J|A|C|S
ARTICLES
Published on Web 01/23/2003

Molecular Dynamics Simulations and Thermodynamics Analysis of DNA−Drug Complexes. Minor Groove Binding between 4′,6-Diamidino-2-phenylindole and DNA Duplexes in Solution

Nad'a Špačková,[†] Thomas E. Cheatham, III,*[,‡] Filip Ryjáček,[§] Filip Lankaš,[§] Luc van Meervelt,[||] Pavel Hobza,[§] and Jiří Šponer*[,†,§]

Contribution from the Institute of Biophysics, Academy of Sciences of the Czech Republic, and National Center for Biomolecular Research, Královopolská 135, 612 65 Brno, Czech Republic, Departments of Medicinal Chemistry and of Pharmaceutics and Pharmaceutical Chemistry, University of Utah, 2000 East 30 South Skaggs Hall 201, Salt Lake City, Utah 84112, J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, and National Center for Complex Molecular Systems and Biomolecules, Dolejškova 3, 182 23 Prague, Czech Republic, and Department of Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200F, B-3001, Heverlee, Belgium

Received January 21, 2002; E-mail: tec3@utah.edu; sponer@ibp.cz

Abstract: An extended set of nanosecond-scale molecular dynamics simulations of DNA duplex sequences in explicit solvent interacting with the minor groove binding drug 4′,6-diamidino-2-phenylindole (DAPI) are investigated for four different and sequence specific binding modes. Force fields for DAPI have been parametrized to properly reflect its internal nonplanarity. Sequences investigated include the binding modes observed experimentally, that is, **AATT** in d(CGCGAATTCGCG)$_2$ and **ATTG** in d(GGCCAATTGG)$_2$ and alternative shifted binding modes **ATTC** and **AATT**, respectively. In each case, stable MD simulations are obtained, well reproducing specific hydration patterns seen in the experiments. In contrast to the 2.4 Å d(CGCGAATTCGCG)$_2$ crystal structure, the DAPI is nonplanar, consistent with its gas-phase geometry and the higher resolution crystal structure. The simulations also suggest that the DAPI molecule is able to adopt different conformational substates accompanied by specific hydration patterns that include long-residing waters. The MM_PBSA technology for estimating relative free energies was utilized. The most consistent free energy results were obtained with an approach that uses a single trajectory of the DNA−DAPI complex to estimate all free energy terms. It is demonstrated that explicit inclusion of a subset of bound water molecules shifts the calculated relative binding free energies in favor of both crystallographically observed binding modes, underlining the importance of structured hydration.

Introduction

Molecular complexes of nucleic acids interacting with small drugs in the minor groove represent an important class of biochemical, pharmacological, and biological systems.[1−5] In this study, we present an extended set of nanosecond-scale explicit solvent molecular dynamics (MD) simulations of multiple DNA duplex sequences interacting with the minor groove binding drug 4′,6-diamidino-2-phenylindole (DAPI) in solution. We investigate four different, and sequence specific, minor groove binding modes in a series of MD simulations complemented by thermodynamic analysis of the binding patterns. DAPI (Figure 1) is a common fluorescent marker for DNA[6] that has numerous applications,[7] including use as an antiparasitic,[8] antibiotic, antiviral, and anti-cancer drug. Its presumed mode of action is by interfering with the activity of some DNA processing enzymes involved in regulatory and structural functions, such as the blocking of RNA polymerase II by inhibition of TATA box binding protein binding to DNA.[9,10]

There are two atomic resolution crystal structures of DAPI bound to DNA available,[11,12] supplemented by numerous NMR,[13−17] spectroscopic,[18−25] and DNA footprinting studies.[26−28]

[†] Institute of Biophysics.
[‡] University of Utah.
[§] Institute of Physical Chemistry.
[||] Katholieke Universiteit Leuven.

(1) Bischoff, G.; Hoffmann, S. *Curr. Med. Chem.* **2002**, *9*, 312−348.
(2) Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291−309.
(3) Han, X.; Gao, X. *Curr. Med. Chem.* **2001**, *8*, 551−581.
(4) Reddy, B. S. P.; Sharma, S. K.; Lown, J. W. *Curr. Med. Chem.* **2001**, *8*, 475−508.
(5) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215−2235.
(6) Kapuscinski, J.; Szer, W. *Nucleic Acids Res.* **1979**, *6*, 3519−3534.
(7) Zimmer, C.; Wahnert, U. *Prog. Biophys. Mol. Biol.* **1986**, *47*, 31−112.
(8) Dann, O.; Bergen, G.; Demant, E.; Volz, G. *Justus Liebigs Ann. Chem.* **1971**, *749*, 68−89.
(9) Chiang, S. Y.; Welch, J.; Rauscher, F. J. I.; Beerman, T. A. *Biochemistry* **1994**, *33*, 7033−7040.
(10) Welch, J. J.; Rauscher, F. J., III; Beerman, T. A. *J. Biol. Chem.* **1994**, *269*, 31051−31058.
(11) Larsen, T. A.; Goodsell, D. S.; Cascio, D.; Grzeskowiak, K.; Dickerson, R. E. *J. Biomol. Struct. Dyn.* **1989**, *7*, 477−491.
(12) Vlieghe, D.; Šponer, J.; van Meervelt, L. *Biochemistry* **1999**, *38*, 16443−16451.

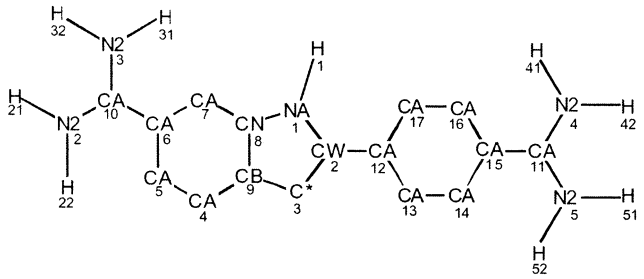

Figure 1. Molecular structure, atom numbering (subscript), and atom types for 4′,6-diamidino-2-phenylindole (DAPI).

The standard binding mode in AT-rich sequences is in the minor groove. As seen in the 2.4 Å resolution crystal structure obtained over a decade ago, DAPI binds directly to the bases in the **AATT** region of the minor groove of the dodecamer sequence d(CGCG**AATT**CGCG)$_2$.[11] This interaction is stabilized by the snug fit of the aromatic rings of DAPI between the walls of the narrow minor groove, the favorable negative electrostatic potential, and the absence of minor groove amino groups in the AT sequences.[29,30] Earlier footprinting studies suggested that, in contrast to Hoechst 33258, DAPI is less accommodating of guanosine residues at the end of its minor groove binding site[26] and instead tends to intercalate or interact with the major groove in GC sequences.[14,15,19,24,31−33] Despite this, a recent 1.9 Å crystal structure shows DAPI shifted to bind into the **ATTG** region of d(GGCCA**ATTG**G).[12] This structure has one of the terminal DAPI amidinium groups tightly bound to guanine in the minor groove. It leads to a close contact with the guanine amino group that is likely facilitated by a substantial sp^3 pyramidalization of the guanine amino group and specific hydration.[12]

A contemporary explicit solvent MD simulation can reliably reproduce varied DNA structure, dynamics, and interactions (for review, see refs 34−36). Despite this, few studies have been published that apply modern MD simulation methods to investigate DNA minor groove binding ligands. These MD simulations have shown a rather accurate representation of drug interaction with the minor groove of DNA, including the observation of netropsin spontaneously shifting from one binding mode to the expected binding mode[37] and freezing of the distribution of DNA backbone conformations (B_I/B_{II} substate populations).[38,39] There has also been an MD evaluation of the free energies of netropsin binding to DNA using the free energy perturbation (FEP) approach.[40] Despite using cutoffs and rather short simulations, the calculated results agreed surprisingly well with experimental estimates. Nevertheless, the prevailing opinion is that such large perturbations perhaps push beyond the accepted range of applicability of FEP and therefore the agreement might be considered as rather fortuitous.

