Navigation in Chemical Space Towards Biological Activity

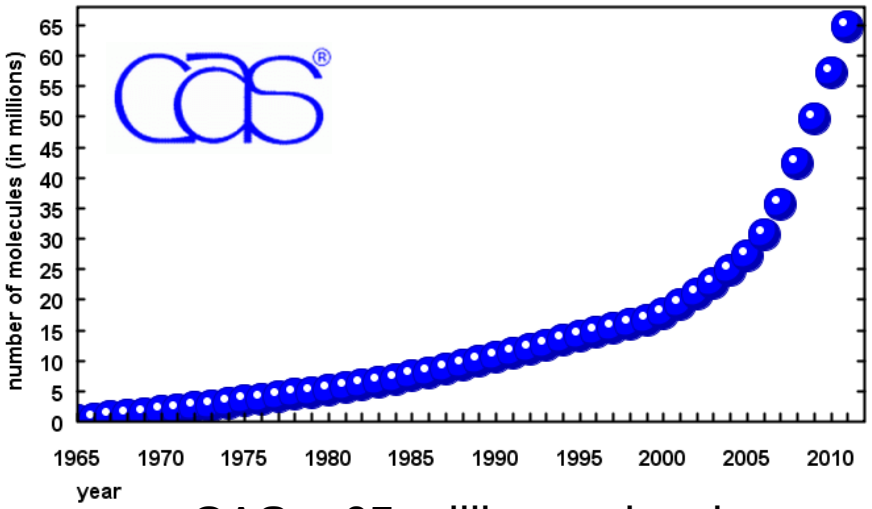
Peter Ertl

Novartis Institutes for BioMedical Research

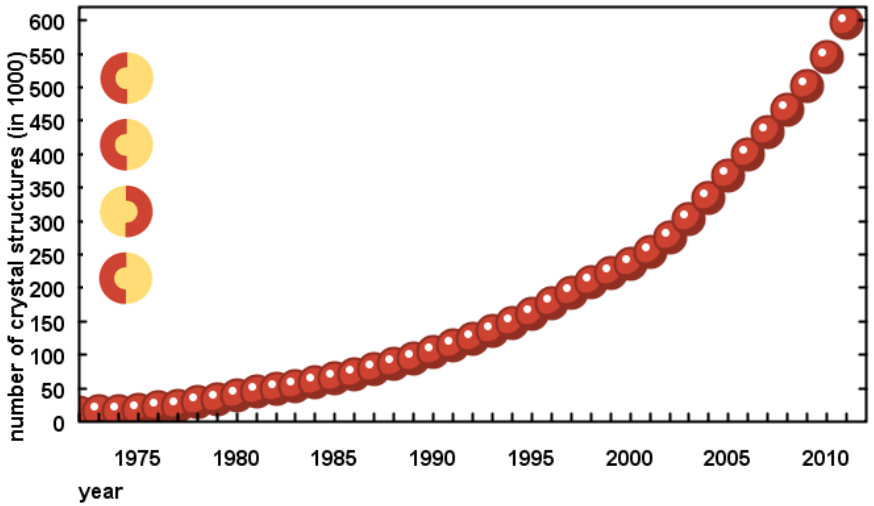
Basel, Switzerland

NOVARTIS

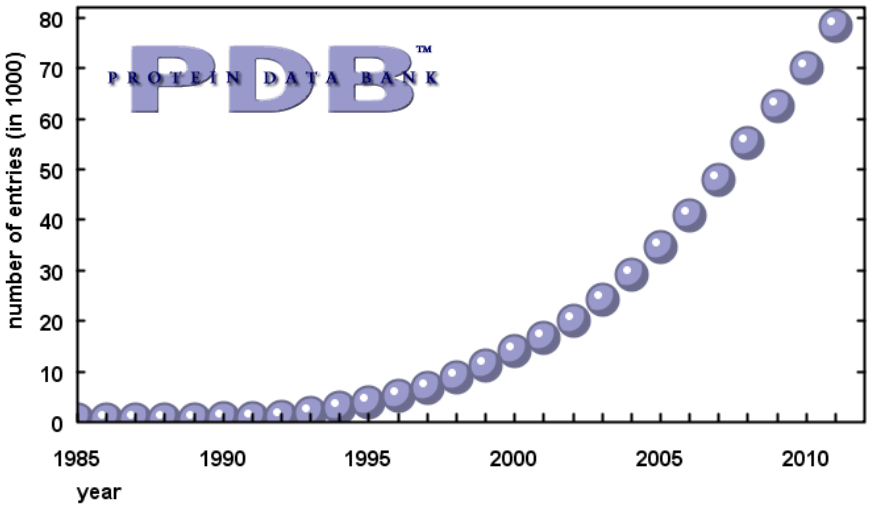
Data Explosion in Chemistry



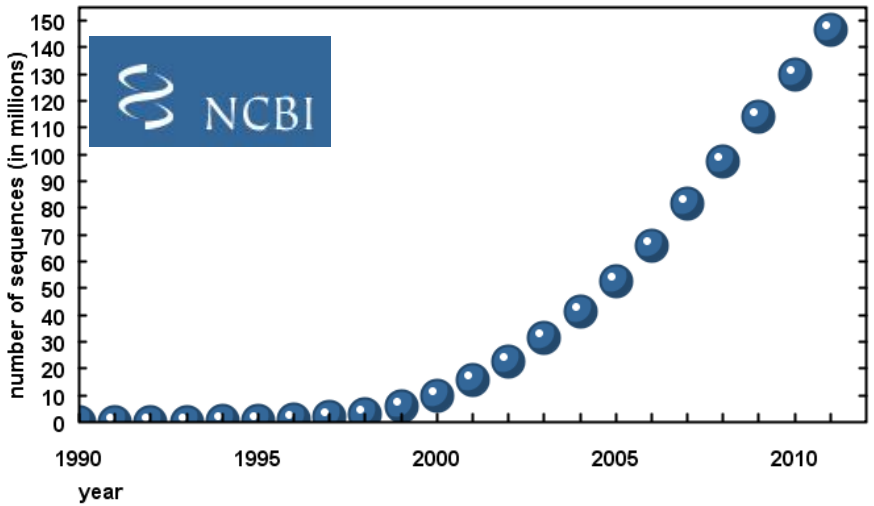
CAS – 65 million molecules



CCDC – 600’000 structures



PDB – 78’000 proteins



GenBank – 145 million sequences

NOVARTIS

Chemical Space is Huge

67 million substances registered in the CAS
33 million compounds in the PubChem database
many millions in archives of pharma / agro companies
many millions available as commercial samples
~1 million molecules with (published) biological activity

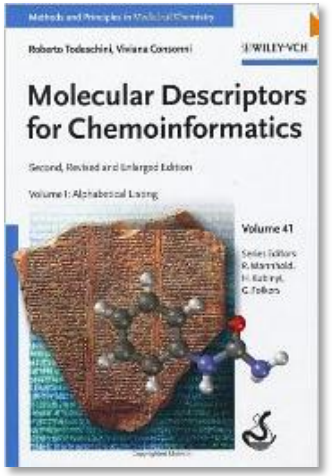
And VERY large number of possible (virtual) molecules.

How to analyze the Chemical Universe?

1. characterization of molecules by proper descriptors
2. dimensionality reduction
3. user friendly visualization

Characterization of Chemical Space

Over 8000 molecular descriptors available:



R. Todeschini, V. Consonni,
Molecular Descriptors for Chemoinformatics, Wiley-VCH, 2009

Descriptors suitable for large-scale analysis:

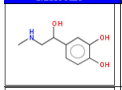
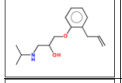
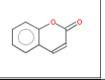
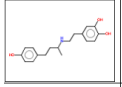
global physicochemical properties = calculated descriptors (logP, PSA, HB donors and acceptors ...)

