Interaction of berenil with the *tyrT* DNA sequence studied by footprinting and molecular modelling. Implications for the design of sequence-specific DNA recognition agents

C.A.Laughton, T.C.Jenkins, K.R.Fox[1] and S.Neidle*
Cancer Research Campaign Biomolecular Structure Unit, Institute of Cancer Research, Sutton, Surrey SM2 5NG and [1]Department of Physiology and Pharmacology, University of Southampton, Southampton SO9 3TU, UK

Received April 11, 1990; Revised and Accepted July 3, 1990

ABSTRACT

We have developed a technique of partially-restrained molecular mechanics enthalpy minimisation which enables the sequence-dependence of the DNA binding of a non-intercalating ligand to be studied for arbitrary sequences of considerable length (> = 60 base-pairs). The technique has been applied to analyse the binding of berenil to the minor groove of a 60 base-pair sequence derived from the *tyrT* promoter; the results are compared with those obtained by DNAse I and hydroxyl radical footprinting on the same sequence. The calculated and experimentally observed patterns of binding are in good agreement. Analysis of the modelling data highlights the importance of DNA flexibility in ligand binding. Further, the electrostatic component of the interaction tends to favour binding to AT-rich regions, whilst the van der Waals interaction energy term favours GC-rich ones. The results also suggest that an important contribution to the observed preference for binding in AT-rich regions arises from lower DNA perturbation energies and is not accompanied by reduced DNA structural perturbations in such sequences. It is therefore concluded that those modes of DNA distortion favourable to binding are probably more flexible in AT-rich regions.

The structure of the modelled DNA sequence has also been analysed in terms of helical parameters. For the DNA energy-minimised in the absence of berenil, certain helical parameters show marked sequence-dependence. For example, purine-pyrimidine (R-Y) base pairs show a consistent positive buckle whereas this feature is consistently negative for Y-R pairs. Further, CG steps show lower than average values of slide while GC steps show lower than average values of rise. Similar analysis of the modelling data from the calculations including berenil highlights the importance of DNA flexibility in ligand binding. We observe that the binding of berenil induces characteristic responses in different helical parameters for the base-pairs around the binding site. For example, buckle and tilt tend to become more negative to the 5'-side of the binding site and more positive to the 3'-side, while the base steps at either side of the centre of the site show increased twist and decreased roll.

INTRODUCTION

Molecules which bind to the minor groove of DNA are of considerable interest, not least for their potential in the development of drug targeting in antiviral and cancer chemotherapy. The search for structural motifs which will permit the rational construction of molecules having high selectivity for an arbitrary DNA sequence will require a clear understanding of the binding phenomenon at the molecular level. A body of relevant information is accumulating from X-ray, nmr and molecular mechanics studies on a number of minor-groove binders, including netropsin (1–5), distamycin (3,6), SN6999 (7), SN18071 (1,3), Hoechst 33258 (3,8–10), NSC57153 (3), DAPI (11) and berenil (3,12–14). In addition, thermodynamic data have been obtained from calorimetric and spectroscopic studies (15). In general, however, these studies have focused on a narrow range of DNA sequences of restricted length and/or sequence variation. Footprinting studies, on the other hand, can yield information on the sequence-specificity of binding for an essentially arbitrary DNA sequence of considerable length (16–19). However, the amount of information which can be obtained from a given site is lower, though recently it has been extended to the calculation of thermodynamic quantities (20,21).

In this paper we present the results of an extended molecular mechanics study on the interaction of berenil (Figure 1) with the minor groove of the *tyrT* promoter DNA sequence. This study aims to bridge the gap between structural methods and footprinting. Berenil was chose for its relatively small size and symmetry, both of which aid the analysis of sequence-dependent variations, and also for the fact that its interactions with DNA have been the subject of several studies (3,12,16). The *tyrT* promoter of DNA sequence was chosen as it is amenable to footprinting analysis and contains a number of discrete binding sites for berenil together with regions for which this ligand

* To whom correspondence should be addressed

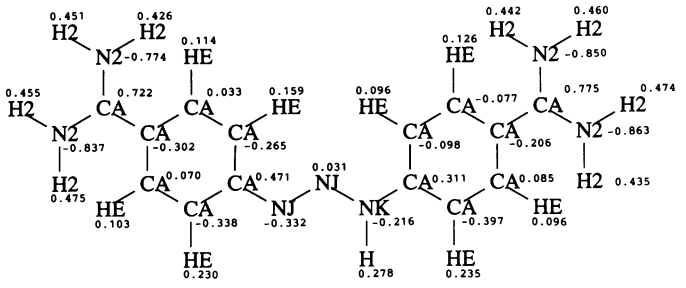

Figure 1. Structure of berenil, showing atom types and partical charges used in the AMBER calculations.

appears to have no particular affinity (16). Berenil binding sites on this DNA fragment have previously been assigned by DNAse I footprinting. In order to compare the footprinting data with molecular mechanics calculations we have performed high-resolution hydroxyl radical footprinting (22) on the interaction of berenil with *tyrT* DNA.

The importance of DNA flexibility in its interaction with other molecules has become evident from a variety of experimental studies. However, in many such studies flexibility can only be inferred, as only distortion of the DNA from some canonical form (23) is directly measurable. The sequence-dependence of DNA distortion has been the subject of several experimental and theoretical investigations; however, in general, these have focused either on small sequences for which there is detailed structural data (24–27) or on larger polymers formed from repeated sequences (28–30).

There is not necessarily a correlation between distortion and flexibility, as has been shown for DNA bending in AT tracts (29). Direct measurements of DNA flexibility have been made using triplet anisotropy decay techniqués (31) and modelled using molecular dynamics (32–34), mechanics (35) and free-energy calculations (36), the last of which have been able to analyse sequence-dependence at the level of individual base-pairs and base-pair steps. Sequence-dependence in flexibility has also been observed in DNA-protein interactions. For instance, Fourier analysis and footprinting studies of the wrapping of DNA around histones suggest that the DNA bends so that at A/T-containing steps the minor groove is directed towards the protein core while at G/C-containing steps the major groove lies innermost (37). Similar conclusions have been reached from a study of DNA bending induced by the catabolite activator protein (38). It was also concluded from this last study that the flexibility of a DNA sequence could be predicted to a good approximation from a knowledge of the flexibility parameters of the constituent dinucleotide steps, in other words, that long-range sequence effects were of minor importance.

As far as the interaction of small molecules with DNA is concerned, it is obvious that DNA flexibility must play a major role in the binding of intercalators, but significant deviations have also been observed between the detailed conformations of DNA sequences in the presence and absence of minor-groove binding molecules (4–11,13). Whilst the requirement for approximate isohelicity remains (39), it appears that the DNA will distort in order to maximise favourable bonding interactions between the inner surface of the minor groove and the ligand.

We present the results of an analysis of the data collected in the molecular mechanics investigation in terms of the variations found in DNA helical parameters, as well as in energetic terms. We have analysed the sequence-dependence of the variation in helical parameters for the energy-minimised DNA alone. We have also analysed the changes in various helical parameters that are induced around the binding site in the presence of a bound berenil ligand. As already noted, the results obtained from such a study do not permit the direct determination of DNA flexibility; however, they do indicate modes of DNA distortion where variations in flexibility with sequence may be expected to contribute most to observed sequence-dependent ligand binding.

