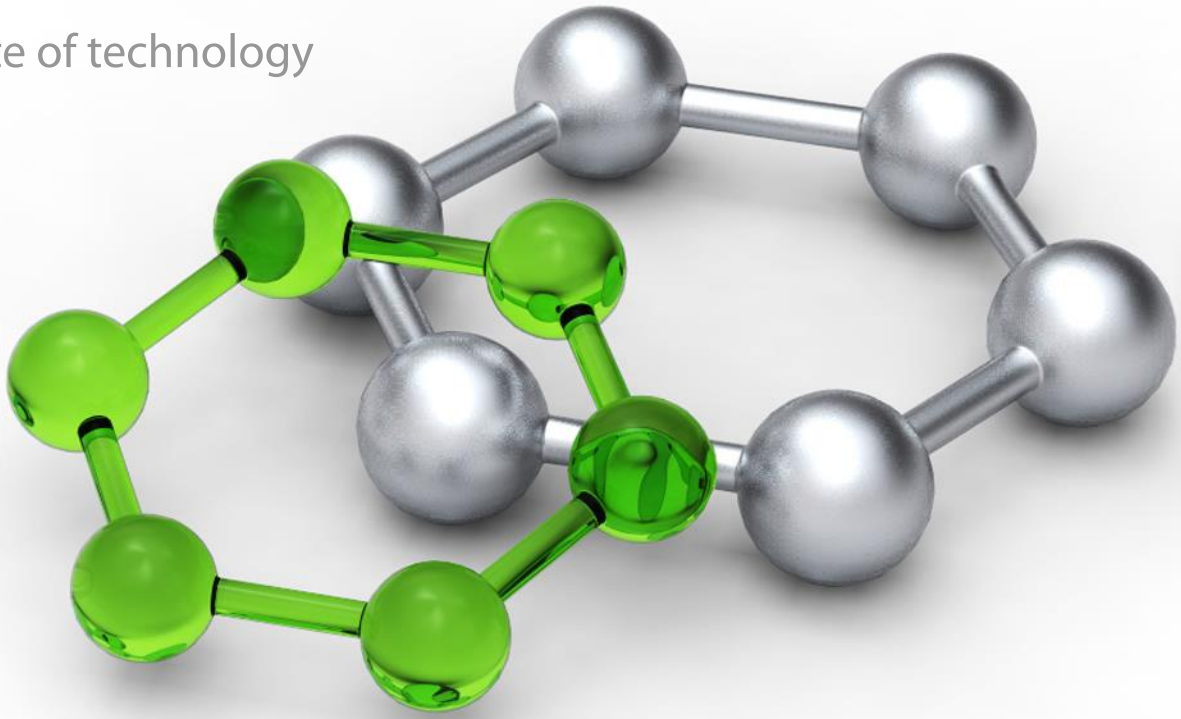
CEITEC

central european institute of technology

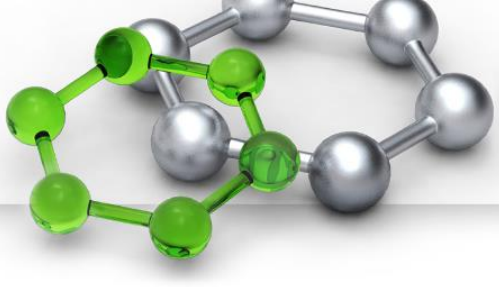
BRNO | CZECH REPUBLIC



Úvod do chemoinformatiky

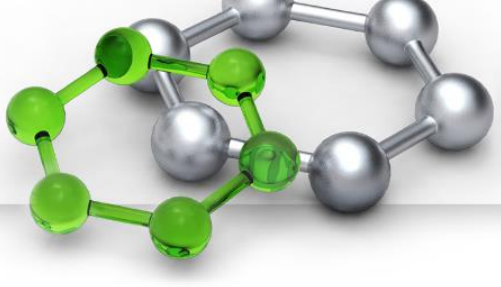
Radka Svobodová

Proč nahrazovat nebo doplňovat experiment výpočtem?

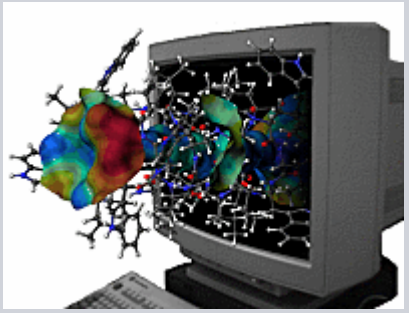


- Vyhneme se práci s toxickými, výbušnými a radioaktivními látkami
- Můžeme pracovat i s nestabilními látkami
- Ušetříme náklady za chemikálie a za realizaci experimentu
- Ušetříme čas experimentálním chemikům :-)
- ...

Chemoinformatika



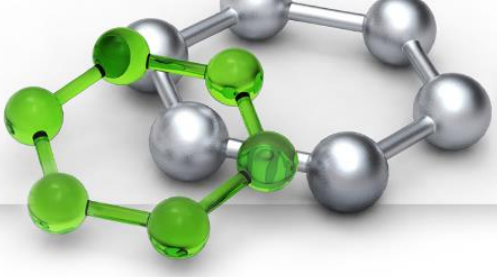
„Chemoinformatika využívá informatických a algoritmických přístupů pro řešení chemických problémů. Převážně se zaměřuje na získání informací z databází malých nebo středně velkých molekul (léků, organických látek, ...).“

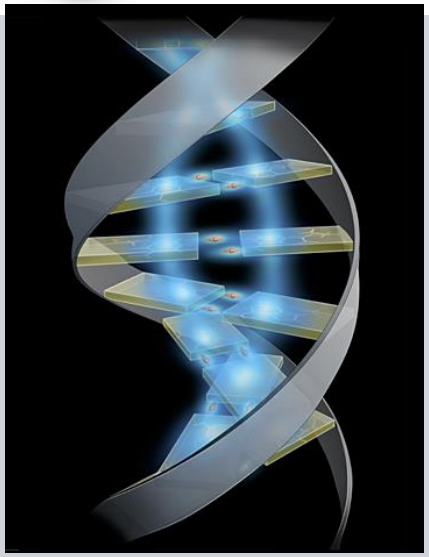


Vzniká v devadesátých létech dvacátého století.

Rozvoj spojen s dostupností velkého množství dat o molekulách léků apod. a s potřebami farmaceutického průmyslu

Bioinformatika

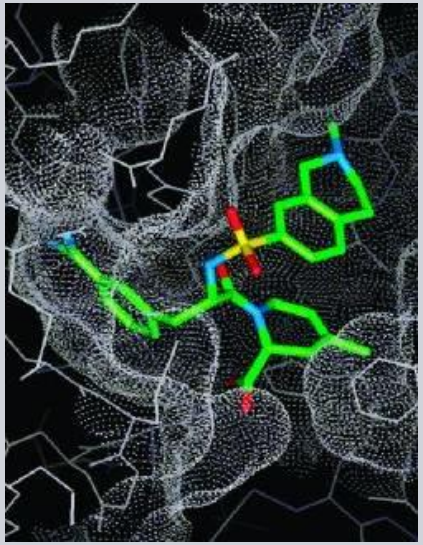




„Bioinformatika je aplikací statistických a výpočetních metod v oblasti molekulové biologie a genetiky. Převážně se specializuje na zpracování dat z databází biomolekul (sekvence DNA, RNA a proteinů).“

Vzniká v osmdesátých létech dvacátého století.

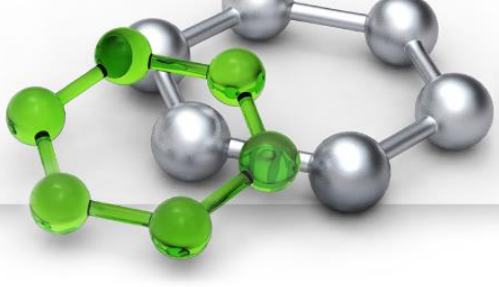
Rozvoj spojen s dostupností nadkritického množství dat o biomolekulách.



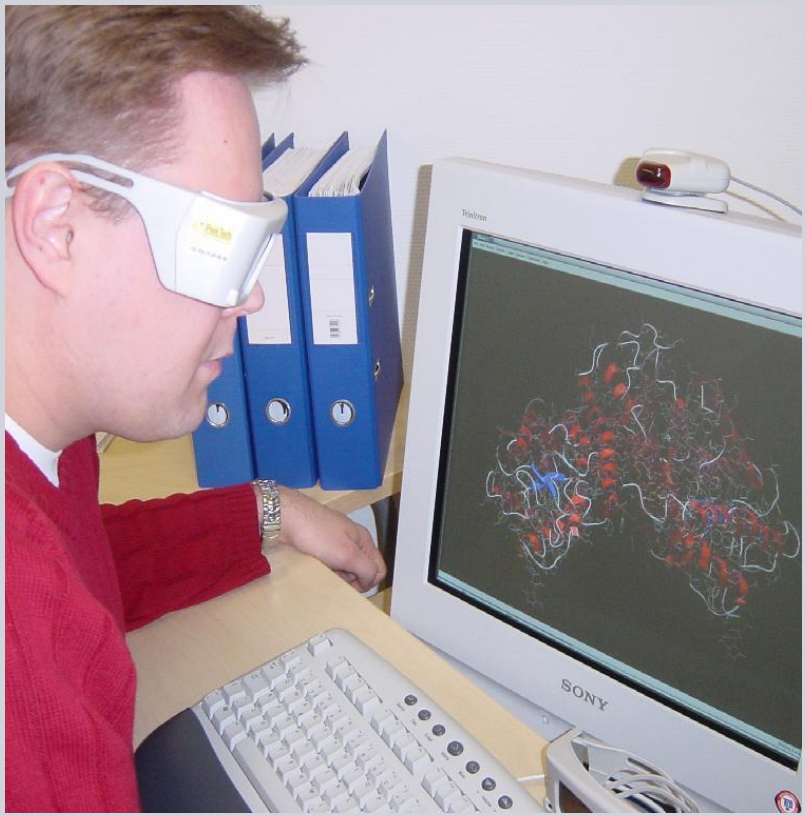
Specializace:

- Strukturní bioinformatika
- Sekvenční bioinformatika, NGS

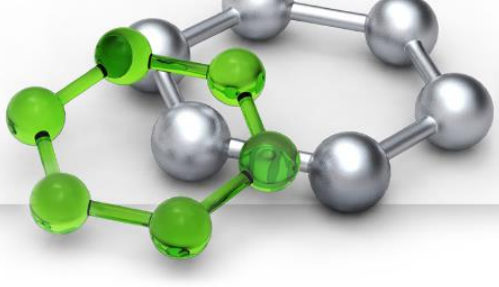
Jak zapsat molekulu v počítači?



- Zjistit, které informace molekulu popisují
- Zapsat je do počítače



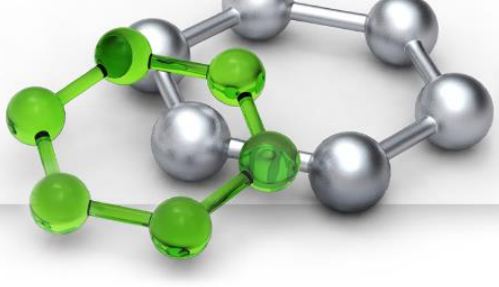
Které informace popisují molekulu?



