

TARGET-FOCUSED LIBRARIES: RAPID SELECTION OF POTENTIAL PDE INHIBITORS FROM MULTI-MILLION COMPOUNDS' REPOSITORIES

*Tünde Tömöri, Sándor Cseh and György
Dormán**

TargetEx Ltd,

May 26, 2010



Outline

Why do we need rapid selection of target-focused libraries?

Design inputs

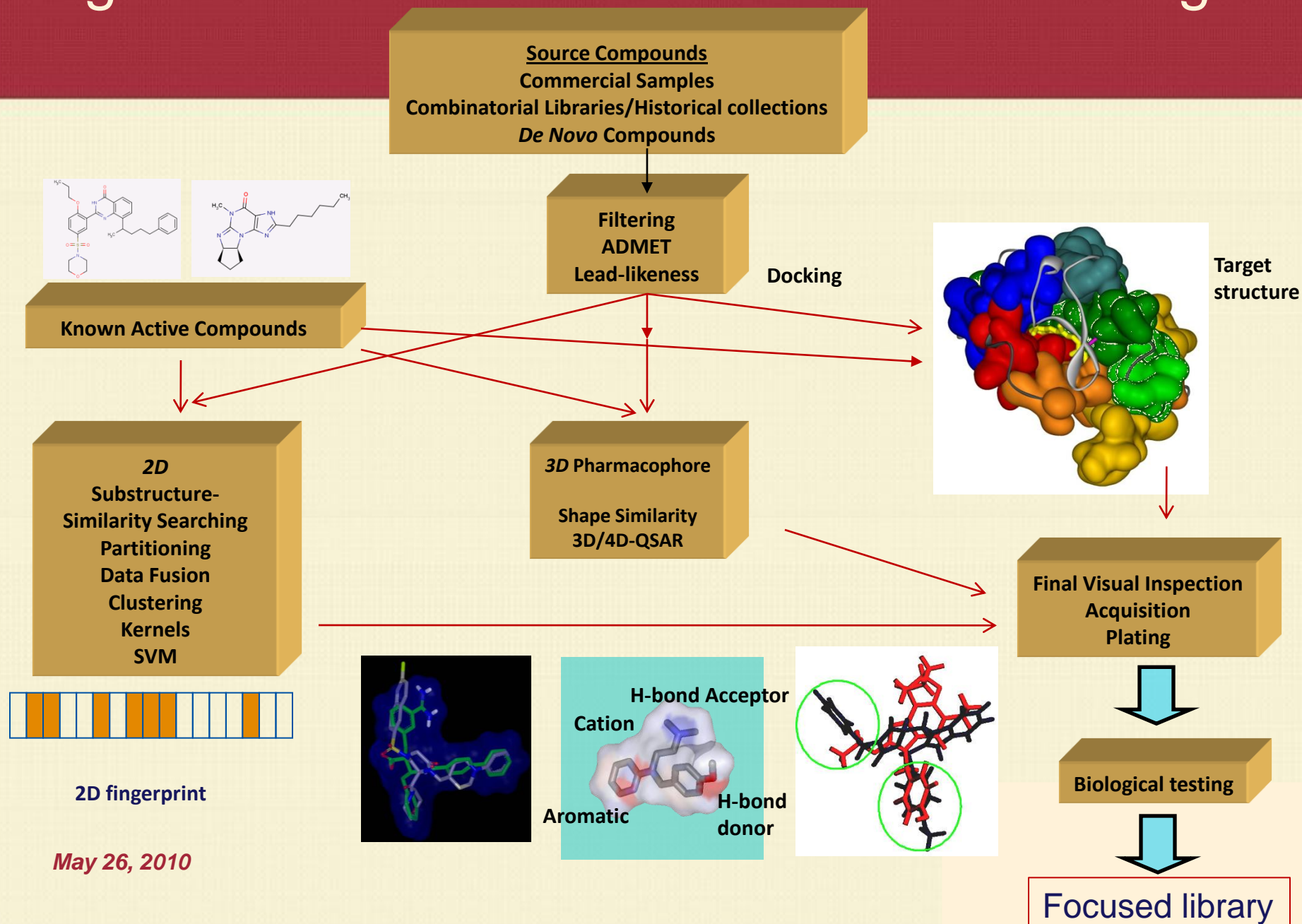
2D similarity searching strategy

Property-based filtering

Seed/ chemotype representation (diversity)

Conclusion/ Proposals

Target -focused libraries via Virtual Screening



Why we need rapid selection of target-focused libraries?

- Target-focused biological screening reduces the synthesis, repository management, screening costs and various assays are still available with lower throughput
- Biotech/ screening compounds do not have synthesis capacity
- The hit rate of the random screening is normally $\leq 0.1\%$, while for the focused library screening it can exceed $\geq 1\%$.
- **Simple robust technique is needed which is capable of selecting focused libraries from multi-million vendor libraries**
- **The solution:** „2D fingerprints are surprisingly effective in many search situations in comparison with more complex 3D designs” (*Jürgen Bajorath, ChemMedChem, 2007, 2, 208-217*)
- Similarity is expressed in Tanimoto coefficients
- Problem: size-dependent (more complex molecules have higher value than the real similarity)