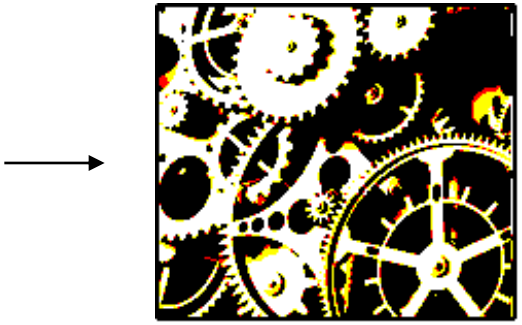
substructure features – fragment counts, structural keys, pharmacophores, fingerprints ...

larger structural features – rings, scaffolds, substituents

NOVARTIS

Handling Complex Data Matrices

Molecule	logP	PSA	natoms	MW	...
	-0.06	72.7	13	183.2	...
	2.58	41.5	18	249.3	...
	2.11	12.9	9	161.6	...
	2.01	20.2	11	146.1	...
	3.31	78.8	30	425.9	...
...					



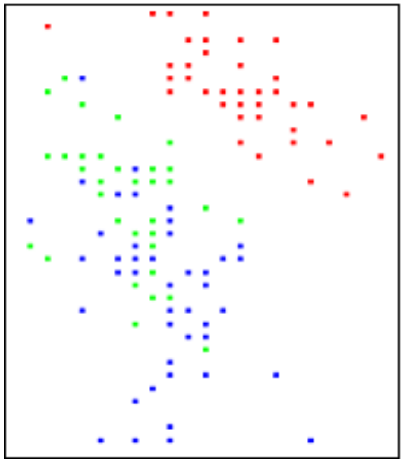


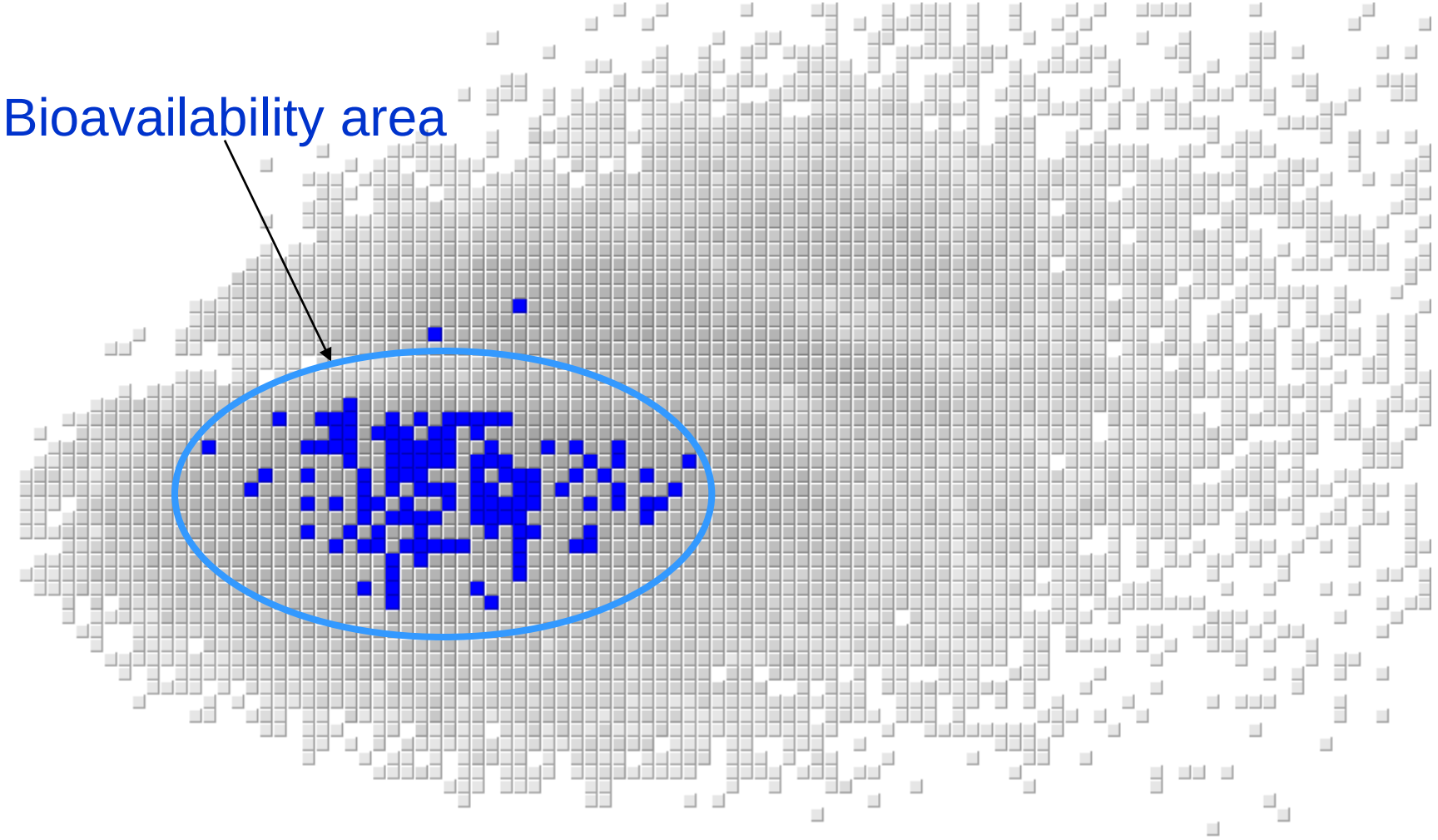
table with properties or fragments

dimensionality reduction

visualization

Highly dimensional data matrices need to be reduced to few dimensions (optimally 2 to enable visual analysis) and the results provided in a visually appealing form – as diagrams, graphs or maps – to help us see and understand complex relationships within the data.

