Protein−Ligand Docking Using Mutually Orthogonal Latin Squares (MOLSDOCK)

S. Nehru Viji, P. Arun Prasad, and N. Gautham*

CAS in Crystallography and Biophysics, University of Madras, Chennai-600025, India

Received June 19, 2009

The theoretical prediction of the association of a flexible ligand with a protein receptor requires efficient sampling of the conformational space of the ligand. Several docking methodologies are currently available. We have proposed a docking technique that performs well at low computational cost. The method uses mutually orthogonal Latin squares to efficiently sample the docking space. A variant of the mean field technique is used to analyze this sample to arrive at the optimum. The method has been previously applied to search through both the conformational space of a peptide as well its docking space. Here we extend this method to simultaneously identify both the low energy conformation as well as a high scoring docking mode for the small organic ligand molecules. Application of the method to 45 protein−ligand complexes, in which the number of rotatable torsions varies from 2 to 19, and comparisons with AutoDock 4.0, showed that the method works well.

INTRODUCTION

Computer-aided analyses of protein−ligand interactions have become important in view of the difficulties in experimentally characterizing them.[1] Automated prediction of molecular interactions is now an important step in rational drug design.[2,3] This is otherwise called molecular docking and may be described as a method of obtaining the low-energy binding modes of a ligand within the active site of a receptor, usually a protein or other biological macromolecule, whose structure is known. Molecular docking refers also to protein−protein and protein−DNA interactions. In the present work, we discuss only the docking of small organic molecules to protein receptors, a situation that arises most commonly in drug design.

Solving the docking problem computationally requires an accurate representation of the intermolecular interactions as well as an efficient algorithm to search for potential binding modes. In the most general form of the problem, both the receptor protein and the small molecule ligand are free to alter their initial conformations to arrive at suitable structures that are optimum for the interaction. This is known as "flexible receptor, flexible ligand" docking[4,5] and requires optimizing not only the intermolecular interactions, but also, simultaneously, the conformational energies of the receptor protein and the ligand. A more restricted version of the docking problem holds the receptor rigid and allows only the ligand to be flexible.[6,7] This requires the simultaneous evaluation of only the intermolecular energy of interaction and the conformational energy of the ligand. The optimization algorithm searches through both the conformational space of the ligand and its "docking space", i.e., the orientation and position of the ligand in the receptor site.

Several functions have been reported in the literature to calculate the energies.[8−10] Several algorithms have also been devised to perform the optimization.[8−10] Among the most widely used are AutoDock,[11] FlexX,[12] GOLD,[13] and DOCK.[14] AutoDock explores the conformational space of the ligand using the Lamarckian genetic algorithm (LGA).[11] The program uses a five-term force field-based function loosely based on the AMBER force field. FlexX[12] employs a deterministic incremental search algorithm to dock a flexible ligand to a rigid protein. The rationale behind the FlexX algorithm is to enumerate all possible interaction sites and then search this list to find matching points between the protein and the ligand. GOLD (Genetic Optimization for Ligand Docking)[13] employs a genetic algorithm to stochastically dock a ligand onto a protein. The algorithm allows full flexibility of the ligand but only partial flexibility of the protein. The fitness function is based partially on the analysis of known 3D-complexes. DOCK[14] finds potential conformations of a possible ligand using either exhaustive search or fragment-based docking. The protein−ligand complexes are scored with respect to steric fit or pharmacophore similarity. The ligand may be divided into small fragments. The first fragment is then placed optimally within the active site and the rest of the fragments appended.[15]

We have developed a method that used mutually orthogonal Latin square sampling to rapidly and exhaustively explore the conformational space of small molecules, in particular peptides.[16−18] We have extended this method to simultaneously search through both the conformational space of a peptide as well as its docking space.[19] An energy function that consisted of a weighted sum of ECEPP/3[20] for the conformational energy and PLP[21] scoring function for the energy of interaction between the protein and the peptide ligand was used, and the MOLS technique was applied to simultaneously optimize the conformation of the peptide and its pose in the receptor site. We showed that this docking technique had advantages as compared to the other techniques above, especially in terms of exhaustiveness of the search. In this work, we present a further extension of the MOLS based docking algorithm, which we call MOLSDOCK, to include other organic molecules as ligands. We have tested the technique on 45 protein−ligand

* For correspondence. (e-mail: n_gautham@hotmail.com).

10.1021/ci900332a
Published on Web 12/07/2009

complexes available in the PDB[22] and report the results here. We also present the results of a study we have carried out to compare the accuracy of the MOLS technique with that of the widely established docking package, AutoDock 4.0.[11]

METHODS

Representation of the Ligand. To generate the ligand, the Simplified Molecular Input Line Entry System (SMILES)[23,24] was used. The three-dimensional structure of the ligand was generated from the SMILES string,[25] and its rotatable torsion angles were identified.

The MOLS Method. A detailed and complete description of the MOLS algorithm as applied to conformational searches of small molecules is given at http://www.unom.ac.in/Gautham_mols.pdf, and elsewhere.[16−18] The MOLS technique treats the conformational search problem as one in experimental design. It utilizes mutually orthogonal Latin Squares (MOLS) to systematically sample the potential energy surface in torsion angle space, and analyses the results of the sampling by a procedure similar to the mean-field technique, to identify the optimal structure. The use of MOLS allows a drastic reduction in the size of the sampled conformational space, while still recovering much of the information content of the entire space. The algorithm consists of four steps: construction of the set of MOLS, calculation of the energy at each point of the sample, analyses of the sample, and identification of the optimal conformation. Each cycle of these four steps identifies one low-energy conformation. The steps may be repeated to identify other equally energetically favorable structures.