Počty atomů?

Ethanol	C_2H_6O
Dimethyl ether	C_2H_6O

Které informace popisují molekulu?

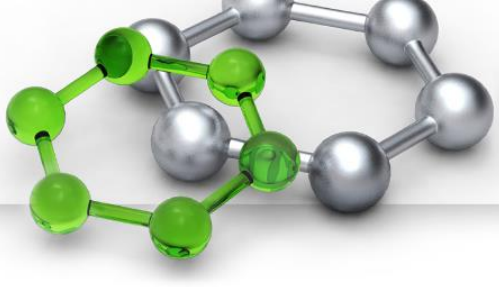


Počty atomů? **Málo**

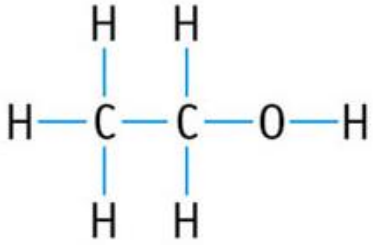
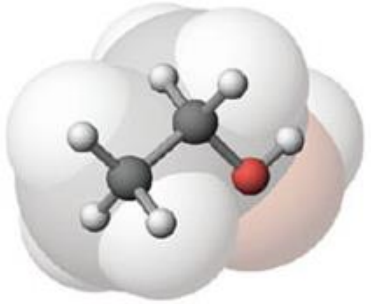
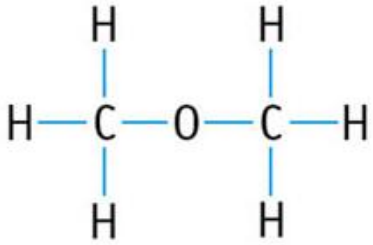
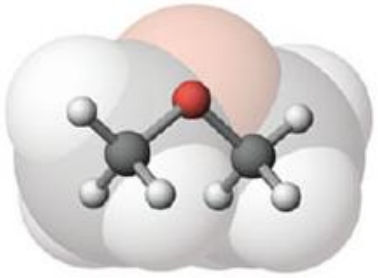
Počty atomů a umístění vazeb?

Ethanol	C_2H_6O	CH_3CH_2OH
Dimethyl ether	C_2H_6O	CH_3OCH_3

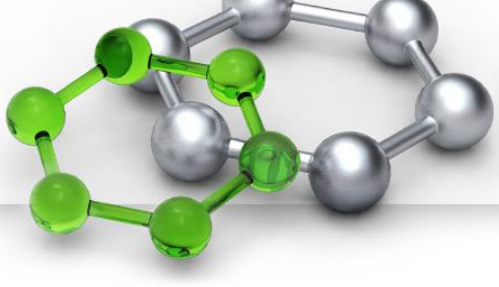
Které informace popisují molekulu?



Počty atomů?	**Málo**
Počty atomů a umístění vazeb?	**Lepší**
Počty atomů, umístění vazeb a poloha atomů v prostoru?	**Ano**

Ethanol	C_2H_6O	CH_3CH_2OH		
Dimethyl ether	C_2H_6O	CH_3OCH_3		

Model molekuly pro počítačové zpracování



Atomy:

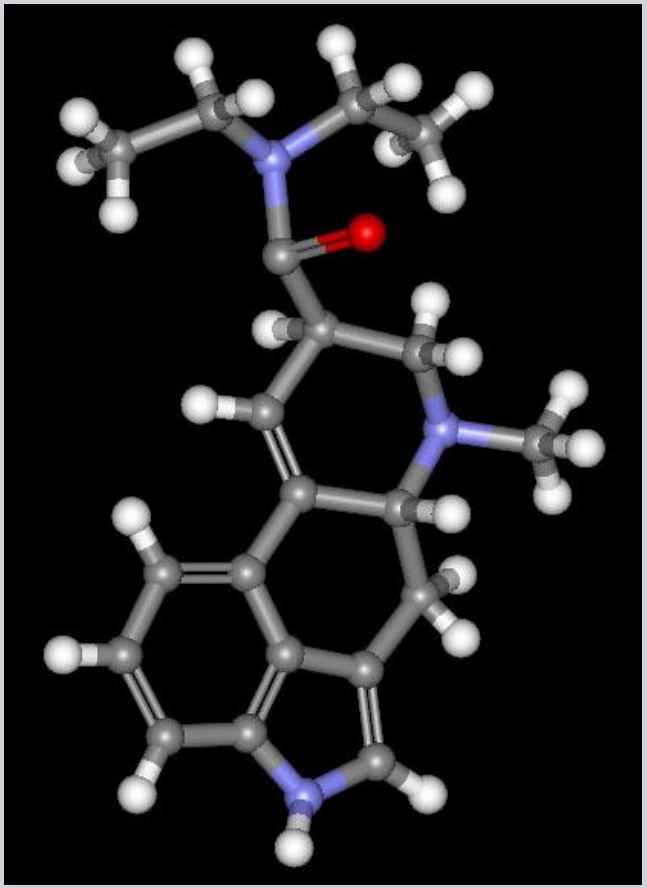
Body v prostoru

U každého uveden chemický symbol prvku

Vazby:

Dvojice atomů, které jsou vázány

Násobnost vazby



Zápis molekuly v počítači

Počet vazeb

Počet atomů

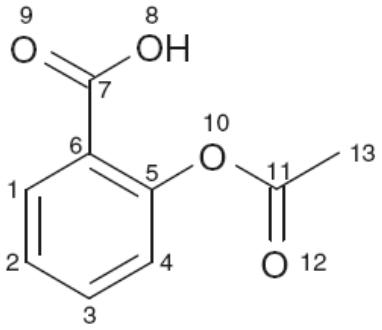
První atom je uhlík

```
  -ISIS-  09270222202D

 13 13  0  0  0  0  0  0  0  0999 V2000
   -3.4639   -1.5375    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.4651   -2.3648    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7503   -2.7777    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.0338   -2.3644    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.0367   -1.5338    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7521   -1.1247    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7545   -0.2997    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.0413    0.1149    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -3.4702    0.1107    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -1.3238   -1.1186    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.6125   -1.5292    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.6167   -2.3542    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    0.1000   -1.1125    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  2  0  0  0  0
  6  7  1  0  0  0  0
  3  4  2  0  0  0  0
  7  8  1  0  0  0  0
  7  9  2  0  0  0  0
  4  5  1  0  0  0  0
  5 10  1  0  0  0  0
  2  3  1  0  0  0  0
 10 11  1  0  0  0  0
  5  6  2  0  0  0  0
 11 12  2  0  0  0  0
  6  1  1  0  0  0  0
 11 13  1  0  0  0  0
M END
```

První tři čísla jsou x, y a z souřadnice atomů

První vazba je mezi atomy 1 a 2 a jde o dvojnou vazbu



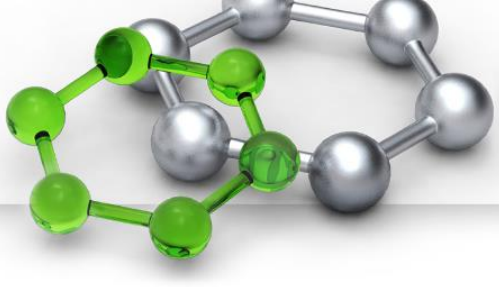
```
  ACD/Labs08061122363D

 21 21  0  0  0  0  0  0  0  0  1 V2000
   18.7769  -15.2504   -0.1032 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.7571  -16.6359   -0.1252 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.5868  -14.5409   -0.1114 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.5465  -17.3106   -0.1545 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.3767  -15.2158   -0.1421 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.3559  -16.6013   -0.1633 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.6081  -13.0313   -0.0880 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.0592  -14.5322   -0.0715 N   0  0  0  0  0  0  0  0  0  0  0  0
   17.5247  -18.7799   -0.1764 N   0  0  0  0  0  0  0  0  0  0  0  0
   15.1150  -14.4620   -0.1527 N   0  0  0  0  0  0  0  0  0  0  0  0
   20.0742  -13.3140   -0.0089 O   0  0  0  0  0  0  0  0  0  0  0  0
   21.1073  -15.1564   -0.0523 O   0  0  0  0  0  0  0  0  0  0  0  0
   16.4750  -19.3759    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   18.5697  -19.4030   -0.2650 O   0  0  0  0  0  0  0  0  0  0  0  0
   14.0496  -15.0560   -0.1515 O   0  0  0  0  0  0  0  0  0  0  0  0
   15.1330  -13.2425   -0.1568 O   0  0  0  0  0  0  0  0  0  0  0  0
   19.7111  -17.2054   -0.1194 H   0  0  0  0  0  0  0  0  0  0  0  0
   15.3860  -17.1427   -0.1873 H   0  0  0  0  0  0  0  0  0  0  0  0
   17.6136  -12.6451   -1.1298 H   0  0  0  0  0  0  0  0  0  0  0  0
   16.7057  -12.6567    0.4410 H   0  0  0  0  0  0  0  0  0  0  0  0
   18.5209  -12.6823    0.4410 H   0  0  0  0  0  0  0  0  0  0  0  0
  2  1  1  0  0  0  0
  3  1  2  0  0  0  0
  4  2  2  0  0  0  0
  5  3  1  0  0  0  0
  6  4  1  0  0  0  0
  6  5  2  0  0  0  0
  3  7  1  0  0  0  0
  1  8  1  0  0  0  0
  4  9  1  0  0  0  0
  5 10  1  0  0  0  0
  8 11  2  0  0  0  0
  8 12  2  0  0  0  0
  9 13  2  0  0  0  0
  9 14  2  0  0  0  0
 10 15  2  0  0  0  0
 10 16  2  0  0  0  0
 17  2  1  0  0  0  0
 18  6  1  0  0  0  0
 19  7  1  0  0  0  0
 20  7  1  0  0  0  0
 21  7  1  0  0  0  0
M  END
```

Kvízová otázka:

Nakresli tuto molekulu. Jak se daná molekula jmenuje?

