

Úvod

C3210 Strukturní bioinformatika

Podzim 2024

Michaela Wimmerová

Přednášející



prof. RNDr. Michaela Wimmerová, Ph.D.



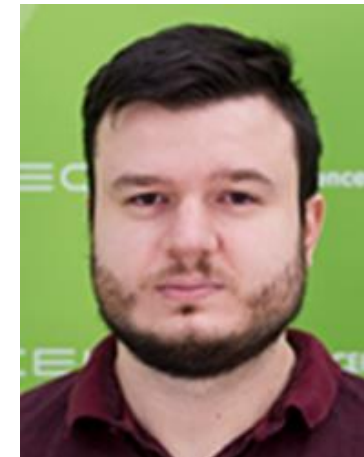
Mgr. Josef Houser, Ph.D.



MVDr. Eva Paulenová, Ph.D.



Mgr. Lenka Malinovská, Ph.D.



Mgr. Filip Melicher

Sylabus

1. Úvod, definice pojmu, náplň předmětu, vztah strukturní bioinformatiky k ostatním vědním disciplínám
2. Struktura proteinů, struktura nukleových kyselin, strukturní motivy
3. Membránové proteiny, signální sekvence, transmembránové oblasti, topologie
4. Struktura sacharidů, oligo- a polysacharidy, zápis sekvence sacharidů, 3D struktura, struktura lipidů
5. Experimentální metody pro určení struktury, difrakční metody, NMR, elektronová mikroskopie
6. Strukturní data, surová data, zpracovaná data, formáty uložení, PDB, CIF, strukturní databáze
7. Validace strukturních dat, kvalita dat, R-faktor, B-faktor, Ramachandranův diagram
8. Proteinové inženýrství, design mutací, stabilita proteinů
9. Predikce struktury proteinů, *ab initio* modelování, homologní modelování, threading
10. Interakce proteinů s malými molekulami, molekulární dokování
11. Interakce proteinů s makromolekulami, interaktom
12. Návrh léčiv, *in silico* screening

Výuka a ukončení

Přednášky ve fyzické podobě – 1 hodina týdně

V případě omezení možný přechod na online

Ukončení:

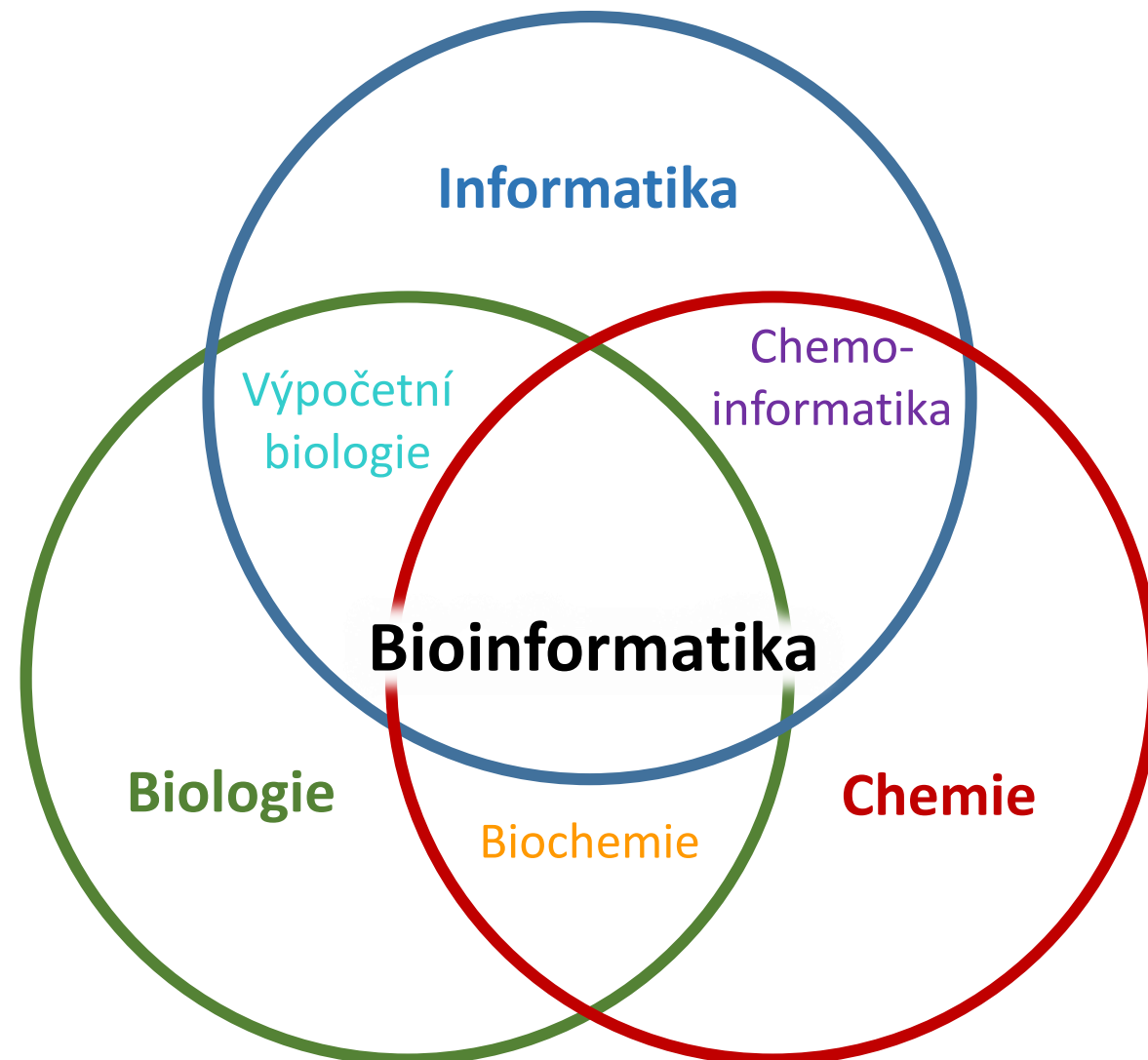
Zkouška – písemný test, případná ústní oprava

Bioinformatika

Bioinformatika je vědní disciplína, která se zabývá metodami pro shromažďování, analýzu a vizualizaci rozsáhlých souborů biologických dat, zejména dat molekulárně-biologických.

Wikipedie

Interdisciplinární věda kombinující **data** (informatiku) a **vědy o živé přírodě** (biologii a chemii)



Bioinformatické kurzy

1. ročník

C2131 Úvod do
bioinformatiky
(jaro)

C2132 Úvod do
bioinformatiky - seminář
(jaro)

2. ročník

C3210 Strukturní
bioinformatika
(podzim)

C2135 Bioinformatika
v praxi
(jaro)

C3211 Aplikovaná
bioinformatika
(jaro)

3. ročník

C2145 Strukturní
bioinformatika v praxi
(podzim)

Mgr. studium

C2138 Pokročilá
bioinformatika
(jaro)

C2139 Pokročilá
bioinformatika - seminář
(jaro)

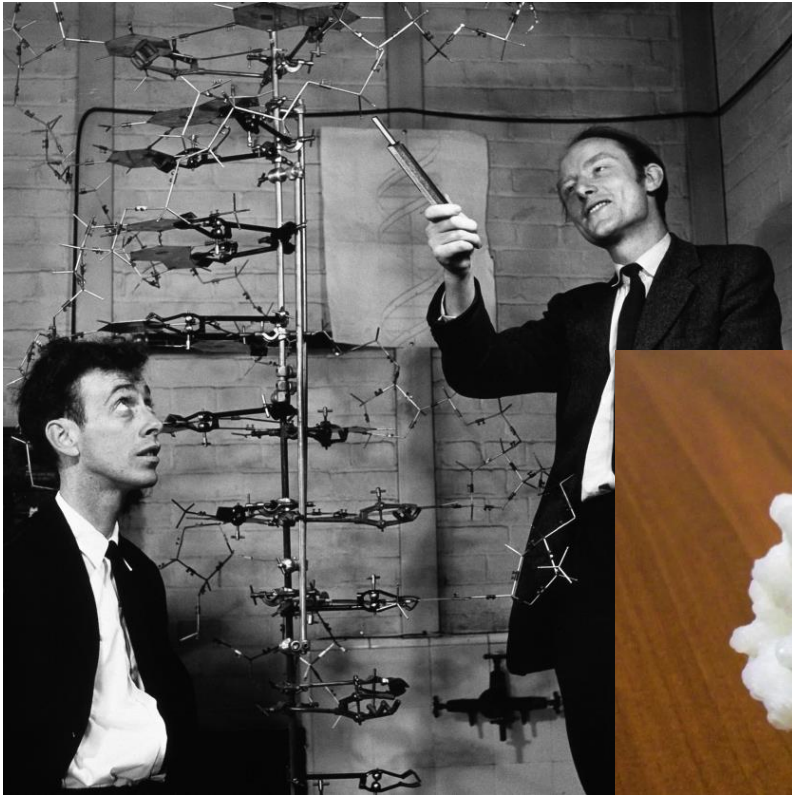
Strukturní bioinformatika

Oblast bioinformatiky zaměřená na strukturu biologických molekul

- **Určení struktury** – zpracování experimentálních dat
- **Analýza** – strukturní popis
- **Porovnání** – podobnost struktur, klasifikace
- **Predikce** – bez použití experimentu
- **Aplikace** – využití poznatků v praxi

