Methylene Blue Binding to DNA with Alternating GC Base Sequence: A Modeling Study

Remo Rohs,[†] Heinz Sklenar,*[,†] Richard Lavery,[‡] and Beate Röder[§]

Contribution from the Arbeitsgruppe Theoretische Biophysik, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, D-13125 Berlin, Germany, Laboratoire de Biochimie Théorique CNRS UPR 9080, Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, F-75005 Paris, France, and Arbeitsgruppe Photobiophysik, Humboldt-Universität zu Berlin, Invalidenstrasse 110, D-10115 Berlin, Germany

Received August 16, 1999. Revised Manuscript Received January 18, 2000

Abstract: Photoactive methylene blue is one of the most efficient singlet oxygen generating dyes. It binds to DNA and induces photosensitized reactions which can be used for sequence-specific cleavage of the DNA backbone. Photophysical data obtained for methylene blue in complexes with DNA indicate different binding modes of the dye depending on base sequences. In this study, the binding of methylene blue to a double-stranded decamer with an alternating GC sequence has been investigated by structural modeling and force field based energy calculations. Solvation and desolvation effects have been treated using an electrostatic continuum model. For each of the three possible binding modes (intercalation and minor and major groove binding), a search of the configurational space resulted in six model structures which were selected by the criterion of lowest total energies. The differences of estimated energies are only a few kilocalories per mole, but suggest a preference for symmetric intercalation at the 5′-CpG-3′ or 5′-GpC-3′ steps. Asymmetric intercalation and minor and major groove binding appear to be less favorable. This result is compatible with published circular dichroism data. An energetic analysis of the model structures gives detailed insight into the interactions involved in the stabilization of the complex and clearly shows the importance of solvent contributions in selecting the most probable structure from an ensemble of structural alternatives.

Introduction

Photosensitization is a promising noninvasive therapeutic tool for modern photomedicine. Type II photosensitization involves the energy transfer from a photoexcited dye molecule to molecular oxygen generating a singlet state[1,2] which causes direct or indirect damages to cells. This mechanism is widely used in so-called "Photodynamic Therapy" (PDT) of tumors and other diseases.[3]

Methylene blue (MB), a phenothiazinium dye, interacts with nucleic acids, proteins, and lipids and induces photosensitized reactions after photoactivation.[4] It generates singlet oxygen very efficiently, causing photooxidative damages in biological systems, including strand breakage in DNA. The photophysical behavior of MB and of its derivatives linked to single- and double-stranded oligonucleotides has been extensively studied experimentally. The results suggest that specifically designed MB derivatives could be used as tools for site-directed cleavage of the DNA backbone. As shown by direct singlet oxygen luminescence measurements, however, the singlet oxygen quantum yield is strongly dependent on the architecture of the dye-DNA complexes. This underlines the need for detailed information on the structure and stability of MB-DNA complexes as a function of base sequences.

There are three main DNA binding modes for the dye: (i) intercalation between two successive base pairs, (ii) insertion into the minor groove, and (iii) insertion into the major groove of the double helix. Experimental studies, notably linear and circular dichroism, carried out over the last 20 years[5−8] have provided a large body of data on the structural properties of MB-DNA complexes. However, a high-resolution structure of a MB-DNA complex derived from X-ray or NMR data is not yet available. Although the spectroscopic data clearly indicate intercalation of MB between two consecutive base pairs as the predominant binding mode,[9−12] neither the geometry of the intercalation complex nor the role of alternative structures as a function of base sequence and environmental conditions (ionic strength, concentration ratio) can be precisely determined on the basis of such data.

The aim of our modeling study was to generate an ensemble of structural models by searching the configurational space,

* To whom correspondence should be addressed.
[†] Max-Delbrück-Centrum für Molekulare Medizin.
[‡] Institut de Biologie Physico-Chimique.
[§] Humboldt-Universität zu Berlin.

(1) Foote, C. S. *Science* **1986**, *162*, 963−970.
(2) Laustriat, G. *Biochimie* **1986**, *68*, 771−778.
(3) Röder, B. *Lasers Med. Sci.* **1990**, *5*, 99−106.
(4) Tuite, E. M.; Kelly, J. M. *J. Photochem. Photobiol. B* **1993**, *21*, 103−124.
(5) Nordén, B.; Tjerneld, F. *Biopolymers* **1982**, *21*, 1713−1734
(6) Wang, J.; Hogan, M.; Austin, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5896−5900.
(7) Hogan, M.; Wang, J.; Austin, R. H.; Monitto, C. L.; Hershkowitz, S. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3518−3522.
(8) Lyng, R.; Härd, T.; Nordén, B. *Biopolymers* **1987**, *26*, 1327−1345.
(9) Bradley, D. F.; Stellwagen, N. C.; O'Konski, C. T.; Paulson, C. M. *Biopolymers* **1972**, *11*, 645−652.
(10) Müller, W.; Crothers, D. M. *Eur. J. Biochem*. **1975**, *54*, 267−277.
(11) Kelly, J. M.; van der Putten, W. J. M.; McConnell, D. J. *Photochem. Photobiol*. **1987**, *45*, 167−175.
(12) OhUigin, C.; McConnell, D. J.; Kelly, J. M.; van der Putten, W. J. M. *Nucleic Acids Res.* **1987**, *15*, 7411−7427.

