Starting Point for Artificial Lead Generation.

Yoshihiko Nishibata and Akiko Itai*

Faculty of Pharmaceutical Sciences, University of Tokyo
7-3-1, Bunkyo-ku, Tokyo, 113, Japan

(*Received in Japan* 12 *September* 1991)

Key Words: Lead Generation; Computational Chemistry; Drug Design

Abstract: We have developed a new method for automatic generation of drug candidate structures based on a known receptor structure. In our method, various structures which fit well to the receptor cavity are generated, by adding atoms one by one using a force field and random numbers. The usefulness of the program was exemplified by application to the E. coli dihydrofolate reductase system. From dozens of generated structures, we could obtain several promising new structures with considerable internal stability and having favorable interactions with the receptor cavity. It is expected that this method will become an essential starting point for artificial lead generation, which has been impossible so far.

Introduction

For the purpose of developing excellent drugs efficiently, it is necessary to establish rational approaches for drug design. To develop methods for finding lead compounds artificially and rationally is especially important, since such compounds have mostly been found by chance so far. The drug-receptor interaction seems to be the most useful basis for that purpose. The real nature of the drug-receptor interaction has been clarified in a number of cases by determination of the three-dimensional structures of protein-ligand complexes at the atomic level by X-ray crystal analyses. There are many drug molecules which are known to bind to the same site of a protein, in spite of large discrepancies in chemical structures. This is because it is neither the chemical structure nor the molecular skeleton, but rather the complementarities in molecular shape and submolecular physical and chemical properties that are important for specific binding to the same receptor site. This strongly suggests that if we can design and synthesize a molecule with a molecular shape and submolecular properties which will complement those of the receptor cavity, it should be able to bind to the receptor specifically. But, it is not easy to design such molecules with new skeletal structures manually due to lack of objectivity.

In recent years, techniques for solving protein crystal structures have made remarkable progress. Moreover, there have been great advances in biochemical techniques such as isolation, purification and protein engineering. A number of biologically important protein structures have been elucidated or are being elucidated. The three-dimensional structures of proteins have been used for interpreting the biological activities and elucidating the biochemical mechanisms involved by docking simulation. One of the rational approaches to the modification of ligand structures (mainly by replacing or adding substituent groups) has been the docking simulation, although it has not been utilized directly for designing new structures so far.

In order to generate lead compounds with new skeletal structures *ab initio*, we require new strategies. If computers can provide us with possible ligand structures which can strongly bind to the protein, we would be able to generate new lead compounds artificially, based on the individual features of the receptor structure.

We have developed a new method and a computer program for this purpose,[1] i.e., for generating drug candidate structures which fit well to the receptor cavity, based on the receptor structure. We named the program "LEGEND". A large number of possible structures generated are selected by another program, named "LORE", based on energetic and structural considerations. Here, we describe the method and the results of its application to the *E. coli* dihydrofolate reductase system.

Methods

For the *de novo* generation of a molecular structure, we must prepare positions and types of atoms, and types of bonds in the molecule. The program provides sixteen atom types which discriminate the combination of atomic element and its hybridized state, i.e., sp^3 carbon, aromatic carbon, carbonyl oxygen, amino nitrogen and so on. It provides five bond types, which are single, double, triple, aromatic and amide.

The size of the structure to be generated is specified by a number of atoms in the input data. The relative ratio of appearance in a structure for each atom type is given in the program. Internal atomic positions can be defined by geometrical parameters (bond lengths and angles) and conformational parameters (dihedral angles). The former can be assumed to be the standard values used in the conventional force field. But, the latter should be determined by using random numbers, as should the atom type and the bond type. In this program, we make use of random numbers in order to determine all the unsettled quantities or to choose items from those prepared in the program. Random numbers used are those output from the computer sequentially.

The atomic coordinates of a protein molecule are read in PDB file format.[2] The preparations before starting LEGEND are as follows: for the high-speed calculation of the intermolecular interaction energies using tabulated data, a three-dimensional grid is generated inside the ligand binding site of the protein.[3] Then, at each grid point, van der Waals interaction energies are calculated between protein atoms and a probe (carbon, nitrogen, oxygen, hydrogen) atom located on the grid. The program uses the MM2[4] force field and parameters. The electrostatic potential at each grid point is also computed using the atomic charges on the protein atoms which are taken from those for individual amino acid residues in the AMBER program.[5] The tabulated data are used for energy estimation at every step of new atom generation and also for structure optimization of the generated raw molecule.

The fundamental process of a structure generation by the LEGEND program consists of three steps, as follows. The process is shown in detail in the flow chart in Fig. 1.

Stage 1 : Generation of the first atom. An anchor atom is selected by use of a random number from among several hydrogen-bonding heteroatoms in the protein, specified beforehand. The position and the atom type of the first atom are determined so as to make a hydrogen bond to the anchor atom.

