RSC Advances

ROYAL SOCIETY OF CHEMISTRY

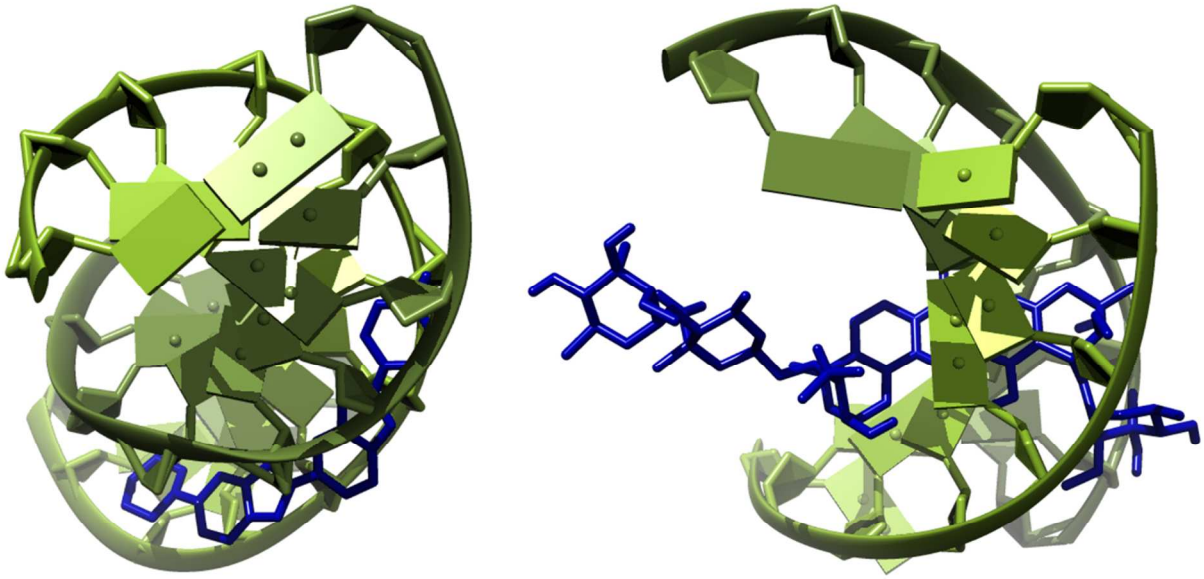
www.rsc.org/advances

In search of a docking protocol to distinguish between DNA intercalators and groove binders: Genetic algorithm Vs shape-complementarity based docking methods

Mariya al-Rashida[a]*, Sana Ahsen[a]

Department of Chemistry, Forman Christian College (A Chartered University), Ferozepur Road, Lahore 54600, Pakistan

Corresponding Author Email: maria_al_rashida@hotmail.com; mariyarashidarashida@fccollege.edu.pk; Phone: +923324213592



Abstract

While molecular docking protocols have been well parameterized for protein-ligand and even protein-protein interactions, there is a significant lack of similar procedures for DNA docking. The accuracy of DNA docking method/protocol is directly linked with the "selection" of a DNA receptor, that is, selecting a DNA receptor with or without an intercalation gap. A molecule, known experimentally to be a DNA groove binder will give many misleading docked conformations when attempted to dock into a DNA receptor containing an intercalation gap, and vice versa for an experimentally known DNA intercalating molecule that will similarly give misleading docked conformations when docked into a DNA receptor that does not contain an intercalation gap. Meaningful DNA docking studies, therefore, require a prior knowledge of "mode-of-binding" (intercalation or groove binding) of the ligand under investigation. This dilemma greatly limits the usefulness of DNA docking studies to search the known chemical libraries for new DNA groove binding or intercalating molecules. This study has been undertaken to investigate whether or not a docking protocol can be developed that will be able to reasonably sort out DNA intercalators from DNA groove binders without any prior knowledge of the mode-of-binding of the ligand.

Introduction

DNA is inarguably an indispensable molecule for continuation and propagation of life in the form of cell replication, among many other functions. Notably, after the development of cis-platin, a platinum containing DNA binding anti-cancer drug, there has been a constant struggle toward the discovery of other, more potent DNA binding molecules as anticancer drugs, and while many examples of such molecules have appeared, the search for even better potential anticancer drugs still continues [1-2]. DNA groove binding (in the major or minor groove of the DNA), and DNA intercalation are two of the most commonly observed modes of interaction of small molecules/drugs with the DNA. In groove binding mode, the molecule is typically flexible, contains rotatable bonds, and is able to orient itself along the major or minor groove of the DNA, thereby inhibiting its regular function [3,4]. The DNA intercalators are typically inflexible planar molecules that stack in-between the DNA base pairs causing an intercalation gap to appear in the

DNA helical structure [5-7]. Figures 1-2 show the structures of some of the well-known DNA groove binders and intercalators respectively.

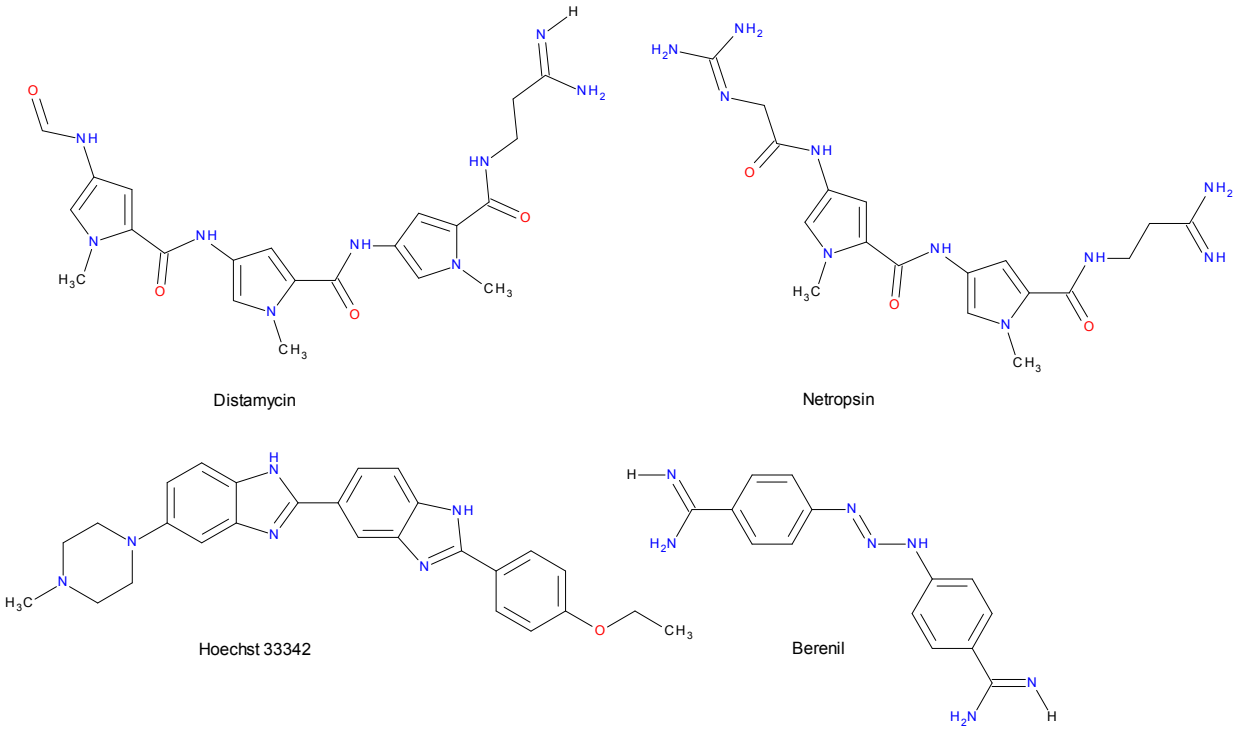


Figure 1. Chemical structures of DNA groove binders.

