Discovery of HIV-1 Integrase Inhibitors by Pharmacophore Searching†

Huixiao Hong,*,‡ Nouri Neamati,§ Shaomeng Wang,‡ Marc C. Nicklaus,‡ Abhijit Mazumder,§ He Zhao,‡ Terrence R. Burke, Jr.,‡ Yves Pommier,§ and George W. A. Milne‡

Laboratories of Medicinal Chemistry and Molecular Pharmacology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Building 37, Bethesda, Maryland 20892-4255

Received October 31, 1996⊗

Based upon a class of known HIV-1 integrase inhibitors, several pharmacophore models were proposed from molecular modeling studies and validated using a 3D database of 152 compounds for which integrase assay data are known. Using the most probable pharmacophore model as the query, the NCI 3D database of 206 876 compounds was searched, and 340 compounds that contain the pharmacophore query were identified. Twenty-nine of these compounds were selected and tested in the HIV-1 integrase assay. This led to the discovery of 10 novel, structurally diverse HIV-1 integrase inhibitors, four of which have an IC_{50} value less than 30 μM and are promising lead compounds for further HIV-1 integrase inhibitor development.

Introduction

Human immunodeficiency virus type 1 (HIV-1) is the causative agent of acquired immunedeficiency syndrome (AIDS).[1,2] Several enzymes important to the replication cycle of this virus, for example, reverse transcriptase, protease, and integrase, are considered to be promising targets for the development of anti-AIDS drugs. In recent years, reverse transcriptase and protease[3,4] have been the primary targets in the development of anti-AIDS drugs. Effective inhibitors have been developed against these two enzymes, and several of them are now marketed in the United States as anti-AIDS drugs.

HIV-1 integrase is an enzyme that mediates the integration of HIV-1 DNA into a host chromosome[5−9] and is essential to replication of the virus.[10,11] The HIV-1 integrase first cleaves the last two nucleotides from each 3′-end of the viral DNA, leaving the terminal dinucleotide sequence CA-OH-3′ at these recessed 3′-ends. This activity is referred to as 3′-processing or dinucleotide cleavage. Transported to the nucleus as a nucleoprotein complex, integrase catalyzes a DNA strand transfer reaction involving the nucleophilic attachment of these ends on a host chromosome. This process is referred to as strand transfer. Finally, the viral 5′-ends are processed, and the gaps between the viral 5′- and target 3′-ends are repaired.[7,9] There is obviously a requirement for a functional integrase in HIV-1 replication. This enzyme is therefore thought to be a suitable target for chemotherapeutic intervention and has become a focus of anti-AIDS drug design efforts in these laboratories.

Over the last few years, several groups have made efforts to develop HIV-1 integrase inhibitors. So far, a number of classes of HIV-1 integrase inhibitors have been identified, including the flavone quercetin and caffeic acid phenyl ethyl ester (CAPE),[12,13] aurintricarboxylic acid and its relatives,[14] the polyanionic drug suramin,[15] curcumin,[16] derivatives of cosalane,[17] copper phenanthroline,[18] tyrphostins,[19] AZT nucleotides,[20] *bis*-catechols,[21] 2*N*-methyl-9-hydroxyellipticinium (NMHE) and oxazolopyridocarbazole (OPC),[22] methyl phosphonodiester and its derivatives,[23] hydroxylated aromatic compounds,[24] dibenzylbutyrolactone type lignanolide (−)-arctigenin,[25] (−)-arctigenin derivatives,[26] 3,5-dicaffeoylquinic acid, 1-(methoxyoxalyl)-3,5-dicaffeoylquinic acid, and L-chicoric acid.[27] None of these HIV-1 integrase inhibitors have entered clinical trials because of either their limited potency or high toxicity. Intensive research efforts are now being made in our laboratories to improve the potency and/or reduce the toxicity of the compounds in these classes.[24,26,28,29] There is also a need to discover lead compounds for new classes of HIV-1 integrase inhibitors, not least because of the high mutation capabilities of the virus, which can lead to a fast buildup of resistance against many individual drugs. The discovery of novel, potent, and structurally simple lead compounds is therefore a highly desirable goal in HIV-1 integrase inhibitor design.

A pharmacophore in a molecule refers to the three-dimensional (3D) arrangement of atoms or functional groups that is necessary for the compound to bind to a specific enzyme or receptor. 3D database pharmacophore searching attempts to identify molecules in a database on the basis of their possessing a particular pharmacophore in their structure, expressed necessarily in three dimensions. This approach has recently gained attention for its ability to discover new leads in drug development programs.[30−35] We have built a searchable 3D database[36] of a total of ca. 408 000 structures from the 2D molecular structures stored in the National Cancer Institute (NCI) Drug Information System (DIS) database,[37] employing the programs Chem-X[38] and Corina,[39] and have subsequently used this 3D database to discover, for example, a number of novel protein kinase C agonists[40] and HIV-1 protease inhibitors.[41]

Recently, we have reported on a pharmacophore that we identified in a significant number of the HIV-1 integrase inhibitors known to date and which we subsequently used to find novel inhibitors for this enzyme in the NCI 3D database.[42] In addition to showing the importance of this (first) pharmacophore, that study, however, presented strong evidence for a second pharmaphore whose presence in a compound might confer anti-integrase activity, presumably due to a different binding mode in the active site of the protein.

† Part 2 in a series of papers describing pharmacophore-based development of HIV-1 integrase inhibitors.

* To whom all correspondence should be addressed.
‡ Laboratory of Medicinal Chemistry.
§ Laboratory of Molecular Pharmacology.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1997.

S0022-2623(96)00754-6

In the present paper, we analyze this second pharmacophore and report on the identification of novel HIV-1 integrase inhibitors through 3D pharmacophore searches in the DIS database.

