Od kofeinu k návrhu léčiv (chemoinformatika)

Kofein a jemu podobní

Studijní materiál k první úloze z ViBuChu 2013/2014

Stanislav Geidl

Pokud studijní text nepomůže, nebojte se napsat **do diskuzního fóra** nebo mi přímo napište **email** (standag@chemi.muni.cz), doporučuji také do kopie uvést mého kolegu Tomáše Bouchala (tbouchal@mail.muni.cz), který všechny úkoly prošel a bude schopný vám také pomoct nebo mě alespoň zaurgovat ;)

Rád bych na tomto místě také poděkoval Mirkovi Brumovskému za jeho skvělou recenzi zadání a Tomášovi za kontrolu a vyzkoušení všech úkolů a pomoc při tvorbě autorského řešení.

Obsah

- základní pojmy
- fytochemická a etnobotanická databáze
- počítačové formáty molekul – SMILES a mol
- práce s databázemi
- podobnostní hledání v PubChemu
- generování struktury pomocí Coriny
- instalace OpenBabelu
- práce s OpenBabelem
- přikládání, zarovnání (alignment)
- přikládání s utilitou obfit

Základní pojmy I

- **logP** – logaritmicky vyjádřený rozdělovací koeficient užívaný k odhadu lipofility látky neboli rozdělovací tendence mezi vodou a organickým rozpouštědlem (lipidy); http://is.muni.cz/th/21631/prif_d/Appendix_III_vysledky.pdf

- **pKa** – logaritmicky vyjádřená rovnovážná kostanta disociace kyseliny (disociační konstanta kyseliny); učebnice chemie nebo http://cs.wikipedia.org/wiki/Disocia%C4%8Dn%C3%AD_konstanta

- **donor** – dárce, **akceptor** – příjemce, **vodíková vazba** (můstek) – učebnice chemie nebo http://chemvazba.moxo.cz/Lekce/lekce9.html

- **donor vodíkové vazby** – prvek, který může použít svůj vodík do vodíkové vazby (např. R–**N**H_2)

- **akceptor vodíkové vazby** – prvek, který dokáže akceptovat vodíkovou vazbu (např. I**O**=R)

  - R–**N**H_2···I**O**=R

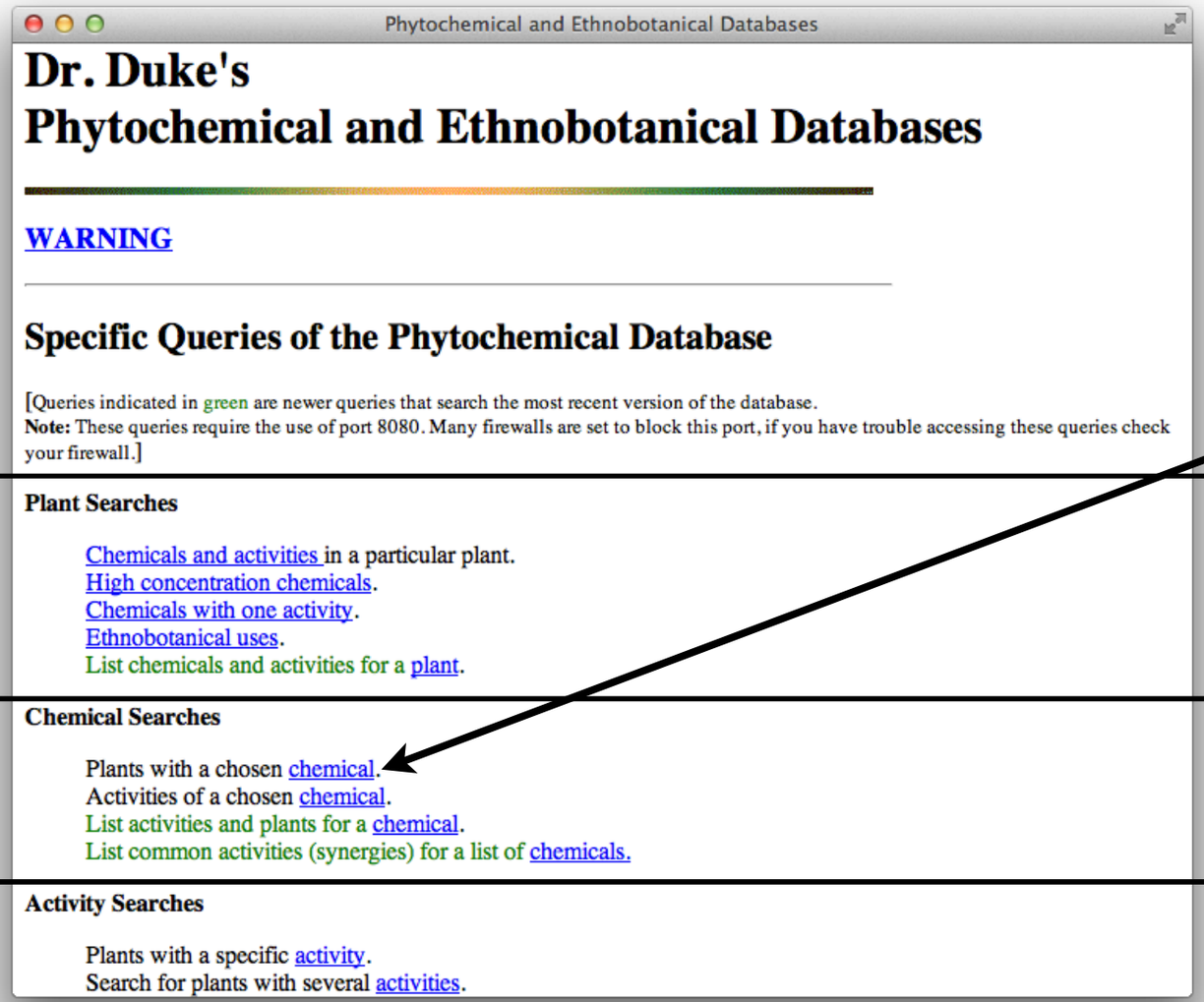
Základní pojmy II

- **Lipinského pravidlo pěti** – empirické pravidlo, které rozhoduje, jestli může být látka léčivem či nikoliv
  http://akademon.cz/clanekDetail.asp?name=Lipinskeho%20pravidlo%20peti&source=0312
  http://en.wikipedia.org/wiki/Lipinski's_rule_of_five
- **RMSD** – viz přikládání (zarovnání, alignment)
- **SMILES** – viz počítačové formáty molekul – SMILES a mol

⇧Obsah

Fytochemická a etnobotanická databáze

- http://www.ars-grin.gov/duke/index.html
- Databáze doktora Jima Duka obsahuje informace o přírodních látkách, jejich obsahu v rostlinách a jejich aktivitách.



Nápověda:
Pokud budete hledat rostliny s obsahem nějaké látky, použijte tento odkaz

Plant Searches	**Hledání rostlin** Přírodní látky a aktivity v konkretních roslinách. Vysoké koncentrace přírodních látek. Přírodní látky s jednou aktivitou. ...
Chemical Searches	**Hledání přírodních látek** Rostliny s vybranou přírodní látkou. Aktivity vybrané přírodní látky. ...
Activity Searches	**Hledání aktivit** Rostliny se specifickou aktivitou. ...

⇧Obsah

Fytochemická a etnobotanická databáze

High-Low Chemical query

Dr. Duke's
Phytochemical and Ethnobotanical Databases

High-Low Chemical query

This query displays either a **HI-LOW** or just a list of plants containing a particular chemical.

Enter the chemical name in the box below (e.g. tyrosine) or select the Chemical link to choose from the list of chemicals in the database.

Chemical: nicotine (required)
View only the top 40 plants with quantitative data.
Check to include plants *without* quantitative data.

Submit Clear Form

WARNING

Send comments or suggestions:

Jim Duke
Green Farmacy Garden
8210 Murphy Road
Fulton, MD 20759
or
Mary Jo Bogenschutz: *godwinm001@hawaii.rr.com*

Written - September 1994

1. Zde zadáme název hledané látky.

2. A potvrdíme.

Chemical Query

Dr. Duke's
Phytochemical and Ethnobotanical Databases

Plants Containing NICOTINE

Ordered by quantity

Species	Part	Quantity	Reference
Nicotiana tabacum L. -- Tobacco	Leaf	40,000 ppm	DUKE1992A
Equisetum arvense L. -- Field Horsetail, Horsetail	Plant		

Sun Sep 15 09:52:27 EDT 2013

Please send questions and comments to:

James A. Duke
Green Farmacy Garden
8210 Murphy Road
Fulton, MD 20759

or **Mary Jo Bogenschutz** (E-Mail: godwinm001@hawaii.rr.com)

Dr. Duke does not recommend self diagnosis or self medication. Please see the disclaimer for more information.

rostlina

její část

obsah v ppm

3. Nakonec si prohlédneme výsledky.

Co je ppm? http://cs.wikipedia.org/wiki/Parts_per_million
Názvy rostlin jsou uváděny v latině a části rostliny v angličtině.

počítačové formáty molekul SMILES a mol

- pokud se podíváme na název nebo lépe na strukturní vzorec, tak je nám jasné o jakou molekulu se jedná
- ale počítači se s těmito informacemi pracuje těžkopádně, takže se nejčastěji v programech pracujícími s molekulami setkáváme se SMILES notací a formáty MOL (označovaný také jako SDF), MOL2, PDB
- **SMILES** – jedná se o řetězec (posloupnost) znaků, ve kterém je uložena struktura molekuly – jaké atomy obsahuje (používají se klasicky atomové symboly, např. C pro uhlík) a jejich topologii (jak jsou atomy pospojovány vazbami)

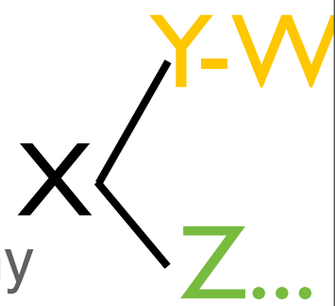
SMILES

- vodíky (které mohou být snadno dopočítány, například v alkanech) se v notaci vynechávájí a dopočítávají se

- atomy jdoucí za sebou jsou spojeny jednoduchou vazbou
  příklady: C (metan), CC (ethan), ..., CO (H_3COH, methanol)

- dvojná vazba je znázorněna "=" a trojná "#"
  příklady: C=C (ethen), C=O (formaldehyd), C#C (ethyn), C#N (kyanovodík)

- pomocí závorek "()" znázorňujeme větvení
  X(YW)Z... - na X je jednoduchou vazbou navázano Y a Z, Y a W jsou spojeny jednoduchou vabou, mezi Y nebo W a Z není žádná vazba
  příklady: CC(CC)CCC (2-ethylpentan), CC(Cl)C (2-chloropropan), CC(=O)C (aceton)



- pomocí čísel jsou označovány kruhy: C1 .... C1 (začátek a konec kruhu)
  příklady: C1CCCCC1 (cyklohexan), C1OC1 (oxiran)

- malými písmeny označujeme aromatické atomy
  příklady: c1ccccc1 (benzen), n1ccccc1 (pyridin)

- [NH] explicitně vyjadřený vodík, [O-] vyjádřený ion, [C@@H] vyznačená chiralita, ...

