Nucleic Acids Research, 1998, Vol. 26, No. 20 **4721–4732**

Identification and hydropathic characterization of structural features affecting sequence specificity for doxorubicin intercalation into DNA double-stranded polynucleotides

Glen E. Kellogg[1,*], J. Neel Scarsdale[1,2] and Frank A. Fornari Jr[1]

[1]Department of Medicinal Chemistry, School of Pharmacy and [2]Department of Biochemistry and Molecular Biophysics, School of Medicine, Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University, Richmond, VA 23298-0133, USA

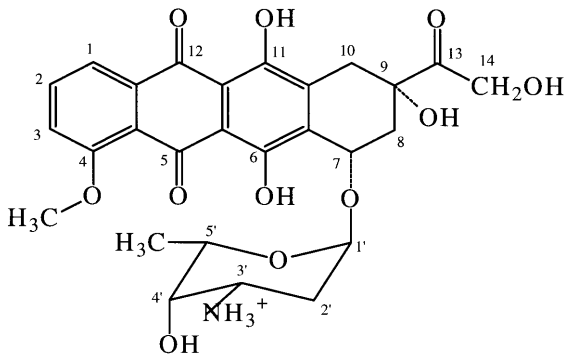
Received June 10, 1998; Revised August 5, 1998; Accepted August 11, 1998

ABSTRACT

The computer molecular modeling program HINT (Hydropathic INTeractions), an empirical hydropathic force field function that includes hydrogen bonding, coulombic and hydrophobic terms, was used to study sequence-selective doxorubicin binding/intercalation in the 64 unique CAxy, CGxy, TAxy, TGxy base pair quartet combinations. The CAAT quartet sequence is shown to have the highest binding score of the 64 combinations. Of the two regularly alternating polynucleotides, d(CGCGCG)$_2$ and d(TATATA)$_2$, the HINT calculated binding scores reveal doxorubicin binds preferentially to d(TATATA)$_2$. Although interactions of the chromophore with the DNA base pairs defining the intercalation site *[I–1] [I+1]* and the neighboring *[I+2]* base pair are predominant, the results obtained with HINT indicate that the base pair *[I+3]* contributes significantly to the sequence selectivity of doxorubicin by providing an additional hydrogen bonding opportunity for the N3′ ammonium of the daunosamine sugar moiety in ~25% of the sequences. This observation, that interactions involving a base pair *[I+3]* distal to the intercalation site play a significant role in stabilizing/destabilizing the intercalation of doxorubicin into the various DNA sequences, has not been previously reported. In general terms, this work shows that molecular modeling and careful analysis of molecular interactions can have a significant role in designing and evaluating nucleotides and antineoplastic agents.

INTRODUCTION

Anthracycline antibiotics such as doxorubicin (Scheme 1) have considerable clinical utility as antineoplastic agents (1,2). Although the exact mechanism of tumor cell cytotoxicity remains unclear (3), many of the proposed mechanisms of action, including DNA intercalation and inhibition of DNA biosynthesis (4), interference with topoisomerase II (5,6) and induction of DNA double-strand breaks (7) and interference with DNA unwinding (8,9), clearly involve interactions between the antibiotic molecule and DNA. Even though there is evidence that these antibiotics have different binding affinities for differing DNA sequences (10,11), to date no comprehensive model has emerged explaining the relationship between sequence and binding affinity; nor have there been any experimental studies aimed at establishing a structural basis for these differential affinities. Rationalization and exploitation of the structure–activity relationships for other classes of therapeutic agents has led to improved medicines for a large variety of disease states. The same kind of approach, in this case understanding and optimizing the structure–selectivity relationships of anthracycline antibiotics, could yield enhanced therapeutic agents for the treatment of cancer.



Scheme 1.

High resolution structural studies of complexes between DNA oligomers and anthracycline antibiotics would seem to form a logical basis for exploring the structural basis for sequence specificity. Indeed, high resolution crystal structures have been reported for complexes between daunomycin and adriamycin

*To whom correspondence should be addressed. Tel: +1 804 828 6452; Fax: +1 804 828 7625; Email: glen.kellogg@vcu.edu

(doxorubicin) bound to the hexanucleotide d(CGATCG)$_2$ (12). Examination of these structures, however, reveals that the intercalation sites are at the *ends* of the DNA oligomer. This preferential binding at the 3′- and 5′-ends of these short oligomers arises because the energetic and structural perturbations associated with disrupting the normal base stacking interactions are smaller for the end bases than for internal bases. In fact, for a short oligomeric sequence, once intercalation has occurred at the terminal base pairs, it is not possible for intercalation to occur at internal bases without complete disruption of the native structure.

In the case of longer DNA sequences such as those used for *in vitro* assays, initial binding events may also involve the terminal bases, but these sequences are sufficiently long that subsequent binding events can occur at internal bases far enough removed from these termini that only local disruption of the native structure may occur. However, it is entirely possible that the initial (terminal base) binding events will result in subtle structural changes at sites remote from the initial binding site which can in turn alter the affinities for subsequent binding. Thus, it is not surprising that there have been significant problems in measuring by assay the sequence discrimination of intercalator binding. As described by Pullman (13), the early experimental data was contradictory. However, the pioneering theoretical treatments of the Pullman group on daunomycin and related anthracycline antibiotics (14,15) produced a model for binding that rationalized the available contemporary experimental data by invoking a base pair triplet model to define sequence selectivity. This triplet is defined as the base pairs on either side of the chromophore intercalation site, *[I–1] [I+1]* (Fig. 1), and a base pair *[I+2]* that interacts with the sugar moieties. The major features of this model were confirmed by the more recent *in vitro* experiments of Graves and Krugh (10), Trist and Phillips (16) and Chaires (17) and crystallographic structures of oligomeric complexes reported by Frederick *et al*. (12).

Despite the fact that >2000 analogs of doxorubicin have been synthesized and tested, no clinical candidates or drugs have emerged with substantially enhanced properties and efficacy (1). While the Chen, Gresh and Pullman triplet model (14) succinctly describes the binding interactions for the major features of the doxorubicin antibiotic, the model offers little information for designing new sequence-specific antibiotic antineoplastic agents. With the continuing need for new anticancer treatments, it would seem, therefore, that re-evaluation of the structure and binding models for doxorubicin intercalation may be in order. The goal of the present study was to build upon the triplet model to add features useful for further and productive molecular design on the doxorubicin framework. Initially we undertook an exhaustive molecular modeling study of all 16 CAx, CGx, TAx and TGx sequences to verify our methodology in the context of the triplet model. Surprisingly, we found in our models that for some sequences the **N3′** ammonium group of the daunosamine sugar can form a quite strong hydrogen bond with a carbonyl oxygen on the *[I+3]* base pair, thus invoking a base pair *quartet* model for sequence selectivity of doxorubicin. Clearly, any improvement that can be made in refining and predicting the selectivity of molecules intercalating into polynucleotides will be ultimately valuable for the design of newer, more selective antineoplastic agents.

In this report we describe the results of a detailed modeling study of doxorubicin binding with 64 bp quartet sequences. We have used the HINT (18–20) model for biomolecular interactions to evaluate the binding efficacy of doxorubicin in each of the DNA model sequences. We have recently shown that HINT results correlate with experimental measurements of free energy for dimer–dimer association for native and mutant hemoglobins (21) and that HINT empirical ligand scoring functions for inhibitors with HIV-1 reverse transciptase can identify potential therapeutic agents in extended database searches (22). Here we show that the HINT model can be extended to small molecule–DNA interactions. One consequence of the applicability of the HINT model in this case is that there is evidence that hydrophobic–hydrophobic interactions can contribute significantly to the binding interactions between ligand and DNA.

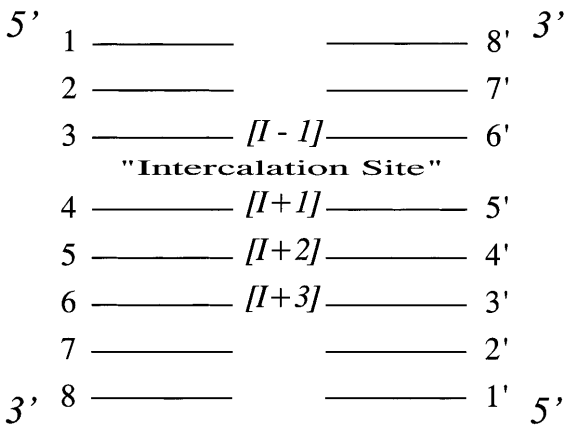


Figure 1. Model for quartet intercalation. The *[I–1]* base pair is above the intercalation site, *[I+1]* is immediately below the intercalation site and *[I+2]* and *[I+3]* are 1 and 2 bp distal of the site, respectively.

