Linking chemical and biological similarity of small molecules

Pekka Tiikkainen

Euro-QSAR 2008, Uppsala

22.9.2008

TURUN YLIOPISTO
UNIVERSITY OF TURKU

VTT

Study goals

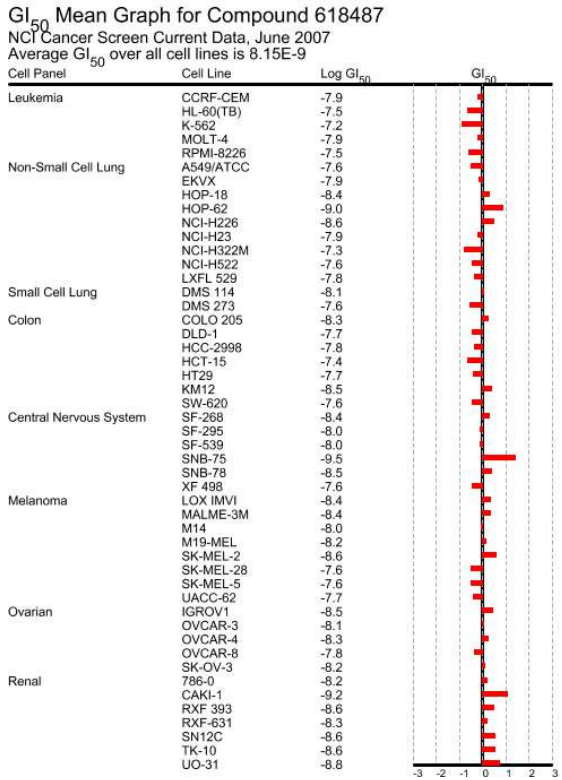
1. Study the relationship between chemical and biological similarity of small molecules
2. Identify biologically similar compound pairs found chemically similar by different ligand-based virtual screening methods
3. Study how combining ligand-based VS methods affects results
4. Apply the achieved results to an external data set

Materials and methods: biological similarity

Over the years, NCI DTP has screened tens of thousands of compounds against the panel.

The most commonly used readout of the screens is GI50. That is, the compound concentration where the cell growth is halved.

In this talk, biological similarity of two small molecule is defined as the **Pearson correlation of their cytotoxicity profiles**.

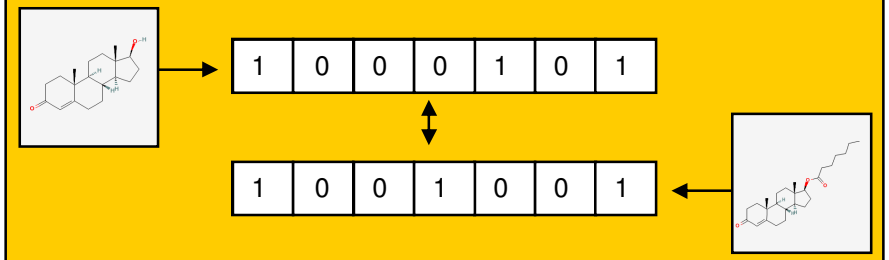


Cytotoxicity profile
of staurosporine

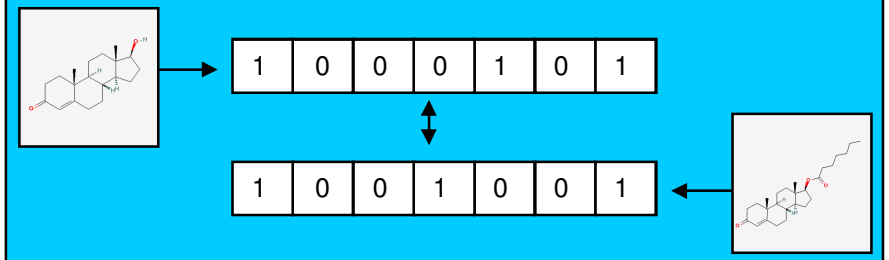
Materials and methods: chemical similarity

Chemical similarity between NCI compound pairs was calculated with four different methods:

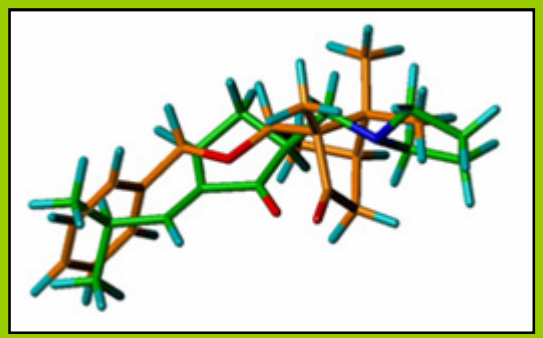
Daylight fingerprints



Unity fingerprints

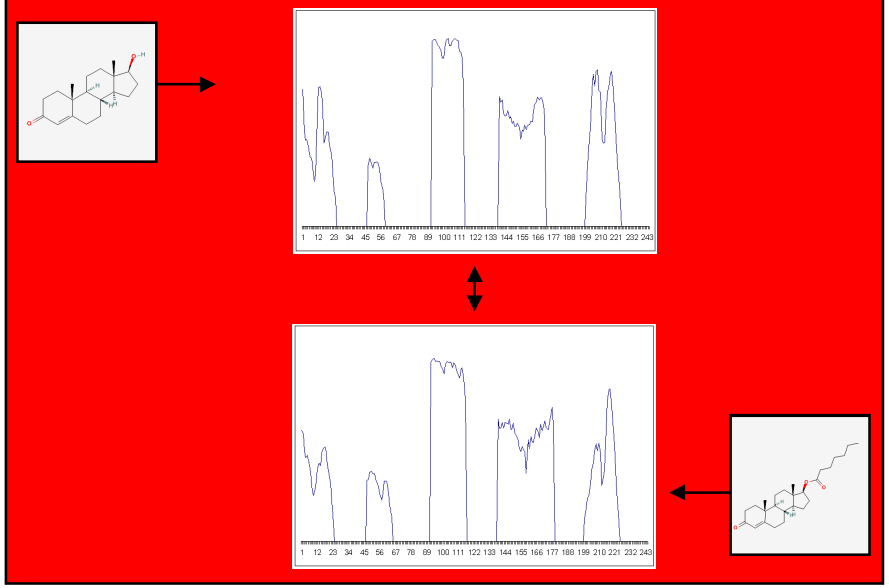


BRUTUS (3D overlay method)



Total score = steric + hbond + vdw + hydro

GRIND descriptors

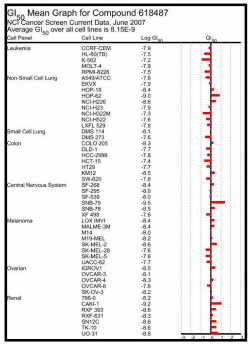


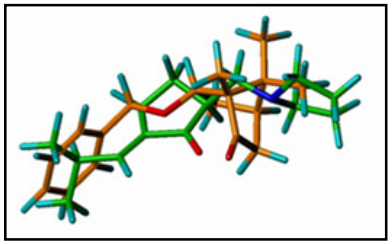
Combining chemical and biological data

After the biological and chemical similarity scores were calculated, the two metrics had to related to each other.

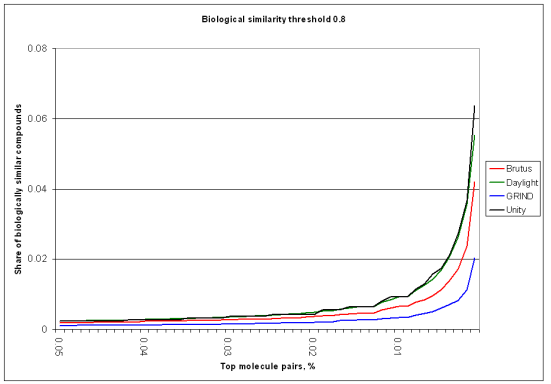
Compound pairs were summed into bins by their chemical similarity. Next the share of biologically similar pairs was calculated for each bin.

Chemical similarity bins were also generated for combinations of similarity metrics.