EXPERIMENTAL PROCEDURES

Footprinting

The *tyrT* DNA fragment (Figure 6) was isolated and 3′-end labelled using reverse transcriptase and α-[^{32}P]-dATP(*Eco*R1 site, bottom strand) of α-[^{32}P]-dCTP(*Ava*I site, top strand). DNAse I and hydroxyl radical footprinting were performed as previously described (16,17,19,22). The products of digestion were resolved on 8% polyacrylamide gels containing 8M urea, run for 2 hours at 1500V. Gels were fixed in 10% acetic acid, dried under vacuum at 80°C and subjected to autoradiography at −70°C with an intensifying screen. Bands in the digests were assigned by comparison with Maxam-Gilbert dimethyl sulphate-piperidine markers specific for guanine. Autoradiographs of hydroxyl radical-induced cleavage were scanned using a Joyce-Loebl Chomoscan 3 microdensitometer to produce profiles from which the relative intensity of each band was measured. Fractional cleavage at each bond, $f=A_i/A_t$, was calculated, where A_i is the area under band i and A_t is the sum of the intensity of all bands in any gel lane. The data are presented as a differential cleavage plot in the form $\ln(f_{antibiotic})-\ln(f_{control})$ as previously described (17). Negative values indicate regions protected by the ligand. A general trend from positive to negative values is seen moving from the 5′ to 3′ direction of each strand, imposed on the true differential plot, due to small differences in the degree of cleavage in the drug-treated and control lanes.

Molecular modelling

The initial coordinates of the DNA fragments used were those of idealised B-DNA (23) and were prepared using the program GENHELIX (40). The initial coordinates of berenil were taken from the crystal structure (12). All molecular mechanics calculations were performed at the all-atom level using the program AMBER 3.1 (41). Solvent and counterions were not included explicitly for reasons of computational expense. Instead their effect was simulated through the use of a simple distance-dependent dielectric constant with $\epsilon=4r_{ij}$. This formulism is well established in the field of protein modelling and has been tested in some detail for a nucleic acid model system with satisfactory results (42). The use of more complex sigmoidal or exponential functions (36,43) may be superior, but there is a sizeable computational cost. Additional parameters required for berenil were obtained by interpolation (44) and from the crystal data (Table 1). A set of partial atomic charges for berenil was provided by Dr. M. Orozco and was obtained from the MNDO wavefunction via the calculation of the molecular electrostatic potential and a scaling and fitting procedure which has been shown to give charges of a quality close to that obtainable from *ab initio* calculations at the 6–31G* level (45). Interactive molecular modelling was performed using the program GEMINI 1.01 (46). All calculations were performed on an Alliant FX40/3 computer, and visualisations on a Silicon Graphics Iris 3130

workstation. As well as the energy-minimised DNA sequence itself, a total of 338 of the DNA/berenil complexes were also analysed using the program NEWHELIX (47).

It is conventional in footprinting studies to label the inter-base linkage rather than the bases themselves. Since labelling the latter is more useful in the description of the modelling studies we have given the base pairs the same number as that of the bond to the 3′ side on the Watson (top) strand. For modelling purposes, the DNA sequence corresponding to bases T60 to A119 of *tyrT* was chosen. A protocol was developed to permit the construction and minimisation of such a length of DNA from smaller segments in a controlled manner. Construction from small segments has the advantages that the minimisations proceed more quickly, can permit larger, energetically favourable shifts (27) and that additions to or subtractions from the overall sequence may be made with the minimum of recalculation . However, there are problems involving end-effects which must be overcome; for example, on minimisation, the termini of DNA duplexes tend to 'fray'.

The first step in the procedure consisted of the minimisation of an 18-mer duplex corresponding to bases 69−86. The only constraints were imposed on the phosphate P atoms, and these were removed in a step-wise fashion. All minimisations were performed to a RMS gradient of < 0.1 kcal/mol/Å. Bases 69−71 and 84−86 were then discarded, and a second 18-mer was then constructed corresponding to bases 81−98. The initial coordinates of bases 81−83 were taken from the previous minimisation and the rest from idealised B-DNA (23). In the minimisation of this fragment the phosphorus atoms were constrained as before, but in addition the 'belly' option of AMBER was used to restrict the minimisation to bases 84−98 at all stages. After discarding the last three base-pairs, the procedure was repeated twice in the 3′-direction and then once in the 5′-direction. The single 12-mer and the four 15-mer duplexes were then combined by overlaying the common, 'frozen' residues followed by a final minimisation of the 60-mer, with the first and last base-pairs alone restrained (belly option). Although only in this final step were restraints applied to both ends of the DNA sequence simultaneously, helix analysis of the 60-mer following energy minimisation showed no global DNA bending.

The ligand binding studies were performed as follows. The berenil molecule was docked to the minor groove of the DNA using interactive graphics such that it spanned bases 64−66 and the central nitrogen atom of the triazene (-N=N-NH-) unit was located 7.0Å radially from the helix axis. The system was then reorientated so that the helix axis coincided with the z-axis. All subsequent initial positioning of the berenil in the minor groove was performed by translation and rotation of the berenil molecule from this position according to the global helical parameters. All minimisations were restricted, using the AMBER belly option, to berenil and the nine base-pairs surrounding the binding site. A small local modification was made to the program to enable the central nitrogen atom of the berenil triazene unit to be restrained, but only from motion in the z-direction. All these minimisations were performed to an RMS gradient < 0.1 kcal/mol/Å. Typically, each minimisation required about 500 cpu seconds on the Alliant FX40/3.

In mapping the interaction of berenil with the 60-mer, successive initial berenil positions corresponded to 0.5Å shifts along the z- or DNA helical axis. At each point two minimisations were performed, with inversion of the berenil between runs to allow for its asymmetry. The results represent the accumulation of data from a total of 678 minimisation procedures.

Table 1. Molecular mechanics parameterization of berenil for AMBER.

Bond parameters

	K_r	r_{eq}
CA−NJ	480.000	1.43000
CA−NK	480.000	1.40000
H −NK	434.000	1.01000
NJ−NJ	480.000	1.26500
NJ−NK	480.000	1.34000

Angle parameters

	K_Θ	Θ_{eq}
CA−CA−N2	75.000	120.000
CA−CA−NJ	85.000	120.000
CA−CA−NK	85.000	120.000
CA−NJ−NJ	85.000	111.000
NJ−NJ−NK	85.000	111.000
CA−NK−H	35.000	120.000
CA−NK−NJ	85.000	120.000
H −NK−NJ	35.000	120.000

Torsion parameters

	NTOR	V/2	δ	γ
X−CA−NJ−X	2	10.000	180.000	2.000
X−CA−NK−X	2	10.000	180.000	2.000
X−NJ−NJ−X	2	10.000	180.000	2.000
X−NJ−NK−X	2	10.000	180.000	2.000
CA−CA−CA−N2	1	10.600	180.000	2.000

Improper torsion parameters

X−X−CA−CA	0	10.500	180.00	2.000
X−X−N2−CA	0	10.500	180.00	2.000

RESULTS

Footprinting

Figure 2 presents the results of DNAse I and hydroxyl radical cleavage of *tyrT* DNA in the presence and absence of varying concentrations of berenil. The DNAse I digests reveal several sites of reduced cleavage located around positions 130, 110, 85 and 70 on the labelled top strand. Inspection of the sequence reveals that, as expected, each of these corresponds to an A/T-rich region. The results are similar to those seen with other A/T-selective minor groove binding ligands such as distamycin and netropsin (16,17,19) but differ in two respects. Firstly no regions of enhanced DNAse I cleavage are apparent, suggesting that berenil does not induce changes in DNA structure in regions remote from its binding sites. Secondly, at elevated drug concentrations DNAse I cleavage is totally inhibited, suggesting that berenil is able to bind to many other DNA sequences.