Význam strukturní bioinformatiky

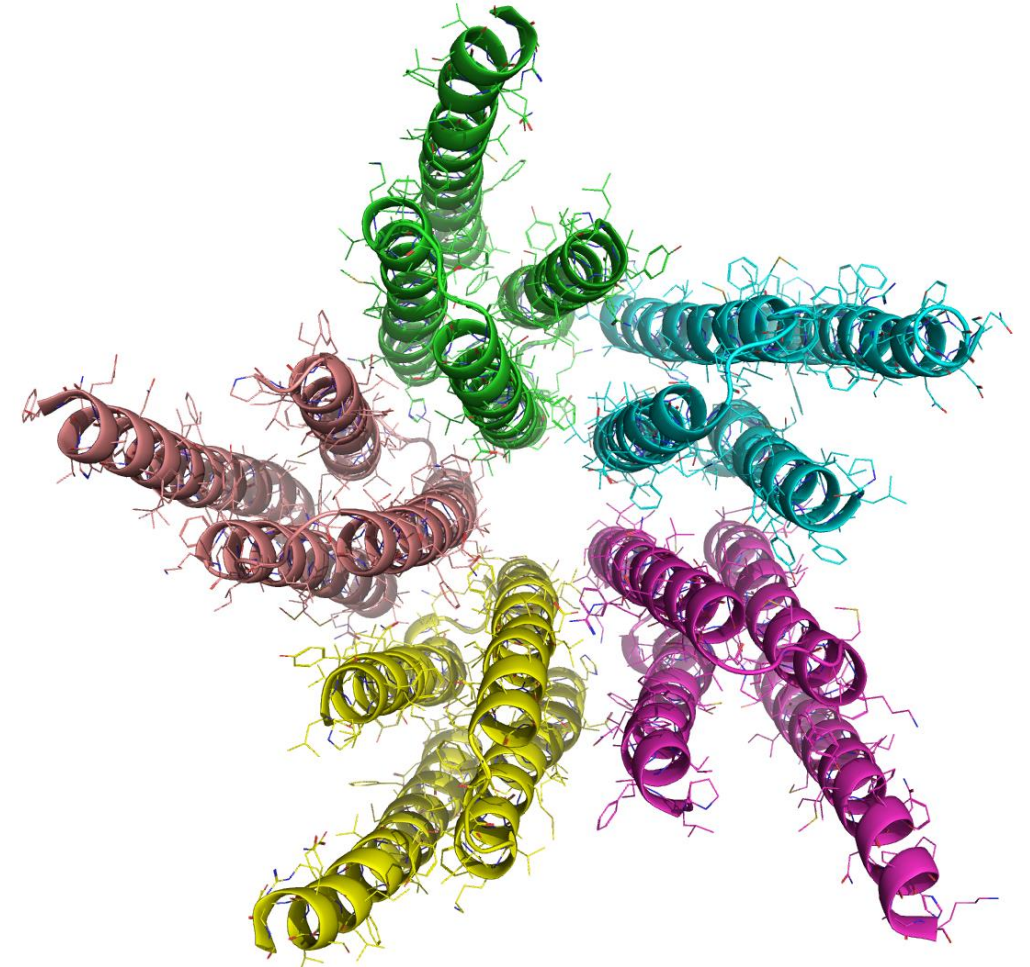
Popis světa řečí čísel



Watson & Crick DNA
(<https://www.sciencehistory.org/>)



Hemoglobin (<https://www.instructables.com/>)

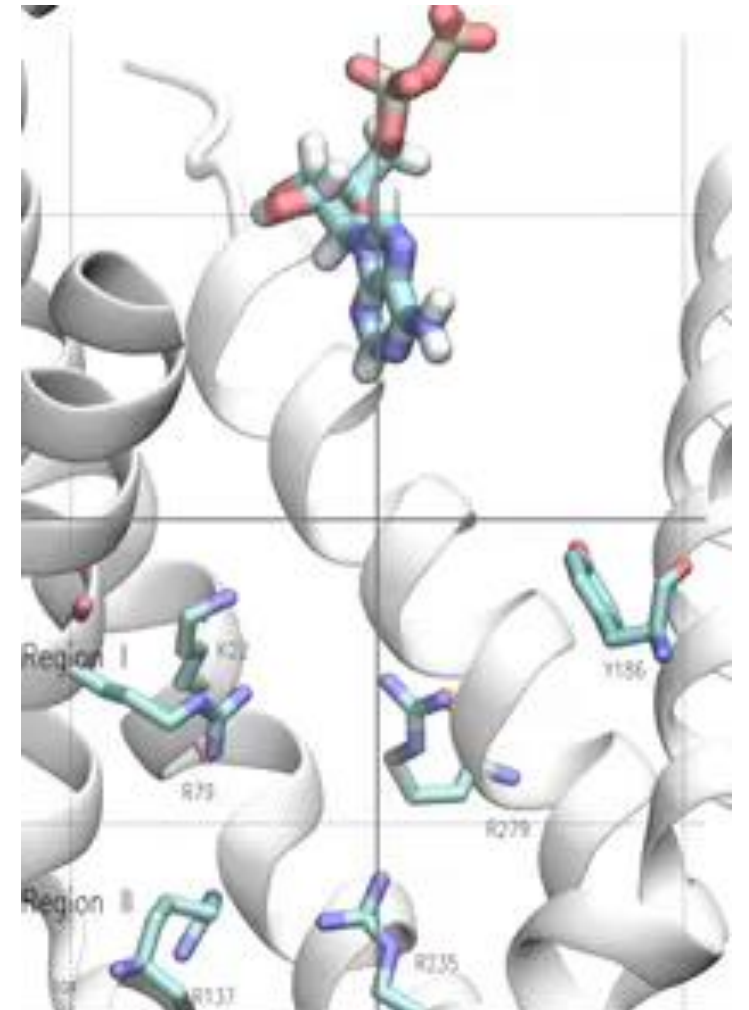
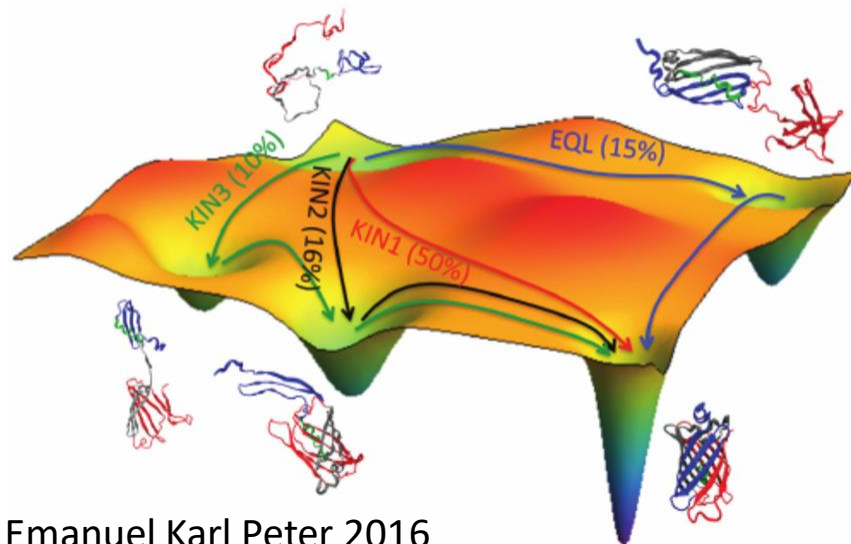


Acetylcholin receptor (PDB 1oed)

Význam strukturní bioinformatiky

Odhalení vztahů mezi molekulami

- Vazba ligandu
- Tvorba vyšších struktur (folding)
- Interakce makromolekul
- Enzymové reakce



Význam strukturní bioinformatiky

Redukce nákladů na experimenty

	Experiment	Výpočet
Čas	Dny – roky	Sekundy – týdny
Spotřební materiál	1 000 – 100 000 Kč	0 – 1000 Kč (energie)
Přístrojové vybavení	100 000 – 10 000 000 000 Kč	10 000 – 10 000 000 Kč
Lidská práce	Střední až náročná	Malá až střední

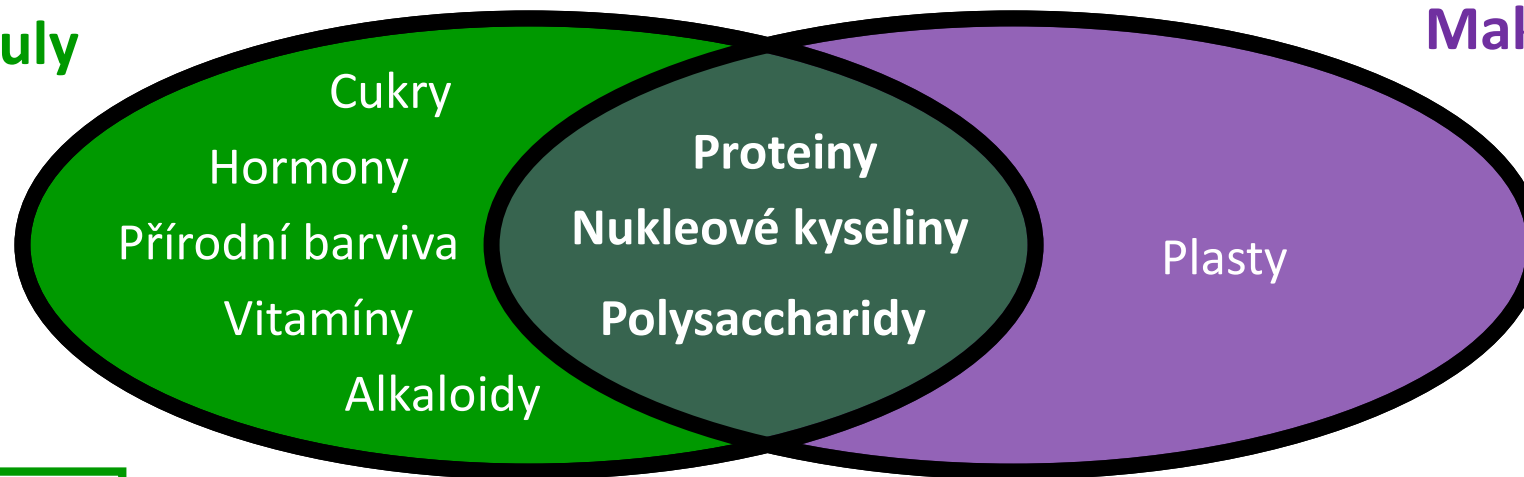
Pozn.: Hodnoty jsou velmi přibližné

Biomakromolekuly

Molekuly:

- Základní stavební jednotky hmoty.
- Tvořeny atomy, které navzájem spojují kovalentní vazby.

