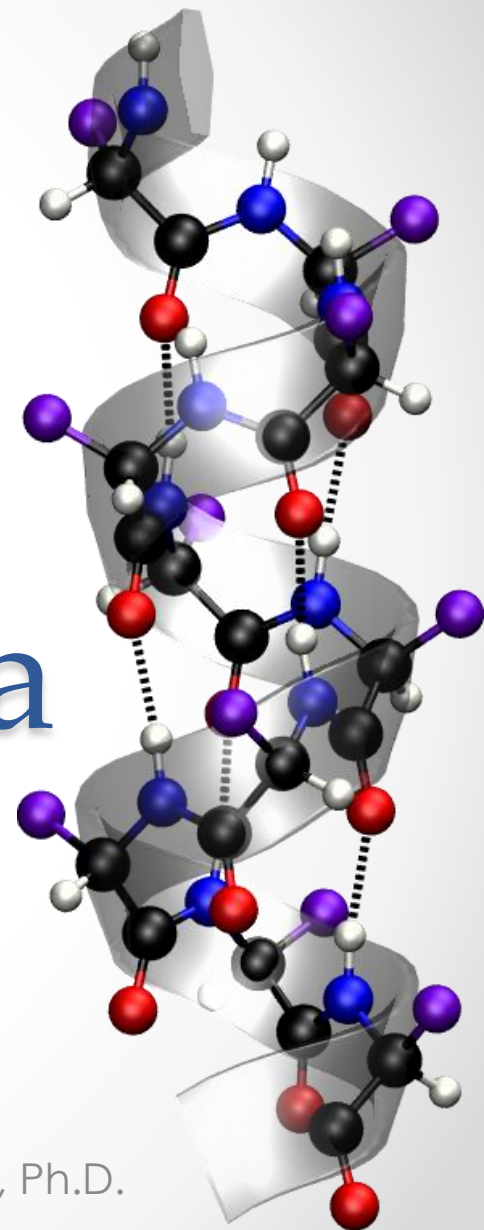
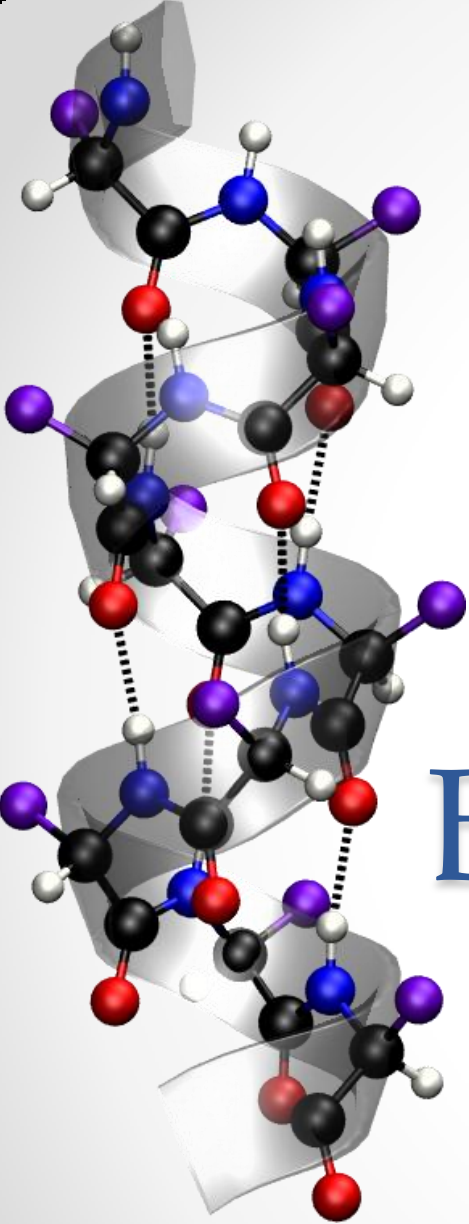




