

## Discovery of Novel Cathepsin S Inhibitors by Pharmacophore-Based Virtual High-Throughput Screening

Patrick Markt,<sup>†</sup> Caroline McGoochan,<sup>‡</sup> Brian Walker,<sup>‡</sup> Johannes Kirchmair,<sup>†,§</sup> Clemens Feldmann,<sup>†</sup> Gabriella De Martino,<sup>||</sup> Gudrun Spitzer,<sup>⊥</sup> Simona Distinto,<sup>#</sup> Daniela Schuster,<sup>†,§</sup> Gerhard Wolber,<sup>§</sup> Christian Laggner,<sup>†</sup> and Thierry Langer<sup>\*,†,§</sup>

Department of Pharmaceutical Chemistry, Institute of Pharmacy and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 52c, 6020 Innsbruck, Austria, School of Pharmacy, Queen's University Belfast, 97 Lisburn Road, BT9 7BL Belfast, Northern Ireland, Inte:Ligand Softwareentwicklungs- and Consulting GmbH, Clemens Maria Hofbauer-Gasse 6, 2344 Maria Enzersdorf, Austria, Facoltà di Farmacia, Università di Roma, Piazzale Aldo Moro 5, 00185 Roma, Italy, Department of Theoretical Chemistry, Institute of General, Inorganic, and Theoretical Chemistry, University of Innsbruck, Innrain 52a, 6020 Innsbruck, Austria, and Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Cagliari, 09124 Cagliari, Italy

Received March 24, 2008

The cysteine protease cathepsin S (CatS) is involved in the pathogenesis of autoimmune disorders, atherosclerosis, and obesity. Therefore, it represents a promising pharmacological target for drug development. We generated ligand-based and structure-based pharmacophore models for noncovalent and covalent CatS inhibitors to perform virtual high-throughput screening of chemical databases in order to discover novel scaffolds for CatS inhibitors. An in vitro evaluation of the resulting 15 structures revealed seven CatS inhibitors with kinetic constants in the low micromolar range. These compounds can be subjected to further chemical modifications to obtain drugs for the treatment of autoimmune disorders and atherosclerosis.

### INTRODUCTION

Cathepsin S (CatS), a cysteine protease of the papain superfamily, is expressed by adipocytes and antigen presenting cells like macrophages, B cells, and dendritic cells.<sup>1,2</sup> The enzyme takes part in the pathogenesis of autoimmune disorders, atherosclerosis, and obesity.<sup>1–3</sup> Antigen presenting cells use major histocompatibility class II (MHC-II) molecules located on the cell surface to load and present antigenic peptides. These molecules are associated with the invariant chain (Ii), a type II membrane protein. In order to allow the binding of antigens onto the MHC-II molecule, Ii is degraded by CatS to the class II-associated invariant chain peptide which dissociates from the MHC-II molecule by forming a complex with the human leucocyte antigen-DM (HLA-DM).<sup>2,4,5</sup> CatS expression is elevated in atherosclerotic lesions. In the case of atherosclerotic lesions, the local extracellular matrix is degraded either by CatS secreted by smooth muscle cells and macrophages or by circulating CatS produced by adipose tissue. This elastolytic activity of CatS causes the migration of blood monocytes and smooth muscle cells into the vascular wall which leads to the progression of atherosclerotic lesions. Furthermore, CatS contributes to the development of adipose tissue by degrading extracellular

components of preadipocytes such as fibronectin and thereby promoting preadipocyte differentiation.<sup>1</sup> Thus, CatS represents a promising therapeutic target for a wide range of diseases. CatS inhibitors are expected to be useful in the treatment of diseases caused by immune hyperresponsiveness such as autoimmune diseases, asthma, multiple sclerosis, and rheumatoid arthritis as well as for the therapy of atherosclerosis.<sup>2,6,7</sup>

CatS is a monomeric protein consisting of 217 amino acids. The structure of CatS is composed of two domains. Residues 12 to 111 and 208 to 211 form the partly helical domain. The other domain consists of a six-stranded  $\beta$  barrel comprising residues 1 to 11 and 112 to 207. The two domains form a long and narrow binding pocket at the protein surface. The active site cleft borders are represented by a helix composed of residues 25 to 40 and two strands including residues 130 to 133 and 157 to 161. Two strands formed by residues 108 to 113 and 204 to 211 stabilize the active site.<sup>8</sup> The active site residues Cys25, His164, and Asn184 represent the catalytic triad of CatS (Figure 1a).

For the proper orientation of the imidazolium moiety of His164 for hydrolysis, a hydrogen bond is formed between Asn184 and His164 (Chart 1).

A proton transfer from the thiol group of Cys25 to the imidazolium moiety of His164 causes a reactive thiolate anion (Chart 1a). The thiolate anion attacks the scissile bond carbonyl atom of the CatS substrate (Chart 1b). The resulting tetrahedral intermediate is stabilized by an oxyanion hole composed of the Gln19 side chain and the Cys25 backbone. Another proton is transferred from the imidazolium cation of His164 to the scissile bond nitrogen atom of the substrate (Chart 1c). Thereby, the aminic substrate part dissociates

\* Corresponding author phone: +43-512-507-5252; fax: +43-512-507-5269; e-mail: Thierry.Langer@uibk.ac.at.

<sup>†</sup> Institute of Pharmacy and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck.

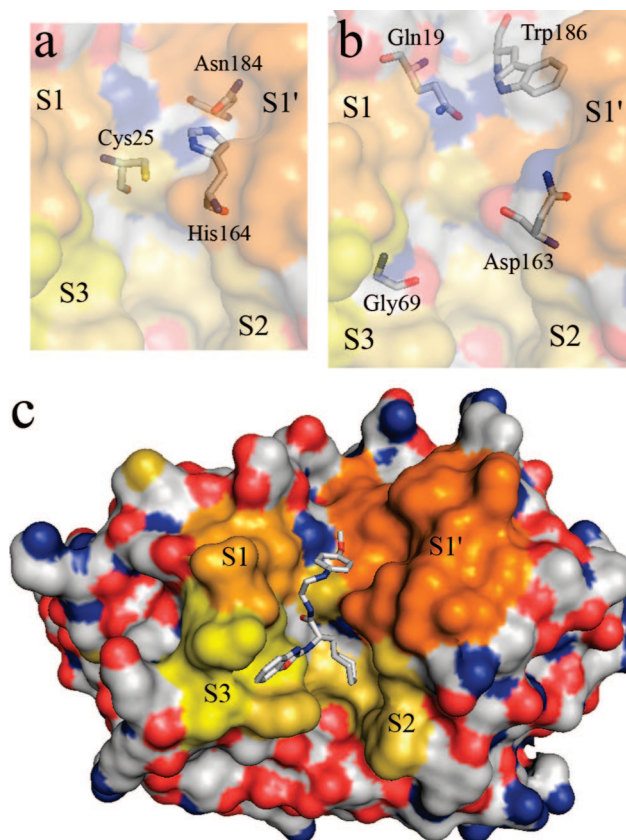
<sup>‡</sup> Queen's University Belfast.

<sup>§</sup> Inte:Ligand Softwareentwicklungs- and Consulting GmbH.

<sup>||</sup> Università di Roma.

<sup>⊥</sup> Institute of General, Inorganic, and Theoretical Chemistry, University of Innsbruck.

<sup>#</sup> Università degli Studi di Cagliari.



**Figure 1.** The catalytic triad (a), the CatS residues involved in the formation of hydrogen bonds to ligands (b), and the four hydrophobic pockets (c). The figure was derived from the Brookhaven Protein Data Bank (PDB, see below) entry 2f1g which is a complex between the noncovalent inhibitor with the CAS registration number 819075-40-6 and CatS.

from CatS (Chart 1d). Afterward, the imidazole ring of His164 polarizes a water molecule which hydrolyzes the thioester bond of the acyl-enzyme (Chart 1e,f). This leads to the release of the cleaved substrate (Chart 1g).<sup>2,9</sup>

Covalent and noncovalent inhibitors are known for cathepsins. Covalent inhibitors possess an electrophilic warhead group, e.g., an aldehyde, hydrazone, or nitrile moiety and bind via nucleophilic addition with the sulfur atom of the active site residue Cys25. The resulting bond can be formed irreversible or reversible. Covalent irreversible binding cathepsin inhibitors are chloromethyl ketones, diacylhydrazides, epoxysuccinyl derivatives, hydrazones, and vinyl sulfones.<sup>2,4,5,10–12</sup> In contrast to that, aldehydes, alkoxymethyl ketones, azepanones, diaminoketones, hydroxymethyl ketones, ketoamides, and nitriles represent cathepsin inhibitors that form a reversible covalent bond with the enzyme.<sup>5,10,13</sup> Noncovalent cysteine protease inhibitors possess more drug-like properties than covalent enzyme inactivators.<sup>14</sup> On that account, noncovalent cathepsin inhibitors like arylaminoethyl amides, arylaminoethyl carbamates, and ketobenzimidazole piperidines that inactivate the enzyme via hydrogen bond interactions were developed recently.<sup>7,15,16</sup>

