

A Scaffold-Tree-Merging Strategy for Prospective Bioactivity Annotation of γ -Pyrones**

Stefan Wetzel, Wolfram Wilk, Samy Chammaa, Bianca Sperl, Anke G. Roth, Aybike Yektaoglu, Steffen Renner, Thorsten Berg, Christoph Arenz, Athanassios Giannis, Tudor I. Oprea, Daniel Rauh, Markus Kaiser, and Herbert Waldmann*

[*] Dr. S. Wetzel,^[§] Dr. W. Wilk,^[+] Dr. S. Chammaa, Dr. S. Renner,^[§]

Prof. Dr. H. Waldmann

Department of Chemical Biology

Max Planck Institute of Molecular Physiology

Otto-Hahn-Strasse 11, 44227 Dortmund (Germany)

and

Faculty of Chemistry, Chemical Biology

Technical University Dortmund, Dortmund (Germany)

E-mail: herbert.waldmann@mpi-dortmund.mpg.de

B. Sperl, Prof. Dr. T. Berg^[++]

Department of Molecular Biology

Max Planck Institute of Biochemistry and

Center for Integrated Protein Science Munich (CIPSM)

Am Klopferspitz 18, 82152 Martinsried (Germany)

A. G. Roth, Prof. Dr. C. Arenz

Institut für Chemie, Humboldt-Universität zu Berlin

Brook-Taylor-Strasse 2, 12489 Berlin (Germany)

A. Yektaoglu, Prof. Dr. A. Giannis

Institute of Organic Chemistry, University of Leipzig

Johannisallee 29, 04103 Leipzig (Germany)

Dr. T. I. Oprea

Division of Biocomputing, University of New Mexico School of

Medicine, Albuquerque, NM 87131 (USA)

Dr. D. Rauh, Dr. M. Kaiser

Chemical Genomics Centre of the Max Planck Society

Otto-Hahn-Strasse 15, 44227 Dortmund (Germany)

[++] New address:

Institute of Organic Chemistry, University of Leipzig (Germany)

[§] Present address:

Novartis Institutes of Biomedical Research, Basel (Switzerland)

[+] These authors contributed equally.

[**] We thank Peter Ertl, Ansgar Schuffenhauer, and Silvio Roggo from the Novartis Institute of Biomedical Research, Basel for stimulating discussions. This research was supported by the Max Planck Society, the State of North Rhine-Westphalia, and the European Regional Development Fund, as well as by the German Federal Ministry for Education and Research through the German National Genome Research Network Plus (NGFN-Plus; grant No. BMBF 01GS08102, D.R., H.W.). S.C. is grateful to the Ministerio de Educación y Ciencia of Spain for a postdoctoral fellowship. Software was generously supplied by Accelrys/SciTEGic (PipelinePilot) and Chapman & Hall/CRC (Informa and Openeye, OEChem). S.W. is grateful to Novartis for a PhD scholarship. Part of this project was supported by NIH grant 1U54MH084690 (T.I.O.). The research of B.S. and T.B. was generously supported by the Department of Molecular Biology (director: Axel Ullrich) at the Max Planck Institute of Biochemistry and the Bundesministerium für Bildung und Forschung (NGFN-2, grant 01GS0451 to T.B.).



Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.200906555>.

Natural products (NPs) can be regarded as biologically relevant compound classes selected during evolution and endowed with structural parameters required for binding to proteins. They are a rich source of small-molecule modulators of protein function.^[1] However, their use is often hampered by their limited accessibility and a lack of knowledge about their biological targets. Synthetically accessible NP-inspired and NP-derived compound collections may offer an attractive possibility to overcome these limitations if their bioactivity could be annotated prospectively. Cheminformatic methods can provide crucial clues and critical insight to guide compound-development programs. Accordingly, several cheminformatic approaches have been described for the analysis of NP structure and properties in the context of bioactive-compound development and to chart NP chemical space.^[2]

