An Anchor-Dependent Molecular Docking Process for Docking Small Flexible Molecules into Rigid Protein Receptors

Thy-Hou Lin* and Guan-Liang Lin

Institute of Molecular Medicine and Department of Life Science, National Tsing Hua University, HsinChu, Taiwan 30013, R.O.C.

Received April 10, 2008

A molecular docking method designated as ADDock, anchor-dependent molecular docking process for docking small flexible molecules into rigid protein receptors, is presented in this article. ADDock makes the bond connection lists for atoms based on anchors chosen for building molecular structures for docking small flexible molecules or ligands into rigid active sites of protein receptors. ADDock employs an extended version of piecewise linear potential for scoring the docked structures. Since no translational motion for small molecules is implemented during the docking process, ADDock searches the best docking result by systematically changing the anchors chosen, which are usually the single-edge connected nodes or terminal hydrogen atoms of ligands. ADDock takes intact ligand structures generated during the docking process for computing the docked scores; therefore, no energy minimization is required in the evaluation phase of docking. The docking accuracy by ADDock for 92 receptor−ligand complexes docked is 91.3%. All these complexes have been docked by other groups using other docking methods. The receptor−ligand steric interaction energies computed by ADDock for some sets of active and inactive compounds selected and docked into the same receptor active sites are apparently separated. These results show that based on the steric interaction energies computed between the docked structures and receptor active sites, ADDock is able to separate active from inactive compounds for both being docked into the same receptor.

INTRODUCTION

In silico drug discovery or virtual screening requires docking a huge number of small molecules into the active site of a protein receptor and then screening out compounds bound with favorable energies, chemical complementarities, or geometries for further experimental tests. Docking is a computationally challenging task[1,2] because there are numerous ways to place a small molecule inside the active site of a receptor and then move and rotate the entire molecule while allowing all the flexible bonds in the molecule to rotate during the docking process. Docking could be even more challenging if all the flexible bonds of receptor and the effect of solvent are taken into account as well. Given the current limitations on computational technology, it is still intractable to use a complicated docking algorithm to conduct virtual screening on a huge database within a reasonable amount of time. A simple docking algorithm has the advantage in computational speed but should be accurate enough for predicting the native binding modes for small molecules.

Most of the early docking studies assume that both small (ligand) and protein (receptor) molecules are rigid such that all the internal degrees of freedom of ligand and receptor are fixed except for the three translational and rotational ones of ligand. Receptor atoms are represented explicitly, and simple potential energy functions are used to evaluate the complementarity of proposed complex configurations.[3,4] Obviously, the rigid body approximation of ligand−receptor interactions does not account for conformational changes,[5] especially when a cryptic site in the uncomplexed receptor opens under the influence of the ligand. A simple approach to sampling some of the ligand degrees of freedom was described by DesJarlais et al.[6] Computation of molecular surfaces is subsequently applied in docking experiments to significantly reduce the number of possible complexes and increase the speed of docking algorithms.[7–10] To further increase the speed of docking algorithms, the physicochemical properties of receptors are computed beforehand and stored on some grid points constructed on the receptor van der Waals surfaces.[11,12] Both DOCK[13] and AutoDock[14] are among the first docking programs to implement the use of grids automatically during the docking process. A conformational search for a ligand is necessary if some flexible bonds of a ligand are allowed to vary during the docking process. There are systematic, deterministic, and stochastic conformational searching methods being developed to treat the ligand flexibility. In the systematic search algorithms, values of each formal degree of freedom are expressed on a grid, and each of these grid values is explored in a combinatorial fashion during the search. The anchor-and-grow or incremental construction algorithm implemented in DOCK[13] is a systematic search method. The energy minimization and molecular dynamics (MD) simulations are the two known deterministic searching methods where the initial state determines the move of the next state.

In the stochastic search algorithms, random changes are made to change the translation and rotation of entire ligand as well as the torsion angles. Both Monte Carlo (MC) and evolutionary or genetic algorithm (GA) are the stochastic search methods. AutoDock implements the simulated annealing

* Corresponding author phone: 886 3 574 2759; fax: 886 3 571 5934, e-mail: thlin@life.nthu.edu.tw.

10.1021/ci800124g CCC: $40.75
Published on Web 07/22/2008

Table 1. Summary of the Docking Results of 92 Ligand (Protein) Complexes Docked by ADDock

lig(pro)	*rmsd*	**g/a**[a]	no. of rot bonds	anc atom[b]	lig(pro)	*rmsd*	**g/a**[a]	no. of rot bonds	anc atom[b]
Ara(1abe)	0.12	96/20	0	16	Dig(1did)	1.48	120/24	2	12
Ntm(1qpq)	0.14	99/16	1	13	Edr (2ack)	1.50	161/28	2	16
Dnj(1die)	0.18	119/21	1	19	Gep(25c8)	1.54	491/52	7	35
Npa(1ngp)	0.28	143/20	1	15	Saa(1mdr)	1.54	156/21	1	12
Aso(1xie)	0.29	99/23	1	12	Gy(3cpa)	1.54	283/31	4	29
Pqq(4aah)	0.30	229/28	0	28	815(1f0r)	1.54	570/50	3	36
Tha(1acj)	0.30	200/28	0	28	Fk5(1fkf)	1.56	1612/126	7	124
Fad(1coy)	0.33	1143/83	11	2	W33(2r07)	1.57	365/45	8	24
Hds(1lic)	0.40	562/53	14	35	Sle(1atl)	1.58	101/23	3	10
C60(1rne)	0.41	1775/109	19	81	Ltr(1wap)	1.58	168/25	2	23
Azm(1azm)	0.46	101/18	1	9	Aah(1kel)	1.58	309/43	7	27
Ae2(1dbj)	0.49	275/51	0	24	Npp(1baf)	1.59	512/55	5	25
Tk4(1c5c)	0.49	324/34	1	26	2am(6rnt)	1.59	353/35	2	3
Dan(2sim)	0.50	246/35	4	21	Otg(2yhx)	1.61	341/39	1	25
Rtl(1rbp)	0.55	535/51	0	36	Tzl(1mup)	1.62	129/22	2	18
Fog(1eed)	0.60	326/31	5	25	Btn(1stp)	1.66	240/32	4	27
Oai(1c83)	0.62	208/24	2	23	Tpm(1f3d)	1.67	233/30	2	28
Hem(1phg)	0.67	665/75	6	50	S5h(2dbl)	1.68	774/68	5	63
Mma(5cna)	0.72	151/27	2	17	Blv(1bbp)	1.68	899/77	6	45
Hac(1d6e)	0.72	118/28	2	20	Apr(2phh)	1.71	812/60	7	25
Nab(1srj)	0.80	359/34	0	33	Fmn(1d3h)	1.73	363/50	6	6
Fmc(1mrk)	0.81	212/32	2	20	Ptp(1snc)	1.75	399/38	4	29
Des(3erd)	0.84	409/40	2	32	Apa(1tpp)	1.76	168/25	2	25
Vac(4phv)	0.84	1228/86	12	84	Pgh(7tim)	1.76	68/13	2	13
C2p(1rob)	0.86	297/33	3	7	Oht(3ert)	1.76	570/58	5	43
Asc(1xid)	0.88	112/19	2	13	Agb(1ejn)	1.78	462/48	4	41
Amp(2ak3)	0.97	265/35	3	35	Nad(1hdy)	1.78	1136/70	8	25
Str(1dbb)	0.98	546/53	1	42	Ndp(1dg5)	1.79	1126/75	10	2
Spc(1d7x)	0.99	342/39	4	25	Tra(1aco)	1.81	77/15	2	13
Nop(1ase)	1.02	171/25	2	17	Vk(1lna)	1.81	380/41	7	4
Stu(1byg)	1.03	506/60	2	24	Mqi(1etr)	1.84	164/27	0	17
Pal(1acm)	1.04	253/24	4	24	Bzs(1cbx)	1.86	112/26	3	20
Fos(1blh)	1.06	206/25	4	25	E64(1aec)	1.89	620/54	11	44
Ben(3ptb)	1.09	66/16	0	14	Lof(2ctc)	1.92	105/19	2	13
Pp1(1qcf)	1.10	330/43	2	29	Uvc(6rsa)	1.93	322/38	3	12
Nev(3hvt)	1.17	301/34	1	27	Pvb(1ckp)	1.94	324/30	2	25
Via(1udt)	1.22	922/68	4	10	Rea(1epb)	1.99	418/50	0	23
Cpm(1cps)	1.25	204/36	4	21	Sc4(1eah)	1.99	400/45	7	30
961(4lbd)	1.29	460/53	2	39	Clt(1tmn)	2.16	165/23	3	16
Sb3(1fkg)	1.31	758/68	9	37	N3t(1tka)	2.70	537/42	7	25
Fds(4fab)	1.37	458/38	0	29	S57(1hri)	2.95	413/42	9	42
Cbo(1hdc)	1.38	737/93	4	86	Gas(2cgr)	3.61	608/45	1	44
Oaa(4cts)	1.38	40/11	1	8	Mid(1dwd)	3.61	728/66	6	53
Hep(1eap)	1.39	391/41	7	35	Nap(1dr1)	4.37	1379/74	9	2
Pfz(1pha)	1.44	378/44	8	36	Gtb(1glq)	5.35	748/48	10	46
Mbh(1bma)	1.44	205/27	3	17	Mtx(4dfr)	5.91	607/57	7	35

[a] **g/a** represents the following ratio: #cavity grid points counted/#ligand atoms. [b] anc atom represents the atom identification number of the anchor chosen.

MC as well as the GA method,[14] while MCDOCK generates the protein−ligand configurations based on an overlap function and then evaluates the MC-based energy.[15] The PRODOCK program[16] is a biased MC method because it uses information about the topology of the potential energy hypersurface to scale the energy after each MC step for docking. Similar to a MC method, PRO_LEADS[17] conducts a Tabu search where the conformational space is randomly explored, and a new solution is kept in a Tabu list if its energy is not lower and it is not similar to anything on the Tabu list. Characterized by some crossover and mutation processes, the GAs are also stochastic methods used to find the global energy minimum. Complex, nondifferentiable scoring functions can be used in the GAs to compute the fitness functions, and the fittest individuals are kept and transferred to the next generation. To increase genetic diversity and prevent premature convergence, some large regions of parents are swapped or crossovered, and random or biased mutations are made during the evolutionary process of the artificial genes. A large single population or multiple subpopulations of parental genes can be generated to make the offspring ones. The GOLD program[18] uses multiple subpopulations and manipulates them simultaneously. The GAs are also implemented in several other docking methods such as GEMDOCK,[19] MolDock,[20] and EADock.[21]

To incorporate ligand flexibility during docking process, the ligand molecule is either incrementally built or flexed during the search or a database of rigid precomputed conformers is oriented in the active site. In the anchor-and-grow method,[22–24] a ligand is divided into rigid and flexible regions based on the rotatable bonds identified between them. Rigid parts are chosen as anchors and docked first, and then flexible parts are sequentially added by systematic scanning of the torsion angles. Greedy algorithms are employed to optimize each subsequent torsion placement to ensure that the overall molecular position obtained is optimal. While the

Table 2. Comparing Some Docking Results[a] by ADDock with Those by Several Other Docking Methods

Pro	ADDock	MolDock	Glide	GOLD	FlexX	Surflex	Pro	ADDock	MolDock	Glide	GOLD	FlexX	Surflex
1abe	0.12	0.26	0.17	0.86	1.16	0.27	1mdr	1.54	1.09	0.52	0.36	0.88	0.68
1acj	0.30	0.78	0.28	4.00	0.49	3.89	1mrk	0.81	1.37	1.20	1.01	3.55	0.85
1acm	1.04	0.56	0.29	0.81	1.39	1.43	1phg	0.67	1.38	4.32	1.35	4.74	4.44
1aco	1.81	0.42	1.02	0.86	0.96	3.39	1rob	0.86	1.13	1.85	3.75	7.70	0.82
1atl	1.58	1.59	0.94	n/a	2.06	7.01	1snc	1.75	1.69	1.91	n/a	7.48	4.92
1baf	1.59	1.60	0.76	6.12	8.27	6.52	1srj	0.80	0.44	0.58	0.42	2.36	0.39
1bbp	1.68	0.99	4.96	n/a	3.75	1.07	1stp	1.66	0.76	0.59	0.69	0.65	0.51
1bma	1.44	1.04	9.31	n/a	13.41	1.00	1tka	2.70	1.35	2.28	1.88	1.17	1.96
1cbx	1.86	1.06	0.36	0.54	1.35	0.70	1tmn	2.16	5.58	2.80	1.68	0.86	1.30
1coy	0.33	0.66	0.28	0.86	1.06	0.54	1wap	1.58	0.48	0.12	n/a	0.57	0.30
1dbb	0.98	1.62	0.41	1.17	0.81	0.54	2ak3	0.97	0.49	0.71	5.08	0.91	0.60
1dbj	0.49	0.93	0.20	0.72	1.22	0.88	2cgr	3.61	0.92	0.38	0.99	3.53	1.63
1dr1	4.37	0.65	1.47	1.41	5.64	1.25	2ctc	1.92	0.37	1.61	0.32	1.97	0.38
1dwd	3.61	1.07	1.32	1.71	1.66	1.68	2dbl	1.68	1.55	0.69	1.31	1.49	0.81
1eap	1.39	2.52	2.32	3.00	3.72	4.89	2phh	1.71	0.69	0.38	0.72	0.43	0.44
1epb	1.99	3.35	1.78	2.08	2.77	2.87	2r07	1.57	1.81	0.48	8.23	11.63	1.35
1etr	1.84	1.96	1.48	4.23	7.24	4.05	2sim	0.50	1.29	0.92	0.92	1.99	1.10
1fkg	1.31	1.89	1.25	1.81	7.59	1.81	3cpa	1.54	1.63	2.40	1.58	2.53	1.90
1glq	5.35	7.09	0.29	1.35	6.43	5.68	3hvt	1.17	0.35	0.77	1.12	10.26	1.64
1hdc	1.38	1.71	0.58	10.49	11.74	1.80	3ptb	1.09	0.17	0.27	0.96	0.55	0.54
1hdy	1.78	1.73	1.74	0.94	n/a	0.66	4cts	1.38	0.60	0.19	1.57	1.53	2.20
1hri	2.95	6.33	1.59	14.01	10.23	1.98	4dfr	5.91	1.39	1.12	1.44	1.40	1.60
1lic	0.40	2.44	4.87	10.78	5.07	3.46	6rnt	1.59	7.48	2.22	1.20	4.79	7.03
1lna	1.81	3.04	0.95	n/a	5.40	0.88	7tim	1.76	0.58	0.14	0.78	1.49	1.20

[a] The *rmsd* value of each docking result by each method[20] is used in the comparison.

