



# BIOINFORMAČNÍ NÁSTROJE



VYUŽITÍ V PRAXI  
VYPRACOVÁNÍ PROJEKTU



## Zkouška:

test 11.5. 2023 9:00 C2-211

+ přednáška/prezentace + nová data (výběr termínu)

## Prezentace – Analýza proteinu

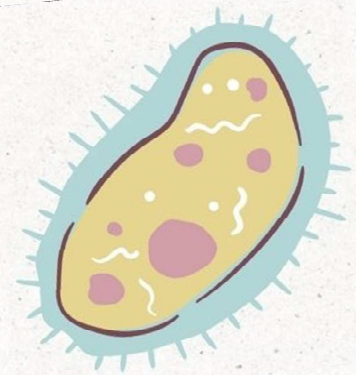
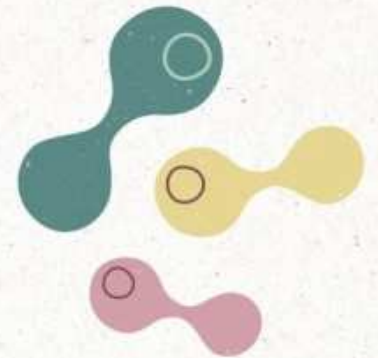
Obsah této přednášky

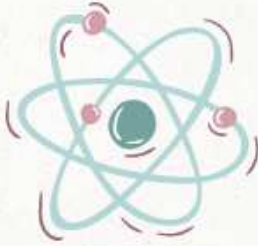
Konkrétní nová data – článek (< 5 let) o komplexu (nebo proteinu)

Ujasnit si souvislosti, rozšířit si znalosti, aplikovat poznatky z přednášek ...

# Analýza proteinu

- Popis a funkce
- Konzervovanost (Alignment)
- Evoluce (fylogenetický strom)
- Zjištění přítomnosti domén daného proteinu
- Struktura proteinu – Alphafold/Colabfold
- Zjištění interakčních partnerů
- Úprava proteinového modelu v PyMol
  - Zvýraznění interakčních aminokyselin
  - Zvýraznění domén





# UniProt a jeho možnosti

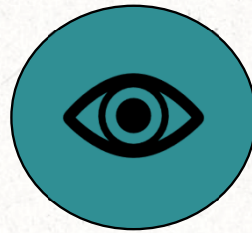
Komplexní, vysoce kvalitní a volně přístupný zdroj sekvenčních a funkčních informací o proteinech



**Hledání  
proteinu**



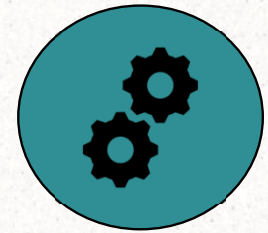
**Souvislosti**  
Funkce  
Interaktom  
Mutace



**Vizualizace**  
AlphaFold  
Lokalizace proteinu



**Stahování  
dat**



**Analýzy**  
BLAST  
Alignment

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

Výběr databáze

## Find your protein

Pokročilé vyhledávání

UniProtKB ▾

Advanced | List Search

Examples: Insulin, APP, Human, P05067, organism\_id:9606

Název genu  
Název proteinu  
Onemocnění  
Search bar

UniProt is the world's leading high-quality, comprehensive and protein sequence and functional information. [Cite UniProt](#)

### Proteins

UniProt Knowledgebase

Reviewed (Swiss-Prot)  
569,213

Unreviewed (TrEMBL)  
245,871,724

### Species

Proteomes

Protein sets for species with sequenced genomes from across the tree of life

### Protein Clusters

UniRef

Clusters of protein sequences at 100%, 90% & 50% identity

### Sequence Archive

UniParc

Non-redundant archive of publicly available protein sequences seen across different databases

# UniProt – Pokročilé vyhledávání

## Advanced Search



Searching in

UniProtKB

	Gene Name [GN]		
	u Gene Name [GN]	ydj1	Remove
AND	Taxonomy [OC]	human	Remove
AND	Keyword [KW]	chromosomal	Remove
AND	All	a4_human, P05067, cdc7 human	Remove

Add Field

Cancel

Search

Type \* in the search box to search for all values for the selected field.

Výběr databáze

## Find your protein

Pokročilé vyhledávání

UniProtKB ▾

Advanced | List Search

Examples: Insulin, APP, Human, P05067, organism\_id:9606

Název genu  
Název proteinu  
Onemocnění  
Search bar

Vyhledávání pomocí „accessions“ nebo ID

UniProt is the world's leading high-quality, comprehensive and protein sequence and functional info

### Proteins

UniProt Knowledgebase

Reviewed (Swiss-Prot) 569,213

Unreviewed (TrEMBL) 245,871,724

### Species

Proteomes

Protein sets for species with sequenced genomes from across the tree of life

### Protein Clusters

UniRef

Clusters of protein sequences at 100%, 90% & 50% identity

### Sequence Archive

UniParc

Non-redundant archive of publicly available protein sequences seen across different databases

# UniProt – ID mapping

Vyhledávání více proteinů najednou

## Retrieve/ID mapping

Enter one or more IDs (100,000 max). You may also [load from a text file](#). Separate IDs by whitespace (space, tab, newline) or commas.

Q08204  
Q12749

Your input contains 2 IDs

S. cerevisiae (2)

### Taxonomy

[Filter by taxonomy](#)

### Proteins with

3D structure (2)

Binary interaction (2)

Binding site (2)

Chain (2)

Coiled-coil (2)

[More items](#)

Protein existence

[BLAST](#) [Align](#) [Map IDs](#) [Download](#) [Add](#) View: [Cards](#) [Table](#) [Customize columns](#) [Share](#)

2 IDs were mapped to 2 results

<input type="checkbox"/>	From	Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/>	Q08204	Q08204	SMC5_YEAST	Structural maintenance of chromosomes protein 5	SMC5, YOLO34W	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	1,093 AA
<input type="checkbox"/>	Q12749	Q12749	SMC6_YEAST	Structural maintenance of chromosomes protein 6[...]	SMC6, RHC18, YLR383W, L3502.2	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	1,114 AA





BLAST Align Peptide search ID mapping SPARQL

Tools

Release 2023\_01 | Statistics



## Find your protein

UniProtKB ▾

Advanced | List

Search

Examples: Insulin, APP, Human, P05067, organism\_id:9606

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)

**Proteins**  
UniProt Knowledgebase

 **Reviewed**  
(Swiss-Prot)  
569,213

 **Unreviewed**  
(TrEMBL)  
245,871,724

**Species**  
Proteomes

Protein sets for species with sequenced genomes from across the tree of life

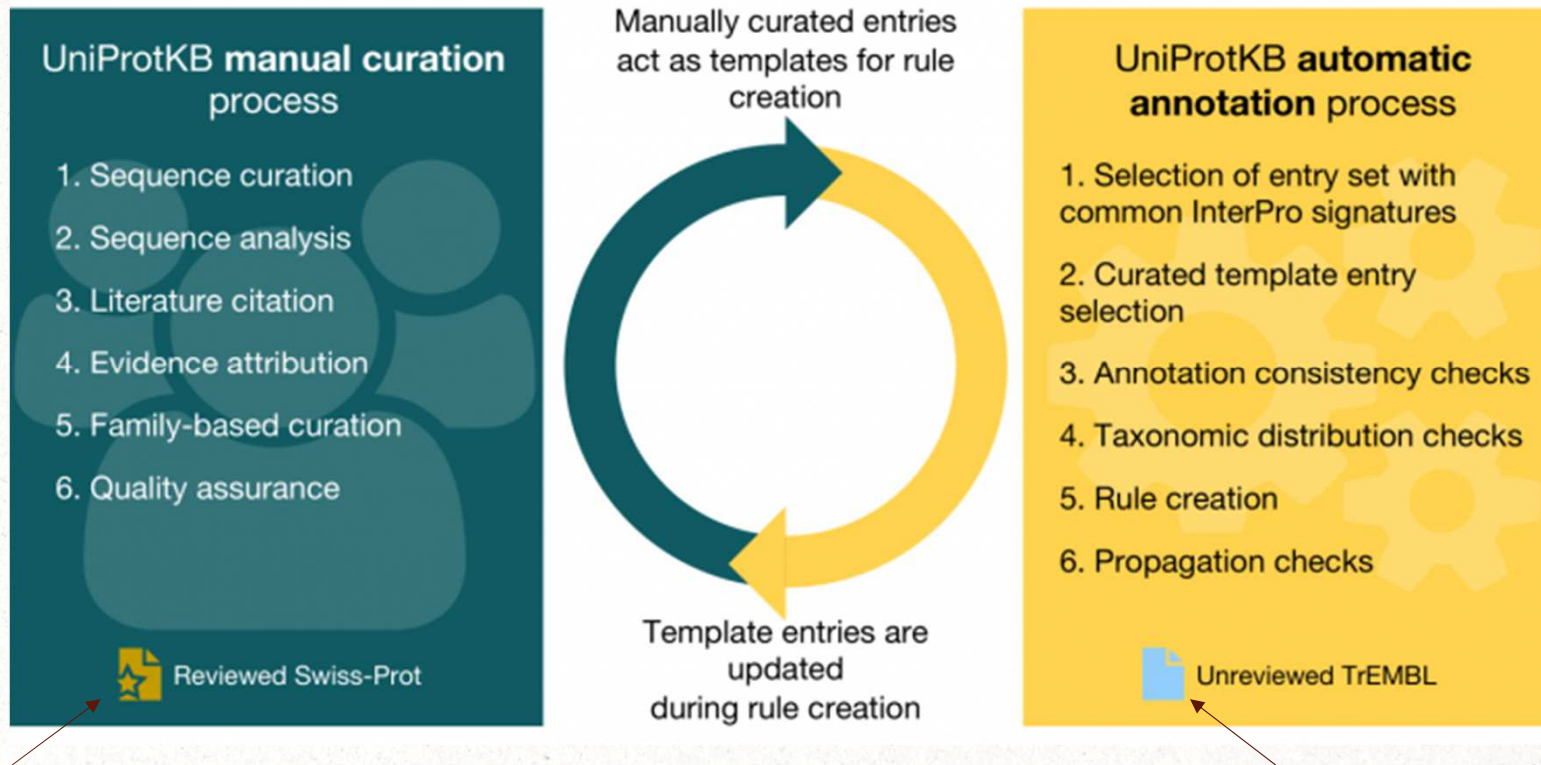
**Protein Clusters**  
UniRef

Clusters of protein sequences at 100%, 90% & 50% identity

**Sequence Archive**  
UniParc

Non-redundant archive of publicly available protein sequences seen across different databases

