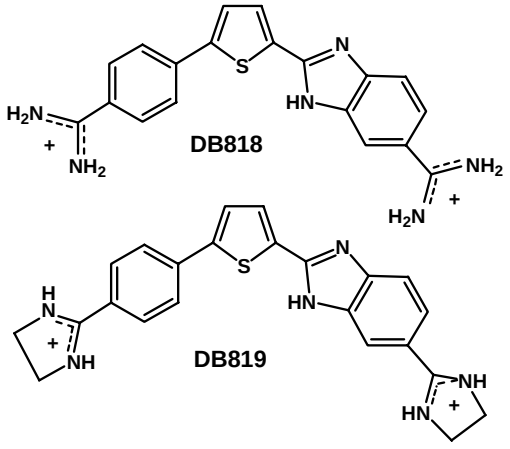
[a]*Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, London WC1N 1AX, UK*
[b]*Spirogen Ltd, 29-39 Brunswick Square, London WC1N 1AX, UK*

Received 29 September 2005; accepted 12 October 2005

Abstract—The crystal structure of the DNA minor groove phenyl benzimidazole diamidine ligand DB819 has been determined, bound to the DNA sequence d(CGCGAATTCGCG)$_2$, at a resolution of 1.36 Å. Conditions for reliable in silico docking that reproduce the observed position of the ligand in the minor groove have been determined.

The minor groove of DNA is the target for a wide range of anticancer, antiviral and antiprotozoal agents. Non-covalently binding molecules tend to bind to the minor groove of A/T-rich sequences,[1] and several are of current clinical use.[2] They are preferentially taken up, often by active transport mechanisms, into susceptible cell types, and exert their effect by blocking topoisomerases, and sometimes by acting as global inhibitors of transcription. Pentamidine is employed against antimony-resistant leishmaniasis, primary stage human African trypanosomiasis (HAT) and for AIDS-related *Plasmodium jiroveci* pneumonia.[3] An orally effective prodrug of the bis-amidino-phenyl-substituted furan drug furamidine is currently in phase II clinical trials against malaria and *Pneumocystis carinii* pneumonia,[4] and is scheduled for phase III trials against HAT.[5]

A series of biphenyl benzimidazole diamidines we have previously reported,[6] with several members having nanomolar activity against *Trypanosoma brucei rhodesiense*, and some showing in vivo activity against a murine model for this disease. In a search for analogues with improved DNA binding and biological properties, the diamidino-phenyl-benzimidazole compound DB818 was recently prepared,[7] with the thiophene ring replacing the furan ring in the compound DB293,[6,8] which has the effect of significantly enhancing DNA binding ability. An X-ray crystallographic study of DB818 bound to the DNA duplex sequence d(CGCGAATT CGCG)$_2$ was able to rationalise the binding and thermodynamic data.[7] It showed that the replacement of the oxygen atom by sulfur resulted in a widening of the concave inner surface so that the amidinium groups of DB818 are in hydrogen bond contact with A/T base pair edges, whereas the geometry of DB293 precludes this occurring.



Keywords: DNA minor groove; Crystallography; Oligonucleotide; In silico screening.
* Corresponding author. Tel.: +44 0207 753 5969; fax: +44 0207 753 5970; e-mail: bmcl@ulsop.ac.uk

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.10.036

The increasing number of diamidinium and related compounds is creating a demand for rapid and reliable ways of screening for DNA binding and biological activity.

DNA binding is an important determinant of biological response to these minor groove drugs, and the crystal structures of drug complexes have provided insight into the factors that govern the binding. We have therefore started to examine whether in silico screening could provide a viable approach to lead compound selection in this area, as it has in other therapeutic areas.[9] This approach requires a high-quality crystal structure against which the results of in silico docking can be assessed. We have used the structure of a complex of the closely related compound DB819, which was co-crystallised[10] with the DNA sequence d(CGCGAATTCGCG)$_2$. The crystal structure was solved to a resolution of 1.36 Å and refined with 13,180 unique reflections to an R factor of 19.6% (and an R_{free} of 24.4%).

