

C2131 Úvod do bioinformatiky
Masarykova univerzita, PřF

Malé molekuly

C2131 Úvod do bioinformatiky, jaro 2023

C2131 Úvod do bioinformatiky
Masarykova univerzita, PřF

Lenka Malinová

Malé molekuly

- Malé molekuly – nízkomolekulární látky: látky o **relativně** nízké molekulové hmotnosti (konkrétní čísla se liší dle zdrojů ☺, < 900, < 1000, < 3000).
- Farmakologická definice: látky s M_r menší než 900, mohou difundovat přes membrány.
- V přírodních vědách nás většinou zajímají **biomolekuly a organické** nízkomolekulární látky.

Malé molekuly:

- **Podjednotky** makromolekul a **stavební prvky** buňky (aminokyseliny, nukleotidy, sacharidy, lipidy)
- Vitaminy, hormony, neurotransmitery
- Kofaktory, koenzymy
- Metabolity
- Sekundární metabolity

Malé molekuly

- Malé molekuly – nízkomolekulární látky: látky o **relativně** nízké molekulové hmotnosti (konkrétní čísla se liší dle zdrojů ☺, < 900, < 1000, < 3000).
- V přírodních vědách nás většinou zajímají **biomolekuly a organické** nízkomolekulární látky.

Malé molekuly:

- **Léčiva/toxiny** (což je většinou ta samá látka, záleží jen na množství...)

"Všechny sloučeniny jsou jedy. Neexistuje sloučenina, která by jedem nebyla. Rozdíl mezi lékem a jedem tvoří dávka."
Paracelsus

Často nás ale zajímají také:

- Ionty kovů, soli, molekuly rozpouštědla (voda), plyny (CO₂, NO) a jejich **vliv** na a **interakce** s biomakromolekulami.

Malé molekuly a chemoinformatika

- **Chemoinformatika:** Vědní odvětví vzniklé z požadavků farmaceutického průmyslu v kombinaci s obrovským nárůstem informací o (nejen) strukturách organických molekul. V současnosti hraje významnou roli také v chemii, biologii a biochemii.
- **Chemoinformatika** je kombinace „of all the information resources that a scientist needs to optimize the properties of a ligand to become a drug“. *Frank Brown, 1998*
- **Chemoinformatika:** Vědní disciplína, která využívá výpočetní techniku pro shromažďování, uchovávání, analýzu a využití chemických dat. Tato chemická data zahrnují informace o vzorcích, strukturách, vlastnostech, spektrech a aktivitách **malých molekul**.

„Cheminformatics“ – severní Amerika

„Chemoinformatics“ – všude jinde

Introduction to Cheminformatics

David S. Wishart¹

¹Departments of Computing Science and Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

Malé molekuly a chemoinformatika

- **Chemoinformatika:** Vědní odvětví vzniklé z požadavků farmaceutického průmyslu v kombinaci s obrovským nárůstem informací o (nejen) strukturách organických molekul. V současnosti hraje významnou roli také v chemii, biologii a biochemii.
- **Chemoinformatika: Chcete vědět víc?**

C2133 Úvod do chemoinformatiky, podzim 2023

▣ Cíle předmětu

Kurs je zaměřen na získání úvodních znalostí v oblasti chemoinformatiky. Základní ideou předmětu je vysvětlit nejdůležitější chemoinformatické metodiky pro práci s informacemi o malých organických molekulách.

Na konci tohoto kurzu bude student schopen:

Porozumět a reprezentovat 1D, 2D a 3D struktury malých organických molekul v počítači;

Porozumět metodikám pro predikci aktivity a vlastností molekuly a pro charakterizaci podobnosti molekuly.

Porozumět základním přístupům předpovědi vztahu struktura – aktivita – biologická funkce

▣ Osnova

1. Co je chemoinformatika - popis oboru, jeho historie a využití
2. Databáze malých organických molekul
3. Zápis molekuly pomocí řetězce (SMILES, InChi, InChiKey), problém unikátního zápisu
4. Reprezentace a manipulace s dvoudimenzionálními (2D) strukturami molekul
5. Způsoby vyjádření 3D struktury molekuly a molekulárního system
6. Molekulární deskriptory
7. Fingerpriny - vektory popisující struktury a vlastnosti molekul
8. Modely pro studium kvantitativních vztahů mezi strukturou a aktivitou látek – QSAR
9. Modely pro studium kvantitativních vztahů mezi strukturou a vlastnostmi látek – QSPR
10. 3D QSAR a 3D QSPR modely
11. Metody pro měření molekulární podobnosti
12. Chemický prostor a metodiky práce s ním

Bioinformatika a chemoinformatika

- **Bioinformatika:**

- Většina zdrojů volně dostupná
- Databáze nevznikly pro farmakologické využití; sekvenční záznamy často **nejsou propojeny** s informacemi o souvisejících léčivech nebo nemocích.

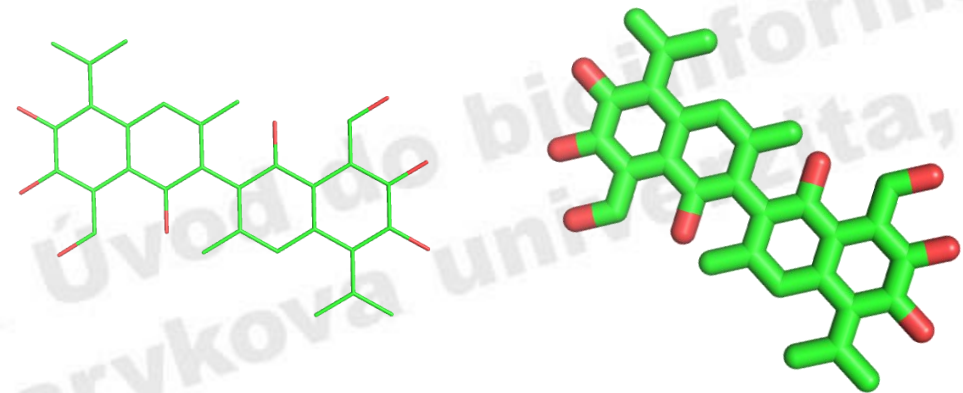
```
ATGCTGGTGATTGTGGATGCCGTTACCCTGCTGAGCGCCTA  
TCCGGAAGCCAGCCGTGATCCGGCCGCCCGACCGTGATTG  
ATGGTCGCCACCTGTATGTTGTTAGCCC GGCGATGCCGC
```

```
MLVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVS PGDA
```

Informace uložené v sekvenci monomerů

- **Chemoinformatika:**

- Mnoho zdrojů bylo (je) komerčních a volně nedostupných
- Databáze nevznikly pro biologické účely; chemická data často **nejsou propojena** s biologickými daty.



Vlastnosti podmíněné strukturou

Bioinformatika a chemoinformatika

- **Bioinformatika:**

- Většina zdrojů volně dostupná
- Databáze nevznikly pro farmakologické využití; sekvenční záznamy často **nejsou propojeny** s informacemi o souvisejících léčivech nebo nemocích.

- **Chemoinformatika:**

- Mnoho zdrojů bylo (je) komerčních a volně nedostupných
- Databáze nevznikly pro biologické účely; chemická data často **nejsou propojena** s biologickými daty.

Makromolekuly jsou složeny z malých molekul (aminokyseliny, nukleotidy).

Makromolekuly a malé molekuly souvisejí i funkčně.

Léčiva cílí na makromolekuly (proteiny, DNA).

Proteiny řídí syntézu a degradaci malých molekul (metabolity).

Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Sequence string format	FASTA (UNIT 3.9, Pearson, 2004), FASTQ	Chemical string format	SMILES, InChI

Bioinformatika: lineární reprezentace sekvence makromolekul.

Chemoinformatika: lineární reprezentace struktur.

ATGCTGGTGATTGTGGATGCCGTTACCCTGCTGAGCGCCTA
TCCGGAAGCCAGCCGTGATCCGGCCGCCCCGACCGTGATTG
ATGGTCGCCACCTGTATGTTGTTAGCCCGGGCGATGCCGC

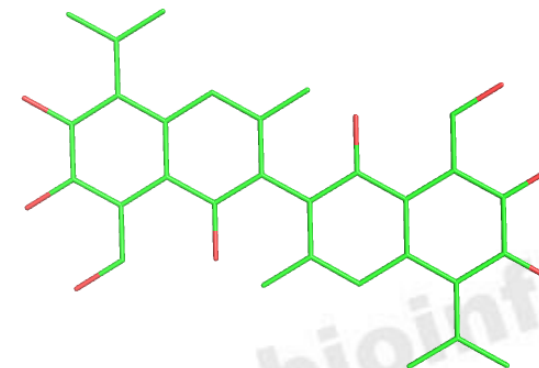
MLVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVSPGDA

FASTA formát

>název(_popis dle vlastní volby)↵

SEKVENCESEKVENCESEKVENCESEKVENCESEKVENCESEKVENCE↵

POVINNÉ VOLITELNÉ



InChI=1S/C30H30O8/c1-11(2)19-15-7-13(5)21(27(35)23(15)17(9-31)25(33)29(19)37)22-14(6)8-16-20(12(3)4)30(38)26(34)18(10-32)24(16)28(22)36/h7-12,33-38H,1-6H3

C12C=C(C(C3C(C)=CC4C(C(C)C)=C(C(O)=C(C=O)C=4C=3O)O)=C(O)C=1C(C=O)=C(C(O)=C2C(C)C)O)C

Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Sequence string format	FASTA (UNIT 3.9, Pearson, 2004), FASTQ	Chemical string format	SMILES, InChI

- **SMILES** (Simplified Molecular Input Line Entry System/Specification)

Bonds are denoted as shown below:

-	Single bond
=	Double bond
#	Triple bond
*	Aromatic bond
.	Disconnected structures

Some examples:

CC	CH ₃ CH ₃	Ethane
C=C	CH ₂ CH ₂	Ethene
CBr	CH ₃ Br	Bromomethane
C#N	C≡N	Hydrocyanic acid
Na.Cl	NaCl	Sodium chloride

A branch from a chain is specified by placing the **SMILES** symbol(s) for the branch between parenthesis. The string in parentheses is placed directly after the symbol for the atom to which it is connected. If it is connected by a double or triple bond, the bond symbol immediately follows the left parenthesis. Some examples:

CC(O)C	2-Propanol
CC(=O)C	2-Propanone
CC(CC)C	2-Methylbutane
CC(C)CC(=O)	2-Methylbutanal
c1c(N(=O)=O)cccc1	Nitrobenzene
CC(C)(C)CC	2,2-Dimethylbutane

2. SMILES

Název formátu SMILES je zkratkou pro Simplified Molecular Input Line Entry Specification⁹. Formát samotný byl navržen pro použití lidmi, zápis sloučeniny se podobá „normálnímu“ zápisu chemických struktur. Byl vytvořen v 80. letech 20. století, přesto nejeví žádné známky zastarávání a je stále velmi populární a využívaný. Nově je dostupná i aktualizovaná formální specifikace, dostupná pod svobodnou licenci¹⁰. Umožňuje – ale nevyžaduje – tzv. kanonickou formu⁹.

Chem. Listy 111, 710–715 (2017)

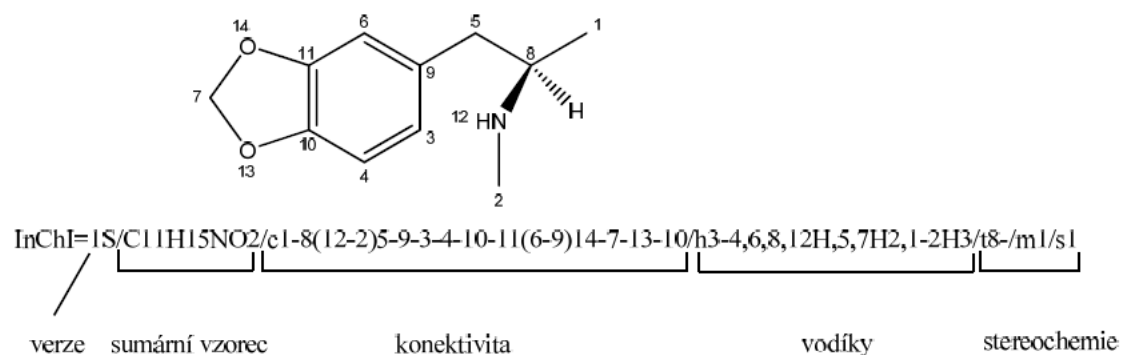
LINEÁRNÍ REPREZENTACE CHEMICKÝCH STRUKTUR

JIŘÍ JIRÁT^{a,b} a DANIEL SVOZIL^{a,b}

Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Sequence string format	FASTA (UNIT 3.9, Pearson, 2004), FASTQ	Chemical string format	SMILES, InChI

- InChI (IUPAC International Chemical Identifier)
- IUPAC - International Union of Pure and Applied Chemistry)



Obr. 3. Struktura identifikátoru InChI, hlavní vrstvy identifikátoru

3. InChI (a InChIKey)

Jedná se o relativně velice mladý formát (vznik kolem r. 2005), je produktem spolupráce organizací IUPAC a NIST (IUPAC International Chemical Identifier)¹¹⁻¹³. Jednou z hlavních ideí bylo vytvoření neutrálního standardního identifikátoru chemických struktur založeného na struktuře sloučeniny – např. CAS RN je vázané na přidělení službou Chemical Abstracts Service, a není dostupné pro látky, které jsou např. jen hypoteticky předpovězené při generování chemického prostoru². Standardy InChI se stále intenzivně vyvíjejí a jsou přidávána nová rozšíření a vlastnosti, např. standard RInChI pro zaznamenávání chemických reakcí¹⁴.

Chem. Listy 111, 710–715 (2017)

LINEÁRNÍ REPREZENTACE CHEMICKÝCH STRUKTUR

Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Archival sequence databases	GenBank	Archival compound databases	PubChem
Curated databases	UniProt (UNIT 1.27, Pundir et al., 2015), RefSeq, FlyBase, SGD (UNIT 1.20, Skrzypek and Hirschman, 2011), HPRD	Curated databases	ChEBI (UNIT 14.9, Degtyarenko et al., 2009), KEGG (UNIT 1.12, Tanabe and Kanehisa, 2012), DrugBank, PharmGKB (UNIT 14.7, Gong et al., 2008), HMDB

PubChem

<https://pubchem.ncbi.nlm.nih.gov/>

Explore Chemistry

Quickly find chemical information from authoritative sources

Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Use Entrez

Compounds

Substances

BioAssays



Draw Structure



Upload ID List



Browse Data



Periodic Table

PubChem is an open chemistry database at the National Institutes of Health (NIH). "Open" means that you can put your scientific data in PubChem and that others may use it. Since the launch in 2004, PubChem has become a key chemical information resource for scientists, students, and the general public. Each month our website and programmatic services provide data to several million users worldwide.

PubChem mostly contains small molecules, but also larger molecules such as nucleotides, carbohydrates, lipids, peptides, and chemically-modified macromolecules. We collect information on chemical structures, identifiers, chemical and physical properties, biological activities, patents, health, safety, toxicity data, and many others.

Where does the data in PubChem come from? PubChem records are contributed by hundreds of data sources. Examples include: government agencies, chemical vendors, journal publishers, and more.

Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Archival sequence databases	GenBank	Archival compound databases	PubChem
Curated databases	UniProt (UNIT 1.27, Pundir et al., 2015), RefSeq, FlyBase, SGD (UNIT 1.20, Skrzypek and Hirschman, 2011), HPRD	Curated databases	ChEBI (UNIT 14.9, Degtyarenko et al., 2009), KEGG (UNIT 1.12, Tanabe and Kanehisa, 2012), DrugBank, PharmGKB (UNIT 14.7, Gong et al., 2008), HMDB

PubChem

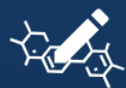
<https://pubchem.ncbi.nlm.nih.gov/>

Explore Chemistry

Quickly find chemical information from authoritative sources

Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Use Entrez Compounds Substances BioAssays



Draw Structure



Upload ID List



Browse Data



Periodic Table

COMPOUND BEST MATCH



Aspirin; ACETYLSALICYLIC ACID; 50-78-2; 2-Acetoxybenzoic Acid; 2-(Acetyloxy)Benzoic Acid; O-Acetylsalicylic Acid; O-Acetoxybenzoic Acid; Acylpyrin; ...

