# RASSE: A New Method for Structure-Based Drug Design†

Zhaowen Luo, Renxiao Wang, and Luhua Lai*

Institute of Physical Chemistry, Peking University, Beijing 100871, P.R. China

Received February 12, 1996⊗

A novel method, RASSE, has been developed to suggest reasonable structures which can fit well to the binding sites of receptors. Molecules are generated by an iterative growing procedure in which atoms are added to existing fragments. Potential ligands are then picked out by special scoring rules. This atom-growing based method is characterized by combinatorial searching of atom types and conformations. To some extent, it is the computer simulation of combinatorial chemistry. This method has been applied to the design of inhibitors for *E. coli* dihydrofolate reductase and human phospholipase $A_2$. The results demonstrate that this program is capable of generating reasonable structures, thus proving its power in drug design.

## 1. INTRODUCTION

Creating molecules with specific properties has been a cherished goal of chemists for generations. Finding new drugs, in particular, is an important part of the new initiatives in health care. The past few years have witnessed exciting progress in the determination of macromolecule structures, while homology based methods for the prediction of protein structure were also improving. With these rapid advances, structure-based drug design strategy has drawn much attention.[1−6] Such method analyzes the active sites of proteins and suggests compounds being able to bind to these sites. This field is still under development, but the underpinnings of the method can be assessed and initial results are being tested.

At the very beginning of structure-based drug design, researchers only employed graphics to analyze the structural features of the target protein. Soon after that, computer-aided drug design proved to be a powerful additional tool. Goodford et al. have developed GRID,[7,8] which places small fragment probes at grid points within the active site, determining the most favorable scores. Since their pioneering work, many methods for structure-based drug design have appeared.[9−44] These methods are, in general, complementary to one another, and many of them can be approximately grouped into three catalogues: database searching, fragment combination, and atom-growing method.

Examples of database searching method include DOCK,[9−11] 3DSEARCH,[12] ALADDIN,[13] CAVEAT,[14] and FLOG,[15] which search through databases of known 3D structures and identify those entries which can fit into the active site. The Cambridge Structure Database (CSD) is one of the most widely used databases. Since several methods for rapid structure conversion from 2D to 3D have become available, e.g., CONCORD, database searching can be extended to screen molecules in chemical directories and even in the CA index. At present, database searching is the widely adopted method for practical uses. The advantages of such approaches are that the molecules retrieved from the databases do exist, and they are fitted into the active site in a low-energy conformation. But an obvious drawback of database searching is that it can only suggest drug candidates among known structures. Moreover, such a procedure of database searching will be rather demanding both in CPU time and in disk space.

A number of programs can be labeled as fragment combination methods, such as LUDI,[16−18] CLIX,[19] SPLICE,[20] GroupBuild,[21] SPROUT,[22−24] and the work by Leach and Kilvinton.[25] The idea is to position molecular fragments into the active site in such a way that favorable interactions can be formed with the protein. These fragments are then connected into a single, complete molecule. Such methods can generate reasonable structures and due to the large number of possible combinations of fragments, the variety of molecules that can be generated is enormous. But they stumble over these problems: First, it is difficult to reconnect functional groups to form complete molecules while maintaining the geometric positions of lowest energy. Secondly, since the molecular fragments are defined by the users, such methods probably cannot result in all feasible structures, especially in a limited space like the binding pocket.

Since molecules are formed by atoms, another way to build up a molecule is to combine atoms. Such approaches include LEGEND[26,27] and GENSTAR.[28] These programs build a structure sequentially from randomly selected atom types which are positioned with random torsion angles. A candidate atom is selected automatically if it is not bumping either the protein or any previous atoms in the growing molecule. After a number of structures have been generated, a post-processing program may be used to select the more interesting structures for graphical analysis. Atom generation is the most direct way to form a molecule. But unfortunately, combinatorial explosion is inherent in the attempt to combine atom types and to carry out conformational searching thoroughly. Because of this, although fairly good results can be obtained, the best ones may be missed.

There are still many other methods[29−44] which cannot be simply classified into these three categories. They either combine the techniques described above or adopt other interesting algorithms to suggest ligands for a specific receptor. These methods are also valuable and instructive in our way to rational drug design.

* Corresponding author address: Institute of Physical Chemistry, Peking University, Beijing 100871, P.R.China. Fax: 86-10-62751725. Tel: 86-10-62751490. E-mail: lai@ipc.pku.edu.cn.

† Key words: drug design, structure-based, atom-growing, RASSE, combinatorial chemistry.

⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0095-2338(95)00277-0 CCC: $12.00 

As far as we know, there have not been any programs yet which can generate chemical structures by atom-growing accompanied with conformational searching. We have developed a new method, Rational Space Searching (RASSE), based on atom-growing with combinatorial searching of atom types and conformations. For a specific enzyme, the number of possible inhibitors is limited. That means the space of inhibitor generation is not boundless. In RASSE, we eliminate the unreasonable cases while growing atoms. Therefore conformational searching can run in a limited, rational space, and combinatorial explosion can be controlled effectively.

To some extent, our method is the computer simulation of combinatorial chemistry. During the procedure of atom growing, there are so many choices among different atom combinations and different conformations. We consider all of the possible situations in our program. This is the reason why combinatorial explosion may happen. But in fact, combinatorial explosion is the spirit of combinatorial chemistry since it offers plenty of diversity. So what we have tried to do is let the explosion happen and then carefully control and utilize it.