Noncovalent as well as covalent CatS inhibitors form hydrogen bonds with residues Gln19, Gly69, Asn163, and Trp186 of the enzyme (Figure 1b).<sup>2</sup> Hydrophobic interactions are established between the S1', S1, S2, and S3 subsites of CatS and the corresponding aliphatic and aromatic parts of the inhibitor (Figure 1c). The S1' pocket is composed of

residues Ala140, Arg141, His142, Pro143, Phe146, Asn163, His164, and Trp186. The S1 subsite consists of mostly main chain atoms and is bordered by Cys22, Gly23, Cys66, Asn67, and Gly68. Residues Phe70, Met71, Gly137, Val162, and Phe211 form the large hydrophobic pocket S2. This pocket is of importance for ligand binding affinity and specificity over other cathepsins. Because of a hinge-like movement of residue Phe211 the pocket is able to accommodate small and large P2 moieties of the inhibitor. If the P2 moiety of the inhibitor represents a small structure such as a Leu side chain, Phe211 closes the entrance to the deeper part of the S2 pocket. In contrast to that, an inhibitor with a P2-Phe side chain leads to a conformational switch of Phe211 which opens the distal region of the subsite. Ultimately, the small S3 pocket comprises residues Gly62, Asn63, Lys64, Gly68, Gly69, and Phe70.<sup>2,7,12,17</sup>

With the release of human CatS X-ray crystal structures complexed with noncovalent inhibitors in the Brookhaven Protein Data Bank (PDB), a basis for structure-based pharmacophore modeling studies on these CatS inhibitors with favorable pharmacological properties is available.<sup>7,17,18</sup> Thus, we created structure-based pharmacophore models for these promising CatS inhibitors. Potent covalent binding cathepsin inhibitors like azepanones and ketoamides are in preclinical and even clinical development. Therefore, we also generated structure-based models for the established concept of covalent binding CatS inhibitors. In addition, ligand-based models for covalent and noncovalent CatS inhibitors were developed. The best models were used as virtual screening tools to select compounds from chemical databases for biological testing.

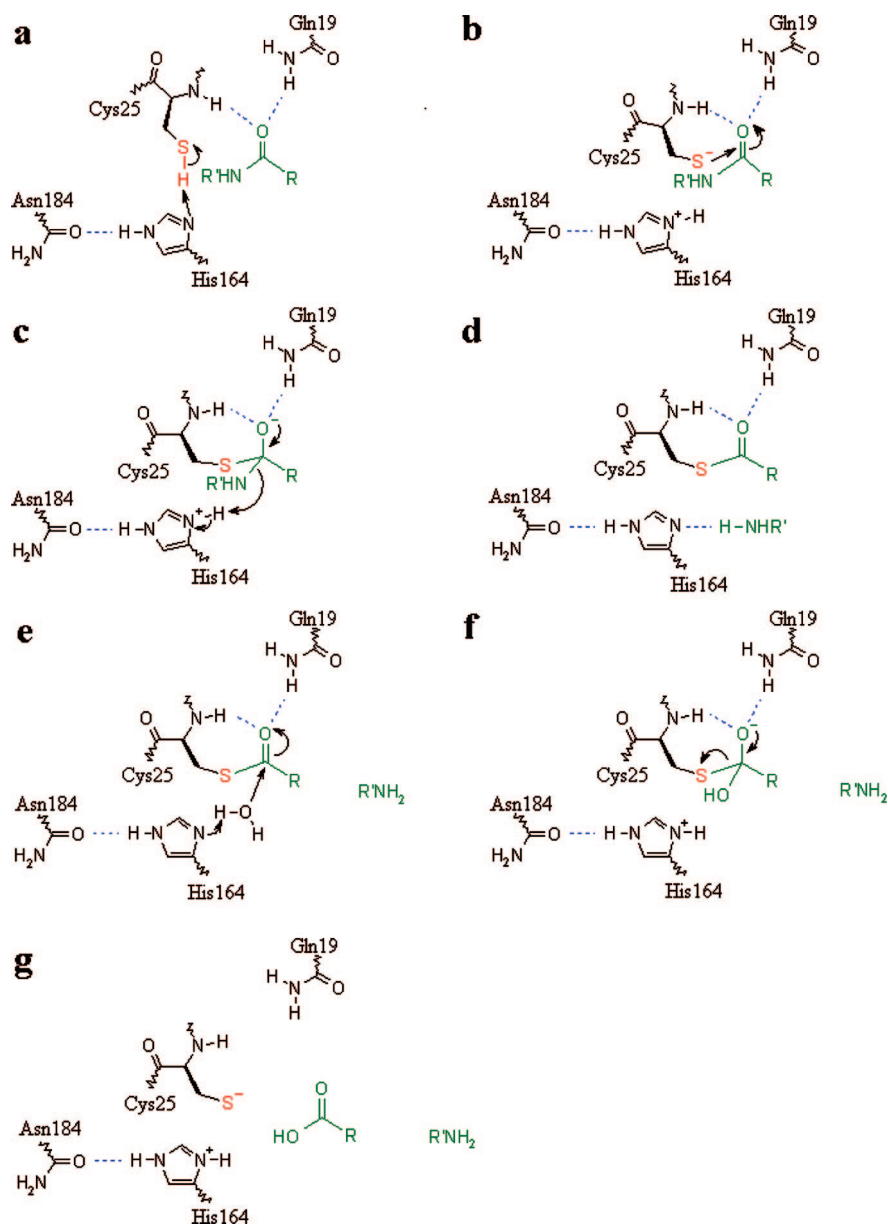
## RESULTS

**Workflow.** Figure 2 displays the virtual screening workflow which resulted in the discovery of novel scaffolds for CatS inhibitors.

**Ligand-Based Pharmacophore Modeling for Noncovalent CatS Inhibitors.** Several ligand-based models were created using the software package Catalyst<sup>19</sup> (see the Experimental Section). The best model was based on the compound with the CAS registration number 819076-13-6 (Table 1).

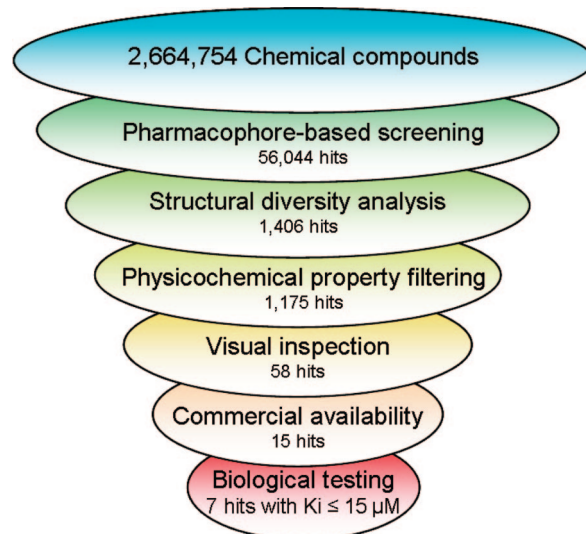
It consists of one hydrogen bond acceptor, one hydrogen bond donor, two ring aromatic features, two hydrophobic features, and one excluded volume sphere which represents an area a compound is not allowed to map. To improve the selectivity of the model, a shape derived from compound 819076-13-6 was added to the model. The final model was named *CatS noncovalent model* (Figure 3a) and validated by screening the noncovalent test set, the noncovalent focused library, and the Derwent World Drug Index 2005 (Derwent WDI).<sup>20</sup>

The noncovalent test set included ten structurally diverse and highly active noncovalent CatS inhibitors (see the Supporting Information). The noncovalent focused library was generated utilizing virtual combinatorial chemistry. Specified chemical fragments were combined using the software ilib:diverse<sup>21</sup> in order to obtain decoys which are putative inactive and structurally similar to CatS inhibitors. To ensure the structural similarity between these decoys and CatS inhibitors, a physicochemical property filter which

**Chart 1.** Reaction Mechanism of CatS<sup>a</sup>

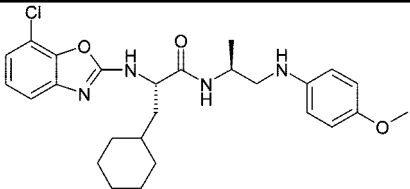
<sup>a</sup> The substrate is colored in green, the reactive thiol group of Cys25 is displayed in red, and hydrogen bonds are shown as blue dotted lines.

included limits for the molecular weight, the number of rotatable bonds, the number of heavy atoms, and the number of hydrogen bond acceptors and donors as well as for the number of oxygens, nitrogens, and halogens was applied (see the Experimental Section). Since the resulting decoys are structurally closely related to CatS inhibitors, some of these compounds could be biologically active. However, according to Kirchmair et al.<sup>22</sup> the probability for this is very low and statistically not significant. Therefore, this focused library was used to assess the ability of the model to discriminate between actives and decoys with similar physicochemical properties. The Derwent WDI is a library of development and marketed drugs. The model was also screened against this library to determine the discriminatory power of the model in a random drug-like compound library with a physicochemical property distribution dissimilar from CatS inhibitors. This may reproduce a virtual screening practice, where chemical databases containing structurally highly diverse and drug-like compounds are searched for new biological active ligands. Furthermore, the integration of

