Chinese Journal of Chemistry, **2004**, *22*, 882—887 **Article**

Discovery of Anti-SARS Coronavirus Drug Based on Molecular Docking and Database Screening

CHEN, Hai-Feng*,[a](陈海峰) YAO, Jian-Hua[a](姚建华) SUN, Jing[a](孙晶)
LI, Qiang[a](李强) LI, Feng[a](李丰) FAN, Bo-Tao[b](范波涛)
YUAN, Shen-Gang[a](袁身刚)

[a] *Laboratory of Computer Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China*
[b] *ITODYS, CNRS UMR 7086, Université Paris 7, 1, rue Guy de la Brosse, 75005 Paris, France*

The active site of 3CL proteinase ($3CL^{pro}$) for coronavirus was identified by comparing the crystal structures of human and porcine coronavirus. The inhibitor of the main protein of rhinovirus (Ag7088) could bind with $3CL^{pro}$ of human coronavirus, then it was selected as the reference for molecular docking and database screening. The ligands from two databases were used to search potential lead structures with molecular docking. Several structures from natural products and ACD-SC databases were found to have lower binding free energy with $3CL^{pro}$ than that of Ag7088. These structures have similar hydrophobicity to Ag7088. They have complementary electrostatic potential and hydrogen bond acceptor and donor with $3CL^{pro}$, showing that the strategy of anti-SARS drug design based on molecular docking and database screening is feasible.

Keywords coronavirus, $3CL^{pro}$, molecular docking

Introduction

The first case of severe acute respiratory syndrome (SARS) was identified in November, 2002, in Guangdong Province, China.[1] In March, 2003, the putative cause of SARS was identified as a new coronavirus.[2,3] SARS-Cov is a member of the coronoviridae family of enveloped, positive-stranded RNA viruses, which have a broad host range. The length of genome sequence for coronaviruses is about 27—32 kb and it could encode 23 putative proteins, including main proteinase (M^{pro}, also called $3CL^{pro}$), nucleocapsid (N), spike (S), membrane (M), and small envelope (E). Because the viral main proteinase ($3CL^{pro}$) controls the activities of the coronavirus replication complex, it is an attractive target for therapy and drug design.[4] A large number of compounds were synthesized and separated, in order to find anti-SARS lead compounds. Virtual screening has the advantages of that searching lead structures is cheaper than the real experiment and the calculation could be performed on compounds that are not yet purchased or synthesized.[5] So virtual screening was used widely to find initial lead structures from large compound collections. Therefore, some amounts of work were done to search the inhibitor of SARS.[6]

In this study, by using molecular docking and other screening filters we have screened several types of databases, such as in-house natural product and ACD screening databases. Some interesting results are reported below.

Methods

Active site identification

The crystal structure of main proteinase for human coronavirus was extracted from Brookhaven Protein Databank (PDB code: 1P9S). Hilgenfeld *et al.*[4] reported an inhibitor complex of porcine coronavirus and found that SARS coronavirus (SARS-Cov) main proteinase reveals a remarkable degree of conservation of the substrate-binding sites with porcine coronavirus. The PDB code of porcine coronavirus is 1LVO. We aligned these two structures with the routine in SYBYL6.9 of "alignment homology" and found that the homologous ratio between two amino acid sequences is 63%. The region of binding site for 1LVO is the same as that of 1P9S. It shows that this region of the substrate binding sites is remarkably conserved.[4] So we could identify the active site of human coronavirus. It is testified by means of the routine MOLCAD.[7]

Molecular docking

Hilgenfeld *et al.*[4] reported that rhinovirus $3CL^{pro}$ in-

* E-mail: hfchen@mail.sioc.ac.cn
Received August 22, 2003; revised March 10, 2004; accepted April 21, 2004.
Project supported by the Minister of Science and Technology of China (Nos. 2003CB114401 and 2002AA231011), the National Natural Science Foundation of China (No. 20073058), Science and Technology Committee of Shanghai (No. 02DJ14013), Chinese Academy of Sciences-National Center of Scientific Research in France Cooperation Program (CNRS/CAS No. 12475) and Embassy of France in China.

hibitors may be modified to make them useful for SARS therapy. Since Ag7088 (see Figure 1) has entered the clinical trials as the inhibitor of human rhinoviruses,[8] it is reasonable to be selected as the screening reference. Molecular dynamics was used to optimize its structure with Tripos force field.[9] Partial atom charges of $3CL^{pro}$ were calculated with Kollman-all-atom[10] approximation and Gasteiger-Hückel for two types of inhibitors. Ag7088 was set in the cavity of binding site. All calculations were performed on a workstation, SGI Origin 300 with 32 CPUs.