Biomolekuly



Makromolekuly

Plasty

Biomolekuly jsou přirozenou součástí živých organismů

Biomakromolekuly

Velké molekuly tvořené tisíci až miliony atomů

Struktura molekuly

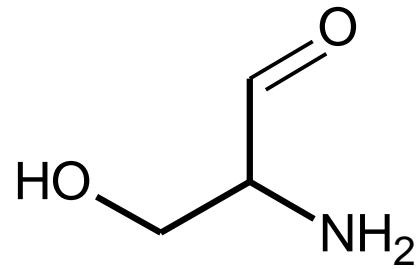
Identita

Konektivita

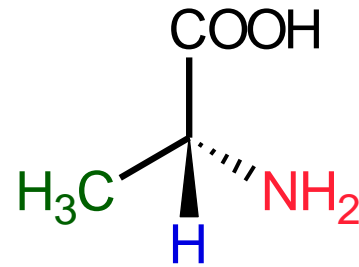
Stereoizomerie

Konformace

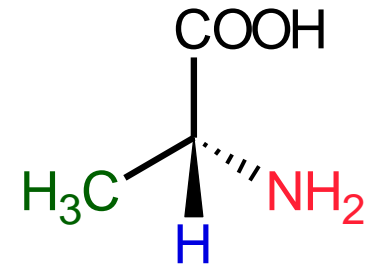
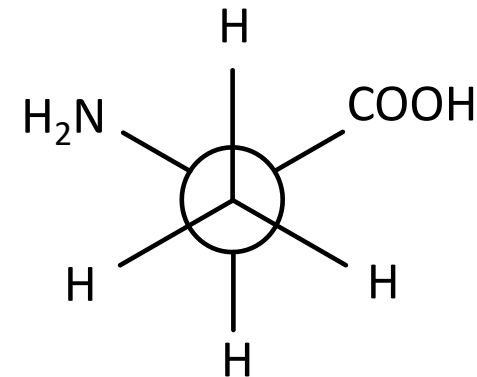
Náboj



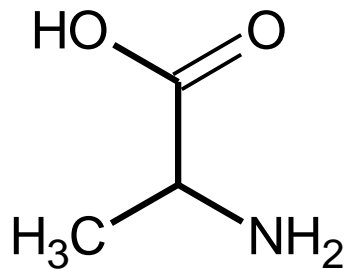
2-amino-3-hydroxypropanal



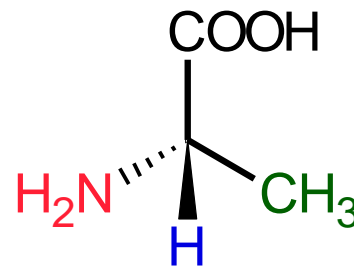
D-alanin



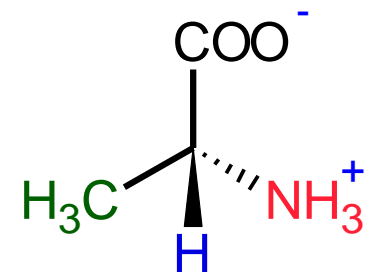
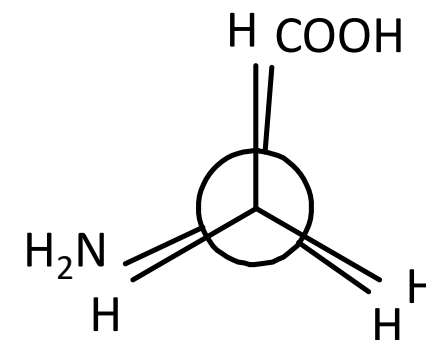
C₃H₇O₂N



2-aminopropanová kys.
(alanin)



L-alanin



Struktura molekuly

Jednoduchá molekula

- unikátní uspořádání atomů a vazeb mezi nimi

Složená molekula – „polymer“

- unikátní uspořádání atomů a vazeb mezi nimi
- tvořená identickými (homopolymer) nebo různými (heteropolymer) podjednotkami
- **struktura podjednotek** + **vazba podjednotek**

Struktura biomakromolekul

2.-4. přednáška

Definované podjednotky

Univerzální vazby mezi podjednotkami

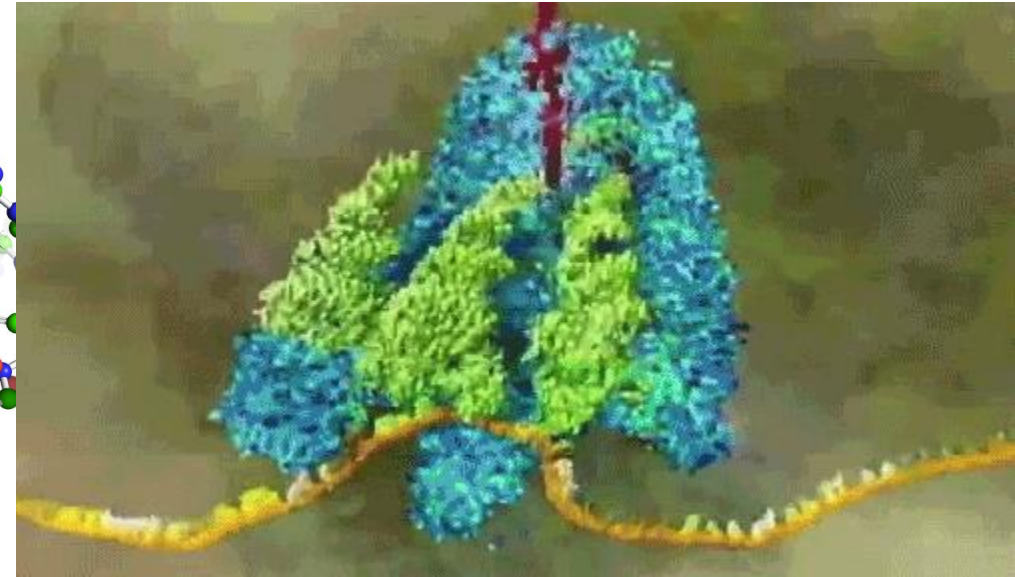
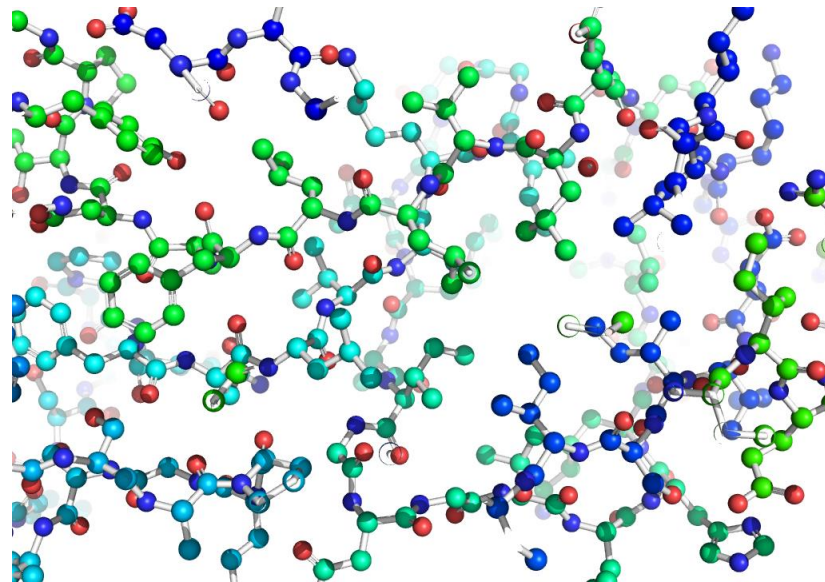
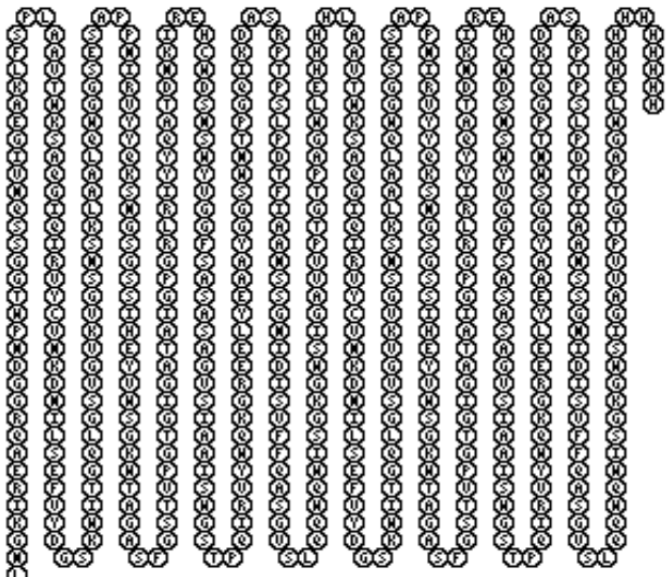
Makromolekula	Stavební jednotky	Počet různých základních jednotek	Typ vazby	Schéma
Protein	Aminokyseliny	22	Peptidová	
Nukleová kyselina	Nukleotidy	5	Esterová	
Polysacharid	Monosacharidy	cca 200	Glykosidická	

