SUPPLEMENTARY MATERIALS FOR

Strong Binding in the DNA Minor Groove by an Aromatic Diamidine With a Shape That Does Not Match the Curvature of the Groove

Binh Nguyen,[†] Michael P. H. Lee,[‡] Donald Hamelberg,[†] Alexandra Joubert,[§] Christian Bailly,[§] Reto Brun,^{||} Stephen Neidle,[‡] and W. David Wilson^{†,*}

[†]Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA, [‡]The Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, London WC1N 1AX, UK, [§]INSERM U-524 et Laboratoire de Pharmacologie Antitumorale du Centre Oscar Lambret, IRCL, Place de Verdun, 59045 Lille, France, ^{||}Swiss Tropical Institute, CH-4002 Basel, Switzerland

DNase I Footprinting

DNase I footprinting was conducted on 117 bp and 265 bp 3'-[³²P]-end labeled fragments from the *EcoRI-PvuII* double digest of the plasmid pBS essentially as previously described with a number of minor groove binding agents.¹⁻³ DNase I digestion was done on the free DNAs and the DNAs complexed with CGP 40215A, berenil, and pentamidine. The labeled digestion products were separated on an 8% polyacrylamide gel under denaturing conditions in TBE buffer (89 mM Tris-borate pH 8.3, 1 mM EDTA). An autoradiogram of the gel from an experiment conducted in this manner is shown in Figure S1. Very similar digestion patterns are produced by the three compounds, all at AT sequences. The weakest footprints are obtained with pentamidine and the strongest footprints are generally obtained with CGP.

Surface Plasmon Resonance (SPR) – Biosensor Binding Studies

The AATT hairpin (Figure 1) was captured on a streptavidine derivatized sensor chip as previously described.³ CGP 40215A and berenil at concentrations from 1 to approximately 700 nM in MES 10 buffer (0.010 M [2-(N-morpholino) ethanesulfonic acid] (MES), 0.001 M EDTA, 0.1 M NaCl, pH 6.25, with 5x10⁻³ % v/v Surfactant P20) were passed over the DNA surface to generate sets of sensorgrams such as those shown in Figure 2. The RU values were converted to *r* (RU/RU_{sat} = moles of compound/mole of DNA) as previously described,⁴ and the results are shown in Figure S2. The data were fitted with a 1:1 binding model:

$$r = (K \cdot C_{\text{free}}) / (1 + K \cdot C_{\text{free}})$$

The binding constant of CGP 40215A is 7.0 x 10⁷ M⁻¹ and that for berenil is 1.1 x 10⁷ M⁻¹.

Crystallography

Materials: The oligonucleotide d(CGCGAATTCGCG) was purchased in HPLC-purified form from Oswel Research Products Ltd. (Southampton, UK). The oligonucleotide was annealed at 358K for 5 minutes in 40mM sodium cacodylate buffer at pH 6.5 before use.

Crystallization: Crystals were grown by vapor diffusion at 285K from hanging drops. The crystal used for data collection was obtained from a drop containing 2 μ l of 75mM MgCl₂, 3 μ l of 2mM double-stranded DNA, 2 μ l of 25% v/v 2-methylpentane-2,4-diol (MPD), 2 μ l of 4.5mM CGP 40215A compound, in 40mM sodium cacodylate buffer at pH 6.5, equilibrated against a reservoir of 700 μ l of 50% v/v MPD solution. Crystals grew as large rods within three weeks. Typical crystals were approximately 0.7 x 0.2 x 0.2 mm in dimension.

Crystallographic Data Collection, Structure Solution and Refinement: A crystal was mounted in a loop and flash-frozen in a nitrogen gas stream at 104K and maintained at that temperature using an Oxford Cryostream (Oxford Cryosystems). Diffraction data were collected on a Rigaku RAXIS IV image plate detector with CuK α radiation from a Rigaku RU200 rotating anode generator (100mA, 50kV) and a Yale focussing mirror system. The crystal-to-detector distance was set at 120mm, with an oscillation angle of 2.5° for the crystal. One hundred and three frames, with exposure time of 1800s for each frame, were collected for the structure, from 0° to 255°, which resulted in a maximum resolution of 1.78Å. Indexing and data processing were carried out using the program package DENZO and SCALEPACK.⁵ The structure was solved by molecular replacement, and refined using the X-PLOR, CNS and SHELX-97 programs.^{6,7} At the top of Figure 3, a view into the minor groove is shown. At the bottom of Figure 3, a view of a 2F_o-F_c electron density map (at 1.78Å resolution) shows the structure of the CGP 40215A – DNA complex looking onto the plane of the ligand. Contours are drawn at a 1.3s level. The specific interactions between the –NH groups on the linker and one of the amidines with the DNA bases are shown in Figure S3 and S4, respectively. A summary of data collection and refinement statistics and resolution ranges for the structure is given in Table 1. Coordinates for the structure have been submitted to the Protein Data Bank: PDB 1M6F.

