Molecular Docking and Molecular Dynamics Study of DNA Minor Groove Binders

Ruchi Mishra[1], Anamika Singh Gaur[2], Ramesh Chandra[3,*], Devesh Kumar[4,*]

[1,3,4]Department of Applied Physics, School for Physical Sciences, Babasaheb Bhimrao Ambedkar University, Vidya Vihar, Rae Bareilly Road, Lucknow, [2]Centre for Molecular Modeling, CSIR-Indian Institute of Chemical Technology, Tarnaka, Hyderabad

***Corresponding Author:**
Email: dkclcre@yahoo.com, ramesh.luphy@gamil.com

ABSTRACT

The fundamental problems in drug discovery are based on the process of molecular recognition by small molecules. The binding specificity of DNA-small molecule is identified mainly by studying the hydrogen bonding and polar interactions. Majority of the minor groove binders and their mechanism of action at the molecular level are not well studied. As these small molecules can act as effective therapeutic agents against many diseases, there is a need to have the detailed mechanistic insights on how they interact with DNA. In this study we have investigated the binding mechanism and stability of the complexes using molecular modeling methods. The molecular docking studies were performed to explore the exact binding sites and affinity inside the DNA minor grove. A 5ns molecular dynamics (MD) simulation for the DNA minor groove binders has been performed using AMBER and GROMACS program. Further, to study the systematic deviation of docked complexes during MD simulation, RMSD as a function of time have been analyzed and it has been found that RMSD variation obtained using AMBER and GROMACS MD simulation are approximately same. The binding free energies between the DNA and minor groove binders were calculated and decomposed by molecular mechanics/generalized born surface area (MM-GBSA) and Molecular Mechanics/Poisson−Boltzmann Surface Area (MM-PBSA) methods. The comparative and systematic analysis presented in this study can provide guidance for the choice of MD methods and the designs of new potent inhibitors targeting DNA.

Keywords: *Minor groove binders, Molecular docking, Molecular dynamics (MD), MM-GBSA method, MM-PBSA method.*

INTRODUCTION

Deoxyribonucleic acid (DNA) is the bio molecule has two complementary helical strands running in anti-parallel directions, which carries genetic information from parents to offspring[1]. DNA plays an important role in cellular processes, including cell division (DNA replication) and protein synthesis (Transcription and translation). Most of the anticancer therapies are involved in the interaction of drugs with DNA. The intercalation and groove binding are the two important modes of binding of drug with DNA. Both covalent and non-covalent types of interactions are possible in these two binding modes. Small molecules that can bind between nucleic acid base pairs are categorized as intercalators. These molecules contain planar heterocyclic groups which stack between adjacent DNA base pairs, which results decrease in the DNA helical twisting and lengthening of the DNA. On the other hand, groove binding does not induce large conformational changes in DNA and may be considered similar to standard lock and key models for ligand-macromolecular binding. Such molecules bind to both major and minor groove of nucleic acid. Minor groove binders are crescent in shape and they complement the shape of minor groove[2-5]. The binding mechanism of drug with the DNA minor groove can be described in mainly two steps. In the first step, the transfer of ligand to the DNA minor groove by the electrostatic and hydrophobic interaction. In the second step, various types of non-covalent interactions occur between the ligand and the functional groups of DNA base pairs. These interactions usually include hydrogen bonds, hydrophobic and van der Waals contacts, and electrostatic interactions. Most of the minor groove binding drugs bind to A/T rich region[6-9].

In the present study, two major DNA minor groove binders classes, polyamides and diary lamidines is undertaken. Different interaction models are available to explain protein-ligand binding but these models is not reasonable for DNA-ligand systems because there is no prescribed active site in the DNA, unlike protein/enzymes. Our study examines the available popular molecular modeling approaches for the

estimation of DNA- drug binding free energies and compares with the experimental results. Molecular modeling methods are a powerful tool to investigate various types of non-covalent interactions, which exist between the receptor and ligand. In the number of studies for the minor groove binders, these computational methods have shown good agreement with the experimental results.[10-14]

The special approaches like molecular dynamics simulation is required to understand the complex systems like nucleic acids. The force field used in various studies to simulate nucleic acid simulation includes, CHARMM, AMBER, GROMOS, OPLS, ENCAD and BMS26. There have been many studies for the comparison of force fields for the nucleic acids but still there is need to analyze them critically at the molecular level. AMBER 03 and CHARMM force fields were chosen for our study to simulate the DNA with small molecules. In this paper, we have used the popular molecular mechanics energies combined with the Generalized-Born surface area (MM-GBSA) and Poisson−Boltzmann Surface Area (MM-PBSA) methods to estimate the binding free energy of the binding of small molecules to DNA. These methods have been applied to wide range of molecules to estimate their ligand binding affinities