# Na co si dát pozor



Data, která jsou ručně zkontrolována  
Označení žlutým listem s hvězdičkou

Automaticky anotované

# UniProt – Vyhledávání proteinu

ScSMC5  
protein

Function

Names & Taxonomy

Subcellular Location

Phenotypes & Variants

PTM/Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar Proteins

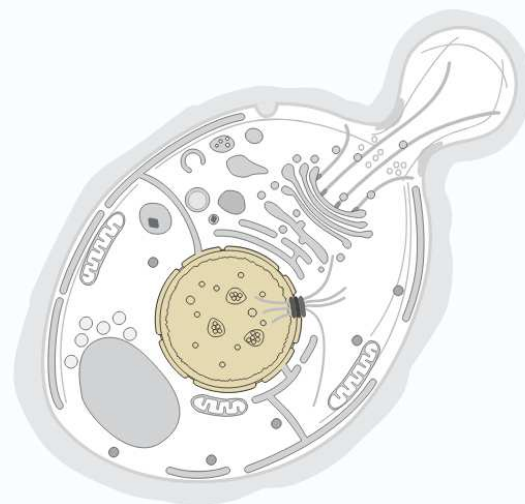
## Function<sup>i</sup>

Acts in a DNA repair pathway for removal of UV-induced DNA damage that is distinct from classical nucleotide excision repair and in repair of ionizing radiation damage. Functions in homologous recombination repair of DNA double strand breaks and in recovery of stalled replication forks.

## Subcellular Location<sup>i</sup>

UniProt Annotation

GO Annotation



Nucleus Curated

Chromosome Curated

## Post-translational modification<sup>i</sup>

Sumoylated by MMS21. 1 Publication

Doporučení: Využit informace z literatury/článků

# UniProt – Vyhledávání proteinu

BRCA1

Function

Names & Taxonomy

Subcellular Location

Disease & Variants

PTM/Processing

Expression

Interaction

Structure

Family & Domains

Sequence & Isoforms

Similar Proteins

## Disease & Variants<sup>i</sup> Breast cancer type 1 susceptibility protein

### Involvement in disease<sup>i</sup>

Breast cancer (BC)

22 Publications

**Note** Disease susceptibility is associated with variants affecting the gene represented in this entry. Mutations in BRCA1 are thought to be responsible for 45% of inherited breast cancer. Moreover, BRCA1 carriers have a 4-fold increased risk of colon cancer, whereas male carriers face a 3-fold increased risk of prostate cancer. Cells lacking BRCA1 show defects in DNA repair by homologous recombination

**Description** A common malignancy originating from breast epithelial tissue. Breast neoplasms can be distinguished by their histologic pattern. Invasive ductal carcinoma is by far the most common type. Breast cancer is etiologically and genetically heterogeneous. Important genetic factors have been indicated by familial occurrence and bilateral involvement. Mutations at more than one locus can be involved in different families or even in the same case.

**See also** MIM:[114480](#)

VAR_007757 61	C>G	in BC and ovarian cancer; no interaction with BAP1; <a href="#">dbSNP:rs28897672</a>	7 Publications
VAR_007758 64	C>G	in BC; no interaction with BAP1; <a href="#">dbSNP:rs80357064</a>	4 Publications
VAR_070460 67	D>Y	in BC; unknown pathological significance; functionally neutral in vitro; <a href="#">dbSNP:rs80357102</a>	1 Publication

# UniProt – Vyhledávání proteinu

|Function

Names & Taxonomy

Subcellular Location

Phenotypes & Variants

PTM/Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar Proteins

## Function<sup>i</sup>

Acts in a DNA repair pathway for removal of UV-induced DNA damage that is distinct from classical nucleotide excision repair and in repair of ionizing radiation damage. Functions in homologous recombination repair of DNA double strand breaks and in recovery of stalled replication forks.

## Subcellular Location<sup>i</sup>

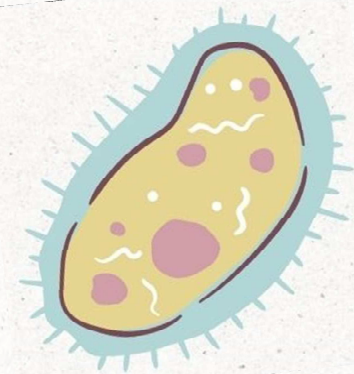
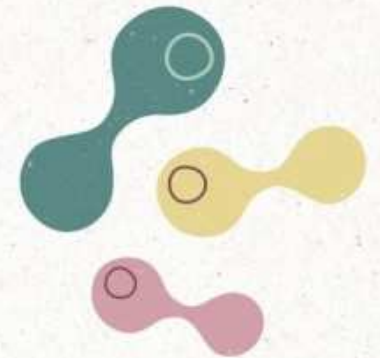
## Post-translational modification<sup>i</sup>

Sumoylated by MMS21.  1 Publication

Probereme během této přednášky

# Přednáška: Analýza proteinu

- Popis a funkce
- Konzervovanost (Alignment)
- Evoluce (fylogenetický strom)
- Zjištění přítomnosti domén daného proteinu



# Alignment

- Porovnání příbuzných proteinů (Multiple sequence alignment)
  - Sekvenční podobnost
  - Strukturní podobnost (Alphafold)
- Souvisí s přítomností konzervovaných domén/aminokyselin
  - Funkční, strukturní souvislosti
  - Evoluční, fylogenetické vztahy
- Strukturní podobnost
  - šroubovice konzervované
  - Hledání podobného „patternu“ – hydrofobní, nabitě aminokyseliny

# Základní alignment

- BLAST – hledání příbuzných sekvencí
  - Lze vybrat i organismy, u kterých bude vyhledávat podobné proteiny
  - Podívat se i na fylogeneticky vzdálenější organismy
  - Grafické rozhraní – napoví o konzervovanosti

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file  Soubor nevybrán. [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

**Choose Search Set**

Databases  Standard databases (nr etc.): [New](#)  Experimental databases [Try experimental clustered nr database](#) [For more info see What is clustered nr?](#)

Compare  Select to compare standard and experimental database [?](#)

**Standard**

Database  [?](#)

Organism   exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude  Models (XM/XP)  Non-redundant RefSeq proteins (WP)  Uncultured/environmental sample sequences

Výběr organismů

Vyloučení organismů

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>





Descriptions

**Graphic Summary**

Alignments

Taxonomy

hover to see the title click to show alignments  Show Conserved Domains

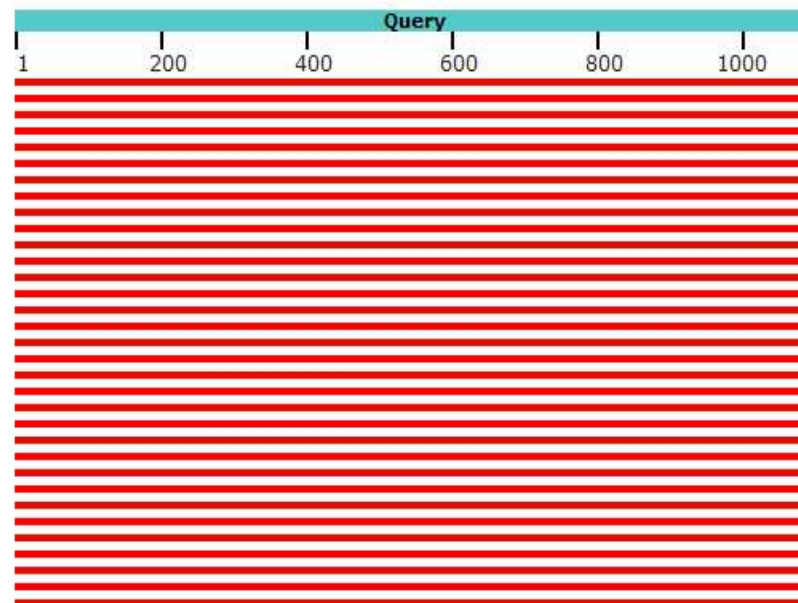
Alignment Scores  < 40  40 - 50  50 - 80  80 - 200  >= 200