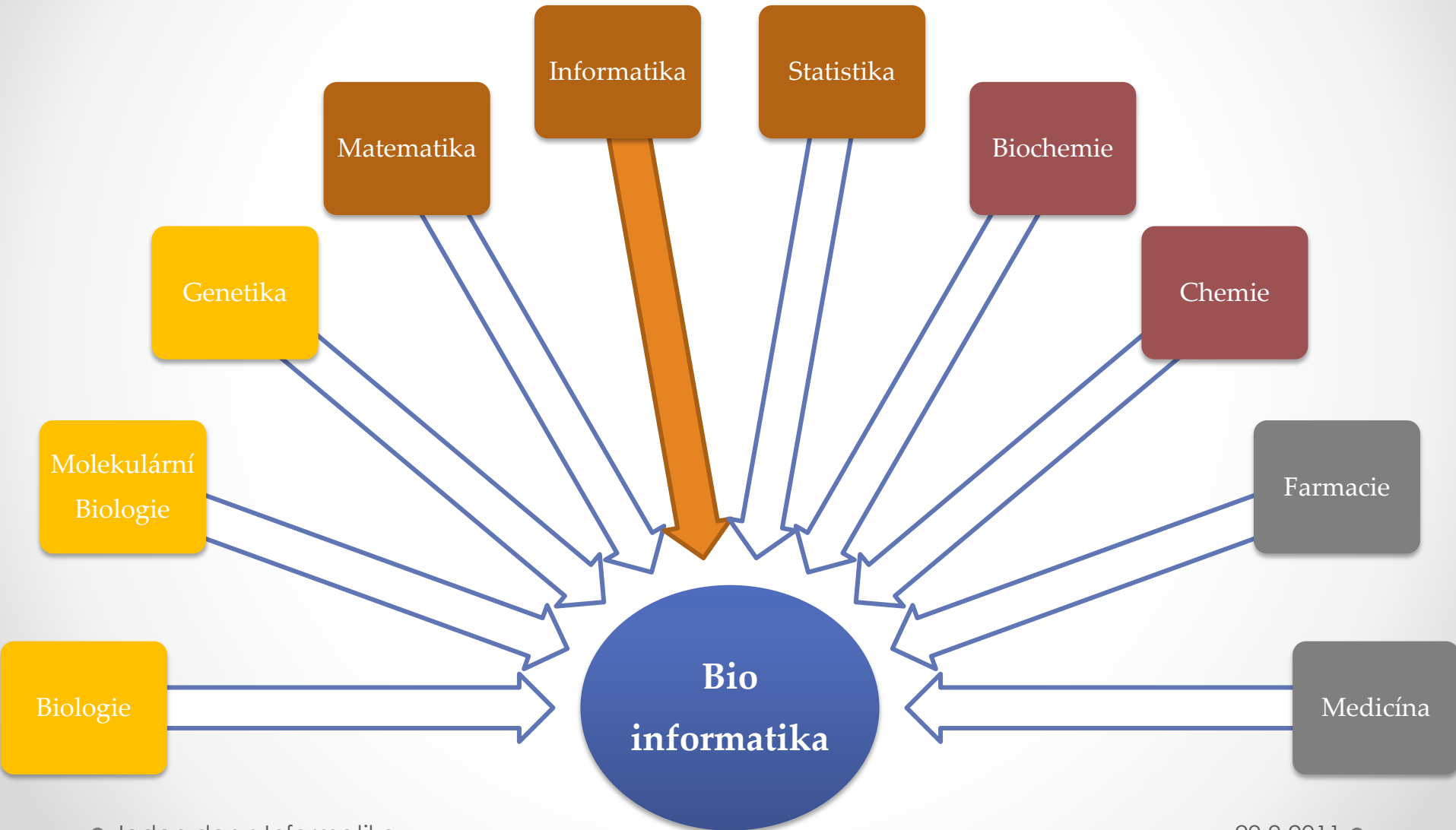
Bioinformatika

Šedá zóna na pomezí informatiky,
biologie a chemie

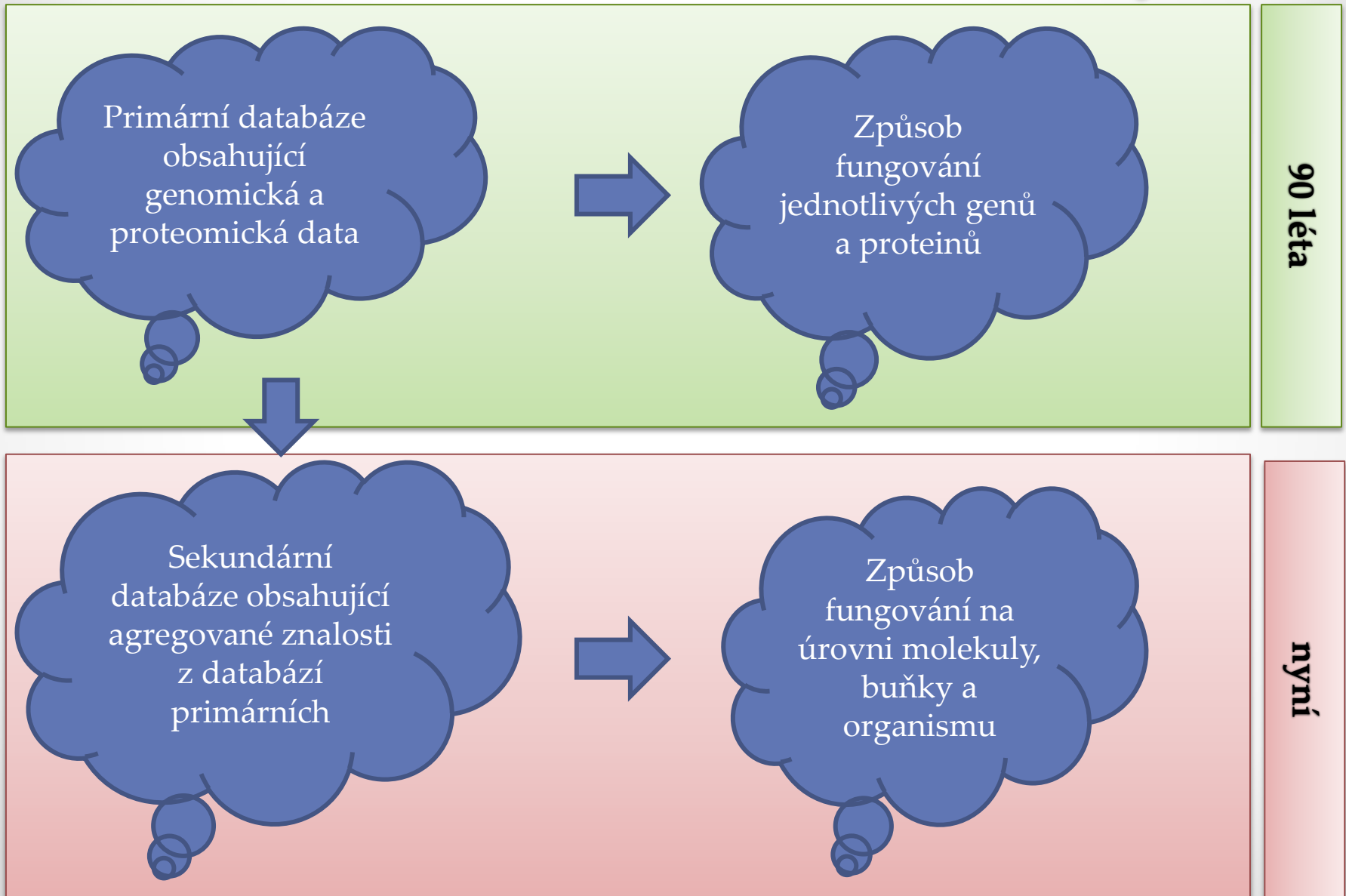
RNDr. David Hoksza, Ph.D.