10.1021/ja992966k CCC: $19.00
Published on Web 03/09/2000

selecting structures on the basis of their conformational energy. In contrast to photophysical experiments using MB derivatives covalently coupled to one of the oligonucleotide strands by a flexible hydrocarbon chain, we have simplified this rather complex system and treated the dye as free ligand with a double-stranded DNA fragment. A decamer with an alternating GC base sequence was chosen as the DNA target of the dye. For this system, circular dichroism (CD) data have been reported[13] which indicate that two distinct binding sites exist. Two CD signals of different sign and strength were attributed to intercalation complexes formed at the 5′-CpG-3′ and 5′-GpC-3′ steps, but the coexistence of structures with different intercalation geometries cannot be completely excluded. Tuite and Kelly determined MB-DNA binding affinities and confirmed a stronger binding to GC than to AT alternating base sequences.[14] This and significant salt effects on the binding mode of MB observed for AT sequences leads us to expect that theoretical predictions for AT are more difficult than those for GC. The present study was therefore restricted to the GC sequences where the clear preference for intercalation has been found to be largely independent of environmental conditions.[13] From binding studies as a function of salt concentration, it was concluded that the orientation of the long axis of MB, which is assumed to be parallel to the averaged long axis of the two base pairs in the case of low salt concentration, changes to a perpendicular orientation at high salt.[5]

Using molecular modeling, a systematic search for low-energy structures has been carried out for each of the binding modes. To account for solvent effects, the energies of the optimized structures resulting from energy minimization were corrected by including the electrostatic reaction field contributions of the solvent. The results show that these interactions are critically important for predicting the relative stability of the optimized structures in the different binding modes. In agreement with the results of CD measurements,[13] the estimated energies suggest both 5′-CpG-3′ and 5′-GpC-3′ intercalation of the dye as the preferred binding modes of MB to DNA with GC alternating sequences. In addition to the lowest energy intercalation structures which obey dyadic symmetry, relatively low-energy stable models involving asymmetric ("gauche") intercalation and groove binding complexes were located and these have also been analyzed in energetic and structural terms. The calculated energy differences are of the order of only a few kilocalories per mole and indicate that such structures could be formed under particular environmental conditions or in the case of other base sequences.

Methods

Structural Modeling and Energy Minimization. The structure of the MB dye was defined using bond lengths and valence angles obtained by geometry optimization with the consistent valence force field (CVFF) method.[15] The planar ring system, shown in Figure 1, obeys 2-fold symmetry and was taken as a rigid body. This restriction results in a structural model of the MB molecule with six conformational degrees of freedom defined by the torsion angles of the two methylated amino groups. In the case of free MB, the force field based energy minimization yielded a structure with both methyl carbon atoms lying almost in the plane of the aromatic ring system.

For structural modeling of MB-DNA complexes, the JUMNA (Junction Minimization of Nucleic Acids) algorithm[16] has been used. This program provides versatile tools for the generation of starting structures, conformational searching, and energy minimization of complexes formed between nucleic acids and ligands. The description of intramolecular flexibility by internal coordinates with fixed bond lengths is complemented by helicoidal variables which relate the translations and rotations of the nucleic acid bases and of the ligands to a helical axis system. This feature facilitates the creation of appropriate starting structures for energy minimization with different orientations of the dye relative to the DNA molecule.

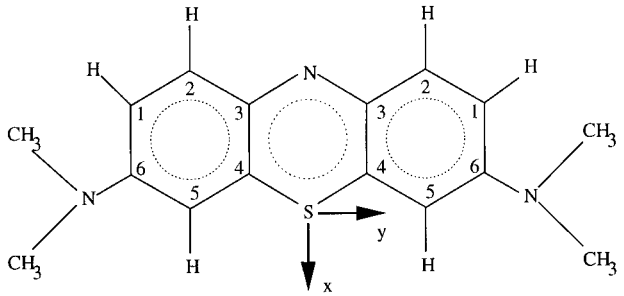

Figure 1. Chemical structure of the MB molecule. The given numbers refer to the assignment of atomic charges in Table 1. The short (*x*-) and long (*y*-) axes are used for describing the position of the dye in the helical coordinate system of the DNA fragment (see below).

The built-in Flex force field[17] used in JUMNA for energy minimization calculates energies and forces in terms of pairwise nonbonded interactions that add to bonded atom terms involving harmonic valence angle contributions and cosine dihedral barriers. Charge distributions are represented by atomic monopoles. For the standard deoxyribonucleotides, these are taken from a library that was calculated using a re-parametrized Hückel−Del Re method.[17,18] To define a consistent charge distribution for the MB dye, SCF ab initio calculations, using the Gaussian 94 program[19] with a standard STO-3G basis set including d-orbitals, have been carried out for extending the re-parametrization of the Hückel−Del Re method to the aromatic sulfur atom. In this way, an optimal fit to the SCF ab initio electrostatic potential and field distributions has been obtained, both for DNA and for the dye.

It is worth mentioning that, in case of DNA, this charge distribution gave very similar results to those obtained with fits to charge distributions from much more extended basis set calculations in the latest Amber force field.[20] A distance-dependent sigmoidal dielectric function[17] and reduced phosphate charges serve to describe solvent-induced electrostatic damping of nonbonded Coulomb interactions. The restriction to pairwise electrostatic interactions of the solute charges makes conformational search and structural refinement by energy minimization computationally feasible, but often fails in the prediction of relative stability of alternative structures.[21] The electrostatic component of the energy-minimized structures was therefore recalculated by using an electrostatic continuum description of the aqueous solvent.

