

# Molecular Binding in Structure-based Drug Design: a Case Study of the Population-based Annealing Genetic Algorithms

Chien-cheng Chen    Leuo-hong Wang    Cheng-yan Kao    Ming Ouhyoung  
Wen-chin Chen

Department of Computer Science and Information Engineering  
National Taiwan University, Taipei, Taiwan, R.O.C.

## Abstract

*The molecular binding problem, one of the most important problems in structure-based drug design, can be formulated as a global energy optimization problem by using molecular mechanics. In this paper, a novel computational algorithm is proposed to address the molecular binding problem. The algorithm is derived from genetic algorithms(GA) plus simulated annealing(SA) hybrid techniques, namely population-based annealing genetic algorithms(PAG). We have applied the algorithm to find binding structures for three drug-protein molecular pairs. One of the three drugs is an anti-cancer drug methotrexate(MTX) and the other two are analogue of antibacterial drug trimethoprim. Moreover, we have also studied two other well-resolved ligand-receptor molecular complex which are obtained from the Protein Data Bank(PDB): Thermolysin-HONH-benzylmalonyl-L-Ala-Gly-p-nitroanilide complex(5tln) and HIV-1 protease-Hydroxyethylene isostere inhibitor complex. Hydroxyethylene isostere inhibitor is one of new potential HIV-1 protease inhibitors synthesized. Through our experiments, all of the binding results not only keep the energy at low levels, but also have a promising binding geometrical structure in terms of number of hydrogen bonds formed.*

## 1 Introduction

Many clinical drugs used today are generated in a lengthy drug discovery cycle. Generally speaking, such a lengthy cycle includes several important steps, including finding good starting molecular structures which are called lead compounds for optimization, refining the starting molecular structures to generate potential drugs, biologically testing potential drugs generated from previous steps and testing new drugs clinically [3]. Each step requires 1-3 years to complete. It requires 6-12 years to bring a new drug from discovery

to market in 1990[10]. Therefore, many researchers desire to shorten the cycle.

One of the efforts in speeding up the time required to find lead compounds is a strategy extended from Cohen's paradigm[9]. In 1977, Cohen proposed a general paradigm to design drugs for infectious diseases. His paper revealed that infectious agents, such as viruses, bind themselves with crucial proteins or nucleic acids. These crucial proteins serve as targets for intervention and can be utilized to design anti-cancer or anti-virus drugs. In other words, if certain drugs can bind appropriately with targets of infectious agents, the diseases will be inhibited. Extending Cohen's paradigm, alternative methods were applied to drug discovery [2][19]. Those methods are categorized in *structure-based drug design*[18]. The key issue of structure-based drug design is to find a small molecule which binds well with the receptor geometrically and biologically. Finding a good lead compound with a desirable mode of action and acceptable biological properties corresponding to a specific receptor has been termed the *molecular binding problem*.

In the past, many researchers have proposed strategies, including simulated annealing-based [1][12] and genetic algorithm-based strategies [5][7][23] to conquer the molecular binding problem. In this paper, we have proposed a novel strategy to address the molecular binding problem. Our strategy is based on genetic algorithms plus simulated annealing hybrids, namely population-based annealing genetic algorithms(PAG).

## 2 Population-based annealing genetic algorithm

Genetic algorithms(GA) [14][15] and simulated annealing(SA) [17][20] are two kinds of useful stochastic techniques which can be used to solve optimization problems approximately. The research incorporating GA with SA can be roughly considered to be two com-

plementary categories, one using genetic approach to design parallel SA [13][22] and the other considering SA algorithm as a neighborhood operator of the genetic algorithms [4][21]. The categories are not apparent for some cases, since these cases can be explained by different ways from different perspectives.

Observing the GA/SA hybrids listed, there is a common characteristic. No matter what the combination between GA and SA is, a partial sequence of SA algorithm was performed on each individual of the population. From the view point of GA, this partial sequence of SA algorithm takes care of improving individuals before or after genetic operators are applied. We can treat this sequence of SA algorithm as a neighborhood operator which searches the neighbors of individuals under the control of Metropolis condition.

