J Mol Model (2012) 18:3705–3722
DOI 10.1007/s00894-012-1369-4

ORIGINAL PAPER

Molecular docking studies of protein-nucleotide complexes using MOLSDOCK (mutually orthogonal Latin squares DOCK)

Shankaran Nehru Viji · Nagarajan Balaji · Namasivayam Gautham

Received: 2 November 2011 / Accepted: 25 January 2012 / Published online: 1 March 2012

Abstract Understanding the principles of protein receptor recognition, interaction, and association with molecular substrates and inhibitors is of principal importance in the drug discovery process. MOLSDOCK is a molecular docking method that we have recently developed. It uses mutually orthogonal Latin square sampling (together with a variant of the mean field technique) to identify the optimal docking conformation and pose of a small molecule ligand in the appropriate receptor site. Here we report the application of this method to simultaneously identify both the low energy conformation and the one with the best pose in the case of 62 protein-bound nucleotide ligands. The experimental structures of all these complexes are known. We have compared our results with those obtained from two other well-known molecular docking software, viz. AutoDock 4.2.3 and GOLD 5.1. The results show that the MOLSDOCK method was able to sample a wide range of binding modes for these ligands and also scores them well.

Keywords Alternate binding modes · Molecular docking · MOLSDOCK · Protein - Nucleotide docking · Structure based drug design

Introduction

The field of *in silico* molecular docking has evolved during the last three decades driven by the needs of structural molecular biology and structure-based drug discovery [1, 2]. The goal of automated molecular docking software is to understand and predict molecular recognition, both structurally, by finding likely binding modes, and energetically, by predicting binding [3–5]. Molecular docking is usually performed between a small molecule and a target macromolecule; this is often referred to as ligand - protein docking. 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. A 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 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 docking energies [8–10]. Several algorithms have also been devised to perform the energy optimizations [8–10]. Among them the most widely used docking programs are AutoDock [11], FlexX [12], GOLD [13], GLIDE [14] and DOCK [15].

We have developed a method that uses mutually orthogonal Latin square sampling to rapidly and exhaustively explore the conformational space of small molecules and peptides [16–18]. The method is extended to simultaneously search through both the conformational space of the small organic molecules as well as its docking space [19, 20]. An energy function that consisted of an unweighted sum of the AMBER force field (FF94) [21] for the conformational energy and the PLP scoring function for the energy of interaction between the protein and the ligand [22–24] was used, and the MOLS technique was applied to simultaneously optimize the conformation of the ligand and its pose in the receptor site. We showed this docking technique had

S. N. Viji · N. Balaji · N. Gautham (✉)
C.A.S. in Crystallography and Biophysics, University of Madras, Maraimalai Campus (Guindy),
Chennai 600025, India
e-mail: n_gautham@hotmail.com

advantages as compared to the other techniques above, especially in terms of exhaustiveness of the search. In this paper we present a further extension of the MOLS based docking algorithm (which we call MOLSDOCK) to nucleotide ligands.

In the Protein Data Bank (PDB) [25] there are about 3293 protein - nucleotide ligand complexes present. These ligands are biologically and chemically distinct from other ligands because they not only sometimes act directly as drug molecules but are involved in various other metabolic processes. They act as mediators for many cellular processes, including signal transduction, protein transport, growth regulation, polypeptide chain elongation and molecular switches [26]. In addition, analogues of nucleotides or nucleosides are used in the treatment of viral diseases, especially those caused by retroviruses, such as HIV. Nucleotides may be modeled as three relatively rigid moieties (i.e., the base, the sugar and the phosphates) connected together. Cyclic nucleotides, such as cAMP and cGMP, have additional bonds between the base and the sugar that increase their rigidity. We have tested the MOLSDOCK technique on 62 protein-ligand complexes available in the PDB and report the results here.

Methods

The MOLS method

A detailed and complete description of the MOLS algorithm as applied to conformational searches and to the docking problem is given at http://www.unom.ac.in/Gautham_mols.pdf, and elsewhere [16–20]. 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 several times to identify other energetically favorable structures. Experience [16–20] indicates that generating about 1500 low energy conformations is sufficient to cover the entire conformational space of small peptides and other small molecules.

While applying this technique to molecular docking [19, 20], the search space is not restricted to the conformational space of the ligand but is expanded to include the 'docking space'. In torsion angle space, the three dimensional structure of the ligand is specified by the 'n' torsion angles θr, r= 1, n. 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 makes a total of n+6 dimensions in the search space (θr, r=1, n+6). The optimal structure of the ligand is defined by the set of θr which yields the minimum of V(θr) 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 'm' intervals, the volume of the search space is m (n+6). The MOLS technique calculates the value of the scoring function at about m^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 structure of the nucleotide ligand was generated from coordinate libraries taken from Insight II software [27]. Since we have applied the method only to experimentally determined structures of protein-nucleotide complexes, the binding site and the search space is defined by a cubic box of 5 Å units centered on the centroid of the nucleotide in the crystal structure of the complex. The rotational and translational parameters inside the box and the conformational parameters of the nucleotide are the variable parameters (i.e., the dimensions) in the search space. (It may be noted that the geometry and pose of the bound nucleotide in the native complex are not taken into account during the MOLS docking calculations. In the MOLS technique no particular 'initial' conformation or pose is required to be specified, since in each cycle of calculations, m^2 conformations and poses covering all the search space are generated.)

Energy function

In most molecular docking calculations, the energy function is composed of two terms, namely the intra-molecular ligand energy and the inter-molecular interaction energy between the ligand and the receptor. In the present application, since the ligand molecules are the nucleotides, the AMBER force field (FF94) [21] is used to calculate the intra-molecular ligand energy. This force field expresses the total energy as a summation of two types of interaction terms: bonded and non-bonded. Bonded interactions typically include bond stretching, angle bending and torsion energy terms whereas non-bonded 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 are not included. The inter-molecular interaction energy is calculated using the PLP scoring function [22–24] and the total potential energy of

Table 1 The 62 protein–nucleotide complexes used for the study are grouped by their bases

Ligand name	PDB	Protein(molecular name)	Resolution in Å
ADENOSINE MONOPHOSPHATE (AMP)	1AER	Exotoxin A	2.30
	1DEL	Deoxynucleoside monophosphate kinase	2.20
	1EFV	Electron transfer flavoprotein	2.10
	1FA9	Glycogen phosphorylase	2.40
ADENOSINE-5′-DIPHOSPHATE (ADP)	1 AM1	Heat shock protein 90	2.00
	1B4S	Nucleoside diphosphate kinase	2.50
	1UW1	Artificial nucleotide binding protein (ANBP)	1.94
	2DLN	D-Alanine–D-alanine ligase	2.30
ADENOSINE-5′-TRIPHOSPHATE (ATP)	1A82	Dethiobiotin synthetase	1.80
	1AQ2	Phosphoenolpyruvate carboxykinase	1.90
	2BUP	Protein (heat shock cognate kda70)	1.70
	2GNK	Protein (nitrogen regulatory protein)	2.00
ADENOSINE-3′,5′-CYCLIC-MONOPHOSPHATE (CMP)	1G6N	Catabolite gene activator protein	2.10
	3I54	Transcriptional regulator, crp/fnr family	2.20
	3KCC	Catabolite gene activator	1.66
	3 N10	Adenylate cyclase 2	1.60
CYTIDINE-5′-MONOPHOSPHATE (C5P)	1H7F	3-Deoxy-manno-octulosonate cytidylyl transferase	2.12
	1H7T	3-Deoxy-manno-octulosonate cytidylyl transferase	2.48
	1QF9	Uridylmonophosphate/cytidylmonophosphate kinase	1.70
	1 W77	2 C-methyl-d-erythritol 4-phosphate	2.00
CYTIDINE-5′-DIPHOSPHATE (CDP)	1EYR	CMP-n-acetylneuraminic acid synthetase	2.20
	1H7H	3-Deoxy-manno-octulosonate cytidylyltransferase	2.30
	1U3L	2-C-Methyl-d-erythritol 2,4-cyclodiphosphate	2.50
	2CMK	Protein (cytidine monophosphate kinase)	2.00
CYTIDINE-5′-TRIPHOSPHATE (CTP)	1H7G	3-Deoxy-manno-octulosonate cytidylyltransferase	2.13
	1I52	4-Diphosphocytidyl-2-c-methylerythritol synthase	1.50
	1RAA	Aspartate carbamoyltransferase catalytic chain	2.50
	1RAD	Aspartate carbamoyltransferase catalytic chain	2.50
GUANOSINE-5′-MONOPHOSPHATE (5GP)	1EX7	Guanylate kinase	1.90
	1G7C	Elongation factor 1-alpha	2.05
	1LVG	Guanylate kinase	2.10
	1ZNX	Guanylate kinase	2.35
GUANOSINE-5′-DIPHOSPHATE (GDP)	1A4R	G25K GTP-Binding protein	2.50
	1CG0	Protein (adenylosuccinate synthetase)	2.50
	1CG1	Protein (adenylosuccinate synthetase)	2.50
	1CIB	Adenylosuccinate synthetase	2.30
GUANOSINE-5′-TRIPHOSPHATE (GTP)	1C1Y	RAS-related protein rap-1a	1.90
	1CKN	MRNA capping enzyme	2.50
	1E96	RAS-related c3 botulinum toxin substrate 1	2.40
	1J2J	ADP-Ribosylation factor 1	1.60
CYCLIC GUANOSINE MONOPHOSPHATE (PCG)	1Q3E	Cyclic nucleotide-gated channel 2	1.90
	3CL1	MLL3241 protein	2.40
	3DYN	CGMP-specific 3′,5′-cyclic phosphodiesterase	2.10
	3DYQ	CGMP-specific 3′,5′-cyclic phosphodiesterase	2.50
THYMIDINE-5′-PHOSPHATE (TMP)	1CY1	DNA Topoisomerase I	2.30
	1GSI	Thymidylate kinase	1.60
	1G3U	Thymidylate kinase	1.95
	1E2F	Thymidylate kinase	1.6
THYMIDINE-5′-DIPHOSPHATE (TYD)	1CR4	DNA Primase/Helicase	2.50

