

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/>

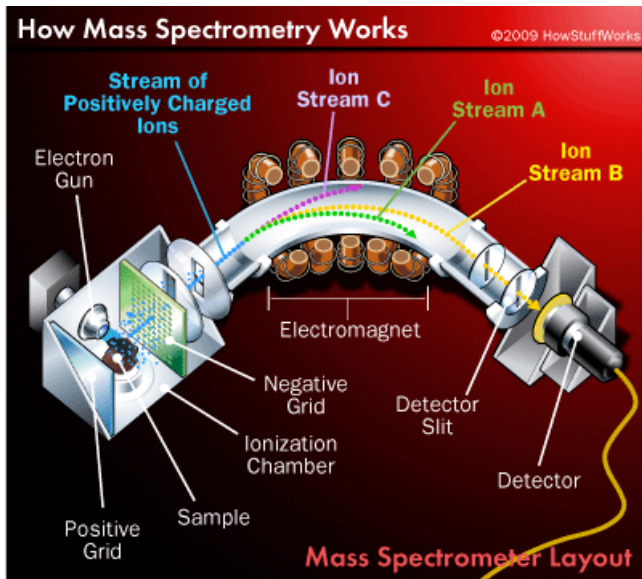
tereza.uhlikova@vscht.cz

O čem to bude

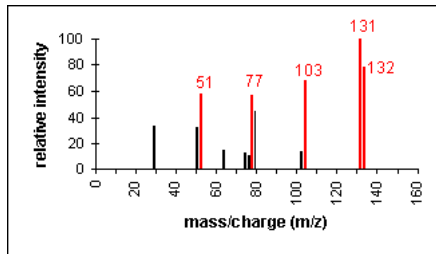
		Téma přednášky	Téma cvičení
1	11.3.	Opakování a seznamování	Hledání odpovědí pomocí AI
2	18.3.	Princip AI	AI databázové nástroje
3	25.3.	Chemické databáze	Analýza spekter pomocí databází
4	1.4.	Forenzní databáze	zkoušková témata
5	8.4.	MS; CCSD	tvorba MS databáze
6	15.4.	společná databáze volně prodejných psych. látek VY a AI	
7	22.4.		
8	29.4.		
9	6.5.	ČTVRTEČNÍ ROZVRH = NENÍ PŘEDNÁŠKA	
10	...	Zkouška	

Hmotnostní spektra

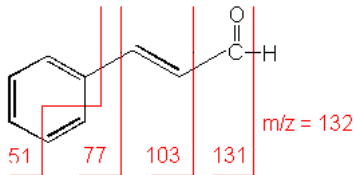
Princip hmotnostních spekter



Hmotnostních spektrum



Hmotnost jednotlivých fragmentů



3-Phenyl-2-propenal

C_9H_8O

MW = 132.16

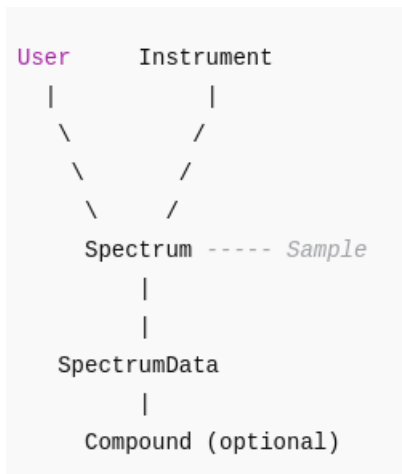
- používá se ve spojení s plynovou chromatografií (GC) - nejprve se vzorek rozseparuje na GC, pak každá složka z GC jde do MS
- při zachování stejných ionizačních podmínek je MS velice dobře reprodukovatelná - hmotnost elementů je neměnná
- s knihovnou je potřeba porovnat celé MS spektrum, nejen konečnou molekulovou hmotnost
- v datech jsou zaznamenány pouze $[x;y]$ souřadnice - program spočítá pravděpodobnost dvou spekter
- člověk pak musí rozhodnout o správnosti