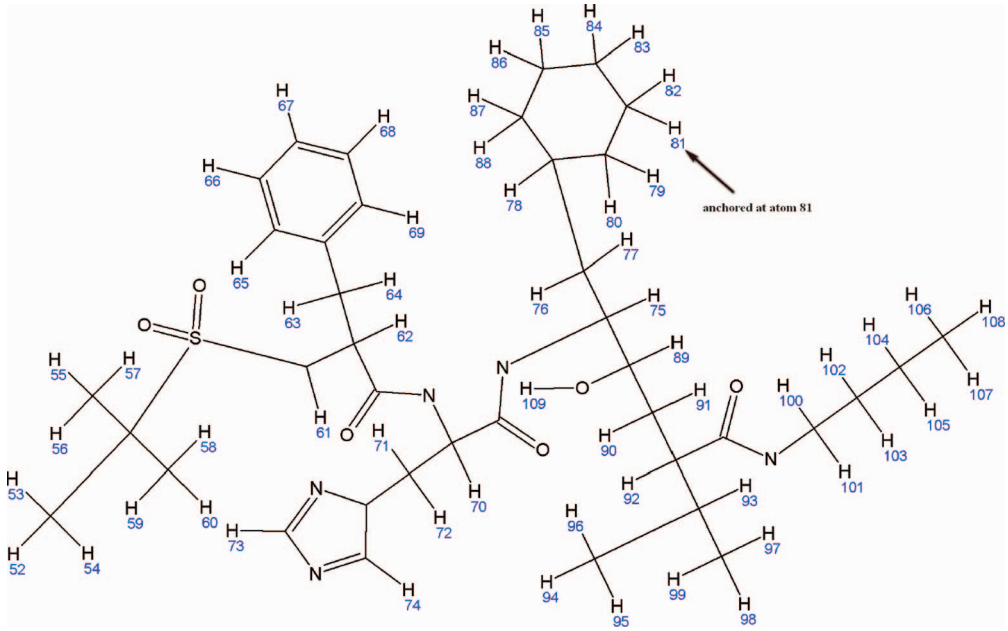


Figure 1. Each of the terminal hydrogen atoms of ligand C60 (1rne)[57] is preliminarily selected as an anchor and labeled by its atomic id number. There are 58 preliminary structures generated for C60 since the corresponding number of terminal hydrogen atoms identified is 58. Each of these structures is docked into the 1rne active site and the best anchor, atom 81 (marked by an arrow on the diagram), is chosen with the minimum docked energy computed among all the preliminary anchors.

redundancy encountered in torsional sampling may be eliminated, setting up databases of precomputed conformers for docking could be prohibitive if there are multiple rotatable bonds existing in a ligand.[25] However, the docking speed can be dramatically increased if ensembles instead of individual conformers are simultaneously docked. FLOG generates conformational libraries called Flexibases using distance geometry and then docks the Flexibases.[26,27] Conformers have also been overlaid on their common largest rigid fragment for the initial placement of the ensemble into the active site.[28] No greedy algorithms are required by this method because all the conformers are scored. However, the docking result of this method depends on the anchor placement because only a small contiguous subset of ligand atoms is used in the initial docking step. This method has been extended to conformers of different molecules sharing the same anchors.[29] In the fragment-based docking method, the most favorable positions and orientations of rigid ligand fragments in the active site are determined first and then used to guide the placement of ligand conformations generated by a GA.[30]

Locating ligand binding poses in receptor active sites is an interesting research issue since in most cases the binding site cavity is considerably larger than the ligand. Moreover,

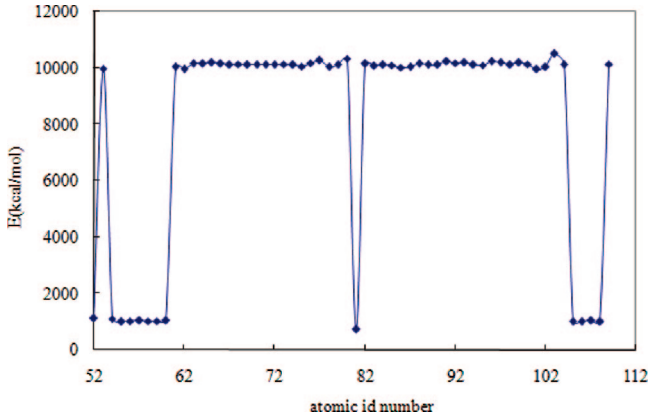


Figure 2. The corresponding docked energy of each of the 58 preliminary dockings described in Figure 1 is plotted as a function of the atomic id numbers. The docked energy obtained for atom 81 chosen as an anchor is 744 kcal/mol (−256 interaction plus 1000 kcal/mol penalty) which is identified to be the smallest among all these energies computed. The number of GA steps is uniquely set as 40000 for these preliminary dockings.

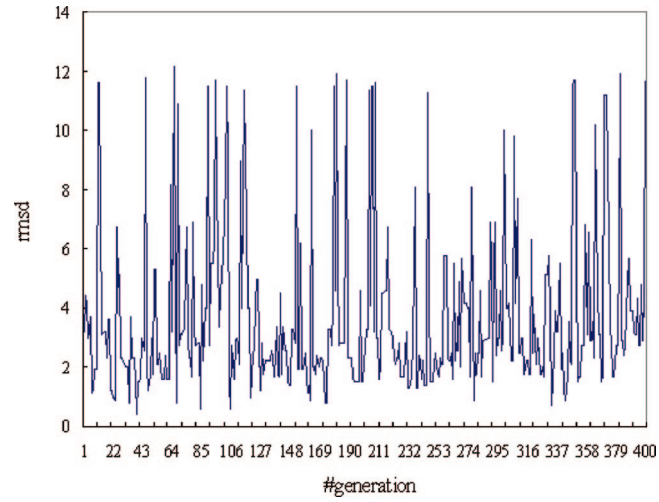


Figure 4. The minimum *rmsd* searched within each generation for C60 is plotted against the generation number during a docking run.

the binding cavity of many enzymes is occupied by more than one molecule (e.g., ligand and cofactor).[31] In this report, we presented an anchor-dependent docking method designated as ADDock for docking flexible ligands into rigid receptor active sites. We chose a ligand atom as an anchor and then constructed the coordinates of the entire ligand based on a bond connection list or topology counted from the anchor. The information for some rotatable bonds identified for the ligand was included in the topology. During the docking process, the entire ligand was rotated by quaterions while being accompanied by simultaneous changes of all the rotatable bonds identified. The four quaterions and values of all the rotatable bonds identified were evolved by a GA. These values were then used in computing a new ligand conformation based on the topology counted from the anchor. The new ligand conformation generated was not subjected to any energy minimization scheme. The interaction energy between each ligand conformation generated and all the receptor atoms identified inside a docking box was scored by a piecewise linear potential scoring scheme.[19,20,32] We found that good docking results were obtained by selecting terminal hydrogen atoms that were deeply buried inside a receptor active site as anchors. We discussed how to build the coordinate of a ligand from the topology counted from an anchor chosen and how to use the GA to evolve the docking parameters in the Methods section. The docking results of 92 receptor−ligand complexes selected from the literatures by ADDock were presented and analyzed in the Results and Discussion section. We have also docked some active and inactive compounds selected from the literature into the same receptors and found that these docked compounds were well separated by the receptor−ligand steric

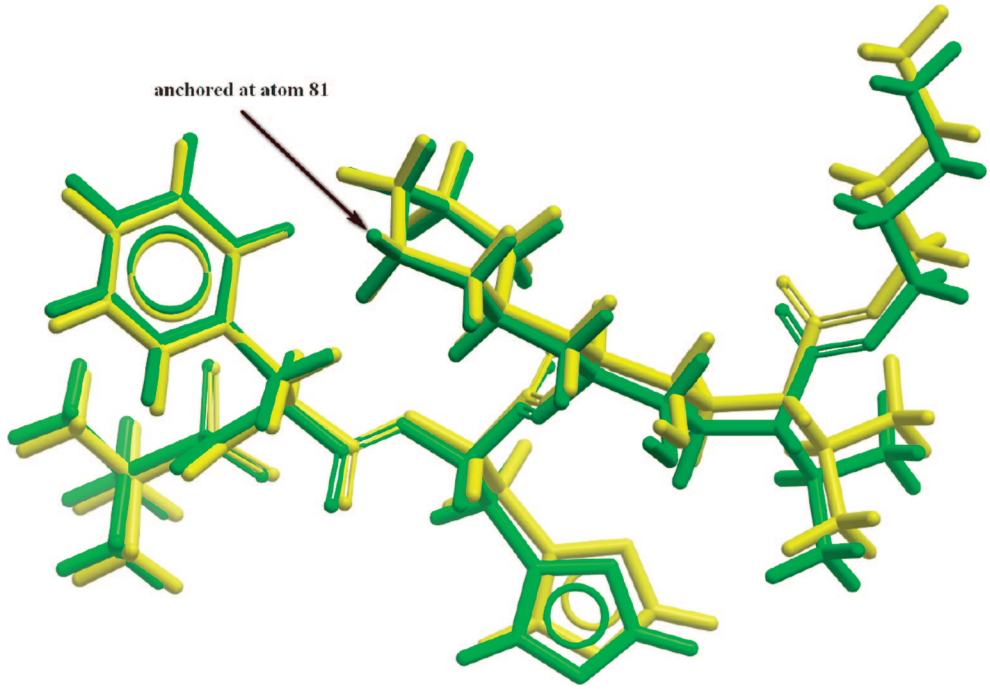


Figure 3. A comparison for the docked conformation (represented by green color) and crystal structure (represented by yellow color) for ligand C60. Both structures are rooted on atom 81 as the anchor chosen which is marked by an arrow. The *rmsd* computed between them is 0.41 Å (Table 1).

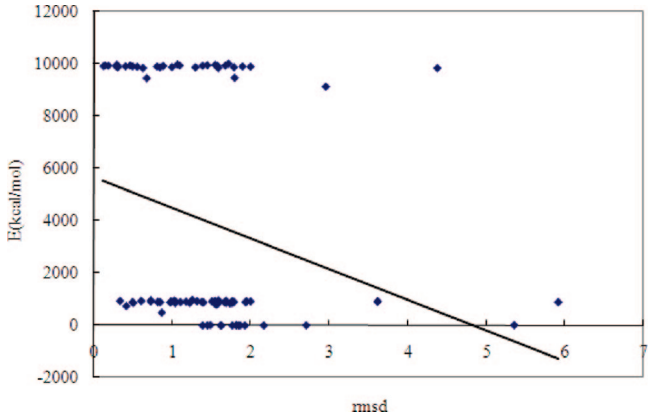


Figure 5. The docked energy for each of the 92 complexes (Table 1) docked by ADDock is plotted against the corresponding *rmsd* computed.

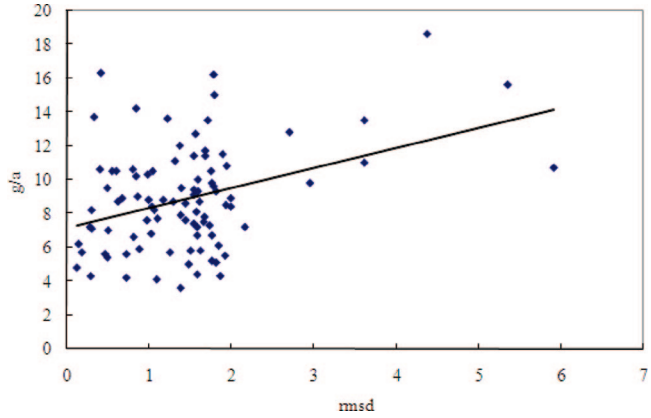


Figure 6. The **g/a** ratio (Table 1) is computed as the ratio between the number of cavity grid points **g** and the number of atoms of each ligand **a** counted and is plotted against the corresponding *rmsd* computed.

interaction energies computed by ADDock. These demonstrated the ability of ADDock to separate the active from inactive compounds during a docking process.

METHODS

Preparing Molecules for Docking. The coordinates of 92 receptor−ligand complexes docked by ADDock were downloaded from the PDB.[33] All of these receptor−ligand complexes have been docked by other groups using other docking methods.[20,34] Hydrogen atoms for both receptors and extracted ligand molecules were added using the SYBYL 7.3 program.[35] All the crystal waters, metal ions, or cofactor molecules identified inside the receptor active sites were kept. Only one set of receptor−ligand coordinates was selected for docking if there were multiple sets present in a PDB file. The coordinates of separated receptor and ligand with hydrogen atoms added were individually saved as a ligand. mol2 file.