- tutoriál v angličtině: http://www.daylight.com/meetings/summerschool98/course/dave/smiles-intro.html

MOL, MOL2, PDB

```
1
2  4504939
3    ChemPy            3D                              0
4
5   24 25  0  0  1  0  0  0  0  0999 V2000
6     11.3995    1.6808    9.6604 N   0  0  0  0  0  0  0  0  0  0  0  0
7      9.7930    1.4933    6.5490 N   0  0  0  0  0  0  0  0  0  0  0  0
8     12.0335    1.5976    7.5192 N   0  0  0  0  0  0  0  0  0  0  0  0
9      8.0344    1.5210    8.1449 N   0  0  0  0  0  0  0  0  0  0  0  0
10    10.2441    1.6220    8.8950 C   0  0  0  0  0  0  0  0  0  0  0  0
11     8.4332    1.4728    6.8021 C   0  0  0  0  0  0  0  0  0  0  0  0
12    10.6751    1.5732    7.5977 C   0  0  0  0  0  0  0  0  0  0  0  0
13     8.8682    1.6062    9.2669 C   0  0  0  0  0  0  0  0  0  0  0  0
14     6.5816    1.5025    8.3796 C   0  0  0  0  0  0  0  0  0  0  0  0
15    10.2685    1.5184    5.1722 C   0  0  0  0  0  0  0  0  0  0  0  0
16    11.4806    1.7343   11.1148 C   0  0  0  0  0  0  0  0  0  0  0  0
17    12.4246    1.6682    8.8040 C   0  0  0  0  0  0  0  0  0  0  0  0
18     8.4213    1.6517   10.4075 O   0  0  0  0  0  0  0  0  0  0  0  0
19     7.6183    1.4048    5.9008 O   0  0  0  0  0  0  0  0  0  0  0  0
20     6.4060    1.5019    9.3438 H   0  0  0  0  0  0  0  0  0  0  0  0
21     6.1959    0.6965    7.9767 H   0  0  0  0  0  0  0  0  0  0  0  0
22     6.1741    2.2967    7.9750 H   0  0  0  0  0  0  0  0  0  0  0  0
23    11.2481    1.4946    5.1644 H   0  0  0  0  0  0  0  0  0  0  0  0
24     9.9578    2.3383    4.7348 H   0  0  0  0  0  0  0  0  0  0  0  0
25     9.9181    0.7394    4.6923 H   0  0  0  0  0  0  0  0  0  0  0  0
26    12.4215    1.7514   11.3886 H   0  0  0  0  0  0  0  0  0  0  0  0
27    11.0444    0.9440   11.4955 H   0  0  0  0  0  0  0  0  0  0  0  0
28    11.0315    2.5430   11.4378 H   0  0  0  0  0  0  0  0  0  0  0  0
29    13.3348    1.7046    9.0738 H   0  0  0  0  0  0  0  0  0  0  0  0
30   1  5  1  0  0  0  0
31   1 11  1  0  0  0  0
32   1 12  1  0  0  0  0
33   2  6  1  0  0  0  0
34   2  7  1  0  0  0  0
35   2 10  1  0  0  0  0
36   3  7  1  0  0  0  0
37   3 12  2  0  0  0  0
38   4  8  1  0  0  0  0
39   4  9  1  0  0  0  0
40   5  7  2  0  0  0  0
41   5  8  1  0  0  0  0
42   6  4  1  0  0  0  0
43   9 15  1  0  0  0  0
44   9 16  1  0  0  0  0
45   9 17  1  0  0  0  0
46  10 18  1  0  0  0  0
47  10 19  1  0  0  0  0
48  10 20  1  0  0  0  0
49  11 21  1  0  0  0  0
50  11 22  1  0  0  0  0
51  11 23  1  0  0  0  0
52  12 24  1  0  0  0  0
53  13  8  2  0  0  0  0
54  14  6  2  0  0  0  0
55 M  END
```

počet atomů; počet vazeb ...

x, y a z souřadnice; symbol prvku; ...

1. atom 12. atom spojený jednoduchou (1) vazbou

MOL (SDF) je formát molekuly, který obsahuje informace o atomech, jejich vlastnostech, poloze v prostoru (x,y,z koordináty v kaztézském prostoru) a informace o vazbách, MOL2 je rozšířením formátu MOL, tyto formáty se používají především pro malé molekuly, PDB formát se pak používá pro biomolekuly, ale o něm až někdy přístě

více o specifikaci formátu mol (CTFile) můžete najít zde: http://c4.cabrillo.edu/404/ctfile.pdf

Databáze

- **PubChem** – http://pubchem.ncbi.nlm.nih.gov/
  - tato databáze obsahuje informace o **malých molekulách a jejich biologických aktivitách** (na bližší informace o biologické aktivitě se můžete tešit v příští úloze)
- **NCI** – http://cactus.nci.nih.gov/ncidb2.2/
  - obsahuje struktury a jejich relevantní informace ke studiu rakoviny
- **ZINC** – http://zinc.docking.org/
  - je “free” databáze komerčně dostupných látek pro **virtuální screening** (o virtuálním screeningu se dovíte v příští úloze více)
  - struktury v této databázi jsou připravené pro **docking**, který si také vyzkoušíme v jedné z příštích úloh
- **KAŽDÁ DATABÁZE obsahuje 3D souřadnice, které jsou generovány empirickým přístupem a ve většině případů se liší od experimentálně získané struktury.**

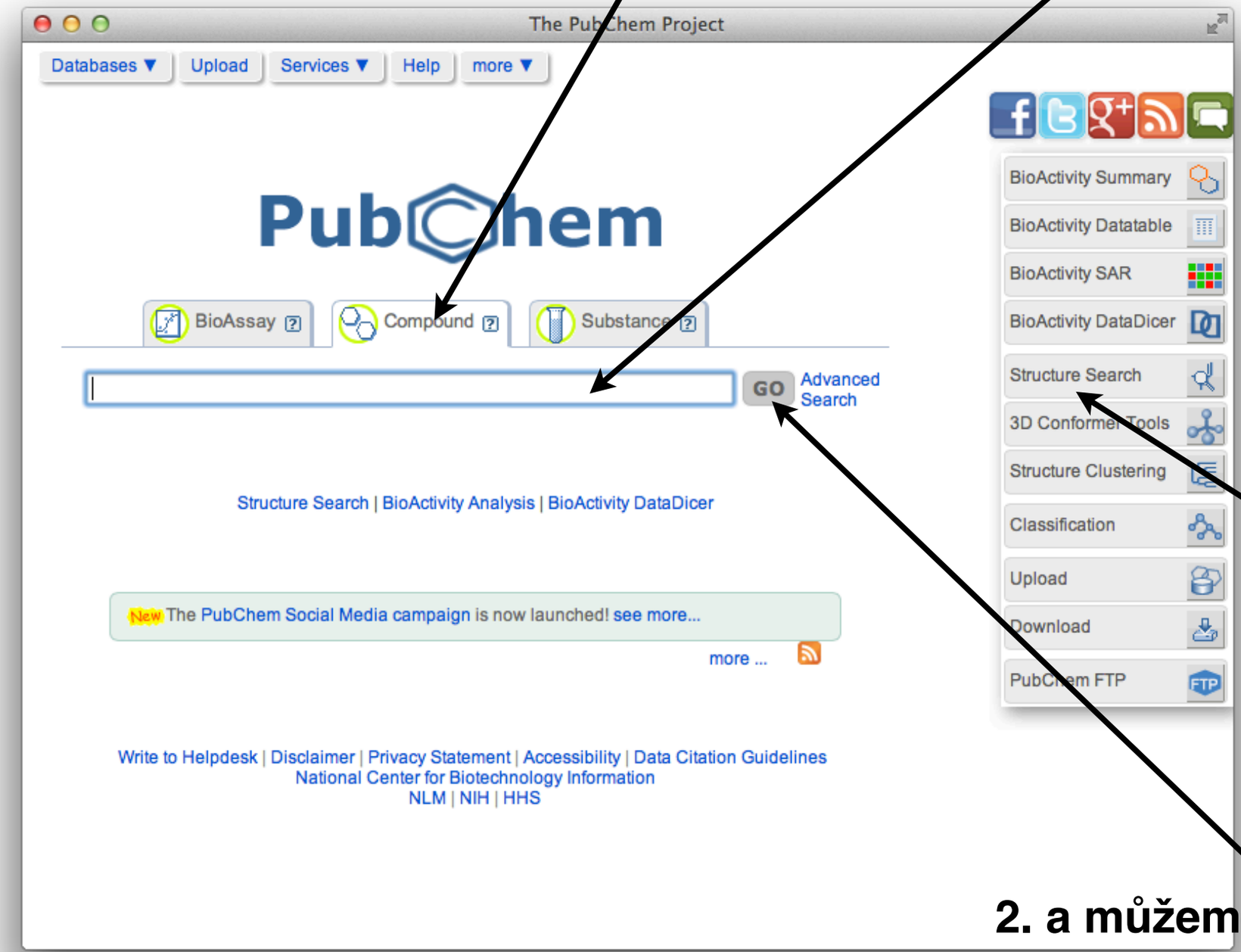
PubChem

- vyhledávání

vyhledáváme sloučeniny

1. zde zadáme požadová klíčová slova, podobně jako v Googlu

tip: používejte anglické názvy sloučenin



zde můžeme hledat v pokročilém režimu (využijeme v podobnostním hledání)

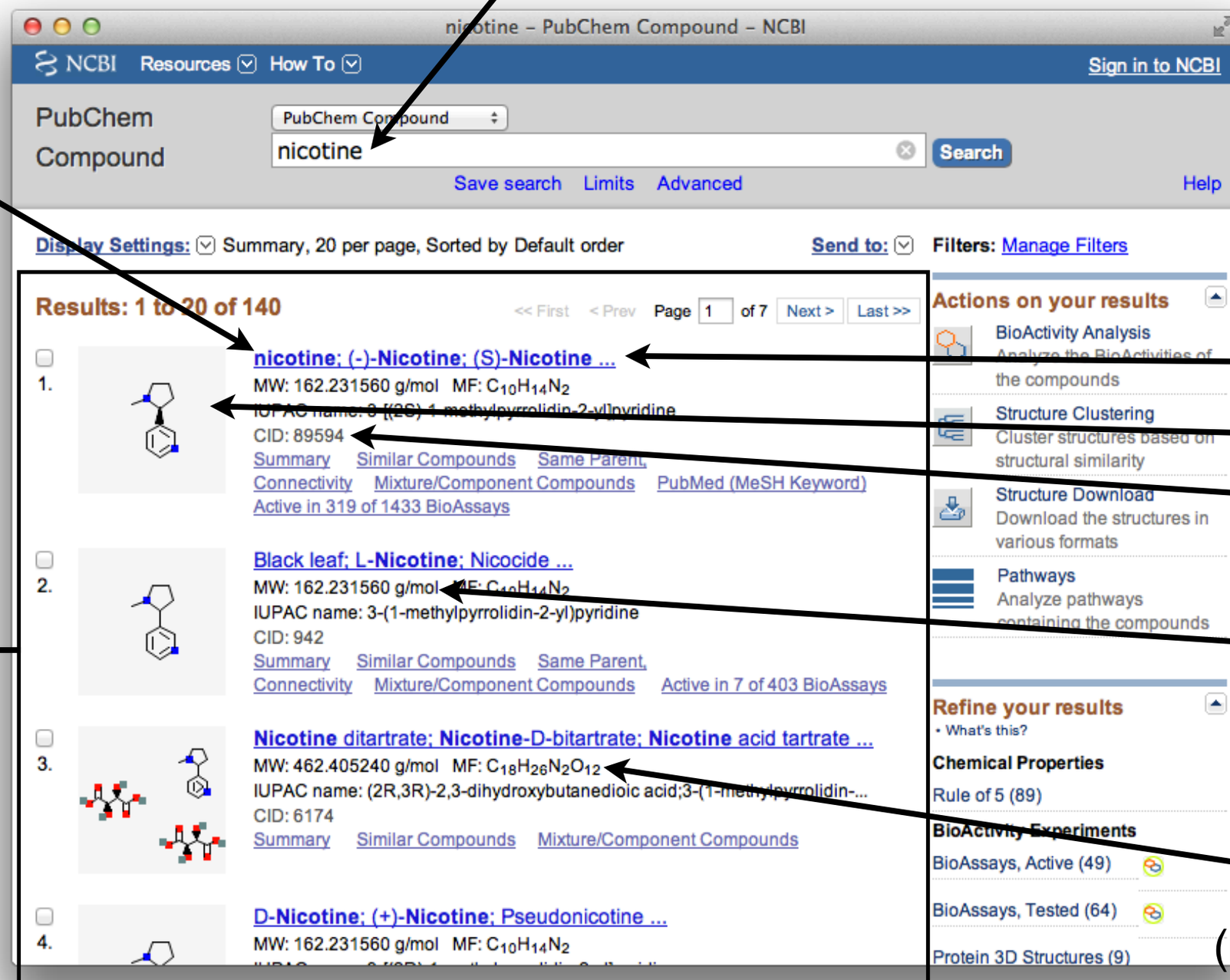
2. a můžeme začít vyhledávat

PubChem

- výsledky

zadaná klíčová slova, v tomto případě nikotin

1. klikneme na název a podíváme se na detail



výsledky

název v databázi

strukturní vzorec

identifikátor (compound ID)

molekulová hmotnost (molecular weight)

