

## DNA minor groove binding of small molecules: Experimental and computational evidence

PRATEEK PANDYA<sup>1</sup>, MD MAIDUL ISLAM<sup>2</sup>, G SURESH KUMAR<sup>2</sup>, B JAYARAM<sup>3</sup> and SURAT KUMAR<sup>1,\*</sup>

<sup>1</sup>Department of Chemistry, Faculty of Science, Dayalbagh Educational Institute, Dayalbagh, Agra 282 005

<sup>2</sup>Biophysical Chemistry Laboratory, Indian Institute of Chemical Biology, Kolkata 700 032

<sup>3</sup>Department of Chemistry, Indian Institute of Technology, New Delhi 110 016

e-mail: kumar.surat@gmail.com

MS received 10 June 2009; accepted 26 August 2009

**Abstract.** Eight indole derivatives were studied for their DNA binding ability using fluorescence quenching and molecular docking methods. These indole compounds have structural moieties similar as in few indole alkaloids. Experimental and theoretical studies have suggested that indole derivatives bind in the minor groove of DNA. Thermodynamic profiles of DNA complexes of indole derivatives were obtained from computational methods. The complexes were largely stabilized by H-bonding and van der Waal's forces with positive entropy values. Indole derivatives were found to possess some Purine (Pu) – Pyrimidine (Py) specificity with DNA sequences. The results obtained from experimental and computational methods showed good agreement with each other, supported by their correlation constant values.

**Keywords.** Indole derivatives; DNA binding; DNA interaction; fluorescence; drug-DNA binding; indole-DNA binding, Pu–Py specificity; non-specific DNA binding.

### 1. Introduction

Many alkaloids contain indole as one of the important component in their structure. Several biologically important molecules like tryptophan, tryptamine, serotonin, gramine, Hoechst-33258, etc. possess indole unit in their structure. Many of these molecules have been extensively used<sup>1,2</sup> in medicinal preparations. It becomes obvious that indole ring plays an important role in the interaction of these molecules with their biological receptors. The binding modes of several biologically important indole derivatives to DNA have been studied and summarized by several workers.<sup>3–6</sup> Rich, Dervan, Wemmer, Lown and other workers have described Netropsin, Distamycin A, and Hoechst-33258, etc. as DNA minor groove binding drugs.<sup>7–10</sup>

It is observed that tryptophan and indole do not intercalate with either single-stranded DNA or double stranded DNA.<sup>3–6</sup> However, the naturally occurring amines tryptamine and gramine have been shown to intercalate between the DNA base pairs.<sup>11</sup>

Transcriptional regulation of abnormal gene expression by chemical means has become an important area in human medicine.<sup>12,13</sup>

There are several ligands derived from pyrrole/imidazole derivatives which are known to bind selectively in the minor groove of DNA. Bisbenzimidazole compounds are known to bind selectively in the DNA minor groove with AT sequence selectivity.<sup>14</sup>

#### 1.1 Chemical rationale for selection of compounds

Studies have provided significant insight into the minor groove binding of small molecules, which possess charged centers.<sup>15,16</sup> In addition, drug like systems also possess long side chains besides the main indole ring in their structures. The knowledge about the DNA binding affinity of main indole ring and of the simpler analogues of indole is not fully understood. We have tried to explore the minor groove binding profile of such indole derivatives with the help of experimental and docking methods.

\*For correspondence

In our study, all the indole derivatives possess only one indole ring. The objective was not to increase the overall stacking forces in the molecule, thereby, specifically targeting the minor groove of DNA. Based on the above considerations two main types of indole derivatives were included, viz. (i) compounds having side chain attached to position 2 of the indole ring and (ii) compounds having side chain present at position 3 on the indole nucleus. The side chains attached in both the cases included 0–3 methylene ( $-\text{CH}_2-$ ) groups present between the terminal  $-\text{COOH}$  or  $-\text{CONH}_2$  group and the indole ring. It is evident that nitrogen present at position 1 of indole ring is a potent H-bond donor. Besides, the side chain attached to the indole ring also contains strong H-bond acceptor ( $>\text{C}=\text{O}$ ) and H-bond donor ( $-\text{OH}$ ,  $-\text{NH}_2$ ) groups. The H-bond donor or acceptor groups are not only facilitating in establishing the intermolecular contacts between small molecule and DNA oligomer, but also determine the DNA sequence specificity.

The side chain position was important in selecting the compounds because side chain attached at position 2 of the indole ring brings two potential H-bond donor/acceptor atoms near each other (indole NH and terminal  $-\text{COOH}/-\text{NH}_2$ ). On the other hand, the side chain attached at position 3 increases the distance between the two H-bond donor/acceptor groups.

Apart from the position, the length of the side chain provided two important characteristics, viz. (i) flexibility of the indole nucleus with respect to the terminal group of the side chain and (ii) increase in the distance between indole N1 atom and the terminal carboxylic acid group of the side chain (e.g. indole-3-acetic acid to Indole-3-butanoic acid). The increase in methylene linker ( $-\text{CH}_2-$ ) length gave the molecules better flexibility and increased distance between two electronegative parts (Indole NH and terminal  $-\text{COOH}/-\text{NH}_2$  of the side chain). This flexibility furnished by methylene linker plays an important role in the orientation of the molecule inside the DNA minor groove while H-bond donor/acceptor atoms in the side chain provide an anchor through H-bonds with the DNA backbone. The indole NH is the main H-bond donor group which forms H-bonds with the DNA bases in the DNA minor groove.

Keeping the above information in mind, a study has been designed for assessing the DNA binding affinity of indole derivatives to generate a comprehensive database of their DNA binding pattern. Such structural information profiling of DNA binding be-

haviour of indole derivatives may help us decipher the biological properties associated with this class of compounds. A group of indole derivatives selected for this study includes 4-(1H-indol-3-yl)butanoic acid (**I**) 3-(1H-indol-3-yl)propanoic acid (**II**), 1H-indol-3-ylethanoic acid (**III**), 1H-indol-2-ylcarboxylic acid (**IV**), 2-(1H-indol-3-yl)ethanamide (**V**) 1H-benzimidazol-2-ol (**VI**), 1H-benzimidazole, (**VII**) and 2,3-dihydro-1H-indole (**VIII**), as shown in figure 1. In these compounds, indole ring has a carboxylate group on its position 2 or 3. This was intended to furnish the H-bond making ability of these indole derivatives so that they may form one or two H-bonds during their interaction with DNA duplex in its minor groove. Research groups of Dervan and Lown have reported the presence of H-bond formation augmenting the minor groove binding of drug molecules in addition to the electrostatic forces.<sup>8,10</sup>

A number of indole derivatives selected for the study have been known for a variety of pharmacological activities, like indol-3-ethanoic acid (**III**) is a plant growth hormone and indol-3-propanoic acid (**II**) protects against free radical-mediated injuries.<sup>17</sup> Few derivatives of 1H-indol-3-ethanamide (**V**) are potent and selective inhibitors of human sPLA<sub>2</sub>.<sup>18</sup> Benzimidazole (**VII**) derivatives like albendazole are known antihelminthic agents. Among them, indole ring is having a carboxylate group present on position 2 or 3 of indole.