Table 1 (continued)

Ligand name	PDB	Protein(molecular name)	Resolution in Å
	1E2G	Thymidylate kinase	1.7
	1EPZ	Dtdp-6-deoxy-d-xylo-4-hexulose 3,5-epimerase	1.75
THYMIDINE-5′-TRIPHOSPHATE (TTP)	1H79	Anaerobic ribonucleotide-triphosphate reductase	2.90
	1N5J	Thymidylate kinase	1.85
URIDINE-5′-MONOPHOSPHATE (U5P)	1FGX	Beta 1,4 galactosyltransferase	2.40
	1G8O	N-Acetyllactosaminide alpha-1, 3- galactosyltransferase	2.30
	1HXP	Hexose-1-phosphate uridylyltransferase	1.80
	1HXQ	Hexose-1-phosphate uridylyltransferase	1.86
URIDINE-5′-DIPHOSPHATE (UDP)	1C3J	Beta-glucosyltransferase	1.88
	1F7P	Pol polyprotein	2.30
	1F7R	Pol polyprotein	2.50
URIDINE 5′-TRIPHOSPHATE (UTP)	1R8C	TRNA nucleotidyltransferase	1.90
	2B56	RNA editing complex protein mp57	1.97

the system is the sum of these two terms. The AMBER energy term is expressed in units of kcal mol^{-1}, while the PLP energy term is expressed as a dimensionless quantity. The total energy is also expressed as a dimensionless quantity.

The method was applied to a set of 62 complexes (Table 1) which were selected from PDB. Only structures of protein–nucleotide complexes with resolution better than 3.0 Å were considered. The receptor protein molecule was held fixed in all the calculations. Thus the present docking protocol falls in the category ‘rigid receptor - flexible ligand docking’. There were atoms in the receptor site with multiple occupancies so the atoms with highest occupancies were selected. Several reports have emphasized the importance of water molecules in the receptor site [28, 29]. 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.

Comparison with AutoDock and GOLD

In order to evaluate the accuracy of the above results in comparison with other established docking programs, we performed the same calculations on two extensively validated packages viz. AutoDock 4.2.3 [11] and GOLD 5.1 [13]. Both programs were used in the ‘flexible ligand - rigid receptor’ docking mode. Both employ genetic algorithms to identify the best docking conformation and pose.

To run AutoDock 4.2.3, both ligand and protein input files for the 62 cases above were prepared using Auto Dock Tool (ADT) by standard protocols described in the literature [30]. Specifically, the rotatable torsion angles were selected explicitly, and were the same as those used in the MOLSDOCK 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 big 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 the Solis and Wets local search. A total of 150 GA runs were performed for each test case. The maximum number of energy evaluations was set at 2.5 million and the maximum number of generations was 27,000. All other docking run options were left at default values for further calculation.

GOLD also uses a genetic algorithm to create putative poses for a single ligand. The program could consider receptor side-chain flexibility and local backbone movement during docking but for our studies the receptor was held rigid. For each complex, the native ligand was removed first, and the binding pocket was defined by the native ligand pose. To aid comparison, the size of the binding site was consistent for GOLD, AutoDock and MOLSDOCK. All scoring function values were kept to their defaults. All GA parameters were set to ‘automatic’. A total of 1500 GA runs were performed for each test case. Since the genetic algorithms used in AutoDock and GOLD are non-deterministic, previous reports have recommended giving at least 100 GA runs to identify the solution [30] for AutoDock. Hence for comparison studies, we fixed the maximum number of GA runs to be 150 for AutoDock and 1500 for GOLD for each test case. As described earlier, the maximum number of MOLSDOCK runs was fixed to 1500 for each test case. The average CPU time required for each complex is 0.98 hours by MOLSDOCK (to generate 1500 low-energy conformations and poses), 2.83 hours by AutoDock (to generate 150 low-energy conformations and poses) and 0.46 hours by GOLD (to calculate 1500 conformations and poses).

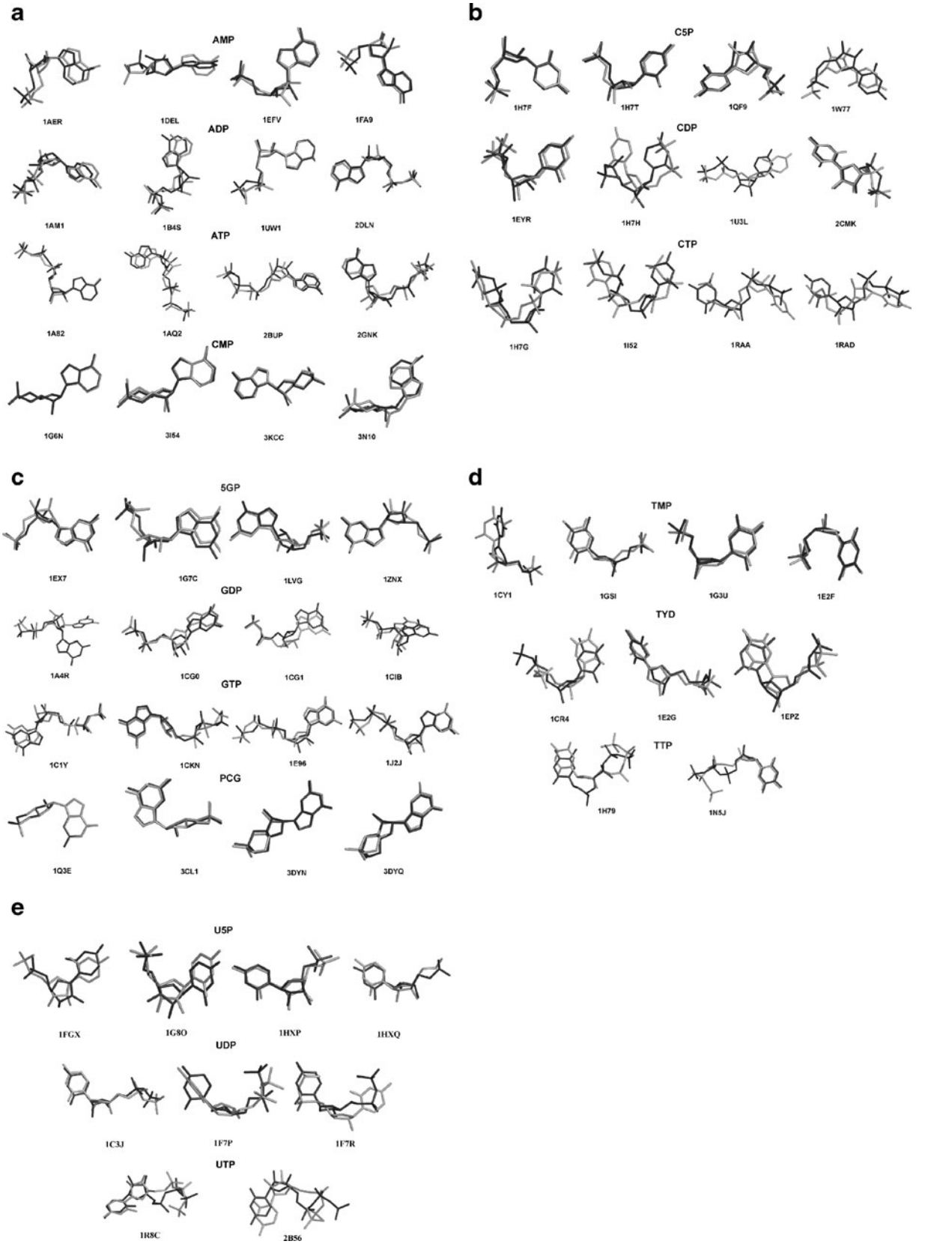

Fig. 1 'Best sampled' ligand structures for all 62 test cases superposed on the native ligand structure. The results are classified in terms of their bases: (**a**) Adenine, (**b**) Cytosine, (**c**) Guanine, (**d**) Thymine, and (**e**) Uracil. The 'best sampled' structures are in gray, and the native ligand structures are in black