Experimental Section

3D HIV-1 Integrase Inhibitors Database. Over the last few years, a number of classes of HIV-1 integrase inhibitors have been identified (see above). We have collected in a 3D database a total of 152 compounds[43] for which HIV-1 integrase inhibition data are known. Out of these 152 compounds, 77 had shown inhibitory activity with an IC_{50} value equal to or less than 100 μM in the 3′-processing assay. This database was built using Chem-X's ChemDBS-3D database build and search module. After the 2D structures of all compounds had been drawn using the Edit/2D-Draw module of Chem-X, they were submitted to the automatic 3D database generation via ChemDBS-3D. We used the program options that ensure that conformational flexibility of the compounds is taken into account and employed keying of this database with 3D distance keys during the build phase (using the 'Search/Generate Keys' option), which allows for rapid screening in subsequent searches with the putative pharmacophores.

NCI 3D Database and Search Software. The details of the NCI 3D database, and of the Chem-X program that was used in both the 3D database build and search processes, have been described elsewhere.[36] The current version of the NCI 3D database consists of 206 876 "open" and 201 036 "discreet" (proprietary) structures,[44] for a total of 407 912 structures. All searches reported in this paper were conducted on only the "open" part of the database. The conformationally flexible search algorithm ('Flexifit') implemented in Chem-X (July 1994 version, running on a Silicon Graphics (SGI) workstation) was employed. For flexible compounds, multiple conformations are generated and analyzed during both building and searching of the database.[36,37]

Molecular Modeling. All molecular modeling studies—other than the building and searching of the databases—were performed with the QUANTA[45] 4.0/CHARMm 2.2 molecular modeling package,[46] running on an SGI IRIS Indigo workstation. 2D structures of compounds were built in QUANTA's ChemNote module and energy-minimized using CHARMm. Energy minimization was typically conducted with 5000 iterations or until convergence was achieved (defined as an energy gradient of 0.001 kcal mol^{-1} $Å^{-1}$ or less), using an adjusted basis Newton−Raphson (ABNR) algorithm as implemented in CHARMm. Conformational searches were performed using either the systematic or the Monte Carlo random search algorithms implemented in QUANTA.

Preparation of Oligonucleotide Substrates. The HPLC-purified oligonucleotides AE117, 5′-ACTGCTAGAGATTTTC-CACAC-3′, and AE118, 5′-GTGTGGAAAATCTCTAGCAGT-3′, were purchased from Midland Certified Reagent Co. (Midland, TX). Dr. R. Craigie, Laboratory of Molecular Biology, NIDDK, NIH, Bethesda, MD, provided the expression system for the wild-type HIV-1 integrase. To analyze the extents of 3′-processing and strand transfer using 5′-end-labeled substrates, AE118 was 5′-end-labeled using T_4 polynucleotide kinase (Gibco BRL) and [γ-^{32}P]-ATP (DuPont-NEN). The kinase was heat-inactivated, and AE117 was added to the same final concentration. The mixture was heated at 95 °C, allowed to cool slowly to room temperature, and run on a G-25 Sephadex quick spin column (Boehringer Mannheim) to separate annealed double-stranded oligonucleotide from unincorporated label.

***In Vitro* HIV-1 Integrase Assays.** Integrase was preincubated at a final concentration of 200 nM with inhibitor in reaction buffer (50 mM NaCl, 1 mM HEPES, pH 7.5, 50 μM EDTA, 50 μM dithiothreitol, 10% glycerol (w/v), 7.5 mM $MnCl_2$ (when specified), 0.1 mg/mL bovine serum albumin, 10 mM 2-mercaptoethanol, 10% dimethyl sulfoxide, 25 mM MOPS, pH 7.2) at 30 °C for 30 min. Then, 20 nM of the 5′-end ^{32}P-labeled linear oligonucleotide substrate was added, and incubation was continued for an additional 60 min. Reactions were quenched by the addition of an equal volume (16 μL) of loading dye (98% deionized formamide, 10 mM EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue). An aliquot (5 μL) was electrophoresed on a denaturing 20 polyacrylamide gel (0.09 M Tris-borate, pH 8.3, 2 mM EDTA, 20% acrylamide, 8 M urea). Gels were dried, exposed in a molecular Dynamics Phosphorimager cassette, and analyzed using a Molecular Dynamics Phosphorimager (Sunnyvale, CA). Percent inhibition was calculated using the following equation:

$$\%I = 100 \times [1 - (D - C)/(N - C)]$$

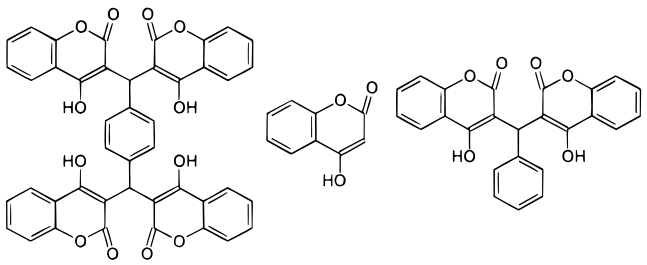
where C, N, and D are the fractions of 21-mer substrate converted to 19-mer (3′-processing product) or strand transfer products for DNA alone, for DNA plus integrase, and for integrase plus drug, respectively. All IC_{50} values were determined by plotting the drug concentration vs percent inhibition and determining the concentration which produced 50% inhibition.

Results and Discussion

Identification of the HIV-1 Integrase Pharmacophore. To perform a pharmacophore search it is first necessary to establish a valid pharmacophore model. In contrast to HIV-1 protease or HIV-1 reverse transcriptase, where crystal structures of enzyme−ligand complexes are available, it is not possible to derive an HIV-1 integrase pharmacophore model directly from such complexes since no experimental structures of the HIV-1 integrase complexed with its substrate or with inhibitors are known to date. The crystal structure of the catalytic core domain of HIV-1 integrase (without any ligand) has been determined;[47] however, residues 141−153, which contain the crucial residue Glu 152 in the DD(35)E motif shown to be essential for activity through site-directed mutagenesis studies, were unresolved due to disorder caused presumably by the flexibility of this loop. It is therefore very difficult to derive a reasonable pharmacophore model directly from the protein X-ray structure. Because of this, one has to resort to indirect approaches for deriving a pharmacophore, which, in this study, was molecular modeling of known potent HIV-1 integrase inhibitors.