## 2. Materials and methods

Indole derivatives, sonicated Calf thymus DNA (CT-DNA) and four DNA decamer sequences as desalted bases were purchased from Sigma-Aldrich Co. DNA-1, DNA-2, DNA-3 and DNA-4 were designed to have 4 base pair specific central core.

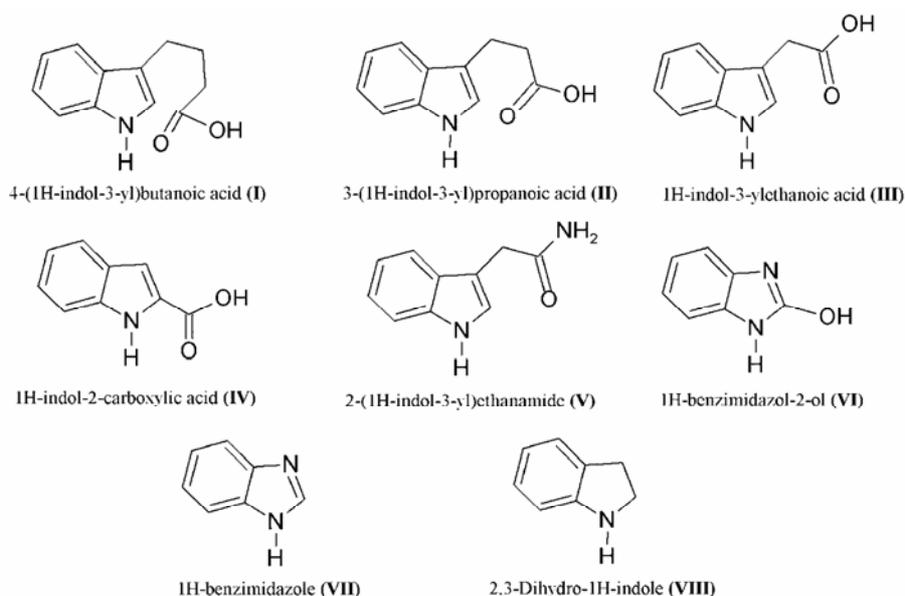
DNA-1: 5'-(d-GATGGCCATC)<sub>2</sub>

DNA-2: 5'-(d-GATCCGGATC)<sub>2</sub>

DNA-3: 5'-(d-GGCAATTGCC)<sub>2</sub>

DNA-4: 5'-(d-GGCTTAAGCC)<sub>2</sub>

The stock solutions of indole derivatives (**I–VIII**) and DNA decamers were made in 20 mM Sodium phosphate buffer (5 mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM (Na)<sub>2</sub>-EDTA and 3 mM NaCl) at pH 7.5. The concentration of DNA decamers (DNA-1 to DNA-4)



**Figure 1.** Chemical structures of indole derivatives used in the study.

were determined spectrophotometrically using the molar extinction coefficients,  $\epsilon_{260} = 95,000$  for DNA-1,  $\epsilon_{260} = 96,600$  for DNA-2,  $\epsilon_{260} = 92,600$  for DNA-3 and  $\epsilon_{260} = 93,200$  for DNA-4. Calf thymus DNA concentration in terms of nucleotide phosphate was estimated using  $\epsilon_{260} = 6600 \text{ M}^{-1} \text{ cm}^{-1}$ .

The concentrations of indole drugs (I–VIII) were calculated volumetrically. All the solutions were freshly prepared and stored below freezing point. All the compounds obeyed Beer's law in the concentration range used in the current study. UV Fluorescence technique was selected for evaluating the drug-DNA binding. Steady state fluorescence measurements were recorded on a Hitachi model F4010 spectrofluorimeter (Hitachi Ltd, Tokyo, Japan), where compounds (I–VIII) in a fixed concentration were titrated with increasing concentration of DNA decamers in the fluorescence-free quartz cuvette of 1 cm path length. The excitation and emission band pass were 5 nm each.

The fluorescence maximum for I was obtained at 364 nm upon excitation at 280 nm, for II was obtained at 367.4 nm upon excitation at 280 nm and that for III at 364 nm upon excitation at 275 nm. Compound IV showed fluorescence maximum at 359 nm upon excitation at 280 nm while V exhibited a fluorescence maximum at 356 nm upon excitation at 279 nm. Upon excitation at 278 nm, compound VI gave fluorescence maximum at 310 nm and VII furnished a maximum at 291.8 nm upon excitation at 243.6 nm. The fluorescence maximum for VIII was observed at 370 nm upon excitation at 237 nm. A

series of detailed fluorescence titration experiments of the selected eight indole derivatives were performed with four DNA decamer sequences (DNA-1 to DNA-4).

Consistent quenching was observed during all the titration experiments, which clearly indicated that the compounds employed in the study bind with DNA decamer sequences. The quenching data obtained from 32 titrations of drug-DNA complexes of eight indole compounds with 4 DNA decamer sequences was analysed by double reciprocal method for the evaluation of binding constants (double reciprocal plot between  $1/[\text{DNA}]$  versus  $1/E_0 - E$ ), since this method was useful in the case of non-specific ligand interactions with DNA oligomers.  $E_0$  is the initial emission maximum while  $E$  is the emission maximum after each addition of DNA aliquot. The titration of I–VIII with DNA decamers (DNA-1 to DNA-4) were performed until an apparent quenching of 10% of the initial emission was observed.

For the purpose of theoretical calculations, the pdb structures of indole derivatives were generated and other molecular modelling analysis was furnished using Discovery Studio<sup>®</sup> Visualizer desktop application program from Accelrys Inc. on an IBM computer with Xeon processor. Drug structures (pdb files) were uploaded along with the DNA base sequence onto the molecular docking server for docking calculations. The program 'DNADock' automatically generates the canonical DNA structure from the input base sequence.<sup>19</sup>

The docking method includes 4 steps: (a) identification of best possible grid/translational points in radius of 3 Å around the reference point (centre of mass); (b) generation of grid and preparation of energy grid in and around the active site of the DNA to pre-calculate the energy of each atom in the candidate ligand; (c) Monte Carlo docking and intensive configurational search of the ligand inside the active site; and (d) identification of the best docked structures on an energy criterion and prediction of the binding free energy of the complex, as reported.<sup>20</sup>

The selected docked complexes were energy minimized in vacuum by using AMBER force field.<sup>21</sup> For vacuum minimization, 1000 steps of steepest descent and 1500 steps of conjugate gradient were carried out. This resulted in some minor conformational changes in the DNA duplex.

