A Computational Model for Anthracycline Binding to DNA: Tuning Groove-Binding Intercalators for Specific Sequences

Derek J. Cashman and Glen E. Kellogg*

Department of Medicinal Chemistry and Institute for Structural Biology & Drug Discovery, School of Pharmacy, Virginia Commonwealth University, P.O. Box 980540, Richmond, Virginia 23219-0540

Received October 17, 2003

Anthracycline antibiotics such as doxorubicin and its analogues have been in common use as anticancer drugs for almost half a century. There has been intense interest in the DNA binding sequence specificity of these compounds in recent years, with the hope that a compound could be identified that could possibly modulate gene expression or exhibit reduced toxicity. To computationally analyze this phenomenon, we have constructed molecular models of 65 doxorubicin analogues and their complexes with eight distinct DNA octamer sequences. The HINT (Hydropathic INTeractions) program was utilized to describe binding, including differences in the functional group contributions as well as sequence selectivity. Of these 65 compounds, two compounds were calculated to have a selectivity (the calculated $\Delta\Delta G_{sel}$ between the sequence with the strongest binding and the second strongest binding sequence) greater than −0.75 kcal mol^{-1} for one sequence over all others, 10 compounds were specific between −0.50 and −0.74 kcal mol^{-1}, 18 compounds were specific between −0.25 and −0.49 kcal mol^{-1}, and 35 compounds were virtually nonspecific with a $\Delta\Delta G$ below −0.24 kcal mol^{-1}. Several compounds have been identified from this study that include features which may enhance sequence selectivity, including several with a halogen in lieu of the 4′-OH in the daunosamine sugar, one compound with a nonaromatic six-membered ring (pirarubicin) in place of the 4′-OH, and a compound with an aromatic ring in the vicinity of the C^{14} region (zorubicin). Removal of the methoxy group at the C^4 position on the aglycone portion also appears to add potency and selectivity (idarubicin). Overall, efficient computational methods are presented that can be utilized to analyze the free energy of binding and sequence selectivity of both known and designed analogues of doxorubicin to identify future lead compounds for further experimental research.

Introduction

Nucleic acids, such as DNA and the closely related molecule, RNA, have been seen as potential therapeutic targets for drugs ever since the discovery of the DNA double helix 50 years ago by James D. Watson and Francis H. C. Crick.[1] To an organic chemist, the unique structural features of DNA due to the planar stacking of aromatic bases along the sugar phosphate backbone make it a particularly interesting target for drug design.[2] To a molecular biologist, the fact that the arrangement of bases in the DNA double helix contains the genetic blueprint of all of the proteins in a cell, combined with the recent completion of the mapping of the human genome, further fuels interest in the discovery of new compounds that bind to DNA and can have some significant therapeutic effect.[3]

Anthracycline antibiotics, such as doxorubicin and daunorubicin, are drugs that are known to bind to DNA and are some of the most common anticancer agents in use today.[4] Their main mode of action is still somewhat unclear, but is generally thought to involve binding to DNA by intercalation and inhibition of DNA biosynthesis, interference with topoisomerase II, and induction of DNA double strand breaks.[5−9] However, the mechanism of action of the anthracyclines is pleiotropic and may involve binding to molecules other than DNA.[4,10] This paper deals specifically with the DNA-binding mechanisms, with specific attention to the sequence specificity.

There has been much interest (and some controversy) in the determination of the sequence-specific binding affinity of doxorubicin. The ultimate goal, of course, is to identify a drug candidate that would have the ability to target a particular gene or promoter sequence and thus regulate its expression. Early X-ray crystallography studies and theoretical studies based on the crystal structures suggested specificity for the CG sequence.[11−13]

DNase I footprinting experiments and induction of stable transcriptional blockage sites suggested a specificity for GC or CA as the site of most favored intercalation.[14−16] Since anthracyclines contain a daunosamine sugar which interacts with the minor groove of the DNA, in addition to the intercalation of the aglycone moiety, a model emerged that suggested a triplet sequence, 5′-(A/T)CG or 5′-(A/T)GC, as the site of highest specificity.[13] This triplet consists of the base pairs immediately adjacent to the intercalation site, as well as a third base pair that interacts with the daunosamine sugar.[13] The major features of this triplet model were confirmed by the more recent in vitro experiments of Graves and Krugh, Trist and Phillips, and Chaires and crystallographic structures of the complexes reported by Frederick et al.[17−20] Recent

* To whom correspondence should be addressed. Phone: +1-804-828-6452, Fax: +1-804-827-3664, E-mail: glen.kellogg@vcu.edu.

10.1021/jm030529h CCC: $27.50
Published on Web 02/14/2004

research in our laboratory using the HINT (Hydropathic INTeractions) program has sought to build upon the structure and binding models for intercalation as defined by the triplet model to add features that would be useful for further and productive molecular design of the doxorubicin scaffold.[21]

However, one of the key difficulties in modeling these compounds has been the lack of a broad array of experimental thermodynamic binding data of these compounds. While there is quite a bit of data available on the toxicity of anthracyclines, as well as on their activity against cancer cells, there have been very few studies conducted to determine the actual binding efficiency of these compounds with DNA.[21,22] Furthermore, while the several crystal structure studies published do provide an excellent view of the 3-D conformation of doxorubicin itself, these structures were performed on a hexamer DNA sequence, i.e., where two doxorubicin molecules are intercalated into a sequence of DNA containing only six base pairs.[20] This creates a fairly crowded complex, and any interaction with base pairs beyond the third base pair is unlikely to be seen since the fourth and fifth base pairs involved in one interaction are also the second and third base pairs involved in the other interaction! Therefore, we believe that molecular modeling can help fill this gap in thermodynamic binding data, by correlating the binding data that is available with computational results obtained from models built upon the known crystal structures.

Our computational tool, the molecular modeling program HINT, calculates empirical, atom-based hydropathic parameters that encode all significant intermolecular and intramolecular noncovalent interactions implicated in the biological environment.[23] HINT is derived from hydrophobic atom constants that are based on thermodynamic, atom-based hydropathy values from solvent partitioning measurements (Log P) of organic molecules.[24] These atom constants are used to quantitatively score the strength of all noncovalent interactions in a molecular system, and these scores have recently been shown to correlate with experimental free energy for several protein systems,[25,26] and for aminoglycoside binding to 16S ribosomal RNA.[27]

The HINT program was also used to calculate binding scores for the interaction of doxorubicin with 64 DNA base pair quartet sequences, and in several HINT models, an interaction between the N3′ ammonium group of the daunosamine sugar and the carbonyl of the fourth base pair was found, invoking a *quartet* model for binding.[21] The highest binding sequence from this study was found to be the C|AAT sequence (the "|" symbol indicates the intercalation site of the doxorubicin chromophore, i.e., between the C and A base pairs).[21] Interestingly enough, a C|AAT sequence has recently been shown to be involved in the upregulation of the *MDR1* promoter sequence in HL60/VCR cells, which is implicated in multi-drug-resistant tumors.[10] So this information, combined with new data obtained from sequencing the human genome, has the potential to identify key promoter sequence sites for possible targeting by DNA-binding drugs.[3]

To further investigate this quartet model, and to calibrate our computational methods, calculated free energies were recently compared to experimental data of eight doxorubicin analogues measured and reported by Chaires.[22,28] The free energies of these eight compounds were calculated as intercalated in 32 DNA quartet sequences, and the difference in free energy ($\Delta\Delta G$) for each compound versus doxorubicin was computationally determined, indicating the relative energetic contribution of the functional group present on that compound.[28] The calculated $\Delta\Delta G$ values from our molecular models were found to correlate with the experimental data in relatively good agreement, with a few notable, yet explainable, differences.[28] Even in this limited test set we nonetheless saw evidence of sequence specificity, as evidenced by large standard deviations of the average calculated free energy from the interaction of each compound with 32 unique sequences of DNA.[28]

The next step is to locate additional doxorubicin analogues to determine if differences in structure would provoke differences or improvements in sequence specificity. The program Unity (Tripos, Inc., St. Louis, MO) was thus employed to perform a 3-D structural search of the 213 628 compounds in the NCI Database for analogues of doxorubicin. The search was ultimately narrowed down to 65 compounds (Schemes 1−5), including our original test set. Since DNA consists of four unique nucleic acid bases and thus 256 possible DNA quartet combinations, the problem is that we need to model quite a number of potential "receptors." Starting with an intercalation site of C|A or C|G, we varied the third and fourth base pair with a set of representative combinations (Figure 1). Complex structures were calculated for these 65 compounds versus eight DNA quartet sequences (C|AAT, C|AAG, C|GAG, C|GAT, C|AAA, C|ATA, C|ACC, and C|AGC). The $\Delta\Delta G$ was then calculated between the highest binding sequence and the second highest binding sequence for each compound to calculate selectivity.

While a truly sequence-specific drug may still be a long way off, these calculations provide detailed and comprehensive insight into the three-dimensional binding of doxorubicin analogues with a variety of sequences. Most significantly, these compounds can certainly be used as molecular tools for the development of future sequence-specific binding agents.

Results and Discussion

Correlation with Biological Data. The first step of this study was to validate our computational procedure against existing biological data. To accomplish this, we modeled a "test set" of a series of eight doxorubicin analogues that had previously been assayed for binding to calf thymus DNA by Chaires (Scheme 1).[22] Not only was an experimental binding value available for each compound, but most significantly, since each compound contained a simple functional group modification from the original doxorubicin molecule itself, Chaires has shown that it was possible to separate each functional group contribution into partial free energy binding interactions.[22] We applied these same techniques to our "test set" of doxorubicin analogues.[28]

Using our previously described methods, we calculated binding free energies for our "test set" of eight doxorubicin analogues versus a series of DNA quartet

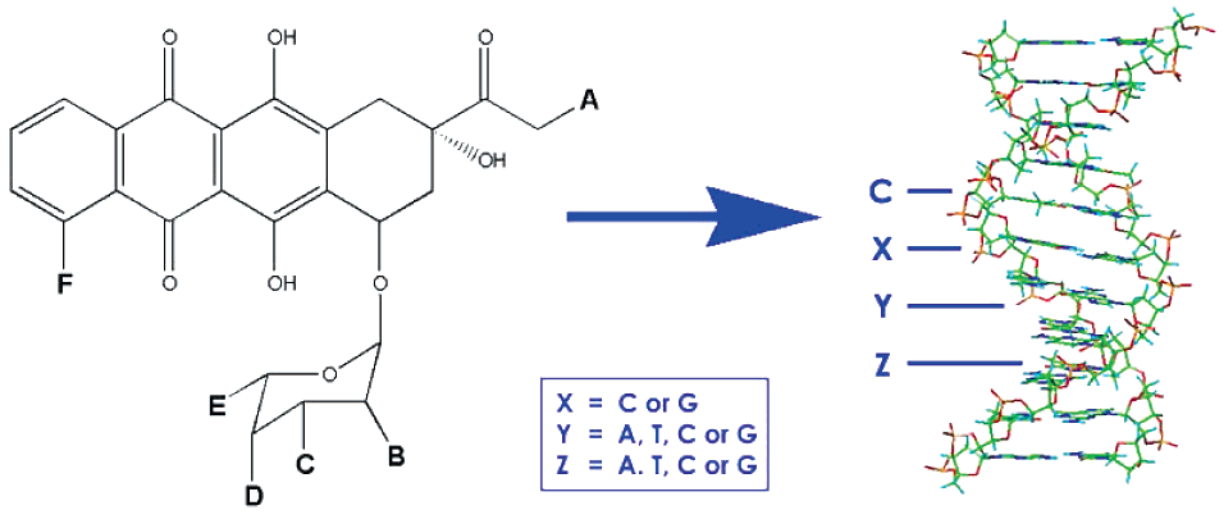
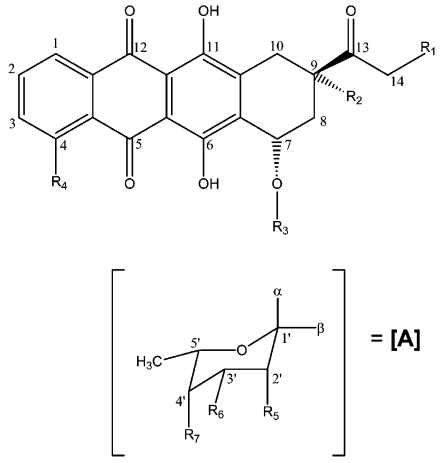


Figure 1. Diagram indicating the various modifications of the doxorubicin molecule as well as the base pairs of the DNA double helix that were mutated. The anthracycline intercalation site is between the "C" and "X" base pairs.

Scheme 1. "Test Set" of Eight Doxorubicin Derivatives



1. Doxorubicin[22]: R_1 = -OH, R_2 = -OH, R_3 = [A] (α), R_4 = -OCH_3, R_5 = -H, R_6 = -NH_3^+, R_7 = -OH
2. Daunorubicin[22]: R_1 = -H, R_2 = -OH, R_3 = [A] (α), R_4 = -OCH_3, R_5 = -H, R_6 = -NH_3^+, R_7 = -OH
3. Hydroxydoxorubicin[22]: R_1 = -OH, R_2 = -OH, R_3 = [A] (α), R_4 = -OCH_3, R_5 = -H, R_6 = -OH, R_7 = -OH
4. 9-Dehydroxydoxorubicin[22]: R_1 = -OH, R_2 = -H, R_3 = [A] (α), R_4 = -OCH_3, R_5 = -H, R_6 = -NH_3^+, R_7 = -OH
5. Adriamycinone[22]: R_1 = -OH, R_2 = -OH, R_3 = -H, R_4 = -OCH_3
6. Daunomycinone[22]: R_1 = -H, R_2 = -OH, R_3 = -H, R_4 = -OCH_3
7. WP-608[22]: R_1 = -H, R_2 = -OH, R_3 = [A] (α), R_4 = -OCH_3, R_5 = -H, R_6 = -OH, R_7 = -NH_3^+
8. Doxorubicin (β-Anomer)[22]: R_1 = -OH, R_2 = -OH, R_3 = [A] (β), R_4 = -OCH_3, R_5 = -H, R_6 = -NH_3^+, R_7 = -OH

sequences.[28] Since our method is structurally based, and therefore, sequence-dependent, we needed to calculate an average value of a representative sample of sequences to correlate this with Chaires's experimental, non-sequence-dependent, binding data. Initially, we took the average of 32 DNA quartet sequences. These sequences started with the C|AAT sequence and first kept the C|A base pairs constant and varied the second two base pairs. Second, we kept the AT base pairs constant and varied the C|A base pairs. This ultimately contributed to 32 DNA quartet sequences in our model.[28]

However, since this is a fairly large number of sequences and sampling a good number of the more than 2000 known doxorubicin analogues would naturally require extraordinarily large amounts of computer and operator time, we also wanted to test a model in which fewer sequences could be modeled. Since there is a reasonably good amount of evidence suggesting C|A or C|G as the site of intercalation,[11−16] we desired to focus our calculations on modifications to the 3rd and 4th base pairs of the quartet. So eight sequences were selected for our smaller test (C|AAT, C|AAG, C|GAG, C|GAT, C|AAA, C|ATA, C|ACC, and C|AGC), in the hope that this reduced set would also be a representative sample of the whole.

