Agaritine and Its Derivatives Are Potential Inhibitors against HIV Proteases

Wei-Na Gao[1], Dong-Qing Wei[1,2,4,*], Yun Li[1], Hui Gao[1], Wei-Ren Xu[2,3],Ai-Xiu Li[1] and Kuo-Chen Chou[2,4,*]

[1]*College of Life Science and Technology, Shanghai Jiaotong University, 800 Donglin Road, Minhang District, Shanghai, 200240, China;* [2]*Institute of Bioinformatics and Drug Discovery, Tianjin Normal University, Tianjin, 300074, China;* [3]*Tianjin Institute of Pharmaceutical Research, Tianjin 300019, China;* [4]*Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA*

Abstract: Agaritine, or β-N-[γ-L(+)-glutamyl]-4-hydroxymethylphenylhydrazine, is a Chinese herbal medicine, known having the antiviral and anticancer function. However, so far no reports whatsoever have been made for its potential as an anti-HIV agent. It was observed by docking experiments for more than 9,000 compounds extracted from various Chinese medicines that the compound agaritine distinguished itself from all the others in binding to the HIV protease with the most favorable free energy. Based on this, a series of derivatives were generated by modifying agaritine. It has been observed thru an extensive docking study that some of agaritine derivatives had markedly stronger binding interaction with the HIV protease than agaritine, suggesting that these derivatives might be good candidates for developing drugs for AIDS therapy.

Key Words: HIV Protease, AIDS, agaritine, docking, binding interaction, derivative of chinese herbal medicine, structure-based drug design.

INTRODUCTION

Functioning as a dimer, the HIV (human immunodeficiency virus) protease is made up of two identical subunits, each having 99 residues, but with only one active site [1,2]. The function of HIV protease is essential for the replication of HIV [3,4], the culprit of AIDS (acquired immunodeficiency syndrome). Accordingly, as a complement to the strategy targeting the HIV reverse transcriptase [5-11], the design of HIV protease inhibitors represents a different and promising approach to AIDS therapy [1,2,12-16].

Agaritine, or β-N-[γ-L(+)-glutamyl]-4-hydroxymethylphenylhydrazine, is a Chinese herbal medicine. As a derivative of glutamic acid, agaritine has the antiviral and anticancer function [17,18], but so far no studies have ever been reported for its potential as an anti-HIV agent. In view of this, the present study was initiated in an attempt to investigate the interaction of agaritine with HIV protease thru docking studies. Based on the results thus obtained, a scenario of modifying the functional groups of agaritine was proposed for designing potent inhibitors against HIV protease.

MATERIALS AND METHODS

1. Selection of the Docking Receptors

In the RCSB Protein Data Bank (http://www.pdb.org/), there are 87 HIV proteases deposited. Therefore, it is prerequisite to find the most appropriate proteases for conducting the docking study. This can be realized as follows. First, those proteases which have no effects on the HIV viruses, such as Abacavir, Didanosine, Dideoxyadenosine, Lamivudine, Stavudine, Zidovudine, Emtricitabine, Zalcitabine, Capravirine, Efavirenz, Nevirapine, Aciclovir, Acyclovir, Valacyclovir, Adefovir, Amantadine, Amidinomycin, Stallimycin, Atevirdine, Delavirdine, Cidofovir, Sorivudine, Cytarabine, Desciclovir, Edoxudine, Famciclovir, Floxuridine, Ganciclovir, Valganciclovir, Idoxuridine, Imiquimod, Kethoxal, Methisazone, MADU, Moroxydine, Oseltamivir, Penciclovir, Pleconaril, Podophyllotoxin, Ribavirin, Rimantadine, Taidingan, Tenofovir, Tilorone, Trifluridine, Tromantadine, Valaciclovir, Vidarabine, and Zanamivir, could be excluded for further consideration. For the remaining proteases, some known HIV-1 protease inhibitors, such as Indinavir, Nelfinavir, Lopinavir, Ritonavir, Saquinavir, Amprenavir, were used to winnow for the desired ones according to the criterion, to wit: only those which had good binding with the known HIV protease inhibitors and poor binding with non HIV protease inhibitors were chosen for further studies. It was derived *via* the aforementioned screening procedures that 1D4H, 1FQX, 1HSG, 1KZK and 4PHV were the best candidates. Furthermore, it was found thru sequence alignment among the above five HIV proteases that 1D4H, 1FQX and 1HSG are exactly the same (Fig. **1**). So the scope of the HIV proteases was further reduced to the following three: 1FQX, 1KZK, and 4PHV. For simplification, let us just take 4PHV as a paradigm for demonstration.

