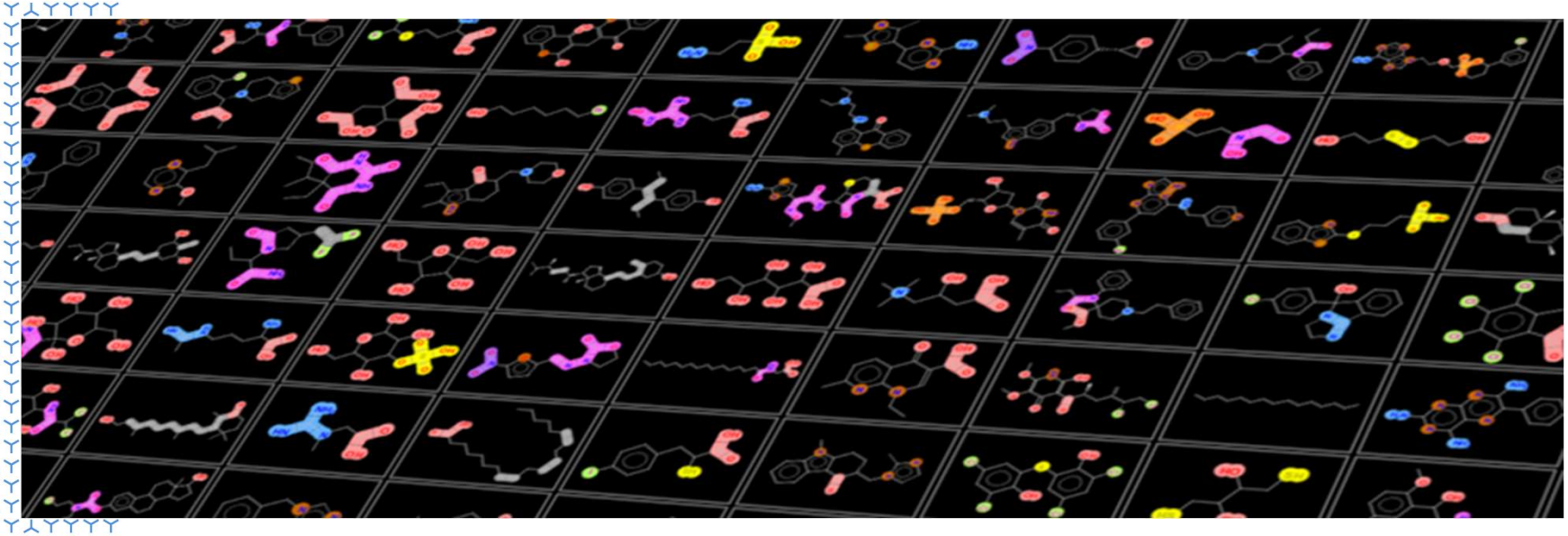


Fun with Functional Groups

Peter Ertl
Novartis Institutes for BioMedical Research, Basel, Switzerland
September 2021

NOVARTIS | Reimagining Medicine

Functional Groups

A functional group (FG) is a group atoms and bonds in a molecule that is responsible for the characteristic chemical reactions of this molecule.

Functional groups form a cornerstone of synthetic chemistry, medicinal chemistry, pharmacokinetics, toxicity, spectroscopy ...

Cheminformatics analysis of FGs provides very useful information about molecular properties and can guide us in navigation in chemical space to identify interesting, underrepresented areas with potentially good biological activity.

How to find the functional groups ?

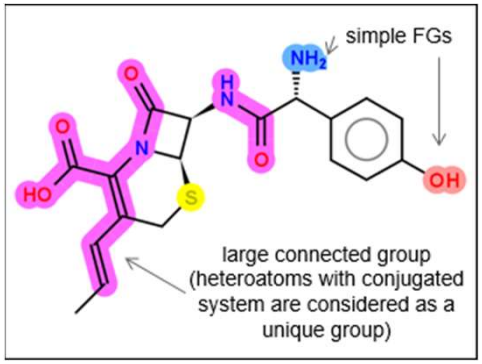
- traditionally, FGs are identified by a set of predefined SMARTS substructure queries
- BUT ... the chemistry is very diverse, and even the largest set of queries published cannot by far cover the whole diversity of FGs space
- particularly for special collections of molecules (natural products, ligands active on specific target ...) the predefined queries do not work
- it was therefore necessary to develop a new method to identify all functional groups in a molecule by an algorithmic approach