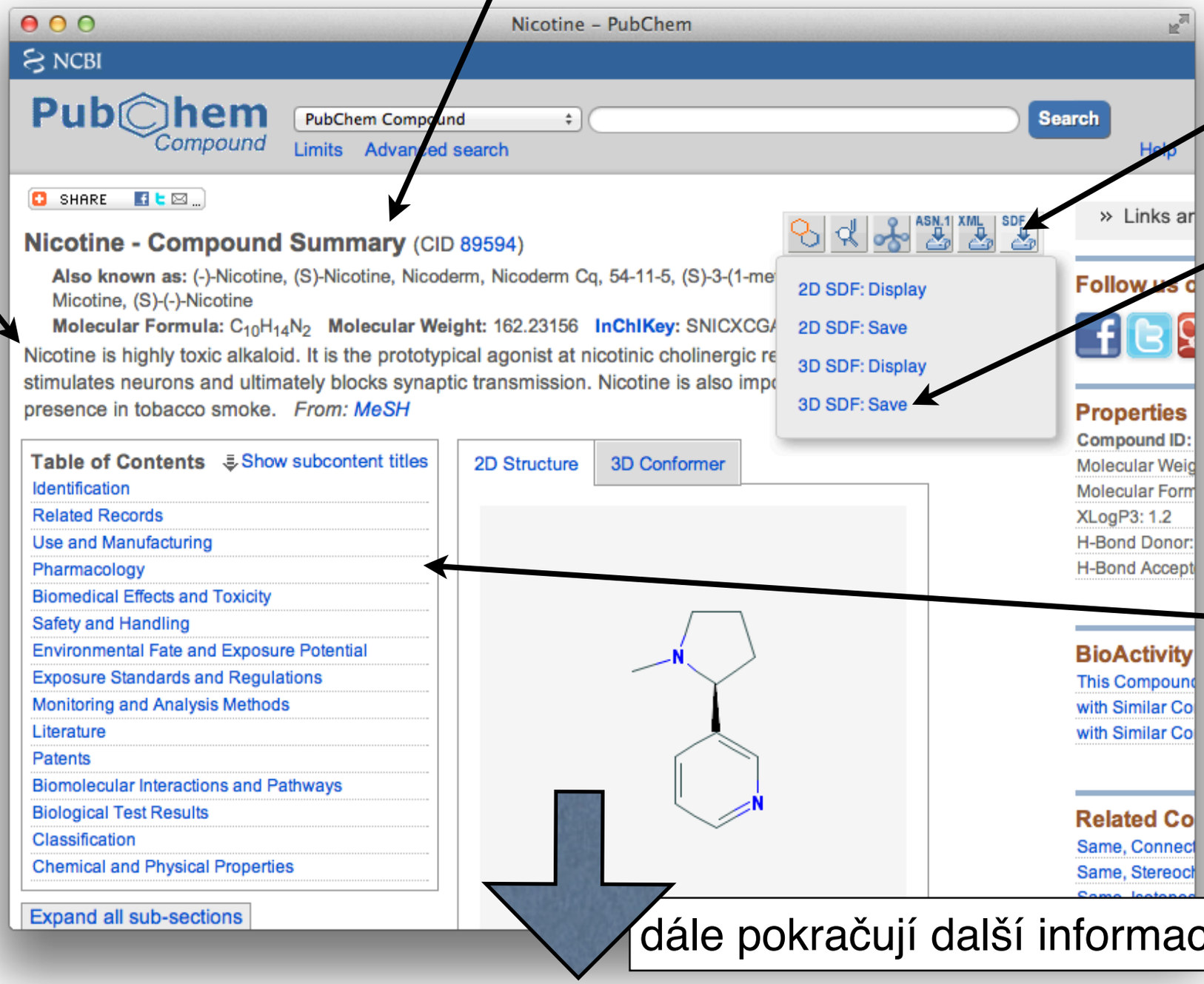
sumarní vzorec (molecular formula)

PubChem

- detail

informace o sloučenině, v našem případě o nikotinu

povídání o nikotinu



1. rozklikneme SDF ikonu

2. stáhneme 3D strukturu do počítače

seznam informací dostupný k dané molekule

dále pokračují další informace

NCI

- vyhledávání

1. vybereme vyhledávací kriterium
"Name Search ..."

2. zde pak můžeme zadat název molekuly

3. stisknutím tlačítka "Start Search" zahajíme hledání

Enhanced NCI Database Browser

Home | About | Contact | Disclaimer | Privacy | Classic Version

Enhanced NCI Database Browser 2.2

Editor | Query | Hitlist | Detail | Display | Downloads | Help | FAQ | News | Credits

Database status: 250250 open structures ready for searching. Bug reports, comments or questions?

Query Type | Negate | Query Data Value

NSC Number(s)
✓ NSC Number(s)
CAS Number(s)
Formula...
Record Number(s)
Molecular Weight Range
Number of Atoms Range
Number of ESSR Rings Range
Number of H-Bond Donors Range
Number of H-Bond Acceptors Range
Number of Rotatable Bonds Range
Number of Catalyst Conformers
Drug Likeness Prediction...
PASS Prediction Range...
logP Range...
Name Search...
Complexity Range
AIDS Screen Result
Yeast Screen Level
Stereocenters
Data Availability Constraints...
LIQCRYST Number(s)
Random Set
Full Structure...
Similarity Search...
Transformation Search...
Functional Group Search...
Substructure and/or 3D Search

nicotine
Editor | (No options)
Editor | (No options)
Editor | Other elements allowed
Editor | (No options)
Choose File | no file selected
All molecules

AND OR XOR
100 hits, 90 seconds
HTML Table with Samples | preferably 3D
NSC Number

NCI

- výsledky

1. klikneme na NSC a podíváme se na detail

strukturní vzorce nalezených sloučenin

identifikátor sloučeniny v databázi (NSC číslo)

výsledky

Enhanced NCI Database Browser

Home | About | Contact | Disclaimer | Privacy | Classic Version

Enhanced NCI Database Browser 2.2

Editor | Query | Hitlist | Detail | Display | Downloads | Help | FAQ | News | Credits

Database status: 250250 open structures ready for searching. Bug reports, comments or questions?

Operations with this Dataset of 1 Structure:

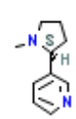
Data Retrieval: Format: SDFile; Output in 3D ☐ Strip H ☐; Max #Records: 100; Fields: NSC Number, Molecular Weight, Name (ACD), All Names; Retrieve

Visualization: GIF Image Gallery; Display

Miscellaneous: Restart Query (at first record); Execute

Structure

5065



	NSC Number	Formula	CAS	#Names	Sample Name
☑	5065	$C_{10}H_{14}N_2$	54-11-5	40	(-)-Nicotine

Invert Selection

Date: 2013-09-15 07:59:33

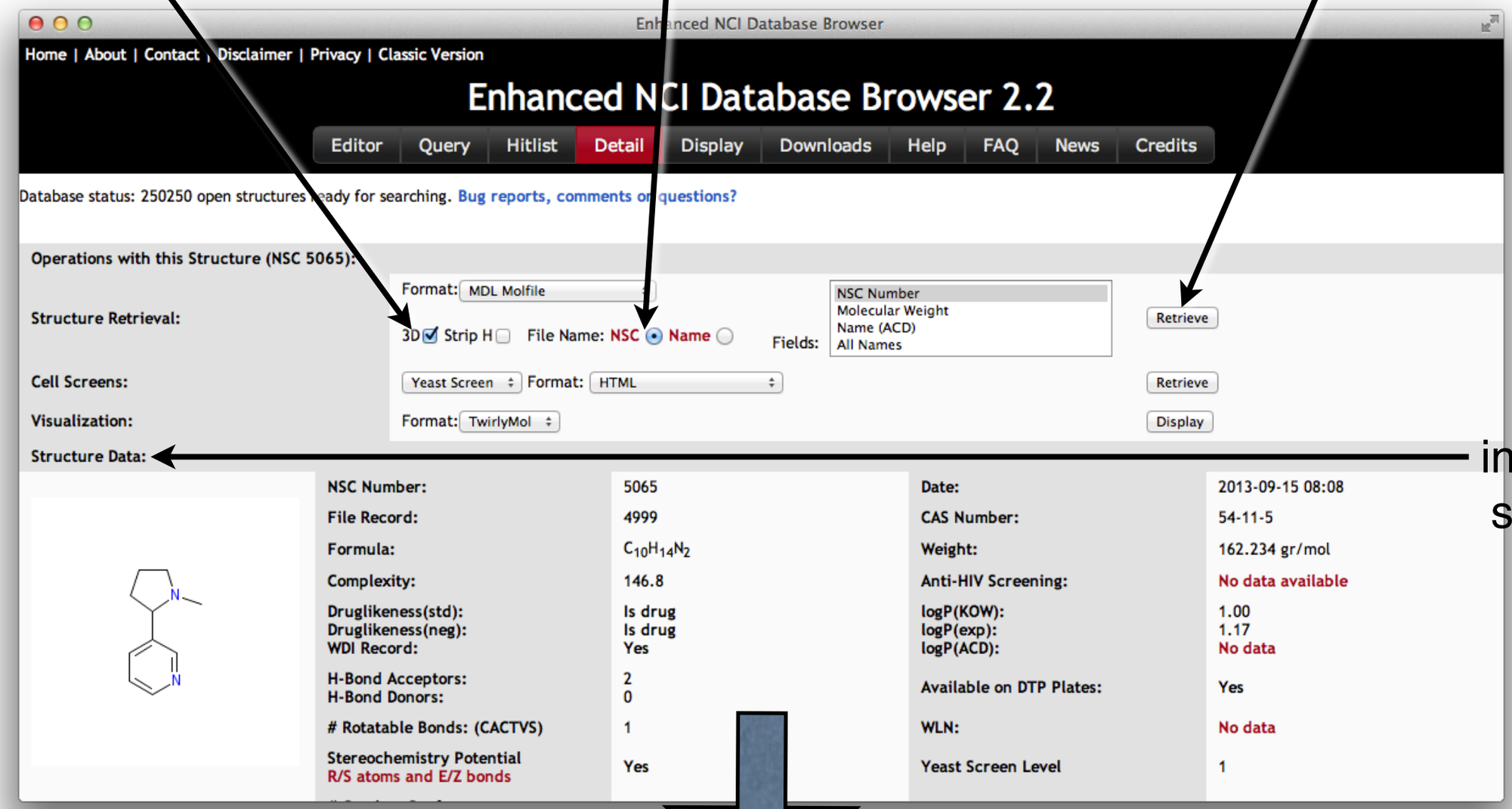
NCI

- detail

1. zaklikněme 3D

2. vybereme NSC

3. stáhneme 3D strukturu do počítače



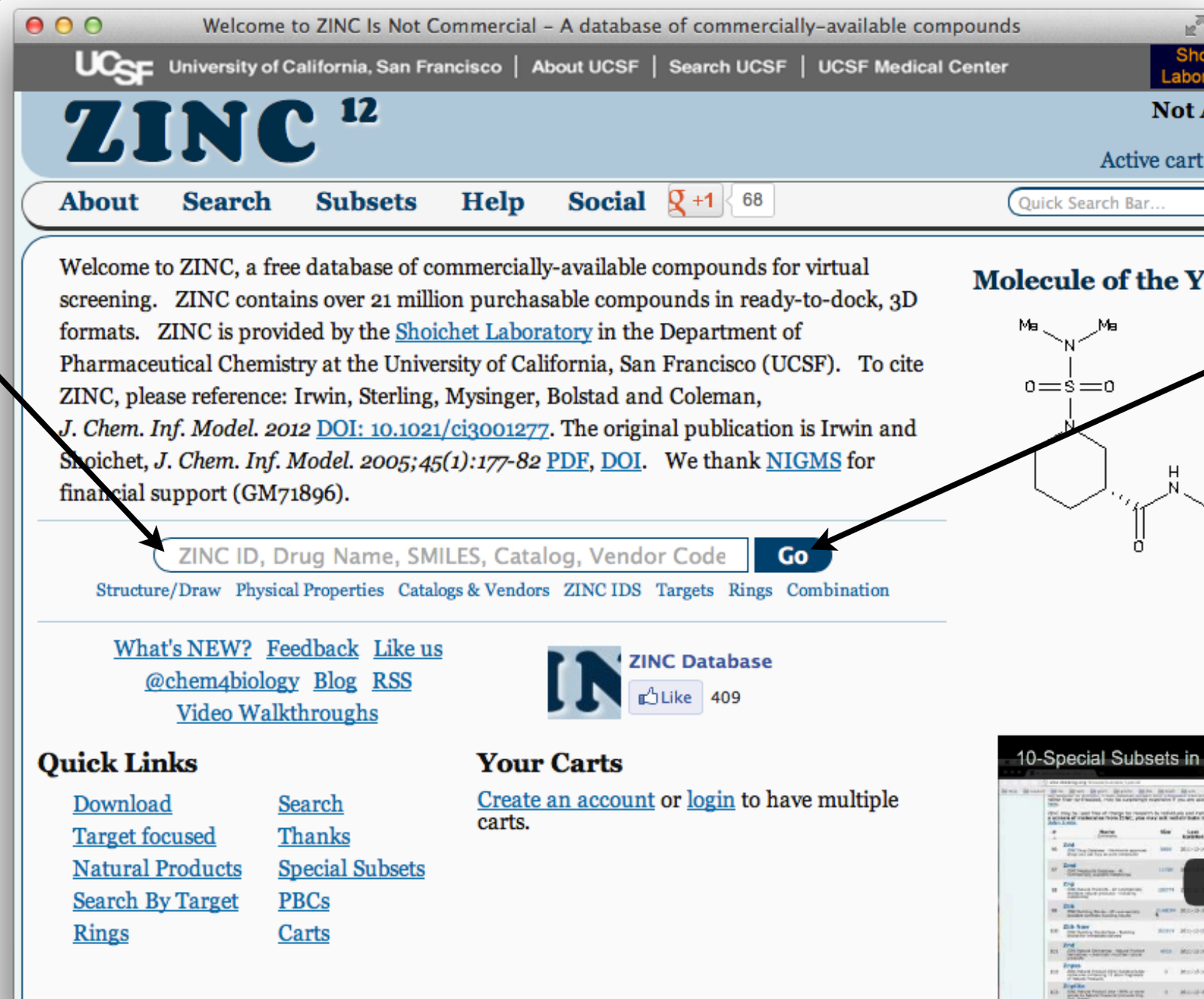
informace o sloučenině

dále pokračují další informace

ZINC

- vyhledávání

1. zadáme název
dále se může zadat ZINC ID, SMILES, ...

