



Komplexné biologické problémy a hľadanie ich riešenia využitím techník a nástrojov informatiky

RNDr. Gabriel Žoldák, PhD.

- Výskumný pracovník, Centrum Interdisciplinárnych Biovied, Technologický a inovačný park UPJŠ
- Spoluzakladateľ DEG Bioresearch s.r.o.
- Spoluzakladateľ DNA Healthcare s.r.o.

Biologické systémy sú organizované hierarchicky



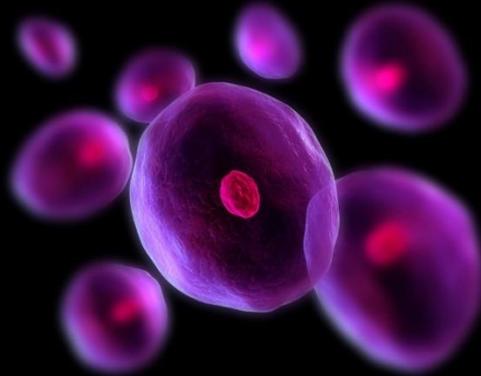
človek



m



10^{13} buniek



10^{-6} m



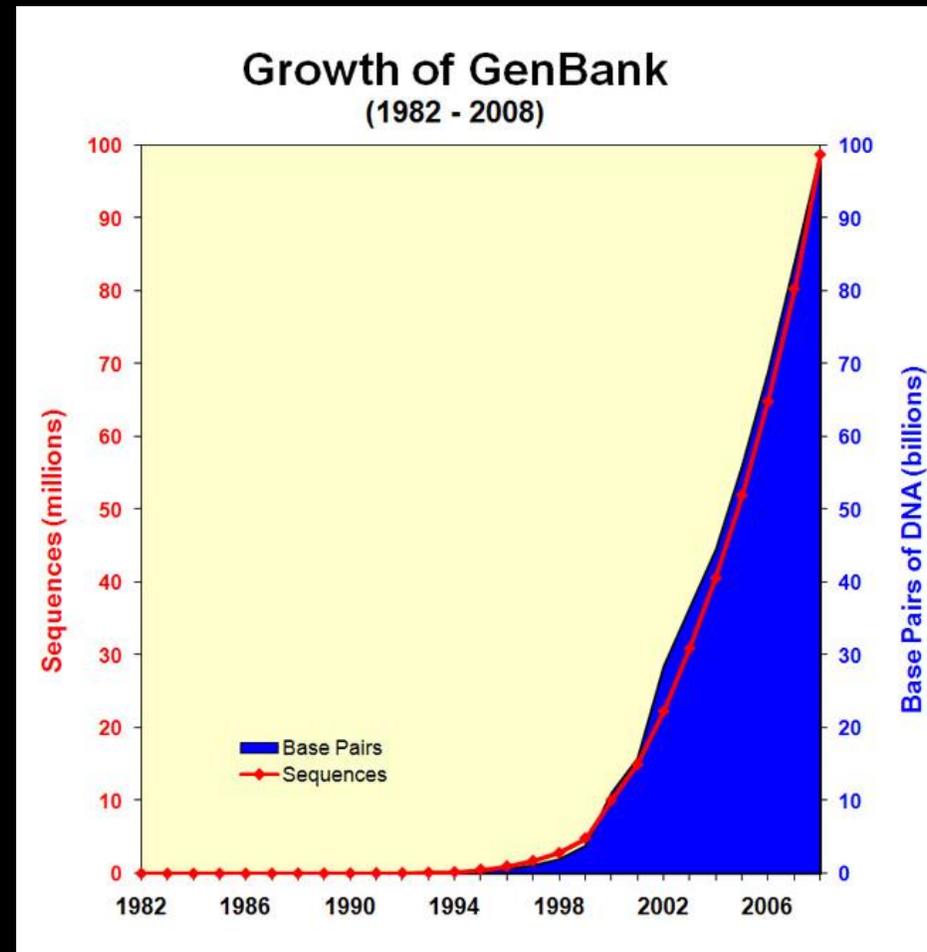
1 bunka: 3×10^9 báz v DNA,
20 tisíc génov



10^{-9} m

Explózia informácií o biologických systémoch

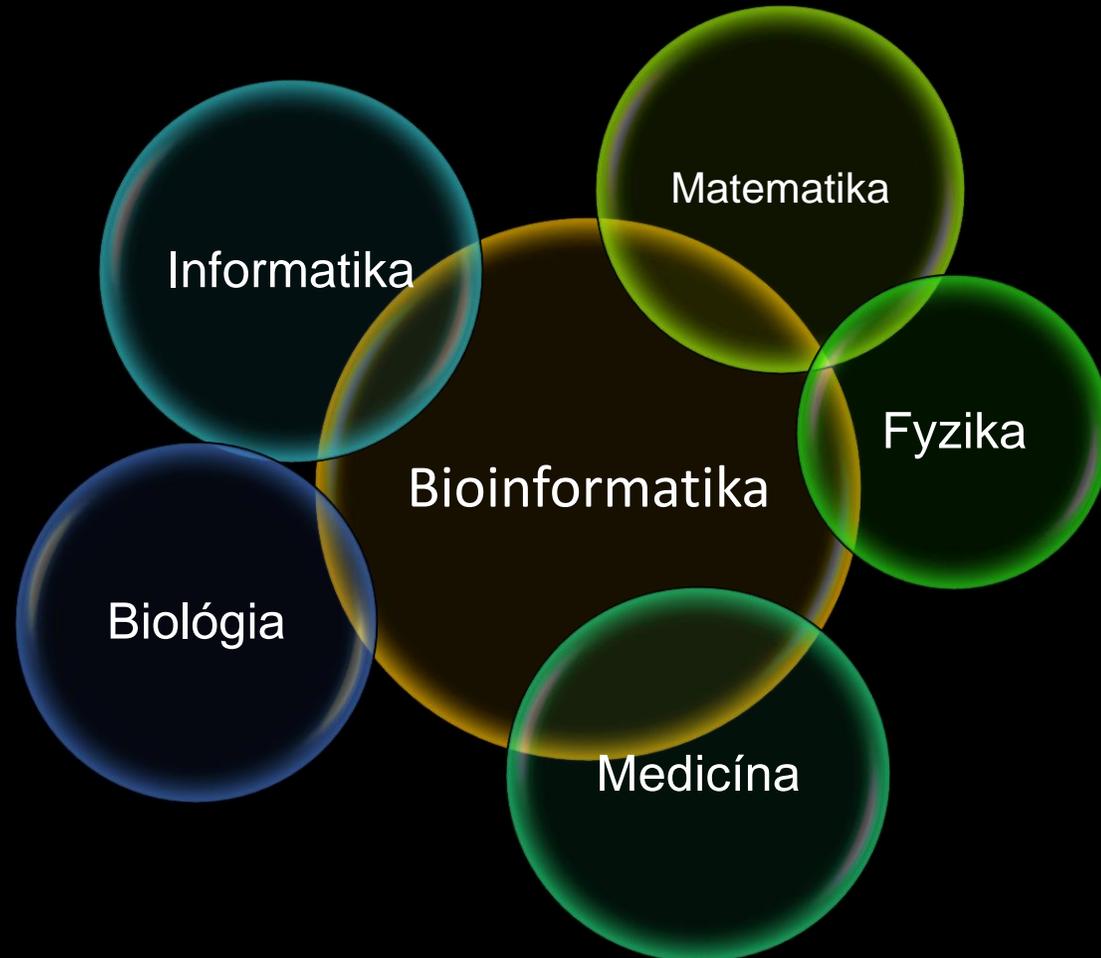
Množstvo dostupnej informácie o biologických objektoch presiahlo analytickú kapacitu ľudstva



Čo je to bioinformatika?

