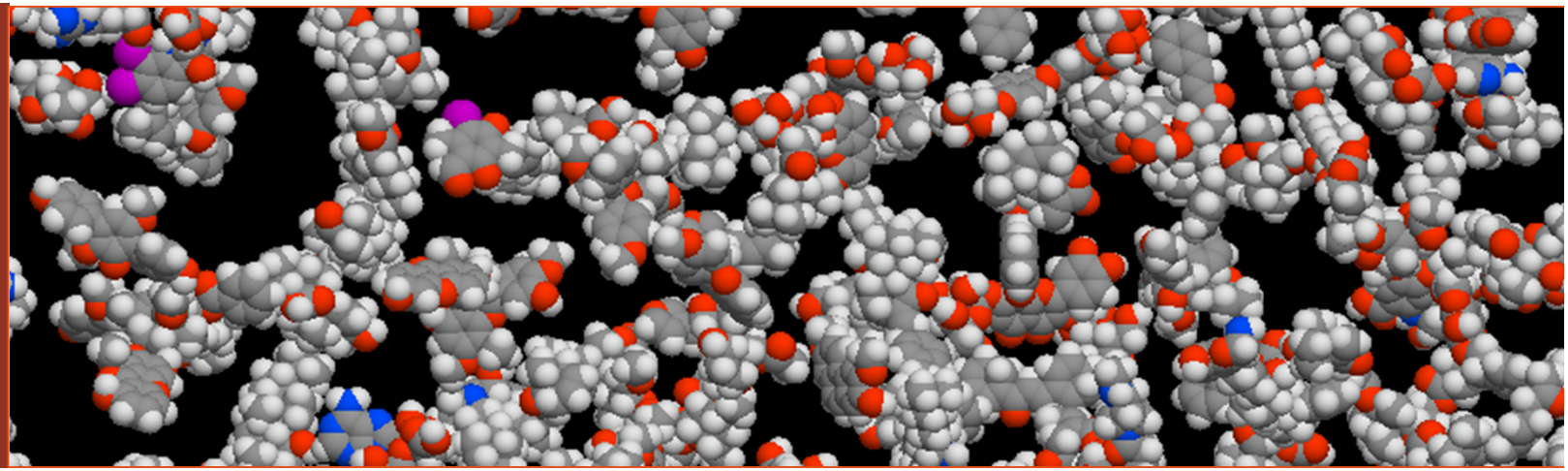


Natural product-likeness score with RDKit

Peter Ertl

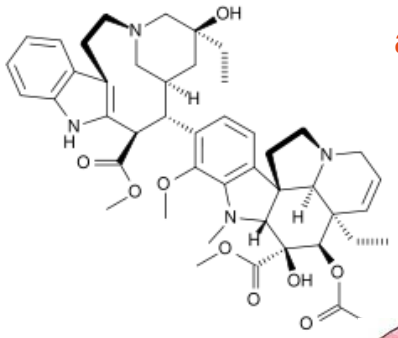
Novartis Institutes for Biomedical Research, Basel, CH

