An Evolutionary Approach for Molecular Docking

Jinn-Moon Yang

Department of Biological Science and Technology
National Chiao Tung University, Hsinchu, 30050, Taiwan
moon@csie.ntu.edu.tw

Abstract. We have developed an evolutionary approach for the flexible docking that is now an important component of a rational drug design. This automatic docking tool, referred to as the GEMDOCK (Generic Evolutionary Method for DOCKing molecules), combines both global and local search strategies search mechanisms. GEMDOCK used a simple scoring function to recognize compounds by minimizing the energy of molecular interactions. The interactive types of atoms between ligands and proteins of our linear scoring function consist only hydrogen-bonding and steric terms. GEMDOCK has been tested on a diverse dataset of 100 protein-ligand complexes from Protein Data Bank. In total 76% of these complexes, it obtained docked ligand conformations with root mean square derivations (RMSD) to the crystal ligand structures less than 2.0 Å when the ligand was docked back into the binding site. Experiments shows that the scoring function is simple and efficiently discriminates between native and non-native docked conformations. This study suggests that GEMDOCK is a useful tool for molecular recognition and is a potential docking tool for protein structure variations.

1 Introduction

The molecular docking problem is the prediction of a ligand conformation and orientation relative to the active site of a target protein. A computer-aided docking process, identifying the lead compounds by minimizing the energy of intermolecular interactions, is an important approach for structure-based drug designs [1]. Solving a molecular docking problem involves two critical elements: a good scoring function and an efficient algorithm for searching conformation and orientation spaces.

A good scoring function should be fast and simple for screening large potential solutions and effectively discriminating between correct binding states and non-native docked conformations. Various scoring functions have been developed for calculating binding free energy, including knowledge-based [2], physic-based [3], and solvent-based scoring functions [4]. In general the binding energy landscapes of these scoring functions are often complex and rugged funnel shapes [5].

Many automated docking approaches have been developed and can be roughly divided into rigid docking, flexible ligand docking, and protein flexible docking methods. The rigid-docking methods, such as DOCK program [6], treated both ligands and proteins as rigid. In contrast the ligand is flexible and the protein is rigid for flexible ligand docking methods including evolutionary algorithms [7,8,9,10], simulated annealing [11], fragment-based approach [12], and other algorithms. For reasonably addressing protein flexible problems, which both ligands and proteins are flexible, most of docking methods

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, pp. 2372–2383, 2003.

often allowed a limited model of protein variations, such as the side-chain flexible or small motions of loops in the binding site [13]. Most of these previous docking methods studied on a small test set (< 20 complexes), by contrast, the GOLD [8] and FlexX [12] were tested on a test set of over 100 complexes.

In this paper, we proposed an automatic program, GEMDOCK (Generic Evolutionary Method for DOCKing molecules), for docking flexible molecules. Our program used a simplified scoring function and a new evolutionary approach which is more robust than standard evolutionary approaches [14,15,16] on some specific domains [17,18,19,20]. Our energy function consisted only of steric and hydrogen-bonding terms with a linear model which was simple and fast enough to recognize potential complexes. In order to balance exploration and exploitation, the core idea of our evolutionary approach is to design multiple operators cooperating with each other by using the family competition which is similar to a local search procedure. We have successfully applied a similar idea to solve optimization problems in some differing fields [17,18,19,20].

In order to evaluate the performance and limitations of GEMDOCK on docking flexible ligands, we have tested it on a diverse dataset of 100 complexes from the Protein Data Bank. GEMDOCK achieved 76 ligands whose structures with RMSD values to the ligand crystal structures are less than 2.0Å. The rate increases to 86% when the structure water is considered. GOLD [8] achieved a 71% success rate in the same dataset and FlexX [12] achieved a 70% success rate on a dataset of 200 complexes extended from the data set of GOLD.

