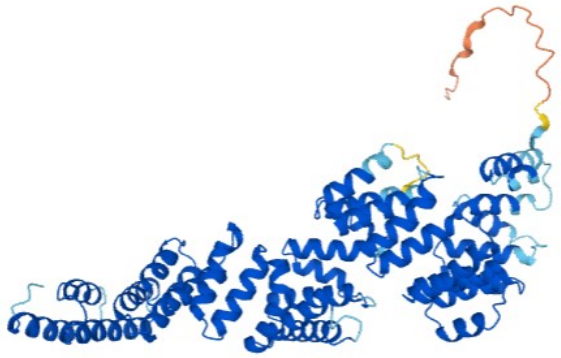
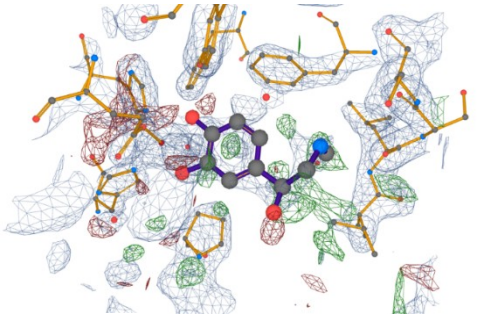
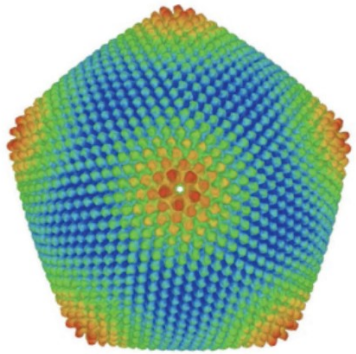
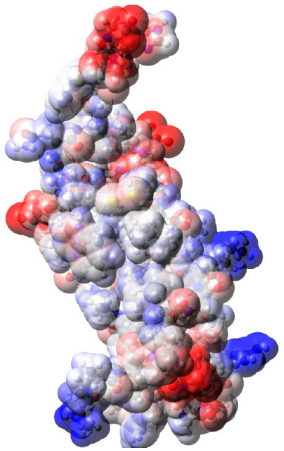
Chemoinformatics



Radka Svobodová







CEITEC
Central European Institute of Technology
BRNO | CZECH REPUBLIC

Radka Svobodová - introduction

2000: Mgr., **Biochemistry**, MU
2003: Mgr., **Informatics**, MU
2006: RNDr., **Physical chemistry**, MU
2007: Ph.D., **Biomolecular chemistry**, MU
2017: Doc., **Biomolecular chemistry**, MU

Definition of chemoinformatics

Cheminformatics

also known as **chemoinformatics** and **chemical informatics**

= the use of computer and informational techniques, applied to a range of problems in the field of chemistry.

Applications:

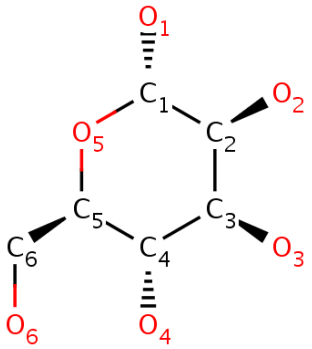
- life science research
- pharmaceutical companies - drug discovery
- chemical and allied industries

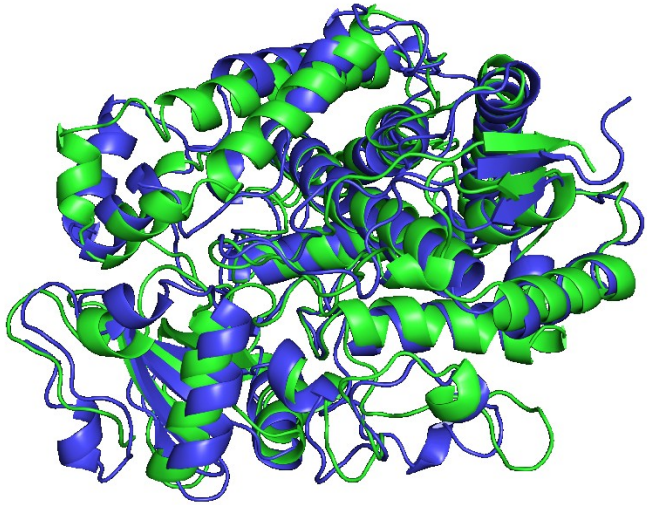
Why theoretical chemistry?

- Avoid working with toxic, explosive and radioactive substances
- We can work with unstable substances
- We can save costs for chemicals and for the implementation of the experiment
- We save time for experimental chemists :-)

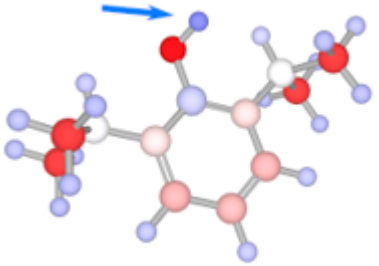
Content

- **Introduction:** concept of chemoinformatics, content of the subject, history of the field
- **Computer model of a molecule:** 1D, 2D and 3D structure, molecule representation using graph and matrix
- **2D structure (topology) of a molecule:**
  - writing a molecule using a string (SMILES, InChi, InChiKey)
  - **Molecular graphs:** Isomorphism and canonical indexing
- **3D structure (geometry) of the molecule:**
  - representation using Cartesian and internal coordinates, data formats, geometry comparison



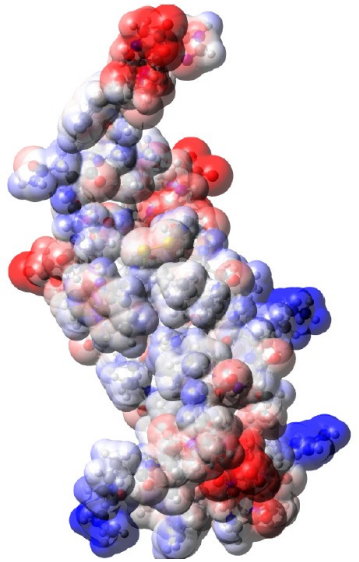


Content



Basic chemoinformatics tasks:

- **Visualization** of structures of molecules and molecular fragments, models for visualization of molecules
- **Molecular descriptors computation:** 1D, 2D and 3D
- **Similarity of molecules:** similarity comparison, similarity coefficients
- **QSAR and QSPR models:** Models for studying quantitative relationships between structure and activity/property of
- **Databases** of small and large molecule structures, searching them
- **Generation of molecule structures** using AI algorithms



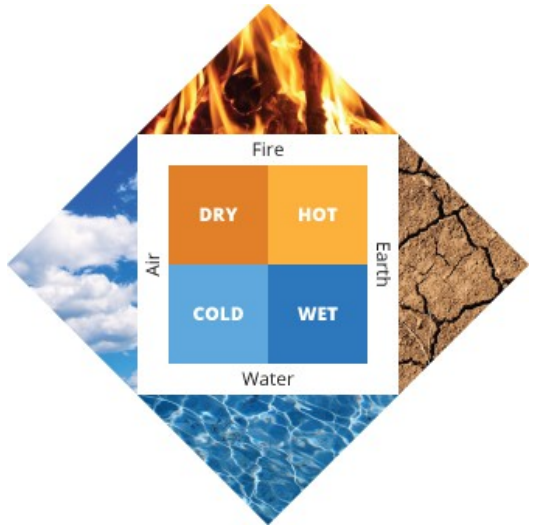
Organization of the course

1 hour lecture + 1 hour praxis
Please, bring your notebooks

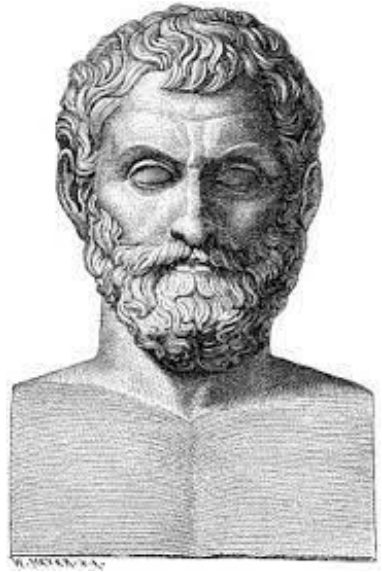
Optional tasks – possibility to obtain additional points

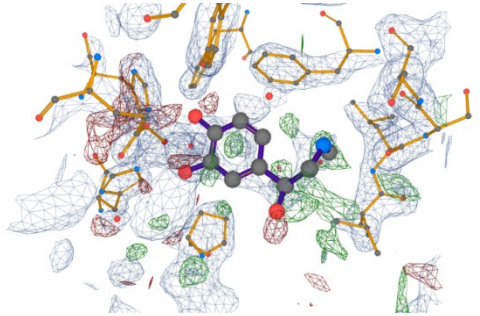
End of the subject:

- **Colloquium**
- Written test
  - Theoretical questions
  - Examples
- All documents allowed