We recently introduced a hierarchical, chemistry-based classification and arrangement of natural products in a natural product scaffold tree^[3,4] to inspire the synthesis of NP-derived compound collections.^[5–7] Brachiation along lines of biological relevance, that is, movement along the branches of the NP tree from complex to simpler structures, enables the identification of structurally simpler scaffolds with retained, yet varying bioactivity.^[6,8,9] We also found that brachiation through scaffold trees built with bioactivity as the key selection criterion is widely distributed among all major pharmaceutical protein target classes.^[8] In the light of these findings, we hypothesized that the correlation of NP scaffolds not annotated for bioactivity with scaffolds of related yet non-natural small molecules with known bioactivity and targets could identify potential proteins targeted by natural product inspired and derived compound collections. We envisaged that merging of trees not annotated for bioactivity (e.g. the NP scaffold tree) with trees annotated for bioactivity would generate novel merged trees from which bioactivity could be assigned prospectively (Figure 1 a).

To investigate this notion, a tree of approximately 60 000 scaffolds was generated from the Dictionary of Natural Products (DNP), version 17.2 according to the chemistry-derived set of rules described by Schuffenhauer et al.^[4] (see Figure 1 in the Supporting Information and associated text) and annotated with biological targets from the WOMBAT database^[10] (see Figure 4 in the Supporting Information) through the analogous generation of a WOMBAT scaffold tree and subsequent tree merging (see Figure 1 in the Supporting Information; for a more detailed analysis of the

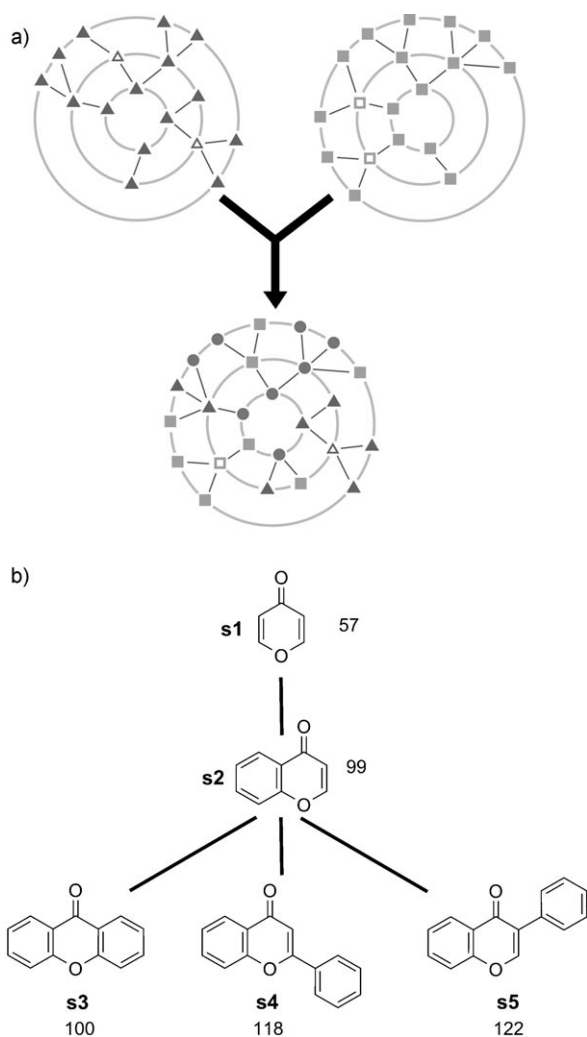


Figure 1. Tree annotation and library composition. a) Schematic illustration of the merging of the scaffold trees from the DNP (triangles) and the WOMBAT database (squares). The symbols denote the scaffolds in the tree. During merging, identical scaffolds are mapped onto each other (circles), and the remaining scaffolds are added to the resulting tree. The nonfilled symbols denote "virtual scaffolds" that are created by scaffold-tree generation but do not represent molecules in the data set. b) Composition of the screening library composed of five γ -pyrone scaffold types.

WOMBAT tree, see the text, Table 2, and Figure 2 in the Supporting Information).

