INSTYTUT CHEMII
UNIWERSYTETU ŚLĄSKIEGO

# Od architektury do topologii chemii

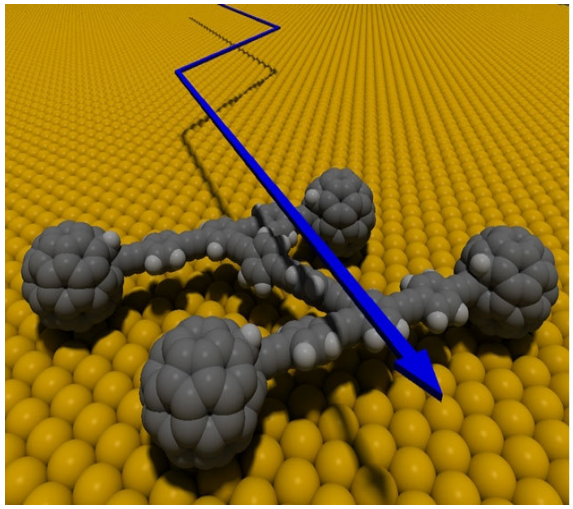
***Jarosław Polański***
*Institute of Chemistry*
*University of Silesia*
*polanski@us.edu.pl*

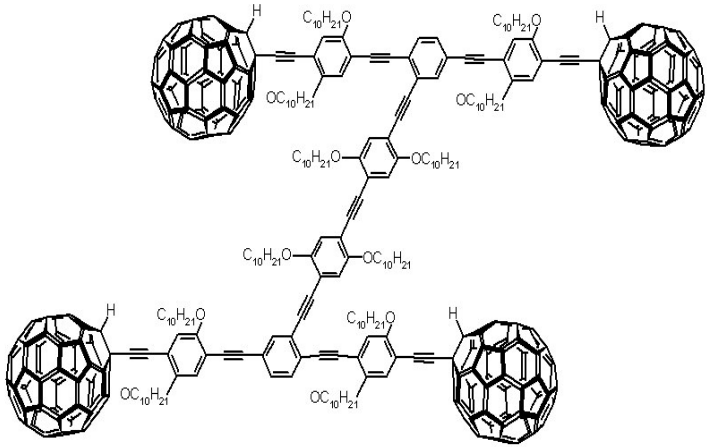
Wisła 2010

# Chemistry paradox

*(...) the most fundamental and lasting objective of (chemical) synthesis is not a production of new compounds but the production of properties.*

*Hammond*

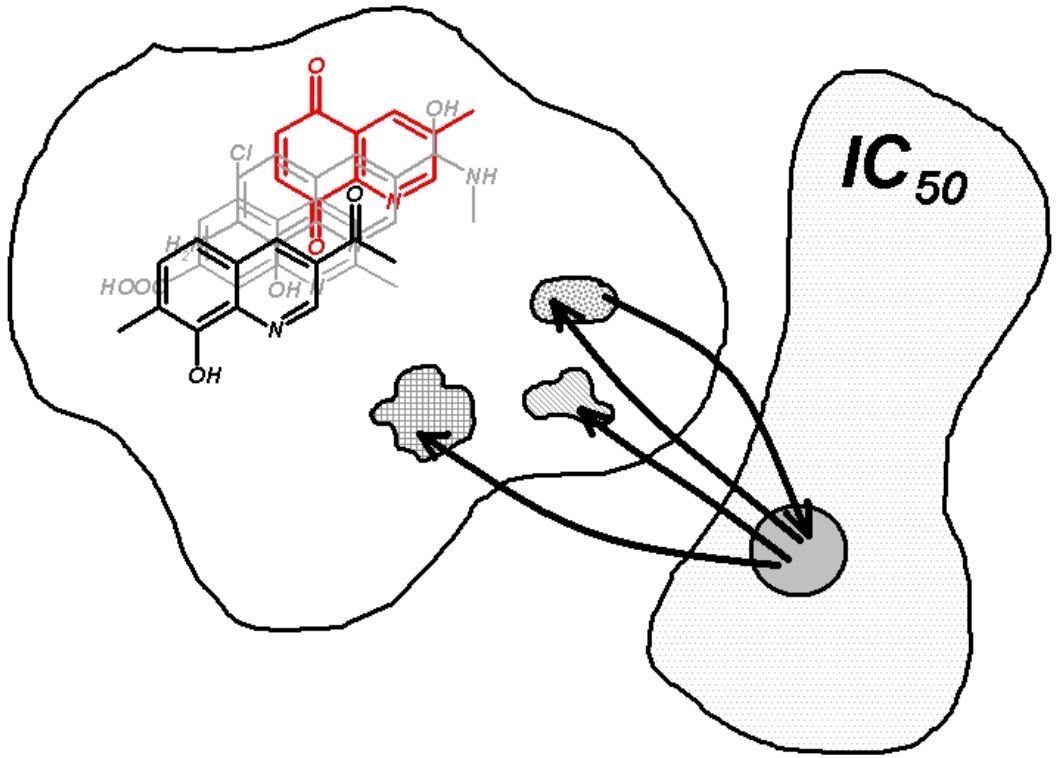




*Kolb, H.C.; Finn, M.G.; Sharpless, K.B. Angew. Chem. Int. Ed.* ***2001****, 40, 2004 – 2021*

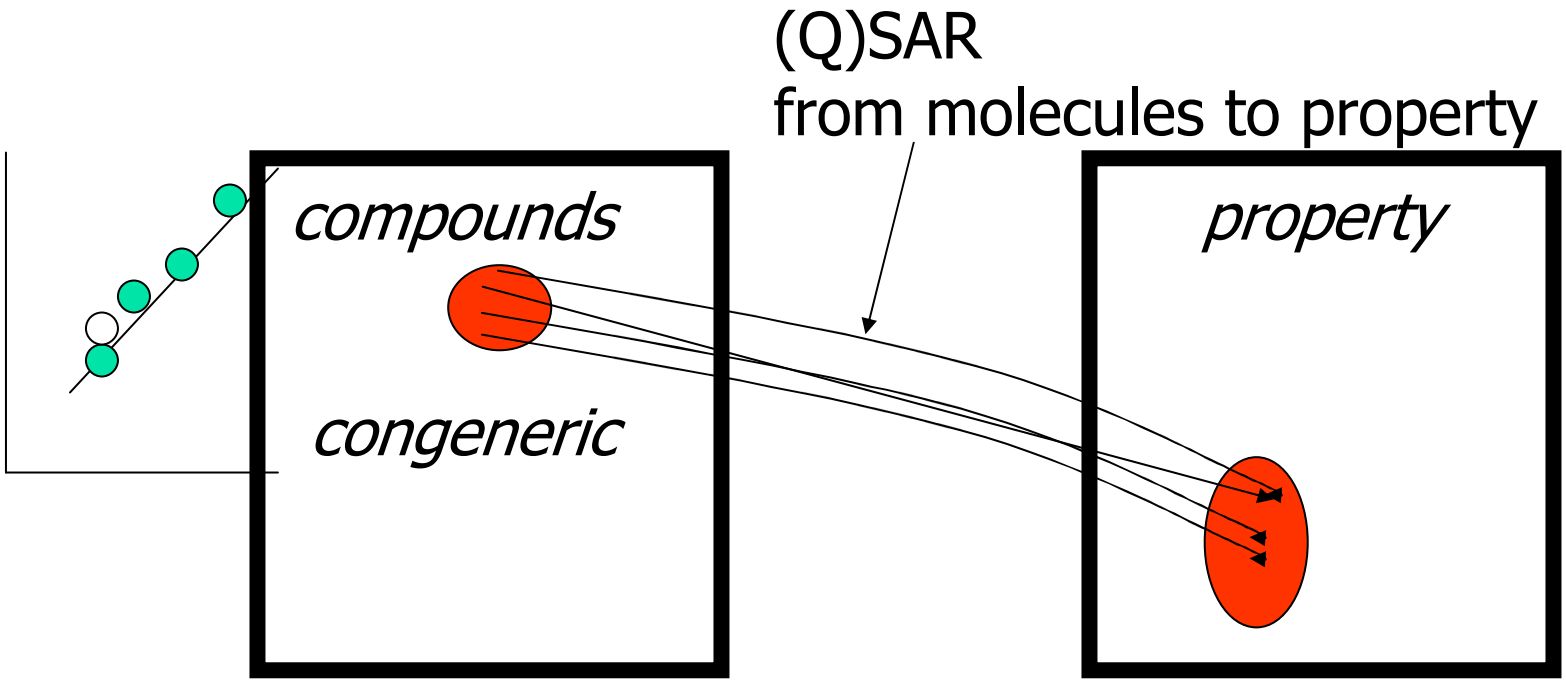
# Structure -property

- A lack of obvious relation P ??? f(S)



# Thus, molecular design needs *exploratory mathematics*

$$P = f(S)$$



# *But …*

*Auguste Compte 1798-1857*

*Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry.*

*If mathematical analysis should ever hold an important place in chemistry (…) it would occasion a swift and general degeneration of that science.*

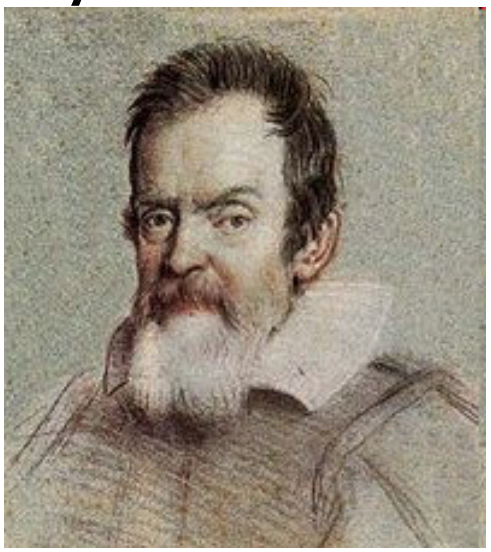
*Current Medicinal Chemistry,* 2009, *16,* 4297-4313

# (Al)chemia

- Molekuła (jej transformacje);
  całe wieki trudne do bezpośredniej obserwacji i zrozumienia (często opis fenomenologiczy)
- Fizyka eksperymentalna;
  duże obiekty dostępne obserwacji (cz. elementarne - możliwy precyzyjny opis)

# Physics-Chemistry-Biology

- Mathematics– formal language for the description of nature (from ancient times)
- Physics– Galileo, Newton



Galileo Galilei (1564-1642)

