JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

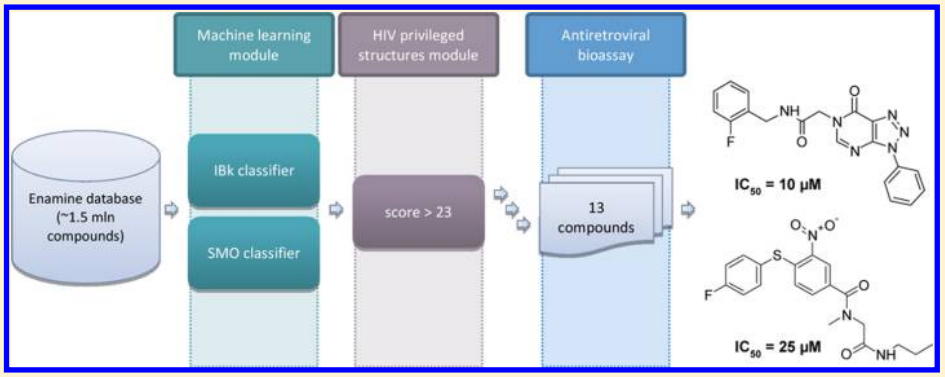
Ligand-Based Virtual Screening in a Search for Novel Anti-HIV-1 Chemotypes

Agata Kurczyk,[†,§,#] Dawid Warszycki,[‡,#] Robert Musiol,[§] Rafał Kafel,[‡] Andrzej J. Bojarski,[‡] and Jaroslaw Polanski*[,§]

[†]Institute of Automatic Control, Silesian University of Technology, 16 Akademicka Street, 44-100 Gliwice, Poland

[‡]Institute of Pharmacology, Polish Academy of Sciences, 12 Smetna Street, 31-343 Kraków, Poland

[§]Institute of Chemistry, University of Silesia, 9 Szkolna Street, 40-006 Katowice, Poland



ABSTRACT: In a search for new anti-HIV-1 chemotypes, we developed a multistep ligand-based virtual screening (VS) protocol combining machine learning (ML) methods with the privileged structures (PS) concept. In its learning step, the VS protocol was based on HIV integrase (IN) inhibitors fetched from the ChEMBL database. The performances of various ML methods and PS weighting scheme were evaluated and applied as VS filtering criteria. Finally, a database of 1.5 million commercially available compounds was virtually screened using a multistep ligand-based cascade, and 13 selected unique structures were tested by measuring the inhibition of HIV replication in infected cells. This approach resulted in the discovery of two novel chemotypes with moderate antiretroviral activity, that, together with their topological diversity, make them good candidates as lead structures for future optimization.

1. INTRODUCTION

Virtual screening (VS) is an important approach in the drug discovery process.[1,2] Methods used in VS can be divided into two classes: structure- or ligand-based techniques. In recent years, a combination of these methods emerged as an interesting and powerful alternative for the individual application of structure- or ligand-based approaches.[3] Thus, a question appeared whether ligand-based methods can still be productive when target structure data are available. Surprisingly, Meslamani et al. discovered that ligand-based methods outperform structure-based methods in the identification of true positives.[4] Thus, despite the increasing availability of crystal structure data for biological targets, ligand-based methods can still be an attractive VS option, particularly with the growing number of ligand data. Technically, VS is typically performed by sequentially filtering down the number of compounds in multistep experiments,[5−10] although single-step campaigns have also been reported.[11]

So far in both structure- and ligand-based VS protocols, machine learning (ML) procedures were often successfully applied.[12] The most common practical ML usage is to classify or prioritize databases of molecules in terms of biological activity,[13] specific ADMET properties[14] or various other chemical purposes.[15] ML classification can be performed as binary classification of molecules (as active or inactive) or as numerical predictions of certain property values determining which compounds possess more "promising" properties than others. This second approach is in fact the ML-based scoring function, which may be used as a decision function (e.g., for distinguishing between active and inactive compounds).

Yet another ligand-based approach for VS is the concept of privileged structures (PS), which was introduced by Evans[16] and further developed by others.[17−19] The PS concept is based on the assumption that certain structural features produce biological effects more often than others.[20] The theory of PS evolved from a pragmatic tendency to simplify the complexity of drug design through fragmentation and has previously been used in explaining structure−activity relationships in diverse groups of drugs.[21] If PS is shown to be useful, then molecular motifs that enrich biological activity can be used when designing new drugs.[22] Some molecular fragments in particular, such as quinoline, were claimed privileged among anticancer,[23]

Received: May 20, 2015
Published: October 2, 2015

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

antifungal,[24] antimicrobial[25] and, recently, HIV integrase (IN) inhibitors.[26] The PS method can be considered as an offspring of fragonomics, which evolved from the experimental measurements of protein−ligand interactions.[27] Recently, a new concept of encoding molecular descriptors for fragonomics into the framework of the molecular database records (FRAGTAL) was applied to disclose new fragments for HIV-1 IN inhibition.[28]

Among current problems in drug design, considerable emphasis is given to HIV IN inhibitors. These drugs have met great expectations that were unfulfilled until 2007 when the first inhibitor, raltegravir, was approved on the market. The importance of HIV IN may be explained by its specificity, with no analogues among human enzymes, and its essential role in the viral replication cycle.[29] The structure of IN has been extensively studied, and even larger sequences of amino acids, including more than one domain, have been resolved.[30,31] However, the full structure of the enzyme is poorly soluble and too flexible to be crystallized and successfully measured. A reasonably good structure (resolution 2.02 Å) was proposed by the Cherepanov group using the prototype foamy virus (PFV) as the model of HIV.[32−36] However, the lack of several details regarding the HIV IN structure and function makes the structure-based design of IN inhibitors ineffective.

In practice, the race for active inhibitors began after observation that topoisomerase II inhibitors, such as mitoxantrone and caffeic acid, may also effectively bind to HIV IN and hamper its reactivity.[37] Then, the identification of the diketoacid moiety (DKA) appeared as one great success of modern medicinal chemistry techniques. DKA was observed in several active compounds during random high throughput screening of 250 000 samples.[38] This functionality has been observed in a great majority of all active experimental drugs that act as IN inhibitors. However, because DKA is a small molecular fragment, it is necessary but not sufficient for activity. DKA must be mounted on another scaffold to become a fully active IN inhibitor. The discovery of the DKA fragment started the era of highly active compounds with prominent selectivity to the strand transfer step.[39,40] Consequently, raltegravir possesses DKA masked in a pyrimidinone carboamide scaffold. Additionally, various modifications of the supportive scaffold resulted in the development of other compounds, such as dolutegravir and GSK-364735.

The main objective of our research was to develop a novel and unique VS strategy to search for anti-HIV-1 chemotypes that would link the performance of methods based on artificial intelligence with the achievements of medicinal chemistry, particularly with regard to the latest concept of privileged fragments. Our VS protocol was based on compounds with determined HIV-1 IN inhibition potentials fetched from the ChEMBL database, and Klekota-Roth fingerprint (KlekFP) was used as molecular representation for developing classification models. KlekFP contains SMARTS patterns of chemical substructural motifs that enrich for biological activity.[22] Our VS protocol consists of two sequential modules: machine learning- (ML-module) and HIV-specific privileged fragments-based (HIV-PS-module). The HIV-PS-module and ML-module can form standalone screening units used for prioritization or classification tasks. However, these modules can be pipelined into a two-step cascade. We evaluated three possible cases: single ML-module, single HIV-PS-module, and both modules arranged in a two-step cascade. The results indicated an advantage of the multistep VS scenario compared with the single components' performances.

Finally, the developed two-step VS cascade was applied for screening a database of 1.5 million commercially available compounds, and 13 unique structures of potential activity were identified. Unlike anti-HIV agents registered in the Opportunistic Infection and Tuberculosis Therapeutics Database (http://chemdb.niaid.nih.gov/), which often indicated compounds with activity for different (HIV) targets,[28,41] only a single such interference in the ChEMBL database was identified. Thus, in our protocol the ChEMBL provides IN-specific references.