Step 2 : Subsequent generation of atoms. The second atom and subsequent atoms are generated one by one by the following procedure, up to the specified number of atoms for a molecule. For every new atom, a root atom is chosen from all the previously generated atoms by using a random number. The atom type of the new atom and the bond type of the bond between the root atom and the new atom are also given by random numbers. Then, the position of the atom is determined by random numbers by choosing a point on the circle which is defined by the bond length and the bond angle from the root atom. The values of the bond length and angle used are assigned according to the array of related atom types, taken from the MM2 program. If the position of the atom is not acceptable due to the violation of van der Waals radii of the previously generated atoms or unstable intermolecular van der Waals interaction energy, the program reassigns the root atom and attempts to find an acceptable new atom. If the attempts fail after a given number of repeats, the program tracks back to the

preceding step, i.e., it withdraws the last one of previously generated atoms and re-generates a new atom.

Step 3 : Completion of the molecular structure. The program completes by adding missing carbon atoms to fragmentary aromatic rings, and supplies hydrogen atoms for all remaining valencies of all nonhydrogen atoms. The atomic charges in the molecule is calculated by Del Re's method.[6] Finally, the structure is optimized by the Simplex method.

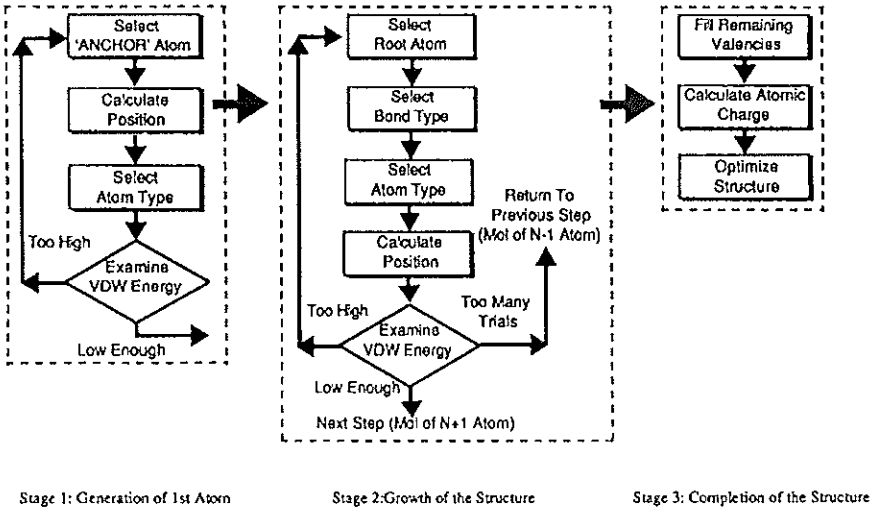

Fig.1. Flow chart of the LEGEND program

Thus, the LEGEND program goes on generating structures one after another up to the maximum number of structures specified in the input data. From among the generated structures, a rather small number of structures are selected by the program LORE. Selections can be made on the basis of various energetic values, as well as some indices related to structural features.

An Application to E. coli dihydrofolate reductase

In order to verify the usefulness of our method, we have applied the program system to *E. coli* dihydrofolate reductase, whose three-dimensional structure has been elucidated by X-ray crystallographic analysis as a ternary complex with coenzyme NADPH and folic acid.[7] The atomic coordinates are available from the Protein Data Bank. We have used the protein structure bound with NADPH, removing the folic acid molecule. Three hydrogen-bonding atoms, the carboxyl oxygen of ASP 27, carbonyl oxygen of ILE 5 and carbonyl oxygen of ILE 94 were chosen as candidates for the anchor atom. A full automatic structure generation by the LEGEND program was performed using the following conditions: the number of atoms in a molecule 30; the number of molecules to be generated 300; the minimum and maximum threshold energies 6.0 kcal/mol and 12.0 kcal/mol, respectively; the maximum number of iterations in atom generation 20; the number of iterations in backtracking 3; the minimum number of rings in a molecule 2.

A total of 300 structures were generated by the LEGEND program. Nine structures were selected with the LORE program by using the following criteria: minimum number of hydrogen bonds

2; the maximum inter-molecular van der Waals and electrostatic energies 50.0 kcal/mol and the maximum total (inter- and intra-molecular) van der Waals energies 50.0 kcal/mol.

```
***** LORE : Legend Output Retrieval Engine *****

 LOAD : Select 9 Mols from 300 Mols

 NO    ID   HB RING    VDW ENERGY          ELECTROSTATIC      TOTAL
                       Inter     Intra     Inter     Intra
-------------------------------------------------------------------
  1   198    2    2    -6.580    12.993    -8.631   -97.114   -99.332
  2   165    2    3    -2.519    38.937     0.034  -113.576   -77.124
  3   130    2    2    -0.036    20.224    -2.122  -134.858  -116.793
  4   279    2    2    -0.794    13.442     2.267  -210.191  -195.276
  5    73    2    2     8.250    28.537    -5.850  -178.076  -147.140
  6   120    2    2    18.722    11.284   -13.345  -131.287  -114.626
  7    48    2    2    23.510     9.330    -1.531  -128.866   -97.556
  8   275    3    2    20.086     7.782     4.122  -113.913   -81.924
  9   269    2    3    50.606     9.926    -7.485   -85.902   -32.855

...

INFORMATION for MOL# 2
VDW:    N( 1) ---      C of ILE5     r =   3.617
HBOND: N( 1) ---      O of ILE5
HBOND: N( 1) ---     OH of TYR100
VDW:    H( 2) ---      C of ILE5     r =   2.588
VDW:    H( 2) ---      O of ILE5     r =   2.029
VDW:    H( 2) ---      N of ALA6     r =   3.120
VDW:    H( 2) ---     CA of ALA6     r =   3.047
VDW:    H( 2) ---     OH of TYR100   r =   2.805
VDW:    H( 2) --- H(CG2) of ILE5     r =   2.592
VDW:    H( 2) ---  H(CA) of ALA6     r =   2.124
...
```

