Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6402–6406, September 1986
Biophysics

Molecular mechanical studies of d(CGTACG)$_2$: Complex of triostin A with the middle A·T base pairs in either Hoogsteen or Watson–Crick pairing

U. C. SINGH, N. PATTABIRAMAN, R. LANGRIDGE, AND P. A. KOLLMAN

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

Communicated by I. Tinoco, Jr., March 24, 1986

ABSTRACT **Computer graphics model building and molecular mechanical calculations have been carried out on d(CGTACG)$_2$ and its bis-intercalation complexes with triostin A and an *N*-Me-Ala analogue of triostin A. Two conformations of the DNA have been considered both for the uncomplexed and for complexed nucleic acid: in one the central A·T base pairs are Watson–Crick base paired; in the other they are Hoogsteen base paired. The calculations offer a clear explanation why Hoogsteen base pairing is not favorable in isolated A+T-rich DNA and also suggest reasons why the bis-intercalation of triostin A might help stabilize the neighboring A·T base pairs into a Hoogsteen form. To our knowledge, this is the first study to use molecular mechanical and dynamical methods to investigate the bis-intercalator–DNA complex.**

The basic features of the double-helical DNA structure proposed by Watson and Crick (1) have stood the test of time and have been validated by the single-crystal structural studies on d(CGCGAATTCGCG)$_2$. However, a few years after the Watson–Crick (WC) structure was solved, the discovery of an alternative base-pairing scheme by Hoogsteen (2) threw some doubt on the basic WC model, particularly for DNA containing adenine and thymine bases. This doubt is further enhanced by the results of theoretical calculations (3, 4), which suggest that Hoogsteen (HG) base pairing for isolated adenine and thymine bases [involving hydrogen bonding between the N7 of adenine and the N3 hydrogen of thymine (N7 · · · H-N3) and between the C6 amino group of adenine and the C4 oxygen of thymine (N6-H · · · O4)] is as favorable or slightly more favorable than WC [involving hydrogen bonding between the N1 of adenine and the N3 hydrogen of thymine (N1 · · · H-N3) and between the C6 amino group of adenine and the C4 oxygen of thymine (N6-H · · · O4)]. Furthermore, molecular mechanical and structural studies on adenosine suggest that the *syn* conformation of the adenine ring (5) relative to the sugar (required for HG hydrogen bonding) is of comparable stability to the *anti* conformation found in the WC structure. Given the intrinsic base pairing and nucleoside conformational preferences, why do A·T base pairs in DNA not exist in HG structures? We examine this question by studying the relative stabilities of d(CGTACG)$_2$ with the central two base pairs of HG and WC structures using a combination of computer graphics model building and molecular mechanics calculations. Our calculations allow a clear and reasonably convincing explanation why A+T DNA should involve WC rather than HG base pairing.

Wang *et al.* (6) have solved the structure of complexes of the bis-intercalators triostin A (Fig. 1) and echinomycin (7) [antibiotics with planar quinoxaline rings attached to cyclic depsipeptides (8–10)] with the above DNA sequence and have found, in the presence of the drug, that the central A·T base pairs are in the HG form. Thus, it was of considerable interest to investigate why this occurred. Would d(CGTACG)$_2$ have HG base pairing in its A·T portion and, if not, how does the presence of the bis-intercalator induce such base pairing? Our calculations strongly predict that NMR studies will show that isolated d(CGTACG)$_2$ has a purely WC structure in solution and suggest how triostin A might preferentially stabilize HG base-paired structures.

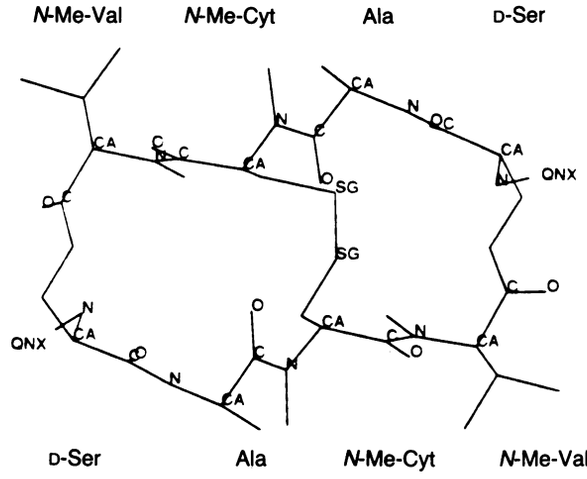

FIG. 1. Skeletal model of triostin A.

