



Protein–small molecule docking with receptor flexibility in iMOLSDOCK

D. Sam Paul¹ · N. Gautham¹

Received: 19 March 2018 / Accepted: 11 August 2018 / Published online: 20 August 2018
© Springer Nature Switzerland AG 2018

Abstract

We have earlier reported the iMOLSDOCK technique to perform ‘induced-fit’ peptide–protein docking. iMOLSDOCK uses the mutually orthogonal Latin squares (MOLSS) technique to sample the conformation and the docking pose of the small molecule ligand and also the flexible residues of the receptor protein, and arrive at the optimum pose and conformation. In this paper we report the extension carried out in iMOLSDOCK to dock nonpeptide small molecule ligands to receptor proteins. We have benchmarked and validated iMOLSDOCK with a dataset of 34 protein–ligand complexes as well as with Astex Diverse dataset, with nonpeptide small molecules as ligands. We have also compared iMOLSDOCK with other flexible receptor docking tools GOLD v5.2.1 and AutoDock Vina. The results obtained show that the method works better than these two algorithms, though it consumes more computer time. The source code and binary of MOLSS 2.0 (under a GNU Lesser General Public License) are freely available for download at <https://sourceforge.net/projects/mols2-0/files/>.

Keywords Molecular docking · Protein–ligand docking · Induced-fit docking · Side-chain flexibility · iMOLSDOCK · Mutually orthogonal Latin squares

Introduction

Molecular docking programs are used to find the best binding pose of a ligand in the protein binding site. Apart from protein–ligand interactions, they are also used to study protein–protein and protein–DNA interactions. In the past 30 years, a large number of docking programs have been developed using various techniques [1]. Docking problems are computationally addressed by combining an accurate representation of the intermolecular interactions, an efficient algorithm and a scoring function. The algorithm in the docking program searches for potential binding modes and the scoring function discriminates and ranks the predicted structures [2].

Initial docking methods treated the ligand and receptor both as fixed structures. The development progressed by including variations in the internal degrees of freedom of the ligand [3]. However, receptor proteins do not remain rigid, but change conformation to accommodate the ligand [4–6]. Active site plasticity of the receptor proteins was also established by comparing the apo and holo forms of proteins. Though main-chain conformations are largely preserved, in most of the proteins significant differences in side-chain conformations occur upon ligand binding [7, 8]. Therefore, docking algorithms which allowed receptor flexibility have been developed [9, 10]. GOLD [11], GLIDE [12], AutoDock Vina [13] and RosettaLigand [2] are some of the docking programs with receptor flexibility. GOLD allows side-chain receptor flexibility and ensemble docking [11]. In GLIDE, a protein structure prediction technique called Prime is used for treating protein flexibility [12]. AutoDock Vina allows side-chain flexibility for the selected protein residues [13]. In RosettaLigand only side-chain conformational changes were initially allowed [2]. Recently, backbone flexibility has also been incorporated [14].

To address the problem of molecular docking, in our laboratory we have developed a docking program called MOLSDOCK [15, 16], which uses the mutually orthogonal

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10822-018-0152-8>) contains supplementary material, which is available to authorized users.

✉ N. Gautham
n_gautham@hotmail.com

¹ Centre of Advanced Study in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai 600025, India

Latin squares (MOLSs) technique [17]. The MOLS sampling algorithm identifies a small sample of the vast multi-dimensional search space, which is nevertheless completely representative of this space. The energy values are calculated at each of the sampled points. A variant of the mean field method [18] analyses these energy values simultaneously to obtain the optimal conformation [19]. MOLSDOCK was tested for docking peptides [15], small molecules [16] and nucleotides [20]. In all the above cases, only the ligand was treated as flexible and the receptor was kept rigid. Later, we upgraded MOLSDOCK by incorporating receptor protein side-chain flexibility [21]. The upgraded version of MOLSDOCK, called iMOLSDOCK, was benchmarked and tested for docking peptides. However nonpeptide small organic molecules are most commonly used as drugs. Therefore we have further extended the ‘induced-fit’ docking in iMOLSDOCK to dock nonpeptide small molecule ligands, and in this paper we report this extension. We have benchmarked iMOLSDOCK with this extension using 34 protein–ligand complexes selected from the Protein Data Bank (PDB) as well as with Astex Diverse dataset [22]. We also have compared the performance of iMOLSDOCK with two popular flexible receptor docking tools: AutoDock Vina [13] and GOLD v5.2.1 [11]. We present these results here.

Materials and methods

iMOLSDOCK is a ‘flexible receptor/flexible peptide’ docking method developed using the MOLS method [17]. The MOLS method has been described in detail elsewhere [17, 23, 24]. For completeness, we give a brief description here. We will first explain the method by applying it to the prediction of the minimum energy structure of a peptide. The MOLS method systematically searches the conformational energy space of the peptide to arrive at the optimum peptide structure. The conformational space of a peptide may be described as the set of all possible combinations of all values of all its variable torsion angles. Consider a peptide having ‘m’ torsion angles, with each torsion angle taking up ‘n’ different values. Then, there are $(n)^m$ conformations for the peptide. The potential energy for any of these conformations of the peptide may be calculated using an energy function such as AMBER force field [25]. The next task is to locate the minimum energy conformation. Any attempt to search through all the possible $(n)^m$ conformations of the peptide will lead to combinatorial explosion. In MOLS method, the method of MOLSs [26] is used to systematically choose a set of $(n)^2$ points (or conformations) from the $(n)^m$ overall conformational space of the peptide. After choosing the $(n)^2$ conformations, the potential energy for each of these chosen conformations is calculated. The $(n)^2$ energy values are then analyzed using a variant of the mean field technique [18, 19]

to arrive at the lowest energy conformation. The $(n)^2$ points could be selected in large number of ways [27], and each choice leads to either the same, or to a different low energy structure. We have shown repeatedly [15, 21, 28] that for small peptides of length up to about 10 residues, choosing the $(n)^2$ points in 1500 different ways (which would yield 1500 low energy conformations) is sufficient to identify all the unique, mutually dissimilar low energy structures.

The MOLS method was extended to address the docking problem [15, 16, 20]. The docking tool, dubbed MOLSDOCK, performed ‘flexible ligand - rigid receptor’ docking. In MOLSDOCK, the variables specifying the position and orientation of the ligand were added to the set of variable torsion angles of the ligand to define the search space. Accordingly, the scoring function for MOLSDOCK was also modified to include the interaction energy between the protein and the ligand, along with the intramolecular energy of the ligand. MOLSDOCK was developed as a program suitable for studying the docking of peptides, nonpeptide small organic molecules and nucleotides to protein receptors [15, 16, 20]. MOLSDOCK was upgraded to iMOLSDOCK [21] by including receptor flexibility. Two major changes were made in iMOLSDOCK. (a) The search space was further increased to include the side-chain torsion angles of the receptor protein. (b) The intra-protein energy, to assess the receptor protein conformation, was added along to the intra-ligand energy and protein–ligand interaction energy.

