Efficient estimation of MMGBSA-based BEs for DNA and aromatic furan amidino derivatives

Hemant Kumar Srivastava* and G. Narahari Sastry*

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

Communicated by Ramaswamy H. Sarma

(Received 29 March 2012; final version received 17 May 2012)

Molecular mechanics with Generalized Born surface area (MMGBSA) based binding energies (BEs) derived from the molecular dynamics (MD) trajectories are highly reliable and extensively used standards to estimate the strength of interactions between ligands and their receptor. MD simulations (5 ns) for 30 aromatic furan amidino derivatives (anti- 
Pneumocystis carinii agents) have been carried out by using Amber program and BEs have been calculated by using Generalized Born (GB) method. Based on the generated data, we present a simple and effective method for the approximation of BEs without performing MD simulations and MMGBSA calculations. Quantum chemical (density functional theory based) and geometrical descriptors are used for the prediction of the BE values. All the developed models are statistically significant with high values of correlation and cross-validation coefficients. The prediction ability and effectiveness of the models are tested by the division of the data-set into four different training and test sets and the average error was only 4–7% (1.56–2.61 kcal/mol) of the actual BEs.

Keywords: DNA minor groove binders; MMGBSA-based BEs; molecular dynamics simulations; thermal stability of DNA–ligand complexes (ΔT_m values); Pneumocystis carinii agents; DFT-based QSAR studies

Introduction

Dicationic aromatic amidine molecules bind noncovalently to the AT rich region of the minor groove of DNA (Boykin et al., 1998; Francesconi et al., 1999; Hopkins et al., 1998). These molecules are known as effective agents against various organisms like 
Pneumocystis carinii, Giardia lambia, Cryptosporidium parvum, etc. (Bell et al., 1993; Boykin et al., 1995, 1998; Brendle et al., 2002; Francesconi et al., 1999; Hopkins et al., 1998; Lombardy et al., 1996; Tidwell et al., 1993). It has been shown that the activity of these molecules is directly proportional to the DNA binding affinity (Boykin et al., 1998). Numerous attempts have been made in order to design 2,5-Bis[4-(N-alkylamidino)phenyl]furans, 2,4-bis (4-aminophenyl)furans, and extended aromatic furan amidino derivatives with enhanced efficacy and DNA minor groove binding affinity (Boykin et al., 1998; Francesconi et al., 1999; Hopkins et al., 1998). Several factors were found important for enhanced DNA binding affinity such as electrostatic interaction, hydrogen bonding, van der Waals interaction, radius of curvature of the molecule, etc. (Francesconi et al., 1999). In general, the increase in aromatic region shows higher binding affinity probably due to increased potential for various kinds of noncovalent interactions (Vijay et al., 2008). Other important factors for determining drug activity include cell membrane permeability and nonspecific binding (Francesconi et al., 1999).

The DNA minor groove is the target for a large number of drugs such as the antiviral agents netropsin and distamycin (Lown et al., 1989; Zimmer, 1975), the anti- 
Pneumocystis carinii drugs pentamidine and furamidines (de-Souza et al., 2004; Stuedli et al., 2011), the antiseptic and disinfectant propamidine (McDonnell et al., 1999), the trypanocidal drug berenil (Othman et al., 2004), the experimental antitumor agent and important DNA stain Hoechst dyes (Gravatt et al., 1994; Purschke et al., 2010), various antitumor compounds (Braithwaite et al., 1980), etc. Several minor groove binders showed potent antitumor activity in preclinical studies, and some of them went to an advanced stage of clinical trials. These include duocarmycin derivatives adozelesin (Burris et al., 1997; Foster et al., 1996), carzelesin (Houghton et al., 1996; Li et al., 1992), bizelesin (Carter et al., 1996; Walker et al., 1994), KW-2189 (Kobayashi et al., 1994), and a distamycin A derivative tallimustine (Pezzoni et al., 1991). Sequence-selective DNA minor groove binding agent SJG-136 was in phase-I clinical trials.
against advanced solid tumors (Hochhauser et al., 2009; Puzanov et al., 2011). Ecteinascidin 743 displayed potent antiproliferative activity against a variety of tumor cells and entered phase II clinical trials (García-Nieto et al., 2000). DNA minor groove binder brostallicin (PNU-166,196, a-bromo-acrylamido tetra-pyrole derivative) showed high cytotoxic potency and antitumor activity (Lorusso et al., 2009). Several groups are involved in the study of interaction of small molecules with minor groove of DNA e.g. Dervin’s group developed the biomimetic approach to DNA recognition which underpins the design of cell-permeable molecules for the regulation of gene expression in vivo and contributed on the interactions of small molecules (imidazopyridines, benzimidazoles, polyamides, etc.) with the DNA (Chenoweth et al., 2009; Renneberg et al., 2003; Warren et al., 2005). The influence of the solvent on the BEs to DNA of various nonintercalating antibiotics (netropsin, distamycin, berenil, stilbamidine, etc.) was estimated by combining the effect of the first hydration shell with that of bulk water (Zakrzewska et al., 1984). The nonintercalative binding of DAPI to the minor groove of double-stranded DNA oligomers was explored by computing related intermolecular interaction energies (Gresh, 1985). The interaction energy of diarylamidines (berenil and stilbamidine) with the DNA (A-T rich region in the minor groove) was computed by using additive procedure (Gresh et al., 1984). The technique of partially restrained molecular mechanics enthalpy minimization, which enables the sequence dependence of the DNA binding of a nonintercalating ligand, was developed and applied to analyze the binding of berenil to the minor groove of DNA (Laughton et al., 1990). Stacking is also an important interaction between the minor groove binders and DNA (Chourasia et al., 2011).

Molecular docking is a simplified computational representation of DNA–ligand interactions and usually fails in accurately predicting the binding affinities (Ferrara et al., 2004; Warren et al., 2006). Although various docking protocols include protein flexibility, the predicted binding affinity is usually based on a single protein–ligand complex structure (Badrinarayan et al., 2011; Srivani et al., 2007). Postprocessing methods are alternatives to rule out the weaknesses of simple scoring functions used in popular docking protocols by including the dynamic information in the energy calculation for comparatively more accurate estimation of free energy of binding. These postprocessing methods include molecular dynamics (MD) or Monte Carlo (MC) simulations for trajectory generation and free energy perturbation (FEP) (Zwanzig et al., 1954), thermodynamic integration (TI) (Kirkwood et al., 1935), linear interaction energy analysis (LIE) (Aqvist et al., 1994) and molecular mechanics Poisson–Boltzmann or Generalized Born surface-area (MMGBSA/GBSA) calculations (Kollman et al., 2000; Srinivasan et al., 1998) for free energy estimation using generated trajectories. MMGBSA and LIE are recent, computationally efficient and appropriate methods for the free energy estimation of a diverse set of ligands and MMGBSA can be applied to any drug–ligand system without additional regression (Thompsan et al., 2011).

Although the combination of MD and MMGBSA seems to be a good approach for the study of drug–ligand interactions, it is computationally expensive and may not be used universally only for the estimation of DNA binding affinities. Considering this, we propose a simple protocol to predict the DNA binding affinity. Thirty aromatic furan amidino derivatives were collected with the experimental DNA binding data (Boykin et al., 1998; Francesconi et al., 1999; Hopkins et al., 1998) for the purpose. MD simulations (5 ns) were performed on all 30 derivatives to generate the trajectories followed by MMGBSA calculations to estimate the BEs. Further, the estimated BE values were used as a data-set of actual BE values for generating density functional theory (DFT) based quantitative structure–activity relationship (QSAR) models for the prediction of those values. QSAR is an important and widely acceptable tool for understanding many aspects of chemical and biological interactions (Katritzky et al., 2001; Leach 2011; McKinney et al., 2000) and has been applied on variety of systems (Bohari et al., 2011; de Jonge et al., 2005; Janardhan et al., 2006; Pasha et al., 2009; Ravindra et al., 2008; Schultz et al., 2003; Srivastava 2009; Srivastava et al., 2009). These methods are based on the axiom that the variance in the activities, toxicities or physicochemical properties of chemical compounds is determined by the variance in their molecular structures. Quantum chemical descriptors have been used frequently (especially in the last years) in QSAR studies, because of the large well-defined physical information content encoded in many theoretical descriptors. After successful prediction of BE values, the study was extended to predict the $\Delta T_m$ (thermal stability of DNA–ligand complexes defined as the difference between the melting point of DNA alone and the melting point of DNA–ligand complex) values of these molecules by using quantum chemical, topological and docking fitness score-based descriptors. The $\Delta T_m$ prediction models were generated on 28 aromatic furan amidino derivatives and validated by generating four different training and test sets. The current study reveals the reliable and effective predictions of MMGBSA-based BE values and also present a valuable model for the prediction of $\Delta T_m$ values.

