Three-Dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-Determined Alignment Rules

Chris L. Waller, Tudor I. Oprea,† Alessandro Giolitti,‡ and Garland R. Marshall*

Center for Molecular Design, Washington University, St. Louis, Missouri 63130

Received August 20, 1993®

Comparative molecular field analysis (CoMFA), a three-dimensional, quantitative structure–activity relationship (QSAR) paradigm, was used to examine the correlations between the calculated physicochemical properties and the *in vitro* activities of a series of human immunodeficiency virus (HIV-1) protease inhibitors. The training set consisted of 59 molecules from five structurally-diverse transition-state isostere classes: hydroxyethylamine, statine, norstatine, keto amide, and dihydroxyethylene. The availability of X-ray crystallographic data for at least one representative from each class bound to the protease provided information regarding not only the active conformation of each ligand but also, via superimposition of protease backbones, the relative positions of each ligand with respect to one another in the active site of the enzyme. Once aligned, these molecules served as templates on which additional congeners were field-fit minimized. Additional alignment rules were derived from minimizations of the ligands in the active site of the semirigid protease. The predictive ability of each resultant model was evaluated using a test set comprised of molecules containing a novel transition-state isostere: hydroxyethylurea. Crystallographic studies (Getman, D. P.; *et al. J. Med. Chem.* **1993**, *36*, 288–291) indicated an unexpected binding mode for this series of compounds which precluded the use of the field-fit minimization alignment technique. The test set molecules were, therefore, subjected to a limited systematic search in conjunction with active-site minimization. The conformer of each molecule expressing the lowest interaction energy with the active site was included in the test set. Field-fit minimization of neutral molecules to crystal ligands and active-site minimizations of protonated ligands yielded predictive correlations for HIV-1 protease inhibitors. The use of crystallographic data in the determination of alignment rules and field-fit minimization as a molecular alignment tool in the absence of direct experimental data regarding binding modes is strongly supported by these results.

Introduction

Human immunodeficiency virus (HIV) is the etiologic agent of acquired immunodeficiency syndrome (AIDS) which expresses its effects through the genetic direction of viral polyproteins prepared by the host. HIV-1 protease is the aspartic proteinase encoded by the virus responsible for the processing of the *gag* and *gag-pol* gene polyproteins. Therapeutic intervention at this proteolytic step of viral replication has been demonstrated to yield immature and noninfectious viral progeny.[1,2] Using the knowledge gained from earlier studies on renin and other aspartic proteinases,[3,4] many peptide-based inhibitors have been developed which mimic the tetrahedral intermediate formed upon hydration of the scissile amide bonds of the substrate. Several transition-state analogs incorporating noncleavable reduced amide,[5] ketomethylene,[6] hydroxyethylamine,[7-10] statine,[11,12] norstatine,[13] or dihydroxyethylene[14] linkages as the isosteric bond have been reported (Figure 1). Many of these analogs express subnanomolar activity *in vitro* and, therefore, are potentially useful in the clinical management of HIV infection.

Conventional structure–activity data indicate a preference for the *S* (or its equivalent)-hydroxyl diastereomer in the isosteric linkage of the inhibitor positioned between the Asp-25/125 carboxylate groups of the protease. Crystallographic data reveal that the residues of the S1 and S1′ pockets [Berger and Schecter nomenclature] of the inhibitor are mostly hydrophobic; therefore, the most potent inhibitors have hydrophobic residues occupying the P1 and P1′ positions. The S2 and S2′ protease subsites are also hydrophobic in nature, but a variety of functionalities are present in the P2 and P2′ positions on ligands since both polar and nonpolar side chains are equally accommodated. The preference of residue is even less well-defined at the S3 and S3′ receptor sites. Furthermore, it is not clear whether P3 and P3′ are necessary requirements since certain inhibitors, including L-689,502,[15] lack residues in both of these areas.

Although structure–activity studies have provided a plethora of information regarding the structural requirements of ligands for the HIV protease, a major deficiency in design approaches has been the difficulty of predicting affinities of designed compounds for the target enzyme. It has been demonstrated that molecular dynamics simulations with free energy perturbations (FEP) are capable of providing information detailing the stabilizing structural features of inhibitor/protease complexes and also make possible calculations of the relative binding free energies of inhibitors to the protease;[16,17] however, the computational requirement is excessive (ca. 100 Cray YMP hours for each novel inhibitor analog). Additionally, in order to adequately sample the relevant configurations of the system and to ensure thermodynamic convergence, the

* To whom correspondence should be addressed.
† Permanent address: Department of Physiology, University of Medicine and Pharmacy, P-ta E. Murgu nr.2, R-1900 Timisoara, Romania.
‡ Present address: A. Menarini, Industrie Farmaceutiche Riunite s.r.l., Firenze, Italy.
® Abstract published in *Advance ACS Abstracts*, November 15, 1993.

0022-2623/93/1836-4152$04.00/0

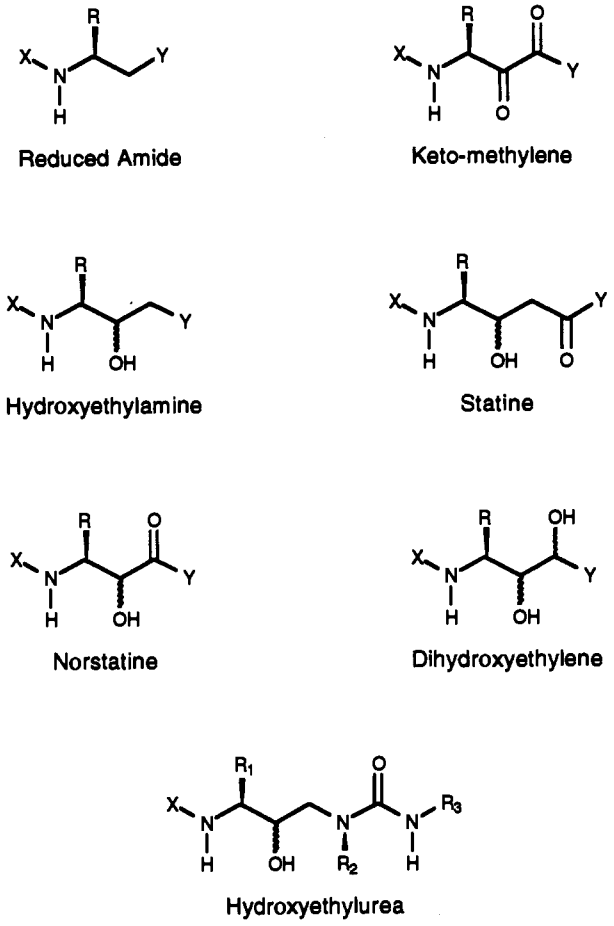


Figure 1. Examples of transition-state isosteres found in HIV-1 protease inhibitors.

novel compound cannot differ greatly from the parent from which it is to be mutated. These and other well-documented[18] limitations preclude the use of FEP as a practical technique for the routine prediction of the affinity of novel inhibitors.

DePriest *et al.*[19] recently reported success in using a 3-D QSAR method, comparative molecular field analysis (CoMFA),[20] to predict the affinities for inhibitors of thermolysin based on the crystal structure of complexes. The success of a CoMFA study is completely determined by the quality of the "alignment rule", or the choice of superimposition of the molecules in the study. Many successful 3-D QSAR's have been generated using the CoMFA technique for conformationally-constrained systems[21-24] and conformationally-flexible systems[25,26] employing a wide variety of alignment procedures. The availability of several structurally-diverse inhibitor structures bound in the active site of the HIV protease provided direct experimental information detailing the molecular alignment of the conformationally-flexible molecules used in the present study. Using this empirical data in conjunction with field-fit minimization[27] techniques (minimization which maximizes the overlap of steric and electrostatic fields between that of the molecule and a standard field, perhaps, that of the molecule with the highest affinity), several alignment rules were tested to develop a statistically-significant QSAR model for HIV-1 protease.

