JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

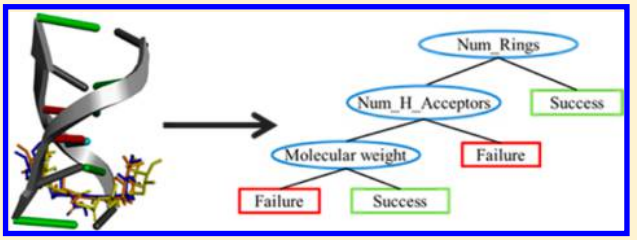
Docking Studies on DNA Intercalators

Yocheved Gilad* and Hanoch Senderowitz

Department of Chemistry, Bar Ilan University, Ramat-Gan 52900, Israel

Supporting Information

ABSTRACT: DNA is an important target for the treatment of multiple pathologies, most notably cancer. In particular, DNA intercalators have often been used as anticancer drugs. However, despite their relevance to drug discovery, only a few systematic computational studies were performed on DNA-intercalator complexes. In this work we have analyzed ligand binding sites preferences in 63 high resolution DNA-intercalator complexes available in the PDB and found that ligands bind preferentially between **G** and **C** and between the **C** and **A** base pairs (70% and 11%, respectively). Next, we examined the ability of AUTODOCK to accurately dock ligands into preformed intercalation sites. Following the optimization of the docking protocol, AUTODOCK was able to generate conformations with RMSD values <2.00 Å with respect to crystal structures in ∼80% of the cases while focusing on the preformed binding site (small grid box) or on the entire DNA structure (large grid box). In addition, a top ranked conformation with an RMSD < 2.00 Å was identified in 75% and 60% of the cases using small and large docking boxes, respectively. Moreover, under the large docking box setting AUTODOCK was able to successfully distinguish between the intercalation site and the minor groove site. However, in all cases the crystal structures and poses tightly clustered around it had a lower score than the best scoring poses suggesting a potential scoring problem with AUTODOCK. A close examination of all cases where the top ranked pose had an RMSD value >2.00 Å suggests that AUTODOCK may overemphasize the hydrogen bonding term. A decision tree was built to identify ligands which are likely to be accurately docked based on their characteristics. This analysis revealed that AUTODOCK performs best for intercalators characterized by a large number of aromatic rings, low flexibility, high molecular weight, and a small number of hydrogen bond acceptors. Finally, for canonical B-DNA structures (where preformed sites are unavailable), we demonstrated that intercalation sites could be formed by inserting an anthracene moiety between the (anticipated) site-flanking base pairs and by relaxing the structure using either energy minimization or preferably molecular dynamics simulations. Such sites were suitable for the docking of different intercalators by AUTODOCK.



1. INTRODUCTION

Due to the role of DNA in controlling cellular functions, it is considered an excellent target for treating genetic diseases most notably cancer.[1] Moreover due to the rapid proliferation of cancer cells, most anticancer drugs target the cell cycle.

Among such drugs, small molecules which bind to double stranded DNA constitute an important research topic, and several binders have been in clinical use for many years (e.g., doxorubicin and daunorubicin).[2,3] Ligand binding to double stranded DNA can occur via three main modes namely, groove binding, intercalation between two base pairs, and covalent binding/metal coordination to the bases (see Figure 1). Furthermore, a single compound can exhibit more than one binding mode (e.g., intercalation and covalent binding). Binding to the sugar−phosphate backbone may also occur but is less common.

Covalent binding can occur in three ways, namely, monoalkylation, interstrand cross-linking, and intrastrand cross-linking.[4] The most potent agents of this class of binders cross-link the two DNA strands (interstrand cross-linking) rather than alkylating only one strand (intrastrand cross-linking or monoalkylation).[5] This covalent binding is irreversible and leads to the inhibition of DNA-related processes and finally to apoptosis. One of the most successful alkylating agents used to treat cancer is cis-platin which forms two metal-coordination bonds to the nitrogen of two adjacent purine DNA bases thereby blocking replication and/or preventing transcription and leading to cell death.[6,7] Other cross-linking agents which showed antibiotic antitumor activity are Azinomycins A and B.[8−10]

When two DNA strands are coiled together they form two different grooves in the B-DNA structure which run parallel to the backbone, namely the major groove and the minor groove. These grooves differ both in size (the major groove being wider and deeper) and in binding elements. While both grooves bind their respective ligands via hydrogen bonds, the major groove can also utilize exposed methyl groups found on **T**. These differences lead to groove selectivity[11] whereby proteins and large molecules (e.g., oligonucleotides[12]) usually bind to the major groove. Small molecules can bind to both grooves (e.g., the antitumor agent pyriplatin[13] binds to the major groove and the antitumor antiviral agent netropsin[14] binds to the minor one). Still, small molecules preferably bind to the minor

Received: June 18, 2013
Published: December 4, 2013

   dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

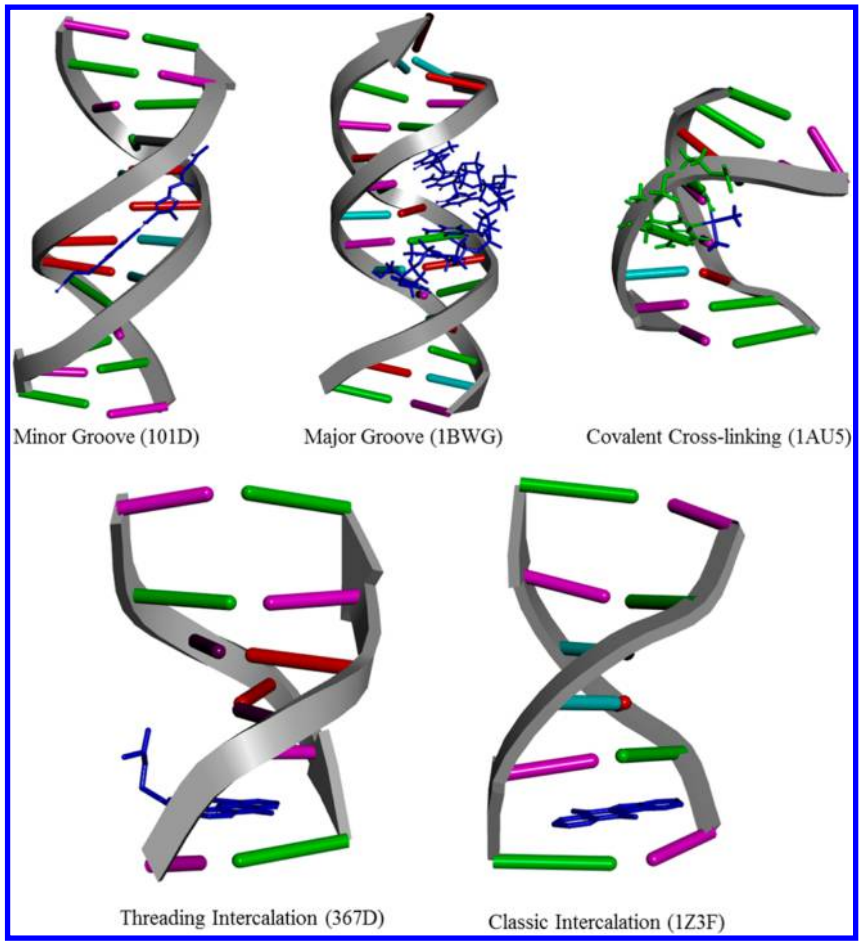


Figure 1. Representative DNA binding modes. The ligand (stick representation) is shown in blue. PDB codes are given in parentheses. In 1AU5 the cross-linked bases are shown in a stick representation.

groove.[15] Minor groove binders feature a flat crescent shape or the ability to adopt such a shape which allows fitting into the groove with little steric hindrance.[16] **AT** rich regions are narrower and slightly deeper than **CG** rich regions allowing for better ligand fit. Thus, many minor groove binders were shown to prefer **AT** rich sequences.[17−19] Overall, groove binders interfere with the interaction of the DNA with DNA binding proteins such as transcription factors and DNA repair proteins. This may lead to the arrest of proliferation and finally to cell death.