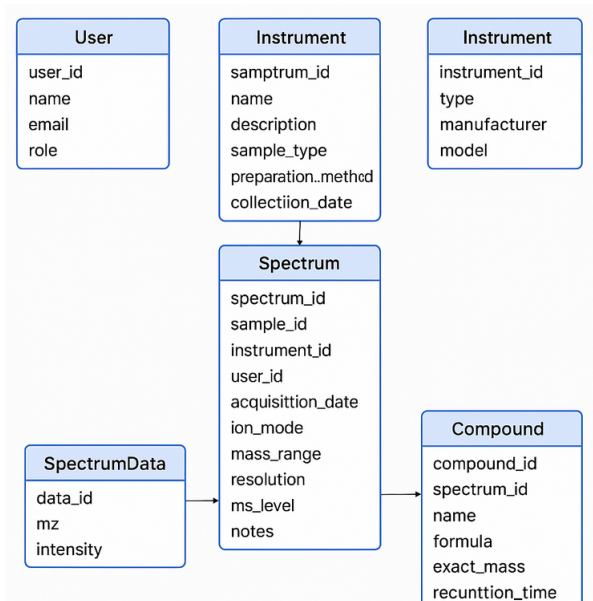
Vytvořme ERA pro databázi MS spekter v malé laboratoři

- Sample** - Represents the physical/chemical sample analyzed.
- Spectrum** - Represents the mass spectrum data generated from the sample.
- Compound** (Optional) - If known, represents the compound(s) identified in the spectrum.
- Instrument** - The device used for measurement.
- User** - The researcher or technician who created or analyzed the spectrum.

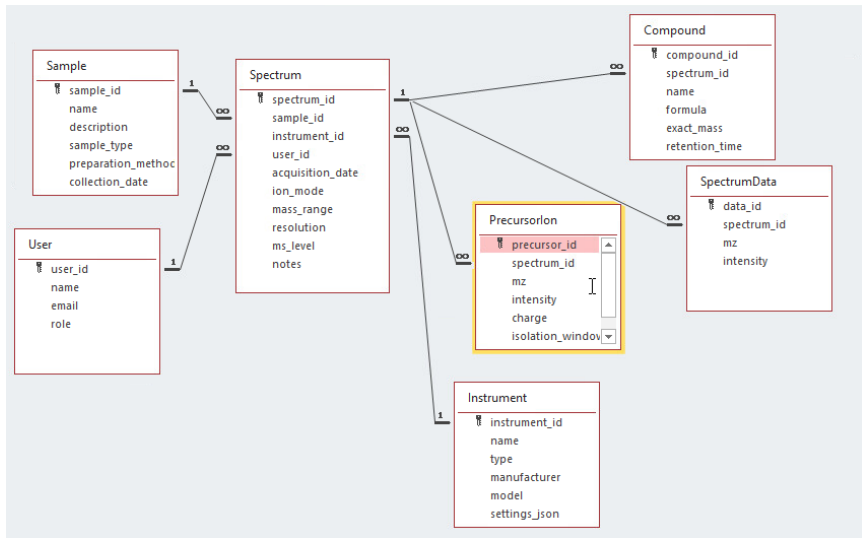
Relationships

- Sample $\leftrightarrow$ Spectrum: One-to-Many (A sample can have multiple spectra)
- Spectrum $\leftrightarrow$ Compound: One-to-Many (A spectrum can be annotated with multiple compounds)
- Instrument $\leftrightarrow$ Spectrum: One-to-Many (One instrument can be used for many spectra)
- User $\leftrightarrow$ Spectrum: One-to-Many (Each spectrum is associated with one user)





ERA - Access



Do Accessu nelze dávat několikanásobné dotazy.

User

user_id	name	email	role
1	Alice Nguyen	alice@lab.org	Technician
2	Bob Silva	bob@lab.org	Researcher
3	Dr. K. Jordan	jordan@lab.org	PI

Instrument

instrument_id	name	type	manufacturer	model	settings_json
1	QTOF 5600+	Q-TOF	AB Sciex	TripleTOF	{"voltage": 5500}
2	Orbitrap Elite	Orbitrap	Thermo	Elite	{"resolution": 60000}
3	TIMS-TOF Pro	TIMS-TOF	Bruker	Pro	{"accumulation": "auto"}