Compound CID: 2244

MF: C₉H₈O₄ MW: 180.16g/mol

InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

IUPAC Name: 2-acetyloxybenzoic acid

Create Date: 2004-09-16

Tagged by PubChem: COVID-19; COVID19; Coronavirus; Corona-virus; SARS; SARS2; SARS-CoV; SARS-CoV-2 [as per clinicaltrial; clinicaltrials; clinical trial; clinical trials; COVID-19 Disease Map]

Summary

Similar Structures Search

Related Records

PubMed (MeSH Keyword)



Aspirin; ACETYLSALICYLIC ACID; 50-78-2; 2-Acetoxybenzoic Acid; 2-(Acetyloxy)Benzoic Acid; O-Acetoxybenzoic Acid; Acylpyrin; O-Acetylsalicylic Acid; ...

Compound CID: 2244

MF: $C_9H_8O_4$ MW: 180.16g/mol

InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

IUPAC Name: 2-acetoxybenzoic acid

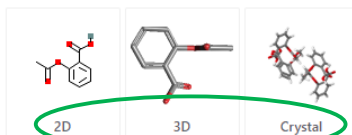
Create Date: 2004-09-16

[Summary](#) [Similar Structures Search](#) [Related Records](#) [PubMed \(MeSH Keyword\)](#)

Aspirin

PubChem CID: 2244

Structure:



[Find Similar Structures](#)

Chemical Safety:



Irritant

[Laboratory Chemical Safety Summary \(LCSS\) Datasheet](#)

Molecular Formula: $C_9H_8O_4$ or $CH_3COOC_6H_4COOH$ or $HC_9H_7O_4$

Synonyms:
aspirin
ACETYLSALICYLIC ACID
50-78-2
2-Acetoxybenzoic acid
2-(Acetyloxy)benzoic acid

[More...](#)

Molecular Weight: 180.16 g/mol

Dates: Modify: 2020-04-25 Create: 2004-09-16

Aspirin is an orally administered non-steroidal antiinflammatory agent. Acetylsalicylic acid binds to and acetylates serine residues in cyclooxygenases, resulting in decreased synthesis of prostaglandin, platelet aggregation, and inflammation. This agent exhibits analgesic, antipyretic, and anticoagulant properties.

Links from PubChem Compound

Items: 1 to 20 of 44549

<< First < Prev Page 1 of 2228 Next > Last >>

- [Essential Thrombocythemia_Reply](#)
 1. Tefferi A, Pardanani A. N Engl J Med. 2020 Mar 19;382(12):e21. doi: 10.1056/NEJMc1917586. No abstract available. PMID: 32187488 [Similar articles](#)
- [Essential Thrombocythemia](#)
 2. Rumi E, Cazzola M. N Engl J Med. 2020 Mar 19;382(12):e21. doi: 10.1056/NEJMc1917586. No abstract available. PMID: 32187487 [Similar articles](#)
- [Essential Thrombocythemia](#)
 3. Fleischman RA. N Engl J Med. 2020 Mar 19;382(12):e21. doi: 10.1056/NEJMc1917586. No abstract available. PMID: 32187486 [Similar articles](#)
- [Ticagrelor with or without Aspirin after PCI_Reply](#)
 4. Baber U, Dangas G, Mehran R. N Engl J Med. 2020 Mar 12;382(11):1076-1077. doi: 10.1056/NEJMc1917041. No abstract available. PMID: 32160679 [Similar articles](#)



Methyl Salicylate; Methyl 2-Hydroxybenzoate; 119-36-8; Wintergreen Oil; Gaultheria Oil; ...

Compound CID: 4133

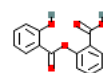
MF: $C_8H_8O_3$ MW: 152.15g/mol

InChIKey: OSWPMRLSEHDHFF-UHFFFAOYSA-N

IUPAC Name: methyl 2-hydroxybenzoate

Create Date: 2005-03-25

[Summary](#) [Similar Structures Search](#) [Related Records](#)



Salsalate; Sasapyrine; 552-94-3; Salicylsalicylic Acid; Disalicylic Acid; ...

Compound CID: 5161

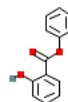
MF: $C_{14}H_{10}O_5$ MW: 258.23g/mol

InChIKey: WVYADZUPLLSGPU-UHFFFAOYSA-N

IUPAC Name: 2-(2-hydroxybenzoyl)oxybenzoic acid

Create Date: 2005-03-25

[Summary](#) [Similar Structures Search](#) [Related Records](#)



Phenyl Salicylate; 118-55-8; PHENYL 2-HYDROXYBENZOATE; Salol; Phenol Salicylate; ...

Compound CID: 8361

MF: $C_{13}H_{10}O_3$ MW: 214.22g/mol

InChIKey: ZQBAKBUEJOMQEX-UHFFFAOYSA-N

IUPAC Name: phenyl 2-hydroxybenzoate

Create Date: 2005-03-26

Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Archival sequence databases	GenBank	Archival compound databases	PubChem
Curated databases	UniProt (UNIT 1.27, Pundir et al., 2015), RefSeq, FlyBase, SGD (UNIT 1.20, Skrzypek and Hirschman, 2011), HPRD	Curated databases	ChEBI (UNIT 14.9, Degtyarenko et al., 2009), KEGG (UNIT 1.12, Tanabe and Kanehisa, 2012), DrugBank, PharmGKB (UNIT 14.7, Gong et al., 2008), HMDB

PubChem

<https://pubchem.ncbi.nlm.nih.gov/>

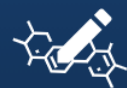
Explore Chemistry

Quickly find chemical information from authoritative sources

C12C=C(C(C3C(C)=CC4C(C(C)C)=C(C(O)=C(C=O)C=4C=3O)O)=C(O)C=1C(C=O)=C(C(O)=C2C(C)O)C

Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Use Entrez Compounds Substances BioAssays



Draw Structure



Upload ID List

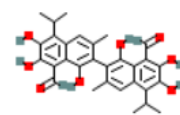


Browse Data



Periodic Table

1 result



Gossypol; 303-45-7; (-)-Gossypol; 90141-22-3; Pogosin; ...

Compound CID: 3503

MF: C₃₀H₃₀O₈ MW: 518.6g/mol

InChIKey: QBKSWRVVCFDOT-UHFFFAOYSA-N

IUPAC Name: 7-(8-formyl-1,6,7-trihydroxy-3-methyl-5-propan-2-yl)naphthalen-2-yl)-2,3,8-trihydroxy-6-methyl-4-propan-2-yl)naphthalene-1-carbaldehyde

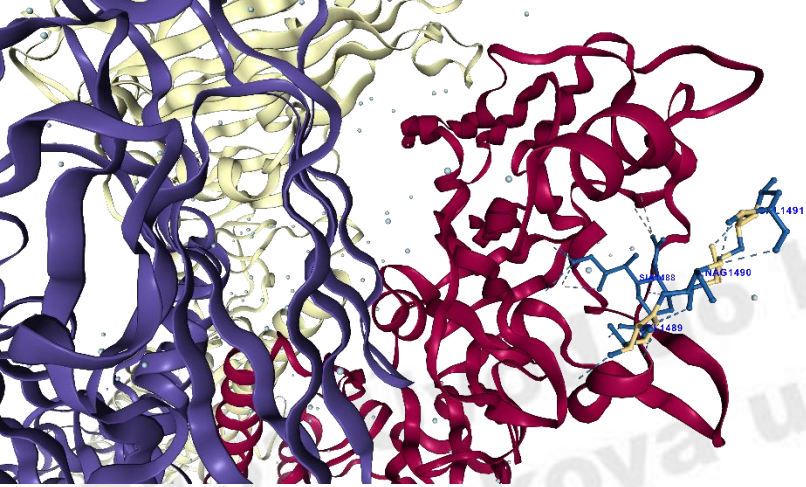
Create Date: 2005-03-25

Summary

Similar Structures Search

Related Records

PubMed (MeSH Keyword)

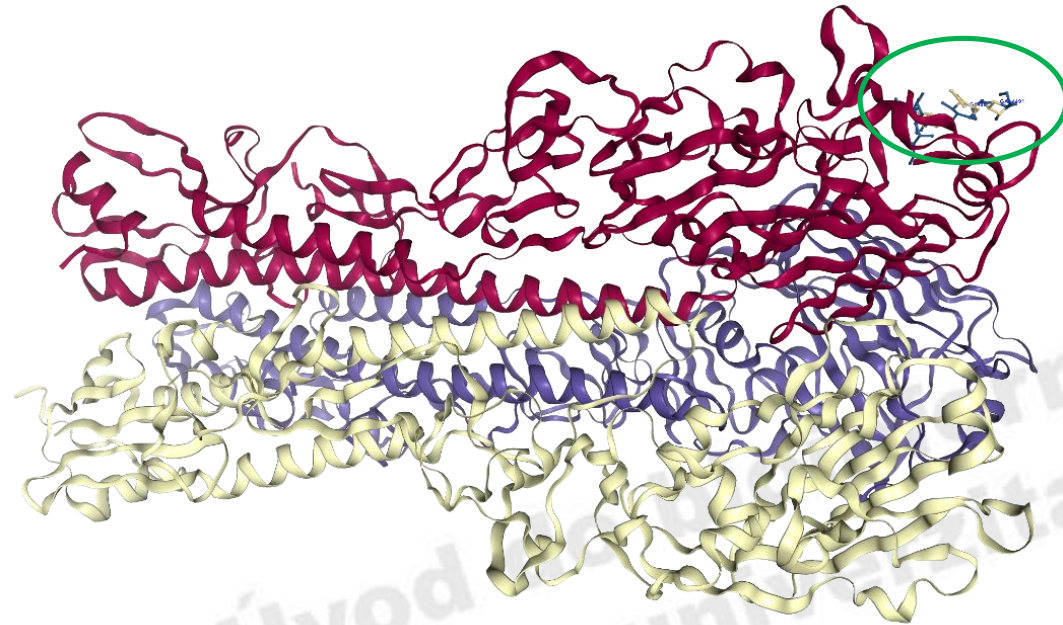
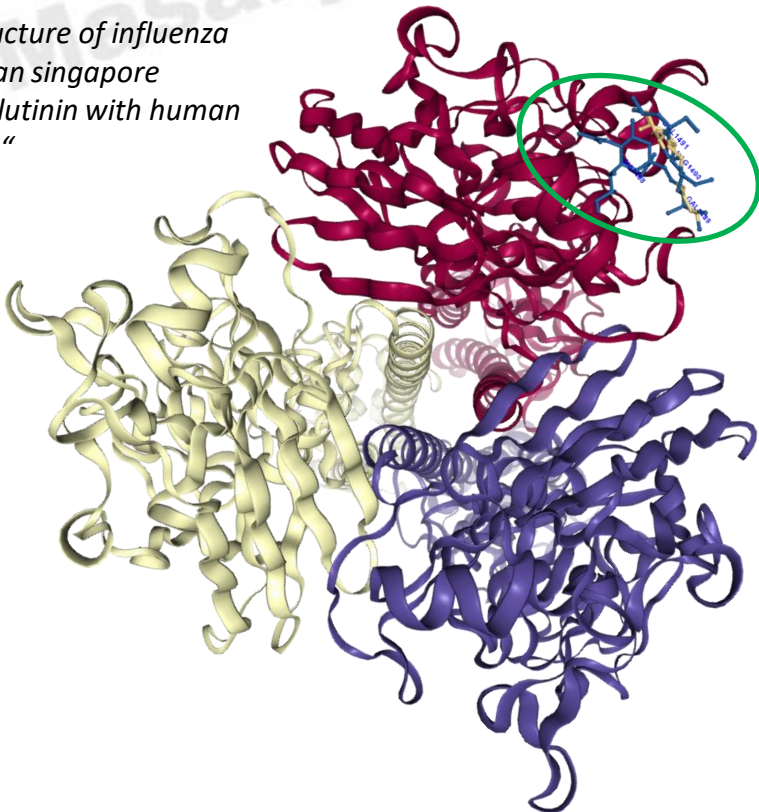


Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Structural databases	PDB (UNIT 1.9, Dutta et al., 2007), PDBe	Structural databases	CSD, Ligand Expo

PDB ID: 2WR7

„The structure of influenza H2 human singapore hemagglutinin with human receptor“



Malé molekuly – **ligandy** makromolekul

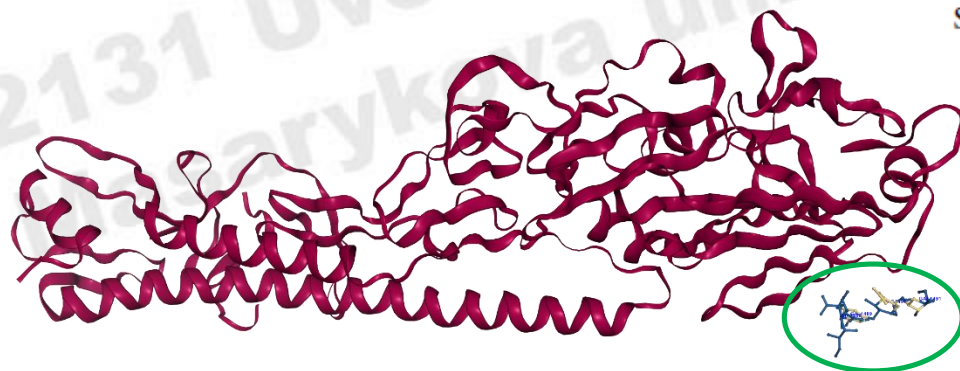
Řešení **3D struktur komplexů** – spojuje chemo a bioinformatiku

Ligand Expo

<http://ligand-expo.rcsb.org/>

Bioinformatika a chemoinformatika

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Structural databases	PDB (UNIT 1.9, Dutta et al., 2007), PDBe	Structural databases	CSD, Ligand Expo



Ligand Expo

Ligand Expo (formerly Ligand Depot) provides chemical and structural information about small molecules within the structure entries of the Protein Data Bank. Tools are provided to search the PDB dictionary for chemical components, to identify structure entries containing particular small molecules, and to download the 3D structures of the small molecule components in the PDB entry. A sketch tool is also provided for building new chemical definitions from reported PDB chemical components.

Getting Started:

The [Ligand Expo Tutorial](#) contains a summary of query, report and download features. Select the [Search](#) menu to find a chemical component or build a new chemical component. Select the [Browse](#) menu to view tabulations of modified amino acids and nucleotides. Select the [Download](#) for links to chemical dictionaries and resource files. Go [here](#) for a brief tutorial on the content of the Chemical Component Dictionary.

Ligandy – důvěřuj ale prověřuj!

Validator^{DB}

Database of validation results for ligands and non-standard residues in the Protein Data Bank.



[Guide](#) [User](#) [Wiki Manual](#)

Database last updated 27/4/2023: 202133 entries from PDBe.org, 750774 molecules relevant for validation, 36260 models from wwPDB CCD.

The molecules deemed relevant for validation are all ligands and non-standard residues with reasonable size (more than six heavy atoms). Standard amino acids and nucleotides are not covered. The validation is performed against models from wwPDB Chemical Component Dictionary (wwPDB CCD). The database is updated weekly.

[Quick Help](#) [Samples](#) [Search](#) [Overview](#) [Details by Molecule](#) [Details by PDB Entry](#)

This is the Validator^{DB} **synopsis page**. Access different tabs for overviews and statistical evaluation of the validation results, in graphical or tabular form. Specific results can be examined in deeper detail by accessing the Validator^{DB} **specifics page**.

Different sections of the web page offer interactive guides indicated by  which give a quick walk through all the main elements of the page. Further help is provided by the info icons . Many tool tips are available by hovering over any graphical or textual element in the interface.

Refer to the [Wiki Manual](#) for any clarifications (especially [Terminology](#) and [User Interface](#)).

First Time User? Check out the [First Time User Presentation](#).

