ASEDock-Docking Based on Alpha Spheres and Excluded Volumes

Junichi Goto and Ryoichi Kataoka

Computational Science Department, Science & Technology Systems Division, Ryoka Systems, Inc., 1-28-38 Shinkawa, Chuo-ku, Tokyo 104-0033, Japan

Hajime Muta and Noriaki Hirayama*

Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1143, Japan

Received September 19, 2007

ASEDock is a novel docking program based on a shape similarity assessment between a concave portion (i.e., concavity) on a protein and the ligand. We have introduced two novel concepts into ASEDock. One is an ASE model, which is defined by the combination of alpha spheres generated at a concavity in a protein and the excluded volumes around the concavity. The other is an ASE score, which evaluates the shape similarity between the ligand and the ASE model. The ASE score selects and refines the initial pose by maximizing the overlap between the alpha spheres and the ligand, and minimizing the overlap between the excluded volume and the ligand. Because the ASE score makes good use of the Gaussian-type function for evaluating and optimizing the overlap between the ligand and the site model, it can pose a ligand onto the docking site relatively faster and more effectively than using potential energy functions. The posing stage through the use of the ASE score is followed by full atomistic energy minimization. Because the posing algorithm of ASEDock is free from any bias except for shape, it is a very robust docking method. A validation study using 59 high-quality X-ray structures of the complexes between drug-like molecules and the target proteins has demonstrated that ASEDock can faithfully reproduce experimentally determined docking modes of various druglike molecules in their target proteins. Almost 80% of the structures were reconstructed within the estimated experimental error. The success rate of ∼98% was attained based on the docking criterion of the root-mean-square deviation (RMSD) of non-hydrogen atoms (≤2.0 Å). The markedly high success of ASEDock in redocking experiments clearly indicates that the most important factor governing the docking process is shape complementarity.

1. INTRODUCTION

Docking simulations between drug candidates and their target proteins are now one of the essential methodologies in many drug discovery projects. The docking simulation is practically an in silico counterpart of high-throughput screening (HTS). The chemical space that can be surveyed by docking simulations is significantly larger than that by HTS. Three-dimensional structures of many biologically important proteins are registered in the Protein Data Bank (PDB)[1] now, and the number of entries is expanding recently. The availability of reliable atomic coordinates of target proteins is also raising great expectations for the contribution of docking simulations to many drug discovery projects.

Many programs such as FLexX,[2] AutoDock,[3] GOLD,[4] Glide,[5] and Surflex[6] have been proposed, and some of them are actually contributing in discovering novel drug molecules. From this sense, the docking technology is edging close to practical use. Nevertheless, many essential problems regarding docking simulations are not solved yet and none of the currently available programs are perfect in predicting the correct binding modes. Therefore, the improvement of the performance of docking simulations is one of the priority subjects in drug discovery technologies.

* To whom correspondence should be addressed. Tel.: +81 463 93 1121. E-mail address: hirayama@is.icc.u-tokai.ac.jp.

We developed a docking program called Ph4Dock.[7] This program exploits the pharamacophoric features both in the ligand and the concave portions of the target protein. Although Ph4Dock gave quite a good performance in the redocking experiments of 43 quality X-ray structures of complexes between drug-like molecules and the target proteins, it failed to give reasonable answers to several structures. After applications of Ph4Dock to many problems, we have learned that Ph4Dock had two important drawbacks. First, in Ph4Dock, the pharmacophore model and the ligand are compared. However, the pharmacophore model cannot always accurately represent the landscape of the concavity, and the poor concordance between the pharmacophore model and the ligand was a reason for failure. Second, the accuracy of pharamacophore was not necessarily adequate in some cases, and manual intervention was required in the actual docking simulations. We have been generally satisfied with the docking performance by Ph4Dock, but, at the same time, the experiences on these problems have led us to think that we should put a high priority on the concordance of the shapes of concavity and the ligand. In this paper, we present a novel docking algorithm that is based on an accurate modeling of concavity shape and a function that evaluates the shape similarity between the model and the ligand. The novel algorithm has proven the markedly high performance

10.1021/ci700352q CCC: $40.75
Published on Web 02/16/2008

of redocking experiments for 59 high-quality X-ray structures of complexes between the drug-like molecules and the target proteins.

2. COMPUTATIONAL METHODS

MOE (Molecular Operating Environment)[8] was used as a developing platform of ASEDock. All of the algorithms of ASEDock were coded using MOE's powerful vector language SVL (Scientific Vector Language). Existing features implemented in MOE were fully applied to realize the functions of ASEDock.

ASEDock is mainly composed of the following five steps: (i) generation of conformations for a ligand, (ii) concavity search on the surface of a target protein, (iii) generation of an ASE model at a concavity, (iv) rigid body alignment of the ligand conformations to the ASE model, and (v) the energy minimization of the posed conformations of the ligand in the concavity.

2.1. Generation of Ligand Conformations. A stochastic conformation generation method is applied, as in Ph4Dock. The hydrogen atoms are added in accordance with the standard protonation states of acids and bases in biological molecules. The generalized Born solvation area model[9] is used in the conformation energy calculations. The MMFF94x force field was used.[10] The following two selection procedures are used to omit redundant conformations. First, the conformations are clustered if the root-mean square deviations (RMSDs) of the corresponding non-hydrogen atoms are <0.5 Å. Second, the most diverse conformations are selected from the clustered conformations as follows. The first conformer to be selected is the one with the minimum energy, the second one is the conformer with the largest RMSD from the first, the third one is selected to be most distant from the previous conformers, and so on.

