History and Challenges of Chemoinformatics

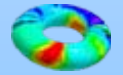
Johann Gasteiger

Computer-Chemie-Centrum

University of Erlangen-Nürnberg

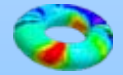
D-91052 Erlangen, Germany

www2.chemie.uni-erlangen.de/



Overview

- the scope of chemoinformatics
- the beginnings
- a field of ist own
- scientific challenges
- political challenges



Synthesis of Properties

The most fundamental and lasting objective of synthesis is not

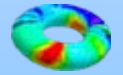
production of new compounds

but

production of properties

George S. Hammond

Norris Award Lecture, 1968



Fundamental Questions in Chemistry

What structure do I need for a certain property?

structure-activity relationships

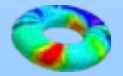
How do I make this structure?

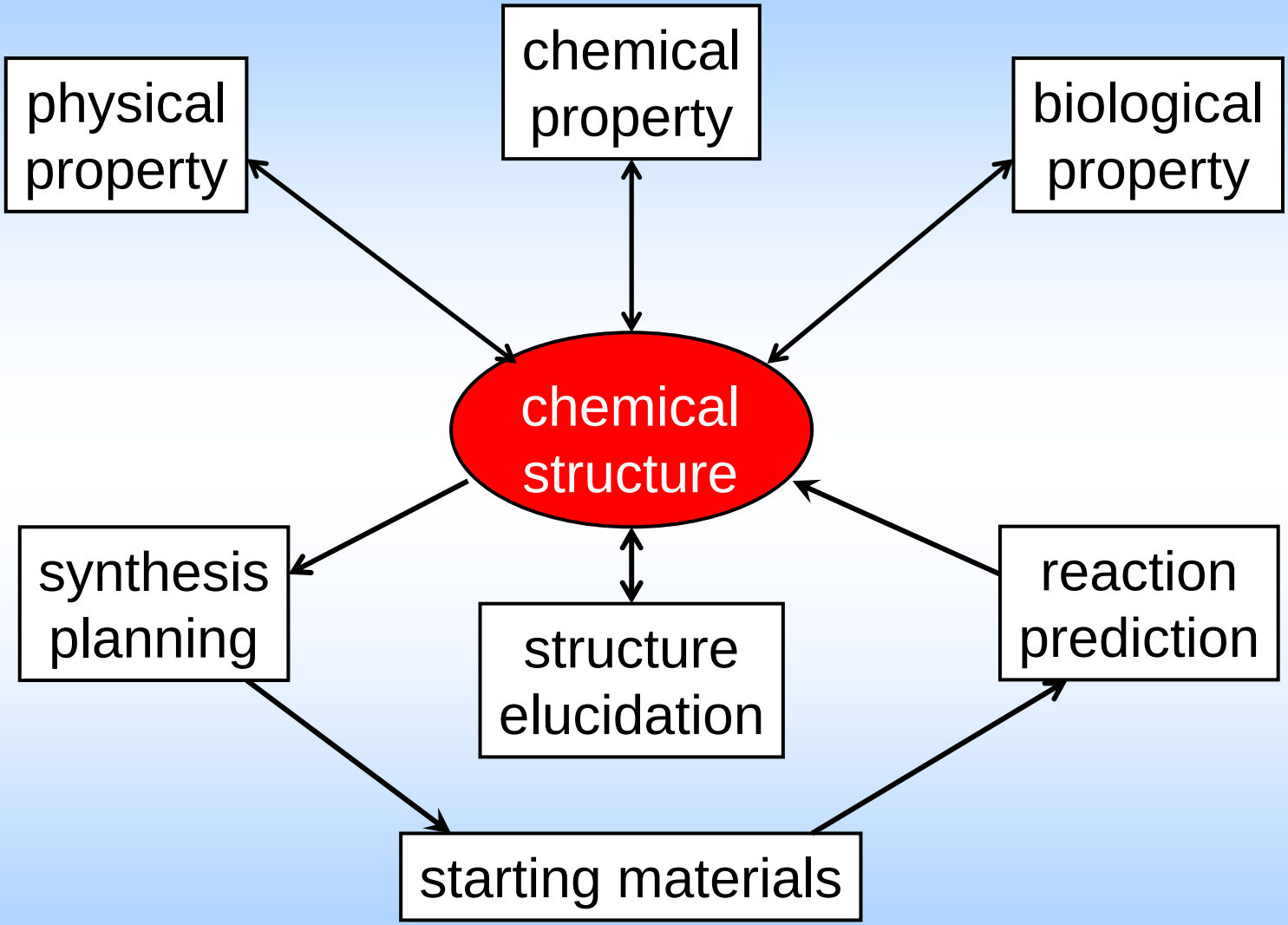
synthesis design

What is the product of my reaction?

reaction prediction

structure elucidation

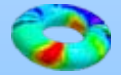



physical property
chemical property
biological property
chemical structure
synthesis planning
structure elucidation
reaction prediction
starting materials

