

Highly Predictive CoMFA and CoMSIA Models for Two Series of Stromelysin-1 (MMP-3) Inhibitors Elucidate S1' and S1–S2' Binding Modes

Elizabeth A. Amin^{*,†} and William J. Welsh[‡]

Department of Chemistry and Minnesota Supercomputing Institute, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455, and Department of Pharmacology, RWJMS-UMDNJ, 675 Hoes Lane, Piscataway, New Jersey 08854

Received March 15, 2006

Three-dimensional quantitative structure–activity relationship models have been derived using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) for two training sets of arylsulfonyl isoquinoline-based and thiazine/thiazepine-based matrix metalloproteinase inhibitors (MMPi). The crystal structure of stromelysin-1 (MMP-3) was used to pinpoint areas on the ligands and receptors where steric and electrostatic effects (for CoMFA) and steric, electrostatic, hydrogen-bond donor, hydrogen-bond acceptor, and hydrophobic effects (for CoMSIA) correlate with an increase or decrease in experimental biological activity. The most predictive CoMFA and CoMSIA models were obtained using training-series subsets that sampled a wide range of activities, together with docking and scoring, inertial alignment, investigation of various partial charge formalisms, and manual adjustment of each compound within the active site. The models developed in this study are in agreement with experimentally observed MMP-3 structure–activity relationship data and offer new insights into binding modes involving the partly solvent-exposed S1–S2' subpocket and certain zinc-chelating groups.

INTRODUCTION

Matrix metalloproteinases (MMPs) are a structurally related family of key zinc-binding endopeptidases responsible for connective tissue remodeling in normal and pathological conditions. Expressed as zymogens, they degrade many extracellular matrix proteins, including gelatin, fibronectin, laminin, basement membrane, interstitial collagens, and proteoglycan.^{1–5} MMPs are active in a number of disease processes including tumor metastasis,^{6,7} periodontal disease,⁸ multiple sclerosis,⁹ congestive heart failure,¹⁰ and degenerative diseases such as osteo- and rheumatoid arthritis.¹¹ Stromelysin-1 (MMP-3) is of particular interest as it hydrolyzes critical extracellular matrix proteins that comprise more than 70% of human cartilage,¹² plays a key role in the spread of metastatic tumors associated with melanoma and breast cancer,¹³ and acts as a precursor to the activity of other endopeptidases.¹⁴ MMPs have, therefore, garnered a great deal of attention as attractive targets for rational drug design, and successful small-molecule inhibitor development would contribute to elucidating their role in processes of normal tissue growth and repair as well as in the disease models noted above.¹⁵ While many classes of matrix metalloproteinase inhibitors (MMPIs) have been designed and synthesized,^{16–20} safety and bioavailability concerns continue to hamper their success in clinical trials.²¹ Advances in computer-aided ligand-based design methods such as those described in the present work can more closely delineate the structural features and binding characteristics of the MMP active sites and thereby minimize MMPI specificity-related side effects.

Matter and Schwab of Hoechst Marion Roussel, Inc., designed and synthesized a congeneric series of active aryl-

sulfonyl isoquinoline MMP inhibitors which display varying degrees of selectivity between MMP-3 and MMP-8.²² Overall inhibitor design was based on established X-ray binding site geometries that point to the S1' pocket as the primary determinant of enzyme specificity. These inhibitors were built on a 1,2,3,4-tetrahydroisoquinoline scaffold, incorporated N-substituted sulfonyl moieties complementary to the MMP-3 S1' binding pocket, and included both hydroxamates and carboxylates as zinc-chelating groups. The sulfonamide grouping, present in all inhibitors, was chosen because of its effectiveness as a hydrogen-bond acceptor. These workers conducted several quantitative structure–activity relationship (QSAR) analyses²² assessing MMP-3 and MMP-8 specificity. Additionally, the synthesis and biological evaluation of a series of highly potent MMPIs based on thiazine and thiazepine skeletons was reported by Almstead and co-workers at Procter & Gamble, Inc.²³ All inhibitors from this set incorporated a hydroxamate zinc-binding group and a sulfonamide moiety attached at the thiazine/thiazepine N1 position. The size and substituents of the thiazepine/thiazine ring were varied to assess binding to the S1–S2' region. A *p*-methoxyphenyl group was attached to the sulfonamide moiety in the majority of inhibitors but was replaced by various aryl groups in a small subset of compounds to further evaluate the nature of the S1' subsite.

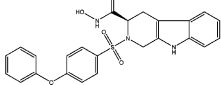
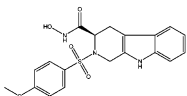
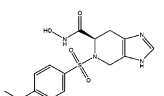
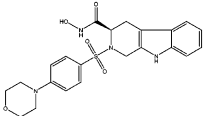
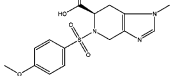
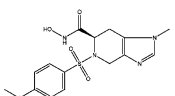
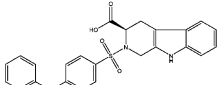
The lack of reliable force-field-bonded parameters for the active-site catalytic zinc(II) cation complicates the molecular modeling of matrix metalloproteinases. Innovative attempts to address this problem include the calculation of binding free energies²⁴ and development of new functional forms;²⁵ however, accurate molecular mechanics binding energies remain elusive. As ligand-based techniques, three-dimensional quantitative structure–activity relationship (3D-QSAR) methods require no initial active-site information and can be used to investigate a variety of enzyme–inhibitor

* Corresponding author phone: 612-626-2387; fax: 612-625-7541; e-mail: amin@chem.umn.edu.

[†] University of Minnesota.

[‡] RWJMS-UMDNJ.

Table 1. Compounds 1–7 (Subset A) of the Training Set of Arylsulfonyl Isoquinoline MMPiS

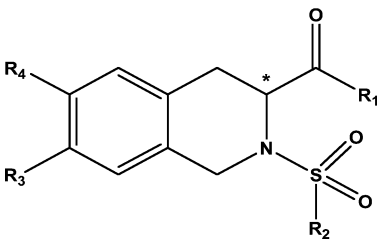
1		7.222
2		7.699
3		7.699
4		6.523
5		5.000
6		7.699
7		5.699

interactions for which an accurate structural rendering of the catalytic domain is unavailable or (as in the present work) a targeted metal center exists which is difficult to model. On the basis of the assumption that properties such as binding affinity are shape- and structure-dependent, 3D-QSAR methods can be utilized to efficiently screen potential drug candidates as well as improve the potency of existing compounds. Comparative molecular field analysis (CoMFA)²⁶ and comparative molecular similarity indices analysis (CoMSIA),²⁷ regarded as industry standards for constructing 3D-QSAR models, have been established in many studies^{28–32} as versatile and powerful tools in rational drug design and related applications. CoMFA, based on the assumption that enzyme–inhibitor interactions are primarily shape-dependent and noncovalent, derives a 3D-QSAR model by systematically sampling the steric and electrostatic fields surrounding a compound set and correlating differences in these fields with experimental biological activity. CoMSIA models are obtained by calculating Gaussian-function similarity indices that represent steric, electrostatic, hydrophobic, and hydrogen-bond donor and hydrogen-bond acceptor interactions and correlating these fields with experimental activity. Together, these two methods cover the chief contributions to ligand binding and can map previously unmodeled active-site properties as well as clarify mechanisms of action.

