

A Priori Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization in the Active Site

M. Katharine Holloway,^{*,†} Jenny M. Wai,[†] Thomas A. Halgren,^{*,†} Paula M. D. Fitzgerald,[‡] Joseph P. Vacca,[§] Bruce D. Dorsey,[§] Rhonda B. Levin,[§] Wayne J. Thompson,[§] L. Jenny Chen,[§] S. Jane deSolms,[§] Neil Gaffin,[§] Arun K. Ghosh,[§] Elizabeth A. Giuliani,[§] Samuel L. Graham,[§] James P. Guare,[§] Randall W. Hungate,[§] Terry A. Lyle,[§] William M. Sanders,[§] Thomas J. Tucker,[§] Mark Wiggins,[§] Catherine M. Wiscount,[§] Otto W. Woltersdorf,[§] Steven D. Young,[§] Paul L. Darke,[⊥] and Joan A. Zugay[⊥]

Departments of Molecular Systems, Biophysical Chemistry, Medicinal Chemistry, and Biological Chemistry, Merck Research Laboratories, West Point, Pennsylvania 19486, and Rahway, New Jersey 07065

Received May 31, 1994[⊗]

We have observed a high correlation between the intermolecular interaction energy (E_{inter}) calculated for HIV-1 protease inhibitor complexes and the observed *in vitro* enzyme inhibition. A training set of 33 inhibitors containing modifications in the P₁' and P₂' positions was used to develop a regression equation which relates E_{inter} and pIC₅₀. This correlation was subsequently employed to successfully predict the activity of proposed HIV-1 protease inhibitors in advance of synthesis in a structure-based design program. This included a precursor, **47**, to the current phase II clinical candidate, L-735,524 (**51**). The development of the correlation, its applications, and its limitations are discussed, and the force field (MM2X) and host molecular mechanics program (OPTIMOL) used in this work are described.

Introduction

There has been considerable recent interest¹ in the application of what has been termed rational or structure-based drug design, in which novel ligands for biological receptors are proposed and/or evaluated prior to synthesis based on structural information. This structural information can be (a) explicit as in the availability of a single crystal X-ray structure of the protein target or (b) implicit as in the availability of enough structure/activity data for a variety of ligands to develop a 3D pharmacophore map. In either case, a prediction of binding affinity for a proposed ligand can be accomplished in a qualitative or quantitative sense. Ideally, one would wish for a quantitative prediction which would also be rapid, thus avoiding the problem of having to make a "prediction" retrospectively, *i.e.*, after a putative ligand has been synthesized and assayed. We report herein the use of a simple, rapid method for predicting *a priori* the *in vitro* enzyme activity of proposed HIV-1 protease inhibitors which employs a single crystal X-ray structure of the enzyme and a limited set of structure/activity data.

The HIV-1 protease is a proteolytic enzyme which is responsible for processing the polyprotein precursors to the structural proteins and enzymes (reverse transcriptase, integrase, and the protease itself)² of the HIV-1 virus. Inactivation of the protease via site-directed mutagenesis of one of the catalytic residues has been shown to result in the production of noninfectious virions.³ Thus, the HIV-1 protease has been an attractive target for chemotherapeutic intervention in the treatment of the acquired immunodeficiency syndrome (AIDS).⁴

Due to widespread interest in developing AIDS therapies, the design of HIV-1 protease inhibitors has been

the subject of intense effort, and concomitantly the HIV-1 protease has become arguably the most studied enzyme crystallographically.⁵ Knowledge of the three-dimensional structure of the HIV-1 protease active site has led to many reports of structure-based inhibitor design in a qualitative sense⁶ but to only a few reports of a more quantitative nature which employed free-energy perturbation (FEP) methods,⁷ comparative molecular field analysis (CoMFA),⁸ or the hypothetical active site lattice (HASL) method.⁹ While FEP methods clearly have the potential for providing accurate evaluation of relative binding free energies of HIV-1 protease inhibitors, they are very expensive in a computational sense. The CoMFA and HASL approaches are more promising in terms of making rapid predictions of activity for proposed inhibitors in a rational design process. However, we have found that simple energy minimization of the proposed inhibitor in the enzyme active site leads to an intermolecular energy (E_{inter}) which correlates highly with enzyme inhibition (pIC₅₀).¹⁰

Modeling

The native,¹¹ acetylpepstatin-inhibited,¹² and L-689,502-inhibited¹³ HIV-1 protease X-ray coordinates were employed in this study. For acetylpepstatin, the enzyme coordinates matching the first orientation of the inhibitor were chosen; for L-689,502, a preliminary set of coordinates was employed which did not discriminate the two orientations of the inhibitor. Based on a pH of 5.5 for the *in vitro* enzyme assay,¹⁴ protonation states for titratable residues were chosen as follows: all titratable residues were charged with the exception of Tyr₅₉ and one of the pair of catalytic aspartic acids, Asp_{A25}. The latter protonation state was chosen based on pH rate profiles which suggest that the catalytic aspartates of (a) the fungal aspartyl proteases Penicillopepsin and *Rhizopus* Pepsin¹⁵ and (b) the HIV-1 protease¹⁶ share one negative charge. For this study the proton on the Asp_{A25} carboxyl was placed on O_{δ1} as proposed by Suguna *et al.*¹⁷ for *Rhizopus* pepsin.

[†] Department of Molecular Systems.

[‡] Department of Biophysical Chemistry.

[§] Department of Medicinal Chemistry.

[⊥] Department of Biological Chemistry.

[⊗] Abstract published in *Advance ACS Abstracts*, December 15, 1994.

For convenience in cross-site comparisons, the two inhibited enzyme structures were aligned onto the native enzyme using the backbone atoms of the catalytic triad, Asp_{A25}-Thr_{A26}-Gly_{A27} and Asp_{B225}-Thr_{B226}-Gly_{B227} (Asp_{B25}-Thr_{B26}-Gly_{B27} in the native enzyme). All crystallographic waters were removed with the exception, in the two inhibited structures, of the tightly bound water (305 in acetylpepstatin and 407 in L-689,502) which is involved in hydrogen bonding with the NH's of the flap residues Ile_{A50} and Ile_{B250} and the P₂ and P₁'¹⁸ carbonyl oxygens of the inhibitors described herein. In the acetylpepstatin-inhibited enzyme, four other waters (303, 304, 319, and 324), which are located in crevices in the floor of the active site, were also included. No significant difference in results was observed when these were included, so they were omitted from the L-689,502-inhibited enzyme.

A model of **1** was constructed in the Merck molecular modeling program AMF¹⁹ based on the X-ray structures of inhibitors of endothiapepsin²⁰ and *Rhizopus* pepsin.¹⁷ Models of all other inhibitors employed the model of **1** as a template. All inhibitor models were neutral (no ionic charge). The flexibility of each inhibitor was manually explored as necessary to obtain a satisfactory fit in the enzyme active site, which also corresponded to a low-energy conformer.

Inhibitor models were minimized in the three enzyme active sites using the MM2X force field implemented in the program OPTIMOL (see below). In all cases, the inhibitor was completely flexible while the enzyme was completely static. Where different favorable conformational or orientational possibilities existed, the final inhibitor model was chosen on the basis of the lowest total energy, *i.e.* a balance between favorable intermolecular and intramolecular energies. Final cartesian coordinates for the training set of inhibitors (**1**, **3**–**34**) as minimized in the L-689,502-inhibited enzyme are provided as supplementary material.

The MM2X force field was developed at the Merck Research Laboratories as an extension of the MM2 force field; it differs from MM2 principally in that lone pairs on heteroatoms are not used and in that electrostatic interactions take place between atom-centered charges, allowing proper treatment of charged systems. For intramolecular vdW interactions, the default choice for MM2X (and the choice made in this work) is to use the MM2 exp-6 potential; for intermolecular interactions, MM2X by default uses the Lennard-Jones 9–6 form (the 12–6 form can also be selected). Electrostatic interactions are calculated from Coulomb's law using atom-centered charges, obtained for intramolecular interactions from MM2's bond-dipole parameters in most cases, and for intermolecular interactions obtained from a companion set of "bond charge increments" chosen to reproduce electrostatic-potential-derived charges. Dielectric constants of 1.5 for intramolecular interactions and 1.0 for intermolecular interactions are normally used. MM2X, the resident force field in OPTIMOL when the work reported in this paper was done, has been parameterized for a wide range of functional groups but shares many parameters with MM2. A more complete description of MM2X and of its implementation in OPTIMOL is given in appendix A, and a complete list of MM2X parameters is included in the supplementary material. The molecular modeling pro-

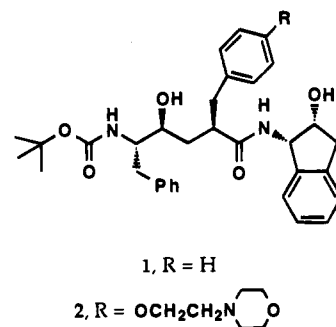
gram OPTIMOL was developed by one of us (T.A.H.) and colleagues at the Merck Research Laboratories.

Graphics visualization was performed using AMF, Quanta,²¹ and *c_view*.²² Correlation plots were produced using Kaleidagraph;²³ however, *R*, *R*², and cross-validated *R*² values for the correlation plots were calculated using RS/1.²⁴ The cross-validated *R*² values were computed using a leave-one-out approach.²⁵

Results and Discussion

Like other known aspartyl protease inhibitors, HIV-1 protease inhibitors have generally been designed to mimic the tetrahedral intermediate formed during hydrolysis of the scissile amide bond of the substrate. The HIV-1 protease inhibitor **1**, containing the hydroxyethylene isostere as tetrahedral intermediate mimic, was optimized from an initial renin inhibitor lead²⁶ and, as a small subnanomolar inhibitor, served as a convenient starting point for molecular modeling studies.