Materials and methods

Twelve derivatives of 2,5-bis[4-(N-alkylamidino)phenyl]furan (Boykin et al., 1998), nine derivatives of 2,4-bis(4-amidinophenyl)furan (Francesconi et al., 1999) and
another nine derivatives of extended aromatic furan (Hopkins et al., 1998) were collected with experimental \( \Delta T_m \) values. All the derivatives were fully optimized by means of DFT, using the hybrid three-parameter Becke–Lee–Yang–Parr (B3LYP) functional (Becke, 1993; Lee et al., 1988) with 6-31G(d) basis set (Ditchfield et al., 1971) on Gaussian 03 program package (Frisch et al., 2003). Frequency calculations characterize that all the obtained stationary points are minima on the potential energy surface. Descriptors were calculated by using Codessa (Comprehensive Descriptors for Structural and Statistical Analysis) program (Katritzky et al., 1994, 1995) unless otherwise stated and regression analysis was carried out on project leader program associated with Scigress explorer package (Scigress, 2008). Codessa is a QSAR/QSPR approach that computes hundreds of structural parameters using the constitutional, topological, geometrical, electrostatic, and quantum chemical descriptors of the chemical compounds. Molecular docking studies of these compounds were carried out on GOLD (Genetic Optimization for Ligand Docking) program (Jones et al., 1997). DNA duplex d(CGCGAATTCGCG) was used as a receptor for all the docking calculations, because the experimental values were observed on this DNA sequence. The MD simulation (with explicit solvent, involving \( \sim 10,000 \) water molecules) was performed up to 5 ns for all the derivatives using Amber 8.0 program (Case et al., 2004; Jakalian et al., 2000; Wang et al., 2004) and binding energies were estimated using MMGBSA calculations.

### Docking

Sybyl 6.9.2 program (Tripos Inc., St Louis, MO, USA) was used for input preparation (Sybyl, 2004). All the derivatives were minimized to 0.001 kcal mol\(^{-1}\)/Å root-mean-square gradients by using MMFF94 force field and point charges. All the considered complexes were optimized at B3LYP/6-31G(d) level and the final geometries were used as input for minimization in Sybyl 6.9.2 program. The DNA duplex d(CGCGAATTCGCG) was extracted from 1VZK (downloaded from protein data bank) (Mallena et al., 2004), explicit hydrogens were added and the energy minimization of hydrogens was carried out using 0.01 kcal mol\(^{-1}\)/Å root-mean-squared gradient. Recently, GOLD docking was proven to be the best docking protocol which reliably reproduces the crystallographic poses for DNA-ligand complexes (Srivastava et al., 2011) and used for the study of various DNA–ligand complexes (Kamal et al., 2009; 2010a; 2010b; 2010c; 2011). Thus in the present study, we performed the docking calculations for current set of derivatives using GOLD3.2 program that uses the genetic algorithm (GA). This method allows a partial flexibility of receptor and full flexibility of ligand in the process of 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) were used and operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. Default cut-off 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 derivative and the best score pose was selected for MD simulations.

### MD simulations

Amber8.0 program was used for MD simulations of the selected docked poses. The ‘leaprc.gaff’ (generalized amber force filed) was used to prepare the ligands, while ‘leaprc.ff03’ was used for the DNA. The ‘addions’ command implemented in ‘xleap’ of AMBER 8.0 was used to add the Na\(^+\) ions explicitly to neutralize the system. Each system was placed in a rectangular box of TIP3P water by using ‘SolvateOct’ command with the minimum distance between any solute atom and the boundary of the box was set to 10 Å. Equilibration of the solvated complex was done by carrying out a short 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. Adequate cut-off sizes are necessary to get reliable MD results and therefore a cut-off of 12.0 Å was used for MD simulations. All long-range electrostatics were included by means of a Particle mesh Ewald (PME) method (Darden et al., 1993). 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 were used for production run, and the coordinates were recorded in every 10 ps. The periodic boundary conditions (PBC) were used during MD simulations. Before submitting for the production run, we verified that the system is equilibrated. Five hundred equally spaced snapshots of the complex (every 10 ps) were generated from the MD trajectories, and all water molecules and counter ions were removed before MMGBSA calculations. Coordinates were extracted by using ‘extract_coords.mmpbsa’ script and the BE values were calculated by using ‘binding_energy.mmpbsa’ script. The reported BE values, obtained from GB (semi-analytical approximations to continuum electrostatics) method (Edinger et al., 1997) are the average of all 500 snapshots. Root mean squared deviation (RMSD)
plots were constructed separately for complexes, DNA and ligands using Xmgrace program and images of snapshots at various time scales were generated using VMD program.

**BE prediction**

B3LYP/6-31G(d) optimized geometries were used for generating descriptors on Codessa program (Katritzky et al., 1994, 1995). Different global reactivity descriptors like chemical potential ($\mu$), electro-negativity ($\chi$) hardness ($\eta$), and electrophilicity index ($\omega$) were calculated as follows (Chattaraj et al., 2006; Parr et al., 1978, 1983, 1999).

$$\chi = -\mu = -\left(\frac{\partial E}{\partial N}\right)_{v(r)}$$
$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(r)}$$
$$\omega = \frac{\mu^2}{2\eta} = \frac{\chi^2}{2\eta}$$

Using finite difference approximation and Koopman’s theorem, above derivatives can be estimated as follows.

$$\mu = \frac{1}{2}(E_{HOMO} + E_{LUMO})$$
$$\eta = (E_{LUMO} - E_{HOMO})$$
$$\omega = \frac{\mu^2}{2\eta}$$

Heuristic method (a statistical technique for descriptor selection and correlation, available in Codessa program) was used for descriptor optimization, which follows a sensible and intuitive pathway for eliminating variables from consideration. The intercorrelation level between descriptors has been tested and highly intercorrelated descriptors were eliminated to avoid chance correlation. The uniqueness of a molecule and its total chemical information cannot be described by very few descriptors, while the large number of descriptors will create confusion and reduce the statistical robustness of the model. Thus, an analysis was done with various models where the number of descriptors was increased from 1 to 10 and three descriptors-based models were found optimum. Almost all possible combinations of descriptors were tested to select the final descriptors. Project leader program associated with Scigress explorer package (Scigress, 2008) was used for multiple linear regression (MLR) analysis. The entire data-set was divided into four different training and test sets for rigorously validating the predictive ability of the QSAR models. The current set of derivatives was collected from three references (Boykin et al., 1998; Francesconi et al., 1999; Hopkins et al., 1998).

The first test set contains all 2,5-bis[4-(N-alkylamidino)phenyl]furan derivatives, the second test set contains all 2,4-bis(4-amidinophenyl)furan derivatives, and the third test set contains all extended aromatic furan amidino derivatives. While fourth test set was constructed by randomly taking 13 compounds from all three references. Leave-one-out, leave-two-out, leave-three-out, and leave-ten-out validation methods were used for cross-validation of the results. Similar methodology was adopted for generating predictive models for $\Delta T_m$ values. However, conceptual DFT descriptors were not used in this case. Three descriptors [MVN, GS, and AIC$_2^2$] based model was found optimum for the prediction of $\Delta T_m$ values.