Sample

sample_id	name	description	sample_type	preparation_method	collection_date
1	Plasma 01	Fasting human plasma	Plasma	Protein crash	2025-03-01
2	Bacterial Lysate	E. coli extract	Lysate	Sonication	2025-03-02
3	Metabolite Mix	Standard mixture	Standard	Dilution	2025-03-05

Spectrum

spectrum_id	sample_id	instrument_id	user_id	acquisition_date	ion_mode	mas:
1	1	1	1	2025-03-01	Positive	50-1
2	1	1	1	2025-03-01	Positive	50-1
3	2	2	2	2025-03-03	Negative	100-

SpectrumData

data_id	spectrum_id	mz	intensity
1	1	445.34	15321
2	1	512.76	9876
3	2	123.45	5432

Precursorlon

precursor_id	spectrum_id	mz	intensity	charge	isolation_window	collision_u
1	2	445.34	15321	1	1.0	35
2	NULL	NULL	NULL	NULL	NULL	NULL
3	NULL	NULL	NULL	NULL	NULL	NULL



Compound (optional)

compound_id	spectrum_id	name	formula	exact_mass	retention_time
1	1	Caffeine	C ₈ H ₁₀ N ₄ O ₂	194.0804	2.13
2	2	Adenine	C ₅ H ₅ N ₅	135.0545	1.98
3	3	Glucose	C ₆ H ₁₂ O ₆	180.0634	0.88

Vkládání dat do tabulek

```
-- Users INSERT INTO User (name, email, role) VALUES  
( 'Alice Nguyen', 'alice@lab.org', 'Technician'),  
( 'Bob Silva', 'bob@lab.org', 'Researcher'),  
( 'Dr. K. Jordan', 'jordan@lab.org', 'PI');
```

Do Accessu není možné :-(
Musí se vkládat řádek po řádku:

```
INSERT INTO Person VALUES (1, "Amir");  
INSERT INTO Person VALUES (2, "Sofia");  
INSERT INTO Person VALUES (3, "Aya");
```

Vyhledávání podle spektra

Vyhledávání podle spektra

Krok 1: Celé spektrum v **SpectrumData**

Krok 2: Přidat **Index** na **mz**

```
CREATE INDEX idx_mz_search ON SpectrumData (mz);
```

rychlejší vyhledávání

```
SELECT spectrum_id FROM SpectrumData WHERE mz BETWEEN  
445.33 AND 445.35;
```

lepší je použít složený index

```
CREATE INDEX idx_mz_spectrum ON SpectrumData (mz,  
spectrum_id);
```

uchovávání celého naměřeného spektra (BLOB, TEXT)

```
ALTER TABLE Spectrum ADD COLUMN spectrum_blob BLOB;
```

-- nebo

```
ALTER TABLE Spectrum ADD COLUMN spectrum_json TEXT;
```

Databáze hmotnostních spekter NIST

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

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

Možno využít také tyto stránky - popis

<http://www.sisweb.com/software/nist-msms.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>

Možnosti vyhledávání

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

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/>

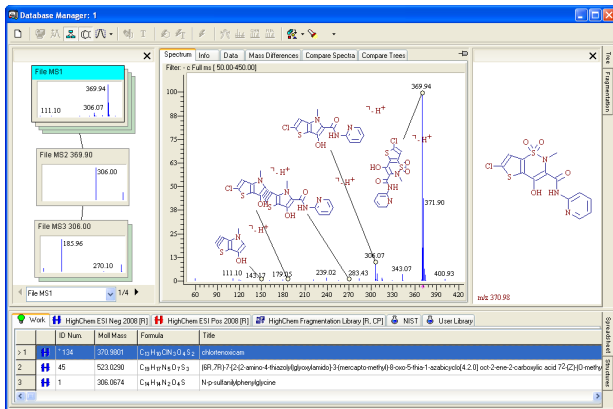
The screenshot displays the Wiley Registry software interface, which is used for searching mass spectral data. The interface is divided into several sections:

- Query area:** The top section, outlined in red, shows a search query for 3-Chloro-aniline. It includes a chemical structure of 3-Chloro-aniline and a table with columns 'Type' and 'St'.
- Result area:** The bottom section, outlined in blue, shows the search results. It includes a chemical structure of 3-Chloro-aniline and a table with columns 'Type' and 'St'.
- Icon bar:** A horizontal bar with various icons for navigating and manipulating the data, located below the search and result areas.
- Hitlist:** A table at the bottom of the interface, outlined in green, listing the search results. It has columns for Rank, Quality, Folder ID, and Property.

