Developing a Dynamic Pharmacophore Model for HIV-1 Integrase

Heather A. Carlson,*,† Kevin M. Masukawa,† Kathleen Rubins,‡ Fredric D. Bushman,‡ William L. Jorgensen,§ Roberto D. Lins,|| James M. Briggs,|| and J. Andrew McCammon†

Department of Chemistry and Biochemistry and Department of Pharmacology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0365, Infectious Diseases Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, Chemistry Department, Yale University, P.O. Box 208107, New Haven, Connecticut 06520-8107, and Department of Biology and Biochemistry, Houston Science Center, Room 402D, University of Houston, Houston, Texas 77204-5513

Received June 22, 1999

We present the first receptor-based pharmacophore model for HIV-1 integrase. The development of "dynamic" pharmacophore models is a new method that accounts for the inherent flexibility of the active site and aims to reduce the entropic penalties associated with binding a ligand. Furthermore, this new drug discovery method overcomes the limitation of an incomplete crystal structure of the target protein. A molecular dynamics (MD) simulation describes the flexibility of the uncomplexed protein. Many conformational models of the protein are saved from the MD simulations and used in a series of multi-unit search for interacting conformers (MUSIC) simulations. MUSIC is a multiple-copy minimization method, available in the BOSS program; it is used to determine binding regions for probe molecules containing functional groups that complement the active site. All protein conformations from the MD are overlaid, and conserved binding regions for the probe molecules are identified. Those conserved binding regions define the dynamic pharmacophore model. Here, the dynamic model is compared to known inhibitors of the integrase as well as a three-point, ligand-based pharmacophore model from the literature. Also, a "static" pharmacophore model was determined in the standard fashion, using a single crystal structure. Inhibitors thought to bind in the active site of HIV-1 integrase fit the dynamic model but not the static model. Finally, we have identified a set of compounds from the Available Chemicals Directory that fit the dynamic pharmacophore model, and experimental testing of the compounds has confirmed several new inhibitors.

Introduction

Combinations of inhibitors of the reverse transcriptase and protease are currently the preferred clinical treatment for HIV infection and AIDS.[1] HIV-1 integrase is the third target for anti-HIV drugs. Inhibitors of the integrase are attractive because it is one of the earliest steps in the viral lifecycle and there is no native homologous process in the host cell. The integrase inserts viral DNA into the host chromosome, catalyzing both the 3′-preprocessing of the viral DNA and its insertion into the host DNA.[1−3]

Integrase is a difficult system for drug development. Roughly 50−100 copies of integrase complex the viral DNA[3] (along with several other host and viral proteins)[2,4] to create the active pre-integration complex that is transported into the nucleus and incorporated into the host genome. The integrase exists as a dimer, tetramer, or higher-order mulitmer.[3,5,6] Domain-domain interactions and enzyme-DNA complexation are only roughly described through photo-cross-linking experiments.[5,7−9] The enzyme is comprised of three domains: the zinc-binding N-terminal multimerization domain (1−50), the catalytic core domain (50−212), and the C-terminal DNA-binding domain (212−288). While structures are unavailable for the complete integrase, the structures of the N- and C-terminal domains have been described by NMR,[10−12] and several crystal structures are now available for the catalytic core.[13−16]

Previously, we reported the first molecular dynamics (MD) studies of the catalytic core domain of HIV-1 integrase.[16] We now report the first receptor-based pharmacophore model for the active site of the integrase based on those MD simulations. We also announce a new method for developing "dynamic" pharmacophore models that account for the inherent flexibility of an active site. The goal of the dynamic model is to identify compounds that complement the receptor while causing a minimal disruption of the conformation and flexibility of the active site, potentially reducing the entropic penalties of binding incurred by the protein.[17,18]

The dynamic method uses a series of snapshots from a MD simulation, creating an ensemble of appropriate protein configurations. Those configurations describe the inherent flexibility of the uncomplexed receptor under the influence of explicit solvent molecules. For the development of the pharmacophore, all solvent molecules and any counterions are removed from the snapshots, providing a collection of "bare" active sites. Binding regions on the surface of each protein are determined for various functional groups through a series of multi-unit search for interacting conformers (MUSIC) simulations. The MUSIC procedure simultaneously performs multiple, gas-phase minimizations for

* To whom correspondence should be addressed. Phone: 858-822-1469. Fax: 858-534-7042. E-mail: hcarlson@mccammon.ucsd.edu.
† University of California, San Diego.
‡ The Salk Institute for Biological Studies.
§ Yale University.
|| University of Houston.

10.1021/jm990322h CCC: $19.00 © 2000 American Chemical Society
Published on Web 05/11/2000

hundreds of probe molecules within the active site using the BOSS program—a leading Monte Carlo (MC) simulation package employing a Boltzmann-weighted, Metropolis sampling algorithm.[19] After minimization, the probe molecules are clustered within the local minima in the bare active site; the clustered probes define appropriate, complementary binding regions for that specific functionality. (For example, methanol molecules were used as probes to define binding positions and orientations for hydroxyl groups.) The results of the MUSIC simulations for each snapshot of the protein are then overlaid to reveal conserved binding regions that are highly occupied over the course of the MD simulation despite the inherent motion of the active site. These conserved binding regions define association to the active site that should, in theory, still allow for almost the same flexible motion.

We pursued compounds that complement the essential residues D64 and D116 without associating to the ion (even potentially displacing the catalytic ion as a means of inhibition). Inhibitors that bind to the essential residues may reduce the rapid rate of escape mutants known to be a serious complication in treating AIDS. Future studies are planned to determine a dynamic pharmacophore model for the active site with the ions present.

HIV-1 integrase is an interesting test case for the dynamic method since it has an active site that is shallow, solvent exposed, and less restricted in conformational sampling. However, we stress the difficulty of working with this system. Only recently has the first crystal structure of the integrase containing an inhibitor in the active site appeared.[20] In that study, the inhibitor was found to associate to essential residue E152 through donating a hydrogen bond. Our pharmacophore models are comprised of hydrogen-bond donor sites for association to D64 and D116. With so little known about the modes of inhibition of HIV-1 integrase, it is possible that binding to D64 and D116 may be a minor mode of inhibition for the system. Even so, identifying such compounds is a benefit to cocrystallization efforts with the catalytic core domain of HIV-1 integrase. These types of compounds would be a wonderful complement to the inhibitor that binds to E152. One could imagine tethering two compounds that bind to different essential residues, creating an outstanding inhibitor.

The use of multiple protein structures is an improvement over the standard technique of using a single crystal structure. The dynamic model is able to fit known inhibitors while a "static" model created from a closely related crystal structure fits very few. Also, we have identified new inhibitors of HIV-1 integrase using the dynamic pharmacophore model to screen the Available Chemicals Directory (ACD).

Computational Methodology

Valid Conformational Structures from the MD Simulations. Comparisons between the MD structures and the new crystal structures are presented. To generate a complete structure for the MD studies,[16] the unresolved flexible loops were modeled after the crystal structure of the catalytic domain of the homologous Avian Sarcoma Virus (ASV) integrase.[21] After completion of the MD simulations, two crystallographic studies were published with complete structures of the catalytic domain in the presence and in the absence of a catalytic divalent ion.[13,14] Crystallization conditions, such as changes in the unit cell and use of cacodylate, were shown to influence the structure; similar behavior was seen previously in ASV integrase.[21]

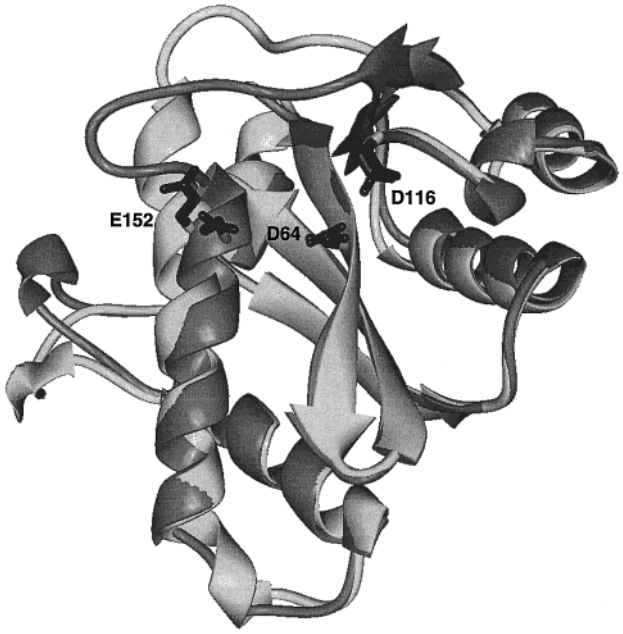


Figure 1. Comparison of the two recent crystal structures that have a resolved loop over the active site (seen at the top of the figure). This view was constructed with MidasPlus[22,23] by a least-squares fit of Cα in similar regions (an RMS of 0.45 Å for residues 60−139 and 155−186). The structure solved by Goldgur et al.[14] (monomer B of 1BIS) is shown in light gray with the active site loop oriented back and away from the active site residues, shown in black. The structure solved by Maignan et al.[13] (monomer C of 1BI4) is shown in dark gray with the loop oriented toward the viewer and over the active site. Though the structures agree very well in the placement of D64 and D116 on the right, the position of the loop significantly alters the placement of E152 on the left.

For example, the two new crystal structures of HIV-1 integrase disagree in the previously unresolved regions, Figure 1. The flexible loop adjacent to the active site is oriented away from the catalytic residues in the crystal structure solved by Goldgur et al.[14] (light gray in Figure 1). The structure solved by Maignan et al.[13] (darker gray in Figure 1) places the flexible loop over the top of the active site and positions E152 closer to the other catalytic residues. The structure solved by Goldgur et al. has much higher B-factors through the flexible region (Cα of residues 141−156 in the range of 57.96−90.03 Å^2) than does the structure by Maignan et al. (22.17−38.37 Å^2).