MATERIALS AND METHODS

Dataset

The crystal data of the B-DNA (1D30, 195D, 1D86 and 102D) were downloaded from the Protein Data Bank[10] and their experimental binding energies were collected from literature[11]. The water molecules and the ligands were removed from the 1D30, 195D, 1D86 and 102D. The drug molecules extracted from these complexes were subjected to geometry optimization using Gaussian 09 at B3LYP/6-31G* level[12].

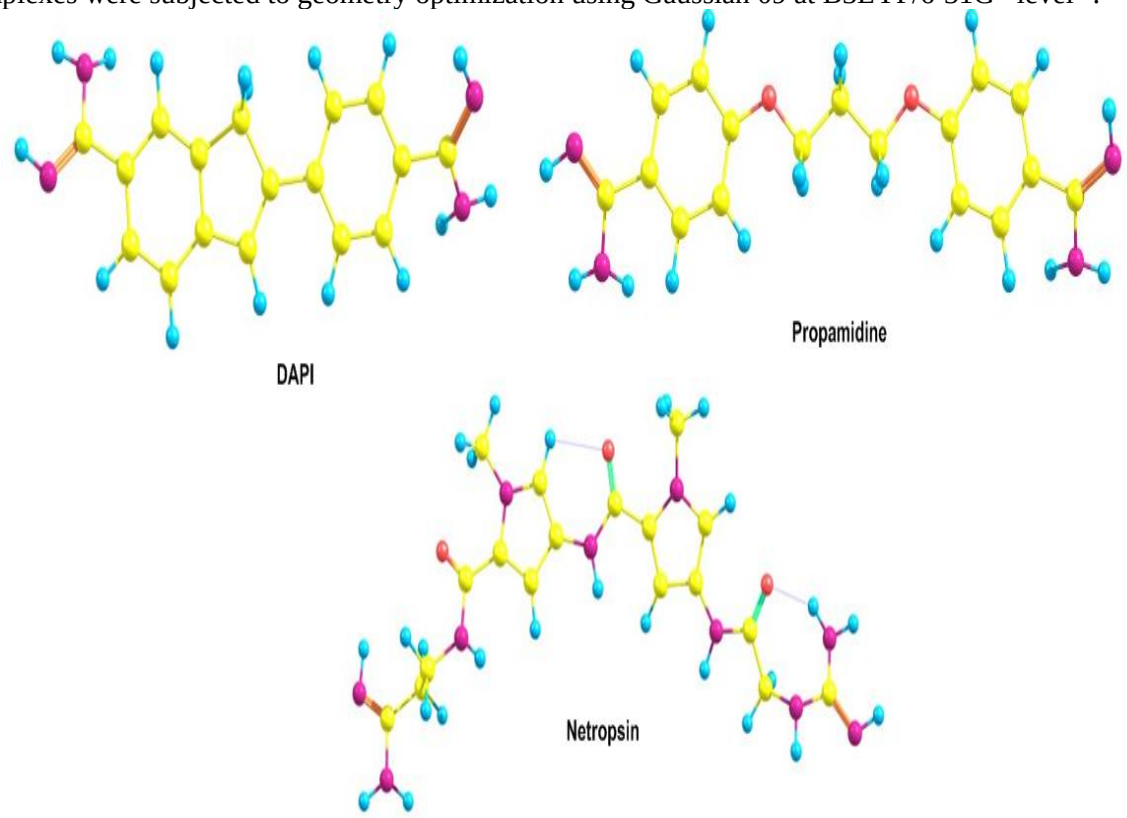

Fig. 1: Optimized structures of DNA minor groove Binders using Gaussian 09 at B3LYP/6-31G* level

Molecular Docking

Molecular docking studies were performed using Autodock 4.2 program[13]. The Gasteiger charges were added to the complex by Autodock Tools (ADT) before performing docking calculations. The binding site was centered on the macromolecule and a grid box was created with 55 × 55 × 60 points and a 0.375 Å grid spacing in which almost covered the entire DNA were involved. Docking simulations were performed using

the classical Lamarckian genetic algorithm (LGA). The 20 LGA runs with maximum of 2500000 energy evaluations performed. In addition, the other parameters were set to default. The pose with lowest energy of binding or binding affinity was extracted and aligned with receptor structure for further analysis.

