DOI: 10.1002/cmdc.200800304

Hit Expansion through Computational Selectivity Searching

Dagmar Stumpfe,[a] Maxim Frizler,[b] Mihiret T. Sisay,[b] José Batista,[a] Ingo Vogt,[a] Michael Gütschow,[b] and Jürgen Bajorath*[a]

Finding small molecules that selectively interact with individual target proteins within target families is a major task in medicinal chemistry and chemical biology.[1] Currently, the identification of suitable small molecules relies primarily on the screening of diverse or specialized compound libraries.[2] In contrast, computational methods have thus far contributed only very little to the identification of molecules that are target selective or that have a selective tendency.[3] We have previously adapted ligand-based computational screening methods for selectivity searching and we now apply these methodologies in the search for cathepsin-K-selective inhibitors. Herein we report the identification of two nonpeptidic cathepsin K inhibitors that are weakly selective over cathepsins L and S. These inhibitors were identified by assaying only 16 candidate molecules taken from ~3.7 million virtually formatted database compounds. One inhibitor lacks an electrophilic "warhead" that is usually a prerequisite for the inhibition of cathepsins.

In virtual screening, computational methods are applied to search large databases for compounds that have a desired biological activity using ligand[4] and/or target structure[5] information as input. However, target selectivity has thus far not been explicitly considered in computational compound screening. We recently evaluated computational approaches for their ability to recognize target-selective compounds.[6–9] The results of these studies suggest that computational screening models may also have the potential to recognize target-selective compounds,[7–9] and this has triggered our current investigation.

Cathepsins K, S, and L are cysteine proteases belonging to the papain superfamily.[10] These closely related enzymes are involved in important physiological processes such as antigen presentation, bone remodeling, and apoptosis.[10,11] Accordingly, cathepsins have become relevant drug targets for the potential treatment of various diseases including cancer, osteoporosis, rheumatoid arthritis, and autoimmune disorders.[10–12] Cathepsin inhibitors are typically substrate analogues with an electrophilic warhead. First-generation inhibitors contain a strongly reactive group that covalently modifies the active site cysteine. More recently, second-generation reversible covalent inhibitors have also been introduced with a less-reactive functional group, typically a nitrile, which renders these inhibitors reversible and more desirable for therapeutic applications.

Among these cysteine proteases, cathepsin K has attracted particular interest. It is predominantly expressed in osteoclasts that mediate bone resorption and is capable of cleaving native type I collagen, the major component of bone matrix, and other components of bone matrix such as osteopontin and osteonectin.[12] Because bone matrix degradation is necessary for osteoclastic bone resorption, cathepsin K constitutes a major therapeutic target for the treatment of osteoporosis and is also implicated in rheumatoid arthritis and osteoarthritis.[10,11] There have been considerable efforts to develop selective cathepsin K inhibitors because simultaneous inhibition of cathepsins S or L is thought to be associated with unwanted side effects.[13,14] However, the design proved to be challenging owing to the high degree of structural and mechanistic similarity between the cathepsins. Only recently, the first inhibitors of human cathepsin K with selectivity over cathepsins S and L, balicatib[13] and odanacatib,[14] have proceeded to clinical evaluation.

Given their therapeutic potential and the difficulties experienced in the design of selective cathepsin K inhibitors, we considered the search for cathepsin K inhibitors that are selective over both cathepsins S and L a challenging and relevant test case for selectivity searching. For our analysis, we implemented a search protocol that involves two conceptually different in silico methods developed in our laboratory, a compound mapping algorithm termed DynaMAD[15] and a specialized type of molecular fingerprint consisting of compound class characteristic substructures, ACCS-FP.[16] These methods were practically applied here because they have been benchmarked for selectivity searching in computational studies.[8,9] Other virtual screening methods could also be applied, but have thus far not been evaluated for selectivity searching. Briefly, DynaMAD is designed to map database compounds to activity-specific consensus positions in chemical space representations of stepwise increasing dimensionality,[15] and ACCS-FP is used for search calculations in which fingerprints of reference and database molecules are compared and fingerprint overlap is quantified as a measure of molecular similarity.[16] Further methodological details including ACCS-FP generation and details of the search calculations are provided in the Supporting Information.