Struktura biomakromolekul

Řetězení stavebních jednotek – **primární struktura (1D)**

Uspořádání stavebních jednotek v prostoru – **terciární struktura (3D)**

Změna struktury v čase – **dynamika**

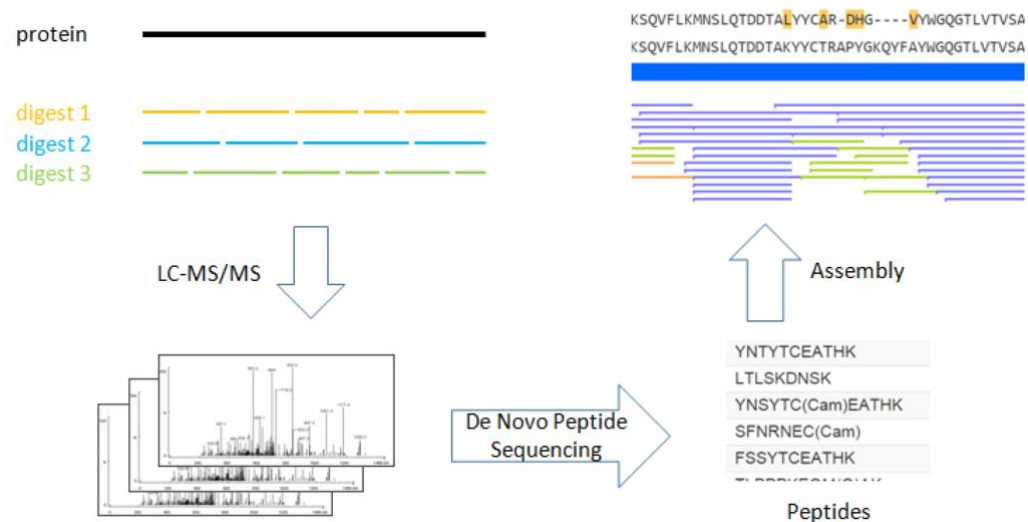


Metody strukturní biochemie

Určení sekvence – sekvenace DNA (RNA, proteinů)

Hmotnostní spektrometrie

- Časově náročnější
- Zejména pro proteiny



Sekvenace DNA

- Založená na polymerázové reakci
- Sekvence proteinu po překladu

The table shows the 64 codons and the amino acid for each. The direction of the mRNA is 5' to 3'.

		2nd base			
		U	C	A	G
U	UUU (Phe/F) Phenylalanine	UCU (Ser/S) Serine	UAU (Tyr/Y) Tyrosine	UGU (Cys/C) Cysteine	
	UUC (Phe/F) Phenylalanine	UCC (Ser/S) Serine	UAC (Tyr/Y) Tyrosine	UGC (Cys/C) Cysteine	
	UUA (Leu/L) Leucine	UCA (Ser/S) Serine	UAA Ochre (Stop)	UGA Opal (Stop)	
	UUG (Leu/L) Leucine	UCG (Ser/S) Serine	UAG Amber (Stop)	UGG (Trp/W) Tryptophan	
C	CUU (Leu/L) Leucine	CCU (Pro/P) Proline	CAU (His/H) Histidine	CGU (Arg/R) Arginine	
	CUC (Leu/L) Leucine	CCC (Pro/P) Proline	CAC (His/H) Histidine	CGC (Arg/R) Arginine	
	CUA (Leu/L) Leucine	CCA (Pro/P) Proline	CAA (Gln/Q) Glutamine	CGA (Arg/R) Arginine	
	CUG (Leu/L) Leucine	CCG (Pro/P) Proline	CAG (Gln/Q) Glutamine	CGG (Arg/R) Arginine	
A	AUU (Ile/I) Isoleucine	ACU (Thr/T) Threonine	AAU (Asn/N) Asparagine	AGU (Ser/S) Serine	
	AUC (Ile/I) Isoleucine	ACC (Thr/T) Threonine	AAC (Asn/N) Asparagine	AGC (Ser/S) Serine	
	AUA (Ile/I) Isoleucine	ACA (Thr/T) Threonine	AAA (Lys/K) Lysine	AGA (Arg/R) Arginine	
	AUG (Met/M) Methionine, Start ^[A]	ACG (Thr/T) Threonine	AAG (Lys/K) Lysine	AGG (Arg/R) Arginine	
G	GUU (Val/V) Valine	GCU (Ala/A) Alanine	GAU (Asp/D) Aspartic acid	GGU (Gly/G) Glycine	
	GUC (Val/V) Valine	GCC (Ala/A) Alanine	GAC (Asp/D) Aspartic acid	GGC (Gly/G) Glycine	
	GUA (Val/V) Valine	GCA (Ala/A) Alanine	GAA (Glu/E) Glutamic acid	GGA (Gly/G) Glycine	
	GUG (Val/V) Valine	GCG (Ala/A) Alanine	GAG (Glu/E) Glutamic acid	GGG (Gly/G) Glycine	

Metody strukturní biochemie

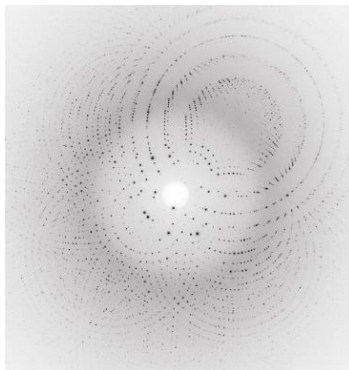
5. přednáška

Určení 3D struktury – atomární model

Difrakční techniky

Rozptyl záření na krystalu

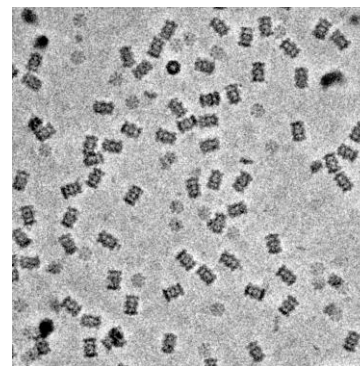
- **Rentgenové** (X-ray)
- **Neutronové**
- **Elektronové**



Mikroskopie

Přímé pozorování molekuly

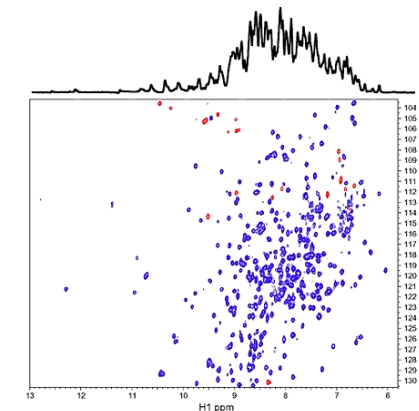
- **Elektronový mikroskop**
- V zamraženém stavu – kryo-mikroskopie



Spektroskopie

Absorpce a emise záření

- **Nukleární magnetická rezonance (NMR)**
- **Infračervená**



Data

6. přednáška

Původní informace (raw data)

- Přímý výstup experimentu
- Značná velikost (GB-TB)
- Problém s uchováváním (datová úložiště, deponitáře)

Odvozené informace (processed data)

- Vznikají zpracováním raw dat
- Řádově menší (MB)
- Depozice (databáze)
- Nutnost dodržování jednotného formátu

Doplňkové informace (metadata)

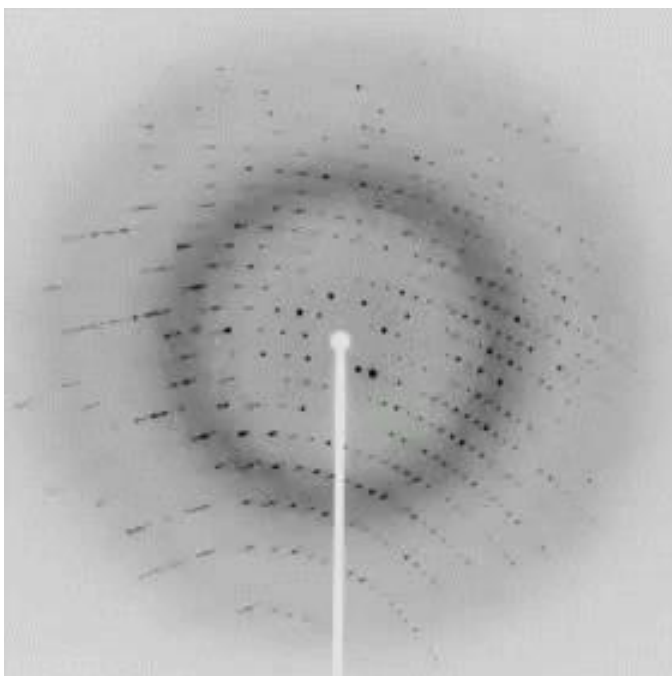
- Často podceňovaná
- Textové informace (kB)
- Nutná pro zařazení do kontextu
- Umožňují globální analýzy