In an earlier paper,[19] we have described the extension of the MOLS method for molecular docking. For this application, the search space is not restricted to the conformational space of the ligand but has been expanded to include the "docking space". In torsion angle space, the three-dimensional structure of the ligand is specified by the M torsion angles θ_r, $r = 1, M$. If the binding site of the ligand on the receptor is known, then six additional parameters describe its pose in the site, three for the position and three for the orientation. This make a total of $M + 6$ dimensions in the search space (θ_r, $r = 1, M + 6$). The optimal structure of the ligand is defined by the set of θ_r which yields the minimum of $V(\theta_r)$ over the entire space, where V is a potential energy function that includes not only the conformational energy of the ligand, but also the interaction energy between ligand and receptor. If each of the dimensions is sampled at N intervals, the volume of the search space is $N^{(M+6)}$. The MOLS technique calculates the value of the scoring function at N^2 points in this space, and analyses them using a variant of the mean field technique, to simultaneously identify the optimum conformation of the ligand as well as its pose.

The binding site can in principle be predefined using experimental data, e.g., site-directed mutagenesis, or through computational predictions of the binding sites.[26] In the present study, since we have applied the method only to experimentally determined structures of protein−ligand complexes, the binding site and the search space is defined by a cubic box of 5 Å units centered on the centroid of the ligand in the crystal structure of the complex. The rotational and translational parameters inside the box and the ligand torsion angles are the variable parameters (i.e., the dimensions) in the search space. (It may be noted that the geometry and pose of the bound ligand in the native complex are not taken into account during the MOLS docking calculations. An extended conformation of the ligand is first generated from the SMILES string. The MOLS algorithm requires the calculations to start from several (37^2 in the present case) conformations simultaneously. As explained above, these 37^2 conformations are chosen by the MOLS method to effectively sample the entire search space.) The range for each torsion angle is set to 0−360°, with a search step size of 10°. The three translation parameters along the x, y, and z axes each have a range of 2.5 Å on either side of the center, and the centroid of the ligand is moved in steps of 0.14 Å. Only the centroid of the ligand is displaced within the defined cubic box during the docking simulation. In the extreme case, the ligand centroid might be placed exactly on the grid boundaries with half of its length stretching outside the grid box. Hence, to calculate the energy of interaction between the ligand and the protein, all atoms of the receptor at a distance less than the sum of half the length of the initial conformation of the ligand and the maximum interaction range of the selected PLP function (6 Å) were considered to be included in the search space.

Three parameters were used to represent the orientation of the ligand inside the cubic box. Two of these represented the position of a rotation axis, and the remaining one was the angle of rotation about this axis. To define the position of the axis, an imaginary unit sphere was constructed around the center of the cubic box. The axis was formed by drawing a line from a point on the surface of the sphere to its center. The two spherical coordinates specifying the point on the surface were then used to represent the position of the axis of rotation. The range of the polar angle was from 0 to π. The range of the azimuthal angle was 0 to 2π. The range of the angle of rotation about the axis was 0 to 2π. All three angles specifying the orientation of the ligand were sampled in steps of 10°.

Since the search space is defined on a discrete grid of $M + 6$ dimensions, in each cycle of calculations the method identifies an optimum point on this grid. However, the actual optimum may lie close to but not actually on the grid. The final step in identifying the optimum was carried out by performing conjugate gradient minimization to find the nearest off-grid optimum.

Energy Function. In most molecular docking calculations, the energy function is composed of two terms, namely the intramolecular ligand energy and the intermolecular interaction energy between the ligand and the receptor. In this application, since the ligand molecules are the small organic compounds, the MMFF94 force field[27,28] is used to calculate the intramolecular ligand energy as it works more efficiently for these type of molecules.[29] This force field expresses the total energy as a summation of two types of interaction terms: bonded and nonbonded. Bonded interactions typically include bond stretching, angle bending and torsion energy terms whereas nonbonded interactions include van der Waals and electrostatic energy terms. In the present calculations, since the search is conducted in torsion angle space, the bond stretching and angle bending energies remain constant and are not included. The intermolecular interaction energy is

Table 1. The 45 Protein−Ligand Complexes Used As Test Cases in This Study, Grouped by the Number of Rotatable Torsion Angles of the Ligand

PDB ID	Ligand	No. of rotatable torsion angles	Protein(Molecular Name)	Resolution (Å)
1F3D		2	Catalytic antibody 4B2	1.87
1TNH		2	Serine protease inhibitors	1.80
1YDR		2	C-AMP-dependent protein kinase	2.20
1AI5		3	Penicillin amidohydrolase	2.36
1AQW		3	Glutathione s-transferase	1.80
1GHB		3	Gamma-chymotrypsin	2.00
1OKL		3	Carbonic anhydrase II	2.10
1TNJ		3	Trypsin	1.80
1TPP		3	Beta-Trypsin	1.40
3MTH		3	Methylparaben insulin	1.90
1A6W		4	B1-8	1.80
1NGP		4	N1G9	2.40
1TNK		4	Trypsin	1.80
4EST		4	Elastase	1.78
1CTR		5	Calmodulin	2.45
1EBG		5	Enolase	2.10
1TNI		5	Trypsin	1.90
1XIE		5	D-Xylose isomerase	1.70
2CTC		5	Carboxypeptidase A	1.40
4TLN		5	Thermolysin	2.30
1APT		6	Penicillopepsin	1.80

calculated using the PLP scoring function.[21] This scoring function was selected based on its evaluation as one of the best of eleven scoring functions used in molecular docking.[30] The total potential energy is an unweighted sum of these two energy terms since they are of approximately the same magnitude. The MMFF94 energy term is expressed in units of kcal/mol and the PLP energy term is expressed as a dimensionless quantity. The total energy is also expressed as a dimensionless quantity.