ValidatorDB: database of up-to-date validation results for ligands and non-standard residues from the Protein Data Bank

David Sehnal^{1,2,3,†}, Radka Svobodová Vařeková^{1,2,†}, Lukáš Pravda^{1,2}, Crina-Maria Ionescu¹, Stanislav Geidl^{1,2}, Vladimír Horský³, Deepti Jaiswal¹, Michaela Wimmerová^{1,2} and Jaroslav Koča^{1,2,*}

ABSTRACT

Following the discovery of **serious errors** in the structure of biomacromolecules, structure validation has become a key topic of research, especially for ligands and non-standard residues. ValidatorDB (freely available at <http://ncbr.muni.cz/ValidatorDB>) offers a new step in this direction, in the form of a database of validation results for all ligands and non-standard residues from the Protein Data Bank (all molecules with seven or more heavy atoms). Model molecules from the wwPDB Chemical Component Dictionary are used as reference during validation. ValidatorDB covers the main aspects of validation of annotation, and additionally introduces several useful validation analyses. The most significant is the classification of chirality errors, allowing the user to distinguish between serious issues and minor inconsistencies. Other such analyses are able to report, for example, completely erroneous ligands, alternate conformations or complete identity with the model molecules. All results are systematically classified into categories, and statistical evaluations are performed. In addition to detailed validation reports for each molecule, ValidatorDB provides summaries of the validation results for the entire PDB, for sets of molecules sharing the same annotation (three-letter code) or the same PDB entry, and for user-defined selections of annotations or PDB entries.

Ligandy – PROVĚŘUJ!

PDBe > 3ib0

Structural basis of the prevention of NSAID-induced damage of the gastrointestinal tract by C-terminal half (C-lobe) of bovine colostrum protein lactoferrin: Binding and structural studies of C-lobe complex with diclofenac

Source organism: *Bos taurus*

Primary publication:

The structural basis for the prevention of nonsteroidal antiinflammatory drug-induced gastrointestinal tract damage by the C-lobe of bovine colostrum lactoferrin.

Mir R, Singh N, Vikram G, Kumar RP, Sinha M, Bhushan A, Kaur P, Srinivasan A, Sharma S, Singh TP

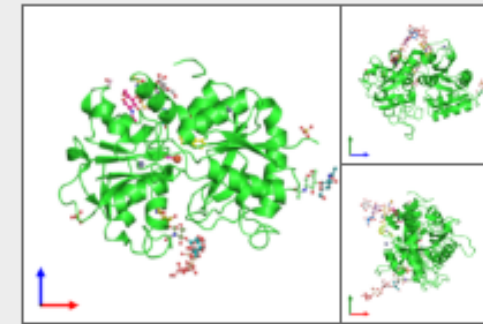
Biophys. J. **97** 3178-86 (2009)

PMID: 20006955

X-ray diffraction
1.4Å resolution

Released: 11 Aug 2009

Model geometry
Fit model/data



Quick links

3ib0 overview

- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

- View
- Downloads
- 3D Visualisation

Function and Biology

Details

- Biochemical function:** not assigned
- Biological process:** not assigned
- Cellular component:** not assigned

Sequence domain:

- Transferrin-like domain

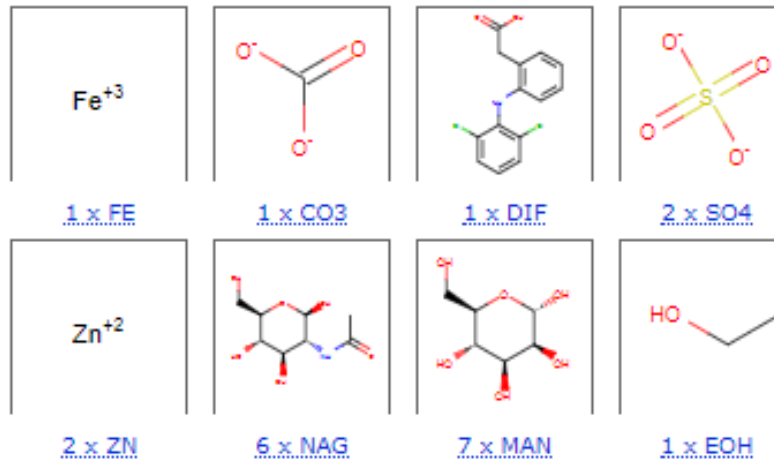
Structure analysis

Details

- Assembly composition:** monomeric (preferred)
- Entry contents:** 1 distinct polypeptide molecule
- Macromolecule:**
- Lactotransferrin

Ligands and Environments

8 bound ligands:



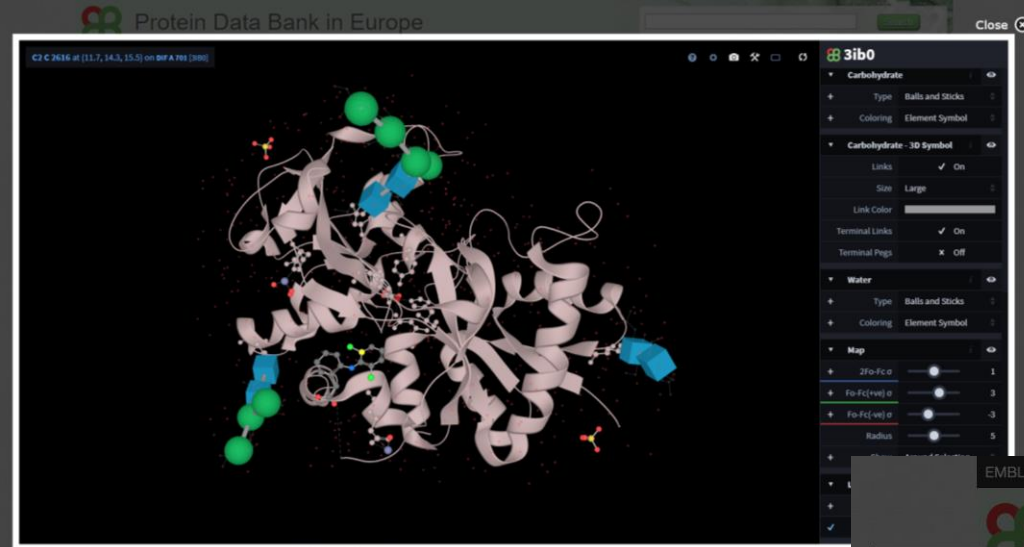
PDB-REDO

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

Model Geometry
Fit model/data

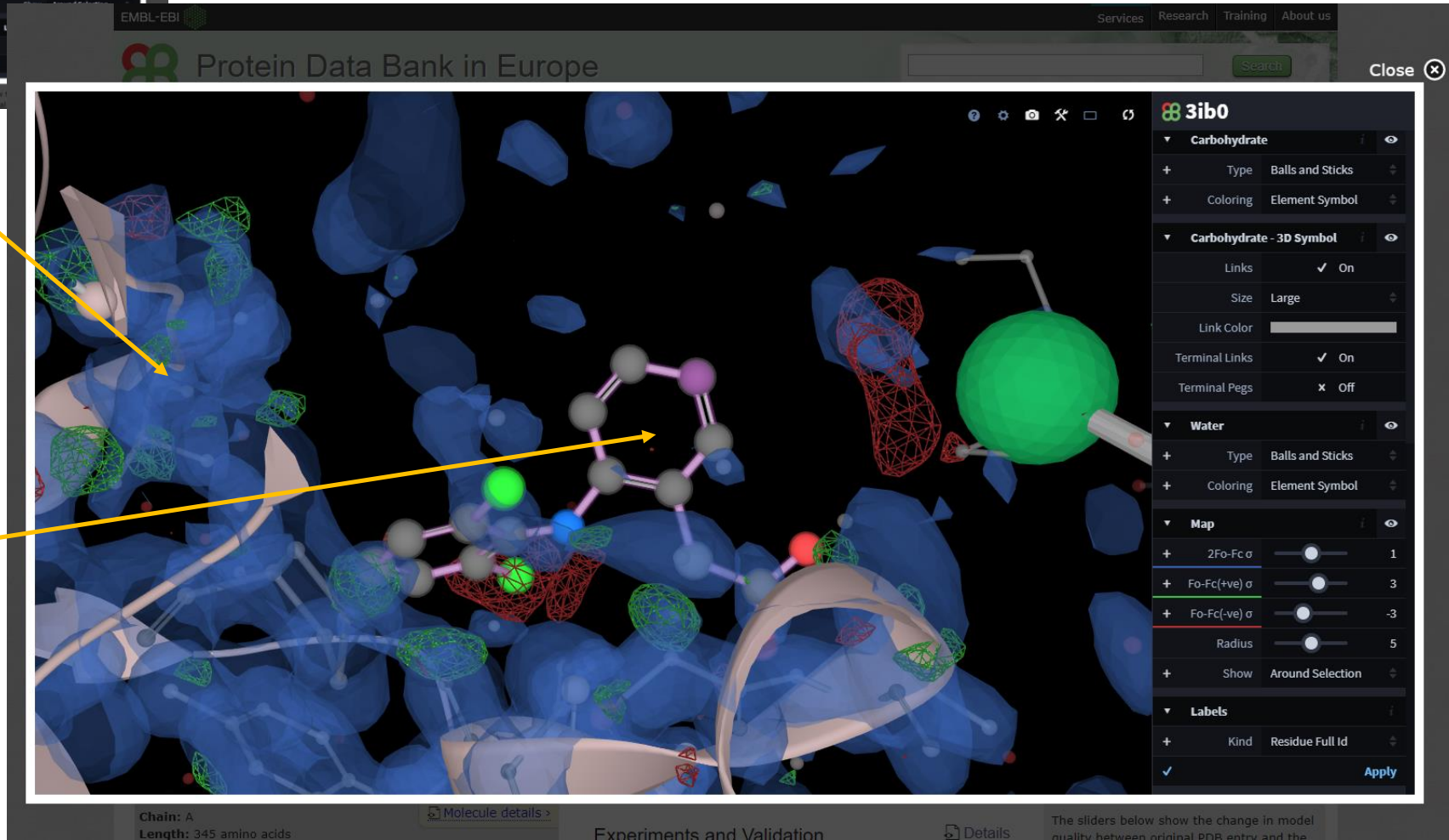
PDB-REDO

Ligandy – PROVĚŘUJ!



Modře je znázorněna elektronová hustota získaná při rentgenostrukturní analýze. Pomocí elektronové hustoty – dosazováním atomů do ní – řešíte strukturu molekuly.

Tam, kde autoři tvrdí, že se váže ligand, elektronová hustota zcela **chybí**. Může se stát, že ligand do elektronové hustoty nesedí – naváže se tam něco jiného než autoři chtějí, rozlišení je špatné, ligand je špatně orientován. Ale tady je přání otcem myšlenky 😊



Validation of ligands in macromolecular structures determined by X-ray crystallography

Oliver S. Smart,^a Vladimír Horský,^{b,c} Swanand Gore,^{a,†} Radka Svobodová Vařeková,^{b,c} Veronika Bendová,^{b,c,d} Gerard J. Kleywegt^a and Sameer Velankar^{a*}

Aby Vaše struktura byla využita jako příklad hned ve dvou článcích o chybách ve strukturách...to doufám nechcete...na své ligandy si dávejte pozor! 😊

Detect, Correct, Retract: How to manage incorrect structural models

Alexander Wlodawer^a, Zbigniew Dauter^b, Przemyslaw J. Porebski^c, Wladek Minor^c, Robyn Stanfield^d, Mariusz Jaskolski^{e,f}, Edwin Pozharski^g, Christian X. Weichenberger^h, and Bernhard Rupp^{h,i,*}

Summary: The crystal structures of the C-lobe of bovine colostrum lactoferrin do not support drug binding

In this series of structures, present in the PDB and the scientific literature, no evidence exists from crystallography that the purported NSAIDs bind where stated. Significant discrepancies exist between the electron density maps displayed in the original publications versus maps that can be generated by accepted standard procedures. Caution must be exercised when maps are constructed in unconventional ways. A possible explanation for the absence of ligands in this entire series of models could be the deposition of structure factor amplitudes corresponding to ligand-free (apo) structure. In this case, the authors would need to produce the correct original data. Here and in any similar cases, the journals where the original work was published should be informed of the problem, and an erratum linked to the new PDB deposit (if any), should be published.

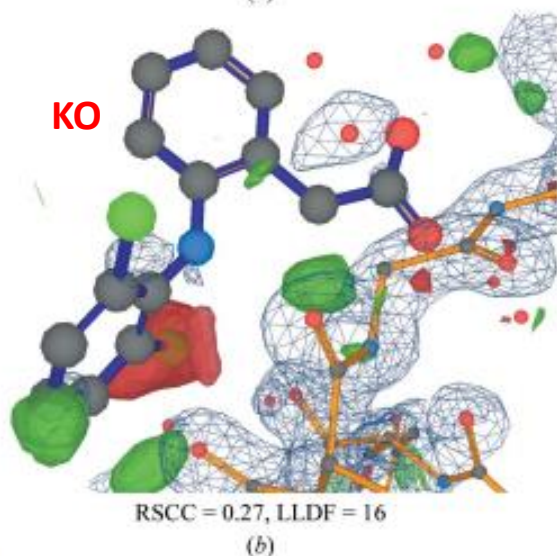
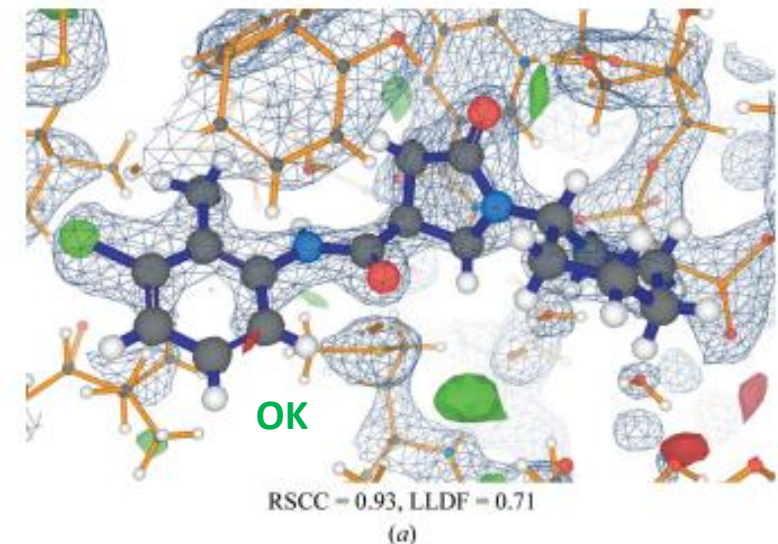
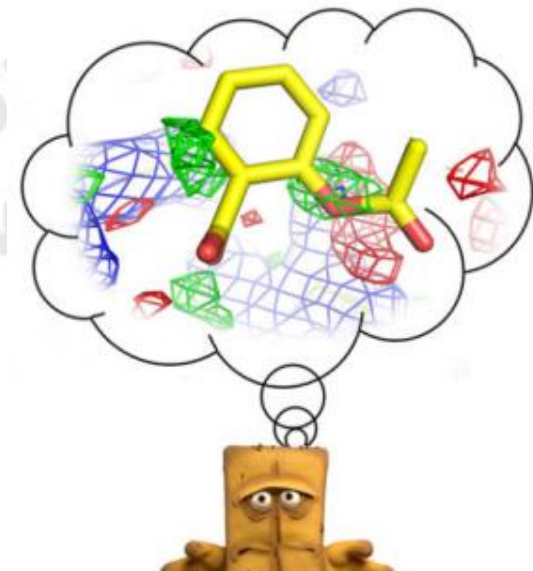


Figure 2
Visualization of two ligands from the PDB together with electron-density maps. The *LiteMol* viewer (Sehna *et al.*, 2017) on the PDB website (Mir *et al.*, 2018) has been used to visualize the ligands with EDS-style (Kleywegt *et al.*, 2004) electron-density maps. In each case, the $2mF_o - DF_c$ map is shown as a blue mesh contoured at $0.39 \text{ e } \text{Å}^{-3}$, whereas the $mF_o - DF_c$ difference map is shown by solid green and red surfaces contoured at $+0.25$ and $-0.25 \text{ e } \text{Å}^{-3}$, respectively. (a) Model and electron density for the well placed ligand 468 from PDB entry 4tzt (He *et al.*, 2006). (b) The ligand DIF in PDB entry 3ib0 (Mir *et al.*, 2009).

Bioinformatika a chemoinformatika

Absorption
Distribution
Metabolism
Excretion

Bioinformatics		Cheminformatics	
Type	Name	Type	Name
Property prediction	Hydrophobicity	Property prediction	LogP
Property prediction	pI	Property prediction	pKa
Property prediction	Solubility	Property prediction	Solubility
Property prediction	Molecular weight	Property prediction	Molecular weight

P – rozdělovací (partition) koeficient, vyjadřuje rozdělení látky mezi vodnou a organickou fází.
Např. voda/oktanol.
Jak hydrofilní/hydrofobní je daná chemická látka?