A model of **1** was developed as described above and is shown in the native enzyme active site in Figure 1. The binding elements (P₂, P₁, P₁', and P₂') are labeled using the Schechter and Berger nomenclature for proteases.¹⁸ Use of this model of **1** led to the successful qualitative design of an improved inhibitor, L-689,502, **2**,¹³ which capitalized on the solvent accessibility of para substituents on the P₁' benzyl.



Our success with such qualitative predictions led us to try a more quantitative approach, *i.e.*, examining correlations between calculated properties and *in vitro* enzyme inhibition (IC₅₀). On the basis of previous promising results²⁷ using a human renin model to aid the design of renin inhibitors, we focused on a possible correlation between the calculated interaction energy (E_{inter}), *i.e.*, the intermolecular component of the total energy, and the observed IC₅₀. Here, E_{inter} corresponds to the sum of the van der Waals (E_{vdw}) and electrostatic (E_{elec}) interactions between the inhibitor and the enzyme when the inhibitor is minimized in the rigid enzyme active site. It does not include the intramolecular energy of the inhibitor, the only other variable component of the total energy in this case.

$$E_{\text{inter}} = E_{\text{vdw}} + E_{\text{elec}}$$

The proposed correlation was premised on two assumptions: (1) that E_{inter} might be proportional to the enthalpy of binding (ΔH_{bind}) and (2) that the entropy of binding (ΔS_{bind}) might be small or more likely constant. Thus, E_{inter} might be proportional to the free energy of binding (ΔG_{bind}) and correlate with the experimental

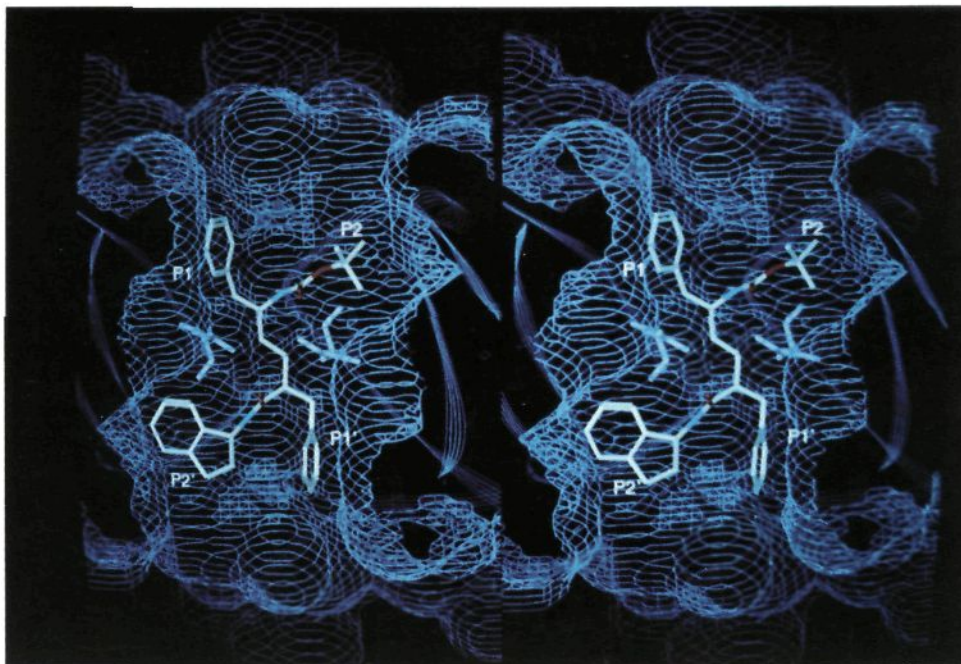


Figure 1. Stereoview of L-685,434, **1**, as modeled in the active site of the native enzyme. The active site cavity is represented by a wire-mesh molecular surface and the enzyme is depicted as a ribbon. The catalytic aspartates (solid blue sticks), Asp_{A25} and Asp_{B25}, are highlighted for reference. The inhibitor binding elements are labeled following the Schechter and Berger nomenclature for proteases.¹⁸

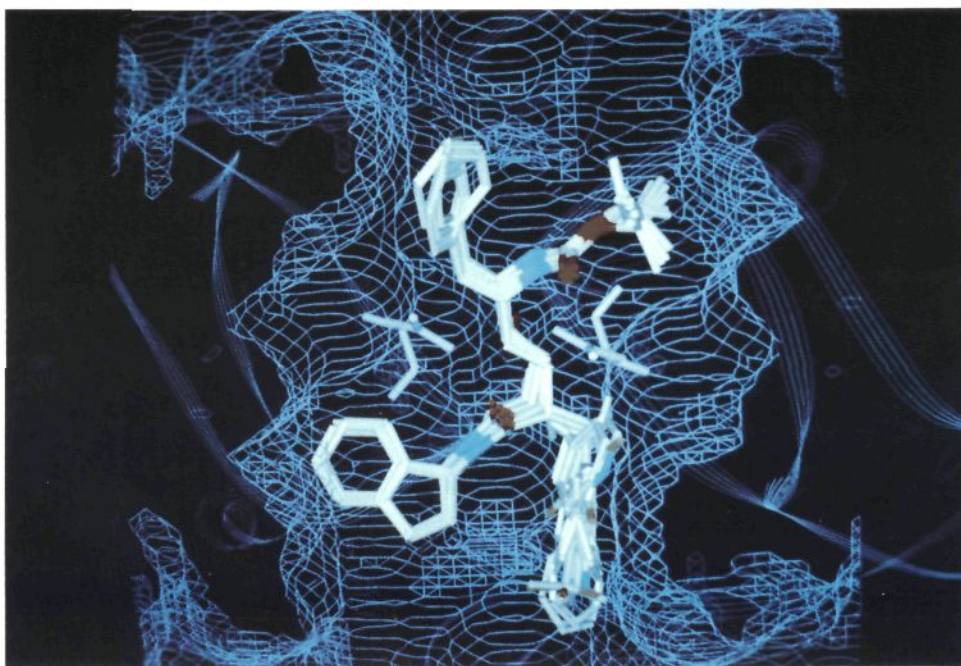


Figure 2. Overlaid models of the training set inhibitors **1** and **3–18**, which contain modifications of the P_{1'} substituent, illustrated in the native enzyme active site. The active site cavity is represented by a wire-mesh molecular surface and the enzyme is depicted as a ribbon. The catalytic aspartates (solid blue sticks), Asp_{A25} and Asp_{B25}, are highlighted for reference.

observation, *i.e.*, IC₅₀.

$$\Delta G_{\text{bind}} = \Delta H_{\text{bind}} - T\Delta S_{\text{bind}}$$

The inhibitors that were chosen for the training set are shown in Table 1. The space they occupy in the active site is illustrated in Figures 2 and 3. They are all related to **1** and were selected for variety of structure (16 inhibitors with modifications in the P_{1'} site and 16 inhibitors with modifications in the P_{2'} site) and for

spread in IC₅₀ (5 orders of magnitude). Initially, these inhibitors were modeled in the native enzyme active site, but subsequently as the acetylpepstatin¹¹ and L-689,502¹² inhibited enzyme active sites became available, these were used to repeat the correlation calculations. As shown in Figure 4, a good correlation was observed between E_{inter} and pIC₅₀ for all three enzyme active sites. Equations 1–3 were derived for the native, acetylpepstatin, and L-689,502 enzyme active sites,

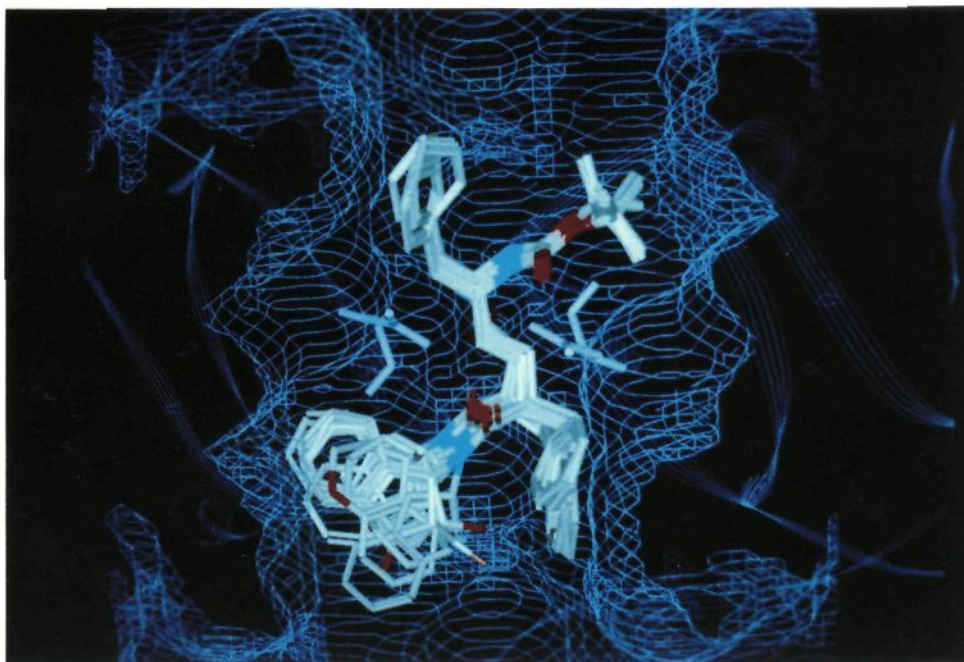


Figure 3. Overlaid models of the training set inhibitors **19–34**, which contain modifications of the P₂' substituent, illustrated in the native enzyme active site. The active site cavity is represented by a wire-mesh molecular surface and the enzyme is depicted as a ribbon. The catalytic aspartates (solid blue sticks), Asp_{A25} and Asp_{B25}, are highlighted for reference.