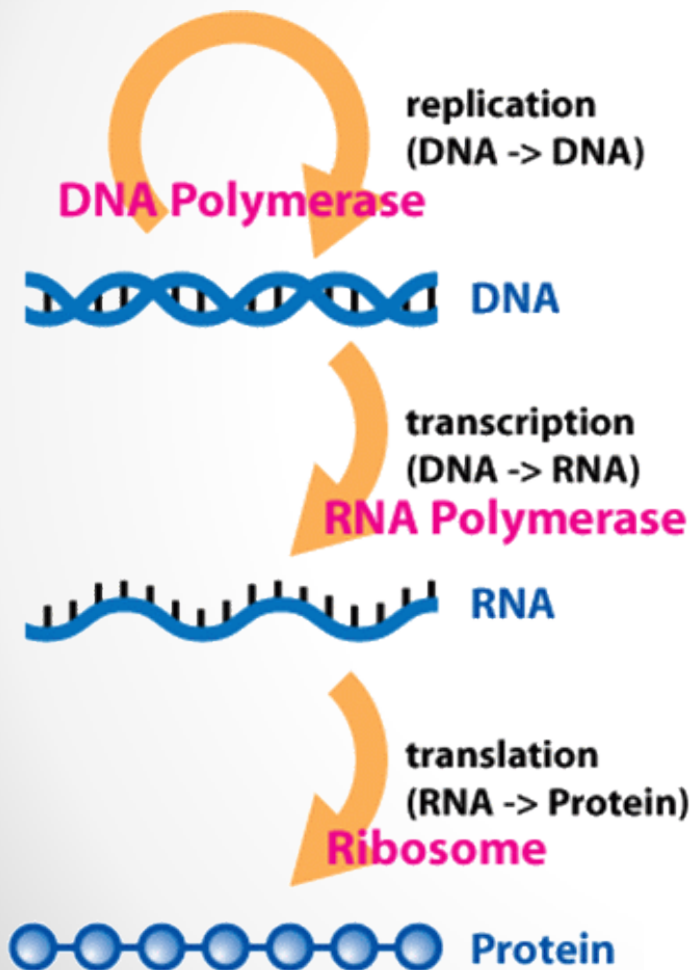
Bioinformatika



Bioinformatika dříve a nyní

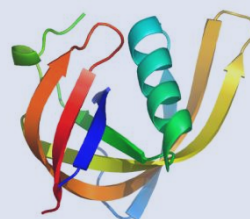
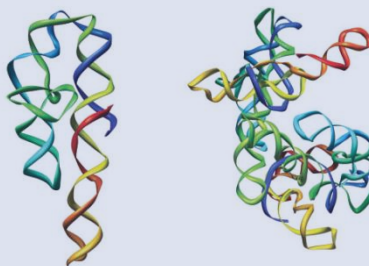
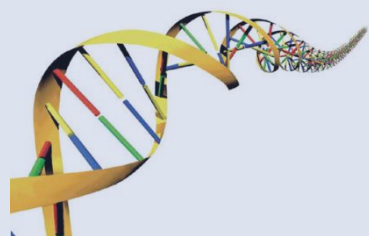


Centrální dogma molekulární biologie



• Jeden den s Informatikou

struktura



sekvence

ID AB000263 standard; RNA; PRI; 368 BP.
 XX
 AC AB000263;
 XX
 DE Homo sapiens mRNA for prepro cortistatin like peptide, complete cds.
 XX
 SQ Sequence 368 BP;

```
acaagatgcc attgtccccc ggctctctgc tgcctgctct ctcggggccc acggccaccg    60
ctgccctgcc ctggagggtt ggccccaccg gccgagacag cgagcatatg caggaagcgg    120
caggaataag gaaaagcagc ctctgacttc tctctgcttg gtggtttgag tggacctccc    180
aggccagtgc cgggcccctc ataggagagg aagctcggga ggtggccagg cggcaggaag    240
gcgcaccccc ccgcaatcc gcgcgccggg acagaatgcc ctgcaggaac ttctctgga    300
agaccttctc ctctgcaaa taaaacctca cccatgaatg ctaacgcaag ttaattaca    360
gacctgaa                                     368
```