The experimental and calculated free energy (ΔG), as well as the difference in free energy between each compound and doxorubicin ($\Delta\Delta G_{dox}$) (both in kcal mol^{-1}), is shown in Table 1. Calculation of the $\Delta\Delta G_{dox}$ for each compound allowed us to determine the net energetic contribution of specific functional groups to binding. Overall, while the ΔG is somewhat overestimated for our "test set," the $\Delta\Delta G_{dox}$ values show relatively good agreement. Most of our calculated free energies show a relatively high standard deviation when compared to the biological data, which can be interpreted as a consequence of sequence-dependent binding for these compounds. It is also important to note that the reported errors between the experimental and modeled data arise from two separate phenomena. The standard deviations reported for the experimental data arise from the calculated average of multiple thermodynamic binding data measurements of each doxorubicin analogue bound in a non-sequence-dependent manner to calf thymus DNA, whereas the standard deviations reported for the modeled data arise from the calculated average of individual free energy calculations from 32 (or 8) specific DNA quartet sequences.[22,28]

We did see a noticeable sign change disagreement with the $\Delta\Delta G_{dox}$ for daunorubicin, in which we predicted that the removal of the C^{14}-OH would provide for a more favorable interaction by 0.4 ± 0.6 kcal mol^{-1}, while experimental data indicated that removal of that group would reduce the overall interaction by 0.9 ± 0.5 kcal mol^{-1}. If we look at an interactive map contour (Figure 2), which visually displays the quantity and magnitude of intermolecular interactions, we see an unfavorable hydrophobic interaction caused by the close proximity of the C^{14} oxygen atom with the methylene atoms of the phosphate backbone of the DNA. Moreover, we detected a fairly large standard deviation in our calculations, so there is likely some uncertainty caused by the flexibility of the C^{14} side chain, in addition to some potential solvent interactions.[28]

Table 1. Experimental Free Energy (ΔG) and the Free Energy Difference ($\Delta\Delta G$) between Each Compound and Doxorubicin

	experimental[a] (kcal mol^{-1})		calculated[b] (kcal mol^{-1}) average of 32 sequences		calculated[b] (kcal mol^{-1}) average of 8 sequences	
	ΔG	$\Delta\Delta G$	ΔG	$\Delta\Delta G$	ΔG	$\Delta\Delta G$
doxorubicin	-7.7 ± 0.3	- -	-9.6 ± 1.1[c]	- -	-9.4 ± 0.9	- -
daunorubicin (loss of C14-OH)	-6.8 ± 0.3	-0.9 ± 0.5	-10.0 ± 0.9[c]	$+0.4 \pm 0.6$[c]	-9.7 ± 0.8	$+0.3 \pm 0.5$
hydroxydoxorubicin (NH_3^+ to OH)	-7.0 ± 0.3	-0.7 ± 0.5	-5.9 ± 0.4[c,d]	-3.7 ± 1.1[c,d]	-6.0 ± 0.3[d]	-3.4 ± 1.0[d]
9-dehydroxydoxorubicin (loss of C9-OH)	-6.5 ± 0.3	-1.2 ± 0.5	-8.9 ± 1.0[c]	-0.7 ± 0.7[c]	-8.6 ± 0.9	-0.7 ± 0.6
adriamycinone (loss of sugar ring)	-5.7 ± 0.3	-2.0 ± 0.5	-6.3 ± 0.6[c]	-3.2 ± 1.1[c]	-6.7 ± 0.3	-2.7 ± 1.0
daunomycinone (loss of sugar ring and C14-OH)	-5.2 ± 0.3	-2.5 ± 0.5	-6.0 ± 0.5[c]	-3.6 ± 1.1[c]	-6.4 ± 0.4	-3.0 ± 0.9
WP-608 (swap C3′-NH_3^+ and C4′-OH)	-6.1 ± 0.3	-1.6 ± 0.5	-7.1 ± 0.5	-2.6 ± 1.2	-7.1 ± 0.5	-2.3 ± 1.1
β-anomer of doxorubicin (sugar is attached β)	-4.5 ± 0.3	-3.2 ± 0.5	-5.3 ± 0.5	-4.2 ± 1.3	-5.3 ± 0.5	-4.1 ± 0.9

[a] Experimental data was obtained from the binding of doxorubicin analogues with calf thymus DNA, from Chaires et al.[27] [b] Two sets of calculated data were obtained from molecular models of each analogue intercalated into a series of 32 and 8 DNA quartet sequences. A HINT interaction score was obtained and divided by -515 (units per kcal mol^{-1}), and the average score from all 32 or 8 models per compound is reported. The $\Delta\Delta G$ for all compounds versus all sequences was also calculated, and the average score from all 32 or 8 models per compound is reported. [c] Reference 28. [d] Assuming doxorubicin N3′ is in the protonated form, we also modeled doxorubicin in the unprotonated (NH_2) form and computed a $\Delta\Delta G$ for hydroxydoxorubicin of -0.2 ± 0.3 kcal mol^{-1}.

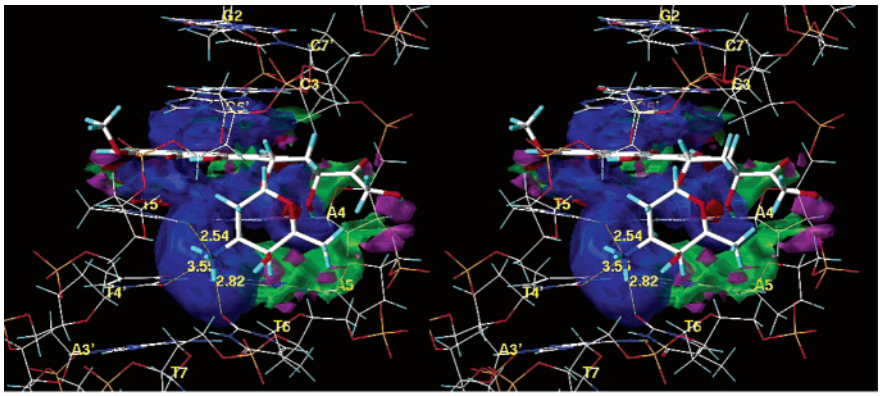


Figure 2. Stereodiagram of a HINT interaction map for the intercalation of doxorubicin with the CAAT base pair sequence of DNA, which displays, visually, the quality and magnitude of the various binding contacts involved in the interaction. The contour surfaces are color-coded by interaction type at a constant map density value of ±80. Blue surfaces represent favorable polar interactions, red surfaces represent unfavorable polar interactions, green surfaces represent favorable hydrophobic interactions, and magenta surfaces represent unfavorable hydrophobic interactions. The interatomic distance between the N3′ ammonium on doxorubicin's sugar ring and the carbonyl oxygen atoms of three surrounding base pairs of DNA is also indicated.

Our calculations also appear to overestimate the binding contribution of the N3′ ammonium by a fairly substantial amount (hydroxydoxorubicin). While we have previously discussed this in much greater detail in an earlier publication,[28] this discrepancy is mainly attributable to the loss of three potent hydrogen bonds between the three hydrogen atoms of the amine cation and three carbonyl oxygen atoms of the DNA bases nearby. The neutral hydroxyl oxygen, with only a single hydrogen atom, does not have the same hydrogen bonding capability as this positively charged amine, and therefore we see a significantly reduced polar interaction for hydroxydoxorubicin. To further analyze the protonation state of this N3′ ammonium, we created a model of an unprotonated doxorubicin vs all 32 DNA quartet sequences, and used this to calculate the $\Delta\Delta G$ for hydroxydoxorubicin. The $\Delta\Delta G$ between hydroxydoxorubicin and this unprotonated form of doxorubicin was calculated to be -0.2 ± 0.3 kcal mol^{-1}. So, we therefore believe that our overestimation of this functional group contribution is due to the fact that there likely is a reduced, or partial, protonation state of this N3′ ammonium when doxorubicin is bound to DNA.[28]

Overall, the calculations correlate quite well to the biological data. We graphed the calculated vs experimental ΔG values for both the 8-sequence and 32-sequence models and found little variation in the plotted points for both average data sets. (Figure 3). Comparison of the data points of both the 8-sequence and 32-sequence models, shows that the 8-sequence model is as significant in our calculations as the 32-sequence model, indicating that the set of eight we have chosen does indeed provide a sufficient representative sample of sequences for our study.

Although the correlation was reasonably good, there was one significant outlier in both models—the estimation in free energy of hydroxydoxorubicin. Naturally, leaving out this outlier would result in a much better correlation ($r^2 = 0.88$). However, if we plot the average free energy of the completely unprotonated forms of the doxorubicin analogues that contain the amine, we can hypothesize that there is a continuum of points on the graph representing partially protonated states of these compounds. We can also see that calculated free energy values for these four compounds at a chosen point along this continuum would result in an improved correlation, indicating that the amine is indeed most likely "partially" protonated.

Functional Group Contribution Calculations. Having demonstrated a good correlation between the $\Delta\Delta G_{dox}$ values for the "test set" of compounds, we extended this procedure to our computational data set to calculate the net energetic contribution of various functional groups of other doxorubicin analogues (Table 2). (Tables 4 and 5, presented as Supporting Information, present the HINT score data for each complex by

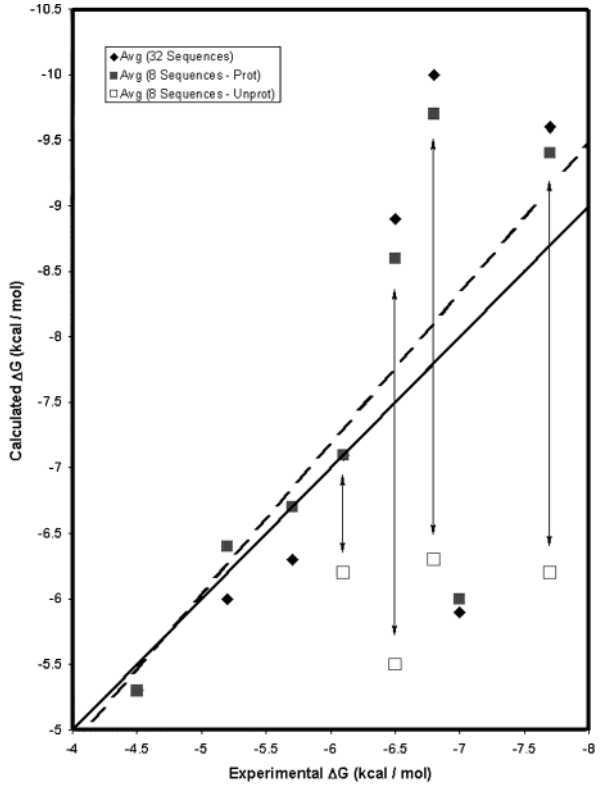


Figure 3. Combined graphs of the calculated ΔG vs experimental ΔG (kcal mol^{-1}) for the "Test Set" of eight doxorubicin analogues modeled in 32 DNA "quartet" sequences and 8 DNA "quartet" sequences (using both protonated and unprotonated models). The dashed line indicates a linear regression fit of the average of 8 sequences (slope = 1.14); the solid line indicates where a perfect linear regression fit (slope = 1) would be. The arrows indicate the range in calculated free energy between the protonated and unprotonated models. The r^2 for the average of 8 sequences is 0.53, and the r^2 for the average of 32 sequences is 0.52.

individual base pair and interaction type contributions, respectively). We must exercise a certain degree of caution, however, since the compounds of the original "test set" differed from doxorubicin by a single functional group, allowing the determination of the net energetic contribution of that specific functional group to binding. There are several compounds in our computational data set, however, that differ by multiple or complex functional groups, and this could potentially add to some confusion in interpreting the data. Nonetheless, this information, combined with the calculations on sequence selectivity also presented here, should still prove useful in the design of new, more potent and selective (vide infra) doxorubicin analogues.

Overall, a decrease in the net energetic contribution ($-\Delta\Delta G_{dox}$) was calculated for 37 of the 65 compounds modeled, while an increase in the net energetic contribution ($+\Delta\Delta G_{dox}$) was calculated for 27 compounds.

1. Effects of Halogenation on Binding. Compounds **9** through **12** each contain a simple modification in which the C4′-OH is replaced with a halogen. We see a gradual increase in the $\Delta\Delta G_{dox}$, from +0.7 for fluorodoxorubicin, to +1.3 to bromodoxorubicin and +1.2 for iododoxorubicin. This increased free energy contribution is likely due to an increased hydrophobic interaction between the larger halogen atoms and the methylene groups of the phosphate backbone of the DNA. Interestingly enough, in compounds **13** to **16**, which contain the respective halogen in place of the C14-OH, we do not see this same trend. As a matter of fact, we really see very little increase in $\Delta\Delta G_{dox}$ for any of these compounds, which further contributes to our earlier argument that this particular area of doxorubicin, being solvent exposed and flexible, does not contribute significantly to binding.