HISTORY



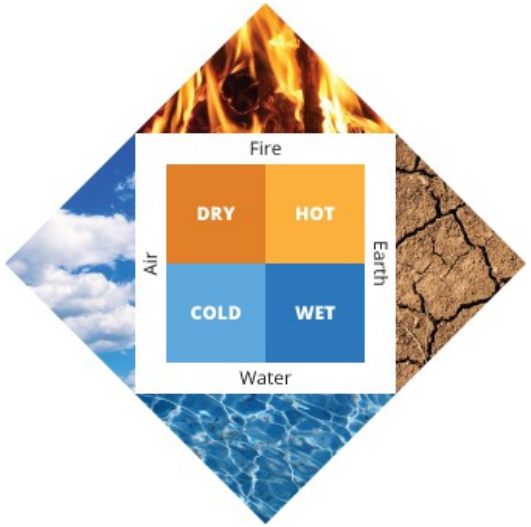


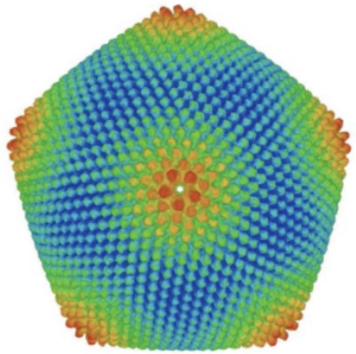
History of theoretical chemistry

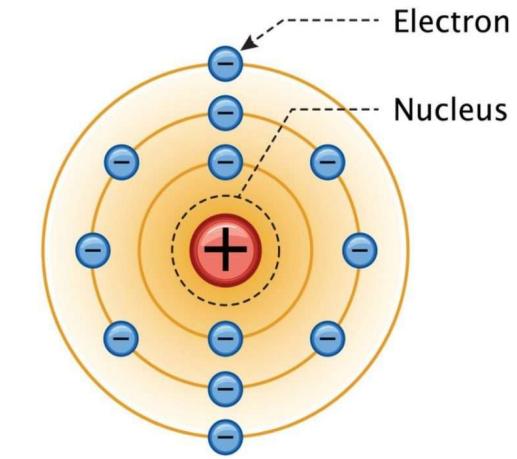
Philosophical theories
Antics - 17. century

First theoretical models
17. century - 1960

Modern life sciences
1960 - now







Planetary model
(Bohr 1913)

Philosophical theories

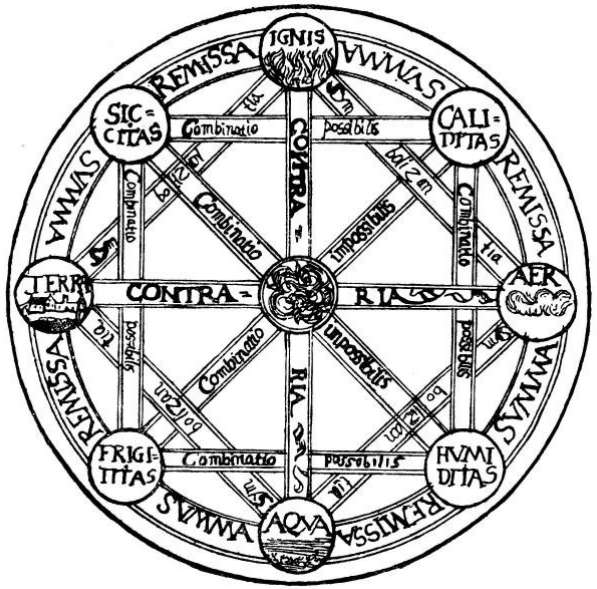
Antics:
The word is composed from a first principle.
It is a basic proposition or assumption that cannot be deduced from any other proposition or assumption.

What is the first principle?:
- Pre-Socratic philosophers
- Theory of 4 elements
- Atomistic theory

Raffael Santi
Aténská škola

Philosophical theories

Theory of 4 elements



Empedocles:

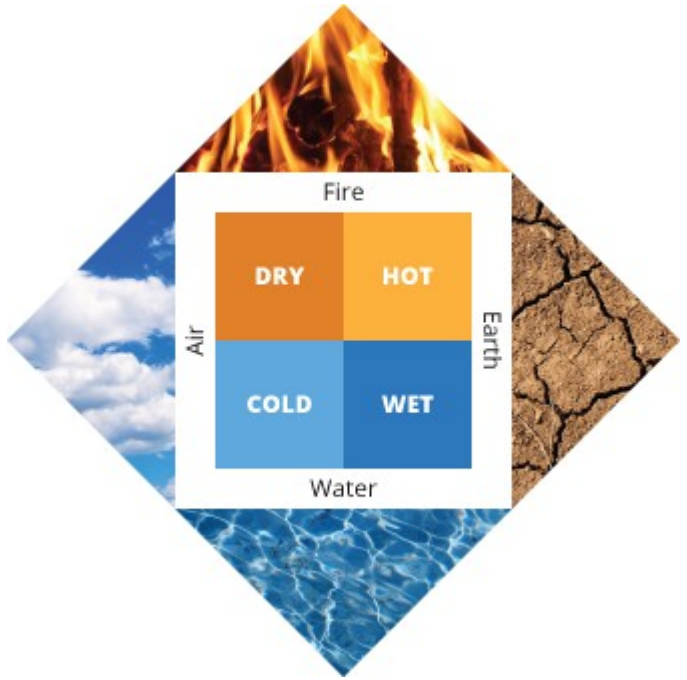
- Established four ultimate elements which make all the structures in the world - **fire, air, water, earth**.
- The four elements are both eternally brought into union and parted from one another by two divine powers, Love and Strife.
- He called them the four "roots"

Plato:

- The first to use the term "element“

Aristotle:

- Related each of the four elements to two of the four sensible qualities:
  - Fire is both hot and dry.
  - Air is both hot and wet (for air is like vapor, □ τμ□ ȼ
  - Water is both cold and wet.
  - Earth is both cold and dry.

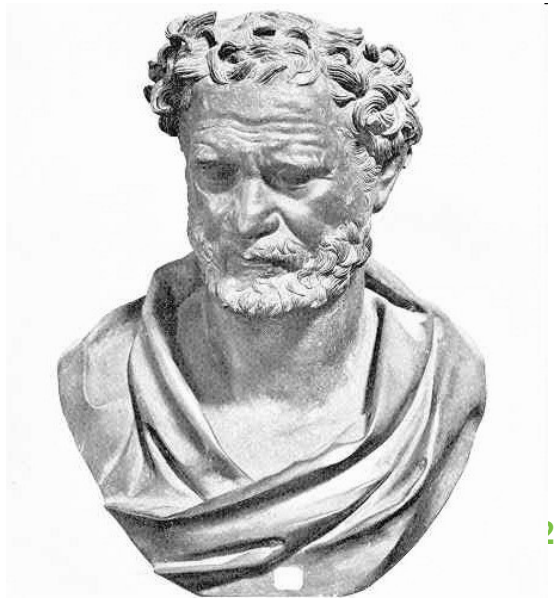


Philosophical theories

Atomistic theory

Leukippos:

- nature consists of two fundamental principles: atom and void
- clusters of different shapes, arrangements, and positions give rise to the various macroscopic substances in the world



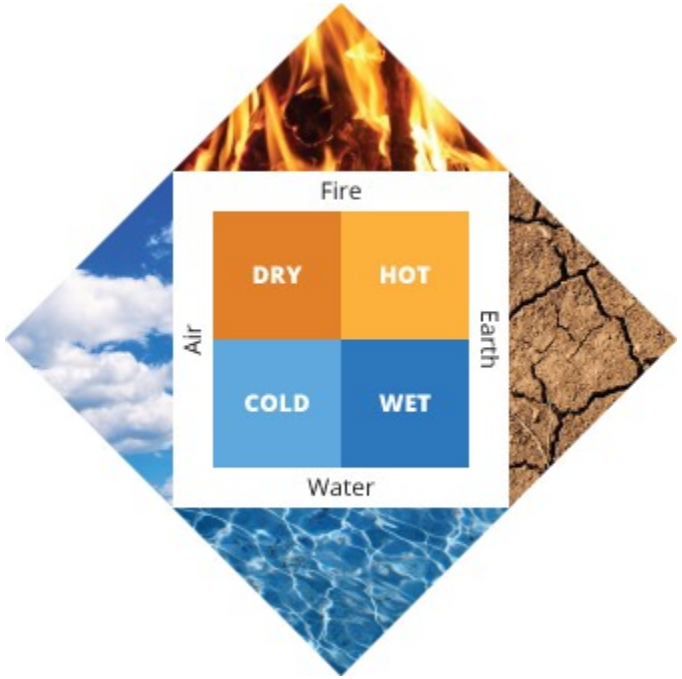
Demokritos:

- all matter was composed of small indivisible particles which they called "atoms„
- atoms are too small for human senses to detect, that they are infinitely many, that they come in infinitely many varieties, and that they have always existed

Which theory was prefered

?

Which theory was prefered?



How long?