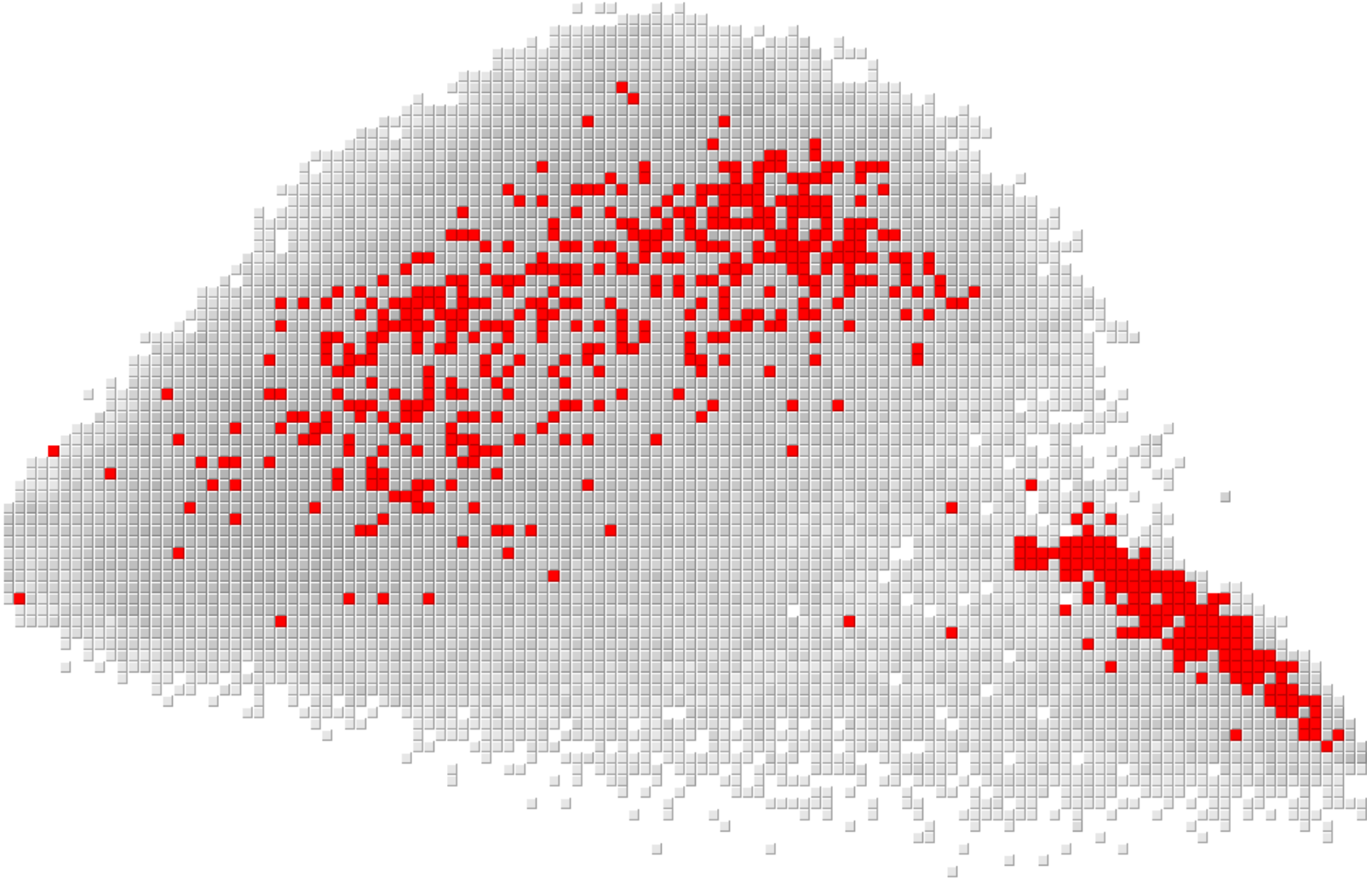
Molecular Property Space



200,000 organic molecules, 10,000 drugs and development candidates

NOVARTIS

Structural Diversity Space



organic molecules, drugs

NOVARTIS

Use of “Molecule Lighthouses” to Navigate in the Chemistry Space

Chem-GPS system developed by Astra-Zeneca

chemistry space is characterized by
set of exotic molecules with “extreme” properties

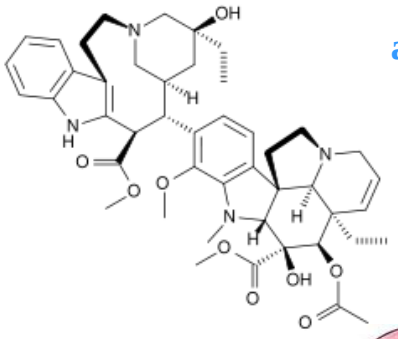
Use of marketed drugs, or other
bioactive molecules

many virtual screening techniques are based
on drug-likeness identification of molecules
similar to know bioactive structures

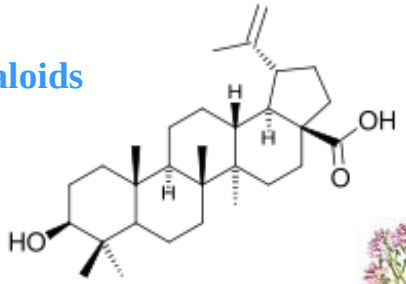
Natural Products as a starting points

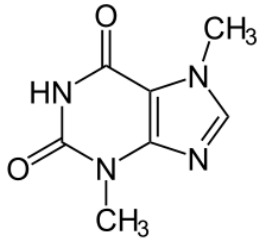
NPs have been optimized in very long natural selection process for optimal interactions with biological macromolecules

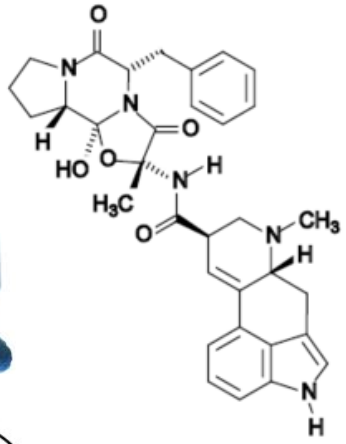
Natural Products - Source of Bioactivity

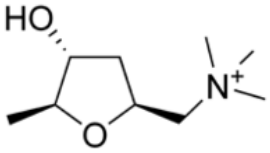


alkaloids







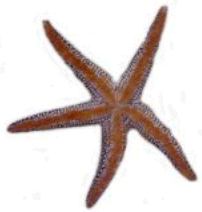


plants

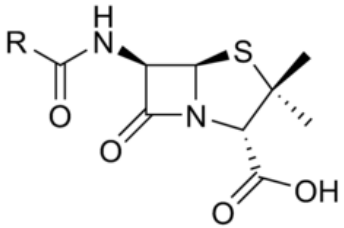
fungi

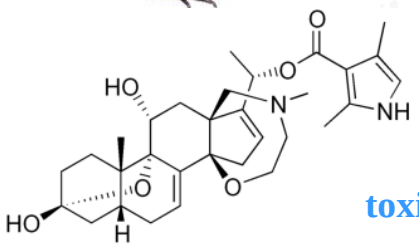
bacteria

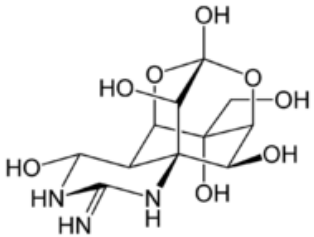
(marine) animals



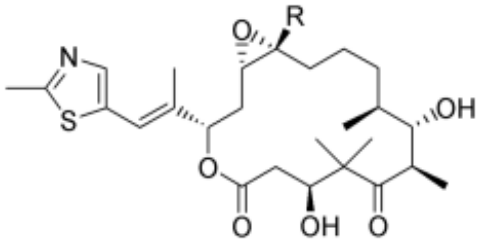
antibiotics







toxins



NOVARTIS

Natural Products as a Source of New Drugs

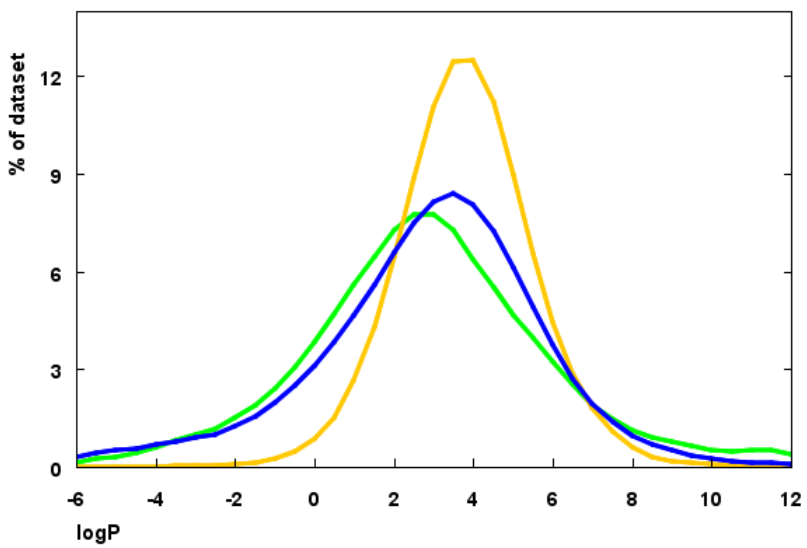
Natural products (NPs) have been optimized in a very long natural selection process for optimal interaction with biological macro-molecules.

