ELSEVIER

Screening of electrophilic compounds yields an aziridinyl peptide as new active-site directed SARS-CoV main protease inhibitor

Erika Martina,[a] Nikolaus Stiefl,[a] Björn Degel,[a] Franziska Schulz,[a] Alexander Breuning,[a] Markus Schiller,[a] Radim Vicik,[a] Knut Baumann,[a] John Ziebuhr[b] and Tanja Schirmeister[a],*

[a]*Institute of Pharmacy and Food Chemistry, Am Hubland, D-97074 Würzburg, University of Würzburg, Germany*
[b]*Institute of Virology and Immunology, Versbacher Street 7, D-97078 Würzburg, University of Würzburg, Germany*

Abstract—The coronavirus main protease, M^{pro}, is considered a major target for drugs suitable to combat coronavirus infections including the severe acute respiratory syndrome (SARS). In this study, comprehensive HPLC- and FRET-substrate-based screenings of various electrophilic compounds were performed to identify potential M^{pro} inhibitors. The data revealed that the coronaviral main protease is inhibited by aziridine- and oxirane-2-carboxylates. Among the *trans*-configured aziridine-2,3-dicarboxylates the Gly-Gly-containing peptide **2c** was found to be the most potent inhibitor.
© 2005 Elsevier Ltd. All rights reserved.

Coronaviruses are important pathogens that mainly cause respiratory and enteric disease in humans, livestock, and domestic animals.[1] In 2003, a previously unknown coronavirus called SARS-CoV was identified as the causative agent of the severe acute respiratory syndrome (SARS), a newly emerging disease that within a few weeks spread from its likely origin in Guangdong Province, China, to neighboring regions and many other countries.[2–4] Coronaviruses are plus-strand RNA viruses that use a complex enzymology to replicate the largest RNA genomes currently known and synthesize an extensive set of 5′ leader-containing subgenomic mRNAs that encode the viral structural proteins and several species-specific proteins with unknown functions.[1,5,6] The enzymatic activities required for viral RNA synthesis are part of two virus-encoded polyproteins of about 450 and 750 kDa, respectively, that are extensively processed by two or three viral proteases to yield up to 16 mature proteins and multiple processing intermediates.[7]

Most of the cleavages are mediated by the coronavirus main protease, M^{pro}, a cysteine protease featuring a two-β-barrel structure (domains I and II) that is linked to a C-terminal α-helical domain III. The structure of domains I and II is similar to that of chymotrypsin-like serine proteases.[7–10] Because of its essential role in proteolytic processing, the M^{pro} is considered an attractive target for antiviral drugs against SARS and other coronavirus infections.[9]

Up to now, a number of potential inhibitors have been proposed employing molecular modeling and virtual screening techniques.[11–19]

However, the inhibitory potency of these compounds has not yet been verified.

Only a small number of potent protease inhibitors were identified by screening assays thus far.[20–24] Most of these studies used commercially available compound libraries for their screening assays. With the exception of a peptidyl chloromethylketone[10] and recently published etacrynic acid derivatives[25] none of the potent compounds identified up to date was developed and is predicted to target the active site cysteine residue. This lack of active-site directed lead structures motivated the search for new leads with proven active-site directed activity.

The scrutinized compounds contain electrophilic building blocks (aziridine,[26,27] epoxide,[26,28–30] see Table 1) which are known to react with nucleophilic amino acids within the active site of proteases. For example, *trans*-configured epoxysuccinyl-based peptides-like E-64,[30]

Table 1. Results of the screening of various protease inhibitors against SARS-CoV M^{pro}

