METHODS: A Companion to Methods in Enzymology **14**, 277–292 (1998)
Article No. ME980584

Stacking Interactions and Intercalative DNA Binding

Federico Gago

Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain

The DNA-binding properties of many ligands can be rationalized on the basis of their structural and electronic complementarity with the functional groups present in the minor and major grooves of particular DNA sequences. Specific hydrogen bonding patterns are particularly useful for the purpose of sequence recognition. Less obvious, however, is the influence of base composition on the conformational preferences of individual base steps and on the binding of intercalating moieties which become sandwiched between contiguous base pairs. Improved knowledge of stacking interactions may lead to a better understanding of the architecture and inherent flexibility of particular DNA sequences and may provide insight into the principles that dictate the structural changes and specificity patterns observed in the binding of some intercalating ligands to DNA. © 1998 Academic Press

The genetic material in living cells is made up of deoxyribonucleic acid (DNA), a storage site for nucleotide sequence information, the integrity and retrieval of which rely on its interaction with proteins and other ligands. These interactions imply mutual recognition that in some cases can be very specific.

At its lowest organization level, the well-defined three-dimensional structure of the DNA molecule is comparatively simpler than that of proteins: hydrogen-bonded base pairs are stacked like coins in a roll along the axis of a right-handed double helix with the sugar–phosphate backbones of each strand on the outside, winding up in antiparallel orientations (1). The interstrand hydrogen bonding patterns naturally impose limitations on the ability of the base pairs to be used as molecular recognition targets, and so the features that first appeared to be recognized by sequence-specific DNA-binding proteins and low-molecular-weight ligands that show sequence selectivity were found along the major and the minor grooves that lie between the phosphodiester linkages of both strands. This simple picture is, however, misleading as there exists a sequence-dependent microheterogeneity in double-helical DNA (2) that provides an additional potential for specific recognition, including differences in the ability to bend (3), and also because DNA can adopt higher-order structures.

Molecular models have proved to be very useful for understanding function of both proteins and nucleic acids at the atomic level. Detailed structural information about these macromolecules and their complexes both with one another and with small ligands (e.g., enzyme inhibitors or DNA-binding drugs) is provided mostly by X-ray crystallography and nuclear magnetic resonance (NMR) techniques.

Many investigators strive to shed light on the forces that stabilize nucleic acids and on those that account for the binding properties of these biopolymers in solution. Computer simulations and theoretical calculations provide information that can be of value for understanding the nature and relative magnitude of these forces. In the case of ligand binding, the aim is to establish a theoretically sound basis for the observed binding specifities, for the measured differences in binding affinity among related ligands, and for the conformational changes that accompany complex formation. A number of predictions can then be made that can be directly tested by experiment.

From the pharmacological standpoint, many effective anticancer drugs in clinical use interact with DNA, and many do so through intercalation. Strictly speaking, the sequences to which these drugs bind have to be viewed as acceptor sites rather than true receptors, as their biological effects are exerted only through interference with the recognition and function of DNA-binding proteins (e.g., polymerases and topoisomerases), which are probably the major determinants of selectivity at the cellular level. Improving our understanding of both the rules that govern the sequence-specific binding to DNA of these and related small ligands and the underlying mechanisms that ultimately lead to cell death could be of considerable help in creating novel chemotherapeuticals using state-of-the-art drug design strategies. To achieve this elusive goal, a close collaboration between workers from many disciplines using different theoretical and experimental ap-

1046-2023/98 $25.00

proaches is now more than ever an essential prerequisite.

Many aspects of intercalation and stacking interactions have come under close scrutiny by theoretical methods in recent years. The initial studies were limited to models with single bases, nucleosides, or, at best, dinucleotides, and solvent effects were neglected. The advent of new methodologies and spectacular advances in high-performance computing and computer graphics have provided added impetus to the pace of research in this area, and progress is booming. A brief historical perspective and a personal view of these advances are given below, expanding somewhat on some recent results from my laboratory.

1. INTERCALATION AS A MODE OF BINDING TO DNA

The conformational flexibility inherent in DNA is well reflected in the intercalation process, the basic postulate of which is that a planar aromatic or heteroaromatic ring system becomes inserted between adjacent base pairs of a DNA molecule without disturbing the overall stacking pattern. On binding of an intercalator, the vertical separation between base pairs (or rise) is increased (Fig. 1) and the double helix is partially unwound. Unwinding involves changes in the degree of rotation between successive base pairs (twist angle) as well as distortions in the sugar–phosphate backbone.

The concept of intercalation was first introduced to explain the reversible and noncovalent binding of some acridines to DNA (4). Many typical intercalators consisting of three to four fused rings have a strong tendency to absorb light at wavelengths less than 300 nm; for this reason their ring systems are usually referred to as chromophores. This class of DNA-binding ligands (Fig. 2) includes dyes (e.g., ethidium), mutagens and carcinogens (e.g., aminoacridines and the reactive metabolites of some polycyclic aromatic hydrocarbons), and antitumor antibiotics (e.g., actinomycin and the anthracyclines), which bind to DNA with varying degrees of sequence selectivity. Ligand binding can affect many biological properties of DNA, including replication and transcription; in the case of intercalators, these effects are modulated not only by the chemical characteristics of the chromophore but also by the presence and position of substituent moieties attached to it.

The lengthening of DNA on intercalation translates into an increase in intrinsic viscosity and overall stiffening which is apparent by a decreased sedimentation coefficient. Typical Scatchard plots representing the binding of simple monointercalators to DNA are concave and extrapolate to a maximum number of bound ligand molecules per nucleic acid base pair of roughly

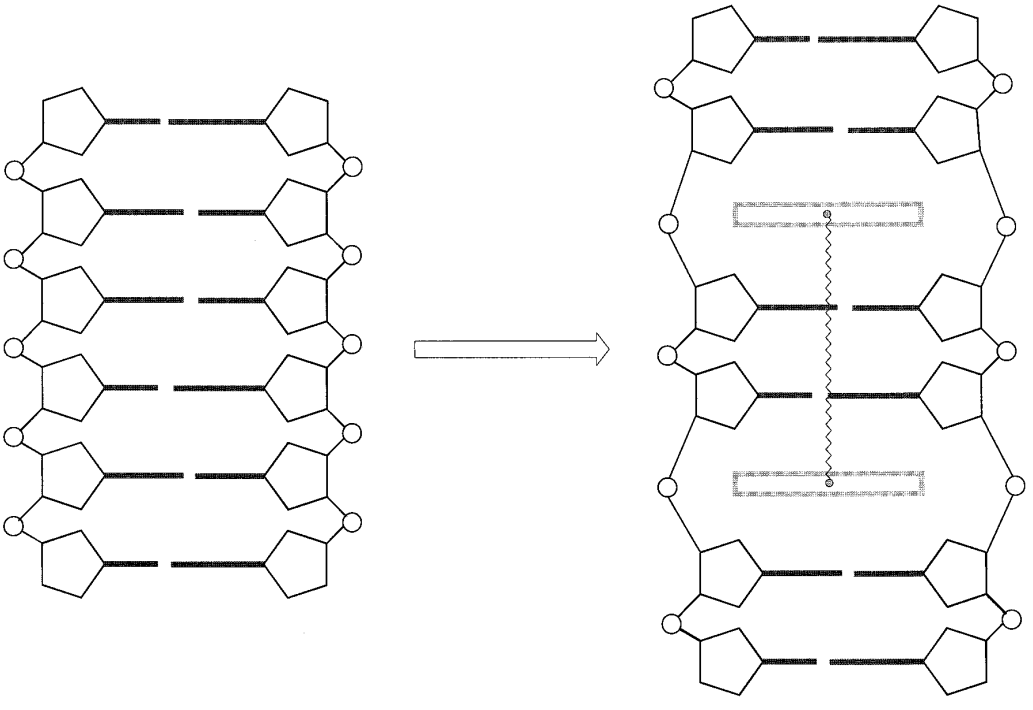

FIG. 1. Intercalation model. Bifunctional agents can be obtained by bridging two chromophores through a suitable linking chain.

0.5. This corresponds to a double helix in which filled and empty binding sites alternate (Fig. 1), and illustrates the so-called nearest-neighbor exclusion principle (1). The fact that both neighboring sites on each side of the intercalation site remain empty is an extreme example of anticooperativity, and a structural explanation was found when the details of intercalation were unveiled by X-ray crystallography.

In the crystal structures of self-complementary dinucleoside monophosphates of the Pyr(3′–5′)Pur type complexed to simple monointercalators (e.g., the alkaloid ellipticine or fluorescent dyes such as acridine orange and ethidium bromide; Fig. 2), the deoxyribose sugar ring of the pyrimidine nucleoside is usually in the C3′-*endo* conformation (characteristic of the monomer unit in A-type DNA), whereas that of the purine nucleoside is predominantly found in C2′-*endo* (characteristic of B-type DNA). The structural parameters derived from this early work stimulated the building of intercalation models in polynucleotides, and a general mode of intercalation was proposed involving formation of a kink, that is, a sharp, localized bend that may provoke the appearance of a wedge-shaped entry notch accessible from the minor groove (5). Initial formation of a kink would be the result of unstacking a single base step and rolling open the base planes, and could be related to the alteration of the normal C2′-*endo*-deoxyribose sugar ring puckering in B-DNA to a mixed puckering pattern of the type C3′-*endo*(3′–5′)C2′-*endo*. These changes do actually bring about DNA unwinding and changes in the direction of the helical axis. However, to open up two adjacent base pairs in A-DNA and create an intercalation site, it is necessary only to increase the value of the glycosidic torsion angle (from 14° to 80°) together with the torsion angle around C5′–O5′ (from 175° to 225°) at the 3′ end of each strand