In Figure 2, an overlay is presented that compares the two crystal structures to the MD structure after 500 ps of simulation time (shown in black). The MD structure is more similar to the structure by Maignan et al.[13] in the flexible loop regions. However, the MD structure has a longer central helix, like the Goldgur et al. structure,[14] with at least one additional turn more than the Maignan et al. structure. The additional turn of the helix formed during equilibration and remained over the entire length of the MD simulation; it was not part of the initial model and is a rare example of an MD

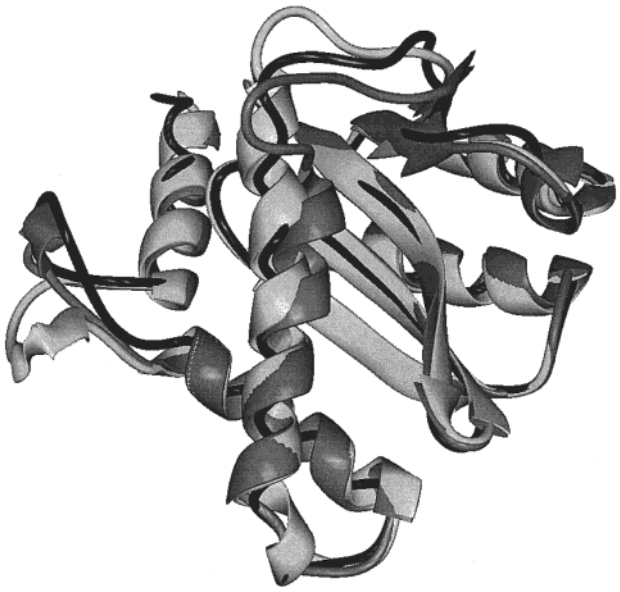


Figure 2. Comparison of the new crystal structures to the protein structure after 500 ps of sampling in the MD simulation (shown in black). Again, the structure solved by Goldgur et al.[14] is shown in light gray (RMS fit of 1.38 Å to MD sturcutre for Cα of reisdues 60−139 and 155−186). The structure solved by Maignan et al.[13] is shown in dark gray (RMS fit of 1.44 Å to MD structure for Cα of residues 60−139 and 155−186). The flexible loop on the left was originally placed in a "down" orientation resembling the structure by Goldgur et al. but relaxed over time to an "up" orientation like the structure by Maignan et al. Discussions involving the flexible loop over the active site appear in the text.

simulation properly predicting secondary structure. Therefore, the collection of structures from the MD simulations appears to provide a reasonable ensemble of configurational states for this study. It is also valid to suggest using the new crystal structures in the pharmacophore modeling. While that was not done in this study (because much of the study was completed prior to their availability), it is the subject of a follow-up study appearing elsewhere.[24]

MUSIC and Multiple-Copy Simulations. Multiple-copy minimizations are a standard practice in computational drug design[17,18,25−30] with a history of successful applications.[31−42] This simulation technique floods the active site of a protein with many copies of noninteracting probe molecules. MD, MC, steepest-descent, or other sampling methods can be used to simultaneously minimize all probe molecules into local minima. In addition to the position and orientation of the clustered probes, information about the relative favorability of the binding regions can be evaluated by the interaction energy or from the number of probes within the minimum. Smaller clusters indicate more stringent requirements for favorable interactions, but larger spread or less ordered orientations will provide a greater degree of flexibility in the ligand model.

MUSIC became available with advances in the 3.8 version of the BOSS program.[43,44] The BOSS program performs MC statistical mechanics simulations for systems of up to 25 solutes in periodic solvent boxes, solvent clusters, binary solvent mixtures, or dielectric continuums. Scalable solvent−solvent, solvent−solute, and solute−solute energetics make MUSIC possible. Probe molecules (designated as solvent in a MUSIC simulation) are user-defined, small molecules such as methanol to represent hydroxyl groups, benzene for aromatic groups, or acetone for carbonyl groups. Given the reliability[45,46] and comprehensive nature of the OPLS force field[47,48] provided with BOSS, options for modeling unique solvent molecules are almost unlimited: united atom, all-atom, fully flexible, partially flexible, or rigid requirements are all properly treated.

It should be noted that the side chains of the protein target could be rigid, partially flexible, or fully flexible in a MUSIC simulation. In this application, the protein configurations were held rigid for the development of the dynamic pharmacophore because the flexibility of the system was represented through the use of many snapshots from the MD simulations. Studies have appeared that demonstrate the need for protein flexibility in ligand docking and multiple-copy simulations in order to achieve proper results;[49,50] however, the large majority of publications continue to report the use of only one rigid structure for multiple-copy simulations.[17,18,29,33,34,36−42] The static model for HIV-1 integrase was created from the crystal structure used to initiate the MD and represents the standard application of multiple copy minimizations in computational drug design.

Results and Discussion

Verifying the Use of MUSIC Simulations with the HIV-1 Integrase System. To verify the reliability of the MUSIC procedure, we have compared the positions of crystallographic water to calculated binding regions for water probe molecules to the surface of the crystal structure. A strict comparison of all calculated binding sites would not be accurate because of the added loop regions. Furthermore, multiple-copy simulations are known to find many more binding sites for solvent than are observed through crystallographic methods.[51] Instead, a more accurate test is to ask if MUSIC can accurately reproduce the most favorable crystallographic waters with the lowest B-factors. Some water molecules have strong interactions with a protein; approximately 50% cannot be removed from a protein surface when introducing neat organic solvent in the multiple solvent crystal structure technique.[51] These tightly-held water molecules are likely to be the most influential for drug binding, and these positions are the most important to identify with MUSIC.

The binding sites determined with the MUSIC procedure compare well to the experimental water sites. For the crystallographic water with the 10 lowest B-factors (19.7−31.7 Å^2), four of the sites for water probes are within 0.5 Å of a crystallographic water molecule and five are reproduced within 1.0 Å. For the 20 lowest B-factors (19.7−35.8 Å^2), 7 sites are reproduced within 0.5 Å and 11 are within 1.0 Å. It is encouraging to note that four of the sites reproduced within 0.5 Å are within cavities, despite the tendency of the probes to minimize to the surface of the protein. For the crystallographic water molecules that were not reproduced, two were on the dimer interface and one was in an interior cavity. These positions are inherently difficult to reproduce given the system setup and simulation protocol.

The Dynamic Pharmacophore Model for HIV-1 Integrase. We have used snapshots from our previous

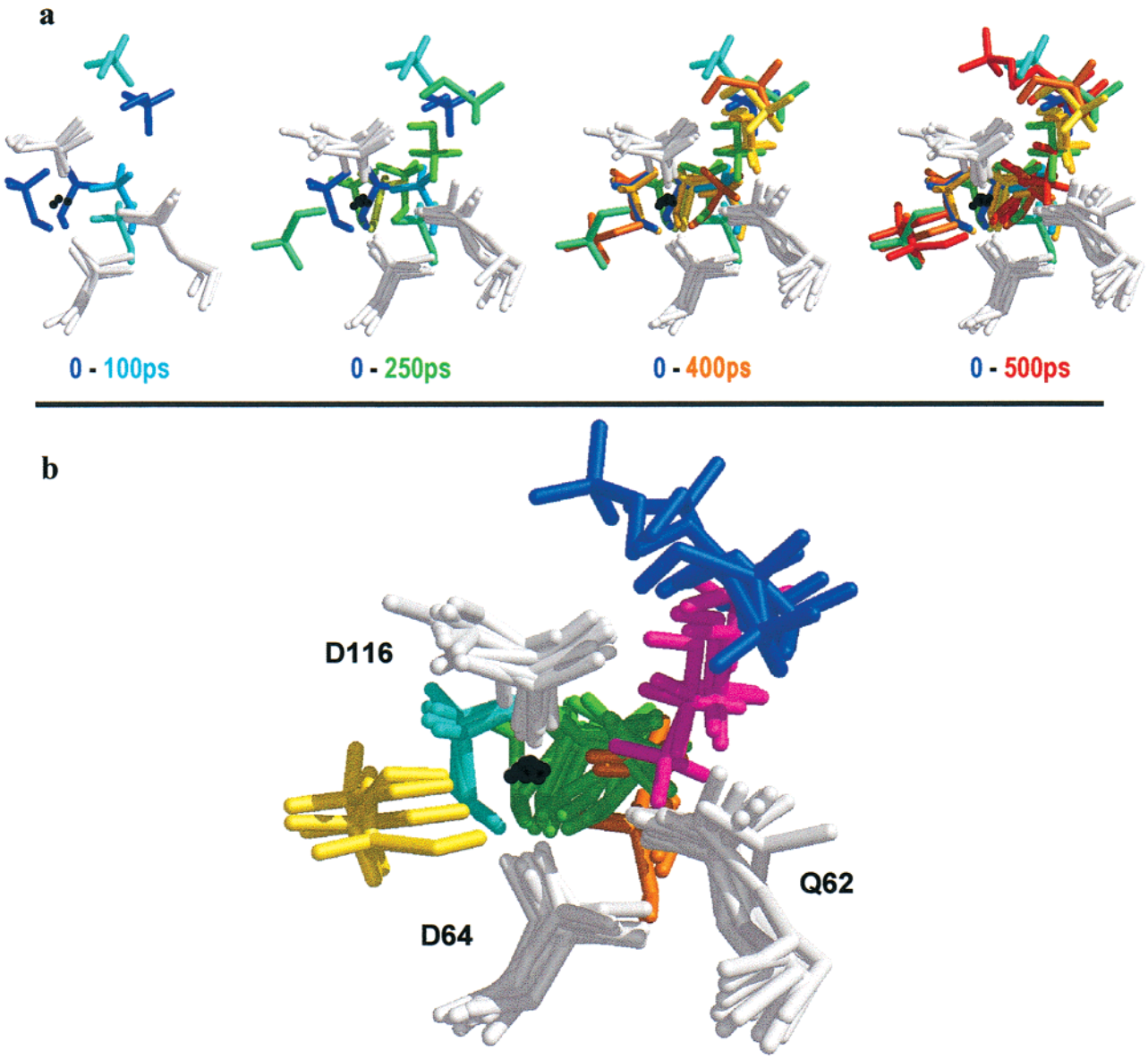
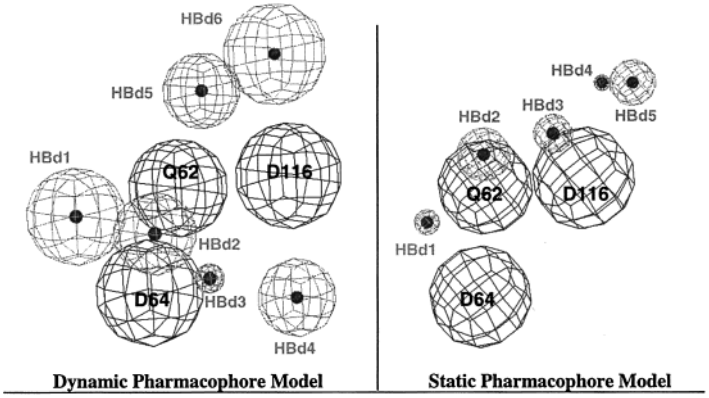


Figure 3. (a) Process for determining conserved binding regions of the methanol probe molecules. The active sites of the proteins are overlaid. Each cluster from the MUSIC simulations is represented by its probe with the most favorable interaction energy. D64, D116, and Q62 from the MD structures are shown in gray (the magnesium ion in light gray was not present during the MUSIC calculations but was used to overlay the active sites of the 11 snapshots from the MD simulations). The methanol are colored from dark blue to yellow to bright red in order of the time reference form the MD simulations (0−500 ps at 50-ps intervals). (b) Final overlay, seen at the far right in panel a, is shown with all six conserved positions for hydrogen-bond donors hightlighted in different colors. The color-coding in panel b does not correspond to the color-coding in panel a.