Ideally, this solvent treatment should be integrated into the energy minimization procedure. This, however, requires an analytical calculation of the corresponding force components, a problem that has not been completely solved yet. Accordingly, we had to confine this study to "single point" corrections. The corrected energies, calculated as the sum of the nonelectrostatic components of the Flex force field, the Coulomb interactions of the solute charges, and the electrostatic solute−solvent interactions (described by the reaction field), however, depend to some extent on the parameters of the electrostatic damping model used in energy minimization. The choice of appropriate parameters has been based on the criterion of lowest corrected total energies. For the slope of the sigmoidal dielectric function, a value of 0.25 $Å^{-1}$ was

(13) Tuite, E. M.; Nordén, B. J. *Am. Chem. Soc.* **1994**, *116*, 7548−7556.
(14) Tuite, E. M.; Kelly, J. M. *Biopolymers* **1995**, *35*, 419−433.
(15) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins: Structure, Functions and Genetics* **1988**, *4*, 31−47.
(16) Lavery, R.; Zakrzewska, K.; Sklenar, H. *Comput. Phys. Commun.* **1995**, *91*, 135−158.
(17) Lavery, R.; Sklenar, H.; Zakrzewska, K.; Pullman, B. *J. Biomol. Struct. Dyn.* **1986**, *3*, 989−1014.
(18) Lavery, R.; Zakrzewska, K.; Pullman, A. *J. Comput. Chem.* **1984**, *5*, 363−373.
(19) Pople, J. A. *Gaussian 94*; Gaussian Inc.: Pittburgh, 1995.
(20) Flatters, D.; Zakrzewska, K.; Lavery, R. *J. Comput. Chem.* **1997**, *18*, 1043−1055.
(21) Zakrzewska, K.; Madami, A.; Lavery, R. *Chem. Phys.* **1996**, *204*, 263−269.

chosen which damps the Coulomb interactions by an effective permittivity starting with $\epsilon = 2$ for short distances and reaching asymptotically the bulk value of $\epsilon = 78$ for distances >20 Å. With use of this damping in combination with a reduction of all phosphate net charges to −0.25e, energy minimization yielded structures with corrected total energies that are close to their minimum values. This has been tested by adiabatic mapping of the corrected total energy with respect to some helical and sugar pucker variables in the vicinity of the optimized geometry.

Continuum Treatment of Solvent Electrostatic Effects. The continuum description of the solvent allows for a simplified treatment of electrostatic solute−solvent interactions by using classical continuum electrostatics. In this approach, the interface between the high dielectric permittivity ($\epsilon = 78$) of an aqueous solvent and the low permittivity ($\epsilon = 1$) of the solute is defined by the van der Waals surface of the solute. The solvent polarization induced by the electric field of solute charges is the source of a reaction field that contributes to the electrostatic energy and results in stabilization or destabilization of the different solute structures. Compared to the electrostatic solvent component, salt-dependent contributions, nonelectrostatic solute−solvent interactions, and entropic terms have generally smaller differential effects on alternative solute structures.[21] This has been shown by the study of several looped-out and stacked-in conformers of a single-base bulge in double-stranded DNA and RNA.[22,23] In this case, which is comparable to our system, both salt and entropic solvent contributions, estimated by the change of the solvent-accessible surface area, have been found to vary only very slightly despite considerable changes in conformation. Thus, both terms are expected to lie in the range of uncertainty introduced by force field parametrization and other approximations used in the structural modeling algorithm, and have therefore been neglected. The dependency on salt concentration observed in experiments[24,25] is however an important aspect that should be investigated in a future study.

The boundary value problem defined by the Poisson equation requires a numerical treatment. We have used the finite difference (FD) algorithm implemented by the Delphi program.[26] A probe sphere radius of 1.4 Å was used for smoothing the solute−solvent interface. To calculate the reaction field energy with an accuracy in the order of 0.1 kcal/mol, we used 150 points in each dimension of the cubic grid and carried out five focusing steps starting with a Coulomb boundary and an initial solute filling factor of 20% (grid spacing of 1.2 Å) and ending with a filling factor of 90% (grid spacing <0.3 Å). With these parameters, the small dependency (of the order of 1 kcal/mol) of the electrostatic energy on the position of the molecules has been removed by averaging over 16 equally spaced angular orientations. The total electrostatic energy can be divided into pairwise additive Coulomb interactions of the solute charges and the reaction field contributions of the polarized solvent continuum. Both components were added to the nonelectrostatic energy of the Flex force field to account for electrostatic solvent effects in estimating the relative stability of modeled structures.

Structural Analysis. Depending on the binding mode, interactions with the dye induce distinct changes of the DNA structure in comparison to the free double helix. The deformations were analyzed in terms of helical parameters, widths and depths of the grooves, sugar puckering states, and backbone torsion angles. The helical parameters refer to the global helical axis system of the DNA calculated by the CURVES program.[27,28] By analogy with the definition of base pair axis parameters,[29] the position and orientation of MB relative to flanking base pairs is described by MB axis parameters (X-displacement, Y-displacement, Rise, Inclination, Tip, and Twist). For this purpose, a rectangular coordinate system with its origin at the sulfur atom and the *x*- and *y*-axes parallel to the short and long axes of MB is introduced.

Table 1. Atomic Monopole Charges (in au) of MB Calculated by Means of the Reparametrized Hückel−Del Re Method[a]

atom	C1	C2	C3	C4	C5	C6	C_{methyl}
charge	−0.12	0.01	0.15	0.29	−0.16	0.32	0.05
atom	S	N_{arom}	N_{amino}	H1	H2	H5	H_{methyl}
charge	−0.26	−0.07	−0.45	0.05	0.05	0.05	0.06

[a] The atom notations refer to the atom numbering indicated in Figure 1.

The direction of the *x*-axis is chosen from nitrogen to sulfur to agree with the dyadic axis of the symmetric intercalation complex (see Figure 1).

Results

Charge Distribution of Methylene Blue. In photochemical applications of MB, singlet oxygen is generated by energy transfer from an excited singlet state of the dye. The lifetime of the excited state is however very short compared to the intercalation time.[30,31] It is therefore justified to use a charge distribution corresponding to its ground state. The atomic monopoles resulting from the Hückel−Del Re method (reparametrized on the basis of SCF ab initio calculations) are given in Table 1 and refer to the atom numbering indicated in Figure 1. These charges have been used both for structural modeling by energy minimization and for the continuum treatment of electrostatic solute−solvent interactions.