Compound	Configuration of the TMR	R^1	R^2	X	R^3	Inhibition of SARS-CoV M^{pro} at 100 μM
1a	*rac-cis (S,S + R,R)*	Phenyl	OMe	Bn-N	H	ni
1b	*rac-cis (S,S + R,R)*	Me	OEt	Bn-N	H	ni
1c (1/1)^c	*rac-cis (S,S + R,R)*	Me	(*S*)-Phe-OBn	Bn-N	H	34 ± 7[a]
1d	*cis (S,S)*	Me	(*S*)-Phe-OBn	Bn-N	H	24 ± 2[a]
1e (1.1/1)^c	*rac-cis (S,S + R,R)*	Phenyl	(*S*)-Phe-OBn	Bn-N	H	30 ± 9[a]
1f (1.4/1)^c	*rac-cis (S,S + R,R)*	Phenyl	(*S*)-Val-OBn	Bn-N	H	22 ± 2[a]
2a	*cis (R,S)*	MeO$_2$C	OMe	HN	H	ni
2b	*cis (R,S)*	EtO$_2$C	OEt	HN	H	ni
2c	*trans (S,S)*	EtO$_2$C	Gly-Gly-OBn	HN	H	54 ± 5[a]/75 ± 7[b] (see also Ref. 37)
2d	*trans (S,S)*	EtO$_2$C	(*S*)-Ile-OBn	HN	H	ni
3a	*trans (S,S)*	BnO$_2$C	OBn	BOC-Gly-(*S*)-Pip-N	H	29 ± 4[a]
3b	*trans (R,R)*	EtO$_2$C	OEt	BOC-(*S*)-Leu-(*R + S*)-Azet-N	H	16 ± 4[a]
3c	*trans (R,R)*	EtO$_2$C	OEt	BOC-(*S*)-Leu-(*S*)-Pro-N	H	16 ± 6[a]
4a	(*R + S*)	H	OMe	MeO$_2$C-CH$_2$-N	H	30 ± 6[a]
4b	(*R + S*)	H	OMe	BnO$_2$C-CH$_2$-N	H	26 ± 4[a]
4c (1.1/1)^c	(*R + S*)	H	OMe	(structure)	H	30 ± 3[a]
4d (1.5/1)^c	(*R + S*)	H	OMe	(structure)	H	20 ± 6[a]
4e (1.1/1)^c	(*R + S*)	H	OMe	(structure)	H	28 ± 7[a]
4f (1.2/1)^c	(*R + S*)	H	OMe	(structure)	H	34 ± 5[a]
4g (2.3/1)^c	(*R + S*)	H	OMe	(structure)	H	30 ± 5[a]
4h (1.2/1)^c	(*R + S*)	H	OMe	(structure)	H	39 ± 6[a]
4i (1.2/1)^c	(*R + S*)	H	OMe	(structure)	H	28 ± 3[a]
4k (1.1/1)^c	(*R + S*)	H	(*S*)-Phe-OMe	MeO$_2$C-CH$_2$-N	H	ni
4l (1.3/1)^c	(*R + S*)	H	OMe	(structure)	H	48 ± 6[a]
5a	(*R + S*)	Me	OMe	BOC-N	CO$_2$Me	ni
5b	(*R + S*)	Me	OMe	EOC-N	CO$_2$Me	ni
5c	(*R + S*)	Me	OMe	HN	CO$_2$Me	ni
6a	*cis (R,R)*	Me	OH	O	H	ni
6b	*cis (R,S)*	MeO$_2$C	OMe	O	H	ni
6c	*cis (R,S)*	EtO$_2$C		O	H	ni
6d	*cis (R,R)*	Me	(*S*)-Phe-OBn	O	H	15 ± 5[a]
6e	*cis (R,R)*	Me	(*R*)-Phe-OBn	O	H	ni
6f	*cis (S,S)*	Me	(*R*)-Phe-OBn	O	H	ni
6g	*cis (S,S)*	Me	(*S*)-Phe-OBn	O	H	28 ± 6[a]
6h	*cis (S,S)*	Me	(*R*)-Phe	O	H	ni
6i	*cis (R,R)*	Me	(*S*)-Phe	O	H	ni
6k	*cis (R,R)*	Me	(*S*)-Val-OBn	O	H	10 ± 3[a]
6l (1.2/1)^c	*rac-cis (R,S + S,R)*	EtO$_2$C	(*S*)-Phe-OBn	O	H	22 ± 5[a]
7a	*cis (R,S)*	For structure see above				ni

