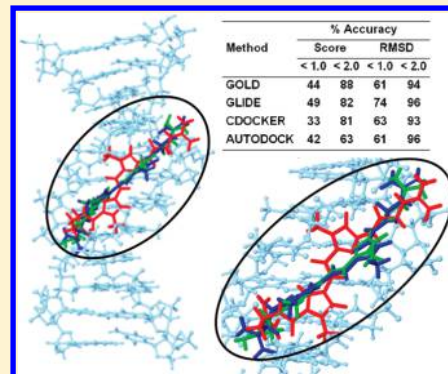ARTICLE

Comparison of Computational Methods to Model DNA Minor Groove Binders

Hemant Kumar Srivastava, Mukesh Chourasia, Devesh Kumar, and G. Narahari Sastry*

Molecular Modelling Group, Indian Institute of Chemical Technology, Taranaka, Hyderabad 500 607, India

S *Supporting Information*

ABSTRACT: There has been a profound interest in designing small molecules that interact in sequence-selective fashion with DNA minor grooves. However, most in silico approaches have not been parametrized for DNA ligand interaction. In this regard, a systematic computational analysis of 57 available PDB structures of noncovalent DNA minor groove binders has been undertaken. The study starts with a rigorous benchmarking of GOLD, GLIDE, CDOCKER, and AUTODOCK docking protocols followed by developing QSSR models and finally molecular dynamics simulations. In GOLD and GLIDE, the orientation of the best score pose is closer to the lowest rmsd pose, and the deviation in the conformation of various poses is also smaller compared to other docking protocols. Efficient QSSR models were developed with constitutional, topological, and quantum chemical descriptors on the basis of B3LYP/6-31G* optimized geometries, and with this ΔT_m values of 46 ligands were predicted. Molecular dynamics simulations of the 14 DNA—ligand complexes with Amber 8.0 show that the complexes are stable in aqueous conditions and do not undergo noticeable fluctuations during the 5 ns production run, with respect to their initial placement in the minor groove region.

Method	% Accuracy			
	Score		RMSD	
	< 1.0	< 2.0	< 1.0	< 2.0
GOLD	44	88	61	94
GLIDE	49	82	74	96
CDOCKER	33	81	63	93
AUTODOCK	42	63	61	96

INTRODUCTION

The interaction of small molecules with DNA recently has drawn the attention of several researchers because DNA is generally the prime target of anticancer drugs. These interactions can cause DNA damage in cancer cells, which inhibits the growth of cancer cells by either inhibiting replication or transcription and results in cell death or apoptosis.[1] The B-form of the DNA right-handed double-helical structure, which is most common in cells under normal conditions, is biologically relevant because of its shallow wide major groove and deep minor groove and thus has been identified as an important target for cancer therapy. Drug—DNA interactions can be classified into two major categories: intercalation and groove binding. Both covalent and noncovalent binding modes are possible in these two categories. Intercalative and nonintercalative drugs represent the major classes of non-covalent binding with DNA.[2] Intercalators are typically made of a planar polyaromatic system, which bind between the layers of nucleic acid base pairs with less sequence specificity and disrupt the organization of the DNA double-helix.[3] The naphthalimides are a family of important DNA intercalators because of their antitumor properties. The DNA alkylating agents like nitrogen mustards,[4] cis-platin,[5] etc. fall into a covalent binding category. These compounds inhibit transcription, translation, and replication of DNA and block the progression of DNA polymerase.[6] The nitrogen mustards represent the extensively studied class of the DNA interstrand cross-linking agent, and mechlorethamine and chlorambucil are two important clinical anticancer agents from this class.[7] The major and minor grooves of DNA differ in very subtle ways in terms of hydrogen bonding characteristics, steric disposition, electrostatic environment, and microenvironmental polarity.[8] This leads to contrasting preferences for the incoming molecules to bind to DNA, with proteins and large molecules preferring the major groove.[9] However, small molecules were found to bind with both the major and minor groove regions of the double-helical DNA with a slight preference toward the later mode of binding.[9] The natural product azinomycin B is an antitumor agent, and its biological activity was traced to the formation of covalent interstrand cross-linking within the major groove via N7 alkylation of two purine bases.[10] Interest in delineating the details of the molecular mechanism of action and effective antitumor activity resulted in several studies involving synthetic, mechanistic, and molecular modeling studies of azinomycin and its derivatives.[11] Table 1 and Figure 1 together depict the various classes of DNA binders, their mode of binding, PDB-ID, and type of activity along with a representative chemical structure.

Minor groove binding agents have long and planar structures that allow them to adopt a crescent shape that fits into the groove. The small molecules bind at AT-rich region in the minor groove in a sequence-selective fashion. The DNA minor groove is the target for a large number of drugs such as the antiviral agents netropsin and distamycin, the antipneumocystis carinii drugs pentamidine and furamidines, the antiseptic and disinfectant propamidine, the trypanocidal drug berenil, the experimental antitumor agent and important DNA stain Hoechst dyes.[12] Moreover, several minor groove binders proved to have potent antitumor activity in

Table 1. List of DNA Binding Drugs, Their Action, and Mode of Binding (Minor/Major Groove and Intercalators) with PDB ID

number	DNA binders	activity	mode of binding		PDB ID
1	netropsin	antitumor, antiviral	minor groove	noncovalent	121D, 195D, 1D86, 1DNE, 261D
2	hoechst	antitumor	minor groove	noncovalent	109D, 127D, 128D, 129D, 1D43, 1D44, 1D45, 1D46, 8BNA, 1DNH, 264D, 302D, 311D, 303D, 442D, 443D, 448D
3	distamycin	antitumor, antiviral	minor groove	noncovalent	2DND
4	DAPI	fluorescence	minor groove	noncovalent	1D30
5	bernil	antitrypanosomal	minor groove	noncovalent	1D63, 2DBE
6	pentamidine	against *Pneumocystis carinii*	minor groove	noncovalent	1D64
7	propamidine	antiseptic and disinfectant	minor groove	noncovalent	102D, 1PRP
8	CC1065	antitumor antibiotic	minor groove	covalent	1DSM
9	bisfuramidine	against *P. carinii*, antitumor	minor groove	noncovalent	1FMQ, 1FMS, 1EEL, 360D, 289D, 298D, 227D
10	SN7167	antitumor, antiviral	minor groove	noncovalent	328D
11	SN6999	against *P. falciparum*	minor groove	noncovalent	144D
12	mithramycin	anticancer antibiotic	minor groove	noncovalent	146D
13	plicamycin	anticancer antibiotic	minor groove	noncovalent	1BP8
14	chromomycin A3	anticancer antibiotic	minor groove	noncovalent	1EKH
15	PBDs	anticancer	minor groove	covalent	2K4L
16	cis-platin	anticancer antibiotic	cross-linking	covalent	1AU5
17	cis-{Pt(NH3)2(pyridine)}2+	anticancer	major groove	covalent	3CO3
18	aminoglycoside (NB33)	broad-spectrum antibacterial	major groove	noncovalent	454D
19	nogalamycin	antitumor	intercalation	noncovalent	182D
20	menogaril	antitumor- topoisomerase II poison	intercalation	covalent	202D

preclinical studies, and some of them are at an advanced stage of clinical trials. These include duocarmycin derivatives adozelesin,[13] carzelesin,[14] bizelesin,[15] KW-2189,[16] and a distamycin A-derivative tallimustine.[17] Distamycin A is a naturally occurring antibiotic agent isolated from *Streptomyces distallicus* that is active against a few viruses, gram-positive bacteria, and protozoa but inactive as an antitumor agent.[18] Recently, synthetic strategies have been developed to modify the distamycin scaffold to enhance the recognition efficiency together with mixed base-pair reading capacity.[18d] Pentamidine, a synthetic diamidine dication, was used against parasitic microorganisms and for the treatment of *Pneumocystis jiroveci* pneumonia (PCP) within infants.[19] A furan-based diamidine called furamidine (DB75, similar in structure to pentamidine) has shown promising therapeutic activity against a variety of parasitic microorganisms.[19] A prodrug of furamidine was in phase II clinical trials against malaria and phase III clinical trials against pneumocystis pneumonia and human African trypanosomiasis (HAT).[20] The antibiotic netropsin is a dicationic amidine—guanidine natural product that has been extremely useful in characterizing the full molecular basis for DNA minor groove recognition.[21] Bisguanidine and especially bis(2-aminoimidazoline)diphenyl compounds displayed potent antitrypanosomal activity in vitro and in vivo against *T. b. rhodesiense* (*Trypanosoma brucei rhodesiense*).[22]

Thus, it is evident that there is a tremendous interest to design the molecule with optimal binding to DNA minor groove. Lead optimization in this case requires a detail understanding of the way in which the small molecules bind to the minor groove. There are several well-established docking procedures for protein—ligand interactions;[23,24] however, in comparison, docking studies involving DNA—ligand interactions are not as popular.[25] This is because most docking methods were developed for targeting

protein—ligand interactions. Developing force field or parametrization for scoring functions is often indispensable to a new class of molecules; however, such a practice is not desirable when the existing scoring function is adequate. Neidle et al. demonstrated that the parametrization is transferable from proteins to DNA in some selected cases.[25e] To the best of our knowledge, no systematic parametrization is available for complexes involving nucleic acid ligand binding (with the exception of DOCK).[26] However, there are a large number of studies in the literature that adopt existing docking protocols to model DNA ligand complexes very effectively.[25e—h,27] Tang et al.[27] in a recent paper have made an exhaustive study on the available RNA ligand complexes and demonstrated the applicability and reliability of GOLD[28] and GLIDE[29] methods. Neidle et al.[25e] evaluated DOCK and AUTODOCK[30] protocols on 28 DNA—ligand complexes and found 40% and 55% success, respectively, in predicting crystallographic poses within a 2.0 Å root-mean-square deviation (rmsd). Recently, Trent et al.[25g] have shown that AUTODOCK and Surflex[31] can reproduce the crystal structure of DNA minor groove binders (distamycin and pentamidine) and DNA intercalators (daunorubicin and ellipticine). Clearly, while the Neidle's study reveals the poor performance of adopted docking methods, Trent's study was far from exhaustive.