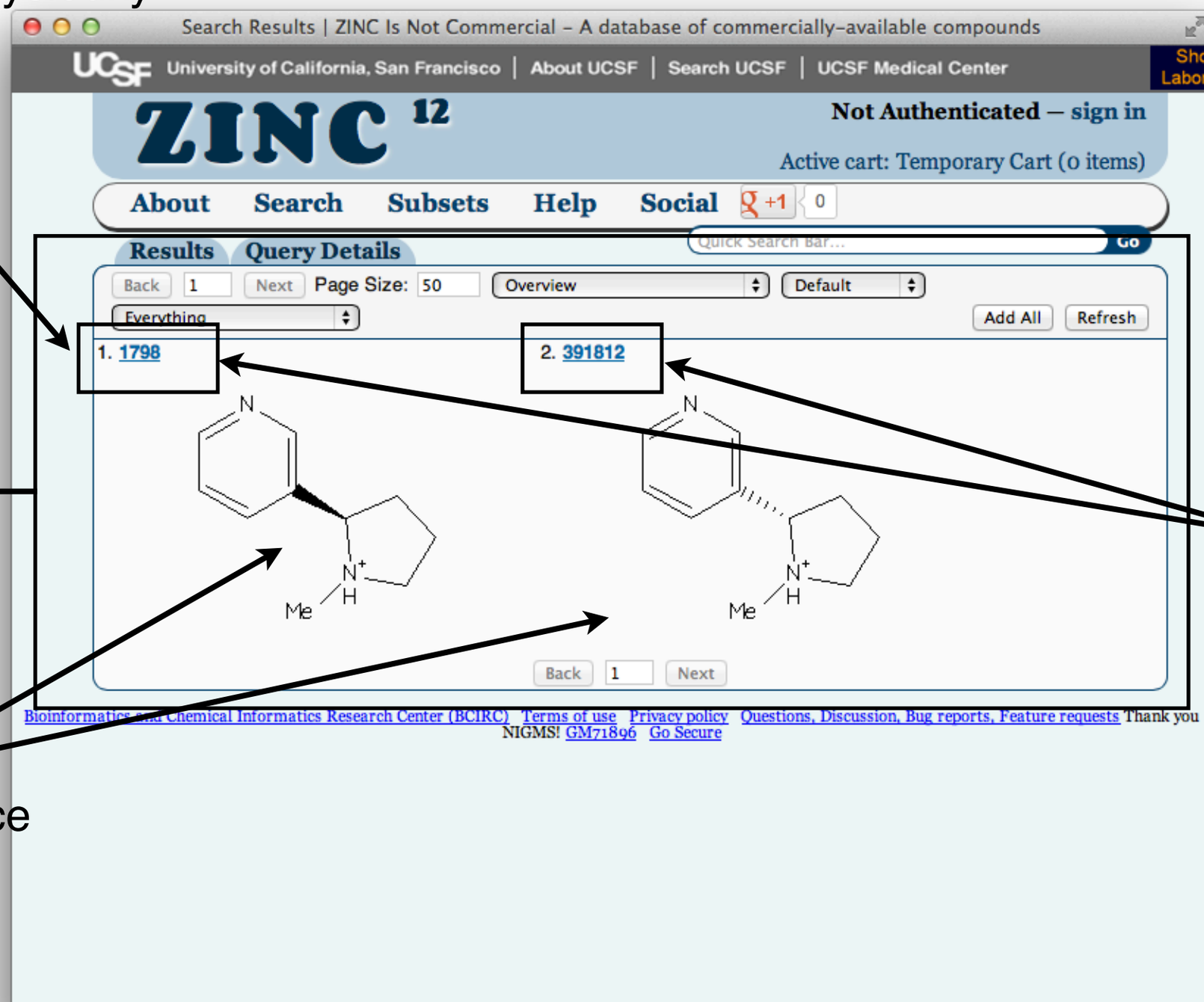


2. a začneme vyhledávat

NCI

- výsledky

1. klikneme na ZINC ID a podíváme se na detail



výsledky

identifikátor sloučeniny v databázi (ZINC ID)

strukturní vzorce nalezených sloučenin

NCI

- detail

ZINC00001798 | ZINC Is Not Commercial – A database of commercially-available compounds

UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center

ZINC 12

Not A

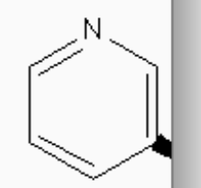
Active cart:

About Search Subsets Help Social +1 0 Quick Search Bar...

Synonyms (143) | Vendors (25) | Annotations (18) | Representations (2) | Notes (14) | Clustered Activity (0) | Targets (0) | Rin

ZINC00001798

In ZINC since	Heavy atoms	Benign functionality
October 4th, 2005	12	Yes



Popular Name: ***Nicotine***
Find On: ***PubMed — Wikipedia — Google***
CAS Numbers: ***22083-74-5 , 25162-00-9; 69782-38-3 , 54-11-5 , 6019-06-3 , 69980-24-1 , [54-11-5]***

Other Names:

(–
(+)-NICOTINE
(+)-nicotine; (R)-3-(1-methyl-2-pyrrolidinyl)pyridine; d-nicotine; pseudonicotine
(+-)-nicotine; (R,S)-nicotine; (RS)-nicotine; Nicotin; Nikotin
(+/-)-NICOTINE-METHYL-D3
(-)-Nicotine

Draw Identit

SMILES: C[NH+]1CCC[C@@H]1c2cccnc2

Download: MOL2 SDF SMILES Flexibase

Vendors

Vitas-M STL146319

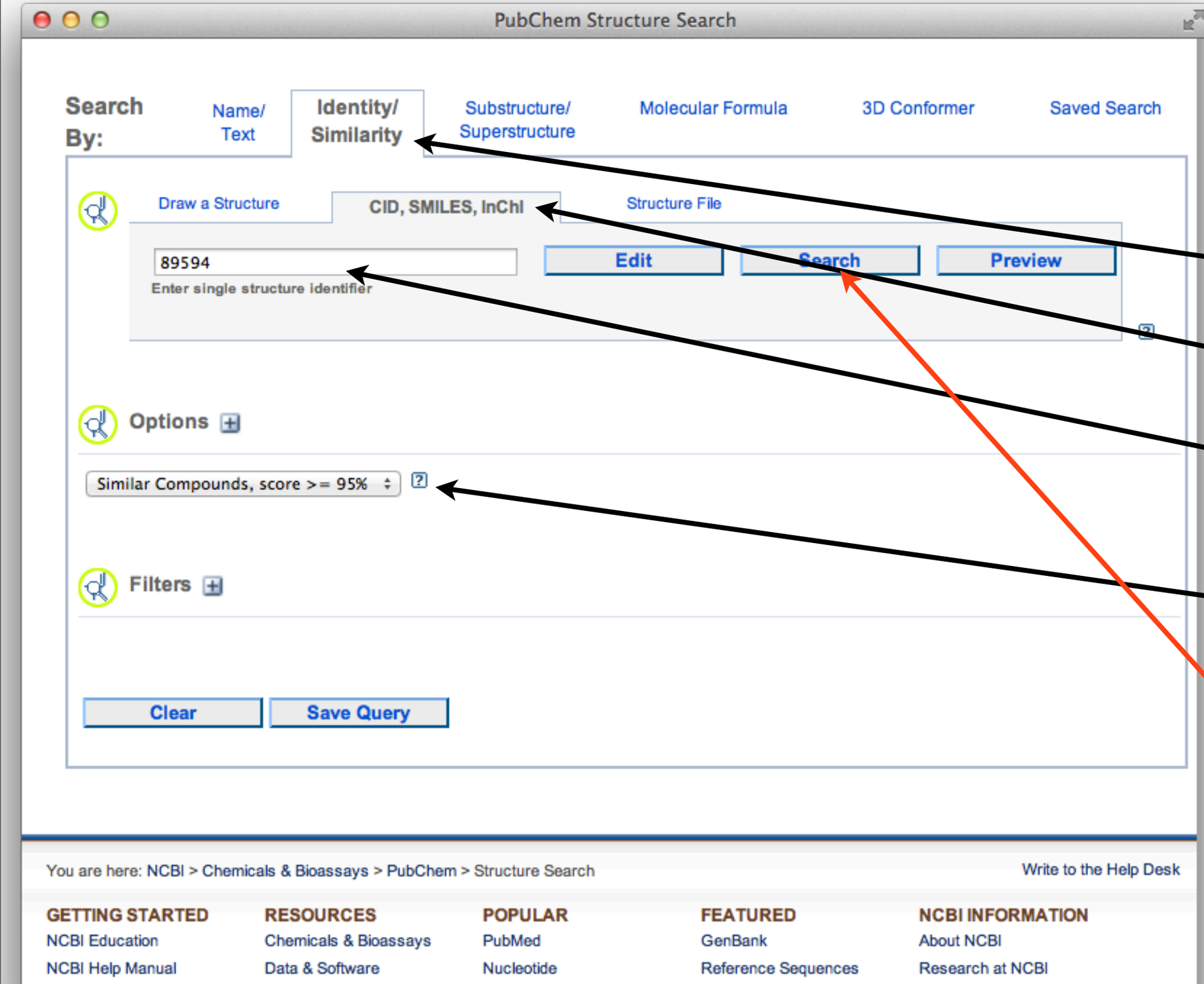
eMolecules 529638

Annotations

1. stáhneme 3D strukturu do počítače

dále pokračují další informace

Podobnostní hledání v PubChemu



0. otevřeme “Structure Search” v PubChemu <u>podle tohoto slidu</u>
1. vybereme “Identity/ Similarity”
2. budeme hledat podle “CID,SMILES,InChI”
3. podle vyhledávání nikotinu zadáme jeho CID
4. v “Options” vybereme “Similar Compounds, score >= 95%”
5. můžeme začít hledat

Podobnostní hledání v PubChemu

526 selected items – PubChem Compound – NCBI

Results: 1 to 20 of 526

Page 1 of 27 Next > Last >>

1. Black leaf; L-Nicotine; Nicocide ...
MW: 162.231560 g/mol MF: $C_{10}H_{14}N_2$
IUPAC name: 3-(1-methylpyrrolidin-2-yl)pyridine
CID: 942
Summary Similar Compounds Same Parent, Connectivity Mixture/Component Compounds Active in 7 of 403 BioAssays

2. nicotine; (-)-Nicotine; (S)-Nicotine ...
MW: 162.231560 g/mol MF: $C_{10}H_{14}N_2$
IUPAC name: 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine
CID: 89594
Summary Similar Compounds Same Parent, Connectivity Mixture/Component Compounds PubMed (MeSH Keyword) Active in 319 of 1433 BioAssays

3. D-Nicotine; (+)-Nicotine; Pseudonicotine ...
MW: 162.231560 g/mol MF: $C_{10}H_{14}N_2$
IUPAC name: 3-[(2R)-1-methylpyrrolidin-2-yl]pyridine
CID: 157672
Summary Similar Compounds Same Parent, Connectivity Mixture/Component Compounds PubMed (MeSH Keyword) Active in 8 of 81 BioAssays

4. AC1L41LR; 94292-33-8
MW: 164.224102 g/mol MF: $C_{10}H_{14}N_2$
IUPAC name: 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine
CID: 175605
Summary Similar Compounds Same Parent, Connectivity

5. AC1L9ONP; MOLI000373
MW: 161.232294 g/mol MF: $C_{10}H_{14}N_2$
IUPAC name: 3-(1-methylpyrrolidin-2-yl)pyridine
CID: 450006
Summary Similar Compounds Same Parent, Connectivity

Actions on your results

BioActivity Analysis
Analyze the BioActivities of the compounds

Structure Clustering
Cluster structures based on structural similarity

Structure Download
Download the structures in various formats

Pathways
Analyze pathways containing the compounds

Refine your results
What's this?