Counting Molecular Topology for Ligands. To count the molecular topology or make a bond connection list for an anchor selected ligand, we treated the molecule as a tree in which atoms were nodes and chemical bonds were edges.[36] Each pair of nodes was connected by an edge. A walk of a tree was an alternating sequence of nodes and edges that begins and ends with nodes. A walk became a path if one node was distinct among all the nodes traversed. The interconnections or the topology of each tree was represented by a connection list.[37] The connection list gave a list of atoms which were linked in each path of a tree, and each path in the tree was rooted on the same anchor. All the bond lengths, bond angles, and dihedral angles on a path were determined based on the coordinates of original crystal structures and stored with the topology counted. Rotatable bonds were determined using the following criteria:[38] acyclic single bonds were rotatable unless they were amide bonds, bonds between two unsaturated atom centers, bonds connected with terminal atoms, bonds of sulfonamides, or bonds between hindered rings.

Converting Ligand Topology To Coordinate Using a Kinematic Model. A molecular kinematic model described in the literature[39,40] was employed to convert ligand topology to coordinate. In this model, a local coordinate frame was attached at the common root of each path of a tree namely, the anchor. If a bond b_i followed a bond b_{i-1} in the path, then the coordinate frame of b_i was related to that of b_{i-1} by the homogeneous (both rotation and translation were performed) transformation described as follows

$$T_i = \begin{bmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i c\alpha_{i-1} & c\theta_i c\alpha_{i-1} & -s\alpha_{i-1} & -l_i s\alpha_{i-1} \\ s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & c\alpha_{i-1} & l_i c\alpha_{i-1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where θ_i represents the dihedral angle, l_i represents bond length, α_{i-1} is the bond angle between bond l_{i-1} and l_i, $c\theta_i$ is $cos\theta_i$, and $s\theta_i$ is $sin\theta_i$. The position of any node (atom) in a path was determined by chaining matrices of the form (1). For example, the *xyz* coordinate of a particular atom a_i in a path with a sequence of bonds (edges) b_i, b_{i-1},..., b_1 from the anchor, a_{anchor}, was computed as follows:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = T_1 T_2 \cdots T_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (2)$$

The anchor was assumed to be located at the origin of the global coordinate system of a receptor−ligand complex. Therefore, a local coordinate frame attached to the anchor could be rotated with respect to the global coordinate frame, or the orientation of the local coordinate frame with respect to the global coordinate frame was computed as follows

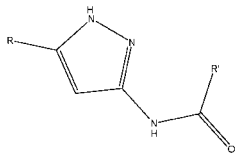
$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = T_R(q_1, q_2, q_3, q_4) T_1 T_2 \cdots T_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (3)$$

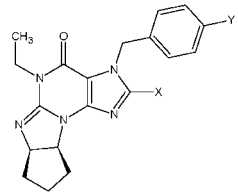
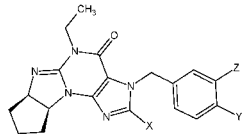
where q_1, q_2, q_3, and q_4 were the four quaternions,[41] and the corresponding rotational matrix T_R was computed as follows[42]

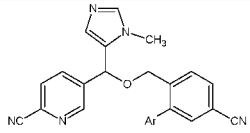
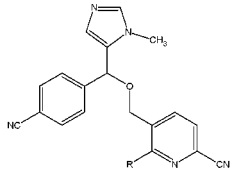
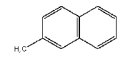
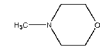
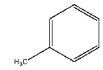
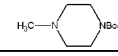
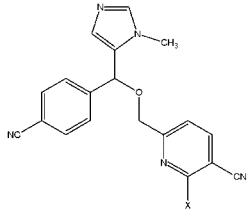
$$T_R = \frac{1}{p} \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1 q_2 - q_3 q_4) & 2(q_1 q_3 + q_2 q_4) & 0 \\ 2(q_1 q_2 + q_3 q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2 q_3 - q_1 q_4) & 0 \\ 2(q_1 q_3 - q_2 q_4) & 2(q_2 q_3 + q_1 q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

where $p = q_1^2 + q_2^2 + q_3^2 + q_4^2$ is the norm of the vector of four quaternions. The rotational matrix T_R was scaled by the norm

Table 3. Structures of the Ten Active Ligands of 1vyz, 1udt, and 1ni1 Docked by ADDock

Core structure of active 1vyz ligands	Ligand name	IC_{50}(nM)	R'	R
	5	34	4-Br-C_6H_4-	cyclopropyl
	20	50	4-Br-C_6H_4-	cyclopentyl
	9	68	4-$CONH_2$-C_6H_4-	cyclopropyl
	18	83	3,4-diCl-C_6H_4-	cyclopropyl
	19	90	4-Br-C_6H_4-	cyclobutyl
	8	165	4-COOH-C_6H_4-	cyclopropyl
	25	200	4-Br-C_6H_4-	isopropyl
	3	224	propyl	cyclopropyl
	12	600	3-OMe-C_6H_4-	cyclopropyl
	7	705	4-OMe-C_6H_4-	cyclopropyl

Core structure of active 1udt ligands	Ligand name	IC_{50}(nM)	X	Y	Z
	34	0.9	OEt	OH	-
	33	1.8	SEt	OH	-
	32	2.1	$CONH_2$	OH	-
	29	4.0	$CONH_2$	MeO	-
	25	18.0	OEt	MeO	-
	30	22.0	NH_2	OH	-
	24	48.0	CF_3	MeO	-
	23	55.0	H	MeO	-
	36	6.5	H	MeO	Br
	37	8.3	H	MeO	Cl

Core structure of active 1ni1 ligands	Ligand name	IC_{50}(nM)	R	Ar	X
	23c	0.4	-	3-OCF_3-Ph	-
	26j	0.16		-	-
	26i	1.9		-	-
	26a	3.9		-	-
	26h	44		-	-
	24c	8.2	-	-	2-naphthyl
	24a	12	-	-	3-Cl-ph
	24e	25	-	-	4-Cl-ph
	24b	58	-	-	4-methyl-piperazin-1-yl
	24f	68	-	-	3-cyano-ph

of the quaternion vector to keep Det(T_R) = 1. The scaling was required to ensure that no correlation between successive rotational matrices was generated. Procedures for generating the random unit-quaternion vectors were detailed elsewhere by other groups.[43,44]

Evolving Parameters Using an Evolutionary Algorithm. The anchor would be located at the origin of the local coordinate frame (the anchor-based frame) once the coordinate of entire ligand was generated from its topology. We then converted the coordinate frame of receptor

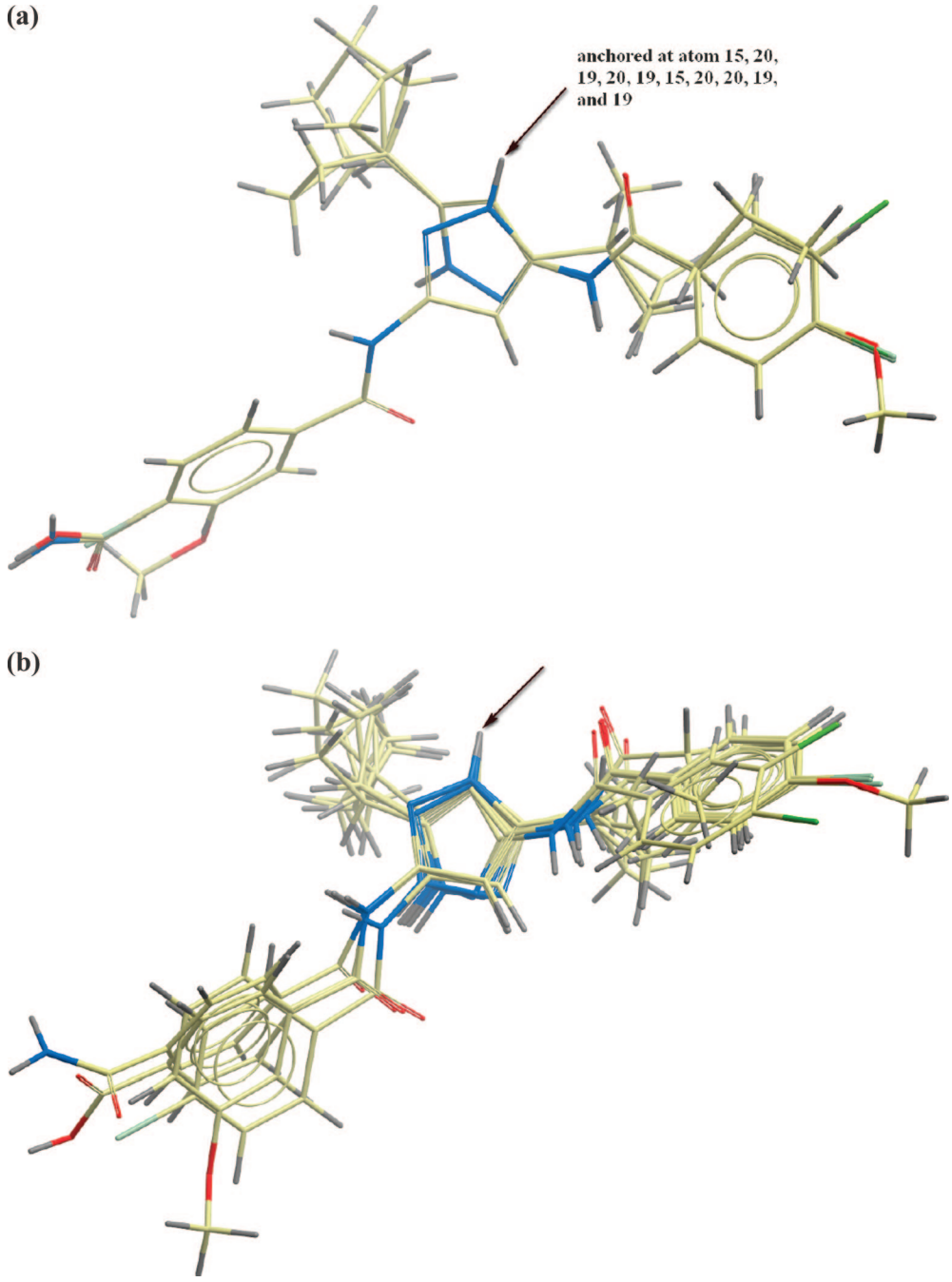


Figure 7. a. Before docking by ADDock, the initial structures prepared for 10 inhibitors of CDK2 (1vyz) of set (i) (Table 3) are all anchored at a hydrogen atom on the pyrazole ring chosen. These are atoms 19, 15, 20, 19, 15, 20, 20, 19, 20, and 19 for inhibitors **5**, **20**, **9**, **18**, **19**, **8**, **25**, **3**, **12**, and **17** (marked by an arrow), respectively. **b.** After docking by ADDock, the docked structures of these 10 inhibitors are still anchored at the anchor atoms chosen (see part **a**) as marked by an arrow.

to the anchor-based frame using a procedure described in the literature.[45] Each of the four quaternions and rotatable bonds identified was treated as a variable and encoded as a 12 bits binary string in an artificial chromosome. The searching range of all the rotatable bonds was set between 180 and −180. There was a certain population of chromosomes generated within each generation, and two out of them were selected by a roulette wheel mechanism[46] for crossover operation. We proceeded with two-points crossover on the two chromosomes randomly selected by first picking out two crossover sites randomly. While bits on the left of the first crossover site were unchanged, those on the right of the first crossover site of the two chromosomes were exchanged. Similarly, bits on the left of the second crossover site were unchanged, while those on the right were exchanged in the second crossover process. The entire population of chromosomes generated was also subjected to a mutation process where every bit in every chromosome was visited, and occasionally bit *1* was flipped to bit *0* or vice versa. The probability of mutation P_m was set as 0.2 throughout the current studies. The crossovered and mutated chromosomes were used to replace the original ones in the same population of chromosomes. The fitness function for each chromosome in each population was calculated as the inverse of the docked energy (interaction energy plus clash penalty) computed. To find the global minimum, the fittest