On the other hand, target activity is only one of the important requirements for the potential drug. Among the others, the activity in the real biological systems (biocompatibility) is probably even more important for the efficient drug candidate. We may assume that more or less biocompatibility is included in the PS concept. Accordingly, our hits should also meet the requirements of biocompatibility. To test this in a single analysis the bioactivity of the new compounds in an *in vitro* model measuring the inhibition of HIV replication in infected cells was tested. This approach allowed the discovery of two novel small-molecule chemotypes with antiretroviral activity.

2. MATERIALS

2.1. Data Sets and Data Preparation. The training sets used for developing the ML- and HIV-PS-modules were prepared as follows: (i) A set of 1140 compounds with determined HIV-1 IN inhibition was fetched from the ChEMBL v.12 database.[42,43] The library of these compounds was further divided according to HIV-1 IN inhibition range (Figure 1) to obtain

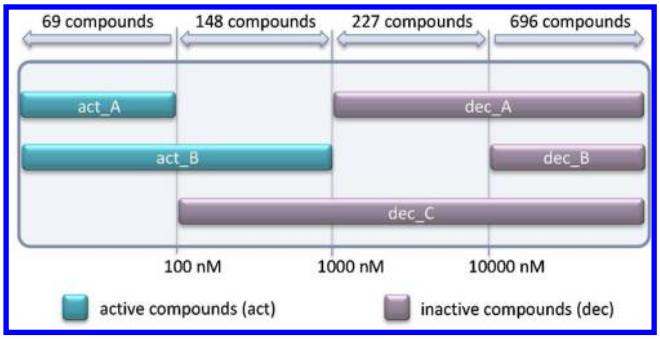


Figure 1. Distribution of HIV-1 IN inhibitors extracted from the ChEMBL database[42,43] according to their activity ranges. Vertical lines indicate corresponding IC_{50} values. Among 1140 IN inhibitors, 69 displayed IC_{50} (or equivalent) lower than 100 nM, 148 between 100 and 1000 nM, 227 between 1000 and 10 000 nM, and 696 above 10 000 nM.

subsets of actives and inactives. (ii) Sets of decoys were generated following the DUD[44] methodology using an in-house script that allowed setting different ratios between actives and inactives. (iii) The set of assumed inactives ("assumed_inact") was prepared from the ChEMBL Drugstore database[45] (consisting of approximately 1500 marketed drugs) by excluding compounds with reported anti-HIV activities. We are aware that those compounds are only putatively inactive;[46] however, because previously marketed drugs are widely examined, we consider the risk of being "false true negative" to be much lower than using randomly selected structures.

We made a comprehensive examination of HIV-1 IN inhibitors deposited in the ChEMBL database (version 12) in the context of other key HIV targets. Exploration of ChEMBL HIV-1 IN biostructural data revealed that the most active IN inhibitors are of high selectivity in their mechanism of action. Within

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

69 chemical structures with determined HIV-1 IN inhibition potentials ($IC_{50} \leq 100$ nM) there is not a single common compound with protease inhibitors and only one common structure with reverse transcriptase inhibitors. The total number of protease and reverse transcriptase ensembles was respectively equal to 3041 and 478 compounds ($IC_{50} \leq 100$ nM). Between protease inhibitors and reverse transcriptase inhibitors there are in turn 10 common structures.

Thirteen unique data sets were composed from subsets of the active and inactive compounds (Table 1) using the Canvas

Table 1. Data Sets Compositions

label	actives	inactives	actives/inactives ratio
I	act_A	dec_A	~1:13
II	act_B	dec_B	~1:3
III	act_A	dec_C	~1:15.5
IV	act_B	dec_A	~1:3
V	act_A	DUD	1:9
VI	act_B	DUD	1:9
VII	act_A	DUD	1:36
VIII	Act_B	DUD	1:36
IX	Act_A	assumed_inact	1:9
X	Act_A	assumed_inact	1:5
XI	Act_B	assumed_inact	1:5
XII	Act_A	assumed_inact	1:1
XIII	Act_B	assumed_inact	1:1

package of Schrödinger software.[47,48] Each of the 13 sets was divided into a training set (1/3 of compounds) and a test set (the remaining 2/3 of the initial collection). The training compounds were selected in three different ways: randomly, the most diverse (diverse-based selection tool) or reflecting population of clusters obtained from hierarchical clustering (performed in Canvas software) using the MOLPRINT2D fingerprint,[49,50] Tanimoto similarity (Tc) metric, average linkage method, and Kelly criterion. Finally, compounds from all sets were coded as KlekFPs (PaDEL-descriptor software)[51] for developing classification models and performing privileged fragments analysis and recognition.

2.2. Test Set for the VS Experiment. Four hundred and 50 compounds from the MDL Drug Data Repository (version from 2010) with anti-HIV activity that were not present in the training set used for methods development were used as active compounds. The test set was supplemented by adding DUD-like decoys at a ratio of 1:36 usually implemented in such experiments. Finally, the test set comprised 16650 records.

3. METHODS

3.1. Machine Learning-Based Module (ML-Module). A comprehensive set of experiments was conducted to assess the performance of seven ML methods (Naïve Bayes classifier,[52] Sequential Minimal Optimization,[53−55] Instance-Based Learning,[56] Decorate,[57,58] Hyper Pipes,[59] PART,[60] and Random Forest;[61] Table 2) tested on all data set compositions described in Section 2.1 (Table 1). Classifiers were constructed using algorithms implemented in the WEKA Data Mining Software version 3.6.[62] Calculations were performed on an Intel Core 2 Duo E840 CPU 3.0 GHz computer system with 8 GB RAM and a 64-bit Linux operating system.

The training process was conducted in accordance with the preassignment of molecules to classes of active and inactive compounds (Figure 2). Classifiers were used to categorize molecules as active or inactive against HIV-1 IN as a binary choice.

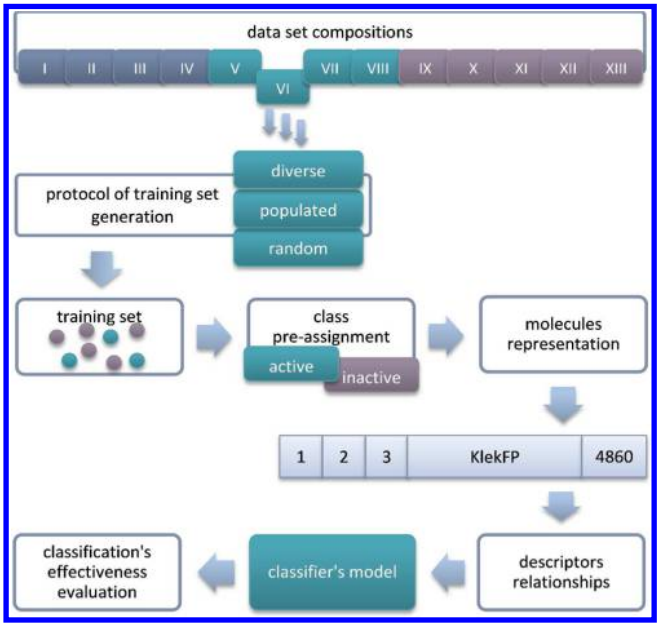


Figure 2. Development of the classifier models.

3.2. HIV-Specific Privileged Fragment-Based Module (HIV-PS-Module). We used the ChEMBL HIV IN specific database and the KlekFP substructures to identify an HIV-specific KlekFP subset. This HIV-specific privileged fragment-based module (HIV-PS-module) could serve as a weight-based scoring function that rates the presence of particular molecular

Table 2. Machine Learning Methods Used for ML-Module Development[a]

classifier	group of ML methods	settings and parameters
Naïve Bayes	bayes	
Sequential Minimal Optimization (SMO)	functions	A linear kernel was used, and an option of normalizing training data was specified. The complexity parameter was set as 1, and the epsilon for a round-off error was equal to 1.0×10^{-12}.
Instance-Based Learning (IBk)	lazy	k-nearest neighbor search algorithm with the Euclidean distance function. The number of neighbors was set as k = 1, 3, 7, and 11.
Decorate	meta	Naïve Bayes was chosen as the base classifier. The number of member classifiers in the ensemble was set as 10, the maximum number of iterations to run was limited to 10, and one artificial example was used during training.
Hyper Pipes	misc	
PART	Rules	PART decision list generated using a partial C4.5 decision tree in each iteration and making the "best" leaf into a rule. The minimum number of instances per rule was set as 2.
Random Forest	trees	Trees with unlimited depth. The number of trees generated was set as 10, and the random number seed used was set as 1.

