Chinese Chemical Letters Vol. 10, No. 7, pp. 615–618, 1999

A Comparison of Three Heuristic Algorithms for Molecular Docking

Ting Jun HOU, Jun Mei WANG, Xiao Jie XU*

Department of Chemistry, Beida-Jiuyuan Molecular Design Laboratory, Peking University, Beijing 100871

Abstract: Three heuristic algorithms: simulated annealing, genetic algorithm, and Tabu search were compared to molecular docking procedure using 3 protein-ligand systems. Statistical analysis of the results indicated that the Tabu search showed the best performance in terms of locating solutions close to the crystallographic ligand conformation. From the comparisons, a hybrid search algorithm was proposed, which gave superior results compared with any one of the algorithms alone.

Keywords: Molecular docking, genetic algorithm, tabu search, simulated annealing.

The docking procedure between inhibitors and protein is a very sophisticated optimization problem; it is very difficult to carry out minimization using gradient methods such as the steepest descent method, Gauss-Newton method, which are very easy to fall into the local potential wells and very difficult to escape from them. So some heuristic methods have been introduced into the studies of molecular recognition. How to choose adequate optimization method in the docking procedure is critical to the calculation results.

In this paper, we describe the implementation and comparison of four search algorithms: random search (RS), simulated annealing (SA), genetic algorithm (GA), and Tabu search (TS) to molecular docking procedure. The algorithms were compared using three protein-ligand systems. From the statistical results, their search ability could be compared.

Methods

In this study, the docking method which was applied to compare the algorithms is the two-stage soft-docking procedure developed by us[1,2]. In our method, geometric complementarity and energetic complementarity are used as the score function to evaluate the binding mode between the receptor and the ligand. The geometric complementarity is evaluated by the score of the matched surface dot areas minus the unfavorable atomic pairs. The energetic complementarity is evaluated by the non-bonded energy between the receptor and the ligand. From the analysis of many cases, using

geometric complementariy is enough to determine the proper binding mode between the receptor and the ligand to a bound complex, and it has been proved that the minimum values of the geometric complementarity corresponded to the preferred binding mode of the ligand[1,2]. Moreover, from our previous study, it was proven that the score function of the geometric complementariy was smoother and simpler than the score function of the energetic complementarity. So using geometric complementarity, it is maybe get more better results when compare these three heuristic algorithms. So in this study, the geometric complementarity was used as the score function to evaluate these three heuristic search algorithms.

Simulated annealing and genetic algorithm are two well-used heuristic algorithms which have successfully appplied in some docking procedures. More recently, Tabu search has begun to attract attention as an effective heuristic search procedure for combinatorial optimazation problems in molecular design field[3]. David firstly applied this search method in docking procedure and proved it was very effective to find the proper binding mode[4]. But docking procedure is a very complicated minimization problem, it is very difficult to find an effective algorithm that can perform well in all conditions, the above three algorithms sometimes showed bad results in some cases. So for the use of a single algorithm, it is very difficult to solve a docking problem thoroughly.

The first goal of this study was to compare these three algorithms. But comparison of algorithms is not very easy. First, each category of algorithm has its own implementations, each of which will perform differently for a given optimization problem. Second, the performance of each algorithm depends on a set of adjustable operational parameters, and the quality of the results depends on if they are optimal for a given test case. So in this study, in order to compare these three algorithms fairly, the methods are all implemented as a traditional manner that is not modified or revised, moreover, the parameters applied in different algorithms are chosen these ones that are suitable in common cases.

All heuristic algorithms will contain stochastic elements, it is necessary therefore to assess performance statistically over a sufficiently large number of independent trials. To ensure a fairly comparison between algorithms, each one was limited to a maximum of 50000(±1%) function evaluations per docking. This number was chosen to be large enough for most algorithms to achieve convergence in most cases. When comparing the heuristic algorithms, the main quantity considered was the median score of the distribution of best scores obtained over 300 independent trials. In this study, the mean value of the maximum score of geometric complementarity was used as a descriptive statistical evaluation criterion to find the best docking mode. Moreover, we also compared algorithms according to their success rate proposed by Gehlhaar interquatile[5], the proportion of the trials which find a solution within 1.5Å (heavy atoms only) of the crystallographic ligand conformation.

The comparision of algorithms was carried out over three test cases as pecified in **Table 1**. These three complex systems comprise two protein-small molecule complexes, one protein-protein complex. The goal of this study was only to compare these three heuristic algorithms, so the systems selected are all bound protein complexes. In order to make certain simplification of the calculations, the flexibility of the ligand and the

receptor is ignored. The degrees of freedom are only six: three rotational degrees of freedom and three translational degrees of freedom of the ligand. Moreover, an active site for the receptor is defined in order to restrict the ligand docking in a small region. In our calculation, the active sites are all defined as a small box with 4×4×4Å in the calculations, the gravity center of the ligand must lie in this box.

