

DOI: 10.1002/cbic.200700254

Structure-Based Design and Synthesis of Highly Potent SARS-CoV 3CL Protease Inhibitors

Yi-Ming Shao,^[a, b, c] Wen-Bin Yang,^[a] Hung-Pin Peng,^[a] Min-Feng Hsu,^[b] Keng-Chang Tsai,^[a] Tun-Hsun Kuo,^[b] Andrew H.-J. Wang,^[a, b, c] Po-Huang Liang,^[a, b, c] Chun-Hung Lin,^[a, b, c] An-Suei Yang,^{*[a]} and Chi-Huey Wong^{*[a, b, c, d]}

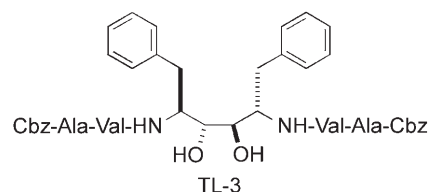
Severe acute respiratory syndrome (SARS), a life-threatening respiratory disease, was first reported in Southern China in November 2002, and in 2003 it spread widely to other Asian countries, North America, and Europe. According to the World Health Organization (WHO), a total of 8098 people worldwide became sick with SARS during the 2003 outbreak, and 774 of the infected patients died. A novel coronavirus associated with cases of SARS (SARS-CoV) was identified as the etiological agent of this endemic atypical pneumonia.^[1] SARS-CoV is a single-stranded positive-strand RNA virus, and its genome structure comprises both replicase and structural regions.^[2] The polyproteins, pp1a (486 kDa) and pp1ab (790 kDa), encoded by the viral replicase gene, are subject to extensive proteolytic processing by viral proteases to produce multiple functional subunits, which are responsible for the formation of the replicase complex. The SARS-CoV 3CL protease, named after the 3C protease of the *Picornaviridae*, is a ~33 kDa cysteine protease that cleaves the replicase polyprotein at 11 conserved sites with canonical Leu-Gln↓(Ser, Ala, Gly) sequences^[3–5] Because of the functional importance of SARS-CoV 3CL protease in the viral life cycle, together with successes in developing efficacious antiviral agents targeting 3C-like proteases in other viruses,^[6] this enzyme has been recognized as a prime target for therapeutic intervention against SARS-CoV infection.

In its X-ray crystal structures the 3CL protease forms a dimer with two protomers, each of which composed of three domains. The active site contains a catalytic dyad (Cys145 and

His41), and the substrate-binding subsite S1 of the enzyme has absolute specificity for Gln-P1 of the substrate.^[7] To date, a large number of inhibitors of 3CL protease have been studied, including molecules identified from high-throughput screening,^[8] electrophilic analogues,^[9] isatin derivatives,^[10] peptidomimetic α,β -unsaturated esters,^[11] peptidic anilides,^[12] and benzotriazole esters.^[13] However, these molecules lack further structural studies to provide in-depth understanding of molecular interactions of the enzyme–inhibitor complex, and/or for structure-based optimization.

There are four reports that describe the 3CL protease structures in complexation with inhibitors.^[7,14–16] In all these complex structures, the ligands are irreversible inhibitors; that is, they are covalently bound to the target protein. Although the S γ atom of Cys145 at the enzyme active site displays bond formation variously with the methylene group of the chloromethyl ketone (CMK),^[7] the C β atoms of Michael acceptors,^[14,15] or the C3 atom of an aza-peptide epoxide (APE),^[16] these mechanism-based inhibitors do not demonstrate satisfactory potency (IC₅₀ = 2 mM for CMK; IC₅₀ = 70 μ M, K_i = 10.7 μ M for Michael acceptors; K_i = 18 μ M for APE) against the 3CL protease.

In general, reversible inhibitors produce fewer side effects than suicide inhibitors and are thus more suitable for therapeutic development.^[17] TL-3, a noncovalent HIV protease inhib-



itor (K_i = 1.5 nM) previously developed in our laboratory,^[18] was found to be an inhibitor of the 3CL protease with a K_i value of 0.6 μ M.^[19] Previous studies have shown that TL-3 is effective against FIV protease and many drug-resistant HIV proteases, has a strong ability to control lentiviral infections in tissue culture, and exerts no adverse effects in ICR mice up to the dose level of 2000 mg kg⁻¹ by gavage during the 14 day study period. The compound is also negative in mouse peripheral blood micronucleus assay. With these considerations taken into account, TL-3 was selected as a lead compound for further optimization in the search for higher inhibition potency.