Databáze malých (organických) molekul



> 1 miliard struktur malých molekul

- Experimentální struktury
- Predikované struktury

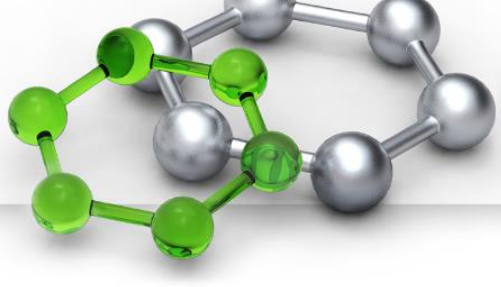
PubChem

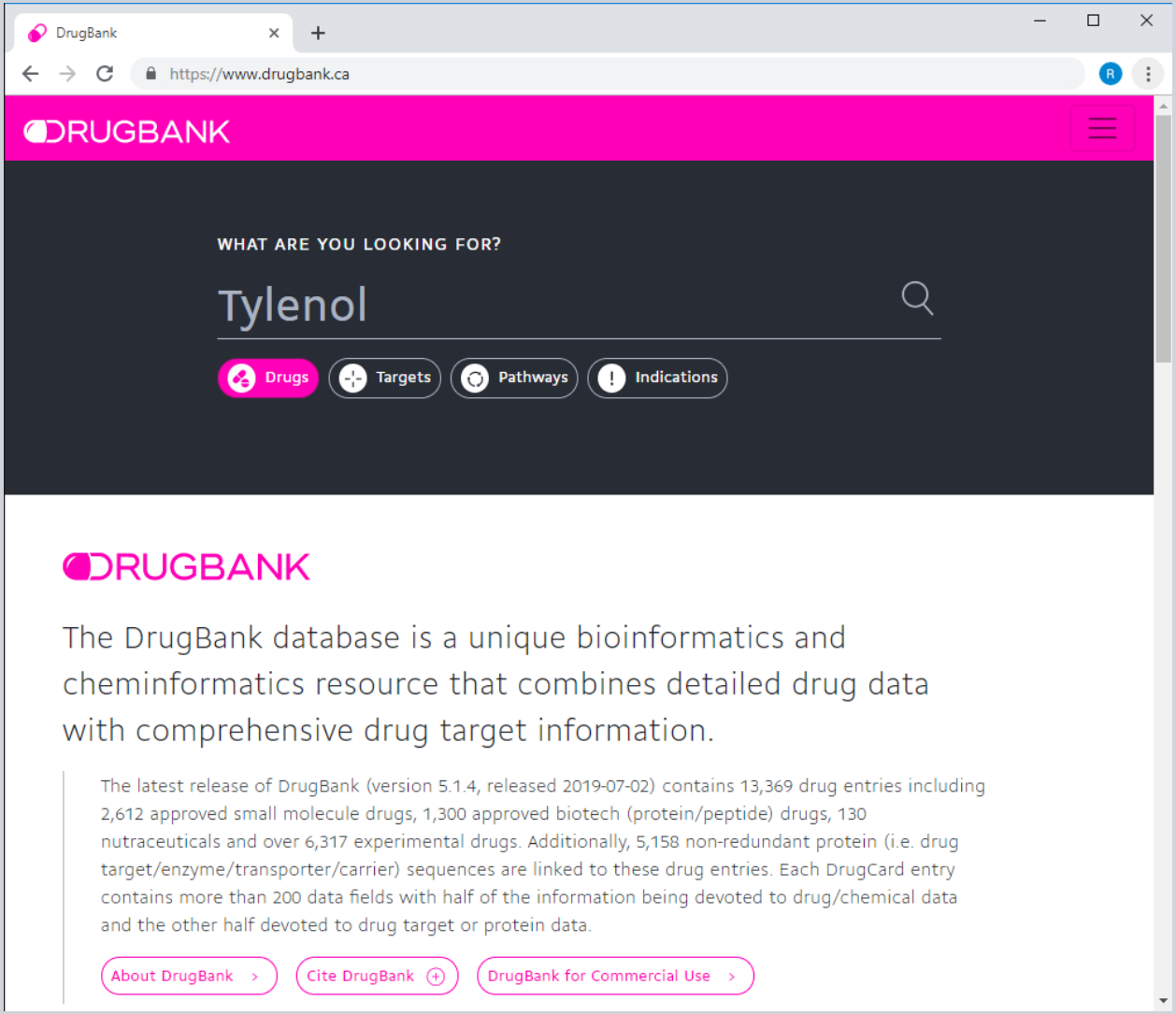
ZINC

ChEMBL

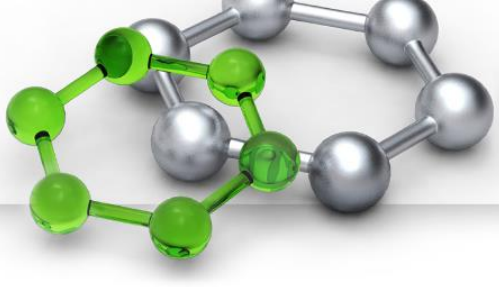
DRUGBANK

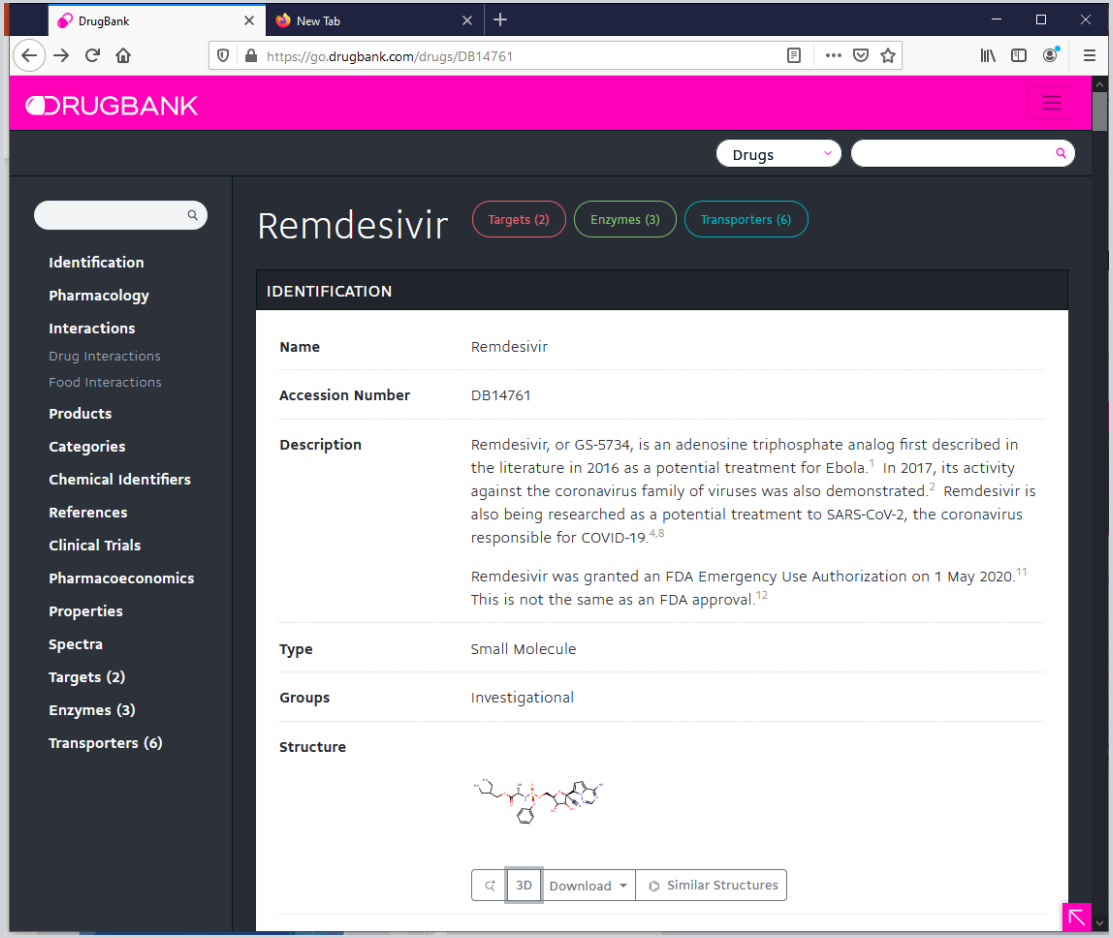
DrugBank – ukázka databáze léků

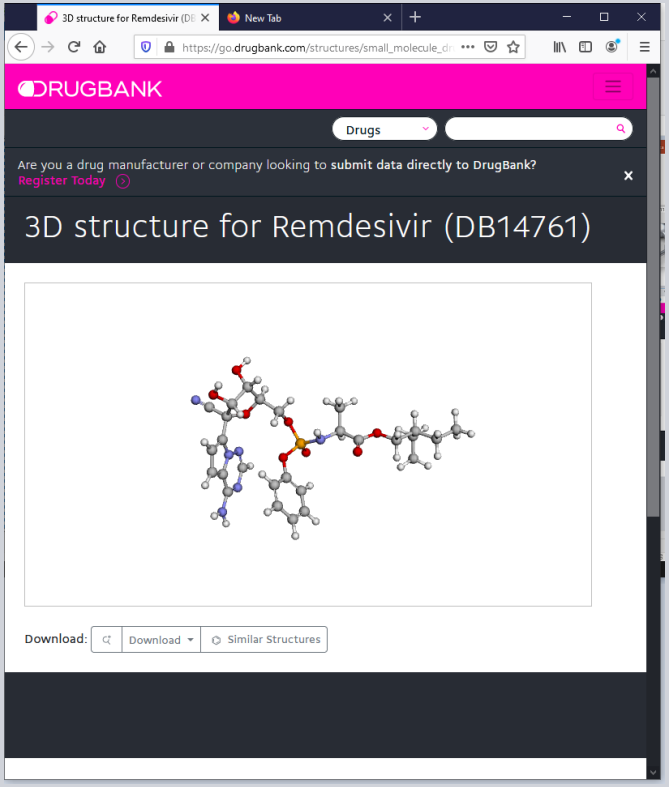




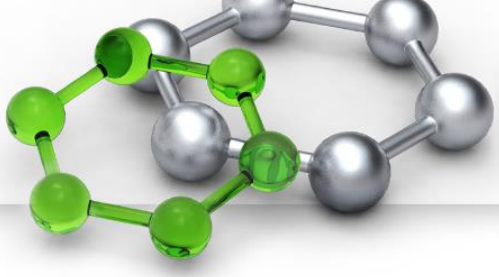
DrugBank – ukázka databáze léků

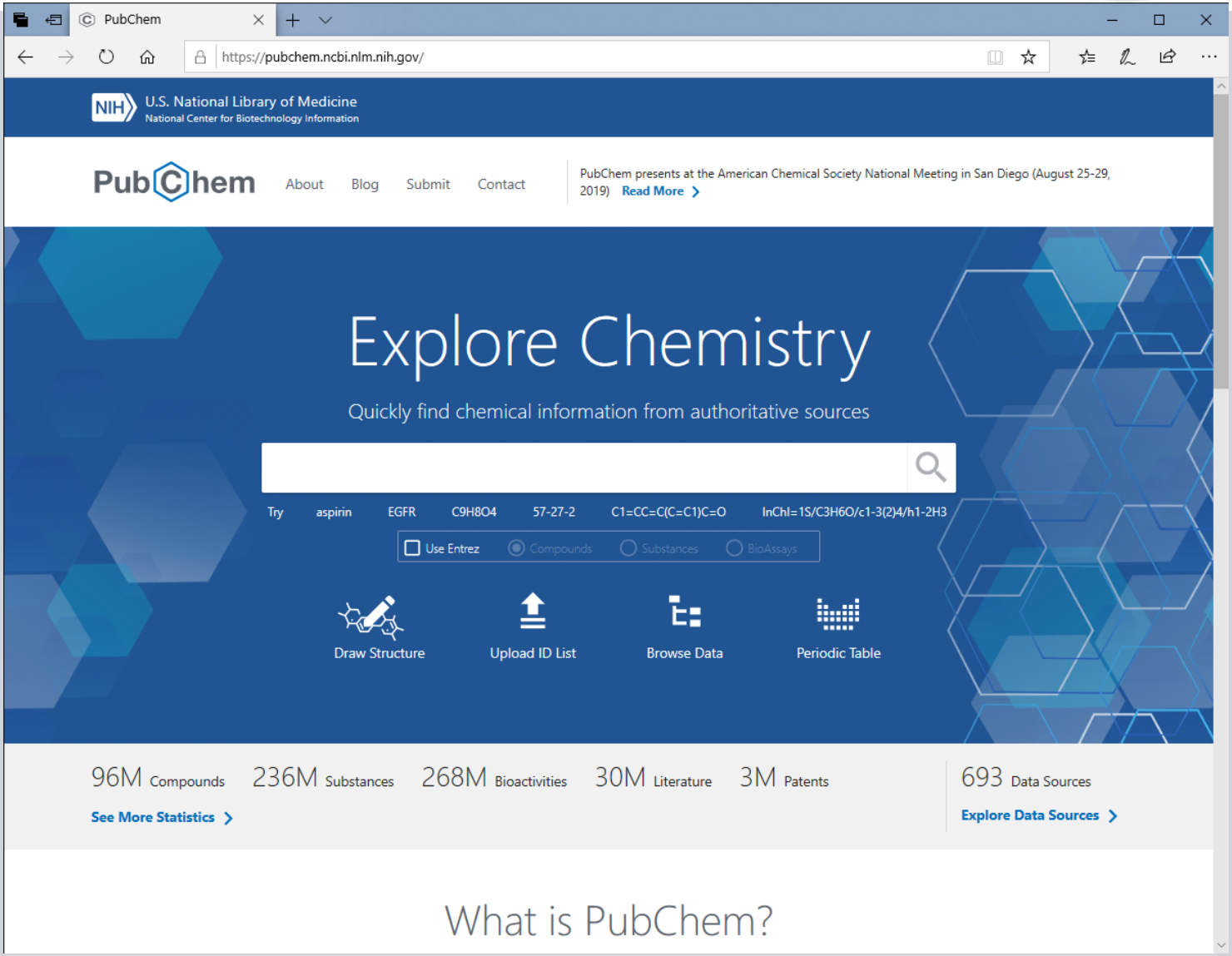




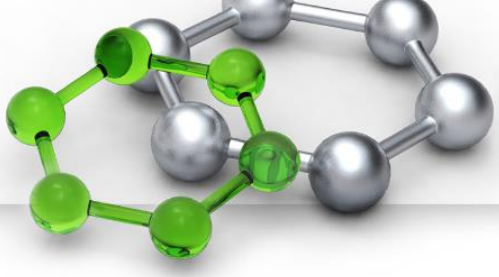


PubChem – ukázka databáze organických molekul





PubChem – ukázka databáze organických molekul



https://pubchem.ncbi.nlm.nih.gov/compound/5288826

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov.
Get the latest research from NIH: https://www.nih.gov/coronavirus.

National Library of Medicine
National Center for Biotechnology Information

PubChem About Blog Submit Contact Search PubChem

COMPOUND SUMMARY

Morphine

Share Tweet Email

Cite Download

PubChem CID: 5288826

Structure: 2D 3D Crystal

Find Similar Structures

Chemical Safety: Irritant

CONTENTS

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Pharmacology and Biochemistry

Ligand Expo – ukázka databáze ligandů

ligand-expo.rcsb.org/ld-search.html

RCSB PDB PROTEIN DATA BANK

RCSB PDB | Contact Us

Ligand Expo

Home | Search | Browse | Download | Ligand Expo Help

Chemical Component Search Tools

Use the forms below to search for chemical components within the PDB Component Dictionary.

- Search for chemical components by *3-letter component identifier code, molecular name, molecular formula, SMILES description, or InChi/InChiKey chemical description*.
  You can also check to see if a 3-letter code is being held by a deposition in progress.
- Chemical substructure searches can also be conducted by *starting from a chemical drawing* created within the MarvinSketch tool.
  Either start with a SMILES description or chemical data file (see drop-menu for acceptable formats), or draw a 2D chemical structure from scratch (**Launch** without input). It can also generate chemical component definitions from your 2D structure.