2.2. Concavity Search and Generation of ASE Model. A concave portion designated as concavity at the surface of a receptor molecule where a ligand is supposed to bind is identified as a collection of spheres, using the modified Delaunay triangulation.[11] The sphere, which is called the "alpha sphere", is used to characterize the concavity as in Ph4Dock. Surrounding a concavity where a ligand is bound, there is a protein region where the ligand cannot access. The volume of such region, which is called the "excluded volume", can be evaluated using the non-hydrogen atoms of the protein within 5.5 Å from the alpha spheres. The excluded volume of each non-hydrogen atom is calculated using a sphere with the radius of 1.4 Å for the lone-pair active atoms (such as nitrogen and oxygen atoms) or 1.8 Å for the non lone-pair active atoms (such as carbon atoms). Hereafter, the sphere is designated as an "exclusion sphere". An ASE model is defined by the combination of alpha spheres and the excluded volumes around the concavity. The ASE model is used as a guide on which various ligand conformations treated as rigid bodies are posed.

2.3. ASE Score. Alpha spheres can be generated at the binding site. Maximizing the overlap of the ligand atoms with the alpha spheres and minimizing the overlap of the ligand atoms with the exclusion spheres would yield the optimal posing of the ligand on the ASE model. To evaluate the goodness of fit of a pose onto the ASE model, we have introduced a simple Gaussian overlap function:

$$\text{ASE Score} = -\sum_{i}^{a,e}\sum_{j}^{l} w_{i,j}\exp(-\alpha\, d_{i,j}^{\,2})$$

Here, d is the distance between a ligand atom and an alpha sphere or an exclusion sphere in the ASE model, and α is an adjustable parameter to attenuate the influence of a ligand atom when it is aligned to the alpha sphere or the exclusion sphere. The sums are taken over the alpha spheres (a), the exclusion spheres (e), and the ligand atoms (l). The parameter w is a weighting factor, being $+r_a r_l$ and $-r_e r_l$ for alpha and exclusion spheres, respectively; here, r_a, r_e, and r_l are the radii of the alpha spheres, the radii of the exclusion spheres, and van der Waals radii of the ligand atoms, respectively. The smaller the ASE score, the larger the overlap between the alpha spheres and the ligand atoms, and the smaller the overlap between the ligand atoms and the exclusion spheres. Because this function is a simple continuous one, it is easy to align a ligand to the ASE model and obtain reasonable posed structures in a short time by minimizing this function.

2.4. Posing and Rigid Body Alignment of Ligand to the ASE Model. The selected independent conformers are posed onto the ASE model. Three non-hydrogen atoms are sampled so that the area of the triangle becomes the largest. On the other hand, three alpha spheres are randomly selected to make multiple triangles. By overlapping these two types of triangles, conformers can be posed to the alpha spheres.

The initial poses of ligands are further optimized with the rigid-body alignment[12] by minimizing the ASE score with the Newton-Lapson method. This process converges relatively faster than that with the standard interaction function. The Gaussian function takes the finite value at the origin and has a gentle gradient, compared to the interaction functions. Therefore, even if the ligand overlaps with the protein, the value of the ASE score would not soar and can easily remove the overlap. The ASE score could be a robust scoring function to evaluate and steadily optimize the initial pose of the ligand at the concavity of the protein.

2.5. Energy Minimization. Posed conformations in a higher rank in the ASE score or the interaction energies are then subjected to be optimized by minimizing the following function:

$$U_{\text{total}} = U_{\text{ele}} + U_{\text{vdw}} + U_{\text{solv}} + U_{\text{ligand}}$$

Here, U_{ele}, U_{vdw}, and U_{solv} are Coulombic, van der Waals, and generalized Born/solvent accessible solvation interactions, respectively. U_{ligand} is the internal conformation energy of the ligand. The MMFF94x force field was used to calculate each energy term.

Because the potential energy calculation is the most time-consuming process in the docking calculations, we divided this process into two steps: rough minimization with a short cutoff distance from the alpha sites at first, and more-precise calculation with longer cutoff distance afterward. In the first step of rough minimization, a shorter cutoff distance of 4.5 Å and a convergence criterion of the root-mean square (RMS) gradient of 10 kcal/(mol Å) are used. The interaction energies are calculated with increasing cutoff distance to 8 Å and then the five structures with the minimum interaction energies or the maximum ASE scores are selected for the second optimization. In the second step of fine minimization, 5−10 structures are further refined by setting the cutoff

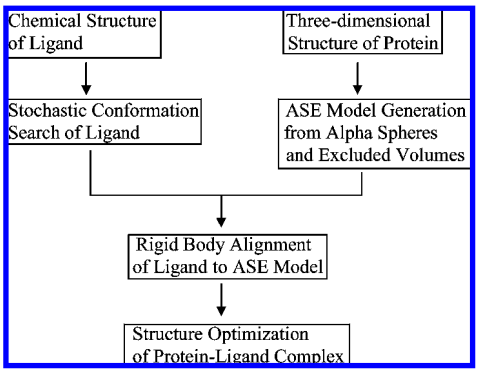

Figure 1. Flowchart of ASEDock.

distance and RMS gradient to be 8 Å and 0.1 kcal/(mol/Å), respectively.

ASEDock is mainly composed of six steps, as schematically shown in Figure 1.