This work presents simulations supplemented with evaluation of the thermodynamics of binding using the technique known as MM_PBSA[41,42] to investigate DAPI binding in the minor groove of DNA. This analysis is equivalent to the ES/IS method of Vorobjev and Hermans,[43,44] similar in spirit to the linear interaction energy approaches,[45,46] and a direct extension of well-known methods for estimating free energies[47−49] that involve averaging over configurations sampled in an MD simulation. The MM_PBSA method has proven useful in the study of a variety of nucleic acid systems.[41,42,50−54] The current work demonstrates that state-of-the-art molecular dynamics simulations are able to complement experimental studies of DNA−drug binding in many aspects, including accurate representation of the structure and hydration. Nonetheless, there are also limitations of the MD technique that in our extended study are considerably more apparent as compared to the preceding reports. Limitations relate to incomplete sampling and inaccuracies of the potential and free energy estimations. We demonstrate that the applied MM_PBSA technology—a simple but fairly approximate method based on a continuum representation of the solvent, simple parametrizations of hydrophobic hydration, and approximate estimates of the solute entropy—in

(13) Trotta, E.; D'Ambrosio, E.; Del Grosso, N.; Ravagnan, G.; Cirilli, M.; Paci, M. *J. Biol. Chem.* **1993**, *268*, 3944−3955.
(14) Trotta, E.; D'Ambrosio, E.; Ravagnan, G.; Paci, M. *Nucleic Acids Res.* **1995**, *23*, 1333−1340.
(15) Trotta, E.; D'Ambrosio, E.; Ravagnan, G.; Paci, M. *J. Biol. Chem.* **1996**, *271*, 27608−27614.
(16) Trotta, E.; Paci, M. *Nucleic Acids Res.* **1998**, *26*, 4706−4713.
(17) Trotta, E.; Del Grosso, N.; Erba, M.; Paci, M. *Biochemistry* **2000**, *39*, 6799−6808.
(18) Manzini, G.; Barcellona, M. L.; Avitabile, M.; Quadrifoglio, F. *Nucleic Acids Res.* **1983**, *11*, 8861−8876.
(19) Barcellona, M. L.; Favilla, R.; von Berger, J.; Avitabile, M.; Ragusa, N.; Masotti, L. *Arch. Biochem. Biophys.* **1986**, *250*, 48−53.
(20) Kubista, M.; Akerman, B.; Norden, B. *Biochemistry* **1987**, *26*, 4545−4553.
(21) Tanious, F. A.; Veal, J. M.; Buczak, H.; Ratmeyer, L. S.; Wilson, W. D. *Biochemistry* **1992**, *31*, 3103−3112.
(22) Eriksson, S.; Kim, S. K.; Kubista, M.; Norden, B. *Biochemistry* **1993**, *32*, 2987−2998.
(23) Jansen, K.; Norden, B.; Kubista, M. *J. Am. Chem. Soc.* **1993**, *115*, 10527−10530.
(24) Norden, B.; Kurucsev, T. *J. Mol. Recognit.* **1994**, *7*, 141−155.
(25) Breusegem, S. Y.; Clegg, R. M.; Loontiens, F. G. *J. Mol. Biol.* **2002**, *315*, 1049−1061.
(26) Portugal, J.; Waring, M. J. *Biochim. Biophys. Acta* **1988**, *949*, 158−168.
(27) Wilson, W. D.; Tanious, F. A.; Barton, H. K.; Jones, R. L.; Fox, K.; Wydra, R. L.; Strekowski, L. *Biochemistry* **1990**, *29*, 8452−8461.
(28) Albert, F. G.; Eckdahl, T. T.; Fitzgerald, D. J.; Anderson, J. N. *Biochemistry* **1999**, *38*, 10135−10146.
(29) Loontiens, F. G.; McLaughlin, L. W.; Diekmann, S.; Clegg, R. M. *Biochemistry* **1991**, *30*, 182−189.
(30) Waring, M. J.; Bailly, C. *J. Mol. Recognit.* **1997**, *10*, 121−127.
(31) Wilson, W. D.; Tanious, F. A.; Barton, H. J.; Strekowski, L.; Boykin, D. W. *J. Am. Chem. Soc.* **1989**, *111*, 5008−5010.
(32) Wilson, W. D.; Tanious, F. A.; Barton, H. K.; Jones, R. L.; Boykin, D. W.; Strekowski, L. *Anti-Cancer Drug Des.* **1990**, *5*, 31−42.
(33) Kim, S. K.; Eriksson, S.; Norden, B. *Biopolymers* **1993**, *33*, 1677−1686.
(34) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182−196.
(35) Cheatham, T. E., III; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−471.
(36) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350−357.
(37) Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Am. Chem. Soc.* **2000**, *122*, 3927−3931.
(38) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Am. Chem. Soc.* **2001**, *123*, 5044−5049.
(39) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Phys. Chem. B* **2001**, *105*, 3135−3142.
(40) Singh, S. B.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 3267−3271.
(41) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(42) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(43) Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins* **1998**, *32*, 399−413.
(44) Vorobjev, Y. N.; Hermans, J. *Biophys. Chem.* **1999**, *78*, 195−205.
(45) Marelius, J.; Hansson, T.; Aqvist, J. *Int. J. Quantum Chem.* **1998**, *69*, 77−88.
(46) Hansson, T.; Marelius, J.; Aqvist, J. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(47) Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293−1301.
(48) Misra, V. K.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 245−263.
(49) Misra, V. K.; Honig, B. *Biochemistry* **1996**, *35*, 1115−1124.
(50) Srinivasan, J.; Miller, J. L.; Kollman, P. A.; Case, D. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 671−682.
(51) Cheatham, T. E., III; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265−280.
(52) Tsui, V.; Case, D. A. *J. Phys. Chem. B* **2001**, *105*, 11314−11325.
(53) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658−12663.
(54) Wu, M.; Yan, S.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *Nucleic Acids Res.* **2002**, *30*, 3422−3432.

Table 1. List of Simulations Carried Out on DNA−DAPI Complexes[a]

simulation	DNA sequence	length of simulation (ns)
M_ATTG	(CCA**ATTG**G)$_2$*GG*	7.5
M_ATTG2	(CCA**ATTG**G)$_2$*GG*	5.0
D_AATT	(CGCG**AATT**CGCG)$_2$	5.8
M_AATT	(CC**AATT**GG)$_2$*GG*	5.0
M_AATT-C	(CC**AATT**CG)*GG*	5.0
M_ATTC	(CCA**ATTC**G)*GG*	5.0
D_ATTC	(CGCG**AATTC**GCG)$_2$	5.0
M_ATTG(r)	(CCA**ATTG**G)$_2$*GG*	3.4
M_AATT2	(CCA**AATT**GG)$_2$*GG*	2.5
M_AATT(r)	(CCA**AATT**GG)$_2$*GG*	5.0
M_ATTG(pl)	(CCA**ATTG**G)$_2$*GG*	4.0

[a] M_ denotes the Van Meervelt sequence, and D_ denotes Dickerson's sequence. M_ATTG/M_ATTG2 and M_AATT/M_AATT2 are each separate, but equivalent, control simulations. For the Van Meervelt sequence, the italic *GG* represents the overhanging bases (see below). In column 1, (r) denotes the DAPI force field with a reduced nonplanarity, and (pl) denotes the planar DAPI force field (described in more detail in the Supporting Information). Other simulations utilized the "s" force field. The DAPI binding site is bold and underlined.

its current parametrization is not able to correctly estimate *absolute* free energies of association. Despite this, the *relative* free energies appear reliable,[42] consistent with a recent study investigating the 1:1 and 2:1 complex of Hoechst 33258 to DNA that demonstrates that the configurational entropy considerably favors the cooperativity of drug binding.[53] Here we show that a marked improvement of the results can be obtained by limited inclusion of explicit solvent in the free energy analysis.

Methods

Parametrization of a DAPI Force Field. The standard method of extending the Cornell et al.[55] force field involves picking atom types for the new molecule by analogy to the Cornell et al. atom types. With this simple approach, the isolated DAPI molecule is completely planar (designated in the following as the "pl" force field), in sharp disagreement with quantum-chemical calculations and the 1.9 Å resolution crystal structure.[12] Thus, we prepared a more sophisticated force field parametrized extensively using the quantum-chemical results. Atom types and van der Waals parameters were chosen by analogy to the original Cornell et al. parametrization (Figure 1). Atomic charges (Supporting Information) were derived for the optimized quantum-chemical geometry with the RESP[56,57] methodology at the HF/6-31G* level with two-stage RESP fitting for the terminal $-C(NH_2)_2$ groups.

The bonded parameters were parametrized by a fit to the ab initio data.[58] The HF/6-31G* optimizations reveal the intersegment dihedral angles of 40.3°, 35.7°, and 42.0° for $N_2-C_{10}-C_6-C_5$, $C_3-C_2-C_{12}-C_{13}$, and $C_{14}-C_{15}-C_{11}-N_5$, respectively (Figure 1). The final force field very closely reproduces the HF/6-31G* structure of isolated DAPI, including the energetics and profiles of the intersegment torsional barriers. This standard force field (marked as "s") and its parametrization are described more fully in the Supporting Information.

All simulations described here (Table 1) and in the Supporting Information were carried out using the AMBER 5 and 6 program packages[59] with the Cornell et al. (PARM94) force field.[55] We have also tested the more recent PARM98/99 force field.[60] However, for reasons outlined in the Supporting Information, herein we discuss only the results using the original PARM94 parametrization with the "s" DAPI force field.