Structural Modeling and Energy Minimization. As a starting point for our study, a systematic conformational search has been performed for the free DNA target structure with a GC alternating base sequence. Taking into account the sequence symmetry, the search was restricted to conformers with dinucleotide symmetry and homonomous strands. The deformations applied to the canonical B-form structure included changes of sugar pucker parameters and of the backbone torsion angle pairs (α,γ) and (ϵ,ζ) in the dinucleotide unit. Subsequent energy minimization resulted in 19 different local minimum structures. From this set, the two lowest energy conformers were selected as appropriate starting structures for our ligand binding study. Both structures show the characteristic features of canonical B-form DNA and are similar to the GC alternating hexamer structure determined by NMR.[32] The structural differences of the two conformers are due to a change of the sugar pucker from C2′-endo to O1′-endo at cytidine in the lowest-energy structure. The significantly higher energies of the discarded DNA structures, which contain unusual backbone conformations, are mainly due to unfavorable reaction field contributions. For each of the three binding modes, structural models of MB-DNA complexes were generated by energy minimization of different starting structures. Six helical variables (three translations and three rotations) have been used for describing the position and orientation of the dye relative to the helical axis of the DNA. In the case of intercalation structures, the dye was inserted into both possible intercalation sites (5′-CpG-3′ and 5′-GpC-3′) created by stretching and unwinding the respective base pair steps within the two lowest energy decamer structures obtained

(22) Zacharias, M.; Sklenar, H. *Biophys. J.* **1997**, *73*, 2990−3003.
(23) Zacharias, M.; Sklenar, H. *J. Mol. Biol.* **1999**, *289*, 261−275.
(24) van der Putten, W. J. M.; Kelly, J. M. *Photochem. Photobiol.* **1989**, *49*, 145−151.
(25) Hagmar, P.; Pierrou, S.; Nielsen, P.; Nordén, B.; Kubista, M. *J. Biomol. Struct. Dyn.* **1992**, *9*, 667−679.
(26) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327−335.
(27) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.
(28) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655−667.
(29) Dickerson, R. E.; Bansal, M.; Calladine, C. R.; Diekmann, S.; Hunter, W. N.; Kennard, O.; Lavery, R.; Nelson, H. C. M.; Olson, W. K.; Saenger, W.; Shakked, Z.; Sklenar, H.; Soumpasis, D. M.; Tung, C.-S.; von Kitzing, E.; Wang, A. A.-J.; Zhurkin, V. B. *Mol. Biol.* **1989**, *205*, 787−791.
(30) Kelly, J. M.; Tuite, E. M.; van der Putten, W. J. M.; Beddard, G. S.; Reid, G. D. *Supramolecular Chemistry*; Balzani, V., De Cola, L., Eds.; Dordrecht: 1992; pp 375−381.
(31) Tuite, E. M.; Kelly, J. M.; Beddard, G. S.; Reid, G. S. *Chem. Phys. Lett.* **1994**, *226*, 517−524.
(32) Lam, S. L.; Au−Yeung, S. C. F. *J. Mol. Biol.* **1997**, *266*, 745−760.

for the free DNA fragment. The translational parameters (X-displacement, Y-displacement, Rise) were used to position the dye in the intercalation pocket. In the starting structures, the heterocyclic ring system of the dye was placed perpendicular to the helical axis (Inclination = 0°, Tip = 0°). For each of the models, an adiabatic mapping of the potential energy as a function of the rotation angle of the MB molecule about the helical axis (Twist) has been performed. In these calculations, we have changed this angle in steps of 10° and searched for the global minimum with respect to in-plane shifts of the dye. The other inter- and intramolecular degrees of freedom were relaxed by energy minimization. In all cases, the resulting potential curves show two deep minima corresponding to structures where the long axis of the dye is either parallel to the averaged long axis of the flanking base pairs or rotated by approximately 140°. The differences of the energy minima are 2.6 kcal/mol (CpG) and 2.3 kcal/mol (GpC), respectively, and the high energy barriers (with respect to overall rotation about the helical axis) are due to close contacts between the methyl groups of MB and the DNA backbone.

The shape of the curve is only slightly affected by the choice of electrostatic parameters (data not shown). This result leads us to believe that actual intercalation structures can be expected to be close to one of the four possible models (two for CpG and two for GpC intercalation). The creation of representative structural models for MB-DNA complexes in the groove binding modes is more difficult. We have explored the space of the six intermolecular variables by stepwise translations and rotations of the dye located in the grooves. Both orientations of the dye with respect to the position of the sulfur atom (inside and outside the grooves) have been tested. However, the structures with the sulfur atom inside the grooves turned out to be extremely unfavorable because of the bulky methyl groups at this edge of the dye. We have also observed that rotations about the short axis of MB are much more restricted in the narrow minor groove than in the wider major groove. Finally a total of 66 starting models, 24 for minor groove and 42 for major groove binding, were obtained and combined with both DNA target structures. In both cases, energy minimization yielded 8 local minimum structures for minor groove binding and 15 for major groove binding.

Energetic Analysis. For the selection of lowest energy structures from the ensemble of model structures, we have used the calculated total energies, corrected by substituting the electrostatic component of the Flex force field by the electrostatic energy resulting from the continuum solvent treatment.