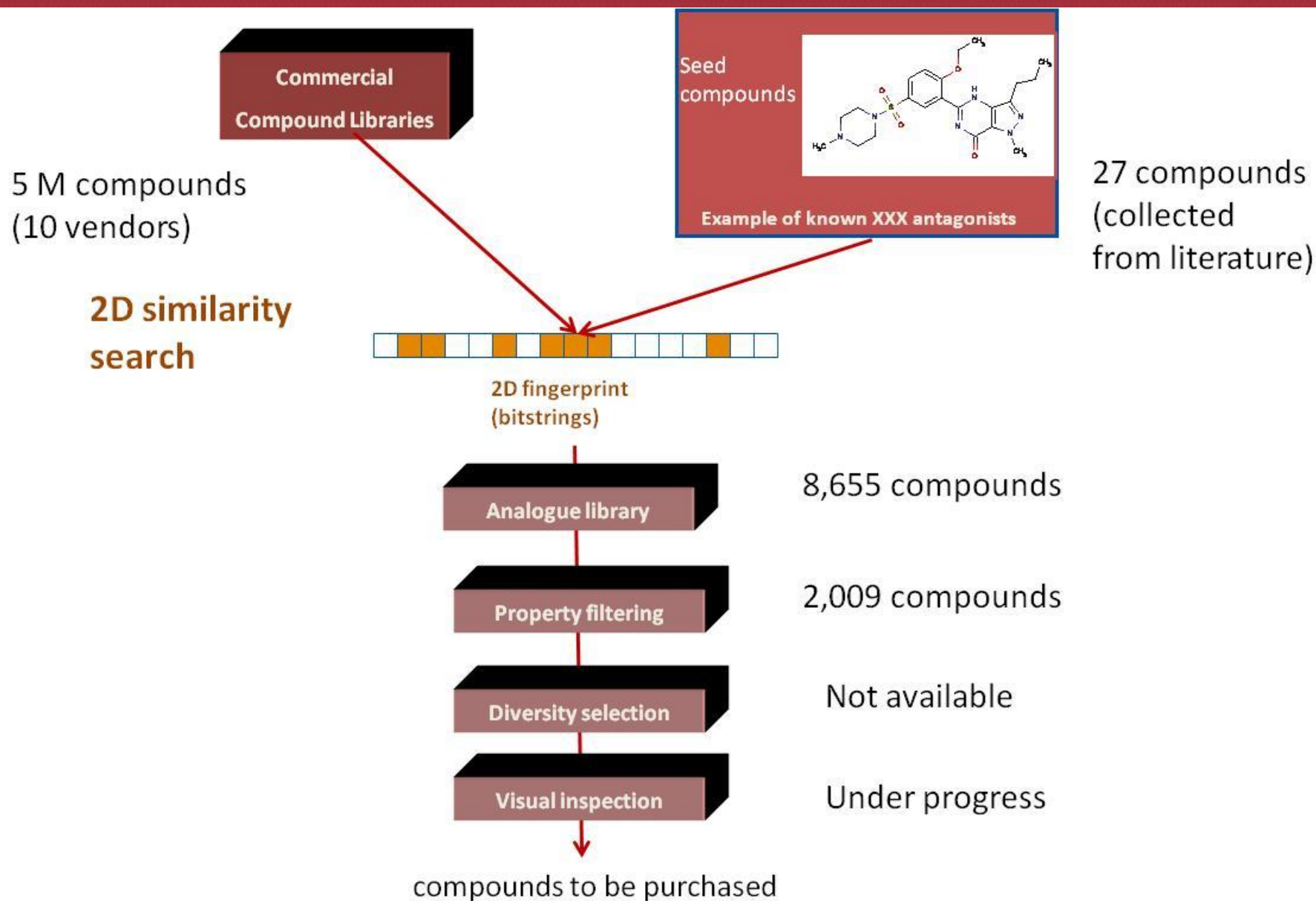
Design inputs I.

- Reference space: Known PDE5 inhibitors (27)
- Literature availability
- Extraction from literature and various commercial databases
- Problems:
 1. Often only the IUPAC or alternative chemical name is provided (IJChem)
 2. The compounds can be clustered (Jklustor?) to major chemotypes to remove the redundant structures (no!)
 3. The property space (Lipinski and Veber rules) can be defined for the „seed” compounds (IJChem)

Design inputs II.

- Chemical Space
 - Existing/ Real compounds
- Non-exclusive Libraries available
 - Zinc DB (25 M compounds)
 - Top10 library vendors: 5 M compounds
 - ChemBridge, ChemDiv, Asinex, Enamine, IfLab, UkrOrg, AMRI, Specs, Maybridge, IBS

2D similarity selection



Similarity searching strategy: execution

- Setting the starting similarity level (dependent on the fingerprint S/W, T= 60-75 % for ChemAxon)
- Iteration based on the results (scenarios):
- the number of virtual hits are between 50 and 500, OK
- the number of virtual hits are <50 or >500
 - if <50 lower the similarity threshold with 5 %
 - if >500 increase the similarity threshold with 5 %
 - This can be continued until the optimal range achieved
 - If 5 % decrease results in >500 compounds the search can be refined by 2% (alternatively a diversity selection would be needed, but that is not available)
 - Duplications can be removed when merged the resulting DBs

How to reduce the number of the hits?

