

Databáze v chemické a forenzní analýze

Tereza Uhlíková, tereza.uhlikova@vscht.cz

Tereza Uhlíková

Ústav analytické chemie

skupina teoretické spektroskopie

místnost A277

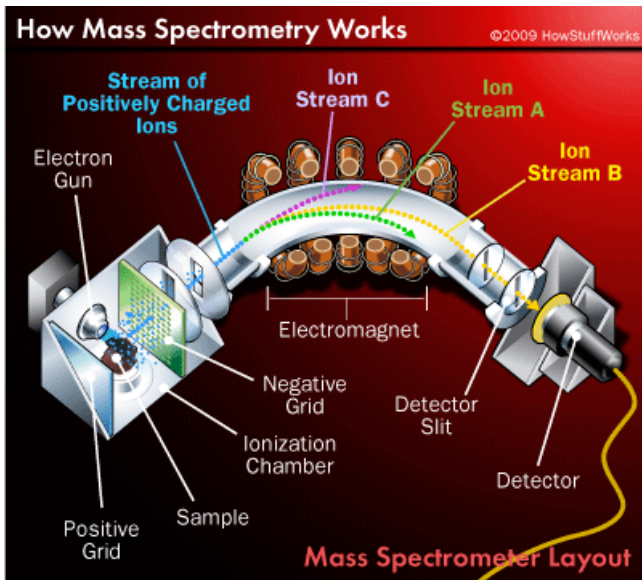
<https://web.vscht.cz/~uhlikovt/>

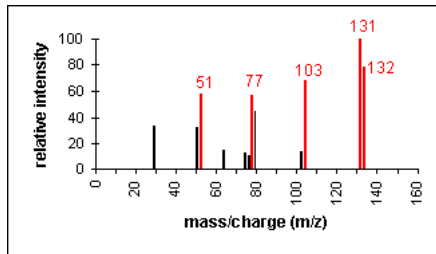
tereza.uhlikova@vscht.cz

O čem to bude

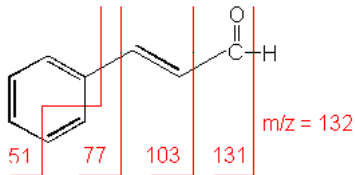
1	19.3.	Opakování databází, normalizace, ERA
2	26.3.	Spektroskopické databáze - atmosférické, vesmírné
3	2.4.	NIST, specDB
4	9.4.	MS, krystalografické
5	16.4.	AI a Kvantové počítače pro práci s daty
6	23.4.	Barvy,laky,sklo,textilie
7	30.4.	Půda, hořlaviny,nukleární materiály
8	7.5.	STŘEDEČNÍ ROZVRH = NENÍ PŘEDNÁŠKA
9	14.5.	Toxikologické
10	21.5.	Zkouška

cvičení: společná databáze volně prodejných psychotropních látek





Hmotnost jednotlivých fragmentů



3-Phenyl-2-propenal

C_9H_8O

MW = 132.16

používá se ve spojení s plynovou chromatografií (GC) - nejprve vzorek rozseparujeme na GC a pak každá složka z GC jde do MS

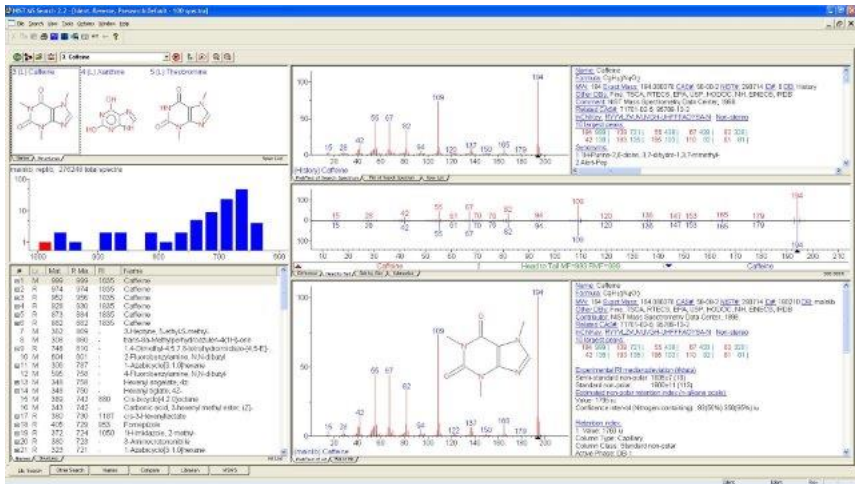
Při zachování stejných podmínek je MS velice dobře reprodukovatelná
hmotnost elementů je neměnná

