A Genetic Algorithm Based Method for Molecular Docking

Chun-lian Li[1], Yu Sun[2], Dong-yun Long[1], and Xi-cheng Wang[3]

[1] School of Computer Science and Technology, Changchun University, Changchun 130022, China
[2] Institute of Special Education, Changchun University, Changchun 130022, China
[3] State Key Laboratory of Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116023, China
chdlichl@sina.com

Abstract. The essential of Molecular docking problem is to find the optimum conformation of ligand bound with the receptor at its active site. Most cases the optimum conformation has the lowest interaction energy. So the molecular docking problem can be treated as a minimization problem. An entropy-based evolution model for molecular docking is proposed in this paper. The model of molecular docking is based on a multi-population genetic algorithm. Two molecular docking processes are investigated to demonstrate the efficiency of the proposed model.

1 Introduction

The molecular docking problem is generally cast as a problem of finding the low-energy binding modes of a small molecule or ligand based on the "lock and key mechanism", within the active site of a macromolecule, or receptor, whose structure is known. It plays an important role in drug design, which is demonstrated by the vast amount of literature devoted to the optimization methods for molecular docking design since the pioneering work of Kuntz et al. [1]. Protein-ligand docking for drug molecular design is an ideal approach to virtual screening, i.e., to search large sets of compounds for putative new lead structure. A fundamental problem with molecular docking is that orientation space is very large and grows combinatorial with the number of degrees of freedom of the interacting molecules. Therefore, simpler and efficient methods are continuously being studied into. An entropy-based model of molecular docking is here presented, and a multi-population genetic algorithm is used to solve optimization problem of molecular conformation for protein-ligand docking.

L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3611, pp. 1159 – 1163, 2005.

2 Entropy-Based Genetic Model of Molecular Docking

Molecular docking is actually a problem of finding the conformation of ligand with the lowest interaction energy. An entropy-based evolutional model for molecular docking is constructed as follows

$$\begin{cases} \min \; -\sum_{j=1}^{m} p_j F(\mathbf{x}) \\ \min \; H = -\sum_{j=1}^{m} p_j \ln(p_j) \\ s.t. \sum_{j=1}^{m} p_j = 1, \; p_j \in [0\ 1] \end{cases} \tag{1}$$

where H is the information entropy, p_j is here defined as a probability that the optimal solution occurs in the population j. $F(\mathbf{x})$ is intermolecular interaction energy

$$F(\mathbf{x}) = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332.0 \frac{q_i q_j}{D r_{ij}} \right) \tag{2}$$

where each term is a double sum over ligand atoms i and receptor atoms j, r_{ij} is distance between atom i in ligand and atom j in receptor, A_{ij}, B_{ij} are Van der waals repulsion and attraction parameters, a, b are Van der waals repulsion and attraction exponents, q_i, q_j are point charges on atoms i and j, D is dielectric function, and 332.0 is factor that converts the electrostatic energy into kilocalories per mole. In the protein-ligand docking process, the binding free energy (2) should be transferred to related to the ligand atoms' Cartesian coordinates only to reduce the computing complexity. This method is based on the pre-calculated energy grid [2]. So as a matter of fact, the equation used in the real optimization can be expressed as follows [3,4]

$$\begin{aligned} &Min \quad F(\mathbf{x}) = f(T_x, T_y, T_z, R_x, R_y, R_z, T_{b1}, \cdots, T_{bn}) \\ &s.t. \quad \underline{X} \le T_x \le \overline{X} \\ &\quad\ \ \underline{Y} \le T_y \le \overline{Y} \\ &\quad\ \ \underline{Z} \le T_z \le \overline{Z} \\ &\quad\ \ -\pi \le angle \le \pi, \\ &\quad\ \ angle = R_x, R_y, R_z, T_{b1}, \cdots, T_{bn} \end{aligned} \tag{3}$$

where design variables $T_{b1}, \cdots, T_{bn}$ are the torsion angles of the rotatable bonds for flexible ligand docking, $T_x, T_y, T_z, R_x, R_y, R_z$, are the position coordinates and rotational angles of the anchor for the matching-based orientation search, and objective function E is intermolecular interaction energy.