We initially incorporated a series of L-amino acids in place of the Val–Ala residues of TL-3. However, none of these enhanced the inhibitory activity (see the Supporting Information). The negative results led us to suspect that the main binding mode of TL-3 was energetically dominated by the two phenyl groups.

Optimization of TL-3 as an inhibitor against the 3CL protease by replacement of the peripheral Val–Ala residues or the two central phenyl groups was based on the rationale that the binding mode of TL-3 in the protein–ligand complex mainly involves at least a dipeptide scaffold. We thus carried out computational modeling methods to explore all 20 \times 20 dipeptides

[a] Y.-M. Shao, Dr. W.-B. Yang, H.-P. Peng, Dr. K.-C. Tsai, Dr. A. H.-J. Wang, Dr. P.-H. Liang, Dr. C.-H. Lin, Prof. A.-S. Yang, Prof. C.-H. Wong
Genomics Research Center, Academia Sinica
No. 128, Sec. 2, Academia Rd., Taipei City 115, Taiwan (R.O.C.)
Fax: (+886)2-2789-8771
E-mail: chwong@gate.sinica.edu.tw

[b] Y.-M. Shao, Dr. M.-F. Hsu, T.-H. Kuo, Dr. A. H.-J. Wang, Dr. P.-H. Liang, Dr. C.-H. Lin, Prof. C.-H. Wong
Institute of Biological Chemistry, Academia Sinica
No. 128, Sec. 2, Academia Rd., Taipei City 115, Taiwan (R.O.C.)

[c] Y.-M. Shao, Dr. A. H.-J. Wang, Dr. P.-H. Liang, Dr. C.-H. Lin, Prof. C.-H. Wong
Institute of Biochemical Sciences, National Taiwan University
No. 1, Sec. 4, Roosevelt Rd., Taipei City 106, Taiwan (R.O.C.)

[d] Prof. C.-H. Wong
Department of Chemistry, The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: wong@scripps.edu

Supporting information for this article is available on the WWW under <http://www.chembiochem.org> or from the author.

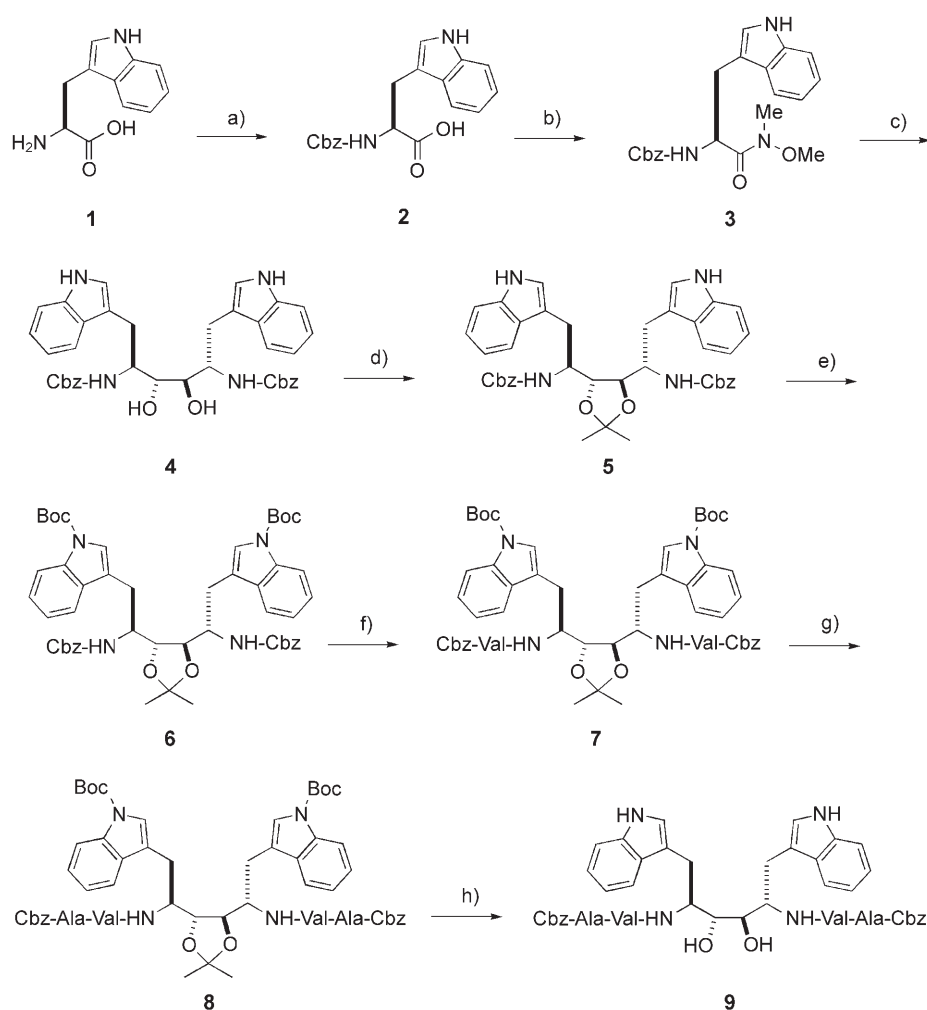
as model ligands for possible protein–ligand interactions. An exhaustive exploration of the dipeptide binding modes using automated computational modeling procedures revealed that the Trp–Trp dipeptide emerged as the strongest binder on the basis of the ranking system used. Moreover, 63 out of the top-ranked 100 protein–ligand complexes contained at least one Trp residue (for details see the Supporting Information). We thus synthesized two compounds to test the binding mode hypothesis: one with two Trp groups adjacent to the central diol (**4**, Scheme 1) and the other with two additional Val–Ala residues as in **9**.