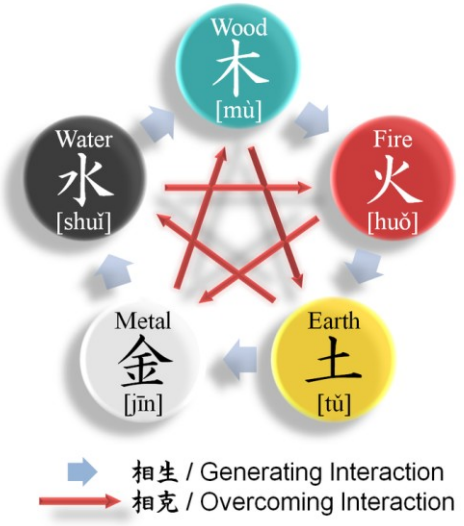
How long?

Till 17th century

How long?


Fire
Aries
Leo
Sagittarius
Earth
Taurus
Virgo
Capricorn
Air
Gemini
Libra
Aquarius
Water
Cancer
Pisces
Scorpio


CHI
(earth)
MIZU
(water)
KA
(fire)
KAZE
(wind)
GAI
(void)


Wood
木
[mù]
Water
水
[shuǐ]
Fire
火
[huǒ]
Metal
金
[jīn]
Earth
土
[tǔ]
相生 / Generating Interaction
相克 / Overcoming Interaction

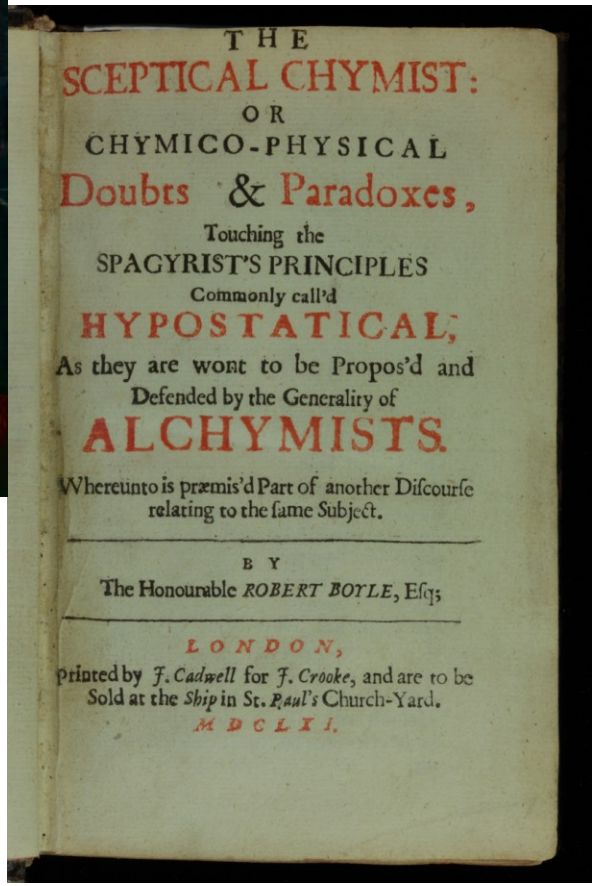
First theoretical models

The four elements theory finally rejected.

Sir Robert Boyle, 1661

The Sceptical Chymist: or Chymico-Physical Doubts & Paradoxes

Comeback of the Democritos theory


THE
SCEPTICAL CHYMIST:
OR
CHYMICO-PHYSICAL
Doubts & Paradoxes,
Touching the
SPAGYRIST'S PRINCIPLES
Commonly call'd
HYPOSTATICAL;
As they are wont to be Propos'd and
Defended by the Generality of
ALCHYMISTS.
Whereunto is præmis'd Part of another Diſcourſe
relating to the ſame Subject.
BY
The Honourable ROBERT BOYLE, Eſq;
LONDON,
printed by J. Cadwell for J. Crooke, and are to be
Sold at the Ship in St. Paul's Church-Yard.
MDCLXI.

First theoretical models

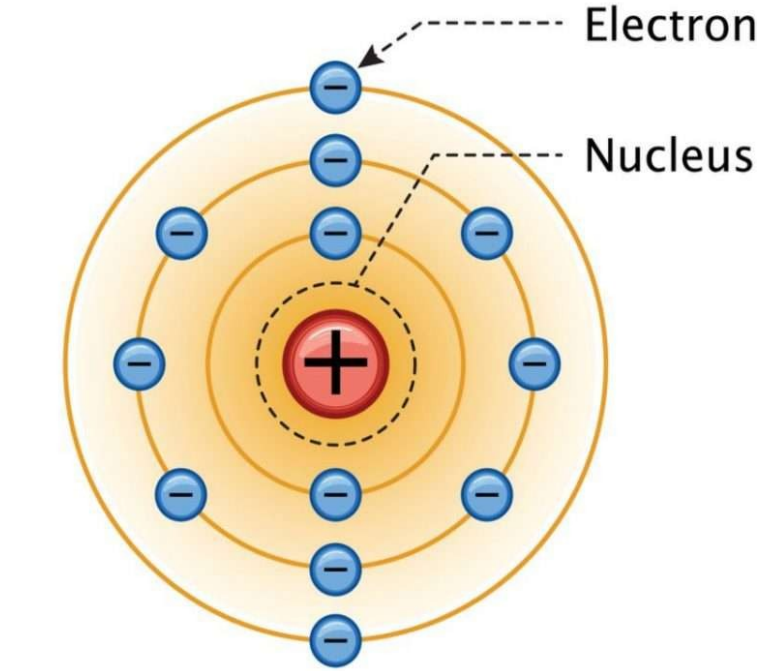
1808 J. Dalton extended the Democritos theory:

- The smallest particles of matter that cannot be physically or chemically divided are **atoms**.
- The **atoms of the same element** have the same quality, size and mass, and in these properties they differ from the atoms of other elements.
- In **chemical combining**, only the whole number of atoms belonging to the elements of which the compound under consideration is composed are always combined.

First theoretical models

Atoms can be divided:

- 1897 J.J. Thompson detected an **electron**
- 1911 E. Rutherford found **atomic core (nucleus)**
- 1913 N. Bohr published **a model of an atom**:
  - The electron can only move around the nucleus in certain orbits without emitting energy.
  - The electron emits or receives energy only when it jumps from one energy level to another.

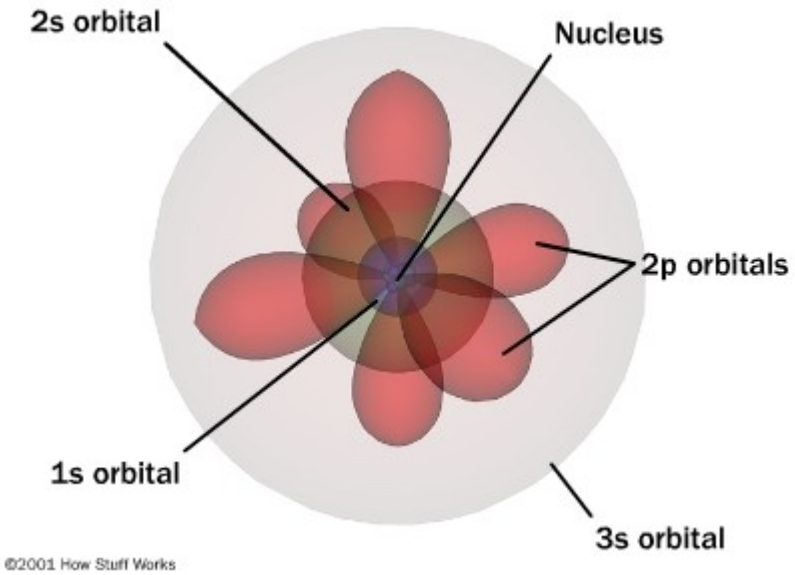


Planetary model
(Bohr 1913)