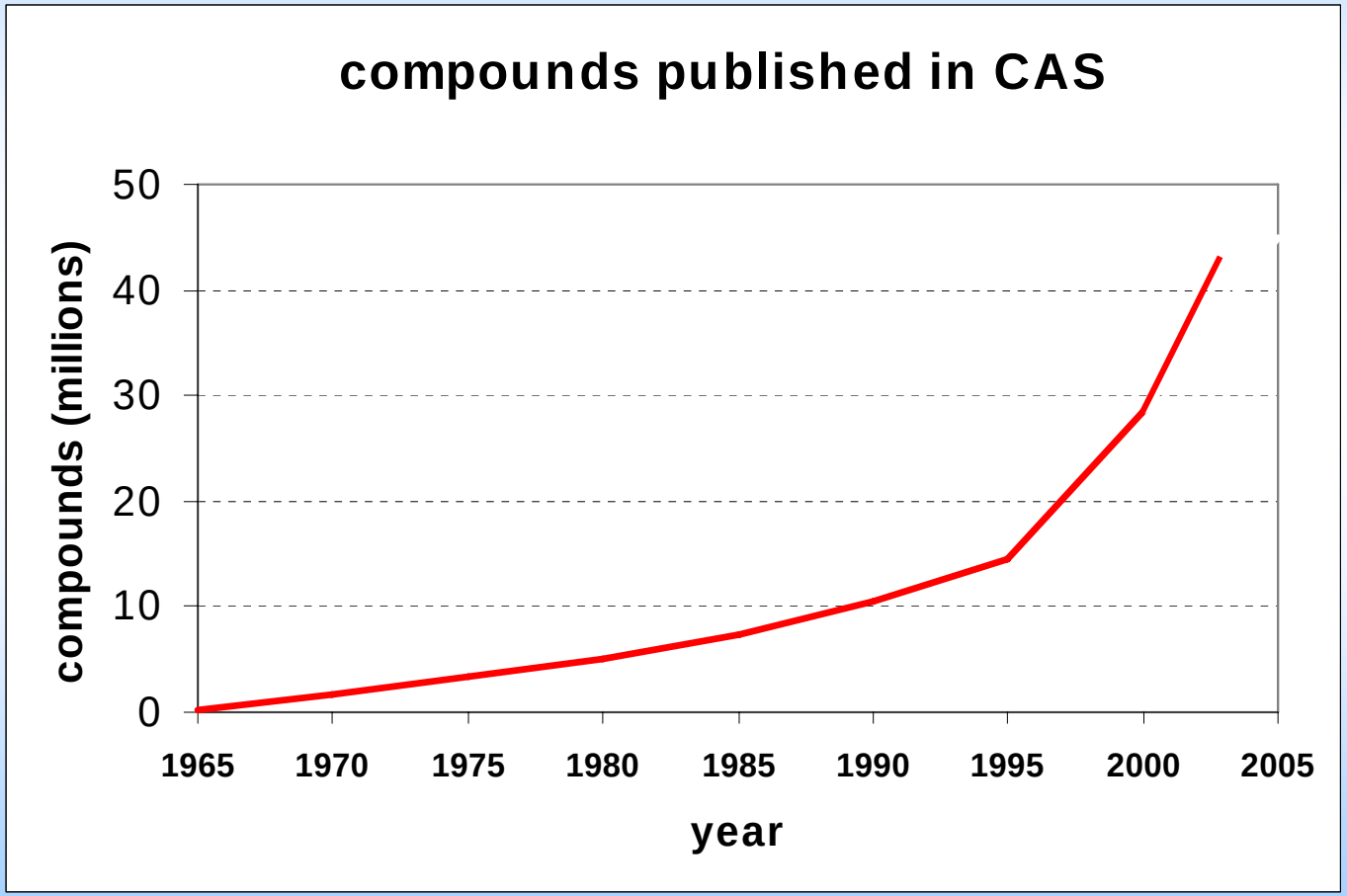
Chemoinformatics - Why?

- complex relationships
  - structure - biological activity
  - chemical reactivity

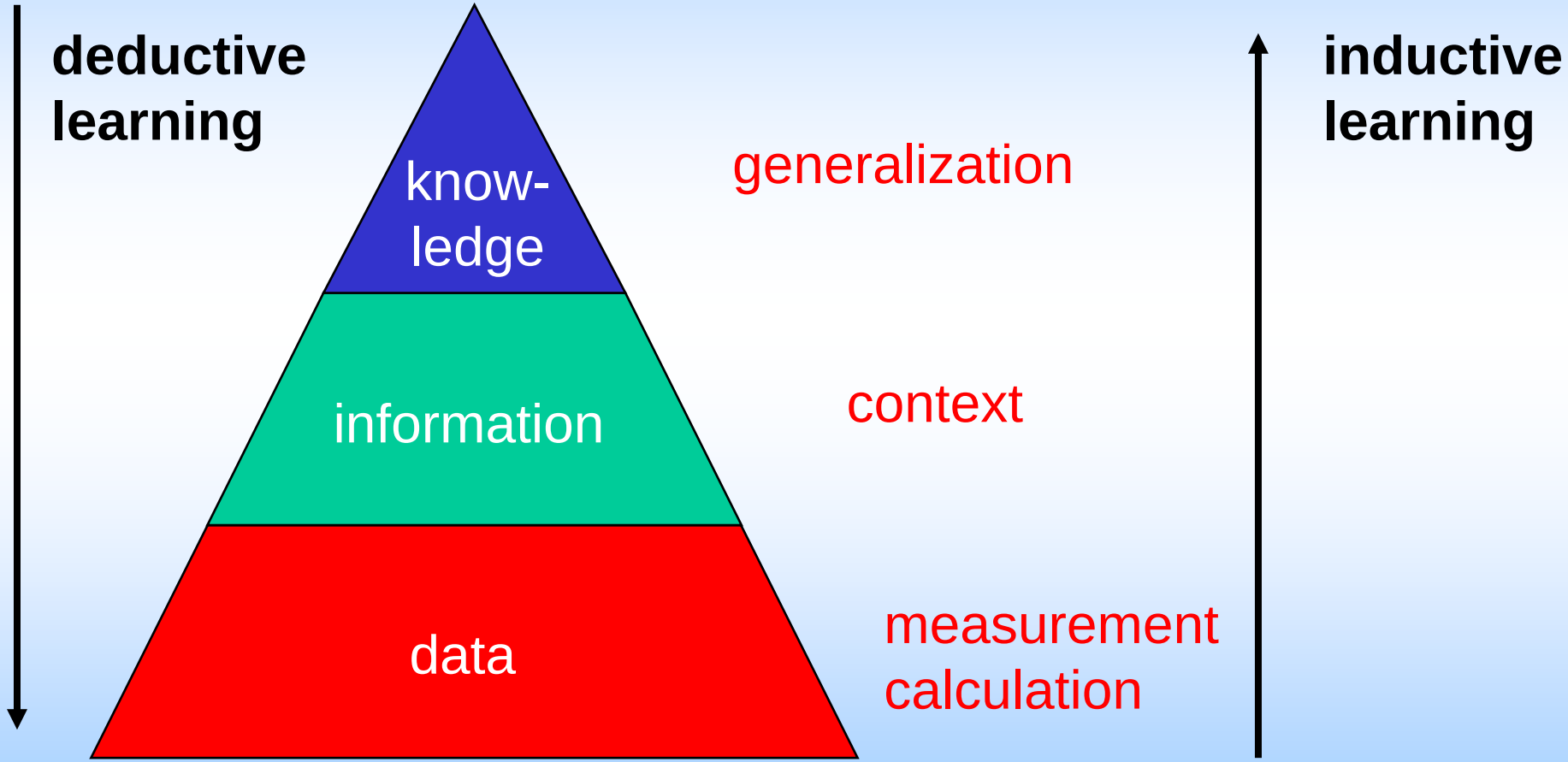
- amount of information
  - many millions of compounds and reactions
  - many millions of publications



