Three-Dimensional Pharmacophores from Binding Data

Arthur M. Doweyko†

CIBA-GEIGY Corporation, Environmental Health Center, Farmington, Connecticut 06032

Received February 2, 1994®

The application of HASL (hypothetical active site lattice) methodology has been successfully extended to generate putative pharmacophoric patterns in three dimensions capable of quantitatively predicting binding activity. The transformation of a HASL model to a pharmacophore is illustrated using pK_i values published for 84 HIV-1 protease inhibitors. Starting with a HASL model generated at 2.00 Å and containing 899 lattice points, a selective trimming process was used to identify significant lattice points. In this manner, a set of 11 points was found which represents a potential pharmacophoric pattern and predicts the pK_i activity of the 84-inhibitor set with a correlation (r^2) of 0.827. Furthermore, the locations of these points were found to coincide with a number of strategic binding areas within the known active site structure HIV-1 protease, thus providing a physical confirmation of their relevancy.

Introduction

A pharmacophore has been defined as a "...minimum collection of atoms spatially disposed in a manner that elicits a biological response."[1-3] This definition has been refined to include topographic or three-dimensional constraints[4,5] of such sets of atoms. Following advances in computationally driven approaches, the pharmacophoric paradigm has been expanded to more effectively accommodate drug design strategies and now includes the "active analog" approach,[6-8] inferred interaction sites (GRID),[9] ensemble distance geometry,[10] and 3D-QSAR (CoMFA, HASL).[11-15] These and other methods correlating three-dimensional structure to activity have been recently reviewed.[16-19]

The present investigation was undertaken to determine whether pharmacologically significant points in three-dimensional space could be located using the hypothetical active site lattice (HASL) 3D-QSAR methodology.[13-15] The first step in the HASL technique converts molecular structures to regularly spaced sets of points (lattices), with each point retaining a fourth-dimensional variable describing the type of atom present at that point. Next, the molecular lattices are merged to form a composite description of the occupied space. Critical to this undertaking is the incorporation of partial activity values associated with each point in such a way that the summation of values for any one set of points, corresponding to a particular molecule, yields the overall activity for that molecule. This approach is particularly suitable for activity values which are linearly proportional to the free energy of binding, e.g., $-\log(K_i)$ or pK_i. Thus, the resulting HASL represents a quantitative and predictive three-dimensional binding model for a given set of molecules. It was found that a further extension of the HASL method, which relies on the systematic removal of points with negligible partial binding values, can result in a relatively small subset of significant points. These HASL-derived points constitute a pharmacophoric pattern capable of quantitatively predicting inhibitor binding.

It was of interest to investigate the HASL methodology and the above-described extension as a platform for pharmacophore building on a relatively well-studied enzyme system having some degree of relevance in medicinal research. Such a system is represented by HIV-1 protease (HIV-1 PR). HIV-1 PR is considered a critical link in the activation of human immunodeficiency virus, the causative agent in AIDS. The structure and inhibition of the enzyme have been reviewed extensively.[20-25] HASL model building was conducted using the HIV-1 PR inhibition data obtained from a variety of published sources.[25-44]

Methods

Hypothetical active site lattice methodology has been successfully used to model the interactions of small molecules with proteins,[13-15] including an inverted application wherein DNA structure–reactivity relationships were modeled for uracil mustard.[45] The strengths and limitations of the HASL approach to 3D-QSAR have been reviewed.[14,16] Since molecular structures are converted to regularly spaced sets of points, the method is particularly sensitive to the conformation of molecules used as templates during superpositioning and the choice of resolution or lattice point spacing in model building. Candidate template molecules can be those having restricted rotational degrees of freedom or, more preferably, those molecules whose bound active site conformation is known. In order to avoid overfitting the structure–activity data, lattice spacing (resolution) just small enough to represent most atoms was used.

Pharmacophore building was approached by first creating a 3D-QSAR HASL model and then trimming the model to the smallest subset of points that retains reasonably predictive properties. This two-step process is illustrated using the 84-HIV-1 PR inhibitor data set shown in Table 1. Since many inhibitors were reported to be competitive and the peptide-like structures were shared by most, all inhibitors were assumed to bind in a similar manner at the HIV-1 PR active site. The HIV-1 PR-bound structures of inhibitors **4** and **51**[31,32,46] were used as templates in the superpositioning and model building of the remaining 82 inhibitors. Structure binding and manipulation were conducted on a VAX platform using MacroModel for least-squares superpositioning and the MacroModel implementation of the AMBER force field energy minimization.[47] Alchemy III was used for structure visualization on a Windows 386-based pc platform.[48] The HASL analyses described herein were conducted using author-written software on VAX and pc platforms.

HASL Creation

The basic premise of HASL modeling is that a molecule can be defined in terms of the space it occupies and the nature of the atoms present in that space. To this end, three-dimensional space is defined as a set of regularly spaced points in a three-dimensional grid. The translation of a molecule to a subset of such points is carried out by

† Current address: Schering-Plough Research Institute, Lafayette, NJ 07848.

® Abstract published in *Advance ACS Abstracts*, May 15, 1994.

0022-2623/94/1837-1769$04.50/0

Table 1. HIV-1 PR Inhibitor Set

inhibitor	structure	pK_i	ref	inhibitor	structure	pK_i	ref
1	qcN [14b] diqNH*tert*-butyl	9.92	35	43	tbocSAA [12] VVome	6.80	26
2	cbzV [17C] Vcbz	9.66	27, 38	44	tbocSAA [7] VVome	6.74	26
3	cbz [17a] Vcbz	9.66	27, 38	45	noaH [22b] Iamp	6.66	44
4	acSLN [14a] PIVome	9.62	32	46	tboc [17b] tboc	6.55	27, 38
5	L-694,746	9.47	41	47	cbz [20a] $NHCH_2CH(CH_3)CH_2CH_3$	6.44	44
6	cbzV [17b] Vcbz	9.42	27, 38	48	tbocFH [20a] $NHCH_2CH(CH_3)CH_2CH_3$	6.44	44
7	tboc [6b] FIFome	9.22	40	49	acSLN [14a] PIome	6.38	40
8	cbzGV [F] $CF_2CH_2CH_2$phenyl	9.00	39	50	tbocFH [20b] Iamp	6.30	44
9	cbzV [F] $CF_2CH_2CH_2$(2-pyridinyl)	8.85	39	51	SAA [8] VVome	6.30	26
10	cbzN [F] $CF_2CH_2CH_2$phenyl	8.68	39	52	acTI [4] $QRNH_2$	6.11	25, 31
11	cpoV [F] $CF_2CH_2CH_2$phenyl	8.66	39	53	SAA [6b] VVome	6.09	26
12	cbzva [F] $CF_2CH_2CH_2$phenyl	8.57	39	54	acSLN [18] PIome	6.01	40
13	cpnV [F] $CF_2CH_2CH_2$phenyl	8.54	39	55	acSLN [1] IVome	5.97	40
14	cbzI [F] $CF_2CH_2CH_2$phenyl	8.47	39	56	tbocSAA [6b] VVome	5.92	26
15	cbzV [16] Vcbz	8.43	27, 38	57	tbocSLN [6b] IVome	5.58	40
16	cbzI [16] Icbz	8.31	27, 38	58	tboc [6c] EF	5.52	43
17	cbzV [F] $CF_2CH_2CH_2$phenyl	8.30	39	59	tboc [19a] EF	5.48	42
18	cbz [20a] Iamp	8.30	44	60	VSQN [9] PIV	5.45	37
19	noaH [20a] Iamp	8.30	44	61	VSQ [15] IV	5.43	37
20	noaH [21] Iamp	8.30	44	62	tboc [6a] EF	5.40	43
21	noaH [22a] Iamp	8.15	44	63	SAA [11] VVome	5.36	26
22	tbocHPFH [2] IH	8.06	25, 28	64	SAA [10] VVome	5.35	26
23	acVV [16] VVac	8.00	27, 38	65	acLN [14a] PIome	5.34	40
24	acSLN [18] PIVome	7.96	40	66	SAA [3b] VVome	5.31	26
25	acV [16] Vac	7.92	27, 38	67	tboc [6d] EF	5.15	43
26	noaH [23] Iamp	7.92	38	68	acSLN [6b] IVome	5.10	40
27	tboc [17c] tboc	7.92	27, 38	69	acSQN [13] $PVVNH_2$	4.89	29
28	ivaVV [1] Astatin	7.77	38	70	acSQN [9] $PVVNH_2$	4.89	40
29	AA [3a] VVome	7.74	26, 34	71	acSQN [9] $PVVNH_2$	4.85	29
30	acLN [14a] PIVome	7.68	40	72	tboc [19b] EF	4.77	42
31	acLN [18] PIVome	7.68	40	73	SAA [9] $PVVNH_2$	4.74	26
32	noaH [24] Iamp	7.59	44	74	lacVV [1] Astatin	4.72	28
33	tbocFH [20a] Iamp	7.55	44	75	tboc [19c] EF	4.66	42
34	acVV [1] Astatin	7.45	26	76	tboc [19b] EF	4.62	42
35	tboc [17a] tboc	7.40	27, 38	77	acSAA [9] $PVVNH_2$	4.42	26
36	cbzAAF [3a] VVome	7.32	25	78	acSQN [6b] VVome	4.41	40
37	SAA [3a] VVome	7.21	26	79	acSQN [6b] $VVNH_2$	4.41	29
38	tboc [6b] EF	7.20	43	80	tboc [19a] PIVome	4.40	42
39	va [5] Iamp	7.15	25, 36	81	haloperidol	4.00	33
40	CTLN [F] PISPI	7.15	30	82	tboc [19d] PIVome	3.46	42
41	SAA [7] VVome	7.10	26	83	tboc [19b] PIVome	3.44	42
42	cbzA [3a] VVome	6.92	34	84	tboc [19c] PIVome	3.41	42