The final drug–DNA complex structures were subjected to the binding free energy analysis using PreDDICTA software tool.<sup>22</sup> The program PreDDICTA appropriately gave the free energy of the DNA binding, the change in the DNA melting temperatures and various energy terms contributing to the overall DNA binding of drug. A total of 32 energy minimized docked structures of drug-DNA complexes were generated and analysed for their binding energies and other structural features.

### 3. Results

Fluorescence DNA binding titrations of eight indole derivatives (I–VII) were carried out first with calf thymus-DNA (CT-DNA). The decrease in the emission maxima of indole derivatives after each addition of CT-DNA aliquot indicated the interaction between these compounds and CT-DNA. Similar titrations were repeated with 4DNA decamer sequences.

The binding constants obtained from fluorescence titrations with DNA decamer sequences revealed the binding affinity of all the indole derivatives in the range of  $10^4$ – $10^5$  per mole. There were 13 drug–DNA complexes which showed binding constant values of the order of  $10^5$  per mole, while 19 complexes showed the binding constant values in the order of  $10^4$  per mole. The maximum binding affinity with all DNA decamer sequences (DNA-1 to DNA-4) was demonstrated by I, IV, while DNA-4 showed highest affinity with VI. This observation is especially significant in the light of theoretical binding energy calculations, since both of these com-

pounds (IV and VI) showed relatively stronger electrostatic contributions in their DNA binding as revealed by docking, using PreDDICTA module.

Molecular docking of indole derivatives on DNA decamer sequences resulted in the formation of 3-D docked structures. These structures were studied for various inter-molecular interactions and later subjected to energy analysis of the complexes.

Veal and Rill (1989) reported that the cleavage of DNA restriction fragments by Cu-complex was showing sequence preference for Py–Pu–Py triplets.<sup>23</sup> Another report on nucleoside transport system exhibited specificity for purine and pyrimidine nucleosides.<sup>24</sup> In these examples, Pu–Py specificity has been underlined rather than individual DNA base specificity.

The pattern for sequence specificity among all indole derivatives (I–VIII) was found to be absent indicating non-specific DNA binding of these compounds, which was also reflected by the similar values of  $K_{\text{exp}}$  with different DNAs (DNA-1 to DNA-4). In the light of non-specific DNA binding of drugs, the energy minimized structures of drug–DNA complexes were evaluated for purine-pyrimidine (Pu–Py) preferences in their DNA binding profiles. Since both purines (G and A) have same structural nucleus and both pyrimidines (T and C) have pyrimidine ring with similar electron density, thus a much broader specificity in terms of purine versus pyrimidine (Pu–Py) was considered for this case of non-specific drug-DNA binding. All eight compounds (I–VIII) showed Pu–Py specificity which is indicated in their preferred binding site on DNA decamer sequences (table I).

The present study underscored the good agreement with the results obtained from both experimental and theoretical methods. For comparing the two techniques for the binding intensity of the ligand–DNA interaction, the theoretical binding constants were secured from the theoretically obtained free energy values of docked structures by employing classical vant Hoff's equation.

$$\Delta G = -RT \ln k. \quad (1)$$

where  $\Delta G$  was the free energy,  $R$  is the gas constant,  $T$  is the temperature in Kelvin (300 K) and  $K$  is the equilibrium constant.

Various sets of  $\Delta G$  and  $K$  values from experimental and theoretical techniques were secured. The DNA binding constants were obtained from the fluorescence titrations henceforth termed as  $K_{\text{exp}}$  which