Isaac Newton 1643-1727

wikipedia

# Physics-Chemistry-Biology

- Chemistry = ch. compounds /molecules

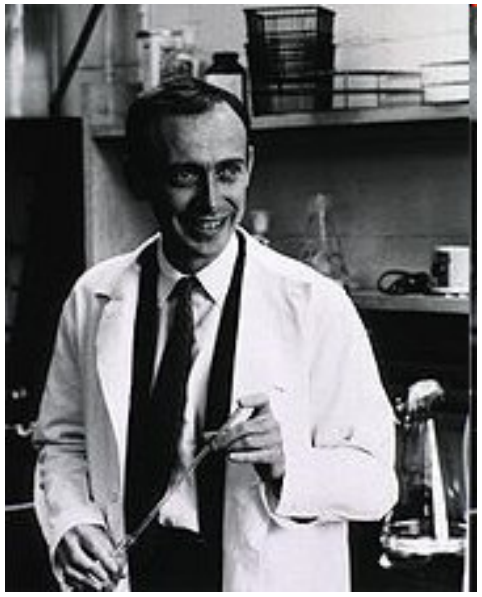
Antoine Lavoisier 1743-1794
wikipedia

John Dalton 1766-1844

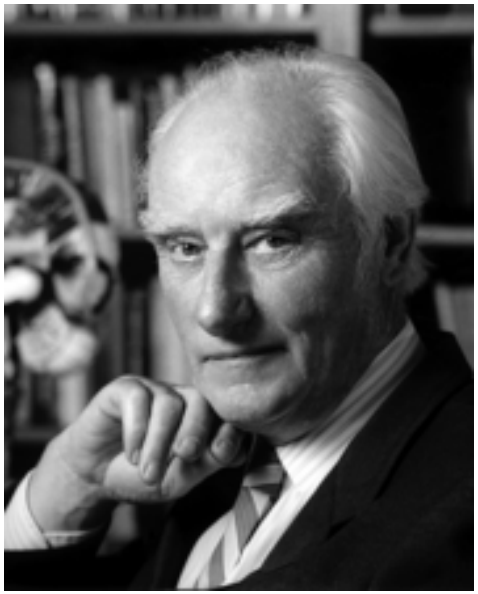
Jöns Berzelius 1779-1848

# Physics-Chemistry-Biology


- Molecular biology – chemistry of living systems



James Dewey Watson



Francis Crick 1916-2004

Craig Venter

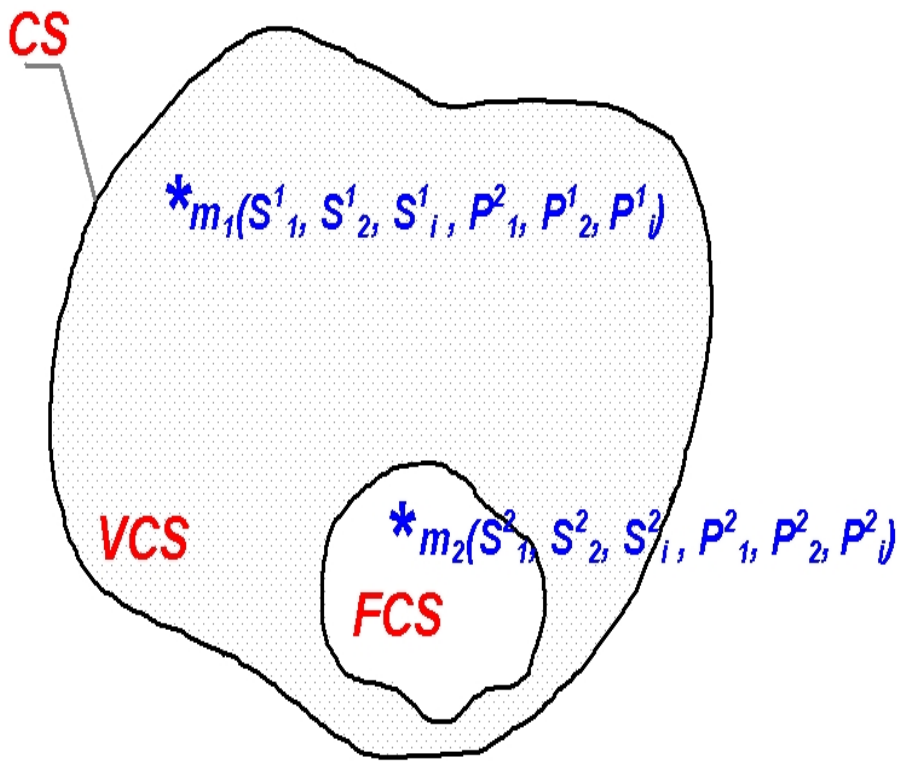
wikipedia

# Physics-Chemistry-Biology

well ordered in
formal structure

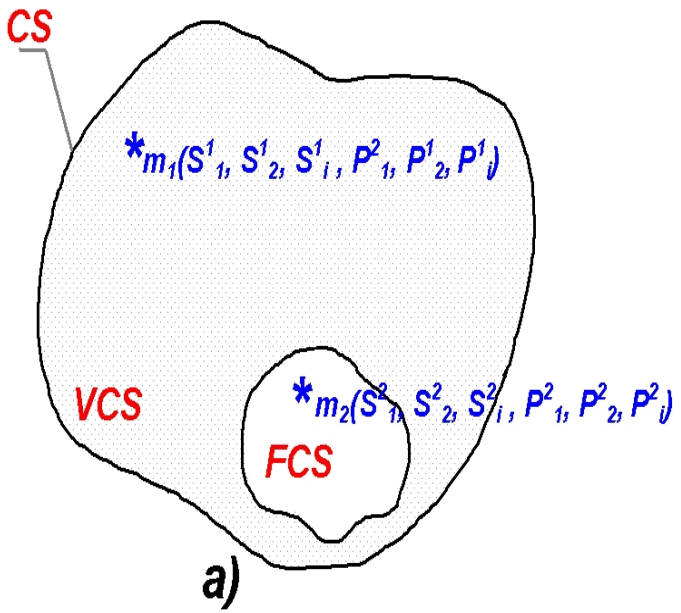
still to be …

# Chemical space CS

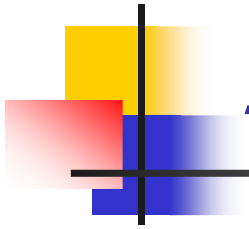



- chemical space CS
- virtual chemical space
- factual chemical space
- *In vitro* or *in silico* chemical operations = mapping molecules in CS

# CS study needs in silico mathematics




- CS (?) between $10^{18}$ and $10^{200}$; $10^{60}$ (being cited most often)
- **FCS 50 mln ca. $10^6$**



# *In silico mathematics*

- *Computer science (informatics)*

*…the science of algorithmic processing, representation, storage and transmission of information*

*Hromkovic, Theoretical Computer Science: Introduction to Automata, Computability, Complexity, Algorithmics Randomization, Communication and Cryptography*, Springer: Berlin, 2003

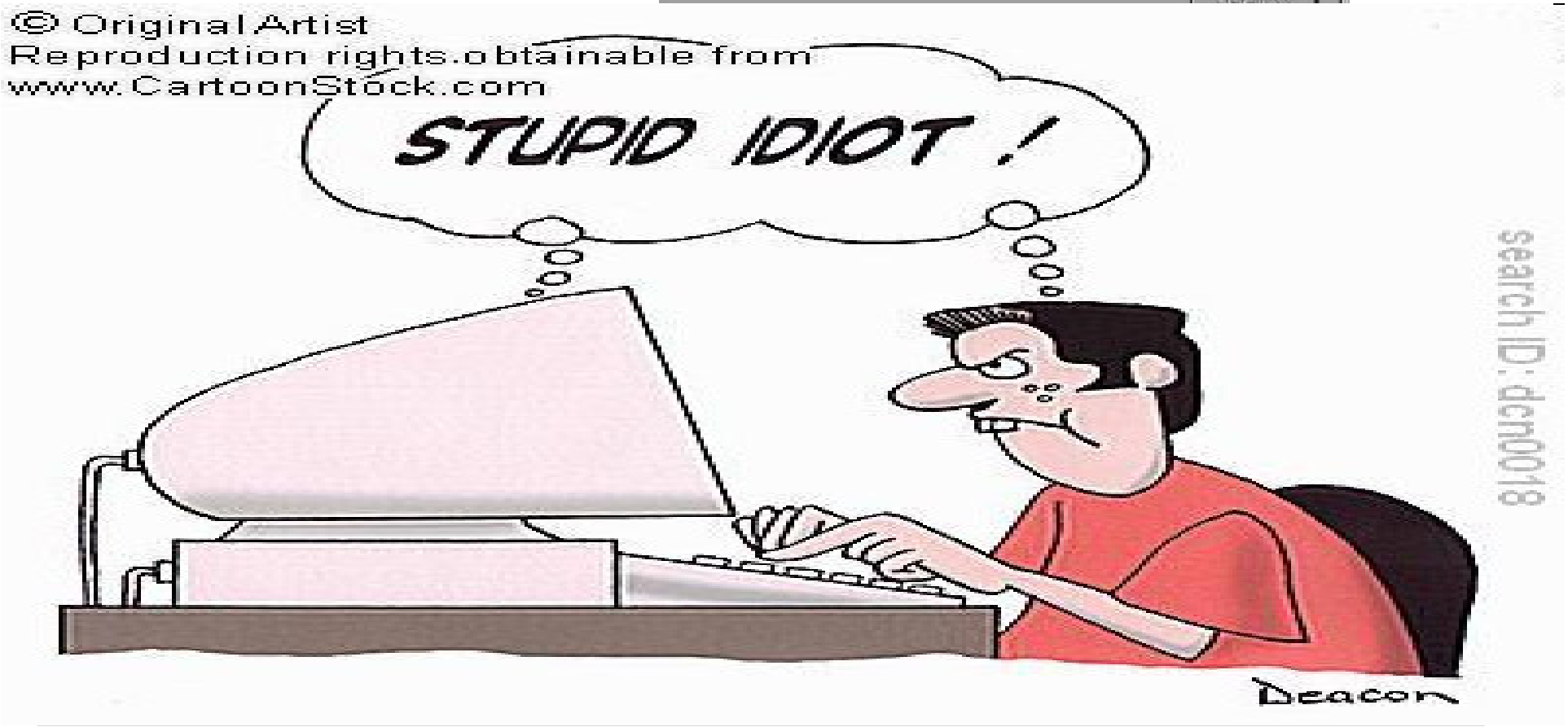
# *In silico* mathematics

- enormous speed and competence in low-level manipulations instead of human intuition

- this allowed to solve "formerly intractable problems and explore areas beyond the reach of human calculation"

# Teaching computers chemistry .... is not easy

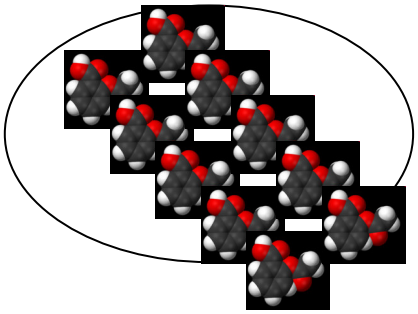
# Computers are not brilliant students

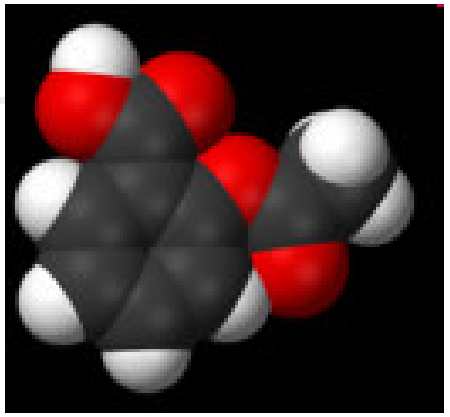
INSTYTUT CHEMII
UNIWERSYTETU ŚLĄSKIEGO

# Intuition (human) + computer = real efficiency

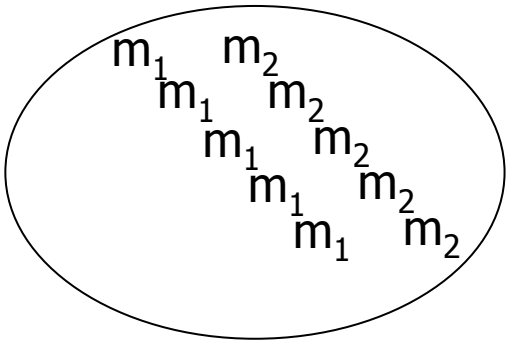
# Molecule vs. chemical compound



chemical compound



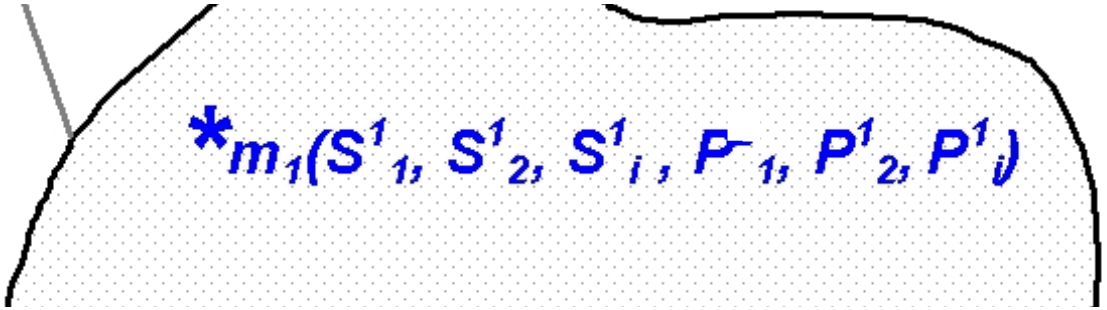
a single molecule




a mixture of chemical compounds

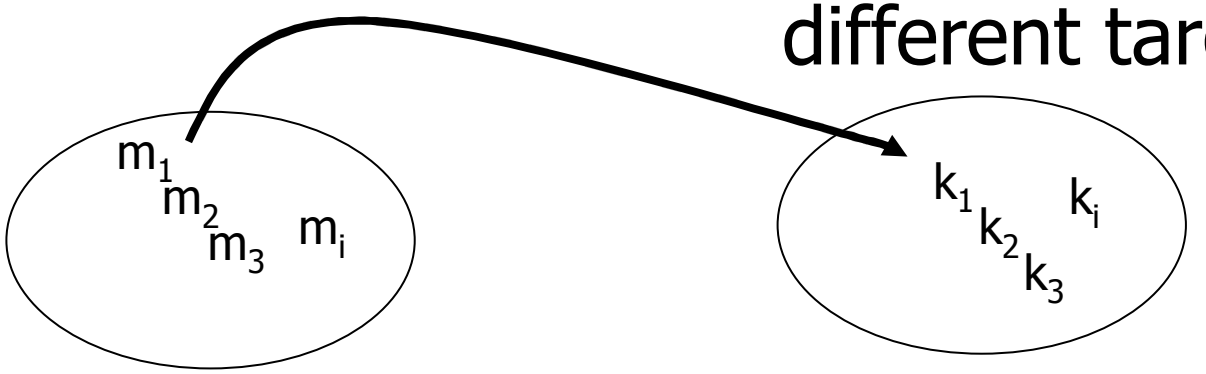
*graphics: wikipedia*

# Descriptors vs. properties




- descriptors (S) – can be *calculated* in virtual operations [mainly molecules]
- properties (P) – can only be measured by experimental approaches [mainly chemical compounds]

# Synteza = odwzorowanie reagent – produkt (in silico; in vitro)




different reagent molecules

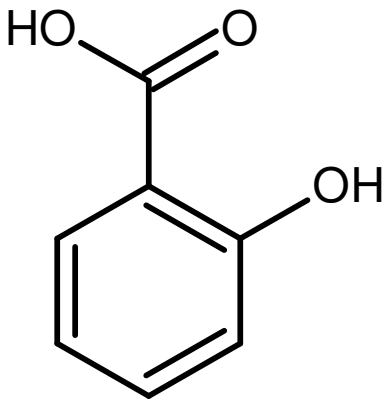
$$\bigwedge_{m \in (FCS)_r} \bigvee_{k \in (CS)_r} m \longrightarrow k$$

r-reaction domain

molecular operation

Polanski, Chemoinformatics w: Comprehensive Chemometrics, Elsevier 2009

# Molecule - still a mess



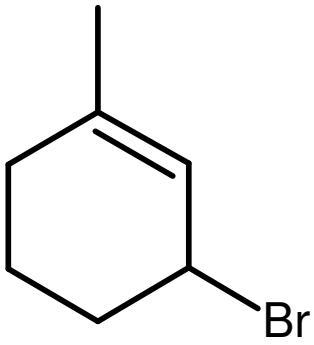

Nazwa systematyczna

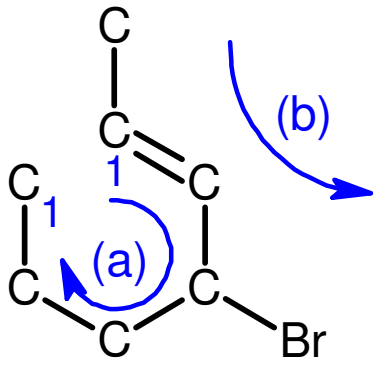
- **IUPAC**: kwas 2-hydroksybenzoesowy
- **SLN**: OHC[1]:CH:CH:CH:CH:C(:@1)C(=O)OH
- **SMILES**: Oc1ccccc1C(=O)O