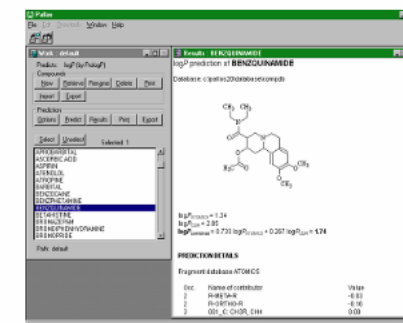
PrologP

Customer Benefits

The program provides an indispensable instrument for predicting logP values before synthesis. This can be found in diversity calculations for combinatorial chemistry, in QSAR, and in situations when the compound displays a logP value which is too low or too high. It is also useful if the compound decomposes during measurement, or when the compound is toxic.

Brief Description

CompuDrug's PrologP calculates the accurate logP values (negative logarithms of the n-octanol/water partition coefficient) for organic compounds, in most cases, within an error range that is not significantly higher than the measurement error of the actual logP value. Three calculation algorithms can perform calculations. Two of them are linear models: one is based on the Rekker fragmental method, the other on the Ghose-Crippen atomic method, while the third one is a neural network model. A further option of the program combines the optimum results obtained by the different models.



CompuDrug - Your Expert in Chemical Informatics

Welcome

CompuDrug is a niche software corporation which has been specializing in developing and generating ADME and drug discovery softwares focusing on ADME prediction and physicochemical data calculation.

The company first in the world developed softwares for prediction of metabolic fate, toxicity and true hydrophobicity of new drug candidates.

The major pharmaceutical, biotechnological and agrochemical research firms, as well as a large number of universities and governmental institutes throughout the world use CompuDrug's discovery softwares in their daily research and developmental practices to accelerate their drug and agrochemical discovery processes and to support regulatory or environmental decisions.

<https://www.compudrug.com/>

Chemoinformatica – další vlastnosti

- Kromě chemických vlastností nás často zajímají vlastnosti biologické např. toxicita, permeabilita (propustnost, gastrointestinální permeabilita, permeabilita mozku).
- Vývoj léčiv
- **Lipinského pravidlo pěti** definující dobré orálně užívané léčivo:

$\log P < 5$

molekulová hmotnost < 500 g/mol

< 5 donorů H-vazeb (OH a NH skupin)

< 10 (2x5) akceptorů H-vazeb (atomů O a N)

Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs

Miniperspective

Michael D. Shultz*

Reporting in the *Journal of Medicinal Chemistry*, Shultz assessed the physicochemical properties of the 409 small molecules that the FDA has approved since 1997. Molecular mass in this cohort has increased considerably from the baseline, he found, and in both 2016 and 2017 the average molecular mass of FDA-approved drugs was greater than 500 daltons. An up-to-date molecular mass cut-off based on the properties of orally available small molecules approved in the past decade would now be more than 600 daltons, he found.

The threshold for hydrogen bond acceptors has also increased substantially.

These data call into question the hypothesis that 'drug-like' properties exist, he concludes. "Repeating the [rule of 5] experiment today, with over twice as many oral drugs approved than were available in 1997, gives different 'rules' than those hypothesized two decades ago ... If our past future predictions have not been accurate with the [rule of 5] parameters, we must call into question our current future predictions."

With new modalities coming online — such as targeted protein degraders and constrained peptides — he urges drug developers to reconsider how they use the rule of 5. "We are in danger of repeating our past mistakes if we assume these new modalities are not 'drug-like' and cannot be oral drugs because they are not [rule of 5] compliant," writes Shultz.

Asher Mullard

Malé molekuly a metabolomika

- **Metabolismus** (látková přeměna): Souhrn všech reakcí v buňce při nichž dochází k přeměně látek. Buňka tak získává energii, stavební a zásobní látky. Reakce jsou uspořádané do sérií – metabolických drah.
- **Metabolit:** Sloučenina zařazená do metabolismu.
- **Sekundární metabolit:** Sloučenina, která není nezbytná pro základní životní funkce. Biosyntéza navazuje na primární metabolismus, její funkce není vždy známá.

- **Metabolom:** Komplexní sada metabolitů. Soubor všech intra- i extracelulárních nízkomolekulárních látek ($M_r < 1000$) v živém systému, které se účastní metabolických reakcí, a které jsou nezbytné pro růst a normální funkci buňky (tkáně, organismu).
- **Metabolomika:** Komplexní analýza metabolomu za konkrétního fyziologického nebo vývojového stádia organismu, tkáně či buňky.

Chem. Listy 105, 745–751 (2011)

METABOLOMIKA – ZÁKLADNÍ POJMY, STRATEGIE A METODOLOGIE

JINDRA MUSILOVÁ a ZDENĚK GLATZ

Ústav biochemie, Přírodovědecká fakulta a Středoevropský technologický institut, Masarykova univerzita, Kotlářská 2, 611 37 Brno
jmusilova@mail.muni.cz

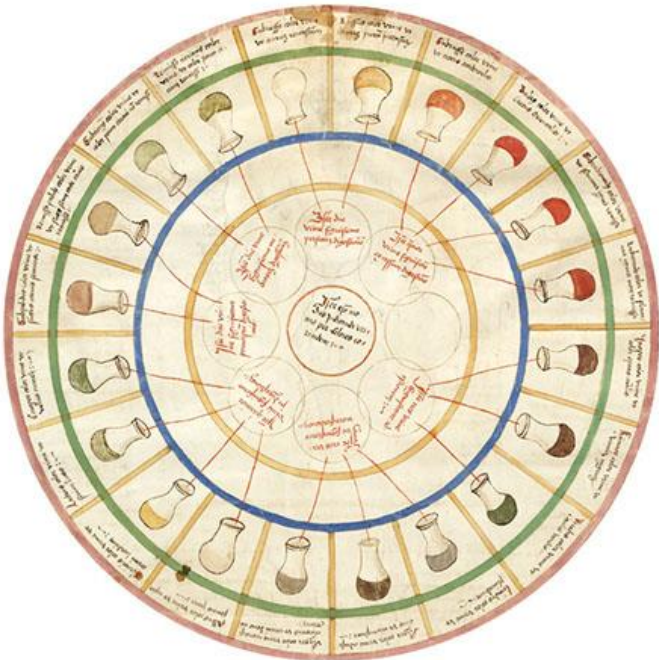
Malé molekuly a metabolomika

- **Metabolomika:** Komplexní analýza metabolomu za konkrétního fyziologického nebo vývojového stádia organismu, tkáně či buňky.
- Více odráží aktuální stav buňky (ve srovnání s transkriptomikou nebo proteomikou).
- Analýza tisíců různých sloučenin s extrémně variabilními strukturami a vlastnostmi.
- Stanovení **biomarkerů** jako indikátorů chorob.

- **Metabolomika:** Mladá věda, definována jako samostatný obor poměrně nedávno (Oliver S., 1998; Fiehn O., 2002).
- Ale s hlubokými kořeny. Za první využití metabolomiky bývá považována analýza tělních tekutin pro hodnocení zdravotního stavu.
- Čína, 2000 př.n.l.: využití **mravenců** ke stanovení koncentrace glukosy v moči diabetiků. (Lze si tady představit využití jednoduché kalibrační křivky podle počtu přilákaných mravenců).
- Egypt, Řecko, 300 př.n.l.: testování moči pacientů **ochutnáním**.

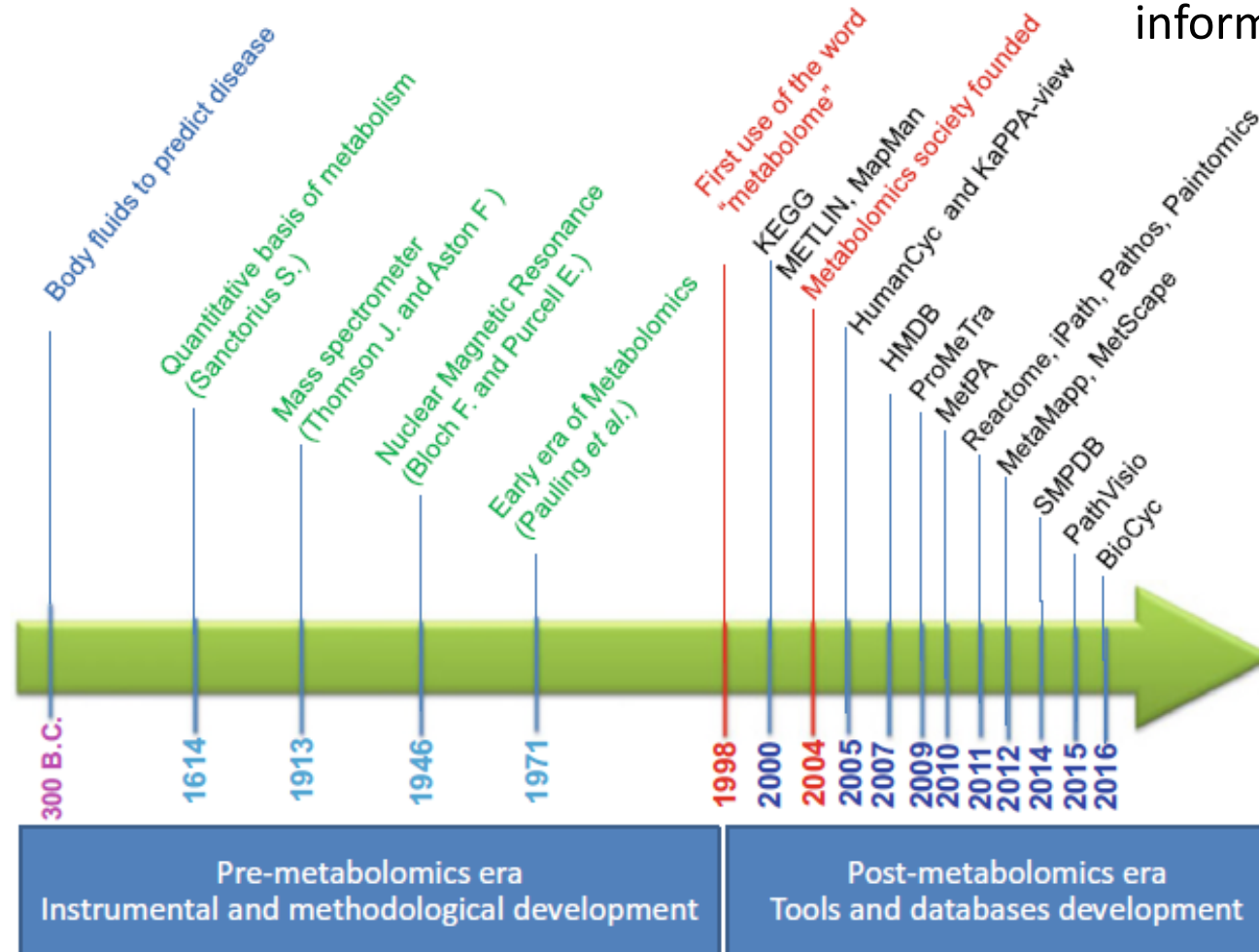
Malé molekuly a metabolomika

„The urine wheel“



Uroscopie (uromancie) – diagnostika chorob z moči

Obrovský nárůst metabolomických dat. Vznik databází, zapojení informačních technologií



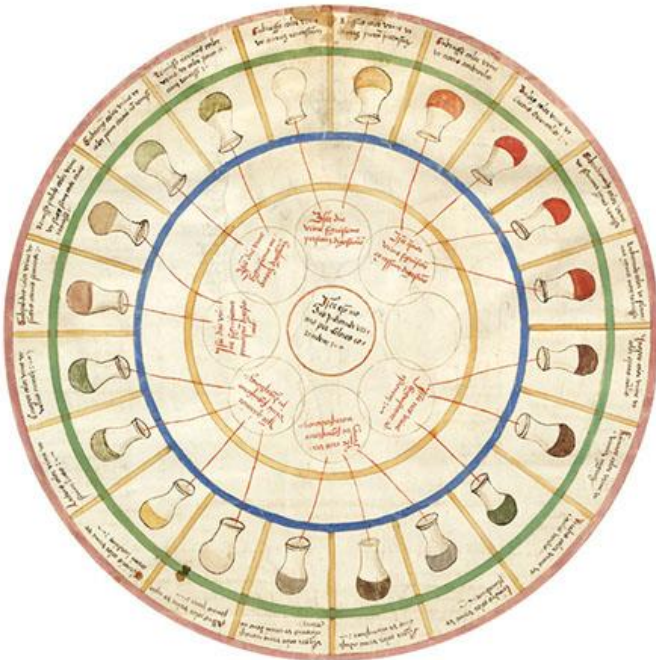
ormatiky
zita, PřF

Uroscopic rainbow: modern matula medicine

C L Foot, J F Fraser

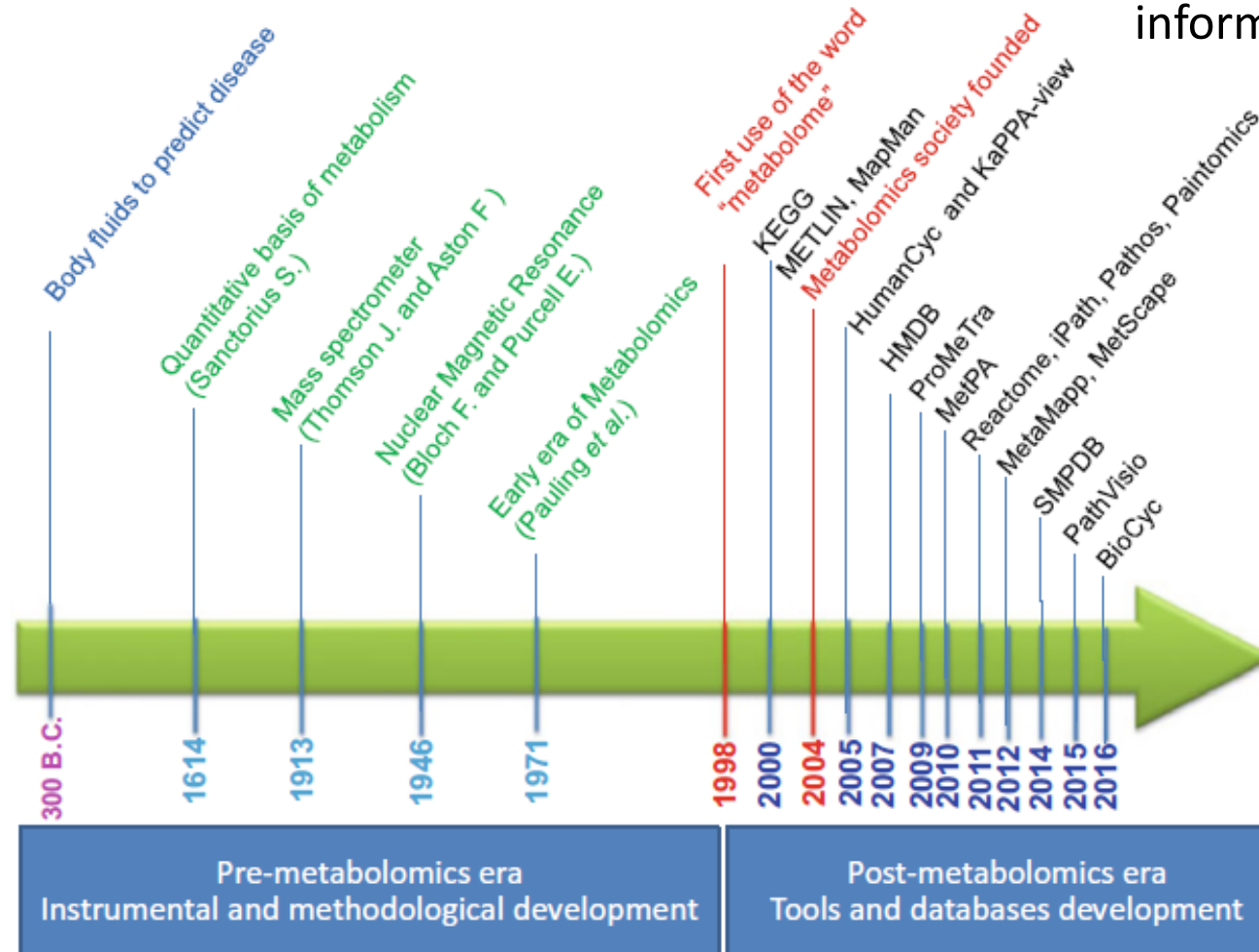


„The urine wheel“



Uroskopie (uromancie) – diagnostika chorob z moči

Obrovský nárůst metabolomických dat. Vznik databází, zapojení informačních technologií