All solutes were surrounded by a periodic rectangular box of water molecules described by TIP3P potential[61] extended to a distance of 10 Å from any solute atom. The system was neutralized by sodium cations[62] initially placed at the most electronegative sites around the molecule using the LEAP module of AMBER 5. All energy minimizations and molecular dynamics simulations were performed using the sander module with the particle mesh Ewald method[63] for the correct treatment of electrostatic interactions.[64] We applied an equilibration protocol that has been described in detail in our preceding studies.[65−68] A summary of the relevant runtime parameters includes the use of ∼1.0 Å grid spacing and fourth-order B-spline interpolation for the reciprocal part of the Ewald sum, an Ewald coefficient of ∼0.32, a cutoff of 9 Å for the Ewald direct space and van der Waals interaction with a 10-step pairlist update (except for the pl and r simulations and also M_AATT2 which used a heuristic pairlist update and a 1 Å buffer on the pairlist), a 2 fs time step, and application of the SHAKE procedure[69] to constrain all bonds involving hydrogen to their correct values. Additional control simulations of the M_AATT, M_ATTG, M_AATT2, and M_ATTG2 were run for 10 ns in a NVE ensemble (starting at 300 K) with a heuristic pairlist update (and a 2 Å buffer); very similar structural and free energetic results were obtained. During the equilibration period, the solvent environment relaxed to reasonable density (∼1.0 g/mL). Additionally, the ion atmosphere relaxed from its initial placement. On the basis of our experience and the experience of others[70] in simulations of DNA duplexes, the structure and dynamics are not significantly biased by the initial ion placement assuming reasonable equilibration. The equilibration was followed by several nanoseconds of production simulation at 300 K with constant temperature and pressure coupling (1.0 ps^{-1} coupling time).[71] Coordinates were written to trajectory files after each picosecond, and the center of mass translational motion was periodically removed.[72,73] Although a rectangular box was used and center of mass rotational motion of the DNA was not inhibited, rotation of the DNA in the periodic unit cell did not have any significant effect on the structure or dynamics. Trajectories were analyzed using the carnal and ptraj modules of AMBER 5.[59] Hydration and distribution of cations were analyzed by a rigorous evaluation of the average atomic densities.[74] In addition, a systematic monitoring of the solute−solvent distances was carried out, calculating minimal distances between selected atoms and all solvent molecules.

Thermodynamic Calculations. Free energy analysis was performed using the MM_PBSA scripts supplied with AMBER 6.0.[41,42] The free

(55) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(56) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(57) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(58) Hopfinger, A. J.; Pearlstein, R. A. *J. Comput. Chem.* **1984**, *5*, 486−499.
(59) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(60) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(62) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(63) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(64) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155−179.
(65) Špačková, N.; Berger, I.; Šponer, J. *J. Am. Chem. Soc.* **2001**, *123*, 3295−3307.
(66) Špačková, N.; Berger, I.; Šponer, J. *J. Am. Chem. Soc.* **2000**, *122*, 7564−7572.
(67) Špačková, N.; Berger, I.; Šponer, J. *J. Am. Chem. Soc.* **1999**, *121*, 5519−5534.
(68) Špačková, N.; Berger, I.; Egli, M.; Šponer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147−6151.
(69) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(70) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997**, *119*, 59−69.
(71) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comput. Phys.* **1984**, *81*, 3684−3690.
(72) Harvey, S. C.; Tan, R. K.-Z.; Cheatham, T. E., III. *J. Comput. Chem.* **1998**, *19*, 726−740.
(73) Chiu, S.-W.; Clark, M.; Subramaniam, S.; Jakobsson, E. *J. Comput. Chem.* **2000**, *21*, 121−131.
(74) Cheatham, T. E., III; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 4805−4825.

energy estimation involves separately evaluating the free energy for the solute and solvent for a series of snapshots and then averaging the results. The solute enthalpic component is estimated from the in vacuo molecular mechanical energy for each configuration of the solute sampled from the MD trajectory (with the solvent atoms removed), evaluating all pairwise interactions and using a dielectric constant of 1. The use of a dielectric constant of 1 in the energy and continuum calculations, rather than a higher value in the 2−4 range, is consistent with our use of a nonpolarizable force field, and also because the solute response to charge fluctuations is estimated through explicit averaging of conformations. Although this may not be sufficient to fully capture the electronic response of the macromolecule, when higher dielectric constants are applied (2 or 4), the absolute free energies of binding are even further off from the experimental values (data not shown). Solute entropic contributions were estimated as discussed in the next section.

The solvation free energies were estimated for each configuration of the solute sampled at 25 ps intervals from 5 ns portions of the MD trajectories. The electrostatic component of the solvation was estimated with a Poisson−Boltzmann electrostatic continuum method using the program Delphi II.[75] The dielectric boundary is the molecular surface defined by a 1.4 Å probe sphere and by spheres centered on each atom with radii taken from the PARSE parameter set[76] (H = 1.0, C = 1.7, N = 1.5, O = 1.4 Å, with a value of 2.0 Å for the phosphorus). We used an interior dielectric of unity (matching the dielectric chosen when evaluating the solute electrostatic interactions), and the outside dielectric was set to 80. For the Poisson−Boltzmann calculation, a cubic lattice with linear dimensions ~50% larger than the longest dimension was applied with a 0.25 Å grid spacing; potentials at the boundaries of the finite-difference lattice were set to a sum of Debye−Huckel potentials.[77] Salt effects were not included implicitly in the continuum model. The hydrophobic component of the solvation free energy ($\Delta G_{\text{nonpolar}}$) was estimated from a surface area dependent term using the solvent accessible surface area (SASA) algorithm and program of Sanner[78] with a parametrization of $\Delta G_{\text{nonpolar}} = \gamma(\text{SASA}) + \beta$, where $\gamma = 0.00542$ kcal/Å^2 and $\beta = 0.92$ kcal/mol. Note that the available simplistic parametrization of the continuum radii is not necessarily fully consistent with the Cornell et al. force field. However, no better set of radii is currently available, and we believe the present parameters are meaningful to correctly calculate the relative binding energies.

The binding energies were estimated using two approaches: single trajectory versus separate trajectories. *Single trajectory* results mean that the thermodynamics data are extracted from a single trajectory of the complex. The assumption here is that the bound DNA and bound DAPI conformations are similar to their free conformations. Estimation of binding free energies in this manner has proven successful previously.[41,42,50−54] Part of the reason for this is a cancellation of errors that hides the effect of incomplete sampling. A logically better approach, limited in practice due to larger fluctuations and errors, is the use of *separate trajectories* where the binding free energy is estimated from three separate MD trajectories of DNA−DAPI complex, free DNA, and free DAPI; because of sampling limitations, this approach appeared to be significantly less stable in the current study. For all of the free energetic analyses, 200 configurations at equally spaced (25 ps) intervals from 5 ns portions of each trajectory were analyzed. Very similar results are obtained if more frames are included in the analysis (data not shown) or if different parts of the trajectory are separately analyzed (block averages; data not shown).

In addition to the standard MM_PBSA approach, we also performed the energetic analysis including some explicit solvent. We included a subset of solvent representing those waters bound to DAPI or in the minor groove of DNA, that is, the 20 closest waters to the DAPI or donors/acceptors in the minor groove of DNA (when DAPI was not bound). The specific 20 water molecules included for each separate snapshot were chosen in an entirely automated way on the basis of geometric criteria (i.e., proximity to all of the atoms in DAPI or to donor/acceptors (O2, N2, N3, O4′) in the minor groove, neglecting the *GG* overhanging bases). This simple extension of the method improved the stability of the calculated numbers.

Estimations of the Solute Entropic Components at 300 K. Solute entropic contributions were estimated using a subset of the sampled structures (at 125−500 ps intervals) based on a harmonic approximation to the normal modes and standard (quantum) formulas at 300 K.[79] Minimization was done through a series of relaxations that first applied the generalized Born method in AMBER 6.0 (1000 steps total, 75 steepest descent, no cutoff) followed by in vacuo minimizations with a dielectric constant of $\epsilon = 4r$ (r is the interatomic distance in Å) and up to 10 000 steps, the first 100 of which are steepest descent followed by conjugate gradient, unless the RMS gradient drops below 10^{-4}. After this, Newton−Raphson minimizations (max of 5000 steps) were carried out on a series of structures until the root-mean-square of the elements of the gradient vector was less than 10^{-4} kcal/mol Å. For estimation of the entropy, the simulations of free DNA and free DAPI in solution were used along with simulations of the M_ATTG and M_AATT complexes with both the "s" and "r" DAPI force fields. The number of frames were as follows: M_ATTG, 25 frames at 250 ps intervals from 0.25 to 6.5 ns; M_AATT, 20 frames at 250 ps intervals from 0.25 to 5.0 ns; M_ATTG(r), 21 frames at 125 ps intervals from 0.5 to 3.0 ns; M_AATT(r), 17 frames at 250 ps intervals from 0 to 3.4 ns; M_94, 9 frames at 500 ps intervals from 1 to 4.5 ns; D_AATT and D_ATTC, 12 frames at 250 ps intervals from 2 to 4.75 ns; free DAPI, 12+ frames at 250 ps intervals. For the M_ATTG and M_AATT trajectories, estimates of the solute entropy at 300 K lead to a contribution of +21.6 and +23.1 kcal/mol for the **ATTG** and **AATT** binding modes, respectively, favoring **ATTG** by 1.5 kcal/mol. These estimates were obtained calculating the entropic contribution from the complex and subtracting out the entropy from the free DNA and free DAPI simulations. For the M_ATTG, the components at 300 K were 15.0, 15.5, and 466.8 kcal/mol, respectively, for the rotational, translational, and vibrational components of the entropy. This compares to 15.0, 15.5, and 465.4 kcal/mol for the M_AATT, 15.0, 15.5, and 447.0 kcal/mol for the free DNA, and 10.4, 12.8, and 18.3 kcal/mol for the free DAPI (at 300 K). Subtracting the values shows that the major component of the (absolute) entropy of binding is unfavorable and comes from the loss of rotational (10.4 kcal/mol) and translational (12.8 kcal/mol) degrees of freedom and is only slightly offset by a favorable vibrational entropic difference (−1.5 kcal/mol). The estimates of rotational and translational entropy losses in the 20−30 kcal/mol range (or 23.2 kcal/mol) as estimated here[80,81] based on ideal gas values and standard formulas (rigid rotor harmonic oscillator approximation) have been seriously questioned. Various other experimental and theoretical groups estimate the loss due to rotation and translation to be much smaller and in the 3−10 kcal/mol range.[82−87] If estimates of the rotational and translational entropy loss in the 3−10 kcal/mol range are used instead, the agreement with absolute free energies of DAPI binding as compared to experiment is much better.