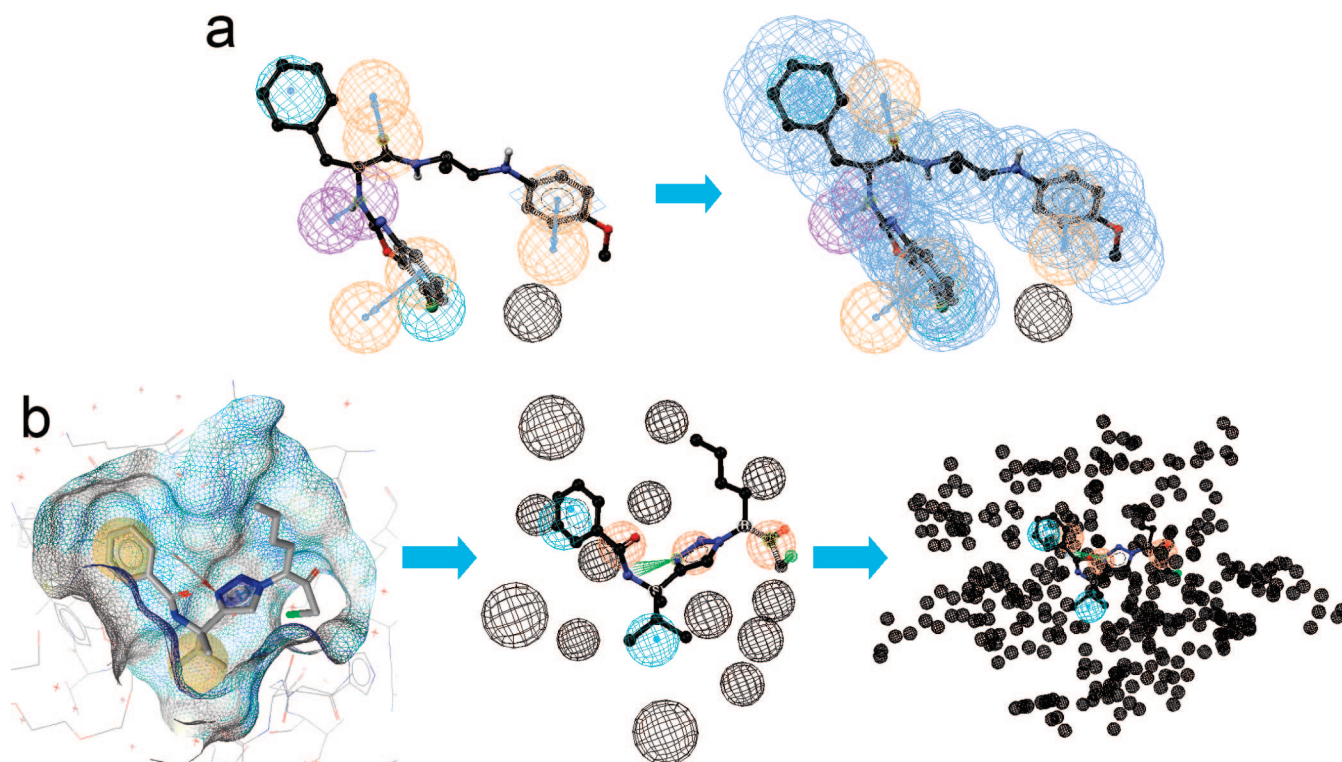
**Figure 2.** Virtual screening workflow.



**Table 1.** CatS Inhibitor Used for the Ligand-Based Generation of *CatS* Noncovalent Model<sup>a</sup>

pharmacophore model	ligand	structure
CatS noncovalent model	819076-13-6 <sup>#</sup>	 <p><math>K_i = 1 \text{ nM}^7</math></p>

<sup>a</sup> The <sup>#</sup> symbol denotes the CAS registration number.



**Figure 3.** Generation of CatS pharmacophore models. The generation of the ligand-based *CatS* noncovalent model using Catalyst (a). Hydrogen bond acceptors and  $\pi$ - $\pi$  interactions are shown as orange features, hydrogen bond donors are displayed as violet features, hydrophobic interactions are depicted in blue, excluded volume spheres are shown in black, and the shape feature is represented by a cloud of dark blue spheres. The structure-based *CatS* covalent model (b) was derived from the PDB entry 2h7j within LigandScout and converted into a Catalyst model which was merged with the EWF (single orange sphere) and refined by adding excluded volume spheres. In LigandScout arrows represent hydrogen bonds,  $\pi$ - $\pi$  interactions are shown as blue rings, hydrophobic interactions are displayed as yellow spheres, and the surface of the ligand-binding pocket is represented by a wireframe which is colored by lipophilicity.

active compounds in a random collection of drug-like molecules represents a common method for validating virtual screening techniques.<sup>23</sup> By using a focused library and a structurally diverse random library for validation, we tried to obtain pharmacophore models which enrich actives not only because they possess a binding site shape (e.g., excluded volume spheres) that makes it impossible for structurally more diverse compounds (e.g., inactive structures with a significantly higher molecular weight than actives) to fit to them but also because they represent the key ligand-protein interactions and therefore can discriminate between structurally similar actives and decoys. In other words, we avoided two extreme cases: (i) models which are more a binding site shape-based filter than an accurate description of the pharmacophore but could lead to significant enrichments in a structurally diverse chemical database with a widespread

physicochemical property distribution and (ii) models which enrich actives among structurally similar decoys but produce a hit list containing a lot of non-drug-like compounds, such as structures with a shape dissimilar from actives as a result of insufficient binding site restrictions. On that account, the enrichment factors (EFs) for both the focused library and the Derwent WDI were calculated to determine the best models. To calculate the EF for the noncovalent focused library, the compounds of the noncovalent test set were integrated into this library. Analogous, the Derwent WDI EF was determined. Five out of ten noncovalent test set compounds (50%) were able to match the model. When the 10,000 decoys were screened, 694 compounds (7%) were retrieved. Therefore, a noncovalent focused library EF of seven was calculated. A Derwent WDI screen resulted in 1,040 matching compounds which led to a Derwent WDI

**Table 2.** Validation Results for the CatS Pharmacophore Models<sup>a</sup>

pharmacophore model	focused library EF	Derwent WDI EF
CatS noncovalent model	7	32
CatS covalent model	7	160

<sup>a</sup> The focused library EF was calculated using the active compounds from the noncovalent or covalent test set and the decoys from the corresponding noncovalent or covalent focused library. Accordingly, the Derwent WDI was determined either utilizing the noncovalent or covalent test set compounds and the Derwent WDI compounds which were not described as CatS inhibitors.

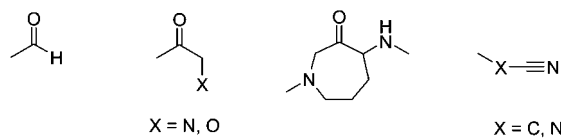
EF of 32 (Table 2, for more details see the Experimental Section). Because of these promising EFs, *CatS noncovalent model* was validated as the best ligand-based model for noncovalent CatS inhibitors.

**Structure-Based Pharmacophore Modeling for Non-covalent CatS Inhibitors.** For structure-based model generation for noncovalent CatS inhibitors, the crystallographic data of the two PDB complexes 2f1g and 2hh5 were analyzed within the software LigandScout<sup>24</sup> (see the Experimental Section). The generated models matched significantly fewer noncovalent test set compounds than *CatS noncovalent model*. None of the models derived from both PDB entries possessed focused library and Derwent WDI EFs comparable to *CatS noncovalent model*. Therefore, the ligand-based *CatS noncovalent model* was determined as the most selective model among all ligand-based and structure based models for noncovalent CatS inhibitors and was utilized for screening chemical databases in order to find new scaffolds for noncovalent CatS inhibitors.

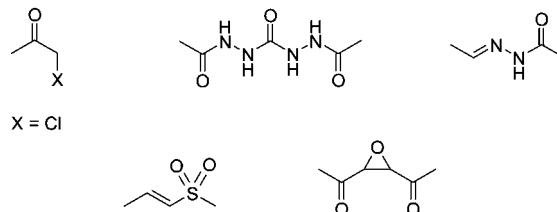
**Creation of an Electrophilic Warhead Function.** As mentioned above, covalent CatS inhibitors include an electrophilic warhead which is attacked by the nucleophilic sulfhydryl group of the active site residue Cys25.<sup>5</sup> Thus, a covalent bond is formed between inhibitor and protein. The reversibility or irreversibility of this nucleophilic addition depends on the properties of the electrophilic warhead. Catalyst supports the creation of user-adapted chemical features by defining sets of chemical substructures which should be matched by this function during virtual screening. To create such a Catalyst function for electrophilic warheads, the electrophilic groups of inhibitors of cysteine proteases, especially cathepsin K, L, and S, were investigated. According to the literature, the carbon atom of the reactive carbonyl group of aldehydes, alkoxymethyl ketones, azepanones, diaminoketones, hydroxymethyl ketones, and ketoamides can be reversibly added to the active site cysteine residue of cysteine proteases.<sup>2,10,13,25</sup> Moreover, the electrophilic carbon atom of nitriles is known to form a reversible thio-imidate bond with the sulfur atom of the active site cysteine residue.<sup>5</sup> In contrast to that, the electrophilic carbon atom of the carbonyl groups of chloromethyl ketones and diacylhydrazides binds irreversibly to the corresponding cysteine residue.<sup>11,12,26</sup> In addition, also epoxysuccinyls, hydrazones, and vinylsulphones form irreversible covalent bonds with cysteine proteases: (i) one of the electrophilic carbon atoms of epoxysuccinyls represents a target for an irreversible nucleophilic attack of cysteine proteases which leads to ring opening of the epoxy moiety and formation of a covalent bond to the sulfur atom of the cysteine residue,<sup>27</sup> (ii) irreversible thioaminal bonds are formed by nucleophilic addition of the active site cysteine residue to the electrophilic