MATERIALS AND METHODS

All molecular modeling calculations and visualizations outlined herein were performed on Silicon Graphics O2 workstations running under the IRIX 6.5 operating system and on an Alienware MJ-12 dual-CPU workstation running under the SuSE Linux Professional 9.3 OS. Training-set

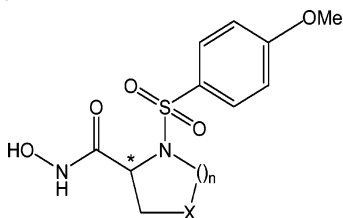
Table 2. Compounds 8–39 of the Training Set of Arylsulfonyl Isoquinoline MMPiS



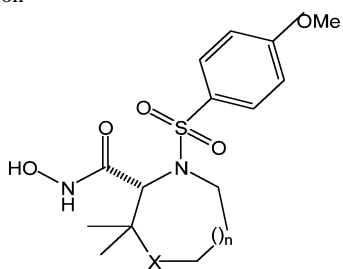
cpd #	isomer	R ₁	R ₂	R ₃	R ₄	MMP-3 pIC ₅₀
8	R	NH–OH	4-phenoxybenzene	H	H	7.699
9	R	NH–OH	benzene	H	H	6.523
10	R	NH–OH	phenylethyl	H	H	7.000
11	R	NH–OH	4-methoxybenzene	NO ₂	H	7.097
12	R	NH–OH	benzyl	H	H	5.000
13	S	NH–OH	4-methoxybenzene	H	H	5.155
14	R	NH–OH	4-methoxybenzene	H	H	7.699
15	R	NH–OH	4-biphenyl	H	H	7.699
16	R	OH	4-methoxybenzene	H	H	5.000
17	R	NH–OH	4-methoxybenzene	NH ₂	H	8.000
18	R	NH–OH	4-butoxybenzene	H	H	7.523
19	R	NH–OH	4-methoxybenzene	OH	H	8.000
20	R	NH–OH	4-trifluoromethylbenzene	H	H	6.301
21	R	OH	4-methoxybenzene	NH ₂	H	5.000
22	R	OH	4-chlorobiphenyl	H	H	7.000
23	R	NH–OH	4-chlorobiphenyl	H	H	7.699
24	R	OH	4-phenoxybenzene	H	H	6.699
25	R	OH	4-chlorobiphenyl	NO ₂	H	6.222
26	R	OH	4-chlorobiphenyl	H	NO ₂	6.301
27	R	OH	4-chlorobiphenyl	NH ₂	H	6.699
28	R	OH	4-biphenyl	H	H	6.301
29	R	OH	4-fluorobiphenyl	NO ₂	H	6.000
30	R	OH	4-fluorobiphenyl	H	NO ₂	6.301
31	R	OH	4-methoxybenzene	OH	H	5.000
32	S	OH	4-chlorobiphenyl	H	H	5.000
33	R	OH	4-fluorobiphenyl	H	H	6.398
34	R	OH	4-fluorobiphenyl	H	NH ₂	6.699
35	S	OH	4-fluorobiphenyl	H	H	5.000
36	R	NH–OH	4-fluorobiphenyl	NH ₂	H	7.699
37	R	OH	4-chlorobiphenyl	OH	H	6.699
38	R	OH	4-(4-trifluoromethyl)-biphenyl	H	H	6.523
39	R	OH	4-bromophenyl	H	H	5.301

compounds were constructed in the SPARTAN '02 Linux/UNIX molecular modeling software.³³ Initial geometry optimizations were performed on each molecule using the SPARTAN PM3 semiempirical methodology. The optimized data set compounds were imported into the SYBYL 6.7 and 7.0 discovery software packages,³⁴ in which the subsequent CoMFA and CoMSIA calculations were performed.

Data Sets. The initial arylsulfonyl isoquinoline training set comprised 39 compounds whose biological activity (reported as IC₅₀) with respect to MMP-3 ranged from 10 to ~10 000 nM. Inhibitors 1–7 included fairly rigid two- and three-ring structures adjacent to the zinc-binding group, whereas compounds 8–39 incorporated a more flexible structure with a wide variety of substituents (Tables 1 and 2). Functional group variation at the R₂ position was designed to test compatibility with the S1' pocket. Zinc-binding groups were alternated between hydroxamate and carboxylate at the R₁ position, and smaller changes were made at the R₃ and R₄ positions to evaluate binding in the S1–S3 regions. The second training set encompassed 35 compounds differing primarily in thiazine/thiazepine ring variations and sulfonamide aryl substituents (Tables 3–6). Following Almstead and co-workers, the compounds are tabulated to highlight the effects of these changes on otherwise similar structures. The biological activity (reported as IC₅₀) of the inhibitor set with respect to MMP-3 ranges from 0.7 to 737 nM, with

Table 3. Partial Training Set of Thiazine- and Thiazepine-Based MMPIs, with the Effect of Ring Size and Ring Heteroatoms on MMP-3 Inhibition


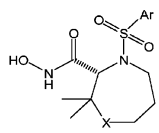
cpd #	<i>n</i>	X	MMP-3 pIC ₅₀
37	1	CH ₂	6.588
38	1	S	6.757
39	2	CH ₂	7.854
40	2	O	8.237
41	2	S	7.824
42	2	SO ₂	6.133
43	2	NH	7.319
21	3	CH ₂	8.155
44	3	S	8.367

Table 4. Partial Training Set of Thiazine- and Thiazepine-Based MMPIs, with the Effect of *gem*-Dimethyl Groups and Ring Size on MMP-3 Inhibition


cpd #	<i>n</i>	X	MMP-3 pIC ₅₀
9	1	CH ₂	9.155
10	1	S	7.734
4a	2	CH ₂	9.155
5a	2	O	8.161
17	2	S	8.481
22	2	SO ₂	8.143

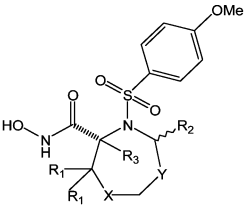
structures incorporating a seven-membered thiazepine ring and a *p*-methoxyphenyl moiety demonstrating the highest activity against that enzyme.

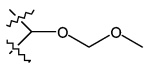
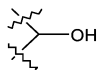
Initial CoMFA and CoMSIA Strategies. For the aryl-sulfonyl isoquinolines, the crystal structure 1BZS³⁵ containing compound 28 provided the alignment templates for all remaining molecules. For the thiazine- and thiazepine-based ligands, the crystal structure of MMP-3 coupled with the inactive R isomer of compound 4a (1D5J²³) was used to build the active S isomer and thereby obtain a template molecule for the construction and alignment of the remaining data series. Gasteiger–Hückel and Gasteiger–Marsili charges were calculated for each compound in the two data sets. Our previously published QSAR investigation of piperazine-based MMPIs³⁶ demonstrated that matching the crystal-structure conformation and docking each compound into the active site were essential prerequisites to obtain likely molecular conformations and alignment conducive to a predictive QSAR model. Prior to inertial alignment, each compound was therefore built directly from the respective crystal structure, matching the bound conformation, and then flexibly docked into the MMP-3 binding groove using the FlexX/CScore docking and scoring modules.³⁷ (FlexX incorporates SYBYL atom typing, which includes parameters for Zn.