**Results and discussion**

Thirty aromatic furan amidino derivatives were collected from literature with experimental $\Delta T_m$ values and presented in Figure 1. These molecules show their activity via noncovalent binding in the AT rich region of the minor groove of DNA (Brovarets et al., 2012; Yurenko et al., 2011). Thus, the molecular docking calculations were performed using GOLD program (see methodology section for details of protocols used for docking calculations). All the derivatives were docked in the AT rich region of the DNA with reasonably good docking fitness scores. The best docking fitness score pose was selected for MD simulations (see methodology section for details of MD simulations protocols). To observe the systematic deviation of docked complexes during MD simulations, RMSD as a function of time is plotted. RMSD plot of complex as well as of DNA and ligand is presented separately in Figure 2 for few selected derivatives (see Supporting information for the plot of all the derivatives). It is clear from Figure 2 that the higher $\Delta T_m$ containing compounds show lower RMSD deviation and vice versa. Figure 3 shows the snapshots at various time scales for the selected derivatives. Figure 4 depicts the hydrogen bonding pattern of the most potent derivative (4) in the minor groove of DNA. This derivative shows six hydrogen bonds with Adenine (A) and Thymine (T) base pairs of DNA (one hydrogen bond each with A-6, A-18, T-7 and T-20, and two with A-17).

Now, we performed MMGBSA calculations by using MD trajectories to obtain BE values. The BE values are very reliable parameter for measuring the strength of interaction between ligands and their receptor. Based on calculated BE values, we present an effective way to predict these values. Various kinds of descriptors (constitutional, geometrical, topological, electrostatic, quantum chemical, conceptual DFT based, etc.) were calculated for each of the derivatives and comparative QSAR studies were performed. Conceptual DFT descriptor (electrophilicity index) and geometrical descriptors (moment of inertia B and YZ shadow) were selected for the final model and the values for these descriptors are presented in Table 1 (see methodology section for details). The descriptors were selected by comparing the correlation coefficient values obtained.
from various combinations of descriptors. The moments of inertia characterize the mass distribution in the molecule and the susceptibility of the molecule to different rotational transitions. Each moment of inertia is defined

Figure 1. Structure, MMGBSA-based BE (Italic), and experimental $\Delta T_m$ values for aromatic furan amidino derivatives considered in this study (− sign in place of $\Delta T_m$ values for derivatives 19 and 20 indicates the unavailability of experimental data).
with respect to a specific rotational axis. Moment of inertia depends on both the mass of the atoms and the geometry of the molecule. Moment of inertia as a descriptor should be used with caution for structurally

Figure 2. Separate RMSD plot for complex, DNA and ligand for the higher and lower active derivatives of 2,5-bis[4-(N-alkylamidino)phenyl]furans (29, 25), 2,4-bis(4-amidinophenyl)furans (15, 16), and extended aromatic furan amidino derivatives (1, 4). (Compound number is depicted in bold, binding energy in italic and ΔTm is underlined).
Figure 3. Snapshots at various time scales for higher and lower stable derivatives of 2,5-bis[4-(N-alkylamidino)phenyl]furans (29, 25), 2,4-bis(4-amidinophenyl)furan (15, 16), and extended aromatic furan amidino derivatives (1, 4). (Compound number is depicted in bold, binding energy in italic and \( \Delta T_m \) is underlined. The chemical structure of all the ligands is also depicted.)
mobile molecules (Hovorun et al., 1999; Nikolaenkov et al., 2011). YZ shadow is the area of the shadows of the molecule as projected on the YZ planes by the orientation of the molecule in the space along the axes of inertia. The electrophilicity index measures the energy lowering of a ligand due to maximal electron flow between donor and acceptor (Parthasarathi et al., 2004; Pearson et al., 1997). Electrophilicity is a reactivity descriptor which allows a quantitative classification of the global electrophilic nature of a molecule within a relative scale and found important for describing the biological activity and toxicity of various inhibitors (Chattaraj et al., 2006; Parr et al., 1999; Parthasarathi et al., 2004; Pasha et al., 2005; Pearson et al., 1997).

Initially, the QSAR model was generated for the whole data-set (using the combination of MI B, YZS, and $\omega$ descriptors) and the results were significant with $R^2 = 0.87$. However, to rigorously validate the prediction model, we have divided the data-set into four different training and test sets. The first test set was constructed by taking all the derivatives of 2,5-bis[4-(N-alkylamidino)phenyl]furan in the test set. The generated prediction model was statistically significant with $R^2 = 0.93$ and $R^2_{cv} = 0.83$, Equation (1). The average residual (AE = average of ‘actual – predicted’ BE values) was 1.3 kcal/mol for training and 1.9 kcal/mol for test set and indicates the good predictive behavior of the generated model-1. Figure 5a shows the plot between actual and predicted BE values and indicates the excellent performance of the model.

$$BE-P^a = -1.17008^*MIB - .395716^*YZS + 11.0375^*\omega + 31.5984$$  (1)

$R^2 = 0.93; R^2_{cv} = 0.83; n = 30$ (training = 18, test = 12)

The second test set was constructed by taking all the derivatives of 2,4-bis(4-amidinophenyl)-furan in the test set. The QSAR model obtained by this test set was very good; however, the statistical parameters for this model was slightly lower ($R^2 = 0.81; R^2_{cv} = 0.68$) compared to model-1 ($R^2 = 0.93; R^2_{cv} = 0.83$). Equation (2) shows the QSAR equation and statistical parameters for model-2. The AE was 1.6 kcal/mol for training and 2.6 kcal/mol for test set for this model. Figure 5(b) depicts the plot between actual and predicted BE values.

$$BE-P^b = -1.05402^*MIB - .395716^*YZS + 11.0375^*\omega + 31.5984$$  (2)

$R^2 = 0.81; R^2_{cv} = 0.68; n = 30$ (training = 21, test = 9)

The third test set was constructed by taking all extended aromatic furan amidino derivatives in the test set. Good correlation coefficient values were obtained for this model also ($R^2 = 0.84; R^2_{cv} = 0.82$), Equation (3). The AE (1.5 kcal/mol for training and 2.6 kcal/mol for test set) for model-3 was comparatively higher for the test set. However, this model can be considered for the prediction as the percent error for the model was only around 7%. Figure 5(c) shows the plot between actual and predicted BE values, and indicates the good performance of the model.

$$BE-P^c = -1.10,116^*MIB - .315532^*YZS + 8.01839^*\omega + 35.8784$$  (3)
Table 1. Actual and predicted BE values with residuals and selected descriptors.

