

## Quantitative Structure–Activity Relationship of Human Neutrophil Collagenase (MMP-8) Inhibitors Using Comparative Molecular Field Analysis and X-ray Structure Analysis

Hans Matter,\* Wilfried Schwab, Denis Barbier, Günter Billen, Burkhard Haase, Bernhard Neises, Manfred Schudok, Werner Thorwart, Herman Schreuder, Volker Brachvogel, Petra Lönze, and Klaus Ulrich Weithmann

Chemical Research & Core Research Functions, Hoechst Marion Roussel, D-65926 Frankfurt am Main, Germany

Received November 6, 1998

A set of 90 novel 2-(arylsulfonyl)-1,2,3,4-tetrahydroisoquinoline-3-carboxylates and -hydroxamates as inhibitors of the matrix metalloproteinase human neutrophil collagenase (MMP-8) was designed, synthesized, and investigated by 3D-QSAR techniques (CoMFA, CoMSIA) and X-ray structure analysis. Docking studies of a reference compound are based on crystal structures of MMP-8 complexed with peptidic inhibitors to propose a model of its bioactive conformation. This model was validated by a 1.7 Å X-ray structure of the catalytic domain of MMP-8. The 3D-QSAR models based on a superposition rule derived from these docking studies were validated using conventional and cross-validated  $r^2$  values using the leave-one-out method, repeated analyses using two randomly chosen cross-validation groups plus randomization of biological activities. This led to consistent and highly predictive 3D-QSAR models with good correlation coefficients for both CoMFA and CoMSIA, which were found to correspond to experimentally determined MMP-8 catalytic site topology in terms of steric, electrostatic, and hydrophobic complementarity. Subsets selected as smaller training sets using 2D fingerprints and maximum dissimilarity methods resulted in 3D-QSAR models with remarkable correlation coefficients and a high predictive power. This allowed to compensate the weaker zinc binding properties of carboxylates by introducing optimal fitting P1' residues. The final QSAR information agrees with all experimental data for the binding topology and thus provides clear guidelines and accurate activity predictions for novel MMP-8 inhibitors.

### Introduction

Matrix metalloproteinases (MMPs) are a family of zinc endopeptidases capable of degrading the extracellular matrix of connective tissues.<sup>1</sup> They are broadly implicated in many degenerative diseases with slow matrix degradation rate, such as cartilage loss in osteoarthritis<sup>2</sup> and rheumatoid arthritis,<sup>3</sup> bone matrix degradation in osteoporosis, or remodeling in Alzheimer disease.<sup>4</sup> All MMPs possess three discrete domains: a propeptide (ca. 80 residues) as activation sequence, a catalytic domain (ca. 180 residues) including a conserved HEXXHXXGXXH zinc binding motif, and a hemopexin-like domain (ca. 250 residues), presumably responsible for substrate recognition.<sup>5</sup> Beside stromelysin-1 (MMP-3), particularly the interstitial collagenase (MMP-1) and the neutrophil collagenase (MMP-8) are responsible for the cleavage of type I, II, and III collagen, and they are seen as key enzymes in the pathology of matrix degradation.

Thus selective collagenase inhibitors are a very attractive goal for rational drug design. At an early stage, peptidic active site inhibitors were designed based on the knowledge of the natural substrate. Combined with a thiol, phosphonic, carboxylic, or hydroxamic acid group as zinc chelating function, they achieve in vitro inhibitory activity in the nanomolar range.<sup>6</sup> However most

substrate-like inhibitors exhibit low oral bioavailability. Therefore recent strategies are directed toward conformationally restricted, nonpeptidic inhibitors in terms of improving the pharmacokinetic properties.

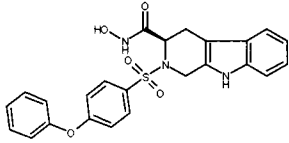
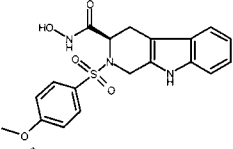
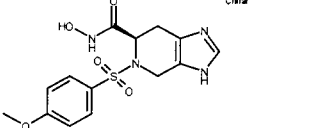
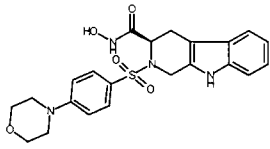
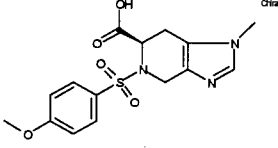
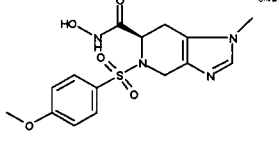
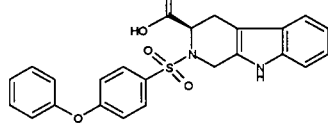
Stimulated by X-ray structure analyses of human MMP-8 in complex with low-molecular-weight ligands,<sup>7–10</sup> we designed a series of nonpeptidic 2-(arylsulfonyl)-1,2,3,4-tetrahydroisoquinoline-3-carboxylates and -hydroxamates and other bicyclic rigidified analogues as inhibitors of MMP-8. In this publication a set of 90 compounds was investigated by 3D-QSAR techniques (CoMFA, CoMSIA) in order to achieve significant, valid, and predictive models as a starting point for further design.

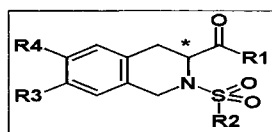
The determination of the active conformation is the crucial step in 3D-QSAR analyses. With the availability of a protein 3D structure, this active conformation can be obtained by two approaches: docking of a reference compound or X-ray crystallography. The first approach was chosen<sup>11</sup> and could be validated by subsequent X-ray structure analyses. The docking calculations were based on 3D structures of the MMP-8 catalytic domain with different peptides.<sup>12,9</sup> This led to an alignment for all other compounds by superimposing them onto the template and relaxing them within the cavity. This superposition produced consistent 3D-QSAR models explaining the key ligand–enzyme interactions.

Comparative molecular field analysis (CoMFA)<sup>13–15</sup> and comparative molecular similarity index analysis

\* To whom all correspondence should be addressed: Chemical Research, Molecular Modeling, G838. Tel: ++49-69-305-84329. Fax: ++49-69-331399. E-mail: hans.matter@hmrag.com.

**Table 1.** Chemical Structures and Biological Activities of 90 MMP-8 Inhibitors

No.	Structure	Isomer	MMP-8 IC 50 [nMol/l]
1		R	7
2		R	3
3		R	4
4		R	70
5		R	10000
6		R	3
7		R	100



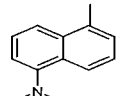
No.	Isomer	R1	R2	R3	R4	MMP-8 IC 50 [nMol/l]
8	R	NH-OH	4-Phenoxybenzene	H	H	2
9	R	NH-OH	Benzene	H	H	20
10	R	NH-OH	4-Acetamidobenzene	H	H	100
11	R	NH-OH	3-Ethoxycarbonylamino benzene	H	H	300
12	R	NH-OH	Dibenzofuran-2-yl	H	H	10
13	R	NH-OH	Phenylethyl	H	H	20
14	R	NH-OH	4-Ethoxycarbonylamino benzene	H	H	30
15	R	NH-OH	Hexadecyl	H	H	3000
16	R	NH-OH	4-Methoxybenzene	NO <sub>2</sub>	H	8
17	R	NH-OH	2-Naphtyl	H	H	70
18	R	NH-OH	Benzyl	H	H	3000
19	R	NH-OH		H	H	200

Table 1. (Continued)

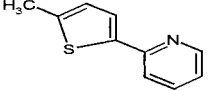
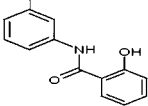
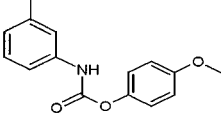
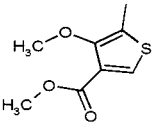
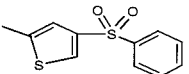
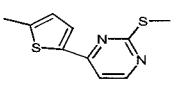
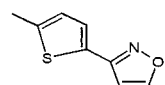
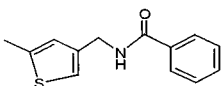
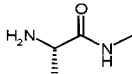
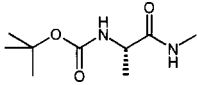
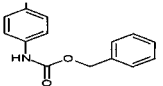
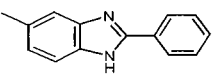
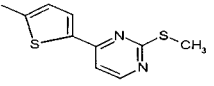
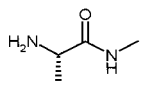
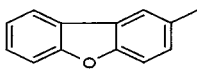
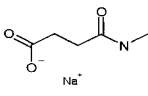
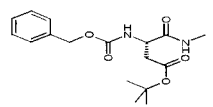
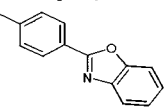
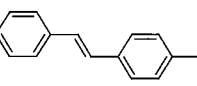
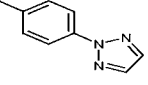
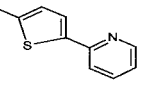
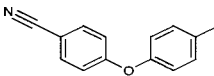
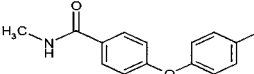
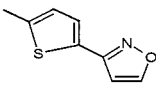
No.	Isomer	R1	R2	R3	R4	MMP-8 IC 50 [nMol/l]
20	S	NH-OH	4-Methoxybenzene	H	H	1000
21	R	NH-OH	4-Methoxybenzene	H	H	4
22	R	NH-OH	4-(4-Dimethylaminophenoxy)-phenyl	H	H	2
23	R	NH-OH	4-Biphenyl	H	H	2
24	R	NH-OH		H	H	2
25	R	NH-OH	4-Benzoylphenyl	H	H	3
26	R	NH-OH		H	H	200
27	R	NH-OH		H	H	200
28	R	NH-OH		H	H	80
29	R	COOH	4-Methoxybenzene	H	H	7000
30	R	NH-OH	4-Methoxybenzene	NH <sub>2</sub>	H	2
31	R	NH-OH	4-Methoxybenzene	tert.-Butoxycarbonylamino	H	3
32	R	NH-OH	4-Pyridylethyl	H	H	30
33	R	NH-OH	4-Butoxybenzene	H	H	2
34	R	NH-OH		H	H	2000
35	R	NH-OH		H	H	20
36	R	NH-OH	4-Morpholinobenzene	H	H	100
37	R	NH-OH	4-Methoxybenzene	Ethoxycarbonyloxy	H	2
38	R	NH-OH	4-Methoxybenzene	OH	H	1
39	R	NH-OH	4-Methoxybenzene	4-Methoxybenzene-sulfonyloxy	H	8
40	R	NH-OH		H	H	8
41	R	NH-OH		H	H	800
42	R	NH-OH	4,5-Dibromthiophene2-yl	H	H	500
43	(S)-Ala-(R)-Tic	NH-OH	4-Methoxybenzene		H	5
44	(S)-Ala-(R)-Tic	NH-OH	4-Methoxybenzene		H	2
45	R	NH-OH		H	H	800
46	R	NH-OH	4-Trifluoromethylbenzene	H	H	50
47	R	NH-OH		H	H	40