[a]Only the IBk method was tuned by different values of nearest neighbors because preliminary studies showed that parameter optimization for other methods did not significantly change the results.

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

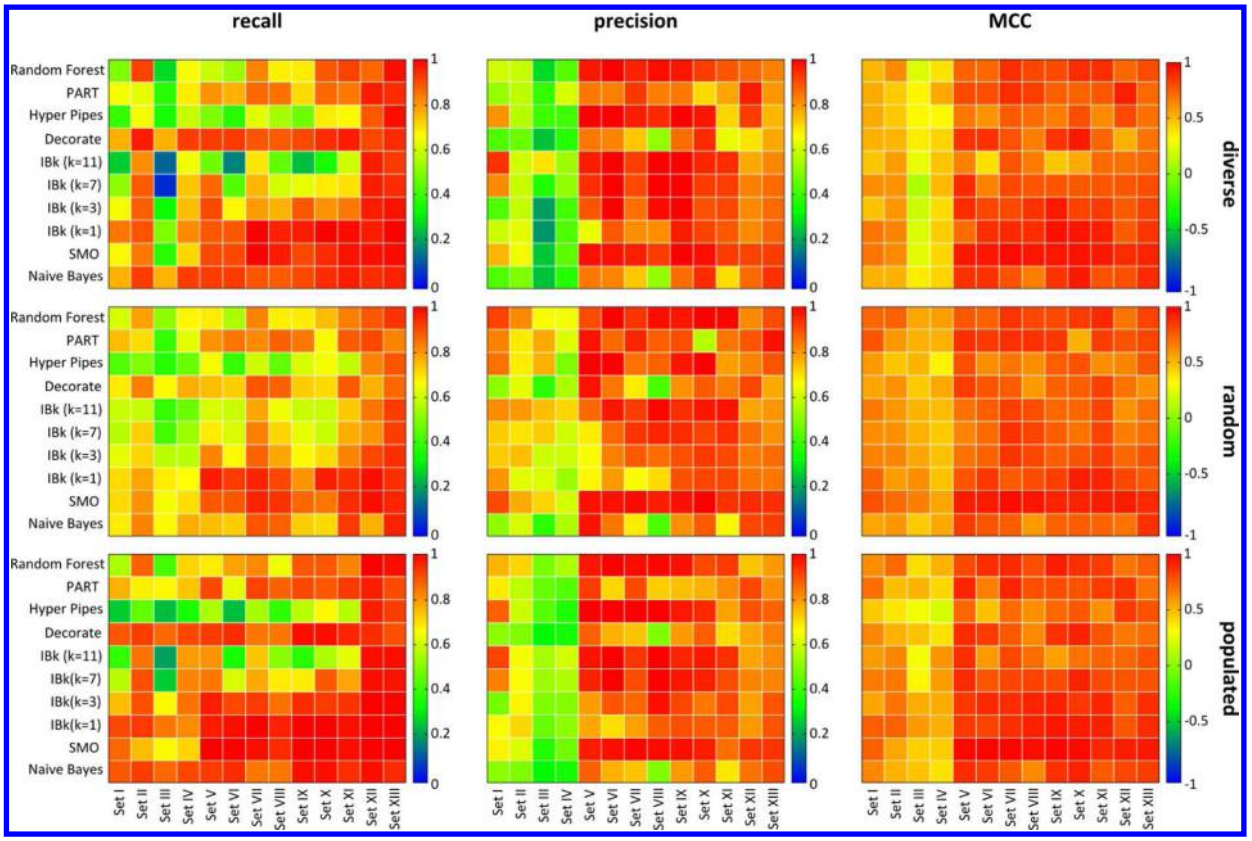


Figure 3. A panel of heat maps visualizing the values of calculated parameters (recall, precision, and MCC) obtained for classifiers tested on 13 different data sets comprised of HIV-1 IN inhibitors using the KlekFP as a numeric representation of molecules. Columns refer to particular data set compositions and considered parameters. Rows correspond to various ML methods and one of three protocols that were applied for training set generation.

fragments previously recognized as privileged in screened compounds. The recognition of privileged fragments was performed using the MI-DSE methodology, which is based on the Shannon entropy formalism originally derived from information theory.[63] MI-DSE analysis enables the identification of molecular descriptors that contain compound class-specific information. Moreover, MI-DSE can be applied to select discriminatory descriptors even for data sets that dramatically differ in size.[63]

3.3. Statistical Parameters To Measure the Quality of VS Scenarios. Binary classifiers were trained on each data set composition described in Section 2.1 (Table 1). The performances of the different methods were evaluated via additional test sets, which were derived from the initial 13 compound ensembles, as described in Section 2.1. Three parameters were calculated for ML methods evaluation: recall (eq 1), precision (eq 2), and the Matthews Correlation Coefficient − MCC (eq 3, Figure 3).

$$\text{recall} = \frac{TP}{TP + FN} = \frac{TP}{P} \tag{1}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{3}$$

The Matthews correlation coefficient (MCC) is a universal metric used in the evaluation of binary classifiers that considers true (T) and false (F) positives (P) and negatives (N). MCC is a balanced measure that can be used even if there is a large disproportion between classes. MCC returns a value between −1 and +1. A coefficient of +1 represents perfect prediction, 0 no better than random prediction, and −1 indicates total disagreement between prediction and observation. Recall measures the ratio of the correctly identified positive objects to the total abundance of positives in the test set. Precision calculates the ratio of correctly identified positives to all objects classified by the model as positive.

3.4. Biological Assay. The anti-HIV activity assays were performed on CEM-T4 cells. All compounds were dissolved in DMSO and diluted in cell culture medium (RPMI 1640 with 10% fetal bovine serum) to ensure the final concentration of the solvent to be <1%. The cells were infected with the wild type HIV-1 SI phenotype.

The cytotoxicity assay. In brief: cells were seeded in 96-well plates (20 000 per well) and incubated in 37 °C, 5% CO_2 in the presence of tested compounds for 7 days. Then, cell viability was measured according to the standard MTT procedure. The results (expressed as CC_{50}) are provided in Table 6 as the mean of three independent experiments.

The anti-HIV assay. The antiretroviral activity of the tested compounds was measured as the replication efficacy expressed by p24 protein concentration. Cells were seeded 20 000 per well in 96-well plates and incubated for 24 h at 37 °C, 5% CO_2 with serial dilution of the compounds. The cells were then inoculated with HIV virus, and incubation was continued for 7 days. Then, 80 μL of the supernatant was removed from each well, and p24 concentration was measured using the Genscreen HIV-1 Ag Assay (Bio-Rad, France). The remaining cells were

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

Table 3. ML Methods Ranked Based on the MCC Metric within One of Three Protocols That Was Applied for the Training Set Generation[a]

ML method	protocol of training set generation		
	diverse	random	populated
SMO	1	1	1
IBk (k = 1)	2	2	2
Naïve Bayes	3	8	7
IBk (k = 3)	4	5	3
PART	5	4	6
Random Forest	6	3	4
IBk (k = 7)	7	7	5
Decorate	8	9	8
Hyper Pipes	9	10	10
IBk (k = 11)	10	6	9

[a]Position in the ranking labeled as "1" indicates the method with the highest efficiency in classification.

