

A Ligand-Based Molecular Modeling Study on Some Matrix Metalloproteinase-1 Inhibitors Using Several 3D QSAR Techniques

Keng-Chang Tsai and Thy-Hou Lin*

Institute of Molecular Medicine & Department of Life Science, National Tsing Hua University,
Hsinchu 30013, Taiwan

Received May 26, 2004

Some three-dimensional quantitative structure–activity relationship (3D-QSAR) models have been constructed using the comparative molecular field analysis (CoMFA) and comparative molecular similarity indices (CoMSIA) for a series of 84 proline-based plus 12 structurally more diversified nonproline matrix metalloproteinase inhibitors. The structures of these inhibitors were built from a structure template extracted from the crystal structure of stromelysin. The structures built were divided into the training and test sets for both the CoMFA and CoMSIA analyses for each being composed of 60 and 24 inhibitors, respectively. The structures in the training set were aligned using some alignment rules derived from the analysis of the Ligplot program on a recent crystal structure of ligand-collagenase-1 complex. Some stepwise CoMSIA's were performed on the aligned training set on which the best CoMFA result was obtained. The best CoMSIA model was identified from the stepwise results, and the corresponding pharmacophore features were used for the construction of a pharmacophore hypothesis by the Catalyst 4.9 program. The training set was extended to include 11 structurally more diversified and nonproline inhibitors. To construct a pharmacophore hypothesis, the conformation of 60 structurally aligned proline-based inhibitors was fixed, while that of the 11 structurally more diversified nonproline inhibitors was allowed to vary during the hypothesis construction process. It was found that the predicted activities by the top hypothesis constructed for both the training and test sets were as good in statistics as those predicted by the best CoMSIA model from which the hypothesis was derived. The top hypothesis was mapped onto the structures of several highly active inhibitors selected from both the training and test sets. The goodness of mapping on each inhibitor was found to be correlated well with the activity of each inhibitor.

INTRODUCTION

Matrix metalloproteinases (MMPs) are a family of zinc-containing endopeptidases that mediate the degradation of extracellular matrix of connective tissues.¹ There are at least 20 enzymes that have been identified for the family, and they are implicated in many degenerative diseases with slow matrix degradation rate, such as cartilage loss in osteoarthritis,^{2–4} rheumatoid arthritis,^{3,5,6} bone matrix degradation in osteoporosis, or remodeling in Alzheimer disease.⁷ Many of these enzymes possess the following three essential domains namely, the propeptide domain as activation sequence, the catalytic domain containing a conserved zinc binding motif, and a hemopexin-like domain involving substrate recognition.⁸ There are three collagenases identified among these enzymes namely, the interstitial collagenase-1 (MMP-1),⁹ stromelysin (MMP-3),¹⁰ and neutrophil collagenase (MMP-8),¹¹ and they are recognized as the key enzymes in the pathology of matrix degradation.

Recently, numerous selective collagenase inhibitors have been developed through a rational design process for MMP-3 and MMP-8 due to the X-ray structures of these two enzymes being available. For example, Cheng and co-workers have synthesized 39 piperazine-based matrix metalloproteinase

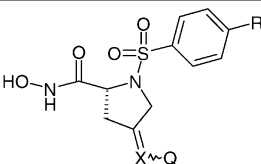
inhibitors (MMPIs) with biological activity (reported as IC₅₀) against MMP-3 ranging from 2.5 to 410 nM.¹² Matter and co-workers have synthesized 90 novel 2-(arylsulfonyl)-1,2,3,4-tetrahydroisoquinoline-3-carboxylates and -hydroxamates and other bicyclic rigidified analogues as inhibitors of MMP-8.¹³ It is also known that several MMPIs have progressed into clinical trials for cancers, rheumatoid, arthritis, and osteoarthritis. However, progression through these trials has been hindered by the appearing of musculoskeletal syndrome (MSS) which manifests itself as musculoskeletal pain after a few weeks of treatment.¹ Inhibition of MMP-1 has been suggested to play a role in the appearance of MSS.² Therefore, based on an aminopyrrolidine scaffold, Natchus and colleagues have designed a series of MMP-1 inhibitors and found that some of these inhibitors are able to protect the degradation of cartilage in a rat model of osteoarthritis.¹⁴ Moreover, several proline-based MMPIs designed by Cheng and colleagues also show substantial activity against MMP-1.¹²

The CoMFA technique¹⁵ has emerged as an industry standard for constructing 3D QSAR models. In using this technique, a structural alignment rule must be established beforehand to bring the orientation of a given molecule in 3D space relative to the other molecules in the analysis. The electrostatic and steric interaction energies between each ligand and a probe atom on a predefined grid are then

* Corresponding author fax: 886-3-571-5934; e-mail: thlin@life.nthu.edu.tw.

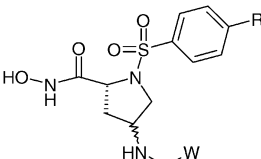
Table 1. Structures of the 84 Proline-Based MMP-1 Inhibitors Studied

Oxime hydroxyl and aromatic sulfonamide substituted modified proline scaffold MMP-1s with experimental biological activity data

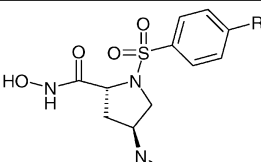
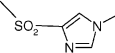


	Inh #	Q	X	R	MMP-1 pIC ₅₀
	13 ^a	-OH	-N	-OMe	7.76
	3 ^a	-OH	-N	-OnBu	6.92
	14 ^a	-OMe	-N	-OMe	8.00
	15 ^a	-OMe	-N	-OEt	7.65
	16 ^a	-OMe	-N	-OnPr	7.88
	4 ^a	-OMe	-N	-OnBu	7.58
	17 ^a	-OMe	-N	-OCH2CH2OMe	6.96
	18 ^a	-OMe	-N	-OPh	8.52
	19 ^a	-OMe	-N	-OC6H4F	8.52
	20 ^a	-OMe	-N	-O-4-Pyr	7.09
	21 ^a	-OEt	-N	-OnBu	7.40
	22 ^a	-OrBu	-N	-OMe	7.69
	23 ^a	-OrBu	-N	-OnBu	7.13
	24 ^a	-OrBu	-N	-O-4-Pyr	7.58
	25 ^a	-OrBu	-N	-OC6H4F	7.61
	26 ^a	-OrBu	-N	-OnBu	7.12
	27 ^a	-OPh	-N	-O-4-Pyr	7.49
	28 ^a	-OCH2Ph	-N	-OnBu	6.89
	29 ^a	-N-piperazine	-N	-OnBu	6.73
	6 ^a	-N-morpholine	-N	-OnBu	6.53
	11 ^a	-H	-CH	-OMe	7.04
	30 ^a	-H	-CH	-OnPr	6.80
	31 ^a	-H	-CH	-OnBu	6.48
	32 ^a	-H	-CH	-OCH2CH2OMe	5.81
	33 ^a	-H	-CH	-OPh	7.82

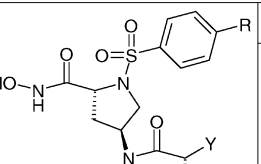
Hydroxyl and aromatic sulfonamide substituted modified proline scaffold MMP-1s with experimental biological activity data

	Inh #	W	R	MMP-1 pIC ₅₀
	34 ^a	-Me	-OMe	6.67
	7 ^a	-Me	-OnBu	6.10
	35 ^a	-Me	-OCH ₂ CH ₂ OMe	5.37
	36 ^a	-Me	-OPh	7.06

Amine-bearing and aromatic sulfonamide substituted new hydroxamate MMP-1s derived from functionalized 4-aminoproline with experimental biological activity data

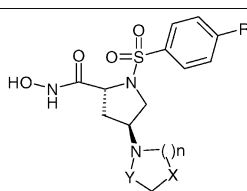
	Inh #	Q	W	R	MMP-1 pIC ₅₀
	19 ^b	-H	-H	-OMe	6.60
20 ^b	-H	-H	-H	-OnBu	6.42
21 ^b	-(CH ₂) ₂ CH ₃	-H	-H	-OMe	6.82
22 ^b	-(CH ₂) ₂ Ph	-H	-H	-OMe	6.25
23 ^b	-SO ₂ CH ₃	-H	-H	-OMe	6.58
24 ^b	-SO ₂ CH ₃	-H	-H	-OnBu	5.88
25 ^b	-SO ₂ CH ₃	-H	-H	-O-4-Pyr	6.34
26 ^b	-SO ₂ CH ₃	-Me	-Me	-OMe	7.72
27 ^b	-SO ₂ CH ₃	-CH ₂ -3-Pyr	-CH ₂ -3-Pyr	-OMe	7.13
28 ^b	-SO ₂ CH ₃	-CH ₂ CH ₂ CH ₃	-CH ₂ CH ₂ CH ₃	-OMe	7.48
29 ^b	-SO ₂ - <i>p</i> -C ₆ H ₄ O	-H	-H	-OMe	6.45
30 ^b	Me	-H	-H	-OnBu	6.15
33 ^b		-CH ₂ -3-Pyr	-CH ₂ -3-Pyr	-OMe	6.56
34 ^b	-CO ₂ iPr	-H	-H	-OnPr	6.30
35 ^b	-COCH ₂ OMe	-H	-H	-OnBu	5.92

Lactic acid amide and aromatic sulfonamide substituted new hydroxamate MMP-1s derived from functionalized 4-aminoproline with experimental biological activity data

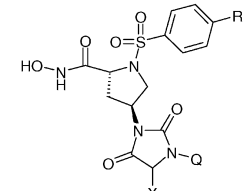
	Inh #	W	Y	Z	R	MMP-1 pIC ₅₀
	36 ^b	-H	-CH3	-OMe	-CH2Ph	6.85
	37 ^b	-H	-CH2Ph	-OMe	-CH2Ph	6.61
	38 ^b	- <i>n</i> Pr	-CH3	-OMe	-H	7.06
	39 ^b	- <i>n</i> Pr	-CH2CH2Ph	-OMe	-H	6.58
	40 ^b	-H	-CH3	- <i>On</i> Pr	-H	6.85
	41 ^b	-H	-CH3	- <i>On</i> Bu	-H	5.92
	42 ^b	-H	- <i>i</i> Pr	- <i>On</i> Bu	-H	6.56
	43 ^b	-H	-CH2 <i>i</i> Pr	- <i>On</i> Bu	-H	6.35

Heterocyclic and aromatic sulfonamide substituted new hydroxamate MMP-1s derived from functionalized 4-aminoproline with experimental biological activity data

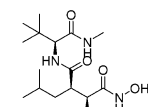
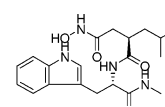
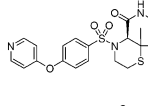
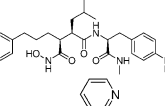
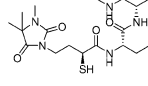
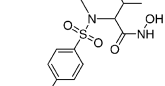
Table 1 (Continued)