Number of Compounds in Chemistry



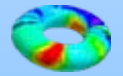
From Data to Knowledge



Chemoinformatics: Definition

„The use of information technology and management has become a critical part of the drug discovery process. ***Chemoinformatics*** is the mixing of those information resources to transform ***data*** into information and ***information*** into ***knowledge*** for the intended purpose of making better decisions faster in the area of drug lead identification and organization.“

F. K. Brown, *Annual Reports in Medicinal Chemistry* **1998**, *33*, 375-384



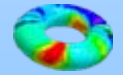
Chemoinformatics: Definition

The application of

informatics methods

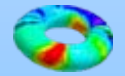
to solve

chemical problems



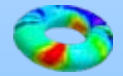
The Scope of Chemoinformatrics

- structure representation and searching
- data analysis and chemometrics
- molecular modeling
- spectra analysis and structure elucidation
- reaction representation and searching
- reaction modeling and synthesis design



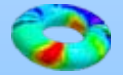
Application Areas for Chemoinformatics

- drug design
- analytical chemistry
- chemical engineering
- inorganic chemistry
- medicinal chemistry
- organic chemistry
- physical chemistry
- theoretical chemistry



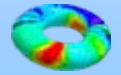
Chemoinformatics – An Old Discipline

- structure representation
  1965, Morgan
- structure elucidation
  1965, Sasaki, Munk, DENDRAL
- synthesis design
  1970, Corey & Wipke, Ugi, Gelernter, Hendrickson
- molecular modeling
  1970, Langridge, Marshall
- data analysis / chemometrics
  1970, Kowalski, Wold



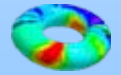
Structure Representation

- European industry: BASF, Hoechst, ICI, Thomae, BASIC, IDC (1965 -)
- Wiswesser Line Notation (1969 -)
- Chemical Abstracts Service: Morgan Algorithm 1965
- Sheffield: M. Lynch, P. Willett (1970 -)
- Paris: J.E.Dubois, DARC (1970 -)
- Munich: I. Ugi, J. Gasteiger, C. Jochum (1972 -)



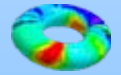
Computer-Assisted Structure Elucidation

- DENDRAL: C.Djerassi, J. Lederberg, D.Feigenbaum (1965)
- CHEMICS: S.Sasaki (1965)
- M.Munk (1968)
- C.Steinbeck (1998)