- Search for *instances* of a chemical component throughout the PDB. The **Display** option allows you to simply see a list of PDB codes, or to download these coordinates in PDB, MOL/SDF and mmCIF formats.
- You can also search for analogs to the standard amino acids, nucleotides, popular drugs, and common aromatic ring systems by using the *Browse* feature in the top menu bar.

Your query results are also searchable! Each hit from your initial query will contain links to continue searching by similar name, chemical formula, or structure (SMILES).

MOLECULAR NAME, FORMULA, AND DESCRIPTOR SEARCH OPTIONS

Search term: chlorophyll Search type: Molecular name (exact sub-string)

Go

SKETCH INPUT AND/OR STRUCTURE SEARCH OPTIONS

File name: Browse... No file selected. File format: MOL/SDF format

-- OR --

SMILES string: Display size: Small

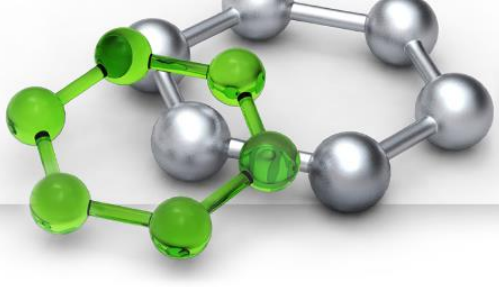
Launch

SEARCH FOR INSTANCES OF CHEMICAL COMPONENTS BY 3-LETTER ID CODE

Component ID code: Display: PDB entry codes + coordinate files

Go

Ligand Expo – ukázka databáze ligandů



ligand-expo.rcsb.org/reports/B/BCL/index.html

RCSB PDB PROTEIN DATA BANK

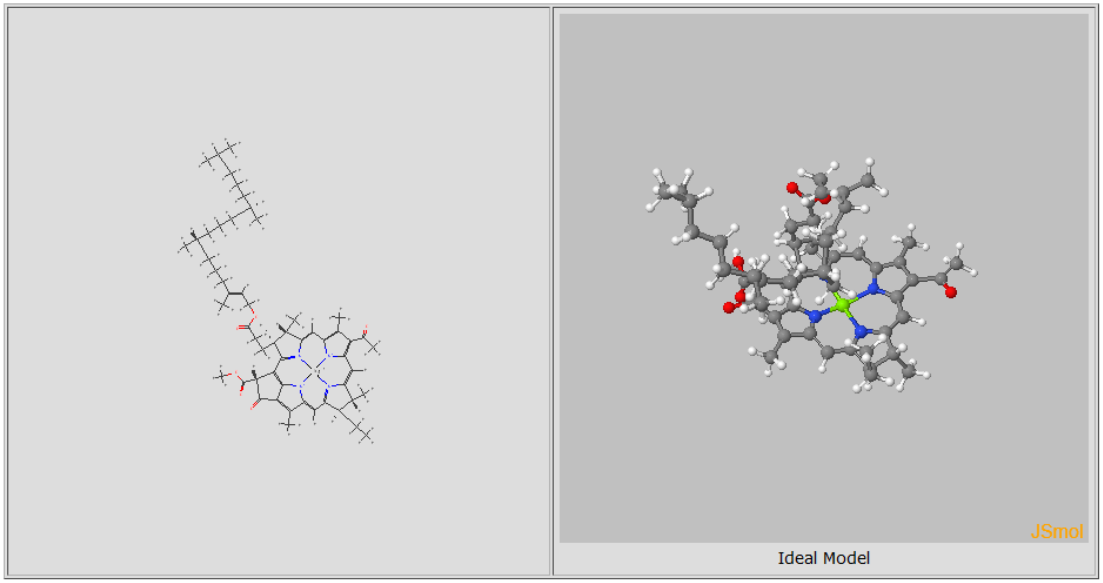
RCSB PDB | Contact Us

Ligand Expo

Home | Search | Browse | Download | Ligand Expo Help

Chemical Details | Geometry | Atom Nomenclature | Downloads | Related Resources

PDB Chemical Component BCL

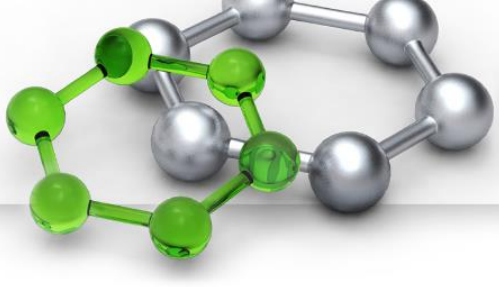


Ideal Model

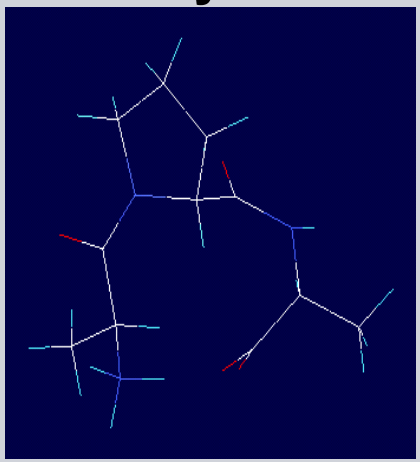
Chemical Description

Name	BACTERIOCHLOROPHYLL A
Formula	C55 H74 Mg N4 O6
Formal charge	0
Molecular weight	911.504 g/mol
Component type	NON-POLYMER
Ambiguous Chemistry Warning	The chemical description of this component is not well described in this definition. Descriptors and chemical names should be used with caution.

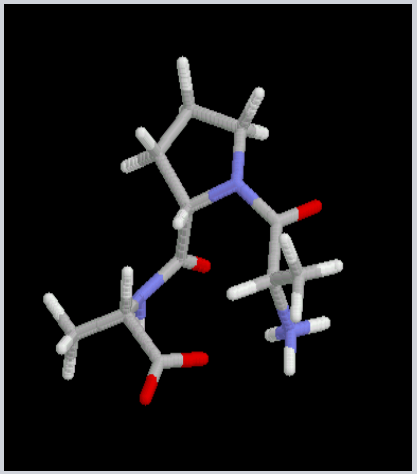
Vizualizace malé molekuly v počítači



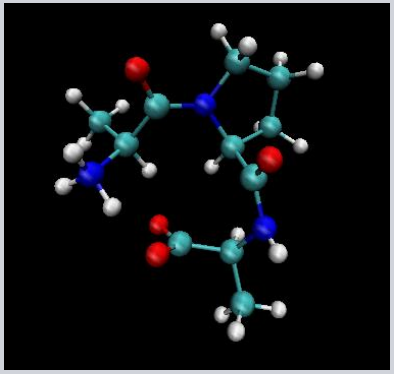
Drátový model:



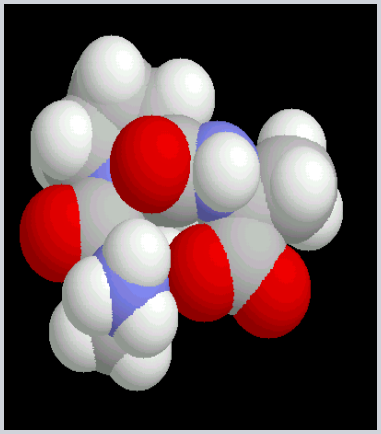
Tyčinkový model:



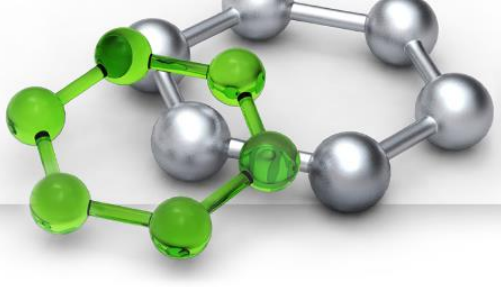
Tyčinky a kuličky:



Kalotový model (CPK):



Databáze biomakromolekul (hlavně proteiny)



> 200 tisíc experimentálních struktur
> 200 millionů predikovaných struktur

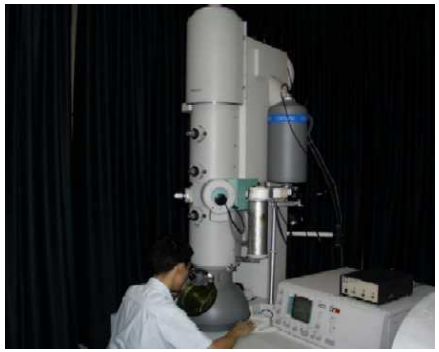
WORLDWIDE PDB
PROTEIN DATA BANK

AlphaFold
Protein Structure Database
Developed by DeepMind and EMBL-EBI

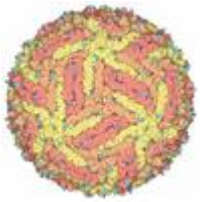
Protein Data Bank – zdroje dat

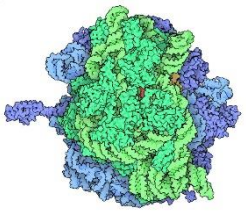
89% Rentgenová krystalografie

10% NMR spektroskopie



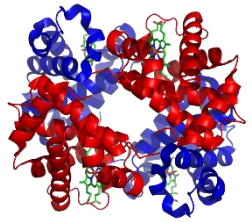
1% kryo-elektronová mikroskopie

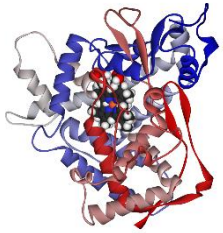




3D struktura

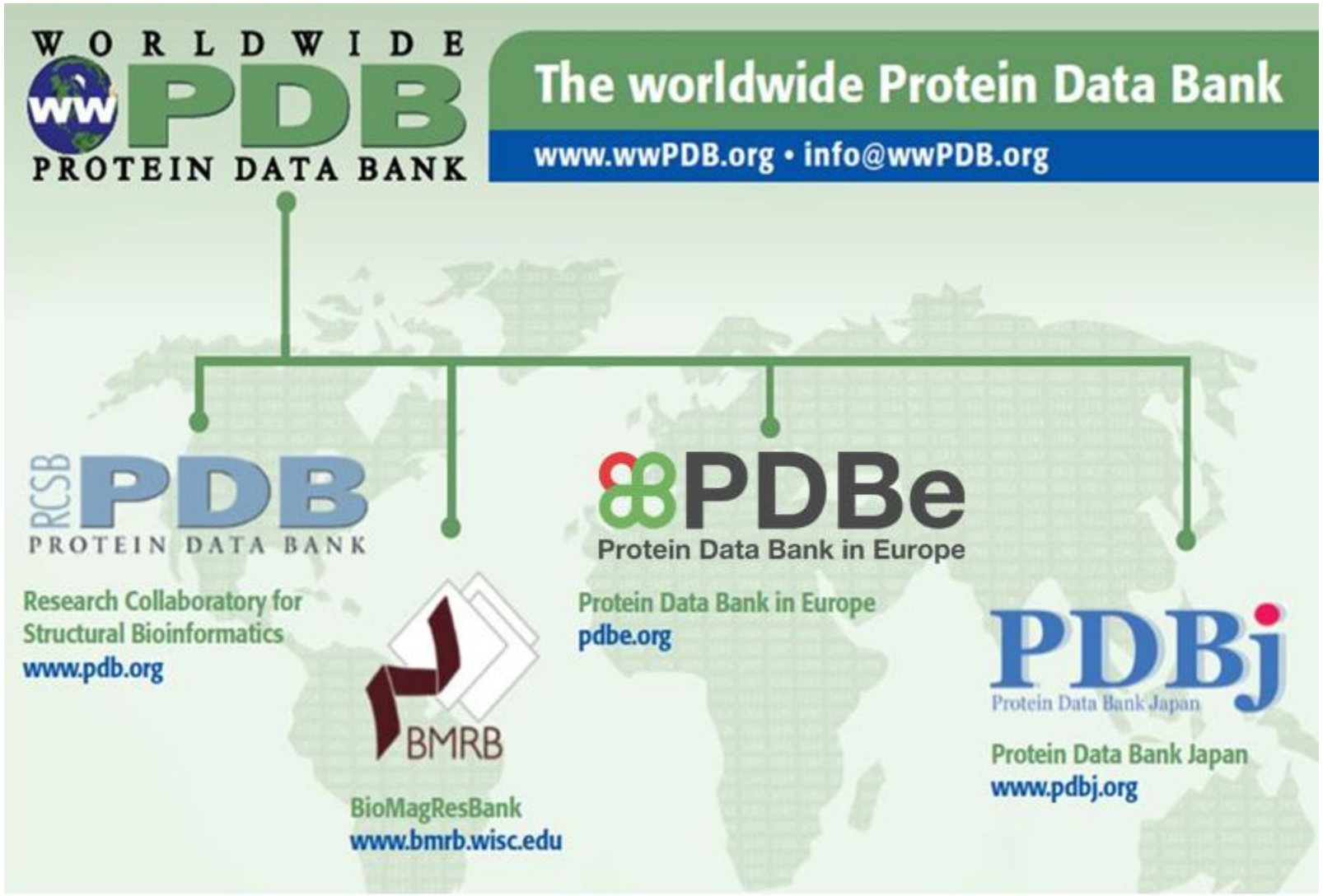
```
...
ATOM    46  C   GLY A  70      51.536  23.360  40.507
ATOM    47  O   GLY A  70      50.947  22.279  40.325
ATOM    48  N   ILE A  71      50.965  24.532  40.270
ATOM    49  CA  ILE A  71      49.595  24.644  39.786
...
```





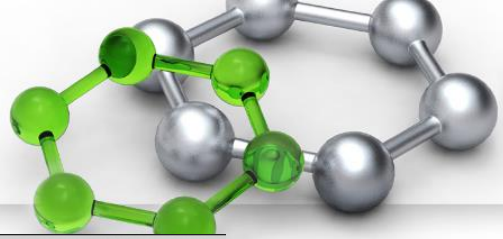
WORLDWIDE PDB
PROTEIN DATA BANK

Protein Data Bank



> 200 000 biomacromolecular structures

Protein Data Bank – ukázka databáze proteinů



PDBe home < EMBL-EB

European Bioinformatics Institute [GB] https://www.ebi.ac.uk/pdbe/

EMBL-EBI

Services | Research | Training | About us

Protein Data Bank in Europe

Bringing Structure to Biology

Search

Examples: hemoglobin, BRCA1_HUMAN

Advanced search

PDBe home | Deposition | PDBe services | PDBe training | Documentation | About PDBe

Share | Feedback

PDBe is the European resource for the collection, organisation and dissemination of data on biological macromolecular structures. Read more about PDBe.

Featured structure

Pygmalion 5x2g

1st September 2019

The September image in our 2019 calendar is inspired by a molecular system that can edit DNA and the story of a statue coming to life.

Read more...

Previous featured structures

Popular

- PDBe-KB
- EMsearch
- PDBeFold
- PDBePISA
- PDBeChem
- Sequence search
- PDBe REST API
- EM resources
- NMR resources
- EMPIAR
- Coordinate Server
- PDB Component Library
- News
- Events
- Training
- Contact us

News

Links added to raw experimental data at PDBe

12 August, 2019

Improve your previously released PDB coordinates with OneDep

1 August, 2019

A celebration of the PDB Art project

26 July, 2019

Mandatory mmCIF format for crystallographic depositions to the PDB

1 July, 2019

More news

Events

Art Exhibition: Molecules of Life

Kendrew Foyer, EBI South Building Wellcome Genome Campus

10 Sep 2019 to 27 Sep 2019

EBI Structural bioinformatics course

EMBL-EBI, Cambridge, UK

16 Sep 2019 to 20 Sep 2019

EBI Exploring Biological Sequence course

EMBL-EBI, Cambridge, UK

10 Oct 2019

More events

Latest archive statistics

As of 11 September 2019 the PDB contains 155830 entries (latest PDB entries, chemistry, biology) and EMDB contains 9016 entries (latest map releases, latest header releases, latest updates).

Tweets by @PDBeurope

Protein Data Bank Retweeted

David Armstrong

@DaveASci

Really looking forward to giving this @PDBeurope training course Medellin in Colombia. Thanks to @raod85, @unubiolac and @CabanaGcrf for supporting this.

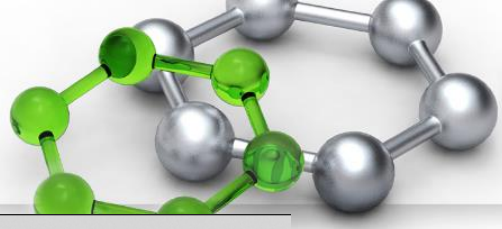
https://twitter.com/raod85/status/1173657623155154944

13h

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notice and Terms of Use.

I agree, dismiss this banner

Protein Data Bank – ukázka databáze proteinů



Search the PDB archive ‹ PD | PDB 3hyu structure sum

European Bioinformatics Institute [GB] https://www.ebi.ac.uk/pdbe/entry/pdb/3hyu

EMBL-EBI

Services | Research | Training | About us

Protein Data Bank in Europe
Bringing Structure to Biology

Search

Examples: hemoglobin, BRCA1_HUMAN

Advanced search

Feedback

PDBe › 3hyu

Crystal structure of the altitude adapted hemoglobin of guinea pig.

Source organism: *Cavia porcellus*

Primary publication:

Structure of the altitude adapted hemoglobin of guinea pig in the R2-state.