//

DSSP

PDB G P G S C P T H A D S L N N L A N I K R E Q G N I E E A V R L Y R K A L E V F P E F A A A H S N

PDB 315 320 330 340 350

DSSP

PDB S N M G N T L K E M Q D V Q G A L Q C Y T R A I Q I N P A F A D A H S N L A S I H K D S G N I F

PDB 389 400 410 420 430

DSSP

PDB W T D Y D E R M K K L V S I V A D Q L E K N R L P S V H P H S M L Y P L S H G F R K A I A E F

PDB 469 480 490 DSSP: no secondary structure assigned

DSSP

PDB G Y V S S D F G N H P T S H L M Q S I P G M H N P D K F E V F C Y A L S P D D G T N F R V K V M

PDB 549 560 570 580 590

DSSP

PDB M N G Y T K G A R N E L F A L R P A P I Q A M W L G Y P G T S G A L F M D Y I I T D Q E T S P A

PDB 629 640 650 660 670

DSSP

PDB AV I D F K S N G H I Y D N R I V L N G I D L K A F L D S L P D V K I V K M K C P D G G D N A E

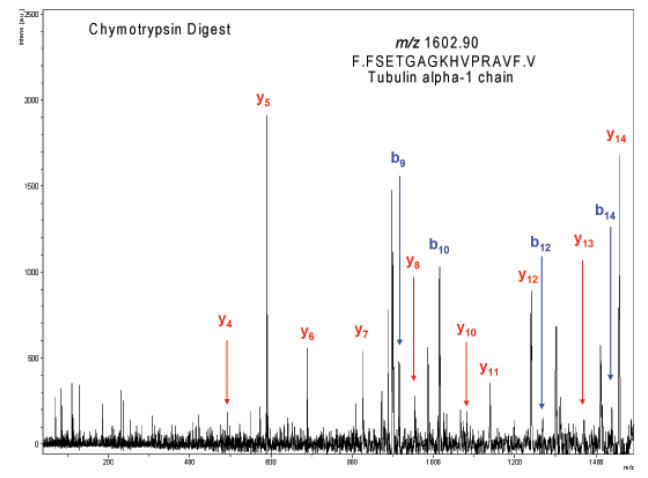
Sekvenční Bioinformatika

- Sekvenování genomu
 - Human Genom Project (HGP)
- Identifikace proteinové sekvence
 - **Hmotnostní spektrometrie**
- Analýza DNA/RNA sekvence
 - Hledání genů
 - Hledání opakujících se motivů
 - Hledání strukturních motivů
- Fylogenetika
 - Kreslení “stromu života”
- ...



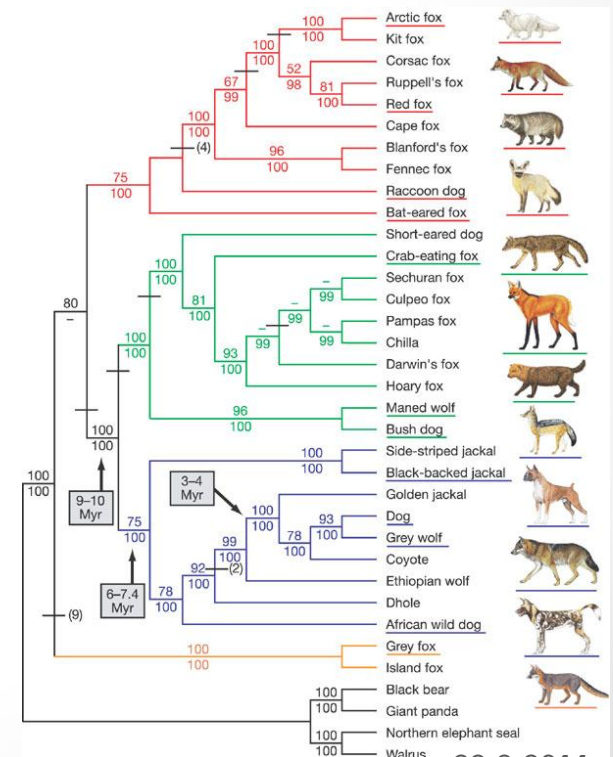
Sekvenční Bioinformatika

- Sekvenování genomu
 - Human Genom Project (HGP)
- Identifikace proteinové sekvence
 - **Hmotnostní spektrometrie**
- Analýza DNA/RNA sekvence
 - Hledání genů
 - Hledání opakujících se motivů
 - Hledání strukturních motivů
- Fylogenetika
 - Kreslení “stromu života”
- ...



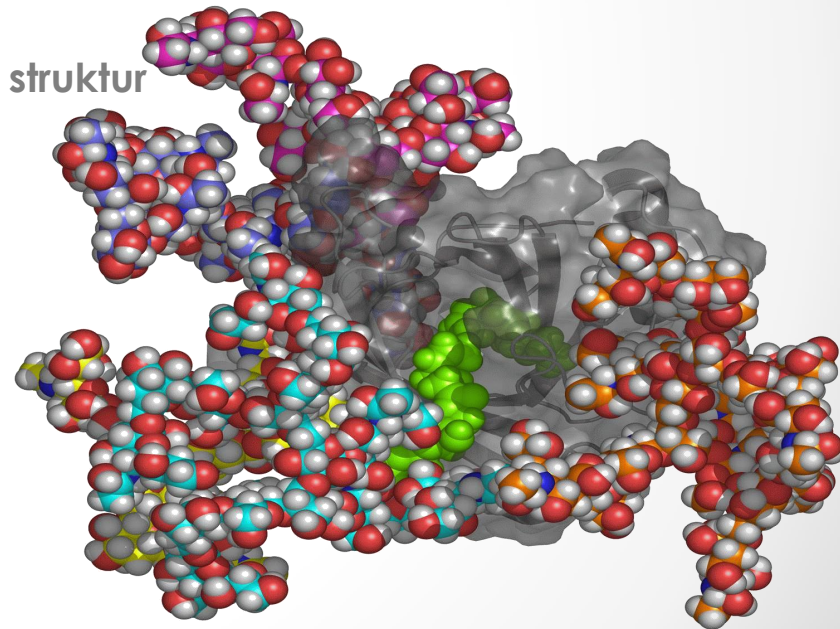
Sekvenční Bioinformatika

- Sekvenování genomu
 - Human Genom Project (HGP)
- Identifikace proteinové sekvence
 - **Hmotnostní spektrometrie**
- Analýza DNA/RNA sekvence
 - Hledání genů
 - Hledání opakujících se motivů
 - Hledání strukturních motivů
- **Fylogenetika**
 - Kreslení “stromu života”
- ...