Table 5. Partial Training Set of Thiazine- and Thiazepine-Based MMPIs, with the Effect of Various Aryl Substituents on MMP-3 Inhibition


cpd #	X	Ar	MMP-3 pIC ₅₀
4a	S		9.155
5a	SO ₂		8.161
4b	S		8.000
5b	SO ₂		7.046
4c	S		7.056
4d	S		8.180
5d	SO ₂		8.569
4e	S		7.456
4f	S		7.602
4g	S		7.770
4h	S		7.959
4i	S		8.959
4j	S		8.201
4k	S		8.187

While these are not highly accurate bonded parameters and cannot be relied upon for quantitative energetics, they proved sufficient to obtain docked ligand conformations for use in subsequent QSAR modeling, as has been previously validated.³⁶ Following standard CoMFA and CoMSIA procedures, each MMPI was inserted into a three-dimensional lattice with grid points separated by 2.0 Å. For the CoMFA analysis, the steric (van der Waals) and electrostatic (Coulombic) field descriptors were calculated at all lattice points by summing the individual energy interactions between each atom of the MMPI molecule and an sp³ carbon probe atom with a +1 charge. A distance-dependent dielectric function ($\epsilon = \epsilon_0/r$, with $\epsilon_0 = 1$) was employed, and maximum field values were truncated to 30 kcal/mol for the steric fields and to ± 30 kcal/mol for the electrostatic fields. CoMSIA analyses were conducted focusing on hydrophobic and hydrogen-bond donor/acceptor fields, with similarity indices computed using an sp³ carbon with a +1 charge, +1 radius, and +1 hydrophobicity. The attenuation factor α was set to 0.3 for the Gaussian-type distance dependence. Throughout this study, we applied the unique “divide-and-conquer” strategy validated in our previous work,³⁶ in which various ligand conformations, alignment schemes, inhibitor subsets,

Table 6. Partial Training Set of Thiazine- and Thiazepine-Based MMPs, with the Effect of Various Ring Substituents on MMP-3 Inhibition


cpd #	R ₁	X	Y	R ₂	R ₃	MMP-3 pIC ₅₀
4a	Me	S	-CH ₂ -	H	H	9.155
34	Me	S		H	H	9.000
35	Me	SO ₂		H	H	8.137
23	Me	S	-CH ₂ OCH ₂ -	H	H	9.155
24	Me	SO ₂	-CH ₂ OCH ₂ -	H	H	7.347
36a	Me	S	-CH ₂ -	α-Ph	H	7.585
36b	Me	S	-CH ₂ -	β-Ph	H	7.481
45	H	S	-CH ₂ -	H	Me	7.509

partial charge formalisms, and column filtering levels were employed in order to optimize the derived models.

Regression Analysis. Linear regression equations were obtained by the partial least-squares technique (PLS),³⁸ correlating changes in the steric and electrostatic fields (for CoMFA) and similarity fields (for CoMSIA) with changes in the IC₅₀ values of the inhibitor sets. PLS reduces the dimensionality of the problem by constructing latent variables that are composed of linear combinations of the original independent variables. “Leave-one-out” cross-validation was used to assess the internal predictive ability of the CoMFA and CoMSIA models. In this technique, compounds are systematically excluded from the data set, and the activity of each removed compound is predicted by a new model derived from the remaining compounds in the set. Cross-validation yielded the optimum number of principal components, or PCs, together with the highest cross-validated r^2 (r_{cv}^2) value. The PLS analyses were then repeated without cross-validation using the optimum number of PCs, giving final CoMFA and CoMSIA models from which the conventional r^2 values, non-cross-validated standard errors, F ratios,³⁹ and related statistical parameters were computed. Column filtering, for any column of computed energies with a variation less than 2.0 kcal/mol, was applied as needed to reduce computation time without negatively affecting the quality of the models. Each CoMFA and CoMSIA model

was represented as a color contour map depicting regions of descriptor fields that contribute significantly to that model.

RESULTS AND DISCUSSION

Arylsulfonyl Isoquinolines. The experimental IC₅₀ data obtained by Matter and Schwab led to several preliminary structure–activity observations. As expected, hydroxamate is a more successful zinc-binding group than carboxylate, and the presence of hydroxamate appears to contribute greatly to overall binding affinity. The QSAR studies done by these workers²² indicate that, in compounds 8–39, the stereochemistry at the sole chiral carbon also affects ligand binding, with the R configuration preferred over the S, and highlights the steric requirements at the metal coordination center. Hydrophobic substituents with limited steric bulk are favored at the R₂ position, which tallies with the known structural characteristics of the hydrophobic S1' area. Substitution of small groups at the R₃ and R₄ positions enhances biological activity, with R₄ substitution increasing binding affinity more than with R₃.

Aligning the complete data set of 39 compounds to inhibitor 28 yielded internally predictive CoMFA models in terms of cross-validated r^2 (0.582) and conventional r^2 (0.807). However, the standard error of estimate for this simulation was high (0.467) and the F ratio low (33.391). The use of Gasteiger–Marsili charges, which proved advantageous in our previous work,³⁶ here resulted in a less predictive model ($r_{cv}^2 = 0.445$, $r^2 = 0.696$, $F = 38.930$, and $SE = 0.568$). Ligands 1–7, designated subset A, exhibited a conformationally restrained ring structure unlike that found in the remainder of the data set. The removal of these compounds from the data set resulted in an increase in internal predictive ability ($r_{cv}^2 = 0.607$) but a slight decrease in self-consistency ($r^2 = 0.792$). A higher F ratio (55.304) and standard error of estimate (0.471) were also obtained. Residual graphs of the first analysis revealed a number of outlier points denoting compounds for which a large difference exists between the experimental and CoMFA-predicted biological activity. The compounds represented by these points, designated subset B, were 16, 21, and 31; all of these contain carboxylate zinc-binding groups and demonstrate virtually no activity against MMP-3 (IC₅₀ ~ 10 000 nM). The removal of subset B and compound 10, a slight outlier, resulted in a significantly more predictive CoMFA model ($r_{cv}^2 = 0.614$, $r^2 = 0.968$, $F = 132.103$, and $SE = 0.180$). Omitting subset A and subset B from the training set, along with 10, yielded the best model of all ($r_{cv}^2 = 0.738$, $r^2 = 0.982$, $F = 202.524$, and $SE = 0.139$). The CoMSIA models for this training set did not prove as predictive as the CoMFA models. The best CoMSIA model ($r_{cv}^2 = 0.608$, $r^2 = 0.835$, $F = 21.986$, and $SE = 0.409$) was obtained using Gasteiger–Hückel charges and by removing subset B and compound 10 from the original data set. The experimental and predicted biological activities of each subset I compound corresponding to the best CoMFA and CoMSIA models are listed in Tables 7 and 8 and plotted in Figures 1 and 2.