100 sequences selected

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of the top 100 Blast Hits on 100 subject sequences



Drtivá většina nalezených sekvencí pochází z kvasinek

Nenapoví to o fylogenezi

Blastování SMC5 sekvence *S. cerevisiae*  
Bez výběru organismů

Descriptions

**Graphic Summary**

Alignments

Taxonomy

hover to see the title click to show alignments  Show Conserved Domains

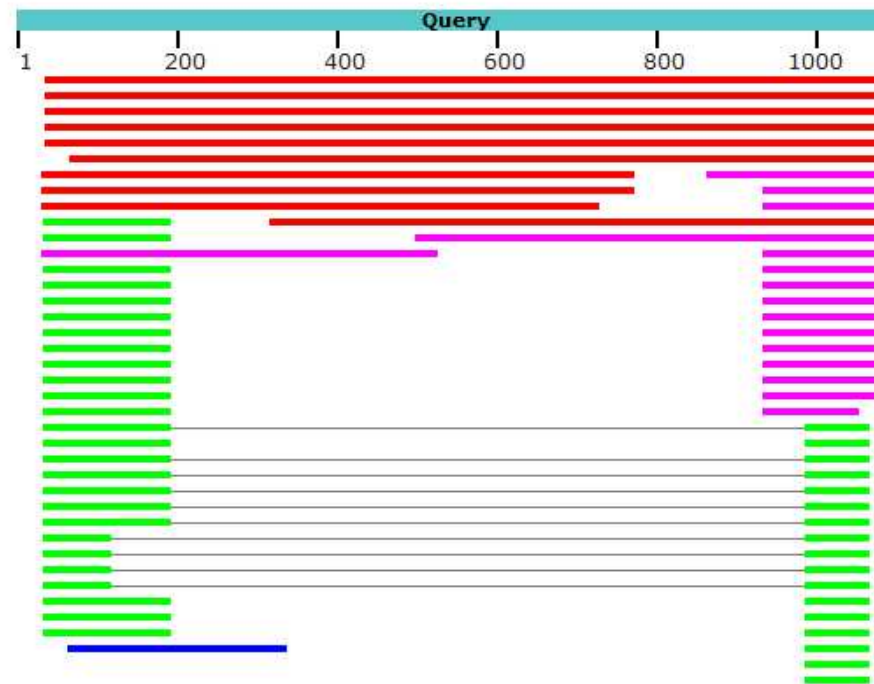
Alignment Scores  < 40  40 - 50  50 - 80  80 - 200  >= 200

60 sequences selected

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of the top 70 Blast Hits on 60 subject sequences



SMC5 proteins

SMC6 proteins

Doporučení:

Vybrat i vzdálenější organismy. Čím více sekvencí, tím lépe

Blastování SMC5 sekvence  
*S. cerevisiae*

Výběr organismů  
*Homo sapiens*  
*Mus musculus*  
*Gallus gallus*  
*Xenopus laevis*  
*Danio rerio*  
*S. cerevisiae*  
*S. pombe*

## Porovnání SMC5 sekvence *S. cerevisiae* a *S. boulardii*

Descriptions | Graphic Summary | **Alignments** | Taxonomy

Alignment view: Pairwise [Restore defaults](#)

3 sequences selected

[Download](#) | [GenPept](#) | [Graphics](#)

**SMC5p Component of the SMC5-SMC6 complex [Saccharomyces boulardii (nom**  
 Sequence ID: [KOH47988.1](#) Length: 1093 Number of Matches: 1  
[See 4 more title\(s\)](#) | [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 1093 [GenPept](#) | [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
2238 bits(5800)	0.0	Compositional matrix adjust.	1091/1093(99%)	1093/1093(100%)	0/1093
Query 1		MTSLIDLGRYVERTHHGEDTEPRSKRVKIAKPDLSFQPGSIIKIRLQDFVITYLTFEFL			60
Sbjct 1		MTSLIDLGRYVERTHHGEDTEPRSKRVKIAKPDLSFQPGSIIKIRLQDFVITYLTFEFL			60
Query 61		SPSLNMIIGPNGSGKSTFVCAVCLGLAGKPEYIGRSKKVEDFIKNGQDVSKIEITLKNSP			120
Sbjct 61		SPSLNMIIGPNGSGKSTFVCAVCLGLAGKPEYIGRSKKVEDFIKNGQDVSKIEITLKNSP			120
Query 121		NVTDIEYIDARDETIKTRIIITRSKRRSDYLINDYQVSESVKTLVAQLNIQLDNLQFL			180
Sbjct 121		NVTDIEYIDARDETIKTRIIITRSKRRSDYLINDYQVSESVKTLVAQLNIQLDNLQFL			180
Query 181		SQERVEEFARLKSVKLLVETIRSIDASLLDVLDELRELQNEQSLQKDLDFKAKIVHLR			240
Sbjct 181		SQERVEEFARLKSVKLLVETIRSIDASLLDVLDELRELQNEQSLQKDLDFKAKIVHLR			240
Query 241		QESDKLRKSVESLRDFQNKKEIELHSQLLPYVVKVDHKEKLNIEEYERAKANLRAIL			300
Sbjct 241		QESDKLRKSVESLRDFQNKKEIELHSQLLPYVVKVDHKEKLNIEEYERAKANLRAIL			300
Query 301		KDKKPFANTKKTLENQVEELTEKCSLKTDFELKAKEKINEIFEKLNIRDEVIKKKNQNE			360
Sbjct 301		KDKKPFANTKKTLENQVEELTEKCSLKTDFELKAKEKINEIFEKLNIRDEVIKKKNQNE			360

## Porovnání SMC5 sekvence *S. cerevisiae* a *Mus musculus*

Descriptions | Graphic Summary | **Alignments** | Taxonomy

Alignment view: Pairwise [Restore defaults](#)

60 sequences selected

[Download](#) | [GenPept](#) | [Graphics](#)

**structural maintenance of chromosomes protein 5 isoform X2 [Mus musculus]**  
 Sequence ID: [XP\\_006527044.1](#) Length: 1086 Number of Matches: 1  
[See 1 more title\(s\)](#) | [See all Identical Proteins\(IPG\)](#)

Range 1: 46 to 1066 [GenPept](#) | [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
340 bits(873)	1e-98	Compositional matrix adjust.	311/1088(29%)	530/1088(48%)	113/1088(10%)
Query 35		SSFQPGSIIKIRLQDFVITYLTFEFLSPSLNMIIGPNGSGKSTFVCAVCLGLAGKPEYIG			94
Sbjct 46		GTFVEGSIVRIAMENFLTYDICEVSPGPHLNMIIGANGTGKSSIVCAICLGLAGKPAFMG			105
Query 95		RSKKVEDFIKNGQDVSKIEITL-KNSPNVTDIEYIDARDETIKTRIIITRSKRRSDYLIN			153
Sbjct 106		RADKVGFFVKGCSKGLVEIELFRTSGNLI-----ITREIDVIKNQSFWFIN			152
Query 154		DYQVSESVKTLVAQLNIQLDNLQFLSQERVEEFARLKSVKLLVETIRSIDASLLDVL			213
Sbjct 153		KKPVTQKIVEEQVAALNIQVGNLCQFLPQDKVGEFALKSKIELLEATEKSVGPPEMHRYP			212
Query 214		-ELRELQNEQSLQKDLDFKAKIVHLRQESDKLRKSVESLRDFQNKKEIELHSQLLPY			272
Sbjct 213		CELKNFREKEKQLETSCKEKTEYLEKMQVQRNERYKQDVERFYERKRHLDLIEMLEAKRPW			272
Query 273		VKVKDHKEKLNIEEYERAKANLRAILKDKKPFAN-----TKKTL-----NQVE			318
Sbjct 273		VEYENVRQYEGVGLIRDRVKEEVRKLEKQIPMTRRIEIDRQRHTLEVRIKESDIIK			332

# Základní alignment

- Pair-Alignment příbuzných sekvencí v BLAST

Align Sequences Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein subjects using a

### Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

MSSDRKRPILDSRSSKRARTSAIRAGDEFLPGNITEIEVHFMITYLYKSK  
PGARLNLVIGPNTGKSSLVCAIGIGLAGEPSLLGRATSIGDYVVKRGEVSG  
SIKITLQDQNPDKKISITRKINKQNKSEWLLLEENKSIHSVTKKEIQEVVARFNI  
QVNNLTQFLPQDRVCEFAKMTPIQLLEETEKAVGDPELSTQHLTLIKKNADL

Query subrange [?](#)

From

To

Or, upload file  Soubor nevybrán. [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

### Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Subject subrange [?](#)

From

To

Or, upload file  Soubor nevybrán. [?](#)

### Program Selection

Algorithm  blastp (protein-protein BLAST)

Choose a BLAST algorithm [?](#)