MATERIALS AND METHODS

Molecular models and energy minimization

All molecular models were created and minimized with the SYBYL 6.2 molecular modeling package (Tripos Inc., St Louis, MO) using the Tripos force field and Gasteiger–Hückel charges. The crystal coordinates for the doxorubicin–d(CGATCG)$_2$ complex (12) were obtained from the Brookhaven Protein Data Bank (23), accession no. 1D12. Modeling of the crystal structure was accomplished by reading the atomic coordinates into SYBYL and holding them as an aggregate. Hydrogens were added and the molecule was solvated using the droplet protocol in SYBYL with a single layer of water molecules. This structure was minimized, first with 300 cycles of steepest descent minimization, then by conjugate gradient minimization, until the energy difference between successive iterations was <0.05 kcal/mol. We found that the water monolayer as added by the droplet protocol was sufficient to retard helix unwinding without adding significant complexity to the system. Sixteen additional structures, the preferred (24) Pyr(3′-5′)Pur sequences CAx, CGx, TAx and TGx (x = A, C, G or T), were created in SYBYL as follows. (i) B-DNA octamers [d(CG-self-complementary hexanucleotide)] were constructed using the builder tool in the Biopolymer module. (ii) Doxorubicin in the crystal structure conformation with the chromophore held as an aggregate was placed within the intercalation site between base pairs *[I+1]* and *[I–1]* as calculated by the program GRID (25). The ionization state of the doxorubicin model was as described in previous modeling studies (13,14) and as seen

experimentally (26), i.e. the ammonium and all hydroxyls are protonated. The placement of doxorubicin was determined by the most favorable site for the NH_3^+ and OH GRID probes. (iii) These structures were solvated and geometry optimized using the minimization protocol described above. To verify that the placement of doxorubicin based on GRID calculations for each of these 16 structures resulted in a structure at a low energy minimum on the potential surface, doxorubicin was translated from its minimized position by ±1.0 Å along the principal axis of the chromophore and these resulting structures were optimized using our minimization protocol. Forty eight additional structures were constructed from this set of 16 by permuting the *[I+3]* base pair over all possible pyrimidine/purine combinations and the resulting structures solvated and minimized as before.

Analysis of hydropathic interactions

The role of hydropathy in doxorubicin binding was analyzed using the program HINT (18–20). In the HINT model specific interactions between small molecules and DNA are described as a double sum over the atoms within each component:

$$B = \sum_{j=1}^{\text{atoms}} \sum_{i=1} b_{ij} = \sum \sum (S_i a_i S_j a_j R_{ij} T_{ij} + r_{ij}) \qquad \mathbf{1}$$

where S is the solvent-accessible surface area, a is the hydrophobic atom constant, T is a descriptor function (*vide infra*) and R and r are functions of the distance between atoms i and j. From this equation, a binding score is calculated where b_{ij} describes the specific interaction between atoms i and j and B describes the total interaction between the two species.

The hydrophobic atom constants (a_j) are derived by reduction (20) of the fragment constants for the water/octanol partition coefficient (27,28). Positive signed atom (fragment) constants indicate hydrophobic atoms (fragments) while negative signed constants indicate polar or hydrophilic atoms (fragments). Partition coefficients (sum of hydrophobic atom constants) for small molecules calculated by HINT are similar to values calculated by other methods. Solvent-accessible surface area (S_i) is a constant describing the shape and accessibility of the atom and its tendency for interaction. Buried atoms have a smaller S and are less involved in interactions.

The descriptor function, T_{ij}, differentiates among the three possibilities for polar–polar interactions (acid–acid, acid–base/hydrogen bonding and base–base) in order to maintain the convention that favorable interactions have positive scores. Each SYBYL atom type is assigned descriptor variables to represent its hydrogen bonding acceptor/donor character, charge, Brønsted acid/base character and Lewis acid/base character. These are used by T_{ij} to calculate a value of +1, –1 or 0 for each atom–atom interaction. We consider the units of T_{ij} to be $Å^{-4}$, so that B and b_{ij} have no units.

The functional form of the range dependence is described by two terms (R_{ij} and r_{ij}). The former scales the hydrophobic atom constant/solvent-accessible surface area product with distance, while the second is independent of hydropathy and responds only to distance variations. For this work R_{ij} has been set to the simple exponential, e^{-r}, where r is the distance between the interacting atoms in Å. The 6–12 Lennard–Jones function,

$$r_{ij} = A\, e_{ij} [(vdw/r)^{-6} – 2(vdw/r)^{-12}] \qquad \mathbf{2}$$

where e_{ij} is the van der Waals parameter (29,30) and A is a scaling factor balancing the contributions of hydropathic and van der Waals forces, was used for r_{ij}. For this study, A = 50/(kcal/mol).

Three-dimensional hydropathic interaction maps were calculated as described previously (21). The contour maps shown in the figures are the result of two independent passes over the map region. The first focused on hydrophobic interactions, while the second focused on polar interactions. The maps were contoured and displayed using SYBYL. In the studies presented herein, the hydropathic interactions between DNA and doxorubicin were examined with HINT. The interactions are color coded as follows: hydophobic–hydrophobic interactions are shown as green contours; polar–polar (favorable) interactions are shown as blue contours (these interactions are due to acid–base, coulombic or hydrogen bonding); polar–polar (unfavorable) interactions are shown as red contours (these are generally due to acid–acid or base–base interactions).

HINT calculation details

This study was performed with HINT v.2.11S (eduSoft LC, Ashland, VA) using adjustable HINT parameters as reported previously (21). HINT v.2.11S has been integrated in SYBYL 6.2. The atom potential types used by HINT v.2.11S are based on the Tripos (SYBYL 6.2) force field. The fragment values and log$P$ calculation method of Hansch and Leo (27) were modified and adapted to the Tripos atom primitive set.

Two principal parameters are assigned to each atom in the HINT model. The first parameter, a_i, is the hydrophobic atom constant and represents the contribution of that atom to the total solvent interactions of the molecule. Each of the nucleotide bases was modeled in SYBYL as phosphate-capped molecular species and subjected to small molecule HINT log$P$ calculations (20). This capping scheme simulates the effects of polar proximity (27,28) between atoms in the current and adjacent bases. S was calculated for each atom using a simple geometric algorithm based on intersecting spheres (atoms) with radii equivalent to the sum of the atomic van der Waal radius (31) and 1.4 Å (presumed to be the radius of water). These resulting values were placed in a dictionary of look-up values keyed to the nucleotide base type and the atom names. The HINT parameters for doxorubicin were calculated using the HINT small molecule partitioning algorithm.

Each of the 64 DNA oligonucleotides were assigned HINT parameters from the dictionary. Doxorubicin and the DNA fragments were partitioned with Hydrogens = Essential, i.e. only polar hydrogens were explicitly used in the model. Then, an interaction score using equation **1** was calculated for each uniquely modeled and structure optimized intercalator complex. Three-dimensional maps that pictorially represent non-covalent interactions were calculated on a 1 Å grid.

RESULTS AND DISCUSSION

HINT hydropathic analysis

There are numerous energetic contributions to a biomolecular event as complex as intercalation of a drug-like molecule into the DNA double helix. Calculation of the free energy (ΔG) for the event would have to include, among others, terms to represent the deformation of the DNA, loss of entropy for the new bimolecular complex and solvent partitioning for the drug from water to the intercalation site, as well as terms specific to the drug–DNA

interaction. This level of computation is beyond the scope of the present work. However, many of these terms can be safely ignored in investigations of $\Delta\Delta G$, the *difference* in free energy of binding between different complexes. This should be a reasonable assumption for the present case where we are investigating the differences in binding energy and interactions for the same drug molecule in DNA oligomers of the same length that have been modeled and optimized in the same manner.