The ligand binding sites are more clearly resolved using hydroxyl radicals as the footprinting probe so that, for example, the protection of DNAse I cleavage around position 85 consists of two distinct binding domains. A differential cleavage plot of the hydroxyl radical cleavage pattern is presented in Figure 3. Examination of Figures 2 and 3 between positions 60−120 shows that protection is consistently found towards the 3′-end of A/T-rich regions, though interestingly no footprints are seen on either strand around position 60 (TAA).

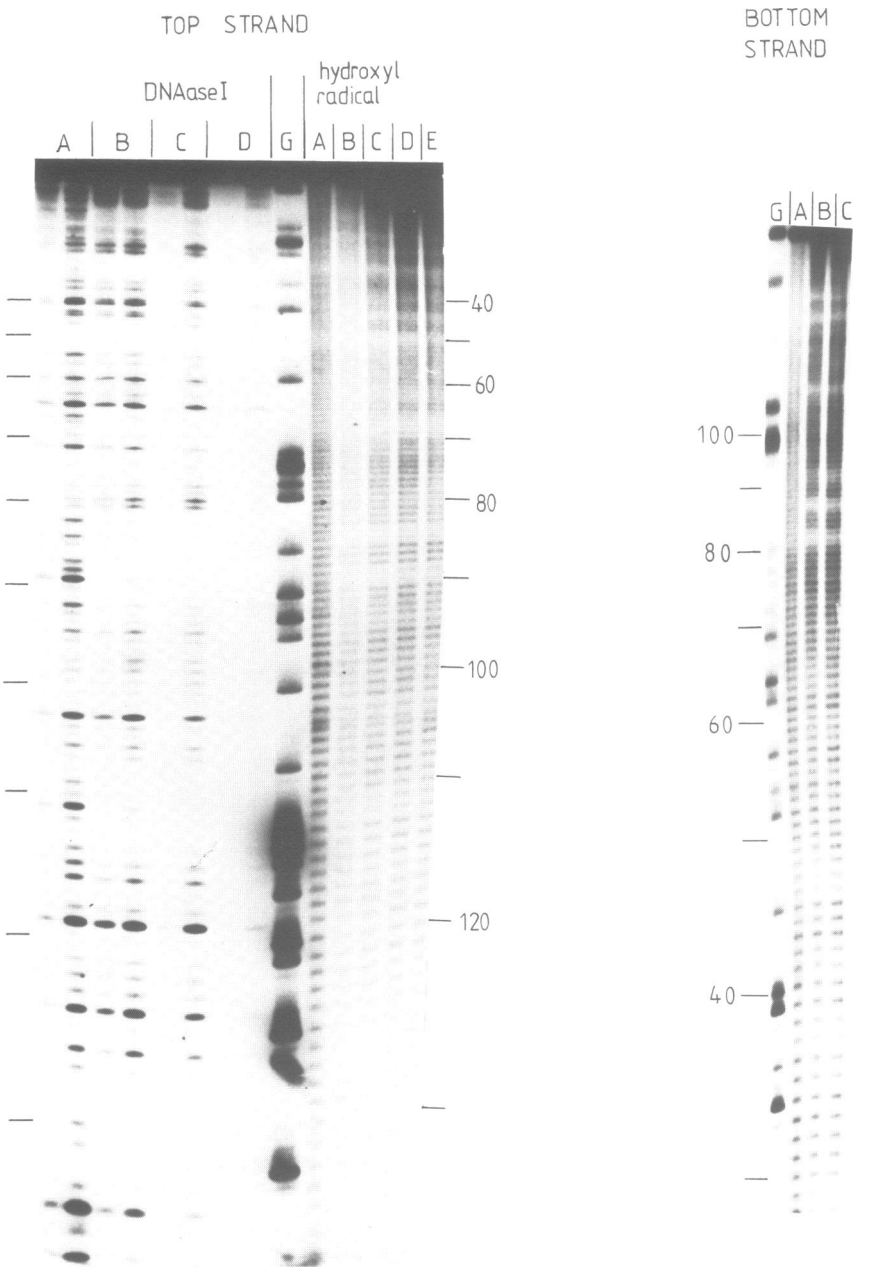

Figure 2. DNAse I and hydroxyl radical footprints of berenil in *tyrT* DNA. The numbers and strand designation correspond to the sequence shown in Figure 5. For DNAse I each pair of lanes corresponds to digestion by the enzyme for 1 and 5 minutes. Berenil concentrations were: A, control (0μM); B, 1μM; C, 5μM; D, 25μM; E, 100μM. The tracks labelled 'G' are Maxam-Gilbert dimethyl sulphate-piperidine markers specific for guanine.

Energetic analysis of berenil binding

The binding energy (E_{bind}) was calculated as the sum of a number of terms:

$$E_{bind} = E_{vdw} + E_q + E_{per}$$

where E_{vdw} and E_q are, respectively, the van der Waals and electrostatic contributions to the interaction energy, and E_{per} is the total induced perturbation energy of the DNA and berenil. Note that E_{bind} is strictly a binding enthalpy; there is no calculation of the entropy changes accompanying binding. This is in contrast to the footprinting experiment, where the results obtained relate to the binding free energy.

The near two-fold symmetry of the berenil molecule was evident from the fact that the difference between the calculated binding energies for the two orientations at a given position averaged only 1.8 kcal/mol. However, on rare occasions it could be much larger (maximum 11.3 kcal/mol); in the subsequent analysis only the best orientation at each position was considered. The variation of E_{bind}, E_{vdw}, E_q and E_{per} with ligand position are shown in Figure 4. The abscissa relates the final z-coordinate of the central nitrogen atom of the berenil triazene unit to that of the mean base-pair planes; the applied restraint ensured that on minimisation this did not alter by more than 0.05Å from its initial value. The total binding energy varies between −35 and −55 kcal/mol. The periodicity in the occurrence of A/T versus G/C rich regions in this portion of the tyrT DNA sequence is mirrored in the general behaviour of the binding energy, being generally greater (more negative) in A/T-rich sequences.

