3D-Quantitative Structure–Activity Relationships of Human Immunodeficiency Virus Type-1 Proteinase Inhibitors: Comparative Molecular Field Analysis of 2-Heterosubstituted Statine Derivatives–Implications for the Design of Novel Inhibitors

Romano T. Kroemer, Peter Ettmayer, and Peter Hecht*

SANDOZ Forschungsinstitut Ges. m. b. H, Brunnerstr. 59, A-1235 Vienna, Austria

Received May 11, 1995[⊗]

A set of 100 novel 2-heterosubstituted statine derivatives inhibiting human immunodeficiency virus type-1 proteinase has been investigated by comparative molecular field analysis. In order to combine the structural information available from X-ray analyses with a predictive quantitative structure–activity relationship (QSAR) model, docking experiments of a prototype compound into the receptor were performed, and the 'active conformation' was determined. The structure of the receptor was taken from the published X-ray analysis of the proteinase with bound MVT-101, the latter compound exhibiting high structural similarity with the inhibitors investigated. The validity of the resulting QSARs was confirmed in four different ways. (1) The common parameters, namely, the cross-validated r^2 values obtained by the leave-one-out (LOO) method ($r^2_{cv} = 0.572–0.593$), and (2) the accurate prediction of a test set of 67 compounds ($q^2 = 0.552–0.569$) indicated a high consistency of the models. (3) Repeated analyses with two randomly selected cross-validation groups were performed and the cross-validated r^2 values monitored. The resulting average r^2 values were of similar magnitudes compared to those obtained by the LOO method. (4) The coefficient fields were compared with the steric and electrostatic properties of the receptor and showed a high level of compatibility. Further analysis of the results led to the design of a novel class of highly active compounds containing an additional linkage between P_1' and P_3'. The predicted activities of these inhibitors were also in good agreement with the experimentally determined values.

Introduction

The proteinase (PR) is a key enzyme in the life cycle of the human immunodeficiency virus (HIV).[1,2] It is responsible for proteolytic processing of the gag precursor protein in the late stage of HIV replication. Thus the inhibition of HIV proteinase (HIV-PR) has become an important therapeutic target with potential usefulness for the chemotherapy of the acquired immune deficiency syndrome (AIDS).[3–5] Recently, first evidence of clinical effectiveness of an HIV-PR inhibitor in HIV-infected patients was reported.[6]

Extensive work has been performed on this enzyme: In this context the structures of more than 170 different HIV-1 proteinase/inhibitor complexes have been elucidated by X-ray analysis.[7] These studies clearly revealed that the proteinase is structurally conserved[8] and highly independent from the cocrystallized inhibitor molecules. A single water molecule (herewith referred to water-511) has been found in many of the complexes and appeared to be important for binding of the inhibitors of the enzyme via H-bond formation. Despite this large body of structural information, its predictive use in drug design has not yet been possible.[8] In an attempt to overcome this obstacle, a combination of different computational chemistry methods, i.e., docking experiments and 3D-quantitative structure–activity relationships (3D-QSAR), was applied.

Comparative molecular field analysis (CoMFA) is a well-established 3D-QSAR method which aims to establish a relationship between biological activities of a set of compounds and their steric and electrostatic properties.[9,10] After the selection of the 'active conformation', 3D-representations of the molecules are generated correspondingly and superpositioned according to predefined rules. Consequently, the steric and electrostatic interaction energies between a probe atom of a given charge/size and each of the structures are calculated at the surrounding points of a predefined grid. In order to derive linear equations from the resulting highly underdetermined matrices, a regression method called partial least squares (PLS) is applied.[11] This statistical method is not sensitive to possibly colinearity of the underlying descriptor matrix as it operates with latent variables and is usually applied in combination with cross-validation,[12] in order to check for consistency (predictiveness) of the model under consideration.

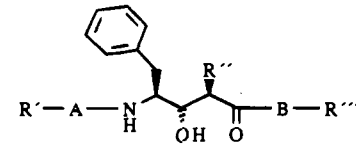
The main use of CoMFA studies is to predict the target properties (the biological activities) of newly designed compounds. Additionally, an inspection of displays of the CoMFA results may help in understanding the steric and electrostatic features of the compounds necessary for good activity, thus aiding in the design of new compounds. In particular, the QSAR coefficient contour maps give a 3D-display of the QSAR with its hundreds or thousands of terms (coefficients). By plotting contour maps of the coefficients of high magnitude, those regions can be identified in which structural (steric) or electrostatic modifications of the compounds have significant influence on the respective target property. Also very useful is the analysis of the so-called σ-fields which reveal the areas of high- and low-electrostatic or steric-variability, indicating those

* To whom correspondence should be addressed at Tripos GmbH, Martin-Kollar-Str. 15, D-81829 Munich, Germany. Tel.: ++49-89-451030-0. Fax: ++49-89-451030-30.

⊗ Abstract published in *Advance ACS Abstracts*, October 1, 1995.

0022-2623/95/1838-4917$09.00/0

Table 1. Structures, Measured Activities, and Residuals of the Training Set



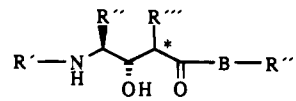
SDZ	R′	A	R″	B	R‴	biol act[a]	residual[b] 1	6	9
282215	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2Ph$	−0.79	0.30	0.20	0.22
282310	$PhCH_2OCO$	Val	$HNCH_2Ph$	Ile	$HNCH_2Ph$	−1.64	−0.05	−0.13	−0.16
282311	$PhCH_2OCO$	Val	$HNCH_2Ph$	Leu	$HNCH_2Ph$	−1.04	0.11	−0.14	−0.12
282312	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	OC2H5	−3.08	−0.50	−0.30	−0.50
282314	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNC(CH_3)_3$	−1.86	0.37	0.37	0.44
282327	$PhCH_2OCO$	Val	$HN(CH_2)_3CH_3$	Val	$HNCH_2Ph$	−1.36	−0.37	−0.25	−0.23
282329	$PhCH_2OCO$	Val	HNPh	Val	$HNCH_2Ph$	−0.84	0.05	0.16	0.27
282349	$PhCH_2OCO$	Val	$HNCH_2Ph$	Asn	$HNCH_2Ph$	−1.62	0.54	0.39	0.39
282350	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2$-2-pyridyl	−1.08	−0.09	−0.06	0.04
282351	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2$-3-pyridyl	−1.13	0.10	−0.08	−0.04
282365	$PhCH_2OCO$	Ile	$HNCH_2Ph$	Val	$HNCH_2Ph$	−0.92	0.23	0.23	0.22
282366	$PhCH_2OCO$	Abu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.23	0.14	0.13	0.11
282382	$PhCH_2OCO$	Val	HN(cyclohexyl)	Val	$HNCH_2Ph$	−1.53	1.10	0.94	0.82
282388	Boc	Val	$HNCH_2Ph$	Val	$HNCH_2Ph$	−2.33	0.20	0.20	0.02
282389	H	Val	$HNCH_2Ph$	Val	$HNCH_2Ph$	−2.87	−0.37	−0.22	−0.11
282390	$PhCH_2OCO$	Gln	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.65	−0.08	−0.13	−0.17
282391	$PhCH_2OCO$	Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.81	−0.42	−0.29	−0.32
282392	$PhCH_2OCO$	Ala	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.53	−0.13	−0.03	−0.05
282395	$PhCH_2OCO$	Asn	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.02	0.34	0.27	0.17
282396	Boc	bond	$HNCH_2Ph$	Val	$HNCH_2Ph$	−2.20	−0.40	−0.09	0.00
282412	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2$-4-pyridyl	−1.04	0.15	0.05	0.12
282423	$PhCH_2OCO$	Val	$HN(CH_2)_2Ph$	Val	$HNCH_2Ph$	−1.53	−0.04	0.06	0.07
282429	$PhCH_2OCO$	Val	$HNCH_2Ph$	*t*-Leu	$HNCH_2Ph$	−1.30	0.10	0.04	0.00
282450	$PhCH_2OCO$	Ser	$HNCH_2Ph$	Val	$HNCH_2Ph$	−2.18	−0.68	−0.55	−0.62
282453	2-quinolinoyl	Val	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.01	0.48	0.49	0.32
282455	$PhCH_2OCO$	His	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.46	−0.01	0.14	0.16
282456	2-$PyCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2Ph$	−0.93	0.28	0.18	0.18
282457	$PhCH_2OCO$	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−0.96	0.11	0.04	−0.04
282479	$PhCH_2OCO$	Val	$HN(CH_2)_2$-2-indolyl	Val	$HNCH_2Ph$	−1.79	0.12	0.09	0.09
282480	$PhCH_2OCO$	Val	$HNCH_2$-1-naphthyl	Val	$HNCH_2Ph$	−1.41	0.27	0.29	0.31
282518	$PhCH_2OCO$	Val	$HNCH_2$-3-pyridyl	Val	$HNCH_2Ph$	−0.85	0.23	0.06	0.13
282529	Boc	bond	SCH_2Ph	Val	$HNCH_2Ph$	−1.43	0.15	−0.07	0.08
282539	2-quinolinoyl	Val	$HNCH_2Ph$	Val	$HNCH_2$-2-pyridyl	−1.63	−0.27	−0.14	−0.24
282540	$PhCH_2OCO$	Val	SCH_2Ph	Val	$HNCH_2Ph$	−1.00	0.28	0.15	0.20
282541	$PhCH_2OCO$	Val	$HN(CH_2)_2$-2-pyridyl	Val	$HNCH_2Ph$	−2.30	−0.76	−0.38	−0.30
282542	$PhCH_2OCO$	Val	$HNCH_2$-*p*-$(OCH_3)Ph$	Val	$HNCH_2Ph$	−0.87	0.05	−0.03	−0.07
282547	2-quinolinoyl	Asn	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.38	−0.30	−0.01	−0.10
282558	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2Ph$	−1.23	−0.08	−0.25	−0.25
282632	$PhCH_2OCO$	Val	$HN(CH_2)_2$-*p*-(OH)Ph	Val	$HNCH_2Ph$	−1.48	0.08	0.21	0.29
282658	Boc	bond	$HNCH_2Ph$	bond	$HNCH_2Ph$	−3.40	0.02	−0.07	0.11
282659	3-methyl-(2*S*)-hydroxybutanoyl	bond	$HNCH_2Ph$	Val	$HNCH_2Ph$	−2.90	−0.38	−0.11	0.02
282664	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2$-2-pyridyl	−0.81	0.24	0.25	0.34
282666	$PhCH_2OCO$	Val	$HNCH_2$-*m*-$(OCH_3)Ph$	Val	$HNCH_2$-2-benzimidazolyl	−0.99	0.11	0.14	0.32
282700	$PhCH_2OCO$	Val	*p*-(HN)biphenyl	Val	$HNCH_2Ph$	−1.77	0.03	−0.07	−0.09
282701	$PhCH_2OCO$	Val	*p*-(HN)biphenyl	Val	$HNCH_2$-2-benzimidazolyl	−0.95	0.07	0.07	0.28
282713	2-pyridylCH_2OCO	Val	$HNCH_2Ph$	Val	$HNCH_2$-2-pyridyl	−1.43	−0.35	−0.28	−0.20
282714	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2$-*p*-(Br)Ph	−1.18	−0.19	−0.33	−0.28
282747	$PhOCH_2CO$	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.20	−0.02	0.01	−0.02
282748	$PhOCH_2CO$	Asn	$HNCH_2Ph$	bond	$HNCH_2Ph$	−2.93	−0.17	−0.21	−0.21
282749	$PhOCH_2CO$	Val	$HNCH_2Ph$	bond	$HNCH_2Ph$	−2.14	0.72	0.63	0.74
282751	Boc	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−2.62	−0.28	−0.27	−0.52
282752	H	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.96	0.24	0.38	0.45
282753	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2$-2-benzimidazolyl	−1.11	−0.08	−0.06	0.10
282756	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(Br)Ph	Val	$HNCH_2$-2-benzimidazolyl	−1.18	−0.28	−0.41	−0.22
282779	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(Br)Ph	Val	$HNCH_2Ph$	−1.36	−0.25	−0.39	−0.39
282796	(2*S*)-$PhCH_2CH(OH)CO$	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.18	−0.09	0.10	−0.07
282797	9-(fluorenyl)CH_2OCO	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.20	0.06	0.11	0.03
282807	$PhCH_2OCO$	Val	*cis*-HN(4-hydroxy-cyclohexyl)	Val	$HNCH_2Ph$	−3.15	−0.48	−0.67	−0.78
282808	$PhCH_2OCO$	Val	*trans*-HN(4-hydroxy-cyclohexyl)	Val	$HNCH_2Ph$	−3.23	−0.42	−0.52	−0.69
282822	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	$HN(CH_2)_2$-3-indolyl	−1.03	0.55	0.35	0.40
282823	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	OCH_3	−3.04	−0.20	−0.07	−0.24
282824	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	OCH_2-2-benzimidazolyl	−1.46	−0.35	−0.06	−0.19
282825	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	5-nitro-2-(OCH_2)-benzimidazolyl	−1.48	−0.20	−0.11	−0.18

