

# Three-Dimensional Quantitative Structure–Activity Relationship Study on Cyclic Urea Derivatives as HIV-1 Protease Inhibitors: Application of Comparative Molecular Field Analysis<sup>†</sup>

Asim Kumar Debnath\*

Biochemical Virology Laboratory, Lindsley F. Kimball Research Institute of The New York Blood Center, 310 East 67th Street, New York, New York 10021

Received June 17, 1998

Three-dimensional quantitative structure–activity relationship (3D-QSAR) models have been developed using comparative molecular field analysis (CoMFA) on a large data set (118 compounds) of diverse cyclic urea derivatives as protease inhibitors against the human immunodeficiency virus type 1 (HIV-1). X-ray crystal structures of HIV-1 protease bound with this class of inhibitors were used to derive the most probable bioactive conformations of the inhibitors. The enzyme active site was used as a constraint to limit the number of possible conformations that are sterically accessible. The test sets have been created keeping in mind structural diversity as well as the uniform simple statistical criteria (mean, standard deviation, high and low values) of the protease inhibitory activities of the molecules compared to the training sets. Multiple predictive models have been developed with the training sets (93 compounds in each set) and validated with the corresponding test sets (25 compounds in each set). All the models yielded high predictive correlation coefficients ( $q^2$  from 0.699 to 0.727), substantially high fitted correlation coefficients ( $r^2$  from 0.965 to 0.973), and reasonably low standard errors of estimates ( $S$  from 0.239 to 0.265). The steric and electrostatic effects have approximately equal contributions, 45% and 55% (approximately), respectively, toward explaining protease inhibitory activities. This analysis yielded models with significant information on steric and electrostatic interactions clearly discerned by the respective coefficient contour plots when overlapped on the X-ray structure of the HIV-1 protease. The HINT CoMFA study revealed significant contribution of hydrophobicity toward protease inhibitory activity. The 3D visualization technique utilizing these contour plots as well as the receptor site geometry may significantly improve our understanding of the inhibitor–protease (HIV-1) interactions and help in designing compounds with improved activity.

## Introduction

The virus-encoded protease plays a crucial role in the life cycle of the human immunodeficiency virus type 1 (HIV-1), the causative agent of AIDS, and is an ideal target for developing anti-HIV-1 drugs for the treatment of AIDS.<sup>1–9</sup> HIV protease cleaves the *gag* and *gag-pol* polyproteins.<sup>10</sup> This processing is essential for the maturation of viral particles and production of infectious virions.<sup>11,12</sup> HIV-1 protease belongs to the family of aspartyl proteases. HIV-1 protease is a homodimer with a 2-fold axis of symmetry. Recently, four protease inhibitors, Invirase (saquinavir; Hoffman-LaRoche), Norvir (ritonavir; Abbott Laboratories), Crixivan (indinavir; Merck), and Viracept (nelfinavir; Agouron), have been approved by the U.S. Food and Drug Administration (FDA). These drugs have shown promising results when used in combination with reverse transcriptase (RT) inhibitors.<sup>8</sup> However, treatment by these drugs leads to development of resistant HIV-1 mutants less sensitive to the inhibitors.<sup>13–16</sup> The emergence of HIV-1 mutant strains suggests that continuation of research to find newer inhibitors is important to counter this problem.

Structure–activity analysis is the foundation for understanding structural features of both the inhibitors and the target receptors responsible for biological activity and helps to design more effective inhibitors.<sup>17,18</sup> Several 3D quantitative structure–activity relationship (QSAR) studies based on comparative molecular field analysis (CoMFA) have been described (only a few of which are listed here).<sup>19–28</sup> The availability of X-ray crystal structures of inhibitors bound with the receptor may contribute to formulating effective predictive 3D-QSAR models as it helps (1) to identify possible bioactive conformations of related inhibitors in the active site and (2) to visualize in three dimensions the interactions of the inhibitors with the receptor. More than 150 HIV-1 protease structures have been solved, most of which are from pharmaceutical companies.<sup>29</sup> A considerable number of HIV-1 protease structures (39 as of June 1998) are available in the Brookhaven Protein Data Bank (PDB).<sup>30</sup> Hundreds of protease inhibitors against HIV-1 have been synthesized and reported by several groups. Despite these large resources, there are only a few reports concerning QSAR<sup>31–33</sup> and 3D-QSAR<sup>34–37</sup> studies on HIV-1 protease inhibitors.

To exploit the wealth of structural and biological data and to contribute toward the understanding of the SAR of HIV-1 protease inhibitors, we have undertaken

<sup>†</sup> Dedicated to Professor Corwin Hansch of Pomona College, California, on the occasion of his 80th birthday.

\* Corresponding author. Tel: (212) 570-3373. Fax: (212) 570-3299. E-mail: adebnath@server.nybc.org.

systematic structure–activity analyses of inhibitors of the cyclic urea class. A large number of these inhibitors have been synthesized in a systematic manner and biologically evaluated.<sup>31,32,38–45</sup> Nine X-ray crystal structures of cyclic urea derivatives complexed with HIV-1 protease are currently available in the PDB. Our recent success in deriving a 3D-QSAR predictive model<sup>37</sup> on a set of symmetrical cyclic urea derivatives against HIV-1 protease prompted us to extend the study to a larger and more diverse set of inhibitors of this class. Though there are a couple of reports on classical QSAR analyses<sup>31,32</sup> on a relatively small set of cyclic urea derivatives as HIV-1 protease inhibitors, no 3D-QSAR on this class of inhibitors has been reported so far. The analyses reported here entail the use of CoMFA as a tool to derive predictive models from 118 diverse cyclic urea derivatives. The results are expected to enhance our knowledge on the molecular recognition of cyclic urea HIV-1 protease inhibitors and help to design more effective inhibitors against this crucial enzyme.

## Materials and Methods

**Biological Data.** The HIV-1 protease inhibitory data, represented by  $K_i$  (nM) values, are taken from the literature<sup>32,33,45</sup> and listed in Table 1. The  $\log 1/K_i$  values were used to derive 3D-QSAR models. From a total of 118 compounds, three individual training sets were created with 93 compounds each and the other 25 compounds were used as the test set. Approximately three-fourths of the inhibitors were chosen for the training set and one-fourth for the test sets. This multi-model approach was used to validate the predictive ability of the training sets. The test sets were created based on the suggestions by Oprea et al.<sup>35</sup> They are (1) the biological assay methods for both training and test sets should be compatible; (2) the biological activity values should span several orders of magnitude but should not exceed activity values in the training set by more than 10%; (3) the test set should represent a balanced number of both active and inactive compounds for uniform sampling of the data.

**Modeling Tools.** Computer modeling was done on a Silicon Graphics Indigo2 Extreme (R4400) computer. The modeling study using the CoMFA methodology was done using the Sybyl 6.4 program.<sup>46</sup> The hydrophobic fields were calculated by the HINT program.<sup>47</sup> The conformational searches were performed using the “systematic search” routine within Sybyl. Structural minimizations were conducted by the Maximin2 minimizer option. The biopolymer module was used to retrieve and analyze all protease structures. The CoMFA feature of the QSAR module was used to derive both the predictive and fitted 3D-QSAR models. Visualization techniques in Sybyl along with stereoglasses (CrystalEyes) from Stereographics<sup>48</sup> were used for all 3D analyses.

**Conformational Analysis.** There are several X-ray crystal structures of the cyclic urea compounds bound with the HIV-1 protease inhibitors available in the PDB. Two of them, 1qbr and 1dmp, containing the ligands [(4*R*)-(4 $\alpha$ ,5 $\alpha$ ,6 $\beta$ ,7 $\beta$ )-3,3'-[[tetrahydro-5,6-dihydroxy-2-oxo-4,7-bis(phenylmethyl)-1*H*-1,3-diazepine-1,3(2*H*)-diyl]bis(methylene)]bis[*N*-2-thiazolylbenzamide] (XV638) and (4*R*,5*S*,6*S*,7*R*)-hexahydro-5,6-dihydroxy-1,3-bis[(3-aminophenyl)methyl]-4,7-bis(phenylmethyl)-2*H*-1,3-diazepin-2-one (DMP450), respectively (Figure 1), were used as templates to construct the molecules. The construction of the molecules was done within the active site of the HIV-1 protease by replacing side chains of the template molecule. Stereoglasses were used in this step of construction to identify any steric clashes, and corrections were made accordingly. A short (100-step) minimization using Maximin2 was used to generate a reasonable starting conformation for systematic conformational search.

The systematic conformational search was carried out in Sybyl 6.4 for all the flexible molecules within the active site

of the HIV-1 protease X-ray crystal structure. The entire protease structure and the portions of the ligands common to all these molecules were kept rigid, and only the flexible substituents were subjected to a systematic conformational search. The Kollman all-atom charges<sup>49</sup> were loaded from the dictionary for the protein portion of the complex, whereas Gasteiger–Marsili<sup>50,51</sup> charges were used for the ligand molecules. All the flexible bonds in the substituents were searched with 15° torsional angle increments in the torsional space.