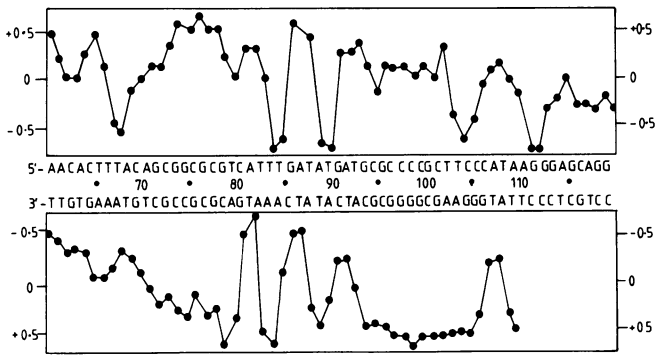

Figure 3. Differential cleavage plot for berenil on the central portion of the tyrT DNA fragment produced by footprinting with hydroxyl radicals. Vertical scales are in units of ln $(f_a) - \ln(f_c)$, where f_a is the fractional cleavage at any bond in the presence of the ligand and f_c the fractional cleavage of the same bond in the control. Negative values, towards the central sequence, indicate blockage.

Examination of the general trends in the variation in E_{vdw} and E_q along the sequence reveals similar periodic behaviour, but the former tends to be greater (more negative) in G/C-rich regions, while the latter tends to be more favourable in A/T-rich regions. The two trends cancel each other out and, with four or five isolated exceptions, the total interaction energy varies over a limited range with little sequence-dependence. The observed general variation in E_{bind} is therefore particularly related to the variation in the total perturbation energy, E_{per}. Since the fluctuations in perturbation energy are dominated by that of the DNA we conclude that the preference for berenil binding in A/T-rich regions stems in no small part from the lower DNA perturbation energies that accompany such binding. We calculated RMS deviations between the DNA structures obtained in the presence and absence of berenil. It was found that such deviations were on average larger in the A/T rich regions. This suggests that the observed reduction in E_{per} in such regions results from increased DNA flexibility in these sequences rather than smaller structural perturbations being required to maximise the binding interaction.

In more detail, we may classify the binding sites according to two criteria. Firstly, they may be divided into classes in which the central nitrogen atom of the berenil triazene (-N=N-NH-) unit lies either approximately in the mean plane or base-pair, so that the berenil molecule as a whole spans three base pairs (class 1), or between two consecutive mean base-pair planes, so the molecule as a whole spans two and two half bases (class 2, see Figure 5). Secondly, they may be divided into 'good' ($E_{bind} \geq -48$ kcal/mol) sites and less good ($E_{bind} < -48$ kcal/mol) binding sites. The average values of the energy components and the RMS deviations for the different classifications are listed in Table 2.

From this we note that, over all sites, class 2 sites have on average a better E_{bind} than class 1 sites by 3.4 kcal/mol, to which differences in E_{vdw}, E_q and E_{per} contribute approximately equally. This generally improved binding energy is reflected in the relative abundance of good binding sites in both classes. In accordance with the trends observed, E_{per} is lower for good sites than the average value, but the RMS deviations are considerably greater, pointing to the importance of DNA flexibility in berenil binding.

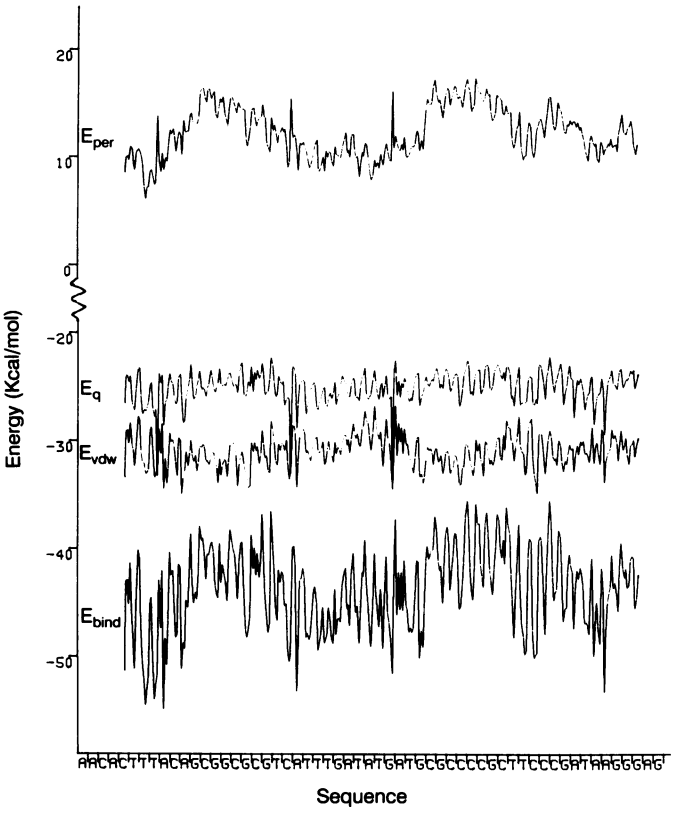

Figure 4. Variation in total binding energy (E_{bind}) and its van der Waals (E_{vdw}), electrostatic (E_q) and perturbation (E_{per}) components, for interaction of berenil with a portion of the *tyrT* DNA sequence.

Comparison with footprinting results

In Figure 6 the positions of regions most clearly protected from hydroxyl radical-induced cleavage in the footprinting study are compared with positions in the sequence having berenil binding sites with an interaction energy ≥ -48 kcal/mol. It can be seen that the overall agreement is good. The footprints are staggered across the DNA strands in the 3'-direction as previously noted (17,22). The only anomaly is the absence of any protection on the bottom strand around bonds 63–65.

Helix analysis of the minimised 60-mer duplex

Because of program limitations, the helix analysis of the 60-mer was restricted to a 48 base pair stretch from C65-A112. The analysis concentrated on the base pair parameters buckle and propeller twist, and the local parameters slide, rise, tilt, twist and roll.

The variation in buckle is shown in Figure 7a. The sequence-dependence is quite simple; pyrimidine-purine (Y-R) base pairs shown a positive buckle of about 5°, whereas R-Y pairs show a negative buckle of a similar magnitude. From the propeller twist data (Figure 7b) there is no evidence that A-T base pairs show greater twists than G-C pairs, but one noticeable feature is the increase in propeller twist towards the centre of runs of pyrimidines (C97–C100, C102–C107). Variations in slide (Figure 7c) are in general rather small, but the consistently low values for CG steps are of note. Rise also shows only slight variation (Figure 7d), and now it is GC steps that show consistently small values; also there are indications of greater than average values of rise for TA steps. Variations in tilt are

in general very small and show no obvious sequence-dependence (results not shown). Twist (Figure 7e) shows greater variation: again, patterns are difficult to discern but AT steps show a consistently low value. For roll more trends are evident (Figure 7f): AT and AG(CT) steps correspond with minima in roll, whilst GC and TC(GA) steps correspond with maxima.

Helix analysis of the response to berenil binding

In view of the important contribution of DNA flexibility to the A/T preference of berenil binding, variations in the helix parameters were investigated in terms of their general response to such binding. Helix parameters for the same 48 base-pair segment of the 60-mer were calculated at each bound ligand position. After subtraction of the values for the unligated DNA, the results were analysed in terms of the average change in the parameter in relation to the berenil binding site. The analysis has been restricted to buckle, propeller twist, tilt, twist and roll, since these parameters show the greatest deviations. The results are presented in Figures (8a−8e): for class 1 sites the berenil molecule is centred over base pair 0, for class 2 sites it is centred between bases −1 and +1.