One of the most common binding modes of small aromatic molecules is intercalation, namely, the insertion of a small molecule or fragment between two adjacent base pairs of the DNA strand either with (e.g., acridine-4-carboxamides;[20] Threading) or without (e.g., benzopyrene;[21] Classic) additional interactions with the grooves. Such an insertion requires a gap opening between the stacked base pairs which results in the formation of a hydrophobic pocket. The insertion of most intercalators occurs into **CG** intercalation site.[22]

DNA intercalators usually contain a planar polyaromatic system which forms $\pi-\pi$ interactions with the two flanking bases. These interactions are thought to be the main driving force of the intercalation binding mode. Other interactions which contribute to the formation of DNA-intercalator complexes are van der Waals, hydrogen bonding, and charge transfer interactions.[23−25] Many synthetic intercalators are positively charged due to the important contribution of electrostatics to the binding energy.[26]

Binding of intercalators induces conformational changes in DNA structures including the opening of a gap between the flanking bases and an elongation and unwinding of the helical twist.[27−29] These structural changes affect the biological functions of DNA including the inhibition of transcription, replication, and DNA repair processes thereby making intercalators potent mutagens and potential antitumor drugs.

Despite their pharmaceutical relevance, only a handful of docking studies have been performed on DNA-ligand complexes, and even fewer studies targeted the intercalation sites in DNA.[30−45] Thus, DOCK has been used for the analysis of RNA-ligand complexes,[45,46] and DNA minor groove binders have been systematically studied by GLIDE,[47] GOLD,[48] CDOCKER,[49] and AUTODOCK.[30,40,50] However, docking simulations of DNA-intercalator complexes have been limited to a handful of nonsystematic examples.[34,50] Furthermore, since most existing docking methods were parametrized for and validated on ligand-protein complexes, their performances for ligand-DNA docking remain largely unknown. DNA is structurally distinct from "typical" proteins having unique features such as high charge density, distinct geometrical symmetry, and exposed binding sites.[46] The lack of docking studies on DNA-intercalators complexes despite their potential usage in cancer treatment urged us to carry out this study as a preface for the design of new potent anticancer drugs. Since most docking studies on ligand-DNA complexes reported to date have been performed using AUTODOCK,[50] we have decided to focus on this program in the current study.

dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

With this in mind we have secured a data set of 63 high resolution DNA-intercalator complexes from the PDB and following the optimization of the docking parameters, evaluated the ability of AUTODOCK to reproduce their crystal structures under two settings differing by the positioning of the docking box relative to the DNA structure. Based on these results, we have identified ligand characteristics required for good performances of AUTODOCK and a potential scoring problem. Finally, we have evaluated the ability of energy minimization and molecular dynamics (MD) simulations to generate sites on canonical B-DNA structures appropriate for the docking of intercalators by AUTODOCK.

2. METHODS

2.1. Complex Selection and Preparation. A search of the protein data bank (PDB) for DNA-ligand complexes identified 660 hits. These were further processed by removing DNA−protein complexes, complexes with a metal ion as a ligand, complexes solved by NMR featuring more than one model, and structures with a resolution >2.00 Å. The remaining structures were visually inspected and divided into intercalators (63 complexes), minor groove binders (44 complexes), covalent ligands (42 complexes), and ligands binding at other sites on the DNA (85 complexes). In this work we focused on the set of 63 DNA-intercalator complexes.

2.2. Docking Simulations. All docking simulations were performed using AUTODOCK version 4.2.[50] Prior to docking the protonation state of each ligand was determined, within a pH range of 7 ± 2, using the Epik program.[51,52]

2.2.1. Parameterization of Docking Protocol. 2.2.1.1. Subset Selection. Each of the 63 ligands was characterized by seven 2D descriptors (AlogP, molecular weight, number of rotatable bonds, number of H-bond acceptors and donors, number of rings and aromatic rings), and a diverse subset of six ligands was selected using the 'Find Diverse Molecules' protocol as implemented in Discovery Studio (DS) version 2.5.[53] This selected subset (PDB codes 1L0R, 1IMS, 1Z3F, 1N37, 385D, and 465D) was used to determine the optimal partial charges and the docking parameters for the subsequent docking simulations.

2.2.1.2. Assignment of Partial Charges. Five different sets of charges were assigned to the diverse subset of six ligands, namely, Charmm charges calculated in DS, OPLS-AA charges calculated by MacroModel[54] as implemented in Maestro 9.0,[55] Hartree−Fock/6-31G* ESP charges calculated through fitting of the electrostatic potential by the Jaguar program[56] also implemented in Maestro, and Gasteiger charges which are the default partial charges of the AUTODOCK program. Since Gasteiger charges are implemented differently in AUTOCOK and in DS, these were calculated by both programs.

2.2.1.3. Defining Parameter Sets. Grid boxes for the docking simulations were defined using AutoDockTools 1.5.6 (ADT).[57] Two boxes were examined for each ligand, focused either on the preformed intercalation sites as available from the crystal structure (small grid box) or on the entire DNA (large grid box). For the small box, the center of the intercalation site was set as the grid center. For the large box the center of the macromolecule was set as the center of the grid box. In both cases the size of the grid box varied among the different DNA structures so as to include the entire desired binding site. The grid spacing of all grid boxes was set to the default value of 0.375 Å. Complexes with two optional intercalation sites were tested for DNA symmetry by halving the DNA structure and by superimposing the two halves using Maestro. Symmetrical cases were defined as such for which the resulting RMSD values were found to be 0.00 Å (considering two significant digits). For these cases only, the crystal structure intercalation site was completely included in the large grid box, while the second site was truncated to avoid symmetry related binding modes.[36] This truncation however maintained the integrity of the minor groove site. All ligands were completely flexible during the docking simulations.

Based on previous works,[36,42] five combinations of docking parameters were considered in which the maximum number of energy evaluations was set to 20,000,000 (2E7), 70,000,000 (7E7), and 5,000,000 (5E6), and the number of docking runs was set to five, ten, or 20 runs. These two parameters were shown to have the largest impact on the length and success of docking simulations.[22] Each of these sets was tested with each of the five charge schemes discussed above for a total of 25 parameter sets (see Table 1). The performances of each of these sets were evaluated for the diverse set of six ligands on both grid boxes (large and small), using the following criteria: (1) The average RMSD of the best scoring poses (criterion 1). (2) The normalized average rank (in %) of the highest ranked poses with an RMSD value with respect to the known crystal structure <2.00 Å (criterion 2). Normalization was performed according to the following eq 1

Table 1. Summary of the 25 Different Docking Parameter Sets

	# docking runs	# energy evaluations	partial charges
Set 1	5	2.00×10^{07}	Charmm
Set 2	5	2.00×10^{07}	Hartree−Fock/6-31G*
Set 3	5	2.00×10^{07}	Gasteiger-DS
Set 4	5	2.00×10^{07}	Gasteiger-AUTODOCK
Set 5	5	2.00×10^{07}	OPLS
Set 6	10	7.00×10^{07}	Charmm
Set 7	10	7.00×10^{07}	Hartree−Fock/6-31G*
Set 8	10	7.00×10^{07}	Gasteiger-DS
Set 9	10	7.00×10^{07}	Gasteiger-AUTODOCK
Set 10	10	7.00×10^{07}	OPLS
Set 11	20	5.00×10^{06}	Charmm
Set 12	20	5.00×10^{06}	Hartree−Fock/6-31G*
Set 13	20	5.00×10^{06}	Gasteiger-DS
Set 14	20	5.00×10^{06}	Gasteiger-AUTODOCK
Set 15	20	5.00×10^{06}	OPLS
Set 16	10	2.00×10^{07}	Charmm
Set 17	10	2.00×10^{07}	Hartree−Fock/6-31G*
Set 18	10	2.00×10^{07}	Gasteiger-DS
Set 19	10	2.00×10^{07}	Gasteiger-AUTODOCK
Set 20	10	2.00×10^{07}	OPLS
Set 21	5	7.00×10^{07}	Charmm
Set 22	5	7.00×10^{07}	Hartree−Fock/6-31G*
Set 23	5	7.00×10^{07}	Gasteiger-DS
Set 24	5	7.00×10^{07}	Gasteiger-AUTODOCK
Set 25	5	7.00×10^{07}	OPLS

$$\overline{\text{rank}} = \frac{\sum_{i,j} \text{rank}_i}{\frac{\text{NumOfComplexes}}{\text{NumOfPoses}}} \times 100\% \quad (1)$$

where i is the highest ranked pose with an RMSD < 2.00 Å, NumOfComplexes is the size of the data set, and NumOfPoses is the number of poses generated in the current run (e.g., 10

dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

docking runs generate 10 poses). (3) The fraction of ligands for which the pose with the best RMSD had a value <2.00 Å (criterion 3). (4) The fraction of ligands for which the best scoring pose had an RMSD value <2.00 Å (criterion 4). The rest of the docking parameters were set to their default values. Lamarckian genetic algorithm (LGA) was selected as the search method since it was shown to be the most efficient, and reliable in reproducing experimental data.[50]