All amino acids are abbreviated in the three-letter code; ni inhibition <10%; TMR three-membered ring.
[a] Percentage inhibition as obtained in the FRET-based assay, values are mean values of at least 2 independent assays.
[b] Percentage inhibition as obtained in the HPLC assay, mean value of four independent assays.
[c] Ratio of diastereomers.
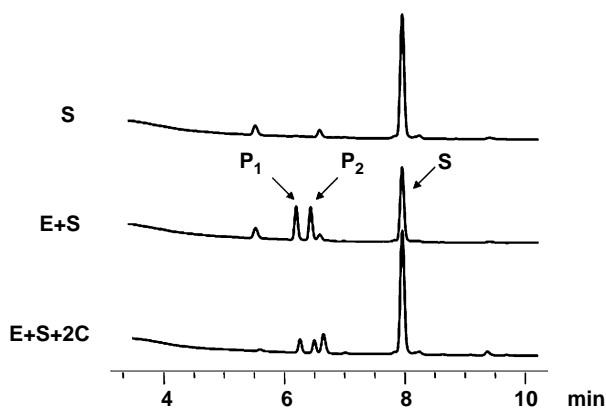
Figure 1. HPLC profiles of proteolytic reactions for determination of enzymatic activity of SARS-CoV M[pro] and of inhibition by **2c**. S, substrate VSVNSTLQ|SGLRKMA; E, enzyme SARS-CoV M[pro]; P, hydrolysis products VSVNSTLQ and SGLRKMA.

and respective aziridines[27] are highly active inhibitors of CAC1[31] cysteine protease. Their proposed inhibition mechanism is the alkylation of the active site cysteine residue. However, E-64 is reported to be inactive against coronaviral main proteases and 3C-like picornaviral proteases.[32–34] *cis*-Configured epoxides, on the other hand, are known to inhibit aspartic proteases[29] by alkylation of one aspartate residue within the active site, and α,β-epoxy ketones are reported to inhibit both serine and cysteine proteases, depending on the stereochemistry of the epoxide ring.[28]

To investigate whether epoxides, and aziridines could serve as electrophilic building blocks for coronaviral main proteases, and to evaluate which stereochemistry will be preferred by these proteases, both, *trans*- and *cis*-configured differently substituted three-membered heterocycles, were included in the screening.

First, a screening was performed with an HPLC-based assay using VSVNSTLQ|SGLRKMA[9,34,35] as substrate (Fig. 1). Second, the screening was extended using a less time-consuming and less intricate fluorimetric assay using a FRET-pair labeled substrate[25] (Table 1). The screening revealed the *trans*-configured N-unsubstituted aziridine-2,3-dicarboxylate **2c** (*S,S*)-(EtO)Azi-Gly-Gly-OBn,[36] which showed 75% inhibition of SARS-CoV M[pro] in the HPLC assay (Fig. 1) and 54% inhibition in the fluorimetric assay at 100 μM,[37] and the aziridine-2-carboxylate **4l**, which showed 48% inhibition, as most potent compounds (Table 1).

Within the series of *trans*-configured aziridines **2c**, **2d**, and **3a–3c**, only the Gly derivative **2c** shows considerable activity. Only weak activity is found for the derivatives-containing larger amino acids (**2d**, **3a–3c**) which, in contrast, are good inhibitors of CAC1 proteases (e.g., inhibition of cathepsin L by **3a**: K_i = 6.4 μM and by **3b**: K_i = 4.8 μM).[27]

The study also revealed that epoxide or aziridine building blocks alone, which do not bear an amino acid

moiety, are not active (**1a**, **1b**, **2a**, **2b**, and **6a–6c**). Within the series of *cis*-configured epoxides and aziridines weak inhibition is exhibited by the *N*-benzyl aziridines-(**1c–1f**) and the PheOBn-containing epoxides **6d**, **6g**, and **6l**. The free acids (**6h**, **6i**) are inactive. The diastereomeric mixture of PheOBn-containing *N*-benzyl aziridines **1c** is slightly more active than the pure compound **1d**, suggesting that the diastereomer with (*R,R*) configured aziridine ring is the more potent isomer. With the exception of **4k** all aziridine-2-carboxylates (**1c–1f**, **4a–4i**, **4l**) are active in the range between 20 and 50%, whereas all aziridine-2,2-dicarboxylates (**5a–5c**) are inactive. Interestingly, epoxide **7a** which is the only compound without an electron-withdrawing substituent at the three-membered ring does not show any activity.