RDKit UGM, September 2015

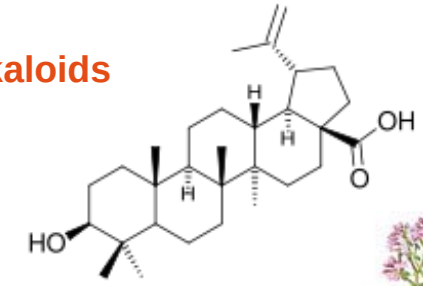
http://peter-ertl.com

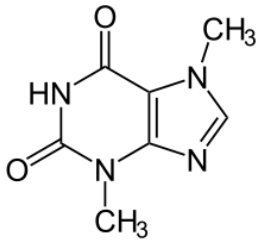
NOVARTIS

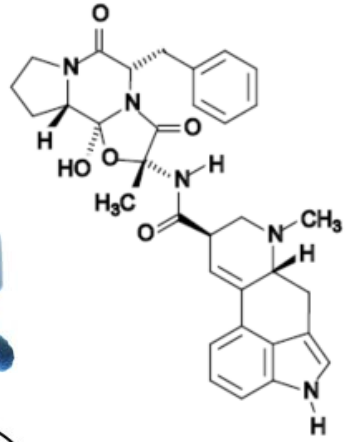
Natural Products 101



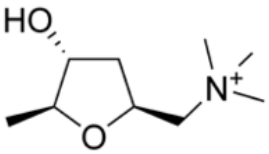
alkaloids







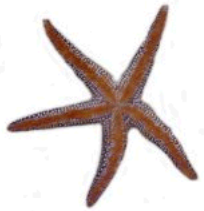
plants



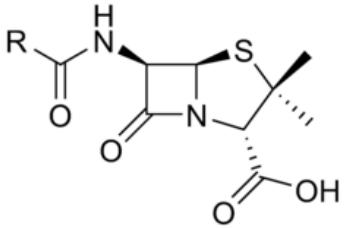
fungi

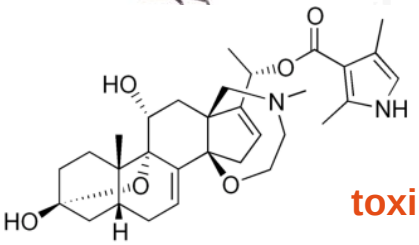
bacteria

(marine) animals

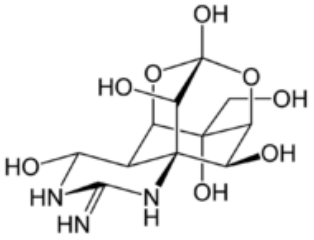


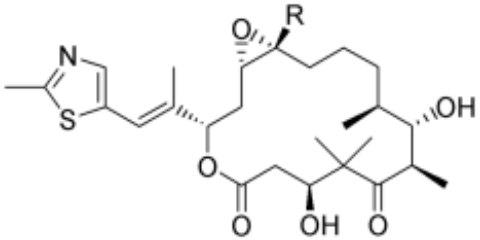
antibiotics





toxins





NOVARTIS

Natural products as a source of new drugs

- natural products (NPs) have been optimized in a very long natural selection process for optimal interaction with biological macromolecules.
- NP molecules are therefore an excellent source of substructures which may be used in the design of new bioactive molecules.
- large portion of drugs on the market are NPs or are of NP origin

Natural product-likeness

- To facilitate application of NP-like chemistry in the drug discovery process one needs to compare somehow the characteristics of studied molecules with those of NPs.
- Natural product-likeness is a measure of similarity to NP molecules (in analogy with well established drug-likeness)
- Calculation of NP-likeness may be based on various molecular characteristics: molecular properties, substructures, 3D pharmacophores ...
- After several tests we decided to use fragment contribution scheme

Calculation of NP-likeness

Step 1

Using a training set of NPs and “average drug-like molecules” build a database of fragments (atom-centered fragments) with their NP-contributions

A fragment NP-likeness contributions is calculated as:

$$f_i = \log (nact_i / ninact_i * ninact_{total} / nact_{total})$$

Step 2

NP-likeness of unknown molecule is calculated by summation of contribution for all fragments in this molecule and normalization relative to molecule size.