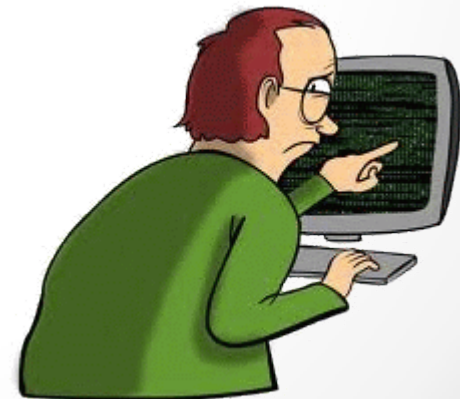
Strukturní bioinformatika

- Strukturně podobné molekuly mají podobnou biologickou funkci
- Identifikace funkce RNA a proteinů
 - **Výzkum podobnosti proteinových a RNA struktur**
 - *Výzkum interakcí protein-protein*
 - *Výzkum interakcí protein-ligand*
- Predikce proteinové struktury
- Molekulární modelování
- ...



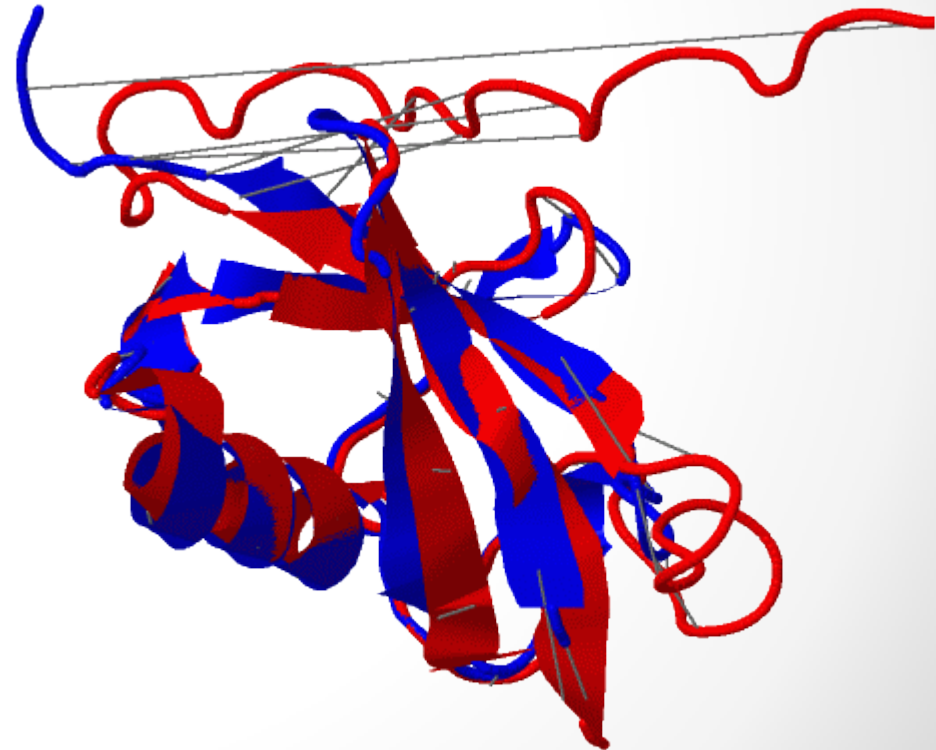
2 pohledy na bioinformatiku

- Bioinformatik – biolog/uživatel
 - Uživatel bioinformatických algoritmů a nástrojů, tj. hledání v databázích, identifikace podobných struktur, ...
- Bioinformatik – informatik/vývojář
 - Vývoj bioinformatických algoritmů a nástrojů