We see another increase in $\Delta\Delta G_{dox}$ for compounds **17** to **20**, that contain a halogen at the C2′ in the sugar, which is somewhat less significant for the fluoro- and chloro-containing compounds (**19**, **20**). This is likely due to the decrease in solvent accessibility which would reduce the solvent's ability to stabilize the amine cation, resulting in a more protonated state and increasing its interaction with the carbonyl oxygens of the T4′ and T5′ base pairs. There also appears to be favorable hydrophobic interactions between the halogen atoms and the methylene backbone of the DNA. Note, however that these compounds also contain an inversion in the stereochemistry of the C4′-OH, as seen in epirubicin (**59**). As the $\Delta\Delta G_{dox}$ for epirubicin is +0.7 kcal mol^{-1}, we can see that a major part of the free energy difference of compounds **17** to **20** is tied to that functional group. However, we can still make the argument that an iodo-substitution at the C2′ position has potential to increase binding.

The other halogenated compounds of our computational data set (**21** to **28**) indicate a significant net energetic decrease ($-\Delta\Delta G_{dox}$). This actually appears to be due to the lack of a protonated functional group instead of the halogenation. However, compounds **21**−**24**, which are identical to compounds **17**−**20**, except that the C3′ amine is replaced with a hydroxyl, show an average increase in calculated ΔG by approximately −6.7 kcal mol^{-1}, which is about 0.7 kcal mol^{-1} more favorable than the ΔG for hydroxydoxorubicin (**3**). So it does appear that C2′ halogenation, combined with the epimerization of C4′, would have a minor effect on free energy. We did not see any significant correlation between the level of halogenation and free energy difference, however.

2. Effects of Converting the C3′ Amine to an Amide with Varying Substituents. Most of the compounds in the computational data set containing amide moieties in place of the C3′ amine (compounds **29** to **39**) also indicated a significant net energetic reduction in binding. Although the $\Delta\Delta G_{dox}$ of most of these compounds is actually better compared to daunorubicin than doxorubicin, since they also lack the C14-OH. However, since the calculated $\Delta\Delta G_{dox}$ between doxorubicin and daunorubicin is only +0.3 kcal mol^{-1}, we can still analyze these compounds reasonably well with our existing calculated $\Delta\Delta G_{dox}$ values.

Like other compounds lacking the C3′ protonated amine, we see a notably reduced $\Delta\Delta G_{dox}$. However, particularly with the urea compounds (**30** to **35**), which contain an extra NH group distal to the amide, the $\Delta\Delta G_{dox}$ is less dramatic than for acetamide (**29**). This is primarily because the NH portion of the amide sits in a position between the carbonyl oxygen atoms of the T′ bases of the second and third base pairs, forming a bifurcated hydrogen bond between the two carbonyl oxygen atoms (Figure 5). While this alone provides a

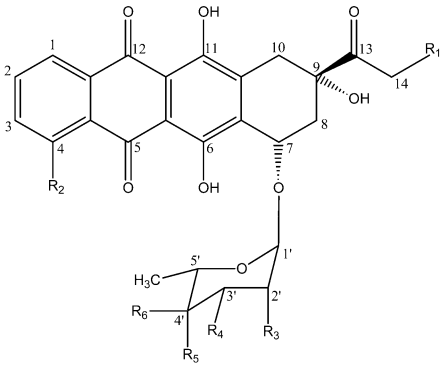
Table 2. Calculated Free Energies (ΔG; in kcal mol^{-1}) for 65 Doxorubicin Derivatives vs Eight DNA Quartet Sequences[a]

compound	CAAT	CAAG	CGAG	CGAT	CAAA	CATA	CACC	CAGC	avg ΔG	avg $\Delta\Delta G_{dox}$
Test Set										
1. doxorubicin	−10.6	−9.1	−8.1	−9.6	−8.7	−9.3	−10.5	−9.2	−9.4 ± 0.9	- -
2. daunorubicin	−10.5	−8.7	−8.4	−10.2	−9.3	−10.5	−10.4	−9.5	−9.7 ± 0.8	+0.3 ± 0.5
3. hydroxydoxorubicin	−6.4	−6.1	−6.3	−6.2	−6.1	−5.7	−5.5	−5.8	−6.0 ± 0.3	−3.4 ± 1.0
4. 9-dehydroxydoxorubicin	−9.5	−8.3	−7.0	−9.4	−8.4	−8.2	−9.0	−9.4	−8.6 ± 0.9	−0.7 ± 0.6
5. adriamycinone	−6.3	−6.8	−6.4	−6.5	−7.1	−6.6	−6.8	−7.0	−6.7 ± 0.3	−2.7 ± 1.0
6. daunomycinone	−6.7	−7.1	−5.8	−5.9	−6.6	−6.3	−6.2	−6.4	−6.4 ± 0.4	−3.0 ± 0.9
7. WP-608	−7.1	−7.2	−6.9	−7.2	−7.4	−8.1	−6.3	−6.6	−7.1 ± 0.5	−2.3 ± 1.1
8. doxorubicin (β-Anomer)	−5.4	−6.1	−5.3	−5.2	−4.2	−5.5	−5.5	−5.4	−5.3 ± 0.5	−4.1 ± 0.9
Halogenated Derivatives										
9. iododoxorubicin	−11.4	−11.1	−11.0	−11.2	−11.1	−8.8	−10.4	−9.5	−10.6 ± 0.9	+1.2 ± 1.2
10. bromodoxorubicin	−11.9	−11.0	−10.9	−11.2	−11.2	−9.1	−10.5	−9.3	−10.7 ± 1.0	+1.3 ± 1.2
11. chlorodoxorubicin	−11.6	−10.9	−10.8	−11.1	−11.0	−9.3	−10.4	−9.0	−10.5 ± 0.9	+1.1 ± 1.1
12. fluorodoxorubicin	−10.6	−10.5	−10.6	−9.5	−9.8	−9.1	−10.1	−10.5	−10.1 ± 0.6	+0.7 ± 1.1
13. C14-iododoxorubicin	−11.4	−11.1	−11.0	−11.2	−11.1	−8.8	−10.4	−9.5	−10.6 ± 0.9	+0.5 ± 0.6
14. C14-bromodoxorubicin	−10.9	−9.1	−8.4	−10.6	−9.9	−9.2	−10.3	−10.0	−9.8 ± 0.8	+0.4 ± 0.5
15. C14-chlorodoxorubicin	−11.6	−10.9	−10.8	−11.1	−11.0	−9.3	−10.4	−9.0	−10.5 ± 0.9	+0.5 ± 0.5
16. C14-fluorodoxorubicin	−10.7	−9.0	−8.0	−10.6	−9.8	−9.2	−10.4	−10.0	−9.7 ± 0.9	+0.3 ± 0.5
17. 2′-iodo-4′-epidaunorubicin	−11.7	−11.5	−9.5	−9.7	−10.6	−10.2	−12.5	−11.5	−10.9 ± 1.1	+1.5 ± 0.8
18. 2′-bromo-4′-epidaunorubicin	−11.4	−11.7	−9.1	−9.8	−10.5	−10.0	−10.9	−11.0	−10.5 ± 0.9	+1.1 ± 0.9
19. 2′-chloro-4′-epidaunorubicin	−11.4	−11.4	−9.1	−9.4	−11.9	−10.0	−9.0	−9.4	−10.2 ± 1.2	+0.8 ± 1.5
20. 2′-fluoro-4′-epidaunorubicin	−11.2	−11.1	−8.6	−9.9	−11.5	−9.5	−8.9	−9.3	−10.0 ± 1.1	+0.6 ± 1.3
21. annamycin (C2′-iodo)	−6.9	−7.0	−6.9	−6.4	−7.2	−7.3	−6.2	−6.4	−6.8 ± 0.4	−2.6 ± 1.1
22. annamycin (C2′-bromo)	−6.8	−7.1	−6.9	−6.5	−7.2	−7.6	−6.1	−6.3	−6.8 ± 0.5	−2.6 ± 1.1
23. annamycin (C2′-chloro)	−6.7	−7.1	−6.7	−6.2	−7.0	−7.5	−5.9	−6.3	−6.7 ± 0.5	−2.7 ± 1.2
24. annamycin (C2′-fluoro)	−6.5	−6.6	−6.7	−6.2	−7.1	−7.1	−6.1	−6.3	−6.6 ± 0.4	−2.8 ± 1.1
25. *N*-(iodoacetyl) doxorubicin	−6.3	−6.5	−6.6	−6.2	−6.2	−5.7	−5.4	−6.1	−6.1 ± 0.4	−3.3 ± 1.2
26. *N*-(bromoacetyl) doxorubicin	−7.2	−7.1	−6.7	−6.6	−6.3	−5.9	−5.3	−5.5	−6.3 ± 0.7	−3.0 ± 1.2
27. *N*-(chloroacetyl) doxorubicin	−6.9	−7.2	−6.9	−6.6	−6.4	−6.1	−5.4	−5.4	−6.4 ± 0.7	−3.0 ± 1.2
28. *N*-(fluoroacetyl) doxorubicin	−6.6	−6.6	−6.3	−6.1	−6.9	−5.7	−5.3	−5.3	−6.1 ± 0.6	−3.3 ± 1.2
Amide Derivatives										
29. N3′-acetamide	−5.9	−5.6	−6.0	−5.8	−6.0	−6.1	−4.8	−5.4	−5.7 ± 0.4	−3.7 ± 1.1
30. N3′-formyl daunorubicin	−7.4	−7.5	−6.9	−6.8	−7.2	−6.2	−5.9	−5.9	−6.7 ± 0.6	−2.7 ± 1.2
31. N3′-methyl Urea	−8.4	−8.0	−6.7	−7.0	−8.4	−7.1	−7.2	−6.3	−7.4 ± 0.8	−2.0 ± 1.0
32. N3′-butyl Urea	−8.7	−8.3	−6.8	−7.0	−7.9	−7.3	−6.6	−7.3	−7.5 ± 0.7	−1.9 ± 1.0
33. N3′-methylthio Urea	−7.1	−7.7	−6.7	−7.8	−7.0	−8.0	−7.2	−6.3	−7.2 ± 0.6	−2.2 ± 0.9
34. N3′-butylthioUrea	−7.3	−6.8	−6.5	−6.8	−6.7	−7.0	−6.5	−7.2	−6.8 ± 0.3	−2.5 ± 0.8
35. N3′-phenylthio Urea	−7.7	−8.2	−8.3	−9.6	−7.5	−8.5	−8.9	−8.4	−8.4 ± 0.7	−1.0 ± 1.0
36. N3′-dimethylglycine daunorubicin	−7.1	−6.6	−7.1	−6.9	−7.4	−7.0	−6.0	−6.7	−6.8 ± 0.4	−2.5 ± 1.1
37. N3′-glycine daunorubicin	−13.2	−10.8	−10.9	−13.3	−11.8	−10.1	−12.0	−12.1	−11.8 ± 1.1	+2.4 ± 1.0
38. doxorubicin 3′-4′-Diacetate	−5.4	−5.2	−5.4	−5.1	−4.3	−4.2	−4.6	−4.6	−4.8 ± 0.5	−4.5 ± 1.0
39. trifluoroacetamide	−7.0	−7.5	−6.8	−6.9	−7.5	−7.1	−5.7	−7.4	−7.0 ± 0.6	−2.4 ± 1.3
Derivatives with Additional Rings										
40. carboxidaunorubicin γ-lactam	−5.9	−6.2	−6.1	−6.1	−5.9	−5.6	−5.6	−6.3	−6.0 ± 0.3	−3.4 ± 1.0
41. cyanomorpholino doxorubicin	−5.4	−5.6	−5.4	−5.6	−5.8	−5.2	−3.9	−5.2	−5.3 ± 0.6	−4.1 ± 1.3
42. isothiocyanatobenzoyl doxorubicin	−8.4	−7.9	−7.4	−8.2	−8.3	−7.2	−6.9	−6.9	−7.7 ± 0.6	−1.7 ± 1.1
43. MX-2	−6.7	−7.0	−6.5	−7.1	−6.9	−6.7	−6.0	−5.9	−6.6 ± 0.4	−2.8 ± 1.0
44. *N*-piperidinoimine	−11.3	−9.9	−9.5	−11.8	−10.2	−9.8	−10.9	−9.5	−10.4 ± 0.9	+1.0 ± 0.7
45. pirarubicin	−9.4	−9.5	−8.8	−9.3	−9.9	−8.6	−11.0	−10.3	−9.6 ± 0.8	+0.2 ± 0.9
46. WPX-01	−8.8	−8.1	−7.8	−8.0	−8.6	−7.3	−7.6	−7.6	−8.0 ± 0.5	−1.4 ± 0.9
47. WPX-02	−9.1	−8.5	−8.0	−8.5	−9.2	−8.4	−7.6	−7.7	−8.4 ± 0.6	−1.0 ± 1.0
48. WPX-03	−8.9	−8.1	−8.0	−8.6	−8.9	−7.1	−7.0	−7.4	−8.0 ± 0.8	−1.4 ± 1.2
49. WPX-04	−8.2	−7.9	−7.6	−8.1	−8.3	−7.3	−7.4	−7.3	−7.8 ± 0.4	−1.6 ± 0.9
50. WP-744	−11.4	−10.8	−10.5	−11.7	−12.2	−9.9	−10.7	−9.5	−10.8 ± 0.9	+1.5 ± 1.2
51. WP-744−01	−11.5	−11.4	−11.3	−11.6	−11.9	−10.1	−10.5	−9.5	−11.0 ± 0.8	+1.6 ± 1.3
52. zorubicin	−12.2	−10.6	−9.3	−11.7	−10.6	−11.5	−11.4	−9.7	−10.9 ± 1.0	+1.5 ± 0.6
53. bromo-zorubicin	−12.7	−9.9	−10.3	−11.1	−11.3	−11.8	−10.1	−10.5	−11.0 ± 0.9	+1.6 ± 1.0
Miscellaneous Derivatives										
54. analogue VIII	−7.3	−7.3	−5.6	−6.4	−7.7	−7.3	−6.8	−7.4	−7.0 ± 0.7	−2.4 ± 0.9
55. analogue IX	−8.8	−8.6	−7.3	−7.4	−9.2	−8.4	−8.4	−8.7	−8.3 ± 0.7	−1.0 ± 0.9
56. analogue X	−7.2	−7.2	−6.9	−6.8	−7.2	−7.2	−6.6	−7.0	−7.0 ± 0.2	−2.4 ± 0.9
57. analogue XI	−10.8	−9.3	−7.8	−10.9	−9.7	−10.8	−10.6	−9.2	−9.9 ± 1.1	+0.8 ± 0.7
58. carminomycin	−11.2	−11.1	−9.7	−11.4	−11.5	−9.8	−10.7	−9.5	−10.6 ± 0.8	+1.2 ± 1.0
59. epirubicin	−11.3	−11.0	−8.0	−9.1	−11.3	−10.8	−9.5	−9.8	−10.1 ± 1.2	+0.7 ± 1.2
60. idarubicin	−11.7	−11.2	−9.6	−10.9	−10.7	−10.1	−10.6	−10.1	−10.6 ± 0.7	+1.2 ± 0.7
61. ME-2303	−6.7	−6.7	−6.2	−6.3	−6.8	−5.7	−5.9	−6.0	−6.3 ± 0.4	−3.1 ± 1.0
62. oxime	−11.5	−11.3	−8.4	−11.1	−11.2	−9.8	−10.6	−9.6	−10.4 ± 1.1	+1.0 ± 0.9
63. semicarbazone	−11.0	−10.0	−9.1	−10.5	−10.0	−9.6	−10.0	−10.5	−10.1 ± 0.6	+0.7 ± 0.6
64. SM-5887	−7.5	−6.8	−6.3	−6.9	−7.4	−7.2	−6.4	−7.3	−7.0 ± 0.5	−2.4 ± 0.9
65. WP-608-01	−14.3	−12.2	−12.4	−14.1	−12.7	−11.2	−14.0	−12.4	−12.9 ± 1.1	+3.5 ± 0.8