(75) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−332.
(76) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(77) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327−335.
(78) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. *Biopolymers* **1996**, *38*, 305−320.
(79) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.
(80) Tidor, B.; Karplus, M. *J. Mol. Biol.* **1994**, *238*, 405−414.
(81) Schwarzl, S. M.; Tschopp, T. B.; Smith, J. C.; Fischer, S. *J. Comput. Chem.* **2002**, *23*, 1143−1149.
(82) Page, M. I.; Jencks, W. P. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1678−1683.
(83) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(84) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.
(85) Luo, R.; Gilson, M. K. *J. Am. Chem. Soc.* **2000**, *122*, 2934−2937.
(86) Yu, Y. B.; Privalov, P. L.; Hodges, R. S. *Biophys. J.* **2001**, *81*, 1632−1642.
(87) Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins* **2002**, *47*, 194−208.

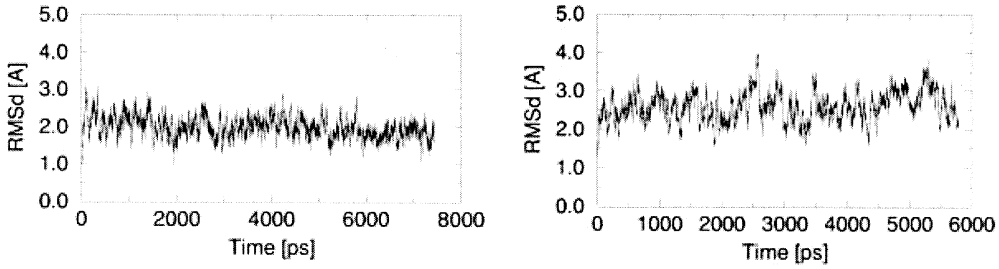

Figure 2. All-atom root-mean-squared deviation to the starting structure of the M_ATTG, omitting the overhanging *GG* residues (left), and D_AATT (right) DNA−DAPI complexes.

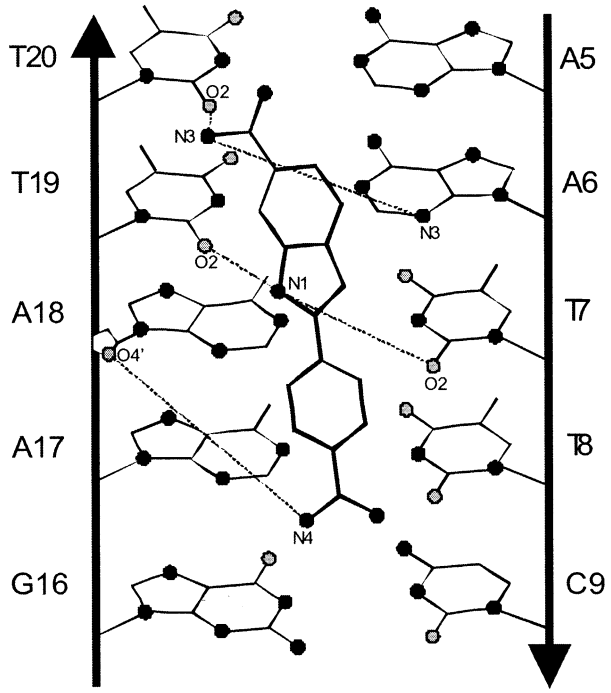

Figure 3. DAPI binding to the AATT region of d(CGCG**AATT**CGCG)$_2$.

Results

We will first comment in detail about the structural data observed in simulations of the two binding sites characterized by X-ray data, G-**AATT**-C and A-**ATTG**-G, and we will follow this by a discussion of several alternate binding sites. After this, a detailed discussion of the relative binding affinities is presented. In general, the structures seemed to equilibrate (based on root-mean-squared deviations) about an equilibria geometry within the first nanosecond of simulation (Figure 2).

Molecular Dynamics of DAPI Complexed with the AATT Region of d(CGCGAATTCGCG)$_2$. A 5.8 ns MD simulation was performed on the basis of the 2.4 Å crystal structure of DAPI bound to d(CGCG**AATT**CGCG)$_2$ (NDB file GDL008)[11] and is denoted as D_AATT (Figure 2, right). As observed in the crystal structure (Figure 3) and reproduced in the simulation, the central N1 (H1) DAPI group interacts with O2 (T19) and O2 (T7), the indole amidinium group forms a bifurcated H-bond to N3 (A6) and O2 (T20), and the phenyl-amidinium N4 nitrogen is bound to O4′ (A18).

The structure of the DNA is stable, and the DAPI remains tightly bound into the minor groove in a single binding conformation. The root-mean-squared deviation for the DNA duplex (neglecting the DAPI) from the starting structure is less than 2.6 Å over the entire trajectory, and the RMSD of the average structure (from 1 to 5.8 ns) to the crystal structure (neglecting the bound DAPI) is 2.3 Å. The DAPI dihedral angles for rotation around the C_6-C_{10}, C_2-C_{12}, and $C_{15}-C_{11}$ bonds are nonplanar (contrary to the constrained planar geometry of the refined crystal structure) and oscillate around 33°, 10°, and 40°, respectively, being close to the optimal geometry of isolated DAPI. Hydrogen bonds between N1 (DAPI) and the O2 groups (T19 and T7) remained conserved (3.0 and 3.2 Å, respectively). The amidinium nitrogens formed H-bonds with O2 groups of T20 and T8 (2.9 Å) and also remained in contact (3.2 Å) with sugar O4′ atoms of adjacent cytosines C21 and C9. An analysis of the water density around the DNA reveals several highly occupied hydration sites (see Figure 4). Three hydration sites are observed in a minor groove pocket near the phenyl end of the DAPI molecule. Two of these sites are almost in the plane of the T8−A17 base pair with the third site above this plane interacting with the two other hydration sites and also the O2/N2 atoms of residues C9/G16, respectively. For the hydration sites in the T8−A17 base pair plane, one of the hydration sites is fairly broad, is located between the two DAPI amines, and interacts with the O4′ atom of residue C9. This hydration site bridges to the N3 atom and base of residue A17 via the second hydration site. In addition to this hydration deep in the groove, there is also a hydration site located on top of the DAPI phenyl ring that represents a three-way bridge between the N5 amine atom of DAPI and the phosphate O1P atoms of residues T19 and T20. At the indol end of DAPI, two hydration sites were in the A5−T20 base pair plane. One of these interacts with the N3 atom of the adenine (A5) and bridges to the interior DAPI amine (N3). These sites are almost 100% occupied by water molecules with residence times of individual water molecules around 0.3 ns. At the indol end, a specific hydration site is found outside the DAPI molecule above the indol that interacts with the outside DAPI amine (N2) bridging the DAPI to the backbone. Water density maps show other hydration sites exhibiting almost 7 times higher than bulk water density near N3 guanine atoms and O2 cytosine atoms, respectively; these hydration sites bridge back to the DAPI via the specific hydration sites of the DAPI terminal amines. This double spine of hydration, in the wider minor groove of the d(CGCG) region, is consistent with the crystal structure. There are several additional and more diffuse hydration sites that may further stabilize the close contacts of DAPI with the sugar−phosphate backbone. These less localized water molecules likely correspond to the second hydration shell.