Table 2 Summary of results for MOLSDOCK for all 62 test cases

S.NO	PDB ID	RMSD in Å		RDE %		Energy		BS rank %	Native energy	CPU time in hours
		BS	LE	BS	LE	BS	LE			
AMP										
1	1AER	0.92	2.41	67	58.53	6.99	3.55	0.27	−0.03	0.59
2	*1DEL*	*1.42*	*7.66*	*50.94*	*8.69*	*-20.21*	*59.75*	*37.8*	*-1.89*	*0.33*
3	1EFV	0.81	2.18	79.57	40.76	−88.36	80.65	81.47	−5.11	1.41
4	1FA9	1.22	2.08	69.64	29.91	1.92	59.81	2.93	4.56	0.6
ADP										
5	1 AM1	1.36	1.37	59.98	64.18	−22.28	−18.40	9.13	−8.11	1.4
6	1B4S	1.52	4.61	55.67	12.22	−14.92	31.10	0.2	−8.48	0.92
7	**1UW1**	**1.25**	**1.25**	**81.13**	**81.13**	**42.32**	**42.32**	**0.07**	**19.22**	**0.55**
8	2DLN	1.3	1.42	67.01	63.42	−22.60	16.02	1.53	12.37	1.9
ATP										
9	1A82	1.33	1.34	70.39	67.21	35.15	−8.38	0.6	18.36	1.25
10	1AQ2	1.4	1.94	51.61	50.18	43.58	−19.26	0.2	−5.33	1.97
11	2BUP	1.22	1.26	62.32	65.64	−13.29	76.24	1.93	−7.25	2.47
12	2GNK	2.15	2.9	29.66	29.61	−12.47	23.66	61.07	−12.38	0.56
CMP										
13	1G6N	0.08	0.24	99.62	96.1	−1.39	−1.38	19.93	−82.33	0.32
14	**3I54**	**0.33**	**0.8**	**94.02**	**72.2**	**−1.58**	**−1.23**	**0.87**	**−11.15**	**0.35**
15	**3KCC**	**0.3**	**0.31**	**94.36**	**93.85**	**−1.78**	**−1.46**	**0.27**	**−100.6**	**0.37**
16	3 N10	0.72	1.07	77.18	64.15	−3.34	−3.48	2.87	−25.62	0.39
C5P										
17	1H7F	0.56	1.47	88.41	78.01	17.04	56.88	3.93	1.88	0.68
18	1H7T	0.72	1.02	80.94	65.26	21.97	24.28	27.47	12.11	0.65
19	1QF9	1.01	1.36	72.38	57.04	14.70	23.69	4.73	1.32	0.74
20	1 W77	2.04	4.22	37.86	18.08	8.92	16.93	25.8	−5.33	0.49
CDP										
21	**1EYR**	**2.43**	**2.43**	**42.68**	**42.68**	**−10.61**	**−10.61**	**0.07**	**15.21**	**1.00**
22	1H7H	3.01	3.72	25.24	20.42	12.52	1.69	39.87	11.78	0.95
23	1U3L	2.74	4.57	25.06	12.2	−2.93	−10.00	65.27	−8.69	0.49
24	2CMK	1.9	2.88	47.39	50.91	12.90	113.26	2.4	−5.72	1.32
CTP										
25	*1H7G*	*6.46*	*7.81*	*31.47*	*5.5*	*-8.69*	*7.85*	*10.27*	*4.83*	*1.24*
26	1I52	2.46	6.5	29.95	5.95	−16.40	19.52	39.33	−9.38	1.32
27	*1RAA*	*5.39*	*9.17*	*27.73*	*5.21*	*-7.96*	*0.29*	*5.47*	*-1.21*	*0.75*
28	*1RAD*	*5.14*	*8.79*	*28.06*	*4.96*	*20.13*	*4.84*	*0.67*	*7.88*	*0.74*
5GP										
29	1EX7	1.11	1.27	69.59	68.75	3.73	6.02	2.6	−14.11	0.85
30	1G7C	1.38	1.62	58.47	53.83	3.98	4.74	24.27	22.89	0.72
31	1LVG	1.08	1.19	72.65	72.22	6.90	6.10	1.27	11.79	1.07
32	1ZNX	0.71	1.16	81.5	60.02	4.69	6.19	0.4	4.58	0.54
GDP										
33	*1A4R*	*4.74*	*6.52*	*25.03*	*25.65*	*-19.93*	*-12.36*	*23.6*	*14.45*	*0.71*
34	*1CG0*	*1.56*	*6.68*	*54.66*	*22.92*	*10.28*	*-10.74*	*3.33*	*-4.51*	*1.65*
35	*1CG1*	*1.37*	*5.16*	*55*	*24.87*	*-19.65*	*14.93*	*2.73*	*-19.32*	*1.57*
36	*1CIB*	*1.58*	*5.6*	*43.97*	*24.91*	*-15.64*	*-15.09*	*0.6*	*-7.25*	*1.74*
GTP										
37	1C1Y	1.47	3.02	58.98	28.48	−29.21	−16.89	0.47	−12.53	1.53

Table 2 (continued)

S.NO	PDB ID	RMSD in Å		RDE %		Energy		BS rank %	Native energy	CPU time in hours
		BS	LE	BS	LE	BS	LE			
38	*1CKN*	*0.94*	*4.14*	*77.87*	*50.55*	*-25.65*	*-25.12*	*9.87*	*-15.52*	*1.96*
39	1E96	1.18	1.26	67.82	62.46	−34.27	66.10	0.2	2.96	1.75
40	**1J2J**	**1.33**	**1.33**	**66.76**	**66.76**	**−24.43**	**−24.43**	**0.07**	**8.38**	**1.42**
PCG										
41	1Q3E	0.03	0.24	99.96	96.55	6.12	6.18	52.07	−35.85	0.28
42	3CL1	0.25	0.45	97.2	91.5	6.13	10.13	0.13	−4.24	0.27
43	3DYN	0.32	0.65	93.9	79.66	5.94	5.95	3.27	−3.33	0.56
44	3DYQ	0.48	7.31	89.09	6.41	6.04	5.98	15.8	11.76	0.52
TMP										
45	1CY1	1.85	2.66	42.15	38.21	−22.53	−3.05	79.13	1.89	1.53
46	1GSI	0.86	1.18	83.26	68.59	66.80	−3.99	4.8	−22.32	1.15
47	1G3U	0.56	1.26	83.24	65.57	−11.60	1.22	2	−8.59	1.00
48	1E2F	0.97	0.98	78.17	74.96	−4.50	4.62	0.47	−19.66	0.91
TYD										
49	1CR4	2.52	2.87	35.61	23.29	6.15	37.34	7.8	−0.22	0.76
50	1E2G	1.47	2.15	57.25	50.6	39.89	83.48	0.27	3.98	1.3
51	1EPZ	2.3	5.66	39.11	8.17	−15.87	−11.13	0.53	−2.22	0.64
TTP										
52	1H79	2.67	5.28	29.83	14.28	−18.26	14.91	70.47	−1.33	0.58
53	1N5J	2.52	3.34	50.51	48.49	−19.02	2.66	0.8	11.41	1.42
U5P										
54	1FGX	1.29	3.568	25.81	12.6	22.72	7.64	63.47	16.32	0.77
55	1G8O	1.07	2.548	29.16	13.9	31.90	7.40	36.67	19.71	0.81
56	1HXP	1.65	2.064	21.48	9.47	−9.94	6.97	87.67	17.81	0.59
57	1HXQ	1.52	1.795	33.32	28.96	7.42	29.79	10.93	11.65	0.72
UDP										
58	1C3J	1.21	5.261	27.32	10.24	35.44	43.58	4.27	−11.11	1.24
59	1F7P	2.17	4.046	18.39	7.32	−10.95	39.41	1.8	−0.11	0.82
60	1F7R	1.57	4.782	36.45	22.79	−2.89	29.38	7.27	−15.22	0.48
UTP										
61	1R8C	2.72	4.09	46.77	12.04	−24.11	7.13	27.67	−16.28	1.44
62	*2B56*	*2.63*	*7.64*	*31.31*	*7.86*	*-38.07*	*-10.64*	*42.93*	*17.89*	*1.86*

BS: Best sampled structure. LE: Lowest energy structure. (See text for details). The 5th column specifies the sum of the ligand conformational energy and the interaction energy as calculated for the crystal structure. The 'exact solutions' are marked in bold and the 'alternate binding modes' are marked in italics

Throughout the manuscript we refer to the crystal structure as the native structure, and the binding mode seen in the crystal as the native binding mode.

Results and discussion

In the following discussions we first present the results obtained by MOLSDOCK, before making comparisons with the results of the other two programs. We specifically identify two structures out of the 1500 structures generated by MOLSDOCK for each complex. One is the best sampled structure, i.e., the prediction that has the lowest RMSD with respect to the native structure (abbreviated as BS in the Tables). The unchanged receptor in the predicted structure of the complex was superposed on the native structure of the receptor. The RMSD of the predicted structure and pose of the ligand was then calculated with respect to the native structure and pose of the ligand. This value is used to identify the best sampled structure. The other structure that we identify in the discussions is the prediction that has the lowest total energy of all 1500 structures for each test case (abbreviated as LE in the Tables).