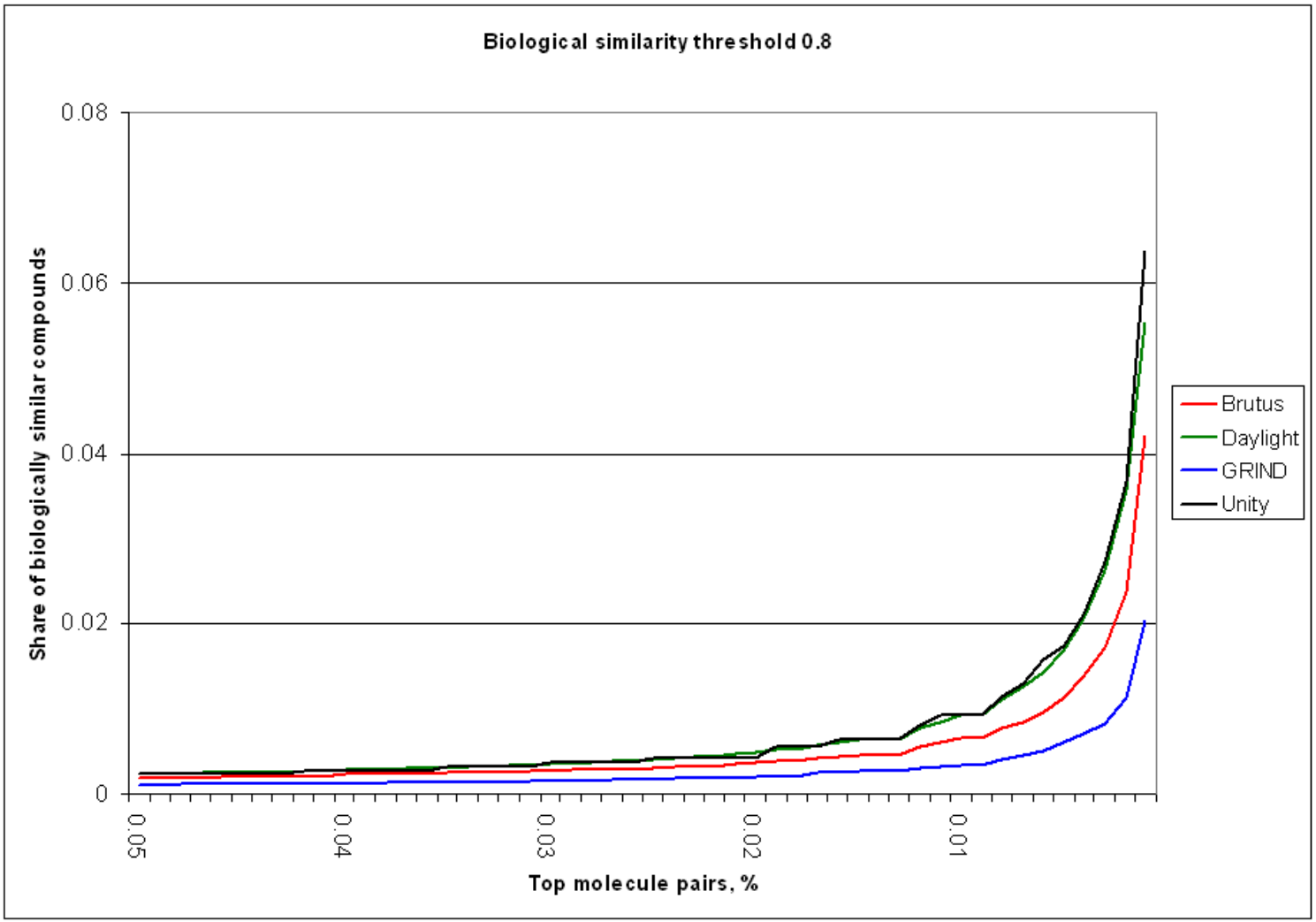


Rank pairs by chemical similarity



Relating chemical similarity to biological similarity: Individual methods

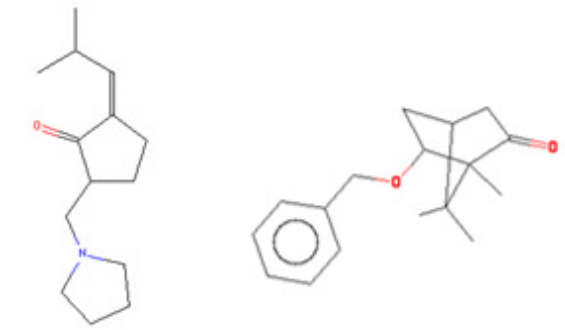
When the very top chemical similarity bins are considered, we see a clear enrichment of biologically similar pairs.



Scaffold hopping example

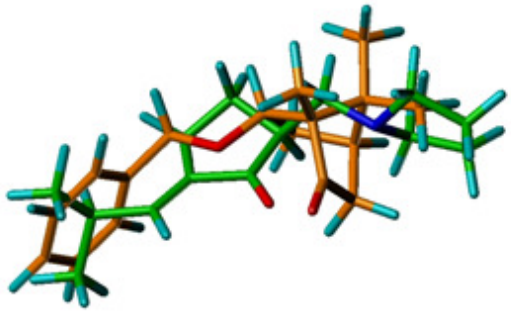
BRUTUS = 2.960, GRIND = 0.915, Daylight = 0.227 and Unity = 0.25

a)