Previously, we reported[29] that 4-hydroxycoumarin-based compounds showed promising activity against HIV-1 integrase. The best inhibitor in this class, a "tetrameric" 4-hydroxycoumarin structure (**1**, NSC 158393), exhibited IC_{50} values of 1.5 and 0.8 μM for 3′-processing and strand transfer, respectively. Subsequent chemical modifications[28] and molecular modeling studies[29] have shown that, although the "monomer" form (**2**) of 4- hydroxycoumarin does not exhibit significant inhibitory activity against HIV-1 integrase, the "dimer" form (**3**) of 4-hydroxycoumarin does possess inhibitory potency (43.4 ± 23.7 and 38.8 ± 25.9 μM, respectively).[28] Therefore, the crucial polar/hydrogen bond interaction sites appear to be contained in the dimer structure **3** but not in the monomer structure **2**. Further molecular modeling of this group of compounds indicated that the two carbonyl groups and two hydroxy groups in the dimer structure **3** can be mapped very well onto the corresponding atoms in the tetramer structure **1**.

We have previously shown that the biologically active conformation of an inhibitor need not be the global energy minimum state (in vacuum).[48] Nevertheless, the conformational energy penalty a compound may have to pay to be able to bind to its target cannot exceed the overall free energy gained in the binding process, and it is thus the low-energy areas of the conformational energy hypersurface that need to be explored to identify



the possible bioactive conformation(s). This was done, for NSC 158393, through a Monte Carlo conformational analysis, in which each of 5000 randomly generated conformations of the compound was minimized using QUANTA/CHARMm. Some 621 distinct conformations were found with a conformational energy of less than 10 kcal/mol above the global minimum, and it was assumed that a reasonable pharmacophore model would be found in this subset.

On the basis of the average values of the two crucial torsion angles, ω_1 and ω_2, shown in Figure 1, the 621 low-energy conformations of NSC 158393 could be partitioned into five distinct clusters. From these, we derived five four-point pharmacophores with the interatomic distances shown in Figure 2. To check if any of these is a reliable pharmacophore model, validation by a search in the small 3D database of 152 HIV-1 integrase inhibitors was performed. Even with relatively wide distance tolerances of 0.7 Å, the hit rates for each of the five four-point pharmacophores, however, were disappointingly low, ranging from 0% to, at best, 17%, and so attempts to define a four-point pharmacophore model were dropped.[49]

Concluding that a three-point pharmacophore, as it is commonly used in enzyme−ligand systems, may be a better choice in the HIV-1 integrase case, too, all possible three-point pharmacophores were extracted from the five conformational clusters of NSC 158393. There are six oxygen−oxygen distances for the four oxygen atoms shown in Figure 2. For each of the conformations, four three-point pharmacophores can be built from these six distances, resulting in 20 possible pharmacophores for this compound. Two of these three-point pharmacophores (5.14, 4.76, 2.72 and 5.14, 4.76, 2.76) are essentially identical and were set as 5.14, 4.76, 2.74. Similarly, the three pharmacophores (6.00, 4.76, 2.72), (6.00, 4.76, 2.76), and (6.00, 4.76, 2.74) were presumed to be identical and replaced by (6.00, 4.76, 2.74). This resulted in the 17 unique possible three-point pharmacophores shown in Table 1, which were then used in a search of the database of 152 integrase inhibitors to give results which are also shown in Table 1. The distances listed in this table show that the 17 possible pharmacophores can be partitioned into four clusters, (1−5), (6−9), (10−13), and (14−17), based upon the pharmacophore dimensions. As can be seen from the search results in Table 1, the first of these clusters was the most successful in retrieving known active compounds, and this was thus presumed to include the most probable of the analyzed HIV-1 integrase pharmacophores.

3D Database Search Results. In order to construct, from the validation database search results shown in Table 1, a reasonable pharmacophore query for a search of the entire (open) NCI 3D database, the multiple sets of distances in each cluster were reduced

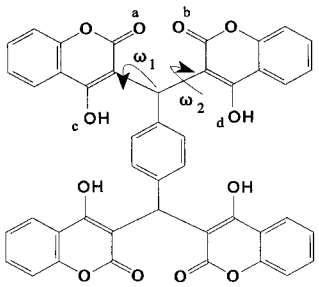


Figure 1. Molecular structure of NSC 158393. Shown are the torsion angles ω_1 and ω_2 as used in the conformational search and analysis. Points a−d are the atoms between which the distances are measured that define the analyzed pharmacophores.

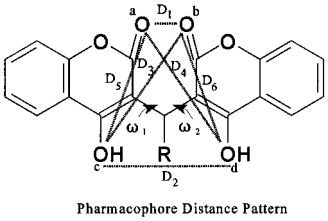


Figure 2. Possible pharmacophores in the hydroxycoumarin 'dimer' structure.

to a single-distance triad in the following way. A weighting factor W_i was defined for every three-point pharmacophore in a cluster depending on the percentage of active compounds containing the pharmacophore in all active compounds (N_+), as given in eq 1:

$$W_i = (N_+)_i / \sum (N_+)_i \; (i = 1, 2, \ldots n) \qquad (1)$$

with n being the size of each cluster (e.g., 5 for cluster 1). W_i represents the contribution of the distance pattern of the ith pharmacophore in the cluster to the combined pharmacophore query distance pattern, which was constructed according eqs 2−4:

$$\mathrm{DQ}_1 = \sum W_i D_{1_i} \; (i = 1, 2, \ldots n) \qquad (2)$$

$$\mathrm{DQ}_2 = \sum W_i D_{2_i} \; (i = 1, 2, \ldots n) \qquad (3)$$

$$\mathrm{DQ}_3 = \sum W_i D_{3_i} \; (i = 1, 2, \ldots n) \qquad (4)$$

The resulting values DQ_1, DQ_2, and DQ_3 then are the dimensions of the pharmacophore query subsequently used in the NCI 3D database search. The pharmacophore query constructed from pharmacophores 1−5 (i.e., from the first cluster) is shown in Figure 3. The distance tolerances for D_1, D_2, and D_3 were set to 0.7, 0.7, and 0.4 Å, respectively. The reason for assigning different distance tolerances in the pharmacophore query was that the distance between two oxygen atoms in the same ring system (D_3) should change very little as the conformation varies, but the distances between oxygen atoms in different ring systems (D_1, D_2) can change much more substantially when the conformation changes.