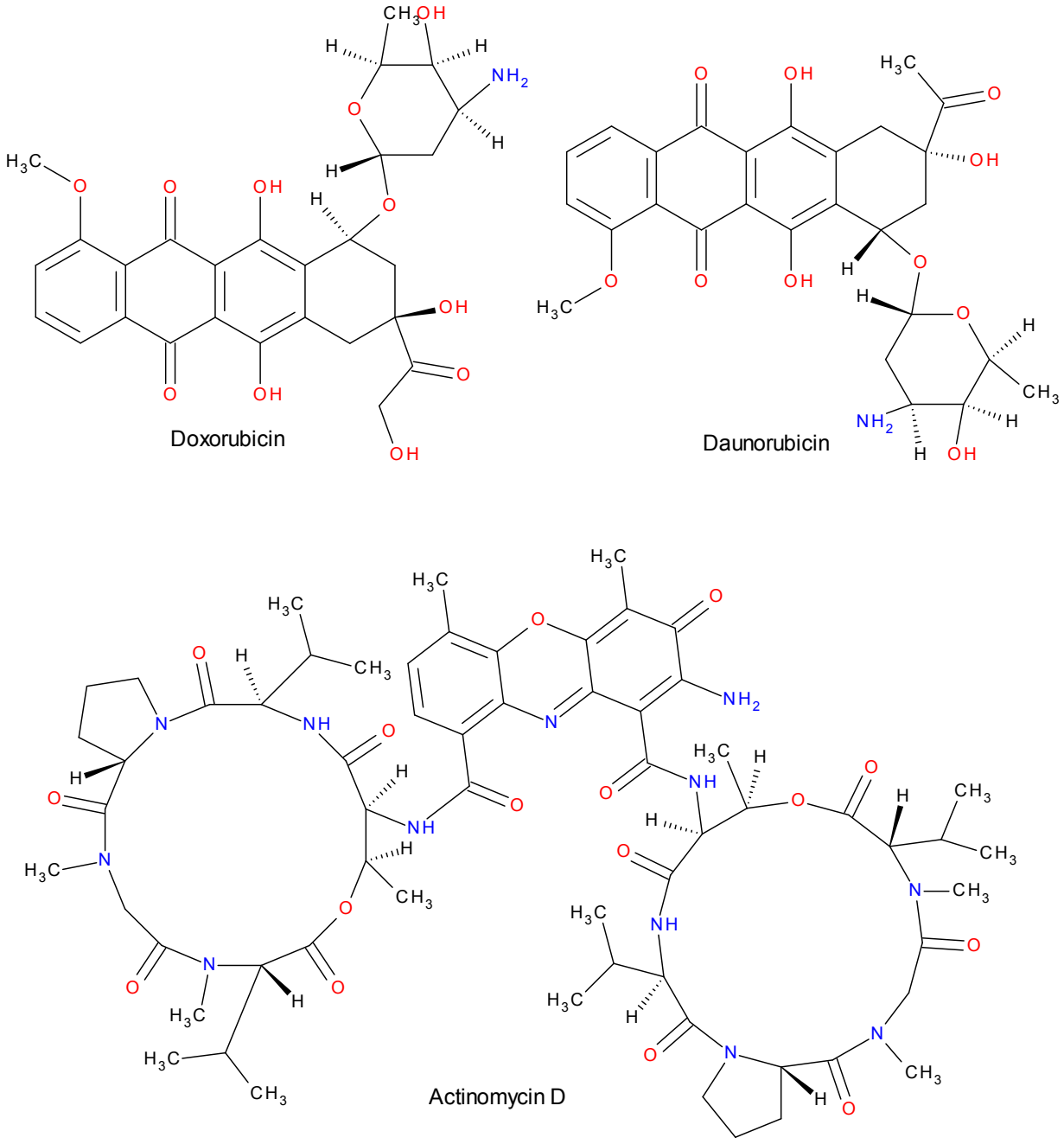


Figure 2. Chemical structures of DNA intercalators.

There is a staggering evidence of successful applications of molecular docking studies in rational drug design [8-14], and while most molecular docking methods have been parameterized for protein-ligand docking, there is a significant lack of similar procedures/protocols for DNA docking studies. Although some studies in this regard have appeared [15-17]. Meaningful DNA docking studies require a prior knowledge of mode of binding (groove binder or intercalator) of ligand with the DNA, since in most docking software, the receptor (here, DNA) is kept rigid and ligand is treated as flexible. If a ligand, that is known to be an intercalator, is docked into a DNA receptor without any preformed intercalation gaps in it, the ligand will give misleading binding modes and vice versa for a known DNA groove binder when docked into a DNA receptor containing intercalation gaps. This dilemma greatly limits the usefulness of DNA docking studies to search the known chemical libraries for new DNA groove binding or intercalating molecules.

We wanted to look for subtle hints in interaction energy/docking score function or other related parameters that could indicate whether the ligand under investigation is a DNA intercalator or a groove binder without any prior knowledge of its mode of binding (groove binder or intercalator) with DNA. For this purpose two main approaches were used, a genetic algorithm based docking approach and shape-complementarity based docking approach. Genetic algorithm based docking approach was applied using AutoDock 4.2 [18] and shape complementarity based docking approach was applied using Hex 8.0 [19].

The first approach simulates the definite docking process in which the ligand-receptors interaction energies are calculated. In this approach, the receptor and the ligand are separated by some physical distance, and the ligand searches its position into the receptor's active site after a number of "moves" in its conformational space. The moves include rigid body transformations such as rotations, translations and internal changes to the ligand's structure. It includes torsion angle rotations. Consequently, total energy of the system is calculated after every move. Simulation is computationally costly. It requires exploring a large energy landscape. Optimization methods, grid-based techniques and improved computer speed have made docking simulation more practical [20, 21].

Second approach uses a matching technique that describes the receptor and the ligand as complementary surfaces. Geometric matching or shape complementarity methods describe the receptor and ligand as a set of features that make them dockable. These features may include molecular surface as complementary surface descriptors. In this case, the receptor's molecular surface is narrated in requisites of its solvent-accessible surface area and the ligand's molecular surface is narrated in requisites of its matching surface description. The complementarity between the two surfaces amounts to the shape matching description that helps in finding the complementary pose of docking the target and the ligand molecules [22]."

Results and Discussion

Crystal structures of DNA intercalators and groove binders were downloaded from the Protein Data Bank, details of pdb files and ligands therein are given in Table 1. Method validation was carried out for DNA intercalators and groove binders. A number of different AutoDock [18] docking parameters were experimented with after which it was found that best docking results could be obtained in reasonable amount of time if the number of GA runs was 20 and maximum number of energy evaluations was 5×10^6. For all self-docking studies the docking methodology was successfully able to reproduce the experimentally observed bound conformation for the ligand. In 1XRW, the ligand is a platinum complex with acridine derivative and ethylene diamine, since Autodock is not been parametrized for use with platinum metal, this part was removed and only the acridine part was kept, the software calculated rmsd of docked ligand (without platinum and ethylene diamine) against the crystallographic ligand (containing platinum and ethylene diamine), that is why there is this much difference between the docked and crystallographic conformation. Cross-docking experiments were designed as depicted in Figures 4-5. In group-I cross-docking, that is, cross-docking within the same binding group (Figure 4), the bound molecule extracted from "groove binder-DNA complex 1" was used as a ligand and docked into the DNA template from "groove binder-DNA complex 2" and vice versa. Thus

ligand from 127D was docked into DNA template from 1QV4 and vice versa. Similarly the bound molecule extracted from "intercalator-DNA complex 1" was used as a ligand and docked into the DNA template from "intercalator-DNA complex 2" and vice versa. Hence ligand from 1XRW was docked into DNA template from 1Z3F and vice versa.

Table 1. Details of DNA intercalators and groove binders downloaded from Protein Data Bank.

Serial No.	**PDB ID**	**Ligand Name**	**Ligand Code**	**Mode of binding**
1	1XRW	1-[2-(acridin-9-ylamino)ethyl]-1,3-dimethylthiourea- platinum(ii)-ethane-1,2-diamine	**2PT**	Intercalator
2	1Z3F	Ellipticine	**EL**	Intercalator
3	127D	2'-(4-hydroxyphenyl)-5-(4-methyl-1-piperazinyl)- 2,5'-bi-benzimidazole (Hoechst)	**HT**	Groove binder
4	1QV4	2'-(3-methyl-4-dimethylaminophenyl)-5-(4- methyl-1-piperazinyl)-2,5'-bi-benzimidazole	**MBB**	Groove binder

2D structures of ligands (groove binders and intercalators) are given in Figure 3.