Molecular Dynamics

In order not to bias the MD simulations in the direction of the crystal structure, CGP 40215A was docked into the DNA that was extracted from a pentamidine crystal structure.⁸ The compound was placed in the minor groove in several positions around the AATT sequence and at about 3 Angstroms from the floor of the groove with the two -NH groups of the linker pointing towards the minor groove and the amino group pointing out. No specific hydrogen bonds were initiated. After equilibration, the simulations were conducted with the AMBER software package⁹ for 10 nanoseconds at a constant temperature of 300 K, a constant pressure of 1 bar and conditions similar to previous MD simulations.^{10,11}

Within the first 500 picoseconds of the simulation, the ligand slides deeply into the groove and forms specific hydrogen bonds with DNA bases. The structure obtained from the unconstrained MD simulations was very similar to the x-ray structure, and an overlay plot of MD and x-ray structures is shown in Figure S5. The amidine group at one end of the molecule interacts with the O2 of T8, while the -NH group on the linker that is closest to the amidine interacts with the O2 of T19 on the other strand. At the same time, the amidine group on the other end of the molecule interacts nonspecifically with the backbone of the duplex and the bases

via water bridges. The reverse complex in which the amidine on the other end of the molecule interacts with the O2 of T7, and the -NH that is closest to it interacts with the O2 of T20 also occurs and is quite favorable. The interactions can be viewed as linked through a "seesaw" motion with linker interactions during the transition. Figure S6 shows an overlay of the two MD structures after a 180-degree end-to-end rotation of one structure. A second simulation was carried out starting with the canonical B-form DNA and quite similar types of interactions were observed. The complex structure thus does not appear to be very sensitive to the initial specific B-form structure of DNA used the docking.

Additional evaluation of the proposed binding sites was done with another MD simulation, also carried out using the canonical B-form duplex d(CGCGAATTTCGCG)₂, but at 400 K to increase the rate of dynamic movements of the ligand. The compound was initially placed in the minor groove over the GA step rather than at the central AT step. No specific hydrogen bonds were formed and CGP 40215A was initially placed about 3 Angstroms away from the floor of the minor groove. At around 300 picoseconds into the simulation, the compound moved to the central AATT sequence of the DNA and formed strong interactions with the minor groove very similar to those seen in the 300 K simulation. The "seesaw" transition from one type of complex to the other was observed multiple times during the simulations.

References

- (1) Waring, M. J.; Bailly, C. *J. Mol. Recognit.* **1997**, *10*, 121-127.
- (2) Lavesa, M.; Fox, K. R. *Anal. Biochem.* **2001**, *293*, 246-250.
- (3) Nguyen, B.; Tardy, C.; Bailly, C.; Colson, P.; Houssier, C.; Kumar, A.; Boykin, D. W.; Wilson, W. D. *Biopolymers* **2002**, *63*, 281-297.
- (4) Davis, T. M.; Wilson, W. D. *Anal. Biochem.* **2000**, *284*, 348-353.
- (5) Otwinowski, Z. M.; Minor, W. In *Methods in Enzymology, Macromolecular Crystallography, Part A*; Charles W. Carter, J., Sweet, R. M., Eds.; Academic Press: New York, NY, 1997; Vol. 276, pp 307-326.
- (6) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr D Biol Crystallogr* **1998**, *54*, 905-921.
- (7) Sheldrick, G. M.; Schneider, T. R. In *Methods in Enzymology, Macromolecular Crystallography, Part B*; Charles W. Carter, J., Sweet, R. M., Eds.; Academic Press: New York, NY, 1997; Vol. 277, pp 319-343.
- (8) Edwards, K. J.; Jenkins, T. C.; Neidle, S. *Biochemistry* **1992**, *31*, 7104-7109.
- (9) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER5*; University of California: San Francisco, 1997.
- (10) Hamelberg, D.; McFail-Isom, L.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2000**, *122*, 10513-10520.
- (11) Hamelberg, D.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2001**, *123*, 7745-7755.

