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

Elizabeth Ambrose Amin and William J. Welsh*,#

Department of Chemistry & Biochemistry, University of Missouri−St. Louis, 8001 Natural Bridge Road, St. Louis, Missouri 63121

Three-dimensional quantitative structure−activity relationship (3D-QSAR) models have been obtained using comparative molecular field analysis (CoMFA) for a novel series of piperazine-based matrix metalloproteinase inhibitors (MMPIs). The crystal structure of stromelysin-1 (MMP-3) was used to identify regions of the enzyme and inhibitors where steric and electrostatic effects correlate strongly with biological activity. A training set composed of a subset of inhibitors (#10−35), which differed only with regards to the substituent (*n*-alkyl, amide, carbamide and sulfonamide) on the piperazine distal nitrogen, yielded the most predictive CoMFA model, with r^2 values of 0.592 (cross-validated) and 0.989 (conventional); this model was further validated using test compounds from two inhibitor subsets. Investigation of various ligand conformations, inhibitor subsets, alignment schemes and partial charge formalisms was required to obtain satisfactory models. The greatest success was achieved by incorporating inertial alignment together with manual adjustment of the enzyme-docked inhibitors to ensure complementarity between the inhibitors' substituent conformations and the structural characteristics of the MMP-3 S1−S2′ binding pockets. Key insights into the structure−activity relationship (SAR) obtained from this analysis for this inhibitor set are in agreement with experimentally observed data on stromelysin-1 biological activity and binding-site topology. In particular, the present study sheds new light on the steric and electrostatic requirements for ligand binding to the partly solvent-exposed S1−S2′ area.

Introduction

Matrix metalloproteinases (MMPs) consist of a class of structurally related zinc-binding endopeptidases which mediate the remodeling and degradation of extracellular matrix proteins (EMPs) such as collagen, gelatin, elastin, and proteoglycan. There are more than 20 known MMPs, most of which act on multiple or unknown substrates.[1] They function through a complex feedback mechanism with their natural inhibitors (tissue inhibitors of matrix metalloproteinases, abbreviated TIMPs) and contribute to many physiologically critical tissue remodeling functions, including cartilage turnover, wound healing, collagen breakdown during cardiac muscle repair, and the growth of new blood vessels via angiogenesis.[2−4] However, many pathological conditions are implicated by uncontrolled matrix degradation[5] that results in excessive tissue destruction and loss of extracellular matrix function. MMP proteolysis contributes to tissue degeneration and inflammation in osteo- and rheumatoid arthritis and may also play a role in the spread of such diseases. Most notably, abnormally high concentrations of MMPs have been identified in human tissue surrounding invasive carcinomas, indicating a local imbalance in the MMP-TIMP equilibrium which directly enables tumor metastasis through EMP degradation and blood vessel formation.[6] In particular, stromelysin-1 (MMP-3) has been the subject of intense research due to its presence in the vicinity of melanomas and metastatic tumors associated with breast cancer; it also functions as an active precursor to the action of other endopeptidases.[1,7,8] MMP-3 has been shown to hydrolyze several key matrix proteins comprising more than 70% of human cartilage, including aggrecan, link protein, and various collagens.[9] MMPs are therefore attractive targets for small-molecule synthetic inhibitors which could allow isolation of primary tumors and thereby enhance the effectiveness of traditional therapeutic methods such as radiation and chemotherapy.[1]

De and co-workers recently reported the synthesis of a promising series of 39 MMPIs based on a piperazine skeleton and derived from *dl*-piperazinecarboxylic acid.[10] The biological activity (reported as IC$_{50}$) of these compounds with respect to MMP-3 ranges from 2.5 to 410 nM, with the sulfonamide-substituted inhibitors demonstrating higher activity against this enzyme than the corresponding *n*-alkyl, carbamide, amide, and urea-substituted compounds. Inhibitor design involved the incorporation of hydroxamic acid as the bidentate chelating agent for the catalytic Zn^{2+}, a phenylsulfonamide moiety at the piperazine-ring N1 position to occupy the hydrophobic S1′ binding pocket, and the addition of various substituents at the piperazine distal nitrogen

* To whom correspondence should be addressed. Tel: 314-516-5318. Fax: 314-516-5342. E-mail: wwelsh@umsl.edu.
# Current address: Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, 675 Hoes Lane, Piscataway, NJ 08854-5635.

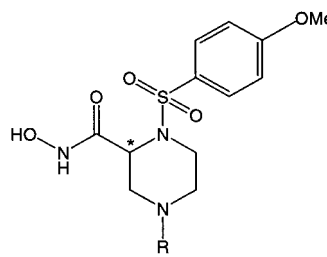position to occupy the S1–S2′ binding areas and to increase bioavailability.