Motivated by the observation mentioned above, SA algorithm being a neighborhood operator seems to be good for GA/SA hybrids. So we fix the number of steps of SA algorithm performed on each individual and define a set of population-based annealing(PA) operators. That is, a PA operator performs K times of *one-step SA* on the whole population. Following the concept, we can define a set of population-based annealing operator:

**Definition 1** A *population-based annealing(PA) operator* is a neighborhood operator of GA. PA can be described as a three tuples operator:  $PA = \{K, \mathcal{N}_g, \mathcal{R}_A\}$ .  $K$  is the number of steps of one-step SA algorithm applied to each individual of current population where  $K \geq 1$ .  $\mathcal{N}_g$  is the neighbor generation method that is used by each individual of current population when performing one-step SA algorithm.  $\mathcal{R}_A$  is the acceptance criterion used by one-step SA algorithm for competition of the current individual and its neighbors.

According to Definition 1, when we incorporate PA with GA, we can derive a new class of GA/SA hybrids, namely population-based annealing genetic algorithms(PAG) [27][28][30]. The pseudo code of the algorithm is given below:

- A. Initialize the parameters, i.e., population\_size,  $T_0$ , and decreasing factor  $\alpha(0 < \alpha < 1)$
- B. Randomly generate initial population.
- C. Repeatedly generate new populations as follows until system is frozen :
  1. For each individual do
    - Best\_point := Current\_point := Current\_individual ;
    - Do K times :
      - a. Generate next\_point from current\_point using the normal probability density function(normal P.D.F.) applied on the whole point space ;
      - b. Accept next\_point as current\_point by probability:
 
$$Prob = \frac{\exp(-C_{next}/T_k)}{\exp(-C_{current}/T_k) + \exp(-C_{next}/T_k)} ;$$

- c. if ( $C_{current} < C_{best}$ ) then
  - Best\_point := Current\_point ;
  - Pick Best\_point into the transient population.
2. Genetic stage.
  - Apply the genetic operators to transient population;
3. if (it is the first stage) then
  - determine the initial temperature  $T_1$ 

$$T_1 = \frac{Max_i(C_{max}^i - C_{min}^i)}{population\_size/2}$$
- else
  - $T_{k+1} = \alpha \cdot T_k ;$

There is something interesting in the initialization of the system temperature. It is recognized that the execution time of SA depends on the initial temperature and the decreasing factor of the temperature. PAG includes SA as an operator, so the efficiency of PAG also depends on the initial temperature. Actually, adjusting the initial temperature is a tradeoff between efficiency and the capability of diversity maintenance of the PA operator. When the initial temperature is high, PA can accept worse solutions with a larger probabilities. That means, much more solution space can be explored. But on the other side, the higher initial temperature is, the more time a system takes to reach to the frozen condition. In contrast, when the initial temperature is low, the exploration capability of PA is inhibited.

In [21], the authors provided a method deciding a reasonable range of initial temperature to prevent the deficiency. Here we provide another strategy. The strategy is defined as follows. For balancing the efficiency and the capability of diversity maintenance of the PA operator, we define the acceptance probability of detrimental move to be 0.6 when the algorithm starts. From Metropolis criterion  $Prob(\Delta C) = \exp(-\Delta C/T)$ , we obtain  $T = \frac{-\Delta C}{\ln 0.6} \cong 2 \cdot \Delta C$  where  $\Delta C$  is determined by the largest possible detrimental move of current generation. Since we are not sure how large a possible detrimental move can be, we try to estimate the possible detrimental move by the values generated from the first cycle of PAG algorithm. In the first cycle of PAG, a piecewise Markov chains are generated from each individual of initial population, these individuals may be located at very different hills. We have to consider all chains to estimate the largest possible detrimental move. Therefore, we calculate the initial temperature as :

$$T_{init} = \frac{Max_i(C_{max}^i - C_{min}^i)}{population\_size/2} \quad (1)$$

where  $C_{max}^i, C_{min}^i$  are the largest and lowest cost of the  $i$ th sequence of Markov chain generated by  $i$ th individual of first generation. We take the maximal

difference of all sequences of Markov chain as  $\Delta C$  to determine the initial temperature.

As for the genetic phase, there are two important issues. One is the selection mechanism and the other is the genetic operators.