We have examined, in detail with computer molecular models, the intermolecular interactions between doxorubicin and 64 base quartet sequences of DNA using the HINT (Hydropathic INTeractions) (18–20) program. This program utilizes the experimental data from small molecule solvent partitioning between 1-octanol and water (log*P*) as the basis for a non-covalent interaction force field. The HINT model is defined around the assertion that the two solvents can be thought of as representations of biological environments, with water a polar environment and 1-octanol a hydrophobic environment. The interactions that the small molecule makes in solvating and partitioning between the two phases are the same ones that ligands make in binding to receptors, etc. Thus, the solvent partitioning data are unique experimental measures of interaction. Especially significant is that these data are related to free energy and thus include entropy. HINT analysis of an interaction for a ligand binding or macromolecular association event produces a detailed list of atom–atom interactions, including a character and score for each. These scores are positive for favorable interactions and negative for unfavorable interactions. However, it should be noted that certain interaction classes favored by HINT are energetically disfavored (largely due to the electrostatic term) by the molecular mechanics force fields used to create the models. The opposite can also occur. This would seem to be a potentially serious limitation of the technique, but there are mitigating factors. First, this affects a relatively small number of interactions in this system. Second, both of the most significant of these scoring 'errors', i.e. hydrophobic–hydrophobic, scored by HINT as favorable but somewhat disfavored electrostatically, and hydrophobic–polar, scored by HINT as unfavorable but electrostatically allowed in some cases (e.g. methyl–carbonyl), *systematically* depress the HINT scores. These two factors suggest that we can examine relative trends and score ordering with reasonable confidence.

The effect of water on the reported scores and, by inference, $\Delta\Delta G$ should also be briefly considered because it is probable that water molecules are mediating the drug–DNA interactions in these systems. First, we should note that this effect would likely be similar for each of these complexes. Second, a limited representation of the effect of water is inherent in the free energy-derived HINT constants. However, 'structural waters', those that are strongly bound, may need to be considered as distinct entities with an appropriate contribution to the total HINT score (32). Our evolving 'rule-of-thumb' is that water molecules having between two and three identifiable macromolecular interactions should be explicitly modeled for HINT interaction analysis. The X-ray crystal structure for the doxorubicin–$d(CGATCG)_2$ complex (12) does not show water molecules between the doxorubicin and DNA meeting this condition. Thus, in the current study, our HINT calculations do not include specific 'structural waters'.

Tables 1 and 2 set out interaction lists for the doxorubicin reaction with the CGCG and CAAT quartet sequences capped as described in Materials and Methods. The data in these tables are filtered to list only the most significant interactions. It turns out that in this case the bulk of the interactions between DNA and the chromophore portion of doxorubicin, i.e. the driving force for the actual intercalation, are individually small on an atom-by-atom basis and thus do not appear particularly prominently in the tables. An alternative method for viewing the interaction profile is with 3D maps that display the contoured interaction fields for the binding event. Figure 2a–c shows contoured maps for the interactions of doxorubicin superimposed on the molecular models for the CGCG, TATA and CAAT quartets extracted from the polynucleotide/doxorubicin models used for the analysis. The contours are color coded by interaction type as described in the figure captions.

From the interaction maps we can see that the chromophore *[I–1][I+1]* region is dominated by polar interactions, where the majority are favorable (blue contours). These arise from the acid–base interactions between the heteroatoms on both the doxorubicin and the DNA bases. It is necessary to point out that unsaturated carbons are also acting as hydrogen bond acceptors and/or Lewis bases (30,33,34). This type of interaction, which is encoded in the HINT model, is clearly important in this system (see Tables 1 and 2). Also note the patches of hydrophobic interactions (in green) more or less paralleling the carbons of the deoxyribose chains. We can assert from this that there are significant hydrophobic–hydrophobic interactions between DNA and this class of ligand, but it remains to be seen whether these interactions contribute to sequence selectivity. While there appear to be some subtle differences in the *[I–1] [I+1]* region between the three sequences (Fig. 2a–c), the major differences are due to the interactions of the sugar portion of doxorubicin with the bases. This is in accord with the analyses of Chen, Gresh and Pullman (14,15), who first proposed a triplet sequence model to explain binding selectivity for this type of agent.

The most significant specific interactions are associated with the **N3′** ammonium group of the sugar, which forms extremely strong hydrogen bonds. It is instructive to compare the size and strength of the blue contours in the maps of Figure 2 for the three sequences. In Figure 2a (CGCG) the contour around **N3′** (lower center) is relatively small as there are only fairly weak hydrogen bonding opportunities for the ammonium ion in this environment (see Table 1). In Figure 2b (TATA) the contour is larger as one of the hydrogens of **N3′** can donate to the **O2** atom of ***T5′***. Finally, in Figure 2c (CAAT) the contour around **N3′** now encloses two significant hydrogen bonds: to the **O2** atom of ***T5′*** and to the **O2** atom of ***T6*** of the *[I+3]* base pair (see Table 2). Close examination of Figure 2c reveals the ***T6*** **O2** atom tipping up towards the daunosamine **N3′**, clearly indicating this to be a *new* substantive interaction. While the Chen *et al*. study did not examine models for the CAAT sequence, the authors did report that ΔE_{inter} for $d(TATATA)_2$ is ~13 kcal/mol more favorable than for $d(CGCGCG)_2$ (14). Thus the graphical results from the present study are in qualitative agreement with previous theoretical treatments.

These results confirm that there is a structural basis for sequence selectivity. The modeling/hydropathic analysis approach we have employed produces results consistent with previous models which examined a small subset of the possible sequences. The observation of interactions at the *[I+3]* base pair is a new result that is in large measure due to our exhaustive examination of the valid sequence combinations.

Table 1. List of atom–atom interactions[a] and HINT scores between doxorubicin and DNA quartet sequence CGCG

Doxorubicin		DNA				Interaction	
Atom	a_i	Base	Atom	a_j	r_{ij} (Å)	Score	Class
C5	0.576	G6′	O6	–1.915	3.09	57	Base–acid[b]
C10	0.585	C3	O2	–1.915	3.45	–62	Hydrophobic–polar
C21	0.800	C5′	C6	0.355	3.68	21	Hydrophobic
C21	0.800	C5′	C5	0.355	3.51	23	Hydrophobic
O5	–0.972	C5′	O2	–1.915	3.49	–86	Base–base
O5	–0.972	C5′	C2	2.255	2.85	53	Acid–base[b]
O5	–0.972	G6′	O6	–1.915	3.59	–78	Base–base
O6	–0.264	C5′	C2’	0.588	3.93	–63	Hydrophobic–polar
O6	–0.264	C5′	C2	2.255	3.29	102	Hydrogen bond[b]
O6	–0.264	G6′	C8	0.920	3.32	167	Hydrogen bond[b]
O6	–0.264	G6′	C6	2.356	3.87	51	Acid–base[b]
O6	–0.264	G6′	C4	0.443	2.83	60	Hydrogen bond[b]
O7	–0.686	G4	N2	–0.641	2.63	73	Hydrogen bond
O9	–0.903	G4	N2	–0.641	3.63	–70	Acid–acid
O9	–0.903	G4	N3	–0.942	2.93	93	Hydrogen bond
O9	–0.903	C5	C5′	0.480	3.54	–56	Hydrophobic–polar
O11	–0.293	C3	C6	0.355	3.56	57	Hydrogen bond[b]
O11	–0.293	C3	C2	2.255	2.82	123	Hydrogen bond[b]
O11	–0.293	G4	C8	0.920	3.37	131	Hydrogen bond[b]
O11	–0.293	G4	C6	2.356	3.50	78	Hydrogen bond[b]
O11	–0.293	G4	C4	0.443	2.92	53	Hydrogen bond[b]
O12	–0.972	G4	O6	–1.915	3.18	–117	Base–base
O13	–1.915	C3	O2	–1.915	4.32	–78	Base–base
O13	–1.915	G6′	N2	–0.641	4.41	54	Acid–base
O14	–1.004	C5	C5′	0.480	3.40	–72	Hydrophobic–polar
C6′	0.765	G6	P	5.086	4.56	40	Hydrophobic
C6′	0.765	G6	O1P	–3.310	4.60	–67	Hydrophobic–polar
O4′	–0.886	G6	C5′	0.480	3.38	–55	Hydrophobic–polar
N3′	–0.998	C5	O2	–1.915	3.96	170	Acid–base
N3′	–0.998	G6	O4′	–0.678	2.74	184	Hydrogen bond
N3′	–0.998	G6	C1′	0.358	3.70	–53	Hydrophobic–polar
N3′	–0.998	G6	N3	–0.942	4.51	52	Acid–base
N3′	–0.998	G4′	N2	–0.641	3.80	–104	Acid–acid
N3′	–0.998	C5′	O2	–1.915	4.21	137	Acid–base

[a]Interaction class definitions: hydrogen bond refers to an interaction between a hydrogen bond donor atom and a hydrogen bond acceptor atom where (i) the atoms are within 3.65 Å and (ii) the HINT score is at least 50; hydrophobic refers to an interaction between two hydrophobic atoms where the HINT score is at least 20; acid–base refers to an interaction between a Lewis acid and a Lewis base where the HINT score is at least 50. Interactions meeting the criterion of hydrogen bond but that are >3.65 Å apart are classified as acid–base; acid–acid are interactions involving two Lewis acid atoms; base–base are interactions involving two Lewis base atoms; hydrophobic–polar are interactions between a hydrophobic atom ($a_i > 0$) and a polar atom ($a_i < 0$).