The usefulness of RASSE was exemplified by applications to the inhibitor design for *E. coli* dihydrofolate reductase and human phospholipase $A_2$. As our test cases have shown, this program suggested novel, reasonable chemical structures which fit the active sites, thus proving its power in structure-based drug design. We also discuss the strengths and weaknesses of this method and our current efforts on developing more advanced methods.

## 2. METHODS

Our program consists of three main modules: SITE, GROW, and SCORE. SITE analyzes the active binding sites. GROW adds atoms to the existing molecular fragments to generate new molecules. SCORE ranks the molecular fragments generated by GROW and selects the best ones for the next atom growing cycle. The last two modules, namely GROW and SCORE, form an iterative procedure of atom-growing. The programs are written in C language. A schematic flow chart of the program is shown in Figure 1. The preparations before running RASSE are as follows:

(1) atomic coordinates of the target protein in PDB format;
(2) pocket residues which define the position of the binding pocket;
(3) an initial molecular fragment

**2.1. Analyzing the Binding Site (SITE).** The main function of SITE is to create a box which defines the binding pocket. The box includes the binding pocket and the pocket residues and is used as the boundary for atom growing. To do this, atomic coordinates of pocket residues are extracted from the PDB file and are analyzed to position the box. Protein structures without hydrogen atoms are used in our program. Conserved bound water molecules should be identified in advance, and they are treated as parts of the protein.

We use a force field to evaluate the quality of ligand binding, which will be described later in this paper. Our program uses nonbonded parameters from the AMBER force field.[45] Atomic charges on the protein are from AMBER too. Amino acid residues Asp, Glu, Lys, and Arg, are assumed to be charged. For efficient calculation of the intermolecular interactions between the ligand and the receptor, we divided the box into regularly spaced grids. In SITE, van der Waals interactions and electrostatic potentials at each grid point are computed and stored. Therefore if a probe atom locates at a certain grid, the steric and electrostatic interactions between the atom and the protein can be calculated conveniently. These tabulated data are used for energy estimation at each step of atom growing in module GROW and also for scoring the generated molecules in module SCORE. Such a grid method can dramatically save the time spent on energy calculation. The distance between two neighboring grid points is usually set to 0.3 Å. This is roughly equal to the accuracy of the atomic positions in a high-resolution protein structure and within the tolerance of most of the results of force field calculations. Therefore we conclude that the use of discrete points does not introduce significant errors into the calculation.

**2.2. Introducing an Initial Molecular Fragment.** Because of the massive conformational space possible for a ligand in the binding pocket of the receptor, it is impractical to carry out a thorough conformational searching. To reduce the conformational space, an initial molecular fragment is introduced into the binding pocket before atom growing. The initial molecular fragment occupies part of the pocket space and acts as a seed for atom growing. In practice, part of the known inhibitor can be used as the initial molecular fragment if the structure of complex is available. The other way to get an initial fragment is to dock a molecule into the binding pocket and then modify it. Keeping some key sites for binding in the initial fragment, e.g., hydrogen bonding sites, will certainly improve the quality of the final results. Using an initial molecular fragment brings RASSE another advantage: you do have choices. When using a small initial fragment which contains only a few atoms, it is really like *de novo* drug design. You can get novel structures beyond your wildest imagination. But when using a relatively large skeleton as the initial fragment, it is similar to the optimization of known structures and you can get various derivatives of it.

**2.3. Adding Atoms to Molecular Fragments (GROW).** The function of GROW, which is the main module of RASSE, is to generate new molecules by adding atoms to an existing molecular fragment. In our program, eight basic atom types have been used, that is, $Csp^3$, $Csp^2$, $Nsp^3$, $Nsp^2$, $Osp^3$, $Osp^2$, aromatic C, and H. Atom types $Csp^2$ and $Nsp^2$ are further divided into two types, respectively, according to their different directions of growing. For each atom type, a growing set is defined which contains the atom types that can be grown from that atom. The atom types and their growing sets are listed in Table 1.

In RASSE, new atoms grow from the atoms with vacant valences, which we refer to as seed atoms. The number of newly grown atoms is determined by the growing sites on the considered seed atom (see Table 1). Unlike the other atom-growing methods, we do not use a random number to determine the types of new atoms. On the contrary, we consider all the possible combinations of the growing sets. For instance, let us suppose there is a $Csp^3$ seed atom. To simplify the case, let us assume its growing set only contains two atom types, namely A and B. According to our growing algorithm, three new atoms should be grown from a $Csp^3$

seed atom at the same time. Thus all feasible growing combinations are AAA, BAA, ABA, AAB, ABB, BAB, BBA, and BBB. Chemically unreasonable combinations are identified and eliminated at this step. New atoms are positioned by standard bond lengths and bond angles which are taken from AMBER force field.[45] Dihedral angles are determined by conformational searching in an interval of 30 degrees. If the tested conformation has no serious van der Waals collisions with the protein, it is considered as reasonable. Therefore each growing combination will probably have several reasonable conformations. All of them form a combination set of the seed atom. The seed atoms will not be seeds in the next growing cycle because they do not have vacant growing sites any more. The newly grown atoms will become seeds.

In principle, aromatic rings can be generated atom by atom in a Kekulè mode. But in practice, it is extremely difficult to let an atom-growing program to do that. Since aromatic rings are most common in organic molecules, we have designed special algorithms to grow six- and five-membered aromatic rings in our program. That is, an aromatic ring is grown out as a whole fragment. We find this has greatly improved the rationality of the generated molecules. Aromatic heterorings can be obtained by substituting carbon for nitrogen at certain sites on the rings.