**Chart 2.** Chemical Fragments Used for the Definition of EWF

Reversible inhibition



Irreversible inhibition



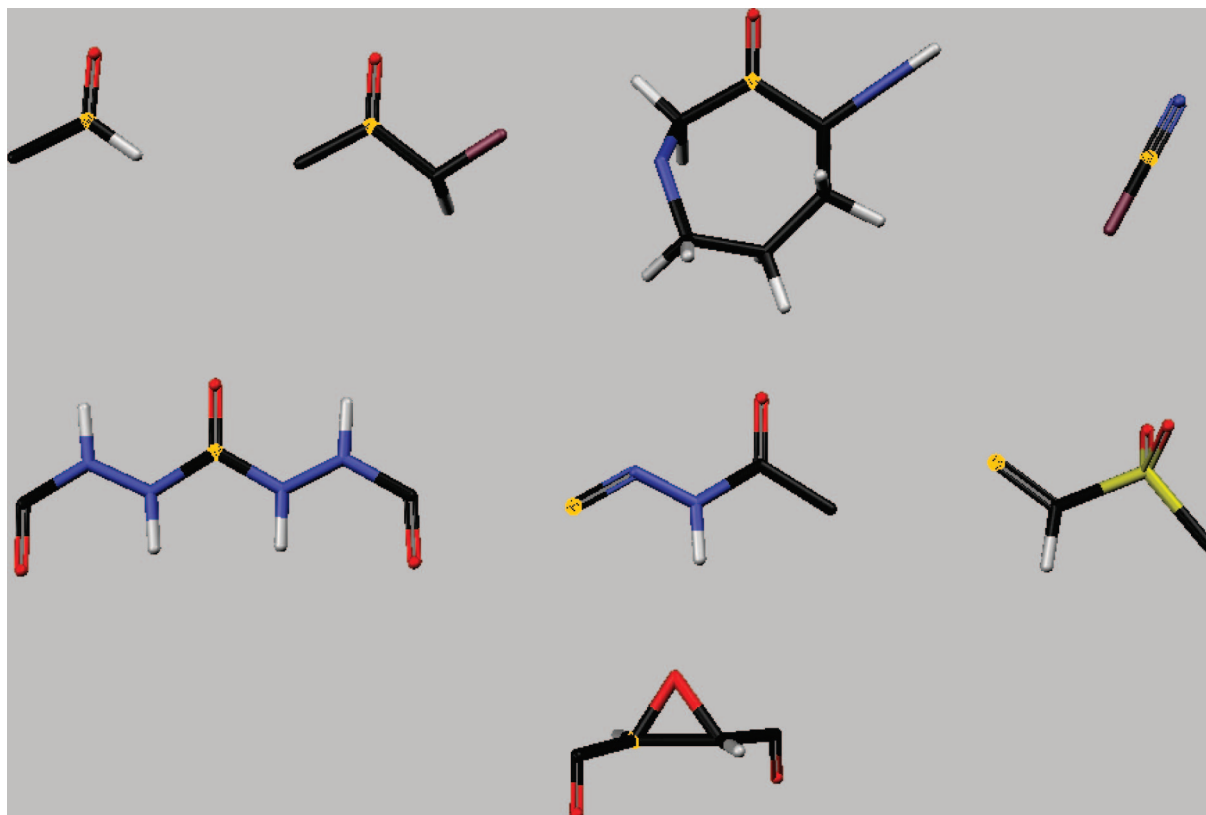
double bond carbon atom of hydrazone inhibitors,<sup>4</sup> and (iii) the double bond carbon atom of the vinyl moiety distal to the sulfonyl group of vinylsulphones is also attacked irreversibly by cysteine proteases.<sup>2</sup>

These reversible and irreversible acting electrophilic warheads were included into the definition of a new Catalyst function which was named electrophilic warhead function (EWF). The structures of these electrophilic chemical fragments are displayed in Chart 2.

The atom and bond properties of the chemical fragments were specified. Thereafter, the electrophilic carbon atom, which serves as target for the nucleophilic attack of the active site cysteine, was set for each fragment. These carbon atoms are used by Catalyst as centers in order to position the feature on the virtual screening compounds. The corresponding carbon atoms are shown as yellow points in Figure 4. Finally, the chemical structures were merged to one Catalyst function.

The resulting EWF was able to match all compounds of the covalent test set. 5,423 hits out of 10,000 decoys were retrieved when EWF was used as a query for a virtual screen of the covalent focused library. Thus, an EF of 1.8 was calculated for the EWF. In addition, a Derwent WDI screen was performed which resulted in 3,204 EWF matching compounds. The resulting Derwent WDI EF was 21. This means that utilizing the EWF without additional pharmacophore features for virtual screening would result in only 1.8 or 21 times more active compounds than a random selection from these compound databases. The best model for covalent CatS inhibitors—*CatS covalent model*—received a four times higher focused library EF (7) and an eight times higher Derwent WDI EF (160). Taking this into consideration, a simple substructure search is not enough for enriching a sufficient small set of compounds which can be considered for biological activity testing. Additional features merged with the EWF significantly increase the virtual screening performance. Therefore, the EWF is more of an additional restriction for CatS pharmacophore models that excludes compounds that do not contain a suitable electrophilic warhead for a nucleophilic attack of CatS rather than a stand-alone virtual screening technique for discovering new CatS inhibitors.

**Structure-Based Pharmacophore Modeling for Covalent CatS Inhibitors.** The PDB complexes 1ms6, 1npz, 1nqc, and 2h7j were utilized for generating structure-based models for covalent binding CatS inhibitors. The most selective model was derived from the PDB entry 2h7j which



**Figure 4.** Feature positioning centers of the chemical fragments utilized for generation of EWF. The carbon atoms which function as origin to position the feature on virtual screening compounds are marked as yellow points.

**Table 3.** Covalent Binding CatS Inhibitor Used for the Generation of the Structure-Based *CatS Covalent Model*<sup>a</sup>

pharmacophore model	PDB entry	ligand	structure
CatS covalent model	2h7j	913981-83-6 <sup>#</sup>	<p><math>K_i = 15 \text{ nM}^{13}</math></p>

<sup>a</sup> The # symbol denotes the CAS registration number.

contains the crystal structure of a complex of CatS and the covalent inhibitor with the CAS registration number 913981-83-6 (Table 3). The initial LigandScout model was imported into Catalyst for database screening. It included a hydrogen bond acceptor formed between the third nitrogen of the ligand's 1,2,3-triazole ring and residue Gly69 and another hydrogen bond acceptor located between the chloromethylketone group of the inhibitor and Gln19. Moreover, the model comprised two hydrophobic features situated at the phenyl ring and the isopropyl moiety of the ligand. The hydrogen bond acceptor at the chloromethylketone group was replaced by the EWF to represent the covalent binding electrophilic group. Afterward, excluded volume spheres representing the heavy atoms of the binding site residues as detected by LigandScout were generated. This binding site

shape replaced the small number of initially by LigandScout derived excluded volume spheres. Thereafter, a shape feature derived from the bioactive conformation of CatS inhibitor 913981-83-6 was generated and merged with the model. However, the focused library and Derwent WDI EFs for the model without the shape were significantly higher than for this merged model. Therefore, the model without the shape was selected. This model was named *CatS covalent model* (Figure 3b). 31 of the 35 covalent test set compounds (89%) matched this model. 1,197 decoys from the covalent focused library (12%) were retrieved which led to a focused library EF of seven. A Derwent WDI search resulted in 340 matching compounds (0.5%) and a Derwent WDI EF of 160 (Table 2). Taking these two EFs into account, *CatS covalent*



*model* was validated as the best virtual screening tool for discovering new scaffolds for covalent CatS inhibitors.

**Virtual Screening of Chemical Databases.** The best model for CatS inhibitors which forms only noncovalent interactions with the protein—*CatS noncovalent model*—and the most selective model for covalent binding CatS inhibitors—*CatS covalent model*—were utilized to screen 18 chemical databases virtually within Catalyst. In other words, a total of 2,664,754 compounds were searched for new scaffolds for noncovalent and covalent CatS inhibitors. The hits were scored using the Best Fit value calculated by Catalyst (see the Experimental Section). Since a Best Fit value higher than 2 was determined for most of the noncovalent and covalent test set compounds that matched the models, this value was used as the cutoff for the virtual screening results. The resulting pharmacophore-based virtual screening hit list included 56,044 compounds.

Afterward, the Pipeline Pilot<sup>28</sup> script based on Tanimoto similarity analysis which was used for the determination of the structural diversity of the two test sets, the two focused libraries, and the Derwent WDI (see the Experimental Section), was executed in order to exclude structurally similar compounds from biological testing. For the pharmacophore-based hits, 1,406 clusters were generated when the maximum allowed dissimilarity was set to 0.7. The compounds representing the cluster centers were selected for the physicochemical property filtering step of the workflow.

For this purpose, a Pipeline Pilot script including a drug-like filter was executed. The drug-like filter criteria corresponded in general with Lipinski's rule of five except for a limit of 600 Da for the molecular weight instead of 500 Da and an AlogP value of six instead of five. Since 15 out of the 45 noncovalent and covalent test set compounds (33%) have a molecular weight slightly greater than 500 Da and since for seven out of 45 structures an AlogP value greater than 5 was determined, these two filter criteria were modified with respect to the physicochemical property distribution of the test set compounds. 1,175 compounds matched these filter criteria.

