

Daisuke Takaya^{1,2}, Genki Terashi², Kazuhiko Kanou², Mayuko Takeda-Shitaka^{1,2}, Hideaki Umeyama^{1,2}

1) RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

2) Biomolecular Design, School of Pharmacy, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

takayad@gsc.riken.jp

1. Introduction

Many protein targets implicated in diseases have been discovered through biochemical experiments. As a result, the competition between pharmaceutical companies and other organizations to discover drug-like compounds which inhibit or activate those protein targets is fierce. Many pharmaceutical companies are using in-silico screening programs such as DOCK, AutoDock and GOLD. These programs use classical mechanical potentials. Here, we report a novel method, ChooseLD¹ (CHOOse biological information Semi-Empirically on the Ligand Docking), which uses simulated annealing (SA) based on bioinformatics for protein-ligand flexible docking.

2. Method

First, we placed a target protein of interest and aim to select one or more ligands with low molecular weight. The query amino acid sequence of the target protein is aligned in the filter of CE² Z-Score (more than 3.7) of Protein Data Bank (PDB) database, which includes ligand molecules, termed the family ligand set. Alignment methods such as PSI-BLAST are then applied. A fingerprint (FP) of a chemical descriptor is determined on the basis of bioinformatics. The FP includes the information about the atom-type such as used in SYBYL and bond-type; single, double, triple or aromatic. The FPs of the same type including different Cartesian

Fig 1. Defined Equation

$$\begin{aligned}
 FPAScore &= F(\text{aligned_fp}, \text{fp_rmsd}, \text{molecule}) \\
 &= \text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd}) \\
 &\quad \times \text{fp_volume}(\text{molecule}) \\
 &\quad \times \text{fp_contact_surface}(\text{molecule})
 \end{aligned} \tag{1}$$

$$\text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd}) = \frac{\text{RawScore}(\text{aligned_fp})}{1 + \ln(\text{fp_rmsd}^{k1} + 1)} \tag{2}$$

$$\text{fp_volume}(\text{molecule}) = \ln \frac{1.0 + n\text{afp}^{k2}}{1.0 + n\text{ap}^{k3}} \tag{3}$$

$$\text{fp_contact_surface}(\text{molecule}) = \frac{\sum_{i=1}^n \text{density_of_atom}(\text{atom}(i))}{\text{total_density_of_atom}(\text{molecule})} \tag{4}$$

coordinates of the docked ligand composes another Cartesian coordinates vector of the redundant FP set. FPAScore (FP Alignment Score) to calculate alignment between target ligand and library ligands is defined by equation (1). The FPAScore is used to determine the most stable docking conformation of the ligand. The value of

“ $RawScore(aligned_fp)$ ” is maximized in the SA calculation process accompanied by variation of the value of “ $\ln(fp_rmsd^{k1+1.0})$ ”. The fp_rmsd is the root mean square deviation (rmsd) value that is the result of the least-square fitting using the FP alignment. The $nafp$ is the number of docking ligand atoms covering the FPs region. The nap is the number of docking ligand atoms covering the target protein region. In the calculation of the benchmark sets, both $k2$ and $k3$ are equal to one. “ $atom(i)$ ” is the sequential atom number of a ligand. “ $density_of_atom(atom(i))$ ” shows the ligand atom number in the region of significance in the area of direct interaction. “ $total_density_of_atom(molecule)$ ” is the denominator for the ligand and is used to standardize the numerator

3. Results and Discussion

In order to test the protein-ligand docking method based upon bioinformatics, two benchmark tests were performed by using two database sets composed of either 85³ or 133⁴ PDB structures respectively. After the ligand molecule was docked to the target protein in the two benchmark sets, rmsd value of Cartesian coordinates between the docked ligand and the X-ray analyzed ligand was calculated. If the docking state is within 2.0 Å, docking is considered successful.

Tc range	Success rate %		Success rate %		
		Docking soft	Corina	MINI ⁴	average
0.08~0.16	12.6	DOCK	21.6	20.6	21.1
0.08~0.24	20.8	AutoDock	26.2	27.0	26.6
0.08~0.36	29.2	GOLD ChemScoreSTD	45.5	45.3	45.4
0.08~0.56	40.1	GOLD ChemScoreLib	44.1	44.9	44.5
0.08~0.76	44.8	GOLD GOLDScoreSTD	45.2	46.7	46.0
0.08~0.96	46.4				

Table 1: Success rates of ChooseLD, DOCK, AutoDock and GOLD.

Using the 85 benchmark set, the constant $k1$ value in equation (2) was optimized to be 4.0 by varying it from 1.0 to 6.0. In the docking calculations carried out for the 133 benchmark set using the $k1$ value of 4.0, the success rates of Tanimoto coefficient (Tc) ranges 0.08~0.16, 0.08~0.24, 0.08~0.36, 0.08~0.56, 0.08~0.76 and 0.08~0.96 were 12.6, 20.8, 29.2, 40.1, 44.8 and 46.4%, respectively. (Table 1)

Compared with the success rates using programs DOCK, AutoDock and GOLD, the success rates of ChooseLD are almost equivalent to those of the docking program. Our program is mainly based upon a bioinformatics basis set called FP. Thus, direct comparison of the success rates of those may be meaningless. Nevertheless, our program is comparably powerful when the researchers want to carry out protein-ligand docking and in-silico screening of a target protein. In the future, the number of PDB codes with the interacting ligand will be increased. It is anticipated that such an increase will improve the success rate of our ChooseLD method.

References

1. Takaya, D., et al *Chem Pharm Bull.* 56(5) 742-744. (2008)
2. Shindyalov, I. N., et al *Protein Engineering* 11, 739-747 (1998)
3. Hartshorn, M. J., et al *J. Med. Chem.* 50, 726-741 (2007).
4. Onodera, K., et al. *Chem. Inf. Model.* 47, 1609-1618 (2007)