Volume 9 Number 6 1981 Nucleic Acids Research

Molecular mechanical studies of proflavine and acridine orange intercalation

Andrew Dearing, Paul Weiner and Peter A.Kollman

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143, USA

Received 5 January 1981

ABSTRACT

Previous workers have reported that proflavine and acridine orange form various structurally different complexes with the dinucleoside phosphates rCpG and dCpG, with uniform C3′-endo and mixed C3′-endo (3′-5′) C2′-endo sugar puckers being observed. We present theoretical calculations, based on the method of molecular mechanics, which support the experimental observations. The results suggest that the mixed C3′-endo (3′-5′) C2′-endo pucker conformation is intrinsically more stable than the uniform C3′-endo conformation, but that the additional stabilisation gained from specific, hydrogen bonding, interactions between nucleic acid and solvent, or intramolecularly within the nucleic acid, can lead to the adoption of the latter conformation, or of variants between the two. The role played by hydrogen bonding between proflavine amino-groups and nucleic acid phosphate appears more subtle than previously supposed.

1. INTRODUCTION

The binding of small molecules to nucleic acid to produce intercalated complexes has been the subject of much investigation since Lerman[1] first proposed a model for the interaction between acridines and DNA. This model involves extension of the nucleic acid double helix so as to double the separation of base pairs at the binding site, simultaneously unwinding the helix in the vicinity of the site. The acridine molecule is then inserted, or intercalated, into the space provided, stacking approximately parallel to the base pairs on either side.

Subsequent experimental work has defined more closely the possible structural properties of the intercalated complex including the unwinding angle at or near the intercalation site, the puckering of the ribose or deoxyribose sugar groups and the torsional angles along the ribophosphate backbone, as well as the disposition of the intercalated molecule within the binding site. In addition, a neighbour-exclusion principle (see for example reference (2)) has been observed, whereby only every other site along a polynucleotide helix can be occupied.

Crystal structure analyses performed on dinucleoside phosphate analogues of the DNA complex have shown the existence of at least two different types of conformation. Sobell[3] has studied an r-i^5CpG:acridine orange complex which possessed mixed C3′-endo (3′-5′) C2′-endo sugar pucker and had a relatively large unwinding angle (properties which are also apparently present in ellipticine and ethidium complexes), and an i^5CpG:proflavine complex which had uniform C3′-endo sugar pucker and a

small unwinding angle. Neidle[4] has found that the two dinucleoside phosphate strands in the complex dCpG:proflavine can adopt different conformations, one uniform C3′-endo and the other mixed C3′-endo (3′-5′) C2′-endo, and in subsequent, more detailed, analysis[5] has determined a very well defined spatial orientation of water molecules around this system which appears to stabilise the complex. Other structure determinations[6,7,8,9,10] have demonstrated uniform C3′-endo puckering with proflavine, and mixed C3′-endo (3′-5′) C2′-endo puckering in other cases.

It has been suggested[11] that the adoption of a mixed sugar pucker C3′-endo (3′-5′) C2′-endo is a factor leading to the existence of an exclusion principle, since this conformation can exist at most at alternate sites along the helix. However, there has been some discussion in general about the extent to which the sugar pucker is dependent on, or a cause of, the flexibility of the nucleic acid backbone. Whilst it is difficult to say how well the results of structural analyses performed on dinucleoside phosphate complexes (which are considerably more flexible than a continuing polynucleotide helix) can be applied to DNA complexes, the data on these small-scale analogues provides the most detailed information currently available on the intercalation phenomenon.

The question we wish to address in this paper is whether the varying conformations observed in these different complexes indicate significant differences in the intrinsic properties of the intercalating agents and dinucleoside phosphates involved, or whether they are merely examples of a family of conformations whose relative stability is determined by external factors such as the water and ion content of the crystals studied. We use a theoretical approach to investigate the properties both of the observed conformations and also of alternative structures which have not to date been found, in order to determine the preferred conformation for each system and the extent of its stability.

Various workers have attempted theoretical studies on the intercalation phenomenon in order to determine the structural attributes which give rise to the observed properties. Most of these have considered the flexibility of the nucleic acid backbone; if anything, their studies have been characterised by the diversity of results obtained. Alden and Arnott studied models of di- and tetra-nucleotide systems to determine the characteristics of intercalation sites compatible with adjacent A-[12] and B-DNA.[13] They found that the most acceptable results showed mixed C3′-endo (3′-5′) C2′-endo pucker and a helical unwinding of 18° spread over three residues if the continuation had a B-like conformation, or uniform C3′-endo pucker and little unwinding if it had an A-like conformation. Miller[14] studied deoxyribose tetranucleotide systems, generating a wide range of conformations and evaluating their internal energies using a mixed quantum mechanical and empirical technique. Acceptable models demonstrated the mixed C3′-endo (3′-5′) C2′-endo pucker and fell into two classes with different helical unwinding, consistent with experimental data. Dearing[15] modelled the process of helix extension in a trinucleotide fragment constrained to form part of a continuing helix, and found that a wide range of models could be produced both with common and unusual sugar puckers covering the whole range of helical unwinding angles observed experimentally; he was unable to find a pathway from an unextended conformation to one with adjacent extended sites, and hence violating neighbour exclusion, regardless of the puckers or helix unwinding adopted. Berman,[16] who used an interactive graphics display to con-

struct intercalation models for dinucleoside phosphate systems, has suggested that the choice of sugar pucker is, in this case, not in itself critical.