Thus, we specify the conformation of the small molecule ligand by ‘s’ torsion angles (θ_r , $r = 1, s$). Six additional parameters describe the ligand’s docking pose, i.e., three for the position and three for the orientation of the ligand in the receptor binding site. If there are ‘p’ torsion angles that describe the flexible residues in the receptor binding site, then we have a total of ‘s + 6 + p’ dimensions in the search space (θ_r , $r = 1, s + 6 + p$). The volume of the search space is $(n)^{s+6+p}$, if each dimension is sampled at ‘n’ intervals. Out of the $(n)^{s+6+p}$ points in the search space, the MOLS method calculates the values of the scoring function only at $(n)^2$ points, and analyzes them to simultaneously locate the optimum conformation of the ligand, its pose, and also the conformation of the side-chains of the receptor flexible residues. Since the search space is defined on a discrete grid of ‘s + 6 + p’ points, the actual optimum may be off-grid and lie close to but not actually on the grid point. Therefore to identify the nearest off-grid optimum, we perform a gradient minimization [29] as the final step.

The ligand (any organic chemical compound) is specified in MDL Molfile (.mol) format. The rotatable bonds in the ligand are identified using *findrotatable.pl* which is available at http://www.ccl.net/cca/software/PERL/Find_Rotatable_Bonds/. All the variable torsion angles of the ligand are sampled from 0° to 360° at intervals of 10°. As described by Arun Prasad and Gautham [15], the

orientation of the ligand in the binding site is specified by three angles, two of which specify the position of a rotation axis for the ligand. The third angle is the angle of rotation about this axis. The ligand translates inside a 5 Å cubic box centered at the midpoint of the receptor binding site. To test successful docking in iMOLSDOCK, it

is necessary to dock the ligands to a pre-defined binding site before blind docking [30]. Therefore, for all the test cases (Table 1), we defined the binding site of the ligand. The binding site information was extracted from the holo protein–ligand crystal structure. (If the binding site in the receptor protein is not known, then binding site may be

Table 1 The 34 protein–ligand complex structure of the benchmarking dataset

PDB ID of holo structure	Resolution of the crystal structure (Å)	PDB ID of apo structure	Resolution of the crystal structure (Å)	No. of ligand torsion angles	Name of protein	RMSD ^a of best sampled ^b structure (Å)	
						GAFF ^c	MMFF ^d
1AI7	2.50	1PNK	1.90	1	Penicillin amidohydrolase	0.25	0.27
1TNH	1.80	5PTP	1.34	2	Serine protease inhibitors	1.29	1.67
1YDR	2.20	1J3H	2.90	2	C-AMP-dependent protein kinase	3.81	3.96
1AI5	2.36	1PNK	1.90	3	Penicillin amidohydrolase	2.50	2.47
1GHB	2.00	4CHA	1.68	3	Gamma-chymotrypsin	1.31	1.75
1OKL	2.10	3GZ0	1.26	3	Carbonic anhydrase II	1.06	1.06
1TNJ	1.80	1S0Q	1.02	3	Trypsin	2.23	2.26
1TPP	1.40	1S0Q	1.02	3	Beta-trypsin	2.95	2.85
3MTH	1.90	4INS	1.50	3	Methylparaben insulin	3.11	2.33
1AJN	2.36	1PNK	1.90	3	Penicillin amidohydrolase	2.90	2.98
1AJQ	2.05	1PNK	1.90	3	Penicillin amidohydrolase	0.72	1.09
1NGP	2.40	1NGQ	2.30	4	NIG9	1.71	1.86
1TNK	1.80	1S0Q	1.02	4	Trypsin	1.31	1.52
4EST	1.78	5AVD	0.86	4	Elastase	2.04	2.25
1AI6	2.55	1PNK	1.90	4	Penicillin amidohydrolase	2.27	2.31
1AJP	2.31	1PNK	1.90	4	Penicillin amidohydrolase	2.59	2.45
1EBG	2.10	3ENL	2.25	5	Enolase	1.97	2.08
1TNI	1.90	1S0Q	1.02	5	Trypsin	2.24	2.17
1XIE	1.70	3KBJ	2.00	5	D-Xylose isomerase	2.65	2.63
2CTC	1.40	1YME	1.53	5	Carboxypeptidase A	1.72	1.87
1AI4	2.35	1PNK	1.90	5	Penicillin amidohydrolase	2.99	3.03
1PDZ	2.20	3ENL	2.25	6	Enolase	2.62	2.60
1TPH	1.80	8TIM	2.50	6	Triosephosphate isomerase	2.71	2.79
1XID	1.70	3KBJ	2.00	6	D-Xylose isomerase	2.95	2.85
2YPI	2.50	7TIM	1.90	6	Triose phosphate isomerase	2.67	2.56
1BLH	2.30	3BLM	2.00	7	Beta-lactamase	2.62	2.70
1CBX	2.00	1YME	1.53	7	Carboxypeptidase A	2.10	2.03
2PK4	2.25	1PK4	1.90	7	Human plasminogen kringle 4	0.49	1.18
2CMD	1.87	3HHP	1.45	9	Malate dehydrogenase	2.71	2.43
2XIS	1.71	3KBJ	2.00	9	Xylose isomerase	2.68	2.80
2PRH	2.40	2PRM	3.00	12	Dihydroorotate dehydrogenase	4.84	4.71
1AEC	1.86	2ACT	1.70	14	Actinidin	4.98	5.09
1ICN	1.74	1IFB	1.96	16	Intestinal fatty acid binding protein	4.84	5.26
1AAQ	2.50	1G6L	1.90	19	HIV-1 protease	5.96	5.95

^aRoot mean square deviation (RMSD) of the ligand docking pose (ligand conformation and ligand orientation) w.r.t. the native ligand docking pose

^bRMSD of the ligand docking pose with the lowest RMSD with respect to the native ligand docking pose from iMOLSDOCK

^cGAFF: results obtained with general AMBER force field as the intraligand energy function

^dMMFF: results obtained with MMFF94 force field as the intraligand energy function

automatically found in iMOLSDOCK using the Fpocket 2.0 algorithm [31, 32].)

The three-dimensional structure of the receptor protein is specified in PDB format. In iMOLSDOCK, side-chain receptor flexibility is allowed [21]. The flexible residues may be specified explicitly. In case, the flexible residues are not known, then residues that are within 4.0 Å from each atom of the ligand are automatically selected. In iMOLSDOCK, a maximum of 50 protein residues may be allowed to be flexible. For the test cases (Table 1), we allowed the side-chain torsion angles of the flexible residues to fluctuate in a range of 40° (i.e., −20° to +20° from their position in the crystal structure).