Table 1. (Continued)

No.	Isomer	R1	R2	R3	R4	MMP-8 IC 50 [nMol/l]
48	R	COOH	4-Methoxybenzene	NH <sub>2</sub>	H	2000
49	R	COOH	4-Chlorobiphenyl	H	H	5
50	R	NH-OH	4-Chlorobiphenyl	H	H	3
51	R	COOH		H	H	1000
52	R	COOH	4-Phenoxybenzene	H	H	9
53	R	COOH	4-(4-Fluorobenzyloxy)-benzene	H	H	30
54	R	COOH	4-Benzoylphenyl	H	H	300
55	(S)-Ala-(R)-Tic	COOH	4-Methoxybenzene		H	2000
56	R	COOH		H	H	500
57	R	NH-OH	4-(4-Fluorobenzyloxy)-benzene	H	H	2
58	R	COOH	4-Chlorobiphenyl	NO <sub>2</sub>	H	30
59	R	COOH	4-Chlorobiphenyl	H	NO <sub>2</sub>	20
60	R	COOH	4-Chlorobiphenyl	NH <sub>2</sub>	H	7
61	R	COOH	4-Phenylmercaptobenzene	H	H	200
62	R	COOH	4-Biphenyl	H	H	10
63	R	COOH	4-(4-Dimethylaminophenyl)-benzene	H	H	1
64	R	COOH	4-Fluorobiphenyl	NO <sub>2</sub>	H	40
65	R	COOH	4-Fluorobiphenyl	H	NO <sub>2</sub>	20
66	R	COOH	4-Methoxybenzene	OH	H	3000
67	S	COOH	4-Chlorobiphenyl	H	H	3000
68	R	COOH	4-(4-Chlorophenoxy)-benzene	H	H	20
69	R	COOH	4-Fluorobiphenyl	H	tert.-Butoxy-carbonyl-amino	30
70	R	NH-OH	4-Fluorobiphenyl	tert.-Butoxycarbonylamino	H	10
71	R	COOH	4-Fluorobiphenyl	H	H	20
72	R	NH-OH	4-Methoxybenzene		H	40
73	(S)-Asp-(R)-Tic	NH-OH	4-Methoxybenzene		H	10
74	R	COOH	4-Fluorobiphenyl	H	NH <sub>2</sub>	80
75	S	COOH	4-Fluorobiphenyl	H	H	4000
76	R	COOH		H	H	80
77	R	NH-OH	4-Fluorobiphenyl	NH <sub>2</sub>	H	2
78	R	COOH		H	H	60
79	R	COOH	4-Chlorobiphenyl	OH	H	20
80	R	COOH		H	H	90
81	R	COOH	4-(4-Dimethylamino)-phenoxybenzene	H	H	30
82	R	COOH		H	H	90
83	R	COOH	1-Styryl	H	H	300

**Table 1.** (Continued)

No.	Isomer	R1	R2	R3	R4	MMP-8 IC 50 [nMol/l]
84	R	COOH		H	H	10
85	R	COOH		H	H	40
86	R	COOH	(4-Biphenyl)-ethyl	H	H	20
87	R	COOH		H	H	600
88	R	NH-OH	4-(4-Dimethylaminophenyl)-benzene	H	H	2
89	R	COOH	4-(4-Trifluoromethyl)-biphenyl	H	H	10
90	R	COOH	4-Bromophenyl	H	H	600

(CoMSIA)<sup>16</sup> are used to derive relationships between molecular property fields of aligned compounds and their biological activities. Electrostatic and steric interaction energies are computed between each ligand and a probe atom located on predefined grid points for CoMFA, while for CoMSIA those interaction fields are replaced by fields based on similarity indices between probe atoms and each molecule. The PLS method (partial least squares)<sup>17</sup> is used to derive a linear relationship for highly underdetermined matrices; cross-validation<sup>18</sup> is used to check for consistency and predictiveness.

The resulting contour maps from 3D-QSAR models enhance the understanding of electrostatic, hydrophobic, and steric requirements for ligand binding, guiding the design of novel inhibitors to those regions, where structural variations altering steric or electrostatic fields reveal a significant correlation to biological properties.

## Methods

**Design of MMP-8 Inhibitors.** Our inhibitor design (Table 1) is based on active site topologies from MMP-8 X-ray structures. The specificity of collagenases to cleave at medium-sized hydrophobic residues resides in the S1' pocket,<sup>19</sup> located next to the catalytic Zn<sup>2+</sup>; its wall is formed by the strand Pro217–Tyr219.<sup>20</sup> The S1' entrance is formed by Gly158–Ala161 plus the Tyr219 side chain, leading into a more hydrophobic interior. The side chain of Arg222 with little steric hindrance at the S1' bottom adopts different conformations in different X-ray structures. Its guanidine side chain is hydrogen-bonded to Pro211-C=O, Gly212-C=O (via solvent), and Ala213-C=O. In most crystal structures the S1' pocket is filled with two or three solvent molecules bonded to Arg222 guanidine and polar backbone atoms.

This topology provided our basis for design: As zinc complexing functionalities hydroxamates and carboxylates were selected, although the latter do not provide the optimal geometry for strong zinc interactions. However, carboxylates show a higher bioavailability in related series of ligands,<sup>21</sup> which prompted us to look for other compensations of this weakness. This led to the design of rigid 2-(arylsulfonyl)-1,2,3,4-tetrahydroisoquinoline-3-carboxylates and corresponding hydroxamates. Several hydrophobic and hydrogen-bond regions in the MMP-8 binding site were identified using the program GRID<sup>22</sup> and further used for design: A hydrogen bond involving Leu160 and Ala161 NH's is a common motif, where a sulfonamide involving the tetrahydroisoquinoline nitrogen provides ideal complementarity plus a convenient synthetic route to a number of variations to probe S1' requirements. The incorporation of the aromatic side chains of Tyr219 and His197

**Table 2.** Data Collection and Refinement Statistics

space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
cell dimensions	<i>a</i> = 33.1 Å, <i>b</i> = 68.9 Å, <i>c</i> = 70.5 Å
observed reflections	129143
unique reflections	18344
resolution	1.7 Å
<i>R</i> -sym	4.8%
completeness	99.8%
protein atoms (non-H)	1302
water molecules	208
inhibitor + HEPES atoms	47
<i>R</i> -factor	19.2%
rmsd bond lengths	0.012 Å
rmsd bond angles	1.54 Å

at opposite sides of the S1' wall suggested that another favorable contribution to binding energy could result from protein–ligand aromatic–aromatic interactions.<sup>23</sup> The final design consideration was to fill the space occupied by crystallographical water molecules in MMP-8 structures in S1' by bulky hydrophobic substituents.

**Chemistry and Enzyme Assay.** As described in parallel publications<sup>24,25</sup> 2-(arylsulfonyl)-1,2,3,4-tetrahydroisoquinoline-3-carboxylates and -hydroxamates and their analogues (Table 1) were synthesized from sulfochlorides and unnatural amino acids under Schotten–Baumann conditions. Sulfonic acids were converted to their sulfochlorides using standard procedures (PCl<sub>5</sub>, POCl<sub>3</sub>, or SOCl<sub>2</sub>). Alternatively, an aromatic ring system was directly sulfochlorinated in the para position using chlorosulfonic acid. (*R*)- or (*S*)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid or ring-modified analogues as amine components were reacted with the appropriate sulfochloride in THF, DMF, or THF/water mixtures in the presence of organic or inorganic base. Conversion to the corresponding hydroxamate was achieved by activation of the acids with isobutyl chloroformate/*N*-ethylmorpholine or oxalyl chloride and reaction with *O*-(trimethylsilyl)hydroxylamine in dry THF followed by treatment with aqueous HCl for deprotection.