Chemical Properties
Rule of 5 (381)

BioActivity Experiments
BioAssays, Active (51)
BioAssays, Tested (75)
Protein 3D Structures (2)
Human Cytochrome P450 2a6 In Complex With Nicotine (2)

BioMedical Annotation
Pharmacological Actions (1)
Ganglionic Stimulants (1)
BioSystems (8)

Depositor Category
Biological Properties (308)

526 podobných molekul

všimněme si ze ve výsledcích se nám vrací i obohacené sloučeniny (sloučeniny obsahující na určitém místě specifický nuklid)

1. vyhledáme pouze molekuly s biologickými vlastnostmi

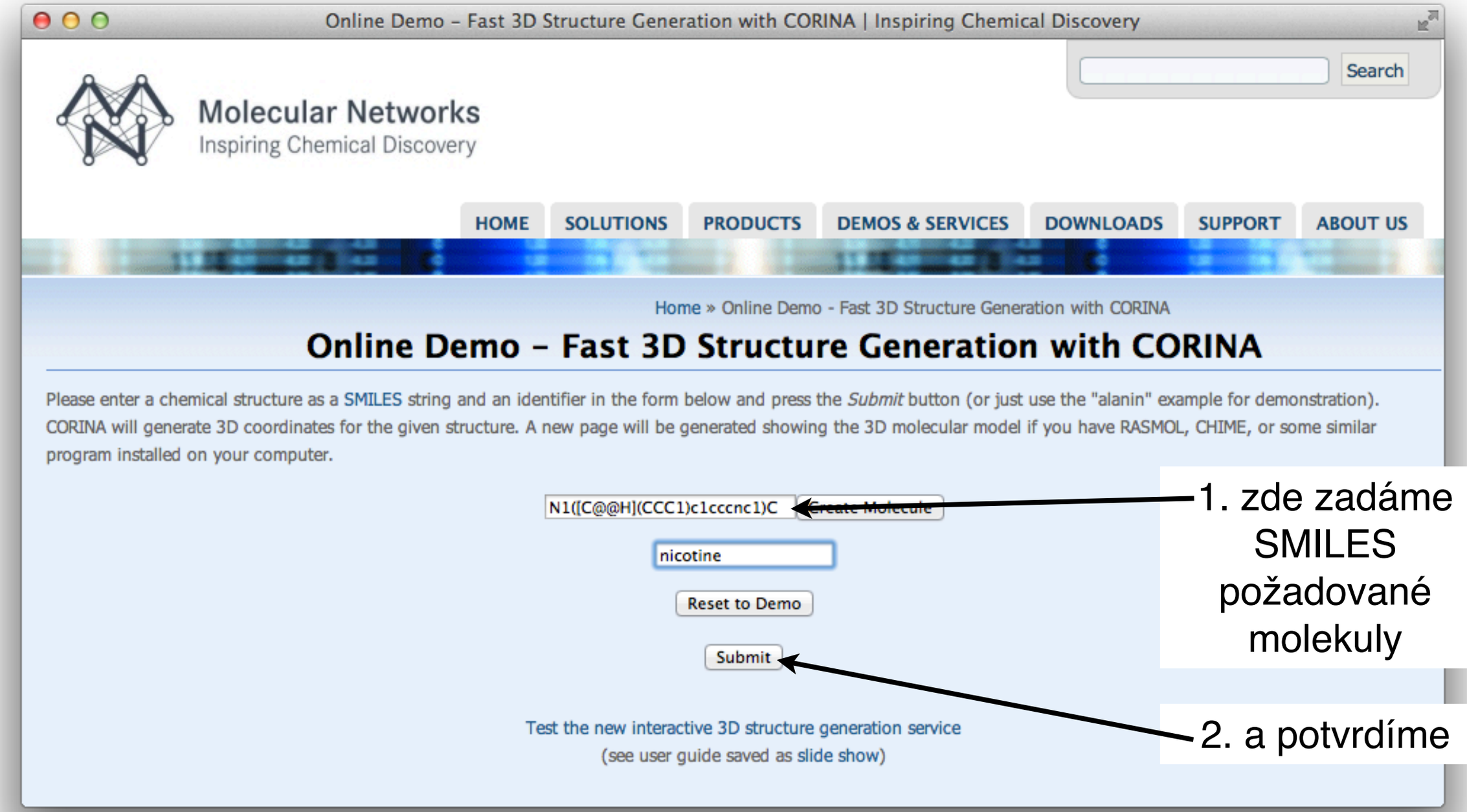
Podobnostní hledání v PubChemu

308 podobných molekul s biologickými vlastnostmi

výsledky můžeme dále filtrovat podle dalších kriterii
- Lipinského pravidlo pěti
- podle znalosti a hodnot bioaktivity
- podle znalosti struktury proteinů
- podle znalosti biologických vlastnosti
...

Corina

- otevřeme v prohlížeči http://www.molecular-networks.com/online_demos/corina_demo

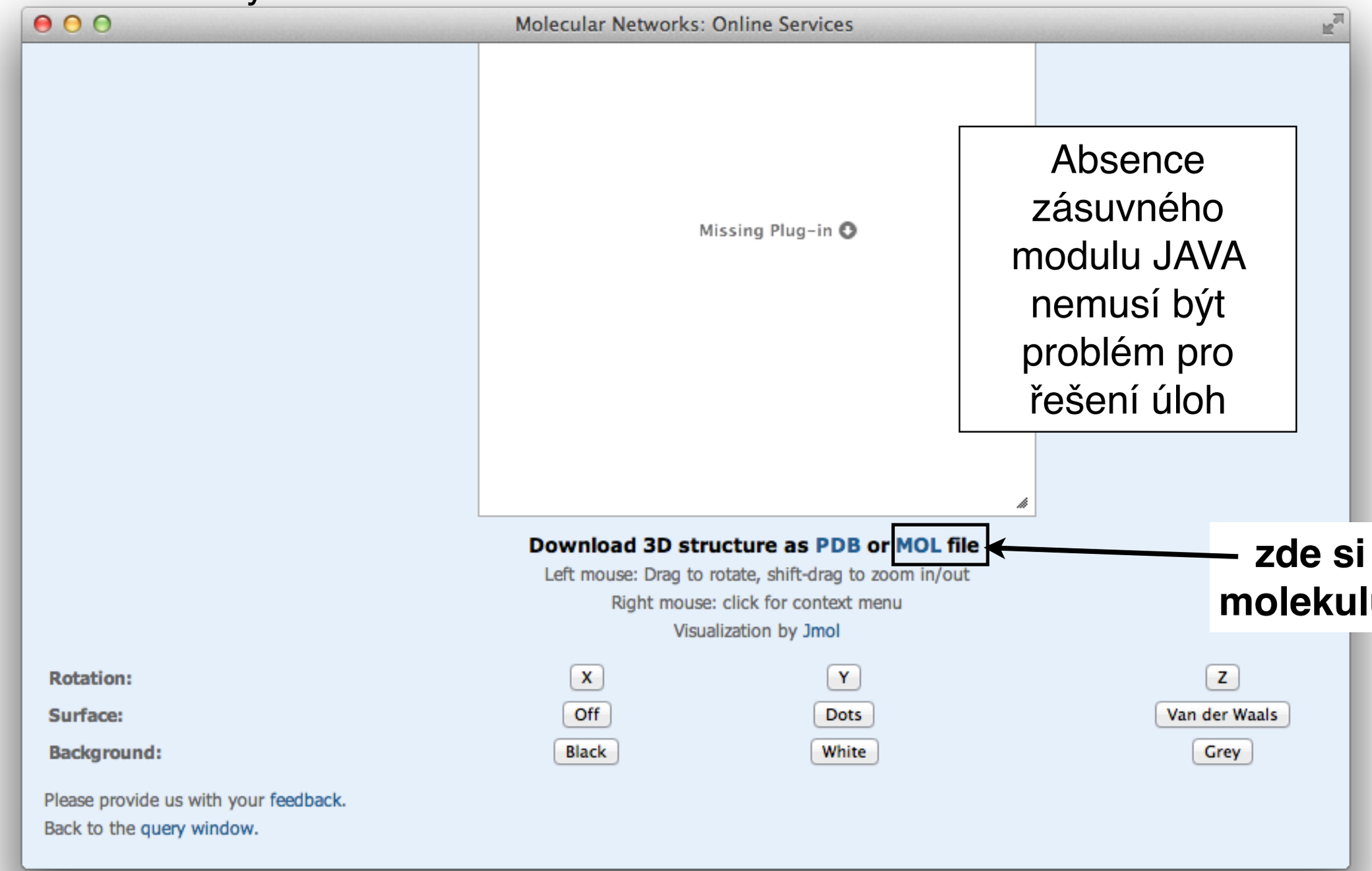


1. zde zadáme SMILES požadované molekuly

2. a potvrdíme

Corina

- výsledek

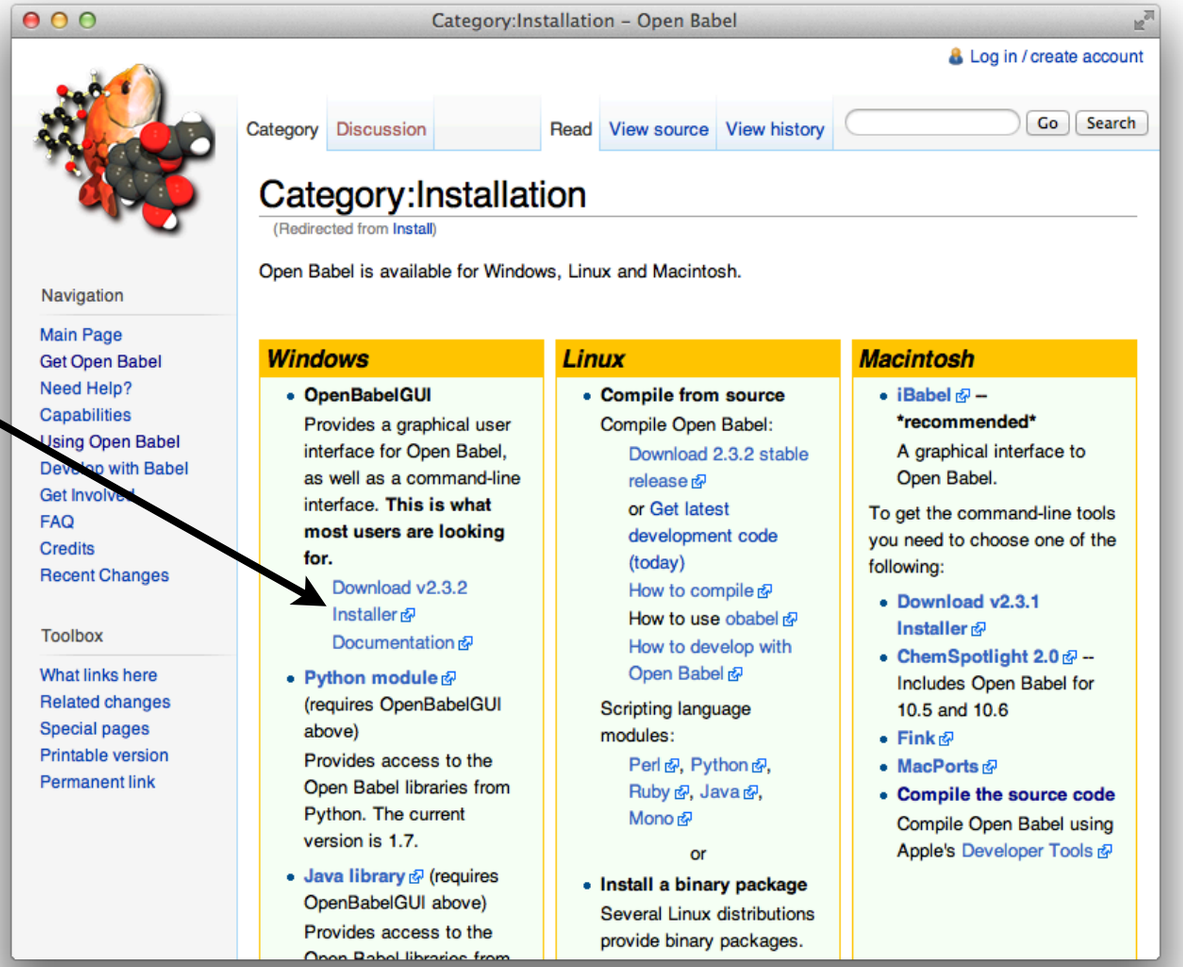


zde si můžeme molekulu stahnout

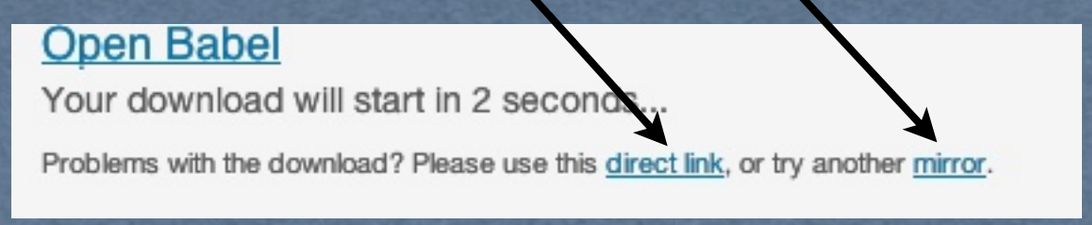
Instalace OpenBabelu

- podrobný návod instalace je popsán pro Windows 7, pokud máte jiný operační systém kontaktujte mě prosím, pomůžu vám s instalací (v případě zájmu připravíme i návod pro Windows XP nebo 8)
- začneme stažením instalačního souboru na adrese http://openbabel.org/wiki/Install

zde

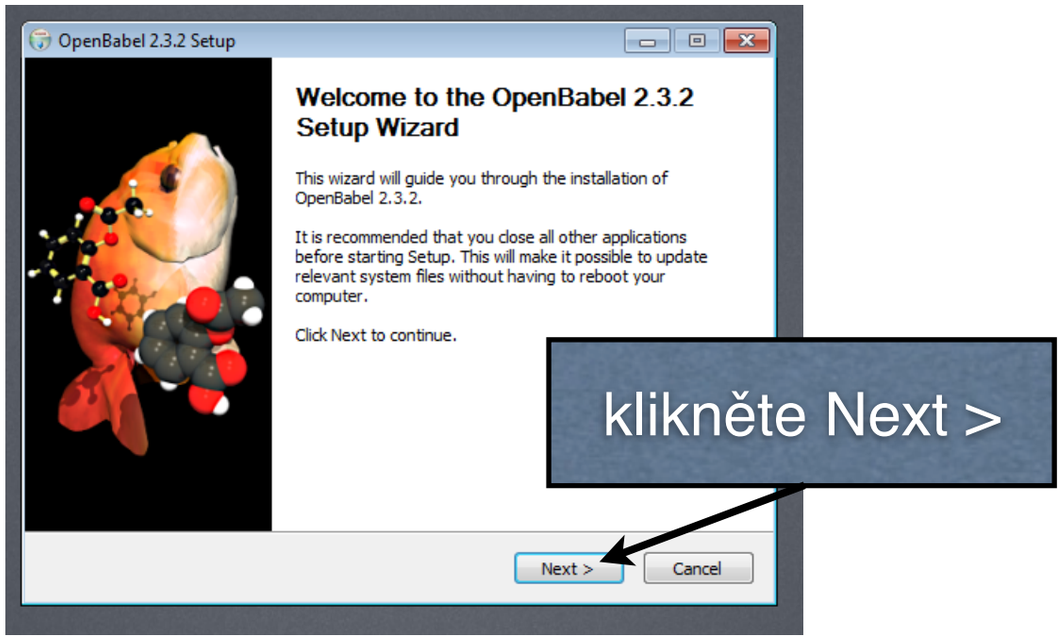


Pokud se vám nezačne program stahovat klikněte zde nebo zde.