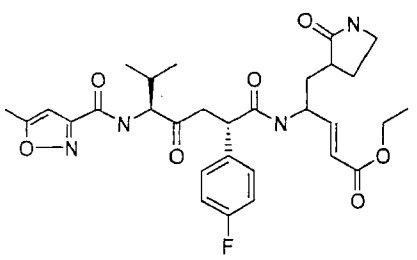


Figure 1 The structure of reference compound Ag7088.

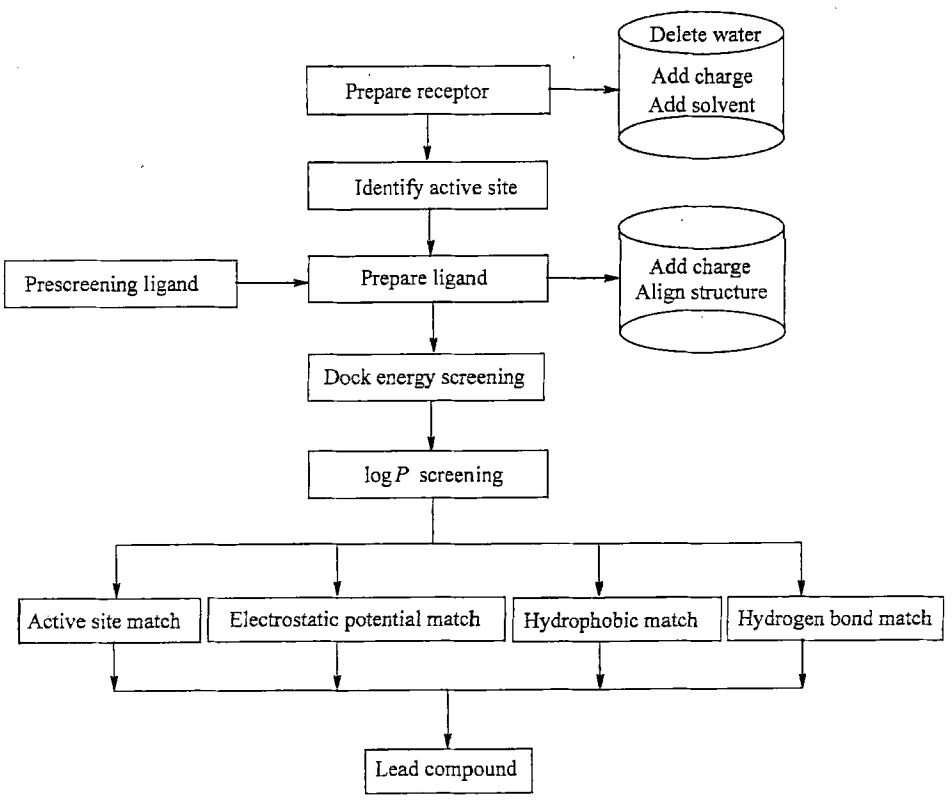
AutoDock3.0[11,12] is a suitable software for performing automated docking of ligands to their macromolecular protein receptors. The individual components of the program include AutoTors, AutoGrid, and AutoDock. AutoTors defines which bonds in the ligand are rotatable, affecting the degrees of freedom (DOF) of the ligand, and thus the complexity of the computations. AutoGrid pre-calculates a three-dimensional grid of interaction energies based on the macromolecular target using the AMBER force field. AutoDock can begin the process of simulation. First, the ligand moves randomly in any one of six degrees of freedom (either translation or rotation) and the energy of the new ligand "state" is calculated. If the energy of the new state is lower than that of the old state, the new one is automatically accepted as the next step in docking. During docking process, a maximum of 50 conformers was considered for each compound (default set is 10 conformers).