Another novel conception employed in RASSE is what we call probability-dependent growing. It can be seen in a rough browse of organic compounds that different atom types have different relative ratios of occurrence. So it is not reasonable to assign even probability of growing to all atom types. This is the request of good chemical structure. In practice, we have used a set of parameters to scale the growing probabilities of different atom types (see Table 1). These parameters can be defined by the users in the program. A recommended way to get such parameters is to perform statistical analyses of the components of common organic compounds and natural products.

There is usually more than one seed atom in one molecular fragment. For all seed atoms, their combination sets are recorded. After the current cycle of atom growing finishes, all the combination sets are further combined to generate new molecular fragments. Thus one important feature of GROW is that a new generation of new molecular fragments will grow out from one old molecular fragment. These new fragments are the derivatives of the old one, and, in principle, all of the possible derivatives can be obtained. That is why we claim our program is the computer simulation of combinatorial chemistry. These new molecular fragments will be subjected to the following tests before they can be output into the next step.

(1) Ring closure: The program will check the distance between each pair of nonbonded atoms in the molecular fragment. Those molecular fragments with internal collisions will be discarded. But if two atoms have overlapped or happened to be at a distance of covalent bond length, a ring will form in the fragment. The program will make necessary modifications to the fragment to close the ring. Only five- to eight-membered rings are considered to be reasonable. Others will be rejected.

(2) Chemical rationality: Here we mainly care about the percentage of heteroatoms, i.e., oxygen and nitrogen, in a molecular fragment. If it is higher than the preset criteria in the program, this molecular fragment will not pass the test because such a molecule might be unstable and difficult to synthesize. If it is necessary, more rules can be added in this module.

(3) Number of seed atoms: If there are too many seed atoms in a molecular fragment, the possible growing combinations will increase rapidly in the next growing cycle, which will probably give rise to combinatorial explosion. So if the number of seed atoms in a molecular fragment is larger than a preset limitation, this molecular fragment will be expunged. This is another important way to control the combinatorial explosion.

Usually there are thousands of molecular fragments which will pass these tests. All of them will be subjected to the next module SCORE.

**2.4. Ranking the Generated Molecular Fragments (SCORE).** Module SCORE can give scores to the generated molecular fragments and rank them in terms of their scores. Only the molecular fragments with top scores will be allowed to enter into the next growing cycle. In our program, we usually allow no more than 100 such fragments to become the templates for further growing. This strategy is similar to natural selection: the good ones continue to develop while the bad ones die out. Due to this "best-first" selection, the final results can be assured to be good enough, and the combinatorial explosion can also be moderated effectively.

Since ranking the candidates molecules is an important task for practical applications in drug design, the philosophy of scoring is worth much consideration. The central assumption of structure-based drug design is that good ligands must possess significant structural and chemical complementarity to their target receptor. In our program, we use force field to evaluate the quality of fitness, and we have uniquely coupled a rule-based approach to the scoring algorithm judging chemical rotionality and synthetic feasibility.

Several features of a generated molecule are considered in SCORE:

(1) Score of binding energy: The nonbonded interaction between the ligand and the receptor is described by the following formula

$$S_b = \sum_i \sum_j \left( \frac{A_{ij}}{d_{ij}^{2n}} - \frac{B_{ij}}{d_{ij}^{n}} + \frac{q_i q_j}{\epsilon\, d_{ij}} \right)$$

where $S_b$, score of nonbond interaction; $i$, the $i$th atom in the ligand; $j$, the $j$th atom in the receptor; $A$, $B$, Lennard-Jones constants, from AMBER force-field;[45] $d_{ij}$, distance between atom $i$ and atom $j$; $n$, power, the default value is 6; $q_i$, $q_j$, partial charges on atom $i$ and atom $j$; and $\epsilon$, dielectric constant. If flexible binding is required, a soft potential can be introduced by varying the power $n$ in the formula. The atomic partial charges of the ligand are computed by the Gasteiger−Hückel method.[46,47]

(2) Score of steric energy: The binding conformation of a ligand should have relatively low energy. Since the molecules have been generated by using standard bond lengths and bond angles, the steric energy is restricted to the interval van der Waals interaction. It is also defined in the Lennard-Jones function

$$S_s = \sum_i \sum_j \left( \frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^{6}} \right)$$

where $S_s$ is the score of steric energy and $i,j$ represent all the pairs of nonbonded atoms in the molecule.

(3) Score of chemical structure: This criterion is intended to assure the chemical rationality and synthetic feasibility. Unreasonable atom combinations on seed atoms have been identified and eliminated while growing atoms in module GROW. But that is not enough. A set of rules have been set up here to check the constitution of a molecule. The principal ones are listed below:

(i) Some functional groups, such as ester and amide groups, are awarded with a positive score. This is because they are generally good building blocks of drug molecules, and they will make the molecules easier to synthesize.

(ii) Some unreasonable groups, such as enol and peroxide, are given a negative score because they are chemically unstable or they will cause many side effects in the body.

(iii) Asymmetric carbon, heteroatom, or double bond are reasonable in a molecule, but too many of them will be penalized too.

(iv) Rings are dealt with by additional rules. When rings are formed in a molecule, they will increase the inner strain. A molecule with rings will become unfavorable especially in steric energy. But it is usually believed that rigid rings will smoothen the binding of ligands with the receptor by dropping the entropy decreases. Provided with this, the compensation for ring formation is necessary, i.e., rings will get extra positive scores. If you care more about that a molecule with rings maybe is more difficult to synthesize, this compensation can be assigned to be negative.