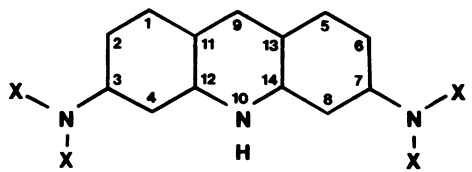
2. MOLECULAR MECHANICS

The calculations we describe here were carried out using a general purpose set of computer programs[17] which enabled us to construct models of the molecular systems in desired conformations and to refine these models automatically using a calculated potential energy as a measure of optimality. The energy is calculated using a set of empirical terms which represent bond-stretching and bending, dispersion interactions,etc:

$$E = \sum_{bonds} k_b(r-r_b)^2 + \sum_{angles} k_a(\theta-\theta_a)^2 + \sum_{dihedrals} \frac{k_d}{2}\left[1+\cos(n\phi-\gamma)\right]$$

$$+ \sum_{non-bonded}\left[B_{ij}r_{ij}^{-12} - A_{ij}r_{ij}^{-6} + \frac{q_i q_j}{\epsilon_{ij} r_{ij}} \right]$$

where the non-bonded terms are summed over all atom-atom pairs, i and j, other than those involved in 1-2 and 1-3 interactions (ie, bonded and next-neighbour atoms). Hydrogen bonding is not modelled



(a) X = Hydrogen : Proflavine

Atom	Type	Charge	Atom	Type	Charge	Atom	Type	Charge
C1	CD	0.1	C2	CD	-0.041	C3	CA	0.238
N3	N2	-0.223	H3	H2	0.140	C4	CD	-0.108
C5	CD	0.1	C6	CD	-0.041	C7	CA	0.238
N7	N2	-0.223	H7	H2	0.140	C8	CD	-0.108
C9	CD	0.178	N10	N*	-0.127	H10	H	0.151
C11	CB	-0.051	C12	CB	0.204	C13	CB	-0.051
C14	CB	0.204						

(b) X = Methyl : Acridine Orange

Atom	Type	Charge	Atom	Type	Charge	Atom	Type	Charge
C1	CD	0.099	C2	CD	-0.040	C3	CA	0.224
N3	N2	-0.125	ME3	C3	0.114	C4	CD	-0.108
C5	CD	0.099	C6	CD	-0.040	C7	CA	0.224
N7	N2	-0.125	ME7	C3	0.114	C8	CD	-0.119
C9	CD	0.178	N10	N*	-0.153	H10	H	0.139
C11	CB	-0.055	C12	CB	0.203	C13	CB	-0.055
C14	CB	0.203						

Figure 1 : Showing residual atomic charges and atom types for proflavine and acridine orange.

Extra parameters needed in addition to those given in reference 19 are as follows: (bonds) CA-CD, CB-CD, CD-CD k_b = 450, r_b = 1.39A, (angles) X-CB-X, X-CD-X, X-N*-X, X-N2-X k_a = 70,θ_a = 120°, (torsions) X-CD-CD-X, X-CA-CD-X, X-CB-CD-X k_d = 25,γ = 180°, n = 2, and (van der Waals) CD A_{ij} and B_{ij} calculated using Slater-Kirkwood parameters r_{vdw} = 1.9, n_{eff} = 6 and α = 2.07.

explicitly, but we find that it is represented adequately by the charge-charge terms between the atoms making up the bonds. Following Hagler et al (18), we do not include the 6-12 van der Waals terms between hydrogen-bonding hydrogens (including the NH_2 and NH groups in the intercalators) and heteroatom (N,O) hydrogen-bond acceptors.

The values we use for the parameters k_a,θ_a,k_b,r_b, k_d,n,ϕ,γ,B_{ij}, A_{ij} and ϵ_{ij} depend on atom type; parameter values and atomic charges for the atoms contained within the nucleic acid will be published elsewhere[19] and the additional terms needed to model the intercalators are given in figure 1 along with atomic charges for these atoms calculated using the CNDO/2 approximation.

Strictly speaking, the energies calculated by this method correspond to gas-phase interactions, which are not, to say the least, the most relevant quantities from a biochemical point of view. Unfortunately, it is not at present practical to model the solvent environment around the complexes adequately, without carrying out Monte Carlo or molecular dynamics calculations in which both solvent and solute are allowed to relax. Monte Carlo calculations have been carried out on complex systems (for example (20)) but these allowed only water coordinates to change. In addition, the technique is expensive and not yet easily applicable. An alternative approach, including only those waters located crystallographically in a conventional molecular mechanics energy refinement, is inappropriate because of the differences both in degree of solvation and degree of refinement of the crystallographic data for the systems considered here. Instead, we attempt to assess the importance of solvation effects, if not to represent them reliably, by performing each numerical refinement with three different force fields:

(1) has a dielectric constant $\epsilon_{ij} = r_{ij}$, the interatomic spacing. This has the effect of reducing electrostatic interactions between groups which are far apart, and which are therefore likely to be separated by water molecules;

(2) has a dielectric constant $\epsilon_{ij} = 4r_{ij}$. This damps electrostatic interactions very strongly, giving the bulk dielectric constant for water at 20; and

(3) has the same dielectric constant as (2) and in addition ignores dispersion (van der Waals) terms between hydrophilic atoms.

The rationale for using a variable dielectric constant is discussed by Hopfinger[21] and for modifying the overall dispersion interaction by Premilat and Maigret.[22]

To some extent, the use of these different force fields avoids the problem of defining precisely how a solvent environment does affect an interaction, whilst allowing us to determine how sensitive our results are to the nature of that environment. In going from (1) to (3), we are modelling an increasingly polarisable solvent.