Coordinates and structure factors for the structure, including 101 water molecules, have been deposited at the Protein Data Bank with ID code 2B3E. The structure (Figs. 1 and 2) shows the ligand bound as expected in the AATT region of the minor groove, covering five base pairs. The inner-facing nitrogen atom of one imidazole ring of DB819 hydrogen bonds to O2 of a thymine (2.9 Å) and more weakly to N3 of an adenine (3.3 Å). The analogous imidazole nitrogen at the other end of the ligand does not contact DNA base edges directly, but hydrogen bonds to two water molecules (Figs. 2 and 3), that together with a third water, also make contacts with DNA base edges and O4′ sugar oxygen atoms. This tight network involves eight hydrogen bonds. The inner-facing nitrogen atom of the benzimidazole ring participates in a pair of hydrogen bonds to thymine O2 atoms, with N···O distances of 2.8 and 3.0 Å. Our previous structure of the amidino analogue DB818 bound to the four base pair AATT in the same DNA sequence showed both amidino groups participating in direct hydrogen bonding to DNA bases in the groove. The replacement of the amidinium groups by imidazoles in DB819 results in an extension of ligand dimension along its long axis, and it is no longer possible for both inner-facing imidazole nitrogen atoms to contact DNA directly; hence, the need for bridging waters at one end.

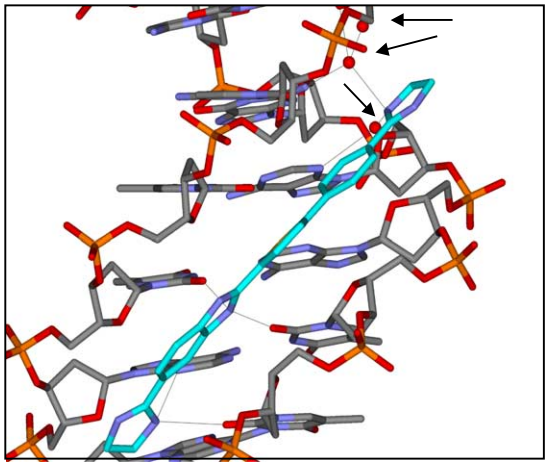


Figure 1. View of the structure of the DB819–d(CGCGAATTCGCG)$_2$ complex, with DB819 coloured cyan and the bound water molecules (as red spheres) indicated by arrows.

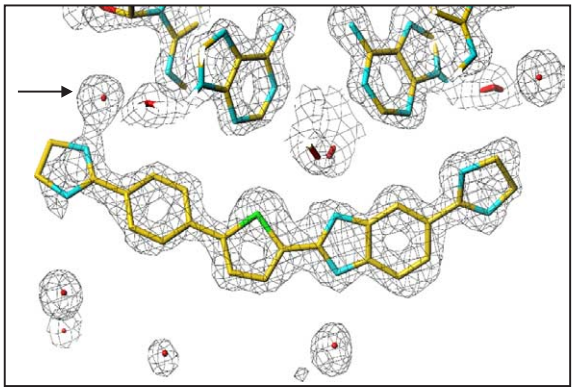


Figure 2. F_o electron density map, drawn at 1.2σ, and looking down onto the plane of the DB819 molecule in the minor groove. A bound water molecule is indicated by an arrow.