The search of the structure database of 206 876 "open" compounds, using this pharmacophore query, yielded a total of 340 compounds which contained the pharmacophore pattern in one or more of their conformations. Samples of 180 of these were available in the NCI repository. Charged molecules which would have likely transport problems *in vivo* were excluded from this set,

Table 1. Three-Point Pharmacophore Validation Results[a]

pharmacophore	interatomic distances (Å)			N_{act}	N_+ (%)	*W*
1	2.74	4.76	6.00	52	67.53	0.208
2	2.69	4.76	5.84	51	66.23	0.204
3	2.78	4.76	5.40	51	66.23	0.204
4	3.39	4.76	5.68	51	66.23	0.204
5	2.74	4.76	5.14	45	58.44	0.180
6	4.19	4.76	5.84	47	61.04	
7	4.01	4.76	6.00	45	58.44	
8	4.07	4.76	5.68	44	57.14	
9	4.35	4.76	5.40	42	54.55	
10	2.74	4.76	4.27	42	54.55	
11	2.69	4.76	4.62	48	62.34	
12	2.78	4.76	4.45	48	62.34	
13	3.39	4.76	4.15	40	51.95	
14	4.01	4.76	4.27	33	42.86	
15	4.19	4.76	4.26	9	11.69	
16	4.07	4.76	4.15	34	44.16	
17	4.35	4.76	4.45	13	18.18	

[a] Search tolerances are ±0.7 Å for the long sides of the triangle, ±0.4 Å for the short side. N_{act}: number of active compounds (IC_{50} < 100 mM) containing the pharmacophore. N_+: percentage of active compounds retrieved with pharmacophore (IC_{50} < 100 mM) containing the pharmacophore in 77 active HIV-1 integrase inhibitors in the database. *W*: value of weighting function (see text).

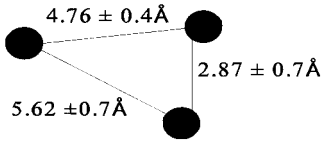


Figure 3. Dimensions of the pharmacophore query used in the search of the 206 876 compounds of the 'open' part of the NCI 3D database. The pharmacophoric atom centers are O and N.

and this, together with exclusions based upon considerations of chemical diversity, led to a set of 29 compounds being submitted to the integrase bioassay.

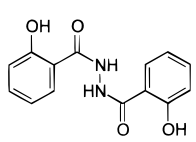
***In Vitro* HIV-1 Integrase Bioassay Evaluations.** The anti-HIV-1 integrase activity of the 29 compounds, shown in Table 2, was measured in the *in vitro* assay as described in the Experimental Section, employing purified recombinant integrase and a 21-mer duplex oligonucleotide corresponding to the U5 region of the HIV LTR sequence. Both 3′-processing and DNA strand transfer reactions are shown in one experiment as described earlier.

All compounds were initially assayed at a concentration of 100 μg/mL; all 29 compounds showed some inhibitory activity. Those that exhibited >50% inhibition at this concentration (for both 3′-processing and strand transfer) were tested at several concentrations to obtain the IC_{50} values shown in Table 2. IC_{50} values of less than 100 μg/mL (see Chart 1) were found for all 10 of these compounds, for which the corresponding micromolar values are listed in Table 2. Their structure, with a highlight on the atoms found by Chem-X to match the pharmacophore query, are depicted in Chart 1. Four of these 10 inhibitors have IC_{50} values of less than 10 μg/mL (equivalent IC_{50} < 30 μM); for NSC 310217, IC_{50}(3′-processing) = 0.6 μg/mL and IC_{50}-(integration) = 0.46 μg/mL, which corresponds to 2.0 and 1.5 μM, respectively, if the molecular weight listed in the NCI database records (302) is used. Upon further investigation, which is reported in detail in the following paper,[50] NSC 310217 proved to be a mixture of two compounds: 1,2-bis(*o*-hydroxylbenzoyl)hydrazine (**4**) and the hydrazide **5**.

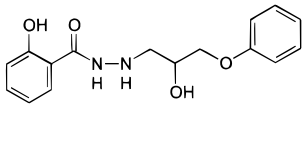
Table 2. Results of HIV-1 Integrase Inhibitory Assay[a]

	% inhibition at 100 μg/mL		IC_{50} (μM)	
NSC no.	3′-processing	strand transfer	3′-processing	strand transfer
4411	36.5 (41.2)	27.6 (48.9)		
16683	14.5 (35.8)	50.3 (67.4)		
23518	20.8	17.7		
26665	40.8 (29.6)	67.7 (34.2)		
30181	25.8 (10.2)	61.5 (55.2)		
48240	100	100	26 ± 2.7	20.6 ± 4.8
49858	99.8	77.4	294 (367)	185 (233)
57586	20.8 (43.2)	54.4 (69.7)		
58327	0.4 (2.2)	0.0 (13.4)		
61916	25.0 (43.9)	52.4 (69.9)		
72097	10.9 (2.6)	17.1 (18.9)		
83080	13.1 (7.1)	25.1 (27.0)		
87566	28.9 (33.3)	25.0 (33.3)		
87863	37.0 (45.8)	77.9 (70.2)		
106723	34.7 (56.7)	70.5 (74.1)		
226058	22.6 (24.9)	63.5 (54.3)		
229488	72.7	69.3	228 (271)	203 (271)
246119	57.9	65.6	350 (289)	268 (135)
305801	43.7 (30.6)	87.4 (56.6)		
310217	100	100	2.0 ± 1.1	1.5 ± 0.8
371056	98.1	92.4	29.9 (71.3)	16.5 (35.8)
408130	25.7 (49.1)	63.3 (66.6)		
612955	22.5 (19.1)	48.1 (43.2)		
626436	63.4	85.7	177 (124)	103 (47.6)
632901	72.5	79.9	182 (238)	101 (142)
633671	45.3 (30.3)	90.6 (47.9)		
635971	23.1 (17.6)	57.2 (37.0)		
641547	62.7 (43.4)	92.2 (66.8)	224 (291)	134 (143)
642710	100	100	5.3 ± 2.5	5.0 ± 1.2

[a] Values with a standard deviation are from at least three independent experiments, and those in parentheses are from a second experiment.