2.2.2. Decision Tree. A decision tree (using the recursive partitioning J48 model[58] as implemented in WEKA[59]) was built to identify ligand characteristics which affect AUTODOCK's performances in terms of docking success. For this method, each ligand was characterized by seven descriptors (independent variables; AlogP, molecular weight, number of rotatable bonds, number of H-bond acceptors and donors, number of rings and aromatic rings), and the dependent variable was defined as docking success (having a pose with an RMSD < 2.00 Å regardless of its rank using the large grid box; criterion 3 above) and could therefore take two values, namely, 1 (successful docking) or 0 (unsuccessful docking). Two models were built, one for each of the two parameter sets selected to be applied to the entire data set with the large grid box, set 13 and set 18 (see Table 5). Models were developed based on a training set of 50 ligands (80%) and were subsequently used to predict the docking success for a test set containing 13 (20%) ligands. The training and the test sets had the same distribution of compounds in terms of AUTODOCK success (78%) and failure (22%).

To estimate the predictive power of the generated classification models the corrected classification rate (CCR) and Matthew's correlation coefficient (MCC) were used. CCR is given by

$$\mathrm{CCR} = \frac{1}{2}\left(\frac{T_N}{N_N} + \frac{T_P}{N_P}\right) \quad (2)$$

where T_N and T_P represent the number of true negative and true positive predictions, respectively, and N_N and N_P represent the total number of class = 0 and class = 1 compounds in the model. MCC is given by

$$\mathrm{MCC} = \frac{T_N T_P - F_N F_P}{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)} \quad (3)$$

where F_N and F_P denote false negative and false positive predictions, respectively.

2.2.3. Docking into Canonical B-DNA Structures. A canonical B-DNA structure with the sequence $d(CGTACG)_2$ (as found in crystal structures 1D36, 1DL8, 1FD5, 1FDG, 1M69, 1XCS, 1XCU, 215D, 288D, 2DES, 308D, 386D, 452D, 465D) was built using DS and subjected to docking simulations. In accord with the preformed intercalator binding sites observed on the crystal structures, AUTODOCK's grid box was centered on the **CG** base pairs, and the corresponding ligands were docked onto it using parameter set 13. Several docking simulations were performed deferring in the degree of flexibility of the target. However, none resulted in ligand insertion into the intercalation binding site.

Intercalation binding sites on the canonical B-DNA structures were formed by manually inserting an anthracene moiety between the appropriate base pairs in parallel to their aromatic moieties and then by refining the structure. Two refinement protocols were considered, namely, energy minimization and MD simulations. Both protocols were performed with GROMACS version 4.5.5[60] using the Amber99-SB-ILDN force field. The canonical B-DNA structure complexed with the anthracene moiety was placed in a TIP3 dodecahedron water box. Periodic boundary conditions were employed, the electrostatic cutoff was set to 10 Å, and long-range electrostatic interactions were treated with the particle-mesh Ewald (PME) method.[61] Energy minimization was performed using the steepest decent algorithm followed by conjugated gradient. The minimized complex was used as a starting point for an MD simulation. The simulation began with two 100 ps equilibration phases (NVT during which the temperature was gradually increased from 1 to 300 K and NPT at 300 K) followed by a 10 ns production phase under NPT conditions. Energies and coordinates were sampled every 2 ps, and the resulting trajectory was clustered. The center of the largest cluster was selected for subsequent docking simulations.

Structures resulting from the energy minimization and the MD simulation were subjected to new docking simulations using the same docking protocol (set 13) and set of ligands as before. The results were evaluated by calculating the number of ligands with at least one intercalating binding mode and the rank of these poses. For comparison, a similar analysis was performed for the docking results targeting the preformed crystallographic binding sites.

3. RESULTS AND DISCUSSION

3.1. Statistical Analysis of the Data Set. As mentioned above all DNA-ligand complexes retrieved from the PDB were analyzed based on the ligand's binding mode, and 27% of the high resolution complexes (63 cases; < 2.00 Å) were found to include a ligand which intercalates with the DNA structure. This data set has not been previously subjected to systematic computational studies.

The selected complexes were further analyzed based on the two nucleotides which form the intercalation site. It was found that in 70% of these complexes DNA intercalation occurs between the **CG** nucleotides (see Table 2). These values are in accord with previous studies which showed a preference of intercalators to bind in an alternating pyrimidine-purine site. This was explained by the assumption that π-stacking stabilization at a pyrimidine-purine site is weaker and easier to disrupt by intercalation than the π-stacking stabilization of a nonalternating site. Furthermore, a slight preference for a **CG** intercalation site was reported in previous studies.[22,43]

Table 2. Binding Site Analysis of 63 DNA-Intercalator Complexes in the PDB

intercalation site	percent
AA	2%
AT	1%
AC	4%
TA	2%
TG	5%
CA	11%
CC	1%
CG	**70%**
GA	1%
GC	2%

3.2. Docking into Preformed Sites. The 63 DNA-intercalator complexes selected from the PDB were subjected to a systematic docking study targeting their preformed binding

dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

Table 3. Parameterization Results Using the Small Grid Box

Small Grid Box Parameterization[a]					
parameter set	run and energy evaluations	RMSD avg. (Å) (criterion 1)	rank avg. (%) (criterion 2)	RMSD < 2.00 Å (%) (criterion 3)	top rank and RMSD < 2.00 Å (%) (criterion 4)
Gasteiger-DS					
Set 3	5_2E7	1.92	43	83	67
Set 8	10_7E7	1.84	30	83	67
Set 13	**20_5E6**	**1.80**	**22**	**83**	**83**
Set 18	**10_2E7**	**1.80**	**28**	**83**	**67**
Set 23	5_7E7	1.91	47	83	50
Gasteiger-AUTODOCK					
Set 4	5_2E7	1.87	40	83	67
Set 9	10_7E7	1.75	28	83	67
Set 14	20_5E6	2.22	30	83	33
Set 19	10_2E7	2.36	35	83	67
Set 24	5_7E7	1.85	47	83	50
Charmm					
Set 1	5_2E7	1.92	50	83	50
Set 6	10_7E7	2.59	47	83	50
Set 11	20_5E6	2.17	31	83	33
Set 16	10_2E7	2.22	47	83	33
Set 21	5_7E7	2.38	50	83	50

[a]The values for the two best parameter sets are shown in bold.

sites. As a first step the 25 parameter sets defined in Table 1 were applied to the subset of six diverse ligands, and their docking results were evaluated according to four criteria detailed in the Method section. These criteria evaluate both the ability of AUTODOCK to accurately reproduce the structure of the complex (e.g., criterion 3) and to place poses with low RMSD values with respect to the crystal structure at or near the top of the ranked poses list (e.g., criterion 4). The most relevant criterion is problem dependent. For example virtual screening campaigns are likely to emphasis criterion 4, whereas lead optimization studies might focus on criterion 3. RMSD values were calculated with AUTODOCK which takes into account molecular symmetry. The 2.00 Å cutoff was selected since it was used in most of the previous studies.[33,42,62]

3.2.1. Docking with a Small Grid Box. 3.2.1.1. Selection of Parameter Sets. All 25 parameter sets were applied to docking the selected subset of six ligands using a grid box focused on the known binding site (small grid box). Docking results of all 25 different sets were examined considering the four criteria mentioned above. In all cases Gasteiger and Charmm partial charges outperformed OPLS and Hartree−Fock/6-31G* charges (data not shown). Set 13 yielded an average RMSD of 1.80 Å for the top ranking poses. Furthermore, for this set, a pose with an RMSD less than 2.00 Å with respect to the crystal structure was found, on average, among the top 22% of the generated poses. In five out of six cases at least one pose had an RMSD less than 2.00 Å, and in five out of six cases the top ranked pose had an RMSD less than 2.00 Å. For set 18, the corresponding values were 1.80 Å, 28%, five out of six and four out of six, respectively. For set 9 the corresponding values were 1.75 Å, 28%, five out of six and four out of six, respectively. However set 9 requires 3.5 and seven times more computer resources than sets 18 and 13, respectively, and was therefore considered to be less favorable. The data is shown in Table 3. Since overall parameter sets 13 and 18 had the best results among all 25 sets, we chose to apply them to the entire data set.