Pre-metabolomics era
Instrumental and methodological development

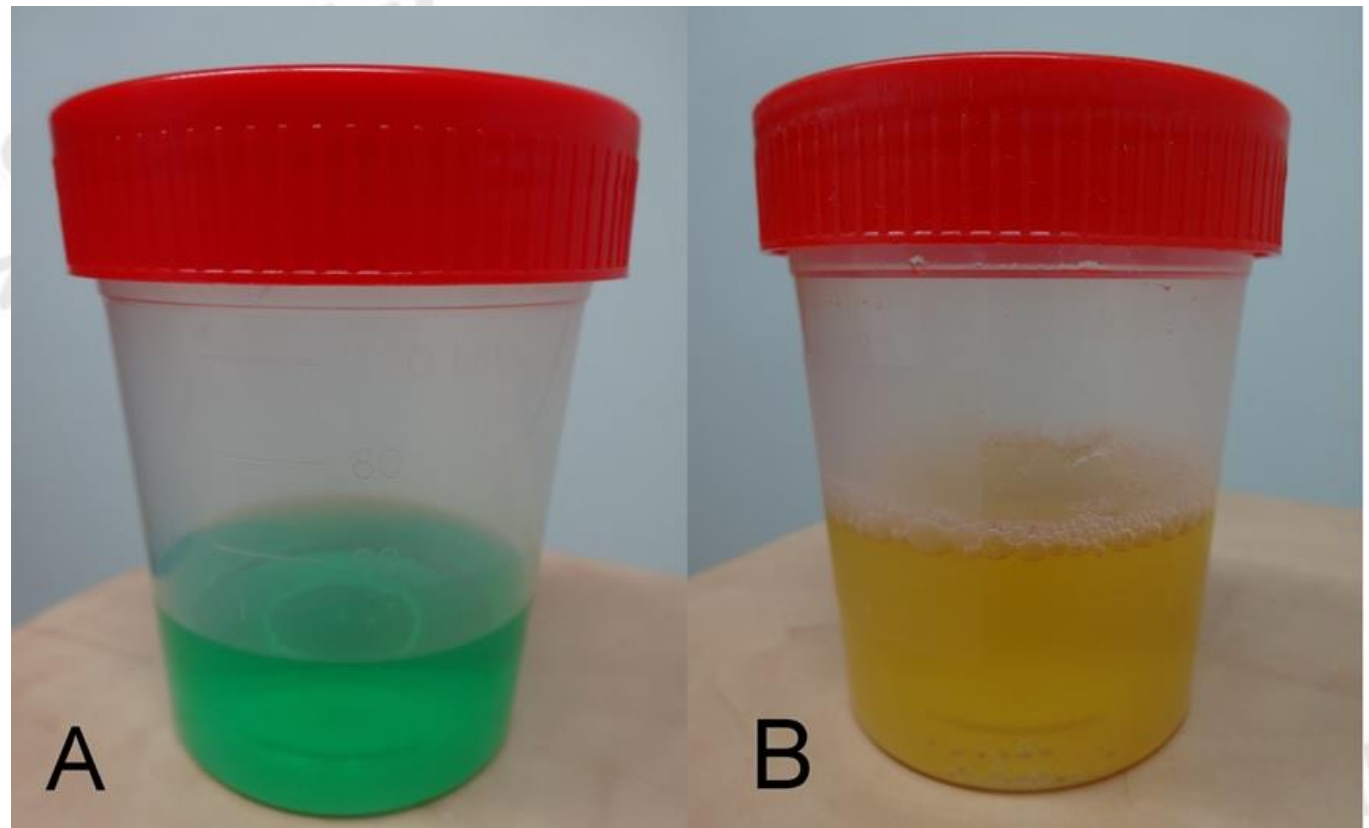
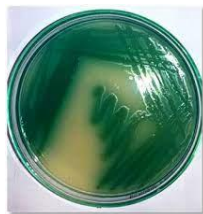
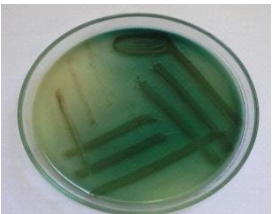
Post-metabolomics era
Tools and databases development

ormatiky
zita, PřF

8. BLUE-GREEN

Green urine arises when blue compounds are excreted by the kidneys and mix with normal yellow urochrome pigments. A number of agents have been known to cause green urine (for example, indomethacin, amitriptyline, triamterene, Listerine, flutamide, and injectable formulations containing phenol such as some preparations of promethazine).²⁶ Propofol is excreted in the urine mainly as the 1-glucuronide, 4-glucuronide and 4-sulfate conjugates of 2,6 di-isopropylphenol. These are phenol derivatives, which cause a pH dependent green urine discoloration that is more prominent in alkaline urine. There are abundant case reports describing green urine in patients with the propofol infusion syndrome, a potentially fatal disorder marked by severe metabolic acidosis, rhabdomyolysis, renal and cardiac failure particularly in association with prolonged high dose propofol infusions, in patients with brain injury.^{27 28} Case reports may also be identified that record green urine in the context of low doses of propofol, administered for sedation in intensive care units and even after single anaesthetic induction doses, unassociated with adverse complications.^{29 30}

Urinary tract infection caused by *Pseudomonas aeruginosa* can cause green urine through the liberation of the pigment pyocyanin.³ In chronic obstructive jaundice green and blue urine may arise attributable to biliverdin excretion, an oxidation product of bilirubin.²

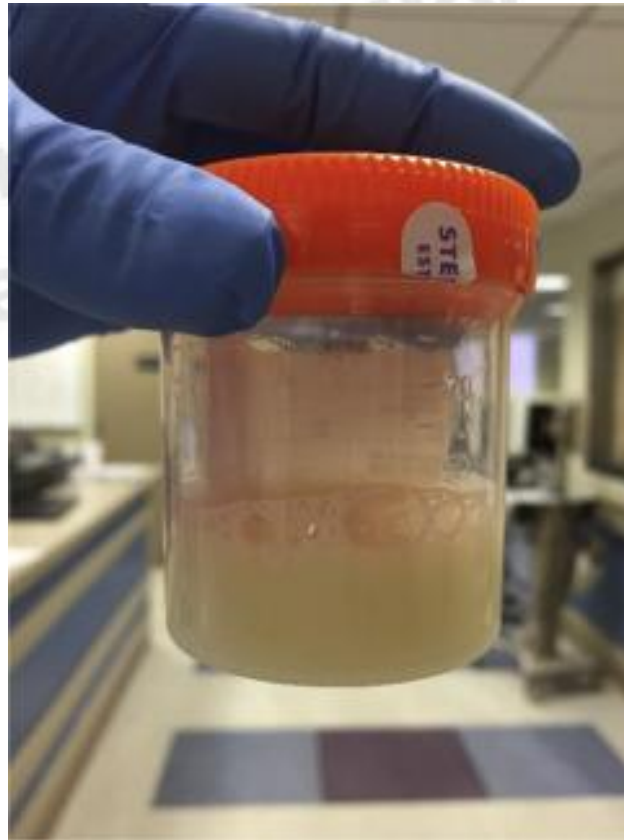


Zelená moč:

Sekrece **modrých** pigmentů (z léčiv, potravin, kosmetiky) do **žluté** moči. Příklad – předávkování preparátem obsahujícím **methylenou modř**.

C2131 Úvod
Masarykova

Small amounts of lipid normally appear in the urine, however, in nephrotic syndrome there are large amounts of high density lipoprotein as well as smaller amounts of other lipoproteins.⁸ Chyle may enter the urinary tract through a traumatic communication between lymphatic vessels and the urinary system.⁹



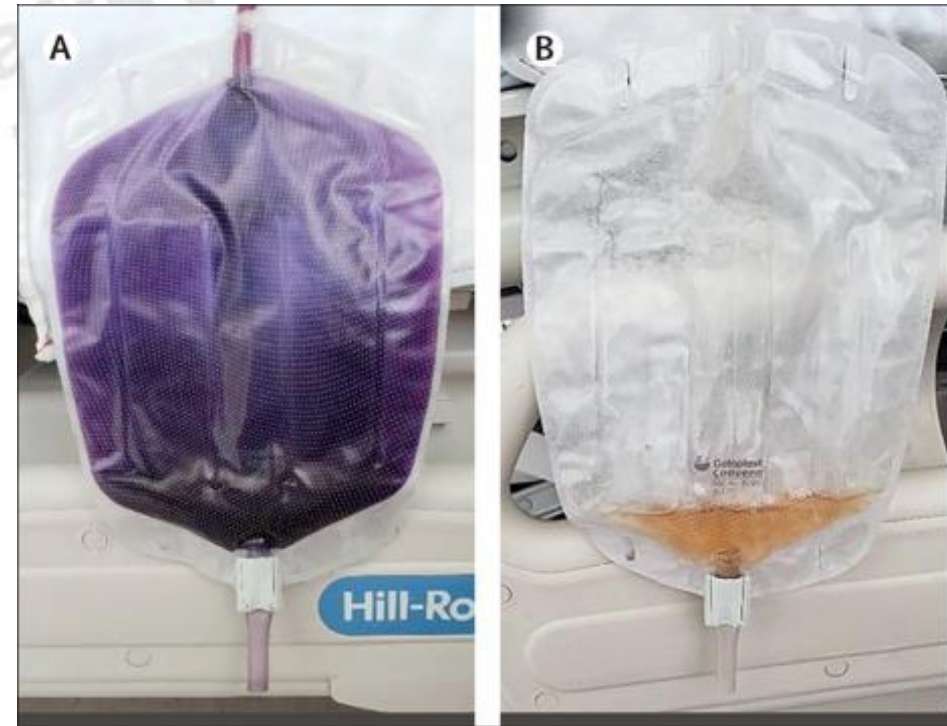
Wuchereria bancrofti



„Mléčná“ moč:

Lipidurie – přítomnost lipidů v moči, chylurie – přítomnost chylu (tuková emulze proudící lymfatickými cévami a obsahující živiny, které byly vstřebány v tenkém střevě) v moči.

Purple urine bag syndrome (PUBS) is rare and was first reported in 1978.[1] and is signified by purple discoloration of the urine usually seen in women and chronically debilitated patients with long term indwelling urinary catheters.[2,3] PUBS can be distressing for patients, family members and healthcare workers who are unaware of this association. This condition is often associated with urinary tract infection. Discolouration of the urine bag is due to the presence of indigo and indirubin pigments which precipitate and react with the synthetic materials of the catheter and urinary bag. We present a case report of this rare phenomenon.



(A) Urine bag showing the characteristic purple colour of urine seen in purple urine bag syndrome. (B) Normal urine after treatment with antibiotics, laxatives, and replacement of an indwelling catheter and urine bag.







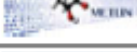






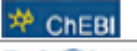



Fialová barva:

Způsobená bakteriální infekcí močových cest (*Proteus mirabilis*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Morganella morganii*, *Klebsiella pneumoniae*), za barvu můžou metabolity **tryptofanu**.

Metabolomika – identifikace metabolitů

- **Metabolomika:** Komplexní analýza metabolomu za konkrétního fyziologického nebo vývojového stádia organismu, tkáně či buňky.
- Více odráží aktuální stav buňky (ve srovnání s transkriptomikou nebo proteomikou).
- Analýza tisíců různých sloučenin s extrémně variabilními strukturami a vlastnostmi.
- Pro identifikaci metabolitů se převážně využívá hmotnostní spektrometrie a NMR.
- Využití databází a vývoj nástrojů pro automatické vyhodnocování experimentálních dat.











Table 5.2 List of metabolite identification databases

Databases			URL
NMR	SDBS		http://sdfs.db.aist.go.jp
	BioMagResBank		http://www.bmrb.wisc.edu/metabolomics/
	NMRShiftDB 2		http://nmrshiftdb.nmr.uni-koeln.de/
	MMCD		http://mmcd.nmrfam.wisc.edu/
MS	GMD		http://gmd.mpimp-golm.mpg.de/
	METLIN		https://metlin.scripps.edu/index.php
	MassBank		http://www.massbank.jp/
	NIST		http://webbook.nist.gov/chemistry/name-ser.html
	LIPIDMAPS		http://www.lipidmaps.org/
	KNAPSAcK		http://kanaya.naist.jp/KNAPSAcK/KNAPSAcK-v1200/KNAPSAcK.php
Compound DBs	ARM		http://metabolomics.jp/wiki/Main_Page
	ChEBI		https://www.ebi.ac.uk/chebi/init.do
	PubChem		https://pubchem.ncbi.nlm.nih.gov/
	ChemSpider		http://www.chemspider.com/
	Ligand Expo		http://ligand-expo.rcsb.org/
	3DMet		http://www.3dmet.dna.affrc.go.jp/
	MyCompoundID		http://www.mycompoundid.org/

Databáze metabolických drah

- **Metabolismus** (látková přeměna): Souhrn všech reakcí v buňce při nichž dochází k přeměně látek. Buňka tak získává energii, stavební a zásobní látky. Reakce jsou uspořádané do sérií – metabolických drah.
- Metabolity jsou součástí metabolických drah. Některé jsou zapojeny do velkého počtu drah, některé jen do několika.
- Metabolické dráhy jsou navzájem propojené.
- Při studiu metabolismu nezkoumáme metabolity odděleně ale jako součásti metabolických drah.
- **Databáze metabolických drah a nástroje pro jejich vizualizaci.**

Table 5.3 List of pathway databases and visualization tools

Databases/tools		URL
KaPPA-View		http://kpv2.kazusa.or.jp/kpv4
KEGG		http://www.genome.jp/kegg
HumanCyc		http://humancyc.org
HMDB		http://www.hmdb.ca/
iPath		http://pathways.embl.de
MapMan		http://mapman.gabipd.org/web/guest/mapman
MetaMapp		http://metamapp.fiehnlab.ucdavis.edu/homepage
MetPA		http://metpa.metabolomics.ca/MetPA/faces/Home.jsp
MetScape		http://metscape.ncibi.org
Paintomics		http://www.paintomics.org/cgi-bin/main2.cgi
Pathos		http://motif.gla.ac.uk/Pathos
PathVisio		http://www.pathvisio.org
ProMeTra		http://omictools.com/prometra-s11541.html
Reactome		http://www.reactome.org
SMPDB		http://smpdb.ca

Databáze metabolických drah











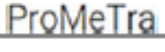


- Databáze metabolických drah a nástroje pro jejich vizualizaci.

- Databáze mohou být širokého zaměření, zkoumají metabolismus mnoha různých druhů, evolučně konzervované dráhy. Záznamy/diagramy jsou obecné a schematické.

- Databáze mohou být specializované, zaměřené na jeden organismus.

- Metabolomika a databáze metabolických drah spojují chemoinformatiku a bioinformatiku a stírají hranice mezi nimi. Spojují chemické vlastnosti sloučenin s jejich biologickým významem.