Thereafter, the structures of the compounds were visually inspected in terms of chemical stability, e.g., compounds containing chemical groups in their main chain that could be hydrolyzed (such as ester moieties) were excluded. This workflow step resulted in 94 compounds which were aligned with the corresponding CatS model within Catalyst in order to visually inspect if the compounds were fitted correctly to the model. For example, compounds on which a hydrogen bond feature was wrongly placed on a  $sp^3$ -hybridized ester oxygen that usually does not accept a hydrogen bond were determined as false positives and discarded.<sup>29,30</sup> The remaining 58 compounds were subjected to a SciFinder database<sup>31</sup> search to explore which of the compounds have been reported as CatS inhibitors which was also an exclusion criterion. None of them was described as a CatS inhibitor in the literature. Finally, with respect to their commercial availability 15 compounds were selected.

**Structural Novelty of Compounds Selected for Biological Testing.** Using the Pipeline Pilot script mentioned above, we also investigated the structural distance of the 15 selected compounds to the 45 noncovalent and covalent CatS inhibitors of the two test sets. Thus, we avoided the testing of

compounds which are structurally diverse but do not represent novel scaffolds for CatS inhibitors.

At a maximum distance of 0.7 singleton clusters for compounds **1**, **4**, **5**, **7**, **9**, **11–13**, and **15** were created, whereas six compounds were assigned to clusters containing test set inhibitors. When the maximum allowed dissimilarity was reduced to 0.5, the Pipeline Pilot script generated for each of the 15 compounds a singleton cluster.

Afterward, we performed a SciFinder similarity search using a Tanimoto similarity score cutoff  $\geq 80$  to further check for similar cathepsin inhibitors reported in the literature. Only for compound **2**, structures with reported inhibitory activity on different cathepsin subtypes were retrieved. Moreover, a more restrictive Tanimoto similarity score cutoff  $\geq 90$ , resulted in two patented cathepsin L (CatL) inhibitors and one structure which was patented as cysteine protease inhibitor.<sup>32,33</sup> However, none of these three inhibitors was reported as CatS inhibitor. Therefore, we also subjected compound **2** to the biological testing.

**Biological Evaluation.** The 15 compounds obtained utilizing our virtual screening approach were investigated in kinetic evaluation studies as described in the Experimental Section. The results of these kinetic studies expressed as kinetic constants  $K_i$  are shown in Table 4.

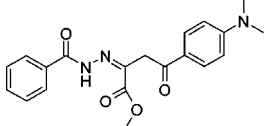
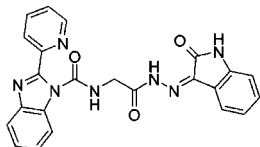
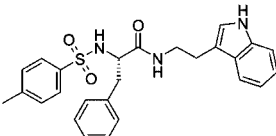
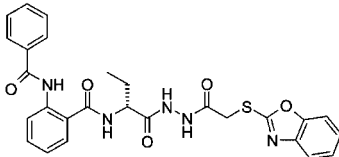
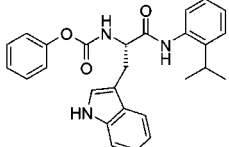
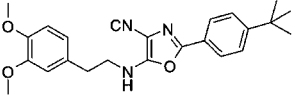
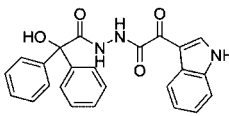
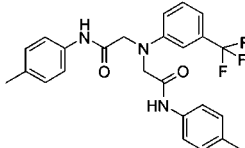
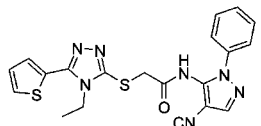
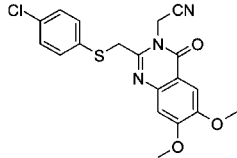
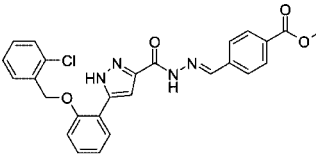
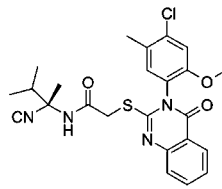
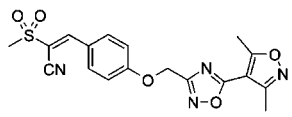
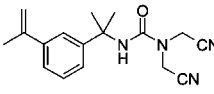
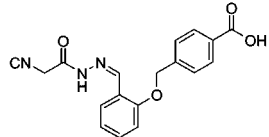
Spectrofluorimetric analysis of the 15 compounds showed that 13 of them possess  $K_i$  values in the micromolar range. From these 13 compounds, three compounds were predicted based on the *CatS noncovalent model*, whereas ten compounds fitted excellently to *CatS covalent model*. The two compounds with kinetic constants above the micromolar range were predicted based on *CatS covalent model* (compounds **8** and **11**). Two out of three compounds (66%, compounds **2** and **12**) which were selected utilizing *CatS noncovalent model*, possess  $K_i$  values equal or below 15  $\mu\text{M}$ . Five out of twelve compounds (42%, compounds **6**, **7**, **9**, **10**, **14**) predicted based on *CatS covalent model* achieved kinetic constants in this low micromolar range.

## DISCUSSION

In this study, we generated several ligand-based and structure-based pharmacophore models for noncovalent and covalent CatS inhibitors in order to screen a total of 2,664,754 compounds from 18 different chemical databases to discover new scaffolds for CatS inhibitors. Because of their outstanding focused library and Derwent WDI EFs, *CatS noncovalent model* and *CatS covalent model* were determined as our best pharmacophore models for noncovalent and covalent CatS inhibitors, respectively. Therefore, the two models were chosen for virtual screening.

The resulting virtual screening hits were investigated using a drug-like filter and visual inspection. The remaining 15 compounds were subjected to a structural similarity analysis to determine their structural distance to the test set compounds. Thus, we tried to avoid the testing of compounds including scaffolds structurally similar to known CatS inhibitors. Since the 15 compounds possessed a sufficient structural distance to the two test sets, they were biologically tested on recombinant CatS expressed in *E. coli* cells. Thirteen of the 15 compounds (87%) possess micromolar  $K_i$  values. Although, the pharmacophore models were generated and validated using highly potent CatS inhibitors, none

**Table 4.** Results of the Kinetic Evaluation of the 15 Compounds Which Were Predicted As CatS Inhibitors Based on Our Virtual Screening Approach<sup>a</sup>

compound	CatS $K_i$ ( $\mu\text{M}$ ) <sup>a</sup>	compound	CatS $K_i$ ( $\mu\text{M}$ ) <sup>a</sup>
	16.63 $\pm$ 3.87		15.27 $\pm$ 0.06
	2.98 $\pm$ 0.96		8.89 $\pm$ 1.21
	19.48 $\pm$ 2.55		146317.8
	56.86 $\pm$ 25.04		14.58 $\pm$ 3.15
	198.25 $\pm$ 44.11		28.17 $\pm$ 7.07
	2.78 $\pm$ 0.05		10.90 $\pm$ 3.34
	8.44 $\pm$ 0.51		437.74 $\pm$ 110.33
	3389.73		

<sup>a</sup> Human recombinant CatS expressed in *E. coli* cells was obtained from Merck Biosciences. For kinetic evaluation of the CatS inhibitors, the hydrolysis of CatS substrate Cbz-Val-Val-Arg-NHMec was determined by spectrofluorimetric analysis. The  $K_i$  values were calculated using nonlinear regression analysis. For more details about the assay conditions see the Experimental Section paragraph.



of the 15 compounds was biologically active in the nanomolar range. However, the EFs determined for the pharmacophore models show that the models are able to enrich CatS inhibitors with nanomolar  $K_i$  values among a large number of decoys. Therefore, the absence of a nanomolar active CatS inhibitor among the 15 compounds is rather related to the limited chemical space of the chemical databases which were virtually screened than to a disability of the virtual screening workflow to retrieve highly potent CatS inhibitors. For seven compounds (47%) the kinetic constant  $K_i$  was determined to be equal or lower to 15  $\mu\text{M}$ . Biological assaying of random compound libraries is supposed to result in a far lower hit rate of 0.1%.<sup>34</sup> Since 18 commercial screening compound databases and not focused libraries of compounds structurally related to cysteine protease inhibitors were virtually screened, the high enrichment of CatS inhibitors among the 15 compounds is more related to the discriminatory power of our pharmacophore-based virtual screening workflow than to random. Two out of the seven compounds with low micromolar activity (29%, compounds **2** and **12**) were predicted based on *CatS noncovalent model*. For compound **2** ( $K_i = 3.0 \mu\text{M}$ ) a SciFinder similarity search resulted in one cysteine protease and two CatL inhibitors that possessed a nearly identical structure to compound **2**, except for an additional aldehyde or methyl alcohol moiety. Therefore, compound **2** seems to be a scaffold not only for CatL inhibitors but also for inhibitors of CatS. In contrast to that, for compound **12** ( $K_i = 14.6 \mu\text{M}$ ) no structurally related compounds with cathepsin activity have been reported. Thus, compound **12** represents a completely new and therefore more interesting scaffold for CatS inhibitors that do not contain an electrophilic warhead than compound **2**.

Five of the compounds with a  $K_i$  value equal or lower to 15  $\mu\text{M}$  (71%, compounds **6**, **7**, **9**, **10**, and **14**) matched *CatS covalent model* and therefore possess an electrophilic warhead. Compound **6** has the highest binding affinity ( $K_i = 2.8 \mu\text{M}$ ) and is the most promising structure retrieved by *CatS covalent model*.