First theoretical models

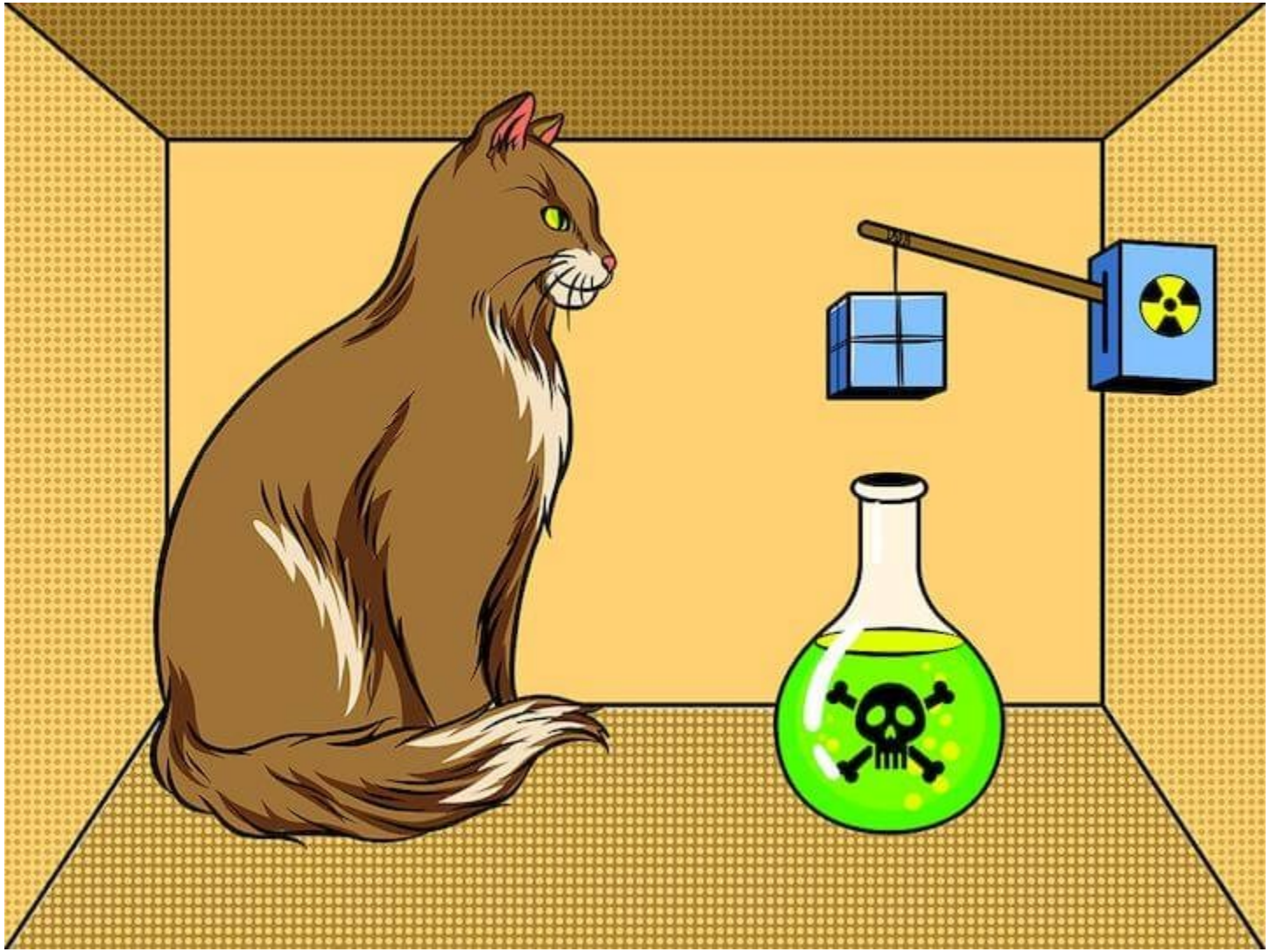
Quantum theory:

- 1925 - 1926 W. Heisenberg and E. Schrodinger published basic ideas of quantum theory
- 1926 E. Schrodinger formed a wave function, describing a probability of electron occurence arround the nucleus.



$$\hat{\mathcal{H}}\Psi = E\Psi$$

First theoretical models

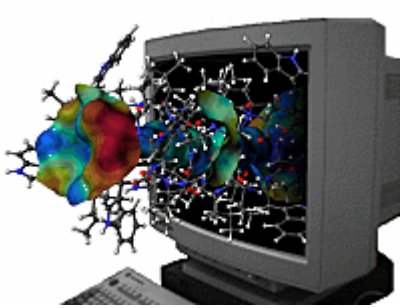


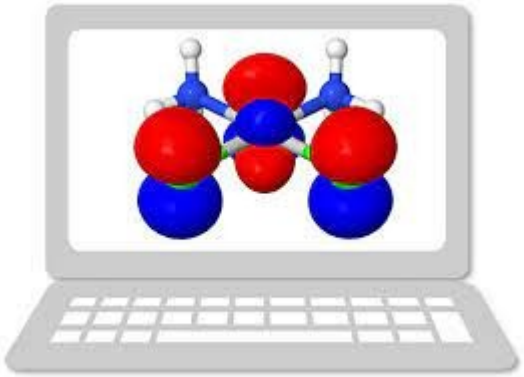
First theoretical models

Modern life sciences

Compute resources available:

- From the 1960s onwards, computers became available for research laboratories
- Available programming languages for scientific calculations (FORTRAN, C, etc.).
- New research field: **Computational chemistry**





Modern life sciences

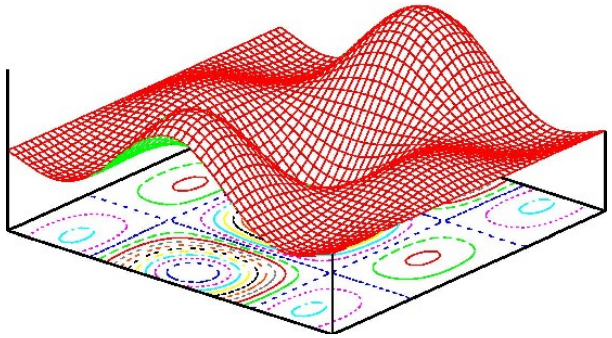
Computational chemistry

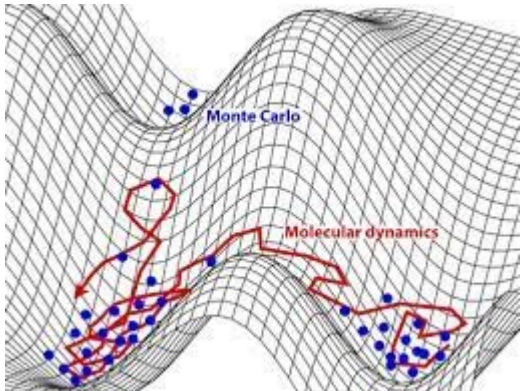
„...an attempt to model all aspects of real chemistry as precisely as possible using computations, instead of experiments.“

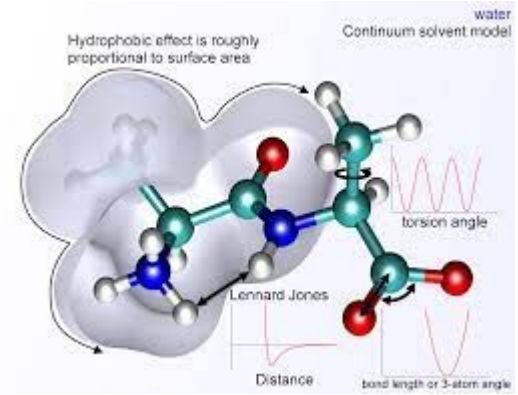
R. Schleyer

Specializations:

- **Molecular mechanics**
- **Quantum mechanics**
- **Molecular dynamics**





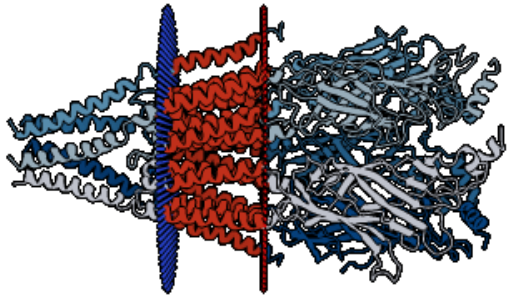


Modern life sciences

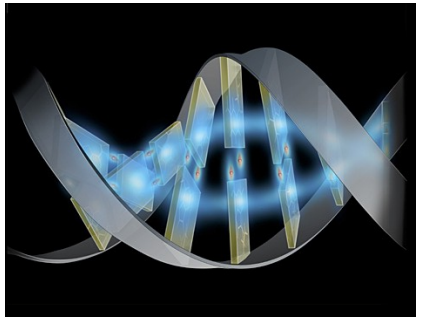
Molecular structure data available:

- From the 1970-1980s onwards, we have available protein structures and DNA sequences
- New field established: **Bioinformatics**

- **Specializations:**
  - **Structural bioinformatics**



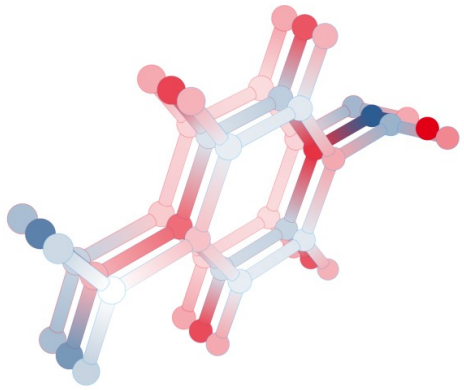
  - **Sequence bioinformatics**

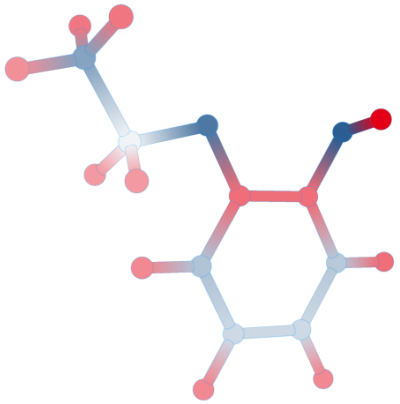


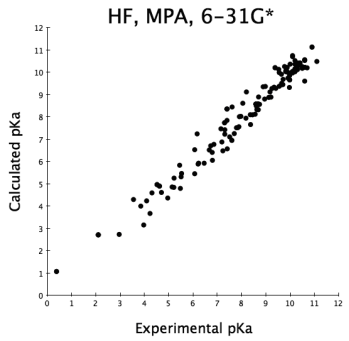
Modern life sciences

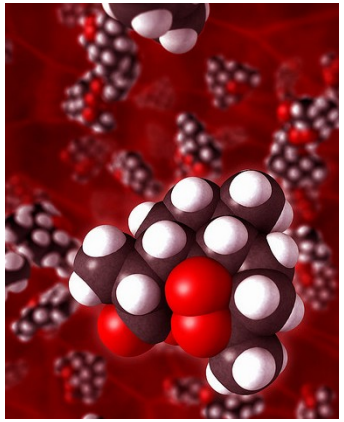
Molecular structure data available:

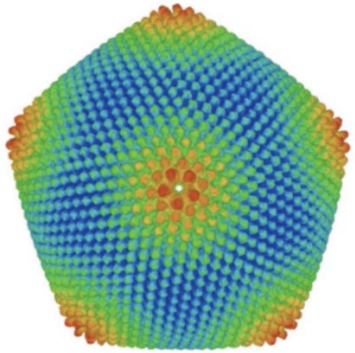
- From the 1980s, structures of small molecules are available
- New field established: **Chemoinformatics**

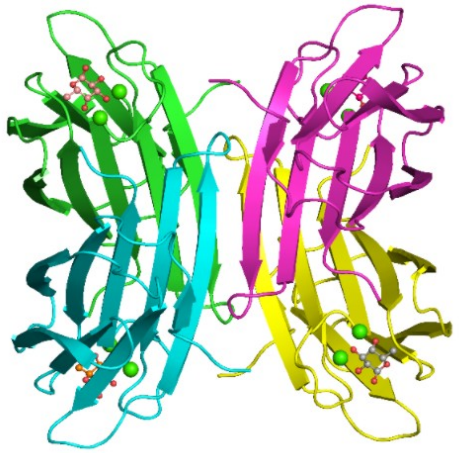




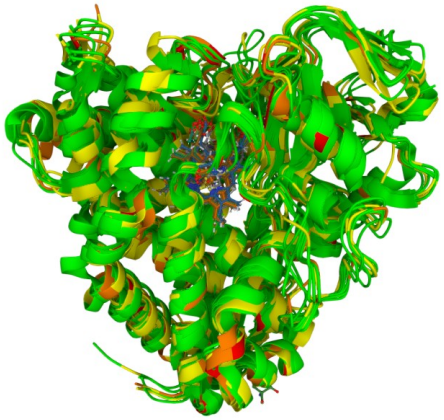


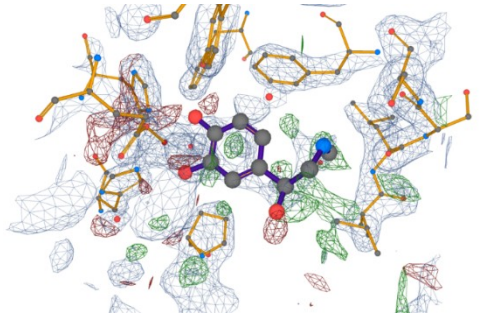






TRENDS

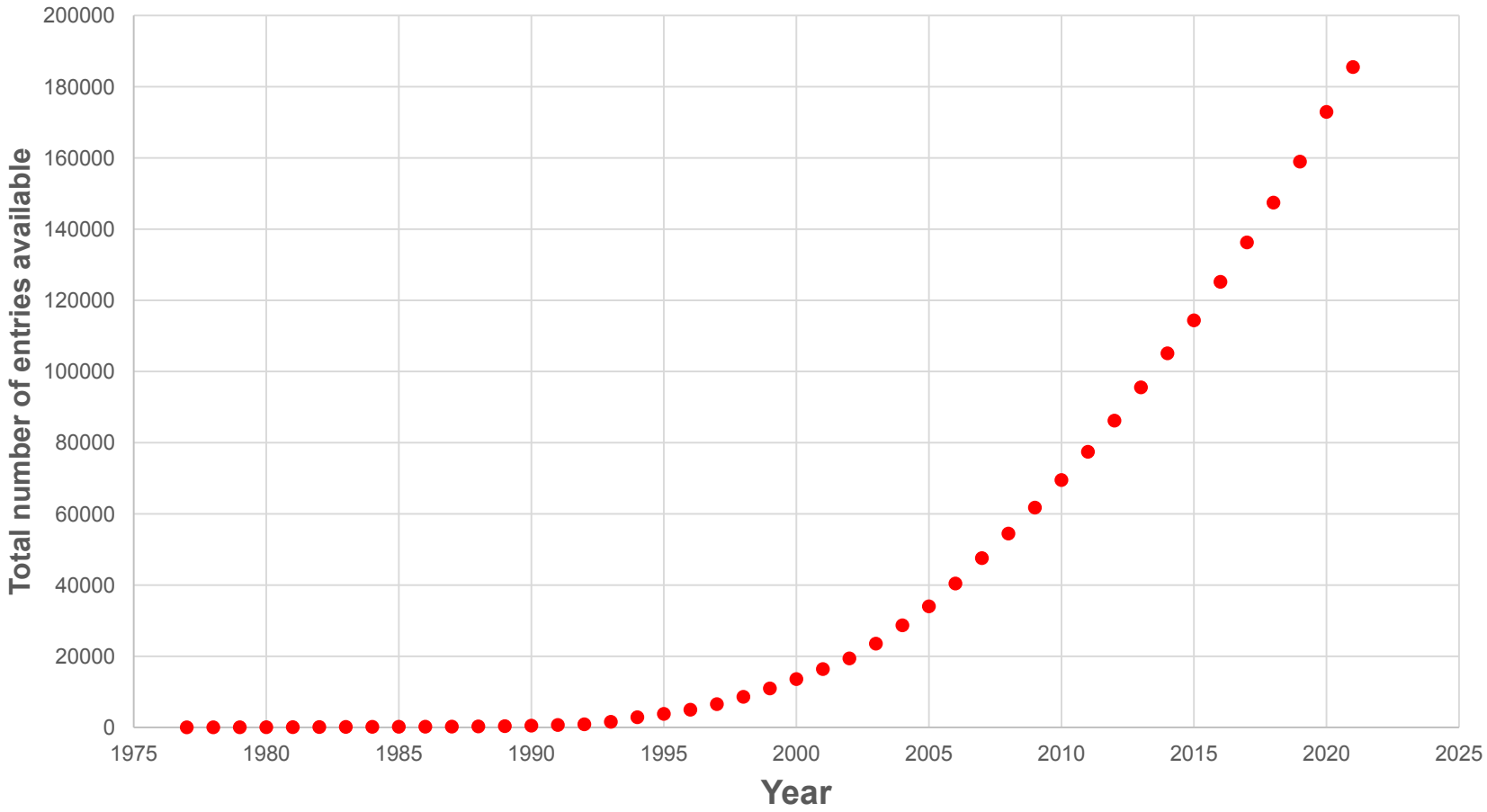




Modern life sciences

TRENDS

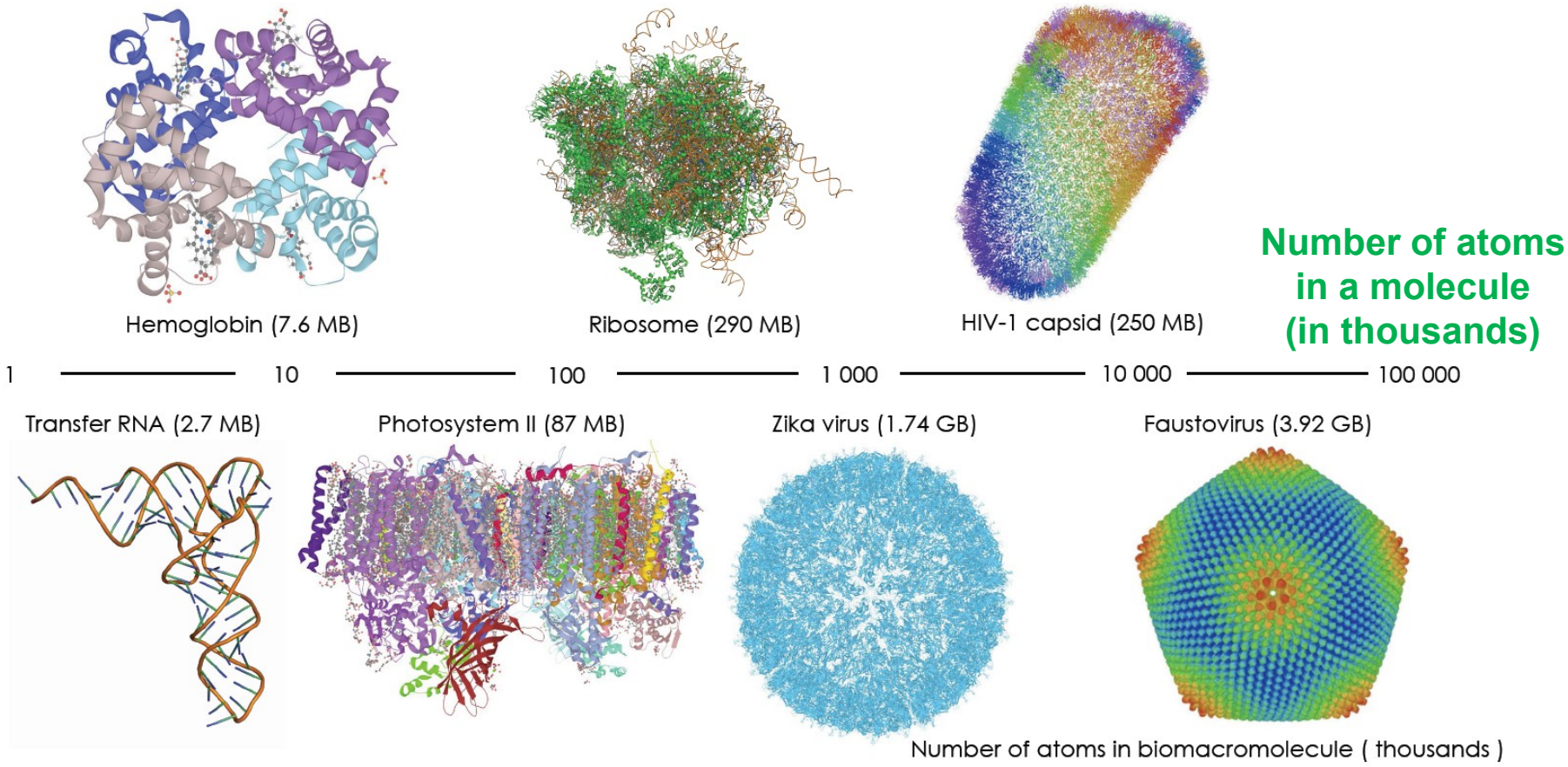
Number of experimental protein structures grows