Design of screening strategy

Binding free energy is an important criterion for reliable virtual screening. The investigated ligand located at the active site of $3CL^{pro}$ is another necessary condition for a screening result. On the other hand, hydrophobic character is another important factor to drug design. It reflects whether a drug molecule could reach the surface of protein. Usually it can be estimated by the ester/water distribution coefficient (log P). Therefore in this study, log P is used as another criterion for virtual screening. At the process of virtual screening, we also think about the standard of electrostatic potential and formation of hydrogen bond. All these features are used as the filters of the virtual screening. The flow chart of virtual screening is shown in Figure 2.

Results and discussion

The active site of main protein ($3CL^{pro}$) is shown in Figure 3, which contains s1 pocket, s2 pocket and a cavity of canal. Because molecular modeling suggests that available rhinovirus $3CL^{pro}$ inhibitors may be modi-



fied to make them useful for treating SARS, we select Ag7088 as the reference of virtual screening.[4] The binding free energy of Ag7088 with $3CL^{pro}$ is −63.42 kJ/mol in the study. Their docking complex is illustrated in Figure 4. In the docking complex, there are hydrophobic and electrostatic interactions between Ag7088 and residues Asp186, Gln187, Pro188, Ser189 and one hydrogen bond with residue Gly167.

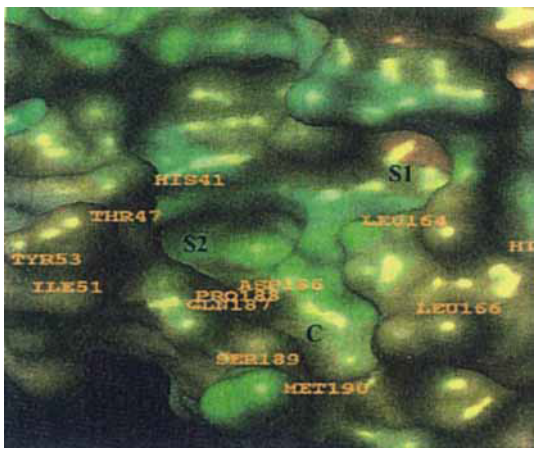


Figure 3 The active site of $3CL^{pro}$.

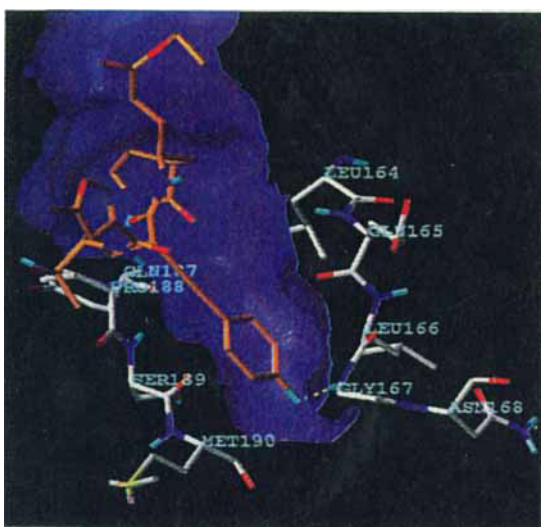


Figure 4 The complex of Ag7088 and $3CL^{pro}$.

In-house natural product database

In this study, 1541 natural product structures, which are selected from an in-house database containing more than 25000 structures collected from recent literature and principally from TCM pharmaceutical components, have been prescreened. From these 25000 structures by using the filters we firstly selected the compounds of macro lactones. Then they were aligned to Ag7088 and the binding free energies with $3CL^{pro}$ were calculated. We considered that their binding free energies for these 43 structures are less than −63.42 kJ/mol as the potential lead structures. The screening results are presented in Figure 5. The ratio between potential lead structures and the total screened structures is 2.79%. This shows that Ag7088 is a good reference structure for screening and our screening strategy is feasible.