To test the performance of MOLSDOCK, we used a set of 45 complexes (Table 1) which were selected from PDB. These proteins belong to 30 different families of the SCOP database[31] and the binding sites are not similar. Only structures of protein−ligand complexes with resolution better than 2.5 Å were considered. The number of variable torsion angles in the ligands varies from 2 to 19. The receptor protein molecule was held fixed in all the calculations. Thus the present docking protocol falls in the category "rigid receptor flexible ligand". Atoms in the receptor site having multiple occupancies were dealt with by selecting the one that had the highest occupancy. Several reports have emphasized the importance of water molecules in the receptor site.[4,32] Therefore, all water molecules in the receptor site that exhibited high occupancy and low temperature factor were retained, and considered part of the rigid receptor. The average CPU time required to generate 1500 structures was 2.48 h, for the test cases selected. It may be noted that the gradient minimization procedure at the final step of each cycle of the MOLS algorithm consumes the largest part (70−80%) of computational time.

Table 1. Continued

PDB ID	Ligand	No. of rotatable torsion angles	Protein(Molecular Name)	Resolution (Å)
1COM		6	Chorismate mutase	2.20
1D3H		6	Dihydroorotate dehydrogenase	1.80
1PDZ		6	Enolase	2.20
1TPH		6	Triosephosphate isomerase	1.80
1XID		6	D-Xylose isomerase	1.70
2YPI		6	Triose phosphate isomerase	2.50
3ERD		6	Triose phosphate isomerase	2.03
4LBD		6	Retinoic acid receptor gamma	2.40
5ABP		6	L-Arabinose-binding protein complex with D-Galactose	1.80
1BLH		7	Beta-Lactamase	2.30
1CBX		7	Carboxypeptidase A	2.00
1IMB		7	Inositol monophosphatase	2.20
2PK4		7	Human plasminogen kringle 4	2.25
2YHX		7	Hexokinase B	2.10
1HDC		8	3-Alpha, 20 beta-hydroxysteroid dehydrogenase	2.20
2CMD		9	Malate dehydrogenase	1.87
2SIM		9	Sialidase	1.60
2XIS		9	Xylose isomerase	1.71
3GPB		9	Glycogen phosphorylase	2.30
1HFC		10	Fibroblast collagenase	1.50
1EAP		12	IGG2B-Kappa 17E8 FAB	2.40
1AEC		14	Actinidin	1.86
1ICN		16	Intestinal fatty acid binding protein	1.74
1AAQ		19	Hiv-1 protease	2.50

Table 2. Summary of Results for MOLSDOCK for All the 45 Test Cases[a]

PDB ID	rmsd (Å)		energy		CPU time in h	energy in the crystal structure
	BS	LE	BS	LE		
1F3D	**0.91**	**0.91**	**−82.89**	**−82.89**	1.58	23.10
1TNH	0.57	0.69	−32.52	−53.57	1.24	19.47
1YDR	1.60	3.18	41.24	15.60	2.03	105.75
1AI5	1.01	4.76	27.29	1.29	3.38	95.46
1AQW	1.68	5.65	52.82	32.48	1.42	91.26
1GHB	1.40	2.10	−2.80	−29.39	0.89	6.60
1OKL	1.75	5.27	23.06	−09.22	3.1	95.85
1TNJ	0.45	1.80	−23.68	−25.02	2.65	60.70
1TPP	1.67	2.59	−20.36	−40.53	2.99	75.48
3MTH	1.30	2.67	12.89	−13.03	1.4	41.55
1A6W	1.54	5.35	15.85	−08.28	1.98	102.42
1NGP	1.06	1.16	16.36	−17.38	3.13	66.73
1TNK	0.70	2.26	−18.37	−38.68	0.31	49.25
4EST	1.72	6.70	77.60	22.21	0.29	168.19
1CTR	2.10	8.99	123.59	105.32	3.76	289.23
1EBG	1.36	3.50	−154.20	−158.01	2.22	−55.10
1TNI	1.58	3.44	146.35	−33.54	2.47	62.23
1XIE	0.76	3.79	99.25	61.43	3.11	196.50
2CTC	1.08	2.09	25.63	00.65	1.99	99.91
4TLN	1.49	4.01	86.87	37.55	1.36	132.64
1APT	2.61	4.20	87.97	67.56	1.52	143.52
1COM	1.73	3.97	11.34	−19.97	3.49	81.30
1D3H	1.76	6.73	41.61	−04.27	2.29	270.14
1PDZ	1.78	1.93	−233.90	−237.79	1.1	−140.66
1TPH	1.21	1.96	−162.88	−214.43	1.73	−104.82
1XID	1.06	4.16	47.72	22.84	1.67	156.15
2YPI	1.38	2.46	−224.60	−249.97	1.42	−129.02
3ERD	0.84	6.75	9.16	−35.66	2.6	57.04
4LBD	1.25	1.42	67.09	31.10	5.81	188.97
5ABP	0.99	1.71	71.64	47.35	0.64	161.00
1BLH	1.44	2.51	−130.14	−162.61	2.09	21.33
1CBX	1.46	1.70	−58.03	−86.31	2.85	17.65
1IMB	1.25	4.35	20.48	−56.80	2.06	25.88
2PK4	1.09	4.38	−24.17	−61.91	2.4	17.90
2YHX	1.80	3.29	78.51	62.29	4.33	101.02
1HDC	2.95	7.83	115.32	62.29	3.95	197.64
2CMD	1.46	2.93	−70.62	−95.63	1.74	11.98
2SIM	1.77	4.28	122.17	13.32	2.73	129.84
2XIS	1.03	2.41	83.15	46.65	1.58	139.23
3GPB	1.73	2.90	−93.20	−153.63	2.72	−19.70
1HFC	1.84	6.57	66.33	−13.38	2.51	110.77
1EAP	2.36	4.99	−30.38	−103.72	3.92	12.36
1AEC	2.02	7.54	−123.14	−184.17	5.08	22.11
1ICN	1.81	10.30	−14.86	−63.98	3.88	96.71
1AAQ	**2.91**	**2.91**	**60.12**	**60.12**	6.29	81.64