At the genetic phase, we use the ranking algorithm as a selection mechanism. As presented by Whitley [29], the ranking algorithm not only prevented GA from premature convergence but also provided a direct control on selective pressure which affects the search speed. We propose a dynamic ranking procedure which adaptively changes the selective bias from generation to generation. The dynamic ranking procedure is given as :

1. Assign initial value of bias. The value biases the ratio of the offspring with expected value between the best and the worst individual in current population.
2. For each individual, evaluate the number of the expected offspring,  $E_i$ , based on the bias and the following linear function from the best one to the worst one.

$$E_i = \frac{(\text{population\_size} \cdot \text{bias}) - (2 \cdot (\text{bias} - 1) \cdot \text{rank}_i)}{(\text{population\_size} + \text{bias} - 1)} \quad (2)$$

where  $\text{rank}_i$  is the rank of each individual. The best one is zero, and the worst one is ( $\text{population\_size}-1$ ). When bias equals 1.2 , the ratio between the best and the worst is about 1.5. When the bias reaches 1.5 , the ratio becomes 3.

3. Raise the bias by a factor  $\alpha$  from generation to generation.

Since we want to keep the diversity at the first few generations of PAG and after that we need to gradually inherit good materials from better individuals, therefore dynamically changing the bias of ranking is appropriate. After a few tests, we use 1.2 as the initial bias and multiply the bias by 1.005 after each generation.

Moreover, the crossover and mutation operators are performed according to the following steps.

1. At first, the parents are selected from the population randomly. The crossover operator is applied with a predefined crossover rate. After that, two offspring are produced.
2. The offspring survive only when the costs of these two offspring are both less than average cost of the previous generation. Otherwise, give up the offspring and continue to apply the mutation

operator to the parents. The mutation operator is an annealing-like mutation operator which mimics the exploration capability of SA. It explores the neighborhood of the parents. Mutation occurs with a logistic probability given as  $\exp(\frac{-E_{new}}{T}) / (\exp(\frac{-E_{old}}{T}) + \exp(\frac{-E_{new}}{T}))$ , where  $E_{new}$  and  $E_{old}$  are the evaluation functions of a problem, and are binding energies in our problem domain.

3. Finally, the offspring or the mutated parents are copied into the new population.

Finally, for the frozen condition, PAG will stop when 80% chromosomes' genes are all the same. Choosing 80% as the stopping criterion is only for the sake of efficiency for our problem domain.

### 3 The Problem

The key issue of the molecular binding problem is to find the binding structure of the drug-receptor complex. To judge whether a binding structure of molecules is stable or not, scientists use the potential energy between binding pairs as a reference.

Generally speaking, the interaction potential energy of a specific binding structure can be approximately evaluated by molecular mechanics[6]. In general, the potential energy of a given system is determined by the summation of all kinds of force fields. The following equation is a canonical form of the total potential energy function:

$$\mathcal{V} = V_b + V_\theta + V_\phi + V_\chi + V_{nb} + V_e + V_{hb} \quad (3)$$

where  $V_b$  is the bond stretching force field,  $V_\theta$  the bond angle bending force field,  $V_\phi$  the bond torsion force field,  $V_\chi$  the out-of-plane bending force field,  $V_{nb}$  the non-bonded interaction force field,  $V_e$  the electrostatic interaction energy and  $V_{hb}$  is the hydrogen bonding force field.

Fortunately, not all of these terms are crucial to solve the molecular binding problem. Actually, only three terms of them are dominant when calculating the intermolecular potential energy. These three terms are non-bonded interaction, bond torsion(steric) and electrostatic force fields. Therefore, without loss generality, we can simplify Equation 3 as follows:

$$\mathcal{V} = V_\phi + V_{nb} + V_e \quad (4)$$

Since we start with a simplified case based on previous work of [24]. Therefore, about the scoring function, we follow authors in [24] to use the Lennard-Jones 6-12 potential function to represent non-bonded

interaction  $V_{nb}$  and electrostatic interaction  $V_e$ . The Lennard-Jones equation is defined as follows[26]:

$$V_{tot}(r, d) = \sum_{r,d} \frac{332q_dq_r}{\varepsilon|\vec{R}_r - \vec{R}_d|} + \sum_{r,d} \frac{A_{rd}}{|\vec{R}_r - \vec{R}_d|^{12}} - \sum_{r,d} \frac{B_{rd}}{|\vec{R}_r - \vec{R}_d|^6} \quad (5)$$

where  $V_{tot}$  is the total binding energy.  $q_r$  and  $q_d$  are the partial charges of atoms belonging to the receptor and the drug respectively.  $\vec{R}_r$  and  $\vec{R}_d$  represent the three dimensional coordinates of atom in the receptor and the drug molecule. Therefore,  $|\vec{R}_r - \vec{R}_d|$  is the distance between atoms of the receptor and drug differently.  $\varepsilon, A_{rd}, B_{rd}$  are the dielectric and non-bond constants. In this function, the first summation simulates the electrostatic interaction between each pair of atoms, and the second and third summation simulate the repulsive and attractive terms in Van der Waals interaction energy respectively.