[a] The average ΔG is the average of the free energies for all eight sequences; the average $\Delta\Delta G_{dox}$ is the average of the difference between the free energies for all compounds versus all sequences and the average free energy for doxorubicin versus all eight sequences.

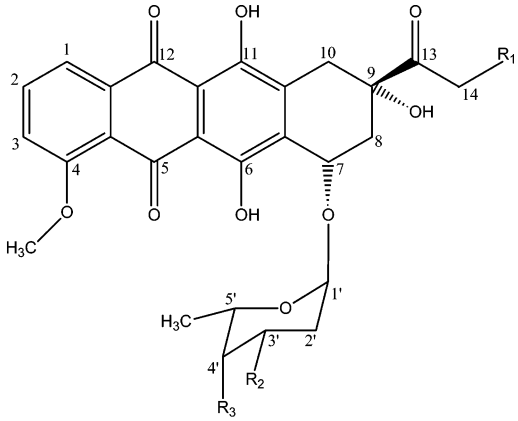
similar interaction strength to the single hydrogen bond of the C3′ hydroxyl in hydroxydoxorubicin, as we can see by comparing the calculated $\Delta\Delta G_{dox}$ between compounds **3** and **29**, the extra NH portion of the urea compounds sits on the other side of the T′ base carbonyl of the third base pair, forming an additional hydrogen

Scheme 2. Halogenated Doxorubicin Derivatives. Compounds Indicated with an Asterisk (*) Were Computationally Designed Based on the Referenced Compound, in Order to Test the Effects of Different Levels of Halogenation



9. Iododoxorubicin[31]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -I, R_6 = -H
10. Bromodoxorubicin[31*]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -Br, R_6 = -H
11. Chlorodoxorubicin[31*]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -Cl, R_6 = -H
12. Fluorodoxorubicin[31*]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -F, R_6 = -H
13. C^{14}-Iododoxorubicin[32*]: R_1 = -I, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -OH, R_6 = -H
14. C^{14}-Bromodoxorubicin[32]: R_1 = -Br, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -OH, R_6 = -H
15. C^{14}-Chlorodoxorubicin[32*]: R_1 = -Cl, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -OH, R_6 = -H
16. C^{14}-Fluorodoxorubicin[32*]: R_1 = -F, R_2 = -OCH_3, R_3 = -H, R_4 = -NH_3^+, R_5 = -OH, R_6 = -H
17. 2'-Iodo-4'-Epidaunorubicin[33]: R_1 = -H, R_2 = -OCH_3, R_3 = -I, R_4 = -NH_3^+, R_5 = -H, R_6 = -OH
18. 2'-Bromo-4'-Epidaunorubicin[33*]: R_1 = -H, R_2 = -OCH_3, R_3 = -Br, R_4 = -NH_3^+, R_5 = -H, R_6 = -OH
19. 2'-Chloro-4'-Epidaunorubicin[33*]: R_1 = -H, R_2 = -OCH_3, R_3 = -Cl, R_4 = -NH_3^+, R_5 = -H, R_6 = -OH
20. 2'-Fluoro-4'-Epidaunorubicin[33*]: R_1 = -H, R_2 = -OCH_3, R_3 = -F, R_4 = -NH_3^+, R_5 = -H, R_6 = -OH
21. Annamycin (C2'-Iodo)[34*]: R_1 = -OH, R_2 = -H, R_3 = -I, R_4 = -OH, R_5 = -H, R_6 = -OH
22. Annamycin (C2'-Bromo)[34*]: R_1 = -OH, R_2 = -H, R_3 = -Br, R_4 = -OH, R_5 = -H, R_6 = -OH
23. Annamycin (C2'-Chloro)[34*]: R_1 = -OH, R_2 = -H, R_3 = -Cl, R_4 = -OH, R_5 = -H, R_6 = -OH
24. Annamycin (C2'-Fluoro)[34*]: R_1 = -OH, R_2 = -H, R_3 = -F, R_4 = -OH, R_5 = -H, R_6 = -OH
25. N-(Iodoacetyl) Doxorubicin[35*]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -$NHCOCH_2I$, R_5 = -OH, R_6 = -H
26. N-(Bromoacetyl) Doxorubicin[35]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -$NHCOCH_2Br$, R_5 = -OH, R_6 = -H
27. N-(Chloroacetyl) Doxorubicin[35]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -$NHCOCH_2Cl$, R_5 = -OH, R_6 = -H
28. N-(Fluoroacetyl) Doxorubicin[35*]: R_1 = -OH, R_2 = -OCH_3, R_3 = -H, R_4 = -$NHCOCH_2F$, R_5 = -OH, R_6 = -H

Scheme 3. Doxorubicin Derivatives Containing Amide Moieties



29. N3'-Acetamide[36]: R_1 = -OH, R_2 = -$NHCOCH_3$, R_3 = -OH
30. N3'-Formyl Daunorubicin[37]: R_1 = -H, R_2 = -NHCHO, R_3 = -OH
31. N3'-Methyl Urea[36]: R_1 = -H, R_2 = -$NHCONHCH_3$, R_3 = -OH
32. N3'-Butyl Urea[36]: R_1 = -H, R_2 = -NHCONHBu, R_3 = -OH
33. N3'-Methylthio Urea[36]: R_1 = -H, R_2 = -$NHCSNHCH_3$, R_3 = -OH
34. N3'-Butylthio Urea[36]: R_1 = -H, R_2 = -NHCSNHBu, R_3 = -OH
35. N3'-Phenylthio Urea[36]: R_1 = -H, R_2 = -NHCSNHPh, R_3 = -OH
36. N3'-Dimethylglycine Daunorubicin[38]: R_1 = -H, R_2 = -$NHCOCH_2N(CH_3)_2$, R_3 = -OH
37. N3'-Glycine Daunorubicin[38]: R_1 = -H, R_2 = -$NHCOCH_2NH_3^+$, R_3 = -OH
38. Doxorubicin 3'-4'-Diacetate[39]: R_1 = -OH, R_2 = -OAc, R_3 = -OAc
39. Trifluoroacetamide[40]: R_1 = -H, R_2 = -$NHCOCF_3$, R_3 = -OH

bond. Additionally, there are more favorable hydrophobic contacts for the urea compounds, particularly those with a bulky butyl (**32**) or phenyl group (**35**) attached.

We calculated an increased $\Delta\Delta G_{dox}$ of +2.7 kcal mol^{-1} for one of the amide compounds, N3′-glycine daunorubicin (**37**). This compound has similar interactions to the acetamide (**29**) involving the amide, but also contains a protonated (or partially protonated) amine at the end which potentially interacts with the T base of the fourth and fifth base pairs, due to the amide group extending the reach of the amine. So this compound may involve a "Quintet" model for binding to doxorubicin, although that is somewhat beyond the scope of this study.

3. Effects of the Addition of Aromatic or Aliphatic Ring Substituents. Most of the compounds containing either an aromatic or aliphatic six-membered ring (compounds **40** to **53**) produced a diminished net energetic contribution. While most of the loss in free energy is due to the loss of hydrogen bonding of the protonated amine, the rings themselves do show evidence of somewhat favorable hydrophobic interactions. The lower $\Delta\Delta G_{dox}$ is evidence of this. Compound **44**, containing a piperidine ring double-bonded to C13, shows a somewhat noticeable increase in its net energetic contribution, mainly due to added hydrophobic contacts with the first base pair (above the intercalation site). Good evidence of actually increasing the binding by modification of the C13−C14 region is also seen in Zorubicin (**52**) and our computationally designed compound, Bromo-Zorubicin (**53**), which contains a phenyl group attached to a nitrogen-bearing chain that is double-bonded to C13. Our calculations show evidence of favorable hydrophobic contacts between the aromatic phenyl group and the methylene atoms of the phosphate backbone of the DNA.

Another interesting item to note is that it appears that compounds with a phenyl group at the C4′ position (**50** and **51**) appear to slightly increase the net energetic contribution to binding. On the other hand, compounds in which the phenyl group is attached to the C3′ position (**46** to **49**) actually indicate a decreased net energetic contribution to binding. While there are favorable hydrophobic contacts between the aromatic phenyl atoms and the methylene groups of the DNA phosphate backbone in both cases, the loss of binding in the C3′ substituents also appears largely due to the loss of the protonated amine functionality.

4. Other Substituent Effects. The $\Delta\Delta G_{dox}$ calculated for carminomycin (**58**) and idarubicin (**60**) also provide interesting information regarding the C4 methoxy. The $\Delta\Delta G_{dox}$ for both of these compounds is calculated to be approximately +1.2 kcal mol^{-1}, and this appears to be due to an adverse steric interaction between the methyl group of the methoxy functionality of doxorubicin and the methyl group of the T′ base of the second base pair (directly below the site of intercalation). This adverse interaction can be visualized by the magenta colored region in the HINT map calculation of doxorubicin vs the C|AAT sequence (Figure 2). Since Carminomycin contains a hydroxyl at C4, and Idarubicin contains only a hydrogen at C4, and both compounds have roughly the same calculated free energy, it appears that the adverse interaction is almost certainly connected to the methyl group of the methoxy functionality, as opposed to the ether oxygen atom.

Compound **65** contains an additional amine functionality at the C4′ position, in addition to the one at C3′. We calculate a net energetic increase of +3.5 kcal mol^{-1}, largely due to the binding of this group (although the C14 hydroxy is also absent, so this compound is also better compared to daunorubicin instead of doxorubicin). Although we are uncertain of the exact protonation state

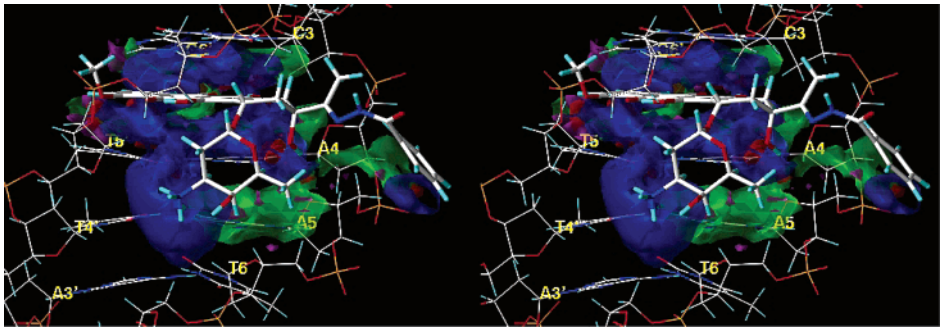


Figure 4. Stereodiagram of a HINT interaction map for the intercalation of zorubicin with the C|AAT base pair sequence of DNA, which displays, visually, the quality and magnitude of the various binding contacts involved in the interaction. See Figure 2 for contour information.

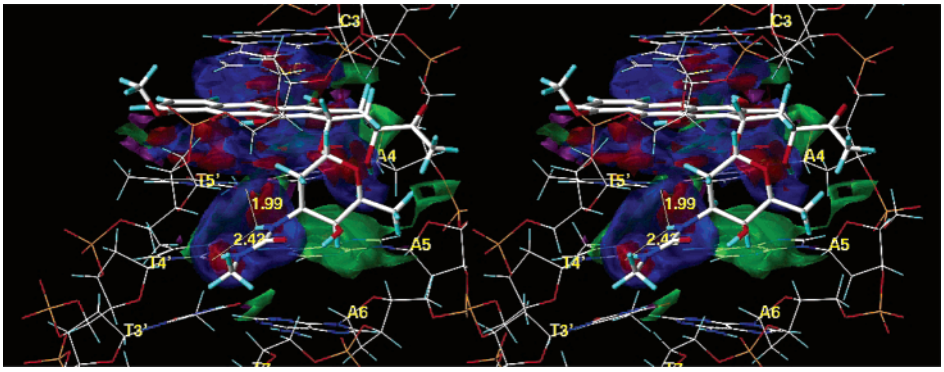


Figure 5. Stereodiagram of a HINT interaction map for the intercalation of N3′-methyl urea (**31**) with the C|AAA base pair sequence of DNA, which displays, visually, the quality and magnitude of the various binding contacts involved in the interaction. See Figure 2 for contour information.

of one or both of these amines, we should still nonetheless see favorable hydrogen bonding between this C4′ amine with one of the phosphate oxygen atoms of the DNA backbone, in addition to hydrogen bonding with the carbonyl oxygen atoms (as seen in doxorubicin; Figure 2).