4



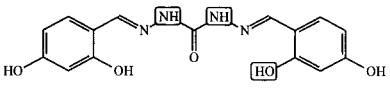
5

Both these structures, as well as the originally provided structure[51] (a closely related hydrazide, see Chart 1), contain the original pharmacophore. Only the bis-hydrazide **4** showed activity (ca. 2 μM) and is thus responsible for the inhibitory activity shown by the assayed sample of NSC 310217. An activity only slightly lower was found for NSC 642710, which inhibited HIV-1 integrase in the 5 μM range. Compounds NSC 48240 and NSC 371056 showed inhibitory activity in the 30 μM range and are notable because they contain a resorcin, i.e., a *m*-dihydroxylphenyl, moiety, which contrasts them with many of the known HIV-1 integrase inhibitors which possess *ortho* pairs of phenolic hydroxyl groups. These four molecules, being nonionic and of low molecular weight and simple structure, are promising lead compounds in the search for integrase inhibitors.

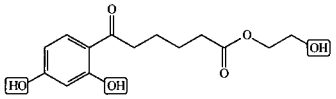
Conformational Energy Penalty. It can be seen from the data in Table 2 that all 29 compounds tested in the HIV-1 integrase assay exhibited some inhibitory activity. However, only 10 compounds had shown inhibition of more than 50% in both integration steps at a concentration of 100 μg/mL and had subsequently shown IC_{50} values of less than 100 μg/mL (only micromolar IC_{50} data shown in Table 2); the remaining 19 compounds, albeit not totally devoid of inhibitory potency, are called "inactives" in the following for the purpose of discussion.

Chart 1. Structures of 10 HIV-1 Integrase Inhibitors Discovered in the 3D Search[a]

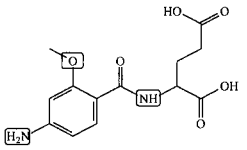
NSC 48240
IC_{50}=8.7±0.9 μg/ml (3'-processing)
IC_{50}=6.8±1.6 μg/ml (strand transfer)



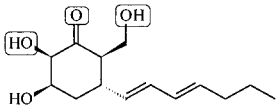
NSC 49858
IC_{50}= 83(104) μg/ml (3'-processing)
IC_{50}= 52(66) μg/ml (strand transfer)



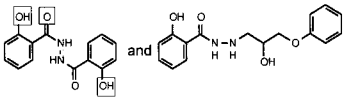
NSC 229488
IC_{50} = 68(80) μg/ml (3'-processing)
IC_{50} = 60(80) μg/ml (strand transfer)



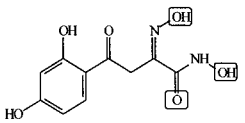
NSC 246119
IC_{50} = 89(74) μg/ml (3'-processing)
IC_{50} = 68(34) μg/ml (strand transfer)



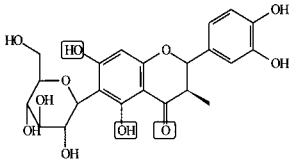
NSC 310217
IC_{50} = 0.6±0.33 μg/ml (3'-processing)
IC_{50} = 0.46±0.25 μg/ml (strand transf



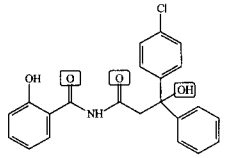
NSC 371056
IC_{50} = 7.6(18) μg/ml (3'-processing)
IC_{50} = 4.2(9.1) μg/ml (strand transfer)



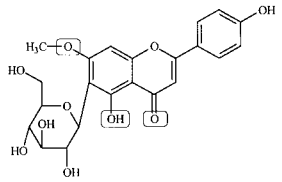
NSC 626436
IC_{50} = 82(58) μg/ml (3'-processing)
IC_{50} = 48(22) μg/ml (strand transfer)



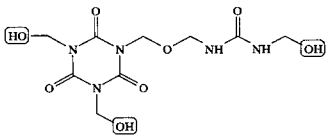
NSC 632901
IC_{50} = 72(94) μg/ml (3'-processing)
IC_{50} = 40(56) μg/ml (strand transfer)



NSC 641547
IC_{50} = 100(130) μg/ml (3'-processing)
IC_{50} = 60(64) μg/ml (strand transfer)



NSC 642710
IC_{50} = 1.7±0.8 μg/ml (3'-processing)
IC_{50} = 1.6±0.4 μg/ml (strand transfer)



This finding suggested that the presence of the pharmacophore that was used in the query, depicted in Figure 3, may be a necessary but not sufficient condition for strong inhibitory activity of a compound in this set and that there are presumably additional factors which are important in the binding of these ligands to HIV-1 integrase. One important such factor is the conformational energy penalty exacted for an inhibitor to achieve the desired conformation containing the pharmacophore query 3D pattern of atoms. The larger this energy penalty is, the lower the free energy of binding will be (all other factors being equal), and the less effectively it will therefore bind to its target.

In the 3D database pharmacophore search in Chem-X, energy was not taken explicitly into account for reasons of computational efficiency. Hence, energy data such as those provided by CHARMm as described in the previous section were not available to narrow down the selection of structures based on conformational energy. Consequently, some pharmacophore-containing compounds found in the search may have failed to bind to the enzyme because it was only in a high-energy conformation that they would have been able to present the pharmacophore to the protein.