The scoring function in iMOLSDOCK is the weighted sum of intra-ligand energy, protein–ligand interaction energy and intra-protein energy. The intra-protein energy is calculated using the AMBER94 force field [25]. The intermolecular interaction energy is calculated using the PLP scoring function [33]. MMFF94 force field works efficiently for small molecule ligands [34]. General AMBER force field (GAFF) is a general force field where all the parameters are available and covers almost all the organic chemical spaces [35]. In iMOLSDOCK, the intra-ligand energy may be calculated either using MMFF94 force field [36, 37] or GAFF [35]. We have tried both MMFF94 and GAFF as intra-ligand energy separately for our test cases. As stated earlier [21], we use the PLP force field [33] for the protein–ligand interaction energy, and AMBER force field [25] for the intra-protein energy. The total energy is a weighted sum of these three terms. To fix the weights for each term, calculations were performed on a small subset of the structures. The optimum weights were chosen to yield maximum positive correlation between the energy and the root mean square deviation (RMSD) of the resulting docked structure with respect to the ‘native’ crystal structure. The AMBER force field [25], MMFF94 [36, 37] and GAFF [35] are expressed in units of kcal/mol, whereas the PLP force field is reported in dimensionless units [33]. Therefore the total potential energy is also reported in dimensionless units.

iMOLSDOCK is a command-line-only FORTRAN-based induced-fit protein–ligand docking tool. To make iMOLSDOCK easily accessible to the scientific community, we have developed a Java-based Graphical User Interface (GUI) and added it to MOLS 2.0 [32]. MOLS 2.0 is a software package developed in our laboratory. The software package is available free at <https://sourceforge.net/projects/mols2-0/files/>.

We have used a set of 34 protein–ligand complexes for which the crystal structures of both *apo* and *holo* forms were available, taken from the PDB, to test the performance of iMOLSDOCK (Table 1). The binding sites in these proteins are not similar to each other, and the proteins belong to different families of the SCOP database [38]. The number of

variable torsion angles in the ligands varies from 1 to 19. For all the test cases, the ligands were docked into the *apo* form of the receptor protein structure. Details of the *apo* and *holo* form of the receptor protein for all test cases are given in Table 1. We carried out induced-fit docking for all the cases, i.e., while docking in iMOLSDOCK, the conformation of the ligand, as well as conformation of the flexible residues lining the binding site in the receptor protein alter simultaneously.

Results and discussion

A total of 1500 structures were generated for all the 34 cases of the benchmarking dataset. Throughout the analysis, the crystal structure was considered as the native structure. In each case, out of the total 1500 structures, two structures were selected for our analysis. The first structure, called the ‘best sampled’ structure, is the prediction that has the lowest RMSD with respect to the native structure. The best sampled structure is found by superimposing all the 1500 predicted protein–ligand complex structures on the respective native protein–ligand complex structure by least squares superposition of all the atoms, without altering either the structure of the ligand, or its position and orientation relative to the protein. The RMSD was then calculated on all the heavy (non-hydrogen) atoms of the ligand. This particular method measures the differences in the docking pose of the ligand with respect to the protein along with the difference in the ligand structure. The second structure we used in our analysis is the prediction with the lowest energy of the total 1500 predictions. This is the energetically top-ranked structure.

Results obtained with GAFF and MMFF94 as intra-ligand energy function are summarized in Supplementary Table 1. The RMSD and energy of the top-ranked structure and the best sampled structure are given in this table. In the results with GAFF as the intra-ligand energy function (hereafter, GAFF-results), the RMSD of the best sampled structure is less than or equal to 2.50 Å in 16 of the 34 cases. In the results with MMFF94 force field as the intra-ligand energy function (hereafter, MMFF-results), the RMSD of the best sampled structures is less than or equal to 2.50 Å in 19 of the 34 cases. A brief summary of the overall results is given in Table 2. The best sampled structure of GAFF-results for all the cases is shown in Supplementary Fig. 1 and the best sampled structure of MMFF-results for all the cases is shown in Supplementary Fig. 2.

Ranking efficiency

An ideal scoring function in a docking tool is expected to top-score the best sampled (near-native) structure. A solution for which the best sampled structure is the same

Table 2 Comparison between GAFF and MMFF94 results in iMOLSDOCK for the 34 test cases

Description	GAFF ^a	MMFF ^b
RMSD of best sampled structures ≤ 1.00	3	1
RMSD of best sampled structures ≤ 2.00	10	9
RMSD of best sampled structures ≤ 2.50	16	19
RMSD of top-ranked structures ≤ 2.50	4	5
Exact solutions ^c	1	Nil
Best sampled structure within top 10% energy ranking	7	7

^aGAFF: results obtained with general AMBER force field as the intraligand energy function

^bMMFF: results obtained with MMFF94 force field as the intraligand energy function

^cSolutions for which the best sampled is the top-ranked structure

as the top-ranked structure is defined as an exact solution. In GAFF-results, exact solution was found for case 2YPI. In both GAFF-results and MMFF-results, for 7 of the 34 cases the best sampled structure is among the top 10% energy-ranked solutions. The top-ranked solution of 1NGP and 2PK4 have RMSD ≤ 2.00 Å from the native structure in GAFF-results. In MMFF-results, the top-ranked solutions of 1TNK, 2PK4 and 1AJQ have RMSD ≤ 2.00 Å from the native structure. It has been observed that in 8 of the 34 cases, in both GAFF-results and MMFF-results, at least one structure with RMSD ≤ 2.00 Å is among the top 10% ranked solutions.

Hydrogen bond interactions

The hydrogen bond interactions between the protein and the ligand were found using HBPLUS [39]. A comparison of these interactions in the native complex and in the top-ranked complex determined by iMOLSDOCK is shown in Fig. 1. Hydrogen bond prediction in GAFF-results is better than MMFF-results for 25 of the 34 cases. In 10 of the 34 cases, the top-rank structure predicted using GAFF has more hydrogen bond interactions than observed in the native structure. In six cases using MMFF94 resulted in more hydrogen bonds than the native structure. For the cases 4EST and 1BLH, no hydrogen bond interactions were predicted by MMFF94 in the top-ranked structure, while GAFF predicted two hydrogen bonds in the top-ranked structure of 4EST and four hydrogen bonds in the top-ranked structure of 1BLH respectively. In many cases the hydrogen bond interactions seen in the native structure are also predicted by iMOLSDOCK in the top-ranked solutions. In cases 1NGP and 2PRH, all the hydrogen bond interactions predicted by MMFF94 as the top-ranked solution are also seen in the native crystal structure (Fig. 2). The RMSD of the top-ranked solution of 1NGP from the native structure is 2.08 Å. The RMSD of the top-ranked structure of 2PRH from the native structure is 5.37 Å.

Alternate binding modes

Alternate solutions that have a lower energy value than the native structure are often detected by algorithms [40].