cycloheksatrien  but hydroxybenzoic acid

# SMILES coding by ASCII

***s**implified **m**olecular*
***i**nput **l**ine **e**ntry **s**ystem*



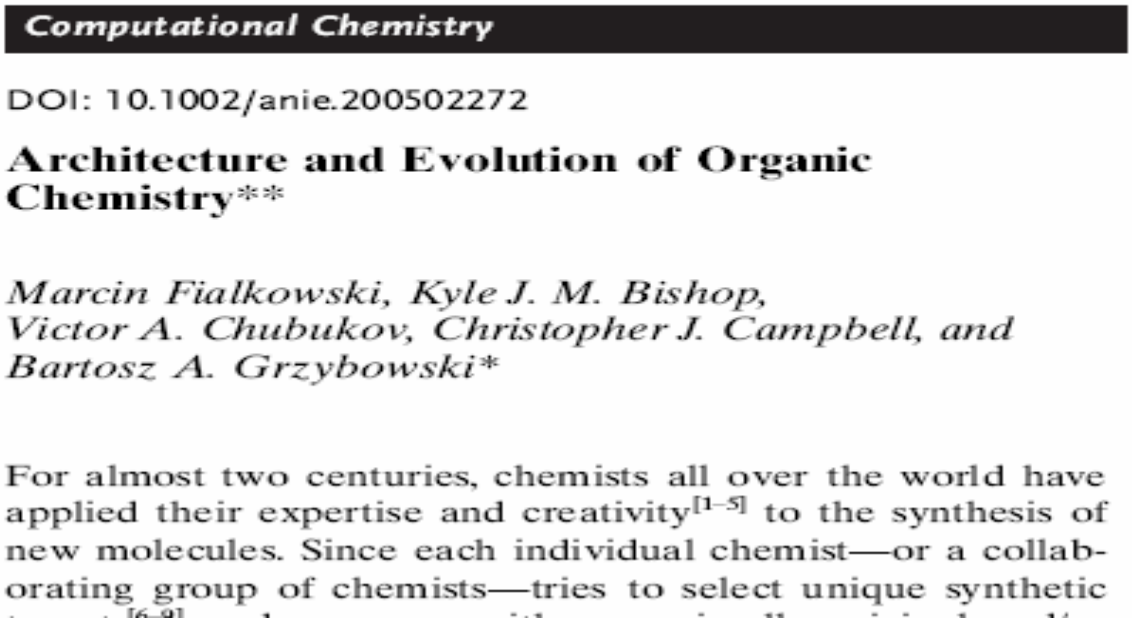




(a) CC1=CC(Br)CCC1

(b) CC1=CC(CCC1)Br

SMILES coding molecules and chemical reaction by *ASCII*.

# Architektura chemii – badanie baz związków chemicznych



Computational Chemistry

DOI: 10.1002/anie.200502272

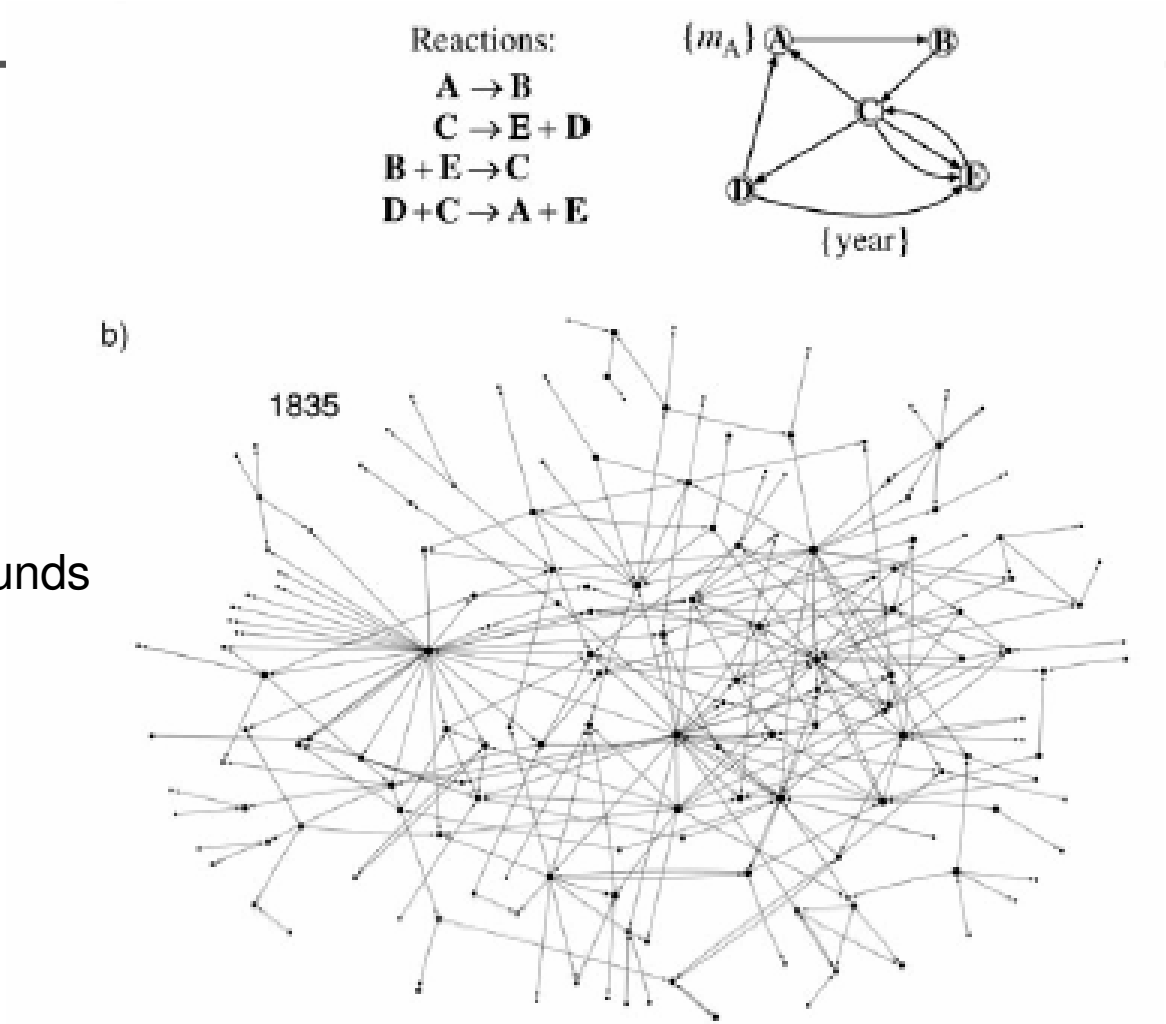
**Architecture and Evolution of Organic Chemistry****

*Marcin Fialkowski, Kyle J. M. Bishop, Victor A. Chubukov, Christopher J. Campbell, and Bartosz A. Grzybowski**

For almost two centuries, chemists all over the world have applied their expertise and creativity[1–5] to the synthesis of new molecules. Since each individual chemist—or a collaborating group of chemists—tries to select unique synthetic

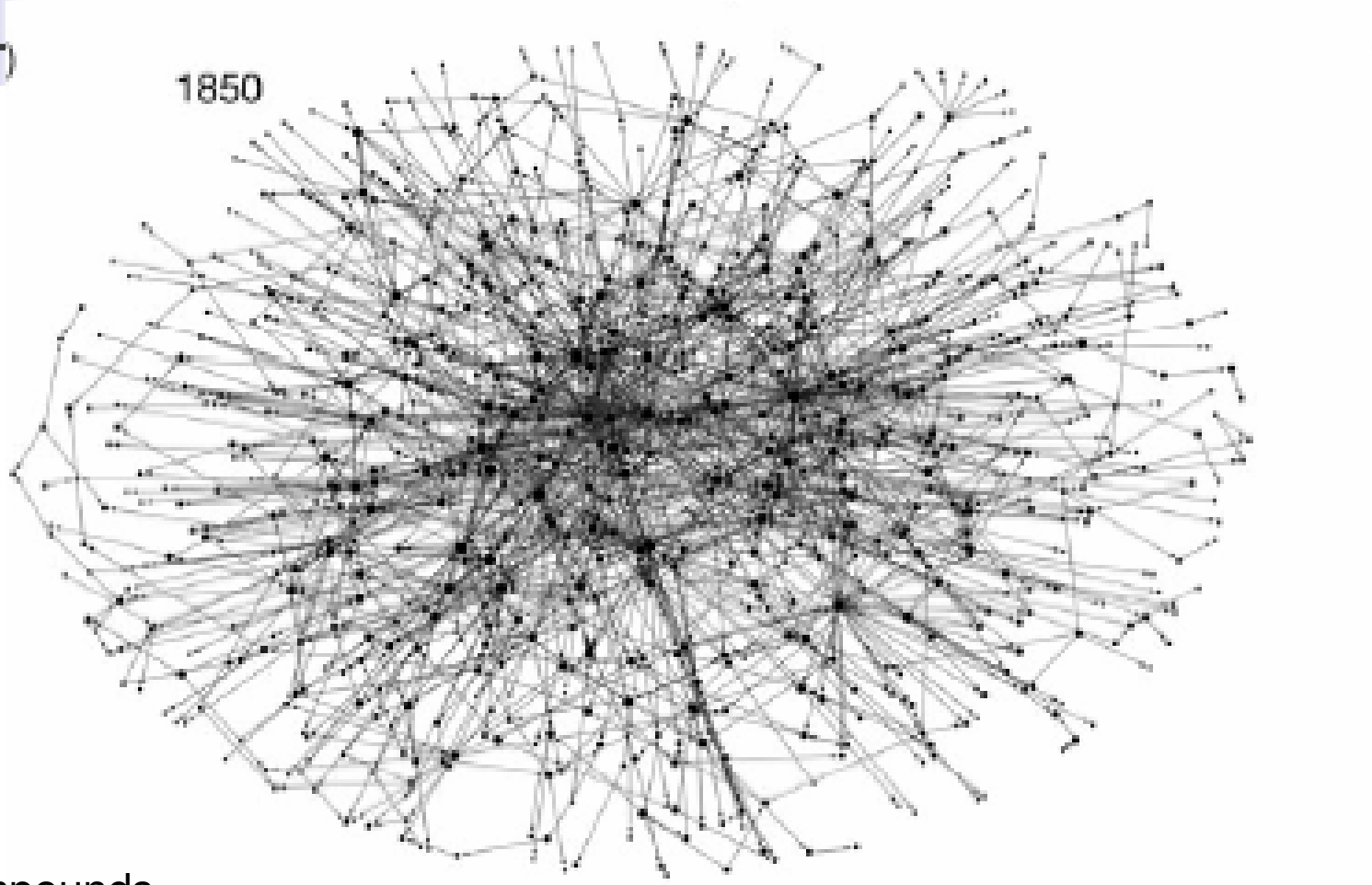
Grzybowski et al., Angew. Chem. Int. Ed. 44, 7263 –7269 (2005);
Grzybowski et al., Nature Chemistry 1, 31 - 36 (2009)