MD simulation of the catalytic core of HIV-1 integrase,[16] stripping away all the solvent molecules and the catalytic magnesium ion in the active site. The MUSIC procedure was used to generate binding sites for methanol probes within the active sites of the 11 protein conformations. Each resulting cluster of probe molecules can be represented by its parent (the probe in the cluster with the most favorable interaction energy to the protein, see the Computational Details section). The active sites of all 11 configurations were overlaid to identify the conserved binding regions for the parents. Conserved binding regions were required to contain at least three parents, and the parent molecules had to be identified with protein structures from early and late in the MD simulation. A site was not considered conserved if it was only observed for structures from the beginning or the end of the MD simulation.

Figure 3a,b shows the overlay process and displays the six conserved binding regions for the methanol probe molecules. None of the conserved sites contained parents from all 11 MUSIC simulations (the maximum was 9 in the green region in Figure 3b). Also, none of the 11 protein configurations identified all 6 conserved regions. This implies that the binding regions are indeed an "averaged" property of the flexible active site. The definition of conserved sites may appear somewhat subjective because they are identified through viewing the overlaid structures. The following criteria, which may not be obvious in the figure, clarify the assignment of the 6 regions. There is no potential ambiguity in the designation of the yellow and orange probes. However, the blue and magenta regions in Figure 3b could have

Table 1. Characteristics of the Dynamic Pharmacophore Model Based on the MD Simulations and the Static Pharmacophore Model Based on the Crystal Structure[a]



	dynamic (Å)				static (Å)			
	radius	*X*	*Y*	*Z*	radius	*X*	*Y*	*Z*
HBdonor1	1.316	7.249	4.783	2.405	0.364	5.127	−1.955	−2.406
HBdonor2	1.088	9.543	4.462	1.849	0.748	7.612	−2.655	−2.819
HBdonor3	0.388	11.037	3.382	1.497	0.534	9.487	−2.311	−3.876
HBdonor4	1.106	10.877	1.845	−2.410	0.250	11.464	−0.353	1.304
HBdonor5	1.088	8.232	7.622	−1.851	0.692	12.149	−0.340	−0.079
HBdonor6	1.494	9.072	8.404	−4.178				
Q62	1.5	5.842	4.291	−3.330	1.5	6.573	0.345	2.229
D64	1.5	7.819	2.218	0.311	1.5	4.765	1.167	−1.907
D116	1.5	10.461	5.219	−2.795	1.5	9.278	0.472	−1.425

[a] Atoms with gray spheres are hydrogen bond donors, and the black spheres are excluded volumes to avoid steric clashes with side chains in the active site.

been specified as one large binding site rather than two smaller ones; the same could be done for the cyan and green probes. The hydroxyl groups of the blue and magenta probes are evenly distributed around one of the Oδ of D116, but they are designated as separate binding sites because the degree of order is different for the two groups. The magenta probes are very ordered with all the methyl groups oriented in the same direction, but the blue probes bind to D116 with a great degree of freedom in the orientation of the methyl groups. Also, the blue and magenta probes are distinctly located in two regions; the order suddenly changes between the probes above D116 and those to the right of the Oδ of D116 in Figure 3b. Two criteria were used to support the cyan group as a separate site from the green probes. First, there is remarkable agreement in the positions and orientations of the three parents that make up the cyan group. Second, MUSIC calculations for the corresponding three snapshots from the MD simulation determined two separate clusters of methanol probes bound to the same Oδ of D64, one for the cyan binding region and the other to the green group. The third frame of Figure 3a shows dark blue, light orange, and light red parents in both the cyan and the green conserved regions in Figure 3b.

On the basis of the Cartesian coordinates from the 11 overlaid protein configurations, the average position for Cγ of D64, Cγ of D116, and Cδ of Q62 were used as the centers for three excluded volumes with radii of 1.5 Å. The excluded volumes were used in the pharmacophore model to eliminate compounds with steric conflicts with the protein. Though the conserved binding regions are specifically for hydroxyl groups, the criteria were extended to include any oxygen, nitrogen, or sulfur that could donate a hydrogen bond. The center of each hydrogen-bond donor site in the pharmacophore model was equal to the average position of the methanol oxygens from the parent molecules in each conserved binding region seen in Figure 3b. The radii of the hydrogen-bond donor sites were set to double the RMS deviation of the parent oxygens. This resulted in a pharmacophore model with nine sites: three excluded volumes and six hydrogen-bond donor sites. The dynamic pharmacophore model is given in Table 1.

The Static Model. Table 1 also presents the static pharmacophore model based on the crystal structure used to initiate the MD (the same crystal structure used to predict crystallographic water positions). Again, a MUSIC calculation with methanol molecules was used to identify hydrogen-bond donor regions. Seeking to develop the static model in as similar a fashion as possible to the dynamic model, the centers of the hydrogen-bond donor sites were placed at the average position of the oxygen atoms of all the probes in an individual cluster. The radius of each site was equal to twice the RMS deviation of all the probes within the cluster. Because the static model is based on a single configuration of the protein, much smaller variability (radii) is seen in the static structure. The static model will reflect the error of using a single protein conformation as well as any errors in the crystal structure that it is based upon. Also, it is clear in Table 1 that the binding sites in the dynamic model are appropriate for two independent aspartates each with its own set of sites resembling bidentate complexation. This independent pattern for the two side chains is expected in the dynamic model but is not seen in the static model. The binding pattern in the static model points to interference between D64 and D116 when minimizing the probe molecules.

Testing against a Set of Inhibitors from the Literature. One of the three-point pharmacophore

models from the literature is a subset of both the dynamic model and the static model. The three-point model was based on the oxygen positions calculated for conformers of NSC 158393; it consists of three oxygen/nitrogen centers in a triangle with sides of 4.76 ± 0.4, 5.62 ± 0.7, and 2.87 ± 0.7 Å.[52] These sites agree within error tolerances to HBd1, HBd2, and HBd5 of the dynamic model and HBd2, HBd3, and HBd5 of the static model. Eight compounds from the study based on NSC 158393 were chosen for our test set. The compounds that fit the three-point model by recognition of hydrogen-bond donor groups (such as hydroxyls) also fit the more specific dynamic model. This implies that the protocol we employ for fitting and generating conformations of the compounds, using the Catalyst program,[53] is in agreement with previously published studies using pharmacophore models and database searching.[52,54−57]

The structures of the 59 compounds in the test set (complete listing in Table 2) are provided as Supporting Information. These compounds from the literature were chosen based on their activities as inhibitors of HIV-1 integrase. To test the selectivity of the pharmacophore models, noninhibitory compounds with structures similar to known inhibitors were included. Twenty compounds in the test set had IC_{50} values under 1 μM for 3′-preprocessing or strand transfer (referred to as very active compounds in Table 2 and the following discussions). An additional 26 compounds had IC_{50} values between 1 and 35 μM for both catalyzed processes (active compounds). Three ineffective inhibitors (IC_{50} values of 88−224 μM) and 10 noninhibitors were also included. More noninhibitors in the set would provide a better counter-test of the model; however, most reports in the literature focus on successful compounds and present few failed inhibitors. All compounds chosen to test the pharmacophore models contain at least four hydrogen-bond donors. Those inhibitors are more likely to bind to D64 and D116 with their many hydrogen-bond donors complementing the carboxylate groups.

There are many known inhibitors of HIV-1 integrase that contain primarily hydrogen-bond acceptor groups and few hydrogen-bond donors. Because of their composition, such compounds do not fit the dynamic and static pharmacophore models and were not included in the test set. Exclusion of these compounds is most likely appropriate because a crystal structure of ASV integrase has shown that an inhibitor with mostly hydrogen-bond acceptors binds to a face far from the active site.[61] Furthermore, if these compounds do interact with the active site, it will require involvement of the divalent ions that have been excluded in the current models.

Though many compounds fit the dynamic queries, *none fit the static model*. Of course, the radii are much smaller in the static model, which may account for its poorer performance. Doubling the radii in the static model provided a set of hydrogen-bond donor sites with radii ranging from 0.5 to 1.496 Å (radii in the dynamic model range from 0.388 to 1.494 Å). With the doubled radii, the static model only fit 5 of the 59 test compounds including one noninhibitor, see Table 2. The poor performance of the static model implies that compounds identified with the dynamic model are more likely to inhibit HIV-1 integrase than compounds identified with the static model. It appears that inclusion of the flexibility of the side chains allows for a more accurate description of binding to the active site (possibly the use of an ensemble of structures smoothes out small inaccuracies in the homology-modeled loop adjacent to the active site).