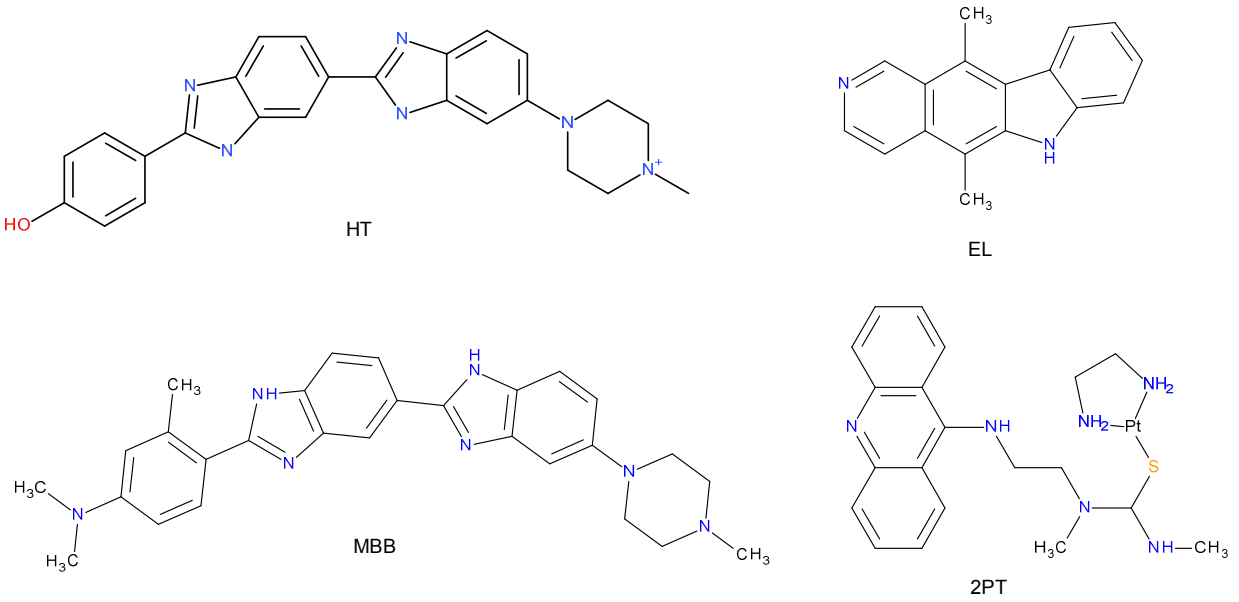


Figure 3. 2D structures of groove binders (HT, MBB) and intercalators (EL, 2PT) used in this study.

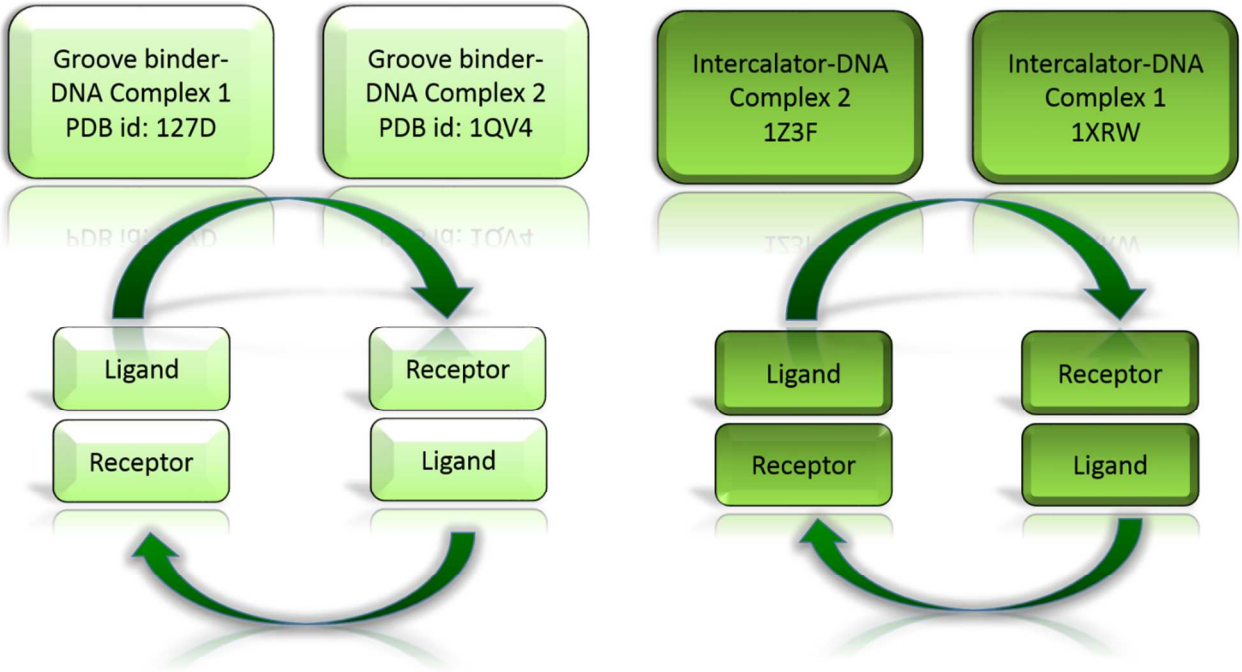


Figure 4. Cross docking scheme within the same binding mode.

In group-II docking, that is cross-docking within different binding group (Figure 5), the bound molecule extracted from "groove binder-DNA complex 1" was used as a ligand and docked into the DNA template derived from "intercalator-DNA complex 1" and "intercalator-DNA complex 2" respectively. Thus ligand (HT) from 127D, was docked into DNA template from 1XRW and 1Z3F. Similarly the bound molecule extracted from "groove binder-DNA complex 2" was used as a ligand and docked into the DNA template from "intercalator-DNA complex 1" and "intercalator-DNA complex 2". Accordingly, the ligand (MBB) from 1QV4, was docked into DNA template from 1XRW and1Z3F. Similar pattern was adopted for DNA intercalator molecules EL and 2PT (extracted from intercalator-DNA complex, 1Z3F and 1XRW, respectively), which were used as ligands and docked into DNA templates derived from groove-binder DNA complexes 1 and 2, that is, 127D and 1QV4 respectively.

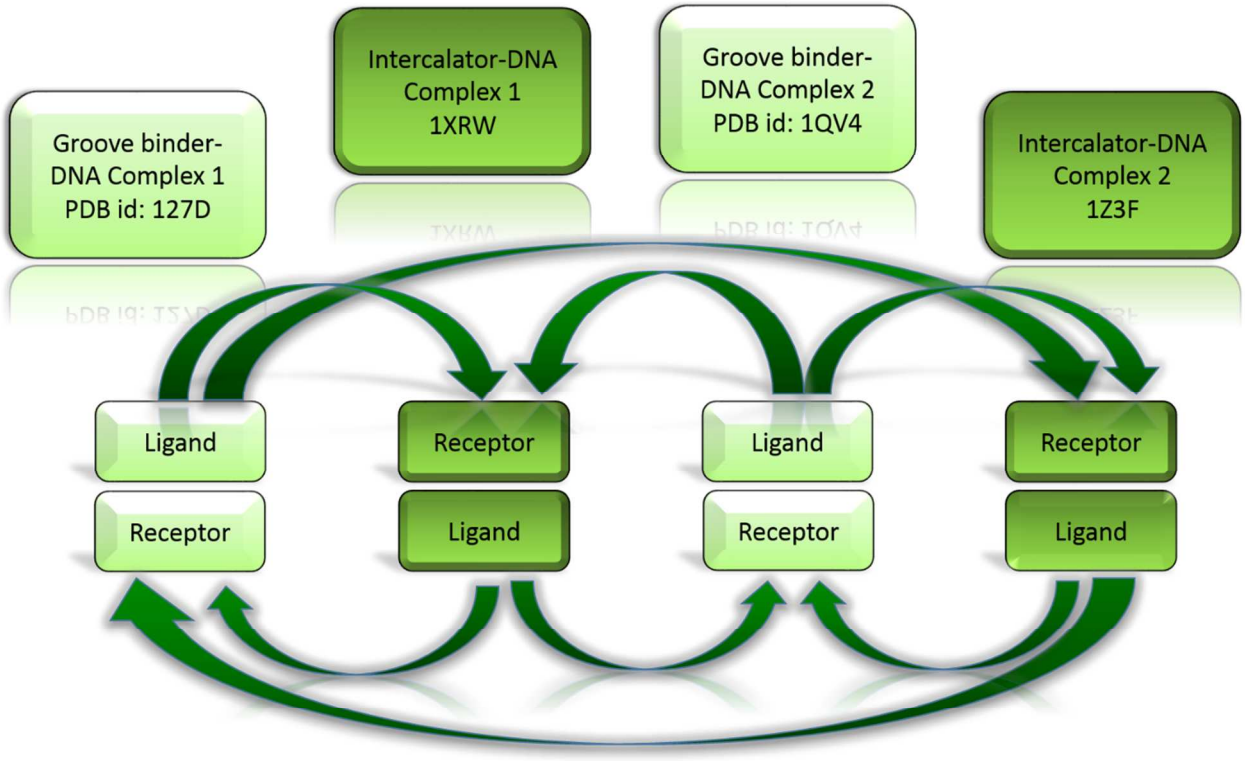


Figure 5. Cross docking scheme within different binding mode.

Out of all docked conformations, the analysis of mode of binding, that is, whether groove-binding and intercalator was based on careful visual examination rather than binding free energy (ΔG, kcal/mol) or docking score. The results of self-docking were as expected and were clearly able to reproduce the original bound conformations of the ligands. The results of self-docking experiments using AutoDock are summarized in Table 2.