To test the utility of the model as a predictive tool, a test set of inhibitors with a unique transition-state isostere (hydroxyethylurea)[28] previously unseen by the model was used. Although direct crystallographic information was not available, it was reported that the P2′ and P1′ subunits were reversed for, at least, one of the inhibitors in this series. A combination of field-fit minimization and minimization within the active site of the protease was employed in the generation of the alignment rule for the test set molecules. On the basis of energetic evidence, multiple binding modes for the hydroxyethylurea class of inhibitors are probable.

Methods

A. CoMFA: Alignment Rules. All conventional and field-fit minimizations were performed in Sybyl[29] using the standard Tripos force field[30] with an energy change convergence criterion of 0.001 kcal/mol. Charges were used as calculated using MOPAC 5.0[31] and the AM1 model Hamiltonian[32] (keywords: 1SCF, MMOK). The minimizations performed in the active site used an energy-change criterion of 1.0 kcal/mol. All backbone atoms and waters of the active site were considered as an aggregate and were not minimized. Active-site side chains and all atoms of the ligand were considered interesting and were fully minimized. For the active site, partial atomic charges were loaded from the dictionary (Kollman all-atom method[33]). The partial charges of all essential water molecules and the nonstandard residues (protonated Asp-125) in the active-site sphere were determined using MOPAC 5.0 and the AM1 model Hamiltonian.

The 59 molecules in the training set represent five structurally-diverse classes (hydroxyethylamine,[7-10] statine,[11,12] norstatine,[13] keto amide,[13] and dihydroxyethylene[14]) of transition-state isosteric protease inhibitors (Tables I–V). Seven crystal structures (Roche,[7] JG365,[34] U-75875,[35] Ag1001,[12] Ag1002,[12] Ag1004,[11,12] and L-689,-502[15]) of enzyme/inhibitor complexes were superimposed via root mean square (RMS) fit of the backbone atoms of the enzyme. Once aligned, the inhibitor molecules were simply extracted from the complex. Additional ligands were then aligned by field-fit minimization using the steric and electrostatic fields of the most closely related crystal structure as template fields (Tables I–V). Subsequent reminimization without the field-fit option allowed the fitted molecules to relax to the nearest local minimum energy structure. This alignment will be referred to as alignment I.

It must be noted that the initial alignment (alignment I) was determined using neutral ligands (no formal ionic charges). Assay conditions indicate a pH between 5.5 and 6.4.[36-40] It is highly likely that the histidine and non-amide nitrogens of the proline, piperidine, and decahydroisoquinoline ring systems of the various inhibitors are protonated under these conditions. However, changes involving deprotonation of the bound ligands could not be excluded, since the pH in the binding site is not necessarily that of the solution. All protonated ligands were submitted to open shell AM1 1SCF calculations (with a charge of +1 on the system). To evaluate the sensitivity of CoMFA with respect to substantial changes in the electronic nature of ligands in the training set, studies were performed on the protonated species without altering the initial alignment. This alignment will be referred to as alignment II. Using the neutral alignment (alignment I) as a starting point, the charged molecules were also subjected to field-fit minimization and subsequent reminimization. This alignment will be referred to as alignment III.

Table I. Compounds Included in Training Set (Alignment Rule: Field-Fit Minimization to Crystal Roche (5) Compound or Active-Site Minimization)

compd	chiral	structure	p(IC_{50}) (μM)	ref
1	*R*	Z.PheΨ[CH(OH)CH_2N]Pro.O^{t}Bu	−0.813	9
2	*R*	Z.Asn.PheΨ[CH(OH)CH_2N]Pro.O^{t}Bu	0.854	9
3	*R*	Z.Asn.PheΨ[CH(OH)CH_2N]Pro.NH^{t}Bu	0.678	9
4	*R*	QC.Asn.PheΨ[CH(OH)CH_2N]Pro.O^{t}Bu	1.638	9
5[a]	*R*	Z.Asn.PheΨ[CH(OH)CH_2N]PIC.NH^{t}Bu	1.745	9
6	*R*	Z.CNA.PheΨ[CH(OH)CH_2N]PIC.NH^{t}Bu	1.638	9
7	*R*	QC.Asn.PheΨ[CH(OH)CH_2N]PIC.NH^{t}Bu	2.699	9
8	*R*	QC.SMC.PheΨ[CH(OH)CH_2N]PIC.NH^{t}Bu	1.921	9
9	*R*	QC.Asn.PheΨ[CH(OH)CH_2N]DIQ.NH^{t}Bu	3.398	9
10	*R*	Z.Asn.PheΨ[CH(OH)CH_2N]DIQ.NH^{t}Bu	2.569	9
11	*R*	Z.Asn.PheΨ[CH(OH)CH_2N]Pro.Ile.Val.Ome	0.500	7
12	*R*	Ac.Ser.Leu.Asn.PheΨ[CH(OH)CH_2N]Pro.Leu.Val.OMe	1.187	8
13	*R*	Boc.Asn.PheΨ[CH(OH)CH_2N]Pro.Leu.Val.Ome	0.071	8
14	*R*	Ac.Ser.Leu.Asn.PheΨ[CH(OH)CH_2N]Pro.O^{t}Bu	1.854	8
15	*S*	Z.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	0.337	13
16		Z.PheΨ[C(O)C(O)N]Pro.NH^{t}Bu	0.222	13
17	*S*	Z.Asn.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	2.131	13
18		Z.Asn.PheΨ[C(O)C(O)N]Pro.NH^{t}Bu	1.699	13
19	*S*	Z.Val.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	2.367	13
20	*S*	Z.Asn.PheΨ[CH(OH)C(O)N]PIC.NH^{t}Bu	1.585	13
21	*S*	Z.Asn.PheΨ[CH(OH)C(O)N]DIQ.NH^{t}Bu	1.076	13
22	*S*	NoA.Asn.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	3.237	13
23	*S*	NoA.Val.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	2.721	13
24	*S*	2-NoA.Asn.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	2.886	13
25	*S*	QC.Asn.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	2.959	13
26	*S*	QC.Asn.PheΨ[CH(OH)C(O)N]DIQ.NH^{t}Bu	1.569	13

[a] Crystal (template) structure.

Table II. Compounds Included in Training Set (Alignment Rule: Field-Fit Minimization to Modified Crystal JG365 Compound (28) or Active-Site Minimization)

compd	chiral	structure	p(IC_{50}) (μM)	ref
27	*S*	QC.Asn.PheΨ[CH(OH)CH_2N]PIC.NH^{t}Bu	0.328	9
28	*S*	Ac.Ser.Leu.Asn.PheΨ[CH(OH)CH_2N]Pro.Leu.Val.OMe	2.469	8
29	*S*	Boc.Asn.PheΨ[CH(OH)CH_2N]Pro.Leu.Val.OMe	1.796	8
30	*S*	Z.Asn.PheΨ[CH(OH)CH_2N]Pro.O^{t}Bu	0.523	8
31	*S*	Ac.Ser.Leu.Asn.PheΨ[CH(OH)CH_2N]Pro.O^{t}Bu	1.854	8
32	*S*	Z.Asn.PheΨ[CH(OH)CH_2N]Pro.Ile.Phe.OMe	2.398	8
33	*S*	QC.Asn.PheΨ[CH(OH)CH_2N]Pro.Ile.Phe.OMe	2.638	8
34	*S*	Z.Asn.PheΨ[CH(OH)CH_2N]DIQ.NH^{t}Bu	0.500	7
35	*S*	Z.Asn.PheΨ[CH(OH)CH_2N]Pro.Ile.Val.OMe	1.886	7
36	*S*	Z.Asn.PheΨ[CH(OH)CH_2N]DIQ.Ile.Val.OMe	0.500	7
37	*R*	Z.Asn.PheΨ[CH(OH)C(O)N]Pro.NH^{t}Bu	−0.491	13
38	*R*	Z.Asn.PheΨ[CH(OH)C(O)N]Pro.Ile.Val.OMe	1.602	13

[a] Crystal (template) structure.