Bioinformatika je interdisciplinárny vedecký odbor

Definícia (pôvodná): Paulien Hogeweg a Ben Hesper: *"the study of informatic processes in biotic systems"*



Bioinformatika je interdisciplinárny vedecký odbor

Techniky a nástroje v bioinformatike:

Uchovávanie biologickej informácie

Databázy

- Tvorba rozhrania
- Prehľadávanie
- Kategorizácia
- Anotácia
- Citácia
- Validácia dát
- Aktualizácia

Interoperabilita

Štandardizácia

Komunikácia

Optimalizácia funkčných procesov

Výskum a Analýzy

Analýza textového reťazca

- Hľadanie textu
- Štatistické analýzy

Hľadanie vzorov

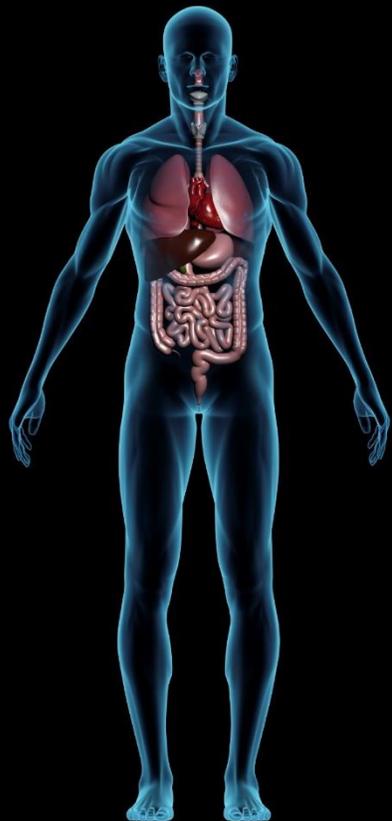
- AI/ machine learning
- Klastrovanie
- Data mining

Geometrické a fyzikálne modely

Má bioinformatika praktické využitie?

Bioinformatika má široké praktické uplatnenie

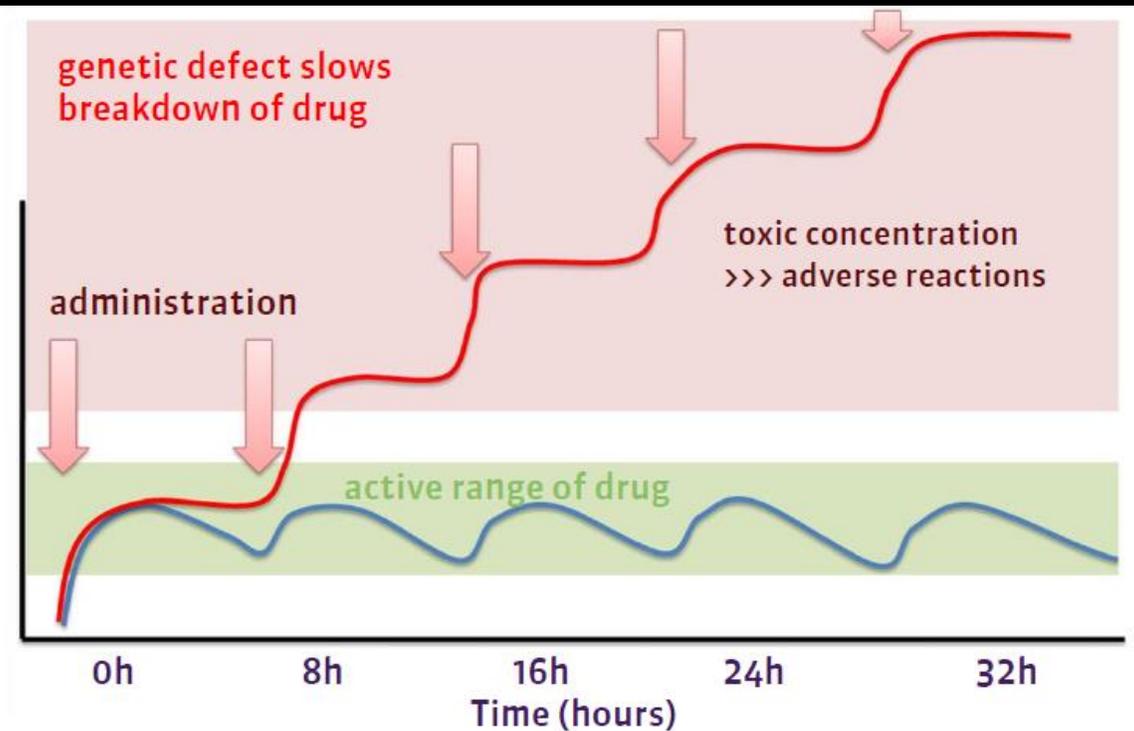
Molekulárna medicína: identifikácia genetickej variability: dávkovanie warfarínu na riedenie krvi.