2.6. Test Dataset. The best way to evaluate the potential of docking algorithms is reconstruction of the X-ray structures between proteins and ligands. If the structure of a ligand docked in the protein concords well with that in the crystal structure, we can judge that the docking algorithm may work well in a virtual screening. The precision of the atomic coordinates, as determined by X-ray analysis, is crucial for this evaluation. Therefore, we selected the best quality structures from PDB. The following five criteria are used to select the quality structure. First, the diffraction resolution and R_{free} are less than or equal to 2.5 Å and 0.24, respectively. These values guarantee that the X-ray analysis was undertaken well. Second, the occupancy factors of all non-hydrogen atoms in the structure are 1.0. This means that the protein structure is reasonably rigid and the docking calculations without including the effect of induced fit should be valid. Third, the ligand bound in the protein is drug-like. The drug-likeness of the ligand is judged by the criteria reported previously.[13] Fourth, the maximum temperature factors of non-hydrogen atoms in the ligand is ≤ 20 Å^2. This condition guarantees that the ligand molecule is tightly bound at the concavity. Fifth, covalently bound ligands and ligands that contain metal atoms were excluded from the scope of the target. Only 59 independent structures that fulfill the aforementioned criteria were screened from all PDB entries downloaded on June 12, 2007. The PDB codes of the structures, together with other information, are given in Table 1. The chemical names of the ligands together with a few physicochemical properties are given in Table 2. The ligands are sufficiently diverse. Therefore, these structures are suitable for redocking experiments.

In 1J4I and 1OIO, small molecules are not bound to the single protein; instead, they are entrapped between multiple proteins in the crystal structures, as shown in Figure 2. Since the small molecules contact with the multiple proteins in the crystals, three and two proteins surrounding the small molecules were considered for 1J4I and 1OIO, respectively, in regard to calculating the concavities.

2.7. Dielectric Constants. The nonbonded interatomic potential functions have crucial roles in molecular simulations. Especially in the docking simulations between a small molecule and a protein, the disposition and the structure of a small molecule can be highly affected by the nonbonded interactions with protein atoms. Therefore, the evaluation of the relevant interatomic interactions is critical in the docking simulations. A balance between electrostatic and van der Waals interactions is a key factor that affects the energy minimization process.

We have studied the effect of dielectric constant (ϵ) on the docking results of 59 complex structures in the test dataset. We especially paid attention to the effect of ϵ on hydrogen bond distances and the RMSDs of non-hydrogen

Table 1. Docking Results Obtained by ASEDock for 59 Protein−Ligand Complexes

PDB code	resolution (Å)	R_{free}	DPI (Å)	$2\sqrt{2}\sigma$ (Å)[a]	RMSD (Å)	lowest U_{total} (kcal/mol)
1az1	1.80	0.232	0.245	0.692	0.864	17.3
1c83	1.80	0.231	0.177	0.501	0.506	0.3
1c88	1.80	0.228	0.174	0.491	0.447	−0.6
1cbk	2.02	0.213	0.252	0.712	0.119	12.1
1cbs	1.80	0.237	0.218	0.615	0.462	8.9
1dbt	2.40	0.228	0.410	1.160	0.200	−10.8
1efi	1.60	0.231	0.215	0.609	0.407	49.0
1efv	2.10	0.222	0.343	0.972	0.159	−15.5
1eiz	1.70	0.237	0.202	0.573	0.543	18.5
1ekx	1.95	0.214	0.219	0.620	0.197	−77.7
1f5v	1.70	0.206	0.195	0.552	0.295	8.8
1f8y	2.40	0.218	0.365	1.033	0.280	15.9
1ftx	2.20	0.211	0.342	0.969	0.299	−31.6
1g2o	1.75	0.212	0.184	0.519	0.240	4.6
1gvo	1.38	0.200	0.110	0.311	0.432	21.8
1hjz	1.70	0.227	0.187	0.529	0.251	−2.6
1iei	2.50	0.199	0.492	1.391	0.341	28.0
1j4i	1.80	0.209	0.215	0.608	0.943	41.9
1ja9	1.50	0.221	0.121	0.342	0.116	12.1
1jak	1.75	0.192	0.125	0.353	0.960	−2.4
1kew	1.80	0.221	0.148	0.417	0.278	−60.7
1kkp	1.93	0.206	0.238	0.673	0.345	−8.2
1kl1	1.93	0.197	0.226	0.638	0.275	−12.2
1kpg	2.00	0.233	0.315	0.890	0.554	22.5
1kt7	1.27	0.195	0.088	0.248	0.163	8.8
1l4g	2.10	0.226	0.345	0.976	0.249	7.8
1l5k	2.00	0.216	0.262	0.740	0.402	−5.1
1l5l	2.00	0.197	0.244	0.690	0.427	−9.8
1l5m	2.00	0.218	0.275	0.779	0.454	−11.2
1lc8	1.80	0.229	0.199	0.564	0.235	−55.8
1li4	2.01	0.236	0.265	0.749	0.200	3.3
1lii	1.73	0.225	0.188	0.530	0.591	27.7
1m2g	1.70	0.220	0.190	0.538	0.950	4.1
1mm7	1.65	0.237	0.180	0.509	0.354	−16.7
1mqj	1.65	0.229	0.160	0.453	0.166	−17.9
1mvc	1.90	0.228	0.070	0.199	0.370	52.7
1n1t	1.60	0.197	0.137	0.388	0.643	9.6
1n3i	1.90	0.213	0.226	0.638	0.229	0.1
1nvr	1.80	0.226	0.193	0.545	0.223	77.4
1oh0	1.10	0.220	0.058	0.165	0.557	40.1
1oio	1.70	0.238	0.123	0.349	0.366	9.8
1owe	1.60	0.236	0.205	0.579	0.427	29.1
1p4j	2.00	0.198	0.231	0.654	0.221	21.0
1pfy	1.93	0.224	0.255	0.721	0.504	21.9
1pg4	1.75	0.211	0.200	0.565	0.291	1.0
1r0p	1.80	0.197	0.190	0.539	0.223	62.8
1sg0	1.50	0.234	0.141	0.398	0.580	19.4
1t40	1.80	0.193	0.190	0.536	0.278	16.8
1tkg	1.50	0.219	0.136	0.384	4.019	19.6
1uf5	1.60	0.198	0.142	0.403	0.222	−22.3
1urg	1.80	0.217	0.202	0.571	0.554	41.8
1v7c	2.00	0.230	0.262	0.740	0.245	−47.1
1wap	1.80	0.225	0.215	0.608	0.166	−6.0
1wu6	1.45	0.180	0.099	0.280	0.178	37.1
1xoz	1.37	0.193	0.101	0.285	0.214	62.7
1yki	1.70	0.167	0.144	0.406	0.245	3.3
2ay3	2.40	0.221	0.506	1.432	0.371	17.7
2prj	2.30	0.237	0.437	1.237	0.253	9.4
2skc	2.40	0.217	0.433	1.225	0.242	10.1

[a] $\sigma = \sigma(r, B_{avg})$.