It means that the optimal conformation of flexible ligand is formed by the translation (T_x, T_y, T_z), rotation (R_x, R_y, R_z) and the torsion motions (T_{bi}, i=1,2, ⋯, n, n is the number of torsion bonds). The former six variables are the six degrees of freedom for rigid body, it can also be seemed as the orientation of the ligand. T_{bi} is the angle of the ith flexible bond. For GA, each chromosome consists of the above three design variables that represent a ligand in a particular conformation and orientation. The design space of (T_x, T_y, T_z) must be limited in the spheres space of the receptor. A Circum-cuboid of the sphere space is here used to confine the rigid body's coordinates, which can greatly avoid the computational complexity of resolving the actual boundary. The rest variables are allowed to vary between $-\pi$ and π rad.

3 The Genetic Approach

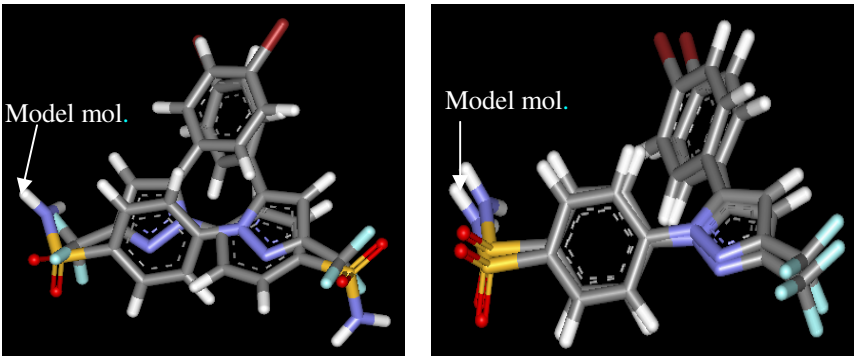
In this paper, an information entropy-based searching technique developed in prior work [5] is used to perform optimization. Design space is defined as initial searching space. M populations with N members are generated in the given space. After two generations are independently evolved in each population, searching space of each population except for the worst one is narrowed according to a coefficient calculated through the application of information entropy (See [5] for detail).

This GA is in binary coded and in each generation, there are three main genetic operators: selection, crossover and mutation. Selection is performed by an integer-decimal method. Crossover is executed by two-point rule, which firstly selecting two individuals as the parents from the current population, then selected randomly two cutting sites, swap 0s and 1s of the strings between the cutting sites of the mating pairs. As to another operator, a uniform mutation is employed to protect against the loss of some useful genetic information, and may help design to get out of local optimization solution. The process of mutation is to select simply a few strings from the population according to probability P_m and change the value of 0s or 1s on each chosen string in terms of some rule.

When taking the elitist maintaining, simple GA can convergent to its optimal solution at probability 1. So, an elitist maintaining mechanism is designed in the proposed GA. It is common for GA to involve a few decades and even more than several hundreds generations before finding the global solution. So a lot of historic information will be generated during the evolutionary process. Among them, we considered most is the elitist of all the populations in each generation. In this paper, the information of the best individual is recorded first in the former generation, then compare with the best one of current generation, and the absolute excellent individual up to now is stored in the contemporary as the elitist. Iterate this process in every generation till the convergence is reached. The final elitist is the solution to the optimization problem.