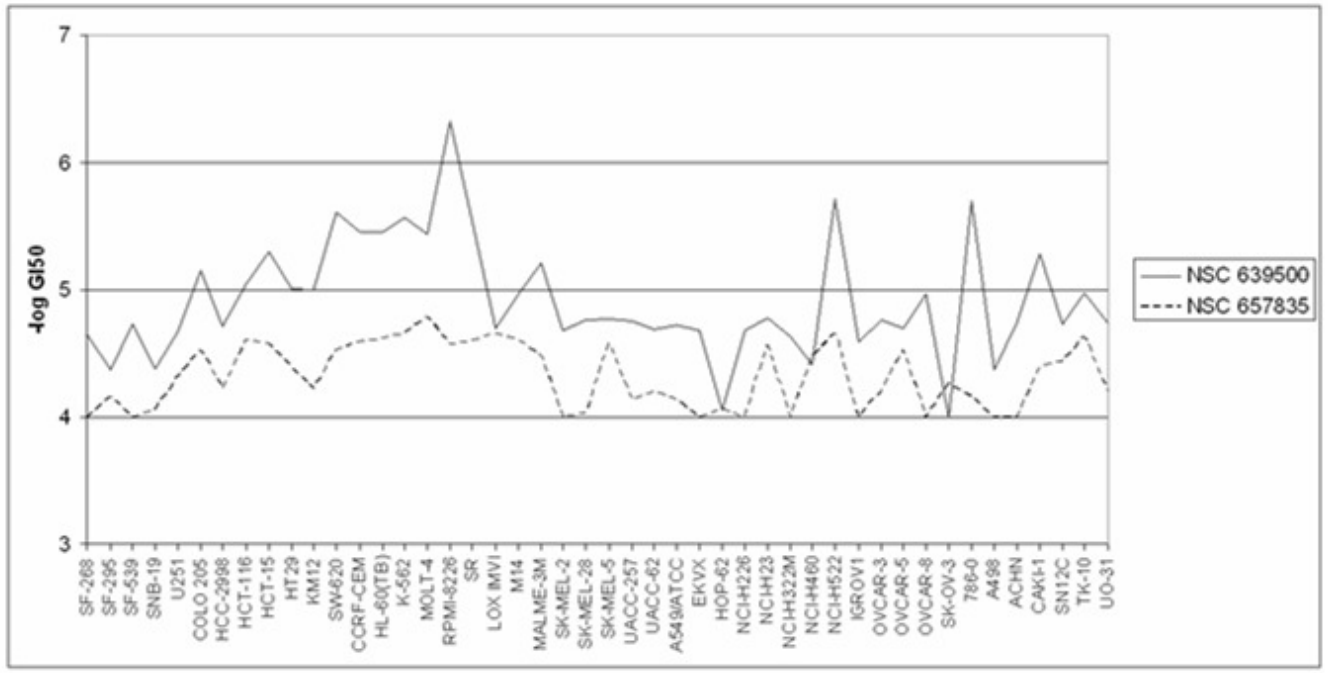


NSC 639500 NSC 657835

b)