As a representative example to explore the approach detailed above, the γ -pyrone branch of the merged tree was chosen because it covers 8171 and 1701 compounds in the DNP and WOMBAT trees, respectively, and includes annotations for 133 targets in WOMBAT (see Figure 3a in the Supporting Information), and because a compound collection with sufficient substituent diversity to guarantee statistical significance was accessible (see below). Moreover, annotation in WOMBAT indicated that γ -pyrones are not promiscuous compounds, as 97 % of the γ -pyrones were annotated for only two targets or less (see Figure S3b in the Supporting Information).

Initial analysis of the target data from WOMBAT identified several proteins which are inhibited with an IC_{50} value of $10\ \mu\text{M}$ or less by γ -pyrones belonging to at least three different hierarchy levels of the γ -pyrone branch, that is, γ -pyrones for which brachiation is validated (see Table 2 in the Supporting Information).

Our experience^[3,7] has shown that at least 100–200 natural product inspired compounds are required to attain a hit rate of 1–2 % in biochemical assays. Therefore, we assembled a collection of 500 γ -pyrones spanning three hierarchy levels in the branch through synthesis and from commercial sources (Figure 1b; see also the Supporting Information for syntheses). For the library, the calculated octanol–water partition coefficient (AlogP^[11]), the number of hydrogen-bond donors and acceptors, and the number of rotatable bonds were within acceptable limits. Thus, 78 % of the library members were compliant with the rule-of-five,^[12] and more than half of the natural product inspired compounds were leadlike^[13] (see Figure 5 and Table 10 in the Supporting Information).

From the annotated targets for the γ -pyrone scaffold branch (see Table 3 in the Supporting Information), we selected proteins that were inhibited by compounds containing at most two of the five scaffolds (see Table 2 in the Supporting Information) and that belonged to three different protein families. The proteins included the monoamine oxidases A and B (MAO A/B), which are documented targets of antidepressive drugs,^[14] the signal transducers and activators of transcription (STATs) 1, 3, and 5B, which are anticancer- and anti-inflammatory-drug targets,^[15] as well as acid and neutral sphingomyelinase, which are involved in apoptosis, Niemann–Pick disease types A and B, and secondary infections in cystic fibrosis.^[16]

Following pre-screens at fixed concentration, IC_{50} values were measured to identify hit candidates (see the Supporting Information for details, including Figure 6). For MAO A, the assay^[17] yielded 60 inhibitors with IC_{50} values of $10\ \mu\text{M}$ or lower; 15 of these inhibitors had an IC_{50} value less than $1\ \mu\text{M}$. For MAO B, 35 inhibitors with IC_{50} values below $10\ \mu\text{M}$ and 11 inhibitors with IC_{50} values below $1\ \mu\text{M}$ were identified. Of these hits, 31 were more than 10-fold and 7 were more than 100-fold selective for MAO A, and 10 were more than 10-fold and 6 were more than 100-fold selective for MAO B (for selected selective hits, see Table 1; for all hits, see Table 4 in the Supporting Information).

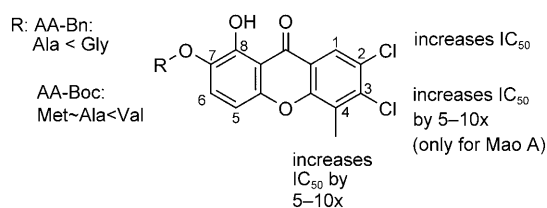
In the merged scaffold tree, mostly isoflavonoid scaffolds are annotated for MAO inhibition, annotation is significantly decreased for flavonoids, and only one xanthone is annotated for MAO inhibition (see Table 5 in the Supporting Information). The screen identified a structurally diverse set of active compounds spanning all scaffold sizes. Analysis of the hit rates per scaffold class identified the xanthone scaffold as a particularly valid starting point for the development of both MAO A and B inhibitors as well as the flavone scaffold for MAO A inhibition. The xanthone scaffold exhibited low ligand-binding efficiency (LBE),^[18] probably owing to the size of the inhibitors; in this respect, the chromone scaffold family may be more attractive, with a hit rate of about 5 % in both assays. This value is high in comparison with the average hit rate of 0.34 % in PubChem assays. The delineation of sensible

Table 1: Selected hits from the monoamine oxidase screen.^[a]