Figure Captions

Figure S1. DNase I footprinting of CGP 40215A, berenil and pentamidine on the *Eco*RI-*Pvu*II 117-bp (A) and 265-bp (B) DNA restriction fragments from plamid pBS. The DNA was 3'-end labeled at the *Eco*RI with [α - 32 P]dATP in the presence of AMV reverse transcriptase. The products of nuclease digestion were resolved on 8% polyacrylamide gels containing 8M urea. The concentration (μ M) of the drug is shown at the top of the appropriate gel lanes. Control tracks (Ct) contained no drug. The tracks labeled "G" represent dimethylsulphate-piperidine markers specific for guanines. Numbers on the left side of the gels refer to the standard numbering scheme for the nucleotide sequence of the DNA fragments. CGP 40215A, berenil and pentamidine all have similar footprinting patterns but the footprints of CGP 40215A and berenil are stronger than those of pentamidine. The DNA sequences in the footprint regions are shown on the left (A) or on the right (B) of the gel.

Figure S2. A plot of SPR binding results of CGP 40215A (squares) and berenil (circles) with the AATT hairpin is shown. Both compounds bind strongly to AT rich sequences in agreement with footprinting results. The fits to the data with a 1:1 binding model are shown as solid lines.

Figure S3. H-bond patterns between the two –NH groups on the linker of CGP 40215A and DNA bases are shown as solid lines. The direct H-bond interactions are between the two –NH groups and the cross-strand O2 groups of T7 and T19 as well as with O2 of T20. DNA is shown in stick representation and the compound in ball and stick.

Figure S4. H-bond patterns between an amidine of CGP 40215A and DNA bases are shown. The direct H-bond interaction is between the amidine and the O2 of T8. The water-mediated interaction is with the complementary base of T8 (A17) as well as to C9. The water molecule is indicated by a sphere. DNA and CGP 40215A are as in Figure S3.

Figure S5. An overlay of a low energy MD structure and the x-ray structure is shown. A consistent agreement between the two structures in terms of the ligand position and the groove-width can be observed. The MD structure is shown in green and the x-ray structure in grey. The RMSD values (considered only heavy atoms) are 0.95Å for the six central bases pairs (GAATTC) with the ligand and 2.63Å for the entire complex.

Figure S6. An overlay plot of two MD structures with opposite amidine groups forming direct H-bonds with DNA bases is shown. The structures were overlaid after one was rotated by 180 degrees end-to-end. The resemblance between the two complexes and the x-ray structure are observed in these views. The RMSD values are 0.85Å for the six central bases pairs (GAATTC) with the ligand and 1.58Å for the entire complex.

Figure S1

Figure S2

Figure S3

Figure S4

Figure S5

Figure S6

Table 1. Data collection and refinement statistics

	CGP Complex
DNA sequence d(CGCGAATTCGCG) ₂	
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	<i>a</i> =24.098Å, <i>b</i> =39.821Å, <i>c</i> =65.620Å, $\alpha=\beta=\gamma=90^\circ$
Number of measured reflections > 1 σ	53082
Number of unique reflections	6040
Completeness (%)	92.8
in the outer shell (%)	90.6 (1.84- 1.78 Å)
<i>R</i> _{merge} (%)	4.7
Resolution range of refinement	10.0-1.78
Number of reflections used in refinement	5636
Reflections with <i>F</i> _o >4 σ (<i>F</i> _o)	5477
Number of parameters refined	2535
Restraints	3761
Maximum resolution of refinement (Å)	1.78
Final <i>R</i> (%)	20.15
Final <i>R</i> for reflections with <i>F</i> _o >4 σ (<i>F</i> _o) (%)	19.88
Final <i>R</i> _{free} (%)	27.13
Number of DNA atoms	486
Number of drug atoms	26
Number of full occupancy waters	111
Number of magnesium ions	1