# BLAST na UniProt



## BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4\_HUMAN or UPI0000000001).

OR

Enter one or more sequences (20 max). You may also [load from a text file](#).

```
SQERVEEFAR LKSVKLLVET IRSIDASLLD VLDELRELQG NEQSLQKDLF FKKAKIVHLR
QESDKLRKSV ESLRDFQNKK GEIELHSQLL PYVKVKDHKE KLNIIYKEEYE RAKANLRAIL
KDKKPFANTK KTLNQQVEEL TEKCSLKTDE FLKAKEKINE IFEKLNITRD EVIKKKNQNE
YYRGRTKKLQ ATIISTKEDE LRSQEILAQT HLPEKSVFED IDIKRKEIIN KEGEIRDLS
EIDAKANAIN HEMRSIQRQA ESKTKSLTTT DKIGILNQDQ DLKEVRDAVL MVREHPMKD
KILEPPIMTV SAINAQFAAY LAQCVDYNTS KALTVVSDS YKLFANPILD KFKVNLRELS
SADTTTPVPA ETVRDLGFEG YLSDFITGDK RVMKMLCQTS KIHTIPVSRRL ELTPAQIKKL
ITPRPNGKIL FKRIIHGNRL VDIKQSAYGS KQVFPTDVTI KQTNFYQGSV MSNEQKIRIE
```

Your input contains 1 sequence

Target database

Restrict by taxonomy

Obnovit

Run BLAST

<https://www.uniprot.org/blast>

# BLAST na UniProt

## BLAST 67 results found in UniProtKB

[Overview](#) [Taxonomy](#) [Hit Distribution](#) [Text Output](#) [Input Parameters](#) [API Request](#)

BLAST [Align](#) [Map IDs](#) [Download](#) [Add](#) [Customize columns](#) [Resubmit](#)

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	
<input type="checkbox"/> Q8CG46	SMCS_MOUSE	Structural maintenance of chromosomes protein 5 [...]	Smc5, Kiaa0594, Smc5l1	Mus musculus (Mouse)	1,101 AA	89.5% (5154)
<input type="checkbox"/> Q8CG46-2	SMCS_MOUSE	Isoform 2 of Structural maintenance of chromosomes protein 5 [...]	Smc5, Kiaa0594, Smc5l1	Mus musculus (Mouse)	1,087 AA	88.3% (5046)
<input type="checkbox"/> Q5ZJY5	SMCS_CHICK	Structural maintenance of chromosomes protein 5 [...]	SMC5, SMC5L1, RCJMB04_14g12, RCJMB04_6o14	Gallus gallus (Chicken)	1,065 AA	63.6% (3487)
<input type="checkbox"/> Q9LFSB	SMCS_ARATH	Structural maintenance of chromosomes protein 5 [...]	SMC5, EMB2782, At5g15920, F1N13.60	Arabidopsis thaliana (Mouse-ear cress)	1,053 AA	31.4% (1329) (1.6e-104)
<input type="checkbox"/> O13710	SMCS_SCHPO	Structural maintenance of chromosomes protein 5 [...]	smc5, spr18, SPAC14C4.02c	Schizosaccharomyces pombe (strain 972 / ATCC 24843) (Fission yeast)	1,076 AA	28.2% (994) (6e-116)
<input type="checkbox"/> Q08204	SMCS_YEAST	Structural maintenance of chromosomes protein 5	SMC5, YOL034W	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	1,093 AA	27.4% (955) (2.6e-110)

### Q8CG46-2 - Isoform 2 of Structural maintenance of chromosomes protein 5 - Mus musculus

"Similarity" highlight  Overview  Wrapped



<https://www.uniprot.org/blast>

# Sekvenční podobnost/alignment na UniProt

## Align results

Overview Trees Percent Identity Matrix Text Output Input Parameters API Request

BLAST Align Map IDs Download Add Resubmit

"Similarity" highlight Select annotation View: Overview  Unannotated

sp|Q08204|SMC5\_YEAST MTS - - - - -  
 sp|O13710|SMC5\_SCHPO - - - - -  
 sp|Q9LF58|SMC5\_ARATH - - - - -  
 sp|Q8CG46|SMC5\_MOUSE MATPSGKAAPPNPQVS  
 sp|Q5ZJY5|SMC5\_CHICK MAV - - - - -

Q08204:Chain

sp|Q08204|SMC5\_YEAST TLTEFNLSPLNMIIG  
 sp|O13710|SMC5\_SCHPO DYCELFPGPYLNLIG  
 sp|Q9LF58|SMC5\_ARATH NHLVCKPGRSLNLVIGPNSGKSSLVCAIALCLGGEPQLLGRATSVGAYVKRGEDSGYVKISLRGN  
 sp|Q8CG46|SMC5\_MOUSE DICEVSPGPHLNMIIGANGTGKSSIVCAICLGLAGKPAFMGRADKVGFFVVRGCSKGLVEIELFRT  
 sp|Q5ZJY5|SMC5\_CHICK DICEVRPGPNLNMIIGANGTGKSSIVCAICLGLAGKPSFLGRAEKVGLFVKQGCLKGLVEIELFKV

Q08204:Chain

sp|Q08204|SMC5\_YEAST PNVTDIEYIDARDETIKITRIITRSKRRSDYLINDYQVSES VVKTLVAQLNIQLDNLCCQLSQERV  
 sp|O13710|SMC5\_SCHPO - - - - -DDETVTITRQISQDKSSSFSINREACATSSITSLMDTFNVQLNNLCHFLPQDRV  
 sp|Q9LF58|SMC5\_ARATH - - - - -TREENTIFRKIDT-RNKS EWMFNGSTVSKKDIVETIQKFNLIQVNLQFLPQDRV  
 sp|Q8CG46|SMC5\_MOUSE - - - - -SGNLIITREIDVIKNQSFWFINKKPVTKIVEEQVAALNIQVGNLCCQLPQDKV  
 sp|Q5ZJY5|SMC5\_CHICK - - - - -PENIITREIQVVTNTSTWHINRKLTTLKTVEEQVAALNIQVDNLCCQLPQDKV

Q08204:Chain

sp|Q08204|SMC5\_YEAST EEFARLKSVKLLVETIRSIDASLLDVL-DELRELQGNEQSLQKDLDFKAKIVHLRQESDKLRKSV  
 sp|O13710|SMC5\_SCHPO AEFALQDPYSRLMETERAIDHEGLLPAHEKLLIDLKREREILQNKNGGSTLNSLKDQRQALEKEV  
 sp|Q9LF58|SMC5\_ARATH CEFAKLTPVQLLEETEKAVGDPQLPVHHRALVEKSRDLKQLERAVAKNGETLNQLKALVDEQEKDV  
 sp|Q8CG46|SMC5\_MOUSE GEFAKLSKIELLEATEKSVGPPPEMHRHYHCELNKRFREKEKQLETSCKEKTEYLEKVMQRNERYKQDV  
 sp|Q5ZJY5|SMC5\_CHICK GEFARMSKIELLEATEKSIGPPPEMYQFHCKLKLKSLKEKERELNVCKDKVNSLEKMKQRAERYKQDV

Q08204:Chain

sp|Q08204|SMC5\_YEAST ESLRDFQNKKGEEIELHSQLLPYVVKVDHKEKLNIIYKEEYERAKA---NLRAILKDKKPFANTKKT  
 sp|O13710|SMC5\_SCHPO NIFKEREKIKSYIEMLGLAKMLVIYREKTNVFNQLRADKKKLLK----KDLKDLVEEFQPIIDKGEE  
 sp|Q9LF58|SMC5\_ARATH ERVRQRELFLTKVDSMKKLPWLKYDMKKA EYMDAKKRMKEAEKLLDEAAKNLNSMKPEIEKQ---  
 sp|Q8CG46|SMC5\_MOUSE ERFYERKRHLDLIEMLEAKRPWVEYENVRQEYEGVKLIRDVRKE---EVRKLLKEGQIPMTRRIEE  
 sp|Q5ZJY5|SMC5\_CHICK DRYHECKRHLDLIDMLQRKRPWVEYETVRQQHEDVKQRDDQAKE---ELKNLKEMQSPLTKKIRE

Q08204:Chain

- tr|A0A2K1IFJ5|A0A2K1IFJ5\_PHYPA
- sp|Q9LF58|SMC5\_ARATH
- sp|Q08204|SMC5\_YEAST
- sp|Q5ZJY5|SMC5\_CHICK
- sp|Q8CG46-2|SMC5\_MOUSE
- sp|Q8IY18|SMC5\_HUMAN

<https://www.uniprot.org/align>



# Multiple Sequence Alignment

<https://www.ebi.ac.uk/Tools/msa/muscle/>

## Multiple Sequence Alignment

MUSCLE stands for **M**Ultiple **S**equence **C**omparison by **L**og-**E**xpectation. MUSCLE is claimed to achieve both better average accuracy and better speed than [ClustalW2](#) or [T-Coffee](#), depending on the chosen options.

**Important note:** This tool can align up to 500 sequences or a maximum file size of 1 MB.

### STEP 1 - Enter your input sequences

Enter or paste a set of sequences in any supported format:

```
>sp|Q08204|SMC5_YEAST Structural maintenance of chromosomes protein 5 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292
GN=SMC5 PE=1 SV=1
MTSLIDLGRYVERTHHGEDTEPRSKRVKIAKPDLSFQPGSIIKIRLQDFVITYTLTEFNL
SPSLNMIIGPNGSGKSTFVCAVCLGLAGKPEYIGRSKKVEDFIKNGQDVSKIEITLKNSP
NVTDIEYIDARDETIKITRIITRSKRRSDYLINDYQVSESVVKTTLVAQLNIQLDNLQFL
SQERVEEFARLKSVKLLVETIRSIDASLLDVLDELRELQGNEQSLQKDLDFKKAKIVHLR
QESDKLRKSVESLRDFQNKKGIEIHSQLLPYVKVKDHKEKLNIIYKEEYERAKANLRAIL
KDKKPFANTKKTLLENQVEELTEKCSLKTDEFLKAKEKINEIFEKLNITIRDEVIKKKNQNE
YYRGRTKKIQATIIUSTKEDELRSEIIL AOTHLPEKSVFEEDIDIKRKEIINKEGEIRDLS
```

Or [upload a file](#):  Soubor nevybrán.

[Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

# Multiple Sequence Alignment

Alignments Result Summary **Phylogenetic Tree** Results Viewers Submission Details

Download Alignment File Show Colors

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

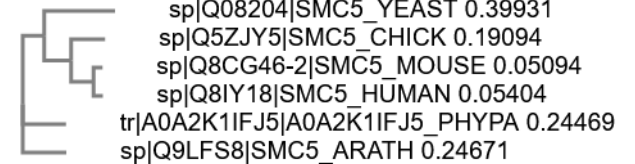
```

sp|Q08204|SMC5_YEAST      -----MTSLIDLGRYVERTHHGEDTE-----PRSKRVKIAKPDLS-SFQPGSIIRLQD
sp|Q5ZJY5|SMC5_CHICK     -----MAVQTRLRAEGSQ-----LRLCDTHHAGKPRS----VEGSIVRIYMEN
sp|Q8CG46-2|SMC5_MOUSE   MATPSGKAAPPNPQVSKRSLPRDASSEVPSKRKNSNPLPTLPRPSGTFVEGSIVRIAMEN
sp|Q8IY18|SMC5_HUMAN     MATPSKKTSTPSPQPSKRALPRDPSSEVPSKRKNSAPQLPLLQSSGPFVEGSIVRISMEN
tr|A0A2K1IFJ5|A0A2K1IFJ5_PHYPA -----MSSDRKRPILDS-----RSSKRARTSAIRAGDEFLPGNITEIEVHN
sp|Q9LFS8|SMC5_ARATH     -----MSE-----RRAKRPKIS--RGEDDFLPGNIIEIELHN
                               .          .          *.* * : :

sp|Q08204|SMC5_YEAST     FVTYTLTEFNLSPSLNMIIGPNGSGKSTFVCAVCLGLAGKPEYIGRSKKVEDFIKNGQDV
sp|Q5ZJY5|SMC5_CHICK     FLTYDICEVVRPGPNLNMIIIGANGTGKSSIVCAICLGLAGKPSFLGRAEKVGLFVKQGCLK
sp|Q8CG46-2|SMC5_MOUSE   FLTYDICEVSPGPHLNMIIIGANGTGKSSIVCAICLGLAGKPAFMGRADKVGFFVKRGCSK
sp|Q8IY18|SMC5_HUMAN     FLTYDICEVSPGPHLNMIVGANGTGKSSIVCAICLGLAGKPAFMGRADKVGFFVKRGCSR
tr|A0A2K1IFJ5|A0A2K1IFJ5_PHYPA FMTYTYLTKSKPGARLNLVIGPNTGKSSLVCAIGIGLAGEPSLLGRATSIGDYVKRGEVS
sp|Q9LFS8|SMC5_ARATH     FMTFNHLVCKPGSRLNLVIGPNTGKSSLVCAIALCLGGEPQLLGRATSVGAYVKRGEVS
*:.*:      .. **::*.**::***:***: : *.* * :**: .: :*:.*

sp|Q08204|SMC5_YEAST     SKIEITLKNSPNVTDIEYIDARDETIKITRIITRSKRRSOYLIND----YQVSESVVKT
sp|Q5ZJY5|SMC5_CHICK     GLVEIELFKVP-----ENIIITREIQVVTNTSTWHINR----KLTTLKTVEEQ
sp|Q8CG46-2|SMC5_MOUSE   GLVEIELFRTS-----GNLIITREIDVIKNQSFWFINK----KPVTQKIVEEQ
sp|Q8IY18|SMC5_HUMAN     GMVEIELFRAS-----GNLVITREIDVAKNQSFWFINK----KSTTQKIVEEK
tr|A0A2K1IFJ5|A0A2K1IFJ5_PHYPA GSIKITLQDQN-----PDKKISITRKINK-QNKSEWLLLENKSIHVSVKKEIQEV
sp|Q9LFS8|SMC5_ARATH     GYVKISLRGNT-----REENLIFRKIDT-RNKSEWFMNG----STVSKKDIVEI
.. :.* *      .: * * * . * : : :      .: .:
    
```

Branch length:  Cladogram  Real



[muscle-l20230503-064457-0144-4283286-p1m.clw](https://www.ncbi.nlm.nih.gov/blast/muscle-l20230503-064457-0144-4283286-p1m.clw)

# Prezentace výsledků o konzervovanosti (inspirace)

- Přenést základní sekvenci do wordu
- Postupně přidávat další sekvence
  - nejlépe i fylogeneticky vzdálenější – napoví o konzervovanosti
- Zaznačit aminokyseliny či pattern, který je podobný
  - Hydrofobní aminokyseliny: L,V,A,M,F,I,W
  - Záporně nabité aminokyseliny: D,E
  - Kladně nabité aminokyseliny: R,K
  - Nenabité aminokyseliny: T,S,C,N,Q
  - Malé aminokyseliny: P,G (otočka)
- Lze proložit i sekundární strukturou

# Alignment CANN domény u podjednotky N5E6

Predikce helixu

```

|OGG--HHHHHHHHH-----HHHHHHHHHHHHHH--HHHHHH--HHHHHHHHH-----HHHHHHHHHHHH-----HHHHHHHHHHHH-----HHHHHHHHHH
Sc---FVNKILKFLMYF-----YSDDNDVREFLKAFTCLILDKVFNAMESDHRLCFKVLLEF-----NEAHFINSYFEIV--DKNDFFLHYRLIQIE-----PHIQSALLRRRFSS-----
Sp---PTEKILKFLMYF-----YKGDG---EEREVFLKSFICLILDNVFNSMEFEFRLCYKILKLF-----DETHFINSYLEIV--CKDDFFLHYRLIQIE-----PIIQNVLLRKLFN
Zb---PVESIMRAFEAY-----SDNCRYFLKSLICFILDAKIWE SMECDAAHACHNVLAKF-----TTQOVLDEYLLN--ERKDYFMHYREVRSF-----PHIQLALVQHLLN
Pp---PLKNLRSILEFIATLFWIRHVQPYYSVVDIESLVIIVHLSLARELLGI---LDKVCDCILALIRYFSPPEEWSEVCFRVTVSISDLT---EGVHNGVFLIGMFSGLG--KRSVDMQRRRALHELKLA--370
At---PPQNIRAWLTLVTTCCQIRCKKPIFTTSQVEQIAEILVLLLDKGLLGL---SILLCECLISVIGSFK---EEWISSCKKIANSASRVFQDINCVRVWESVSGVD--ARSKHRSSTIAHQMLVWLL
Bn---PPQNIRAWLKFVAACQIRCKNPIFTTSQVEQIAEILVLLLDKGLLGL---SILLCECLISVIGSFK---EEWVSSCKEIANSLASRVFQDMNCLRIVESVSGVDT--RSKYRRSSTINQMIVWLL
Xl---PEMQNTNVIKFLGYCTAVC--PEIFSDQDTLLLVVLLIKIYLEKQRDS--FVVDLHVLVGNLHNFR--DWEIKMPELSLALSEISA---HHHNYVRLVQVLPPTESRGRQRCHLSLVFTAKLL
Gg---PESNLMNVKFLDFCTTVV--QDGYTDEEILLLLLLLEKLSLEKQKHV--SLIDFCCLTKLMSIK--DWCTKMPELCLGUSGSS---QHNNLLNIVQLVPSWITRGREVRRLSLVITAKLL
Md---PETNILNVKFLGLCTAIY--PEGYTDKEIILLMLFKMSLEKQKQI--ALVDFCSILLINLIRNFR--DWDTKMPELCLAINESSS---HHNNLLNIVQFVFNWTSRGRQRCHLSLITAKFL
Oa---PENNFNVKFLGLCTAIY--PEGYKDQEIFLLLILFKMSLEKQKQI--PLVDFCSILLINLKNIR--DWKAKMPELCLAINESSS---HHNNLLNIVQLIPNWIARGRQRCHLSLVITAKFL
Dn---PETNILNVKFLGLCTSIH--PEGYQDHEIMLLILMLFKMSLEKQKQI--PLVDFCSILLINLKNIR--DWNKVPPELCLAINESSS---HFNHLLNIVQLVFNWTSRGRQRRECLSLVITAKLL
Mm---PETNILNVKFLGLCTSIH--PEGYQDGEIMLLILMLFKMSLEKQKQI--PLVDFCSILLINLKNIR--DWNKQVHELCLGINESS---HFNHLLNIVQLVFNWTSRGRQRCHLSLVITAKLL
Hs---PETNILNVKFLGLCTSIH--PEGYQDREIMLLILMLFKMSLEKQKQI--PLVDFCSILLINLKNIR--DWNKVPPELCLGINESS---HFNHLLNIVQLVFNWTSRGRQRCHLSLVITAKLL
Alpha--HHHHHHHHH-----HHHHHHHHHHHHHH--HHHH--HHHHHHHHHHHHHH-----HHHHHHHHHHHH-----HHHHHHHHHHHH-----HHHHHHHHHHHHHHHHHHHH
  