3. PROCEDURE

We considered the following model structures:

(1) dCpG with a $\mathbf{B}_{(10,3.36)}$ structure[23]

(2) rCpG with a $A_{(11,2.81)}$ structure[24]

(3) proflavine.H^+

(4) acridine orange.H^+

(5) based on Sobell's structure[3] for i^5CpG:proflavine.H^+

(6) based on Sobell's structure[3] for i^5CpG:acridine orange.H^+ and

(7) based on Neidle's structure[4] for dCpG:proflavine.H^+,

ignoring solvent and externally bound dye molecules where given. We also ignored the iodine atoms in models 5 and 6, assuming that structural differences between hydrogenated and iodinated forms would be minimal and well taken care of by the numerical refinement. We generated a total of four complexes for models 5 and 6, and two for model 7, the additional forms having proflavine occupying acridine orange sites or vice-versa (by adding or deleting methyl groups as necessary) and (for models 5 and 6) deoxyribose groups replacing ribose groups (by deleting the hydroxyl groups at atom C2). In addition, a limited number of calculations were performed using Berman's coordinates for proflavine:CpG,[10] to check that no bias was being introduced by using the (less well refined) data of Sobell.

The published crystal structures do not include hydrogen positions. Our molecular mechanics force field, in common with those of some other groups, does not explicitly consider hydrogen atoms bonded to carbon, but does consider those bonded to hetero-atoms. Necessary hydrogen atoms were added to the models, with standard[19] bond-lengths and angles, and, where necessary, the torsional angles were adjusted to values which would facilitate hydrogen bonding or alleviate steric strain. It proved necessary to adjust only the angle C3-C2-O2-H to about 330° in each ribose group at the 5′ end of the helix in rCpG models, in order to direct these hydrogen atoms away from the intercalation site and towards the adjacent C3 hydroxyl.

These models, fourteen in all, were used as starting points for numerical refinement by the molecular mechanics software described above. (Refinement was necessary, not so much because the starting conformations were unreasonable, but more to adjust each of these conformations to a common basis. It would not have been correct merely to base our analyses on the calculated energies of the starting conformations, since the structure determinations were probably subject to different errors and the observed conformations were certainly due partly to different crystal packing and solvation forces which are not represented in our calculations.) The structure of each complex was refined three times using force fields (1) to (3), in each case until the calculation produced a (local) minimum energy conformation. Convergence was ensured by checking that the rms derivative of the energy function with respect to changes in atomic coordinates was less than 0.1 kcal $^{-1}$, and that the relative changes in atomic position and energy value in the last stages of the refinement were small.

4. RESULTS

There were no significant differences between the results obtained from refining models based on Sobell's proflavine:5'CpG and Berman's proflavine:CpG structures, and to simplify the following analysis, we present only the results obtained from the first of these.

Initial and optimised energies for the complexes are shown in table 1. Tables 2 and 3 give a breakdown of the optimised energies in terms of intra- and intermolecular terms, and table 4 lists the important conformational properties of the models, with torsional angle changes of more than 25° and sugar pucker changes marked in bold type. Energies are quoted (notionally) in kilocalories. Figure 3

Table 1 : Calculated energies of model complexes.

Complex	Model	Force field 1	Force field 2	Force field 3
dCpG	1	-68.8	30.5	86.4
rCpG	2	-63.2	12.4	65.2
proflavine.H$^+$	3	62.6	68.5	72.3
acridine orange.H$^+$	4	73.1	79.0	82.6
proflavine:dCpG	5	112.9	233.5	313.0
	6	12.1	127.0	199.0
	7	115.0	229.0	304.0
proflavine:rCpG	5	154.0	258.0	343.0
	6	45.3	147.0	222.0
acridine orange:dCpG	5	859.0	871.0	879.0
	6	32.2	148.0	220.0
	7	174.0	289.0	365.0
acridine orange:rCpG	5	8700	8810	8900
	6	71.0	173.0	250.0

(a) Initial energies

Complex	Model	Force field 1		Force field 2		Force field 3	
dCpG	1	-123.7		-22.9		33.1	
rCpG	2	-106.0		-25.6		36.7	
proflavine.H$^+$	3	24.9		31.0		34.8	
acridine orange.H$^+$	4	43.0		48.8		52.4	
proflavine:dCpG	5	-167.0	(-68.2)	-36.5	(-44.6)	45.2	(-22.9)
	6	-161.8	(-63.0)	-37.0	(-45.1)	46.3	(-21.8)
	7	-164.4	(-65.6)	-36.9	(-45.0)	45.4	(-22.7)
proflavine:rCpG	5	-147.0	(-65.9)	-32.7	(-38.1)	53.7	(-17.8)
	6	-142.0	(-60.9)	-33.9	(-39.3)	55.2	(-16.3)
acridine orange:dCpG	5	-133.6	(-52.9)	-14.8	(-40.7)	61.8	(-23.7)
	6	-136.7	(-56.0)	-17.7	(-43.6)	59.0	(-26.5)
	7	-135.3	(-54.6)	-17.6	(-43.5)	59.8	(-25.7)
acridine orange:rCpG	5	-111.0	(-48.0)	-8.7	(-31.9)	73.1	(-16.0)
	6	-118.0	(-55.0)	-14.4	(-37.6)	67.6	(-21.5)

(b) After refinement. Figures in parentheses for each complex A:X indicate the complexation energy E(A:X) - (E(A)+E(X))

Table 2 : Breakdown of optimised energies into group:group terms.