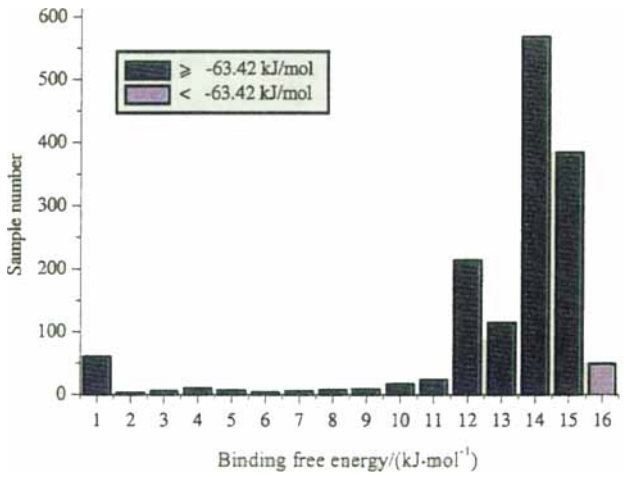


Figure 5 Histogram of virtual screening (1: >462 kJ/mol; 2: <462 and ≥420 kJ/mol; 3: <420 and ≥378 kJ/mol; 4: <378 and ≥336 kJ/mol; 5: <336 and ≥294 kJ/mol; 6: <294 and ≥252 kJ/mol; 7: <252 and ≥210 kJ/mol; 8: <210 and ≥168 kJ/mol; 9: <168 and ≥126 kJ/mol; 10: <126 and ≥84 kJ/mol; 11: <84 and ≥42 kJ/mol; 12: <42 and ≥0 kJ/mol; 13: <0 and ≥−21 kJ/mol; 14: <−21 and ≥−42 kJ/mol; 15: <−42 and ≥−63.42 kJ/mol; 16: <−63.42 kJ/mol).

Among 43 potential lead structures only 6 structures bind with $3CL^{pro}$ receptor at active site. Their binding free energies, log *P*, the size of ring and comments are gathered in Table 1. Because the influence of hydrophobicity is important, log *P*, values of structures **N3** and **N5** approach to that of Ag7088. The ring of those structures is from 18 to 19 membered cycle in size. It suggests that suitable large ring could locate at the active site of receptor. Figure 6 is an example of docking complex for structure **N1**. Structure **N1** locates at the active site of $3CL^{pro}$. The electrostatic potential of **N1** and $3CL^{pro}$ is complementary to each other. This could increase the binding affinity between ligands and the receptor and the interaction is favorable to bioactivity. There are one hydrogen bond between oxygen atom (O=C) of **N1** and the residue Glu165, and two hydrogen bonds between NH and residues Phe139 and His171.

Table 1 Results of virtual screening for in-house database

No.	Binding free energy/(kJ•mol^{-1})	log *P*	Size of ring	Comment[a]
Ag7088	−63.42	3.13		One hydrogen bond
N1	−79.97	1.639	19	+++++
N2	−82.57	0.608	13	++++
N3	−365.78	3.614	18	+++++
N4	−78.04	1.639	19	+++++
N5	−588.80	4.079	18	+++++
N6	−258.59	5.158	19	+++++

[a] ++++ good; +++++ very good.

For **N3**, two oxygen atoms of carbonyl are linked by hydrogen bond to residue Glu165. From investigation on the complex of **N4**-3CL^{pro}, there are four hydrogen bonds between the ligand and residues Glu165, Gln187, Phe139 and His171. There is also one hydrogen bond between **N5** and the residue Gln191, **N6** and the residue His163, respectively. These hydrogen bonds seem to be favourable to the activity.