[a] BS: Best Sampled structure. LE: Lowest Energy structure. (See text for details). The last column specifies the ligand conformational energy and the interaction energy as calculated for the crystal structure. The "exact solutions" are marked in bold. The sixth column shows the CPU time taken by MOLSDOCK (1500 cycles).

RESULTS AND DISCUSSION

In the following discussions we specifically identify two structures out of the 1,500 predictions for each complex. One is the best sampled structure, i.e., the prediction that has the lowest all-atom rmsd with respect to the native structure (abbreviated as BS in the Tables). The predicted structure and pose of the ligand and protein is superposed on the native structure to obtain the best overall rmsd. The rmsd is then calculated for the ligand structure alone without any further rotation or translation. The other structure that we identify in the discussions is the prediction that has the lowest total energy (abbreviated as LE in the Tables), as defined in the Methods section, of all the 1500 predictions for each test case.

Overall Results. The most important requirement of a docking calculation is its ability to distinguish the real binding conformation and pose of the ligand on the protein from nonspecific and/or energetically unfavorable ones. Ideally, the method should predict the crystal structure (or a structure with very low root-mean-square deviation from the crystal structure) as the one with optimum energy. In other words, the best sampled structure and the lowest energy structure should be the same. Here, this is the case in 2 of the 45 test structures (Table 2). In the other 43 cases, the method finds at least one solution that has a low energy, as well rmsd less than 2.95 Å with respect to the crystal structure. The best sampled structures, positioned and oriented as in the receptor site and superposed without rotation or translation on the native structure, for all the 45 test cases are shown in Figure 1. As mentioned above, the method found "exact" solutions to two of the test cases in which the best sampled structure is the same as the lowest

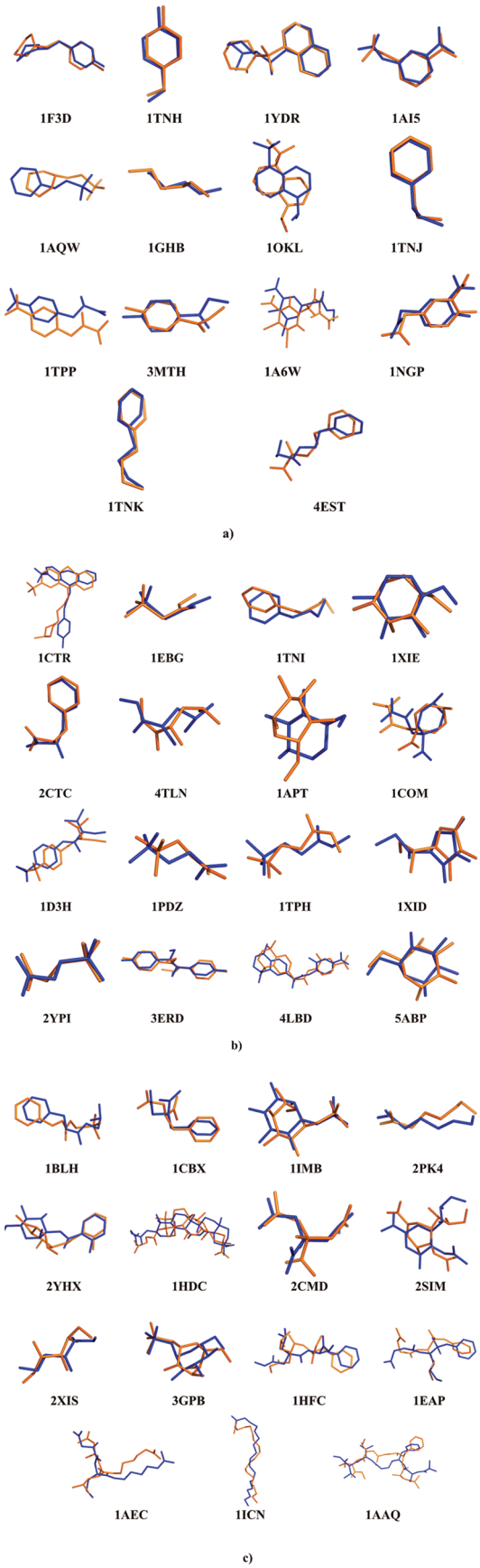

Figure 1. Best sampled ligand structures for all the 45 test cases are shown superposed without rotation and translation on the native ligand structure. The results are classified in terms of their rotatable torsions: (a) torsions 2−4, (b) torsions 5−6, and (c) torsion 7−19. The best sampled structures are in orange, and the native ligand structures are shown blue.