```

Hydrofobní aminokyseliny: L,V,A,M,F,I,W

Záporně nabitě aminokyseliny: D,E

Kladně nabitě aminokyseliny: R,K

Nenabitě (neutrální) aminokyseliny: T,S,C,N,Q

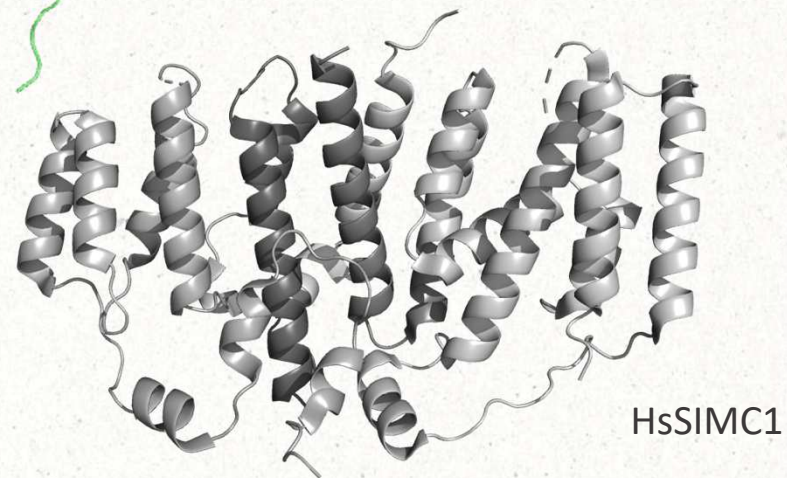
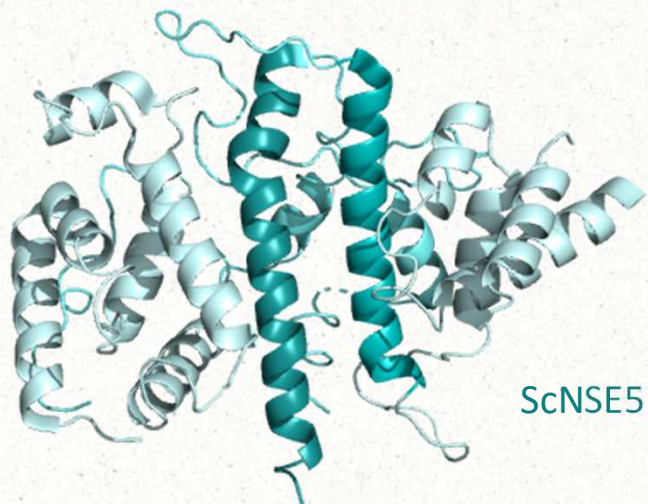
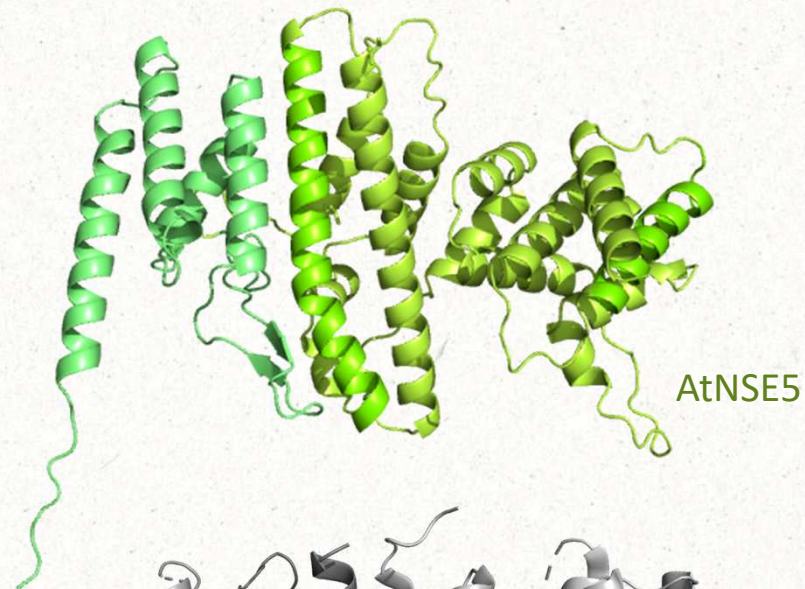
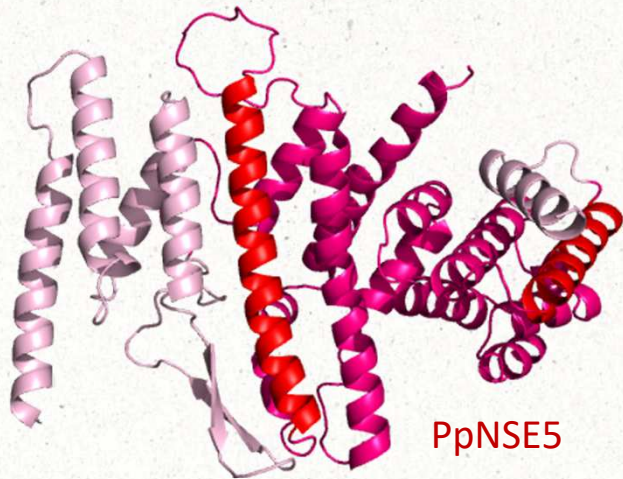
Malé aminokyseliny: P,G (otočka)

doc. Jan Paleček  
(Lelkes et al. 2023)

# Pokročilý alignment

- BLAST nenajde příbuzné sekvence s vaším proteinem
- Nalezení jiného proteinu, který by měl být homologem/ortologem
- Alignment příbuzných sekvencí v BLAST
- Přenesení výsledků do dokumentu (word)
- Postupné přidávání dalších sekvencí
- Využití informace
  - sekundární struktura (PSIPRED)
  - terciální struktura (Alphafold)
- Značení patternu

# Strukturní podobnost podjednotky NSE5



AlphaFold  
doc. Jan Paleček  
(unpublished)

# Alignment podjednotky NSE5

```

-----190-----HHHHHHHHHHHH-----230--HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-----HHHHHH--HHHHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHH--
Pp187-TQK---ANEKFTRVITDLYSLLENADEDGRQESKKKGDFGVDSASKNEDPLNRLVQFLMLRWLSLFESLLRVRHQAFQGNPE---NSELQESLIFRFMGQF-----NVKTELTINLRLRARVDKLDDLNDG-312
Cp ---TQKS-----EESAKVTKDLYLLGNADDEGKETLKKRG--ASDFAVRNDDPQNRLVQFLVLRWLSLCSDFQVRQQAFQDNPE---KLDLQESLIREMMKQD-----NSTKKLENLRIAKPGELDDLKGG
Ps --AYSDRNSTNGDLDELNPNGFSLLTQSAQAADTGF-----KLSGIKFLADMFLFQYLKVLERDFLRNSVYKESS-----EWSLRESLINTLGSR-----INVSLFKDCLSILCKGSHAGFFDIP
Ls ---TEFSPNEKEDHGKTSGSLDATAHDIKDVEVADEATT-----EVLDIKFLRNMLLFQYLSFESGDFVIRNLAFTENS-----DWVTRESLINMLIVSR-----ITYTLIKDCLSAMCQLSQFSMDSSNV
At ---SDVD-----SDEKDLSVRSLFFQQLIQNNKRVNES-----EESDLKILGNMFLEKYLAHVLKLDFTFERNQVYEETM-----NWSLAKESINLLIASR-----VNFLLMKDLSTMCASIDADEKS
Br ---SDVD-----SGEKDISVRSSLFQQLIQNNKGVES-----EESDLKILGNMFLEKYLAHVLKLDFTFERNQVYEETM-----NWSLAKESINLLIASR-----VNFLLMKDLSTMCSPDADEKS
Me ---LDIYGEKESAVKSYCGILDSSVHLIQDAEVVNETDL-----QKMQIKSLGKMLLFQYLINVLEGDFVERNKAYEETM-----NWMLRESLISMLISSR-----INVVLVKDCLSINCGLGQFNSELSD
Pa ---LENASREREAGGRSISGPLDSIGLQMQEAEVANEGNL-----QALEIKSLRNMLLFQYLNAEEGDELRNRVYEETM-----NWTHRESLINMLISAR-----INVYGLMKDCLSINCLFDVSAGISE
Nt -----LDRDDEINKGSSGSIDSVGFHSIQELGKIVNQKS-----KVLETEILRKMLLHYLVSVLEADFVERNYAFKEKM-----RWTLRDSLNMLISSR-----IIYGLVKDCLLVICDVFSDLSDQTES
Gs ---VDKATAVSEAGGKQSGGSLDHSSPQAIIEEAEILSETKF-----EAASMEPLRNMLIFQYLVVEDDELRNATL-----NWSMQRESLSLISSR-----INVYSLMKYFMAIICQLSQLQEAKDDYE
Zm ---NGAVSSEIGNVRSDHLFDSRSSLAIEAVEPINGTD-----DNNGIKAIENMLFQYLVNTEADFVRQIATKESL-----DWVIRESLISMLISSR-----LVFSSVKNCMCILLNQYHCHKDLKD
Dc ---SGNSHVEASGICRDSKSLLDSASSIVEETAIG-----RHSAIKTVQNMLLFQYLRAEADELRHTLYKETL-----NWVFRESVLSLVASR-----MNYSLVHNCSVSTRFIEANI
Xs ---DAVCSERQAARANSIGPPDFSGETLIQELGGVVNGTNF-----QAVDTEYLRNMLLQYLNVLEGDELF-----QTS-----NWNLRESLINLIGSR-----INVYVLMKDCLSISGLQARAGINNDHG
Rg -----MVLLQYLISVLEGDFVERNSVYKENM-----DWNLRVSLNMLISSR-----ISYDLIKGCLSICESELSKHPVL
Cc -----TVQWDLAYHENKDPHRKHSLTK-----NDRLLFIQYVQTEDDFQLKHDM-----PQCTIAKTVISCDDK-----FSNVCVIKNLIDISMSHLLAE
Ss -----TTIFWDWELITSVMAEQSRGRRLCC-----EVVRMLLQYVQTEDDERVKLSFRD-----HHSIAKATISCDVR-----FTQVADVINWLEAIVKDSEDTQ
Dr -----STVWDWELKSVKDQDDTWRLRT-----EVQVMLLQYVQEDDEHLKLSTQC-----QHSLAKKVISCGET-----FGQVADVISWIMTAKESVKHSK
Xl ---DAVTWYDLSNVENKEG-----FVLCLFIQYVQTEDDFQLCLQRRC-----HTCLSKVMSWDF-----NVSHIADVIKWLIKTINDSQIGLE
Gg -----TTVGWDMLKYTENEEKL-----PGRLLFIQYVQTEDDFQQNLRLHL-----QKSITKKVISCDT-----CFNNVEVWELVAVTGIRFFQL
Oa ---NSVQWNEMRSVEAQAPP-----PRRLYFLRYVQTEDDFQNLRRHWQ-----LQGSIVSSVISCDF-----QPNNIADVISWLREVTGMGTDQD
Md ---ASVEWDKLIYVEEKQEEM-----PGRIFFLRYVQTEDDFQQVLRHQQH-----LQQSIACTMISCDF-----QPQNIDITIQWLGATRGGFQKE
La ---KTVEWDKLIYVEEEGQNL-----PGRVLFLRYVQTEDDFQQMLRRQRQH-----LQQSIASTVISCDF-----QPHNVDVIKWLKVTEDELTEP
Mm ---KTVEWDKLIYVEEEGQTL-----PGRVLFLRYVQTEDDFQQILRRQRQH-----LQQSIANTVISCDF-----QPHNVDVIKWLKATENELTP
SIMC1---KTVEWDKLIYVEEEGQTL-----PGRVLFLRYVQTEDDFQQTLRRQRQH-----LQQSIANMVISCDF-----QPHNVDVIKWLKATEDGLTQP
-----HHHHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-----HH--HHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHH-----