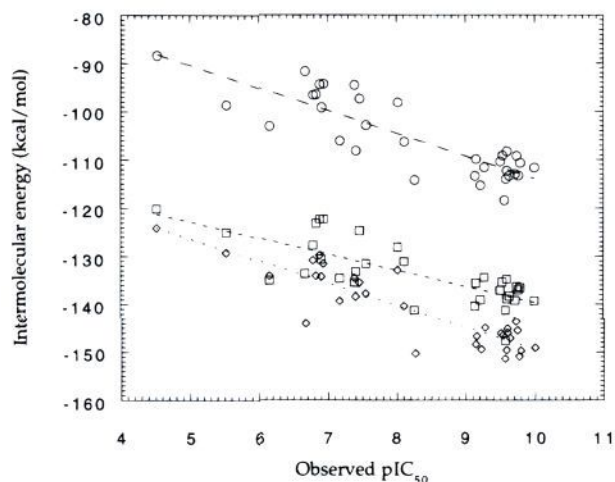


Figure 4. Plot of calculated enzyme–inhibitor interaction energy vs experimental enzyme inhibition (pIC₅₀) for the training set of inhibitors **1, 3–34** (circles = native enzyme active site, squares = acetylpepstatin-inhibited active site, diamonds = L-689,502-inhibited active site).

respectively. The corresponding R , R^2 , and cross-validated R^2 values are given following each equation.
native:

$$\text{pIC}_{50} = -0.15435(E_{\text{inter}}) - 8.069 \quad (1)$$

$$R = 0.8524, \quad R^2 = 0.7265, \quad \text{cross-validated } R^2 = 0.6910$$

acetylpepstatin inhibited:

$$\text{pIC}_{50} = -0.17302(E_{\text{inter}}) - 14.901 \quad (2)$$

$$R = 0.7623, \quad R^2 = 0.5811, \quad \text{cross-validated } R^2 = 0.5244$$

L-689,502 inhibited:

$$\text{pIC}_{50} = -0.16946(E_{\text{inter}}) - 15.707 \quad (3)$$

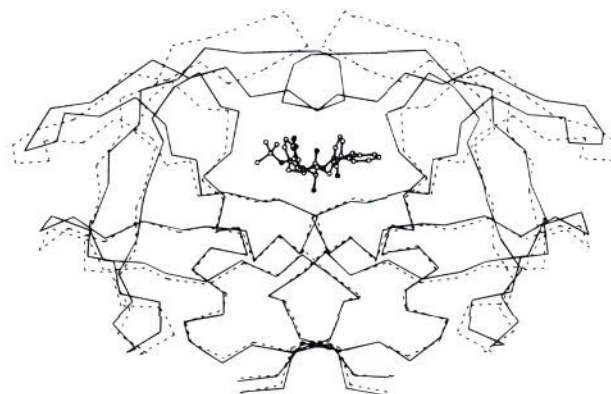

















Figure 5. A comparison of the X-ray coordinates of the native enzyme (dashed) and the complex (solid) with L-689,502, **2** (ball and stick). The two enzymes are represented as C_α traces to illustrate the difference in position of the flaps (at the top of the figure) which in the native enzyme X-ray structure are too distant to interact with the inhibitor.

$$R = 0.8852, \quad R^2 = 0.7835, \quad \text{cross-validated } R^2 = 0.7551$$

In all three cases the R^2 and cross-validated R^2 values are comparable, indicating that no one data point is overly influential in deriving the correlation equation; thus, the derived model should be predictive.

Note that the correlation in eq 1 compares favorably with that in eq 3. This is somewhat surprising since in the native enzyme the flaps which contribute to binding in the S₂, S₁, S₁', and S₂' pockets are involved in crystal packing interactions and are not within reach of the modeled inhibitor, as illustrated in Figure 5. The correlation thus seems to be relatively independent of the position of the flaps. This may be explained by two factors: (1) there are specific hydrogen bonds between the inhibitor and the floor of the active site which force the inhibitor to adopt an appropriate bioactive confor-

Table 1. Calculated Enzyme–Inhibitor Intermolecular Energies and Experimental IC₅₀ Values for the Training Set of HIV-1 Protease Inhibitors


| No | R ₁ | R ₂ | R ₃ | R ₄ | E _{Native} ^a kcal/mol | E _{AcPep} ^b kcal/mol | E ₅₀₂ ^c kcal/mol | IC ₅₀ (nM) | pIC ₅₀ ^d |
|----|--|-----------------|----------------|---|--|---|---|--------------------------|--------------------------------|
| 1 | CH ₂ Ph | H | H | | -108.3 | -134.8 | -145.1 | 0.25 | 9.6021 |
| 3 | CH ₂ Ph | CH ₃ | H | | -106.3 | -131.1 | -140.4 | 7.7 | 8.1135 |
| 4 | CH ₂ CH ₂ CH ₂ Ph | H | OH | | -113.0 | -139.2 | -143.6 | 0.19 | 9.7212 |
| 5 | CH ₂ -4-CF ₃ Ph | H | H | | -114.0 | -141.3 | -149.6 | 0.26 | 9.5850 |
| 6 | (E)-CH ₂ CH=CHPh | H | H | | -113.4 | -138.3 | -147.1 | 0.23 | 9.6383 |
| 7 | CH ₂ C ₆ F ₅ | H | H | | -115.3 | -139.1 | -149.4 | 0.6 | 9.2218 |
| 8 | CH ₂ -4-CH ₃ Ph | H | H | | -109.1 | -135.4 | -146.5 | 0.29 | 9.5376 |
| 9 | CH ₂ -4-NH ₂ Ph | H | H | | -110.4 | -137.1 | -146.1 | 0.31 | 9.5086 |
| 10 | CH ₂ -4-NO ₂ Ph | H | H | | -118.4 | -147.7 | -151.4 | 0.27 | 9.5686 |
| 11 | H | H | H | | -98.6 | -125.0 | -129.2 | 2934 | 5.5325 |
| 12 | CH ₂ -4-OHPh | H | H | | -110.7 | -136.6 | -149.7 | 0.16 | 9.7959 |
| 13 | CH ₂ CH=CH ₂ | H | H | | -102.9 | -131.5 | -137.8 | 27.5 | 7.5607 |
| 14 | CH ₂ -4-IPh | H | H | | -113.4 | -140.4 | -148.4 | 0.72 | 9.1427 |
| 15 | CH ₂ C(O)Ph | H | H | | -114.2 | -141.3 | -150.3 | 5.42 | 8.2660 |
| 16 | CH ₂ -4-pyridyl | H | H | | -111.6 | -134.4 | -144.9 | 0.53 | 9.2757 |
| 17 | CH ₂ SPh | H | H | | -112.3 | -138.9 | -146.0 | 0.25 | 9.6021 |
| 18 | CH ₂ -4-t-butylPh | H | H | | -113.3 | -137.0 | -150.9 | 0.17 | 9.7696 |
| 19 | | | | NHCH ₂ Ph | -94.3 | -122.3 | -131.5 | 114 | 6.9431 |
| 20 | | | |  | -98.2 | -128.1 | -132.9 | 9.53 | 8.0209 |
| 21 | | | |  | -97.4 | -124.7 | -135.5 | 34.25 | 7.4653 |
| 22 | | | |  | -103.1 | -135.0 | -134.1 | 690 | 6.1612 |
| 23 | | | |  | -96.6 | -127.7 | -130.8 | 161 | 6.7932 |
| 24 | | | |  | -106.1 | -134.6 | -139.3 | 66.3 | 7.1785 |
| 25 | | | |  | -91.7 | -133.6 | -144.0 | 212.42 | 6.6728 |
| 26 | | | |  | -99.2 | -130.6 | -134.2 | 121.8 | 6.9144 |
| 27 | | | |  | -109.9 | -135.6 | -146.7 | 0.7 | 9.1549 |
| 28 | | | |  | -109.2 | -136.4 | -145.5 | 0.18 | 9.7447 |
| 29 | | | |  | -94.5 | -135.5 | -134.5 | 40.5 | 7.3925 |
| 30 | | | | NHCH ₂ C ₆ H ₁₁ | -88.4 | -120.1 | -124.1 | 30000 | 4.5229 |
| 31 | | | |  | -94.4 | -122.3 | -129.8 | 130 | 6.8861 |
| 32 | | | |  | -96.5 | -123.2 | -134.1 | 146 | 6.8356 |
| 33 | | | |  | -111.7 | -139.3 | -149.1 | 0.1 | 10.000 |
| 34 | | | |  | -108.1 | -133.2 | -138.4 | 38.6 | 7.4134 |

^a Intermolecular energy calculated in the native HIV-1 protease active site. ^b Intermolecular energy calculated in the Acetylpepstatin inhibited HIV-1 protease active site. ^c Intermolecular energy calculated in the L-689,502 inhibited HIV-1 protease active site. ^d pIC₅₀ = -log (IC₅₀).

mation in the absence of the flaps and (2) the binding elements contributed by the flaps are relatively constant for this series of inhibitors. However, the best correla-

tion is obtained in eq 3, with the L-689,502-inhibited enzyme, in which the flaps are positioned to interact with the inhibitor models and involved in binding of an

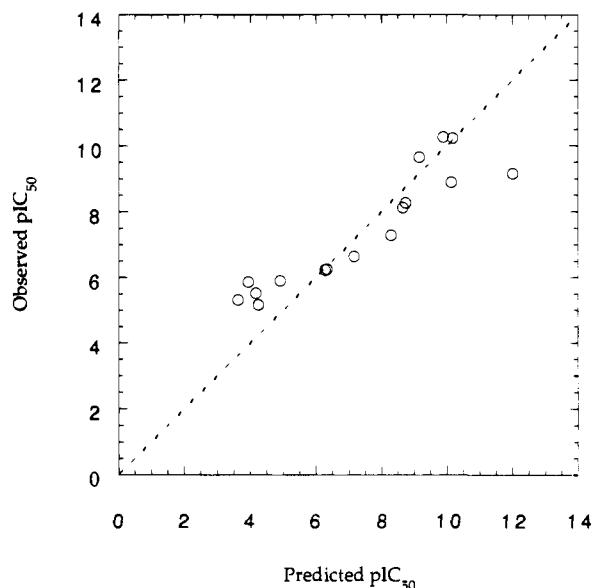


Figure 6. Plot of predicted pIC_{50} vs. observed pIC_{50} values for the predicted set of inhibitors **35**–**50**. The line is one of unit slope, *i.e.*, predicted pIC_{50} = observed pIC_{50} .

inhibitor, **2**, which is structurally comparable to those in the training set.