Table 4. Privileged Structural Fragments Recognized within HIV-1 IN Inhibitors Specified According to Their KlekFP Bit Numbers with Corresponding SMARTS Pattern as Well as Privilegeness Expressed As a Weight

bit	SMARTS	counts
KR590	[!#1][CH2]c1[cH][cH]c(F)[cH][cH]1	13
KR1798	[!#1]c1[cH][cH]c(F)[cH][cH]1	11
KR1148	[!#1][OH]	10
KR2949	[OH]	9
KR3548	Cc1ccc(F)cc1	9
KR4067	Fc1ccccc1	9
KR582	[!#1][CH2]c1[cH][cH]c([!#1])[cH][cH]1	8
KR2260	[!#1]F	7
KR4032	F	7
KR1634	[!#1]c1[cH][cH][cH]nc1[!#1]	6
KR297	[!#1][CH2][!#1]	5
KR669	[!#1][CH3]	3
KR1193	[!#1]C(=O)[!#1]	3
KR1418	[!#1]C(=O)C(=O)[!#1]	3
KR1642	[!#1]c1[cH][cH]c([!#1])[cH][cH]1	3
KR3025	C=O	3
KR3706	CCCCC	3
KR4472	O=CC=O	3
KR1	[!#1][CH]([!#1])[!#1]	2
KR677	[!#1][NH][!#1]	2
KR1406	[!#1]C(=O)[OH]	2
KR2983	C(NCc1ccccc1)c2ccccc2	2
KR3750	CCN	2
KR4772	Oc1cccnc1	2
KR298	[!#1][CH2][CH]([!#1])[!#1]	1
KR3943	CNc1ccccc1	1
KR4080	N	1
KR4763	Oc1ccccc1C=O	1
KR4770	Oc1ccccc1O	1

subsequently tested in the MTT assay. The results (expressed as IC_{50}) are the mean of three independent experiments.

4. RESULTS AND DISCUSSION

4.1. Evaluation of the Performance of the ML Classifiers. Figure 3 presents the results of three evaluating parameters calculated by all ML methods for each of the 13 data sets generated in three different ways. To identify the most effective ML method, the arithmetic means of MCC values from all 13 data set compositions were calculated and ranked (Table 3). The results indicated the SMO classifier as the method with the highest efficiency for the classification of HIV-1 IN inhibitors. Only slightly worse results were obtained using the IBk classifier with the nearest neighbor set as 1.

4.2. Privileged Fragments of HIV-1 IN Inhibitors. MI-DSE analysis performed on 13 data sets described in Section 2.1 (Table 1) identified privileged molecular fragments (defined in SMARTS format corresponding to particular bits from KlekFP) that were recognized as particularly effective in distinguishing chemical structures of HIV-1 IN inhibitors from other compounds. Within each data set, recognized fragments were ranked according to their MI-DSE scoring. The number of training runs that indicates the same fragment as privileged defines its weight from 1 to 13. All runs resulted in 29 different structural subunits (Table 4).

4.3. Evaluation of Single Modules and Two-Step Virtual Screening Scenario. ML- and HIV-PS- modules were generated for use in developing the HIV-1 IN inhibitors VS protocol. Modules were tested as standalone screening units for prioritization (HIV-PS-module which can be interpreted as feature selection in the KlekFP substructure set) or classification (ML-module) tasks. Alternatively, we also designed a pipeline including both modules as a two-step cascade (Figure 4). It is particularly interesting whether such a cascade can provide any

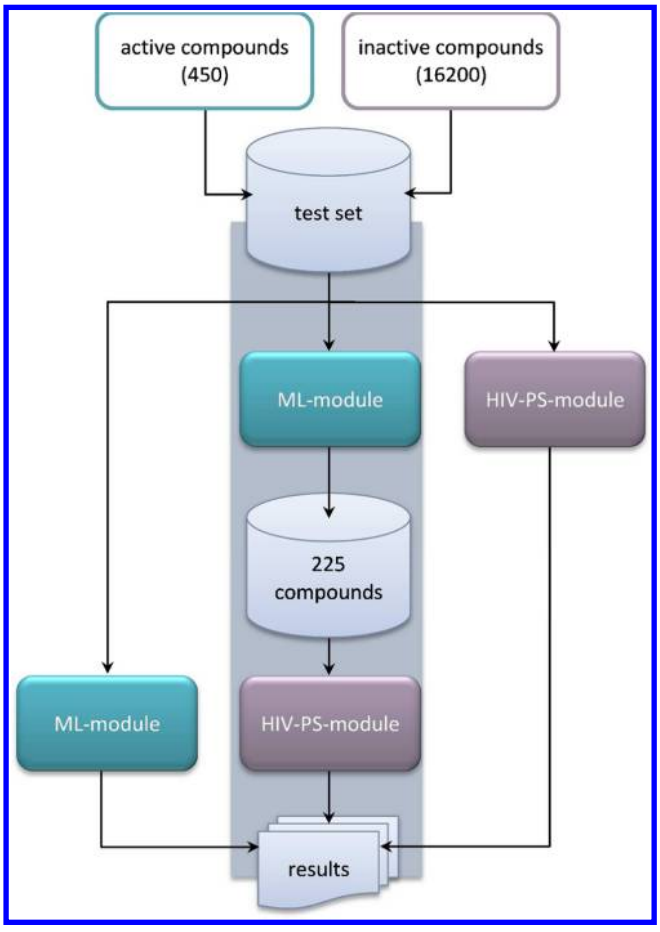


Figure 4. Virtual screening protocol performed on data extracted from the MDL Drug Data Repository. Both modules were evaluated separately and connected as a pipeline (labeled blue).

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

advantage because the pipeline formally including both modules could simply repeat the task of the single HIV-PS-module.

Evaluations of (i) single ML-module (Table 5), (ii) single HIV-PS-module, and (iii) cascade (Figure 5) in the VS experiment were performed using an external test set comprised of 450 anti-HIV drugs from MDDR and 16200 DUD decoys (see Methods). The sequential combination of developed modules was the most efficient method for real screening purposes, although both single modules also reached high statistical parameters (MCC for ML-module and AUC for HIV-PS-module). Unlike the ML module providing a binary classification, the HIV-PS-module and cascade output a continuous score that can be filtered to tune the precision and selectivity (recall) of the predictions. This result is illustrated through receiver operating characteristics (ROC) graphs (Figure 5).

Table 5. Evaluation of Single ML-Modules

classifier	TP	FP	FN	TN	recall	precision	MCC
IBk	147	78	303	16122	0.33	0.65	0.45
SMO	5	0	445	16200	0.01	1.00	0.10

The high true negative rate (despite the accompanying low values of other statistical parameters) makes the ML-module a perfect initial stage for any multistep screening purposes (Table 5). The HIV-PS-module is also characterized by a similar AUC value as a two-step cascade, and its potential application as a single filter bears significant time-cost benefits. Moreover, the initial slope of the ROC curve is higher, making the HIV-PS-single step screening scenario more relevant when only the ranking of compounds determines which instances will be tested. However, we observed that the sequential combination of modules was more efficient in the "liberal" classification range, i.e., for higher recall (northeastern part of the graph), in which the cascade mode provided higher precision than a single HIV-PS-module.

These results suggested that the cascade will be advantageous when we aim to diversify the drug candidate chemotypes under design. Because the search for novel chemotypes was also a priority in this study, we decided to apply a two-step screening cascade. In particular, to further increase precision, we combined compound rankings with expert evaluation.