These results show that in contrast to CAC1 proteases which are only inhibited by *trans*-configured three-membered heterocycles[28,30,38] *cis*-configured analogues can serve as building blocks for inhibitors of PAC30 proteases as well.

To better understand the relevant interactions between the most potent inhibitor **2c** and the SARS-CoV M[pro], docking experiments using FlexX[TM39] were carried out. The binding site was extracted from the recently published structure of the complex of SARS-CoV M[pro] with a peptpeptidyl chloromethyl ketone (CMK) (PDB code: 1UK4).[10]

Figure 2 shows a docking overlay of the CMK (green), with the aziridinyl peptide **2c** (orange) and the proposed binding mode of **2c**.

The substrate analogue hexapeptidyl chloromethyl ketone inhibitor (Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK) is shown as found after elimination of the covalent bond and subsequent minimization of the active site of the SARS-CoV M[pro].

The docking of the aziridine derivative **2c** (orange) suggests that the reactive center of the compound is located in close proximity to the sulfur of Cys145 (co-crystallized ligand: 3.14 Å; **2c**, 3.91 Å). The main part of **2c** is located in the S1 pocket of the enzyme.

For this compound the interactions with the enzyme are described by hydrogen bonds to amino acids of the B-chain (Ser1) and A-chain (Ser144, His163, and His172) (Fig. 2) suggesting that **2c** should better fit into the protein dimer which is formed in solution at high concentrations and which is supposed to be the active enzyme form.[40] For the docked conformations, hydrophobic interactions are found for the terminal ethyl group only. This group is positioned in proximity to the S1′ pocket. In the docked conformations, the terminal benzyl residue is solvent exposed, suggesting that this group is not overly important for a high biological activity. Since neither the ethyl nor the benzyl groups show optimal fit into the enzyme, a number of possibilities for synthetical optimization are conceivable. These include enlargement of the ethyl group and replacement of the benzyl group by an amide group mimicking the side chain of Gln
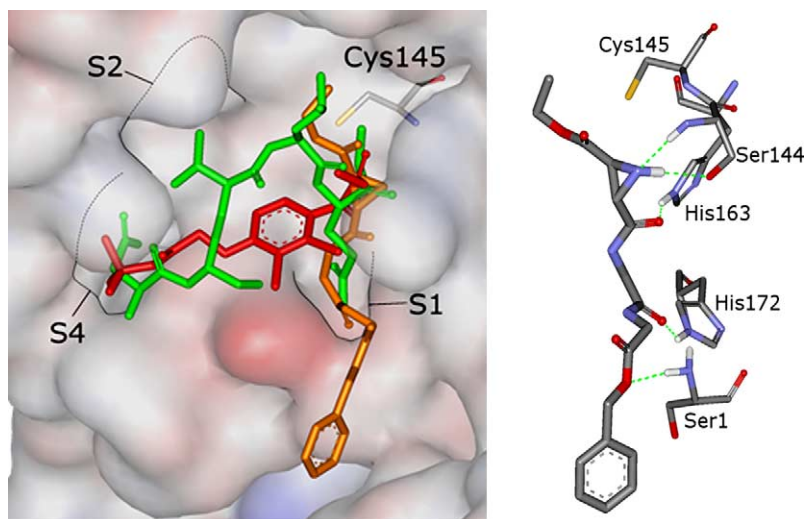
Figure 2. Docking overlay of hexapeptidyl CMK (green) and aziridinyl peptide **2c** (orange) (left). Predicted H-bonds between **2c** and SARS-CoV M^{pro} (right).

which is supposed to be the optimal residue for the S1 pocket. In this context, a modification of the peptidic nature of **2c** into a peptidomimetic one has also to be kept in mind due to pharmacokinetic reasons.

In summary, a comprehensive screening of electrophilic compounds has revealed the *trans*-configured aziridine-2,3-dicarboxylate **2c** as modest active-site directed SARS-CoV M^{pro} inhibitor with potential for further optimization. In addition, aziridine- and oxirane-2-carboxylic acid-containing compounds also show weak inhibitory activity. This activity might be enhanced when the electrophilic building blocks are linked to appropriate amino acids (e.g., Gln), substrate analogue peptides or peptidomimetics.