Pairet B, Jaenicke E

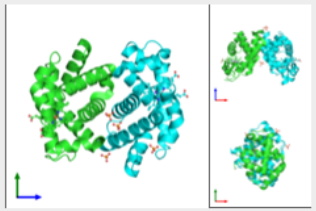
PLoS ONE **5** e12389 (2010)

PMID: 20811494

X-ray diffraction
1.67Å resolution

Released: 23 Jun 2010

Model geometry
Fit model/data



Quick links

- 3hyu overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation
- View
- Downloads
- 3D Visualisation

Function and Biology

Details

Biochemical function: heme binding

Biological process: oxygen transport

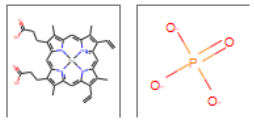
Cellular component: hemoglobin complex

Sequence domains:

- Haemoglobin, alpha-type
- Haemoglobin, beta-type
- Globin
- Globin/Protoglobin
- Globin-like superfamily

Ligands and Environments

2 bound ligands:



2 x HEM 4 x PO4

No modified residues

Structure analysis

Details

Assembly composition: hetero tetramer (preferred)

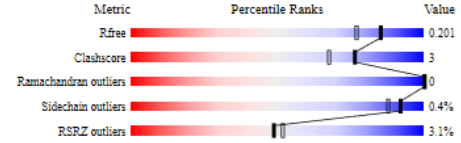
Entry contents: 2 distinct polypeptide molecules

Macromolecules (2 distinct):

Hemoglobin subunit alpha

Experiments and Validation

Details



PDB-REDO

The sliders below show the change in model quality between original PDB entry and the PDB-REDO entry

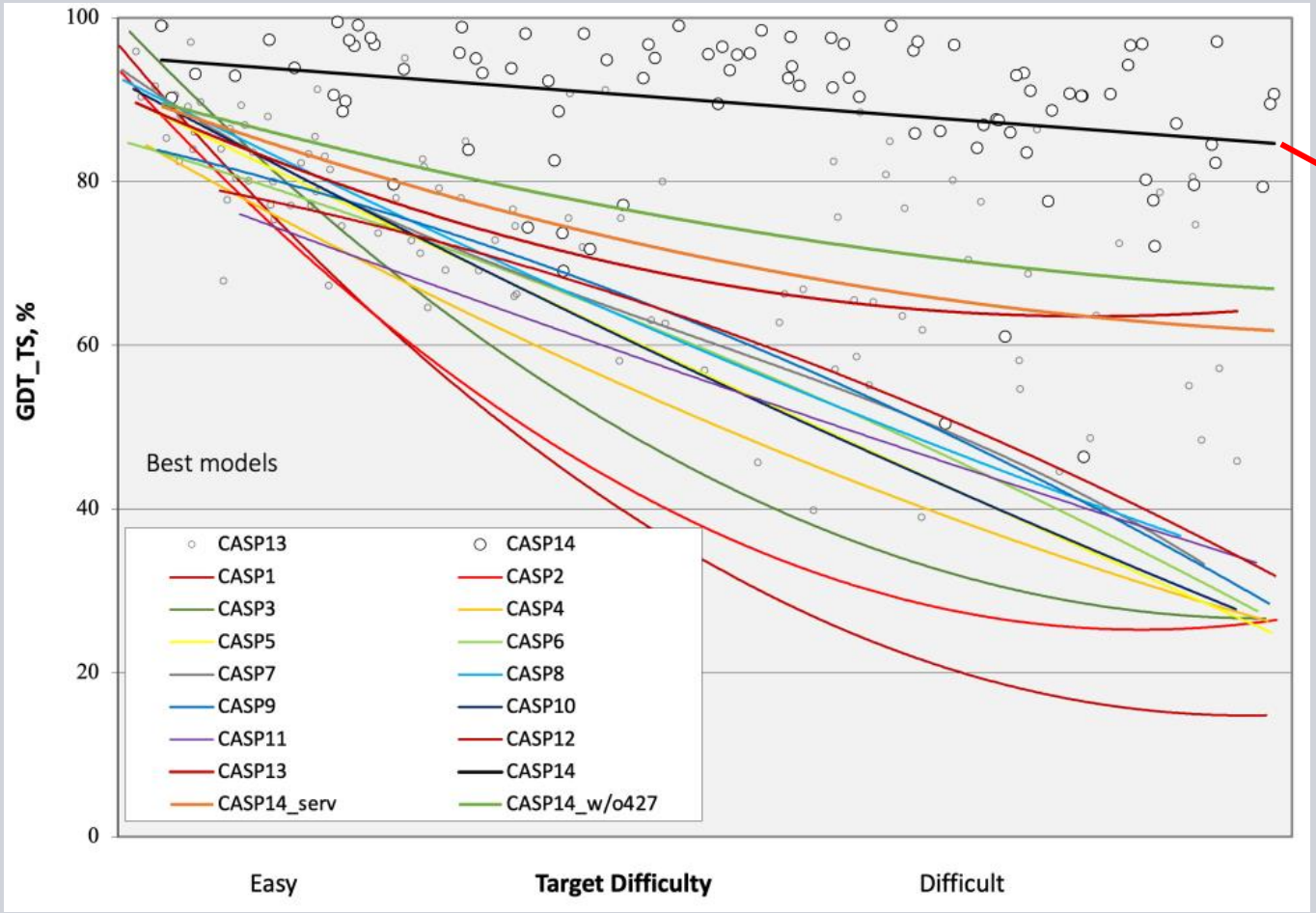
Model Geometry
Fit model/data

PDB-REDO

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notice and Terms of Use.

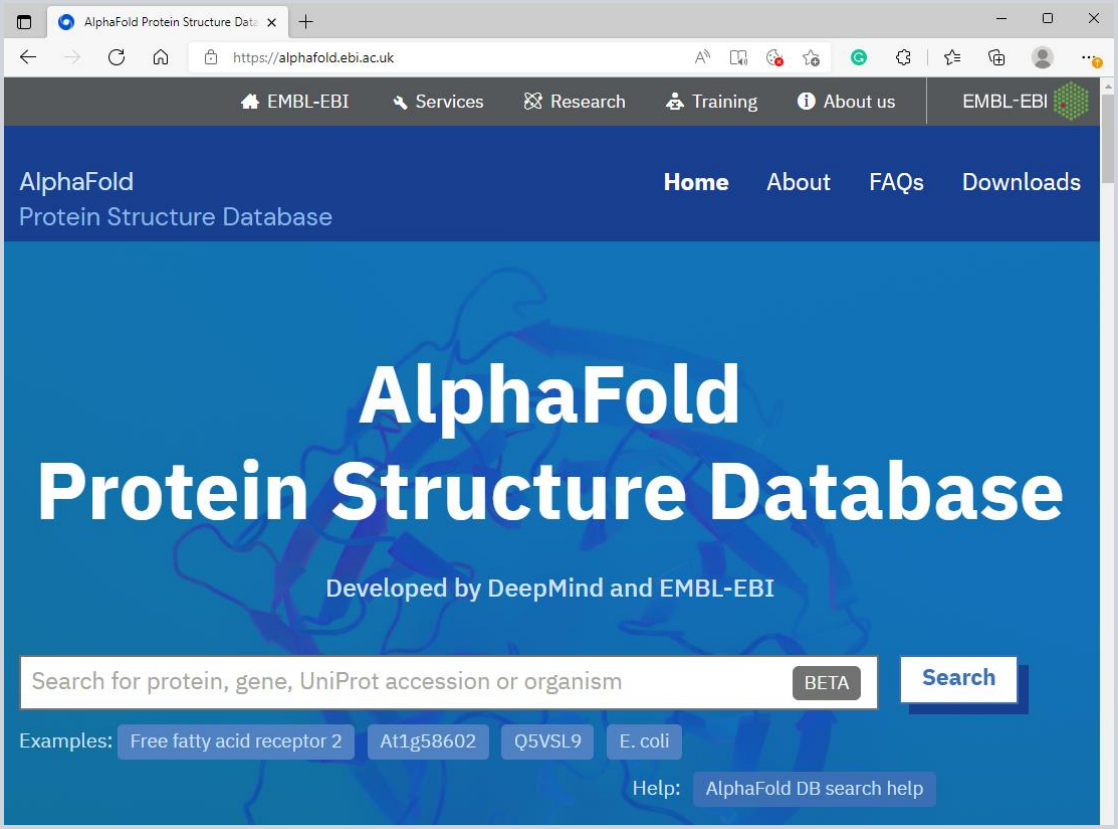
I agree, dismiss this banner

AlphaFold algoritmus



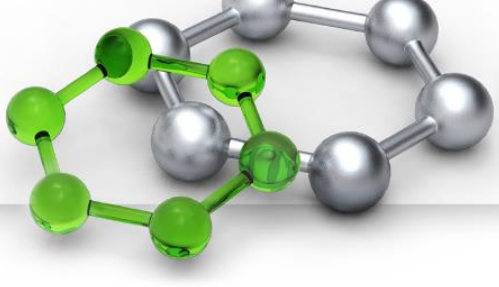
Structure prediction challenge 2020: AlphaFold2 wins

AlphaFold DB

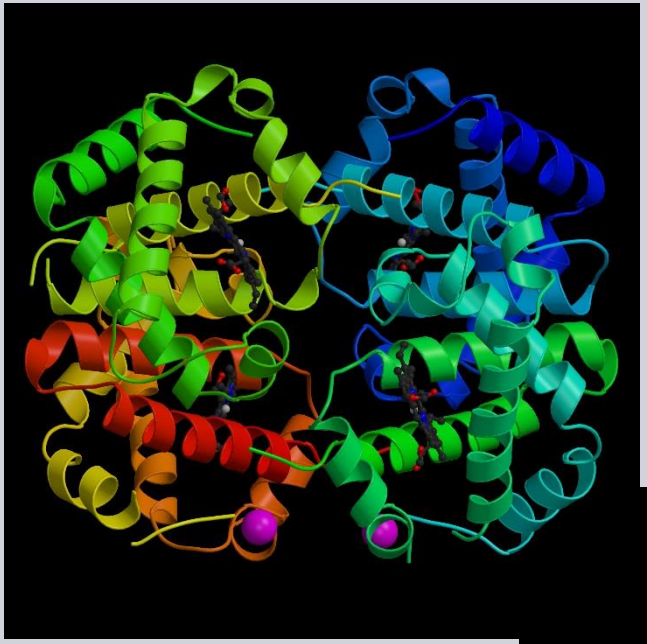


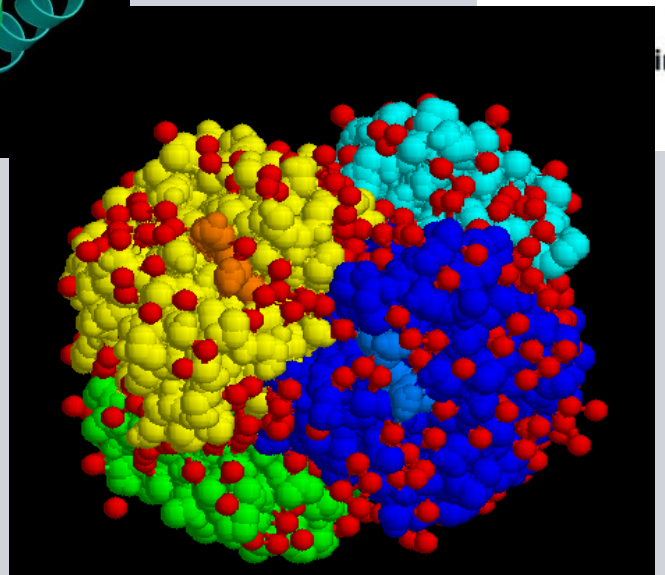
> 200 milionů proteinových struktur

Vizualizace biomolekuly v počítači

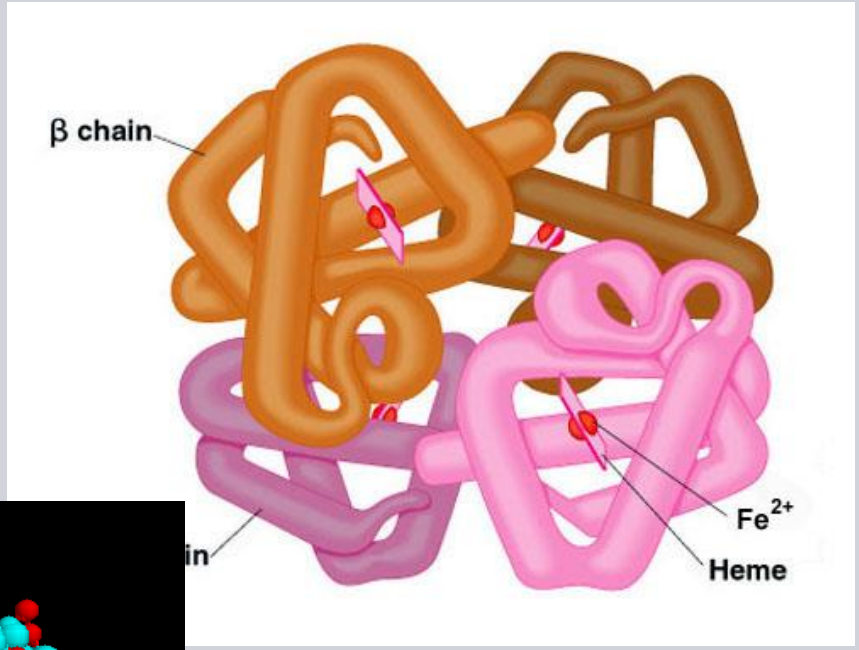


Cartoon model:

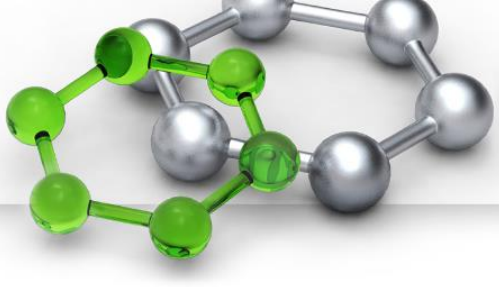




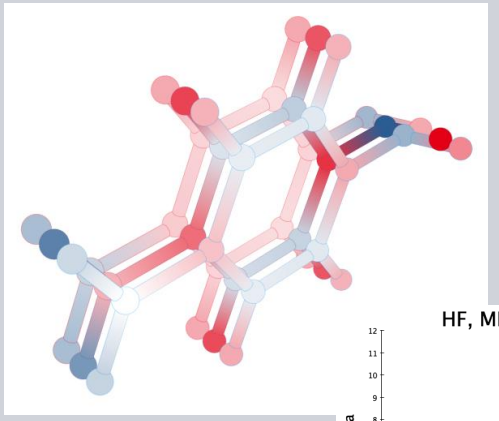
Schématický model:

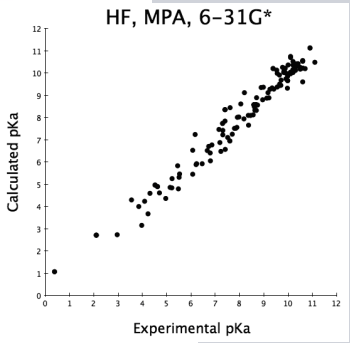


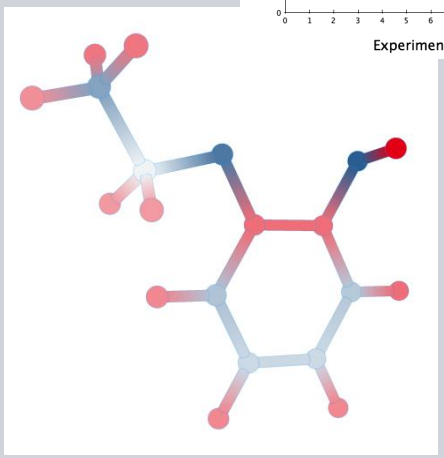
Chemoinformatika – hlavní oblasti



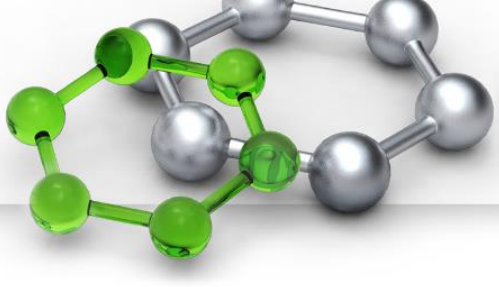
- Podobnostní vyhledávání v databázích
- Výpočty a aplikace deskriptorů
- QSAR / QSPR
- Vytváření a aplikace virtuálních knihoven molekul
- Virtuální screening







Proces návrhu a vývoje léku



Uvedení nového léku na trh stojí v průměru 900 milionů dolarů a trvá více než 10 let.

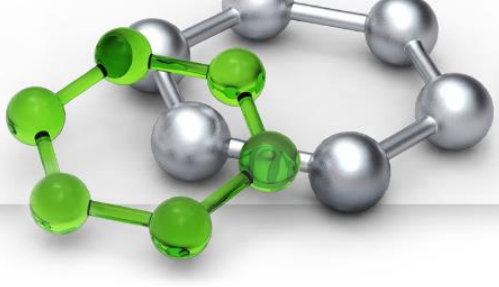
Farmaceutické společnosti často zkoumají a testují 10 000 – 30 000 rozličných látek předtím, než je jedna z nich úspěšně uvedena na trh.

Látky jsou nejdříve navrženy v základním výzkumu.

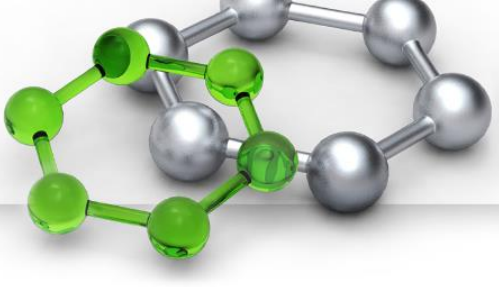
Poté musí projít předklinickými a klinickými zkouškami.

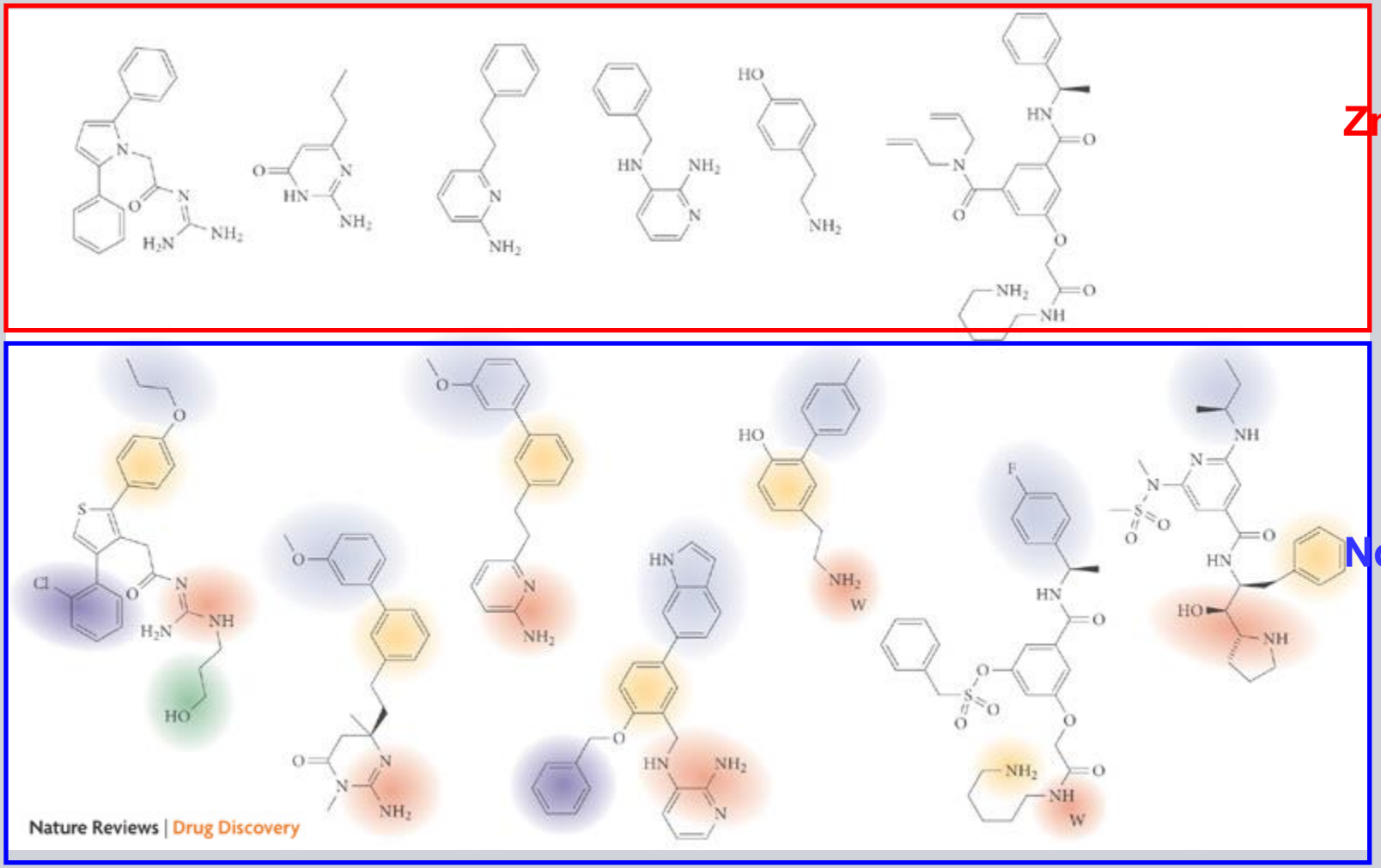
Většina nepostoupí dále, ale ty, které postoupí, mohou nabídnout šanci na kvalitnější život pacientů.