	Inh #	n	Y	X	R	MMP-1 pIC ₅₀
	44 ^b	2	CH ₂	CH ₂	-OMe	6.55
	45 ^b	2	CH ₂	CH ₂	-OnPr	6.67
	46 ^b	2	CH ₂	CH ₂	-nPent	5.69
	47 ^b	2	CH ₂	CH ₂	-OPh	7.76
	48 ^b	2	CH ₂	O	-OMe	6.65
	49 ^b	2	CH ₂	O	-OnPr	6.30
	50 ^b	2	CH ₂	O	-OnBu	6.03
	51 ^b	2	CH ₂	O	-nPent	5.69
	52 ^b	2	CH ₂	O	-OPh	7.72
	53 ^b	2	CH ₂	SO ₂	-OMe	6.52
	55 ^b	1	CH ₂	CH ₂	-OnPr	6.50
	56 ^b	1	CH ₂	CH ₂	-O-4-C ₆ H ₄ F	7.69
	57 ^b	1	CH ₂	CH ₂	-OPh	7.72
	58 ^b	1	SO ₂	CH ₂	-OMe	7.37
	59 ^b	1	SO ₂	CH ₂	-OnPr	7.00
	60 ^b	1	SO ₂	CH ₂	-OnBu	6.63

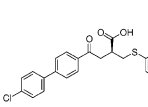
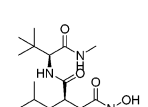
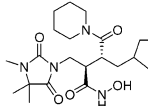
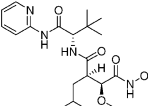
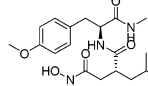
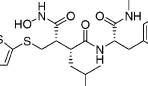
Hydantoin moieties and aromatic sulfonamide substituted new hydroxamate MMPiS derived from functionalized 4-aminoproline with experimental biological activity data

	Inh #	Q	X	R	MMP-1 pIC ₅₀
	61 ^b	-CH ₃	-H	-OMe	6.69
	62 ^b	-CH ₃	-H	-OEt	6.23
	63 ^b	-CH ₃	-H	-OnPr	6.58
	64 ^b	-CH ₃	-H	-OnBu	6.03
	65 ^b	-CH ₃	-H	-OCH ₂ CH(CH ₃) ₂	6.52
	66 ^b	-CH ₃	-H	-OCH ₂ CH ₂ OCH ₃	5.74
	67 ^b	-CH ₃	-H	-OPh	7.67
	68 ^b	-CH ₃	-H	-O-4-Pyr	6.22
	69 ^b	-H	-SCH ₃	-OnBu	6.12
	70 ^b	-H	-(CH ₃) ₂	-OnBu	6.10
	71 ^b	-CH ₂ CH ₂ CH ₂	-H	-OnPr	6.45
	72 ^b	-CH ₂ CH ₂ CH ₂	-H	-OnBu	6.31
	73 ^b	-CH ₂ CH ₂ CH ₂	-H	-OCH ₂ CH ₂ OCH ₃	5.53
	74 ^b	-CH ₂ CH ₂ CH ₃	-H	-OnBu	6.24
	75 ^b	-CH ₂ CH ₂ CH ₃	-H	-OCH ₂ CH ₂ OCH ₃	5.60

c. The 12 structurally more diversified non-proline MMP-1 inhibitors

Inh#	IC ₅₀ (nM) pIC ₅₀	Inh#	IC ₅₀ (nM) pIC ₅₀
	01 ^c 0.77 9.11		07 ^c 1.5 8.82
	02 ^c 5.7 8.34		08 ^c 6.0 8.22
	03 ^c 25 7.60		09 ^c 33 7.48

c. The 12 structurally more diversified non-proline MMP-1 inhibitors

Inh#	IC ₅₀ (nM) pIC ₅₀	Inh#	IC ₅₀ (nM) pIC ₅₀
	04 ^c 5000 5.30		10 ^c 3.0 8.52
	05 ^c 3.0 8.52		11 ^c 10 8.00
	06 ^c 20 7.69		12 ^c 0.99 9.00

^a Matrix metalloproteinase inhibitors derived from a modified proline scaffold. ^b New hydroxamate matrix metalloproteinase inhibitors derived from functionalized 4-aminoproline. ^c The 12 structurally more diversified nonproline MMP-1 inhibitors.

computed. Both the CoMFA and GRID/Generating Optimal Linear PLS Estimations (GOLPE) approaches¹⁶ have been used by Matter and Schwab to study the binding affinity and selectivity for 90 MMPIs synthesized for MMP-3 and MMP-8.¹⁷ Meanwhile, Amin and Welsh have used CoMFA to construct some 3D QSAR models for the 39 piperazine-based MMPIs synthesized by Cheng and co-workers¹² to against MMP-3.¹⁸ In this report, we have employed several 3D QSAR techniques on the aligned structures of 30 and 54 MMPIs of MMP-1 designed respectively by Cheng et al.¹² and Natchus et al.¹⁴ for constructing some 3D QSAR models for these compounds. While both series of compounds were derived from *cis*-hydroxy-D-proline, the major difference in structural features between the two was that a sp² center for the former rather than a sp³ one for the latter introduced at C-4 of the pyrrolidine ring. To explore the binding with the

subsites in the catalytic domain, S1' and S2' pockets, both series of compounds were varied at R-4 of the aromatic sulfonamide part or at C-4 of the pyrrolidine ring.^{12,14} These proline-based inhibitors were divided into the training and test sets. Both the training and test sets were structurally aligned and analyzed by the CoMFA and CoMSIA methods to derive the best 3D QSAR model for the inhibitors. Further, the pharmacophore features obtained from the best CoMSIA model were used to construct some pharmacophore hypotheses using the Catalyst 4.9 program¹⁹ on a SGI Origin 3800 workstation equipped with 48 × 400 MHz MIPS R12000 processors. Some structurally more diversified nonproline inhibitors were also included in the training set for the construction of these pharmacophore hypotheses. The top hypothesis thus generated was mapped onto the structures of several highly active inhibitors selected from both the

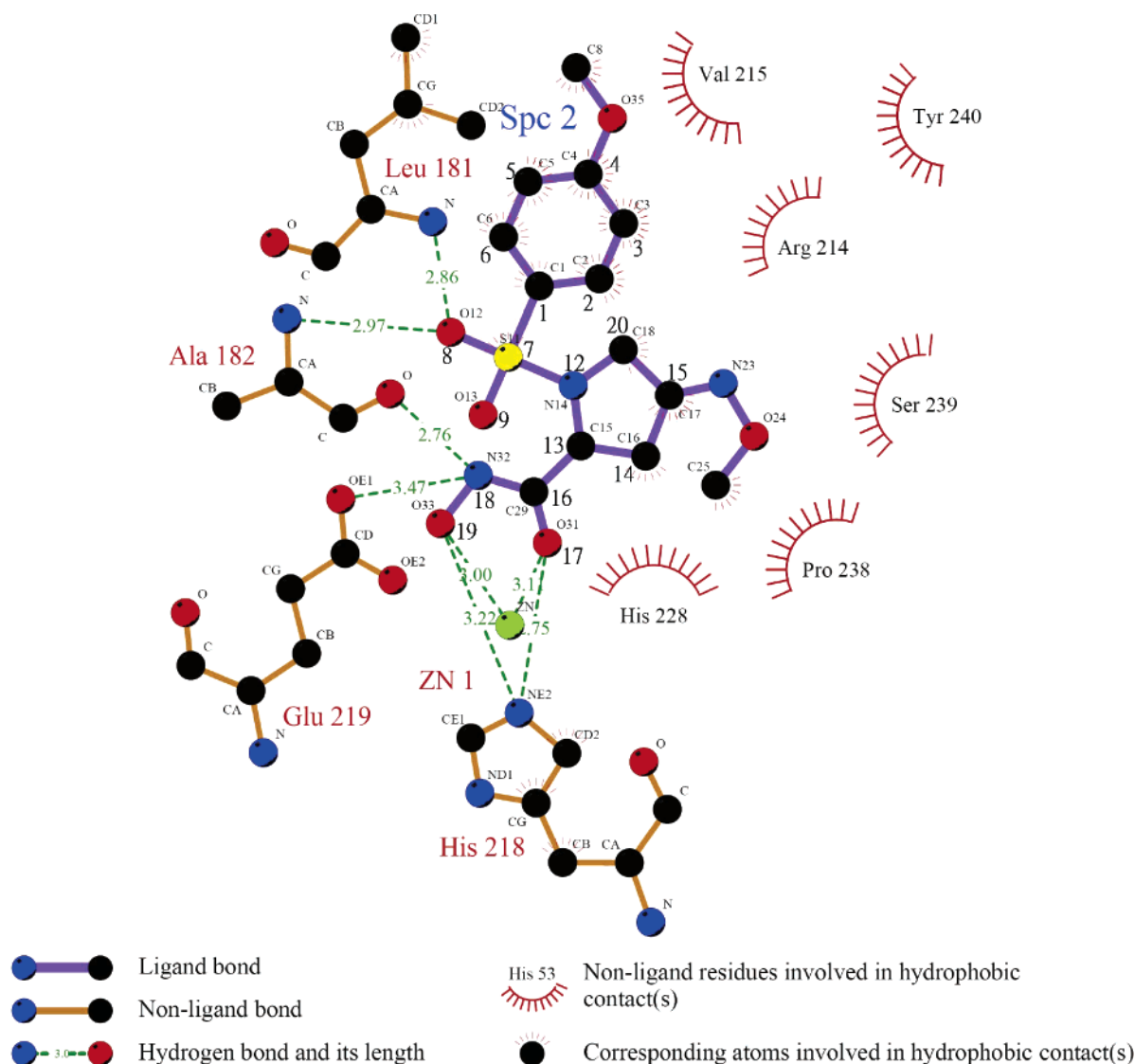


Figure 1. The structure of inhibitor **14^a** (Spc2)-MMP-1 receptor complex²⁹ was analyzed using the Ligplot 4.22 program¹⁹ to identify some specific contacts between atoms of ligand and receptor. The ligand atoms served as the correspondence points in the subsequent structural alignment processes were marked with numbers alongside.

training and test sets. The predicted activities for both the training and test sets by the top hypothesis were found to be as good in statistics as those predicted by the best CoMSIA model from which the hypothesis was derived. The feasibility of using a 3D QSAR model to assist the construction of a pharmacophore hypothesis by the Catalyst 4.9 program¹⁹ was discussed.