2 Method

The basic structure of the GEMDOCK (Figure 1) is as follows: Randomly generate a starting population with N solutions by initializing the orientation and conformation of the ligand relating to the center of the receptor. Each solution is represented as a set of four n-dimensional vectors $(x^i, \sigma^i, \upsilon^i, \psi^i)$, where n is the number of adjustable variables of a docking system and $i = 1, \ldots, N$ where N is the population size. The vector x represents the adjustable variables to be optimized in which x_1, x_2, and x_3 are the 3-dimensional location of the ligand; x_4, x_5, and x_6 are the rotational angles; and from x_7 to x_n are the twisting angles of the rotatable bonds inside the ligand. σ, υ, and ψ are the step-size vectors of decreasing-based Gaussian mutation, self-adaptive Gaussian mutation, and self-adaptive Cauchy mutation. In other words, each solution x is associated with some parameters for step-size control. The initial values of x_1, x_2, and x_3 are randomly chosen from the feasible box, and the others, from x_4 to x_n, are randomly chosen from 0 to 2π in radians. The initial step sizes σ is 0.8 and υ and ψ are 0.2. After GEMDOCK initializes the solutions, GEMDOCK enters the main evolutionary loop which consists of three main stages in every iteration: decreasing-based Gaussian mutation, self-adaptive Gaussian mutation, and self-adaptive Gaussian mutation. Each stage is realized by generating a new quasi-population (with N solutions) as the parent of the next stage. As shown in Figure 1, these stages apply a general procedure "FC_adaptive" with only different working population and the mutation operator.

The FC_adaptive procedure (Figure 1) employs two parameters, namely, the working population (P, with N solutions) and mutation operator (M), to generate a new quasi-population. The main work of FC_adaptive is to produce offspring and then conduct the

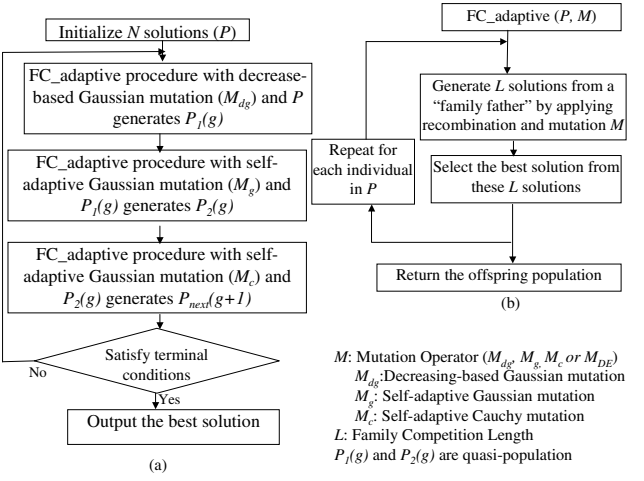


Fig. 1. Overview of GEMDOCK for molecular docking: (a) Main procedure (b) FC_adaptive procedure.

family competition. Each individual in the population sequentially becomes the "family father." With a probability p_c, this family father and another solution that is randomly chosen from the rest of the parent population are used as parents for a recombination operation. Then the new offspring or the family father (if the recombination is not conducted) is operated on by a mutation. For each family father, such a procedure is repeated L times called the family competition length. Among these L offspring and the family father, only the one with the lowest scoring function value survives. Since we create L children from one "family father" and perform a selection, this is a family competition strategy. This method avoids the population prematureness but also keeps the spirit of local searches. Finally, the FC_adaptive procedure generates N solutions because it forces each solution of the working population to have one final offspring.

In the following, genetic operators are briefly described. We use $a = (x^a, \sigma^a, v^a, \psi^a)$ to represent the "family father" and $b = (x^b, \sigma^b, v^b, \psi^b)$ as another parent. The offspring of each operation is represented as $c = (x^c, \sigma^c, v^c, \psi^c)$. The symbol x_j^s is used to denote the jth adjustable optimization variable of a solution s, $\forall j \in \{1, \ldots, n\}$.