The most predictive CoMFA and CoMSIA models were applied to derive the biological activity values of four MMPs originally omitted from the training set: T1–T4, ranging from inactive (pIC₅₀ = 5.000) to moderately active (pIC₅₀ = 7.699) (Table 9). The CoMSIA-predicted activities of the

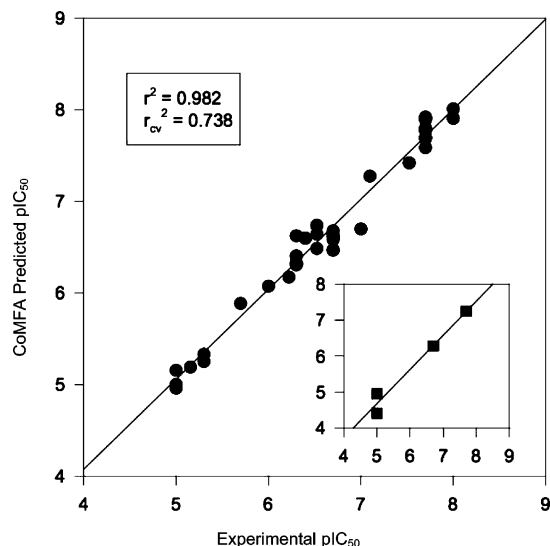


Figure 1. Plot of the CoMFA-predicted vs experimental biological activities for the partial training set (subsets A and B omitted) of arylsulfonyl isoquinoline MMPiS, using Gasteiger–Hückel charges. The inset shows four test compounds.

Table 7. Comparison of Experimental and CoMFA-Predicted Biological Activity Values (Subsets A and B and Compound 10 Omitted, Gasteiger–Hückel Charges) of Arylsulfonyl Isoquinoline MMPiS

cpd #	experimental (pIC ₅₀)	predicted (pIC ₅₀)	residual
8	7.699	7.686	0.013
9	6.523	6.484	0.039
11	7.097	7.274	-0.177
13	5.155	5.189	-0.034
14	7.699	7.797	-0.098
15	7.699	7.694	0.005
17	8.000	8.007	-0.007
18	7.523	7.418	0.105
19	8.000	7.904	0.096
20	6.301	6.326	-0.025
22	7.000	6.696	0.304
23	7.699	7.770	-0.071
24	6.699	6.627	0.072
25	6.222	6.171	0.051
26	6.301	6.403	-0.102
27	6.699	6.607	0.092
28	6.301	6.622	-0.321
29	6.000	6.073	-0.073
30	6.301	6.305	-0.004
32	5.000	5.000	0.000
33	6.398	6.598	-0.200
34	6.699	6.675	0.024
35	5.000	4.959	0.041
36	7.699	7.585	0.114
37	6.699	6.469	0.230
38	6.523	6.640	-0.117
39	5.301	5.329	-0.028

test set compounds were, except in the case of compound T3, much closer to the experimental values than those predicted by CoMFA. The color contour map of the most predictive CoMFA model together with compound 8 (MMP-3 pIC₅₀ = 8.000) are shown in Figure 3. This map was inserted into the MMP-3 binding pocket to associate the colored polyhedra with particular regions and residues within the binding groove.

This CoMFA model illustrates several broadly defined locations that offer generalized yet accurate information about the steric and electrostatic features of the binding site. The

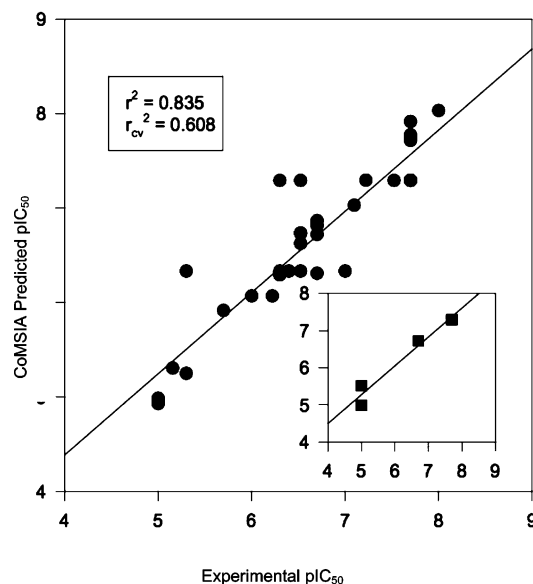


Figure 2. Plot of the CoMSIA-predicted vs experimental biological activities for the partial training set (subset B and compound 10 omitted) of arylsulfonyl isoquinoline MMPiS, using Gasteiger–Hückel charges. The inset shows four test compounds.

Table 8. Comparison of Experimental and CoMSIA-Predicted Biological Activity Values (Subset B and Compound 10 Omitted, Gasteiger–Hückel Charges) of Arylsulfonyl Isoquinoline MMPiS

cpd #	experimental (pIC ₅₀)	predicted (pIC ₅₀)	residual
1	7.222	7.296	-0.076
2	7.699	7.296	0.404
3	7.699	7.747	-0.047
4	6.523	6.627	-0.107
5	5.000	4.987	0.013
6	7.699	7.717	-0.017
7	5.699	5.917	-0.217
8	7.699	7.294	0.406
9	6.523	7.294	-0.774
11	7.097	7.032	0.068
13	5.155	5.307	-0.147
14	7.699	7.294	0.406
15	7.699	7.294	0.406
17	8.000	8.035	-0.035
18	7.523	7.294	0.226
20	6.301	7.294	-0.994
22	7.000	6.334	0.666
23	7.699	7.294	0.406
24	6.699	6.311	0.389
25	6.222	6.071	0.149
26	6.301	6.296	0.004
27	6.699	6.816	-0.116
28	6.301	6.334	-0.034
29	6.000	6.071	-0.071
30	6.301	6.296	0.004
32	5.000	4.933	0.067
33	6.398	6.334	0.066
34	6.699	6.868	-0.168
35	5.000	4.933	0.067
36	7.699	7.776	-0.076
37	6.699	6.722	-0.022
38	6.523	6.334	0.186
39	5.301	6.334	-1.034

two large yellow areas near the distal end of the R₂ substituent denote the correlation of steric bulk with decreased binding affinity; they highlight the steric requirements of the narrow and constrained S1' pocket entrance, which is formed by Pro217, His207, His197, and Val194. Beyond these yellow areas lies a fairly large blue region

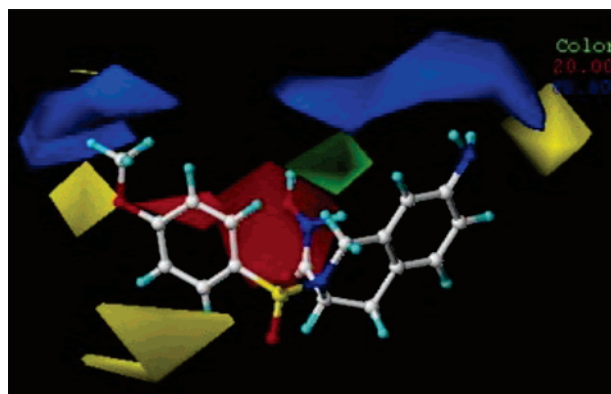


Figure 3. CoMFA contour map for the partial training set (subsets A and B omitted) of arylsulfonyl isoquinoline MMPis, using Gasteiger–Hückel charges and shown with compound 8 ($\text{pIC}_{50} = 8.000$). Colored polyhedra represent areas on or near the ligand where properties correlate strongly with an increase in binding affinity. Red = negative electrostatic potential; blue = positive electrostatic potential; yellow = negative steric potential; green = positive steric potential.