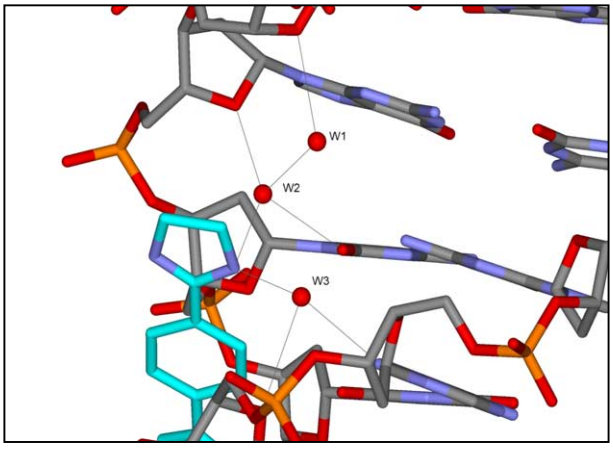


Figure 3. View of the three water molecules and their contacts with an imidazole group in DB819 and the DNA. Hydrogen bonds are indicated by thin black lines.

There are numerous programs designed to predict the bound state of ligands in a macromolecular receptor by in silico screening of compound libraries[9] but relatively few of them have been developed with drug–nucleic acid interactions in mind,[11,12] and applications to the prediction of DNA minor groove binding molecules are rare.[13] The DOCK program (version 5.1.1) was used in this study, as there have been recent attempts to optimise the scoring function of this program for DNA minor groove binders,[14] and there are previous examples of the use of DOCK with DNA.[15–17] DOCK uses a rigid-receptor, flexible-ligand approach.[18,19] In particular, we wished to see if in silico screening could predict the correct binding position for DB819 in view of the observation from the crystal structure that one end of the ligand does not contact the DNA directly, but instead is hydrogen bonded with a cluster of three water molecules.

The DNA coordinates from the crystal structure of berenil bound to the d(CGCGAATTCGCG)$_2$ duplex (PDB ID: 2DBE[20]) were used as the receptor for the docking runs. This complex was chosen because it was taken as a starting point in the structure solution of the DB818 crystal structure.[7] It was decided to use the structure of a previously determined analogous complex in the docking studies, rather than the DNA from the DB819 co-crystal structure, in order to more realistically mimic the process of structure prediction in the absence of crystallographic information. Partial atomic charges in the DNA were assigned according to the Amber94 all-atom force field,[21] although a united-atom model was applied in the construction of grids, whereby the charge on non-polar hydrogen atoms was subsumed into the adjacent carbon atom. The non-polar interaction grid was also defined using a united-atom model. DOCK requires that the binding site be defined by a set of spheres, and a sphere cluster was defined which included the entire minor groove. All other preparation steps were carried out using default settings.[22]

The two ligands were drawn using the Sybyl 7.0 modelling package[23] and minimized to an rms gradient of 0.05 kcal mol^{-1} Å^{-1} using the MMFF94 force field.[24–28] Two possible atom type assignments that are available in the SYBYL package, each with distinct charge distributions, were investigated for each ligand. Charge assignments were DB818 for each nitrogen atom in the amidinium groups, of −0.75442e (model A with Sybyl atom types N2, N3 for the amidinium nitrogen atoms) and −1.1924e (model B with Sybyl atom type Nar for the imidazole nitrogen atoms); and DB819 for each nitrogen atom attached to a nitrogen in the imidazole rings, of −0.7939e (model A with Sybyl N2, N3 for the amidinium nitrogen atoms) and −1.0129e (model B with Sybyl atom type Nar for the imidazole nitrogen atoms). We chose MMFF94 charges as these are provided in the 'mol2' format files downloadable from the current version of the ZINC database[29] and we are interested in extending the docking methodology to screen larger collections of molecules for their DNA binding ability. All-atom representations of the ligands were used.