[b]Interactions between an unsaturated carbon (which is a potential Lewis base and/or hydrogen bond acceptor) and a Lewis acid/hydrogen bond donor.

Contributions to sequence selectivity

How important are hydrophobic interactions in determining sequence selectivity? In order to address this issue we separated the contributions to the HINT score for each base pair quartet into three groups: hydrophobic–hydrophobic interactions, hydrogen bonds and ‘all other’ (which includes acid–base, acid–acid, base–base and hydrophobic–polar) interactions. The result of this analysis is presented in the bar chart graph of Figure 3, where the light green portion of each bar represents the contribution of hydrophobic interactions, the blue portion represents hydrogen bonds and magenta represents the remainder (i.e. ‘all other’). From this we can see that the hydrophobic interactions contribute ~25% to the overall interaction score, but add little, if any, selectivity. The magenta bars, contributing 5–15% to the interaction score, have some variability as a function of sequence, but no pattern readily emerges. However, the blue bars, representing hydrogen bonding contributions, show significant variation and appear to be the source of most selectivity. This confirms the qualitative graphical analysis afforded by the HINT interaction maps of Figure 2.

Table 2. List of atom–atom interactions[a] and HINT scores between doxorubicin and DNA quartet sequence CAAT

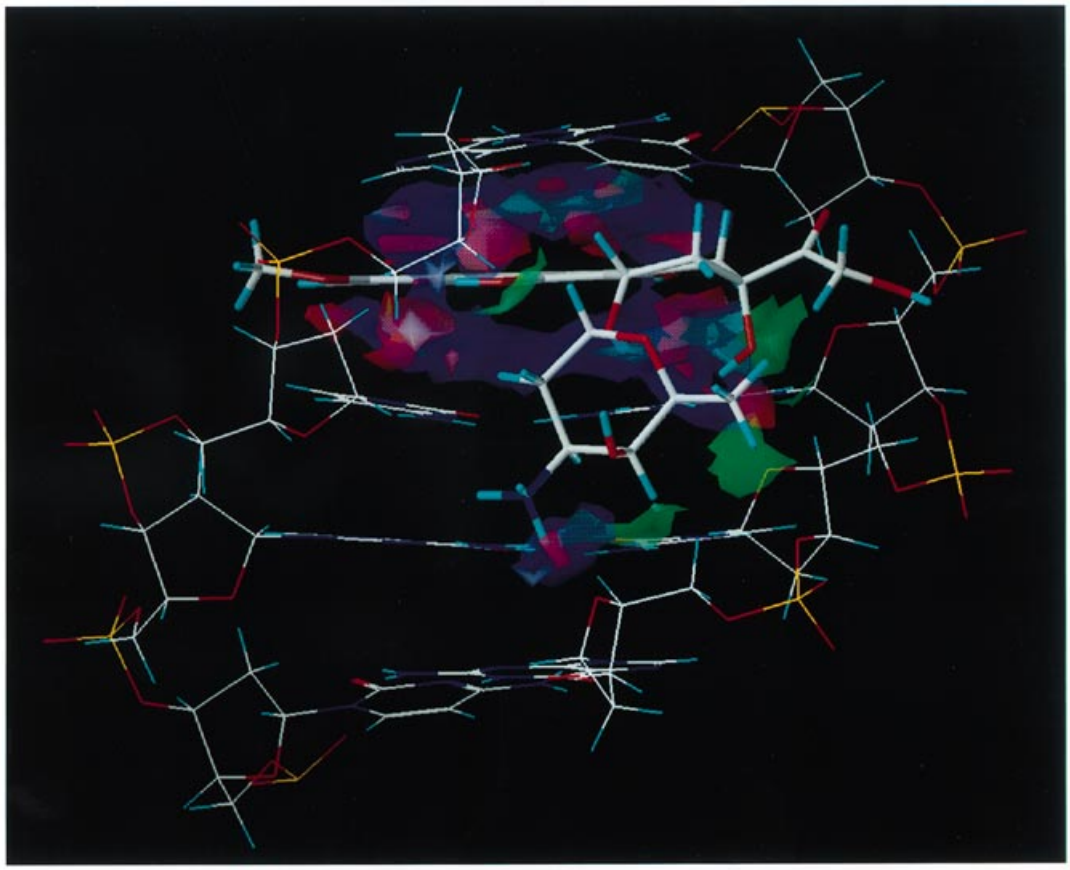
Doxorubicin		DNA				Interaction	
Atom	a_i	Base	Atom	a_j	r_{ij} (Å)	Score	Class
C3	0.355	T5′	C5M	0.794	3.61	23	Hydrophobic
C5	0.576	G6′	O6	–1.915	3.05	59	Base–acid[b]
C10	0.585	C3	O2	–1.915	3.62	–52	Hydrophobic–polar
C10	0.585	A4	C2	0.920	3.76	26	Hydrophobic
C21	0.800	T5′	C5M	0.794	3.71	67	Hydrophobic
O4	–0.406	T5′	C5M	0.794	2.74	–94	Hydrophobic–polar
O5	–0.972	T5′	O4	–1.302	3.59	–52	Base–base
O5	–0.972	T5′	O2	–1.915	3.75	–67	Base–base
O5	–0.972	G6′	O6	–1.915	3.52	–84	Base–base
O6	–0.264	T5′	C2	1.783	3.30	79	Hydrogen bond[b]
O6	–0.264	G6′	C8	0.920	3.22	172	Hydrogen bond[b]
O6	–0.264	G6′	C6	2.356	3.70	60	Acid–base[b]
O6	–0.264	G6′	C4	0.443	2.90	56	Hydrogen bond[b]
O9	–0.903	A4	C2	0.920	3.26	235	Hydrogen–bond[b]
O9	–0.903	A4	N3	–0.873	2.74	94	Hydrogen bond
O9	–0.903	A5	C8	0.920	4.41	77	Acid–base[b]
O11	–0.293	C3	C6	0.355	3.71	51	Acid–base[b]
O11	–0.293	C3	C2	2.255	2.80	132	Hydrogen bond[b]
O11	–0.293	A4	C8	0.920	3.20	160	Hydrogen bond[b]
O11	–0.293	A4	C2	0.920	3.92	59	Acid–base[b]
O12	–0.972	A4	N6	–0.503	3.21	58	Hydrogen bond
O14	–1.004	A5	C5′	0.480	3.30	–75	Hydrophobic–polar
C6′	0.765	T6	P	5.086	4.74	33	Hydrophobic
C6′	0.765	T6	C5′	0.468	3.79	26	Hydrophobic
N3′	–0.998	A5	C2	0.920	3.54	83	Hydrogen bond[b]
N3′	–0.998	A5	N3	–0.873	3.78	78	Acid–base
N3′	–0.998	T6	O2	–1.915	2.80	789	Hydrogen bond
N3′	–0.998	T4′	O2	–1.915	4.29	176	Acid–base
N3′	–0.998	T5′	O2	–1.915	2.62	793	Hydrogen bond

[a,b]See notes to Table 1.