**Table 1.** DNA binding pattern of indole derivatives.

Drug	Complex	Binding site	Pu-Py Specificity	Binding pattern	$K_{\text{exp}}$ ( $\text{mol}^{-1}$ )	Inter-molecular H-bonds
Indole-3-butanoic acid (I)	DNA-1	5'-C <sub>6</sub> C <sub>7</sub> A <sub>8</sub>	5'-Py-Py-Pu	Same	$1.23 \times 10^5$	1
	DNA-2	5'-C <sub>4</sub> C <sub>5</sub> G <sub>6</sub>	5'-Py-Py-Pu		$1.96 \times 10^5$	1
	DNA-3	5'-T <sub>7</sub> G <sub>8</sub> C <sub>9</sub>	5'-Py-Pu-Py	Same	$1.33 \times 10^5$	1
	DNA-4	5'-A <sub>1</sub> G <sub>8</sub> C <sub>9</sub>	5'-Py-Pu-Py		$1.11 \times 10^5$	1
Indole-3-propanoic acid (II)	DNA-1	5'-C <sub>6</sub> C <sub>7</sub> A <sub>8</sub>	5'-Py-Py-Pu	3-Bases altered	$7.33 \times 10^4$	0
	DNA-2	5'-G <sub>7</sub> A <sub>8</sub> T <sub>9</sub>	5'-Pu-Pu-Py		$7.30 \times 10^4$	1
	DNA-3	5'-T <sub>6</sub> T <sub>7</sub> G <sub>8</sub>	5'-Py-Py-Pu	Same	$1.41 \times 10^5$	1
	DNA-4	5'-T <sub>4</sub> T <sub>5</sub> A <sub>6</sub>	5'-Py-Py-Pu		$1.38 \times 10^5$	1
Indol-3-ethanoic acid (III)	DNA-1	5'-C <sub>7</sub> A <sub>8</sub> T <sub>9</sub>	5'-Py-Pu-Py	1-Base altered	$5.81 \times 10^4$	1
	DNA-3	5'-A <sub>4</sub> A <sub>5</sub> T <sub>6</sub>	5'-Pu-Pu-Py		$1.04 \times 10^5$	1
	DNA-2	5'-G <sub>6</sub> G <sub>7</sub> A <sub>8</sub>	5'-Pu-Pu-Pu	Same	$1.11 \times 10^5$	1
	DNA-4	5'-A <sub>6</sub> A <sub>7</sub> G <sub>8</sub>	5'-Pu-Pu-Pu		$6.39 \times 10^4$	1
Indole-2-carboxylic acid (IV)	DNA-2	5'-T <sub>3</sub> C <sub>4</sub> C <sub>5</sub>	5'-Py-Py-Py	Same	$1.19 \times 10^5$	2
	DNA-3	5'-C <sub>3</sub> T <sub>4</sub> T <sub>5</sub>	5'-Py-Py-Py		$1.48 \times 10^5$	1
	DNA-4	5'-A <sub>6</sub> A <sub>7</sub> G <sub>8</sub>	5'-Pu-Pu-Pu	1-Base altered	$9.44 \times 10^4$	0
	DNA-1	5'-T <sub>3</sub> G <sub>4</sub> G <sub>5</sub>	5'-Py-Pu-Pu		$1.00 \times 10^5$	2
Indol-3-ethanamide (V)	DNA-1	5'-A <sub>2</sub> T <sub>3</sub> G <sub>4</sub>	5'-Pu-Py-Pu	Position Altered	$5.28 \times 10^4$	2
	DNA-2	5'-G <sub>7</sub> A <sub>8</sub> T <sub>9</sub>	5'-Pu-Pu-Py		$7.64 \times 10^4$	1
	DNA-3	5'-A <sub>5</sub> T <sub>6</sub> T <sub>7</sub>	5'-Pu-Py-Py	Reverse Position	$5.28 \times 10^4$	1
	DNA-4	5'-T <sub>4</sub> T <sub>5</sub> A <sub>6</sub>	5'-Py-Py-Pu		$6.47 \times 10^4$	1
Benzimidazol-2-ol (VI)	DNA-1	5'-A <sub>2</sub> T <sub>3</sub> G <sub>4</sub>	5'-Pu-Py-Pu	Same	$5.93 \times 10^4$	2
	DNA-2	5'-A <sub>2</sub> T <sub>3</sub> C <sub>4</sub>	5'-Pu-Py-Pu		$7.49 \times 10^4$	1
	DNA-3	5'-T <sub>6</sub> T <sub>7</sub> G <sub>8</sub>	5'-Py-Py-Pu	Same	$5.78 \times 10^4$	2
	DNA-4	5'-T <sub>6</sub> T <sub>7</sub> G <sub>8</sub>	5'-Py-Py-Pu		$1.68 \times 10^5$	2
Benzimidazole (VII)	DNA-1	5'-A <sub>2</sub> T <sub>3</sub> G <sub>4</sub>	5'-Pu-Py-Pu	Position Altered	$6.27 \times 10^4$	1
	DNA-2	5'-G <sub>7</sub> A <sub>8</sub> T <sub>9</sub>	5'-Pu-Pu-Py		$6.33 \times 10^4$	0
	DNA-3	5'-T <sub>6</sub> T <sub>7</sub> G <sub>8</sub>	5'-Py-Py-Pu	2-Bases altered	$5.30 \times 10^4$	1
	DNA-4	5'-A <sub>6</sub> A <sub>7</sub> G <sub>8</sub>	5'-Pu-Pu-Pu		$7.67 \times 10^4$	0
2,3-Dihydro-1H-indole (VIII)	DNA-1	5'-T <sub>3</sub> G <sub>4</sub> G <sub>5</sub>	5'-Py-Pu-Pu	Reverse Position	$5.27 \times 10^4$	0
	DNA-3	5'-A <sub>5</sub> T <sub>6</sub> T <sub>7</sub>	5'-Pu-Py-Py		$8.23 \times 10^4$	0
	DNA-2	5'-T <sub>3</sub> C <sub>4</sub> C <sub>5</sub>	5'-Py-Py-Py	1-Base altered	$1.05 \times 10^5$	0
	DNA-4	5'-T <sub>4</sub> T <sub>5</sub> A <sub>6</sub>	5'-Py-Py-Pu		$1.02 \times 10^5$	1

were used to calculate the  $\Delta G$  with the help of the (1), appropriately called as  $\Delta G_{\text{calc}}$ . The PreDDICTA module gave theoretical  $\Delta G_{\text{PreDD}}$  and other energy term values (table 2). These  $\Delta G_{\text{PreDD}}$  values were in turn used to calculate K values for docked structures, which were defined as  $K_{\text{PreDD}}$ . Therefore, 3 sets of values were furnished, viz.  $K_{\text{exp}}$ ,  $\Delta G_{\text{calc}}$ ,  $\Delta G_{\text{PreDD}}$  and  $K_{\text{PreDD}}$  (table 3).

Only the representative examples of double reciprocal plot of IV with 4 DNA decamer sequences (IV: DNA complexes) have been shown in figure 2. These plots supplied the  $K_{\text{exp}}$  values for a given drug-DNA complex which were used to obtain

$\Delta G_{\text{calc}}$  values. Such an exercise furnished the comprehensive data sets of  $\Delta G$  and binding constant  $K$  values of drugs (I–VIII) individually binding to 4 DNA duplexes (DNA-1 to DNA-4) summarized in tables 2 and 3. It was later used to assess the statistical agreement of the theoretical and experimental values. Correlation plots were drawn between  $\Delta G_{\text{PreDD}}$  versus  $\Delta G_{\text{calc}}$  values, and between  $K_{\text{exp}}$  values versus  $K_{\text{PreDD}}$  values. Consequently,  $R^2$  values were also obtained between  $K_{\text{exp}}$  values and  $\Delta G_{\text{PreDD}}$  values. These values were in the range of 0.82–0.83. In addition to the correlation analysis of theoretical and experimental values, a detailed structural

**Table 2.** Energy values of docked structures of drug-DNA complexes of indole derivatives obtained from PreD-DICTA.

Drug	Complex with	$\Delta G_{\text{preDD}}$ (kcal/mol)	Total electrostatics (kcal/mol)	Total van der Waals (kcal/mol)	Rotational translational entropy (kcal/mol)	Hydration free energy (kcal/mol)	Total energy (kcal/mol)
Indole-3-butanoic acid (I)	DNA-1	-7.0	-0.8	-7.1	23.1	-8.5	6.6
	DNA-2	-7.3	-1.0	-8.3	23.1	-9.1	4.7
	DNA-3	-7.2	-1.3	-7.3	23.1	-9.3	5.1
	DNA-4	-7.2	-1.5	-7.3	23.1	-9.3	4.9
Indole-3-propanoic acid (II)	DNA-1	-6.7	-0.2	-5.9	23.0	-7.6	9.2
	DNA-2	-6.7	-0.6	-5.1	23.0	-7.9	9.4
	DNA-3	-7.1	-0.3	-7.4	23.0	-8.8	6.4
	DNA-4	-6.9	-0.5	-6.2	23.0	-8.3	7.9
Indol-3-ethanoic acid (III)	DNA-1	-6.5	-0.7	-3.6	22.7	-7.2	11.2
	DNA-2	-6.9	-0.8	-7.2	22.7	-7.1	7.5
	DNA-3	-6.9	-0.7	-7.0	22.7	-7.5	7.4
	DNA-4	-6.6	-0.9	-4.6	22.7	-7.3	9.8
Indole-2-carboxylic acid (IV)	DNA-1	-6.9	-1.2	-7.7	22.6	-6.3	7.4
	DNA-2	-7.0	-1.2	-7.7	22.6	-6.6	7.1
	DNA-3	-7.2	-2.8	-7.4	22.6	-7.0	5.4
	DNA-4	-7.1	-1.4	-8.4	22.6	-6.8	5.9
Indol-3-ethanamide (V)	DNA-1	-6.5	-0.4	-4.6	22.7	-7.0	10.8
	DNA-2	-6.7	-0.7	-5.1	22.8	-7.7	9.3
	DNA-3	-6.5	-0.8	-3.8	22.8	-7.5	10.7
	DNA-4	-6.6	-0.6	-5.6	22.8	-6.8	9.7
Benzimidazol-2-ol (VI)	DNA-1	-6.6	-1.5	-5.5	22.0	-5.4	9.7
	DNA-2	-6.7	-1.3	-5.6	22.0	-5.6	9.5
	DNA-3	-6.5	-1.8	-5.2	22.0	-4.5	10.6
	DNA-4	-7.2	-3.5	-8.0	22.0	-5.0	5.5
Benzimidazole (VII)	DNA-1	-6.6	-0.8	-4.6	21.6	-6.0	10.3
	DNA-2	-6.6	-0.5	-5.1	21.6	-5.7	10.4
	DNA-3	-6.5	-1.1	-4.2	21.6	-5.9	10.4
	DNA-4	-6.7	-0.7	-6.2	21.6	-5.7	9.0
2,3-Dihydro-1H-indole (VIII)	DNA-1	-6.8	-0.4	-4.6	21.9	-8.7	8.2
	DNA-2	-6.9	-0.5	-5.1	21.9	-8.7	7.6
	DNA-3	-6.8	-0.3	-4.3	21.9	-9.1	8.2
	DNA-4	-6.9	-0.3	-5.8	21.9	-8.1	7.6