It appears to be extremely important to evaluate various popular docking protocols on a reasonable number of compounds for DNA—ligand interaction studies. Here, we present a systematic computational analysis of 57 noncovalent DNA minor groove binders (available in PDB)[32] using four popular docking protocols: GOLD, GLIDE, CDOCKER,[33] and AUTODOCK (Table 2). One of the main objectives of the current paper is to address the question, "Is it necessary to develop a new scoring function for nucleic acid ligand interaction or can one use the
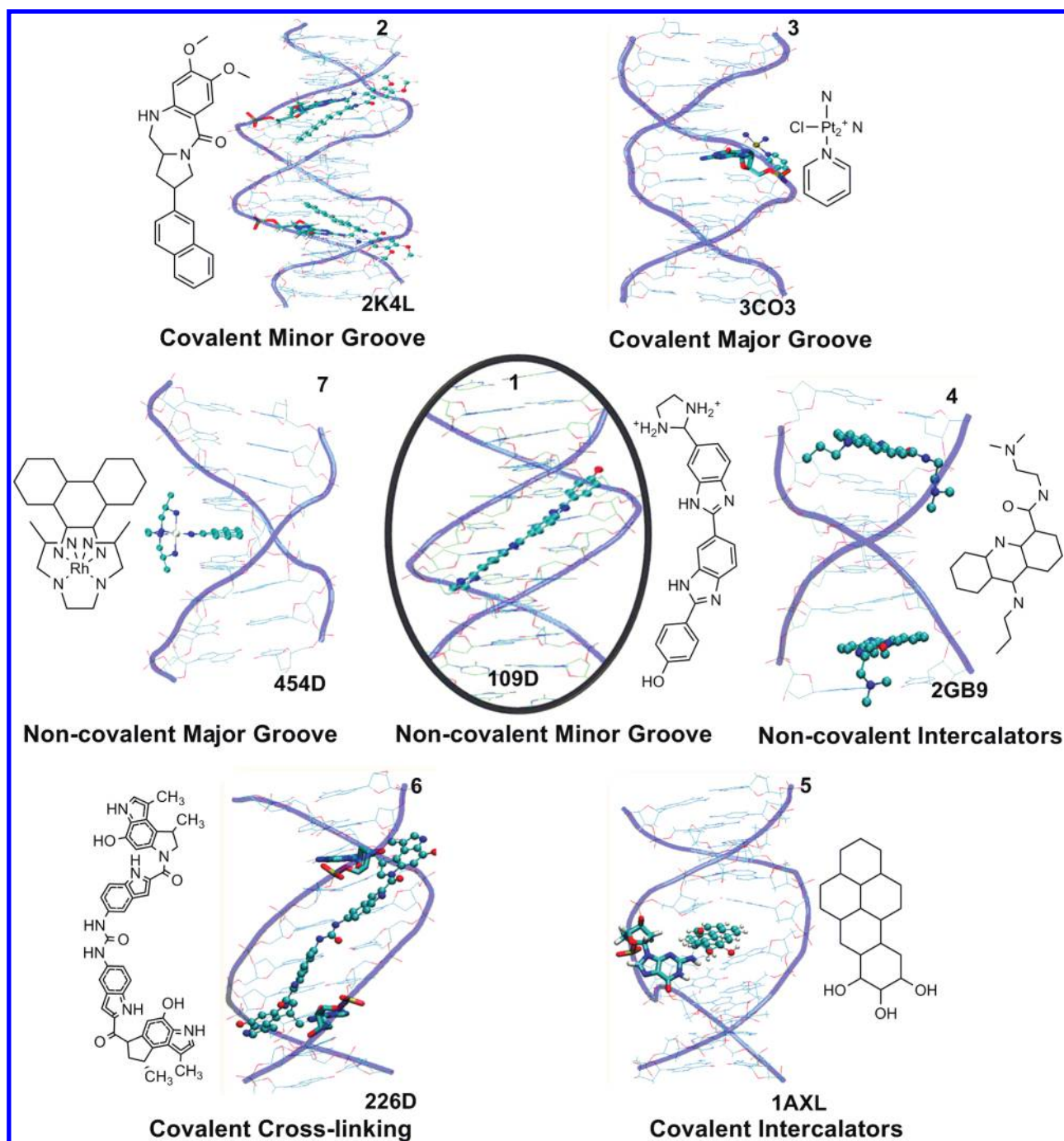
Figure 1. Representation of various classes of DNA binders and their mode of binding.

existing parameters?" Experimentally observed ΔT_m is an important parameter to measure the thermal stability of the DNA—ligand complex. The ΔT_m values of 11 available compounds (among the structures considered) were collected from the literature[34] for QSSR (quantitative structure stability relationship) studies. QSAR/QSPR/QSTR/QSSR methods are based on the axiom that the variance in the activities or physicochemical properties of chemical compounds is determined by the variance in their molecular structures.[35] QSAR has been widely used for the prediction of activities of diverse series of compounds.[36] In this work, we present a valuable method to predict ΔT_m values of DNA minor groove binders on the basis of

QSSR approaches. To address the fundamental issues relating the DNA rigidity and structural deviations upon the ligand binding, we did an extensive 5 ns molecular dynamics simulation on a set of 14 selected complexes. The MD simulation provide useful information on the time-dependent behavior, including atomic level fluctuations and conformational changes of a macromolecule.[37] MM-PBSA based interaction energies derived from the MD simulation are broadly accepted to estimate interactions between small and large molecules.[38] The interaction energy of all the complexes were calculated by using the Poisson—Boltzmann continuum electrostatic method in which the influence of the solvent is incorporated implicitly.[39]

Table 2. Structure of All Ligands with Their PDB ID and Resolution, along with rmsd values of B3LYP/6-31G* Optimized Geometries with the Crystallographic Ligands[a]
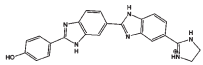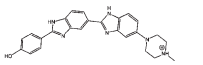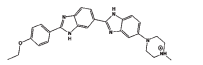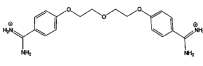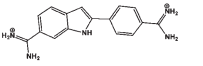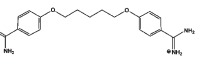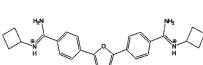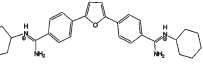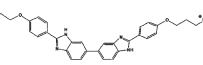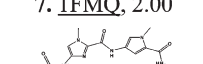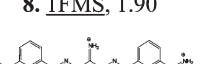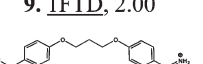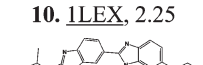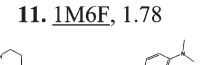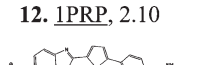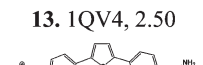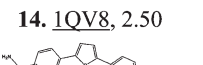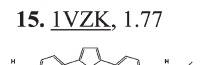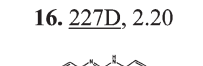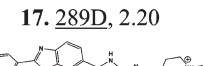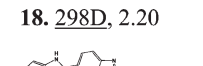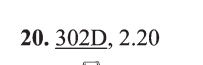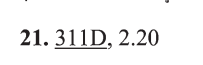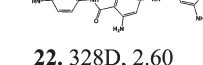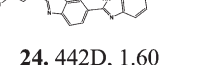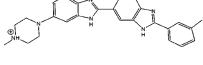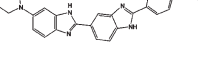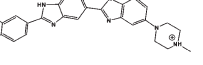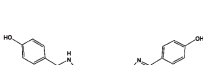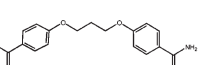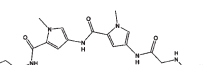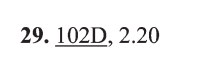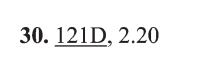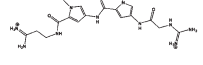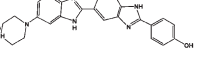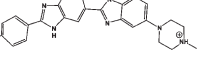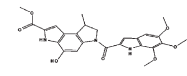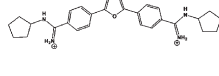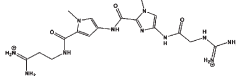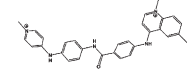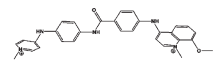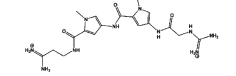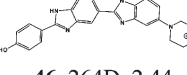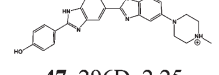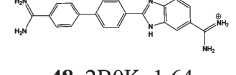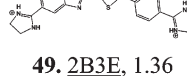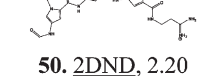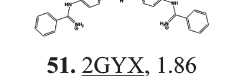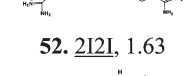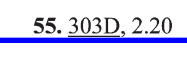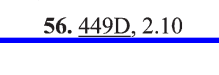
Ligand, PDB-ID and Resolution	RMSD	Ligand, PDB-ID and Resolution	RMSD	Ligand, PDB-ID and Resolution	RMSD
1. 109D, 2.00	4.09	**2.** 127D, 2.00	0.53	**3.** 129D, 2.25	0.89
4. 166D, 2.20	0.85	**5.** 1D30, 2.40	0.60	**6.** 1D64, 2.10	1.43
7. 1FMQ, 2.00	1.21	**8.** 1FMS, 1.90	1.80	**9.** 1FTD, 2.00	1.55
10. 1LEX, 2.25	1.16	**11.** 1M6F, 1.78	1.01	**12.** 1PRP, 2.10	0.86
13. 1QV4, 2.50	4.35	**14.** 1QV8, 2.50	6.14	**15.** 1VZK, 1.77	3.08
16. 227D, 2.20	0.78	**17.** 289D, 2.20	1.23	**18.** 298D, 2.20	1.04
19. 2DBE, 2.50	0.77	**20.** 302D, 2.20	1.15	**21.** 311D, 2.20	0.61
22. 328D, 2.60	2.27	**23.** 360D, 1.85	0.59	**24.** 442D, 1.60	0.53
25. 443D, 1.60	0.71	**26.** 447D, 2.20	1.14	**27.** 448D, 2.20	0.59
28. 453D, 1.80	0.26	**29.** 102D, 2.20	0.72	**30.** 121D, 2.20	1.09
31. 195D, 2.30	0.77	**32.** 1D43, 2.00	1.15	**33.** 1D44, 2.00	0.90
34. 1D45, 1.90	1.08	**35.** 1D46, 2.00	0.84	**36.** 1D63, 2.00	0.37
37. 1D86, 2.20	1.39	**38.** 1DNE, 2.40	1.74	**39.** 1DNH, 2.25	0.63

Table 2. Continued

Ligand, PDB-ID and Resolution	RMSD	Ligand, PDB-ID and Resolution	RMSD	Ligand, PDB-ID and Resolution	RMSD
40. <u>1DSA</u>, NA	0.53	**41.** <u>1EEL</u>, 2.40	1.06	**42.** <u>1LEY</u>, 2.25	0.98
43. <u>1ZPH</u>, 1.80	0.99	**44.** <u>1ZPI</u>, 1.60	1.25	**45.** <u>261D</u>, 2.40	0.88
46. <u>264D</u>, 2.44	0.59	**47.** <u>296D</u>, 2.25	0.55	**48.** <u>2B0K</u>, 1.64	0.28
49. <u>2B3E</u>, 1.36	0.47	**50.** <u>2DND</u>, 2.20	0.48	**51.** <u>2GYX</u>, 1.86	1.05
52. <u>2I2I</u>, 1.63	0.25	**53.** <u>2I5A</u>, 1.65	0.51	**54.** <u>2NLM</u>, 2.05	0.53
55. <u>303D</u>, 2.20	0.80	**56.** <u>449D</u>, 2.10	0.68	**57.** <u>8BNA</u>, 2.20	0.49

[a] Ligands from DNA sequences 5′-CGCGAATTCGCG-3′ (1-28, 32-35, 37, 39, 41-44, 48-49, 51-57), 5′-CGCAAATTTGCG-3′ (29, 30, 36, 46, 47, 50), 5′-GACTAATTGAC-3′ (40), 5′-CGCGTTAACGCG-3′ (31), 5′-CGCGATATCGCG-3′ (38), and 5′-CGCAATTCGC-3′ (45) have been collected (see references in the Supporting Information for citations of all the considered structures with PDB ID).