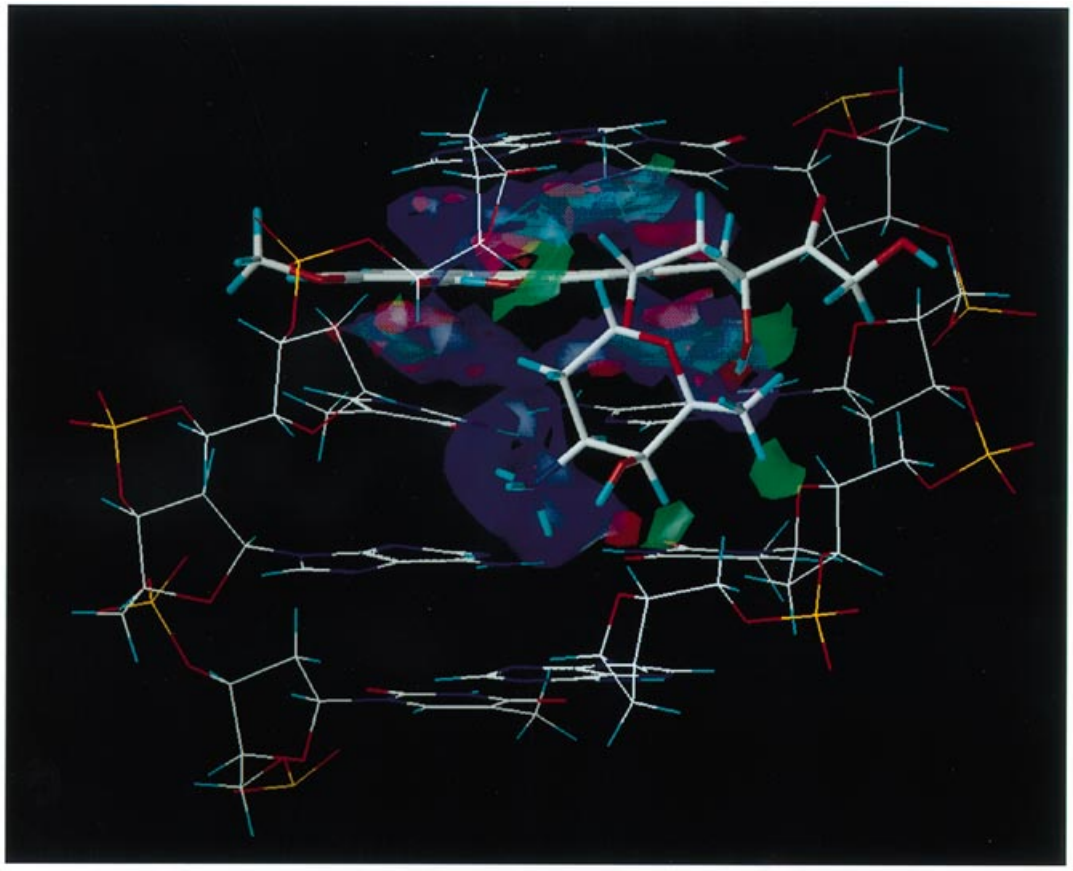
From Figure 4, a bar chart showing the contributions of *[I–1]* (blue), *[I+1]* (yellow), *[I+2]* (red) and *[I+3]* (green) base pairs to the total interaction score, we can assess the sequence discrimination for all 64 CAxy, CGxy, TAxy and TGxy quartets. It is plain that there is a significant effect from the fourth base pair, i.e. there are often significant differences among the members of each triplet family. For example, CAAt has a HINT interaction score ~1000 more than that for CAAa. It is relevant to discuss uncertainty and error of HINT interaction scores at this point. In most of our previous experience with HINT we have used crystallographically determined structures as the basis of our molecular models and have reported uncertainties in the vicinity of ±100–200 for total interaction scores (18,21,32). For the present case, where the model structures are themselves created with molecular mechanics force fields, assessing uncertainty is more difficult. However, our modeling procedure was partially verified by reproducing the crystallographically determined structure of d(CGATCG)$_2$ with an RMS deviation of 1.34 Å. [The RMS calculation was performed on the heavy (non-hydrogen) atoms of the CGAT–doxorubicin portion of both the crystal and molecular mechanics models. Note that the intercalation site for the crystal model is at the 3′-end of the hexanucleotide and some unraveling of the DNA double helix has likely occurred which accounts for a portion of the RMS deviation.] Thus we believe the uncertainty in HINT scores in this study to be similar to that reported before and differences of the order of 1000 are likely to be statistically significant. In previous studies (20,21) we have found that 300–500 score units corresponds to 1 kcal/mol free energy difference. Therefore, a HINT score difference of the order of 1000 may represent a 1–2 order of magnitude difference in the equilibrium constant of binding.

What position(s) of the base pair quartet controls selectivity? There is little selectivity at *[I–1]* (above the intercalation site; Fig. 1). There is significant variability with the *[I+1]* base pair, however, this is largely due to **N3′** forming a hydrogen bond with some sequences on the ‘backside’ of the pair, not because of intercalation differences. Selectivity at the *[I+2]* base pair is modest. The **O9** hydroxy of doxorubicin can find an acceptor for some nucleotides and hydrogens attached to **N3′** can interact favorably or unfavorably with available atoms in base pairs at this position. For example, consider CGCG (Table 1), where **O9** interacts unfavorably with ***C5*** **C5′**; **N3′** interacts favorably with ***C5*** **O2** but unfavorably with ***G4′*** **N2**. In CAAT (Table 2) **O9** interacts favorably with ***C5*** **C8**; **N3′** interacts favorably with ***A5*** **C2** and favorably with ***T4′*** **O2**. The *[I+3]* base pair yields only one significant interaction in ~25% of the 64 sequences we examined. That interaction is the hydrogen bond between the doxorubicin **N3′** and the **O2** atom of either ***C6*** or ***T6***. Neither

a



b



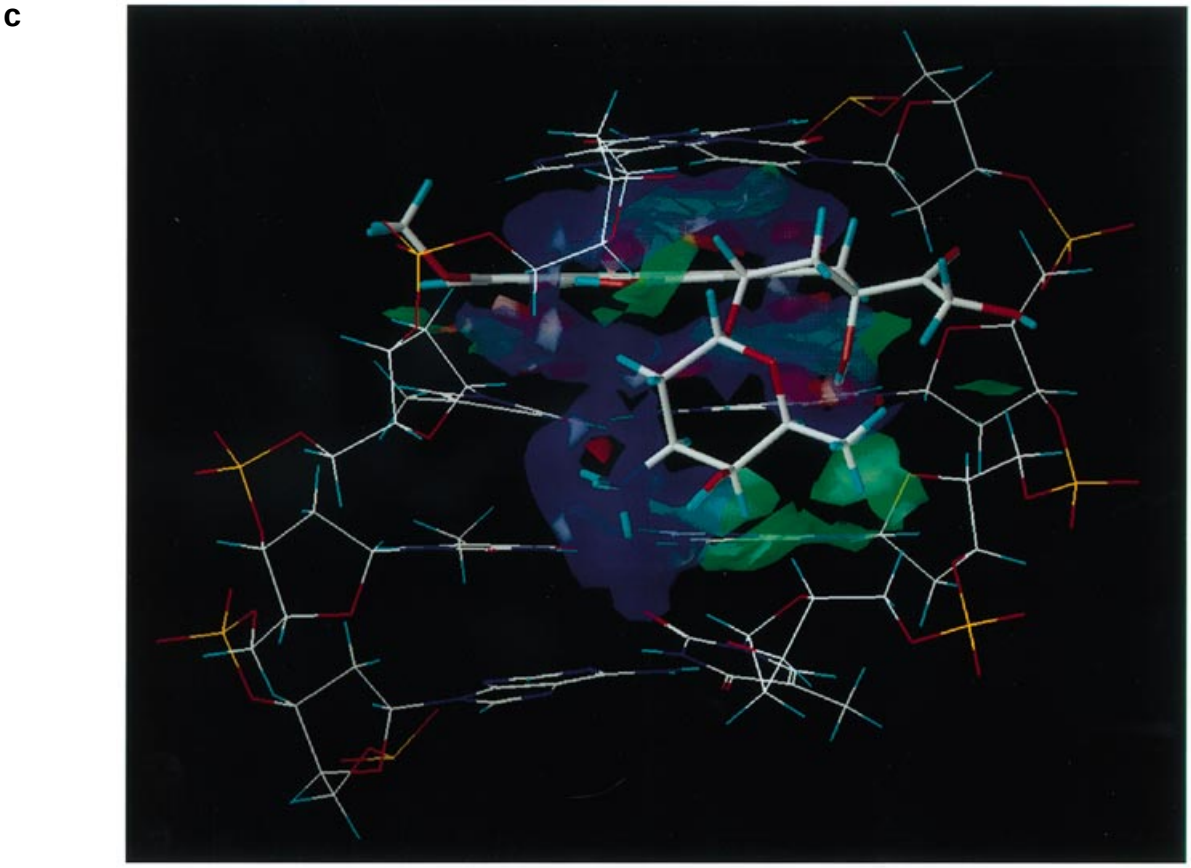


Figure 2. (Above and previous page). HINT interaction maps for the intercalation of doxorubicin into various base pair sequences of DNA. HINT interaction contour maps display, visually, the quality and magnitude of the binding contacts: the contour surfaces are color coded by interaction type at a constant map density value of ±125; the relative volume of the enclosed contour surface can be correlated with the relative magnitude of the interaction. Blue surfaces represent favorable polar–polar contacts, which are generally hydrogen bonds; red surfaces represent unfavorable polar–polar contacts such as base–base or acid–acid interactions; green surfaces represent regions where there are favorable hydrophobic–hydrophobic contacts. (**a**) Contacts of doxorubicin with the DNA CGCG model; (**b**) contacts of doxorubicin with the DNA TATA model; (**c**) contacts of DNA with the DNA CAAT model. The new hydrogen bond between the doxorubicin N3′ ammonium nitrogen and the *[I+3]* base pair is evident.

adenine nor guanine can make this hydrogen bond, as they lack an appropriate acceptor atom in this region. For the five sequences examined by the Pullman group in their 1986 theoretical study (14), in which the reported preferential affinity for doxorubicin was CGTa > TATa > TGAt > CGCg > TACg, we find the ordering TGAT > CGTA ~ TATA > TACG > CGCG.