MATERIALS AND METHODS

The structures and activities of all 84 MMPIs of MMP-1 studied were listed in Table 1. Construction of structures for these MMPIs was based on the X-ray structure 1d7x¹² which contains ligand *N*-hydroxy-1*N*-(4-methoxyphenyl)-sulfonyl-4-(*Z,E*-*N*-methoxyimino)pyrrolidine-(2*R*)-carboxamide (inhibitor **14^a** of Table 1) treated as the template and obtained from the protein data bank.²⁰ The construction of the molecules was done within the active site of 1d7x by replacing side chains of the template with other groups as

has been described by others. The hydrogen atoms were added for each structure. Each of these structures was rotated into the coordinate frame of ligand RS2 of a MMP-1 receptor (966C)⁸ before being merged with the ligand depleted MMP-1 receptor. Each structure of the ligand–receptor complex constructed was subjected to a brief energy minimization using the SYBYL 6.9.1 program²¹ and the Tripos60 Force Field Engine. The Gasteiger–Hückel²² and KOLL_ALL²³ charges were deployed respectively for ligands and receptor, and a nonbonding cutoff of 8 Å was used for each structure complex. Each ligand structure constructed was extracted from each structure complex using the SYBYL Extract module.²¹ The total number of structures constructed was 84 which include 30 proline- (series a)¹² and 54 4-aminoproline- (series b)¹⁴ based inhibitors. The biological activity expressed in pIC₅₀ for the compound set was from 5.37 to 8.52. The compound set was divided into two sets, namely, the training and test set with each containing 60 and 24 proline-based inhibitors, respectively. The structures in the test set were selected randomly, and they were as

Table 2. Stepwise Development of the Training Set CoMFA and CoMSIA Models^a

CoMFA MMP-1 60		CoMSIA	MMP-1 cross-validation			MMP-1 no validation		
			NC	q ²	CF	SEP	r ²	F
leave one out q ² _{loo}	0.650	S	5	0.525	2	0.308	0.812	46.514
cross-validated (q ² _{cv})	0.649	E	3	0.283	2	0.399	0.672	38.188
conventional r ²	0.954	H	4	0.648	2	0.247	0.876	97.535
standard error of estimate	0.152	D	1	0.103	2	0.651	0.096	6.136
principal components	5	A	3	0.277	2	0.449	0.585	26.280
F-values _(5,54)	224.167	H + S	6	0.655	2	0.171	0.943	146.569
		H + E	5	0.633	2	0.203	0.918	120.724
		H + D	5	0.608	2	0.215	0.908	106.596
		H + A	5	0.651	2	0.192	0.926	135.859
		H + S + E	6	0.657	2	0.141	0.961	218.344
		H + S + D	6	0.604	2	0.142	0.961	216.116
		H + S + A*	6	0.730	2	0.157	0.952	175.639
		H + S + A + D	6	0.622	2	0.134	0.964	244.380
		H + S + A + E	6	0.626	2	0.149	0.957	194.251
		all fields	6	0.649	2	0.166	0.946	154.599

^a SEP: standard error of prediction, S: steric, E: electrostatic, H: hydrophobic, A: H-bond acceptor, D: H-bond donor H + S + A* is represented S:H:A = 26.0%:47.8%:26.2%.

Table 3. Information of Statistical Significance and Predictive Power Presented in Cost Values for Top 10 Hypotheses Generated from Three Different Approaches

set	hypothesis no.	total cost	Δcost	rms deviation	correlation (r)
I^a	1	235.835	25.601	0.812	0.818
	2	237.133	24.303	0.828	0.810
	3	239.457	21.979	0.902	0.768
	4	239.554	21.882	0.873	0.787
	5	239.618	21.818	0.879	0.783
	6	239.699	21.737	0.880	0.783
	7	240.702	20.734	0.922	0.757
	8	242.147	19.289	0.947	0.741
	9	242.312	19.124	0.944	0.743
	10	244.042	17.394	0.984	0.716
II^b	1	252.479	16.120	1.079	0.691
	2	254.958	13.641	1.113	0.666
	3	255.214	13.385	1.120	0.661
	4	255.297	13.302	1.120	0.661
	5	255.352	13.247	1.122	0.659
	6	255.521	13.078	1.125	0.657
	7	255.940	12.659	1.128	0.655
	8	256.595	12.004	1.141	0.645
	9	256.634	11.965	1.141	0.644
	10	256.733	11.866	1.142	0.643
III^c	1	278.117	71.826	0.773	0.900
	2	284.203	65.740	0.881	0.868
	3	284.648	65.295	0.900	0.861
	4	286.615	63.328	0.937	0.848
	5	287.723	62.220	0.937	0.849
	6	291.222	58.721	0.998	0.826
	7	292.658	57.285	1.025	0.815
	8	293.008	56.935	1.019	0.818
	9	293.640	56.303	1.020	0.818
	10	293.755	56.188	1.015	0.821

^a I set: 60 proline-based inhibitors of MMP-1 aligned for CoMFA and CoMSIA and without any free conformations generated for pharmacophore hypotheses. Null cost of top-ten score hypotheses is 261.436 bits. Fixed cost is 215.016 bits. Configuration cost is 12.0748 bits. ^b II set: 60 proline-based inhibitors of MMP-1 aligned for CoMFA and CoMSIA and with free conformations generated for pharmacophore hypotheses. Null cost of top-ten score hypotheses is 268.599 bits. Fixed cost is 217.539 bits. Configuration cost is 14.5983 bits. ^c III set: 60 proline-based inhibitors (without any free conformations generated) of MMP-1 aligned for CoMFA and CoMSIA plus 11 nonproline inhibitors with free conformations generated for pharmacophore hypotheses. Null cost of top-ten score hypotheses is 349.943 bits. Fixed cost is 255.305 bits. Configuration cost is 15.3642 bits.

follows: **18^a**, **33^a**, **57^b**, **67^b**, **04^a**, **27^a**, **21^a**, **23^a**, **11^a**, **03^a**, **37^b**, **37^a**, **33^b**, **65^b**, **31^a**, **29^b**, **72^b**, **34^b**, **62^b**, **69^b**, **64^b**, **35^b**, **66^b**, **75^b**, and **12^c** (Table 1).

Table 4. Validation of the Hypo1 Hypothesis Using the CatScramble Program Implemented in Catalyst Software Package^a

validation no.	total cost	fixed cost	rms deviation	correlation (r)	configuration cost
Results for Unscrambled					
	278.117	255.305	0.773	0.900	15.364
Results for Scrambled					
trial 01	349.943	238.816	1.769	0.000	0.000
trial 02	345.439	253.925	1.599	0.429	13.984
trial 03	335.950	253.774	1.521	0.511	13.833
trial 04	337.302	247.226	1.580	0.452	7.285
trial 05	341.413	248.145	1.619	0.405	8.205
trial 06	349.943	238.816	1.769	0.000	0.000
trial 07	337.181	247.729	1.578	0.457	7.788
trial 08	343.542	254.880	1.577	0.455	14.939
trial 09	349.943	238.816	1.769	0.000	0.000
trial 10	335.711	246.541	1.582	0.449	6.600
trial 11	341.706	248.376	1.608	0.420	8.434
trial 12	336.356	248.620	1.570	0.462	8.680
trial 13	337.140	247.716	1.584	0.446	7.775
trial 14	336.071	247.716	1.577	0.454	7.775
trial 15	330.848	246.774	1.530	0.503	6.833
trial 16	342.425	246.281	1.634	0.386	6.340
trial 17	322.837	247.742	1.453	0.570	7.801
trial 18	347.996	247.830	1.678	0.319	7.889
trial 19	343.888	247.441	1.638	0.381	7.500

^a Null cost = 349.943.

The structure of inhibitor **14^a** (Spc2)-MMP-1 receptor complex⁸ was analyzed using the Ligplot 4.22 program²⁴ to identify some specific contacts between atoms of ligand and receptor. Some of these ligand atoms were served as the correspondence points in the subsequent structural alignment processes. The structure of inhibitor **19^a** (*N*-hydroxy-1*N*-[4-(4-fluorophenoxy)phenyl]sulfonyl-4-(*Z,E*-*N*-methoxyimino)-pyrrolidine-(2*R*)-carboxamide) (Table 1) which was the most active one among the series¹² studied was chosen as the target against which all the other structures were aligned.

The steric and electrostatic potential fields of CoMFA were calculated at each lattice intersection for each molecule being submerged into a regularly spaced lattice of 2.0 Å. The lattice was extended 4 Å units beyond the van der Waals volume of each molecule in the X, Y, and Z directions. The steric and electrostatic fields were calculated using the Tripos force field.²⁵ A C.3 carbon atom with radius 1.52 Å and charge +1.0 was used as the probe to calculate both the steric and electrostatic fields. The truncation for both the steric and

Table 5. Actual and Predicted Activities of the Training Set Based on the Best Pharmacophore Hypothesis Hypo1 and CoMFA and CoMSIA (S+H+A) Models

MMP-1 Inh #	Act pIC ₅₀	CoMFA Pred pIC ₅₀	CoMSIA S+H+A Pred pIC ₅₀	catalyst pharmacophore hypothesis Hypo1					
				Act IC ₅₀ (nM)	Pred IC ₅₀ (nM)	error	act activity scale ^a	pred activity scale ^a	principal
19 ^a	8.52	8.21	8.34	3	16	5.3	+++	++	2
14 ^a	8.00	7.82	7.91	10	15	1.5	++	++	2
16 ^a	7.88	8.06	7.94	13	28	2.1	++	++	2
13 ^a	7.76	7.68	7.68	17	200	12	++	++	0
47 ^b	7.76	7.54	7.56	17	18	1.1	++	++	2
26 ^b	7.72	7.46	7.41	19	53	2.8	++	++	2
52 ^b	7.72	7.57	7.60	19	22	1.2	++	++	2
56 ^b	7.69	7.76	7.63	20	18	1.1	++	++	2
22 ^a	7.69	7.70	7.81	20	11	1.8	++	++	2
15 ^a	7.65	7.84	7.74	22	32	1.4	++	++	2
25 ^a	7.61	7.74	8.15	24	33	1.4	++	++	2
24 ^a	7.58	7.33	7.44	26	41	1.6	++	++	2
28 ^a	7.48	7.39	7.38	33	47	1.4	++	++	2
58 ^b	7.37	7.35	7.39	42	24	1.7	++	++	2
27 ^b	7.13	7.26	7.14	74	62	1.2	++	++	2
26 ^a	7.12	7.24	7.00	75	93	1.2	++	++	2
20 ^a	7.09	7.31	7.13	81	70	1.2	++	++	2
36 ^a	7.06	7.25	7.22	86	120	1.4	++	++	2
38 ^b	7.06	6.88	6.88	87	410	4.8	++	++	2
59 ^b	7.00	7.03	6.95	100	330	3.3	++	++	0
17 ^a	6.96	6.70	6.59	109	36	3	++	++	0
28 ^a	6.89	6.87	6.77	127	320	2.5	++	++	0
36 ^b	6.85	6.79	6.80	140	330	2.3	++	++	0
40 ^b	6.85	6.79	6.80	140	330	2.3	++	++	0
21 ^b	6.82	6.77	6.79	150	53	2.8	++	++	0
30 ^a	6.80	7.03	6.89	155	340	2.2	++	++	0
29 ^a	6.73	6.90	6.80	184	400	2.2	++	++	0
61 ^b	6.69	6.73	6.81	200	360	1.8	++	++	0
45 ^b	6.67	6.57	6.81	210	140	1.5	++	++	0
34 ^a	6.67	6.75	6.53	212	200	1.1	++	++	0
48 ^b	6.65	6.78	6.67	220	200	1.1	++	++	0
60 ^b	6.63	6.75	6.79	230	330	1.4	++	++	0
19 ^b	6.60	6.78	6.75	250	120	2	++	++	0
63 ^b	6.58	6.46	6.56	260	170	1.5	++	++	0
39 ^b	6.58	6.44	6.62	260	150	1.7	++	++	0
23 ^b	6.58	6.84	6.84	260	370	1.4	++	++	0
42 ^b	6.56	6.67	6.59	270	330	1.2	++	++	0
44 ^b	6.55	6.66	6.60	280	190	1.4	++	++	0
06 ^a	6.53	6.50	6.67	293	200	1.5	++	++	0
53 ^b	6.52	6.57	6.57	300	250	1.2	++	++	0
55 ^b	6.50	6.63	6.57	310	200	1.5	++	++	0
71 ^b	6.45	6.41	6.43	350	330	1.1	++	++	0
20 ^b	6.42	6.30	6.36	380	320	1.2	++	++	0
43 ^b	6.35	6.40	6.42	440	320	1.4	++	++	0
25 ^b	6.34	6.25	6.14	450	1100	2.5	++	+	0
49 ^b	6.30	6.33	6.38	490	330	1.5	++	++	0
22 ^b	6.25	6.32	6.28	560	370	1.5	++	++	0
74 ^b	6.24	6.03	6.06	570	340	1.7	++	++	0
68 ^b	6.22	6.28	6.13	600	460	1.3	++	++	0
30 ^b	6.15	6.01	6.08	700	340	2	++	++	0
70 ^b	6.10	6.11	6.24	780	230	3.4	++	++	0
07 ^a	6.10	5.78	6.01	790	420	1.9	++	++	0
50 ^b	6.03	5.99	5.86	920	710	1.3	++	++	0
41 ^b	5.92	6.12	6.05	1200	570	2.1	+	++	1
24 ^b	5.88	6.03	6.18	1300	470	2.8	+	++	1
32 ^a	5.81	5.86	5.80	1548	2700	1.7	+	+	1
51 ^b	5.69	5.74	5.75	2000	1400	1.4	+	+	1
46 ^b	5.69	5.57	5.77	2000	560	3.6	+	++	1
73 ^b	5.53	5.66	5.43	2900	970	3	+	++	1
35 ^a	5.37	5.39	5.55	4235	2500	1.7	+	+	1
Inh #	name or code		ref						
01 ^c	Marimastat		29	0.77	1.6	2.1	+++	+++	2
02 ^c	Prinomastat		29	5.7	9	1.6	+++	+++	2
03 ^c	BMS-275291		29	25	2	13	++	+++	2
04 ^c	BAY-129566		29	5000	330	15	+	++	0
05 ^c	Cipemastat		30	3.0	4.3	1.4	+++	+++	2
06 ^c	SC-44463		31	20	17	1.2	++	++	2
07 ^c	Ilomastat		32	1.5	2.3	1.5	+++	+++	2
08 ^c	FYK-1388		33	6.0	9.5	1.6	+++	+++	2
09 ^c	CGS-27023A		34	33	31	1.1	++	++	2
10 ^c	Ro-319790		32	3.0	16	5.2	+++	++	2
11 ^c	BB-3644		35	10	13	1.3	++	++	2