The first result of this energy comparison is that the choice of the DNA target structure (alternating C2′-endo/O1′-endo or all C2′-endo sugar pucker) does not affect the order of stability of corresponding complex structures. Thus, we have confined our analysis to structures resulting from the lowest energy target structure obtained for free DNA, with alternating sugar conformations, although the alternative set of higher energy structures is in slightly better agreement with the structure of the free CpG alternating hexamer derived from NMR data.[32]

In the following we compare the relative stability of the four intercalation structures (denoted by ic1-CpG, ic2-CpG, ic1-GpC, and ic2-GpC) and the lowest energy structures found in minor groove (min-g) and major groove (maj-g) binding, respectively. The calculated energies and their decompositions into nonelectrostatic and electrostatic components are given in Table 2. The energy values were calculated as differences relative to the lowest calculated total energy, ic1-CpG. The results can be summarized in the following points: (i) Using continuum electrostatics (E_{cm}^{elec}) instead of sigmoidal damping (E_{Ed}) changes the order of stability from "min-g/maj-g/ic1-CpG/ic1-GpC/ic2-CpG/ic2-GpC" to "ic1-CpG/ic1-GpC/min-g/ic2-GpC/ic2-CpG/maj-g". (ii) This change is due to the unfavorable total electrostatic energies obtained for both groove binding modes. Although groove binding is much more stabilized by reaction field contributions (E_{Rf}) than intercalation (because of a higher solvent accessibility of the dye), this effect is overcompensated by larger Coulomb repulsion (E_{Coul}). (iii) In contrast to electrostatic damping, the continuum treatment of solvent electrostatics almost completely counterbalances the differences of the nonelectrostatic energy contributions calculated for CpG and GpC intercalation.

Table 2. Decomposition of Total Energies (in kcal/mol) Obtained for the Modeled MB-DNA Complexes in the Three Different Binding Modes[a]

structure	ΔE_{Ne}	ΔE_{Ed}	ΔE_{Flex}^{tot}	ΔE_{Coul}	ΔE_{Rf}	ΔE_{cm}^{elec}	ΔE_{cm}^{tot}
ic1-CpG	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ic2-CpG	1.9	0.7	2.6	7.8	−7.1	0.7	2.6
ic1-GpC	−3.1	4.2	1.1	58.9	−55.7	3.2	0.1
ic2-GpC	3.6	4.9	8.5	40.8	−42.0	−1.2	2.4
min-g	−12.0	0.2	−11.8	158.2	−144.2	14.0	2.0
maj-g	−6.7	2.0	−4.7	152.9	−141.7	11.2	4.5

[a] Notations of the structures indicate the binding mode (ic: intercalation, min-g: minor groove insertion, maj-g: major groove insertion), the intercalation site (CpG: 5′-CpG-3′ step; GpC: 5′-GpC-3′ step) and distinguished different orientations of the dye (ic1: the long axis of MB is parallel to the long axis of the base pair; ic2: the long axis of MB is rotated about the helix axis). The energies are given as differences relative to the lowest energy structure (ic1-CpG) and include the nonelectrostatic contributions (ΔE_{Ne}) of the Flex force field, the electrostatic energy obtained by electrostatic damping (ΔE_{Ed}), the total flex energy (ΔE_{Flex}^{tot}), the Coulomb energy of solute charges (ΔE_{Coul}), the reaction field energy (ΔE_{Rf}), the total electrostatic energy (ΔE_{cm}^{elec}) of the continuum model, and the final total energy including nonelectrostatic contributions (ΔE_{cm}^{tot}).

Table 3. Decomposition of Binding Energies (in kcal/mol) Obtained for the Modeled MB-DNA Complexes in the Three Different Binding Modes[a]

structure	ΔE_{DNA}^{def}	ΔE_{MB}^{def}	E_{Int}	E_{Bdg}
ic1-CpG	5.6	0.1	−20.5	−14.8
ic2-CpG	6.1	0.1	−18.4	−12.2
ic1-GpC	2.8	0.2	−17.7	−14.7
ic2-GpC	5.6	0.1	−18.1	−12.4
min-g	2.9	0.1	−15.8	−12.8
maj-g	1.9	0.1	−12.3	−10.3

[a] Notation of structures corresponds to Table 1. The deformation energies of DNA (ΔE_{DNA}^{def}) and methylene blue (ΔE_{MB}^{def}) are differences of total energies (ΔE_{cm}^{tot}) obtained for the deformed structures in the complexes and the respective free structures. E_{Int} is the interaction energy of the deformed structures and E_{Bdg} the resulting binding energy.

Binding energies of MB-DNA complexes can be estimated by calculation of differences between their total energies and the sum of lowest energies found for the optimized structures of free DNA and MB. Although the values given in Table 3 have a correct order of magnitude, we should stress that they result from differences between large numbers and are certainly affected by the approximations used in our simplified energy calculations and, in particular, by the neglect of salt effects. The decomposition of binding energies into deformation and interaction components shows that the deformation contribution of DNA found in intercalation structures is consistently larger than that in groove binding complexes. The interaction energy of the deformed structures, however, is more favorable in the intercalation structures. The lower stability of the secondary intercalation structures, where the dye is rotated about the helical