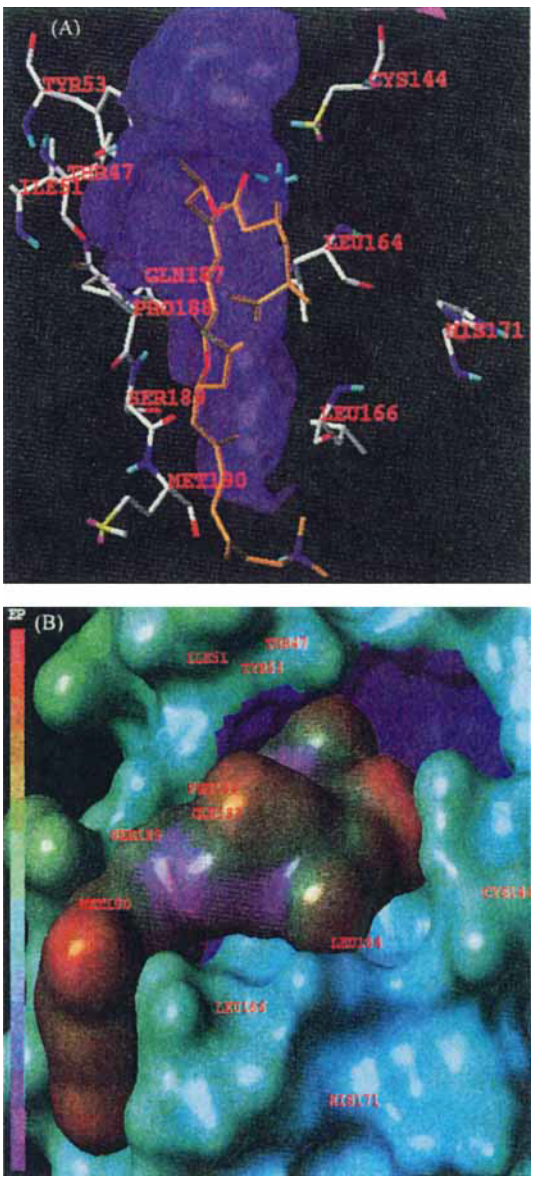


Figure 6 The complex of **N1** and 3CL^{pro}. (A) **N1** in active site; (B) the electrostatic potential surface of **N1** and 3CL^{pro}.

ACD screening database

After the docking of the in-house database into the 3CL^{pro} was finished, the screening for ACD-SC database was then performed. 16000 compounds were selected under the limitation of molecular weight and other conditions. Then docking research was done. The results of the virtual screening are organized in Table 2. Their complexes are shown in Figure 7. They could bind with 3CL^{pro} receptor tightly. Structures **A1**—**A3** include the functional group of sulphone and their molecular volumes are less than those of natural products. Structure **A4** has the antiviral activity and has the same scaffold with Ag7088.[13] The flexibility of an active site might make different sizes of molecule located at this position.

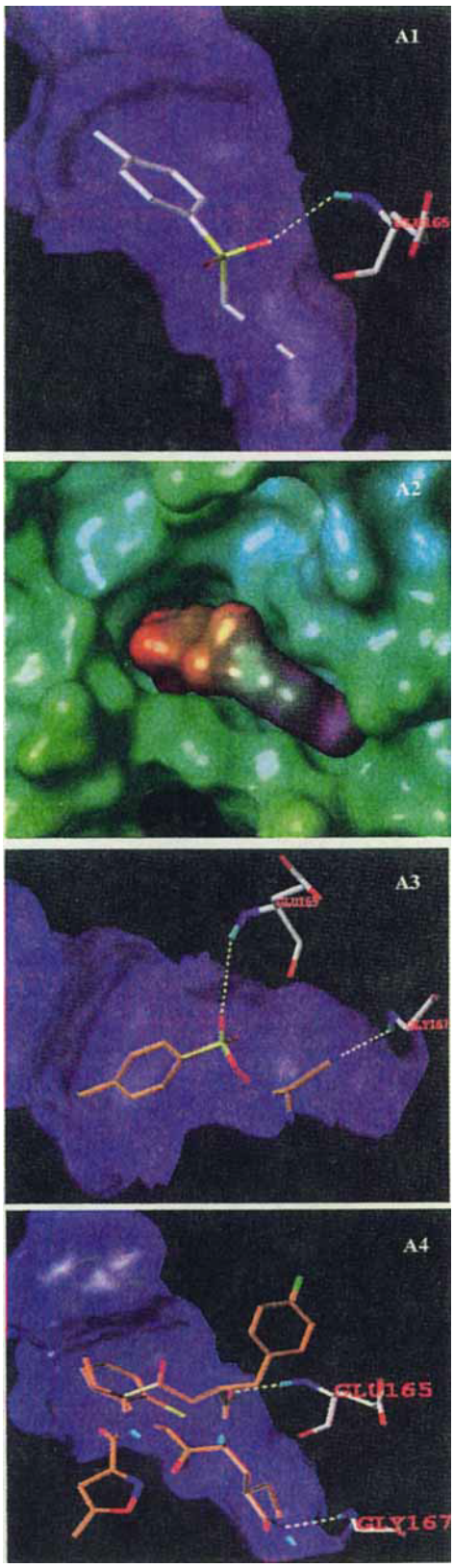
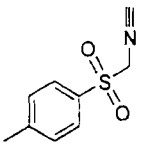
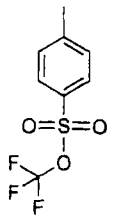
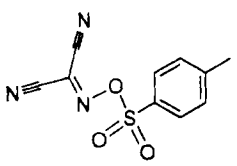
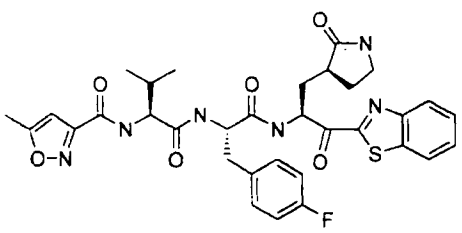


Figure 7 The complex of ligand and 3CL^{pro}.