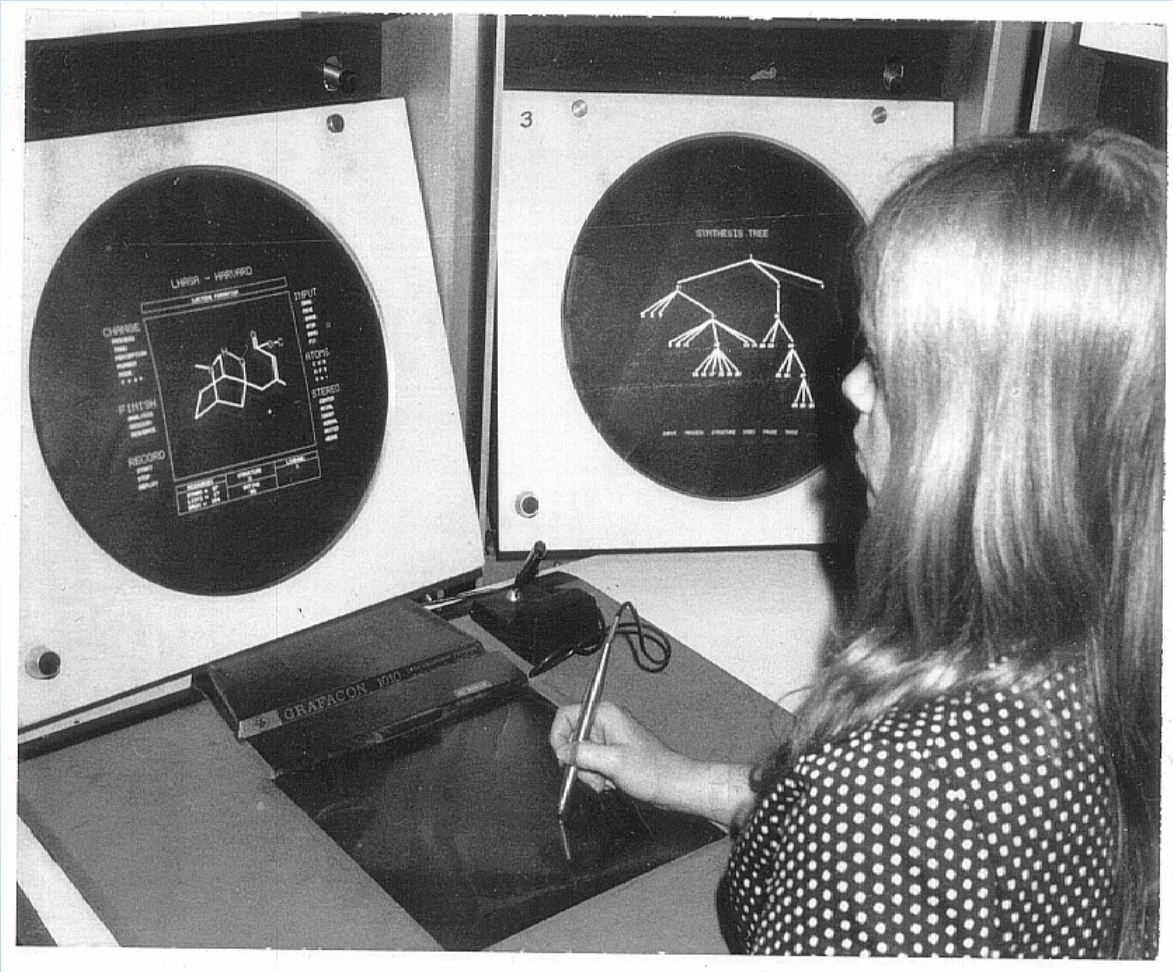


Computer-Assisted Synthesis Design

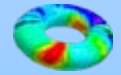
1969	Corey + Wipke	OCSS	⟶ LHASA, SECS
1973	Ugi + Gasteiger	CICLOPS	⟶ WODCA, THERESA
1971	Hendrickson	SYNGEN	
1976	Bersohn	SYNSUP	
1977	Gelernter	SYNCHEM	
1985	Hanessian	CHIRON	
1988	Zefirov	FLAMINGOES	
1988	Sasaki + Funatsu	AIPHOS	



Visualzation of Chemical Structures

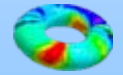


LHASA 1970



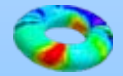
Data Analysis Methods

- Chemometrics: B.Kowalski (1970)
- PLS: S. Wold (1978)
- Self-organizing neural network: Kohonen (1983)
- Backpropagation Algorithm: Rumelhart, Hinxton (1987)



Databases

- Chemical Abstracts Service (1975)
- DARC System (1980)
- Cambridge CSD (1984)
- Inorganic Structures Database (1985)
- Beilstein (1990)
- Gmelin (1990)
- ChemInformRX (1991)
- SpecInfo (1991)

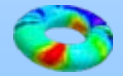


Common Topics: Structure Representation

- data storage and retrieval
- property prediction
- drug design
- synthesis design
- spectra analysis and prediction

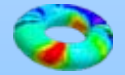
Common Topics: Data Analysis Methods

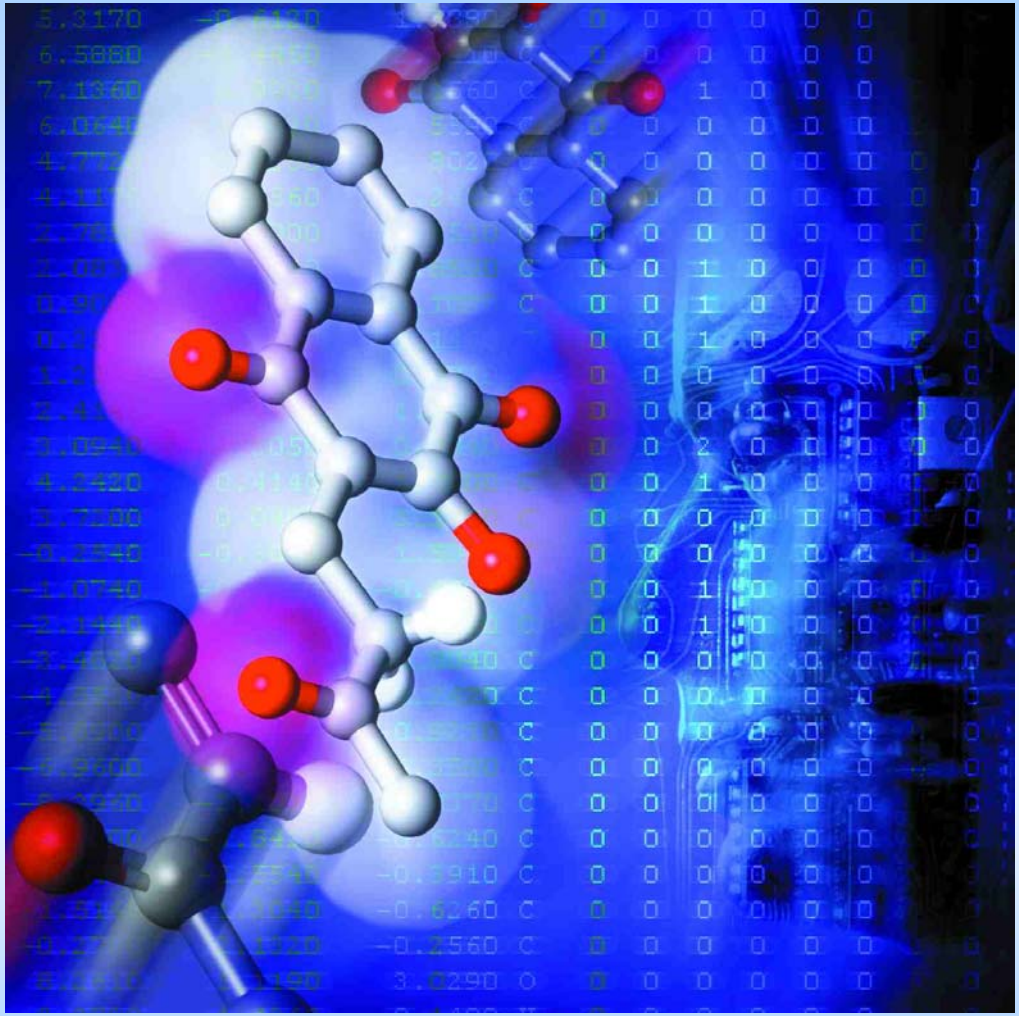
- property prediction
- drug design
- analytical chemistry
- spectra analysis and prediction