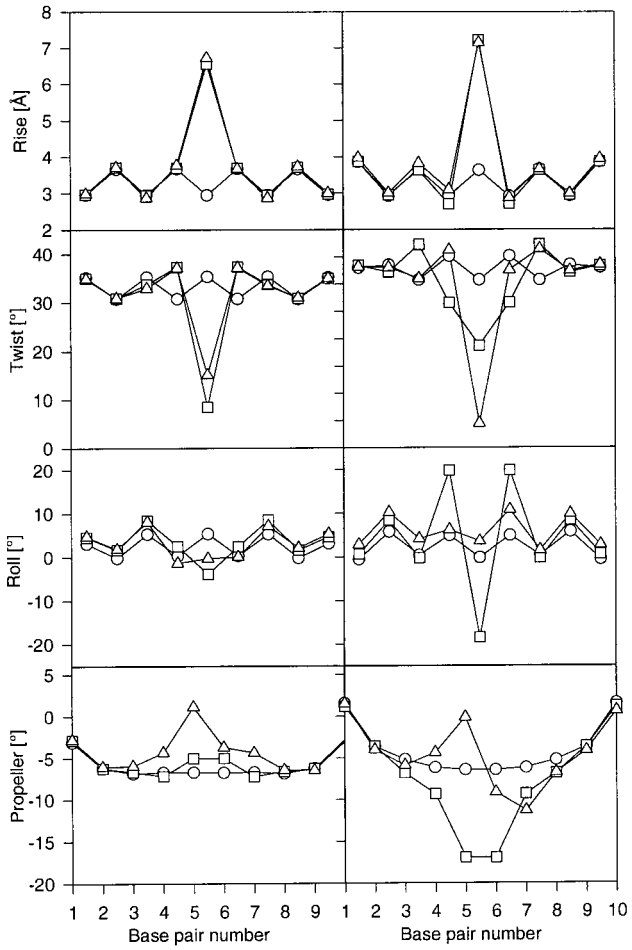

Figure 2. Helical parameters of the DNA decamers in the 5′-CpG-3′ (left panel) and 5′-GpC-3′ (right panel) intercalation complexes with intercalation sites between the central base pairs 5 and 6. The parameters obtained for symmetric (squares) and gauche (triangles) intercalation are compared with those of free DNA (circles) and refer to the global helical axis system calculated by the CURVES program.[27,28]

axis by an angle between 135° and 140° compared to the symmetric orientation, is mainly due the larger deformation energy of DNA.

Structural Analysis. The inversion symmetry of pCpG dinucleotide repeats is reflected by a dyadic symmetry of the lowest energy structures of the free DNA decamer. The helical parameter plots (Figure 2) show that this symmetry is conserved in the intercalation complexes ic1-CpG and ic1-GpC. In these structures, the 2-fold symmetry axis of MB coincides with the dyadic axis of the DNA fragment. Although the estimated stability of the complexes ic1-CpG and ic1-GpC is similar, there are remarkable differences in structural terms: The intercalation cavities resulting from stretching and untwisting the central base pair step differ in the calculated *Rise* and *Twist* parameters which are 8° and 6.6 Å at the 5′-CpG-3′ step and 19° and 7.2 Å at the 5′-GpC-3′ step (see Figure 2). An even more distinct variation of the Roll angles and of the Propeller twist angles of the flanking base pairs is also seen in the molecular models displayed in Figure 3. The differences of helix morphology are probably due to a different sugar pucker at the flanking nucleotides. In contrast to the sugar pucker sequence 5′-G(C2′-endo)pC(C3′-endo)>p<G(C2′-endo)pC(O1′-endo)-3′ of ic1-CpG, the sequence 5′-C(C3′-endo)pG(C3′-endo)>p<C(C3′-endo)pG(C2′-endo)-3′ was found in ic1-GpC. Another characteristic feature of the backbone geometry at intercalation sites, the so-called α/γ flip (this torsion angle pair changes from the usual gauche−/gauche+ state to the unusual trans/trans state), is observed in both intercalation structures.

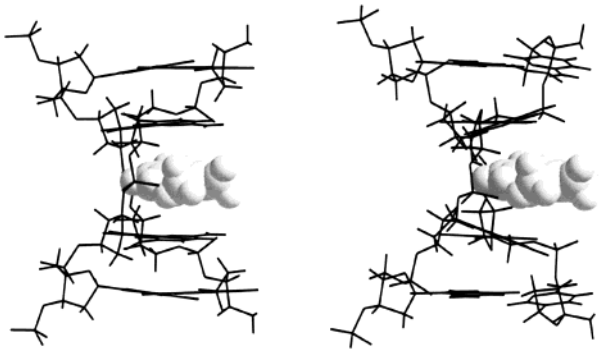

Figure 3. Molecular graphics representations of the symmetric intercalation complexes ic1-CpG (left) and ic1-GpC (right). Note the different Propeller angles of the flanking base pairs.

Table 4. Helical Parameters (translations in Å, Rotations in deg) Describing the Position and Orientation of MB Relative to Flanking Base Pairs in the Global Helical Axes System of DNA[a]

structure	*Xdisp*	*Ydisp*	*Rise*	*Inc*	*Tip*	*Twist*
ic1-CpG	0.9	−0.1	3.3	−5.1	−0.3	4.0
ic2-CpG	3.4	1.0	3.1	3.7	9.6	142.5
ic1-GpC	1.4	−0.2	3.7	−5.8	−5.1	8.6
ic2-GpC	2.8	0.0	3.4	10.0	3.9	145.5
min-g	9.6	−1.2	0.0	34.4	0.9	−170.8
maj-g	5.9	−0.2	1.4	−14.1	3.3	15.6

[a] Notation of structures corresponds to Table 1. The helical parameters describing the MB axis system with respect to the axis system associated with the flanking base pair were calculated according to the definitions given by Lavery and Sklenar.[27,28] Let $\vec{P}$ be the origin of the base pair system and $\vec{u}$, $\vec{v}$, $\vec{w}$, the associated helical axis system (calculated with CURVES) and $\vec{O}$ the origin of the MB (fixed) system with $\vec{x}$, $\vec{y}$, $\vec{z}$ being the axis vectors (*n* indicates normalized vectors):

$$Xdisp = (\vec{O} - \vec{P}) \cdot (\vec{y} \times \vec{u})_n; \quad Ydisp = (\vec{O} - \vec{P}) \cdot (\vec{u} \times (\vec{y} \times \vec{u})_n); \quad Rise = (\vec{O} - \vec{P}) \cdot \vec{u};$$

$$Inc = \arcsin(\vec{y} \cdot \vec{u}); \quad Tip = \begin{cases} \arccos((\vec{y} \times \vec{u})_n \cdot \vec{x}) & \text{if} \quad (\vec{u} \cdot \vec{x}) \le 0 \\ -\arccos((\vec{y} \times \vec{u})_n \cdot \vec{x}) & \text{if} \quad (\vec{u} \cdot \vec{x}) > 0 \end{cases};$$

$$Twist = \begin{cases} \arccos((\vec{y} \times \vec{u})_n \cdot \vec{v}) & \text{if} \quad ((\vec{y} \times \vec{u})_n \cdot \vec{w}) \le 0 \\ -\arccos((\vec{y} \times \vec{u})_n \cdot \vec{v}) & \text{if} \quad ((\vec{y} \times \vec{u})_n \cdot \vec{w}) > 0 \end{cases}$$