Table 2. AutoDock calculated binding free energies and rmsd for self-docking experiments.

Sr. #	DNA Groove Binders				DNA Intercalators			
	PDB ID: 127D		PDB ID: 1QV4		PDB ID: 1XRW		PDB ID: 1Z3F	
	ΔG kcal/mol	RMSD Aº	ΔG kcal/mol	RMSD Aº	ΔG kcal/mol	RMSD Aº	ΔG kcal/mol	RMSD Aº
1	-12.69	0.38	-13.67	3.26	-7.26	5.12	-9.03	1.80
2	-12.69	0.36	-13.64	2.94	-7.22	4.69	-9.03	1.80
3	-12.66	0.4	-13.64	3.38	-7.13	4.52	-9.03	1.80
4	-12.53	0.49	-13.61	2.97	-7.11	5.17	-9.03	1.81
5	-12.52	0.5	-13.59	1.69	-7.05	4.5	-9.03	1.80
6	-12.51	1.06	-13.59	2.94	-7.01	4.87	-9.03	1.80
7	-12.51	1.04	-13.57	1.39	-6.88	4.73	-9.03	1.80
8	-12.5	0.52	-13.55	1.42	-6.83	5.08	-9.03	1.81
9	-12.49	1.06	-13.52	1.77	-6.81	5.35	-9.03	1.80
10	-12.48	1.06	-13.52	1.44	-6.79	5.28	-9.03	1.80
11	-12.44	0.44	-13.51	1.47	-6.78	5.39	-9.03	1.80
12	-12.42	1.03	-13.51	1.4	-6.6	2.66	-9.03	1.80
13	-11.97	1.63	-13.5	1.48	-6.6	4.37	-9.03	1.80
14	-11.95	1.59	-13.49	2.95	-6.59	4.64	-9.03	1.80
15	-11.9	1.8	-13.48	1.49	-6.56	1.93	-9.03	1.80
16	-11.44	1.15	-13.48	1.47	-6.55	4.49	-9.03	1.80
17	-11.43	1.09	-13.47	2.47	-6.5	5.27	-9.03	1.80
18	-11.33	1.94	-13.44	1.02	-6.47	4.98	-9.03	1.81

19	-11.05	1.39	-13.36	1.03	-6.44	4.47	-9.02	1.80
20	-10.54	1.07	-12.7	1.3	-6.32	4.95	-9.02	1.80

Cross-Docking within Same Group

After method validation and selecting the best protocol for Autodock, cross docking experiments were carried out. Cross docking experiments were carried out under two categories (a) group-I docking, i.e., cross docking within the same binding group (Fig 4), and (b) group-II docking, i.e., cross docking within different binding modes (Fig 5). Before carrying out docking experiments, it was hypothesized that molecules, which are experimentally known to be DNA groove binders (i.e., MBB from 1QV4, and HT from 127D), should give all docked modes indicating groove binding irrespective of the DNA receptor into which they are docked (as long as the DNA does not contain any intercalation gaps). Similarly, for DNA intercalators (EL from 1Z3F, and 2PT from 1XRW), it was hypothesized that when docked, such molecules should give all docked modes indicating intercalative binding with DNA irrespective of the DNA receptor into which they are docked (as long as the DNA contains intercalation gaps). Accordingly, cross docking experiments within the same binding mode (as depicted in Fig 4) were carried out, the results are summarized in Table 3. When MBB (a known DNA groove binder, extracted from 1QV4) was docked into DNA receptor from 127D, all docked modes indicated groove binding (rmsd ranging from 0.90 to 2.11Å), as expected, since the DNA receptor derived from 127D does not contain any intercalation gaps.

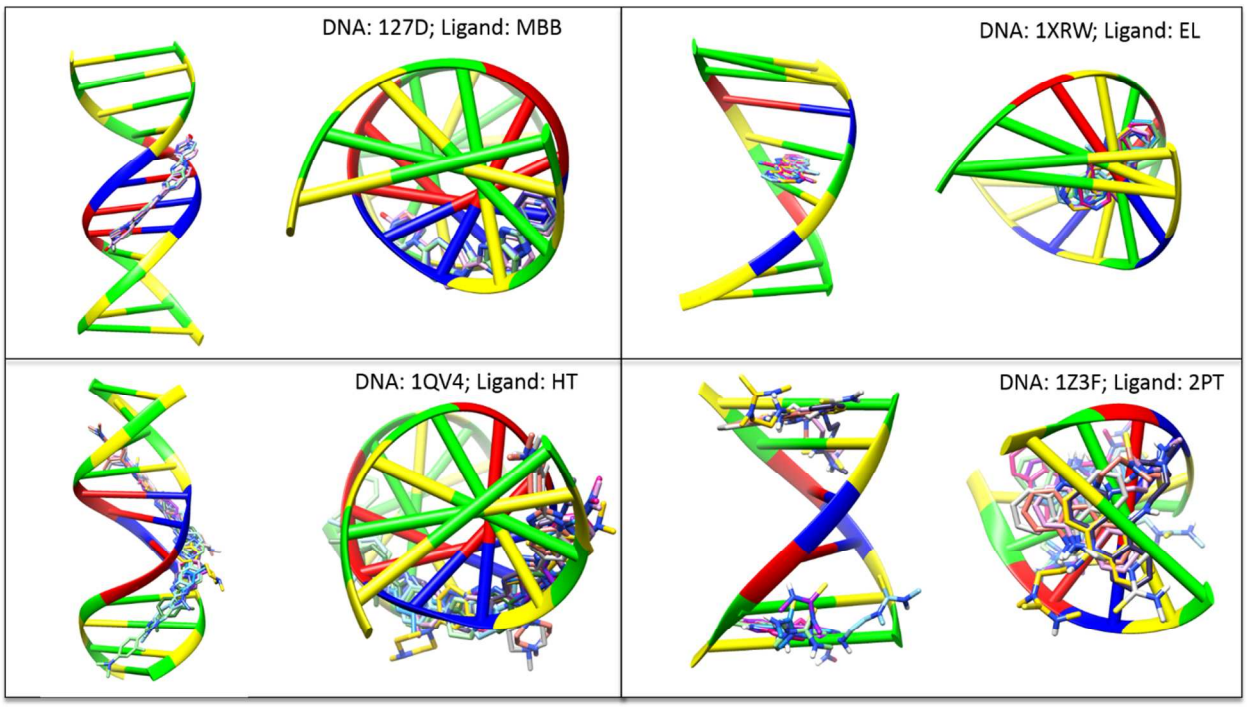


Figure 6. AutoDock results of cross docking within the same group. Known groove binders MBB and HT are docked into each other's respective DNA, all docked modes indicate groove binding as the only docked mode (left, top and bottom). Similarly, known intercalators EL and 2PT are docked into each other's respective DNA, all docked modes indicate intercalation as the only docked mode (right, top and bottom). DNA is in ladder representation, A/T = Red/Blue; G/C = green/yellow.

Similarly, when HT (a known DNA groove binder, extracted from 127D) is docked against DNA receptor derived from 1QV4 (without intercalation gaps), all groove binding docked modes were observed (rmsd ranging from 0.12-1.24Å). Similar trends were observed for DNA intercalators EL and 2PT, when docked into each other's respective DNA template. Since now both DNA receptors contained pre-formed intercalation gaps, all docked modes for EL (rmsd ranging from

0.45 to 0.46Å) and 2PT (rmsd ranging from 1.35 to 2.64Å) indicated intercalation into the pre-formed intercalation sites.

Table 3. AutoDock calculated binding free energies (ΔG, kcal/mol) for cross docking experiments within the same binding mode.

Receptor	**127D**[a]	**1QV4**[a]	**1XRW**[b]	**1Z3F**[b]
Ligand	**MBB**[c(a), c(b)]	**HT**[c(a), c(b)]	**EL**[d(a), d(b)]	**2PT**[d(a), d(b)]
# of docked conformations	ΔG kcal/mol	ΔG kcal/mol	ΔG kcal/mol	ΔG kcal/mol
1	-13.22	-12.75	-7.51	-8.05
2	-13.16	-12.75	-7.5	-7.64
3	-12.92	-12.74	-7.5	-7.56
4	-12.91	-12.74	-7.48	-7.55
5	-12.87	-12.74	-7.48	-7.54
6	-12.87	-12.73	-7.48	-7.46
7	-12.85	-12.72	-7.47	-7.45
8	-12.74	-12.68	-7.45	-7.32
9	-12.41	-12.55	-7.4	-7.27
10	-12.37	-12.52	-7.39	-7.26
11	-12.36	-12.5	-7.39	-7.24
12	-12.35	-12.42	-7.39	-7.21
13	-12.26	-12.31	-7.39	-7.16
14	-12.2	-12.09	-7.39	-7.07
15	-12.11	-11.9	-7.39	-6.99
16	-12.1	-11.64	-7.36	-6.88
17	-11.97	-11.6	-7.21	-6.87
18	-11.96	-11.49	-7.19	-6.83
19	-11.72	-11.45	-7.18	-6.54

20	-11.44	-11.44	-7.17	-6.31

[a]DNA receptor does not contain any intercalation gaps; [b]DNA receptor contains intercalation gaps; [c(a)]Known groove binder; [c(b)]All docked modes indicated groove binding; [d(a)]Known intercalator; [d(b)] All docked modes indicated intercalation.