3.2.1.2. Application to the Entire Data Set. The two best parameter sets shown above were applied to the entire data set (a total of 63 ligands). For set 18 the RMSD values of the entire data set ranged from 0.49 to 8.76 Å, and the average RMSD of the top ranked poses was 1.75 Å (Table S1, Supporting Information). Furthermore, for this set, a pose with an RMSD less than 2.00 Å with respect to the crystal structure was found, on average, among the top 28% of the generated poses. In 84% of the cases at least one pose had an RMSD less than 2.00 Å, and in 76% of the cases the top ranked pose had an RMSD less than 2.00 Å. For set 13, the RMSD values were lower (0.53−7.63 Å), but the average RMSD for the top ranked pose was slightly higher at 1.84 Å (Table S2, Supporting Information). Moreover, for this set, a pose with an RMSD less than 2.00 Å with respect to the crystal structure was found, on average, among the top 25% of the generated poses. In 83% of the cases there was at least one pose with an RMSD less than 2.00 Å, and in 73% of the cases the top ranked pose had an RMSD less than 2.00 Å. Based on all four criteria, these results indicated that both sets produced very similar results for this data set and by extension, for cases where the location of the binding site is known.

3.2.2. Docking with a Large Grid Box. A large grid box represents a more realistic case where the exact location of the binding site is unknown.

3.2.2.1. DNA Symmetry. Symmetrical DNA structures may result in the presence of two identical binding sites. In such cases, the docking procedure might place the ligand in the "wrong" binding site which will result in very high RMSD values. Since when using a small grid box which is focused only on the "correct" binding site, DNA symmetry is not an issue, we address this potential symmetry problem only for the large grid box case.

All 63 DNA structures were examined for symmetry by halving their structures and superimposing the two halves. In twelve cases DNA symmetry was considered irrelevant either since the intercalator was found to bind in the center of the structure (PDB codes 1L0R, 1N37, 1QCH, and 2GWA) or since only a single intercalation site exists (PDB codes 1FN1, 1FN2, 1EG6, 1FD5, 1FDG, 1RQY, 1XCS, and 2GWA). The

dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

remaining 51 cases could be divided into two groups with RMSD values of 0.00 Å (40 cases; see Figure 2) and 0.03−0.53

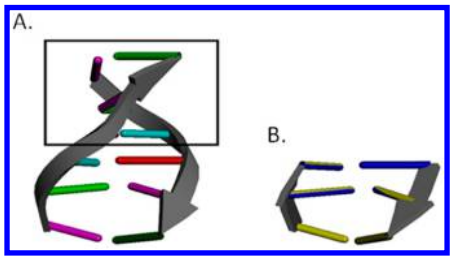


Figure 2. A. The complete DNA structure of 1D11 (without the ligand). The box distinguishes between the two halves of the structure. B. The two halves (top half in blue and bottom in yellow) are superposed with an RMSD of 0.00 Å.

Å (11 cases; PDB codes: 182D, 198D, 1D21, 1D22, 1Z3F, 224D, 245D, 258D, 2DES, 2GB9, and 308D; see Figure 3). In

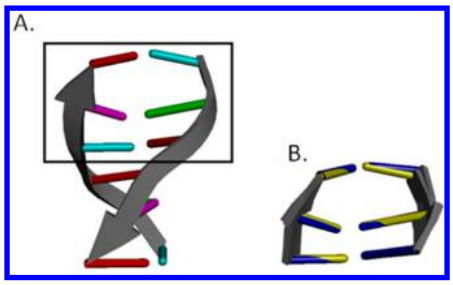


Figure 3. A. The complete DNA structure of 245D (without the ligand). The box distinguishes between the two halves of the structure. B. The two halves (top half in blue and bottom in yellow) are superposed with an RMSD of 0.18 Å. There is a small deviation between the two halves.

cases for which no symmetry was identified (RMSD > 0.00 Å) the complete DNA structure was included in the docking simulation. For the rest of the cases only the crystal structure intercalation site was completely included in the simulation box, whereas the second site was truncated. This was done to avoid docking in the second binding site which will result in a large RMSD value from the crystal structure and hence will affect the docking results. Importantly, even in these cases the groove site was kept intact.

3.2.2.2. Parameter Set Selection. All 25 parameter sets were applied to the selected subset of six ligands using a large grid box, and the results were evaluated according to the four criteria defined above (see Table 4). As before, Gasteiger and Charmm partial charges outperformed OPLS and Hartree−Fock/6-31G* partial charges (data not shown). Similar to the small grid box, sets 13 and 18 provided the best results while considering all four criteria. Using set 13 the average RMSD value for the top ranked pose was 1.80 Å. Furthermore, for this set, a pose with an RMSD less than 2.00 Å with respect to the crystal structure was found, on average, among the top 30% of the generated poses. In five out of six cases at least one pose had an RMSD less than 2.00 Å, and in three out of six cases the top ranked pose had an RMSD less than 2.00 Å. For set 18 the corresponding values were 1.79 Å, 43%, four out of six, and four out of six. Both sets were therefore applied to the rest of the data set.

3.2.2.3. Results on Full Data Set. The selected parameter sets were applied to the docking of the full data set using a large grid box. Not unexpectedly, the RMSD variations were much larger than when using a small grid box (0.55−16.83 Å for set 13 and 0.60−16.25 Å for set 18), and the average RMSD of the top ranked pose were higher (3.83 Å and 3.96 Å for sets 13 and 18, respectively; these numbers improved to 3.04 Å and 2.60 Å, upon treating cases where the RMSD between the two DNA halves was smaller than 0.60 Å as symmetrical). These larger RMSD values are readily explained by the higher number of potential docking poses available in the larger grid box. A pose with an RMSD < 2.00 Å with respect to the crystal structure was found, on average, among the top 31% and 38% of the

Table 4. Parameterization Results Using the Large Grid Box

		Large Grid Box Parameterization[a]			
parameter set	run and energy evaluations	RMSD avg. (Å) (criterion 1)	rank avg. (%) (criterion 2)	RMSD < 2.00 Å (%) (criterion 3)	top rank and RMSD < 2.00 Å (%) (criterion 4)
Gasteiger-DS					
Set 3	5_2E7	2.32	53	67	67
Set 8	10_7E7	1.82	43	67	67
Set 13	**20_5E6**	**1.80**	**30**	**83**	**50**
Set 18	**10_2E7**	**1.79**	**43**	**67**	**67**
Set 23	5_7E7	1.85	60	67	50
Gasteiger-AUTODOCK					
Set 4	5_2E7	3.01	73	50	33
Set 9	10_7E7	2.05	53	67	50
Set 14	20_5E6	2.67	32	83	50
Set 19	10_2E7	2.18	50	67	33
Set 24	5_7E7	2.67	60	67	50
Charmm					
Set 1	5_2E7	2.53	87	33	33
Set 6	10_7E7	3.34	62	50	33
Set 11	20_5E6	2.16	35	83	33
Set 16	10_2E7	2.69	50	83	17
Set 21	5_7E7	2.11	73	50	33

[a]The values for the two best parameter sets are shown in bold.

dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

generated poses for sets 13 and 18 (improved to 30% and 33%). For set 13, in 78% of the cases at least one pose had an RMSD < 2.00 Å with respect to the crystal structure, and in 60% of the cases the top ranked pose had an RMSD < 2.00 Å (Table S3, Supporting Information). These numbers increased to 78% and 65% upon treating cases where the RMSD between the two DNA halves was smaller than 0.60 Å as symmetrical. For set 18, the corresponding values were 78%, 60%, 78%, and 70%, respectively (Table S4, Supporting Information). Overall the performances of both sets were very similar when considering the four criteria used to evaluate AUTODOCK's results.