Table III. Compounds Included in Training Set (Aligment Rule: Field-Fit Minimization to Crystal U-75875 (4) Compound or Active-Site Minimization)

compd	structure [*R*,*R*]	p(IC_{50}) (μM)[b]	ref
39	NoA.His.LeuΨ[CH(OH)CH(OH)]Val.Ile.Amp	1.950	14
40[a]	NoA.His.CalΨ[CH(OH)CH(OH)]Val.Ile.Amp	2.745	14
41	NoA.His.CalΨ[CH(OH)CH(OH)]Leu.Ile.Amp	1.950	14
42	NoA.His.LeuΨ[CH(OH)CH(OH)]Leu.Ile.Amp	1.803	14
43	NoA.His.PheΨ[CH(OH)CH(OH)]Phe.Ile.Amp	1.569	14
44	NoA.His.CalΨ[CH(OH)CH(OH)]Cal.Ile.Amp	1.233	14
45	Boc.Phe.His.LeuΨ[CH(OH)CH(OH)]Val.Ile.Amp	1.201	14
46	Boc.Phe.His.LeuΨ[CH(OH)C[*S*]H(OH)]Val.Ile.Amp	−0.051	14
47	NoA.His.LeuΨ[CH(OH)CH(OH)]D-Leu.Ile.Amp	0.305	14
48	Boc.Phe.His.LeuΨ[CH(OH)CH(OH)]Val.Mba	0.092	14
49	PoA.HisΨ[CH(OH)CH(OH)]Val.Ile.Amp	1.950	14
50	PoA.HisΨ[CH(OH)CH(OH)]Val.Mba	0.092	14
51	PoA.His.CalΨ[CH(OH)CH(OH)]Val.Ile.Amp	2.347	14
52	Hac.His.CalΨ[CH(OH)CH(OH)]Val.Ile.Amp	2.046	14
53	Gly.His.CalΨ[CH(OH)CH(OH)]Val.Ile.Amp	1.606	14
54	Cha.CalΨ[CH(OH)CH(OH)]Val.Ile.Amp	1.305	14

[a] Crystal (template) structure. [b] See Appendix.

The availability of crystallographic data for the protease offered another possibility for alignment rule—minimization within the active site. To reduce the computational complexity of this task, a substructure sphere 24 Å in diameter centered on the ligand of the Roche inhibitor/protease complex was extracted and defined as the active site. All molecules of both field-fit aligned training sets (alignment I and alignment III) were then sequentially merged into the active site, minimized as described above, extracted from the active site, and reminimized to the nearest local minimum energy conformation. Active-site-minimized neutral molecules will be referred to as align-

Table IV. Compounds Included in Training Set (Alignment Rule: Crystal Structures or Active-Site Minimization)

compd	structure	p(IC_{50}) (μM)	ref
55[a]	Sta.Ala.Sta.Val.Val.Iva	−1.079	12
56[a]	Ile.Gln.Sta.Asn.Phe.Ser	−1.079	12
57[a]	Gln.Val.Ile.Sta.Asn.Gln.Ser	−0.903	11, 12

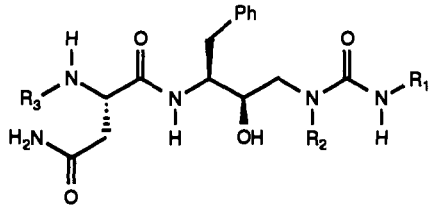
[a] Crystal structure.

Table V. Compounds Included in Training Set (Alignment Rule: Field-Fit Minimization to L-689,502 Crystal Structure or Active-Site Minimization)

compd	structure[a]	p(IC_{50}) (μM)	ref
58	Boc.PheΨ[CH(OH)CH_2]Phe.Ahi	0.599	10
59	Boc.Asn.PheΨ[CH(OH)CH_2]Phe.Ahi	0.530	10

[a] Ahi = 1(*S*)-amino-2(*R*)-hydroxyindan.

Table VI. Compounds Included in Test Set (Alignment Rule: Active-Site Minimization)



compd	R_1	R_2	R_3	p(IC_{50}) (μM)	ref
m3	$CH_2CH(CH_3)_2$	CH_3	Cbz	−0.176	28
m4a	$CH_2CH(CH_3)_2$	$(CH_2)_3CH_3$	Cbz	0.026	28
m4b	$CH_2CH(CH_3)_2$	$(CH_2)_3CH_3$	Qua	0.899	28
m5	$CH_2CH(CH_3)_2$	$(CH_2)_3CH_3$	Cbz	0.285	28
m6	$CH_2CH(CH_3)_2$	CH_2CH_3	Cbz	0.481	28
m7	$CH_2CH(CH_3)_2$	$CH(CH_3)_2$	Cbz	0.585	28
m8a	$CH_2CH(CH_3)_2$	$C(CH_3)_3$	Cbz	1.456	28
m8b	$CH_2CH(CH_3)_2$	$C(CH_3)_3$	Qua	2.221	28
m9a	$CH_2CH_2CH(CH_3)_2$	$C(CH_3)_3$	Cbz	1.886	28
m9b	$CH_2CH_2CH(CH_3)_2$	$C(CH_3)_3$	Qua	2.523	28
m10a	$CH_2C_6H_{11}$	$C(CH_3)_3$	Cbz	1.537	28
m10b	$CH_2C_6H_{11}$	$C(CH_3)_3$	Qua	2.301	28
m11a	CH_2Ph	$C(CH_3)_3$	Cbz	1.721	28
m11b	CH_2Ph	$C(CH_3)_3$	Qua	2.523	28
m12	*R*-$CH(CH_3)Ph$	$C(CH_3)_3$	Cbz	−0.813	28
m13	*S*-$CH(CH_3)Ph$	$C(CH_3)_3$	Cbz	−0.707	28
m14a	CH_2(4-pyridyl)	$C(CH_3)_3$	Cbz	0.978	28
m14b	CH_2(4-pyridyl)	$C(CH_3)_3$	Qua	1.721	28

ment IV and active-site-minimized charged molecules, alignment V.

To evaluate the predictive ability of the resulting models, a test set of 18 additional inhibitors[28] (Table VI) was utilized. These molecules represent another class of transition-state isostere: hydroxyethylurea. Due to our lack of crystallographic structures for this particular class of compounds, the available information concerning the bound conformation of one representative of this novel class of transition-state inhibitors was used to generate an alternate alignment. The test molecules were built and field-fit minimized to the most closely related crystal structure (Roche). Using this conformation as the initial guess, all 18 molecules of the test set were then subjected to a limited systematic search routine which rotated the residues in the P1′ and P2′ positions. All resultant conformers were active-site minimized, and the "active conformation" of each molecule to be included in the test set was then selected on a purely enthalpic basis as that yielding the lowest total energy for the enzyme/inhibitor complex. The selection criterion was based on the assumption that the difference in entropy for different conformations of a bound inhibitor between the complexed and uncomplexed inhibitor would be approximately constant. Therefore, the dominant component of free energy of binding could be assumed to be the enthalpy of complex formation, thus justifying the selection of the enzyme/inhibitor complex with the lowest enthalpy for inclusion in the test set. For the test set, only neutral molecules were considered (i.e. pyridyl substituents were not protonated since the pK_a of pyridine is less than 5.5).

Table VII. Summary of CoMFA Results

	alignment rule				
	I	II	III	IV	V
r^2_{cross}	0.778(6)	0.653(8)	0.607(8)	0.659(7)	0.642(7)
sep	0.552	0.704	0.749	0.684	0.707
r^2	0.984(6)	0.990(8)	0.991(8)	0.988(7)	0.983(7)
s	0.146	0.122	0.112	0.129	0.156
F test	549.838	597.130	703.933	592.234	413.512
p value	0.000	0.000	0.000	0.000	0.000
contributions					
steric	0.36	0.48	0.50	0.32	0.47
electro	0.64	0.52	0.50	0.68	0.53
r^2_{pred}	0.662	0.625	0.660	0.327	0.512

B. CoMFA: Interaction Energies and Regression Techniques. All CoMFA studies were performed on Silicon Graphics Iris Indigo R4000 computers running SYBYL 6.0.[29] The steric and electrostatic field energies were calculated using an sp^3 carbon probe atom with a charge of −1 and a distance-dependent dielectric constant at all intersections of a regularly-spaced (2 Å) grid of dimensions 36 × 26 × 22 Å. Steric and electrostatic contributions were truncated to a value of ±30 kcal/mol, and the electrostatic contributions at lattice intersections yielding maximal (±30 kcal/mol) steric values were ignored.