NP-likeness ranges from about 5 to -5, score above 1 = NP-like

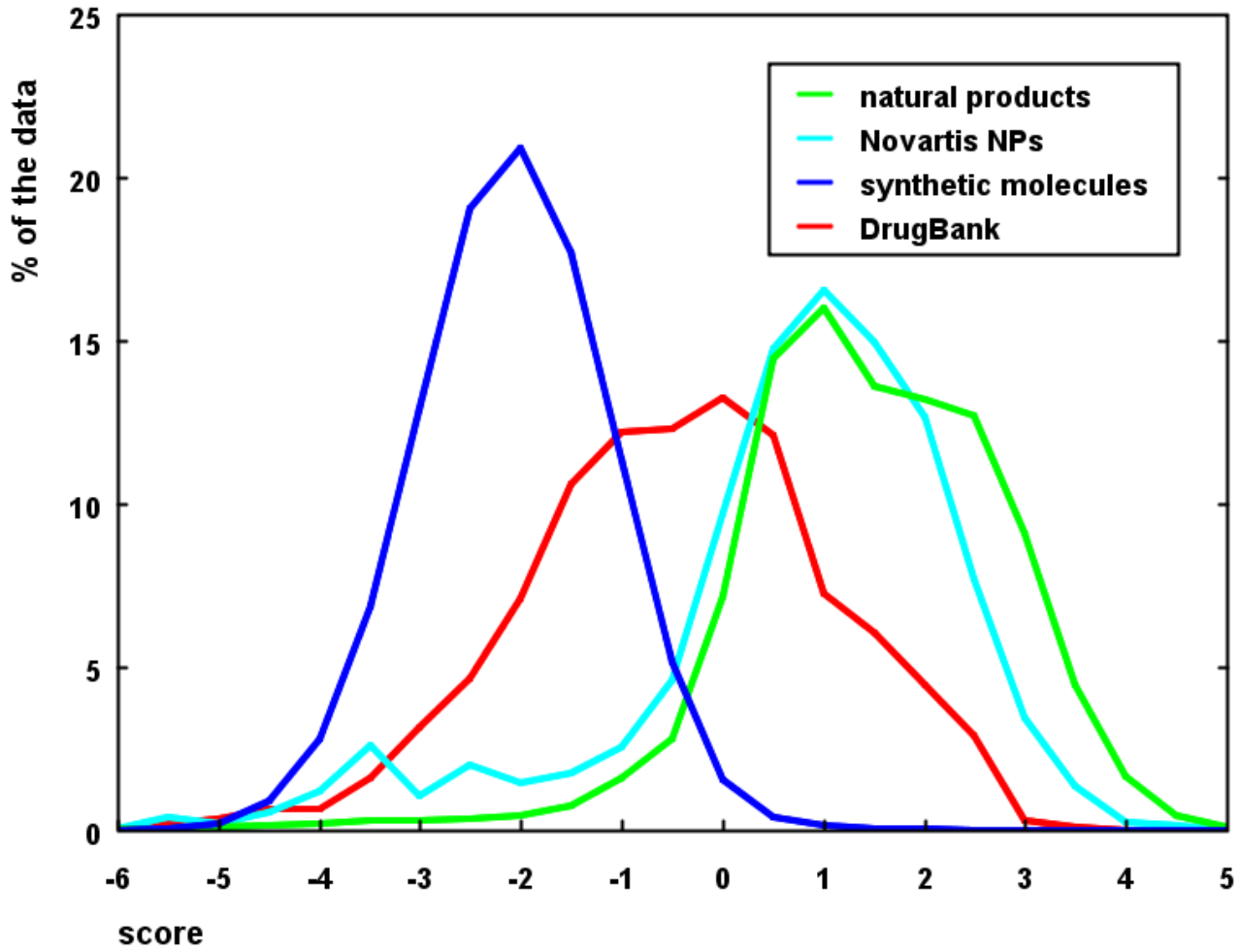
Calculation of NP-likeness

Methodology details are described in:
Natural Product-likeness Score and Its Application for Prioritization of Compound Libraries
Peter Ertl, Silvio Roggo, and Ansgar Schuffenhauer
Journal of Chemical Information and Modeling, 48, 68-74 (2008)

2nd incarnation:
Natural product-likeness score revisited: an open-source, open-data implementation
Kalai Vanii Jayaseelan, Pablo Moreno, Andreas Truszkowski, Peter Ertl and Christoph Steinbeck
BMC Bioinformatics 13:106 (2012)

And finally:
Open source implementation using RDKit – in the contrib directory
trained on ~45,000 NPs from various open sources and ~1M molecules from ZINC as “non-NP-like background”, August 2015

Natural Product-likeness score



NOVARTIS

Application of NP-likeness

Prioritization of molecules, virtual screening

selection of molecules for screening and company archive enhancement

Design of combinatorial libraries

selection of NP-like scaffolds and substituents (BBs) supporting design of NP-like combinatorial libraries; compromise between complexity and NP-likeness