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>

Krystalografie

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

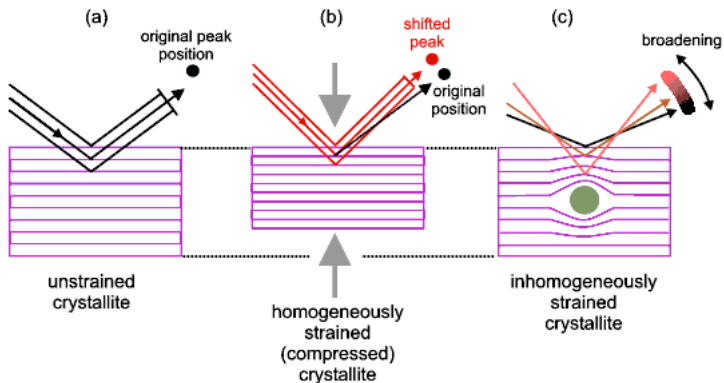
dle rozdílných krystalografických struktur (polymorfů) lze určit původ, padělky, změny vlivem působení různých chemicko-fyzikálních podmínek, biologickou aktivitu

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

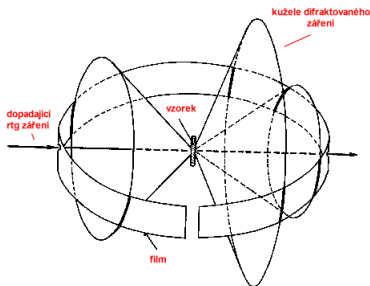
Aplikace a její výsledky – stupeň krystalinity, fázová identifikace, indexování základních buněk vyřešení krystalové struktury studium jedné látky v různých fázích (polymer)

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

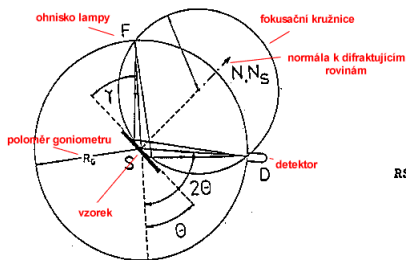
Prášková difrakce

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

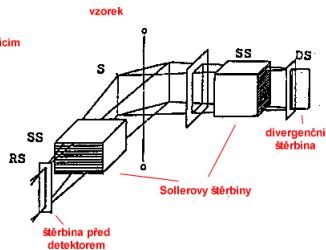
$$2d_{hkl} \sin \theta_{hkl} = \lambda$$



Prášková difrakce



Schema Braggovy-Brentanovy geometrie



Systém štěrbin v BB geometrii

PDF - Powder Diffraction Files (není portable document od Adobe)
standardně se dodávána 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>

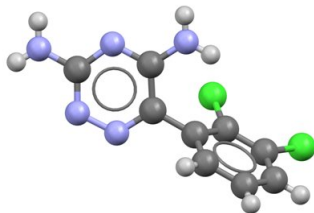
The Cambridge Crystallographic Data Centre

<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
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)