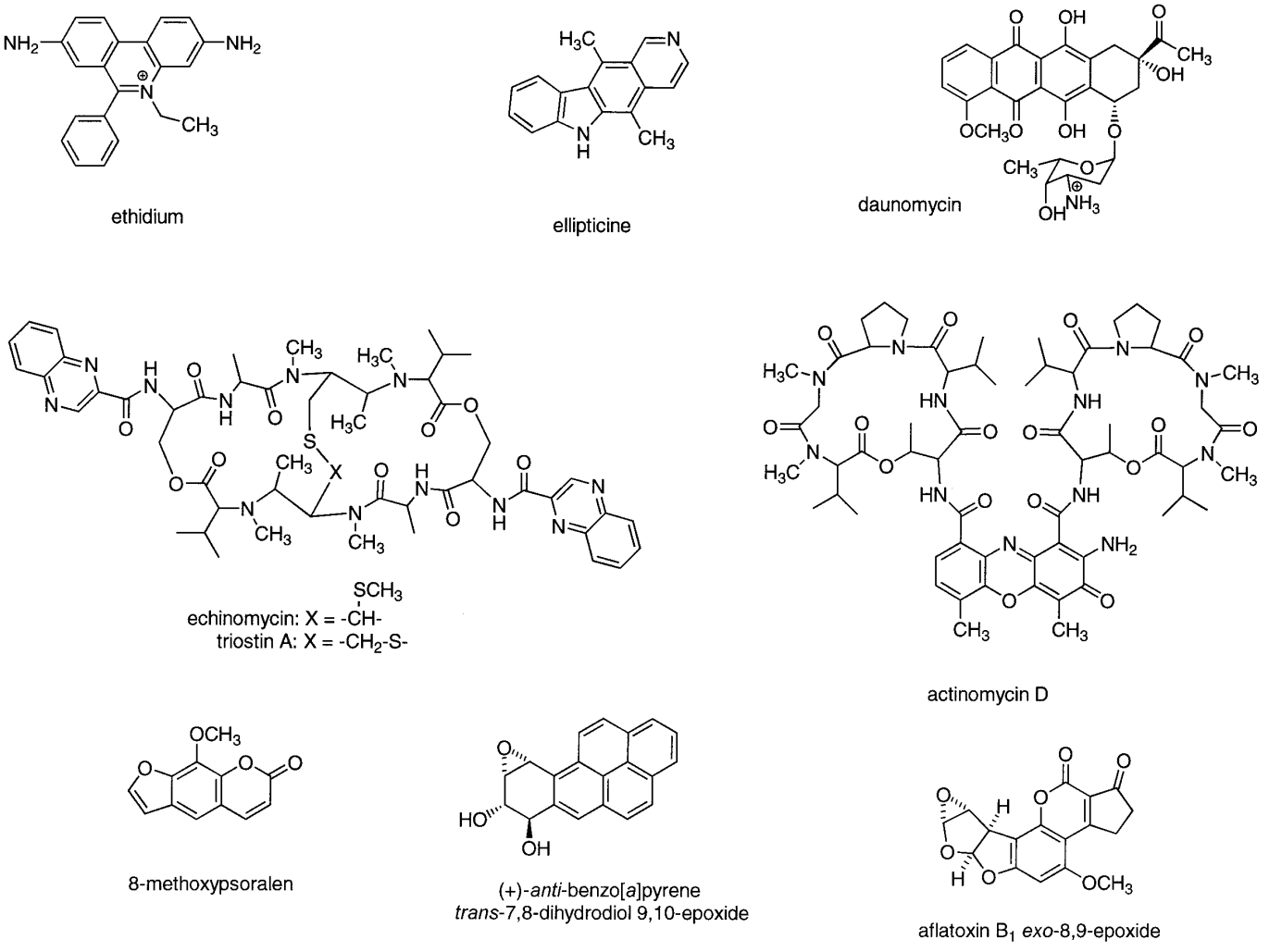

FIG. 2. Chemical structures of some natural and synthetic mono- and bisintercalators: ethidium, ellipticine, daunomycin, echinomycin, and triostin A, actinomycin, 8-methoxypsoralen, (+)-*anti*-benzo[*a*]pyrene *trans*-7,8-dihydrodiol 9,10-epoxide, and aflatoxin B_1 *exo*-8,9-epoxide.

(6). These are precisely the geometrical parameters that differ most from the average of those torsion angles in uncomplexed dinucleoside monophosphate structures.

Kinking of the DNA by minor groove intercalation has actually been observed not only on binding of small ligands but also in complexes with proteins involved in transcriptional regulation (7). A single base step is pried open by a wedge comprising two to five side chains on the protein surface whose leading edge consists of a single hydrophobic residue (Ile, Leu, Phe, or Trp), but alternative modes of intercalation are possible in different functional environments, as shown by the base pair rearrangement that takes place in the GGCC recognition element of a DNA molecule bound to a methyltransferase (8). There is also good experimental evidence that the Tyr residue of the heptad repeat unit SPTSPSY of RNA polymerase II can intercalate (9), and the YSPTSPSY peptide has indeed been found to adopt nonrandom conformations in solution that are consistent with a bisintercalation binding mode (10).

Despite the evidence that the conformational changes associated with intercalation of small ligands into DNA were not restricted to just the two base pairs above and below the intercalation site, inherent uncertainties in molecular structures modeled from dinucleoside monophosphate–intercalator complexes due to end effects precluded a detailed analysis of changes extending beyond first neighbors. This long-range influence was first observed in the 2:1 complex of the anthracycline antibiotic daunomycin (Fig. 2) with the hexanucleotide $d(CGTACG)_2$, and is in harmony with the nearest-neighbor exclusion principle. The aglycon chromophores of daunomycin are intercalated between the CpG sites at both ends of the duplex so that the amino sugars occupy the minor groove. Interestingly, unwinding in this complex is not associated with the bases on either side of the intercalating ring system but with the next two base pairs, and is limited to just 8° per daunomycin molecule (11).

The general Pyr–Pur sequence preference exhibited by simple monointercalators is overridden in the case of actinomycin D (Fig. 2), which binds with high affinity to GpC steps. A 1:2 crystalline complex of this antibiotic with deoxyguanosine (12) revealed specific hydrogen bonds between the peptide group of the L-threonines and N-3 and amino N-2 of the guanines stacked above and below the phenoxazone ring system. By placing a deoxycytidine-5′-monophosphate opposite the deoxyguanosine to form a G:C base pair and connecting adjacent C and G residues immediately above and below the chromophore by a phosphodiester linkage, a GpC intercalation step was created. Further extension by adding two base pairs on either side led to a hexanucleotide complex which appeared to be stabilized by both the intercalation of the chromophore and interactions between the peptide subunits of actinomycin and the minor groove of the DNA helix. This proposal, initially supported by Corey–Pauling–Koltun (CPK) space-filling models (12) and computer-assisted molecular modeling (13), reconciled experimental data suggesting that recognition of the GpC site took place through stacking interactions as well as by specific hydrogen bonding to both guanine bases. These models have stood up well in the light of subsequent crystallographic work (14).

The use of triplex-forming oligonucleotides has opened a completely new field of studies aimed at the design of sequence-specific gene regulators. Intercalators such as acridine or benzo[*e*]pyridoindole and benzo[*h*]pyridoquinoxaline derivatives have been covalently linked to these oligonucleotides to enhance their DNA binding properties, confer on them increased resistance to nuclease attack, and improve their cellular uptake (15). The intercalating moiety has been attached not only at one of the ends of the oligonucleotide but also at internucleotide positions, and molecular modeling tools were found useful in this latter case to predict the optimal length of the linking chain (16). Given that, to intercalate into a triplex, two base triplets rather than two base pairs (as in duplex DNA) must be unstacked, the subsequent stabilizing interactions that must compensate for the higher-energy penalty associated with this process are likely to depend on a different set of requirements (as to size, shape, and electronic characteristics of the intercalator) from those taking place in double-stranded helices. In fact, the affinity of ethidium and other simple intercalators for triplex DNA has been shown to be lower than that for duplex DNA (15) whereas benzo[*e*]pyridoindole and a series of substituted imidazothioxanthones (17), among other chromophores, bind preferentially to triplexes.

2. BINDING ENERGETICS AND MODEL BUILDING

The binding of an intercalating ligand to a particular DNA sequence depends on multiple factors. The terms that need be considered include appropriate steric and electronic complementarity of the contact surfaces, hydrophobic effects, ease of unstacking of the DNA, nature of the bases flanking the intercalation site, possible dehydration of the grooves, release of counterions, and entropic effects. Elucidating the magnitude and relative importance of all the factors involved in the energetics of complex formation is essential for a thorough understanding of the rules that govern high-affinity binding and selection of a particular sequence.

The interaction energy between a ligand L and its receptor R (ΔE^{L-R}) is determined as the difference between the energy of the complex (E^{L-R}) and the sum of the individual energies of ligand ($E^{L'}$) and receptor ($E^{R'}$) in the unbound state. If the conformational energy changes undergone by both the ligand and the target molecule on binding are neglected (as is often done when series of ligands are compared), the interaction energy can be approximated by

$$\Delta E^{L-R} = E^{L-R} - (E^{L} + E^{R}),$$

where E^{L} and E^{R} are the intramolecular energies of ligand and receptor in the bound state.

When R is a DNA target molecule and L is an intercalating ligand, insertion of the chromophore requires that the bases that make up the intercalation site separate, which brings about conformational changes in the sugar–phosphate backbone. In addition, as a result of the increased distance between the stacked bases, there is a loss of dispersion and electrostatic intramolecular energy in the DNA that needs to be compensated by the interactions established between the ligand and the DNA if intercalation is to occur. It is therefore important to characterize stacking interactions both in nucleic acid bases and in their complexes with planar heteroaromatics, and this has been attempted by both quantum mechanical and molecular mechanical methods. Within the former category, which aims at solving the Schrödinger equation, *ab initio* calculations should in principle offer the best theoretical level and the most accurate results, especially when including electron correlation effects. The quality of the results depends heavily on the basis set of atomic orbitals used (and whether or not polarization functions are included) and on the extent of correlation energy covered (18). Their primary and most serious limitation, however, is the size of the systems to which they can be applied, especially when the most accurate methods are employed. For this reason most *ab initio* studies have been performed on the free nucleic acid bases, and both hydrogen bonding and base stacking interactions have been studied. It has been noted (1), however, that N-methylated bases are more appropriate for this sort of study as the absence of the glycosidic C–N bond may introduce a significant perturbation of the electronic states in the free isolated bases (e.g., tautomerization). When used to evaluate interaction energies, it is important to realize that conformations taken from crystal structures or standard DNA geometries are not real minima on the gas-phase potential energy surface (19), and care should be taken to correct for the basis set superposition error (18, 19). An important outcome of these studies has been that electrostatic interactions control the orientational dependence of the stacking energy although the stability of the stacked bases originates in the dispersion interaction. In addition, it was found that the sequence-dependent variability of the total interaction energies is relatively small for the stacked base pairs, whereas intrastrand and interstrand contributions vary more significantly (18).