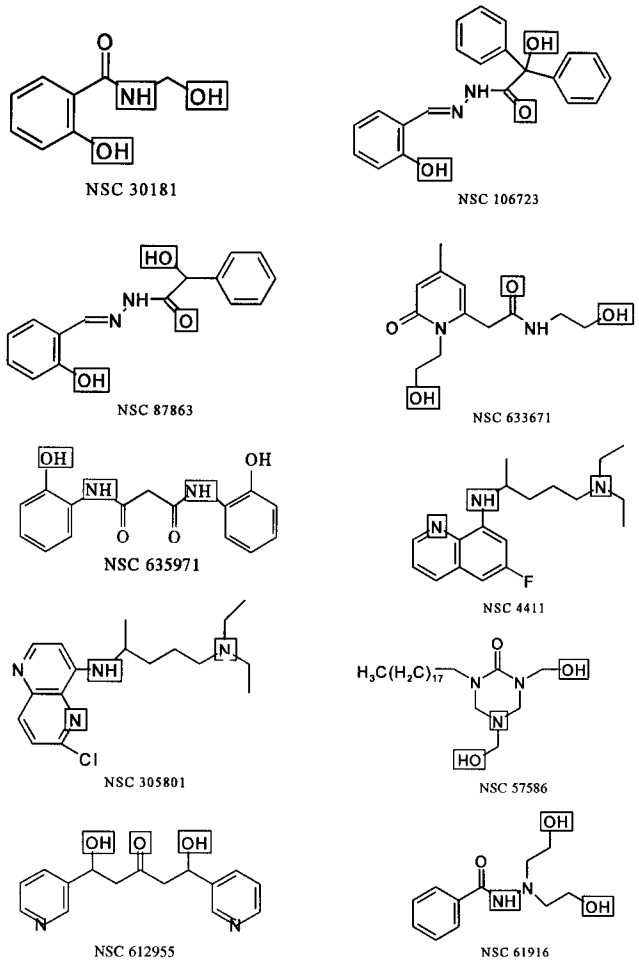
In a systematic investigation of this possibility, conformational searching and pharmacophore mapping were carried out for all 10 active compounds and, as a control, for 10 of the inactive compounds. For each compound, 4000 conformations were generated using random sampling, and each conformation was subsequently energy-minimized. These conformations were then mapped onto the pharmacophore query constructed in Figure 3, and the results are shown in Table 3. It can be seen that all but one of the 10 active compounds contain the pharmacophore in their global energy minimum conformation. All 10 inactive compounds, on the other hand, can adopt conformations containing the pharmacophore only at higher energies (average ΔE = 5.14 kcal/mol), and thus, to bind to the enzyme, they

Table 3. Lowest Energy Conformation That Contains the Pharmacophore[a]

NSC no.	rms (Å)[b]	ΔE (kcal/mol)[c]
48240	0.215	0
49858	0.062	0
229488	0.137	0
246119	0.212	0
310217	0.236[d]	0.657[d]
371056	0.107	0
626436	0.134	0
632901	0.157	0
641547	0.138	0
642710	0.101	0
4411	0.080	7.560
30181	0.214	2.920
57586	0.249	2.163
61916	0.299	8.424
87863	0.168	2.354
106723	0.196	2.971
305801	0.197	13.511
612955	0.171	8.380
633671	0.074	1.908
635971	0.276	1.215

[a] First 10 rows are the 'active' compounds; last 10 rows are 'inactive' compounds (see text). [b] rms: root-mean-square value of rigid-body fitting of three oxygens or nitrogens in a conformation of the compound to the pharmacophore query. [c] ΔE: difference between the energy of the conformation containing the pharmacophore and the global energy minimum of the compound. [d] These values were calculated for the original structure in the DIS database. Subsequent analysis showed that the global energy minimum conformation of the active molecule **4** contains a triad of oxygens that matches two of the three pharmacophore distances very exactly and the distance range c (Figure 3) within 0.46 Å. Distorting this conformation somewhat allows to bring the third distance within the range used in the search. The energy penalty for doing so was 1.75 kcal/mol, and the rms between that conformation and the pharmacophore was 0.214 Å.

Chart 2. Ten 'Inactive' Compounds Containing the HIV-1 Integrase Pharmacophore Query (See text)[a]



face a significant energy penalty. The structure and pharmacophore centers identified by Chem-X for the 10 inactive compounds are shown in Chart 2. The two compounds with the lowest percentage of inhibition, NSC 4411 and NSC 612955, also show particularly high energy penalties (ΔE > 7.5 kcal/mol), but beyond this, the small sample size and the accuracy of the assay data preclude further statistical analysis. However, it seems plausible that the energy penalty of at least 1 kcal/mol, and typically more than 2 kcal/mol, for each of these compounds is at least partly responsible for their low inhibitory activity. The effect of such energy penalties on binding constants has been discussed previously;[52] briefly, every kcal/mol in energy penalty corresponds to a reduction of approximately 175 in the binding constant, a rule of thumb which is consistent with the data in Tables 2 and 3.

Search Efficiency. Chem-X employs a rule-based algorithm in both the database build and search stages. Each single bond is rotated with a default step size of 120°, and each conjugated bond is rotated with a default step size of 180°. Chem-X thus covers a much wider conformational space than other 3D database build programs in which only a single conformer is stored. But with such a large step size, the program provides only a limited coverage of the entire conformational space for each compound. Chem-X offers another conformation generation algorithm, the 'Flexifit' method, which is particularly appropriate for structures with a large number of degrees of freedom. In this method each structure is modified by rotation of nonrigid bonds to try to fit it to the pharmacophore query, without taking into account internal energy changes. The process can be repeated from randomly chosen, different starting geometries. Before the 3D search was carried out on the NCI database, these two methods were compared. It was found that the Flexifit method was better suited to the search: more hits were obtained, reducing the possibility of missing active compounds.

Conclusion

Ten inhibitors of HIV-1 integrase have been discovered by a search of the open part of the NCI 3D database (206 876 compounds), using a pharmacophore query developed through a molecular modeling study of a class of known HIV-1 integrase inhibitors. Four of these 10 inhibitors, which are particularly potent with IC_{50} < 30 μM, constitute useful lead compounds for the development of anti-AIDS drugs and are under active investigation.

Despite a number of deficiencies in the current 3D database pharmacophore search technology, the successful discovery of a number of novel, structurally diverse HIV-1 integrase inhibitors shows that computer 3D database pharmacophore searching is an effective and promising tool in the drug development process.