Complex	Model	Nuc-nuc	Dye-dye	Nuc-dye	Force field
proflavine:dCpG	5	-99.7	26.4	-93.7	1
		2.5	31.8	-70.8	2
		36.9	35.1	-26.7	3
proflavine:dCpG	6	-93.9	25.7	-93.6	1
		4.1	31.2	-72.4	2
		36.4	34.9	-24.9	3
proflavine:dCpG	7	-97.2	25.7	-92.9	1
		3.6	31.2	-71.8	2
		36.5	35.0	-26.1	3
acridine orange:dCpG	5	-91.2	44.2	-86.6	1
		8.6	49.9	-73.2	2
		37.4	52.7	-28.3	3
acridine orange:dCpG	6	-92.7	43.4	-87.4	1
		5.8	49.1	-72.6	2
		37.2	52.5	-30.7	3
acridine orange:dCpG	7	-91.6	43.6	-87.3	1
		6.5	49.2	-73.4	2
		37.6	52.6	-30.3	3

shows best fit superimpositions of the complexes proflavine:rCpG (uniform C3′-endo pucker) and acridine orange:rCpG (uniform and mixed puckers) before and after refinement.

It can be seen that the only structures which are very unfavourable energetically before refinement are those in which the acridine orange has been placed in a nucleic acid model with a uniform C3′-endo sugar pucker. Similarly, the unrefined models based on the acridine orange crystal structure have a somewhat lower energy than the alternative models. The reason for this is simple and obvious: the proflavine crystal structure has a particularly compact intercalation site (the distance between phosphorus atoms in the two strands is about 15.3), and it is not possible to fit acridine orange (with its bulky methyl groups) into the site without causing undue steric strain. Likewise, the acridine orange crystal structure has a large space available for the intercalator (P-P distance about 17.3), and there is no difficulty in inserting the dye. Neidle's[4] mixed structure is intermediate between these two extremes.

After optimisation, the energy values for each model complex become much more similar. Relatively little structural change occurs in complexes based on model 6 or 7; those based on model 5 (uniform sugar pucker) change somewhat more, with the acridine orange:dCpG and rCpG complexes having one sugar repucker. Except for the proflavine complexes refined using force field (2) (dielectric constant $\epsilon_{ij}=4r_{ij}$), the predicted order of stability is (for proflavine) model 5 > model 7 > model 6, and (for acridine orange) model 6 > model 7 > model 5. In other words, the calculations are predicting that the acridine orange complex will preferentially adopt a mixed C3′-endo (3′-5′) C2′-endo pucker and (in two cases out of three) that the proflavine complex will adopt a uniform C3′-endo pucker.

Table 3 : Detailed analysis of group-group interaction for force field 1. Groups are defined in figure 2. Note that the total energy for each complex is obtained by adding the totals (a) and (b) to half the total (c) and then adding the dye-dye terms from table 2.

Group	prof-dCpG Model 5	prof-dCpG Model 6	prof-dCpG Model 7	acr.or-dCpG Model 5	acr.or-dCpG Model 6	acr.or-dCpG Model 7
1	1.14	0.76	0.78	1.20	0.68	0.74
2	-18.88	-14.54	-16.12	-20.29	-13.34	-15.47
3	-6.07	-0.98	-0.71	-0.80	-1.38	-1.27
4	-23.84	-29.23	-29.07	-23.72	-23.91	-23.94
5	0.82	0.70	0.90	0.76	0.61	0.64
6	1.09	0.93	0.95	1.10	1.16	1.14
7	-18.76	-17.19	-17.58	-19.39	-20.51	-20.18
8	-6.16	-7.04	-6.71	-0.88	-3.56	-2.67
9	-23.89	-27.70	-26.04	-25.25	-27.89	-26.93
10	0.82	0.73	0.67	0.64	0.73	0.68
Total	-93.7	-93.6	-92.9	-86.6	-87.4	-87.3

(a) : Interactions between intercalator and nucleic acid

Group	prof-dCpG Model 5	prof-dCpG Model 6	prof-dCpG Model 7	acr.or-dCpG Model 5	acr.or-dCpG Model 6	acr.or-dCpG Model 7
1	0.36	0.36	0.35	0.36	0.38	0.42
2	-3.35	-2.76	-3.21	-2.36	-3.43	-3.33
3	3.60	3.17	3.35	3.51	2.88	2.56
4	-12.42	-14.54	-12.36	-10.37	-15.07	-14.44
5	0.77	0.55	0.71	0.69	0.55	0.56
6	0.36	0.36	0.36	0.36	0.35	0.36
7	-3.51	-2.73	-3.44	-2.54	-3.25	-3.39
8	3.44	2.70	3.13	3.14	3.23	3.36
9	-12.73	-15.09	-13.87	-14.15	-14.21	-13.78
10	0.76	0.55	0.55	0.55	0.55	0.55
Total	-22.7	-27.4	-24.4	-20.8	-28.0	-27.1