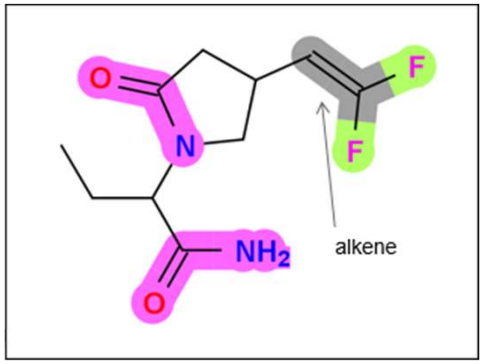
Methodology to identify functional groups

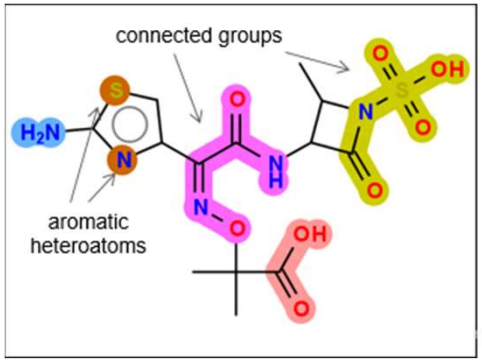
Algorithmic definition of a functional group:

- nonaromatic heteroatoms connected with other heteroatoms or through multiple bonds with carbon
- set of connected carbons (C_{ar} and C_{al} are distinguished) form an environment of the FG (but such carbons are not part of the FG itself)
- alkene and alkyne groups are also FGs, as well as 3-atomic reactive heterocycles like oxirane or aziridine
- heteroatoms in aromatic systems are not considered to be FGs

Examples of functional groups



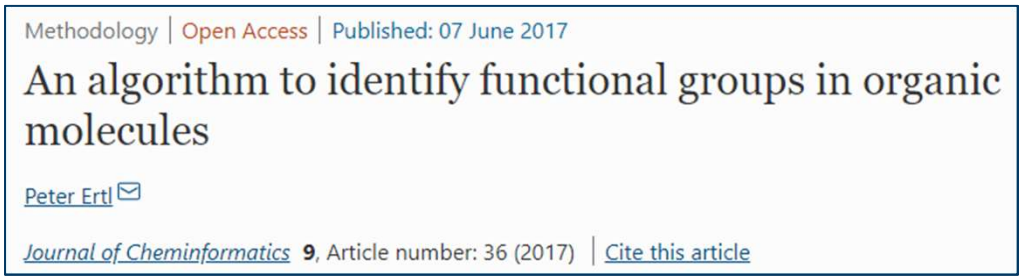




The algorithm may be fine-tuned by defining the level of FG environment (C_{ar}, C_{al}, H). In some cases (alcohols / phenols, amines / anilines) this is important, for larger groups this information needs to be generalized, since too detailed data would lead to too high fragmentation of FGs.

Methodology details and availability

Full details of the method are described in:


Methodology | Open Access | Published: 07 June 2017

An algorithm to identify functional groups in organic molecules

Peter Ertl

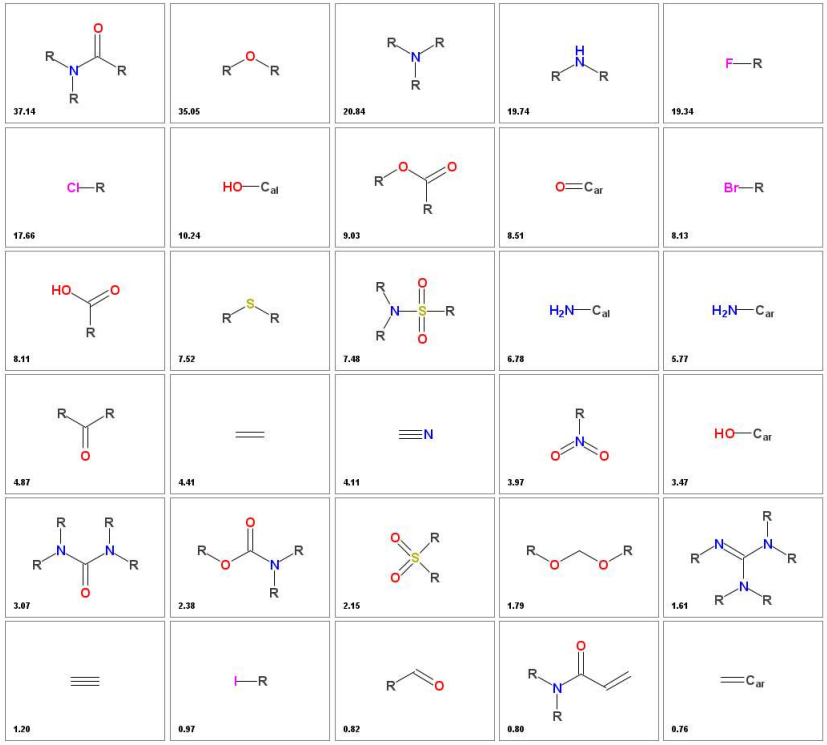
Journal of Cheminformatics **9**, Article number: 36 (2017) | Cite this article