2. Design of Agaritine Derivatives

Using 4PHV as HIV protease receptor, docking experiments were conducted for more than 9,000 compounds extracted from various Chinese medicines. It was found that the agaritine distinguished itself from all the other compounds in binding to the receptor with the most favorable mode. Here our efforts were focused on how to modify aga-

*Address correspondence to Dr. Wei at the College of Life Science and Technology, Shanghai Jiaotong University, 800 Donglin Road, Minhang District, Shanghai, 200240, China; E-mail: lifescience@san.rr.com; and Dr. Chou at the Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA; E-mail: kchou@san.rr.com

1573-4064/07 $50.00+.00

(a) The sequence alignment of chain A of the 5 HIV proteases

(b) The sequence alignment of chain B of the 5 HIV proteases

Fig. (1). Sequence alignment of 1D4H, 1FQX, 1HSG, 1KZK and 4PHV for **(a)** chain A and **(b)** chain B, where sequences 1, 2, 3, 4, and 5 refer to 1D4H, 1FQX, 1HSG, 1KZK, and 4PHV, respectively. The amino acids are colored according to their function: acidic-red; basic-blue; neutral hydrophilic-pink; aliphatic-dark green; aromatic-light green; thiol containing-yellow; and imino-orange. The active-site triad is Asp-25, Thr-26, Gly-27. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper).

ritine in order to determine if its derivatives could be even better in inhibiting HIV proteases.

To realize this, let us decompose the structure of agaritine into four regions: A, B, C, and D (Fig. **2**), where some important positions of the compound are marked as 1, 2, 3, and 4, respectively. It is quite clear that region B and D are

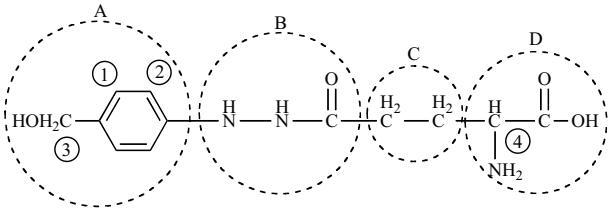


Fig. (2). Schematic illustration to show the structure of agaritine decomposed into four regions: A, B, C, and D. Marked as 1, 2, 3 and 4 are the four important positions of the compound.

hydrophilic, and A, hydrophobic. It was observed that the hydrophilic groups of regions B and D matched very well with the hydrophilic surface of the receptor. However, because some part of the receptor surfaces near region A is hydrophilic, it does not match with the hydrophobic groups of regions A, particularly in the lower left region. To improve their binding interaction, a logic strategy is to modify the agaritine molecule by adding some hydrophilic group to region A. Shown in Table **1** are the 19 chemical groups used to substitute those in the positions 1, 2, 3, and 4. The details of the structure modification (SM) of agaritine are as follows.

SM1

Substituting hydrogen atom at position 1 of region A in turn by each of the 19 groups listed in Table **1** while keeping all the other parts exactly the same.

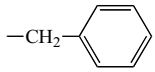
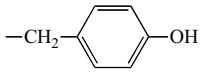
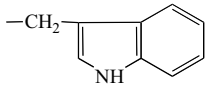
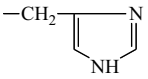
SM2

Substituting hydrogen atom at position 2 of region A in turn by each of the 19 groups listed in Table **1** while keeping all other parts exactly the same.