Using eqs 1–3, we were able to make predictions of activity for many proposed inhibitors prior to synthesis, *i.e.*, true predictions and not *post hoc* explanations of activity. Since this was done over a period of time, the earliest predictions of activity were made using eq 1, while the later predictions were made using either eq 2 or 3. We have chosen to report only the original prediction for the compounds listed in Table 2 in order to illustrate that useful predictions of activity were made prior to synthesis using this simple model. Thus, only two compounds are reported which were predicted with eq 1 and only three compounds are reported which were predicted with eq 2. This is due to a variety of factors, including (a) the availability of the L-689,502-inhibited (**2**) HIV-1 protease X-ray coordinates in a timely manner, (b) the natural lag between prediction and synthesis which favored prediction with the newer model, and (c) the fact that in many cases the exact compound which was predicted was not synthesized, but rather a structural analog.

The relative accuracy of these predictions is illustrated graphically in Figure 6, in which the line is one of unit slope, *i.e.*, predicted pIC_{50} = observed pIC_{50} . There is only one significant outlier, **36**, on the far right hand side of the graph, which will be discussed in more detail below. The average absolute error in the predicted values in Table 2 is 1.01 log units across a range of 5.10 log units. When **36** is omitted, the average absolute error drops to 0.79 log units, or less than a factor of 10.

The compounds listed in Table 2 illustrate a natural progression of activity predictions for this program and span a range of tetrahedral intermediate mimics, *e.g.*, constrained hydroxyethylene (**35**, **38**, and **40**); statine (**37**); hydroxyethylamine (**41**–**44**, **46**–**49**); and symmetrical hydroxyethylene (**36**), hydroxyethylamine (**45**), and diol (**50**) isosteres. Each was designed to answer specific structural questions or to address pharmacokinetic properties. However, only a few of the predictions of activity will be discussed in greater detail in order to illustrate the scope and the limitations of this simple approach.

As mentioned above, the significant outlier in Figure 6 is **36**, which was designed as a symmetrical version of **1**. Like many others,²⁸ we believed that a symmetrical inhibitor might bind more tightly to the symmetrical active site of the HIV-1 protease. The activity of **36** was predicted using eq 1. In this case, it was clear even prior to the enzyme assay of **36** that the predicted activity was substantially exaggerated, *i.e.*, it was predicted to have an IC_{50} of 0.0009 nM, while the subsequently determined IC_{50} was 0.69 nM. At the time, we were concerned that our prediction was in error due to differences in binding between our model and experiment. However, when the X-ray structure of **36** was solved,²⁹ there was good agreement between the modeled and X-ray structures as illustrated in Figure 7. There are several other possible explanations for the overprediction of activity: (1) the use of the native enzyme model in eq 1, rather than one of the inhibited enzyme models in eqs 2 or 3; (2) the presence of an additional hydrogen bond to the active site which would be overemphasized in a gas-phase molecular mechanics calculation; or (3) the existence of a higher barrier to obtaining the bioactive conformation necessary for binding. Of the three possibilities, the last two seem most

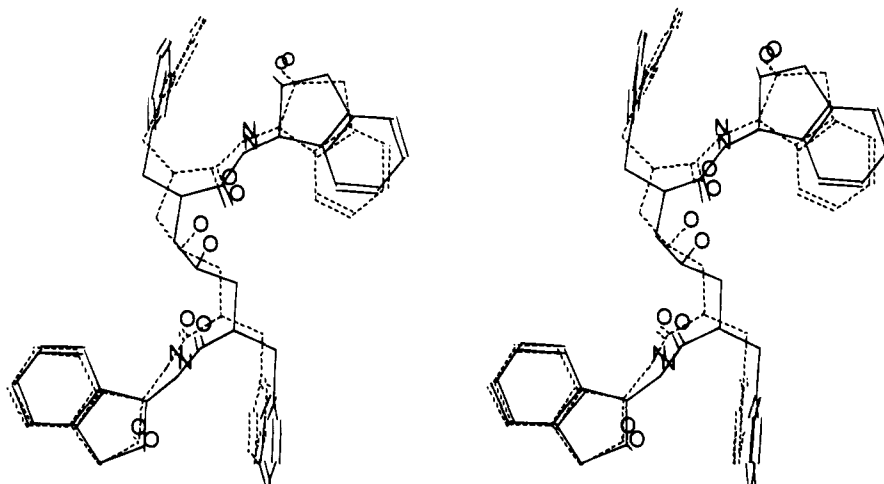


Figure 7. A comparison of the modeled (dashed) and X-ray (solid) structures of **36**.

Table 2. Calculated Enzyme-Inhibitor Intermolecular Energy and Predicted and Observed pIC₅₀ Values for the Predicted Set of HIV-1 Protease Inhibitors

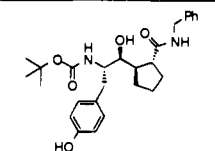
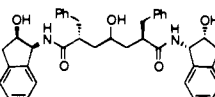
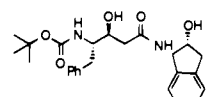
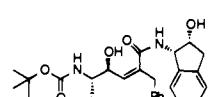
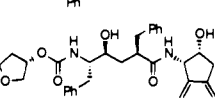
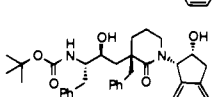
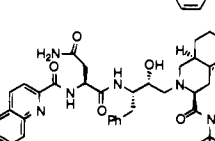
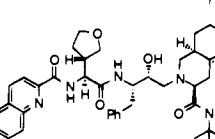
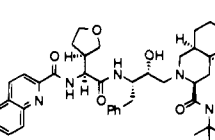
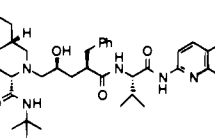
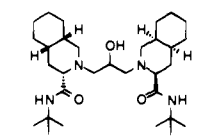
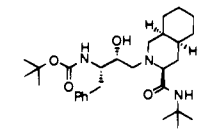
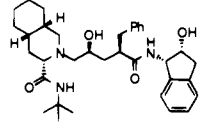
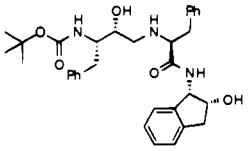
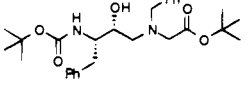
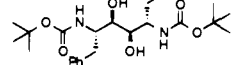
| No | Structure | E _{inter} ^a kcal/mol | Active Site ^b | Predicted pIC ₅₀ | Observed pIC ₅₀ ^c |
|----|---|---|--------------------------|--------------------------------|--|
| 35 |  | -96.4 | Native | 6.2760 | 6.2299 |
| 36 |  | -123.4 | Native | 12.012 | 9.1612 |
| 37 |  | -127.3 | AcPepstatin | 6.3302 | 6.2457 |
| 38 |  | -140.1 | AcPepstatin | 10.141 | 8.8861 |
| 39 |  | -140.2 | AcPepstatin | 10.171 | 10.222 |
| 40 |  | -125.9 | L-689,502 | 4.9099 | 5.8965 |
| 41 |  | -145.5 | L-689,502 | 9.1493 | 9.6383 |
| 42 |  | -143.6 | L-689,502 | 8.7383 | 8.2676 |
| 43 |  | -148.9 | L-689,502 | 9.8847 | 10.2676 |
| 44 |  | -141.6 | L-689,502 | 8.3058 ^d | 7.2774 |
| 45 |  | -122.8 | L-689,502 | 4.2394 | 5.1675 |
| 46 |  | -122.5 | L-689,502 | 4.1745 | 5.5229 |
| 47 |  | -143.2 | L-689,502 | 8.6518 | 8.1163 |

Table 2 (Continued)

| No | Structure | E_{inter}^a kcal/mol | Active Site ^b | Predicted pIC_{50} | Observed pIC_{50}^c |
|----|---|----------------------------------|--------------------------|--------------------------------|---------------------------------|
| 48 |  | -136.3 | L-689,502 | 7.1594 | 6.6402 |
| 49 |  | -120.0 | L-689,502 | 3.6338 | 5.3279 |
| 50 |  | -121.4 | L-689,502 | 3.9366 | 5.8617 |

^a Calculated intermolecular energy in the indicated active site. ^b Active site in which minimization was performed and whose correlation equation was used for prediction. ^c $\text{pIC}_{50} = -\log(\text{IC}_{50})$. ^d The prediction of activity was originally made for a closely related compound; the "predicted" activity reported here was calculated at a later date for the compound which was actually synthesized, **44**.

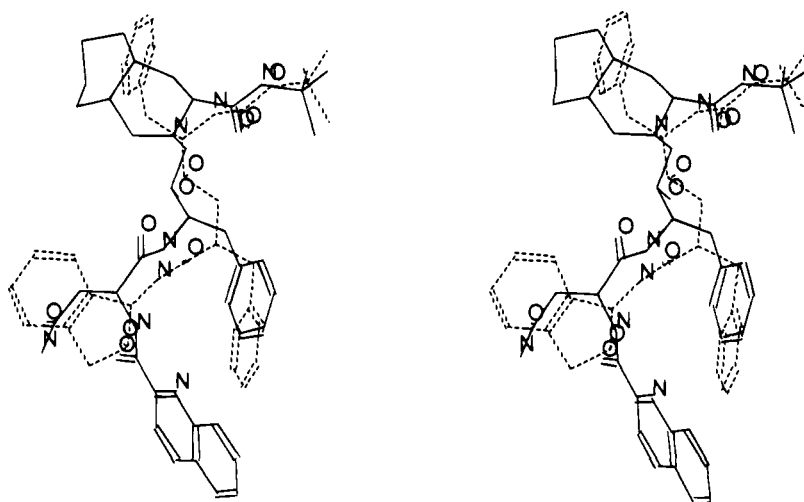


Figure 8. An overlay of the model of **1** oriented in an N→C fashion (dashed) and the X-ray structure of **41** oriented in a C→N fashion (solid).

likely since the activity predicted using the other two models is also exaggerated. In particular, factor (3) may play a key role, since **36** experienced a much larger decrease (13.7 kcal/mol) than other inhibitors (2.5 kcal/mol for **2**) in its intramolecular energy when minimized outside of the active site, an indication that the bound conformation may be significantly higher in energy than the global minimum.