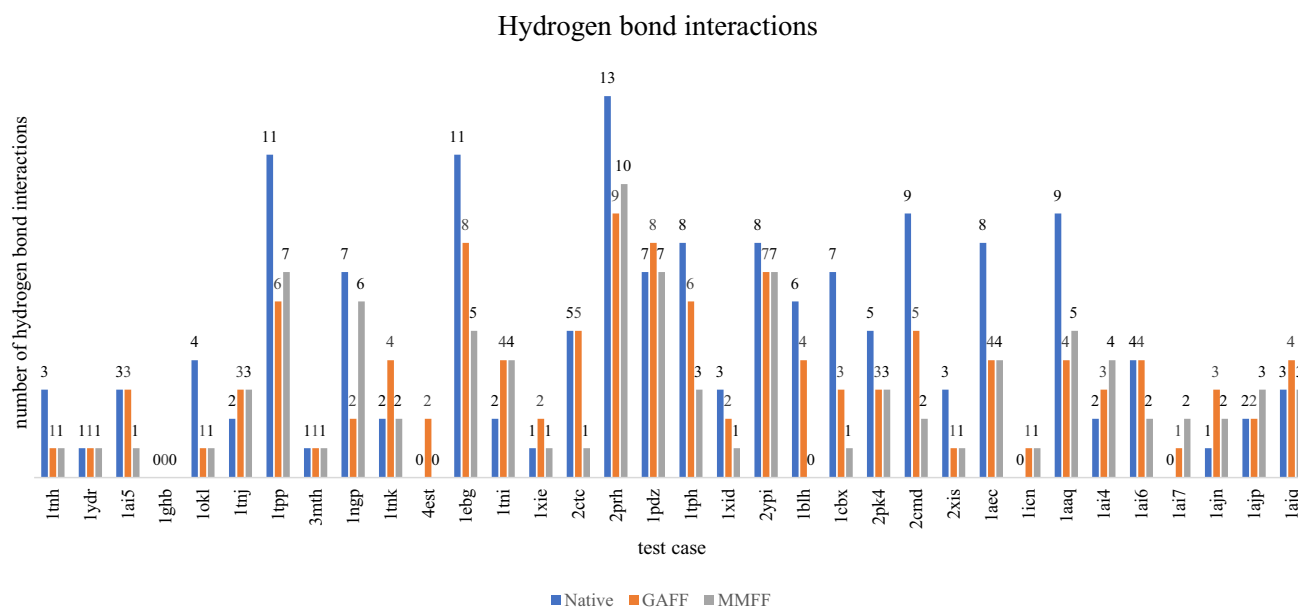
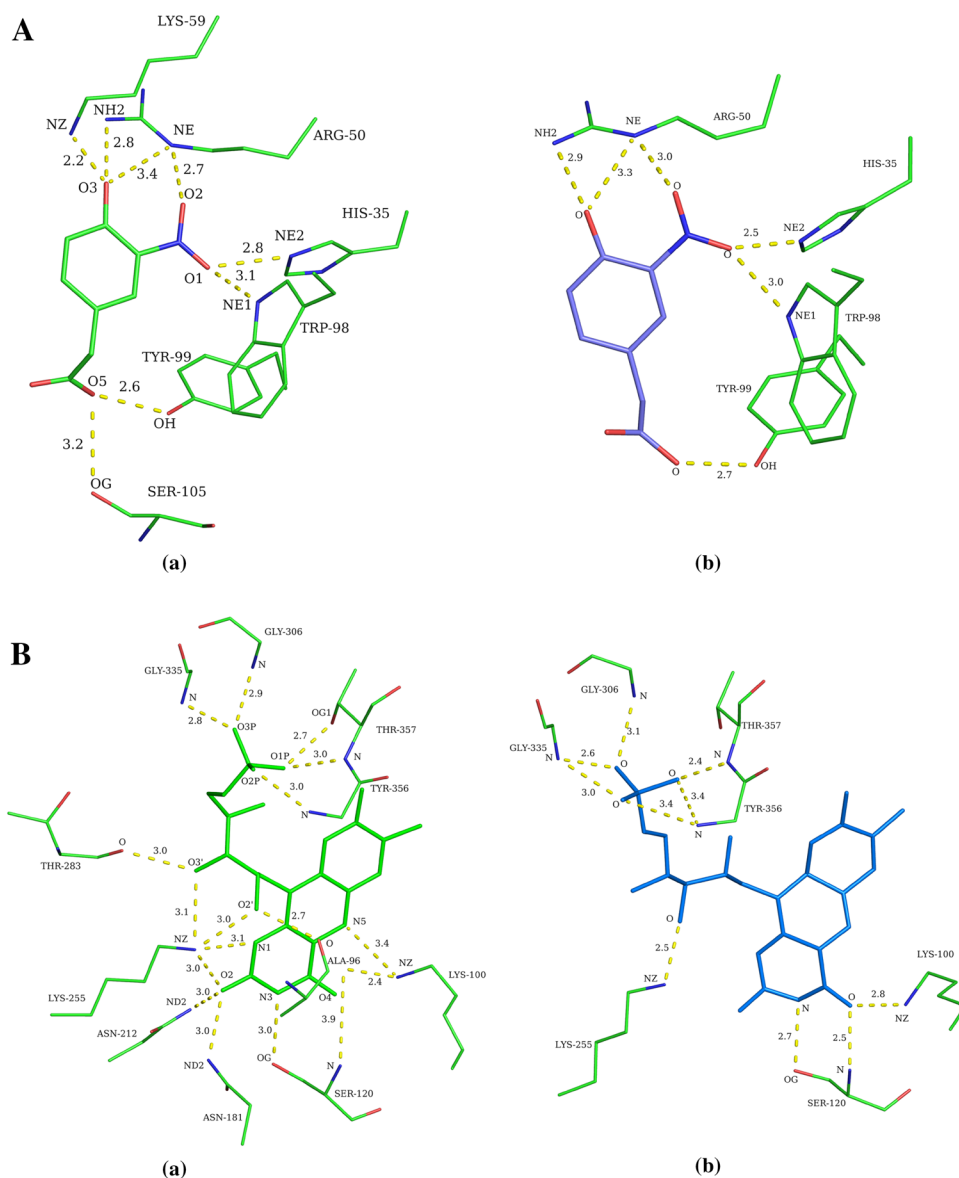


Fig. 1 Comparison of hydrogen bond interactions in the top-ranked iMOLSDOCK structure (with GAFF and MMFF94 force field as intraligand energy function) with the hydrogen bond interactions found in the native structure

Fig. 2 **A** Case: 1NGP—*a* hydrogen bond interactions in the native protein–ligand complex. *b* Conserved hydrogen bonds in the top-ranked structure of iMOLSDOCK with MMFF94 force field as the intra-ligand energy function. The RMSD of the top-ranked solution w.r.t. the native structure is 2.08 Å. **B** Case: 2PRH—*a* hydrogen bond interactions in the native protein–ligand complex. *b* Conserved hydrogen bonds in the top-ranked structure of iMOLSDOCK with MMFF94 force field as the intra-ligand energy function. The RMSD of the top-ranked structure w.r.t. the native structure is 5.37 Å