Normally screening companies would like to buy 100-1000 compounds

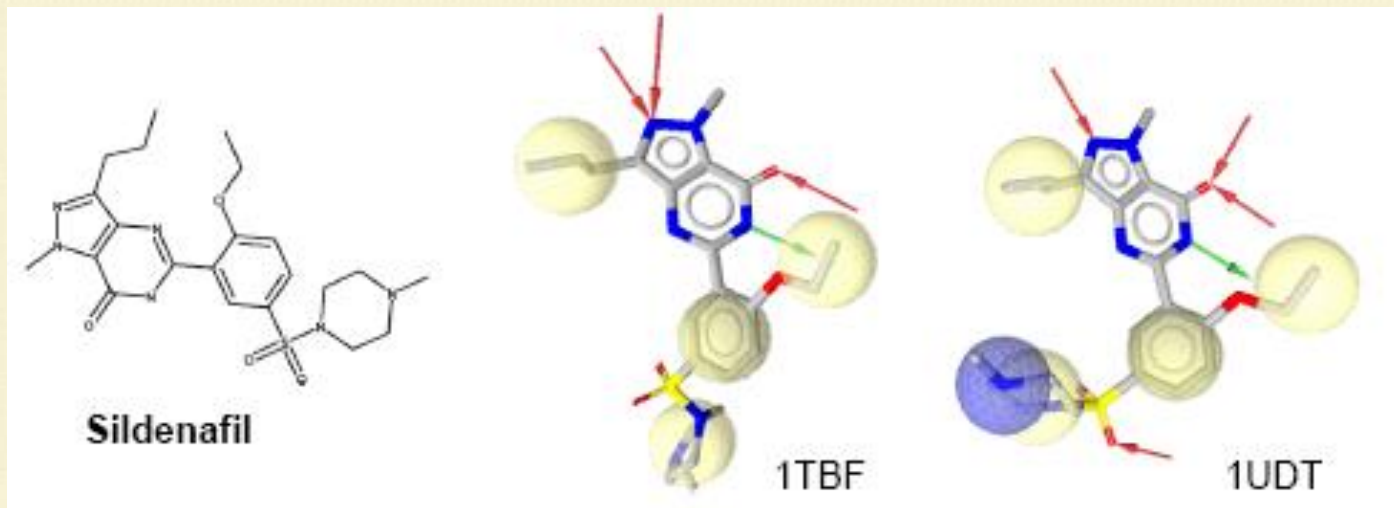
- Since from the various vendor DBs we can obtain 2000-10.000 virtual hits their number can be reduced
- 1. Applying the reference property space (Lipinski and Veber rules) (IJchem OK)
- 2. There are overrepresented seeds thus virtual hits coming from those seeds can be reduced (IJchem OK)
- 3. Applying an optimal distribution of the resulting chemotypes (removing the overrepresented compounds) (Limited with Jklustor)
- 4. Simple diversity analysis (No for IJchem)

May 26, 2010



1. Applying the reference property space:

Structural determinants: H-bond donor/ acceptor, hydrophobic interactions (property space determination)

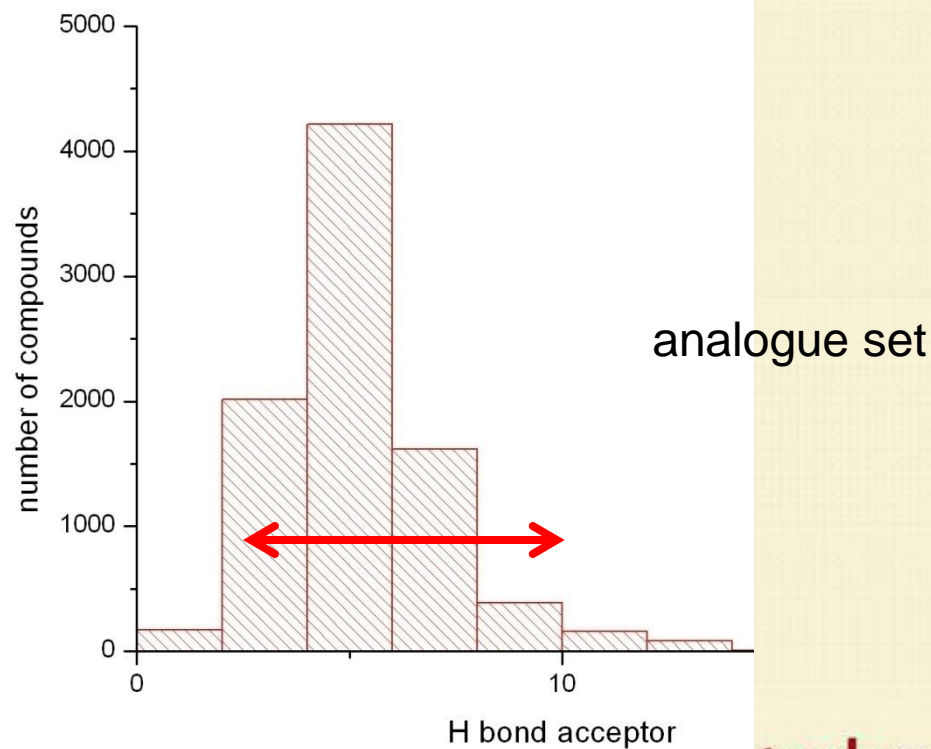
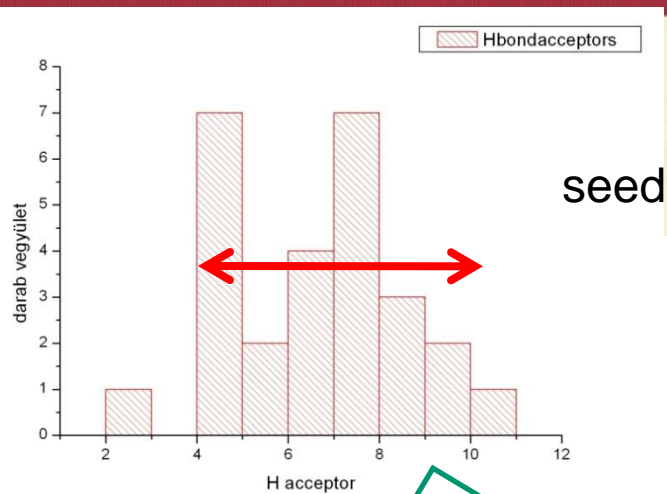


Pharmacophore fingerprints requires more computation and time consuming

In simple similarity search pharmacophore features can only be considered as statistical features (not connected to structures)