	Compound	IC ₅₀ [μM]		LBE
		MAO A	MAO B	
1		0.95 ± 0.07	> 50	0.30
2		1.31 ± 0.04	> 50	0.31
3		0.94 ± 0.03	> 50	0.35
4			0.34 ± 0.17	0.28
5			0.31 ± 0.03	0.24
6			0.54 ± 0.06	0.24

[a] The IC₅₀ value for the other isoenzyme is above 50 and 100 μM for hits 1–3 and 4–6, respectively. This indicates that these inhibitors are selective.

structure–activity–relationship (SAR) patterns proved difficult, since a set of diverse compounds had been tested. However, for the xanthone scaffold, an SAR could be established (Figure 2). Substitution at the 2- or 4-position generally led to reduced potency of the compounds; substitution at the 3-position led to reduced potency only for MAO A inhibition.

**Figure 2.** Structure–activity–relationship pattern derived for the xanthones. Bn = benzyl, Boc = *tert*-butoxycarbonyl; AA = amino acid.

STATs 1, 3, and 5 contain a conserved Src homology 2 (SH2) domain, by which they bind to activated cytokine receptors or growth-factor receptors. Small-molecule inhibitors are mainly known for STAT3,^[15] for STAT5 and STAT1, ligands for the SH2 domains with a chromone structure have been described.^[19] The screen (see the Supporting Information) revealed compounds **7**, **8**, and **9** as selective STAT inhibitors (Table 2; see also Table 6 in the Supporting Information).

Table 2: In vitro activity against the SH2-domain functions of STAT proteins, as determined by fluorescence polarization assays.^[a]

	Compound	IC ₅₀ (apparent) [μM] against		
		STAT5B	STAT3	STAT1
7		32 ± 1	70 ± 3	> 500
8		38 ± 2	31 ± 1	29 ± 1
9		14 ± 1	22 ± 1	39 ± 1

[a] Compound **7**, which was discovered in a previous study,^[19] was characterized to enable direct comparison of the data.

The biochemical results for compounds **7** and **8** were confirmed by means of cellular experiments (Figure 3). Schiff base **9** was not active in cells, probably as a result of hydrolysis of the imine. We found that compound **8** was stable under the conditions of the assay for at least 3 h, which is well within the time required for the experiment. Compounds **8** and **9** share their core structures with inhibitors recently discovered independently by means of a high-volume screening approach by Berg and co-workers^[19] (see Table 7 in the Supporting Information) but with different selectivity patterns. The chromones identified independently by two fairly different approaches define one of the very few known nonpeptidic small-molecule STAT-inhibitor classes.

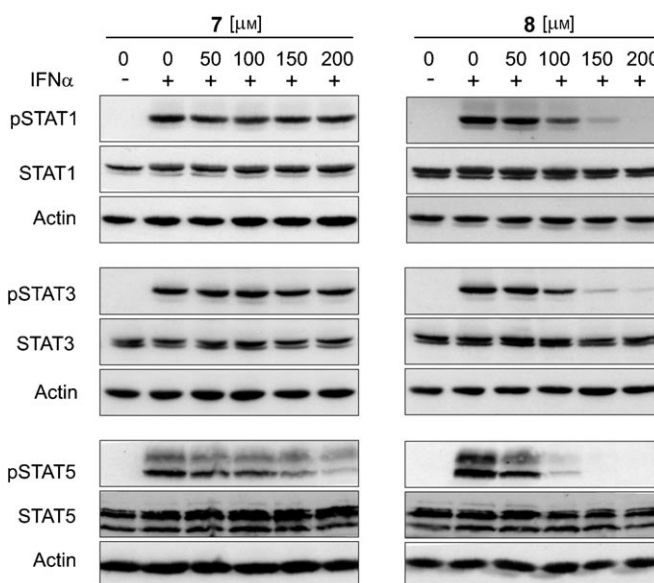
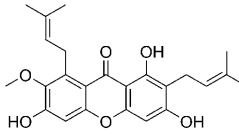
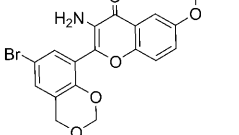
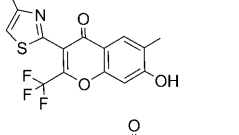
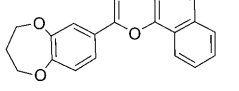
**Figure 3.** Cellular activities of the inhibitors **7** and **8** as determined by western blotting with phosphotyrosine-specific antibodies (see the Supporting Information for experimental details). Since STAT phosphorylation requires prior STAT binding to activated receptors through the STAT SH2 domain, an inhibitor of STAT will lead to reduced levels of tyrosine-phosphorylated STAT protein (pSTAT). IFN = interferon α_{2A}.