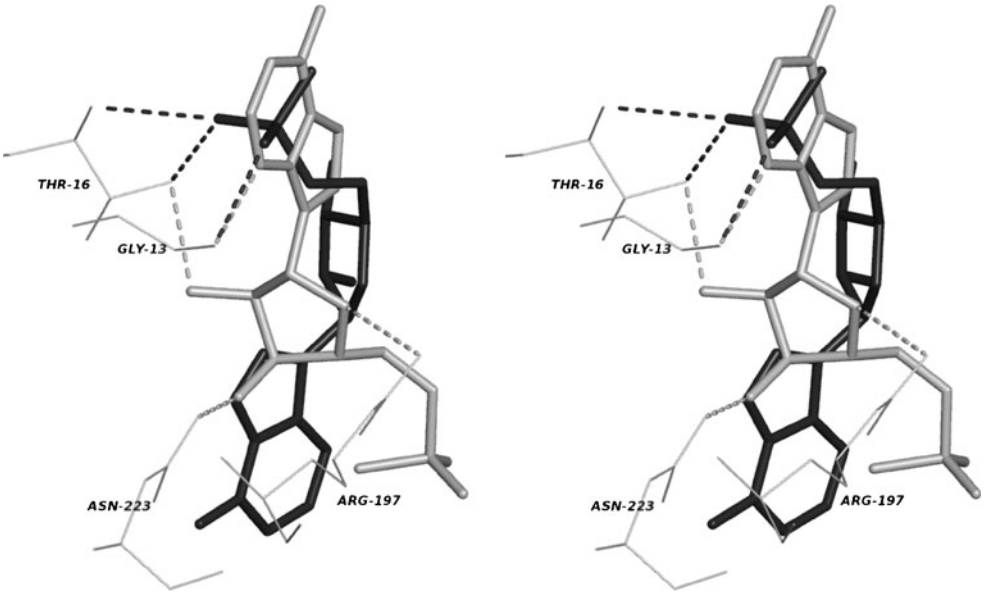

Fig. 2 Stereo view of an alternate binding mode achieved by the 'lowest energy' structure of 1DEL. The figure shows the superposition of this structure on the native structure. The ligand molecule is shown as sticks. The lowest energy ligand as obtained by MOLSDOCK is in gray and the native ligand is in black. Interacting protein residues are labeled and shown as lines. Hydrogen bonds formed by the lowest energy ligand and native ligand are shown in gray and black respectively

Overall results

For any molecular docking tool the most important requirement is its ability to differentiate the real binding conformation and pose of the ligand on the protein from nonspecific and/or energetically unfavorable ones. Based on the restrictions of experimental structure resolution, Gohlke *et al.* [31] suggested a threshold value of 2.0 Å for a docking pose to be correct. This was used as the criterion for the evaluation in the present studies. The RMSD of the best sampled structure is within 2.0 Å in 45 of the 62 cases, and within 2.5 Å in 51 of the 62 cases. Out of the former 45 cases, 15 of 16 complexes with an adenine base, 4 of the 12 complexes with a cytosine base, 15 of the 16 complexes with a guanine base, 5 of the 9 complexes with a thymine base, and 6 of the 9 complexes with a uracil base have their RMSD within 2.0 Å. In the latter group, i.e., when the evaluation criterion is taken to be 2.5 Å, the number of cases that meet this standard are 16 of 16 complexes with an adenine base, 7 of 12 with a cytosine base, 15 of 16 with a guanine base, 6 of 9 with a thymine base and 7 of 9 with a uracil base. Similarly the RMSD of the lowest energy structure is within 2.0 Å in 26 of the 62 cases, and within 2.5 Å in 33 of the 62 cases. Out of the former 26 cases, 10 of 16 complexes with an adenine base, 3 of the 12 complexes with a cytosine base, 9 of the 16 complexes with a guanine base, 3 of the 9 complexes with a thymine base, and 1 of the 9 complexes with a uracil base have their RMSD within 2.0 Å. In the latter group, the number of cases that meet this standard are 13 of 16 complexes with an adenine base, 4 of 12 with a cytosine base, 9 of 16 with a guanine base, 4 of 9 with a thymine base and 3 of 9 with a uracil base. The average RMSD values of the best sampled structure and the lowest energy structure were 1.50 Å and 3.15 Å respectively, for the entire set of 62 complexes. The best sampled structures, positioned and oriented as in the receptor site and superposed without rotation or translation on the native structure, for all 62 test cases are shown in Fig. 1.

Table 2 summarizes the RMSD, relative displacement error (RDE) [32], and docking energy for the best sampled structure and the lowest energy structure in all 62 test cases. The table also gives the energy of the native complex. Both RDE as well as RMSD values were used to evaluate the docking solution. RDE is used to rank different conformations of a ligand of N atoms (i=1, N) docked to the receptor with respect to the known native ligand atoms (j=1, N). The relative displacement error (or fraction correct) is calculated for all N heavy atoms of a ligand by using the following formula as adapted from Abagyan and Totrov [32].

$$\mathrm{RDE} = 100 * \left(1 - \frac{\mathrm{L}}{\mathrm{N}} * \left(\sum_{\substack{\mathrm{i}=1,\mathrm{N} \\ \mathrm{j}=1,\mathrm{N}}} \frac{1}{\mathrm{L} + \mathrm{Dij}}\right)\right), \qquad (1)$$

where L is the scale parameter, N is the number of ligand atoms, Dij is the deviation in the position of the ligand atom i from the corresponding ligand atom j in the crystal structure. The scale parameter defines the accuracy scale. Values of L between 1.5 and 3.0 Å are reasonable, since at larger distances specific interactions of ligand atoms with the receptor atoms are significantly reduced and possibly replaced by different interactions. In this study L value is set to be 2 Å. The above formula has the following properties: if all the deviations are 0, RDE is 0%, if deviations are equal to L, RDE is about 50%, the same result may be achieved if half of the ligand atoms are predicted correctly (or deviate by much less than L), while the other half deviate by much more than L. Hence, in the context of molecular

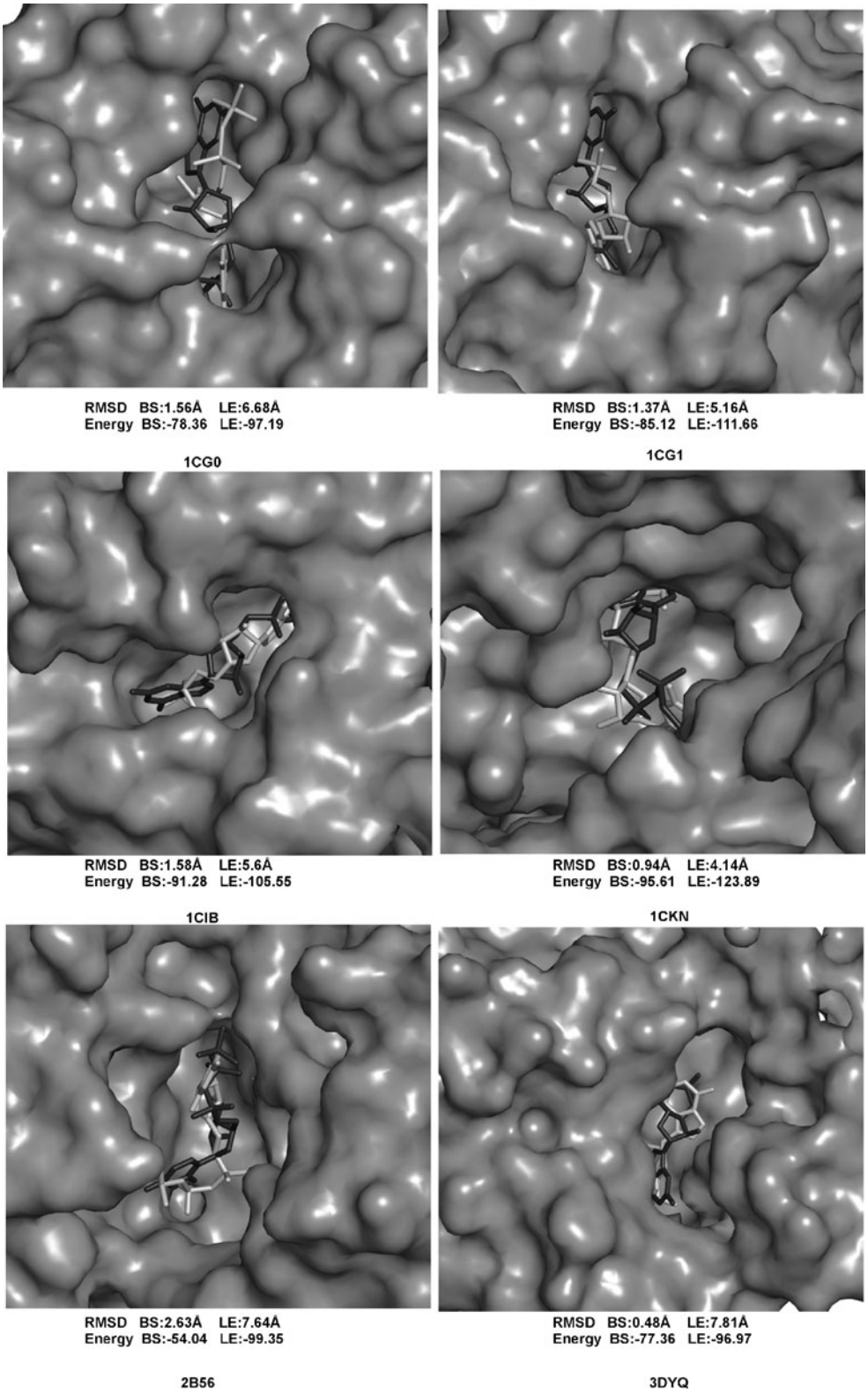

Fig. 3 The alternate binding modes exhibited by the other six lowest energy structures. The lowest energy structure of the ligand as obtained by MOLSDOCK is shown in gray and the native structure is shown in black. The protein molecule is shown as a surface

docking, RDE values are useful to capture specific ligand receptor interactions

Table 2 shows that in 90% of cases the RDE value for the lowest energy structure is lower when compared with the best sampled structure. The difference in the RDE values in some cases is particularly large, indicating the lowest energy structure is much closer to the native complex than the best sampled structure, even though the RMSD values indicate