energy structure. These two structures are 1F3D and 1AAQ. The corresponding all atom rmsd with the respective native structures are 0.91 Å and 2.91 Å. According to the criterion of Chung and Subbiah,[33] for a prediction to be considered correct, its backbone rmsd as compared to the experimental structure should be within 2.00 Å. This is the case for the best sampled structure in 39 of the 45 cases. It is within 2.50 Å in 42 of the 45 cases.

In 17 of the 45 cases, the best sampled structure is in the top 10% when ranked in terms of the energy. The low energy docking solutions consist both of the conformations that belong to the native binding mode, as well as solutions that are quite different. For example, the lowest energy prediction for the test case 1ICN that has a large rmsd of 10.30 Å with the native. The energy of this prediction is −64.0. (This is lower than the energy of the native complex, which is 96.7 when calculated using the same formula.) One hydrogen bond was observed between the protein and the lowest energy ligand, but there were none in the native structure. A total of 19 nonbonded contacts were observed between the protein and the native ligand and only 11 between the protein and the lowest energy ligand (Hydrogen bonds and nonbonded contacts were analyzed using HBPLUS[34]). We conclude this section by reiterating that despite the large size and the extreme unevenness of the search space, the MOLS search algorithm identifies reasonable solutions at reasonable computation cost in all the test cases. Particular examples of these are discussed below. This is followed by a comparison of the performance of this method with that of AutoDock version 4.0.[11]

Alternate Binding Modes. Table 2 shows that there are many structures whose low energy structure has higher rmsd value with that of the native structure. Since the method does not converge to a single solution, but generates hundreds of low-energy possibilities, it often detects alternate solutions that have a lower energy value than the native structure. An analysis was carried out to check for the presence of alternate binding modes. It was found that out of the 45 cases, 11 cases exhibited such binding modes. (These are considered "alternate binding modes" since they fit into the binding site cavity in a manner similar to the native structure). An example of this is the structure of the glucocorticoid receptor interacting protein[35] in complex with the ligand DES (3ERD). Here, the lowest energy structure identified by the algorithm probably represents an alternate binding mode. The docked energies of the native, best sampled and lowest energy structures are 54.0, 9.2, and −35.7 respectively, i.e., they may be considered approximately iso-energetic.[36] The rmsd of the best sampled and the lowest energy structure as compared to the crystal structure are 0.84 and 6.76 Å, respectively. In both modes, the ligand makes the same type of hydrogen bonding patterns as seen in the native structure. However, the position of the oxygen atom of the ligand is exactly interchanged between the two alternate modes. Thus, the O1 atom of the native ligand forms a hydrogen bond with the residue His 524 of the protein and the O2 atom forms a hydrogen bond with Arg 394. This is reversed in the case of the other binding mode (Figure 2). In 10 other cases also (viz., 1AI5, 1AQW, 1A6W, 4EST, 1CTR, 1XIE, 1D3H, 1HDC, 2SIM, and 1HFC), the lowest energy structures are positioned in the

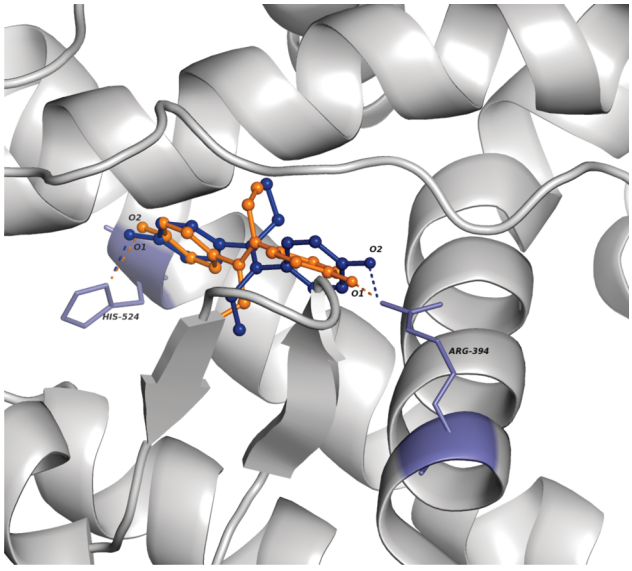

Figure 2. A perfectly symmetrical alternate binding mode achieved is shown by the superposition of the lowest energy structure of 3ERD on the respective low energy structure. The lowest energy ligand is shown in orange and the native structure in blue. Hydrogen bonds formed by the lowest energy ligand and native ligand are shown in orange and blue, respectively. The protein molecule is shown in light gray.