Simulations

Molecular Dynamics simulation of 5 ns was carried out using the AMBER 15[14-17] and GROMACS 4.5.6 packages[18]. AMBER15 program was used for MD simulations of the selected docked poses. The 'leaprc.gaff" (generalized amber force field) was used to prepare the ligands, while the 'leaprc.ff03' was used for the DNA. The 'add ions' command implemented in 'tleap' of AMBER15 was used to add the Na+ ions explicitly to neutralize the system. Each system was placed in a box of TIP3P water by using 'solvateOct' command with the minimum distance between any solute atom and the boundary of the box is set to 8Å. Energy minimization was then performed to achieve the nearest stable low energy conformations (500 steps of each steepest descent and conjugate gradient method), 50 ps of heating and 50ps of density equilibration with weak restraints on the complex followed by 500ps of constant pressure equilibration at 300K. Cut off size of 12 Å was used for MD simulations. All long –range electrostatics were included by means of a Particle mesh Ewald (PME) method. All hydrogen and heavy atom bonds were constrained by the shake method, and simulations were performed with a 2fs time step and langvenin dynamics was used for temperature control. The same conditions for the final phase of equilibration were used for the production run and the coordinates were recorded at every 10 ps. Periodic boundary conditions were used for the final production run. Five hundred snapshots of the complex are obtained at every 10 ps from the MD trajectories, and all the water molecules and ions were removed before MMPBSA/MMGBSA[19] calculations using the "extract_coords.mmpbsa" script and the $\Delta G_{bind} - PB/GB$ values were calculated using the "binding_energy.mmpbsa" script.

In GROMACS MD simulation the topology and co-ordinate files for the DNA were generated by pdb2gmx program of the GROMACS package taking parameters from CHARMM[20] all atom force field and for the ligand using SwissParam[21] Web server. The coordinate and topology files of DNA and ligand were merged to obtain the final starting structure and topology file for each complex. The drug-DNA complex was placed in the center of dodecahedron periodic box. The system was then solvated in TIP3P water Molecules. The total charge on the system was then neutralized by adding counter ions. The energy was minimized using steepest descent algorithm. Then the system was heated to 300K during 50ps of constant volume simulation with 2fs time step. The pressure was equilibrated to 1 atm during 50ps NPT simulation with 2 fs time step. In both the simulations a restrained with force constant of 1000kJ/ (mol/nm^2). Both temperature and pressure were regulated using Berendsen algorithm. Production simulations were performed 5ns with a 2fs time step. The temperature and pressure were maintained at 300 K and 1 atm using the v-rescale temperature and Parninello-Rahman pressure coupling method. The binding energy was calculated using g_mmpbsa[22]. Binding energy of each snapshot was calculated for each complex. The entropy contribution was not included in the binding energy.

RESULTS AND DISCUSSIONS

Molecular docking studies

Molecular docking calculations shows that all the ligands bind to AT-rich region of DNA with good docking fitness score (Table 1). Comparison between experimental and calculated binding energies obtained from docking studies shows that the complex having lowest experimental binding energy also has the lowest calculated binding energy and vice-versa. This shows the ability of Autodock in predicting the correct binding modes for drug DNA complex. The PDB ID **1D30** have lowest binding energy, this shows that the interaction between DNA duplex of sequence d (CGCGAATTCGCG) $_2$ and DAPI (fig. 2) is more stable in comparison to other complexes. The further studies were performed on the best pose with lowest binding energy for all complexes.

Table 1: List of PDB IDs taken for the study with their DNA sequence, calculated binding energies (kcal/Mol) and experimental binding energies obtained from literature

PDB ID	DNA sequence	Ligand	$\Delta G_{calc.}$[a]	$\Delta G_{exp.}$[b]
1D30	5'-CGCGAATTCGCG-3'	DAPI	-9.07	-8.8
1D86	5'-CGCGAATTCGCG-3'	Netropsin	-5.42	-8.7
195D	5'-CGCGTTAACGCG-3'	Netropsin	-4.35	-8.0
102D	5'-CTTTTGCAAAAG-3'	Propamidine	-4.61	-8.2

[a] Calculated binding Energy in kcal/mol.
[b] Experimental binding energy kcal/mol.