3-4 z 20 tisíc génov



Drug in blood



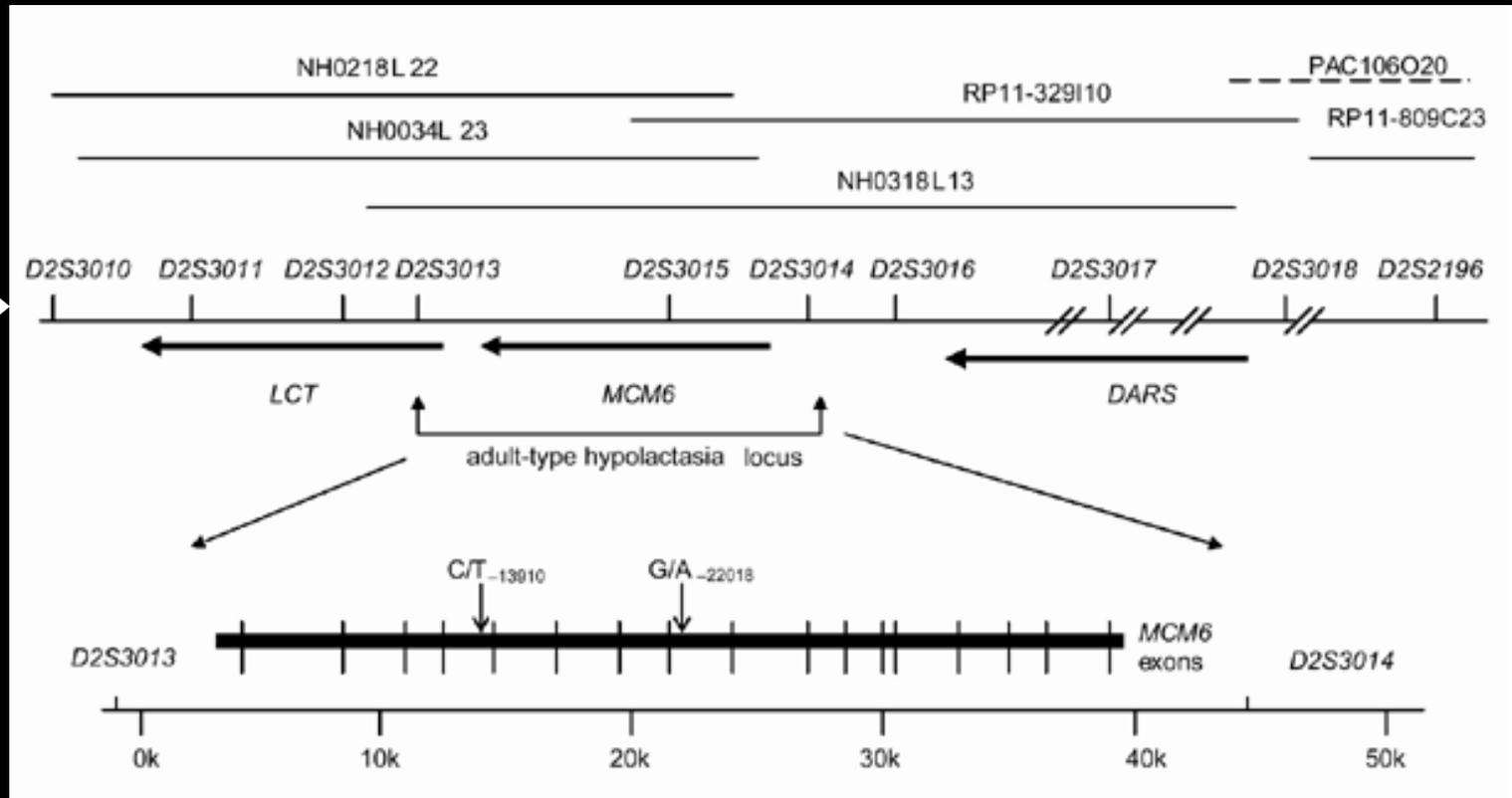
Bioinformatika má široké praktické uplatnenie

Molekulárny detektív odhalenie generácie X-men: laktózová tolerancia v dospelosti

Pred viac ako
10 000 rokov



1 z 20 tisíc génov



Bioinformatika má široké praktické uplatnenie

Bioinformatické analýzy poskytujú racionálny základ molekulárnej medicíny

● Včasná detekcia a genetická predispozícia k chorobám

● Génová terapia

● Spresnená diagnostika chorôb

● Farmakogenetika

- Individualizované liečivá
- Špecifický výber liečiv
- Určenie množstva/dávky liekov
- Obmedzenie vedľajších účinkov

Bioinformatika má široké praktické uplatnenie

(1) Analýza rizík a ich znižovanie

- Čističky vôd
- Znižovanie pravdepodobnosti dedičných mutácií

(2) Zdroje energie a prostredie

- Biopalivá
- Detekcia polutantov
- Zmena klímy

(3) Poľnohospodárstvo

- Úroda, rezistencia voči suchu
- Zvyšovanie nutričnej hodnoty
- Biopesticídy

(4) Forézne analýzy

- DNA
- Hmotnostná spektroskopia

Bioinformatika má široké praktické uplatnenie

PROTEINOVÉ INŽINIERSTVO

(1) Analýza rizík a ich znižovanie

- Čističky vôd
- Znižovanie pravdepodobnosti dedičných mutácií

(2) Zdroje energie a prostredie

- Biopalivá
- Detekcia polutantov
- Zmena klímy

(3) Poľnohospodárstvo

- Úroda, rezistencia voči suchu
- Zvyšovanie nutričnej hodnoty
- Biopesticídy

(4) Forézne analýzy

- DNA
- Hmotnostná spektroskopia

Bioinformatika má široké praktické uplatnenie

Cielená modifikácia vlastností biologických objektov: proteínové inžinierstvo

Návrhy riešenia

Bioinformatika

- Databázy
- Modelovanie
- Big data

Spôsob dosiahnutia

Testovanie
nových
molekúl

- Experimentálny
biochemický a
biofyzikálny
prístup

Dosiahnutý efekt

Príprava nových katalyzátorov

- Príprava biopalív
- Syntéza nových zlúčenin
- Odbúravanje mikroplastov

Zlepšovanie úrodnosti

- Biodegradovateľné pesticídy
- Stratégie zvyšovanie odolnosti voči suchu

Ako sa využívajú nástroje informatiky pri riešení
biologických problémov?

Príklad 1. Molekulová evolúcia proteínov

Príklad 2. Molekulárna brána ako regulátor enzýmovej aktivity

Príklad 3. Analýza komplexných sietí metabolický dráh

Príklad 4. Jurský park na molekulovej úrovni: rekonštrukcia pravekých enzýmov

Príklad 5. Biofarmaceutický priemysel: predikcia rozpustnosti proteínov

Bioinformatika ako nástroj pochopenia biologických objektov

Príklad 1. Molekulová evolúcia proteínov

doc. Ing. Martin Tomášek PhD.
Bc. Samuel Pristiak

RNDr. Gabriel Žoldák PhD
Bc. Michal Gala



Fakulta elektrotechniky
a informatiky



Katedra
počítačov
a informatiky



Technologický
a inovačný park
UPJŠ



CENTER FOR
INTERDISCIPLINARY
BIOSCIENCES

Bioinformatika ako nástroj pochopenia biologických objektov

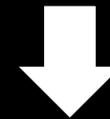
Príklad 1. Molekulová evolúcia proteínov

Bc. Michal Gala:

Diplomová práca – Fylogenéza bakteriálnych stresových proteínov a ich úloha v adaptácii na extrémne podmienky.