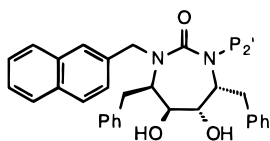
When the systematic conformational search yielded a number of conformations, the lowest energy conformation was selected for further minimization. The minimization was carried out within the active site of the HIV-1 protease using the Maximin2 minimization routine within Sybyl with a gradient change criterion of 0.05 kcal/mol·Å. During minimization, only the cyclic urea portion and the entire protein were kept rigid. The rest of the ligand was allowed to move within the receptor site to maximize the interactions of the ligand side chains with the amino acid side chains of the receptor. The minimization also removed any steric clashes that might have been present initially. The energy-minimized ligands were extracted from the complex, and the partial charges were calculated using Mopac 6.0 with the AM1 Hamiltonian<sup>52</sup> adopted in the Sybyl 6.4 program. No geometry optimization was done in this step, rather the keywords 1SCF and MMOK were used to calculate the charges and for the molecular mechanics corrections of the amidic bonds of the ligands, respectively.

**Alignment Rule.** The cyclic urea derivatives were all constructed from two template ligands (XV638 and DMP450) in two X-ray crystal structures (1qbr and 1dmp), respectively, as all the cyclic urea inhibitors belong broadly to these two classes of structures. The rms deviation of the C- $\alpha$  atoms between these two HIV-1 protease structures is only 0.326 Å. The distances of common hydrogen bond interactions between HIV-1 protease and these two inhibitors are also similar as documented in detail.<sup>15,53</sup> The conformational search and geometry optimizations were all done within the active site of the HIV-1 protease to find out the best possible conformations for biological activity. The cumulative information led us to believe that the alignment of the C- $\alpha$  atoms of the minimized inhibitor–protease complex on the corresponding C- $\alpha$  atoms of the template–protease structure is the most logical choice for alignment. This rule ensured (1) superposition of the atoms in the cyclic urea portion of the inhibitors to the cyclic urea atoms of the corresponding template structure, as these atoms in each inhibitor were kept rigid during minimization steps, and (2) formation of four hydrogen bonds by the two diol oxygen atoms of the cyclic urea moiety with the two catalytic aspartic acid residues (Asp25 and Asp25') and two hydrogen bonds between the carbonyl oxygen of the cyclic urea moiety and the isoleucine (Ile50 and Ile50') residues of the HIV-1 protease active site which are common for all 118 compounds used in this study.<sup>15,53</sup> Cyclic urea derivatives are designed to utilize the carbonyl oxygen to displace the structural water molecule (Wat301) from the active site as a strategy to design HIV-1 protease inhibitors.<sup>38</sup> Figure 2 shows the stereoview of all the aligned molecules (including test sets) to generate the CoMFA column.

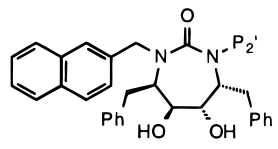
**CoMFA Study: Interaction Energy Calculation and Regression Analyses.** Standard steric and electrostatic CoMFA field energies of each inhibitor were calculated using an sp<sup>3</sup> carbon probe atom with a +1 charge at all intersections in regularly spaced (1.5 and 2.0 Å) grids in a grid box of 23.3 Å × 30.1 Å × 56.5 Å surrounding each molecule. Lennard–Jones 6–12 potential and Coloumbic potential functions, respectively, within the Tripos force field<sup>54</sup> and a distance-dependent (1/*r*) dielectric constant were used in the calculation. The grid box dimension was determined by the “create automatically” feature in the CoMFA module within the Sybyl program. This ensured that region boundaries were extended beyond 4 Å in each direction from the coordinates of each molecule. The same grid box was used in all calculations. An energy cutoff of 30 kcal/mol for both steric and electrostatic

**Table 1.** Observed Protease Inhibitory Activities and Residuals of Cyclic Urea Derivatives in the Training and Test Sets

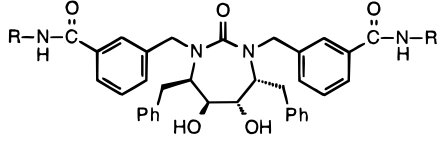
no.	P2/P2'	$K_i^a$ (nM)	obsd	$\log 1/K_i^a$					
				model 1 residual		model 2 residual		model 3 residual	
				training set	test set	training set	test set	training set	test set
1	methyl	5700	-3.75	-0.23			-2.43	-0.10	
2	ethyl	100	-2.00	0.25		-0.41		-0.03	
3	<i>n</i> -propyl	8	-0.90	0.08		0.00		0.04	
4	<i>n</i> -butyl	1.4	-0.15	0.12		0.35		0.20	
5	<i>n</i> -pentyl	1.6	-0.20		0.66	0.36		0.53	
6	<i>n</i> -hexyl	4.6	-0.66	0.42		0.16		0.15	
7	<i>n</i> -heptyl	260	-2.41		-0.30	-0.03		-0.10	
8	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	800	-2.90		-0.72	0.03		-0.19	
9	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>2</sub> CH <sub>3</sub>	1100	-3.04	-0.10		-0.41		-0.32	
10	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	7700	-3.89	0.36		0.26			-0.55
11	<i>i</i> -butyl	49	-1.69	-0.65			-0.95	-0.65	
12	<i>i</i> -pentyl	12	-1.08		-0.86	-0.09		-0.04	
13	<i>i</i> -hexyl	7	-0.84		0.19	-0.11		0.07	
14	<i>i</i> -heptyl	30	-1.48	0.10		0.19		0.19	
15	<i>i</i> -octyl	110	-2.04	-0.15		0.11		0.02	
16	neohexyl	36	-1.56		-0.99	-0.36		-0.23	
17	allyl	5.2	-0.72	-0.07			-0.24	-0.04	
18	2-methylpropen-3-yl	7.3	-0.86	-0.06		-0.33			0.02
19	isoprenyl	1.8	-0.25	0.09		0.36		0.59	
20	CH <sub>2</sub> CH <sub>2</sub> OCHCH <sub>2</sub>	60	-1.78		0.65	0.13		0.38	
21	3-propynyl	22	-1.34	0.37			-0.26	0.39	
22	cyclopropylmethyl	2.1	-0.32		0.16	-0.06		-0.02	
23	cyclobutylmethyl	1.3	-0.11	-0.03		0.29			0.66
24	cyclopentylmethyl	4.3	-0.63		-0.59	-0.35		0.12	
25	cyclohexylmethyl	37	-1.57	-0.52		-0.18		-0.48	
26	<i>N</i> -morpholino-2-ethyl	4000	-3.60		-3.78	0.05		0.05	
27	benzyl	3.0	-0.48	0.65		0.59		0.65	
28	2-picoyl	145	-2.16	-0.74		-0.74		-0.68	
29	3-picoyl	9.7	-0.99	0.14		0.18		0.09	
30	4-picoyl	90	-1.95	-0.40			-1.73	-0.28	
31	$\alpha$ -naphthylmethyl	86	-1.93	-0.24		0.07			-1.44
32	$\beta$ -naphthylmethyl	0.31	0.51	0.22		0.11			1.96
33	<i>o</i> -fluorobenzyl	34	-1.53	0.51			0.57	0.08	
34	<i>m</i> -fluorobenzyl	3.0	-0.48		-0.72	-0.22		-0.23	
35	<i>p</i> -fluorobenzyl	1.4	-0.15	0.47		0.10			1.14
36	<i>o</i> -chlorobenzyl	240	-2.38	0.10		0.48			-0.59
37	<i>m</i> -chlorobenzyl	0.89	0.05	-0.22			-0.11	-0.18	
38	<i>p</i> -chlorobenzyl	5.2	-0.72	-0.19		-0.13		0.25	
39	<i>m</i> -bromobenzyl	1.4	-0.15	0.16			0.22	-0.07	
40	<i>p</i> -bromobenzyl	27	-1.43	-0.02			-0.96	-0.17	
41	<i>m</i> -methylbenzyl	7.0	-0.84	-0.37		-0.69		-0.92	
42	<i>p</i> -methylbenzyl	5.7	-0.75	0.15			0.74	0.25	
43	<i>m</i> -(trifluoromethyl)benzyl	22	-1.34	-0.41		-0.46			-1.13
44	<i>p</i> -(trifluoromethyl)benzyl	51	-1.71	-0.03		-0.04		0.05	
45	<i>o</i> -methoxybenzyl	1870	-3.27	-0.09		-0.26			-1.82
46	<i>m</i> -methoxybenzyl	1.6	-0.20	-0.05		0.10		-0.06	
47	<i>p</i> -methoxybenzyl	157	-2.19		-0.72	0.10		-0.06	
48	<i>m</i> -nitrobenzyl	2.8	-0.45	-0.27		-0.05			-0.39
49	<i>m</i> -iodobenzyl	0.42	0.38		0.80	0.08		0.00	
50	<i>p</i> -(hydroxymethyl)benzyl (DMP323)	0.34	0.47	0.07			0.03	0.00	
51	<i>m</i> -(hydroxymethyl)benzyl	0.14	0.85		1.39	0.21		0.15	
52	<i>p</i> -hydroxybenzyl	0.12	0.92	0.41		0.03			1.76
53	<i>m</i> -hydroxybenzyl	0.12	0.92	0.06		0.14		0.34	
54	<i>m</i> -aminobenzyl (DMP450)	0.28	0.55	-0.02		-0.11		-0.09	