These results suggest that, depending on the exact success criterion, AUTODOCK is often able to successfully dock ligands into DNA targets with preformed binding sites yet in an *a priori* unknown location (see Figure 4 for representative

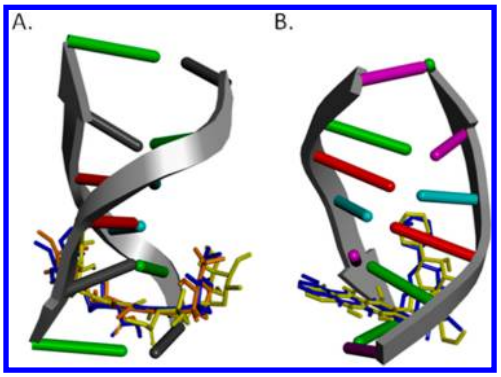


Figure 4. A. Complex 1D17: Three conformations are presented, the crystal ligand (score = −8.68 kcal/mol, blue), the top ranked pose generated by AUTODOCK (score = −10.58 kcal/mol, RMSD = 2.59 Å, yellow), and the pose with the best RMSD (score = −10.37 kcal/mol, RMSD = 0.85 Å, orange). B. Complex 235D: Two conformations are presented, the crystal conformation (score = −6.86 kcal/mol, blue) and the top ranked pose generated by AUTODOCK (which is also the pose with the best RMSD, score = −10.14 kcal/mol, RMSD = 1.08 Å, yellow).

examples). Furthermore, in most cases AUTODOCK is able to distinguish between the intercalation site and the close by groove binding site and will rarely dock intercalators as groove binders. There were several cases in which AUTODOCK placed a ligand as a groove binder rather than as an intercalator. However such poses had a lower score than poses which were correctly docked in the intercalation site (see Figure 5).

3.2.3. Performances of AUTODOCK as a Function of Ligand Structure. In order to gain insight into the effect of ligand structure on the performances of AUTODOCK we have developed a binary classifier in the form of a decision tree (DT). Two models were built using docking results obtained with sets 13 and 18. The model based on set 13 was able to correctly classify 94% of the training set with a CCR value of 0.88 and a MCC value of 0.84 (Table 5). Applying this DT to the test set resulted in 100% success, all ligands being correctly classified (CCR and MCC are 1). The model based on set 18 (see Figure 6 and Table 5) was able to correctly classify 92% of the training set with a CCR value of 0.82 and a MCC value of 0.77. Applying this DT to the test set also resulted in 100% success (CCR and MCC equal 1).

Both decision trees show a split into two nodes based on the number of rings in the molecule with successfully docked ligands characterized by num_rings larger than four and

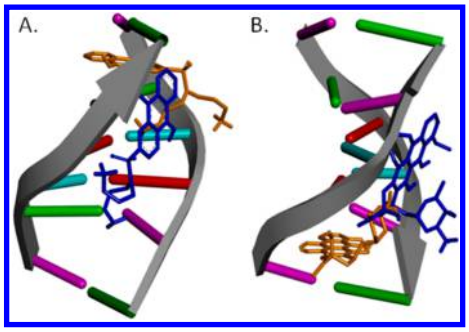


Figure 5. A. Complex 386D. Two conformations are presented, the top ranked conformation docked in the intercalation site (score = −7.76 kcal/mol, orange) and a conformation which was docked into the groove binding site (score = −7.09 kcal/mol, blue). B. Complex 1D10. Two conformations are presented, the top ranked conformation docked in the intercalation site (score = −12.04 kcal/mol, orange) and a conformation which was docked into the groove binding site (score = −7.17 kcal/mol, blue).

unsuccessfully docked ligands characterized by less than four rings. For set 18 this second node was further split based on the number of H-bond acceptors and subsequently by molecular weight. For set 13 the second node was split based on the number of rotatable bonds, and molecules with less than five rotatable bonds were further split based on the number of H-bond acceptors followed by molecular weight as in set 18 (see Figure 6). Overall the models suggest that the number of rings in the ligand is the main factor governing docking success so that ligands with more than four rings will most probably be accurately docked by AUTODOCK. The presence of multiple (aromatic) rings allows for π-stacking interactions with the nucleotides which are the main driving force in the intercalation binding mode. Ligands with less than four rings require additional features to compensate for the relative lack of aromatic interactions, for example, through a high molecular weight (>210 Da) which is suggestive of larger hydrophobic interactions. Ligands with low molecular weights do not fill the binding site very well and thus are less well docked by the program. Furthermore molecules with more than five rotatable bonds seem to be too flexible for AUTODOCK to correctly dock.

3.2.4. Analysis of Poses' Scores. Our results suggest that in most cases, AUTODOCK is able to reproduce the crystal structure of intercalator-DNA complexes within an RMSD of 2.00 Å. However, the program is unable to consistently identify low RMSD poses with respect to the crystal structures as top scorers. In fact, only in a few cases the best scoring pose corresponds to the lowest RMSD one (small grid box: 5% for set 13 and 10% for set 18; large grid box: 8% for set 13 and 10% for set 18). This RMSD/energy "mismatch" was also observed in previous works.[42] Furthermore, when examining the score values, all 63 ligands scored better in (at least one of) their docked poses than in their crystal conformation. This indicates a potential scoring problem in AUTODOCK as was suggested in a previous study.[32] To gain insight into the magnitude of this problem we tested whether ligand conformations which are very similar to the crystal structure will have better scores than poses obtained by the docking simulations. Furthermore, to identify possible sources of the scoring problem we have analyzed high-RMSD docking poses.

3.2.4.1. Score Analysis of Crystal-like Poses. Ten representative complexes with varying energy differences

Table 5. Confusion Matrices[a]

Train Set				Test Set			
		AUTODOCK				AUTODOCK	
		Successful docking	Unsuccessful docking			Successful docking	Unsuccessful docking
Decision Tree	Successful docking	39/38[b]	0/0	Decision Tree	Successful docking	10/10	0/0
	Unsuccessful docking	4/3	7/9		Unsuccessful docking	0/0	3/3

[a]The table on the left gives the results for the training set in which 46 out of 50 ligands (for set 18) or 47 out of 50 ligands (for set 13) were correctly classified based on docking results obtained for these sets. The table on the right provides the results for the test set where the models correctly predicted all 13 ligands. The resulting CCR and MCC values for the training sets corresponding to parameter sets 18 and 13 are 0.82, 0.77 and 0.88, 0.84, respectively, and both values are 1 for the test sets. [b]The first number corresponds to set 18 and the second to set 13.

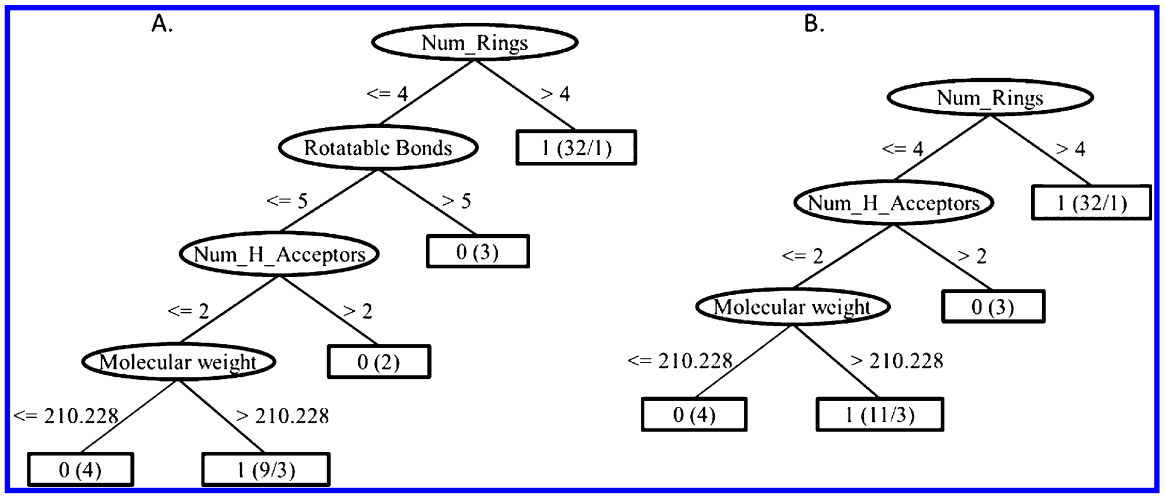


Figure 6. Decision trees for predicting DNA intercalators which will be successfully docked by AUTODOCK. A. using set 13. B. using set 18.

between the crystallographic pose and the lowest energy pose generated by AUTODOCK were selected (1D12, 1DA0, 1M69, 1QCH, 1R68, 1XCU, 215D, 278D, 2DES, 308D). Starting from the crystal structure a conformational search with a ±1 Å flat bottom Cartesian constrain on all atoms was performed using MacroModel with the OPLS 2005 force field in GBSA water. The flat-bottom constrain defines the distance along each of the Cartesian coordinates an atom is allowed to move from its initial position before an energy penalty is applied. A minimum number of energy minimizations steps (five) were performed in order to sample the conformational space around the ligand's crystal structure as densely as possible. This conformational search generated an ensemble of conformations for each ligand with an RMSD from the crystal structure which varied between 0.17 and 0.77 Å. The resulting conformations were placed into the binding site by superimposing them on the original crystal structure (see Figure 7). These conformations were scored within the binding site using the "epdb" command. The results showed that the best energy for each ligand was still lower than the energies of these crystal-like conformations.