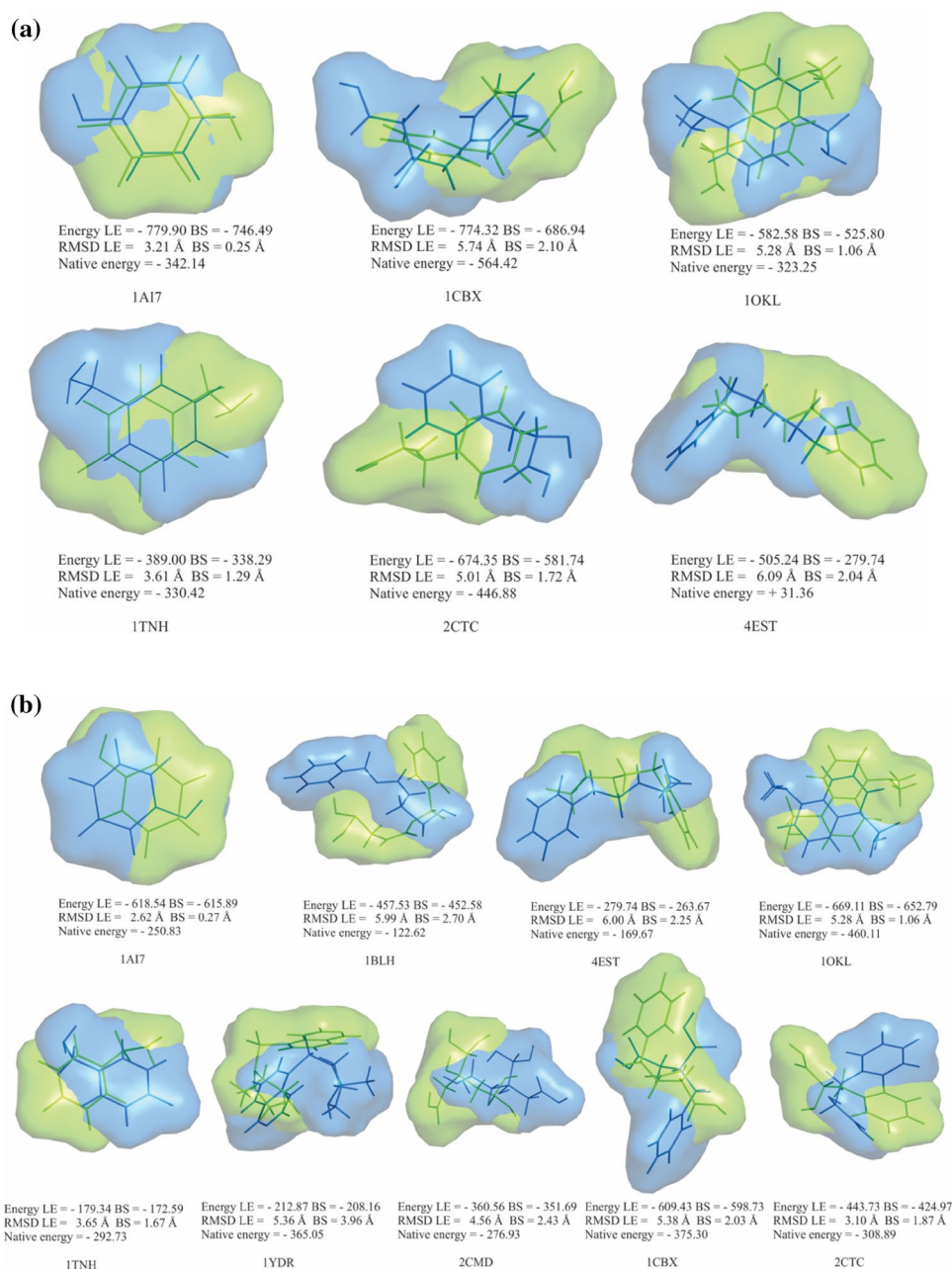
iMOLSDOCK does not converge to just a single solution but generates hundreds of low-energy possibilities. In certain cases, the native structure, the best sampled structure and the lowest energy structure (which is the top-ranked structure) are iso-energetic, i.e., they are energetically indistinguishable [41]. In GAFF-results, for 6 of the 34 test cases, the lowest energy structure shows an equally favourable alternate binding mode (Fig. 3a). In MMFF-results, for 9 of the 34 test cases, the lowest energy structure shows an equally favourable alternate binding mode (Fig. 3b). These alternate binding modes are characterized by energy values lower than native energy, but with larger values of RMSD as compared to the native complex. Alternate binding modes are important in the study of the possible toxicity of putative drugs [42].

Comparison studies between Vina, GOLD and iMOLSDOCK

We chose AutoDock Vina [13]—a free and open-source docking tool—and GOLD [11]—a commercial and widely used docking tool—for the comparison. The key aspects that are taken into consideration during the comparison are sampling efficiency, ranking efficiency, prediction of non-bonded interactions and computation time of iMOLSDOCK, Vina and GOLD. The iMOLSDOCK results obtained with GAFF [35] as the intra-ligand energy are considered for the comparison studies.

We carried out small molecule docking in Vina for all the 34 protein–ligand complexes (Table 1). The input structure of ligand and protein were prepared using

Fig. 3 a Alternate binding modes predicted by iMOLSDOCK top-ranked structures with GAFF as the intra-ligand energy function. The top-ranked structure of iMOLSDOCK is shown in blue and the native structure is shown in green. LE is the lowest energy structure. BS is the best sampled structure. Refer text for more information. **b** Alternate binding modes predicted by iMOLSDOCK top-ranked structures with MMFF94 force field as the intra-ligand energy function. The top-ranked structure of iMOLSDOCK is shown in blue and the native structure is shown in green. LE is the lowest energy structure. BS is the best sampled structure. Refer text for more information



AutoDock Tools (ADT; [43]). The input structure of the ligand was taken from the bound complex crystal available in the PDB. We used ADT to randomize the starting conformation, position and orientation of the ligand. The apo protein structure taken from the PDB (see Table 1) was used as the receptor protein structure for all the cases. The binding site was predefined for all the test cases. On an average, 10 residues lining the binding site in the receptor protein were selected to be flexible. We set the search space to 22.5 Å, the maximum dimension suggested by the authors [13], in all the three axes. The other parameters (exhaustiveness, number of modes, etc.) were fixed to the default values.

The crystal structure of the ligand was taken from the protein–ligand complex and was used as the input ligand structure. We defined the binding site for all the cases. To allow receptor flexibility, we allowed 10 flexible residues in the binding site to be flexible. GOLD allows partial protein flexibility in the neighbourhood of the protein active site. We chose the ‘Crystal’ option under ‘Rotamer Library Operations’ for protein side-chain flexibility. GoldScore was selected as the fitness function for all the cases. GoldScore is formulated in such a way that, the larger the score, the better the docking result is likely to be. GoldScore is dimensionless. Throughout the small molecule docking, the parameters of the fitness function were fixed to their

default values. A total of 150 GA runs were performed for each test case. All the other operations were set to ‘automatic’.

Sampling efficiency

Supplementary Table 2 shows a comparison of the results obtained from the three programs. Out of all the docked structures for each test case, we chose the best sampled structure and the top-ranked structure for our analyses. The best sampled structure is the structure with the lowest RMSD with respect to the native complex. The lowest energy structure is the top-ranked structure.

With Vina, the best sampled solutions of all the cases except 1TNH are 2.00 Å or more away from the crystal structure. With GOLD, in 11 of the 34 cases, the best sampled structures are within 2.00 Å from the native structure. With iMOLSDOCK, in 10 of the 34 cases, the best sampled structures are within 2.00 Å from the native structure.

With Vina, the top-ranked structure of 1TNH is within 2.00 Å from the crystal structure. With GOLD the top-ranked solution in four cases (1TPP, 1NGP, 1AEC, and 1AJQ) are within the RMSD of 2.00 Å from the native structure. With iMOLSDOCK the top-ranked solution in two cases (1NGP and 2PK4) are within the RMSD of 2.00 Å from the native structure.

Thus, both in terms of the best sampled solutions, as well as the top-ranked structures, iMOLSDOCK performs almost as well as, if not better than, the other two algorithms.