The dynamic pharmacophore model has an excellent rate for fitting the very active compounds (15 correct

Table 2. Performance of the Pharmacophore Models in Fitting Compounds from the Literature That Have Been Tested for Inhibitory Activity[a]

compound[a]	IC_{50} values 3′	IC_{50} values ST	dynamic	static	static with doubled radii
chicoric acid[b]	0.15	0.13	√		
107[b]	0.23	0.11	√		
4,5-DCQA[b]	0.25	0.46	√		
81[b]	0.4	0.2	√		√
67[b]	0.5				
71[b]	0.5				
85[b]	0.5	0.2	√		√
3,5-DCQA[b]	0.64	0.66	√		
3,4-DCQA[b]	0.79	0.54	√		
1,5-DCQA[b]	0.68	1.08			
NSC 118695[c]	0.9	0.3	√		
quercategetin[d]	1.3	0.6			
105[b]	0.98	0.81	√		
NSC 64205[c]	1.1	0.5	√		
NSC 158393[e]	1.5	0.8	○	○	○
NSC 607319[c]	1.4	1.0	√		
NSC 309121[c]	1.7	1.0	√		√
68[b]	1.7		√		
myricetin[d]	2	0.6			
doxorubicin[b]	0.9	2.4	√		very active compds
purpurogallin[f]	2.1				active compds
111[b]	2.3	1.1			
NSC 310217[e]	2	1.5			
NSC 64452[c]	1.2	3.6	√		
quinalizarin[d]	4	1			
83[b]	3.3	1.9	√		
NSC 261045[c]	2.3	4.1			
90[b]	1.38	4.71	√		
NSC 642710[e]	5.3	5.0	√		
NSC 115290[g]	5				
ellagic acid[f]	5.1				
115[b]	6.7	5.2			
mitoxantrone[b]	3.8	8.0	√		√
rosmarinic acid[b]	9		√		
tyrophostin A51[d]	10	3			
110[b]	9.1	5.8			
1,4-DCQA[b]	9.5	7.8	√		
141[b]	11.6	7.9			
hypericin[f]	10		○	○	○
TMS[f]	17	5			
NSC 318213[c]	23.9	14.0			
66[b]	21.4	5.4	√		
NSC 233026[c]	20.6	19.7			
NSC 371056[e]	29.9	16.5	√		
NSC 48240[e]	26	20.6	√		
97[b]	33	33	√		active compds
chlorogenic acid[b]	87.8	45.8			ineffective compds
13[g]	120	96			
NSC 641547[e]	224	134			
NSC 674503[c]					
NSC 635971[e]			√		
NSC 642651[c]			√		
112[b]			√		
9[g]			√		
52[g]			√		√
NSC 281311[c]			√		
103[b]			√		
5[e]					
NDGA[f]			√		

[a] Compounds that fit are marked with √, and the two compounds that do not and should not fit are marked with ○. The numbers listed for some of the compounds are the labels given in the referenced papers. Compounds are ordered by inhibitory activity. [b] Ref 58. [c] Ref 56. [d] Ref 59. [e] Ref 52. [f] Ref 60. [g] Ref 54.

Table 3. Analysis of the Ability of the Dynamic Pharmacophore Model To Fit Known Inhibitors and Discriminate against Ineffective Compounds

% yield = (correct predictions/all compds fit to model)
% actives = (correct predictions/all active compds in test set)
enrichment = (% yield)/(all active compds in test set/all in test set)[a]

	% yield	% active	enrich-ment
active compds = 20 very active compds[b]	44	75	1.30
active compds = all 46 active compds[c]	77	59	0.99

[a] Enrichment of 1.0 indicates that the model is fitting compounds with the same ratio of active compounds as exist in the original test set. Enrichment values over 1 reveal that the model is identifying a higher percentage of active compounds (model is preferentially identifying active compounds). Enrichment less than 1 imples that model is preferentially fitting noninhibitors. [b] 15 correct predictions, counting nonfit of NSC 158393. [c] 27 correct predictions, counting nonfit of NSC 158393 and hypericin.

predictions out of 20 compounds), see Table 2. For the active compounds, 12 out of 23 are correctly predicted, and 62% of the ineffective compounds fit the dynamic model. The enrichment values in Table 3 show that the model is specific for the very active compounds, but there is no discrimination against the ineffective compounds when all active compounds are included in the analysis (and no preference for them either). Additional factors that influence binding, such as desolvation effects or conformational penalties of the ligand, could be added in future studies to better discriminate against less effective compounds.

Compounds That Fit the Dynamic Pharmacophore Model. All three tetracycline compounds in the test set fit the dynamic pharmacophore model. This is particularly encouraging because tetracyclines are thought to bind in the active site.[56] Furthermore, their activity is metal independent, implying that they may interact directly with side chains in the active site.

Many of the inhibitors with the lowest IC_{50} values are bis-catechols. Of the 24 bis-catechol compounds in the test set, the dynamic model fits 19 (including the one noninhibitory bis-catechol). Methylation of the hydroxyls of bis-catechol compounds eliminates their activity, so the ability to donate hydrogen bonds is most likely central to their inhibitory activity.[58] For the related cinnamoyl derivatives, a general correlation was found for increased inhibitory activity with increasing numbers of hydroxyl groups.[62] Studies have indicated that bis-catechol compounds bind within the active site by noting their indiscriminate blockage of any nucleophilic attack on the viral DNA and their ability to inhibit homologous integrase enzymes with the same pattern of catalytic residues.[63] Escape mutants of the most effective bis-catechol inhibitors (L-chicoric acid and dicaffeoylquinic acids, DCQAs) are the most compelling evidence that these compounds associate to the active site, see Figure 4. The G140S mutation gives rise to resistance to L-chicoric acid; also, the C65A mutation decreases the activity of L-chicoric acid and of 3,4-DCQA.[63] The G140 and C65 residues lie—almost linearly—on opposite sides of D116 and D64. It appears that L-chicoric acid and DCQAs must associate to integrase in the vicinity of the two essential residues used in the development of the dynamic pharmacophore model. There are mixed reports regarding the metal dependence of inhibition by bis-catechol compounds.[63,65−67] While a lack of metal dependence negates the role of chelating ions during inhibition, some metal dependence can be related to the known conformational changes that the ions induce in the integrase.[68,69]

One of the most interesting characteristics of the bis-catechol inhibitors reported throughout the literature is that the linkers between the two aromatic rings can vary widely in length, composition, and chirality.[58,62,67] This implies that the linker may not be part of the pharmacophore.[70] As can be seen in Figure 5, the fit of these compounds into the dynamic pharmacophore does not involve the linker region, which would be oriented out into solution. In fact, this may explain why nordihydroguaiaretic acid (NDGA) is the only inactive example of a bis-catechol compound. NDGA has a completely hydrophobic linker region, which would be unfavorable if oriented into solution. Furthermore, Figure 5 reveals an association of the aromatic rings in sort of skewed T-shape stacking. This is in agreement with Mazumder et al., who proposed a conformer for curcumin with π-stacking of the phenyl rings to mimic a catechol arrangement of the hydroxyl groups.[71]

Of the five bis-catechol compounds that were not fit, four are relatively rigid (TMS, NSC 318213, and cyclolignans 67 and 71). More flexibility in the pharmacophore model could possibly allow it to fit those compounds as well as some flavones (see below).

Compounds That Do Not Fit the Dynamic Model. Many of the compounds that cannot be fit to the dynamic pharmacophore are relatively rigid, polycyclic aromatic compounds such as the one seen to bind far from the active site of ASV integrase[61] (for example: NSC 158393, purpurogallin, ellagic acid, hypericin, NSC 233026, and several flavones and anthraquinones). Two studies have concluded that hypericin most likely does not interact with the active site,[60,72] and we can assume that hypericin (and possibly related compounds) *should not* fit the model. Also, NSC 158393 has been shown to bind synergistically with bis-catechol compounds;[65] since the bis-catechols most likely bind to the active site, NSC 158393 must bind elsewhere on the protein to bind simultaneously. Polycyclic aromatic compounds with many hydroxyl groups do not all act by binding to the same site,[60] so it is possible that not all of these compounds will bind outside the active site like hypericin and NSC 158393. This family of compounds can chelate metal ions, and it has been suggested that they may form ternary complexes with the protein and Mg^{2+} or Mn^{2+} ions.[54,60,66,73] None of the flavones fit the dynamic model (quercategetin, myricetin, purpurogallin, and NSC 641547), and only two of the anthraquinones would fit the dynamic model (doxorubicin and mitoxantrone fit the model, but quinalizarin, NSC 261045, and NSC 115290 did not). In general, a large number of hydroxyl groups are correlated with inhibitory activity of flavones and anthraquinones; but negatively charged derivatives are also active. The negatively charged derivatives could only bind to the active site through coordination to the divalent ions. To test the mechanism of inhibition by flavones and anthraquinones, methylated derivatives should be screened for activity as was done for the bis-catechol compounds; this would indicate whether the hydroxyl groups act as hydrogen-bond donors or acceptors.

Table 4. Experimental Testing of Compounds Identified from the ACD Using the Dynamic Pharmacophore Model[a]

compd		1 mM	200 μM	25 μM
1	**gluconic acid hydrazide**	+	−	−
2	*N*-α-carbamyl-L-arginine	−	−	
3	**1,4-phenylenebis(thiosemicarbazide)**	+	−	−
4	RCL S3,225-6	−	−	
6	4-propyl-1-(2-(3-propylureido)benzoyl)semicarbazide	−	−	
7	Bis-Tris propane	−	−	
9	hirudonin sulfate	−	−	
10	**imidurea**	+	−	−
12	Gly-Arg-AMC	−	−	
14	dihydrostreptomycin	−	−	
16	**chlorhexidine**	+	−	−
17	D-gluconyl-Val-Leu-Gly-Lys-NHET	−	−	
18	streptomycin sulfate	−	−	
20	***N-tert*-BOC-*O*-benzyl-Ser-Gly-Arg p-nitroanilide**	+	−	−
21	monohydroxyethyl trihydroxypropyl ethylenediamine	−	−	
22	pentaethylenehexamine	−	−	
23	1,4-bis(tris(hydroxymethyl)methyl)oxamide	−	−	
24	α-D-galactosamine 1-phosphate	−	−	
25	α-D-glucose 1-phosphate	−	−	
26	α-D-(+)-mannose 1-phosphate	−	−	
27	2,3-dehydro-2-deoxy-*N*-acetylneuraminic acid	−	−	
28	*N*-acetylneuraminic acid methyl ester	−	−	
29	fructosazine	−	−	
30	famotidine	−	−	
31	isomaltose	−	−	
32	maltitol	−	−	
33	thiodigalactoside	−	−	
34	lactose	−	−	
35	**psychosine**	+	−	−
36	4-(3-chlorophenyl)-1-(3-furfurylureidoacetyl)semicarbazide	−	−	
37	4,4′-hexamethylenebis(1-phenylsemicarbazide)	−	−	
38	RCL S17,722-9	−		
39	**dihydrofolic acid**	+	+	−
40	**isoarterenol D-bitartrate**	+	−	−
41	elastatinal	−	−	
42	viomycin	−	−	
43	**1,1′-(4,6-dibromo-1,3-phenylene)bis(3-(2-carbamoylphenyl))urea**	+	+	−
44	**(−)-epigallocatechin gallate**	+	+	+
45	**direct red 173**	+	+	+

[a] Inhibitory activity of the compounds are listed below as greater than (+) or less than (−) 50% inhibition at three concentrations. From the first set of 20 compounds, only 14 were still commercially available. Of the 189 compounds identified from the second, wider search of the ACD, 25 were chosen at random for experimental testing. Compounds unavailable: **5,** RCL S18,262-1; **8**, Maybridge S 01147; **11**, ethylenebisbiguanide sulfate; **13**, Leu-Arg-Pro-Aza-Gly; **15**, RCL S17,720-2; **19**, Arg-Arg-AMC.