# Graf reakcji chemicznych



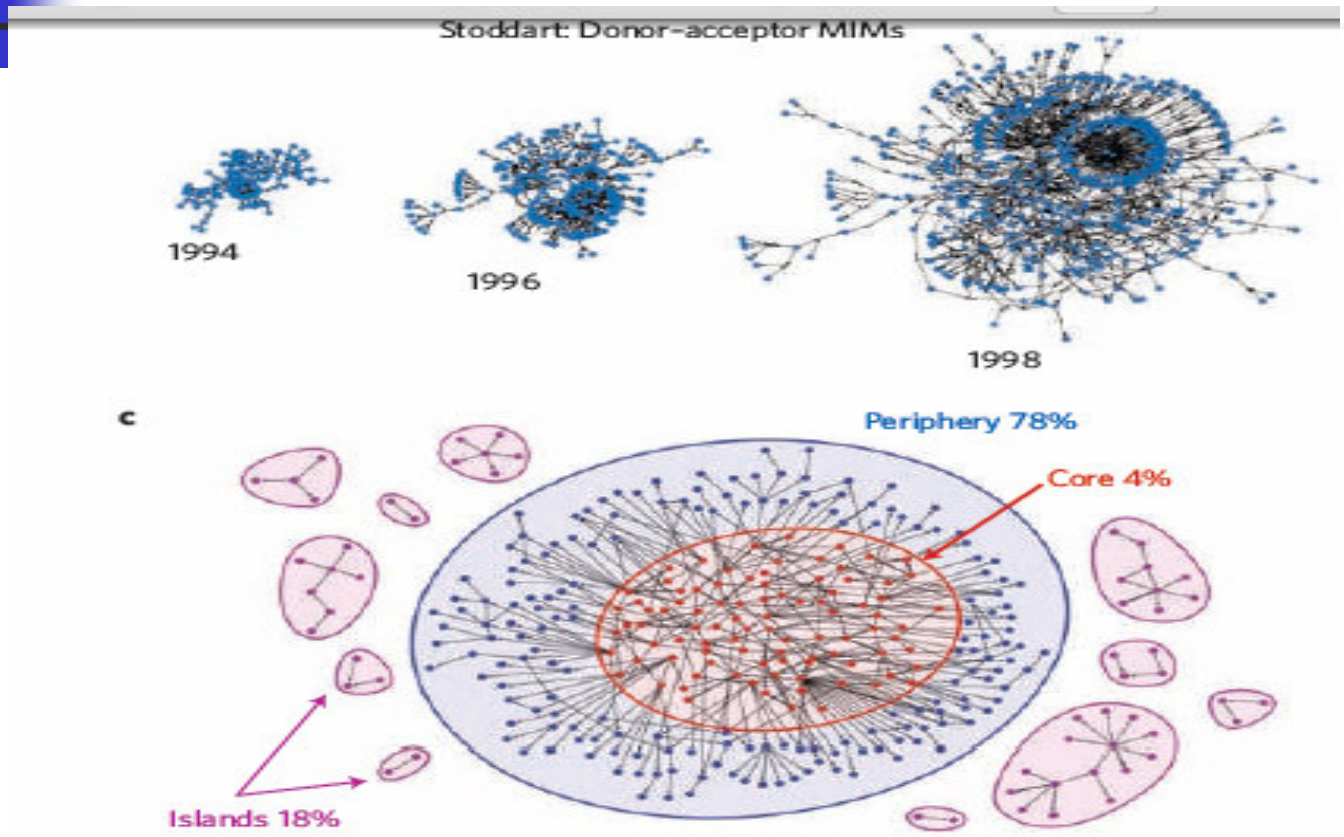

176 compounds

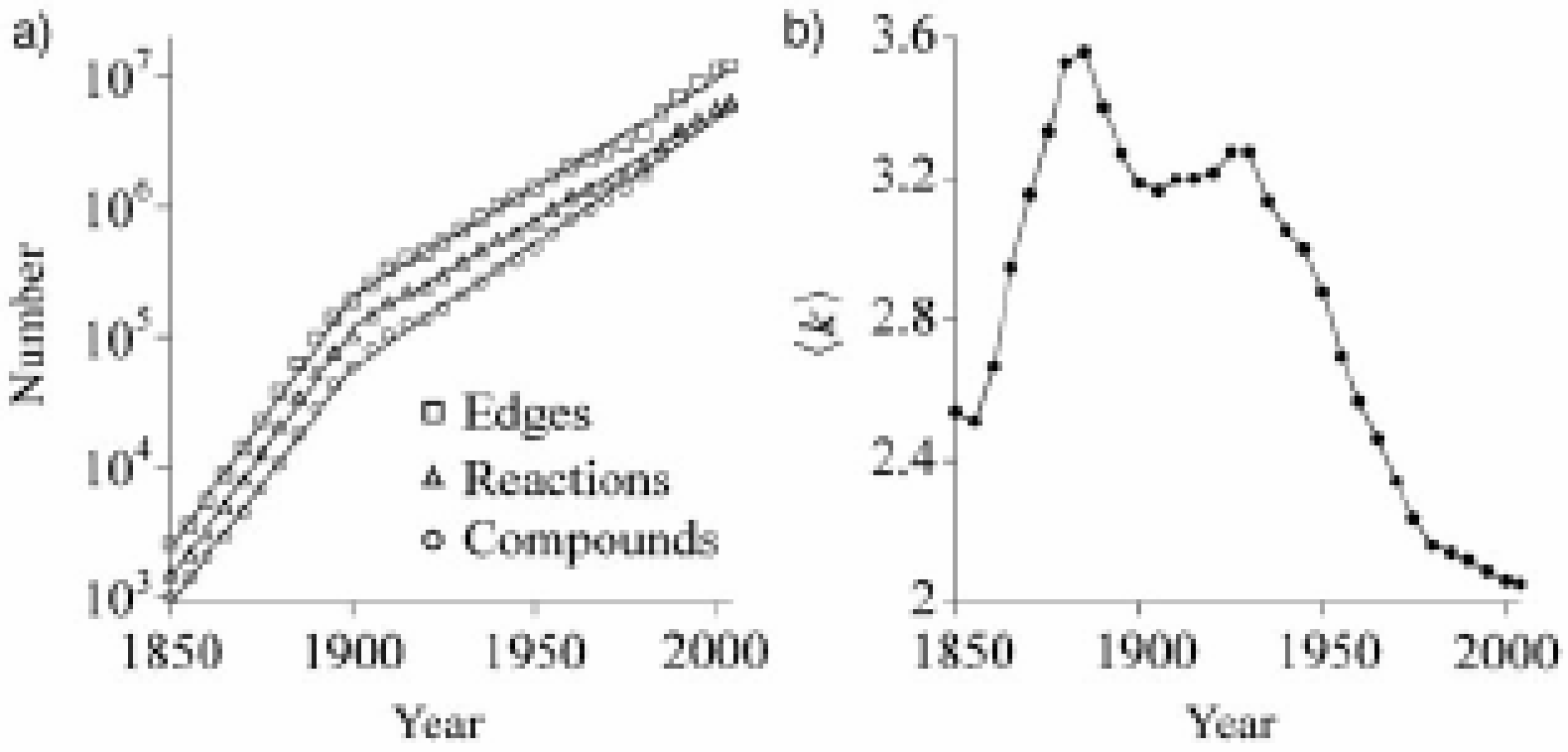
# Graf reakcji chemicznych




876 compounds

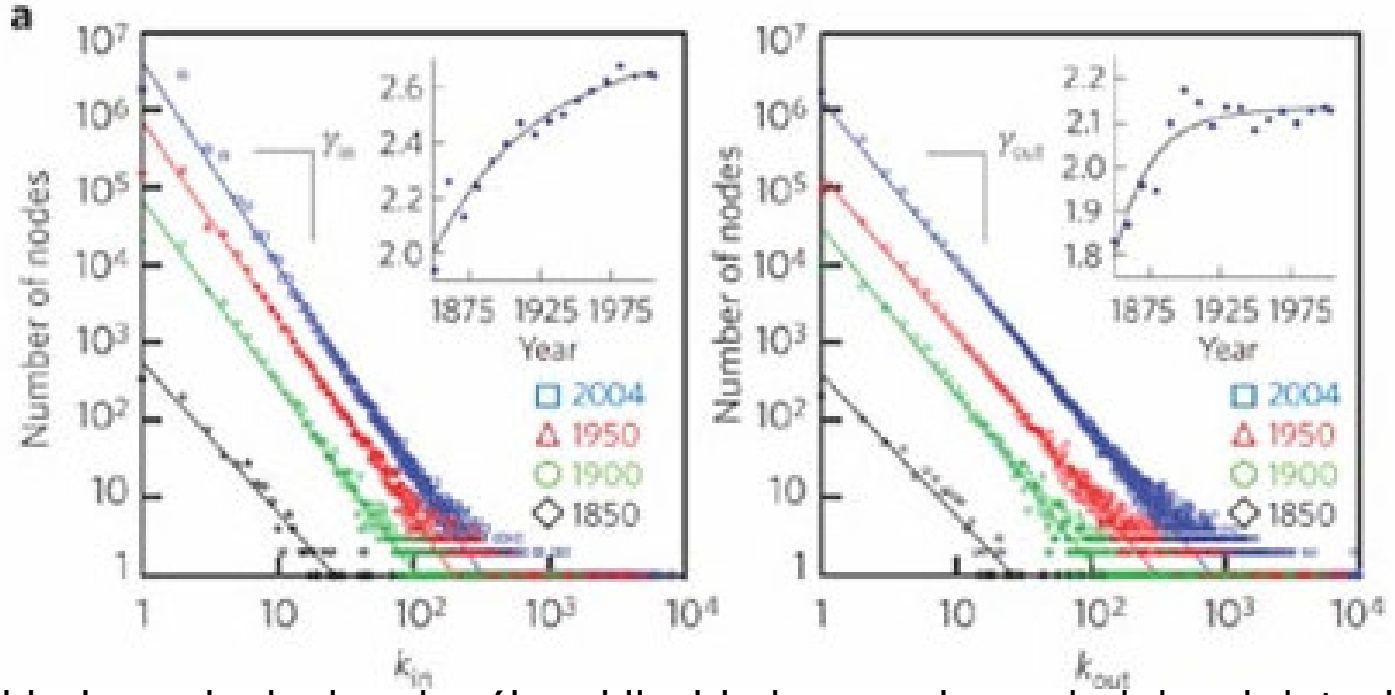
# Chemia


Stoddart: Donor–acceptor MIMs
1994
1996
1998
c
Periphery 78%
Core 4%
Islands 18%




K = compounds/reactions

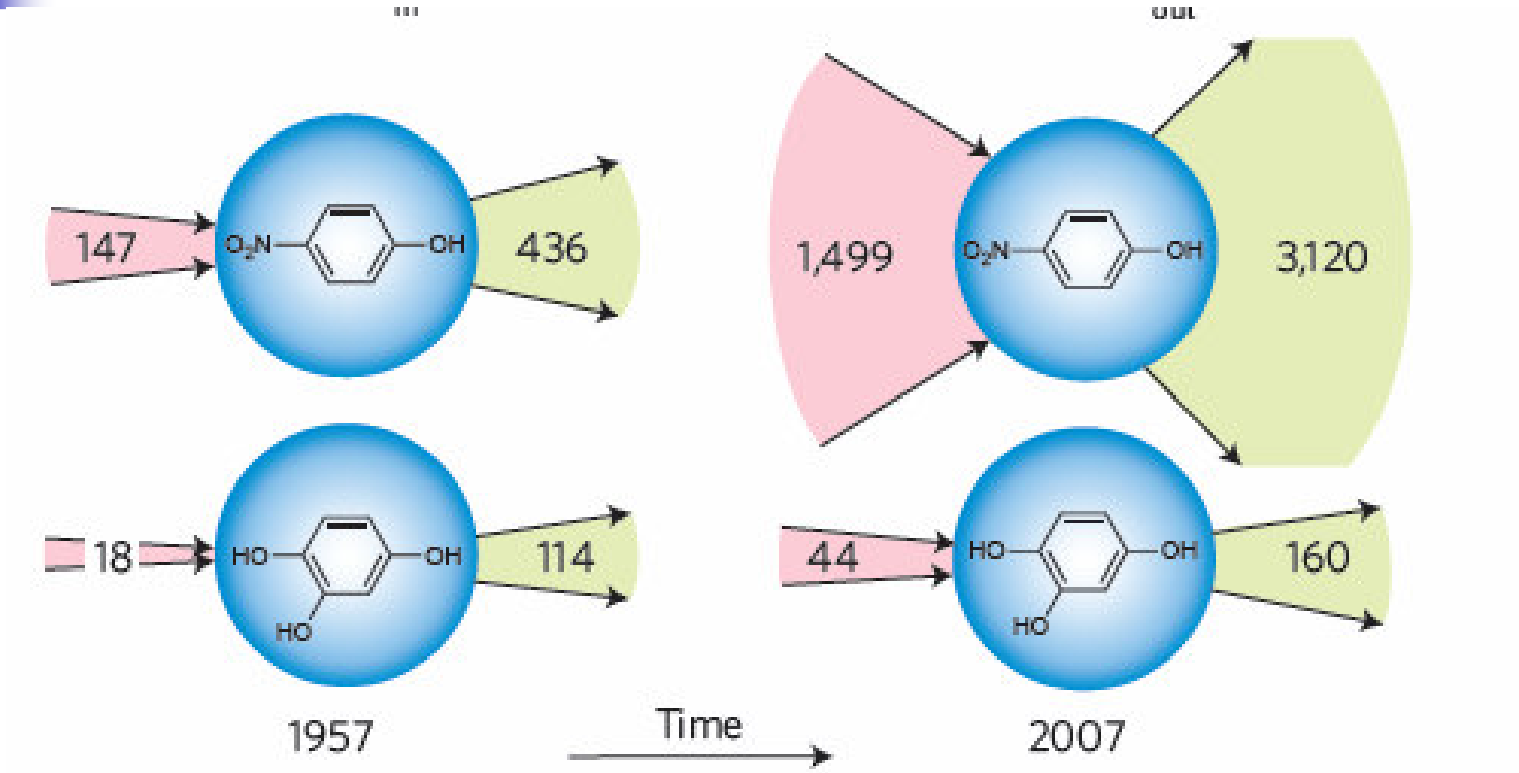
# Architektura chemii



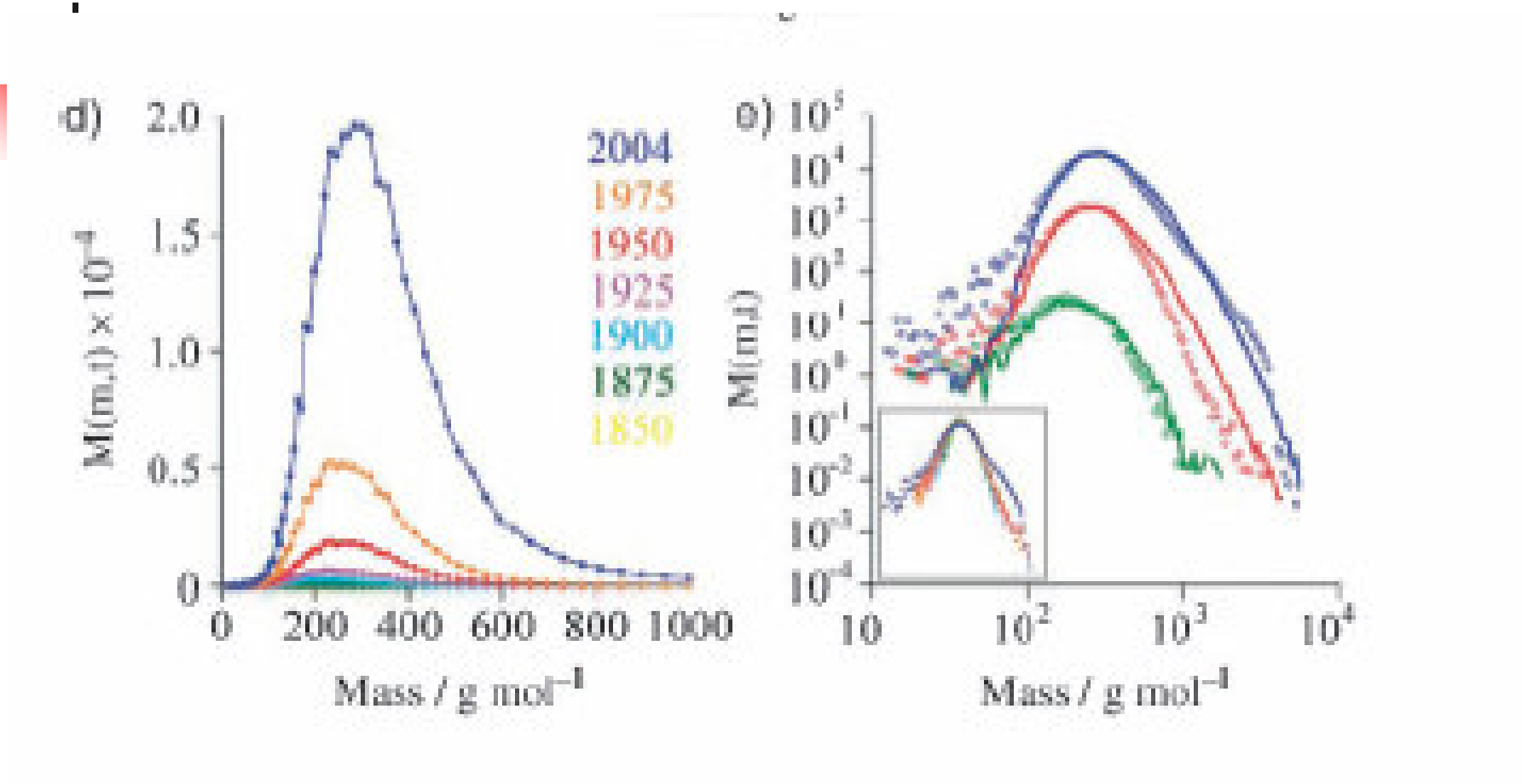

rozkład oczek sieci o określonej liczbie $k_{in}$ oraz $k_{out}$ w kolejnych latach