Molecular Dynamics of DAPI Complexed with the ATTG Region of the d(CCAATTGG)$_2$*GG*. We have also investigated the binding of DAPI to the **ATTG** of d(GGCCA**ATTG**G) based on the 1.9 Å resolution crystal structure of the d(GGCCAAT-

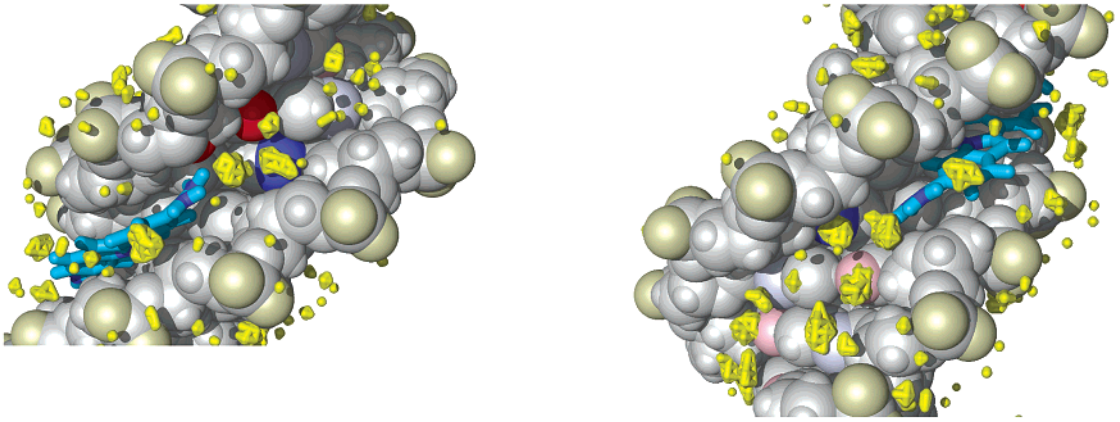

Figure 4. Two views of hydration in the d(CGCGAATTCGCG)$_2$ DAPI−DNA complex. Average water densities at ∼3 times bulk solvation (in yellow), accumulated around the best fit structures to the shown average structure. The phenyl end is shown on the left, and the indole end is shown on the right. In the phenyl end picture, the bridging O2 and O4′ from C9 are red, and the N3 (A17) is shown in blue. Right, the O2 (C21) is shown in red, and the N3 (A5) is shown in blue. The DAPI molecule is shown in cyan with the nitrogens in blue.

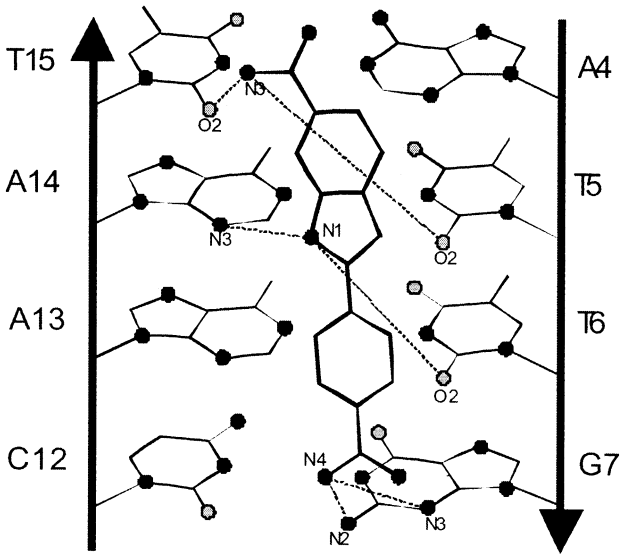

Figure 5. DAPI binding to the ATTG region of dGGCCA**ATTG**G.

TGG)−DAPI complex (NDB code DD0002, Figure 5).[12] The DAPI molecule in this crystal is nonplanar with $C_7-C_6-C_{10}-N_3$, $C_{17}-C_{12}-C_2-N_1$, and $C_{16}-C_{15}-C_{11}-N_4$ intersegment dihedral angles of 21°, 16°, and 15°, respectively. The N1−(H1) (DAPI) group interacts with N3 (A14) and O2 (T6) (bond lengths of 3.1 and 3.2 Å), and the N3 amidinium group attached to the indole ring of DAPI interacts with O2 (T15) and O2 (T5) (with heteroatom distances of 3.2 and 2.8 Å, respectively). The amidinium nitrogen N4 at the phenyl ring of DAPI forms a strong H-bond with N3 of guanine G7 (3.0 Å) with a 3.0 Å contact between the N4 (DAPI) and N2 amino group nitrogen of G7. The starting geometry was taken directly from the crystal structure. However, the two 5′-end overhanging guanine residues that are not adjacent to the binding site were omitted. On the other strand, however, because the two 5′ overhanging guanines are proximal to the binding site and could influence binding, these residues were included in the simulations. Thus, the simulations were actually done for the DNA duplex sequence d(CCA**ATTG**G)·d(*GG*C**CAAT**TGG) (the overhang residues are denoted in italic). We abbreviate this molecule as d(CCA**ATTG**G)$_2$*GG*. Throughout the entire 7.5 ns simulation, the DAPI molecule remains bound in the ATTG binding region. The all-atom root-mean-squared-deviation (RMSD) of the duplex (without overhanging residues G9 and G10 and neglecting the DAPI) with respect to the crystal structure oscillates at a value around 2 Å.

As discussed in greater detail in the Supporting Information, an additional set of simulations were performed with various DAPI force fields to probe the sensitivity of the DAPI force field. In all of these, qualitatively similar behavior is seen in stable simulation of DAPI−DNA binding. Modest differences relate to the sampling of substates and DAPI conformation.

Substates of DAPI Binding in the ATTG Region of the Minor Groove. Contrary to what was observed in the AATT site, the DAPI molecule oscillates within three distinct binding substates in the ATTG binding pocket, designated as S1, S2, and S3 (Table 2). These conformational substates are characterized by rapid conformational switches in the phenyl-amidinium segment bound to the GC base pair, specifically of the $C_{16}-C_{15}-C_{11}-N_4$ dihedral angle (abbreviated as T3). The T3 angle adopted a value of ∼40° in states S1 and S2, while it switches to −40° in state S3 (Figure 6). The N4 (DAPI)−N3 (G7) distance was ∼3 Å in state S1 and 5 Å in states S2 and S3.

None of these computed binding modes is identical to the local geometry reported in the crystal structure. First, the dihedral angle T3 in the substates S1 and S2 has a substantially larger value than is observed in the crystal structure where its value is essentially identical to that of the optimal isolated DAPI. Further, in substate S2, H-bonding toward N3 (G7) is abolished, whereas substate S3 shows the angle T3 to be completely inverted to −40°. However, when considering the average structure from 2 to 7 ns, the central angle T2 at 24.8° is consistent with experiment, and the observed differences in the diamidino groups relate largely to small differences in the relative planarity (at 32.1° and −15.7° for the T1 and T3 angles). Another apparent difference is the lack of close contacts between the guanine amino group and the N4 amidinium nitrogen of DAPI (Table 2). This is true even in the 2−7 ns average structure, where although the N4 (DAPI) is within 3.1 Å of the N3 (G7), it is 4.6 Å from the N2. Among the three substates, S1 is closest to the crystal structure.

The indol-amidinium fragment of DAPI interacts with the upper ATT part of the binding domain throughout the whole simulation in essentially the same manner as in the crystal. Both

Table 2. Three Distinct Binding Patterns Observed during Simulation M_ATTG[a]

substate	T1	T2	T3	N_4 (D)···N_3 (G7)	N_4 (D)···N_2 (G7)	occurrence (ns)	population (%)
X-ray	**21°**	**16°**	**15°**	**3.0 Å**	**3.0 Å**		
S1	30°	10−20°	40°	3.1 Å	3.4 Å	0−2.1, 6.4−7.4	41
S2	30°	−10 to 10°	40°	5.5 Å	5.8 Å	3.4−3.7, 3.9−4.6	13
S3	30°	30°	−40°	5.0 Å	4.6 Å	2.1−3.4, 3.7−3.9, 4.6−6.4, 7.4−7.5	46

[a] T1, T2, and T3 stand for dihedral angles $C_7-C_6-C_{10}-N_3$, $C_{17}-C_{12}-C_2-N_1$, and $C_{16}-C_{15}-C_{11}-N_4$, respectively.

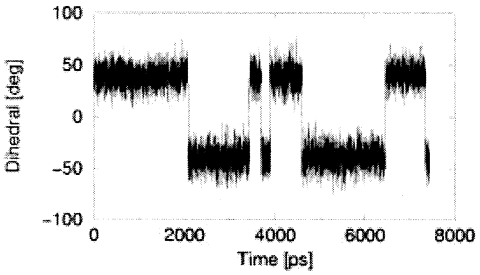

Figure 6. Time dependence of the $C_{16}-C_{15}-C_{11}-N_4$ dihedral of DAPI in the M_ATTG simulation.

Table 3. Main Hydration Sites Observed during Simulation M_ATTG in the T6G7 DNA−DAPI Binding Region

site	location	substates	occupancy	residence times (ns)
H1	O2 (C12)	S1, S2, S3	75−100%	0.1−0.5
H2	N3 (A13), O4′ (A14), N4 (DAPI)[a]	S3,[a] (S1)[b]	∼100%	>1.0
H2	N3 (A13) and O4′ (A14)	S2	50%	0.1−0.3
H3	connects sites H1 and H2	S1, S2, S3	50−100%	0.1−0.3
H4	N2 (G7) and N4 (DAPI)	S1	∼100%	0.1−0.4
H5	N3 (G7), N4 (DAPI), O4′ (G8)	S2	∼100%	0.2−0.4

[a] The N4 (DAPI) interaction is only observed in the S3 substate. [b] Only during period 6.4−7.4 ns (cf. Table 2).

bifurcated hydrogen bonds between the drug and the duplex are well conserved. The averaged N1 (DAPI)−N3 (A14), N1 (DAPI)−O2 (T6), and N3 (DAPI)−O2 (T5) H-bond distances amount to 3.2 Å, while the N3 (DAPI)−O2 (T15) contact is 2.8 Å.