The outlier **36** clearly demonstrates the limitations of this simple model. However, in the best case, we were able to accurately predict the relative and absolute activities of **42** (predicted $\text{IC}_{50} = 1.8$ nM, observed $\text{IC}_{50} = 5.4$ nM) and **43** (predicted $\text{IC}_{50} = 0.13$ nM, observed $\text{IC}_{50} = 0.05$ nM), which differ only in the stereochemistry of the furan ring in P_2 ,³⁰ a novel surrogate for the Asn residue of Ro-31-8959, **41**.³¹ We were also able to predict that the 6-membered lactam ring in **40** would fit poorly (predicted $\text{IC}_{50} = 12.3$ μM , observed $\text{IC}_{50} = 1.269$ μM) in the active site, although the analogous 5-membered lactam was a nanomolar inhibitor ($\text{IC}_{50} = 37$ nM).³²

A last example, **47**, serves to showcase the utility of this model in a structure-based drug design program in which crystallographers, molecular modelers, and medicinal chemists work closely together. An interesting feature of HIV-1 protease complexes with inhibitors

is that, due to the symmetrical nature of the enzyme, some inhibitors are observed^{12,13,28e,29,33} to bind in the active site in two directions, both in an N→C and a C→N orientation with respect to the flaps, which in their closed H-bonded form introduce the asymmetry which is the direction marker. Comparing models of **41**, oriented in the N→C and C→N fashions in the active site, led to the design of an active "reversed" Roche analog, **44**, in which all the binding elements remain the same, but the amide bond directionality is reversed. Comparing **1** oriented in an N→C fashion and **41** oriented in a C→N fashion, as illustrated in Figure 8, led to the hypothesis that novel hybrid inhibitors such as **47** could be developed which incorporated the P_1' and P_2' binding groups from both **1** and **41**.

As predicted, **47** was an active HIV-1 protease inhibitor (predicted $\text{IC}_{50} = 2.23$ nM, observed $\text{IC}_{50} = 7.65$ nM) and via excellent medicinal chemistry³⁴ it led to the current clinical candidate, L-735,524 (**51**).³⁵ Figure 9 illustrates a comparison between the models of **47**, **51**, and the X-ray structure³⁴ of a related compound, **52**, as complexed to the HIV-1 protease.

Conclusion

In summary, we have observed a good correlation between a simple calculated property, E_{inter} , and the

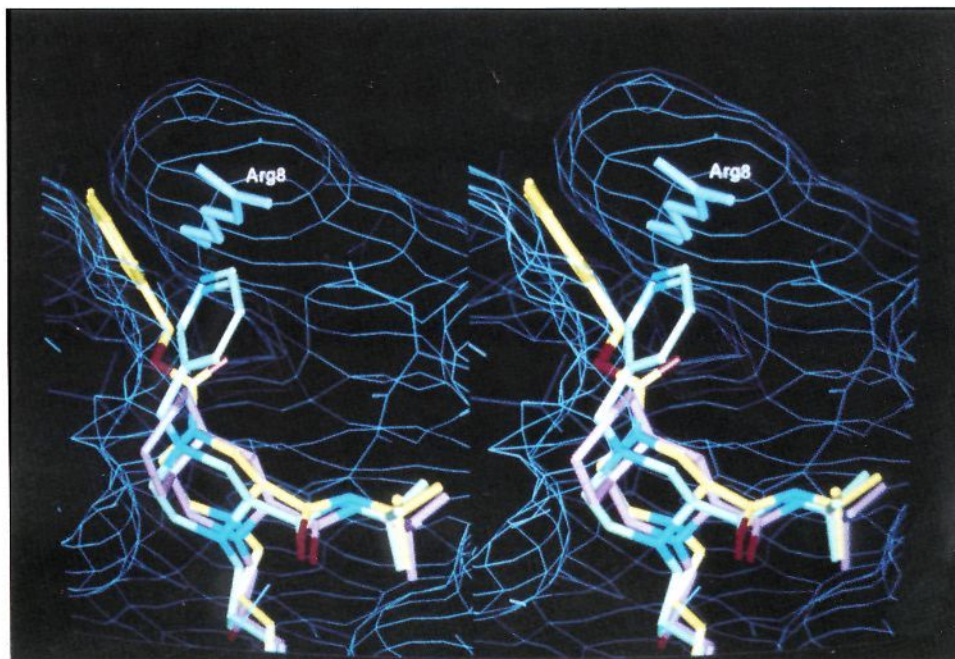
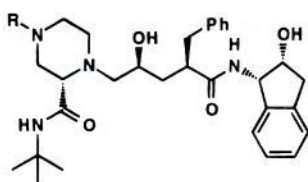


Figure 9. A comparison of the models of **47** and **51** and the X-ray structure of a related compound, **52**, as complexed to the HIV-1 protease. The side chain of Arg₈ (solid blue) contacts the aromatic ring of the Cbz substituent in **52** and may be involved in a specific hydrogen bond with the pyridyl nitrogen of **51**.



51, R = 3-pyridylmethyl
52, R = benzyloxycarbonyl

observed IC_{50} for a series of HIV-1 protease inhibitors which differed in their P_1' and P_2' groups. This correlation, in particular eq 3, was subsequently used to successfully predict the activities of proposed HIV-1 protease inhibitors. This included inhibitors which contain modifications in P_2 , P_1 , P_1' , and P_2' , as well as different backbone structures, *e.g.*, hydroxyethylamine, statine, and symmetrical diol isosteres, in addition to the hydroxyethylene isostere in the test dataset.

It must be emphasized that we made many more predictions of activity than are reported here. Happily, one reason for the difference is that frequently when a prediction was unfavorable, the compound was not synthesized. In addition, in many cases when a prediction was favorable, a close analog, rather than the exact compound which was modeled, was synthesized.

Although this simple model is easy, rapid, and has proven successful in aiding the structure-based drug design process, it has clearly been limited in this study to inhibitors which are neutral and approximately the same size. However, preliminary work indicates that a successful correlation may also be derived employing charged inhibitors, *e.g.*, those containing basic amines, or water-displacing templates, *e.g.*, the DuPont Merck cyclic ureas,³⁶ as well as other classes of inhibitors.

This model obviously neglects some factors which are key to binding, *e.g.*, (1) the flexibility of the enzyme active site which can adjust to fit different inhibitor

structures; (2) the difference in energy between the solution and bound conformations of the inhibitor, which may lead to an initial energy barrier that must be surmounted by the inhibitor prior to binding; and (3) the solvation/desolvation of the inhibitor and the enzyme, which may oppose or enhance the enzyme-inhibitor interaction. In at least one case which we have illustrated, **36**, one or more of these factors must be important for the accurate prediction of activity. However, attempts to incorporate these effects into a predictive model have been thus far unsuccessful in improving the observed correlation. We continue to examine the effect of these factors on our correlation in the hope of deriving an even more accurate model which still allows for timely prediction of activity.

Acknowledgment. We would like to acknowledge R. B. Nachbar for providing the RS/1 procedure to calculate cross-validated R^2 values and R. B. Nachbar, D. J. Underwood, and G. M. Smith for contributions related to the development of the MM2X parameter set. We also thank B. L. Bush, E. F. Fluder, R. B. Nachbar, and G. M. Smith for providing some of the code used in the implementation of MM2X in OPTIMOL.

Supplementary Material Available: Cartesian coordinates for the training set of inhibitors (**1**, **3–34**) and MM2X force field parameters (112 pages). Ordering information is given on any current masthead page.

Appendix A. The MM2X Force Field and the Molecular Mechanics Program OPTIMOL

1. The MM2X Energy Expression. The MM2X energy expression can be written as

$$E_{\text{MM2X}} = \sum EB_{ij} + \sum EA_{ijk} + \sum EBA_{ijk} + \sum EOO_{ijk;l} + \sum ET_{ijkl} + \sum EvdW_{ij} + \sum EQ_{ij} \quad (\text{A1})$$

where the constituent terms, each expressed in kilocalories per mole, are defined as shown below.

1.1. Bond Stretching. Like MM2,³⁷ MM2X employs the cubic function

$$EB_{ij} = 143.88 \frac{kb_{IJ}}{2} \Delta r_{ij}^2 (1 + cs \Delta r_{ij}), \quad (A2)$$

where kb_{IJ} is the force constant in md/Å, $\Delta r_{ij} = r_{ij} - r_{IJ}^0$ is the difference in angstroms between actual and reference bond lengths, and $cs = -2 \text{ Å}^{-1}$ is the "cubic stretch" constant. For large Δr_{ij} , as is well-known, this function diverges to $-\infty$. Accordingly, the MM2X implementation in OPTIMOL removes all cubic contributions whenever the quadratic contribution for any particular bond exceeds a preset maximum of 3 kcal/mol. A special reference bond length (1.47 Å) and force constant (5.0 md/Å) is used when $i-j$ is formally a single bond between sp^2 -hybridized carbon atoms. In this and the following subsections, the use of capital letters as subscripts denotes a quantity which depends on the MM2X atom types (e.g., r_{IJ}^0) rather than on the atomic coordinates (e.g., r_{ij}).

1.2. Angle Bending. MM2X uses the MM2 expansion

$$EA_{ijk} = 0.043 \ 828 \frac{ka_{IJK}}{2} \Delta \vartheta_{ijk}^2 (1 + sb \Delta \vartheta_{ijk}^4) \quad (A3)$$

where ka_{IJK} is the force constant in md Å/rad², $\Delta \vartheta_{ijk} = \vartheta_{ijk} - \vartheta_{IJK}^0$ is the difference between actual and reference bond angles in degrees, and $sb = 7 \times 10^{-8} \text{ deg}^{-4}$ is the "sextic-bend" constant. As noted in the supplementary material, special sets of parameters are used when the three atoms comprise a 3-membered ring or lie in a 4-membered ring.