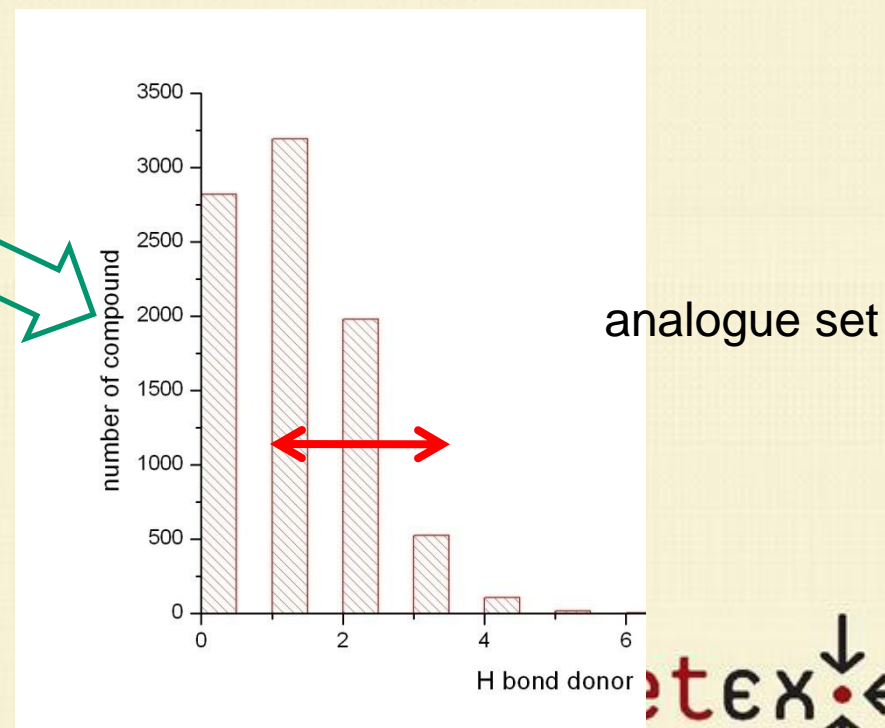
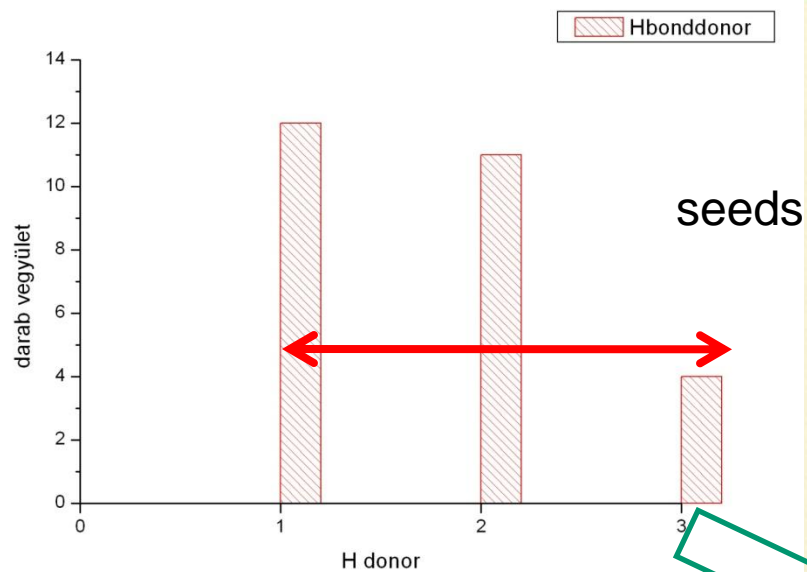
The similarity search results can be filtered based on the physico-chemical parameter space of the seed compounds (+10/-10 % range applied)

H-bond acceptors



May 26, 2010

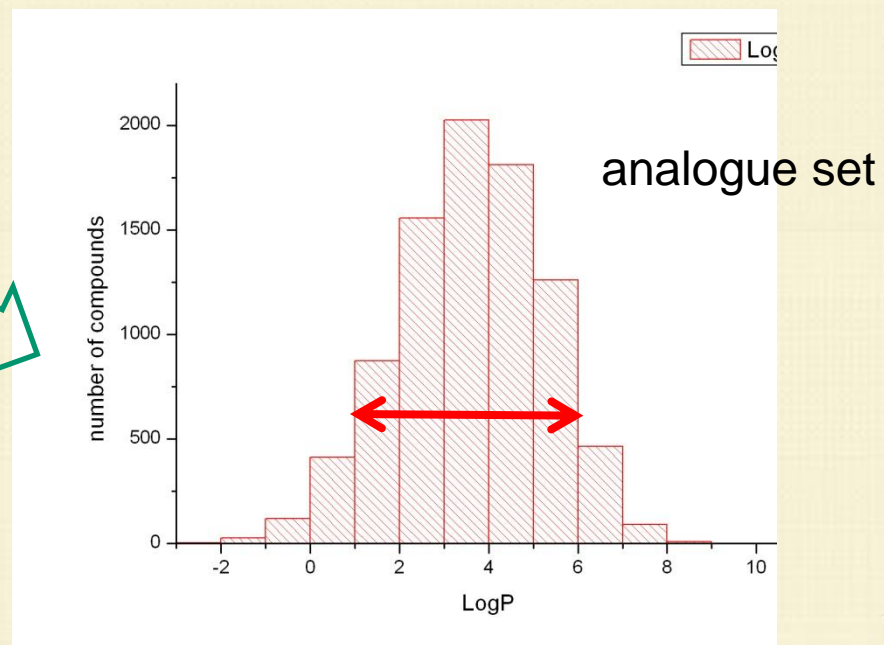
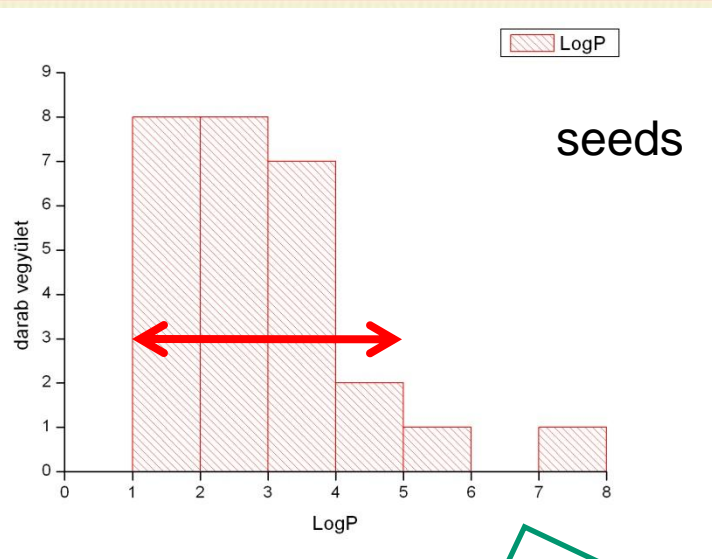
H-bond donors