While we do not have actual biological data to confirm the correlation between the calculated $\Delta\Delta G_{dox}$ and the actual net energetic contribution of the functional groups of our computational data set, we have demonstrated a relatively good correlation for the eight compounds in the "test set." It is very important to take note of the relatively large standard deviation calculated for the ΔG as well as the $\Delta\Delta G_{dox}$. We believe that the large standard deviations that we calculated are indicative of significant sequence selectivity for many of these compounds, so the above $\Delta\Delta G_{dox}$ calculations are likely biased by sequence selectivity.

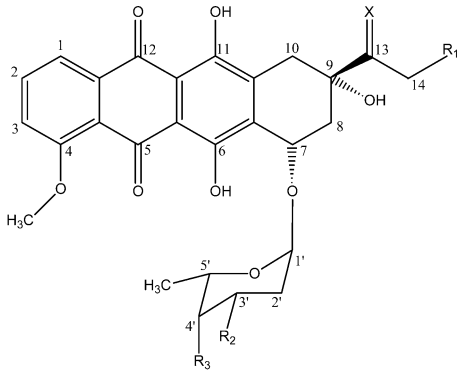
Selectivity Calculations

In a similar manner that we used to calculate the difference in average free energy of binding between the various analogues of doxorubicin, we were likewise able to calculate a difference in free energy of binding between each analogue of doxorubicin and each of the eight quartet sequences in our study. Table 3 shows the calculated $\Delta\Delta G_{sel}$ for all 65 doxorubicin analogues vs all eight quartet sequences. The sequence for which the binding is the strongest is indicated by two dashes (- -), and the $\Delta\Delta G_{sel}$ for all eight quartet sequences is listed. The lowest $\Delta\Delta G_{sel}$ for each analogue is in **bold**, which tells us the overall selectivity for the particular "bolded" sequence over the others. While we are mainly interested in the overall selectivity in this study, we can get a feeling for how selective compounds are by also looking at the $\Delta\Delta G_{sel}$ for the other sequences. If the $\Delta\Delta G_{sel}$ values are very similar, there may not be much selectivity for that particular compound; while if the $\Delta\Delta G_{sel}$ values are dramatically different, we will likely see good evidence of selectivity for that compound.

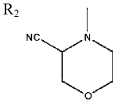
Out of the 65 compounds that we analyzed, two have a calculated $\Delta\Delta G_{sel}$ greater than 0.75 kcal mol^{-1} (**17** and **53**). Ten compounds have a calculated $\Delta\Delta G_{sel}$ between 0.50 and 0.74 kcal mol^{-1}, 18 compounds are selective between 0.25 and 0.49 kcal mol^{-1}, and 35 compounds are virtually nonselective with a calculated $\Delta\Delta G_{sel}$ of less than 0.25 kcal mol^{-1}. Overall, 27 out of 65 compounds are calculated to have some reasonable degree of selectivity for one of the eight sequences tested. Out of these 27 compounds, 10 are selective for the C|AAT sequence, three for C|AAG, two for C|GAT, seven for C|AAA, three for C|ATA, two for C|ACC, and none are selective for the C|GAG and C|AGC sequences.

1. Selectivity of Halogenated Derivatives. An interesting trend in selectivity is seen with the C4′ halogenated derivatives of doxorubicin, similar to the trend in average free energy discussed earlier. All of these analogues are calculated to be selective for the C|AAT sequence, except fluorodoxorubicin (**12**), which does not appear particularly selective at all. The calculated $\Delta\Delta G_{sel}$ increases as we move from fluoro to chloro to bromo, and then decreases with the iodo derivative (Figure 7). This increase in selectivity, as well as the increase in binding (discussed earlier), is most likely due to an increase in favorable hydrophobic interactions caused by the halogen interacting with the DNA back-

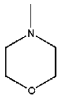
Scheme 4. Doxorubicin Derivatives Containing Various Aromatic and Nonaromatic Rings. Compounds with an Asterisk Were Computationally Designed Based on the Referenced Compound

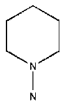


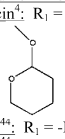
40. Carboxydaunorubicin γ-Lactam[36]: R_1 = -H, R_2 = -NHCOO- = R_3, X = O

41. Cyanomorpholino Doxorubicin[41]: R_1 = -OH, R_2  R_3 = -OH, X = O

42. Isothiocyanatobenzoyl Doxorubicin[42]: R_1 = -H, R_2 = -NHCOPh-*p*-N=C=S, R3 = -OH, X = O

43. MX-2[43]: R_1 = -H, R_3 = -OH, X = O, R_2 = 

44. N-piperidinoimine[36]: R_1 = -H, R_2 = $-NH_3^+$, R_3 = -OH, X = 

45. Pirarubicin[4]: R_1 = -OH, R_2 = $-NH_3^+$, X = O, R_3 = 

46. WPX-01[44]: R_1 = -H, R_2 = $-NHCH_2Ph$, R_3 = -OH, X = O

47. WPX-02[44]: R_1 = -H, R_2 = $-NHCH_2Ph$-*o*-NH_2, R_3 = -OH, X = O

48. WPX-03[44]: R_1 = -H, R_2 = $-NHCH_2Ph$-*m*-NH_2, R_3 = -OH, X = O

49. WPX-04[44]: R_1 = -H, R_2 = $-NHCH_2Ph$-*p*-NH_2, R_3 = -OH, X = O

50. WP-744[44]: R_1 = -OH, R_2 = $-NH_3^+$, R_3 = $-OCH_2Ph$, X = O

51. WP-744-01[44]: R_1 = -OH, R_2 = $-NH_3^+$, R_3 = $-OCH_2Ph$-*p*-NH_2, X = O

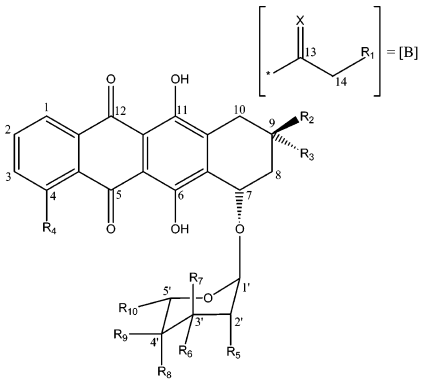
52. Zorubicin[4]: R_1 = -H, R_2 = $-NH_3^+$, R_3 = -OH, X = =NNHCOPh

53. Bromo-Zorubicin[4*]: R_1 = -H, R_2 = $-NH_3^+$, R_3 = -Br, X = =NNHCOPh

bone. It is also likely that the larger halogen is interfering with the solvent's ability to access and stabilize the C3′ amine cation, thereby allowing for an increased positive charge on the amine.

Halogenation at the C14 position of doxorubicin (**9** to **12**) does not show any particularly significant selectivity for one of the eight sequences tested. This is not terribly surprising due to the uncertainties with this region, as well as the added flexibility of this arm of doxorubicin. Halogenation at the C2′ position on the ring (**17** to **24**) appears to produce some notably selective compounds, particularly with the larger halogen atoms such as iodine and bromine. It appears that the large size of the iodine atom has a perturbing effect on the positioning of the ring, allowing good hydrogen bonding interactions with the T5′ carbonyl of one strand of the DNA, and the **C5** and **C6** carbonyls of the other strand of the DNA, which appears to give selectivity by about −0.79 kcal mol^{-1} for the C|ACC sequence. The bromine does not appear to have this same interaction, although it does appear to reduce the solvent accessibility for the amine cation, thereby increasing its interaction with the T4′ and T5′ carbonyls of the C|AAG sequence. The epimerization of the C4′-OH hydrogen bonds with the aromatic nitrogen and carbon atoms of the G6 residue, more than compensating for the reduced interaction with the carbonyl of the T6 residue (C|AAT). So we see a

Scheme 5. Miscellaneous Doxorubicin Derivatives



54. Analog VIII[32]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = $-OCH_3$, R_5 = -OH, R_6 = -H, R_7 = -OH, R_8 = -OH, R_9 = -H, R_{10} = $-CH_2OH$, X = O

55. Analog IX[32]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = $-OCH_3$, R_5 = $-NH_3^+$, R_6 = -H, R_7 = -OH, R_8 = -OH, R_9 = -H, R_{10} = $-CH_2OH$, X = O

56. Analog X[32]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = -H, R_5 = -OH, R_6 = -H, R_7 = -OH, R_8 = -OH, R_9 = -H, R_{10} = $-CH_2OH$, X = O

57. Analog XI[32]: R_5 = -H, R_3 = -H, R_4 = -H, R_5 = -H, R_6 = $-NH_3^+$, R_7 = -H, R_8 = -OH, R_9 = -H, R_{10} = $-CH_3$

58. Carminomycin[45]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = -OH, R_5 = -H, R_6 = $-NH_3^+$, R_7 = -H, R_8 = -OH, R_9 = -H, R_{10} = $-CH_3$, X = O

59. Epirubicin[4]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = $-OCH_3$, R_5 = -H, R_6 = $-NH_3^+$, R_7 = -H, R_8 = -H, R_9 = -OH, R_{10} = $-CH_3$, X = O

60. Idarubicin[4]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = -H, R_5 = -H, R_6 = $-NH_3^+$, R_7 = -H, R_8 = -OH, R_9 = -H, R_{10} = $-CH_3$, X = O

61. ME-2303[46]: R_1 = $-OCO(CH_2)_5COO^-$, R_2 = [B], R_3 = -OH, R_4 = $-OCH_3$, R_5 = -F, R_6 = -OH, R_7 = -H, R_8 = -OH, R_9 = -H, R_{10} = $-CH_3$, X = O

62. Oxime[36]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = $-OCH_3$, R_5 = -H, R_6 = $-NH_3^+$, R_7 = -H, R_8 = -OH, R_9 = -H, R_{10} = $-CH_3$, X = =N-OH

63. Semicarbazone[36]: R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = $-OCH_3$, R_5 = -H, R_6 = $-NH_3^+$, R_7 = -H, R_8 = -OH, R_9 = -H, R_{10} = $-CH_3$, X = =NNHCONH$_2$

64. SM-5887[47] R_1 = -H, R_2 = [B], R_3 = $-NH_3^+$, R_4 = -H, R_5 = -H, R_6 = -OH, R_7 = -H, R_8 = -OH, R_9 = -H, R_{10} = -H, X = O

65. WP-608-01[44] R_1 = -H, R_2 = [B], R_3 = -OH, R_4 = $-OCH_3$, R_5 = -H, R_6 = $-NH_3^+$, R_7 = -H, R_8 = $-NH_3^+$, R_9 = -H, R_{10} = $-CH_3$, X = O

somewhat moderate selectivity of −0.30 kcal mol^{-1} for the C|AAG sequence with compound **18**. Compounds **19** and **20** appear to have moderate selectivity for the C|AAA sequence, which appears to be due to some steric and hydrophobic effects in positioning the ring, as well as reduced solvent accessibility (compound **19**) or minor inductive effects (compound **20**), in contributing to a slightly more protonated amine cation.

2. Selectivity of Derivatives with Aromatic or Aliphatic Rings. Another compound of interest was the compound zorubicin (**52**), which we discussed earlier and which displayed a $\Delta\Delta G_{dox}$ of approximately +1.5 kcal mol^{-1}, due largely to the presence of an aromatic ring in the C14 region of the molecule. Not only is it a stronger binding compound than doxorubicin, but the lowest nonzero $\Delta\Delta G_{sel}$ was −0.51 kcal mol^{-1}, indicating a moderate degree of selectivity. This good selectivity of zorubicin, combined with the selectivity trend seen with the halogenated derivatives, leads us to explore the possibility of "computationally designing" an analogue of zorubicin that might have increased selectivity. So we added a bromine in place of the C4′ hydroxyl of zorubicin and calculated its free energy of binding and free energy differences. This hypothesis was confirmed computationally, as the lowest nonzero $\Delta\Delta G_{sel}$ for "bromo-zorubicin" is −0.96 kcal mol^{-1} (Figure 7), an increase from the $\Delta\Delta G_{sel}$ for zorubicin by 0.45 kcal mol^{-1}.