Hydration and Ion Interaction at the Phenyl-amidinium Binding Site. The M_ATTG simulation revealed five significant ordered hydration sites in the region of the DAPI (phenyl amidinium end)−DNA binding site referenced as H1−H5 in the following text (Table 3 and Figure 7).

Hydration sites H1−H3 do not appear to depend on the DAPI geometry, while sites H4 and H5 are substate specific. Hydration site H1 is present near O2 (C12) through the entire simulation. Hydration site H2 is formed by N3 (A13) and O4′ (A14) in substates S1 and S2 and involves an additional N4 (DAPI) interaction during DAPI substate S3. This site was vacant in the initial interval of 0−2.1 ns (during S1 binding). When sites H1 and H2 are occupied, another water molecule forms hydration site H3 interconnecting sites H1 and H2. Hydration site H4 was observed only during the S1 substate of DAPI and manifests as a water bridge between N2 (G7) and N4 (DAPI), likely stabilizing the rather close 3.4 Å contact between them. If the site H2 near N3 (A13) is occupied by a water molecule, the H4 hydrating water molecule is slightly shifted to O4′ (G8). Hydration site H5 occurs only for the S2 substate and involves interactions with N3 (G7), O4′ (G8), and N4 (DAPI).

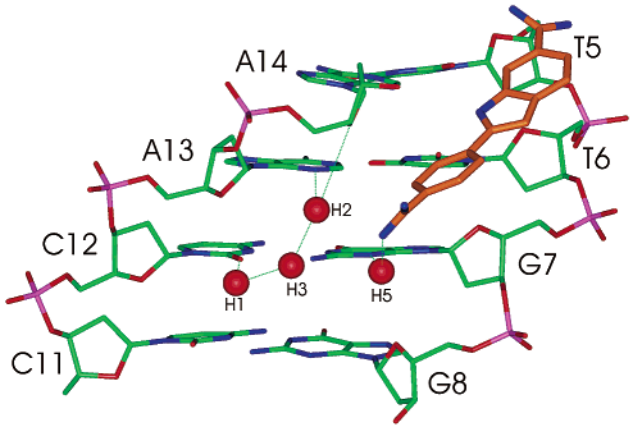

Figure 7. Schematic of the various hydration sites in the phenyl-amidinium DAPI region (simulation M_ATTG, substate S2).

All hydration sites except H2 show residence times of individual water molecules in the usual range of 0.1−0.3 ns,[66,88−90] with a few longer residency times near 0.5 ns. Distinct behavior was observed for the H2 hydration site near N3 (A13). Because of poor accessibility, this site was vacant during the initial 2.1 ns of simulation. At 2.1 ns into the simulation, this site became hydrated and remained occupied for the rest of the simulation despite sampling all of the distinct binding substates. During substates S1 and S3, this site was nearly fully occupied by only two distinct long-residing water molecules (2.1−3.4 ns, S3, and 4.6−7.4 ns, S3 + S1). These long-residing and presumably tightly bound water molecules show considerably attenuated fluctuations as compared to other sites of specific hydration. Occupancy of the H2 during S2 substate is only around 50% with short residency times. Whereas transitions between substates are usually accompanied with exchange of water in the hydration sites, the S3 to S1 substate transition observed at 6.4 ns did not cause any water switch in H2. During sampling of the S2 substate, slightly different behavior of the water was observed as the main hydration site was formed at H5 and involved N3 (G7) and N4 (DAPI) instead of H2. During the water exchange in the H5 site, whenever a new water molecule enters this site, the water formerly residing there moves to the H2 site. Hydration sites H1 and H2 can be clearly identified as crystal water molecules W46 and W42. Hydration sites H3 and H4 appear to correspond to crystal water positions W43 and W76, although here the overlap is not perfect. Site H5 could not be related to any crystal water molecule. The additional simulations confirmed the above hydration picture.

We have also analyzed hydration outside of the DAPI binding sites and in the free duplexes. There is, in the free DNA

(88) Zhou, D.; Bryant, R. G. *J. Biomol. NMR* **1996**, *8*, 77−86.
(89) Phan, A. T.; Leroy, J. L.; Gueron, M. *J. Mol. Biol.* **1999**, *286*, 505−519.
(90) Nagan, M. C.; Kerimo, S. S.; Musier-Forsyth, K.; Cramer, C. J. *J. Am. Chem. Soc.* **1999**, *121*, 7310−7317.

duplexes, formation of a single spine of hydration in the central AATT region that extends to a double spine in adjacent terminal regions. We have been able to identify several hydration sites outside of the binding site that are also seen in the 1.9 Å resolution crystal structure; however, none of them shows extended water residency times. Binding of the DAPI dication inhibits Na^+ interaction with the minor groove. In the absence of DAPI, we have noticed modest interactions of cations with ApT steps in the minor groove and with guanines in the major groove. All of these results are consistent with the literature data.[34,91] We have found only minor effects of the drug binding on the DNA geometry with a tendency to stabilize the canonical B-like conformation (consistent with experiment[28]). For space reasons, further details are not presented.

The Effect of Shifting the DAPI Binding Mode by One Base Pair Step or Altering the Sequence Outside of the Canonical AATT Binding Mode. (a) Shifting from the ATTG to the AATT Region in d(CCAATTGG)$_2$*GG*. To highlight the principles underlying DAPI binding, we have carried out a 5 ns simulation of the d(CCAATTGG)$_2$*GG*−DAPI complex with the DAPI molecule docked one base pair step up into the **AATT** binding site (simulation M_AATT). This leads to a binding similar to that seen in the dodecamer sequence, except that the GC base pairs immediately adjacent to the **AATT** region are reversed. During the time periods from 0 to 1.4 ns and from 3.2 to 3.8 ns, DAPI adopted a binding mode very similar to that observed when it interacts with the **AATT** region of the dodecamer with a persistent N4 (DAPI)−O2 (T6) hydrogen bond. During the time period from 1.4 to 3.2 ns and from 3.8 to 5.0 ns, the complex adopts a second substate where the N4 (DAPI)−O2 (T6) distance increased to 5 Å and oscillations of the $C_{16}-C_{15}-C_{11}-N_4$ DAPI dihedral angle between values +40° and −40° on a time scale of 20−50 ps were observed. The N4 (DAPI)−N3 (A14) distance is 4.5−5 Å for both substates. In the second substate, an unusual hydration site involving the N4 (DAPI), O2 (T6), and N3 (A14) atoms develops with water residency times in the range 0.5−1.5 ns. When the $C_{16}-C_{15}-C_{11}-N_4$ angle adopts a value around 40°, a water bridge is formed between N4 (DAPI) and O2 (T6). When the angle is around −40°, this water bridge shifts to link N4 (DAPI) and N3 (A14). Nevertheless, for a significant fraction of the time, the long-residing water molecule can be considered as connecting N4 (DAPI) with both N3 (A14) and O2 (T6). Despite oscillations of the C15−C11 torsion which are correlated with *local* movement of water between N3 and O2 sites on the DNA, the water molecules remain firmly bound to N4 (DAPI).

(b) Replacing AATTG by AATTC in d(CCAATTCG)*GG*. To investigate the influence of orientation of the GC pair adjacent to the **AATT** region, we have carried out a 5 ns simulation with DAPI bound to the **AATT** sequence but with a reversed GC base pair. This binding site is identical to the dodecamer complex in the whole **AATT**CG region. In the time periods of 0−0.2, 1.7−3.0, 3.8−4.3, and 4.5−5.0 ns, the binding geometry is very similar to that observed for the dodecamer complex with identical main hydration sites. During the remainder of the simulation, fast oscillations of the $C_{16}-C_{15}-C_{11}-N_4$ dihedral angle between values +40° and −40° were observed. Thus, the overall behavior of DAPI is more similar to that observed in the d(CC**AATT**GG)$_2$*GG* simulation than to that in the (CGCG**AATT**CGCG)$_2$ simulation. As before, water molecules were found in the O2 (T6), N3 (A14), and N4 (DAPI) pockets with residency times in the range of 0.2−1.5 ns.