2 open source implementations are available:

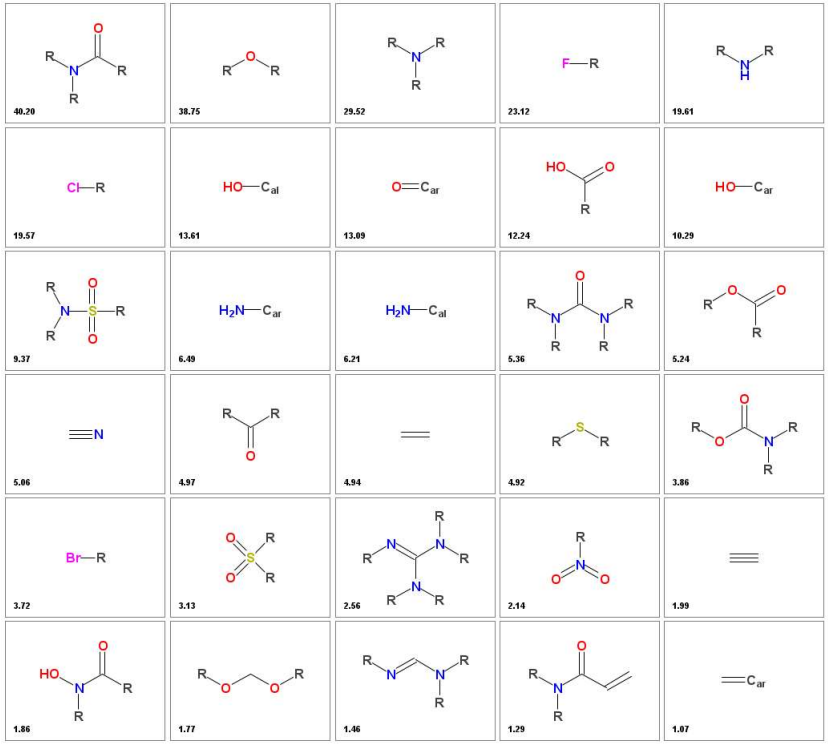
- Python implementation in RDKit
- Java implementation in CDK (J. Cheminformatics 11 (37) 2019)