Scoring efficiency

With Vina, we generated the default number of binding modes, 9 in each of the 34 cases. Vina is able to identify exact solutions—the lowest RMSD structure is also the top-ranked structure—in four cases (2PRH, 2YPI, 2XIS, and 1AAQ). GOLD predicted exact solutions in seven cases (1NGP, 4EST, 1PDZ, 1BLH, 1AEC, 1AI4, and 1AJQ). Of these, three cases are within 2.00 Å from the native structure. iMOLSDOCK identified exact solution in one case (2YPI).

With Vina, except for the exact solutions, none of the cases have the best sampled structure among the top 10% energy-ranked solutions. With GOLD, in 27 out of the 34 cases the best sampled structure is among the top 10% of the ranked structures. With iMOLSDOCK, the best sampled structures in 8 of the 34 cases are among the top 10% of the energy-ranked solutions.

Clearly GOLD performs the best in this measure, but iMOLSDOCK results are comparable to those obtained with Vina.

Hydrogen bond prediction

The number of hydrogen bond interactions in the top-ranked structures of Vina and GOLD and iMOLSDOCK are shown in Fig. 4. iMOLSDOCK predicted more hydrogen bonds in the top-ranked structures than Vina or GOLD in 15 and 11, respectively, of the 34 cases. In five cases (1TNH, 1NGP, 1TPH, 1CBX, and 1AI7), the number of hydrogen bonds in the top-ranked structure are the same in both iMOLSDOCK and Vina. In seven cases (1TPP, 3MTH, 4EST, 1EBG, 2PRH, 2YPI, and 1BLH), the number of hydrogen bonds in the top-ranked structure are equal in both iMOLSDOCK and GOLD. Overall, GOLD predicts a larger number of hydrogen bonds for each structure than iMOLSDOCK.

Computation time comparison between iMOLSDOCK, Vina and GOLD

The computation time of Vina, GOLD, and iMOLSDOCK with respect to the number of torsion angles (ligand torsion angles + protein torsion angles) are shown in Fig. 5 (see also Supplementary Table 3). The average computation time for one run in Vina is 3.1 m. In GOLD it is 1.8 m. Both these values are much smaller than for iMOLSDOCK (1.3 h per structure). iMOLSDOCK however handles more variable parameters (torsion angles) than Vina or GOLD. In our test cases, we allowed 20 flexible residues in iMOLSDOCK, 10 flexible residues in Vina and in GOLD.

iMOLSDOCK performance against the Astex Diverse set of protein–ligand complex

Of the 85 cases of the Astex Diverse set [43], we selected 55 protein–ligand complexes that had no ligand–metal ion interaction. We tested the performance of iMOLSDOCK for 55 cases of the Astex Diverse set taken from the Cambridge Crystallographic Data Centre and then compared the performance of iMOLSDOCK with other docking tools: GOLD [11] and AutoDock Vina [13]. The overall comparison results are shown in Supplementary Table 4. For all the 55 cases, 1500 models were generated as discussed in the Methods section. In 27 of the 55 cases the best sampled structure has an RMSD less than 2.00 Å. In 10 of the 55 cases, the best sampled structure has an RMSD less than 1.00 Å. For one case (1s3v), iMOLSDOCK predicted the best sampled structure as the lowest energy structure, i.e., the top-ranked structure. However, for 30 of the 55 cases the best sampled solution is within the top 10% when ranked in terms of energy. The top-ranked structure is within 2.00 Å from the native structure in 13 of the 55 cases. Of the 55 cases, GOLD predicted exact solutions, i.e., the best sampled structure is the top-ranked structure, for 32 cases and Vina predicted exact solutions for 38 cases respectively. Of the

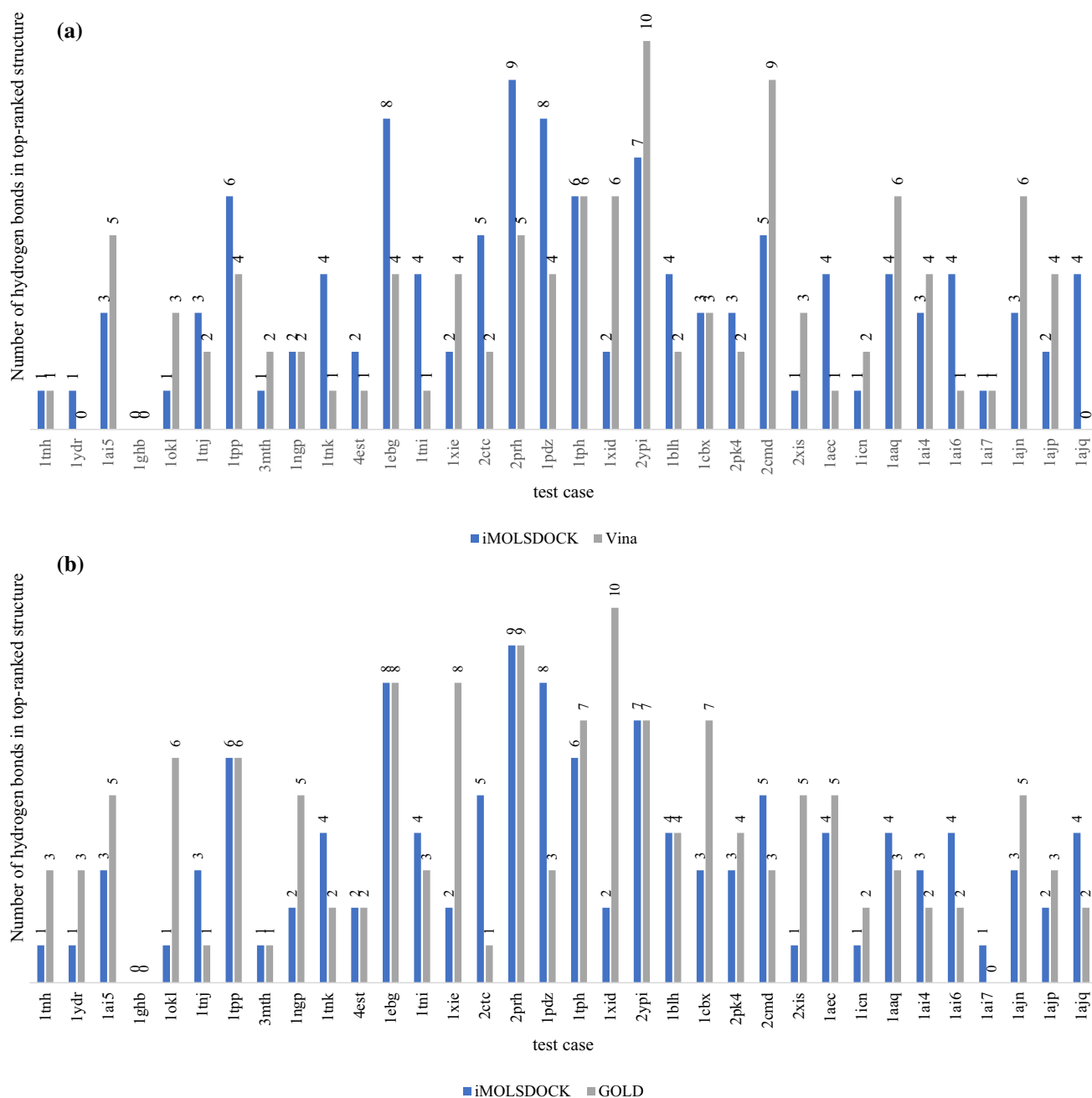


Fig. 4 **a** Number of hydrogen bond interactions (HB) found in the top-ranked structures of Vina and iMOLSDOCK. HB in the top-ranked iMOLSDOCK structure is shown in blue color. HB in the top-ranked structure of Vina is shown in gray colour. **b** Number of

hydrogen bond interactions (HB) found in the top-ranked structures of GOLD and iMOLSDOCK. HB in the top-ranked structure of iMOLSDOCK is shown in blue color. HB in the top-ranked structure of GOLD is shown in gray colour

55 cases, the best sampled structure predicted by GOLD and Vina is within 2.00 Å for 40 and 43 cases, respectively.