References

(1) Gallo, R. C.; Sarin, P. S.; Gelmann, E. P.; Robert-Guroff, M.; Richardson, E.; Kalyanaraman, V. S.; Mann, D.; Sidhu, G. D.; Stahl, R. E.; Zolla-Prazner, S.; Leibowitch, J.; Popovic, M. Isolation of a T-cell leukemia virus in acqired immune deficiency syndrome (AIDS). *Science* **1983**, *220*, 865−867.
(2) Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220*, 868−871.
(3) Johnston, M. I.; Hoth, D. F. Present status and future prospects for HIV therapies. *Science* **1993**, *260*, 1286−1293.
(4) Yarchoan, R.; Pluda, J. M.; Perno, C.-F.; Mitsuya, H.; Broder, S. Antiretroviral therapy of human immunodeficiency virus infection: current strategies and challenges for the future. *Blood* **1991**, *78*, 859−884.
(5) Varmus, H.; Brown, P. Retroviruses. In *Mobile DNA*; Berg, D., Howe, M., Eds.; American Society of Microbiology: Washington, DC, 1989; pp 53−108.
(6) Katz, R. A.; Skalka, A. M. The retroviral enzymes. *Annu. Rev. Biochem.* **1994**, *63*, 137−173.
(7) Goff, S. P. Genetics of retroviral integration. *Annu. Rev. Genet.* **1992**, *26*, 527−544.
(8) Vink, C.; Plasterk, R. H. A. The human immunodeficiency virus integrase protein. *Trends Genet.* **1993**, *9*, 433−437.
(9) Craigie, R. Hotspots and Warm Spots: Integration Specificity of Retroelements. *Trends Genet.* **1992**, *8*, 187−190.
(10) LaFemina, R. L.; Schneider, C. L.; Robibins, H. L.; Callahan, P. L.; LeGrow, K.; Roth, E.; Schleif, W. A.; Emini, E. A. Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. *J. Virol.* **1992**, *66*, 7414−7419.
(11) Sakai, H.; Kawamura, M.; Sakuragi, J.; Sakuragi, S.; Shibata, R.; Ishimoto, A.; Ono, N.; Veda, S.; Adachi, A. Integration is essential for efficient gene expression of human immunodeficiency virus type 1. *J. Virol.* **1993**, *67*, 1169−1174.
(12) Fesen, M. R.; Kohn, K. W.; Leteurtre, F.; Pommier, Y. Inhibitors of human immunodeficiency virus integrase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2399−2403.
(13) Fesen, M.; Pommier, Y.; Leteurtre, F.; Hiroguchi, S.; Yung, J.; Kohn, K. W. Inhibition of HIV-1 integrase by flavones, Caffeic Acid Phenethyl Ester (CAPE) and related compounds. *Biochem. Pharmacol.* **1994**, *48*, 595−608.
(14) Cushman, M.; Sherman, P. Inhibition of HIV-1 integration protein by aurintricarboxylic acid monomers, monomer analogs, and polymer fractions. *Biochem. Biophys. Res. Commun.* **1992**, *185*, 85−90.
(15) Carteau, S.; Mouscadet, J. F.; Goulaouic, H.; Subra, F.; Auclair, C. Inhibitory effect of the polyanionic drug suramin on the in vitro HIV DNA integration reaction. *Arch. Biochem. Biophys.* **1993**, *305*, 606−610.