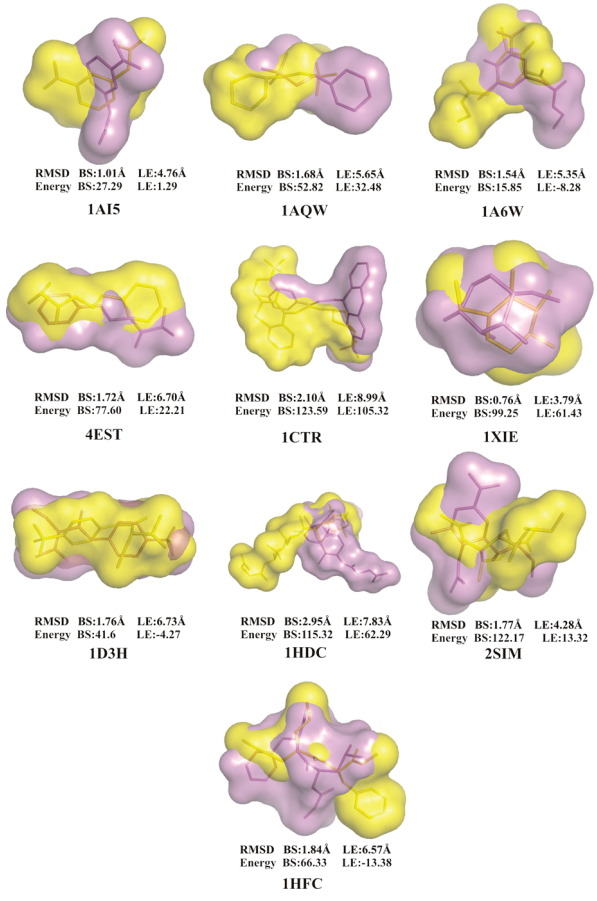

Figure 3. The alternate binding modes exhibited by other 10 structures are shown. The low energy structure of the ligand is shown in yellow and the best sampled structure is shown in violet.

receptor site in an alternate binding mode (Figure 3), with large rmsd as compared to the native structure (Table 2). In all of these structures, about 57% of the hydrogen bonds and about 69% of the nonbonded contacts were the same as in the native complex,

Comparison with AutoDock 4.0. The present MOLS algorithm was compared with AutoDock version 4.0.[11] For the 45 cases treated in the calculations described above, both ligand and protein input files were prepared using Auto Dock Tool (ADT) by standard protocols described in the literature.[37] Specifically, the rotatable torsion angles were selected explicitly, and were the same as those used in the MOLS method. Both the Grid Parameter File (GPF) and the Docking Parameter File (DPF) were prepared using ADT. The number of grid points in the grid box was set large enough to accommodate the extended conformation of the native ligand completely inside the grid box. The center of the grid box was set to the center of the ligand. All other grid parameter options were left at their default values. Docking was carried out using the default Genetic Algorithm (GA) parameters, along with Solis and Wets local search. A total of 150 GA runs were performed for each test case. The maximum number of energy evaluations was kept to be 2.5 million and the maximum number of generations was kept to be 27 000, respectively. All other docking run options were left at default values for further calculation.

Table 3 shows the results of the AutoDock runs for the 45 cases, and the energy ranking of best sampled structures for both AutoDock and MOLSDOCK methods. The best sampled structure had rmsd from the crystal structure of less than 2.50 Å in 30 of the 45 complexes in the AutoDock results. As seen from Table 2, this number is 42 in the case of the MOLSDOCK results. The three cases for which MOLSDOCK best sampled structures have rmsd greater than 2.5 Å, the AutoDock result shows a best rmsd value of 1.14 Å for one of these (1HDC). The other two have rmsd greater than 5.5 Å in AutoDock. In 9 cases, the structure best sampled by AutoDock was found within the top 10% when ranked in terms of energy. This was true in 17 cases when the MOLS method was used. Six of these were common to both algorithms. Two exact solutions, i.e., solutions in which the best sampled are the same as the lowest energy structure were found by both AutoDock and by MOLSDOCK, though these two cases were not the same. AutoDock was able to identify alternate binding modes, as described above, in five cases (1BLH, 1EBG, 1HDC, 2CTC, and 3MTH) while, as mentioned previously, MOLS identified eleven alternate binding modes In only one case, viz., 1HDC, the alternate mode identified by AutoDock is the same as the one identified by MOLS. In all the above, we could not identify any particular chemical or structural feature in the molecules which made the docking more amenable to either of the methods. These results show that the MOLS method samples a wider range of binding modes and ranks them better than AutoDock.

The AutoDock genetic algorithm is nondeterministic. Previous reports have recommended giving a minimum of 100 GA runs to identify the solution.[38] To compare the running times, for each test case, the maximum number of GA runs was fixed at 150, and the maximum number of MOLS runs was fixed at 1500. With these parameters, the average CPU time for each structure is 2.06 h for AutoDock (150 GA cycles) and 2.48 h for MOLS (1500