Table 2 Result of virtual screening for ACD screening database

No.	Structure	Binding free energy/(kJ•mol^{-1})	log *P*	Comment[a]
A1		−86.31	1.52	+++++
A2		−72.79	2.91	+++++
A3		−78.92	3.42	+++++
A4		−63.29	0.10	++++

[a] ++++ good; +++++ very good.

Conclusion

We have presented a novel approach based on the molecular docking and database screening to search for inhibitors of SARS-Cov. The first step is to identify the active site of $3CL^{pro}$ by comparing the crystal structure of human and porcine coronavirus. Since Ag7088 could inhibit the main protein of rhinovirus, it was selected as the screening reference. Before docking research, pre-screening database of compounds was built. Then, these compounds were screened in the putative pocket. Known antiviral inhibitors like **A4** could be screened within the best-scoring list. This shows that our screening strategy is feasible.

The binding free energy between Ag7088 and $3CL^{pro}$ is −63.42 kJ/mol. Several structures from natural products and ACD-SC databases are found to have lower binding free energy than that of Ag7088. These structures have similar hydrophobicity to Ag7088. Their electrostatic potential and hydrogen bond acceptor and donor are complementary with $3CL^{pro}$. These structures are potential lead inhibitors to anti-SARS. The synthesis and bioactivity of these compounds will be done later.

Acknowledgements

The authors thank Professor Arthur J. Olson for his kindness in offering us the AutoDock 3.0.3 program.

References

1 Ruan, Y. J.; Wei, C. L.; Ling, A. E.; Vinsensius, B. V.; Herve, T.; Se, T. S. Y.; Chia, J. M.; Patrick, N.; Chiu, K. P.; Lim, L.; Zhang, T.; Chan, K. P.; Lynette, O.; Lin, E.; Lee, N. M.; Leo, Y. S.; Lisa, F. P. N.; Ren, E. C.; Lawrence, W. S.; Philip, M. L.; Edison, T. L. *Lancet* **2003**, *361*, 1779.

2 Drosten, C.; Gunther, S.; Preiser, W. *N. Engl. J. Med.* **2003**, *348*, 1967.

3 Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S. *N. Engl. J. Med.* **2003**, *348*, 1953.

4 Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. *Science* **2003**, *300*, 1763.

5 Rayer, M. In *Bioinformatics —From Genomes to Drugs*, Vol. I, Ed.: Thomas, L., Wiley-VCH Verlag GmbH, Weinheim, **2002**. pp. 318—319.

6 Xiong, B.; Gui, C. S.; Xu, X. Y.; Luo, C.; Chen, J.; Luo, H. B.; Chen, L. L.; Li, G. W.; Sun, T.; Yu, C. Y.; Yue, L. D.; Duan, W. H.; Shen, J. K.; Qin, L.; Shi, T. L.; Li, Y. X.; Chen, K. X.; Luo, X. M.; Shen, X.; Shen, J. H.; Jiang, H. L. *Acta Pharmacol. Sin.* **2003**, *24*, 497.

7 Heiden, W.; Moeckel, G.; Brickmann, J. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 503.

8 Matthews, D.; Dragovich, P. S.; Webber, S. E.; Fuhrman, S. A.; Patick, A. K.; Zalman, L. S.; Hendrickson, T. F.; Love, R. A.; Prins, T. J.; Marakovits, J. T.; Zhou, R.; Tikhe, J.; Ford, C. E.; Meador, J. W.; Ferre, R. A.; Brown, E. L.; Binford, S. L.; Brothers, M. A.; Delisle, D. M.; Worland, S. T. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11000.

9 Clark, M.; Cramer, III R. D.; Van, O. N. *J. Comput. Chem.*

1989, *10*, 982.

10 Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S. J.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

11 Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.

12 Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293.

13 *MDL Drug Data Report 3D database*, MDL Inc., **2001**.

(E0308221 PAN, B. F.)