Table 1 (Continued)

SDZ	R′	A	R″	B	R‴	biol act[a]	residual[b] 1	6	9
282826	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	5,6-dichloro-2-(OCH_2)-benzimidazolyl	−1.36	−0.15	0.01	−0.09
282828	$BocNH(CH_2)_5CO$	Val	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.32	−0.03	0.17	0.03
282832	$PhCH_2OCO$	bond	$HNCH_2Ph$	bond	$HNCH_2Ph$	−3.43	0.12	−0.05	0.14
282833	$PhCH_2OCO$	*t*-Leu	$HNCH_2Ph$	bond	OC_2H_5	−2.66	0.04	0.27	0.23
282834	$PhNCH_3CO$	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−2.81	−0.67	−0.67	−0.93
282835	$CH_3(CH_2)_7OCO$	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.15	0.02	0.11	−0.08
282870	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	$HNCH_2$-2-benzimidazolyl	−0.53	0.16	0.09	0.22
282915	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	$HN(CH_2)_2$-2-morpholinyl	−1.38	−0.16	−0.19	−0.28
282916	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	$HN(CH_2)_2$-*p*-(OH)Ph	−1.26	−0.12	−0.12	−0.15
282939	*p*-(OH)Ph(CH_2)$_2$CO	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.11	0.02	0.15	0.05
282943	$PhCH_2OCO$	*t*-Leu	$HNCH_2Ph$	bond	(1*S*)-HNCHPhCH_2OH	−2.03	−0.02	−0.33	−0.39
282944	$H2N(CH_2)_5OCO$	Val	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.52	−0.04	0.17	0.08
282946	Boc	bond	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2$-2-benzimidazolyl	−1.04	0.61	0.43	0.64
282967	Boc	bond	$HNCH_2$-*p*-(OCH_3)Ph	bond	$HNCH_2$-2-benzimidazolyl	−3.36	−0.39	−0.33	−0.21
282969	H	bond	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2$-2-benzimidazolyl	−2.26	−0.11	−0.11	0.15
282978	4-f-2-benzimidazolyl	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.43	−0.12	−0.05	−0.33
282979	*c*	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−1.04	0.28	0.07	0.22
282981	2-pyridylCH_2OCO	*t*-Leu	$HNCH_2Ph$	Val	$HNCH_2Ph$	−0.92	0.27	0.26	0.20
282985	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	(1*S*)-HNCH(Ph)CH_2OH	−1.15	0.07	0.12	0.14
282987	$PhCH_2OCO$	Val	$HNCH_2Ph$	Val	*d*	−2.74	−0.31	−0.31	−0.37
283004	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2$-2-benzimidazolyl	−0.79	0.15	−0.02	0.18
283005	$PhCH_2OCO$	*t*-Leu	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2$-2-benzimidazolyl	−0.88	−0.07	−0.20	−0.09
283010	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	HN-2-benzothiazolyl	−1.02	0.04	0.14	0.09
283011	$PhCH_2OCO$	Val	$HNCH_2$-*m*,*p*-(OCH_3)$_2$Ph	Val	$HNCH_2$-2-benzimidazolyl	−1.01	−0.25	−0.32	−0.16
283012	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	$HNCH_2$-*o*,*p*-(OCH_3)$_2$Ph	−0.62	0.43	0.46	0.59
283013	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	2(HN)benzimidazolyl	−1.00	−0.03	−0.01	0.05
283043	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	*e*	−1.18	0.14	0.15	0.25
283044	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	$HN(CH_2)_2$-*ph*(SO_2NH_2)Ph	−1.15	−0.02	−0.19	−0.17
283045	$PhCH_2OCO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Gly	$HNCH_2$-2-benzimidazolyl	−1.19	−0.18	−0.25	−0.10
283046	$PhCH_2OCO$	*t*-Leu	$HNCH_2Ph$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.16	0.37	0.31	0.31
283047	$PhCH_2OCO$	*t*-Leu	$HNCH_2Ph$	bond	HN((2*S*)-hydroxyindan-(1*R*)-yl)	−3.20	0.38	0.30	0.19
283051	Boc	bond	$HNCH_2$-*p*-(OCH_3)Ph	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−2.17	−0.40	−0.40	−0.27
283052	$PhCH_2OCO$	*t*-Leu	$HNCH_2Ph$	bond	(1*S*)-HNCH(Ph)CH_2OH	−1.56	0.10	−0.05	−0.18
283053	$PhCH_2OCO$	*t*-Leu	$HNCH_2$-*p*-(OCH_3)Ph	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−0.98	0.15	0.23	0.12
283054	$PhCH_2OCO$	*t*-Leu	$HNCH_2$-*p*-(OCH_3)Ph	bond	HN((2*S*)-hydroxyindan-(1*R*)-yl)	−3.26	0.04	0.11	−0.07
283055	(2-benzimidazolyl)-$CH_2N(CH_3)CO$	Val	$HNCH_2$-*p*-(OCH_3)Ph	Val	$HNCH_2Ph$	−1.01	0.06	−0.04	−0.14

[a] Biological activity expressed as $-\log K_i$. [b] Residuals for the predicted activities based on the non-cross-validated analyses of 1, 6, and 9. [c] (2,5-Dioxoimidazolidin-4-yl)CH_2CO. [d] $O(CH_2)_2$N[2,4-dioxo-1-(2-hydroxyethyl)-5-methyl-1*H*-pyrimidin-2-yl]. [e] HN(2,3-dihydrobenzo[1,4]dioxin-6-yl).