Molecular modeling has emerged as a valuable tool for rational design of inhibitors that specifically target the enzyme in question and, thereby, are more likely effective at low doses and devoid of negative side effects. Modeling of MMPs is nevertheless nontrivial. In this study we have considered the catalytic Zn(II) cation to be covalently bonded to its three chelating histidines; however, accurate and reliable force field bonded parameters for the Zn(II) cation are not easily obtainable. Comparative Molecular Field Analysis (CoMFA),[11] generally regarded as the industry standard for constructing three-dimensional quantitative structure–activity relationship (3D-QSAR) models, is particularly effective in cases when the targeted receptor is unknown, difficult to model, or, in the case of MMPI design, does not lend itself to accurate binding-energy calculations. CoMFA has previously been successfully utilized in 3D-QSAR analyses of MMPIs composed of 2-arylsulfonyl-1,2,3,4-tetrahydro-isoquinoline-3-carboxylates and related hydroxamates.[12,13] In the present study, CoMFA has been employed in an attempt to pinpoint steric and electrostatic interactions between this series of piperazine-based inhibitors and specific areas within the MMP-3 binding cleft. The primary objective was to develop a new SAR model which, in addition to elucidating binding characteristics, can be used to accurately predict the biological activity of novel structurally related inhibitors.
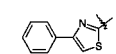
Materials and Methods

All molecular modeling calculations and visualizations were performed on a Silicon Graphics O2 workstation running under the IRIX 6.5 operating system. The molecules were initially constructed using the SPARTAN SGI Version 5.13 molecular modeling software.[14] To evaluate various schemes for assigning partial atomic charges, geometry optimizations were performed on each molecule using the SPARTAN AM1 and PM3 semiempirical methodologies; ab initio geometry optimizations (HF, 6-31G* basis set) were also calculated to an energy convergence criterion of 0.00014 kcal/mol per iteration, yielding natural and Mulliken charges. The ab initio-optimized data set compounds were imported into the SYBYL 6.7 modeling software[15] in which all remaining calculations, including CoMFA, were performed.

Data Sets. The initial training set comprised 39 piperazine-based sulfonamide compounds differing principally in the substituent(s) attached to the piperazine distal nitrogen (Tables 1–4). Crystal structures of the carbamate-substituted MMPI #20 and the cyclic-urea-substituted MMPI #47 complexed with MMP-3 were generously provided by Dr. Fei Gu at Procter & Gamble, Inc. All compounds incorporated a hydroxamate zinc-chelating group and a sulfonamide designed to hydrogen-bond to the enzyme backbone of Leu164 and Ala165 in MMP-3.[10] The two binary complexes of MMP-3 with inhibitors #20 and #47 reveal that the hydroxamate substituent on the piperazine ring adopts the R configuration. Most compounds also included a *p*-methoxysulfonamide group which has previously been shown to bind strongly within the deep MMP-3 S1′ pocket.[16] The initial data set was then divided into two subsets: subset I containing compounds #10–35, with *n*-alkyl, sulfonamide, carbamate, and amide groupings, and subset II containing compounds #36–49, with cyclic and noncyclic urea substituents as well as minor variations in the S1′-binding group. Values of the experimentally determined binding affinities (in units of pIC$_{50}$) to MMP-3 for this series of compounds[10] served as the dependent variables to develop the CoMFA models in this study.
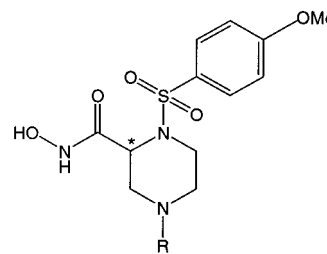
Table 1. *N*-Alkyl- and Heterocycle-Substituted Piperazine-Based MMPIs (Subset I) with Experimental Biological Activity Data[a]
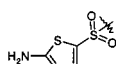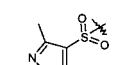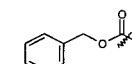

Inh #	R	MMP-3 pIC$_{50}$
10	H-	7.3188
11	Me-	6.5702
12	*n*-C$_6$H$_{13}$	7.1487
13	Bn-	7.1249
14		7.9586

[a] Compound numbering begins at #10, following the order given in De et al.