MOLECULAR MODELING AND MOLECULAR MECHANICS

Molecular mechanics calculations were carried out on the following DNA structures and DNA–drug complexes: (*i*) DNA with the sequence (CGTACG)·(CGTACG), (*ii*) a DNA–triostin complex, and (*iii*) a DNA–Me-Ala-triostin complex. For the above structures, both HG and WC base-pairing schemes for the middle A·T base pairs were considered. The G·C base pairs were kept as WC base pairs.

The structure of the triostin–DNA complex was model-built from the torsion angles of the DNA, and the triostin molecule and also the pucker conformations of the deoxyribosugar of DNA from the x-ray crystal structure published by Wang *et al.* (6). The triostin molecules were "docked into" DNA using the Evans and Sutherland PS2 at the University of California, San Francisco (11, 12). We then used the molecular mechanics program AMBER (13, 14) with constraints on the hydrogen bonds between the base pairs to create coordinates very close to those in the published structure (6). The WC A·T base pairs at the middle of the DNA–triostin complex were model-built using computer

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HG, Hoogsteen; WC, Watson–Crick.

graphics from the coordinates of B DNA (15). The glycosidic conformation of the adenine base was assumed to be *anti* instead of *syn* as observed in the DNA–triostin complex. The electrostatic potential charges for the methylated amino acids and the quinoxaline ring were calculated using the Gaussian 80-UCSF program with the STO-3G bases set used for the nucleic acid fragments (16). The Me-Ala-triostin was generated by replacing the hydrogen of alanine with a methyl group.

The structures were energy refined using molecular mechanics until the rms gradient was less than 0.1 kcal/mol Å using a distance-dependent dielectric constant, $\varepsilon = R_{ij}$; molecular mechanics calculations were also carried out placing counterions bisecting the phosphates atom P and the two nonester oxygens. The counterions with van der Waals parameters $\varepsilon = 0.10$ and $R^* = 5.0$ have been found to be a useful way to ensure charge neutrality without necessitating strong "contact" ion pairs or reducing the charges of the phosphate groups of DNA (17).

RESULTS

WC and HG DNA Structure. From Table 1 (columns 2 and 3), it is clear that the structure of the uncomplexed DNA with WC A·T middle base pairs is more stable than the structure with HG A·T base pairs by 10.8 kcal/mol. The base stacking energies between the base pairs plotted against the base-pair sequence are shown in Fig. 2. The base stacking energy between the HG A·T base pairs is less favorable than that of the WC A·T base pairs by 0.5 kcal/mol. However, the total base stacking energy between the two models is almost the same. The major contribution to the destabilization of HG structure is due to the local interactions among the 3′ and 5′ phosphates with the sugar and the base of deoxyriboadenine. For an HG A·T base pair, the distance between the C1′ atoms of the two sugars is shortened by 2 Å compared to the WC A·T base pair. Due to the shortening of the distance, the sugar-phosphate backbone has been pulled "inside" the helix compared to the rest of the backbone; and since the phosphates remain as far apart as possible, this movement reduces the sugar-phosphate attraction. Fig. 3 shows the various components of interaction energy among different groups around the deoxyriboadenine for the HG and WC models.

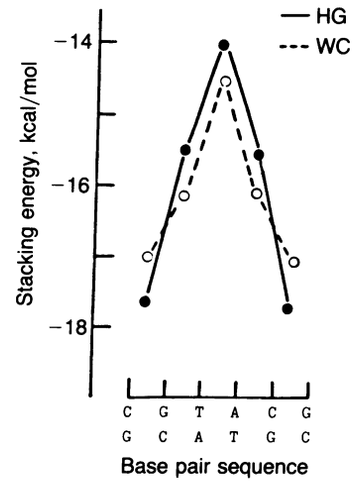

FIG. 2. A plot of the stacking energies of the base pairs versus the base-pair sequences for the WC (broken line) and HG (solid line) models of uncomplexed DNA.