s knihovnou je potřeba porovnat celé MS spektrum, nejen konečnou molekulovou hmotnost

v datech jsou zaznamenány pouze $[x;y]$ souřadnice - program spočítá pravděpodobnost dvou spekter

člověk pak musí rozhodnout o správnosti

Databáze hmotnostních spekter - NIST



Možnosti vyhledávání

- Nominální hmotnosti
- Přesné hmotnosti
- Vzorce/ nebo části
- Nějakáho píku
- Okrajové podmínky
- CAS číslo
- ID číslo
- NIST číslo
- Jméno/ i triviální

NIST knihovna <https://chemdata.nist.gov/>

<http://www.nist.gov/srd/nist1a.cfm>

Možno využít také tyto stránky - popis

<http://www.sisweb.com/software/ms/nist.htm>

demo verze <https://www.sisweb.com/software/ms/nist.htm#demo>

Online hledání sloučeniny, je-li v databázi

<http://www.sisweb.com/software/ms/nistsearch.htm>

Databáze MS peptidů

<http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:mspepsearch>

Wiley Registry

Wiley Registry of mass spectral data – už je placená
775500 hmotnostních spekter, 741000 chemických struktur za 6335 liber
<https://sciencesolutions.wiley.com/solutions/technique/gc-ms/wiley-registry-of-mass-spectral-data/>

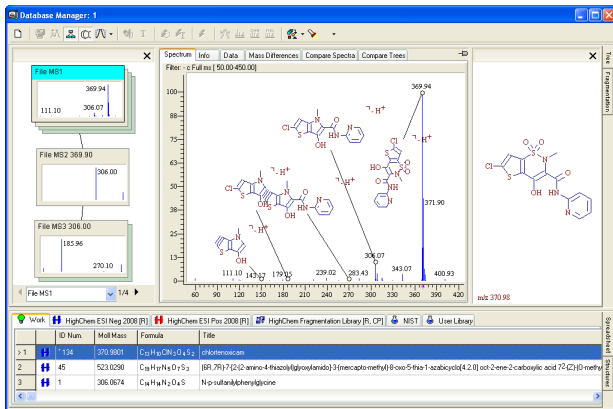
The screenshot displays the Wiley Registry software interface. The top menu bar includes File, Edit, Search, Predict, Show, Float, Clear, Settings, and Help. The interface is divided into several sections:

- Query area (red box):** Shows an empty mass spectrum plot with the x-axis labeled 'ppm' and values 140, 130, 120. The y-axis is labeled 'Arbitrary Unit' with values 0, 25, 50, 100. A search bar contains the text 'Query area'. To the right is a chemical structure of 3-Chloro-aniline (Nc1cccc(Cl)c1).
- Result area (blue box):** Shows a mass spectrum plot with peaks at approximately 112, 119, 130, 140, 150, and 160 ppm. The x-axis is labeled 'ppm' and the y-axis is 'Arbitrary Unit'. A search bar contains the text 'Result area'. To the right is the same chemical structure of 3-Chloro-aniline.
- Hitlist (green box):** A table listing search results.