Cross-docking

Cross-docking experiment is carried out in situations where a crystal structure of the protein with a ligand bound in the binding site is known. During drug discovery, the binding of

different ligands to the same protein needs to be evaluated accounting for potential changes in the protein side-chain or backbone conformation [2]. Firstly, for our cross-docking experiment, we randomly selected the crystal structure of *Penicillin acylase enzyme* bound with 10 different ligands from the Astex non-native dataset [44]. The ligand was cross-docked to the non-native conformer of the same protein using iMOLSDOCK. The results are shown in Supplementary Table 5. For

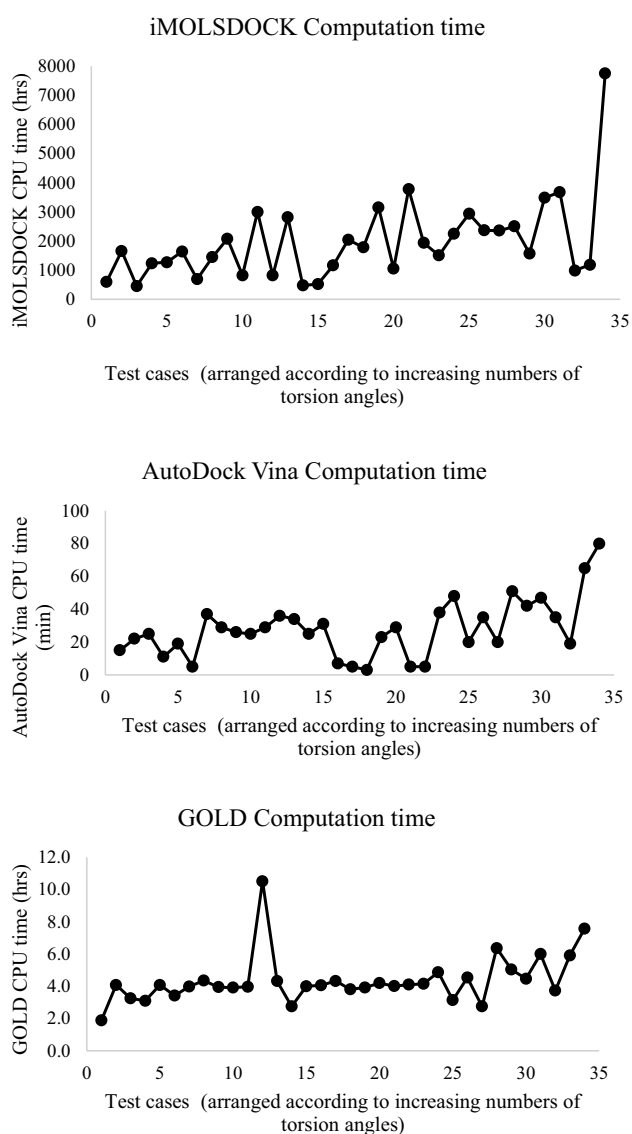


Fig. 5 Computation time comparison between iMOLSDOCK, Vina and GOLD

all the 10 cases, the best sampled structure of the ligand is within 2.01 Å from the native structure. Secondly, for our cross-docking experiment, we randomly selected 10 cases from the SEQ17 dataset [45] which contains apo and holo pairs for a diverse set of receptors [45]. This cross-docking experiment is carried to check whether iMOLSDOCK is able to perform the protein side-chain conformational change that is required for ligand binding as rigid docking is likely to fail for these cases. The results obtained were compared with the results obtained from GOLD docking tool. RMSD of the best sampled solution of iMOLSDOCK and GOLD are shown in Supplementary Table 6. Except for one case (1it8), iMOLSDOCK was not able to bind the ligand at least within 2.00 Å from the native structure. In case 1it8, the ligand could approach the active

site, as iMOLSDOCK modified the side-chain torsion angle of the active site residue Phe229 from 99.4° (apo form which is the starting structure) to 60.0° (native = 68.8°) thereby enabling the ligand to bind with RMSD = 1.43 Å from the native ligand. However, for case 1k4h, iMOLSDOCK was not able to alter the conformation of Tyr106, the active site residue blocking the entry of the ligand into the active site. Whereas GOLD, using the ‘rotamer’ option for receptor flexibility, was able to give a rotameric shift to Tyr106 enabling ligand binding with RMSD = 0.85 Å from the native structure. Side-chain torsion angle difference between the apo and the holo form of the active site residue Tyr106 is 109.2°. iMOLSDOCK is currently equipped to provide small side-chain fluctuations to the flexible residues in the receptor active site.

The aim of the current paper is to improve iMOLSDOCK by enabling small molecule-protein docking with induced-fit receptor flexibility. The docking failures observed has shown us areas where iMOLSDOCK and the scoring function may be improved. Especially, the addition of rotameric side-chain flexibility along with ring-flipping [11, 46] algorithm might enhance iMOLSDOCK especially for cross-docking experiments.

Conclusion

We have extended the ‘induced-fit’ docking feature in iMOLSDOCK, from docking peptide ligands to dock drug-like small molecule ligands. Test runs using 34 protein–ligand complexes from the PDB show that the method performs well. We have tested iMOLSDOCK with two different force fields (GAFF and MMFF94) for calculating the intra-ligand energy of the scoring function. Comparison of iMOLSDOCK with AutoDock Vina and GOLD shows that sampling in iMOLSDOCK is better than the other two tools. The search space in iMOLSDOCK is much greater than in Vina or GOLD as it allows for more flexible residues than Vina and GOLD. iMOLSDOCK is also able to predict alternate binding modes.