The computational screening methods applied here extrapolate from known ligand information in order to identify structurally diverse compounds with desired properties. Thus, the design of compound reference sets is generally very important for the outcome of the search calculations. For the assembly of a reference set for selectivity searching, we used differential

[a] D. Stumpfe,[+] Dr. J. Batista, Dr. I. Vogt, Prof Dr. J. Bajorath
Department of Life Science Informatics
Bonn–Aachen International Center for Information Technology
Rheinische Friedrich-Wilhelms-Universität Bonn
Dahlmannstr. 2, 53113 Bonn (Germany)
Fax: (+49) 228-2699-341
E-mail: bajorath@bit.uni-bonn.de

[b] M. Frizler,[+] M. T. Sisay, Prof Dr. M. Gütschow
Pharmazeutisches Institut, Pharmazeutische Chemie I
Rheinische Friedrich-Wilhelms-Universität Bonn
An der Immenburg 4, 53121 Bonn (Germany)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.200800304.

compound potency against cathepsins as a selectivity criterion. Accordingly, a total of 69 known inhibitors were assembled from the literature that had at least 50-fold higher potency against cathepsin K over cathepsins S and L, representing a relatively large reference set. The potency and selectivity distribution within this reference set is illustrated in Figure 1, and its

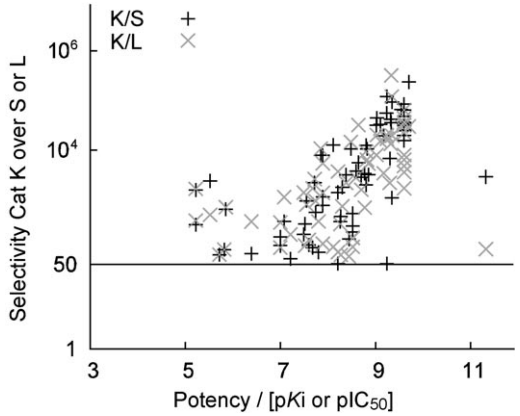


Figure 1. The potency and selectivity distribution of the compound reference set used for selectivity searching. The set consists of published inhibitors that display at least 50-fold higher potency for cathepsin K over cathepsins S or L. Compounds were only considered if potency against all three enzymes was reported. Each chosen reference inhibitor is represented by a K/S (+) as well as a K/L (×) symbol. The *x* axis reports the potency of each reference compound against cathepsin K, and the *y* axis the potency ratio used as a measure of selectivity.

exact composition and the literature sources are reported in the Supporting Information. Typically, the more reference compounds are used for search calculations, the more chemical information is available to facilitate hit or lead expansion. However, the number of reference compounds is usually less important than their chemical nature; for example, including many active analogues that represent the same chemotype adds only little structure–activity information.

The reference set was used for both DynaMAD and ACCS-FP calculations, which were carried out in parallel to screen ~3.7 million compounds with unique 2D structures taken from the publicly available ZINC database.[17] A summary of the selectivity search results is shown in Figure 2. On the basis of DynaMAD and ACCS-FP calculations, only 50 and 49 compounds were selected, respectively. No compound was selected by both methodologies. These sets were combined, and from pairs of molecules sharing the same core structure a compound was omitted, leading to the removal of 32 molecules. Of the remaining 67 candidate compounds, which are listed in table 1 of the Supporting Information, 16 could be obtained from commercial sources.