```

doc. Jan Paleček  
(unpublished)

# Zjištění konzervovaných domén proteinu

- NCBI Conserved domain search <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>
- UniProt <https://www.uniprot.org/>
- Protein může být popsán a charakterizovaný v publikacích
  - Příbuzné organismy - porovnání
- Interakční motivy – kratší než domény (jednotky aminokyselin)
  - Online tools/programy na jejich vyhledávání



# Zjištění konzervovaných domén proteinu

SUMO-TARGETED UBIQUITIN E3 LIGASE 2

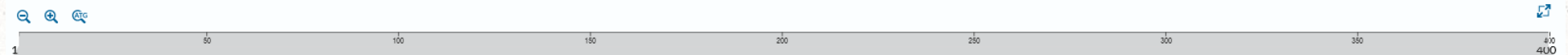
STUbl2

UniProt

## Family & Domains<sup>1</sup>

### Features

Showing features for domain<sup>1</sup>, region<sup>1</sup>, compositional bias<sup>1</sup>.



TYPE	ID	POSITION(S)	DESCRIPTION	
<input type="text" value="-- Select --"/>				
▶ Domain		1-84	BRCT <small>InterPro Annotation</small>	BLAST <a href="#">Add</a>
▶ Region		148-219	Disordered <small>Automatic Annotation</small>	BLAST <a href="#">Add</a>
▶ Compositional bias		184-219	Basic and acidic residues <small>Automatic Annotation</small>	BLAST <a href="#">Add</a>
▶ Region		270-291	Disordered <small>Automatic Annotation</small>	BLAST <a href="#">Add</a>
▶ Compositional bias		272-291	Basic and acidic residues <small>Automatic Annotation</small>	BLAST <a href="#">Add</a>
▶ Domain		300-345	RING-type <small>InterPro Annotation</small>	BLAST <a href="#">Add</a>

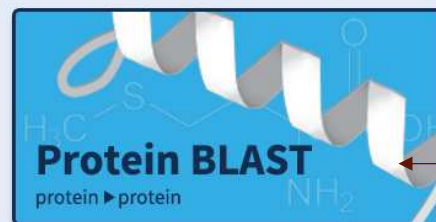
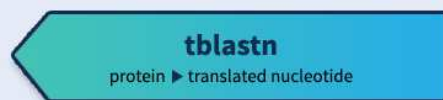
Expand table

## Keywords<sup>1</sup>

Domain [#Zinc-finger](#) Automatic Annotation

# Zjištění konzervovaných domén proteinu

## Web BLAST



Zadat sekvenci proteinu

## Program Selection

### Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm ?

Descriptions

**Graphic Summary**

Alignments

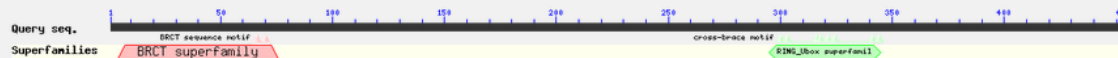
Taxonomy

hover to see the title click to show alignments  Show Conserved Domains

Alignment Scores  < 40  40 - 50  50 - 80  80 - 200  >= 200 ?

100 sequences selected ?

Putative conserved domains have been detected, click on the image below for detailed results.



Propojení na NCBI  
Conserved domains

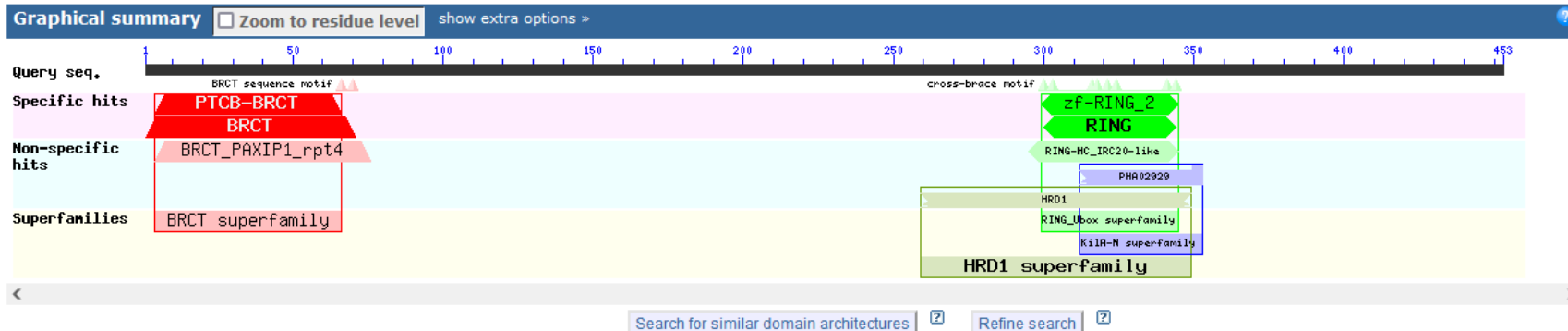
# Zjištění konzervovaných domén proteinu



## Conserved domains on [cl|Query\_17726]

View **Standard Results** ?

Local query sequence



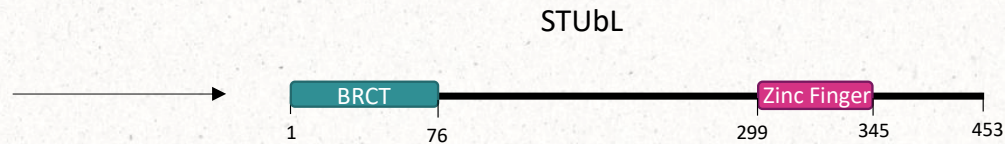
### List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	BRCT_PAXIP1_rpt4	cd17730	fourth BRCT domain of PAX-interacting protein 1 (PAXIP1) and similar proteins; PAXIP1, also ...	4-76	4.20e-23
[+]	PTCB-BRCT	pfam12738	twin BRCT domain; This is a BRCT domain that appears in duplicate in most member sequences. ...	4-66	2.16e-22
[+]	RING-HC_IRC20-like	cd23135	RING finger, HC subclass, found in Saccharomyces cerevisiae increased recombination centers ...	295-345	3.01e-17
[+]	zF-RING_2	pfam13639	Ring finger domain;	299-345	2.02e-09
[+]	RING	smart00184	Ring finger; E3 ubiquitin-protein ligase activity is intrinsic to the RING domain of c-Cbl and ...	300-344	2.77e-09
[+]	BRCT	smart00292	breast cancer carboxy-terminal domain;	1-71	5.57e-08
[+]	PHA02929	PHA02929	N1R/p28-like protein; Provisional	312-353	1.24e-05
[+]	HRD1	COG5243	HRD ubiquitin ligase complex, ER membrane component [Posttranslational modification, protein ...	259-349	6.99e-05

<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

# Schémata genů/proteinů

- Powerpoint
- BioRender
- CorelDRAW



## Schémata genů

- Zaznačení exonů, intronů, UTR sekvencí
- Místa mutací – původní nukleotid/pozice/ změněn na jaký nukleotid (A325G)
- Jaké mají mutace efekt na proteinové úrovni?