Table 9. Comparison of Experimental, CoMFA-Predicted (Subsets A and B Omitted, Gasteiger–Hückel Charges), and CoMSIA-Predicted (Subset B and Compound 10 Omitted, Gasteiger–Hückel Charges) Biological Activity Values for Arylsulfonyl Isoquinoline Test Set Compounds

cpd #	experimental (pIC_{50})	predicted (pIC_{50})	residual
T1 CoMFA	5.000	4.411	0.589
T2 CoMFA	6.699	6.280	0.419
T3 CoMFA	7.699	7.246	0.453
T4 CoMFA	5.000	4.955	0.045
T1 CoMSIA	5.000	4.987	0.013
T2 CoMSIA	6.699	6.722	−0.023
T3 CoMSIA	7.699	7.294	0.405
T4 CoMSIA	5.000	5.515	−0.515

representing the hydrophobic S1' pocket itself. The medium-sized green area adjacent to the hydroxamate indicates a low tolerance for steric bulk near the catalytic zinc, as well as confirms the preferred stereochemistry at carbon C₃; the inversion of this chiral center would orient the aromatic isoquinoline ring toward this green area. A long blue region and a smaller yellow polygon are found in the area corresponding to the S1–S2' subsite, pointing to a correlation between positive electrostatic potential, low steric bulk, and biological affinity in that area.

Although the best CoMSIA model for this training set was not as predictive as CoMFA, the CoMSIA contour map offers valuable information as to the hydrogen-bonding environment of the MMP-3 active site (Figure 4), especially the zinc-chelating region. The two cyan regions near the protonated hydroxamate nitrogen point to favorable hydrogen-bond donation from the hydroxamate NH to the Ala165 backbone carbonyl as well as from the hydroxamate OH to the Glu202 charged side chain, illustrating the superiority of hydroxamate over carboxylate as a zinc-binding group. Moreover, a medium-sized magenta region directly behind the hydroxamate carbonyl underscores the necessity of a hydrogen-bond acceptor at this location to engage in zinc binding. Finally, large magenta polygons adjacent to the sulfonamide moiety and near the hydrophobic entrance of the S1' subsite indicate that hydrogen-bond donors are disfavored in this area.

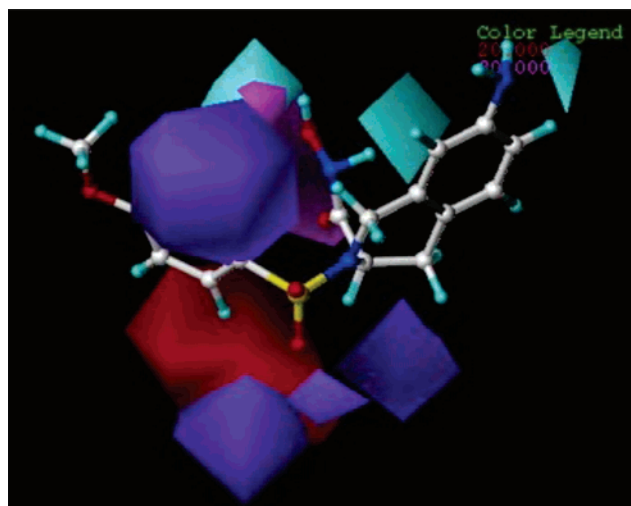


Figure 4. CoMSIA contour map for the partial training set (subset B and compound 10 omitted) of arylsulfonyl isoquinoline MMPis, using Gasteiger–Hückel charges and shown with compound 8 ($\text{pIC}_{50} = 8.000$). Colored polyhedra represent areas on or near the ligand where hydrogen bonding correlates strongly with binding affinity. Cyan = hydrogen-bond donors favored; purple = hydrogen-bond donors disfavored; magenta = hydrogen-bond acceptors favored; red = hydrogen-bond acceptors disfavored.

Thiazine- and Thiazepine-Based Compounds. X-ray crystallographic structural data provided by Almstead and co-workers reveal that the aryl (usually methoxyphenylsulfonamide) group is strongly attached to the S1' site, most likely forming a hydrogen bond with Leu164, while the hydroxamate NH hydrogen-bonds to Ala165. This is consistent with our CoMFA and CoMSIA results obtained from the first training set as well as with previously obtained structural information.^{22,36} The inactivity of the R isomer of compound 4a (which was not included in the CoMFA/CoMSIA training set) is due to unfavorable interactions between the hydroxamate group and the catalytic zinc; significant reorganization of the ligand and substrate would have to occur in order for chelation to take place.²³ Experimental binding-affinity data indicates that the thiazepines inhibit MMP-3 much more strongly than do the thiazines; the MMP3–4a crystal structure shows that the thiazepine is oriented in a chairlike conformation with the ring sulfur and one methyl group facing Val163. Adding substituents to the thiazepine ring did not affect biological activity as much as the researchers expected, which is not surprising because the anterior S1–S2' area is open and solvent-exposed. The addition of a *gem*-dimethyl functionality on the thiazepine ring, however, radically increases potency. The thiazepine ring has been hypothesized to undergo hydrophobic interactions with the Pro221 ring structure, and the increased activity of the methyl-substituted thiazepines has been traced to a second strong hydrophobic interaction between CH₃ and Val163.²³ The crystal structure also reveals the proximity of Phe210 and Phe186 to the thiazepine ring, which supports our previous conclusion³⁶ that these two hydrophobic residues play an important role in ligand binding to the S1–S2' area.

A CoMFA analysis of the complete data set, using Gasteiger–Marsili charges and 2.0 column filtering, was only mildly predictive ($r_{\text{cv}}^2 = 0.207$, $r^2 = 0.881$, $F = 41.622$, and $\text{SE} = 0.274$). Three compounds with large residual values were omitted from the training set: 5d (0.564), 43

($-0.649\ 65$), and 17 ($-0.504\ 337$); the obtained CoMFA model was greatly improved ($r_{cv}^2 = 0.490$, $r^2 = 0.976$, $F = 178.016$, and $SE = 0.123$). Eliminating column filtering from the PLS analysis of this model resulted in a slight loss of predictive ability ($r_{cv}^2 = 0.408$, $r^2 = 0.974$, $F = 150.208$, and $SE = 0.134$), as did the use of Gasteiger–Hückel charges ($r_{cv}^2 = 0.454$, $r^2 = 0.949$, $F = 93.005$, and $SE = 0.184$); a column filtering of 2.0 and Gasteiger–Marsili charges were therefore used for all remaining analyses of this data set.