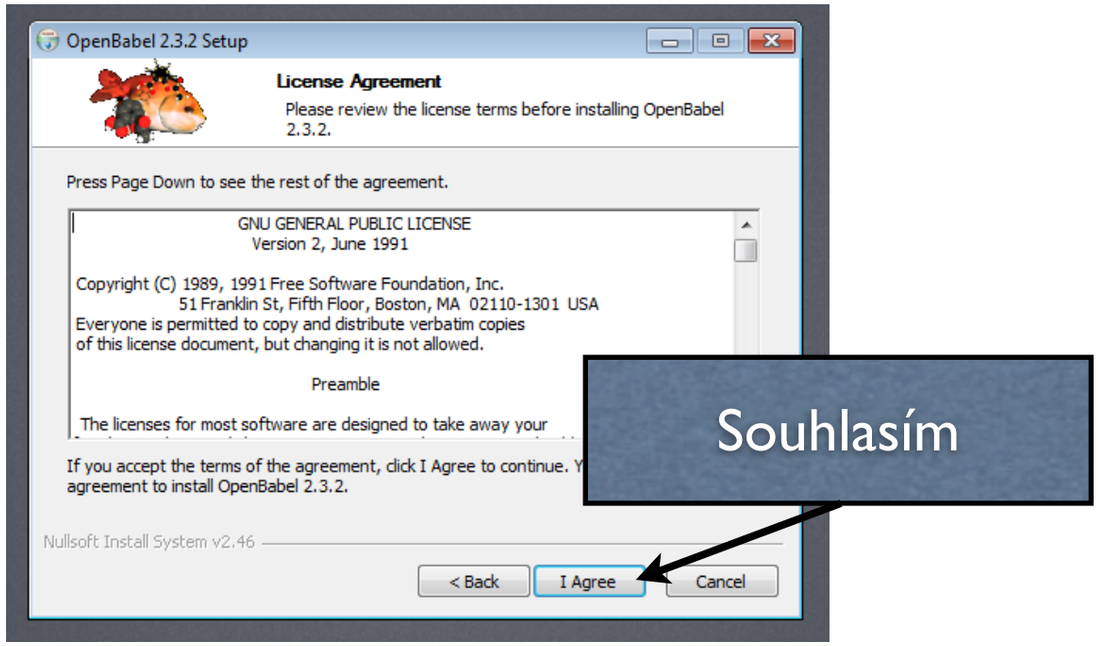


Instalace OpenBabelu

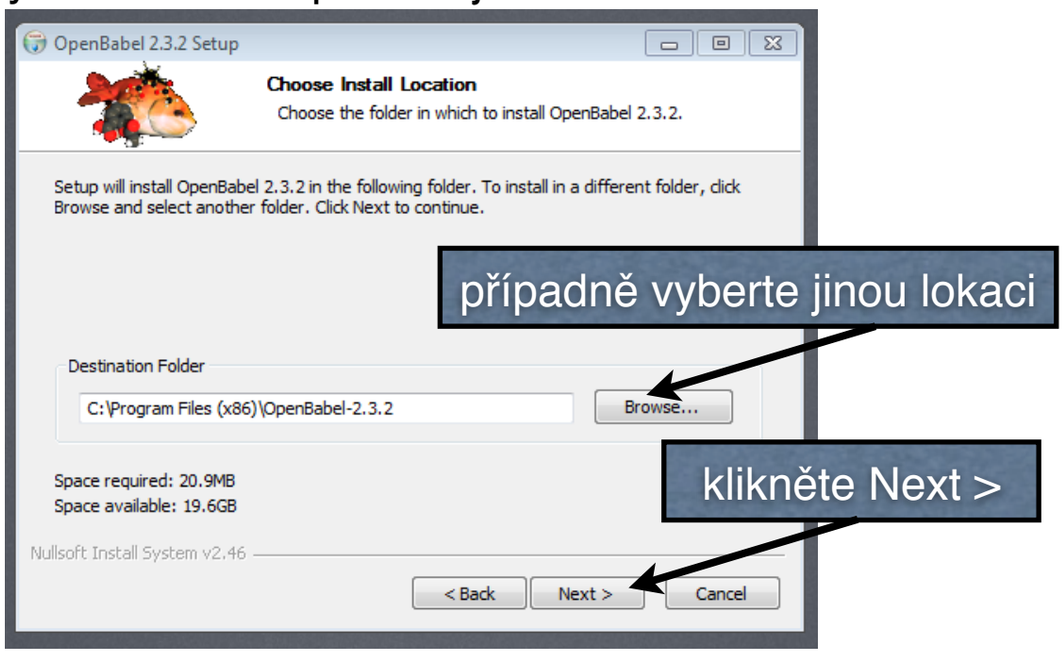
1. otevřeme stažený soubor



2. potvrdíme licenční podmínky



3. vyberte lokaci a pokračujte



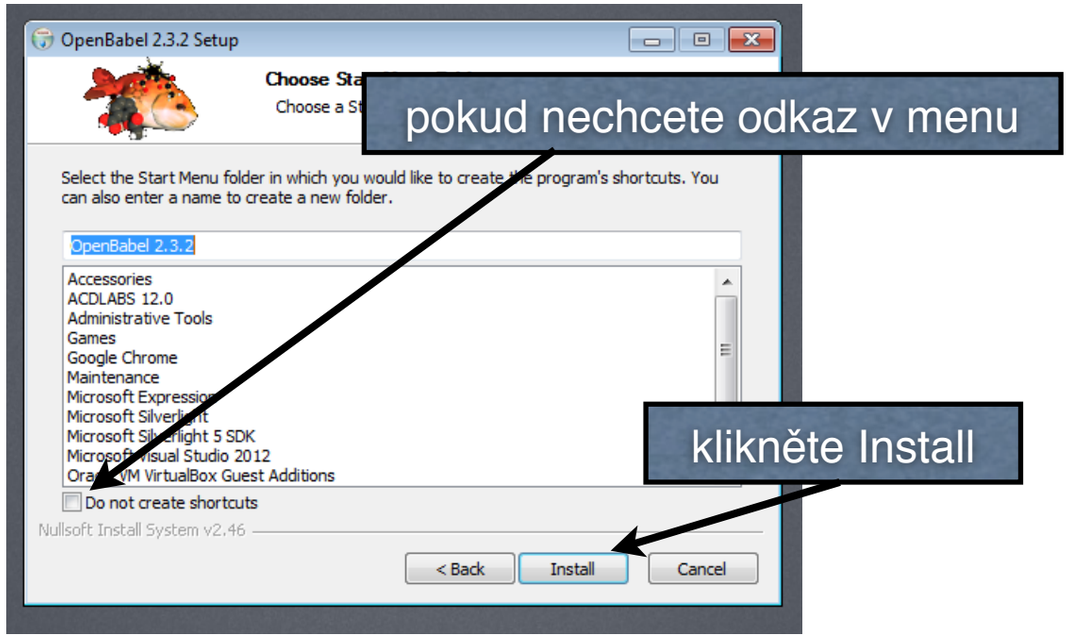
Licenční podmínky:

- program je zdarma a musí být dále šířen zdarma s danou licencí!
- může být modifikováno a dále šířen pouze pod odvozenou licencí
- bližší informace:

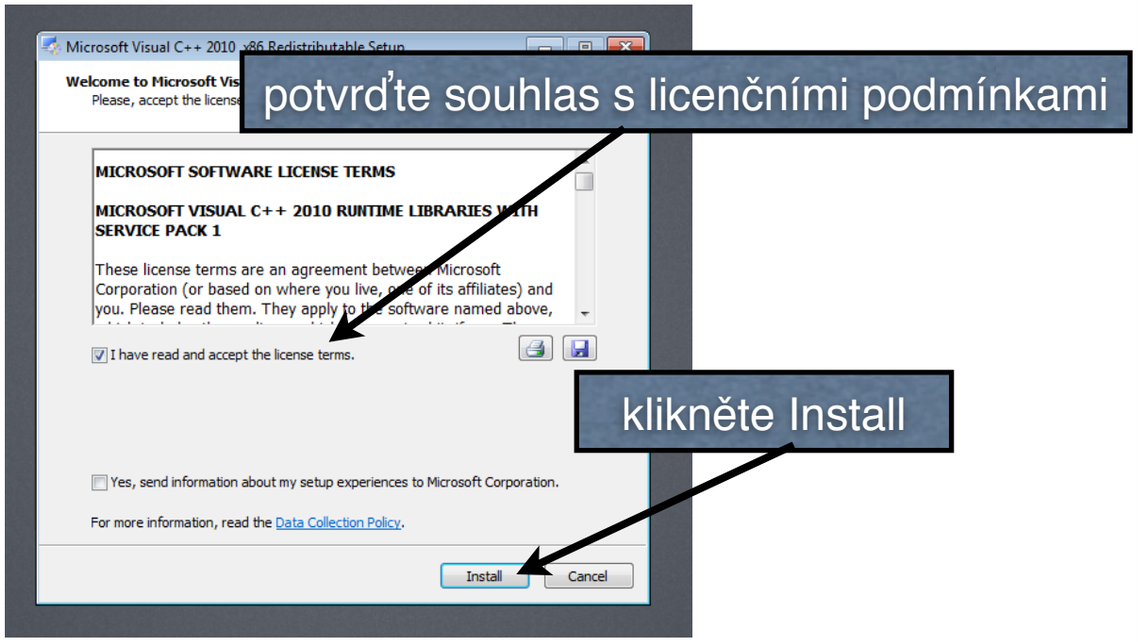
http://cs.wikipedia.org/wiki/GNU_General_Public_License
http://cs.wikipedia.org/wiki/GNU_General_Public_License#Verze_2