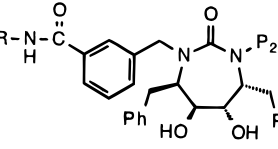
no.	P2'	$K_i^a$ (nM)	obsd	$\log 1/K_i^a$					
				model 1 residual		model 2 residual		model 3 residual	
				training set	test set	training set	test set	training set	test set
55	<i>n</i> -propyl	1.1	-0.04	-0.08			0.42	-0.22	
56	<i>n</i> -butyl	0.6	0.22	0.06			0.79	0.09	
57	allyl	1.4	-0.15		0.01	0.29		-0.05	
58	cyclopropylmethyl	1.5	-0.18	-0.37			0.12	-0.43	
59	cyclopentyl	0.28	0.55	0.39			0.97	0.31	
60	benzyl	2.3	-0.36	0.21		0.25		0.14	
61	3-picoyl	5.2	-0.72	-0.11		-0.10			-0.33

**Table 1** (Continued)


no.	P2'	$K_i^a$ (nM)	obsd	$\log 1/K_i^a$					
				model 1 residual		model 2 residual		model 3 residual	
				training set	test set	training set	test set	training set	test set
<b>62</b>	4-picoyl	6.9	-0.84	-0.24			-0.95	-0.24	
<b>63</b>	<i>p</i> -fluorobenzyl	3.6	-0.56	-0.24		-0.05			0.02
<b>64</b>	<i>p</i> -(hydroxymethyl)benzyl	0.93	0.03	-0.30		-0.20			1.49
<b>65</b>	<i>m</i> -aminobenzyl	1.0	0.00		-0.13	-0.22		-0.09	
<b>66</b>	<i>m</i> -hydroxylbenzyl	0.33	0.48	0.45		0.37		0.39	



no.	R	$K_i^a$ (nM)	obsd	$\log 1/K_i^a$					
				model 1 residual		model 2 residual		model 3 residual	
				training set	test set	training set	test set	training set	test set
<b>67</b>	H	0.039	1.41	-0.11		0.02		-0.06	
<b>68</b>	NH <sub>2</sub>	0.018	1.74	0.04		-0.07			0.03
<b>69</b>	OH	0.020	1.70	0.19		0.14		0.02	
<b>70</b>	OCH <sub>3</sub>	0.045	1.35	-0.06		-0.13		-0.06	
<b>71</b>	CH <sub>3</sub>	0.066	1.18	-0.12		-0.23		-0.30	
<b>72</b>	CH <sub>2</sub> CH <sub>3</sub>	0.210	0.68	-0.05		0.01			-0.09
<b>73</b>	CH(CH <sub>3</sub> ) <sub>2</sub>	0.579	0.24	-0.14		0.02		0.20	
<b>74</b>	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	0.359	0.44	0.05		0.13		0.17	
<b>75</b>	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	0.424	0.37	-0.05			0.13	0.16	
<b>76</b>	C(CH <sub>3</sub> ) <sub>3</sub>	2.400	-0.38	-0.25		-0.13			-0.42
<b>77</b>	CH <sub>2</sub> C <sub>3</sub> H <sub>5</sub>	0.741	0.13	-0.04		0.17			-0.45
<b>78</b>	CH <sub>2</sub> CF <sub>3</sub>	0.210	0.68	-0.05		0.01		-0.09	
<b>79</b>	CH <sub>2</sub> CN	0.063	1.20	-0.11			-0.17	-0.23	
<b>80</b>	benzyl	0.430	0.37	0.14		-0.10		-0.13	
<b>81</b>	4-pyridinyl	0.410	0.39	0.13		0.01		0.09	
<b>82</b>	3-pyridinyl	0.290	0.54	0.08		-0.04		0.11	
<b>83</b>	2-pyridinyl	0.043	1.37	-0.13		-0.34		-0.25	
<b>84</b>	2-(3-CH <sub>3</sub> -pyridinyl)	0.260	0.59		-0.78	-0.13		-0.22	
<b>85</b>	2-(4-CH <sub>3</sub> -pyridinyl)	0.027	1.57		0.84	-0.30		0.02	
<b>86</b>	2-(5-CH <sub>3</sub> -pyridinyl)	0.011	1.96	0.24			1.06	-0.01	
<b>87</b>	2-(6-CH <sub>3</sub> -pyridinyl)	0.020	1.70		0.30	-0.08		0.01	
<b>88</b>	2-(4,6-di-CH <sub>3</sub> -pyridinyl)	0.016	1.80	-0.11			0.57	0.00	
<b>89</b>	2-(5-Cl-pyridinyl)	0.012	1.92	0.27		0.30		0.17	
<b>90</b>	2-(3,5-di-Cl-pyridinyl)	0.245	0.61	-0.18			0.48	0.00	
<b>91</b>	2-(5-Br-pyridinyl)	0.035	1.46	-0.29		-0.18		-0.40	
<b>92</b>	2-(4-CH <sub>3</sub> -pyrimidinyl)	0.115	0.94		-0.65	-0.08		-0.02	
<b>93</b>	2-(5-CF <sub>3</sub> -pyridinyl)	0.085	1.07	-0.13			-0.12	-0.26	
<b>94</b>	2-pirazinyl	0.018	1.74	-0.13		0.28			0.59
<b>95</b>	2-pyrimidinyl	0.152	0.82	-0.11		0.16		-0.09	
<b>96</b>	4-methyl-2-oxazolyl	0.064	1.19	-0.24		-0.27			-1.32
<b>97</b>	5-(trifluoromethyl)-1,3,4-thiadiazol-2-yl	0.180	0.74	0.06		0.00		-0.07	
<b>98</b>	1,3,4-thiadiazol-2-yl	0.110	0.96	-0.05		0.09			0.09
<b>99</b>	2-thiazolyl	0.027	1.57	-0.15		-0.04		0.12	
<b>100</b>	4-methylthiazol-2-yl	0.025	1.60	0.35		0.26			-0.18
<b>101</b>	5-methylthiazol-2-yl	0.014	1.85		0.87	0.12		-0.03	
<b>102</b>	2-imidazolyl	0.014	1.85	0.05		0.08		0.02	
<b>103</b>	2-benzimidazolyl	0.024	1.62	0.38			1.35	0.57	



no.	R	P2'	$K_i^a$ (nM)	obsd	$\log 1/K_i^a$					
					model 1 residual		model 2 residual		model 3 residual	
					training set	test set	training set	test set	training set	test set
<b>104</b>	2-pyrazinyl	3,5-dimethoxyphenyl	2.511	-0.40	-0.09		-0.10		0.16	
<b>105</b>	2-(5-CH <sub>3</sub> -pyridinyl)	3,5-dimethoxyphenyl	1.900	-0.28	0.27			-0.05	0.06	
<b>106</b>	2-(6-CH <sub>3</sub> -pyridinyl)	3,5-dimethoxyphenyl	0.700	0.15		0.58	0.02		0.10	
<b>107</b>	2-pyridinyl	3,5-dimethoxyphenyl	0.860	0.06	0.23			-0.87	-0.07	
<b>108</b>	2-pyrazinyl	3-methoxyphenyl	0.038	1.42	0.17		-0.06		0.02	
<b>109</b>	2-(5-CH <sub>3</sub> -pyridinyl)	3-methoxyphenyl	0.069	1.16		0.11	0.16		-0.11	
<b>110</b>	2-(6-CH <sub>3</sub> -pyridinyl)	3-methoxyphenyl	0.053	1.27	0.06		0.09		-0.17	
<b>111</b>	2-pyridinyl	3-methoxyphenyl	0.047	1.33		0.36	0.14		0.21	



Table 1 (Continued)

no.	R	P2'	$K_i^a$ (nM)	obsd	log 1/ $K_i^a$					
					model 1 residual		model 2 residual		model 3 residual	
					training set	test set	training set	test set	training set	test set
112	2-(5-CH <sub>3</sub> -pyridinyl)	3-aminophenyl	0.075	1.12	-0.10		-0.12			
113	2-pyridinyl	3-nitrophenyl	0.096	1.02	0.27		0.05		0.15	
114	1,1-dimethylethyl	3-aminophenyl	0.410	0.39	-0.12		-0.12		0.13	
115	2-pyrazinyl	3-aminophenyl	0.016	1.79	0.35		0.02			-0.48
116	2-benzimidazolyl	3-aminophenyl	0.023	1.64	0.11		0.10		0.15	
117	2-imidazolyl	3-aminophenyl	0.012	1.92		0.81	0.11		0.29	
118	CH <sub>2</sub> CN	benzoylglycin-3-yl	0.024	1.62	-0.25		0.24		-0.04	

<sup>a</sup> SD  $K_i < \pm 40\%$ .