The interaction between the adenine base with its sugar is almost the same for both the *syn* and *anti* conformations for the base. From Fig. 3 it is clear that the sugar-phosphate interaction in the HG model is less favorable than that in the WC model by 2 kcal per pair or more. The interaction between the 5′ phosphate and the adenine base, which is in the *syn* conformation for the HG model, is less favorable by about 3 kcal than the interaction for the WC model in which the adenine base is *anti*.

We have also carried out the molecular mechanics optimization including counterions in the calculations, and the DNA energy differences are similar to those found in the calculations without counterions (Table 1). The sugar pucker conformation for the adenine sugar has repuckered from C2′ *endo* to O1′ *endo* in the case of HG model (both with and without counterions). In the case of WC model, all the sugar pucker conformations are in the C2′ *endo* region.

DNA–Triostin Complex. In Table 1, the energy components for the interaction of DNA with triostin for both the WC (column 6) and HG (column 7) models are given. The WC

Table 1. Components of the interaction energies for the uncomplexed DNA and the DNA–drug complexes

	Interaction energies, kcal/mol							
	Uncomplexed DNA				Triostin–DNA complex*			
	Without counterions		With counterions		Without counterions		With counterions	
	WC	HG	WC	HG	WC	HG	WC	HG
Total energy	−572.5	−561.7	−795.7	−787.0	−635.3	−629.2	−843.0	−840.5
Intra-energy of DNA	−572.5	−561.7	−570.1	−560.9	−503.7	−491.0	−504.0	−489.6
Intra-energy of drug 1					45.4	45.4	48.2	48.3
Intra-energy of drug 2					45.4	45.6	48.5	48.6
Interaction energy of drug 1 with DNA					−111.1	−114.6	−109.7	−113.9
Interaction energy of drug 2 with DNA					−111.1	−114.4	−109.9	−113.7
Counterion interaction with DNA			−281.2	−290.1			−255.6	−261.4
Counterion–counterion interaction			55.6	64.0			48.2	50.6
Counterion interaction with drug 1							−4.3	−3.9
Counterion interaction with drug 2							−4.2	−5.1

Drug 1, molecule 1 of triostin A. Drug 2, molecule 2 of triostin A.
*Drug–drug interaction (−0.3 kcal) is omitted.

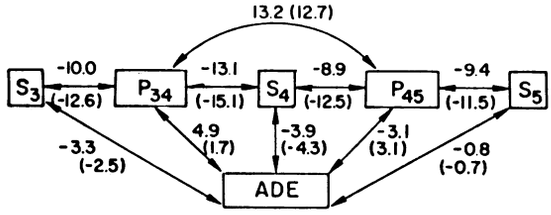

FIG. 3. Component analysis of the interaction energies for the sugar-phosphate backbone in the vicinity of the adenine base in the HG and the WC (shown in brackets) models.

model is more stable by about 7 kcal/mol than the HG model. The intramolecular energy of DNA in the WC model is more stable than that in the HG by about 15 kcal/mol, but the interaction energy of the drugs with the DNA is more favorable by 6.5 kcal/mol in the HG model than in the WC model. The difference in the intramolecular energy of DNA between the two models is essentially due to the same type of interactions among the 3′ and 5′ phosphates with the sugar and the base of deoxyriboadenosine as found in uncomplexed DNA. The results for the *N*-Me-Ala-triostin complex with DNA are qualitatively similar to those for the triostin–DNA complex. The DNA portion is more stable in WC, and the drug–DNA energy favors the HG model.

Fig. 4 shows the base-pairs stacking energy plotted against the base pair of the DNA. Q_1, Q_2, Q'_1, and Q'_2 denote the quinoxaline rings. The stacking energy between the middle A·T base pairs is 3 kcal more favorable for the HG model than the WC model. The stacking energies between the quinoxaline rings and the base pairs are similar for both the models. It is clear from Table 1 that the intramolecular energy difference of the DNA between the HG and WC models is similar for both complexed and uncomplexed structures.