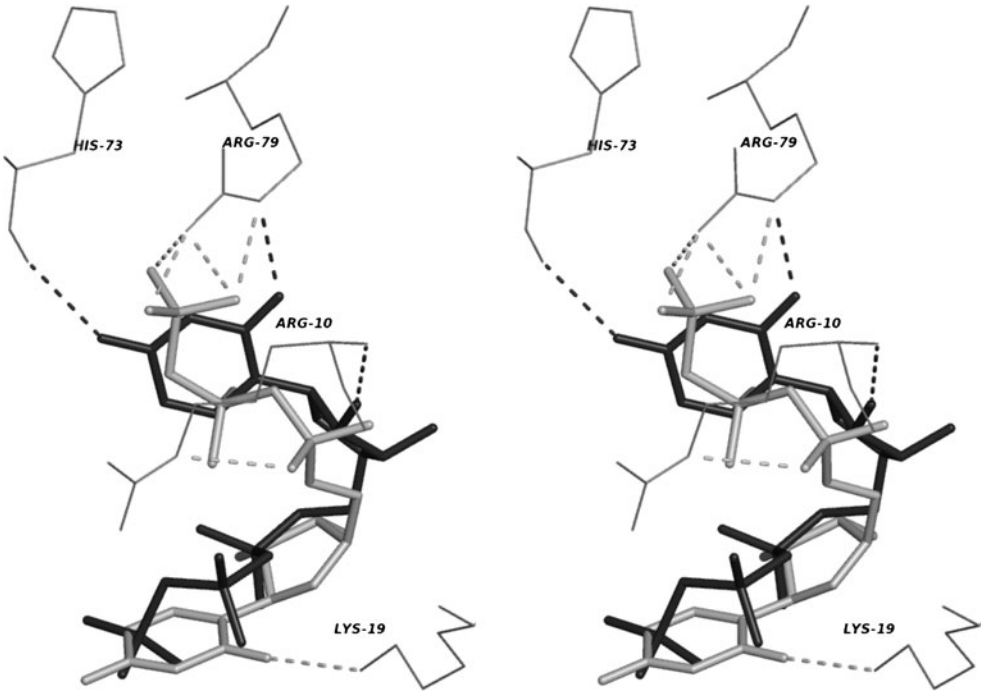

Fig. 4 Stereo view of an alternate binding mode achieved by the 'best sampled' structure of 1H7G. The figure shows the superposition of this structure on the native structure. The ligand molecule is shown as sticks. The best sampled ligand as obtained by MOLSDOCK is in gray and the native ligand is in black. Interacting protein residues are labeled and shown as lines. Hydrogen bonds formed by the best sampled ligand and native ligand are shown in gray and black respectively

just the opposite. For example, in the test case 1DEL, the RMSD of best sampled structure is 1.42 Å, while the lowest energy structure has RMSD of 7.66 Å. The corresponding RDE values are 50.94% and 8.69%, respectively. In order to be consistent in our discussions, however, we use the RMSD as a measure of similarity between two structures.

Exact solutions

As mentioned earlier, 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 non-specific 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. (Only solutions with RMSD less than 2.5 Å as compared to the native are considered.) Here, this is the case in five of the 62 test structures. Table 2 shows these 'exact solutions' in bold. In the other 57 cases, the method finds at least one solution that has a low energy, as well RMSD less than 6.5Å with respect to the crystal structure. The five exact solutions are 1EYR, 1E2F, 1J2J, 1UW1 and 3KCC in which the ligands are cytidine diphosphate, thymidine monophosphate, guanosine triphosphate, adenosine diphosphate and adenosine cyclic monophosphate respectively. The corresponding RMSD with the respective native structures are 2.43, 0.97, 1.33, 1.25 and 0.30 Å.

Alternate binding modes

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. (Such alternate binding modes have been seen experimentally earlier [33].) There are many structures (in Table 2) whose lowest energy structures have high RMSD when superposed with the native structure. An analysis was carried out to check for the presence of alternate binding modes. It showed that 7 out of the 62 cases exhibit alternate binding modes. (These are considered 'alternate binding modes' since they fit into the binding cavity in a pose that is radically different from the native mode, and yet make equally good contacts and interactions, if not better.) An example of this is the structure of deoxynucleoside monophosphate kinase protein in complex with the ligand AMP (PDB ID: 1DEL) [34]. 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 -1.89, -46.85 and -85.87 respectively, i.e., they may be all considered approximately equal [35]. The RMSD of the best sampled and the lowest energy structure as compared to the crystal structure are 1.46 and 7.66 Å respectively. As shown in Fig. 2, in the MOLSDOCK solution the positions of the phosphate group and the base are exchanged as compared to the crystal structure. This lowest energy MOLSDOCK structure makes four hydrogen bonds with the receptor, while the native structure makes three. In six other structures

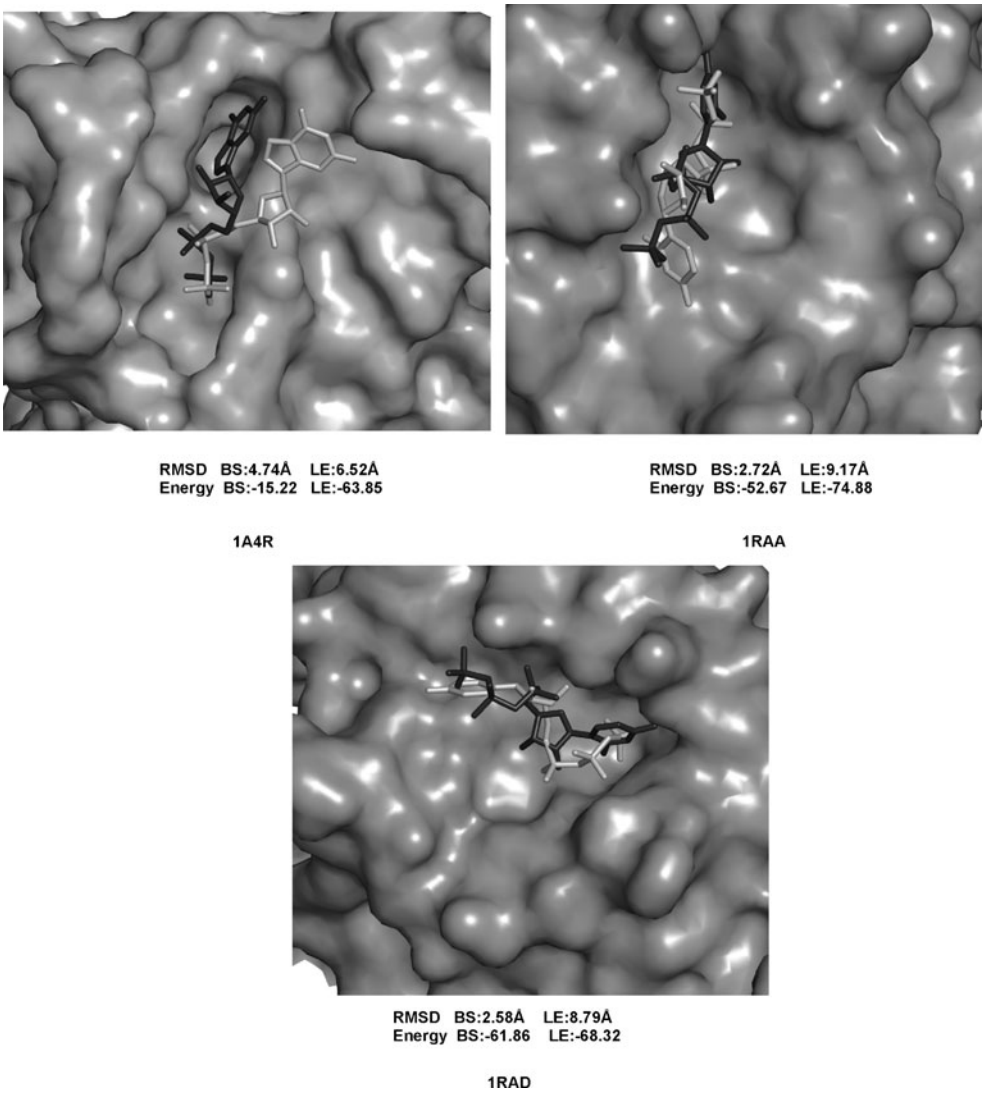

Fig. 5 The alternate binding modes exhibited by the other three best sampled structures. The best sampled structure of the ligand as obtained by MOLSDOCK is shown in gray and the native structure is shown in black. The protein molecule is shown as a surface in gray

also (1CG0, 1CG1, 1CIB, 1CKN, 2B56 and 3DYQ), the lowest energy structures were found to be positioned in the receptor site in an alternate binding mode (Fig. 3) with large RMSD as compared to the native structure (Table 2).