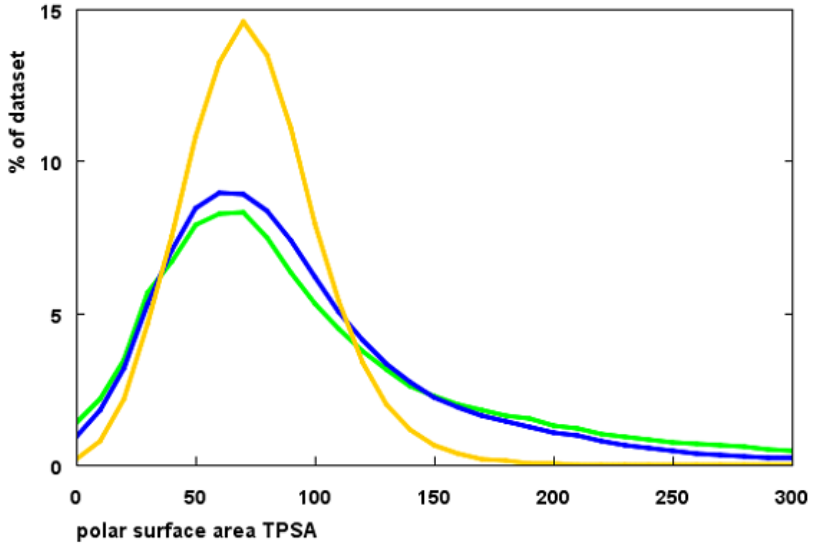
NPs are therefore an excellent source of validated substructures which may be used in the design of new bioactive compounds.

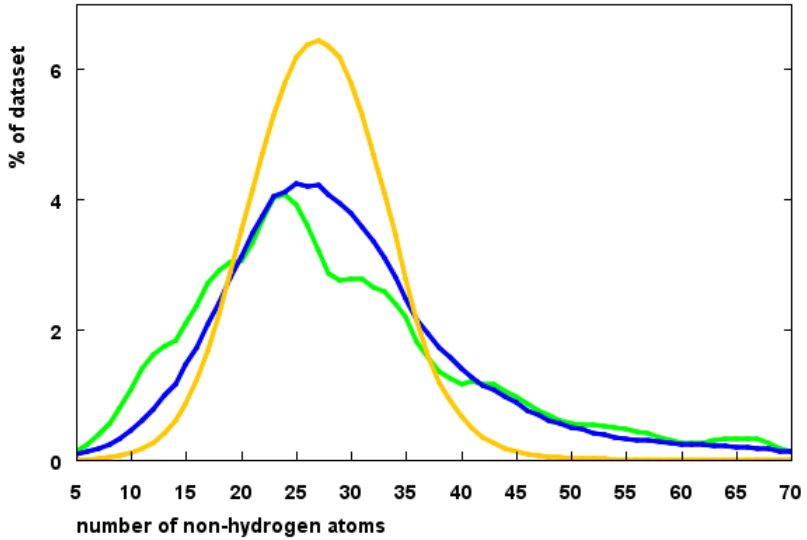
Large part of current pharmacopeia consists of NPs and many other NPs are under development as new drugs.

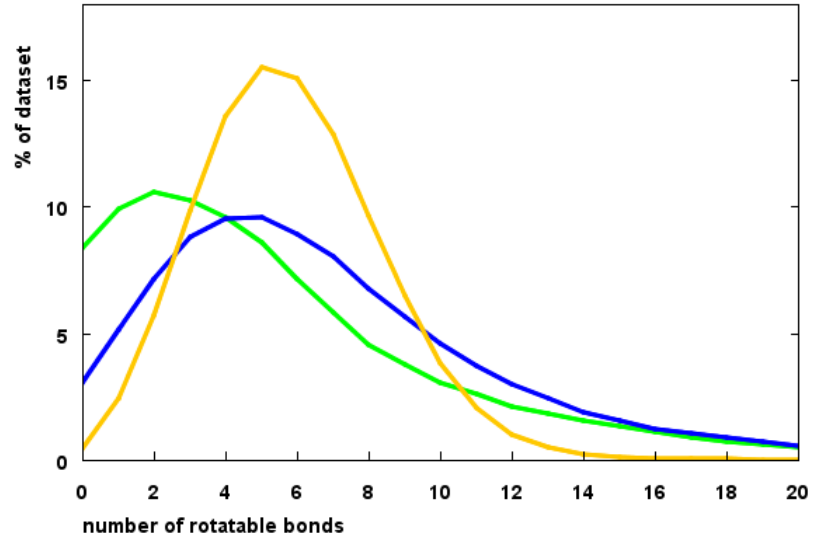
But what makes NP so successful in interacting with protein targets?