Less computationally demanding than conventional *ab initio* methods are those based on density functional theory (DFT), which have become very popular in the past few years. Although current DFT calculations provide good dipole moments and reliable charge distributions for DNA bases, their use is not recommended for the study of interactions between stacked base pairs as present DFT functionals are not able to properly represent the dispersion energy that originates mainly from electron correlation effects (18).

Much more computationally feasible and widely applicable are semiempirical methods (e.g., MNDO, AM1, PM3) that neglect many of the integrals by introducing a number of empirical parameters. Although useful for a variety of purposes, including the calculation of electrostatic potential-derived atomic charges (20), they use Hartree–Fock methods and therefore also ignore dispersion attraction. Thus, they cannot be used to estimate base stacking interactions accurately, even though they appear to assess the angular dependency of this energy correctly (21).

Molecular mechanics force fields represent the main category of empirical methods: molecules are treated as classical systems composed of atoms held together by bonds. The total energy is calculated as the sum of bond stretching, bond bending, bond torsion, and attraction and repulsion between nonbonded atoms. The nonbonded interaction is divided into a Lennard–Jones part (van der Waals interaction, calculated with pairwise transferable potentials) and a Coulombic term (electrostatic interaction, calculated using atom-centered point charges). Some of these empirical potentials (e.g., AMBER 4.1) (22) have been shown to provide hydrogen bonding and stacking stabilization energies of nucleic acid base dimers in reasonable agreement with those calculated *ab initio* incorporating electron correlation effects (18, 19). This type of assessment is critical for checking the validity of the parameters incorporated in new-generation force fields that are widely used in simulations of oligonucleotides and their complexes with ligands.

Essentially two classes of procedures have been used on the computer for constructing intercalation geometries in oligonucleotides using molecular models in combination with computational chemistry tools. One alternative is to extend a dinucleotide intercalation site by fitting fragments of a canonical double helix to each of the base pairs. The phosphodiester bonds are then added, and the structure is progressively refined by

gradually minimizing the empirical energy function associated with the molecule using a molecular mechanics force field, so as to dissipate smoothly strains that inevitably accompany the initial model building. An alternative option is to start with a canonical double helix and use internal restraints in conjunction with energy minimization techniques to force the base pairs into an intercalation geometry. Following a mixed approach, the crystal structure of a *cis-syn* monoadduct of thymine and the photosensitizing intercalator 8-methoxypsoralen (Fig. 2) was used to construct a diadduct which was subsequently incorporated into a DNA duplex (23). The resulting model predicted a substantial kink at the site of covalent damage which was later proved to be correct (24). Nevertheless, a problem with modeling techniques based on energy minimization is that they are not very efficient at detecting minima far removed from the starting structure, and the final model may not be a good reflection of reality. An extreme example is the possible stabilization of extrahelical structural elements by intercalators, as recently observed in the complex of a metalloporphyrin with the hexamer $d(CGATCG)_2$ (25), which was not anticipated by molecular modeling techniques.

To explore the conformational space more efficiently, molecular dynamics (MD) or Monte-Carlo (MC) simulations can be performed. If the goal is only to obtain a representative conformational sample, and no equilibrium thermodynamic properties or free energy changes are going to be computed, generation of Boltzmann-weighted ensembles is not required (26). If the simulations are carried out *in vacuo,* some restraining function is needed to prevent large-scale deformation of the DNA, particularly in the absence of experimental information (27); distance restraints to enforce interstrand hydrogen bonds, positional restraints, or both are relatively common. With the advent of two-dimensional NMR spectroscopy and the development of sequential resonance assignment strategies (28), virtually complete and unambiguous resonance assignments for oligonucleotides and their complexes with different ligands can be obtained (29). The interproton distances derived from nuclear Overhauser enhancement (NOE) measurements, usually defined by lower and upper bounds, can then be incorporated into a MD simulation by adding to the empirical energy function an expression for the difference between observed and computed data which is equal to zero if the model matches the data perfectly. Likewise, information about NOESY volumes and *J*-coupling constants can also be incorporated into the simulation. Structural refinement of model-built DNA–intercalator complexes by using MD/simulated annealing (26, 30) protocols restrained by NMR data has become widespread and is not restricted to drugs. An interesting application has been to the determination of the complexes of carcinogenic metabolites of polycyclic aromatic hydrocarbons (Fig. 2), e.g., (+)-*trans*-7,8-dihydroxy-*anti*-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene (31), or fungal products, e.g., aflatoxin B1 *exo*-8,9-epoxide (32), with DNA, to which they bind covalently following intercalation.

The complete neglect of solvent effects limits the accuracy of the biomolecular properties obtained from *in vacuo* simulations, especially in the absence of NOE-derived restraints, even though use of distance-dependent dielectric models or reduced charge models is common for dampening electrostatic interactions. On the other hand, the large computational cost associated with the incorporation of explicit water molecules and counterions in the simulations may preclude extensive searches of conformational space. A standard procedure to speed up the computations and to minimize edge effects in MD simulations of fully solvated systems is to treat them under periodic boundary conditions (33), so that they are surrounded by identical translated images of themselves. Care must be taken to ensure that the size of the box is sufficiently large to avoid interactions between molecules and their periodic images (this, on the other hand, is desirable for simulations of condensed phases or crystals). Unrealistic behavior can still be observed in simulations of DNA due to the high local ionic density of this molecule, the long range of electrostatic forces, and the use of a finite cutoff distance in the evaluation of pairwise Coulombic interactions to shorten the list of nonbonded atoms. This problem can be circumvented by application of the recently developed particle mesh–Ewald method (34), which allows DNA simulations to be extended to the nanosecond time scale without significant structural disruption. The method has been shown to be able to accommodate interstrand phosphate–phosphate repulsions when tested on a DNA dodecamer in a crystal unit cell (35), and has been used to provide a reliable model of the d(TAT) triple helix in aqueous solution (36).

Interaction energies in ligand–DNA complexes are commonly calculated using molecular mechanics either on a single structure that is usually taken to represent the ensemble average of each complex or along a MD simulation. These approaches, however, neglect the effect of the discontinuity between the low-dielectric solute and the high-dielectric solvent and do not take into account that the electrostatic interaction term is opposed by the desolvation of parts of the ligand and the DNA binding site on complex formation. Macroscopic solvent models based on the Poisson–Boltzmann (PB) equation can efficiently describe the dynamically averaged dielectric behavior of the solvent environment, and continuum methods based on the PB equation are currently used to compute the unfavorable desolvation free energies that accompany the removal of polar and

charged groups from water, as well as the screening effects of the surrounding solvent (37, 38).

From the perspective of the ligand, double-helical DNA may be viewed as a rather regular exterior array of phosphate charges surrounding an interior lattice of stacked base pairs which are accessible to ligand binding from either the major or the minor groove. A nonspecific electrostatic binding mode is expected between the phosphates and positively charged ligands, which can then be followed by redistribution to higher-affinity sites along the DNA helix on recognition of particular functional groups of the stacked bases. For actinomycin and echinomycin (Fig. 2), it has been suggested that the lack of a net positive charge could be compensated by the high dipole moment calculated for these molecules, which arises from an asymmetric distribution of positive and negative electrostatic potential regions (39). Binding to DNA of an intercalating ligand with a net positive charge (located either on the chromophore or on an attached side chain) will favor the release of bound counterions, so that the observed binding constant is likely to depend on the ionic strength. These electrostatic aspects of ligand binding have been treated according to counterion condensation theory or PB models. In the former, two distinct populations around DNA are considered: a condensed layer, vicinal to the nucleic acid and invariant to salt concentration, and a more diffuse, salt-dependent ion atmosphere (40). The PB model, instead, assumes that the ions are continuously distributed in a single population defined by a Boltzmann factor governed by the electrostatic potentials; in this case, the binding of a cationic ligand leads to the redistribution of ions around each interacting molecule. The nonlinear solution to the PB equation has been successfully used to describe the change in pK_a of the intercalator 3,8-diamino-6-phenylphenanthridine on binding to a DNA dodecamer as well as the dependence of the pK_a shift on bulk salt concentration (41).

3. FROM MONO TO BIS: STRUCTURE-BASED DESIGN OF BIFUNCTIONAL INTERCALATORS

Interest in bifunctional, or even polyfunctional, intercalators originally stemmed from the possibility of enhancing the binding constant over that of the corresponding monomers. On the other hand, increasing the size of the site occupied by the ligand can afford greater opportunities for imposing selective binding to defined nucleotide sequences: there are only 10 unique dinucleotide steps at which monointercalation can take place, but for a binding site covering 4 base pairs the number of distinguishable sequences is 136, and this number increases to 2080 if the binding site size is 6 base pairs. The main problem is that it is not at all clear what sequences to target.

Some well-known fluorescent intercalators, e.g., ethidium and acridine, were early exploited to synthesize homo- and heterodimers (Fig. 3) which behave as bifunctional intercalators (42). Other dyes, such as thiazole orange (TO) and its benzoxazole analog oxazole yellow (YO), have also been bridged through a biscationic linker resembling spermine and spermidine to produce dimeric compounds (TOTO and YOYO) that form stable and highly fluorescent complexes with double-stranded DNA (43). Following a trial-and-error strategy, TOTO was found to exhibit site selectivity in its binding to DNA, in contrast to ethidium homodimer and other previous synthetic bisintercalators. The capacity for preferential binding was attributed to the thiazole orange chromophores and not to the linker as binding of spermidine-like molecules had been previously shown to be nonselective (44).