The major difference comes from our observation of the hydrogen bond from **N3′** to the **O2** atom of ***T6*** in TGAT which was not observed by the Pullman group. This interaction of TGAT, which is the only sequence of the five possessing it, gives this sequence the highest score in our model. It is also interesting that this sequence, TGAT, is recommended by Trist and Phillips for further study of high affinity doxorubicin–DNA complexes (16).

What are the factors that rule whether ***C6*** or ***T6*** can make this unique hydrogen bond? It only appears in about half of the sequences containing these bases. It is likely there is a complex balance of steric and energetic effects in this region of the oligonucleotide. (i) There are at least five hydrogen bond acceptors accessible to **N3′** in this region for some sequences, for example, in CACT (a sequence for which we do not see a strong hydrogen bond from the ammonium to the *[I+3]* base pair) ***T5′*** **O2**, ***T5′*** **O4′**, ***C5*** **O2**, ***T6*** **O2** and ***T6*** **O4′** (see Fig. 5). (ii) In order for the ***T6*** **O2** to make a bid for the sugar **N3′** it must release a portion of its hydrogen bonding to the ***3′*** base. (iii) Small atom translations can affect hydrogen bond formation by changing the angle between the acceptor lone pairs and the donor's hydrogen. (iv) There must be water molecules present in this region; since they can act as both hydrogen bond acceptors and donors, these waters are clearly a confounding factor (32). At this level of analysis it is impossible to sort and prioritize these multiple complex effects. However, since all the models in the present study were built using a consistent methodology, it would seem reasonable to assume that the resulting structures are themselves self-consistent and that the observed differences (i.e. $\Delta\Delta G$) are real within the confines of the force field and minimization procedures.

Correlation with prior sequence specificity studies

Integration of these results into the constellation of previously reported experimental and theoretical sequence specificity invest-

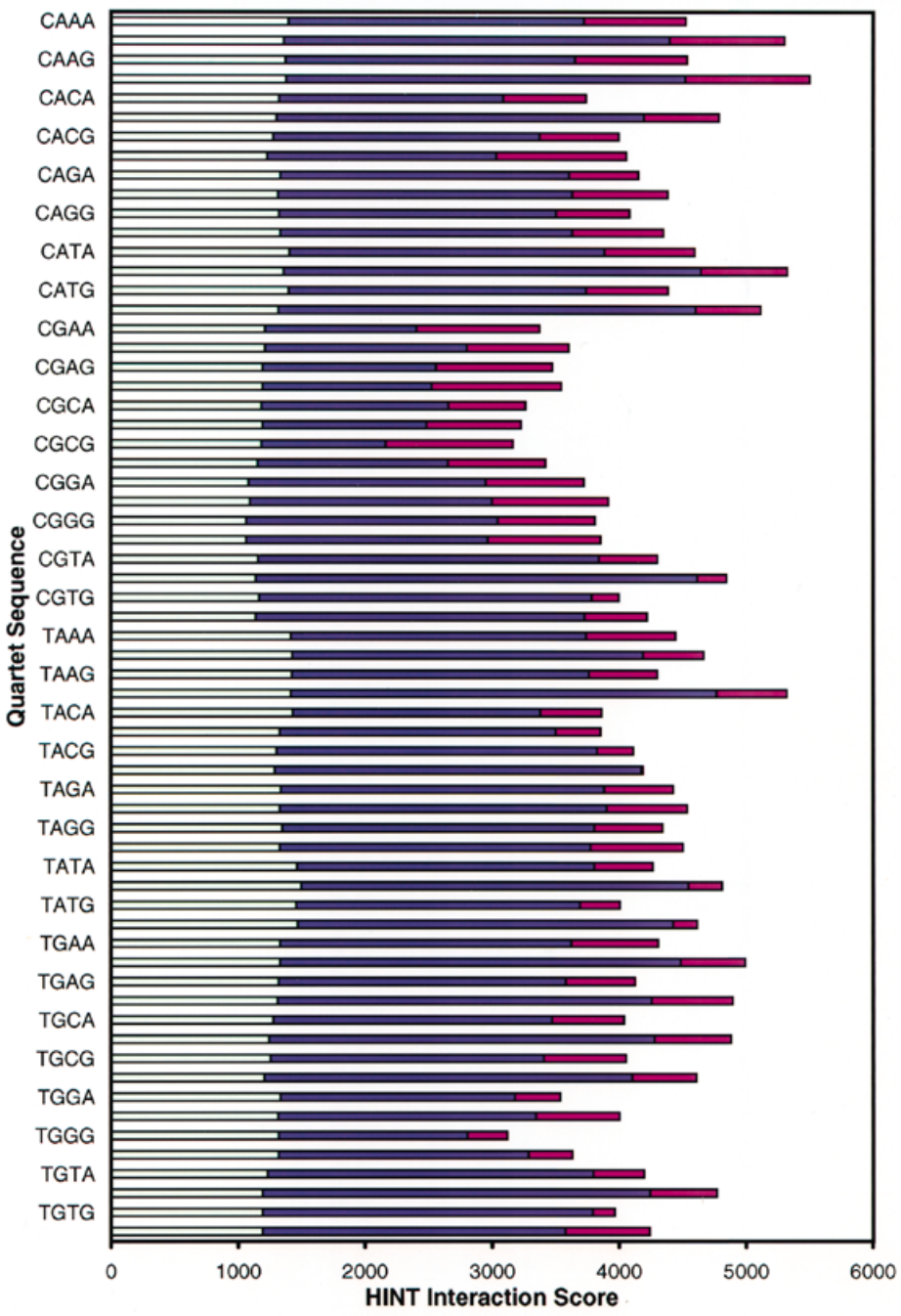


Figure 3. HINT interaction score by base pair quartet sequence (interaction type). Each bar has three segments representing the contribution from: light green, hydrophobic–hydrophobic interactions; blue, hydrogen bonds; magenta, sum of all other favorable and unfavorable interactions, which includes acid–base (favorable), acid–acid, base–base and hydrophobic–polar (all unfavorable). For clarity on the quartet sequence axis, only xyzA and xyzG sequences are labeled. The unlabeled sequences are xyzC and xyzT, as alphabetically appropriate.

igations is a difficult exercise: (i) the experimental DNase I footprinting studies have a limited basis for determining the orientation of the drugs, i.e. which base pairs define the intercalation site and which base pair(s) is involved in the minor groove binding interactions; (ii) since interpretation of footprinting results is dependent on the supposed site size (35), can the results from studies that assumed a triplet be fairly reconciled with a quartet model?; (iii) should daunorubicin (in which the **O14** hydroxide has been replaced by a hydrogen) (35) give the same results in high resolution footprinting titration studies as doxorubicin?; (iv) the theoretical studies (13–15), upon which virtually all arguments about triplet specificity models are based, examined only 5 out of the possible 16 energetically reasonable (24) triplets (and thus only 5 out of the possible 64 quartets) and were performed on hexamer DNA double helix segments where end effects are significant.