retaining those points found within the van der Waals radius of each atom in the molecule. Thus, the resulting molecular lattice of points consists of the classical x,y,z coordinates, where each point carries two other types of information: atom type and partial binding value.

The atom type is simply defined as electron-rich, electron-poor, or electron-neutral and is represented by H-values of 1, −1, and 0, respectively. The H-values used in the present investigation are defined in Table 2 and represent an arbitrary atom-type definition set herein loosely based on electron density or atomic hydrophobicity parameters.[49] The process of lattice construction is illustrated in Figure 1. At a resolution of 2.0 Å, inhibitor 1 (Figure 1A) was converted to 261 points (x,y,z,H; Figure 1B) which represented 97.0% of inhibitor 1 atoms. The superpositioning of 10 inhibitors is shown in Figure 1C to illustrate the overall β-sheet configuration of these inhibitors. Ultimately, the superpositioning of all 84 inhibitor molecules resulted in a merged set of 899 lattice points as depicted in Figure 1D.

The effect of point spacing or resolution value (Å) on the extent of molecular representation is shown in Figure 2. Using inhibitor 1 as an example, the number of lattice points and the percent of the molecule represented were found to increase exponentially (in a concave/convex relationship) as the point spacing decreased (Figure 2A). For example, decreasing the resolution from 3.0 to 2.0 Å results in an increase in the number of lattice points from 71 to 261, while molecular representation increases from 61.6% to 97.0%, respectively. Using atomic volumes as defined by van der Waals radii, inhibitor 1 was calculated to contain atom types H = 1, −1, and 0 at levels of 12.7%, 8.6%, and 78.7%, respectively. A comparison of the distribution of lattice point H-values (as percentages) determined at different resolutions was made with the calculated values (lines) for inhibitor 1 and is shown in Figure 2B. These analyses demonstrate that the point descriptions of inhibitor 1 represent the molecule most accurately at resolution values below 2.5 Å, where lattice point percentages most closely agree with the calculated H-value distribution.

The distribution of partial binding values (partial pK_i) among the composite 899-point HASL was determined by using an improved version of the previously described algorithm,[13,14] which now entails randomly scrambling the sequence of molecules during the corrective phase, thus avoiding a repeated sequence-dependent bias. Iterations were carried out until the average error in prediction was minimized. The HASL model which emerged as a result was found to have a self-consistent correlation (r^2) of 0.988 (predicted vs actual pK_i), as shown in Figure 3, and reflects the potential of the method to model the HIV-1 PR active site/inhibitor interaction. However, since there are 899 points in the model serving to account for the activities of 84 inhibitors, it would appear prudent to reduce the model to a smaller, more robust subset of points which

Chart 1

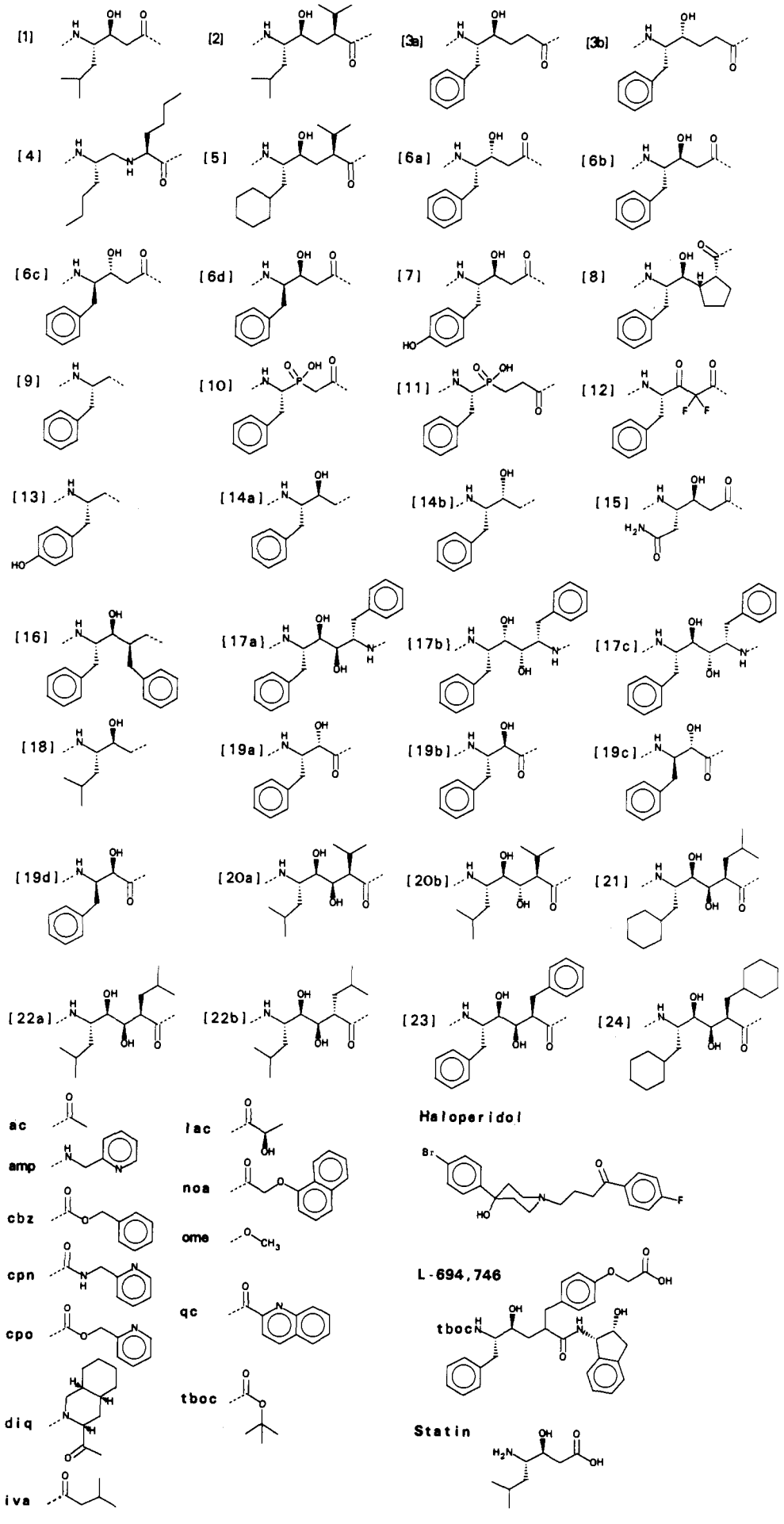
[1]
[2]
[3a]
[3b]
[4]
[5]
[6a]
[6b]
[6c]
[6d]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14a]
[14b]
[15]
[16]
[17a]
[17b]
[17c]
[18]
[19a]
[19b]
[19c]
[19d]
[20a]
[20b]
[21]
[22a]
[22b]
[23]
[24]
ac
amp
cbz
cpn
cpo
diq
iva
lac
noa
ome
qc
tboc
Haloperidol
L-694,746
Statin