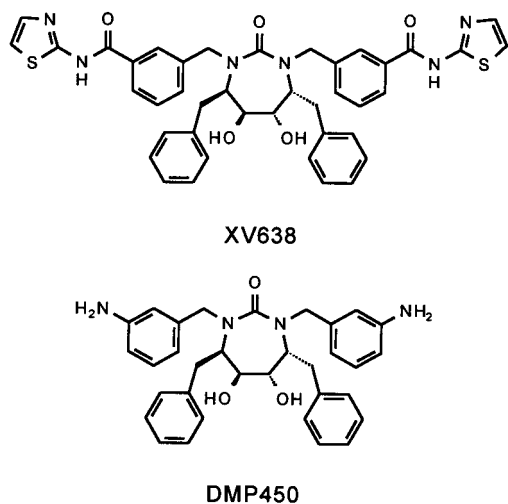


Figure 1. Chemical structures of XV638 and DMP450.

contributions was set, and the electrostatic terms were dropped within regions of steric maximum, i.e., 30 kcal/mol. The results are shown in Table 2.

The regression analyses of the CoMFA field energies were performed by the PLS algorithm<sup>55</sup> adopted in the QSAR module within Sybyl. Due to the large data set, 10 orthogonal components (also known as latent variables) were first extracted by the PLS method using the leave-one-out cross-validation technique. The steric and electrostatic fields were scaled using the CoMFA Standard to equalize their weight in the CoMFA models. The minimum  $\sigma$  of 2.0 kcal/mol was set as column filter to reduce the noise in the PLS analyses. The optimal number of components was derived from the cross-validated runs having the lowest standard error of predictions ( $s$ ). The same optimum number of components was used for deriving the best fitted correlation model using the non-cross-validated run. A minimum  $\sigma$  of 0.0 and CoMFA scaling of CoMFA Standard were used.

**3D-QSAR Study Using HINT<sup>56</sup> Methodology.** HINT methodology was used in order to verify whether the steric effect derived from the standard Sybyl CoMFA study has any relationship to the hydrophobic contribution from the substituents in the protease inhibitors used in the study. The training set from model 3 was only used, and all other conditions were identical to those used for the CoMFA study in model 3.

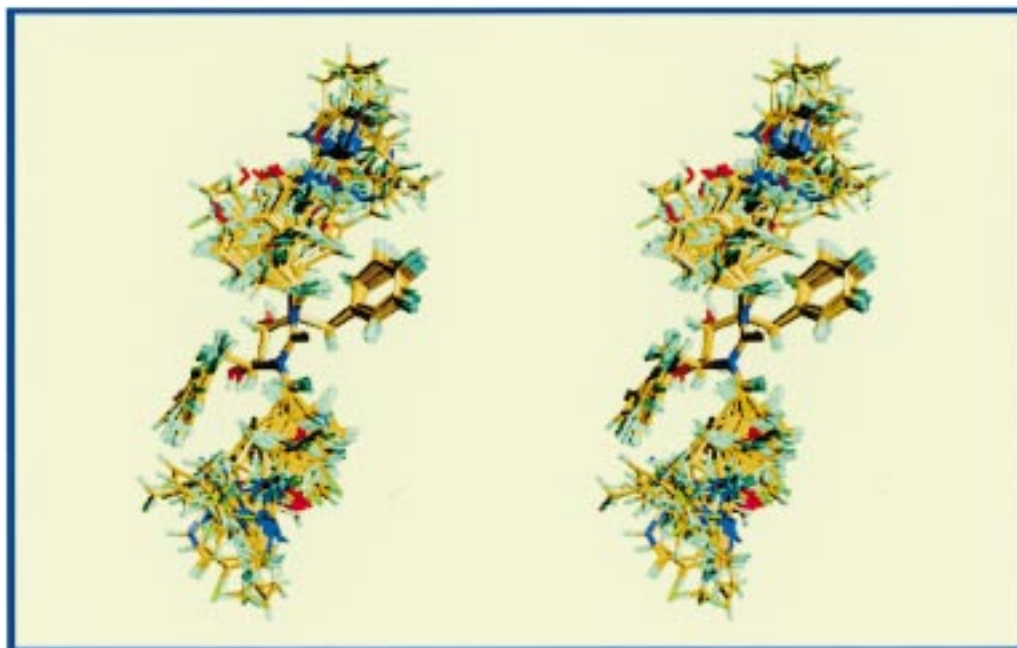
## Results and Discussion

**Training and Test Sets.** Due to the availability of a large data set of inhibitors with diverse structures within the same cyclic urea class having a wide range of activities, a multimodel approach was used to validate rigorously the overall approach for model development

and its utility as a predictive tool. Three different training sets were created with 93 molecules in each set. Three corresponding test sets were also created to validate the predictive capability of the models derived from the training sets. Though there were overlaps among molecules in the three training sets, no overlap was allowed in the test sets to ensure rigorous validation for each predictive model. Great care was exercised to create the test sets to ensure uniform sampling of biological activity. Table 3 shows in detail the simple statistics of the biological data (log 1/ $K_i$ ) for the training and test sets of models 1–3. Figure 3 shows the histograms of biological activity (log 1/ $K_i$ ) data for all the molecules, the training sets, and the test sets of models 1–3. The results indicate that the biological activity data distribution in both the training and test sets was uniform.

**Conformational Analyses.** Identification of the biologically relevant conformations in any 3D-QSAR study is one of the important criteria in developing models. X-ray crystal coordinates of the inhibitor bound with the receptor may help substantially in this crucial step. The systematic conformational search within the active site of HIV-1 protease was used as a key constraint to search for the biologically relevant conformations. This approach was used quite extensively by Marshall's group in their work on CoMFA on a series of HIV-1 protease inhibitors.<sup>34,35</sup> During the course of our study, it became evident that if conformational search is done in the absence of the receptor, many low-energy conformations would not fit the receptor cavity. In other words, those conformations were not sterically accessible in the receptor site for optimal interactions. Due to limited computer resources and time limitations, the entire receptor structure was kept rigid during conformational search and energy minimization steps, although admittedly the structure is flexible in reality. The lowest energy conformers were then energy minimized within the receptor cavity. Due to the active site structural constraints, unnecessary movement of the inhibitor molecule can be avoided during minimization. The structures of the ligand from the X-ray crystallographic ligand–protease structures were used 'as is' without any conformational search or minimization.

**CoMFA Study.** Three different models using both 1.5- and 2.0-Å grid spacings were developed. Table 2 indicates that all three models yielded consistent results with each grid spacing, but the 1.5-Å grid spacing



**Figure 2.** Stereoview of all the aligned cyclic urea inhibitors in the training and test sets.

**Table 2.** Results of the CoMFA PLS Analyses<sup>a</sup> of the Training Sets

statistics	model 1		model 2		model 3		HINT	HINT-CoMFA
	CoMFA		CoMFA		CoMFA			
	1.5 Å <sup>c</sup>	2.0 Å	1.5 Å	2.0 Å	1.5 Å	2.0 Å		
optimum no. of components	7	7	7	7	7	6	4	7
$q^2_{\text{cross}}^b$	0.713	0.699	0.727	0.725	0.705	0.699	0.715	0.703
standard error of predictions ( <i>s</i> )	0.753	0.770	0.767	0.770	0.764	0.766	0.737	0.766
$r^2$	0.965	0.952	0.973	0.969	0.964	0.944	0.805	0.965
standard error of estimate ( <i>S</i> )	0.263	0.302	0.239	0.257	0.265	0.330	0.610	0.261
<i>F</i> test	334.2	241.8	445.2	383.0	329.3	241.8	90.0	338.9
contributions:								
steric	0.429	0.466	0.454	0.479	0.453	0.509		0.347
electrostatic	0.571	0.534	0.546	0.521	0.547	0.491		0.425
HINT							1.000	0.228