Although bisintercalation constitutes a powerful way in which a relatively small molecule can interact with DNA and modify its structure, the complete structural requirements for bifunctional intercalators to show antitumor activity are not completely understood. An empirical approach is to dimerize monointercalators that show good antitumor activity. Thus, dimers that incorporate the 6*H*-pyridocarbazole chromophore of ellipticine (Fig. 2) or the related 10-methoxy-7*H*-pyridocarbazolium have been prepared (45). The antitumor activity of these dimers has been shown to depend strongly on the nature of the linking chain and its site of attachment on the ring system. For ditercalinium (Fig. 3), probably the most interesting derivative, a novel mechanism of action was described involving enzymes that locate and repair damaged sites within large domains of DNA, and particularly the nucleotide excision repair system (46). Interestingly, DNA binding affinity and antitumor activity were found not to be correlated in this series. Furthermore, dimerization strategies do not always lead to DNA bisintercalators, as illustrated by the antineoplastic imidazoacridinones. Although several bisimidazoacridones have been described with potent and selective activity against colon cancer (47), their interaction with DNA does not appear to involve bisintercalation (48). On the contrary, the linking of two naphthalimide groups with a polyamine bridge led to bisnaphthalimides such as LU-79553 (Fig. 3) (49) and DMP-840 (50), both currently in clinical trials and displaying an improved pharmacological profile over those of the intercalating mononaphthalimides amonafide and mitonafide. DMP-840 has been shown to interact directly with DNA and produce single-strand breaks. LU-79553 has been conclusively shown to bisintercalate (presumably from the major groove), and its footprinting profile suggests a definite preference for GpT (ApC) and TpG (CpA) steps,

which is reminiscent of the recognition pattern observed with the anthracycline nogalamycin (51). Efforts are currently underway to elucidate the solution structure of a high-affinity oligonucleotide–LU-79553 complex by NMR spectroscopy (52). Interestingly, some recently reported polyintercalating molecules that incorporate a number of units of the related 1,4,5,8-naphthalenetetracarboxylic diimide chromophore connected in series by peptide linkers have been shown to bind in a cooperative manner to the major groove of double-stranded DNA, with a clear preference for GC over AT sequences (53).

The conception and design of new drugs now more than ever rest on the availability of three-dimensional coordinates of the target macromolecule. The well-known 2:1 stoichiometry observed in the complexes of several DNA hexanucleotides with different anthracycline antibiotics offers the opportunity to link the two drug molecules through a suitable spacer. When the 3′ amino groups of the daunosamine moieties of two

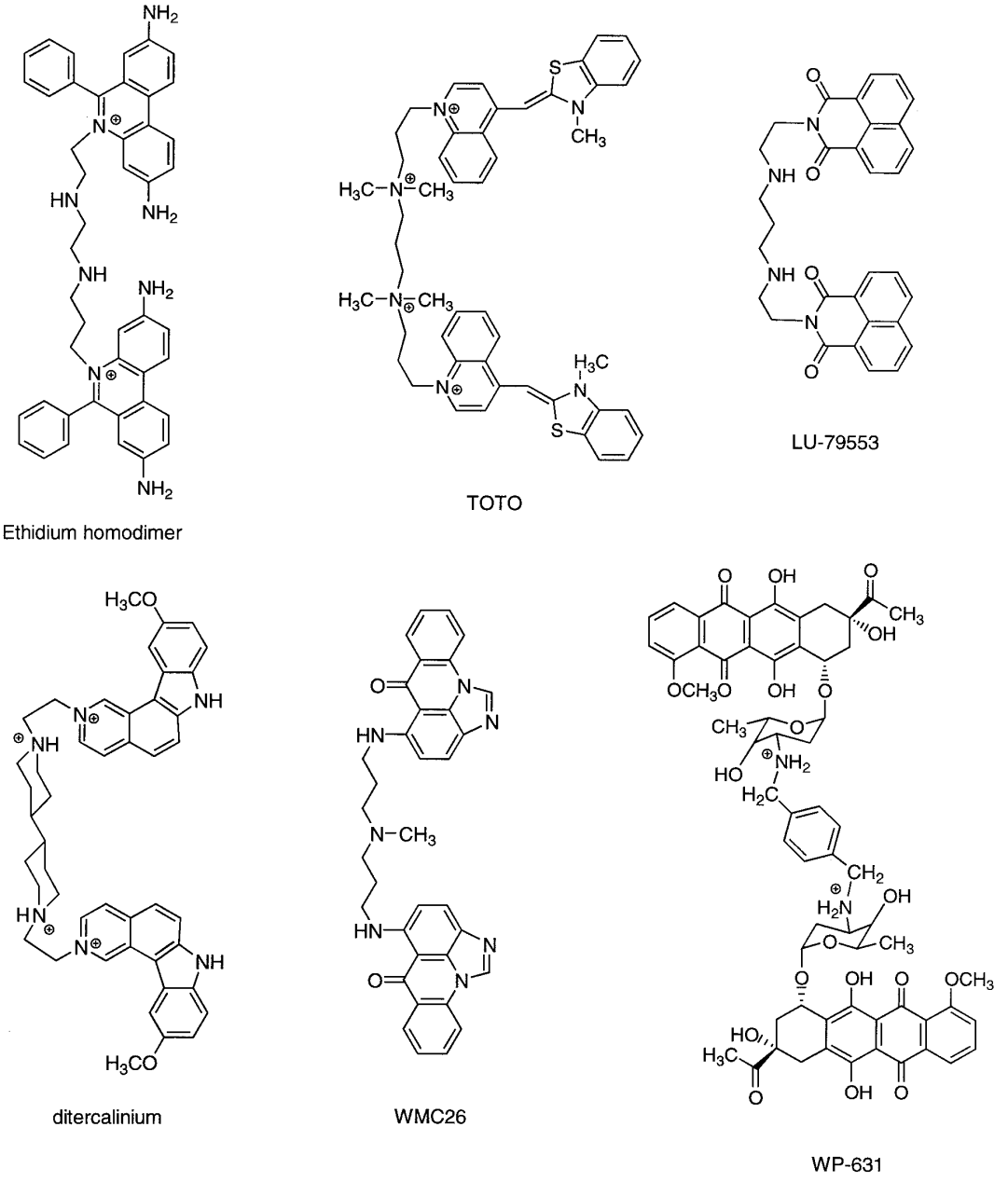

FIG. 3. Examples of dimerization strategies: ethidium homodimer, thiazole orange homodimer (TOTO), bisnaphthalimide (LU-79553), ditercalinium, bisimidazoacridinone (WMC26), and bisdaunomycin (WP631).

daunomycin molecules located in the minor groove were joined by a *p*-xylenyl tether (Fig. 3), a novel bisdaunorubicin (WP631) was developed that behaves as an effective bisintercalator and shows promising antitumor activity (54). The complexes of WP631 with the oligonucleotides $d(CGATCG)_2$ and $d(ACGTACGT)_2$ did indeed show the predicted binding mode, with the chromophores intercalated at the CpG steps embracing the central tetranucleotide. However, when a daunomycin analogue bearing the amino group on C4′ was dimerized with the same linker to yield WP652, the complex of the drug with $d(TGTACA)_2$ showed the chromophores sandwiching the GpT step in a binding mode reminiscent of that of the natural antibiotics triostin A and echinomycin (55).

Bisintercalators with rigid linkers have also been synthesized, aimed at crosslinking different duplexes. In addition to being able to knot and catenate DNA, some of them have been shown to bind in a sequence- and structure-specific manner to X- and Y-shaped DNA molecules (56). The dye Stains-All has also been reported to color branched DNA molecules differently from linear duplexes (57). Although molecular models of these 4H and 3H junctions have been built (58, 59), atomic details pertaining to their interaction with this sort of intercalators remain to be determined.

4. MOLECULAR BASIS OF SPECIFICITY IN BINDING OF INTERCALATORS TO DNA

When DNA–ligand complexes are subjected to the action of reagents that degrade DNA, such as DNase I, and the DNA is electrophoresed on a high-resolution polyacrylamide gel, sites protected by the ligand from cleavage show up as blank spaces on the autoradiograph of the gel. These so-called footprints are taken as evidence of preferential binding, and footprinting studies are very useful in showing the sequence preference of many antitumor antibiotics that bind to DNA or synthetic polynucleotides (60). By use of these techniques, characteristic patterns of nucleotide sequence selectivity were early recognized for a family of quinoxaline antibiotics and derivatives that were shown to intercalate bifunctionally into DNA via the minor groove (61). A starting point for confirming the proposed binding mode as well as the origin of specificity for members of this class of compounds was provided by the crystal structures of triostin A (Fig. 2), whose canonical binding site is CpG, and its demethylated derivative TANDEM, which binds with high affinity to TpA steps. Although the quinoxaline-2-carboxamide chromophores do not appear to have intercalative properties per se, they are held by the cyclic depsipeptide backbone in a position such that simultaneous intercalation is possible. This high degree of preorganization strongly suggested formation of complexes with DNA in which two base pairs are sandwiched between the chromophores and sequence specificity is achieved through the establishment of hydrogen bonding interactions between the alanine residues of the depsipeptides (Fig. 2) and the bases in the minor groove (61). Definite proof that the proposed binding mode was correct was provided by X-ray and NMR studies on complexes between short oligonucleotides and echinomycin, triostin A, and TANDEM (62). The strong affinity of echinomycin and triostin A for CpG steps is decisively determined by a number of hydrogen bonds between its alanine residues and both the N-3 and 2-amino groups of guanine. One striking feature of some of these complexes was the sequence- and context-dependent formation of a Hoogsteen hydrogen bonding scheme (Fig. 4) in the A:T base pairs adjacent to the CpG binding sites in oligonucleotides such as $d(A\mathbf{CG}T)_2$, $d(\mathbf{CG}TA\mathbf{CG})_2$, $d(A\mathbf{CG}TA\mathbf{CG}T)_2$, and $d(A\mathbf{CG}TATA\mathbf{CG}T)_2$, but not in $d(T\mathbf{CG}A)_2$, $d(T\mathbf{CG}AT\mathbf{CG}A)_2$ and $d(AAA\mathbf{CG}TTT)_2$. Despite the structural evidence, the reason why some purine bases rotate 180° about the glycosidic bond so as to adopt a *syn* orientation relative to the sugar remained unclear. Detailed computational analysis of the complexes of echinomycin and the DNA tetramers $d(A\mathbf{CG}T)_2$ and $d(T\mathbf{CG}A)_2$ in aqueous solution using MD simulations allowed us (63) to propose that (i) the dipole moment of an A:T base pair increases from a value of about 2 D in the Watson–Crick conformation to more than 5 D in a Hoogsteen arrangement (Fig. 4), (ii) the quinoxaline-2-carboxamide chromophore of echinomycin and triostin has a high dipole moment of about 4 D (Fig. 4), later confirmed by experiment (64), and (iii) when Hoogsteen base pairs are observed, stabilization most likely arises from an improved electrostatic interaction between these stacked systems due to the antiparallel arrangement of their respective dipole moments (Fig. 5).