- 244 baktériálnych Hsp70 proteínov
- Bioinformatická analýza
- Konzervované/variabilné oblasti

```
1 MAPQVPPLLA FPLQPPTMDS PVQIFRGEPE PTCAPRACQL PNSSAWFPGW AEPDGNASA
2 METPVQIFRG EPGPTCAPSA CLLPNSSAWF PGWAESGNGS AGSEEAQLEP AHISPAIPV
3 MESPVQIFRG EPGPTCAPST CLLPNDSGWF PGWAEPDGNS SAGSEDALE PAHISPAIP
  MESPVQIFRG EPGPTCAPSA CLLPNSSAWF PGWAESGNGS AGSEEAQLEP AHISPAIPV
  MESPVQIFRG EPGPTCAPST CLLPNDSGWF PGWAEPDGNS SAGSEDALE PAHISPAIP
  MESPVQIFRG DPGPTCAPSA CLLPNSSAWF PGWAEQDND SASSEPAQLE PAHISPAIP
  MMESPVQIFR GEPGPTCAPS ACLLPNGSTW FPDWAEPSK GNGTASAEDA QLETAHISP
  MESPVQIFRG EPGPTCAPST CLLPNDSGWF PGWAEPDGNS SAGSEDALE PAHISPAIP
  MDSPVQIFRG EPGPTCAPRA CLLPNSSAWF SGWAESDSNG SASPEDEPLE PAHISPAIP
  MESPVQIFRG EPGPTCAPSA CLLPNSSAWF PGWAESGNGS SAGSEDALE PAHISPAIP
```



Sekvenčné
zarovnanie

```
TNPHNTLYAVKRLIG-RRFDEDVVQKDI-QMVPYKIVKADNNDWAVEV-----N-G-----QK-MA
LNSERTIKSIKRMG-SDY-----K-VRI-----D-D-----KE-YT
TNPNTLFAIKRLIG-RRFDDKVVQKDI-GMVPYKIAKADNGDAWVEI-----N-G-----KK-LA
TNPKNITFSIKRFMG-RFY--DEVTEEI-SMVPYKIVRGENNTRVEVEVGG-----KRLYT
TNPRTIYAVKRLIG-RNFTDPMVRKDQ-GLVPYNIKADNGDAWVEA-----D-N-----HK-YS
TNVDRTISSVKRHMGT-DW-----TTEI-----D-G-----KK-YT
TNPRTVYSIKRFMG-MRYNDVKDERTS-----Y-NV-SANASGDVQVKI-----E-D-----KE-YS
TNPRTVYSIKRFMG-RRYND--VKQEA-ERVG YQVWPGPHGDVVRVRL-----G-D-----KD-YS
```

Príklad 1. Molekulová evolúcia proteínov

Evolúcia Hsp70 proteínov: vývoj nových nástrojov na analýzu lokálnej mechanickej plasticity štruktúry