Strukturní data

- Difrakční snímky
- NMR spektra
- Mikroskopické snímky

- 3D souřadnice
- Elektronová hustota

- Identifikace molekuly
- Detaily experimentu



ATOM	1133	O	GLU	A	148	3.095	-2.529	14.985	1.00	26.53	O
ATOM	1134	N	GLY	A	149	2.058	-3.665	16.604	1.00	23.46	N
ATOM	1135	CA	GLY	A	149	0.759	-3.683	15.930	1.00	21.50	C
ATOM	1136	C	GLY	A	149	-0.149	-2.717	16.654	1.00	21.02	C
ATOM	1137	O	GLY	A	149	0.278	-1.625	17.036	1.00	22.79	O
ATOM	1138	N	THR	A	150	-1.387	-3.126	16.876	1.00	19.16	N
ATOM	1139	CA	THR	A	150	-2.394	-2.276	17.487	1.00	19.56	C
ATOM	1140	CB	THR	A	150	-3.785	-2.884	17.241	1.00	20.01	C
ATOM	1141	OG1	THR	A	150	-4.047	-2.962	15.841	1.00	20.52	O
ATOM	1142	CG2	THR	A	150	-4.869	-2.051	17.865	1.00	18.35	C
ATOM	1143	C	THR	A	150	-2.198	-2.224	19.019	1.00	21.59	C
ATOM	1144	O	THR	A	150	-1.879	-3.271	19.623	1.00	20.76	O
ATOM	1145	N	ASN	A	151	-2.435	-1.056	19.649	1.00	19.17	N
ATOM	1146	CA	ASN	A	151	-2.570	-0.979	21.118	1.00	18.28	C
ATOM	1147	CB	ASN	A	151	-2.145	0.382	21.592	1.00	18.35	C
ATOM	1148	CG	ASN	A	151	-1.978	0.454	23.079	1.00	18.22	C
ATOM	1149	OD1	ASN	A	151	-2.978	0.566	23.868	1.00	18.69	O
ATOM	1150	ND2	ASN	A	151	-0.723	0.377	23.504	1.00	17.30	N
ATOM	1151	C	ASN	A	151	-4.021	-1.149	21.519	1.00	20.04	C
ATOM	1152	O	ASN	A	151	-4.886	-0.474	20.947	1.00	22.82	O
ATOM	1153	N	LEU	A	152	-4.292	-2.003	22.515	1.00	19.09	N
ATOM	1154	CA	LEU	A	152	-5.642	-2.363	22.898	1.00	18.77	C
ATOM	1155	CB	LEU	A	152	-5.785	-3.893	22.899	1.00	20.18	C
ATOM	1156	CG	LEU	A	152	-5.438	-4.501	21.523	1.00	20.57	C
ATOM	1157	CD1	LEU	A	152	-5.297	-5.997	21.622	1.00	21.44	C
ATOM	1158	CD2	LEU	A	152	-6.472	-4.139	20.457	1.00	19.49	C
ATOM	1159	C	LEU	A	152	-6.058	-1.799	24.224	1.00	19.31	C
ATOM	1160	O	LEU	A	152	-7.062	-2.206	24.769	1.00	19.35	O
ATOM	1161	N	GLY	A	153	-5.305	-0.841	24.746	1.00	20.82	N
ATOM	1162	CA	GLY	A	153	-5.801	0.007	25.831	1.00	20.04	C
ATOM	1163	C	GLY	A	153	-5.400	-0.499	27.213	1.00	21.28	C
ATOM	1164	O	GLY	A	153	-4.695	-1.544	27.348	1.00	20.92	O
ATOM	1165	N	GLY	A	154	-5.850	0.228	28.243	1.00	20.30	N

EXPERIMENT TYPE :	X-RAY
TEMPERATURE (KELVIN) :	100
NUMBER OF CRYSTALS USED :	1
SYNCHROTRON (Y/N) :	Y
RADIATION SOURCE:	ESRF
BEAMLINE:	ID23-1
X-RAY GENERATOR MODEL :	NULL
MONOCHROMATIC OR LAUE (M/L) :	M
WAVELENGTH OR RANGE (A) :	0.9795
MONOCHROMATOR :	SI111
OPTICS :	MIRRORS

Databáze

6. přednáška

Primární strukturní databáze – PDB, NDB, EMDB, BMRB

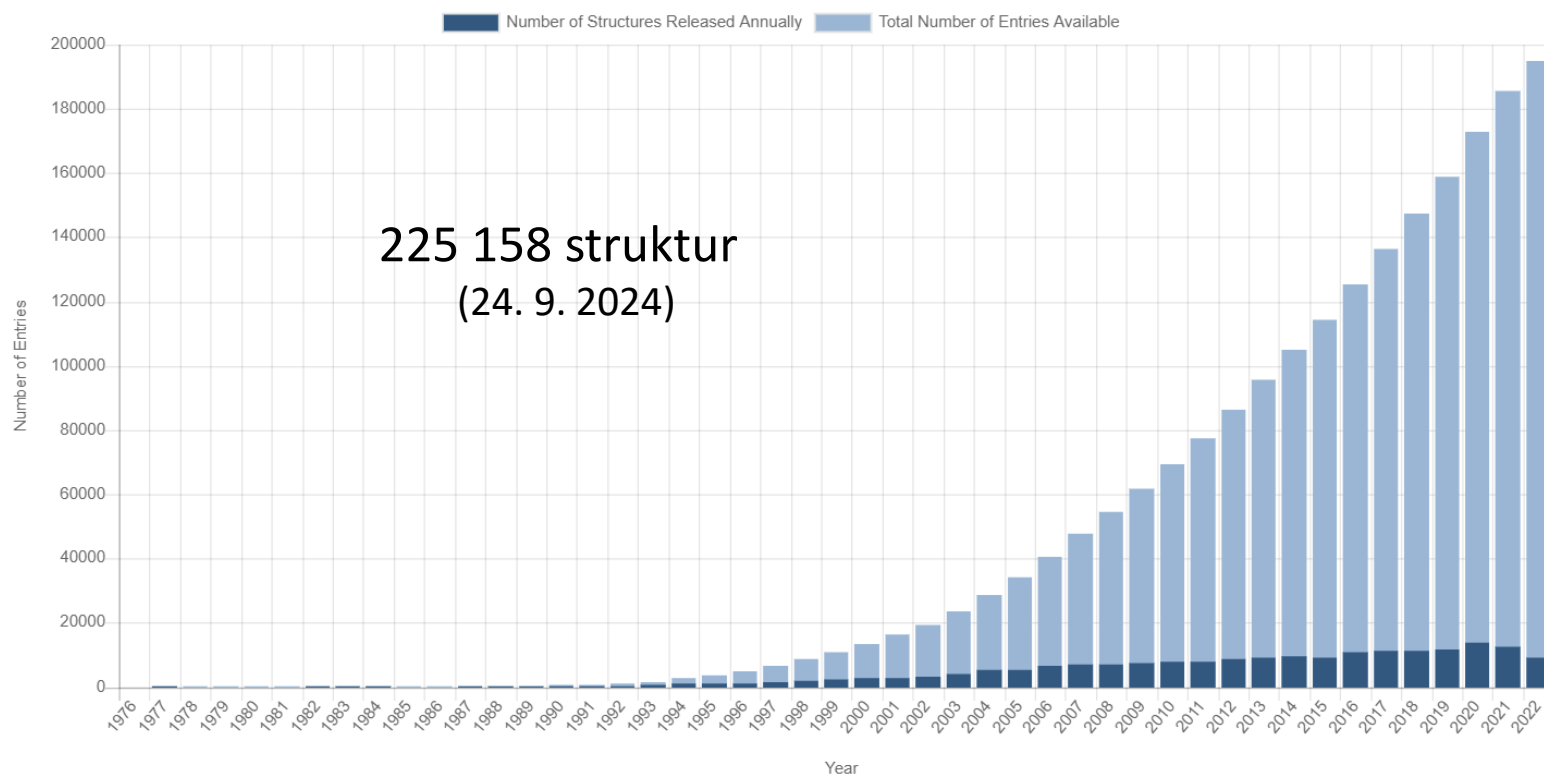


<https://www.rcsb.org/>



<https://www.wwpdb.org/>

PDB Statistics: Overall Growth of Released Structures Per Year



6. přednáška

Databáze

http://ndbserver.rutgers.edu/

18 369 struktur

Home | About NDB | Standards | Education | Tools | Software | Download



A Portal for Three-dimensional Structural Information about Nucleic Acids
As of 8-Sep-2021 number of released structures: 11677

Search DNA | Search RNA | Advanced Search

Enter an NDB ID or PDB ID

Welcome to the NDB

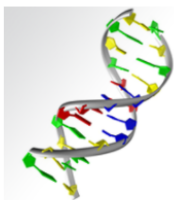
The NDB contains information about experimentally-determined nucleic acids. Use the NDB to perform searches based on annotations relating to sequence, and learn about nucleic acids.