The interaction energies of individual amino acids of molecule 1 and molecule 2 of triostin A with DNA are given in Table 2 for the HG and WC models. The stacking energies between the ring and the base pairs contribute significantly to the binding energy of the drugs. The interaction energies between serine residues and the DNA are similar for both the models. Another important interaction between triostin A and DNA is due to alanine residues (7). In the triostin A molecule, only the alanine residue has the amino group capable of forming a hydrogen bond. The calculations for both the models show that there are four hydrogen bonds between a triostin and the DNA molecules (Fig. 5 *a* and *b*).

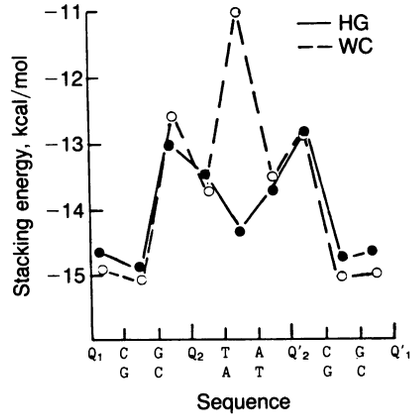

FIG. 4. A plot of the stacking energies in kcal/mol of the base pairs and the quinoxaline rings versus the sequences and the quinoxaline rings for the DNA–triostin complex. The solid lines represent the HG model, and the dashed lines represent the WC model.

In both the WC and HG models, there are two strong hydrogen bonds of length within 1.9 Å between the amino group of alanines and N3 of guanines (N-H · · · N3), also two other hydrogen bonds between the carbonyl group of alanines and the C2 amino group of guanines (O · · · H-N), one of which is slightly weaker because of the asymmetrical binding of the triostin molecule with DNA. The interaction energies between the *N*-Me-Cys and DNA are small for both the models. The role of these residues in the triostin molecule is apparently to form the disulfide bridge to give rigidity to the backbone. *N*-Me-Val interacts quite asymmetrically with DNA. In the case of the HG model, one of the *N*-Me-Val (particularly side chain) of the triostin molecule has a good van der Waals contact with the sugar and 3′ phosphate of adenine. The difference in this interaction is illustrated in Fig. 5, where the dotted van der Waals molecular surface (18) is shown for one of the side chains of valines (shown as a higher density of dots) and the sugar-phosphate backbone (shown as a lower density of dots) in contact with the valine. This interaction is absent in the WC model because in the WC model the minor groove is 2 Å larger than in the HG model. This gives rise to a stabilization energy difference between HG and WC of about 3 kcal per interaction or a total of 5.6 kcal. This interaction is the dominant reason why the drug–DNA energy (Table 1) is more favorable in the HG than in the WC model.

The difference in the helical twists of the middle A·T base pairs between the uncomplexed and the complexed DNA with the drug molecules is similar for HG and WC models. The average helical twist between the middle A·T base pairs and C·G base pairs is approximately 24° and that between the G·C base pairs and the C·G base pairs at both the ends is approximately 12°. In the HG model of drug–DNA complex, the sugar pucker conformations are as follows: C1, O1′ *endo*; G2, O1′ *endo*; T3, C2′ *endo*; A4, C2′ *endo*; C5, O1′ *endo*; and G6, C2′ *endo*. It may be noted that the sugar pucker conformations of C1 is O1′ *endo* and not the O1′ *exo* as reported in the crystal (7). In the WC model, the sugar pucker conformations are as follows: C1, O1′ *endo*; G2, O1′ *endo*; T3, C2′ *endo*; A4, C2′ *endo*; C5, C3′ *endo*; and G6, C2′ *endo*.

When counterions are included in the calculations the difference between the total energies of the two models is narrowed to 2.5 kcal from the 6.1 kcal found without counterions, still favoring the WC model. These calculations have similar intra- and intermolecular energies as those without counterions. A least-squares fit of the structures of the two models (Fig. 6) shows that the rms deviations for the drug molecules and the C·G base pairs between the two models are less than 0.5 Å, with the main difference coming from the sugar-phosphate backbone for the middle A·T base pairs.