Table 2. Sulfonamide- and Carbamide-Substituted Piperazine-Based MMPIs (Subset I) with Experimental Biological Activity Data
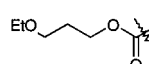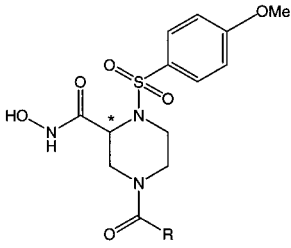

	R	MMP-3 pIC$_{50}$
15	MeSO$_2$-	8.4559
16	*p*-MeOPhSO$_2$-	7.8861
17		8.6021
18		8.4949
19	Boc-	7.2218
20	CBz-	7.7447
21		8.0506
22		8.2757
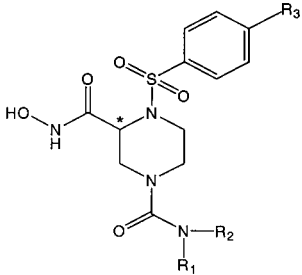
Initial CoMFA Procedure. Crystal structures of MMPI #20 and MMPI #47 served as alignment templates for all remaining molecules. Gasteiger-Huckel charges were calculated for each compound in the data set, which was inertially aligned first to #20 and then to #47. Following standard procedures in CoMFA, each MMPI was inserted into a three-dimensional lattice with grid points separated by 2.0 Å. The steric (van der Waals) and electrostatic (Coulombic) potential energy fields were calculated at each grid point by summing

Table 3. Amide-Substituted Piperazine-Based MMPIs (Subset I) with Experimental Biological Activity Data

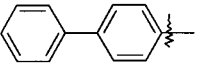| Inh #	R	MMP-3 pIC$_{50}$
23	CH$_3$-	7.7447
24	n-C$_5$H$_{11}$-	7.4089
25	n-C$_6$H$_{11}$-	7.5686
26	(isopropyl ketone)	7.2218
27	PhOCH$_2$-	8.0223
28	Ph-	7.1805
29	(pyridyl)	6.7352
30	(thienyl)	7.8239
31	(furyl)	8.1249
32	(isoxazolyl)	6.8327
33	(methylthiadiazolyl)	6.7620
34	(phenyl-methylisoxazolyl)	6.3872
35	(biphenyl)	7.6383

Table 4. Urea-Substituted Piperazine-Based MMPIs (Subset II) with Experimental Biological Activity Data

Inh #	R1	R2	R3	MMP-3 pIC$_{50}$
36	H-	H-	OMe	7.6990
37	n-C$_6$H$_{13}$-	H-	OMe	7.5528
38	n-C$_6$H$_{13}$-	Me-	OMe	8.3665
39	n-C$_6$H$_{13}$-	Me-	Br	6.7570
40	PhCH$_2$-	H-	OMe	7.7696
41[a]	PhCH$_2$-	H-	OMe	7.3872
42	PhCH$_2$-	Me-	OMe	7.4202
43	PhCH$_2$CH$_2$-	H-	OMe	7.5686
44	m-MeOPh-	H-	OMe	7.8861
45	c-C$_3$H$_5$CH$_2$-	n-Pr-	OMe	7.1871
46	-(CH$_2$)$_6$-		OMe	7.2676
47	CH$_2$CH$_2$OCH$_2$CH$_2$-		OMe	6.9066
48	CH$_2$CH$_2$NMeCH$_2$CH$_2$-		OMe	6.9431
49	MeOCH$_2$CH$_2$-	MeOCH$_2$CH$_2$-	Br	6.4145

[a] Compound #41 is a thiourea.

the individual energy interactions between each atom of the MMPI molecule and a probe atom represented by an sp^3 carbon carrying a +1 charge. A distance-dependent dielectric function ($\epsilon = \epsilon_0 r$, with $\epsilon_0 = 1$) was employed in applying Coulomb's Law, and maximum field values were truncated to 30 kcal/mol for the steric fields and to ±30 kcal/mol for the electrostatic fields.

Regression Analysis. The technique of partial least squares (PLS)[17] was employed to generate a linear regression equation that correlates changes in the steric and electrostatic CoMFA fields, representing the independent variables, with changes in the IC$_{50}$ values of the MMPIs, representing the dependent variables. PLS constructs latent variables (commonly known as principal components or PCs) composed of linear combinations of the original independent variables, thereby reducing the dimensionality of the problem. The internal predictive ability of the CoMFA models was first assessed using "leave-one-out" cross-validation. In this procedure, each compound is extracted one at a time from the data set after which its activity is predicted by a new model derived from the remaining compounds in the set. This procedure yielded the optimum number of PCs, which is associated with the highest cross-validated r^2 (r_{cv}^2) value. The PLS analysis was then repeated without cross-validation using the optimum number of components to produce a final model from which the conventional r^2 value, and related statistical parameters are computed.