Table 2. Structures, Measured Activities, and Residuals of the Training and Test Sets



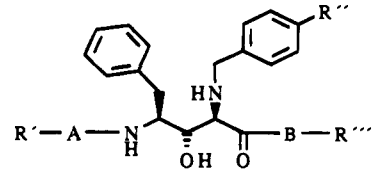
SDZ	R′	R″	*	R‴	B	R″″	biol act[a]	residual[b] 1	6	9
282730[c]	$PhCH_2OCO$-Val	$(CH_2)_2Ph$	*R*	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2$-2-benzimidazolyl	−1.23	0.10	0.15	0.08
282364[d]	$PhCH_2OCO$-Val	CH_2Ph	*S*	$HNCH_2Ph$	Val	$HNCH_2$-3-pyridinyl	−2.15	−0.92	−1.06	−1.10
282665[d]	$PhCH_2OCO$-Val	CH_2Ph	*R*	$HNCH_2$-*m*-(OCH_3)Ph	Val	$HNCH_2Ph$	−0.98	0.44	0.67	0.66
283186[d]	Boc	CH_2-*p*-(OCH_3)Ph	*R*	$HNCH_2$-*p*-(Cl)Ph	bond	OC2H5	−3.57	−0.58	−0.42	−0.50
283194[d]	Boc	CH_2-*p*-(OCH_3)Ph	*R*	$HNCH_2$-*p*-(Cl)Ph	Val	$HNCH_2$-2-benzimidazolyl	−0.95	0.63	0.64	0.43
283253[d]	$PhCH_2OCO$-Val	CH_2Ph	*R*	$HN(CH_2)_2$-*p*-(OCH_3)Ph	Val	$HNCH_2$-2-benzimidazolyl	−1.10	−0.07	0.06	−0.11
283489[d]	$PhCH_2OCO$-Val	CH_2Ph	*R*	(2*S*)-$HNCH_2CH$(OH)Ph	Val	$HNCH_2$-2-benzimidazolyl	−1.45	−0.19	−0.17	−0.32

[a] Biological activity expressed as $-\log K_i$. [b] Residuals for the predicted activities based on the non-cross-validated analyses of 1, 6, and 9. [c] Member of the training set. [d] Member of the test set.

regions where the structural and electrostatic modifications in the underlying data set are high.

The determination of the 'active conformation' of the compounds under investigation is recognized to be the crucial step in CoMFA. If the receptor is structurally available, this information can be obtained either from X-ray data directly or by docking experiments of a reference molecule. Other options include a pharmacophoric analysis of the most active compounds or the 'active analog approach'.[13,14] Although these 'inhibitor-based' methods for determination of an active conformation aim to create a high level of internal consistency, it is not guaranteed that the receptor-bound conformation will be deduced. Therefore, the 'receptor-based' approach was chosen for the current study: We made use of the published X-ray structure of the HIV-1 proteinase/MVT-101 complex[15] and replaced the inhibitor by our prototype compound SDZ-282870 for docking experiments in order to derive the 'active conformation'. Based on this template, the 3D-structures of all the other compounds were generated, and a CoMFA study was performed. To check the validity of such an

Table 3. Structures, Measured Activities, and Residuals of The Test Set



SDZ	R′	A	R″	B	R‴	biol act[a]	residual[b] 1	6	9
283134	$PhCH_2OCO$	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−0.66	−0.12	−0.10	−0.17
283143	(2*S*)-$PhCH_2CH(OH)CO$	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−0.78	−0.31	−0.23	−0.24
283206	$PhCH_2OCO$	Val	OCH_3	Val	6-(HN)benzothiazolyl	−1.20	0.40	0.47	0.41
283209	$PhCH_2OCO$	Val	OCH_3	Val	2-($HN(CH_2)_2HN$)-5-nitropyridinyl	−0.98	0.29	0.21	0.10
283239	$PhCH_2OCO$	Val	OCH_3	Val	$HN(CH_2)_2N(CH_3)_2$	−1.41	−0.18	−0.05	−0.03
283240	$PhCH_2OCO$	Val	OCH_3	Val	$HNCH_2$-4-([1,2,3]thiadiazolyl)-phenyl	−1.15	−0.17	−0.33	−0.40
283245	Boc	bond	$O(CH_2)_2$(4-morpholinyl)	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−2.11	0.05	0.27	0.10
283249	$PhCH_2OCO$	Val	OCH_3	Gly	HN(4-amino-3-pyridinyl)	−3.18	−1.09	−1.17	−1.25
283254	$PhCH_2OCO$	Val	OCH_3	Val	HN(4-hydroxy-6-methyl-2-pyrimidinyl)	−1.20	0.28	0.43	0.57
283260	(2-benzimidazolyl)-$(CH_2)_2CO$	Val	H	Val	$HNCH_2Ph$	−1.00	0.41	0.22	0.22
283261	$PhCH_2CH(R)OHCO$	*t*-Leu	OCH_3	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.15	−0.06	0.03	0.17
283262	Boc	Val	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−1.32	0.66	0.52	0.52
283263	(*p*-hydroxyphenyl)-$(CH_2)_2CO$	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−1.11	−0.67	−0.68	−0.73
283265	$PhCH_2OCO$	*t*-Leu	$O(CH_2)_2$-4-morpholinyl	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−0.98	0.45	0.26	0.22
283266	$PhCH_2OCO$	*t*-Leu	$O(CH_2)_2$-4-morpholinyl	bond	$HNCH_2Ph$	−1.84	0.09	−0.41	−0.33
283267	$PhCH_2OCO$	*t*-Leu	$O(CH_2)_2$-4-morpholinyl	bond	(1*S*)-$HNCH(Ph)CH_2OH$	−2.11	−0.32	−0.21	−0.21
283268	$PhCH_2OCO$	*t*-Leu	$O(CH_2)_2$-4-morpholinyl	bond	(1*S*)-$HNCH(Ph)CH_2OH$	−2.00	0.02	−0.22	−0.15
283321	$PhCH_2OCO$	Val	OCH_3	bond	$HN(CH_2)_2NHCH_2$-2-benzimidazolyl	−3.11	−2.13	−1.97	−2.37
283336	Boc	bond	$O(CH_2)_2OH$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.90	0.06	0.16	0.10
283337	Boc	bond	$O(CH_2)_2OH$	bond	$HNCH_2Ph$	−3.08	−0.67	−0.34	−0.46
283341	$PhCH_2OCO$	Val	OCH_3	Glu(OMe)	$HNCH_2$-2-benzimidazolyl	−1.04	−0.24	−0.06	−0.22
283342	$PhCH_2OCO$	Val	OC_2H_5	Val	$HNCH_2$-2-benzimidazolyl	−1.38	−0.60	−0.56	−0.73
283343	$PhCH_2OCO$	Val	OCH_3	bond	$HN(CH_2)_2CONHCH_2$-2-benzimidazolyl	−1.67	−0.83	−0.52	−0.69
283353	$PhCH_2OCO$	Val	OCH_3	Val	$HNCH_2$-2-cyclohexan-(1*R*)-ol	−1.20	−0.51	−0.07	−0.01
283356	Boc	bond	OC_2H_5	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.96	−0.08	0.26	0.15
283364	H	bond	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−2.36	1.57	2.13	1.86
283365	$PhCH_2OCO$	Phg	H	Val	$HNCH_2Ph$	−1.71	−0.46	−0.69	−0.68
283366	Boc	bond	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−0.89	0.54	0.60	0.39
283374	$PhCH_2OCO$	Val	OCH_3	Val	$HNCH_2$-1-benzotriazolyl	−1.18	0.14	0.19	0.00
283378	$PhCH_2OCO$	Val	OCH_3	Val	$HN(CH_2)_2$-1-benzimidazolyl	−1.28	0.05	−0.01	−0.01
283406	Boc	bond	$CH_2NHCO(CH_2)_2$-2-benzimidazolyl	bond	$HNCH_2Ph$	−2.83	−1.57	−1.79	−1.90
283407	Boc	bond	$CH_2NHCO(CH_2)_2$-2-benzimidazolyl	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.61	−0.57	−0.93	−1.10
283410	$PhCH_2OCO$	*t*-Leu	$O(CH_2)_2OH$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.20	0.06	−0.01	0.07
283412	$PhCH_2OCO$	*t*-Leu	OC_2H_5	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.00	0.19	0.31	0.36
283424	2-imidazo[4,5-*c*]pyridyl-CH_2OCO	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−1.28	−0.67	−0.88	−0.96
283426	2-imidazo[4,5-*c*]pyridyl-CH_2OCO	(D)*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−2.65	−1.34	−0.72	−0.71
283434	$PhCH_2OCO$	Val	OCH_3	bond	$HNCH_2$-2-hydroxy-4-methoxyphenyl	−3.48	−0.99	−0.91	−0.84
283436	$PhCH_2OCO$	Val	OCH_3	Val	$HN(CH_2)_2$-1-imidazolyl	−1.46	−0.07	−0.08	−0.07
283440	$PhCH_2OCO$	*t*-Leu	OC2H5	bond	$HNCH_2Ph$	−1.57	−0.03	0.12	0.22
283441	$PhCH_2OCO$	*t*-Leu	$O(CH_2)_2OH$	bond	$HNCH_2Ph$	−1.53	0.07	0.03	0.14
283451	$PhCH_2OCO$	Val	OCH_3	Val	$HNCH_2$-2-hydroxy-4-methoxyphenyl	−1.15	−0.11	−0.26	−0.39
283471	(3-iodo-4-hydroxyphenyl)-$(CH_2)_2CO$	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−0.63	−0.28	−0.27	−0.31