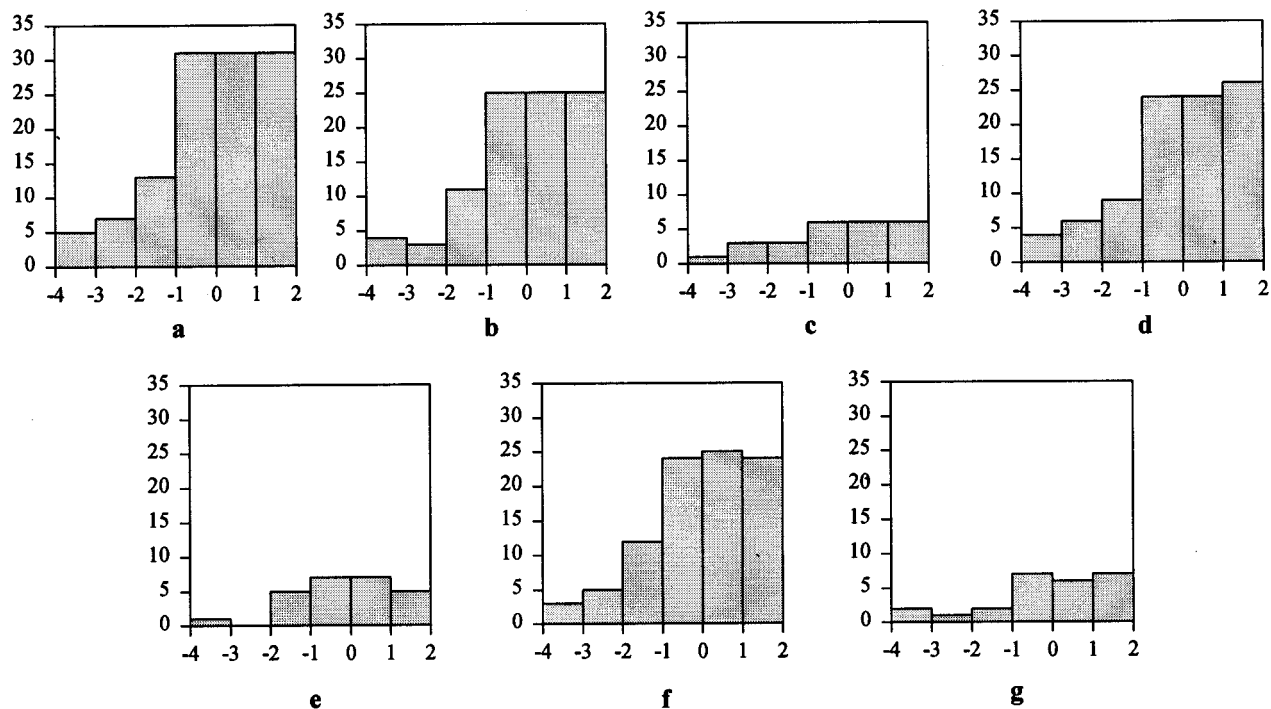
<sup>a</sup> CoMFA-Standard was used as the scaling method in all CoMFA PLS analyses. Energy cutoff of 30 kcal/mol was used for both steric and electrostatic contributions. <sup>b</sup> Minimum  $\sigma = 2.0$  was used for leave-one-out cross-validated runs, whereas minimum  $\sigma = 0.0$  was used for non-cross-validated runs. <sup>c</sup> Lattice spacings.

**Table 3.** Simple Statistics of the Protease Inhibitory Activities of the Training and Test Sets

model	set	log 1/ $K_i$			
		mean	SD	high	low
1	training	-0.01	1.35	1.96	-3.89
	test	-0.23	1.54	1.92	-3.60
2	training	-0.02	1.41	1.92	-3.89
	test	-0.19	1.31	1.96	-3.75
3	training	-0.05	1.35	1.96	-3.75
	test	-0.09	1.55	1.79	-3.89

resulted in slightly better results in terms of both cross-validated and non-cross-validated statistics. The models with the 1.5-Å grid spacing were selected as the models of choice for the predictive purpose. At a higher grid spacing (2.0 Å), some information on H-bonding and Lennard-Jones potentials may be lost as they are dependent on distances.<sup>57</sup> A lower grid spacing (1.0 Å) may generate more noise in the PLS calculations and would require more computational effort. The 2-Å grid was shifted by +0.5 Å to find out the effect of altered lattice point locations on the results of the CoMFA study. This resulted in a somewhat lower correlation (optimum number of components = 7;  $q^2 = 0.662$ ,  $s =$

0.816;  $r^2 = 0.946$ ,  $S = 0.328$ ) as compared to the original study with 2 Å (optimum number of components = 6;  $q^2 = 0.699$ ,  $s = 0.766$ ;  $r^2 = 0.944$ ,  $S = 0.330$ ). Therefore, the following model descriptions and detailed analyses will be limited to models with 1.5-Å grid spacing. The best predictive CoMFA models were selected based on the lowest standard error of predictions (*s*) for cross-validated PLS analyses as shown in Table 4. The optimal number of components in each case was 7. Considering high standard deviations in  $K_i$  values (SD < ±40%) and the diversity of cyclic urea inhibitors used to derive the models, significantly high cross-validated correlation coefficients ( $q^2$  from 0.699 to 0.727) and reasonable standard error of predictions (*s* from 0.753 to 0.767) were obtained. The non-cross-validated PLS runs with the same number of components yielded high fitted correlation coefficients ( $r^2$  from 0.964 to 0.973) and relatively low standard error of estimates (*S* from 0.239 to 0.265). The models explained approximately 97% of the variance in protease inhibitory activities of the compounds used in each training set. The contributions of steric and electrostatic interactions to the protease inhibitory activities are almost equal, i.e., approximately



**Figure 3.** Histogram of protease inhibitory activity (abscissa:  $\log 1/K_i$  in nM) vs number of molecules (ordinate): (a) all 118 molecules; (b and c) model 1; (d and e) model 2; (f and g) model 3. Panels b, d, and f are training sets, and panels c, e, and g are tests sets.

**Table 4.** Statistics of Cross-Validated PLS CoMFA Using 1.5-Å Lattice Space

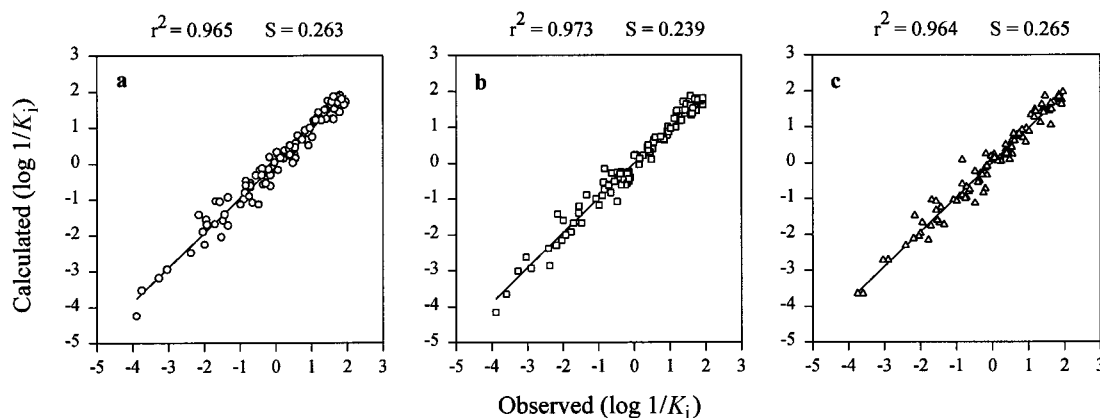
model	statistics	components									
		1	2	3	4	5	6	7 <sup>a</sup>	8	9	10
1	<i>s</i>	0.961	0.908	0.856	0.810	0.792	0.765	0.753	0.756	0.760	0.771
	<i>q</i> <sup>2</sup>	0.498	0.557	0.611	0.655	0.674	0.700	0.713	0.713	0.714	0.719
2	<i>s</i>	1.016	0.961	0.929	0.851	0.840	0.802	0.767	0.768	0.755	0.755
	<i>q</i> <sup>2</sup>	0.488	0.547	0.584	0.653	0.665	0.698	0.727	0.730	0.742	0.745
3	<i>s</i>	0.940	0.866	0.843	0.858	0.840	0.802	0.764	0.765	0.784	0.795
	<i>q</i> <sup>2</sup>	0.520	0.598	0.623	0.614	0.635	0.671	0.705	0.707	0.696	0.691

<sup>a</sup> Selected components. *s*, standard error of predictions.

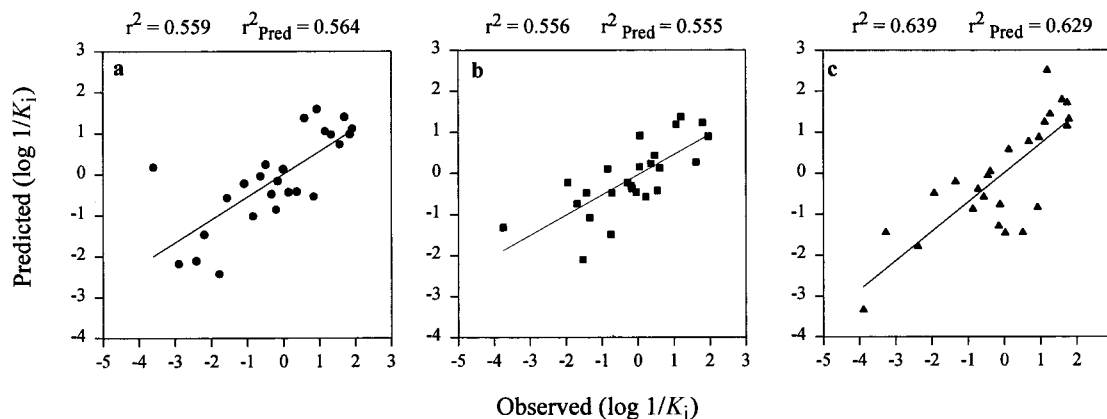
45% and 55%, respectively. The HINT model alone yielded a high cross-validated correlation ( $q^2 = 0.715$ ) with a standard error of prediction of 0.737 with four components. A fitted PLS regression with the same number of components yielded a somewhat lower  $r^2$  (0.805) value with a significantly higher standard error of estimate ( $S = 0.610$ ). When the HINT column was used with steric and electrostatic interactions, the overall statistics was similar to that obtained without HINT data ( $q^2 = 0.703$ ,  $s = 0.766$ ;  $r^2 = 0.965$ ,  $S = 0.261$ ). In other words, there was no improvement of the quality of the PLS regression analysis. However, the relative contribution data (steric, 0.347; electrostatic, 0.425; HINT, 0.228) indicate that the hydrophobic effect is correlated partially with the steric and electrostatic contributions in the CoMFA and suggests that hydrophobicity of the substituent groups contribute to biological activity.