SM3

Substituting hydrogen atom on the carbon atom of hydroxy methyl at position 3 of region A in turn by each of the 19 groups listed in Table **1** while keeping all the other parts exactly the same.

Table 1. List of Chemical Groups Used for Modifying Agaritine

1	—H	11	$-CH_2-C(=O)-OH$
2	$-CH_3$	12	$-CH_2-CH_2-C(=O)-OH$
3	$-CH(CH_3)-CH_3$	13	$-CH_2-C(=O)-NH_2$
4	$-CH_2-CH(CH_3)-CH_3$	14	$-CH_2-CH_2-C(=O)-NH_2$
5	$-CH(CH_3)-CH_2-CH_3$	15	$-CH_2-SH$
6		16	$-CH_2-CH_2-S-CH_3$
7		17	$-CH_2-CH_2-CH_2-CH_2-NH_2$
8		18	$-CH_2-CH_2-CH_2-NH-C(=NH)-NH_2$
9	$-CH_2-OH$	19	
10	$-CH(OH)-CH_3$		

SM4

The hydrophilic group amido in region D is replaced in turn by each of the 19 groups listed in Table **1** while keeping all the other parts exactly the same.

SM5

The hydrophilic group carboxyl in region D is replaced in turn by each of the 19 groups listed in Table **1** while keeping all the other parts exactly the same.

SM6 – SM10

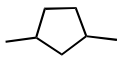
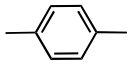
Doing the same operations as described in SM1-SM5 after replacing substructure-1 of Table **2** in region C (see Fig. **2**) by substructure-2 of Table **2**.

SM11 – SM15

Doing the same operations as described in SM1-SM5 after replacing substructure-1 of Table **2** in part C (see Fig. **2**) by substructure-3 of Table **2**.

The molecules thus created were optimized, and their partial charges were assigned by electric equilibrium electric charge method using Sybyl 7.0 in AMBER99 [19]. Then the structures were optimized again using the AMBER99 force field and saved in mol2 format to be used for docking.

Table 2. List of Substructures Mentioned in the Text

Substructure-1	$-CH_2CH_2-$
Substructure-2	
Substructure-3	

3. Docking Studies and Results

The FlexX/Run muti-ligand mode of Sybyl 7.0 software package [20,21] was used to perform docking studies. Several techniques developed for doing the structure-based drug design are incorporated in the package. During the docking process, the base fragment (the ligand core) is automatically selected and is placed into the active site using a new algorithmic approach based on a pattern recognition technique called pose clustering. The remainder of the ligand is built up incrementally from other fragments. The new fragment is added in all possible conformations to all placements found in the previous iteration, but only the *k* best placements are taken on to the next construction step. The conformational flexibility of the ligand is included by generating multiple conformations for each fragment and including all in the

ligand building steps. Placement of the ligand is scored on the basis of protein-ligand interactions, as the shape alone is a weak descriptor, especially for small or flexible ligands. Finally, the binding energies are estimated, and the placements are ranked.

The residues within 8Å around the ligand were taken into account for computing the binding energy. The free energy was evaluated and scored as a measure for binding: the smaller (more negative) the score, the better the binding. Each of the 15x19 = 285 compounds modified from agaritine was docked to the 4PHV protease receptor, and their corresponding docking free energies are given in Table **3**.

As is well known, saquinavir is a powerful protease inhibitor and is also the first HIV protease inhibitor authorized by the FDA [22-24]. It was observed that saquinavir was tethered to Gly27 and Asp29 of 4PHV by three hydrogen bonds. In contrast, agaritine was tethered to the residues around the active site of 4PHV by four hydrogen bonds (Fig. **4**), and the modified agaritine by five hydrogen bonds (Fig. **5**), suggesting that the hydrogen bonding interaction of the HIV protease with the modified agaritine was even stronger. It is instructive to note that the modified structure forms a hydrogen bond with Asp 25, which is one of the most important residues in the active site.