Cross-Docking within Different Group

Cross docking experiments within different binding mode were carried out, as depicted in Fig 5. When known DNA intercalators EL from 1Z3F and 2PT from 1XRW, were docked into DNA receptors containing no intercalation gaps (derived from 127D and 1QV4), all docked modes were found to be groove binding in nature (Table 4). Since, now the DNA contains no intercalation gaps, no true intercalative docked modes can be expected.

Table 4. AutoDock calculated binding free energies (ΔG, kcal/mol) for cross docking experiments within different binding mode.

Receptor	**127D[a]**		**1QV4[a]**		**1Z3F[b]**		**1XRW[b]**	
Ligand	EL	2PT	EL	2PT	MBB	HT	MBB	HT
Conf #	d(a), d(b)	d(a), d(b)	d(a), d(b)	d(a), d(b)	c(a), c(b)	c(a), c(b)	c(a), c(b)	c(a), c(b)
1	-7.71	-10.76	-8.1	-9.84	-10.13	-9.5	-10.15	-8.78
2	-7.71	-10.59	-8.1	-9.77	-10.13	-9.47	-10.13	-8.6
3	-7.71	-10.5	-8.1	-9.69	-10.12	-9.45	-10.12	-8.58
4	-7.7	-10.48	-8.1	-9.64	-10.11	-9.44	-10.11	-8.57
5	-7.7	-10.47	-8.1	-9.63	-10.08	-9.41	-10.05	-8.55
6	-7.7	-10.43	-8.1	-9.6	-10.06	-9.37	-10.02	-8.54
7	-7.7	-10.42	-8.08	-9.5	-10.06	-9.37	-10.0	-8.45
8	-7.7	-10.37	-8.08	-9.48	-10.05	-9.3	-9.98	-8.44
9	-7.7	-10.33	-8.08	-9.47	-10.03	-9.29	-9.94	-8.41
10	-7.69	-9.96	-8.08	-9.44	-9.94	-9.29	-9.76	-8.39
11	-7.69	-9.81	-8.08	-9.43	-9.88	-9.27	-9.74	-8.38

12	-7.68	-9.8	-8.08	-9.43	-9.77	-9.26	-9.73	-8.37
13	-7.68	-9.76	-8.08	-9.42	-9.74	-9.17	-9.72	-8.36
14	-7.68	-9.47	-8.08	-9.41	-9.72	-9.14	-9.7	-8.3
15	-7.68	-9.41	-8.07	-9.4	-9.71	-9.11	-9.59	-8.28
16	-7.68	-9.39	-8.07	-9.22	-9.58	-9.09	-9.59	-8.27
17	-7.68	-9.35	-8.07	-9.06	-9.57	-9.07	-9.44	-8.27
18	-7.68	-9.34	-8.07	-9.0	-9.53	-9.01	-9.43	-8.26
19	-7.68	-9.25	-8.07	-8.85	-9.36	-8.97	-9.19	-8.24
20	-7.63	-8.6	-8.06	-8.62	-9.32	-8.71	-9.1	-8.21

[a]DNA receptor does not contain any intercalation gaps; [b]DNA receptor contains intercalation gaps; [c(a)]Known groove binder; [c(b)]All docked modes indicate intercalation; [d(a)]Known intercalator; [d(b)] All docked modes indicate groove binding.

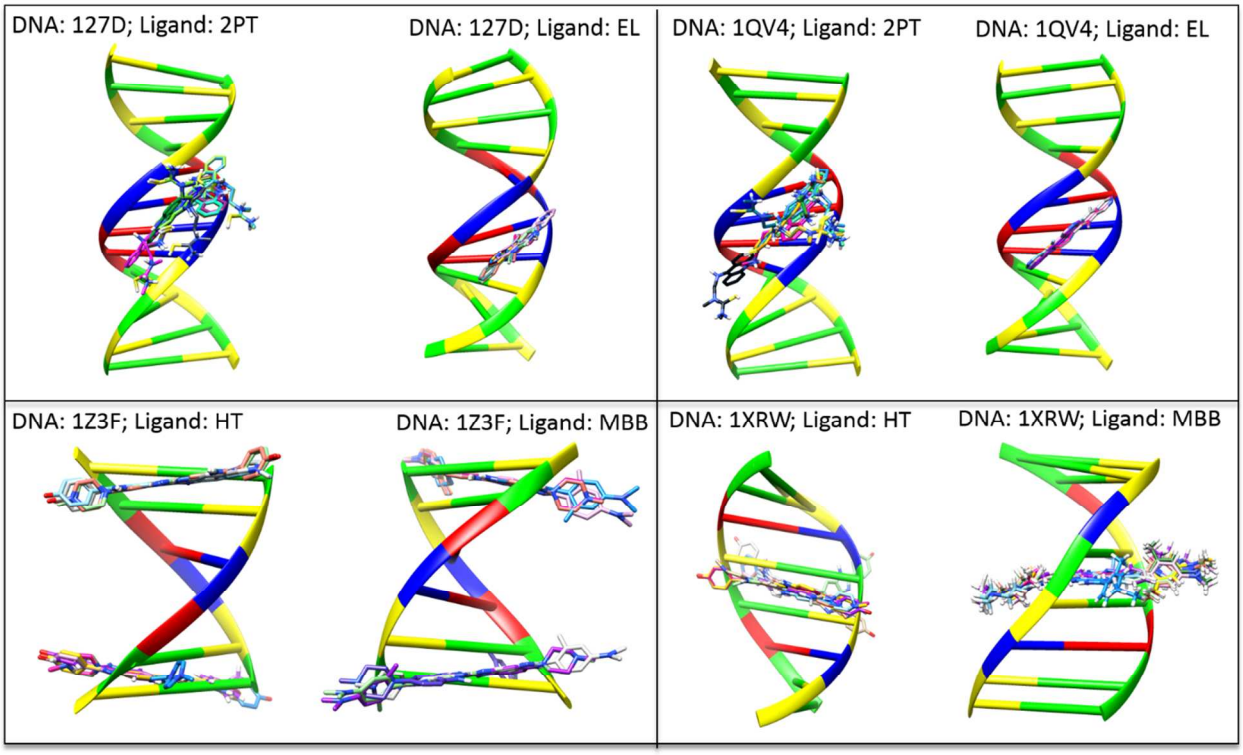


Figure 7. AutoDock results of cross docking within different group. Known intercalators 2PT and EL are docked into DNA receptor (derived from 127D and 1QV4) without intercalation gaps, all docked modes indicate groove binding (top, left and right). Known groove binders are docked into DNA receptor (derived from 1Z3F and 1XRW) containing intercalation gaps, all

docked modes indicate intercalation (bottom, left and right). DNA is in ladder representation, A/T = Red/Blue; G/C = green/yellow.

Similarly, known DNA groove binders (MBB from 1QV4, and HT from 127D), were docked into DNA containing pre-formed intercalation gaps. The DNA receptors containing pre-formed intercalation gaps were of two types, (a) DNA receptor (PDB id: 1Z3F) containing two intercalation gaps, and (b) DNA receptor (PDB id: 1XRW) containing only one intercalation gap. When MBB and HT (both known DNA groove binders), were docked into these DNA receptors, all docked modes indicated intercalation as the only mode of binding, no groove binding modes were observed in any of the docked conformations (Table 4). These intercalative docked modes indicate false binding modes for MBB and HT, which are known to be DNA groove binders and not DNA intercalators, yet when the DNA receptor contains intercalation gaps, many misleading docked modes are obtained. These results are in agreement with previous studies on AutoDock [12].