Table 4. Energy Components of the Docked Energy of 10 Active Plus 35 Nonactive Ligands Docked into Receptor 1vyz by ADDock

ligand	IC_{50}(nM)	atom no.[a]	tol-E[b]	rl-ster	ll-tor	ll-Hb	ll-ster	ll-clash	rl-Hb	rl-ele	rl-clash
20	50	36	9889.6	−120.9	1.1	0.	3.1	0.	6.2	0.	10000.
12	600	34	9892.9	−112.7	1.4	0.	0.8	0.	3.4	0.	10000.
7	705	34	9897.4	−111.1	0.1	0.	2.5	0.	5.8	0.	10000.
8	165	33	9896.5	−109.4	0.	0.	3.3	0.	2.7	0.	10000.
5	34	30	9898.9	−109.0	1.7	0.	1.9	0.	4.3	0.	10000.
19	90	33	9901.0	−108.6	1.3	0.	5.1	0.	3.1	0.	10000.
9	68	34	9902.1	−105.3	4.2	0.	−0.3	0.	3.4	0.	10000.
25	200	32	9907.7	−102.7	2.6	0.	4.2	0.	3.6	0.	10000.
18	83	30	9902.3	−102.6	0.0	0.	2.0	0.	2.9	0.	10000.
3	224	29	914.9	−88.8	0.8	0.	−2.8	1000.	5.7	0.	0.
Des	-	40	913.7	−88.8	0.	0.	7.0	1000.	−4.5	0.	0.
Pp1	-	43	919.2	−83.8	3.2	0.	8.5	1000.	−8.7	0.	0.
Agb	-	48	960.0	−82.8	4.2	0.	40.7	1000.	−2.0	0.	0.
2am	-	35	863.4	−82.4	0.	−5.7	−0.5	1000.	−8.4	−39.6	0.
Clm	-	31	914.6	−81.7	0.	−7.1	17.8	1000.	−6.1	−8.4	0.
Uvc	-	38	895.2	−74.4	0.	−11.0	15.6	1000.	−19.0	−16.0	0.
Pfz	-	44	931.9	−71.7	0.	−0.7	7.5	1000.	−0.1	−3.2	0.
Dan	-	35	882.2	−70.4	0.	−9.5	6.6	1000.	−4.8	−39.6	0.
Ptp	-	38	859.4	−69.2	0.	0.	25.3	1000.	−11.5	−85.2	0.
Gy	-	31	928.1	−69.0	0.	−2.5	12.2	1000.	−2.5	−10.1	0.
Fmc	-	32	894.1	−68.2	0.	−8.9	5.5	1000.	−6.8	−27.5	0.
Tk4	-	34	929.5	−67.9	0.	−2.0	6.2	1000.	−6.8	0.	0.
Otg	-	39	9919.2	−67.0	0.	−12.5	18.6	0.	−5.4	−14.5	10000.
Tpm	-	30	952.7	−66.9	3.0	0.	16.6	1000.	0.	0.	0.
C2p	-	33	903.1	−66.8	0.	−9.9	20.5	1000.	−7.1	−33.6	0.
Fog	-	31	933.7	−66.6	0.	−4.2	11.4	1000.	0.	−6.9	0.
Ufp	-	31	875.0	−60.7	0.	−2.5	16.0	1000.	−8.9	−68.9	0.
N3t	-	42	920.1	−57.0	3.3	0.	4.5	1000.	−1.0	−29.6	0.
Vk	-	41	945.9	−52.9	0.	−2.5	3.9	1000.	0.9	−3.5	0.
Spc	-	39	933.7	−49.0	0.	−2.2	−0.9	1000.	−1.3	−12.9	0.
Cpm	-	36	949.6	−48.7	0.	−2.5	10.3	1000.	−6.3	−3.2	0.
Amp	-	35	862.9	−40.2	0.	−9.3	0.4	1000.	−10.0	−78.1	0.
T44	-	34	−1.0	−34.7	0.	0.	30.4	0.	0.	3.3	0.
Fds	-	38	−0.7	−33.1	0.	0.	36.0	0.	−1.0	−2.6	0.
Hep	-	41	−1.0	−32.9	5.4	0.	35.7	0.	0.	−9.1	0.
Nab	-	34	−0.9	−32.4	0.	0.	41.3	0.	−5.2	−4.6	0.
Nev	-	34	−1.0	−28.8	2.5	0.	22.3	0.	3.1	0.	0.
Sc4	-	45	−0.3	−25.7	12.7	0.	19.7	0.	0.	−7.0	0.
Btn	-	32	−1.0	−20.0	0.	0.	20.5	0.	−2.3	0.8	0.
Gtb	-	48	1032.4	−18.9	17.2	0.	29.9	1000.	0.3	3.9	0.
Gas	-	45	−0.5	−16.6	8.7	0.	7.9	0.	−0.5	0.	0.
S57	-	42	−1.0	−16.4	15.6	0.	6.8	0.	0.	−6.9	0.
Pvb	-	30	−1.0	−6.8	2.4	0.	3.4	0.	0.	0.	0.
W33	-	45	−1.0	−6.4	10.9	0.	−2.0	0.	−0.7	−2.8	0.
Aah	-	43	39.8	14.8	0.	−0.9	21.0	0.	−1.6	6.4	0.

[a] Number of atoms of each ligand docked. [b] All the energy components are in units of kcal/mol, tol-E expresses the total docked energy, rl-ster expresses the receptor−ligand steric interaction energy, ll-tor expresses the ligand internal torsion energy, ll-Hb expresses the ligand internal hydrogen-bond energy, ll-ster expresses the ligand internal steric interaction energy, ll-clash expresses the ligand internal clash energy, rl-Hb expresses the receptor−ligand hydrogen bond energy, rl-ele expresses the receptor−ligand electrostatic interaction energy, and rl-clash expresses the receptor−ligand clash energy.

chromosome in the entire population was searched and propagated to the next generation.

Scoring the Interaction between Ligand and Receptor. Each ligand conformation was generated during the evolutionary process using the values of four quaternions and all the rotatable bonds generated. The scoring function used by ADDock was the PLP scoring function adopted from MolDock which was originally proposed by Gehlhaar et al.[32,47] and later extended in GEMDOCK and modified extensively in MolDock especially for the part of estimating the hydrogen-bond interactions. The interaction energy between ligand and receptor was scored by E_{inter} and E_{intra}, where the former was the intermolecular interaction energy between ligand and receptor and the latter was the intramolecular interaction energy between ligand atoms. E_{inter} consists of E_{PLP}, a piecewise linear potential described both in GEMDOCK and MolDock, and an electrostatic interaction energy term between charged ligand and receptor atoms. Ligand atoms and selected receptor atoms inside a docking box were classified as donor, acceptor, both (donor and acceptor) for formation of hydrogen bonds, or nonpolar types according to those described in GEMDOCK and MolDock. For both receptor−ligand and ligand−ligand interactions, only 7 out of 16 (4 × 4) matching types namely, donor−acceptor, donor−both, acceptor−donor, acceptor−both, both−donor, both−acceptor, both−both were considered as hydrogen-bond interactions, while the rest of them were steric interactions. The SYBYL atomic types were converted to the atomic types described in GEMDOCK and MolDock to compute the formal charge for each atom. The dielectric constant was fixed as 4. The formal charges for crystal waters or metal ions identified inside the docking box were also assigned according to those described in GEMDOCK and MolDock.

There were two different parameter sets used in function E_{PLP}[19,20,32] to evaluate both the steric (van der Waals) and

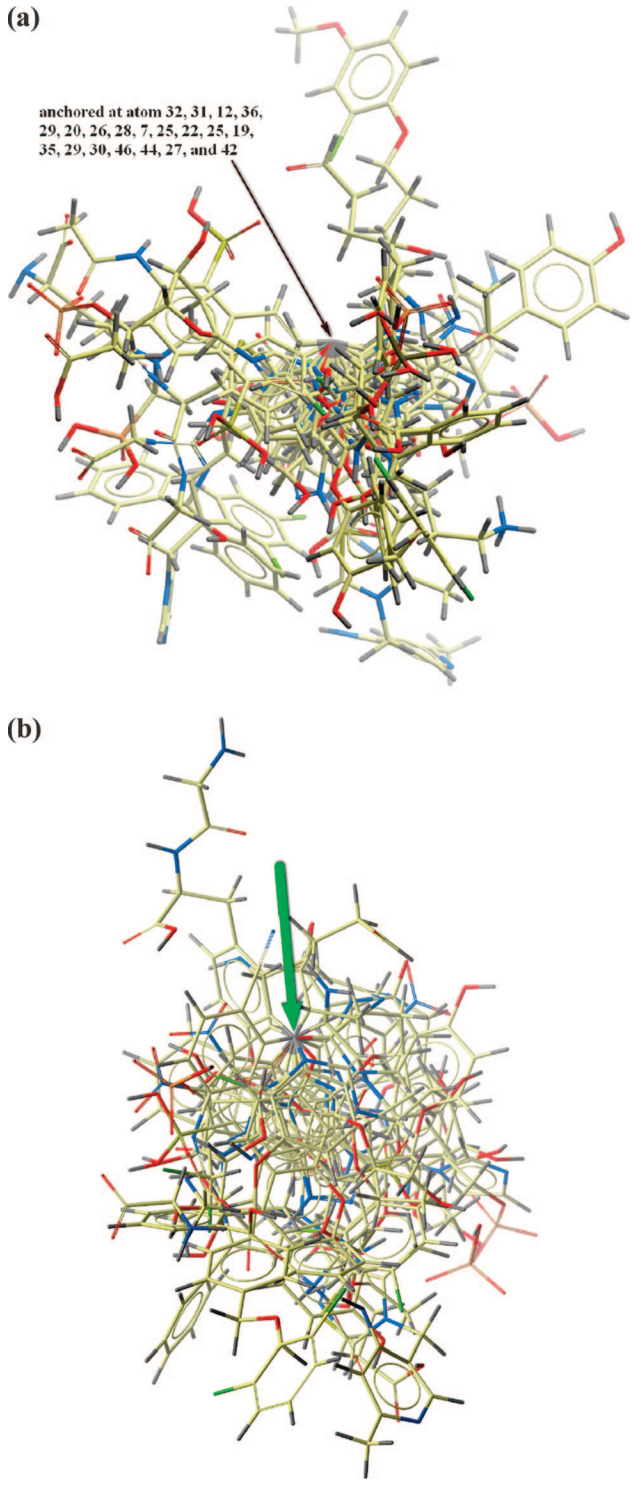


Figure 8. a. Before docking by ADDock into the CDK2 active site, the structures prepared for 20 inactive compounds namely, Des, Clm, Uvc, Pfz, Gy, Fmc, Tk4, Tpm, C2p, Fog, Ufp, N3t, Vk, Amp, Fds, Sc4, Gtb, Gas, Aah, and S57 chosen from Table 1 are all anchored at the same anchors (marked by an arrow) chosen for docking these crystal ligands into their native receptors (Table 1). The coordinate of each inactive compound rooted on each anchor chosen is brought by ADDock to the same origin of the anchor chosen for the docking crystal ligand N5b into the CDK2 active site. The corresponding anchors chosen for these inactive compounds are atoms 32, 31, 12, 36, 29, 20, 26, 28, 7, 25, 22, 25, 19, 35, 29, 30, 46, 44, 27, and 42, respectively (Table 1). For brevity, only 20 structures of the inactive compounds of set (i) are selected for such a presentation. **b**. After docking by ADDock, the docked structures of these 20 inactive compounds are still anchored at the anchor atoms chosen (see part **a**) as marked by a green arrow.

hydrogen-bond interaction energies between atoms. While the steric interaction depends only on the distance, the hydrogen-bond part depends on both the distance and directionality (geometry) of hydrogen bonding. While the distance-dependent parameter sets were adopted from GEMDOCK and MolDock, the strength of formation of hydrogen-bond was scaled by some geometry scaling factors adopted from MolDock. Once a potential hydrogen bond was identified, atoms playing the roles of donor (D), acceptor (A), and acceptor antecedent (AA) were identified. The coordinates of these atoms were used to compute the geometry scaling factors defined in MolDock. The calculation of these factors was omitted if it was impossible to calculate one of these hydrogen-bond angles as described in MolDock. The ligand intramolecular interaction energy E_{intra} was also composed of a E_{PLP} term plus a torsional energy term for all the rotatable bonds identified.[19,20] The E_{PLP} interaction energy between any atomic pairs of ligand was computed by a double summation except for atoms that were connected by two bonds or less.[19,20] The averaged parameter sets for sp^2-sp^3 and sp^3-sp^3 dihedral types described in MolDock were employed for the computation of torsional energy. The clash penalty E_{clash} was defined as 10000 kcal/mol if the distance between any receptor and ligand atom was shorter than 2.0 Å or as 1000 kcal/mol if the distance between any two ligand atoms that were separated by more than two bonds was shorter than 2.0 Å.

Estimating the Cavity Size for Ligand and Comparing the Docked Structures. We defined a docking box for each receptor−ligand complex by recruiting any receptor atom including crystal waters and metal ions that were within a distance of 8 to 13 Å with any ligand atom into the box such that the total number of receptor atoms included was around 900. We then constructed a grid on the docking box using a grid space of 1 Å for estimating the size of cavity. The grid was centered at the center of mass of ligand, and then the *xyz* coordinate extremes of ligand were identified. The number of grid points inside a ligand cavity was counted by excluding grid points that were within the van der Waals envelope of the receptor atoms of the docking box or that their distances with the center of mass of ligand were greater than the *xyz* coordinate extremes of ligand. The atomic radii of the Gaff force field parameter set of the Amber package[48] were used in this calculation. To compare the docked ligand conformation by ADDock with that of the original ligand extracted from a PDB file, we computed the root-mean-square-deviation (*rmsd*) between the coordinates of them as follows

$$rmsd = \left(\frac{\sum_i^N ||x_i - y_i||^2}{N} \right)^{\frac{1}{2}} \quad (5)$$

where x_i and y_i represents the coordinates of docked and original ligand respectively, and N was the total number of ligand atoms including all the hydrogen atoms added.