Included in these parameters is the *F-ratio*, defined as $r^2/(1-r^2)$ and representing the ratio of properties explained by the QSAR model to those not explained by it.[18,19] Column filtering, in which any column of grid-point energies having a variation less than a designated value (usually 2.0 kcal/mol) is omitted from the PLS analysis, was applied as needed to reduce computation time. The 3D-QSAR models were also represented as CoMFA color contour maps which visually depict those regions of steric and electrostatic fields that contribute significantly to a given model.

Results and Discussion

Aligning the initial data set of 39 compounds to inhibitor #20 or, alternatively, to inhibitor #47 yielded poorly predictive models in terms of statistical parameters ($r_{cv}^2 = -0.078$, $r^2 = 0.783$, and $F = 42.137$ for the best model using template = #20, inertial alignment, no column filtering, and GH charges). The molecules were consequently reconstructed in SYBYL 6.7, with subset I built directly from the crystal structure of #20 and subset II from the crystal structure of #47, reproducing the crystal-structure conformations as closely as possible. Aligning all reconstructed inhibitors to #47 yielded even less satisfactory results ($r_{cv}^2 = -0.064$, $r^2 = 0.044$, *F*-ratio = 1.690 using no column filtering and GH charges). It became apparent that the significant structural differences between compounds #46−48 and all other compounds in the data set as well as substantial conformational flexibility of the cyclic urea ring

Table 5. Summary of Statistical Results from CoMFA-PLS Analysis of Biological Activity (pIC_{50}) of Piperazine-Based MMPIs Using the Complete Data Set with Gasteiger-Huckel Charges, Various Alignment Templates, and Column Filtering Levels

	47^a, 2.0^b	20, 2.0	47, none	20, none	47, none, rec^c
cross-validated r^2 (r_{cv}^2)	−0.088	−0.101	−0.083	−0.078	−0.064
conventional r^2	0.796	0.783	0.796	0.783	0.044
SE of estimate	0.276	0.285	0.276	0.285	0.582
principal components	3	3	3	3	1
F-ratio	45.613	42.137	45.613	42.137	1.690

^{a} Template structure used; all molecules inertially aligned. ^{b} Level of column filtering used. ^{c} Reconstructed from crystal structure templates: subset I from #20 and subset II from #47.

Table 6. Summary of Statistical Results from CoMFA-PLS Analysis of Biological Activity (pIC_{50}) of Piperazine-Based MMPIs Using Various Inhibitor Subsets, Gasteiger-Huckel Charges, Column Filtering Levels, and Alignment Templates

	I,^{a} none,^{b} 20^c	I, 2.0, 20	II, 2.0, 47
cross-validated r^2 (r_{cv}^2)	0.206	0.276	−0.216
conventional r^2	0.787	0.808	0.994
SE of estimate	0.296	0.281	0.057
principal components	2	2	6
F-ratio	42.500	48.327	168.416

^{a} Inhibitor subset used: I = #10−35 and II = #36−49. ^{b} Level of column filtering used. ^{c} Template molecule used.

systems themselves rendered #47 an inappropriate template for alignment of the complete set. Subsets I and II of the reconstructed data set were therefore subjected to separate CoMFA analyses,[20] with subset I aligned to #20 and subset II aligned to #47. The resulting models were much improved in terms of statistical confidence and predictability (e.g., r_{cv}^2 = 0.276, r^2 = 0.808, and F = 48.327 for subset I using template = #20, inertial alignment, 2.0 kcal/mol column filtering, and GH charges). Statistical outputs for these more successful attempts are summarized in Tables 5 and 6.

Various charge formalisms were implemented in order to improve the internal predictive ability of the CoMFA model: MMFF94, Gasteiger-Marsili,[21] "natural" and Mulliken charges from HF 6-31 g*, AM1, and PM3. SYBYL 6.7 was used to calculate MMFF94 charges for each MMPI. The GM charges resulted in marginal improvement of the model (Table 7).