analysis of docked pdb structures was also furnished (figure 3).

As reported,<sup>25</sup> the minor groove widened at the binding site in a drug-conjugated DNA structure to accommodate the drug molecule in a sequence specific binding in the minor groove of DNA oligomer. Distances between phosphorus atoms of phosphates of opposite residues across the strands obtained from I-DNA complex as a representative

example have been summarized in table 4. These distances have reflected the DNA base pair width. The change in the base-pair width along all the residues of DNA oligomer as a result of binding of indole derivatives (I–VIII) with four DNA sequences were analysed using the 32 drug–DNA complex structures. The overall opening of base pair was observed across the whole length of DNA oligomers in all the 32 cases of drug–DNA

**Table 3.**  $K_{\text{exp}}$  and  $\Delta G_{\text{PreDD}}$  values of indole derivatives and DNA complexes.

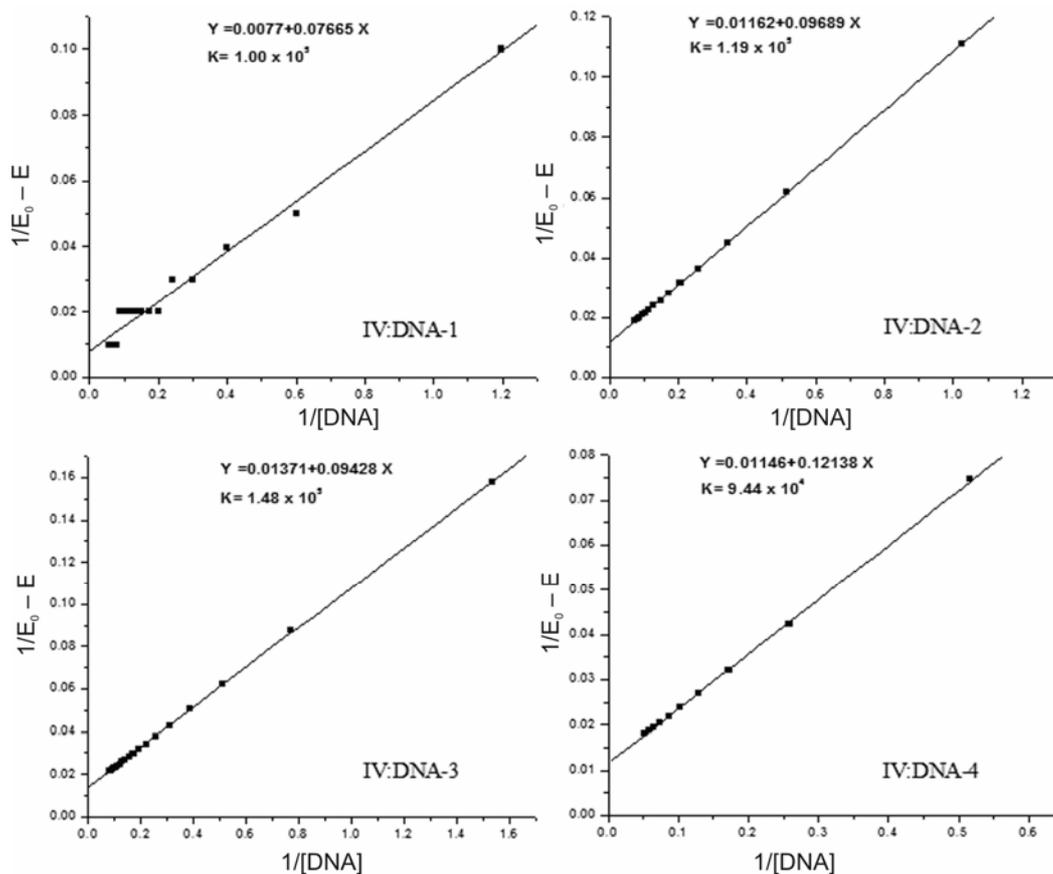
Drug	Complex with	$\Delta G_{\text{calc}}$ (kcal/mol)	$K_{\text{exp}}$ (mol <sup>-1</sup> )	$\Delta G_{\text{PreDD}}$ (Kcal/mol)	$K_{\text{PreDD}}$ (mol <sup>-1</sup> )
Indol-3-butanoic acid (I)	DNA-1	-6.9	$1.23 \times 10^5$	-7.0	$1.26 \times 10^5$
	DNA-2	-6.9	$1.96 \times 10^5$	-7.3	$2.10 \times 10^5$
	DNA-3	-6.8	$1.33 \times 10^5$	-7.2	$1.76 \times 10^5$
	DNA-4	-6.9	$1.11 \times 10^5$	-7.2	$1.76 \times 10^5$
Indol-3-propionic acid (II)	DNA-1	-6.7	$7.33 \times 10^4$	-6.7	$7.61 \times 10^4$
	DNA-2	-6.7	$7.30 \times 10^4$	-6.7	$7.61 \times 10^4$
	DNA-3	-7.1	$1.41 \times 10^5$	-7.1	$1.49 \times 10^5$
	DNA-4	-7.1	$1.38 \times 10^5$	-6.9	$1.06 \times 10^5$
Indol-3-ethanoic acid (III)	DNA-1	-6.5	$5.81 \times 10^4$	-6.5	$5.44 \times 10^4$
	DNA-2	-6.9	$1.11 \times 10^5$	-6.9	$1.06 \times 10^5$
	DNA-3	-6.9	$1.04 \times 10^5$	-6.9	$1.06 \times 10^5$
	DNA-4	-6.6	$6.39 \times 10^4$	-6.6	$6.43 \times 10^4$
Indol-2-carboxylic acid (IV)	DNA-1	-6.9	$1.00 \times 10^5$	-6.9	$1.06 \times 10^5$
	DNA-2	-7.0	$1.19 \times 10^5$	-7.0	$1.26 \times 10^5$
	DNA-3	-7.1	$1.48 \times 10^5$	-7.2	$1.76 \times 10^5$
	DNA-4	-6.8	$9.44 \times 10^4$	-7.1	$1.49 \times 10^5$
Indol-3-ethanamide (V)	DNA-1	-6.5	$5.28 \times 10^4$	-6.5	$5.44 \times 10^4$
	DNA-2	-6.7	$7.64 \times 10^4$	-6.7	$7.61 \times 10^4$
	DNA-3	-6.5	$5.28 \times 10^4$	-6.5	$5.44 \times 10^4$
	DNA-4	-6.6	$6.47 \times 10^4$	-6.6	$6.43 \times 10^4$
Benzimidazol-2-ol (VI)	DNA-1	-6.6	$5.93 \times 10^4$	-6.6	$6.43 \times 10^4$
	DNA-2	-6.7	$7.49 \times 10^4$	-6.7	$7.61 \times 10^4$
	DNA-3	-6.5	$5.78 \times 10^4$	-6.5	$5.44 \times 10^4$
	DNA-4	-7.2	$1.68 \times 10^5$	-7.2	$1.76 \times 10^5$
Benzimidazole (VII)	DNA-1	-6.6	$6.27 \times 10^4$	-6.6	$6.43 \times 10^4$
	DNA-2	-6.6	$6.33 \times 10^4$	-6.6	$6.43 \times 10^4$
	DNA-3	-6.5	$5.30 \times 10^4$	-6.5	$5.44 \times 10^4$
	DNA-4	-6.7	$7.67 \times 10^4$	-6.7	$7.61 \times 10^4$
2,3-Dihydro-1H-indole (VIII)	DNA-1	-6.5	$5.27 \times 10^4$	-6.8	$8.99 \times 10^4$
	DNA-2	-6.9	$1.05 \times 10^5$	-6.9	$1.06 \times 10^5$
	DNA-3	-6.7	$8.23 \times 10^4$	-6.8	$8.99 \times 10^4$
	DNA-4	-6.9	$1.02 \times 10^5$	-6.9	$1.06 \times 10^5$

