Development of multistep ligand-based virtual screening cascade methodology in a search for novel HIV-1 integrase inhibitors: 2. Privileged structures

Dawid Warszycki[a], Agata Kurczyk[b], Rafał Kafel[a], Robert Musioł[b], Andrzej J. Bojarski[a], Jarosław Polański[b]

[a]Institute of Pharmacology, Polish Academy of Sciences, 12 Smetna Street, 31-343 Kraków, Poland
[b]Institute of Chemistry, University of Silesia, 9 Szkolna Street, 40-006 Katowice, Poland
e-mail: warszyc@if-pan.krakow.pl

Introduction

HIV integrase which is essential in the virus replication cycle and has no homologue among human enzymes [1], became an important target for drug development more than twenty years ago. Nevertheless, progress has been hampered by the lack of assays suitable for high throughput screening. Thus, a real breakthrough was only observed in 2007 with the introduction of the first integrase inhibitor, raltegravir, into treatment.
Crystal structure for HIV-1 integrase is already known and thus, both techniques commonly used in VS campaigns (structure and ligand-based) could be developed. Here we introduced a multistep ligand-based screening cascade because it is suggested that ligand-based methods outperform structure-based in true positives identification [2]. Our strategy consists of two sequential modules: machine learning-based (ML-based) and privileged fragments-based (PF-based).