Thus, the current computational study rigorously examines the available popular modeling approaches for the DNA minor groove binding and compares them with experimental results where available. Finally, the paper provides important insights into the lead optimization of this class of compounds, which should be of high interest to medicinal chemists.

■ MATERIALS AND METHODS

A total of 57 crystal structures of DNA—ligand complexes were downloaded from the protein data bank (PDB). These 57 complexes are divided into six different groups on the basis of the extent of sequence similarity of DNA. Complexes **1**−**28**, **32**−**35**, **37**, **39**, **41**−**44**, **48**−**49**, and **51**−**57** have 5′-CGCGAATTCGCG-3′, complexes **29**, **30**, **36**, **46**, **47**, and **50** have 5′-CGCAAATTTGCG-3′, complex **40** has 5′-GAC-TAATTGAC-3′, complex **31** has 5′-CGCGTTAACGCG-3′, complex **38** has 5′-CGCGATATCGCG-3′, and complex **45** has 5′-CGCAATTCGC-3′ DNA sequences. Considering this, we used six different DNA sequences (PDB ID: 1VZK, 121D, 1DSA, 195D, 1DNE, and 261D) as receptors for docking of the respective group of complexes. Docking with the above-mentioned six different receptors does not show any significant variation in absolute rmsd values in comparison with that of a single receptor (see Table S13 of the Supporting Information). The four widely accepted docking protocols, GOLD, GLIDE, CDOCKER, and AUTODOCK, were considered for the docking study. The ability to reproduce a ligand pose close to the

crystallographic pose is a critical determinant for the effectiveness of the docking program. Constitutional, topological, and conceptual DFT-based descriptors were used to generate QSSR models for the prediction of ΔT_m values. Gaussian 03,[40] CODESSA,[41] and Scigress Explorer[42] programs were used for geometry optimization (B3LYP/6-31G* level), descriptor generation, and regression analysis, respectively. MD simulation was performed with the Amber 8.0[43] program, and interaction energies were calculated with MM-PBSA[39] calculations.

■ COMPUTATIONAL DETAILS

Ligands and Receptors Preparation. The ligands were extracted from DNA-bound crystal structures. Explicit hydrogens were added and were subjected to energy minimization on a Sybyl 6.9.2 (Tripos, Inc., St. Louis, MO) program by using MMFF94 force field and point charges.[44] The DNA was extracted from six considered complexes. Explicit hydrogens were added, and the energy minimization of hydrogens was carried out. Ligands were optimized by using 0.01 kcal mol^{-1}/Å, whereas hydrogens of all six receptors were optimized by using 0.1 kcal mol^{-1}/Å root-mean-squared gradient.

■ METHODOLOGY FOR DOCKING

GOLD. The entire conformational space of the DNA was scanned by selecting the central atom as the binding site for the molecular docking calculations in GOLD 3.2 software that uses

the genetic algorithm (GA) to perform flexible docking. For each of the 10 independent GA runs, a maximum number of 100,000 operations were performed with a population size of 100 individuals. Default values of niche size (2) and selection pressure (1.1) was used, and operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. Default cutoff values of 2.5 Å for hydrogen bonds and 4.0 Å for van der Waals distance were employed. The 10 best conformations were generated for each ligand without putting any constraints in the docking.

GLIDE. The Sybyl 6.9.2 generated receptors were optimized in gas phase in Maestro by using OPLS2005 force field before docking study using GLIDE 4.5. All heavy atoms were constrained to their original coordinates with the "Truncated Newton" algorithm, and 100 iterations were used with 0.01 gradient criterion. A receptor grid was generated using a 1.00 van der Waals (vdW) radius scaling factor and 0.25 partial charge cutoff. The radius of 30 Å from the central atom of DNA was defined as the binding site that covers the entire DNA. Ligands were docked using the standard precision mode without using any constraints and a 0.80 van der Waals (vdW) radius scaling factor and 0.15 partial charge cutoff. Twenty poses per ligand were generated, and post-docking minimization was carried out.

AUTODOCK. The binding site in AUTODOCK is defined by the grid of interaction points. The N3 atom of minor groove adenine was defined as the center of grid, and the grid was made up of the grid points $70(X)$, $80(Y)$, and $126(Z)$ with grid spacing of 0.375 Å. The central axis of DNA is along the Z-axis. The number of generations, energy evaluations, and individuals in the population are set to 27000, 5×10^6, and 150, respectively. The 50 deg/step was used for quaternion and torsion, while 2 Å/step was used for the translation during the docking run. The Lamarckian genetic algorithm was adopted for sampling ligand conformations in the minor groove of DNA. The default parameters of free energy scoring function were used for the docking studies. The best scoring and lowest rmsd solutions from 20 runs were considered as the predicted binding conformers. The rmsd values of each docked conformer were calculated from the corresponding crystal structure as reference to access the accuracy of poses with respect to their binding energy.

CDOCKER. CDOCKER has an all-atom CHARMm force field-based docking algorithm and uses soft-core potentials with an optional grid representation to dock ligands into active site of the receptor. The nucleotide falls within the 20 Å radius from the center atom of the DNA as defined in AUTODOCK and was considered an active site. CHARMm force field was applied to both DNA and ligands, and DNA was kept rigid during the docking study. CDOCKER generates the random ligand conformations through molecular dynamics simulation. A variable

number of rigid-body rotations/translations are applied to each conformation to generate the initial ligand poses. The final poses of the ligands were then subjected to simulated annealing to refine the poses in the DNA minor groove. A total of 10 final poses were saved, and rmsd values were also calculated for the same.

rmsd Calculations. The all-atom rmsd values were calculated by superimposing all the docked poses obtained from various docking protocols on their respective crystallographic structure using the superimposition tool of the Maestro 8.5 package. While calculating the rmsd values of the ligands, the receptor of the docked structure was fitted on the receptor of the crystal structure employing least-squares fitting.

■ MOLECULAR DYNAMICS SIMULATION

Molecular dynamics simulations were carried out using the Amber 8.0 program on 14 DNA ligand complexes. The antechamber module of the program was used to generate parameters for the cocrystal from the GAFF force field using AM1-BCC charges. The "addions" command as implemented in "xleap" was used to add the Na^+ ions explicitly to neutralize the system. The TIP3P water box of 10 Å distance from the edge atom of the DNA–ligand complex was constructed. Equilibration of the solvated complex was completed by carrying out a short energy minimization (500 steps of each steepest descent and conjugate gradient method), 50 ps of heating, and 50 ps of density equilibration, with weak restraints on the complex followed by 500 ps of constant pressure equilibration at 300 K. Application of cutoffs to include noncovalent interactions is subjective, and
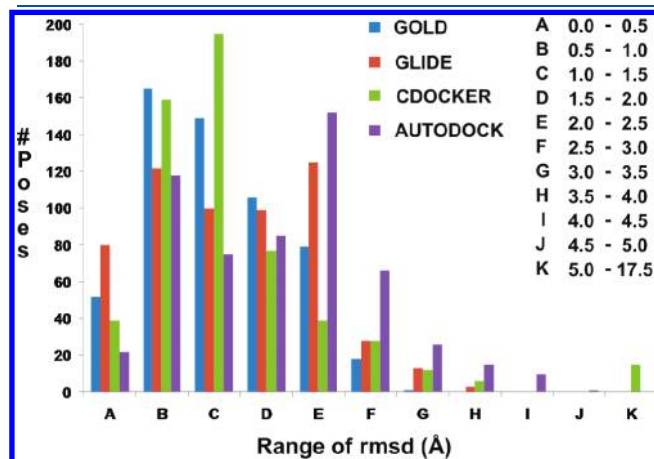
Figure 2. Comparison of rmsd values of all the generated ligand poses obtained from the GOLD, GILDE, CDOCKER, and AUTODOCK docking protocols from the crystallographic pose.

Table 3. Success Rate of Prediction of Crystallographic Poses within 2 Å rmsd by Various Docking Protocols, rmsd Variation, and Mean Average Deviations (MAD)

protocols	best score[a] (%)	lowest rmsd[b] (%)	rmsd variation[c]	MADb[d]	MADl[e]
GOLD	88	94	0.29−3.02	1.24	0.95
GLIDE	82	96	0.22−3.55	1.23	0.81
CDOCKER	81	93	0.24−17.65	1.44	0.95
AUTODOCK	63	96	0.23−4.97	1.57	0.92

[a] Success rate, while considering best score pose. [b] Success rate, while considering the pose having the lowest rmsd values. [c] The variation of rmsd values in different generated poses. [d] Average rmsd of the poses having the best score. [e] Average rmsd of the poses having the lowest rmsd from crystallographic poses.