Docking with canonical B Form DNA

For minor groove binders (HT and MBB), docking with canonical B-form DNA was also carried out, for comparison. Canonical B-form DNA having 12 base pairs (CGGCCGGCATATTATACGGCCGGC) was constructed using online resource, 3D DART (3DNA Driven DNA Analysis and Rebuilding Tool) [23]. 3D-DART server provides an easy way of generating and modelling custom structural models of DNA, it utilizes the renowned DNA analysis software 3DNA [24]. As can be expected, all groove binding modes were observed for both HT (rmsd from 0.98-1.18Å) and MBB (rmsd from 1.47-2.02Å). The AutoDock calculated binding free energies (ΔG, kcal/mol, Table 5) were found to be less than that obtained via self-docking of HT and MBB in their own crystallographic DNA. The binding free energies were also less than those observed for cross docking experiments for groove binders within the same binding group. Figure 8 shows overlap of docked conformations of HT and MBB in B-DNA. Docked conformation were visualized using Discovery Studio Visualizer 4.0 [25].

Table 5. Binding free energies for docking of groove binders HT and MBB against B-form DNA.

Conf. #	**HT** **ΔG (kcal/mol)**	**MBB** **ΔG (kcal/mol)**
1	-11.81	-11.09
2	-11.79	-11.07
3	-11.79	-11.07
4	-11.77	-11.06
5	-11.76	-11.06
6	-11.76	-11.05
7	-11.71	-10.98
8	-11.66	-10.96
9	-11.64	-10.89
10	-11.59	-10.88
11	-11.51	-10.75
12	-11.51	-10.73
13	-11.45	-10.72
14	-11.37	-10.72
15	-11.36	-10.71
16	-11.33	-10.68
17	-11.01	-10.64
18	-10.98	-10.51
19	-10.52	-10.51
20	-10.26	-10.24

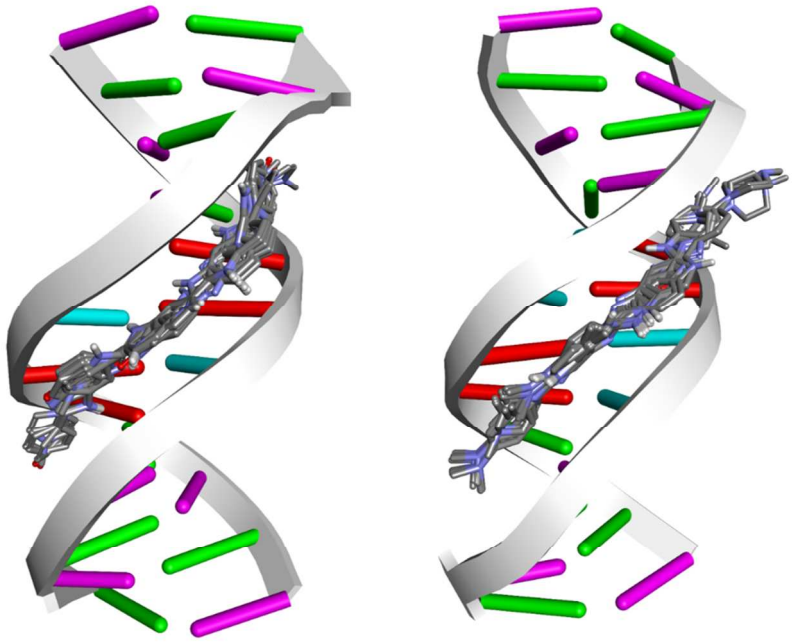


Figure 8. Overlap of all docked conformations of HT (left) and MBB (right), when docked against generic B-form DNA, A/T = Red/Blue; G/C = green/purple.

Docking using Hex 8.0: Shape Complementarity Based Docking

In this research, we wanted to investigate how shape-based docking methods (which make use of a completely different algorithm than AutoDock) would perform in comparison to Autodock. For this purpose Hex 8.0 [19] was selected. Hex is a well-known shape-based docking software that is ideal for predicting interactions based on shape complementarity of receptor (here, DNA) and ligand. In Hex, the docking search is started by rotating the ligand and receptor about their centroids at a range of intermolecular distances. Two Euler rotation angles are assigned to each receptor and ligand, and the final rotation is defined as a twist of the ligand about the intermolecular axis, afterwards a full six-dimensional search is performed over the full rotational ranges [19]. For docking, shape-only correlation was selected. The default settings of the software were kept to perform an initial Steric Scan at N=16, followed by a Final Search at N=25, this part uses just the steric contribution to the docking energy. The first step of search, that is Steric Scan (at N = 16), is fast but of low resolution than the second step, Final Search (at

N=25), which is of high resolution and uses smaller distance increments. This strategy allows the search space to be covered more rapidly (but relatively coarsely) in the first phase, and more meticulously in the final phase. In other words, the default behavior is essentially to scan the search space at 1Å resolution, but to perform the high resolution scoring at 0.5 Å resolution. The resulting orientations are then sorted by calculated energy, and a new set of trial orientations is generated for the top scoring 10,000–20,000 orientations. The software then calculates the surface skin coefficients after which docking correlation scores are calculated at each of the specified angular and intermolecular increments. A Cartesian grid is used to sample the molecular skins numerically, this grid only needs to contain the larger of the two molecules (receptor or ligand) so that much finer sampling grids can be applied afterwards.

While over-sampling of search space is recommended over under-sampling (in which case a good solution can be missed by the search space), however, this can cause multiple similar but incorrect orientations to push good solutions down the list. In order to overcome this problem, Hex uses a simple clustering algorithm to group spatially similar docking orientations. Each docking solution is first ordered by energy, and the lowest energy solution is made the seed orientation for the first cluster. The list is then searched down to a given depth for other similar orientations whose rmsd is within a given threshold (default 3Å) of the seed orientation, these orientations are then assigned to the first cluster. The process is then repeated starting from the next lowest unassigned orientation, until all solutions have been assigned to a cluster. A total of 2000 docked solutions are generated which are arranged into 10 clusters. First conformation (most favorable energy) from each of the 10 clusters was selected. There are 200 solutions in each cluster. By writing expressions for the overlap of pairs of parametric functions, an overall docking score is obtained as a function of the six degrees of freedom in a rigid body docking search [19].

Cross docking experiments within the same binding mode and within different binding mode, as designed earlier for AutoDock (Fig 4-5), were performed using Hex 8.0. Results for cross docking experiments within the same binding mode are depicted in Table 5. As expected, when known DNA groove binding molecules, MBB and HT were docked into each other's respective DNA receptor (without any intercalation gaps), all docked modes indicated groove binding as shown in Fig. 8. These results are in agreement with the results from similar cross docking

experiments using AutoDock. Similarly when known DNA intercalator molecules EL and 2PT were docked into each other's respective DNA, all docked modes were found to be intercalating in nature (Fig 9). Again these results were expected and in agreement with the results from similar docking experiments using AutoDock.

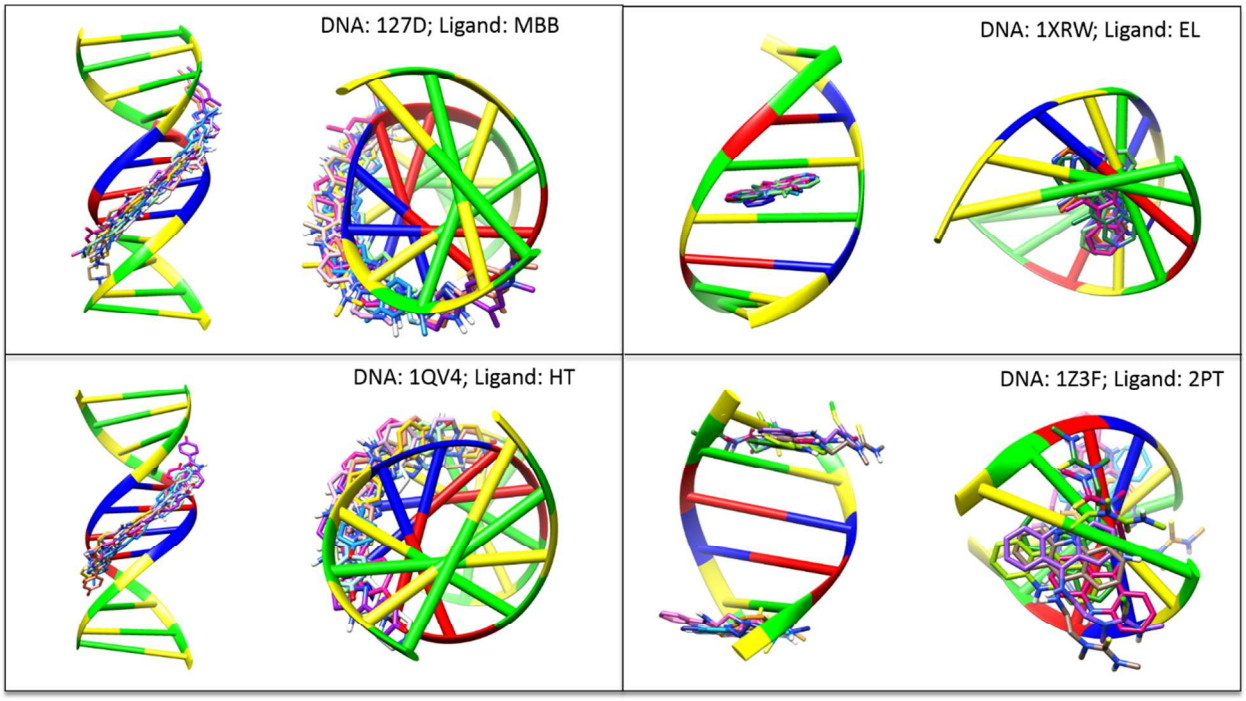


Figure 9. Hex results of cross docking within the same group. Known groove binders MBB and HT are docked into each other's respective DNA, all docked modes indicate groove binding as the only docked mode (left, top and bottom). Similarly, known intercalators EL and 2PT are docked into each other's respective DNA, all docked modes indicate intercalation as the only docked mode (right, top and bottom). DNA is in ladder representation, A/T = Red/Blue; G/C = green/yellow.