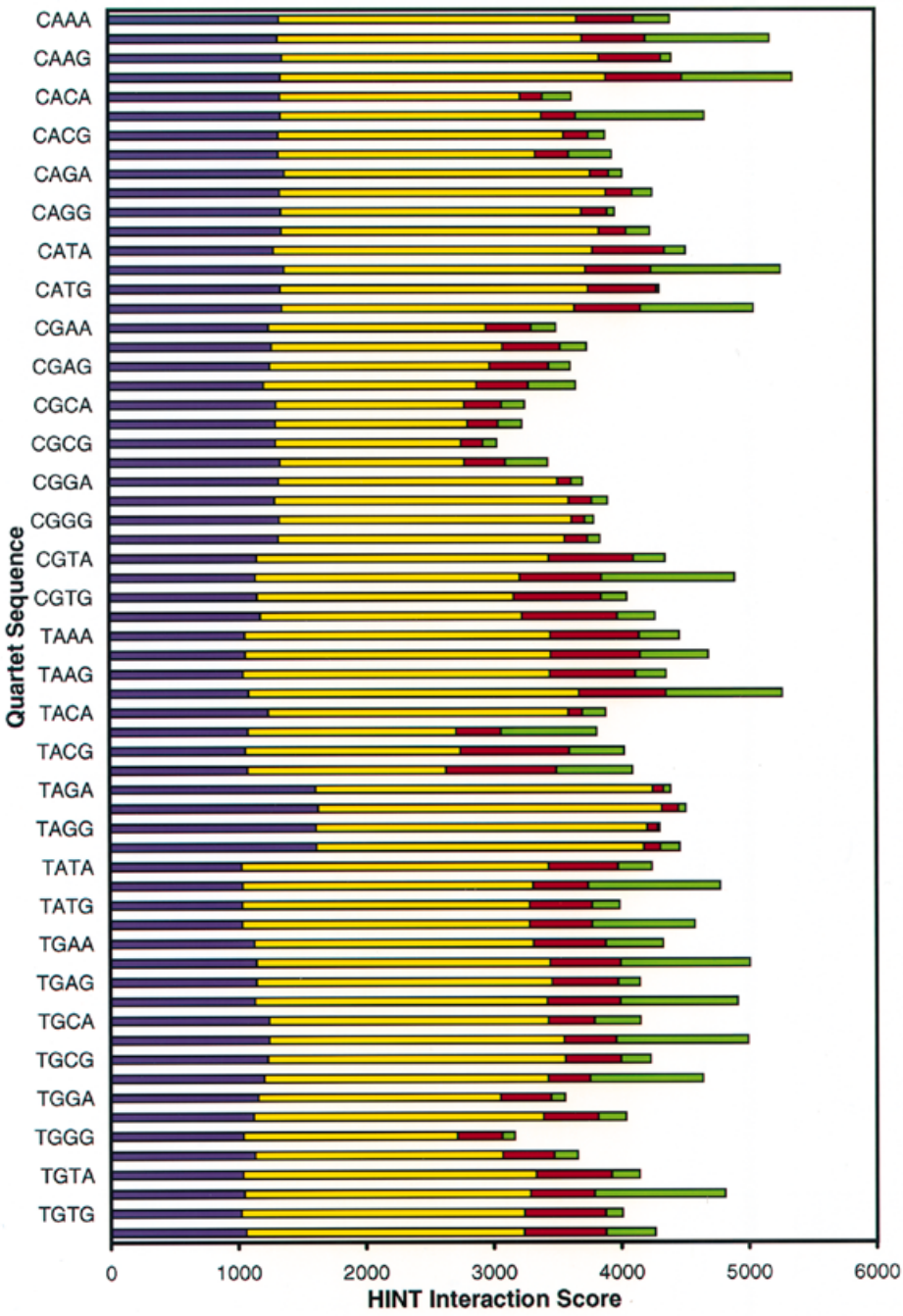


Figure 4. HINT interaction score by base pair quartet sequence (base position). Each bar has four segments representing the contribution from: blue, *[I–1]* base pair (above the intercalation site); yellow, *[I+1]* base pair (immediately below the intercalation site); red, *[I+2]* base pair; green, *[I+3]* base pair. For clarity on the quartet sequence axis, only xyzA and xyzG sequences are labeled. The unlabeled sequences are xyzC and xyzT, as alphabetically appropriate.

By examining the thermodynamic free energy of binding for doxorubicin and strategic analogs, Chaires *et al.* (26) recently reported estimates of group contributions to the overall binding free energy. For example, $\Delta\Delta G_t$ for **O9** is ~1.1 kcal/mol; $\Delta\Delta G_t$ for the sugar is ~2.0 kcal/mol; $\Delta\Delta G_t$ for **O14** is ~0.9 kcal/mol. We have recently reported that HINT scores can be related to $\Delta\Delta G$ for protein–protein associations and protein–ligand associations by equating 300–500 HINT score units with 1 kcal/mol (21). Analysis of the fractional HINT scores for these groups (averaged over all sequences) yields 370 ± 80 for **O9** and 1600 ± 500 for the daunosamine sugar, which are both in reasonable agreement with the Chaires results with this simplistic conversion of HINT score differences to $\Delta\Delta G$. However, the third group contribution, that of **O14**, does not appear significantly in the HINT analysis. Our models show **O14** at least 4.5 Å from the nearest potential hydrogen bond acceptor. The reason for this discrepancy is unclear, but it appears that the OH has assumed a neutral non-interacting position in our models. We tried to manually

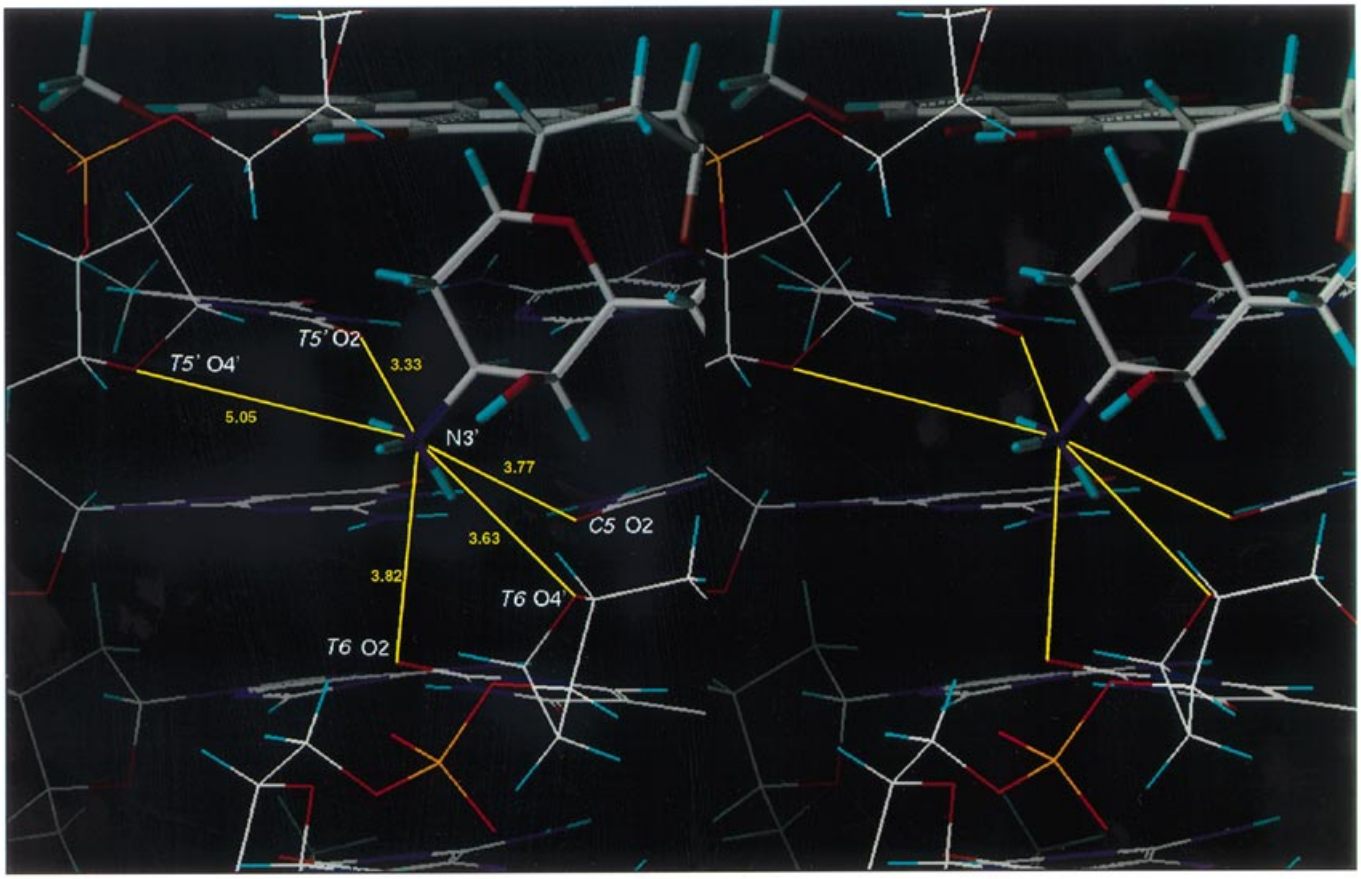


Figure 5. Stereoview of the region surrounding the N3′ ammonium in sequence CACT. The five potential hydrogen bond acceptor atoms of the nucleotide bases are labeled.

force hydrogen bond formation with nearby backbone oxygens, but models of this type had consistently poorer total HINT scores largely due to a host of induced hydrogen–polar interactions.