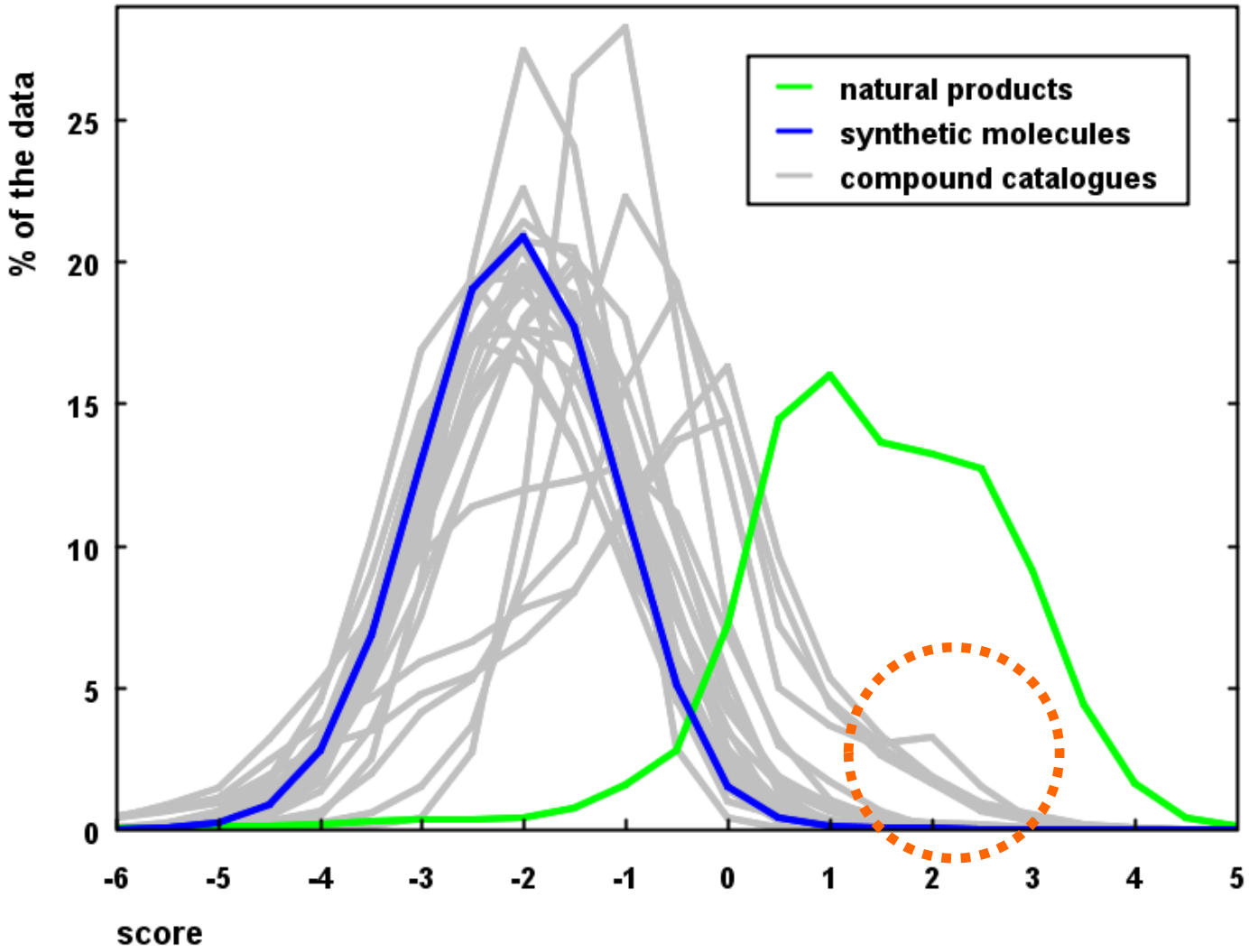
Fragment-based screening

fragments are used as a starting point for design of new bioactive molecules

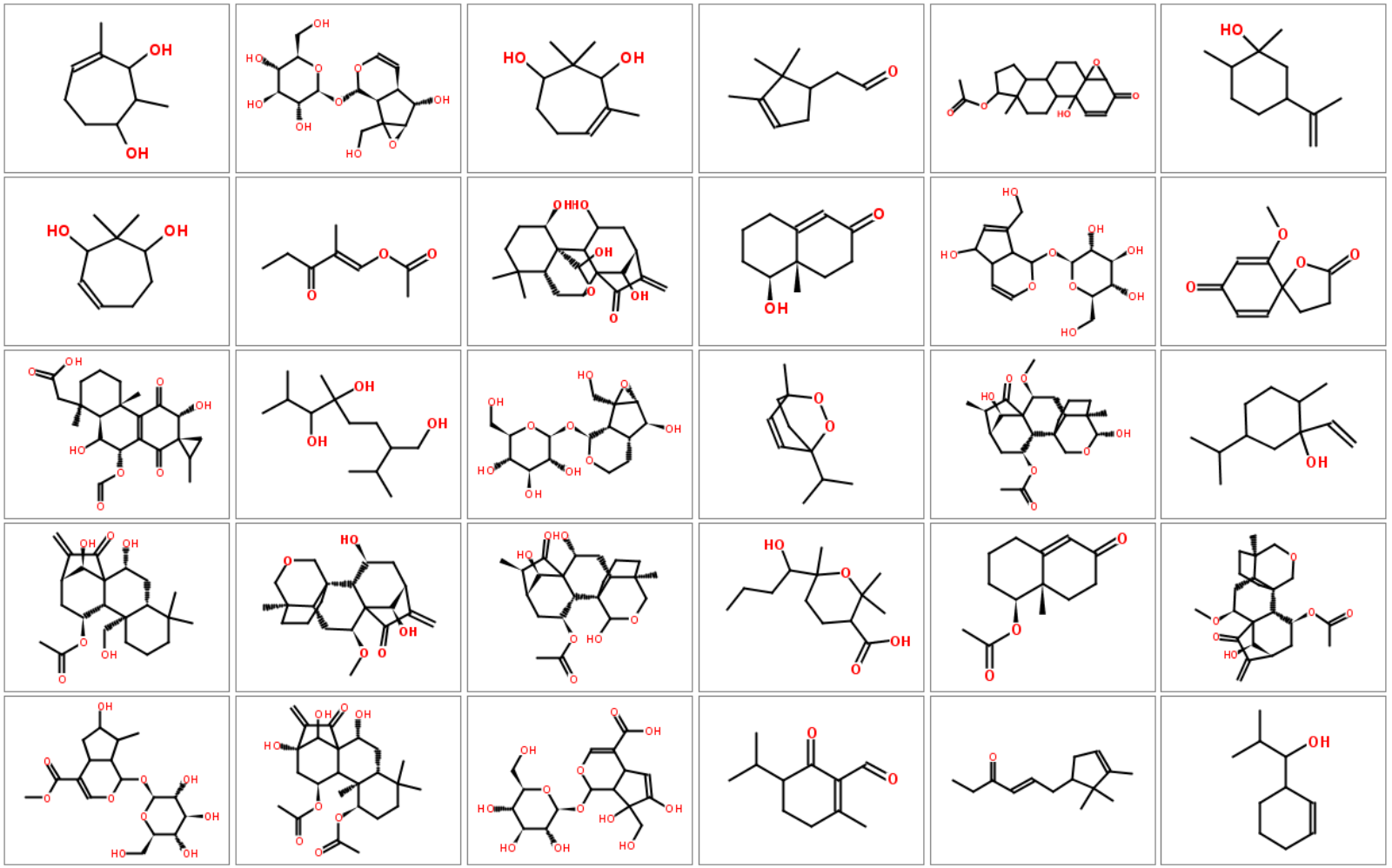
Evolutionary design of molecules with desired properties

procedure for automatic evolutionary design of molecules optimizing at the same time multiple properties