> 200k protein structures

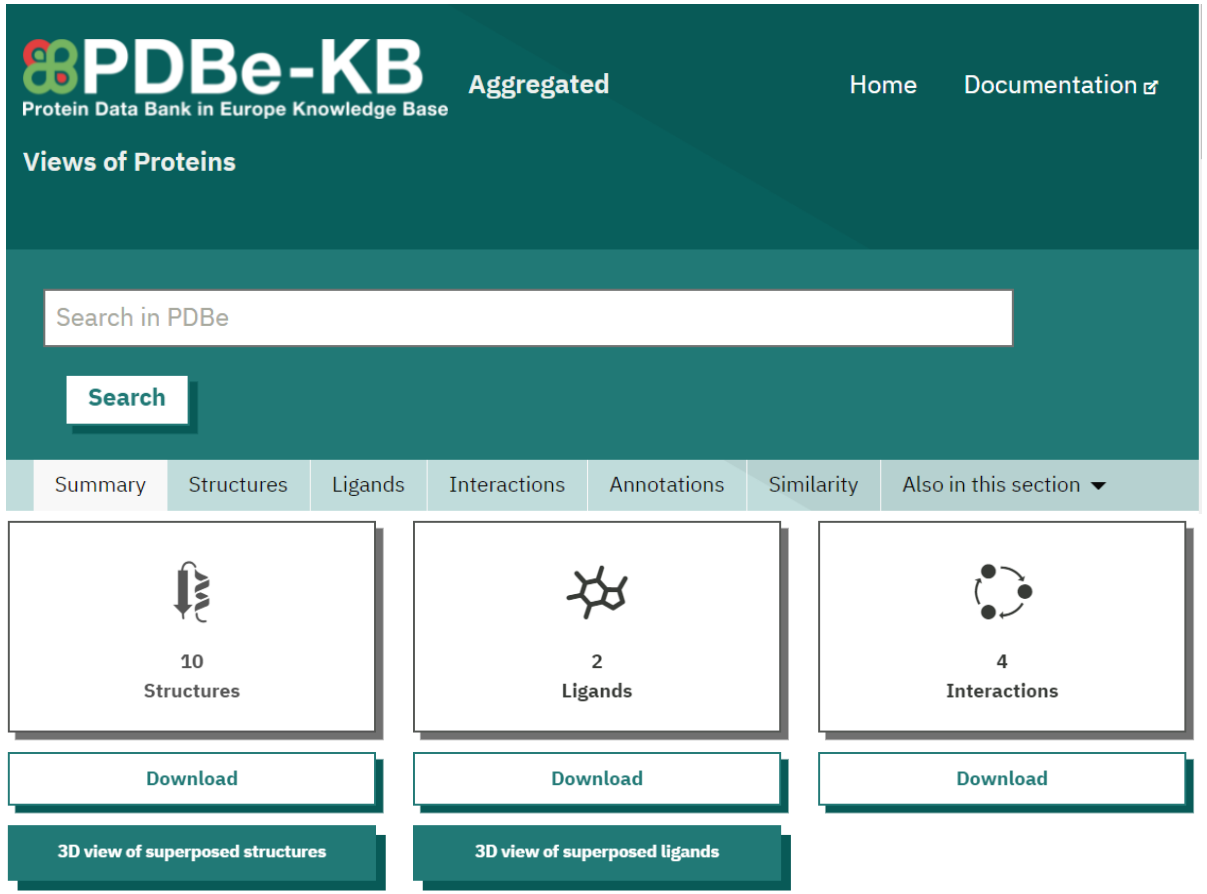
Modern life sciences
TRENDS

Size of experimental protein structures grows



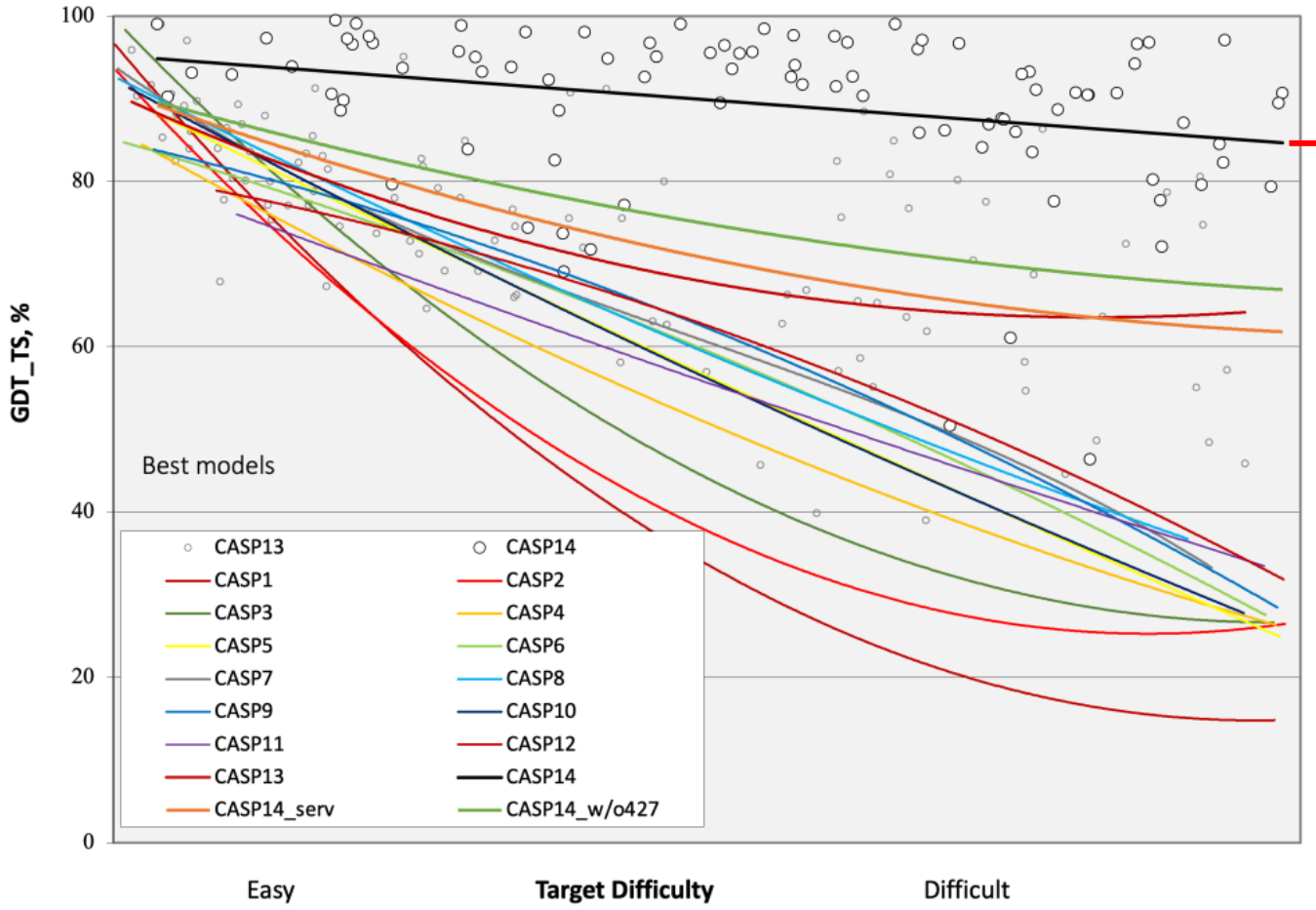
Modern life sciences
TRENDS

Not only structures, also properties



Varadi M., …, Svobodova R., … et al,., 2022. *PDBe-KB: collaboratively defining the biological context of structural data*. **Nucleic Acids Research**, 50(D1), D534-D542.