May 26, 2010

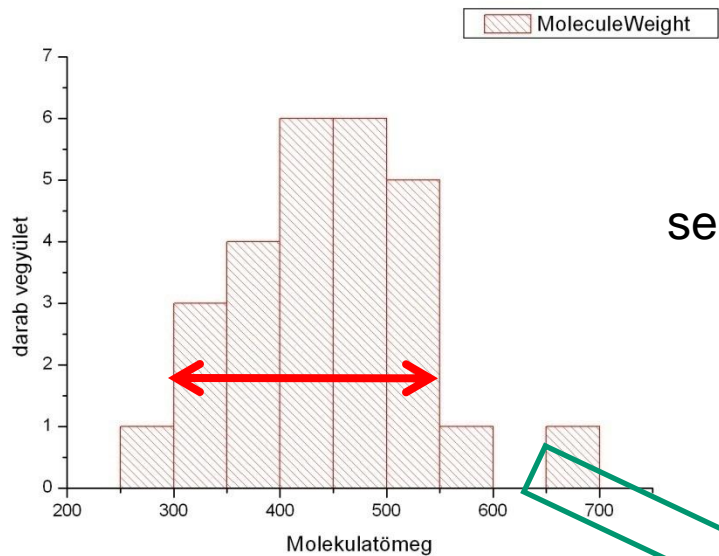


Log P

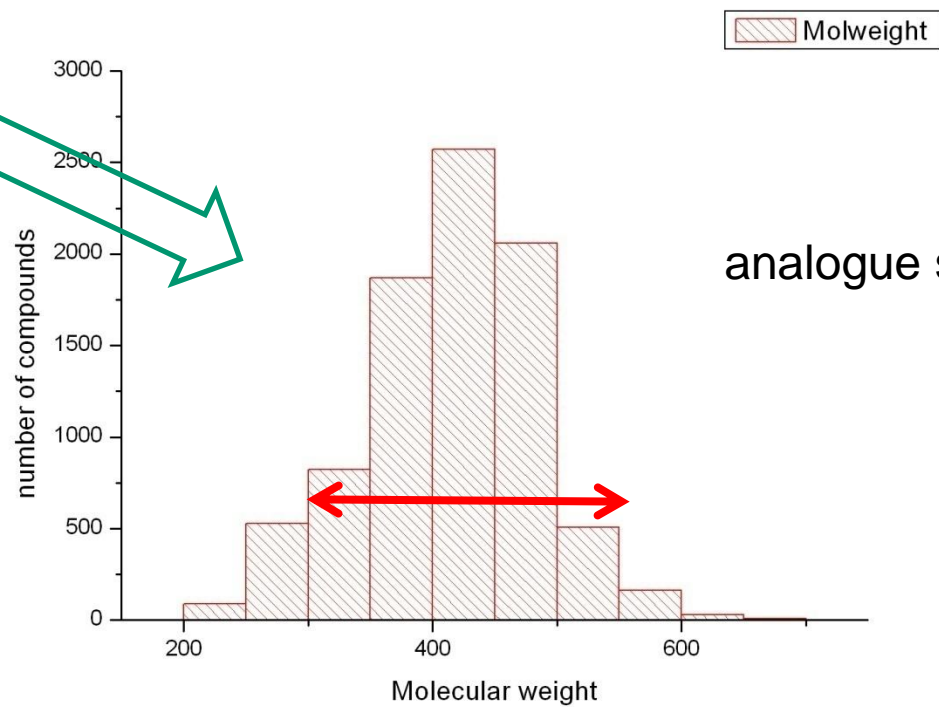


May 26, 2010

Molecular weight



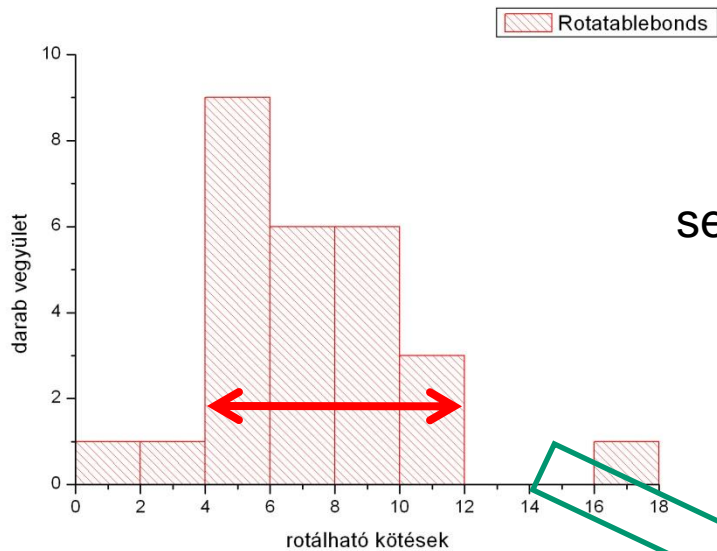
seeds



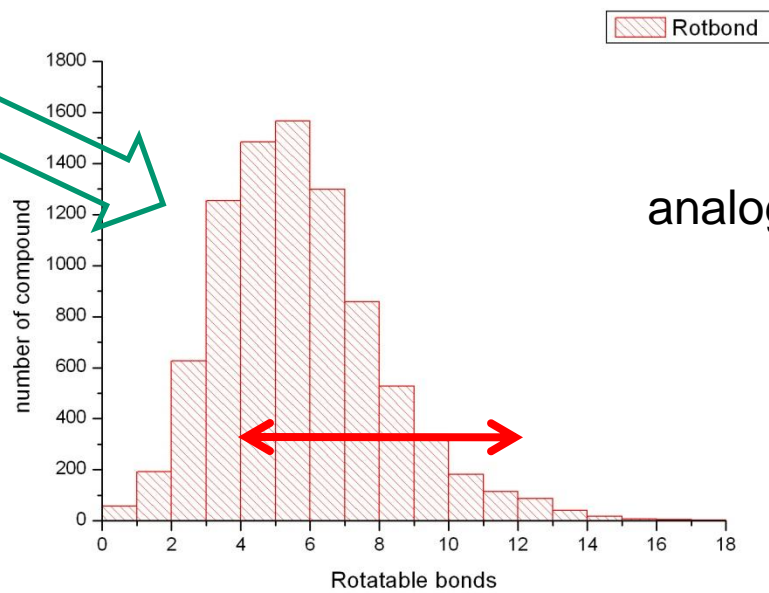
analogue set

May 26, 2010

Rotatable bonds



seeds

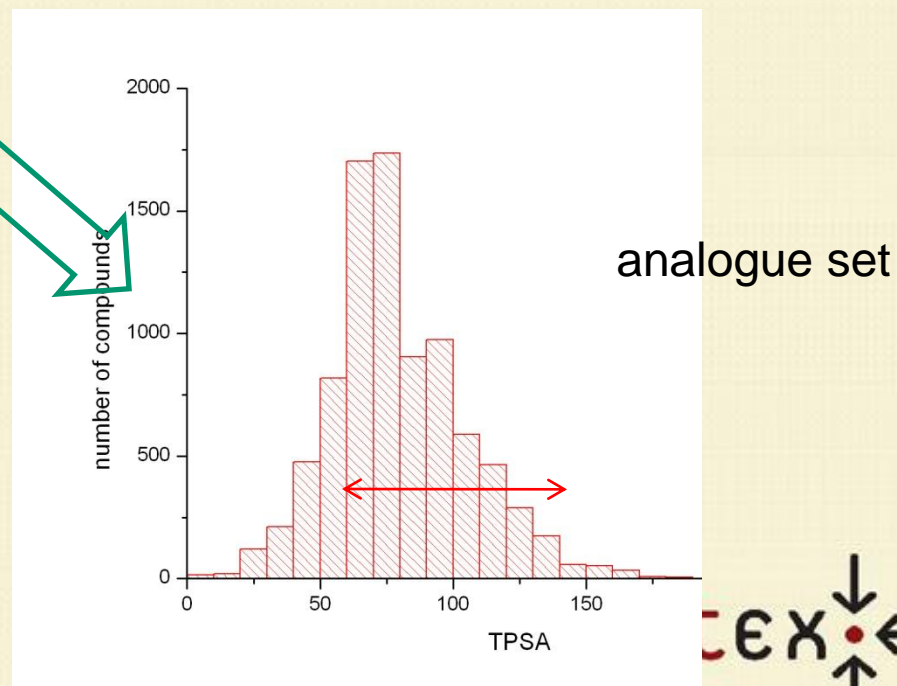
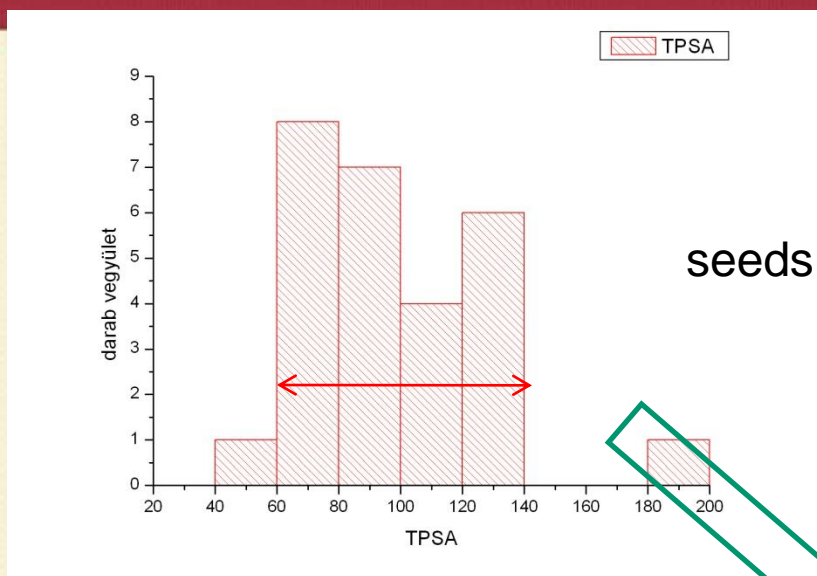


analogue set

May 26, 2010



Topological polar surface area