The standard DOCK energy score, based on a molecular-mechanics estimation of the binding energy, was used as the primary and secondary scoring function in the docking runs. A generalized-Born scoring function has been incorporated into DOCK,[30] but this was found to be too slow to be used for adequate conformational searching over the multiple trial runs carried out and may be better suited for use as a filter on pre-docked positions of multiple ligands.[31] Default run parameters were used for the DOCK runs,[32] except for three parameters which were systematically varied to determine their effect: m_a, the minimum anchor size as a number of heavy atoms; n_o, the number of orientations of each anchor attempted; and n_c, the number of conformations per cycle of ligand growth. These parameters all relate to the anchor-first flexible-ligand algorithm.[32] Increasing n_o and n_c will increase the sampling carried out in each docking run, which should improve the accuracy of the method if the scoring function is adequate, but naturally also increases the run time. Decreasing m_a should also increase sampling, as more anchors are used. The accuracy of each docking run was assessed by calculating the root-mean-squared deviation (rmsd) of the non-hydrogen atom coordinates of the docked ligand compared with those of the ligand actually in the final refined crystal structure, after the non-hydrogen atoms of the two DNA duplexes had been placed in maximal alignment. This procedure was automated via a script written for the VMD program.[34] Only the best-scoring orientation from each docking run was considered.

The rmsd results are presented in Table 1. DOCK energy scores for DB819 with model A vary between −41 and −55 kcal mol^{-1}, and between −48 and −59 kcal mol^{-1} for model B. The most striking result is the difference between the rmsd values for ligand models A and B. In model A, neither DB818 nor DB819 could be docked to an rmsd to the crystal structures that was <2 Å in any of the runs. Moreover, there is no systematic improvement in the performance of the method with increased sampling, with a plateau of rmsd values of around 6 Å. This indicates that the scoring function is inadequate to accurately predict the binding position. By contrast, the results for model B show both ligands approaching a low rmsd value relatively systematically as the sampling is improved. This was correlated with the lowest values for the DOCK energy scores. For DB818 a value just above 2 Å in rmsd was achieved reliably, and for DB819, the ligand in this crystallographic study, DOCK predicted the binding mode to <0.7 Å. The series of structures with low rmsds in Table 1 has the ligand in a position very close to that observed crystallographically (Fig. 4), with the ligand–DNA hydrogen-bonding pattern being reproduced (apart from those involving water molecules). This indicates that a careful consideration of the method of charge and atom type assignment used in model B is necessary to give an accurate scoring function for these two ligands.

It is also noteworthy that a relatively extensive sampling is required to reliably locate the correct binding position, even with the correct scoring function. n_o = 100,000, $n_c \geqslant 100$ appear to be minimal requirements in this case, compared to the program defaults of n_o = 1000, n_c = 25. We note that this sampling required 550 s on a 2.4 GHz Intel CPU, which is not significant if one is only interested in a handful of ligands, but problematic if it is required to scale the methodology to screen very large databases of compounds, even with a computer cluster. The m_a variable, controlling the number and size of anchors, did not appear to significantly affect the accuracy of the docking results.

We conclude from these docking results that it was possible to use the DOCK 5.1.1 program[33] to predict the binding position of the DB819 ligands bound to the d(CGCGAATTCGCG)$_2$ duplex to an rmsd of <0.7 Å, without a knowledge of the crystal structure. For both DB818 and DB819, the increased charge polarization of the N–H groups and particular atom type assign-

Table 1.

m_a	n_o	n_c	Model A DB818 rmsd/Å	Model A DB819 rmsd/Å	Model B DB818 rmsd/Å	Model B DB819 rmsd/Å
4	1000	10	2.02	13.08	1.51	12.27
4	10000	10	10.42	13.05	4.33	4.78
4	100000	10	5.95	13.06	13.47	12.33
4	1000000	10	10.65	11.83	—	—
4	10000	100	5.83	13.08	1.78	12.31
4	100000	100	10.40	6.48	1.74	12.34
4	10000	1000	6.03	6.52	5.97	**0.63**
4	100000	1000	6.08	6.54	2.10	0.69
6	1000	10	5.40	6.71	1.53	12.27
6	10000	10	10.42	6.72	10.42	4.69
6	100000	10	5.59	12.29	13.47	12.33
6	1000000	10	10.63	11.81	—	—
6	10000	100	5.79	12.33	2.20	**0.72**
6	100000	100	6.05	2.17	2.16	**0.63**
6	10000	1000	5.97	6.48	2.04	**0.70**
6	100000	1000	6.00	11.93	2.16	0.64
10	1000	10	5.40	6.02	1.51	11.94
10	10000	10	10.42	6.36	4.35	5.72
10	100000	10	5.59	12.29	13.47	12.33
10	1000000	10	10.63	11.82	13.43	12.34
10	10000	100	5.99	11.91	2.02	12.32
10	100000	100	6.00	13.09	2.18	**0.63**
10	10000	1000	6.03	13.07	2.02	**0.63**
10	100000	1000	5.95	11.90	2.05	0.71