By setting these rules, we have made an attempt to develop a rule-based expert system to judge chemical rationality and synthetic feasibility. This requires the knowledge background of both organic and medicinal chemistry. In this way, molecules with unreasonable structures or unwanted features will be prevented from entering the next growing cycle. Those rules employed by our present program are by no means perfect, and we are still working on developing more elaborate ones.

The total score of a molecule is the summation of all the terms above. As described previously, all the terms are adaptable to the users' favor and the weights of each terms can also be adjusted. Thus the program will be induced on some extent to generate what the users like. At the end of the procedure of SCORE, the molecular fragments are sorted by their scores. A number of molecular fragments with top scores are selected out and put into the next growing cycle.

**2.5. Final Results.** During the process of atom growing, there are a number of molecular fragments with no seed atoms left, e.g., all the growing sites have ended in hydrogen or $sp^2$ oxygen atoms. They are not molecular fragments anymore but complete molecules. If the number of atoms in a molecule ranges from 20 to 40, which is a common feature of good drug molecules, the molecule will be recorded in a special file. After the whole iterative procedure of growing finishes, these molecules will be ranked by SCORE, and the top 100 ones will be put out in PDB format for further consideration.

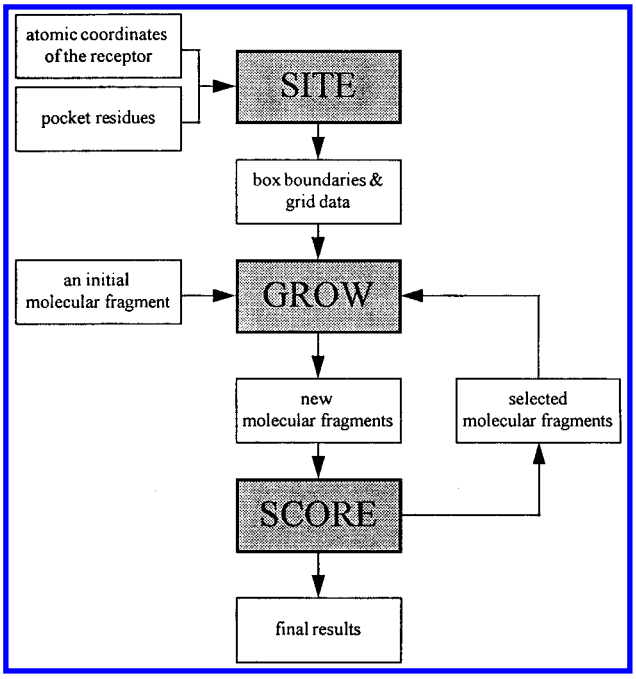


**Figure 1.** Flow chart of the computer program RASSE.

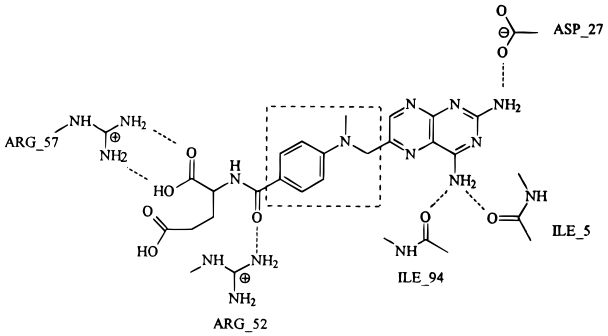


**Figure 2.** Chemical structure of MTX and the initial fragment used in the present calculations.

## 3. APPLICATIONS

We envision three basic uses for RASSE: (1) simulation of combinatorial chemistry, providing chemical diversity; (2) modification of known structures; and (3) an iterative build-and-choose process, possibly from a small core. We have applied RASSE to the design of inhibitors for *E. coli* dihydrofolate reductase and human phospholipase $A_2$ to test these properties.

**3.1. *E.coli* Dihydrofolate Reductase.** The 3D structure of *E. coli* dihydrofolate reductase (DHFR) complexed with the anticancer drug methotrexate (MTX) was solved by Bolin et al.[48] (entry 4DFR in the Brookhaven Protein Data Bank[49]). The complex structure, without water molecules, was used in the present calculations. The chemical structure of MTX together with its hydrogen bonding with DHFR is shown in Figure 2. The purpose of the calculations was to examine whether RASSE would suggest structures bound in approximately the same way of MTX.

We have noticed that MTX adopts an unusual L-shaped conformation in the complex structure. This clearly reflects the image of the binding pocket. So we used the middle part of MTX, which locates right at the corner of the "L",

**Table 1.** Atom Types Used in RASSE

| Atom type | Pattern | Description | Growing set | Growing sites | Growing probability |
|---|---|---|---|---|---|
| C1 | R—C< | $sp^3$ carbon | C1,C2, N1, N2, O1,CA,H | 3 | 1.0 |
| C2 | R—C< | $sp^2$ carbon | double bond: C3, N3, O2 single bond: C1, C2, N1, N2,O1,CA,H | 2 | 0.8 |
| C3 | R=C< | $sp^2$ carbon | C1,C2,N1,N2,O1,CA,H | 2 | 0.8 |
| N1 | R—N< | $sp^3$ nitrogen | C1,C2,N1,N2,CA,H | 2 | 0.5 |
| N2 | R—N< | $sp^2$ nitrogen | C3,N3 | 1 | 0.1 |
| N3 | R=N< | $sp^2$ nitrogen | C1,C2,N1,N2,CA,H | 1 | 0.1 |
| O1 | R—O< | $sp^3$ oxygen | C1,C2,CA,H | 1 | 0.5 |
| O2 | R=O | $sp^2$ oxygen | | 0 | 0.5 |
| CA | | aromatic carbon | C1,C2,N1,N2,O1,CA,H | 1 | 0.3 |
| H | R—H | hydrogen | | 0 | 1.0 |