Modern life sciences
TRENDS
Structures generated by artificial intelligence

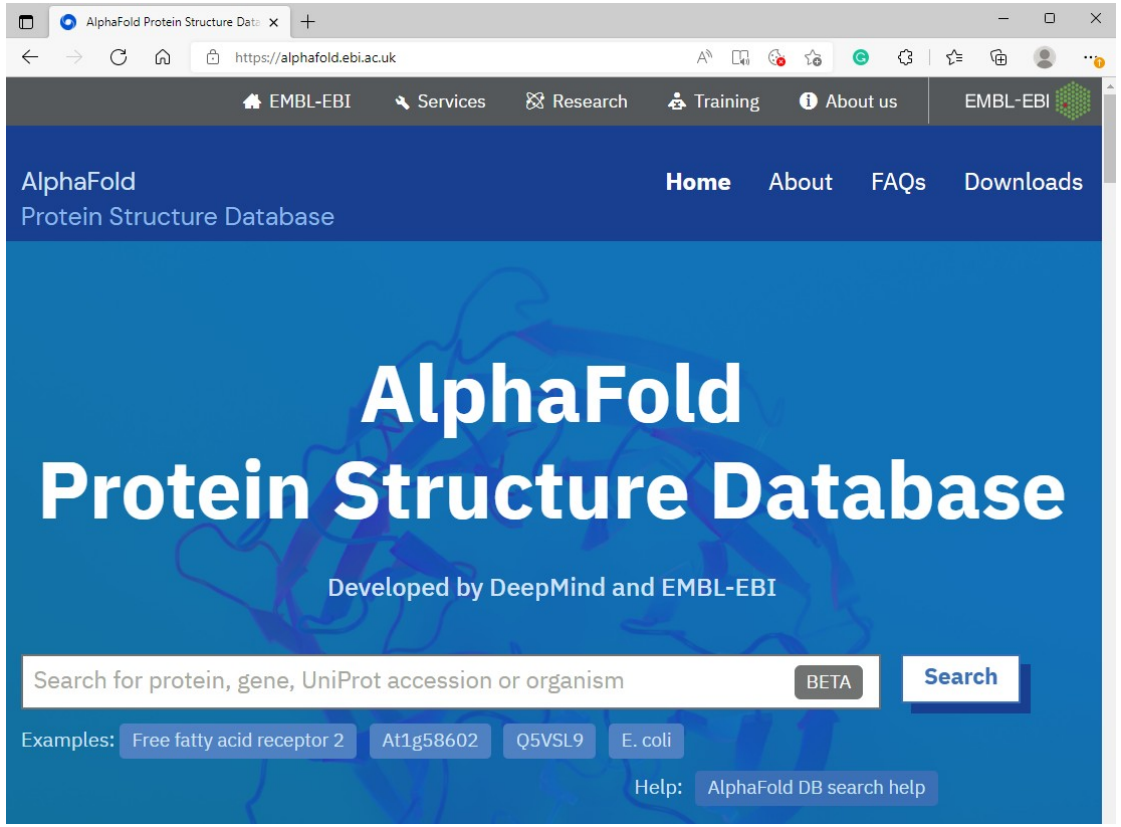


Structure prediction challenge 2020: AlphaFold2 wins

https://predictioncenter.org/casp14/doc/presentations/2020_11_30_CASP14_Introduction_Moult.pdf