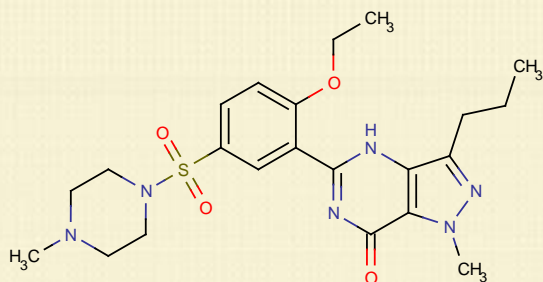
May 26, 2010

Results and further reduction

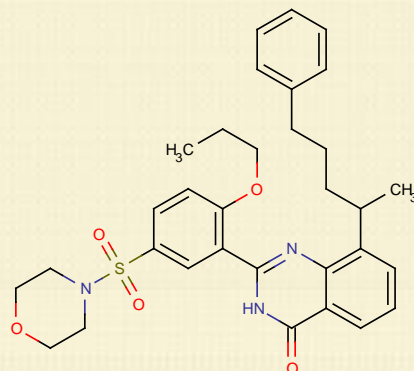
- Similarity search results: 8655
- After property filtering: 2009
- 2. There are overrepresented seeds thus virtual hits coming from those seeds can be reduced
- When combining the similarity search the contribution of the seeds can be controlled
- (or set the number of analogues derived from certain seeds)