Following the work in [24], we will first bind three ligands to dihydrofolate reductase(DHFR) to demonstrate our algorithms. Three kinds of ligands are used to be ‘‘docked’’ to DHFR: methotrexate, inhibitor 91 and inhibitor 309. Methotrexate is an anti-cancer drug which is used clinically to cure patients. Inhibitor 91 and 309 are analogues of trimethoprim which is an anti-bacterial drug. To reduce computational complexity in this pilot research, we only fetch the active site of DHFR as our target molecule. The constitution of all molecules including the receptor and the number of degrees of freedom are listed on Table 1.

Molecule	Atoms	Single bonds	Degrees of freedom
MTX	54	12	18
91	54	13	19
309	45	12	18
DHFR	557	-	-

Table 1: *The data set of our simulation model.*

In addition, we also study two other well-resolved ligand-receptor molecular complex which are obtained from the Protein Data Bank(PDB): Thermolysin-HONH-benzylmalonyl-L-Ala-Gly-p-nitroanilide complex(5tln) and HIV-1 protease-Hydroxyethylene isostere inhibitor complex. Thermolysin is one of the important enzymes that exist widely and participate in a number of important biological and physiological processes. HONH-benzylmalonyl-L-Ala-Gly-p-nitroanilide is one of the well-known inhibitors which bind well to thermolysin

[16]. Hydroxyethylene isostere inhibitor is one of new potential HIV-1 protease inhibitors synthesized[11]. The DOF and numbers of atom of receptors and ligands are listed on Table 2.

Molecule	Atoms	Single bonds	DOF
Inhibitor for HIV-1	88	26	32
Inhibitor for TLN	42	10	16
HIV-1	3122	-	-
TLN	4711	-	-

Table 2: *The data set of our simulation model, where DOF is the number of single bonds plus six degrees of freedom in translation and rotation. TLN is thermolysin, HIV-1 is Human Immunodeficiency Virus-1 Protease.*

All test molecules listed above can be viewed as a number of atoms defined by their three-dimensional coordinates. Using these data sets as well as scoring functions mentioned previously, we will use PAG to regenerate binding structures of these molecular complex.

Intuitively, the three-dimensional location of the drug and its three rotational angles relative to 3 axes are all adjustable. Moreover, the molecules have a number of deformable single bonds. Each single bond is a degree of freedom. Based on the description, the energy minimization between a molecular system becomes:

1. Fix the location of the receptor. Initialize the structure of the drug molecule. Evaluate the interaction energy based on the scoring function.
2. Repeatedly adjust different degrees of freedom, including translating and rotating the drug molecule and twisting single bonds inside the drug, to fit the receptor. Evaluate the interaction energy of each new configuration.
3. Find the best configuration with the lowest interaction energy from these configurations.

By adjusting the value of each degree of freedom, a new configuration will be generated. The whole search space is the combination of possible values of all degrees of freedom. Therefore we encode all degrees of freedom as a chromosome:

$$(t_x, t_y, t_z, r_x, r_y, r_z, s_1, \dots, s_m)$$

where  $t_x, t_y$  and  $t_z$  represent the position of the drug molecule relative to the centroid of the receptor,  $r_x,$

Drug	Population size	Times in sec.	Generations	Evaluation(times)	Energy(kcal/mol)
MTX	50	52042	165	261818	-102.60
	100	137149	241	757069	-110.49
91	50	61406	173	309147	-71.83
	100	138991	269	946534	-73.13
309	50	41272	215	363772	-61.32
	100	106849	284	958634	-66.58

Table 3: The results evaluated by PAG, where the binding energy need to be globally minimized. Evaluation means the count of evaluation about scoring function.

$r_y$ ,  $r_z$  are the rotational angles of the drug, and  $s_i$  are the twisting angles about the single bonds inside the drug molecule.