The most common functional groups

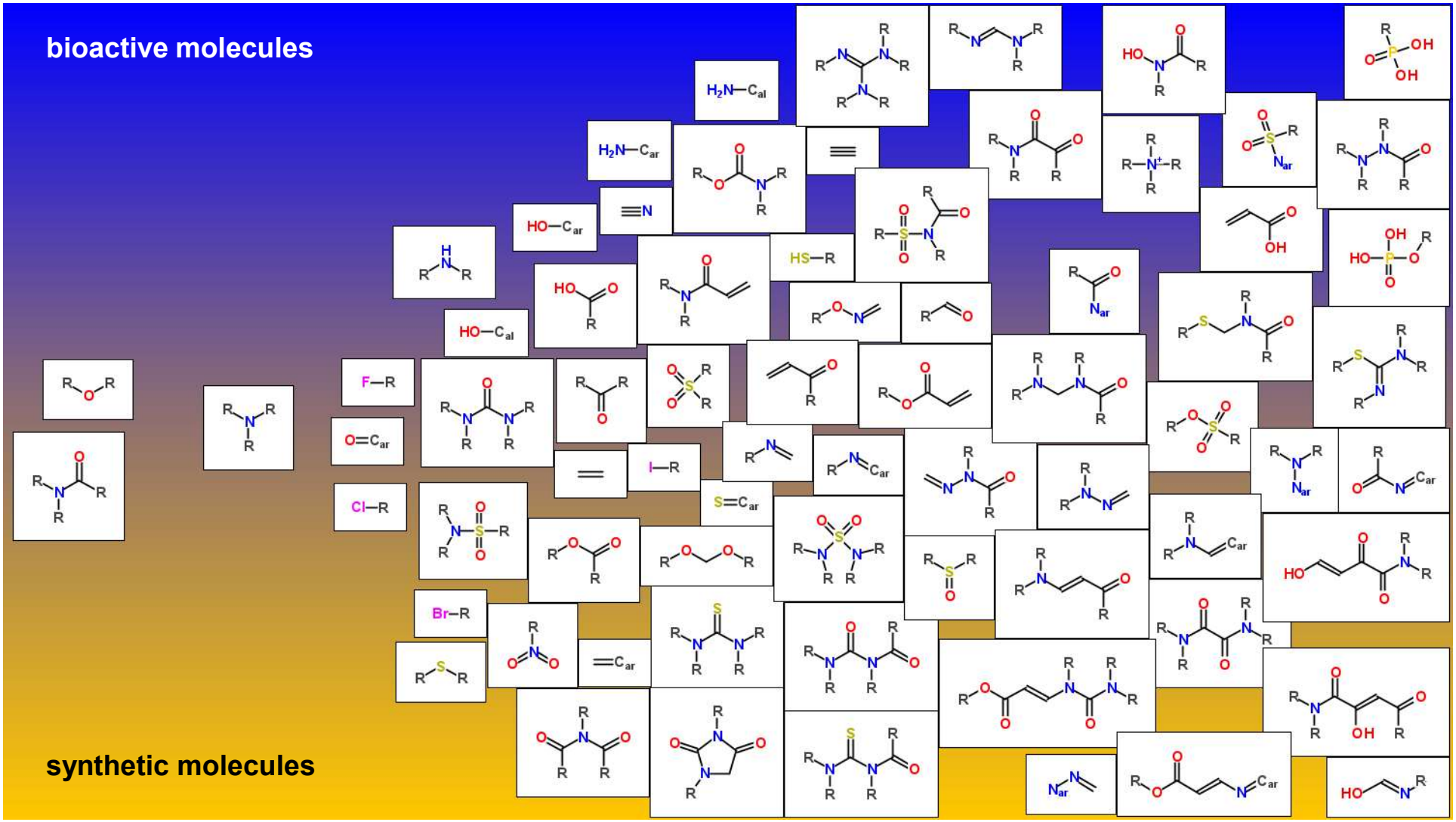
PubChem



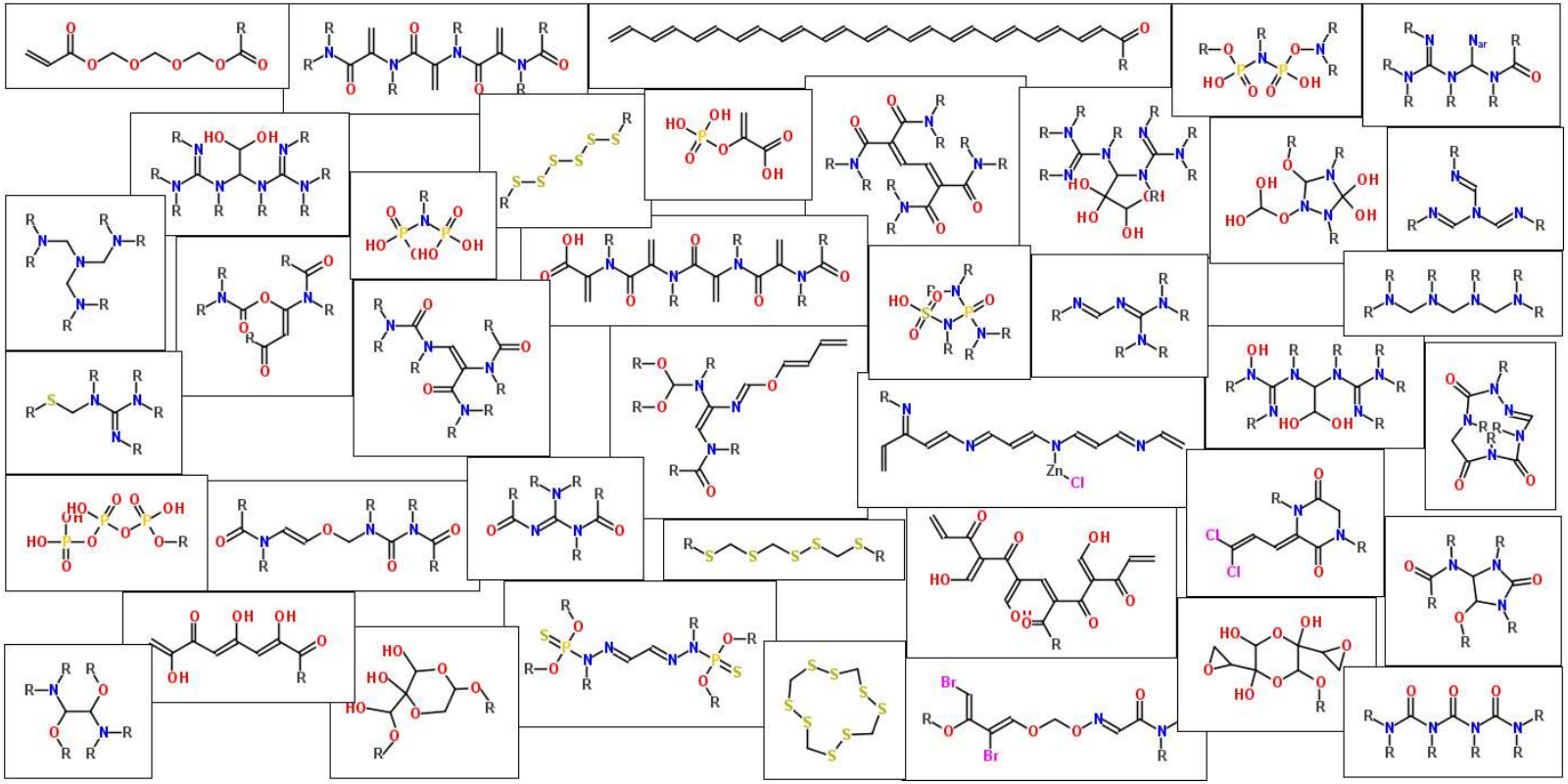
ChEMBL



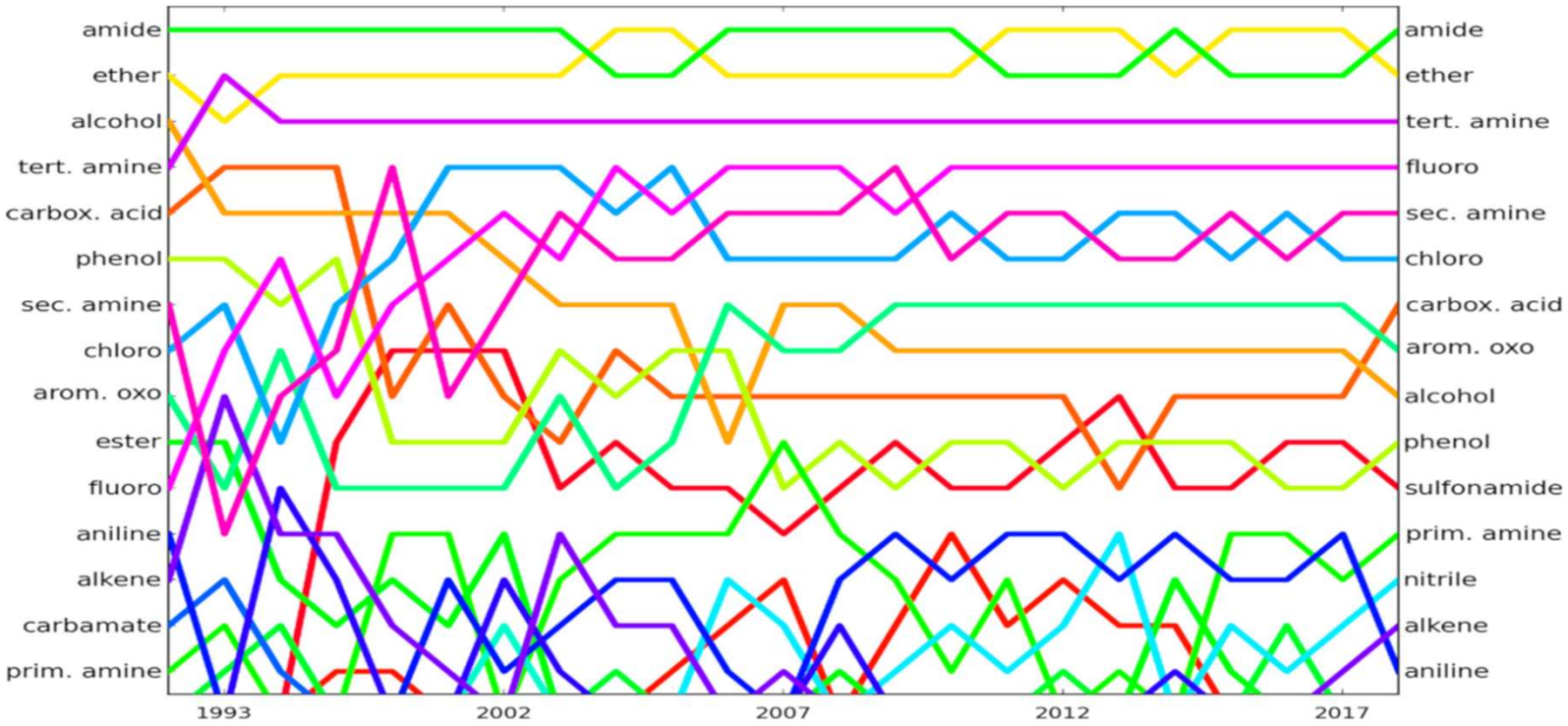
NOVARTIS | Reimagining Medicine


bioactive molecules
synthetic molecules

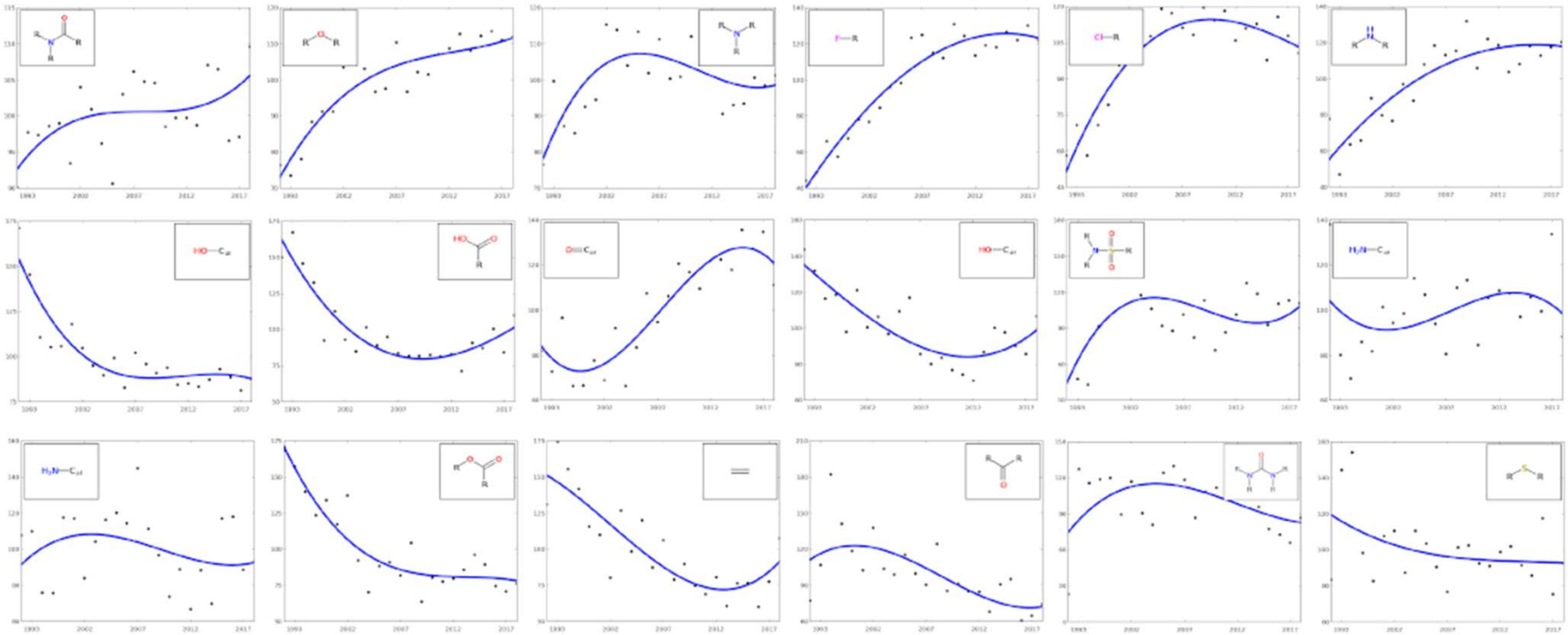