Distribution of NP-likeness

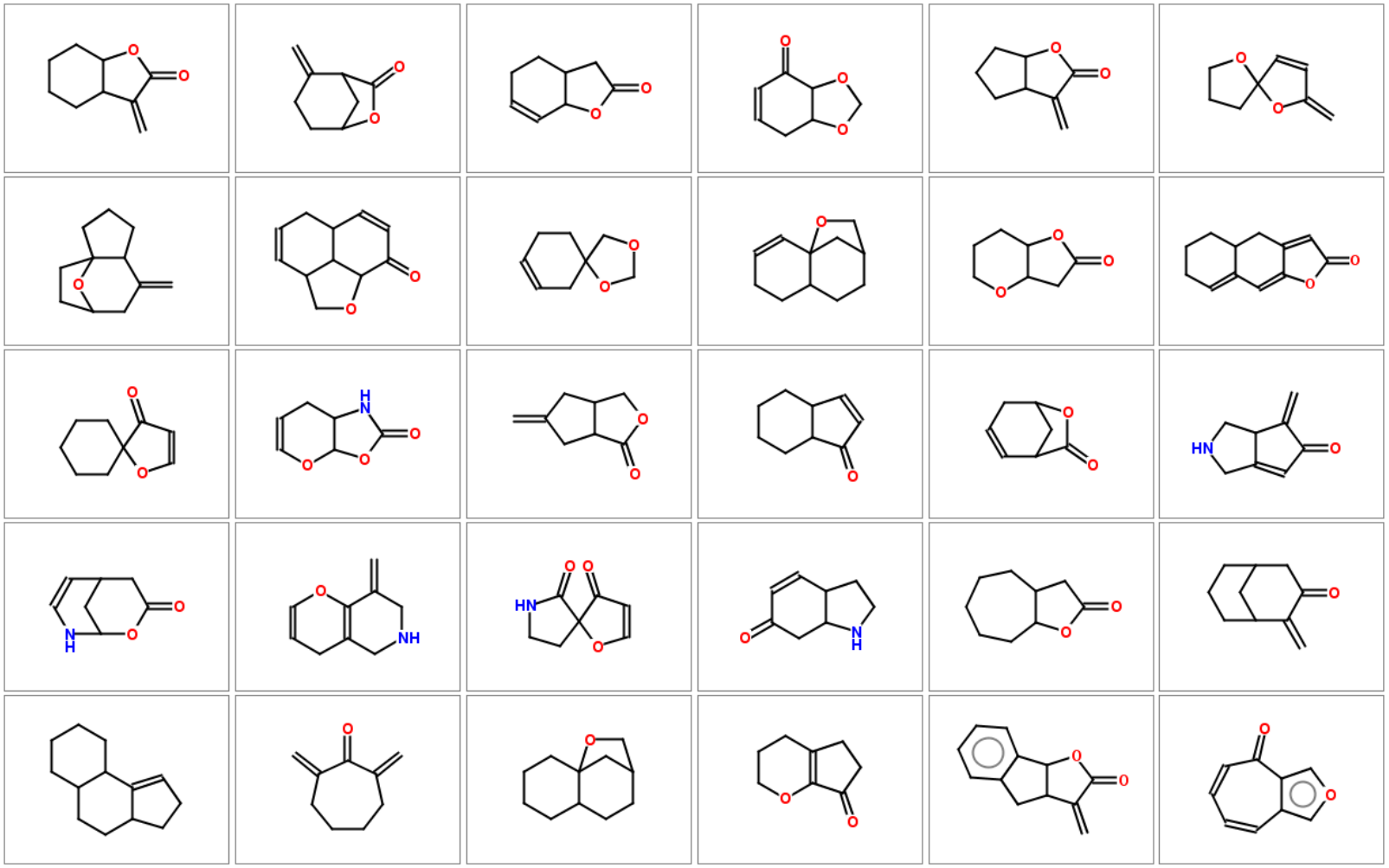


Examples of NPs hits from the MDPI molMall



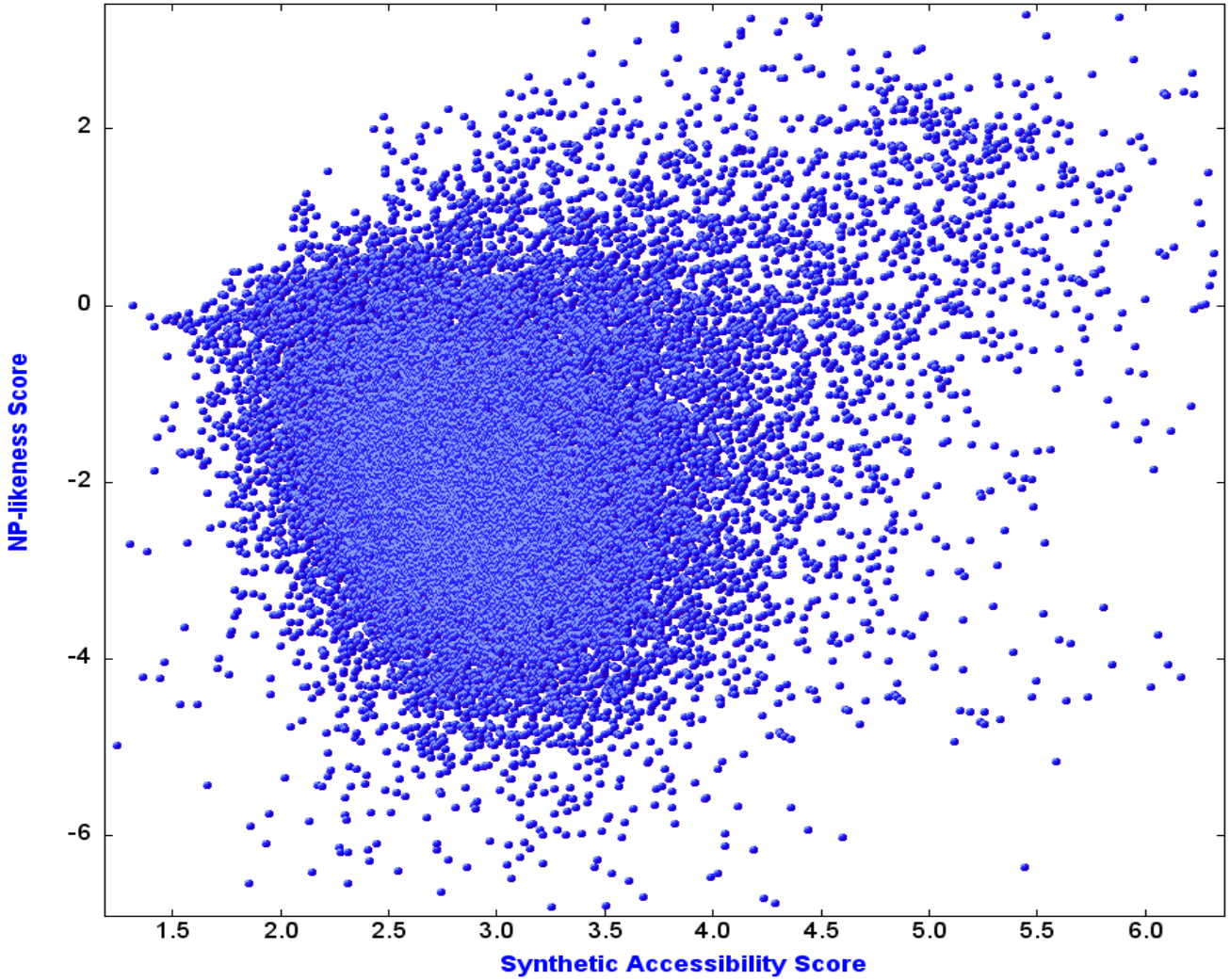
NOVARTIS

Scaffolds with high NP-likeness from PubChem



NOVARTIS

Prioritization of compound libraries

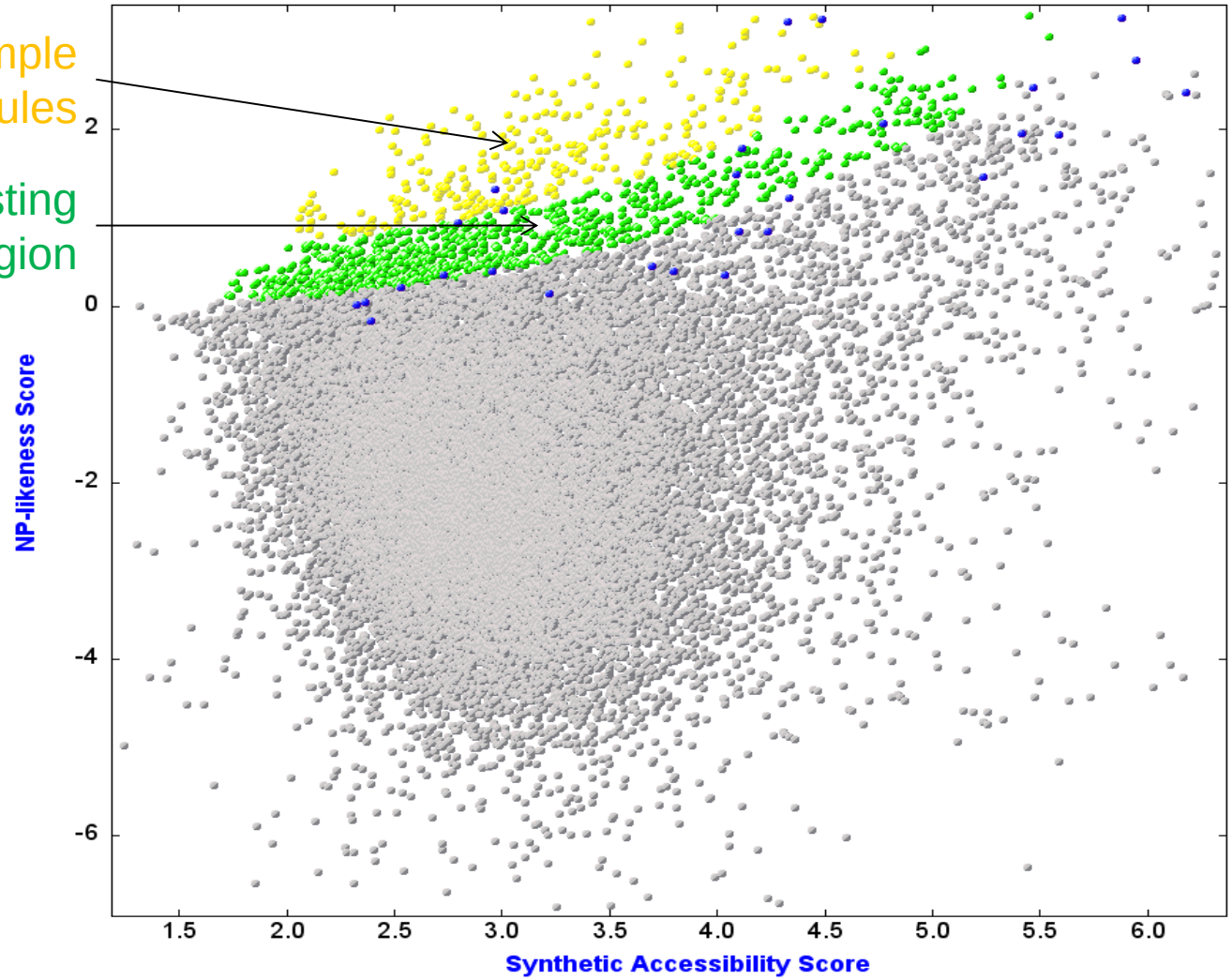


NOVARTIS

Prioritization of compound libraries

too simple molecules

interesting region



NOVARTIS

Summary

- NP-likeness score is a useful measure helping to guide design of new molecules towards interesting regions of chemical space that have been identified as “bioactive” by natural evolution
- the calculation is simple and fast
- results may be used in compound prioritization, virtual screening, purchasing of molecules and library design
- code to calculate the NP-likeness implemented in RDKit based on open data is available in the RDKit contrib directory

Acknowledgement - Data Sources

Thank to Greg Landrum for reviewing (and improving) my code.

Natural products (~45,000 unique structures)

- ChEMBL structures originating from the J.Nat.Prod. https://www.ebi.ac.uk/chembl/
- natural product collections from the ZINC database
  - http://zinc.docking.org/browse/catalogs/natural-products
  - AfroDb, Herbal Ingredients, NPACT, NuBBE, TCM Database@Taiwan, UEFS NPs
- CamMedNP http://www.biomedcentral.com/1472-6882/13/88
- South African NP Database https://sancdb.rubi.ru.ac.za/

Reference non-NP structures

- 1 million representative molecules from the ZINC “drug-like” subset http://zinc.docking.org/