Rank	Quality	Folder ID	Property
1/1	1000	1:FO0000000039	3-Chloro-aniline

HighChem Mass Frontier™; mzCloud <https://www.mzcloud.org/>

založeno jednou skupinou vědců v laboratoři v Bratislavě



Japonská massBank

<http://www.massbank.jp/?lang=en>

manuál

http://www.massbank.jp/manuals/MassBankUserManual_en.pdf

MassBank of North America

<http://mona.fiehnlab.ucdavis.edu>

<http://mona.fiehnlab.ucdavis.edu/spectra/search>

X-RAY DIFFRACTION ANALYSIS IN THE FORENSIC SCIENCE: THE LAST RESORT IN MANY CRIMINAL CASES

International Centre for Diffraction Data 2003, Advances in X-ray Analysis, Volume 46

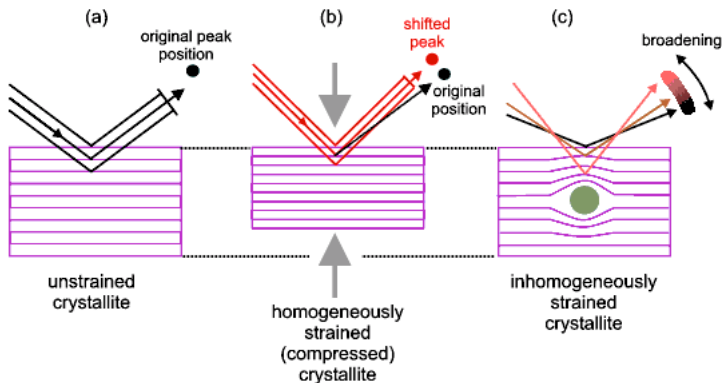
dle rozdílných krystalografických struktur (polymorfů) lze určit původ, padělky, změny vlivem působení různých chemicko-fyzikálních podmínek, biologickou aktivitu

existuje 230 krystalografických grup

<https://crystalsymmetry.wordpress.com/2014/08/15/the-space-group-list-project-as-a-poster/>

Anorganická – minerály (1913) a ideální krystal

- Struktura krystalu - Konkrétní rozmístění stavebních částic (atomů, iontů) krystalických látek v prostoru
- Základní vzor
- Krystalová mříž (mřížka)
- Základní buňka (např. plošně centrovaná buňka)
- Krystalové (krystalografické) soustavy
- Symetrie vnějšího tvaru a struktury krystalů
- Bodové grupy symetrie krystalů
- Prostorové grupy symetrie krystalů
- Významné krystalografické směry



Je těžké vytvořit monokrystal – **Prášková difrakce**

Spektrum poskytuje – úhlovou polohu difrakčního maxima, intenzitu, tvar a šířku píků

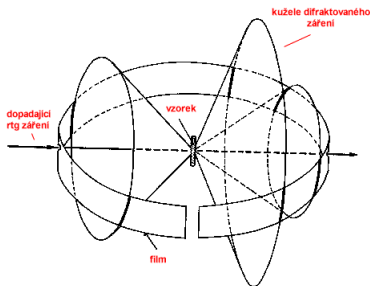
Aplikace a její výsledky – stupeň krystalinity, fázová identifikace, indexování základních buněk vyřešení krystalové struktury studium jedné látky v různých fázích (polymer)

Princip metody <http://www.xray.cz/kfes/vyuka/lp/pd.htm>

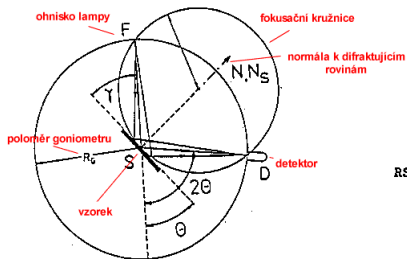
Prášková difrakce

Podmínkou pro difrakci rentgenového záření o vlnové délce λ na osnově rovin s difrakčními indexy hkl a mezirovinnou vzdáleností d_{hkl} je, aby záření dopadalo na tyto roviny pod úhlem θ_{hkl} , který splňuje Braggovu rovnici

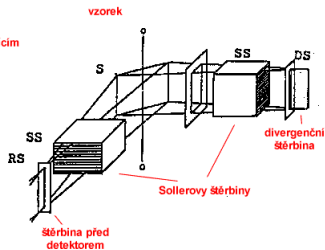
$$2d_{hkl} \sin \theta_{hkl} = \lambda$$