Table 3 Hydrogen bonds and non-bonded contacts found in the native structure, best sampled structure and in the predicted lowest energy alternate binding modes. HB – hydrogen bonds; NB – non-bonded contacts

PDB ID	Native		Total	Best sampled Structure		Total	Lowest energy structure		Total
	HB	NB		HB	NB		HB	NB	
1DEL	3	11	14	2	13	15	4	9	13
1CG0	6	9	15	9	14	23	3	11	14
1CG1	8	10	18	3	18	21	7	14	21
1CIB	4	8	12	4	15	19	4	10	14
1CKN	4	9	13	3	14	17	4	11	15
2B56	4	11	15	3	22	25	3	18	21
3DYQ	4	9	13	4	17	21	3	14	17
1H7G	4	16	20	6	18	24	4	17	21
1A4R	10	13	23	3	16	19	2	18	20
1RAA	2	9	11	2	18	20	3	12	15
1RAD	0	11	11	5	21	26	3	18	21

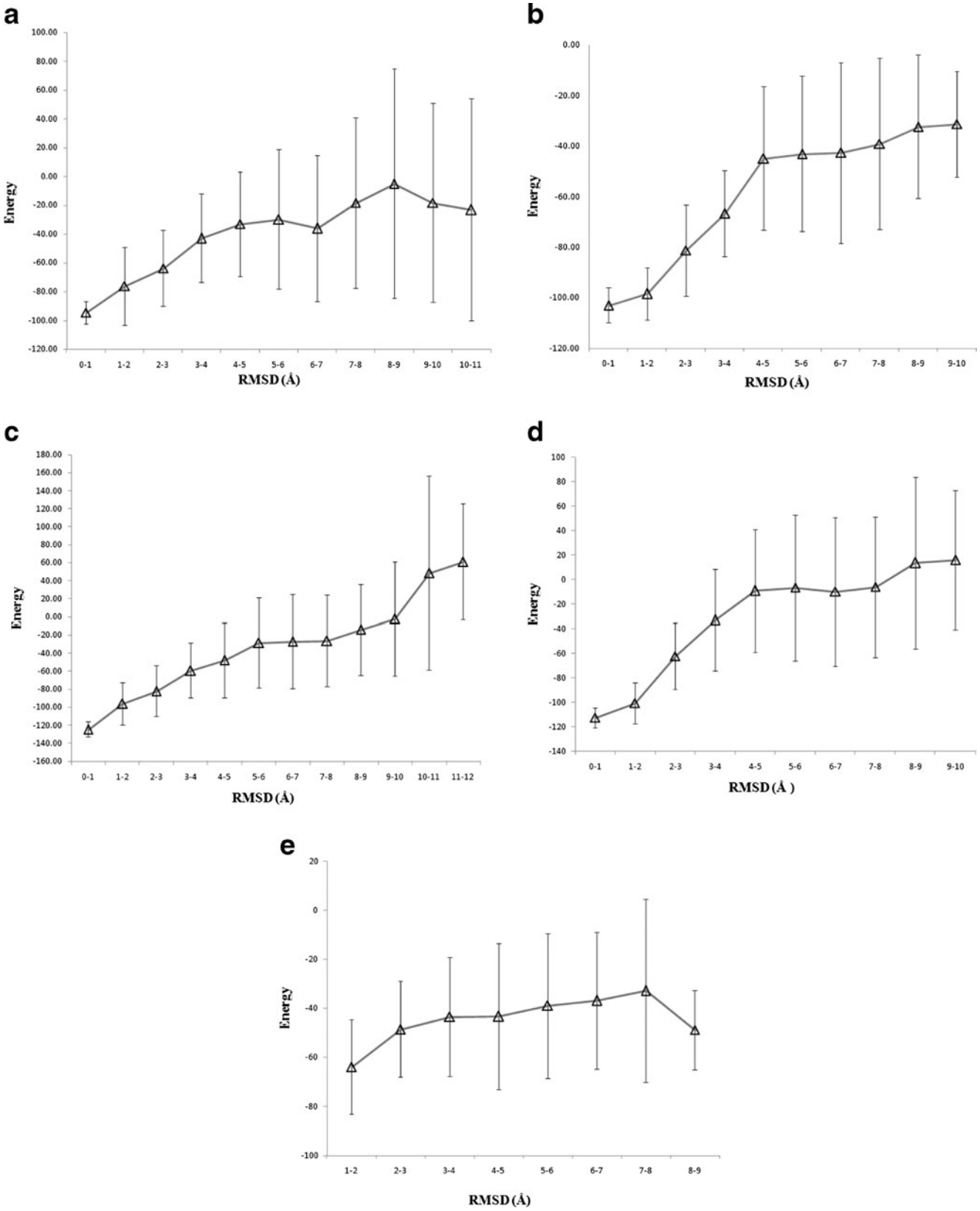

Fig. 6 The average energy of the structures in each bin plotted against different bins of RMSD. The RMSD values are binned at intervals of 1 Å. The error bars show one standard deviation in the average energy. The numbers are plotted for the ligands containing Adenine (**a**), Cytosine (**b**), Guanine (**c**), Thymine (**d**), and Uracil (**e**)

Table 4 Results of the AutoDock runs for the 62 test cases. The table shows the root mean square deviation (RMSD) for the best sampled structures (BS) and lowest energy structures (LE), energy for the (BS) and (LE), and energy of the native structures and the ranking of best sampled structures of AutoDock for all 62 test cases. 'Exact solutions' are shown BOLD

S.NO	PDB ID	RMSD in Å		Energy		BS rank %	Native energy	CPU time in hours
		BS	LE	BS	LE			
AMP								
1	1AER	5.26	14.62	−4.95	−7.15	33.33	−4.55	2.40
2	1DEL	5.15	14.10	−4.5	−6.87	90.00	−4.76	2.40
3	1EFV	3.94	4.52	−6.36	−7.71	82.00	−7.46	2.42
4	1FA9	4.12	5.81	−6.2	−6.64	35.33	−6.35	2.45
ADP								
5	1 AM1	3.50	4.63	−4.74	−5.71	22.67	−5.36	3.22
6	1B4S	3.99	5.05	−6.86	−7.8	78.00	−6.15	3.15
7	1UW1	4.40	5.26	−6.93	−9.6	19.33	−8.69	3.10
8	2DLN	9.66	11.11	−4.83	−5.93	27.33	−4.4	3.07
ATP								
9	1A82	6.44	8.48	−3.9	−5.65	17.33	−3.82	3.87
10	1AQ2	6.50	6.87	−7.31	−8.1	65.33	−7.6	3.90
11	2BUP	8.44	10.52	−4.63	−7	59.33	−6.44	3.92
12	2GNK	4.71	8.35	−6.63	−7.79	47.33	−6.18	4.02
CMP								
13	1G6N	0.29	0.38	−7.26	−7.43	46.00	−7.12	1.83
14	3I54	0.28	0.46	−7.41	−7.44	72.00	−7.09	1.83
15	3KCC	0.92	16.36	−8.05	−8.52	69.33	−7.78	1.85
16	3 N10	2.1	12.4	−5.45	−7.35	18.00	−5.9	1.88
C5P								
17	1H7F	1.31	4.53	−6.24	−8.78	66.00	−7.05	2.20
18	1H7T	1.82	4.63	−6.07	−8.64	54.00	−6.71	2.20
19	1QF9	0.72	9.70	−9.26	−10.02	76.67	−9.37	2.15
20	1 W77	2.9	5.68	−6.47	−7.4	58.00	−6.17	2.17
CDP								
21	1EYR	1.41	11.38	0.99	−2.01	27.33	−6.03	2.62
22	1H7H	4.78	5.92	−4.7	−6.3	94.00	−3.88	2.87
23	1U3L	9.84	17.90	−1.88	−4.31	79.33	−3.08	2.80
24	2CMK	4.24	4.24	−7.31	−7.31	43.33	−6.28	2.88
CTP								
25	1H7G	4.11	5.80	−4.44	−5.83	71.33	−3.54	3.73
26	1I52	3.67	6.10	−7.66	−10.69	73.33	−7.8	3.68
27	1RAA	4.38	6.84	−4.46	−5.51	5.33	−4.38	3.63
28	1RAD	5.75	7.21	−2.48	−4.68	12.67	−3.77	3.63
5GP								
29	1EX7	0.52	1.46	−6.36	−8.66	60.67	−8.38	2.53
30	1G7C	1.38	11.16	−5.91	−6.62	15.33	−5.89	2.50
31	1LVG	1.11	5.99	−8.56	−10.5	24.67	−9.75	2.47
32	1ZNX	1.36	10.17	−5.14	−6.22	44.67	−5.66	2.48
GDP								
33	1A4R	1.23	1.54	−4.84	−9.8	50.67	−9.06	3.27
34	1CG0	2.5	6.31	−5.94	−7.77	50.67	−6.19	3.20
35	1CG1	3.05	7.41	−4.54	−7.76	38.67	−6.02	3.17
36	1CIB	1.34	1.41	−6.62	−8.39	22.00	−7.92	3.22