Each subset I inhibitor was then docked into the MMP-3 binding cleft, after which its conformation was manually altered to achieve the best fit to the S1′ and S1−S2′ binding pockets. The features of the MMP-3 S1′ binding pocket have been well characterized.[16,22] The CoMFA model was recalculated using the settings of the best model attained at that point in time: GM charges,
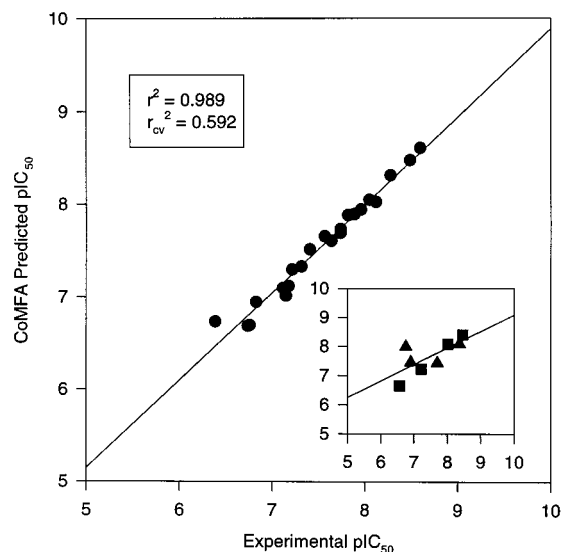
Figure 1. A plot of the CoMFA-predicted vs experimental biological activities for the subset I training set of piperazine-based MMPIs, using Gasteiger-Marsili charges, 2.0 column filtering and manual conformation adjustment. Inset shows four subset I (squares) and subset II (triangles) test set compounds.

2.0 kcal/mol column filtering, and inertial alignment. This produced a model with exceptionally high self-consistency (r^2 = 0.989) and acceptable internal predictive ability (r_{cv}^2 = 0.592) (Table 7). The corresponding experimental and CoMFA-predicted biological activities of each subset I compound corresponding to this model are listed in Table 8 and plotted in Figure 1.

Validation of the CoMFA Model. The most predictive model was applied to derive the pIC_{50} values of four subset I MMPIs originally omitted from the training set: #11 (amide-substituted), #15 (sulfonamide-substituted), #19 (carbamate-substituted), and #27 (amide-substituted) as well as four subset II urea-substituted MMPIs (Table 9). The CoMFA-predicted activities of the subset I compounds were extremely close to the experimentally determined values. While the residual values were somewhat larger for the subset II compounds, the model still demonstrated fair predictive ability for structurally related compounds outside the training set. The somewhat large residual of −1.2661 pIC_{50} units for #39 can be explained by the fact that the *p*-methoxy-phenyl-sulfonamide moiety found in all subset I compounds is replaced by a *p*-bromophenylsulfonamide moiety in #39.

Interpretation of the CoMFA Model. Inspection of the experimental binding data (IC_{50} values) reported by De and co-workers reveals that steric bulk does not significantly alter the in vitro profile of the MMPI series

Table 7. Summary of Statistical Results from CoMFA-PLS Analysis of Biological Activity (pIC_{50}) of Piperazine-Based MMPIs Using Subset I Inertially Aligned to #20 with 2.0 Column Filtering and Various Charge Formalisms

	GH^a	MMFF94^b	GM^c	AM1^d	PM3^e	Ab in Nat^f	Ab in Mul^g	GM (docked)^h
cross-validated r^2 (r_{cv}^2)	0.276	0.191	0.350	0.132	0.219	0.137	0.240	0.592
conventional r^2	0.808	0.788	0.832	0.756	0.756	0.639	0.830	0.989
SE of estimate	0.281	0.296	0.263	0.317	0.317	0.385	0.265	0.072
principal components	2	2	2	2	2	2	2	6
F-ratio	48.327	42.652	57.112	35.711	35.656	20.375	56.094	297.358

^{a} Gasteiger-Huckel. ^{b} Merck Molecular Force Field. ^{c} Gasteiger-Marsili. ^{d} Semiempirical Austin Model 1. ^{e} Semiempirical Parametric Model 3. ^{f} Ab initio natural. ^{g} Ab initio Mulliken. ^{h} Gasteiger-Marsili, with inhibitors docked into the MMP-3 binding cleft and manually adjusted for best fit.

Table 8. Comparison of Experimental and CoMFA-Predicted Biological Activity Values (Subset I, Gasteiger-Marsili Charges, Manual Alignment) of Piperazine-Based MMPIs

inhibitor no.	experimental (pIC$_{50}$)	predicted (pIC$_{50}$)	residual
10	7.3188	7.332	−0.014
11	6.5702	6.655	−0.085
12	7.1487	7.017	0.132
13	7.1249	7.100	0.025
14	7.9586	7.941	0.018
15	8.4559	8.395	0.061
16	7.8861	7.893	−0.007
17	8.6021	8.607	−0.005
18	8.4949	8.474	0.021
19	7.2218	7.228	−0.006
20	7.7447	7.734	0.011
21	8.0506	8.046	0.004
22	8.2757	8.311	−0.035
23	7.7447	7.694	0.051
24	7.4089	7.517	−0.108
25	7.5686	7.655	−0.086
26	7.2218	7.301	−0.079
27	8.0223	8.068	−0.046
28	7.1805	7.123	0.057
29	6.7352	6.686	0.050
30	7.8239	7.883	−0.059
31	8.1249	8.023	0.102
32	6.8327	6.946	−0.113
33	6.7620	6.695	0.067
34	6.3872	6.372	0.016
35	7.6383	7.609	0.029