Table 5.3 List of pathway databases and visualization tools

Databases/tools		URL
KaPPA-View		http://kpv2.kazusa.or.jp/kpv4
KEGG		http://www.genome.jp/kegg
HumanCyc		http://humancyc.org
HMDB		http://www.hmdb.ca/
iPath		http://pathways.embl.de
MapMan		http://mapman.gabipd.org/web/guest/mapman
MetaMapp		http://metamapp.fiehnlab.ucdavis.edu/homepage
MetPA		http://metpa.metabolomics.ca/MetPA/faces/Home.jsp
MetScape		http://metscape.ncibi.org
Paintomics		http://www.paintomics.org/cgi-bin/main2.cgi
Pathos		http://motif.gla.ac.uk/Pathos
PathVisio		http://www.pathvisio.org
ProMeTra		http://omictools.com/prometra-s11541.html
Reactome		http://www.reactome.org
SMPDB		http://smpdb.ca



Databáze metabolických drah

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps molecular interaction, reaction and relation networks for:

1. Metabolism

Global/overview **Carbohydrate** Energy Lipid Nucleotide Cofactor/vitamin Terpenoid/PK Other secondary metabolite

2. Genetic Information Processing

3. Environmental Information Processing

4. Cellular Processes

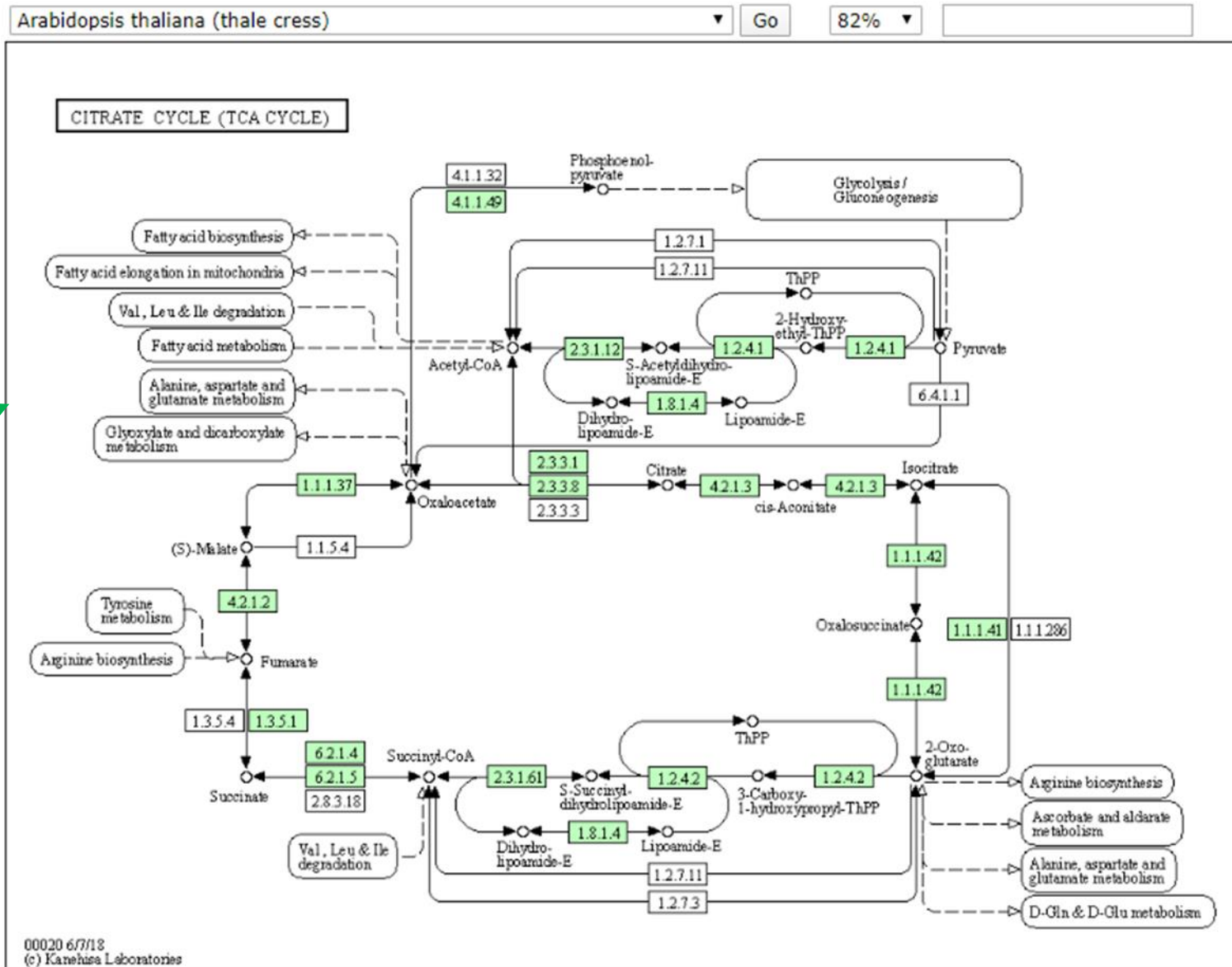
5. Organismal Systems

6. Human Diseases

7. Drug Development

1.1 Carbohydrate metabolism

- 00010 Glycolysis / Gluconeogenesis
- 00020 Citrate cycle (TCA cycle)
- 00030 Pentose phosphate pathway
- 00040 Pentose and glucuronate interconversions
- 00051 Fructose and mannose metabolism
- 00052 Galactose metabolism
- 00053 Ascorbate and aldarate metabolism
- 00500 Starch and sucrose metabolism
- 00620 Pyruvate metabolism
- 00630 Glyoxylate and dicarboxylate metabolism
- 00640 Propanoate metabolism
- 00650 Butanoate metabolism
- 00660 C5-Branched dibasic acid metabolism
- 00562 Inositol phosphate metabolism





- Arabidopsis thaliana (thale cress)
- Canis lupus (chicken)
- Xenopus laevis (African clawed frog)
- Danio rerio (zebrafish)
- Oryzias latipes (Japanese medaka)
- Drosophila melanogaster (fruit fly)
- Caenorhabditis elegans (nematode)
- Arabidopsis thaliana (thale cress)



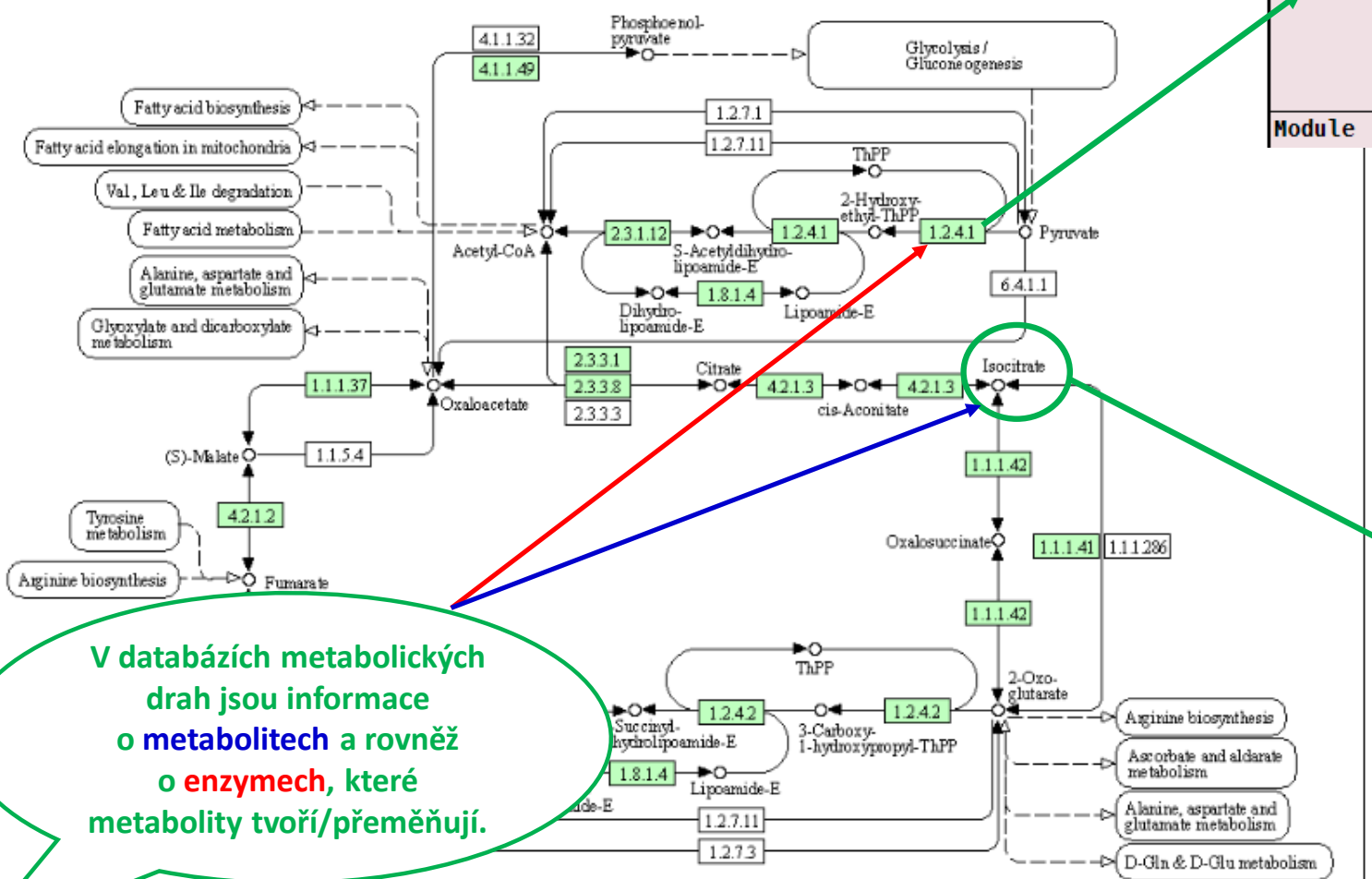
Arabidopsis thaliana (thale cress): AT1G01090

Help

Entry	AT1G01090	CDS	T00041
Gene name	PDH-E1_ALPHA		
Definition	(RefSeq) pyruvate dehydrogenase E1 alpha		
K0	K00161 pyruvate dehydrogenase E1 component alpha subunit [EC:1.2.4.1]		
Organism	ath Arabidopsis thaliana (thale cress)		
Pathway	ath00010 Glycolysis / Gluconeogenesis ath00020 Citrate cycle (TCA cycle) ath00620 Pyruvate metabolism ath01100 Metabolic pathways ath01110 Biosynthesis of secondary metabolites ath01200 Carbon metabolism		
Module	ath_M00307 Pyruvate oxidation, pyruvate => acetyl-CoA		

Arabidopsis thaliana (thale cress) Go 82%

CITRATE CYCLE (TCA CYCLE)



COMPOUND: C00311

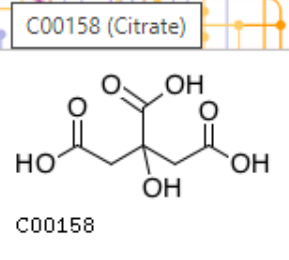
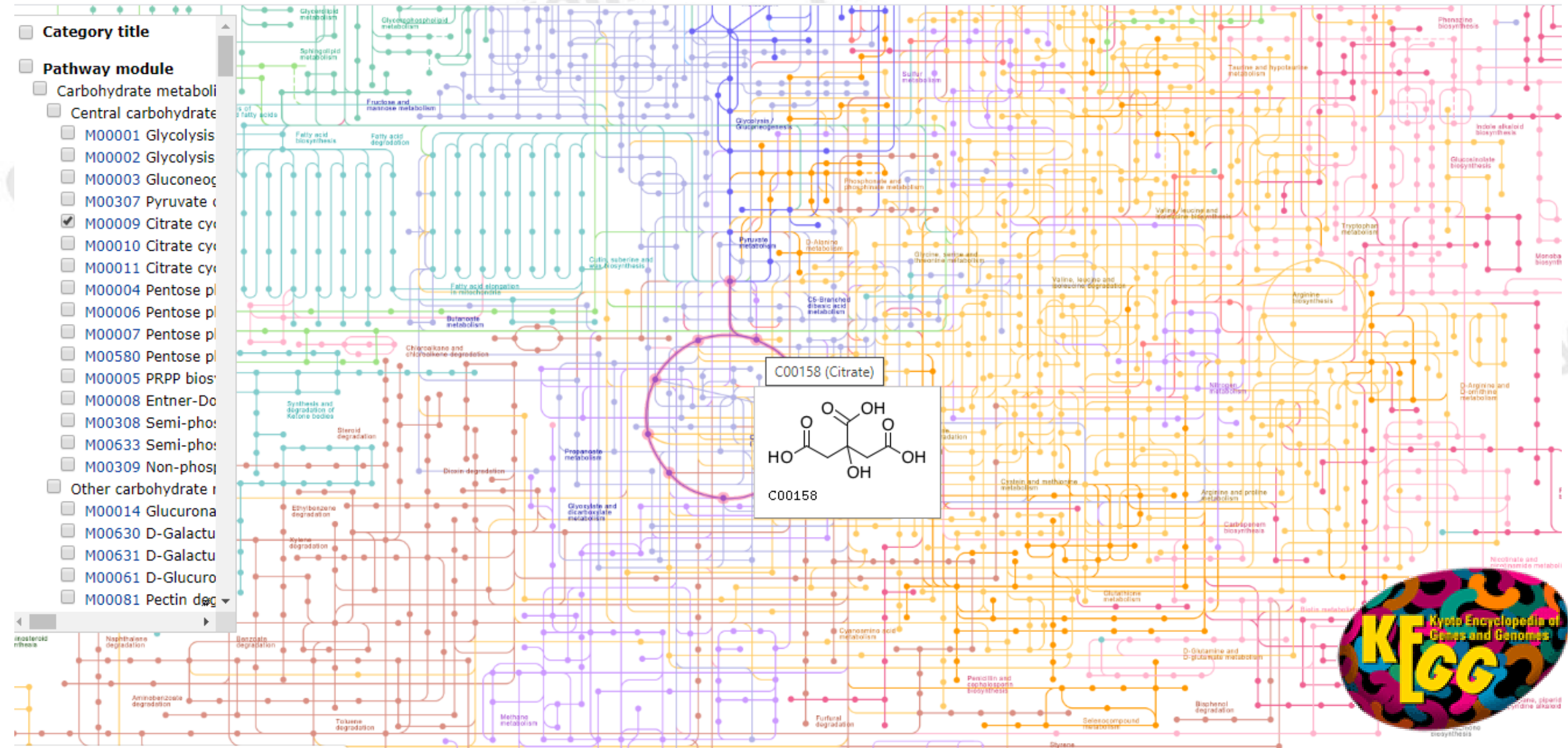
Help

Entry	C00311	Compound
Name	Isocitrate; Isocitric acid; 1-Hydroxytricarballic acid; 1-Hydroxypropane-1,2,3-tricarboxylic acid	
Formula	C6H8O7	
Exact mass	192.027	
Mol weight	192.1235	
Structure	<p>C00311</p> <p>Mol file KCF file DB search</p>	
Reaction	R00267 R00479 R00709 R01324 R01899 R01900 R01946	
Pathway	map00020 Citrate cycle (TCA cycle) map00630 Glyoxylate and dicarboxylate metabolism map00720 Carbon fixation pathways in prokaryotes map01060 Biosynthesis of plant secondary metabolites map01061 Biosynthesis of phenylpropanoids map01062 Biosynthesis of terpenoids and steroids map01063 Biosynthesis of alkaloids derived from shikimate pathway	

V databázích metabolických drah jsou informace o metabolitech a rovněž o **enzymech**, které metabolity tvoří/přeměňují.