Base Sequence Specificity. Triostin A generally prefers G+C-rich DNA to A+T-rich DNA (9). As noted above and by Wang *et al.* (6), there are two key hydrogen bonds between each triostin A and DNA, and one is a C=O · · · H-N bond involving the exocylic C2 amino group of guanine (Fig. 5). This hydrogen bond alone can rationalize a G·C preference in the drug since an A·T base pair has no proton donor in the minor groove. Changing the *N*-methyl groups of *N*-Me-Ala, *N*-Me-Cys, and *N*-Me-Val to hydrogens changes triostin to TANDEM. These changes have a dramatic effect on base specificity, and TANDEM prefers A-T sequences (19) and has an unusually strong affinity for poly d(A-T)·poly d(A-T). An examination of Fig. 5 suggests two possible reasons for this observation; first, the amide group in TANDEM can form a strong electrostatic/hydrogen-bonding interaction with the electron-rich A·T minor groove, just as the A·T preferring netropsin does (20, 21). Second, this amide might form an intramolecular hydrogen bond with the carbonyl group in TANDEM, thus pulling this group away from DNA,

Table 2. Interaction energies of individual amino acids of triostin-A with DNA

Model	Drug molecule	Interaction energy, kcal/mol				
		QNX	D-Ser	Ala/*N*-Me-Ala*	*N*-Me-Cys	*N*-Me-Val
Triostin–DNA (HG model)	Molecule 1	−35.2	−10.9	−12.8 (−7.7)	−2.7	−4.5
		−19.3	−4.6	−14.9 (−9.6)	−2.9	−7.8
	Molecule 2	−35.2	−10.8	−13.3 (−7.9)	−2.4	−4.6
		−19.2	−4.5	−14.8 (−9.7)	−2.0	−7.6
Triostin–DNA (WC model)	Molecule 1	−34.4	−11.2	−12.6 (−7.6)	−3.2	−4.4
		−19.1	−5.3	−14.3 (−9.1)	−1.7	−4.8
	Molecule 2	−34.4	−11.2	−12.6 (−7.6)	−3.1	−4.3
		−19.1	−5.4	−14.3 (−9.1)	−1.8	−5.0

QNX, Quinoxaline ring.
*The numbers in parentheses represent the electrostatic interaction energies.

preventing its interaction with the exocyclic C2 amino group of guanine and reducing any carbonyl group repulsion with the electron-rich A·T minor groove.

DISCUSSION AND CONCLUSIONS

In this paper we have presented molecular mechanics calculations that, to our knowledge, are the first molecular mechanics calculations performed on a bis-intercalator–DNA complex. We have concluded from our calculations that uncomplexed DNA with the A·T base pairs in the HG base-pairing scheme is less stable compared to the model with the WC base-pairing scheme. When $d(CGTACG)_2$ forms a complex with triostin A, its HG structure is stabilized relative to WC by a number of specific and nonspecific interactions. First, the total base and drug stacking energy in the HG structure is more favorable than in the WC model. Second, the *N*-Me-Val side chain of triostin can form a much more favorable van der Waals interaction in the HG than WC structure. Ughetto *et al.* (7) point out the possible importance of Mg^{2+} in stabilizing the crystal and the drug–DNA complex in the crystal. Our counterion calculations also suggest a role for cations in stabilizing the drug–DNA complex in general and the HG over WC structure in particular.

Our calculated total energies for our most complete (counterion included) model for the triostin complex with the HG base pair are slightly higher (2.5 kcal/mol) than that for the WC structure, even though this energy difference is much less than the 8.0 kcal/mol difference found for the isolated DNA fragment. We note that the 2.5 kcal/mol energy difference comes from a balance between stacking and drug–DNA van der Waals terms (which favor the HG structure) and DNA sugar-phosphate and phosphate-base interactions (which favor WC). The former two terms should be less solvent and dielectric constant dependent, and thus it is not unreasonable to image "weighting" them more and concluding that such a model "reproduces" the experimental observation of an HG triostin–DNA complex. On the other hand, crystal-lattice forces may be important in making the HG structure in net more stable than WC, and so our calculated energies may be more relevant for the situation in solution. "Footprint" experiments (22–24) on various DNAs using echinomycin show a strong preference for binding to a central G·C site, usually flanked by A·T sites. The G·C preference is readily explained by the hydrogen bonds observed in the crystal complex (6, 7) and also found here in both HG and WC structures. The preference for A·T flanking sites could indicate a net preference for an HG structure in solution or could merely reflect the fact that A·T base pairs,