Prášková difrakce



Schema Braggovy-Brentanovy geometrie



Systém štěrbin v BB geometrii

PDF - Powder Diffraction Files (není portable document od Adobe)
standardně se dodává jako součást software pro rentgenové difraktometry
obsahuje úhly reflexi (d_{hkl}) a jejich intenzity a slouží pro identifikaci
jednotlivých krystalických fází (fázovou analýzu)

David Sedmidubský

The International Centre for Diffraction Data \$5775.00 na rok

<http://www.icdd.com/>

http://www2.fiz-karlsruhe.de/icsd_home.html

Obsahuje kompletní strukturní data (grupa symetrie, mřížové parametry, polohy jednotlivých atomů a jejich obsazovací faktory, teplotní faktory)
199000 prověřených dat včetně jejich atomových souřadnic od roku 1913
2019 krystalových struktur prvků
33809 data pro dvojčlenné směsi
65126 trojčetné
60669 čtyřčetné a pětičetné
okolo 138500 položek (79.8%) má přiřazený strukturní typ a dalších 8230
má aspoň prototyp
.. Cena a demoverze

<http://www.nist.gov/srd/nist84.cfm>

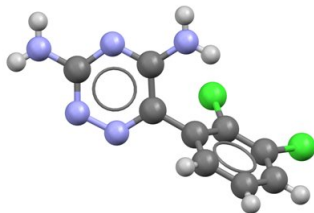
The Cambridge Crystallographic Data Centre

<http://www.ccdc.cam.ac.uk/pages/Home.aspx>

Cambridge Structural Database

<http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/CSD.aspx>

Založena v 1965, CSD je světová úschovna krystalických struktur malých organických a organo-kovových. Nyní obsahuje přes 1 110 837 struktur.



Dělí se na několik částí:

- CSD – samotná databáze
- ConQuest – vyhledávání v CSD
- IsoStar – knihovna mezimolekulových interakcí
- WebCSD – online přístup k nově publikovaným strukturám
- Mercury – zobrazování a analýza CSD dat
- PreQuest – prostředí pro vytvoření vaší vlastní databáze
- Mogul – ověřování molekulové geometrie

http:

[//www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/WebCSD.aspx](http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/WebCSD.aspx)

ConQuest

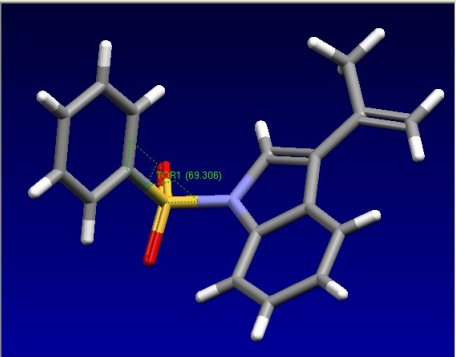
- vyhledávání podle jména, vzorce, prvkového zastoupení, reference v literatuře, experimentálních detailů
- chemické podmínky jako náboj, hybridizace a počet cyklů
- mezimolekulové a nevazebné interakce

CCDC ConQuest (1) : search2 [Search]

File Edit Options View Databases Results Help

Build Queries Combine Queries Manage Hitlists View Results

Refcode: ABUSAR CSD version 5.33 (November 2011)