These 16 compounds were tested for enzyme inhibition in spectrophotometric assays for cathepsins S and L and a fluorimetric assay for cathepsin K. Assay details are provided in the Supporting Information. Assays were repeated multiple times, yielding consistent results. As shown in Table 1, three of the 16 compounds inhibited cathepsin activity with IC_{50} values in the micromolar range, which is typically observed for structurally

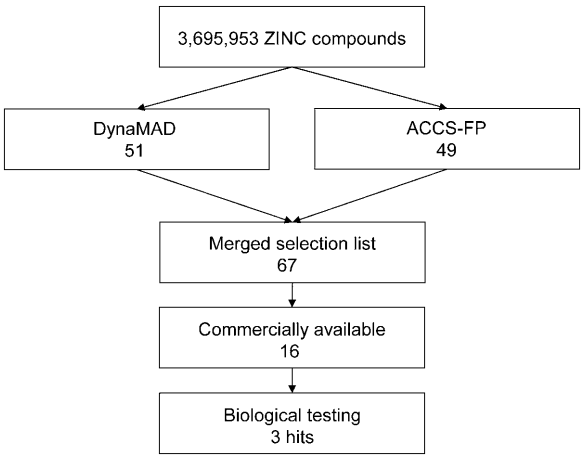
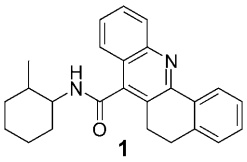
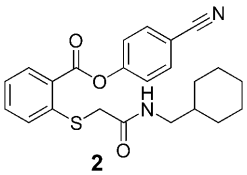
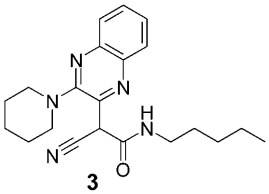


Figure 2. Selectivity searching. This diagram summarizes the results of selectivity searching, compound selection, and acquisition.

Table 1. Cathepsin K inhibitors[a]

Compound	IC_{50} [μM]
Cat K selective	
 **1**	Cat K: 48 Cat S: 200 Cat L: 290
 **2**	Cat K: 34 Cat S: 150 Cat L: > 500
Active compound	
 **3**	Cat K: 31 Cat S: 48 Cat L: 25

[a] Shown are the structures of three of 16 tested compounds that inhibit cathepsins K, S, and L at various levels. Compounds **1** and **2** are selective for cathepsin K over cathepsins S and L, while compound **3** is comparably active against all three enzymes.

diverse hits identified by virtual screening,[4,5] and two of these molecules, compounds **1** and **2**, were weakly selective for cathepsin K over both cathepsins S and L. Compound **1** was about five- to sixfold selective for cathepsin K over cathepsins S and L, and compound **2** was roughly fivefold selective over cathepsin S and essentially inactive against cathepsin L. Compound **1** was identified by ACCS-FP search, and compounds **2** and **3** using DynaMAD. Compound **1** has very low structural similarity to known cathepsin inhibitors. Its maximum conventional MACCS structural keys[18] Tanimoto similarity[19] to reference set

compounds was only 0.53. The corresponding values for compounds **2** and **3** were 0.79 and 0.80, respectively, which also indicate only limited structural similarity, below the level at which similarity in biological activity is expected.[20]

Importantly, no pre-existing pharmacophore information was taken into account in our selectivity search calculations, because the methods we applied evaluate similarity relationships on a whole-molecule basis.[4] In particular, no nitrile group constraint was present, and compounds with or without nitrile groups were computationally selected (eight of 16 tested compounds had a nitrile group). Compound **2** and the active but not selective compound **3** contain nitrile groups that are a hallmark of reversible cathepsin inhibitors, as mentioned above. In contrast, no nitrile function nor any other electrophilic warhead is present in compound **1**. For competitive inhibition, as suggested by our experiments, the IC_{50} values of compound **1** obtained for cathepsins K, S, and L correspond to K_i values of 6.1, 87, and 40 μM, respectively. This selective cathepsin K inhibitor is structurally distinct from known cathepsin inhibitors and is therefore a candidate for further chemical optimization.