The synthesis started with L-Trp as shown in Scheme 1. Cbz-L-tryptophan (**2**) was converted into its Weinreb amide (**3**), and this was followed by reduction to give the protected L-tryptophanal. The aldehyde was then subjected to stereoselective pi-

naccol homocoupling according to the reported method^[20] to yield diol **4**. Successive protections of the diol with isopropylidene and of the indole nitrogens with *tert*-butoxycarbonyl (Boc) provided **6** in 53% overall yield. Removal of the terminal Cbz groups by hydrogenolysis and subsequent coupling with Cbz-Val-OH produced adduct **7** in 78% yield. Compound **8**, containing additional Ala residues, was prepared by a similar process in 98% yield. Final deprotection under acidic conditions afforded the desired diol **9** in 43% yield. Compounds **4** and **9** were tested against the 3CL protease and proved to be more effective competitive inhibitors of the 3CL protease than TL-3, with K_i values of 0.34 and 0.073 μM , respectively. It is noted that compound **9** is highly selective for the 3CL protease and that no inhibition was observed against HIV protease at 100 μM .

In order to understand the molecular interactions of the enzyme–inhibitor complex in detail, we carried out an X-ray crystallography analysis of the SARS-CoV 3CL protease complex containing compound **4**. The preliminary results revealed that a conformational change, presumably induced by the presence of the ligand, was observable at the loop region near the active site (Figure 1A). Differential electron density maps in the presence of compound **4** indicated that the ligand was partially visible, but the molecular structure of the ligand could not be precisely defined.

However, computational structures of compound **4** with the 3CL protease were constructed by use of the preliminary X-ray protein structure obtained in the presence of compound **4**. This structure was different from the 3CL protease structure used in the previous modeling task in that the previous structure was derived with chloromethyl ketone complexed with the protease. We expected that this round of computational modeling would produce a more realistic protein–ligand complex model. The predicted model ligand structure would then be superimposed onto the differential electron density maps to confirm the validity of the binding mode of the ligand on the target protein. As de-



Scheme 1. Synthesis of compound **4** and **9**. Reagents and conditions: a) CbzCl, Na₂CO₃, NaHCO₃, MeCN/H₂O (2:3), 0 → 23 °C, 2 h, 90%; b) HCl-NHMe(OMe), Et₃N, HOBt, WSC, DMF, 0 → 23 °C, 15 h, 86%; c) i: DIBAL-H, PhMe, THF, -78 °C, 15 min; ii: VCl₃(thf)₃, Zn, CH₂Cl₂, 23 °C, 17 h, 34%; d) cat. *p*-TsOH-H₂O, (MeO)₂CMe₂, CHCl₃, 23 °C, 2 h, 79%; e) Boc₂O, 4-DMAP (cat.), CH₃CN, 23 °C, 26 h, 67%; f) i: H₂ (1 atm), cat. Pd/C (10%), THF/MeOH (1:1); ii: Cbz-Val-OH, HBTU, DIPEA, DMF, 23 °C, 13.5 h, 78%; g) i: H₂ (1 atm), cat. Pd/C (10%), THF/MeOH (1:1); ii: Cbz-Ala-OH, HBTU, DIPEA, DMF, 23 °C, 14.5 h, 98%; h) 4 N HCl/dioxane/H₂O, 23 °C, 14 h, 43%. Cbz = benzyloxycarbonyl; HOBt = *N*-hydroxybenzotriazole; WSC = 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride; DMF = *N,N*-dimethylformamide; DIBAL-H = diisobutylaluminum hydride; THF = tetrahydrofuran; Ts = *p*-toluenesulfonyl; Boc = *tert*-butoxycarbonyl; 4-DMAP = 4-(dimethylamino)pyridine; HBTU = (1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; DIPEA = diisopropylethylamine.