2.1 Recombination Operators

GEMDOCK implemented modified discrete recombination and intermediate recombination [15]. A recombination operator selected the "family father (a)" and another solution (b) randomly selected from the working population. The former generates a child as follows:

$$x_j^c = \begin{cases} x_j^a \text{ with probability } 0.8 \\ x_j^b \text{ with probability } 0.2. \end{cases} \quad (1)$$

The generated child inherits genes from the "family father" with a higher probability 0.8. Intermediate recombination works as:

$$w_j^c = w_j^a + \beta(w_j^b - w_j^a)/2, \tag{2}$$

where w is σ, v, or ψ based on the mutation operator applied in the FC_adaptive procedure. The intermediate recombination only operated on step-size vectors and the modified discrete recombination was used for adjustable vectors (x).

2.2 Mutation Operators

After the recombination, a mutation operator, the main operator of GEMDOCK, is applied to mutate adjustable variables (x).

Gaussian and Cauchy Mutations: Gaussian and Cauchy Mutations are accomplished by first mutating the step size (w) and then mutating the adjustable variable x:

$$w'_j = w'_j A(\cdot), \tag{3}$$
$$x'_j = x_j + w'_j D(\cdot), \tag{4}$$

where w_j and x_j are the ith component of w and x, respectively, and w_j is the respective step size of the x_j where w is σ, v, or ψ. If the mutation is a self-adaptive mutation, $A(\cdot)$ is evaluated as $\exp[\tau' N(0,1) + \tau N_j(0,1)]$ where $N(0,1)$ is the standard normal distribution, $N_j(0,1)$ is a new value with distribution $N(0,1)$ that must be regenerated for each index j. When the mutation is a decreasing-based mutation $A(\cdot)$ is defined as a fixed decreasing rate $\gamma = 0.95$. $D(\cdot)$ is evaluated as $N(0,1)$ or $C(1)$ if the mutation is, respectively, Gaussian mutation or Cauchy mutation. For example, the self-adaptive Cauchy mutation is defined as

$$\psi_j^c = \psi_j^a \exp[\tau' N(0,1) + \tau N_j(0,1)], \tag{5}$$
$$x_j^c = x_j^a + \psi_j^c C_j(t). \tag{6}$$

We set τ and τ' to $(\sqrt{2n})^{-1}$ and $(\sqrt{2\sqrt{n}})^{-1}$, respectively, according to the suggestion of evolution strategies [15]. A random variable is said to have the Cauchy distribution ($C(t)$) if it has the density function: $f(y;t) = \frac{t/\pi}{t^2+y^2}$, $-\infty < y < \infty$. In this paper t is set to 1. The formulation of the self-adaptive Gaussian mutation is similar to the self-adaptive Cauchy mutation and is given

$$v_j^c = v_j^a \exp[\tau' N(0,1) + \tau N_j(0,1)], \tag{7}$$
$$x_j^c = x_j^a + v_j^c N_j(0,1). \tag{8}$$

Our decreasing-based Gaussian mutation uses the step-size vector σ with a fixed decreasing rate $\gamma = 0.95$ and works as

$$\sigma^c = \gamma \sigma^a, \tag{9}$$
$$x_j^c = x_j^a + \sigma^c N_j(0,1). \tag{10}$$

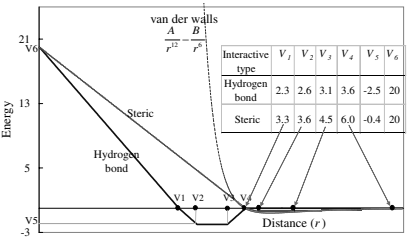


Fig. 2. The linear energy function of the pair-wise atoms for steric and hydrogen bonds in GEM-DOCK (bold line) with a standard Lennard- Jones potential (light line).

Rotamer-Mutation: This operator is only used for x_7 to x_n to find the conformations of the rotatable bonds inside the ligand. For each ligand, this operator mutates all of the rotatable angles according to the rotamer distribution and works as:

$$x_j = r_{ki} \text{ with probability } p_{ki}, \tag{11}$$

where r_{ki} and p_{ki} are the angle value and the probability, respectively, of ith rotamer of kth bond type including $sp^3 - sp^3$ and $sp^3 - sp^2$ bond. The values of r_{ki} and p_{ki} are based on the energy distributions of these two bond types.