Table 2. Chemical Names and Some Physicochemical Properties of the Ligands

PDB code	ligand name	MW[a]	SlogP[b]	TPSA[c]	Nrot[d]
1az1	alrestatin	254.2	0.19	77.5	2
1c83	6-(oxalyl-amino)-1H-indole-5-carboxylic acid	246.2	−1.78	125.2	3
1c88	2-(oxalyl-amino)-4,5,6,7-tetrahydro-thieno [2,3-c]pyridine-3-carboxylic acid	269.3	−3.31	126.0	3
1cbk	7,8-dihydro-7,7-dimethyl-6-hydroxypterin	209.2	−1.90	108.6	0
1cbs	retinoic acid	299.4	4.27	40.1	5
1dbt	uridine-5′-monophosphate	322.2	−4.73	159.5	4
1efi	4′-aminophenyl-alpha-D-galactopyranoside	271.3	−1.55	125.4	3
1efv	adenosine monophosphate	346.2	−3.47	182.9	4
1eiz	*S*-adenosylmethionine	399.5	−3.88	187.1	7
1ekx	*N*-(phosphonacetyl)-L-aspartic acid	251.1	−6.80	160.6	6
1f5v	flavin mononucleotide	454.3	−3.30	195.3	7
1f8y	5-methyl-2′-deoxypseudouridine	282.2	−1.44	99.1	2
1ftx	{1-[(3-hydroxy-methyl-5-phosphonooxymethyl-pyridin-4-ylmethyl)-amino]-ethyl}-phosphonic acid	350.2	−5.24	161.6	6
1g2o	1,4-dideoxy-4-aza-1-(*S*)-(9-deazahypoxanthin-9-yl)-D-ribitol	267.3	−2.79	134.6	2
1gvo	2,4-dinitrophenol	183.1	1.65	114.7	2
1hjz	2-(*N*-morpholino)-ethanesulfonic acid	195.2	−2.55	64.9	3
1iei	zenarestat	440.6	3.21	80.8	4
1j4i	4-methyl-2-{[4-(toluene-4-sulfonyl)- thiomorpholine-3-carbonyl]-amino}-pentanoic acid	413.5	0.15	106.6	7
1ja9	pyroquilon	173.2	1.52	20.3	0
1jak	(2R,3R,4S,5R)-2-acetamido-3,4-dihydroxy-5-hydroxymethyl-piperidinium	205.2	−3.64	106.4	2
1kew	thymidine-5′-diphosphate	400.2	−4.26	185.8	6
1kkp	Ser-PLP[e]	331.2	−3.83	166.3	7
1kl1	Gly-PLP[e]	301.2	−3.19	146.1	6
1kpg	*S*-adenosyl-L-homocysteine	384.4	−3.62	187.1	7
1kt7	Retinol	286.5	5.51	20.2	5
1l4g	nicotinate mononucleotide	333.2	−5.17	154.1	5
1l5k	1-alpha-D-ribofuranosyl-benzimiazole-5′-phosphate	329.2	−3.05	129.2	4
1l5l	7-alpha-D-ribofuranosyl-purine-5′-phosphate	331.2	−4.26	155.0	4
1l5m	7-alpha-D-ribofuranosyl-2-aminopurine-5′-phosphate	346.2	−4.68	181.0	4
1lc8	{3-[(3-hydroxy-2-methyl-5-phosphonooxymethyl-pyridin-4-ylmethyl)-amino]-2-methyl-propyl}-phosphonic acid	381.2	−4.36	167.6	8
1li4	neplanocin	261.2	−1.09	127.2	2
1lii	adenosine	267.2	−1.88	139.5	2
1m2g	adenosine-5-diphosphoribose	557.3	−6.59	285.2	9
1mm7	quisqualate	189.1	−3.58	126.4	3
1mqj	willardiine	199.2	−3.59	117.2	3
1mvc	4-[2-(5,5,8,8-tetramethyl-6,7-dihydronaphthalen-2-yl)-1,3-dioxolan-2-yl]benzoate (BMS649)	379.5	3.96	58.6	3
1n1t	2-deoxy-2,3-dehydro-*N*-acetyl-neuraminic acid	290.2	−4.40	159.4	5
1n3i	3-hydroxy-4-hydroxymethyl-1-(4-oxo-4,4a,5,7a-tetrahydro-3H-pyrrolo[3,2-d]pyrimidin-7-ylmethyl)-pyrrolidinium	265.3	−1.95	102.2	3
1nvr	staurosporine	467.5	4.00	74.03	2
1oh0	equilenin	266.3	3.94	37.3	0
1oio	*N*-acetyl-D-glucosamine	221.2	−3.08	119.3	2
1owe	6-[(Z)-amino(imino)methyl]-*N*-phenyl-2-naphthamide	290.3	1.56	80.7	3
1p4j	C-(1-hydrogyl-beta-D-glucopyranosyl) formamide	223.2	−4.37	153.5	2
1pfy	5′-O-[(L-methionyl)-sulfamoyl]adenosine	478.5	−3.36	219.4	9
1pg4	adenosine-5′-monophosphate-propyl ester	388.3	−2.04	171.9	7
1r0p	methyl(5S,6R,8R)-6-hydroxy-5-methyl-13-oxo-5,6,7,8,14,15-hexahydro-13H-5,8-epoxy-4b,8a,14-triazadibenzo[b,h]cycloocta[1,2,3,4-jkl]cyclopenta[e]-as-indacene-6-carboxylate (K-252A)	467.5	4.33	94.7	1
1sg0	resveratrol	228.2	2.97	60.7	2
1t40	2-[5-fluoro-2-[(4,5,7-trifluoro-1,3-benzothiazol-2-yl)methylcarbamoyl]phenoxy]ethanoic acid (IDD552)	413.3	3.26	91.4	6
1tkg	5′-O-(*N*-(L-seryl)-sulfamoyl)adenosine	434.4	−4.90	239.7	7
1uf5	*N*-carbamyl-D-methionine	191.2	−1.70	95.3	5
1urg	maltose	342.3	−5.40	189.5	4
1v7c	(2E)-2-[({3-hydroxy-2-methyl-5-[(phosphonooxy)methyl]pyridin-4-yl}methyl)amino]-5-phosphonopent-2-enoic acid	421.2	−4.78	197.3	10
1wap	L-tryptophan	204.2	−0.93	83.6	3
1wu6	xylobiose	282.2	−4.12	149.1	2
1xoz	tadalafil	389.4	2.31	74.9	1
1yki	nitrofurazone	198.1	0.19	126.4	3
2ay3	3-(3,4-dimethoxyphenyl)propionic acid	209.2	0.39	58.6	5
2prj	1-*N*-acetyl-beta-D-glucosamine	221.2	−3.08	119.3	2
2skc	glucose	180.2	−3.22	110.4	1