The interactions of **O14** are key to our purposes because they illuminate the question of whether the high resolution daunorubicin footprinting results (35) are directly applicable to doxorubicin as the difference between the two molecules is that daunorubicin lacks the **O14** hydroxide. As noted above, HINT is inconclusive. The Chen, Gresh and Pullman theoretical studies (13–15) show different results for the sequence specificity of daunorubicin [$d(CGATCG)_2 \geq d(CGTACG)_2 > d(TGATCA)_2 \geq d(TATATA)_2 >> d(CGCGCG)_2 > d(TACGTA)_2$] versus doxorubicin [$d(CGTACG)_2 > d(TATATA)_2 > d(TGATCA)_2 >> d(CGCGCG)_2 > d(TACGTA)_2$] Thus, based on their studies we cannot expect that daunorubicin and doxorubicin would prefer the same sequences. Nevertheless, while table S-1 (supplementary) of the Chaires *et al*. high resolution footprinting study on binding of daunorubicin (35) indicates that 9 of the 21 protected sequences contain the 'putative triplet binding sequence(s)' CG(A/T), GC(A/T) and (A/T)C(A/T) in agreement with the Pullman theoretical results, *eight contain the CAx and TAx triplets generally favored by HINT for doxorubicin*. All this indicates that doxorubicin/daunorubicin–DNA binding and associated sequence specificity is a very complex process.

Strategies for exploitation of structure–sequence selectivity relationships

The ultimate goal in defining sequence selectivity for a DNA intercalator drug is being able to use this information to predict, as accurately as possible, where the drug will bind. Actually, the invocation of a base pair quartet model to fully define the sequence selectivity for doxorubicin is a very positive result as it implies that it *may* be possible to restrict the binding of this class of drug to only one of the 256 potential DNA quartet sequence combinations. Two points are obvious, however: (i) both the results in this report and available experimental evidence from footprinting and other assay studies (11,16,17,35–38) show that doxorubicin is not a particularly selective intercalator; (ii) it is difficult to experimentally verify selectivity by structural means since intercalation into short chain oligonucleotides suitable for crystallographic or spectroscopic investigations is likely to be biased by preferential binding to the nucleotide ends.

Now that we have a more detailed model for doxorubicin–DNA binding interactions, it may be possible to increase the sequence selectivity by making chemical modifications to the doxorubicin 'lead' compound (1). The goal here would be to add structural features that would either make new site-specific interactions with a particular DNA sequence(s) or enforce known potential

interactions via a rigid analog approach. It would appear from this and previous studies that little can be gained by modifying the chromophore. It is the sugar portion of doxorubicin, not the chromophore, from which its sequence selectivity is derived. The daunosamine sugar does have several potential sites for chemical modification. It is also important to not neglect the potential utility of adding new hydrophobic–hydrophobic interactions. Our previous investigations unequivocally confirm that hydrophobic interactions contribute to robust binding environments in ligand–protein (18), inhibitor–enzyme (22) and protein–protein (21) systems. We are currently performing additional modeling studies of doxorubicin analogs to identify target structures for experimental investigation.

ACKNOWLEDGEMENTS

Significant portions of this project were performed by F.A.F. as part of the requirements for a course in molecular modeling offered by the Department of Medicinal Chemistry at VCU. We acknowledge the helpful comments and support of Drs Gwen B. Bauer, Donald J. Abraham and David A. Gewirtz in performing this research and preparing this manuscript. We also wish to thank Dr Nagarajan Pattabiraman for critical reading of the manuscript. G.E.K. and J.N.S. acknowledge the support of VCU. The SYBYL software has been made available through a University Software grant from Tripos Inc.

REFERENCES

1 Weiss,R.B. (1992) *Semin. Oncol.*, **19**, 670–686.
2 Henderson,I.C. and Canellos,G.P. (1990) *New Engl. J. Med.*, **302**, 78–90.
3 Fornari,F.A., Randolph,J.K., Yalowich,J.C., Ritke,M.K. and Gewirtz,D.A. (1994) *Mol. Pharmacol.*, **45**, 649–656.
4 Momparler,R.L., Karon,M., Siegel,S.E. and Avila,F. (1976) *Cancer Res.*, **36**, 2891–2895.
5 Tewey,K.M., Rowe,T.C., Yang,L., Halligan,B.D. and Liu,L.F. (1984) *Science*, **226**, 466–468.
6 Schneider,E., Hsiang,Y. and Liu,L.F. (1990) *Adv. Pharmacol.*, **21**, 149–183.
7 Sander,M. and Tsieh,T.-S. (1983) *J. Biol. Chem.*, **258**, 8421–8428.
8 Bachur,N.F., Yu,F., Johnson,R., Hickey,R., Wu,Y. and Malkas,L. (1992) *Mol. Pharmacol.*, **41**, 993–998.
9 George,J.W., Ghate,S., Matson,S.W. and Besterman,J.M. (1992) *J. Biol. Chem.*, **267**, 10683–10689.
10 Graves,D.E. and Krugh,T. (1983) *Biochemistry*, **22**, 3941–3947.
11 Cullinane,C., Cutts,S.M., van Rosmalen,A. and Phillips,D.R. (1994), *Nucleic Acids Res.*, **22**, 2296–2303.
12 Frederick,C.A., Williams,L.D., Ughetto,G., van der Marel,G., van Boom,J.H., Rich,A. and Wang,A.H.-J. (1990) *Biochemistry*, **29**, 2538–2549.
13 Pullman,B. (1991) *Anti-Cancer Drug Des.*, **7**, 95–105.
14 Chen,K.-X., Gresh,N. and Pullman,B. (1986) *Nucleic Acids Res.*, **14**, 2251–2267.
15 Chen,K.-X., Gresh,N. and Pullman,B. (1985) *J. Biomol. Struct. Dyn.*, **3**, 445–465.
16 Trist,H. and Phillips,D.R. (1989) *Nucleic Acids Res.*, **17**, 3673–3688.
17 Chaires,J.B. (1990) *Biophys. Chem.*, **35**, 191.
18 Wireko,F.C., Kellogg,G.E. and Abraham,D.J. (1991) *J. Med. Chem.*, **34**, 758–767.
19 Kellogg,G.E., Semus,S.F. and Abraham,D.J. (1991) *J. Comput. Aided Mol. Design*, **5**, 545–552.
20 Kellogg,G.E., Joshi,G.S. and Abraham,D.J. (1992) *Med. Chem. Res.*, **1**, 444–453.
21 Abraham,D.J., Kellogg,G.E., Holt,J.M. and Ackers,G.K. (1997) *J. Mol. Biol.*, **272**, 613–632.
22 Gussio,R., Pattabiraman,N., Zaharevitz,D.W., Kellogg,G.E., Topol,I.A., Rice,W.G., Schaeffer,C.A., Erickson,J.W. and Burt,S.K. (1996) *J. Med. Chem.*, **39**, 1645–1650.
23 Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.
24 Ornstein,R.L. and Rein,R. (1979) *Biopolymers*, **18**, 1277–1291.
25 Goodford,P.J. (1985) *J. Med. Chem.*, **28**, 857–864.
26 Chaires,J.B., Satyanarayana,S., Suh,D., Fokt,I., Przewloka,T. and Priebe,W. (1996) *Biochemistry*, **35**, 2047–2053.
27 Hansch,C. and Leo,A.J. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley & Sons, New York, NY.
28 Abraham,D.J. and Leo,A.J. (1987) *Proteins Struct. Funct. Genet.*, **2**, 130–152.
29 Levitt,M. (1983) *J. Mol. Biol.*, **168**, 595–620.
30 Levitt,M. and Perutz,M.F. (1988) *J. Mol. Biol.*, **201**, 751–754.
31 Huheey,J.E. (1978) *Inorganic Chemistry: Principles of Structure and Reactivity*. Harper and Row, New York, NY.
32 Wright,C.S. and Kellogg,G.E. (1996) *Protein Sci.*, **5**, 1466–1476.
33 Perutz,M.F., Fermi,G., Abraham,D.J., Poyart,C. and Bursaux,E. (1986) *J. Am. Chem. Soc.*, **108**, 1064.
34 Burley,S.K. and Petsko,G.A. (1986) *FEBS Lett.*, **201**, 751.
35 Chaires,J.B., Herrera,J.E. and Waring,M.J. (1990) *Biochemistry*, **29**, 6145–6153.
36 Skorobogaty,A., White,R.J., Phillips,D.R. and Reiss,J.A. (1988) *FEBS Lett.*, **227**, 103–106.
37 Skorobogaty,A., White,R.J., Phillips,D.R. and Reiss,J.A. (1988) *Drug Des. Del.*, **3**, 125–151.
38 Li,L. and Yu,F.-H. (1993) *Biochem. Mol. Biol. Int.*, **31**, 879–887.