^a Activity scale: highly active (<10 nM, +++), moderately active (10–1000 nM, ++), and inactive (>1000 nM, +).

Table 6. Actual and Predicted Activities of the Testing Set Based on the Best Pharmacophore Hypothesis Hypo1 and CoMFA and CoMSIA (S+H+A)

MMP-1 Inh #	Act pIC ₅₀	CoMFA Pred pIC ₅₀	CoMSIA S+H+A Pred pIC ₅₀	catalyst pharmacophore hypothesis Hypo1				
				Act IC ₅₀ (nM)	Pred IC ₅₀ (nM)	error	act activity scale*	pred activity scale*
18 ^a	8.52	8.05	8.29	3	19	6.2	+++	++
33 ^a	7.82	7.25	7.22	15	120	7.6	++	++
57 ^b	7.72	7.13	6.96	19	53	2.8	++	++
67 ^b	7.67	7.18	7.36	21	37	1.8	++	++
04 ^a	7.58	7.27	7.49	26	36	1.4	++	++
27 ^a	7.49	7.42	7.31	32	65	2	++	++
21 ^a	7.40	7.26	7.26	39	23	-1.7	++	++
23 ^a	7.13	6.80	7.13	74	48	-1.5	++	++
11 ^a	7.04	7.19	6.96	90	610	6.7	++	++
03 ^a	6.92	6.89	7.08	119	500	4.2	++	++
37 ^b	6.61	6.85	7.13	240	280	1.2	++	++
37 ^a	6.59	6.78	6.47	255	53	-4.8	++	++
33 ^b	6.56	6.53	6.69	270	100	-2.7	++	++
65 ^b	6.52	6.63	6.55	300	450	1.5	++	++
31 ^a	6.48	6.59	6.32	329	340	1	++	++
29 ^b	6.45	7.09	6.76	350	130	-2.7	++	++
72 ^b	6.31	5.99	6.24	350	340	-1	++	++
34 ^b	6.30	6.73	6.99	500	390	-1.3	++	++
62 ^b	6.23	6.59	6.74	580	350	-1.6	++	++
69 ^b	6.12	5.93	6.06	750	110	-6.8	++	++
64 ^b	6.03	6.06	6.09	920	450	-2	++	++
35 ^b	5.92	6.01	6.19	1200	350	-3.4	+	++
66 ^b	5.74	5.52	5.77	1800	820	-2.2	+	++
75 ^b	5.60	6.01	5.97	2500	450	-5.5	+	++
Inh # 12 ^c	name or code Batimastat	ref 31		0.99	7.7	7.8	+++	+++

electrostatic contributions computed was at ± 30 kcal/mol. The electrostatic contribution at the lattice intersections where maximum steric interactions were computed was ignored. Both the CoMFA steric and electrostatic fields computed were scaled by the standard option given in the program.

The same lattice where each molecule was submerged for CoMFA was used for CoMSIA. A C.3 atom of radius 1.0 Å and charge +1 was used as the probe to compute the CoMSIA similarity indices defined by Klebe et al.¹⁵ The similarity indices were calculated using Gaussian-type distance dependence between the probe and the atoms of the molecules of the data set. This functional form requires no arbitrary definition of cutoff limits and the similarity indices can be calculated at all lattice points inside and outside the molecule. The value of attenuation factor α was set to 0.3. In the SYBYL CoMSIA module,²⁶ the third power of the atomic radii was computed as the steric indices, the atomic partial charges were treated as the electrostatic indices, the atom-based parameters developed by Viswanadhan et al.²⁷ were used as the hydrophobic indices, and a rule-based method derived experimentally²⁸ was used as the hydrogen bond donor and acceptor indices.

Both the CoMFA or CoMSIA results were cross-validated using the SYBYL PLS (partial-least-squares)²¹ module. The column filtering was set at 2.0 kcal/mol for reducing the noise. The CoMFA and CoMSIA descriptors were treated as the independent variables, while the pIC₅₀ values were treated as the dependent ones in all the PLS regression analyses for deriving the 3D QSAR models. The optimum number of components used to derive the nonvalidated model was defined as the number of components leading to the highest cross-validated r^2 (q^2) and the lowest standard error of prediction. The non-cross-validated models were assessed by the conventional correlation coefficient r^2 , standard error

of estimate, and F -values. The non-cross-validated analyses were used to make predictions of the binding affinities of the inhibitors from the test set and to display the coefficient contour maps.

The structures of 12 more diversified nonproline MMP-1 inhibitors^{29–35} were drawn and constructed using the SYBYL 6.9.1 program²¹ for the purpose of external validation and pharmacophore characterization using the Catalyst 4.9 program.¹⁹ The biological activity expressed in pIC₅₀ was from 5.33 to 9.11 for the compound set. Each structure in the set was energy minimized using the CHARMM-like force field³⁶ within the Catalyst 4.9 program¹⁹ and subjected to a conformational analysis using the Poling algorithm.³⁷ All the parameters used were default settings except that the maximum number of conformers allowed for each molecule was set at 250. Based on the CoMSIA results, there were four pharmacophore features namely, E (E1,E2,E3), the three excluded volumes; H, hydrophobic; D, hydrogen-bond donor; and A, hydrogen-bond acceptor group; selected for the hypothesis generation process. The pharmacophores were manually or automatically generated for the 60 structurally aligned proline-based MMP-1 inhibitors in the training set or a set composed of the training set plus 11 structurally more diversified nonproline MMP-1 inhibitors constructed using the HypoGen module¹⁹ of the Catalyst 4.9 program.¹⁹ The top 10 scored hypotheses composed of the four pharmacophore features for each set were exported. The significance of the top (best) hypotheses generated was cross-validated using the CatScramble module of the Catalyst 4.9 program.¹⁹ To obtain a 95% confidence level, 19 random spreadsheets were generated, and every generated spreadsheet was submitted to the HypoGen module using the same experimental conditions (functions and parameters) as those used in the initial run.

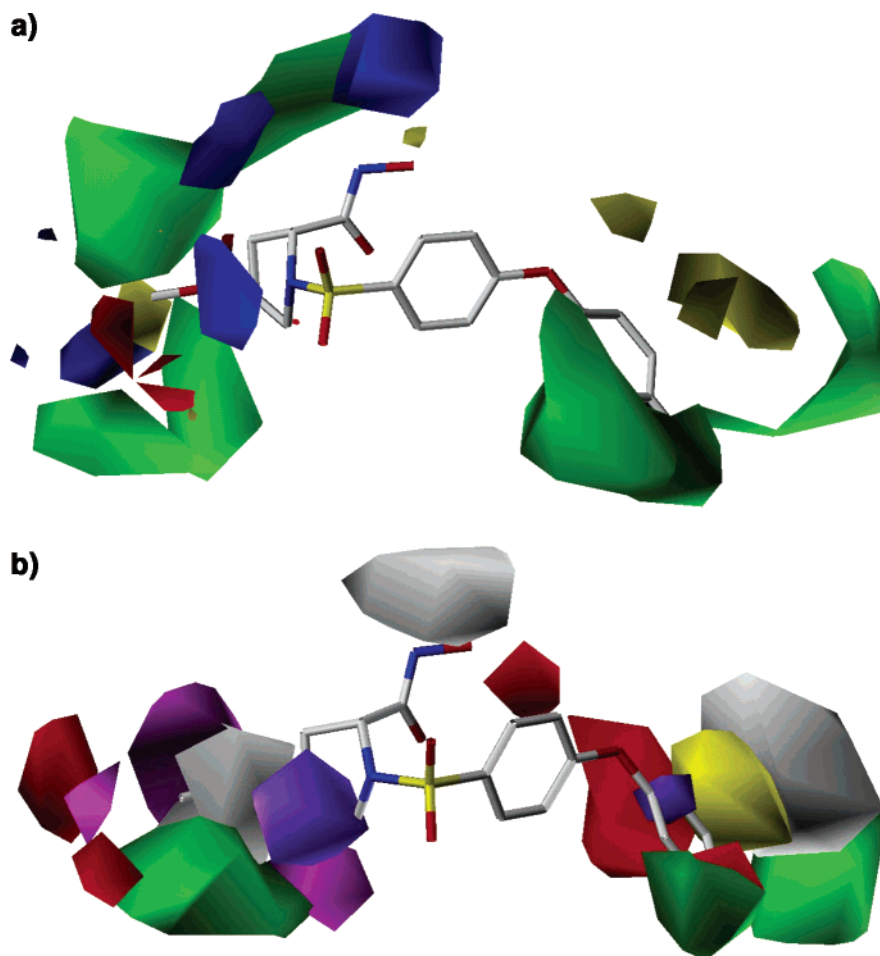


Figure 2. a. The CoMFA contours obtained from the aligned structures of 60 proline-based inhibitors of the training set was mapped onto the structure of inhibitor **19a**. Green contours represent favor, while yellow ones represent disfavor regions for steric interactions. Blue contours represent favor, while red contours represent disfavor regions for electrostatic interactions. b. The CoMSIA contours obtained from the aligned structures of 60 proline-based inhibitors of the training set were mapped onto the structure of inhibitor **19a**. Green contours represent favor, while yellow contours represent disfavor regions for steric interactions. Purple contours represent favor, while white contours represent disfavor regions for hydrophobic interactions. Magenta contours represent favor, while red contours represent disfavor regions for H-bond acceptor.