Buckle (Figure 8a) is observed to become more negative to the 5′-side of the binding site and more positive to the 3′-side, with increased deviation apparent at 'good' binding sites. For propeller twist (Figure 8b) we observe that berenil binding tends to produce a marked negative change (increase) in this parameter at the centre of the berenil binding site, particularly in class 2 sites. The difference between the change observed in good sites and that averaged over all sites is, however, only marked at bases more remote from the binding site centre. The influence of binding upon tilt (Figure 8c) is not well defined, though it seems that there is a tendency for negative induced deviations to the 5′-side and positive ones to the 3′-side. For twist (Figure 8d), we observe two regions of positive deviation located either side of the centre of the binding site, though the pattern for class 1 sites is markedly asymmetric. Only for this class are the differences in induced perturbation markedly greater than the average in good sites. Roll (Figure 8e) shows the reverse pattern, with two regions either side of the centre of the site showing particularly large negative deviations upon interaction with berenil.

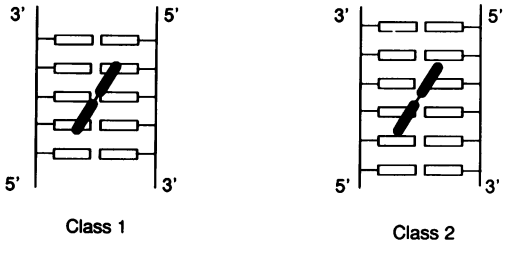

Figure 5. Schematic representation of the two classes of binding site of berenil in the minor groove of DNA.

DISCUSSION

Comparison between footprinting data and molecular mechanics energies

In this study, for the first time, the results of a molecular mechanics study can be compared directly with those obtained from a footprinting experiment. The agreement is in general good, despite the simplicity of the modelled system (e.g. no explicit consideration of solvent or counter-ions, and ignorance of entropic factors). The few 'anomalous' results observed (apparent fairly good binding sites in G/C-rich regions) all seem to involve the partial expulsion of the berenil molecule from the groove and strong electrostatic interactions between one of the amidinium groups and a backbone phosphate. This type of interaction was only observed by Caldwell and Kollman (2) in their study of netropsin-DNA interactions. It may well be an artifact, the result of the limitations of the distance-dependent dielectric constant approximation as a representation of the effects of solvent on electrostatic interactions.

In broad terms, the preference for binding in A/T-rich regions is observed to stem from the improved electrostatic interactions in such areas, coupled with lower induced DNA perturbation energies. That electrostatic factors would favour binding of positively-charged ligands to the minor groove in A/T-rich regions has been the conclusion of other studies. Thus, *ab initio* calculations have revealed a more negative electrostatic potential in the minor groove of A/T sequences compared with G/C ones (1,48,49). Although the question of the sequence-dependence of DNA conformation has been the subject of many investigations, fewer studies have concentrated on the sequence-dependence of DNA flexibility. Sarai *et al* (36) have used conformational free energy calculations to show that the flexibility of DNA varies in a complex manner both with the helical parameter being measured and the base-sequence, but the generally greater flexibility of A- and T-containing steps seems to be a feature.

At the level of greatest detail, the binding energy plots reveal that even within favourable regions, the enthalpy wells can be very sharp; and there can be more than one local energy minimum over the base-pair step. From the relative widths of the minima it appears that helical motion of the berenil along the minor groove should be particularly restrained at binding sites containing alternating A/T sequences, and less so at sites containing A_n/T_n stretches.

In previous studies of berenil-DNA interaction, models postulating a binding site covering either two or three base-pairs

Table 2. Analysis of averaged binding energies and component energies (kcal/nol) by classification of binding site.

Classification[a]	No.	E_{vdw}	E_q	E_{per}	E_{int}	E_{bind}	RMS
All sites	339	−31.0	−25.2	−12.4	−56.2	−43.8	0.069
All good sites	52	−32.6	−27.4	9.9	−60.0	−50.1	0.096
All class 1 sites	171	−30.6	24.5	12.9	−55.0	−42.1	0.063
All good cl.1 sites	14	−33.0	−27.8	10.7	−60.7	−50.0	0.097
All class 2 sites	168	−35.1	−25.9	11.9	−57.4	−45.5	0.074
All good cl.2 sites	38	−32.5	−27.1	9.6	−59.7	−50.1	0.096

a: 'Good' sites have $E_{bind} \geq -48$ kcal/mol (see text).

have been presented. Whilst the class 1 site does indeed correspond with the three base-pair spanning model presented previously (13), in class 2 sites the angle between the berenil axis and the helix axis is similar to that in class 1 sites, such that the molecule extends beyond both of the two base pairs which define the site. In contrast, the two base-pair models (3,12,14), require that the berenil makes a greater angle with the helix axis and lies somewhat across the line of the minor groove in order to maximise hydrogen-bonding interactions between the amidinium groups and acceptor groupings on the spanned bases. However, this has a marked effect upon the degree to which the minor groove can narrow. From the energetic analysis of our models it appears that generalised electrostatic interactions are more important than specific hydrogen bond formation in stabilising the complexes.

It is interesting to compare these models with the recent X-ray structure of the complex between berenil and d(CGCGAATTCGCG) (13). The drug binds asymmetrically but close to the centre of the sequence, i.e. in an approximate class 2 site. However, only one end of the drug lies deep in the minor groove, and at the other end of the molecule the amidinium group interacts with the DNA via a bridging water molecule. This feature has also been observed in the X-ray structure of one other minor groove ligand (11). This result points out an obvious limitation of the present method; as the solvent is not included explicitly in the calculations no such binding mode could be predicted.

The structure of the minimised *tyrT* fragment

Of the helical parameters analysed for this fragment, that showing the clearest correlation with base sequence is buckle. Without exception Y-R base pairs show a positive buckle whilst R-Y base pairs are characterised by a negative buckle. Although the source of this trend remains obscure, a qualitatively similar pattern of behaviour for this parameter has been observed experimentally, for instance in the crystal structure of the dodecamer d(CGCGAATTCGCG) duplex (50).