All regression analyses were done using the Partial Least Squares (PLS)[41] algorithms in SYBYL.[29] Initial analyses were performed using full cross-validation[42] (leave-one-out method) and 10 principal components (PCs). The optimal number of components to be used in the non-cross-validated (conventional) analyses was defined as that which yielded the highest cross-validated r^2 value. For component models with identical values, the component number producing the smallest standard error of prediction (SEP) was selected. To minimize the influence of noisy columns, all cross-validated analyses were performed with a minimum σ (column filter) value of 2.00 kcal/mol.

C. "Predictive" r^2 Values. The "predictive" r^2 was based only on molecules not included in the training set and is defined as: predictive r^2 = (SD − "press")/SD where SD is the sum of the squared deviations between the affinities of molecules in the test set and the mean affinity of the training set molecules and "press" is the sum of the squared deviations between predicted and actual affinity values for every molecule in the test set. It should be obvious from the equation that prediction of the mean value of the training set for each member of the test set would yield a predictive r^2 = 0. This is analogous to the cross-validated r^2 definition[42] and can result in a negative value reflecting a complete lack of predictive ability of the training set for the molecules included in the test set.

Results

A. CoMFA of HIV-1 Protease Inhibitors. The results of the CoMFA studies are presented in summarized form in Tables VII and VIII. The initial analysis based on alignment I yielded a correlation with a cross-validated r^2 of 0.778 (F = 549.838) using six principal components

Table VIII. Differences between Predicted and Actual Activities for Test Set Molecules

	alignment rule					
compd	I	II	III	IV	V	actual
m3	0.302	0.565	0.482	0.718	0.571	−0.176
m4a	0.067	0.633	0.703	0.042	0.170	0.026
m4b	0.846	0.547	−0.177	−0.373	−0.094	0.899
m5	−0.324	−0.088	0.082	−0.336	−0.408	0.285
m6	−0.256	0.291	0.281	−0.353	0.079	0.481
m7	−0.142	0.194	0.190	−0.561	−0.442	0.585
m8a	1.223	0.861	0.757	0.910	0.345	1.456
m8b	0.264	−0.032	−0.472	−1.110	−1.071	2.221
m9a	−0.883	−0.625	−0.407	−1.520	−0.784	1.886
m9b	0.013	−0.002	−0.307	−0.602	−0.512	2.523
m10a	−1.271	−1.253	−0.829	−1.098	−1.462	1.537
m10b	−1.007	−0.874	−0.902	−0.998	−0.547	2.301
m11a	−0.627	−0.559	−0.737	−1.553	−1.134	1.721
m11b	−0.408	−1.543	−0.800	−1.240	−1.045	2.523
m12	0.257	0.875	0.933	1.043	1.035	−0.813
m13	0.525	1.265	1.105	0.294	0.791	−0.707
m14a	−0.339	−0.403	0.571	−0.817	−0.849	0.978
m14b	−0.030	0.191	0.002	0.131	0.044	1.721
	Average Absolute Errors					
	0.486	0.533	0.541	0.754	0.630	

(PCs). The conventional r^2 for this analysis was 0.984. This model expressed good predictive ability for the test set of hydroxyethylurea compounds (r^2_{pred} = 0.662) with all compounds predicted within 1.27 log unit (1.7 kcal/mol in binding affinity) of their actual activities with an average absolute error of 0.49 log units (0.6 kcal/mol) across a range of 3.03 log units.

Alignment II yielded poorer cross-validated (r^2_{cross} = 0.653 with 8 PCs), conventional (r^2 = 0.990), and predictive results (r^2_{pred} = 0.625). These results are possibly indicative of poor molecular alignment caused by structural differences represented by additional protons on charged ligands and the resultant changes in charge distributions. These results are useful in that the analysis provided an indication of the sensitivity of CoMFA to such changes.

It was somewhat anticipated that, by realigning the charged molecules of the alignment II traning set using the field-fit minimization procedure, the internal consistency of the model represented by the cross-validated r^2 value would improve. This was not to be the case as another decrease in the cross-validated r^2 value was noted (r^2_{cross} = 0.607 and 8 PCs). The conventional statistical results (r^2 = 0.991) and the predictive ability of this model for the 18 test compounds (r^2_{pred} = 0.660) were not affected by the changes in the molecular alignment of the training set.

Alignment IV produced an internally predictive (r^2_{cross} = 0.659 with 7 PCs) and statistically significant (r^2 = 0.988) model. Although this model yielded a high cross-validated r^2, the predictive ability for the test set molecules was poor (r^2_{pred} = 0.327). The model mispredicted the activities of the test set molecules by as much as 1.6 log units (2.1 kcal/mol) and expressed the highest average absolute error (0.75 log units, 1.0 kcal/mol) in predictability. Furthermore, an underpredictive tendency was noted in that the activities of 12 out of the 18 molecules in the test set were underpredicted.

Active-site minimization of protonated ligands, alignment V, produced a correlation with a cross-validated r^2 of 0.642 using 7 PCs. The conventional r^2 for this model was 0.983. A respectable degree of predictability for the 18 test set molecules was produced (r^2_{pred} = 0.512) with an average absolute error in predictions of 0.54 log units (0.7 kcal/mol). The underpredictive trend of the previous model also based on active-site minimization (alignment IV) was not as pronounced in this analysis.

B. CoMFA Fields. The CoMFA steric and electrostatic fields for the analysis based on alignment I are presented as contour plots in Figures 2 and 3. The field values were calculated as the scalar product of the β-coefficient and the standard deviation associated with a particular column in the QSAR table (stdev*coeff). The values corresponding to steric columns are plotted as the percentage of contribution to the QSAR equation. In Figure 2, areas of high steric bulk tolerance (80% contribution), represented by the green polyhedra, are noted in the vicinity of the P1 and P1′ positions of the ligand. Both the active molecule (blue stick structure) (**10**) and the less active molecule (red stick structure) (**34**) have phenyl and isoquinoline substituents in those positions, respectively. Areas of low steric bulk tolerance (20%), represented by the yellow polyhedra in Figure 2, are noted near the P1′ position. The isoquinoline group of the less active molecule (**34**) is seen to extend into these contours.

The electrostatic contribution contour plot from the analysis based on alignment I is plotted as the stdev*coeff field in Figure 3. Areas of decreased tolerance for negative charge (20% contribution), designated by the red polyhedra, are noted in the vicinity of the scissile bond (P1–P1′) of the inhibitor opposite the hydroxy substituent of both the crystal ligands (yellow stick structure is JG365 (**28**) and green stick structure is U-75875 (**40**)). It is of interest to note that an area of increased tolerance for negative charge (80% contribution), designated by blue polyhedra, is localized in P2 near the histidine of **40**. With respect to the active site (not shown), the areas of decreased tolerance for negative charge are located in the vicinity of Gly-49, Ile-50, Gly-149, and Ile-150 and opposite the active-site aspartates (Asp-25 and Asp-125). These polyhedra overlap the area previously defined as being tolerant to increases in steric bulk (Figure 2). This suggests that this area of the active site may express a preference for hydrophobic interactions which is supportive of the preference for the *R* (or equivalent) isomer in the transition-state isostere of the ligand as determined from X-ray crystal structures.[43]