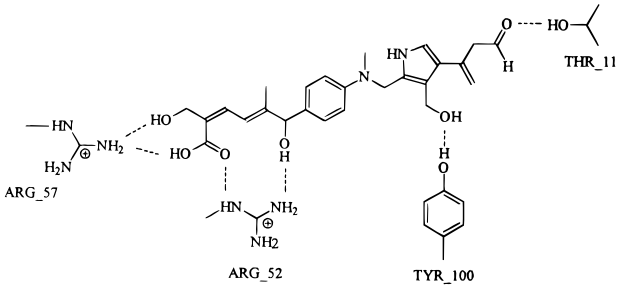


**Figure 3.** A designed inhibitor for DHFR

as the initial fragment (see Figure 2, in dotted frame). Because the peteridine ring was not used as the initial fragment, we eliminated the conserved water molecule which binds to the peteridine ring of MTX. After eight cycles of growing, RASSE selected 70 structures among 1324 generated molecules. Most of them have similar orientations to that of MTX and reproduce the hydrogen bonds with DHFR. Figure 3 shows the suggested structure with the highest score. We found this molecule could form hydrogen bonds with ARG_52 and ARG_57, which were similar to those of MTX. Moreover, it explored two new hydrogen bonding sites, namely TYR_100 and THR_113. A stereoview of the superimposed structures of this molecule and MTX is shown in Figure 4a where the solid line and the dotted line represent the designed inhibitor and MTX, respectively. It indicates that the designed inhibitor closely mimics MTX. We optimized the conformation of this molecule in the binding pocket by QUANTA/CHARMm. The optimized structure was compared with the nonoptimized one in Figure 4b, where the solid line and the dotted line shows the nonoptimized and optimized structures, respectively. The root mean square deviation of these two conformations is 0.93 Å. From the similarity of these two conformations, it is strongly suggested that the original one is sufficiently reasonable.

**3.2. Human Phospholipase $A_2$.** The second test case involves human phospholipase $A_2$ ($PLA_2$) complexed with a transition-state analogue GEL[50,51] (entry 1POE in the Brookhaven Protein Data Bank[49]). The chemical structure of GEL is shown in Figure 5. We found the binding pocket of $PLA_2$ was rather large. So in this case, we focused on testing the *de novo* design ability of RASSE. A relatively small initial fragment, that is, the glyceryl group in GEL which only contains four atoms, has been used in the calculations (see Figure 5 and Figure 6, in dotted frame). We ran RASSE several times, each time using a different set of regulating parameters, and we were interested to see that RASSE had suggested several families of novel structures, as shown in Figure 6 where $R_1$, $R_2$, $R_3$, and $R_4$ represent various substitution groups on the backbones of the designed structures. Most of the structures seem to have reasonable conformation and prove to have shapes well fitting the receptor cavity (see Figure 7).

## 4. DISCUSSION AND CONCLUSION

We have developed a new computer program RASSE for automated structure-based drug design. The program generates molecules by growing atoms in the binding pocket. All the candidate molecules are then subjected to a scoring module where the best ones will be picked out. Two applications of RASSE have been presented in this paper. The results have shown that RASSE can generate molecules with rational conformations, reasonable chemical structures, and strong abilities of binding to the specific receptor.

The ultimate goal of the present approach is an effective automated design of protein ligands. Because of the insufficient knowledge of the interactions between a ligand and its target protein, one should realize that we are still far away from getting the best answer in one run. The usefulness of our method lies in providing "raw material" for further collaborative modeling studies between synthetic and computational chemists. This has been recognized satisfactorily in our program.

We have adopted the strategy of combinatorial chemistry to explore the potential ligands for a specific receptor. In our program, different types of atoms serve as the "building-blocks" and we combine them to generate molecules. In the test cases, we established that our program can generate a wide variety of reasonable structures. The number of generated molecules is so large that it usually takes several hours to run RASSE with seven or eight growing cycles on SGI INDIGO2/R4000. In order to control the combinatorial