Table 2. HASL Atom-Type Definitions (*H*)

MM2	*H*	atom	type	MM2	*H*	atom	type
1	0	C	sp^3)C(	16	−1	S^+	)S—
2	0	C	sp^2)C=	17	−1	S	)SO
3	−1	C	sp^2)C=O	18	−1	S	$)SO_2$
4	0	C	sp —C≡C—	19	0	Si	silane
5	0	H	hydrogen	21	−1	H	—OH
6	1	O	COC, COH	22	0	C	cyclopropyl
7	1	O	)C=O	23	−1	H	)N—H
8	1	N	sp^3)N—	24	−1	H	COOH
9	1	N	sp^2 —N=	25	1	P	)P—
10	1	N	sp C≡N	26	−1	B	)B—
11	1	F	fluoride	27	0	B	)B(
12	1	Cl	chloride	28	0	H	vinyl
13	1	Br	bromide	29	−1	N^+	)N(+
14	1	I	iodide	37	1	N	)C=N—
15	1	S	—S—				

retains useful predictivity while minimizing the potential to overfit data. The objective of this exercise would be to focus the model on the remaining points which may retain a spatial relevance when compared to actual active site geometry.

HASL Trimming

The process of trimming the HASL as a means to locate the most significant points (relevant to binding) in the model must insure that (1) the initial model has incorporated all potentially relevant points and (2) the process itself does not remove relevant points. Therefore, unlike most structure–activity paradigms wherein efforts are made to reduce the number of descriptors, premise 1 suggests that the process of building a three-dimensional pharmacophore from a HASL requires an initial, detailed model (at small lattice spacing) in order to insure that all or most of the atoms in each molecule are represented. Premise 2 can be addressed heuristically by examining the effects of different trimming methods on the resultant models' predictivities.

The following two steps involved in trimming are cycled through as many times as necessary to reduce the original HASL model to a subset of points, culminating in a HASL-derived pharmacophore: (1) removal of those HASL points that currently represent the least significant partial pK_i values (e.g., 10% of all HASL points in the current model that have partial pK_i values nearest to zero) and (2) iterative distribution of the partial pK_i values among the remaining lattice points to achieve the best correspondence between actual and predicted pK_i. This approach is illustrated in Figure 4 wherein four HASL models (at resolutions of 2.0, 2.5, 3.0, and 3.5 Å) were gradually reduced using a 10% trimming cycle as just described. The data demonstrate that the correlation (self-consistency) of the models at these four resolutions holds up quite well until the models drop below 100 points. In particular, the model at 3.5 Å collapses ($r^2 < 0.9$) at ca. 60 points, while that at 2.0 Å reaches a similar collapse at ca. 25 points. These results suggest that a potential pharmacophore may emerge when starting with HASL models at lower resolution which represent a greater percentage of the atoms in each molecule and are therefore more likely to retain atom sites directly relevant to inhibitor binding.

The risk of removing relevant points using the two-step trimming cycle was addressed by focusing on its magnitude.

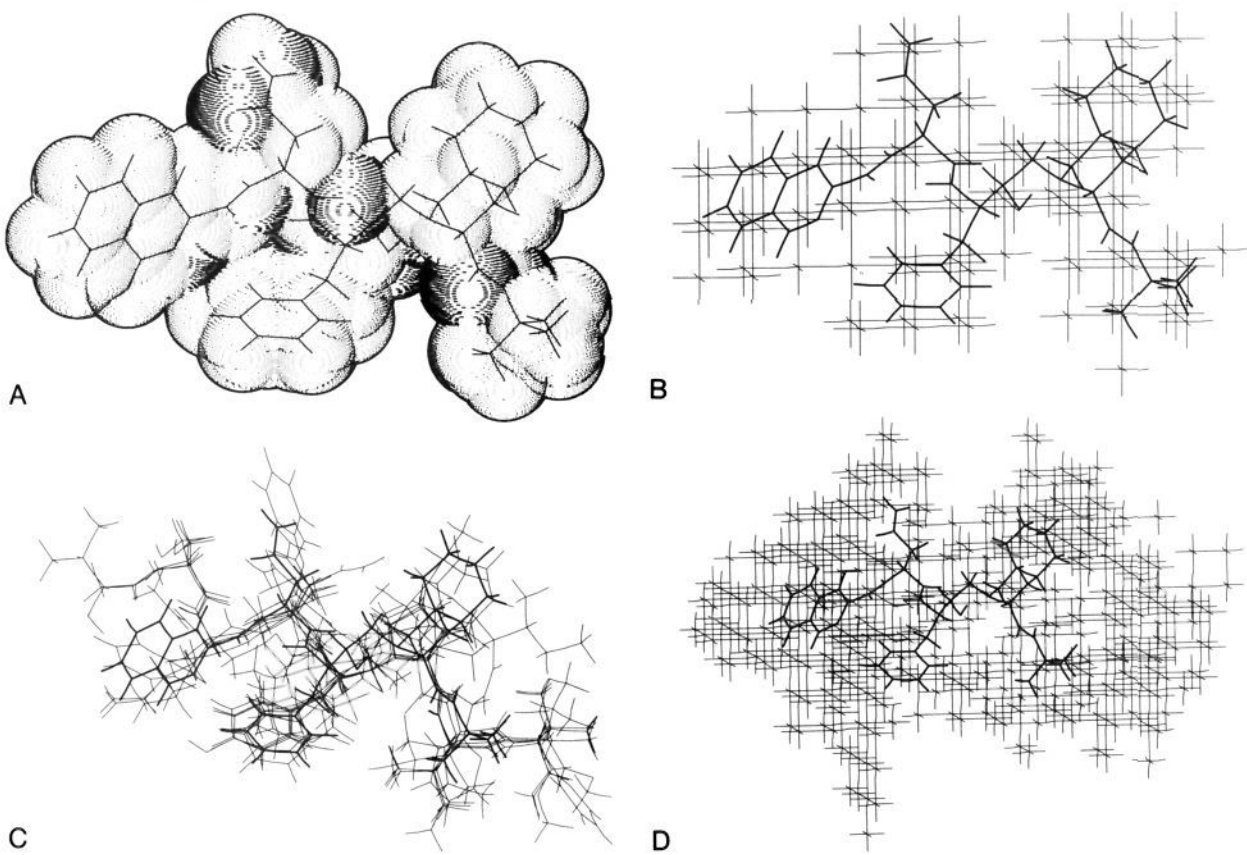

Figure 1. (A) van der Waals space-filling model of inhibitor **1**. (B) 261-Point molecular lattice developed for inhibitor **1** at 2.0 Å representing 97.0% of the molecule. (C) Superposed set of HIV-1 PR inhibitors (**1**, **10**, **20**, **30**, **40**, **50**, **60**, **70**, and **84**) illustrating overall β-pleated configuration. (D) Complete 899-point HASL developed for the 84-HIV-1 PR inhibitor set. Inhibitor **1** is highlighted in panels B–D.