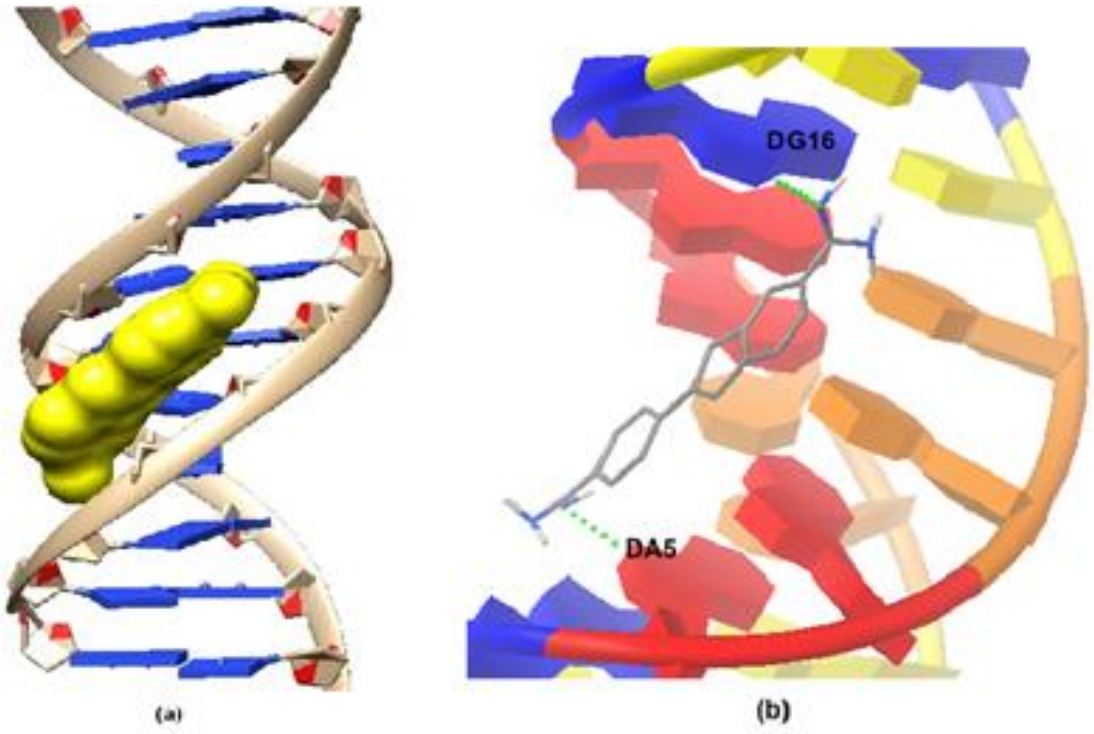

Fig. 2: A molecular docked model for DAPI with DNA duplex of sequence d (CGCGAATTCGCG) $_2$ (PDB ID: 1D30). (a)The full view of docking between DAPI and 1D30; (b) The binding mode between DAPI and 1D30 and the green dashed line showing hydrogen bond interactions between them