Instalace OpenBabelu

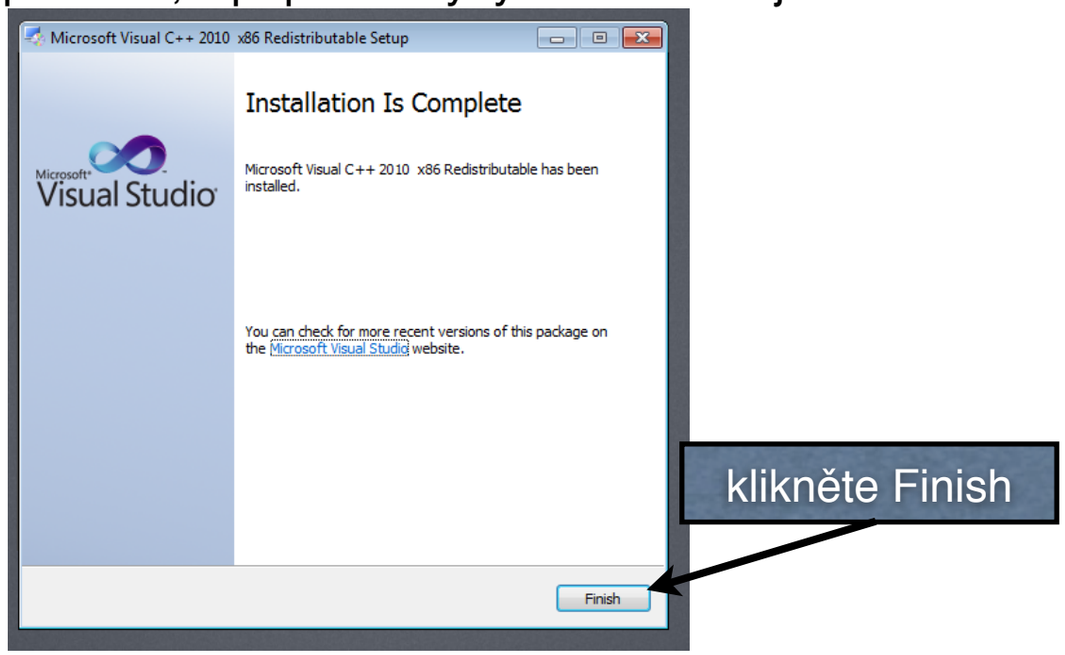
4. vytvořit odkaz v menu Start



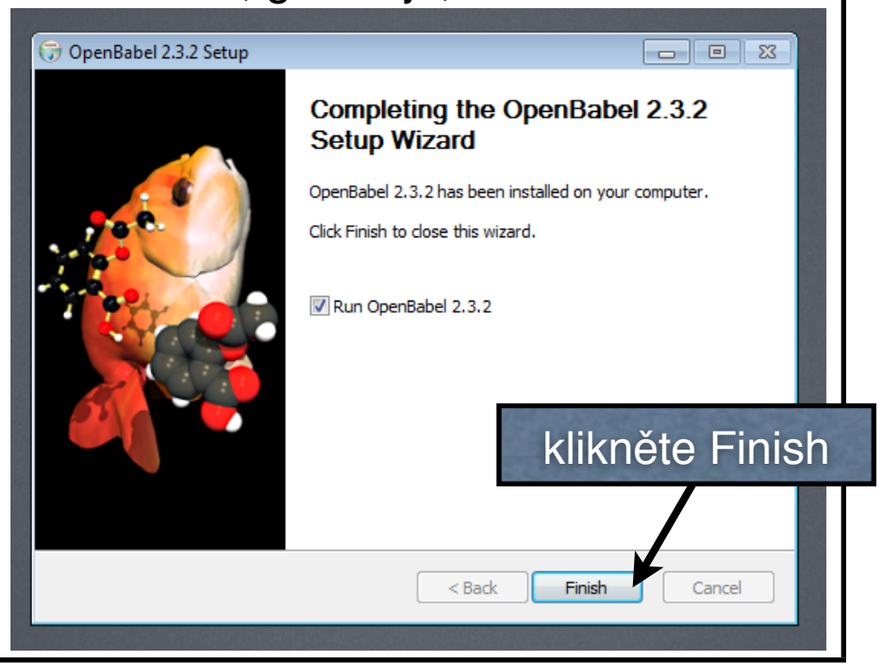
5. pokud chybí knihovny Microsoftu je nutné je doinstalovat



6. potvrďte, v případě chyby mě kontaktujte



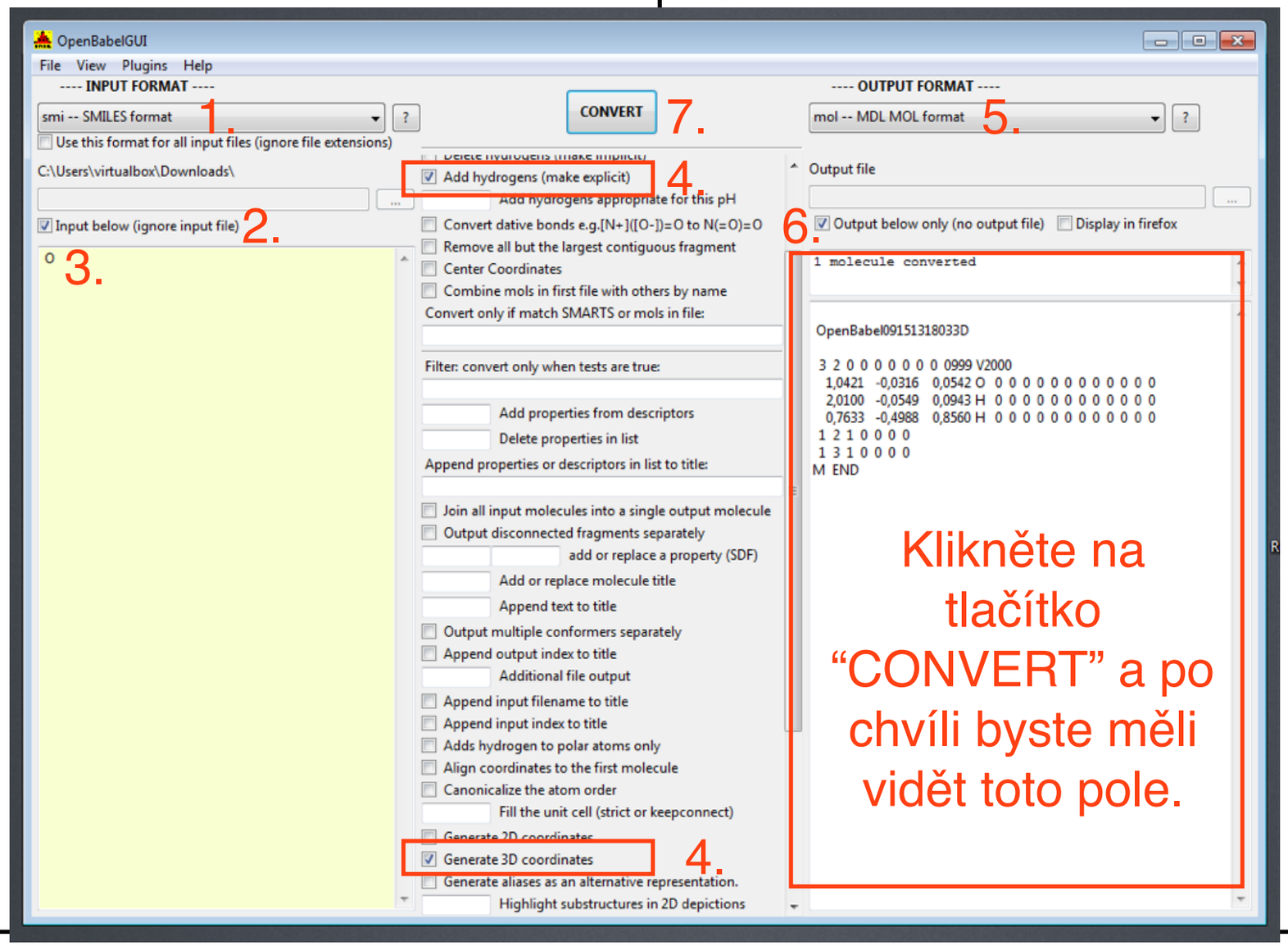
7. pokud jste se dostali až sem, gratuluju, máte naistalováno



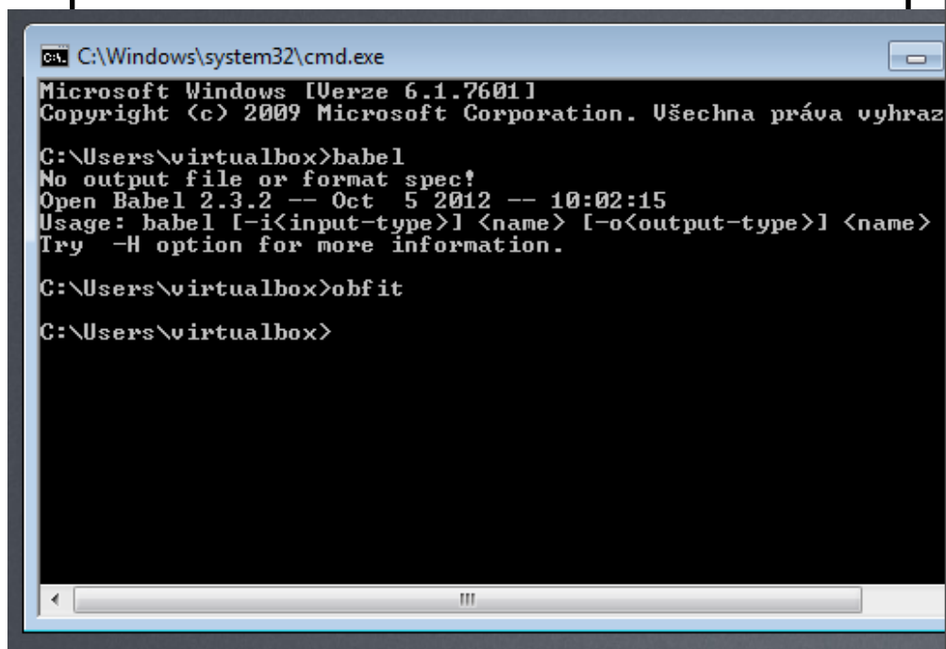
Otestování instalace

VSTUP (INPUT)
1. vyberte “smi ...”
2. zaškrtněte “Input below (ignore input file)”
3. zadejte “O”

VÝSTUP (OUTPUT)
4. zaškrtněte (přidání vodíků a tvorba 3D struktury)
5. vyberte “mol ...”
6. zaškrtněte “Output below only (no output file)”



Klikněte na tlačítko Start s zadejte “cmd” a potvrďte enterem.
Pak zkuste do příkazové řadky zadat “babel” a potvrďte enterem.
Pak zkuste do příkazové řadky zadat “obfit” a potvrďte enterem.
Výsledkem by mělo být něco podobného:



OpenBabel

- tvoříme SMILES ze stažené struktury

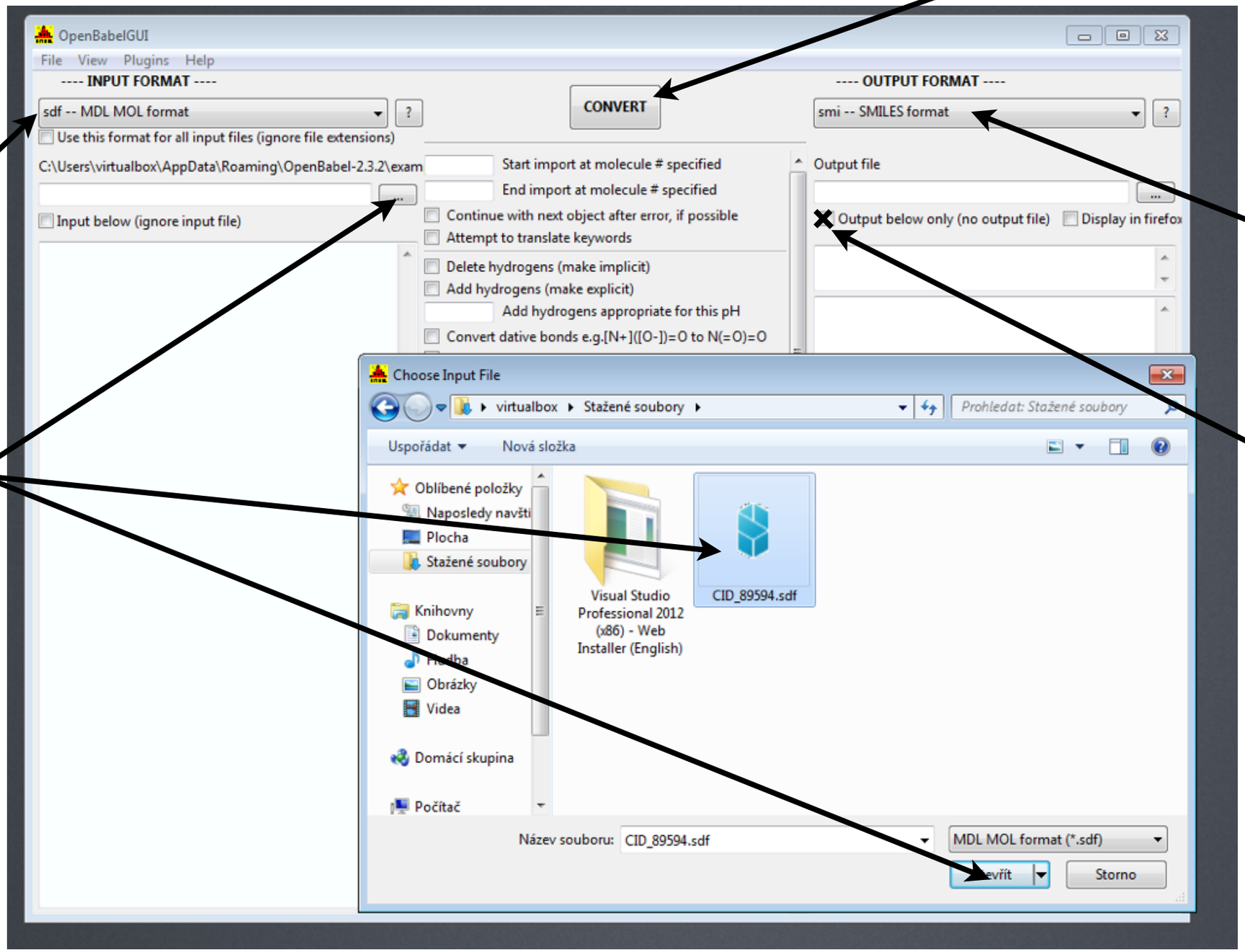
5. CONVERT

1. vybereme vstupní formát sdf

2. najdeme staženou molekulu v počítači

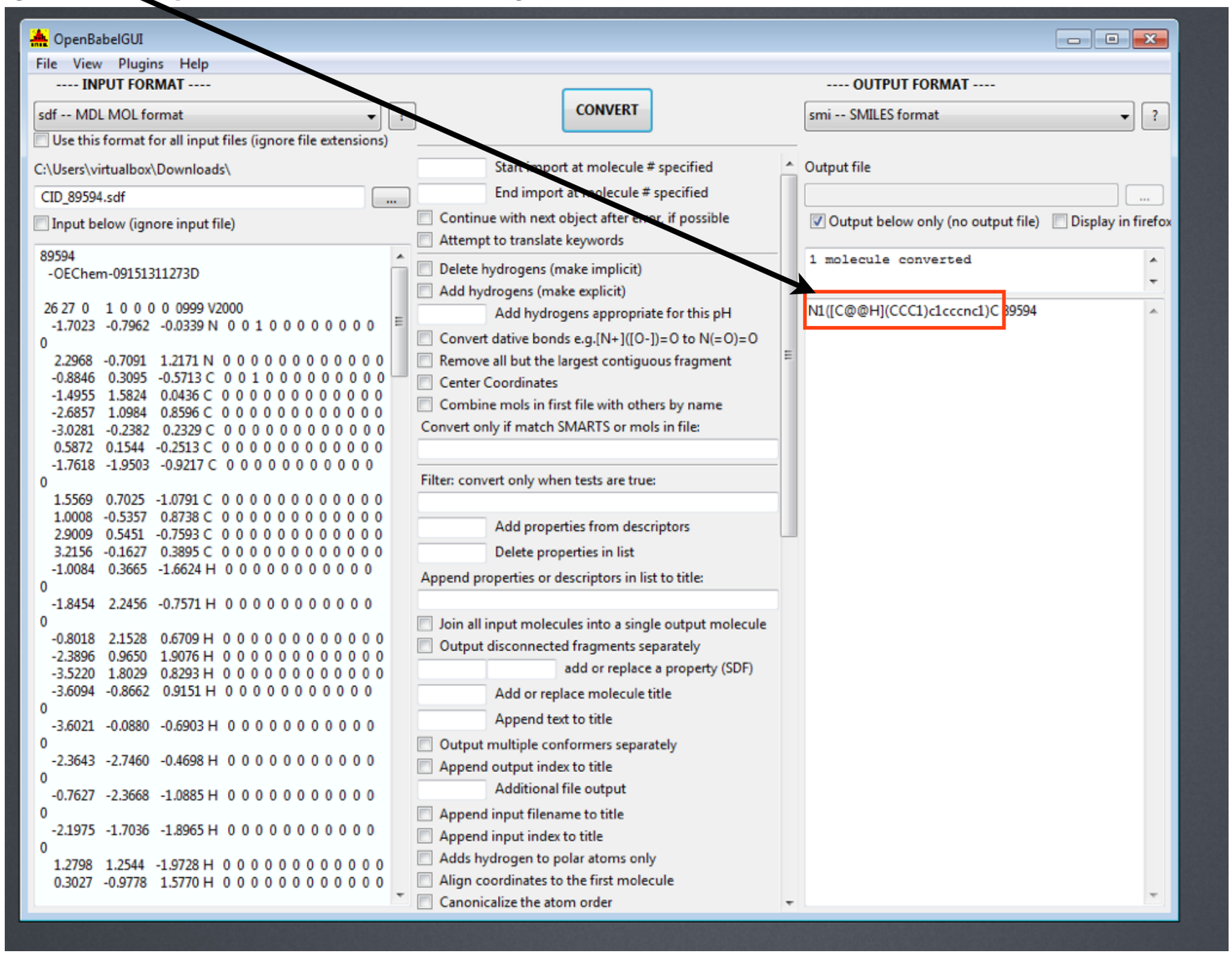
3. vybereme výstupní formát smi (SMILES format)

4. nebudeme potřebovat výstup, proto zaklikneme “Output below only”

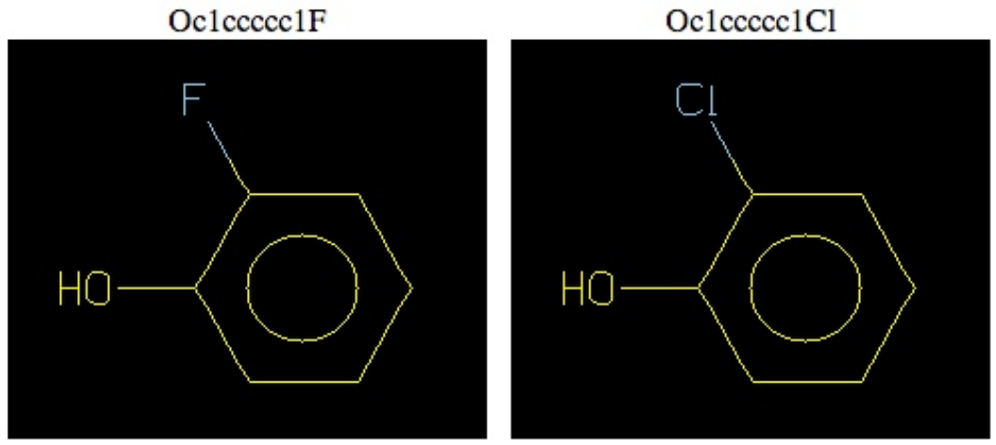


OpenBabel

- výsledek pro nikotin stažený z PubChemu



Přiložení

- nebo také zarovnání, v angličtině alignment
- o co jde? snažíme se co nejlépe přiložit dva objekty, aby si byly co nejblíže, ale nesmíme žádný z těchto objektů změnit
- co můžeme přikládat? tvary, strukturní vzorce, DNA nebo proteinové sekvence, 3D struktury
- **přiložení strukturních vzorců na základě podobnosti**
  - snažíme se přiložit na sebe dva obrázky, tak aby si co nejvíce atomů odpovídalo, například tyto dvě struktury na sebe přiložíme podle zvýrazněné žluté barvy
  - 

Přiložení

- **přiložení proteinové sekvence**

```
AAB24882        TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881        --------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
                                     ****: .***:  * *:** * :****.:* *******..

AAB24882        PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881        HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                 **** *:***********:***:**.: .****************    :  *.: :
```

  - jedná se o přiložení dvou proteinových sekvencí (posloupnost aminokyselin), kde aminokyseliny jsou zapsané v jednopísmenných zkratkách, všimněte si, že někde aminokyseliny chybí, někde jsou i rozdílné

- **přiložení 3D struktur**

  - představme si, že atomy jsou body v prostoru, snahou tohoto přiložení je to, aby si stejné atomy byly blízko, to provedeme tak, že jednu molekulu zafixujeme a s druhou molekulou (případně s ostatními) hýbeme v prostoru (rotujeme a posunujeme celou molekulou), matematicky můžeme rozdílnost popsat pomocí RMSD

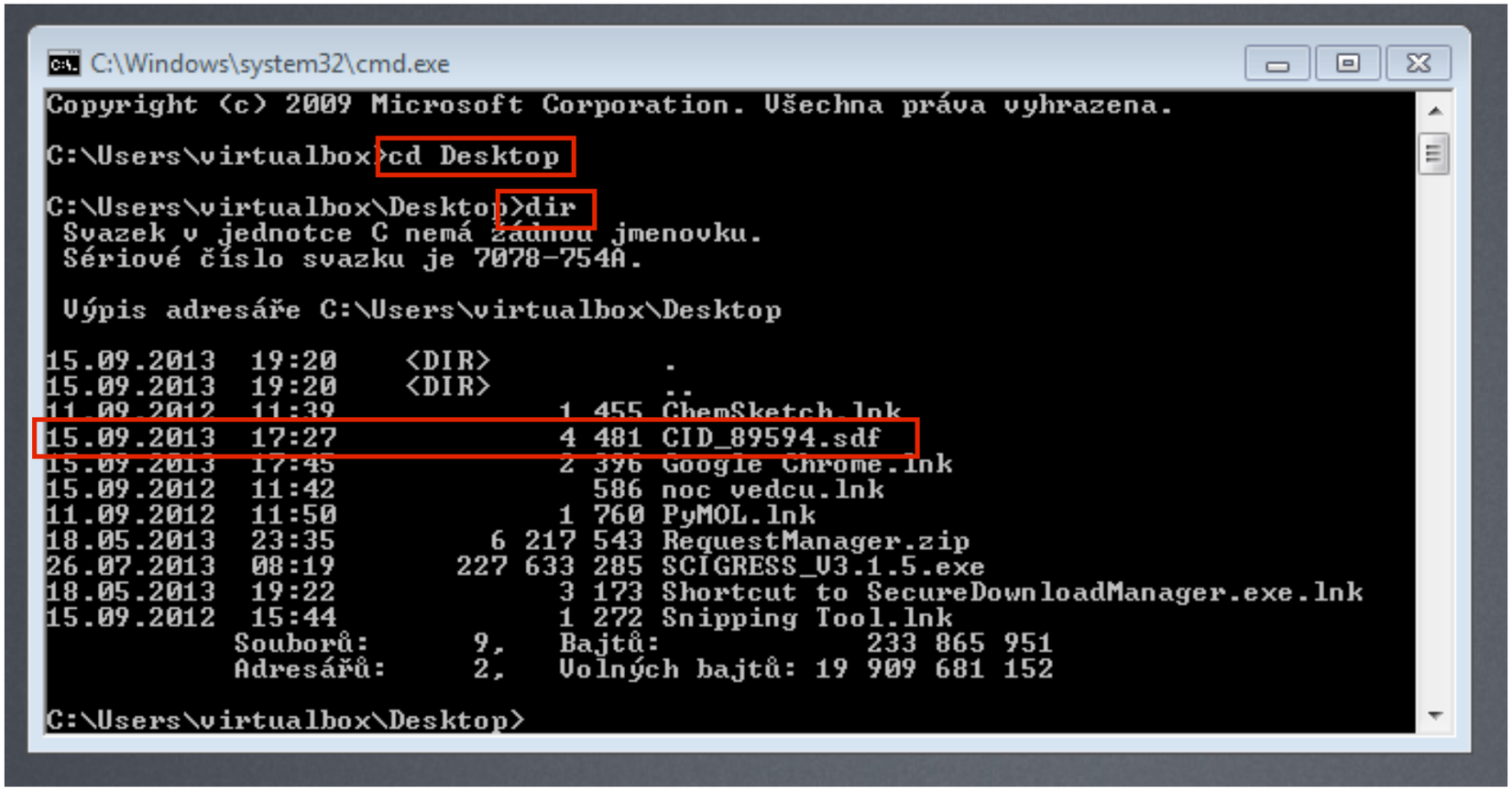
RMSD

- RMSD (root mean square deviation)

$$RMSD(x,y) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|x_i - y_i\|^2},$$

  - kde *x* je bod reprezentující molekulu X a *y* molekulu Y, *N* je počet molekul
  - takže sparujeme porovnávané atomy molekuly X a Y, projedeme všechny tyto páry a napočítáme jejich vzdálenost v prostoru $||x_i - y_i||$,
    tuto hodnotu umocníme, poté je všechny sečteme a podělíme počtem atomů, nakonec odmocníme
  - vzdálenost dvou bodu v prostoru spočítáme
    - $r = \sqrt{(x_x-y_x)^2+(x_y-y_y)^2+(x_z-y_z)^2}$

obfit

- nakopírujte si všechny struktury na Plochu
- otevřeme příkazovou řádku (Menu Start, napsat “cmd” a enter)
- zadejte “dir”, pokud máte v seznamu Desktop, zadejte “cd Desktop” nebo pokud máte v seznamu Plocha, zadejte “cd Plocha”
- zadejte příkaz “dir” a zkontrolujte, jestli máte molekuly ve složce
- 

obfit

- program obfit potřebuje celkem 3 parametry, vzor, podle kterého bude přikládat (SMILES) a dvě struktury, první zafixuje a druhou se snaží hýbat

- stáhnul jsem molekulu nikotinu z PubChemu (CID_89594.sdf) a ze ZINCu (zinc_1798.sdf), SMILES jsem použil z toho slidu, takže zadání zní:

  - `obfit “N1([C@@H](CCC1)c1cccnc1)C” CID_89594.sdf zinc_1798.sdf`

```
C:\Users\virtualbox\Desktop>obfit "N1([C@@H](CCC1)c1cccnc1)C" CID_89594.sdf zinc
_1798.sdf
RMSD: 1.528723
```

- možné výsledky:

správný výsledek	nic se nevrátilo chyba ve vzoru (SMILES), nebyl nalezen v první struktuře	nevrátilo se RMSD chyba ve vzoru (SMILES), nebyl nalezen v druhé struktuře

`M CHG 1 2 1`
`M END`
`> <RMSD>`
`1.528723`
`$$$$`
`C:\Users\virtualbox\Desktop>`
	`C:\Users\virtualbox\Desktop>obfit "N1([C`
	`c_1798.sdf`
	`C:\Users\virtualbox\Desktop>`
		`7 13 1 0 0 0 0`
		`7 14 1 0 0 0 0`
		`8 15 1 0 0 0 0`
		`M END`
		`$$$$`
		`C:\Users\virtualbox\Desktop>`