2.3 Scoring Function

In this work, we used a simple scoring function given as

$$E_{tot} = E_{inter} + E_{intra} + E_{penal}, \tag{12}$$

where E_{inter} and E_{intra} are the intermolecular and intramolecular energy, respectively, E_{penal} is a large penalty value if the ligand is out of range of the search box. In this paper, E_{penal} is set to 10000.

The intermolecular energy is defined as

$$E_{inter} = \sum_{i=1}^{lig} \sum_{j=1}^{pro} \left[F(r_{ij}^{B_{ij}}) + 332.0 \frac{q_i q_j}{4 r_{ij}} \right], \tag{13}$$

where r_{ij} is the distance between the atoms i and j, q_i and q_j are the formal charges and 332.0 is a factor that converts the electrostatic energy into kilocalories per mole. The lig and pro denote the numbers of the heavy atoms in the ligand and receptor, respectively. $F(r_{ij}^{B_{ij}})$ is a simple atomic pair-wise potential function (Figure 2) modified from previous works [7,21] and given as

Table 1. Atom types of GEMDOCK

Atom type	Heavy atom name
Donor	primary and secondary amines, sulfur, and metal atoms
Acceptor	oxygen and nitrogen with no bound hydrogen
Both	structural water and hydroxy1 groups
Nonpolar	other atoms (such as carbon and phosphorus)

$$F(r_{ij}^{B_{ij}}) = \begin{cases} V_6 - \frac{V_6 r_{ij}^{B_{ij}}}{V_1} & \text{if } r_{ij}^{B_{ij}} \leq V_1 \\ \frac{V_5(r_{ij}^{B_{ij}} - V_1)}{V_2 - V1} & \text{if } V_1 < r_{ij}^{B_{ij}} \leq V_2 \\ V_5 & \text{if } V_2 < r_{ij}^{B_{ij}} \leq V_3 \\ V_5 - \frac{V_5(r_{ij}^{B_{ij}} - V_3)}{V_4 - V_3} & \text{if } V_3 < r_{ij}^{B_{ij}} \leq V_4 \\ 0 & \text{if } r_{ij}^{B_{ij}} > V_4 \end{cases} . \quad (14)$$

$r_{ij}^{B_{ij}}$ is the distance between the atoms i and j with bond type B_{ij} which is the interaction bonding type forming by the pair-wise heavy atoms of a ligand and a protein. B_{ij} is either hydrogen binding or steric state. The values of parameters, $V_1, \ldots, V_6$, are given in Figure 2. In this atomic pair-wise model, the interactive types are only hydrogen binding and steric potential which have the same function form but with different parameters, $V_1, \ldots, V_6$. The energy value of hydrogen binding should be larger than the one of steric potential. In this model, the atom is divided into four different atom types (Table 1) : donor, acceptor, both, and nonplar. The hydrogen binding can be formed by the following pair atom types: donor-acceptor (or acceptor-donor), donor-both (or both-donor), acceptor-both (or both-acceptor), and both-both. Other pair-atom combinations are to form the steric state.

The intramolecular energy of a ligand is

$$E_{intra} = \sum_{i=1}^{lig} \sum_{j=i+2}^{lig} F(r_{ij}^{B_{ij}}) + \sum_{k=1}^{dihed} A[1 - \cos(m\theta_k - \theta_0)], \quad (15)$$

where $F(r_{ij}^{B_{ij}})$ is defined as Equation 14 except the value is set to 1000 when $r_{ij}^{B_{ij}} <$ 2.0 Å and $dihed$ is the number of rotatable bonds. We followed the work of Gehlhaar et al. (1995) to set the values of A, m, and θ_0. For the $sp^3 - sp^3$ bond A, m, and θ_0 are set to 3.0, 3, and π; and $A = 1.5$, $m = 6$, and $\theta_0 = 0$ for the $sp^3 - sp^2$ bond.