Table 3 (Continued)

SDZ	R′	A	R″	B	R‴	biol act[a]	residual[b] 1	6	9
283472	H	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−0.89	0.66	0.76	0.57
283478	$PhCH_2OCO$	Val	OCH_3	bond	(2*R*)-$HNCH_2CH(OH)Ph$	−3.40	−1.25	−1.36	−1.13
283479	[4-bis(benzylphosphono)-phenyl]-$(CH_2)_2CO$	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−0.72	−0.49	−0.26	−0.30
283480	(4-phosphonophenyl)-$(CH_2)_2CO$	*t*-Leu	OCH_3	Val	$HNCH_2$-2-benzimidazolyl	−0.89	−0.55	−0.41	−0.45
283481	Boc	bond	OCH_3	bond	NH_2	−3.66	−0.79	−0.50	−0.48
283490	$PhCH_2OCO$	Val	OCH_3	Glu(OMe)	$HNCH_2$-2-hydroxy-4-methoxyphenyl	−0.78	0.34	0.62	0.43
283494	$PhCH_2OCO$	Val	OCH_3	Val	$HNCH_2$-2-imidazo[4,5-*c*]-pyridyl	−0.62	0.13	0.18	0.08
283497	Boc	bond	$CH_2NHCO(CH_2)_2$-2-benzimidazolyl	bond	(1*S*)-HNCH(Ph)CH_2OH	−3.40	−2.24	−2.50	−2.64
283498	$PhCH_2OCO$	*t*-Leu	$CH_2NHCO(CH_2)_2$-2-benzimidazolyl	bond	$HNCH_2Ph$	−1.73	−1.25	−1.59	−1.54
283516	$PhCH_2OCO$	*t*-Leu	$CH_2NHCO(CH_2)_2$-2-benzimidazolyl	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−0.61	−0.38	−0.71	−0.73
283519	Boc	bond	$CH_2NHCOOCH_2Ph$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.28	−0.27	−0.61	−0.69
283520	Boc	bond	$O(CH_2)_2OH$	bond	HN-1-indanyl	−2.89	−0.53	−0.34	−0.34
283521	Boc	bond	$CH_2NHCOCH_3$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.85	0.24	0.32	0.17
283522	Boc	bond	$CH_2NHCONHPh$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.52	0.18	0.28	0.10
283532	$PhCH_2OCO$	*t*-Leu	$CH_2NHCOOCH_2Ph$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.00	−0.74	−1.29	−1.20
283537	$PhCH_2OCO$	Val	OCH_3	Val	$HN(CH_2)_2$-2-benzimidazolyl	−0.95	0.19	0.23	0.10
283549	$PhCH_2OCO$	*t*-Leu	$CH_2NHCONHPh$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−0.52	0.49	0.27	0.23
283550	$PhCH_2OCO$	*t*-Leu	$CH_2NHCOCH_3$	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−0.76	0.68	0.58	0.58
283559	H	bond	OCH_3	Phg	$HNCH_2$-2-benzimidazolyl	−1.95	1.80	2.93	2.57
283560	$PhCH_2OCO$	Val	CH_2NH_2	bond	OC_2H_5	−3.43	−1.03	−1.57	−1.31
283567	$PhCH_2OCO$	Val	CH_2NHCH_2Ph	bond	OC_2H_5	−3.28	−1.76	−1.09	−1.07
283568	Boc	bond	OCH_3	Val	$N(CH_3)CH_2$-2-benzimidazolyl	−1.08	0.25	0.32	0.20
283569	2-quinolinoyl	*t*-Leu	Br	Val	$HNCH_2$-2-benzimidazolyl	−0.75	0.25	−0.22	−0.21
283570	Boc	bond	OCH_3	Val	$HNCH_2$-2-imidazo[4,5-*c*]-pyridyl	−1.18	0.18	0.27	0.07
283573	$PhCH_2OCO$	Val	$CH_2NHCOPh$	bond	OC_2H_5	−2.97	−1.12	−0.98	−0.66
283579	(imidazo[1,2-*a*]-pyrimidin-2-yl)CO	*t*-Leu	OCH_3	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−1.36	−0.21	0.11	1.07
283580	Boc	*t*-Leu	OCH_3	bond	HN((2*R*)-hydroxyindan-(1*S*)-yl)	−2.32	−0.67	−0.56	−0.42

[a] Biological activity expressed as $-\log K_i$. [b] Residuals for the predicted activities based on the non-cross-validated analyses of 1, 6, and 9.

approach, the results obtained were then compared with the active site of the proteinase. In the present study, different CoMFA models were generated for a training set of 100 compounds. Subsequently a test set of 75 inhibitors was predicted using the best analyses. Finally, after the evaluation of the models, the results were analyzed and novel inhibitor types were designed.

Methods

The compounds belonging to the class of 2-heterosubstituted statine derivatives were synthesized and tested for enzyme inhibition according to previously described methods.[16] Enzyme inhibition is expressed as the negative logarithm of the IC_{50} (Tables 1–3).

For the docking experiments, we made use of the published X-ray structure of the HIV-1 proteinase/MVT-101 complex.[15] The inhibitor was removed and compound SDZ-282870 manually built into the active site. Subsequently, the resulting complex was partially minimized. The protein backbone and all side chains, except the residues constituting the active site (residues 8–18, 23–32, 46–56, 76–87, 108–110, 123–132, 146–156, 176–187), were excluded from the minimization as performed in two steps. First, distance constraints corresponding to a set of hydrogen bonds (Figure 1) were applied. These were removed before the second cycle. The DISCOVER force field was used and the conjugate gradient minimization technique applied;[17] convergence criteria were set to be 0.001 kcal/Å.

The compounds were built starting from the docked structure of SDZ-282870 using SYBYL 6.0 and 6.04[18] and their structures minimized using the standard TRIPOS force field without electrostatics, applying the Powell minimization technique.[19,20] The convergence criteria were set to be an energy change of <0.05 kcal/mol in two consecutive steps. Finally, the compounds were refitted to SDZ-282870 using the carbonyl functions (C and O) in P_1 and P_1' which form H-bonds to water-511.

In order to find a quick and reliable charge calculation method,[21] a comparison of semiempirical (MNDO, AM1) with empirical procedures (Gasteiger–Marsilli, Gasteiger–Hückel, Delre, Pullmann) was performed for molecule SDZ-282870.[22–27] Of the empirical methods, the Gasteiger–Marsilli approach showed the highest correlation (Table 4) with the semiempirical ones and was, therefore, selected for the calculation of partial charges.

The CoMFA region was defined to extend the Van der Waals radii of the assembly of superimposed molecules by 4 Å along the principal axes of a Cartesian coordinate system. The grid space was set to 2 Å, and as a probe atom served a C sp^3 with a formal charge of +1. The maximum field values were truncated to 30 kcal/mol for the steric and +/−30.0 kcal/mol for the electrostatic interaction energies. For the points 'inside'

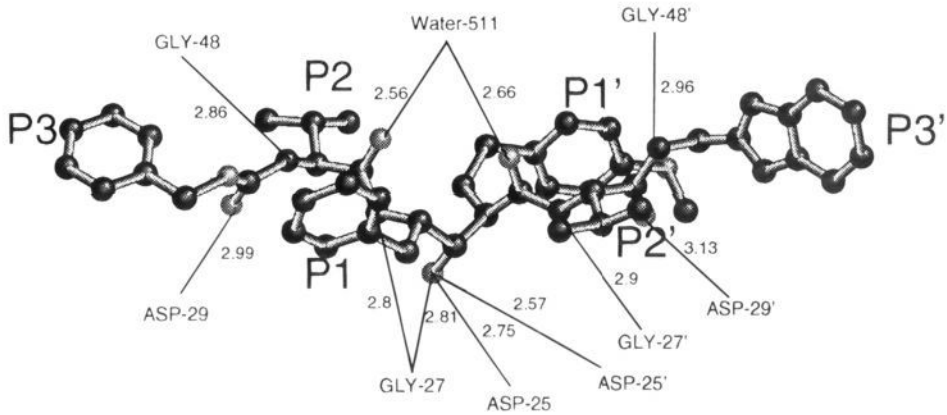


Figure 1. Results of the docking experiments with SDZ-282870. The lines indicate the H-bonds used as distance constraints; the distances shown represent the final values after minimization. Nomenclature of substituents is according to Schechter and Burger.[34]

Table 4. Correlations of the Net Atomic Charges for SDZ-282870 Using Different Calculation Methods

method	MNDO	AM1	PM3
Gasteiger–Marsili	0.865	0.686	0.373
Gasteiger–Hückel	0.827	0.647	0.392
Delre	0.649	0.569	0.247
Pullmann	0.717	0.572	0.357

the molecule (determined by a steric energy value of 30 kcal/mol), no electrostatic energy was calculated. These field values were set to the mean of the corresponding column in the subsequent PLS analyses. In addition to calculating the energy values at the grid points (the standard procedure), a second method, referred to as volume averaging (vol avg), was applied.[28] The interaction energy at a given grid point was replaced by the mean of the values at the surrounding points, the latter ones being located at the vertices of a cube, one-third of the grid spacing away from the point in its center. In order to speed up the analysis and reduce the amount of noise,[21] a column filter was used excluding the columns with a variance smaller than 1.0 or 2.0 ('minimum σ'). Equal weights for the steric and electrostatic descriptors were assigned using the CoMFA scaling option.[21]

The overall quality of the analyses was expressed by the corresponding cross-validated r^2 value (r^2_{cv}) which is defined as

$$r^2_{cv} = \frac{SD - PRESS}{SD} \quad (1)$$

where SD is the variance of the biological activities of the molecules around the mean value. "PRESS" represents the sum of the squared differences between the predicted and actual target property values for every compound. By definition, the r^2_{cv} can take up values in the range from $-\infty$ to 1.0. The ideal value of 1.0 is reached when "PRESS" becomes 0.0 (i.e., the internal prediction is perfect). Therefore, the r^2_{cv} is considered to be a very critical indicator for the internal consistency of the analyses. The calculation of the predictive r^2 value (q^2) was based on the molecules in the test set and defined in analogy to the r^2_{cv}. SD is the variance of the biological activities of the molecules in the test set around the mean activity of the training set molecules.