Fig. 2. Output from the LORE program

In addition to the file output of the three-dimensional atomic coordinates of the selected structures, the LORE program outputs a summary of the selected structures as shown in Fig. 2. The various energy values and some structural features for the nine structures (upper), and the intermolecular distance information (and interaction type) for each structure (lower) are listed up. Some of chemical structures of the output structures from LORE are shown in Fig. 3.

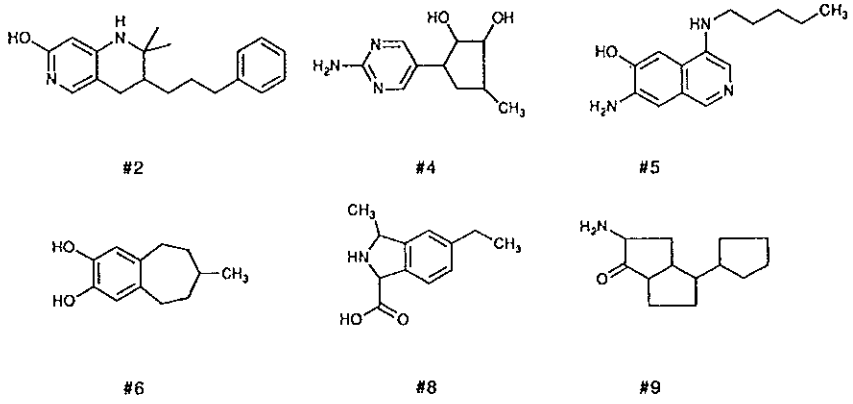

Fig. 3. Chemical structures chosen with LORE from 300 generated structures

In order to examine the conformational and geometrical stability of the generated structures, we have optimized one of the structures by the PM3 method using the MOPAC program.[8] (In this example, structure #8 was selected.) The optimized structure was compared with the non-optimized one by the least-squares superposing method. A stereoview of the superposed structures is shown in Fig. 4. The solid line and the dotted line shows the non-optimized and optimized structures respectively. From the high similarity of the two structures, it is strongly suggested that the original, non-optimized structure is sufficiently stable.

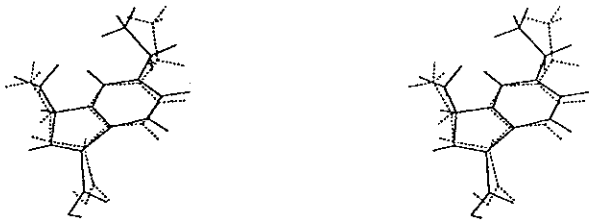

Fig. 4. Comparison of the PM3-optimized and non-optimized structures

In Fig. 5, aspects of the intermolecular interactions with the target protein are shown for structure #8. The hydrogen bonds reported in the LEGEND output are represented by dotted lines. These hydrogen bonds were searched by atom type, distance of heteroatoms and bond angle of hydrogen bonds. In this case, the anchor atom is the carboxyl oxygen of ASP 27. Besides the hydrogen bond between the first atom and the anchor atom, the structure forms additional hydrogen bonds to the carbonyl oxygen of ALA 6 and the nitrogen of the indole group of TRP 30.

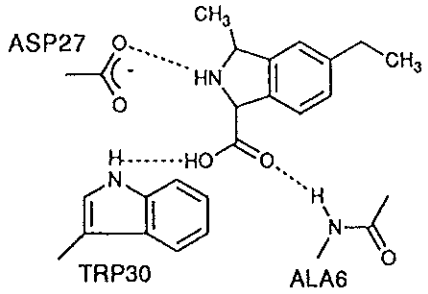

Fig. 5. Aspects of intermolecular hydrogen bonding with target protein

Results and Discussions

The purpose of this study was to develop a method to obtain diverse receptor-binding structures with suitable molecular shapes and with suitable functional groups at the proper positions and orientations in the molecule, covering all possible structures without prejudice. Lewis has proposed a method for the same purpose.[9] This is the only paper related to the present problem so far published. But, because his method places atoms in a molecule on the lattice points of a diamond lattice with a ridge line of the carbon-carbon covalent bond length, it produces only a limited kind of structures: the structures cannot contain sp^2 hybridized atoms, geometries apart from exact tetrahedral angles or conformations except for exact *trans* or *gauche* torsion angles. As our method is based on a new algorithm using random numbers and a force field, the structures generated are not only unlimited, but also not unstable internally.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.