Search Structures

Search DNA
Search DNA and its complexes

Search RNA
Search for RNA structures in the NDB archive or in the Non-Redundant list

Advanced Search
Search for structures based on structural features, chemical features, binding modes, citation and experimental information



Featured

RNA 3D Model collection of hairpin loop

Non-redundant 3D structure

RNA Base Pair motifs conservation

WebFR3D, a geometric

R3D Align, alignments

https://www.ebi.ac.uk/emdb/

38 626 struktur

EMBL-EBI | Services | Research | Training | About us | EMBL-EBI

EMDB
Electron Microscopy Data Bank

Enter your search term(s) in the box below or [build an advanced search query](#)

Search EMDB...

Examples: 1001, Apoferritin, Tomography, Rossmann MG, 5A1A

Home | Deposition | Documentation | Resources | FTP Archive | REST API | About | Feedback | Share

https://bmr.io/

8 325 struktur

BMRB | ABOUT | DEPOSIT | SEARCH | VISUALIZE | ANALYZE | DATA | LEARN



Searches all entries

BMRB makes bio-NMR data FAIR.
Findable, Accessible, Interoperable, Re-usable.

BMRB collects, annotates, archives, and disseminates spectral and quantitative data derived from NMR spectroscopic investigations of biological macromolecules and metabolites.

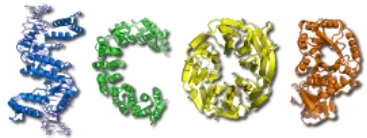
--	--	--	--

Odvozené databáze

Vychází ze strukturních dat a dále je analyzují



PDBsum (<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html>)
- analýza záznamů z PDB



SCOP 2 (<https://www.ebi.ac.uk/pdbe/scop/>)
- strukturní klasifikace proteinů



CATH/Gene3D (<http://www.cathdb.info/>)
- klasifikace proteinových domén z PDB

Kvalita dat, důvěryhodnost

7. přednáška

Data nevznikají sama od sebe

Přesnost měření, přesnost výpočtu, lidská chyba

Dezinterpretace, falzifikace

**Validace
nutná !**

Kvalita dat, důvěryhodnost

7. přednáška

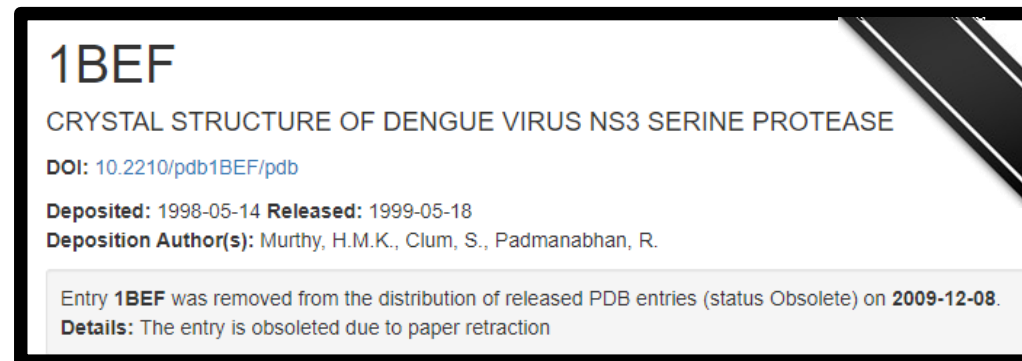
Falzifikace...

Protein structure fabrication scandal

In total, twelve of H.M. Krishna Murthy's structures and ten of his papers were affected by this scandal. The papers were retracted by their respective journals and upon retraction. All of the structures have been removed from the Protein Data Bank. The protein structures in question were [1BEF](#), [1CMW](#), [1DF9](#), [2QID](#), [1G40](#), [1G44](#), [1L6L](#), [2OU1](#), [1RID](#), [1Y8E](#), [2A01](#), and [2HR0](#). Murthy's falsified data ended up affecting 449 papers. These papers covered a range of topics; everything from dengue to Taq DNA polymerase.

*In multiple instances, they found **impossible electron densities and highly unlikely geometries**. In some cases, the information Murthy cited as proof of the structures was **not a piece of original work**. This confirmed the fictitious nature of his structures by showing that he had used another protein as a model for his creation. Specifically, Murthy tried to pass some elements of a known structure of TAQ polymerase (PDB entry [1TAQ](#)) as a new structure (PDB entry [1CMW](#)). When the misconduct inquiry looked for Murthy's work on several of these proteins, **no experimental data was made available to them**.*

https://en.wikipedia.org/wiki/H.M._Krishna_Murthy



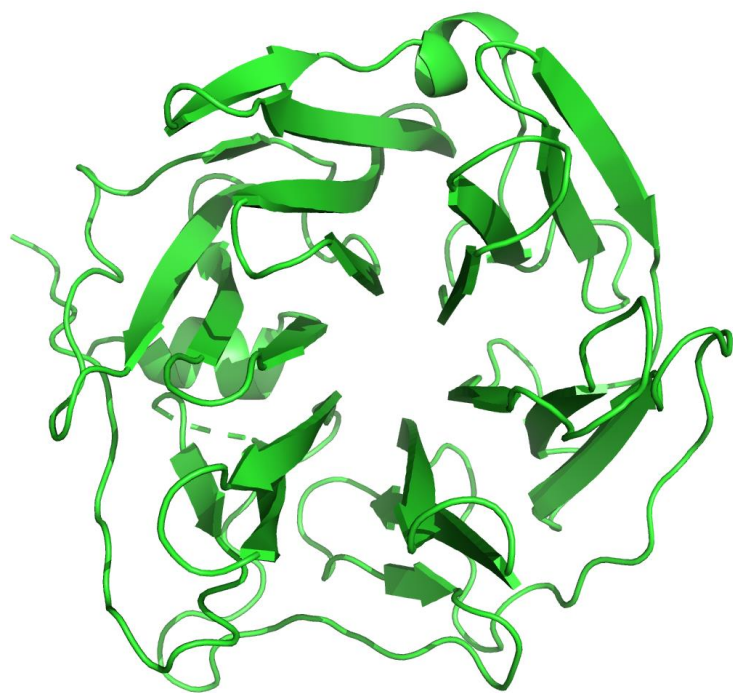
1BEF
CRYSTAL STRUCTURE OF DENGUE VIRUS NS3 SERINE PROTEASE
DOI: [10.2210/pdb1BEF/pdb](https://doi.org/10.2210/pdb1BEF/pdb)
Deposited: 1998-05-14 Released: 1999-05-18
Deposition Author(s): Murthy, H.M.K., Clum, S., Padmanabhan, R.
Entry **1BEF** was removed from the distribution of released PDB entries (status Obsolete) on 2009-12-08.
Details: The entry is obsoleted due to paper retraction

The UAB team that investigated Murthy's publications said that 1BEF is “**an improbable structure...[with] an unacceptable level of inter-atomic clashing.**”

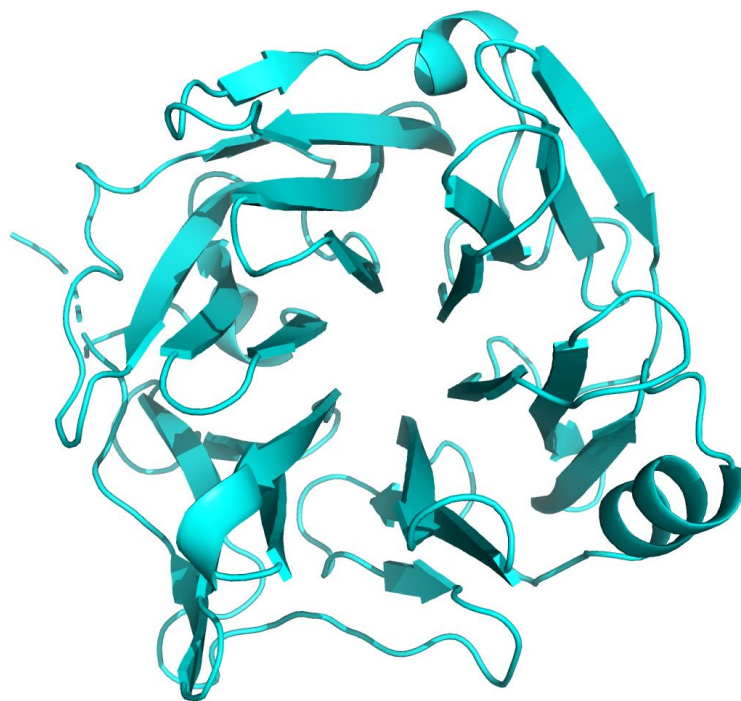
**Validace
nutná !**

Aplikace – predikce struktury

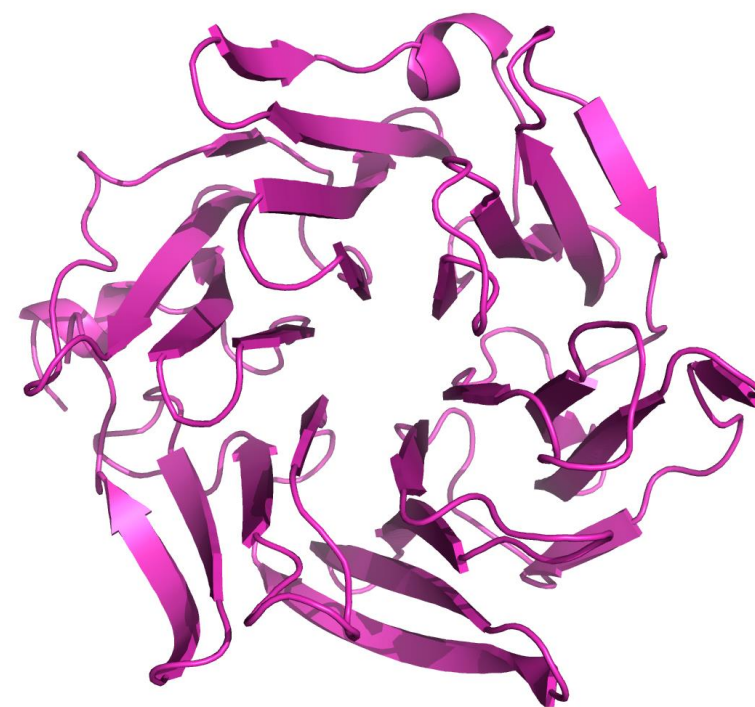
Ab initio, homologní modelování, „threading“