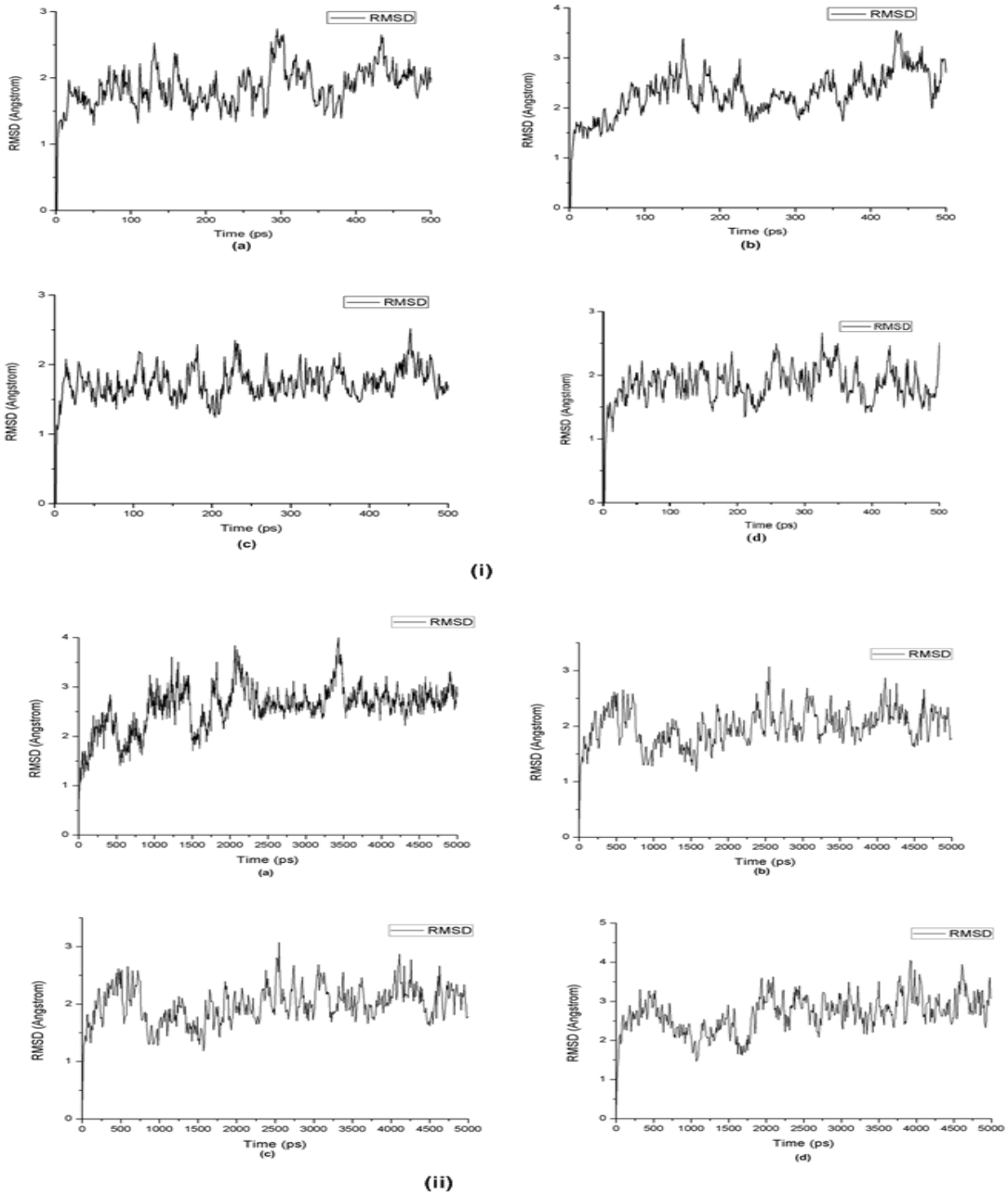

Fig. 3: Plots of RMSD vs. time of their trajectory for PDB ID (a) 1D30 (b) 1D86 (c) 102D (d) 195D obtained from (i) AMBER (ii) GROMACS

Molecular dynamics studies

MD simulation from AMBER15 and GROMACS 4.5.6 were performed for the best poses selected from the docking studies. RMSD as a function of time is plotted for all the complexes in fig 3 to obtain the systematic deviation of complexes. The RMSD profile shows that the ligands remain bound to the DNA near the preferential binding site. It has been observed that the all the four complexes show RMSD variation from both types of simulations. The range of RMSD from fig. 3 is between 1Å-3Å from AMBER and 1Å-3.5Å for GROMACS during the course of simulation. Convergence of RMSD values shows the stability of complex. In order to identify the conformational changes during 5ns of MD simulation, the snapshots were extracted from the MD trajectory at every 10 ps for both AMBER and GROMACS (fig 4). It has been

observed from snapshots that ligand forms interactions with AT-rich region of DNA duplex and bounded in minor groove up to the end of simulation.