Because of a radical change in the thiophene–pyridine ring pair conformation in compound 4i during docking studies, this compound did not seem to align well with the remaining data set. Omitting compound 4i from the complete set, however, yielded a very poor CoMFA model ($r_{cv}^2 = -0.029$, $r^2 = 0.243$, $F = 9.959$, and $SE = 0.635$). Compounds 4c and 5b were slight outliers and comprised two of the three compounds in this series incorporating a *p*-bromophenyl aryl substituent. Compound 4c in particular did not dock well into the binding site, despite a moderately high experimental biological activity value ($pIC_{50} = 7.056$); this was probably due to steric conflicts encountered in the docking procedure involving the *o*-methyl group on the phenyl ring. Removing compound 5b from the partial training set which yielded the best model to date did not improve upon that model ($r_{cv}^2 = 0.457$, $r^2 = 0.979$, $F = 93.005$, and $SE = 0.184$); however, removing 4c resulted in a significant increase in internal predictivity ($r_{cv}^2 = 0.522$, $r^2 = 0.985$, $F = 256.039$, and $SE = 0.101$). Excluding both 5b and 4c from the partial training set produced slightly less favorable results ($r_{cv}^2 = 0.458$, $r^2 = 0.985$, $F = 239.388$, and $SE = 0.101$). The CoMSIA models for this training set were not predictive; the “best” CoMSIA model only yielded an r_{cv}^2 of -0.103 and an r^2 of 0.157, with an extremely high standard error of estimate (0.670). The experimental and predicted biological activities of each subset I compound corresponding to the best CoMFA model are listed in Table 10 and plotted in Figure 5. The most predictive CoMFA model was again applied to derive the biological activity values of five MMPIs originally omitted from the training set: T1–T5, ranging from fairly active ($pIC_{50} = 6.757$) to very active ($pIC_{50} = 8.201$) (Table 11). With the exception of compound T3, the CoMFA-predicted values were close to the experimental biological activity data, with very low residuals. A color contour map of the most predictive CoMFA model together with compound 4a (MMP-3 $pIC_{50} = 9.155$) is shown in Figure 6. As in the first training-set analysis, these maps were inserted into the MMP-3 binding pocket to associate the colored polyhedra with particular regions and residues within the binding groove. This CoMFA map focuses specifically on binding requirements at either end of the ligand, that is, at the S1' and S1–S2' subsites. The large green and blue areas surrounding the thiazepine ring indicate, respectively, the positive steric potential at the top of the S1–S2' subsite and the hydrophobic nature of the sides and bottom of that subsite. The area of positive steric potential in this CoMFA map is larger than those in the maps generated by the arylsulfonfyl isoquinolines, probably because the favorable interactions of the *gem*-dimethyl group on the thiazepine ring with hydrophobic residues are interpreted by the model as a tolerance for steric bulk. A small yet critical area of negative steric potential is located near the blue hydrophobic region; we found that the addition of a hydro-

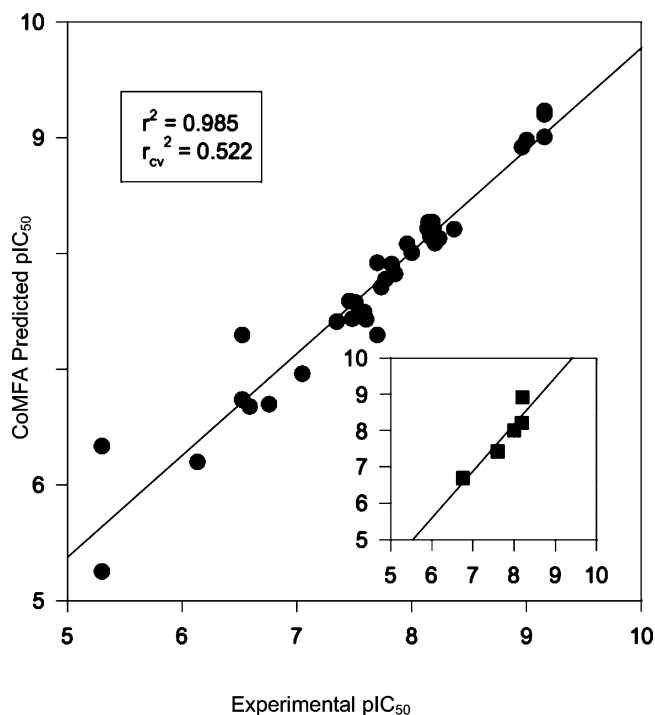


Figure 5. Plot of the CoMFA-predicted vs experimental biological activities for the partial training set (compounds 4c, 5d, 17, and 43 omitted) of thiazine- and thiazepine-based MMPIs, using Gasteiger–Marsili charges. The inset shows five test set compounds.

Table 10. Comparison of Experimental and CoMFA-Predicted Biological Activity Values for the Partial Training Set (Gasteiger–Marsili Charges; Compounds 4c, 5d, 17, and 43 Omitted) of Thiazine- and Thiazepine-Based MMPIs

cpd #	experimental (pIC_{50})	predicted (pIC_{50})	residual
4a	9.155	9.200	−0.050
4b	8.000	8.004	−0.004
4d	8.180	8.268	−0.088
4e	7.456	7.587	−0.127
4f	7.602	7.429	0.171
4g	7.770	7.775	−0.005
4h	7.959	8.081	−0.121
4i	8.959	8.917	0.043
4j	8.201	8.087	0.113
4k	8.187	8.213	−0.023
5a	8.161	8.142	0.018
5b	7.046	6.959	0.091
9	9.155	9.230	−0.080
10	7.734	7.708	0.022
21	8.155	8.231	−0.081
22	8.143	8.266	−0.126
23	9.155	9.004	0.146
24	7.347	7.409	−0.059
34	9.000	8.977	0.023
35	8.137	8.214	−0.074
36a	7.585	7.491	0.099
36b	7.481	7.434	0.046
37	6.588	6.676	−0.086
38	6.757	6.696	0.064
39	7.854	7.820	0.030
40	8.237	8.128	0.112
41	7.824	7.905	−0.085
42	6.133	6.197	−0.067
44	8.367	8.209	0.161
45	7.509	7.574	−0.064

philic functionality with even a small amount of steric bulk at this location can radically alter binding affinity. The area around the catalytic zinc is not clarified by this model, mainly

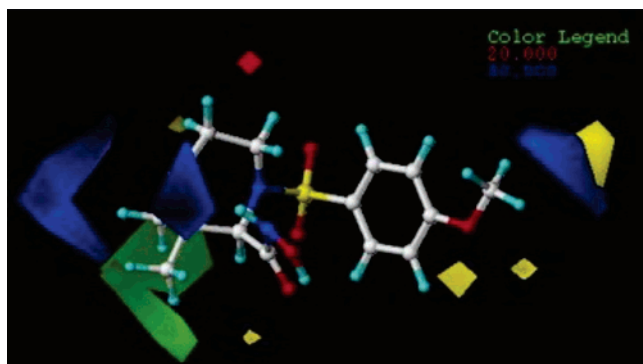


Figure 6. CoMFA contour map for the partial training set (Gasteiger–Marsili charges; compounds 4c, 5d, 17, and 43 omitted) of thiazine- and thiazepine-based MMPs, shown with compound 4a ($pIC_{50} = 9.155$). Colored polyhedra represent areas on or near the ligand where properties correlate strongly with an increase in binding affinity. Red = negative electrostatic potential; blue = positive electrostatic potential; yellow = negative steric potential; green = positive steric potential.