Table 4. Single Trajectory Free Energy Estimates for DAPI Interaction with d(CCAATTGG)$_2$*GG*[a]

trajectory	DNA + DAPI	DNA	DAPI	Δ*G*	ΔΔ*G*	ΔΔ*G**
M_ATTG	−3860.3	−3689.6	−150.4	−20.3	+2.3	+0.8
M_AATT	−3864.9	−3691.4	−150.8	−22.6		

[a] All of the energies (kcal/mol) were estimated from a single trajectory calculating each component (DNA + DAPI, free DNA, and free DAPI) separately. Solute entropic effects are omitted except from the very last column. Δ*G* represents the difference in energy between the complex and free subsystems, $G_{DNA+DAPI} - (G_{DNA} + G_{DAPI})$. The ΔΔ*G* values are the differences between the $\Delta G_{AATT} - \Delta G_{ATTG}$; positive values favor the AATT binding mode.

(c) Investigating Changed Sequence and Shifting with ATTC Binding in d(CGCGAATTC**GCG)$_2$ and d(CCA**ATTC**G)*GG*.** To complete our investigation, we have carried out two 5 ns simulations of DAPI bound to the **ATTC** binding site. In the d(CGCGA**ATTC**GCG)$_2$ simulation, the DAPI molecule was manually shifted one base pair down. In the case of the octamer complex, DAPI remained in the initial position, and the GC pair was reversed. The DAPI molecule remained bound in both simulations to the **ATTC** region with 2.9 Å hydrogen bonds between amidinium nitrogens and O2 atoms of T19 (T15 for octamer) and C9 (C7 for octamer), respectively. The DAPI adopts a single geometry with dihedral angles similar to those of the S1 substate described above. Average values of dihedrals $C_7-C_6-C_{10}-N_3$, $C_{17}-C_{12}-C_2-N_1$, and $C_{16}-C_{15}-C_{11}-N_4$ for the dodecamer (octamer) simulations are 32° (30°), 11° (16°), and 38° (36°), respectively. There is a major hydration site between N4 (DAPI) and N2 (G16) (G12 for the octamer) with residence times around 0.1−0.5 ns. During one-half of the MD simulation of the octamer complex, we see a more structured N3 (A13), N4 (DAPI), and N2 (G12) hydration pocket with waters that have residence times up to 1 ns. In the dodecamer simulation, no water molecule was found near the corresponding atom N3 (A17).

Thermodynamic Analysis

As a complement to the structural data discussed above, estimates of the relative free energy of binding of DAPI to the minor groove of DNA in the various binding modes have been performed. For space reasons, we outline below only the main points of our analysis and present additional details in the Supporting Information.

Thermodynamic Analysis of DAPI Binding to AATT versus ATTG in the d(CCAATTGG)$_2$*GG* from Single Trajectories. Table 4 compares estimated free energies of the DNA−DAPI complex, free DNA, and free DAPI from MD trajectories of the DNA−DAPI complex for both the observed **ATTG** and the hypothetical **AATT** binding sites in the decamer crystal structure.

The calculations are based on single trajectories of the complexes and (except in the last column of Table 4) do not include the entropic effects of the solute. The absolute values of binding free energies without the solute entropy term are in the range around −20 kcal/mol. Estimates of the solute entropy at 300 K (see Methods section) lead to a contribution of +21.6 and +23.1 kcal/mol for the **ATTG** and **AATT** binding modes,

(91) Cheatham, T. E., III; Young, M. A. *Biopolymers* **2001**, *56*, 232−256.

Table 5. Separate Trajectory Free Energy Estimates for DAPI Interaction with d(CCAATTGG)$_2$*GG*[a]

trajectory	DNA + DAPI	DNA	DAPI	ΔG	$\Delta\Delta G$	$\Delta\Delta G^*$
M_ATTG	−3860.3	−3691.7	−149.7	−18.9	+4.6	+3.1
M_AATT	−3864.9	−3691.7	−149.7	−23.5		

[a]The energies were estimated from trajectories of the DNA−DAPI complex as compared to a simulation of free DNA and free DAPI in solution, without solute entropic effects (except for the last column). The ΔG value represents the free energy difference between the complex and subsystems, $G_{DNA+DAPI} - (G_{DNA} + G_{DAPI})$. The $\Delta\Delta G$ values are the difference between the $\Delta G_{AATT} - \Delta G_{ATTG}$; positive values favor the AATT binding mode. See the Supporting Information for further details.

respectively, favoring **ATTG** by 1.5 kcal/mol. Adding together the ΔG values in column 5 of Table 4 with the solute entropic contribution listed above leads to an overall free energy of DAPI binding of ∼0 kcal/mol (not shown). This significantly underestimates the experimental free energies of binding which are around −9 to −12 kcal/mol for DAPI[92] and other minor groove binders.[93−96] Better results (data not shown) are obtained when entropic estimates are obtained directly from the MD covariance matrix or, as discussed in the Methods section, when alternative estimates of the rotational and translational entropy losses on DAPI binding are considered.[83−87] At present, given the uncertainties in the entropic estimates and balance of the continuum model, we are not able to reproduce *absolute* DAPI binding energies to a DNA duplex. However, we believe that the *relative* binding free energies are meaningful, as has been shown in related work.[41,42] These results suggest that the **AATT** site is weakly favored (contrary to what is observed in the crystal).

Thermodynamic Analysis of DAPI Binding to AATT versus ATTG in the d(CCAATTGG)$_2$*GG* from Multiple Trajectories. The above free energy estimates were based on the analysis of single trajectories; that is, the configurations of the "free DNA" were taken from the simulation of the complex simply by removing the DAPI. In principle, more reliable free energies can be obtained using separate trajectories for the complex and each subsystem (Table 5).

For this to be reliable, equivalent sampling should be achieved in all of the simulations independent of binding-related structural effects. Unfortunately, this appears not to be the case even in these relatively long (∼2.5−7.5 ns) simulations, where the DNA structure may be relaxing to a different extent (as represented by B_I/B_{II} and (α,γ) backbone transitions, sugar repuckering, and other motions) in the DNA-drug and DNA-free simulations. As shown in Tables S5 and S6 in the Supporting Information, the free energies of the DNA part of the complexes in various simulations span a range from −3689.6 to −3697.5 kcal/mol. This lack of equilibration or incomplete sampling is hidden in the single trajectory approach due to fortuitous cancellation. When analyzing the *separate* trajectories in longer simulations (Tables 5 and S6), the isolated DNA trajectory does not cancel the variances of the DNA in the complex. This leads to a considerably larger variability of the calculated free energy estimates. Considering these results, we suggest that in the limit of incomplete sampling, it may be better to use the results from single trajectories rather than separate trajectories.

Table 6. Thermodynamic Analysis of the d(CCAATTGG)$_2$*GG* Structure with DAPI Bound, Including the 20 Closest Waters to the Drug (kcal/mol)[a]

trajectory	$E_{complex}$	$E_{DNA+20w}$	DAPI	ΔG	$\Delta\Delta G$	$\Delta\Delta G^*$
M_ATTG	−4085.0	−3915.6	−149.7	−19.7	−0.9	−2.4
M_AATT	−4086.4	−3917.9	−149.7	−18.8		

[a] The ΔG values represent the difference between the DNA−DAPI complex (with 20 closest waters) and the sum of the free DNA (with 20 closest waters) and free DAPI based on a single trajectory. Solute entropy is not included, except for in the last column. The $\Delta\Delta G$ results represent the difference in binding free energies between ATTG and AATT binding of DAPI, with positive values favoring the AATT site.

Including Bound Water. Obviously drug binding is strongly influenced by specific interaction with highly structured water molecules. This specifically bound water is typically neglected by continuum solvent treatments. To see if more consistent results would be obtained, we performed the free energy analysis including some explicit water while treating the rest of the solvent with the continuum model.

To allow for complete solvation of the DAPI in the minor groove, 20 waters (representing more than the first solvation shell beyond the bound DAPI) were chosen on the basis of geometric criteria in each snapshot. This bound water was considered for both the DNA−DAPI complex and the DNA subsystem to allow for proper cancellation, while no explicit water was included when evaluating the separate free energy term of DAPI. Including the bound water does not change the results dramatically, but it clearly leads to more consistency of the simulations; specifically, the differences between the various simulations are reduced (see Supporting Information, Table S7). Further, there is a marked free energy shift in favor of the ATTG binding site. If the calculated solute entropy correction (see above) is included, we now get a consistent free energy favoring the experimentally observed **ATTG** binding (Tables 6 and S7).

Thermodynamics of AATT versus ATTC Binding. We have also compared the binding free energies in the observed dodecamer binding site **AATT** to those of its hypothetical shifted site, **ATTC**. Standard single-trajectory thermodynamic analysis gives absolute binding free energies (omitting solute entropy) that are nearly isoenergetic for the d(CGCGAATTCGCG)$_2$ sequence (D_AATT = −26.1 kcal/mol and D_ATTC = −26.3 kcal/mol). When the 20 closest waters to the DAPI are included, the D_AATT is favored by 0.8 kcal/mol and further stabilized by ∼0.9 kcal/mol due to solute entropic effects, leading to a preference for the *observed* **AATT** binding site by ∼1.7 kcal/mol. When we analyze the d(CCAATTCG)$_2$*GG* sequence, without bound water and omitting the solute entropic effects, the **ATTC** site is favored (**ATTC** ≈ −25.3 and **AATT** ≈ −23.3 kcal/mol). However, when the 20 closest water molecules to DAPI are explicitly included, the *expected* **AATT** binding site becomes favored by ∼1.5 kcal/mol.