Bestiarium of functional groups



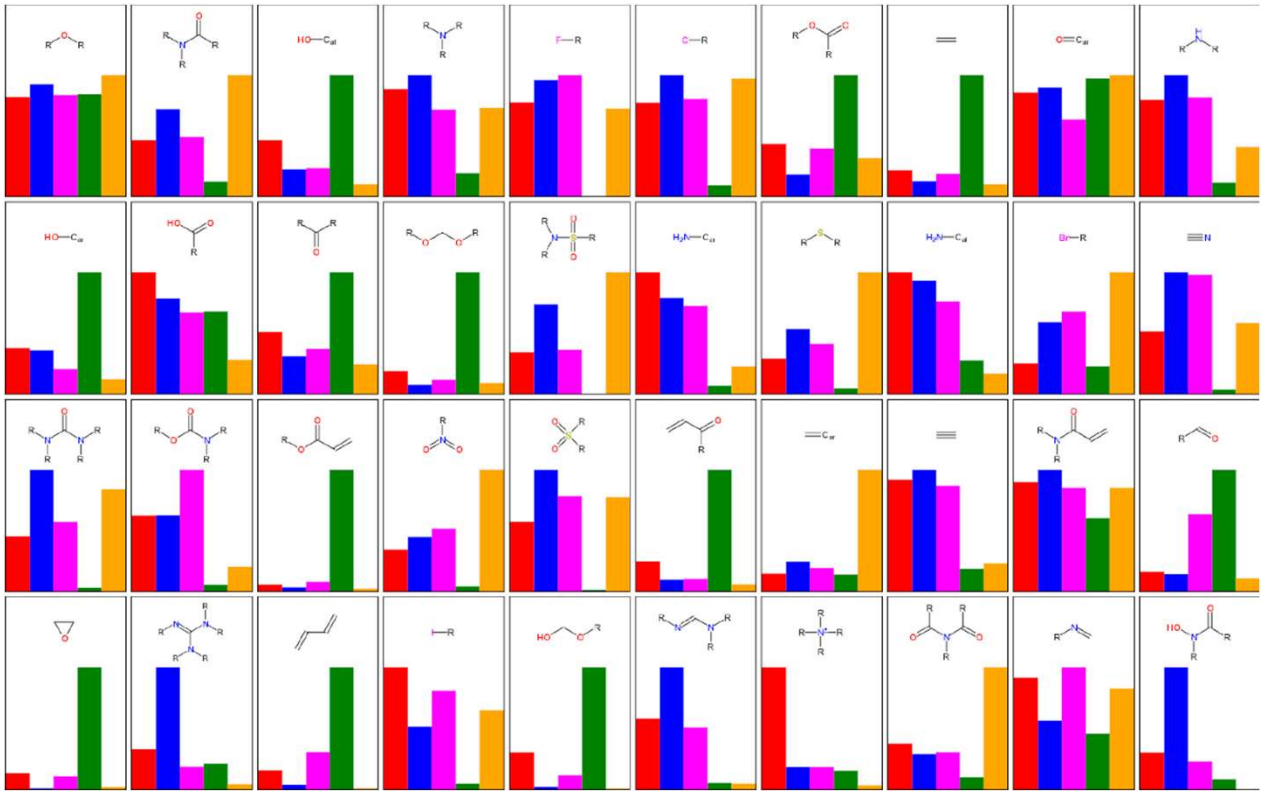
Popularity of FG over time



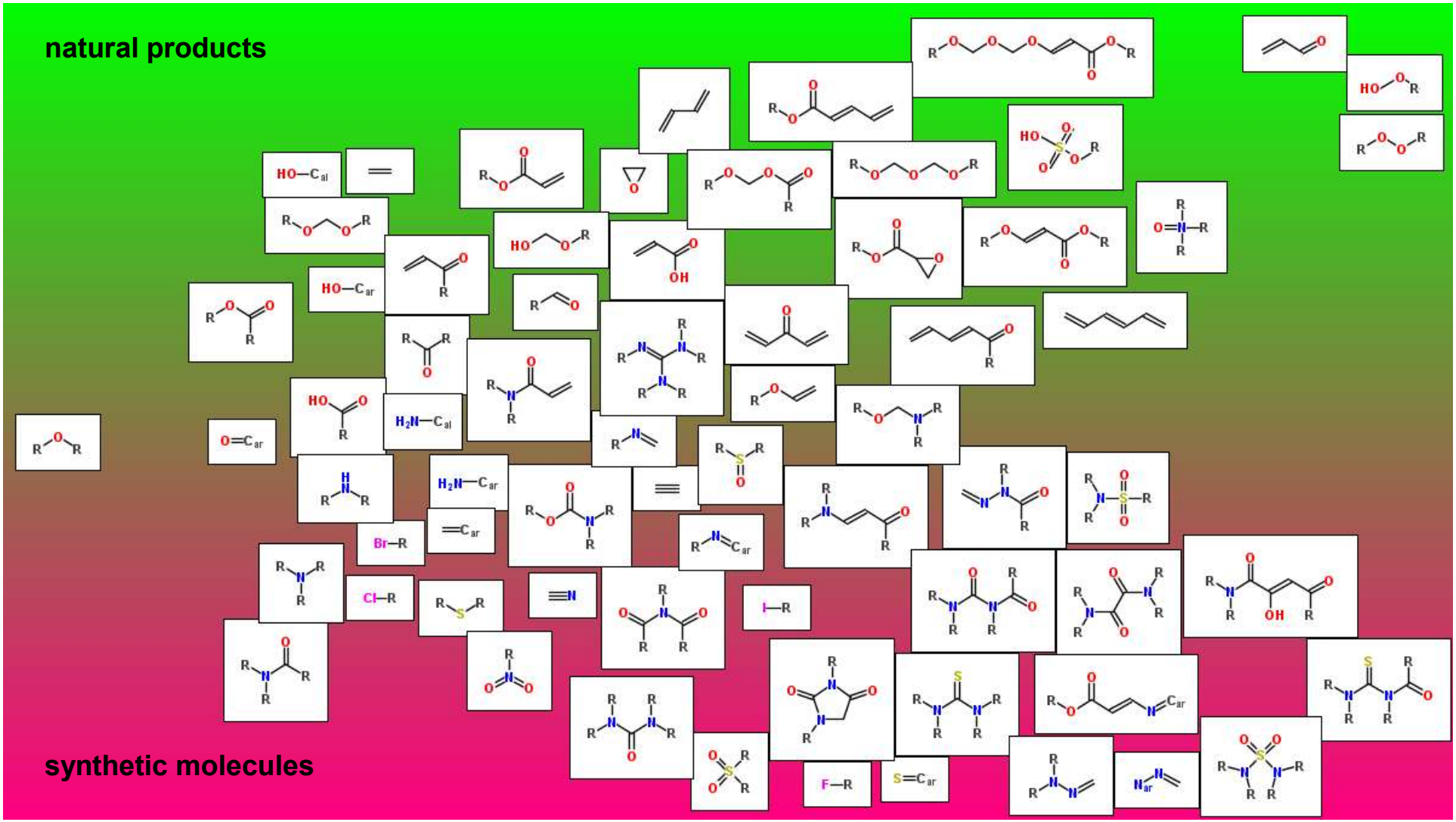
Popularity of FG over time



FGs in different molecular collections



drugs
bioactive molecules
patents
natural products
synthetic molecules


natural products
synthetic molecules

Summary

- a novel method to identify functional groups was developed and applied to a large collection of natural products
- the most common FGs in NPs were identified, distribution of FGs in NPs of different origins and also differences between FGs in NPs and bioactive molecules were analyzed

JOURNAL OF NATURAL PRODUCTS

More details are available at: P. Ertl and T. Schuhmann, A systematic cheminformatics analysis of functional groups occurring in natural products, *J. Nat. Prod.* 82, 1258 (2019)

Journal of Medicinal Chemistry

Application in MedChem: P. Ertl, E. Altmann, J. McKenna, The most common functional groups in bioactive molecules and how their popularity has evolved over time, *J. Med. Chem.* 63, 8408 (2020)

NOVARTIS | Reimagining Medicine