Table 3: Results of the screens for sphingomyelinase inhibition.^[a]

	Compound	IC ₅₀ /K _i [μ M]		LBE	AlogP
		acid	neutral		
10		12.5	113.3	0.16	5.9
11		3.1 $\pm$ 2.1	$\geq$ 50	0.21	2.6
12		9.5 $\pm$ 2.8	$\geq$ 50	0.22	3.4
13		21.5 $\pm$ 6.8	> 50	0.18	3.9

[a] Compound **10** is one of the xanthones from the WOMBAT database that were used to annotate acid sphingomyelinase inhibition to the γ -pyrone branch.^[10,21] Compounds **11–13** are inhibitors from our screening that exhibited an IC₅₀ value below 25 μ M. We also determined the IC₅₀ value against the neutral enzyme to assess the isoenzyme selectivity. The newly discovered hits also possess more promising properties than those of the known inhibitors, that is, improved LBE and improved AlogP values.

Only a few inhibitors of the second-messenger-generating sphingomyelinases are known.^[20] We used a fluorescence-based assay for acid sphingomyelinase and determined the inhibition of neutral sphingomyelinase by a radioactivity-based assay. The screen revealed two potent compounds with a benzopyran scaffold (Table 3; see also Table 8 in the Supporting Information).

The best inhibitor, **11**, showed an IC₅₀ value of 3.1 μ M for acid sphingomyelinase and a pronounced selectivity for acid sphingomyelinase over neutral sphingomyelinase. Remarkably, our screen identified no compounds from the xanthone scaffold group, although α -mangostin (**10**)^[21] and its derivatives, the only γ -pyrone-containing inhibitors of sphingomyelinase described so far, guided the prospective target annotation (see Table 9 in the Supporting Information). It had been hypothesized that the prenyl groups in these compounds might be required for selective inhibition of acid sphingomyelinase, as they make the compounds lipophilic. The new hits are promising starting points for further optimization given their better potency and LBE with higher selectivity and significantly lower AlogP values.

These examples clearly indicate that the use of hierarchical scaffold trees, together with brachiation and tree merging, may be an efficient means for the prospective identification of targets for natural product inspired library scaffolds. The bioactivity-/target-annotation approach advanced in this study focuses on the compound-class scaffold rather than on individual compounds. Since individual members of a compound collection differ in their substituents and the positioning of these substituents on the scaffold, they will also differ in

their precise interaction with target proteins and their bioactivity. However, as a whole, the library should yield protein ligands with a relatively high hit rate. By analogy, the approach may not be able to identify the targets of individual guiding natural products. Nevertheless, given the conservation of protein structure during evolution,^[22] similarity or homology between the protein(s) targeted by this collection and others may guide the identification of further possible target-protein candidates.

The scaffold-tree-merging approach should primarily be regarded as a hypothesis-generating method best applied in the initial steps of inhibitor development. The initial hit compounds identified by our approach may frequently be only of limited potency and require further development. The scaffold tree-merging approach offers new opportunities for protein-ligand discovery over and above those available through brachiation alone. Whereas brachiation in general leads to smaller scaffolds within a given tree branch, after tree merging and annotation, larger scaffolds or even scaffolds from neighboring branches can be identified. Thus, for monoamine oxidases and sphingomyelinases, scaffolds from neighboring branches as well as scaffolds with more rings than the guiding scaffolds were identified, and for the STATs a hit class with a larger scaffold was found.