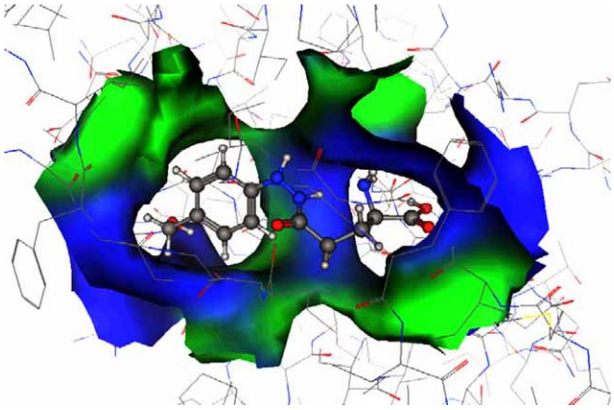


Fig. (3). A close view of the binding interaction of agaritine with the residues of the HIV protease (4PHV). Agaritine is in the ball-stick representation. The receptor surface near the ligand is displayed with the blue and green colors referring to hydrophobic and hydrophilic areas, respectively. It is quite clear that the ligand is well situated in the receptor pocket and that the hydrophilic features of the two molecules are well matched to each other, except a small part of the A region (cf. Fig. **2**). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper).

Table 3. The Docking Free Energy (Kcal/mol) of Agaritine Derivatives after Various Structure Modifications (SM)

CHEMICAL GROUP[A]	SM (STRUCTURAL MODIFICATION) SCHEME														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	——	——	——	-21.7	-33.3	——	——	——	-25.3	-22.3	——	——	——	-26.4	-26.6
2	-30.5	-26.1	-24.8	-24.8	-25	-33.7	-30.3	-31.3	-22.5	-24.2	-36.3	-34.7	-33.7	-29.2	-27.8
3	-26.8	-25	-23.6	-20.2	-22.2	-25.7	-24.3	-25.5	-19.8	-20.8	-37.1	-33.2	-31.6	-30.6	-28.2
4	-28	-25.7	-25.4	-21.5	-22.5	-35.7	-25.8	-30.4	-21.9	-26.3	-37.2	-32.8	-32.2	-27.5	-28.5
5	-26.2	-26	-25.5	-17.9	-19.8	-29.9	-19.4	-26.6	-31	-24	-36.8	-37.6	-36.8	-36	-26.6
6	-27.9	-27.9	-28.6	-27.4	-25.5	-27.9	-27.6	-28.9	-27.6	-21.9	-38.3	-36.1	-33.7	-30.3	-26.1
7	-27	-26.5	-28.9	-26.6	-23.3	-28.3	-24.9	-28.6	-25.6	-24.3	-35.9	-32.9	-29.6	-33.3	-28.5
8	-32	-27.9	-32.1	-29.1	-26.2	-29.3	-39.2	-35	-23.4	-24.6	-33.9	-34.4	-34.8	-31.8	-31.7
9	-29.5	-25.8	-31.2	-24.1	-27	-29.8	-25	-27	-25.5	-28.7	-33	-37.3	-33.5	-28	-28.8
10	-31.2	-27	-28.7	-26.4	-26.7	-32.0	-23.2	-27.9	-21.4	-27.9	-36.1	-35.3	-33.3	-29.2	-35.6
11	-30.3	-29.3	-32	-33.6	-29.1	-34.0	-22.9	-34.8	-31.2	-33.5	-38.1	-36.2	-36.9	-33.5	-37
12	-33.1	-29.3	-30.3	-33.2	-25.2	-32.4	-32.8	-35.3	-26.1	-27.4	-36.5	-38.1	-39	-29.6	-31.3
13	-32.8	-29.2	-41.4	-33.4	-29.3	-32.1	-31.5	-35.4	-34.9	-30.6	-38.3	-30.0	-34.9	-35.8	-38.2
14	-36.3	-30.9	-36.2	-33.9	-29.6	-34.5	-31.3	-36.3	-29.4	-28.5	-36.7	-34.5	-31.7	-32	-33.4
15	-26.8	-28.6	-27.7	-22.1	-21.1	-29.5	-28.7	-27.1	-25.4	-25.7	-35.2	-35.7	-29.6	-27.5	-23.8
16	-24.6	-25.1	-26.3	-20.3	-21.4	-26.4	-24.7	-26.4	-19.7	-28.7	-32.1	-32.2	-28.8	-27.1	-26.4
17	-30.1	-25.2	-30.8	-22	-24.8	-32	-27.2	-28.3	-26.4	-23.3	-33.4	-40.2	-32.6	-28.5	-28.5
18	-39.7	-31.8	-34.9	-29.5	-29.4	-29.9	-32.9	-31.9	-31.2	-28	-34.9	-38.4	-28.1	-33.7	-33.2
19	-29.7	-30.3	-33.7	-31.5	-26.6	-31.3	-27.8	-26.8	-26.1	-24	-32.2	-35.5	-32.8	-27.4	-25.7