The observed anti-protease activities and residuals of molecules in the training sets and the test sets from each model (excluding HINT and HINT CoMFA) are listed in Table 1. Plots of observed versus calculated activities of the training set molecules are shown in Figure 4. The plots show consistent linear relationships between observed and calculated activities for all three models developed. The predictive power of all three

models was further validated by calculating their "predictive  $r^2$ " ( $r^2_{\text{pred}}$ ) values. Models 1 and 2 yielded  $r^2_{\text{pred}}$  values of 0.564 and 0.555, respectively, whereas model 3 showed relatively better predictive power ( $r^2_{\text{pred}} = 0.629$ ). Plots of observed versus predicted activities of the test set molecules are shown in Figure 5. A region-focusing technique, available in the advanced CoMFA module in Sybyl, was used using model 3 and yielded very similar results to the cross-validated run ( $q^2 = 0.737$ ,  $S = 0.720$ ). Thus this technique did not improve substantially the quality of the model. It is interesting to note that model 2 had the test set which contained one of the most active protease inhibitors DMP323 (**50** in Table 1), an initial candidate for phase I clinical trials, and the model predicted the activity remarkably well (obsd 0.47 vs calcd 0.44). The agreement between the calculated and experimentally determined anti-protease activity of the other clinical candidate DMP450 (**54** in Table 1), not part of any test set, was also excellent (obsd 0.55 vs 0.57, 0.66, and 0.68, respectively, for models 1–3). All three models made worse predictions for one of the least active compounds in each case: e.g., (**26**, obsd  $-3.60$  vs calcd 0.18 by model 1; **1** obsd  $-3.75$  vs calcd  $-1.32$  by model 2 and **45** obsd  $-3.27$  vs calcd  $-1.45$  by model 3). The possible explanation for this discrepancy is that some of the substituents



**Figure 4.** Observed vs calculated protease inhibitory activities of the training sets in (a) model 1, (b) model 2, and (c) model 3. PLS correlation coefficients and standard error of estimates ( $S$ ) are shown for each plot.



**Figure 5.** Observed vs predicted protease inhibitory activities of the test sets in (a) model 1, (b) model 2, and (c) model 3. Correlation coefficients and the predictive  $r^2$  are shown for each plot.

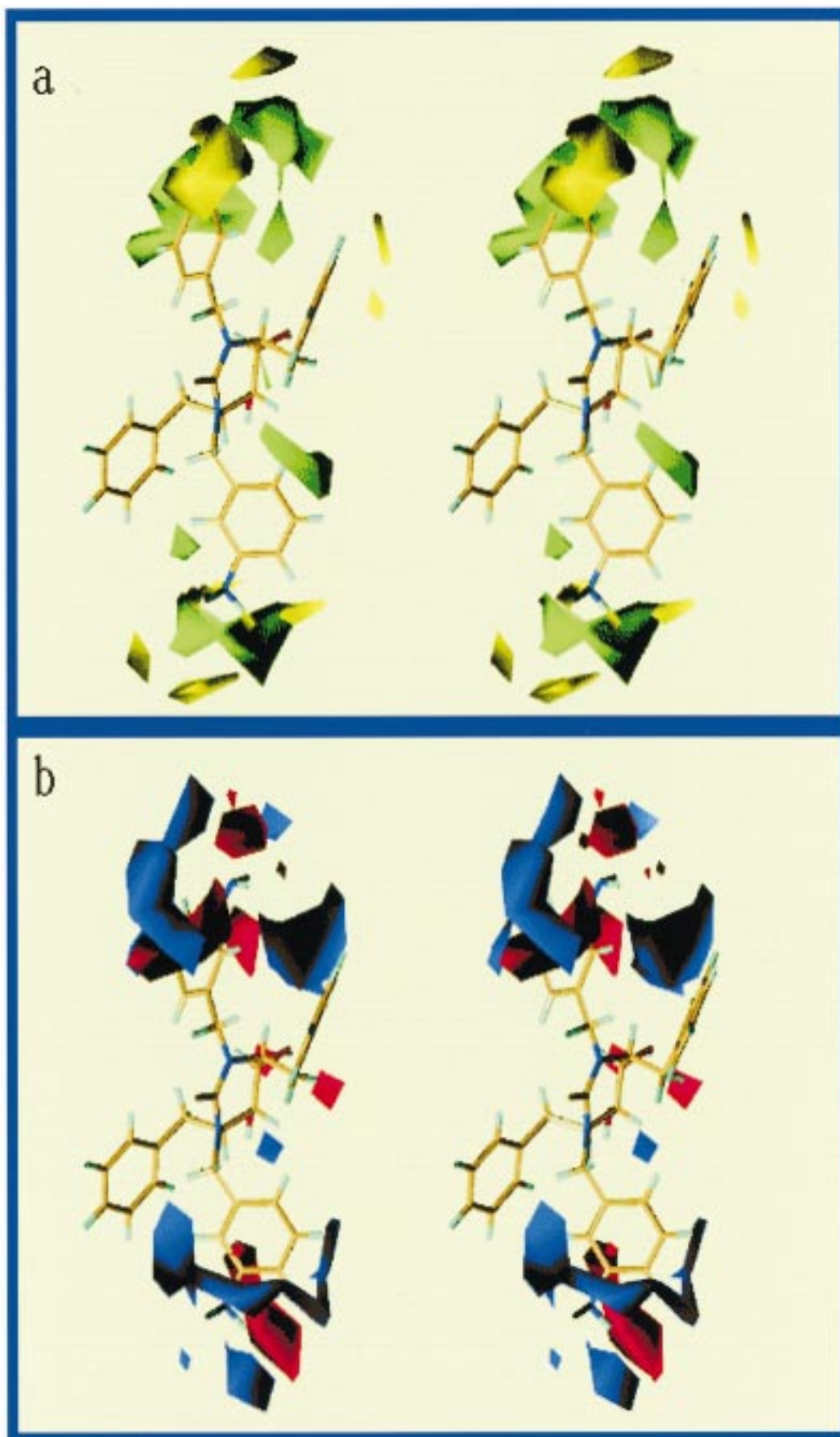
in compounds with inaccurately predicted activity occupy the space not occupied by any substituents in compounds selected for the corresponding training set by which the predictions were made.

An analysis of the percent extrapolation needed to predict anti-protease activity of all the 25 compounds in the test set in model 3 revealed that 18 inhibitors needed only 0–1% extrapolation whereas 6 needed 2–3% extrapolation. Only one compound (**48**) needed about 12% extrapolation for predicting anti-protease activity. The low requirement of extrapolation needed to predict protease inhibitory activity by the training set in model 3 gives additional confidence on its usefulness as a predictive model.

Two major advantages of 3D-QSAR using the CoMFA approach are (1) a fundamental understanding of the interactions of the ligands with the binding site can be achieved in real 3D perspective; (2) the 3D coefficient contour plots from the non-cross-validated PLS analyses of the training sets can be viewed and compared in three dimensions with respect to the real receptor site structure. A critical analysis of contour plots in relation to the receptor site structure may provide valuable insight into drug–receptor interactions and help in designing inhibitors with improved activity. The coefficient contour plots of steric and electrostatic interactions from model 3 (selected based on the  $r^2_{\text{pred}}$  values) are shown in Figure 6. One of the most active compounds, DMP450, is shown in the background. The green contours in Figure 6a represent favorable steric effects in those

regions; i.e., incorporation of bulkier groups most likely will increase the protease inhibitory activity. On the other hand, yellow contours indicate that bulkier groups in those regions are detrimental to activity. HINT CoMFA analysis revealed that there are overlaps between the steric and HINT hydrophobic fields. Similar findings have been reported by Oprea et al.<sup>36</sup> in a CoMFA study on other protease inhibitors. Blue contours in Figure 6b indicate that negatively charged substituents in those areas will affect the protease inhibitory activity adversely whereas negative charge in the red regions will contribute favorably toward the activity. As the conformational searches and geometry optimizations were all done within the active site of HIV-1 protease, the question arises whether these contours reflect real features of inhibitor–HIV-1 protease interactions. Close visual inspections using stereographic eye glasses reveal several important aspects of the contour plots. In Figure 6a, two of the large green contours are in the close proximity to the hydrophobic regions created by Ile50, Ile47, Val56, and Leu76. Some of the yellow contours are immediately next to the green contours, and they are closest to the residues Asp29 and Asp30, suggesting that the bulkier groups should not extend beyond the green regions. A qualitative SAR indicates that a parabolic relationship between the length of hydrophobic linear substituents and protease inhibitory activity exists for compounds **1–10**. Compound **1** with just one methyl group has the least inhibitory activity, whereas the activity reaches a peak