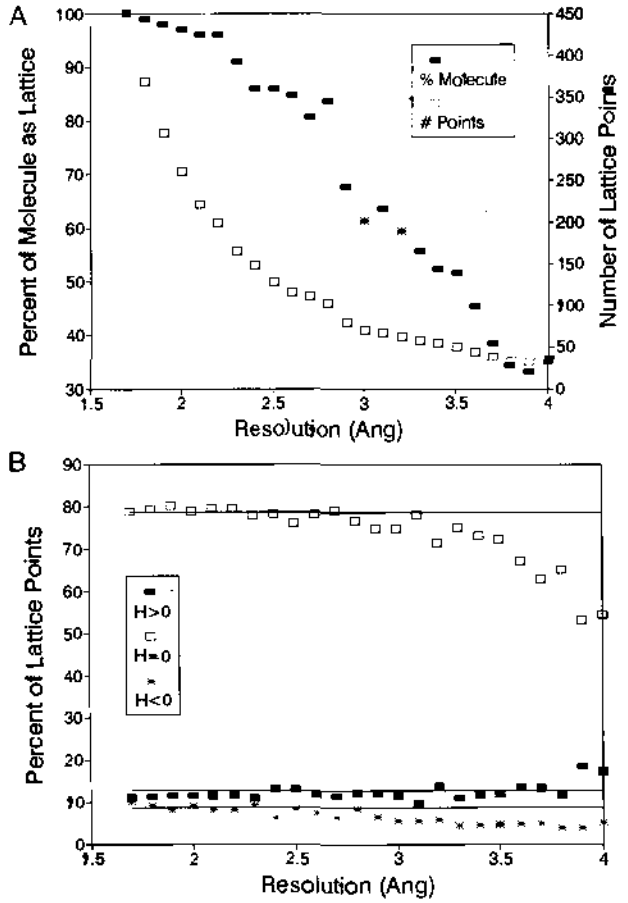

Figure 2. (A) Dependence of molecular representation and number of lattice points on the resolution. (B) Dependence of lattice point representation on the resolution. Analyses illustrated in panels A and B were developed using inhibitor **1** and the *H*-value definitions listed in Table 2.

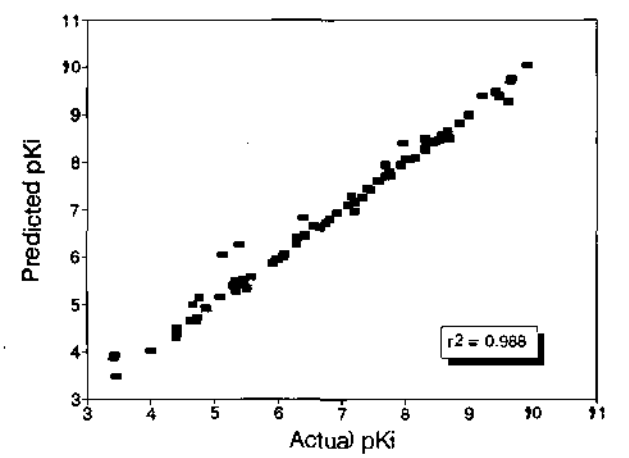

Figure 3. 899-Point HASL predicting pK_i for all 84 inhibitors and yielding a self-consistency $r^2 = 0.988$.

The data in Figure 5 were obtained by starting with the same 899-point HASL model originally derived at 2.0 Å and varying the trimming cycle (2.5%, 5.0%, 10%, and 20%). As might be expected, the most conservative trimming option (2.5%) yielded the most robust sets of models. Using an r^2 target of >0.8, a 2.5% trimming cycle yielded an **11**-point pharmacophore (Figure 5, inset) while a 20% trimming cycle yielded a **15**-point pharmacophore. It can be concluded that the **11**-point model, obtained by using the conservative 2.5% trimming cycle, must contain

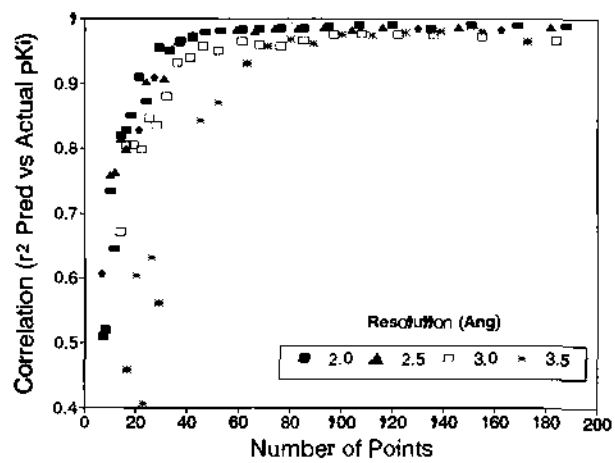

Figure 4. Comparison of the editing process (10% trimming cycle) conducted on four HASL models (2.0, 2.5, 3.0, and 3.5 Å) illustrating the dependence of r^2 on the number of lattice points.

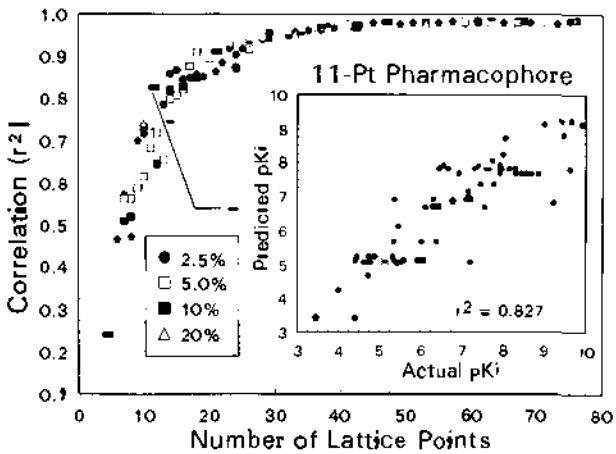

Figure 5. Comparison of editing processes using trimming cycles set at 2.5%, 5.0%, 10%, and 20% illustrating the dependence of r^2 on the number of lattice points.

more relevant binding information than the **15**-point model, obtained from a 20% trimming cycle, since both models have comparable r^2-correlation values of 0.8.

Predictivity

HASL and pharmacophore predictivities were assessed by using half the 84-inhibitor data set as a learning set (odd-numbered inhibitors) and the other half as a test set (even-numbered inhibitors). Using the previously determined optimum resolution of 2.0 Å, a 784-point HASL was obtained from the 42-inhibitor learning set having a self-consistency r^2 of **1.00**. When the even-numbered 42-inhibitor test set was predicted by this HASL model, a predictivity r^2 of 0.726 was obtained (Figure 6).