Privileged fragments-based module development

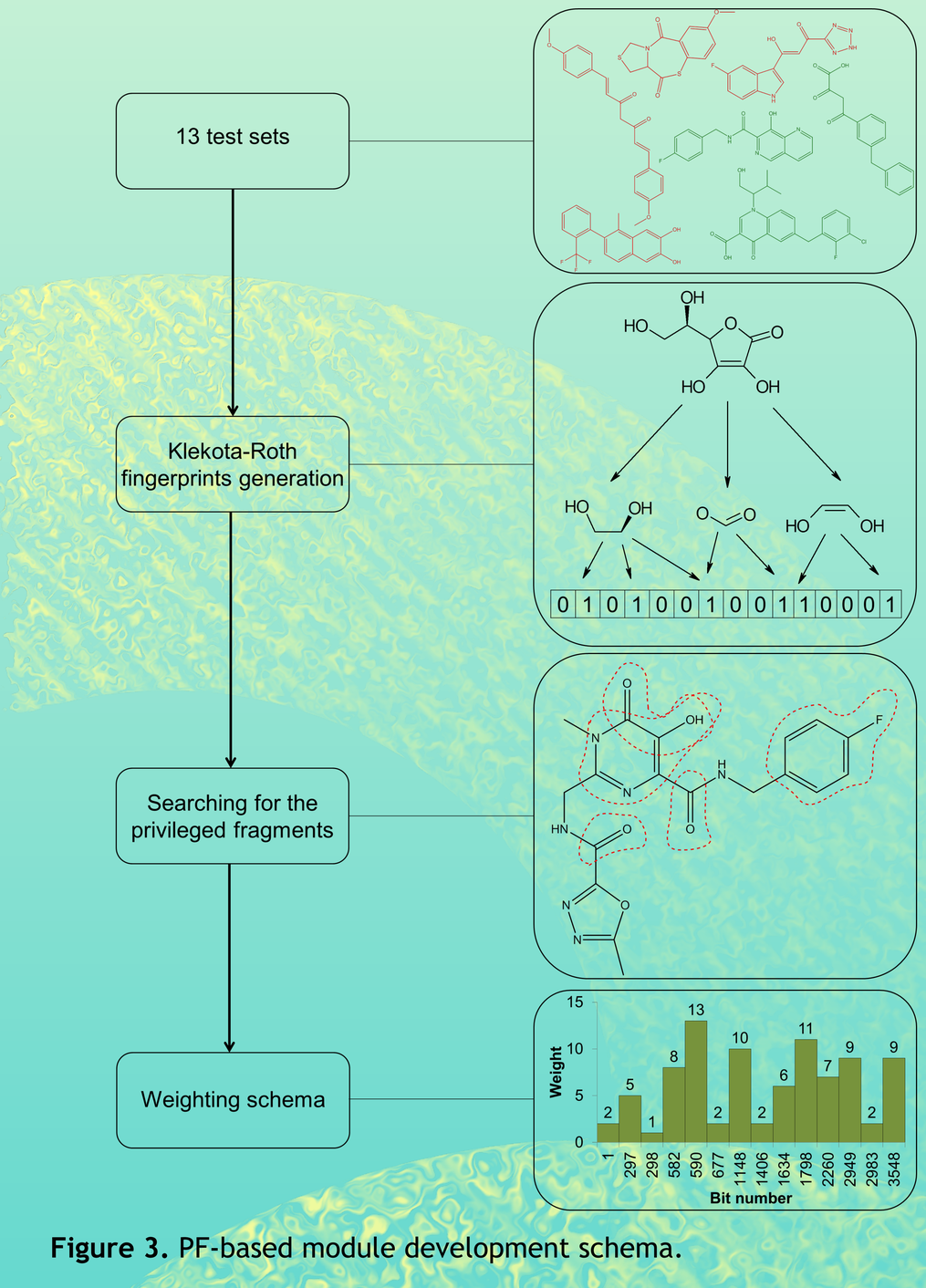


Figure 3. PF-based module development schema.

Conclusions

The enhancement of privileged structures theory, application of Klekota-Roth fingerprints for translating chemical structures and use of MI-DSE [5] formalism for search for the most discriminating fragments led to novel scoring function. VS experiments show that PF-based module can be used as a standalone screening filter, however, the best enrichment factors were reached when PF-based module was supported by machine learning in screening cascade. The developed methodology was applied to search of new HIV integrase inhibitors in databases of commercialy available compounds.

Priviliged fragments

Idea described herein is an extention of privileged structures (PS) concept intruduced by Johnson [4] and further developed by others. PS are understood as a chemical moieties (e.g. biphenyl etc.) which are effective in design of bioactive structures which are able to interact with more than one target whilst PF are chemical subunits (definied in SMARTS format corresponding to particular bits from Klekota-Roth fingerprint) specially effective in distinguishing active compounds from inactives for a particular target.

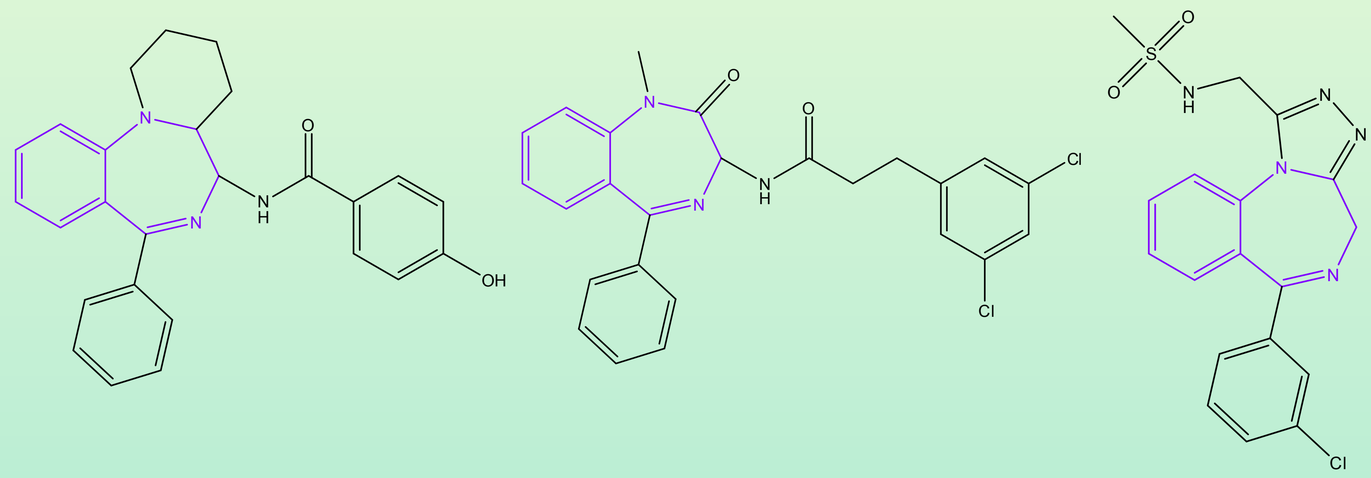


Figure 1. Examples of compounds interacting with different targets (chelocistokinin antagonist, potassium channel blocker and GABA-A agonist respectively) and containing benzoazephane as a privileged structure.