[a] Molecular weight. [b] The calculated log *P* value,[14] based on an atomic contribution model. [c] Polar surface area,[15] calculated using group contributions to approximate the polar surface area only from the connection table information. [d] Number of rotatable single bonds. Conjugated single bonds are not included. [e] PLP = pyridoxal phosphate.

atoms in the docked structure from the X-ray structure. For simplicity, the amino acids located within 6.0 Å from the non-hydrogen atoms of the small molecule were included in the calculations. The atomic coordinates of all atoms of the small molecule and hydrogen atoms of the selected amino acids were optimized. The dielectric constants were varied over a range of $\epsilon = 1-20$. In Figure 3, the effect of ϵ on the $N^+\cdots O^-$ distance is illustrated. The mean distances in 59

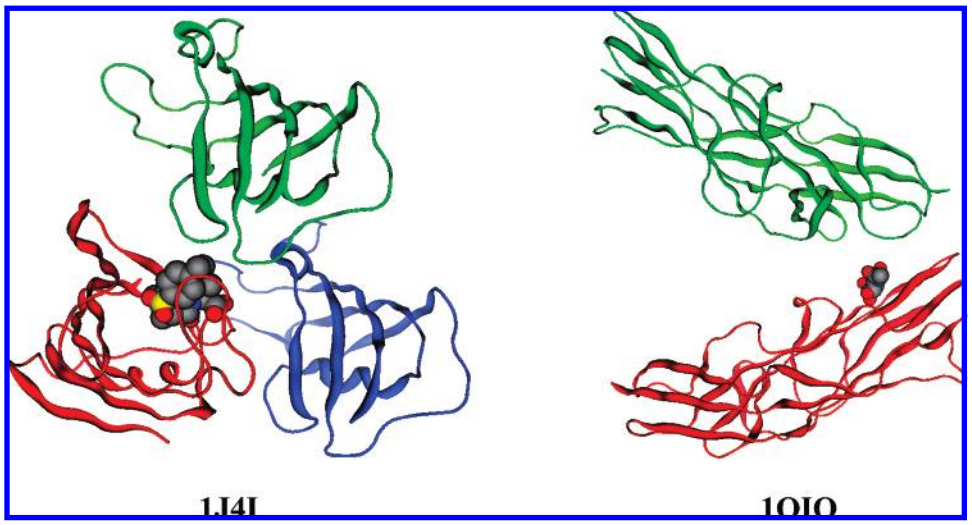

Figure 2. Small molecules entrapped between multiple proteins in the crystal structures.

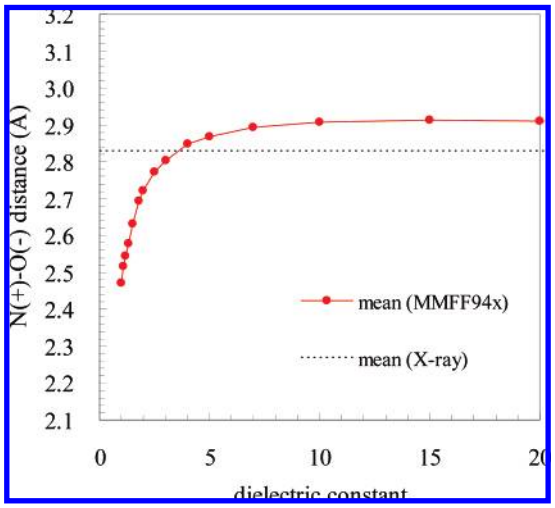

Figure 3. Effect of dielectric constant (ϵ) on the intermolecular $N^+\cdots O^-$ distances.