The effect of applying a 2.5% trimming cycle to the 784-point HASL was examined for both the learning set self-consistency r^2 and the test set predictivity r^2 and is illustrated in Figure 7. No significant changes in r^2 were observed until the lattice point number fell below 30, whereupon both self-consistency and predictivity r^2-values appeared to decrease in a parallel manner. As expected, points that are required for a self-consistent learning set model are also required for predicting the binding of members of the test set. Since the trimming cycle affects both properties in a similar manner, as evidenced by the relatively constant test set/learning set r^2 ratio, the points

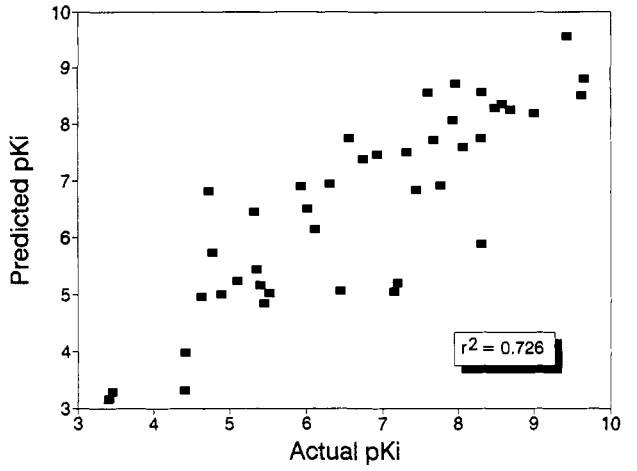

Figure 6. 784-Point HASL, developed from a learning set of 42 odd-numbered inhibitors, predicting the pK_i values of 42 even-numbered inhibitors yielding a predictivity r^2 = 0.726.

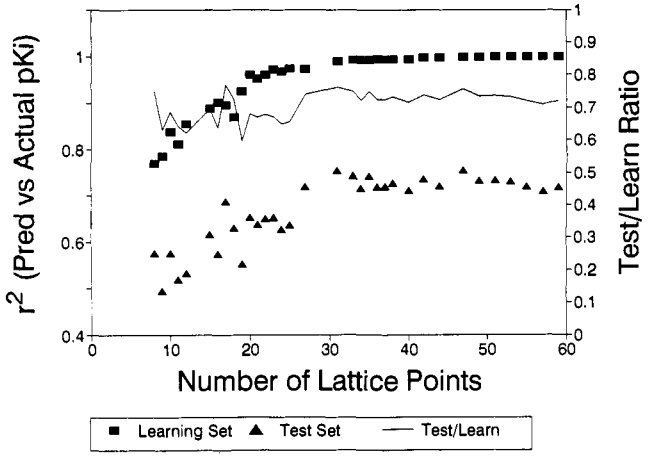

Figure 7. Comparing the effects of a 2.5% trimming cycle applied to the odd-numbered inhibitor learning set and the even-numbered inhibitor test set. Learning set r^2 represents self-consistency and test set r^2 represents predictivity.

Table 3. 11-Point HIV-1 PR Pharmacophore Derived from 899-Point HASL at 2.0 Å

point	x (Å)	y (Å)	z (Å)	H	partial pK_i
1	6	12	0	0	6.37
2	4	10	14	0	1.30
3	0	6	8	0	1.55
4	10	-2	10	0	0.89
5	4	8	14	-1	1.45
6	8	6	10	-1	1.69
7	0	6	14	0	-1.32
8	14	2	12	0	-2.15
9	-2	10	12	-1	-1.14
10	-2	12	12	1	-1.56
11	2	4	12	1	-1.65

that remain (<30) must contain significant binding information and are likely to provide valuable insight regarding the active site/inhibitor interaction.

Pharmacophore and Active Site

The Cartesian coordinates, H-values, and partial pK_i (ppK_i) values of the 11-point pharmacophore obtained from the full 84-inhibitor, 899-point HASL constructed at 2.0 Å are listed in Table 3. Regions representing positive binding interactions were found centered at site points 1–4 (H = 0, electron-neutral, potentially lipophilic) and site points 5 and 6 (H = −1, electron-poor, potentially hydrogen-bond acceptors). The other five sites represent regions that result in negative binding or repulsive interactions when occupied by the designated atom type.

The inhibitor/pharmacophore relationship can be further explored by examining which site points play the greatest role in determining inhibitor potency. This was done by superposing the strongest five inhibitors (**1–5**) and the weakest five inhibitors (**80–84**) on the 11-point pharmacophore and determining the minimum distance between a given site point and a corresponding atom type in the inhibitor molecule. Correspondence occurs when both the site point and the atom type share the same H-value. The results of this analysis are shown in Table 4. The average distances between a site point and the nearest corresponding inhibitor atom appear to be correlated to the differences in binding activities between the two sets of inhibitors. The stronger inhibitor set clearly occupies positions nearer the first six site points where interactions are favored (positive ppK_i) and is further removed from unfavorable (negative ppK_i) interactions at sites 7–11. The exceptions observed at points 9 and 10 are not significant since distances beyond 2 Å are not within the critical van der Waal's interaction range assumed in the development of a HASL model.

Further examination of the trends evident in the data of Table 4 suggests that it may be possible to relate pK_i to a function of the distances between pharmacophore site points and corresponding inhibitor atoms. Such a relationship was developed using the distance value (D_{ij}) for every site point (i) obtained for each inhibitor (j) according to eq 1.

$$D_{ij} = 1 = \left(\frac{\text{dist}}{\text{tol}}\right)^n \qquad (1)$$

where

dist = distance in angstroms between site point and nearest corresponding atom in molecule j

tol = tolerance, defined as 2 Å

n = 6

The value of D_{ij} exponentially approaches unity as the distance approaches zero and is defined as zero when the distance exceeds the tolerance (tol = 2.0 Å). This operation results in a matrix of D_{ij} values for i = 1, 11 site points and j = 1, 84 inhibitors. When this matrix was subjected to multiple regression analysis using all 11 site points, the two equations shown in Table 5 were obtained wherein the regression line either used a calculated y intercept or was forced through the origin, resulting in eqs 2 or 3, respectively.

The use of the sixth-order exponential function in eq 1 is based on the arbitrary assumption that the proximity interactions are comparable to classical electrostatic forces between two charged points. It should be noted, however, that a simpler second-order (n = 2) assumption yields similar, albeit slightly poorer, correlations. The exact mathematical form of the relationship is empirical and not in itself the point of this excercise; rather, it is the independent statistical confirmation that these HASL-derived site points could be physically related to the activities of the HIV-1 PR inhibitor set.