Supplementary data

Experimental section: syntheses and analytical data of compounds, enzyme assays, docking procedures. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2005.09.012.

References and notes

1. Siddell, S. G.; Ziebuhr, J.; Snijder, E. J. Coronaviruses, Toroviruses, and Arteriviruses. In *Topley's Microbiology and Microbial Infections*; 10th Ed.; Mahy, B. W. J., ter Meulen, V., Eds.; Hodder Arnold: London, 2005; Vol. 1, pp 823–856.
2. Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H. R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A.; Berger, A.; Burguiere, A. M.; Cinatl, J.; Eickmann, M.; Escriou, N.; Grywna, K.; Kramme, S.; Manuguerra, J. C.; Muller, S.; Rickerts, V.; Sturmer, M.; Vieth, S.; Klenk, H. D.; Osterhaus, A. D.; Schmitz, H.; Doerr, H. W. *N. Engl. J. Med.* **2003**, *348*, 1967.
3. Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J. A.; Lim, W.; Rollin, P. E.; Dowell, S. F.; Ling, A. E.; Humphrey, C. D.; Shieh, W. J.; Guarner, J.; Paddock, C. D.; Rota, P.; Fields, B.; DeRisi, J.; Yang, J. Y.; Cox, N.; Hughes, J. M.; LeDuc, J. W.; Bellini, W. J.; Anderson, L. J. *N. Engl. J. Med.* **2003**, *348*, 1953.
4. Peiris, J. S.; Lai, S. T.; Poon, L. L.; Guan, Y.; Yam, L. Y.; Lim, W.; Nicholls, J.; Yee, W. K.; Yan, W. W.; Cheung, M. T.; Cheng, V. C.; Chan, K. H.; Tsang, D. N.; Yung, R. W.; Ng, T. K.; Yuen, K. Y. *Lancet* **2003**, *361*, 1319.
5. Ziebuhr, J. *Curr. Opin. Microbiol.* **2004**, *7*, 412.
6. Ziebuhr, J. *Curr. Top. Microbiol. Immunol.* **2005**, *287*, 57.
7. Ziebuhr, J.; Snijder, E. J.; Gorbalenya, A. E. *J. Gen. Virol.* **2000**, *8*, 853.
8. Anand, K.; Palm, G. J.; Mesters, J. R.; Siddell, S. G.; Ziebuhr, J.; Hilgenfeld, R. *EMBO J.* **2002**, *21*, 3213.
9. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. *Science* **2003**, *300*, 1763.
10. Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. *Proc. Natl. Acad Sci. U.S.A.* **2003**, *100*, 13190.
11. Liu, B.; Zhou, J. *J. Comput. Chem.* **2005**, *26*, 484.
12. Rajnarayanan, R. V.; Dakshanamurthy, S.; Pattabiranam, N. *Biochem. Biophys. Res. Commun.* **2004**, *321*, 370.
13. Chou, K.-C.; Wei, D.-Q.; Zhong, W.-Z. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 148.
14. Jenwitheesuk, E.; Samudrala, R. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3989.
15. Zhang, X. W.; Yap, Y. L.; Altmeyer, R. M. *Eur. J. Med. Chem.* **2005**, *40*, 57.
16. Du, Q.; Wang, S.; Wei, D.; Sirois, S.; Chou, K.-C. *Anal. Biochem.* **2005**, *337*, 262.
17. 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.; Li, Y.-X.; Chen, K.-X.; Luo, X.-M.; Shen, X.; Shen, J.-H.; Jiang, H.-L. *Acta. Pharmacol. Sin.* **2003**, *24*, 497.
18. Zhang, X.-W.; Yap, Y. L. *Bioorg. Med. Chem.* **2004**, *12*, 2517.

E. Martina et al. / Bioorg. Med. Chem. Lett. 15 (2005) 5365–5369

5369

19. Toney, J. H.; Navas-Martin, S.; Weiss, S. R.; Koeller, A. *J. Med. Chem.* **2004**, *47*, 1079.

20. Wu, C. Y.; Jan, J. T.; Ma, S. H.; Kuo, C. J.; Juan, H. F.; Cheng, Y. S.; Hsu, H. H.; Huang, H. C.; Wu, D.; Brik, A.; Liang, F. S.; Liu, R. S.; Fang, J. M.; Chen, S. T.; Liang, P. H.; Wong, C. H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10012.