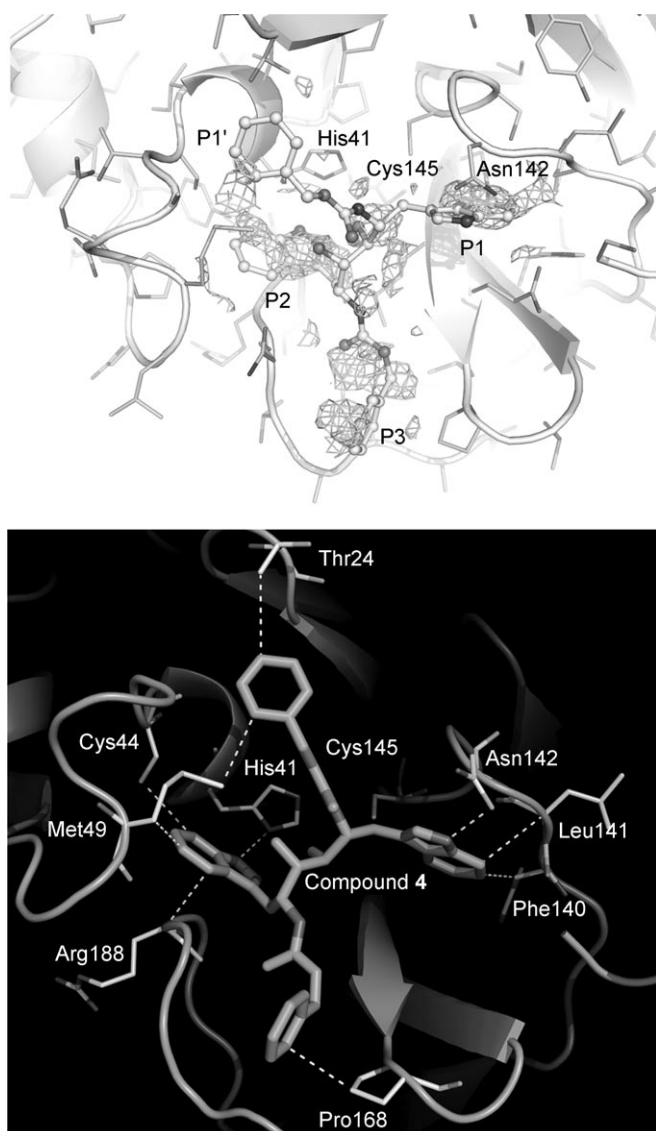


Figure 1. A) X-ray differential density maps and the modeled conformation of ligand **4**. B) Compound **4** in the active site of SARS-CoV 3CLpro, generated by computer modeling. The secondary structure elements are shown as a ribbon drawing, and the important residues involved in inhibitor binding are labeled. The dotted lines represent H-bonding and hydrophobic (from Asn142 and His41) interactions of compound **4** with SARS-CoV 3CLpro.

scribed in the Supporting Information, the prediction of the model binding structure of compound **4** did not use any information from the differential electron density map; the purpose was to use the differential electron density map as an independent test to assess the validity of the predicted binding mode.

Visual examination of the top-ranked protein–ligand complex structures generated from all three docking software packages revealed a consensus complex structure, suggesting a tight packing of the two tryptophan sidechains into the S1 and S2 binding pockets, with the two Cbz moieties occupying the neighboring S3 and S1' sites. The consensus was particularly strong with regard to the tryptophan sidechain conformation, for which precise hydrogen-binding configurations satisfy-

ing hydrogen-bonding donors in the tryptophan side chains were conserved throughout the top-ranked structures, while the hydrophobic packing between the tryptophan side chains and the ligand-binding sites was tightly matched. The induced hydrogen-bonded local conformation involving side chain Asn142 fitted perfectly to accommodate the tryptophan side-chain as part of the S1 binding site in the model structure, while His41 formed another hydrogen-bonding pattern with another indole group at the S2 site (Figure 1 B). The binding conformations of the Cbz moieties, however, were less certain from the consensus complex structures. The uncertainty in modeling Cbz moieties in the ligand–protein complex was the result of a lack of favorable interactions between the Cbz moieties and substrate binding site in the protein. This uncertainty was also manifested in the incomplete X-ray crystallographic data for the Cbz groups. The paucity of favorable Cbz–protein interactions in the complex model suggested the possibility for further improvement in the binding affinity and specificity through replacement of the Cbz moieties with other functional groups.