complex structures calculated using different values of ϵ were compared with the mean distance of 2.83 Å observed in 59 X-ray structures. If $\epsilon = 1.0$, the mean distance is <2.5 Å, and it is significantly shorter than that in the X-ray structures. The mean distance sharply increases with ϵ and approaches the experimental value. Above 7, it reaches a plateau. Judging from the curve, a dielectric constant of $\epsilon = 3-5$ is more preferable. It clearly suggests that ϵ values of <3 or >7 are not suitable, as long as we use MMFF94x as a potential function suite. We adopted a dielectric constant of $\epsilon = 5.0$ in this study. The balance of the absolute values of electrostatic and van der Waals interactions is generally a good rule of thumb to judge the validity of the nonbonded potential functions qualitatively. In Figure 4, the electrostatic and van der Waals interactions calculated with two different dielectric constants of 1.0 and 5.0 are compared. This figure shows the distribution of the values for 59 structures. Obviously, the balance is poor when 1.0 is applied for ϵ. The electrostatic interactions take mostly large negative values. On the other hand, the estimated van der Waals interactions are mostly positive. When 5.0 is applied for ϵ, however, the signs of the both interaction energy are mostly negative. It indicates that the ϵ value of 5.0 is more appropriate for the docking simulations. Since we apply a

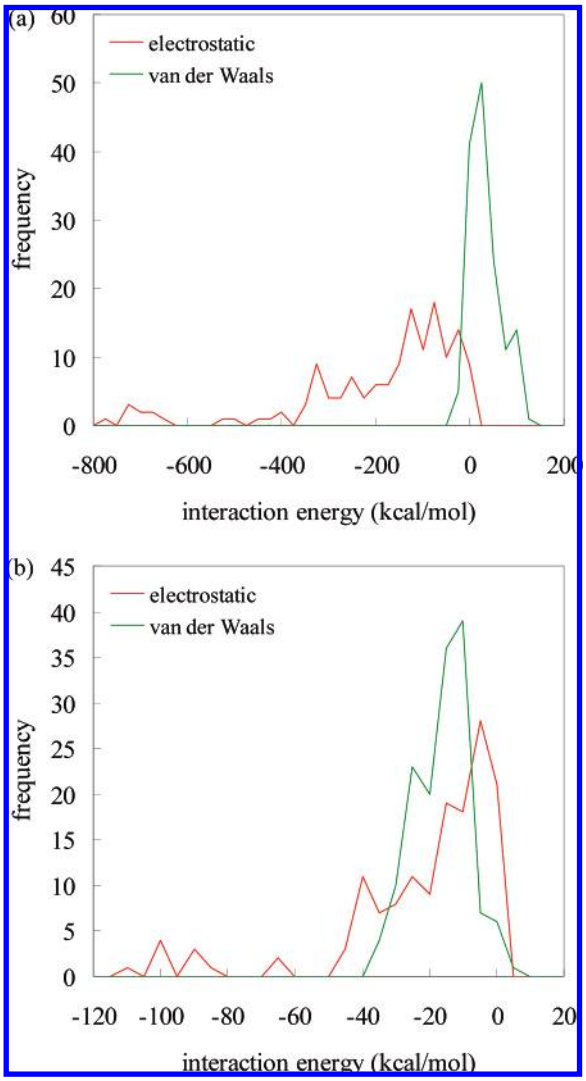

Figure 4. Effect of dielectric constant (ϵ) on the electrostatic and van der Waals interaction energies.

common dielectric constant for all 59 docking experiments, we thought further tuning of ϵ is not meaningful and 5.0 was used as the value of ϵ throughout this study. The relevance of this value can be validated by the docking performance.

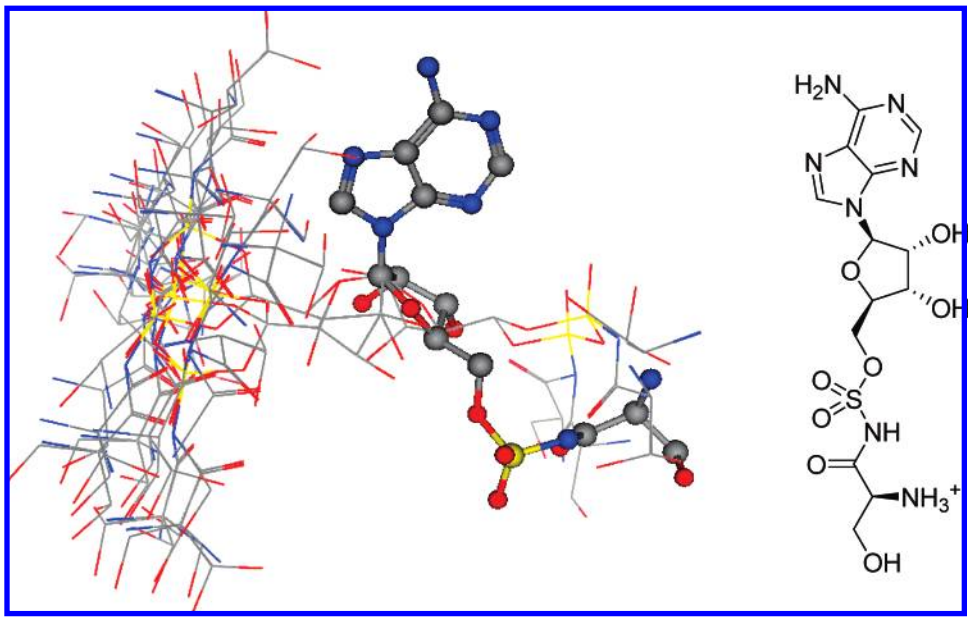

Figure 5. Ligand structures of 1TKG. Conformations generated by ASEDock (line) and the X-ray structure (ball-and-stick), as well as the chemical structure, are shown.