For further evaluation of the electrostatic characteristics of the ligands with respect to pharmacological potency, an analysis based on charged (protonated) molecules was utilized. In order to facilitate a direct comparison with the preceding analysis and to minimize any steric effect, the molecules of alignment I were protonated (if applicable) but not subjected to realignment. The steric plots from the analysis based on alignment II (not shown) were, therefore, not appreciably different from the neutral analysis (alignment I). As previously noted, increased negative charge appears to be desired near the P2 position. Both hydrophilic and hydrophobic groups (i.e. leucine and asparagine) groups are commonly found in the P2 position of molecules included in the training set. However, certain molecules of the training set (**39–44, 47,** and **49–53**) contain a positively charged residue (histidine) in this position. In the analysis based on alignment I, all molecules were represented as the neutral species; therefore, the electrostatic plots were not well-defined in this region, and the effect of noncharged histidine residues could not be ascertained. In the analysis based on alignment II, all histidine residues of the training set molecules were

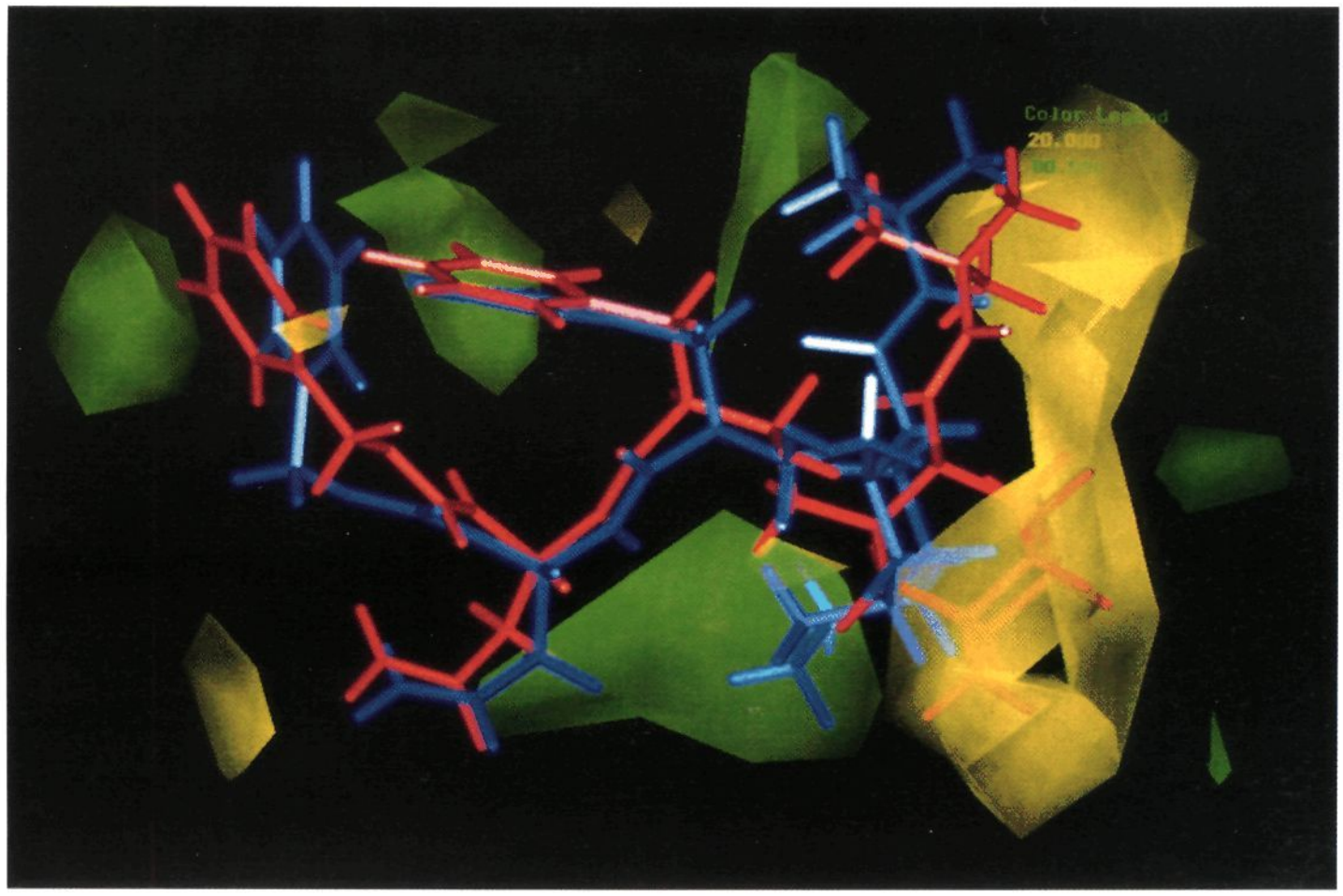


Figure 2. The CoMFA steric stdev*coeff contour plot from the analysis based on alignment I. Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. Active molecule (**10**) is represented as blue stick structure. Less active molecule (**34**) is represented as red stick structure.

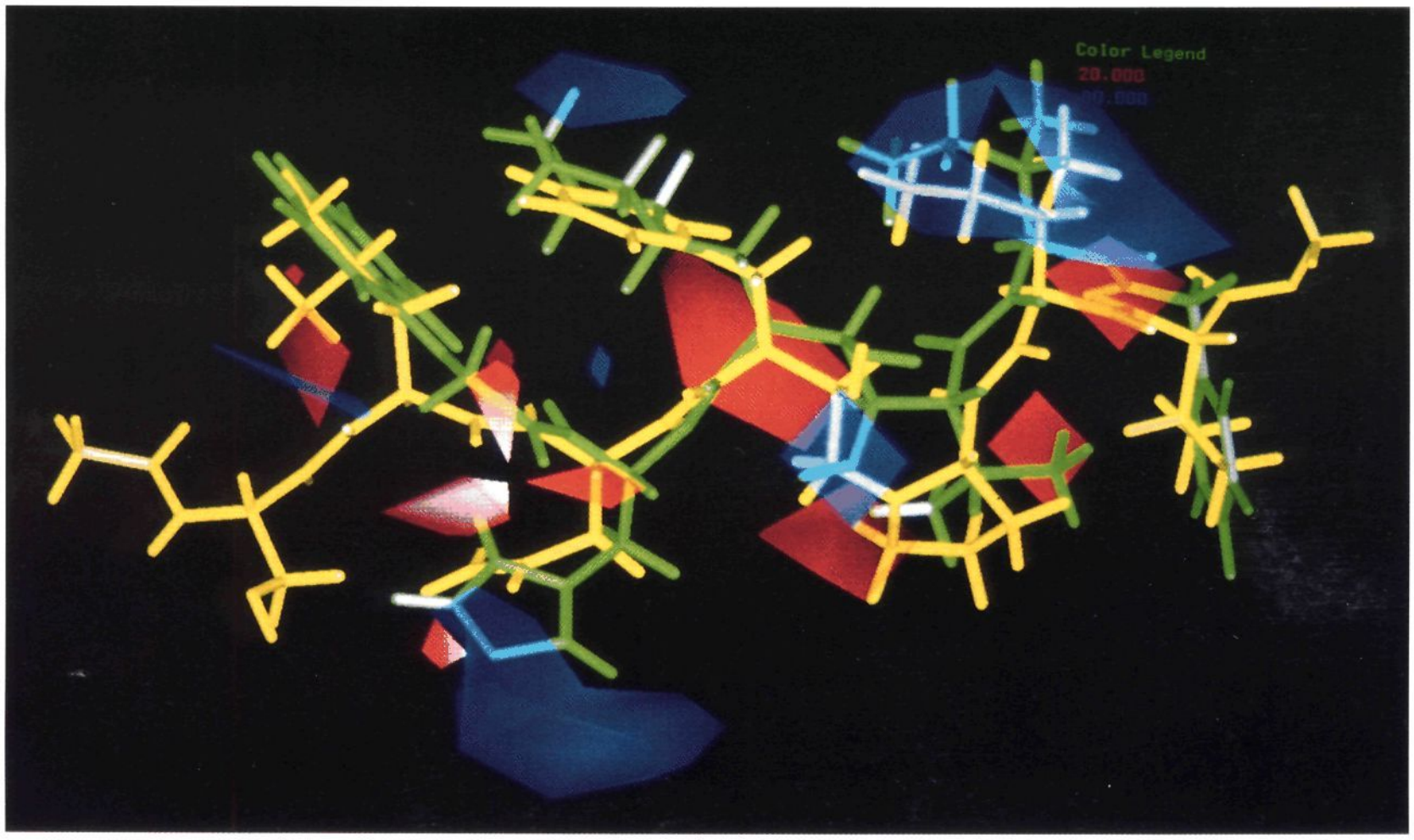


Figure 3. The CoMFA electrostatic stdev*coeff contour plot from the analysis based on alignment I. Negative charge favored areas (contribution level of 80%) are represented by blue polyhedra. Negative charge disfavored areas (contribution level of 20%) are represented by red polyhedra. JG365 (**28**) is represented as yellow stick structure. U-75875 (**40**) is represented as green stick structure.

protonated. This clearly indicates that positive charge in this area is not desired; thus, histidine in the P2 position is detrimental to the potency of the inhibitor.