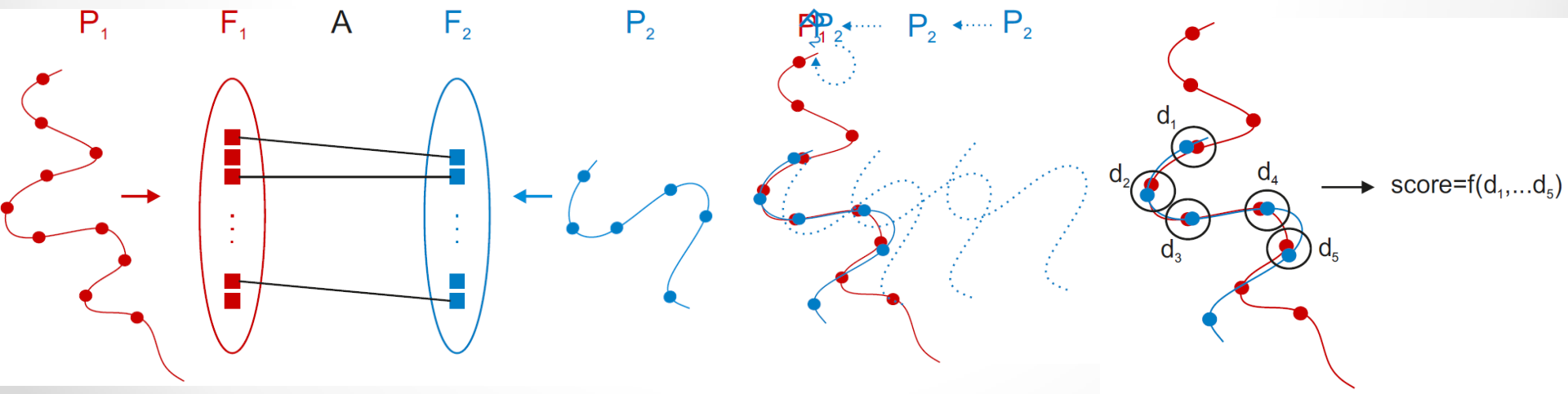
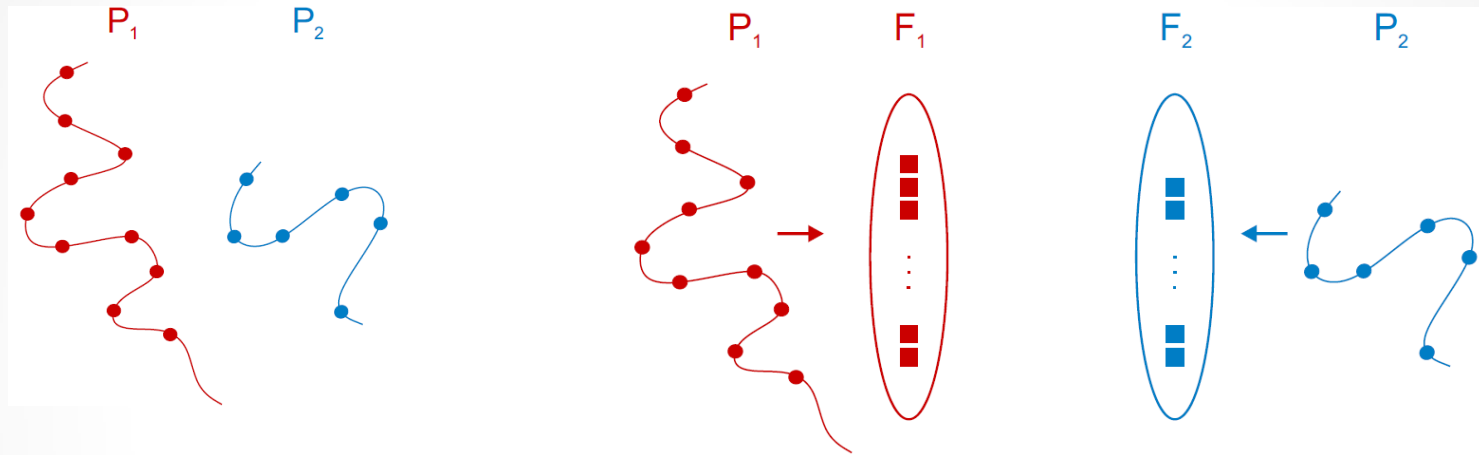


Podobnost proteinových struktur

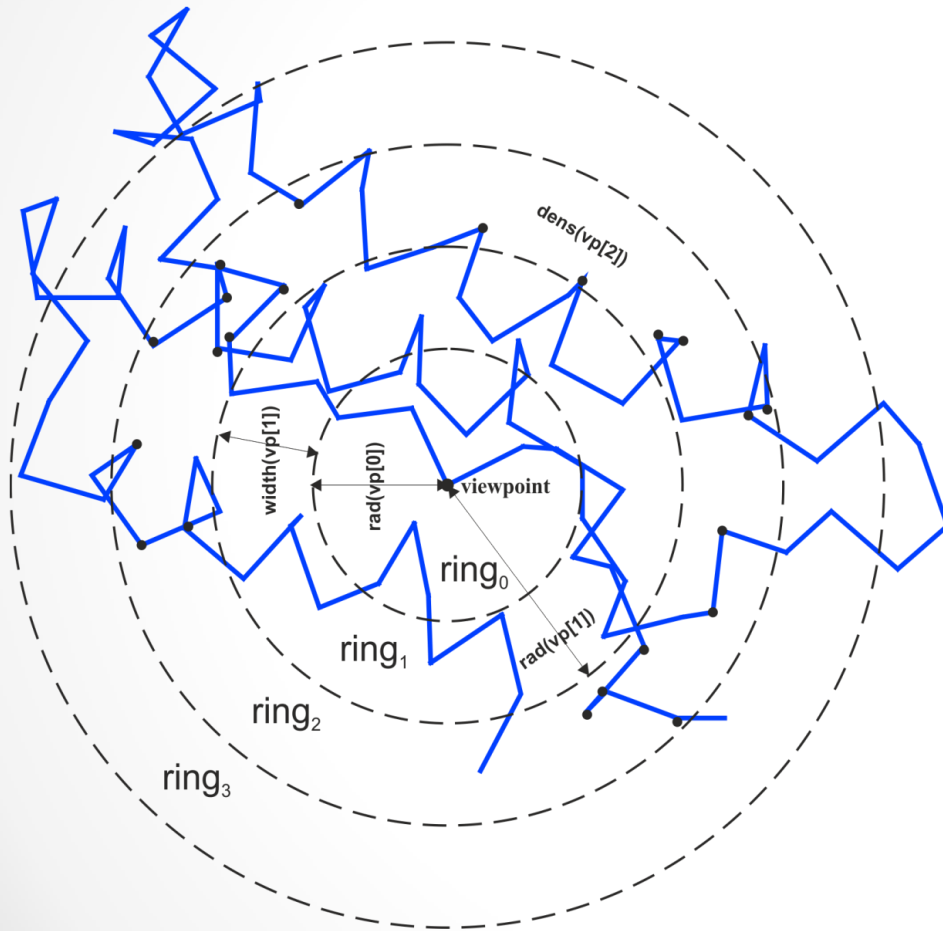
- Motivace
 - Funkční analýza proteinů
 - Predikce proteinové struktury
- Algoritmus
 - Extrakce vlastností
 - Zarovnání
 - 3D superpozice
 - Aplikace vzdálenostní funkce