Compound **45**, pirarubicin, provides a good example of selectivity modification for a sequence other than the C|AAT sequence. Pirarubicin has a calculated selectivity of −0.74 kcal mol^{-1}, which makes it one of the most selective compounds in our computational data set. This compound shows good selectivity for the C|ACC sequence. The primary reason for its apparent selectivity for C|ACC is due to three strong hydrogen bonds with

Table 3. Calculated Free Energy Differences ($\Delta\Delta G_{sel}$ in kcal mol^{-1}) between Each DNA Quartet Sequence for 65 Doxorubicin Derivatives

compound	CAAT	CAAG	CGAG	CGAT	CAAA	CATA	CACC	CAGC
Test Set								
1. doxorubicin	--	−1.53	−2.55	−1.02	−1.94	−1.27	**−0.14**	−1.42
2. daunorubicin	--	−1.81	−2.11	−0.32	−1.27	**−0.06**	−0.09	−1.08
3. hydroxydoxorubicin	--	−0.23	**−0.10**	−0.19	−0.27	−0.71	−0.89	−0.54
4. 9-dehydroxydoxorubicin	--	−1.21	−2.53	**−0.12**	−1.07	−1.24	−0.48	**−0.12**
5. adriamycinone	−0.81	−0.34	−0.73	−0.64	--	−0.48	−0.29	**−0.16**
6. daunomycinone	**−0.35**	--	−1.26	−1.25	−0.54	−0.82	−0.88	−0.69
7. WP-608	−1.01	−0.90	−1.20	−0.88	**−0.71**	--	−1.76	−1.48
8. doxorubicin (β-anomer)	−0.69	--	−0.86	−0.92	−1.91	−0.67	**−0.59**	−0.75
Halogenated Derivatives								
9. iododoxorubicin	--	−0.30	−0.32	**−0.20**	−0.24	−2.52	−0.95	−1.82
10. bromodoxorubicin	--	−0.86	−0.97	**−0.67**	−0.68	−2.82	−1.33	−2.54
11. chlorodoxorubicin	--	−0.68	−0.77	**−0.43**	−0.58	−2.27	−1.14	−2.53
12. fluorodoxorubicin	**−0.08**	−0.10	--	−1.19	−0.87	−1.53	−0.51	−0.12
13. C14-iododoxorubicin	**−0.12**	−1.72	−2.49	--	−1.02	−1.70	−0.47	−0.80
14. C14-bromodoxorubicin	--	−1.72	−2.48	**−0.22**	−0.99	−1.65	−0.56	−0.81
15. C14-chlorodoxorubicin	--	−1.64	−2.57	**−0.30**	−0.97	−1.74	−0.44	−0.85
16. C14-fluorodoxorubicin	--	−1.70	−2.70	**−0.10**	−0.86	−1.52	−0.27	−0.74
17. 2′-iodo-4′-epidaunorubicin	**−0.79**	−1.06	−3.04	−2.79	−1.93	−2.29	--	−1.07
18. 2′-bromo-4′-epidaunorubicin	**−0.30**	--	−2.63	−1.94	−1.23	−1.73	−0.85	−0.76
19. 2′-Chloro-4′-epidaunorubicin	−0.54	**−0.46**	−2.84	−2.53	--	−1.88	−2.90	−2.47
20. 2′-fluoro-4′-epidaunorubicin	−0.29	**−0.38**	−2.92	−1.55	--	−1.99	−2.57	−2.15
21. annamycin (C2′-iodo)	−0.43	−0.29	−0.40	−0.90	**−0.13**	--	−1.08	−0.92
22. annamycin (C2′-bromo)	−0.77	−0.46	−0.70	−1.03	**−0.38**	--	−1.44	−1.26
23. annamycin (C2′-chloro)	−0.75	−0.38	−0.80	−1.26	−0.49	--	−1.57	−1.18
24. annamycin (C2′-fluoro)	−0.58	−0.49	−0.38	−0.90	--	**−0.01**	−1.05	−0.83
25. *N*-(iodoacetyl) doxorubicin	−0.38	**−0.11**	--	−0.48	−0.40	−0.92	−1.26	−0.50
26. *N*-(bromoacetyl) doxorubicin	--	**−0.11**	−0.57	−0.62	−0.95	−1.32	−1.96	−1.75
27. *N*-(chloroacetyl) doxorubicin	**−0.22**	--	−0.28	−0.61	−0.73	−1.02	−1.81	−1.75
28. *N*-(fluoroacetyl) doxorubicin	**−0.23**	−0.26	−0.54	−0.75	--	−1.19	−1.57	−1.57
Amide Derivatives								
29. N3′-acetamide	−0.15	−0.42	**−0.03**	−0.26	−0.04	--	−1.27	−0.63
30. N3′-formyl daunorubicin	**−0.10**	--	−0.64	−0.68	−0.26	−1.31	−1.60	−1.56
31. N3′-methyl urea	**−0.04**	−0.47	−1.75	−1.48	--	−1.32	−1.22	−2.15
32. N3′-butyl urea	--	−0.38	−1.83	−1.68	−0.71	−1.38	−2.03	−1.34
33. N3′-methylthio urea	−0.93	−0.33	−1.40	**−0.23**	−1.09	--	−0.81	−1.75
34. N3′-butylthiourea	--	−0.52	−0.79	−0.47	−0.61	−0.31	−0.79	**−0.09**
35. N3′-phenylthio urea	−1.91	−1.40	−1.26	--	−2.11	−1.11	**−0.65**	−1.19
36. N3′-dimethylglycine daunorubicin	−0.30	−0.71	−0.29	−0.47	--	−0.35	−1.37	−0.61
37. N3′-glycine daunorubicin	**−0.10**	−2.46	−2.38	--	−1.46	−3.18	−1.29	−1.19
38. doxorubicin 3′-4′-Diacetate	**−0.01**	−0.16	--	−0.28	−1.10	−1.17	−0.82	−0.78
39. trifluoroacetamide	−0.50	**−0.04**	−0.69	−0.63	--	−0.45	−1.83	−0.17
Derivatives with Additional Rings								
40. carboxidaunorubicin γ-lactam	−0.40	**−0.10**	−0.21	−0.19	−0.40	−0.69	−0.69	--
41. cyanomorpholino doxorubicin	−0.49	**−0.26**	−0.44	−0.28	--	−0.63	−1.95	−0.68
42. isothiocyanatobenzoyl doxorubicin	--	−0.44	−1.02	−0.15	−0.03	−1.17	−1.50	−1.52
43. MX-2	−0.40	**−0.08**	−0.60	--	−0.21	−0.43	−1.10	−1.17
44. *N*-piperidinoimine	**−0.53**	−1.98	−2.32	--	−1.61	−2.06	−0.96	−2.33
45. pirarubicin	−1.60	−1.60	−2.25	−1.70	−1.17	−2.41	--	**−0.74**
46. WPX-01	--	−0.65	−1.00	−0.73	**−0.14**	−1.40	−1.16	−1.19
47. WPX-02	**−0.12**	−0.70	−1.25	−0.74	--	−0.84	−1.58	−1.55
48. WPX-03	--	−0.80	−0.92	−0.38	**−0.03**	−1.86	−1.89	−1.53
49. WPX-04	**−0.03**	−0.37	−0.70	−0.22	--	−0.96	−0.88	−0.98
50. WP-744	−0.74	−1.32	−1.63	**−0.48**	--	−2.25	−1.44	−2.70
51. WP-744−01	−0.47	−0.54	−0.62	**−0.33**	--	−1.85	−1.41	−2.41
52. zorubicin	--	−1.66	−2.90	**−0.55**	−1.64	−0.78	−0.82	−2.53
53. bromo-zorubicin	--	−2.79	−2.43	−1.61	−1.43	**−0.96**	−2.58	−2.25
Miscellaneous Derivatives								
54. analogue VIII	−0.37	−0.38	−2.08	−1.27	--	−0.38	−0.84	−0.24
55. analogue IX	−0.45	−0.59	−1.96	−1.83	--	−0.81	−0.79	−0.51
56. analogue X	−0.05	−0.06	−0.28	−0.47	--	**−0.04**	−0.59	−0.27
57. analogue XI	--	−0.78	−3.70	−0.61	−1.73	−0.69	−0.93	−2.28
58. carminomycin	−0.38	−0.44	−1.82	−0.13	--	−1.78	−0.84	−2.01
59. epirubicin	**−0.04**	−0.26	−3.27	−2.20	--	−0.50	−1.84	−1.53
60. idarubicin	--	**−0.50**	−2.10	−0.81	−1.02	−1.63	−1.18	−1.65
61. ME-2303	**−0.14**	−0.19	−0.66	−0.49	--	−1.14	−0.94	−0.84
62. oxime	--	**−0.26**	−3.16	−0.46	−0.38	−1.77	−0.95	−1.96
63. semicarbazone	--	−1.01	−1.88	**−0.50**	−0.98	−1.36	−0.97	**−0.50**
64. SM-5887	--	−0.64	−1.20	−0.59	**−0.05**	−0.32	−1.11	−0.16
65. WP-608-01	--	−2.12	−1.92	**−0.23**	−1.57	−3.13	−0.28	−1.91

[a] The $\Delta\Delta G_{sel}$ between each additional sequence and the most favorable is indicated for each sequence. The sequence with the lowest (most favorable) free energy of interaction is indicated by the dashed lines. The lowest $\Delta\Delta G_{sel}$ for each compound is in **bold**, indicating the overall calculated selectivity for the sequence with the lowest free energy.

the C3′ amine and three carbonyl atoms of nearby base pairs (Figure 8A). There is a good hydrogen bond with the carbonyl of the thymine of the second base pair, which also is seen in C|AAT and other sequences.

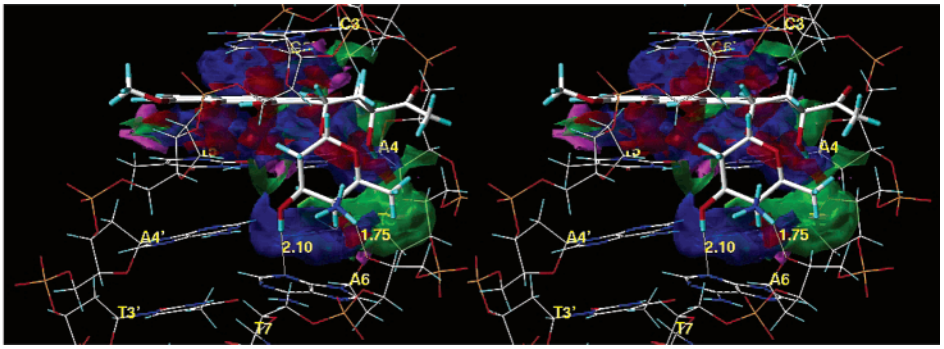


Figure 6. Stereodiagram of a HINT interaction map for the intercalation of WP-608 (**7**) with the C|ATA base pair sequence of DNA, which displays, visually, the quality and magnitude of the various binding contacts involved in the interaction. See Figure 2 for contour information.

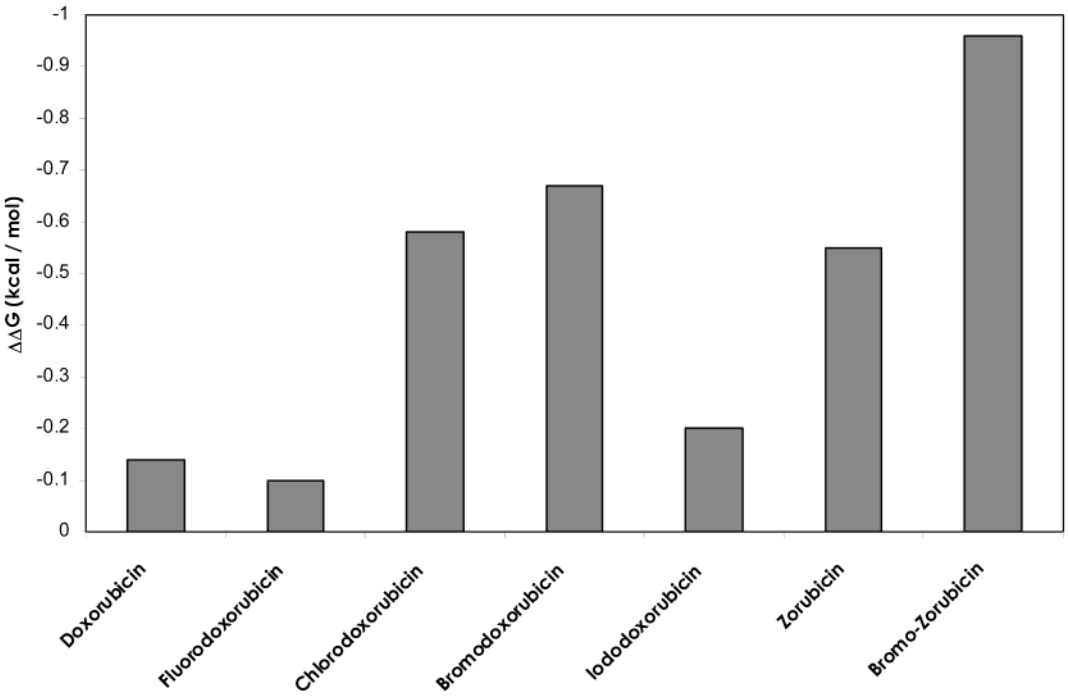


Figure 7. Calculated $\Delta\Delta G_{sel}$ (kcal mol^{-1}) for the second strongest sequence of halogenated derivatives of doxorubicin. Doxorubicin, zorubicin, and "bromo-zorubicin" are also included for comparison.

However, since the third and fourth bases are both cytosines in C|ACC, there are two carbonyl atoms that are positioned approximately 2.4 Å from the C3′ amine. This causes a slight rotation in the positioning of the daunosamine sugar (compared to the sugar positioning in doxorubicin, Figure 2) leading to an adverse interaction caused by the proximity of the C3′ amine to the anilinic amine of the guanine base of the third base pair. While this interaction prevents the overall calculated free energy of binding from significantly exceeding the calculated free energy of binding of doxorubicin, the proximity to the aforementioned carbonyl atoms contributes to the overall favorable interaction of the complex. If we compare the complex of C|ACC vs pirarubicin to the complex of C|AAT vs pirarubicin (Figure 8B), we can see the slight rotation in the ring more clearly. We can also see that, due to the difference in rotation, the second six-membered ring of pirarubicin prevents the C3′ amine from getting in close proximity to the carbonyl oxygen of the thymine of the fourth base pair, which results in a reduction in the overall calculated free energy of this complex.

We even see a significant drop in free energy between the most favorable (C|ACC) and the second most favorable (C|AGC) complexes for pirarubicin. The only difference between these sequences is the reversal of the CG base pair in the third slot, which places the anilinic amine of the guanine base where there was a favorable, hydrogen bond accepting carbonyl oxygen in C|ACC. However, this also places a carbonyl oxygen in the third slot, where the adverse anilinic amine was in C|AGC. The six-membered ring of pirarubicin also appears to prevent any interaction with the fourth base pair (carbonyl on cytosine), similar to the loss of the carbonyl interaction in thymine of the C|AAT sequence.