Table 3. Summary of Results for Autodock for All the 45 Test Cases[a]

	rmsd (Å)		energy				BS energy ranking (%)	
PDB ID	BS	LE	BS	LE	CPU time in h	energy in the crystal structure	AutoDock	MOLS
1F3D	9.22	19.15	−4.82	−6.60	1.78	12.33	82.67	0.07
1TNH	0.35	0.41	−6.25	−6.30	1.00	26.59	80.67	54.4
1YDR	0.23	0.28	−9.64	−9.66	1.68	−1.24	76.00	14.6
1AI5	1.30	1.43	−7.32	−7.55	1.50	−3.27	2.00	29.2
1AQW	2.21	2.76	−5.27	−5.45	1.32	46.88	34.00	17.27
1GHB	0.21	3.70	−3.21	−3.24	0.87	12.37	74.00	42.6
1OKL	1.05	1.37	−7.40	−8.13	1.62	3.86	96.67	0.13
1TNJ	0.58	2.26	−6.58	−6.99	1.17	−1.9	92.67	8.00
1TPP	1.62	2.12	−7.91	−8.03	1.88	−8.69	65.33	78.93
3MTH	1.04	13.81	−4.27	−5.40	1.12	32.9	79.33	66.07
1A6W	0.77	1.21	−8.92	−9.21	1.92	15.53	28.67	12.00
1NGP	1.20	3.69	−9.28	−10.29	1.67	−8.44	73.33	13.40
1TNK	0.68	1.60	−6.86	−7.12	1.30	−9.47	10.67	2.07
4EST	1.99	16.37	−5.20	−7.87	1.62	6.55	15.33	84.13
1CTR	1.28	4.96	−7.74	−8.41	2.70	−12.02	3.33	1.53
1EBG	1.97	13.99	−5.48	−6.10	1.37	−10.82	88.00	0.60
1TNI	0.91	3.33	−6.96	−7.79	1.50	−1.39	76.00	6.73
1XIE	0.47	8.27	−5.48	−7.12	1.25	2.88	62.67	33.47
2CTC	0.64	14.78	−6.41	−6.86	1.42	29.71	88.67	10.20
4TLN	1.83	4.47	−9.11	−9.77	1.33	−3.32	48.67	76.13
1APT	5.70	14.69	−4.77	−5.89	1.35	−0.84	51.33	63.00
1COM	1.69	3.34	−5.32	−6.86	1.28	−1.32	28.67	12.33
1D3H	9.25	12.38	−6.83	−7.19	1.90	5.43	94.67	26.73
1PDZ	22.31	26.23	−4.34	−5.24	1.02	2.32	42.00	0.20
1HAK	3.72	7.714	−9.32	−10.55	3.17	2.26	100.00	35.60
1XID	2.79	9.48	−6.53	−8.53	1.28	−1.9	27.33	10.87
2YPI	1.60	3.39	−4.35	−4.84	1.30	−0.99	6.67	29.67
3ERD	0.39	6.80	9.27	−10.71	2.13	−2.33	41.33	7.13
4LBD	14.30	18.70	−7.65	−10.44	3.43	−0.66	7.33	1.67
5ABP	11.49	16.16	−5.85	−6.77	1.48	3.51	41.33	11.87
1BLH	1.86	7.03	−5.99	−6.99	1.95	−4.44	98.00	11.73
1CBX	5.44	16.51	−4.54	−5.46	1.73	1.13	91.33	4.87
1IMB	7.16	7.23	−1.58	−2.38	1.47	0.32	96.00	39.67
2PK4	0.68	1.35	−7.66	−8.13	1.45	2.55	34.67	52.40
2YHX	2.07	4.43	−8.97	−9.35	2.62	−3.33	13.33	5.27
1HDC	1.14	13.69	10.74	−10.28	4.12	−11.45	8.67	7.53
2CMD	0.71	3.50	11.01	−9.54	1.68	45.65	2.00	6.33
2SIM	**4.99**	**4.99**	**−9.83**	**−9.83**	2.77	23.65	92.67	89.07
2XIS	1.68	4.66	−7.25	−7.83	1.62	−6.26	4.67	32.73
3GPB	1.01	1.46	−8.14	−9.33	2.10	−1.32	31.33	12.80
1HFC	**2.78**	**2.78**	**−11.00**	**−11.00**	3.20	−1.66	65.33	24.47
1EAP	3.23	4.62	−10.80	−10.97	3.08	35.63	57.33	30.07
1AEC	4.85	4.87	−7.79	−10.74	4.43	−1.34	26.00	17.00
1ICN	1.31	8.37	−8.56	−9.33	1.78	3.87	84.00	13.73
1AAQ	8.98	16.05	9.35	−11.88	7.87	12.28	6.00	0.07

[a] BS: Best Sampled structure. LE: Lowest Energy structure. (See text for details). The sixth column shows the CPU time taken by AutoDock (150 GA cycles). The seventh column specifies the ligand conformational energy and the interaction energy as calculated for the crystal structure. The exact solutions are marked in Bold.

predictions). However, in order to complete a single GA run, AutoDock samples 250 000 ligand conformations. To obtain a single low energy conformation, MOLS samples only $N^2 = 1369$ ligand conformations. Here, N is the order (or size) of the MOLS square. This is set to 37 in all of the test cases considered. (It should be noted that with the AutoDock genetic algorithm, as with all evolutionary algorithms,[11] it is a common practice to have a very high count for the number of energy evaluation ($\sim 1.5 \times 10^6$) thereby sampling more ligand conformations). Thus to complete one single cycle, AutoDock takes 44 s, whereas MOLS takes only 8 s.