3 Results

3.1 Parameters of GEMDOCK

Table 2 indicates the setting of GEMDOCK parameters, such as initial step sizes, family competition length ($L = 3$), population size ($N = 200$), and recombination probability ($p_c = 0.3$) in this work. The GEMDOCK optimization stops when either the convergence is below certain threshold value or the iterations exceed a maximal preset value which

Table 2. Parameters of GEMDOCK

Parameter	Value of parameters
Initial step sizes	$\sigma = 0.8, \upsilon = \psi = 0.2$ (in radius)
Family competition length	L = 3
Population size	$N = 200$
Recombination rate	p_c =0.3
# of the maximum generation	100

was set to 100. Therefore, GEMDOCK generated 2400 solutions in one generation and terminated after it exhausted 240000 solutions in the worse case. These parameters were decided after experiments conducted to recognize complexes of test docking systems with various values.

3.2 Test Data Set

In order to evaluate the strength and limitation of GEMDOCK, we tested it on a highly diverse dataset of 100 protein-ligand complexes proposed by Jones et al. [8] (Table 3). The ligand input files were generated by GENLIG which assigned the formal charge and atom type (donor, acceptor, both, or nonplar) of each atom and the bond type ($sp^3 - sp^3$, $sp^3 - sp^2$, or others) of a rotatable bond inside a ligand. These materials were used in Equation 12 to calculate the scoring value of a solution. Table 3 shows the ligand summary, including the minimum, average, and maximum values of the number of rotatable bonds and the number of heavy atoms.

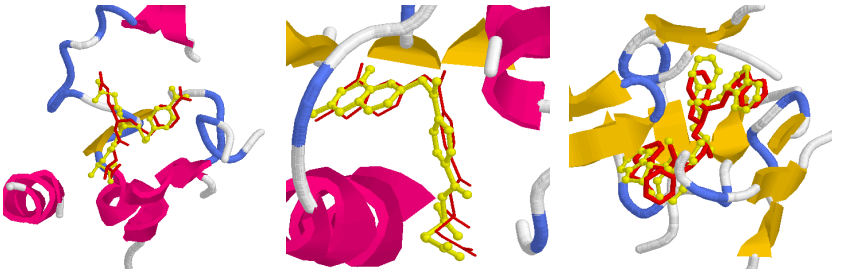
When preparing the proteins, we removed all structural water molecules and metal atoms except we discussed the influence of considering these hetero atoms. In order to decide the size of active site, GEMDOCK was tested on four different sizes (d Å): 6Å, 8Å, 10Å, and 12Å. The size with d Å means that all protein atoms in the active site are selected if they are located less than d Å apart from each ligand atom. GEMDOCK automatically decide the cube of a binding site based on the maximum and minimum of coordinates of these selected protein atoms. Experiments shows that GEMDOCK had little influence on the different sizes. In this paper, the distance d is set to 10Å when a ligand is docked back into the active site. Among these 100 test systems, the minimum cube is 23Å×24Å×20Å (2mcp) and the maximum cube is 41Å×40Å×30Å (2r07).

3.3 Results on the Dataset of 100 Complexes

GEMDOCK executed 10 independent runs for each complex. The solution with lowest scoring function was then compared with the observed ligand crystal structure. Table 3 shows the summary information and performance. We based the results on root mean square deviation (RMSD) error in ligand heavy atoms between the docked conformation and the crystal structure. By contrast, Jones et al [8] used four subjective categories (good, close, error, and wrong) to evaluate the performance. Because they found all of good and close solutions with RMSD were below 2.0 Å, we considered that a docking result is acceptable if the RMSD value is less than 2.0 Å [8,12].