Acknowledgements We thank the Department of Science and Technology, Government of India, for financial support. We also thank the University Grants Commission for support under the CAS Program.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- Sousa SF, Ribeiro AJM, Coimbra JTS et al (2013) Protein–ligand docking in the new millennium—a retrospective of 10 years in the field. *Curr Med Chem* 20:2296–2314
- Meiler J, Baker D (2006) ROSETTALIGAND: protein–small molecule docking with full side-chain flexibility. *Proteins Struct Funct Bioinform* 65:538–548. <https://doi.org/10.1002/prot.21086>

3. DesJarlais RL, Sheridan RP, Seibel GL et al (1988) Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J Med Chem* 31:722–729
4. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA* 44:98–104
5. Monod J, Changeux J-P, Jacob F (1963) Allosteric proteins and cellular control systems. *J Mol Biol* 6:306–329. [https://doi.org/10.1016/S0022-2836\(63\)80091-1](https://doi.org/10.1016/S0022-2836(63)80091-1)
6. Monod J, Wyman J, Changeux J-P (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12:88–118. [https://doi.org/10.1016/S0022-2836\(65\)80285-6](https://doi.org/10.1016/S0022-2836(65)80285-6)
7. Najmanovich R, Kuttner J, Sobolev V, Edelman M (2000) Side-chain flexibility in proteins upon ligand binding. *Proteins Struct Funct Bioinform* 39:261–268. [https://doi.org/10.1002/\(SICI\)1097-0134\(20000515\)39:3%3C261::AID-PROT90%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0134(20000515)39:3%3C261::AID-PROT90%3E3.0.CO;2-4)
8. Zavodszky MI (2005) Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis. *Protein Sci* 14:1104–1114. <https://doi.org/10.1110/ps.041153605>
9. Kuntz ID (1992) Structure-based strategies for drug design and discovery. *Science* 257:1078–1082
10. Morris GM, Goodsell DS, Halliday RS et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662
11. Jones G, Willett P, Glen RC et al (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748. <https://doi.org/10.1006/jmbi.1996.0897>
12. Sherman W, Day T, Jacobson MP et al (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 49:534–553
13. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. <https://doi.org/10.1002/jcc.21334>
14. Lemmon G, Meiler J (2012) Rosetta Ligand docking with flexible XML protocols. *Methods Mol Biol Clifton NJ* 819:143–155. https://doi.org/10.1007/978-1-61779-465-0_10
15. Arun Prasad P, Gautham N (2008) A new peptide docking strategy using a mean field technique with mutually orthogonal Latin square sampling. *J Comput Aided Mol Des* 22:815–829
16. Viji SN, Prasad PA, Gautham N (2009) Protein–ligand docking using mutually orthogonal Latin squares (MOLSDOCK). *J Chem Inf Model* 49:2687–2694
17. Vengadesan K, Gautham N (2003) Enhanced sampling of the molecular potential energy surface using mutually orthogonal Latin squares: application to peptide structures. *Biophys J* 84:2897
18. Koehl P, Delarue M (1996) Mean-field minimization methods for biological macromolecules. *Curr Opin Struct Biol* 6:222–226. [https://doi.org/10.1016/S0959-440X\(96\)80078-9](https://doi.org/10.1016/S0959-440X(96)80078-9)
19. Kanagasabai V, Arunachalam J, Arun Prasad P, Gautham N (2007) Exploring the conformational space of protein loops using a mean field technique with MOLS sampling. *Proteins Struct Funct Bioinform* 67:908–921
20. Viji SN, Balaji N, Gautham N (2012) Molecular docking studies of protein–nucleotide complexes using MOLSDOCK (mutually orthogonal Latin squares DOCK). *J Mol Model* 18:1–18
21. Paul DS, Gautham N (2017) iMOLSDOCK: induced-fit docking using mutually orthogonal Latin squares (MOLS). *J Mol Graph Model* 74:89–99. <https://doi.org/10.1016/j.jmkgm.2017.03.008>
22. Hartshorn MJ, Verdonk ML, Chessari G et al (2007) Diverse, high-quality test set for the validation of protein–ligand docking performance. *J Med Chem* 50:726–741. <https://doi.org/10.1021/jm061277y>
23. Vengadesan K (2004) Sampling the molecular potential energy surface using mutually orthogonal Latin squares and application to peptide structures. Dissertation, University of Madras
24. Vengadesan K, Gautham N (2005) A new conformational search technique and its applications. *Curr Sci* 88:1759–1770
25. Cornell WD, Cieplak P, Bayly CI et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
26. Ito K (1987) Encyclopedic dictionary of mathematics. MIT Press, Cambridge
27. Liu CL (1968) Introduction to combinatorial mathematics. McGraw-Hill Book Co., New York
28. Vengadesan K, Gautham N (2004) Energy landscape of Met-enkephalin and Leu-enkephalin drawn using mutually orthogonal Latin squares sampling. *J Phys Chem B* 108:11196–11205
29. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1993) Numerical recipes in FORTRAN; the art of scientific computing, 2nd edn. Cambridge University Press, New York
30. Ben-Shimon A, Niv MY (2015) AnchorDock: blind and flexible anchor-driven peptide docking. *Structure* 23:929–940. <https://doi.org/10.1016/j.str.2015.03.010>
31. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. *BMC Bioinform* 10:168
32. Paul DS, Gautham N (2016) MOLS 2.0: software package for peptide modeling and protein–ligand docking. *J Mol Model* 22:1–9. <https://doi.org/10.1007/s00894-016-3106-x>
33. Gehlhaar DK, Verkhivker GM, Rejto PA et al (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 2:317–324
34. Gundertofte K, Liljefors T, Norrby P, Pettersson I (1996) A comparison of conformational energies calculated by several molecular mechanics methods. *J Comput Chem* 17:429–449. [https://doi.org/10.1002/\(SICI\)1096-987X\(199603\)17:4%3C429::AID-JCC5%3E3.0.CO;2-W](https://doi.org/10.1002/(SICI)1096-987X(199603)17:4%3C429::AID-JCC5%3E3.0.CO;2-W)
35. Wang J, Wolf RM, Caldwell JW et al (2004) Development and testing of a general AMBER force field. *J Comput Chem* 25:1157–1174. <https://doi.org/10.1002/jcc.20035>
36. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 17:490–519
37. Halgren TA (1996) Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J Comput Chem* 17:520–552. [https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5%3C520::AID-JCC2%3E3.0.CO;2-W](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5%3C520::AID-JCC2%3E3.0.CO;2-W)
38. Hubbard TJ, Murzin AG, Brenner SE, Chothia C (1997) SCOP: a structural classification of proteins database. *Nucleic Acids Res* 25:236–239
39. McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238:777–793. <https://doi.org/10.1006/jmbi.1994.1334>
40. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein–small molecule docking methods. *J Comput Aided Mol Des* 16:151–166
41. Taylor RD, Jewsbury PJ, Essex JW (2003) FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J Comput Chem* 24:1637–1656
42. Guengerich FP (2011) Mechanisms of drug toxicity and relevance to pharmaceutical development. *Drug Metab Pharmacokin* 26:3–14
43. Morris GM, Huey R, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791. <https://doi.org/10.1002/jcc.21256>

44. Verdonk ML, Mortenson PN, Hall RJ et al (2008) Protein–ligand docking against non-native protein conformers. *J Chem Inf Model* 48:2214–2225. <https://doi.org/10.1021/ci8002254>
45. Ravindranath PA, Forli S, Goodsell DS et al (2015) AutoDockFR: advances in protein–ligand docking with explicitly specified binding site flexibility. *PLoS Comput Biol* 11:e1004586. <https://doi.org/10.1371/journal.pcbi.1004586>
46. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Improved protein–ligand docking using GOLD. *Proteins Struct Funct Bioinform* 52:609–623