RESULTS AND DISCUSSION

Overall Docking Performance by ADDock. The overall docking performance on 92 ligand−receptor complexes by ADDock is presented in Table 1. The docking accuracy of ADDock on these complexes is 91.3% since the *rmsd* values

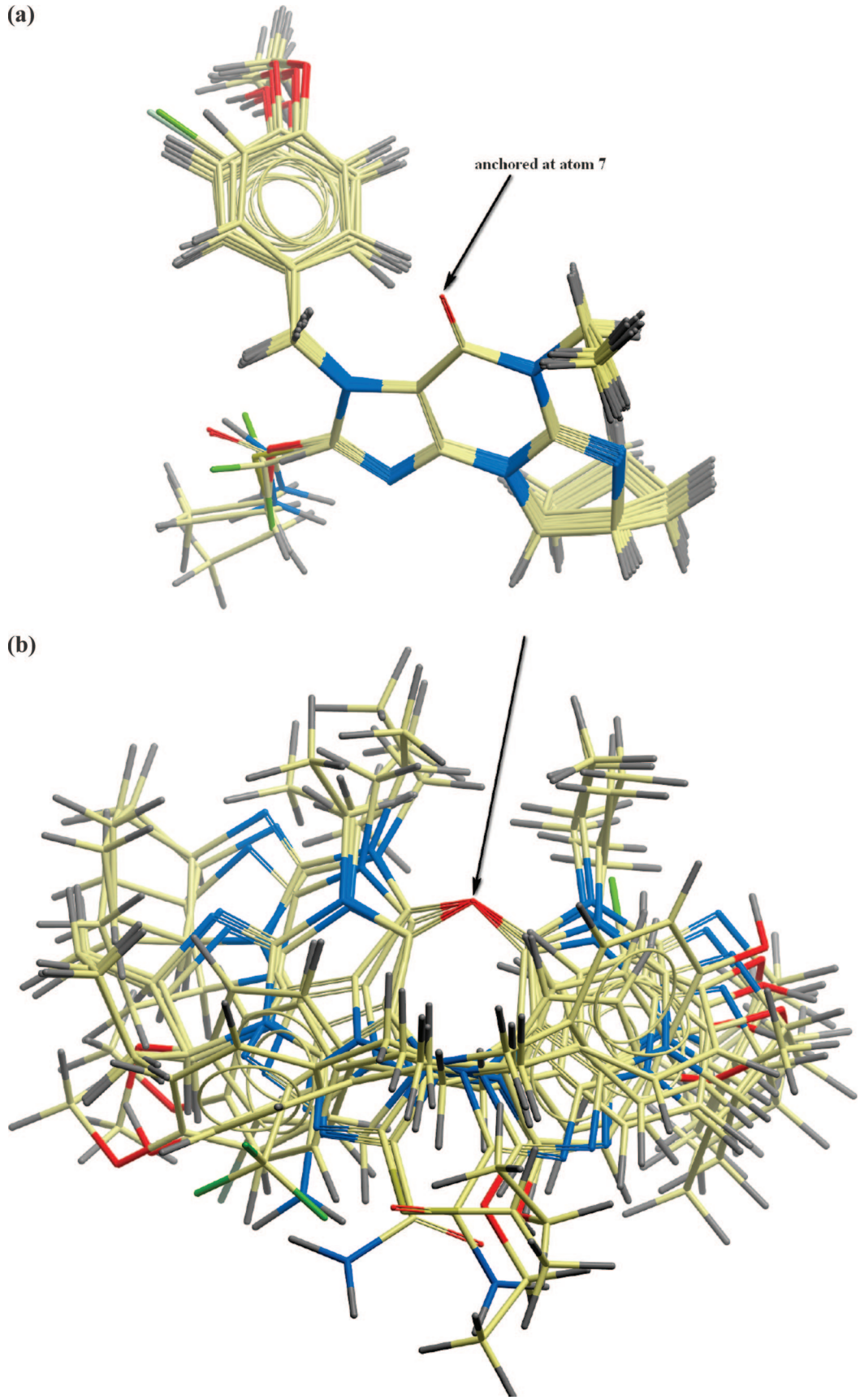


Figure 9. a. Before docking by ADDock, the initial structures prepared for 10 inhibitors of PDE5 (1udt) of set (ii) (Table 3) are all anchored at atom 7, the oxygen atom on the cyclic guanine ring (Table 3). **b**. After docking by ADDock, the docked structures of these 10 inhibitors are still anchored at the anchor atoms chosen (see Figure 9a) as marked by an arrow.

computed for 84 out of 92 complexes are below 2.0 Å (Table 1). The averaged *rmsd* computed for the 92 complexes is 0.84 Å. The largest (Fk5) and smallest (Oaa) ligand docked consisted of 126 and 11 atoms, respectively (Table 1). The

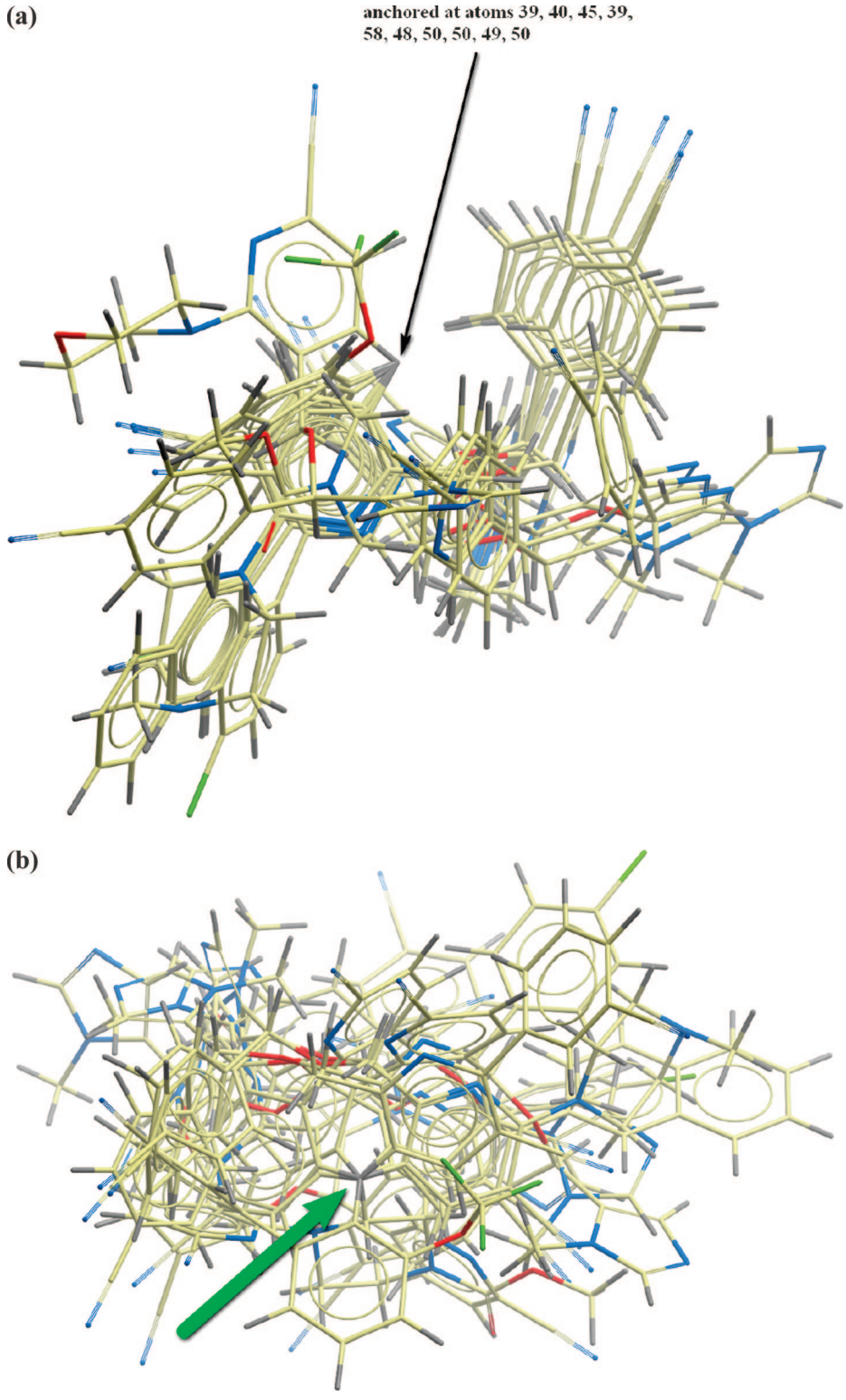


Figure 10. a. Before docking by ADDock, the initial structures prepared for 10 inhibitors of FTase (1ni1) of set (iii) (Table 3) are all anchored at a hydrogen atom on the cyanopyridine ring namely, atoms 48, 50, 50, 49, 50, and 45 for inhibitors **24c**, **24a**, **24e**, **24b**, **24f**, and **26i** (Table 3), or a hydrogen atom on the phenyl ring of group Ar (atom 39), biphenyl ring of group R (atom 40), phenyl ring of group R (atom 39), and ring structure of group R (atom 58), for inhibitors **23c**, **26j**, **26a**, and **26h** (Table 3), respectively. **b**. After docking by ADDock, the docked structures of these 10 inhibitors are still anchored at the anchor atoms chosen (see part **a**) as marked by a green arrow.

best *rmsd* obtained by ADDock is 0.12 Å on complex Ara (1abc), while the worst one obtained is 5.91 Å on complex Mtx (4dfr) (Table 1). In fact, the *rmsd* computed for 29 or 31.5% complexes are below 1.0 and are between 1.0 and 2.0 Å for 55 or 59.8% complexes (Table 1), respectively. The largest number of rotatable bonds identified among the 84 accurately docked complexes by ADDock is 19 (for ligand C60) while that for the 8 poorly docked complexes is 10 (for ligand Gtb) (Table 1). The *rmsd* obtained for Fk5 and Oaa, the largest (126 atoms) and smallest (11 atoms) ligands

Table 5. Energy Components of the Docked Energy of 10 Active Plus 33 Nonactive Ligands Docked into Receptor 1udt by ADDock

ligand	IC_{50}(nM)	atom no.[a]	tol-E[b]	rl-ster	ll-tor	ll-Hb	ll-ster	ll-clash	rl-Hb	rl-ele	rl-clash
Sb3	-	68	9900.5	−131.3	7.0	0.	42.7	0.	0.	−17.9	10000.
36	6.5	50	867.1	−121.6	2.4	0.	10.4	1000.	0.	−24.1	0.
25	18.0	57	9847.6	−121.5	8.9	0.	−2.3	0.	0.	−37.6	10000.
Gel	-	72	9801.1	−119.5	22.3	0.	−6.7	0.	−3.1	−91.9	10000.
Hem	-	75	9861.8	−111.3	4.6	0.	−29.3	0.	−2.2	0.	10000.
Bzt	-	77	864.3	−111.1	0.	0.	−7.0	1000.	0.	−17.6	0.
24	48.0	53	897.4	−109.6	4.1	0.	9.5	1000.	0.	−6.6	0.
32	2.1	51	891.0	−109.1	0.	0.	2.7	1000.	−2.6	0.	0.
29	4.0	54	9857.5	−108.0	3.3	0.	−0.4	0.	−4.3	−33.0	10000.
33	1.8	54	918.3	−106.9	0.	0.	25.2	1000.	0.	0.	0.
Mtx	-	57	881.4	−106.8	0.	−3.2	17.0	1000.	−5.1	−20.5	0.
Blv	-	77	9885.5	−106.5	19.0	0.	−19.1	0.	−7.9	0.	10000.
Psi	-	86	9835.1	−105.8	0.	−5.0	−2.6	0.	−11.8	−39.8	10000.
34	0.9	54	892.2	−105.5	0.	0.	10.7	1000.	0.	−13.0	0.
30	22.0	49	911.4	−102.4	0.	0.	13.8	1000.	0.	0.	0.
Ndp	-	75	900.0	−100.4	0.	−11.5	15.6	1000.	−3.6	0.	0.
23	55.0	50	908.3	−97.9	2.6	0.	9.3	1000.	0.	−5.7	0.
37	8.3	50	931.8	−88.7	3.5	0.	22.8	1000.	0.	−5.9	0.
Str	-	53	933.3	−84.8	0.7	0.	17.5	1000.	0.	0.	0.
961	-	53	9890.7	−84.6	0.	−3.3	14.9	0.	−10.7	−25.6	10000.
Fmn	-	50	891.9	−82.8	0.	−10.1	17.4	1000.	−10.5	−22.1	0.
Hds	-	53	9907.1	−81.5	0.3	0.	−2.1	0.	−9.5	0.	10000.
E64	-	54	869.7	−77.8	0.	−3.5	−6.0	1000.	−11.4	−31.6	0.
Oht	-	58	937.8	−77.6	0.	0.	24.5	1000.	0.	−9.2	0.
Dhg	-	69	1036.7	−77.3	0.	−7.2	172.3	1000.	−2.8	−48.3	0.
Xk2	-	84	9934.1	−76.0	0.	−2.5	24.9	0.	−2.5	−9.8	10000.
Apr	-	60	823.1	−75.1	0.	−22.4	18.5	1000.	−14.4	−83.5	0.
Dgx	-	81	884.2	−74.8	0.	−2.5	4.4	1000.	−6.0	−36.8	0.
Stu	-	60	9923.5	−74.4	1.4	0.	−3.5	0.	0.	0.	10000.
Ae2	-	51	9923.1	−73.9	0.	0.	10.8	0.	0.	−13.8	10000.
Mid	-	66	6.3	−71.0	10.6	0.	66.9	0.	−0.3	0.	0.
Npp	-	55	973.6	−70.4	6.8	0.	24.5	1000.	0.	12.7	0.
Ola	-	54	896.4	−66.2	0.	0.	−5.8	1000.	−5.0	−26.6	0.
Rea	-	50	940.7	−58.3	0.	0.	5.1	1000.	0.	−6.2	0.
Sph	-	60	923.8	−56.4	0.	−2.5	−4.4	1000.	−1.1	−11.8	0.
Coa	-	81	917.6	−52.6	0.	−8.7	−16.8	1000.	−4.3	0.	0.
Gep	-	52	−1.0	−44.1	13.3	0.	14.0	0.	−4.1	20.0	0.
815	-	50	−1.0	−41.1	5.5	0.	34.6	0.	0.	0.	0.
S5h	-	68	7.0	−38.1	0.	0.	49.0	0.	0.	−4.0	0.
Nad	-	70	1363.0	−37.6	0.	−9.3	414.8	1000.	−4.9	0.	0.
Fad	-	83	9868.5	−34.1	0.	−16.0	5.0	0.	−6.4	−80.0	10000.
Vac	-	86	10012.5	57.0	0.	−5.0	−3.5	0.	−0.4	−35.5	10000.
Rtl	-	51	9784.0	72.1	0.	0.	0.4	0.	7.6	−296.1	10000.