Table 9. Comparison of Experimental and CoMFA-Predicted Biological Activity Values for Subset I and II Test Set Compounds

inhibitor no.	experimental (pIC$_{50}$)	predicted (pIC$_{50}$)	residual
11	6.5702	6.656	−0.085
15	8.4559	8.395	0.061
19	7.2218	7.228	−0.006
27	8.0223	8.068	−0.046
36	7.6990	7.426	0.273
38	8.3665	8.078	0.288
39	6.7570	7.984	−1.227
47	6.9066	7.457	−0.551

and that a basic nitrogen adjacent to the piperazine ring is not well tolerated. These researchers also hypothesized that the orientation of substituents located on the piperazine distal nitrogen may play a crucial role in MMPI binding. The enzyme-ligand crystal structure containing #20 orients the Cbz moiety approximately midway between the S1 and S2′ pockets; hydrophobic interactions were observed between Cbz and two phenylalanine residues located at the edge of the MMP binding cleft (Phe210 and Phe186). Since the S1−S2′ area is partly exposed to solvent, certain polar substituents (or substituent regions) were theorized to orient themselves away from the enzyme. The CoMFA results support this assumption as well as the experimental data and point to six major areas on and around the ligand where the orientation and electrostatic properties of 4N-substituents are especially important.

The color contour maps of the most predictive CoMFA model, with subset I inhibitor #17 (the most active compound in the data set) inserted for visual clarity, are shown in Figure 2a−c. These contour maps, together with #17, were also inserted into the MMP-3 binding pocket to associate the various color contours with specific regions and/or residues within the binding cleft. The colored polyhedra represent three-dimensional regions corresponding to zones on the ligand which
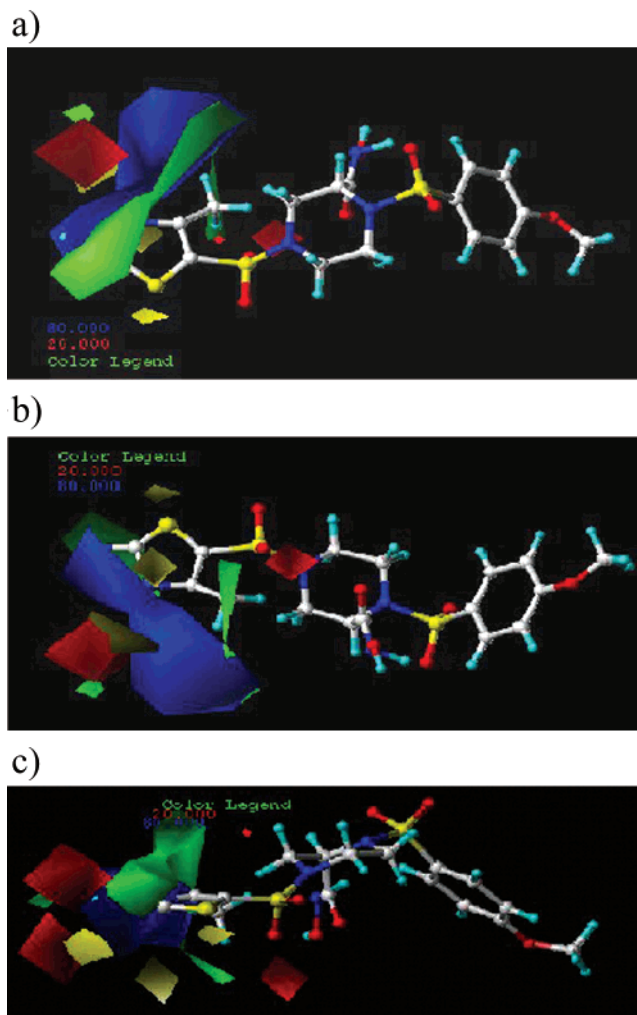
Figure 2. a−c. Graphical renderings (top, bottom, and side views) of the most predictive CoMFA model shown with the most active subset I MMPI (#17).