The observation that propeller twists were generally no higher for A-T base pairs than G-C ones was unexpected, as the inherently greater hydrogen bond involvement in G-C base-pairing might be predicted to increase the energy penalty accompanying propeller twisting. However, the results are in accord with those of Sarai *et al* (36) who found that it was the flexibility, rather than average value, of propeller twist which

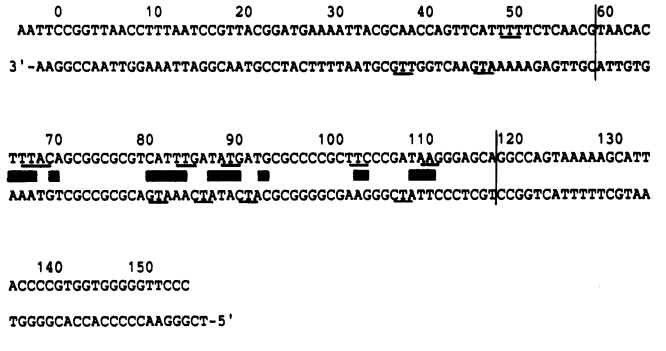

Figure 6. Correspondence between inter-base linkages found to be protected in the footprinting study (underlined) and areas predicted to contain good berenil binding sites from the modelling study (horizontal bars). The limits of the portion of the *tyrT* sequence that was modelled are indicated by vertical lines.

was greater for A-T base pairs. The increase in propeller twist towards the centre of homopyrimidine-homopurine runs (C97–C100,C102–C107) may be explained in terms of reduced steric clash between base pairs that will accompany concerted movements in such sequences (51).

The origin of the low values of slide for CG steps remains obscure, but the observation that GC steps show low values of rise may be linked in part to the variation in buckle. If two consecutive base pairs buckle towards each other, as GC steps do, this will reduce the mean base pair plane separation. However, this can only be a partial explanation, as the effect is not particularly noticeable for other RY steps, though the behaviour of TA steps (somewhat greater than average rise) provides some support for this.

The observation of particularly small values of twist for AT steps is of interest in relation to the well-studied behaviour of $(AT)_n$ tracts (52). Such sequences may adopt an abnormal conformation in which the DNA is underwound (53). Our results suggest that such underwinding may be expected even in the absence of large conformational changes in the backbone.

The correlations we observe in the sequence-dependence of roll do not match those found by Sarai *et al* (36), who calculated that CG steps in particular should show larger than average values of roll. However, they noted that the behaviour of tilt was strongly dependent on the electrostatic parameters. In their calculations

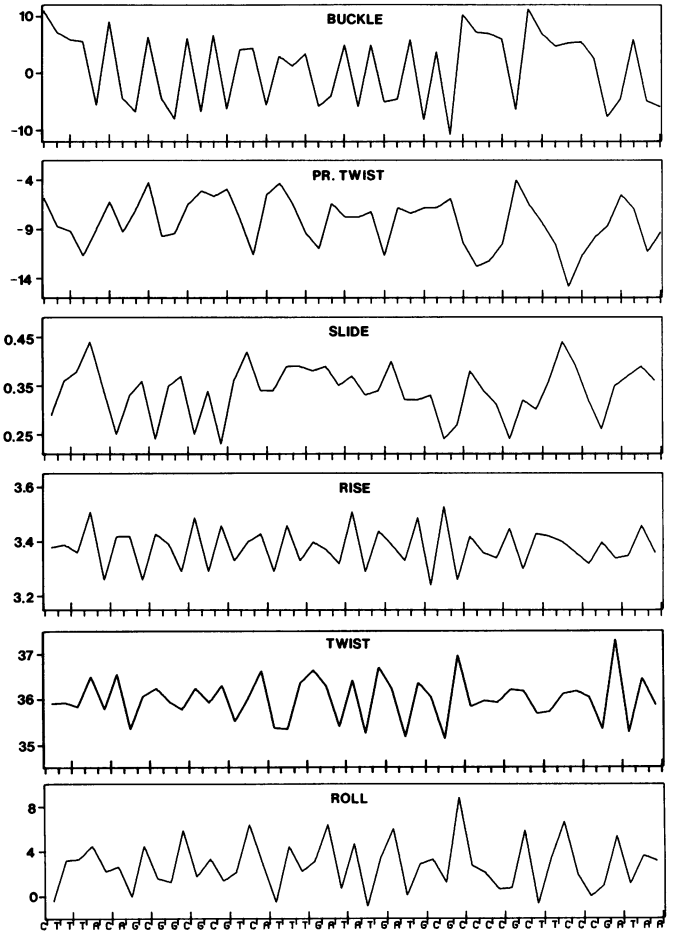

Figure 7. Helical parameters for the DNA minimised in the absence of berenil: a) buckle, b) propeller twist, c) slide, d) rise, e) twist and f) roll.

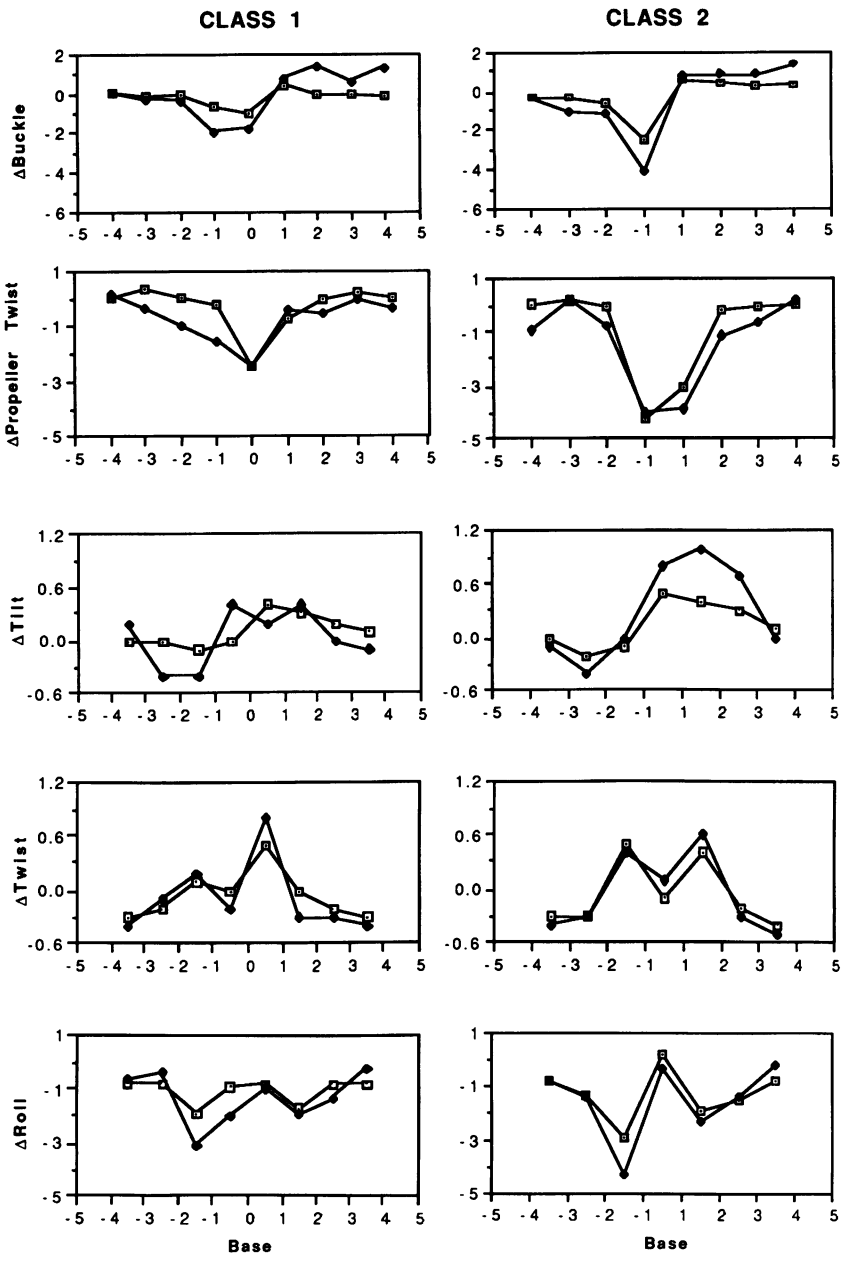

Figure 8. DNA perturbations induced by berenil binding for class 1 (left) and class 2 (right) binding sites (see text for definitions). Averages over all sites ( ) and for 'good' binding sites (binding energy > −48 kcal/mol). Diagrams show: a) buckle, b) propeller twist, c) tilt, d) twist and e) roll.

they made use of an exponential shielding function to simulate the effect of solvent on electrostatic interactions, whereas in our study a simple distance-dependent form of the dielectric constant was used.