[a] Number of atoms of each ligand docked. [b] All the energy components are in units of kcal/mol, tol-E expresses the total docked energy, rl-ster expresses the receptor−ligand steric interaction energy, ll-tor expresses the ligand internal torsion energy, ll-Hb expresses the ligand internal hydrogen-bond energy, ll-ster expresses the ligand internal steric interaction energy, ll-clash expresses the ligand internal clash energy, rl-Hb expresses the receptor−ligand hydrogen bond energy, rl-ele expresses the receptor−ligand electrostatic interaction energy, and rl-clash expresses the receptor−ligand clash energy.

docked by ADDock, is 1.38 and 1.56 Å, respectively (Table 1). ADDock does not need to predefine the docking poses for a ligand since the intact rather than fragmental ligand structure is generated during the docking process. No energy minimization is required either during the predocking phase, namely the generation of ligand structure, or the after-docking phase where the docked conformation is produced. Therefore, the scoring scheme used by ADDock directly determines the docked poses or conformation of a ligand during the docking process. ADDock requires only a random number and selection of a ligand atom as an anchor to initiate the docking process. Any single-edge connected node or terminal ligand atom can be selected as an anchor to prepare the topology for a ligand. The anchors chosen for each docked ligand shown in Table 1 are either the terminal hydrogen or oxygen atoms of a carboxylic group. The structure of entire ligand is treated as a topological tree and rooted on the anchor selected. The topology created also includes the number of rotatable bonds identified. Being randomly varied between 180 and −180 by the GA, these rotatable bonds generated are plugged into eq 3, the kinematic equation, to compute the coordinate for a ligand for docking. ADDock sets no limitation on the number of rotatable bonds allowed for a ligand.

A comparison for the docking performance on 48 complexes by ADDock with those by MolDock, Glide,[49] GOLD, FlexX,[50] and Surflex[51] is shown in Table 2. The docking results for these 48 complexes by the five aforementioned methods are adopted from MolDock. Apparently, with a docking accuracy of 83.3 (40/48), 83.3, and 81.3% (39/48) obtained, ADDock, MolDock, and Glide give the best docking results over the other methods compared on these complexes. The docking accuracy by the other methods compared is ranked decreasingly as follows: 73.8% (31/42) for GOLD, 72.9% (35/48) for Surflex, and 51.1% (24/47) for FlexX. The best *rmsd* obtained by ADDock is 0.12 for

Table 6. Energy Components of the Docked Energy of 10 Active Plus 33 Nonactive Ligands Docked into Receptor 1ni1 by ADDock

ligand	IC_{50}(nM)	atom no.[a]	tol-E[b]	rl-ster	ll-tor	ll-Hb	ll-ster	ll-clash	rl-Hb	rl-ele	rl-clash
26h	44	60	26.3	−158.7	11.1	0.	173.9	0.	0.	0.	0.
24b	58	57	54.2	−141.1	3.8	0.	191.6	0.	0.	0.	0.
24f	68	51	3.0	−103.3	9.3	0.	97.0	0.	0.	0.	0.
24c	8.2	56	9.5	−102.8	5.8	0.	106.5	0.	0.	0.	0.
26i	1.9	53	24.6	−99.7	5.2	0.	119.0	0.	0.	0.	0.
24e	25	50	−1.0	−93.2	6.1	0.	86.1	0.	0.	0.	0.
24a	12	50	−1.0	−88.7	5.0	0.	82.6	0.	0.	0.	0.
Fmn	-	50	706.5	−86.6	0.	−13.9	20.4	1000.	−8.7	−204.8	0.
Bzt	-	77	837.3	−84.4	0.	0.	−20.6	1000.	−2.8	−54.8	0.
26j	0.16	56	−0.8	−84.3	5.3	0.	78.3	0.	0.	0.	0.
Blv	-	77	927.6	−83.6	17.8	0.	−1.5	1000.	−5.1	0.	0.
Apr	-	60	607.4	−81.5	0.	−15.6	16.9	1000.	−12.6	−299.8	0.
Hem	-	75	881.9	−79.9	3.6	0.	−35.5	1000.	−6.4	0.	0.
Rea	-	50	881.3	−73.9	0.	0.	3.0	1000.	−4.8	−43.0	0.
Sph	-	60	890.7	−72.9	0.	−5.6	−1.6	1000.	−0.5	−28.8	0.
Gep	-	52	13.8	−72.8	9.4	0.	15.9	0.	−5.8	67.1	0.
Nad	-	70	877.5	−72.4	0.	−23.4	−26.0	1000.	−0.6	0.	0.
Gel	-	72	698.6	−70.3	20.8	0.	12.3	1000.	−9.1	−255.1	0.
26a	3.9	50	−1.0	−68.4	5.1	0.	62.4	0.	0.	0.	0.
Mid	-	66	−0.1	−67.6	8.8	0.	61.2	0.	−2.5	0.	0.
23c	0.40	54	−1.0	−66.4	1.5	0.	63.9	0.	0.	0.	0.
Hds	-	53	9931.9	−64.8	0.3	0.	−2.1	0.	−1.5	0.	10000.
Mtx	-	57	781.4	−62.1	0.	−7.2	13.1	1000.	−9.6	−152.9	0.
Psi	-	86	886.5	−62.0	0.	−4.9	−5.0	1000.	−2.3	−39.2	0.
Dgx	-	81	783.5	−60.4	0.	−2.5	13.7	1000.	−2.3	−164.9	0.
Dhg	-	69	710.7	−58.2	0.	−2.5	−16.4	1000.	−3.8	−208.5	0.
Ae2	-	51	9919.3	−57.4	0.	0.	10.8	0.	−2.1	−32.0	10000.
Ndp	-	75	909.5	−54.8	0.	−22.2	−7.1	1000.	−6.4	0.	0.
Sb3	-	68	−1.0	−47.0	13.8	0.	58.5	0.	0.	−26.2	0.
Fad	-	83	9772.9	−39.6	0.	−16.0	5.0	0.	1.5	−178.1	10000.
Xk2	-	84	0.9	−30.7	0.	−2.5	80.9	0.	0.	−46.8	0.
Npp	-	55	−1.0	−28.0	4.5	0.	7.8	0.	−0.5	15.2	0.
Str	-	53	−1.0	−19.0	3.0	0.	15.0	0.	0.	0.	0.
815	-	50	−1.0	−12.5	6.1	0.	5.3	0.	0.	0.	0.
961	-	53	−1.0	−12.4	0.	−3.7	35.7	0.	0.	−20.6	0.
E64	-	54	0.4	−10.5	0.	−5.0	41.7	0.	0.	−25.8	0.
Stu	-	60	−1.0	−7.2	5.2	0.	1.0	0.	0.	0.	0.
Coa	-	81	−1.0	−5.5	0.	−7.1	14.2	0.	−2.5	0.	0.
S5h	-	68	−1.0	−5.4	0.	0.	23.2	0.	0.	−18.9	0.
Oht	-	58	−0.3	9.7	0.	0.	20.5	0.	0.	−30.5	0.
Ola	-	54	1.0	32.1	0.	0.	−6.9	0.	−2.2	−22.0	0.
Vac	-	86	9591.1	324.9	0.	−5.0	−3.5	0.	2.2	−727.3	10000.
Rtl	-	51	9215.0	334.0	0.	0.	0.4	0.	15.1	−1134.6	10000.

[a] Number of atoms of each ligand docked. [b] All the energy components are in units of kcal/mol, tol-E expresses the total docked energy, rl-ster expresses the receptor−ligand steric interaction energy, ll-tor expresses the ligand internal torsion energy, ll-Hb expresses the ligand internal hydrogen-bond energy, ll-ster expresses the ligand internal steric interaction energy, ll-clash expresses the ligand internal clash energy, rl-Hb expresses the receptor−ligand hydrogen bond energy, rl-ele expresses the receptor−ligand electrostatic interaction energy, and rl-clash expresses the receptor−ligand clash energy.

complex 1abe, while that obtained by MolDock is 0.17 Å for complex 3ptb. On the contrary, the worst *rmsd* obtained by ADDock is 5.91 Å for complex 4dfr, while that obtained by MolDock is 7.48 Å for complex 6rnt. It is necessary to carefully compare the docking results given by ADDock and MolDock since both methods use the same scoring function to score the docking results. In fact, both methods have 24 winning cases over the other if one takes smaller *rmsd* obtained for a complex by one method over the other as the winning case. Both methods also agree with each other on 73% or 35 complexes docked since all these *rmsd* computed by both are below 2.0 Å and are close to each other. Some badly docked complexes such as 1glq, 1hri, and 1tmn by ADDock are also badly docked by MolDock. However, complexes 1eap, 1epb, 1lic, 1lna, and 6rnt are well docked by ADDock but poorly docked by MolDock. On the other hand, complexes 1drl, 1dwd, 1tka, 2cgr, and 4dfr are poorly docked by ADDock but well docked by MolDdock. With an averaged *rmsd* of 1.38 Å computed, the overall docking accuracy by MolDock for 77 complexes is 87.01%. We conclude that ADDock is as accurate as MolDock on docking these 48 complexes despite the fact that each uses a different docking strategy for a similar scoring function.

Selection of Anchors by ADDock. The other main feature of ADDock is that no translational motion is implemented for a ligand during the docking process. In some known docking methods, a ligand is placed randomly in an active site and randomly moved inside the active site through a combination of translational and rotational motion,[52–55] or ligand conformation is generated by a conformational search and then placed in the active site through a systematic search.[56] Because only rotational motion of the whole molecule is implemented and the conformation is randomly generated in ADDock, we search the best binding pose for a ligand through systematic changing of the anchor selected. We also intend to distinguish the intramolecular clash

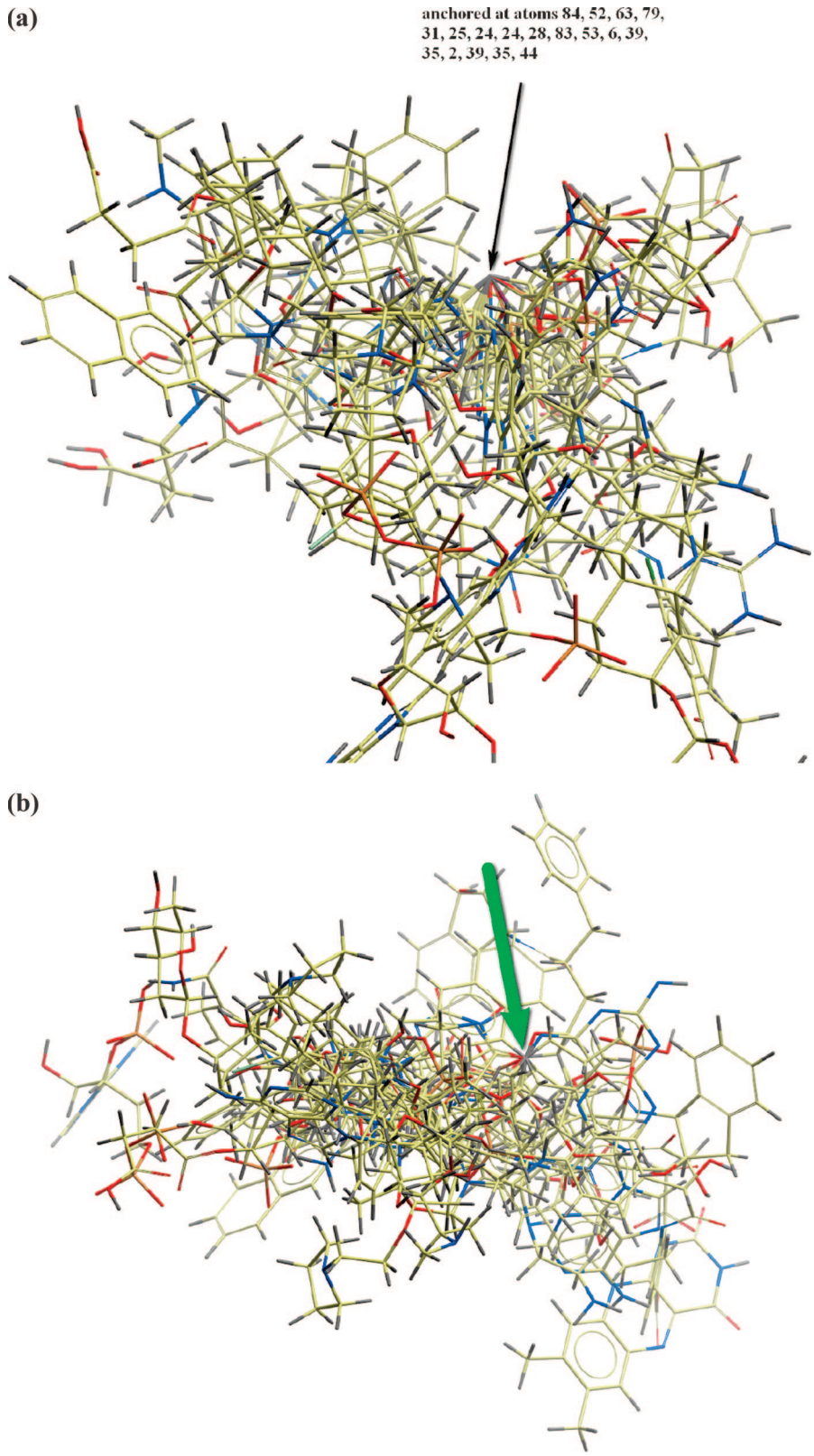


Figure 11. a. Before docking by ADDock into PDE5 or the FTase active site, the structures prepared for 18 inactive compounds namely, Vac, Fad, S5h, Coa, Ola, Npp, Ae2, Stu, Dgx, Xk2, Dhg, Fmn, 961, Hds, Ndp, Bzt, Mtx, and E64 chosen from Table 1 are all anchored at the same anchors (marked by an arrow) chosen for docking these crystal ligands into their native receptors (Table 1). The coordinate of each inactive compound rooted on each anchor chosen is brought by ADDock to the same origin of the anchor chosen for docking crystal ligand VIA or 2C5 into PDE5 or FTase active site. The corresponding anchors chosen for these inactive compounds are atoms 84, 52, 63, 79, 31, 25, 24, 24, 28, 83, 53, 6, 39, 35, 2, 39, 35, and 44, respectively (Table 1). For saving brevity, only 18 structures of the inactive compounds of sets (ii) and (iii) are selected for such a presentation. **b**. After docking by ADDock, the docked structures of these 18 inactive compounds are still anchored at the anchor atoms chosen (see part **a**) as marked by a green arrow.