4.3. Virtual Screening of a Commercial Database. One and a half million commercially available compounds in the Enamine database (in May 2012) were used as input for screening cascade consisting of both previously developed modules (Figure 6). In the ML-module, the IBk and SMO methods were applied in parallel, passing through 10665 and 3 instances, respectively. Because only one compound was identified by both methods, 10667 unique structures were evaluated in the second step of VS. The hits ranked with weights below 23 (the average weight of HIV-PS-module amounted to 23.53) obtained in the HIV-PS-module were rejected. For the remaining 5290 structures, MOLPRINT2D fingerprints[49,50] were generated and hierarchically clustered in the Canvas software using the default settings (Tc similarity metric and average cluster linkage method). Finally, using the Kelly criterion,[64] 385 clusters were obtained. From each cluster, structures with only the highest HIV-PS-module scores were further investigated, reducing the subset of compounds to 978 structures (often several compounds in the same cluster possessed the

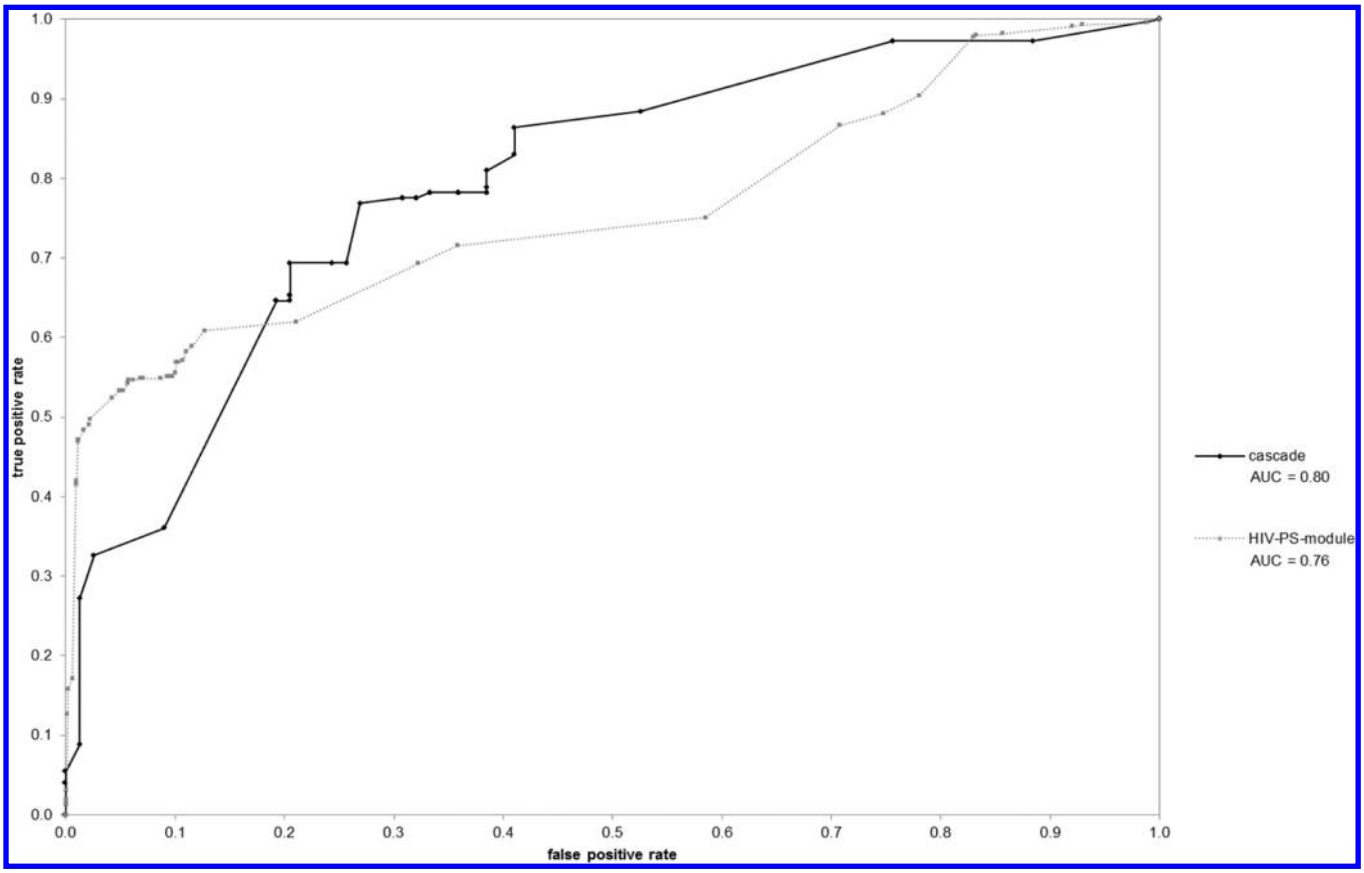


Figure 5. ROC graph showing the single HIV-PS-module and cascade.

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

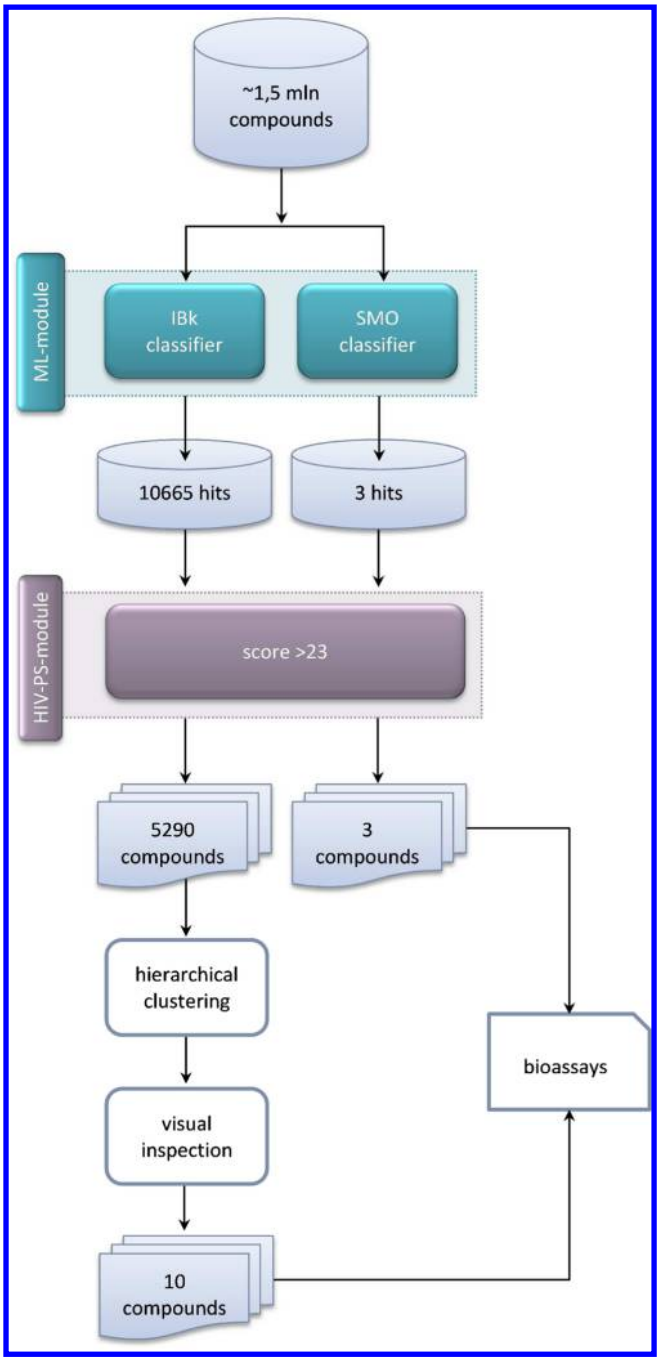


Figure 6. Ligand-based virtual screening protocol for searching for novel HIV-1 IN inhibitors.

same HIV-PS-score). Then, they were divided into ten groups based on their chemical similarities (hierarchical clustering forced to produce 10 clusters) and evaluated by team members to select compounds with the most diverse chemotypes (one per cluster). Each selected compound exhibited Tc similarity lower than 0.64 against any compounds with determined HIV-1 IN activity. These structures and structures positively evaluated by the SMO method in the ML-module were finally acquired and tested *in vitro* for inhibiting HIV replication in infected cells (Table 6).