Response of helical parameters to berenil binding

Analysis of the effect of berenil binding on the local DNA conformation reveals that trends in the behaviour of different helical parameters are to an extent sequence-independent, relating more to the relative dispositions of base pair and berenil. However, it should be appreciated that the results obtained here are from data encompassing only a fraction of the possible sequences that could constitute a binding site.

Buckle is observed to become more negative to the 5'-side of the binding site and more positive to the 3'-side. This corresponds to the base pairs distorting away from the centre of the site and may serve (i) to reduce steric clashes between the bases ad the inward-facing hydrogens of the berenil phenyl groups, and (ii) to facilitate hydrogen-bonded interactions involving the bases ad the terminal amidinium groups of the ligand. Deviations are greater at both 'good' binding sites and class 2 sites, presumably reflecting the balance of forces at the central base pair in class 1 (3 b.p.) sites. Similarly, the deviations observed for propeller twist are greater for class 2 sites than for class 1 sites. In contrast to buckle, however, deviations in propeller twist appear to be symmetrical with respect to the centre of the site. The observed increases in propeller twist may permit deeper penetration of the berenil into the minor groove and so enhance the overall interaction as well as improving specific hydrogen-boding interactions between the bases and the berenil amidinium groups.

It is apparent that the maximum distortions induced do not differ greatly between good and less good sites. The main difference, particularly noticeable for class 1 sites, is that deviations in propeller twist are greater for bases more remote from the centre in 'good' (energetically favoured) binding sites.

The behaviour of twist is complicated by the fact that as only nine base-pairs are free to move in each minimisation, the total helical twist from the beginning to the end of each 9-mer is constant and hence deviations induced at any step must be compensated for elsewhere. For class 1 sites the behaviour of twist is markedly asymmetric, with a large average increase in twist for the base step 0 to 1, and a smaller increase for the step −2 to −1. In contrast, for class 2 sites, deviations in twist are symmetric with respect to the centre of the binding site; it appears that increases in twist for the three central base pair steps are compensated for by reductions in twist towards the ends of the binding site. This behaviour may represent a process in which the minor groove is narrowed around the phenyl and amidinium groups of the berenil ligand, leading to improved binding interactions.

The behaviour of tilt is not well-defined, but appears to be reduced to the 5′-side of the binding site and increased to the 3′-side. This behaviour may be interpreted as a tendency of those base pairs closest to the amidinium groups to rotate towards the plane of the berenil. This may be in response to the tendency of the berenil to lie slightly across the minor groove, with its 5′-end closer to the 5′-3′ strand and its 3′-end closer to the 3′-5′ strand.

Either side of the centre of the binding site, roll is observed to become more negative. Since the initial conformation shows on average a small positive roll, this corresponds to a move towards closer co-planarity of successive base pairs. The effect is to move the edges of the base pairs that are in the minor groove towards the centre of the binding site. This may be in response to steric clashes with the inner edges of the phenyl rings of the berenil, but would appear to be in conflict with the behaviour of buckle, where movement of the base pairs away from the centre of the binding site is observed.

Sequence-specific ligand design

This study raises implications for the design of agents which bind to DNA in a sequence-specific manner. A major target of many efforts in this area is the design of ligands with good G/C-recognition characteristics. Success in this area to date has stemmed from the joint application of two strategies.

The first strategy has been to reduce the ligand charge. Monocationic ligands will bind more weakly than equivalent dicationic ligands to DNA as the size of E_q will be reduced. However, as trends in E_q lead to A/T selectivity, reducing its contribution to E_{bind} may be effective, but only if the other contributions to E_{bind} show a different sequence selectivity. Our results suggest that this is true for E_{vdw}, which favours G/C-binding, but not for E_{per}. The second strategy has been to design ligands that will show active recognition of the guanine N2 group through, for example, hydrogen bonding. Our results here suggest that the active avoidance of steric clashes with this group and others, as opposed to positive interaction, may be the more important factor. Clearly, in the design of G/C-recognition elements, a lack of DNA flexibility means that careful attention will have to be paid to the accurate matching of groove and ligand surfaces, whereas a much cruder match may suffice for A/T-recognition.

Our results suggest that charge reduction (54), by increasing the relative contribution of E_{vdw} to E_{bind}, should be particularly effective. However, such a modification is by itself clearly not sufficient, as evidenced by the A/T-selectivity of, for example, distamycin (6). Apart from general electrostatic differences, the most obvious difference between the minor groove in G/C as opposed to A/T sequences is the presence of the guanine N2, and particularly its attached hydrogen, in the groove floor. This moiety has been expected to inhibit deep penetration of the ligand into the groove and hence reduce both electrostatic and van der Waals interactions. However, our results here, in particular the behaviour of E_{vdw}, suggests a slightly different explanation. Because of the inherently lower flexibility of G/C-rich regions, steric clashes between a ligand and, for instance, the N2-hydrogens, will not be readily moderated through DNA distortion. As a result perturbation energies will be higher and binding affinities lower. However, through careful ligand design, the guanine N2 group may be persuaded to play a positive role in increasing G/C selectivity. This idea, combined with the effect of the reduction in ligand charge, is evident in the enhanced G/C sequence-dependent DNA binding of the monocationic imidazole lexitropsins (55).

The importance of DNA flexibility in the sequence-dependence of ligand binding is becoming clear. In order to explain, predict and control sequence-specific DNA recognition it will be necessary to combine knowledge of the sequence-dependence of DNA flexibility with an understanding of the strategies of structural deformation adopted by DNA sequences in response to the introduction of a ligand. The results of this study suggest that these basic strategies of DNA structural deformation are in broad terms sequence-independent, but that DNA flexibility, which *is* sequence-dependent, determines to what extent these strategies can be implemented. Several questions remain, for example which modes of DNA distortion are the most important for the enhancement of the binding interaction. Another issue which must be addressed is the extent to which the modes of DNA distortion are ligand-independent, only then will it be possible to establish what predictive power the analysis may have.

ACKNOWLEDGEMENTS

We are grateful for grants from the Cancer Research Campaign (to SN) and the Medical Research Council (to KRF). KRF is a Lister Institute Research Fellow.