Received: November 20, 2009

Revised: March 4, 2010

Published online: April 14, 2010

Keywords: chemical space · cheminformatics · natural products · scaffold trees · target space

[1] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2007**, 70, 461.

[2] a) T. Henkel, R. M. Brunne, H. Müller, F. Reichel, *Angew. Chem.* **1999**, 111, 688; *Angew. Chem. Int. Ed.* **1999**, 38, 643; b) M. Feher, J. M. Schmidt, *J. Chem. Inf. Comput. Sci.* **2003**, 43, 218; c) P. Ertl, A. Schuffenhauer in *Natural Products as Drugs* (Eds.: F. Petersen, R. Amstutz), Birkhäuser Verlag, Basel, **2007**; d) K. Grabowski, G. Schneider, *Curr. Chem. Biol.* **2007**, 1, 115; e) T. Langer, C. Laggner, J. M. Rollinger, H. Stuppner, *Chimia* **2007**, 61, 350; f) K. Grabowski, K.-H. Baringhaus, G. Schneider, *Nat. Prod. Rep.* **2008**, 25, 892; g) K.-D. Entian, A. Gockele, M. Karas, R. Kirschner, P. Koetter, M. Piepenbring, T. Stein, G. Schneider, *BIOspektrum* **2009**, 15, 408; h) J. M. Rollinger, D. Schuster, B. Danzl, S. Schwaiger, P. Markt, M. Schmidtke, J. Gertsch, S. Raduner, G. Wolber, T. Langer, H. Stuppner, *Planta Med.* **2009**, 75, 195; i) Y. Tanrikulu, O. Rau, O. Schwarz, E. Proschak, K.