ABUSAR

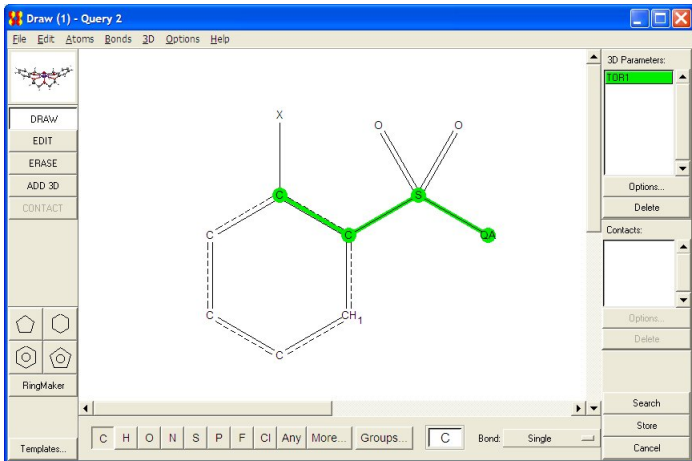
Analyse Hitlist

- ✓ ABADOX
- ✓ ABADUD
- ✓ ABEBIS
- ✓ ABEBOY
- ✓ ABEMOJ
- ✓ ABEMUP
- ✓ ABUKEN
- ✓ ABUPOC
- ✓ ABURUK
- ✓ **ABUSAR**
- ✓ ABUSOF
- ✓ ABZNP S
- ✓ ABZSLM
- ✓ ABZTCX
- ✓ ACANEY
- ✓ ACANIC
- ✓ ACANOI
- ✓ ACAPEAW
- ✓ ACAPEA
- ✓ ACENEB
- ✓ ACIFAT
- ✓ ACIROT

1910 hits 34%

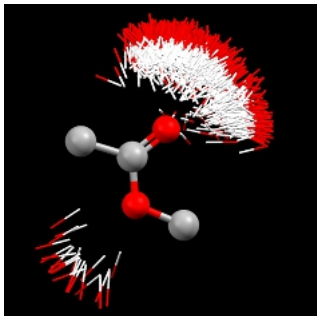
Stop Search

Show substructure matches
 Q2 Param: All No Objects Detach

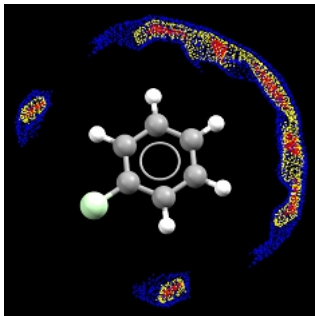


Isostar – Knihovna mezimolekulových interakcí

O-H skupiny okolo alifatické esterové skupiny

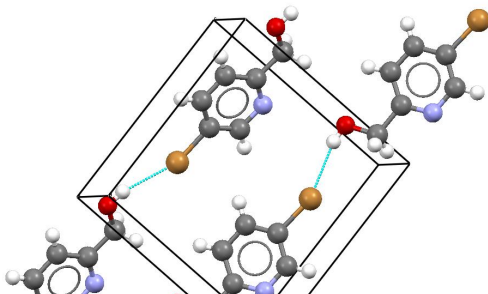


Elektronová hustota okolo fenylové skupiny



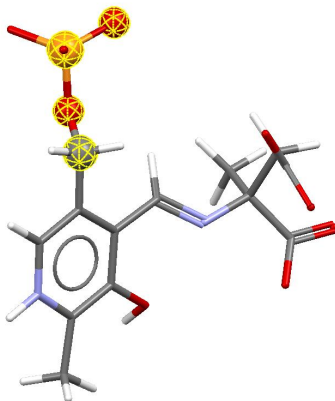
Mercury – zobrazování a analýza CSD dat

- Zobrazuje elementy prostorových grup
- Vypočítá a zobrazí volná místa v krystalech, založená na kontaktech povrchů nebo s rozpouštědlem
- Ukazuje i vypočtenou geometrii jedné neinteragující molekuly pomocí MOPAC
- Počítá mezimolekulový potenciál a zobrazuje nejsilnější interakci v krystalu



Mogul – ověřování molekulové geometrie

Poskytuje přesné informace o dané molekulové geometrii prohledáváním chemicky klasifikovaných vazebných délek, valenčních úhlů, torzních úhlů a prstencových uspořádání odvozených z CSD



WebCSD – vyhledávání přes web

File Filter Help

Find Entry

Entry

- BUFPQAE01
- BUVGII
- CAACTY
- CACWOS
- CADVEI
- CAFINE**
- CAFROR
- CAMHFA
- CAMXAP01
- CAQTET
- CARQOB
- CARTEN
- CARTEN02
- CATCOL13
- CBMZPN01
- CBMZPN02
- CBMZPN03
- CBMZPN10
- CBMZPN11
- CBMZPN12
- CCXAPT
- CEBGUL
- CECZEP
- CECZIT
- CEFXOA
- CEHTAK10