Helical parameters describing the position and orientation of MB relative to flanking base pairs are given in Table 4. The calculated values measure the translations and rotations of the dye relative to the helical axis system of flanking base pairs. The parameters obtained for ic1-CpG and ic1-GpC are similar and indicate that MB is positioned in the center of the intercalation cavity with its methyl groups lying in the major groove.

The inversion symmetry of the lowest energy intercalation structures is lost if the dye is rotated about the helical axis (a rotation by 180° would reestablish the symmetry, but this structure is not stable because the bulky methyl groups cannot be accommodated in the minor groove). The structures ic2-CpG and ic2-GpC result from a right-handed rotation (or an equivalent left-handed rotation if the DNA strands are exchanged) of the dye by approximately 140° relative to the orientation in the symmetric complexes. According to our force field based adiabatic mapping and energy evaluation, these structures are located in a well-defined local energy minima (~2.5 kcal/mol above the lowest energy state) and are the only alternatives to the symmetric intercalation complexes. We propose the notation "gauche intercalation" which reflects the fact that both the backbone geometry of the two DNA strands and the stacking interactions of the dye with the two flanking base pairs are different. The different stacking pattern observed

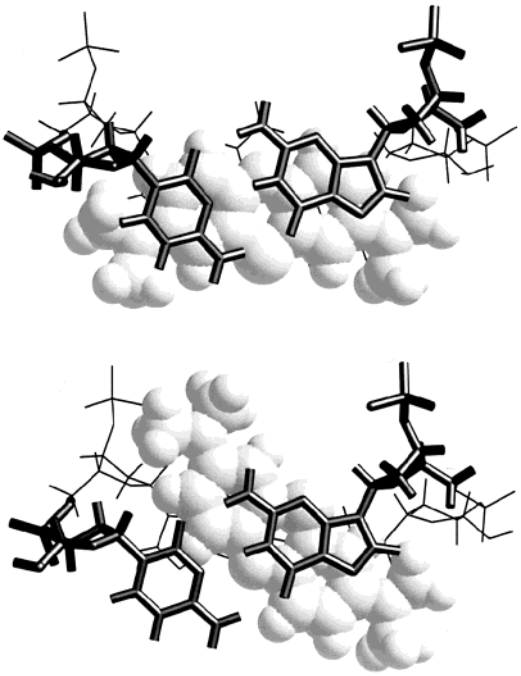

Figure 4. Stacking pattern in the symmetric (upper figure) and asymmetric "gauche" (lower figure) intercalation complexes ic1-CpG and ic2-CpG. The view is perpendicular to the plane of the dye. In contrast to the symmetric stacking of MB on both base pairs, the stacking is reduced to one base pair (above the MB plane) and to the guanine base of the other base pair (below the MB plane) in the asymmetric case.

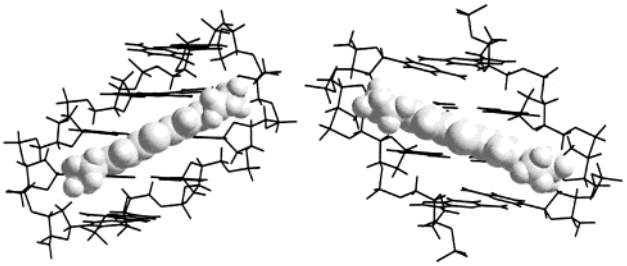

Figure 5. Structural models obtained for minor (left) and major (right) groove binding of MB to DNA. Whereas the long axis of the dye is oriented parallel to the DNA strands in the minor groove complex, it spans the groove in the major groove complex.

in the symmetric (ic1-CpG) and gauche (ic2-CpG) intercalation complexes is shown in Figure 4. Compared to the symmetric stacking interactions with both flanking base pairs, the gauche intercalation allows an optimal stacking with only one base pair and the guanine base of the other. The helical parameters of the DNA in the gauche intercalation complexes (Figure 2) indicate a considerable flexibility of the DNA structure in adapting to the dye. The structural differences (*Twist*, *Roll*, *Propeller*) between symmetric and gauche intercalation sites are more drastic for the GpC step than for CpG. It should also be noted that the dye is slightly shifted from the center of the cavity (helical axis position) toward the minor groove. Corresponding translational parameters *Xdisp* and *Ydisp* are given in Table 4.

In contrast to the intercalation complexes, the structures obtained by insertion of MB into the grooves of DNA are less well-defined. The two representatives, one for minor groove binding and one for major groove binding, shown in Figure 5 were selected from an ensemble of local minima structures with comparable energies. According to our energy evaluation, these

Table 5. Groove Widths and Depths (in Å) of the DNA Fragments Deformed by Interactions with MB in the Three Different Binding Modes[a]

structure	w_{min}	d_{min}	w_{maj}	d_{maj}
free helix	6.9	4.2	12.5	7.7
ic1-CpG	9.5	3.2	18.1	4.0
ic2-CpG	8.8	3.6	17.1	3.8
ic1-GpC	10.8	3.2	16.0	7.6
ic2-GpC	11.2	2.9	16.9	7.1
min-g	4.7	5.5	13.7	5.6
maj-g	7.9	3.6	9.2	9.3