3D Visualiser

ABUSAR

Analyse Hitlist

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

1910 hits

34%

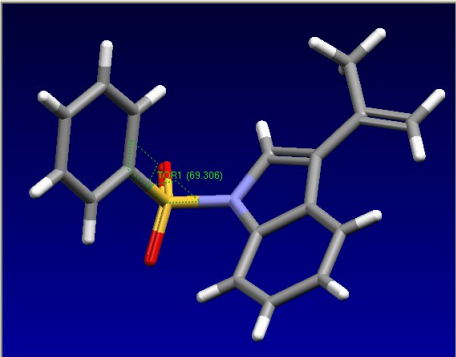
Stop Search

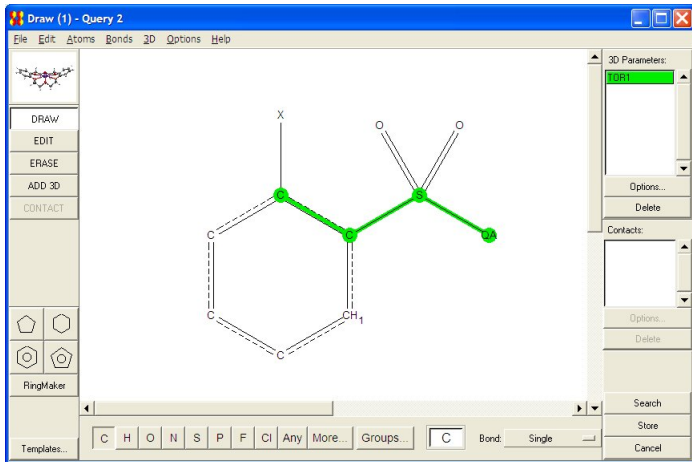
Detach

✓ Show substructure matches

Q2 Param: All

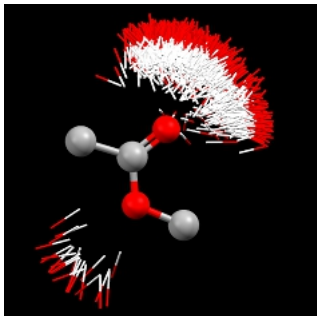
No Objects



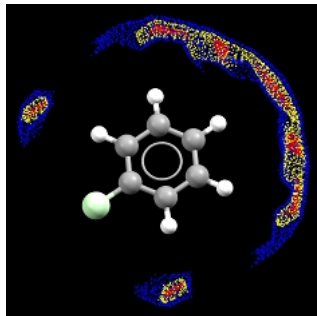


Isostar – Knihovna mezimolekulových interakcí

O-H skupiny okolo alifatické esterové skupiny

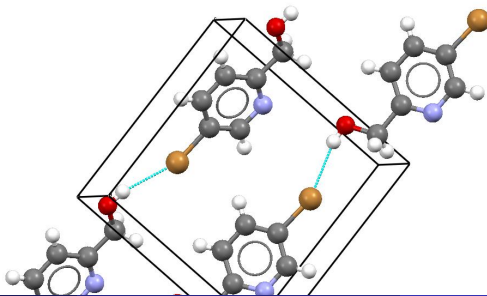


Elektronová hustota okolo fenylové skupiny



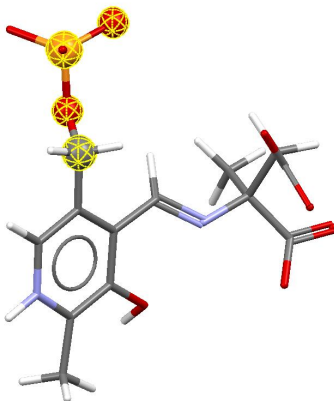
Mercury – zobrazování a analýza CSD dat

- 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



Cambridge Structural Database

WebCSD – vyhledávání přes web

<https://www.ccdc.cam.ac.uk/structures/Home/>

File Filter Help

Find Entry

Entry

OTHER003

BUPQAE01

BUVGII

CAACTY

CACWOS

CADVEI

CAFINE

CAFROR

CAMHFA

CAMKAP01

CAQTET

CARGOB

CARTEN

CARTEN02

CATCOL13

CBMZPN01

CBMZPN02

CBMZPN03

CBMZPN10

CBMZPN11

CBMZPN12

CCXAPT

CEBGUL

CECZEP

CECZIT

CEFXOA

CEHTAK10

<

>

500 Hits

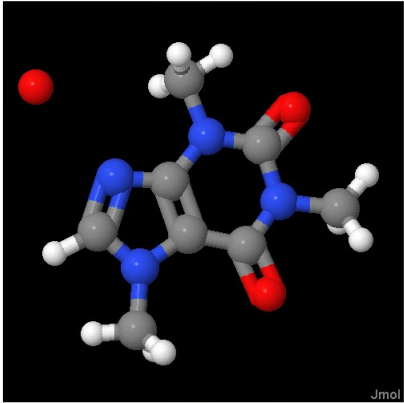
100%

Stop Search

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

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

Hide Viewer



Diagram

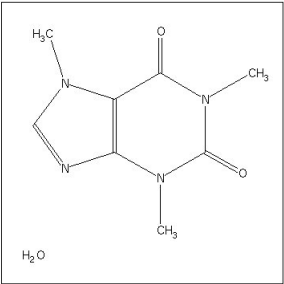
Details

Viewer

Export

Options

Help



H_2O

$\text{C}_8\text{H}_{10}\text{N}_4\text{O}_2\cdot\text{H}_2\text{O}$

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. (293-299)