Global Molecular Properties







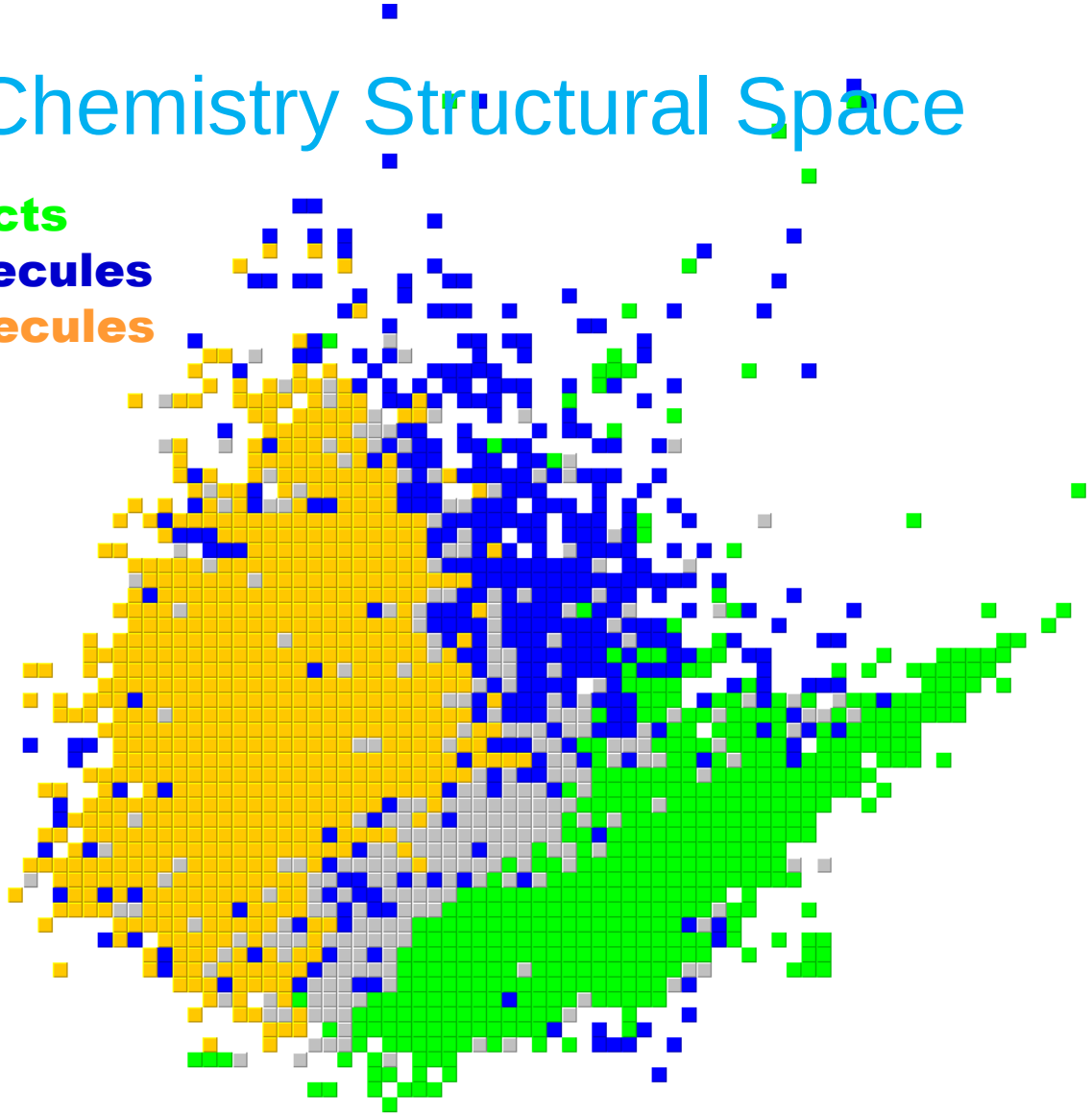


Natural Products (130k) Bioactive molecules (120k) Synthetic molecules (150k)

NOVARTIS

Chemistry Structural Space

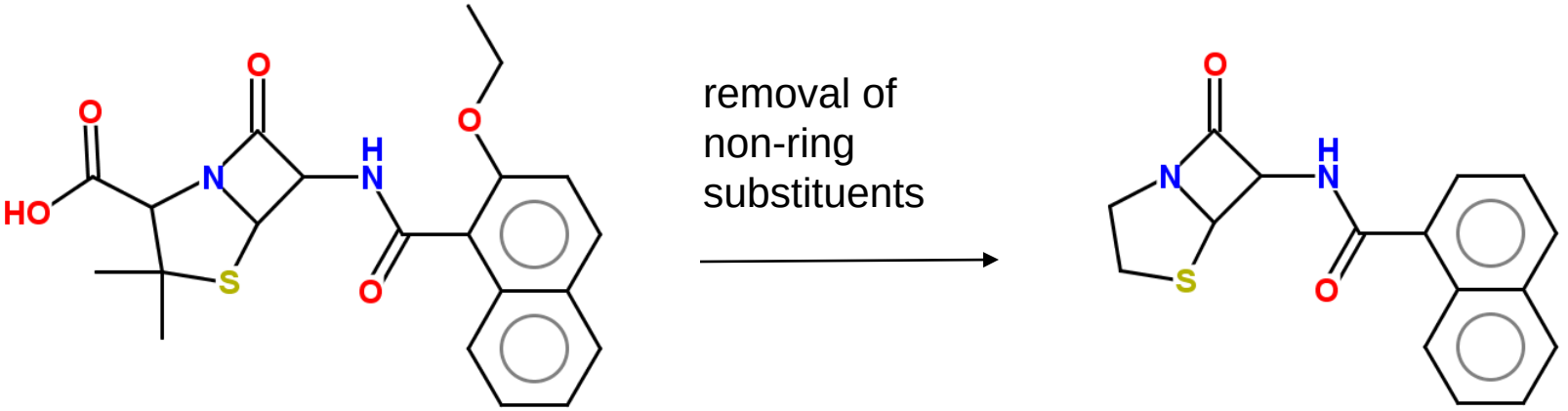
Natural Products
Bioactive molecules
Synthetic molecules
Mixed



P. Ertl, A. Schuffenhauer, Cheminformatics Analysis of Natural Products**, in:** ***Natural Compounds as Drugs, Vol. II*****, Springer, 2008**

NOVARTIS

Molecule Scaffold as Classifying Element

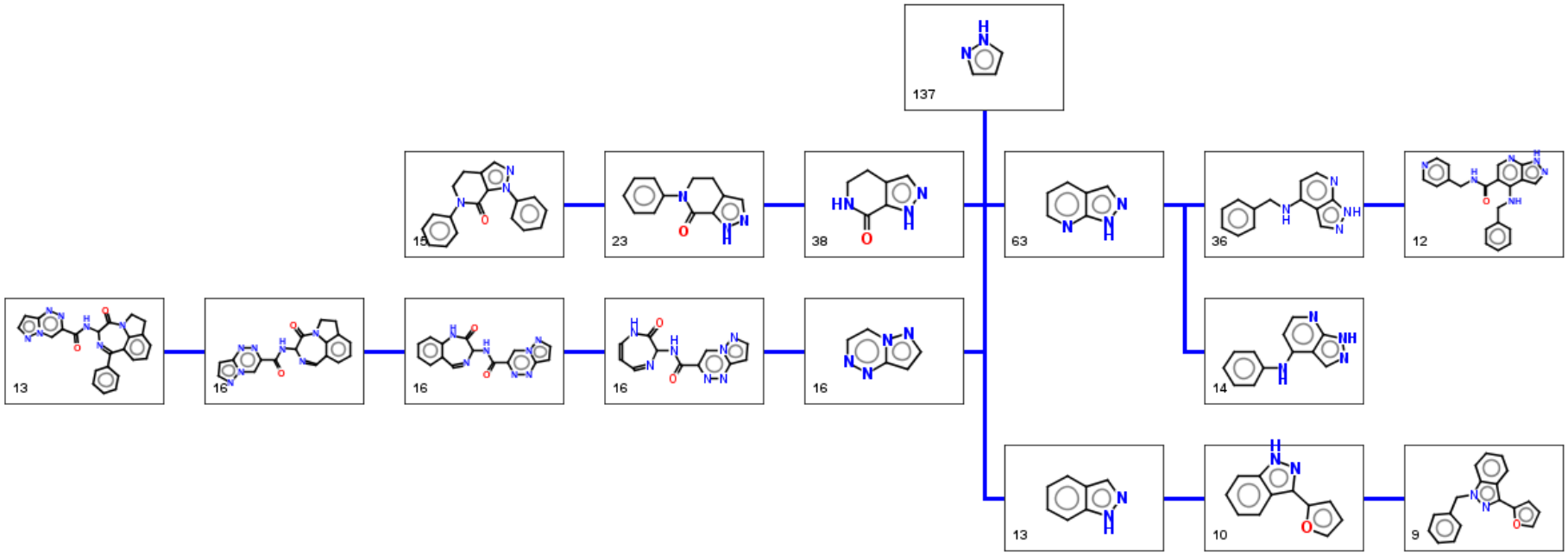


molecule scaffold

- scaffold is the most important part of the molecule, giving it its shape and keeping substituents in their proper positions
- scaffold influence global molecular properties
- scaffolds often determine biological activity of the parent molecule
- provide easy understandable, common natural language between synthetic and computational chemists
- scaffolds play important role in combinatorial chemistry and scaffold hopping applications