doc. Ing. Martin Tomášek PhD.
Bc. Samuel Pristiak

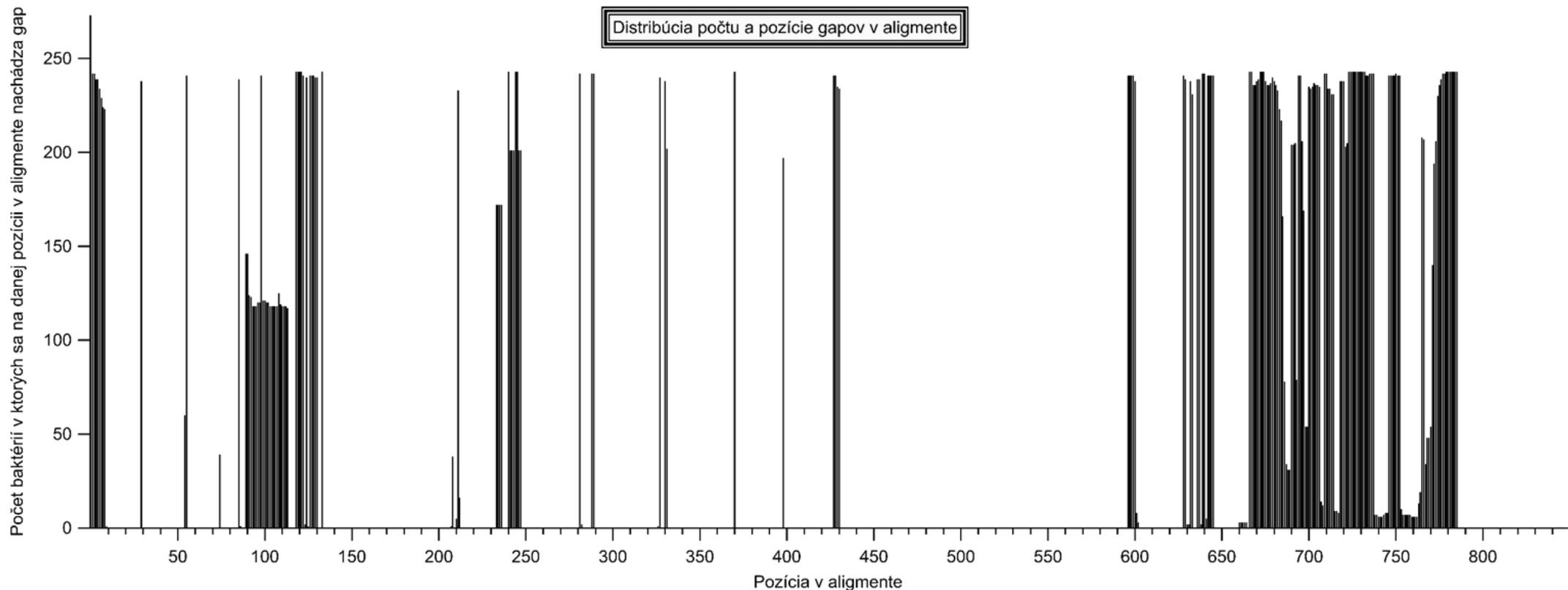
-----MGKIIIGIDLTTNSCVSILE-NGNVKVLNAEGARTTPSIIAYAND-GEILVGSQAKRQAVTNPNTLYAVKRLIG-RRFDEEDVVQKDI-QMVPYKIVKADNNDAAWVEV-----N-G-----QK-MAPPQISAEVLKMKKTAEDYLGEVPTAEVITVPAYFNDSQRQATKDAGRIAGLDVVKRIINEPTAAALAYGMDKAKG-DHTVIVYDLGGGTFDVSIEIEAENVG
-----MSDRFVVGIDLTTNSVIAWMPKPGSIEVIPNAEGSRTPSIVAFKTK-GEILVGEPAKRQMIILNERTIKSIKRRMG-SDY-----KVRI-----D-D-----KE-YTPQOISAYILMKMKKDAEYLGKIKKAVITCPAYFNDAQRQATKEAGQIAGFEVLRINEPTAAALAYGLDKKKE--QKVLVYDLGGGTFDVSILEI----S
-----MGKVIIGIDLTTNSCVAVME-GDKVKVIENAEGTRTTPSIVAYKN--DEILVGSQAKRQAVTNPNTLFAIKRLIG-RRFDDKVVQKDI-GMVPYKIAKADNGDAWVEI-----N-G-----KK-LAPPQVSAEILKMKKTAEDYLGETVTEAVVTPAYFNDSQRQATKDAGRIAGLDVVKRIINEPTAAALAYGMDKQKQ-DSTVAVYDLGGGTFDVSIEIEADVDG
-----MGKIIIGIDLTTNSVAVME-GGEPKVIINPEGRVTPSVVAFAD-GEPLVGAPAKRQAITNPKNTIFSIRKRFMG-RFY--DEVTEEI-SMVPYKVVIRGENNTVRVEVEVGGE-----KRLYTPQEISAMILQKLKQTAEEYLVGPVTEAVITVPAYFNDAQRQATKEAGEIAGLVKRRILNEPTAAALAYGLDKKDK-ELKIAVYDLGGGTYDISILEL----G
-----MGKVIIGIDLTTNSCVAVME-GKEPKVIDNAEGERTTPSIIAFAN--SERLVGQPAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----HK-YSPSQISAFILQKMKETAENYLGEKVTQAVITVPAYFNDAQRQATKDAGRIAGLEVLRIINEPTAAALAYGFEKSSS--KTIAVYDLGGGTFDVSILEI----S
-----MARAVGIDLTTNSVSTLE-GGPEVIPNAEGARTTPSVVAFSKD-GDVLVGSVAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----TTEI-----D-G-----KK-YTPQEISARILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKEAGEIAGLVNLRIVNEPTAAALAYGLDKGKE-DELILVFDLGGGTFDVSILLEV----G
-----MGKSIIGIDLTTNSCVAVLE-GGDPVVIQNSEGERTTSPVAFDRKSGERLVGQLARRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YSPPEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGKSIIGIDLTTNSCVAVLE-GGDPVVIQNSEGERTTSPVAFDRKSGERLVGQLARRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YSPPEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MSKIIIGIDLTTNSCVAVME-GGESKVIIPNPEGNRTAPSVAFAFN--GERQVGEVAKRQAITNPN-TVQSIKRRMG-TDY-----KVEI-----E-G-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLEVLRIINEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGKVIIGIDLTTNSVAVME-GDDPEVIENAEGSRTPSVVAYKDD-GERLVGAPAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----TTEI-----D-G-----KK-YTPQEISAVLQKLKQTAEDYLGEVPTAEVITVPAYFNDAQRQATKEAGEIAGLVNLRIVNEPTAAALAYGLDDES--DQWVAVYDLGGGTFDVSILEI----G
-----MSKIIIGIDLTTNSCVAIMD-GTQARVLENAEGDRTPSIIAYTQD-GETLVGQPAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGKIIIGIDLTTNSCVAIMD-GAKARVLENAEGDRTPSIIAYTQD-GETLVGQPAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGKIIIGIDLTTNSCVAIMD-GTTPRVLLENAEGDRTPSIIAYTQD-GETLVGQPAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MAKVIIGIDLTTNSCVAVME-GGPKVVIENAEGARTTPSIVAFKDK-GERLVGQPAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MSKIIIGIDLTTNSCVAVME-GGPKVVIENAEGARTTPSIVAFKDK-GERLVGQPAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGKVIIGIDLTTNSCVSILE-GDEPKVIQNPPEGARTTPSVVAFKFN--GETQVGEVAKRQAITNPN-TVQSIKRRMG-TDY-----KVDI-----E-G-----KS-YTPQELSAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLEVLRIINEPTAAALAYGLDKTEK-DEKILVFDLGGGTFDVSILEI----G
-----MGKIIIGIDLTTNSCVAIMD-GGKARVLENAEGDRTPSIVAYTKD-GEVLVGSQAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MSKIIIGIDLTTNSAVAVLE-GTESKIIANPEGNRTTPSVVAFKFN--GEIIVGDAAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGKVIIGIDLTTNSCVAVME-GGQPVVIPNPEGRVTPSVVAFKFN--GEIIVGDAAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGKVIIGIDLTTNSCVAVME-GGQPVVIPNPEGRVTPSVVAFKFN--GEIIVGDAAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MGRVIGIDLTTNSVMAVIE-GGEPVVIQNSEGERTTSPVAFKFN--GEIIVGDAAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MSDRFVVGIDLTTNSVIAWMPKPGSIEVIPNAEGSRTPSIVAFKTK-GEILVGEPAKRQMIILNERTIKSIKRRMG-SDY-----KVRI-----D-D-----KE-YTPQOISAFILMKMKKDAEYLGKIKKAVITCPAYFNDAQRQATKEAGQIAGFEVLRINEPTAAALAYGLDKKKE--QKVLVYDLGGGTFDVSILEI----S
-----MAEKKEFVVGIDLTTNSVIAWMPKPGTVEVIPNAEGSRVTPSVVAFKTKS-GEILVGEPAKRQMIILNERTIKSIKRRMG-TDY-----KVRI-----D-D-----KE-YTPQEISAFILMKMKKDAEYLGKIKKAVITCPAYFNDAQRQATKEAGQIAGFEVLRINEPTAAALAYGLDKKAG-EEKVLVYDLGGGTFDVSILEI----G
-----MAEKKEFVVGIDLTTNSVIAWMPKPGTVEVIPNAEGSRITPSVAFKTKS-GEILVGEPAKRQMIILNERTIKSIKRRMG-TDY-----KVRI-----D-D-----KE-YTPQEISAFILMKMKKDAEYLGKIKKAVITCPAYFNDAQRQATKEAGQIAGFEVLRINEPTAAALAYGLDKKAG-EEKVLVYDLGGGTFDVSILEI----G
-----MLVRRGEVMAKAVGIDLTTNSVIAVME-GGKPVVLENAEGERTTPSVVAFKFN--GEIIVGDAAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MAKAVGIDLTTNSVIAVME-GGKPVVLENAEGERTTPSVVAFKFN--GEIIVGDAAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G
-----MAKAVGIDLTTNSVIAVLE-GGKPVVLENAEGERTTPSVVAFKFN--GEIIVGDAAKRQAVTNPNTIYAVKRLIG-RNFDPMPVRKQD-GLVPYNIIVKADNGDAWVEA-----D-N-----YNVSANASGDVQVKI-----E-D-----KE-YTPQEISAMILQKLKNDAEAYLNDKVTDAVITVPAYFNDAQRQATKDAGRIAGLVNLRIVNEPTAAALAYGLDK-EN-DQTLVFDLGGGTFDVSILEI----G