complexes. This observation plausibly pointed towards non-specific binding of indole derivatives in our study.

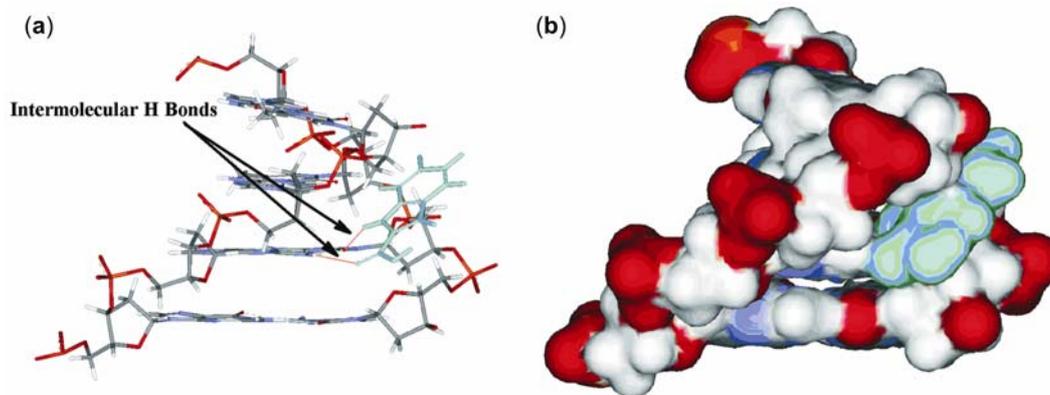
#### 4. Discussion

The DNA binding constant values,  $K_{\text{exp}}$  obtained from the double reciprocal method (table 1) were in the range of  $10^4$ – $10^5$  per mole for all the indole derivatives indicated a similar binding affinity for these compounds which is relevant to their small size. The possibility of intercalation of these indole

derivatives between the DNA base pairs was little, mainly due to less-effective stacking forces between the indole nucleus and the DNA bases. The side chain associated with indole derivatives (I–VIII) provided them with additional H-bond donor/acceptor atoms and thereby enabling the molecule to stay in the minor groove with the help of van der Waal's forces and H-bonds. The H-bond distances were manually measured keeping in mind the H-bond distance between a pair of atoms to be 1.5–3.0 Å.<sup>26</sup> It can be possibly concluded from table 2 that these compounds bind to DNA largely because



**Figure 2.** Double reciprocal plot of indol-2-carboxylic acid (IV) with 4 DNA decamer sequences.



**Figure 3.** Docked structure of indol-2-carboxylic acid (IV) with DNA decamer. (a) Close-up view of Drug–DNA binding with 2 intermolecular H-bonds shown in red. (b) van der Waal's surface view of the same complex.

of van der Waal's interactions as compared to electrostatic forces. This was the reason why most of the complexes showed 1–2 H-bonds while others did not have H-bonds.

The DNA sequences (DNA-1 to DNA-4) chosen in this study were designed to evaluate base sequence specificity among chosen compounds. How-

ever, the study revealed that there was no such specificity in terms of base sequences of DNA duplexes. Nevertheless, all the 8 indole compounds exhibited some interesting purine-pyrimidine (Pu–Py) sequence preferences as shown in table 1. It was noted that compounds I, II, III, IV and VI possessed very interesting Pu–Py specificity patterns.

**Table 4.** Distances between phosphates of opposite residues across strands obtained from indole-3-butanoic acid (I): DNA complexes.