Discussion

CoMFA is a shape-dependent 3-D QSAR technique which is sensitive to both conformational and configurational changes within a series of molecules. The various alignment rules selected for this study oriented the P1′ hydroxy substituent of all molecules in the training and test sets toward the Asp-25/Asp-125 and away from the "hydrolytic" water. This is consistent with the mechanistically-derived tetrahedral transition state of the normal substrate and supported by crystallographic evidence (Roche (**5**) and JG365 (**28**) express different chirality but the hydroxy is oriented similarly when the structures are

superimposed). Clearly, crystallographic data provided for additional insight into the development of the alignment rule. Without such evidence, it is possible that the inversion of chirality would have been represented by a mistaken reorientation of the hydroxy of the inhibitor in the active site.

The orientation of the P1′ hydroxy substituent produces a pronounced directing effect on the P2′ and P3′ (if present) substituents of the inhibitors. In compounds possessing the *R*-isomer in the P1′ position (i.e. crystal structure of **5**), the P2′ nitrogen (proline) is found as the *S*-isomer. The converse is found to be true as revealed by the crystal JG365 (**28**) structure. This effect is most clearly reflected in the dramatic differences in potency for equivalently substituted hydroxyethylamines with opposing P1′ hydroxy chiralities. CoMFA is sensitive to this conformational change as depicted in Figure 2. The isoquinoline group, a common feature of compounds **10** and **34**, is redirected depending on the chirality at the P1′ and corresponding P2′ positions.

The CoMFA model also provides an explanation for the antiparallel effects on the potencies of equivalently substituted hydroxyethylamine (**3**, **5**, and **10**) and norstatine-containing (**19**, **20**, and **21**) inhibitors with identical P1′ chirality (*R*). Within the hydroxyethylamine series, as the steric bulk of the P2′ position is increased, increases in biological potency are noted. The opposite trend is found in the norstatine series. Examination of the conformations of comparable molecules (i.e. **10** and **21**) reveals that the P2′ substituent of compound **21** is positioned much like that of JG365 (**28**) (Figure 4). This may be attributed to the achiral nature of the P2′ nitrogen. The nitrogen is amidic in the statine-based inhibitors and as such is trigonal-planar regardless of the chirality at the P1′ position. All CoMFA models presented in this paper accurately predict this trend.

The alignment rule for the test set molecules was active-site minimization of neutral ligands regardless of the alignment rule used for the training set. In the absence of crystallographic data for this novel class of inhibitors, active-site minimization seems to be a reasonable approach. The use of neutral ligands is justified since the pK_a of urea is 0.1. It was previously reported that the P1′ and P2′ substituents in at least one of the inhibitors from this class (**m4b**) were the reverse of the anticipated binding mode based on crystallographic data for similarly substituted inhibitors from other classes. The systematic search and energy-evaluation routine utilized herein is validated in that the conformer of **m4b** yielding the lowest interaction energy with the active-site complex expressed this "flipped" binding mode. Only seven test set molecules (**m3**, **m8a**, **m8b**, **m9b**, **m10b**, **m12**, and **m14b**) were predicted on the basis of molecular mechanical derived interaction energies to interact with the receptor in the "normal" mode.

Analyses based on the field-fit minimization alignment rule are better (i.e. more predictive) than the corresponding active-site-minimized models because field-fitting of a ligand to a representative crystal template allows for the subtle changes in the binding mode between chemical classes of inhibitors to be better represented. That is, molecules of a particular class are forced to conform to the structural characteristics of the given template crystal structure and then allowed to relax to the nearest local minimum-energy structure. Molecular mechanics minimization within an active site cannot possibly be expected to exactly simulate all of the ligand/receptor interactions and are limited by omission of solvent. Most recently, Sansom *et al.*[44] reported that, in order to maintain the experimental geometry and the important inhibitor/protease interactions during molecular mechanical minimization procedures on complexes, specific constraints must be placed on hydrogen bonds and α-carbon atoms. Furthermore, during the active-site minimization procedure employed herein, only side-chain atoms of the receptor fragment were allowed to relax. The assumption is that the spatial positioning of the backbone atoms of the receptor and the distribution of essential waters do not change during ligand binding. This may not be the case, since by holding the backbone and water atoms rigid, the ligand may be forced to assume a nonoptimal alignment within the active site. A partial solution to this problem may lie in the use of algorithms designed to solvate molecules.[45] This would effectively allow for the redistribution of waters within the active site. The details of such a procedure are being explored.

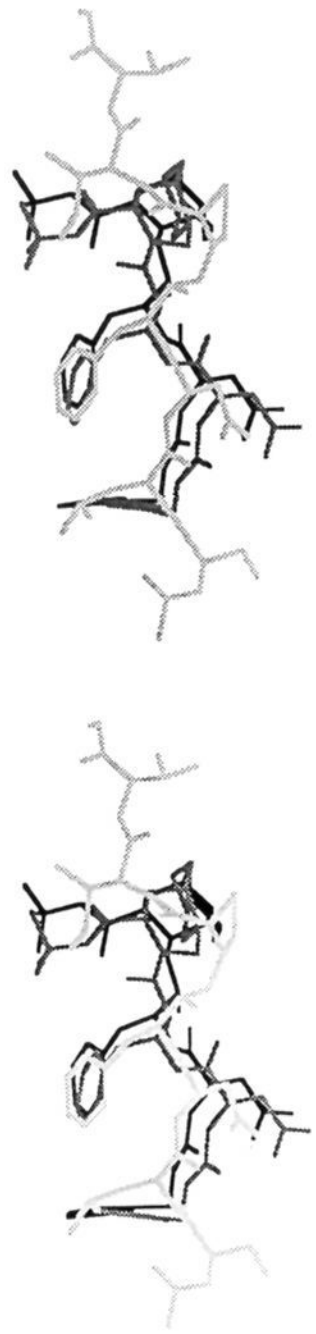


Figure 4. Superimposed stereoview of hydroxyethylamine (**10**, black; **21**, dark gray) and statine-based (**28**, light gray) inhibitors as found in alignment I.

CoMFA was designed to evaluate the steric and electrostatic environments of a series of molecules and to correlate changes in these two properties with changes in biological activities. CoMFA steric and electrostatic interactions alone do not give explicit consideration to the effects of solvation and lipophilicity, two factors which are known to strongly influence the free energy of binding.

Several applications, including Amsol[46] and Delphi,[47] have been developed to calculate (estimate) the solvation free energies of molecules. Likewise, many methods exist for the calculation of lipophilicity ranging from simplistic additive constituitive algorithms (c log P[48]) to grid-type algorithms which are capable of directly interfacing with CoMFA (HINT![49]). Currently, studies are being undertaken to examine the effects of the addition of regressors such as these to existing CoMFA models.