Table 4. Regression Equations and Statistical Significance of Various QSSR Models Developed for the Prediction of ΔT_{m} Values[a]

model	regression equation	R^2	R^2_{cv}	AE	outliers
1	$\Delta T_{m} = -79.7666 \times RNSB + 1.49384 \times SIC_1 + 22.0308$	0.98	0.94	0.54	1
2	$\Delta T_{m} = 1.61315 \times BIC_1 + 84.8372 \times RNAB - 62.2725$	0.99	0.97	0.46	1
3	$\Delta T_{m} = 7.40497 \times AIC_2 - 5.80395 \times \eta - 3.64542$	0.84	0.71	1.56	0
4	$\Delta T_{m} = -71.3385 \times RNSB + 1.37832 \times SIC_1 + 19.7289$	0.95	0.87	0.76	0
5	$\Delta T_{m} = 1.47582 \times BIC_1 + 75.0745 \times RNAB - 55.1181$	0.95	0.89	0.77	0
6	$\Delta T_{m} = 0.454204 \times GOLD - 21.411$	0.74	0.53	1.82	0
7	$\Delta T_{m} = -3.35634 \times GLIDE - 25.8601$	0.57	0.56	2.52	0
8	$\Delta T_{m} = 0.159975 \times CDOCKER + 5.22858$	0.37	0.14	2.68	0
9	$\Delta T_{m} = -1.26609 \times AUTODOCK + 3.51885$	0.26	0.06	3.46	0
10	$\Delta T_{m} = 9.58562 \times AIC_1 - 16.1747$	0.72	0.59	1.95	0
11	$\Delta T_{m} = -0.83609 \times \omega + 33.4477$	0.79	0.68	2.40	1
12	$\Delta T_{m} = 7.2866 \times AIC_2 - 12.4702$	0.74	0.63	2.04	0

[a] RNSB is relative number of single bonds. SIC_1 is structural information content (order 1). BIC_1 is bonding information content (order 1). RNAB is relative number of aromatic bonds. AIC_2 is average information content (order 2). η is chemical hardness [$\eta = (E_{LUMO} - E_{HOMO})/2$]. GOLD is best fitness score obtained from GOLD docking protocols. GLIDE is best fitness score obtained from GLIDE docking protocols. CDOCKER is best fitness score obtained from CDOCKER docking protocols. AUTODOCK is best fitness score obtained from AUTODOCK docking protocols. AIC_1 is average information content (order 1). ω is electrophilicity index [$\omega = \mu^2/2\eta$, where μ is chemical potential, $\mu = (E_{LUMO} + E_{HOMO})/2$]. R^2 is correlation coefficient. R^2_{cv} is cross validation coefficient. AE is average of residuals (observed−predicted).
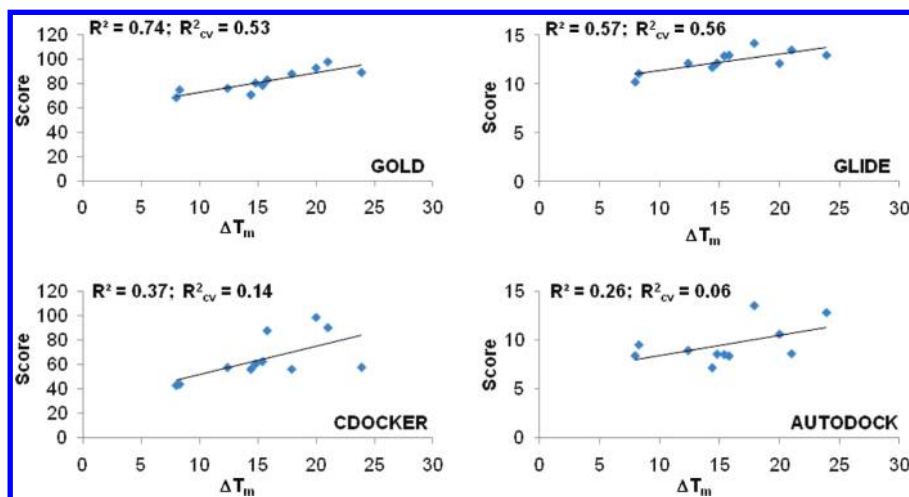
Figure 3. Comparison of various docking scores with experimental ΔT_m values.

adequate cutoff sizes are necessary to get reliable results. The larger cutoff sizes for the long-range nonbonded interactions are computationally more demanding, and a perusal of the available literature indicates that a cutoff size of 8.0 Å is more frequently used.[44] To verify the appropriate cutoff size, we performed a systematic study with 6.0, 8.0, 10.0, and 12.0 Å cutoff sizes on three different systems with 5 ns time period, and found a cutoff size of 8.0 Å is reasonable for these complexes. Therefore, a cutoff of 8.0 Å was used for MD simulations for all 14 complexes. All long-range electrostatics were included by means of a Particle−Mesh Ewald (PME) method. All hydrogen-heavy atom bonds were constrained by the SHAKE method, and simulations were performed with a 2 fs time step and langevin dynamics for temperature control. The same conditions as the final phase of equilibration are used for production run of 5 ns, and the coordinates are recorded in every 10 ps. The periodic boundary conditions (PBC) were used during MD simulations. MM-PBSA calculations were performed to extract 500 snapshots from production runs by using "extract_coords.mmpbsa" script, and

the interaction energies were calculated by using "binding_energy. mmpbsa" script (see Table S14 of the Supporting Information for the details of the parameters used for the IE calculation). The reported interaction energies are the average of all 500 snapshots.

■ RESULTS AND DISCUSSION

Docking of Noncovalent Minor Groove Binders. This study was designed to evaluate four widely accepted docking protocols for the study of 57 available DNA minor groove binders in the PDB. From the investigation of the docked poses obtained from various docking protocols, in most cases it was found that the best score pose does not correspond to the lowest rmsd pose (see Tables S6 and S7 of the Supporting Information). Initially, 10 docked poses were generated in all the docking programs. In those cases where the success rate while evaluating docking protocols was found to be less than 95% on the basis of the first 10 docked poses, the number of poses has been increased to 20. An analysis of all the docked poses (irrespective to their docking

564

dx.doi.org/10.1021/ci100474n |J. Chem. Inf. Model. 2011, 51, 558−571

Table 5. Comparison of Experimental vs Predicted ΔT_m Values of Ligands by QSSR Models of Current Study with the Available Previous Data[a]

PDB ID	Exp. ΔT_m	charge	pred. ΔT_m (J1)	pred. ΔT_m (J2)	pred. ΔT_m (M2)	pred. ΔT_m (M5)	Res. (J1)	Res. (J2)	Res. (M2)	Res. (M5)
109D	23.9	2	22.5	25.5	23.73	22.49	1.40	−1.60	0.17	1.42
2DBE	8.0	1	5.00	5.60	8.55	8.57	3.00	2.40	−0.55	−0.57
121D	20.0	2	16.8	15.5	20.11	19.74	3.20	4.50	−0.11	0.26
2DND	21.0	1	21.3	22.6	21.18	20.72	−0.30	−1.60	−0.18	0.28
227D	8.3	2	5.2	7.1	7.69	7.76	3.10	1.20	0.61	0.54
298D	14.4	2	15.3	14.6	14.70	14.51	−0.90	−0.20	−0.30	−0.11
289D	12.4	2	13.5	13.5	13.53	13.41	−1.10	−1.10	−1.13	−1.01
127D	17.9	1	19.4	20.4	20.47	20.47	−1.50	−2.50	−	−2.57
1FMQ	14.8	2	15.8	15.2	14.23	14.13	−1.00	−0.40	0.57	0.67
1EEL	15.8	2	17.1	17.1	14.84	14.75	−1.30	−1.30	0.96	1.05
1FMS	15.4	2	19.5	18.2	15.45	15.36	−4.10	−2.80	−0.05	0.04
Avg.							1.90	1.78	0.46	0.77

[a] Exp. is experimental. Charge is formal charge on ligand. J1 and J2 are predicted values collected from reference 45. M2 is QSSR model-2 and M5 is QSSR model-5 of this work. Res. is the difference in experimental and predicted ΔTm values. Avg. is the average res. M5 is without any outlier. M2 is with 127D as outlier.
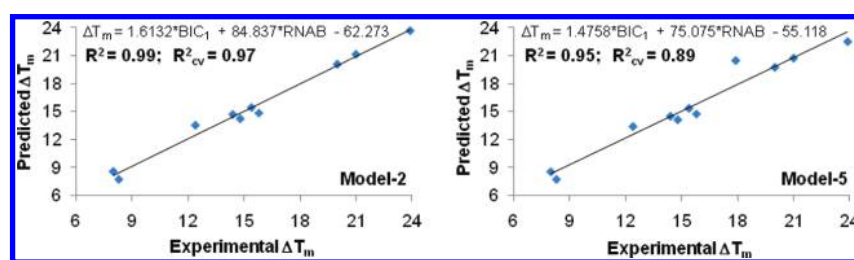
Figure 4. Plots of experimental versus predicted ΔT_m values of the DNA minor groove binders by two best QSSR models.

score) showed that GOLD, GLIDE, CDOCKER, and AUTODOCK have 30%, 63%, 36%, and 76% cases where at least one pose shows rmsd > 2.0 Å. This suggests that AUTODOCK and GLIDE generate diverse poses, whereas GOLD and CDOCKER generate similar poses. This is probably the reason for the requirement of 20 docked poses in the first two docking protocols. A success rate of 88%, 82%, 81%, and 63% has been obtained for GOLD, GLIDE, CDOCKER, and AUTODOCK protocols, respectively, with the best score pose (see Tables S1—S7of the Supporting Information). Encouragingly, the lowest rmsd score poses are within the acceptable deviations in all docking protocols employed (Table 3). Thus, all the docking protocols are able to produce at least one pose that is close to the crystallographic pose, although it may not be the one with the best score, which is reflected in most cases. In addition, all the docking protocols demonstrate similar results (GOLD 12%, GLIDE 9%, CDOCKER 11%, and AUTODOCK 14%) when we consider the cases where the best score pose and lowest rmsd pose are the same. Figure 2, which provides the graphical representation of rmsd values of all the generated poses obtained in various docking protocols, indicates that the performance of GLIDE and GOLD docking protocols is highly satisfactory. Lower rmsd variations were obtained in GOLD (0.29—3.02 Å) and GLIDE (0.22—3.55 Å) compared to those of CDOCKER (0.24—17.65 Å) and AUTODOCK (0.23—4.97 Å) as shown in Table 3. The summary of docking results in terms of mean values is also presented in Table 3. If we consider the mean average deviation in rmsd of the best score pose (MADb, the average rmsd, while considering the best score pose), GOLD (1.24 Å) and GLIDE (1.23 Å) provide better results compared to those of

CDOCKER (1.44 Å) and AUTODOCK (1.57 Å) protocols. At the same time if we consider the mean average deviation on lowest rmsd pose (MADl, the average rmsd, while considering the poses have lowest rmsd values), similar results are observed in all docking protocols. The rmsd of the B3LYP/6-31G* optimized geometry with the crystal structure of ligands were calculated and are presented in Table 2. It has been observed that 37% (>2.0 Å 30%, >3.0 Å 2%, and >4.0 Å 5%) ligands show rmsd values higher than 2.0 Å.