strongly affect binding affinity. Areas of negative (red) and positive (blue) electrostatic potential are shown together with areas of negative (yellow) and positive (green) steric potential. The CoMFA map displays several sharply delineated polyhedra within a relatively small volume and thereby offers detailed information about specific ligand requirements in that area. Regions where electrostatic properties correlate strongly with an increase or decrease in binding affinity are larger and more numerous than those representing steric effects. The distal-nitrogen substituents are surrounded by a large blue region oriented toward the bottom edge of the MMP-3 binding cleft and by a comparably sized green region facing the open top of the cleft. Two smaller yellow regions are found below this blue area, and two larger and more prominent red areas are located on the side of the substituent area facing the solvent. The blue region indicates a strong preference for hydrophobic groups facing the protein, specifically toward the two phenylalanine residues near the bottom of the binding cleft; in some cases this blue region includes the 4N-substituents themselves (or parts of them). The two yellow areas directly below point to a modest intolerance for steric bulk; a small green wedge representing a correlation of larger substituent size with higher binding affinity is visible adjacent to the yellow areas. The small

size and close proximity of these green and yellow areas indicate that the influence of steric bulk in that region on biological activity is complex and not uniform; this information was not obvious from mere inspection of the docked inhibitors within the enzyme binding site and can be utilized to precisely fine-tune the orientation and conformation of bulkier distal-nitrogen substituents. The side of the 4N-area facing the solvent contains no yellow polyhedra at all but rather a long green region representing a strong tolerance for steric bulk and two adjacent broad red areas indicating a preference for negative electrostatic potential. This observation suggests that an MMP-3 inhibitor with strong binding affinity at the S1−S2′ location would incorporate, in combination, a medium-size polar group oriented toward the solvent/top of the MMP-3 binding cleft, and a hydrophobic moiety directed toward the bottom of the cleft, with minimal, carefully oriented areas of steric bulk. The experimental data for #17 (MMP-3 pIC_{50} = 8.6021) correlate especially well with these results.

Summary and Conclusions

A highly predictive CoMFA model was derived for a subset of *n*-alkyl-, sulfonamide-, carbamide-, and amide-substituted piperazine-based MMPIs which should be useful for guiding the design and refinement of additional piperazine-based MMPIs, most especially the optimization of 4N-substituents. Docking each inhibitor inside the MMP-3 binding cleft and manually adjusting inhibitor conformations to match the structural characteristics of each MMP-3 binding pocket proved to be essential in obtaining an accurate and predictive model. The choice of subset I as the CoMFA training set was advantageous in that all molecules were identical except for the distal nitrogen substituents, allowing a more focused and detailed investigation of the S1−S2′ binding areas. The CoMFA contour maps not only emphasize the importance of the piperazine 4N-substituent conformation to ligand binding but also include sharply delineated regions that may play essential roles in overall MMPI biological activity. This model should therefore prove useful in MMPI design starting from a variety of scaffolds. To maximize S1−S2′ binding affinity, an MMPI will need to incorporate a carefully oriented combination of polar and nonpolar substituents at that region, with lower steric bulk toward the bottom of the MMPI binding cleft. Two additional models were fairly predictive for MMPI biological activity, both using 2.0 kcal/mol column filtering: subset I with Gasteiger-Huckel charges (r_{cv}^2 = 0.276, r^2 = 0.808, *F*-ratio = 48.327) and subset I with ab initio Mulliken charges (r_{cv}^2 = 0.240, r^2 = 0.830, *F*-ratio = 56.094). The most predictive model also demonstrated fair applicability to a second subset of urea-substituted inhibitors and could therefore be used to roughly predict and rank-order binding affinities for structurally related compounds.