RESULTS AND DISCUSSION

There are three functional domains namely, the *N*-terminal propeptide, the protease or catalytic, and the *C*-terminal hemopexin-like domains have been characterized in MMPs.^{8,12} The catalytic domain is the most critical one in determining both the MMP substrate and MMPI specificity.¹² The only significant structural differences within the catalytic domains of various MMPs are the conformation of surface loops. The conserved MMP core has a rms deviation of 0.3 Å when residues 113–119, 128–133, 148–152, 160–164, 181–185, 194–198, and 211–225 of MMP-1 are superimposed with residues 117–123, 132–137, 152–156, 164–168, 185–189, 198–202, and 215–229 of MMP-3.⁸ Therefore, a series of proline-based ligands built from inhibitor **14a** which is the ligand of a MMP-3 receptor¹² can be docked into the ligand-depleted MMP-1 receptor. The structural features of the sulfone and diphenyl parts of the two are nearly the same, and they are superimposed well on each other (data not shown here).

As shown in Table 1, the biological activity of the 84 MMP-1 inhibitors studied is varied with two major substitutions on their structures namely, one is at the S1' group or the diphenyl ether substituent and the other one is at the C-4

carbon on the proline ring.^{8,12,14} The former and latter substitutions are designed for exploring the binding with the S1' and S2' pockets of the MMP-1 receptor, respectively. An analysis by Ligplot 4.22 program²⁴ on the structure of inhibitor **14a**-MMP-1 complex is presented in Figure 1. The result shows that both the diphenyl- and proline-based substituents of **14a** are properly positioned into the S1' and S2' pockets of the receptor. In addition, one oxygen atom of the sulfone group is positioned to serve as the hydrogen bond acceptor to the amide nitrogens of Leu 181 (2.86 Å) and Ala 182 (2.97 Å). The nitrogen atom of the hydroxamic acid group of **14a** also serves as hydrogen bond acceptors with the carbonyl oxygens of Ala 182 (2.76 Å) and Glu 219 (3.47 Å). Both the carbonyl and *N*-hydroxyl oxygens of the group also form a bidentate chelation with the Zn²⁺ ion with distances of 3.1 and 3.22 Å, respectively (Figure 1). However, these two distances are slightly larger than those between the same oxygens on the original ligand RS2 with the metal ion in which they are determined to be 1.9 and 2.1 Å, respectively.⁸

To proceed with 3D QSAR studies, some atoms on the structure of inhibitor **19a** which is the most active one in the series (Table 1) are treated as the correspondence points for

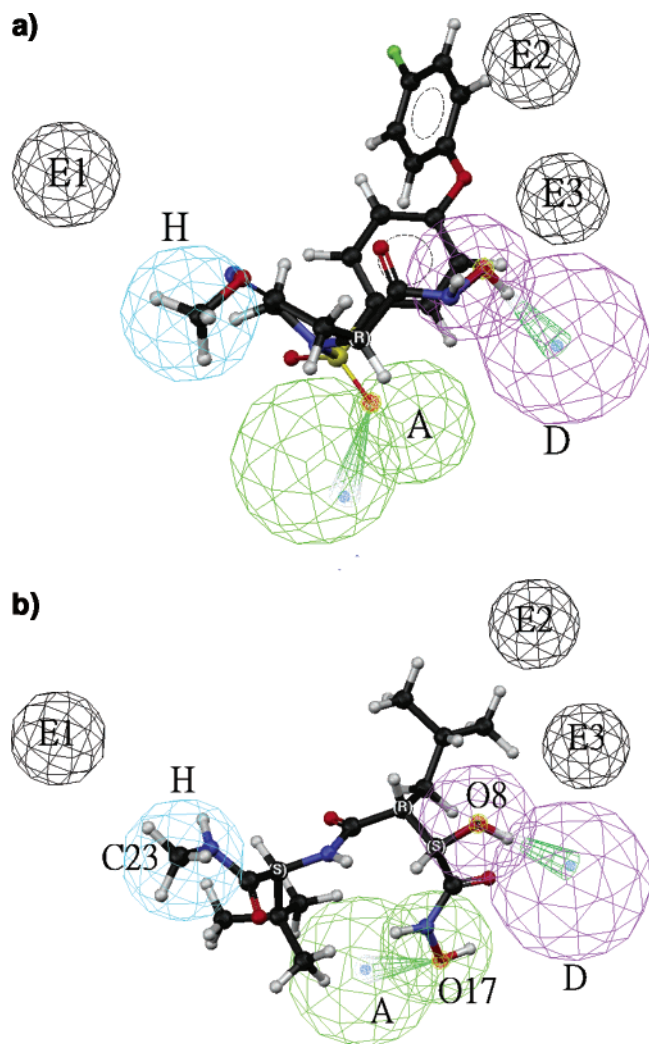


Figure 3. a. The Hypo1 hypothesis is mapped onto the structure of the most active proline-based inhibitor **19^a** of the training set. The pharmacophore features are color coded as follows: black spheres, three excluded volume (E1, E2, and E3); blue spheres, hydrophobic (H); violet spheres, hydrogen-bond donor (D); and green spheres, hydrogen-bond acceptor (A). b. The Hypo1 hypothesis is mapped onto the structure of the most active nonproline inhibitor **01^c** of the training set. The pharmacophore features are color coded as follows: black spheres, three excluded volume (E1, E2, and E3); blue spheres, hydrophobic (H); violet spheres, hydrogen-bond donor (D); and green spheres, hydrogen-bond acceptor (A).

performing the point alignment using the SYBYL FIT program.²¹ As shown in Figure 1, atoms on the proline and phenyl rings as well as those on the sulfone and hydroxamic acid groups are all included as the correspondence points. Some preliminary alignments for the training set are conducted by using at least three correspondence points arbitrarily selected in Figure 1. These aligned structural sets are then analyzed by the SYBYL CoMFA, CoMSIA, and PLS programs,²¹ and only those which give significant statistics are kept for further analyses. The best CoMFA result of these point alignments is presented in Table 2. The best alignment is obtained by using all the atoms labeled in Figure 1 on the structure of inhibitor **19^a** as the correspondence points. The value of cross-validated r^2 (q^2) computed for the best aligned structural set is 0.649, while that computed for the corresponding conventional r^2 (r^2) is 0.954 (Table 2).

A stepwise CoMSIA on the 60 structurally aligned MMP-1 inhibitors in the training set for CoMFA was performed, and the results are presented in Table 2. The five different fields (steric denoted as S, electrostatic denoted as E, hydrophobic denoted as H, H-bond acceptor denoted as A, and H-bond donor denoted as D) available were chosen one by one, or a combination of different fields was chosen for CoMSIA on the 60 structurally aligned MMP-1 inhibitors of the training set. As judged by q^2 values computed, no statistically significant CoMSIA result is obtained for each single field selected except for the H one (Table 2). A combination of H field with each of the other four ones was then conducted. Apparently, the statistics of these results is far better than those of the single field ones (Table 2). Since the combination by H and S fields gives the best CoMSIA statistics among these results, a third step was conducted by adding each of the rest of the three fields to the combined fields of H and S. This step creates an even better CoMSIA result with a q^2 value of 0.73 (the best CoMSIA model) due to the combination by H, S, and A fields (Table 2). However, no apparent improvement in CoMSIA statistics is obtained for further steps using a combination of either D or E field with the combined H+S+A fields or even a combination of all the five fields (Table 2). Therefore, the stepwise CoMSIA results indicate that the interaction of the MMP-1 inhibitors of the training set with their common receptor is best described by a combination of H, S, and A fields. The participation of A field is not picked by the CoMFA result. In fact, with a q^2 value of 0.655, the CoMSIA result of the combined H+S fields is nearly the same as that given by the CoMFA one (Table 2).

To further explore the nature of pharmacophore for the MMP-1 inhibitors, the Catalyst 4.9 program¹⁹ was used to generate some pharmacophore hypotheses using the following four features: the three excluded volumes (E1, E2, and E3), hydrophobic (H), hydrogen-bond donor (D), and hydrogen-bond acceptor (A) selected from the best stepwise CoMSIA results. There were three types of hypotheses generated namely, (I) hypotheses generated for the 60 structurally aligned and conformation fixed inhibitors of the training set, (II) hypotheses generated for the 60 inhibitors of the training set with conformations being freely generated within the Catalyst 4.9 program,¹⁹ and (III) hypotheses generated for a combination of the 60 structurally aligned and conformation fixed inhibitors of the training set plus 11 structurally more diversified nonproline inhibitors (**01^c**, **02^c**, **03^c**, **04^c**, **05^c**, **06^c**, **07^c**, **08^c**, **09^c**, **10^c**, and **11^c**) listed in Table 1 in which conformations of the latter are freely generated within the Catalyst 4.9 program.¹⁹ A comparison for the statistical significance of the top 10 hypotheses generated for each type of hypotheses is given in Table 3. The two cost differences namely between null and fixed or null and total costs should be greater than 70 or 60, and the configuration cost should be smaller than 17 bits for a good hypothesis generated.¹⁹ While the configuration cost computed for each hypothesis type is always smaller than 17 bits, the two cost differences computed for the top hypothesis generated for types I, II, and III hypotheses are 25.601, 46.320; 16.120, 51.06; and 71.826, 94.638; respectively. Therefore, only the top hypothesis of type III hypotheses (designated as Hypo1) generated meets the criteria of being a good hypothesis. The Hypo1 hypothesis was then cross-