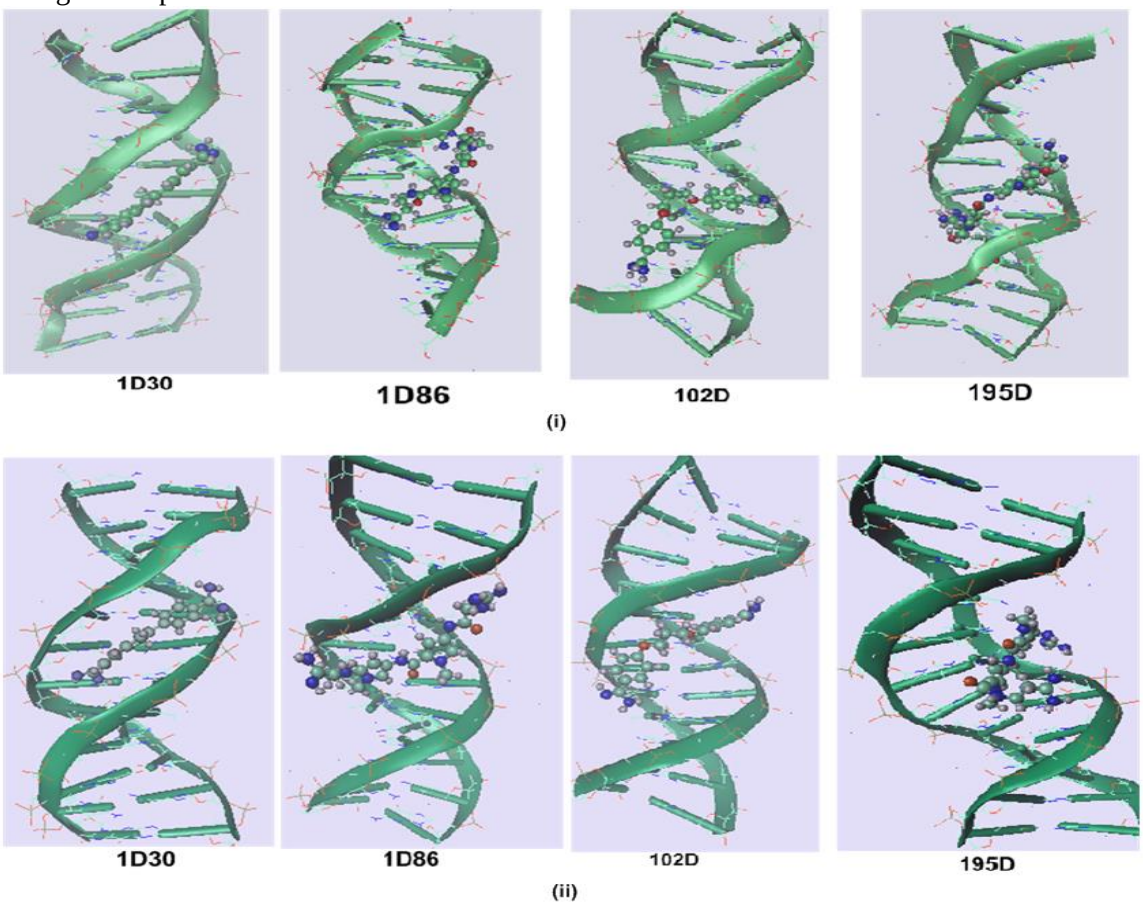

Fig. 4: Snapshots of different DNA-ligand complexes at 5 ns MD simulation for PDB ID (a) 1D30 (b) 1D86 (c) 102D (d) 195D obtained from (i) AMBER (ii) GROMACS

Further, MMPBSA/MMGBSA calculations were performed from AMBER 15 by using MD trajectories to obtain binding energy values. Binding energy values predict the strength of ligand with their respective receptors. Calculated binding free energy of all four drug-DNA complexes with the contribution of van der Waals, electrostatic, salvation energy etc. is shown in fig. 5 and it was found that the total Binding free energy of complex with PDB ID 1D30 is lowest binding energy (-25.52 kcal/mole), so this complex is slightly more stable than others.