Grzybowski et al., Nature Chemistry 1, 31 - 36 (2009)

# Preferential attachment

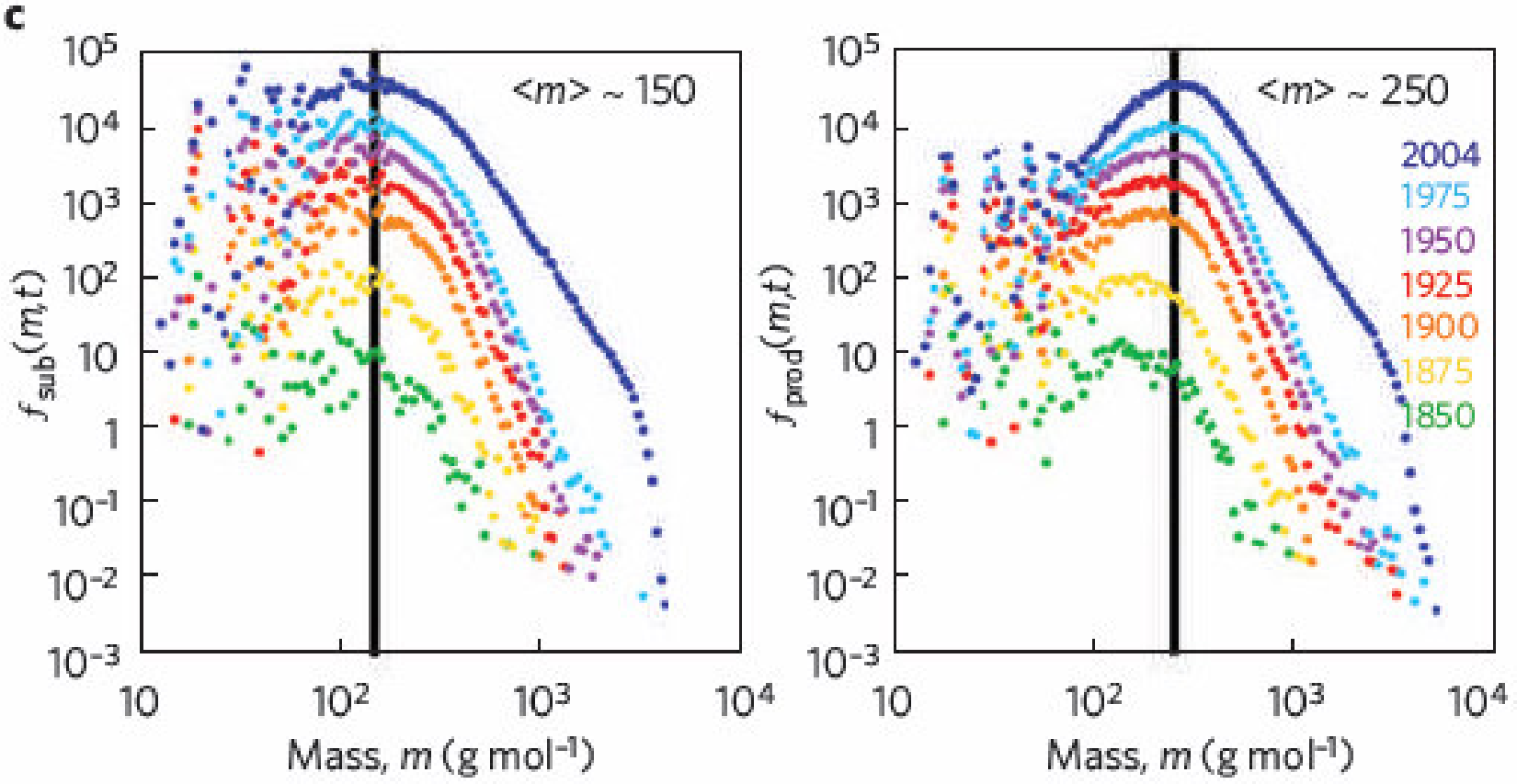



Grzybowski et al., Nature Chemistry 1, 31 - 36 (2009)

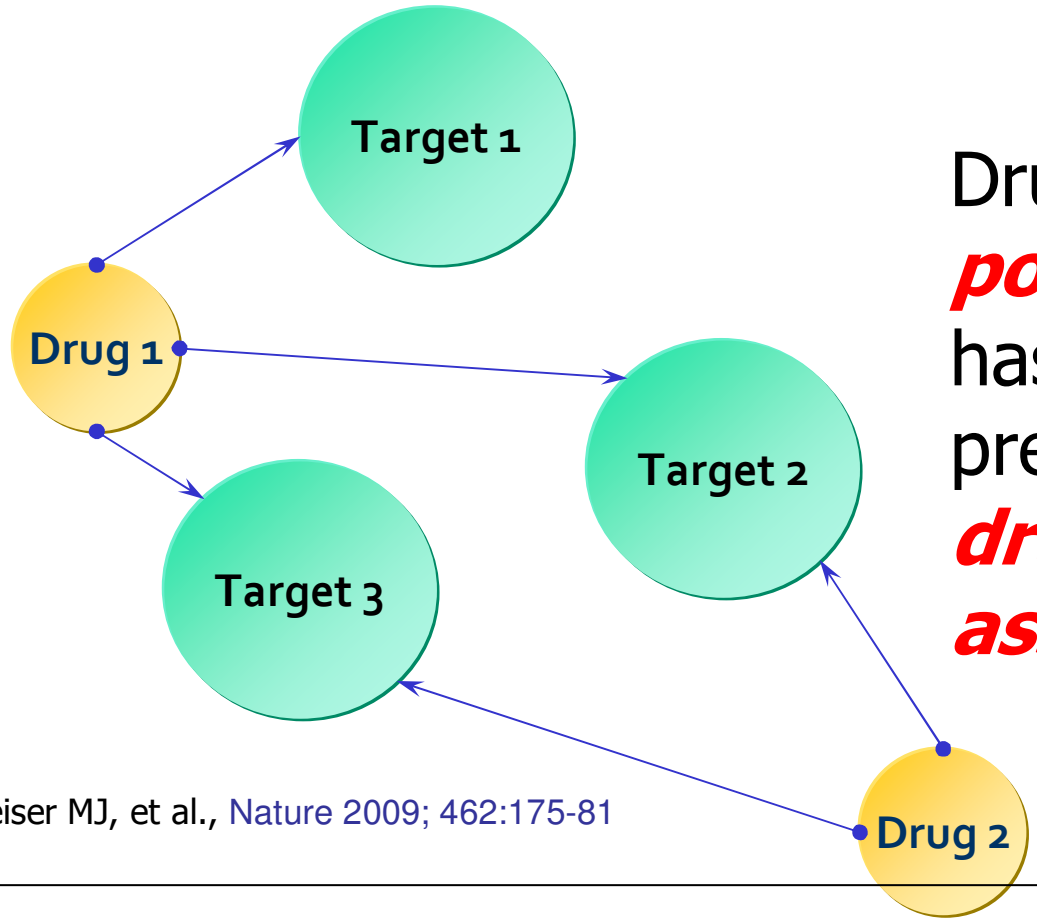



Rozkład masy związków oryginalny oraz po przeskalowaniu
Średnia masa rośnie od 200 (1850) do 350 (2004)

# Masy $k_{in}$ i $k_{out}$



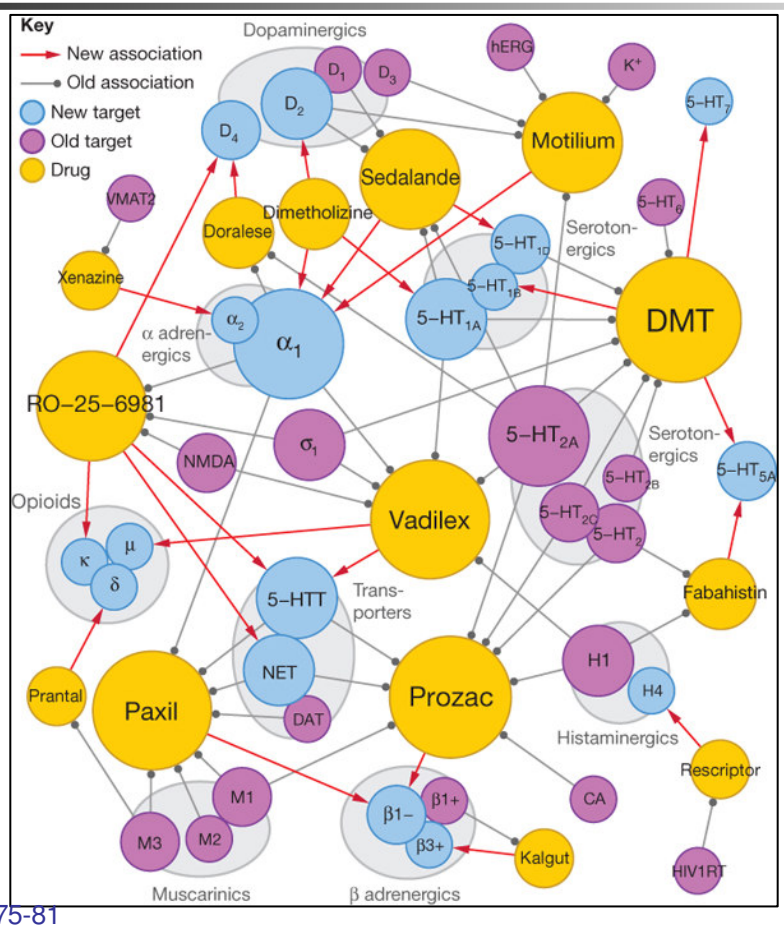
# Global mapping of pharmacological space




Keiser MJ, et al., Nature 2009; 462:175-81

Drug ***polypharmacology*** has inspired efforts to predict and characterize ***drug-target associations***

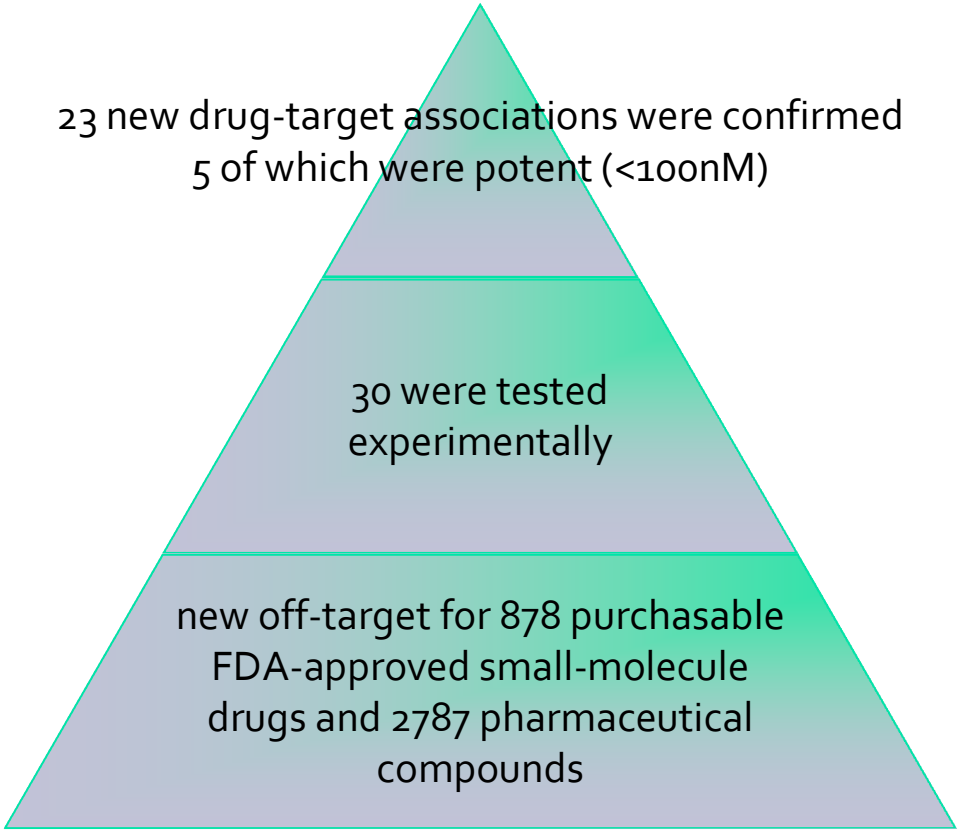
# Discovered off-targets network



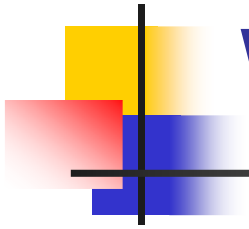

Nature 2009; 462:175-81

- Bipartite network where drugs (gold) are linked by grey edges to their known targets (violet) and by red arrows to their discovered off-targets (cyan).
- Node sizes increase with number of incident edges.