Distance between residues	Standard B-DNA	DNA-1	DNA-2	DNA-3	DNA-4
R2-R19	18.23	19.07	19.22	19.43	18.98
R3-B18	18.23	19.62	19.95	19.21	19.99
R4-B17	18.23	19.40	19.67	19.39	19.37
R5-B16	18.23	19.64	19.77	19.76	19.71
R6-B15	18.23	<u>19.03</u>	<u>19.18</u>	20.33	19.24
R7-B14	18.23	18.80	<u>19.52</u>	<u>19.39</u>	<u>20.10</u>
R8-B13	18.23	<u>19.58</u>	<u>19.97</u>	<u>19.47</u>	<u>18.96</u>
R9-B12	18.23	19.82	19.51	<u>19.40</u>	<u>19.74</u>
Binding site	–	5'-C <sub>6</sub> C <sub>7</sub> A <sub>8</sub>	5'-C <sub>6</sub> C <sub>7</sub> A <sub>8</sub>	5'-T <sub>7</sub> G <sub>8</sub> C <sub>9</sub>	5'-A <sub>7</sub> G <sub>8</sub> C <sub>9</sub>

Among them, drugs **I** and **VI** exhibited same Pu–Py specific patterns in the pairs of drug–DNA complexes. Compounds **II**, **III** and **IV**, possessed at least one pair of complexes having same Pu–Py specific pattern in DNA binding. Whereas in ligands **V**, **VII** and **VIII**, there was also some sort of Pu–Py specificity. In compounds **V**, **VII** and **VIII**, Pu–Py specificity has revealed similar patterns with 1-base altered, 2-bases altered, reverse position or position altered in pairs of complexes, as summarized in table 1. These Pu–Py specific patterns could be very interesting in the light of Pu–Py specificity of DNA cleavage.<sup>23</sup> This Pu–Py specificity in the view of non-specific DNA binding nature of drug (**I**–**VIII**) presents a broader tool to look into the DNA sequence specific binding, where the electron density patterns would be similar, individually in purine and pyrimidine bases on the binding site. Therefore, it encompasses such broader sequence preference.

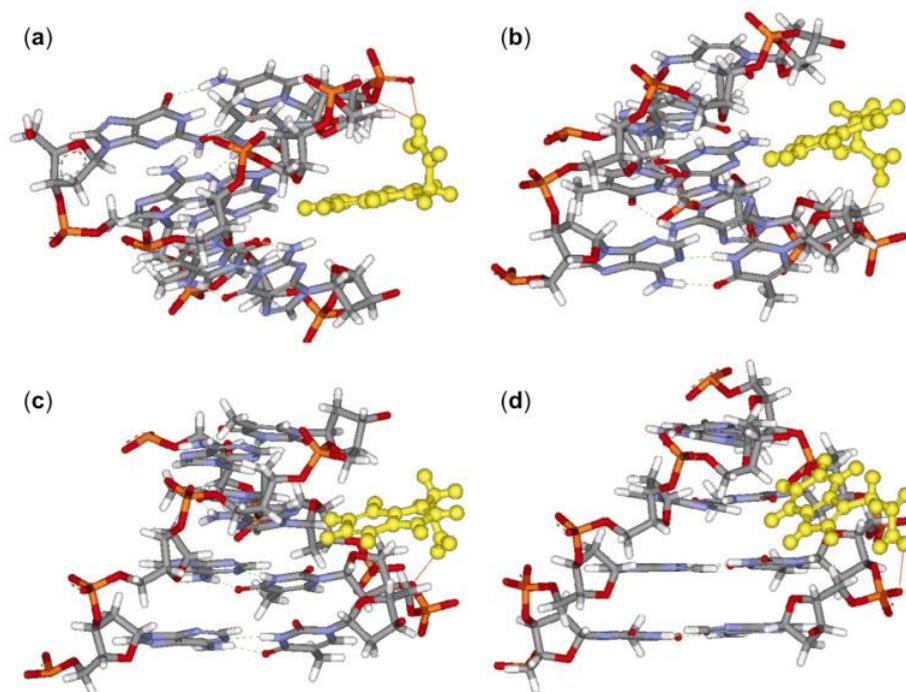
Docking experiments were performed in order to find out the chosen binding site along with the preferred orientation of the ligand inside the DNA minor groove. Docking experiments resulted in the energy minimized docked structures suggesting the best possible geometry of the ligand inside the DNA minor groove. Although in the case of minor groove binding ligands, the possibility of DNA structural perturbations was very small. However, all the docked structures in this study have shown slight perturbation in the DNA structure as a result of drug binding.

The structural analysis of docked structures divulges significant details about the binding pattern of these indole derivatives. It has been observed that binding of all the drug molecules was stabilized by one or more H-bonds either with the DNA bases or with the backbone PO<sub>2</sub> group in addition to van der Waal's interactions. Compound **I** demonstrated binding of  $1.96 \times 10^5$  per mole with DNA-2 which

was highest among all the compounds. Similarly, **IV** gave binding constant of  $1.48 \times 10^5$  per mole and **VI** showed binding constant of  $1.68 \times 10^5$  per mole. Additionally, these compounds have shown the formation of intermolecular H-bonds. Thermodynamic profile of drug–DNA complexes obtained from PreDDICTA analysis shows that the complexation is accompanied by large positive values of entropy which is normally observed in case of other minor groove binding ligands.<sup>27</sup> Similar trend was also observed in case of drug-tRNA binding.<sup>28,29</sup> It can be concluded that intermolecular interactions in all the 32 drug–DNA complexes were also supported by intermolecular H-bonds, besides other factors like van der Waal's forces, etc. as revealed by the higher values of energy terms for van der Waal's interactions and electrostatic (table 2).

The statistical similarity of these results was established by correlation plots. Correlation between binding constant  $K_{\text{exp}}$  and  $K_{\text{PreDD}}$  values shown good agreement with each other as evident from the  $R^2$  value of 0.83. Similar  $R^2$  value was obtained for  $G_{\text{PreDD}}$  and  $\Delta G_{\text{calc}}$  values (table 4). It was deduced from the correlation constants and values of  $K_{\text{exp}}$ ,  $K_{\text{PreDD}}$ ,  $G_{\text{PreDD}}$  and  $\Delta G_{\text{calc}}$  (table 3) that the theoretical method for assessing minor groove binding of indole derivatives predicted  $K_{\text{PreDD}}$  and  $\Delta G_{\text{PreDD}}$  quite accurately, which was further confirmed by values obtained from the experimental methods. It could be emphasized at this point that the theoretical method was originally designed for the prediction of crystal structures. However, the present study has revealed that these molecular modelling approaches were adequately capable in assessing the results from solution studies like fluorescence, UV-absorbance and other similar techniques.