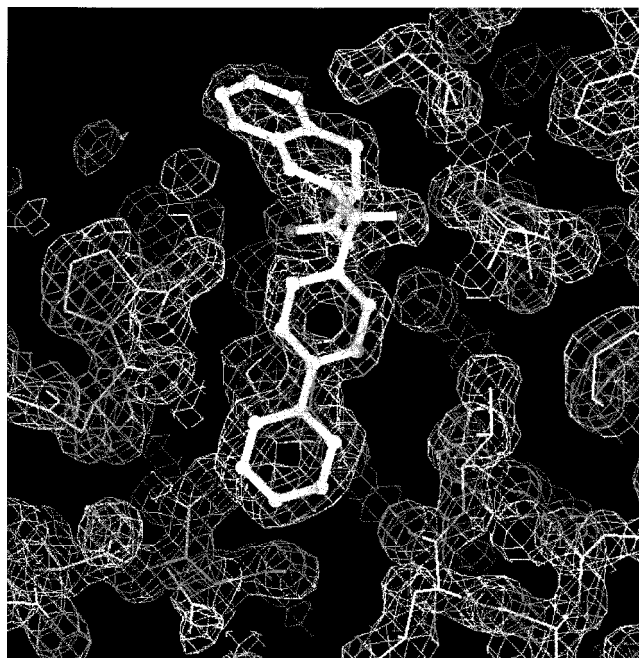
The catalytic domain of human recombinant MMP-8 (hMMP-8cd; Gly99–Gln271; MW 19.9 kDa) was obtained as described previously.<sup>26</sup> Biological activities were determined in 96-well titer plate format in a luminescence spectrometer (LS 50B, Perkin-Elmer, Langen, FRG) using the quenched fluorogenic substrate<sup>27</sup> 7-methoxycoumarin-4-yl acetyl-Pro-Leu-Gly-Leu-3-(2',4'-dinitrophenyl)-L-2,3-diaminopropionyl-Ala-Arg-NH<sub>2</sub> ( $\lambda_{\text{ex}}$  = 328 nm,  $\lambda_{\text{em}}$  = 393 nm; available from Bachem, Heidelberg, Germany). Each well contained 70  $\mu$ L of buffer (0.1 mol/L Tris/HCl, pH 7.5; 0.1 mol/L NaCl; 0.01 mol/L CaCl<sub>2</sub>; 0.05% Brij 35), 10  $\mu$ L of enzyme, and 10  $\mu$ L of drug in 10% DMSO. After preincubation for 15 min at room temperature the reaction was started by addition of 10  $\mu$ L of substrate (1 mmol/L in 10% DMSO). The initial velocity of enzymatic reaction was determined without ligand (=100%) and with seven different ligand concentrations, ranging from 10<sup>-4</sup> to 10<sup>-10</sup> mol/L. Each





**Figure 1.** Superposition of the MMP-8–**62** complex with the Pro-Leu-Gly-hydroxamate complex from 1jan. The protein fold, taken from our study, is indicated by a magenta ribbon–tube display encoding secondary structure elements. The fold of 1jan is virtually identical and not shown. The Pro-Leu-Gly-hydroxamate inhibitor is shown with gray carbons, while **62** is displayed with green carbons. Gray spheres indicate the structurally conserved zinc ions.

measurement was done in duplicate to ensure statistically significant results ( $p < 0.05$ ). The ligand-dependent inhibition of the initial velocity was plotted, and the  $IC_{50}$  values (concentration necessary to obtain 50% inhibition) were determined using standard software.



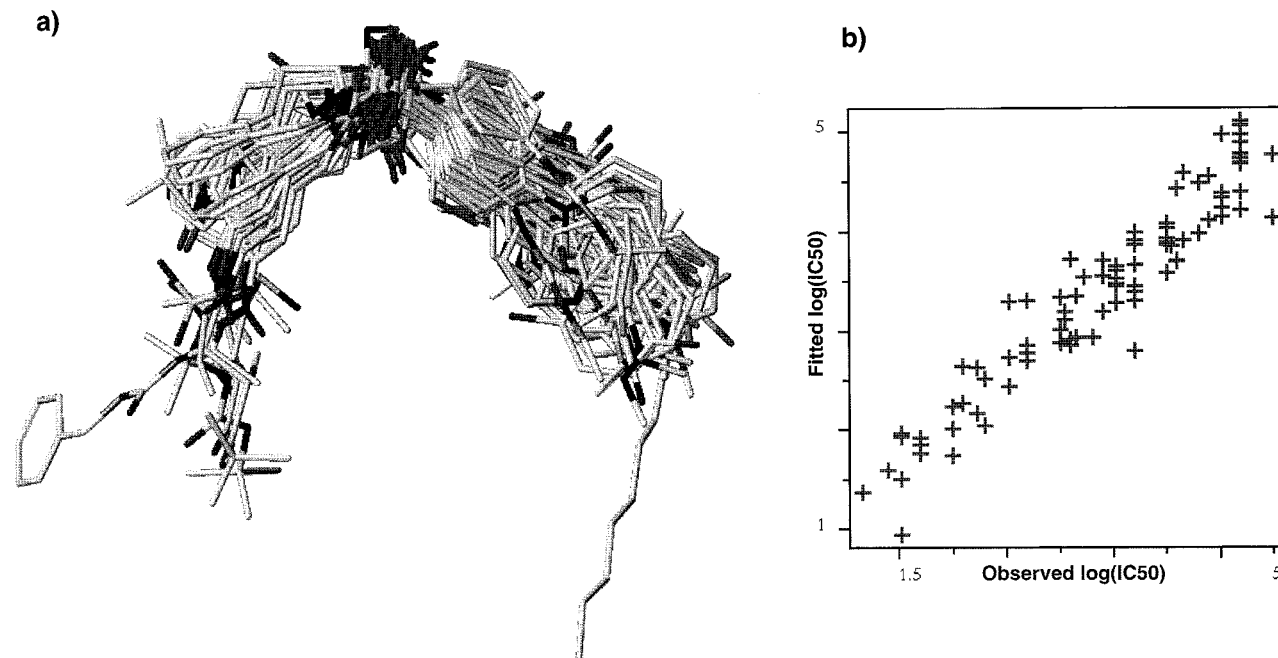
**Figure 2.** Final 1.7 Å electron density map of the MMP-8 active site region cocrystallized with **62**. Shown is the  $2F_o - F_c$  map contoured at  $1\sigma$ .

**X-ray Structure Analysis.** The purified MMP-8 was crystallized using the hanging drop method. A 3-fold excess of **62** in DMF was added to a solution of 10 mg/mL protein in buffer A (25 mM MES, 100 mM NaCl, 10 mM  $CaCl_2$ , 0.1 mM  $ZnCl_2$ , pH 6.0). The solution contained 10–25% (w/w) PEG6000. Needle-shaped crystals appeared in 1 week. Data were collected to 1.7 Å using a 345-mm image plate detector (Mar Research, Hamburg, Germany), mounted on a rotating anode

**Table 3.** Summary of 3D-QSAR Analysis Results Obtained Using CoMFA or CoMSIA Molecular Fields for 90 MMP-8 Inhibitors

(a) Statistical Results per PLS Component for Various Analyses <sup>a</sup>									
	component								
	1	2	3	4	5	6	7		
CoMFA_2									
$r^2$ (cv)	0.375	0.423	0.445	0.535	<b>0.569</b>				
SD(cv)	0.806	0.779	0.768	0.708	<b>0.685</b>				
$r^2$	0.549	0.660	0.772	0.874	<b>0.905</b>				
SD	0.685	0.598	0.492	0.367	<b>0.321</b>				
CoMFA_1									
$r^2$ (cv)	0.384	0.433	0.448	0.500	<b>0.516</b>				
SD(cv)	0.800	0.772	0.766	0.733	<b>0.726</b>				
$r^2$	0.564	0.665	0.760	0.863	<b>0.911</b>				
SD	0.673	0.593	0.505	0.383	<b>0.311</b>				
CoMSIA_2									
$r^2$ (cv)	0.344	0.389	0.345	0.372	0.415	0.444	<b>0.478</b>		
SD(cv)	0.826	0.802	0.834	0.822	0.798	0.782	<b>0.763</b>		
$r^2$	0.579	0.678	0.777	0.831	0.883	0.907	<b>0.924</b>		
SD	0.661	0.574	0.487	0.426	0.376	0.320	<b>0.292</b>		
CoMSIA_1									
$r^2$ (cv)	0.327	0.359	0.343	0.348	0.384	0.416	<b>0.447</b>		
SD(cv)	0.836	0.821	0.836	0.838	0.819	0.802	<b>0.786</b>		
$r^2$	0.579	0.669	0.773	0.829	0.883	0.908	<b>0.924</b>		
SD	0.661	0.589	0.491	0.429	0.357	0.319	<b>0.290</b>		
(b) Statistical Results for the PLS Analysis with the Optimal Number of Components									
analyses	grid spacing (Å)	min $\sigma$	$r^2$	no. of var./field	no. of compnts	$F$ -value	contributions		
							steric	electr	hydrophob
CoMFA_2	2	2	0.905	2688	5	160.5	0.543	0.457	
CoMFA_1	1	2	0.911	19251	5	172.1	0.504	0.496	
CoMSIA_2	2	0	0.924	2688	7	141.5	0.161	0.476	0.363
CoMSIA_1	1	0	0.924	19251	7	143.1	0.165	0.474	0.361

<sup>a</sup> Bold letters indicate the PLS analyses with the optimal number of components after cross-validation.