Takhle vypadá (skoro) celý
metabolismus ☺

Databáze metabolických drah





Databáze metabolických drah

<https://smpdb.ca/>

Welcome
to the

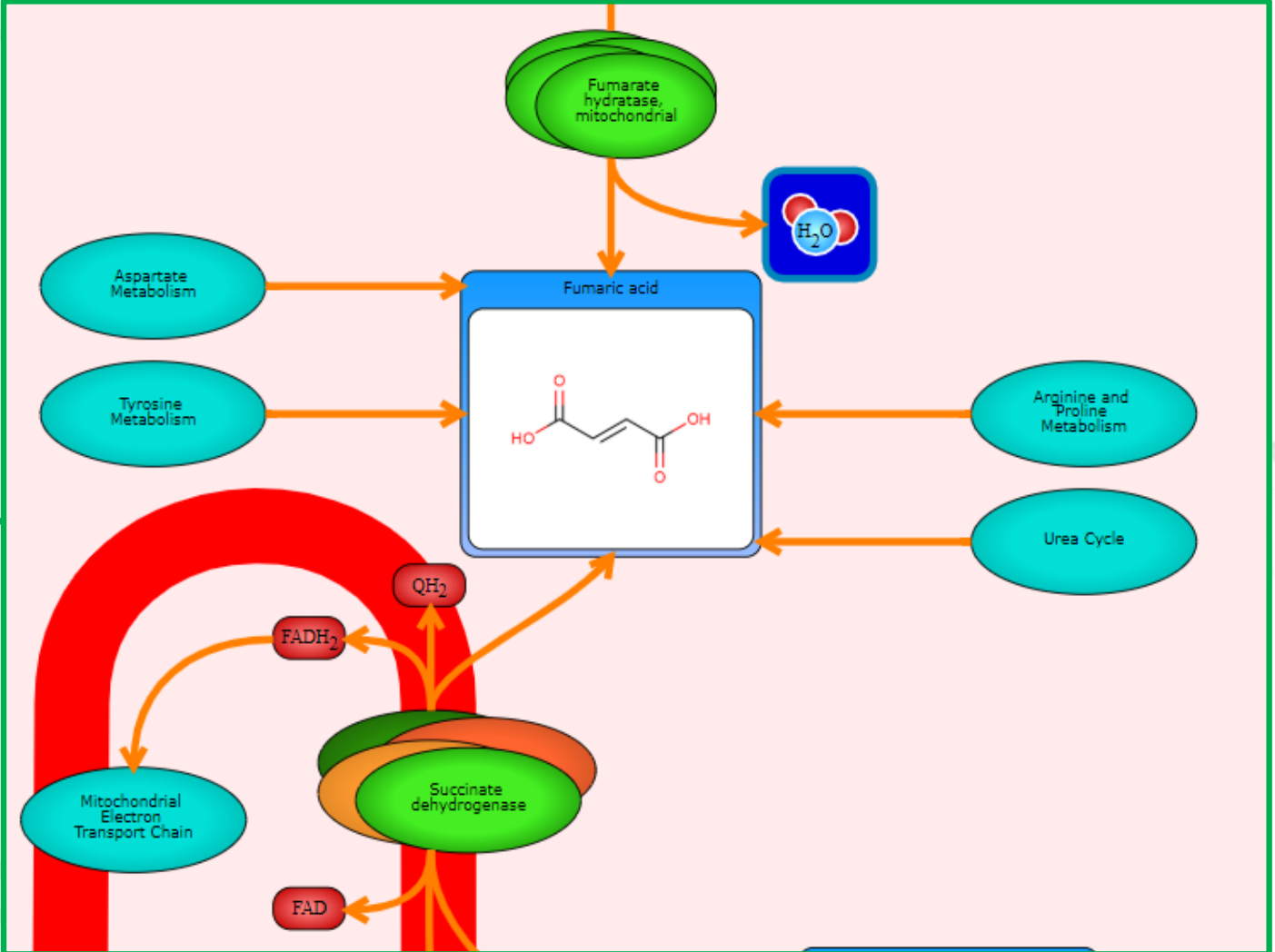
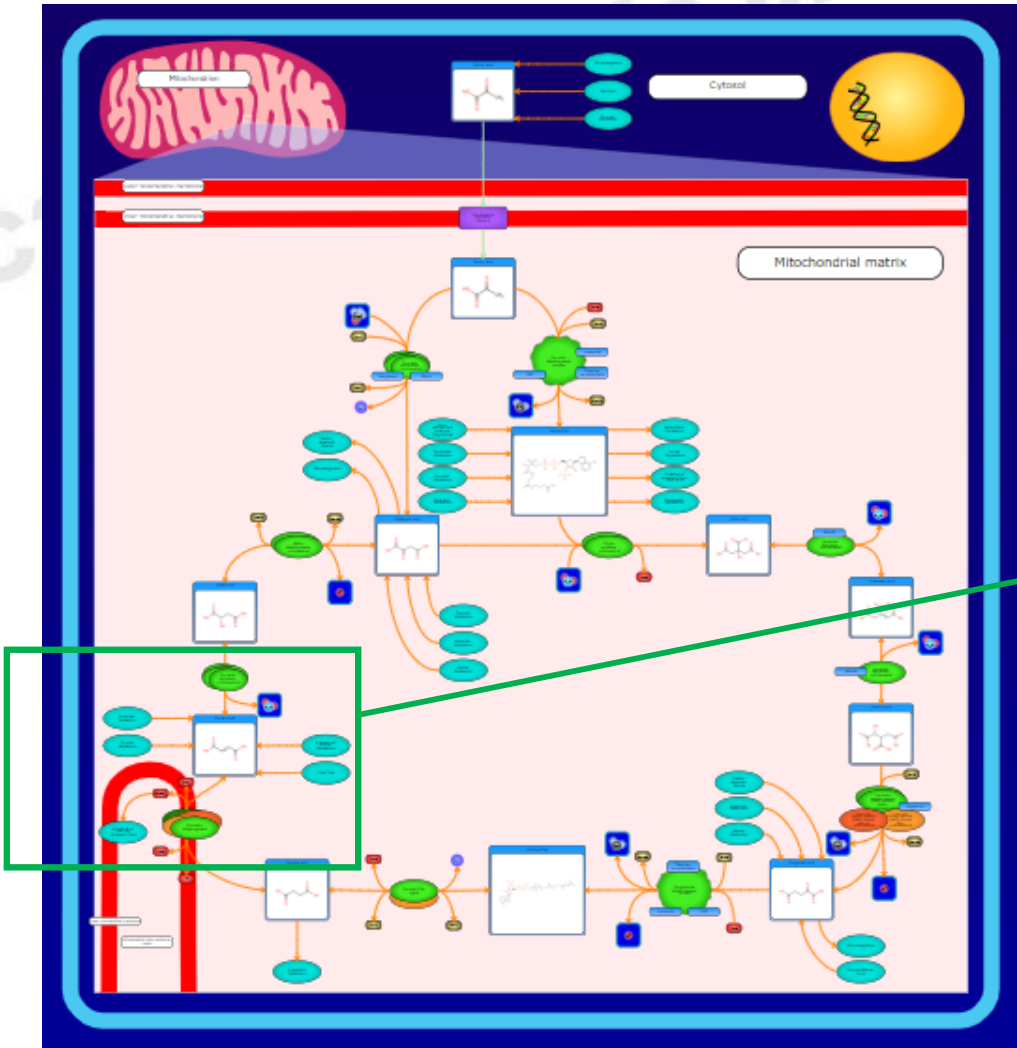
Small Molecule Pathway Database

Version
2.0



Brought to you by the creators of the [Human Metabolome Database \(HMDB\)](#) and [DrugBank](#):

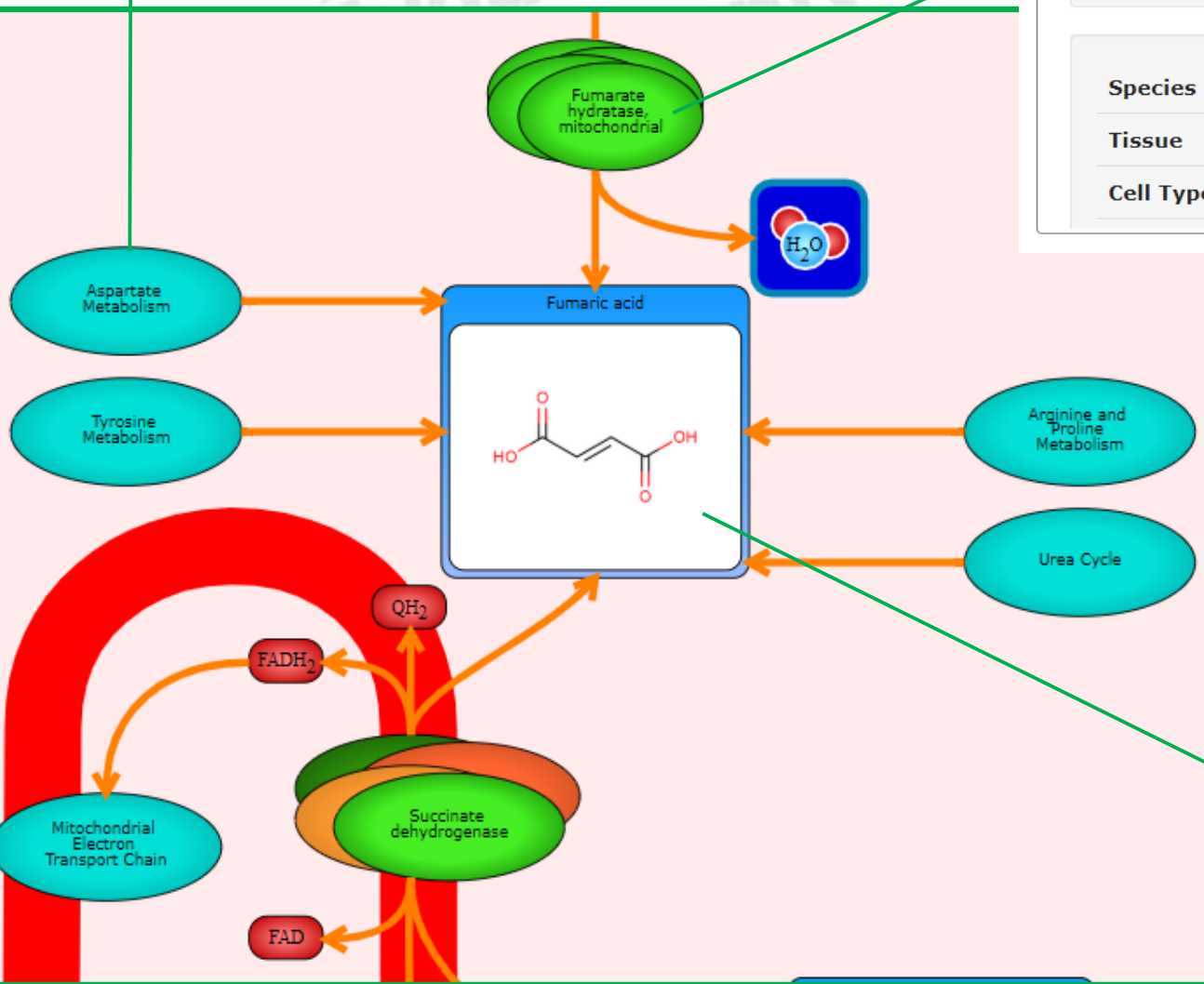
SMPDB (The Small Molecule Pathway Database) is an interactive, visual database containing more than 30 000 small molecule pathways found in humans only. The majority of these pathways are not found in any other pathway database. SMPDB is designed specifically to support pathway elucidation and pathway discovery in metabolomics, transcriptomics, proteomics and systems biology. It is able to do so, in part, by providing exquisitely detailed, fully searchable, hyperlinked diagrams of human metabolic pathways, metabolic disease pathways, metabolite signaling pathways and drug-action pathways. All SMPDB pathways include information on the relevant organs, subcellular compartments, protein_complex cofactors, protein_complex locations, metabolite locations, chemical structures and protein_complex quaternary structures. Each small molecule is hyperlinked to detailed descriptions contained in the [HMDB](#) or [DrugBank](#) and each protein_complex or enzyme complex is hyperlinked to [UniProt](#). All SMPDB pathways are accompanied with detailed descriptions and references, providing an overview of the pathway, condition or processes depicted in each diagram. The database is easily browsed and supports full text, sequence and chemical structure searching. Users may query SMPDB with lists of metabolite names, drug names, genes/protein_complex names, SwissProt IDs, GenBank IDs, Affymetrix IDs or Agilent microarray IDs. These queries will produce lists of matching pathways and highlight the matching molecules on each of the pathway diagrams. Gene, metabolite and protein_complex concentration data can also be visualized through SMPDB's mapping interface. All of SMPDB's images, image maps, descriptions and tables are [downloadable](#).



Aspartate Metabolism

Aspartate is synthesized by transamination of oxaloacetate by aspartate aminotransferase or amino acid oxidase. Aspartyl-tRNA synthetase can then couple aspartate to aspartyl tRNA for protein synthesis. The aspartate content in human proteins is about 7%. Asparagine synthase can convert aspartate to the polar amino acid asparagine. Aspartate is also a precursor for cellular signaling compounds such as, N-acetyl-aspartate, beta-alanine, adenylysuccinate, arginino-succinate and N-carbamoylaspartate. Aspartate is also a metabolite in the urea cycle and involved in gluconeogenesis. Additionally, aspartate carries the reducing equivalents in the mitochondrial malate-aspartate shuttle, which utilizes the ready interconversion of aspartate and oxaloacetate. The conjugate base of L-aspartic acid, aspartate, also acts as an excitatory neurotransmitter in the brain which activates NMDA receptors.

[View Pathway \(SMP000067\)](#)



Fumarate hydratase, mitochondrial

Components	Description	Reactions
Fumarate hydratase, mitochondrial		
Details in UniProt (P07954)		
Species	Homo sapiens	
Tissue		
Cell Type	Cell	

<https://smpdb.ca/>

Fumaric acid

Structure	Description	Reactions
Details in HMDB (HMDB0000134)		
Details in DrugBank (DB04299)		

Bioinformatika + chemoinformatika

- Metabolomika a databáze metabolických drah spojují chemoinformatiku a bioinformatiku a stírají hranice mezi nimi. Spojují chemické vlastnosti sloučenin s jejich biologickým významem.
- Bioinformatika a chemoinformatika se postupně sblíží a prolínají, někteří autoři používají výraz „chemická bioinformatika“ pro průnik těchto oborů (spojení studia chemických vlastností s molekulárně-biologickými rolemi). Souvisí rovněž s rozšiřujícími se znalostmi o významu malých molekul.
- Vzpomeňte na sacharidy a lipidy: jsou to nízkomolekulární látky a metabolity, ale zároveň jsou součástí makromolekul (glykoproteiny, lipoproteiny), mají významné strukturní funkce (polysacharidy, buněčné membrány), mají informační role (antigeny, interakce s patogeny).

OPEN ACCESS Freely available online

 PLOS | COMPUTATIONAL BIOLOGY

Education

Chapter 3: Small Molecules and Disease

David S. Wishart^{1,2,3*}

1 Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada, 2 Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, 3 National Research Council, National Institute for Nanotechnology (NINT), Edmonton, Alberta, Canada

Malé molekuly - léčiva

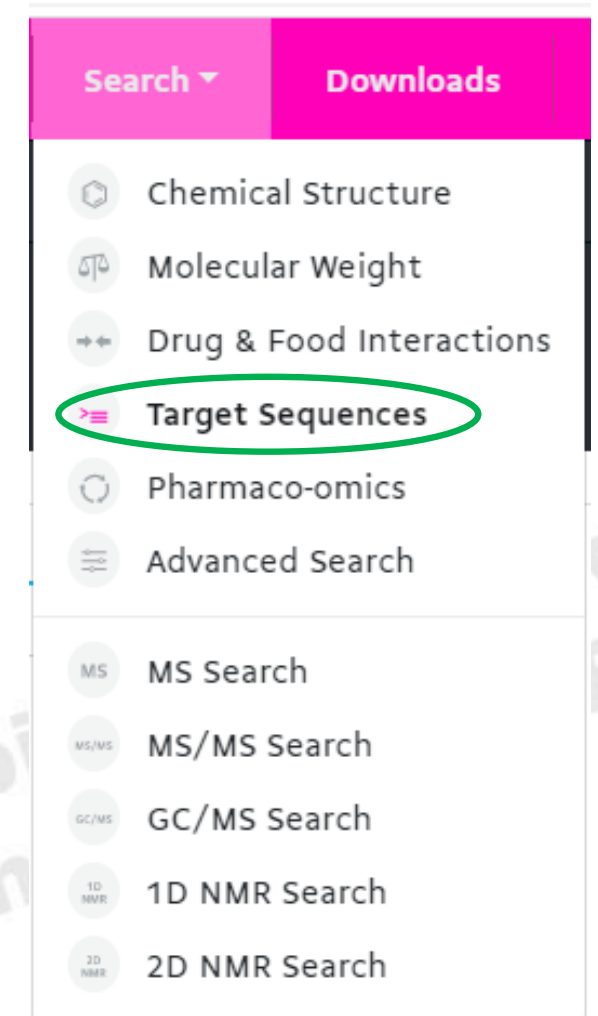


The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information.

LATEST RELEASE

The latest release of DrugBank Online (version 5.1.10, released 2023-01-04) contains 15,585 drug entries including 2,742 approved small molecule drugs, 1,579 approved biologics (proteins, peptides, vaccines, and allergenics), 134 nutraceuticals and over 6,720 experimental (discovery-phase) drugs. Additionally, 5,296 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each entry contains more than 200 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

<https://www.drugbank.ca/>



Name	Cetirizine
Accession Number	DB00341 (APRD00630)
Type	Small Molecule
Groups	Approved

Synonyms

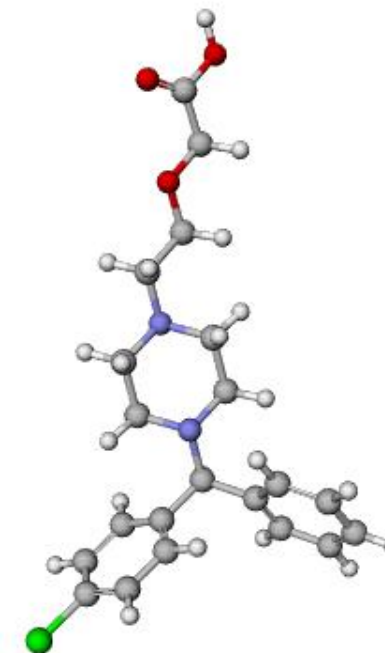
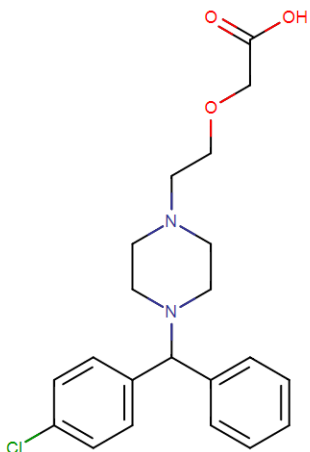
Cetirizin	
Cetirizina	
Cetirizine	
Cetirizinum	

Description

Cetirizine, also commonly known as **Zyrtec**, is an orally active second-generation histamine H1 antagonist proven effective in the treatment of various allergic symptoms, such as sneezing, coughing, nasal congestion, hives, and other symptoms ^{1, 2}.