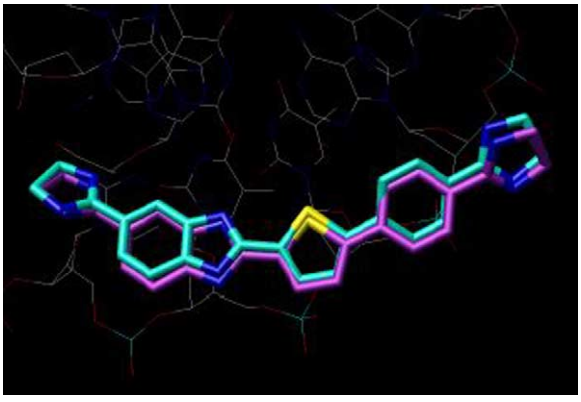


Figure 4. Overlay of the DNA in the crystal structure with that used in the docking calculations, showing the two positions of the DB819 molecule, in the complex (cyan) and in the docked position (purple). The rmsd between the two drug positions is 0.63 Å.

ments of the amidinium groups in the model B ligand set were critical for correct positional prediction in the groove, as was the extent of sampling undertaken, demonstrating the need for caution in applying default parameters when using DOCK and other in silico programs.

The crystal structure of the DB819–d(CGCGAATTCG CG)$_2$ complex shows one water molecule bridging between the DNA and an amidinium group at one end of the ligand. We speculate that the highly polar character of the amidinium groups in model B was able to compensate in some way for the absence of discrete water molecules in the docking model, even without use of knowledge-based potentials that incorporate structural information about water positions.[35] Preliminary tests with DOCK have also been carried out on seventeen minor groove binders with amidinium groups that are in the PDB as dodecanucleotide complexes. Positional prediction with charge model A was successful (to <2.5 Å rmsd) in 59% of cases, whereas charge model B was successful in 76% of cases, suggesting that the charges on amidinium group atoms are important generally, but not overriding factors for all compounds. It is encouraging that at least in the case of DB819, the DOCK program is able to correctly predict minor groove binding position even in the absence of the explicit solvent molecules that are present in the binding site of the crystalline complex.

Acknowledgments

This work was supported by Cancer Research UK (programme grant to S.N.). We are grateful to Professors D. W. Boykin, W. D. Wilson and Dr. A. Kumar (Georgia State University) for provision of material and much useful discussion.

References and notes

1. Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291.
2. Wilson, W. D.; Nguyen, B.; Tanious, F. A.; Mathis, A.; Hall, J. E.; Stephens, C. E.; Boykin, D. W. *Curr. Med. Chem. Anti-Cancer Agents* **2005**, *5*, 389.
3. Fairlamb, A. H. *Trends Parasitol.* **2003**, *19*, 488.
4. Ismail, M. A.; Brun, R.; Easterbrook, J. D.; Tanious, F. A.; Wilson, W. D.; Boykin, D. W. *J. Med. Chem.* **2003**, *46*, 4761.