**Figure 3.** (a) Superposition of 90 MMP-8 inhibitors for 3D-QSAR studies (cf. Table 1), built on the basis of the docked, *bioactive* conformation of the reference compound **62** and fitted into the protein cavity. Hydrogens are omitted for clarity. (b) Graph of observed versus fitted biological activity for the final CoMFA model with 2 Å grid spacing.

generator model FR571 (Nonius, Delft, The Netherlands) using Cu K $\alpha$  radiation at cryo-temperatures. Processing was done using the XDS package.<sup>28</sup> The structure was determined using the program X-PLOR version 3.1<sup>29</sup> and a model derived from coordinates in the Brookhaven Protein Data Bank (1jan<sup>9</sup>). The space group of 1jan is identical to ours, and the cell dimensions differ less than 3%. Rigid body refinement reduced the *R*-factor from 0.447 to 0.355 (15.0–3.0 Å data). Energy minimization and temperature factor refinement reduced the *R*-factor to 0.232 (8.0–2.8 Å data). At this stage, clear and unambiguous electron density for the inhibitor was present and the inhibitor was fitted. After nine more cycles of map inspection, energy minimization, and temperature factor refinement, a final model was obtained with an *R*-factor of 0.192 and excellent geometry (Table 2). (The coordinates of the X-ray structure analysis for MMP-8–**62** have been deposited in the Brookhaven Protein Data Bank (1bzs).)

**Computational Procedures.** Modeling work was performed using the program SYBYL<sup>30</sup> on Silicon Graphics workstations. All energy calculations were based on the TRIPOS 6.0 force field<sup>31</sup> including Gasteiger–Marsili charges.<sup>32</sup> Conformations of ligands and complexes were minimized using quasi-Newton–Raphson (BFGS) or conjugate gradient (CG) procedures. For docking the X-ray structures of catalytic domains of MMP-8 were used.<sup>12,14</sup> After analysis of key protein–ligand interactions using GRID,<sup>22</sup> candidate molecules were manually docked into the active site. Subsequently the protein–ligand complex for **62** was minimized treating all ligand atoms and protein residues within a sphere of 4 Å as flexible, while the remaining receptor was used to compute nonbonded interactions. Convergence criteria were set to 0.001 kcal/Å. All other compounds were built accordingly, docked into MMP-8, and minimized using a rigid receptor from the first minimization. In parallel, all compounds were minimized treating the receptor as flexible within a 4 Å sphere. However, subsequent CoMFA analyses did not lead to improved statistical results compared to the rigid alignment.

After superposition steric and electrostatic interaction energies between a probe atom and every structure are calculated at surrounding points of a predefined grid, using a volume-dependent lattice with 1 or 2 Å grid spacing, a positively charged carbon atom, and a distance-dependent dielectric constant. The magnitude of the regions was defined to extend the superimposed conformers by 4.0 Å along the principal axes.

The maximum field values were truncated to 30 kcal/mol for steric and  $\pm 30$  kcal/mol for electrostatic energies. For points “inside” a molecule (steric energy of 30 kcal/mol), no electrostatic energy was computed, and those values were set to the mean of the related column.

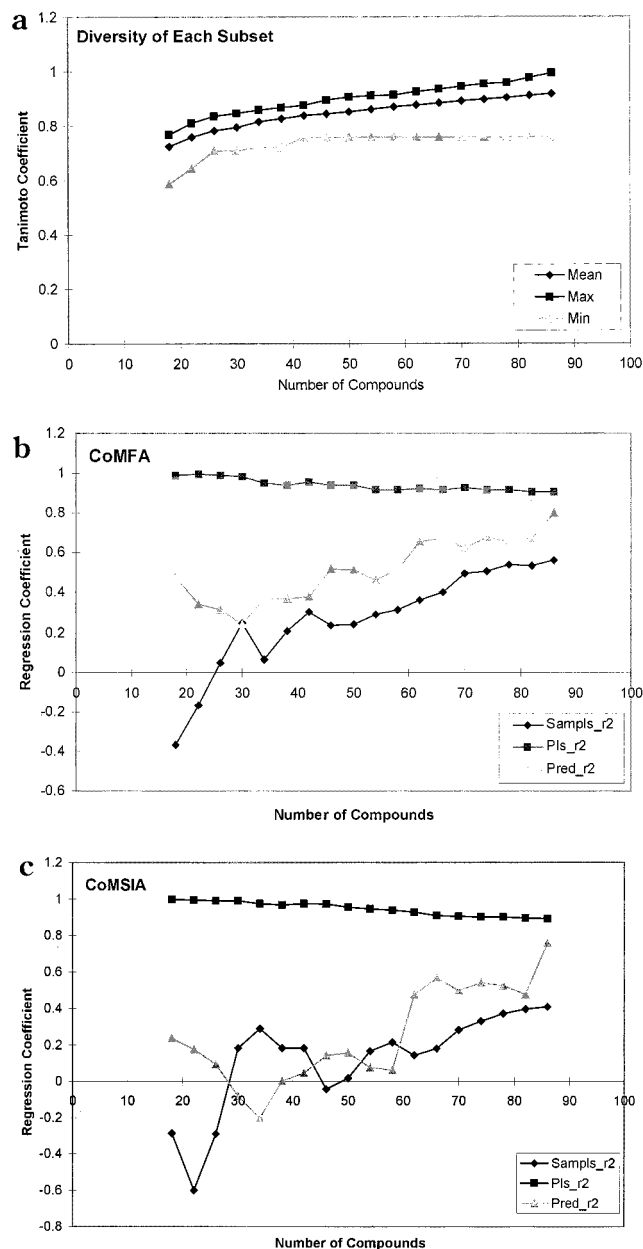
This alignment and region also served to compute steric, electrostatic, and hydrophobic similarity index fields for CoMSIA. The latter field description is based on Crippen’s partial atomic hydrophobicities.<sup>33</sup> The advantage of CoMSIA fields is that no singularities occur at atomic positions due to a Gaussian-type distance dependence of the physicochemical properties; thus no arbitrary cutoffs are required. Similarity indices<sup>34</sup> were computed using a probe with a charge of +1, a radius of +1, a hydrophobicity of +1, and 0.3 as attenuation factor  $\alpha$  for the Gaussian-type distance dependence.

Equal weights for CoMFA or CoMSIA fields were assigned using the CoMFA STD scaling option.<sup>35</sup> Enzyme inhibition is expressed as  $\log(1/IC_{50} \times 100000)$ . Cross-validated analyses were run using the leave-one-out method in SAMPLS<sup>36</sup> and two cross-validation groups with random selection of group members. PLS analyses using two randomly selected cross-validation groups were averaged over 100 runs. While for CoMFA columns with a variance smaller than 2.0 were excluded prior to the PLS analysis (*minimum-sigma*), no column-filtering was used for CoMSIA. The overall quality of all PLS analyses was expressed using the cross-validated  $r^2$  value  $r^2(cv)$ , defined as:

$$r^2(cv) = (SD - PRESS)/SD \quad (1)$$

where SD is the variance of the biological activities around the mean values and PRESS represents the sum of squared differences between predicted and target property values. The ideal value of 1.0 is reached when PRESS becomes 0.0 (perfect internal prediction).

2D Fingerprints for compound selection, generated using the program UNITY,<sup>37</sup> contain information about the presence of molecular fragments in a binary format. Their similarity is computed using the Tanimoto coefficient,<sup>38</sup> defined by the number of bit sets in both bitstrings normalized by the number of bit sets in common. Compound selections were done using *maximum dissimilarity* methods and 2D fingerprints.<sup>39,40</sup> After random selection of a seed compound, every new compound is chosen such that it is maximally dissimilar from all previous



**Figure 4.** Validation and assessment of the predictivity of 3D-QSAR models with several smaller training sets designed using maximum dissimilarity methods. (a) Mean and maximum Tanimoto coefficient for training sets on the  $y$ -axis versus training set size. (b) Comparison of cross-validated, conventional, and predictive  $r^2$  values for individual PLS models for a particular training set on the  $y$ -axis versus training set size using CoMFA as descriptor. (c) Comparison of cross-validated, conventional, and predictive  $r^2$  values for individual PLS models for a particular training set on the  $y$ -axis versus training set size using CoMSIA as descriptor.

members. The first three compounds are rejected after the fourth selection but are allowed for later picking. The mean Tanimoto coefficient is computed as an average from Tanimoto coefficients for every structure to its nearest neighbor.

## Results and Discussion

**X-ray Structure Analysis.** Crystals of the MMP-8-inhibitor **62** complex diffracted well up to 1.7 Å resolution. All  $\varphi/\psi$  angle combinations are in allowed regions; other geometric parameters are as expected or better for this resolution.<sup>41</sup> The overall structure of MMP-8 is very similar to 1jan with an rmsd of 0.24 Å for 164

equivalent C $\alpha$  atoms. The inhibitor is well-defined in the 1.7 Å electron density map (Figure 2), and the tetrahydroisoquinoline ring is situated at the entrance to the active site between Ile159 and His207, while its aromatic part is exposed to the solvent. Both oxygens of the 3-carboxyl group are in coordination with the catalytic zinc. One of the sulfonamide oxygens is exposed to the solvent, while the other is involved in hydrogen bonds to Leu160-NH and Ala161-NH. The biphenyl system is buried inside the S1' pocket, stacked on top of the His197 imidazole. At the other side, the loop between Leu214 and Ala220 wraps around the inhibitor. This experimental binding mode is very similar to that used as a template for defining the alignment rule.