3. RESULTS AND DISCUSSION

3.1. Docking Results. A summary of the docking results is given in Table 1. In this study, the solution that gives the minimum U_{total} is assumed to be the best solution. The RMSD value that was used to evaluate the docking precision is derived from the best solution. The diffraction-component precision index (DPI) (hereafter denoted as σ) is "a good and rough guide" to coordinate the precision of experimentally obtained structures[16] and can be used to evaluate the reliability of the docking results. Suppose the standard uncertainty of the observed and predicted molecular model is the same in magnitude and is equal to σ, the estimated standard uncertainty of the RMSD between the corresponding non-hydrogen atoms in the observed and predicted molecules can be approximated to be $\sqrt{2}\sigma$.[7] The value of $2\sqrt{2}\sigma$ is given in this table as a guide of standard uncertainty of the RMSD. The RMSD values of 46 structures, which comprise ~78% of the entire dataset, are less than twice the corresponding estimated error of the observed structures. This indicates that the level of precision of the redocking results attained by ASEDock is mostly comparable to the expected experimental errors of the X-ray structures. An RMSD value of ≤ 2 Å is generally used as a rough guide to judge the correctness of the redocked structure. If we use this criterion as a measure of successful redocking, 58 out of 59 structures were redocked successfully. A success rate of ~98% was attained. This markedly high success rate demonstrates that if the high-quality X-ray structures of the complexes are available, ASEDock can precisely predict the dispositions and the structures of various small molecules at the concavities in the target proteins.

A dielectric constant of $\epsilon = 5.0$ was used in these docking experiments. The markedly good reconstruction of the X-ray structures strongly indicates that, in the concavities of these 59 structures, a dielectric constant of $\epsilon = 5.0$ is appropriate. Moreover, the high success rate of docking for various protein−ligand systems indicates that the dielectric constant of 5.0 should be generally acceptable in the docking simulation of druglike molecules at druggable concavities.

3.2. A Case of Failure. The redocking of 1TKG was in failure, and there may be multiple reasons for this failure. In Figure 5, the conformations of the ligand generated by ASEDock and the X-ray structure are compared. Obviously, ASEDock failed to generate a conformation that corresponds to the X-ray structure. The RMSD between the best generated conformation and the X-ray structure was 1.784 Å. The magnitude of the deviation seems to be beyond the radius of convergence by molecular mechanics calculations. However, it is intriguing that the very similar ligand of 1PFY was successfully redocked. In this case, an almost correct conformation with the RMSD of 0.906 Å was generated, as shown in Figure 6, and was successfully optimized to give the final RMSD of 0.504 Å. Although the ligand in 1TKG takes a more-folded conformation than that in 1PFY, most of the generated conformations are relatively extended. The generation method of conformation adopted in ASEDock is possibly slightly biased for this particular class of structure. It indicates that improvement of the method of conformation generation promises further improvement of the docking simulations.

3.3. Applications of ASEDock to the Structures for Which Ph4Dock Failed To Give Reasonable Answers. In our previous paper,[7] after a merging of the structures in a CCDC/Astex validation set[17] and a dataset proposed by Wang et al.,[18] we selected 43 structures for evaluation of Ph4Dock. Ph4Dock marked an appreciably good success rate of 86%, based on an rmsd of ≤ 2.0 Å, in redocking experiments. Nevertheless, Ph4Dock failed in redocking of six complexes, i.e., 1B9V, 1C1E, 1D3D, 1D3P, 1YEE, and 25C8. These structures are not included in the 59 structures, because they violate the criteria of the high quality used in this study. A summary of the docking results for these structures is given in Table 3. For all structures, the RMSDs were <2.0 Å, and the X-ray structures were reasonably reconstructed. The

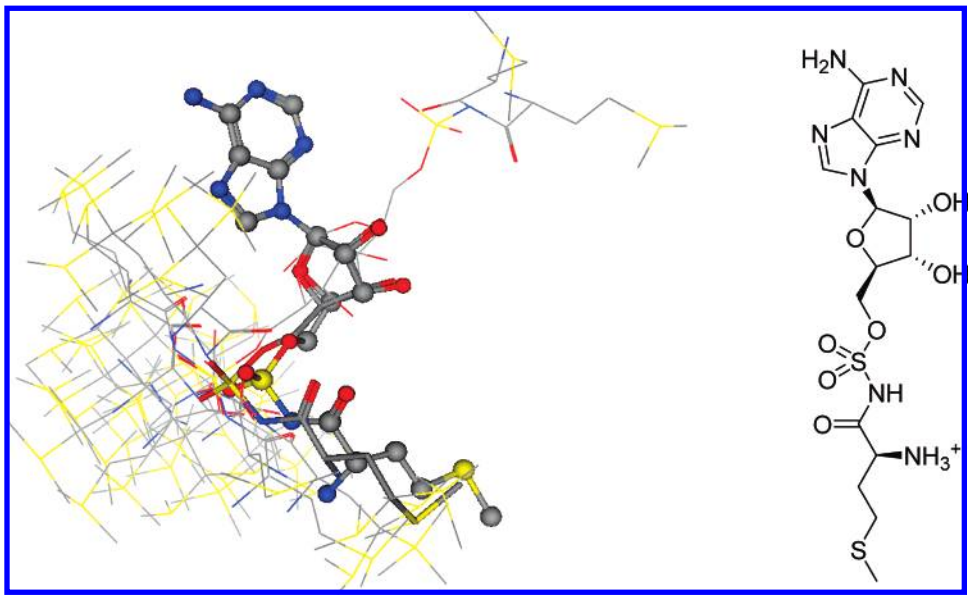

Figure 6. Ligand structures of 1PFY. Conformations generated by ASEDock (line) and the X-ray structure (ball-and-stick), as well as the chemical structure, are shown.

Table 3. Docking Results Obtained by ASEDock for 1B9V, 1C1E, 1D3D, 1D3P, 1YEE, and 25C8

PDB code	resolution (Å)	R_{free}	DPI (Å)	$2\sqrt{2}\sigma$ (Å)[a]	RMSD (Å)	lowest U_{total} (kcal/mol)
1B9V	2.35	0.271	0.467	1.321	0.945	48.6
1C1E	1.90	0.294	0.347	0.982	0.223	61.4
1D3D	2.04	0.223	0.288	0.815	1.940	62.4
1D3P	2.10	0.214	0.295	0.834	1.370	66.9
1YEE	2.20	0.260	0.141	0.398	0.993	−11.6
25C8	2.00	0.292	0.326	0.923	1.786	13.6

[a] $\sigma = \sigma(r, B_{avg})$.