(b) : Intra-residue energies within nucleic acid

Group	prof-dCpG Model 5	prof-dCpG Model 6	prof-dCpG Model 7	acr.or-dCpG Model 5	acr.or-dCpG Model 6	acr.or-dCpG Model 7
1	-3.13	-3.04	-2.99	-3.09	-3.28	-3.69
2	-21.67	-22.55	-22.48	-22.03	-22.06	-21.85
3	-16.89	-13.93	-16.26	-15.94	-11.86	-10.62
4	-22.91	-25.72	-22.00	-22.01	-24.15	-23.21
5	-12.59	-2.96	-13.41	-13.26	-2.67	-3.12
6	-3.06	-3.14	-3.00	-3.11	-2.85	-2.85
7	-22.32	-22.33	-22.43	-21.86	-21.55	-22.03
8	-16.66	-13.16	-14.13	-12.13	-12.91	-13.22
9	-22.54	-23.83	-26.16	-24.52	-24.99	-25.27
10	-12.21	-2.38	-2.72	-2.85	-3.04	-2.82
Total	-154.0	-133.0	-145.6	-140.8	-129.4	-128.7

(c) : Total interaction energy of each nucleic acid residue with other nucleic acid residues

Neidle's[4] mixed/uniform model remains intermediate in energy between the other two models in each case, but is always closer in energy to the more stable form. The fact that attempts to refine a model containing acridine orange in a uniform C3′-endo site produce a final structure in which one strand has

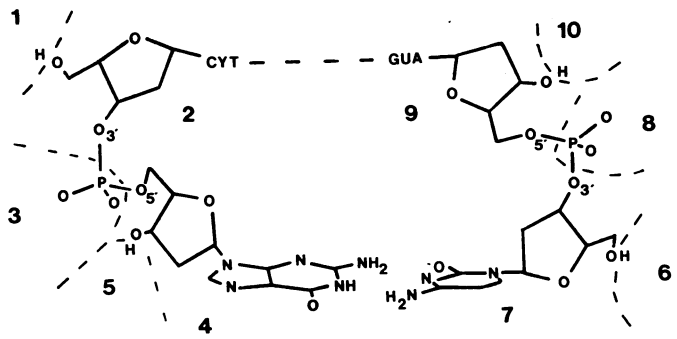


Figure 2 : Defining the groups used in table 3 to break down the interaction energy into its components.

switched to a mixed pucker indicates that the initial geometry cannot relax to accommodate the acridine orange molecule without undergoing a relatively major conformational change. The results obtained with rCpG and dCpG models are similar, although the preference shown by acridine orange for the mixed pucker geometry is more marked with rCpG than with dCpG. It is worth noting, though, that the overall interaction energies obtained with the ribose models are, in all cases, smaller than the corresponding energies obtained with the deoxyribose models. (This may in part be due to the different reference structures used for dCpG and rCpG.) However, in view of the similarity, we shall henceforth limit discussion to the results found for dCpG models.

Examination of table 2, in which each of the refined energies has been broken down into three contributions giving intra-nucleic acid, intra-intercalator and intermolecular terms, shows that variations in the internal energy of the nucleic acid molecule are as important in predicting the preferred conformation as are the intermolecular terms. The relative importance of these two contributions varies depending on the force field, but by breaking down the terms still further, it is possible to see certain interactions which remain critical, regardless of the force field used. This breakdown is given for force field 1 (dielectric constant $\epsilon_{ij} = r_{ij}$) in table 3.

The results indicate that, on the basis of the internal (intra-residue) energies of the residues comprising the nucleic acid, the mixed C3′-endo (3′-5′) C2′-endo conformation is intrinsically more stable than the uniform C3′-endo conformation (-27.4 kcal against -22.7 kcal, the difference coming mainly from the sugar residues at the 3′-hydroxyl end of the dinucleoside phosphates). (Energy values quoted in the text refer to dCpG models refined using force field 1 and are taken directly from table 3.) However, when inter-residue interactions within the nucleic acid are examined, it is seen that these favour the uniform pucker, and the extra stabilisation is seen to be due to terms involving the hydroxyl group at the 3′-terminus and the phosphate oxygens (for example, -16.89 kcal and -12.59 kcal for groups 3 and 5, phosphate and hydroxyl respectively, with model 5 compared with -13.93 kcal and -2.98 kcal with model 6). In other words, hydrogen bonding is occurring between these two groups with the

Table 4 : Conformational properties before and after refinement using force field 1. Angles are defined as follows: ω= P-O3-C3-C4, ϕ= O5-P-O3-C3, ψ= C5-O5-P-O3, θ= C4-C5-O5-P, ξ= C3-C4-C5-O5 and χ= C8(6)-N9-C1-C2.

Angle (degrees)	Model 5		Model 6		Model 7	
	Strand 1	Strand 2	Strand 1	Strand 2	Strand 1	Strand 2
ω	224	220	217	226	210	203
ϕ	294	273	295	284	290	300
ψ	305	323	291	299	290	287
θ	273	206	235	228	219	218
ξ	3	53	58	60	46	72
$\chi_{5'}$	4	19	19	15	15	10
$\chi_{3'}$	103	85	95	90	80	113
ribose$_{5'}$	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*
ribose$_{3'}$	C_3'–*endo*	C_3'–*endo*	C_2'–*endo*	C_2'–*endo*	C_3'–*endo*	C_2'–*endo*

(a) Starting conformations

Angle	Model 5		Model 6		Model 7	
	1	2	1	2	1	2
ω	199	197	200	205	198	197
ϕ	299	300	302	286	304	296
ψ	269	275	289	312	285	292
θ	260	257	250	211	250	240
ξ	47	46	64	70	37	71
$\chi_{5'}$	13	14	19	20	27	18
$\chi_{3'}$	88	87	105	91	86	107
ribose$_{5'}$	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*
ribose$_{3'}$	C_3'–*endo*	C_3'–*endo*	C_2'–*endo*	C_2'–*endo*	C_3'–*endo*	C_2'–*endo*