QSSR Analysis for the Prediction of ΔT_m Values. We made an attempt to develop QSSR models with the experimental ΔT_m values of 11 available compounds. Twelve diverse QSSR models have been generated by using two different schemes. In the first scheme, docking scores obtained from various docking protocols were used as descriptors for making four different QSSR models. In the second scheme, descriptors generated from CODESSA program along with conceptual DFT-based descriptors were used for making eight different QSSR models. Table 4 summarizes the statistical significance of all 12 QSSR models (see Table S8 and S9 of the Supporting Information for the values of various descriptors and comparison of experimental and predicted ΔT_m values and Figure S1 for the plot of experimental vs predicted values obtained from all 12 QSSR models). GOLD docking scores ($R^2 = 0.74$ and $R^2_{cv} = 0.53$) show the best correlation with ΔT_m values, while the correlation for GLIDE docking scores ($R^2 = 0.57$ and $R^2_{cv} = 0.56$) is acceptable to a certain extent. No correlations were found for CDOCKER ($R^2 = 0.37$ and $R^2_{cv} = 0.14$) and AUTODOCK ($R^2 = 0.26$ and $R^2_{cv} = 0.06$) docking scores with the ΔT_m values (Figure 3). These QSSR studies validate our observations that GOLD and GLIDE
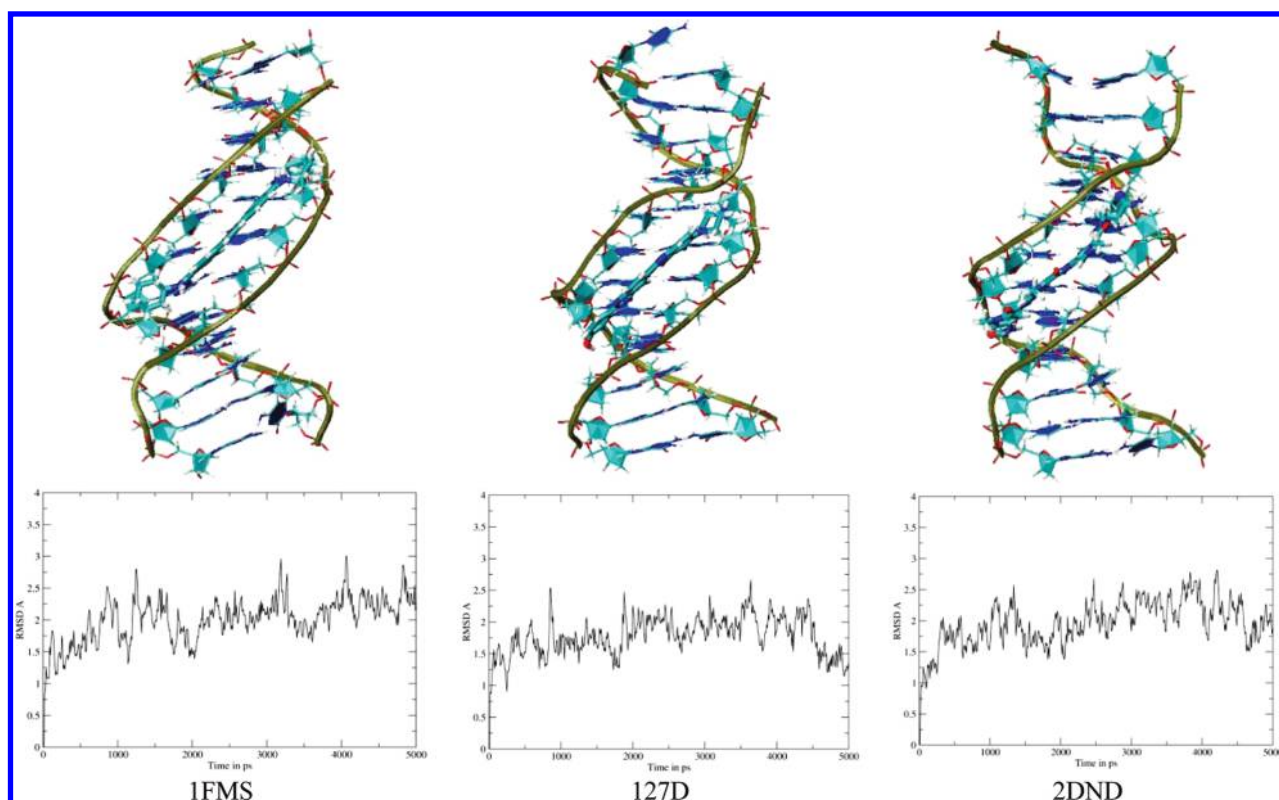
Figure 5. Snapshots of three different DNA−ligand complexes at 5 ns MD simulation, which show the highest, moderate, and lowest ΔT_m values and plots of rmsd vs time of their trajectory.

Table 6. Interaction Energies of DNA−Ligand Complexes at Various Time Scales and Comparison with Experimental ΔT_m Values

number	PDB ID	Exp. ΔT_m	IE[a] 0−1 ns	IE 1−2 ns	IE 2−3 ns	IE 3−4 ns	IE 4−5 ns	IE 0−5 ns	IE[b] 0−5 ns
1	2DBE	8.00	−43.38	−34.65	−38.09	−36.40	−35.22	−37.55	−44.77
2	227D	8.30	−36.76	−39.10	−39.54	−35.46	−36.07	−37.39	−45.57
3	289D	12.40	−39.71	−51.56	−51.51	−51.93	−47.59	−48.46	−55.01
4	298D	14.40	−51.04	−50.26	−49.42	−49.97	−51.41	−50.42	−56.71
5	1FMQ	14.80	−51.70	−51.25	−39.44	−38.60	−39.26	−44.05	−51.92
6	1FMS	15.40	−54.82	−54.08	−55.65	−51.47	−54.49	−54.10	−59.34
7	1EEL	15.80	−53.85	−51.95	−51.82	−51.45	−48.68	−51.55	−60.53
8	127D	17.90	−53.12	−54.64	−55.24	−57.46	−55.66	−55.22	−56.47
9	121D	20.00	−70.21	−71.00	−69.66	−67.65	−67.71	−69.25	−73.45
10	2DND	21.00	−63.29	−62.72	−64.79	−64.69	−62.12	−63.52	−64.74
11	109D	23.90	−47.76	−48.43	−47.14	−48.31	−47.55	−47.84	−53.61
12	1D86	99.89[c]	−70.00	−69.02	−69.13	−70.12	−68.88	−69.43	−73.29
13	1D63	76.70[c]	−36.07	−40.49	−45.01	−45.59	−45.11	−42.45	−47.20
14	1D46	36.45[c]	−45.91	−46.48	−47.01	−44.55	−41.65	−45.32	−46.23

[a] IE (interaction energy) is the final estimated binding free energy calculated by the MM-PBSA method at various time intervals. [b] Final estimated binding free energy calculated by the MM-GBSA method. [c] GOLD docking fitness score.

docking protocols are good for DNA−ligand complexes and provide reliable scores and poses. QSSR model-5 is the best model with correlation coefficient (R^2) of 0.95 and cross-validation coefficient (R^2_{cv}) of 0.89 along with an average residual (AE) of 0.77. Only two descriptors, BIC_1 (topological descriptor, bonding information content, order 1) and RNAB (constitutional descriptor, relative number of aromatic bonds), were used for making QSSR model-5. QSSR model-2 is the modification of model-5 by taking out one outlier. This model has very good regression summary with $R^2 = 0.99$, $R^2_{cv} = 0.97$, and AE = 0.46 only. The experimental and predicted ΔT_m values by using QSSR model-2 and model-5 are given in Table 5 along

with the residual values. The plot of experimental versus predicted ΔT_m values show the linear relationship (Figure 4). Table 4 clearly shows that the topological descriptors (SIC_1, AIC_1, and AIC_2) and conceptual DFT-based descriptors (η and ω) can also be used in different combinations for making valuable QSSR models. Some properties like ω, AIC_1, and AIC_2 even as a single descriptor provide good regression coefficients, but the average residual is on the higher side (models 10, 11, and 12, respectively). A systematic, swift all-atom energy-based computational protocol for the prediction of ΔT_m values has been presented earlier.[45] Clearly, the QSAR equations derived here are robust and have a high predictive ability compared to the earlier studies, as evidenced from the statistical parameters (Table 5). On the basis of three selected QSSR models, the ΔT_m values for the remaining 46 complexes used in this study are predicted (see Table S10 of the Supporting Information). We found that the ligands of 1FTD, 328D, and 1ZPI complexes may bind very strongly to the minor groove of DNA as these ligands have highest predicted ΔT_m values.

Molecular Dynamics Simulation. Following the validation of various docking protocols on 57 DNA–ligand complexes and QSSR analysis for the prediction of ΔT_m values, more rigorous MD simulations were performed on 14 selected complexes. These complexes were chosen on the basis of the availability of experimental ΔT_m values (11 complexes) and variation of GOLD scores (highest, moderate, and lowest score). MD simulations with explicit solvent involving around 8000 water molecules have been performed up to 5 ns time period. Snapshots at 5 ns for the selected complexes are presented in Figure 5, along with the rmsd variations during the simulation (see Figures S2 and S3 of the Supporting Information for rmsd values and snapshots at various time scales, respectively, for all the complexes considered for MD simulations). It is clear from Figure 5 that the ligands remain bound to the DNA near the preferential binding position and do not experience considerable deviations with respect to their initial placement in the minor groove region of DNA. A separate plot of rmsd versus time interval for each complex, DNA, and ligand (see Figure S2 of the Supporting Information for details) also indicate that all the complexes are

stable and do not show any unusual fluctuations during the simulation. It has been observed that 1FMQ and 109D complexes show higher rmsd (~3.5 Å) variations. In the case of the 1FMQ complex, both DNA (3.5 Å) and ligand (2.0 Å) are responsible for the higher rmsd variation, while in the case of 109D complex, ligand (0.5 Å) is the most stable structure compared to all the complexes, and only DNA (3.5 Å) is responsible for the higher rmsd variation. Except for the 1FMQ and 109D complexes, all other complexes show rmsd variation less than 3.0 Å during a 5 ns production run. Further, there is no considerable deviation in the rmsd of complexes and DNA with an increase in time period. If we compare the rmsd variation of ligands separately, ligands for 1FMQ, 1FMS, and