Virtual screening workflow

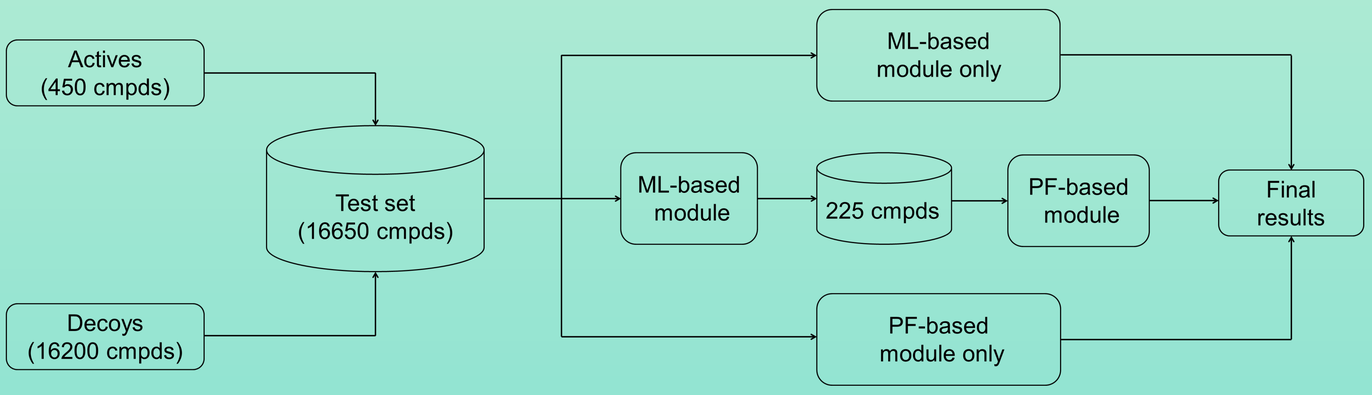
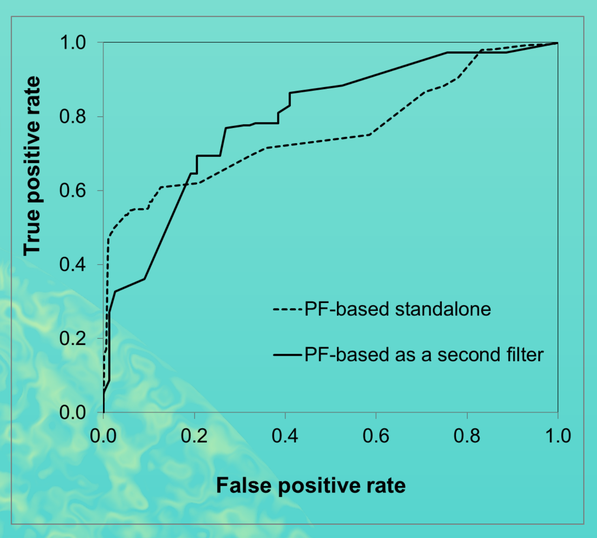


Figure 2. Virtual screening workflow. Test set was composed by actives not used in modules development. Decoys were generated basing on Directory of Useful Decoys methodology introduced by Huang et al. [4].

Results

The performance of VS protocols composed with the use of PF-based module is definied as an area under the curve (AUC), enrichment factor (specified as a change of actives fraction among all compounds population after and before screening stage) for 1% of top-scored instances ($EF_{1\%}$) and maximum value of EF.



Parameter	PF-based module		VS cascade
	standalone	second step	
$EF_{1\%}$	27.7	6.9	142.9
EF_{max}	288.0	21.2	1440.0
AUC	0.760	0.796	0.661

Figure 4. Left: the results of PF-based module performance. True poistive rate is fraction of correctly classified actives out of all actives, and false positive rate is fraction of incorrectly classified inactives out of all inactives. Right: statistical comparison of PF-based and full cascade (ML-based + PF-based) protocols.

[1] Deleis et al. Integrase and integration: biochemical activities of HIV-1 integrase., *Retrovirology*, 2008, 5, 114.
[2] Meslamani et al. Protein Ligand-Based Pharmacophores: Generation and Utility Assessment in Computional Ligand Profiling., *J. Chem. Inf. Model.*, 2012; 52, 943-55.
[3] Johnson et al. Concepts and applications of molecular similarity., Wiley & Sons, 1990.
[4] Huang et al. Benchmarking sets for molecular docking., *J. Med. Chem.*, 2006, 49, 6789-801.
[5] Wassermann et al. Identification of descriptors capturing class-specific features by mutual information analysis., *J. Chem. Inf Model.*, 2010, 50, 1935-40.

This poster is a continuation of poster entitled "Development of multistep ligand-based virtual screening cascade methodology in a search for novel HIV-1 integrase inhibitors: 1. Machine learning".