Common Topics

- representation of chemical structures
- searching structures in databases
- visualization of chemical structures
- representation of chemical reactions
- data analysis methods



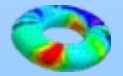


Handbook of Chemoinformatics

From Data to Knowledge

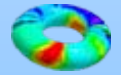
J. Gasteiger (Editor)

65 authors
73 contributions
4 volumes
1900 pages
Wiley-VCH, Weinheim
(August 2003)



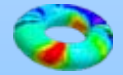
Chemical Structures - Challenges

- representation of Markush structures (patents)
- representation of polymers
- conformational flexibility (bioactive conformation)
- similarity searching (beyond fingerprints and Tanimoto coefficient)



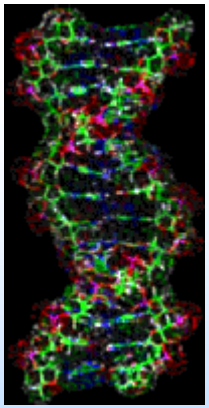
Proteins - Challenges

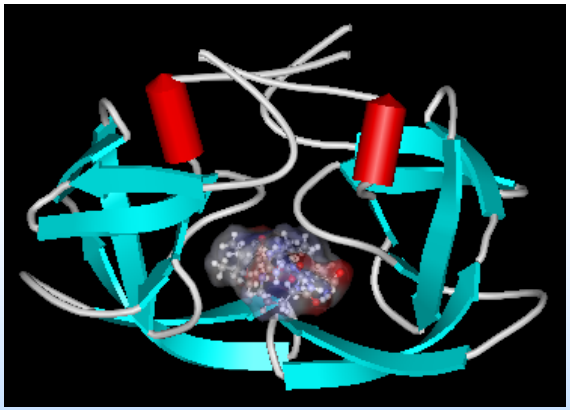
- scoring functions for docking
- flexibility of proteins

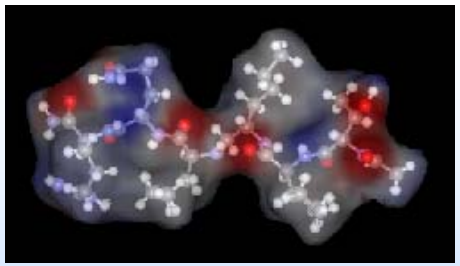


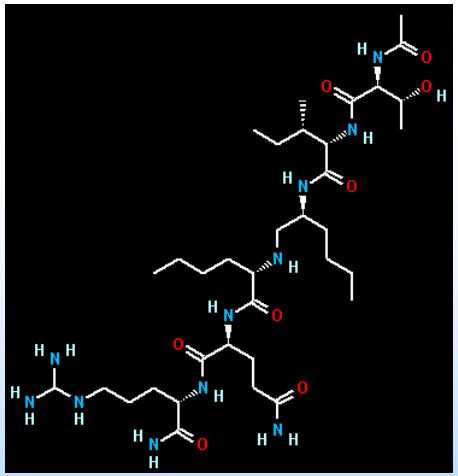
Bioinformatics Chemoinformatics

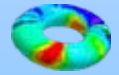
gene ⟷ protein ⟷ drug ⟷ lead





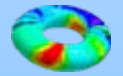






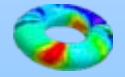
Information Acquisition - Challenges

- electroníc laboratory notebooks
- publishing chemical information (3D structures, spectra)
- publishing and searching on the internet
- text mining
- optical character recognition
- input of chemical structure (hand writing, voice)



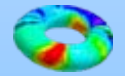
Data Mining - Challenges

- descriptor elimination
- model validation
- automatic model building
- definition of applicability domain

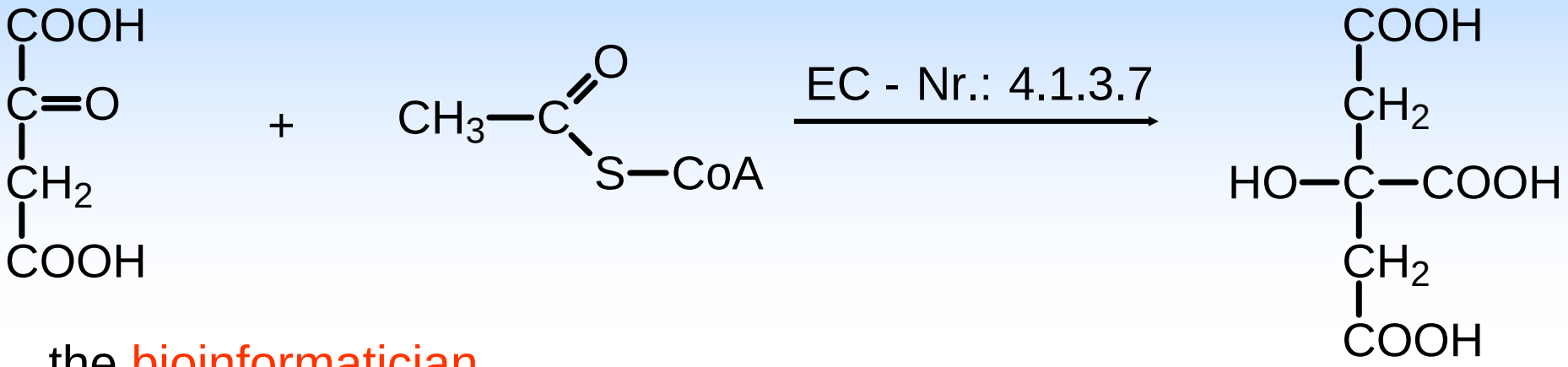


Chemical Reactions - Challenges

- modeling of chemical reactivity
- prediction of the course of chemical reactions
- synthesis design
- prediction of metabolism/degradation (abiotic and biotic)
- analysis of biochemical pathways