Table 7. Comparison of Docking Scores with Binding Free Energies of DNA–Ligand Complexes That Have Experimental ΔT_m Values and Ligands, Showing Low, Moderate, and High GOLD Fitness Scores

| PDB ID | Exp. ΔT_m | docking score | | | | |
		GOLD	GLIDE	CDOCKER	AUTODOCK	IE[a]
227D	8.3	74.52	−11.08	43.68	−9.53	−37.39
2DBE	8.0	68.05	−10.23	42.79	−8.40	−37.55
1D63	_	76.70	−10.16	70.28	−7.33	−42.45
1FMQ	14.8	80.13	−12.16	60.57	−8.56	−44.05
1D46	_	36.45	−8.21	77.08	−15.91	−45.32
109D	23.9	88.95	−12.97	57.75	−12.88	−47.84
289D	12.4	75.87	−12.15	57.48	−8.95	−48.46
298D	14.4	70.44	−11.72	56.10	−7.15	−50.42
1EEL	15.8	82.78	−12.97	88.15	−8.37	−51.55
1FMS	15.4	78.22	−12.89	62.54	−8.53	−54.10
127D	17.9	87.71	−14.18	56.16	−13.59	−55.22
2DND	21.0	97.73	−13.50	90.60	−8.61	−63.52
121D	20.0	92.60	−12.12	99.20	−10.64	−69.25
1D86	_	99.89	−12.13	99.74	−9.53	−69.43

[a] IE (interaction energy) is the final estimated binding free energy calculated by MM-PBSA method.
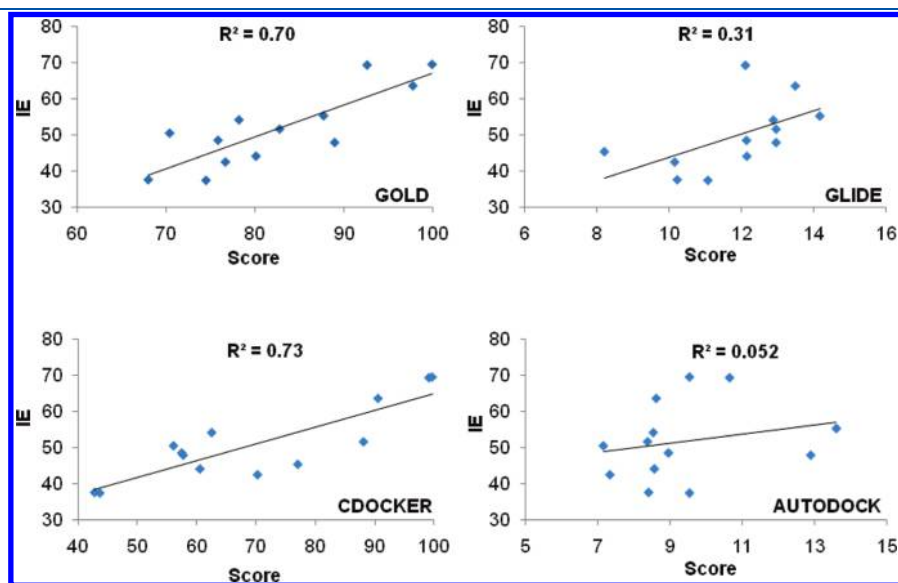

Figure 6. Regression analysis of docking scores of DNA minor groove binders obtained from various docking protocols with their IE (interaction energy) obtained from the MM-PBSA method.

Table 8. Interaction Eenergy (IE) Deviation Based on Different Cutoff Values Used in the Molecular Dynamics Simulation at 5 ns Time and Average IE Deviation for Three Complexes Considered for the Study

PDB ID	cutoff Å/IE GB				cutoff Å/IE PB			
	6	8	10	12	6	8	10	12
127D	54.29	56.47	58.55	57.53	51.51	55.22	57.49	55.67
298D	52.36	56.71	56.02	54.75	45.24	50.42	50.84	48.72
1FMQ	49.23	51.92	51.90	50.28	40.90	44.05	44.15	43.49
Avg.	51.96	55.03	55.49	54.19	45.88	49.90	50.83	49.29

127D complexes show higher rmsd variation ($\sim$2.0 Å), while other ligands show rmsd variation less than 2.0 Å.

Interaction Energy of DNA–Ligand Complexes. The IE values as obtained from MM-PBSA calculations are presented in Table 6 with the respective ΔT_m values. Higher deviation in IE values with respect to time was observed for ligands having lower ΔT_m values (<15.00). Ligands with higher ΔT_m values show consistent interaction until the 5 ns period without much deviation in the IE values. Table 6 clearly shows that the ligands with the lowest ΔT_m (2DBE, 227D complexes; ΔT_m < 10.0) have lowest IE values ($\sim$−37.0 kcal/mol), the ligands with moderate ΔT_m (289D, 298D, 1FMQ, 1FMS, 1EEL and 127D complexes; ΔT_m 10.0 to 20.0) have IE values in the range of −50.0 to −60.0 kcal/mol, and the ligands with the highest ΔT_m (121D, 2DND complexes; ΔT_m > 20.0) have highest IE values ($\sim$−70.0 kcal/mol). However, the ligand of the 109D complex is the only exception with ΔT_m = 23.9, and IE = −47.84 kcal/mol. The reason for the exceptional behavior of the 109D complex may be a consequence of a higher rmsd (3.5 Å during simulation in case of DNA; see Figure S2 of the Supporting Information) compared to those of all other ligands. However, the ligand of the 109D complex shows the lowest rmsd (0.5 Å), which correlates well with the highest ΔT_m of this compound. We have observed a moderate correlation between IE values and docking score for GOLD (R^2 = 0.70) and CDOCKER (R^2 = 0.73) protocols, where a docking score >90.00 is always associated with IE >−62.0 kcal/mol, though such a correlation is not observed in the case of GLIDE (R^2 = 0.31) and AUTODOCK (R^2 = 0.052) protocols (Table 7). Moreover a GOLD docking score of <80.00 is always associated with the IE values <−51.0 kcal/mol. An exception to this behavior is observed in the case of 109D complex probably because of the highest rmsd during simulation in the case of DNA (see Figure S2 of the Supporting Information). From the above discussion and Figure 6, it is clear that the GOLD and CDOCKER docking score correlates better than other docking scores with the IE values obtained from MM-PBSA calculations. The effect of various cutoff sizes (6.0, 8.0, 10.0, and 12.0 Å) on the IE values were also studied on three different DNA–ligand complexes (127D, 298D, 1FMQ) between the 0 and 5 ns time period. The deviation in IE values for the considered complexes with average values at 5 ns time is presented in Table 8 (see Table S12 of the Supporting Information for more details). Higher deviations in IE values were observed between 6.0 to 8.0 Å cutoff values (average deviation in IE value is 4.02 and 3.07 kcal/mol in PB and GB methods, respectively). However, the deviation in IE values is marginal from 8.0 Å and onward cutoff values as apparent from Table 8.

■ CONCLUSIONS

DNA minor groove binders are an important class of compounds with high therapeutic potential, and this class of compounds provided several FDA approved drugs and offers a great potential for structure-based computer-aided drug design methods. The present comprehensive computational study with 57 DNA minor groove binders reveals the strength and limitations of popular docking protocols. This study shows that GOLD and GLIDE docking protocols seem very reliable in modeling nucleic acid ligand complexes. Molecular dynamics studies suggest that the DNA duplex skeleton undergoes only a minor deviation upon ligand binding in most cases, and this provides the basis for employing the docking protocols in which the receptor is kept rigid. Therefore, the current exhaustive study consolidates the existing evidence to apply the currently available docking protocols to nucleic acid ligand binding, albeit they are not specifically parametrized for nucleic acids. While the studies are directed to the DNA minor groove noncovalent binders, further validation is required before generalization to intercalators, covalent binders, etc. Surprisingly, ΔT_m values, which are a measure of duplex stability, have shown good correlation with docking fitness scores and MM-PBSA based interaction energies. Considering that similar good correlations are reported earlier between the ΔT_m and binding affinity, it seems that the stronger binding of ligands to DNA impart higher stability to the duplex. The MM-PBSA based interaction energies calculated from the MD simulations are in good agreement with the experimental ΔT_m values and the GOLD docking fitness scores. While we observe subtle deviations in the backbone and cleft of DNA upon ligand binding, in most cases (except for the 109D complex) the dynamical nature of DNA does not affect the binding strength. Our attempts to build QSSR models based on the 11 available experimental ΔT_m values yielded robust statistical models using one (R^2 = 0.79; R^2_{cv} = 0.68) and two (R^2 = 0.99; R^2_{cv} = 0.97) descriptors. Considering the high pharmaceutical relevance of the minor groove binders and also the range of flexible options available for lead optimization, we feel that this will be a fertile ground for the interaction between molecular modeling and medicinal chemists. The present study illustrates and validates the performance of docking, molecular dynamics, and structure stability relationship approaches that should trigger the interplay further.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. The rmsd values corresponding to various generated poses for each of the cocrystals from GOLD3.2, GLIDE4.5, CDOCKER, and AUTODOCK docking protocols. Lowest and highest docking score from all the docking protocols. Descriptors and their values used to generate QSSR models. Experimental and predicted ΔT_m values of different QSSR models. Comparison of docking score with MD interaction energies and ΔT_m values. DNA sequence, DOI, NDB ID, PubMed ID, and chemical name of all the ligands. Interaction energy deviation on the basis of different cutoff values at various time scale for three DNA ligand complexes. Variation in lowest and highest rmsd values with respect to receptor. Details of parameters used for snapshot extraction and IE calculation. Plot of predicted versus experimental ΔT_m values. The rmsd variations with time in MD for complex, DNA, and ligand separately. Snapshots of the complexes at various time scales in the simulation process. Structural variations in the ligands considered for QSSR analysis.

References for all the PDB codes used in the study. This material is available free of charge via Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +91 40 27193016; fax: +91 40 27160512; e-mail: gnsastry@gmail.com.

■ ABBREVIATIONS

QM = quantum mechanical; QSSR = quantitative structure stability relationship; rmsd = root mean squared deviation, which is the measure of the average distance between the backbones of superimposed structures; MM-PBSA/GBSA = a postprocessing method to evaluate binding/interaction energies by using Poisson−Boltzmann or Generalized-Born methods; ΔT_m = the change in melting temperature of DNA upon drug binding (a standard for duplex stability); QSPR = quantitative structure property relationship; MD = molecular dynamics; Success rate = the rate of success in getting rmsd less than 2.0 Å, while superimposing docked poses obtained from various docking protocols on the crystallographic poses; IE (interaction energy) = the final estimated binding free energy calculated by MM-PBSA method after molecular dynamics simulations; CODESSA (comprehensive descriptors for structural and statistical analysis) = a software for descriptor calculation, regression and QSAR analysis; MMFF = Merck molecular force field (first published force field in this family is MMFF94); OPLS = optimized potential for liquid simulations (force field); GOLD = genetic optimization for ligand docking; GLIDE= a complete solution for ligand−receptor docking; CDOCKER = a CHARMm-based MD docking algorithm; AUTODOCK = suite of automated docking tools; PDB = protein data bank

■ REFERENCES

(1) (a) Song, Y.-M.; Wu, Q.; Yang, P.-J.; Luan, N.-N.; Wang, L.-F.; Liu, Y.-M. DNA Binding and cleavage activity of Ni(II) complex with all-trans retinoic acid. *J. Inorg. Biochem.* **2006**, *100*, 1685–1691. (b) Kožurková, M.; Sabolová, D.; Janovec, L.; Mikeš, J.; Koval', J.; Un-gvarskcy', J.; Stefanisinová, M.; Fedorocko, P.; Kristian, P.; Imrich, J. Cytotoxic activity of proflavine diureas: Synthesis, antitumor, evaluation and DNA binding properties of 1′,1′′-(acridin-3,6-diyl)-3′,3′′-dialkyl-diureas. *Bioorg. Med. Chem.* **2008**, *16*, 3976–3984. (c) Reddy, B. S. P.; Sondhi, S. M.; Lown, J. W. Synthetic DNA minor groove-binding drugs. *Pharmacol. Ther.* **1999**, *84*, 1–111.