Since for each ligand the conformational ensemble included at least one conformer with an RMSD with respect to the crystal structure as high as 0.55 Å, we conclude that AUTODOCK is unable to locate a pose within 0.55 Å of the crystal structure as the lowest energy one. Stated otherwise, when looking at the lowest energy pose generated by AUTODOCK for intercalators docked into DNA, one cannot

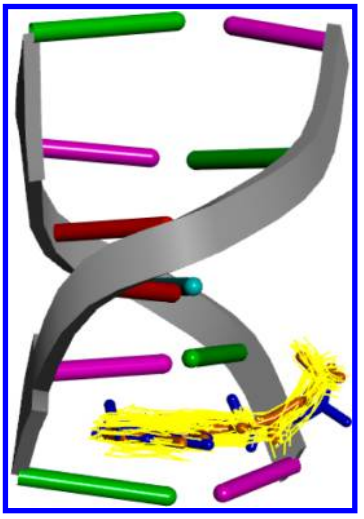


Figure 7. Ensemble resulting from a flat bottom conformational search for 1M69. The crystal conformation, AUTODOCK's top ranked pose, and the poses obtained by the flat bottom search are depicted in orange, blue, and yellow, respectively.

dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

expect to get closer than 0.55 Å to the crystal structure. Furthermore, in 80% of the cases RMSD values were as high as 0.60 Å.

3.2.4.2. Analysis of High RMSD Poses. For several ligands AUTODOCK was unable to generate any pose with an RMSD smaller than 2.00 Å with respect to the crystal structures. Analysis of these cases may provide insight into AUTODOCK's scoring deficiencies. Nine of these DNA structures are characterized by distorted intercalation sites, namely, sites in which one nucleotide is oriented away from all others or is altogether missing (PDB codes: 1FN1, 1FN2, 1EG6, 1FD5, 1FDG, 1RQY, 1XCS, 1XCU, and 2GWA; see Figure 8 for a representative example). In these cases most of the ligand poses generated by AUTODOCK were located in the groove site, rather than in the intercalation site (Figure 8).

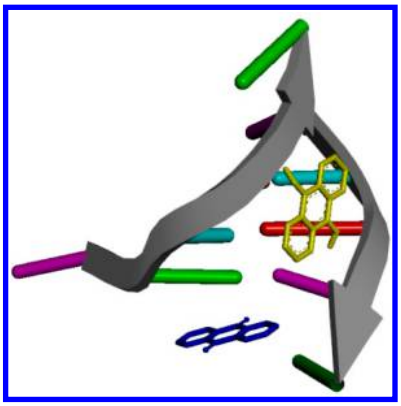


Figure 8. DNA structure with a distorted preformed intercalation site (PDB code 1FDG). In this case AUTODOCK failed to dock the ligand in the intercalation site and instead docked the ligand in the groove site. (Crystal ligand in blue, top ranked pose generated by AUTODOCK in yellow).

Inappropriate positioning of nucleotides' side chains reduces the magnitude of DNA-ligand $\pi-\pi$ interactions since the ligand can now interact with three rather than with four aromatic moieties. AUTODOCK's poor performances in such cases may suggest that the scoring function underemphasizes $\pi-\pi$ interactions. Indeed, AUTODOCK does not have a specific term for such interactions, and they are considered as part of the hydrophobic term. To test this hypothesis, we used crystal symmetry to build the entire unit cell. In eight cases this led to the placement of an alternative base in the space left unoccupied by the rotation of the original base and to the restoration of the site's π system. Yet ligand redocking into the restored site did not improve the results. Thus, factors other than inappropriate treatment of aromatic interactions may be responsible to AUTODOCK's scoring problem, at least as manifested in these cases.

With this in mind, we have analyzed the DNA-ligand hydrogen bonding patterns in these cases and found the number of H-bonds to be consistently higher in the best ranking poses than in the crystal structures. This may suggest that AUTODOCK's scoring function overemphasizes the hydrogen bond term. To further probe this assumption we have extended this analysis to all cases where AUTODOCK was unable to find a pose with an RMSD < 2.00 Å as the top ranked one. Indeed we found that this trend holds for all of these cases (21 including the nine distorted structures). Furthermore, in ten of these cases there is a pose with an RMSD < 2.00 Å. These poses also feature a larger number of H-bonds than the crystal structures, yet this number is smaller than for the top ranking poses. These data point to a clear tendency of high RMSD poses to participate in a large number of H-bonds with the intercalation site, which supports our hypothesis that H-bonds are overemphasized, at least for DNA intercalators. This is in fact in line with some features of the decision trees (Figure 6). Specifically ligands with larger numbers of H-bond acceptors are less probable to be correctly docked to the intercalation sites.

3.3. Docking into Canonical B-DNA Structure. As noted in the Introduction, the presence of an intercalator affects the DNA structure by opening a gap between two consecutive base pairs. Indeed, all of our docking simulations targeted preformed sites. However, when targeting DNA sequences which lack crystal structures, such sites are unavailable. These sequences could be modeled by canonical B-DNA structures. Evaluating the performances of AUTODOCK on canonical B-DNA structures is therefore important.

With this in mind, a canonical B-DNA structure with the sequence $d(CGTACG)_2$ was built using DS. This sequence is common to 14 DNA-intercalator complexes in our data set (PDB codes: 1D36, 1M69, 1DL8, 1FD5, 1FDG, 1XCU, 1XCS, 215D, 288D, 2DES, 308D, 386D, 452D, and 465D). Initially, rigid docking was performed for each of the 14 ligands using this B-DNA structure as the target. Not surprisingly, in all cases AUTODOCK was unable to dock the ligands as intercalators. Similar results were obtained upon introduction of torsional flexibility into the DNA structure in course of the docking simulations. Clearly, torsional flexibility, at least as implemented in AUODOCK, is not enough to induce gap opening between DNA base pairs. These results demonstrate that AUTODOCK

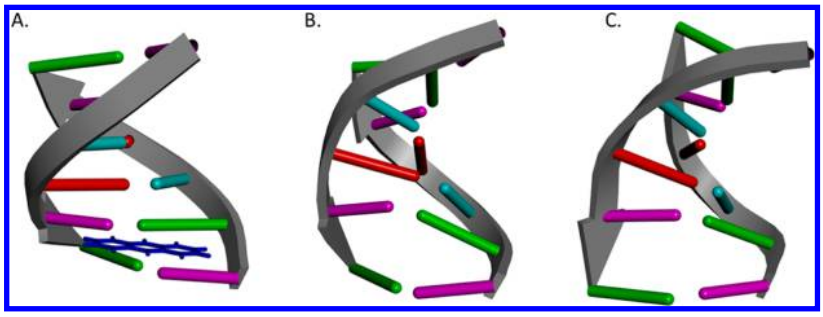


Figure 9. Canonical B-DNA with the sequence $d(CGTACG)_2$. A. Original canonical B-DNA with the anthracene moiety placed between the **CG** base pairs. B. Canonical B-DNA with a preformed intercalation site between the **CG** base pairs generated by energy minimization. C. Canonical B-DNA with a preformed intercalation site between the **CG** base pairs generated by MD simulations.

dx.doi.org/10.1021/ci400352t | J. Chem. Inf. Model. 2014, 54, 96−107

is unable to dock intercalators without a preformed intercalation site.