What is a Chemical Reaction?



the bioinformatician

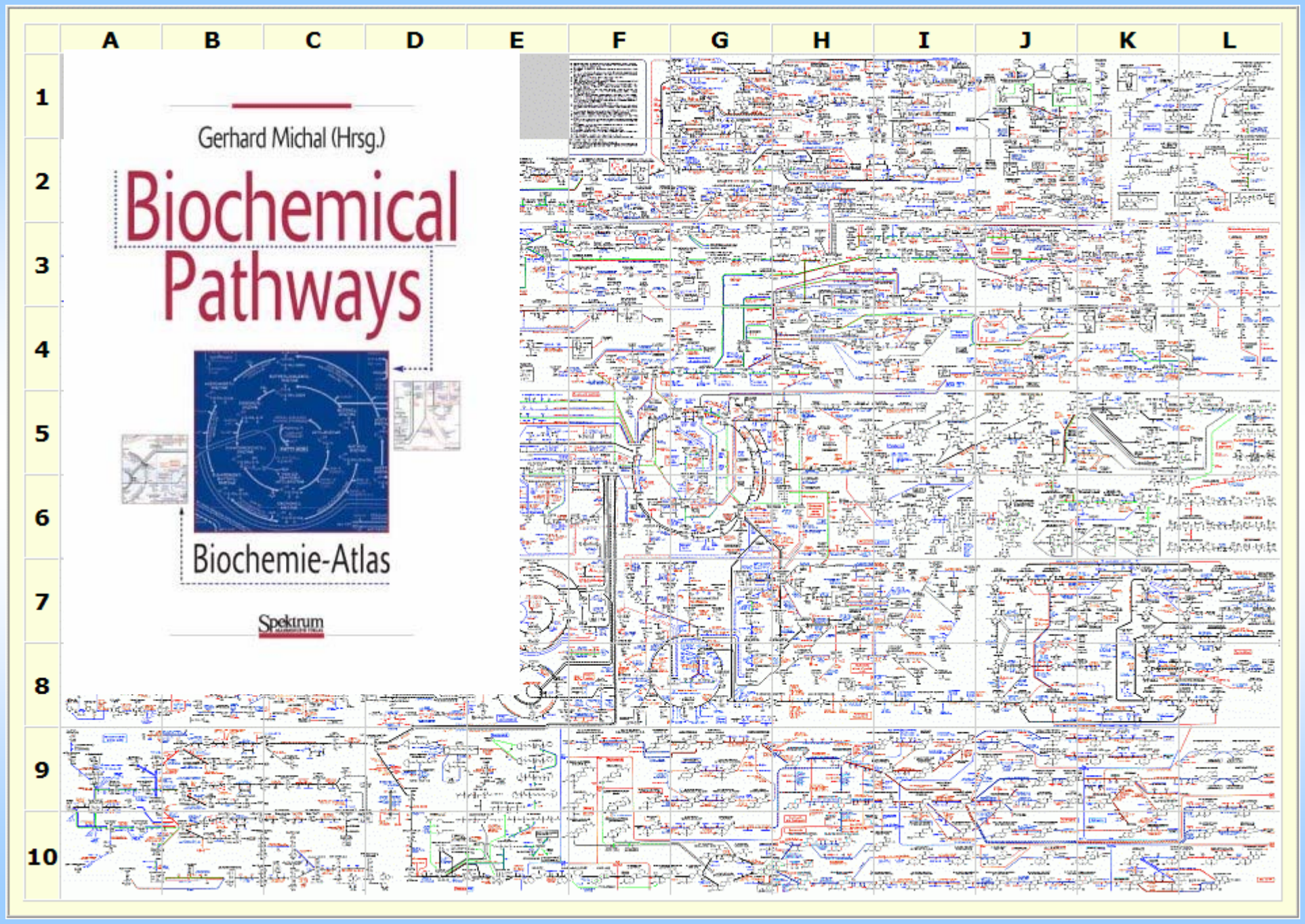
an event influenced by a gene, a protein