Drug-induced Hoogsteen pairing is not restricted to A:T base pairs. In the crystal structure of $d(G\mathbf{CG}TA\mathbf{CG}C)_2$ with two triostin A molecules bound (62), the terminal G:C base pairs flanking the quinoxaline rings also display this alternative scheme. Moreover, comparison of the solution complexes of echinomycin with either $d(GCGC)_2$ or $d(CCGG)_2$ by NMR spectroscopy reveals that the terminal G:C pairs in the $d(G\mathbf{CG}C)_2$ complex, which are Watson–Crick paired at neutral pH, shift to Hoogsteen pairing as the pH is lowered, whereas no Hoogsteen base pairs are observed in the $d(C\mathbf{CG}G)_2$ complex at any pH (65). The need for an acidic pH is related to the fact that a Hoogsteen hydrogen bonding mode in G:C pairs requires the protonation at the N-3 of the cytosine base. The observed conformational preferences in these complexes also appear to arise mainly from differential stacking interac-

tions in which the electrostatic component is shown to play a dominant role (66).

A rather unexpected result from the analysis of all of these complexes was that the dipole moments of the G:C base pairs sandwiched between the chromophores of the drug, which are also notably high ($\mu \approx 5$ D) in a Watson–Crick conformation (Fig. 4), give rise to an unfavorable electrostatic stacking interaction with the quinoxaline-2-carboxamide groups. This led us to think that modulation of the dipole moments of the intercalating chromophores could be an additional element to be taken into account in the recognition process; in fact, different studies had already shown that the sequence recognition properties of these substances can be radically altered by defined changes not only in the depsipeptide ring structure (as in going from triostin A to TANDEM) but also in the chromophores themselves [e.g., bis-7-chlorotriostin A binds very tightly to poly (dA)·poly(dT)] (67). Following an approach equivalent to site-directed mutagenesis in proteins, the structure–affinity relationships for this class of intercalating ligands were expanded by introducing site-specific modifications into a DNA fragment. These experiments showed that whereas removal of the exocyclic amino group of guanine (by replacement of this base with hypoxanthine) leads to the predicted loss of echinomycin's specificity for CpG steps, introduction of an extra amino group in the minor groove of AT regions [by replacing adenine with 2,6-diaminopurine (DAP)] favors high-affinity binding of echinomycin to any pyrimidine–purine (YpR) combination other than the usual CpG step (68).

To explore to what extent hydrogen bonding and stacking interactions play a role in determining the binding preferences of echinomycin, we generated a series of molecular models for the 1:1 complexes of echinomycin with standard and modified DNA sequences to evaluate the different contributions to their relative stability (64). The complexes of echinomycin with 13 double-helical DNA hexamers were built based on the solution structure of a complex between CysMeTANDEM and d(GA**TA**TC)$_2$ (69). The choice of this template was guided by the fact that this is the only experimentally determined structure formed between a quinoxaline antibiotic and an oligonucleotide in which all the base pairs are Watson–Crick paired. The models were intended to represent sections of longer DNA fragments for which there is no evidence of Hoogsteen pairs at internal positions surrounding CpG binding sites. A common terminal G:C base pair on both 5′ and 3′ ends

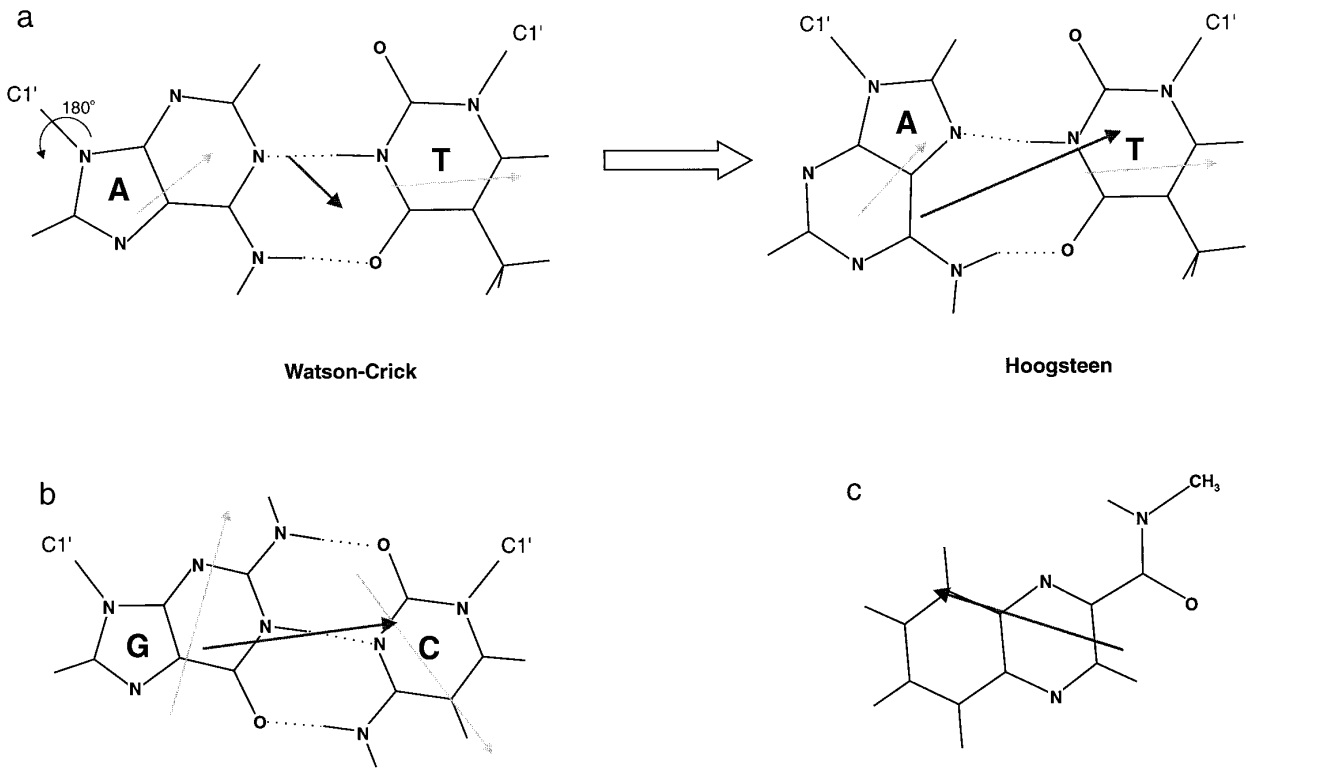

FIG. 4. Dipole moments (boldface arrows) of (a) an A:T base pair in both Watson–Crick and Hoogsteen pairing arrangements, (b) a G:C base pair in Watson–Crick conformation, and (c) the quinoxaline-2-carboxamide chromophore of triostin and echinomycin (63, 64). The dipole moments of individual bases are drawn as light arrows, and all vectors have their midpoints centered on the geometrical center of the system considered.

was added to the central tetranucleotide so as to avoid possible end effects on the central region where the most intimate interactions with the drug take place. The hexanucleotides studied were grouped into four different categories: (i) a first family of general sequence d(GA**XZ**TC)$_2$, where the central XpZ step was CpG, TpA, GpC, or ApT; (ii) a second subset of d(GX**CG**ZC)$_2$ sequences, where the canonical CpG central step was conserved and the X–Z combinations were T–A, G–C, and C–G; (iii) three modified DNA sequences in which inosine substituted for guanosine in either one or both strands [d(IA**CI**TC)·d(GA**CG**TC), d(GA**CG**TC)·d(IA**CI**TC), and d(IA**CI**TC)·d(IA**CI**TC)]; and (iv) three modified DNA sequences containing DAP (D) in place of adenine [d(GD**TD**TC)$_2$, d(GD**TG**TC)·d(GD**CD**TC), and d(GD**CD**TC)·d(GD**TG**TC)]. The whole set of sequences thus contained every combination of YpR and RpY binding sites experimentally probed by echinomycin and was expected to provide information not only on all the possible hydrogen bonding arrangements between the depsipeptide and the DNA atoms in the minor groove but also on the stacking interactions between the quinoxaline-2-carboxamide system and any DNA base pair.

Examination of the ligand–receptor interaction energies showed that the different hexamers could be readily clustered into two distinct families with respect to their affinity for echinomycin: a first subset of good binding sites presenting both a central dinucleotide step endowed with full hydrogen bonding capabilities in the minor groove and a base pair arrangement that gives rise to an overall favorable stacking interaction with the quinoxaline chromophores, and a second subset of excluded sites sharing poorer hydrogen bonding possibilities and overall weaker interactions with the quinoxaline-2-carboxamide systems. This result was in consonance with available data from footprinting experiments and supported our previous hypothesis that, in addition to the crucial intermolecular hydrogen bonds in the central region, an important role in modulating the binding specificity of this class of bisintercalating agents is played by the stacking interactions involving the quinoxaline-2-carboxamide chromophores of the drug and the DNA base pairs. The importance of the stacking term was most clearly seen when sequences with similar minor groove environments were compared (e.g., CpI vs TpA or CpG vs TpD) and detailed energy decomposition was performed. The important role played by the stacking interactions, which is partially masked by the more prominent hydrogen bonding interactions, manifests itself most dramatically at TpD and CpG central steps in the modified DNA containing DAP:T pairs in place of standard A:T base pairs. These two sequences present a similar minor groove environment, and the differences in their binding affinities for echinomycin are a reflection of the distinct stacking properties of DAP:T and G:C base pairs, which favor binding of this antibiotic to a TpD step over the canonical CpG step.