Table 3. GEMDOCK results and summary of ligands in the dataset of 100 complexes

RMSD(Å)	rank 1	any rank	PDB code with rank 1
≤ 0.5	24	41	1abe 1acm 1aco 1did 1epb 1fki 1hdy 1hsl 1ida 1lst 1pbd 1pha 1rob 1stp 1tpp 2ada 2cgr 2cht 2dbl 3aah 3tpi 4phv 6abp 6rsa
$> 0.5, \leq 1.0$	39	32	1acj 1ack 1acl 1aha 1dbb 1dbj 1die 1dr1 1dwd 1eap 1etr 1fkg 1ghb 1hri 1hyt 1ldm 1lic 1mrk 1nis 1phd 1phg 1rds 1slt 1srj 1tka 1tmn 1ulb 1glq 2ak3 2ctc 2mcp 2pk4 2r07 2sim 3cpa 3hvt 4cts 4dfr 4est
$> 1.0, \leq 1.5$	10	11	1aaq 1apt 1cbx 1cps 1hdc 1icn 1poc 4fab 5p2p 8gch
$> 1.5, \leq 2.0$	3	6	1aec 1tdb 3ptb
$> 2.0, \leq 2.5$	1	0	1mdr
$> 2.5, \leq 3.0$	3	6	1ase 1blh 2yhx
> 3.0	20	10	1eta 1eed 1azm 1rne 6rnt 1mcr 2mth 1mup 2plv 1baf 1ive 2phh 3gch 1hef 1igj 1coy 1xid 1xie 3cla 7tim

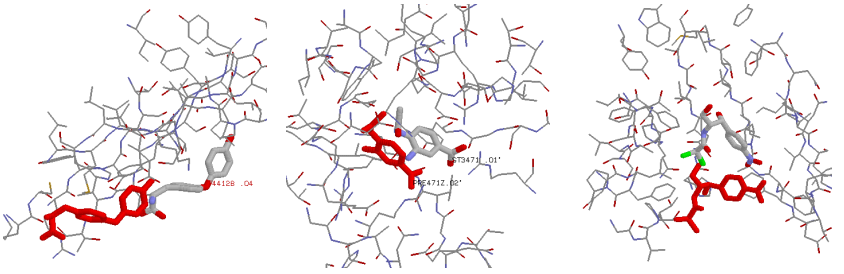


(a) 1glq (b) 4dfr (c) 4phv

Fig. 3. Acceptable docking examples: The docked ligand conformation (red) is similar to the crystal ligand structure (yellow). The RMSD values are 0.78 Å for 1glq, 0.56 Å for 4dfr, and 0.42 Å for 4phv.

Table 3 shows that GEMDOCK achieved a 76% success rate in identifying the experimental binding model if the solutions at the first rank are considered. The RMSD values of 63 complexes are less than 1.0 Å. This rate further rises to 84% based on the solutions with any rank. The performance of GEMDOCK was little influenced by the number of rotatable bonds and the number of heavy atoms of a ligand. When the structural water and metal atoms are considered, the success rate is improved to 86% with the first rank. On average GEMDOCK took 305 seconds for a docking run on Pentium 1.4 GHz personal computer with single processor. The maximum time was 883 seconds for the complex, 1rne, and the shortest time was 102 seconds for 2pk4. In the following, we discussed some acceptable examples and unacceptable examples.

Figure 3 shows three typical acceptable solutions in which GEMDOCK predicted correct positions of all ligand groups. The predicted ligand is red and crystal ligand is yellow. All of these three examples are identical with the crystal structures and the RMSD values are 0.78 Å (1glq: nitrophenyl ligand for glutathione S-transferase), 0.56 Å



(a) 1eta (b) 1ive (c) 3cla

Fig. 4. Results of three poor docking examples. The docked ligand conformation is red and the crystal ligand structure is cpk. The RMSD values are 3.62 Å for 1eta, 6.32 Å for 1ive, and 7.56 Å for 3cla.

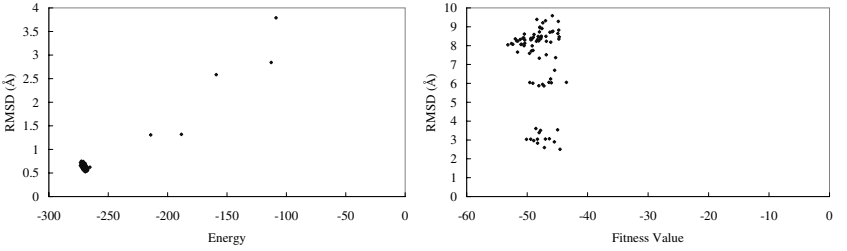
(4dfr: methotrexate ligand for dihydrofolate reductase), and 0.42 Å (4phv: peptide-like ligand for HIV-1 protease).