None of the hydrazide-containing compounds in the test set (111, NSC 310217, 115, 110, 5) fit the dynamic pharmacophore model. In this family of compounds, the hydrazide functionality formed a link between two aromatic rings with the hydrazide always ortho-substituted to a hydrogen-bond donor group.[74] Even though these compounds did not fit the dynamic pharmacophore model, several compounds containing hydrazide and carbazide functionalities were identified when using the model to search the ACD.

Identifying Potentially Inhibitory Compounds in the Available Chemicals Directory. The Catalyst program[53] was used to screen the ACD against the dynamic pharmacophore model, but it was too specific and obtained no hits. Six queries were systematically generated by removing one of the hydrogen-bond donor sites. The interaction energies between the protein and each of the parent molecules were very similar, so none of the sites appeared more or less influential. Each of the queries contained five hydrogen-bond donor sites and three excluded volumes. Four of the six queries (removal of HBd1, HBd2, HBd3, or HBd4) successfully identified potential inhibitory compounds in the ACD; the queries without HBd5 or HBd6 did not identify any compounds in the ACD. Some compounds that were identified in the database searching were inappropriate and were removed from the hit list (reactive compounds, compounds with protonated acids being counted as hydrogen-bond donors, large peptides). Also, the many small peptides identified in the database searching were not pursued. This resulted in 20 novel compounds identified for experimental testing (including a few unnatural amino acid derivatives), see compounds **1**−**20** in Figure 6 Compounds **5**, **8**, **11**, **13**, **15**, and **19** were prohibitively expensive or unavailable; this limited the original set to 14 candidates for testing.

More compounds were necessary for a statistically significant testing of the dynamic model. The first set of 20 compounds was identified using a stringent comparison routine employed in the Catalyst program (the "fast" search option) which strictly compares "native" conformations of the molecules against the model. In an effort to widen the fit criteria and find additional compounds for screening, a more forgiving search routine was used in a second search of the ACD (the so-called "best" routine allows for an additional cycle of fitting to adjust the conformation of the compounds into better agreement with the model). The wider search resulted in an additional 189 possible inhibitors. A subset of 25 compounds was chosen for experimental

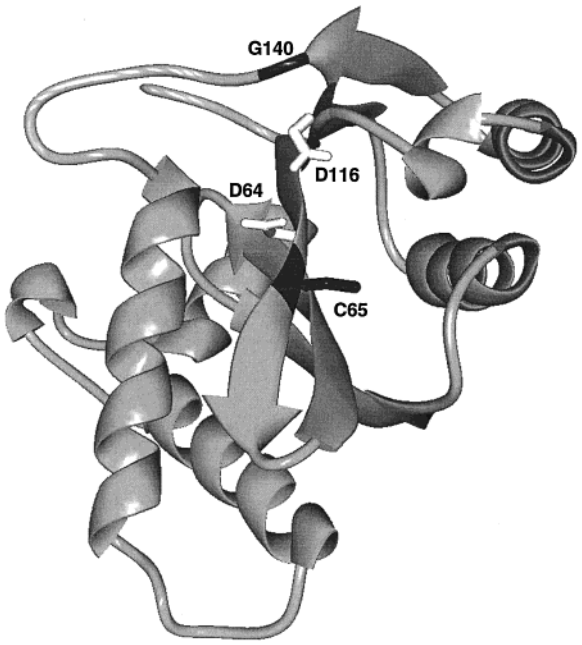


Figure 4. Mutations of C65 and G140 reduce the inhibition of L-chicoric acid and DCQAs. C65 and G140 (in black) lie on either side of essential residues D64 and D116 (in white).[63,64] It is unlikely that these two inhibitors can be influenced by these two mutations without binding in the region of the essential residues used to generate the dynamic pharmacophore.

testing based on diversity, availability, solubility, and cost (compounds **21**−**45** in Figure 6).

Experimental Testing of Compounds from the ACD. Assays of the integration reaction were carried out essentially as described in Craigie et al.,[75] except that integration product was detected using digoxygenin labeled oligonucleotides and an antidigoxygenin antibody conjugated to horseradish peroxidase. The oligo nucleotide substrates used were of sequence:

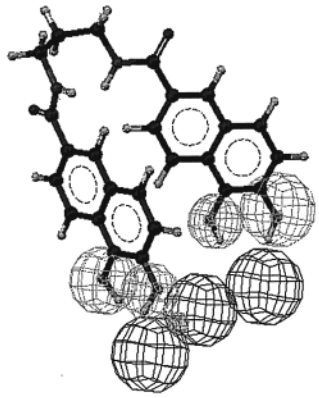
5′-ACTGCTAGAGATTTCCACACTGACTAAAGGGTC-(BIO)-3′

5′-(DIG)-GACCCTTTAGATCAGTGTGGAAAATCTCTCTAGCA-3′

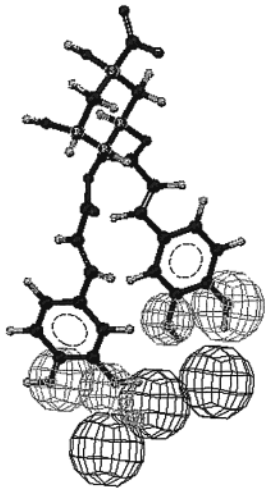
where (BIO) indicates the position of the biotin modification and (DIG) indicates the location of the digoxygenin modification. Note that this assay is different from the testing used to evaluate the inhibitors presented in Table 2. It uses pre-assembled integrase/DNA complexes and Mg^{2+} instead of Mn^{2+} as the catalytic ion. These factors increase the discrimination of the assay and probably lower the IC_{50} values.[59,76] For this reason, we define an active compound as any with an IC_{50} below 1 mM in Table 4.

Testing the compounds revealed that 11 out of 39 were active. Notably, there is a higher success rate for the compounds held more strictly to the dynamic model. The first set of compounds produced 5 successful compounds from 14 (36%). The second set of 25 compounds had a lower success rate (24%).

There is a common motif to the tested compounds. Many contain urea groups, carbazides, and guanidinium (including 7 of the 11 successful compounds). Each of these groups provides bidentate complementary to carboxylates, as would be expected. But more importantly, they are capable of donating hydrogen bonds in the same fashion as the hydroxyl groups of a catechol. It is possible that this method as applied here (and elsewhere to the new crystal structures)[24] can produce compounds that may inhibit in the same fashion as the bis-catechol compounds without the same level of toxicity.



1,3-bis(5,6-dihydroxy-2-naphthalene carboxamido)propane
(107 in Table 2)



4,5 dicaffeoylquinic acid
(4,5 DCQA in Table 2)

Figure 5. Fit of the two bis-catechol compounds with a query based on the dynamic pharmacophore model. The linker regions are oriented out into solution; this can explain why the linkers vary widely in the bis-catechol class of inhibitors.

Conclusion

We have described a new method for generating dynamic pharmacophore models that may compensate for the inherent flexibility of an active site. This method identifies conserved binding regions over multiple configurations of the active site and uses those regions to define a complementary model. The dynamic model of HIV-1 integrase compares very well to known inhibitors of the system, particularly those with very low IC_{50} values. None of the known inhibitors fit the static model based on the crystal structure alone.

Of course, we would like to continue to develop and improve the method. It could be extended to include a phase of docking studies for the compounds from the database searches. Using a comprehensive docking equation would provide the most accurate ranking of the identified compounds. Also, the searches could be

restricted to databases of drug-like molecules. Pommier and co-workers[52,54−57] have reported several successful searches of the National Cancer Institute's database, which is comprised of compounds tested for anti-viral and anti-cancer activity.

In closing, note that the dynamic model was developed from an incomplete crystal structure. Unresolved regions in crystal structures make computer-assisted drug design particularly difficult. The method of developing dynamic pharmacophore models with MD simulations and homology-modeled missing loops appears to be an effective tool—particularly for systems such as HIV-1 integrase where incomplete crystal structures were available for several years before complete structures were solved.

Computational Details

Molecular Dynamics Simulations. The full report of two 1-ns simulations, for the catalytic core of the integrase both with and without the magnesium ion present in the active site, has recently appeared.[16] Only a brief account is given here for completeness. The initial crystal structure of the catalytic core of HIV-1 integrase (residues 50−212) was completed using the InsightII program[77] to add missing side chains and to homology model the missing loops based on the alignment of backbones with ASV integrase.[21] Hydrogens were added with the HBUILD module of CHARMM v.25,[78] and the system was relaxed for 200 steps of steepest-descent minimization.

For the MD simulation, the NWchem v3.2 program[79] was used. The AMBER95 force field[80] was employed for the solutes, and a large cubic cell of SPC/E water[81] with counterions solvated the system. Nonbonded terms were truncated at 10 Å, but the Coulombic contribution was calculated with a Particle Mesh Ewald scheme.[82] The system was stepwise equilibrated in the NVT ensemble and brought to 298 K. The sampling phase was simulated under periodic boundary conditions in the NPT ensemble. A total of 500 ps of sampling was available for this study; but later the simulation was extended to 1 ns. The dynamics of the system with the magnesium ion (the system used for this study) were consistent over the 1 ns, and the behavior of the first 500 ps is representative of the entire simulation.