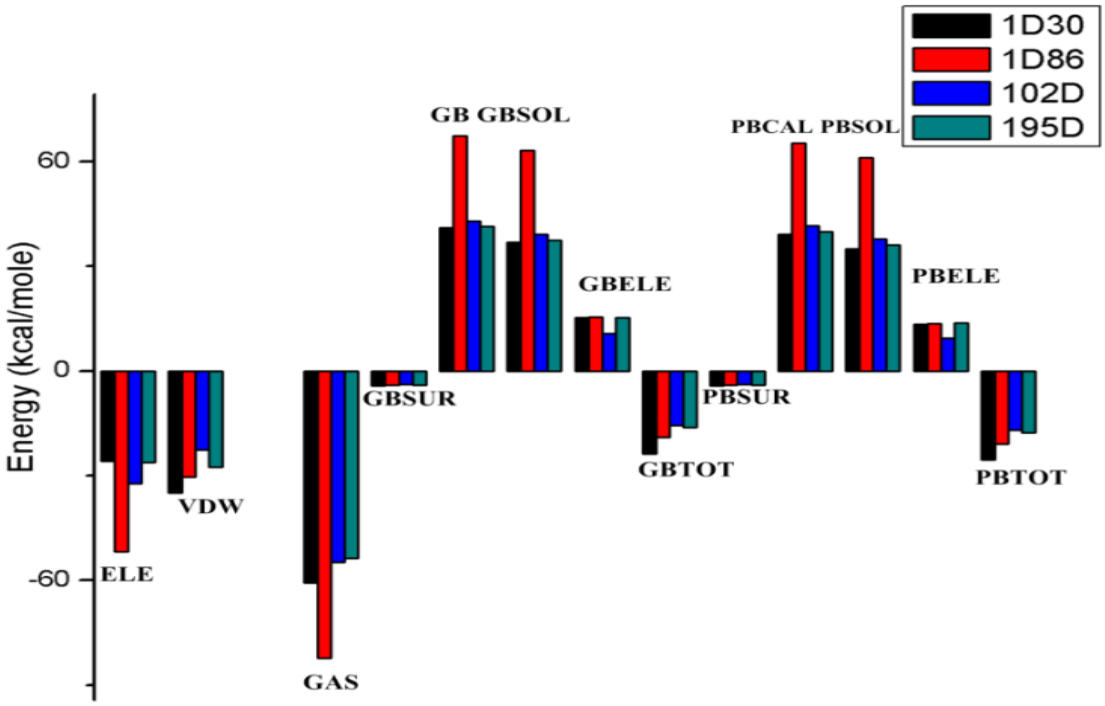

Fig. 5: Histogram depicting view of the contribution of various energy components to the binding free energy

ELE: electrostatic energy; **VDW**: van der Waals energy; INT: internal energy; GAS: total gas-phase energy; PBSUR/GBSUR: nonpolar contribution to the solvation free energy; PBCAL/GB: electrostatic contribution to the solvation free energy calculated by PB or GB, respectively; PBSOL/GBSOL: sum of non-polar and polar contribution to solvation; PBELE/GBELE: sum of the electrostatic solvation free energy and MM electrostatic energy; PBTOT/GBTOT: final estimated binding free energy.
The binding free energy has been also recalculated for all the four complexes using the g_mmpbsa for the MD trajectories obtained from GROMACS MD simulation. It has been observed that binding energy obtained from both types of programs that the PDB ID 1D30 having lowest binding energy. To solve the PB equation g_mmpbsa uses the APBS package whereas mm_pbsa.pl uses the PBSA program of the AMBER suite. It has been observed that the energies calculated using g_mmpbsa and the AMBER MM-PBSA package is approximately similar and the difference of 1-3 kcal/mol has been observed due to the difference ΔG_{polar} (Table 3). The difference in ΔG_{polar} observed because of different algorithms, implemented in APBS and PBSA.

The binding energy calculation shows that DNA-DAPI complex is more stable than others. So the MD structures of above complex were selected to study the hydrogen bonds which retained throughout the MD simulation. The results show that only few hydrogen bonds that were present in the original complex (energy minimized) were retained during the simulations. In the DNA-DAPI complex two hydrogen bonds between LIG25:H4-DT19:O2 and LIG25:H5-DT20-O4' were retained throughout the simulation using AMBER and GROMACS. It has been also observed that hydrogen bonds, mainly forms between minor groove binders and the functional groups of the bases exposed in the grooves via their end groups and also through their amide and linker group.

Table 2: Comparison of binding energy components obtained from AMBER MM-PBSA and g_mmpbsa

PDB Id	program	ΔE_{elec} [a]	ΔE_{vdw} [b]	ΔG_{polar} [c]	$\Delta G_{nonpolar}$ [d]	$\Delta G_{binding}$ [e]
1D30	mm_pbsa.pl	-25.80±6.54	-34.83±3.39	39.27±6.40	-4.15±0.22	-25.52±3.64
	g_mmpbsa	-12.01±5.18	-16.42±7.24	17.84±7.12	-1.71±0.58	-25.95±10.11
1D86	mm_pbsa.pl	-51.80±13.59	-30.43±4.80	65.35±12.44	-4.10±0.33	-20.98±6.58
	g_mmpbsa	-17.60±12.47	-27.36±12.61	25.41±14.16	-2.92±1.16	-22.46±13.18
102D	mm_pbsa.pl	-32.25±10.30	-22.56±3.63	41.63±8.77	-3.76±0.20	-16.94±4.06
	g_mmpbsa	-11.91±4.10	-26.51±5.13	20.65±5.75	-2.89±0.41	-20.66±7.33
195D	mm_pbsa.pl	-26.25±14.81	-27.46±3.69	39.99±14.48	-3.93±0.39	-17.65±4.36
	g_mmpbsa	-26.26±4.31	-34.99±3.61	37.72±6.65	-3.75±0.27	-27.30±4.17

[a] Electrostatic component to the binding energy in kcal/mol.
[b] van der Waals component to the binding energy in kcal/mol.
[c] Polar solvation energy.
[d] Non-polar solvation energy.
[e] Binding energy.