Having failed to demonstrate gap opening between the relevant base pairs during the docking simulations we turned our attention to a second alternative, namely, forming an intercalation binding site prior to docking. Previous reports have demonstrated that gap opening could be achieved through energy minimization.[62] With this in mind, we have manually placed an anthracene moiety between the **C** and **G** base pairs of the canonical B-DNA (see Figure 9A). This moiety was selected since it is common to many intercalators. The resulting complex was minimized leading to a DNA structure with an intercalation gap which was subjected to docking simulations (see Figure 9B). Furthermore, since DNA is known to be a flexible molecule capable of adopting several distinct isoforoms,[63] the B-DNA structure with the newly formed gap was further refined by MD simulations. The resulting trajectory was clustered, and the conformation representing the center of the largest cluster (without the anthracene moiety) was selected as input to additional docking simulations (see Figure 9C). The above-discussed 14 ligands were docked into the two new structures using two grid boxes (small and large) and parameter set 13 as before. The results are presented in Table S5, Supporting Information. For the minimized structure, an intercalation binding mode was found in 93% and 64% of the cases using the small grid box and the large grid box, respectively. For the MD-refined structure these numbers increased to 100% for both grid boxes. The corresponding numbers obtained when using the crystal structures for all 14 complexes are 100% and 93% for the small and large grid boxes, respectively. These results demonstrate that when using a canonical B-DNA structure with a preformed intercalation site created by energy minimization or MD simulations, AUTODOCK is able to dock ligands as intercalators. Furthermore, a (nonspecific) intercalator model system (i.e., anthracene) is sufficient to induce the formation of an intercalator site suitable for the docking of multiple ligands. A more quantitative evaluation of the performances of AUTODOCK in terms of its ability to dock intercalators into preformed intercalation sites requires a comparison between the resulting poses and the corresponding crystal structures. Such a comparison however also depends on the ability of the energy minimization and MD simulations to reproduce the DNA crystal structure and is beyond the scope of the current work. Such a comparison will be undertaken in a subsequent study.

4. CONCLUSIONS

In this work, we examined the ability of the docking program AUTODOCK which was mainly parametrized for and applied to ligand-protein complexes to reproduce DNA-ligand complexes. Following the optimization of the docking protocol, docking simulations were performed on a data set of 63 DNA-intercalator complexes with a preformed intercalation site obtained from the PDB. Overall, satisfactory results were obtained while docking ligands into DNA structures with *a priori* known and unknown binding sites. Thus, AUTODOCK was able to generate a conformation with an RMSD < 2.00 Å in 84% of the cases with a known binding site and in 78% of the cases with an unknown binding site using two different parameter sets. A top ranked conformation with an RMSD < 2.00 Å was identified in 76% of the cases with a known binding site and in 60% of the cases with an unknown binding site using again two different parameter sets. Furthermore, AUTODOCK was able to distinguish between the two major binding modes, namely intercalation and groove binding. Even though both sites were included in the simulation box the intercalators were preferentially docked into the intercalation site and not into the DNA grooves. Using the docking results and a set of ligand-derived molecular descriptors, we have developed decision trees to predict the likelihood of successful docking.

Despite these results AUTODOCK is unable to dock intercalators into canonical B-DNA structures lacking a preformed intercalation site. Encouragingly, we have demonstrated that intercalation sites suitable for the docking of different ligands could be preformed by minimizing the structure in the presence of an intercalator "prototype" (anthracene) or, preferably, by further refining it by means of MD simulations. Using a prototype intercalator may provide less biased intercalating sites than sites which could be obtained with more specific intercalators. Such sites may therefore be more suitable for virtual screening.

Still, the docking simulations performed in this work suggest that due to a scoring problem, AUTODOCK is unable to produce binding modes with RMSD values with respect to the crystal structure lower than 0.60 Å as the lowest energy ones. This scoring problem is also manifested in the relative small number of top scorer ligands which are also the most similar to the crystal structures. An analysis of cases with top scoring poses having RMSD > 2.00 Å revealed that at least part of this scoring problem may be attributed to an over emphasis of H-bond interactions.

The results presented in this work point to the potential usefulness of AUTODOCK in the computer aided drug design of new DNA intercalators which may serve as pharmaceutical agents. However, the limitations of the scoring function may require alternative strategies for calculating DNA-ligand binding free energies. Such strategies, based on molecular dynamics simulations, are currently being developed in our laboratory.

■ ASSOCIATED CONTENT

Supporting Information

Docking results of the entire data set (RMSD values and ranks) using set 18 with the small grid box (Table S1), set 13 with the small grid box (Table S2), set 13 with the large grid box (Table S3), and set 18 with the large grid box (Table S4). Docking results into preformed canonical B-DNA structures (Table S5). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ybaime@gmail.com.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Gurova, K. New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents. *Future Oncol.* **2009**, *5*, 1685−1704.

(2) Gewirtz, D. A. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem. Pharmacol.* **1999**, *57*, 727−741.

(3) Ghosh, D.; Hossain, M.; Saha, C.; Dey, S. K.; Kumar, G. S. Intercalation and induction of strand breaks by adriamycin and daunomycin: A study with human genomic DNA. *DNA Cell Biol.* **2011**, *31*, 378−387.