(b) proflavine:dCpG complexes after refinement

Angle	Model 5		Model 6		Model 7	
	1	2	1	2	1	2
ω	209	209	210	196	221	195
ϕ	315	314	296	310	285	313
ψ	279	286	298	279	307	274
θ	259	239	224	247	192	255
ξ	19	48	59	58	52	54
$\chi_{5'}$	10	13	18	13	27	14
$\chi_{3'}$	95	110	97	106	66	115
ribose$_{5'}$	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*	C_3'–*endo*
ribose$_{3'}$	C_3'–*endo*	C_2'–*endo*	C_2'–*endo*	C_2'–*endo*	C_3'–*endo*	C_2'–*endo*

(c) acridine orange:dCpG complexes after refinement

C3′-endo conformation, which is not possible with the mixed pucker conformation.

There has been speculation[3] for some time about the role that hydrogen bonding between the amino-groups in compounds like proflavine and the phosphate groups in DNA might play in stabilising an intercalation complex. Certainly, a uniform C3′-endo pucker conformation has a smaller intercalation site (allowing substituents at positions 3 and 7 on the intercalated acridine to approach more closely to the nucleic acid phosphate) than the mixed pucker conformation. Examination of the energy components between proflavine and the phosphate oxygens in the uniform C3′-endo conformation shows that there is a considerable interaction between these groups (-6.07-6.16 = -12.23 kcal for proflavine:dCpG in model 5 compared with -0.98-7.04 = -8.02 kcal for proflavine:dCpG in model 6).

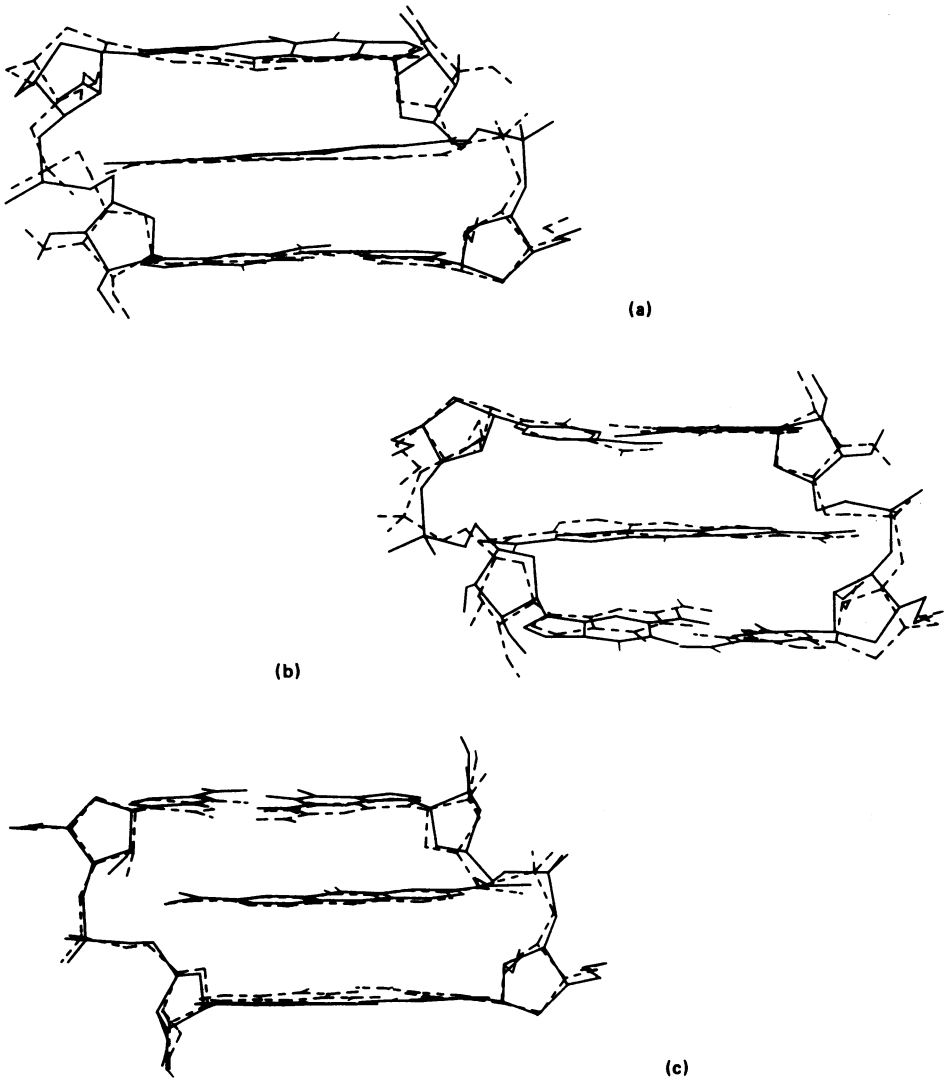


Figure 3 : Showing initial (solid) and refined (dashed) structures for (a) proflavine:rCpG, (b) acridine orange:rCpG both starting with uniform C3'-endo sugar pucker, and (c) acridine orange:rCpG starting with mixed C3'-endo (3'-5') C2'-endo pucker, using force field 1.

However, this stabilisation is compensated for by a decrease in base-intercalator overlap (-18.88-23.84-18.76-23.89 = -85.4 kcal for the former and -88.7 kcal for the latter) and it is seen that (with a dielectric constant $\epsilon_{ij} = r_{ij}$, at least) the overall intermolecular interaction term is almost independent of geometry (-93.7 kcal against -93.6 kcal).