Four tested compounds appeared to actively inhibit HIV replication in infected cells ($IC_{50} \leq 32$ μM). Their levels of activity were comparable within an order of magnitude to the reference compound FZ-41 used here as a relatively moderate activity compound that is a typical styrylquinoline-type HIV IN inhibitor.[65,66] Among the four most active inhibitors observed, two structures possessed the DKA motif, confirming the importance of this structural feature. Notably, although theoretical analysis indicated the SMO classifier as of the highest precision, the indicated hits (T6412586, T6412424, T553419) were the least active ($102 \leq IC_{50} \leq 112$ μM). Noticeably, all tested compounds were considerably toxic against normal uninfected cells. Only T5561007 displayed a therapeutic index above 2. Structural analysis of the most active compounds revealed the importance of two amide groups −C(O)NH− preferably separated by one sp^3 carbon atom. If the two motifs were separated by two atoms, the activity drops dramatically (compare T5894322 and T6070186 or T6802914).

Bearing in mind the results of the statistical analysis of ChEMBL which showed the specificity against HIV targets, it is prospective that the compounds indicated by our protocol VS and having measured anti-HIV-1 activity in the cell-based assay will operate through the mechanism of HIV-1 IN inhibition. At the same time testing our hits against real infected cells provides the results which could be treated as more realistic, i.e., having much higher potential for further application in the drug design pipeline. For example, a number of natural (i.e., caffeic acid phenethyl ester, quercetin, 3,5-dicaffeoylquinic acid) or synthetic (catechols, flavones) polyhydroxylated aromatic IN inhibitors inhibit isolated enzyme *in vitro* but failed in cell-based assays. While chicoric acid and curcumin provided antiviral activity in HIV-1 infected cells.[67] In the same context it is interesting to compare the activity of the DKA based inhibitors, e.g., the activity of L-731988 (lead compound) and 5CITEP determined as anti-IN activity amounted to 0.05 μM and 2.1 μM, respectively; however, antiviral activity of L-731988 in cell assays decreased to 1.0 μM.[68,69] The IC_{50} of the commercial HIV DKA based IN inhibitors have been optimized at the very low level, e.g., for raltegravir the IC_{50} amounts to 10 nM (see Figure 7).[70]

Interestingly, the activity of the reference FZ-41 determined vs isolated target IN were reported to amount to 0.7 to 1.7 μM for 3′-P reactions and strand transfer, respectively, while this increased to 12 μM in the cell-based assay[71] which compares well with the *in vitro* cell test determined in the current study 2 μM.

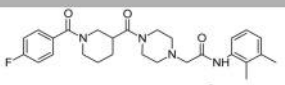
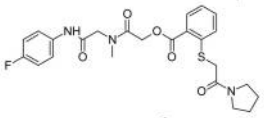
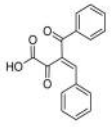
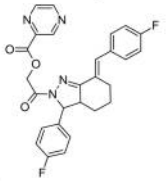
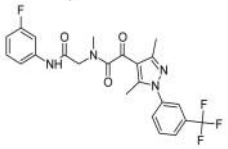
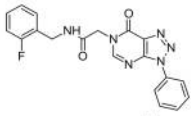
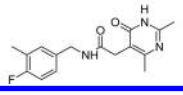
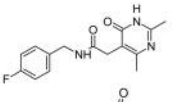
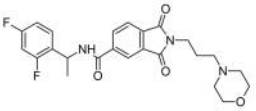
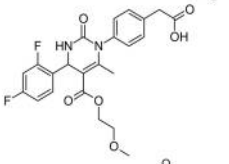
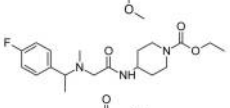
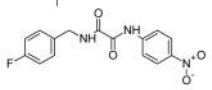
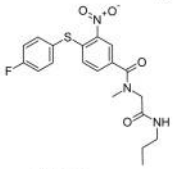
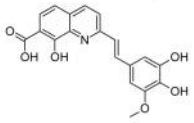
We believe that our approach can be an answer to the drawbacks in the target-oriented drug development approaches. It is the target, mode-of action, target validation, screening cascade that identifies current drug design dogma. However, we observe productivity crisis in pharma. According to some hypotheses, a fully rational target-based design is too optimistic at the current state of pharma technology, and we are more and more aware that understanding the mode-of-action of a drug is not a prerequisite for successful molecular design.[72] Our approach provide a scheme that connects both a target based scenario with a more pragmatic PS concept.

5. CONCLUSIONS

We combined a machine learning approach with the privileged structures concept to provide a multistep, ligand-based virtual screening protocol. In this protocol, we used both the machine learning approach and an HIV-specific subset of KlekFP substructures (HIV-PS-module) that we selected by mining the ChEMBL database for HIV IN inhibitors. Our results demonstrate that such a combination provides an efficient protocol that outperforms ML-modes alone. Moreover, our results indicated that feature selection KlekFP substructures that include more

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

Table 6. Results of anti-HIV and Cytotoxicity Assays

Compound ID[a]	Identified by SMO/IBk	Chemical structure	IC_{50} [μM]	CC_{50} [μM]	TI[b]
T5894322	IBk		65	> 65	> 1
T5537982	IBk		65	> 68.3	> 1.05
T5553419	SMO, IBk		105	> 119	> 1.13
T0510-0122	IBk		ND[c]	< 1.4	ND[c]
T6070186	IBk		22	15	0.68
T6802914	IBk		10	15	1.50
T6412424	SMO		102	> 110	> 1.08
T6412586	SMO		112	> 115.3	> 1.03
T6002810	IBk		> 72.5	> 72.7	~ 1.00
T6492823	IBk		> 72.3	> 72.3	~ 1.00
T6114980	IBk		85	> 91.3	> 1.07
T6420392	IBk		32	> 35	> 1.09
T5561007	IBk		25	70	2.80
FZ-41 (used as standard)	-		2	6	3.00

[a]Compound ID in accordance with indexing of products used in the Enamine database (http://www.enamine.net/). [b]Therapeutic index. [c]ND – not determined; antiviral activity was not determined because the compound displayed cytotoxicity at the lowest tested concentration (1.4 μM).

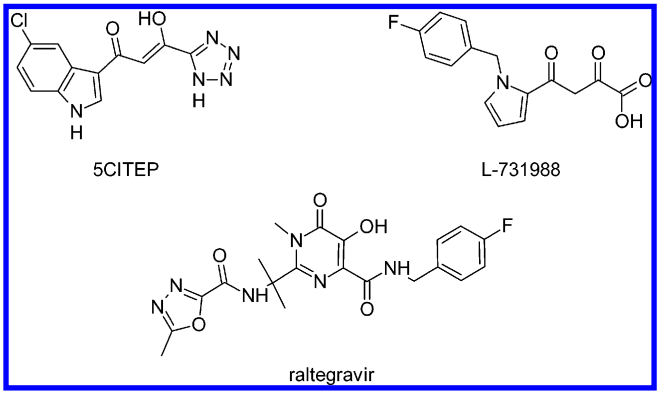


Figure 7. Structures of known HIV-1 IN inhibitors.

specific activity can be an interesting and efficient option in VS procedures.

From a practical point of view, our protocol was applied to virtually screen a database of commercially available 1.5 million compounds (the Enamine database). As a result, 13 potentially anti-HIV compounds were selected. Subsequently, these substances were tested *in vitro* for inhibiting HIV replication. Finally, two novel small-molecule chemotypes that act as antiretroviral compounds were discovered. These compounds have little similarity (Tc similarity less than 0.5) with respect to all HIV-1 chemotypes deposited in the ChEMBL database.[42,43] Although the reported molecules displayed only moderate potencies and exhibit some degree of cytotoxicity, their novel nature and topological diversity make them good candidates as lead structures for future optimization and successful development into highly potent anti-HIV compounds. On the other hand, we should understand that the compounds appeared to be active against HIV in the tests against real infected cells.

In conclusion, our work introduced the successful combination of machine learning methods with the privileged structures concept to create component modules of screening protocols. Our VS protocol application resulted in enriching the chemical space of anti-HIV compounds with two novel, structurally diverse anti-HIV agents.

AUTHOR INFORMATION

Corresponding Author

*E-mail: polanski@us.edu.pl.