Conclusions

The development of CoMFA models should consist of two distinct phases which need emphasis. The primary goal in the construction of a successful CoMFA QSAR is most certainly the development of an internally self-consistent model, that is, a model which yields an acceptable cross-validated r^2 value. Self-consistency alone has traditionally been taken as a measure of predictability for a data set. Along these lines, others[50] have attempted to minimize the occurrence of spurious correlations in the training set by selecting only those variables which contribute most significantly to the prediction ability of the model. While the utilization of this advanced variables selection procedure has yielded improvements in the self-consistency of models versus normal linear PLS applications, it has been demonstrated that a self-consistent analysis does not necessarily possess predictive ability for novel analogs. The accurate prediction of biological activity (i.e. affinity) of novel compounds most assuredly is the ultimate goal of any QSAR analysis. As a measure of the "true" predictive ability of a CoMFA QSAR, we recommend the use of chemically novel test sets (ideally structurally and electronically different from the members of the training set) as originally proposed by DePriest *et al.*[19]

The use of chemically diverse test sets as measures of predictive ability is not exclusively limited to the CoMFA QSAR technique. CoMFA would, however, seem to be ideally suited for test sets of this type and, in this sense, possess an advantage over more conventional QSAR techniques in that molecules are represented in the analysis by molecular fields—specific atom types and functional groups are not specified per se. A successful CoMFA QSAR would be one that contains sufficient information regarding the molecular field so as to be capable of extrapolating to the effect of other chemical functionalities, not represented in the training set, in a given area.

Several models which sufficiently provide an internally self-consistent view of the variance in the biological potencies of a training set of structurally-diverse HIV-1 protease inhibitors and accurately predict the activities of novel structures using a 3-D QSAR technique (CoMFA) have been presented. Presently, the output from these analyses in the form of contour plots is being used to aid in the design of novel, third-generation HIV-1 protease inhibitors. Once constructed, candidate structures are aligned by active-site minimization. Using selected QSAR models, potential synthetic lead compounds are identified as those expressing high predicted potencies. The results of these rational design efforts will be reported in future publications.

Acknowledgment. The authors wish to thank Alexander Wlodawer (National Cancer Institute) and William Stallings (Monsanto) for supplying the crystal coordinates for the inhibitor/protease complexes. All authors gratefully acknowledge support for this research from National Institutes of Health Research Cooperative Agreement Grant No. 5 UO1 AI27302 and Grant GM 24483. One author (C.L.W.) wishes to acknowledge support from NIH Cardiovascular Training Grant No. T32HLO7275.

Appendix

All biological activities used in the present CoMFA study were expressed as

$$\text{Bio} = -\log_{10} \text{IC}_{50} \quad (1)$$

where Bio is the biological activity, and IC_{50} is the micromolar concentration of the inhibitor producing 50% inhibition.

Biological activities for the peptide series shown in Table III were published as K_i values.[14] In order to obtain IC_{50} values, the Cheng and Prusoff equation (Cheng, Y.-c.; Prusoff, W. H. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108) determined for the case involving one substrate and one competitive inhibitor present was used:

$$\text{IC}_{50} = K_i(1 + S/K_m) \quad (2)$$

where K_i is the dissociation constant of the enzyme-inhibitor complex, S is the substrate concentration, and K_m is the Michaelis constant of the substrate. This equation is valid when the velocity in the presence of the inhibitor is half the velocity in the absence of the inhibitor. The following numerical values were used: $K_m = 2.0$ mM[38] (Tomasselli, A. G. Personal communication, 1993), S = 2.5 mM,[14] and K_i as published for each inhibitor.[14]

On the basis of eq 2, IC_{50} values for all the other compounds in this series (Table III) were determined using the calculated K_m value. Results are listed in Table III.

Note Added in Proof. The atomic coordinates for test set molecule (**m4b**) have been obtained allowing a prediction of $p(IC_{50})$ (μM) = 0.928 (actual $p(IC_{50})$ = 0.899) based on alignment rule I. The RMS deviation between the crystal and active-site-minimized structures is 0.740 Å.

Supplementary Material Available. The atomic coordinates and partial atomic charges of all molecules are available from the authors upon request.

References

(1) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. Active Human Immunodeficiency Virus Protease is Required for Viral Infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686–4690.
(2) Krausslich, H.-G.; Wimmer, E. Viral Proteinases. *Annu. Rev. Biochem.* **1988**, *57*, 701.
(3) Rich, D. H. *Peptidase Inhibitors In Comprehensive Medicinal Chemistry*; Sammes, P. G., Ed.; Pergamon Press; Oxford, 1990; Vol. 2, pp 391–441.
(4) Rich, D. H. In *Proteinase Inhibitors*; Barret, A. J., Salvesen, G., Eds.; Research Monographs in Cell and Tissue Physiology; Elsevier Science Publ.: Amsterdam, 1986; pp 179–217.
(5) (a) Toth, M. V.; Chin, F.; Clawson, L.; Selk, L.; Kent, S. B. H.; Marshall, G. R. Unpublished results. (b) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, B. H.; Wlodawer, A. Structure of Complex of Synthetic HIV-1 Protease with a Substrate-Based Inhibitor at 2.3 Å Resolution. *Science* **1989**, *246*, 1149–1152.
(6) Marinler, A.; Toth, M. V.; Podlogar, B. L.; Ho, C. M. W.; Houseman, K.; Marhsall, G. R. HIV-1 Protease Inhibitors Derived from MVT-101: Kto-methylene Isosteres. Unpublished data.
(7) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. Novel Binding Mode of Highly Potent HIV-Proteinase Inhibitors Incorporating the (R)-Hydroxyethylamine Isostere. *J. Med. Chem.* **1991**, *34*, 3440–3442.
(8) Rich, D. H.; Sun, C.-Q.; Prasad, J. V. N. V.; Pathiasseril, A.; Toth, M. V.; Marshall, G. R.; Clare, M.; Mueller, R. A.; Houseman, K. Effect of Hydroxyl Group Configuration in Hydroxyethylamine Dipeptide Isosteres on HIV Protease Inhibition. Evidence of Multiple Binding Modes. *J. Med. Chem.* **1991**, *34*, 1222–1225.