3.4 Examples of Unacceptable Solutions

Table 3 shows 24 unacceptable docking complexes with RMSD values more than 2.0 Å. Three poor examples are shown in Figure 4 in which the structures of predicted ligands and crystal ligands are displayed with red and cpk color, respectively. The RMSD values are 3.62 Å for 1eta (tetraiodo L-thyronine ligand for transthyretin), 6.32 Å for 1ive (acetylamino ligand for influenza), and 7.56 Å for 3cla (chloramphenicol ligand for Type III chloramphenicol acetyltransferase).

We have analyzed these poor examples to understand why GEMDOCK failed to recognize the binding models by using numerical experiments. These experiments were based on three main factors: the scoring functions, the docking materials, and the search methods. For the scoring functions we tested various uses and parameter values (Equation 12) on the dataset of 100 complexes. According to our experimental results, the E_{inter} was the main factor in our system, the E_{intra} and E_{penal} were minor factors that influenced some specific docking cases. The element, $F(r_{ij}^{B_{ij}})$, of the E_{inter} (Equation 13) dominated the performance. Figure 5 shows the relationship between the RMSD values and scoring values with 100 independent runs. For the good docking example (4dfr) 95 solutions with RMSD values are less than 1.0 Å and the scoring value is similar in each run (Figure 5(a)). By contrast, for a poor example the RMSD value is more than 3.0 Å and the score is diverse (Figure 5(b)). These experiments indicate that our scoring function seems to be simple and fast to discriminate native binding state and non-native docked conformations for 90% testing complexes.

For the docking materials we have discussed the influences of the sizes of the binding site (see Subsection Test data set) and of the hetero atoms in the binding site. Among these 100 complexes there are 17 proteins with metal atoms, and 84 proteins with structural water atoms. When the metal atoms are included, GEMDOCK can consistently im-



(a) good examples- 4dfr (b) poor example- 1ive

Fig. 5. Typical relationship between the values of the scoring function and the RMSD on 100 independent runs. (a) For a good docking example the 95 solutions with RMSD values are less than 0.5 Å and the scoring values are similar. (b) For a poor docking example the RMSD values are often more than 3.0 Å and the scoring values are diverse.

prove docking accuracy, such as the complexes 1xid and 1xie. In general GEMDOCK is able to improve the docking accuracy when both structural water and metal atoms are considered.

3.5 Comparison with Other Approaches

According to our best survey, most of previous docking works studied on a small dataset (< 20 complexes) except the GOLD [8] and FlexX [12] used. Here we compared GEMDOCK with GOLD and FlexX on the dataset of 100 complexes. Table 4 shows the summary of these three docking tools. The rates of FlexX based on a test set of 200 complexes enlarged the GOLD test set. GOLD was a steady-state genetic algorithm and FlexX was an incremental approach. GEMDOCK obtained a 76% success rate based on the condition of an RMSD value less than 2 Å. In contrast GOLD [8] achieved a 71% success rate in identifying the experimental binding model based on their assessment categories, and the rate was 66% if based on the RMSD condition. FlexX [12] achieved a 70% success rate based on solutions with any rank and the RMSD condition. The rate was 46.5% if the solutions at the fist rank were considered. The results of FlexX were often sensitive to the choice of the base fragment and its placement and the number of the fragments.

4 Conclusions

In this work, we have developed a robust evolutionary approach with a simple fitness function for docking flexible molecules. Experiments on 100 test systems verify that the proposed approach achieved a 76% success rate in recognizing the binding models.

Table 4. Comparsion GEMDOCK with GOLD and FlexX on the dataset of 100 complexes

RMSD(Å)	GEMDOCK	GOLD[a]	FlexX[b]
≤ 0.5	24%	8%	12.5%
$> 0.5, \leq 1.0$	39%	27%	38.5%
$> 1.0, \leq 1.5$	10%	20%	12.5%
$> 1.5, \leq 2.0$	3%	11%	5.5 %
$> 2.0, \leq 2.5$	1%	2%	7.5 %
$> 2.5, \leq 3.0$	3%	4%	2 %
> 3.0	20%	28%	21.5%

[a]: GOLD [8] is a steady-state genetic algorithm.
[b]: The rate of FlexX [12], a fragment-based approach, is based on any rank with a dataset of 200 complexes extended from the GOLD data set.