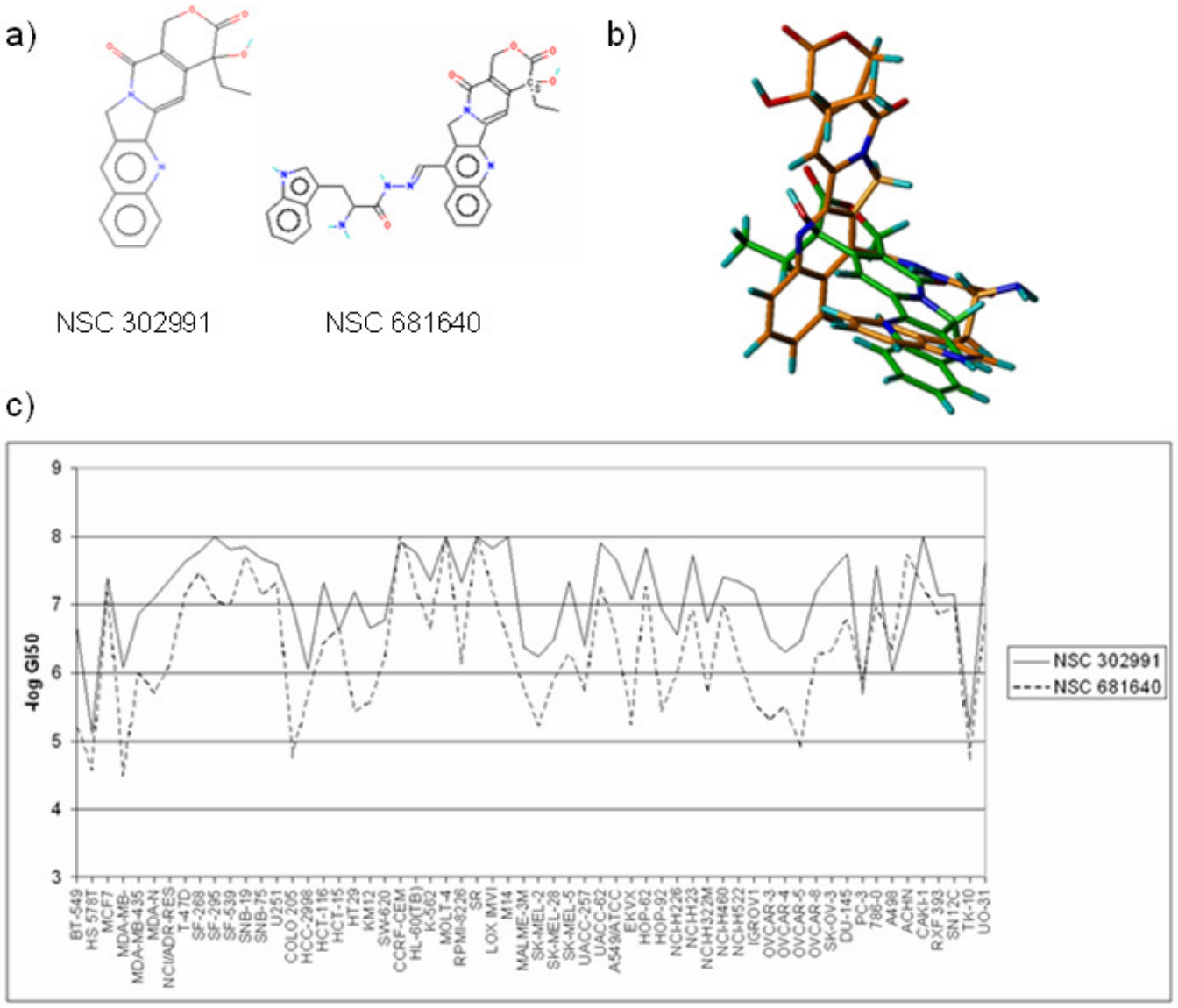


c)