The number of cross-validation groups was set either equal to the number of rows included (leave-one-out method, LOO) or equal to a value of 2. In the latter case, the cross-validation groups were randomly selected. The lowest 'standard error of prediction' value determined the optimum number of components which is often lower than the number determined by the highest r^2_{cv}. This procedure was chosen since the number of components at which the initial steep increase of the r^2_{cv} starts to level off has proven to give better predictive CoMFA models.[29]

For the calculation of the Connolly surface of the active site of the proteinase, all amino acids (including hydrogens) located within 3 Å of the spheres around all atoms of the inhibitor SDZ-282870 were considered. Charges were calculated using the Gasteiger–Marsilli method. For each oxygen of the free carboxyl group of the aspartates, a formal charge of −0.5 was assigned. Electrostatic properties were mapped onto the surface using the MOLCAD software.[30]

Results and Discussion

CoMFA Models. In general, the CoMFA models derived showed a high degree of internal consistency (Table 5). For standard CoMFA settings (3: both field types, no volume averaging, minimum σ of 2), a r^2_{cv} of 0.559 could be obtained. The contributions of the two field types revealed that the steric descriptors predominate in model 3. However, since the magnitude of the contributions might be affected by scaling of the descriptor matrices or might partially arise from noise, we performed separate analyses using one field type only. These CoMFAs (1: $r^2_{cv} = 0.593$) (2: $r^2_{cv} = 0.312$) confirmed that the quality of model 3 is mainly based on steric information. Application of vol avg for the calculation of the fields changed results only slightly. The number of included columns increased for the analysis with the steric fields (4) and decreased for the one with the electrostatic field (5), leading to a higher r^2_{cv} for the electrostatics (4: $r^2_{cv} = 0.343$) and a lower one for the sterics (5: $r^2_{cv} = 0.546$). The CoMFA including both field types (6: $r^2_{cv} = 0.574$) showed a higher level of internal consistency compared to the standard method. Lowering the column filter from 2.0 to 1.0 yielded an identical r^2_{cv} for the analyses of the steric fields (7), while the r^2_{cv} for the electrostatics improved slightly (8: $r^2_{cv} = 0.366$). For the analysis including both field types, a r^2_{cv} of 0.572 was achieved (9). The best analyses (1, 6, and 9) were selected for the subsequent prediction of test compounds and repeated without cross-validation.

Although the cross-validation method should reflect the predictive power of a given analysis, the LOO method might lead to high r^2_{cv} values which do not necessarily reflect a general predictiveness of the underlying model.[12,31] Therefore, analyses with two cross-validation groups were performed; each of the respective submodels consisted of 50% of the compounds (randomly selected), and the remaining ones were predicted. As the random formation of cross-validation groups might

Table 5. Summary of the CoMFA Results

analysis	field type	vol avg	min σ	no. of col[a]	no. comp	r^2_{cv}	STDP[b]	r^2 [c]	F value	q^2 75[d]	q^2 67[e]	steric contrib[g]
1	ster	no	2	334	5	0.593	0.488	0.837	96.507	0.258	0.569	
2	elec	no	2	88	3	0.312	0.627	0.566	41.737			
3	both	no	2	422	4	0.559	0.505	0.804	97.864			0.712
4	ster	yes	2	432	5	0.546	0.515	0.816	83.578			
5	elec	yes	2	41	5	0.343	0.620	0.574	25.321			
6	both	yes	2	473	5	0.574	0.499	0.841	99.592	0.080	0.555	0.772
7	ster	yes	1	460	5	0.546	0.515	0.817	83.982			
8	elec	yes	1	179	4	0.366	0.605	0.692	53.419			
9	both	yes	1	639	6	0.572	0.503	0.870	103.746	0.094	0.552	0.749

[a] Number of columns used in the analysis after the application of column filtering (min σ). [b] Standard error of prediction. [c] Conventional r^2 determined with the optimum number of components. [d] Predictions made for 75 compounds. [e] After exclusion of eight outliers. [f] The sum of steric and electrostatic contributions equals 1.0.

Table 6. Summary of the CoMFAs with the Number of Cross-Validation Groups = 2 (cv = 2) and with Randomized Biological Activities

	cv = 2, r^2_{cv}[a]			random, r^2_{cv}[a]		
	1	6	9	1	6	9
mean[b]	0.452	0.433	0.436	−0.122	−0.129	−0.132
std dev[c]	0.094	0.108	0.104	0.078	0.091	0.098
high[d]	0.629	0.659	0.643	0.042	0.027	0.028
low[e]	0.022	−0.063	−0.034	−0.301	−0.424	−0.490

[a] Highest r^2_{cv} found within the first five components extracted. [b] Mean of all 100 runs. [c] Standard deviation. [d] Highest value found. [e] Lowest value found.

have an impact on the results, this kind of analysis was repeated 100 times for 1, 6, and 9 (Table 6, Figure 2) with an identical set of cross-validation groups, respectively. The mean r^2_{cv} for each of the 100 runs was slightly lower compared to the values obtained with the LOO method, and the standard deviation for these values was rather low. Nevertheless, in all three cases a few analyses with a rather poor r^2_{cv} could be obtained, indicating a certain degree of inconsistency in the underlying dataset. On the other hand, a few higher r^2_{cv} values were obtained, too. Noteworthy, these 'extrema' were found with identical cross-validation groups within the three sets of analyses.

As another test, analyses 1, 6, and 9 were repeated 100 times with LOO cross-validation, each time after randomly interchanging the biological activities between the compounds (Table 6). This was done in order to determine whether the quality of the models was due to chance correlation. In all cases the mean r^2_{cv} values were negative. The highest r^2_{cv} values in these runs were only slightly higher than zero, indicating that analyses 1, 6, and 9, using the correct assignment of biological activities, were not based on chance correlation.

In order to evaluate the general validity of the three selected models, a test set of 75 additional compounds was predicted (Tables 2 and 3). Interestingly, the q^2 values for the complete test set were rather low. However, removal of eight compounds from the test set yielded q^2 values of comparable magnitude to the corresponding r^2_{cv} of 1, 6, and 9. An analysis of these eight 'outliers' revealed some unique structural features which were not present in the training set: Compounds SDZ-283516, SDZ-283406, SDZ-283497, and SDZ-283498 are characterized by a rather extended substituent in position P_1' containing a benzimidazole. SDZ-283560 is the only molecule containing a primary amine in P_1'. The activities of these five compounds were predicted too high. This might be due to an unfavorable interaction with the receptor which was not explored

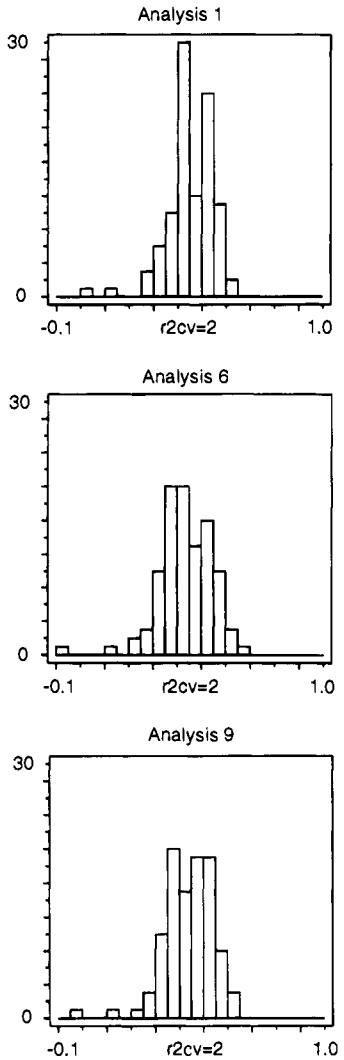


Figure 2. Distribution of the r^2_{cv} values for analyses 1, 6, and 9.

by the training set. The biological activity of SDZ-283321 was overestimated as well. It contains a benzimidazolyl moiety in P_2' similar to the compound SDZ-282967 of the training set but contains an additional substituent in P_3. SDZ-283364 and SDZ-283559, the only test compounds predicted significantly too low, do not carry substituents in P_2/P_3 and are therefore exceptionally small. In this case either different binding modes or exceptionally different entropic properties of these compounds might be encountered.