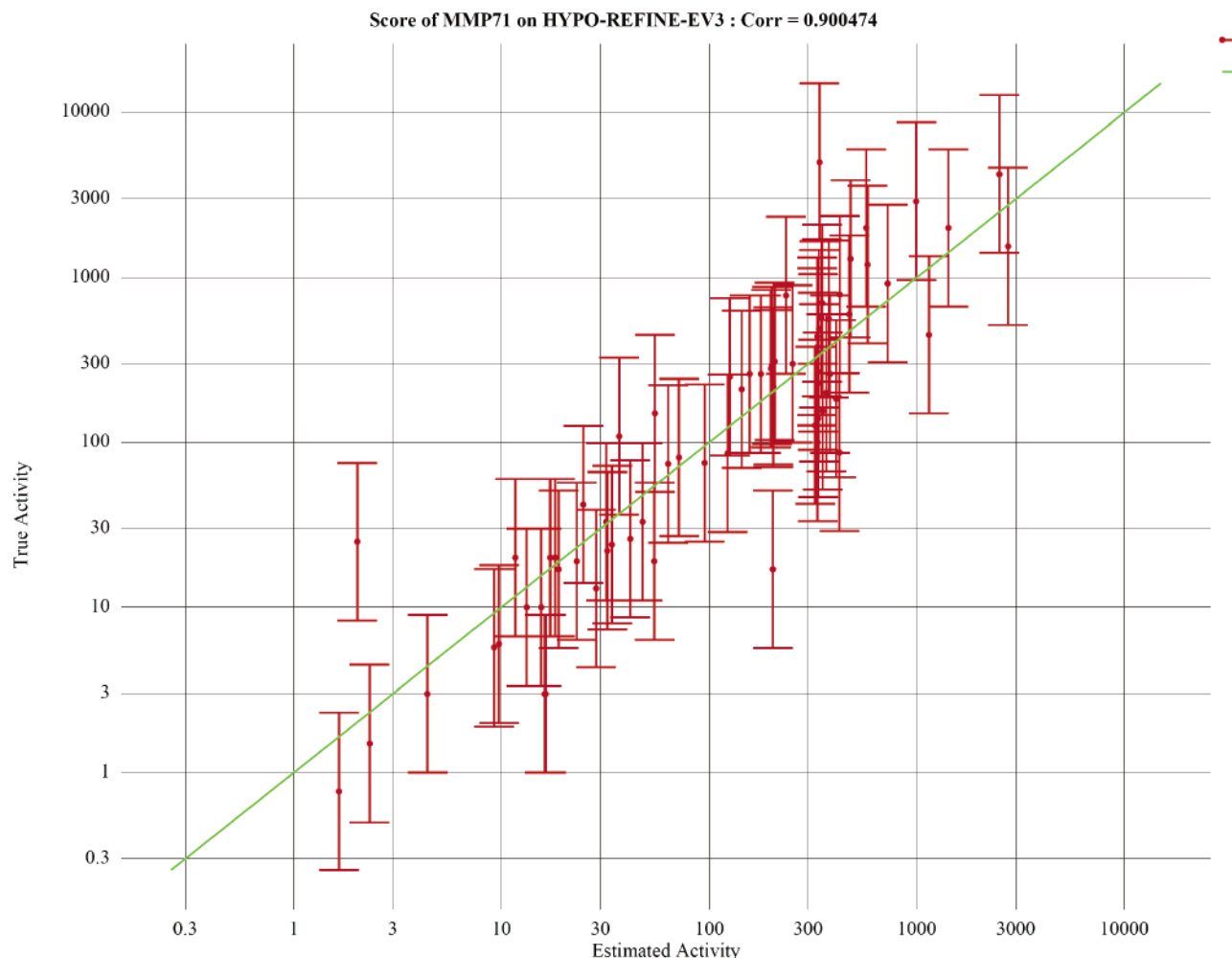


Figure 4. The regression of actual versus predicted activities by the Hypo1 hypothesis for the training set inhibitors onto a linear relationship. The training set includes 60 proline-based plus 11 nonproline inhibitors.

validated using the CatScramble module of the Catalyst 4.9 program.¹⁹ The validation proceeds with a random reassignment of activity values e.g., generation of random spreadsheets among the molecules of the training set. To achieve a confidence level of 95%, 19 random spreadsheets (random hypotheses) were generated, and the corresponding statistics are listed in Table 4. The validation clearly shows that the Hypo1 hypothesis is not generated by chance since its statistics are far more superior to those of the 19 random hypotheses generated (Table 4).

The S1' binding pocket of MMP-1 has been characterized as a relative small one and whose depth is defined by Arg214 at the bottom of the pocket.^{8,12} Both the CoMFA and CoMSIA contour maps also show that there are disfavor regions for steric interaction somewhere around the tip of the bulky diphenyl substituent on the MMP-1 inhibitors (Figure 2a,b). The HypoRefine module of the Catalyst 4.9 program¹⁹ was used to refine the Hypo1 hypothesis generated since the steric interaction appears to play an important role in the binding activity measured for the MMP-1 inhibitors. The algorithm examines for differences in the steric interaction between some active and inactive compounds selected and labeled in the "Principal" column of the input spreadsheet. The Hypo1 hypothesis was built based on the structures of 71 MMP-1 inhibitors where the 28 most and 7 least active compounds were chosen as "Active" (labeled with digit 2) or "Inactive" (labeled with digit 1) by the

HypoRefine run (Tables 5 and 6). The refined Hypo1 hypothesis was mapped onto the structures of the two most active inhibitors **19^a** and **01^c** (Table 1) in the training set, and the results are presented in Figure 3 (parts a and b, respectively). The CoMFA and CoMSIA contour maps agree with each other on the identification of favor regions for steric interaction (displayed with green contours by both CoMFA and CoMSIA) and which are around the diphenyl substituent or around the tip of the hydroxamic acid group (Figure 2a,b). The favor regions for electrostatic interaction identified by CoMFA (displayed with blue contours) also agree with the disfavor regions for hydrophobic interaction identified by CoMSIA (displayed with white contours) which are around the oxime substituent on the pyrrolidine ring or the hydroxamic acid group (Figure 2a,b). However, the favor or disfavor regions for hydrogen-bond acceptor are only picked by CoMSIA and are around the hydroxamic acid group (Figure 2b). This would agree with those given by the Hypo1 hypothesis where the structure of inhibitor **19^a** is mapped (Figure 3a) since both features of hydrogen-bond donor (displayed by violet spheres) and hydrogen-bond acceptor (displayed by green spheres) are identified to be around the hydroxamic acid group. The features of hydrophobic (displayed by blue spheres) and excluded volumes (displayed by black spheres) of the Hypo1 hypothesis (Figure 3a) are also matched with those given by the CoMSIA contours (Figure 2b) where the former is expressed with

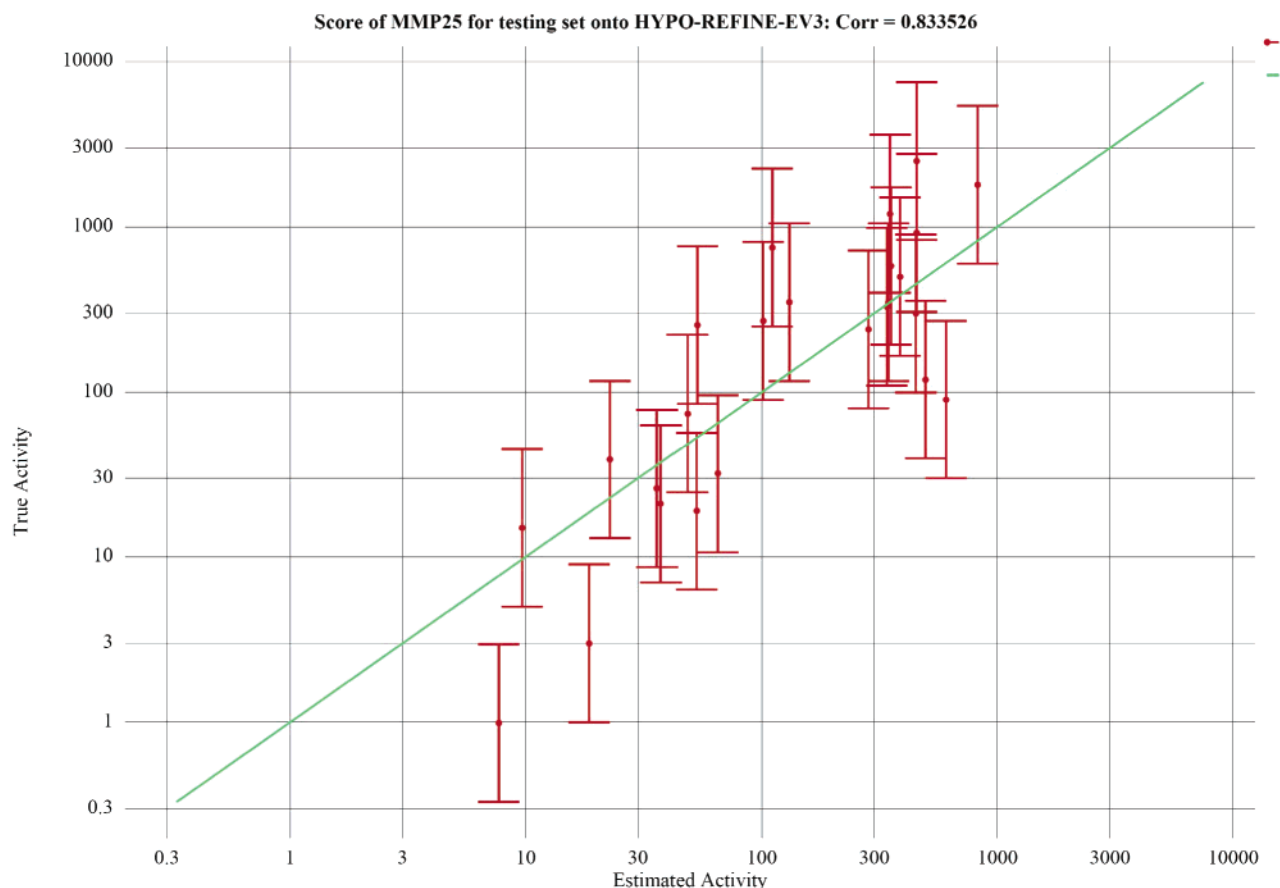


Figure 5. The regression of actual versus predicted activities by the Hypo1 hypothesis for the test set inhibitors onto a linear relationship. The test set includes 24 proline-based plus 1 nonproline inhibitors.

Table 7. Mapping of Hypo1 Hypothesis to Several Highly Active Inhibitors Selected from Both the Training and Test Sets

Inh #	E1	E2-E3	H	D	A
01^c (++++) ^a	methylamine	2-methylbutane	C ₂₃ (0.15) ^b	O ₈ (0.28,0.51) ^b	O ₁₇ (0.43,0.23) ^b
12^c (++++)	<i>N</i> -methyl-acetamide	<i>N</i> -hydroxy-acetamide	C ₂₉ (0.14)	O ₁₇ (0.17,0.51)	S ₁₀ (0.86,1.30)
07^c (++++) ^a	<i>N</i> -methyl-formamide	3-methyl-1H-indole	C ₁₆ (0.19)	N ₁₃ (0.08,0.84)	O ₂₈ (0.29,0.67)
05^c (++++) ^a	1-piperidin-1-yl-ethanone	<i>N</i> -hydroxy-acetamide	piperidine (0.19)	O ₁₉ (0.16,0.62)	O ₂₆ (0.78,0.90)
18^a (++)	O-methyl-hydroxylamine	methoxybenzene	C ₂₃ (0.34)	O ₁₉ (0.42,0.49)	O ₈ (1.16,1.17)
02^c (++++) ^a	4-methylpyridine	<i>N</i> -hydroxy-formamide	pyridine (0.11)	O ₂₁ (0.90,0.86)	O ₂ (0.71,0.70)
08^c (++++) ^a	butylbenzene	<i>N</i> -hydroxy-acetamide	benzene (0.25)	O ₁₇ (0.77,1.17)	O ₈ (0.59,0.69)
14^a (++) ^a	1-methoxy-4-methyl-benzene	<i>N</i> -hydroxy-formamide	C ₁₁ (0.27)	N ₁₈ (0.54,1.37)	O ₈ (0.83,0.79)
16^a (++) ^a	O-methyl-hydroxylamine	3-phenoxy-propan-1-ol	C ₂₃ (0.21)	O ₁₉ (0.64,0.49)	O ₈ (1.13,1.42)
33^a (++)	1-methyl-4-phenoxybenzene	<i>N</i> -hydroxy-acetamide	benzene (0.23)	N ₁₈ (0.70,2.25)	O ₈ (0.88,0.59)
57^b (++)	methoxybenzene	<i>N</i> -hydroxy-acetamide	benzene (0.33)	N ₁₈ (0.82,1.77)	O ₈ (0.67,1.00)
06^c (++) ^a	1-methoxy-4-methylbenzene	<i>N</i> -methyl-acetamide	benzene (0.07)	O ₂₇ (0.78,0.89)	O ₂₃ (0.97,1.02)
67^b (++)	methoxybenzene	<i>N</i> -hydroxy-acetamide	benzene (0.27)	N ₁₈ (0.77,1.87)	O ₈ (0.70,0.48)
03^c (++++) ^a	isobutane	2-methylbutane	C ₃₄ (0.05)	S ₈ (0.42,0.50)	O ₉ (0.43,0.55)
04^a (++)	pentane	<i>N</i> -hydroxy-acetamide	C ₂₃ (0.17)	N ₁₈ (0.53,1.61)	N ₂₁ (0.83,1.22)
04^c (++) ^a	benzenethiol	4'-chloro-4-methylbiphenyl	benzene (0.02)	(not mapped)	O ₃ (0.06,0.19)