5. Tidwell, R. R.; Boykin, D. W. In Demeunynck, M., Bailly, C., Wilson, W. D., Eds.; DNA and RNA Binders From Small Molecules to Drugs; Wiley-VCH, 2003; Vol. 2, pp 414–460, Chapter 16.
6. Ismael, M. A.; Brun, R.; Wenzler, T.; Tanious, F. A.; Wilson, W. D.; Boykin, D. W. *Bioorg. Med. Chem.* **2004**, *12*, 5405.
7. Mallena, S.; Lee, M. P.; Bailly, C.; Neidle, S.; Kumar, A.; Boykin, D. W.; Wilson, W. D. *J. Am. Chem. Soc.* **2004**, *126*, 13659.
8. Wang, L.; Bailly, C.; Kumar, A.; Ding, D.; Bajic, M.; Boykin, D. W.; Wilson, W. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12.
9. Shoichet, B.; McGovern, S.; Wei, B.; Irwin, J. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439.
10. The crystal used for data collection was obtained from a 17 μl drop containing 11.765 mM $MgCl_2$, 0.176 mM double-stranded DNA, 0.235 mM compound DB819, 2.941% v/v (±)-2-methyl-2,4-pentanediol (MPD) and 7.059 mM sodium cacodylate buffer at pH 7.0. The DNA:ligand ratio was 3:4. The drop was equilibrated against a reservoir of 0.5 ml of 50% v/v MPD solution.
11. Morley, S.; Afshar, M. *J. Comput. Aid. Mol. Des.* **2004**, *18*, 189.
12. Brooijmans, N.; Kuntz, I. D. *Ann. Rev. Biophys.* **2003**, *32*, 335.
13. Pang, Y. *Bioorg. Med. Chem.* **2004**, *12*, 3063.
14. Kang, X.; Shafer, R.; Kuntz, I. *Biopolymers* **2004**, *73*, 192.
15. Grootenhuis, P.; Kollman, P.; Seibel, G.; DesJarlais, R.; Kuntz, I. *Anticancer Drug Des.* **1990**, *5*, 237.
16. Grootenhuis, P.; Roe, D.; Kollman, P.; Kuntz, I. *J. Comput. Aid. Mol. Des.* **1994**, *8*, 731.
17. Watts, C.; Kerwin, S.; Kenyon, G.; Kuntz, I.; Kallick, D. *J. Am. Chem. Soc.* **1995**, *117*, 9941.
18. Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. *J. Mol. Biol.* **1982**, *161*, 269.
19. Ewing, T.; Kuntz, I. *J. Comput. Chem.* **1997**, *18*, 1175.
20. Brown, D. G.; Sanderson, M. R.; Skelly, J. V.; Jenkins, T. C.; Brown, T.; Garman, E.; Stuart, D. E.; Neidle, S. *EMBO J.* **1990**, *9*, 1329.
21. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. W., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
22. *Dock 4.0 manual*, **1998**, URL http://dock.compbio.ucsf.edu.
23. *Tripos website*, **2005**, URL http://www.tripos.com.
24. Halgren, T. *J. Comput. Chem.* **1996**, *17*, 490.
25. Halgren, T. *J. Comput. Chem.* **1996**, *17*, 520.
26. Halgren, T. *J. Comput. Chem.* **1996**, *17*, 553.
27. Halgren, T.; Nachbar, R. *J. Comput. Chem.* **1996**, *17*, 587.
28. Halgren, T. *J. Comput. Chem.* **1996**, *17*, 616.
29. Irwin, J.; Shoichet, B. *J. Chem. Inf. Model.* **2005**, *45*, 177.
30. Zuo, X.; Sun, X.; Kuntz, I. *J. Am. Chem. Soc.* **1999**, *121*, 8033.
31. Liu, H.; Kuntz, I.; Zou, X. *J. Phys. Chem. B* **2004**, *108*, 5453.
32. Leach, A.; Kuntz, I. *J. Comput. Chem.* **1992**, *13*, 730.
33. *Dock 5.1.1 manual*, **2003**, URL http://dock.compbio.ucsf.edu.
34. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.
35. Ge, W.; Schneider, B.; Olson, W. K. *Biophys. J.* **2005**, *88*, 1166.