| No. | MIB  | YZS | ω   | BE-C | BE-Pa | BE-Pb | BE-Pc | BE-Pd | Res.a | Res.b | Res.c | Res.d |
|-----|------|-----|-----|------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1   | 27.337 | 29.2603 | 3.5880 | −28.66 | −27.64 | −26.98 | −25.31* | −26.79 | 1.02 | 1.68 | 3.35 | 1.87 |
| 2   | 14.789 | 44.2406 | 3.4183 | −34.82 | −34.52 | −33.38 | −33.04* | −33.85* | 0.3 | 1.44 | 1.78 | 0.97 |
| 3   | 10.384 | 47.0407 | 3.4111 | −39.38 | −38.48 | −37.04 | −36.95* | −37.86* | 0.90 | 2.34 | 2.43 | 1.53 |
| 4   | 17.621 | 36.8805 | 4.2431 | −45.76 | −43.22 | −42.73 | −38.86* | −42.44 | 2.54 | 3.03 | 6.90 | 3.32 |
| 5   | 12.070 | 54.3609 | 4.1415 | −41.06 | −41.68 | −41.78 | −38.64* | −41.10* | −0.6 | −0.72 | 2.42 | −0.04 |
| 6   | 9.5160 | 26.2809 | 4.1378 | −43.46 | −43.86 | −43.81 | −40.82* | −43.31* | −0.40 | −0.35 | 2.64 | 0.15 |
| 7   | 21.720 | 34.5204 | 3.9191 | −32.93 | −35.78 | −35.24 | −32.49* | −34.98 | −2.85 | −2.31 | 0.44 | −2.05 |
| 8   | 13.343 | 52.6208 | 3.7795 | −35.07 | −36.88 | −36.61 | −34.89* | −36.30 | −1.83 | −1.54 | 0.18 | −1.23 |
| 9   | 09.923 | 56.2609 | 3.7735 | −40.79 | −39.37 | −38.98 | −37.46* | −38.83 | 1.42 | 1.81 | 3.33 | 1.96 |
| 10  | 23.634 | 29.9603 | 2.9905 | −25.38 | −25.10 | −23.43* | −24.38 | −24.29 | 0.28 | 1.95 | 1.00 | 1.09 |
| 11  | 14.894 | 30.3003 | 2.7801 | −36.19 | −32.87 | −29.98* | −32.21 | −32.07* | 3.32 | 6.21 | 3.98 | 4.12 |
| 12  | 11.405 | 36.7205 | 2.8633 | −35.59 | −35.33 | −32.62* | −34.69 | −34.61 | 0.26 | 2.97 | 0.90 | 0.99 |
| 13  | 11.466 | 32.0604 | 2.9038 | −33.49 | −37.55 | −34.53* | −36.42 | −36.77* | −0.06 | −1.04 | −2.93 | −3.28 |
| 14  | 09.103 | 39.7405 | 2.8372 | −36.71 | −36.54 | −33.77* | −36.07 | −35.85* | 1.71 | 2.94 | 0.65 | 0.86 |
| 15  | 07.742 | 36.0404 | 2.8470 | −39.20 | −39.70 | −36.50* | −38.81 | −38.98* | −0.50 | 2.70 | 0.39 | 0.22 |
| 16  | 23.281 | 36.4405 | 2.6706 | −18.28 | −19.41 | −17.86* | −20.16 | −18.69 | −1.13 | 0.42 | −1.88 | −0.41 |
| 17  | 11.268 | 40.7406 | 2.5462 | −32.08 | −30.40 | −27.65* | −31.03 | −29.73* | 1.68 | 4.44 | 1.05 | 2.35 |
| 18  | 11.613 | 26.1002 | 1.9942 | −28.66 | −29.47* | −20.24 | −30.49 | −28.40 | −0.54 | 3.94 | −1.83 | 0.27 |

MIB is a geometrical descriptor named moment of inertia B; YZS is a geometrical descriptor named YZ shadow; ω is a conceptual DFT descriptor named electrophilicity index and the values for ω are given in eV, BE-C is binding energy calculated by GB method after 5 ns MD simulations; BE-Pa is predicted binding energy by using QSAR Equation 1 (all the molecules from reference-1, i.e., 19–30, were used as test set); BE-Pb is predicted binding energy by using QSAR Equation 2 (all the molecules from reference-2, i.e., 10–18, were used as test set); BE-Pc is predicted binding energy by using QSAR Equation 3 (all the molecules from reference-3, i.e., 1–9, were used as test set); BE-Pd is predicted binding energy by using QSAR Equation 4 (test set molecules were selected randomly i.e. five from reference-1, four from reference-2, and four from reference-3); Res. is residual (actual – predicted) of binding energies; and 'show the molecular descriptors in the test set. All the MIB values are in kg·m². BE values and residuals are in kcal/mol.

\[ R^2 = 0.84; \quad R_{cv}^2 = 0.82; \quad n = 30 \quad (\text{training} = 21, \text{test} = 9) \]

The fourth test set was constructed by randomly taking 13 compounds in the test set. This model shows very good correlation coefficient values (\( R^2 = 0.90; R_{cv}^2 = 0.85 \)), Equation (4). The AE for the test set of model-4 was lowest (1.6 kcal/mol for test set) compared to all other models. Interestingly, the AE for test (1.6 kcal/mol) and training (1.6 kcal/mol) sets was essentially very similar, which shows the wide applicability of this model. Figure 5d shows the plot between actual and predicted BE values, and indicates a good performance of the model.

\[
\text{BE-P}^d = -1.17081 \times \text{MIB} - 0.384569 \times \text{YZS} + 10.9957 \times \omega + 30.5957 \\
(4)
\]

\[ R^2 = 0.90; \quad R_{cv}^2 = 0.85; \quad n = 30 \quad (\text{training} = 17, \text{test} = 13) \]

The docking fitness score was more than 80.0 for compounds with experimental \( \Delta T_m \) values greater than 12.0 with one exception. However, a very good qualitative or quantitative (\( R^2 = 0.34 \)) relationship between docking fitness scores and \( \Delta T_m \) values was not visible. Thus, we also present the predictive models for \( \Delta T_m \) values. Quantum chemical descriptor called 'maximum valency of a N atom,' ‘GOLD fitness scores,’ and topological descriptor called 'average information content order-2' were used for generating \( \Delta T_m \) prediction models, Table 2. The generated models were validated by dividing them into four different training and test sets, Equations (5)–(8). All the models show similar performance for the prediction of \( \Delta T_m \) values (AE for training and test set for all the models was around
However, model-4 was statistically the best predictive model ($R^2 = .82$ and $R^2_{cv} = .80$). Figure 6 shows the plots between experimental and predicted $\Delta T_m$ values for all four sets.

Typically, an outlier is defined as any derivative for which the experimentally observed value substantially

$$
\Delta T_m^{\text{Pc}} = 16.5045^{*}\text{MVN} + .177179^{*}\text{GS} - 1.0192^{*}\text{AIC}_2 - 16.9343
$$

(7)

$$
R^2 = .84; R^2_{cv} = .53; n = 28; \text{ AE for Training Set } = .99; \text{ AE for Test Set } = 1.09
$$

$$
\Delta T_m^{\text{Pd}} = 20.981^{*}\text{MVN} + .201079^{*}\text{GS} - 11.6519^{*}\text{AIC}_2 - 26.8286
$$

(8)

$$
R^2 = .82; R^2_{cv} = .80; n = 28; \text{ AE for Training Set } = 1.06; \text{ AE for Test Set } = 1.11
$$

Typically, an outlier is defined as any derivative for which the experimentally observed value substantially
differs from the computed value. When removing one or
more data point from the data-set dramatically improves
the statistical quality of a given QSAR model, the outlier
concept is employed. However, in the current study there
was no such problem and outlier concept was not used
while generating QSAR models for binding energy pre-
diction. Leave-one-out and leave-n-out (basically cross-
validation methods) techniques estimate the validity and
applicability of a model by testing its prediction ability
(Kiralj & Ferreira, 2009). Leave-one-out can be performed
by excluding each derivative of the data-set once and
constructing a new model for rest of the derivatives
followed by the prediction of the activity of excluded
derivative. However, leave-n-out excludes the set of n-
umber of derivatives each time while constructing the
model. We also observed the effect of this validation on
the correlation coefficient values. This practice may
reduce the possibility of chance correlation. The gener-
ated QSAR equations were statistically validated using
leave-one-out, leave-two-out, leave-five out, and leave-
ten-out techniques. Table 3 shows the statistical coeffi-
cient values for leave-n-out validation and indicates the
wide applicability of the generated models.