the computer scientist

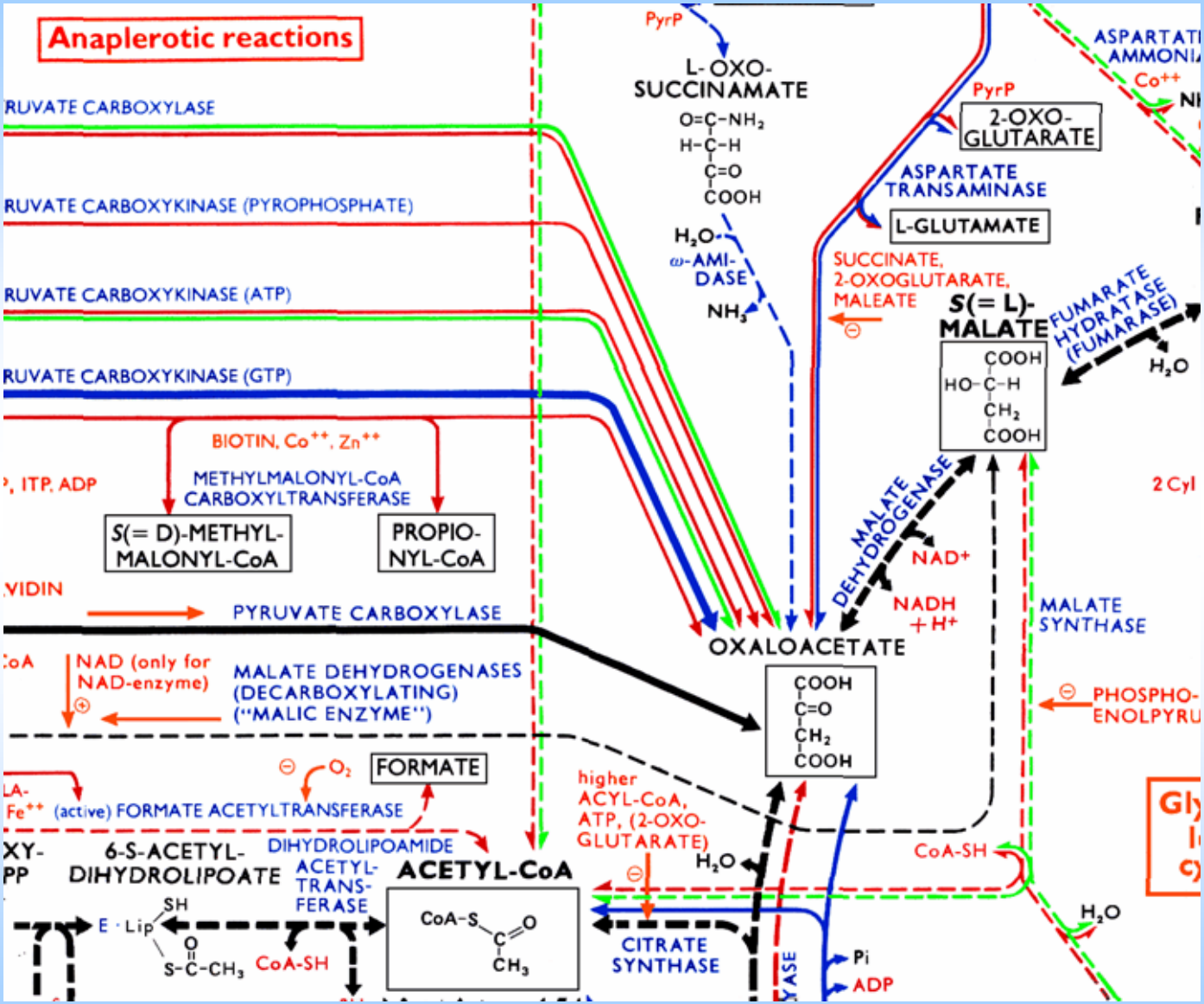
a context sensitive graph rewriting rule

the chemist

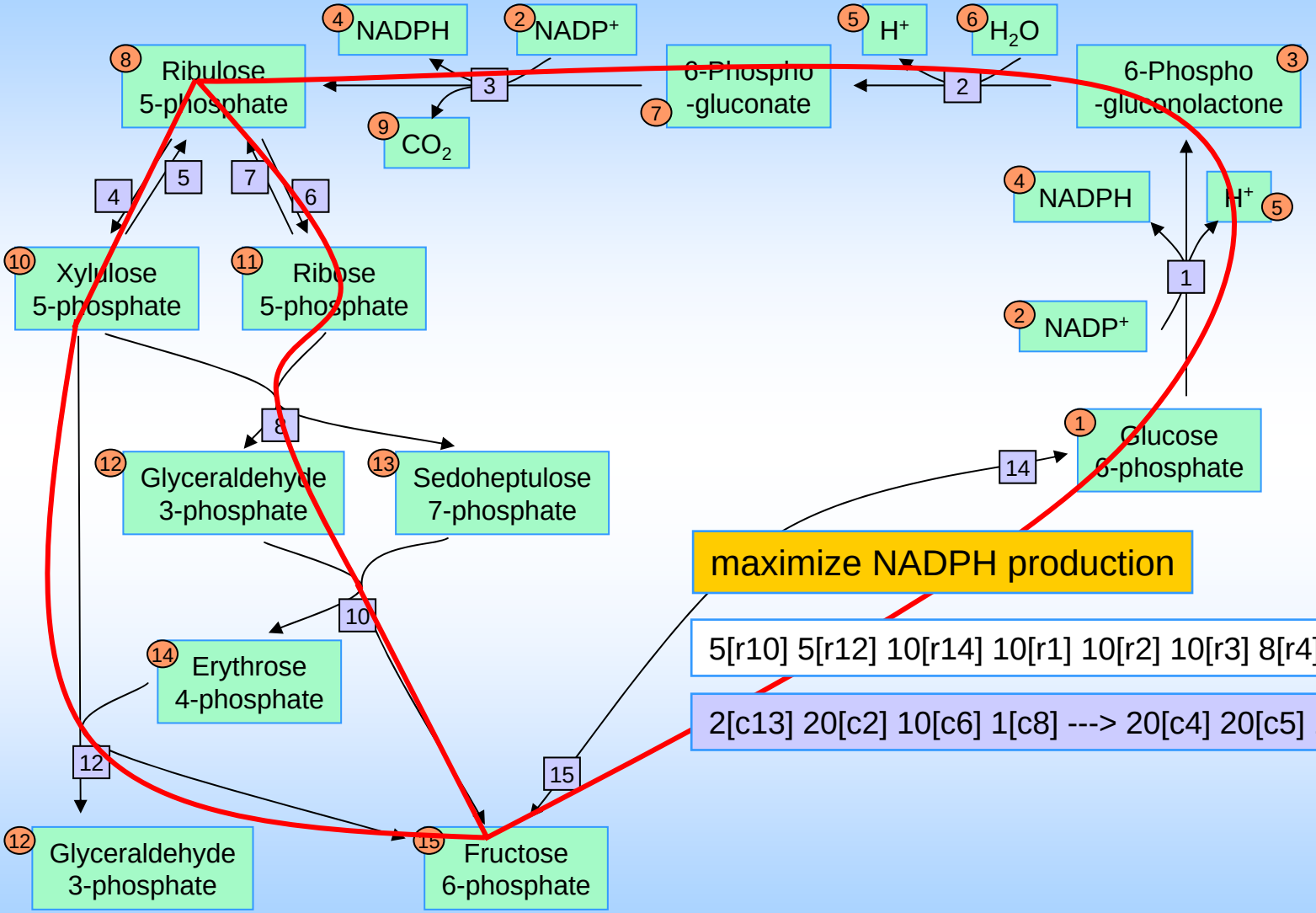
an event breaking and making bonds

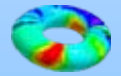

A
B
C
D
E
F
G
H
I
J
K
L
1
2
3
4
5
6
7
8
9
10
Gerhard Michal (Hrsg.)
Biochemical Pathways
Biochemie-Atlas
Spektrum