Guidelines for New Structures. The structural variability of a given dataset spans the parameter space

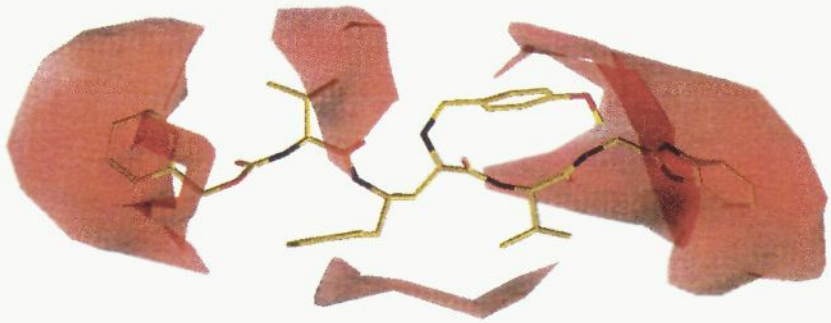


Figure 3. Variance of the steric descriptors. All areas with a variance > 10 kcal/mol are colored orange. SDZ-282870 is presented as a color-coded stick structure with undisplayed hydrogens.

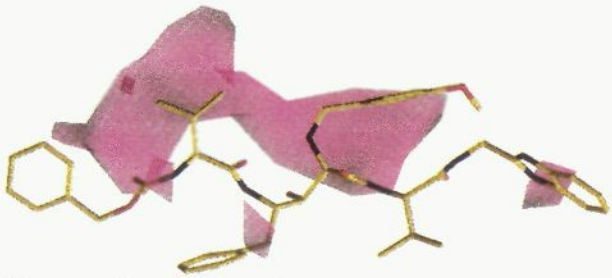


Figure 4. Variance of the electrostatic descriptors. All areas with a variance > 2.5 kcal/mol are colored magenta. SDZ-282870 is presented as a color-coded stick structure with undisplayed hydrogens.

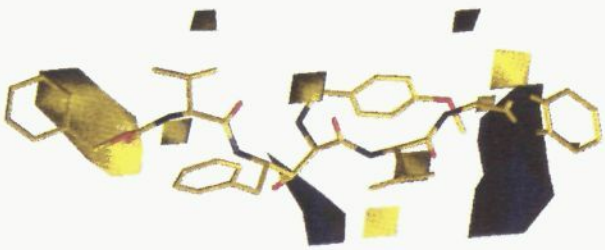


Figure 5. Electrostatic std*coeff contour map of analysis 6. Areas where negatively charged substituents are disfavored (values > 0.010) are colored blue; areas where negatively charged substituents are favored (values < −0.022) are colored yellow. SDZ-282870 is presented as a color-coded stick structure with undisplayed hydrogens.

within which prediction of newly designed compounds are reliable. Therefore the fields expressing the variance in the underlying descriptor matrix ('σ fields'; Figures 3 and 4) determine those regions for which predictions and proposals for new structures are possible. For the dataset investigated, most of the sterical variations could be found in P_3' and P_3, minor variations in P_2 and P_1', and virtually no modifications in P_2' and P_1. The variability of the electrostatic fields is mostly focused on one side of the molecule (P_1'–P_2). Minor variabilities could be detected in P_1 and P_3'.

Another field type, the 'standard-deviation-times-coefficient fields' (std*coeff fields; Figures 5–7), combines information about the regions being thoroughly investigated and the impact of putative modifications on the biological activity. For the analyses 1, 6, and 9, these fields were compared by a pointwise correlation (Table 7). The high correlation coefficient for the corresponding fields of 6 and 9 indicates that a reduction of the column filter from 2.0 to 1.0 has only a minor impact on the results. Therefore, only analysis 6 was considered for further interpretation. The steric std*coeff fields of analyses 1 and 6 were significantly different, indicating that the exclusion of the electrostatic descriptors had a high influence on the analysis of the steric descriptors.

In order to design novel compounds with high biological activity, a detailed inspection of the std*coeff fields was performed. Substituents with positive charges should be located at position P_3', whereas a negative polarization in P_3 should increase binding affinity. In the area of the transition-state mimetic, a positive charge appears advantageous. Despite the low correlation coefficient of the steric std*coeff fields for 1 and 6, the overall appearance of these fields is very similar (Figures 6 and 7). On both ends of the molecule, steric extension is favorable. Bulky substituents in P_2 should lead to compounds with increased activity. This was exemplified by the substitution of valine by *tert*-leucine in the training set.

Of special interest were the structural features at P_1': Further substitution in P_1' perpendicular to the molecule axis will lead to repulsive interactions with the receptor. On the other hand, these coefficient maps indicated that a linkage of P_1' with P_3' (the C-terminus of P_2') should be possible. If no unfavorable interaction is caused, the introduction of such a linkage should lead to increased activity due to entropic reasons. Based on these findings, a novel class of compounds, referred to as 'paracyclophanes', was designed, synthesized, and tested (Table 8).[32] These structures are characterized by an ethylene glycol ether linkage between P_1' (para or meta) and the C-terminus of the P_2' amino acid.[33] Linkers of different length in the range of 4–11 atoms were introduced. The biological activities were predicted for two of these compounds and turned out to be slightly too low (40–50 nM predicted vs 8–10 nM experimentally). An explanation for this underprediction may be that CoMFA is a method which normally utilizes calculated enthalpies and correlates them with experimentally determined properties. The latter are very often related to the free energy (ΔG) of binding, as in this particular study. The introduction of an additional linkage in the 'paracyclophanes' reduces the flexibility of these compounds compared to the members of the training set, and it consequently affects the entropic contributions to the ΔG of binding. Therefore, the slight underprediction of these novel inhibitor types seems to be related to their different entropic properties.

Comparison of the Results with the Receptor Topology. As this CoMFA was based on docking experiments of a prototype compound into the receptor, we were interested to see whether the results would be

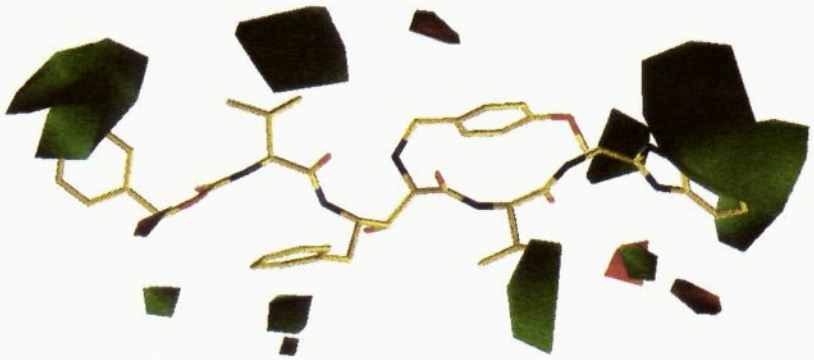


Figure 6. Steric std*coeff contour map of analysis 6. Sterically favored areas (values > 0.02) are colored green; disfavored areas (values < −0.02) are colored red. SDZ-282870 is presented as a color-coded stick structure with undisplayed hydrogens.

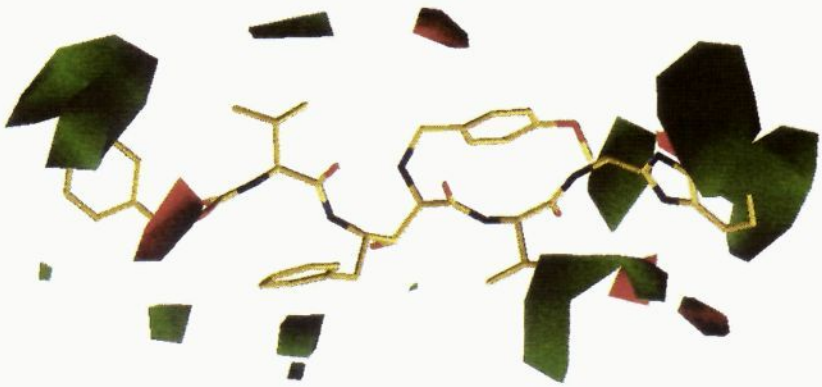


Figure 7. Steric std*coeff contour map of analysis 1. Sterically favored areas (values > 0.02) are colored green; disfavored areas (values < −0.02) are colored red. SDZ-282870 is presented as a color-coded stick structure with undisplayed hydrogens.

Table 7. Comparison of the Equations of Analyses 1, 6, and 9

	steric[a]			electrostatic[a]	
analysis	1	6	9	6	9
1	1.00			1.00	
6	0.31 (2560[b]	1.00		0.98(41)[b]	1.00
9	0.27 (303)[b]	0.99 (432)[b]	1.00		

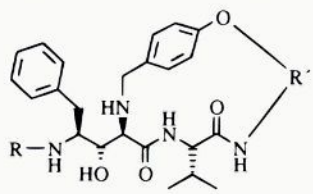
[a] Correlation coefficients for pointwise comparison of standard deviation × coefficient fields. [b] Number of points used for comparison is given in parentheses.

consistent with the steric and electrostatic properties of the active site of the enzyme. Therefore, the coefficient fields of analysis 6 were compared with the surface of the active site (Figures 8 and 9). The level of agreement was very high: Sterically forbidden regions (negative coefficients) stretched close to the receptor surface, while the 'favored' regions (positive coefficients) showed a good fit into the receptor. Only for small patches with repulsion according to the CoMFA could a direct contact surface from the receptor side not be found. A possible interpretation of this data is that the binding process involves also sliding of the inhibitor into the cavity which would contribute to additional steric restrictions. The agreement of the electrostatic fields of inhibitor and receptor was excellent; prominent regions for electrostatic interaction around the inhibitor corresponded with complementary areas on the surface of the proteinase.