Structural optimization of the most interesting compounds, especially of compounds **6** and **12** and a more detailed biological characterization in terms of cathepsin subtype selectivity of the resulting compounds, will be part of future studies.

## CONCLUSION

A virtual screening workflow based on our two best pharmacophore models for CatS inhibitors resulted in the selection of 15 out of 2,664,754 chemical database compounds. The 15 compounds were biologically tested on human recombinant CatS expressed in *E. coli* cells. Seven out of these 15 compounds (47%) possess a  $K_i$  value equal or lower to 15  $\mu\text{M}$ . In particular, compound **6** which is the structure with the highest binding affinity among the seven compounds and compound **12** which does not contain an electrophilic warhead as known for the majority of published CatS inhibitors represent promising novel scaffolds for CatS inhibitors. These positive biological results confirm that our pharmacophore-based virtual high-throughput screening strategy allows the identification of novel scaffolds for CatS inhibitors, which can be further chemically modified in order to develop potential drugs for the therapy of autoimmune disorders and atherosclerosis.

**Table 5.** Similarity Analysis Results for the Two Test Sets, the Two Focused Libraries, and the Derwent WDI<sup>a</sup>

compound set	number of clusters	
	maximum distance 0.7	maximum distance 0.5
covalent test set	11	21
noncovalent test set	4	7
covalent focused library	279	9509
noncovalent focused library	250	8732
Derwent WDI	4534	25334

<sup>a</sup> A larger number of clusters correlates with a higher structural diversity of the compound set.

## EXPERIMENTAL SECTION

**Hardware Specifications.** Molecular modeling studies were carried out on an Intel Pentium Core 2 Duo 6400 equipped with 1 GB RAM running Linux Fedora Core 6.

**Software Specifications.** For this study the following software was applied: Catalyst 4.11 for ligand-based pharmacophore modeling and virtual screening of chemical databases, LigandScout 2.0 for structure-based pharmacophore modeling, and PipelinePilot 5.0.1.100 for structural similarity analysis and physicochemical property filtering.

**Generation of Compound Sets.** The Catalyst software package was utilized to build and energetically minimize structures of CatS inhibitors derived from literature (see the Supporting Information references). The conformational models for the compounds were created using catConf—the conformer generator of Catalyst—with the following settings: maximum number of conformers = 100, generation type = fast quality, and energy range = 20 kcal/mol above the calculated lowest energy conformation.

Several highly active and structurally diverse CatS inhibitors were used for generation of ligand-based models.

For validation of structure-based and ligand-based models, two test sets were generated containing active CatS inhibitors: 35 covalent CatS inhibitors were forming the covalent test set, whereas the noncovalent test set was composed of ten noncovalent CatS inhibitors (see the Supporting Information table).

A Pipeline Pilot script that clusters molecules based on Tanimoto similarities between a set of descriptors was used to determine the structural diversity of the compounds of the two test sets. As a set of descriptors Scitegic's method to calculate structural fingerprints—the extended connectivity fingerprints (ECFP) as implemented in the Pipeline Pilot software—was selected. The ECFP method characterizes molecules by indexing the environment of every atom of a compound. For this purpose, up to 4 billion different structural features are utilized. The setting ECFP\_6 was applied to analyze the structural diversity. This setting considers all neighbor atoms within a diameter of six bonds for the feature calculation of each atom.<sup>35</sup> For the investigation of the covalent and the noncovalent test set, the maximum allowed dissimilarities were set to 0.7 and 0.5, respectively. Structurally diverse compounds are clustered if the maximum distance is set to 0.7. In contrast to that, clusters of structurally more related compounds are formed using the more restrictive value of 0.5. The results of the similarity analysis indicated a sufficient structural diversity for both test sets (Table 5).

**Table 6.** Filter Properties for the Generation of the Covalent and Noncovalent Focused Library

property	covalent focused library				noncovalent focused library			
	standard distribution		limits		standard distribution		limits	
	mean	variance	min.	max.	mean	variance	min.	max.
molecular weight	475	7707			512	5388		
rotatable bonds	13	8			12	7		
heavy atoms	34	41			36	19		
hydrogen bond acceptors			3	8			2	5
hydrogen bond donors			0	6			0	4
nitrogens			2	6			3	6
oxygens			0	7			2	4
halogens			0	4			0	4

**Generation of Focused Libraries.** Only a few compounds which are structurally similar to CatS inhibitors but do not show biological activity were found in the literature. A validation of the models using the active test set compounds and a small number of these CatS inactives would not represent virtual screening practice, where the fraction of active molecules is far smaller than the fraction of inactives. Therefore, two focused libraries containing decoys were created using the Java-based software *ilib:diverse*. The so-called covalent focused library contained decoys that mimicked covalent CatS inhibitors, whereas the noncovalent focused library included decoys that corresponded to the structures of noncovalent CatS inhibitors. For the generation of both focused libraries, organic acids, alcohols, aliphatics, amides and imides, amines, amino acids, benzenes, carbocycles, ethers, esters, halogens, heterocycles, ketones and aldehydes, and sulfur containing groups were selected from the *ilib:diverse* chemical fragment database. Additional chemical fragments were used for the generation of the covalent focused library: azepanones, diamino ketones, hydrazones, hydroxymethyl ketones, ketoamides, nitriles, and vinyl sulfones. These fragments represent the electrophilic warheads of covalent CatS inhibitors. They were not included in the original *ilib:diverse* chemical fragment database. Therefore, these substructures were imported into *ilib:diverse* as MDL mol-files and added to the set of fragments which was selected for the creation of the covalent focused library. To increase the structural similarity between the decoys and the corresponding CatS inhibitors, the physicochemical properties of the CatS inhibitors from the covalent and noncovalent test set were investigated and consequently used as a filter during the focused library generation process. On that account, standard distributions of the molecular weight, the number of rotatable bonds, and the number of heavy atoms were utilized as filter properties. Moreover, limits for the number of hydrogen bond acceptors and donors as well as limits for the number of nitrogens, oxygens, and halogens served as filter criteria. Finally, the *ilib:diverse* reactivity filter was selected to reject unstable compounds. The filter properties which were used to create the covalent and noncovalent focused library are shown in Table 6.

The application of these drug-likeness filters and user-defined sets of chemical fragments led to the generation of two focused libraries each containing 10,000 decoys which possessed highly comparable structural characteristics to covalent and noncovalent CatS inhibitors, respectively. The two focused libraries were exported as MDL sd-files. Generation of 3D coordinates and energy minimization of the focused library structures utilizing the Clean force field

was performed by executing a Pipeline Pilot script. Afterward, the conformational models of the decoys were generated using the conformer generator of Catalyst with the same settings as mentioned for the generation of the training and test sets (100 conformers, fast generation, energy range 20 kcal/mol).

The structural diversity of the two focused libraries was analyzed by executing the Pipeline Pilot script that was applied for the structural dissimilarity investigation of the covalent and noncovalent test sets. The results are displayed in Table 5 and show that the two focused libraries comprise a widespread variety of decoys structurally related to CatS inhibitors.

The two focused libraries were screened using our CatS pharmacophore models as query in order to assess the number of decoys that match the models and consequently to calculate the focused library EF for these models. For EF calculations, the covalent test set compounds were integrated as actives subset into the covalent focused libraries, and the noncovalent test set compounds were included in the noncovalent focused library. The EF was calculated using the following equation<sup>36–38</sup>

$$EF = \frac{TP/n}{A/N} \quad (1)$$

where EF is the enrichment factor, TP is the number of CatS actives matched by the CatS model, n is the number of CatS active and inactive compounds matched by the CatS model, A is the number of CatS actives in the library, and N is the number of all CatS compounds in the library. The resulting EF describes for how many times the virtual screening technique performs better in enriching active compounds in a hit list compared to a random-based compound selection.<sup>37,39</sup>

**Generation of a Random Library.** Analogous to the generation of the training and test sets, Catalyst was used for energetical minimization and conformational model generation of the 65,070 Derwent WDI compounds (100 conformers, fast generation, energy range 20 kcal/mol). Thereafter, the Pipeline Pilot script mentioned above was applied to determine the structural diversity of the random drug-like library. The results depicted in Table 5 show that the Derwent WDI represents a structural diverse compound library. Before the noncovalent and covalent test sets were generated, the Derwent WDI was examined for known CatS inhibitors. Two compounds were found which were mentioned as CatS inhibitors in the literature: (i) JNJ 10329670 represents a noncovalent CatS inhibitor with an IC<sub>50</sub> value of 100 nM<sup>15</sup> and (ii) paecilopeptin inhibits 50% of CatS activity at a concentration of 2.1 nM.<sup>40</sup> Thus, JNJ 10329670

was added to the noncovalent test set, whereas paecilopeptin was assigned to the covalent test set. In order to calculate the Derwent WDI EF of our noncovalent and covalent CatS models, the corresponding test sets including the two Derwent WDI CatS inhibitors, served as actives subset (10 noncovalent and 35 covalent CatS inhibitors), whereas the remaining 65,068 Derwent WDI structures were used as subset of random drug-like structures with no reported biological activity on CatS.

The two focused libraries of decoys served as an objective tool for validating the ability of the CatS models to differ between structural similar decoys and actives, whereas the Derwent WDI was used to assess the screening performance of our CatS models if they are screened against structurally diverse chemical databases. Both the focused library EF and the Derwent WDI EF were used for model validation.