The consensus complex structure was compared with the differential density map, which shows the superimposition of the differential electron density map and the modeled binding mode of compound **4**. The comparison indicated that all the visible fragments in the differential electron density maps were accounted for by the model ligand conformation in the active site, especially the positions of the indole groups. The consistency of the consensus model structures and the differential electron density map supported the validity of the model complex structure.

The ligand–protein complex shown in Figure 1 B was different from the top-ranked Trp–Trp binding mode predicted from the preliminary modeling task (for differences see Figure S2). In retrospect, although the discrepancies could arise from the differences in the target protein structures and the ligand molecule structures used in the modeling tasks, the partial agreement between the two models (Figure S2) nevertheless suggested that the preliminary modeling effort provided a productive direction for improvement of the ligand–protein binding affinity. The subsequent refinement of the complex structure significantly improved the accuracy of the binding model, and provided a working model for further optimization of the lead compounds.

In summary, we have used computational technology to optimize a lead inhibitor (TL-3) of the 3CL protease. The C_2 -symmetric diol **9** was then designed and synthesized, and showed higher affinity than the previous lead compound (TL-3), by one order of magnitude in its inhibition constant ($0.6 \rightarrow 0.073 \mu\text{M}$). Compounds **4** and **9** represent the most potent noncovalent 3CL protease inhibitors with structural analyses reported to date. This was a successful example of lead optimization by computer modeling prediction, and both potency and selectivity (over HIV protease) were achieved. Though the electron-density map of the complex has not been well defined, the preliminary X-ray structure can be used to support the computer modeling analysis of diol **4** in complexation with the 3CL protease, which indicated that both NHs of the indole rings

formed hydrogen bonds with the side chains of Asn142 and His41. We believe that this approach has provided a platform for further lead optimization.

Acknowledgements

This work is supported by the National Science Council, Taiwan and the Genomics Research Center, Academia Sinica. The SYBYL computation was conducted at the National Center for High Performance Computing, Taiwan. The DISCOVERY STUDIO 1.2 SBD computation was conducted at the computational center of the Academia Sinica.

Keywords: computer modeling · inhibitors · medicinal chemistry · proteases · structure-based design