MUSIC Simulations. All MUSIC calculations were performed with the BOSS program, version 3.8,[43] using the all-atom OPLS force field.[47,48] Large cutoff radii were used so that no nonbonded interactions were neglected. Two minor additions were made to the BOSS code for this study. First, a flag was introduced to generate a dense collection of probe molecules with a short cutoff to the protein (1.5 Å) when setting the solvent−solvent (probe−probe) interactions to zero. Second, the output was modified to assign each probe to a given cluster and to order the probe molecules by favorable interaction energies with the protein. The first cluster is identified by the probe with the most favorable interaction energy (the parent of cluster 1). All other probes within a user-defined RMS deviation from parent 1 were assigned to cluster 1 (RMS deviations of 2.0 Å appeared to be best for water and methanol). The parent of the second cluster is the probe molecule with the most favorable interaction energy that is not part of cluster 1; all probes within the appropriate RMS that are not part of cluster 1 are assigned to cluster 2. The pattern is repeated until all probes are assigned to a local minimum. In the case of flexible probe molecules, internal strain energies should be considered in addition to the interaction energy with the protein when ordering the probes by favorable energy.[40]

Eleven configurations from the MD calculation were obtained by saving the equilibrated structure and additional configurations of the system at 50-ps intervals. Only the protein structure in each snapshot was used for the MUSIC studies. Two additional changes were made to the structures for use in the MUSIC simulations: H66 was protonated to match its ionization state without divalent ions present,[83] and the N- and C-termini were changed from charged termini to neutral methyl amides. The later change was made to avoid any false minima associated with the terminal charges at residues 57 and 210 because they are not present in the full integrase enzyme. The protein structures were not rotated or translated during the MUSIC studies, and their internal coordinates were held fixed for the simulations. An additional benefit to using a rigid protein is that domain definitions can be used in BOSS to avoid calculating the numerous internal energy contributions of the protein, which do not change over the course of the simulation, resulting in significant speed-up.

The probe molecules were also held internally rigid since the rotation of the methyl group is of minimal consequence for methanol. The maximum allowed sampling moves were displacements of 0.15 Å and rotations of 15°. A 17.0-Å sphere of probe molecules was centered at the active site (Oϵ of Q62), resulting in 358−423 copies of methanol depending on the protein configuration. A half harmonic potential was applied at the boundary of the 17.0-Å sphere (force constant of 5 kcal/mol·Å^2) to keep the probes from dissociating into the surrounding vacuum. The simulated annealing protocol was 10^6 iterations of sampling at 300 °C, 10^6 iterations at 200 °C, 10^6 iterations at 100 °C, 10^6 iterations at 0 °C, and finally 2×10^6 iterations at −100 °C.

For the reproduction of crystallographic water sites, a 27-Å sphere of TIP3P water molecules[84] was used to completely cover the protein monomer, resulting in 1967 water molecules. It should be noted that the starting positions for the water molecules came from preequilibrated water boxes available with BOSS rather than being generated through the custom solvent routine. The sampling protocol was the same as given above except that the larger number of probes with wider distribution required more iterations for minimization: 2×10^6 iterations of sampling at 300 °C, 2×10^6 iterations at 200 °C, 2×10^6 iterations at 100 °C, 3×10^6 iterations at 0 °C, and finally 8×10^6 iterations at −150 °C.

Pharmacophore Development. The clusters of probe molecules, resulting from the MUSIC studies for each snapshot, were represented by their parent molecules. The MidasPlus[22,23] program was used to overlay all 11 results to identify conserved binding regions. Each protein structure was overlaid with the 250-ps structure by an RMS fit of Cγ of D64, Cγ of D116, Cδ of Q62, and the Mg^{2+} position from the MD simulation. Four sites are required for the RMS fit, and the four atoms above are central to the active site. Though E152 is an essential catalytic residue, its position was widely sampled and was not appropriate for incorporation into the RMS fitting or the final pharmacophore model. The average positions for Cγ of D64, Cγ of D116, and Cδ of Q62 were calculated from the output of the Cartesian coordinates from the overlays. The average position and RMS deviation of the oxygens of the parent molecules from each conserved binding site were also calculated from the Cartesian coordinates from the overlays. The variability (radius) of each binding site was defined as double the RMS deviation. Each conserved binding site was described as a hydrogen-bond donor in the Catalyst program[53] by setting the atom type to oxygen, nitrogen, or sulfur with one, two, or three hydrogens.

In the case of the static model, only one conformer of the protein was used (the same model used in identifying crystallographic water positions with MUSIC using TIP3P water as probes). All protocol was exactly the same for the MUSIC studies except that the centers and radii of the hydrogen-bond donor sites were based on all methanol probes within a cluster (rather than on multiple parent molecules as in the dynamic model).

Using the Catalyst Program. When first searching the ACD with the Catalyst program, the "fast compare/fit" routine was used.[53] This is the most stringent routine that simply evaluates the geometry of the conformers of the compounds to test for fit. It was chosen to reduce the number of compounds to a reasonable size for testing. A second search routine, the

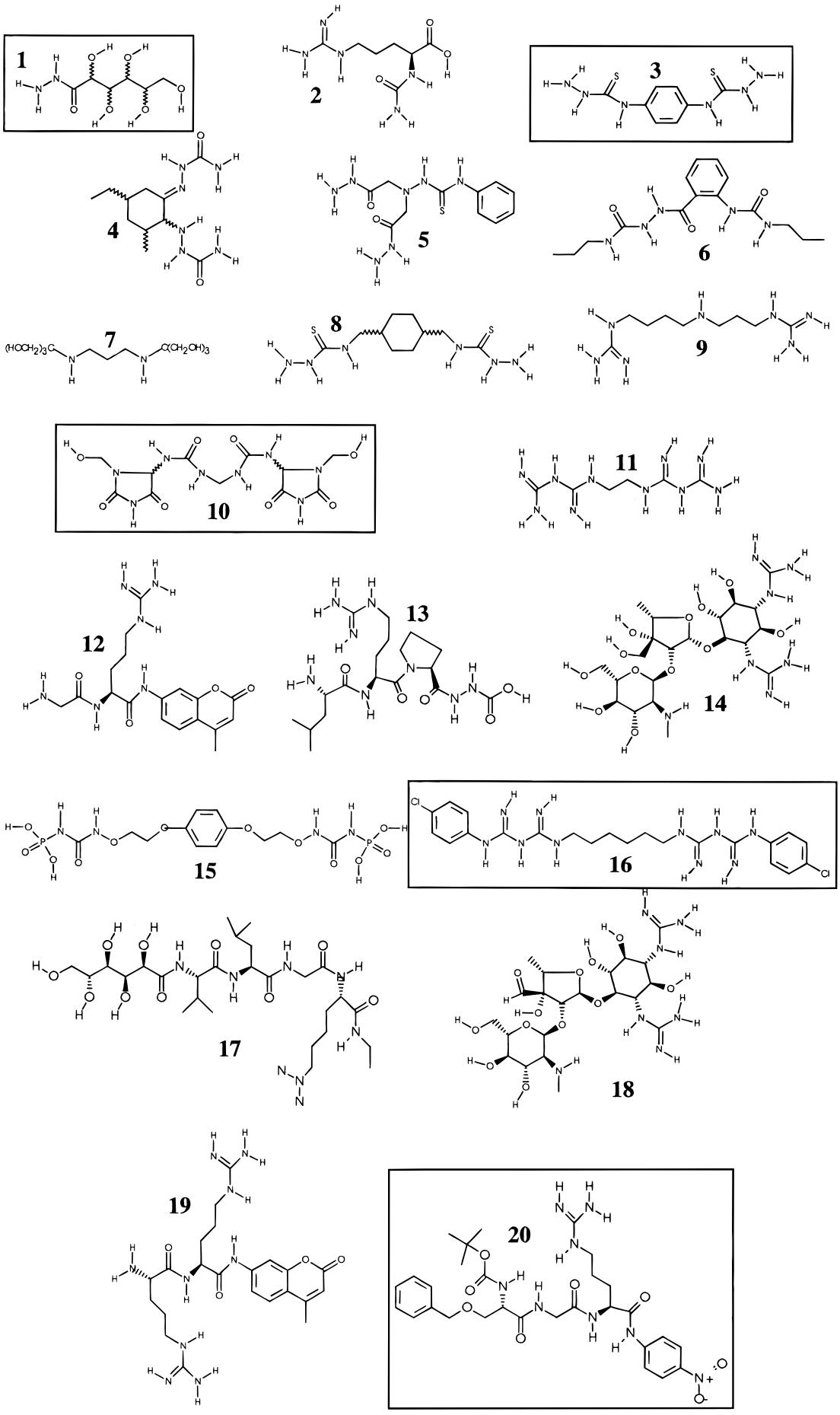


Figure 6.