between ligand atoms from the intermolecular one between ligand and receptor atoms during the docking process by imposing two different clash penalties into the overall docking score. The intermolecular clash penalty is set 10

times larger than that of the intramolecular one such that the former will be easily recognized by GA during the docking process. To demonstrate how anchors are selected by ADDock for docking a ligand into its receptor active site, we present in Figure 1 a diagram for ligand C60 (1rne)[57] where all the terminal hydrogen atoms are preliminarily selected as an anchor and labeled by atomic id numbers. The corresponding docked energy of each of these preliminary dockings is plotted in Figure 2 as a function of the atomic id numbers. There are 58 terminal hydrogen atoms being selected as an anchor in these preliminary dockings (Figure 1). Apparently as shown in Figure 2, the docked energy for atom 81 being chosen as an anchor is among the smallest of all the terminal hydrogen atoms chosen. The number of GA steps (generations × populations) is uniquely set as 40000 for these preliminary dockings. The docked energy obtained for atom 81 chosen as an anchor is 744 kcal/mol (−256 interaction plus 1000 kcal/mol penalty). Therefore, we decide to choose atom 81 as the anchor for subsequent docking for C60. A comparison for the docked conformation obtained from choosing atom 81 as the anchor with the crystal structure of C60 is shown in Figure 3. The *rmsd* computed for this docking is 0.41 Å. As shown in Figure 3, the docked substructure of C60 nearby the anchor is closer in structural feature to its crystal counterpart than the rest of the structure compared. This shows the fact that the docked substructure given by ADDock nearby the anchor is usually bearing a greater resemblance in structural features to its crystal counterpart even for the poorly docked ligand Gtb (1glq)[58] (Table 1) (data not shown here). The docking trajectory given by ADDock for C60 with atom 81 chosen as the anchor is followed by a plot of the minimum *rmsd* computed and searched in each GA generation against the generation number and presented in Figure 4. Apparently, the *rmsd* computed during the docking process is widely varied between 0.41 and 11.9 Å (Figure 4). This reflects the fact that C60 is a very flexible ligand since the number of rotatable bonds identified on it is 19 (Table 1).

Dependence of Docking Results by ADDock on the Cavity Size. The docked energy for each of the 92 complexes docked by ADDock is plotted against the corresponding *rmsd* computed and shown in Figure 5. The docked energy obtained for 36.9% (31/84), 50% (42/84), and 13.09% (11/84) of 84 correctly docked complexes by ADDock is around 10000, 1000, and 0 kcal/mol, respectively (Table 1 and Figure 5). This shows that docked energy of the most correctly docked complexes by ADDock is around 10000 or 1000 kcal/mol (Table 1 and Figure 5). Only 11 out of 84 complexes of this category give a docked energy that is around or smaller than 0. kcal/mol. Further, the docked energy obtained for 58.6% (17/29) and 41.4% (12/29) complexes of the 29 best docked complexes with the corresponding *rmsd* obtained smaller than 1.0 Å is around 10000 and 1000 kcal/mol, respectively (Table 1 and Figure 5). None of these complexes docked with a docked energy that is around or smaller than 0. kcal/mol (Table 1 and Figure 5). Therefore, the docked energy obtained by ADDock tends to decrease with the increase of *rmsd* as shown in Figure 5. We also observe the fact that a larger binding cavity for a ligand may give more freedom and eventually lead to a bad docked result for the ligand which is in accord with that reported previously by others.[59] To estimate the size of the binding cavity for each ligand, we count the number of cavity grid points for a grid constructed with a 1 Å grid space to surrounding each ligand and filling up each corresponding cavity. The ratio between the number of cavity grid points and the number of atoms of each ligand counted is designated as the **g/a** ratio presented in Table 1 which is also being plotted against the corresponding *rmsd* obtained for each ligand and shown in Figure 6. The larger the **g/a** ratio computed means the greater the binding cavity perceived by a ligand. Apparently, the **g/a** ratio computed for each complex docked by ADDock tends to increase with the increase of *rmsd* (Figure 6). Moreover, the **g/a** ratios of the 8 badly docked complexes are somewhat larger than those of the other accurately docked complexes by ADDock (Table 1 and Figure 6).

Evaluating the Compound Screening Ability of ADDock. To test whether or not ADDock is capable of screening virtual compounds, we have docked 10 active plus 35 or 33 inactive compounds into the following three receptors randomly chosen from the literature: (i) 1vyz (N5b),[60] the type 2 cyclin-dependent kinase CDK2 complexes with ligand N5b, (ii) 1udt (VIA),[61] the type 5 phosphodiesterase PDE5 complexes with ligand VIA, and (iii) 1ni1 (2C5),[62] the farnesyltransferase FTase complexes with ligand 2C5. The number of compounds docked by ADDock for each of these mixed active and inactive compound sets is 45 (10 active plus 35 inactive) for set (i), 43 (10 active plus 33 inactive) for set (ii), and 43 (10 active plus 33 inactive) for set (iii), respectively. The structures of each set of 10 inhibitors docked plus their corresponding activity measured are shown in Table 3. The activity measured in IC_{50} for each set of 10 inhibitors docked into CDK2 (set (i)),[60] PDE5 (set (ii)),[61] and FTase (set (iii))[62] ranges from 34 to 705, 0.9 to 55.0, and 0.16 to 68 nM, respectively (Table 3). The crystal ligand of each complex is used as a template to generate the structure for each of these active ligands as described previously.[63] The *rmsd* obtained by ADDock for docking each crystal ligand into its corresponding receptor are 1.23 Å for N5b (anchored at atom 19), 1.17 Å for VIA (anchored at atom 41), and 1.26 Å for 2C5 (anchored at atom 11), respectively. The corresponding docked energies obtained are 9898.9 (−101.1 interaction plus 10000 clash penalty) for N5b, 850.7 (−149.27 interaction plus 1000 clash penalty) for VIA, and 856.5 (−143.46 interaction plus 1000 clash penalty) kcal/mol for 2C5, respectively. The active site of PDE5 and FTase appears to be much larger than that of CDK2 since the corresponding **g/a** ratios computed are 13.1, 14.4 for the formers, and 9.4 for the latter. In docking 2C5 into the FTase active site, the hydroxyl farnesyl phosphate group bound to the active site is kept and treated as a part of the active site atoms identified. The GA steps used for docking these crystal ligands as well as the three sets of combined active and inactive compounds are fixed as 250000.

There is only one structural type of the 10 inhibitors of CDK2 docked in set (i) (Table 3). The crystal ligand N5b with measured IC_{50} of 34 nM and designated as **5**[60] is also among the 10 inhibitors docked (Table 3). After some preliminarily docking tries, the anchors for each of these inhibitors chosen are atoms 19, 15, 20, 19, 15, 20, 20, 19, 20, and 19 for inhibitor **5**, **20**, **9**, **18**, **19**, **8**, **25**, **3**, **12**, and **17** (Table 3 and Figure 7a), respectively. The corresponding

numbers of rotatable bonds identified and implemented in the docking process are 3, 1, 4, 1, 1, 3, 3, 3, 4, and 2, respectively. Before docking by ADDock, all 10 initial structures prepared are all anchored at a hydrogen atom on the pyrazole ring chosen as shown in Figure 7a. After docking by ADDock, the docked structures of these 10 inhibitors are still anchored at the anchor atoms chosen as shown in Figure 7b. However, the docked conformations are significantly deviated from those before docking (Figure 7a,b). Except for inhibitor **3**, the docked energies of all these inhibitors are close to that obtained for crystal ligand N5b (Table 4). However, the docked energies computed for these inhibitors are not ranked in parallel with the IC_{50} measured (Table 4). The 35 inactive compounds of the set are chosen from the 92 docked crystal ligands described in Table 1. We choose the inactive ligands such that their molecular masses are as close to the 10 inhibitors docked as possible. As shown in Table 4, the number of atoms of 10 inhibitors docked varies from 29 to 36 while that of 35 inactive ligands selected varies from 30 to 48. We perform a series of preliminary dockings for each of these inactive compounds by choosing several terminal atoms as the anchors but subsequently find that the corresponding docked results are no better than those obtained for docking these inactive compounds into their native receptors (Table 1). Therefore, we choose the same anchors for docking these inactive compounds into their native receptors (Table 1) as the anchors for these dockings. The numbers of rotatable bonds implemented in docking these inactive compounds are similar to those described in Table 1. To dock these inactive compounds into the CDK2 active site, we rotate and transfer the coordinate of each inactive compound generated from its anchor to the coordinate frame of crystal ligand N5b. As shown in Figure 8a before docking, the coordinate of each inactive compound rooted on each anchor chosen is brought by ADDock to the same origin of the anchor chosen for docking crystal ligand N5b into the CDK2 active site. After docking as shown in Figure 8b, all the docked conformations generated by ADDock are still anchored although they are significantly different from the undocked ones (Figure 8a). The docked energies computed for these inactive compounds are varied from −1.0 to 9919.2 kcal/mol (Table 4). The energy components of the docked energies of the whole set are listed and compared in Table 4. Among all the energy components compared, we find that the receptor−ligand steric interaction energies (the rl-steric scores) are apparently correlated with the IC_{50} measured for the 10 inhibitors if they are sorted in order (Table 4). Moreover, we find that the sorted rl-steric scores varied from −120.9 to −88.8 for the 10 inhibitors are well separated from those varied from −88.8 to 14.8 kcal/mol for the 35 inactive compounds docked (Table 4). In other words, these 10 active compounds are completely separated by ADDock from the inactive ones after the docking process.

There are two major structural types identified in the 10 inhibitors selected and docked in set (ii) by ADDock into the PDE5 active site (Table 3). Since the structures of 10 PDE5 inhibitors are much different from that of crystal ligand VIA, we have performed a series of preliminary dockings on these inhibitors to find that atom 7, the oxygen atom on the cyclic guanine ring (Table 3), can be chosen as a unique anchor for these compounds (Figure 9a). The numbers of rotatable bonds implemented for inhibitors **34**, **33**, **32**, **29**, **25**, **30**, **24**, **23**, **36**, and **37** (Table 3) during docking are 4, 4, 3, 4, 5, 3, 4, 3, 4, and 3, respectively. However, there are three structural types identified for the 10 inhibitors of set (iii) (Table 3). After some preliminary dockings, a hydrogen atom on the cyanopyridine ring, namely atoms 48, 50, 50, 49, 50, and 45 of inhibitors **24c**, **24a**, **24e**, **24b**, **24f**, and **26i** (Table 3), are chosen respectively as the anchors (Figure 10a). For inhibitors **23c**, **26j**, **26a**, and **26h** (Table 3), a hydrogen atom on the phenyl ring of group Ar (atom 39), the biphenyl ring of group R (atom 40), the phenyl ring of group R (atom 39), and the ring structure of group R (atom 58) are chosen as the anchors, respectively (Figure 10a). The number of rotatable bonds identified and implemented in the docking processes for these 10 inhibitors is uniquely set as two. The structures of these two inhibitor sets before (Figures 9a and 10a) and after (Figures 9b and 10b) docking by ADDock are all anchored and rooted on the same anchors chosen for docking the two corresponding crystal ligands as similar to those performed for set (i). Since the molecular sizes for set (ii) or (iii) inhibitors are varied from 49 to 57 or 50 to 60 atoms (Tables 5 and 6), we select 33 compounds with molecular sizes varied from 50 to 86 atoms out of the 92 docked crystal ligands (Table 1) as the inactive compound set for both sets (ii) and (iii). In other words, the 33 inactive compounds selected for both sets (ii) and (iii) are docked into receptors PDE5 and FTase, respectively. We also take the same anchors chosen for docking these inactive compounds into their native receptors (Table 1) as the anchors for these dockings (Figure 11a). The docked structures of all these inactive compounds are anchored as shown in Figure 11b. Apparently, the sorted rl-steric scores computed for the 10 inhibitors of both sets (ii) and (iii) are also separated from those computed for most of the 33 inactive compounds docked as shown in Tables 5 and 6, respectively. The sorted rl-steric scores computed for the 10 inhibitors of set (i) or (iii) are varied from −121.6 to −88.7 or −158.7 to −66.4 kcal/mol, respectively (Tables 5 and 6). However, the rl-steric scores computed for 8 (Sb3, Gel, Hem, Bzt, Mtx, Blv, Psi, and Ndp) or 11 (Fmn, Bzt, Blv, Apr, Hem, Rea, Sph, Gep, Nad, Gel, and Mid) inactive compounds of set (ii) or (iii) (Tables 5 and 6) are also falling into these two ranges. The molecular sizes of these inactive compounds are varied from 57 to 86 or 50 to 77 atoms for set (ii) or (iii), respectively (Table 5 and 6). Apparently, the molecular sizes of some of these inactive compounds are far larger than those of the 10 inhibitors docked especially for set (ii). This implies that molecular size can be used as a supplementary criterion for accompanying the rl-steric scores computed in a virtual screening experiment conducted by ADDock to screen the active compounds. For example, the successful screening rate for active compounds is only 56 (10/18) or 47% (10/21) for set (ii) or (iii), respectively, if the rl-steric score computed is the only criterion used in the screening (Tables 5 and 6). However, the successful screening rate for active compounds can be enhanced respectively to 91 (10/11) or 67% (10/15) for set (ii) or (iii), if a cutoff in molecular size for screened compounds of both sets is set as 60 atoms (Tables 5 and 6). The choice of rl-steric scores computed by ADDock as a screening criterion is not accidental. We find that the rl-steric scores computed by ADDock are discriminative even for inactive ligands that are similar or smaller in molecular size

than the active ones. Furthermore, we find that some inactive compounds of the same series docked with poor rl-steric scores into PDE5 are also poorly docked into the FTase active site by ADDock. For examples, the rl-steric scores of the following poorly docked inactive compounds, namely Dhg, Xk2, Dgx, Stu, Ae2, Npp, Ola, Coa, S5h, Fad, Vac, and Rtl, by ADDock into both active sites are ranked correlatively (Tables 5 and 6).