**Structure-Based Pharmacophore Modeling.** For examination of seven CatS inhibitor-protein complexes from the PDB, the software LigandScout was utilized. The software analyzes all the possible ligand-protein interactions and automatically generates a pharmacophore model based on this information. The resulting model was imported into Catalyst for refinement and database screening. To refine the model, excluded volume spheres were derived from all heavy atoms of the surrounding amino acids of the binding site and merged with the model. This binding site shape replaced the original excluded volume spheres generated by LigandScout. In order to further increase the selectivity of the model, a shape feature which represents the spatial information of an active compound fitted to the model, was added. To validate the usability of the final model for CatS inhibitor discovery, the corresponding focused library and the Derwent WDI were screened.

**Ligand-Based Pharmacophore Modeling.** Several highly active CatS inhibitors derived from literature (see the Supporting Information) were used to create ligand-based models. For this purpose, Catalyst was utilized to identify chemical functionalities and to place pharmacophore model features on the corresponding chemical functions of the compounds. CatS inactives found in the literature (see the Supporting Information) were fitted to the generated models in order to determine the areas active compounds are not allowed to map. Consequently, excluded volume spheres were added to the models. The resulting models were further improved within Catalyst by adding a shape feature. According to ref 41, structure-based models provide a better description of the key ligand-protein interactions than ligand-based models. Therefore, only ligand-based models with significant higher focused library and Derwent WDI EFs than structure-based models were used for screening chemical databases to find new scaffolds for CatS inhibitors.

**Database Screening Using Catalyst.** Catalyst was utilized for 3D database searching. The Catalyst parameters were set as follows: (i) compare.minInterBlobDistance—the minimum distance between features parameter—was set to 0 which allows tighter feature placements on a compound, (ii) the FunctionMapping.Hydrophobe.Neighbor.numBondLessEqualFromOWithDoubleBond parameter was also set to 0 which makes it possible to fit hydrophobic features to phenyl sulfonamide moieties, and (iii) a value of 10,000 was applied for the ViewDatabase.max-Hits parameter which means that a maximum number of 10,000 hits can be retrieved by

database screening. A fluorine hydrogen bond donor feature was added to the default Catalyst feature dictionary as described by Wolber et al.<sup>24</sup> Moreover, our customized Catalyst function, which describes the substructures of the electrophilic warheads of covalent binding CatS inhibitors—the EWF—was added to the dictionary. All the other Catalyst settings were kept default. Only compounds were retrieved as hits that matched all features of the pharmacophore model which was used for virtual screening. The Best Flexible Search algorithm of Catalyst was applied for database screening. In contrast to the Fast Flexible Search, which only uses the existing conformational model of the screening compounds to fit them to the pharmacophore model, the Best Flexible search algorithm performs an energy minimization of the compound conformers and therefore offers a more comprehensive coverage of the conformational space of the screening compounds during database search.

Within Catalyst virtual screening hits can be ranked according to their fitting to the pharmacophore model. For this purpose, fit values which are based on the feature weight and on the mapping of the pharmacophore feature on the corresponding chemical group of the compound are calculated. For our study, the feature weight was set to the default value of 1. Thus, a fit value of 1 is only assigned to a pharmacophore feature if its feature center is directly placed on the chemical group. In contrast to that, a fit value between 0 and 1 corresponds to a pharmacophore feature center that does not exactly match the chemical group. If all single pharmacophore feature fit values are added, the fit value for the alignment of the compound to the pharmacophore model is obtained. With respect to the Fast and Best Flexible Search algorithm, Catalyst allows the calculation of Fast Fit and Best Fit values. Thus, we used Best Fit values as cutoff for our pharmacophore-based virtual screening hit list.<sup>19</sup>

**Kinetic Evaluation of Inhibitors.** Human recombinant CatS expressed in *E. coli* cells was purchased from Merck Biosciences. Dithiothreitol (DTT), 2-[N-morpholino]ethanesulfonic acid (MES), and sodium acetate were purchased from Sigma (Poole, U.K.). Disodium ethylenediamine tetra acetate (EDTA) was obtained from Gibco-BRL (Paisley, U.K.). Cbz-Val-Val-Arg-NHMec was supplied by Bachem Ltd. (Bubendorf, Switzerland). Candidate compounds were purchased from different companies: Compounds **1** and **2** were obtained from ChemBridge Corporation (San Diego, U.S.A.); compounds **3** and **4** were supplied by Asinex Ltd. (Moscow, RU); compounds **5** and **7** were purchased from Maybridge (Cornwall, U.K.); compound **6** was obtained from Pharmeks Ltd. (Moscow, RU); compound **8** was supplied by Vitas-M Laboratory Ltd. (Moscow, RU); compounds **9**, **10**, and **11** were purchased from Interbioscreen Ltd. (Moscow, RU); compounds **12**, **13**, and **14** were obtained from Enamine Ltd. (Kiew, UA); and compound **15** was purchased from Tripos Discovery Research (Cornwall, U.K.). No experimental tests were performed to confirm the purity or identity of the candidate compounds obtained from the companies. Human recombinant CatS was activated in 100 mM sodium acetate pH 5.5 containing 1 mM EDTA and 2 mM DTT for 30 min at 37 °C. CatS was then added to 150  $\mu$ L assay buffer (0.5 M MES pH 6.0 containing 2 mM DTT) and the inhibitors under study (25, 50, 100 and 200  $\mu$ M). CatS substrate, Cbz-Val-Val-Arg-NHMec (25  $\mu$ L to each



well; final concentration 50  $\mu$ M), was added, and the rate of substrate hydrolysis at 37 °C was monitored continuously over a period of 60 min by measuring the rate of increase of fluorescence (at excitation  $395 \pm 25$  nm, emission  $460 \pm 40$  nm) in a CYTOFLUOR Multiwell plate reader Series 4000 spectrofluorimeter. The  $K_i$  values were calculated using nonlinear regression analysis according to Cha et al. (1975) and Morrison et al. (1982).<sup>42,43</sup>

#### ACKNOWLEDGMENT

We thank Dr. Rémy D. Hoffmann, Accelrys SARL Paris, for screening the Derwent WDI database.

**Supporting Information Available:** Structures of the CatS inhibitors and CatS inactives used for ligand-based pharmacophore modeling as well as structures of the ten noncovalent and 35 covalent test set compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### REFERENCES AND NOTES