Bioinformatika ako nástroj pochopenia biologických objektov

Príklad 1. Molekulová evolúcia proteínov

Evolúcia Hsp70 proteínov: vývoj nových nástrojov na analýzu lokálnej mechanickej plasticity štruktúry

doc. Ing. Martin Tomášek PhD.
Bc. Samuel Pristiak



Bioinformatika ako nástroj pochopenia biologických objektov

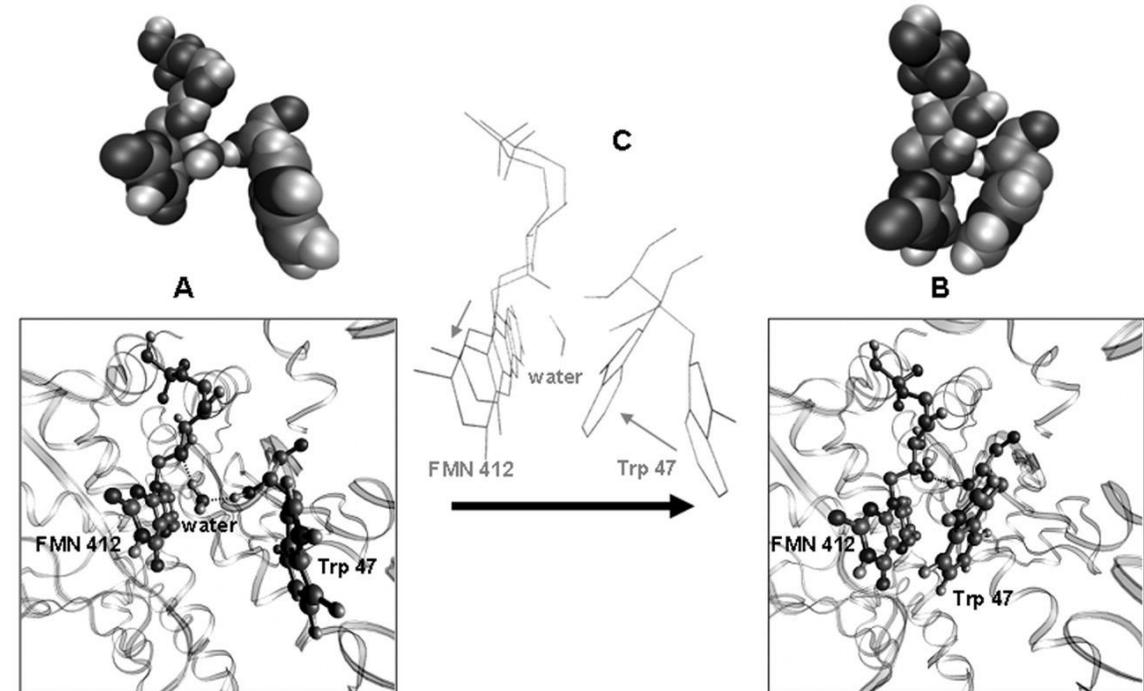
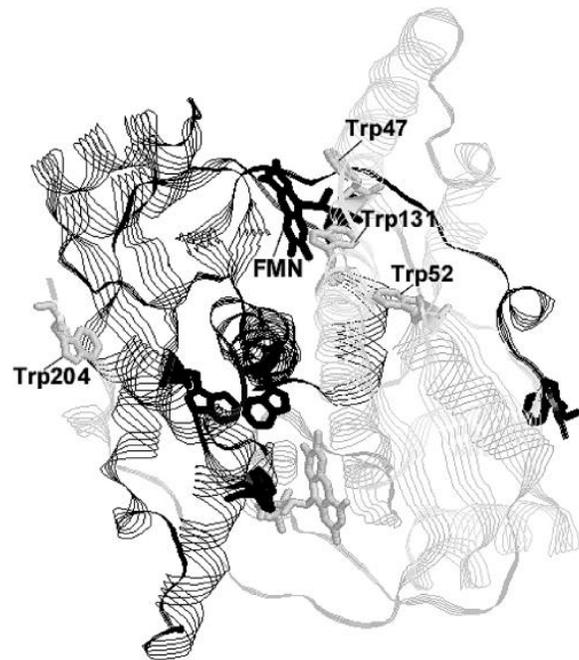
Príklad 2. Molekulárna brána ako regulátor enzýmovej aktivity

Bioinformatika ako nástroj pochopenia biologických objektov

Príklad 2. Molekulárna brána ako regulátor enzýmovej aktivity

Nanoštruktúra enzýmu NADH oxidáza

- Simulácie molekulovej dynamiky
- Fyzikálny model: klasická mechanika, silové polia