**3D-QSAR Models.** All resulting 3D-QSAR models showed a high degree of internal consistency. Using a 2 Å grid spacing, a CoMFA model with an  $r^2(\text{cv})$  value of 0.569 for 5 components and a conventional  $r^2$  of 0.905 was obtained with the alignment obtained using the MMP-8 binding site topology (see Table 3 and Figure 3). The graph of observed versus the fitted biological activities<sup>42</sup> is also shown in Figure 3.

The steric field descriptors (2 688 variables) explain 54.3% of the variance, while the electrostatic descriptors explain 45.7%, showing a balance between both effects. When reducing the grid spacing to 1 Å (19 251 variables), the quality of the PLS model changes only slightly. Again a 5-component model was obtained with  $r^2(\text{cv})$  of 0.516,  $r^2$  of 0.911 (cf. Table 3), and a balanced relationship between steric and electrostatic descriptors (50.4%/49.6%).

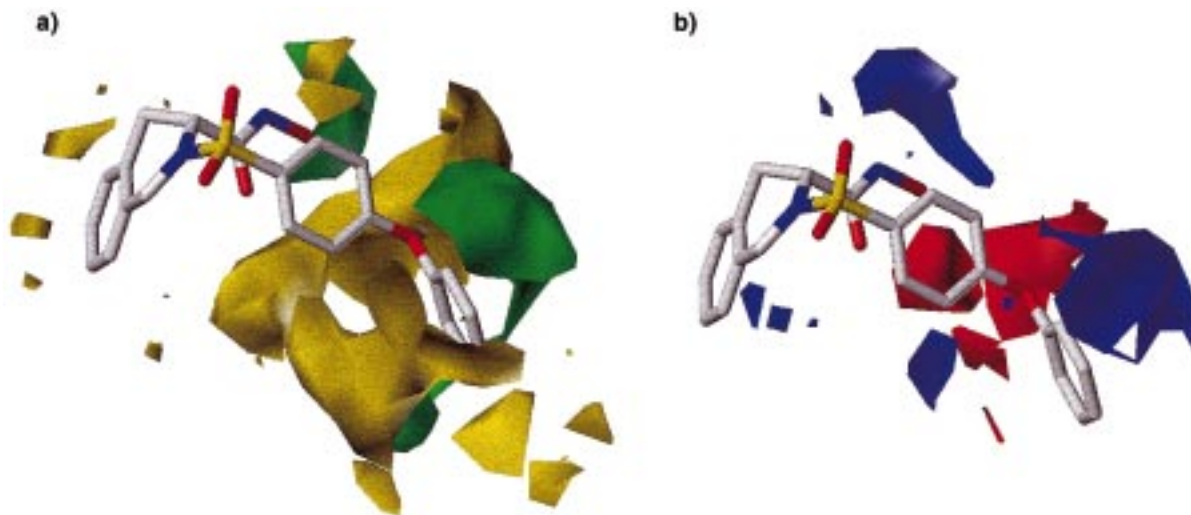
The model with 2 Å grid spacing was further validated to assess its predictive power. The effect of the alignment relative to the grid position was investigated by consistently moving all compounds in increments of 0.5 Å. The  $r^2(\text{cv})$  values for each orientation range from 0.42 to 0.58, suggesting a slight dependence of the model on the absolute orientation.

The effect of different choices of the origin of the grid was investigated using 14 atom types in addition to C3 as probe atoms with a 2 Å grid spacing. The  $r^2(\text{cv})$  values for each probe atom range from 0.543 (O2, O3, Oco2; 8 components) to 0.583 (C3), showing only a slight dependence on the chosen probe (mean  $r^2(\text{cv})$ , 0.56; SD, 0.02). The C3-based model was obtained for 6 PLS components, but the last component is not required for interpretation and was skipped in further analyses.

Subsequently the biological activities were randomly assigned to molecular structures and subjected to a leave-one-out statistical analysis.<sup>43</sup> The mean  $r^2(\text{cv})$  for 50 randomizations is -0.13 (SD 0.11). No model was obtained with more than 3 components. Thus the reported PLS model is significantly better than a random model.

Although cross-validation reflects the predictive power of a model, the leave-one-out method might produce too high  $r^2(\text{cv})$  values.<sup>44</sup> Thus PLS analyses were run with two randomly chosen cross-validation groups each containing 45 randomly selected compounds. Due to the random group formation, the analyses were repeated 100 times. The mean  $r^2(\text{cv})$  value of 0.438 (SD, 0.08; high, 0.617; low, 0.235) was found to be slightly lower than using the leave-one-out method (0.569) with a low standard deviation, which also counts for a stable,





**Figure 5.** Contour maps from the final CoMFA analysis with 1 Å grid spacing in combination with the inhibitor **8**. Hydrogens are omitted for clarity. (a) Steric std\*coeff contour map. Green contours (>85% contribution) refer to sterically favored regions; yellow contours (<15% contribution) indicate disfavored areas. (b) Electrostatic std\*coeff contour map. Blue contours (>85% contribution) refer to regions where negatively charged substituents are disfavored; red contours (<15% contribution) indicate regions where negatively charged substituents are favored.

predictive PLS model. However some analyses with a low  $r^2(\text{cv})$  value were obtained, which points to some inconsistency in this dataset.

A final study was done using smaller, representative subsets to derive PLS models for prediction, which should allow for another assessment of the predictive power of a model.<sup>45,46</sup> Representative subsets with 18–86 members were selected using maximum dissimilarity methods and 2D fingerprints.<sup>47</sup> For each subset the similarity among individual structures (mean Tanimoto coefficients) is plotted in Figure 4a. A cross-validated PLS analysis using leave-one-out serves to extract the  $r^2(\text{cv})$  plus the optimal number of components. Then a PLS analysis without cross-validation using a number of 5 components as best compromise led to the  $r^2$  and a model for predicting the remaining compounds, resulting in a predictive  $r^2$  value.

The different  $r^2$  values versus the number of components for each subset are plotted in Figure 4b. The  $r^2(\text{cv})$  value is low with less than 40 diverse compounds. Although the conventional  $r^2$  is high, the predictive ability is not sufficient. When increasing the subset size, the cross-validated  $r^2$  reaches values between 0.5 and 0.6, leading to stable and predictive models. The predictive  $r^2$  value decreases again with more than 78 molecules, as the prediction set is then statistically no longer significant.

Similar results were obtained using CoMSIA. Using a 2 Å grid spacing, a CoMSIA model with an  $r^2(\text{cv})$  value of 0.478 for 7 components and a conventional  $r^2$  of 0.924 was obtained (Table 3). The steric field descriptors (2 688 variables) explain only 16.1% of the variance, while the proportion of the electrostatic descriptors remains in the same range with 47.6%. Now the additional hydrophobic field explains the remaining 36.3% of the variance. Thus the CoMFA steric field contribution can be seen as a balance between pure steric plus hydrophobic effects. With a grid spacing of 1 Å (19 251 variables), the model remains stable, and a 7-component model was obtained with  $r^2(\text{cv})$  of 0.447 and  $r^2$  of 0.924.

The CoMSIA 2 Å model was subjected to further validation. When randomizing the biological activities 50 times, a mean  $r^2(\text{cv})$  of  $-0.18$  (SD, 0.10; high, 0.04; low,  $-0.44$ ) is observed, revealing the significance of this model. When averaging over 100 PLS analyses with two random cross-validation groups, a mean  $r^2(\text{cv})$  of 0.321 (SD, 0.08; high, 0.523; low, 0.123) results, which again is only slightly lower than the  $r^2(\text{cv})$  value obtained using leave-one-out. Finally the same maximum dissimilarity selected subsets were used to establish PLS models using CoMSIA as descriptor. The obtained  $r^2$  values for each subset are plotted in Figure 4c on the  $y$ -axis versus the subset population. For CoMSIA the first reliable PLS models emerge with more than 70 molecules, thus showing that CoMFA models have a somewhat better predictive power than CoMSIA in this particular case.

**Comparison with the Receptor Topology.** The steric and electrostatic std\*coeff CoMFA fields with a 1 Å grid are shown as contour maps in Figure 5 with the potent inhibitor **8** containing a hydroxamate and a biphenyl ether sulfonamide attached to the 1,2,3,4-tetrahydroisoquinoline ( $\text{IC}_{50}$  2 nM). Those maps are similar for different CoMFA or CoMSIA analyses. In all figures the MMP-8 standard view<sup>9</sup> was rotated by  $-90^\circ$ , which allows for a view into the S1' pocket.

The steric field contributions are displayed in Figure 5a. Green contours (>85% contribution) represent regions where steric bulk is favorable for biological potency, while yellow contours (<15% contribution) indicate regions affecting the biological activity in a negative way: a bulky substituent will significantly lower the activity. Although PLS results were obtained taking only ligand information into account, the contour maps are discussed in combination with the receptor topology to show their consistency with steric, electrostatic, and hydrophobic binding requirements.