We also see a difference in selectivity with two compounds that contain a phenyl ring at the C4′ position (**50** and **51**). The phenyl ring is connected with an additional methyl linker between the ring and the C4′ oxygen, which places the ring itself a bit further away than the aliphatic ring of pirarubicin. Consequently, we do not see the same rotation in the daunosamine sugar that we saw with pirarubicin. The selectivities are also different as well; **50** is −0.48 kcal mol^{-1} selective, and

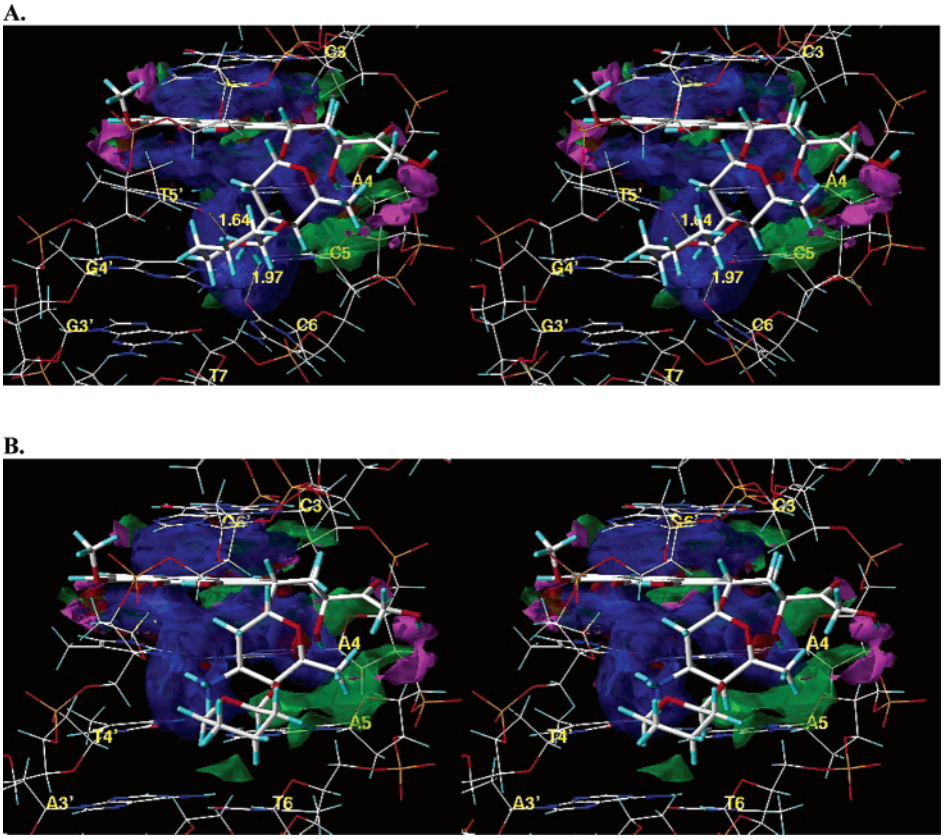


Figure 8. Stereodiagram of a HINT interaction map for the intercalation of pirarubicin (**45**) with the C|ACC (A) and C|AAT (B) base pair sequences of DNA, which displays, visually, the quality and magnitude of the various binding contacts involved in the interaction. See Figure 2 for contour information.

51 is −0.33 kcal mol^{-1} selective, both for C|AAA. So we do see the inhibition of the fourth base pair interaction with the C|AAT sequence in both of these compounds, but we do not see the rotation and selectivity for C|ACC that we saw for pirarubicin. Also of interest, compounds in which the phenyl ring was placed at the C3′ position instead of the C4′ position (**46** to **49**) did not produce any significant selectivity.

3. Selectivity by Modifications to the C13/C14 Region. We discussed earlier that compounds with relatively minor modifications to the C13/C14 region, such as daunorubicin and the C14 halogenated derivatives, do not appear to be particularly selective one way or another. Yet, a compound with a relatively large substituent, such as zorubicin, that includes a bulky phenyl group in this region, displays considerably more selectivity. So what about other compounds with modifications to this region? Is there any apparent selectivity that can be gained by modifying here?

We noted a similar increase in binding and selectivity in compounds **44** and **63**. Analogue **44**, *N*-piperidinoimine, contains a piperidine ring double-bonded to **C13** and shows decent selectivity of −0.53 kcal mol^{-1} for the C|GAT sequence. Analogue **63**, Semicarbazone, contains an amine-bearing moiety that interacts favorably with base pairs above the intercalation site, and demonstrates selectivity of −0.50 kcal mol^{-1} for the C|AAT sequence. Also, it is interesting to note that analogues **57** and **60** demonstrated selectivity for the C|AAT sequence as well. Analogue **57** lacks both the C4 methoxy as well as the entire C13/C14 region and analogue **60** (idarubicin) lacks the C4 methoxy. Both compounds have calculated selectivities of −0.61 and −0.50 kcal mol^{-1}, respectively. Therefore, we can see that, while the intercalation interaction itself does not contribute significantly to selectivity, adding or taking away functional groups connected to the intercalating chromophore region can affect selectivity by modifying the positioning of the daunosamine sugar.

4. Other Examples of Selectivity Modulation. Another excellent example of a shift in selectivity from the common C|AAT sequence to one of the other eight sequences can be seen in our original "test set." Compound **7**, WP-608, contains two minor differences from doxorubicin; the positions of the C3′ amine and the C4′ hydroxyl have been swapped, and like daunorubicin, the C14-OH is missing. Since doxorubicin and daunorubicin have virtually identical selectivities in our data, we can therefore assume that the difference in selectivity for WP-608 is due to the C3′/C4′ swap. In the interaction of WP-608 versus the C|ATA sequence, the C4′ amine is in position to interact with the ether oxygen of the ribose on the 4th (adenine) base pair (Figure 6). While this particular interaction is not sequence specific, since the amine interacts with the same atom in all eight structures, there is a sequence-specific interaction of the C3′-OH. In C|ATA, the oxygen of the C3′-OH is positioned approximately 4.18 Å from the aromatic nitrogen atom of the adenine base (opposite the thymine in the 3rd base pair), versus a distance of 3.05 Å from the aromatic nitrogen of the adenine of the 4th base pair. In the C|AAT sequence, there are thymines in these two positions, containing carbonyl oxygen atoms that would ordinarily be able to serve as hydrogen bond donors for

the C3′-OH. However, the distance between the C3′ oxygen and each of these carbonyl oxygen is 4.18 Å to the carbonyl of the 3rd base pair thymine, and 4.92 Å to the carbonyl oxygen of the 4th base pair thymine. So very little interaction is seen here, and WP-608 turns out to be -0.71 kcal mol^{-1} selective for the C|ATA sequence.

It should also be noted that compound **65**, WP-60801, which contains an amine at both the C3′ and C4′ position on the ring, does not appear to be particularly selective for any sequence ($\Delta\Delta G_{sel} = -0.23$ kcal mol^{-1}). So it appears that the, "brute force approach," of adding two protonated (or partially protonated) amines to the compound, while significantly contributing to free energy by $+3.5$ kcal mol^{-1}, does not do much for selectivity. However, changing the position of the amine can have an effect on selectivity.

Summary

This study of 65 doxorubicin analogues is, to our knowledge, one of the most comprehensive structural studies of such compounds to date, and the vast amount of data generated from this project suggest good potential for the future design of sequence selective DNA binding agents. By correlating experimental binding data and information from a crystal structure with computational data, it is possible to obtain useful data for making predictions and designing new drugs. Additionally, we have provided an example demonstrating that the HINT model can be utilized for predicting the free energy of binding of nucleic acid complexes, supplementing predictions on protein complexes from previous studies.[25,26]

We have established a reasonably good correlation with experimental binding data, with a few explainable discrepancies. The lack of substantive experimental binding data on many of these complexes remains, however, a serious handicap for unequivocal validation of computational methods, and the rational design of new analogues. These results do indicate the reproducibility and accuracy of the computational studies; however, the major source of uncertainty arises from conformational uncertainty in model building, while the lack of understanding of chemical effects and our ability to model them, e.g. the protonation of the C3′ amine, is a vexing cause of discrepancy between the calculated and experimental studies.

Nonetheless, our calculations on the functional group contributions of the various substitutions of doxorubicin analogues, as well as selectivity calculations, can be a guide for developing future sequence selective compounds. While a $\Delta\Delta G_{sel}$ of -1.0 kcal mol^{-1} for the most selective compounds only corresponds to a 5.40-fold increase in affinity (when converted to K_i), the computational data provided here should nonetheless be valuable for designing molecular tools to further drug design.

Materials and Methods

Molecular Models and Energy Minimization. Molecular models were created and minimized using the SYBYL v. 6.7 molecular modeling package (Tripos, Inc., St. Louis, MO). The starting point was the complex structure of doxorubicin intercalated into the DNA octamer sequence d(CGC|AATCG/CGATT|GCG). The sequence that is underlined is the DNA quartet that is hereafter referred to, with the doxorubicin analogue intercalated between the first two base pairs in the sequence (indicated by the "|"). Double-helical B-DNA was constructed using the SYBYL Biopolymer module. Doxorubicin, in the same conformation as found in the crystal structure (pdb accession coordinates: 1D12),[19] was placed in the intercalation site. The doxorubicin molecule was placed into the DNA according to probable binding sites for an amine cation and a hydroxyl oxygen as calculated by the program GRID (GRID v. 17, Molecular Discovery, Ltd., London, UK), as previously described.[29] The ionization state of the compounds was initially the same as described in previous modeling studies (i.e. the ammonium and all hydroxyls were protonated).[13,21,27,28]

The SYBYL Unity module was employed to perform a 3-D structural search of the 213 628 compounds in the NCI Database for analogues of doxorubicin. The template for this search was the anthracycline ring system of the doxorubicin molecule, with the sugar removed, the C4 methoxy removed, the C9 hydroxy removed, and the C13−C14 side chain removed. Approximately 230 compounds were initially identified from the Unity database search, and this was then reduced to 65 compounds (Schemes 1−5) upon more detailed analysis.

The structures of these 65 compounds were modified from the model of bound doxorubicin, using the SYBYL Builder module. The DNA oligomer sequences were then modified using the SYBYL Biopolymer, "Mutate Monomers," function. Each doxorubicin analogue obtained from the Unity database search was modeled in eight DNA quartet sequences (C|AAT, C|AAG, C|GAG, C|GAT, C|AAA, C|ATA, C|ACC, and C|AGC). These sequences were selected based on our previous modeling study of doxorubicin intercalated into 64 DNA quartet sequences, as well as a more recent calibration study of six doxorubicin analogues intercalated into 32 DNA quartet sequences and compared to experimental binding data.[21,28]

Each structure was solvated using the droplet protocol in SYBYL with a single layer of water molecules to simulate the hydration layers that surround duplex DNA.[30] We have also found, from previous experiments, that this water monolayer is necessary to prevent helix unwinding during the complex minimization process without adding significant complexity to the system.[21,28]

All structures were energy minimized using the Tripos Force Field and Gasteiger-Hückel charges with 300 cycles of steepest descent energy minimization followed by conjugate gradient energy minimization until the energy difference was less than 0.05 kcal mol^{-1}, as reported previously.[21,28] We earlier showed that this protocol builds model structures with good agreement to available crystal structures.[21,28] To ensure the integrity of our model, we superimposed our final structure of doxorubicin intercalated with the C|GAT quartet sequence onto the crystal structure and determined the RMS deviation to be 1.14 Å.[20] We should also highlight, as it is one of our primary assumptions, that although the models thus generated are very likely not at the global energy minima, as they were all built with consistent protocols the relative differences between them should be significant.

Hydropathic Interaction Analysis. The interactions of each complex were analyzed using the HINT program v. 2.35S (Tripos, Inc., St. Louis, MO), as described previously.[20,21,23−25,28] The HINT model describes specific interactions between two molecules, DNA and the anthracycline in our case, as a double sum over the atoms within each component (eq 1).

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

The variable a is the hydrophobic atom constant, S is the solvent accessible surface area, T is a descriptor function, and R and r are distance functions.[21,23,24−28] The binding score, b_{ij}, describes the specific interaction between two atoms in the complex, i and j, and B describes the total interaction between both molecules.[21,23,24−28] The application of this equation has previously been described.[21,23,24−28]

ΔG and ΔΔG Calculation Details. Equation 1 was used to calculate an interaction score for each structurally minimized anthracycline/DNA complex. We have found through previous experiments of numerous protein−ligand complexes that approximately −515 HINT score units is approximately equal to 1 kcal mol^{-1}.[25,26] To extend this correlation to nucleic acid complexes, we can take the average calculated HINT score for each compound in our "test set" and divide this by the experimental Δ*G* as determined by Chaires.[22] Taking the average of these results for all eight compounds in our "test set", we get a correlation of approximately −590 HINT score units equal to 1 kcal mol^{-1}, with a standard deviation of 112 HINT score units. Since these two approximations of HINT versus free energy only differ by 75 HINT score units, which is well within our standard deviation, we used the original approximation of −515 HINT score units equal to 1 kcal mol^{-1} for this study and can expect accuracy to within 0.2 kcal mol^{-1}.

Each HINT score was thus converted to an approximate free energy score in kcal mol^{-1} by dividing the total HINT score for each complex by −515. This empirical score factor has been determined from the slope of Δ*G* vs HINT score plots for a series of protein−ligand complexes, so is more appropriate for estimating ΔΔ*G* than Δ*G*.[25,26]

Furthermore, there are numerous energetic contributions to a biomolecular event as complex as DNA/doxorubicin intercalation, including terms to represent the deformation of the DNA, loss of entropy for the new complex and solvent partitioning for the drug from water to the intercalation site, as well as terms specific to the drug−DNA interaction.[21] Thus, the calculated Δ*G* scores must be regarded with a certain degree of caution. However, the ΔΔ*G*, that is, the difference in free energy of binding between different complexes, should allow for a reasonably accurate model that describes relative differences in binding for the system, especially for complexes of similar doxorubicin analogues intercalated in eight DNA oligomers of identical length that have been modeled and optimized using an identical procedure. And the ΔΔ*G* should be more than sufficient for an analysis into both the functional group contributions and the selectivity of doxorubicin analogues.