Podobnost proteinových struktur



Extrakce vlastností



- upstream backbone
- downstream backbone
- upstream neighborhood
- downstream neighborhood

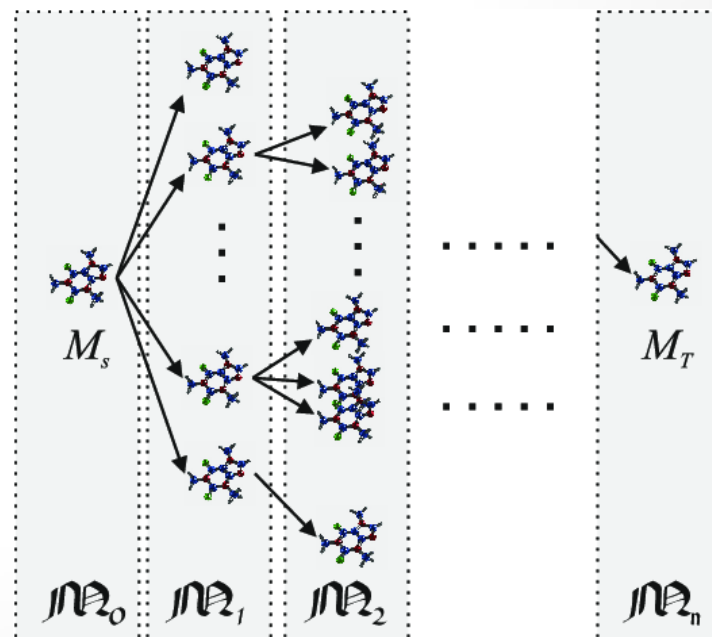
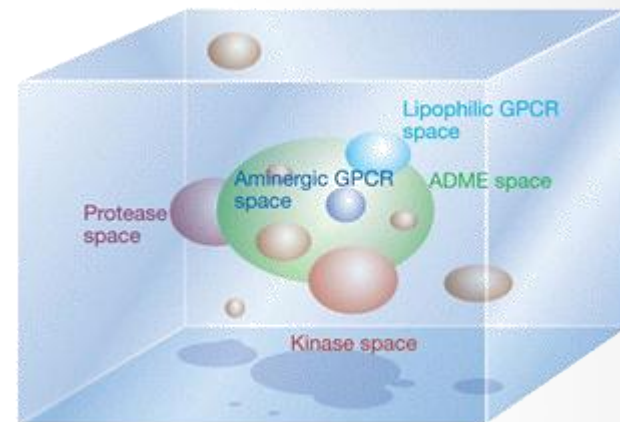
<11;29;23;19;16>

Cheminformatika

- Vědní disciplína úzce spjatá s vývojem léčiv pomocí počítačových metod
 - relativně uzavřená disciplína
- Cena nového léku je v průměru \$100,000,000
- Oblasti
 - Vytěžování znalostí z chemických databází
 - Generování virtuálních knihoven
 - Virtuální screening
 - ...

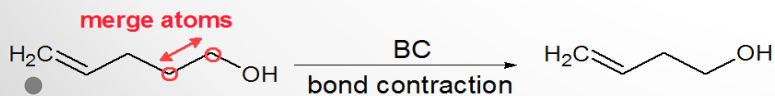
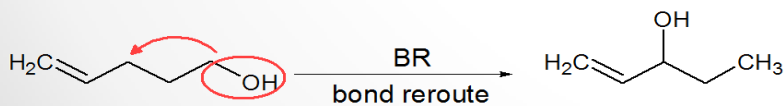
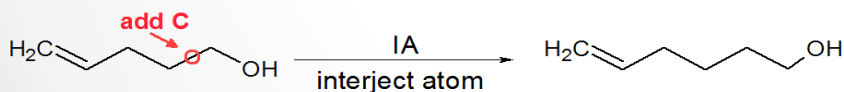
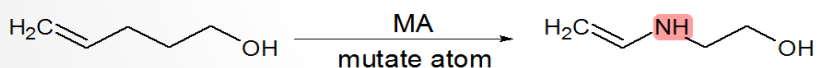
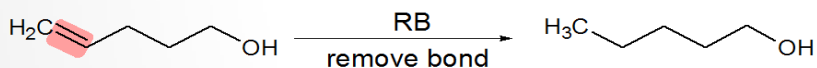
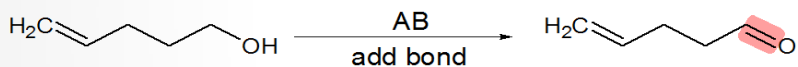
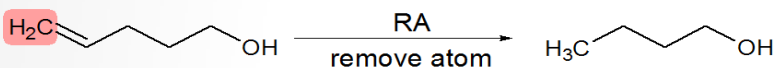
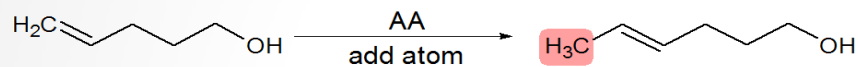
Hledání cesty v chemickém prostoru