References

(1) Whitaker, M.; Floyd, C.; Brown, P.; Gearing, A. J. Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem. Rev.* **1999**, *99*, 2735−2776.
(2) Chapman, K. T.; Kopka, I. E.; Durette, P. L.; Esser, C. K.; Lanza, T. J.; Izquierdo-Martin, M.; Niedzwiecki, L.; Chang, B.; Harrison, R. K.; Kuo, D.; Lin, T.-Y.; Stein, R. L.; Hagmann, W. K. Inhibition of matrix metalloproteinases by *N*-carboxyalkyl peptides. *J. Med. Chem.* **1993**, *36*, 4293−4301.
(3) Gomis-Rüth, F.-X.; Maskos, K.; Betz, M.; Bergner, A.; Huber, R.; Suzuki, K.; Yoshida, N.; Nagase, H.; Brew, K.; Bourenkov, G. P.; Bartunik, H.; Bode, W. Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* **1997**, *389*, 77−81.
(4) Dhanaraj, V.; Ye, Q.; Johnson, L. L.; Hupe, D. J.; Ortwine, D. F.; Dunbar, J. B.; Rubin, J. R.; Pavlovsky, A.; Humblet, C.; Blundell, T. L. X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. *Atm. Environ.* **1994**, *28*, 547−554.
(5) Botos, I.; Scapozza, L.; Zhang, D.; Liotta, L.; Meyer, E. F. Batimastat, a potent matrix metalloproteinase inhibitor, exhibits an unexpected mode of binding. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2749−2754.
(6) Fernandez-Catalan, C.; Bode, W.; Huber, R.; Turk, D.; Calvete, J. J.; Lichte, A.; Tschesche, H.; Maskos, K. Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor. *EMBO J.* **1998**, *17*, 5238−5248.
(7) Becker, J. W.; Marcy, A. I.; Rokosz, L. L.; Axel, M. G.; Burbaum, J. J.; Fitzgerald, P. M. D.; Cameron, P. M.; Esser, C. K.; Hagmann, W. K.; Hermes, J. D.; Springer, J. P. Stromelysin-1: Three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. *Prot. Sci.* **1995**, *4*, 1966−1976.
(8) Gooley, P. R.; O'Connell, J. F.; Marcy, A. I.; Cuca, G. C.; Axel, M. G.; Caldwell, C. G.; Hagmann, W. K.; Becker, J. W. Comparison of the structure of human recombinant short form stromelysin by multidimensional heteronuclear NMR and X-ray crystallography. *J. Biomol. NMR* **1996**, 7, 8−28.
(9) MacPherson, L. J.; Bayburt, E. K.; Capparelli, M. P.; Carroll, B. J.; Goldstein, R.; Justice, M. R.; Zhu, L.; Hu, S.; Melton, R. A.; Fryer, L.; Goldberg, R. L.; Doughty, J. R.; Spirito, S.; Blancuzzi, V.; Wilson, D.; O'Byrne, E. M.; Ganu, V.; Parker, D. T. 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.
(10) Cheng, M.; De, B.; Pikul, S.; Almstead, N. G.; Natchus, M. G.; Anastasio, M. V.; McPhail, S. J.; Snider, C. E.; Taiwo, Y. O.; Chen, L.; Dunaway, C. M.; Gu, F.; Dowty, M. E.; Mieling, G. E.; Janusz, M. J.; Wang-Weigand, S. Design and synthesis of piperazine-based matrix metalloproteinase inhibitors. *J. Med. Chem.* **2000**, *43*, 369−380.
(11) Cramer, R.; Patterson, D.; Bunce, J. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, 110, 5959−5967.
(12) Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; Lönze, P.; Weithmann. K. U. 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.
(13) 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.
(14) Wavefunction, Inc., Irvine, CA.
(15) Tripos, Inc., St. Louis, MO.
(16) Almstead, N. G.; Bradley, R. S.; Pikul, S.; De, B.; Natchus, M. G.; Taiwo, Y. O.; Gu, F.; Williams, L. E.; Hynd, B. A.; Janusz, M. J.; Dunaway, C. M.; Mieling, G. E. Design, synthesis, and biological evaluation of potent thiazine- and thiazepine-based matrix metalloproteinase inhibitors. *J. Med. Chem.* **1999**, *42*, 4547−4562.
(17) Wold, S.; Albano, C.; Dunn, W.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjosstrom, M. Multivariate data analysis in chemistry. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984.
(18) Harnett, D.; Murphy, J. *Introductory Statistical Analysis*. Reading, MA: Addison-Wesley: 1980.
(19) *The Tripos Bookshelf Version 4.0*; Tripos: St. Louis, MO, 1999; Section 4.5−4.6.
(20) A similar "divide and conquer" strategy has been employed previously by other workers in which the data set was divided into subsets according to structural similarity in order to improve the statistical quality and/or predictability of CoMFA models.

See, for example: Waller, C. L.; Oprea, T. I.; Chae, K.; Park, H. K.; Korach, K. S.; Laws, S. C.; Wiese, T. E.; Kelce, W. R.; Gray, L. E. Ligand-based identification of environmenal estrogens. *Chem. Res. Toxicol.* **1996**, *9*, 1240−1248.

(21) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(22) Finzel, B. C.; Baldwin, E. T.; Bryant, G. L.; Hess, G. F.; Wilks, J. W.; Trepod, C. M.; Mott, J. E.; Marshall, V. P.; Petzold, G. L.; Poorman, R. A.; O'Sullivan, T. J.; Schostarez, H. J.; Mitchell, M. A. Structural characterizations of nonpeptidic thiazole inhibitors of matrix metalloproteinases reveal the basis for stromelysin selectivity. *Prot. Sci.* **1998**, *7*, 2118−2126.