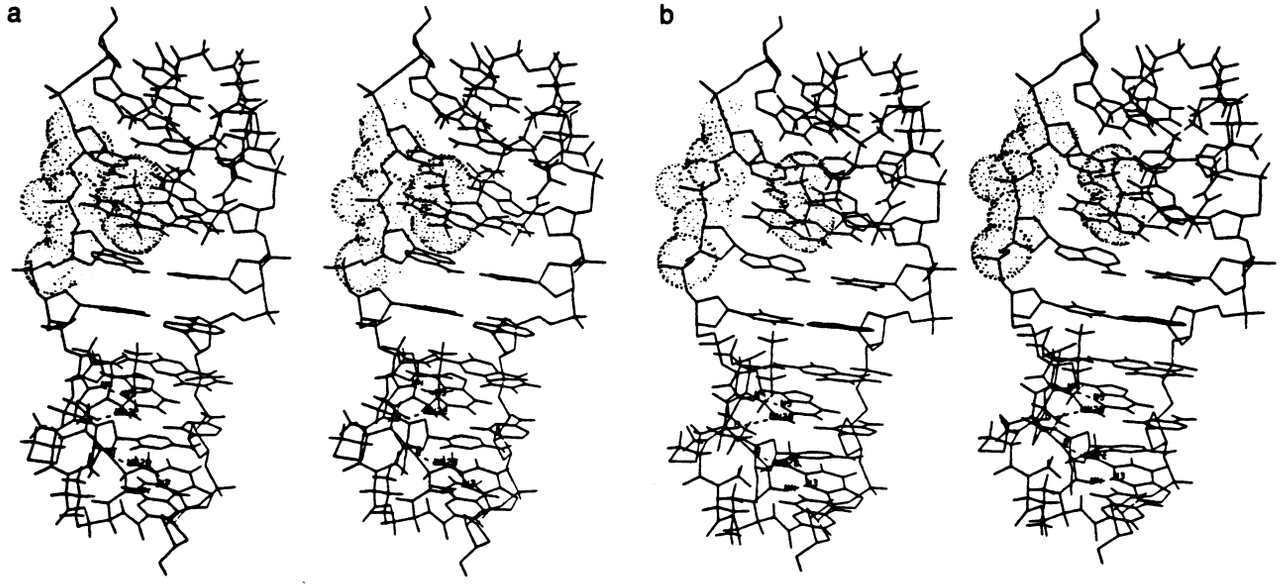

FIG. 5. Stereo views of the van der Waals molecular surfaces for one of the valine side chains and the sugar-phosphate chain in contact with the valine side chain for DNA–triostin complex. Dashed lines represent the hydrogen bonds between the drug and the DNA for one (lower) of the drug molecules for clarity. (*a*) HG model and (*b*) WC model.

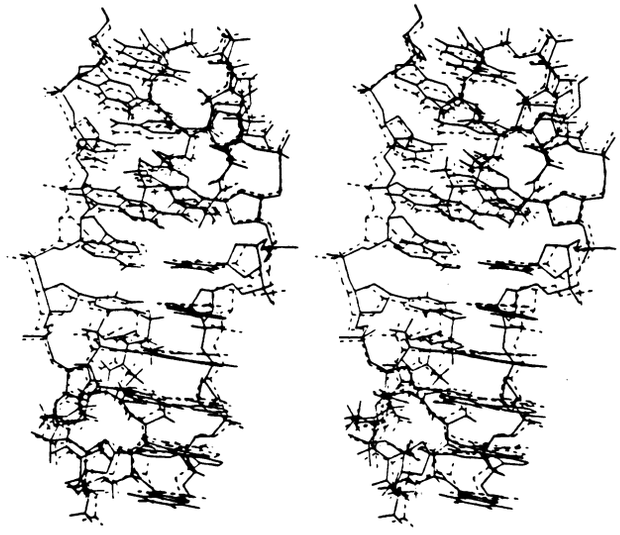

FIG. 6. A stereo view of the superimposed structure of the HG and the WC model for the DNA-triostin complex. HG, solid lines; WC, dashed lines.