For convenience, we use real-coded scheme to encode each parameter. That is, all parameters are real numbers. In addition, we use one-point crossover and blend crossover as genetic operators.

## 4 Experimental results and discussion

### 4.1 Molecular binding results

We have implemented PAG algorithms on a Sun UltraSparc workstation. (Moreover, we also have implemented another PC version of PAG algorithms.) Three pairs of receptor-ligand: DHFR-MTX, DHFR-Inhibitor 309 and DHFR-Inhibitor 91 are evaluated by using the algorithm listed in previous section. The preliminary results in this pilot research are given in Table 3.

In the table, PAGs are executed 10 times for each case. All of the parameter settings are the same. The decreasing factor of temperature is 0.9. K, the number of steps performing PA operator in each generation is equal to degrees of freedom in each problem. We use one-point crossover and blend crossover as the genetic operators. The probability of crossover and blend crossover are both 0.5. Based on the design of genetic stage explained in the previous section, the mutation rate is dependent on the success of the crossover operators. Moreover, the final results of PAG is improved by a steepest-descent local minimizer. Since we use the real number encoding, each DOF is a 64-bit real number, the total number of possible candidates for each DOF are  $2^{64}$ . Therefore, if a drug molecule possesses 12 single bonds, the total number of possible position and conformation will be  $2^{64 \times 12} = 6.07 \times 10^{346}$ . We also compare our results with similar results from Oshiro, Kuntz and Dixon [23]. The following is a table of comparison (Table 4). Please note that Kuntz's method is

a pure genetic algorithm, and uses binary coding for the solution space, while we use real number coding; furthermore, Kuntz's algorithm needs to know the active site first, but ours can use random position for the starting points. For initial population, we use a much smaller size (100) as compared to 12167 from Kuntz. The results show that our PAG is 24% better in terms of number of energy function evaluations, even if our solution space is much bigger.

Drug	Population	Evaluation	Rmsd	Solution space
Kuntz	12167	940332	0.8 Å	active site
PAG	100	757069	0.8 Å	whole space

Table 4: Table of comparison with the binding of an inhibitor MTX to a receptor DHFR, where Rmsd is root-mean-square deviation.

### 4.2 Comparison of experimental results

There are many criteria to judge whether the binding structures generated by the program are reasonable. The first criterion is the binding energy. Since we use the same data set and energy functions as the experiment done in [24][25], the results can be compared with the data listed in [24]. In [24], the best binding energy of these three drugs are from -40Kcal/mol to -120Kcal/mol. PAGs have obtained the minimal energy at the range from -40Kcal/mol to -110Kcal/mol. The binding energy of the pairs MTX-DHFR, 91-DHFR and 309-DHFR are in the range of -70Kcal/mol to -110Kcal/mol, -40Kcal/mol to -80Kcal/mol and -50Kcal/mol to -70Kcal/mol respectively.

The results listed above roughly verify that PAGs can find out the binding structures with reasonable low binding energy. However, the corresponding drug structures of approximative binding energy may be different in orientation and conformation.

In addition to the binding energy, the existence of hydrogen-bonds is the other stronger criterion that determines the goodness of fitting between ligands and proteins. The hydrogen bond is formed when the hydrogen atom is close enough ( $2.4\text{\AA} \sim 2.8\text{\AA}$  in distance) to another atom such as oxygen. Since the hydrogen bond involves a strong “binding energy”, two to five pairs of such hydrogen bonds becomes dominant in the total binding energy. There are 4 primary hydrogen bonds formed between MTX and DHFR. These 4 primary hydrogen bonds construct the elementary conformational structure of MTX-DHFR complex. The same situation happens to the 309-DHFR and 91-DHFR complex.

Table 5 shows our experimental results obtained from the PAG. From the analysis of hydrogen bond-

Inhibitor	MTX		91		309	
	50	100	50	100	50	100
Total runs	10		10		10	
hb threshold <sup>a</sup>	2.5 Å		2.4 Å		2.8 Å	
4 hb <sup>b</sup>	3	7	7	7	2	8
3 hb <sup>c</sup>	7	8	8	10	9	10

Table 5: *The hydrogen bonds found by PAG.*

<sup>a</sup>The largest distance of forming a hydrogen bond

<sup>b</sup>Forms 4 hydrogen bonds

<sup>c</sup>Forms at least 3 hydrogen bonds

ing between drugs and DHFR, we can claim the presented PAG have found some good fitting molecular structures between drugs-DHFR complex, since most of the resulting drug structures getting from PAG are buried deeply into the correct pocket of DHFR.