Conclusions
A computational approach that combines quantum
mechanics, docking, molecular dynamics, and statistical
analysis to model the energetic patterns of the interac-
tions between DNA and their ligands is presented. It
has been demonstrated that conceptual DFT descriptor
(electrophilicity index, \( \omega \)) along with two geometrical
descriptors (moment of inertia B and YZ shadow) is
able to predict the MMGBSA-based binding energies
with statistically significant values (\( R^2 = .93, R^2_{cv} = .83
\) and \( AE = 1.32 \text{ kcal/mol for the best model} \)). The predic-

Table 2. Experimental and predicted \( \Delta T_m \) values with residuals and selected descriptors.

| No. | MVN | GS | AIC2 | \( \Delta T_m \) | \( \Delta T_m \)P\(^a\) | \( \Delta T_m \)P\(^b\) | \( \Delta T_m \)P\(^c\) | \( \Delta T_m \)P\(^d\) | Res.a | Res.b | Res.c | Res.d |
|-----|-----|----|------|----------------|----------------|----------------|----------------|----------------|------|------|------|------|
| 1   | 2.9575 | 65.73 | 3.2029 | 11.7 | 12.38 | 12.10 | 11.43* | 11.14 | −68 | −40 | .27 | .56 |
| 2   | 3.1629 | 74.67 | 3.548 | 14.4 | 13.59 | 13.33 | 12.95* | 13.23* | .82 | 1.07 | 1.45 | 1.17 |
| 3   | 3.1679 | 86.62 | 3.5728 | 15.8 | 15.48 | 14.98 | 14.90* | 15.45 | .32 | .82 | .90 | .35 |
| 4   | 3.2997 | 84.84 | 4.0696 | 14.4 | 11.64 | 11.71 | 11.78* | 12.07 | 2.76 | 2.69 | 2.62 | 2.33 |
| 5   | 3.2990 | 95.04 | 4.1833 | 12.9 | 12.06 | 12.05 | 12.44* | 12.78 | .84 | .85 | .46 | .12 |
| 6   | 3.2990 | 98.46 | 4.1497 | 11.7 | 13.06 | 12.93 | 13.38* | 13.86* | −1.36 | −1.23 | −1.68 | −2.16 |
| 7   | 3.3001 | 76.08 | 3.9604 | 11.4 | 11.41 | 11.53 | 11.33* | 11.59* | −.01 | −.13 | .07 | −.19 |
| 8   | 3.2995 | 80.02 | 4.0876 | 10.0 | 10.57 | 10.78 | 10.75* | 10.89* | −.57 | −.78 | −.75 | −.89 |
| 9   | 3.2995 | 85.84 | 4.056 | 0.97 | 11.97 | 12.00 | 12.09* | 12.43 | −2.27 | −2.30 | −2.39 | −2.73 |

MVN is a quantum chemical descriptor named maximum valency of a N atom; GS is GOLD fitness score; AIC2 is a topological descriptor named average information content (order 2); \( \Delta T_m \) is the experimental thermal stability of DNA-ligand complexes.; \( \Delta T_m \)P\(^a\) is predicted thermal stability of the complex by using QSAR Equation 5 (all the molecules from reference-1, i.e. 19–31, were used as test set); \( \Delta T_m \)P\(^b\) is predicted thermal stability of the complex by using QSAR Equation 6 (all the molecules from reference-2, i.e. 10–18, were used as test set); \( \Delta T_m \)P\(^c\) is predicted thermal stability of the complex by using QSAR Equation 7 (all the molecules from reference-3, i.e. 1–9, were used as test set); \( \Delta T_m \)P\(^d\) is predicted thermal stability of the complex by using QSAR Equation 8 (test set molecules were selected randomly i.e. five from reference-1, four from reference-2, and four from reference-3); Res. is residual (experimental – predicted) of \( \Delta T_m \) values; and *show the molecules in the test set. All the \( \Delta T_m \) are given in °centigrade.
The predictive ability of the model is also validated by generating four different training and test sets (average $R^2 = .87$, average $R^2_{cv} = .80$), and leave-one-out and leave-n-out (2, 5, and 10) methods. Proposed methodology is applicable to screen selected derivatives from a large pool of molecules prior to carry out highly expensive MD simulations. We believe that, the current approach may be helpful for the development of a general methodology. The RMSD plots show that the ligands remain bound to the DNA duplex near the preferential binding position and do not experience substantial fluctuations with respect to their initial placements in the DNA minor groove. However, higher fluctuations are observed for lower active compounds and vice versa. While the studies are directed to the aromatic furan amidino derivatives, further validation is required before generalization of this approach to other classes of compounds.

**Acknowledgments**

HKS and GNS thank the Department of Science and Technology (DST), New Delhi for the financial assistance through Fast-Track (SR/FT/CS-031/2009) and Swarnajayanti projects respectively.

**Supplementary Information**

The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2012.703071.

**Abbreviations**

- **MI$_B$**: geometrical descriptor named moment of inertia B
- **YZS**: geometrical descriptor named ‘YZ shadow’
- **$\omega$**: conceptual DFT descriptor named ‘electrophilicity index’
- **MVN**: quantum chemical descriptor named ‘maximum valency of a nitrogen atom’
- **GS**: GOLD fitness score obtained from docking calculations using GOLD program
- **AIC$_2$**: topological descriptor named ‘average information content - order 2’

Figure 6. Plots between experimental and predicted $\Delta T_m$ values. (a) all the derivatives of 2,5-bis[4-(N-alkylamidino)phenyl]furan (19–30) in the test set, b all the derivatives of 2,4-bis(4-amidinophenyl)furan (10–18) in the test set, c all extended aromatic furan amidino derivatives (1–9) in the test set, and d test set was constructed by randomly taking 13 compounds from all three type of derivatives in the test set. $\Delta T_m$ is the experimental thermal stability of the DNA-ligand complexes. $\Delta T_m$-P is predicted thermal stability of the DNA-ligand complexes using Equations (5)–(8). Circles show the molecules in the test set.
| Out | $R^2$ | $R_{cv}^2$ | Out | $R^2$ | $R_{cv}^2$ | Out | $R^2$ | $R_{cv}^2$ | Out | $R^2$ | $R_{cv}^2$ |
|-----|-------|-----------|-----|-------|-----------|-----|-------|-----------|-----|-------|-----------|
| None | 86.20 | 84.60     | None | 86.20 | 84.60     | None | 86.20 | 84.60     | None | 86.20 | 84.60     |
| 1   | 86.07 | 84.40     | 1–2 | 86.21 | 84.49     | 1–5 | 86.25 | 84.28     | 1–10 | 82.26 | 78.93     |
| 2   | 86.30 | 84.66     | 2–3 | 86.24 | 84.52     | 2–6 | 83.54 | 81.19     | 2–11 | 84.74 | 81.88     |
| 3   | 86.12 | 84.45     | 3–4 | 86.29 | 84.58     | 3–7 | 83.38 | 81.01     | 3–12 | 84.50 | 81.59     |
| 4   | 86.37 | 84.73     | 4–5 | 85.53 | 83.72     | 4–8 | 83.71 | 81.38     | 4–13 | 86.19 | 83.60     |
| 5   | 85.47 | 83.73     | 5–6 | 83.83 | 81.81     | 5–9 | 84.25 | 82.00     | 5–14 | 86.59 | 84.08     |
| 6   | 84.80 | 82.97     | 6–7 | 85.33 | 83.50     | 6–10 | 83.62 | 81.27     | 6–15 | 87.11 | 84.70     |
| 7   | 86.68 | 85.09     | 7–8 | 86.85 | 85.21     | 7–11 | 87.43 | 85.64     | 7–16 | 81.85 | 78.44     |
| 8   | 86.34 | 84.70     | 8–9 | 86.25 | 84.53     | 8–12 | 87.17 | 85.33     | 8–17 | 82.10 | 78.74     |
| 9   | 86.16 | 84.50     | 9–10 | 84.67 | 82.76     | 9–13 | 88.31 | 86.65     | 9–18 | 80.60 | 76.96     |
| 10  | 84.73 | 82.90     | 10–11 | 86.89 | 85.25     | 10–14 | 88.28 | 86.61     | 10–19 | 79.58 | 75.75     |
| 11  | 88.06 | 86.63     | 11–12 | 88.22 | 86.74     | 11–15 | 89.02 | 87.45     | 11–20 | 83.52 | 80.42     |
| 12  | 86.26 | 84.61     | 12–13 | 87.75 | 86.22     | 12–16 | 81.62 | 78.99     | 12–21 | 84.35 | 81.42     |
| 13  | 87.74 | 86.27     | 13–14 | 87.69 | 86.15     | 13–17 | 81.88 | 79.29     | 13–22 | 84.79 | 81.94     |
| 14  | 86.18 | 84.53     | 14–15 | 85.81 | 84.01     | 14–18 | 78.39 | 75.30     | 14–23 | 81.71 | 78.29     |
| 15  | 85.83 | 84.13     | 15–16 | 79.51 | 76.95     | 15–19 | 77.23 | 73.97     | 15–24 | 81.55 | 78.10     |
| 16  | 80.09 | 77.81     | 16–17 | 80.41 | 77.96     | 16–20 | 77.56 | 74.35     | 16–25 | 84.74 | 81.88     |
| 17  | 86.62 | 85.01     | 17–18 | 86.02 | 84.27     | 17–21 | 88.36 | 86.70     | 17–26 | 91.91 | 90.39     |
| 18  | 85.62 | 83.90     | 18–19 | 85.30 | 83.46     | 18–22 | 88.15 | 86.46     | 18–27 | 92.94 | 91.37     |
| 19  | 85.93 | 84.24     | 19–20 | 85.95 | 84.19     | 19–23 | 88.87 | 87.28     | 19–28 | 93.08 | 91.78     |
| 20  | 86.23 | 84.57     | 20–21 | 88.66 | 87.24     | 20–24 | 89.10 | 87.54     | 20–29 | 92.91 | 91.49     |
| 21  | 88.65 | 87.29     | 21–22 | 88.90 | 87.51     | 21–25 | 91.03 | 89.74     | 21–30 | 92.98 | 91.66     |
| 22  | 86.38 | 84.75     | 22–23 | 86.50 | 84.81     | 22–26 | 89.20 | 87.66     | –      | –      | –         |
| 23  | 86.30 | 84.66     | 23–24 | 86.31 | 84.60     | 23–27 | 89.81 | 88.35     | –      | –      | –         |
| 24  | 86.20 | 84.55     | 24–25 | 87.86 | 86.34     | 24–28 | 89.60 | 88.12     | –      | –      | –         |
| 25  | 87.84 | 86.38     | 25–26 | 88.70 | 87.29     | 25–29 | 89.28 | 87.75     | –      | –      | –         |
| 26  | 86.96 | 85.39     | 26–27 | 87.67 | 86.13     | 26–30 | 87.37 | 85.56     | –      | –      | –         |
| 27  | 86.85 | 85.27     | 27–28 | 86.83 | 85.18     | –      | –      | –         | –      | –      | –         |
| 28  | 86.19 | 84.53     | 28–29 | 85.95 | 84.20     | –      | –      | –         | –      | –      | –         |
| 29  | 85.96 | 84.27     | 29–30 | 86.08 | 84.34     | –      | –      | –         | –      | –      | –         |
| 30  | 86.25 | 84.60     | –      | –      | –         | –      | –      | –         | –      | –      | –         |