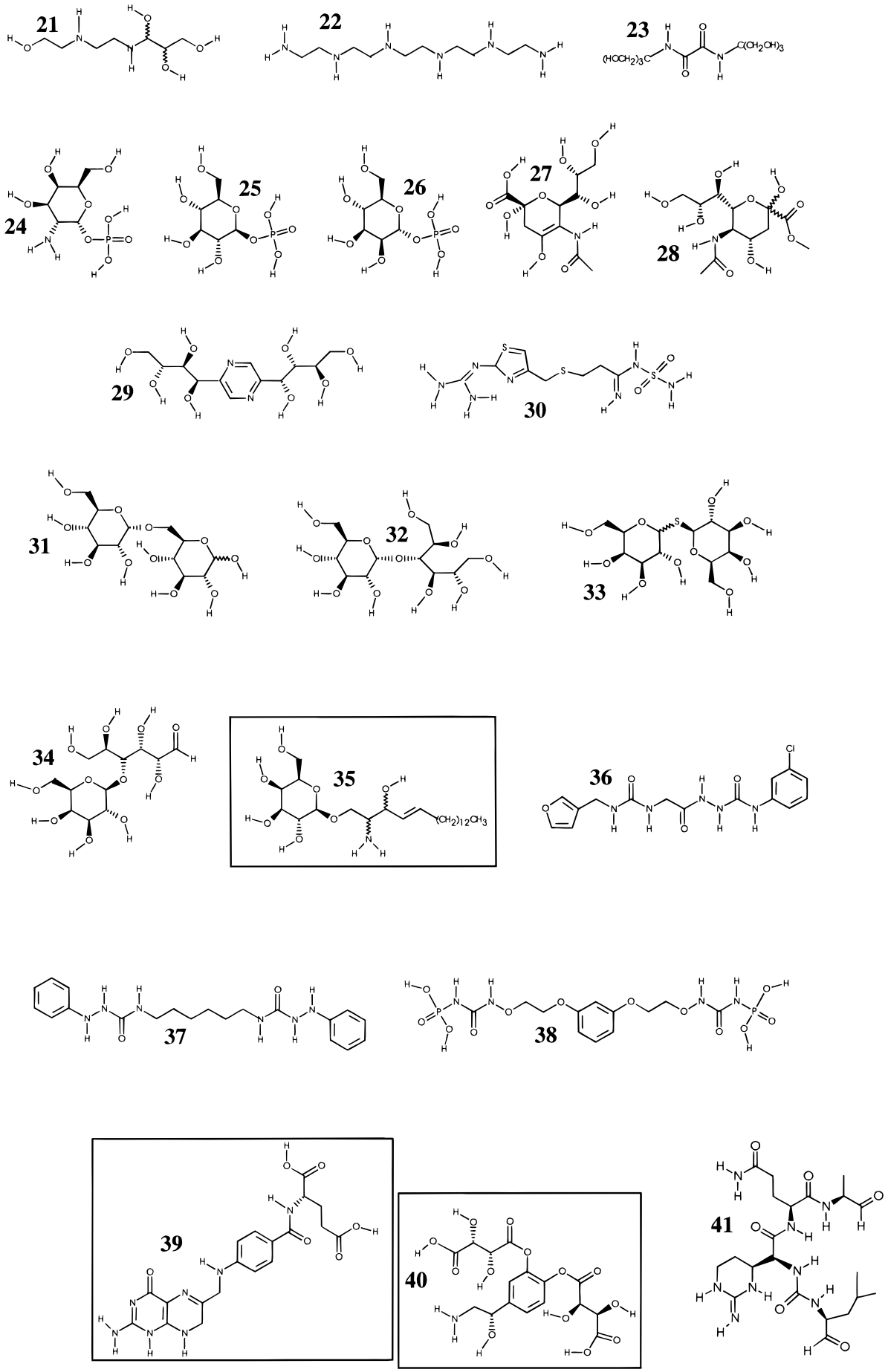


Figure 6.

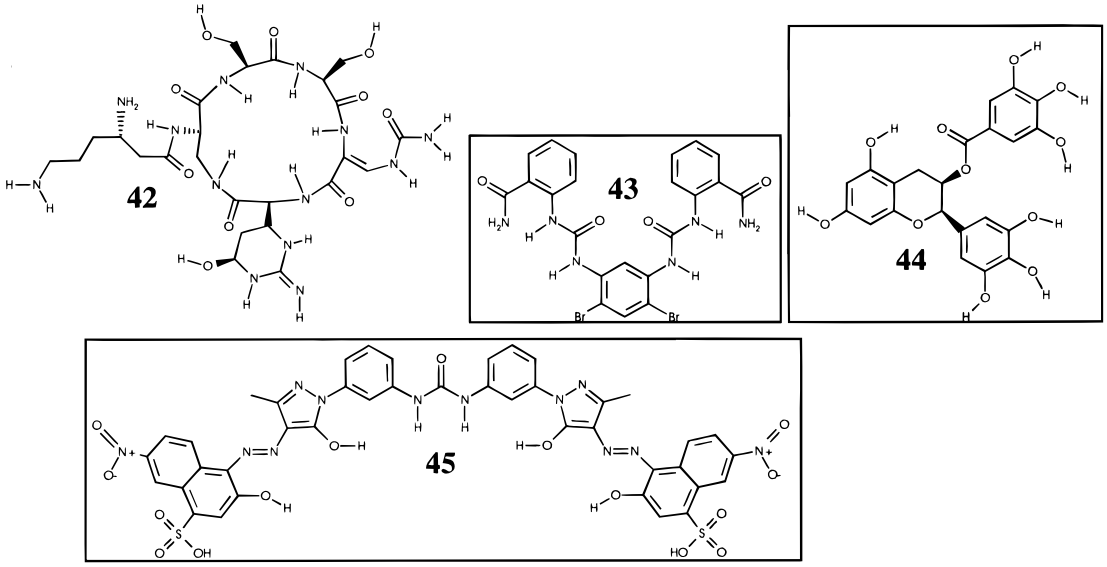


Figure 6. 20 compounds identified from a stringent screening of the ACD against the dynamic phamacophore are given in the first section (**1**−**20**). From the larger, more generous search of the ACD, 25 compounds were chosen for experimental testing (**21**−**45**). Compounds are named and drawn in the fashion used in Catalyst.[53] Groups are shown in their neutral state as they appear in the ACD. Successful compounds are highlighted with boxes.

"best" routine, is more generous and was used to generate a larger set of compounds.

For the test set of known inhibitors, the compounds were built into a user-defined database through the 3D viewer interface of the Catalyst program (structures of the compounds are given in the Supporting Information). Conformations of the inhibitors were created with the fast conformational generator employing the default limits and the intrinsic force field available within Catalyst. Inhibitors were fit to the pharmacophore models using the same fast routine that was employed in the database searching of the ACD. This was to evaluate the performance of the models in the same fashion that they were used to identify compounds for experimental testing. Since the search was later widened, it should be noted that the following additional compounds in Table 2 fit the dynamic model when using the "best" fitting routine: 71, 1,5-DCQA, myricetin, NSC 261045, 141, chlorogenic acid, NSC 641547, and NSC 674503.

Acknowledgment. We thank MSI for their generous donation of the InsightII and Catalyst software packages and the ACD database. We also thank Dr. Paul D. Kirchhoff of MSI for his assistance with Figure 6. H.A.C. is grateful to the American Cancer Society for a postdoctoral fellowship (PF-4427) and is also thankful for participation in the La Jolla Interfaces in Science Training Program, funded through the generosity of the Burroughs-Wellcome Fund. K.M.M. thanks the UCSD Student Affiliate of the American Chemical Society for a summer undergraduate research award. This work was supported by NIH Grants GM56553 and GM31749, by NSF Grant MCB-9722173, and through generous grants of supercomputer time from the National Partnership for Advanced Computational Infrastructure.

Supporting Information Available: Structures of all 59 compounds that comprise the test set of compounds from the literature (16 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Coffin, J. M.; Hughes, S. H.; Varmus, H. E. *Retroviruses*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1997.

(2) Hansen, M. S. T.; Carteau, S.; Hoffman, C.; Li, L.; Bushman, F. D. Retroviral cDNA integration: Mechanism, applications, and inhibition. In *Genetic Engineering. Principles and Methods*, Vol. 20; Setlow, J. K., Ed.; Plenum Press: New York, 1998; pp 41−63.

(3) Asante-Appiah, E.; Skalka, A. M. Molecular mechanisms in retrovirus DNA integration. *Antiviral Res.* **1997**, *36*, 139−156.

(4) Miller, M. D.; Farnet, C. M.; Bushman, F. D. Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition. *J. Virol.* **1997**, *71*, 5382−5390.

(5) Heuer, T. S.; Brown, P. O. Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* **1998**, *37*, 6667−6678.

(6) Zheng, R.; Jenkins, T. M.; Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13659−13664.

(7) Esposito, D.; Craigie, R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interactions. *EMBO J.* **1998**, *17*, 5832−5843.

(8) Jenkins, T. M.; Esposito, D.; Engelman, A.; Craigie, R. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.* **1997**, *16*, 6849−6859.

(9) Heuer, T. S.; Brown, P. O. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* **1997**, *36*, 10655−10665.

(10) Eijkelenboom, A. P. A. M.; Puras Lutzhe, R. A.; Boelens, R.; Plasterk, R. H. A.; Kaptein, R.; Hård, K. The DNA-binding domain of HIV-1 integrase has and SH3-like fold. *Nat. Struct. Biol.* **1995**, *2*, 807−810.

(11) Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **1997**, *4*, 567−577.

(12) Eijkelenboom, A. P. A. M.; van den Ent, F. M. I.; Vos, A.; Doreleijers, J. F.; Hård, K.; Tullius, T. D.; Plasterk, R. H. A.; Kaptein, R.; Boelens, R. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: A three-helix bundle stabilized by zinc. *Curr. Biol.* **1997**, *7*, 739−746.

(13) Maignan, S.; Guilloteau, J.-P.; Zhou-liu, Q.; Clément-Mella, C.; Mikol, V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: High level of similarity of the active site with other viral integrases. *J. Mol. Biol.* **1998**, *282*, 359−368.

(14) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M., Craigie, R.; Davies, D. R. Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9150−9154.

(15) Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* **1994**, *266*, 1981−1986.