Figure 1 illustrates the superimposing of MTX crystal structure with the best of our results. The rmsd(root-mean-square deviation) value of this structure is  $0.8\text{\AA}$ .

### 4.3 Experimental results of more general cases

We have applied PAG to solve more complicated cases. These cases are molecular complex obtained from PDB, including Thermolysin-HONH-benzylmalonyl-L-Ala-Gly-p-nitroanilide complex(5tln) and HIV-1 protease-Hydroxyethylene isostere inhibitor complex. Since the problem size is bigger than the cases of DHFR, we use larger population size for PAG. The experimental results of more general cases are listed in Table 6.

All of the cases listed above are the best cases. The parameter settings and genetic operators are all the

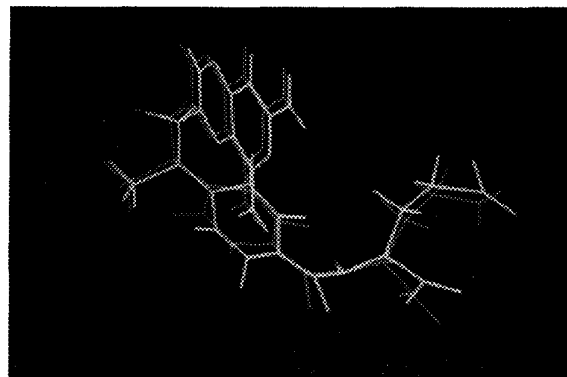


Figure 1: *The superimposing of our resulting structure with the crystal structure of MTX with rmsd value of  $0.8\text{\AA}$ . The brighter line segments are the crystal structures of MTX, and the darker line segments are our resulting structure.*

same with the cases of DHFR. We will discuss the results with the following issues:

**Binding energy** When the population size is 200, the binding energy of HIV-1 protease with its inhibitor obtained from PAG are in the range of  $-110\text{ Kcal/mol}$  to  $-130\text{ Kcal/mol}$ . As to the cases of Thermolysin, the binding energy are in the range of  $-40\text{ Kcal/mol}$  to  $-80\text{ Kcal/mol}$ . Comparison with the binding energy of crystal structures which are  $-145\text{ Kcal/mol}$ (HIV-1) and  $-88\text{ Kcal/mol}$ (TLN) respectively, our results are near enough. Moreover, we also try another experiments of parameter settings of PAG. We have focused on the initial temperature  $T_0$  and temperature decreasing factor  $\alpha$  respectively. Since raising the initial temperature and the temperature decreasing factor could strengthen the effects of simulated annealing to prevent from trapping into local optimal. Due to the stochastic property of PAG, the results aren't always much better. We will try more experiments about parameter settings of PAG in the future.

**Hydrogen bonding** According to the results presented in [11] and [16], there are a few primary hydrogen bonds formed between HIV-1 protease and Thermolysin with their inhibitors. There are 5 primary hydrogen bonds between HIV-1 protease with its inhibitor. In one experiment done by Wang [30] has showed 5 out of 10 runs could locate these 5 hydrogen bonds. In the case of Thermolysin, there are 4 primary hydrogen bonds be-

Drug	Population size	Generations	Evaluation(times)	Energy(kcal/mol)
Inhibitor for HIV-1	100	473	2608561	-124.11
	200	538	5384201	-126.19
Inhibitor for TLN	100	171	451049	-77.87
	200	436	2204649	-80.20

Table 6: The results evaluated by PAG. Evaluation means the count of evaluation about scoring function.

tween Thermolysin with its inhibitor. 6 out of 10 runs in his results locate these 4 hydrogen bonds.

From the analysis of hydrogen bonds, PAG still have found good fitting molecular structures over 50% even when the problem size is much larger than the cases of DHFR.

From the discussion listed above, we can claim PAG still have found the reasonable fitting structures over 50% even when the problem size is getting much larger. The following figure(Figure 2) shows the illustration of the binding structures.