[a]See Table **1** for the corresponding chemical groups.

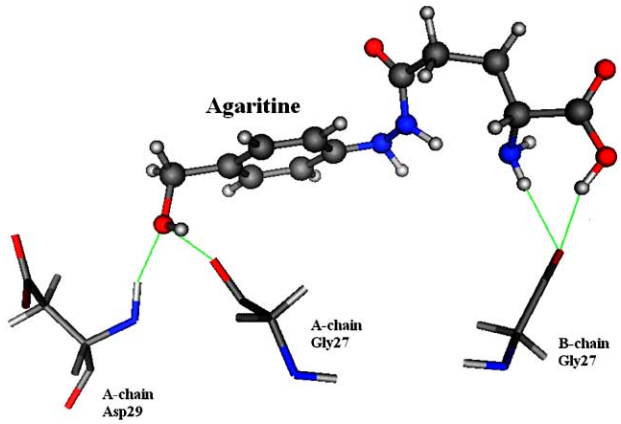


Fig. (4). A close view of the binding interaction between agaritine and the residues of receptor 4PHV. The agaritine is in the ball-stick representation, the residues of the receptor is in the stick representation with oxygen, nitrogen, carbon, and hydrogen colored in red, blue, gray, and white, respectively. The green lines represent the hydrogen bonds which are formed with Asp29 and Gly27 of 4PHV, respectively. Gly27 is the key residue for the enzyme's active site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

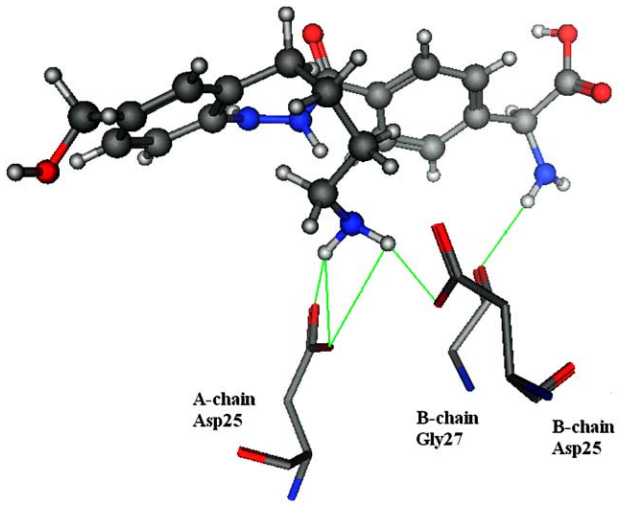


Fig. (5). A close view of the binding interaction between the modified agaritine and the residues of receptor 4PHV. The modified agaritine is in the ball-stick representation, the residues of the receptor is in the stick representation with oxygen, nitrogen, carbon, and hydrogen colored in red, blue, gray, and white, respectively. The green lines represent the hydrogen bonds which are formed with Asp25 and Gly27 of 4PHV, respectively. Both Asp25 and Gly27 are the key residues for the enzyme's active site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper).