1.3. Stretch-Bend Interactions. MM2X employs the same form as is used in MM2, namely

$$EBA_{ijk} = 2.511 \ 18 (kba_{IJK} \Delta r_{ij} + kba_{KJI} \Delta r_{kj}) \Delta \vartheta_{ijk} \quad (A4)$$

where kba_{IJK} and kba_{KJI} are force constants in md/rad which couple the $i-j$ and $k-j$ stretches to the $i-j-k$ bend, and Δr and $\Delta \vartheta$ are as defined above. The MM2 parameters and rules governing when stretch-bend terms are excluded (e.g., when i or k is a hydrogen atom) are used.

1.4. Out-of-Plane Bending at Tricoordinate Centers. MM2X uses the form

$$EOOP_{ijk;l} = 0.043 \ 828 \frac{koop_{IJK;L}}{2} \chi_{ijk;l}^2 (1 + sb \chi_{ijk;l}^4) \quad (A5)$$

where $koop_{IJK;L}$ is the force constant in md Å/rad², $\chi_{ijk;l}$ is the Allinger (MM2) angle in degrees between the bond $k-l$ and the plane $i-j-l$ and k is the central atom,³⁸ and sb is as in eq A3. All three out-of-plane angle components (all permutations of base atoms i, j , and l) are included in the energy expression and share the same value for the koop force constant. For atom types having out-of-plane interactions, the "in-plane" angles to the central atom k used in eq A5 are the projections on the $i-j-l$ base plane, as in MM2.

1.5. Torsion Interactions. MM2X uses the 3-fold representation employed in MM2, where Φ is the $i-j-$

$k-l$ dihedral angle:

$$ET_{ijkl} = 0.5(V_1(1 + \cos \Phi) + V_2(1 - \cos 2\Phi) + V_3(1 + \cos 3\Phi)) \quad (A6)$$

Torsion interactions are defined for atom quartets i, j, k , and l , where $i-j, j-k$ and $k-l$ are bonded pairs and $i \neq l$. As noted in the supplementary material, separate sets of torsion parameters are used when one of the three bonds is formally a single bond between sp^2 -hybridized carbon atoms or when the four atoms comprise a 4-membered ring. The parameters V_1, V_2 , and V_3 depend on the MM2X atom types I, J, K , and L .

1.6. Van der Waals Interactions. In a departure from most force field programs, the implementation of MM2X in OPTIMOL recognizes both what it calls a "subject molecule" (SM) and what it calls a "context molecule" (CM). All terms in the MM2X energy expression apply to a SM, but a CM, when supplied (e.g., an enzyme active site), is held rigid and interacts with the SM (e.g., bound ligand) only through vdW and electrostatic nonbonded interactions. For vdW interactions between pairs of SM atoms, MM2X employs the MM2 exp-6 representation

$$E_{vdw_{ij}} = \sum \epsilon_{IJ} \{ 290 \ 000 \exp(-12.5 R_{ij}/R_{IJ}^*) - 2.25 (R_{IJ}^*/R_{ij})^2 \} \quad (A7)$$

where ϵ_{IJ} and R_{IJ}^* depend on the MM2X atom types I and J for the interacting atoms i and j . Like MM2, MM2X computes the minimum-energy separation R_{IJ}^* as the sum of the vdW radii R_I^* and R_J^* , computes ϵ_{IJ} as the geometric mean of ϵ_{II} and ϵ_{JJ} , excludes 1,2- and 1,3-interactions, foreshortens C-H bonds by a factor of 0.915, and uses MM2's special ϵ_{IJ} and R_{IJ}^* values for interactions between aliphatic carbon and hydrogen.

For interactions between SM and CM atoms, in contrast, MM2X by default uses the Lennard-Jones 9-6 form

$$E_{vdw_{ij}} = \sum \epsilon_{IJ} \{ 2(R_{IJ}^*/R_{ij})^9 - 3(R_{IJ}^*/R_{ij})^6 \} \quad (A8)$$

and employs different sets of well depths and vdW radii (but uses the same combination rules for ϵ_{IJ} and R_{IJ}^*). Alternatively, the Lennard-Jones 12-6 form can be selected. In addition, MM2X as implemented in OPTIMOL can also use a Lennard-Jones form rather than the exp-6 form of eq A7 for intramolecular interactions between pairs of SM atoms.

1.7. Electrostatic Interactions. In place of MM2's dipole-dipole interaction energy, MM2X uses the simple Coulombic form

$$EQ_{ij} = 332.0538 q_i q_j / (DR_{ij}), \quad (A9)$$

to facilitate the description of systems containing non-zero formal or net ionic charges. Also unlike MM2, MM2X makes no use of lone pairs on heteroatoms. In this expression, q_i and q_j are partial atomic charges, R_{ij} is the internuclear separation in angstroms, and D is the "dielectric constant" (normally taken as constant $D = 1.5$, though use of a distance-dependent dielectric constant is also supported). For interactions between pairs of SM atoms, the partial atomic charges q_i (SM \rightarrow SM) are constructed from initial full or fractional formal

atomic charges (usually zero, but, *e.g.*, -0.5 for carboxylate oxygens) by adding contributions from bond dipole increments δ_{IK} , which describe the polarity of the bonds to atom i from attached atoms k . Specifically, MM2X computes $q_i(\text{SM} \rightarrow \text{SM})$ as

$$q_i(\text{SM} \rightarrow \text{SM}) = q_i^0 + \sum \delta_{IK} \quad (\text{A10})$$

where $\delta_{IK} = -\delta_{KI}$. As usual, 1,2- and 1,3-interactions are excluded. The bond dipole charges δ_{IK} are themselves computed from stored MM2-like "bond dipoles" μ_{IK} as

$$\delta_{IK} = -\mu_{IK}/(4.803r_{IK}^0) \quad (\text{A11})$$

where μ_{IK} is in Debye, r_{IK}^0 is in angstroms, and the sign change reflects the difference between the MM2 and MM2X sign conventions; for MM2X, δ_{IK} with $I < K$ is the charge contributed to atom k of atom type K from the $i-k$ bond by, and at the expense of, the atom i of atom type I (thus, $\delta_{KI} = -\delta_{IK}$). For example, a C=O bond (atom types 3 for C and 7 for O) has a bond dipole moment $\mu_{3,7}$ of 2.95 D (vs 2.6 D in MM2 proper), which with $r_{3,7}^0 = 1.229 \text{ \AA}$ yields $\delta_{3,7} = -0.50$, thus placing a charge of -0.50 on O and contributing a charge of $+0.50$ toward the final partial atomic charge on C. Although not true in this case, in many other cases the MM2X and MM2 bond dipole moments are the same.

For interactions between SM and CM atoms, the requisite partial atomic charges are computed as

$$q_i(\text{SM} \rightarrow \text{CM}) = q_i^0 + \sum \omega_{IK} \quad (\text{A12})$$

from a separately defined set of bond charge increments ω_{IK} . These bond charge increments were chosen in so far as possible to make the resultant partial atomic charges consistent in value and pattern with the electrostatic potential fit charges derived by Cox and Williams.³⁹ This use of different electrostatic and vdW parameters for interactions between SM and CM atoms represented an attempt to combine a soundly based intramolecular force field patterned after MM2 with an intermolecular component which could describe intermolecular interactions more accurately than we felt MM2 could. As specified in the supplementary material, the partial atomic charges $q_i(\text{CM} \rightarrow \text{SM})$ and vdW parameters for the CM itself are assigned by table look up from recognized atom and residue names. In general, however, these parameters are similar, if not identical, to those that would be assigned as SM \rightarrow CM parameters for a SM of similar constitution.

2. Implementation of MM2X in OPTIMOL. OPTIMOL, the host molecular-mechanics platform for MM2X, has been widely used in molecular-modeling applications at Merck for nearly 10 years. We have already noted that OPTIMOL treats "subject molecules" both in isolation and in the context of an enzyme active site. In this section, we summarize some other features and discuss some pertinent elements related to the implementation of MM2X.

2.1. Major OPTIMOL Features. OPTIMOL provides a variety of techniques for energy minimization, including full or selective optimization in either Cartesian or torsion space using a BFGS⁴⁰ variable-metric algorithm. Rigid-body optimization with or without simultaneous torsional optimization is also supported.

The latter capabilities are useful in docking a ligand to a receptor such as an enzyme site. Unlike the "subject molecule" with which OPTIMOL is primarily concerned, an enzyme site is represented without hydrogens on carbon atoms and is held rigid. Enzyme sites are also employed in creating "hydrophobic", "hydrophilic", and other *active-site maps* for three-dimensional color graphics display. These active-site maps depict regions in the *accessible space* for ligand binding considered to be appropriate for occupancy by portions of the ligand having matching physiochemical properties.⁴¹ As such, these maps provide qualitative guidance for inhibitor docking and for inhibitor design.⁴² OPTIMOL also provides limited conformational search capabilities. One or two specified torsion angles can be incremented through a specified range of values, either in "scan" mode (rigid rotation) or in "search" mode (with simultaneous optimization of other degrees of freedom, subject, if desired, to imposed penalty-function restraints on specified internal coordinates). More comprehensive conformational searches normally are conducted by employing OPTIMOL to refine a series of conformers generated using the distance-geometry programs DGEOM⁴³ or JIGGLE⁴⁴ and by using various procedures to group the optimized conformers into geometrically related families. OPTIMOL also provides a number of analysis capabilities, including a geometric analysis of hydrogen bonds and an energy decomposition analysis of the energy of the subject molecule or of its interaction with the enzyme site. Also included is an "Analyze Differences" facility which allows contributing force-field interaction energies to be compared for two related sets of Cartesian coordinates (*e.g.*, coordinates obtained before and after energy minimization or ligand docking).