Although the 11-point HASL pharmacophore appears to result in a self-consistent correlation between predicted and actual inhibitor pK_i (r^2 = 0.827), some interpretive

Table 4. Distance Matrix Developed for Selected Inhibitors and 11-Point Pharmacophore

site point	1	2	3	4	5	6	7	8	9	10	11
ppK_i	6.37	1.30	1.55	0.89	1.45	1.69	−1.32	−2.15	−1.14	−1.56	−1.65
H	0	0	0	0	−1	−1	0	0	−1	1	1
	distance from pharmacophore site point to nearest corresponding inhibitor atom (Å)										
inhibitor	pK_i										
1	9.92	0.65	1.09	3.38	3.74	1.68	1.44	2.01	1.73	3.76	3.32	3.28
2	9.66	0.57	0.87	1.09	2.78	2.56	1.40	3.56	1.70	2.46	4.69	2.61
3	9.66	0.57	0.87	1.09	2.78	2.56	1.40	3.56	1.70	2.46	4.69	2.61
4	9.62	1.01	1.03	3.03	2.55	1.56	1.82	1.13	1.64	1.48	1.85	3.58
5	9.47	0.64	2.03	1.13	0.45	1.95	1.84	1.91	2.24	6.80	6.72	2.42
average	9.67	0.69	1.18	1.94	2.46	2.06	1.58	2.43	1.80	3.39	4.25	2.90
80	4.40	1.01	2.30	3.03	2.55	2.28	1.24	0.74	1.64	6.27	6.27	1.52
81	4.00	1.06	4.05	3.62	5.50	5.25	2.60	4.00	0.77	8.84	9.31	2.19
82	3.46	1.01	2.30	4.00	2.55	2.28	1.82	0.74	1.64	6.27	6.27	1.52
83	3.44	1.01	2.30	3.03	2.55	2.28	1.82	0.74	1.64	6.27	6.27	1.52
84	3.41	1.01	2.30	4.00	2.55	2.28	1.24	0.74	1.64	6.27	6.27	1.52
average	3.74	1.02	2.65	3.54	3.14	2.87	1.74	1.39	1.47	6.78	6.88	1.65

Table 5. Site Point Coefficients Obtained by Regression Analysis Compared with Partial pK_i (ppK_i)

		site point											
	constant	1	2	3	4	5	6	7	8	9	10	11	r^2
ppK_i		6.37	1.30	1.55	0.89	1.45	1.69	−1.32	−2.15	−1.14	−1.56	−1.65	0.827
eq 2[a]	8.38	−1.99	1.40	0.65	1.08	1.64	1.43	−1.33	−0.44	−1.67	−1.27	−2.47	0.728
(t-stat)	(4.23)	(−0.97)	(3.61)	(1.53)	(2.06)	(4.18)	(3.94)	(−3.82)	(−1.37)	(−4.13)	(−2.81)	(−3.97)	
eq 3[b]		6.50	1.98	−0.12	2.08	1.21	1.41	−1.47	−0.54	−1.70	−0.75	−2.32	0.665
(t-stat)		(14.85)	(4.92)	(−0.27)	(4.00)	(2.88)	(3.52)	(−3.83)	(−1.51)	(−3.79)	(−1.56)	(−3.37)	

[a] Equation 2 uses all site point coefficients derived from eq 1 assuming a best linear least-squares fit. [b] Equation 3 uses all site point coefficients derived from eq 1 assuming a best linear least-squares fit through the origin.

insight can be obtained by using an independent multiple regression analysis of this structure–activity data. For example, the site 1 ppK_i of 6.37 would appear unusually large compared to other site ppK_i values. This point may represent a necessary commonality among all inhibitors, since the average ppK_i difference in the strong vs weak inhibitor comparison (Table 4) is relatively small. The regression analysis would appear to confirm this interpretation since eq 2 assigns a large and statistically significant (95% confidence $|t\text{-stat}| \geq 2.20$) common or residual p$K_i$ value of 8.38, while eq 3 assigns an equally significant cofficient of 6.50 to the point 1 variable. The significant site point coefficients in eqs 2 and 3 and the HASL ppK_i values share the same signs. Since the positive/negative effects on binding are assigned in the same way for most of the pharmacophore sites, the regression analysis is consistent with the HASL-derived physical relevance of the site points with the possible exception of points 3, 8, and 10 (where $|t\text{-stat}| < 2.20$).

The relative positions of pharmacophore site points, inhibitor, and active site components are detailed in the illustrations shown in Figure 8. Inhibitor 1 and the 11-point pharmacophore are shown in Figure 8, top. The nature of the interactions occurring at sites 1–6 is of particular interest since they represent a positive contribution to inhibitor binding. As previously noted, site 1 (H = 0, lipophilic), which should represent a structural commonality for most inhibitors, is centered at the P1–P1′ region (located near the substrate scissile bond). This observation is not unexpected since all the inhibitors in this study extend out from the P1–P1′ region and necessarily occupy this space. The other significant lipophilic interaction site (2) is found near the P2 region and corresponds to inhibitor alkyl side chains frequently encountered at this position (e.g., Ala, Asp, or Val).

Occupation of sites 5 and 6 (where H = −1) by electron-deficient atoms in the inhibitor is predicted by the model to enhance binding activity. As is illustrated for inhibitor 1, sites 5 and 6 are near an amide carbonyl carbon (1.68 Å) and hydroxyl hydrogen (1.44 Å), respectively. These two features are found among the many peptidyl backbone hydrogen-bonding sites present in this HIV-1 PR inhibitor set[31,32] and are consistent with a positive contribution to inhibitor potency. The four sites (6, 1, 2, and 5) comprise an arc which sweeps through the P1–P2 region and effectively captures the significant binding elements present in most inhibitors through that region.

The relationship of pharmacophore sites 1–6 to the HIV-1 PR active site components is illustrated in Figure 8, bottom. Potential interactions between an internalized HASL model (the pharmacophore) and the external active site components can be identified by comparing complementary H-values. Electron-neutral (H = 0) site points and active site atoms are considered complementary since they are capable of undergoing attractive, lipophilic van der Waals interactions. Electron-rich (H = 1) and electron-poor (H = −1) atoms would be expected to interact with sites having the opposite H-value, e.g., a hydroxyl hydrogen (H = −1) hydrogen bonding with a carbonyl oxygen (H = 1).

The relevance of site 5 (H = −1) may be related to the central role played by the active site water molecule in providing hydrogen bonding between the lower active site flap and inhibitor. Site 5 appears to correspond to the presence of the carbonyl carbon of a peptidyl amide moiety at the P1 site. The implied carbonyl oxygen atom can undergo hydrogen bonding with the active site water molecule. Site 5–HOH distance was found to be 2.6 Å. This is well within hydrogen-bonding range since the carbonyl carbon–oxygen bond length is 1.2 Å, making the potential hydrogen-bonding distance as low as 1.4 Å. Thus, site 5 represents an important hydrogen-bonding interaction region between the inhibitor and the lower HIV-1 PR