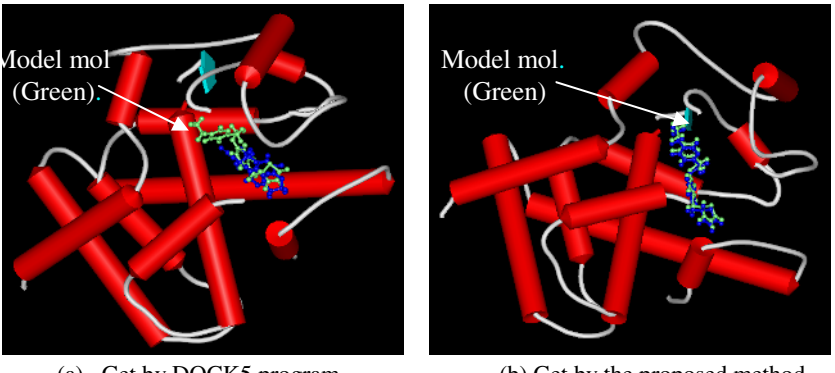
4 Application

To test the efficiency of the molecular docking, two docking examples are investigated. The first is, the protein-ligand docking process between sc-558 and its



(a) Get by DOCK5 program (b) Get by the proposed method

Fig. 1. Cox-2 result conformations against the model molecule



(a) Get by DOCK5 program (b) Get by the proposed method

Fig. 2. PPARλresult conformations against the model molecule and its receptor

receptor in known cyclooxygenase-2 (COX-2) inhibitors, and the second is one of the peroxisome proliferator-activated receptors (PPARs) –PPARλand small molecule: pioglitazone. In both cases, docking program runs at the same environment (Pentium® 4 CPU 2.4GHz, 256 MB ram, Red hat linux 7.1). In the first example, the docking results are: energy score of the best conformation obtained by the proposed method is –2. 0 kcal/mol with 2.91 seconds against the result of 187.7 kcal/mol with 68 seconds got by DOCK5.0.0 program [6]. In the second example, the docking results are: energy score of the best conformation obtained by the proposed method is -32.5 kcal/mol with 2.56s seconds against the result of –17 kcal/mol with 172.2 seconds got by DOCK5.0.0 program. Fig.1 and Fig.2 show the result conformations against the model molecule.

5 Conclusions

The use of docking as a virtual screening tool is more challenging than using it as a ligand design tool. It has to be developed so that docking algorithms can find the

correct binding mode of a compound by effectively sampling its available conformational and orientational space in the binding pocket within 10 seconds of central processing unit (CPU) time. To date, the best performing docking algorithms take about 1-3 min of CPU time for a ligand-protein docking experiment [7]. An evolutionary design model of molecular conformation for PLD is presented in this paper. The application examples show that the proposed design model and method are suitable for drug molecular design, and can get good accuracy and efficiency.

Acknowledgements

The authors gratefully acknowledge financial support for this work from the National Natural Science Foundation (10272030) and the Subsidized by the Special Funds for Major State Basic Research Project (G1999032805) of China. Additionally, appreciate the Kuntz group (UCSF) for supplying DOCK program.

References

1. I. D. Kuntz, J. M. Blaney, S. J. Oatley, and R. L. Langridge, A Geometric Approach to Macromolecule-Ligand Interactions. J. Mol. Biol., 161:269-288,1982.
2. E. C. Meng, B. K. Shoichet, I. D. Kuntz. Automated Docking with Grid-Based Energy Evaluation. J. Comput. Chem. **13(4)**:505-524,1992.
3. Li chun-lian, Wang Xi-cheng, Zhao Jin-cheng, Li Wen, Yu En-sheng, A Parallel Computing of Drug Molecular Docking Design, International Conference on Parallel Algorithms and Computing Environments (ICPACE)□2003:229-231
4. Li chunlian, Wang xicheng, Zhao jincheng, Drug molecular docking design based on optimal conformation search, Computers and Applied Chemistry, vol.21 (2), 2004
5. Li chun-lian, Wang Xi-cheng, Zhao Jin-cheng, et al. An Information Entropy-based Multi-population Genetic Algorithm. Journal of Dalian University of Technology, vol.44(4):589-593, 2004(in Chinese)
6. DOCK5.0.0. Demetri Moustakas. Kuntz Laboratory, UCSF, 4,15,2002.
7. I. Muegge, M. Rarey. Small molecular docking and scoring. In Reviews in computational chemistry, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 2001, Vol.17:1-60