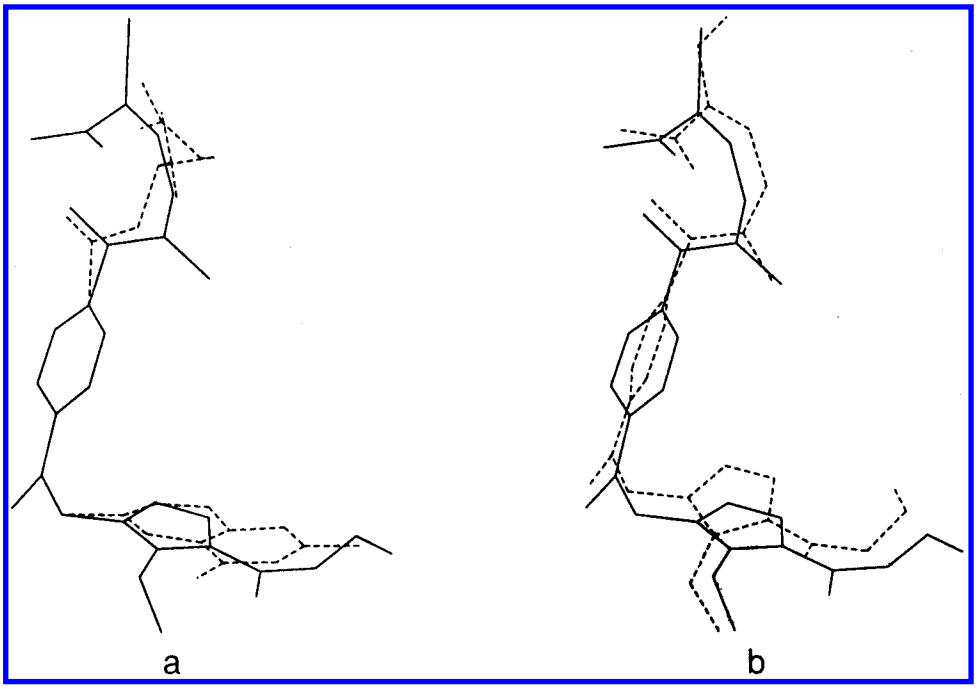


**Figure 4.** (a) Superimposed structures of the designed inhibitor and MTX (b) comparison of the optimized and nonoptimized structures.

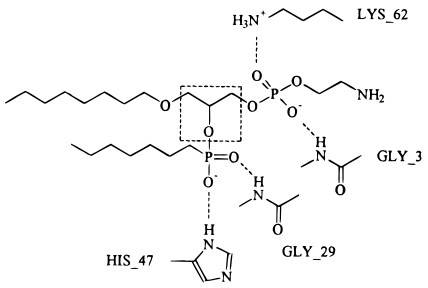


**Figure 5.** Chemical structure of GEL and the initial fragment used in the present calculations.

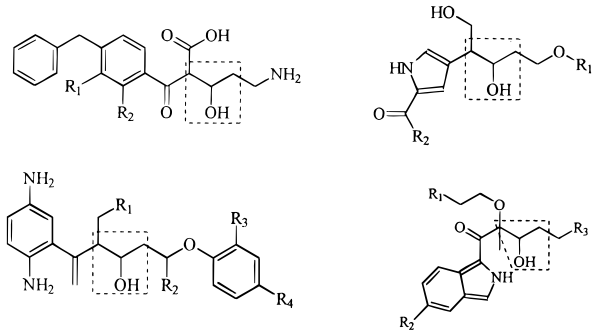

**Figure 6.** Families of designed inhibitors for $PLA_2$.

explosion, we have taken several measures: (1) Introduce an initial fragment. This fragment will occupy part of the pocket and reduce the conformational space to search. (2) Eliminate the unreasonable fragments generated during growing. These fragments either have unreasonable chemical structure or have van der Waals collisions with the receptor. Eliminating them will prevent additional meaningless calculations. (3) Limit the number of seed atoms in one molecular fragment. More seed atoms mean more possible derivatives. Limiting the number of them helps to control the combinatorial explosion. This measure will also simplify the chemical structure of the generated molecules, which makes them easier to synthesize. (4) Limit the number of generated molecular fragments. Although the first three measures have been taken, combinatorial explosion does occasionally happen. That will be a disaster both for CPU time and for disk space. So we have to set a limitation of the number of generated molecular fragments in every growing cycle. In practice, we usually set this number of 20 000. It needs approximately 30MB disk space to store these molecules.

Another idea implanted in RASSE is that only a number of molecules with top scores are allowed to enter into the next growing cycle. Thus the bad ones are removed, while the best ones are kept for further growing. This is similar to natural selection. The criteria of selection is the scores for molecular fragments. Ranking the candidate molecules is of vital importance in drug design, and it remains a developing area in theoretical chemistry. The factors which contribute to binding include hydrogen bonding, hydrophobic effect, van der Waals interactions, solvation effects, and other electrostatic interactions. Our present algorithm of scoring evaluates the nonbond interactions between ligand and receptor by the force field calculations. How to take into account hydrogen bonding, hydrophobic effect, and solvation effects explicitly and effectively is still in consideration. Another drawback underlying our algorithm is that the receptor is assumed to be rigid. Flexible binding is not accounted for. It is well-known that ligand-induced conformational changes of the protein occur upon ligand binding. We make a compromise by introducing "soft atoms" in our program, which allows two atoms to overlap to some extent.

It is worth remembering that tight binding to an receptor is a necessary but not sufficient condition for a successful drug. A good ligand should also have good chemical properties and can be easily synthesized. We have attempted to adopt a set of rules to evaluate the chemical rationality and syntheses feasibility of the candidates. These rules are based on our knowledge of organic and medicinal chemistry. Our program will identify the functional groups in a molecule, award the "good" ones, and penalize the "bad" ones. This method has greatly helped RASSE to generate chemically reasonable molecules. But these rules are not