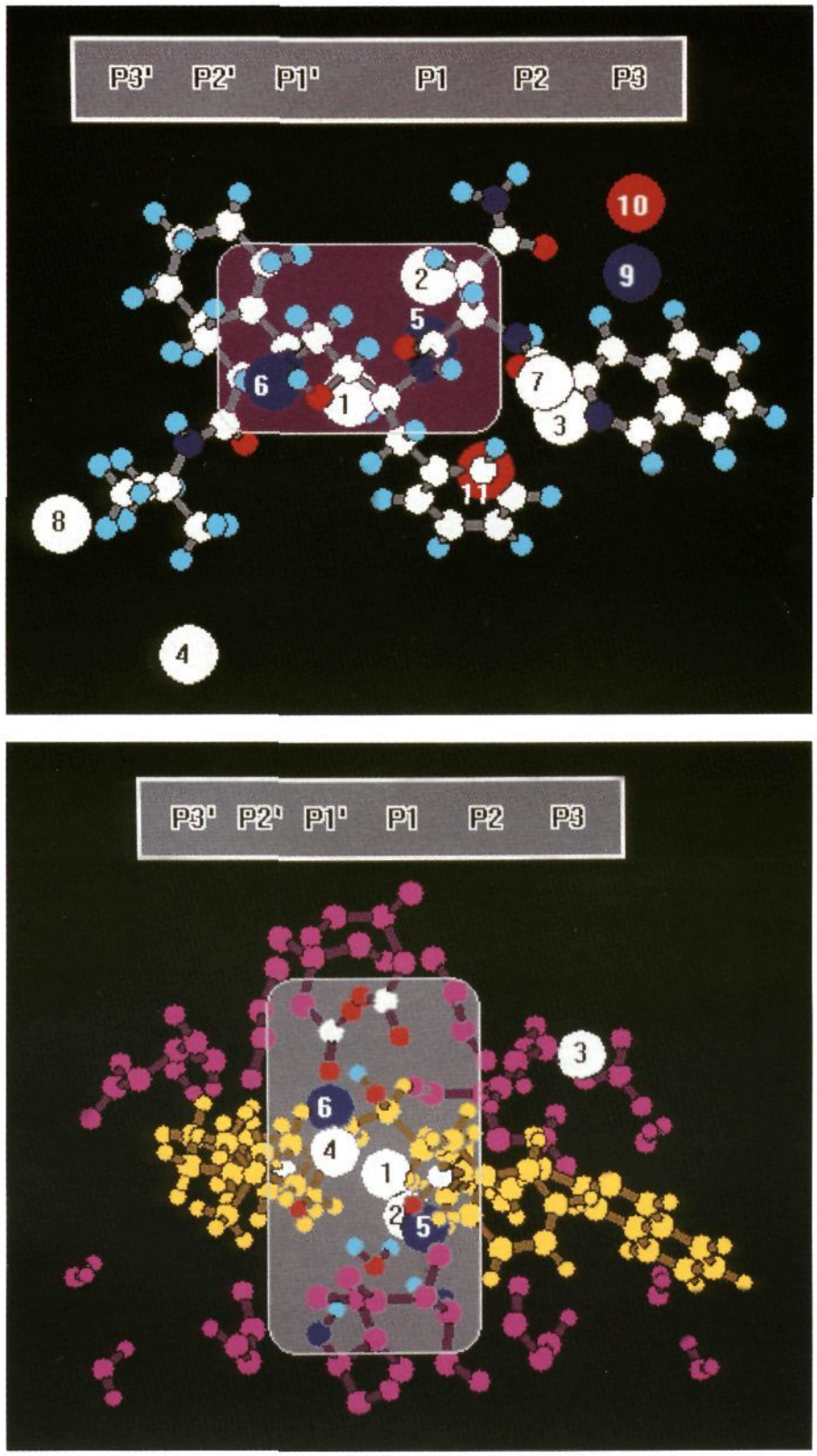

Figure 8. (Top) Inhibitor **1** superposed on the 11-point HASL-derived pharmacophore, illustrating site 6–hydroxyl and site 5–carbonyl correspondence. (Bottom) HIV-1 PR active site residues (magenta), inhibitor **1** (yellow), and site points (1–6) superposed, illustrating site 5–water molecule and site 6–aspartyl carboxyl correspondence. Site points having *H*-values of 1, 0, and −1 are colored red, white, and blue, respectively.

active site flap and is consistent with the known role and position of the active site water molecule.[32]

Site 6 ($H = -1$) essentially represents the average position for a hydroxyl hydrogen present in many of the inhibitors under study. Inhibitors containing this moiety are considered examples of multisubstrate analogs wherein the hydroxyl functionality mimics the transition state during the course of general acid–general base catalysis of

the substrate peptide bond by an incoming water molecule.[50] Site 6 is found 3.6–4.0 Å from the aspartate carboxyl oxygens of the upper active site flap. A distance range of 2.5–3.6 Å was observed in the case of HIV-1 PR-bound inhibitor 4.[32] Thus, site 6 appears to reflect the need for a electron-deficient atom located in a similar manner to the incoming water molecule needed by HIV-1 PR to hydrolyze the substrate.

Conclusions

The application of HASL to the study of 84 HIV-1 PR inhibitors serves to illustrate a useful extension of the method to yield a subset of HASL points which corresponds to a predictive, three-dimensional pharmacophore. The 11-point HASL subset obtained from a 899-point HASL determined at 2.0 Å represents a predictive pharmacophore (r^2 = 0.827, 84 inhibitors) wherein each point is assigned a partial pK_i value denoting its contribution to the overall activity of an inhibitor. The pharmacophore was found to be consistent with several critical hydrogen-bonding sites between the inhibitor and the active site. The use of such a predictive tool in the rational design of novel inhibitors awaits application.