< >

500 Hits

100%

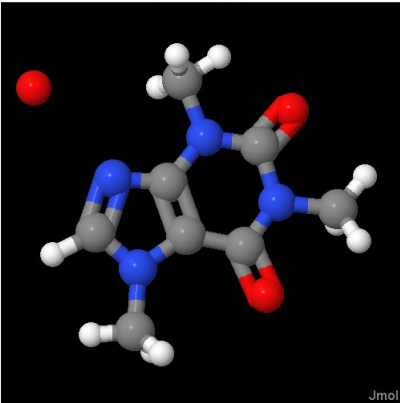
Stop Search

Entry loaded

CAFINE : 1,3,7-Trimethyl-purine-2,6-dione monohydrate

D.J.Sutor, *Acta Crystallogr.* (1958), **11**, 453, doi:10.1107/S0365110X58001286

Hide Viewer

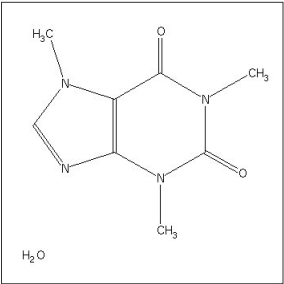


Jmol

Ball and Stick No Labels

Hydrogens Bond types Disorder

Diagram Details Viewer Export Options Help



H₂O

C₈ H₁₀ N₄ O₂ · H₂O

Space Group: P 2₁/a

a 14.8(1) **b** 16.7(1) **c** 3.97(3)

α 90 **β** 97.0(5) **γ** 90

R-Factor: 14.6%

Temperature (K): Room Temp.(283-303)

Krystalografické - proteinové

1958 – první krystal myoglobin

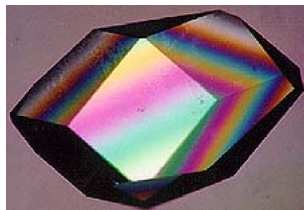
1962 - Nobelova cena v chemii Kendrew a Perutz



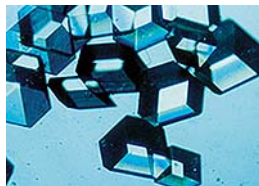
Krystalizace proteinů je obtížná – velká citlivost na teplotě, pH, iontové síle. . .

Proteinové krystaly jsou molekulární krystaly s velkým obsahem rozpouštědla

Lysozym

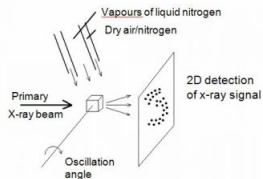
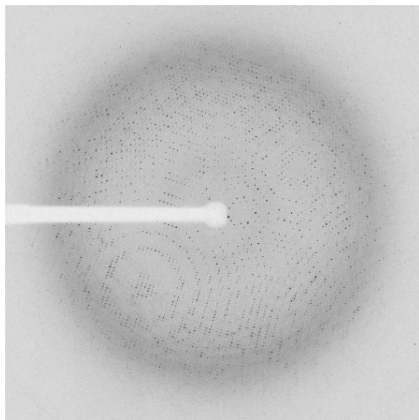


Inzulin

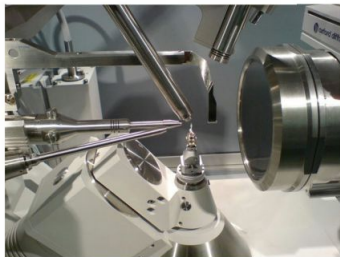


Krystalografické - proteinové

<https://old.fzu.cz/oddeleni/oddeleni-strukturni-analyzy/research-subjects/krystalografie-proteinu>
difrakce



Experimental set-up



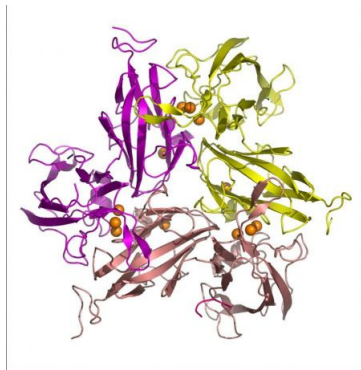
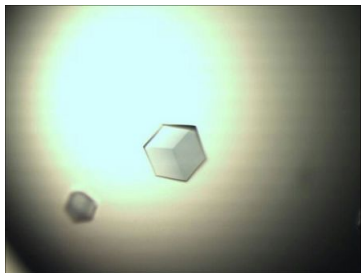
Krystalografické - proteinové