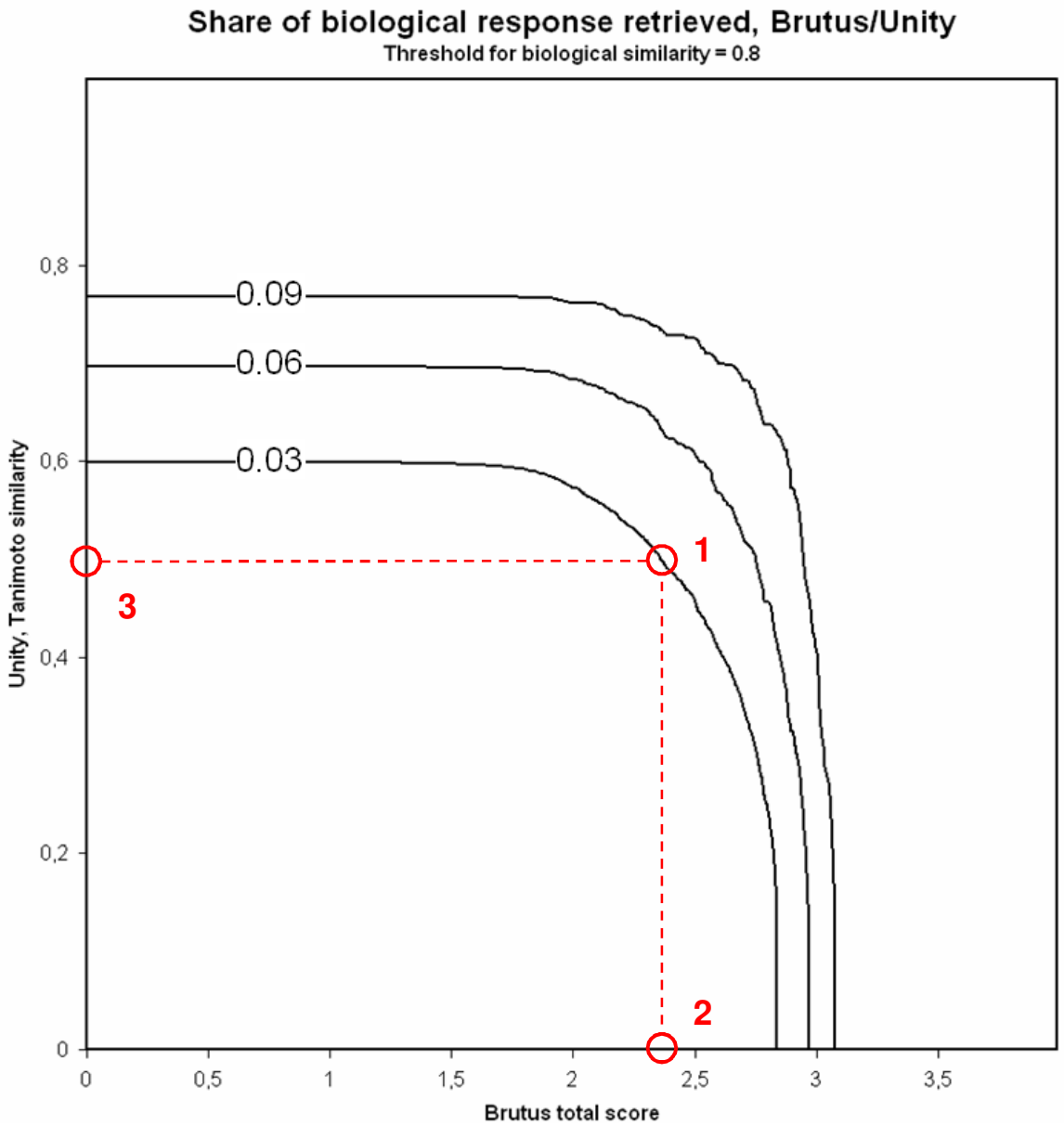


When 3D methods fail (while fingerprints do not)

Brutus 1.923, GRIND 0.845, Unity 0.81 and Daylight 0.999



Relating chemical similarity to biological similarity: combining methods



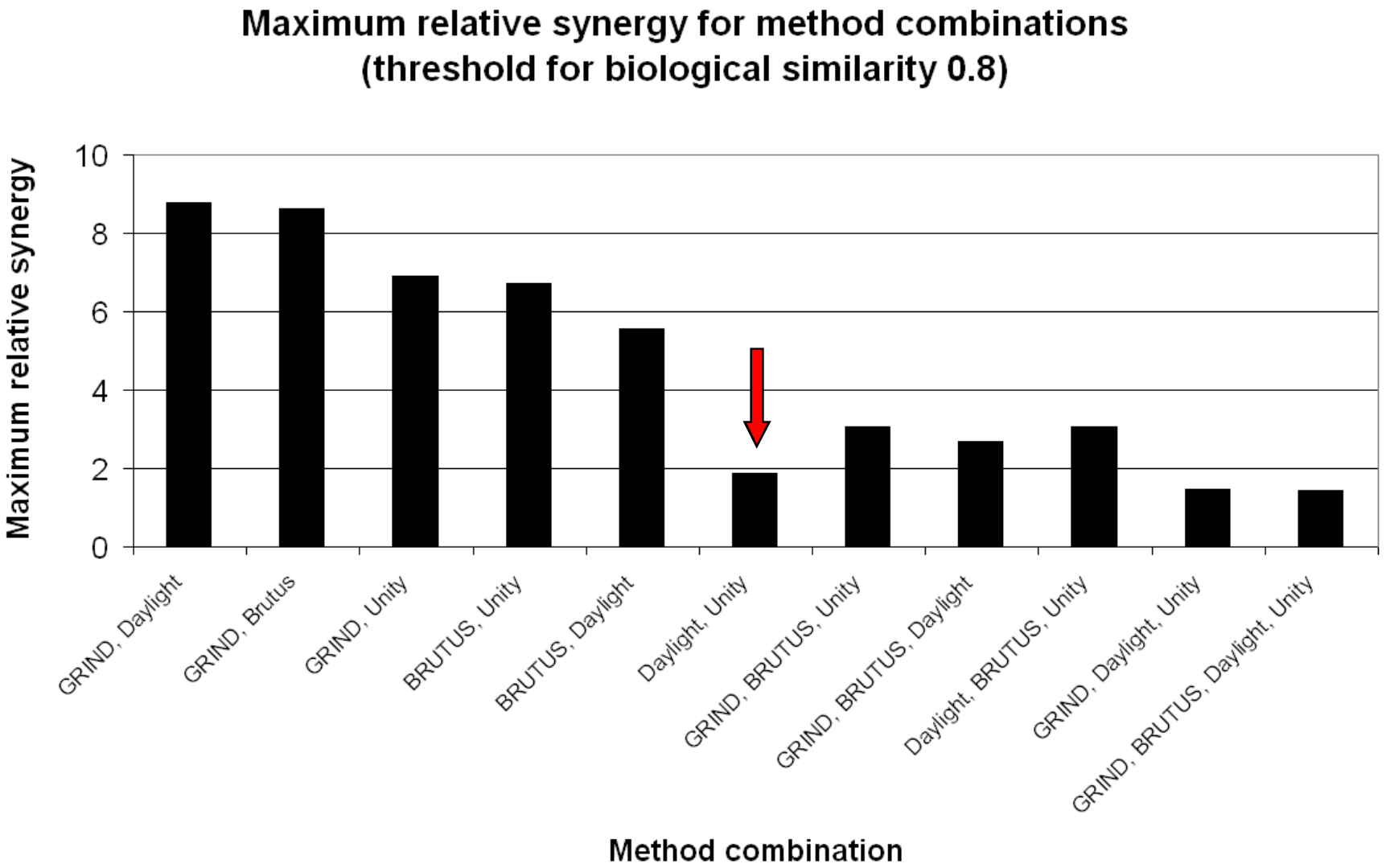
Point 1:
Brutus = 2.36, Unity = 0.50
Biol. similarity = 0.030