[a] Notation of structures corresponds to Table 1. The groove widths (w_{min} and w_{maj}) and depths (d_{min} and d_{maj}) of the minor and major grooves (averaged over the central base pair step) were calculated by using the algorithm of Stofer and Lavery[33] as implemented by the CURVES program.[27,28]

structures are the lowest energy variants of the minor and major groove binding complexes. In both structures, the sulfur atom and the methyl groups of MB face the outer edges of the grooves (avoiding the steric hindrance associated with the reverse orientation). Interestingly, in the minor groove complex the center of the MB ligand is positioned in the base pair plane (*Rise* = 0.0, see Table 4) so that the structure does not obey inversion symmetry. In this case, a hydrogen bond between the central nitrogen atom of MB and the amino group of guanine is formed. On the other hand, the lowest energy major groove complex (2.5 kcal/mol higher in energy than the best minor groove site), where the short axis of MB agrees with the dyadic axis (between the two central base pairs) of the DNA decamer, is perfectly symmetric. The orientation of the ligand in the minor and major groove is also different. Whereas in the minor groove the long axis of the dye is nearly parallel to the backbones of both DNA strands (*Tilt* = 34°, see Table 4), it spans the groove in the major groove complex (*Tilt* = −14°) allowing favorable contacts with both guanines of the central base pair step. The calculated groove geometries,[33] shown in Table 5, indicate a narrowing of the respective grooves at the position where the ligand is bound, in comparison to the groove widths of free DNA and to the values obtained for the intercalation complexes where both grooves are wider.

Discussion

Our modeling study of MB binding to DNA with an alternating GC base sequence resulted in six structural models which are representatives of the different binding modes of the MB-DNA complex. The critical question is, to what extent do these models reflect the structural behavior of the real system? This can only be answered by a comparison with the experimental data. For DNA with an alternating GC base sequence, it has been shown by absorption spectroscopy[13] that intercalation of MB is the only binding mode. This result is independent of ionic strength and concentration ratio. Our modeling results, where salt effects have been neglected, can therefore be compared directly with the results of CD studies, where two signals of opposite sign and shift[13] were interpreted as two different binding sites corresponding to the 5′-CpG-3′ and 5′-GpC-3′ intercalation pockets. The equal stability estimated for the lowest energy structures, ic1-CpG and ic1-GpC, is in correspondence with these data. Our structural and energetic analysis suggests that the symmetric (parallel) intercalation structures are energetically more favorable than the asymmetric (gauche) variants (ic2-CpG and ic2-GpC), but this could not be verified from the CD study because of the strong induced

(33) Stofer, E.; Lavery, R. *Biopolymers* **1993**, *34*, 337−346.

CD signal dependence of lateral and rotational displacements. Geometrical data on DNA helix lengthening (2.9 Å) and unwinding (24°) upon MB intercalation, derived from electric dichroism measurements[7] and topoisomerization experiments,[12] agree very well with the helical parameters (see Figure 2) of the modeled intercalation complexes. Although there is no experimental evidence for the different sugar pucker sequences found in the ic1-CpG and ic1-GpC intercalation complexes, it should be mentioned that both types of sugar pucker have been observed in the X-ray crystal structures of other nucleoside−intercalator complexes[34]

An advantage of the molecular modeling approach is that it allows a detailed energetic analysis of both the structural models which are assumed to be close to those occurring in the real system and of unobserved structures. The data given in Tables 2 and 3 show that the relative stability of the alternative structural models (intercalation vs groove binding) results from the fine balance of very different energy contributions. Thus, the favorable solvent contributions in the groove binding complexes are counterbalanced by much less favorable Coulomb interactions of the solute charges, and the larger deformation energy of the DNA in the intercalation complexes is compensated by the favorable (stacking) interactions of the dye with the bases. The relatively small differences of the total energies calculated for the alternative model structures indicate that the structure of MB-DNA complexes could be sensitive to environmental conditions and base sequences, but also underline the need of rather precise energy calculations. Indeed, it should be noted that the existence of both gauche intercalation complexes and groove binding complexes (for AT sequences) has been discussed in several experimental studies.[5,8,13] Base sequence effects on MB binding to DNA as a function of salt concentration are of considerable interest and will be the subject of a forthcoming study. In this case, however, reliable predictions have to be based on estimates of free energy differences which can only be derived from dynamic simulations.

(34) Sobell, H. M. *Biological Macromolecules and Assemblies, Volume 2: Nucleic Acids & Interactive Proteins*; Jurnak, F. A., MacPherson, A., Eds.; Wiley: New York, 1985; pp 171−232.

Conclusions

On the basis of a modeling study of MB binding to a DNA decamer with an alternating GC base sequence, six structural models have been derived. The conclusion of detailed structural and energetic analysis is that these structures could be considered as representatives of MB-DNA complexes in different binding modes (intercalation, minor and major groove binding). The lowest energies (including a continuum treatment of solvent electrostatic effects) have been found for the symmetric intercalation complexes, at either the 5′-CpG-3′ or the 5′-GpC-3′ sites. This result is in agreement with the structural interpretation given to experimental data by other authors. Small differences of total energies calculated for the alternative gauche intercalation and groove binding complexes indicate that such structures may exist under other environmental conditions or in the case of other base sequences. Although a relatively simple modeling technique and solvent treatment have been used, the qualitative behavior of the model seems to be in good agreement with the data obtained experimentally. The predicted structures may therefore be useful for a more detailed interpretation of experimental results. They could serve as a starting point for molecular dynamics simulations and for studying base sequence effects in view of the photochemical applications of MB.

Acknowledgment. R. Rohs acknowledges funding by the Deutscher Akademischer Austauschdienst (DAAD-HSP III) and the Deutsche Forschungsgemeinschaft (DFG-GK 80).

JA992966K