2. Overrepresented seeds

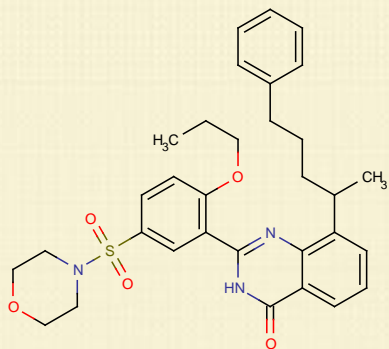
Seeds leading to highest number of similar hits



#4 (Sildenafil)
238 analogues
(60 % similarity or above)



#13
328 analogues
(60 % similarity or above)

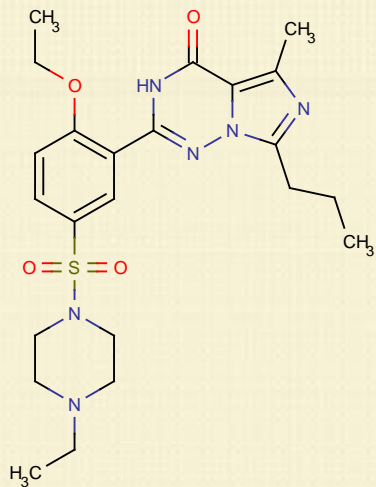


#18 (desantafil)
4494 analogues
(60- 80 % similarity)

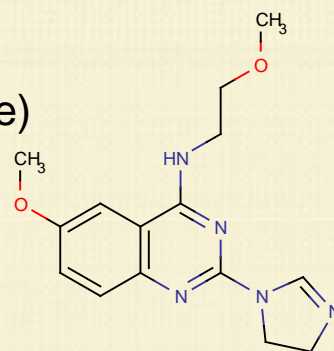
May 26, 2010

2. Overrepresented seeds

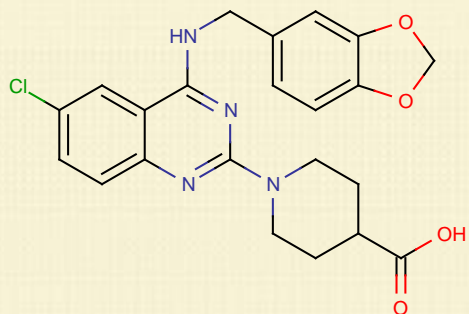
Seeds leading to highest number of similar hits



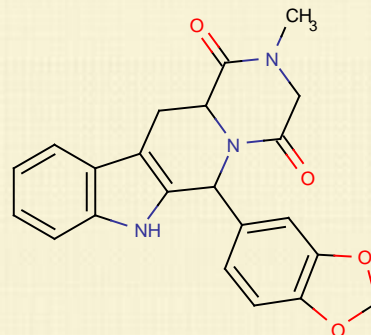
#27
237 analogues
(60 % similarity or above)



#28
272 analogues
(60 % similarity or above)



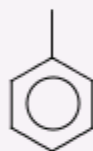
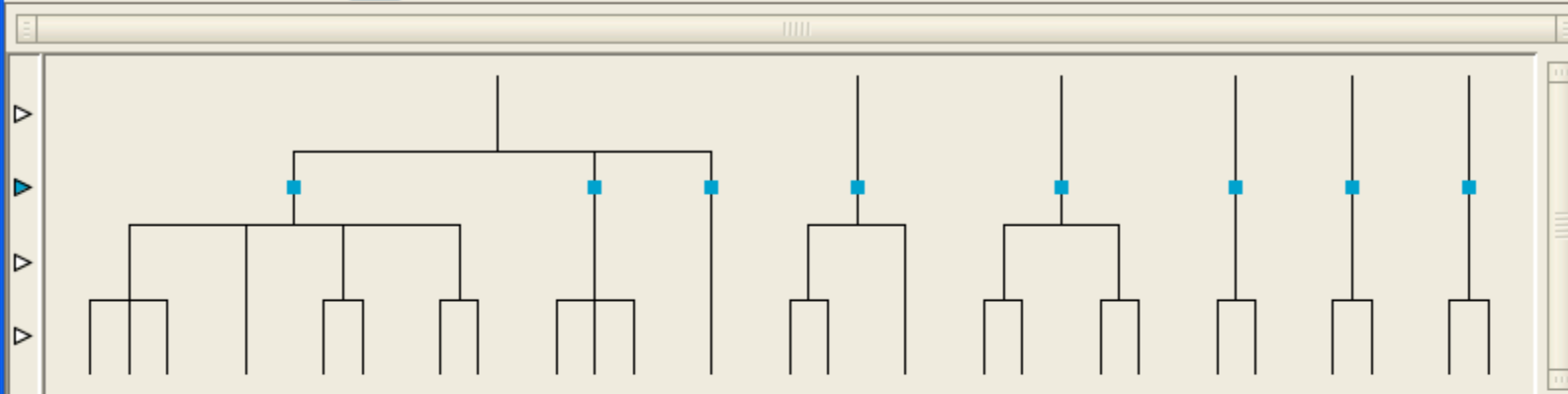
#30
466 analogues
(60 % similarity or above)



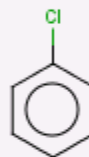
#44
2726 analogues
(60 % similarity or above)



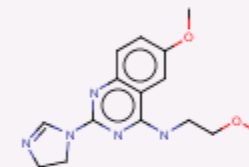
Recurring structural motifs in the seed structures



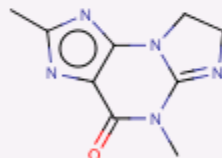
ClusterID: 40
Mol Weight: 270,29, 575,72, 397,94
Cld: 3, 23, 12



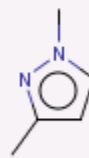
ClusterID: 42
Mol Weight: 423,9, 449,73, 438,17
Cld: 4, 14, 8



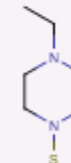
ClusterID: 45
Mol Weight: 301,34, 301,34, 301,34
Cld: 2, 2, 2