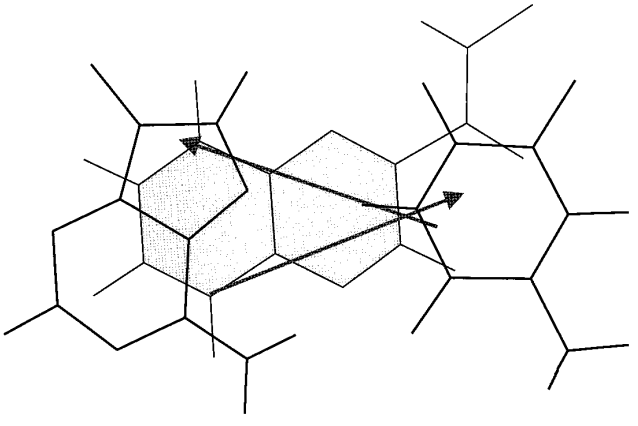

FIG. 5. Antiparallel arrangement of the dipole moments (arrows) of quinoxaline-2-carboxamide (shaded) and a terminal A:T base pair in Hoogsteen conformation as found in the complex of echinomycin with d(ACGT)$_2$ (63).

A qualitative description of the electrostatic stacking interactions reported above that improves on the simple representation of dipole moments can be formulated based on interpretation of the molecular electrostatic potentials (MEPs) of the stacked systems, i.e., hydrogen-bonded base pairs and drug chromophores. A color-coded graphical representation of the MEP on a plane close to the recognition surface is particularly useful to visualize charge distributions and to understand electrostatic complementarity and repulsive interactions (70). To this end, a cubic lattice is defined and at each point surrounding the molecule the electrostatic energy between the molecule and a unitary positive point charge is calculated (71). The resulting MEP values are then contoured as a two-dimensional slice 1.7 Å below or above the plane of the ring atoms as it is approximately at this distance that van der Waals contacts take place between successive base pairs or between a base pair and a drug chromophore. For a G:C pair, for example, the most negative MEP region is found in the surroundings of the N-7 and O-6 atoms of guanine, whereas the most positive MEP region is located around the cytosine ring.

This computational approach to the study of binding specificity and induced conformational changes was then extended to ditercalinium (Fig. 3), a synthetic bifunctional intercalator with antitumor activity. The pyridocarbazolium rings of this drug intercalate into each of the contiguous CpG steps of a d(CGCG)$_2$ oligonucleotide, as shown by both NMR spectroscopy (72) and X-ray crystallography (73). In contrast, binding of this drug to d(GCGC)$_2$ shows one ring intercalated at the CpG step and the other stacked on top of one of

the external base pairs (72). In both complexes, the bis(ethylpiperidinium) linker lies diagonally across the major groove, and the convex edge of each chromophores is oriented toward the minor groove. Visualization of the MEP calculated for the positively charged chromophore of ditercalinium and for a G:C base pair in Watson–Crick conformation initially suggested that the orientation of the intercalated chromophores would determine good electrostatic complementarity with the central G:C base pairs clamped by the drug but electrostatic repulsion with the G:C base pairs that make up the boundaries of the bisintercalation site, the orientation of which is reversed (70). When we analyzed the intermolecular interactions present in the crystal structure of ditercalinium bound to $d(CGCG)_2$ and in the modeled complexes in solution of ditercalinium and a more flexible analogue bound to $d(CGCGCG)_2$, we concluded that complex stabilization was strongly dependent on the electrostatic contributions to the stacking energies between the 10-methoxy-7*H*-pyridocarbazolium rings and the DNA bases sandwiched between them (74). These stacking specificities for the central GpC step are in concert with the attractive electrostatic interactions between the protonated nitrogens of the linkers (Fig. 3) and the N-7 and O-6 atoms of guanines which shape two regions of very negative electrostatic potential in the major groove. The summation of these effects is that the sandwiched base pairs are very effectively stapled by the drugs, irrespective of the hydrogen bonding potential of the linkers, which appeared to be substantially reduced in the aqueous medium due to competing interactions with water. On the other hand, the electrostatic interactions between the chromophores and the base pairs that delimit the bisintercalation site are repulsive and may be largely responsible for the notable structural features reported for these complexes in the solid state (73).

The consistent picture emerging from these computational studies is that intercalating chromophores should be viewed not as simple hydrophobic plates that become sandwiched indiscriminately between the base pairs but as molecular fragments that show marked orientational preferences in their stacking interactions. Due to the distinct electrostatic properties of G:C and A:T base pairs (70), it would appear that additional sequence selectivity can be achieved for bisintercalators by combining specific groove interactions with judicious modulation of the electronic properties of the intercalating chromophores.

In line with this reasoning is an early report that demonstrates a correlation between specificity of binding to G:C base pairs and polarizability of the chromophore in a series of monointercalating heteroaromatic ligands lacking special hydrogen bonding functions (75). As a matter of fact, the specificity of actinomycin D for guanine over the other common bases is retained in two derivatives that cannot form hydrogen bonds with the 2-amino groups of guanine: 2-aminophenoxazin-3-one and actinomine, an actinomycin analogue in which the cyclic pentapeptides (Fig. 2) have been replaced with *N,N*-diethylethylenediamine side chains (76). The recent finding that the sequence specificity of actinomycin drastically changes from the standard GpC site to any other 5′-purine–pyrimidine-3′ (RpY) step different from GpC when DAP substitutes for adenine in a tyrT DNA fragment (68) strongly supports the intervention of additional elements of recognition apart from the hydrogen bonds and prompted us to investigate the stacking interactions involved. In contrast to the quinoxaline-2-carboxamide chromophores of echinomycin, the phenoxazone-1,9-dicarboxamide chromophore of actinomycin D (Fig. 2) does not have a marked dipolar charge distribution. Instead, the two carboxamide groups attached perpendicularly to the aromatic ring system and in opposite orientations give rise to two alternate negative and positive electrostatic potential regions above and below the plane of the chromophore (39). To gain a more detailed understanding of the interactions that govern the sequence-specific binding of actinomycin to DNA, we built molecular models of 14 double-helical DNA hexamers with the drug intercalated at the central dinucleotide step, using as template the X-ray crystal structure of the actinomycin–$d(GAAGCTTC)_2$ complex (14). The complexes studied varied in composition at the central intercalation step, in flanking sequences, and in the presence of DAP:T pairs in place of G:C pairs, so as to contemplate all possible hydrogen bonding arrangements between the drug and the DNA minor groove as well as all possible combinations of stacking interactions (39). A good correlation was found between the intermolecular interaction energies calculated for the refined complexes and the relative preferences of actinomycin binding on standard and modified DNA. These energies were decomposed into van der Waals and electrostatic components and further broken down into interactions involving the chromophore and the peptidic part of the antibiotic. When the resulting energy matrix was subjected to principal component analysis, it was shown that actinomycin D does indeed discriminate among different DNA sequences by an interplay of hydrogen bonding and stacking interactions. Whereas the electrostatic and hydrogen bonding interactions of the peptides dominate the discrimination between G:C- and A:T-containing sites, the importance of the stacking term increases for distinguishing among G:C-containing sites, such as GpC, GpG, and CpG, and becomes crucial for favoring DpT over GpC central steps.

The interaction energies reported above do not take into account the changes in energy undergone by either the drug or the DNA molecules on binding. For the ligands, these changes are usually small and of similar

magnitude for all the complexes considered. The situation for the DNA molecules can be rather different, but an accurate computation of these energy values is hampered by uncertainties regarding their particular conformation in the unbound state. As described below, the sequence-dependent microheterogeneity of this biopolymer is becoming more and more apparent (2), and the possibility exists that intercalation at certain sequences is favored because of the inherent tendency of some dinucleotide steps to underwind and roll.

5. INTERCALATION IN RELATION TO ANATOMY OF DNA MOLECULE

If we are to understand the principles underlying the intercalation of small ligands between DNA base pairs, we should improve our knowledge of the rules that dictate the conformational preferences of individual steps. To shed light on the choice of target sequence at which intercalation takes place, useful information can be obtained from the analysis of dinucleotide geometries taken from databases of DNA crystal structures, such as the Nucleic Acid Data Base at Rutgers University (77; NDB: *http://ndbserver.rutgers.edu:80/*). Work aimed in this direction necessarily starts with use of a common nomenclature for describing the structural parameters of nucleic acids and valid mathematical expressions for calculating them (78). According to the Cambridge convention (79), six parameters are used to define the orientation of one base pair relative to the other (twist, rise, roll, slide, tilt, and shift) and six more to define the relative orientation of the bases within each base pair (opening, stagger, propeller twist, stretch, buckle, and shear). Systematic analyses of the DNA structures deposited with the NDB have revealed that of all these parameters, those that show large variations are slide (relative translation of the base pairs about the long axis of the base step), roll (relative rotation of the base pairs about the long axis of the base step), twist (relative rotation of the base pairs about an axis perpendicular to the plane of the base step), and propeller twist (relative rotation of the bases about the long axis of the base pair). In fact, the three major conformational families of DNA molecules (A, B, and Z) can be separated on the basis of their distinct values of slide, roll, and twist.

Different attempts, ranging from simple functions related to steric clash (80, 81) to more complex energy calculations, have been made to link the primary chemical structure of DNA to the three-dimensional features of base pair architecture. Thus, relative to other dinucleotide steps, CpG, GpG (=CpC), ApT, ApG (=CpT), and, to a lesser extent, ApA (=TpT) steps have been shown to have lower than average twist values (81). Moreover, the helical twist at dinucleotide steps appears to be inversely related to the roll angle; that is, a lower average twist is usually correlated with an increased positive roll, especially at Pyr–Pur steps (81). The possibility exists, therefore, that intercalation at certain sequences be based, at least in part, on the inherent tendency of some dinucleotide steps to underwind and roll and/or on the predilection of a particular DNA sequence toward specific helical distortions. In fact, two early attempts to use an empirical potential energy function to study the interaction of several intercalating ligands with the base-paired dinucleoside monophosphates GpC and CpG (82) revealed that the known preference of many intercalators for the 5′-Pyr–Pur-3′ isomer sequence was accounted for by the differences found in the electrostatic component that accompanies unstacking rather than by the calculated intermolecular interaction energies.