Table 4 (continued)

S.NO	PDB ID	RMSD in Å		Energy		BS rank %	Native energy	CPU time in hours
		BS	LE	BS	LE			
GTP								
37	1C1Y	4.98	8.27	−7.05	−9.96	38.67	−6.99	4.08
38	1CKN	3.32	5.89	−8.67	−10.64	8.67	−7.12	4.08
39	1E96	5.04	9.88	−6.73	−9.51	11.33	−5.55	4.10
40	1J2J	3.09	5.13	−5.13	−6.87	78.67	−5.5	4.07
PCG								
41	1Q3E	0.24	0.59	−7.67	−8.57	100.00	−8.21	1.97
42	**3CL1**	**0.49**	**0.49**	**−8.78**	**−8.78**	**41.33**	**−8.41**	**1.98**
43	3DYN	3.43	8.90	−5.57	−6.38	64.67	−5.55	1.97
44	3DYQ	7.00	7.12	−7.04	−7.76	2.67	−6.87	1.93
TMP								
45	1CY1	2.42	6.18	−6.15	−8.95	88.00	−8.09	2.20
46	1GSI	4.54	9.65	−9.52	−12.92	82.67	−9.52	2.18
47	1G3U	4.42	10.32	−8.13	−9.69	27.33	−7.63	2.22
48	1E2F	6.47	11.30	−11.68	−13.68	36.67	−11.03	2.22
TYD								
49	1CR4	3.71	15.02	−8.28	−10.07	28.00	−7.58	2.90
50	1E2G	1.88	6.29	−10.7	−12.63	46.67	−10.7	2.92
51	1EPZ	4.44	5.85	−3.19	−5.63	63.33	−3.72	2.88
TTP								
52	1H79	2.1	5.37	−4.88	−7.32	48.00	−5.76	3.63
53	1N5J	3.98	8.97	−8.79	−12.1	26.67	−11.03	3.67
U5P								
54	1FGX	1.09	5.35	−6.86	−8.01	28.67	−6.28	2.20
55	1G8O	1.34	4.73	−6.88	−6.88	92.67	−6.48	2.12
56	1HXP	1.73	14.38	−4.47	−7.03	66.67	−5.19	2.17
57	1HXQ	1.6	13.70	−6.9	−7.47	74.00	−6.38	2.05
UDP								
58	1C3J	0.74	7.03	−7.4	−10.15	38.67	−7.88	2.90
59	1F7P	2.27	6.50	−5.31	−7.33	61.33	−6.38	2.77
60	1F7R	2.65	13.92	−4.77	−6.22	29.33	−4.4	2.68
UTP								
61	1R8C	5.25	17.09	−2.97	−6.68	39.33	−3.34	3.63
62	2B56	5.19	7.13	−5.51	−7.34	54.00	−5.43	3.55

In addition to the above, there are few cases where even the best sampled structures have high RMSD values, i.e., greater than 2.5 Å. When these were analyzed for alternate binding modes, it was found that four of them exhibit such a mode. An example of this is the structure of 3-deoxy-manno-octulosonate cytidyl transferase protein in complex with ligand CTP (PDB ID: 1H7G) [36]. The docked energies of the native, best sampled and lowest energy structures are 4.83, -51.4 and -130.05 respectively. The RMSD of the best sampled and the lowest energy structures as compared to the crystal structure are 6.46 and 7.81 Å respectively (Fig. 4). The native structure makes four hydrogen bonds whereas the best sampled structure makes six hydrogen bonds with the receptor. The other three cases are 1A4R, 1RAA and 1RAD, where the best sampled structures are positioned in the receptor site in an alternate binding mode (Fig. 5) with large RMSD to the native structure (Table 2). In all these cases, (except 1A4R), the ligand is rotated by about 180º in the alternate binding mode as compared to the native structure, i.e., the base and the phosphate group interchange places. The number of hydrogen bonds and the non-bonded contacts in the native structure, the best sampled structure and the lowest energy structure, are given in the Table 3. In two of the cases, viz. 11DEL and 1CG0, the lowest energy alternate binding modes show a

Table 5 Results of the GOLD runs for the 62 test cases. The table shows the root mean square deviation (RMSD) for the best sampled structures (BS) and lowest energy structures (LE), energy for the (BS) and (LE) and the ranking of best sampled structures of GOLD for all 62 test cases. 'Exact solutions' are shown BOLD

S. No	PDB ID	RMSD in Å		Energy		BS rank %	CPU time in hours
		BS	LE	BS	LE		
AMP							
1	1AER	1.35	10.06	53.97	53.97	33.33	0.009
2	1DEL	2.32	6.30	44.26	49.68	75.00	0.018
3	1EFV	0.33	0.38	61.21	64.16	66.67	0.010
4	1FA9	4.54	4.54	42.79	42.79	0.12	0.010
ADP							
5	1 AM1	1.42	4.00	39.98	41.76	100.00	0.011
6	1B4S	2.48	3.05	53.20	55.03	60.00	0.014
7	1UW1	4.64	4.81	64.35	65.37	44.44	0.041
8	**2DLN**	**0.98**	**0.99**	**59.98**	**60.66**	**9.33**	**0.011**
ATP							
9	1A82	2.52	2.68	42.68	58.57	98.29	2.420
10	1AQ2	0.71	1.19	54.78	57.47	74.36	0.201
11	2BUP	0.38	6.99	46.27	62.69	31.02	0.674
12	2GNK	3.40	11.10	43.11	52.66	81.05	0.579
CMP							
13	**1G6N**	**0.24**	**0.24**	**47.03**	**47.03**	**0.16**	**0.007**
14	**3I54**	**1.02**	**1.02**	**44.76**	**44.79**	**8.22**	**0.007**
15	**3KCC**	**0.23**	**0.23**	**49.94**	**49.94**	**0.16**	**0.009**
16	3 N10	0.64	1.04	43.42	43.91	33.33	0.009
C5P							
17	1H7F	4.84	5.76	43.96	49.30	100.00	0.012
18	1H7T	1.97	5.03	38.13	45.93	0.53	0.477
19	1QF9	0.49	4.14	36.76	45.41	40.00	0.017
20	1 W77	2.80	2.80	43.54	43.54	18.52	0.039
CDP							
21	1EYR	4.49	4.49	51.97	51.97	9.11	0.087
22	1H7H	6.71	6.95	48.34	51.87	80.00	0.017
23	1U3L	6.23	6.23	46.66	46.66	0.31	0.027
24	2CMK	1.68	2.13	47.85	52.41	68.18	0.075
CTP							
25	1H7G	5.26	5.84	64.40	69.12	93.12	0.595
26	1I52	6.60	7.20	52.57	59.03	75.00	0.019
27	1RAA	1.92	7.69	41.39	44.05	8.60	3.391
28	1RAD	2.18	8.92	36.73	43.62	2.78	2.240
5GP							
29	1EX7	0.48	0.78	68.76	69.32	66.67	0.009
30	1G7C	0.99	6.65	32.83	47.81	0.33	1.008
31	1LVG	2.61	2.71	62.98	66.94	66.67	0.011
32	1ZNX	2.11	6.06	43.82	52.58	41.46	0.053
GDP							
33	1A4R	1.31	9.89	44.17	46.08	63.64	0.024
34	1CG0	1.29	1.54	61.03	63.89	100.00	0.012
35	1CG1	0.92	1.21	37.33	57.43	7.69	0.050
36	1CIB	0.47	7.90	42.22	55.12	1.89	0.373

Table 5 (continued)

S. No	PDB ID	RMSD in Å		Energy		BS rank %	CPU time in hours
		BS	LE	BS	LE		
GTP							
37	1C1Y	0.58	5.32	62.81	72.59	62.50	0.277
38	1CKN	2.93	4.49	47.22	63.94	83.33	0.021
o39	1E96	1.06	7.01	57.68	75.34	2.58	1.428
40	1J2J	5.82	9.08	49.24	69.70	2.80	4.976
PCG							
41	1Q3E	0.31	0.35	57.54	58.76	66.67	0.007
42	**3CL1**	**0.20**	**0.21**	**49.38**	**49.64**	**4.33**	**0.007**
43	3DYN	0.64	1.25	44.48	46.93	33.33	0.008
44	3DYQ	0.66	0.95	41.00	42.63	33.33	0.008
TMP							
45	1CY1	2.93	4.80	34.20	37.43	65.15	0.065
46	1GSI	1.15	1.31	60.79	61.67	100.00	0.008
47	1G3U	1.31	1.35	62.32	63.28	100.00	0.008
48	1E2F	1.38	2.81	50.88	66.57	97.95	0.564
TYD							
49	1CR4	1.18	5.99	38.20	46.35	55.74	0.325
50	1E2G	1.71	2.31	60.45	64.87	30.00	0.020
51	1EPZ	4.80	8.10	28.60	44.00	80.00	2.006
TTP							
52	1H79	1.65	4.84	47.88	47.88	100.00	0.014
53	1N5J	1.16	2.97	75.16	86.50	75.45	2.902
U5P							
54	1FGX	1.46	7.09	47.45	47.45	100.00	0.010
55	**1G8O**	**1.89**	**1.89**	**53.58**	**53.58**	**7.17**	**0.012**
56	1HXP	1.21	1.60	30.13	39.25	6.42	0.109
57	1HXQ	1.76	6.70	31.84	36.83	22.22	0.022
UDP							
58	1C3J	1.00	7.93	37.80	45.18	26.32	0.133
59	1F7P	0.88	4.22	42.16	47.64	10.26	0.181
60	1F7R	2.46	2.55	48.79	50.41	100.00	0.010
UTP							
61	1R8C	3.72	7.52	45.39	50.41	31.82	0.058
62	2B56	1.86	7.71	49.73	52.97	57.58	0.088

smaller number of interactions than the respective native counterpart. In the other cases, namely 1CG1, 1CIB, 1CKN, 2B56, 3DYQ, 1H7G, 1RAA and 1RAD, the total number of interactions found in the native complex is less than in the predicted best sampled and the lowest energy alternate binding modes.