It can be seen from Table **3** how the substitution with various chemical groups affect the binding of the modified agaritine to the protease. Let's first compare the average docking results of SM1 and SM2. The average docking free energy over the 19 substituting groups of SM1 and SM2 are -30.14 Kcal/mol and -27.64 Kcal/mol, respectively, suggesting that it is more effective to add a functional group in position 1. This is understandable because the groups in position 1 have more space to adapt to the receptor. It is interesting to see that SM6 and SM10 have much lower average free energy, -35.67 Kcal/mol and -35.28 Kcal/mol, which is achieved by replacing substructure-1 (see Table **2**) of region C (see Fig. **2**) by substructure-2 and substructure-3 (see Table **2**), respectively, suggesting that the introduction of an aromatic group in region C would improve the binding interaction.

The second type of modification scheme (SM3) involves replacing hydrogen atom on the carbon atom of hydroxy methyl in region 1 in turn by each of the 19 groups listed in Table **1**. The average docking free energy was -30.12 Kcal/mol, which was slightly lower than that of SM2 and very close to that of SM1. After the C region was replaced by substructure-2 and substructure-3 of Table **2**, the corresponding average docking free energies were -30.19 Kcal/mol and -32.98 Kcal/mol, indicating that free energy changes were not as much as in the case of SM1 and SM2. This is because replacing hydrogen on carbon by an aromatic group might impose stricter special and geometric requirement for the binding interaction.

The third type of modification scheme is to replace the hydrophilic amido at position 4. The average docking free energy is -26.27 Kcal/mol. And after changing the framework structure for substructures 2 and 3, the corresponding docking free energies were -26.02 Kcal/mol and -30.39 Kcal/mol, respectively. Similar result (-25.68 Kcal/mol) was obtained by replacing carboxyl at position 4. After changing the framework structure for substructures 2 and 3, the corresponding docking free energies were -26.04 Kcal/mol and -29.78 Kcal/mol respectively. It indicates that the hydrophilic groups in position 4 should be kept and introduction of other type groups did not improve the binding. The existence of the benzene ring was favorable for the binding, while substructure-2 did not much help the binding interaction.

We found thru the above analysis that replacing region C with substructure 3 of Table **2** would certainly improve the binding interaction. Introducing the functional groups in regions A could make the ligand more active. The best result with the free docking energy -40.2 Kcal/mol was achieved by following the SM12 scheme; i.e., substituting chemical group 17 of Table **1** for the position 2 of region A and substituting substructure-3 for substructure-1 in region C.

CONCLUSION

Agaritine is a compound extracted from Chinese herbal medicine. It has been observed through an extensive docking experimental study that some of its derivatives have very favorable binding interaction with the HIV protease. The findings reported here might provide useful clues and footings for designing powerful drugs against AIDS.

ACKNOWLEDGEMENTS

This work was supported by the grants from Chinese National Science Foundation under the Contract No. 10376024, Education Foundation under the contract No.

2003001, and the Tianjin Commission of Sciences and Technology under the Contract No. 033801911. Supports were also from the special fund for intensive computation, Virtual Laboratory for Computational Chemistry of CNIC, and Supercomputing Center of CNIC, Chinese Academy of Sciences.