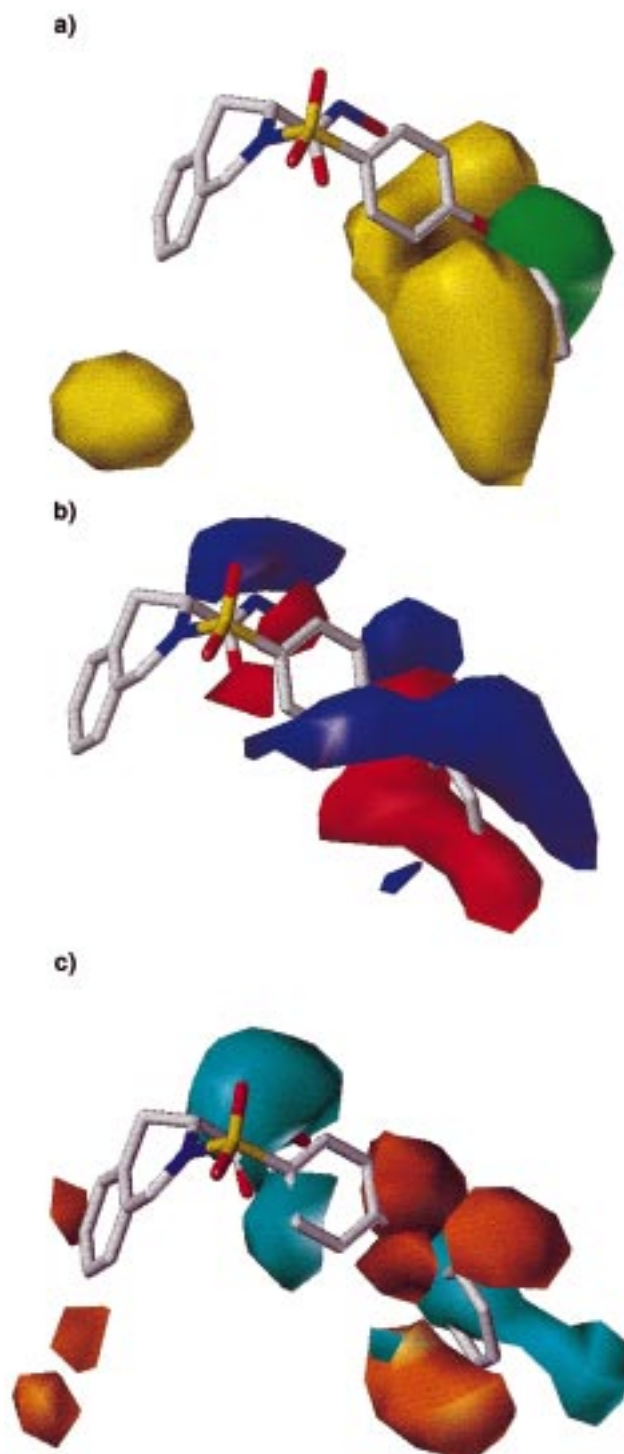
Enhancing steric bulk in the inhibitor molecule toward the bottom of the S1' pocket increases biological activity. A green contour in Figure 5 at the distal biphenyl ether ring highlights another hydrophobic cleft formed by the side chains of Tyr219, Leu193, and

Val194, which is filled with structurally conserved water in several MMP-8–inhibitor complexes. The ortho position of this terminal ring is highlighted by another green contour pointing to the “Met-turn”, a type I  $\beta$ -turn encompassing Ala213(*i*)-Leu214-Met215-Tyr216(*i*+4). This motif represents a conserved topological fragment in the “metzincins”<sup>5</sup> as hydrophobic basis for histidines coordinating the catalytic zinc. The green contour is located parallel to the plane of the Met215-Tyr216 amide bond, where the observed side chain rotamer of Tyr216 tolerates bulky ligands. This region is restricted by some yellow contours indicating hindrance by Ala213 and Leu214. The third favorable region indicates the preferred zinc binding geometry, as the optimal distance of two oxygens for zinc coordination is better realized in hydroxamates (2.7 Å) than in carboxylates (2.2 Å).<sup>7</sup>

A yellow contour in Figure 5 at the first biphenyl ether ring indicates a close proximity of low-affinity ligands to Pro217, His207, Zn<sup>2+</sup>, His197, or Val194, which correspond to the narrow, polar S1' entrance. At the edge of S1' unfavorable steric interactions at the Asn218-Tyr219 peptide backbone are identified. The tolerated steric requirements at the S1' bottom at Arg222 are highlighted by another yellow region, as this side chain cannot rotate unconstrained and imposes restrictions on the ligand's size. Finally a region at the upper side of the 1,2,3,4-tetrahydroisoquinoline ring is in agreement with the preferred stereochemistry at C3. With inverted chirality, this aromatic ring is moved toward the yellow contour indicating a decrease of biological activity. The favorable location of the aromatic ring close to the protein surface is in agreement with interactions of a hydrophobic probe to the binding cavity identified using GRID. The difference in affinity is caused by hydrophobic interactions to the plane of the His207 imidazole.

The inspection of CoMSIA steric std\*coeff contour maps shows a high correspondence to the CoMFA results (see Figure 6), although the steric field explains only 16.1% of the overall variance. One advantage of CoMSIA is that entropic contributions to binding, which cannot be completely treated using Lennard–Jones and Coulombic fields, are treated using a hydrophobic similarity index field. The steric std\*coeff contour maps point directly to the favorable regions at the ligand (green, >80% contribution), indicating size and shape of substituents binding into the S1' specificity pocket, especially for the second biaryl ether ring. Sterically unfavorable regions are indicated by yellow contours (<20% contribution), e.g., Asn218-Tyr219, the S1' entrance region. In general CoMSIA contours are more compact and centered on the ligand atoms,<sup>16</sup> while CoMFA contours are scattered at the edge of the ligand surface, thus highlighting complementary features at the receptor site.

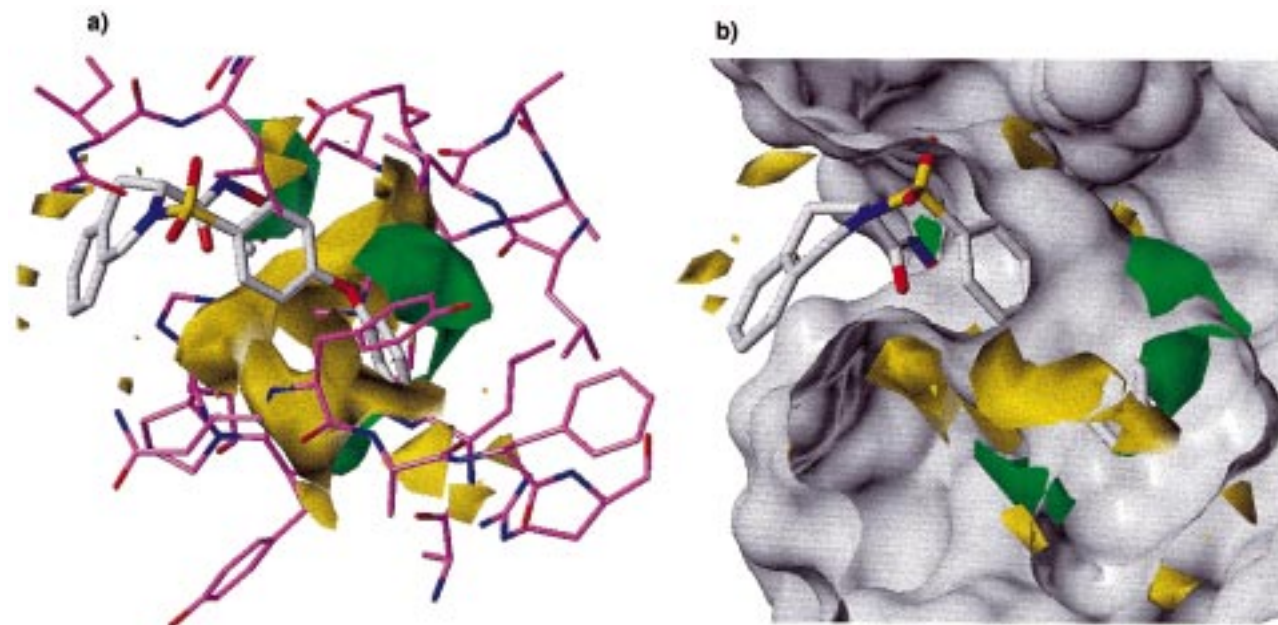
The electrostatic std\*coeff contour maps are also similar for CoMFA (Figure 5) and CoMSIA (Figure 6). Blue contours (>85% contribution for CoMFA, >80% for CoMSIA) represent regions where more positive charge or a reduction of negative charge is favorable to enhance biological potency, while red contours (<15% contribution for CoMFA, <20% for CoMSIA) indicate regions where more negative charge is favored. Red CoMFA and CoMSIA contours point toward the position of the



**Figure 6.** Contour maps from the final CoMSIA analysis with 1 Å grid spacing in combination with the inhibitor **8**. Hydrogens are omitted for clarity. (a) Steric std\*coeff contour map. Green contours (>80% contribution) refer to sterically favored regions; yellow contours (<20% contribution) indicate disfavored areas. (b) Electrostatic std\*coeff contour map. Blue contours (>80% contribution) refer to regions where negatively charged substituents are disfavored; red contours (<20% contribution) indicate regions where negatively charged substituents are favored. (c) Hydrophobic std\*coeff contour map. Cyan contours (>80% contribution) refer to regions where hydrophilic substituents are favored; orange contours (<20% contribution) indicate regions where hydrophobic substituents are favored.

catalytic zinc, indicating the optimal geometry of the bidental zinc binding group. From CoMSIA even the





**Figure 7.** Steric std\*coeff contour maps from the final CoMFA analysis with 1 Å grid spacing in combination with the inhibitor **8** mapped onto selected amino acids from the MMP-8 active site, taken from our X-ray analysis. Green contours (>85% contribution) refer to sterically favored regions; yellow contours (<15% contribution) indicate disfavored areas. (a) Color-coded stick representation of selected MMP-8 residues. (b) Connolly surface of the MMP-8 binding site. To allow a view toward the S1' pocket, depth-cueing was applied.

optimal distance of 2.7 Å better realized for both hydroxamate oxygens is obvious. Another red CoMSIA region is located at the biphenyl ether oxygen atom, suggesting that a reduction of positive charge is favorable. The corresponding CoMFA region points toward the plane of the His97 aromatic ring, while hydrophobic interactions are the dominant reason for ligand binding here (hydrophobic CoMSIA contour maps in Figure 6).