Point 2:
Brutus = 2.36, Unity = 0
Biol. similarity = 0.002

Point 3:
Brutus = 0, Unity = 0.50
Biol. similarity = 0.011

Relative synergy:
0.030 / max(0.002; 0.011) = 2.7

Synergy achieved vs. number of methods combined



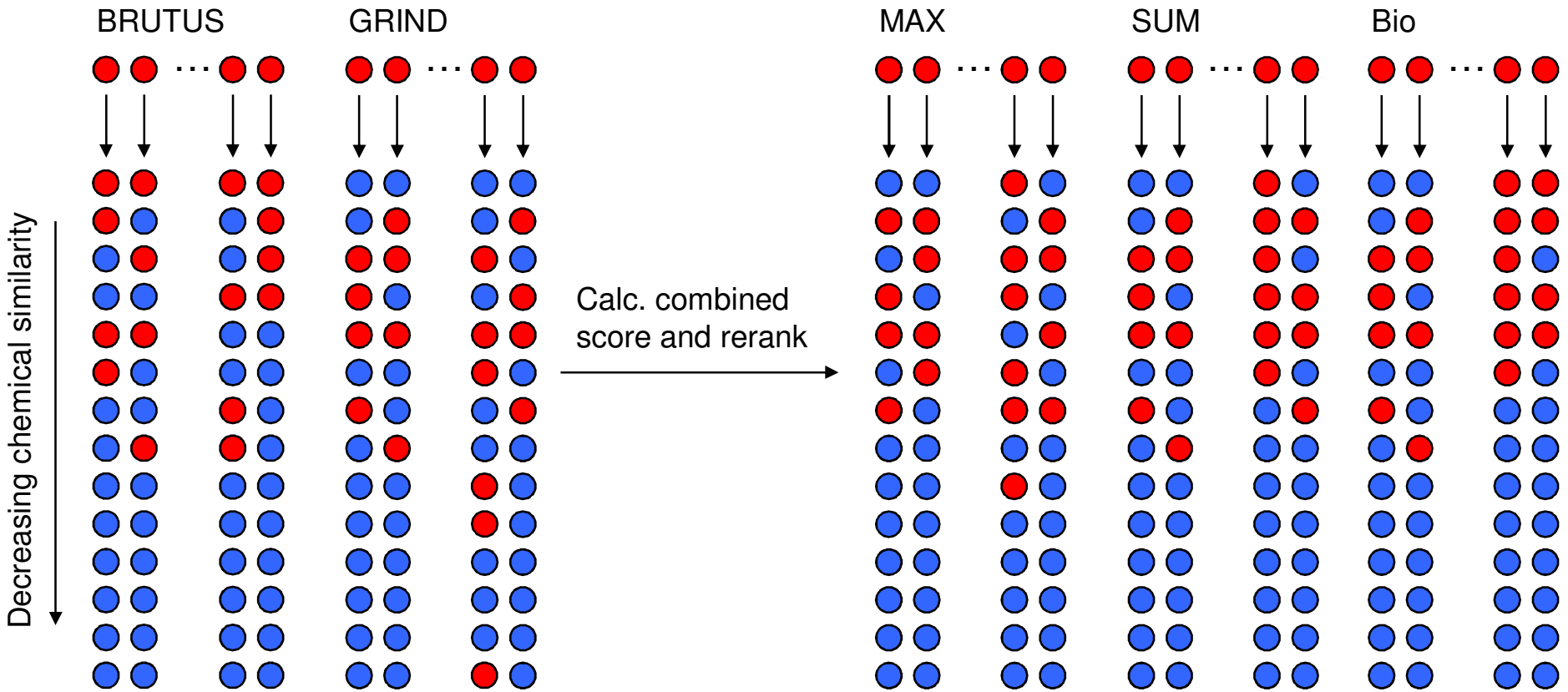
Putting the data into good use

- Logical implementation for the results described above is their application to virtual screening.
- Repeating the same virtual screen with two methods and replacing the two scores with an estimate of biological similarity should lead to an improved retrieval of active molecules.
- We chose DUD (Directory of Useful Decoys) compound set for this purpose. The dataset contains known ligands and decoy compounds divided into 40 target groups.