Strukturální model 1



Strukturální model 2



Reálná struktura

Aplikace – interakce molekul

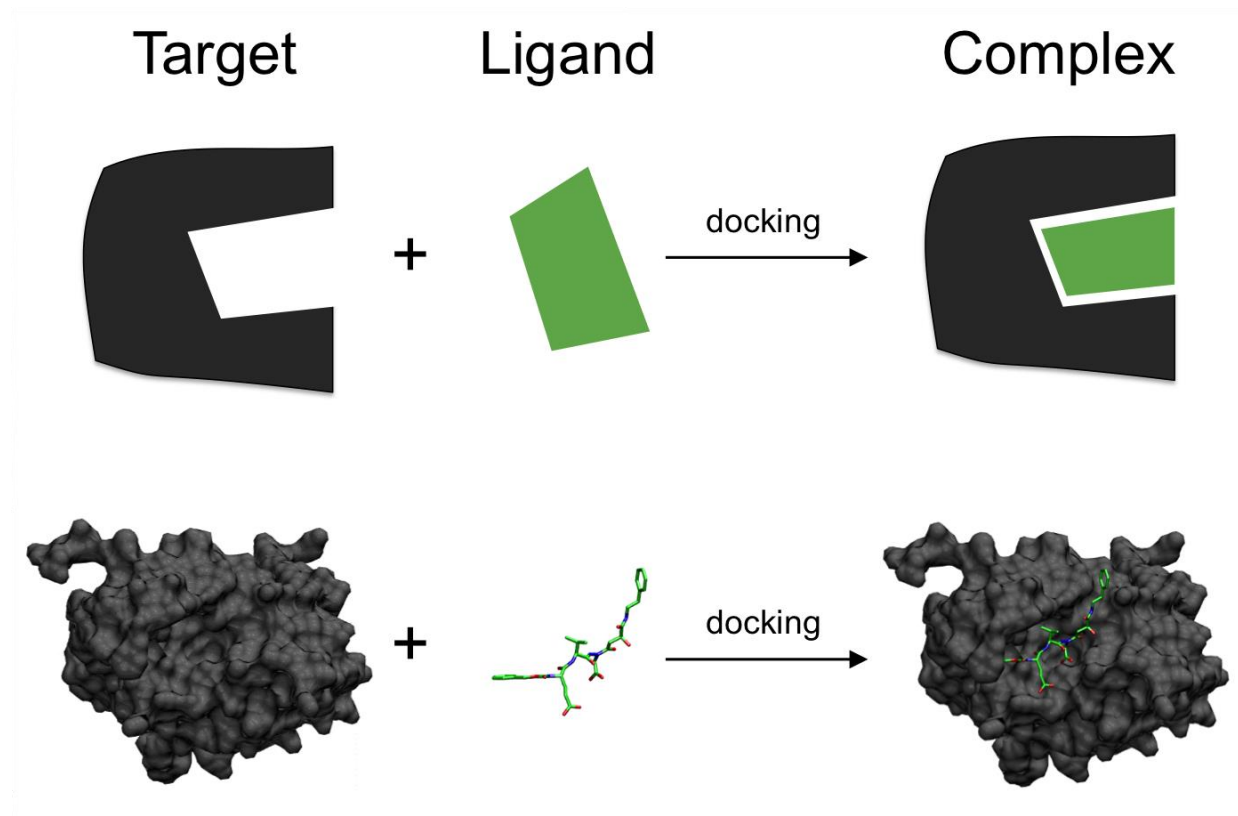
10.-11. přednáška

„**Docking**“ – „zkoušení, jak zapadne molekula do druhé“

- rigidní
- „induced fit“
- flexibilní
- výpočet vazebné energie

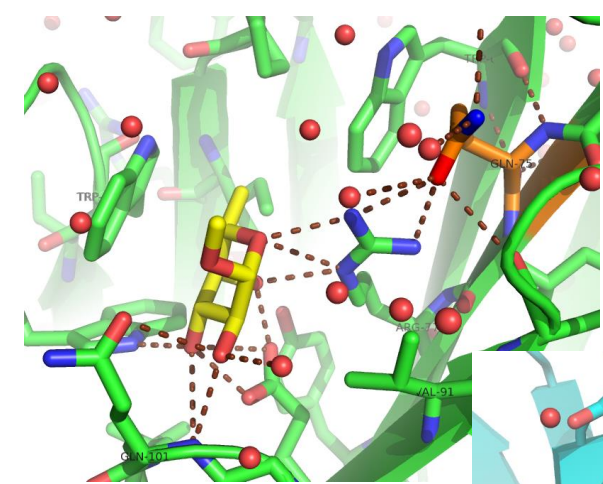
Molekulová dynamika

- pohyby
- energetika

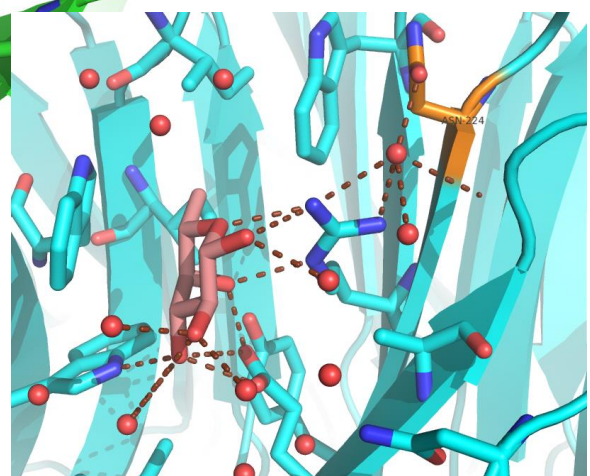
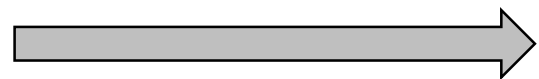


Aplikace – design proteinů

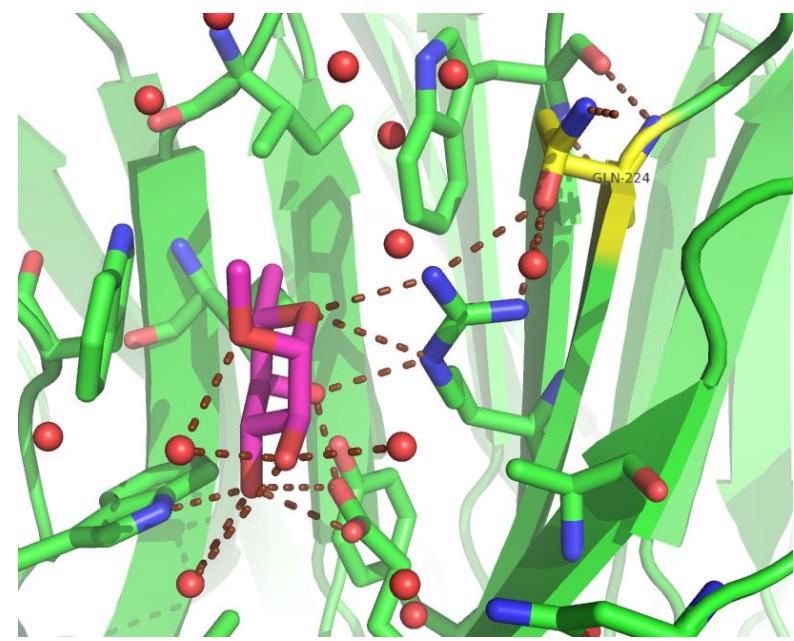
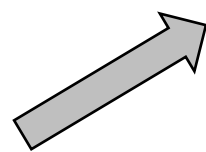
Zvýšení stability, změna vazebných vlastností, enzymová specifita,...