(9) Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.; Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Krohn, A.; Lambert, R. W.; Merrett, J. H.; Mills, J. S.; Parkes, K. E. B.; Redshaw, S.; Ritchie, A. J.; Taylor, D. L.; Thomas, G. J.; Machin, P. J. Rational Design of Peptide-Based HIV Proteinase Inhibitors. *Science* **1990**, *248*, 258–261.
(10) Tucker, T. J.; Lumma, W. C., Jr.; Payne, L. S.; Wai, J. M.; de Solms, S. J.; Giuliani, E. A.; Darke, P. L.; Heimbach, J. C.; Zugay, J. A.; Schleif, W. A.; Quintero, J. C.; Emini, E. A.; Huff, J. R.; Anderson, P. S. A Series of Potent HIV-1 Protease Inhibitors Containing a Hydroxyethyl Secondary Amine Transition State Isostere: Synthesis, Enzyme Inhibition, and Antiviral Activity. *J. Med. Chem.* **1992**, *35*, 2525–2533.
(11) Appelt, K., Agouron Pharmaceutical Co., San Diego, CA. Unpublished results.
(12) Provided by A. Wlodawer, NCI, Frederick, MD.
(13) Tam, T. F.; Carriere, J.; MacDonald, I. D.; Castelhano, A. L.; Pliura, D. H.; Dewdney, N. J.; Thomas, E. M.; Bach, C.; Barnett, J.; Chan, H.; Krantz, A. Intriguing Structure-Activity Relations Underlie the Potent Inhibition of HIV Protease by Norstatine-Based peptides. *J. Med. Chem.* **1992**, *35*, 1318–1320.
(14) Thaisrivongs, S.; Tomasselli, A. G.; Moon, J. B.; Hui, J.; McQuade, T. J.; Turner, S. R.; Stronbach, J. W.; Howe, J. B.; Tarpley, W. G.; Heinrikson, R. L. Inhibitors of the Protease from Human Immunodeficiency Virus: Design and Modeling of a Compound Containing a Dihydroxyethylene Isostere Insert with a High Binding Affinity and Effective Antiviral Activity. *J. Med. Chem.* **1991**, *34*, 2344–2356.
(15) Thompson, W. J.; Fitzgerald, P. M. D.; Holloway, K. M.; Emini, E. A.; Darke, P. L.; McKeever, B. M.; Schleif, W. A.; Quintero, J. C.; Zugay, J. A.; Tucker, T. A.; Schwering, J. E.; Homnick, C. F.; Nunberg, J.; Springer, J. P.; Huff, J. R. Synthesis and Antiviral Activity of a Series of HIV-1 Protease Inhibitors with Functionality Tethered to the P1 or P1′ Phenyl Substituents: X-ray Crystal Structure Assisted Design. *J. Med. Chem.* **1992**, *35*, 1685–1701.
(16) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. Determination of the Relative Binding Free Energies of Peptide Inhibitors to the HIV-1 Protease. *J. Med. Chem.* **1991**, *34*, 2654–2659.
(17) Reddy, M. R.; Vishwanadhan, V. N.; Weinstein, J. N. Relative Differences in the Binding Free Energies of Human Immunodeficiency Virus 1 Protease Inhibitors: A Thermodynamic Cycle-Perturbation Approach. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10287–10291.
(18) Lybrand, T. P.; McCammon, J. A. Computer Simulation Study of the Binding of an Antiviral Agent to a Sensitive and a Resistant Human Rhinovirus. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 259–266.
(19) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors: A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active-site Geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372–5384.
(20) 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.
(21) Thomas, B. F.; Compton, D. R.; Martin, B. R.; Semus, S. F. Modeling the Cannabinoid Receptor: A Three-Dimensional Quantitative Structure-Activity Analysis. *Mol. Pharm.* **1991**, *40*, 656–665.
(22) Waller, C. L.; McKinney, J. D. Comparative Molecular Field Analysis of Polyhalogenated Dibenzo-p-dioxins, Dibenzofurans, and Biphenyls. *J. Med. Chem.* **1992**, *35*, 3660–3666.
(23) Loughney, D. A.; Schwender, C. F. A Comparison of Progestin and Androgen Receptor Binding using the CoMFA Technique. *J. Comput.-Aided Molec. Des.* **1992**, *6*, 569–581.
(24) Bjorkroth, J.-P.; Pakkanen, T. A.; Lindroos, J. Comparative Molecular Field Analysis of Some Clodronic Acid Esters. *J. Med. Chem.* **1991**, *34*, 2338–2343.
(25) Nicklaus, M. C.; Milne, G. W. A.; Burke, T. R., Jr. QSAR of Conformationally Flexible Molecules. Comparative Molecular Field Analysis of Protein-Tyrosine Kinase Inhibitors. *J. Comput.-Aided Molec. Des.* **1992**, *6*, 487–504.
(26) Klebe, G.; Abraham, U. On the Prediction of Binding Properties of Drug Molecules by Comparative Molecular Field Analysis. *J. Med. Chem.* **1993**, *36*, 70–80.
(27) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative Molecular Field Analysis (CoMFA). 2. Toward its use with 3D-Structural Databases. *Tet. Comp. Met.* **1990**, *3*, 47–59.
(28) Getman, D. P.; DeCrescenzo, G. A.; Heintz, R. M.; Reed, K. L.; Talley, J. J.; Bryant, M. L.; Clare, M.; Houseman, K. A.; Marr, J. J.; Mueller, R. A.; Vazquez, M. L.; Shieh, H.-S.; Stallings, W. C.; Stegeman, R. A. Discovery of a Novel Class of Potent HIV-1 Protease Inhibitors Containing the (R)-Hydroxyethylurea Isostere. *J. Med. Chem.* **1993**, *36*, 288–291.
(29) The program Sybyl 6.0 is available from Tripos Associates, 1699 South Hanley Road, St. Louis, MO 63144.
(30) Clark, M.; Carmer, R. D., III; Van Opdenbosh, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
(31) The program MOPAC 5.0 is available from Quantum Chemistry Program Exchange, no. 455.
(32) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(33) Weiner, S.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230–252.
(34) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. X-ray Crystallographic Structure of a Complex between a Synthetic Protease of Human Immunodeficiency Virus 1 and a Substrate-based Hydroxyethylamine Inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.
(35) Thanki, N.; Rao, J. K. M.; Foundling, S. I.; Howe, W. J.; Moon, J. B.; Hui, J. O.; Tomasselli, A. G.; Heinrikson, R. L.; Thaisrivongs, S.; Wlodawer, A. Crystal Structure of a Complex of HIV-1 Protease with a Dihydroxyethylene-containing Inhibitor: Comparisons with Molecular Modeling. *Protein Sci.* **1992**, *1*, 1061–1072.
(36) Toth, M. V.; Marshall, G. R. A Simple, Continuous Fluorometric Assay for HIV Protease. *Int. J. Peptide Protein Res.* **1990**, *36*, 544–550.
(37) Broadhurst, A. V.; Roberts, N. A.; Ritchie, A. J.; Handa, B. K.; Kay, C. Assay of HIV-1 Proteinase: A Colorimetric Method Using Small Pepetide Substrates. *Anal. Biochem.* **1991**, *193*, 280–286.
(38) Tomasselli, A. G.; Olsen, M. K.; Hui, J.; Staples, D. J.; Sawyer, T. K.; Heinrikson, R. L.; Tomich, C.-S. C. Substrate Analog Inhibition and Active-site Titration of Purified Recombinant HIV-1 Protease. *Biochemistry* **1990**, *29*, 264–269.
(39) Protease from the BRU strain of HIV-1 virus was expressed microbially and used to monitor the HIV-1 protease-mediated hydrolysis of an octapeptide substrate, VSQN-β-Naphthyl-alanine-PIV, by modifications to the method of Heimbach, J. C.; Garsky, V. M.; Michelson, S. R.; Dixon, R. A.; Sigal, I. S.; Darke, P. L. Affinity Purification of the HIV-1 Protease. *Biochem. Biophys. Res. Commun.* **1989**, *164*, 995–960.
(40) Heimbach, J. C.; Garsky, V. M.; Michelson, S. R.; Dixon, R. A.; Sigal, I. S.; Darke, P. L. Affinity Purification of the HIV-1 Protease. *Biochem. Biophys. Res. Commun.* **1989**, *164*, 995–960.
(41) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. The Covariance Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comp.* **1984**, *5*, 735–743.
(42) 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.
(43) Oprea, T. I.; Waller, C. L.; Marshall, G. R. 3D-QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 3. Interpretation of CoMFA Results. Submitted for publication in *Drug Des. Discov.*
(44) Sansom, C. E.; Wu, J.; Weber, I. T. Molecular Mechanics Analysis of Inhibitor Binding to HIV-1 Protease. *Protein Eng.* **1992**, *5*, 659–667.
(45) Blanco, M. Molecular Silverware. I. General Solutions to Excluded Volume Constrained Problems. *J. Comput. Chem.* **1991**, *12*, 237–247.
(46) (a) Cramer, C. J.; Truhlar, D. G. General Parameterized SCF Model for Free Energies of Solvation in Aqueous Solution. *J. Am. Chem. Soc.* **1991**, *113*, 8305–8311. (b) Cramer, C. J.; Truhlar, D. G. An SCF Solvation Model for the Hydrophobic Effect and Absolute Free Energies of Aqueous Solvation. *Science* **1992**, *256*, 213–216.
(47) (a) The program Delphi is available from Biosym Technologies, 9685 Scranton Road, San Diego, CA 92121. (b) Gilson, M.; Sharp, K.; Honig, B. Calculating the Electrostatic Potential of Molecules in Solution. *J. Comput. Chem.* **1988**, *9*, 327–335.
(48) (a) The program c log P is available from Daylight Chemical Information Systems, Irvine, CA. (b) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and their Uses. *Chem. Rev.* **1971**, *71*, 525–616.
(49) (a) Kellog, G. E.; Semus, S. F.; Abraham, D. J. HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J. Comput.-Aided Molec. Des.* **1991**, *5*, 545–552. (b) Kellogg, G. E.; Abraham, D. J. Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298.
(50) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20.