Conclusions

So far, the major parameter for the quality of CoMFA studies has been the cross-validated r^2 value, being a measure for the level of internal consistency. However, a CoMFA with a high r^2_{cv} does not necessarily reflect

Table 8. Structures and Measured and Predicted Activities of Selected Paracyclophanes



			activity			
				predicted[b]		
SDZ	R	R′	measured[a]	1	6	9
284386	(Z)-Val	$(CH_2)_2$-O-$(CH_2)_2$-O-$(CH_2)_2$	−0.84	−1.78	−1.51	−1.45
284652	(Z)-Val	$(CH_2)_2$-O-$(CH_2)_2$-O-$(CH_2)_2$-O-$(CH_2)_2$	−0.88	−1.65	−1.45	−1.51

[a] Measured activity expressed as $-\log K_i$. [b] Predicted activity according to 1, 6, and 9.

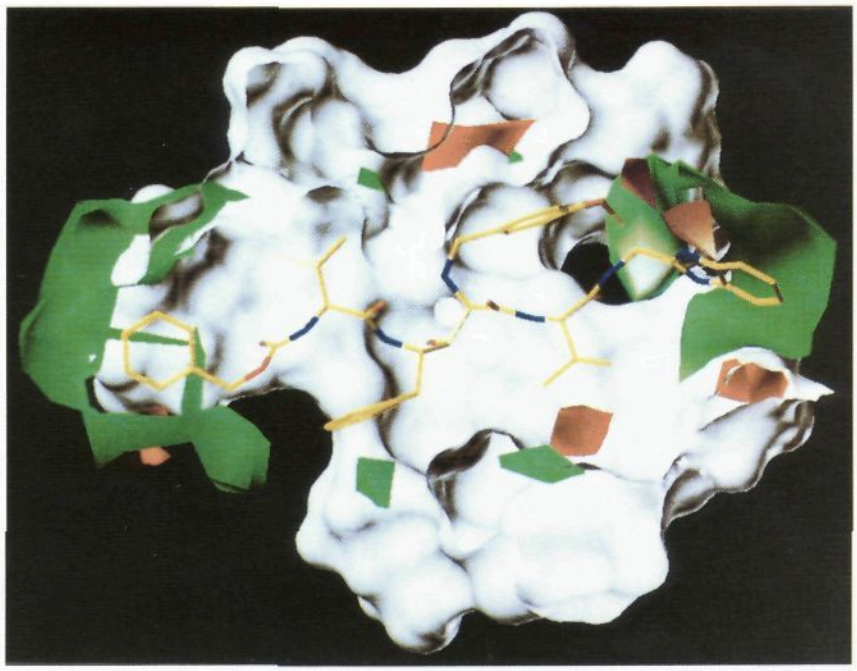


Figure 8. Connolly surface of the proteinase with the steric std*coeff field of analysis 6 (displayed as in Figure 6). SDZ-282870 is presented as a color-coded stick structure with undisplayed hydrogens.

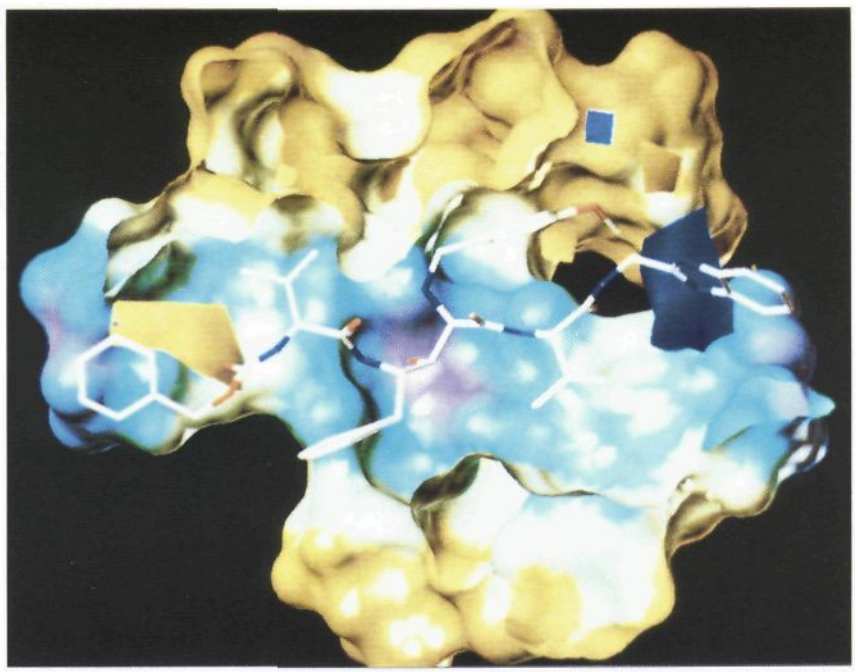


Figure 9. Connolly surface of the proteinase color-coded by the charge distribution. The electrostatic std*coeff field is displayed as in Figure 5. SDZ-282870 is presented as a color-coded stick structure with undisplayed hydrogens.

the actual physicochemical properties of the receptor. This accounts especially for datasets whose molecules are built and aligned consistently, but the 'active' conformation cannot be deduced from experimental data. In contrast, molecular modeling studies based on X-ray analysis and/or NMR experiments mainly focus on the understanding of ligand–receptor interactions and do not provide any predictive models. To date, in most of the cases, these two approaches have been applied separately. In the current study we successfully combined these strategies. Starting from an X-ray structure of the receptor, we obtained a highly predictive QSAR model. Besides other quality assessments which showed a high consistency of the models generated, we could successfully map the results back onto the starting point of this study—the receptor itself. Furthermore, we were able to use these models for successful design and prediction of a novel class of inhibitors.

With respect to the interpretation and application of a QSAR model, one has to keep in mind that QSAR methods are based on the statistical analysis of the variance in a given set, thus defining the parameter space for which the corresponding model is valid. If an attempt is made to predict the target properties of

compounds being significantly different from those in the training set, the prediction becomes uncertain and is likely to fail. This was, e.g., indicated by the error in prediction for eight compounds in the test set which showed significant structural differences to those of the training set.

Extensive tests were performed in order to verify the robustness of the models. The r^2_{cv} values obtained by the LOO method were slightly higher than the mean values resulting from 100 runs with two randomly selected cross-validation groups. This indicates that the LOO procedure overestimated the quality of the CoMFAs generated. Randomization tests revealed that the models were not based on chance correlation but showed a high predictive correlation between the calculated fields and the observed biological activities.

Acknowledgment. A. Billich and G. Winkler are gratefully acknowledged for the determination of the biological activities of the compounds described in this study. In addition, the authors would like to thank H. Gstach and P. Stuetz for critical reading of the manuscript.

References

(1) (a) Gallo, R. C.; Salahuddin, S. Z.; Popovic, M.; Shearer, G. M.; Kaplan, M.; Haynes, B. F.; Palker, T. J.; Redfield, R.; Oleske, J.; Safai, B.; White, G.; Foster, P.; Markham, P. D. Frequent Detection and Isolation of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and at Risk for AIDS. *Science* **1984**, *224*, 500–503. (b) Barre-Sinoussi, F.; Checmann, J. C.; Rey, F.; Nugeyre, M. T.; Chamoret, S.; Gruest, J.; Daugnet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science* **1983**, *220*, 868–871.

(2) (a) Seelmeier, S.; Schmidt, H.; Turk, V.; von de Helm, K. Human Immunodeficiency Virus has an Aspartic-type Protease that can be Inhibited by Pepstatin A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6612–6616. (b) Mous, J.; Heimer, E. P.; Le Grice, S. F. J. Processing Protease and Reverse Transcriptase from Human Immunodeficiency Virus Type I Polyprotein in Escherichia Coli. *J. Virol.* **1988**, *62*, 1433–1436.

(3) (a) Huff, J. R. HIV Protease: A Novel Chemotherapeutic Target for AIDS. *J. Med. Chem.* **1991**, *34*, 2305–2314. (b) Meek, T. D. Inhibitors of HIV-1 Protease. *J. Enzym. Inhib.* **1992**, *6*, 65–98. (c) Robins, T.; Plattner, J. HIV Protease Inhibitors: Their Anti-HIV Activity and Potential Role in Treatment. *J. Acquired Immune Defic. Syndr.* **1993**, *6*, 162–170.

(4) Debouck, C. The HIV-1 Protease as a Therapeutic Target for AIDS. *AIDS Res. Hum. Retroviruses* **1992**, *8*, 153–164.