2.2. Implementation of MM2X. OPTIMOL has been designed to be as easy as possible for both computational chemists and laboratory scientists to use. The user (or the invoking modeling platform) simply represents the subject molecule to OPTIMOL in language familiar to the organic chemist, *i.e.*, as a collection of atoms joined by single, double, or triple bonds, some atoms of which may have a non-zero formal charge.⁴⁵ Aromatic systems may be supplied in any constituent Kekule form. Thus, the user is *not* required to designate atom types, specify hybridization, detect aromaticity, or assign partial atomic charges. Rather, OPTIMOL uses the supplied structural information to generate all additional information needed to carry out the calculation. It automatically "sets up" the calculation by determining the torsional "tree structure", perceiving and classifying rings, defining symbolic atom types based on local connectivity, detecting aromaticity, and creating appropriate lists of bond, angle, and torsional interactions. As previously described,⁴⁶ the symbolic atom types are then translated into the numeric values used to assign force-field parameters to the force-field interaction terms.

In establishing the relationship between parameters and force-field interactions, the parameter files, which are kept in "canonical order" based on indices derived from the numerical atom types, are processed using a rapid binary search algorithm. If present, the *fully qualified* parameter corresponding to the precise set of atom types supplied is used. For vdW, bond stretching, stretch-bend interaction, bond charge increment, and

bond dipole parameters, no *equivalences* are recognized. For angle bending, out-of-plane bending, and torsion interactions, however, whenever the fully qualified parameter is not found, OPTIMOL executes a staged "step down" procedure in which increasingly generic values are sought. This protocol is governed by the entries in the MM2XDEF.PAR parameter file, where the "Level 1" atom types define the fully qualified parameters. Entries from Levels 2–5 are employed as needed in subsequent searches. Those at Level 5, always "0", serve as *wild cards*. Such wild card values are used only for "wing atoms" in an angle bending or torsional interaction or for non-central atoms in an out-of-plane interaction. Level 4 generally corresponds to the atomic species, Level 3 to atomic species plus hybridization. Currently, the first two levels employ identical numerical atom types in nearly all instances. The protocol used in the step down procedure depends on the interaction type (angle, torsion, out-of-plane).⁴⁷ Finally, if no parameter is found, a default rule may be invoked. This staged-search/default-rule procedure allows applications to go forward when specific parameters are unavailable, though inevitably with a loss in reliability.

The MM2X parameter files and information regarding their usage in OPTIMOL may be found in the supplementary material.

References

- (1) (a) Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. Structure-Based Molecular Design. *Acc. Chem. Res.* **1994**, *27*, 117–123. (b) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *J. Med. Chem.* **1994**, *37*, 1035–1054. (c) Bugg, C. E.; Carson, W. M.; Montgomery, J. A. Drugs by Design. *Sci. Am.* **1993**, December, 92–98.
- (2) Farmerie, W. G.; Loeb, D. D.; Casavant, N. C.; Hutchinson, C. A., III; Edgel, M. H.; Swanstrom, R. Expression and Processing of the AIDS Virus Reverse Transcriptase in *Escherichia coli*. *Science* **1987**, *236*, 305–308.
- (3) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. Active Human Immunodeficiency Virus is Required for Viral Infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686.
- (4) Huff, J. R. HIV Protease: A Novel Chemotherapeutic Target for AIDS. *J. Med. Chem.* **1991**, *34*, 2305–2314.
- (5) It was recently estimated that as many as 160 HIV-1 protease and HIV-1 protease/inhibitor complex structures have been solved thus far. Wlodawer, A.; Erickson, J. W. Structure-Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543–85.
- (6) For a recent review of structure-based design of HIV protease inhibitors, see Clare, M. HIV Protease: Structure-Based Design. *Perspect. Drug Discovery Des.* **1993**, *1*, 49–68.
- (7) (a) Reddy, M. R.; Viswanadhan, 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. (b) 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. (c) Cieplak, P.; Kollman, P. A. Peptide Mimetics as Enzyme Inhibitors: Use of Free Energy Perturbation Calculations To Evaluate Isosteric Replacement for Amide Bonds in a Potent HIV Protease Inhibitor. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 291–304.
- (8) (a) 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. (b) Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-Dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. I. A CoMFA Study Employing Experimentally-Determined Alignment Rules. *J. Med. Chem.* **1993**, *36*, 4152.
- (9) Dowejko, A. M. Three-Dimensional Pharmacophores from Binding Data. *J. Med. Chem.* **1994**, *37*, 1769–1778.
- (10) One other group has employed molecular mechanics minimization in the enzyme active site for a set of only three inhibitors and found no correlation between activity and energy. See: Sansom, C. E.; Wu, J.; Weber, I. T. Molecular Mechanics Analysis of Inhibitor Binding to HIV-1 Protease. *Protein Eng.* **1992**, *5*, 659–667.
- (11) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. Three-dimensional Structure of Aspartyl Protease from Human Immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615–620.
- (12) Fitzgerald, P. M. D.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C.-T.; Herber, W. K.; Dixon, R. A. F.; Darke, P. L. Crystallographic Analysis of a Complex between Human Immunodeficiency Virus Type 1 Protease and Acetyl-Pepstatin at 2.0-Å Resolution. *J. Biol. Chem.* **1990**, *265*, 14209–14219.
- (13) Thompson, W. J.; Fitzgerald, P. M. D.; Holloway, M. K.; Emini, E. A.; Darke, P. L.; McKeever, B. M.; Schleif, W. A.; Quintero, J. C.; Zugay, J. A.; Tucker, T. J.; Schering, 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.
- (14) Heimbach, J. C.; Garsky, V. M.; Michelson, S. R.; Dixon, R. A. F.; Sigal, I. S.; Darke, P. L. Affinity Purification of the HIV-1 Protease. *Biochem. Biophys. Res. Commun.* **1989**, *164*, 955–960.
- (15) Hofmann, T.; Hodges, R. S.; James, M. N. G. Effect of pH on the Activities of Penicillopepsin and Rhizopus Pepsin and a Proposal for the Productive Substrate Binding Mode in Penicillopepsin. *Biochemistry* **1984**, *23*, 635–643.
- (16) Hyland, L. J.; Tomaszek, T. A., Jr.; Meek, T. D. Human Immunodeficiency Virus-1 Protease. 2. Use of pH Rate Studies and Solvent Kinetic Isotope Effects to Elucidate Details of Chemical Mechanism. *Biochemistry* **1991**, *30*, 8454–8463.
- (17) Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; Davies, D. R. Binding of a Reduced Peptide Inhibitor to the Aspartic Proteinase from *Rhizopus chinensis*: Implications for a Mechanism of Action. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7009–7013.
- (18) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
- (19) AMF, the Advanced Modeling Facility, is an extension of work described previously: (a) Gund, P.; Andose, J. D.; Rhodes, J. B.; Smith, B. M. Three-dimensional Molecular Modeling and Drug Design. *Science* **1980**, *208*, 1425–31. (b) Smith, G. M.; Hangauer, D. G.; Andose, J. D.; Bush, B. L.; Fluder, E. M.; Gund, P.; McIntyre, E. F. Intermolecular Modeling Methods in Drug Design: Modeling the Mechanism of Peptide Cleavage by Thermolysin. *Drug Inf. J.* **1984**, *18*, 167–78. The major authors of the AMF program are J. D. Andose, R. A. Blevins, E. F. Fluder, and J. Shpungin.
- (20) Founding, S. I.; Cooper, J.; Watson, F. E.; Cleasby, A.; Pearl, L. H.; Sibanda, B. L.; Hemmings, A.; Wood, S. P.; Blundell, T. L.; Valler, M. J.; Norey, C. G.; Kay, J.; Boger, J.; Dunn, B. M.; Leckie, B. J.; Jones, D. M.; Atrash, B.; Hallet, A.; Szelke, M. High Resolution X-ray Analyses of Renin Inhibitor-Aspartic Proteinase Complexes. *Nature* **1987**, *327*, 349–352.
- (21) Molecular Simulations Inc., 16 New England Executive Park, Burlington, MA 01803-5297.
- (22) J. Christopher Culberson, Merck Research Laboratories, unpublished work.
- (23) Synergy Software, 2457 Perkiomen Ave., Reading, PA 19606.
- (24) BBN Software Products Corporation, 10 Fawcett St., Cambridge, MA 02238.
- (25) (a) Efron, B.; Tibshirani, R. Statistical Data Analysis in the Computer Age. *Science* **1991**, *253*, 390–394. (b) Efron, B.; Gong, G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *Am. Stat.* **1983**, *37*, 36–48.
- (26) Lyle, T. A.; Wiscourt, C. M.; Guare, J. P.; Thompson, W. J.; Anderson, P. S.; Darke, P. L.; Zugay, J. A.; Emini, E. A.; Schleif, W. A.; Quintero, J. C.; Dixon, R. A. F.; Sigal, I. S.; Huff, J. R. Benzocycloalkyl Amines as Novel C-Terminal HIV Protease Inhibitors. *J. Med. Chem.* **1991**, *34*, 1228–1230.
- (27) (a) Williams, P. D.; Perlow, D. S.; Payne, L. S.; Holloway, M. K.; Siegl, P. K. S.; Schorn, T. W.; Lynch, R. J.; Doyle, J. J.; Strouse, J. F.; Vlasuk, G. P.; Hoogsteen, K.; Springer, J. P.; Bush, B. L.; Halgren, T. A.; Richards, A. D.; Kay, J.; Veber, D. F. Renin Inhibitors Containing Conformationally Restricted P₁-P₁' Dipeptide Mimetics. *J. Med. Chem.* **1991**, *34*, 887–900. (b) Williams, P. D.; Payne, L. S.; Perlow, D. S.; Holloway, M. K.; Siegl, P. K. S.; Schorn, T. W.; Lynch, R. J.; Doyle, J. J.; Strouse, J. F.; Vlasuk, G. P.; Hoogsteen, K.; Springer, J. P.; Bush, B. L.; Halgren, T. A.; tenBroeke, J.; Greenlee, W. J.; Richards, A. D.; Kay, J.; Veber, D. F. Design of Renin Inhibitors Containing Conformationally Restricted Mimetics of the P₁-P₁' and P₁ through P₂' Sites. In *Structure and Function of Aspartic Proteinases*; Dunn, B. M., Ed.; Plenum Press: New York, 1991; pp 325–334.