REFERENCES

[1] Wlodawer, A.; Erickson, J. W. *Annu. Rev. Biochem.*, **1993**, *62*, 543-585.
[2] Chou, K. C. *Analytical Biochem.*, **1996**, *233*, 1-14.
[3] Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 4686-4690.
[4] Seelmeier, S.; Schmidt, H.; Turk, V.; von der Helm, K. *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 6612-6616.
[5] Althaus, I. W.; Chou, J. J.; Gonzales, A. J.; Diebel, M. R.; Chou, K. C.; Kezdy, F. J.; Romero, D. L.; Aristoff, P. A.; Tarpley, W. G.; Reusser, F. *J. Biol. Chem.*, **1993**, *268*, 6119-6124.
[6] Althaus, I. W.; Gonzales, A. J.; Chou, J. J.; Diebel, M. R.; Chou, K. C.; Kezdy, F. J.; Romero, D. L.; Aristoff, P. A.; Tarpley, W. G.; Reusser, F. *J. Biol. Chem.*, **1993**, *268*, 14875-14880.
[7] Althaus, I. W.; Chou, J. J.; Gonzales, A. J.; Diebel, M. R.; Chou, K. C.; Kezdy, F. J.; Romero, D. L.; Aristoff, P. A.; Tarpley, W. G.; Reusser, F. *Biochemistry*, **1993**, *32*, 6548-6554.
[8] Althaus, I. W.; Chou, J. J.; Gonzales, A. J.; Diebel, M. R.; Chou, K. C.; Kezdy, F. J.; Romero, D. L.; Thomas, R. C.; Aristoff, P. A.; Tarpley, W. G.; Reusser, F. *Biochem. Pharmacol.*, **1994**, *47*, 2017-2028.
[9] Althaus, I. W.; Chou, K. C.; Franks, K. M.; Diebel, M. R.; Kezdy, F. J.; Romero, D. L.; Thomas, R. C.; Aristoff, P. A.; Tarpley, W. G.; Reusser, F. *Biochem. Pharmacol.*, **1996**, *51*, 743-750.

Received: 16 August, 2006 Revised: 14 September, 2006 Accepted: 15 September, 2006

[10] Chou, K. C.; Kezdy, F. J.; Reusser, F. *Analytical Biochem.*, **1994**, *221*, 217-230.
[11] Althaus, I. W.; Chou, J. J.; Gonzales, A. J.; Diebel, M. R.; Chou, K. C.; Kezdy, F. J.; Romero, D. L.; Aristoff, P. A.; Tarpley, W. G.; Reusser, F. *Experientia*, **1994**, *50*, 23-28.
[12] Debouck, C. *AIDS Res. Hum. Retroviruses*, **1992**, *8*, 153-164.
[13] Chou, K. C. *J. Biol. Chem.*, **1993**, *268*, 16938-16948.
[14] Chou, J. J. *Biopolymers*, **1993**, *33*, 1405-1414.
[15] Chou, J. J. *J. Prot. Chem.*, **1993**, *12*, 291-302.
[16] Thompson, T. B.; Chou, K. C.; Zheng, C. *J. Theoretical Biol. 177*, **1995**, 369-379.
[17] Espin, J. C.; Jolivet, S.; Overeem, A.; Wichers, H. J. *Phytochemistry*, *50*, 555-563.
[18] Baumgartner, D.; Lienhard, H.; Rast, D.M. *Phytochemistry*, **1998**, *49*, 465-474.
[19] Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.*, **2005**, *26*, 1668-1688.
[20] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.*, **1996**, *261*, 470-489.
[21] Klebe, G.; Mietzner, T. *J. Comput. Aided Mol. Des.*, **1994**, *8*, 583-606.
[22] Boffito, M.; Kurowski, M.; Kruse, G.; Hill, A.; Benzie, A. A.; Nelson, M. R.; Moyle, G. J.; Gazzard, B. G.; Pozniak, A. L. *AIDS*, **2004**, *18*, 1291-1297.
[23] Cameron, D. W.; Japour, A. J.; Xu, Y.; Hsu, A.; Mellors, J.; Farthing, C.; Cohen, C.; Poretz, D.; Markowitz, M.; Follansbee, S.; Angel, J. B.; McMahon, D.; Ho, D.; Devanarayan, V.; Rode, R.; Salgo, M.; Kempf, D. J.; Granneman, R.; Leonard, J. M.; Sun, E. *AIDS*, **1999**, *13*, 213-224.
[24] Gray, A.; Karim, S. S.; Gengiah, T. N. *AIDS*, **2006**, *20*, 302-303.