- Siems, L. Müller-Kuhrt, M. Schubert-Zsilavecz, G. Schneider, *ChemBioChem* **2009**, *10*, 75; j) M. Rupp, T. Schroeter, R. Steri, H. Zettl, E. Proschak, K. Hansen, O. Rau, O. Schwarz, L. Müller-Kuhrt, M. Schubert-Zsilavecz, K.-R. Müller, G. Schneider, *ChemMedChem* **2010**, *5*, 191; k) D. Schuster, L. Kern, D. P. Hristozov, L. Terfloth, B. Bienfait, C. Laggner, J. Kirchmair, U. Grienke, G. Wolber, T. Langer, H. Stuppner, J. Gasteiger, J. M. Rolling, *Comb. Chem. High Throughput Screening* **2010**, *13*, 54.
- [3] a) M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17272; b) A. Nören-Müller, I. Reis-Corrêa, Jr., H. Prinz, C. Rosenbaum, K. Saxena, H. J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10606.
- [4] A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. A. Koch, H. Waldmann, *J. Chem. Inf. Model.* **2007**, *47*, 47.
- [5] a) B. Meseguer, D. Alonso-Díaz, N. Griebenow, T. Herget, H. Waldmann, *Angew. Chem.* **1999**, *111*, 3083; *Angew. Chem. Int. Ed.* **1999**, *38*, 2902; b) O. Barun, S. Sommer, H. Waldmann, *Angew. Chem.* **2004**, *116*, 3258; *Angew. Chem. Int. Ed.* **2004**, *43*, 3195.
- [6] a) I. R. Corrêa, Jr., A. Nören-Müller, H.-D. Ambrosi, S. Jakupovic, K. Saxena, H. Schwalbe, M. Kaiser, H. Waldmann, *Chem. Asian J.* **2007**, *2*, 1109; b) O. Schwarz, S. Jakupovic, H.-D. Ambrosi, L. O. Haustedt, C. Mang, L. Müller-Kuhrt, *J. Comb. Chem.* **2007**, *9*, 1104.
- [7] a) A. Nören-Müller, W. Wilk, K. Saxena, H. Schwalbe, M. Kaiser, H. Waldmann, *Angew. Chem.* **2008**, *120*, 6061; *Angew. Chem. Int. Ed.* **2008**, *47*, 5973; b) K. Kumar, H. Waldmann, *Angew. Chem.* **2009**, *121*, 3272; *Angew. Chem. Int. Ed.* **2009**, *48*, 3224.
- [8] S. Renner, W. van Otterlo, M. D. Seoane, S. Möcklinghoff, B. Hofmann, S. Wetzel, A. Schuffenhauer, P. Ertl, T. Oprea, D. Steinhilber, L. Brunsveld, D. Rauh, H. Waldmann, *Nat. Chem. Biol.* **2009**, *5*, 585.
- [9] S. Wetzel, K. Klein, S. Renner, D. Rauh, T. I. Oprea, P. Mutzel, H. Waldmann, *Nat. Chem. Biol.* **2009**, *5*, 581.
- [10] M. Olah, M. Mracec, L. Ostopovici, R. Rad, A. Bora, N. Hadaruga, I. Olah, M. Banda, Z. Simon, M. Mracec, T. I. Oprea in *Methods and Principles in Medicinal Chemistry*, Vol. 23 (Ed.: T. I. Oprea), Wiley-VCH, Weinheim, **2005**, p. 223.
- [11] A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, *J. Phys. Chem. A* **1998**, *102*, 3762.
- [12] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **2001**, *46*, 3.
- [13] H. J. Verheij, *Mol. Diversity* **2006**, *10*, 377.
- [14] a) M. B. H. Youdim, D. Edmondson, K. F. Tipton, *Nat. Rev. Neurosci.* **2006**, *7*, 295; b) M. Bortolato, K. Chen, J. C. Shih, *Adv. Drug Delivery Rev.* **2008**, *60*, 1527.
- [15] a) T. Berg, *ChemBioChem* **2008**, *9*, 2039; b) L. Costantino, D. Barlocco, *Curr. Med. Chem.* **2008**, *15*, 834; c) S. Fletcher, J. Turkson, P. T. Gunning, *ChemMedChem* **2008**, *3*, 1159.
- [16] a) N. Marchesini, Y. A. Hannun, *Biochem. Cell Biol.* **2004**, *82*, 27; b) V. Teichgräber, M. Ulrich, N. Endlich, J. Riethmüller, B. Wilker, C. C. De Oliveira-Munding, A. M. van Heeckeren, M. L. Barr, G. von Kürthy, K. W. Schmid, M. Weller, B. Tümmler, F. Lang, H. Grassme, G. Döring, E. Gulbins, *Nat. Med.* **2008**, *14*, 382.
- [17] L. Novaroli, M. Reist, E. Favre, A. Carotti, M. Catto, P.-A. Carrupt, *Bioorg. Med. Chem.* **2005**, *13*, 6212.
- [18] C. H. Reynolds, B. A. Tounge, S. D. Bembenek, *J. Med. Chem.* **2008**, *51*, 2432.
- [19] J. Müller, B. Sperl, W. Reindl, A. Kiessling, T. Berg, *ChemBioChem* **2008**, *9*, 723.
- [20] a) F. Nara, M. Tanaka, T. Hosoya, K. Suzuki-Konagai, T. Ogita, *J. Antibiot.* **1999**, *52*, 525; b) R. Uchida, H. Tomoda, Y. Dong, S. Omura, *J. Antibiot.* **1999**, *52*, 572; c) C. Arenz, A. Giannis, *Angew. Chem.* **2000**, *112*, 1498; *Angew. Chem. Int. Ed.* **2000**, *39*, 1440; d) M. Kölzer, C. Arenz, K. Ferlinz, N. Werth, H. Schulze, R. Klingenstein, K. Sandhoff, *Biol. Chem.* **2003**, *384*, 1293; e) V. Wascholowski, A. Giannis, *Angew. Chem.* **2006**, *118*, 841; *Angew. Chem. Int. Ed.* **2006**, *45*, 827.
- [21] C. Okudaira, Y. Ikeda, S. Kondo, S. Furuya, Y. Hirabayashi, T. Koyano, Y. Saito, K. Umezawa, *J. Enzyme Inhib.* **2000**, *15*, 129.
- [22] M. A. Koch, L.-O. Wittenberg, S. Basu, D. A. Jeyaraj, E. Gourzoulidou, K. Reinecke, A. Odermatt, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16721.