- (28) (a) Erickson, J.; Neidhart, D. J.; VanDrie, J.; Kempf, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Widburg, N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M. Design, Activity, and 2.8 Å Crystal Structure of a C₂ Symmetric Inhibitor Complexed to HIV-1 Protease. *Science* **1990**, *249*, 527–533. (b) Humber, D. C.; Cammack, N.; Coates, J. A. V.; Copley, K. N.; Orr, D. C.; Storer, R.; Weingarten, G. G.; Weir, M. P. Penicillin Derived C₂-Symmetric Dimers as Novel Inhibitors of HIV-1 Proteinase. *J. Med. Chem.* **1992**, *35*, 3080–3081. (c) Jadhav, P. K.; Woerner, F. J. Synthesis of C₂-Symmetric HIV-1 Protease Inhibitors from D-Mannitol. *Bioorg. Med. Chem. Letters* **1992**, *2*, 353–356. (d) Spaltenstein, A.; Leban, J. J.; Furfine, E. S. Synthesis of C₂-Symmetric HIV-Protease Inhibitors with Sulfur-containing Central Units. *Tetrahedron Lett.* **1993**, *34*, 1457–1460. (e) Dreyer, G. B.; Boehm, J. C.; Chenera, B.; DesJarlais, R. L.; Hassell, A. M.; Meek, T. D.; Tomaszek, T. A., Jr. A Symmetric Inhibitor Binds HIV-1 Protease Asymmetrically. *Biochemistry* **1993**, *32*, 937–947.
- (29) Bone, R. F.; Vacca, J. P.; Anderson, P. S.; Holloway, M. K. X-ray Crystal Structure of the HIV Protease Complex with L-700,417, an Inhibitor with Pseudo C₂ Symmetry. *J. Am. Chem. Soc.* **1991**, *113*, 9382–9384.
- (30) (a) Thompson, W. J.; Ghosh, A. K.; Holloway, M. K.; Lee, H.-Y.; Munson, P. M.; Schwering, J. E.; Wai, J.; Darke, P. L.; Zugay, J.; Emini, E. A.; Schleif, W. A.; Huff, J. R.; Anderson, P. S. 3'-Tetrahydrofuranlyglycine as a Novel, Unnatural Amino Acid Surrogate for Asparagine in the Design of Inhibitors of the HIV Protease. *J. Am. Chem. Soc.* **1993**, *115*, 801–803. (b) Ghosh, A. K.; Thompson, W. J.; Holloway, M. K.; McKee, S. P.; Duong, T. T.; Lee, H.-Y.; Munson, P. M.; Smith, A. M.; Wai, J. M.; Darke, P. L.; Zugay, J. A.; Emini, E. A.; Schleif, W. A.; Huff, J. R.; Anderson, P. S. Potent HIV Protease Inhibitors: The Development of Tetrahydrofuranlyglycines as Novel P₂-Ligands and Pyrazine Amides as P₃-Ligands. *J. Med. Chem.* **1993**, *36*, 2300–2310.
- (31) 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*, 358–61.
- (32) Vacca, J. P.; Fitzgerald, P. M. D.; Holloway, M. K.; Hungate, R. W.; Starbuck, K. E.; Chen, L. J.; Darke, P. L.; Anderson, P. S.; Huff, J. R. Conformationally Constrained HIV-1 Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 499–504.
- (33) (a) Fitzgerald, P. M. D.; McKee, B. M.; VanMiddlesworth, J. F.; Springer, J. P. Binding of a Reduced-Peptide Inhibitor and a Statine-containing Inhibitor to the Protease from the Human Immunodeficiency Virus. In *Advances in Molecular Biology and Targeted Treatment for AIDS*; Kumar, A., Ed.; Plenum Press: New York, 1991; pp 245–249. (b) Dreyer, G. B.; Lambert, D. M.; Meek, T. D.; Carr, T. J.; Tomaszek Jr., T. A.; Fernandez, A. V.; Bartus, H.; Cacciavillani, E.; Hassel, A. M.; Minnich, M.; Petteway, S. R., Jr.; Metcalf, B. W.; Lewis, M. Hydroxyethylene Isostere Inhibitors of Human Immunodeficiency Virus-1 Protease: Structure-Activity Analysis Using Enzyme Kinetics, X-ray Crystallography, and Infected T-cell Assays. *Biochemistry* **1992**, *31*, 6646. (c) Murthy, K. H. M.; Winborne, E. L.; Minnich, M. D.; Culp, J. S.; Debouck, C. The Crystal Structure at 2.2 Å Resolution of Hydroxyethylene-based Inhibitors Bound to Human Immunodeficiency Virus Type 1 Protease Show That the Inhibitors Are Present in Two Distinct Orientations. *J. Biol. Chem.* **1992**, *267*, 22770–22778.
- (34) (a) Dorsey, B. D.; Levin, R. B.; McDaniel, S. L.; Vacca, J. P.; Darke, P. L.; Zugay, J. A.; Emini, E. A.; Schleif, W. A.; Quintero, J. C.; Lin, J. L.; Chen, I. W.; Ostovic, D.; Fitzgerald, P. M. D.; Holloway, M. K.; Anderson, P. S.; Huff, J. R. L-735,524: The Rational Design of a Potent and Orally Bioavailable HIV Protease Inhibitor, 206th ACS National Meeting, Chicago, IL, August 22–27, 1993. (b) Dorsey, B. D.; Levin, R. B.; McDaniel, S. L.; Vacca, J. P.; Guare, J. P.; Darke, P. L.; Zugay, J. A.; Emini, E. A.; Schleif, W. A.; Quintero, J. C.; Lin, J. H.; Chen, I.-W.; Holloway, M. K.; Fitzgerald, P. M. D.; Axel, M. G.; Ostovic, D.; Anderson, P. S.; Huff, J. R. L-735,524: The Design of a Potent and Orally Bioavailable HIV Protease Inhibitor. *J. Med. Chem.* **1994**, *37*, 3443–3451.
- (35) Vacca, J. P.; Dorsey, B. D.; Schleif, W. A.; Levin, R. B.; McDaniel, S. L.; Darke, P. L.; Zugay, J.; Quintero, J. C.; Blahey, O. M.; Roth, E.; Sardana, V. V.; Schlabach, A. J.; Condra, J. H.; Gotlib, L.; Holloway, M. K.; Lin, J.; Chen, I.-W.; Ostovic, D.; Anderson, P. S.; Emini, E. A.; Huff, J. R. L-735,524: An Orally Bioavailable HIV-1 Protease Inhibitor. *Proc. Nat. Acad. Sci. U.S.A.* **1994**, *91*, 4096–4100.
- (36) The DuPont-Merck inhibitor DMP 323 is predicted using eq 3 and the L-689,502-inhibited active site *sans* water to have a pIC₅₀ value of 9.6977 versus an observed pIC₅₀ value of 8.4437, as measured under the assay conditions described in ref 14. Additionally, in a more even comparison, a strong correlation ($R = 0.88357$) is observed between the E_{inter} calculated using the L-689,502-inhibited active site *sans* water and the reported IC₅₀ values for the four cyclic urea inhibitors contained in Table I of Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bachelier, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C.-H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors. *Science* **1994**, *263*, 380–384.
- (37) (a) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127. (b) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
- (38) Allinger, N. L.; Tribble, M. T.; Miller, M. A. An Improved Force Field for the Calculation of Structures and Energies of Carbonyl Compounds. *Tetrahedron* **1972**, *28*, 1173–1190.
- (39) Cox, S. R.; Williams, D. E. Representation of the Molecular Electrostatic Potential by a Net Atomic Charge Model. *J. Comput. Chem.* **1981**, *3*, 304–323.
- (40) Fletcher, R. *Practical Methods of Optimization: Unconstrained Optimization*; Wiley: New York, 1980; Vol. 1, Chapter 3.
- (41) Halgren, T. A.; Bush, B. L. Unpublished research; cf. ref 42.
- (42) Weber, A. E.; Halgren, T. A.; Doyle, J. J.; Lynch, R. J.; Siegl, P. K. S.; Parsons, W. H.; Greenlee, W. J.; Patchett, A. A. Design and Synthesis of P₂-P₁'-Linked Macrocyclic Human Renin Inhibitors. *J. Med. Chem.* **1991**, *34*, 2692–2701.
- (43) DGEOM (Andose, J. D. Unpublished research) is based on subroutine EMBEDD written by G. M. Crippen; cf.: (a) Crippen, G. M., *Distance Geometry and Conformational Calculations*; John Wiley and Sons: New York, 1981. (b) Crippen, G. M.; Havel, T. F. *Acta Crystallogr. A* **1987**, *34*, 282.
- (44) JIGGLE is a multifaceted distance-geometry based facility developed at the Merck Research Laboratories by S. K. Kearsley.
- (45) Ordinarily, the molecular data structure is read from an external file or is created using graphical sketching and other tools provided by AMF (cf. ref 19).
- (46) Halgren, T. A. Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.
- (47) For bending of the $i-j-k$ angle, a five-stage process based on the level combinations 1-1-1, 2-2-2, 3-2-3, 4-2-4, and 5-2-5 is used. For $i-j-k-l$ torsion interactions, a five-stage process based on level combinations 1-1-1-1, 2-2-2-2, 3-2-2-5, 5-2-2-3, and 5-2-2-5 is used, where stages 3 and 4 correspond to "half default" or "half wild card" entries. For out-of-plane bending ijk/l , where j is the central atom (cf. eq 5), the three-stage protocol 1-1-1; 1, 2-2-2; 2, 5-2-5; 5 is used. The final stage provides wild-card defaults for all except the central atom.