Out is the derivatives kept out for generating models; $R^2$ is correlation coefficient; and $R_{cv}^2$ is cross-validation coefficient values.
References

Aqvist, J., Medina, C., & Samuelsson, J.E. (1994). A new method for predicting binding affinity in computer-aided drug design. Protein Engineering, 7, 385–391.

Badrinarayan, P., & Sastry, G.N. (2011). Sequence, analysis of p38 MAP kinase: Exploiting DFG-out conformation as a strategy to design new type II leads. Journal of Chemical Information and Modeling, 51, 115–129.

Beeke, A.D. (1993). Density-functional thermochemistry III. The role of exact exchange. Journal of Chemical Physics, 98, 5648–5652.

Bell, C.A., Dykstra, C.C., Naiman, N.N., Cory, M., Fairley, T.A., & Tidwell, R.R. (1993). Structure-activity studies of dicationic substituted bis-benzimidazoles against Giardia lamblia: correlation of antiagdial activity with DNA binding affinity and giardial topoisomerase II inhibition. Antimicrobial Agents and Chemotherapy, 37, 2668–2673.

Bohari, M.H., Srivastava, H.K., & Sastry, G.N. (2011). Analogue-based approaches in anti-cancer compound modeling: The relevance of QSAR models. Organic and Medicinal Chemistry Letters, 1, 3.

Boykin, D.W., Kumar, A., Spychala, J., Zhou, M., Lombardy, R.J., Wilson, W.D., … Neidle, S. (1995). Dicationic diarylfurans as anti-Pneumocystis carinii agents. Journal of Medicinal Chemistry, 38, 912–916.

Boykin, D.W., Kumar, A., Xiao, G., Wilson, W.D., Bender, B.C., McCurdy, D.R., … Tidwell, R.R. (1998). 2,5-bis-[4-(N-alkylamidino)phenyl]furans as anti-Pneumocystis carinii agents. Journal of Medicinal Chemistry, 41, 124–129.

Braithwaite, A.W., & Baguley, B.C. (1980). Existence of an extended series of antitumor compounds which bind to deoxyribonucleic acid by nonintercalative means. Biochemistry, 19, 1101–1106.

Brendle, J.J., Outlaw, A., Kumar, A., Boykin, D.W., Patrick, D.A., Tidwell, R.R., & Werbovetz, K.A. (2002). Antileishmanial activities of several classes of aromatic dications. Antimicrobial Agents and Chemistry, 46, 797–807.

Brovarets, O.O., Yurenko, Y.P., Dubey, I.Y., & Hovorun, D.M. (2012). Can DNA-binding proteins of resipimeline tautomerize nucleotide bases? Ab initio model study. Journal of Biomolecular Structure and Dynamics, 29, 597–605.

Burris, H.A., Diers, V.C., Tuncu, M., Earhart, R.H., Eckardt, J.R., Rodriguez, G.I., … Von-Hoff, D.D. (1997). Phase I study with the DNA sequence-specific agent adozelesin. Anti-Cancer Drugs, 8, 588–596.

Carter, C.A., Waud, W.R., Li, L.H., DeKoning, T.F., McGovern, J.P., & Ploswam, J. (1996). Preclinical antitumor activity of bizelesin in mice. Clinical Cancer Research, 2, 1143–1149.

Case, D.A., Darden, T.A., Cheatham, T.E.III, Simmerling, C.L., Wang, J., Duke, R.E., … Kollman, P.A. (2004). AMBER 8. San Francisco, CA: University of California, San Francisco.

Chattaraj, P.K., Sarkar, U., & Roy, D.R. (2006). Electrophilicity index. Chemical Reviews, 106, 2065–2091.

Chenoweth, D.M., & Dervan, P.B. (2009). AllostERIC modulation of DNA by small molecules. Proceedings of the National Academy of Sciences USA, 106, 13175–13179.

Chourasia, M., Sastry, G.M., & Sastry, G.N. (2011). Aromatic – aromatic database, A2ID: An analysis of aromatic networks in proteins. International Journal of Biological Macromolecules, 48, 540–552.

Darden, T., York, D., & Pederson, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. Journal of Chemical Physics, 98, 10089–10092.

de Jonge, M.R., Koymans, L.M., Vinkers, H.M., Daeyaert, F.F., Heeres, J., Lewi, P.J., & Janssen, P.A. (2005). Structure-based activity prediction of HIV-1 reverse transcriptase inhibitors. Journal of Medicinal Chemistry, 48, 2176–2183.

De Souza, E.M., Lansiaux, A., Bailly, C., Wilson, W.D., Hu, Q., Boykin, D.W., … Soeiro, M.N. (2004). Phenyl substitution of furamidine markedly potentiates its anti-parasitic activity against Trypanosoma cruzi and Leishmania amazonensis. Biochemical Pharmacology, 68, 593–600.

Ditchfield, R., Hehre, W.J., & Pople, J.A. (1971). Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules. Journal of Chemical Physics, 54, 724.

Edinger, S.R., Cortis, C., Shenkin, P.S., & Friesner, R.A. (1997). Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the poisson-boltzmann equation. Journal of Physical Chemistry B, 101, 1190–1197.

Ferrara, P., Gohlke, H., Price, D.J., Klebe, G., & Brooks, C.L. III (2004). Assessing scoring functions for protein–ligand interactions. Journal of Medicinal Chemistry, 57, 3032–3047.