- [1] C. Drosten, S. Günther, W. Preiser, S. van der Werf, H.-R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R. A. M. Fouchier, A. Berger, A.-M. Burguière, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J.-C. Manuguerra, S. Müller, V. Rickerts, M. Stürmer, S. Vieth, H.-D. Klenk, A. Osterhaus, H. Schmitz, H. W. Doerr, *N. Engl. J. Med.* **2003**, *348*, 1967–1976.
- [2] K. Stadler, M. Vega, M. Eickmann, S. Becker, S. Abrignani, H.-D. Klenk, R. Rappuoli, *Nat. Rev. Microbiol.* **2003**, *1*, 209–218.
- [3] V. Thiel, K. A. Ivanov, Á. Putics, T. Hertzog, B. Schelle, S. Bayer, B. Weißbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, J. Ziebuhr, *J. Gen. Virol.* **2003**, *84*, 2305–2315.
- [4] M. Bartlam, H. Yang, Z. Rao, *Curr. Opin. Struct. Biol.* **2005**, *15*, 664–672.
- [5] J. Ziebuhr, *Curr. Top. Microbiol. Immunol.* **2005**, *287*, 57–94.
- [6] P. S. Dragovich, T. J. Prins, R. Zhou, T. O. Johnson, Y. Hua, H. T. Luu, S. K. Sakata, E. L. Brown, F. C. Maldonado, T. Tuntland, C. A. Lee, S. A. Fuhrman, L. S. Zalman, A. K. Patick, D. A. Matthews, E. Y. Wu, M. Guo, B. C. Borer, N. K. Nayyar, T. Moran, L. Chen, P. A. Rejto, P. W. Rose, M. C. Guzman, E. Z. Doval Santos, S. Lee, K. McGee, M. Mohajeri, A. Liese, J. Tao, M. B. Kosa, B. Liu, M. R. Batugo, J.-P. R. Gleeson, Z. P. Wu, J. Liu, J. W. Meador, R. A. Ferre, *J. Med. Chem.* **2003**, *46*, 4572–4585.
- [7] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13190–13195.
- [8] R. Y. Kao, W. H. W. Tsui, T. S. W. Lee, J. A. Tanner, R. M. Watt, J.-D. Huang, L. Hu, G. Chen, Z. Chen, L. Zhang, T. He, K.-H. Chan, H. Tse, A. P. C. To, L. W. Y. Ng, B. C. W. Wong, H.-W. Tsoi, D. Yang, D. D. Ho, K.-Y. Yuen, *Chem. Biol.* **2004**, *11*, 1293–1299.
- [9] a) R. P. Jain, H. I. Pettersson, J. Zhang, K. D. Aull, P. D. Fortin, C. Huitema, L. D. Eltis, J. C. Parrish, M. N. G. James, D. S. Wishart, J. C. Vederas, *J. Med. Chem.* **2004**, *47*, 6113–6116; b) S. I. Al-Gharabli, S. T. Ali Shah, S. Weik, M. F. Schmidt, J. R. Mesters, D. Kuhn, G. Klebe, R. Hilgenfeld, J. Rademann, *ChemBioChem* **2006**, *7*, 1048–1055.
- [10] L.-R. Chen, Y.-C. Wang, Y.-W. Lin, S.-Y. Chou, S.-F. Chen, L.-T. Liu, Y.-T. Wu, C.-J. Kuo, T. S.-S. Chen, S.-H. Juang, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3058–3062.
- [11] J.-J. Shie, J.-M. Fang, T.-H. Kuo, C.-J. Kuo, P.-H. Liang, H.-J. Huang, Y.-T. Wu, J.-T. Jan, E. Y.-S. Cheng, C.-H. Wong, *Bioorg. Med. Chem.* **2005**, *13*, 5240–5252.
- [12] J.-J. Shie, J.-M. Fang, C.-J. Kuo, T.-H. Kuo, P.-H. Liang, H.-J. Huang, W.-B. Yang, C.-H. Lin, J.-L. Chen, Y.-T. Wu, C.-H. Wong, *J. Med. Chem.* **2005**, *48*, 4469–4473.
- [13] C.-Y. Wu, K.-Y. King, C.-J. Kuo, J.-M. Fang, Y.-T. Wu, M.-Y. Ho, C.-L. Liao, J.-J. Shie, P.-H. Liang, C.-H. Wong, *Chem. Biol.* **2006**, *13*, 261–268.
- [14] a) A. K. Ghosh, K. Xi, K. Ratia, B. D. Santarsiero, W. Fu, B. H. Harcourt, P. A. Rota, S. C. Baker, M. E. Johnson, A. D. Mesecar, *J. Med. Chem.* **2005**, *48*, 6767–6771; b) A. K. Ghosh, K. Xi, M. E. Johnson, S. C. Baker, A. D. Mesecar, *Annu. Rep. Med. Chem.* **2006**, *41*, 183–196.
- [15] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K. Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, *PLoS Biol.* **2005**, *3*, e324.
- [16] T. W. Lee, M. M. Cherney, C. Huitema, J. Liu, K. E. James, J. C. Powers, L. D. Eltis, M. N. G. James, *J. Mol. Biol.* **2005**, *353*, 1137–1151.
- [17] T. P. Smyth, *Bioorg. Med. Chem.* **2004**, *12*, 4081–4088.
- [18] T. Lee, G. S. Laco, B. E. Torbett, H. S. Fox, D. L. Lerner, J. H. Elder, C.-H. Wong, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 939–944.
- [19] C.-Y. Wu, J.-T. Jan, S.-H. Ma, C.-J. Kuo, H.-F. Juan, E. Y.-S. Cheng, H.-H. Hsu, H.-C. Huang, D. Wu, A. Brik, F.-S. Liang, R.-S. Liu, J.-M. Fang, S.-T. Chen, P.-H. Liang, C.-H. Wong, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10012–10017.
- [20] A. W. Konradi, S. F. Pedersen, *J. Org. Chem.* **1992**, *57*, 28–32.

Received: May 10, 2007

Published online on August 23, 2007