The $\Delta\Delta G_{dox}$, or the difference in calculated Δ*G* between each compound (ΔG_{seq}) and the calculated Δ*G* of doxorubicin (ΔG_{dox}), was calculated for each sequence, using eq 2. The average of $\Delta\Delta G_{dox}$ for all 65 compounds was then calculated, with standard deviations.

$$\Delta\Delta G_{dox} = \Delta G_{dox} - \Delta G_{seq} \quad (2)$$

To calculate selectivity for each compound versus a particular sequence, we first determined the sequence that produced the lowest (most favorable) Δ*G* for each compound (indicated by ΔG_{min}). Next, we subtracted the Δ*G* for each of the other eight sequences (ΔG_{seq}) from this ΔG_{min} value, giving a ΔΔ*G* of selectivity for each sequence ($\Delta\Delta G_{sel}$), using eq 3. The quartet sequence that any given compound is selective for is indicated by a $\Delta\Delta G_{sel} = 0.00$ kcal mol^{-1}, and the relative degree of selectivity is indicated by the smallest $\Delta\Delta G_{sel}$ of each of the seven remaining sequences.

$$\Delta\Delta G_{sel} = \Delta G_{min} - \Delta G_{seq} \quad (3)$$

Acknowledgment. This work was supported in part by the Massey Cancer Center of V.C.U. through an American Cancer Society Institutional Research Grant. We would like to acknowledge the helpful comments and discussions from Drs. J. P. Rife and J. N. Scarsdale in preparing this paper. We would also like to thank Dr. J. B. Chaires for measuring and publishing the valuable biophysical data that allowed us to calibrate our models. The SYBYL software has been made available through a University Software Grant from Tripos, Inc. The HINT program was developed in the laboratory of G.E.K. and may be obtained from Tripos, Inc., as a SYBYL program module.

Supporting Information Available: Table 4 includes the HINT score data for each complex by base pair contribution. Table 5 includes the HINT score data for each complex by interaction type (hydrogen bonding, acid/base, hydrophobic). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids. *Nature* **1953**, *171*, 737−738.
(2) Graves, D. E.; Velea, L. M. Intercalative Binding of Small Molecules to Nucleic Acids. *Curr. Org. Chem.* **2000**, *4 (9)*, 915−929.
(3) The International Human Genome Mapping Consortium. A Physical Map of the Human Genome. *Nature* **2001**, *401*, 860−921.
(4) Weiss, Raymond B. The Anthracyclines: Will We Ever Find a Better Doxorubicin? **1992**, *Semin. Oncol. 19* (6), 670−686.
(5) Pigram, W. J.; Fuller, W.; Hamilton, L. D. Stereochemistry of Intercalation: Interaction of Daunomycin with DNA. *Nat. N. Biol.* **1972**, *235*, 17−19.
(6) Momparler, R. L.; Karon, M.; Siegel, S. E.; Avila, F. Effect of Adriamycin on DNA, RNA, and Protein Synthesis in Cell-Free Systems and Intact Cells. *Cancer Res.* **1976**, *36*, 2891−2895.
(7) Tewey, K. M.; Rowe, T. C.; Yang, L.; Halligan, B. D.; Lui, L. F. Adriamycin-Induced DNA Damage Mediated by Mammalian DNA Topoisomerase II. *Science* **1984**, *226*, 466−468.
(8) Schneider, E.; Hsiang, Y.; Lui, L. F. DNA Topoisomerases as Anticancer Drug Targets. *Adv. Pharmacol.* **1990**, *21*, 149−183.
(9) Sander, M.; Tsieh, T.-S. Double strand DNA cleavage by type II DNA Topoisomerase from *Drosophila melanogaster*. *J. Biol. Chem.* **1983**, *258*, 8421−8428.
(10) Ogretmen, B.; Safa, A. R. Identification and Characterization of the MDR1 Promoter-Enhancing Factor 1 (MEF1) in the Multidrug Resistant HL60/VCR Human Acute Myeloid Leukemia Cell Line. *Biochemistry* **2000**, *39*, 194−204.
(11) Wang, A. H. J.; Ughetto, G.; Quigley, G. J.; Rich, A. Interactions Between an Anthracycline Antibiotic and DNA: Molecular Structure of Daunomycin Complexed to d(CpGpTpApCpG) at 1.2-Å Resolution. *Biochemistry* **1987**, *26*, 1152−1163.
(12) Moore, M. H.; Junter, W. N.; d'Estaintot, B. L.; Kennard, O. DNA-Drug Interactions. The Crystal Structure of d(CGATCG) Complexed With Daunomycin. *J. Mol. Biol.* **1989**, *206*, 693−705.
(13) Chen, K. X.; Gresh, N.; Pullman, B. A Theoretical Investigation on the Sequence Selective Binding of Daunomycin to Double-Stranded Polynucleotides. *J. Biomol. Struct. Dyn.* **1985**, *3*, 445−466.
(14) Chaires, J. B.; Fox, K. R.; Herrerra, J. E.; Britt, M.; Waring, M. J. Site and Sequence Specificity of the Daunomycin-DNA Interaction. *Biochemistry* **1987**, *26*, 8227−8236.
(15) Skorobogaty, A.; White, R. J.; Phillips, D. R.; Reiss, J. A. The 5′-CA DNA-Sequence Preference of Daunomycin. *FEBS Lett.* **1988**, *227* (2), 103−106.
(16) Cullinane, C. R.; Phillips, D. R. Induction of Stable Transcriptional Blockage Sites By Adriamycin: GpC Specificity of Apparent Adriamycin-DNA Adducts and Dependence On Iron(III) Ions. *Biochemistry* **1990**, *29*, 5638−5646.
(17) Graves, D. E.; Krugh, T. R. Adriamycin and Daunorubicin Bind in a Cooperative Manner to Deoxyribonucleic Acid. *Biochemistry* **1983**, *22 (16)*, 3941−3947.
(18) Trist, H.; Phillips, D. R. In Vitro Transcription Analysis of the Role of Flanking Sequence on the DNA Sequence Specificity of Adriamycin. *Nucleic Acids Res.* **1989**, *17*, 3673−3688.
(19) Chaires, J. B.; Herrerra, J. E. Preferential Binding of Daunomycin to 5′-(A/T)CG and 5′-(T/A)GC Sequences Revealed by Footprinting Titration Experiments. *Biochemistry* **1990**, *29*, 6145−6153.
(20) Frederick, C. A.; Williams, L. D.; Ughetto, G.; Van der Marel, G. A.; Van Boom, J. H.; Rich, A.; Wang, A. H. J. Structural Comparison of Anticancer Drug−DNA Complexes: Adriamycin and Daunomycin. *Biochemistry* **1990**, *29* (10), 2538−2549.
(21) Kellogg, G. E.; Scarsdale, J. N.; Fornari, F. A. Identification and Hydropathic Characterization of Structural Features Affecting Sequence Specificity for Doxorubicin Intercalation into DNA Double-Stranded Polynucleotides. *Nucleic Acids Res.* **1998**, *26* (20), 4721−4732.
(22) Chaires, J. B.; Satyanarayana, S.; Suh, D.; Fokt, I.; Przewloka, T.; Priebe, W. Parsing the Free Energy of Anthracycline Antibiotic Binding to DNA. *Biochemistry* **1996**, *35* (7), 2047−2053.
(23) Kellogg, G. E.; Abraham, D. J. Hydrophobicity: Is Log $P_{o/w}$ More Than The Sum Of Its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651−661.

(24) Hansch, C.; Leo, A. J. Extension of the Fragment Method to Calculate Amino Acid Zwitterion and Side Chain Partition Coefficients. *Proteins: Struct., Funct., Genet.* **1987**, *2*, 130−152.
(25) Burnett, J. C.; Kellogg, G. E.; Abraham, D. J. Computational Methodology for Estimating Changes in Free Energies of Biomolecular Association upon Mutation. The Importance of Bound Water in Dimer-Tetramer Assembly for *β*-37 Mutant Hemoglobins. *Biochemistry* **2000**, *39* (7), 1622−1633.
(26) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Simple, Intuitive Calculations of Free Energy of Binding for Protein−Ligand Complexes. 1. Models without Explicit Constrained Water. *J. Med. Chem.* **2002**, *45* (12), 2469−2483.
(27) Cashman, D. J.; Rife, J. P.; Kellogg, G. E. Which Aminoglycoside Ring Is Most Important For Binding? A Hydropathic Analysis of Gentamicin, Paromomycin, and Analogues. *Bioorg., Med. Chem. Lett.* **2001**, *11*, 119−122.
(28) Cashman, D. J.; Scarsdale, J. N.; Kellogg, G. E. Hydropathic Analysis of the Free Energy Differences in Anthracycline Antibiotic Binding to DNA. *Nucleic Acids Res.* **2003**, *31* (15), 4410−4416.
(29) Goodford, P. J. A Computational Procedure For Determining Energetically-Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(30) Qu, X.; Chaires, J. B. Hydration Changes for DNA Intercalation Reactions. *J. Am. Chem. Soc.* **2001**, *123* (1), 1−7.
(31) Berger, I.; Su, L.; Spitzner, J. R.; Kang, C.; Burke, T. G.; Rich, A. Molecular Structure of the Halogenated Anti-Cancer Drug Iododoxorubicin Complexed with d(TGTACA) and d(CGATCG). *Nucleic Acids Res.* **1995**, *23*, 4488−4494.
(32) Arcamone, F.; Penco, S.; Vigevani, A. Adriamycin (NSC-123127): New Chemical Developments and Analogues. *Cancer Chemotherapy Rep., Pt. 3*, **1975**, *6* (2), 123−129.
(33) Gao, Y.-G.; Priebe, W.; Wang, A. H.-J. Substitutions at C2′ of Daunosamine in the Anticancer Drug Daunorubicin Alter its DNA-Binding Sequence Specificity. *Eur. J. Biochem.* **1996**, *240*, 331−335.
(34) Dziewiszek, K.; Priebe, W. U.S. Pat. No. 5, 977, 327, 1999.
(35) Farquhar, D.; Newman, R. A.; Zuckerman, J. E.; Anderson, B. S. Doxorubicin Analogues Incorporating Chemically Reactive Substituents. *J. Med. Chem.* **1991**, *34*, 561−564.
(36) Yamamoto, K.; Acton, E. M.; Henry, D. W. Antitumor Activity of Some Derivatives of Daunorubicin at the Amino and Methyl Ketone Functions. *J. Med. Chem.* **1972**, *15* (8), 872−875.
(37) Komiyama, T.; Matsuzawa, Y.; Oki, T.; Inui, T.; Takahashi, Y.; Naganawa, H.; Takeuchi, T.; Umezawa, H. Baumycins, New Antitumor Antibiotics related to Daunomycin. *J. Antibiot.* **1977**, *30* (7), 619−621.
(38) Gabbay, E. J.; Grier, D.; Fingerle, R. E.; Reimer, R.; Levy, R.; Pearce, S. W.; Wilson, W. D. Interaction Specificity of the Anthracyclines. *Biochemistry* **1976**, *15* (10), 2062−2070.
(39) Horton, D.; Priebe, W.; Turner, W. R. Synthesis and Antitumor Activity of 7-O-(3,4-Di-O-Acetyl-2,6-Dideoxy-α-L-*lyxo*-Hexopyranosyl)Adriamycinone. *Carbohydrate Res.* **1981**, *94*, 11−25.
(40) Smith, T. H.; Fujiwara, A. N.; Henry, D. W. Synthesis of daunorubicin analogues with novel 9-acyl substituents. *J. Med. Chem.* **1979**, *22* (1), 40−44.
(41) Acton, E. M.; Tong, G. L.; Mosher, C. W.; Wolgemuth, R. L. Intensely Potent Morpholinyl Anthracyclines. *J. Med. Chem.* **1984**, *27*, 638−645.
(42) Rosik, L.; Sweet, F. Electrophilic Analogues of Daunorubicin and Doxorubicin. *Bioconjugate Chem.* **1990**, *1*, 251−256.
(43) Watanabe, M.; Komeshima, N.; Nakajima, S.; Tsuruo, T. MX2, A Morpholino Anthracycline, As A New Antitumor Agent Against Drug-Sensitive and Multidrug-Resistant Human and Murine Tumor Cells. *Cancer Res.* **1988**, *48*, 6653−6657.
(44) Rudnicki, W. R.; Kurzepa, M.; Szczepanik, T.; Priebe, W.; Lesyng, B. A Simple Model for Predicting the Free Energy of Binding Between Anthracycline Antibiotics and DNA. *Acta Biochim. Pol.* **2000**, *47* (1), 1−9.
(45) Horton, D.; Turner, W. R. Adriamycin Analogues Hydroxylated at C-3′: Synthesis and Antitumor Activity. *Carbohydrate Res.* **1979**, *77*, C8 − C11.
(46) Tsuruo, T.; Yusa, K.; Sudo, Y.; Takamori, R.; Sugimoto, Y. A Fluorine-Containing Anthracycline (ME−2303) As A New Antitumor Agent Against Murine and Human Tumors and Their Multidrug-Resistant Sublines. *Cancer Res.* **1989**, *49*, 5537−5542.
(47) Inouse, K.; Ogawa, M.; Horikoshi, N.; Mukaiyama, T.; Itoh, Y.; Imajoh, K.; Ozeki, H.; Nagamine, D.; Shinagawa, K. Phase I and Pharmacokinetic Study of SM-5887, A New Anthracycline Derivative. *Invest. N. Drugs* **1989**, *7*, 213−218.

JM030529H