RMSDs for 1B9V and 1C1E were $<2\sqrt{2}\sigma$, and their structures were rebuilt within experimental error. One example of the redocking results is shown for 1C1E in Figure 7. In this case, there are no hydrogen bonds between the ligand and the protein. This ligand is compact in shape and has no specific directionality. Therefore, it was difficult to pose the molecule in the concavity by Ph4Dock, which best uses pharmacophores as docking guides. As shown in this figure, the molecule well covers the portion where alpha spheres are densely populated. This example illustrates the crucial advantage of ASEDock that adapts the ASE score.

4. CONCLUSIONS

The successful redocking is an indispensable prerequisite to apply a docking algorithm to virtual screening of compounds that could specifically bind a target protein. In addition, such docking software is usually composed of several complex routines and a selection of appropriate parameters is important to obtain the optimum docking results. Therefore, the redocking experiments are essential in determining an appropriate set of parameters adapted to a particular problem.

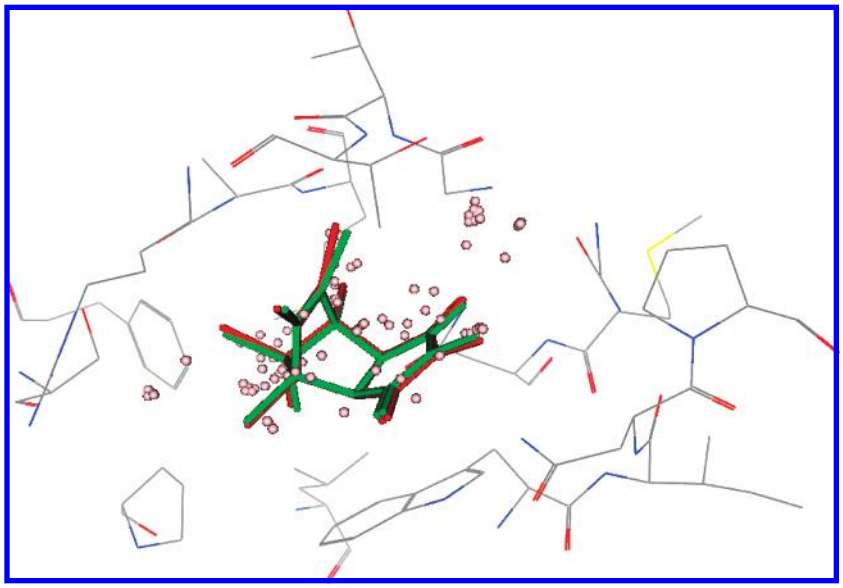

Figure 7. Redocked result of 1C1E. Alpha spheres calculated in the concavity are shown by small spheres. The ligand structure obtained by X-ray analysis and ASEDock are depicted by red and green sticks, respectively.

ASEDock is a novel docking program based on an accurate modeling of concavity shape and an effective function that evaluates the shape similarity between the model and the ligand. To evaluate the docking accuracy of ASEDock, redocking experiments for a selection of 59 highest-quality protein−ligand complexes from PDB were undertaken.

All the current docking algorithms involve several thorny problems, such as the quality of the potential functions, the accuracy of atomic charge, the generation of the ligand conformations, the effect of induced fit, and the effect of water molecules. Despite these inherent problems, however, the redocking performance of ASEDock is reasonably good, as far as the high-quality crystal structures are used. The markedly high success of ASEDock in redocking experiments clearly indicates that the most important factor governing the docking/binding process is shape complementarity.

ACKNOWLEDGMENT

This work was partially supported by the Grant-in-Aid for Scientific Research (A) (19200024) from MEXT (Ministry of Education, Culture, Sports, Science and Technology) for N.H.

REFERENCES AND NOTES

(1) Berman, H. M.; Westbrook, J.; Fenz, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(2) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(3) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Hey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(4) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(5) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A. Glide: A New Approach for Rapid Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(6) Jain, A. N. Surflex: Fully Automatic Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003**, *46*, 499−511.

(7) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock:Pharmacophore-Based Protein-Ligand Docking. *J. Med. Chem.* **2004**, 47, 6804−6811.

(8) MOE (Molecular Operating Environment), Version 2004.0301; Chemical Computing Group, Inc.: Montreal, Quebec, Canada, 2006.

(9) Wojciechowski, M.; Lesyng, B. Generalized Born Model: Analysis, Refinement and Applications to Proteins. *J. Phys. Chem. B* **2004**, *B108*, 18368−18376.

(10) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(11) Edelsbrunner, H.; Facello, M.; Fu, P.; Liang, J. Measuring Proteins and Voids in Proteins. In *Proceedings of the 28th Hawaii International Conference on System Sciences (HICSS'95)*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, 1995; Vol. 5, pp 256−264.

(12) Kearsley, S. K.; and Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.

(13) Horio, K.; Goto, J.; Muta, H.; Hirayama, N. A Simple Method to Improve the Odds in Finding 'Lead-like' Compounds from Chemical Libraries. *Chem. Pharm. Bull.* **2007**, *55*, 980−984.

(14) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868−873.

(15) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.

(16) Blow, D. M. A rearrangement of Cruickshank's formulae for the diffraction-component precision index. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *D58*, 792−797.

(17) Cambridge Crystallographic Data Centre. hhtp:/www.ccdc.cam.ac.uk/products/life_sciences/validate/astex/pdb_entries (accessed July 10, 2004).

(18) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.

CI700352Q