**Figure 6.** Stereoview of the CoMFA contour plots of (a) steric fields ( $\text{stdev} \times \text{coeff}$ ), the green polyhedra (80% contribution) represent areas where bulkier groups may enhance activity, whereas yellow polyhedra (20% contribution) indicate areas where bulk may have detrimental effects on activity; (b) electrostatic fields ( $\text{stdev} \times \text{coeff}$ ), the favored electrostatic areas with positive charges are indicated by blue polyhedra (80% contribution), whereas the favored electrostatic areas with the negative charges are indicated by red polyhedra (20% contribution). One of the most active compounds, DMP450, is shown as the reference compound.

with *n*-butyl/*n*-pentyl substituents and again decreases as the bulk (as well as length) increases. The steric contour plots are in agreement with this observation; i.e., compound **1** shows that the methyl group is further away from the green contour regions while bulkier groups should enhance activity. As the substituent bulk and length increase in compounds **2–6**, their activity increases considerably, whereas further increase of both bulk and length in compounds **7–10**, which gradually approaches the yellow regions, reduces activity considerably. For compounds **50** (DMP323) and **54** (DMP450), two of the most potent earlier candidates for clinical trials, the phenyl portions of the *p*-(hydroxymethyl)-benzyl and *m*-aminobenzyl, in the respective compounds, fill out the area of green contours adequately. The other interactions, e.g., electrostatic and hydrogen bonding, probably distinguish themselves in terms of their activities from the rest of the compounds with similar features. A CoMFA study on the training set compounds in model 3 using only an H-bond field yielded a good correlation ( $q^2 = 0.624$ ,  $s = 0.842$ ;  $r^2 = 0.803$ ,  $S = 0.610$ ). However, in Sybyl, H-bond fields are treated similarly to the indicator variable, and no attempt was made here to interpret the corresponding data in terms of hydrogen-donating or -accepting properties of the compounds used in the data set. The H-bond interactions of DMP 450 with the receptor site amino acids are well-documented and have been considered as one of the contributing factors to the protease inhibitory activity.<sup>53</sup> These visual analyses of the models provide support for their utility as predictive tools for designing new compounds with improved protease inhibitory activity.

## Conclusion

Multiple CoMFA 3D-QSAR models have been developed from a large data set (118 compounds) of cyclic urea HIV-1 protease inhibitors and validated with multiple test sets. The objective of these analyses were three-fold: (1) utilization of the wealth of knowledge on the X-ray crystal structures of HIV-1 protease–inhibitor complexes to find the most probable inhibitor conformations within the binding site of the receptor needed for optimal interactions, (2) utilization of the reported biological data for a diverse series of cyclic urea derivatives to develop predictive models to assist in designing new improved inhibitors, and (3) validation of predictive models and extraction of information from the coefficient contour plots in relation to the geometry of the protease active site side-chain amino acids responsible for molecular recognition of the inhibitors.

The value of CoMFA using X-ray structure-guided alignment rule is extraordinary as one can avoid the most controversial aspect of CoMFA study, the alignment rule. The purpose of CoMFA is not only to study the ligand–receptor interactions but also to predict quantitatively the biological activity of the compounds to be synthesized. This helps the chemist in their decision making process to prioritize the compounds to be synthesized first. It is not possible to predict the biological activity of a yet to be synthesized compound from any existing X-ray crystal structure. The models generated in this study validated the approach and not only yielded significant statistical results but also

helped in understanding the inhibitor–receptor interactions contributing to protease inhibitory activities.

The reported study provides confidence for the utility of CoMFA-based 3D-QSAR analyses, when used in concert with other relevant 2D-QSAR models and knowledge of the 3D structure of the receptor. The combined approach is expected to facilitate rational drug design.

**Acknowledgment.** This study was supported by a Grant for AIDS Research from Philip Morris Companies, Inc. The author is indebted to Dr. A. Robert Neurath for his encouragement to undertake this project and for reading the manuscript. I also thank Dr. Cathy Falk for reading the manuscript and J. Ng Pack for preparing the tables and figures.

## References

- (1) Debouck, C. The HIV-1 Protease as a Therapeutic Target for AIDS. *AIDS Res. Hum. Retroviruses* **1992**, *8*, 153–164.
- (2) McQuade, T. J.; Tomasselli, A. G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T. K.; Heinrikson, R. L.; Tarpley, W. G. A Synthetic HIV-1 Protease Inhibitor with Antiviral Activity Arrest HIV-like Particle Maturation. *Science* **1990**, *247*, 454–456.
- (3) Meek, T. D.; Lambert, D. M.; Dreyer, G. B.; Carr, T. J.; Tomaszek, T. A., Jr.; Moore, M. L.; Strickler, J. E.; Debouck, C.; Hyland, L. J.; Mathews, T. J.; et al. Inhibition of HIV-1 Protease in Infected T-lymphocytes by Synthetic Peptide Analogues. *Nature* **1990**, *343*, 90–92.
- (4) Winslow, D. L.; Otto, M. J. HIV Protease Inhibitors. *AIDS* **1995**, *Suppl. A*, S183–S192.
- (5) Prasad, J. V.; Lunney, E. A.; Para, K. S.; Tummino, P. J.; Ferguson, D.; Hupe, D.; Domagala, J. M.; Erickson, J. W. Nonpeptide Potent HIV-1 Protease Inhibitors. *Drug Des. Discov.* **1996**, *13*, 15–28.
- (6) Deeks, S. G.; Volderding, P. A. HIV-1 Protease Inhibitors. *AIDS Clin. Rev.* **1997**, 145–185.
- (7) Korant, B. D.; Rizzo, C. J. The HIV Protease and Therapies for AIDS. *Adv. Exp. Med. Biol.* **1997**, *421*, 279–284.
- (8) Deeks, S. G.; Smith, M.; Holodny, M.; Kahn, J. O. HIV-1 Protease Inhibitors: A Review for Clinicians. *J. Am. Med. Assoc.* **1997**, *277*, 145–153.
- (9) Flexner, C. HIV–Protease Inhibitors. *N. Engl. J. Med.* **1998**, *338*, 1281–1292.
- (10) Kohl, N. E.; Ermini, E. A.; Scheif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnik, 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.
- (11) Henderson, L. E.; Sowerder, R. C.; Copeland, T. D.; Oroszlan, S.; Benveniste, R. E. Gag Precursors of HIV and SIV are Cleaved into Six Proteins Found in the Mature Virions. *J. Med. Primatol.* **1990**, *19*, 411–419.
- (12) Vogt, V. M. Proteolytic Processing and Particle Maturation. *Curr. Top. Microbiol. Immunol.* **1996**, *214*, 95–132.
- (13) Luque, I.; Todd, M. J.; Gomez, J.; Semo, N.; Freire, E. Molecular Basis of Resistance to HIV-1 Protease Inhibition: A Plausible Hypothesis. *Biochemistry* **1998**, *37*, 5791–5797.
- (14) Ała, P. J.; Huston, E. E.; Klabe, R. M.; McCabe, D. D.; Duke, J. L.; Rizzo, C. J.; Korant, B. D.; DeLoskey, R. J.; Lam, P. Y. S.; Hodge, C. N.; Chang, C.-H. Molecular Basis of HIV-1 Drug Resistance: Structural Analysis of Mutant Protease Complexed With Cyclic Urea Inhibitors. *Biochemistry* **1997**, *36*, 1573–1580.
- (15) Jadav, P. K.; Ala, P.; Woerner, F. J.; Chang, C.-H.; Garber, S. S.; Anton, E. D.; Bacheler, L. T. Cyclic Urea Amides: HIV-1 Protease Inhibitors with Low Nanomolar Potency against both Wild-Type and Protease Inhibitor Resistant Mutants of HIV. *J. Med. Chem.* **1997**, *40*, 181–191.
- (16) Ridky, T.; Leis, J. Development of Drug Resistance to HIV-1 Protease Inhibitors. *J. Biol. Chem.* **1995**, *270*, 29621–29623.
- (17) Cho, S. J.; Serrano Gracia, M. L.; Bier, J.; Tropsha, A. Structure-based Alignment and Comparative Molecular Field Analysis of Acetylcholinesterase Inhibitors. *J. Med. Chem.* **1996**, *39*, 5064–5071.
- (18) Boyd, D. B. Successes of Computer-Assisted Molecular Design. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1990; pp 355–371.
- (19) 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.
- (20) Kim, K. H. Comparison of Classical and 3D QSAR. In *3D QSAR in Drug Design: Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 619–642.