Blue CoMFA and CoMSIA contours at the hydroxamate NH indicate favorable interactions by hydrogen bonds to Ala161-C=O and the Glu198 side chain in accord with X-ray studies.<sup>48</sup> Another blue CoMFA and CoMSIA contour next to the meta position of the first biphenyl ether ring corresponds to a favorable hydrogen bond to Val194-C=O and Leu193-C=O. Interestingly this is a region where structural water has been observed for ligands in X-ray structures, which do not fill the S1' pocket. Thus CoMFA indicates the replacement of these water molecules.

To estimate entropic contributions to ligand binding, hydrophobic similarity index fields were analyzed in CoMSIA. Regions where enhanced hydrophilicity is favorable are indicated by cyan contours in Figure 6c (>80% contribution), e.g., the zinc coordinating substituent, the ortho position of the first biphenyl ether ring close to the Pro217 carbonyl oxygen, and the ortho/meta position of the distal ring of the Arg222 side chain. This phenyl ring replaces two or three structural water molecules being observed in X-ray structures with smaller ligand substituents within S1'.

Preferred hydrophobic interactions are indicated by orange regions (<20% contribution) in Figure 6. One region corresponds to the biphenyl ether oxygen close to the His197 imidazole. The hydrophobic cavity formed by Tyr219, Leu193, and Val194 is in accord with CoMFA results. The third important region points to the meta/para position of the distal biphenyl ether ring with the neighboring residues Asn218-Tyr219-Ala220.

The hypothetical binding mode based on literature and our own X-ray data could be confirmed. The binding modes of all other than the reference compound are validated using 3D-QSAR models; their complementarity to the receptor topology is obvious from inspection of Figure 7, showing MMP-8 X-ray structure in complex with the reference inhibitor and CoMFA steric std\*coeff contour maps. It can be seen that favored regions show a good fit to the receptor, while a similar complementarity is observed for electrostatic and hydrophobic interactions.

## Conclusions

Structure-based design is typically focused on the understanding of protein–ligand interactions but does not always lead to predictive models. In contrast 3D-QSAR with acceptable  $r^2(\text{cv})$  values does not necessarily reflect topological features of the protein structure. These models are usually built using alignment rules, which are not always similar to the bioactive conformation. In this study we successfully combined both approaches: Based on the enzyme's X-ray structure, consistent and highly predictive 3D-QSAR models were derived, which could be mapped back to the original protein topology. This leads to a better understanding of important protein–ligand interactions and thus provides guidelines for ligand design plus a predictive model for scoring novel synthetic candidates. 3D-QSAR results based on CoMFA and CoMSIA allow to focus on those regions, where steric, electronic, or hydrophobic effects play a dominant role in ligand–receptor interactions. Extensive testing of the models has validated their robustness, like randomization tests, 100 runs with two randomly selected cross-validation groups and predictions based on maximum dissimilarity designed subsets with different sizes. The comparable results from different 3D-QSAR approaches and their complementarity to the receptor topology further indicate the

validity of these models. Finally the proposed binding mode was validated by X-ray structure analysis.

The application of these models for quantitative prediction of binding affinities is possible within a structurally limited range, as shown using designed training sets. A similarity threshold of 0.75–0.8, expressed as 2D fingerprint-based Tanimoto coefficient, can be estimated (see Figure 4b). Hence, for new candidates within this threshold, reliable binding affinities can be computed by “interpolation”, while less reliable affinities by “extrapolation” are obtained for candidates with a lower similarity to any training set molecule.

While carboxylates are known to bind weaker to MMPs than hydroxamates, they show a better oral bioavailability.<sup>21,49</sup> We could show by interpretation of 3D-QSAR results that inhibitors with weaker zinc binding groups require a compensation by an optimal fit into the S1' pocket,<sup>50,51</sup> which allows to design and predict the affinity of novel candidates with enhanced oral bioavailability. Thus new, highly potent MMP-8 inhibitors may emerge for therapy of several diseases.

**Acknowledgment.** C. Giegerich is gratefully acknowledged for assistance during the implementation of CoMSIA molecular fields. We thank A. Liesum for expert technical assistance and collection of the diffraction data and Prof. R. Huber, Dr. W. Bode, and Dr. F. Grams (MPI Martinsried) for early access to X-ray coordinates of MMP-8 ligand complexes.

## References

- (1) (a) Woessner, J. F., Jr. Matrix metalloproteinases and their inhibitors in connective tissue remodeling. *FASEB J.* **1991**, *5*, 2145–2154. (b) Birkedal-Hansen, H.; Moore, W. G. I.; Bodden, M. K.; Windsor, L. J.; Birkedal-Hansen, B.; DeCarlo, A.; Engler, J. A. Matrix metalloproteinases: a review. *Crit. Rev. Oral Biol. Med.* **1993**, *4*, 197–250. (c) Murphy, G.; Docherty, A. J. P. The matrix metalloproteinases and their inhibitors. *A. J. Res. Cell. Mol. Biol.* **1992**, *7*, 120–125. (d) Matrisian, L. M. Metalloproteinases and their inhibitors in matrix remodeling. *Trends Genet.* **1990**, *6*, 121–125.
- (2) Lohmander, L. S.; Hoernner, L. A.; Lark, M. W. Metalloproteinases, tissue inhibitor, and proteoglycan fragments in knee synovial fluid in human osteoarthritis. *Arthritis Rheum.* **1993**, *36*, 181–189.
- (3) Murphy, G.; Hembry, R. M. Proteinases in rheumatoid arthritis. *J. Rheumatol.* **1992**, *19*, 61–64.
- (4) Peress, N.; Perillo, E.; Zucker, S. Localization of tissue inhibitor of matrix metalloproteinases in Alzheimer's disease and normal brain. *J. Neuropathol. Exp. Neurol.* **1995**, *54*, 16–22.
- (5) Bode, W.; Reinemer, P.; Huber, R.; Kleine, T.; Schnierer, S.; Tschesche, H. The X-ray crystal structure of the catalytic domain of human neutrophil collagenase inhibited by a substrate analogue reveals the essentials for catalysis and specificity. *EMBO J.* **1994**, *13*, 1263–1269.
- (6) (a) Nishino, N.; Powers, J. C. Design of potent reversible inhibitors for thermolysin. Peptides containing zinc coordinating ligands and their use in affinity chromatography. *Biochemistry* **1979**, *18*, 4340–4347. (b) Powers, J. C.; Harper, J. W. Inhibitors of metalloproteinases. In *Proteinase Inhibitors*; Barrett, A. J., Salvesen, G., Eds.; Elsevier: Amsterdam, New York, Oxford, 1986; pp 219–298. (c) Johnson, W. H.; Roberts, N. A.; Borkakoti, N. Collagenase inhibitors: Their design and potential therapeutic use. *J. Enzyme Inhib.* **1987**, *2*, 1–22. (d) Zask, A.; Levin, J. I.; Killar, L. M.; Skotnicki, J. S. Inhibition of Matrix Metalloproteinases: Structure Based Design. *Curr. Pharm. Des.* **1996**, *2*, 624–661.
- (7) Reinemer, P.; Grams, F.; Huber, R.; Kleine, T.; Schnierer, S.; Pieper, M.; Tschesche, H.; Bode, W. Structural implications for the role of the N-terminus in the 'superactivation' of collagenases. *FEBS Lett.* **1994**, *338*, 227–233.
- (8) Stams, T.; Spurlino, J. C.; Smith, D. L.; Wahl, R. C.; Ho, T. F.; Qoronfle, M. W.; Banks, T. M.; Rubin, B. Structure of human neutrophil collagenase reveals large S1' specificity pocket. *Nature Struct. Biol.* **1994**, *1*, 119–123.