(16) 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.
(17) Cushman, M.; Golebiewski, W. M.; Pommier, Y.; Mazumder, A.; Reymen, D.; De Clerq, E.; Graham, L.; Rice, W. G. Cosalane analogs with enhanced potencies as inhibitors of HIV-1 protease and integrase. *J. Med. Chem.* **1995**, *38*, 443−452.
(18) Mazumder, A.; Gupta, M.; Perrin, D. M.; Sigman, D. S.; Rabinovitz, M.; Pommier, Y. Inhibition of human immunodeficiency virus type 1 integrase by a hydrophobic cation: the phenanthroline-cuprous complex. *AIDS Res. Hum. Retroviruses* **1995**, *11*, 115−125.
(19) Mazumder, A.; Gazit, A.; Levitzki, A.; Nicklaus, M.; Yung, J.; Kohlhagen, G.; Pommier, Y. Effects of Tyrphostins, Protein Kinase Inhibitors, on Human Immunodeficiency Virus Type 1 Integrase. *Biochemistry* **1995**, *34*, 15111−15122.
(20) Mazumder, A.; Cooney, D.; Agbaria, R.; Gupta, M.; Pommier, Y. Inhibition of human immunodeficiency virus type 1 integrase by 3′-azido-3′-deoxythymidylate. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5771−5775.
(21) LaFemina, R. A.; Graham, P. L.; Legrow, K.; Hastings, J. C.; Wolfe, A.; Young, S. D.; Emini, E. A.; Hazuda, D. J. Inhibition of human immunodeficiency virus integrase by bis-catechols. *Antimicrob. Agents Chemother.* **1995**, *39*, 320−324.
(22) Carteau, S.; Mouscadet, J. F.; Goul Goulaouic, H.; Subra, F.; Auclair, C. Effect of the topoisomerase inhibitors on the in vitro HIV DNA integration reaction. *Biochem. Biophys. Res. Commun.* **1993**, *192*, 1409−1414.
(23) Mazumder, A.; Gupta, M.; Pommier, Y. Methylphosphonodiester substitution near the conserved CA dinucleotide in the HIV LTR alters both extent of 3′-processing and choice of nucleophile by HIV-1 integrase. *Nucleic Acids Res.* **1994**, *22*, 4441−4448.
(24) Burke, T. R.; Fesen, M. R.; Mazumder, A.; Wang, J.; Carothers, A. M.; Grunberger, D.; Driscoll, J.; Kohn, K.; Pommier, Y. Hydroxylated aromatic inhibitors of HIV-1 integrase. *J. Med. Chem.* **1995**, *38*, 4171−4178.
(25) Pfeifer, K.; Merz, H.; Steffen, R.; Mueller, W. E. G.; Trumm, S.; Schulz, J.; Eich, E.; Schroeder, H. C. In Vitro anti-HIV activity of lignans-differential inhibition of HIV-1 integrase reaction, topoisomerase activity and cellular microtubules. *J. Pharm. Med.* **1992**, *2*, 75−97.
(26) Eich, E.; Pertz, H.; Kaloga, M.; Schulz, J.; Fesen, M. R.; Mazumder, A.; Pommier, Y. (−)-Arctigenin as a lead structure for inhibitors of human immunodeficiency virus type-1 integrase. *J. Med. Chem.* **1996**, *39*, 86−95.
(27) Robinson, W. E.; Reinecke, M. G.; Abdel-Malek, S.; Jia, Q. and Chow, S. A. Inhibitors of HIV-1 replication that inhibit HIV integrase. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6326−6331.
(28) Zhao, H.; Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Milne, G. W. A.; Pommier, Y.; Burke, T. R. Coumarin-Based Inhibitors of HIV Integrase. *J. Med. Chem.* **1997**, *40*, 242−249.
(29) Mazumder, A.; Wang, S.; Neamati, N.; Nicklaus, M.; Sunder, S.; Chen, J.; Milne, G. W. A.; Rice, W. G.; Pommier, Y. Antiretroviral agents as inhibitors of both human immunodeficiency virus type 1 integrase and protease. *J. Med. Chem.* **1996**, *39*, 2472−2481.
(30) Martin, Y. C. 3D database searching in drug design. *J. Med. Chem.* **1992**, *35*, 2145−2154.
(31) Martin, Y. C.; Bures, M. G.; Willett, P. Searching databases of three-dimensional structures. In *Review in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1990; pp 213−263.
(32) Sheridan, R. P.; Rusinko, A.; Nilakantan, R.; Venkataraghavan, R. Searching for pharmacophores in large coordinate data bases and its use in drug design. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8165−8169.
(33) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Oritz 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) Bures, M. G.; Hutchins, C. W.; Maus, M.; Kohlbrenner, W.; Kadam, S.; Erichson, J. W. Using three-dimensional substructure searching to identify novel, nonpeptide inhibitors of HIV-1 protease. *Tetrahedron Comput. Methodol.* **1990**, *3*, 673−680.
(35) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380−384.
(36) Milne, G. W. A.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. W. The NCI Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1219−1224.
(37) The National Cancer Institute Drug Information System was designed and built by NCI; see: Milne, G. W. A.; Miller, J. A.; et al. The NCI Drug Information System. Parts 1−6. *J. Chem. Inf. Comput. Sci*. **1986**, *26*, 154−197.
(38) Chem-X is provided by Chemical Design Ltd., Roundway House, Cromwell Park, Chipping Norton, Oxon OX7 5SR, U.K. Chem-Model is the model builder module in Chem-X. ChemDBS-3D is the 3D database build and search module in Chem-X.
(39) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(40) Wang, S.; Zaharevitz, D. W.; Sharma, R.; Marquez, V. E.; Lewin, N. E.; Du, L.; Blumberg, P. M.; Milne, G. W. A. The Discovery of novel, structurally diverse protein kinase C agonists through computer 3D-database search. Molecular modeling studies. *J. Med. Chem.* **1994**, *37*, 4479−4489.
(41) Wang, S.; Milne, G. W. A.; Yan, X.; Posey, I.; Nicklaus, M. C.; Graham, L.; Rice, W. G. Discovery of novel, non-peptide HIV-1 protease inhibitors by pharmacophore searching. *J. Med. Chem.* **1996**, *39*, 2047−2054.
(42) 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.
(43) Most of the 152 compounds in the validation database have been previously published. For references, see ref 42. Structures are available as Supporting Information to ref 42.
(44) Compounds which are provided to NCI upon a confidential basis are called "discreet", and all data associated with "discreet" compounds are restricted. Compounds which are provided without conditions or restrictions are called "open", and all data associated with "open" compounds are publicly accessible.
(45) QUANTA/CHARMm, a molecular modeling system, is supplied by Molecular Simulations Inc., 9685 Scranton Rd, San Diego, CA 92121-3752.
(46) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMm: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(47) 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.
(48) Nicklaus, M. C.; Wang, S.; Driscoll, J. D.; Milne, G. W. A. Conformational Changes of Small Molecules Binding to Proteins. *Bioorg. Med. Chem.* **1995**, *3*, 411−428.
(49) The failure of four-point pharmacophore searches where searching with three of the four points is successful has been noted previously in this laboratory. The reason for this is presumably that, whereas a three-point search needs to satisfy three distance constraints at the same time, going to four points *doubles* the number of interatomic distances. Having to fulfill six distance constraints simultaneously lowers the hit rate to zero in many instances. This is not surprising given the typical hit rate we have observed for three-point searches. This is very often a few hundred compounds out of 200 000, i.e., on the order of 10^{-3}. If extrapolation is permitted, one would expect a hit rate of $\sim 10^{-6}$ for six distances, or less than one compound out of the 200 000 on average. Sampling the conformational space of each compound in the database with smaller torsion angle steps during the search and/or build phases (not usually feasible, however, for reasons of computation time) might help increase the hit rates to some extent, but obvious geometrical boundary conditions will always pose an upper limit to the number of simultaneously fulfilled distance constraints. We do not take this to prove that no compound can ever attach to its protein binding site through more than three (noncovalent) attachment points but, rather, that the NCI 3D database, in spite of its size, covers the entire universe of possible compounds only in a sparse manner.
(50) Zhao, H.; Neamati, N.; Hong, H.; Milne, G. W. A.; Pommier, Y.; Burke, T. R., Jr. Hydrazide-Containing Inhibitors of HIV-1 Integrase. *J. Med. Chem.* **1997**, *40*, 937−941.
(51) The 2D structures entered in the NCI DIS database are those provided by the supplier of the samples to the NCI cancer screening program. Because of the volume of compounds entered (ca. 20 000/year), NCI does not conduct structural analysis of the samples submitted.
(52) Wang, S.; Kazanietz, M. G.; Blumberg, P. M.; Marquez, V. E.; Milne, G. W. A. Molecular Modeling and Site-Directed Mutagenesis Studies of a Phorbol Ester-Binding Site in Protein Kinase C. *J. Med. Chem.* **1996**, *39*, 2541−2553.

JM960754H