21. Blanchard, J. E.; Elowe, N. H.; Huitema, C.; Fortin, P. D.; Cechetto, J. D.; Eltis, L. D.; Brown, E. D. *Chem. Biol.* **2004**, *11*, 1445.

22. Kao, R. Y.; To, A. P. C.; Ng, L. W. Y.; Tsui, W. H. W.; Lee, T. S. W.; Tsoi, H.-W.; Yuen, K.-Y. *FEBS Lett.* **2004**, *576*, 325.

23. Hsu, J. T.-A.; Kuo, C.-J.; Hsieh, H.-P.; Wang, Y.-C.; Huang, K.-K.; Lin, C. P.-C.; Huang, P.-F.; Chen, X.; Liang, P.-H. *FEBS Lett.* **2004**, *574*, 116.

24. Bacha, U.; Barrila, J.; Velazquez-Campoy, A.; Leavitt, S. A.; Freire, E. *Biochemistry* **2004**, *43*, 4906.

25. Käppler, U.; Stiefl, N.; Schiller, M.; Vicik, R.; Breuning, A.; Schmitz , W.; Rupprecht, D.; Schmuck, C.; Baumann, K.; Ziebuhr, J.; Schirmeister, T.; *J. Med. Chem.* **2005** accepted.

26. Schirmeister, T.; Klockow, A. *Mini Rev. Med. Chem.* **2003**, *3*, 585.

27. Vicik, R.; Busemann, M.; Gelhaus, C.; Stiefl, N.; Scheiber, J.; Schmitz, W.; Schulz, F.; Mladenovic, M.; Engels, B.; Leippe, M.; Baumann, K.; Schirmeister, T.; *J. Med. Chem.* **2005** in revision.

28. Powers, J. C.; Asgian, J. L.; Ekici, Ö. D.; James, K. E. *Chem. Rev.* **2002**, *102*, 4639.

29. Ro, S.; Baek, S. G.; Lee, B.; Ok, J. H. *J. Pept. Res.* **1999**, *54*, 242.

30. Hanada, K.; Tamai, M.; Morimoto, S.; Adachi, T.; Ohmura, S.; Sawada, J.; Tanaka, I. *Agric. Biol. Chem.* **1978**, *42*, 537.

31. For classification of proteases see: http://merops.sanger. ac.uk/index.htm.

32. Kleina, L. G.; Grubman, M. J. *J. Virol.* **1992**, *66*, 7168.

33. Seybert, A.; Ziebuhr, J.; Siddell, S. G. *J. Gen. Virol.* **1997**, *78*, 71.

34. Ziebuhr, J.; Heusipp, G.; Siddell, S. G. *J. Virol.* **1997**, *71*, 3992.

35. Hegyi, A.; Ziebuhr, J. *J. Gen. Virol.* **2002**, *83*, 595.

36. Schirmeister, T.; Breuning, A.; Murso, A.; Stalke, D.; Mladenovic, M.; Engels, B.; Szeghalmi, A.; Schmitt, M.; Kiefer, W.; Popp, J. *J. Phys. Chem. A* **2004**, *108*, 11398.

37. **2c** is an irreversible inhibitor ($k_{obs}/I = 311 \pm 60\ M^{-1}\ min^{-1}$) which means that inhibition is time-dependent. Hence the higher value for inhibition by **2c** obtained in the HPLC assay compared to the fluorimetric assay is due to the much longer incubation time of the former.

38. Schirmeister, T. *Arch. Pharm. Pharm. Med. Chem.* **1996**, *329*, 239.

39. FlexX® 112 BioSolveIT, An der Ziegelei 75, 53757 St. Augustin, Germany 2003.

40. Chou, C. Y.; Chang, H. C.; Hsu, W. C.; Lin, T. Z.; Lin, C. H.; Chang, G. G. *Biochemistry* **2004**, *43*, 14958.