CONCLUSIONS

The novel "rigid receptor flexible ligand" docking algorithm, named MOLSDOCK, was tested on 45 protein−ligand complexes, in which the number of rotatable torsion angles of the ligand ranges from 2 to 19. The test shows that the MOLS method identifies the correct structure and pose of the ligand in the receptor site well. The MOLS method is also capable of identifying alternate binding modes. In comparison with AutoDock, it was found that both AutoDock and MOLS could find similar solutions in similar computational times. In general, it is a suitable method when it is desirable to extensively explore both conformational space and docking space simultaneously, at reasonable computational cost. The method may be adapted for "flexible receptor flexible ligand" docking by including the conformation of the residues lining the receptor site. Further, the method could be improved by using various scoring functions thereby increasing the docking accuracy and also its ranking accuracy.

ACKNOWLEDGMENT

We thank the Department of Biotechnology, Government of India for financial support. We also thank the University Grants Commission, and the Department of Science and Technology, Government of India for support under the CAS program and the FIST program, respectively.

REFERENCES AND NOTES

(1) Shoichet, B. K.; Kuntz, I. D. Predicting the structure of protein complexes: a step in the right direction. *Chem. Biol.* **1996**, *3* (3), 151–156.
(2) Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **1996**, *6* (3), 402–406.
(3) Lybrand, T. P. Ligand-protein docking and rational drug design. *Curr. Opin. Struct. Biol.* **1995**, *5* (2), 224–228.
(4) Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J. Principles of flexible protein-protein docking. *Proteins* **2008**, *73* (2), 271–289.
(5) Andrusier, N.; Nussinov, R.; Wolfson, H. J. FireDock: Fast interaction refinement in molecular docking. *Proteins* **2007**, *69* (1), 139–159.
(6) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43* (25), 4759–4767.
(7) Kontyianni, M.; McClellan, L. M.; Slkol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47* (3), 558–565.
(8) Glick, M.; Rayan, A.; Goldblum, A. A stochastic algorithm for global optimization and for best populations: A test case of side chains in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (2), 703–708.
(9) Pappu, R. V.; Hart, R. K.; Ponder, J. W. Analysis and application of potential energy smoothing and search methods for global optimization. *J. Phys. Chem. B.* **1998**, *102* (48), 9725–9742.
(10) Nair, N.; Goodman, J. M. Genetic algorithms in conformational analysis. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (2), 317–320.
(11) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662.
(12) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261* (3), 470–489.
(13) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
(14) Makino, S.; Kuntz, I. D. Automated flexible ligand docking method and its application for database search. *J. Comput. Chem.* **1997**, *18* (14), 1812–1825.
(15) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15* (5), 411–428.
(16) Vengadesan, K.; Gautham, N. Enhanced sampling of the molecular potential energy surface using mutually orthogonal Latin squares: application to peptide structures. *Biophys. J.* **2003**, *84* (5), 2897–2906.
(17) Vengadesan, K.; Gautham, N. Conformational studies on Enkephalins using the MOLS technique. *Biopolymers.* **2004**, *74* (6), 476–494.
(18) Vengadesan, K.; Gautham, N. An application of experimental design using mutually orthogonal Latin squares in conformational studies of peptides. *Biochem. Biophys. Res. Commun.* **2004**, *316* (3), 731–737.
(19) Prasad, P. A.; Gautham, N. A new peptide docking strategy using a mean field technique with mutually orthogonal Latin square sampling. *J. Comput.-Aided Mol. Des.* **2008**, *22* (11), 815–829.
(20) Ne'methy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* **1992**, *96* (15), 6472–6484.
(21) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AC-1343 by HIV-l protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2* (5), 317–324.
(22) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids. Res.* **2000**, *28* (1), 235–242.
(23) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31–36.
(24) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (2), 97–101.
(25) Gilbert, G.; Guha, R. Smi23D, http://www.chembiogrid.org/cheminfo/smi23d, (accessed Oct 16, 2007).
(26) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct., Genet.* **2002**, *47* (4), 409–443.
(27) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 490–519.
(28) Halgren, T. A. Merck molecular force field. II. MMFF94van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17* (5−6), 520–552.
(29) Gundertofte, K.; Liljefors, T.; Norrby, P.; Pettersson, I. A comparison of conformational energies calculated by several molecular mechanics methods. *J. Comput. Chem.* **1996**, *17* (4), 429–449.
(30) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 Scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46* (12), 2287–2303.
(31) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **1995**, *247* (4), 536–540.
(32) Dean, P. M.; Poornima, C. S. Hydration in drug design. 1.Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 500–512.
(33) Chung, S. Y.; and Subbiah, S. How similar must a template protein be for homology modeling by side-chain packing methods? In *Pacific Symposium on Biocomputing*; Hunter, L., Klein, T. E., Eds.; World Scientific: Hawaii, USA. 1996, 126−141.
(34) McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238* (5), 777–793.
(35) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95* (7), 927–937.
(36) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. FDS: Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J. Comput. Chem.* **2003**, *24* (13), 1637–1656.
(37) Tame, J. R.; Dodson, E. J.; Murshudov, G.; Higgins, C. F.; Wilkinson, A. J. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* **1995**, *3* (12), 1395–1406.
(38) Rosenfeld, R. J.; Goodsell, D. S.; Musah, R. A.; Morris, G. M.; Goodin, D. B.; Olson, A. J. Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. *J. Comput.-Aided Mol. Des.* **2003**, *17* (8), 525–536.

CI900332A