http://dud.docking.org/r2/

Putting the data into good use

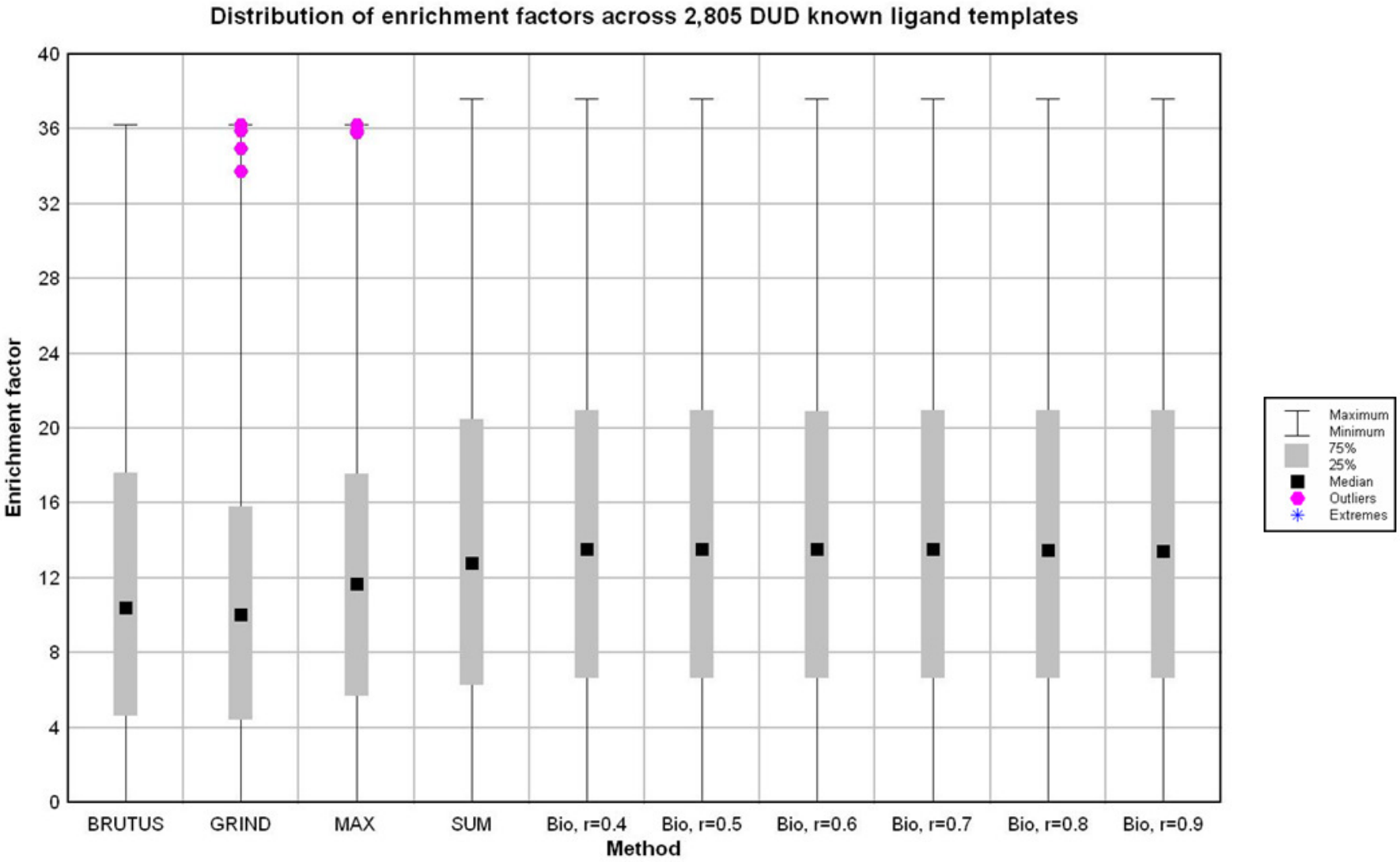
First the chemical similarity of each known ligand (2,805) was calculated to all the molecules in its target set with BRUTUS and GRIND. Next, the results were merged with three data fusion methods…



…and once all the lists were ranked, enrichment of known ligands was measured with a set of metrics.

Putting the data into good use

Enrichment of known ligands at top 1 % of the ranked lists shows that all data fusion methods are superior to both individual methods.



Conclusions

- As a result of the current work, we have an objective method to link chemical and biological similarity
- Combining methods creates synergy
  - gain in synergy drops as more methods are combined
  - similar methods are less synergistic than those with more different basis
- The estimated biological similarity of compounds can be applied to virtual screening where it is superior or equal to more simple data fusion methods
- Recently, another article* using a different dataset came to very much the same conclusions

* Muchmore et al. Application of Belief Theory to Similarity Data Fusion for Use in Analog Searching and Lead Hopping. *J. Chem. Inf. Model.* **2008**.

Acknowledgements

Olli Kallioniemi,
University of Turku, VTT

Antti Poso, University of Kuopio

Center for Scientific
Computing (CSC)