- (21) Debnath, A. K.; Jiang, S.; Strick, N.; Lin, K.; Haberfield, P.; Neurath, A. R. Three-Dimensional Structure–Activity Analysis of a Series of Porphyrin Derivatives with Anti-HIV-1 Activity Targeted to the V3 Loop of the gp120 Envelope Glycoprotein of the Human Immunodeficiency Virus type 1. *J. Med. Chem.* **1994**, *37*, 1099–1108.
- (22) Martin, Y. C.; Lin, C. T.; Wu, J. Application of CoMFA to D1 Dopaminergic Agonists: A Case Study. In *3D QSAR in Drug Design: Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 643–660.
- (23) Pajeva, I.; Wiese, M. Molecular Modeling of Phenothiazines and Related Drugs As Multidrug Resistance Modifiers: A Comparative Molecular Field Analysis Study. *J. Med. Chem.* **1998**, *41*, 1815–1826.
- (24) Lieske, S. F.; Yang, B.; Eldefrawi, M. E.; MacKerell, A. D., Jr.; Wright, J. (–)-3-Beta-Substituted Ecgonine Methyl Esters As Inhibitors For Cocaine Binding and Dopamine Uptake. *J. Med. Chem.* **1998**, *41*, 864–876.
- (25) Swaan, P. W.; Szoka, F. C., Jr.; Oie, S. Molecular Modeling of The Intestinal Bile Acid Carrier: A Comparative Molecular Field Analysis Study. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 581–588.
- (26) Moro, S.; van Rhee, A. M.; Sanders, L. H.; Jacobson, K. A. Flavonoid Derivatives As Adenosine Receptor Antagonists: A Comparison of the Hypothetical Receptor Binding Site Based on A Comparative Molecular Field Analysis Model. *J. Med. Chem.* **1998**, *41*, 46–52.
- (27) Schmetzer, S.; Greenidge, P.; Kovar, K. A.; Schulze-Alexandru, M.; Folkers, G. Structure–Activity Relationships of Cannabinoids: A Joint CoMFA and Pseudoreceptor Modelling Study. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 278–292.
- (28) Kaminski, J. J.; Doweiko, A. M. Antiulcer Agents. 6. Analysis of the in vitro Biochemical and in vivo Gastric Antisecretory Activity of substituted imidazol[1,2-*a*]pyridines and Related Analogues Using Comparative Molecular Field Analysis and Hypothetical Active Site Lattice Methodologies. *J. Med. Chem.* **1997**, *40*, 427–436.
- (29) Appelt, K. Crystal Structures of HIV-1 Protease-Inhibitor Complexes. *Perspect. Drug Discov. Des.* **1993**, *1*, 23–48.
- (30) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (31) Wilkerson, W. W.; Akamike, E.; Cheatham, W. W.; Hollis, A. Y.; Collins, R. D.; DeLucca, I.; Lam, P. Y. S.; Ru, Y. HIV Protease Inhibitory Bis-benzamide Cyclic Urea: A Quantitative Structure–Activity Relationship Analysis. *J. Med. Chem.* **1996**, *39*, 4299–4312.
- (32) Wilkerson, W. W.; Dax, S.; Cheatham, W. W. Nonsymmetrically Substituted Cyclic Urea HIV Protease Inhibitors. *J. Med. Chem.* **1997**, *40*, 4079–4088.
- (33) Wei, D. T.; Meadows, J. C.; Kellogg, G. E. Effects of entropy on QSAR equations for HIV-1 protease: 1. Using hydrophobic binding descriptors. 2. Unrestrained complex structure optimization. *Med. Chem. Res.* **1997**, *7*, 259–270.
- (34) 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.
- (35) Oprea, T. I.; Waller, C. L.; Marshall, G. R. Three-Dimensional Quantitative Structure–Activity Relationship of Human Immunodeficiency Virus (I) Protease Inhibitors. 2. Predictive Power Using Limited Exploration of Alternate Binding Modes. *J. Med. Chem.* **1994**, *37*, 2206–2215.
- (36) Oprea, T. I.; Waller, C. L.; Marshall, G. R. 3D-QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. III. Interpretation of CoMFA Results. *Drug Des. Discov.* **1994**, *12*, 29–51.
- (37) Debnath, A. K. Comparative Molecular Field Analysis (CoMFA) of a Series of Symmetrical Bis-Benzamide Cyclic Urea Derivatives as HIV-1 Protease Inhibitors. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 761–767.
- (38) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; et al. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas As HIV Protease Inhibitors. *Science* **1994**, *263*, 380–384.
- (39) Nugiel, D. A.; Reese, K. Synthesis of Novel Cyclic Urea Based HIV-1 Protease Inhibitors. *Drug Des. Discov.* **1995**, *13*, 83–87.
- (40) Sham, H. L.; Zhao, C.; Stewart, K. D.; Betebenner, D. A.; Lin, S.; Park, C. H.; Kong, X. P.; Rosenbrook, W., Jr.; Herrin, T.; Madigan, D.; Vasavanonda, S.; Lyons, N.; Molla, A.; Saldivar, A.; Marsh, K. C.; McDonald, E.; Wideburg, N. E.; Denissen, J. F.; Robins, T.; Kempf, D. J.; Plattner, J. J.; Norbeck, D. W. A Novel, Picomolar Inhibitor of Human Immunodeficiency Virus Type 1 Protease. *J. Med. Chem.* **1996**, *39*, 392–397.
- (41) Hodge, C. N.; Aldrich, P. E.; Bacheler, L. T.; Chang, C. H.; Eyermann, C. J.; Garber, S.; Grubb, M.; Jackson, D. A.; Jadhav, P. K.; Korant, B.; Lam, P. Y.; Maurin, M. B.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Reid, C.; Sharpe, T. R.; Shum, L.; Winslow, D. L.; Erickson-Viitanen, S. Improved Cyclic Urea Inhibitors of The HIV-1 Protease: Synthesis, Potency, Resistance Profile, Human Pharmacokinetics and X-ray Crystal Structure of DMP 450. *Chem. Biol.* **1996**, *3*, 301–314.
- (42) Nugiel, D. A.; Jacobs, K.; Kaltenbach, R. F.; Worley, T.; Patel, M.; Meyer, D. T.; Jadhav, P. K.; De Lucca, G. V.; Smyser, T. E.; Klabe, R. M.; Bacheler, L. T.; Rayner, M. M.; Seitz, S. P. Preparation and Structure–Activity Relationship of Novel P1/P1'-Substituted Cyclic Urea-Based Human Immunodeficiency Virus Type-1 Protease Inhibitors. *J. Med. Chem.* **1996**, *39*, 2156–2169.
- (43) Lam, P. Y.; Ru, Y.; Jadhav, P. K.; Aldrich, P. E.; DeLucca, G. V.; Eyermann, C. J.; Chang, C. H.; Emmett, G.; Holler, E. R.; Daneker, W. F.; Li, L.; Confalone, P. N.; McHugh, R. J.; Han, Q.; Li, R.; Markwalder, J. A.; Seitz, S. P.; Sharpe, T. R.; Bacheler, L. T.; Rayner, M. M.; Klabe, R. M.; Shum, L.; Winslow, D. L.; Kornhauser, D. M.; Hodge, C. N.; et al. Cyclic HIV Protease Inhibitors: Synthesis, Conformational Analysis, P2/P2' Structure–Activity Relationship, and Molecular Recognition of Cyclic Ureas. *J. Med. Chem.* **1996**, *39*, 3514–3525.
- (44) Jadhav, P. K.; Ala, P.; Woerner, F. J.; Chang, C. H.; Garber, S. S.; Anton, E. D.; Bacheler, L. T. Cyclic Urea Amides: HIV-1 Protease Inhibitors With Low Nanomolar Potency Against Both Wild-Type and Protease Inhibitor Resistant Mutants of HIV. *J. Med. Chem.* **1997**, *40*, 181–191.
- (45) Nugiel, D. A.; Jacobs, K.; Cornelius, L.; Chang, C. H.; Jadhav, P. K.; Holler, E. R.; Klabe, R. M.; Bacheler, L. T.; Cordova, B.; Garber, S.; Reid, C.; Logue, K. A.; Gorey-Feret, L. J.; Lam, G. N.; Erickson-Viitanen, S.; Seitz, S. P. Improved P1/P1' Substituents For Cyclic Urea Based HIV-1 Protease Inhibitors: Synthesis, Structure–Activity Relationship, and X-ray Crystal Structure Analysis. *J. Med. Chem.* **1997**, *40*, 1465–1474.
- (46) SYBYL Molecular Modeling System (Version 6.4) is available from Tripos Associates Inc., 1699 South Hanley Rd, St. Louis, MO 63144.
- (47) HINT 2.25S is available from eduSoft, LC, P.O. Box 1811, Ashland, VA 23005.
- (48) CrystalEyes 2-Stereo Eyewear is available from StereoGraphics Corp. 2171 East Francisco Blvd, San Rafael, CA 94901.
- (49) 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.
- (50) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity – A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- (51) Marsili, M.; Gasteiger, J. Pi-Charge Distributions from Molecular Topology and Pi-Orbital Electronegativity. *Croat. Chem. Acta* **1980**, *53*, 601–614.
- (52) Dewar, M. J. S.; Zoelbis, 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.
- (53) Ala, P. J.; DeLoskey, R. J.; Huston, E. E.; Jadhav, P. K.; Lam, P. Y.; Eyermann, C. J.; Hodge, C. N.; Schadt, M. C.; Lewandowski, F. A.; Weber, P. C.; McCabe, D. D.; Duke, J. L.; Chang, C. H. Molecular Recognition of Cyclic Urea HIV-1 Protease Inhibitors. *J. Biol. Chem.* **1998**, *273*, 325–12331.
- (54) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (55) 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. Comput.* **1994**, *5*, 735–743.
- (56) Kellogg, G. E.; Semus, S.; Abraham, D. J. HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 545–552.
- (57) Folker, G.; Merz, A.; Rognan, D. CoMFA: Scope and Limitations. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 583–618.