^a The representative of testing set. ^b Hydrophobic (H); hydrogen-bond donor (D); and hydrogen-bond acceptor (A) displacement of the feature from the center of the location constraint.

purple contours around the oxime group on the pyrrolidine ring and the latter is represented with yellow contours (disfavor region for steric) around the diphenyl substituent. The Hypo1 hypothesis was also mapped onto the structure of **01^c**, the most active compound of the 11 structurally more diversified nonproline based MMP-1 inhibitors (Table 1) being used for the construction of the hypothesis. There are two hydroxyl groups on the structure, and both are correctly mapped onto the violet spheres representing the feature of hydrogen-bond donor (Figure 3b). Both the 2-methylbutane and methylamine moieties of the inhibitor are also identified

as the favor regions of three excluded volume (black spheres) or hydrophobic (blue sphere).

The C-4 substituent namely, the C=X bond (Table 1) on the proline ring is flat or in sp² orientation which is extended into the S_{2'} surface pocket of the MMP-1 receptor and is also considered to be hydrophobic.¹²⁻¹⁴ It has been suggested that these substituents could not directly affect the overall binding but would alter the conformation of proline ring to a more efficiently bound orientation.¹² This feature is also identified by the Hypo1 hypothesis on inhibitor **01^c** in which a blue sphere is mapped onto the methylamine group of the

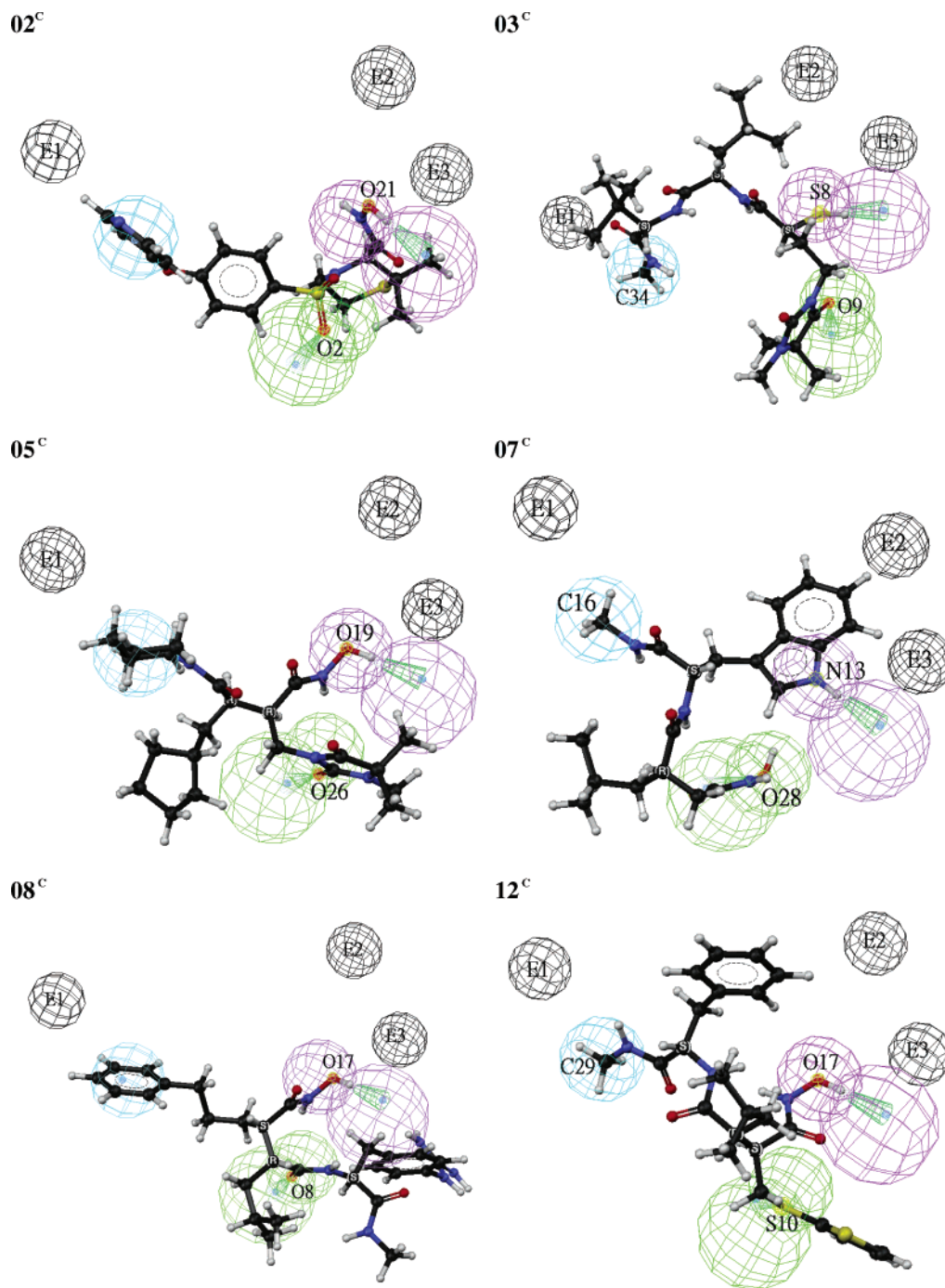


Figure 6. The mapping of Hypo1 hypothesis onto the structures of 6 highly active nonproline inhibitors (**02^c**, **03^c**, **05^c**, **07^c**, **08^c**, and **12^c**) selected from both the training and test sets. The pharmacophore features are color coded as follows: black spheres, three excluded volume (E1, E2, and E3); blue spheres, hydrophobic (H); violet spheres, hydrogen-bond donor (D); and green spheres, hydrogen-bond acceptor (A).

inhibitor (Figure 3b). The actual and predicted biological activities by the best CoMFA (Table 2 and Figure 2a) and CoMSIA (Table 2 and Figure 2b) models and the Hypo1 hypothesis for each training set inhibitors are listed and compared in Table 5, while those for test set inhibitors are given in Table 6, respectively. Regression of the predicted activities for both the training and test sets by the best CoMSIA model onto a linear relationship versus the actual ones gives correlation coefficients R^2 of 0.952 and 0.836, respectively. The regression results by the same relationship for those predicted by the Hypo1 hypothesis are shown in Figures 4 and 5 for the training and test set inhibitors,

respectively. The corresponding R^2 values obtained for each plot are 0.900 and 0.834 (Figures 4 and 5), respectively. This indicates that the prediction results given by the best CoMSIA model are well reproduced by the Hypo1 hypothesis constructed using the pharmacophore features selected by the best CoMSIA model. Note that most of the highly active inhibitors correctly predicted by the Hypo1 hypothesis are the 11 structurally more diversified nonproline inhibitors of the training set (Table 5). Moreover, the only structurally more diversified nonproline inhibitor grouped into the test set is inhibitor **12^c** and which is also correctly predicted by Hypo1 hypothesis as a highly active compound (Table 6).

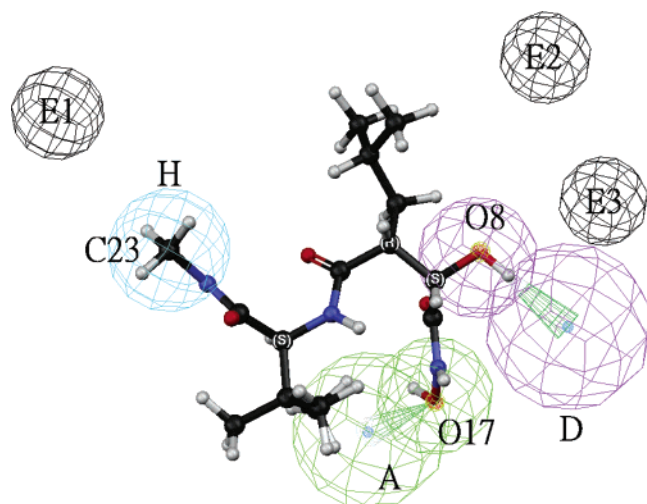


Figure 7. The Hypo1 hypothesis is mapped onto the flexibly docked conformation of inhibitor **01c**.

The mapping of Hypo1 hypothesis onto the structures of some highly and medium active inhibitors predicted for both the training and test sets is presented in Table 7. The inhibitors listed from top to bottom in the table are arranged in the order of actual activities measured. The following 8 inhibitors namely, **01c**, **12c**, **07c**, **05c**, **18a**, **02c**, **08c**, and **14a**, are considered as highly active since their pIC_{50} values measured are greater than 8.0 (Tables 1 and 7). Apparently, most of these highly active inhibitors are predicted by the Hypo1 hypothesis as highly active and labeled with the '+++' scale (Table 7). The only two inhibitors predicted incorrectly and labeled with the '++' scale are the proline-based ones namely **18a** and **14a** (Tables 1 and 7). However, there is also a medium active inhibitor namely **03c** being incorrectly predicted by the Hypo1 hypothesis as highly active and labeled with the '+++' scale (Table 7). Therefore, the overall percentage of correct prediction by the Hypo1 hypothesis for these inhibitors is 81% (Table 7). Each of the structural features of the Hypo1 hypothesis is also mapped correctly onto the corresponding functional group of each inhibitor as can be judged by the displacement values parenthesized underneath each functional group (Table 7). The mapping of Hypo1 hypothesis onto the structures of 6 highly active nonproline based inhibitors **02c**, **03c**, **05c**, **07c**, **08c**, and **12c** is also presented in Figure 6. The most diversified structural features of these nonproline based inhibitors are identified by the Hypo1 hypothesis as the regions mapped with the three excluded volume (represented with the black spheres) (Table 7 and Figure 6). These regions may be worth further exploration since not much difference is found in the mapping of other structural features of these inhibitors by the Hypo1 hypothesis. A further validation for the Hypo1 hypothesis is also performed by docking the most active inhibitor **01c** studied into the active site of the MMP-1 receptor (PDB entry 966c)⁸ using the DOCK 4.0.2 program.^{38,39} The pocket for the active site is constructed by including all the residue atoms that are within a distance of 6 Å from every ligand atom. While Gasteiger-Hückel²² charges are assigned for each ligand atom, the AMBER95 charges⁴⁰ are used for each receptor one. The parameters of the flexible docking process such as the maximum anchor orientations and configuration per cycle are set as 300 and 50, respectively. The Hypo1 hypothesis is then mapped onto

the docked conformation of inhibitor **01c**, and the result is presented in Figure 7. The predicted pIC_{50} by the Catalyst program¹⁹ for the mapped inhibitor is 2.21 which is in close agreement with that given for the original conformation of the same compound (see Figure 3b and Table 5). Except the D (hydrogen-bond donor) feature, both the H (hydrophobic) and A (hydrogen-bond acceptor) features of the two conformations are well matched (Figures 3b and 7), indicating that the Hypo1 hypothesis constructed can be used to represent the MMP-1 receptor.