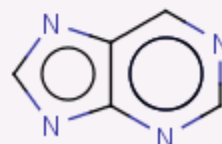
ClusterID: 38
Mol Weight: 315,41, 389,37, 340,74
Cld: 7, 17, 10,67



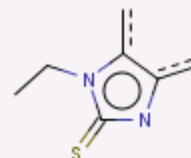
ClusterID: 39
Mol Weight: 353,42, 516,66, 434,02
Cld: 10, 22, 16,25



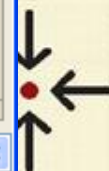
ClusterID: 41
Mol Weight: 488,6, 519,66, 504,13
Cld: 1, 21, 11



ClusterID: 43
Mol Weight: 522,39, 692,9, 607,65
Cld: 19, 25, 22

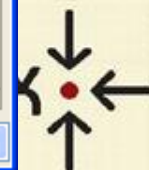
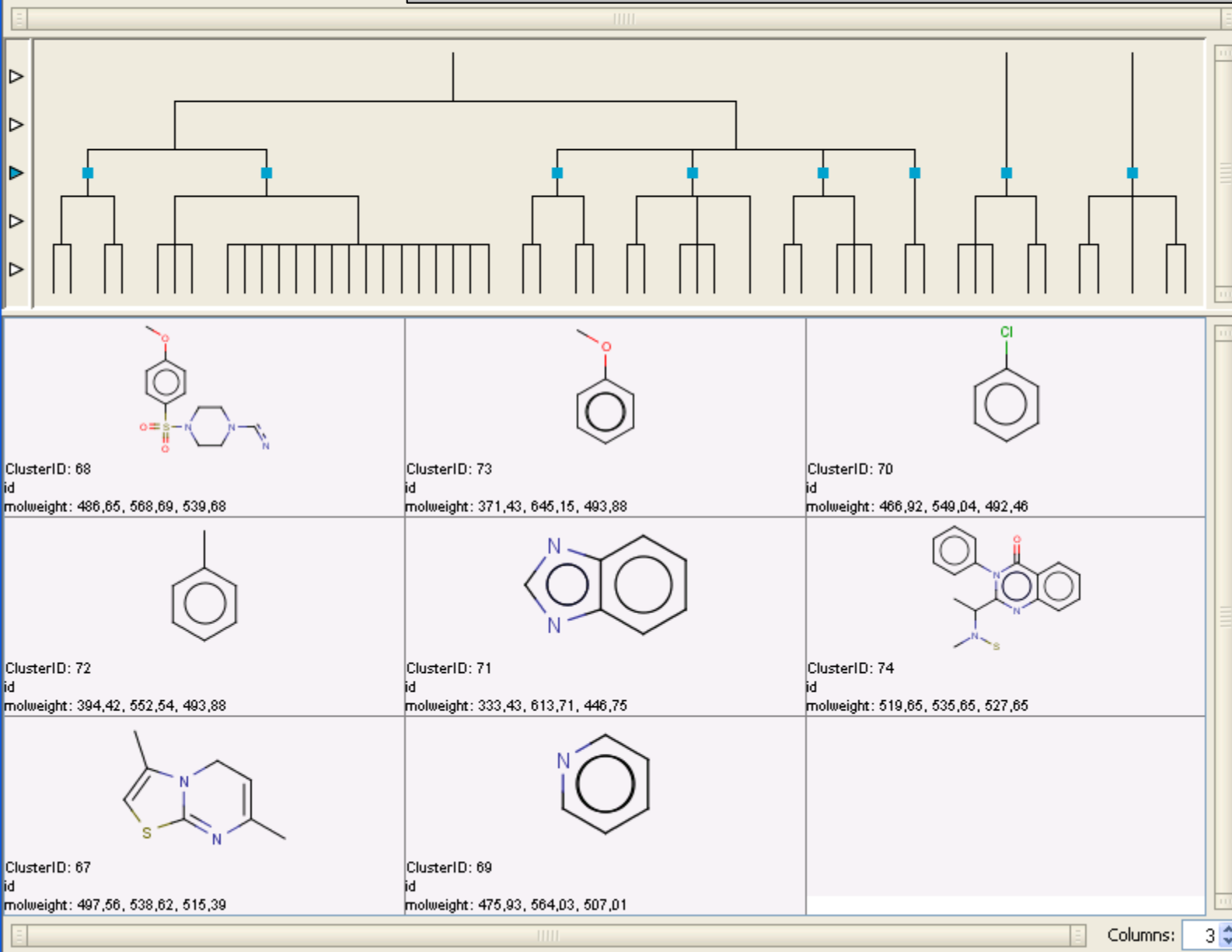


ClusterID: 44
Mol Weight: 430,61, 448,58, 439,6
Cld: 16, 24, 20





Recurring structural motifs in the similarity search results



3. Applying an optimal distribution of the resulting chemotypes

Proposed application of Jklustor/LibMCS

- Taking into consideration of the substructure where the maximum number of connection (bond) is found
 - it can be an option
 - Maybe difficult to define
- Using such option the „real” core structure can be found easier

Conclusion/ Proposals

Combining the elements of the ChemAxon softwares into a dedicated *in silico* tool can be devised which could help the screening companies that rely on compound purchasing

Specific features required (proposal)

- Similarity search
 - Multiple compound search since normally there are more seeds than DBs, all the seed compounds can be searched on one DB while automatically removing the duplications (multiplications).
 - Possibility to set the minimum number of hits and/or the similarity treshold
 - Taking into consideration of the seed representation
- Automated seed property space determination and direct application in library filtering
- Clustering the results into chemotypes/ diversity selection



TargetEx team



Sándor Cseh Zsolt Lőrincz
Tünde Tömöri

This work was partially supported by the National Development Agency (NFÜ) Grant: #KMOP 1.1.1-07/1-2008-0029)

For further info: Thank you for your attention!
dorman@targetex.com