Table 6. Hex results within same binding mode showing groove binding (GB) or intercalative mode (IC) of binding.

Receptor	**1Z3F**	**Etotal**	**1XRW**	**Etotal**	**127D**	**Etotal**	**1QV4**	**Etotal**
Ligand	**2PT**		**EL**		**MBB**		**HT**	
1	IC	-257.3	IC	-239.7	GB	-422.4	GB	-334.5
2	IC	-222.6	IC	-233.0	GB	-394.0	GB	-320.8

3	IC	-216.7	IC	-231.2	GB	-385.7	GB	-315.6
4	IC	-213.1	IC	-229.9	GB	-379.2	GB	-312.5
5	IC	-210.7	IC	-228.7	GB	-374.0	GB	-309.6
6	IC	-208.7	IC	-227.7	GB	-369.6	GB	-307.1
7	IC	-206.9	IC	-226.8	GB	-366.5	GB	-304.9
8	IC	-205.4	IC	-226.5	GB	-362.9	GB	-302.9
9	IC	-204.3	IC	-226.2	GB	-359.8	GB	-301.0
10	IC	-203.2	IC	-226.0	GB	-357.3	GB	-299.5

When known DNA intercalator molecules EL and 2PT were docked into DNA receptor, without any intercalation gaps, a slight difference in nature of docked modes (as compared to same obtained via AutoDock) was observed. When a completely planar, rigid molecule EL (a known intercalator) was docked into DNA receptors without any intercalation gaps (derived from 127D and 1QV4), some of the docked modes indicated that the elipticine ring was not properly aligned along the grooves of DNA (Fig 10), although none of these modes can be considered as true intercalating modes, these modes are not true groove binding either. Similar trends were observed when another intercalator molecule 2PT was docked into DNA receptors without any intercalation gaps.

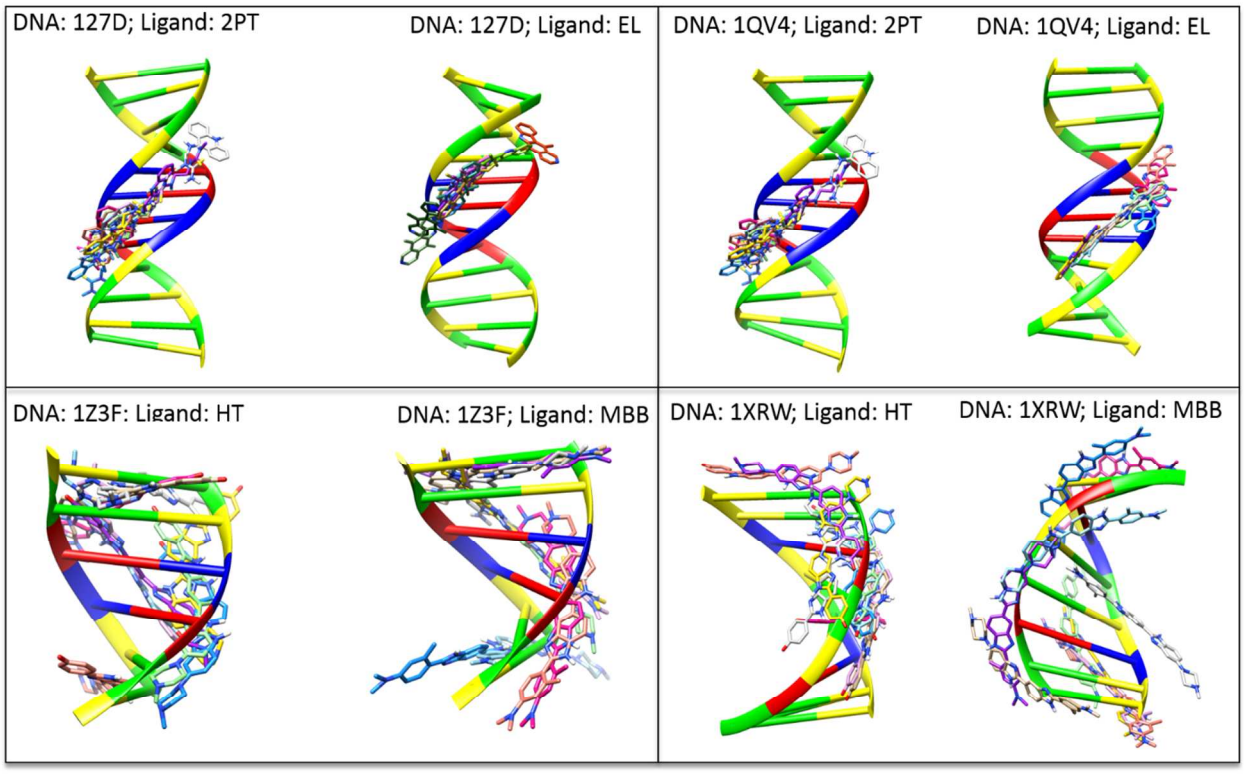


Figure 10. Hex results of cross docking within different group. Known intercalators 2PT and EL are docked into DNA receptor (derived from 127D and 1QV4) without intercalation gaps (top,

left and right). Known groove binders are docked into DNA receptor (derived from 1Z3F and 1XRW) containing intercalation gaps (bottom, left and right). DNA is in ladder representation, A/T = Red/Blue; G/C = green/yellow.

Hex docking results for cross docking within different binding mode are given in Table 7. When known groove binder molecules, HT and MBB were docked into a DNA receptor containing intercalation gaps, a marked improvement in number of docked modes that indicated groove binding was noted as compared to AutoDock. When HT is docked into DNA containing intercalation gaps (derived from 1XRW), only 20% docked modes were found to be intercalative and 80% were groove binding in nature. When HT was docked into another DNA containing intercalation gaps (derived from 1Z3F), 40% of docked modes indicated intercalation while 60% indicated groove binding (Table 7). Similarly, MBB (known groove binder) gave only 20% intercalating modes and 80% groove binding modes when docked against 1XRW. When MBB was docked against 1Z3F, 50% intercalating and 50% groove binding modes were observed (Table 7). This indicates a marked improvement as compared to AutoDock Hex was shown to outperform AutoDock when the DNA contains intercalation gaps. This has been compared in table 8.

Table 7. Cross docking within diff mode when DNA receptor contains intercalation gaps.

Receptor	**1Z3F**	**Etotal**	**1Z3F**	**Etotal**	**1XRW**	**Etotal**	**1XRW**	**Etotal**
Ligand	**HT**		**MBB**		**HT**		**MBB**	
1	IC	-286.6	IC	-300.3	GB	-317.8	GB	-328.5
2	GB	-268.7	IC	-269.0	GB	-272.3	GB	-292.2
3	GB	-264.5	GB	-261.8	GB	-264.6	GB	-284.5
4	GB	-261.2	GB	-258.0	GB	-259.8	GB	-280.1
5	IC	-258.7	GB	-255.0	GB	-256.0	GB	-276.2
6	IC	-256.8	IC	-252.4	IC	-252.9	IC	-273.2
7	IC	-255.2	GB	-250.2	GB	-250.3	IC	-270.7
8	GB	-253.7	GB	-248.2	IC	-248.4	GB	-268.6
9	GB	-252.3	IC	-246.3	GB	-246.7	GB	-266.8
10	GB	-251.1	IC	-244.9	GB	-245.2	GB	-265.2

Table 8. % Distribution of docked intercalating (IC) and groove binding (GB) modes obtained from AutoDock and Hex.