CONCLUSIONS

The computational studies were performed to evaluate and analyze the binding energies of DNA minor groove binders. The focus of the study is to provide a detailed perspective on drug-DNA interactions at the molecular level. Thus, our study attempts to give detail insight on the complexity in binding modes of small molecules to DNA. Theses analysis will be helpful for the improvement of existing minor groove binders, and also in designing novel chemical entities which can act as good DNA inhibitor. This will provide a theoretical protocol for complementing experimental techniques, generation of database for structure-energy relationship in drug-DNA complexes.

ACKNOWLEDGMENTS

RM likes to acknowledge UGC for the financial support. DK is DST-Ramanujan Fellow (SR/S2/RJN-11/2008). The authors are thankful to Dr G. Narahari Sastry for his support and discussions.

REFERENCES

1. Watson JD, Crick FHC, Nature, 1953,171:737–738.
2. Song YM, Wu Q, Yang PJ, Luan NN, Wang LF, Liu YM, J. Inorg. Biochem., 2006,100:1685–1691.
3. Dervan PB, Bioorg. Med. Chem., 2001,9:2215–2235.
4. Dervan PB, Edelson BS, Curr. Opin. Struct. Biol., 2003,13:284–299.
5. Masta A, Gray PJ, Phillips DR, Nucleic Acids Res., 1995, 23:3508–3515.
6. Tse WC, Boger DL, Chemistry&Biology, 2004,11:1607-1617.
7. Nunn MC, Neidle S, J. Med. Chem., 1995,38:2317-2325.
8. Kennard O, Pure &Appl. Chem., 1993,65:1213-1222.
9. Srivastava HK, Chourasaia H, Kumar D, Sastry GN, Chem. Inf. and Model, 2011,51:558-571.

10. Kamal A, Shetti RV, Rmaiah MJ, Swapna P, Reddy KS, Mallareddy A, Rao MPN, Chourasia M, Sastry GN, Juvekar A, Zingde S, Sarma P, Pushpavalli SN, Bhadra MP, Med. Chem. Comm., 2011,2:780-788.
11. Kamal A, Shankaraiah N., Reddy Ch. R, Prabhakar S, Markandeya N, Srivastava HK, Sastry GN, Tetrahedro*n*, 2010,66:5498-5506.
12. Kamal A, Reddy KS, Khan MNA, Shetti RVCRNC, Ramaiah MJ, Pushpavalli SNC VL, Srinivas, C, Pal-Bhadra, M, Chourasia M, Sastry GN, Juvekar A, Zingde S, Barkume M, Bioorg. Med. Chem., 2010,18:4747-4761.
13. Kamal A, Bharathi EV, Ramaiah MJ, Dastagiri D, Reddy JS, Viswanath A, Sultana F, Pushpavalli SNCVL, Pal-Bhadra M, Srivastava HK, Sastry GN, Juvekar A, Sen S, Zingde S. Bioorg. Med. Chem., 2010**,**18:526-542.
14. Kamal A, Rajender, Reddy DR, Reddy MK, Balakishan G, Shaik, TB, Chourasia M, Sastry GN, Bioorg. Med. Chem., 2009,17:1557-1572.
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE ."The Protein Data Bank". Nucleic Acids Res. 2000,28:235–242.
16. Shaikh SA, Jayaram B, J. Med. Chem., 2007,50:2240–2244.
17. Gaussian 09, Revision E.01, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, .Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J , Fox DJ, Gaussian, Inc., Wallingford CT, 2009.
18. Morris G, Goodsell D, Halliday R, Huey R, Hart W, Belew R, Olson AJ, J. Comput. Chem., 1998,19:1639–1662.
19. Case D. et al. AMBER 15, University of California, San Francisco, 2015.
20. Salomon-Ferrer R, Case DA, Walker RC, WIREs Comput. Mol. Sci., 2013,3:198-210.
21. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B and Woods R. J. Computat. Chem., 2005,26:1668-1688.
22. Cheatham TE, III and Case DA. Biopolymers, 2013,99:969-977.
23. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hes B, Lindahl E, Bioinformatics, 2013,29:845−854.
24. Thompsan JJ, Lill MA, Journal of Chemical Information and Modeling, 2011,51:2680–2689.
25. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, uvench OG, Lopes P, Vorobyov I, MacKerell Jr. AD, Journal of Computational Chemistry, 2010,31:671-690.
26. Zoete V, Cuendet MA, Grosdidier A, Michielin O, J. Comput. Chem., 2011,32:2359-68.
27. Kumari et al J. Chem. Inf. Model., 2014,54:1951-1962.