(2) Neidle, S. The molecular basis for the action of some DNA-binding drugs. *Prog. Med. Chem.* **1979**, *16*, 151–221.

(3) Hurley, L. H. DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 188–200.

(4) Bauer, G. B.; Povrik, L. F. Specificity and kinetics of interstrand and intrastrand bifunctional alkylation by nitrogen mustards at a G-G-C sequence. *Nucleic Acids Res.* **1997**, *25*, 1211–1218.

(5) Wang, K.; Lu, J.; Li, R. The events that occur when cisplatin encounters cells. *Coord. Chem. Rev.* **1996**, *151*, 53–88.

(6) (a) Masta, A.; Gray, P. J.; Phillips, D. R. Nitrogen mustard inhibits transcription and translation in a cell free system. *Nucleic Acids Res.* **1995**, *23*, 3508–3515. (b) Fuertes, M. A.; Castilla, J.; Alonso, C.; Perez, J. M. Novel concepts in the development of platinum antitumor drugs. *Curr. Med. Chem.: Anti-Cancer Agents* **2002**, *2*, 539–551.

(7) Haskel, C. M. *Cancer Treatment*, 2nd ed.; Saunders: Philadelphia, PA, 1990.

(8) Blackburn, G. M. DNA and RNA Structure. In *Nucleic Acids in Chemistry and Biology*, 2nd ed.; Blackburn, G. M., Gait, M. J., Eds.; Oxford University Press: New York, 1996; pp 15−81

(9) Morávek, Z; Neidle, S; Schneider, B. Protein and drug interactions in the minor groove of DNA. *Nucleic Acids Res.* **2002**, *30*, 1182–1191.

(10) (a) Nagaoka, K.; Matsumoto, M.; Oono, J.; Yokoi, K.; Ishizeki, S.; Nakashima, T.; Azinomycins, A; New Antitumor, B. Antibiotics. I. Producing organism, fermentation, isolation and characterization. *J. Antibiot.* **1986**, *39*, 1527–1532. (b) Hodgkinson, T. J.; Shipman, M. Chemical synthesis and mode of action of the azinomycins. *Tetrahedron* **2001**, *57*, 4467–4488. (c) Rajski, S. R.; Williams, R. M. DNA cross-linking agents as antitumor drugs. *Chem. Rev.* **1998**, *98*, 2723–2795.

(11) (a) Fujiwara, T.; Saito, I.; Sugiyama, H. Highly efficient DNA interstrand cross-linking induced by an antitumor antibiotic, carzino-philin. *Tetrahedron Lett.* **1999**, *40*, 315–318. (b) Coleman, R. S. Issues of orthogonality and stability: Synthesis of the densely functionalized heterocyclic ring system of the antitumor agents azinomycins A and B. *Synlett* **1998**, *10*, 1031–1039. (c) Coleman, R. S.; Li, J.; Navarro, A. Total synthesis of azinomycin A. *Angew. Chem., Int. Ed.* **2001**, *40*, 1736–1739. (d) Coleman, R. S.; Kong, J. S.; Richardson, T. E. Synthesis of naturally occurring antitumor agents: Stereocontrolled synthesis of the azabicyclic ring system of the azinomycins. *J. Am. Chem. Soc.* **1999**, *121*, 9088–9095. (e) Alcaro, S.; Ortuso, F.; Coleman, R. S. Molecular modeling of DNA cross-linking analogues based on the azinomycin scaffold. *J. Chem. Inf. Model.* **2005**, *45*, 602–609.

(12) (a) Henderson, D.; Hurley, L. H. Molecular struggle for transcriptional control. *Nature Med.* **1995**, *1*, 525–527. (b) Thurston, D. E. Nucleic acid targeting: Therapeutic strategies for the 21st century. *Br. J. Cancer* **1999**, *80*, 65–85. (c) Marchini, S.; Ciro, M; Gallinari, F.; Geroni, C.; Cozzi, P.; D'Incalci, M.; Broggini, M. α-Bromoacryloyl derivative of distamycin A (PNU 151807): A new non-covalent minor groove DNA binder with antineoplastic activity. *Br. J. Cancer* **1999**, *80*, 991–997.

(13) (a) Burris, H. A.; Dieras, V. C.; Tunca, M.; Earhart, R. H.; Eckardt, J. R.; Rodriguez, G. I.; Shaffer, D. S.; Fields, S. M.; Campbell, E.; Schaaf, L.; Kasunic, D.; Von Hoff, D. D. Phase I study with the DNA sequence-specific agent adozelesin. *Anticancer Drugs.* **1997**, *8*, 588–596. (b) Foster, B. J.; LoRusso, P. M.; Poplin, E.; Zalupski, M.; Valdivieso, M.; Wozniak, A.; Flaherty, L.; Kasunic, D. A.; Earhart, R. H.; Baker, L. H. Phase I trial of adozelesin using the treatment schedule of daily 35 every 3 weeks. *Invest. New Drugs.* **1996**, *13*, 321–326.

(14) (a) Li, L. H.; DeKoning, T. F.; Kelly, R. C.; Krueger, W. C.; McGovren, J. P.; Padbury, G. E.; Petzold, G. L.; Wallace, T. L.; Ouding, R. J.; Prairie, M. D.; Gebhard, I. Cytotoxicity and antitumor activity of carzelesin, a prodrug cyclopropylpyrroloindole analogue. *Cancer Res.* **1992**, *52*, 4904–4913. (b) Houghton, P. J.; Cheshire, P. J.; Hallman, J. D. N.; Houghton, J. A. Therapeutic efficacy of the cyclopropylpyrro-loindole, carzelesin, against xenografts derived from adult and childhood solid tumors. *Cancer Chemother. Pharmacol.* **1995**, *36*, 45–52.

(15) (a) Carter, C. A.; Waud, W. R.; Li, L. H.; DeKoning, T. F.; McGovren, J. P.; Plowman, J. Preclinical antitumor activity of bizelesin in mice. *Clin. Cancer Res.* **1996**, *2*, 1143–1149. (b) Walker, D. L.; Reid, J. M.; Ames, M. M. Preclinical pharmacology of bizelesin, a potent bifunctional analog of the DNA-binding antibiotic CC-1065. *Cancer Chemother. Pharmacol.* **1994**, *34*, 317–322.

(16) Kobayashi, E.; Okamoto, A.; Asada, M.; Okabe, M.; Nagamura, S.; Asai, A.; Saito, H.; Gomi, K.; Hirata, T. Characteristics of antitumor activity of KW-2189, a novel water-soluble derivative of duocarmycin, against murine and human tumors. *Cancer Res.* **1994**, *54*, 2404–2410.

(17) Pezzoni, G.; Grandi, M.; Biasoli, G.; Capolongo, L.; Ballinari, D.; Giuliani, F. C.; Barbieri, B.; Pastori, A.; Pesenti, E.; Mongelli, N.;

Spreafico, F. Biological profile of FCE 24517, a novel benzoyl mustard analogue of distamycin A. *Br. J. Cancer* **1991**, *64*, 1047–1050.

(18) (a) Yang, F.; Belitsky, J. M.; Villanueva, R. A.; Dervan, P. B.; Roth, M. J. Inhibition of Moloney murine leukemia virus integration using polyamides targeting the long-terminal repeat sequences. *Biochemistry* **2003**, *42*, 6249–6258. (b) Dyatkina, N. B.; Roberts, C. D.; Keicher, J. D.; Dai, Y.; Nadherny, J. P.; Zhang, W.; Schmitz, U.; Kongpachith, A.; Fung, K.; Novikov, A. A.; Lou, L.; Vellingan, M.; Khorlin, A. A.; Chan, M. S. Minor groove DNA binders as antimicrobial agents. 1. Pyrrole tetraamides are potent antibacterials against vancomycin resistant *Enteroccoci* and methicillin resistant *Staphylococcus aureus*. *J. Med. Chem.* **2002**, *45*, 805–817. (c) Broyles, S. S.; Kremer, M.; Knutson, B. A. Antiviral activity of distamycin A against vaccinia virus is the result of inhibition of postreplicative mRNA synthesis. *J. Virol.* **2004**, *78*, 2137–2141. (d) Ghosh, S.; Usharani, D.; Paul, A.; De, S.; Jemmis, E. D.; Bhattacharya, S. Design, synthesis, and DNA binding properties of photoisomerizable azobenzene—distamycin conjugates: An experimental and computational study. *Bioconjugate Chem.* **2008**, *19*, 2332–2345.

(19) Tidwell, R. R.; Boykin, D. W. Dicationic, D. N. A. Minor groove binders as antimicrobial agents. *Small Molecule DNA and RNA Binders: From Synthesis to Nucleic Acid Complexes*; Wiley-VCH: New York, 2003; pp 416–460.

(20) (a) Fairlamb, A. H. Chemotherapy of human African trypanosomiasis: Current and future prospects. *Trends Parasitol.* **2003**, *19*, 488–494. (b) Bouteille, B.; Oukem, O.; Bisser, S.; Dumas, M. Treatment perspectives for human African trypanosomiasis. *Fundam. Clin. Pharmacol.* **2003**, *17*, 171–181. (c) Yeramian, P.; Meshnick, S. R.; Krudood, S.; Chalermurut, K.; Silachamroon, U.; Tangukdee, N.; Allen, J.; Brun, R.; Kweik, J. J.; Tidwell, R. R.; Looareesuwan, S. Efficacy of DB289 in Thai patients with *Plasmodium* vivax or acute, uncomplicated *Plasmodium falciparum* infections. *J. Infect. Dis.* **2005**, *192*, 319–322.