In our study, ligand-based computational screening approaches were successfully used for the first time to intentionally identify compounds that have at least a target-selective tendency. The majority of our 67 computational candidate compounds could not be acquired, but testing of only 16 of our candidates confirmed two selective inhibitors. With compound **1**, a previously unobserved chemotype was identified. Although the compounds we identified are only weakly selective, they expand the current repertoire of cathepsin K inhibitors by adding new chemotypes with a target-selective tendency that might be useful as starting points for the development of selective and nonpeptidic cathepsin K inhibitors.

Acknowledgements

We thank D. Brömme (University of British Columbia, Canada) and K. Schilling (University of Jena, Germany) for providing cathepsin K. D.S. is supported by Sonderforschungsbereich (SFB) 704 of the Deutsche Forschungsgemeinschaft (DFG), and M.T.S. by a fellowship from Graduiertenkolleg (GRK) 677 of the DFG.

Keywords: cathepsin k inhibitors · molecular similarity · selectivity searching · small molecules · virtual screening

[1] B. R. Stockwell, *Nature* **2004**, *432*, 846–854.
[2] D. S. Tan, *Nat. Chem. Biol.* **2005**, *1*, 74–84.
[3] J. Bajorath, *Curr. Opin. Chem. Biol.* **2008**, *12*, 352–358.
[4] J. Bajorath, *Nat. Rev. Drug Discovery* **2002**, *1*, 882–894.
[5] B. K. Shoichet, *Nature* **2004**, *432*, 862–865.
[6] D. Stumpfe, H. E. A. Ahmed, I. Vogt, J. Bajorath, *Chem. Biol. Drug Des.* **2007**, *70*, 182–194.
[7] I. Vogt, D. Stumpfe, H. E. A. Ahmed, J. Bajorath, *Chem. Biol. Drug Des.* **2007**, *70*, 195–205.
[8] I. Vogt, H. E. A. Ahmed, J. Auer, J. Bajorath, *Mol. Diversity* **2008**, *12*, 25–40.
[9] D. Stumpfe, H. Geppert, J. Bajorath, *Chem. Biol. Drug Des.* **2008**, *71*, 518–528.
[10] D. Brömme, J. Kaleta, *Curr. Pharm. Des.* **2002**, *8*, 1639–1658.
[11] O. Vasiljeva, T. Reinheckel, C. Peters, D. Turk, V. Turk, B. Turk, *Curr. Pharm. Des.* **2007**, *13*, 387–403.
[12] S. A. Stoch, J. A. Wagner, *Clin. Pharmacol. Ther.* **2008**, *83*, 172–176.
[13] J. P. Falgueyret, S. Desmarais, R. Oballa, W. C. Black, W. Cromlish, K. Khougaz, S. Lamontagne, F. Massé, D. Riendeau, S. Toulmond, M. D. Percival, *J. Med. Chem.* **2005**, *48*, 7535–7543.
[14] J. Y. Gauthier, N. Chauret, W. Cromlish, S. Desmarais, T. Duong le, J. P. Falgueyret, D. B. Kimmel, S. Lamontagne, S. Léger, T. LeRiche, C. S. Li, F. Massé, D. J. McKay, D. A. Nicoll-Griffith, R. M. Oballa, J. T. Palmer, M. D. Percival, D. Riendeau, J. Robichaud, G. A. Rodan, S. B. Rodan, C. Seto, M. Thérien, V. L. Truong, M. C. Venuti, G. Wesolowski, R. N. Young, R. Zamboni, W. C. Black, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 923–928.
[15] H. Eckert, I. Vogt, J. Bajorath, *J. Chem. Inf. Model.* **2006**, *46*, 1623–1634.
[16] J. Batista, J. Bajorath, *ChemMedChem* **2008**, *3*, 67–73.
[17] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
[18] *MACCS Structural Keys*, Symyx Software, San Ramon, CA (USA), **2005**.
[19] P. Willet, J. M. Barnard, G. M. Downs, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.
[20] Y. C. Martin, J. L. Kofron, L. M. Traphagen, *J. Med. Chem.* **2002**, *45*, 4350–4358.

Received: September 18, 2008
Revised: October 16, 2008
Published online on December 3, 2008