A large body of theoretical work has consistently demonstrated that the position of the energy minima for two stacked bases is not necessarily at the angles the successive base planes assume in a regular double helix (83). Moreover, in many cases, this optimal position is not compatible with the constraints imposed by the sugar–phosphate backbone, which highlights the interplay that takes place in a DNA molecule between stacking and backbone conformational preferences. The largest attractive contribution to stacking is the dispersion energy, which is dependent mostly on the extent of base overlap. The electrostatic energy, which depends on the polarity of the charge distribution across the stacked base pairs, is generally weaker than the van der Waals energy (dispersion + repulsion) but is the major contributor to the stacking dependence on orientation and composition of the respective bases (18, 83, 84) and largely modulates interstrand cross-interactions (19). The sequence-dependent local variability of DNA would thus be the result of a rather delicate balance between intrastrand and interstrand contributions (19, 21).

A simple molecular model that represents the nucleic acid bases as a positively charged σ framework sandwiched between two regions of negatively charged π-electron density has been very successful in correlating calculated interaction energies with the preferred geometries of all ten dinucleotide step types and provides insight into the origins of the major trends found in DNA crystal structures (84, 85). An extension of this approach including the sugar–phosphate backbone to couple the conformational behavior of neighboring steps could be useful in understanding the sequence-selective binding of mono- and bisintercalators. For example, it has been proposed (85) that unwinding is a way to accommodate the opposed conformational tendencies of adjacent TA and AT steps. This can be important for the origin of replication at the TATA box,

but could also account for the highly cooperative binding of TANDEM to poly[d(A**TA**T)$_2$] or could explain why the binding of two molecules of echinomycin to d(A**CG**TA**CG**T)$_2$ and d(A**CG**TATA**CG**T)$_2$ is cooperative, but binding to d(T**CG**AT**CG**A)$_2$ is not (62). It has also been suggested that binding of ditercalinium to d(CGCG)$_2$ or selection by TOTO of CTAG as the preferred binding sequence might be related to the facts that ApG, in common with CpG, is one of the DNA steps with the lowest average value of twist in B-DNA (81), which presumably favors intercalation of the chromophores, and TpA, in common with GpC, is a naturally overwound step (74).

In footprinting experiments, enhanced susceptibility to cleavage by DNase I at sequences flanking the binding sites of many intercalators is a common observation reflecting that the conformational changes produced by ligand binding can be propagated over neighboring regions. This means that the energy penalty for altering DNA conformation can be distributed over several base pairs within one helical turn. In fact, it is still a matter of debate whether the particular conformation of a DNA region is simply the result of the combined intrinsic features of individual steps (possibly modulated by the environment) (86) or there is a definite influence of flanking sequences on base pair geometries (87). Understanding these issues in detail may influence our choice of sequence to be targeted by rationally designed ligands.

6. CONCLUSIONS AND FUTURE PROSPECTS

Different sorts of computations can be of help in rationalizing the known sequence-dependent binding of several intercalating ligands and in uncovering the origin of the conformational changes that accompany complex formation. Stacking interactions can be regarded as a more "hidden" element of recognition compared with interactions with the functional groups readily accessible from the minor and major grooves. In particular, the directional aspects of the electrostatic contribution to stacking interaction energies appear to make them much more important for inducing structural changes and accounting for specificity patterns than their absolute magnitudes would in principle suggest.

It is also worth bearing in mind that for optimal intercalative binding, especially in the case of bisintercalators, the intermolecular interactions may need be complemented with suitable accommodation of the distortion inflicted on the DNA double helix, which will depend on the conformational preferences of individual base steps. The DNA binding profiles of these ligands can shed additional light on the inherent flexibility and structural characteristics of particular sequences in different contexts (62, 87).

Electrostatic interactions between molecules in solution are largely offset by changes in solvation of both the ligand and the DNA on binding. To approximate calculated and experimental values, the energy penalty associated with desolvation of the interacting surfaces should also be included in the computation of the binding energies. Continuum methods based on the PB equation can be used for this purpose (37, 38, 41). Alternatively, incorporation into the models of explicit solvent molecules and use of free energy perturbation or thermodynamic integration methods (88) in the context of MD or MC simulations can be of value in obtaining binding free energy differences. These calculations involve nonphysical conversions between pairs of rather similar molecules (either ligands or target sites in macromolecules) in both the free and the bound state, and they are computer-intensive. Two thermodynamic cycles may be envisaged for quinoxaline antibiotics: A first one, involving the disappearance of all or half of the *N*-methyl groups of triostin to yield TANDEM or *cis*-methyl-TANDEM, both in the free state and in the complexes of these drugs with a nucleotide containing a CpG step, would describe the loss of affinity of the semisynthetic derivative(s) for the canonical binding step for quinoxaline antibiotics. A second one could study the decreased binding of triostin A or echinomycin when each 2-amino group of the guanines that make up its preferred CpG step is mutated to a hydrogen atom so as to yield two inosine residues. This strategy was successfully used in the computational study of netropsin binding to the minor grooves of poly[d(IC)]·poly[d(IC)] and poly[d(GC)]·poly[d(GC)] (89) and, even though binding free energies have not been determined for any of these quinoxaline antibiotics, qualitative agreement with experiment should be possible.

Finally, there are still tremendous voids in our understanding of how existing drugs that are known to interact with DNA really do work (90), although the gap is being progressively narrowed by spectacular advances in the molecular biology and crystallography of DNA-binding proteins (7). It is hoped that the near future will provide us with details of ternary complex formation with DNA topoisomerase II, helicase, or any other enzyme the activity of which is impaired or altered by drug binding.

ACKNOWLEDGMENTS

I am greatly indebted to my former collaborators J. Gallego, A. R. Ortiz, and B. de Pascual-Teresa for their enthusiasm, hard work, and many stimulating discussions and I thank M. Orozco and C. González for their critical reading of the manuscript. Biosym/MSI contributed a software license, and the Spanish CICYT provided partial support for this research (Projects SAF94/630 and SAF 96/231).