Biochemical Pathways



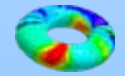
Pathway Searching





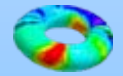
Prediction of Properties - Challenges

- physical
  - spectra (CASE)
  - color of dyes etc
- chemical
  - chemical reactivity
- biological
  - toxicity
- risk assessment (chemical + biological) REACH



Application Areas for Chemoinformatics - Challenges

- drug design
- analytical chemistry
- chemical engineering
- inorganic chemistry
- medicinal chemistry
- organic chemistry
- physical chemistry
- theoretical chemistry



Teaching

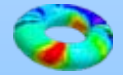
Sheffield

UMIST

Strasbourg

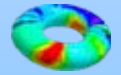
Erlangen

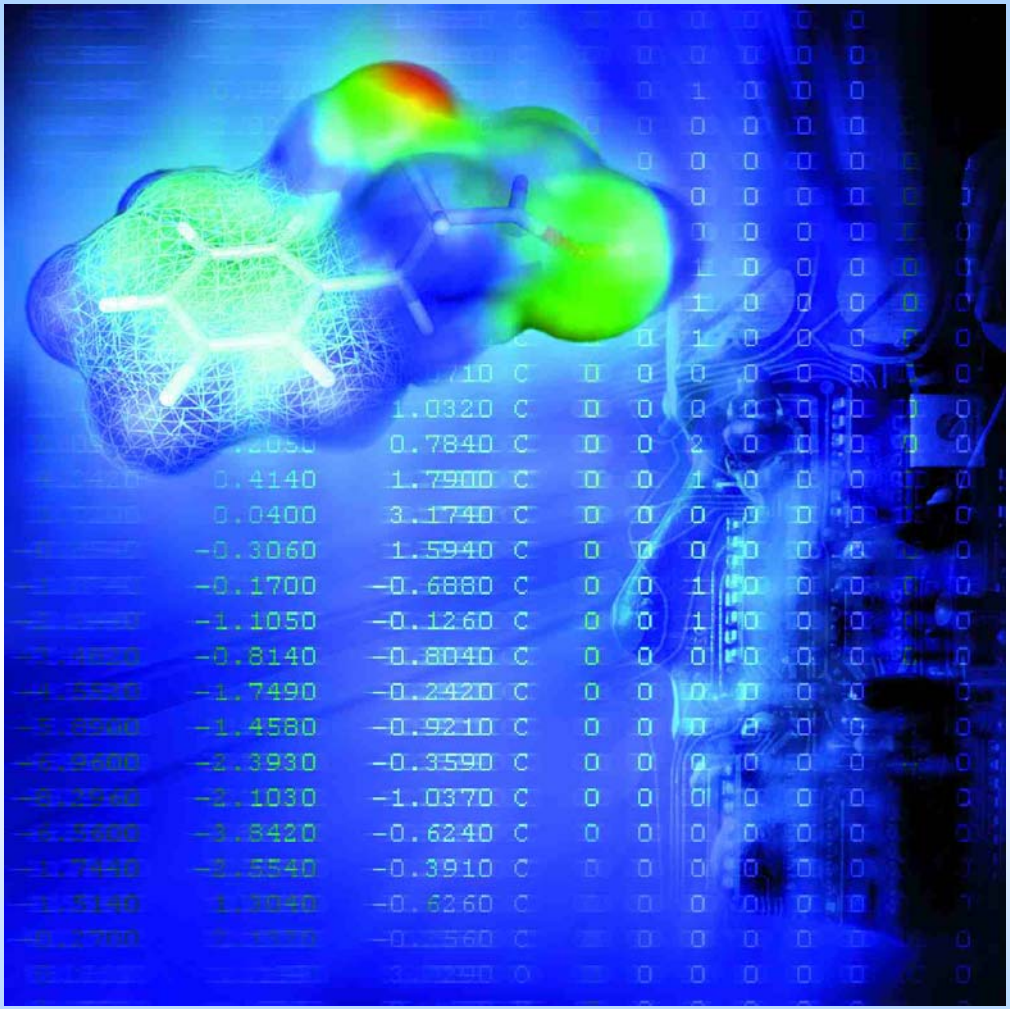
Indiana University



Textbooks on Chemoinformatics

- V. Gillet, A. Leach
- J. Gasteiger, T. Engel
- J. Bajorath



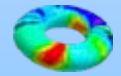


Chemoinformatics
- A Textbook -

J. Gasteiger, T. Engel
(Editors)

650 pages

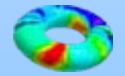
Wiley-VCH, Weinheim
(September 2003)



Teaching

- define curriculum in chemoinformatics

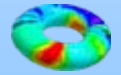
- what contents of chemoinformatics have to go into regular chemistry curricula



Cooperation Industry - Academia

- industry: generate data
- academia: develop methods

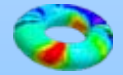
➡ provide academia access to data



Funding

- increase awareness for importance of Chemoinformatics

- go into committees



Get Organized!

Chemometrics Society

QSAR Society

FECS Working Party: Computational Chemistry

➡ Chemoinformatics Society