Discussion

We have carried out an extended set of explicit solvent MD simulations of DNA−DAPI complexes, to complement the existing experimental data and to highlight advantages and limitations of MD simulations in studies of minor groove binding.

(92) Lan, T.; McLaughlin, L. W. *J. Am. Chem. Soc.* **2001**, *123*, 2064−2065.
(93) Lane, A. N.; Jenkins, T. C. *Q. Rev. Biophys.* **2000**, *33*, 255−306.
(94) Mazur, S.; Tanious, F. A.; Ding, D.; Kumar, A.; Boykin, D. W.; Simpson, I. J.; Neidle, S.; Wilson, W. D. *J. Mol. Biol.* **2000**, *300*, 321−337.
(95) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997**, *271*, 244−257.
(96) Haq, I.; Jenkins, T. C.; Chowdhry, B. Z.; Ren, J.; Chaires, J. B. *Methods Enzymol.* **2000**, *323*, 373−405.

Structural Agreement with X-ray Structures. The binding of DAPI in the MD simulations is fully consistent with the available X-ray data for the **AATT** region. While the refinement of the lower resolution crystal structure imposed DAPI planarity constraints due to medium resolution of the data, the simulations reveal that DAPI is nonplanar. In the **ATTG** site, the overall binding is again stable, with the indol-amidinium DAPI fragment binding closely resembling the crystal. However, the 3.04 Å N4 (DAPI)−N2 (G) contact formed in the crystal by the phenyl-amidinium fragment of DAPI is absent throughout the entire simulation. For a part of the simulation, even the N3 (G)−N4 (DAPI) interaction observed in the crystal is broken and replaced by a water bridge. The simulations show three binding conformational substates of the DAPI, one of them with reversal of the $C_{16}-C_{15}-C_{11}-N_4$ DAPI angle. One of the sources of the discrepancy is the force field approximation neglecting the partial sp^3 hybridization of the guanine amino group, leading to exaggeration of the guanine amino group−DAPI repulsion.[12,97] However, as shown, although the different DAPI substates differ from the crystal geometry, the structure averaged over all substates becomes reasonable close.

DAPI Parametrization. A simple force field prepared (as is common in the literature) by analogy with available Cornell et al. force field parameters leads to an entirely planar DAPI molecule in disagreement with the 1.9 Å X-ray structure and ab initio calculations. To overcome this, a considerably more realistic DAPI force field was obtained by fitting to ab initio data. This provides a DAPI force field that is in excellent agreement with the target ab initio structures and torsional energy profiles (see Supporting Information). In simulations with this force field, the DAPI retains essentially the same values of its three intersegment dihedral angles as are observed in the optimal gas phase structure of DAPI, even in simulations of free DAPI in water. Thus, the geometry of DAPI shows little response to the polar solvent. As compared to the crystal structure, the DAPI nonplanarity actually appears to be somewhat exaggerated. To probe this, we have carried out simulations with a scaled force field for DAPI that reduces its intrinsic nonplanarity. The DAPI structure is again unaffected by the environment but is closer to the X-ray data (see Supporting Information). It should be noted, nevertheless, that a Cambridge Database search revealed substantial variability of the C−C−C−N dihedral angles of DAPI and related molecules.[12] This is one of the reasons why we do not provide any final judgment justifying such DAPI force field scaling. Work is in progress to test several additional adjustments of the DAPI force field and modifications of the DNA force field to better mimic the partial sp^3 pyramidalization of the guanine amino group.

Occurrence of Distinct Substates in the Simulations. The simulations reveal the simultaneous population of several distinct local binding patterns (substates), with interconversion on a nanosecond time scale and conformational changes correlated with specific changes in hydration. Although the balance of substates actually seen in our simulations is likely to be affected by the force field and sampling limitations, we believe that interconverting substates may occur with many minor groove binders.

Occurrence of Integral Hydration Sites. The simulations reveal integral hydration sites with close to 100% occupancies that are involved in the stabilization of the drug binding via formation of water bridges between the DNA and DAPI. Some hydration sites exhibit water residency times above 1 ns; this is significantly longer than residence times typically observed in MD simulations (in the 0.05−0.3 ns range). Very long nanosecond-scale residency hydration sites stabilizing local conformational variations have also been reported recently in simulations of a DNA zipper duplex,[66] a RNA pseudoknot,[98] and other systems.[90,99,100]

Thermodynamics Analysis of the Different Binding Sites. We have applied the MM_PBSA technique,[41,42] an approximate method similar to others in the literature[43−46], to estimate binding free energies. As presently parametrized, especially given the poor estimates of rotational and translational entropy loss in binding, these methods do not accurately predict the absolute drug binding free energies. Nevertheless, the technique provides useful insight into the relative binding energies. Because of sampling limitations, the less rigorous single trajectory approach is more stable and should be preferred over the separate trajectory analysis within the present simulation time scale. A promising way to increase the accuracy of the MM_PBSA approach is an inclusion of a limited number of *explicit* water molecules. This extension of the method clearly favored the observed binding sites for both crystal structures and improved stability of the calculated numbers. The explicit inclusion of water molecules around the binding site does not bring any subjectivity into the analysis. The water molecules are included in an entirely automated way based on geometric criteria. It is a logical extension of the MM_PBSA method to account for the effects of specific structured hydration that, as clearly shown, are significant and specific for each local geometry and substate.

The MM_PBSA calculations allowed us to better understand the recent crystallographic observation of a tight binding of DAPI to the **ATTG** sequence, shifted off of the presumably more favorable **AATT** binding region that is still present in this sequence. These two binding sites are almost isoenergetic with a preference for the *observed* site when bound water molecules are considered. The **ATTG** binding pattern would likely further profit from a proper inclusion of the guanine amino group flexibility into the force field. Notably, the calculations favor the observed **AATT** binding site over the shifted one in the dodecamer sequence, again in full agreement with the crystal data. We are aware that footprinting experiments and other solution data clearly support a preferential binding of DAPI to AT-rich regions. Nevertheless, in the decamer investigated here, DAPI binds to three AT pairs and to one CG (instead of four AT pairs). As the DAPI is still primarily binding to an AT-rich region, in our opinion, this is compatible with the solution data. As noted in the footprinting study by Portugal and Waring,[26] "DAPI appears to bind fairly specifically to a run of 3−4 AT base pairs," which is in clear agreement with our results for all binding sites studied. Additionally, spectroscopic results suggest a profound similarity in CD and absorption spectra when three or more AT base pairs are present, consistent with the ATTG binding mode observed in the crystal.[23] It is to be noted that in the 1.9 Å decamer crystal structure,[12] there are no close contacts between DAPI and the neighboring DNA molecules. The DAPI

(97) Hobza, P.; Šponer, J. *Chem. Rev.* **1911**, *11*, 3247−3276.
(98) Csaszar, K.; Špačková, N.; Stefl, R.; Šponer, J.; Leontis, N. B. *J. Mol. Biol.* **2001**, *313*, 1073−1091.
(99) Schneider, C.; Brandl, M.; Suhnel, J. *J. Mol. Biol.* **2001**, *305*, 659−667.
(100) Guo, J. X.; Gmeiner, W. H. *Biophys. J.* **2001**, *81*, 630−642.

is deep in the minor groove of the AATTG region, which is suspended between the triplet forming parts of the crystal packing. This suggests that DAPI binding in the minor groove is not markedly influenced by crystal packing. Our simulations together with our preceding X-ray structure[12] reveal that a single guanine amino group at the end of a AT stretch does not prevent DAPI binding in the minor groove, and our results further support the preferential binding of DAPI to AT-rich regions.

Conclusions

State-of-the-art molecular dynamics simulation methods accurately describe drug−DNA structure and hydration in the minor groove. Application of methods to estimate approximate free energies of binding, with inclusion of some explicit water around the binding sites, provide useful insights into the binding of DAPI to various sequences in the minor groove of DNA. Although there are clear limitations in sampling, force fields, and balance of the approximate models, we are excited by the progress and look forward to calculations on a much wider series of DNA minor groove binding drugs.

Acknowledgment. The study was supported by grants LN00A016 MSMT CR (N.S.) and LN00A032 MSMT CR (F.R., F.L., P.H., J.S.). Simulations were carried out using the computational resources available at the Supercomputer Center, Brno, Czech Republic, with the thermodynamic analysis and supplemental simulations performed using facilities at the Center for High Performance Computing, University of Utah, and NRAC and friendly computer resources at the Pittsburgh Supercomputing Center and the University of Kentucky allocated to T.E.C. (MCA01S027S).

Supporting Information Available: (i) Details about force field parametrization, including a complete set of DAPI torsional profiles obtained by all force fields and the quantum-chemical data, (ii) a complete listing of the DAPI force field parameters employed in the current work, (iii) discussion of the performance of the PARM98/99 force field and various DAPI force fields including a complete set of thermodynamics data from control simulations and simulations with the (r) force field (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA025660D