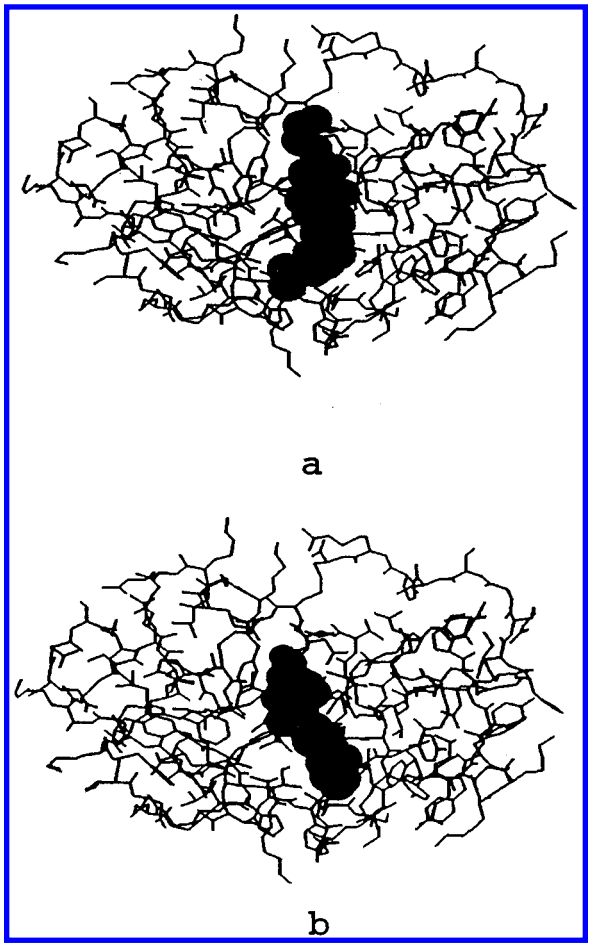


**Figure 7.** Binding mode of GEL(a) and that of a designed inhibitor (b).

perfect yet. Some strange-looking molecules still appear occasionally. An organic chemist may point out easily whether a compound is easy to synthesize or not or whether a compound is stable or not, but it is difficult to teach a computer to do that. More elaborate rules are currently under development.

Despite the limitations, RASSE has proven to be a helpful new tool for structure-based drug design. It offers several interesting features to the drug designer:

(1) It simulates combinatorial chemistry on the computer. An unbiased list of molecules can be provided, and this may keep the drug designer from overlooking possibilities in drug design.
(2) This program can perform *de novo* design as well as proposing new substituents for an existing ligand.
(3) Rule-based evaluation of chemical rationality and syntheses feasibility is employed in the program. It helps the program give more reasonable results.
(4) The program is easy to use. It can be applied to different receptors without any modification. The usage of the PDB format for output files makes interfacing to existing molecular modeling software simple.

## ACKNOWLEDGMENT

This work was supported by the Chinese High-Technology Project, the State Key Projects of Fundamental Research, and the Trans-Century Training Program Foundation for the Talents.

## REFERENCES AND NOTES

(1) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078−1082.
(2) Dixon, J. S. Computer-aided drug design: getting the best results. *TIBTECH* **1992**, *10*, 357−363.
(3) Christophe, L. M. J.; Hol, W. G. J. Structure-based drug design: progress, results and challenges. *Structure* **1994**, *2*, 577−587.
(4) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *J. Med. Chem.* **1994**, *37*, 1035−1054.
(5) Kuntz, I. D.; Meng, E. C.; Shoichet, B. K.; Structure-Based Molecular Design. *Acc. Chem. Res.* **1994**, *27*, 117−123.
(6) Lybrand, T. P. Ligand-protein docking and rational drug design. *Curr. Opin. Struct. Biol.* **1995**, *5*, 224−228.
(7) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(8) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. *J. Med. Chem.* **1989**, *32*, 1083−1094.
(9) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(10) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722−729.
(11) Meng, E. C.; Schoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(12) Sheridan, R. P.; Nilakantan, R.; Rusinko, III, A.; Bauman, N.; Haraki, K. S.; Venkataraghavan, R. 3DSEARCH: A System for Three-Dimensional Substructure Searching. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 255−260.
(13) Van Drie, J. H.; Weininger, D.; Martin, Y. C. ALLADIN: an integrated tool for computer-assisted molecular design and pharcophore recognition from geometric, steric and substructure searching of three-dimensional molecular structures. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 225−254.
(14) Lauri, G.; Bartlett, P. A. CAVEAT: A program to facilitate the design of organic molecules. *J. Comput.-Aided Mol. Des* **1994**, *8*, 51−66.
(15) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. FLOG: A system to select "quasi-flexible" ligands complementary to a receptor of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 153−174.
(16) Böhm, H.-J. The computer program LUDI: A new method for the *de novo* design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61−78.
(17) Böhm, H.-J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593−606.
(18) Böhm, H.-J. On the use of LUDI to search the Fine Chemicals Directory for ligands of proteins of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 623−632.
(19) Lawrence, M. C.; Davis, P. C. CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins: Struct. Funct Genet.* **1992**, *12*, 31−41.
(20) Ho, C. M. W.; Marshall, G. R. SPLICE: A program to assemble partial query solutions from three-dimensional database into novel ligands. *J. Comput-Aided Mol. Des.* **1993**, *7*, 623−647.
(21) Rostein, S. H.; Murcko, M. K. GroupBuild: A Fragment-Based Method for *De Novo* Drug Design. *J. Med. Chem.* **1993**, *36*, 1700−1710.
(22) Gillet, V.; Johnson, A. P.; Mata, P.; Sike, S.; Williams, P. SPROUT: A program for structure generation. *J. Comput-Aided Mol. Des.* **1993**, *7*, 127−153.
(23) Gillet, V. J.; Newell, W.; Mata, P.; Myatt, G. J.; Sike, S.; Zsoldos, Z.; Johnson, A. P. SPROUT: Recent Developments in the de Novo Design of Molecules. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 207−217.
(24) Mata, P.; Gillet, V. J.; Johnson, A. P.; Lampreia, J.; Myatt, G. J.; Sike, S.; Stebbings, A. L. SPROUT: 3D Structure Generation Using Templates. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 479−493.
(25) Leach, A. R.; Kilvington, S. R. Automated molecular design: A new fragment-joining algorithm. *J. Comput-Aided Mol. Des.* **1994**, *8*, 283−298.
(26) Nishibata, Y.; Itai, A. Automatic Creation of Drug Candidate Structures Based on Receptor Structure: Starting Point for Artificial Lead Generation. *Tetrahedron* **1991**, *47*, 8985−8990.
(27) Nishibatta, Y.; Itai, A. Confirmation of Usefulness of a Structure Construction Program Based on Three-Dimensional Receptor Structure for Rational Lead Generation. *J. Med. Chem.* **1993**, *36*, 2921−2928.