Proces návrhu a vývoje léků



Návrh léku (drug design)

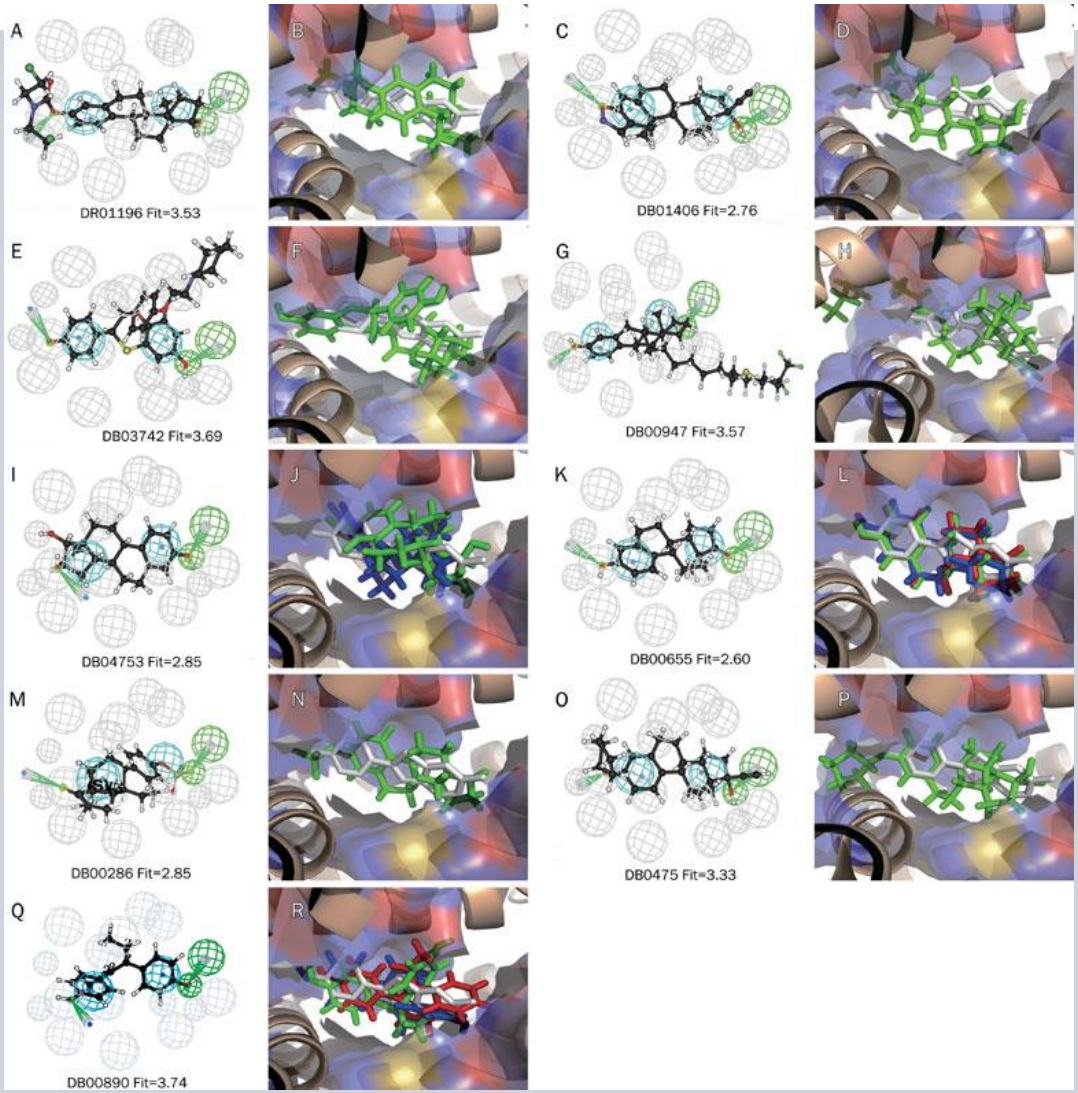




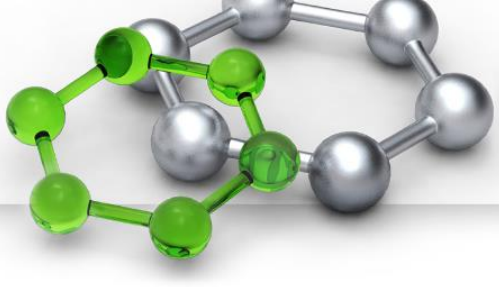
Známé léky

Nově navržené léky

Návrh léku (drug design)



Návrh léku (drug design)

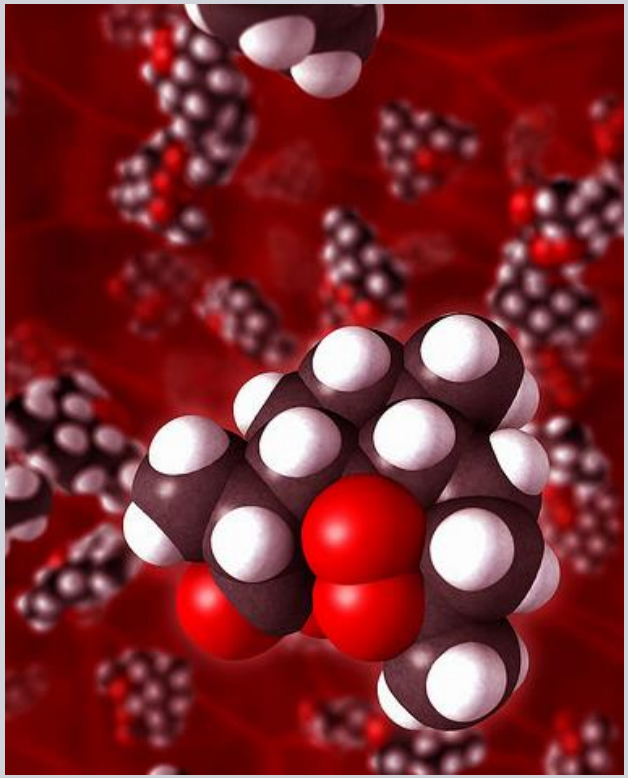


Na základě několika známých molekul léků můžeme vytvořit (ručně nebo automaticky) rozsáhlé sady molekul.

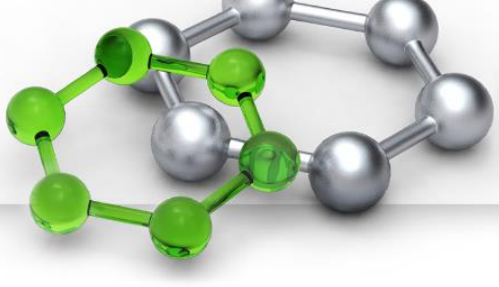
Tyto sady se nazývají virtuální knihovny a obsahují desetiticíce, staticíce i miliony molekul.

Několik z molekul ve virtuální knihovně může být velmi účinnými léky.

Ale které to jsou ???



Jak zjistit, která z navržených molekul je lékem?



Navržené molekuly existují jen na papíře nebo v počítači a nebyly zatím syntetizovány.

Nemáme tedy naměřeny jejich fyzikální a chemické vlastnosti ani nevíme nic o jejich aktivitě.

Jak tedy určit, která z nich bude vhodným lékem?

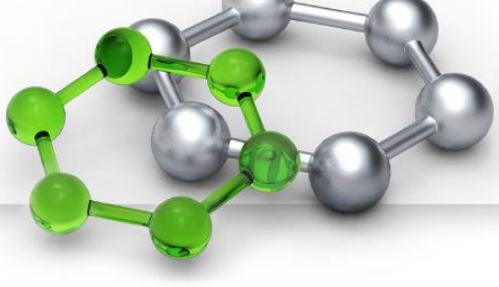
Máme dvě možnosti:

a) Molekuly synetizovat a jejich vlastnosti i aktivitu změřit.

b) Vlastnosti i aktivitu molekul odhadnout (predikovat) na základě jejich struktury.

Chemoinformatika

Orgranizace výuky



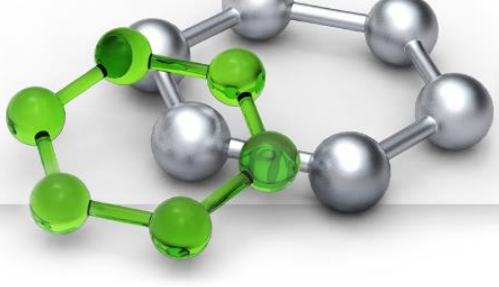
Přednáška + praktická cvičení
Prosím, noste si notebooky

Občas nepovinné úkoly – možnost získat procenta navíc

Ukončení předmětu:
Písemka
- Teoretické otázky
- Příklady a cvičení

Povoleny všechny materiály

Orgranizace výuky 2



Klasifikace:

<100%, 90%>: A

(90%, 80%>: B

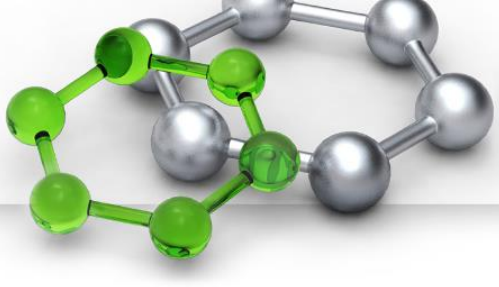
(80%, 70%>: C

(70%, 60%>: D

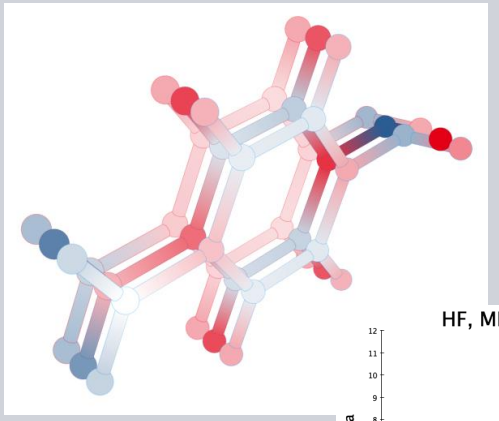
(60%, 50%>: E

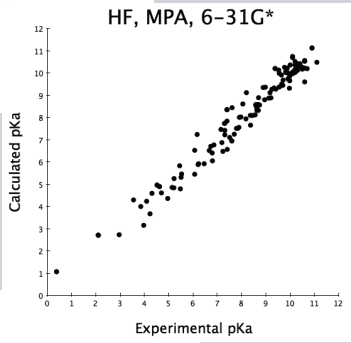
(50%, 0%>: F

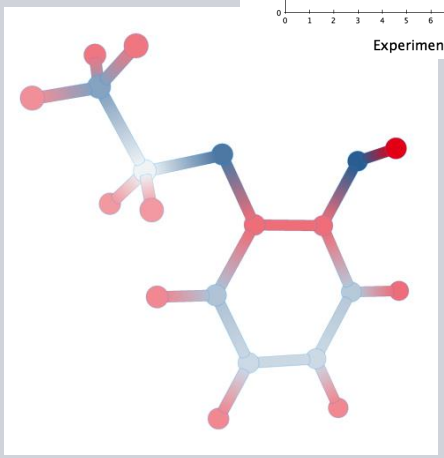
Osnova výuky

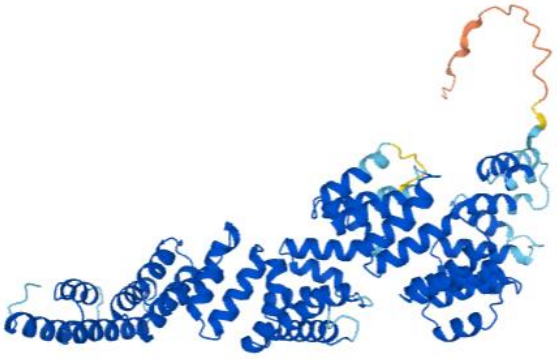


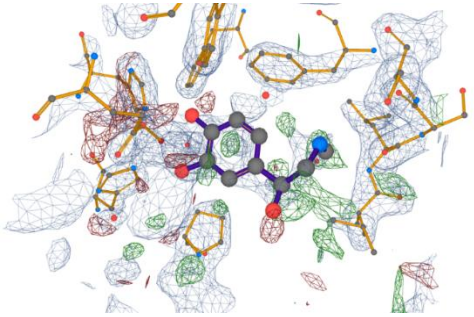
- Struktura molekuly (2D, 3D)
- Databáze molekul
- Deskriptory
- Podobnost molekul
- Podobnostní vyhledávání v databázích
- QSAR / QSPR
- 3D QSAR











Děkuji za pozornost