REFERENCES

1. Zakrzewska, K., Lavery, R. and Pullman, B. (1983) *Nucleic Acids Res.*, **11**, 8825−-8839.
2. Caldwell, J. and Kollman, P. (1986) *Biopolymers*, **25**, 249−266.
3. Gago, F., Reynolds, C. A. and Richards, W. G. (1989) *Mol. Pharmacol.*, **35**, 232−241.
4. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R. E. (1985) *J. Mol. Biol.*, **183**, 553−563.
5. Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A. and Wang, A. H.-J. (1989) *Biochemistry*, **28**, 310−320.
6. Coll, M., Frederick, C. A., Wang, A.H.-J. and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8385−8389.
7. Leupin, W., Chazin, W. J., Hyberts, S., Denny, W. A. and Wuthrich, K. (1986) *Biochemistry*, **25**, 5902−5910.
8. Carrondo, M. A. A. F. de C. T., Coll, M., Aymami, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H. and Rich, A. (1989) *Biochemistry*, **28**, 7849−7859.
9. Teng, M., Usman, N., Frederick, C. A. and Wang, A. H.-J. (1988) *Nucleic Acids Res.*, **16**, 2671−2690.

10. Pjura, P. E., Grzeskowiak, K. and Dickerson, R. E. (1987) *J. Mol. Biol.*, **197**, 257–271.
11. Larsen, T. A., Goodsell, D. S., Cascio, D., Grzeskowiak, K. and Dickerson, R. E. (1989) *J. Biomolec. Structure and Dynamics*, **7**, 477–491.
12. Pearl, L. H., Skelly, J. V., Hudson, B. D. and Neidle, S. (1987) *Nucleic Acids Res.*, **15**, 3469–3478.
13. Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I. and Neidle, S. (1990) *EMBO Journal*, **9**, 1329–1334.
14. Gresh, N. and Pullman, B. (1984) *Molec. Pharmacol.*, **25**, 452–455.
15. Marky, L. A. and Breslauer, K. J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4359–4363.
16. Portugal, J. and Waring, M. J. (1987) *Eur. J. Biochem.*, **167**, 281–289.
17. Fox, K. R. and Waring, M. J. (1984) *Nucleic Acids Res.*, **12**, 9271–9285.
18. Van Dyke, M. W. and Dervan, P. B. (1983) *Nucleic Acids Res.*, **10**, 5555–5567.
19. Portugal, J. and Waring, M. J. (1987) *FEBS Lett.*, **225**, 195–220.
20. Ward, B., Rehfuss, R., Goodisman, G. and Dabrowiak, J. C. (1988) *Biochemistry*, **27**, 1198–1205.
21. Rehfuss, R., Goodisman, J. and Dabrowiak, J. C. (1990) *Biochemistry*, **29**. 777–781.
22. Cons, E. M. and Fox, K. R. (1989) *Nucleic Acids Res.*, **17**, 5447–5489.
23. Arnott, S., Chandrasekaran, R., Hall, I. and Puigjaner, L. C. (1983) *Nucleic Acids Res.*, **11**,4141–4155.
24. Tilton, R. F., Weiner, P. K. and Kollman, P. A. (1983) *Biopolymers*, **22**, 969–1002.
25. Peticolas, W. L., Wang, Y. and Thomas, G. A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2579–2583.
26. Nadeau, J. G. and Crothers, D. M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2622–2626.
27. von Kitzing, E. and Dickmann, S. (1987) *Eur. Biophys. J.*, **15**, 13–26.
28. Hagerman, P. J. (1986) *Nature*, **321**, 449–450.
29. Koo, H.-S., Wu, H.-M. and Crothers, D. M. (1986) *Nature*, **320**, 501–506.
30. Lavery, R., Zakrzewska, K. and Pullman, B. (1986) *J. Biomolec. Structure and Dynamics*, **3**, 1155–1170.
31. Hogan, M., LeGrange, J. and Austin, B. *Nature*, **304**, 752–754.
32. Seibel, G. L., Singh, U. C. and Kollman, P. A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6537–6540.
33. Tidor, B., Irikura, K. K., Brooks, B. R. and Karplus, M. (1983) *J. Biomolec. Structure and Dynamics*, **1**, 231–252.
34. Van Gunsteren, W. F., Berendsen, H. J. C., Geurtsen, R. G. and Zwinderman, H. R. J. (1986) *Annals New York Acad. Sci.*, **482**, 287–303.
35. Srinivasan, A. R., Torres, R., Clark, W. and Olson, W. K. (1987) *J. Biomolec. Structure and Dynamics*, **5**, 459–496.
36. Sarai, A., Mazur, J., Nussinov, R. and Jernigan, R. L. (1989) *Biochemistry*, **28**, 7842–7849.
37. Satchwell, S., Drew, H. R. and Travers, A. A. (1986) *J. Mol. Biol.*, **191**, 659–675.
38. Gartenburg, M. R. and Crothers, D. M. (1988) *Nature*, **333**, 824–829.
39. Goodsell, D. and Dickerson, R. E. (1986) *J. Med. Chem.*, **29**, 727–733.
40. GENHELIX, a program for the generation of coordinates for A- and B-form DNA, writted by S. Neidle and L. H. Pearl at the Institute of Cancer Reseearch.
41. Weiner, S. J., Kollman, P. A., Nguyen, D. T. and Case, D. A. (1986) *J. Comp. Chem.*, **7**, 230–252.
42. Orozco, M., Herzyk, P., Laughton, C. A. and Neidle, S., (1990) *J. Biomolec. Structure and Dynamics* in press.
43. Hingerty, B., Riche, R. H., Ferrel, T. L. and Turner, J. E. (1985) *Biopolymers*, **24**, 427–439.
44. Kollman, P. AMBER technical note, February 1989.
45. Orozco, M. and Luque, F. J. (1990) *J. Comp. Chem.* (in press).
46. The molecular modelling program GEMINI, written by Dr. A. Beveridge, at the Institute of Cancer Research, is available from Hampden Data Services Ltd., Foxcombe Court, Wyndyke Furlong, Abingdon Business Park, Abingdon, Oxon OX14 1DZ.
47. We thank Professor R. Dickerson for a copy of NEWHELIX.
48. Weiner, P. K., Langridge, R., Blaney, J. M., Schaefer, R. and Kollman, P. A. (1982) *Proc. Natl. Acad Sci. USA*, **79**, 3754–3758.
49. Jayaram, B., Sharp, K. A. and Honig, B. (1989) *Biopolymers*, **28**, 975–993.
50. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2179–2183.
51. Calladine, C. R. (1982) *J. Mol. Biol.*, **161**, 343–352.
52. Vorlickova, M. and Kypr. J. (1985) *J. Biomolec. Structure and Dynamics*, **3**, 67–83.
53. McClellan, J. A., Palecek, E. and Lilley, D. M. J. (1986) *Nucleic Acids Res.*, **14**, 9291–9309.
54. Randrianarivelo, M., Zakrzewska, K. and Pullman, B. (1989) *J. Biomolec. Structure and Dynamics*, **6**, 769–799.
55. Kissinger, K., Krowiciki, K., Dabrowiak, J. C. and Lown, J. W. (1987) *Biochemistry*, **26**, 5590–5595.