being more flexible, can respond with less cost in energy. Definitive high-resolution NMR studies on complexes of triostin with various small DNA fragments are required to resolve this issue. Our calculations suggest that counterion effects may be of interest as well since they may influence the HG ⇄ WC equilibrium. Despite the uncertainties in molecular mechanical energies/model due in large part to our neglect of specific solvent effects, one can conclude that an HG base-paired structure is unlikely to be found in uncomplexed DNA with neutral bases, unless specific ion effects can overcome the significant preference for WC pairing we find due to sugar–phosphate and phosphate-base interactions. Second, a stereochemically acceptable complex of triostin A with $d(CGTACG)_2$ made up only of WC base pairs can be built and is of comparable energy to the structure with T·A HG base pairs. Drug–DNA interactions (better base stacking of the A·T base pairs and a more favorable van der Waals contact of the *N*-Me-Val residue of echinomycin with the DNA backbone in the more compact HG structure) favor the HG structure.

This research was supported by Grants CA-25644 (P.A.K.) and RR-1081 (R.L.) from the National Institutes of Health.

1. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171,** 737–738.
2. Hoogsteen, K. (1959) *Acta Crystallogr.* **12,** 822–823.
3. Langlet, J., Claverie, P., Cavon, F. & Boeuve, J. C. (1981) *Int. J. Quant. Chem.* **20,** 299–338.
4. Weiner, S., Kollman, P. A., Case, D., Singh, U. C., Ghio, C., Alagona, G. & Weiner, P. K. (1984) *J. Am. Chem. Soc.* **106,** 765–784.
5. Kollman, P., Weiner, P., Quigley, G. & Wang, A. H.-J. (1982) *Biopolymers* **21,** 1945–1969.
6. Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H. & Rich, A. (1984) *Science* **225,** 1115–1121.
7. Ughetto, G., Wang, A. H.-J., Quigley, G. J., van der Marel, G. A., van Boom, J. H. & Rich, A. (1985) *Nucleic Acids Res.* **13,** 2305–2323.
8. Katagiri, K., Yoshida, T. & Sato, K. (1975) in *Antibiotics, Mechanism of Action of Antimicrobial and Antitumor Agents,* eds. Corcoran, J. W. & Hahn, F. E. (Springer, Heidelberg), Vol. 3, pp. 234–251.
9. Waring, M. J. (1979) in *Antibiotics, Mechanism of Action of Antieukaryotic and Antiviral Compounds,* ed. Hahn, F. E. (Springer, Berlin), Vol. 5, Part 2, pp. 173–194.
10. Waring, M. J. & Wakelin, L. P. G. (1974) *Nature (London)* **252,** 653–657.
11. Dearing, A. (1981) *CHEM*—A Molecular Display Program (Univ. of California, San Francisco).
12. Jarvis, L., Huang, C., Ferrin, T. & Langridge, R. (1984) *MIDAS*—Molecular Interactive Display and Simulation Program (Computer Graphics Laboratory, Univ. of California, San Francisco).
13. Weiner, P. & Kollman, P. (1981) *J. Comp. Chem.* **2,** 287–303.
14. Weiner, P. K., Singh, U. C., Kollman, P. A., Caldwell, J. & Case, D. A. (1984) *AMBER* (UCSF)—A Molecular Mechanics and Dynamics Program (Univ. of California, San Francisco).
15. Arnott, S. & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47,** 1504–1510.
16. Singh, U. C. & Kollman, P. A. (1984) *J. Comp. Chem.* **5,** 129–145.
17. Singh, U. C., Weiner, S. J. & Kollman, P. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 755–759.
18. Bash, P. A., Pattabiraman, N., Huang, C., Ferrin, T. E. & Langridge, R. (1983) *Science* **222,** 1325–1327.
19. Lee, J. S. & Waring, M. J. (1978) *Biochem. J.* **173,** 129–144.
20. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1376–1380.
21. Caldwell, J. & Kollman, P. A. (1986) *Biopolymers* **25,** 249–266.
22. Low, C. M. L., Olsen, R. K. & Waring, M. J. (1984) *FEBS Lett.* **176,** 414–420.
23. Low, C. M. L., Drew, H. R. & Waring, M. J. (1984) *Nucleic Acids Res.* **12,** 4865–4879.
24. van Dyke, M. M. & Dervan, P. B. (1984) *Science* **225,** 1122–1127.