Author Contributions

#These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was partially supported by the National Centre for Research and Development, Poland, project POIG.02.03.01-00-040/13. The authors thank Dr. A. Piasek for antiretroviral bioassay.

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168–2177

■ ABBREVIATIONS

HIV-1, human immunodeficiency virus type 1; VS, virtual screening; ML, machine learning; PS, privileged structures; IN, integrase; DUD, directory of useful decoys; KlekFP, Klekota-Roth fingerprint

■ REFERENCES

(1) Song, C. M.; Lim, S. J.; Tong, J. C. Recent Advances in Computer-Aided Drug Design. *Briefings Bioinf.* **2009**, *10*, 579−591.
(2) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.
(3) Drwal, M. N.; Griffith, R. Combination of Ligand- and Structure-Based Methods in Virtual Screening. *Drug Discovery Today: Technol.* **2013**, *10*, e395−e401.
(4) Meslamani, J.; Li, J.; Sutter, J.; Stevens, A.; Bertrand, H.-O.; Rognan, D. Protein-Ligand-Based Pharmacophores: Generation and Utility Assessment in Computational Ligand Profiling. *J. Chem. Inf. Model.* **2012**, *52*, 943−955.
(5) Luzhkov, V. B.; Selisko, B.; Nordqvist, A.; Peyrane, F.; Decroly, E.; Alvarez, K.; Karlen, A.; Canard, B.; Qvist, J. Virtual Screening and Bioassay Study of Novel Inhibitors for Dengue Virus mRNA Cap (nucleoside-2′O)-Methyltransferase. *Bioorg. Med. Chem.* **2007**, *15*, 7795−7802.
(6) Jacobsson, M.; Gäredal, M.; Schultz, J.; Karlén, A. Identification of Plasmodium Falciparum Spermidine Synthase Active Site Binders through Structure-Based Virtual Screening. *J. Med. Chem.* **2008**, *51*, 2777−2786.
(7) Lee, H. S.; Choi, J.; Kufareva, I.; Abagyan, R.; Filikov, A.; Yang, Y.; Yoon, S. Optimization of High Throughput Virtual Screening by Combining Shape-Matching and Docking Methods. *J. Chem. Inf. Model.* **2008**, *48*, 489−497.
(8) Knox, A. J. S.; Price, T.; Pawlak, M.; Golfis, G.; Flood, C. T.; Fayne, D.; Williams, D. C.; Meegan, M. J.; Lloyd, D. G. Integration of Ligand and Structure-Based Virtual Screening for the Identification of the First Dual Targeting Agent for Heat Shock Protein 90 (Hsp90) and Tubulin. *J. Med. Chem.* **2009**, *52*, 2177−2180.
(9) Kurczab, R.; Nowak, M.; Chilmonczyk, Z.; Sylte, I.; Bojarski, A. J. The Development and Validation of a Novel Virtual Screening Cascade Protocol to Identify Potential Serotonin 5-HT(7)R Antagonists. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2465−2468.
(10) Swann, S. L.; Brown, S. P.; Muchmore, S. W.; Patel, H.; Merta, P.; Locklear, J.; Hajduk, P. J. A Unified, Probabilistic Framework for Structure- and Ligand-Based Virtual Screening. *J. Med. Chem.* **2011**, *54*, 1223−1232.
(11) Song, C.-H.; Yang, S. H.; Park, E.; Cho, S. H.; Gong, E.-Y.; Khadka, D. B.; Cho, W.-J.; Lee, K. Structure-Based Virtual Screening and Identification of a Novel Androgen Receptor Antagonist. *J. Biol. Chem.* **2012**, *287*, 30769−30780.
(12) Melville, J. L.; Burke, E. K.; Hirst, J. D. Machine Learning in Virtual Screening. *Comb. Chem. High Throughput Screening* **2009**, *12*, 332−343.
(13) Hammann, F.; Gutmann, H.; Baumann, U.; Helma, C.; Drewe, J. Classification of Cytochrome p(450) Activities Using Machine Learning Methods. *Mol. Pharmaceutics* **2009**, *6*, 1920−1926.
(14) Schroeter, T.; Schwaighofer, A.; Mika, S.; Laak, A. T.; Suelzle, D.; Ganzer, U.; Heinrich, N.; Müller, K.-R. Machine Learning Models for Lipophilicity and Their Domain of Applicability. *Mol. Pharmaceutics* **2007**, *4*, 524−538.
(15) Podolyan, Y.; Walters, M. a; Karypis, G. Assessing Synthetic Accessibility of Chemical Compounds Using Machine Learning Methods. *J. Chem. Inf. Model.* **2010**, *50*, 979−991.
(16) Evans, B.; Rittle, K.; Bock, M.; DiPardo, R.; Freidinger, R.; Whitter, W.; Lundell, G.; Veber, D.; Anderson, P. Methods for Drug Discovery: Development of Potent, Selective, Orally Effective Cholecystokinin Antagonists. *J. Med. Chem.* **1988**, *31*, 2235−2246.
(17) Patchett, A. A.; Nargund, R. P. Privileged Structures — An Update. *Annu. Rep. Med. Chem.* **2000**, *35*, 289−298.
(18) Bondensgaard, K.; Ankersen, M.; Thøgersen, H.; Hansen, B. S.; Wulff, B. S.; Bywater, R. P. Recognition of Privileged Structures by G-Protein Coupled Receptors. *J. Med. Chem.* **2004**, *47*, 888−899.
(19) Duarte, C. D.; Barreiro, E. J.; Fraga, C. A. M. Privileged Structures: A Useful Concept for the Rational Design of New Lead Drug Candidates. *Mini-Rev. Med. Chem.* **2007**, *7*, 1108−1119.
(20) Kubinyi, H. Privileged Structures and Analogue-Based Drug Discovery. In *Analogue-based Drug Discovery*; Triggle, D. J., Taylor, J. B., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2006; pp 53−68.
(21) Polanski, J.; Kurczyk, A.; Bak, A.; Musiol, R. Privileged Structures - Dream or Reality: Preferential Organization of Azanaphthalene Scaffold. *Curr. Med. Chem.* **2012**, *19*, 1921−1945.
(22) Klekota, J.; Roth, F. P. Chemical Substructures That Enrich for Biological Activity. *Bioinformatics* **2008**, *24*, 2518−2525.
(23) Solomon, R.; Lee, H. Quinoline as a Privileged Scaffold in Cancer Drug Discovery. *Curr. Med. Chem.* **2011**, *18*, 1488−1508.
(24) Musiol, R.; Serda, M.; Hensel-Bielowka, S.; Polanski, J. Quinoline-Based Antifungals. *Curr. Med. Chem.* **2010**, *17*, 1960−1973.
(25) Musiol, R.; Magdziarz, T.; Kurczyk, A. Quinoline Scaffold as a Privileged Substructure in Antimicrobial Drugs. In *Science against microbial pathogens: communicating current research and technological advances*; Méndez-Vilas, A., Ed.; Formatex Research Center: 2011; Vol. *1*, pp 72−83.
(26) Musiol, R. Quinoline-Based HIV Integrase Inhibitors. *Curr. Pharm. Des.* **2013**, *19*, 1835−1849.
(27) Zartler, E. R.; Shapiro, M. J. Fragonomics: Fragment-Based Drug Discovery. *Curr. Opin. Chem. Biol.* **2005**, *9*, 366−370.
(28) Bak, A.; Magdziarz, T.; Kurczyk, A.; Serafin, K.; Polanski, J. Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments. *Comb. Chem. High Throughput Screening* **2013**, *16*, 274−287.
(29) Delelis, O.; Carayon, K.; Saïb, A.; Deprez, E.; Mouscadet, J.-F. Integrase and Integration: Biochemical Activities of HIV-1 Integrase. *Retrovirology* **2008**, *5*, 114.
(30) Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. Structure of a Two-Domain Fragment of HIV-1 Integrase: Implications for Domain Organization in the Intact Protein. *EMBO J.* **2002**, *20*, 7333−7343.
(31) Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, a H.; Leavitt, a D.; Stroud, R. M. Crystal Structure of the HIV-1 Integrase Catalytic Core and C-Terminal Domains: A Model for Viral DNA Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8233−8238.
(32) Hare, S.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. Retroviral Intasome Assembly and Inhibition of DNA Strand Transfer. *Nature* **2010**, *464*, 232−236.
(33) Li, X.; Krishnan, L.; Cherepanov, P.; Engelman, A. Structural Biology of Retroviral DNA Integration. *Virology* **2011**, *411*, 194−205.
(34) Hare, S.; Vos, A. M.; Clayton, R. F.; Thuring, J. W.; Cummings, M. D.; Cherepanov, P. Molecular Mechanisms of Retroviral Integrase Inhibition and the Evolution of Viral Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20057−20062.
(35) Maertens, G. N.; Hare, S.; Cherepanov, P. The Mechanism of Retroviral Integration from X-Ray Structures of Its Key Intermediates. *Nature* **2010**, *468*, 326−329.
(36) Cherepanov, P. Integrase Illuminated. *EMBO Rep.* **2010**, *11*, 328.
(37) 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.
(38) Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. Inhibitors of Strand Transfer That Prevent Integration and Inhibit HIV-1 Replication in Cells. *Science* **2000**, *287*, 646−650.
(39) Espeseth, a S.; Felock, P.; Wolfe, a; Witmer, M.; Grobler, J.; Anthony, N.; Egbertson, M.; Melamed, J. Y.; Young, S.; Hamill, T.; Cole, J. L.; Hazuda, D. J. HIV-1 Integrase Inhibitors That Compete with the Target DNA Substrate Define a Unique Strand Transfer