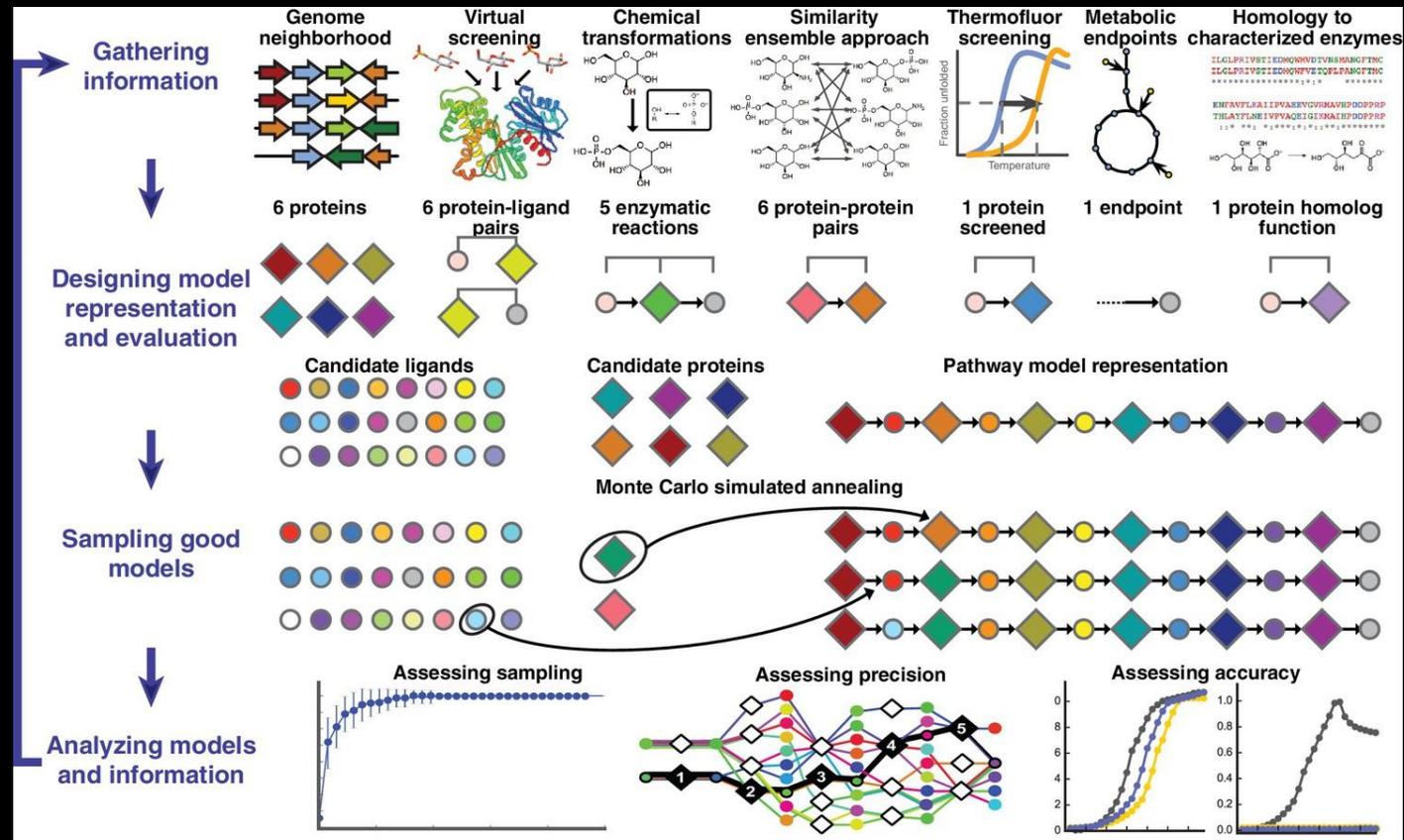


Bioinformatika ako nástroj pochopenia biologických objektov

Príklad 3. Analýza komplexných sietí metabolický dráh

Bioinformatika ako nástroj pochopenia biologických objektov

Príklad 3. Analýza komplexných sietí metabolických dráh



Calhoun S, Korczynska M, Wichelecki DJ, San Francisco B, Zhao S, Rodionov DA, Vetting MW, Al-Obaidi NF, Lin H, O'Meara MJ, Scott DA, Morris JH, Russel D, Almo SC, Osterman AL, Gerlt JA, Jacobson MP, Shoichet BK, Sali A. Prediction of enzymatic pathways by integrative pathway mapping. *Elife*. 2018 Jan 29;7. pii: e31097. doi: 10.7554/eLife.31097.

Bioinformatika ako nástroj pochopenia biologických objektov

Príklad 4. Jurský park na molekulovej úrovni: rekonštrukcia pravekých enzýmov

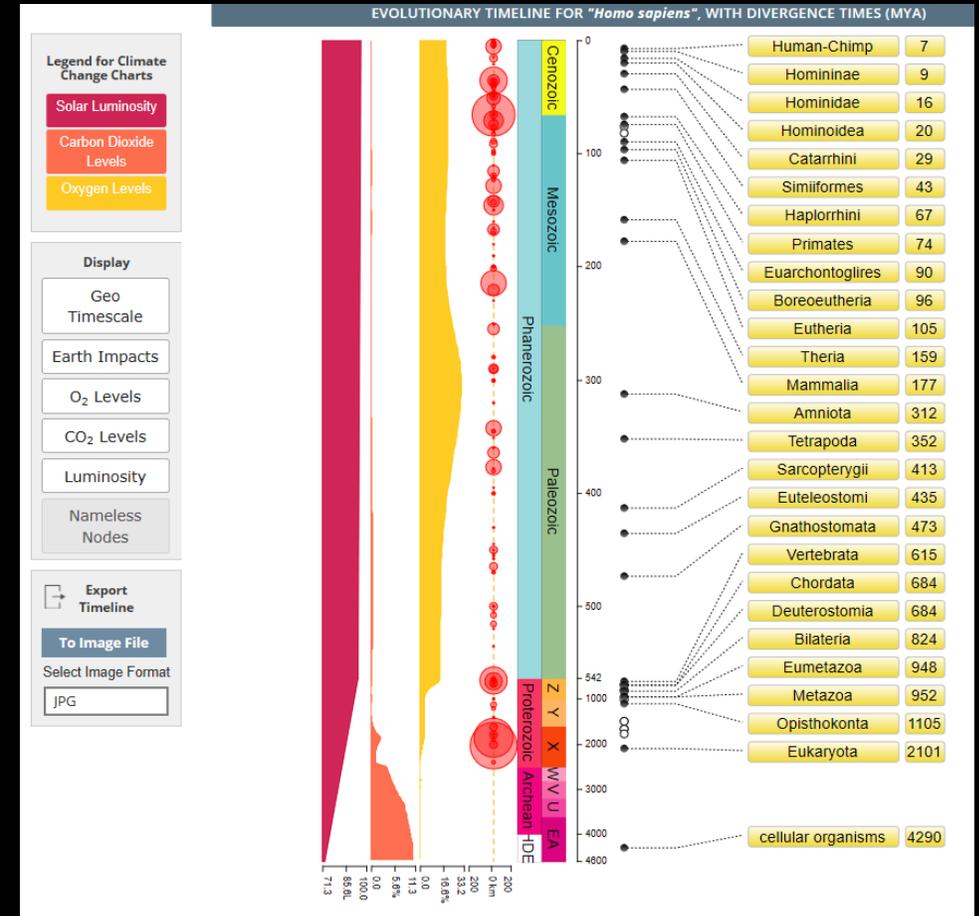
Bioinformatika ako nástroj pochopenia biologických objektov

Rekonštrukcia pravekých enzýmov

Fylogenetický strom

- Molekulárne hodiny sú známe
- Ako sa menila proteínová sekvencia v čase
- Rekonštrukcia spoločného predchodcu
- Väčšinou odolnejší voči zvýšenej teplote

```
MAPQVPPLLA FPLQPPTMDS PVQIFRGEFG PTCAPRACQL PNSSAWFPGW AEPDGNASA  
METPVQIFRG EPGPTCAPSA CLLPNSSAWF PGWAESGNGS AGSEEAQLEP AHISPAIPV  
MESPVQIFRG EPGPTCAPST CLLPNDSGWG PGWAE PDGNS SAGSEDALE PAHISPAIP  
MESPVQIFRG EPGPTCAPSA CLLPNSSAWF PGWAESGNGS AGSEEAQLEP AHISPAIPV  
MESPVQIFRG EPGPTCAPST CLLPNDSGWG PGWAE PDGNS SAGSEDALE PAHISPAIP  
MESPVQIFRG DPGPTCAPSA CLLPNGSAWF PGWAEQD GND SASSEPAQLE PAHISPAIP  
MMESPVQIFR GEPGPTCAPS ACLLPNGSTW FPDWAE PDSK GNGTASAEDA QLETAHISP  
MESPVQIFRG EPGPTCAPST CLLPNDSGWG PGWAE PDGNS SAGSEDALE PAHISPAIP  
MDSPVQIFRG EPGPTCAPRA CLLPNGSAWF SGWAE SDSNG SASPEDEPLE PAHISPAIP  
MESPVQIFRG EPGPTCAPSA CLLPNGSAWF PGWAE PDGNS SAGSEDALE PAHISPAIP
```



Bioinformatika ako nástroj pochopenia biologických objektov

Príklad 5. Biofarmaceutický priemysel: predikcia rozpustnosti proteínov

doc. Ing. Martin Tomášek PhD.