- (1) Taleb, S.; Cancellato, R.; Clement, K.; Lacasa, D. Cathepsin S promotes human preadipocyte differentiation: possible involvement of fibronectin degradation. *Endocrinology* **2006**, *147*, 4950–4959.
- (2) Pauly, T. A.; Sulea, T.; Ammirati, M.; Sivaraman, J.; Danley, D. E.; Griffor, M. C.; Kamath, A. V.; Wang, I. K.; Laird, E. R.; Seddon, A. P.; Menard, R.; Cygler, M.; Rath, V. L. Specificity determinants of human cathepsin S revealed by crystal structures of complexes. *Biochemistry* **2003**, *42*, 3203–3213.
- (3) Taleb, S.; Lacasa, D.; Bastard, J.-P.; Poitou, C.; Cancellato, R.; Pelloux, V.; Viguerie, N.; Benis, A.; Zucker, J.-D.; Bouillot, J.-L.; Coussieu, C.; Basdevant, A.; Langin, D.; Clement, K. Cathepsin S a novel biomarker of adiposity: relevance to atherogenesis. *FASEB J.* **2005**, *19*, 1540–1542.
- (4) Cywin, C. L.; Firestone, R. A.; McNeil, D. W.; Grygon, C. A.; Crane, K. M.; White, D. M.; Kinkade, P. R.; Hopkins, J. L.; Davidson, W.; Labadia, M. E.; Wildeson, J.; Morelock, M. M.; Peterson, J. D.; Raymond, E. L.; Brown, M. L.; Spero, D. M. The design of potent hydrazones and disulfides as cathepsin S inhibitors. *Bioorg. Med. Chem.* **2003**, *11*, 733–740.
- (5) Ward, Y. D.; Thomson, D. S.; Frye, L. L.; Cywin, C. L.; Morwick, T.; Emmanuel, M. J.; Zindell, R.; McNeil, D.; Bekkali, Y.; Girardot, M.; Hrapchak, M.; DeTuri, M.; Crane, K.; White, D.; Pav, S.; Wang, Y.; Hao, M.-H.; Grygon, C. A.; Labadia, M. E.; Freeman, D. M.; Davidson, W.; Hopkins, J. L.; Brown, M. L.; Spero, D. M. Design and synthesis of dipeptide nitriles as reversible and potent cathepsin S inhibitors. *J. Med. Chem.* **2002**, *45*, 5471–5482.
- (6) Liu, J.; Ma, L.; Yang, J.; Ren, A.; Sun, Z.; Yan, G.; Sun, J.; Fu, H.; Xu, W.; Hu, C.; Shi, G.-P. Increased serum cathepsin S in patients with atherosclerosis and diabetes. *Atherosclerosis* **2006**, *186*, 411–419.
- (7) Tully, D. C.; Liu, H.; Alper, P. B.; Chatterjee, A. K.; Epple, R.; Roberts, M. J.; Williams, J. A.; Nguyen, K. T.; Woodmansee, D. H.; Tumanut, C.; Li, J.; Spraggon, G.; Chang, J.; Tuntland, T.; Harris, J. L.; Karanewsky, D. S. Synthesis and evaluation of arylaminoethyl amides as noncovalent inhibitors of cathepsin S. Part 3: Heterocyclic P3. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1975–1980.
- (8) McGrath, M. E.; Palmer, J. T.; Bromme, D.; Somoza, J. R. Crystal structure of human cathepsin S. *Protein Sci.* **1998**, *7*, 1294–1302.
- (9) Rzychon, M.; Chmiel, D.; Stec-Niemczyk, J. Modes of inhibition of cysteine proteases. *Acta Biochim. Pol.* **2004**, *51*, 861–873.
- (10) Mendonca, R. V.; Venkatraman, S.; Palmer, J. T. Novel route to the synthesis of peptides containing 2-amino-1'-hydroxymethyl ketones and their application as cathepsin K inhibitors. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2887–2891.
- (11) Thompson, S. K.; Smith, W. W.; Zhao, B.; Halbert, S. M.; Tomaszek, T. A.; Tew, D. G.; Levy, M. A.; Janson, C. A.; D'Alessio, K. J.; McQueney, M. S.; Kurdyla, J.; Jones, C. S.; DesJarlais, R. L.; Abdel-Meguid, S. S.; Veber, D. F. Structure-based design of cathepsin K inhibitors containing a benzyloxy-substituted benzoyl peptidomimetic. *J. Med. Chem.* **1998**, *41*, 3923–3927.
- (12) Patterson, A. W.; Wood, W. J. L.; Hornsby, M.; Lesley, S.; Spraggon, G.; Ellman, J. A. Identification of selective, nonpeptidic nitrile inhibitors of cathepsin S using the substrate activity screening method. *J. Med. Chem.* **2006**, *49*, 6298–6307.
- (13) Marquis, R. W.; Ru, Y.; LoCastro, S. M.; Zeng, J.; Yamashita, D. S.; Oh, H.-J.; Erhard, K. F.; Davis, L. D.; Tomaszek, T. A.; Tew, D.; Salyers, K.; Proksch, J.; Ward, K.; Smith, B.; Levy, M.; Cummings, M. D.; Haltiwanger, R. C.; Trescher, G.; Wang, B.; Hemling, M. E.; Quinn, C. J.; Cheng, H. Y.; Lin, F.; Smith, W. W.; Janson, C. A.; Zhao, B.; McQueney, M. S.; D'Alessio, K.; Lee, C.-P.; Marzulli, A.; Dodds, R. A.; Blake, S.; Hwang, S.-M.; James, I. E.; Gress, C. J.; Bradley, B. R.; Lark, M. W.; Gowen, M.; Veber, D. F. Azepanone-based inhibitors of human and rat cathepsin K. *J. Med. Chem.* **2001**, *44*, 1380–1395.
- (14) Le, G. T.; Abbenante, G.; Madala, P. K.; Hoang, H. N.; Fairlie, D. P. Organic azide inhibitors of cysteine proteases. *J. Am. Chem. Soc.* **2006**, *128*, 12396–12397.
- (15) Gustin, D. J.; Schon, C. A.; Wei, J.; Cai, H.; Meduna, S. P.; Khatuya, H.; Sun, S.; Gu, Y.; Jiang, W.; Thurmond, R. L.; Karlsson, L.; Edwards, J. P. Discovery and SAR studies of a novel series of noncovalent cathepsin S inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1687–1691.
- (16) Tully, D. C.; Liu, U.; Chatterjee, A. K.; Alper, P. B.; Williams, J. A.; Roberts, M. J.; Mutnick, D.; Woodmansee, D. H.; Hollenbeck, T.; Gordon, P.; Chang, J.; Tuntland, T.; Tumanut, C.; Li, J.; Harris, J. L.; Karanewsky, D. S. Arylaminoethyl carbamates as a novel series of potent and selective cathepsin S inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5107–5111.
- (17) Tully, D. C.; Liu, H.; Chatterjee, A. K.; Alper, P. B.; Epple, R.; Williams, J. A.; Roberts, M. J.; Woodmansee, D. H.; Masick, B. T.; Tumanut, C.; Li, J.; Spraggon, G.; Hornsby, M.; Chang, J.; Tuntland, T.; Hollenbeck, T.; Gordon, P.; Harris, J. L.; Karanewsky, D. S. Synthesis and SAR of arylaminoethyl amides as noncovalent inhibitors of cathepsin S: P3 cyclic ethers. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5112–5117.
- (18) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (19) Catalyst, version 4.11; Accelrys: San Diego, CA, 2005.
- (20) Derwent World Drug Index 2003, version Thomson; Scientific: 2003.
- (21) ilib:diverse, version 1.0.2; Inte:Ligand: Maria Enzersdorf, Austria, 2006.
- (22) Kirchmair, J.; Ristic, S.; Eder, K.; Markt, P.; Wolber, G.; Lagner, C.; Langer, T. Fast and efficient in silico 3D screening: Toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J. Chem. Inf. Model.* **2007**, *47*, 2182–2196.
- (23) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2547.
- (24) Wolber, G.; Langer, T. LigandScout: 3D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–169.
- (25) Smith, R. A.; Bhargava, A.; Browe, C.; Chen, J.; Dumas, J.; Hatoum-Mokdad, H.; Romero, R. Discovery and parallel synthesis of a new class of cathepsin K inhibitors. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2951–2954.
- (26) Thompson, S. K.; Halbert, S. M.; Bossard, M. J.; Tomaszek, T. A.; Levy, M. A.; Zhao, B.; Smith, W. W.; Abdel-Meguid, S. S.; Janson, C. A.; D'Alessio, K. J.; McQueney, M. S.; Amegadzie, B. Y.; Hanning, C. R.; DesJarlais, R. L.; Briand, J.; Sarkar, S. K.; Huddleston, M. J.; Ijames, C. F.; Carr, S. A.; Barnes, K. T.; Shu, A.; Heys, J. R.; Bradbeer, J.; Zembyki, D.; Lee-Rykaczewski, L.; James, I. E.; Lark, M. W.; Drake, F. H.; Gowen, M.; Gleason, J. G.; Veber, D. F. Design of potent and selective human cathepsin K inhibitors that span the active site. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14249–14254.
- (27) Katunuma, N.; Tsuge, H.; Nukatsuka, M.; Fukushima, M. Structure-based development of cathepsin L inhibitors and therapeutic applications for prevention of cancer metastasis and cancer-induced osteoporosis. *Adv. Enzyme Regul.* **2002**, *42*, 159–172.
- (28) Pipeline Pilot version 5.0.1.100; Scitegic: San Diego, CA, 2006.
- (29) Lommerse, J. P. M.; Price, S. L.; Taylor, R. Hydrogen bonding of carbonyl, ether, and ester oxygen atoms with alkanol hydroxyl groups. *J. Comput. Chem.* **1997**, *18*, 757–774.
- (30) Kubinyi, H. In search for new leads. In *EFMC - Yearbook*, 2003 ed.; Ferran, S., Ed.; LD Organisation: Louvain-la-Neuve, Belgium, 2003; pp 14–28.
- (31) SciFinder, version 2006; American Chemical Society: Washington, DC, 2006.
- (32) Chatterjee, S. Preparation of D-amino acid derivatives as cysteine and serine protease inhibitors. PCT Int. Appl. WO 9721690, Nov 27, 1996.
- (33) Sohda, T.; Fujisawa, Y.; Yasuma, T.; Mizoguchi, J.; Kori, M.; Takizawa, M. Peptide alcohol or aldehyde derivatives as cathepsin L inhibitors and bone resorption inhibitors. Eur. Pat. Appl. EP 94-102404, Feb 17, 1994.

- (34) Edwards, B. S.; Bologa, C.; Young, S. M.; Balakin, K. V.; Prossnitz, E. R.; Savchuck, N. P.; Sklar, L. A.; Oprea, T. I. Integration of virtual screening with high-throughput flow cytometry to identify novel small molecule formylpeptide receptor antagonists. *Mol. Pharmacol.* **2005**, *68*, 1301–1310.
- (35) Schuster, D.; Laggner, C.; Steindl, T. M.; Paluszczak, A.; Hartmann, R. W.; Langer, T. Pharmacophore modeling and in silico screening for new P450 19 (aromatase) inhibitors. *J. Chem. Inf. Model.* **2006**, *46*, 1301–1311.
- (36) Jacobsson, M.; Liden, P.; Stjernschantz, E.; Bostroem, H.; Norinder, U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.* **2003**, *46*, 5781–5789.
- (37) Hecker, E. A.; Duraiswami, C.; Andrea, T. A.; Diller, D. J. Use of catalyst pharmacophore models for screening of large combinatorial libraries. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1204–1211.
- (38) Diller, D. J.; Li, R. Kinases, homology models, and high throughput docking. *J. Med. Chem.* **2003**, *46*, 4638–4647.
- (39) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–1519.
- (40) Shindo, K.; Suzuki, H.; Okuda, T. Paecilopeptin, a new cathepsin S inhibitor produced by *Paecilomyces carneus*. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 2444–2448.
- (41) Steindl, T.; Langer, T. Influenza virus neuraminidase inhibitors: generation and comparison of structure-based and common feature pharmacophore hypotheses and their application in virtual screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1849–1856.
- (42) Cha, S. Tight-binding inhibitors. I. Kinetic behavior. *Biochem. Pharmacol.* **1975**, *24*, 2177–2185.
- (43) Morrison, J. F. The slow-binding and slow, tight-binding inhibition of enzyme-catalyzed reactions. *Trends Biochem. Sci.* **1982**, *7*, 102–105.

CI800101J