- (9) Grams, F.; Reinemer, P.; Powers, J.; Kleine, T.; Pieper, M.; Tschesche, H.; Huber, R.; Bode, W. X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors; implications for substrate binding and rational drug design. *Eur. J. Biochem.* **1995**, *228*, 830–841. The coordinates were obtained prior to submission to PDB due to a collaboration with this group.
- (10) Grams, F.; Crimmin, M.; Hinnes, L.; Huxley, P.; Pieper, M.; Tschesche, H.; Bode, W. Structure determination and analysis of human collagenase complexed with a hydroxamate inhibitor. *Biochemistry* **1995**, *34*, 14012–14020.
- (11) Kroemer, R. T.; Etmayer, P.; Hecht, P. 3D-Quantitative Structure–Activity Relationships of Human Immunodeficiency Virus Type-1 Proteinase Inhibitors: Comparative Molecular Field Analysis of 2-Heterosubstituted Statine Derivatives – Implication for the Design of Novel Inhibitors. *J. Med. Chem.* **1995**, *38*, 4917–4928.
- (12) PDB files obtained from Protein Database (National Brookhaven Laboratories): 1MMB, 1MNC, 1KBC, 1JAN, 1JAO, 1JAP, and 1JAQ (<http://www.pdb.bnl.gov>). Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (13) Cramer, R. D., III; Patterson, D. E.; Bunce, J. E. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (14) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative Molecular Field Analysis (CoMFA). 2. Towards 1st use with 3D-Structural Databases. *Tetrahedron Comput. Methods* **1990**, *3*, 47–59.
- (15) *3D-QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993. This includes many applications and cross-references of the CoMFA methodology in medicinal chemistry.
- (16) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- (17) (a) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindberg, W.; Sjöström, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984; pp 17–95. (b) Dunn, W. J.; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. Multivariate Structure–Activity Relationship Between Data from a Battery of Biological Tests and an Ensemble of Structure Descriptors: The PLS Methodol. *Quant. Struct.-Act. Relat.* **1984**, *3*, 31–137. (c) Geladi, P. Notes on the History and nature of Partial Least Squares (PLS) Modelling. *J. Chemom.* **1988**, *2*, 231–246.
- (18) (a) Wold, S. Cross-Validatory Estimation of the Number of Component in Factor and Principal Component Models. *Technometrics* **1978**, *4*, 397–405. (b) Diaconis, P.; Efron, B. Computer-Intensive Methods for Statistics. *Sci. Am.* **1984**, *116*, 96–117. (c) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. Crossvalidation, Bootstrapping and Partial Least Squares Compared with Multile Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.
- (19) Schlechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
- (20) Residues and residue numbering rely on the Pro HNC nomenclature: Hasty, K. A.; Pourmotabbed, T. F.; Goldberg, G. I.; Thompson, J. P.; Spinella, D. G.; Stevens, R. M.; Mainardi, C. L. Human neutrophil collagenase. A distinct gene product with homology to other matrix metalloproteinases. *J. Biol. Chem.* **1990**, *265*, 11421–11424.
- (21) Hodgson, J. Remodeling MMPs. *Bio/Technology* **1995**, *13*, 554–557.
- (22) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (23) Burley, S. K.; Petsko, G. A. Aromatic–aromatic interaction: a mechanism of protein structure stabilization. *Science* **1985**, *229*, 23–28.
- (24) Schwab, W.; Thorwart, W.; Barbier, D.; Billen, G.; Haase, B.; Matter, H.; Neises, B.; Schudok, M.; Weithmann, K. U. Tetrahydroisoquinoline-3-carboxylate based Matrix Metalloproteinase Inhibitors: Design, Synthesis and Structure–Activity Relationship. *Bioorg. Med. Chem.* **1999**, to be submitted.
- (25) Thorwart, W.; Schwab, W.; Schudok, M.; Haase, B.; Bartnik, E.; Weithmann, K. U. Preparation of cyclic N-substituted  $\alpha$ -imino-hydroxamates as matrix metalloproteinase inhibitors. Ger. Offen., 17 pp; Coden: GWXXBX. DE 19542189 A1 970515. CAN 127:50547.



- (26) Weithmann, K. U.; Schlotte, V.; Jeske, V.; Seiffge, D.; Laber, A.; Haase, B.; Schleyerbach, R. Effects of tiaprofenic acid on urinary pyridinium cross-links in adjuvant arthritic rats: Comparison with doxycycline. *Inflamm. Res.* **1997**, *46*, 246–252.
- (27) Knight, C. G.; Willenbrock, F.; Murphy, G. A novel coumarin-labeled peptide for sensitive continuous assays of the matrix metalloproteinases. *FEBS Lett.* **1992**, *296*, 263–266.
- (28) Kabsch, W. Evaluation of single-crystal X-ray diffraction data from a position sensitive detector. *J. Appl. Crystallogr.* **1988**, *21*, 916–924.
- (29) Brünger, A. T. X-PLOR version 3.1, A system for X-ray crystallography and NMR; Yale University Press: New Haven, CT, 1992.
- (30) SYBYL Molecular Modelling Package, version 6.33; Tripos: St. Louis, MO, 1996.
- (31) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982–1912.
- (32) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219–3228. Details of the implementation are given in: *Sybyl 6.3 Theory Manual*; Tripos: St. Louis, MO, 1996; p 67.
- (33) (a) Ghose, A.; Crippen, G. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure–Activity Relationships. 1. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565–577. (b) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure–Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- (34) Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Method* **1990**, *3*, 615–633.
- (35) (a) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A.; Rognan, D. Recommendations for CoMFA Studies and 3D QSAR Publications. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 711–717. (b) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and Limitations. *Ibid.*; pp 583–616. (c) Cramer, R. D., III; DePriest, S. A.; Patterson, D. E.; Hecht, P. The Developing Practice of Comparative Molecular Field Analysis. *Ibid.*; pp 443–485.
- (36) Sheridan, R. P.; Nachbar, R. B.; Bush, B. L. Extending the trend vector: The trend matrix and sample-based partial least squares. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 323–340.
- (37) For details to compute fingerprints, see UNITY Chemical Information Software, Version 3.0, Reference Guide; Tripos Inc., 1699 S Hanley Rd, St. Louis, MO 63144.
- (38) (a) Willett, P. *Similarity and Clustering in Chemical Information Systems*; Research Studies Press: Letchworth, 1987. (b) Willett, P.; Winterman, V. A. Comparison of Some Measures for the Determination of Intermolecular Structural Similarity. *Quant. Struct.-Act. Relat.* **1986**, *5*, 18–25.
- (39) Lajiness, M.; Johnson, M. A.; Maggiora, G. M. Implementing Drug Screening Programs using Molecular Similarity Methods. In *QSAR: Quantitative Structure–Activity Relationships in Drug Design*; Fauchere, J. L., Ed.; Alan R. Liss Inc.: New York, 1989; pp 173–176.
- (40) Taylor, R. Simulation Analysis of Experimental Design Strategies for Screening Random Compounds as Potential New Drugs and Agrochemicals. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 59–67.
- (41) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (42) Biological activities are expressed in general as  $\log(1/IC_{50} \times 100000)$ .
- (43) Clark, M.; Cramer, R. D., III. The Probability of Chance Correlation using Partial Least Squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137–145.
- (44) G. Cruciani, S. Clementi, personal communication.
- (45) Carpignano, R.; Dolci, M.; Scarfi, D. Design of more informative molecules for QSAR study in a series of juvenile hormones. In *Trends QSAR Mol. Modelling 92, Proc. Eur. Symp. Struct.-Act. Relat.*; Wermuth, C.-G., Ed.; ESCOM: Leiden, The Netherlands, 1993.
- (46) (a) Norinder, U.; Hoegberg, T. PLS-based quantitative structure–activity relationship for substituted benzamides of clebopride type. Application of experimental design in drug research. *Acta Chem. Scand.* **1992**, *46*, 363–366. (b) Norinder, U. Experimental design-based quantitative structure–toxicity relationship of some local anesthetics using the PLS method. *J. Appl. Toxicol.* **1992**, *12*, 143–147. (c) Norinder, U. An experimental design based quantitative structure–activity relationship study on  $\beta$ -adrenergic blocking agents using PLS. *Drug Des. Discov.* **1991**, *8*, 127–136. (d) Norinder, U. Experimental design based 3D QSAR analysis of steroid–protein interaction: application to human CBG complexes. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 381–389.
- (47) Pötter, T.; Matter, H. Random or Rational Design? Evaluation of Diverse Compound Subsets from Chemical Structure Databases. *J. Med. Chem.* **1998**, *41*, 478–488.
- (48) Betz, M.; Huxley, P.; Davies, S. J.; Mushtaq, Y.; Pieper, M.; Tschesche, H.; Bode, W.; Gomis-Rüth, F. X. 1.8-A crystal structure of the catalytic domain of human neutrophil collagenase (matrix metalloproteinase-8) complexed with a peptidomimetic hydroxamate primed-side inhibitor with a distinct selectivity profile. *Eur. J. Biochem.* **1997**, *247*, 356–363.
- (49) Chapman, K. T.; Durette, P. L.; Caldwell, C. G.; Sperow, K. M.; Niedzwiecki, L. M.; Harrison, R. K.; Saphos, C.; Christen, A. J.; Olszewski, J. M.; Moore, V. L.; Maccoss, M.; Hagmann, W. K. Orally Active Inhibitors of Stromelysin-1 (MMP-3). *Bioorg. Med. Chem. Lett.* **1996**, *6*, 803–806.
- (50) (a) Caldwell, C. G.; Sahoo, S. P.; Polo, S. A.; Eversole, R. R.; Lanza, T. J.; Mills, S. G.; Niedzwiecki, L. M.; Izquierdo-Martin, M.; Chang, B. C.; Harrison, R. K.; Kuo, D. W.; Lin, T. Y.; Stein, R. L.; Durette, P. L.; Hagmann, W. K. Phosphinic Acid Inhibitors of Matrix Metalloproteinases. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 323–328. (b) Chapman, K. T.; Wales, J.; Sahoo, S. P.; Niedzwiecki, L. M.; Izquierdo-Martin, M.; Chang, B. C.; Harrison, R. K.; Stein, R. L.; Hagmann, W. K. Inhibition of Matrix Metalloproteinases by P1 Substituted N-Carboxyalkyl Dipeptides. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 329–332. (c) Beeley, N. R. A.; Ansell, P. R. J.; Docherty, A. J. P. *Curr. Opin. Ther. Pat.* **1994**, *4*, 7–17.

JM980631S