# Results




Nature 2009; 462:175-81



# Wnioski

- Lepszy opis FCS pozwala na lepszą chemię (CS: VCS);
- konieczne nowe metody badania i opisu od intuicyjnej architektury do matematycznej topologii
- miejsce dla matematyki

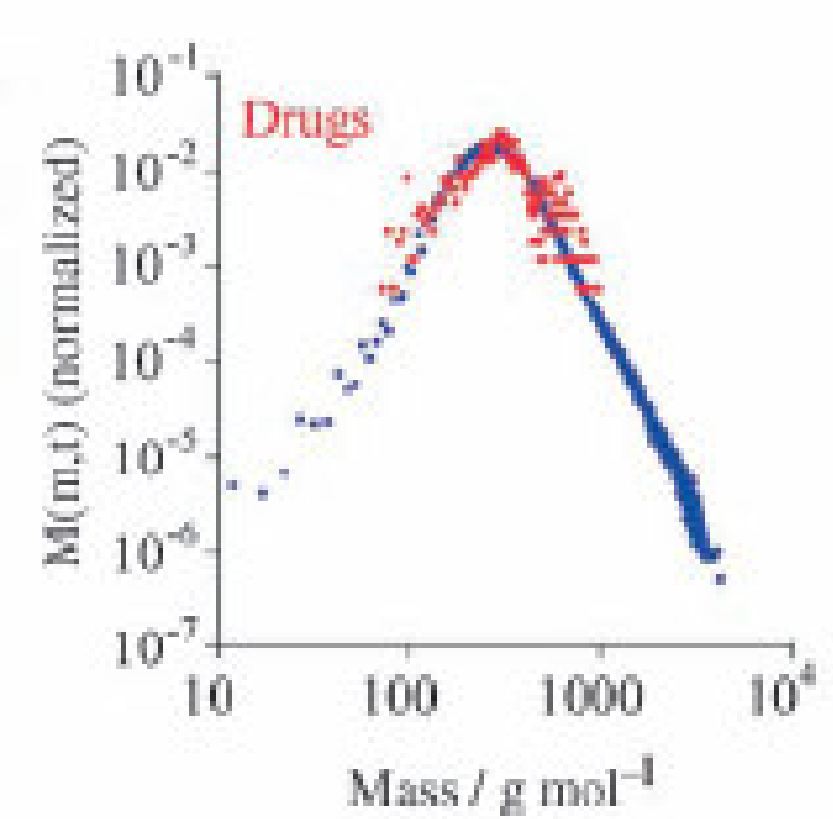
# Architektura leków - problemy

- Projektowanie nowych właściwości
  Paradoks: związki nie są interesujące, interesują nas właściwości
- Np. leki / nie_leki [zbiory; przestrzenie; otocznia; granice; etc]
- Uprzywilejowane struktury leków
- Polifarmakologia (uprzywilejowane układy lek-cel działania leku)
- .....

# leki vs. nie_leki – RO5

- Lek
  - nie więcej niż 5 HBD
  - nie więcej niż 10 HBA
  - masa poniżej 500 Daltonów
  - wsp. P (podział woda-oktanol) poniżej 5

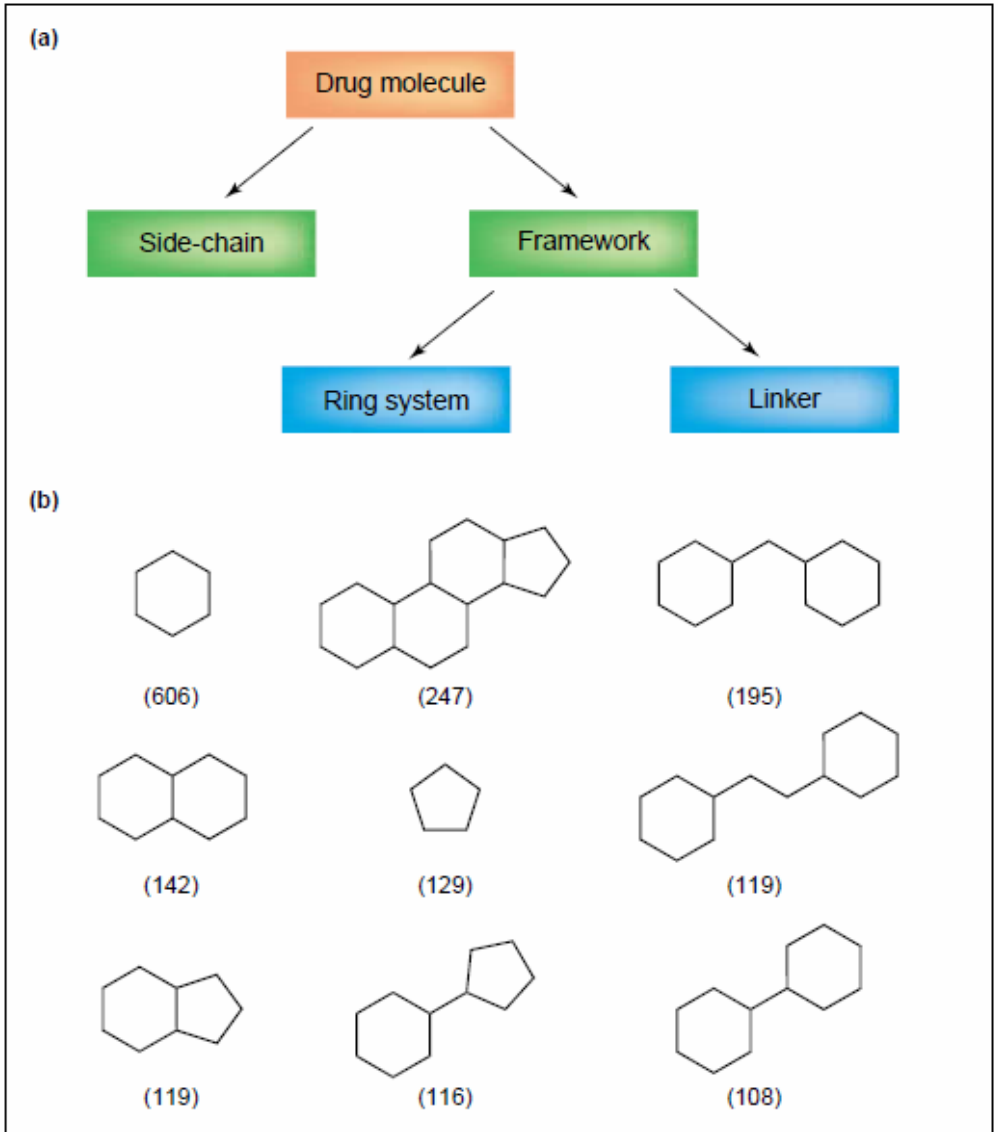
# Leki – „Rule of five”




Grzybowski et al., Angew. Chem. Int. Ed. 2005, 44, 7263 –7269 ;

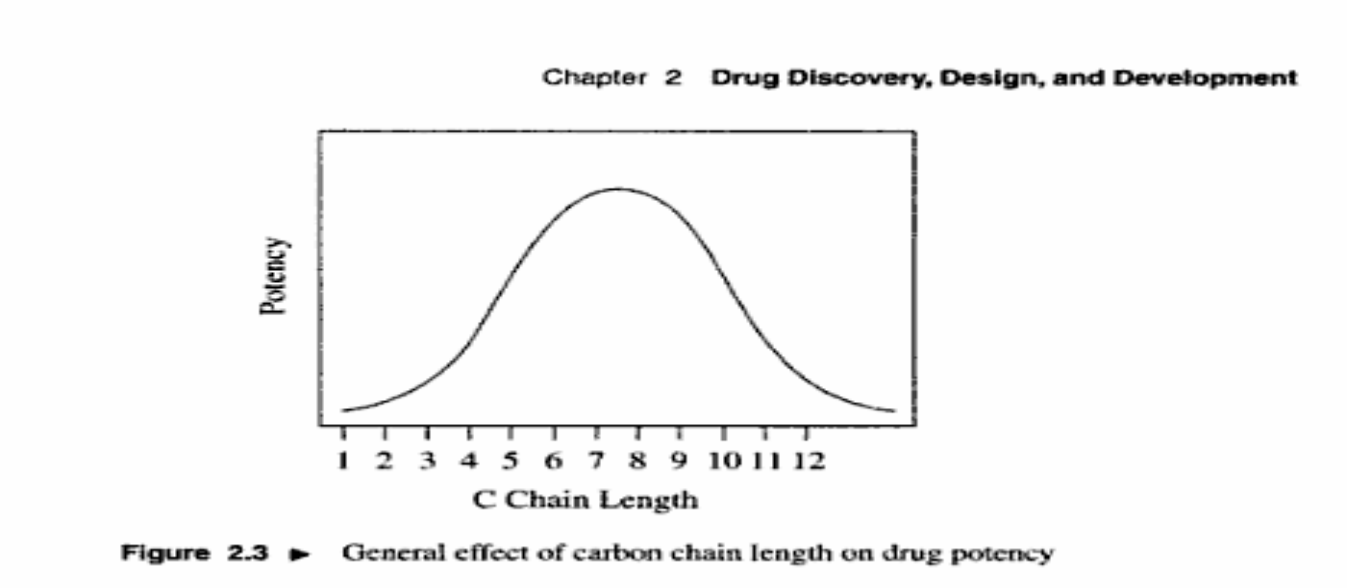
# Privileged structures

Database
*Comprehensive Medicinal Chemistry*



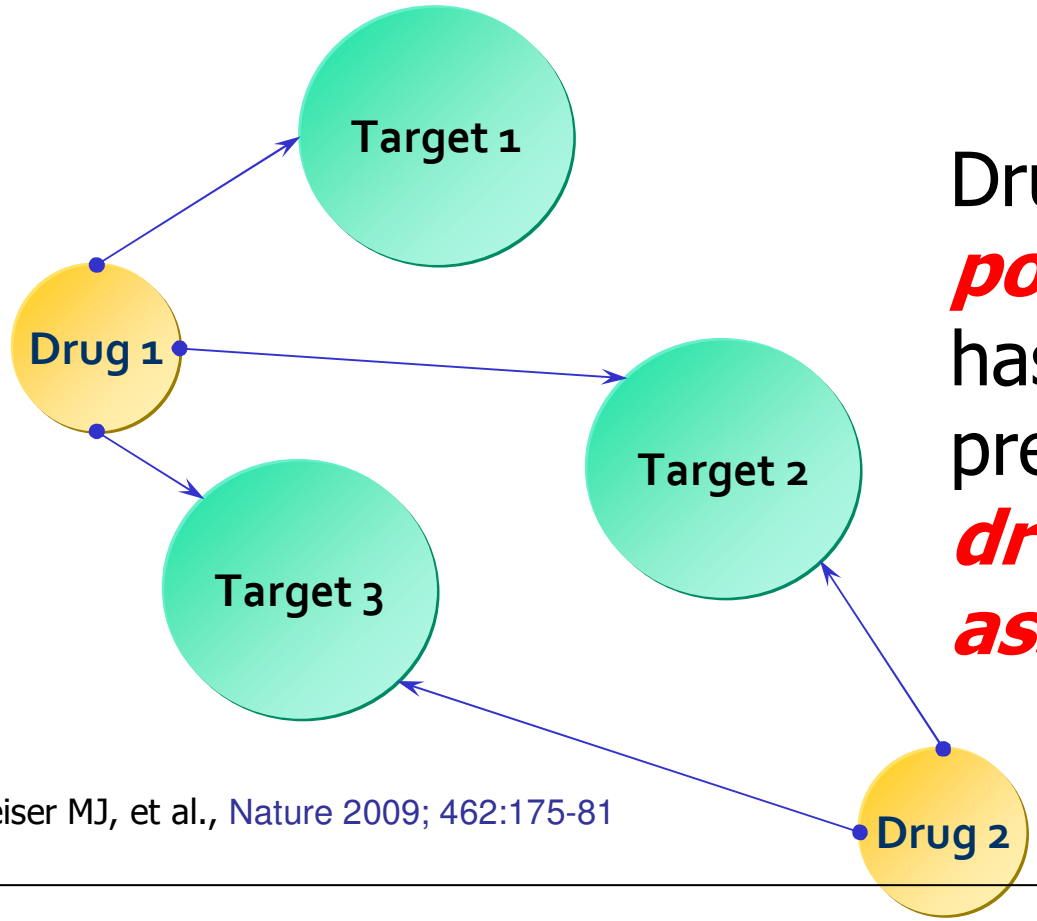

Fattori, D. D. Molecular Recognition:
The Fragment Approach in Lead Generation. Drug Discov. Today 2004, 9, 229–238.