(5) (a) Overton, H. A.; McMillan, D. J.; Gridley, S. J.; Brenner, J.; Redshaw, S.; Mills, J. S. Effects of Two Novel Inhibitors of the Human Immuno-deficiency Virus Protease on the Maturation of the HIV *gag* and *gag-pol* Polyproteins. *Virology* **1990**, *179*, 508–511. (b) Ashorn, P.; McQuade, T. J.; Thaisrivongs, S.; Tomasselli, A. G.; Tarpley, W. G.; Moss, B. An Inhibitor of the Protease Blocks Maturation of Human and Simian Immunodeficiency Viruses and Spread of Infection. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7472–7476. (c) Lambert, D. M.; Petteway, S. R.; McDanal, C. E.; Hart, T. K.; Leary, J. J.; Dreyer, G. B.; Meet, T. D.; Bugelski, P. J.; Bolognesi, D. P.; Metcalf, B. W.; Matthews, T. J. Human Immunodeficiency Virus Type 1 Protease Inhibitors Irreversibly Block Infectivity of Purified Virions from Chronically Infected Cells. *Antimicrob. Agents Chemother.* **1992**, *36*, 982–988. (d) Kaplan, A. H.; Zack, J. A.; Knigge, M.; Paul, D. A.; Kempf, D. J.; Norbeck, D. W.; Swanstrom, R. Partial Inhibition of the Human Immunodeficiency Virus Type 1 Protease Results in Aberrant Virus Assembly and the Formation of Noninfectious Particles. *J. Virol.* **1993**, *67*, 4050–4055.

(6) Delfraissy, J. F.; Sereni, D.; Brun-Vezinet, F.; Dussaix, E.; Krivine, A.; Dormont, J.; Bragman, K. A Phase I-II Dose Ranging Study of the Safety and Activity of Ro 31-8959 (HIV Protease Inhibitor) in Previously Zidivudine (ZDV) Treated HIV-Infected Individuals. Abstracts at the 9th Int. Conf. on AIDS, Berlin, Germany, 1993. Kitchen, V.; Skinner, C.; Sedwick, A.; Bragman, K.; Pinching, A. J.; Weber, J. A Phase I-II Dose Ranging Study of the Safety and Activity of Ro 31-8959 (HIV Protease Inhibitor) in Asymptomatic or Midley Sympthomatic HIV-Infection. *Ibid.*

(7) Wlodawer, A.; Erickson, J. W. Structure-Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543–585.

(8) Appelt, K. Crystal Structures of HIV-1 Protease-Inhibitor Complexes. *Perspect. Drug Discovery Des.* **1993**, *1*, 23–48.

(9) 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.

(10) (a) Minor, D. L.; Wyrick, S. D.; Charifson, P. S.; Watts, V. J.; Nichols, D. E.; Mailman, R. B. Synthesis and Molecular Modeling of 1-Phenyl-1,2,3,4-tetrahydroisoquinolines and Related 5,6,8,9-Tetrahydro-13bH-dibenzo[a,h]quinolizines as D1 Dopamine Antagonists. *J. Med. Chem.* **1994**, *37*, 4317–4328. (b) Myers, A. M.; Charifson, P. S.; Owens, C. E.; Kula, N. S.; McPhail, A. T.; Baldessarini, R. J.; Booth, R. G.; Wyrick, S. D. Conformational Analysis, Pharmacophore Identification, and Comparative Molecular Field Analysis of Ligands for the Neuromodulatory Sigma 3 Receptor. *J. Med. Chem.* **1994**, *37*, 4109–4117. (c) Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-determined Alignment Rules. *J. Med. Chem.* **1993**, *36*, 4152–4160.

(11) (a) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindberg, W.; Sjöström, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984; pp 17–95. (b) Dunn, W. J., III; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. Multivariate Structure-Activity Relationship Between Data from a Battery of Biological Tests and an Ensemble of Structure Descriptors: The PLS Method. *Quant. Struct.-Act. Relat.* **1984**, *3*, 31–137. (c) Geladi, P. Notes on the History and Nature of Partial Least Squares (PLS) Modelling. *J. Chemom.* **1988**, *2*, 231–246.

(12) (a) Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Component Models. *Technometrics* **1978**, *4*, 397–405. (b) Diaconis, P.; Efron, B. Computer-Intensive Methods for Statistics. *Sci. Am.* **1984**, *116*, 96–117. (c) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. Crossvalidation, Bootstrapping and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.

(13) (a) Martin, Y. C.; Bures, M. G.; Dahner, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A Fast Approach to Pharmacophore Mapping and its Application To Dopaminergic and Benzodiazepine Agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83–102. (b) Brint, A. T.; Willett, P. Algorithms for the Identification of Three-dimensional Maximal Common Substrates. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 152–158.

(14) (a) Motoc, I.; Dammkoehler, R. A.; Mayer, D.; Labanowski, J. Three-Dimensional Structure-Activity Relationships. I. General Approach to the Pharmacophore Model Validation. *Quant. Struct.-Act. Relat.* **1986**, *5*, 99. (b) Hibert, M. F.; Gittos, M. W.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. Graphics Computer-Aided Receptor Mapping as a Predictive Tool for Drug Design: Development of Potent, Selective, and Stereospecific Ligands for the 5-HT_{1A} Receptor. *J. Med. Chem.* **1988**, *31*, 1087–1093. (c) Holtje, H. D.; Mauerhofer, E. Conformational Analysis on Calcium Channel Active Diphenylalkylamines, Diphenylbutylpiperidines, Phenylalkylamines, and Perhexiline. *Quant. Struct.-Act. Relat.* **1989**, *8*, 259–265.

(15) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, S.; Kent, S. B. H.; Wlodawer, A. Structure of the Complex of Synthetic HIV-1 Protease with a Substrate-Based Inhibitor at 2.3 Angstroems Resolution. *Science* **1989**, *246*, 1149–1152.

(16) Scholz, D.; Billich, A.; Charpiot, B.; Ettmayer, P.; Lehr, P.; Rosenwirth, B.; Schreiner, E.; Gstach, H. Inhibitors of the HIV-1 Proteinase Containing a 2-Heterosubstituted 4-Amino-3-hydroxy-5-phenylpentanoic Acid: Synthesis, Enzyme Inhibition, and Antiviral Activity. *J. Med. Chem.* **1994**, *37*, 3079–3089.

(17) The program DISCOVER is available from Biosym Technologies, 9685 Scranton Rd., San Diego, CA 92121.

(18) The program SYBYL 6.0/6.04 is available from Tripos Assoc., 1699 S. Hanley Rd., St. Louis, MO 63144.

(19) Clark, M.; Cramer, R. D., III; Van Opdenbosh, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.

(20) Powell, M. J. D. Restart Procedure for the Conjugate Gradient Method. *Math. Program* **1977**, *12*, 214.

(21) (a) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A.; Rognan, D. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 711–717. (b) Folkers, G.; Merz, A.; Rognan, D. *Ibid.* pp 583–618. (c) Cramer, R. D., III; DePriest, S. A.; Patterson, D. E.; Hecht, P. *Ibid.* pp 443–485.

(22) The program MOPAC is available from Quantum Chemical Program Exchange No. 455.

(23) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Chemical Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(24) Gasteiger, J.; Marsilli, M. Iterative Partial Equalization of Orbital Electronegativity - A rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.
(25) *SYBYL Molecular Modelling Software Version 6.1., Theory Manual, 1994*; Tripos Assoc.: St. Louis, MO, 1994; pp 66–68.
(26) Streitwieser, A. *Molecular Orbital Theory for Organic Chemists*; Wiley: New York, 1961.
(27) Berthod, H.; Giessner-Prettre, C.; Pullman, A. Role of .sigma and .pi. Electrons on the Properties of Halogens-substituted Conjugated Molecules. Application to the Study of Uracil and Fluorouracil. *Theoret. Chim. Acta* **1967**, *8*, 212–222.
(28) *SYBYL Molecular Modeling Software Version 6.1, Command Manual, 1994*; Tripos Assoc.: St. Louis, MO, 1994; p 391.
(29) *SYBYL Molecular Modelling Software Version 6.1, Theory Manual, 1994*; Tripos Assoc.: St. Louis, MO, 1994; p 368.
(30) Hieden, W.; Moeckel, G.; Brickmann, J. A New Approach to the Display of Local Lipophilicity/Hydrophilicity Mapped on Molecular Surfaces. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 163–172. The program MOLCAD is available from Tripos Assoc., 1699 S. Hanley Rd., St. Louis, MO 63144.
(31) Clementi, S. Personal communication.
(32) Ettmayer, P.; Hübner, M.; Billich, A.; Hecht, P.; Rosenwirth, R.; Gstach, H. Paracyclophanes as Water-soluble Inhibitors of HIV-1 Protease: Synthesis, Enzyme Inhibition, and Antiviral Activity. Manuscript in preparation.
(33) For P_1–P_3 cyclizations, see: (a) Smith, R. A.; Coles, P. J.; Chen, J. J.; Robinson, V. J.; MacDonald, D.; Carriere, J.; Krantz, A. Design, Synthesis, and Activity of Conformationally Constrained Macrocyclic Peptide-based Inhibitors of the HIV Protease. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2217. (b) Podlogar, B. L.; Farr, R. A.; Friedrich, D.; Tarnus, C.; Huber, E. W.; Cregge, R. J.; Schirlin, D. Design, Synthesis, and Conformational Analysis of a Novel Macrocyclic HIV-Protease Inhibitor. *J. Med. Chem.* **1994**, *37*, 3684–3692.
(34) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.

JM950347Q