Table 11. Comparison of Experimental and CoMFA-Predicted Biological Activity Values for Test Set Compounds (Gasteiger–Marsili Charges; Compounds 4c, 5d, 17, and 43 Omitted) of Thiazine- and Thiazepine-Based MMPs

cpd #	experimental (pIC_{50})	predicted (pIC_{50})	residual
T1	6.757	6.696	0.061
T2	7.602	7.429	0.173
T3	8.201	8.917	−0.716
T4	8.000	8.004	−0.004
T5	8.187	8.213	−0.026

because no variation exists in the compounds at that point. Differences in aryl groups attached to the sulfonamide give rise to relatively small yellow and blue areas denoting an intolerance for steric bulk at the entrance to the S1' pocket and a preference for positive electrostatic potential within the pocket.

CONCLUSION

Predictive and self-consistent CoMFA models were derived for subsets of arylsulfonyl isoquinoline and thiazine/thiazepine-based MMPs, which should be useful for assisting the design of structurally similar compounds, and which shed light on important ligand–receptor binding characteristics. These models correlate well with experimental results and support as well as build upon previous MMP-3 QSAR analyses. The new models validate our “divide and conquer” method³⁶ as well as our finding that docking each compound into the MMP-3 active site and matching ligand conformations to the MMP-3 binding-site structure greatly enhances QSAR model predictivity. CoMSIA models were developed which, although not as statistically sound as the CoMFA models, are quite detailed and underscore crucial hydrogen-bonding requirements for successful MMPs, such as a sulfonamide or similar moiety such as a hydrogen-bond acceptor, and specify the hydrogen-bonding environment near the catalytic zinc, pointing to residues interacting with the protonated hydroxamate nitrogen. The arylsulfonyl isoquinoline models should prove especially valuable for optimizing the design of S1' substituents, and the thiazine/thiazepine models can be of use in fine-tuning hydrophobic interactions within the S1–S2' binding region. The important contribution

of the ill-defined S1–S2' active-site region to ligand binding is emphasized by the fact that all S-isomer compounds in the second training set exhibited broad-spectrum activity against MMPs, regardless of the structure/polarity of the P1' substituent. The best model from the first training set yielded color contour maps that confirm assumptions made from careful observations of X-ray and NMR structures, most notably the narrow hydrophobic nature of the S1' pocket and the necessity to avoid steric bulk near the zinc-binding group. The best thiazine/thiazepine CoMFA model indicates that binding affinity would be most enhanced by the careful addition of small hydrophobic substituents to specific sites on the thiazepine ring; the model also points to hydrophobic residues that interact with these substituents. The CoMFA maps derived from both training sets correlate well with each other and strongly indicate the need for small hydrophobic substituents oriented toward the bottom of the active site to enhance binding in that region.

REFERENCES AND NOTES

- (1) Whittaker, M.; Floyd, C.; Brown, P.; Gearing, A. Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem. Rev.* **1999**, *99*, 2735–2776.
- (2) Woessner, J. Matrix metalloproteinases and their inhibitors in connective tissue remodeling. *FASEB J.* **1991**, *5*, 2145–2154.
- (3) Hartenstein, B.; Dittich, B.; Stickens, D.; Heyer, B.; Vu, T.; Teurich, S.; Schorpp-Kistner, M.; Werb, Z.; Angel, P. Epidermal development and wound healing in matrix metalloproteinase 13-deficient mice. *J. Invest. Dermatol.* **2006**, *126*, 486–496.
- (4) Wang, J. Importance of plasma matrix metalloproteinases (MMP) and tissue inhibitors of metalloproteinase (TIMP) in development of fibrosis in agnogenic myeloid metaplasia. *Leuk. Lymphoma* **2005**, *46*, 1261–1268.
- (5) Malemud, C. Matrix metalloproteinases (MMPs) in health and disease: An overview. *Front. Biosci.* **2006**, *11*, 1696–1701.
- (6) Mendes, O.; Kim, H.; Stoica, G. Expression of MMP2, MMP9 and MMP3 in breast cancer brain metastasis in a rat model. *Clin. Exp. Metastasis* **2005**, *22*, 237–246.
- (7) Li, M.; Yamamoto, H.; Adachi, Y.; Maruyama, Y.; Shinomura, Y. Role of matrix metalloproteinase-7 (matrilysin) in human cancer invasion, apoptosis, growth, and angiogenesis. *Exp. Biol. Med.* **2006**, *231*, 20–27.
- (8) Pozo, P.; Valenzuela, M.; Melej, C.; Zaldivar, M.; Puente, J.; Martinez, B.; Gamonal, J. Longitudinal analysis of metalloproteinases, tissue inhibitors of metalloproteinases and clinical parameters in gingival crevicular fluid from periodontitis-affected patients. *J. Periodontol. Res.* **2005**, *40*, 199–207.
- (9) Kanesaka, T.; Mori, M.; Hattori, T.; Oki, T.; Kuwabara, S. Serum matrix metalloproteinase-3 levels correlate with disease activity in relapsing-remitting multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* **2006**, *77*, 185–188.
- (10) Wagner, D.; Delagardelle, C.; Ernens, I.; Rouy, D.; Vaillant, M.; Beissel, J. Matrix metalloproteinase-9 is a marker of heart failure after acute myocardial infarction. *J. Card. Failure* **2006**, *12*, 66–72.
- (11) Fichter, M.; Korner, U.; Schomburg, J.; Jennings, L.; Cole, A.; Mollenhauer, J. Collagen degradation products modulate matrix metalloproteinase expression in cultured articular chondrocytes. *J. Orthop. Res.* **2006**, *24*, 63–70.
- (12) Wu, J.; Lark, M.; Chun, L.; Eyre, D. Sites of stromelysin cleavage in collagen types II, IX, X, and XI of cartilage. *J. Biol. Chem.* **1991**, *266*, 5625–5628.
- (13) Bachmeier, B.; Albini, A.; Vene, R.; Benelli, R.; Noonan, D.; Weigert, C.; Weiler, C.; Lichtinghagen, R.; Jochum, M.; Nerlich, A. Cell density-dependent regulation of matrix metalloproteinase and TIMP expression in differently tumorigenic breast cancer cell lines. *Exp. Cell. Res.* **2005**, *305*, 83–98.
- (14) Knauper, V.; Wilhelm, S.; Seperack, P.; DeClerck, Y.; Langley, K.; Osthus, A.; Tschesche, H. Direct activation of human neutrophil procollagenase by recombinant stromelysin. *Biochem. J.* **1993**, *295*, 581–586.
- (15) Morgunova, E.; Tuuttila, A.; Bergmann, U.; Isupov, M.; Lindqvist, Y.; Schneider, G.; Tryggvason, K. Structure of human pro-matrix metalloproteinase-2: Activation mechanism revealed. *Science* **1999**, *284*, 1667–1670.