(28) Rotstein, S. H.; Murcko, M. A. GenStar: A method for *de novo* drug design. *J Comput.-Aided Mol. Des.* **1993**, *7*, 23−43.
(29) Pearlman, D. A.; Murcko, M. A. CONCEPTS: New Dynamic Algorithm for *De Novo* Drug Suggestion. *J. Comput. Chem.* **1993**, *14*, 1184−1193.
(30) Ivanov, J.; Karabunarliev, St.; Mekenyan, O. 3DGEN: A System for Exhausive 3D Molecular Design Proceeding from Molecular Topology. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 234−243.
(31) Gehlhaar, D. K.; Moerder, K. E.; Kichi, D.; Sherman, C. J.; Ogden, R. C.; Freer, S. T. *De Novo* Design of Enzyme Inhibitors by Monte Carlo Ligand Generation. *J. Med. Chem.* **1995**, *38*, 466−472.
(32) Glen, R. C.; Payen, A. W. R. A genetic algorithm for the automated generation of molecules within constraints. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 181−202.
(33) Roe, D. C.; Kuntz, I. D. BUILDER v.2: Improving the chemistry of a de novo design strategy. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 269−282.
(34) Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. PRO_LIGAND: An approach to de novo molecular design. 1. Application to the design of organic molecules. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 13−32.
(35) Waszkowycz, B.; Clark, D. E.; Frenkel, D.; Li, J.; Murray, C. W.; Robson, B.; Westhead, D. R. PRO_LIGAND: An Approach to *de novo* Molecular Design. 2. Design of Novel Molecules from Molecular Field Analysis(MFA) Models and Pharmacophores. *J. Med. Chem.* **1994**, *37*, 3994−4002.
(36) Westhead, D. R.; Clark, D. E.; Frenkel, D.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B. PRO_LIGAND: An approach to de novo molecular design. 3. A genetic algorithm for structure refinement. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 139−148.
(37) Frenkel, D.; Clark, D. E.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. PRO_LIGAND: An approach to de novo molecular design. 4. Application to the design of peptides. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 213−225.
(38) Clark, D. E.; Murray, C. W. PRO_LIGAND: An approach to de novo molecular design. 5. Tools for the Analysis of Generated Structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 914−923.
(39) Murray, C. W.; Clark, D. E.; Byrne, D. G. PRO_LIGAND: An approach to de novo molecular design. 6. Flexible fitting in the design of peptides. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 381−395.
(40) Barakat, M. T.; Dean, P. M. The atom assignment problem in automated de novo drug design. 1. Thansferability of molecular fragment properties. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 341−350.
(41) Barakat, M. T.; Dean, P. M. The atom assignment problem in automated de novo drug design. 2. A method for molecular graph and fragment perception. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 351−358.
(42) Barakat, M. T.; Dean, P. M. The atom assignment problem in automated de novo drug design. 3. Algorithms for optimization of fragment placement onto 3D molecular graphs. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 359−372.
(43) Barakat, M. T.; Dean, P. M. The atom assignment problem in automated de novo drug design. 4. Tests for site-directed fragment placement based on molecular complementarity. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 448−456.
(44) Barakat, M. T.; Dean, P. M. The atom assignment problem in automated de novo drug design. 5. Tests for envelop-directed fragment placement based on molecular similarity. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 457−462.
(45) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(46) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity − A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(47) Purcel, W. P.; Singer, J. A. A Brief Review and Table of Semiempirical Parameters Used in the Huckel Molecular Orbital Method. *J. Chem. Eng. Data* **1967**, *2*, 235−246.
(48) Bolin, J. T.; Filman, D. A.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. Crystal Structures of *Escherichia coli* and *Lactobacillus casei* Dihydrofolate Reductase Refined at 1.7 Å Resolution. *J. Biol. Chem.* **1982**, *257*, 13650−13662.
(49) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, M. D., Jr.; Brice, E. F.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(50) Scott, D. J.; White, S. P.; Browning, J. L.; Rosa, J. J.; Gelb, M. H.; Sigler, P. B. Structures of Free and Inhibited Human Secretory Phospholipase $A_2$ from Inflammatory Exude. *Science*, **1991**, *254*, 1007−1010.
(51) Scott, D. L.; White, S. P.; Otwinowski, Z.; Yuan, W.; Gelb, M. H.; Sigler, P. B. Interfacial Catalysis: The Mechanism of Phospholipase $A_2$. *Science* **1990**, *250*, 1541−1546.

CI950277W