Foster, B.J., LoRusso, P.M., Poplin, E., zalupski, M., Valdivieso, M., Wozniak, A., … Baker, L.H. (1996). Phase I trial of adozelesin using the treatment schedule of daily 35 every 3 weeks. Investigational New Drugs, 13, 321–326.

Francesconi, I., Wilson, W.D., Tansous, F.A., Hall, J.E., Bender, B.C., Tidwell, R.R., … Boykin, D.W. (1999). 2,4-Diphenyl furan diamidines as novel anti-Pneumocystis carinii pneumonia agents. Journal of Medicinal Chemistry, 42, 2260–2265.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., … Pople, J.A. (2003). Gaussian 03 Revision E.0.1. Pittsburgh, PA: Gaussian.

Garcia-Nieto, R., Manzanares, I., Cuevas, C., & Gago, F. (2000). Increased DNA binding specificity for antitumor ecteinascidin 743 through protein-DNA interactions? Journal of Medicinal Chemistry, 16, 4367–4369.

Gravitt, G.L., Baguley, B.C., Wilson, W.R., & Denny, W.A. (1994). DNA-directed alkylation agents. 6. Synthesis and antitumor activity of DNA minor groove-targeted antilrive mustard analogs of phenibezimol (Hoechst 33258). Journal of Medicinal Chemistry, 37, 4338–4345.

Gresh, N. (1985). A theoretical study of the interaction of 4′,6-diamidino-2-phenylindole (DAPI) with the double-stranded oligonucleotides (dA-dT)11 and (dA)11·(dT)11. Journal of Biological Macromolecules, 7, 199–202.

Gresh, N., & Pullman, B. (1984). A theoretical study of the non-intercalative binding of berenil and stilbamidine to double-stranded (dA-dT)11 oligomers. Molecular ZPharma-

logy, 25, 452–458.

Hoehuhauser, D., Meyer, T., Spanswicks, V.J., Wu, J., Clingen, P.H., Loadman, P., … Jodrell, D. (2009). Phase I study of sequence-selective minor groove DNA binding agent SJG-136 in patients with advanced solid tumors. Clinical Cancer Research, 15, 2140–2147.

Hopkins, K.T., Wilson, W.D., Bender, B.C., McCurdy, D.R., Hall, J.E., Tidwell, R.R., … Boykin, D.W. (1998). Extended aromatic furan amidino derivatives as anti-Pneumocystis carinii agents. Journal of Medicinal Chemistry, 41, 3872–3878.
Houghton, P.J., Cheshire, P.J., Hallman, J.D.N., & Houghton, J.A. (1996). Therapeutic efficacy of the cyclopropylypyrroloindole, carzelesin, against xenografts derived from adult and childhood solid tumors. *Cancer Chemotherapy and Pharmacology, 36*, 45–52.

Hovorun, D.M., Gorbi, L., & Leszcynski, J. (1999). From the nonplanarity of the amino group to the structural nonrigidity of the molecule: A post-Hartree–Fock *ab initio* study of 2-aminoimidazole. *International Journal of Quantum Chemistry, 74*, 245–253.

Jakalian, A., Bush, B.L., Jack, D.B., & Bayly, C.I. (2000). Fast, efficient generation of high-quality atomic charges. *AM1-BCC Model. Journal of Computational Chemistry, 21*, 132–146.

Janardhan, S., Srivani, P., & Sastry, G.N. (2006). 2D and 3D synthesis of bis-1,2,3-triazolo-bridged unsymmetrical thiazole/benzoxazole conjugates. *Bioorganic & Medicinal Chemistry, 17*, 1557–1572.

Kamal, A., Reddy, K.S., Khan, M.N., Shetti, R.V., Ramaiah, M.J., Pushpavalli, S.N., & Barkume, M. (2010c). Remarkable enhancement in the DNA-binding ability of C2-fluoro substituted pyrrolo[2,1-c][1,4]benzodiazepine (PBD) conjugates: Design, synthesis and biological evaluation as potential anticancer agents. *Bioorganic & Medicinal Chemistry, 18*, 526–542.

Kamal, A., Rajender Reddy, D.R., Reddy, M.K., Balakishan, G., Shaik, T.B., Chourasia, M., & Sastry, G.N. (2009). Remarkable enhancement in the DNA-binding ability of C2-fluoro substituted pyrrolo[2,1-c][1,4]benzodiazepines and their anticancer potential. *Bioorganic & Medicinal Chemistry, 17*, 1557–1572.

Kamal, A., Shankaraiah, N., Reddy, C.R., Prabhakar, S., Markandeya, N., Srivastava, H.K., & Sastry, G.N. (2010b). Synthesis of bis-1,2,3-triazolo-bridged unsymmetrical pyrrolo-benzodiazepine: trimers via ‘click’ chemistry and their DNA-binding studies. *Tetrahedron, 66*, 5489–5506.

Kamal, A., Shetty, R.V.C.R.N.C., Ramaiah, M.J., Swapna, P., Reddy, K.S., Mallareddy, A., ... Pal-Bhadra, M. (2011). Carbazole-pyrrolo[2,1-c][1,4]benzodiazepine conjugates: Design, synthesis, and biological evaluation. *MedChemComm, 2*, 780–788.

Katritzky, A.R., Lobanov, V.S., & Karelson, M. (1994). Copyright 1994–1996, CODESSA 2.0, Comprehensive descriptors for structural and statistical analysis, University of Florida, USA.

Katritzky, A.R., & Lobanov, V.S. (1995). QSAR: The correlation and quantitative prediction of chemical and physical properties from structure. *Chemical Society Reviews, 24*, 279–287.

Katritzky, A.R., Petrukhin, R., Tatham, D., Basak, S., & Benfenati, E. (2001). Interpretation of quantitative structure–property and activity relationships. *Journal of Chemical Information and Computer Sciences, 41*, 679–685.

Kiralj, R., & Ferreira, M.M.C. (2009). Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *Journal of the Brazilian Chemical Society, 20*, 770–787.

Kirkwood, J.G. (1935). Statistical mechanics of fluid mixtures. *Journal of Chemical Physics, 3*, 300–313.

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

Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Hsu, S., Chong, L., ... Cheatham, T.E.III (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research, 33*, 889–897.

Laughton, C.A., Jenkins, T.C., Fox, K.R., & Neidle, S. (1990). Interaction of beneril with the ty3T DNA sequence studied by footprinting and molecular modelling. Implications for the design of sequence-specific DNA recognition agents. *Nucleic Acids Research, 18*, 4479–4488.

Leach, A.R. (2001). Molecular modelling: Principles and applications. Harlow: Addison–Wesley Longman Ltd.

Lee, C., Yang, W., & Parr, R.G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B, 37*, 785–789.

Li, L.H., DeKoning, T.F., Kelly, R.C., Krueger, W.C., McGovern, J.P., Padbury, G.E., ... Gebhard, I. (1992). Cytotoxicity and antitumor activity of carzelesin, a prodrug cyclopropylpyrroloindole analogue. *Cancer Research, 52*, 4904–4913.

Lombardy, R.L., Tanius, F.A., Ramachandran, K., Tidwell, R. R., & Wilson, W.D. (1996). Synthesis and DNA interactions of benzimidazole dications which have activity against opportunistic infections. *Journal of Medicinal Chemistry, 39*, 1452.

Lorusso, D., Mainenti, S., Pietragalla, A., Ferrandina, G., Foco, G., Masciuollo, V., & Scambia, G. (2009). Brostallicin (PNU-166196), a new minor groove DNA binder: Preclinical and clinical activity. *Expert Opinion on Investigational Drugs, 18*, 1939–1946.

Lown, J.W., Krowicki, K., Balzarini, J., Newman, R.A., & Declercq, E. (1989). Novel linked antiviral and antitumor agents related to netropsin and distamycin: Synthesis and biological evaluation. *Journal of Medicinal Chemistry, 32*, 2368–2375.