# Drug architecture




Richard B. Silverman, The organic chemistry of drug design and drug action, Elsevier, 2004

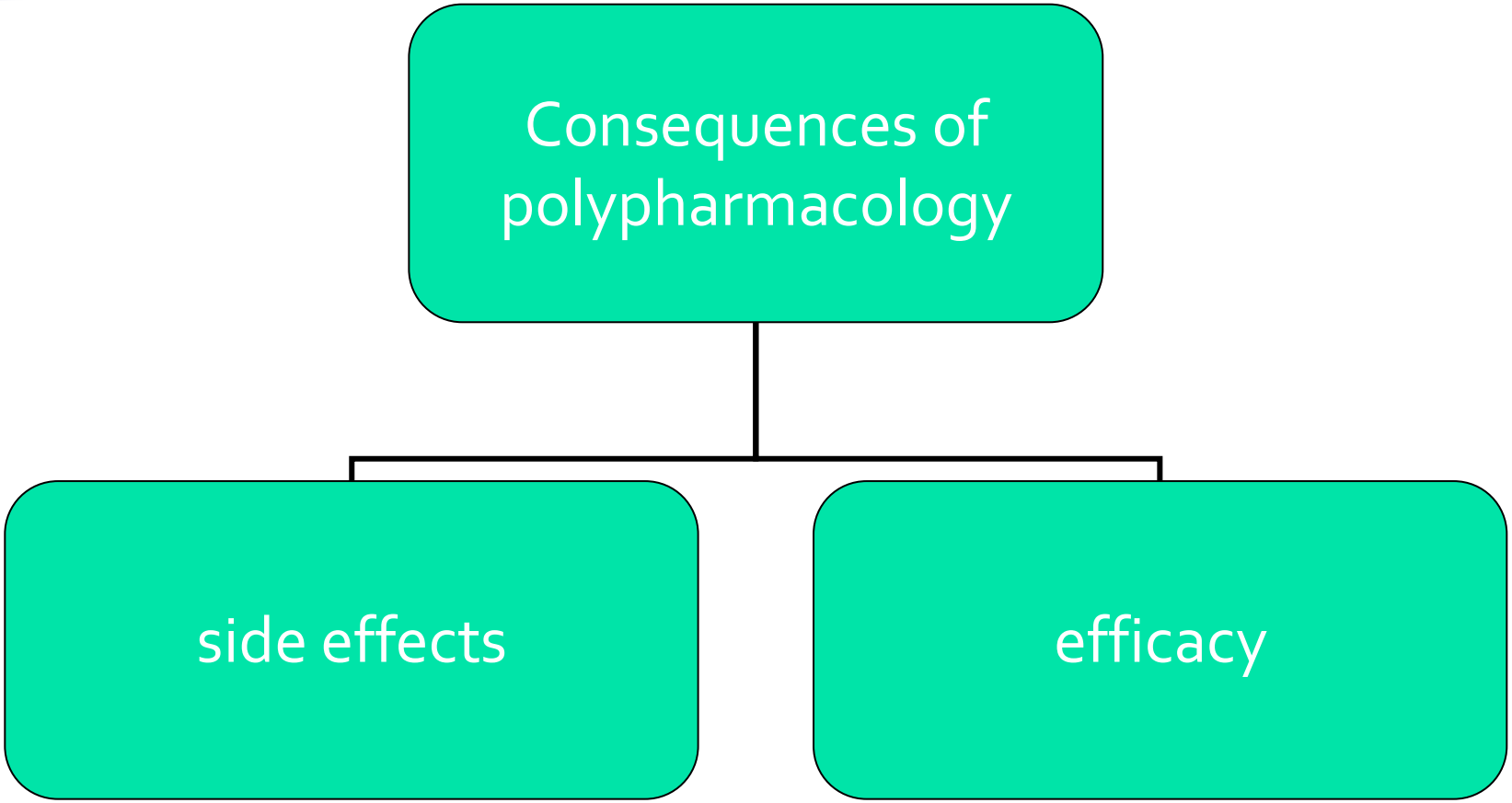
# Global mapping of pharmacological space



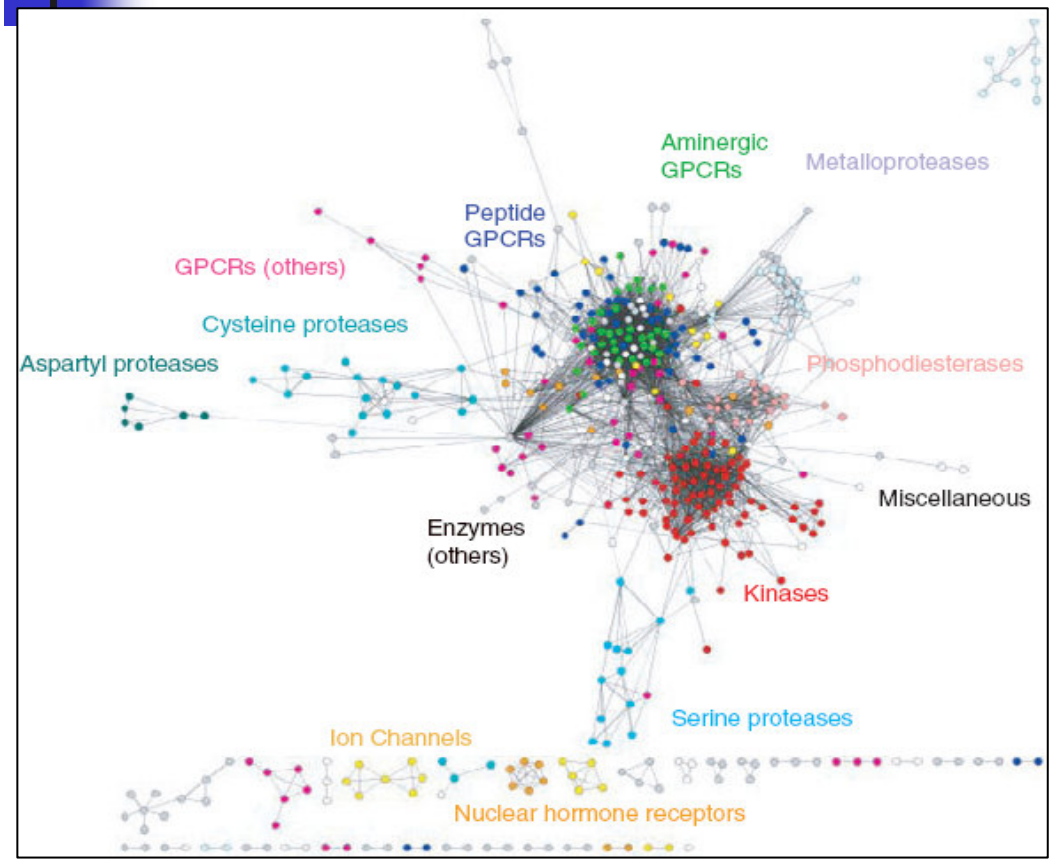

Keiser MJ, et al., Nature 2009; 462:175-81

Drug ***polypharmacology*** has inspired efforts to predict and characterize ***drug-target associations***

# Polypharmacology



# Relationships between proteins in chemical space




- *Human polypharmacology interaction network*
- Two proteins are deemed interacting (joined by an edge) if both bind one or more compound within a defined difference in binding energy threshold
- Nodes are colored by gene family

Paolini GV,et al., Nat. Biotechnol. 2006; 24:805-15

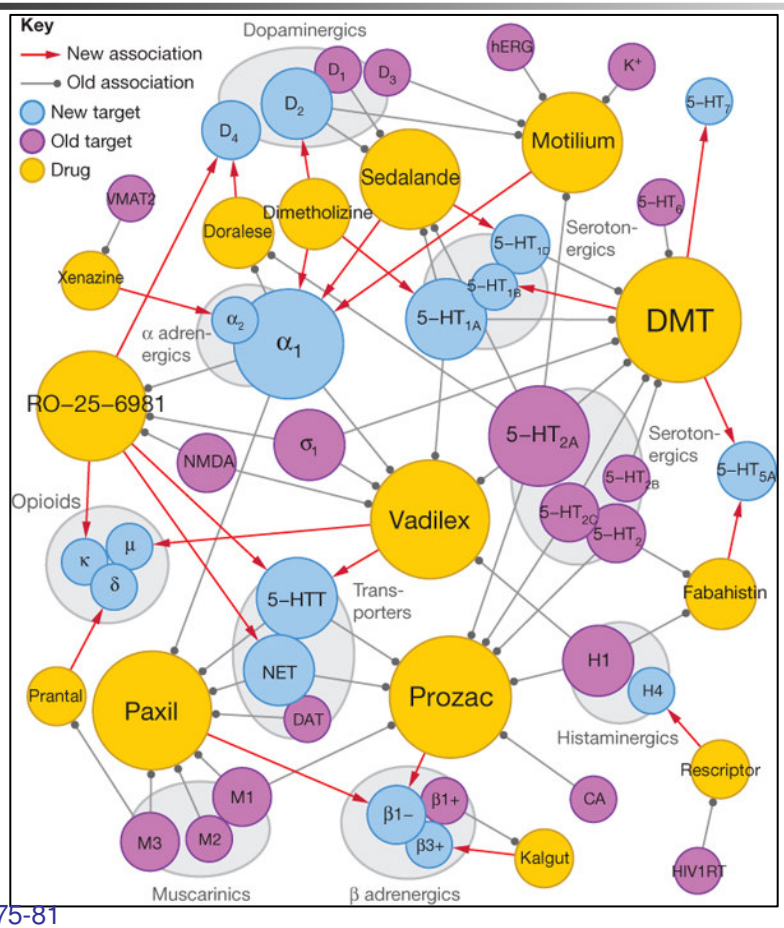
# How to predict drug polypharmacology ?

- A collection of 3665 FDA-approved and investigational drug structures was computationally screened against a panel of more than 1400 protein targets. The drug collection was extracted from the MDL Comprehensive Medicinal Chemistry database. Each target was represent solely by its set of known ligands, which were extracted from three sources: the MDL Drug Data Report, the WOMBAT, and the StARlite databases. The ***two-dimensional structural similarity of each drug to each target's ligand set was quantified as an E-value using the SEA***.

- SEA, similarity ensemble approach
  ***compares targets by the similarity of the ligands that bind to them***, expressed as expectations values (E-values), adapting the BLAST algorithm.

Michael J. Keiser et al., Nature 2009; 462:175-81

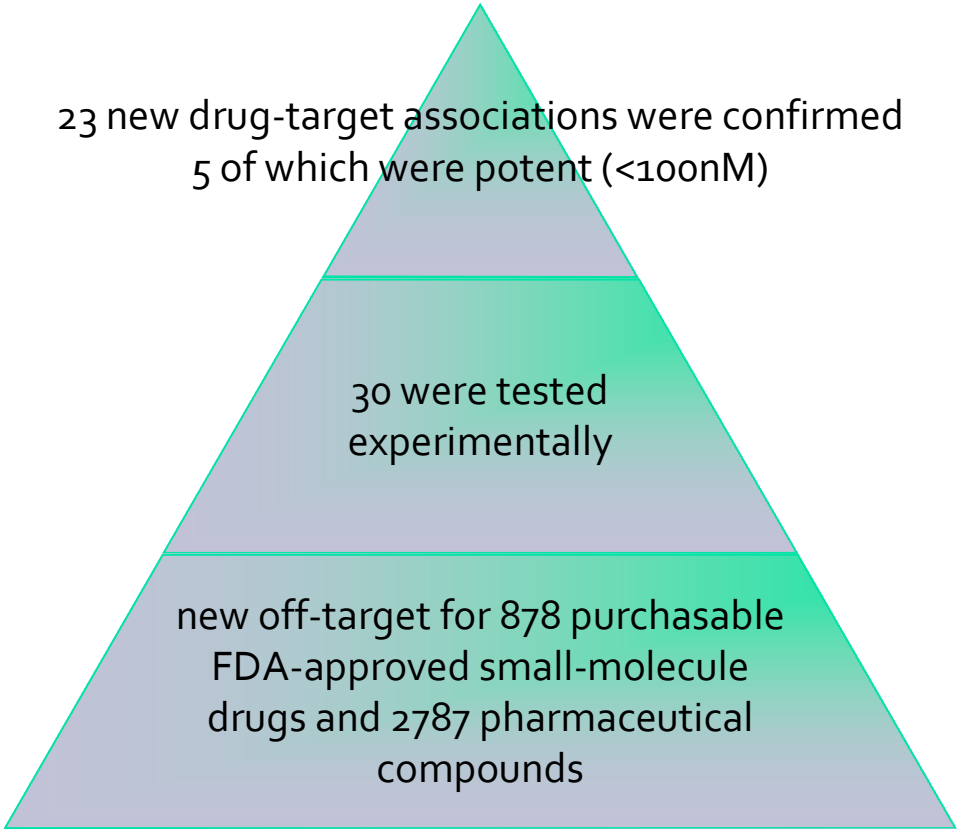
# Discovered off-targets network




Nature 2009; 462:175-81

- Bipartite network where drugs (gold) are linked by grey edges to their known targets (violet) and by red arrows to their discovered off-targets (cyan).
- Node sizes increase with number of incident edges.

# Results




Nature 2009; 462:175-81

# Quinoline

antineoplastics activity

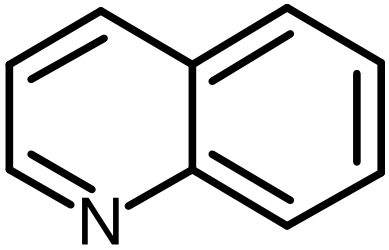
antifungals

antiprotozoic drugs

antituberculotic agents

quinine

antibacterials




- Quinoline scaffold can be found in many classes of biologically active compounds used as
  - antifungals
  - antibacterials
  - antiprotozoic drugs
  - antituberculotic agents
- Some quinoline analogues showed also antineoplastics activity
- Quinine molecule - Nature preference for the system?