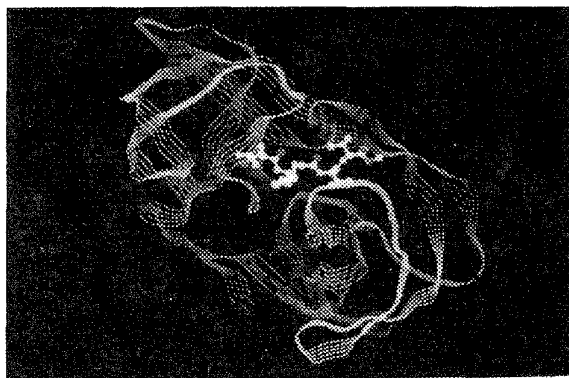


Figure 2: The binding structure of Inhibitor with HIV-1, where the central small molecule is the inhibitor, and the outside ribbon model is the HIV-1.

## 5 Conclusion

The importance of the molecular binding problem is without questioning. In this paper, we have developed a class of neighborhood operators called population-based annealing operators for genetic algorithms. Incorporating a different PA operator with genetic algorithms will form a new GA/SA hybrid. We design the PAGs to solve the level-2 [8] molecular binding problem. Experimental studies not only indicate the usefulness of PAGs but also the properties of different PA operators.

The near-optimal solutions obtained by PAG can be used for the work of drug design. We have verified that the genetic algorithms incorporating with SA are suitable to solve the molecular binding problem and believe these results are very useful for the biotechnology community. In summary, the contribution of this paper are:

1. We have discussed the GA/SA hybrids in details. After some elementary analysis, we design the GA/SA hybrids. The algorithm can be used to solve the level-2 molecular binding problem. Furthermore, in the binding problem, we don't need to know the binding pocket site at first, where most researchers have to know for their algorithms to reduce the problem size. We also tried the case of HIV-1 protease with its inhibitor, where the size of the problem have kept others from automatically finding its binding position.
2. Our PAG algorithms are extensible since we have focused on balancing the efficiency and enhancing the capability of diversity maintenance of the algorithms. Based on the extensibility, we have a great chance to examine the level 3 binding problem.

## References

- [1] R. Abagyan, M. Totrov, D. Kuznetsov, *ICM-A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation*, Journal of Computational Chemistry, Vol. 15, No. 5, pp. 488-506, 1994.
- [2] K. Appelt, et al., *Design of Enzyme Inhibitors Using Iterative Protein Crystallographic Analysis*, Journal of Medicinal Chemistry, Vol.34, No. 7, pp. 1925-1934, 1991.
- [3] L. M. Balbes, S. W. Mascarella, D. B. Boyd, *A Perspective of Modern Methods in Computer-Aided Drug Design*, Reviews in Computational Chemistry, Volume V, K. B. Lipkowitz and D. B. Boyd, Ed. VCH Publishers, Inc. New York, 1994.