Modern life sciences
TRENDS

Structures generated by artificial intelligence



> 200M protein structures

Modern life sciences
TRENDS

> 1 G structures of small molecules

PubChem

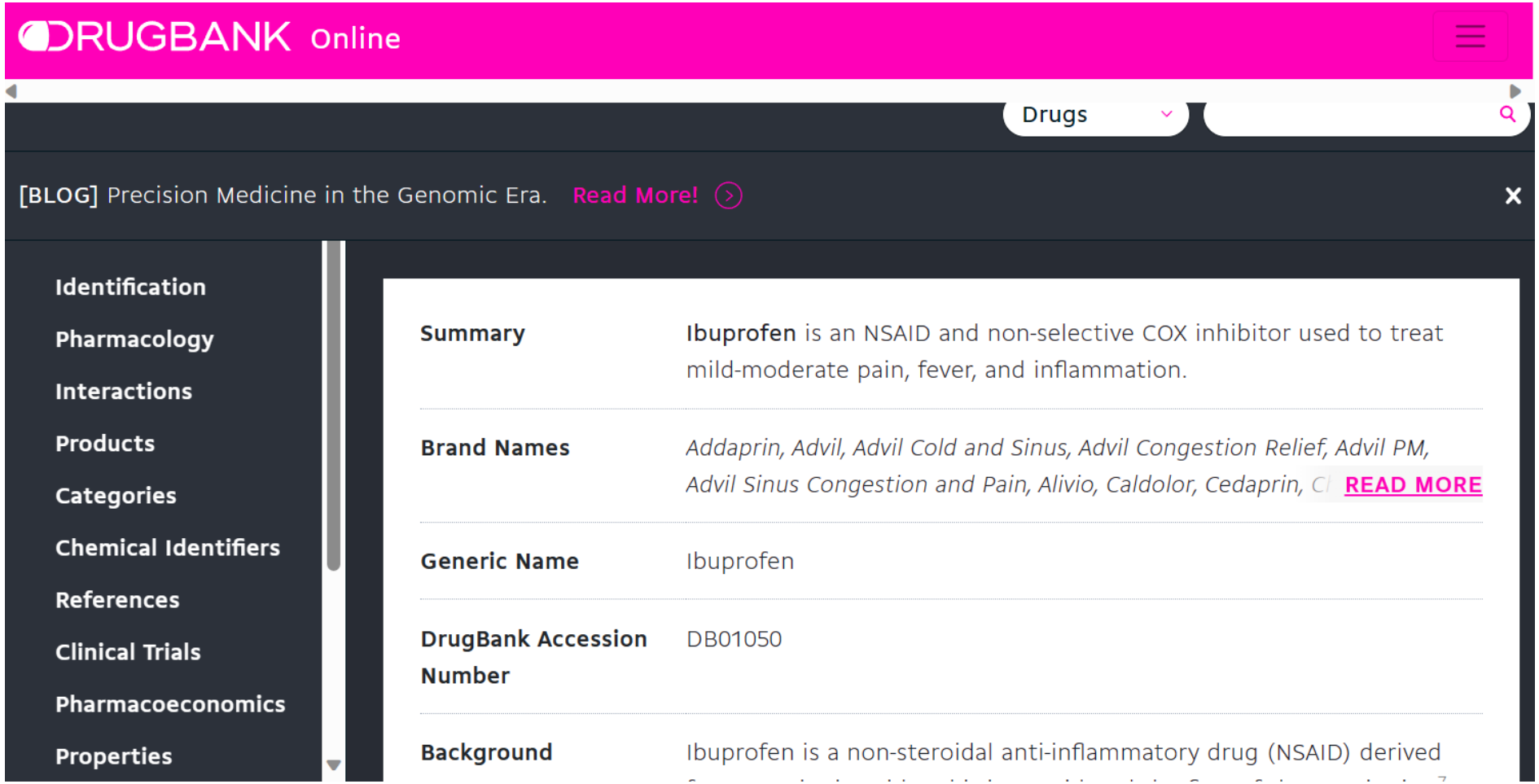
ZINC

ChEMBL

DRUGBANK

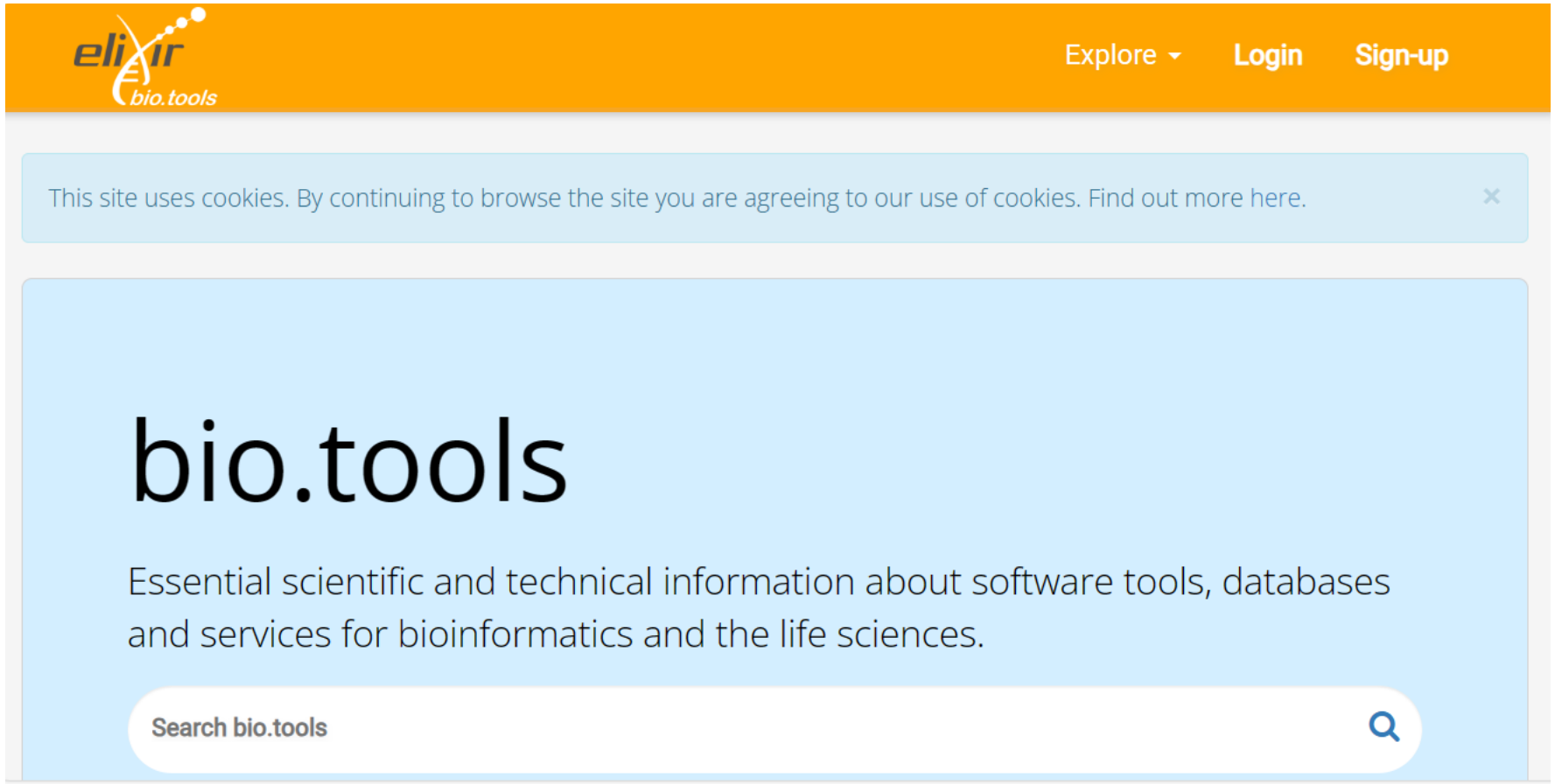
Modern life sciences
TRENDS

Not only structures, also properties



Modern life sciences
TRENDS

Web based databases and tools



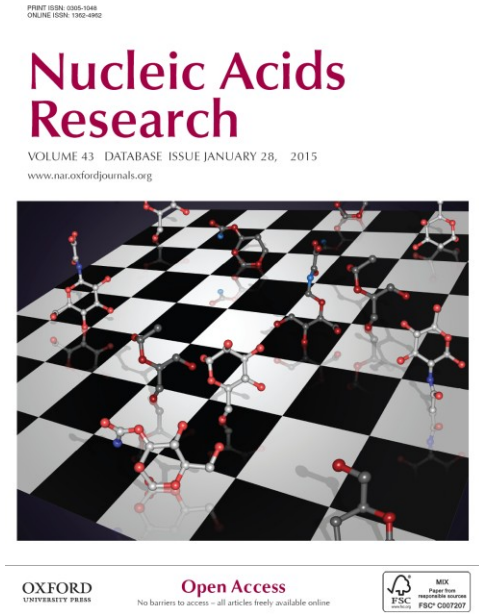
~ 30k tools and databases

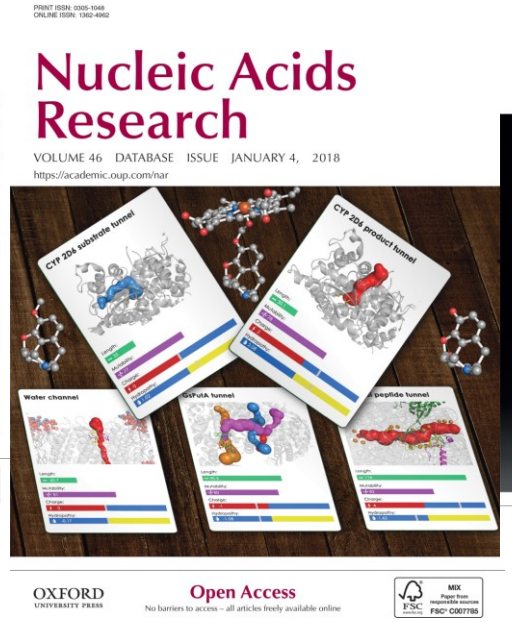
Modern life sciences

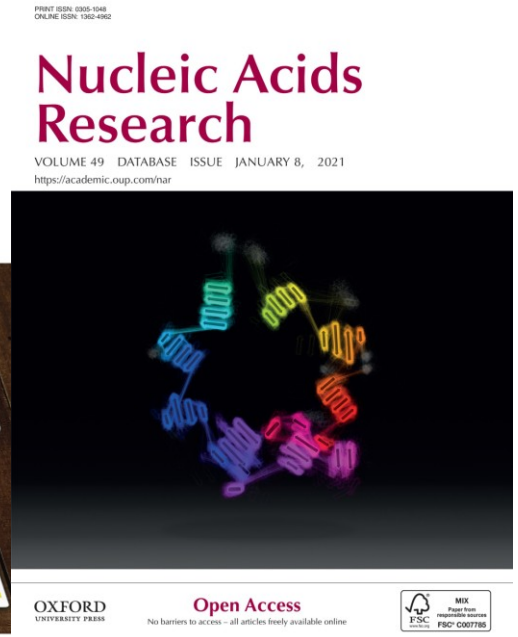
TRENDS

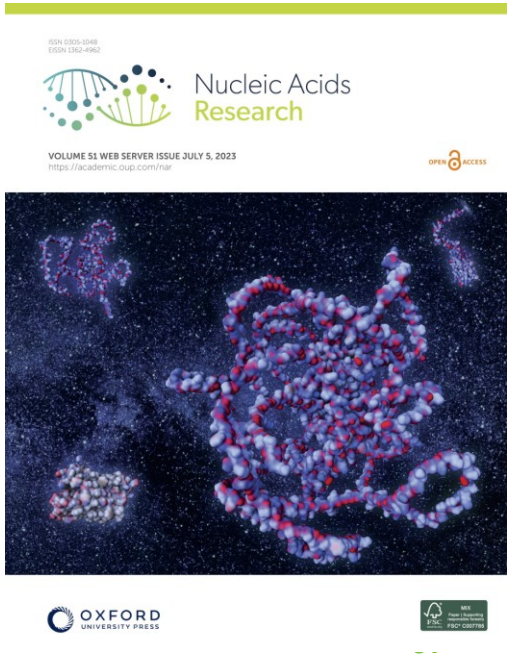
Tools from our research group :-):

- Validation: ValidatorDB, MotiveValidator, ValTrendsDB
- Detection: MoleOnline, ChannelsDB, PatternQuery
- Properties: Atomic Charge Calculator, aCharges
- Visualization: Mol*, LiteMol, 2DProts, OverProt





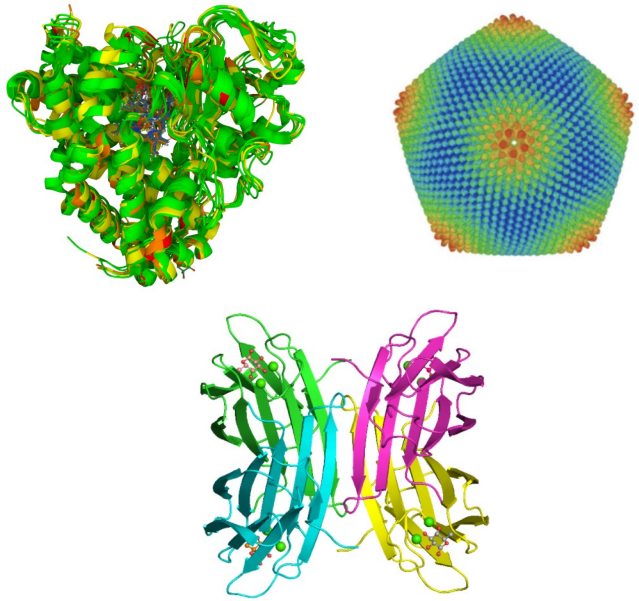




Conclusions

Great era!
We can enjoy it :-)))

Bioinformatics



Tools
Databases
AI

Chemoinformatics