	Receptor	Ligand	% IC modes (AutoDock; Hex)	% GB modes (AutoDock; Hex)
Within same binding mode	1QV4	HT	0; 0	100; 100
	127D	MBB	0; 0	100; 100
	1Z3F	2PT	100; 100	0; 0
	1XRW	EL	100; 100	0; 0
Within different binding mode	1XRW	HT	100; 20	0; 80
	1XRW	MBB	100; 20	0; 80
	1Z3F	HT	100; 40	0; 60
	1Z3F	MBB	100; 50	0; 50

AutoDock treats receptor as rigid and ligand as flexible. As a result when the DNA receptor contains pre-formed intercalation gaps, the ligand, since it is treated as a flexible molecule with many possible torsions and conformations, the genetic algorithm implemented in AutoDock gives a "misfit" or in other words "makes" the ligand "fit" the intercalation gap, in this case many misleading intercalating docked poses are generated for known groove binder molecules, even the AutoDock calculated binding free energies are favorable and cannot always reliably serve as a parameter to sort out "false-positive" docked modes. On the other hand the shape complementarity based docking method (as implemented in Hex), treats both receptor and ligand as rigid, this is a fundamental difference from the AutoDock (where receptor is rigid and ligand is kept flexible). While generally, long chain, crescent shaped molecules are known to interact with DNA as groove binders (since owing to their shape they typically align themselves along the grooves of DNA), molecules containing rigid, planar rings are DNA intercalators, since owing to the presence of rigid, planar rings, it now becomes easy for such molecules to stack in-between the DNA base pairs, causing an intercalation gap to appear in the DNA. Although, false-

positive docked modes were also obtained in case of shape based docking, they were less frequent as compared to AutoDock (Table 8). To further resolve these ambiguities, use of molecular dynamics simulation is highly recommended.

Experimental

X-ray crystal structures of DNA intercalators, PDB IDs 1XRW [26] and 1Z3F [27], and DNA groove binders PDB IDs 127D [28] and 1QV4 [29], deposited at the Protein data bank (PDB) [30] were selected for the study. In order to carry out docking studies using genetic algorithm, AutoDock 4.2 [18] was used. Series of self-docking and cross-docking studies were carried out for both DNA intercalators and groove binders. Self-docking studies served the purpose of method validation. For self-docking studies the molecule (intercalator or groove binder) that had co-crystallized with the DNA fragment was extracted and prepared as for docking by adding hydrogen atoms, merging non-polar hydrogen atoms and adding Gasteiger charges using MGL Tools [18]. Since AutoDock is not parametrized for use with platinum metal, this was removed from the ligand 2PT (PDB id: 1XRW). The remaining DNA fragment, from which the ligands were extracted, was similarly prepared as receptor by deleting all solvent and hetero molecules, adding hydrogen atoms and Gasteiger charges. The AutoDock 4.2 parameters used were 20 GA runs and $5x10^6$ energy evaluations, the grid box was large enough to cover the entire DNA fragment [15]. The docking method selected was able to replicate the experimentally observed bound conformation of ligand. To carry out docking using shape complementarity based methods, Hex 8.0 [19] was used with default parameters where correlation type was set to Shape Only, 3D FFT mode was selected, both receptor and ligand range were 180 (each at a step size of 7.5), grid dimension was 0.6 (default), twist range was 360 (step size 5.5), and distance range was 40 Å. A total of 2000 docking solutions were generated, which were arranged in 10 clusters (rmsd threshold was 3Å within each cluster), each cluster contained 200 docked solutions.

Conclusion

In this study we find that the accuracy of DNA docking is directly linked with the selection of a correct DNA template, that is, with or without pre-formed intercalation gaps. It is suggested that if the experimental evidence regarding the nature of interaction of a molecule with DNA is known (that is, whether intercalator or groove binder), a DNA receptor (with or without pre-formed intercalation gaps) should accordingly be selected, AutoDock can then be used reliably and can give very accurate results. The problem arises when the nature of interaction of ligand with DNA is not known. In such case, we have presented a comparison between AutoDock and Hex. We find that the overall number of false-positive docked modes obtained via AutoDock are greater than that obtained via Hex, however an unambiguous preference for shape-based docking methods cannot be ascertained, as a result we recommend the use of other techniques such as molecular dynamics simulation and thermodynamics integration to further resolve such ambiguities.

Acknowledgements

The authors wish to thank Dr. Muhammad Shahid for helpful discussions.

References

1. D. Lebwohl, R. Canetta. *Eur. J. Cancer*. 1998, **34**(10), 1522–34.
2. E.M. Walker Jr, S.M. Walker. *Ann. Clin. Lab. Sci*. 1999, **29**(4), 263–74.
3. P.R. Turner, W.A. Denny. *Mutat. Res*. 1996, **355**(1-2), 141–69.
4. P.G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M.A. Tabrizi, M.G. Pavani, R. Romagnoli. *Med. Res. Rev*. 2004, **24**(4), 475–528.

5. M.F. Brana, M. Cacho, A. Gradillas, B. de Pascual-Teresa, A. Ramos. *Curr. Pharm. Des*. 2001, **7**(17), 1745–80.
6. N.J. Wheate , C.R. Brodie, J.G. Collins, S. Kemp, J.R. Aldrich-Wright JR. *Mini Rev. Med. Chem*. 2007, **7**(6), 627–48.
7. R. Martínez, L. Chacón-García. *Curr. Med. Chem*. 2005, **12**(2), 127–51.
8. X.Y. Meng, H.X. Zhang, M. Mezei, M. Cui. *Curr. Comput. Aided Drug Des*. 2011, **7**(2), 146–57.
9. R.T. Kroemer. *Curr. Protein Pept. Sci*. 2007, **8**(4), 312–28.
10. G.D. Geromichalos. *J. B.U.O.N*. 2007, 12 Suppl 1:S101–18.
11. B.K. Shoichet, S.L. McGovern, B. Wei, J.J. Irwin. *Curr. Opin. Chem. Biol*. 2002, **6**(4), 439–46.
12. D. L. Ma, T.S. Lai, F.Y. Chan, W.H. Chung, R. Abagyan, Y. C. Leung, K. Y. Wong. *ChemMedChem*. 2008, **3**(6), 881 – 4.
13. D.S. Chan, H.M. Lee, F. Yang, C. M. Che, C. C. Wong, R. Abagyan, C. H. Leung, D. L. Ma. Angew Chem Int Ed Engl. 2010, 49(16), 2860 – 4.
14. D. L. Ma, D. S. H. Chana, C. H. Leung. *Chem. Sci*., 2011, **2**, 1656–1665.
15. C.G. Ricci, P.A. Netz. *J. Chem. Inf. Model*. 2009, **49**(8), 1925–35.
16. Y. Gilad and H. Senderowitz. *J. Chem. Inf. Model*., 2014, **54** (1), 96–107.
17. P. A. Holt, J. B. Chaires, and J. O. Trent. *J. Chem. Inf. Model*. 2008, **48**(8), 1602–1615.
18. G.M. Morris, R. Huey, W. Lindstrom, et al. *J. Comput. Chem*. 2009, **16**, 2785–91.
19. L. Mavridis and D.W. Ritchie. *J. Chem. Inf. Model*. 2007, **47**(5), 1787–1796. (http://hex.loria.fr/hex.php, including Hex manual)
20. E.C. Meng, B. K. Shoichet, I.D. Kuntz. *J. Comput. Chem*., 2004, **13**, 505–524.

21. G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson. *J. Comput. Chem*, 1998, **19**(14), 1639–1662.

22. B.K. Shoichet, I.D. Kuntz, D.L. Bodian. Molecular docking using shape descriptors. *J. Comput. Chem*, 2004, **13**(3), 380–397.

23. M. van Dijkm A.M.J.J. Bonvin. *Nucl. Acids Res*., 2009, **37** (Web Server Issue):W235-W239.

24. X. J. Lu, W. K. Olson. *Nucl. Acids Res*., 2003, **31**(17), 5108-21.

25. Accelrys Sofware Inc., Discovery Studio Visualizer, Release 4.0, San Diego, Accelrys Sofware Inc., 2013.

26. H. Baruah, M.W. Wright, U. Bierbach. *Biochemistry*. 2005, **44**(16), 6059–70.

27. A. Canals, M. Purciolas, J. Aymamí, M. Coll. *Acta Crystallogr D Biol Crystallogr*. 2005, **61**(Pt 7), 1009–12.

28. M. Sriram, G.A. van der Marel, H.L. Roelen, J.H. van Boom, A.H. Wang. *EMBO J.* 1992, **11**(1), 225–32.

29. R.F. Martin, S. Broadhurst, M.E. Reum, C.J. Squire, G.R. Clark, P.N. Lobachevsky, J.M. White, C. Clark, D. Sy, M. Spotheim-Maurizot, D.P. Kelly. *Cancer Res*. 2004, **64**(3), 1067–70.

30. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. *Nucleic Acids Res.*, 2000, **28**, 235–242.