Mallena, S., Lee, M.P., Bailly, C., Neidle, S., Kumar, A., Boykin, D.W., & Wilson, W.D. (2004). Thiophene-based diamide forms a ‘super’ at binding minor groove agent. *Journal of the American Chemical Society, 126*, 13659–13669.

McDonnell, G., & Russell, A.D. (1999). Antiseptics and disinfectants: Activity, action, and resistance. *Clinical Microbiology Reviews, 12*, 147–179.

McKinney, J.D., Richard, A., Waller, C., Newman, M.C., & Gerberick, F. (2000). The practice of structure activity relationships (SAR) in toxicology. *Toxicological Sciences, 56*, 8–17.

Nikolaenko, T.Y., Bulavin, L.A., & Hovorun, D.M. (2011). How flexible are dna constituents? The quantum mechanical study. *Journal of Biomolecular Structure and Dynamics, 29*, 563–575.

Othman, O.E., & Ahmed, S. (2004). Cytogenetic effect of the trypanocidal drug beneril in blood cultures of river buffalo. *Journal of Biological Sciences, 4*, 180–184.
Parr, R.G., Donnelly, R.A., Levy, M., & Palke, W.E. (1978). Electronegativity: The density functional viewpoint. Journal of Chemical Physics, 68, 3801.

Parr, R.G., & Pearson, R.G. (1983). Absolute hardness: Companion parameter to absolute electronegativity. Journal of the American Chemical Society, 105, 7512–7516.

Parr, R.G., Szentpaly, L., & Liu, S. (1999). Electropolarity index. Journal of the American Chemical Society, 121, 1922–1924.

Parthasarathi, R., Subramanian, V., Roy, D.R., & Chatteraj, P.K. (2004). Electropolarity index as a possible descriptor of biological activity. Bioorganic & Medicinal Chemistry, 12, 5533–5543.

Pasha, F.A., Muddassar, M., & Cho, S.J. (2009). Molecular docking and 3D QSAR studies of Chk2 inhibitors. Chemical Biology and Drug Design, 73, 292–300.

Pasha, F.A., Srivastava, H.K., & Singh, P.P. (2005). Comparative QSAR study of phenol derivatives with the help of density functional theory. Bioorganic & Medicinal Chemistry, 13, 6823–6829.

Pearson, R.G. (1997). Chemical hardness: Applications from molecules to solids. Weinheim: Wiley-VCH.

Pezzioli, G., Grandi, M., Biasoli, G., Capolongo, L., Ballinari, D., Giuliani, F.C., … Spreafico, F. (1991). Biological profile of FCE 24517, a novel benzoyl mustard analogue of distamycin A. British Journal of Cancer, 64, 1047–1050.

Purschke, M., Rubio, N., Held, K.D., & Redmond, R.W. (2010). Phototoxicity of Hoechst 33342 in time-lapse fluorescence microscopy. Photochemical & Photobiological Sciences, 9, 1634–1639.

Puzanov, I., Lee, W., Chen, A.P., Calcutt, M.W., Hachev, D.L., Vermeulen, W.L., … Rothenberg, M.L. (2011). Phase I pharmacokinetic and pharmacodynamic study of SIG-136, a novel DNA sequence selective minor groove cross-linking agent, in advanced solid tumors. Clinical Cancer Research, 17, 3794–3802.

Ravindra, G.K., Achaiah, G., & Sastry, G.N. (2008). Molecular modeling studies of phenoxyoxymidimyl imidazoles as p38 kinase inhibitors using QSAR and docking. European Journal of Medicinal Chemistry, 43, 830–838.

Renneberg, D., & Dervan, P.B. (2003). Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove binders. Journal of the American Chemical Society, 125, 5707–5716.

Schultz, T.W., Cronin, M.T.D., Walker, J.D., & Aptula, A.O. (2003). Quantitative structure-activity relationships (QSARs) in toxicology: A historical perspective. Journal of Molecular Structure: THEOCHEM, 622, 1–22.

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

Srinivasan, J., Miller, J., Kollman, P.A., & Case, D.A. (1998). Continuum solvent studies of the stability of RNA hairpin loops and helices. Journal of Biomolecular Structure and Dynamics, 16, 671–682.

Srivani, P., Srinivas, E., Raghu, R., & Sastry, G.N. (2007). Molecular modeling studies of pyridopurinone derivatives – Potential phosphodiesterase 5 inhibitors. Journal of Molecular Graphics and Modelling, 26, 378–390.

Srivastava, H.K. (2009). A comparative QSAR study of alkanes with the help of computational chemistry. Bulletin of the Korean Chemical Society, 30, 67–76.

Srivastava, H.K., Chourasia, M., Kumar, D., & Sastry, G.N. (2011). Comparison of computational methods to model DNA minor groove binders. Journal of Chemical Information and Modeling, 51, 558–571.

Srivastava, H.K., Pasha, F.A., Mishra, S.K., & Singh, P.P. (2009). Novel applications of atomic softness and QSAR study of testosterone derivatives. Medicinal Chemistry Research, 18, 455–466.

Stuedli, A.N., Boykin, D., Tidwell, R.R., & Brun, R. (2011). Novel diaminides with activity against Babesia Divergens in vitro and Babesia Microti in vivo. Antimicrobial Agents and Chemotherapy, 55, 3439–3445.

SYBYL 6.9.2; Tripos Inc. (2004). 1699 South Hanley Road, St. Louis, MO 63144–2917.

Thompson, J.J., & Lill, M.A. (2011). Solvent interaction energy calculations on molecular dynamics trajectories: Increasing the efficiency using systematic frame selection. Journal of Chemical Information and Modeling, 51, 2680–2689.

Tidwell, R.R., Jones, S.K., Naiman, N.A., Berger, L.C., Brake, W.B., Dykstra, C.C., & Hall, J.E. (1993). Activity of cationically substituted bis-benzimidazoles against experimental Pneumocystis carinii pneumonia. Antimicrobial Agents and Chemotherapy, 37, 1713–1716.

Vijay, D., & Sastry, G.N. (2008). Exploring the size dependence of cyclic and acyclic -systems on cation-pi binding. Physical Chemistry Chemical Physics, 10, 582–590.

Walker, D.L., Reid, J.M., & Ames, M.M. (1994). Preclinical pharmacology of bizelesin, a potent bifunctional analog of the DNA-binding antibiotic CC-1065. Cancer Chemotherapy and Pharmacology, 34, 317–322.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., & Case, D.A. (2004). Development and testing of a general Amber force field. Journal of Computational Chemistry, 25, 1157–1174.

Warren, G.L., Andrews, C.W., Capelli, A.M., Clarke, B., Lalonde, J., Lambert, M.H., … Head, M.S. (2006). A critical assessment of docking programs and scoring functions. Journal of Medicinal Chemistry, 49, 5912–5931.

Warren, C.L., Kratochvil, N.C., Hauschild, K.E., Foister, S., Brezinski, M.L., Dervan, P.B., … Ansari, A.Z. (2005). Defining the sequence-recognition profile of DNA-binding molecules. Proceedings of the National Academy of Sciences USA, 103, 867–872.

Yarungko, Y.P., Zhurakivsky, R.O., Samijlenko, S.P., & Hovorun, D.M. (2011). Intramolecular CH···O hydrogen bonds in the AI and BI DNA-like conformers of canonical nucleosides and their Watson-Crick pairs. Quantum chemical and AIM analysis. Journal of Biomolecular Structure and Dynamics, 29, 51–65.

Zakrzewska, K., Lavery, R., & Pullman, B. (1984). The solvation contribution to the binding energy of DNA with non-intercalating antibiotics. Nucleic Acids Research, 12, 6559–6574.

Zimmer, C. (1975). Effects of the antibiotics nitropepa and distamycin-A on the structure and function of nucleic acids. Progress in Nucleic Acid Research and Molecular Biology, 15, 285–318.

Zwanzig, R.W. (1954). High-temperature equation of state by a density functional theory. Physical Chemistry Chemical Physics, 22, 3780–3790.