RNDr. Gabriel Žoldák PhD
Doc. RNDr. Erik Sedlák, PhD



Fakulta elektrotechniky
a informatiky



Katedra
počítačov
a informatiky



Technologický
a inovačný park
UPJŠ



CENTER FOR
INTERDISCIPLINARY
BIOSCIENCES

Bioinformatika ako nástroj pochopenia biologických objektov

Biofarmaceutický priemysel: predikcia rozpustnosti proteínov

- Komplexná štruktúra terapeutických proteínov a vakcín
 - Limitujúca koncentrácia pre aktívne použitie
 - Prísne regulačné kritéria pre množstvo/stav terapeutika
 - Predikcia na základe proteínovej sekvencie

```
MAPQVPPLLA FPLQPPTMDS PVQIFRGEFG PTCAPRACQL PNSSAWFPGW AEPDGNASA
METPVQIFRG EPGPTCAPSA CLLPNSSAWF PGWAESGNGS AGSEEAQLEP AHISPAIPV
MESPVQIFRG EPGPTCAPST CLLPNDSGWF PGWAEPDGNS SAGSEDAPLE PAHISPAIP
MESPVQIFRG EPGPTCAPSA CLLPNSSAWF PGWAESGNGS AGSEEAQLEP AHISPAIPV
MESPVQIFRG EPGPTCAPST CLLPNDSGWF PGWAEPDGNS SAGSEDAPLE PAHISPAIP
MESPVQIFRG DPGPTCAPSA CLLPNGSAWF PGWAEQDGND SASSEPAQLE PAHISPAIP
MMESPVQIFR GEPGPTCAPS ACLLPNGSTW FPDWAEPDSK GNGTASAEDA QLETAHISP
MESPVQIFRG EPGPTCAPST CLLPNDSGWF PGWAEPDGNS SAGSEDAPLE PAHISPAIP
MDSPVQIFRG EPGPTCAPRA CLLPNGSAWF SGWAESDSNG SASPEDEPLE PAHISPAIP
MESPVQIFRG EPGPTCAPSA CLLPNGSAWF PGWAEPDGNS SAGSEDAPLE PAHISPAIP
```

ok



not ok



Aké sú veľké výzvy v bioinformatike? I.

TOP10 problémov v proteínovej bioinformatike

1. Algoritmy pre sekvenčné zarovnávanie proteínov
2. Predikcia vlastností proteínov z ich primárnej sekvencie
3. Predikcia funkcie proteínov
4. Predikcia 3D štruktúry
5. Membránové proteíny
6. Identifikácia funkčného/regulačného miesta
7. Proteín-proteín interakcia
8. Predikcia proteín-malá molekula (dokovanie)
9. Proteínový dizajn
10. Proteínové inžinierstvo

Aké sú veľké výzvy v bioinformatike? II.

GRAND CHALLENGES IN COMPUTATIONAL BIOLOGY

- Protein structure prediction
- Homology searches
- Multiple alignment and phylogeny construction
- Genomic sequence analysis and gene-finding

MODELING OF BIOLOGICAL SYSTEMS

Integrating data and developing models of complex systems across multiple spatial and temporal scales

1. Scale relations and coupling
2. Temporal complexity and coding
3. Parameter estimation and treatment of uncertainty
4. Statistical analysis and data mining
5. Simulation modeling and prediction

1. Structure-function relationships

1. Large and small nucleic acids
2. Proteins
3. Membrane systems
4. General macromolecular assemblies
5. Cellular, tissue, organismal systems

OPPORTUNITIES IN MOLECULAR BIOMEDICINE IN THE ERA OF TERAFLIP COMPUTING

- Study protein-protein and protein-nucleic acid recognition and assembly
- Investigate integral functional units (dynamic form and function of large macromolecular and supramolecular complexes)
- Bridge the gap between computationally feasible and functionally relevant time scales
- Improve multiresolution structure prediction
- Combine classical molecular dynamics simulations with quantum chemical forces
- Sample larger sets of dynamical events and chemical species
- Realize interactive modeling
- Foster the development of biomolecular modeling and bioinformatics
- Train computational biologists in teraflop technologies, numerical algorithms, and physical concepts
- Bring experimental and computational groups in molecular biomedicine closer together.

1. Ecological and evolutionary systems

1. Image analysis and visualization

1. Image interpretation and data fusion
2. Inverse problems
3. Two-, three- and higher-dimensional visualization and virtual reality

2. Basic mathematical issues

1. Formalisms for spatial and temporal encoding
2. Complex geometry
3. Relationships between network architecture and dynamics
4. Combinatorial complexity
5. Theory for systems that combine stochastic and nonlinear effects often in partially distributed systems

3. Data management

1. Data modeling and data structure design
2. Query algorithms, especially across heterogeneous data types
3. Data server communication, especially peer-to-peer replication
4. Distributed memory management and process management

4. NEXT-GENERATION BIOLOGY: THE ROLE OF NEXT-GENERATION COMPUTING

Exemplar Challenges for Bioinformatics and Computational Biology

1. Full genome-genome comparisons
2. Rapid assessment of polymorphic genetic variations
3. Complete construction of orthologous and paralogous groups of genes
4. Structure determination of large macromolecular assemblies/complexes
5. Dynamical simulation of realistic oligomeric systems
6. Rapid structural/topological clustering of proteins
7. Prediction of unknown molecular structures; protein folding
8. Computer simulation of membrane structure and dynamic function
9. Simulation of genetic networks and the sensitivity of these pathways to component stoichiometry and kinetics
10. Integration of observations across scales of vastly different dimensions and organization to yield realistic environmental models for basic biology and societal needs

GRAND CHALLENGES IN BIOMEDICAL COMPUTING: Biomedical Applications from Coupling Imaging and Modeling

- Real-time noninvasive three-dimensional imaging of many body systems
- Real-time generation of three-dimensional patient-specific models
- Multiple-technology (multimodal) imaging and modeling
- Whole-organ modeling
- Multiple-organ system modeling
- Patient-specific modeling of organ anomalies
- Model support for (partial) restoration of hearing, coarse vision, and locomotion (via both paralyzed and artificial limbs)
- All of these applications make use of:
 - Three-dimensional models
 - Increasingly refined grids and increasing levels of tissue discrimination
 - Anatomically realistic models
 - Special-purpose hardware for visualization
 - Distributed computing techniques.

EMERGING FIELDS IN BIOINFORMATICS

- 1.Data storage and retrieval, database structures, annotation
- 2.Analysis of genomic/proteomic/other high-throughput information
- 3.Evolutionary model building and phylogenetic analysis
- 4.Architecture and content of genomes
- 5.Complex systems analysis/genetic circuits
- 6.Information content in DNA, RNA, protein sequences and structure
- 7.Metabolic computing
- 8.Data mining using machine learning tools, neural nets, artificial intelligence
- 9.Nucleic acid and protein sequence analyses

HIGH-PERFORMANCE COMPUTING, COMMUNICATION, AND INFORMATION TECHNOLOGY GRAND CHALLENGES

- Understanding protein folding
- Predicting structure of native protein
- Exhaustive discovery and analysis of cancer genes
- Molecular recognition and dynamics
- Drug discovery

Spolupráca:
doc. Ing. Martin Tomášek PhD.



Fakulta elektrotechniky
a informatiky



Katedra
počítačov
a informatiky

Ďakujem za pozornosť.

DEG Bioresearch



Technologický
a inovačný park
UPJŠ



CENTER FOR
INTERDISCIPLINARY
BIOSCIENCES