(16) Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J.; Choe, S.; McCammon, J. A. Molecular dynamics studies on the HIV-1 integrase catalytic domain. *Biophys. J.* **1999**, *76*, 2999−3011.
(17) Böhm, H.-J.; Klebe, G. What can we learn from molecular recognition in protein−ligand complexes for the design of new drugs? *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588−2614.
(18) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening−an overview. *DDT* **1998**, *3*, 160−178.
(19) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1994; pp 110−126.
(20) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040−13043.
(21) Bujacz, G.; Jaskólski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* **1995**, *253*, 333−346.
(22) *MidasPlus*; Computer Graphics Lab, University of California, San Diego.
(23) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The MIDAS display system. *J. Mol. Graphics* **1988**, *6*, 13−27.
(24) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A**.** Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A* **1999**, *103*, 10213−10219.
(25) Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078−1082.
(26) Kuntz, I. D.; Meng, C. E.; Shoichet, B. K. Structure-based molecular design. *Acc. Chem. Res.* **1994**, *27*, 117−123.
(27) Colman, P. M. Structure-based drug design. *Curr. Opin. Struct. Biol.* **1994**, *4*, 868−874.
(28) Blundell, T. L. Structure-based drug design. *Nature* **1996**, *S384*, 23−26.
(29) Zheng, Q.; Kyle, D. J. Computational screening of combinatorial libraries via multicopy sampling. *DDT* **1997**, *6*, 229−234.
(30) Marrone, T. J.; Briggs, J. M.; McCammon, J. A. Structure-based drug design: Computational advances. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71−90.
(31) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(32) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195−202.
(33) Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* **1991**, *11*, 29−34.
(34) Böhm, H.-J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61−78.
(35) Caflisch, A.; Miranker, A.; Karplus, M. Multiple copy simultaneous search and construction of ligands in binding sites: Application to inhibitors of HIV-1 Aspartic Proteinase. *J. Med. Chem.* **1993**, *36*, 2142−2167.
(36) Miranker, A.; Karplus, M. An automated method for dynamic ligand design. *Proteins* **1995**, *23*, 472−490.
(37) Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. PRO_LIGAND: An approach to de novo molecular design. 1. Application to the design of organic molecules. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 13−32.
(38) Pearlman, D. A.; Murko, M. A. CONCERTS: Dynamic connection of fragments as an approach to de novo ligand design. *J. Med. Chem.* **1996**, *39*, 1651−1663.
(39) Joseph-McCarthy, D.; Fedorov, A. A.; Almo, S. C. Comparison of experimental and computational functional group mapping of an RNase A structure: Implications for computer-aided drug design. *Protein Eng.* **1996**, *9*, 773−780.
(40) Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. *Proteins* **1997**, *29*, 32−58.
(41) Burt, S.; Hutchins, C.; Zielinski, P. J. A Monte Carlo method for finding important ligand fragments from receptor data. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 243−255.
(42) Castro, A.; Richards, W. G.; Lyne, P. D. The design of novel acetylcholinesterase inhibitors using the multiple copy simultaneous search method. *Med. Chem. Res.* **1999**, *9*, 98−107.
(43) Jorgensen, W. L. *BOSS Version 3.8*; Yale University: New Haven, CT, 1997.
(44) Probes in the first MUSIC simulations were used to identify catalytic sites around transition structures from *ab initio* molecular orbital calculations. Carlson H. A. Chapter 5: Design of catalytic hosts to promote the rearrangement of chorismate to prephenate. In Methodological Development and Bioorganic Applications of Computational Simulations. Thesis, Yale University, New Haven, CT, 1997.
(45) Kaminski, G.; Jorgensen, W. L. Performance of the AMBER94, MMFF94, and OPLS-AA force fields for modeling organic liquids. *J. Phys. Chem.* **1996**, *100*, 18010−18013.
(46) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.
(47) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. Free-energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. *J. Phys. Chem.* **1994**, *98*, 13077−13082.
(48) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(49) Zheng, Q.; Kyle, D. J. Multiple copy sampling: Rigid versus flexible protein. *Proteins* **1994**, *19*, 324−329.
(50) Totrov, M.; Abagyan, R. Flexible protein−ligand docking by global energy optimization in internal coordinates. *Proteins* **1997**, *Suppl. 1*, 215−220.
(51) Ringe, D. What makes a binding site a binding site? *Curr. Opin. Struct. Biol.* **1995**, *5*, 825−829.
(52) Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazumder, A.; Zhao, H.; Burke, T. R., Jr.; Pommier, Y.; Milne, G. W. A.: Discovery of HIV-1 integrase inhibitors by pharmacophore searching. *J. Med. Chem.* **1997**, *40*, 930−936.
(53) *Catalyst*; Molecular Simulations Inc.: San Diego, CA, 1996.
(54) Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W. A.; Pommier, Y. HIV-1 integrase pharmacophore: Discovery of inhibitors through three-dimensional database searching. *J. Med. Chem.* **1997**, *40*, 920−929.
(55) Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Nicklaus, M. C.; Milne, G. W. A.; Proska, B.; Pommier, Y. HIV-1 integrase pharmacophore: Discovery of inhibitors through three-dimensional database searching. *J. Med. Chem.* **1997**, *40*, 942−951.
(56) Neamati, N.; Hong, H.; Sunder, S.; Milne, G. W. A.; Pommier, Y. Potent inhibitors of human immunodeficiency virus type 1 integrase: Identification of a novel four-point pharmacophore and tetracyclines as novel inhibitors. *Mol. Pharmacol.* **1997**, *52*, 1041−1055.
(57) Hong, H.; Neamati, N.; Winslow, H. E.; Christensen, J. L.; Orr, A.; Pommier, Y.; Milne, G. W. A.: Identification of HIV-1 integrase inhibitors based on a four-point pharmacophore. *Antiviral Chem. Chemother.* **1998**, *9*, 461−472.
(58) Neamati, N.; Sunder, S.; Pommier, Y. Design and discovery of HIV-1 integrase inhibitors. *DDT* **1997**, *2*, 487−498.
(59) Farnet, C. M.; Wang, B.; Lipford, J. R.; Bushman, F. D. Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9742−9747.
(60) Farnet, C. M.; Wang, B.; Hansen, M.; Lipford, J. R.; Zalkow, L.; Robinson, W. E., Jr.; Siegel, J.; Bushman, F. D. Human immunodeficiency virus type 1 cDNA integration: New aromatic hydroxylated inhibitors and studies of the inhibition mechanism. *Antimicrob. Agents Chemother.* **1998**, *42*, 2245−2253.
(61) Lubkowski, J.; Yang, F.; Alexandratos, J.; Wlodawer, A.; Zhao, H.; Burke, T. R., Jr.; Neamati, N.; Pommier, Y.; Merkel, G.; Skalka, A. M. Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4831−4836.
(62) Artico, M.; Di Santo, R.; Costi, R.; Novellino, E.; Greco, G.; Massa, S.; Tramontano, E.; Marongiu, M. E.; De Montis, A.; La Colla, P. Geometrically and conformationally restrained cinnamoyl compounds as inhibitors of HIV-1 integrase: Synthesis, biological evaluation, and molecular modeling. *J. Med. Chem.* **1998**, *41*, 3948−3960.
(63) Zhu, K.; Cordeiro, M. L.; Atienza, J.; Robinson, W. E., Jr.; Chow, S. A. Irreversible inhibition of human immunodeficiency virus type 1 integrase by dicaffeoylquinic acids. *J. Virol.* **1999**, *73*, 3309−3316.
(64) King, P. J.; Robinson, W. E., Jr. Resistance to the anti-human immunodeficiency virus type 1 compound L-chicoric acid results from a single mutation at amino acid 140 of integrase. *J. Virol.* **1998**, *72*, 8420−8424.
(65) Mazumder, A.; Neamati, N.; Sunder, S.; Schultz, J.; Pertz, H.; Eich, E.; Pommier, Y. Curcumin analogs with altered potencies against HIV-1 integrase as probes for biochemical mechanisms of drug action. *J. Med. Chem.* **1997**, *40*, 3057−3063.
(66) Mekouar, K.; Mouscadet, J.-F.; Desmaële, D.; Subra, F.; Leh, H.; Savouré, D.; Auclair, C.; d'Angelo, J. Styrylquinoline derivatives: A new class of potent HIV-1 integrase inhibitors that block HIV-1 replication in CEM cells. *J. Med. Chem.* **1998**, *41*, 2846−2857.

(67) Lin, Z.; Neamati, N.; Zhao, H.; Kiryu, Y.; Turpin, J. A.; Aberham, C.; Strebel, K.; Kohn, K.; Witvrouw, M.; Pannecouque, C.; Debyser, Z.; De Clercq, E.; Rice, W. G.; Pommier, Y.; Burke, T. R., Jr. Chicoric acid analogues as HIV-1 integrase inhibitors. *J. Med. Chem.* **1999**, *42*, 1401−1414.
(68) Asante-Appiah, E.; Skalka, A. M. A metal-induced conformational change and activation of HIV-1 integrase. *J. Biol. Chem.* **1997**, *272*, 16196−16205.
(69) Asante-Appiah, E.; Seeholzer, S. H.; Skalka, A. M. Structural determinants of metal-induced conformational changes in HIV-1 integrase. *J. Biol. Chem.* **1998**, *273*, 35078−35087.
(70) Mazumder, A.; Gazit, A.; Levitzhi, A.; Nicklaus, M.; Yung, J.; Kohlhagen, G.; Pommier, Y. Effects of tyrophostins, protein kinase inhibitors, on human immunodeficiency virus type 1 integrase. *Biochemistry* **1995**, *34*, 15111−15122.
(71) Mazumder, A.; Raghavan, K.; Weinstein, J.; Kohn, K. W.; Pommier, Y. Inhibition of human immunodeficiency virus type-1 integrase by curcumin. *Biochem. Pharmacol.* **1995**, *49*, 1165−1170.
(72) Sánchez-Cortés, S.; Miskovsky, P.; Jancura, D.; Bertoluzza, A. Specific interactions of antiretroviraly active drug hypericin with DNA as studied by surface-enhanced resonance Raman spectroscopy. *J. Phys. Chem.* **1996**, *100*, 1938−1944.
(73) Mazumder, A.; Wang, S.; Neamati, N.; Nicklaus, M.; Sunder, S.; Chen, J.; Milne, G. W. A.; Rice, W. G.; Burke, T. R., Jr.; Pommier, Y. Antiretroviral agents as inhibitors of both human immunodeficiency virus type-1 integrase and protease. *J. Med. Chem.* **1996**, *39*, 2472−2481.
(74) Zhao, H.; Neamati, N.; Sunder, S.; Hong, H.; Wang, S.; Milne, G. W. A.; Pommier, Y.; Burke, T. R., Jr. Hydrazide-containing inhibitors of HIV-1. *J. Med. Chem.* **1997**, *40*, 937−941.
(75) Craigie, R.; Mizuuchi, K.; Bushman, F. D.; Engelman, A. A rapid in vitro assay for HIV DNA integration. *Nucleic Acids Res.* **1991**, *19*, 2729−2734.
(76) When using an assay (ref *59*) employing purified integrase and Mn^{2+}, IC_{50}'s for compound **3** are much lower: 57 and 37 μM for 3′-preprocessing and strand transfer, respectively.
(77) *InsightII*; Molecular Simulations Inc.: San Diego, CA, 1997.
(78) *HBUILD*; Polar hydrogen parameter set for CHARMm Version 22; Molecular Simulations Inc.: Waltham, MA, 1992.
(79) Anchell, J.; Apra, E.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dupuis, M.; Dyall, K. H.; Fruchtl, G. F.; Gutowski, M.; Harrison, R.; Hess, A.; Jaffe, J.; Kendall, R.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nichols, J.; Nieplocha, J.; Rendall, A.; Stave, M.; Straatsma, T. P.; Taylor, H.; Thomas, G.; Wolinski, K.; Wong, A. *NWChem, A computational chemistry package for parallel computers, Version 3.2*; High Performance Computational Chemistry Group, Pacific Northwest Laboratory: Richland, WA, 1998.
(80) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(81) Berendsen, H. J. C.; Grigerra, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(82) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(83) Briggs, J. M. Unpublished data.
(84) Jorgensen, W. L.; Impey, R. W.; Chandrasekhar, J.; Madura, J. D.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

JM990322H