The Scaffold Tree

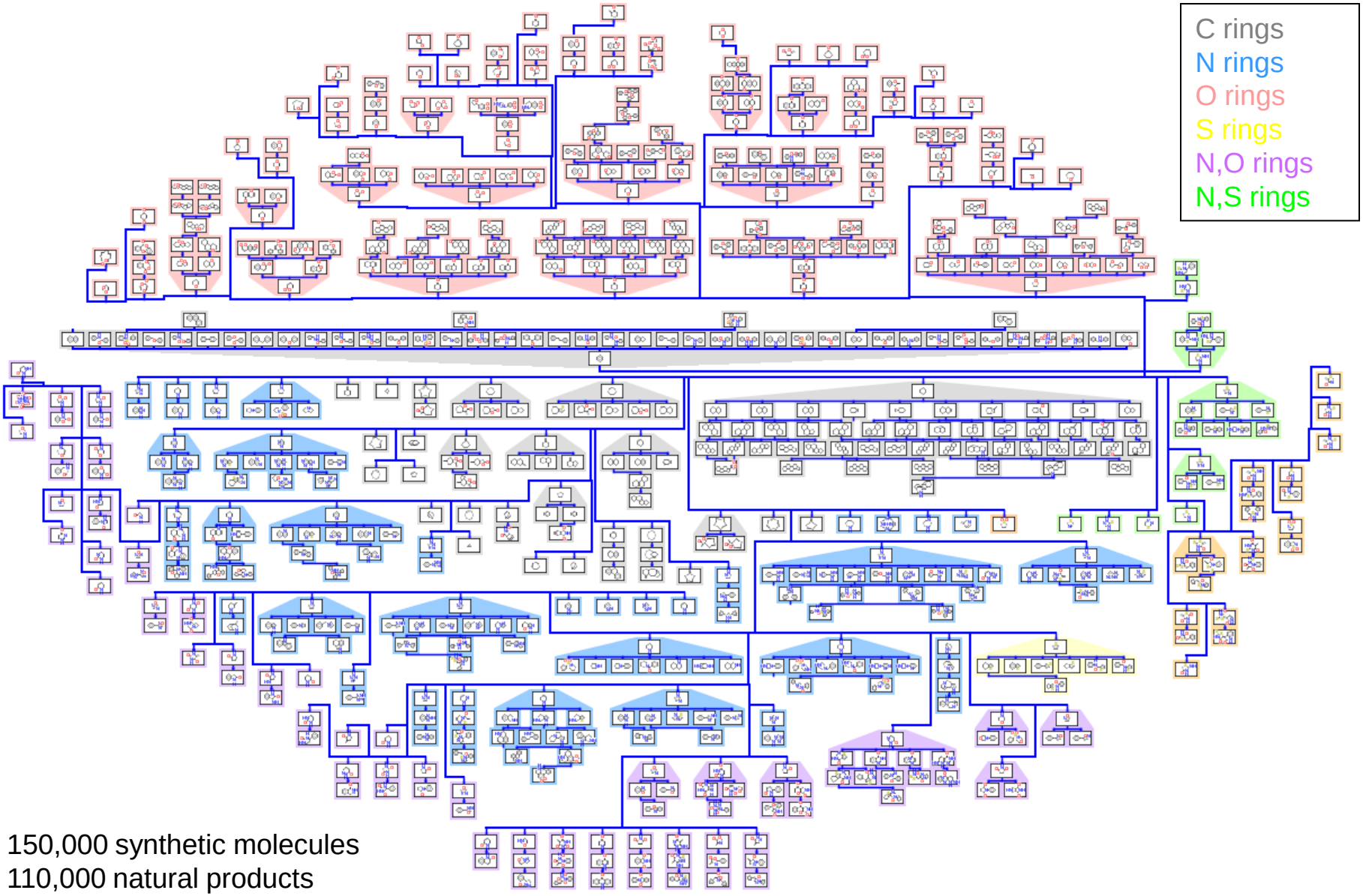
A method to classify molecules based on their scaffolds. Molecules are converted to their frameworks, then rings are removed one-by-one based on a set of predefined rules, creating such a scaffold hierarchy



The Scaffold Tree – Visualization of the Scaffold Universe by Hierarchical Scaffold Classification
A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. Koch, H. Waldmann, *J. Chem. Inf. Model.* 47, 47, 2007