Correlation between energy and RMSD

When the best sampled structure was scored in terms of the energy, it was found that 40 of the 62 complexes were in the top 10% ranking. These are found predominately in the lowest energy regions. Figure 6 shows a plot of the number of structures that fall in specific bins of total docked energy and as well as of RMSD values for the 16 complexes with adenosine, 12 complexes with cytosine, 16 complexes with guanidine, nine complexes with thymine and nine complexes with uracil nucleotides. Out of the 40 cases, 12 cases with adenosine, six with cytosine, 12 with guanidine, seven with thymine and three with uracil nucleotide fall in top 10% of the lowest energy bins.

However, the reverse is generally not true, and it is observed that the lowest 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 1C1Y that has an RMSD of 3.03 Å with the native. The energy of this prediction is -122.84. (This is lower than the energy of the native complex, which is -12.53 when calculated using the same formula). Nine hydrogen bonds were observed between the protein and the lowest energy ligand, but the native structure had eight hydrogen bonds. A total of nine non-bonded contacts were observed between the protein and the native ligand and but 12 non-bonded contacts were found between the protein and the lowest energy ligand.

Similarly, the lowest energy predictions of two other cases, 1B4S and 1F7R have large RMSD of 4.61 and 4.78 Å respectively with the native. In the case of the native ligand of 1B4S and 1F7R, the total number of non bonded contacts found is seven and 19 respectively, whereas 19 and 11 were found in their respective lowest energy prediction. In the case of the native ligand of 1B4S and 1F7R, the total number of hydrogen bonds found is five and zero respectively, whereas three and four were found in their respective lowest energy prediction. In two more cases, 1C3J and 1FGX the lowest energy prediction have large RMSD of 5.26 and 3.56 Å respectively with the native. The energy of the native and the lowest energy prediction of 1C3J are -11.11 and -91.19 respectively. The energy of the native and the lowest energy prediction of 1FGX are 16.32 and -91.11 respectively. In the above two cases, the energy of the lowest energy prediction is less than its native energy. In the case of lowest energy prediction of 1C3J and 1FGX, the total number of non bonded contacts found is 16 and 10 respectively, whereas seven and 11 were found in their respective native structure.

The overall energy/RMSD correlation coefficients for the complexes with adenosine, cytosine, guanidine, thymine and uracil nucleotides are 0.84, 0.91, 0.83, 0.92 and 0.63 respectively.

Comparison with AutoDock and GOLD

Tables 4 and 5 show the results of the AutoDock and GOLD runs, respectively, for the 62 cases. Of the 62 cases, the best sampled structure had RMSD from the crystal structure of less than 2.50 Å in 25 complexes in the case of AutoDock results, 45 complexes in GOLD. (In the MOLSDOCK results there are 51 such complexes). In three cases, the structure best sampled by AutoDock was found within the top 10% when ranked in terms of energy and 18 cases in GOLD. (There are 40 such cases by the MOLSDOCK method). AutoDock was able to identify one exact solution (3CL1), i.e., the solutions in which the best sampled structure is the same as the lowest energy structure. GOLD was able to find six complexes with an exact solution. (MOLSDOCK found five exact solutions.). AutoDock was able to identify alternate binding modes, in six cases. The cases were 1C3J, 1EYR, 1H7F, 1H7T, 3KCC and 3N10 respectively. GOLD also identified alternate binding modes in four cases. They are 1AER, 1EX7, 1FGX and 2B56 respectively. MOLSDOCK identified seven alternate binding modes. The alternate modes identified by AutoDock and MOLSDOCK are not the same. In one of the cases (2B56) both GOLD and MOLSDOCK identified the same alternate binding mode.

Conclusions

The MOLSDOCK ‘rigid receptor - flexible ligand’ docking algorithm was tested on 62 protein–ligand (nucleotide) complexes. In general, it is a suitable method when it is desirable to explore both conformational space and docking space simultaneously and exhaustively, 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.

Acknowledgments We thank the University Grants Commission, and the Department of Science and Technology, Government of India for support under the Centre of Advanced Study (CAS) program and the Fund for Improvement of S&T Infrastructure (FIST) program, respectively.

References

1. Shoichet BK, Kuntz ID (1996) Predicting the structure of protein complexes: a step in the right direction. Chem Biol 3:151–156
2. Klebe G (2000) Recent developments in structure-based drug design. J Mol Med 78:269–281
3. Lybrand TP (1995) Ligand-protein docking and rational drug design. Curr Opin Struct Biol 5:224–228
4. Lengauer T, Rarey M (1996) Computational methods for biomolecular docking. Curr Opin Struct Biol 6:402–406
5. Huang SY, Zou X (2010) Advances and challenges in protein-ligand docking. Int J Mol Sci 11:3016–3034
6. Bissantz C, Folkers G, Rognan D (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 43:4759–4767
7. Kontyianni M, McClellan LM, Slkol GS (2004) Evaluation of docking performance: comparative data on docking algorithms. J Med Chem 47:558–565
8. Glick M, Rayan A, Goldblum A (2002) A stochastic algorithm for global optimization and for best populations: A test case of side chains in proteins. Proc Natl Acad Sci USA 99:703–708
9. Pappu RV, Hart RK, Ponder JW (1998) Analysis and application of potential energy smoothing and search methods for global optimization. J Phys Chem B 102:9725–9742
10. Nair N, Goodman JM (1998) Genetic algorithms in conformational analysis. J Chem Inf Comput Sci 38:317–320
11. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian

genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662
12. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261:470–489
13. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748
14. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749
15. Makino S, Kuntz ID (1997) Automated flexible ligand docking method and its application for database search. J Comput Chem 18:1812–1825
16. Vengadesan K, Gautham N (2003) Enhanced sampling of the molecular potential energy surface using mutually orthogonal Latin squares: application to peptide structures. Biophys J 84:2897–2906
17. Vengadesan K, Gautham N (2004) Conformational studies on Enkephalins using the MOLS technique. Biopolymers 74:476–494
18. Vengadesan K, Gautham N (2004) An application of experimental design using mutually orthogonal Latin squares in conformational studies of peptides. Biochem Biophys Res Commun 316:731–737
19. Prasad PA, 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
20. Viji SN, Prasad PA, Gautham N (2009) Protein-ligand docking using mutually orthogonal Latin squares (MOLSDOCK). J Chem Inf Model 49:2687–2694
21. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197
22. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST (1995) Molecular recognition of the inhibitor AC-1343 by HIV-l protease: conformationally flexible docking by evolutionary programming. Chem Biol 2:317–324
23. Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 46:2287–2303
24. Wang R, Lu Y, Fang X, Wang S (2004) An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. J Chem Inf Comput Sci 44:2114–2125
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
26. Wittinghofer A (1992) Three-dimensional structure of $p21^{H\text{-}ras}$ and its implications. Cancer Biol 3:189–198
27. Biosym/MSI Release 95.0 (1995) San Diego, CA 92121 – 3752, USA
28. Andrusier N, Mashiach E, Nussinov R, Wolfson HJ (2008) Principles of flexible protein-protein docking. Proteins 73:271–289
29. Dean PM, Poornima CS (1995) Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein–ligand interactions. J Comput Aided Mol Des 9:500–512
30. Rosenfeld RJ, Goodsell DS, Musah RA, Morris GM, Goodin DB, Olson AJ (2003) Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. J Comput Aided Mol Des 17:525–536
31. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 295:337–356
32. Abagyan RA, Totrov MM (1997) Contact area difference (CAD): A robust measure to evaluate accuracy of protein models. J Mol Biol 268:678–685
33. Moliner ED, Brown NR, Johnson LN (2003) Alternative binding modes of an inhibitor to two different kinases. Eur J Biochem 270:3174–3181
34. Teplyakov A, Sebastiao P, Obmolova G, Perrakis A, Brush GS, Bessman MJ, Wilson KS (1996) Crystal structure of bacteriophage T4 deoxynucleotide kinase with its substrates dGMP and ATP. EMBO J 15:3487–3497
35. 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
36. Jelakovic S, Schulz GE (2001) The structure of CMP: 2-keto-3-deoxy-manno-octonic acid synthetase and of its complexes with substrates and substrate analogs. J Mol Biol 312:143–155