<https://old.fzu.cz/oddeleni/oddeleni-strukturni-analyzy/research-subjects/krystalografie-proteinu>

Fázový problém u proteinů

Upřesňování proteinových struktur

Monokrystaly malé lakázy pěstované v malé kapce. Velikost krystalu uprostřed je asi 100 mikronů.



Laktáza z bakterie *Streptomyces coelicolor*, oranžové kuličky ionty mědi v aktivních místech enzymu, 12000 atomů

Protein Data Bank

<http://www.rcsb.org/pdb/home/home.do>

Obsahuje 151579 biologických makromolekulových struktur

Podle čeho se vyhledává

Příklad Hemoglobin

Ovo → ovotransferin

<http://www.rcsb.org/pdb/explore/explore.do?structureId=2D3I>

Najít příklady třeba viru Herpes

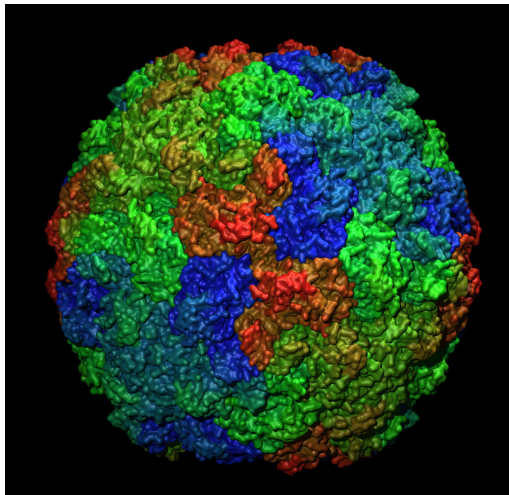
The Biologically Interesting Molecule Reference Dictionary (BIRD)

<http://www.wwpdb.org/data/bird>

Krystalografické - proteinové

symetrie obalu lidského rhinoviru typu 14 složená z 240 jednotek (PDB 1k5m)

Vítěz soutěže: Věda je krásná



Doplňky ke cvičení

Zadávání RGB

datový typ při zadávání barev jako RGB

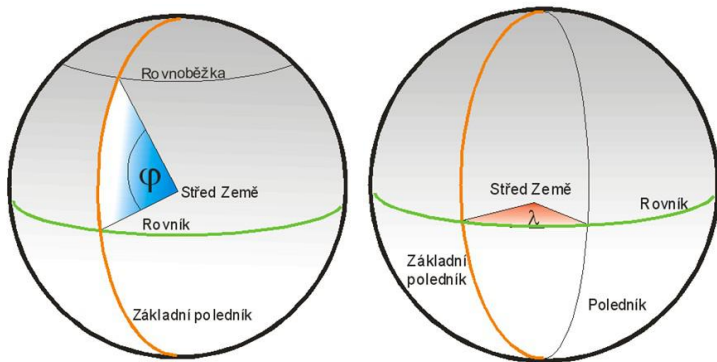
name	hexCode	red10	green10	blue10
acid green	#B0BF1A	176	191	26
Amaranth purple	#AB274F	171	39	79
Canary yellow	#FFEF00	255	239	0

name	hexCode	refFraction	greenFrac	blueFrac
acid green	#B0BF1A	0.69	0.75	0.1
Amaranth purple	#AB274F	0.67	0.15	0.31
Canary yellow	#FFEF00	1	0.94	0

<https://www.back4app.com/database/back4app/rgb-color-codes-and-names/color-dataset-via-api>

<https://encycolorpedia.com/ffef00>

zeměpisná šířka & zeměpisná délka



Jako GPS šířka: Decimal(8,6), délka: Decimal(9,6)

formát bude vypadá takto `##.#####` and `###.#####`
přesnost polohy ~ 10 cm