NOVARTIS

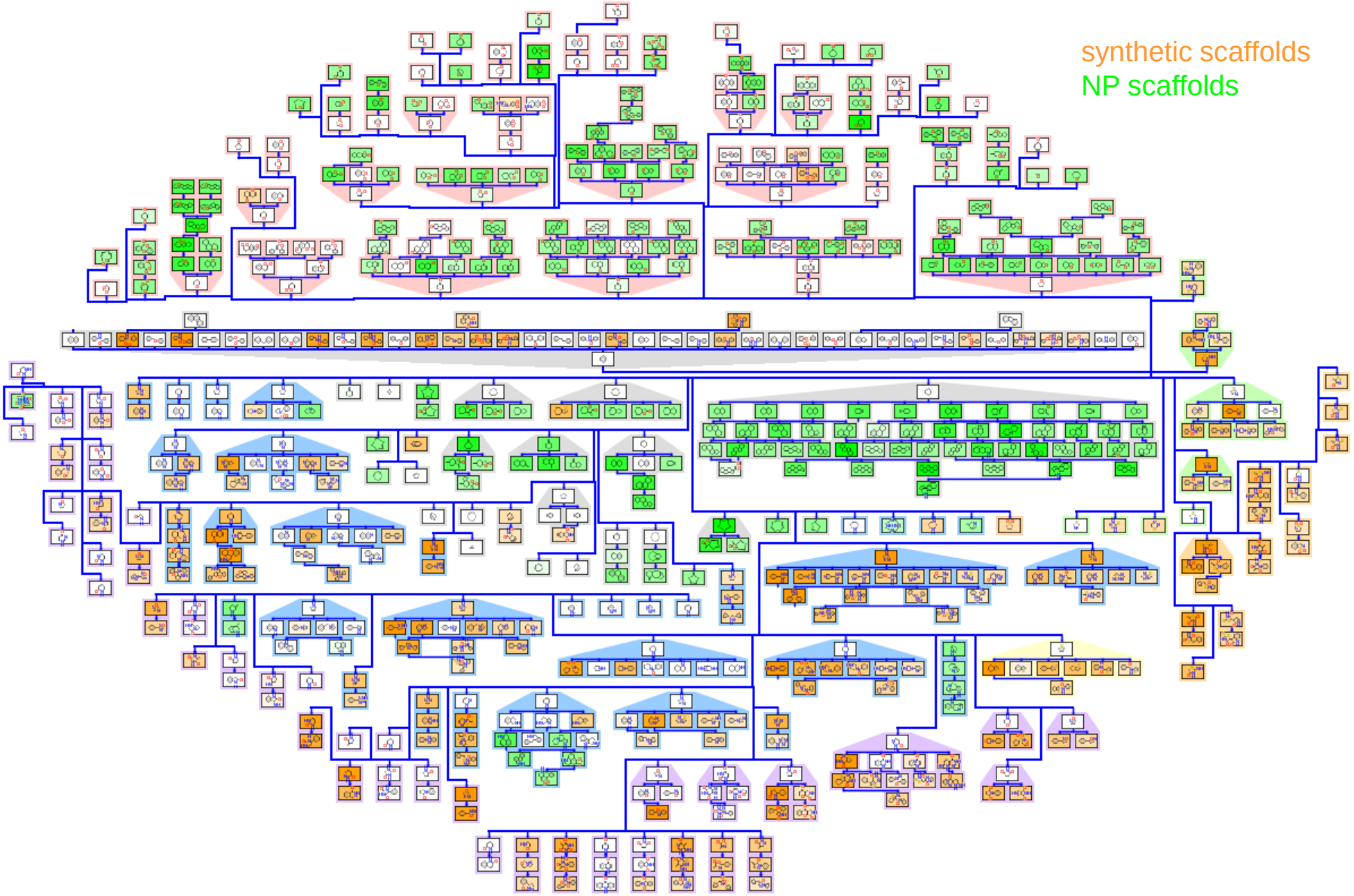
Natural Product Scaffold Tree



150,000 synthetic molecules
110,000 natural products

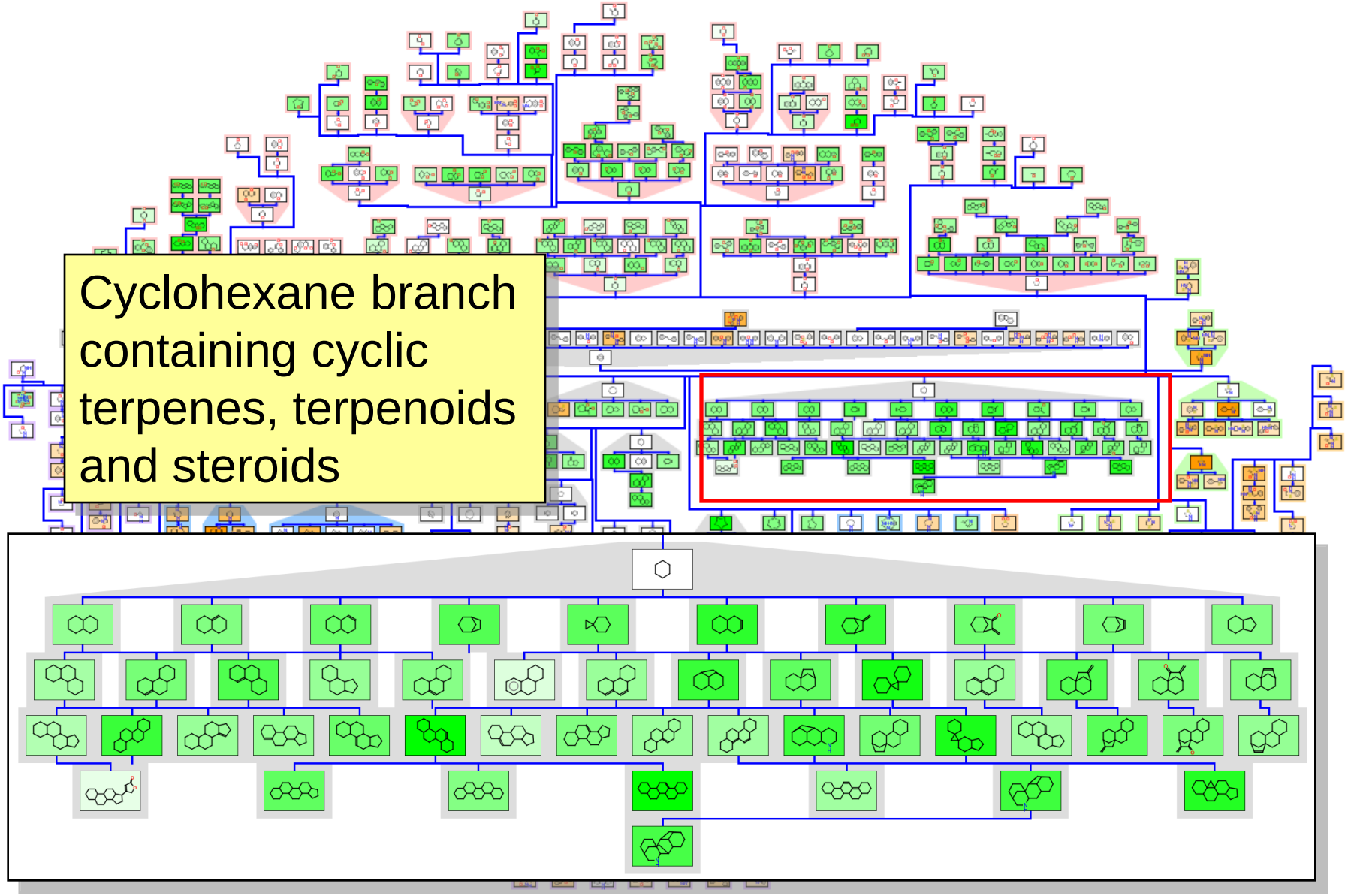
NOVARTIS

Natural Product Scaffold Tree



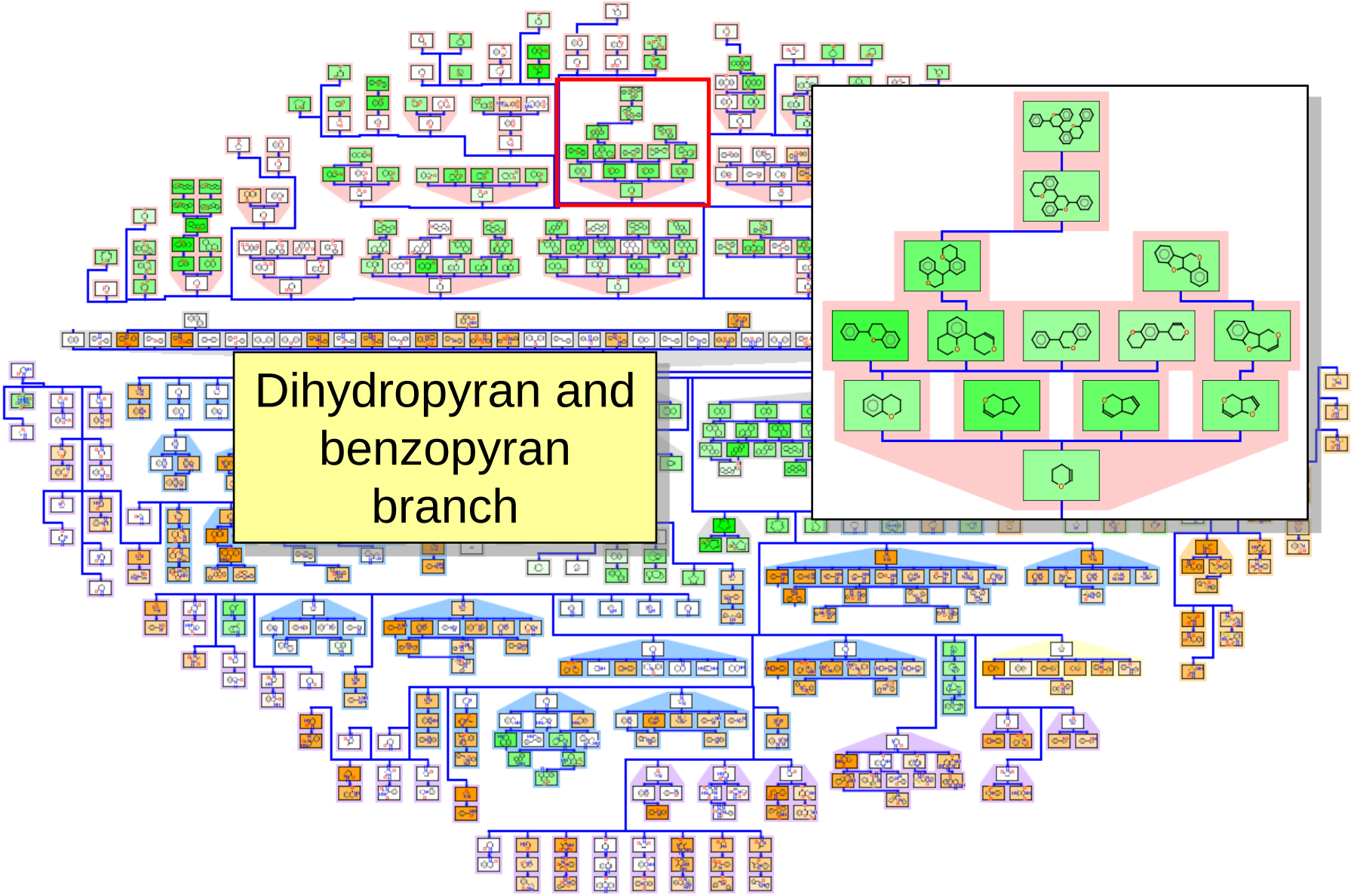
NOVARTIS

Natural Product Scaffold Tree



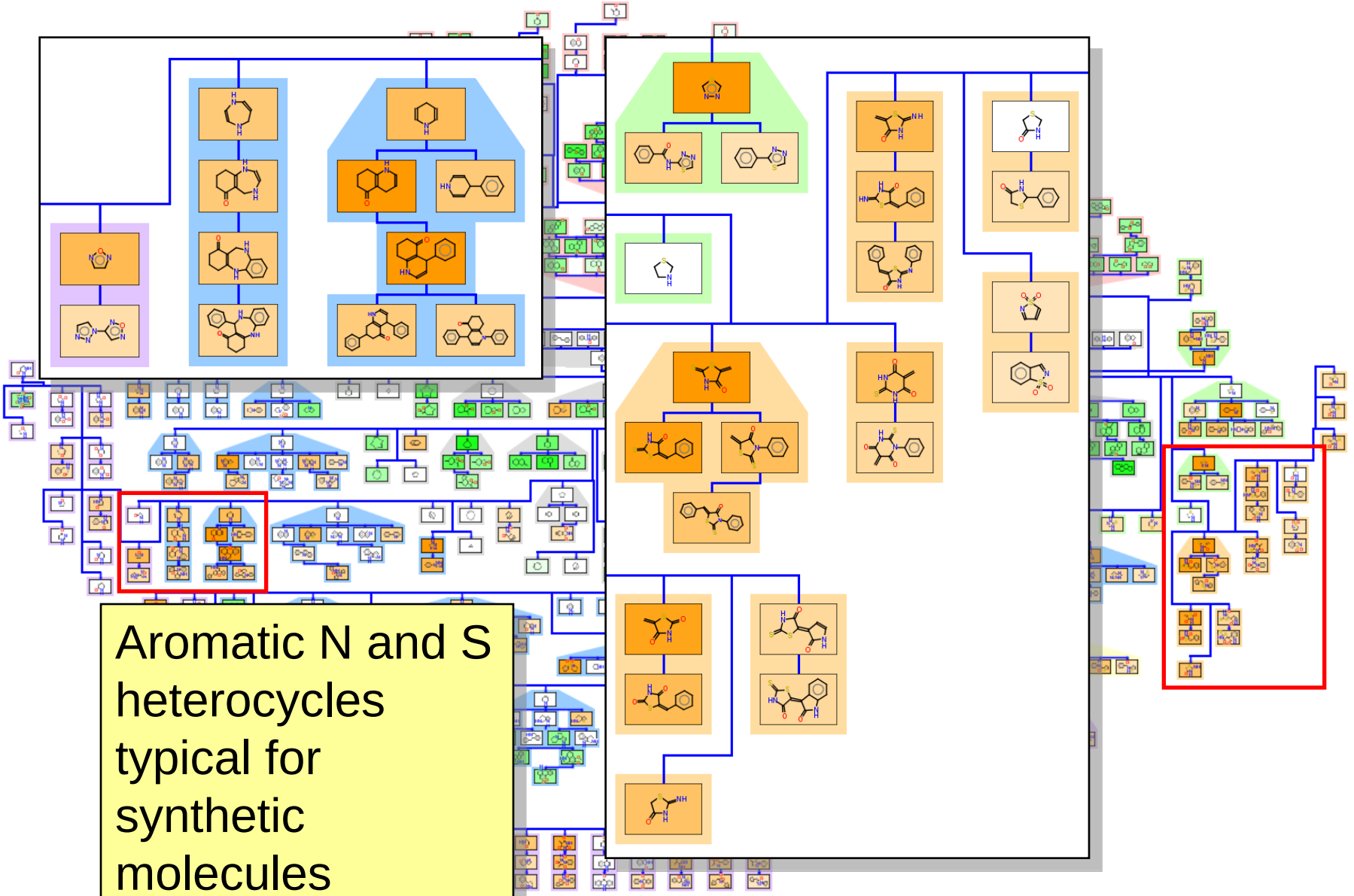
NOVARTIS

NP Scaffold Tree



NOVARTIS

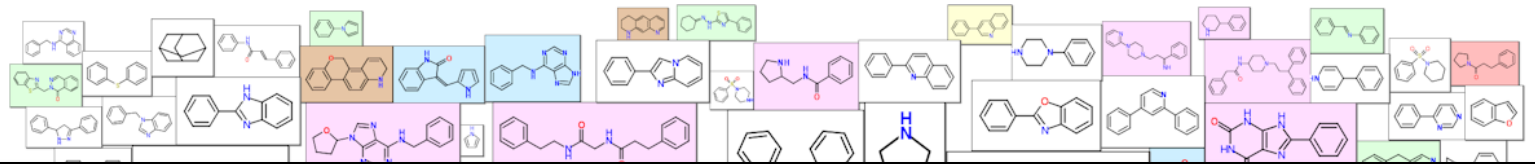
Natural Product Scaffold Tree



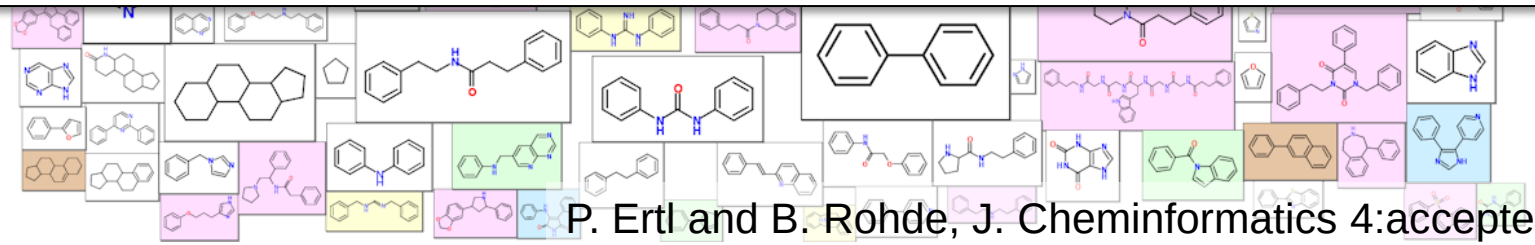
Aromatic N and S heterocycles typical for synthetic molecules

NOVARTIS

Molecule Cloud



activity algorithm analysis areas available bioactive cheminformatics chemistry
cloud collections common compact containing data database diagram
display example features figure frequency implemented information java
largest layout method million molecular molecule cloud
molecules novartis number overlap particular present processing provided
pubchem scaffolds sets size smiles source structures substructure text
typical visualization zinc



P. Ertl and B. Rohde, J. Cheminformatics 4:accepted (2012)

NOVARTIS

Chemistry Space - Summary

- The chemistry space is huge and is growing nearly exponentially, we need reliable cheminformatics and data mining methods to navigate in this maze
- Chemical space may be characterized by numerous parameters, select those which are relevant to your problem
- Classification of chemistry space based on scaffolds is intuitive and provides good results
- We can use known areas of chemistry space as reference; natural products provide useful, evolutionary validated starting point for design of new bioactive molecules
- Visualization techniques are indispensable when analyzing large molecule collections - visualization in 2D (maps, trees, clouds) of complex data can help to understand the data