5. DISCUSSION

It is probably important to emphasise once again that we are not proposing that the structures produced by optimisation, starting from crystallographically-determined coordinates, should be regarded as superior to the starting structures, nor that the typical change in energy of about 200-300 kcal mol^{-1} on refinement should be taken to indicate an inherent instability of this order of magnitude in the published structures. The existence of crystal packing forces, counter-ions, etc, in the crystal structures will, in any case, modify the stability and conformation to some extent. The force field used in our calculations has been set up to model the behaviour of nucleic acid systems in a typical polar environment and appears to predict relative stabilities correctly (see also reference (19) which is concerned with the stabilities of different DNA sequences, and reference (25) on A- versus B-DNA). In view of the possible role of crystal forces, and the simplicity of our force field, differences between the crystallographic and our refined dihedral angles (table 4) of the order of 25° are certainly not unreasonable.

On the basis of the analysis given above, we conclude that the mixed C3′-endo (3′-5′) C2′-endo conformation is the intrinsically more stable geometry for intercalation sites: this conformation has a somewhat larger binding site, allowing intercalation of more bulky compounds, has a lower energy for the residues making up the nucleic acid, and its greater unwinding leads to a larger stabilisation of the complex from stacking interactions between the intercalator and the base-pairs.

However, the preference for this geometry is not great, and is capable of being overcome by two types of hydrogen-bonding interaction. Firstly, suitable disposition of amino-groups on the intercalator can allow hydrogen bonding between these groups and phosphate, provided that the nucleic acid adopts a uniform C3′-endo sugar pucker. The extent of the stabilisation introduced in this way is approximately equal to the loss of stacking interactions between base-pairs and intercalator. Secondly, the ability of the terminal hydroxyl groups to undergo specific hydrogen-bond interactions can stabilise the uniform pucker conformation significantly. Our calculations have produced models in which this hydrogen bonding has been intramolecular, with the phosphate oxygens, but we can perhaps generalise this conclusion, especially in view of Neidle's work,[5] to say that a suitably ordered water cage around the complex can fulfill the same function as the nucleic acid phosphate in stabilising one conformation relative to another.

It is very illuminating that the one case where the calculations do not appear to give correct predictions is with the proflavine complexes modelled with force field 2 (dielectric constant $\epsilon_{ij} = 4r_{ij}$). As noted above, the two factors which stabilise the proflavine complex with model 5 are proflavine-phosphate and intra-nucleic acid hydrogen bonds. Increasing the dielectric constant reduces these interactions, since hydrogen bonding terms are modelled using an electrostatic approach, and thus preferentially stabilises the model 6 complex. Changing from force field 2 to force field 3 reduces the stacking contribution, which had preferentially favoured intercalation into model 6. Force field 3 therefore gives the same qualitative prediction as force field 1, because we have reduced both hydrogen-bonding (favouring model 5) and stacking (favouring model 6) interactions.

Finally, we note that these results demonstrate once again the care which should be taken before results obtained from dinucleoside phosphate complexes are used to predict the properties of polynucleotide complexes. Since our analysis has shown the crucial role played by terminal hydroxyl groups in determining the stable conformation, due allowance must be made when predicting the properties of those parts of a larger system which lack these "end-effects".

6. SUMMARY AND CONCLUSIONS

To summarise:

(1) Calculations have shown that the mixed and uniform pucker conformations are comparable in energy, but that there is a preference for the mixed pucker geometry with acridine orange, and for the uniform geometry with proflavine.

(2) Factors favouring the mixed geometry are the better stacking interactions obtained between base-pairs and intercalator, the greater intrinsic stability of the nucleic acid in this conformation and the larger size of the intercalation site, minimising steric strain with intercalators such as acridine orange.

(3) The factor favouring the uniform geometry is the better possibility for hydrogen bonding which can compensate for the loss of stacking (but bear in mind the dependence of this on dielectric constant: an increased dielectric constant results in smaller hydrogen bonding terms).

(4) The implication is that the mixed geometry will generally be preferred, except when hydrogen bonding and/or the detailed water structure act against it.

The conclusion that the mixed and uniform pucker conformations are comparable in energy is perhaps trivial, since these are both observed experimentally. However, we believe that the rationalisation of the results in terms of the relative magnitudes of the energy components is new. In particular, we have demonstrated that the important factors which stabilise intercalation complexes with the mixed sugar pucker are the internal energy of the nucleic acid and the stacking interactions between drug and intercalator. Drug-nucleic acid and intramolecular nucleic acid hydrogen bonding interactions can, on the other hand, stabilise the structure with uniform sugar pucker.

We also hope to have demonstrated the usefulness of using energy refinement methods as a complement to model building methods, such as employed by others.[12,13,16] Model building studies determine allowed geometries, but often do not give insight into which of the (perhaps many) is likely to be the most stable. We do not claim to be making precise energy estimates here, but feel that the general agreement with experiment found here and elsewhere lends credibility to the techniques used. It is also encouraging that our calculations based on crystal structures from two independent laboratories give consistent results - the Neidle[5] structure with one chain adopting a mixed pucker and the other adopting a uniform pucker has an energy intermediate between the wholly mixed and wholly uni-

form dinucleoside phosphate complexes of Sobell et al.[3]

What do we have to say at this point regarding the discussion in the literature concerning the sugar pucker and structure at the intercalation sites? Recall that Sobell has favoured[11] a mixed pucker whilst Berman[16] has noted that a uniform sugar pucker is entirely reasonable. Our calculations seem to indicate that the mixed conformation is to be preferred, except when specific hydrogen bonding terms can play a role, and the experimental results of Neidle[5] also indicate an important role of differential solvation in this preference. Since our calculations were done on a dinucleoside phosphate model, and even here the mixed sugar pucker appeared generally more favourable, we expect that intercalation into the interior (ie away from any free ends) of a nucleic acid will also favour this geometry.