Polanski, J., et al., Quinoline-Based Antifungals, Curr. Med. Chem., 2010, 1960-1973

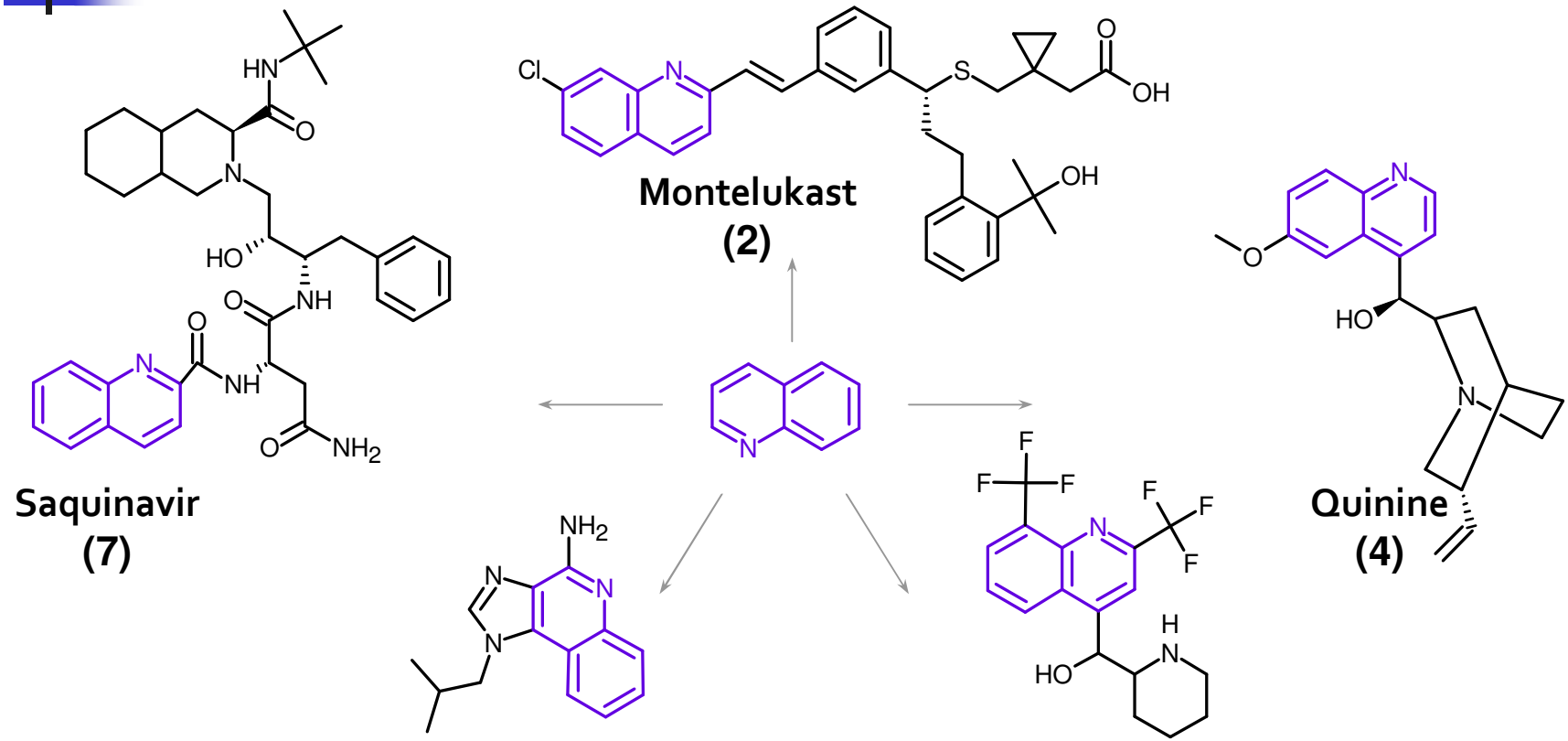
# Small molecule databases

| | |
|---|---|
| **General databases** | The Available Chemical Directory (ACD) |
| | Chemical Entities of Biological Interest |
| | ChemBank |
| | PubChem |
| **Medicinal agents** | The Comprehensive Medical Chemistry |
| | MDL Drug Data Report (MDDR) |
| | The Derwent World Drug Index (WDI) |
| | The National Cancer Institute database |
| | TheWOrld of Molecular BioActivities database (WOMBAT) |
| | DrugBank |
| **Physico-chemical properties** | CrossFire Beilstein |

# Database mining

- DrugBank is a unique bioinformatics and cheminformatics resource
- detailed drug data and drug target information

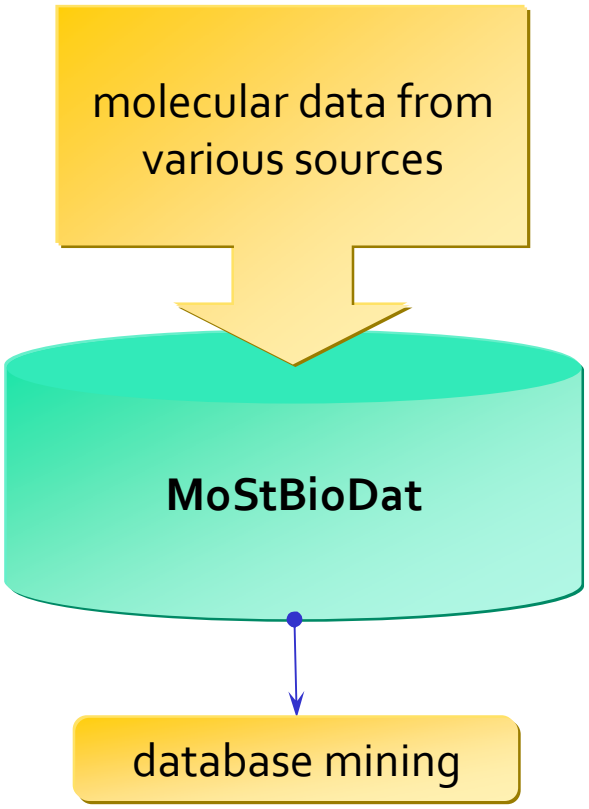
# Quinoline scaffold – a privileged structure for drug discovery






*The structures and data were extracted from DrugBank database*

# The Molecular and Structural Bioinformatics Database




The MoStBioDat platform is an integrated software system for storing data in a unified format with an ensemble of tools for data manipulation.

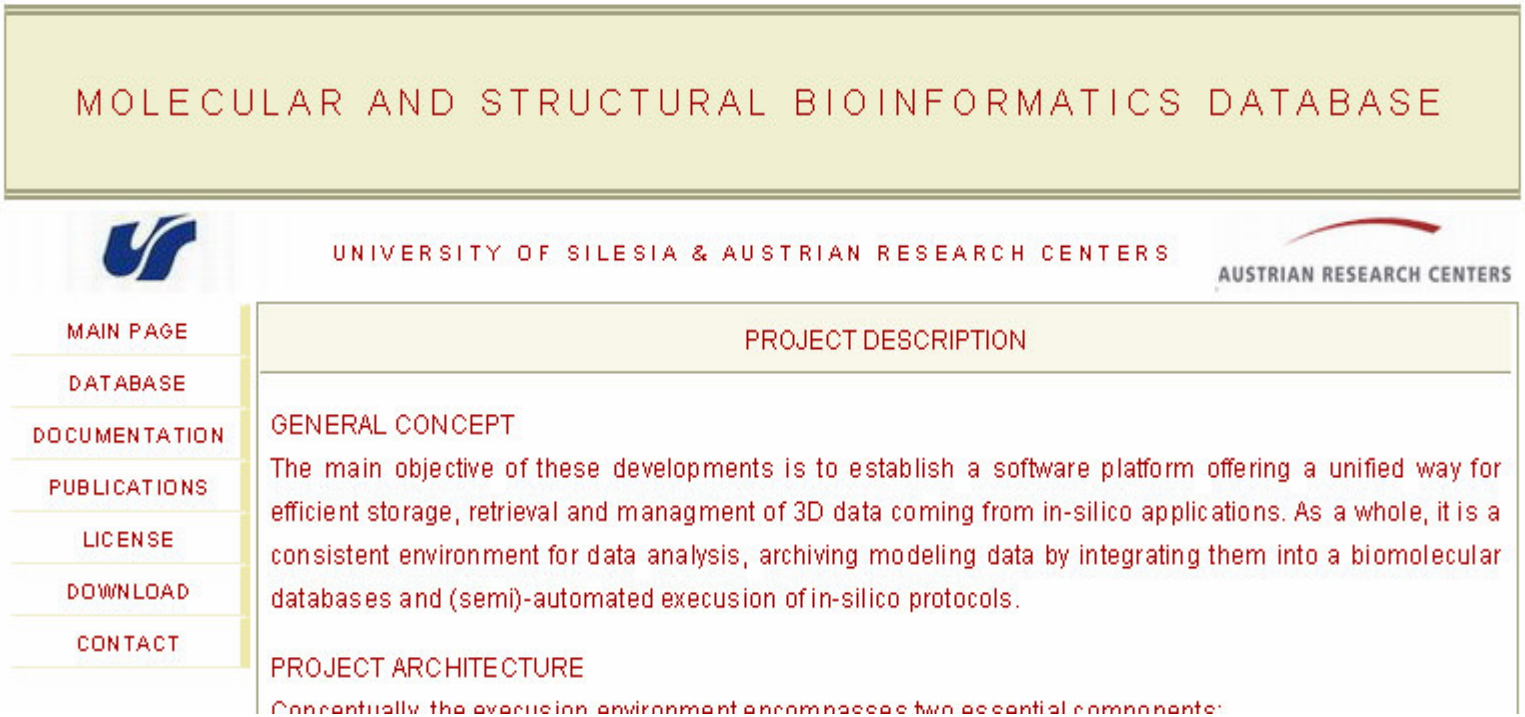
**Application**

- data management platform
- substructure search

**Advantages**

- possibility of being installed locally with a pretty simple database driven by the Python package installation procedure
- fully integrated data at a scientific level for data mining
- investigation of the large ligand and receptor data

# Mostbiodat



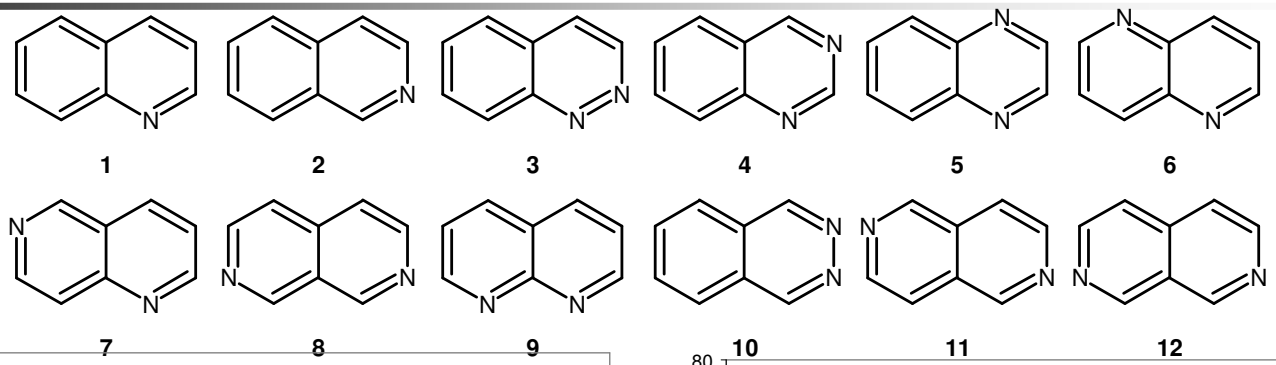

www.chemoinformatyka.us.edu.pl

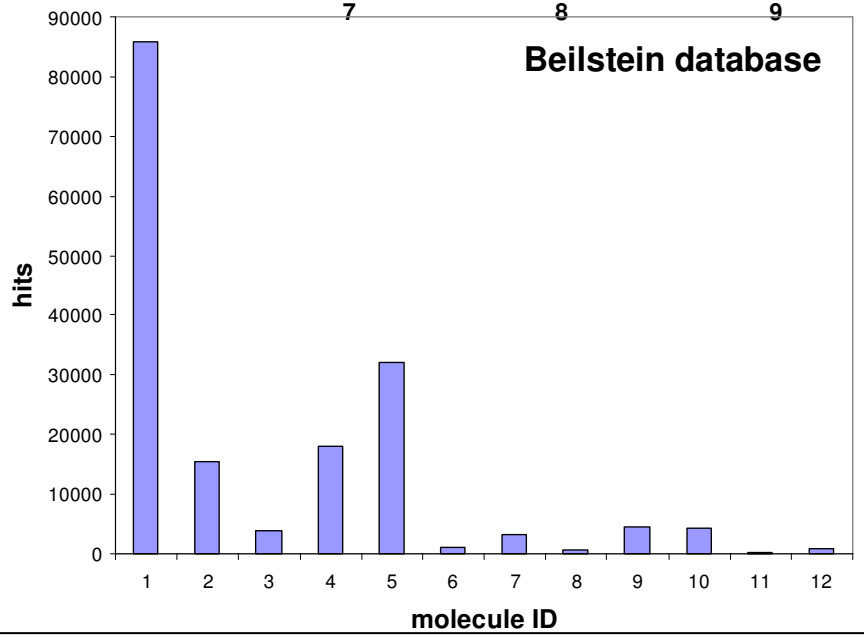
A. Bak, J. Polanski, T. Stockner, A. Kurczyk, MoStBioDat – Molecular and Structural Bioinformatics Database, *Comb. Chem. High Throughput Screen.,* 13, 2010, pp. 366-374; *Molecules,* 14(9), 3436-3445, 2009

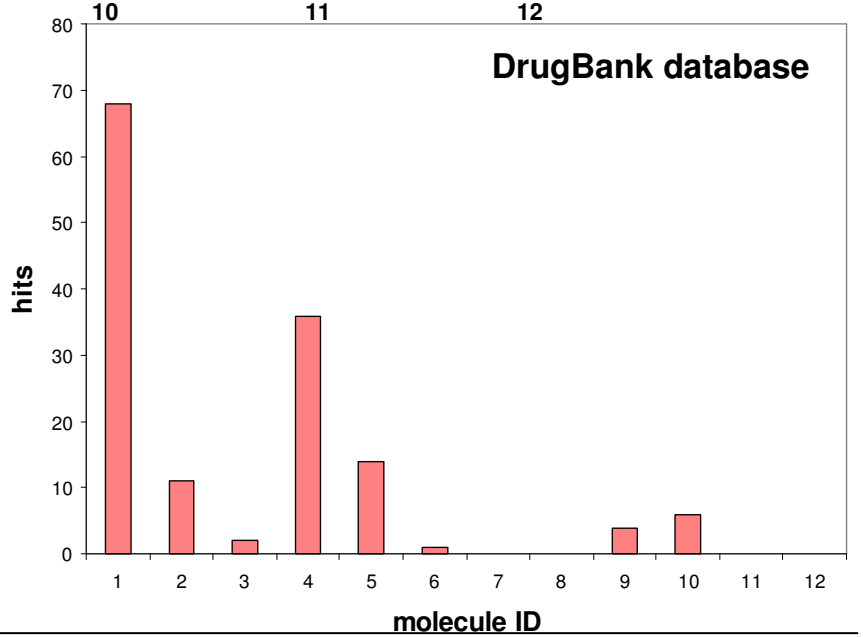
# Quinoline vs. isoquinoline

| Fragment | CrossFire Beilstein [hits] | DrugBank [hits] |
|---|---|---|
| Quinoline (q) | 85857 | 32 |
| Isoquinoline (i) | 15391 | 1 |
| Parameter q/i | **5,58** | **32** |

# Synthetic availability vs. drug population



# Conclusions

- Lepszy opis FCS pozwala *projektować lepszą chemię* (CS: VCS);
- konieczne nowe metody badania i opisu od intuicyjnej architektury do matematycznej topologii
- miejsce dla matematyki

# Instytut Chemii
## Uniwersytetu Śląskiego

*Organic Chemistry Katowice*

*chemoinformatyka.us.edu.pl*