- (16) MacPherson, L.; Bayburt, E.; Capparelli, M.; Carroll, B.; Goldstein, R.; Justice, M.; Zhu, L.; Hu, S.; Melton, R.; Fryer, L.; Goldberg, R.; Doughty, J.; Spirito, S.; Blancuzzi, V.; Wilson, D.; O'Byrne, E.; Ganu, V.; Parker, D. Discovery of CGS 27023A, a non-peptidic, potent, and orally active stromelysin inhibitor that blocks cartilage degradation in rabbits. *J. Med. Chem.* **1997**, *40*, 2525–2532.
- (17) Pikul, S.; Dunham, K.; Almstead, N.; De, B.; Natchus, M.; Anastasio, M.; McPhail, S.; Snider, C.; Taiwo, Y.; Chen, L.; Dunaway, C.; Gu, F.; Mieling, G. Design and synthesis of phosphinamide-based hydroxamic acids as inhibitors of matrix metalloproteinases. *J. Med. Chem.* **1999**, *42*, 87–94.
- (18) Chapman, K.; Kopka, I.; Durette, P.; Esser, C.; Lanza, T.; Izquierdo-Martin, M.; Niedzwiecki, L.; Chang, B.; Harrison, R.; Kuo, D.; Lin, T.; Stein, R.; Hagmann, W. Inhibition of matrix metalloproteinases by *n*-carboxyalkyl peptides. *J. Med. Chem.* **1993**, *36*, 4293–4301.
- (19) Cheng, M.; De, B.; Pikul, S.; Almstead, N.; Natchus, M.; Anastasio, M.; McPhail, S.; Snider, C.; Taiwo, Y.; Chen, L.; Dunaway, C.; Gu, F.; Dowty, M.; Mieling, G.; Janusz, M.; Wang-Weigand, S. Design and synthesis of piperazine-based matrix metalloproteinase inhibitors. *J. Med. Chem.* **2000**, *43*, 369–380.
- (20) Finzel, B.; Baldwin, E.; Bryant, G.; Hess, G.; Wilks, J.; Trepod, C.; Mott, J.; Marshall, V.; Petzold, G.; Poorman, R.; O'Sullivan, T.; Schostarez, H.; Mitchell, M. Structural characterizations of nonpeptidic thiazole inhibitors of matrix metalloproteinases reveal the basis for stromelysin selectivity. *Protein Sci.* **1998**, *7*, 2118–2126.
- (21) Close, D. Matrix metalloproteinase inhibitors in rheumatic diseases. *Ann. Rheum. Disease* **2001**, *60*, 62–67.
- (22) Matter, H.; Schwab, W. Affinity and selectivity of matrix metalloproteinase inhibitors: A chemometrical study from the perspective of ligands and proteins. *J. Med. Chem.* **1999**, *42*, 4506–4523.
- (23) Almstead, N.; Bradley, R.; Pikul, S.; De, B.; Natchus, M.; Taiwo, Y.; Gu, F.; Williams, L.; Hynd, B.; Janusz, M.; Dunaway, C.; Mieling, G. Design, synthesis, and biological evaluation of potent thiazine- and thiazepine-based matrix metalloproteinase inhibitors. *J. Med. Chem.* **1999**, *42*, 4547–4562.
- (24) Donini, O.; Kollman, P. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J. Med. Chem.* **2000**, *43*, 4180–4188.
- (25) Sakharov, D.; Lim, C. Zn protein simulations including charge transfer and local polarization effects. *J. Am. Chem. Soc.* **2005**, *127*, 4921–4929.
- (26) Cramer, R.; Patterson, D.; Bunce, J. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (27) 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.
- (28) Shim, J.; Collantes, E.; Welsh, W.; Subramaniam, B.; Howlett, A.; Eissenstat, M.; Ward, S. Three-dimensional quantitative structure–activity relationship study of the cannabimimetic (aminoalkyl)indoles using comparative molecular field analysis. *J. Med. Chem.* **1998**, *41*, 4521–4532.
- (29) Kunick, C.; Laurenroth, K.; Wieking, K.; Zie, X.; Schultz, C.; Gussio, R.; Zaharevitz, D.; Leost, M.; Meijer, L.; Weber, A.; Jorgensen, F.; Lemcke, T. Evaluation and comparison of 3D-QSAR CoMSIA models for CDK1, CDK5, and GSK-3 inhibition by paullones. *J. Med. Chem.* **2004**, *47*, 22–36.
- (30) Paula, S.; Tabet, M.; Keenan, S.; Welsh, W.; Ball, W. Three-dimensional structure–activity relationship modeling of cocaine binding to two monoclonal antibodies. *J. Mol. Biol.* **2003**, *325*, 515–530.
- (31) Johnson, T.; Khan, I.; Avery, M.; Grant, J.; Meshnick, S. Quantitative structure–activity relationship studies of a series of sulfa drugs as inhibitors of *Pneumocystis carinii* dihydropteroate synthetase. *Antimicrob. Agents Chemother.* **1998**, *42*, 1454–1458.
- (32) Tong, W.; Lowis, D.; Perkins, R.; Chen, Y.; Welsh, W.; Goddette, D.; Heritage, T.; Sheehan, D. Evaluation of quantitative structure–activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669–677.
- (33) Wavefunction, Inc., Irvine, CA.
- (34) Tripos, Inc., St. Louis, MO.
- (35) Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; Lonze, P.; Weithmann, K. Quantitative structure–activity relationship of human neutrophil collagenase (MMP-8) inhibitors using comparative molecular field analysis and X-ray structure analysis. *J. Med. Chem.* **1999**, *42*, 1908–1920.
- (36) Amin, E.; Welsh, W. Three-dimensional quantitative structure–activity relationship (3D-QSAR) models for a novel class of piperazine-based stromelysin-1 (MMP-3) inhibitors: Applying a “divide and conquer” strategy. *J. Med. Chem.* **2001**, *44*, 3849–3855.
- (37) Rarey, M.; Kramer, B.; Lengauer, T. Multiple automatic base selection: Protein–ligand docking based on incremental construction without manual intervention. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 369.
- (38) Wold, S.; Albano, C.; Dunn, W.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjostrom, M. Multivariate Data Analysis in Chemistry. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984.
- (39) Harnett, D.; Murphy, J. *Introductory Statistical Analysis*; Addison-Wesley: Reading, MA, 1980.

CI060089D