GEMDOCK seamlessly blends local search and global search to work cooperatively by the integration of a number of genetic operators, each having unique search mechanism. In summary, we have demonstrated the robustness and adaptability of GEMDOCK for exploring the conformational space of a molecular docking problem and efficiently finding the solution under the constraint of the fitness function used. Our scoring function seems to be simple and fast to discriminate native binding states and non-native docked conformations. We believe that GEMDOCK is an effective tool for docking flexible molecules.

Acknowledgements. This work was supported in part by grant 91-2320-B-009-001 from the National Science Council of Taiwan and in part by grand DOH92-TD-1132 from Department of Health, Taiwan.

References

1. I. D. Kuntz. Structure-based strategies for drug design and discovery. *Science*, 257:1078–1082, 1992.
2. H. Gohlke, M. Hendlich, and G. Klebe. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, 295:337–356, 2000.
3. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106:765–784, 1984.
4. B. K. Shoichet, A. R. Leach, and I. D. Kuntz. Ligand solvation in molecular docking. *Proteins: Structure, Function, and Genetics*, 34:4–16, 1999.
5. D. W. Miller and K. A. Dill. Ligand binding to proteins: the binding landscape model. *Protein Science*, 6:2166¡V2179, 1997.
6. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. A geometric approach to macromolecular-ligand interactions. *Journal of Molecular Biology*, 161:269–288, 1982.
7. D. K. Gehlhaar, G. M. Verkhivker, P. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel, and S. T. Freer. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chemistry and Biology*, 2(5):317–324, 1995.

8. G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267:727–748, 1997.
9. J. S. Taylor and R. M. Burnett. Darwin: A program for docking flexible molecules. *Proteins: Structure, Function, and Genetics*, 41:173V191, 2000.
10. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a lamarckian genetic algorithm and empirical binding free energy function. *Journal of Computational Chemistry*, 19:1639–1662, 1998.
11. C. J. Sherman, R. C. Ogden, and S. T. Freer. *De Novo* design of enzyme inhibitors by monte carlo ligand generation. *Journal of Medicinal Chemistry*, 38(3):466–472, 1995.
12. B. Kramer, M. Rarey, and T. Lengauer. Evaluation of the flexX incremental construction algorithm for protein-ligand docking. *Proteins: Structure, Function, and Genetics*, 37:228–241, 1999.
13. F. Österberg, G. M. Morris, M. F. Sanner, A. J. Olson, and D. S. Goodsell. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock. *Proteins: Structure, Function, and Genetics*, 46:34–40, 2002.
14. D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Inc., Reading, MA, USA, 1989.
15. T. Bäck. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, USA, 1996.
16. D. B. Fogel. *Evolutionary Computation: Toward a New Philosophy of Machine Intelligent*. IEEE Press, New York, 1995.
17. J.-M. Yang and C.-Y. Kao. Flexible ligand docking using a robust evolutionary algorithm. *Journal of Computational Chemistry*, 21(11):988–998, 2000.
18. J.-M. Yang, C.-H. Tsai, M.-J. Hwang, H.-K. Tsai, J.-K. Hwang, and C.-Y. Kao. GEM: A gaussian evolutionary method for predicting protein side-chain conformations. *Protein Science*, 11:1897–1907, 2002.
19. J.-M. Yang and C.-Y. Kao. A robust evolutionary algorithm for training neural networks. *Neural Computing and Application*, 10(3):214–230, 2001.
20. J.-M. Yang, J.-T. Horng, C.-J. Lin, and C.-Y. Kao. Optical coating designs using an evolutionary algorithm. *Evolutionary Computation*, 9(4):421–443, 2001.
21. R. M. A. Knegtel, J. Antoon, C. Rullmann, R. Boelens, and R. Kaptein. Monty: a monte carlo approach to protein-dna recongnition. *Journal of Molecular Biology*, 235:318–324, 1994.