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177

Conformation for Integrase. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11244−11249.
(40) Serrao, E.; Odde, S.; Ramkumar, K.; Neamati, N. Raltegravir, Elvitegravir, and Metoogravir: The Birth of "Me-Too" HIV-1 Integrase Inhibitors. *Retrovirology* **2009**, *6*, 25.
(41) Serafin, K.; Mazur, P.; Bak, A.; Laine, E.; Tchertanov, L.; Mouscadet, J.-F.; Polanski, J. Ethyl Malonate Amides: A Diketo Acid Offspring Fragment for HIV Integrase Inhibition. *Bioorg. Med. Chem.* **2011**, *19*, 5000−5005.
(42) ChEMBL_12, ChEMBL-EBI. http://www.ebi.ac.uk/chembldb/Index.php (accessed December 6, 2011).
(43) Warr, W. A. ChEMBL. An Interview with John Overington, Team Leader, Chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). *J. Comput.-Aided Mol. Des.* **2009**, *23*, 195−198.
(44) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(45) Drugstore, ChEMBL-EBI. http://www.ebi.ac.uk/chembldb/drugstore (accessed December 12, 2011).
(46) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing Pitfalls in Virtual Screening: A Critical Review. *J. Chem. Inf. Model.* **2012**, *52*, 867−881.
(47) *Canvas*, Version 1.5; Schrödinger, LLC: New York, NY, 2012.
(48) Duan, J.; Dixon, S. L.; Lowrie, J. F.; Sherman, W. Analysis and Comparison of 2D Fingerprints: Insights into Database Screening Performance Using Eight Fingerprint Methods. *J. Mol. Graphics Modell.* **2010**, *29*, 157−170.
(49) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *J. Chem. Inf. Model.* **2004**, *44*, 170−178.
(50) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Model.* **2004**, *44*, 1708−1718.
(51) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466−1474.
(52) John, G. H.; Pat, L. *Estimating Continuous Distributions in Bayesian Classifiers*. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*; Morgan Kaufmann Publishers: San Mateo, 1995.
(53) Keerthi, S. S.; Shevade, S. K.; Bhattacharyya, C.; Murthy, K. R. K. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Comput.* **2001**, *13*, 637−649.
(54) Hastie, B. T.; Tibshirani, R. Classification by Pairwise Coupling. *Ann. Stat.* **1998**, *26*, 451−471.
(55) Platt, J. C. Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In *Advances in Kernel Methods - Support Vector Learning*; Schölkopf, B., Burges, C., Smola, A., Eds.; MIT Press: Cambridge, MA, 1999; pp 185−208.
(56) Aha, D. W.; Kibler, D.; Albert, M. K. Instance-Based Learning Algorithms. *Mach. Learn.* **1991**, *6*, 37−66.
(57) Melville, P.; Mooney, R. J. *Constructing Diverse Classifier Ensembles Using Artificial Training Examples*. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-2003)*; Acapulco, Mexico, 2003; pp 505−510.
(58) Melville, P.; Mooney, R. J. Creating Diversity In Ensembles Using Artificial Data. *Inf. Fusion* **2004**, *6*, 99−111.
(59) Witten, I. H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann Publishers: 2005.
(60) Frank, E.; Witten, I. H. *Generating Accurate Rule Sets Without Global Optimization*. In *ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning*; Morgan Kaufmann Publishers: 1998; pp 144−151.
(61) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(62) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA Data Mining Software. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10.
(63) Wassermann, A. M.; Nisius, B.; Vogt, M.; Bajorath, J. Identification of Descriptors Capturing Compound Class-Specific Features by Mutual Information Analysis. *J. Chem. Inf. Model.* **2010**, *50*, 1935−1940.
(64) Kelly, J. A New Interpretation of Information Rate. *IEEE Trans. Inf. Theory* **1956**, *2*, 185−189.
(65) 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.
(66) Polanski, J.; Zouhiri, F.; Jeanson, L.; Desmaële, D.; d'Angelo, J.; Mouscadet, J.-F.; Gieleciak, R.; Gasteiger, J.; Le Bret, M. Use of the Kohonen Neural Network for Rapid Screening of Ex Vivo Anti-HIV Activity of Styrylquinolines. *J. Med. Chem.* **2002**, *45*, 4647−4654.
(67) Di Santo, R.; Costi, R.; Artico, M.; Ragno, R.; Greco, G.; Novellino, E.; Marchand, C.; Pommier, Y. Design, Synthesis and Biological Evaluation of Heteroaryl Diketohexenoic and Diketobutanoic Acids as HIV-1 Integrase Inhibitors Endowed with Antiretroviral Activity. *Farmaco* **2005**, *60*, 409−417.
(68) Fujishita, T.; Yoshinaga, T. Indole derivatives with antiviral activity (Osaka, JP), PCT Int. Appl.WO 9950245, 1999.
(69) Di Santo, R.; Costi, R.; Roux, A.; Miele, G.; Crucitti, G. C.; Iacovo, A.; Rosi, F.; Lavecchia, A.; Marinelli, L.; Di Giovanni, C.; Novellino, E.; Palmisano, L.; Andreotti, M.; Amici, R.; Galluzzo, C. M.; Nencioni, L.; Palamara, A. T.; Pommier, Y.; Marchand, C. Novel Quinolinonyl Diketo Acid Derivatives as HIV-1 Integrase Inhibitors: Design, Synthesis, and Biological Activities. *J. Med. Chem.* **2008**, *51*, 4744−4750.
(70) Evering, T. H.; Markovitz, M. HIV-1 Integrase Inhibitors. *PRN Notebook* **2008**, *13*, 1−9.
(71) Zouhiri, F.; Mouscadet, J. F.; Mekouar, K.; Desmaële, D.; Savouré, D.; Leh, H.; Subra, F.; Le Bret, M.; Auclair, C.; d'Angelo, J. Structure-Activity Relationships and Binding Mode of Styrylquinolines as Potent Inhibitors of HIV-1 Integrase and Replication of HIV-1 in Cell Culture. *J. Med. Chem.* **2000**, *43*, 1533−1540.
(72) Sams-Dodd, F. Is poor research the cause of the declining productivity of the pharmaceutical industry? An industry in need of a paradigm shift. *Drug Discovery Today* **2013**, *18*, 211−217.

DOI: 10.1021/acs.jcim.5b00295
J. Chem. Inf. Model. 2015, 55, 2168−2177