Varianta A



Varianta B

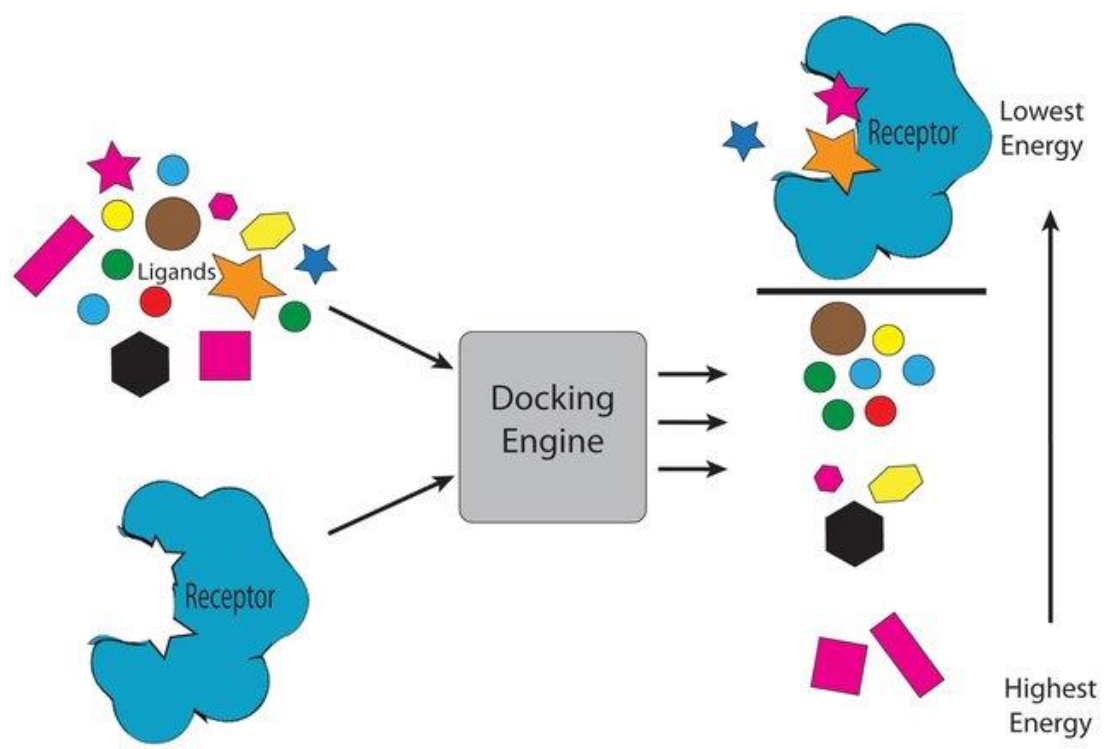


Výsledný mutant – kombinace vlastností (vazebná specifita, afinita)

Aplikace – návrh léčiv

Virtuální screening

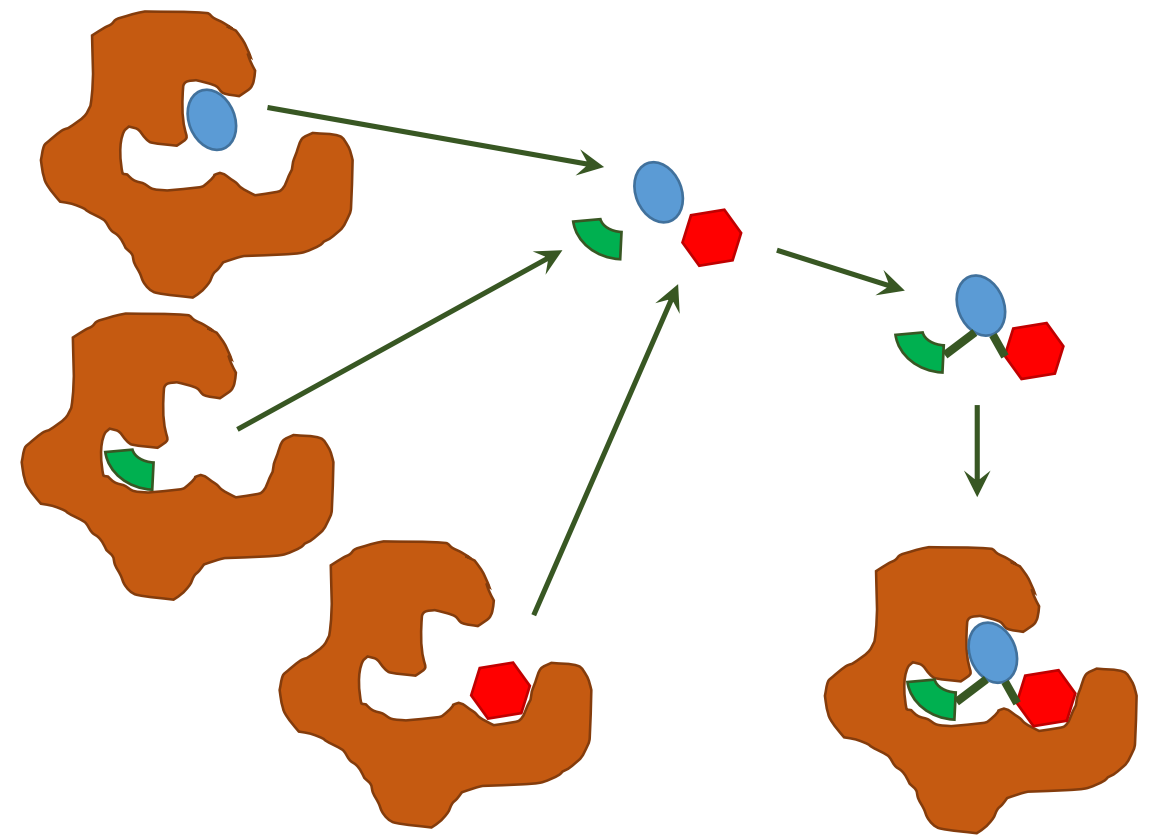
- výběr možných interagujících látek



Reed B Jacob et al 2012

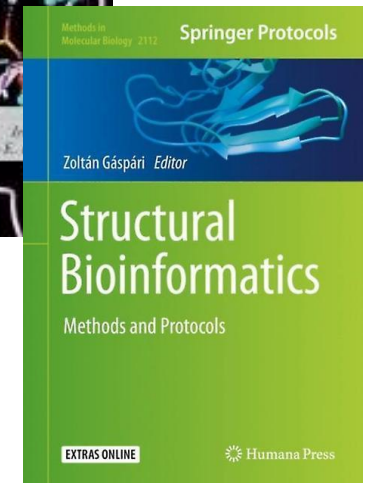
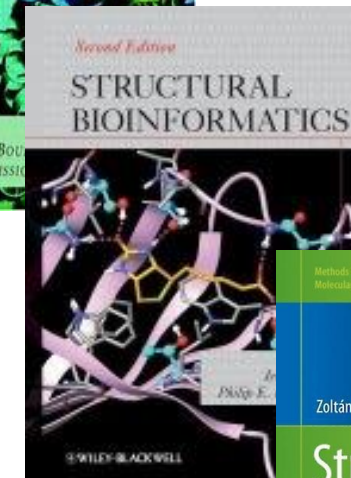
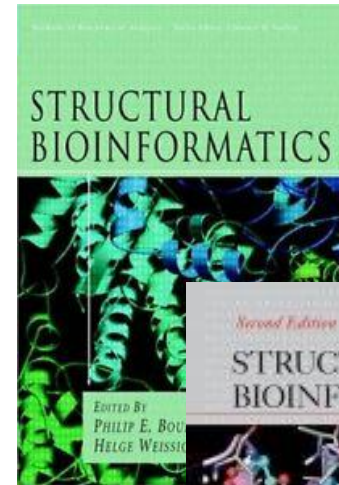
„Fragment screening“

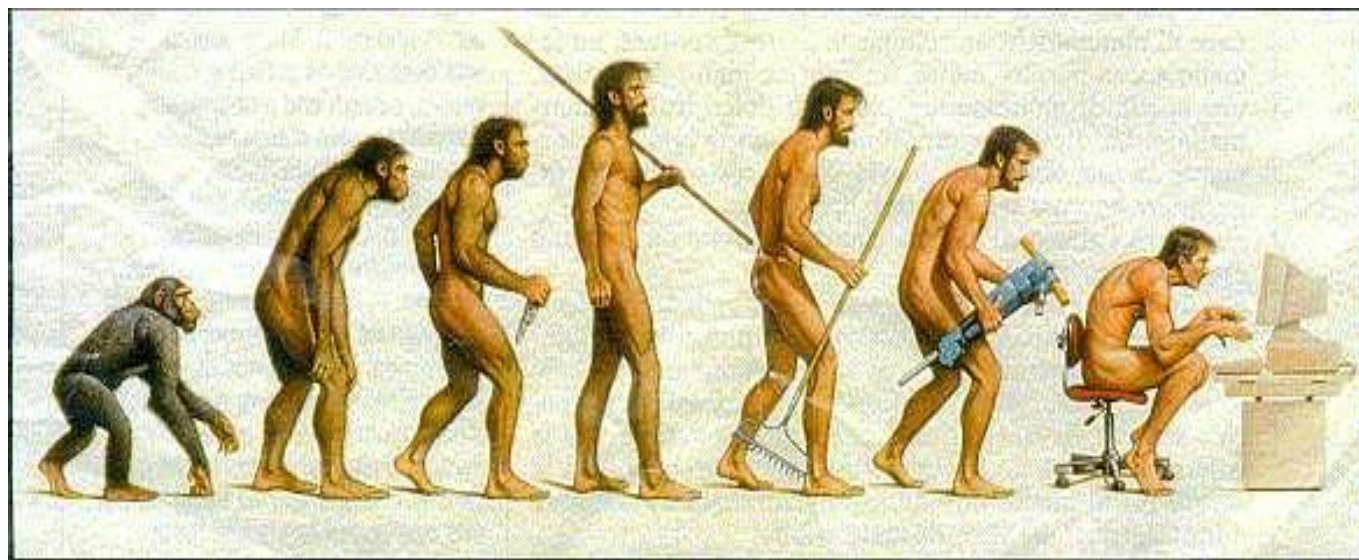
- kombinace malých ligandů do komplexních molekul



Literatura

- **Přednášky** – základ
- **Knihy** – souhrnné informace
- **Odborné články** – detaily, aktuálnost





Evoluce bioinformatika