(4) Hurley, L. H. DNA and its associated processes as target for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 188−200.
(5) Goldarce, R. J.; Loveless, A.; Ross, W. C. J. Mode of production of chromosome abnormalities by the nitrogen mustards the possible role of cross-linking. *Nature* **1949**, *163*, 667−669.
(6) Cepeda, V.; Fuertes, M. A.; Castilla, J.; Alonso, C.; Quevedo, C.; Pérez, J. M. Biochemical mechanisms of cisplatin cytotoxicity. *Anticancer Agents Med. Chem.* **2007**, *7*, 3−18.
(7) Fuertes, M. A.; Castilla, J.; Alonso, C.; Perez, J. M. Cisplatin biochemical mechanism of action: From cytotoxicity to induction of cell death through interconnections between apoptotic and necrotic pathways. *Curr. Med. Chem.* **2003**, *10*, 257.
(8) Alcaro, S.; Coleman, R. S. Molecular modeling of the antitumor agents azinomycins A and B: Force-field parametrization and DNA cross-linking-based filtering. *J. Org. Chem.* **1998**, *63*, 4620−4625.
(9) Alcaro, S.; Ortuso, F.; Coleman, R. S. DNA cross-linking by azinomycin B: Monte Carlo simulations in the evaluation of sequence selectivity. *J. Med. Chem.* **2002**, *45*, 861−870.
(10) Nagaoka, K.; Matsumoto, M.; Oono, J.; Yokoi, K.; Ishizeki, S.; Nakashima, T. Azinomycins A and B, new antitumor antibiotics III. Antitumor activity. *J. Antibiot.* **1986**, *39*, 1527−1532.
(11) Pabo, C. O.; Sauer, R. T. Protein-DNA recognition. *Annu. Rev. Biochem.* **1984**, *53*, 293−321.
(12) DA Ros, T.; Giampiero, S.; Prato, M.; Saison-Behmoaras, T.; Boutorine, A.; Cacciari, B. Oligonucleotides and oligonucleotide conjugates: A New approach for cancer treatment. *Curr. Med. Chem.* **2005**, *12*, 71−88.
(13) Hollis, L. S.; Sundquist, W. I.; Burstyn, J. N.; Heiger-Bernays, W. J.; Bellon, S. F.; Ahmed, K. J.; Amundsen, A. R.; Stern, E. W.; Lippard, S. J. Mechanistic studies of a novel class of trisubstituted platinum(II) antitumor agents. *Cancer Res.* **1991**, *51*, 1866−1875.
(14) Wartell, R. M.; Larson, J. E.; Wells, R. D. Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid. *J. Biol. Chem.* **1974**, *249*, 6719−6731.
(15) Morávek, Z.; Neidle, S.; Schneider, B. Protein and drug interactions in the minor groove of DNA. *Nucleic Acids Res.* **2002**, *30*, 1182−1191.
(16) Lewis, E. A.; Munde, M.; Wang, S.; Rettig, M.; Le, V.; Machha, V.; Wilson, W. D. Complexity in the binding of minor groove agents: netropsin has two thermodynamically different DNA binding modes at a single site. *Nucleic Acids Res.* **2011**, *39*, 9649−9658.
(17) Hampshire, A. J.; Fox, K. R. The effects of local DNA sequence on the interaction of ligands with their preferred binding sites. *Biochimie* **2008**, *90*, 988−998.
(18) Neidle, S. DNA minor-groove recognition by small molecules. *Nat. Prod. Rep.* **2001**, *18*, 291−309.
(19) Wilson, W. D.; Tanious, F. A.; Mathis, A.; Tevis, D.; Hall, J. E.; Boykin, D. W. Antiparasitic compounds that target DNA. *Biochimie* **2008**, *90*, 999−1014.
(20) Todd, A. K.; Adams, A.; Thorpe, J. H.; Denny, W. A.; Wakelin, L. P. G.; Cardin, C. J. Major groove binding and 'DNA-induced' fit in the intercalation of a derivative of the mixed topoisomerase I/II poison N-(2-(dimethylamino)ethyl)acridine-4-carboxamide (DACA) into DNA: X-ray structure complexed to d(CG(5-BrU)ACG)2 at 1.3-Å Resolution. *J. Med. Chem.* **1999**, *42*, 536−540.
(21) Yeh, H. J. C.; Sayer, J. M.; Liu, X.; Altieri, A. S.; Byrd, R. A.; Lakshman, M. K.; Yagi, H.; Schurter, E. J.; Gorenstein, D. G.; Jerina, D. M. NMR solution structure of a nonanucleotide duplex with a dG mismatch opposite a 10S adduct derived from trans addition of a deoxyadenosine N6-amino group to (+)-(7R,8S,9S,10R)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene: An unusual syn glycosidic torsion angle at the modified dA. *Biochemistry* **1995**, *34*, 13570−13581.
(22) Boer, D. R.; Canals, A.; Coll, M. DNA-binding drugs caught in action: the latest 3D pictures of drug-DNA complexes. *Dalton Trans.* **2009**, 399−414.
(23) Baginski, M.; Fogolari, F.; Briggs, J. M. Electrostatic and non-electrostatic contributions to the binding free energies of anthracycline antibiotics to DNA. *J. Mol. Biol.* **1997**, *274*, 253−267.
(24) Rehn, C.; Pindur, U. Model building and molecular mechanics calculations of mitoxantrone-deoxytetranucleotide complexes: Molecular foundations of DNA intercalation as cytostatic active principle. *Monatsh. Chem.* **1996**, *127*, 631−644.
(25) Waring, M. J.; Bailly, C. The purine 2-amino group as a critical recognition element for binding of small molecules to DNA. *Gene* **1994**, *149*, 69−79.
(26) Hannon, M. J. Supramolecular DNA recognition. *Chem. Soc. Rev.* **2007**, *36*, 280−295.
(27) Bauer, W.; Vinograd, J. The interaction of closed circular DNA with intercalative dyes: III. Dependence of the buoyant density upon superhelix density and base composition. *J. Mol. Biol.* **1970**, *54*, 281−298.
(28) Lerman, L. S. Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.* **1961**, *3*, 18−30.
(29) Waring, M. Variation of the supercoils in closed circular DNA by binding of antibiotics and drugs: Evidence for molecular models involving intercalation. *J. Mol. Biol.* **1970**, *54*, 247−279.
(30) Chen, K.; Adelstein, S. J.; Kassis, A. I. Molecular modeling of the interaction of iodinated Hoechst analogs with DNA: implications for new radiopharmaceutical design. *J. Mol. Struct.: THEOCHEM* **2004**, *711*, 49−56.
(31) Chen, Q.; Shafer, R. H.; Kuntz, I. D. Structure-based discovery of ligands targeted to the RNA double helix. *Biochemistry* **1997**, *36*, 11402−11407.
(32) Detering, C.; Varani, G. Validation of automated docking programs for docking and database screening against RNA drug targets. *J. Med. Chem.* **2004**, *47*, 4188−4201.
(33) Evans, D. A.; Neidle, S. Virtual screening of DNA minor groove binders. *J. Med. Chem.* **2006**, *49*, 4232−4238.
(34) Grootenhuis, P. D. J.; Kollman, P. A.; Sieble, G. L.; DesJarlais, R. L.; Kuntz, I. D. Computerized selection of potential DNA binding compounds. *Anti-Cancer Drug Des.* **1990**, *5*, 237−242.
(35) Grootenhuis, P. D. J.; Roe, D. C.; Kollman, P. A.; Kuntz, I. D. Finding potential DNA-binding compounds by using molecular shape. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 731−750.
(36) Holt, P. A.; Chaires, J. B.; Trent, J. O. Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex. *J. Chem. Inf. Model.* **2008**, *48*, 1602−1615.
(37) Kamal, A.; Vijaya Bharathi, E.; Janaki Ramaiah, M.; Dastagiri, D.; Surendranadha Reddy, J.; Viswanath, A.; Sultana, F.; Pushpavalli, S. N. C. V. L.; Pal-Bhadra, M.; Srivastava, H. K.; Narahari Sastry, G.; Juvekar, A.; Sen, S.; Zingde, S. Quinazolinone linked pyrrolo[2,1-c][1,4]-benzodiazepine (PBD) conjugates: Design, synthesis and biological evaluation as potential anticancer agents. *Bioorg. Med. Chem.* **2010**, No. 18, 526−542.
(38) Lauria, A.; Patella, C.; Ippolito, M.; Almerico, A. M. Docking and synthesis of pyrrolopyrimidodiazepinone derivatives (PPDs) and their precursors: New scaffolds for DNA-interacting agents. *J. Mol. Struct.: THEOCHEM* **2007**, *819*, 26−31.
(39) López Navarrete, J. T.; Casado, J.; Ramírez, F. J. Electronic spectroscopy study and molecular docking simulation of the interaction of terthiophene with DNA. *J. Mol. Struct.* **2007**, *834*−*836*, 176−181.
(40) Rohs, R.; Bloch, I.; Sklenar, H.; Shakked, Z. Molecular flexibility in ab initio drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations. *Nucleic Acids Res.* **2005**, *33*, 7048−7057.
(41) Sobhani, A. M.; Amini, S. R.; Tyndall, J. D. A.; Azizi, E.; Daneshtalab, M.; Khalaj, A. A theory of mode of action of azolylalkylquinolines as DNA binding agents using automated flexible ligand docking. *J. Mol. Graphics Modell.* **2006**, *25*, 459−469.
(42) Srivastava, H. K.; Chourasia, M.; Kumar, D.; Sastry, G. N. Comparison of computational methods to model DNA minor groove binders. *J. Chem. Inf. Model.* **2011**, *51*, 558−571.
(43) Tse, W. C.; Boger, D. L. Sequence-selective DNA recognition: Natural products and nature's lessons. *Chem. Biol.* **2004**, *11*, 1607−1617.

(44) Tuttle, T.; Kraka, E.; Cremer, D. Docking, triggering, and biological activity of dynemicin A in DNA: A computational study. *J. Am. Chem. Soc.* **2005**, *127*, 9469−9484.
(45) Yan, Z.; Sikri, S.; Beveridge, D. L.; Baranger, A. M. Identification of an aminoacridine derivative that binds to RNA tetraloops. *J. Med. Chem.* **2007**, *50*, 4096−4104.
(46) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7−S26.
(47) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(48) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(49) Wu, G.; Robertson, D. H.; Brooks, C. L.; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER—a CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549−1562.
(50) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(51) *Epik*, version 2.0; Schrödinger, LLC: New York, NY, 2009.
(52) Shelley, J.; Cholleti, A.; Frye, L.; Greenwood, J.; Timlin, M.; Uchimaya, M. Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681−691.
(53) *Discovery Studio Modeling Enviroment*, version 3.5; Accelrys Software Inc.: San Diego, 2012.
(54) *MacroModel*, version 9.7; Schrödinger, LLC: New York, NY, 2009.
(55) *Maestro*, Version 9.0; Schrödinger, LLC: New York, NY, 2009.
(56) *Jaguar*, Version 7.6; Schrödinger, LLC: New York, NY, 2009.
(57) Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graphics Modell.* **1999**, *17*, 57−61.
(58) Quinlan, J. R. *C4.5: programs for machine learning*; Morgan Kaufmann Publishers: San Mateo, 1993.
(59) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.* **2009**, *11*, 10−18.
(60) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(61) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(62) Ricci, C. G.; Netz, P. A. Docking studies on DNA-ligand interactions: Building and application of a protocol to identify the binding mode. *J. Chem. Inf. Model.* **2009**, *49*, 1925−1935.
(63) Dickerson, R.; Drew, H.; Conner, B.; Wing, R.; Fratini, A.; Kopka, M. The anatomy of A-, B-, and Z-DNA. *Science* **1982**, *216*, 475−485.