References

(1) Ehrlich, P. Present Status of Chemotherapy. *Chem. Ber.* **1909**, *42*, 17–47.
(2) Ariens, E. J. Molecular Pharmacology, a Basis for Drug Design. In *Progress in Molecular and Subcellular Biology*; Hahn, F. E., Ed.; Springer-Verlag: New York, 1977; Vol. 5, pp 429–529.
(3) Gund, P. Pharmacophoric pattern searching and receptor mapping. *Annu. Rep. Med. Chem.* **1979**, *14*, 298–308.
(4) Gund, P. Three-Dimensional Pharmacophoric Pattern Searching. *Prog. Drug. Res.* **1966**, *10*, 117–143.
(5) Humblet, C.; Marshall, G. R. Pharmacophore Identification and Receptor Mapping. *Annu. Rep. Med. Chem.* **1980**, *15*, 267–276.
(6) Marshall, G. R.; Naylor, C. B. Use of Molecular Graphics for Structural Analysis of Small Molecules. In *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Series Eds., Ramsden, L. A., Vol. Ed.; Pergammon Press: Oxford and New York, 1990; Vol. 4, pp 431–458.
(7) Langridge, R.; Klein, T. E. Introduction to Computer Graphics and Its Use for Displaying Molecular Structures. In *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Series Eds., Ramsden, L. A., Vol. Ed.; Pergammon Press: Oxford and New York, 1990; Vol. 4, pp 413–430.
(8) Cohen, N. C.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. Molecular Modeling Software and Methods for Medicinal Chemistry. *J. Med. Chem.* **1990**, *33*, 883–894.
(9) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(10) Ghose, A. K.; Crippen, G. M. The Distance Geometry Approach to Modeling Receptor Sites. In *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Series Eds., Ramsden, L. A., Vol. Ed.; Pergammon Press: Oxford and New York, 1990; Vol. 4, pp 715–734.
(11) 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.
(12) Marshall, G. R.; Cramer, R. D., III. Three-dimensional structure-activity relationships. *Trends Pharmacol. Sci.* **1988**, *9*, 285–289.
(13) Doweyko, A. M. The Hypothetical Active Site Lattice. An Approach to Modeling Active Sites from Data on Inhibitor Molecules. *J. Med. Chem.* **1988**, *31*, 1396–1406.
(14) Doweyko, A. M. New Tool for the Study of Structure-Activity Relationships in Three Dimensions. In *Probing Bioactive Mechanisms*; Magee, P., Henry, D. R., Block, J. H., Eds.; ACS Symposium Series 413; American Chemical Society: Washington, DC, 1989; pp 82–104.
(15) Doweyko, A. M. The Hypothetical Active Site Lattice – In Vitro and In Vivo Explorations Using a Three Dimensional QSAR Technique. *J. Math. Chem.* **1991**, *7*, 273–285.
(16) Wiese, M. The Hypothetical Active-Site Lattice. In *3D QSAR In Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM, 1993; pp 431–442.
(17) Molecular Design and Modeling: Concepts and Applications, Part B, Section III, Drugs. *Methods in Enzymology*; Langone, J. J., Ed.; Academic Press: New York, 1991; Vol. 203, pp 587–693.
(18) *QSAR: Quantitative Structure-Activity Relationships in Drug Design*, Proceedings of the 7th European Symposium on QSAR; Fauchere, J. L., Ed.; Alan R. Liss: New York, 1989.
(19) Loew, G. H.; Villar, H. O.; Alkorta, I. Strategies for indirect computer-aided drug design. *Pharm. Res.* **1993**, *10*, 475–486.
(20) Meek, T. D. Inhibitors of HIV-1 Protease. *J. Enzyme Inhib.* **1992**, *6*, 65–98.
(21) Huff, J. R. HIV Protease: A Novel Chemotherapeutic Target for AIDS. *J. Med. Chem.* **1991**, *34*, 2305–2314.
(22) Norbeck, D. W.; Kempf, D. J. HIV Protease Inhibitors. *Annu. Rep. Med. Chem.* **1991**, *26*, 141–149.
(23) Martin, J. A. Recent advances in the design of HIV-proteinase inhibitors. *Antiviral Res.* **1992**, *17*, 265–278.
(24) Wlodawer, A.; Erickson, J. W. Structure-based inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543–585.
(25) Norbeck, D. W. Recent Advances in Anti-retroviral Chemotherapy for AIDS. In *Annual Reports in Medicinal Chemistry*; Bristol, J. A., Ed.; Academic Press: New York, 1989; pp 149–158.
(26) Dreyer, G. B.; Metcalf, B. W.; Tomaszek, T. A., Jr.; Carr, T. J.; Chandler, A. C., III; Hyland, L.; Fakhoury, S. A.; Magaard, V. W.; Moore, M. L.; Strickler, J. E.; Debouck, C.; Meek, T. D. Inhibition of human immunodeficiency virus 1 protease *in vitro*: Rational design of substrate analogue inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9752–9756.
(27) Kempf, D. J.; Norbeck, D. W.; Codacovi, L.; Wang, X. C.; Kohlbrenner, W. E.; Wideburg, N. E.; Paul, D. A.; Knigge, M. F.; Vasavanonda, S.; Craig-Kennard, A.; Saldivar, A.; Rosenbrook, W., Jr.; Clement, J. J.; Plattner, J. J.; Erickson, J. Structure-Based, C_2 Symmetric Inhibitors of HIV Protease. *J. Med. Chem.* **1990**, *33*, 2687–2689.
(28) Richards, A. D.; Roberts, R.; Dunn, B. M.; Graves, M. C.; Kay, J. Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases. *FEBS Lett.* **1989**, *247*, 113–117.
(29) Moore, M. L.; Bryan, W. M.; Fakhoury, S. A.; Magaard, V. W.; Huffman, W. F.; Dayton, B. D.; Meek, T. D.; Hyland, L.; Dreyer, G. B.; Metcalf, B. W.; Strickler, J. E.; Gorniak, J. G.; Debouck, C. Peptide Substrates and Inhibitors of the HIV-1 Protease. *Biochem. Biophys. Res. Commun.* **1989**, *159*, 420–425.
(30) Billich, S.; Knoop, M.-T.; Hansen, J.; Strop, P.; Sedlacek, J.; Mertz, R.; Moelling, R. Synthetic Peptides as Substrates and Inhibitors of Human Immune Deficiency Virus-1 Protease. *J. Biol. Chem.* **1988**, *263*, 17905–17908.
(31) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Structure of Complex of Synthetic HIV-1 Protease with a Substrate-Based Inhibitor at 2.3 Å Resolution. *Science* **1989**, *246*, 1149–1152.
(32) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. X-ray crystallographic Structure of a complex Between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.
(33) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Ortiz de Montellano, P. R.; DeCamp, D. L.; Babe, L. M.; Craik, C. S. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.
(34) 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.; Matthews, T. J.; Metcalf, B. W.; Petteway, S. R. Inhibition of HIV-1 protease in infected T-lymphocytes by synthetic peptide analogues. *Nature* **1990**, *343*, 90–92.
(35) 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. F. Rational Design of Peptide-Based HIV Proteinase Inhibitors. *Science* **1990**, *248*, 358–361.
(36) McQuade, T. J.; Tomasselli, A. G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T. K.; Heinrickson, R. L.; Tarpley, W. G. A Synthetic HIV-1 Protease Inhibitor with Antiviral Activity Arrests HIV-Like Particle Maturation. *Science* **1990**, *247*, 454–456.
(37) Tomasselli, A. G.; Olsen, M. K.; Hui, J. O.; Staples, D. J.; Sawyer, T. K.; Heinrickson, R. L.; Tomich, C.-S. C. Substrate Analogue Inhibition and Active Site Titration of Purified Recombinant HIV-1 Protease. *Biochemistry* **1990**, *29*, 264–269.
(38) Matayoshi, E. D.; Wang, G. T.; Kraft, G. A.; Erickson, J. Novel Fluorogenic Substrates for Assaying Retroviral Proteases by Resonance Energy Transfer. *Science* **1990**, *247*, 954–958.
(39) Sham, H. L.; Betebenner, D. A.; Wideburg, N. E.; Saldivar, A. C.; Kohlbrenner, W. E.; Vasavanonda, S.; Kempf, D. J.; Norbeck, D. W.; Zhao, C.; Clement, J. J.; Erickson, J. E.; Plattner, J. J. Potent HIV-1 Protease Inhibitors with Antiviral Activities *in Vitro*. *Biochem. Biophys. Res. Commun.* **1991**, *175*, 914–919.
(40) Rich, D. H.; Green, J.; Toth, M. V.; Marshall, G. R.; Kent, S. B. H. Hydroxyethylamine Analogues of the p17/p24 Substrate Cleav-

age site Are Tight-Binding Inhibitors of HIV Protease. *J. Med. Chem.* **1990**, *33*, 1288–1295.
(41) Lingham, R. B.; Arison, B. H.; Colwell, L. F.; Hsu, A.; Dezeny, G.; Thompson, W. J.; Garrity, G. M.; Gagliardi, M. M.; Hartner, F. W.; Drake, P. L.; Balani, S. K.; Pitzenberger, S. M.; Murphy, J. S.; Ramjit, H. G.; Inamine, E. S.; Treiber, L. R. HIV-1 Protease Inhibitor Activity of L-694,746, A novel Metabolite of L-689,502. *Biochem. Biophys. Res. Commun.* **1991**, *181*, 1456–1461.
(42) Raju, B.; Deshpande, M. S. Substrate Analog Inhibitors of HIV-1 Protease Containing Phenylnorstatine as a Transition State Element. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 181–186.
(43) Raju, B.; Deshpande, M. S. Investigating the Stereochemistry of Binding to HIV-1 Protease with Inhibitors Containing Isomers of 4-Amino-3-hydroxy-5-phenylpentanoic Acid. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 187–190.
(44) Thaisrivongs, S.; Tomasselli, A. G.; Moon, J. B.; Hui, J.; McQuade, T. J. Turner, S. R.; Strohbach, J. W.; Howe, W. J.; Tarpley, W. G.; Heinrickson, R. L. Inhibitors of the Protease from Human Immunodeficiency Virus: Design and Modeling of a Compound Containing a Dihydroxyethylene Isostere Insert with High Binding Affinity and Effective Antiviral Activity. *J. Med. Chem.* **1991**, *34*, 2344–2356.
(45) Doweyko, A. M.; Mattes, W. B. An Application of 3D-QSAR to the Analysis of the Sequence Specificity of DNA Alkylation by Uracil Mustard. *Biochemistry* **1992**, *31*, 9388–9392.
(46) Cartesian coordinates of HIV-1 PR and bound inhibitors **4** and **52** were provided by A. Wlodawar.
(47) *MacroModel, Interactive Molecular Modeling System, V 2.5*; Clark Still, Columbia University, 1989.
(48) *Alchemy III, 3D Molecular Modeling Software*; TRIPOS Associates, Inc., 1992.
(49) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565–577.
(50) 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.