REFERENCES

1. Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer-Verlag, New York.
2. (a) Dickerson, R. E. (1992) *Methods Enzymol.* **211,** 67–111. (b) Young, M. A., Ravishanker, G., Beveridge, D. L., and Berman, H. M. (1995) *Biophys. J.* **68,** 2454–2468. (c) Ulyanov, N. B., and James, T. L. (1995) *Methods Enzymol.* **261,** 90–120.
3. Trifonov, E. N. (1991) *Trends Biochem. Sci.* **16,** 467–470.
4. Lerman, L. S. (1961) *J. Mol. Biol.* **3,** 18–30.
5. Sobell, H. M., Tsai, C.-C., Jain, S. C., and Gilbert, S. G. (1977) *J. Mol. Biol.* **114,** 333–365.
6. Berman, H. M., and Young, P. R. (1981) *Annu. Rev. Biophys. Bioeng.* **10,** 87–114.
7. Werner, M. H., Gronenborn, A. M., and Clore, G. M. (1996) *Science* **271,** 778–784.
8. Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. (1995) *Cell* **82,** 143–153.
9. Suzuki, M. (1990) *Nature* **344,** 562–565.
10. Khiat, A., Lamoureux, M., and Boulanger, Y. (1996) *J. Med. Chem.* **39,** 2492–2498.
11. Quigley, G. J., Wang, A. H. J., Ughetto, G., van der Marel, G., van Boom, J. H., and Rich, A. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7204–7208.
12. Sobell, H. M., Jain, S. C., Sakore, T. D., and Nordman, C. E. (1971) *Nature New Biol.* **231,** 200–205.
13. Lybrand, T. P., Brown, S. C., Creighton, S., Shafer, R. H., and Kollman, P. A. (1986) *J. Mol. Biol.* **191,** 495–507.
14. Kamitori, S., and Takusagawa, F. (1992) *J. Mol. Biol.* **225,** 445–456.
15. Mergny, J. L., Duval-Valentin, G., Nguyen, C. H., Perrouault, L., Faucon, B., Rougee, M., Montenay-Garestier, T., Bisagni, E., and Hélène, C. (1992) *Science* **256,** 1681–1684.
16. Silver, G. C., Sun, J.-S., Nguyen, C. H., Boutorine, A. S., Bisagni, E., and Hélène, C. (1997) *J. Am. Chem. Soc.* **119,** 263–268.
17. Fox, K. R., Thurston, D. E., Jenkins, T. C., Varvaresou, A., Tsotinis, A., and Siatra-Papastaikoudi, T. (1996) *Biochem. Biophys. Res. Commun.* **224,** 717–720.
18. Sponer, J., Leszczynski, J., and Hobza, P. (1996) *J. Biomol. Struct. Dyn.* **14,** 117–135.
19. Alhambra, C., Luque, F. J., Gago, F., and Orozco, M. (1997) *J. Phys. Chem. B* **101,** 3846–3853.
20. (a) Besler, B. H., Merz, K. M., Jr., and Kollman, P. A. (1990) *J. Comp. Chem.* **11,** 431–439. (b) Ferenczy, G. G., Reynolds, C. A., and Richards, W. G. (1990) *J. Comp. Chem.* **11,** 159–169. (c) Orozco, M., and Luque, F. J. (1990) *J. Comp. Chem.* **11,** 909–923.
21. Hobza, P., Kabelac, M., Sponer, J., Mejzlik, P., and Vondrasek, J. (1997) *J. Comput. Chem.* **18,** 1136–1150.
22. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc.* **117,** 5179–5197.
23. Pearlman, D. A., Holbrook, S. R., Pirkle, D. H., and Kim, S. H. (1985) *Science* **227,** 1304–1308.
24. Tomic, M. T., Wemmer, D. E., and Kim, S.-H. (1987) *Science* **238,** 1722–1725.
25. Lipscomb, L. A., Zhou, F. X., Presnell, S. R., Woo, R. J., Peek, M. E., Plaskon, R. R., and Williams, L. D. (1996) *Biochemistry* **35,** 2818–2823.
26. Brünger, A. T., and Karplus, M. (1991) *Acc. Chem. Res.* **24,** 54–61.
27. Alhambra, C., Luque, F. J., Portugal, J., and Orozco, M. (1995) *Eur. J. Biochem.* **230,** 555–566.
28. (a) Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York. (b) Brünger, A. T., and Nilges, M. (1993) *Q. Rev. Biophys.* **26,** 49–125.
29. (a) Schmitz, U., and James, T. L. (1995) *Methods Enzymol.* **261,** 3–44. (b) Gilbert, D. E., and Feigon, J. (1991) *Curr. Opin. Struct. Biol.* **1,** 439–445. (c) Keniry, M. A., and Shafer, R. H. (1995) *Methods Enzymol.* **261,** 575–604.
30. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983) *Science* **220,** 671–680.
31. Cosman, M., Fiala, R., Hingerty, B. E., Amin, S., Geacintov, N. E., Broyde, S., and Patel, D. J. (1994) *Biochemistry* **33,** 11518–11527.
32. Johnston, D. S., and Stone, M. P. (1995) *Biochemistry* **34,** 14037–14050.
33. Van Gunsteren, W. F., Luque, F. J., Timms, D., and Torda, A. E. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23,** 847–863.
34. Darden, T., York, D., and Pedersen, L. (1993) *J. Chem. Phys.* **98,** 10089–10092.
35. York, D. M., Yang, W., Lee, H., Darden, T., and Pedersen, L. G. (1995) *J. Am. Chem. Soc.* **117,** 5001–5002.
36. Shields, G. C., Laughton, C. A., and Orozco, M. (1997) *J. Am. Chem. Soc.* **119,** 7463–7469.
37. Honig, B., and Nicholls, A. (1995) *Science* **268,** 1144–1149.
38. Checa, A., Ortiz, A. R., de Pascual-Teresa, B., and Gago, F. (1997) *J. Med. Chem.* **40,** 4136–4145.
39. Gallego, J., Ortiz, A. R., de Pascual-Teresa, B., and Gago, F. (1997) *J. Comput.-Aided Mol. Design* **11,** 114–128.
40. Record, M. T., Mazur, S. J., Melancon, P., Roe, J.-H., Shaner, S. L., and Unger, L. (1981) *Annu. Rev. Biochem.* **50,** 997–1024.
41. Misra, V. K., and Honig, B. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 4691–4695.
42. Gaugain, B., Barbet, J., Oberlin, R., Roques, B. P., and Le Pecq, J. B. (1978) *Biochemistry* **17,** 5071–5078.
43. Rye, H. S., Yue, S., Wemmer, D. E., Quesada, M. A., Haugland, R. P., Mathies, R. A., and Glazer, A. N. (1992) *Nucleic Acids Res.* **20,** 2803–2812.
44. Jacobsen, J. P., Pedersen, J. B., Hansen, L. F., and Wemmer, D. E. (1995) *Nucleic Acids Res.* **23,** 753–760.
45. Pelaprat, D., Delbarre, A., Le Guen, I., and Roques, B. P. (1980) *J. Med. Chem.* **23,** 1336–1343.
46. Lambert, B., Rocques, B. P., and Le Pecq, J.-B. (1988) *Nucleic Acids Res.* **16,** 1063–1078.
47. Cholody, W. M., Hernández, L., Hassner, L., Scudiero, D. A., Djurickovic, D. B., and Michejda, C. J. (1995) *J. Med. Chem.* **38,** 3043–3052.
48. Hernández, L., Cholody, W. M., Hudson, E. A., Resau, J. H., Pauly, G., and Michejda, C. J. (1995) *Cancer Res.* **55,** 2338–2345.
49. Braña, M. F., Catellano, J. M., Morán, M., Pérez de Vega, M., Romerdahl, C. A., Qian, X. D., Bousquet, P., Emmling, F., Schlick, E., and Keilhauer, G. (1993) *Anti-Cancer Drug Design* **8,** 257–268.
50. Kirshenbaum, M. R., Chen, S. F., Behrens, C. H., Papp, L. M., Stafford, M. M., Sun, J. H. Behrens, D. L., Fredericks, J. R., Polkus, S. T., Sipple, P., Patten, A. D., Dexter, D., Seitz, S. P., and Gross, J. L. (1994) *Cancer Res* **54,** 2199–2206.
51. Bailly, C., Braña, M., and Waring, M. J. (1996) *Eur. J. Biochem.* **240,** 195–208.
52. Gallego, J. personal communication.

53. Lokey, R. S., Kwok, Y., Guelev, V., Pursell, C. J., Hurley, L. H., and Iverson, B. L. (1997) *J. Am. Chem. Soc.* **119,** 7202–7210.
54. Chaires, J. B., Leng, F., Przewloka, T., Fokt, I., Ling, Y. H., Pérez-Soler, R., and Priebe, W. (1997) *J. Med. Chem.* **40,** 261–266.
55. Robinson, H., Proebe, W., Chaires, J. B., and Wang, A. H.-J. (1997) *Biochemistry* **36,** 8663–8670.
56. Carpenter, M. L., Low, G., and Cook, P. R. (1996) *Nucleic Acids Res.* **24,** 1594–1601.
57. Lu, M., Guo, Q., Seman, N. C., and Kallenbach, N. R. (1990) *Biochemistry* **29,** 3407–3412.
58. Ouporov, I. V., and Leontis, N. B. (1995) *Biophys. J.* **68,** 266–274.
59. Von Kitzing, E., Lilley, D. M., and Diekmann, S. (1990) *Nucleic Acids Res.* **18,** 2671–2683.
60. Bailly, C., and Waring, M. J. (1995) *J. Biomol. Struct. Dyn.* **12,** 869–898.
61. Waring, M. J. (1990) Molecular Basis of Specificity in Nucleic Acid–Drug Interactions, p. 225, Kluwer Academic, Dordrecht.
62. (a) Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H., and Rich, A. (1984) *Science* **225,** 1115–1121. (b) Addess, K. J., and Feigon, J. (1994) *Nucleic Acids Res.* **22,** 5484–5491.
63. Gallego, J., Ortiz, A. R., and Gago, F. (1993) *J. Med. Chem.* **36,** 1548–1561.
64. Gallego, J., Luque, F. J., Orozco, M., Burgos, C., Alvarez-Builla, J., Rodrigo, M. M., and Gago, F. (1994) *J. Med. Chem.* **37,** 1602–1609.
65. Gao, X., and Patel, D. J. (1989) *Q. Rev. Biophys.* **22,** 93–138.
66. Gallego, J., Luque, F. J., Orozco, M., and Gago, F. (1994) *J. Biomol. Struct. Dyn.* **12,** 111–129.
67. Cornish, A., Fox, K. R., and Waring, M. J. (1983) *Antimicrob. Agents Chemother.* **23,** 221–231.
68. Bailly, C., Marchand, C., and Waring, M. J. (1993) *J. Am. Chem. Soc.* **115,** 3784–3785.
69. Addess, K. J., Sinsheimer, J. S., and Feigon, J. (1993) *Biochemistry* **32,** 2498–2508.
70. Gallego, J., de Pascual-Teresa, B., Ortiz, A. R., Pisabarro, M. T., and Gago, F. (1995) QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications, p. 274, J. R. Prous, Barcelona.
71. Sanz, F., Manaut, F., Rodríguez, J., Lozoya, E., and López-de-Briñas, E. (1993) *J. Comput.-Aided Mol. Design* **7,** 337–347.
72. Delepierre, M., Milhe, C., Namane, A., Dinh, T. H., and Roques, B. P. (1991) *Biopolymers* **31,** 331–353.
73. Gao, Q., Williams, L. D., Egli, M., Rabinovich, D., Chen, S. L., Quigley, G. J., and Rich, A. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2422–2426.
74. de Pascual-Teresa, B., Gallego, J., Ortiz, A. R., and Gago, F. (1996) *J. Med. Chem.* **39,** 4810–4824.
75. (a) Müller, W., and Crothers, D. M. (1975) *Eur. J. Biochem.* **54,** 267–277. (b) Krugh, T. R., Reinhardt, C. G. (1975) *J. Mol. Biol.* **97,** 133–162.
76. Müller, W., and Crothers, D. M. (1968) *J. Mol. Biol.* **35,** 251–290.
77. Berman, H., Olson, W., Beveridge, D., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., and Schneider, B. (1992) *Biophys. J.* **63,** 751–759.
78. (a) Lavery, R., and Sklenar, H. (1989) *J. Biomol. Struct. Dyn.* **6,** 655–667, (b) Babcock, M. S., Pednault, E. P. D., and Olson, W. K. (1994) *J. Mol. Biol.* **237,** 125–156, (c) El Hassan, M. A., and Calladine, C. R. (1995) *J. Mol. Biol.* **251,** 648–664.
79. Dickerson, R. E., *et al.* (1989) *Nucleic Acids Res.* **17,** 1797–1803.
80. Calladine, C. R. (1982) *J. Mol. Biol.* **161,** 343–352.
81. Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1995) *J. Mol. Biol.* **247,** 34–48.
82. (a) Pack, G. R., and Loew, G. (1978) *Biochim. Biophys. Acta* **519,** 163–172, (b) Nuss, M. E., Marsh, F. J., and Kollman, P. A. (1979) *J. Am. Chem. Soc.* **101,** 825–833.
83. Rein, R. (1978) Perspectives in Quantum Chemistry and Biochemistry, Vol. II, p. 307, Wiley, New York.
84. (a) Hunter, C. A. (1993) *J. Mol. Biol.* **230,** 1025–1054, (b) Hunter, C. A., and Lu, X. J. (1997) *J. Mol. Biol.* **265,** 603–619.
85. Hunter, C. A. (1996) *Bioessays* **18,** 157–162.
86. Subirana, J. A., and Faria, T. (1997) *Biophys. J.* **73,** 333–338.
87. Lam, S. L., and Au-Yeung, S. C. F. (1997) *J. Mol. Biol.* **266,** 745–760.
88. Kollman, P. (1993) *Chem. Rev.* **93,** 2395–2417.
89. Gago, F., and Richards, W. G. (1990) *Mol. Pharmacol.* **37,** 341–346.
90. Hurley, L. H. (1989) *J. Med. Chem.* **32,** 2027–2033.