(21) (a) Wemmer, D. E. Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439–461. (b) Wemmer, D. E. Ligands recognizing the minor groove of DNA: Development and applications. *Biopolymers* **1999**, *52*, 197–211. (c) Neidle, S. DNA minor-groove recognition by small molecules. *Nat. Prod. Rep.* **2001**, *18*, 291–309. (d) Bailly, C.; Chaires, J. B. Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. *Bioconjugate Chem.* **1998**, *9*, 513–538.

(22) (a) Dardonville, C.; Barrett, M. P.; Brun, R.; Kaiser, M.; Tanious, F.; Wilson, W. D. DNA binding affinity of bisguanidine and bis(2-aminoimidazoline) derivatives with in vivo antitrypanosomal activity. *J. Med. Chem.* **2006**, *49*, 3748–3752. (b) Dardonville, C.; Brun, R. Bisguanidine, bis(2-aminoimidazoline), and polyamine derivatives as potent and selective chemotherapeutic agents against *Trypanosoma brucei rhodesiense*. Synthesis and in vitro evaluation. *J. Med. Chem.* **2004**, *47*, 2296–2307.

(23) (a) 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. (b) Sternberg, M. J. E.; Gabb, H. A.; Jackson, R. M. Predictive docking of protein—protein and protein—DNA complexes. *Curr. Opin. Struct. Biol.* **1998**, *8*, 250–256.

(24) (a) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759–4767. (b) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303. (c) Warren, G.; Andrews, C.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M.; Lindvall, M.; Nevins, N.; Semus, S.; Senger, S.; Tedesco, G.; Wall, I.; Woolven, J.; Peishoff, C.; Head, M. A. Critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931. (d) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. Virtual screening in drug discovery: A computational perspective. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–351.

(25) (a) Grootenhuis, P. D. J.; Kollman, P. A.; Seibel, G. L.; DesJarlais, R. L.; Kuntz, I. D. Computerized selection of potential DNA binding compounds. *Anti-Cancer Drug Des.* **1990**, *5*, 237–242.

(b) 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. (c) 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. (d) Tuttle, T.; Kraka, E.; Cremer, D. J. Docking, triggering and biological activity of dynemicin A in DNA: A computational study. *J. Am. Chem. Soc.* **2005**, *127*, 9469–9484. (e) Evans, D. A.; Neidle, S. Virtual screening of DNA minor groove binders. *J. Med. Chem.* **2006**, *39*, 4232–4238. (f) 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. (g) 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. (h) Kamal, A.; Bharathi, E. V.; Ramaiah, M. J.; Dastagiri, D.; Reddy, J. S.; Viswanath, A.; Sultana, F.; Pushpavalli, S. N. C. V. L.; Pal-Bhadra, M.; Srivastava, H. K.; Sastry, G. N.; 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**, *18*, 526–542.

(26) DOCK 5.1.1 Manual, 2003. http://dock.compbio.ucsf.edu.

(27) Li, Y.; Shen, J.; Sun, X.; Li, W.; Liu, G.; Tang, Y. *J. Chem. Inf. Model.* **2010**, *50*, 1134–1146.

(28) 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.

(29) 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.

(30) Morris, G.; Goodsell, D.; Halliday, R.; Huey, R.; Hart, W.; Belew, R.; Olson, A. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(31) *Surflex*, version 2.11; Tripos, Inc.: St. Louis, MO, 2007.

(32) PDB Web site, 2009. http://www.rcsb.org/pdb/.

(33) Wu, G.; Roberston, D. H.; Brooks, C. L., III; 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*, 549–562.

(34) (a) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373. (b) Warren, G. L.; Andrews, C. W.; Capelli, A; Clarke, B; LaLonde, J; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M; Peishoff, C. E; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931. (c) Ge, W.; Schneider, B.; Olson, W. K. Knowledge-based elastic potentials for docking drugs or proteins with nucleic acids. *Biophys. J.* **2005**, *88*, 1166–1190. (d) Shaikh, S. A.; Ahmed, S. R.; Jayaram, B. A molecular thermodynamic view of DNA-drug interactions: A case study of 25 minor-groove binders. *Arch. Biochem. Biophys.* **2004**, *429*, 81–99. (e) Beveridge, D. L.; McConnell, K. J. Nucleic acids: Theory and computer simulation, Y2K. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182–196. (f) Orozco, M.; Perez, A.; Agnes, N.; Javier, L. F. Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.* **2003**, *6*, 350–364. (g) Cheatham, T. E. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360–367. (h) Young, M. A.; Jayaram, B.; Beveridge, D. L. Local dielectric environment of B-DNA in solution: Results from a 14 ns molecular dynamics trajectory. *J. Phys. Chem. B.* **1998**, *102*, 7666–7669. (i) Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **1978**, *11*, 179–246. (j) Cornell, W. D.; Cieplak, P.; Bayly, C. F.; Gould, I. R.; Kenneth, M. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second-generation force field for the simulation of proteins, nucleic acids

570

dx.doi.org/10.1021/ci100474n |*J. Chem. Inf. Model.* 2011, 51, 558–571

and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197. (k) Yu, Y. B.; Privalov, P. L.; Hodges, R. B. Contribution of translational and rotational motions to molecular association in aqueous solution. *Biophys. J.* **2001**, *81*, 1632–1642. (l) Ren, J.; Jenkins, T. C.; Chaires, J. B. Energetics of intercalation reactions. *Biochemistry* **2000**, *39*, 8439–8447.

(35) (a) Schultz, T. W.; Cronin, M. T. D.; Walker, J. D.; Aptula, A. O. Quantitative structure—activity relationships (QSARs) in toxicology: A historical perspective. *J. Mol. Struct. — THEOCHEM* **2003**, *622*, 1–22. (b) Karcher, W.; Devillers, J. SAR and QSAR in the Environmental Chemistry and Toxicology: Scientific Tool or Wishful Thinking? In *Practical Applications of Quantitative Structure—activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht, Boston, London, 1990; pp 1—12. (c) Singh, P. P.; Pasha, F. A.; Srivastava, H. K. Novel application of softness parameter for regioselectivity and reaction mechanism. *Indian J. Chem. B.* **2004**, *43*, 983–991. (d) Katritzky, A. R.; Petrukhin, R.; Tatham, D.; Basak, S.; Benfenati, E. Interpretation of Quantitative Structure-Property and -Activity Relationships. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 679–685.

(36) (a) Ravindra, G. K.; Achaiah, G.; Sastry, G. N. Molecular modeling studies of phenoxypyrimidinyl imidazoles as p38 kinase inhibitors using QSAR and docking. *Eur. J. Med. Chem.* **2008**, *43*, 830–838. (b) Janardhan, S; Srivani, P; Sastry, G. N. 2D and 3D quantitative structure—activity relationship studies on a series of bis-pyridinium compounds as choline kinase inhibitors. *QSAR Combi. Sci.* **2006**, *25*, 860–872. (c) Srivastava, H. K. A Comparative QSPR study of alkanes with the help of computational chemistry. *Bull. Korean Chem. Soc.* **2009**, *30*, 67–76. (d) Srivani, P.; Srinivas, E.; Raghu, R.; Sastry, G. N. Molecular modeling studies of pyridopurinone derivatives-potential phosphodiesterase 5 inhibitors. *J. Mol. Graph. Modell.* **2007**, *26*, 378–390. (e) Pasha, F. A.; Muddassar, M.; Cho, S. J. Molecular docking and 3D QSAR studies of Chk2 inhibitors. *Chem. Biol. Drug. Des.* **2009**, *73*, 292–300. (f) Srivastava, H. K.; Pasha, F. A.; Singh, P. P. Atomic softness based QSAR activity of testosterone derivatives. *Int. J. Quantum Chem.* **2005**, *103*, 237–245. (g) Srivani, P.; Sastry, G. N. Potential choline kinase inhibitors: A molecular modeling study of bis-quinolinium compounds. *J. Mol. Graph. Modell.* **2009**, *27*, 676–688. (h) de Jonge, M. R.; Koymans, L. M.; Vinkers, H. M.; Daeyaert, F. F.; Heeres, J.; Lewi, P. J.; Janssen, P. A. Structure based activity prediction of HIV-1 reverse transcriptase inhibitors. *J. Med. Chem.* **2005**, *48*, 2176–2183.

(37) (a) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652. (b) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. Molecular dynamics simulations of RNA: An in silico single molecule approach. *Biopolymers.* **2007**, *85*, 169–184.

(38) (a) Lu, N. D.; Kofke, D. A. Accuracy of free energy perturbation calculations in molecular simulation. I. Modeling. *J. Chem. Phys.* **2001**, *114*, 7303–7311. (b) Florian, J.; Goodman, M. F.; Warshel, A. Free energy perturbation calculations of DNA destabilization by base substitutions: The effect of neutral guanine thymine, adenine cytosine and adenine difluorotoluene mismatches. *J. Phys. Chem. B.* **2000**, *104*, 10092–10099. (c) Kamal, A.; Shankaraiah, N.; Reddy, C. R.; Prabhakar, S.; Markandeya, N.; Srivastava, H. K.; Sastry, G. N. Synthesis of bis-1,2,3-triazolo-bridged unsymmetrical pyrrolobenzodiazepine trimers via 'click' chemistry and their DNA-binding studies. *Tetrahedron.* **2010**, *66*, 5489–5506. (d) Brandsdal, B. O.; Smalas, A. O. Evaluation of protein—protein association energies by free energy perturbation calculations. *Protein Eng.* **2000**, *13*, 239–245. (e) Beveridge, D. L.; Dicapua, F. M. Free-energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biol.* **1989**, *18*, 431–492.

(39) (a) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1–20. (b) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.

(40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.;

Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision E.0.1; Gaussian, Inc.: Pittsburgh, PA, 2003.

(41) Katritzky, A. R.; Lobanov, V. S.; Karelson, M.*CODESSA 2.0, Comprehensive Descriptors for Structural and Statistical Analysis*; University of Florida: Gainesville, FL, 1994—1996.

(42) *Scigress Explorer*, version 7.7; Fujitsu: Tokyo, 2008.

(43) (a) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004. (b) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146. (c) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, A. P.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. (d) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast efficient generation of high-quality atomic charges. AM1-BCC model: II. Parametrization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(44) Beck, D. A. C.; Armen, R. S.; Daggett, V. Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochemistry* **2005**, *44*, 609–616.

(45) Shaikh, S. A.; Jayaram, B. A Swift all-atom energy-based computational protocol to predict DNA—ligand binding affinity and ΔT_m. *J. Med. Chem.* **2007**, *50*, 2240–2244.

571

dx.doi.org/10.1021/ci100474n |*J. Chem. Inf. Model.* 2011, 51, 558–571