CONCLUSION

Searching initial docking poses for a ligand is required by some available docking methods. These indecisive procedures are avoided by ADDock by anchoring the ligand at an energetically favorable position inside the active site for subsequent docking. Since ligands are anchored at the same position before and after docking, ADDock guarantees that during the docking process all the ligands are moving within the same subspace of active site. This may be advantageous for docking a series of different compounds into the same active site since the corresponding docked energies computed are compared on the same ground. To reduce the *rmsd* or improve the docking accuracy for the 8 badly docked complexes (Table 1) by ADDock, we have gradually increased the GA steps from 20000 to 640000. However, we failed to bring down the *rmsd* of any of these 8 badly docked complexes to that below 2 Å within 640000 GA steps used. This shows that it is disadvantageous for ADDock to dock small ligands into large receptor active sites that do not provide enough geometric constraints on ligands during a docking process. The docking speed given by ADDock is reasonably quick since it takes about 126 s on an HP workstation running on a 1.5 GHz Intel Itanium2 processor to complete a docking of 40000 GA steps for a ligand consisting of 50 atoms. To rank the random compounds searched, ADDock takes the rl-steric scores computed for each compound docked as the major criterion. This screening criterion may be accompanied by the molecular size of screened ligands to achieve a better screening accuracy.

The probability for docking ligands accurately by ADDock relies on the choice of an anchor and a random seed number entered at the beginning. In fact, selection of an anchor is much more important than entering a random number at the beginning. The probability of being accurately docked by ADDock for an anchor chosen will be quite different from that for a different anchor chosen. Since no translational motion for ligands is implemented, ADDock chooses anchors by changing all the terminal atoms identified on a ligand systematically. In other words, a series of structures are generated at the beginning based on each preliminary anchor chosen, and then the best anchor is chosen as the minimum docked energy computed among all the preliminary anchors. We find that terminal atoms deeply buried inside an active site are usually selected as the best anchor. We also find that most of the 8 badly docked ligands by ADDock are loosely bound within their corresponding active sites. In this regard, ADDock agrees with those reported in the literature[59,64] that it is rather difficult to accurately dock a ligand that is loosely bound inside its active site. For all 92 complexes docked, the GA parameters such as crossover and mutation probabilities are determined by some preliminary runs and then fixed throughout the docking processes. To obtain greater docking accuracy by ADDock, the GA steps are literally increased while keeping all the GA parameters fixed.

ADDock has been commercialized by a local firm. Interested parties can contact jeff@biodelight.com.tw or http://www.biodelight.com.tw to obtain a copy of the program.

ACKNOWLEDGMENT

This work is supported in part by a grant (NSC95-2313-B007-002) from the National Science Council, Taiwan, ROC. The SYBYL 7.3 program is provided by the National Center for High Performance Computing, Taiwan, ROC.

Supporting Information Available: *rmsd* and steric interaction energies computed from a cross-docking process for 15 ligand−receptor complexes by ADDock (Tables S.1 and S.2) and a list of program and file systems used and generated by ADDock during a self- or cross-docking process. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
(2) Stahura, F. L.; Bajorath, J. New methodologies for ligand-based virtual screening. *Curr. Pharm. Des.* **2005**, *11*, 1189–1202.
(3) Levinthal, C.; Wodak, S. J.; Kahn, P.; Dadivanian, A. K. Hemoglobin interaction in sickle cell fibers. I. Theoretical approaches to the molecular contacts. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1330–1334.
(4) Salenme, F. R. A hypothetical structure for an intermolecular electron transfer complex of cytochromes C and b5. *J. Mol. Biol.* **1976**, *102*, 563–568.
(5) Koshland, D. E., Jr. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–104.
(6) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.* **1986**, *29*, 2149–2153.
(7) Gabb, J.; Jackson, R. M.; Sternberg, M. J. E. Modeling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997**, *272*, 106–120.
(8) Jiang, F.; Gao, Y.; Mao, F.; Liu, Z.; Lai, L. Potential of mean force for protein-protein interaction studies. *Proteins* **2002**, *46*, 190–196.
(9) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecular-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(10) Norel, R.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Shape complementarity at protein-protein interfaces. *Biopolymers* **1994**, *34*, 933–940.
(11) Busetta, B.; Tickle, I. J.; Blundel, T. L. DOCKER, an interactive program for simulating protein receptor and substrate interactions. *J. Appl. Crystallogr.* **1983**, *16*, 432–437.
(12) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(13) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
(14) Goodsell, D. S.; Olson, A. J. Automated docking of structures to proteins by simulated annealing. *Proteins* **1990**, *8*, 195–202.
(15) Liu, M.; Wang, S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 435–451.
(16) Trosset, J. Y.; Scheraga, H. A. Prodock: software package for protein modeling and docking. *J. Comput. Chem.* **1999**, *20*, 412–427.
(17) Murray, C. W.; Baxter, C. A.; Frankel, A. D. The sensitivity of the results of molecular docking to induced effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547–562.
(18) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(19) Yang, J. M.; Chen, C. C. GEMDOCK: A genetic evolutionary method for molecular docking. *Proteins* **2004**, *55*, 288–304.

(20) Thomsen, R.; Christensen, M. H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49*, 3315–3321.
(21) Grosdidier, A.; Zoete, V.; Michielin, O. EADock: Docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins* **2007**, *67*, 1010–1025.
(22) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424–440.
(23) DesJarlais, R. L.; Scheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **1988**, *31*, 722–729.
(24) 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*, 411–428.
(25) Brooijmans, N. Theoretical studies of molecular recognition. Ph.D. thesis, University of California San Francisco,January 24, 2003, 250 pp.
(26) Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P.; Miller, M. D. Flexibases: A way to enhance the use of molecular docking methods. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 565–582.
(27) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. FLOG: A system to select quasi-flexible ligands complementary to receptor of known three-dimensional structures. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 153–174.
(28) Boehm, H. J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309–323.
(29) Brown, R. D.; Martin, Y. C. Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572–584.
(30) Budin, N.; Majeux, N.; Caflisch, A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.* **2001**, *382*, 1365–1372.
(31) Pazos, F.; Sternberg, M. J. Automated prediction of protein function and detection of functional sites from structures. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14754–14759.
(32) Gehlhaar, D. K.; Verkhivker, G.; Rejto, P. A.; Foget, D. B.; Fogel, L. J.; Freer, S. T. Docking conformationally flexible small molecules into a protein binding site through evolutionary programming. In Evolutionary Programming IV, Proceedings of the Fourth International Conference on Evolutionary Programming, San Diego, CA, March 1−3, 1995; McDonnell, J. R., Reynalds, R. G., Fogel, D. B., Eds.; MIT Press: Cambridge, MA, 1995, pp615−627.
(33) Berman, H. M.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980.
(34) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: Pharmacophore-based protein-ligand docking. *J. Med. Chem.* **2004**, *47*, 6804–6811.
(35) SYBYL 7.3. The Tripos Associates, 1699 Hanley Rd., St. Louis, MO, 2006.
(36) Lin, T. H.; Yu, Y. S.; Chen, H. J. Classification of some active compounds and their inactive analogues using two three-dimensional molecular descriptors derived from computation of three-dimensional convex hulls for structures theoretically generated for them. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1210–1221.
(37) Randic, M.; Wilkins, C. L. Graph theoretical approach to recognition of structural similarity in molecules. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 31–37.
(38) Moock, T. E.; Henry, D. R.; Ozkabak, A. G.; Alamgir, M. Conformational searching in ISIS/3D Databases. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 184–189.
(39) Singh, A. P.; Latombe, J. C.; Brutlag, D. L. A motion planning approach to flexible ligand binding. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1999**, *7*, 252–261.
(40) Hartenburg, R. S.; Denavit, J. A. Kinematic notation for lower pair mechanisms based on matrices. *J. Appl. Mechanics* **1955**, *77*, 215–221.
(41) Shoemake,K. Uniform random rotations. In *Graphics Gems III*, 1st ed.; Kirk, D., Ed.; AP Professional: Boston, MA, 1992; pp 124−132.
(42) Trosset, J. Y.; Scheraga, H. A. PRODOCK: Software package for protein modeling and docking. *J. Comput. Chem.* **1999**, *20*, 412–427.
(43) Hestenes, D. Proper dynamics of a rigid point particle. *J. Math. Phys.* **1974**, *15*, 1778–1791.
(44) Juffer, A. H.; Shepherd, C. M.; Vogel, H. J. Protein-membrane electrostatic interactions: application of the lekner summation technique. *J. Chem. Phys.* **2001**, *114*, 1892–1905.
(45) Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr.* **1978**, *A34*, 827–828.
(46) Lin, T. H.; Chiu, S. H.; Tsai, K. C. Supervised feature ranking using a genetic algorithm optimized artificial neural network. *J. Chem. Inf. Model.* **2006**, *46*, 1604–1614.
(47) Gehlhaar, D. K.; Bouzida, D.; Rejto, P. A. Fully automated and rapid flexible docking of inhibitors covalently bound to serine proteases. In Evolutionary Programming VII, Proceedings of the Seventh International Conference on Evolutionary Programming,San Diego, CA, March 1−3, 1998; Eiben, A., Porto, V., Saravanan, N., Eds.; Springer-Verlag:New York, 1998; pp 449−461.
(48) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(49) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T A. Glide: A new approach for rapid accurate docking and scoring. I. method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
(50) Rarey, M.; Krame, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(51) Jain, A. N. Surflex: Fully automatic molecular docking using a molecular similarity-based searching engine. *J. Med. Chem.* **2003**, *46*, 499–511.
(52) Trosset, J. Y.; Scheraga, H. A. Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8011–8015.
(53) Abagyan, R.; Totrov, R. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.
(54) Abagyan, R.; Totrov, R.; Kuznetsov, D. ICM-a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488–506.
(55) Totrov, R.; Abagyan, R. Flexible protein-ligand docking by global energy minimization in internal coordinates. *Proteins (Suppl.)* **1997**, *1*, 215–220.
(56) Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* **2001**, *22*, 1750–1771.
(57) Gruetter, M. G.; Rahuel, J.; Priestle, J. P. The crystal structure of recombinant glycosylated human rennin alone and in complex with a transition state analog inhibitor. *J. Struct. Biol.* **1991**, *107*, 227–236.
(58) Garcia-Saez, I.; Parraga, A.; Phillips, M. F.; Mantle, T. J.; Coll, M. Molecular structure at 1.8 Å of mouse liver class pi glutathione S-transferase complexed with S-(p-nitrobenzyl)glutathione and other inhibitors. *J. Mol. Biol.* **1994**, *237*, 298–314.
(59) Onodera, K.; Satou, K.; Hirota, H. Evaluations of molecular docking programs for virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1609–1618.
(60) Pevarello, P.; Brasca, M. G.; Amici, R.; Orsini, P.; Traquandi, G.; Corti, L.; Piutti, C.; Sansonna, P.; Villa, M.; Pierce, B. S.; Pulici, M.; Giordano, P.; Martina, K.; Fritzen, E. L.; Nugent, A.; Casale, E.; Cameron, A.; Ciomei, M.; Rolette, F.; Isacchi, A.; Fogliatto, G.; Pesenti, E.; Pastori, W.; Marsiglio, A.; Leach, K.; Clare, P. M.; Fiorentini, F.; Varasi, M.; Vulpetti, A.; Warpehoski, M. A. 3-aminopyrazole inhibitors of CDK2/Cyclin A as antitumor agents I. lead finding. *J. Med. Chem.* **2004**, *47*, 3367–3380.
(61) Pissamitski, D. A.; Asberom, T.; Boyle, C. D.; Chackalamannil, S.; Chintala, M.; Clader, J. W.; Greenlee, W. J.; Hu, Y.; Kurowski, S.; Myers, J.; Palamanda, J.; Stamford, A. W.; Vemulapalli, S.; Wang, Y.; Wang, P.; Wu, P.; Xu, R. SAR development of polycyclic guanine derivatives targeted to the discovery of a selective PDE5 inhibitor for treatment of erectile dysfunction. *Bioorg. Med. Chem.* **2004**, *14*, 1291–1294.
(62) Tong, Y.; Lin, N. H.; Wang, L.; Hasvold, L.; Wang, W.; Leonard, N.; Li, T.; Li, Q.; Cohen, J.; Gu, W. Z.; Zhang, H.; Soll, V.; Bauch, J.; Marsh, K.; Rosenberg, S. H.; Sham, H. L. Discovery of potent imidazole and cyanophenyl containing farnesyltransferase inhibitors with improved oral bioavailability. *Bioorg. Med. Chem.* **2003**, *13*, 1571–1574.
(63) Wei, H. Y.; Tsai, K. C.; Lin, T. H. Modeling ligand-receptor interaction for some MHC class II HLA-DR4 peptide mimetic inhibitors using several molecular docking and 3D QSAR techniques. *J. Chem. Inf. Model.* **2005**, *45*, 1343–1351.
(64) Sato, H.; Shewchuk, L. M.; Tang, J. Prediction of multiple binding modes of the CDK2 inhibitors, anilinopyrazoles, using the automated docking programs GOLD, FlexX, and LigandFit: an evaluation of performance. *J. Chem. Inf. Model.* **2006**, *46*, 2552–2562.

CI800124G