The effect of methylene linker length on the overall binding affinity of these ligands was also considered



**Figure 4.** Docked structure of indol-2-acetamide (**III**) with four DNA decamer sequences showing anchoring mechanism. (a) Docked structure with 5'-d(GGCAATTGCC)<sub>2</sub> (DNA-1), (b) with 5'-d(GATCCGGATC)<sub>2</sub> (DNA-2), (c) with 5'-d(GGCAATTGCC)<sub>2</sub> (DNA-3) and (d) with 5'-d(GGCTTAAGCC)<sub>2</sub> (DNA-4).

in selecting the compounds for this DNA binding study. Longer methylene chain length in case of **I** facilitates proper alignment of functional groups whereas in compounds **II** and **III**, fewer methylene groups in the linker chain would not as favourably allow the proper alignment of drug in the DNA minor groove. This could be a plausible explanation for an order of magnitude stronger DNA-binding affinity of drug **I**, which was further augmented by intermolecular H-bonds with NH group of indole ring. It was also observed in the case of 2-(1H-indol-3-yl)ethanamide (**V**), that the methylene linker provides an anchor to the molecule to remain within the minor groove of DNA (figure 3). The ethanamide derivative **V** of indole forms two intermolecular H-bonds between NH<sub>2</sub> group of drug and PO<sub>2</sub> group of DNA. The presence of rather strong intermolecular H-bonds clearly indicates the anchoring of the drug molecule with its side chain, while the main indole ring remains within the minor groove of DNA and reorients itself according to the shape of the minor groove. Similar phenomenon may also be possible in case of other indole derivatives possessing methylene linker side chain.

However, in **IV** the alignment of carboxylic acid terminus with respect to indole NH group facilitated as strong DNA binding as in ligand **I**. From the analysis of the drug **IV**-DNA complex structure, it was revealed that DNA interaction of **IV** was supported by at least 2 intermolecular H-bonds (figure 4). Nevertheless, similar pattern of the formation of intermolecular H-bond was also observed in the ligand **VI**-DNA complex. However, the extent of DNA binding of **VI** was an order of magnitude lower than in ligand **I** (table 1), as indicated by its DNA binding constant *K*. It was also concluded by evaluating the energy terms for drug-DNA complexes (table 2), that electrostatics played an important role in the DNA binding of ligands **I**, **IV** and **VI** as compared to DNA binding of other compounds.

It could be precisely suggested from the observations that electrostatics was not the only primary factor in the DNA binding of these ligands. Consequently, the alignment of carboxylic acid terminus with respect to NH group of indole ring was another important factor in the DNA binding of indole derivatives (**I**-**VIII**). Longer the length of methylene linker, greater was the facilitation in DNA binding

of ligands. It was also reflected by an order of magnitude larger binding constant for ligand I. Here, no sequence specificity for individual DNA bases was revealed, meaning ligand I was binding to DNA-1 to DNA-4 with the same strength. Moreover, similar pattern was also seen in ligands II–VIII. This further confirmed that DNA binding of these indole derivatives was non-specific.

## 5. Conclusions

This work illustrated the non-specific minor groove DNA binding of indole derivatives with a precise order of Pu–Py specificity. Uniform opening of the minor groove also underscored the non-specific DNA binding of these indole molecules. This binding was supported by intermolecular H-bonding and van der Waal's forces. Experimental values of binding constants and free energy of drug–DNA binding matched adequately with values obtained by PreDDCTA module. Such study opens an avenue for testing other systems supplemented by both experimental and theoretical approaches.

## Acknowledgements

S K acknowledges the consistent support of Prof. P S Satsangi, Chairman, Advisory Committee on Education, Dayalbagh, Agra, India. This work was supported by a Major Research Project grant to SK from University Grants Commission of India (No. F.30-52/2004 (SR)). Project Fellowship to PP is also gratefully acknowledged.

## References

- Singh, M P, Joseph T, Kumar S, Bathini Y and Lown J W 1992 *Chem. Res. Toxicol.* **5** 597
- Sartorius J and Schneider H J 1995 *FEBS Lett.* **374** 387
- Brun F, Toulme J J and Helene C 1975 *Biochem.* **14** 558
- Dimicoli J L and Helene C 1974 *Biochem.* **13** 714
- Helene C, Dimicoli J L and Brun F 1971a *Biochemistry* **10** 3802
- Helene C and Maurizot J C 1981 *CRC Crit. Rev. Biochem.* **10** 213
- Coll M, Aymami J, van der Marel G A, van Boom J H, Rich A and Wang A H J 1989 *Biochemistry* **28** 310
- Dwyer T J, Geierstanger B H, Mrksich M, Dervan P D and Wemmer D E 1993 *J. Am. Chem. Soc.* **115** 9900
- Klevit R E, Wemmer D and Reid B R 1986 *Biochemistry* **25** 3296
- Lown J W 1994 *J. Mol. Rec.* **7** 79
- Rajeswari M R, Montenay-Garestier T and Helene C 1987 *Biochemistry* **26** 6825–6831 and references cited therein
- Darnell J E 2002 *Nature Rev.* **2** 740
- Pandolfi P P 2001 *Oncogene* **20** 3116
- Baraldi P G *et al* 2004 *Med. Res. Rev.* **24** 475
- Tse W C and Boger D L 2004 *Chem. Biol.* **11** 1607
- Haq I 2002 *Arch. Biochem. Biophys.* **403** 1
- Chyan Y-J, Poeggeler B, Rawhi A O, Chain D G, Frangione B, Ghiso J and Pappolla M A 1999 *J. Biol. Chem.* **274** 21937
- Bach N J, Dillard R D, Draheim S E, Hermann R B and Schevitz R W 2001 Patent No. 6252084, USA
- Jayaram B 2007 *DNA Ligand Docking*. Retrieved 2008, from supercomputing Facility of bioinformatics and computational biology; <http://www.scfbio-iitd.res.in/dock/dnadock.jsp>
- Gupta A, Gandhimathi A and Sharma P and Jayaram B 2007 **14** 632
- Pearlman D A, Case D A, Caldwell J W, Ross W S, Cheatham III J E, Debolt S, Ferguson D, Seibel G and Kollman P 1995 *Comp. Phys. Comm.* **91** 1
- Shaikh S A and Jayaram B 2007 *J. Med. Chem.* **50** 2240
- Veal J M and Rill R 1989 *Biochemistry* **28** 3243
- Huang Q Q, Harvey C M, Paterson A R, Cass C E and Young J D 1993 *J. Biol. Chem.* **268** 20613
- Kumar S, Reed M, Gamper H, Gorn V, Kutayavin I, Lukhtanov E, Meyer R and Schweitzer B I 1998 *Nucleic Acids Res.* **26** 831
- Jeffrey G A and Saenger W 1991 *Hydrogen bonding in biological structures*(Berlin: Springer-Verlag)
- Chairs J B 2006 *Arch. Biochem. Biophys.* **453** 26
- Islam M M, Pandya P, Choudhary S R, Kumar S and Kumar G S 2008 *J. Mol. Struct.* **891** 489
- Islam M M, Pandya P, Kumar S and Kumar G S 2009 *J. Mol. Biosystems.* **5** 244