- [4] D. E. Brown, C. L. Huntley, A. R. Spillane, *A parallel genetic heuristic for the quadratic assignment problem*, Proceeding of the Third International Conference in Genetic Algorithms, pp. 406-415, 1989.
- [5] T. Brodmeier, E. Pretsch, *Application of Genetic Algorithms in Molecular Modeling* Journal of Computational Chemistry, Vol. 15, No. 6, pp. 588-595, 1994.
- [6] U. Burkert, N. L. Allinger, *Molecular Mechanics*, American Chemical Society, Washington D.C., 1982.
- [7] K. P. Clark, *Ajay Flexible Ligand Docking without Parameter Adjustment across Four Ligand-Receptor Complexes*, Journal of Computational Chemistry, Vol. 16, No. 10, pp. 1210-1226, 1995.
- [8] D. E. Clark, D. R. Westhead, *Evolutionary algorithms in computer-aided molecular design*, Journal of Computer-Aided Molecular Design, Vol. 10, pp. 337-358, 1996.
- [9] S. S. Cohen, *A Strategy for the Chemotherapy of Infectious Disease*, Science, Vol. 197, No. 4302, pp. 431-432, 1977.
- [10] J. A. DiMasi, R. W. Hansen, H. G. Grabowski, L. Lasagna *Cost of Innovation in the Pharmaceutical Industry*, Journal of Health Economics.
- [11] G. B. Dreyer, et al. *Hydroxyethylene Isostere Inhibitors of Human Immunodeficiency Virus-1 Protease: Structure-Activity Analysis Using Enzyme Kinetics, X-ray Crystallography, and Infected T-Cell Assays*, Biochemistry, vol. 31, pp. 6646-6659, 1992.
- [12] D. K. Gehlhaar, K. E. Moerder, D. Zichi, C. J. Sherman, R. C. Ogden, S. T. Freer, *De Novo Design of Enzyme Inhibitors by Monte Carlo Ligand Generation*, Journal of Medicinal Chemistry, Vol. 38, No.3, pp. 466-472, 1995.
- [13] D. E. Goldberg, *A Note on Boltzmann Tournament Selection for Genetic Algorithms and Population-Oriented Simulated Annealing*, Complex Systems, 4, pp. 445-460, 1990.
- [14] D. E. Goldberg, *Genetic algorithms : In Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, 1989.
- [15] J. H. Holland, *Adaptation in natural and artificial systems*, The university of Michigan Press, 1975.
- [16] M. A. Holmes, B. W. Matthews, *Binding of Hydroxamic Acid Inhibitors to Crystalline Thermolysin Suggests a Pentacoordinate Zinc Intermediate in Catalysis*, Biochemistry, vol. 20, pp. 6912-6920, 1981.
- [17] S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, *Optimization by simulated annealing*, Science, Vol. 220, No. 4598, pp. 671-680, 1983.
- [18] I. D. Kuntz, *Structure-Based Strategies for Drug Design and Discovery*, Science, Vol. 257, pp. 1078-1082, 1992.
- [19] L. F. Kuyper, et al., *Receptor-based Design of Dihydrofolate Reductase Inhibitors: Comparison of Crystallographically Determined Enzyme Binding with Enzyme Affinity in a Series of Carboxy-Substituted Trimethoprim Analogues*, Journal of Medicinal Chemistry., Vol 28, pp.303-311, 1985.
- [20] P. J. M. van Laarhoven, E. H. L. Aarts, *Simulated Annealing : theory and applications*, D. Reidel Publishing Company, Dordrecht, Holland, 1987.
- [21] F. T. Lin, C. Y. Kao, C. C. Hsu, *Applying the Genetic Approach to simulated Annealing in Solving Some NP-Hard Problems*, IEEE Transaction on System, Man, Cybernetics, Vol. 23 No. 6, pp. 1752-1767, 1993.
- [22] S. W. Mahfoud, D. E. Goldberg, *Parallel Recombinative Simulated Annealing: A Genetic Algorithm*, Parallel Computing, Vol. 21, No. 1, pp. 1-28, 1995.
- [23] C. M. Oshiro, I. D. Kuntz, J. S. Dixon, *Flexible Ligand Docking Using a Genetic Algorithm* Journal of Computer-Aided Molecular Design, Vol 9, pp. 113-130, 1995.
- [24] M. Ouhyoung, *Force Display in Molecular Binding*, Ph. D dissertation, CS dept., University of North Carolina at Chapel Hill, 1990.
- [25] F. P. Brooks, Jr., M. Ouhyoung, J. J. Batter, P. J. Kilpatrick, *Project Grope: Haptic Display for Scientific Visualization*, ACM SIGGRAPH'90, pp. 177-186, (also Computer Graphics Vol. 24, No.4, 1990).
- [26] N. Pattabiraman et al., *Computer Graphics and Drug Design: Real Time Docking, Energy Calculation and Minimization*, Journal of Computational Chemistry, Vol. 6, pp. 432-436, 1985.
- [27] L. H. Wang, C. Y. Kao, M. Ouhyoung, W. C. Chen, *Using an Annealing Genetic Algorithm to Solve Global Energy Minimization Problem in Molecular Binding*, IEEE International Conference on Tools with Artificial Intelligence, pp. 404-410, New Orleans, USA, 1994.
- [28] L. H. Wang, C. Y. Kao, M. Ouhyoung, W. C. Chen, *Molecular Binding: a Case Study of the Population-based Annealing Genetic Algorithms*, IEEE International Conference on Evolutionary Computing, Perth, Australia, 1995.
- [29] D. Whitley, *The Gentitor Algorithm and Selection Pressure : Why Rank-Based allocation of Reproductive Trials is Best*, Proceedings of the Third International Conference on Genetic Algorithms, pp. 116-121, 1989.
- [30] L. H. Wang, *Molecular Binding in Structure-based Drug Design: a Case Study of the Population-based Annealing Genetic Algorithms*, Ph.D. Thesis, CSIE dept., National Taiwan University, 1997.