One of the most common uses for this drug is for a condition called *allergic rhinitis*. The prevalence of allergic rhinitis in the United States is about 15% according to physician diagnoses, and up to 30%, according to self-reported nasal symptoms. Allergic rhinitis is associated with multiple missed or unproductive days at work and school, problems with sleep, and other difficulties with day to day activities for many individuals ³. Furthermore, some antihistamine agents that are used to treat this condition cause undesirable, sedating effects ⁴.

Cetirizine is one of the first second-generation H1 antihistamines (SGAHs) formulated to selectively inhibit the H1 receptor without sedating effects ².



Name	Cetirizine
Accession Number	DB00341 (APRD00630)
Type	Small Molecule
Groups	Approved

Carcinogenesis and mutagenesis: In a 2-year carcinogenicity study in rats, cetirizine was not shown to be carcinogenic at dietary doses up to 20 mg/kg (approximately 15 times the maximum recommended daily oral dose in adults). In a 2-year carcinogenicity study in mice, cetirizine administration lead to an incidence of benign liver tumors in males at a dietary dose of 16 mg/kg (approximately 6 times the maximum recommended daily oral dose in adults). The clinical significance of these findings during long-term use of cetirizine is unknown at this time ^{Label}.

Cetirizine was not mutagenic in the Ames test, and not clastogenic in the human lymphocyte assay, the mouse lymphoma assay, and in vivo micronucleus test in rats ^{Label}.

Kind	Protein	General Function	Histamine receptor activity
Organism	Humans	Specific Function	In peripheral tissues, the H1 subclass of histamine receptors mediates the contraction of smooth muscles, increase in capillary permeability due to contraction of terminal venules, and catecholamin...
Pharmacological action	Yes		
Actions	Antagonist		
		Gene Name	HRH1
		Uniprot ID	P35367
		Uniprot Name	Histamine H1 receptor
		Molecular Weight	55783.61 Da

Pregnancy Category B:

In mice, rats, and rabbits, cetirizine was not teratogenic at oral doses up to 96, 225, and 135 mg/kg, respectively (approximately 40, 180 and 220 times the maximum recommended daily oral dose in adults). There are no adequate and well-controlled studies in pregnant women. Because animal studies are not always predictive of human response, cetirizine should be used in pregnancy only if clearly needed ^{Label}.

Use in breastfeeding/nursing

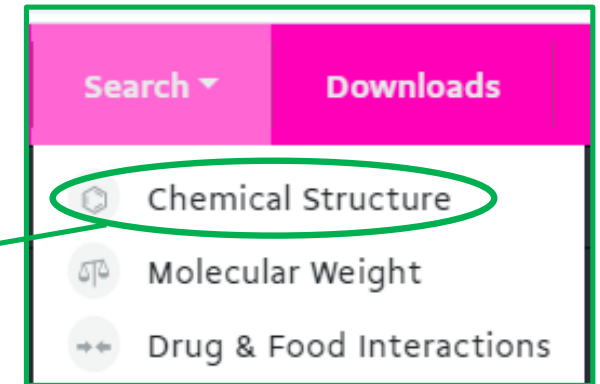
Cetirizine has been reported to be excreted in human breast milk. The use of cetirizine in nursing mothers is not recommended ^{Label}.

Malé molekuly - léčiva

<https://www.drugbank.ca/>

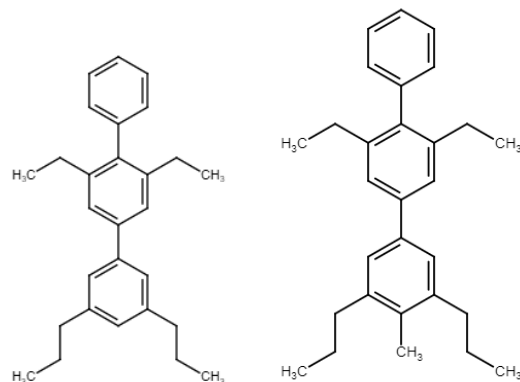


The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information.



STRUCTURE SEARCH MOLECULAR WEIGHT

Search by structure



DB07412 Score: 0.516		1-biphenyl-2-ylmethanamine experimental
DB12794 Score: 0.5		p-Quaterphenyl 135-70-6 investigational

Malé proteiny

- **Protein vs. peptid** – často (historicky) rozdělovány podle **velikosti**. Jako proteiny jsou označovány polymery s minimální délkou 50-100 aminokyselin, kratší bývají označovány jako peptidy.
- **Protein vs. peptid** – dělení podle **struktury a původu**. Peptidy vznikají proteolýzou většího prekursoru nebo jsou syntetizovány uměle. Jako peptidy se také někdy označují přirozeně neuspořádané řetězce aminokyselin, které nemají stabilní prostorové uspořádání a mohou flexibilně měnit konformaci.
- **Malé proteiny** – proteiny, které vznikají přímo **translací krátkého ORF** (kratší než 100 kodonů). Při predikci genů mohou unikát, cut off používaný při predikci vycházející z průměrné délky ORF je často větší (při snížení bychom museli kontrolovat mnohonásobně více ORF).

Malé proteiny

- **Malé proteiny** – proteiny, které vznikají přímo **translací krátkého ORF** (kratší než 100 kodonů). Při predikci genů mohou unikát, cut off používaný při predikci vycházející z průměrné délky ORF je často větší (při snížení bychom navíc museli kontrolovat mnohonásobně více ORF).
- **Malé proteiny** – unikají i při biochemických analýzách, metody „nepočítají“ s proteiny menšími než cca 5 kDa. Často objeveny jen shodou okolností.

Příklady malých proteinů:

- protein SpoVM (**26 aminokyselin!**), podílí se na sporulaci grampozitivních bakterií. Delece genu pro SpoVM způsobuje těžké defekty sporulace, díky tomu byl objeven. Předpokládá se, že při sporulaci plní minimálně čtyři různé funkce (malý, ale šikovný).
- Protein SidA (**29 aminokyselin**), ovlivňuje buněčné dělení u bakterií (*Caulobacter crescentus*).
- Protein MgtR (**30 aminokyselin**), regulace membránově vázaných enzymů (*Salmonella*).

Malé proteiny

Predikce malých proteinů (krátkých ORF):

- krátkých ORF se náhodně vyskytuje v genomu **mnoho**, zároveň se zmenšující se délkou klesá přesnost predikce založené na statistických parametrech.
- ORF jsou prostě příliš krátké, aby např. obsahovaly dost hexamerů (pro HMM pátého řádu), problém vyhodnotit i frekvenci výskytu tripletů – **využití kodonů**.
- **Problematické** odlišení „pravých“ krátkých ORF od mnoha náhodných krátkých ORF.
- Mnoho zatím charakterizovaných malých proteinů je **hydrofobní** – cílené vyhledávání v membránových frakcích?

Small Proteins Can No Longer Be Ignored

Gisela Storz,¹ Yuri I. Wolf,²
and Kumaran S. Ramamurthi³

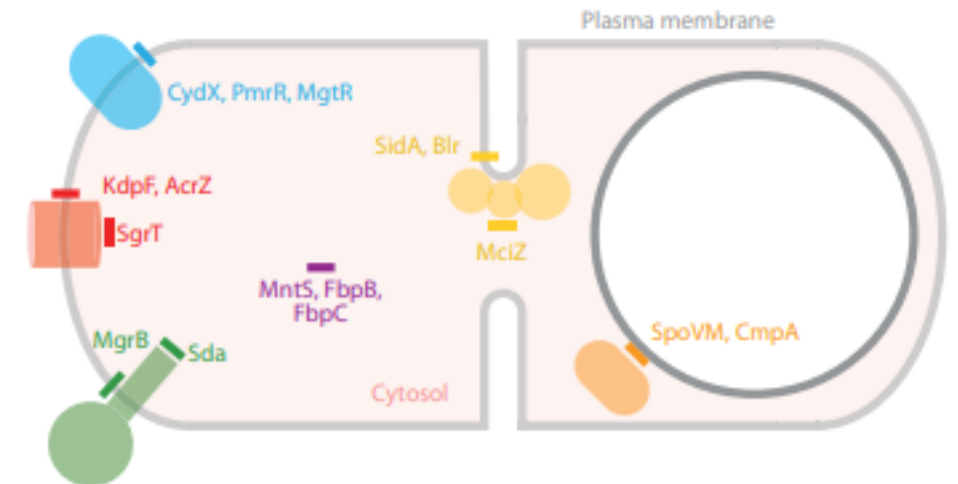


Figure 1

Sites of small protein action. The cytosol of a composite gram-positive or gram-negative bacterium bounded by the plasma membrane (*light gray*) is shown. The outer forespore membrane during sporulation is depicted in dark gray. Proteins associated with various cell functions are colored as follows: kinases, green; transporters, red; membrane-bound enzymes, blue; cell division septum, yellow; forespore during sporulation, orange; soluble chaperones, purple. Small proteins are depicted as rectangles. Transmembrane small proteins are depicted as rectangles that traverse the plasma membrane, amphipathic helical small proteins that are peripherally membrane associated are drawn as rectangles that are parallel to the plane of the membrane, and soluble small proteins are shown in the cytosol (*pink*).

Malé proteiny

CNCB-NGDC

SmProt

a reliable repository with comprehensive annotation of small proteins derived from ribosome profiling

Welcome to SmProt v2.0!

<http://bigdata.ibp.ac.cn/SmProt/>

Small proteins are the general term for proteins with length shorter than 100 amino acids. **SmProt** contains records of **Small Proteins** encoded by genes, especially for ones from UTRs and non-coding RNAs. The selected small proteins were identified from ribosome profiling data, literature, mass spectrometry (MS), etc., carried out in eight species including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Danio rerio*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Escherichia coli*. Moreover, **SmProt** contains features for the collected small proteins on their sequences, genomic locations, tissues/cell lines, assessment reflecting coding potential, function, variants, and related diseases that have been verified or predicted, etc.

Proteiny kódované nekódující RNA je zvláštní výraz. 😊 Kódující status je u některých ncRNA nyní diskutabilní...krátké ORF byly prostě přehlédnuty nebo ignorovány.

Peptides/Proteins Encoded by Non-coding RNA: A Novel Resource Bank for Drug Targets and Biomarkers

Song Zhu^{1,2†}, Jizhong Wang^{1,2†}, Yutian He^{1,2†}, Nan Meng^{1,2†} and Guang-Rong Yan^{1,2*}

Malé proteiny

SORFS.ORG

A REPOSITORY OF SMALL ORFS IDENTIFIED BY RIBOSOME PROFILING

<http://www.sorfs.org/>

Welcome

Introduction

Small open reading frames (sORFs) can be defined as open reading frames smaller than or equal to 300 nucleotides (100 amino acids). These “sORFs”, while inherent to all genomes, are historically ignored in gene annotation studies, stating that these lack any coding potential. Exclusion of these sORFs has emerged as a side effect during the development of different (gene prediction) tools in the field of bioinformatics/genomics/proteomics trying to reduce noise, imposed by technological limitations. However, recent scientific breakthroughs discovered coding potential of several sORFs with clinical significance, indicating their importance. ^{1, 2, 4}. In particular, the advent of ribosome profiling ⁵ (RIBO-seq), a next generation deep sequencing technique, providing a genome-wide snapshot of the translating machinery in a cell, provided evidence of translation in sORFs. The value and importance of sORFs is becoming widely recognized ^{6, 7} furthermore ribosome profiling data is becoming more abundant. The creation of a public repository for sORFs, providing information resulting from various tools and metrics, seems a necessity in aiding functional research in the micropeptide field.

sORFs.org: a repository of small ORFs identified by ribosome profiling

Volodimir Olexiouk^{1,*}, Jeroen Crappé¹, Steven Verbruggen¹, Kenneth Verhegen^{2,3}, Lennart Martens^{2,3} and Gerben Menschaert^{1,*}

An update on sORFs.org: a repository of small ORFs identified by ribosome profiling

Volodimir Olexiouk^{*}, Wim Van Criekeing and Gerben Menschaert^{*}

Malé molekuly - shrnutí

- Malé molekuly – heterogenní skupina látek s nejrůznějšími funkcemi
- SMILES, InChI
- Mnoho informatických zdrojů a nástrojů je komerčních – farmaceutický průmysl
- Souvisejí funkčně i strukturně s makromolekulami
- Metabolomika – mladé vědní odvětví zabývající se komplexním studiem metabolitů
- Databáze celých metabolických drah
- Malé proteiny/mikropeptidy – nová oblast studia proteinů
- Chemoinformatika/bioinformatika/chemická bioinformatika – původně oddělené obory s neprovázanými informacemi se začínají prolínat

Použitá a doporučená literatura

Introduction to Cheminformatics

David S. Wishart¹

¹Departments of Computing Science and Biological Sciences, University of Alberta,
Edmonton, Alberta, Canada

Chem. Listy 111, 710–715 (2017)

LINEÁRNÍ REPREZENTACE CHEMICKÝCH STRUKTUR

JIŘÍ JIRÁT^{a,b} a DANIEL SVOZIL^{a,b}

Validation of ligands in macromolecular structures determined by X-ray crystallography

Oliver S. Smart,^a Vladimír Horský,^{b,c} Swanand Gore,^{a,‡} Radka Svobodová Vařeková,^{b,c} Veronika Bendová,^{b,c,d} Gerard J. Kleywegt^a and Sameer Velankar^{a*}

Detect, Correct, Retract: How to manage incorrect structural models

Alexander Wlodawer^a, Zbigniew Dauter^b, Przemyslaw J. Porebski^c, Wladek Minor^c, Robyn Stanfield^d, Mariusz Jaskolski^{e,f}, Edwin Pozharski^g, Christian X. Weichenberger^h, and Bernhard Rupp^{h,i,*}

Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs

Miniperspective

Michael D. Shultz^{*†}

ValidatorDB: database of up-to-date validation results for ligands and non-standard residues from the Protein Data Bank

David Sehnal^{1,2,3,†}, Radka Svobodová Vařeková^{1,2,†}, Lukáš Pravda^{1,2}, Crina-Maria Ionescu¹, Stanislav Geidl^{1,2}, Vladimír Horský³, Deepti Jaiswal¹, Michaela Wimmerová^{1,2} and Jaroslav Koča^{1,2,*}

Použitá a doporučená literatura

OPEN ACCESS Freely available online



Education

Chapter 3: Small Molecules and Disease

David S. Wishart^{1,2,3*}

¹ Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada, ² Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, ³ National Research Council, National Institute for Nanotechnology (NINT), Edmonton, Alberta, Canada

Chem. Listy 105, 745–751 (2011)

METABOLOMIKA – ZÁKLADNÍ POJMY, STRATEGIE A METODOLOGIE

Chapter 5 Informatics for Metabolomics

Kanthida Kusonmano, Wanwipa Vongsangnak,
and Pramote Chumnanpuen

© Springer Science+Business Media Singapore 2016
B. Shen et al. (eds.), *Translational Biomedical Informatics*, Advances in
Experimental Medicine and Biology 939, DOI 10.1007/978-981-10-1503-8_5

Small Proteins Can No Longer Be Ignored

Gisela Storz,¹ Yuri I. Wolf,²
and Kumaran S. Ramamurthi³

Peptides/Proteins Encoded by Non-coding RNA: A Novel Resource Bank for Drug Targets and Biomarkers

Song Zhu^{1,2†}, Jizhong Wang^{1,2†}, Yutian He^{1,2†}, Nan Meng^{1,2†} and Guang-Rong Yan^{1,2*}

SOFTWARE

MiPepid: MicroPeptide identification tool using machine learning

Mengmeng Zhu^{1,2} and Michael Gribskov^{2*}