An important proviso in the above statements concerns the role of stacking. This favours the mixed pucker for the three-ring chromophores studied here, but will be relatively less important for small two ring intercalators such as 4-nitroquinoline N-oxide,[26] and for substances like daunomycin[27] which cannot stack as effectively between the bases.

There has also been debate on the biological relevance of modelling dinucleoside phosphate complexes or performing crystallographic studies on these systems. We feel that studies on dinucleoside phosphates could be as relevant as those on polynucleotides when one is considering a biological activity near replicating DNA. Obviously, one needs a precise delineation of the similarities and differences between intercalation in the centre and at the ends of double helices, and to know the dependence of each of these on base sequence and on the precise intercalating drug. We are continuing our studies along these lines, focusing on the intercalation of ethidium bromide, 4-nitroquinoline N-oxide and daunomycin, and on the energetic basis of the neighbour exclusion principle.

7. ACKNOWLEDGEMENTS

This work was supported by NIH grant CA25644 to PAK. AD acknowledges leave of absence granted by Shell International Petroleum Company to carry out this work. Calculations, graphical work and preparation of this paper were performed in the departmental Computer Graphics Laboratory, which is supported by USPHS NIH grant RR-1081-03 (principal investigator R. Langridge). Graphics software was written using facilities provided by T.E. Ferrin and R. Langridge.[28]

8. REFERENCES

1. Lerman, L.S., (1961) *J Mol Biol.* **3**,18.
2. Peacocke, A.R., (1973) p. 723 in *The Acridines*, ed. R.M. Acheson, Wiley. (Vol 9 in *The Chemistry of Heterocyclic Compounds*)
3. Sakore, T.D., Jain, S.C., Tsai, C.C., and Sobell, H.M., (1977) *Proc Nat Acad Sci U.S.A.* **74**,188.
4. Shieh, H-S., Berman, H.M., Dabrow, M., and Neidle, S., (1980) *Nucleic Acid Research* **8**,85.
5. Neidle, S., Berman, H.M., and Shieh, H-S., (1980) *Nature* **288**,129.
6. Tsai, C.C., Jain, S.C., and Sobell, H.M., (1977) *J Mol Biol.* **114**,301.
7. Tsai, C.C., Jain, S.C., and Sobell, H.M., (1977) *J Mol Biol.* **114**,317.
8. Wang, A.H.J., Quigley, G.J., and Rich, A., (1979) *Nucleic Acid Research* **6**,3879.
9. Wang, A.H.J., Nathans, J., Marel, G. van der, Boom, J.H. van, and Rich, A., (1978) *Nature* **276**,471.

10. Berman, H.M., Stallings, W., Carrell, H.L., Glusker, J.P., Neidle, S., Taylor, G., and Achari, A., (1979) *Biopolymers* **18**,2405.
11. Sobell, H.M., Tsai, C.C., Gilbert, S.G., Jain, S.C., and Sakore, T.D., (1976) *Proc Nat Acad Sci U.S.A.* **73**,3068.
12. Alden, C.J. and Arnott, S., (1975) *Nucleic Acid Research* **2**,1701.
13. Alden, C.J. and Arnott, S., (1977) *Nucleic Acid Research* **4**,3855.
14. Miller, K.J. and Pycior, J.F., (1979) *Biopolymers* **18**,2683.
15. Dearing, A., "Computer Studies of Molecular Structure and Function," (1978) D.Phil. Thesis, Oxford.
16. Berman, H.M., Neidle, S., and Stodola, R. King, (1978) *Proc Nat Acad Sci U.S.A.* **75**,828.
17. Weiner, P. and Kollman, P.A., (in press 1981) *J.Comput.Chem.*.
18. Hagler, A., Huler, E., and Lifson, A., (1974) *J.Am.Chem.Soc.* **96**,5319.
19. Kollman, P.A., Weiner, P., and Dearing, A., (submitted) *Biopolymers.*
20. Hagler, A. and Moult, J., (1978) *Nature* **272**,222.
21. Hopfinger, A.J., (1973) p. 59 in *Conformational Properties of Macromolecules*, Academic Press.
22. Premilat, S. and Maigret, B., (1977) *J.Chem.Phys.* **66**,3418.
23. Langridge, R. and Wilkins, M., (1960) *J Mol Biol.* **2**,38.
24. Arnott, S., Smith, P.J. Campbell, and Chandrasekaran, R., (1976) p. 411 in *Handbook of Biochemistry and Molecular Biology.* Nucleic Acids vol 2 (3rd Edition)
25. Tilton, R. and Kollman, P.A., (submitted) *Biopolymers.*
26. Winkle, S.A. and Tinoco, I., (1979) *Biochemistry* **18**,3833.
27. Pigram, W.J., Fuller, W., and Hamilton, L.D., (1972) *Nature New Biology* **235**,17.
28. Ferrin, T.E. and Langridge, R., (1980) *Computer Graphics* **13**,320.