CONCLUSION

In this report we have shown that structural features extracted from a 3D QSAR model can be used to assist the construction of a top pharmacophore hypothesis by the Catalyst program. The two most commonly used methods for building a 3D QSAR model are CoMFA and CoMSIA. While each of the two has its own merit, we found that the latter one is more versatile in identifying some important molecular features contributed to the overall activity of a molecule analyzed. However, both rely on the preliminary structure alignment and have a limitation on the use of structurally more diversified molecules in the analysis. The disadvantages may be complemented by using the Catalyst program, a pharmacophore hypothesis building method which allows the identification of structural features from numerous conformations generated for a series of molecules studied. No preliminary structure alignment is required for the series of molecules analyzed by the method. The statistics or the correlation coefficient R^2 of a linear regression model of the predicted versus actual activities given by the Catalyst program was around 0.6 if free conformations for all the 60 proline-based inhibitors of the training set were used in the hypothesis generation process. However, these statistics were enhanced to 0.8 if fixed or aligned conformations for all the 60 proline-based inhibitors were used in the hypothesis generation process. This reveals that one cannot simply count on the generation of a variety of conformations to achieve the construction of a better pharmacophore hypothesis. By including 11 structurally more diversified nonproline inhibitors in the training set and by allowing the conformations of only these inhibitors to change freely during the hypothesis generation process, a top hypothesis of R^2 of 0.9 was constructed.

The other difficulty encountered in using the Catalyst program is to choose some representative pharmacophore features for building a hypothesis. The task is laborious since there are 11 default features provided by the program, and only some of them are selected in one run. We are aware of a recent development of some 4D-QSAR techniques^{41,42} in choosing the representative pharmacophore features for building a hypothesis. However, we have shown here that a top hypothesis can be constructed using some structural features extracted from some stepwise CoMSIA models. These CoMSIA features are produced from the fixed or aligned conformation of 60 proline-based inhibitors and may be essential for the activities of these molecules. These CoMSIA features may be slightly modified in the top hypothesis generated since some structurally more diversified nonproline inhibitors were added into the hypothesis generation process. A hypothesis generated using structures of

purely nonproline inhibitors would not achieve the best prediction accuracy of the top hypothesis. Most of the nonproline inhibitors are more active than those of the proline-based ones, and yet they are better predicted by the top hypothesis constructed from the CoMSIA features extracted from the proline-based ones.

ACKNOWLEDGMENT

This work is supported in part by a grant from the National Science Council, (NSC92-2313-B-007-002). The SYBYL CoMFA, CoMSIA, and PLS plus the Catalyst 4.9 studies were conducted at the National Center for High Performance Computing, Taiwan.

REFERENCES AND NOTES

- Whittaker, M.; Floyd, C. D.; Brown, P.; Gearing, A. J. H. Design and Therapeutic Application of Matrix Metalloproteinase Inhibitors. *Chem. Rev.* **1999**, *99*, 2735–2776.
- Leff, R. L. Clinical trials of a stromelysin inhibitor. Osteoarthritis, matrix metalloproteinase inhibition, cartilage loss, surrogate markers, and clinical implications *Ann. N. Y. Acad. Sci.* **1999**, *878*, 201–207.
- Clark, I. M.; Rowan, A. D.; Cawston, T. E. Matrix metalloproteinase inhibitors in the treatment of arthritis. *Curr. Opin. Anti-inflammatory Immunomodulatory Invest. Drugs* **2000**, *2*, 16–25.
- Shlopov, B. V.; Lie, W. R.; Mainardi, C. L.; Cole, A. A.; Chubinskaya, S.; Hasty, K. A. Osteoarthritic lesions: involvement of three different collagenases. *Arthritis Rheum.* **1997**, *40*, 2065–2074.
- Ahrens, D.; Koch, A. E.; Pope, R. M.; Steinpicarella, M.; Niedbala, M. J. Expression of matrix metalloproteinase 9 (96-kd gelatinase B) in human rheumatoid arthritis. *Arthritis Rheum.* **1996**, *39*, 1576–1587.
- Blaser, J.; Triebel, S.; Maajosthusmann, U.; Rimisch, J.; Krahmteblowski, U.; Freudenberg, W.; Fricke, R.; Tschesche, H. Determination of metalloproteinases, plasminogen-activators and their inhibitors in the synovial fluids of patients with rheumatoid arthritis during chemical synoviothetesis. *Clin. Chim. Acta* **1996**, *244*, 17–33.
- 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.
- Lovejoy, B.; Welch, A. R.; Carr, S.; Luong, C.; Broka, C.; Hendricks, R. T.; Campbell, J. A.; Walker, K. A. M.; Matin, R.; Van Wart, H.; Browner, M. F. Crystal structures of MMP-1 and -13 reveal the structural basis for selectivity of collagenase inhibitors. *Nature Struct. Biol.* **1999**, *6*, 217–221.
- Knight, C. G.; Willenbrock, F.; Murphy, G. A. A novel coumarin-labeled peptide for sensitive continuous assays of the matrix metalloproteinases. *FEBS Lett.* **1992**, *296*, 263–266.
- Okada, Y.; Nagase, H.; Harris, E. D. A metalloproteinase from human rheumatoid synovial fibroblasts that digests connective tissue matrix components. Purification and characterization. *J. Biol. Chem.* **1986**, *261*, 14245–14255.
- 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.
- Cheng, M.; De, B.; Almstead, N. G.; Pikul, S.; Dowty, M. E.; Dietsch, C. R.; Dunaway, C. M.; Gu, F.; Hsieh, L. C.; Janusz, M. J.; Taiwo, Y. O.; Natchus, M. G.; Hudlicky, T.; Mandel, M. Design, Synthesis, and Biological Evaluation of Matrix Metalloproteinase Inhibitors Derived from a Modified Proline Scaffold. *J. Med. Chem.* **1999**, *42*, 5426–5436.
- Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwat, 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.
- Natchus, M. G.; Bookland, R. G.; De, B.; Almstead, N. G.; Pikul, S.; Janusz, M. J.; Heitmeyer, S. A.; Hookfin, E. B.; Hsieh, L. C.; Dowty, M. E.; Dietsch, C. R.; Patel, V. S.; Garver, S. M.; Gu, F.; Pokross, M. E.; Mieling, G. E.; Baker, T. R.; Foltz, D. J.; Peng, S. X.; Bornes, D. M.; Strojnowski, M. J.; Taiwo, Y. O. Development of New Hydroxamate Matrix Metalloproteinase Inhibitors Derived from Functionalized 4-Aminoproline. *J. Med. Chem.* **2000**, *43*, 4948–4963.
- Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- Baron, M.; Constantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimation (GOLPE) – An advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct. Act. Relat.* **1993**, *12*, 9–20.
- 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.
- Amin, E. A.; Welsh, W. J. 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.
- Catalyst, version 4.9 (software package); Accelrys, Inc. (previously known as Molecular Simulations, Inc.): San Diego, CA, 2003, <http://www.accelrys.com>.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourno, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- SYBYL 6.9.1; The Tripos Associates; 1699 S. Hanley Rd., St. Louis, MO.
- Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **1995**, *8*, 127–134.
- Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- 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.
- 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.
- Klebe, G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **1994**, *237*, 212–235.
- Wada, C. K.; Holms, J. H.; Curtin, M. L.; Dai, Y.; Florjancic, A. S.; Garland, R. B.; Guo, Y.; Heyman, H. R.; Stacey, J. R.; Steinman, D. H.; Albert, D. H.; Bouska, J. J.; Elmore, I. N.; Goodfellow, C. L.; Marcotte, P. A.; Tapang, P.; Morgan, D. W.; Michaelides, M. R.; Davidsen, S. K. Phenoxypheyl sulfone N-formylhydroxylamines (retrohydroxamates) as potent, selective, orally bioavailable matrix metalloproteinase inhibitors. *J. Med. Chem.* **2002**, *45*, 219–232.
- Lewis, E. J.; Bishop, J.; Bottomley, K. M. K.; Bradshaw, D.; Brewster, M.; Broadhurst, M. J.; Brown, P. A.; Budd, J. M.; Elliott, L.; Greenham, A. K.; Johnson, W. H.; Nixon, J. S.; Rose, F.; Sutton, B.; Wilson, K. British. Ro 32–3555, an orally active collagenase inhibitor, prevents cartilage breakdown in vitro and in vivo. *J. Pharmacol.* **1997**, *121*, 540–546.
- Fray, M. J.; Dickinson, R. P.; Huggins, J. P.; Occleston, N. L. A potent, selective inhibitor of matrix metalloproteinase-3 for the topical treatment of chronic dermal ulcers. *J. Med. Chem.* **2003**, *46*, 3514–3525.
- Yamamoto, M.; Tsujishita, H.; Hori, N.; Ohishi, Y.; Inoue, S.; Ikeda, S.; Okada, Y. Inhibition of membrane-type 1 matrix metalloproteinase by hydroxamate inhibitors: an examination of the subsite pocket. *J. Med. Chem.* **1998**, *41*, 1209–1217.
- Fujisawa, T.; Igeta, K.; Odake, S.; Morita, Y.; Yasuda, J.; Morikawa, T. Highly water-soluble matrix metalloproteinases inhibitors and their effects in a rat adjuvant-induced arthritis model. *Bioorg. Med. Chem.* **2002**, *10*, 2569–2581.
- 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 nonpeptidic, potent, and orally active stromelysin inhibitor that blocks cartilage degradation in rabbits. *J. Med. Chem.* **1997**, *40*, 2525–2532.
- Wall, L.; Talbot, D. C.; Bradbury, P.; Jodrell, D. I. Br. A phase I and pharmacological study of the matrix metalloproteinase inhibitor BB-3644 in patients with solid tumours. *J. Cancer.* **2004**, *90*, 800–804.

- (36) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (37) Smellie, A.; Teig, S. L.; Tobwin, P. Poling: Promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171–187.
- (38) Ewing, T.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175–89.
- (39) Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078–82.
- (40) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–97.
- (41) Senese, C. L.; Hopfinger, A. J. Receptor-Independent 4D-QSAR Analysis of a Set of Norstatine Derived Inhibitors of HIV-1 Protease. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1297–1307.
- (42) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, G. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.

CI049824G