- Motivace pro procházení chemického prostoru
 - hledání nových léčiv s předpokládanými vlastnostmi
 - virtual screening
- Algoritmus
 - Procházení prostoru s přibližováním se k cíli
- Výzvy
 - Procházení prostoru
 - “morfovací” operátory
 - Vzdálenost mezi molekulami
 - otisky molekul
 - Vizualizace prostoru
 - Dimenze prostoru
 - Sloučenin o velikosti až 13 s atomů C, N, O, S, Cl je ~ 970 milionů
 - Paralelizace

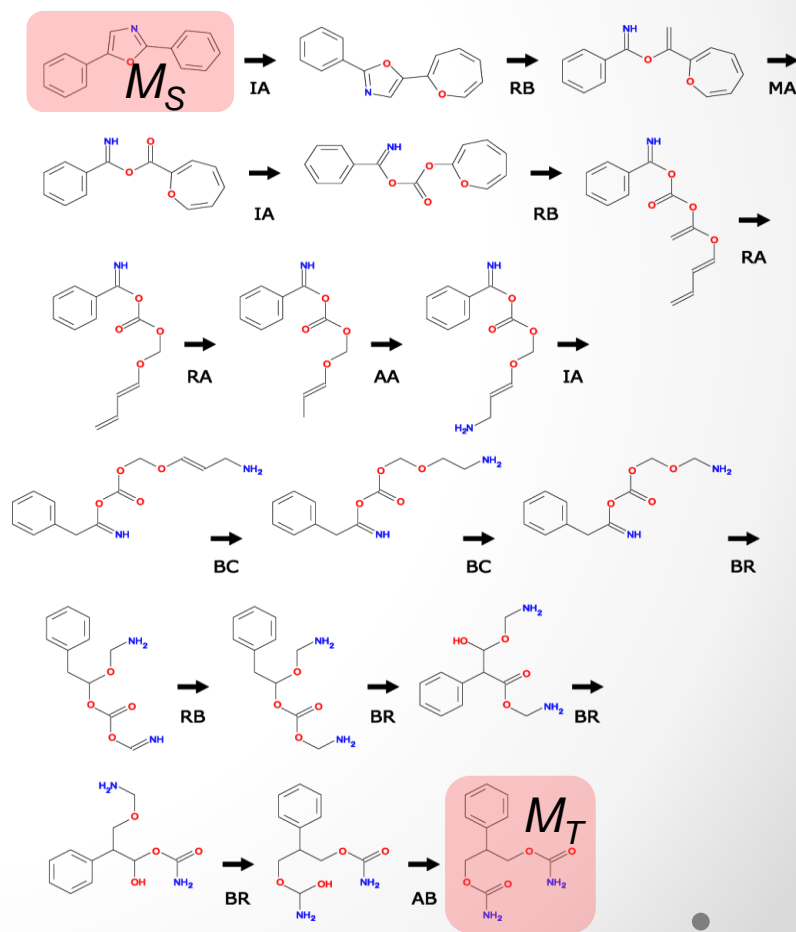


Hledání cest v chemickém prostoru

morfovací operátory

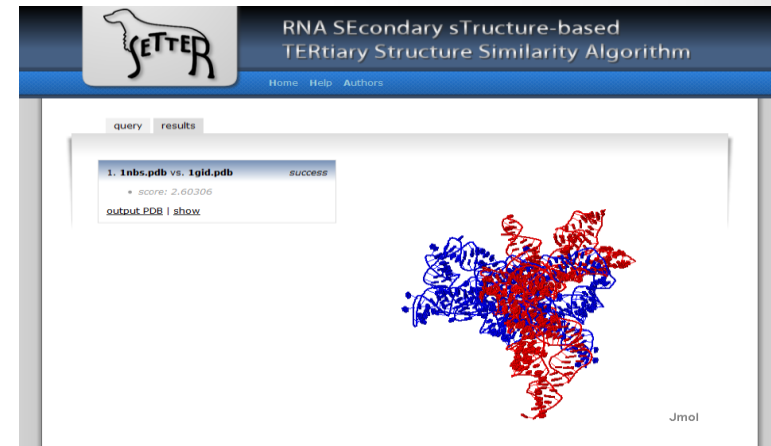


příklad cesty v prostoru



Veřejně přístupné bioinformatické aplikace vyvinuté na KSI MFF UK

- Strukturní podobnost na úrovni proteinů a RNA
 - <http://siret.cz/p3s/>
 - <http://siret.cz/setter/>
- Identifikace proteinových sekvencí
 - <http://siret.cz/simtandem>



SimTandem.org

General settings

Search title [?](#)

Database [?](#)

Protein sequences digestion enzyme [?](#)

Max. missed cleavage sites [?](#)

Modifications [?](#)

Max. modifications per spectrum [?](#)

Data file [?](#)

My search

SwissProt

Trypsin

0

+57 Carbamidomethyl, H (3) C (2) NO

+28 Lys->Arg, H2

+16 Oxidation, O

+14 H (2) C

-14 H (-2) C (-1)

Add [?](#)

Select *.mcf file to upload...

Advanced database settings

Advanced query settings

Report settings

Search

Demo search [?](#)

- Jeden den s Informatikou

SimTandem is a tool for protein and peptide sequences identification from tandem mass spectra based on similarity search in the protein sequence database. SimTandem works with spectra in [Mascot Generic Format \(*.mcf\)](#). You can download sample files [here](#).

SimTandem is free for non-commercial use. If you use SimTandem in your work, please cite us using the following publication: J. Novák, T. Skopal, D. Hoksza, J. Lokoč: **Improving the Similarity Search of Tandem Mass Spectra using Metric Access Methods**, SISAP 2010, Istanbul, Turkey, ISBN 978-1-4503-0420-7, pp. 85-92.

Otázky?