Krystalografické - proteinové

1958 – první krystal myoglobin

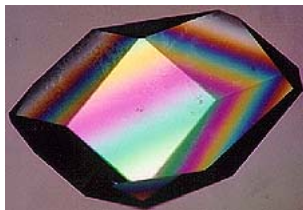
1962 - Nobelova cena v chemii Kendrew a Perutz



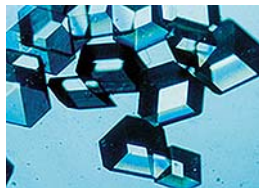
Krystalizace proteinů je obtížná – velká citlivost na teplotě, pH, iontové síle. . .

Proteinové krystaly jsou molekulární krystaly s velkým obsahem rozpouštědla

Lysozym

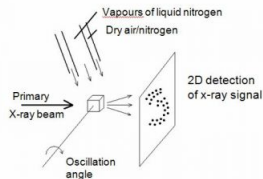
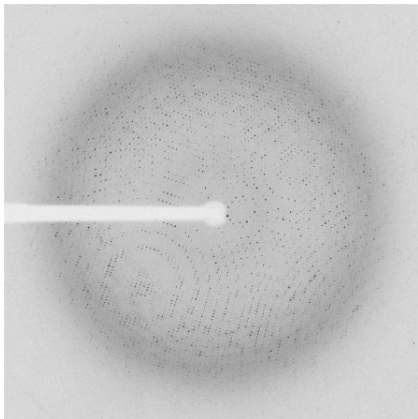


Inzulin

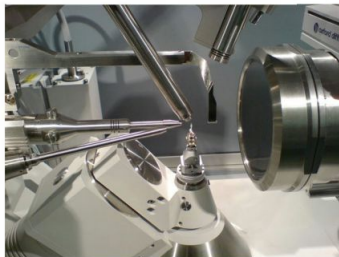


Krystalografické - proteinové

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



Experimental set-up



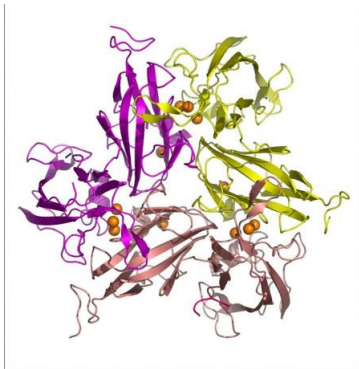
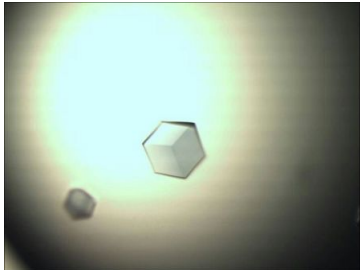
Krystalografické - proteinové

<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

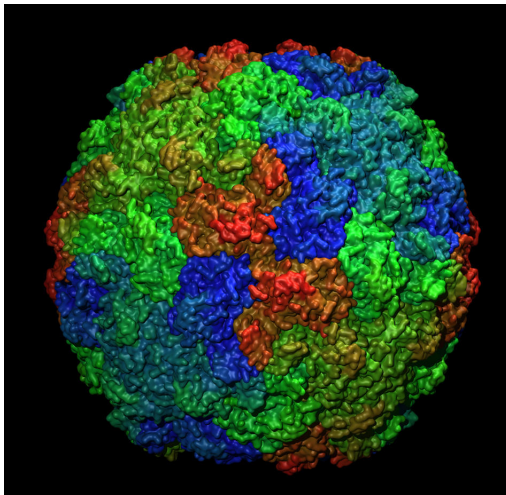
The Biologically Interesting Molecule Reference Dictionary (BIRD)

<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



AI pomocníci

- <https://www.sqlai.ai/app>
- <https://www.text2sql.ai/>
- <https://ai2sql.io/>
- <https://www.askyourdatabase.com/>
- <https://www.blazesql.com/>
- <https://deepnote.com>