Table 1. The test cases using in our calculation

Molecular names	Probe atoms[a]	Targe atoms	Probe dots[b]	Target dots
3DFR	33	1343	128	2957
4MBN	44	1294	150	2871
2PTC	454	1629	1184	3347

[a] The number of probe and targe atoms only represent the number of the heavy atoms.
[b] The probe radius to generate the connolly surface is defined as 1.5Å

Results and Discussion

Table 2 shows the results for the different algorithms on the test cases. It is obvious that RS performs very poorly. The low mean score of the geometric complementarity and success rate reflect the fact that RS is ineffectual in a search space of this size, and that is the reason that we want to use good heuristic algorithms in docking procedure. The success rate of 3DFR and 2WRP shows that GA performed best, TS performs a litter worse. But from the mean value of the 2WRP shows TS performs better than GA, so from the overall results, TS performs a little better in these cases. The reason is TS converge more slowly than GA, in some cases, it may already reach near the best solution, but it can not achieve the region that we have defined. Complex 2PTC is a protein-protein complex, from the mean value of the geometric complementarity and the success rate, the TS performs best. The contact surface between protein and protein is relatively large, so the score function of the geometric complementarity is more complex than that between protein and small molecule or protein and peptide. In this case, GA can fall into local minima easily and can not escape from it sometimes. But TS can do much better, it can avoid falling into local minima effectively.

From the comparison of these three algorithms, GA and TS both have its merits and shortages. GA converges faster, when near the best solution, it can find it very quickly, but GA can fall into local minima very easily. In contrast with GA, TS can avoid falling into local minima, but it converges relatively slower. So according to their merits and shortages, we propose a hybrid algorithm (HA), which combined GA and TS together. The basic procedure of the hybrid algorithm is similar with TS, but compared with traditional TS, it is different in two points. The first difference is when after N possible moves from the current solution, some extra steps of crossover and mutation operations, which come from GA, are added. The second modification is after every crossover operation, N new solutions are ranked, and the best several solutions were compared with the solutions in the Tabu list to check if they are Tabu, if they are, these Tabu solutions are replaced by new solutions generated randomly. The new hybrid algorithm holds the merits of GA and TS at the same time; it not only converges fast, but also does not fall

into local minima easily. We also applied the hybrid algorithm on these three test cases, the calculation results showed this new algorithm did much better than three heuristic algorithms alone.

Table 2. Docking results for the test cases given in **Table 1**

PDB code	Algorithm	Maximum score	rms/rmd(ligand)[a]	Mean score	Success rate(%)
3DFR	SA	1309.04	1.06	625.31	21
3DFR	GA	1340.37	1.01	701.51	33
3DFR	TS	1178.05	0.76	812.45	31
3DFR	RS	508.32	5.51	405.10	0
3DFR	HA	1268.11	1.12	1102.23	81
4MBN	SA	1181.34	1.36	657.12	37
4MBN	GA	1230.45	0.77	724.50	51
4MBN	TS	1143.49	0.88	756.23	42
4MBN	RS	876.54	2.35	489.23	2
4MBN	HA	1223.23	0.68	1108.90	96
2PTC	SA	2177.14	0.70	1703.23	59
2PTC	GA	2184.62	0.78	1650.45	50
2PTC	TS	2194.56	1.07	1745.56	48
2PTC	RS	1935.87	1.89	1345.70	4
2PTC	HA	2209.67	1.09	1956.98	100

[a] rmd means root mean distance, All the distances of atomic pairs of the ligand molecule have been calculated. The root mean of the different of relative distances (rmd) of bound complexes is used to evaluate the effectivity of the soft-docking calculations.

Acknowledgment

This work is supported by NCSF.

References

1. J.M. Wang, X.J. Xu, F. Jiang, The 1996 Chinese Peptide Symp. **1996**, 106.
2. J.M. Wang, T.J. Hou, X.J. Xu, *Chemometrics and Intelligent Laboratory Systems*, (In press).
3. V., Kvasnicka, Pospichal, J., *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 1109.
4. R.W. David, *et al. Jounal of Computer-Aided Molecular Design*, **1997**, *11*, 209.
5. D.K. Gehlhaar, *et al.*, *Chem. Bio.*, **1995**, *2*, 317.

Received 26 January 1999