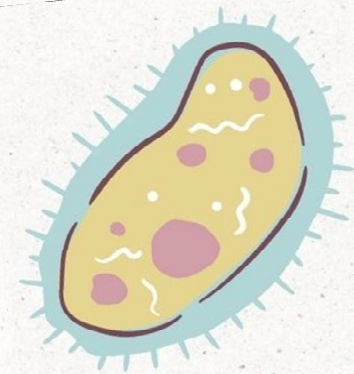
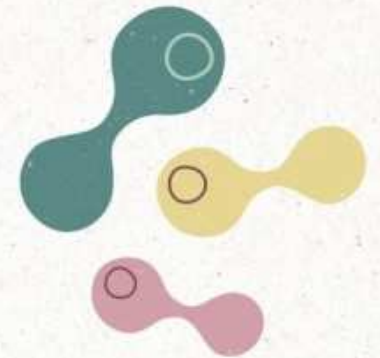
## Schémata proteinů

- Domény
- Lze zaznačit i efekt mutací

Při publikacích pozor, ať obrázky neztrácí kvalitu při zoomu.

# Přednáška: Analýza proteinu

- Popis a funkce
- Konzervovanost (Alignment)
- Evoluce (fylogenetický strom)
- Zjištění přítomnosti domén daného proteinu
- Struktura proteinu – Alphafold/Colabfold



# Predikce sekundární struktury

- PSIPRED: <http://bioinf.cs.ucl.ac.uk/psipred>

## Submission details

### Protein Sequence

### Help...

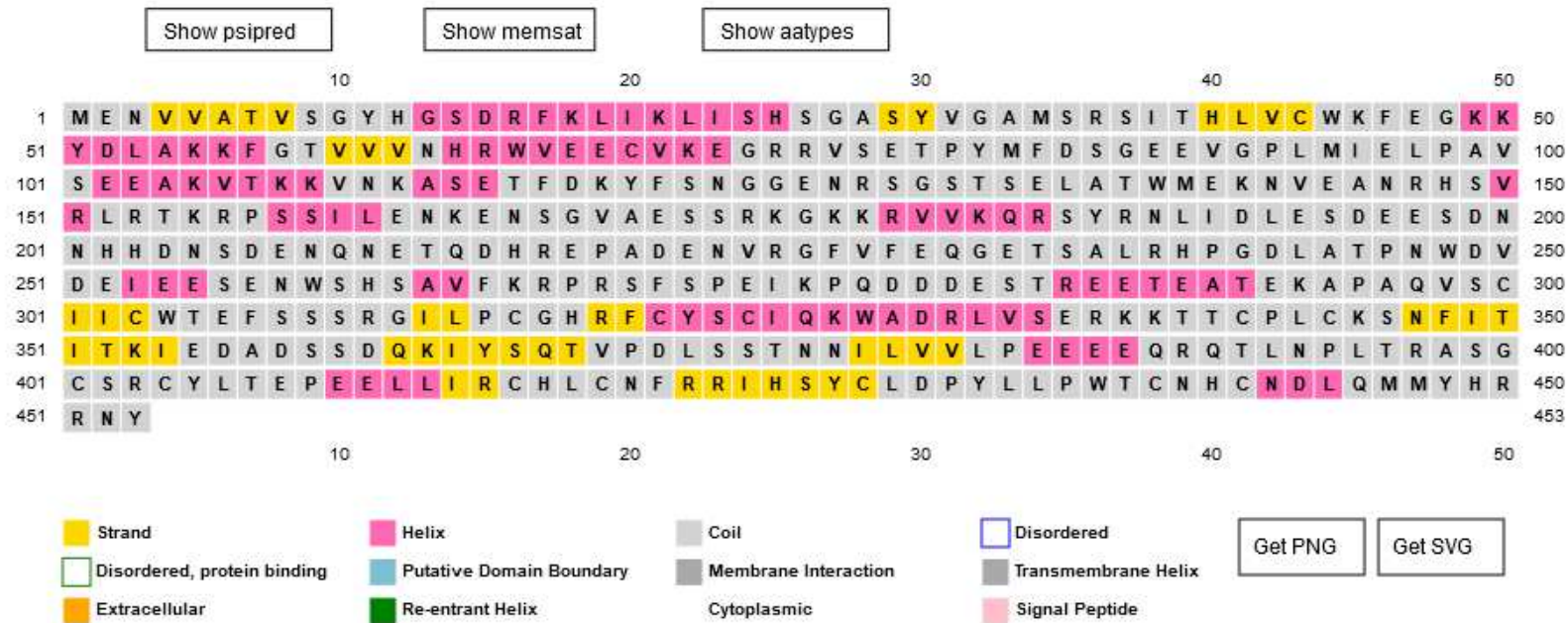
If you wish to test these services follow this link to retrieve [a test fasta sequence](#).

### Job name

### Email (optional)

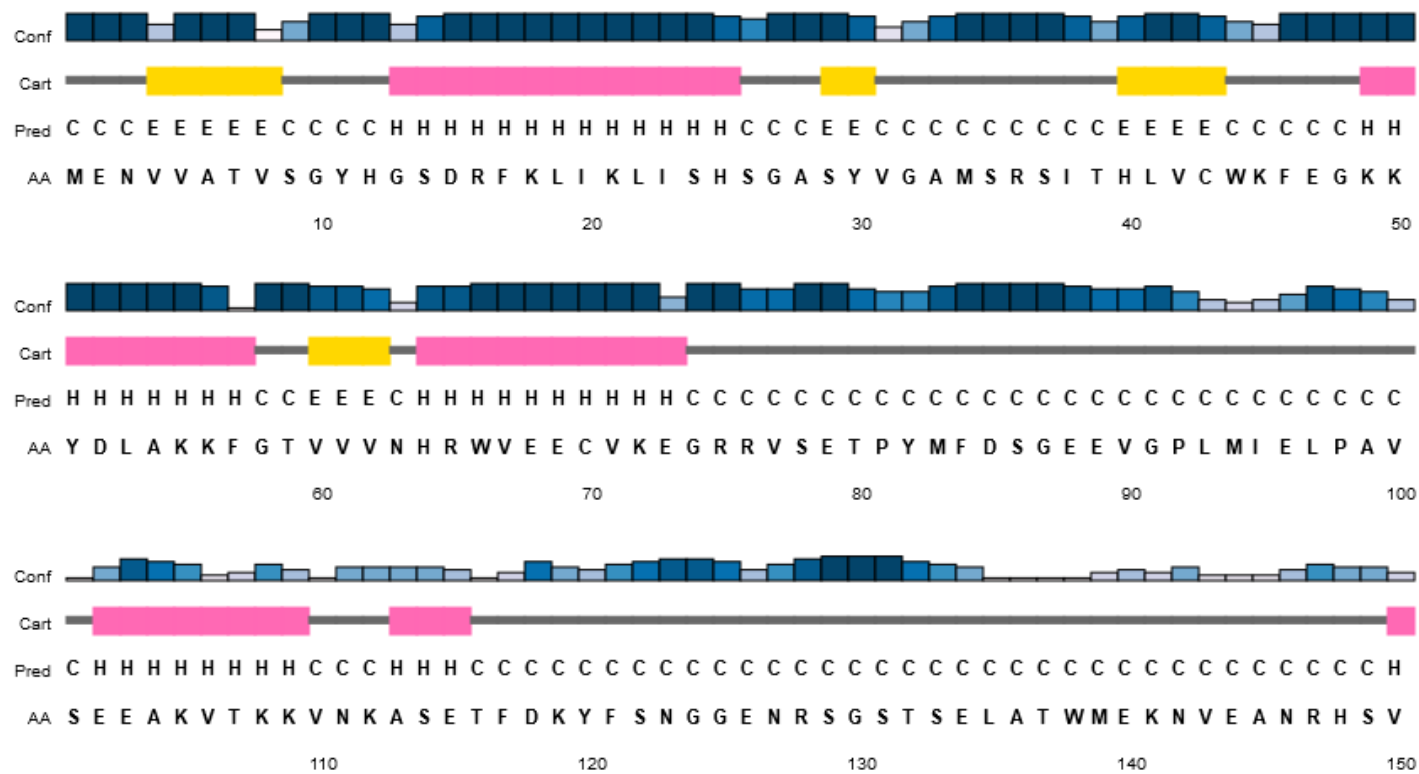
# PSIPRED – predikce sekundární struktury

Sequence Plot



# PSIPRED – predikce sekundární struktury

PSIPRED Cartoon



**Legend:**

- Strand (Yellow)
- Helix (Pink)
- Coil (Grey)

**Conf:** - + Confidence of prediction  
**Cart:** 3-state assignment cartoon  
**Pred:** 3-state prediction  
**AA:** Target Sequence



# AlphaFold – 3D struktury proteinu

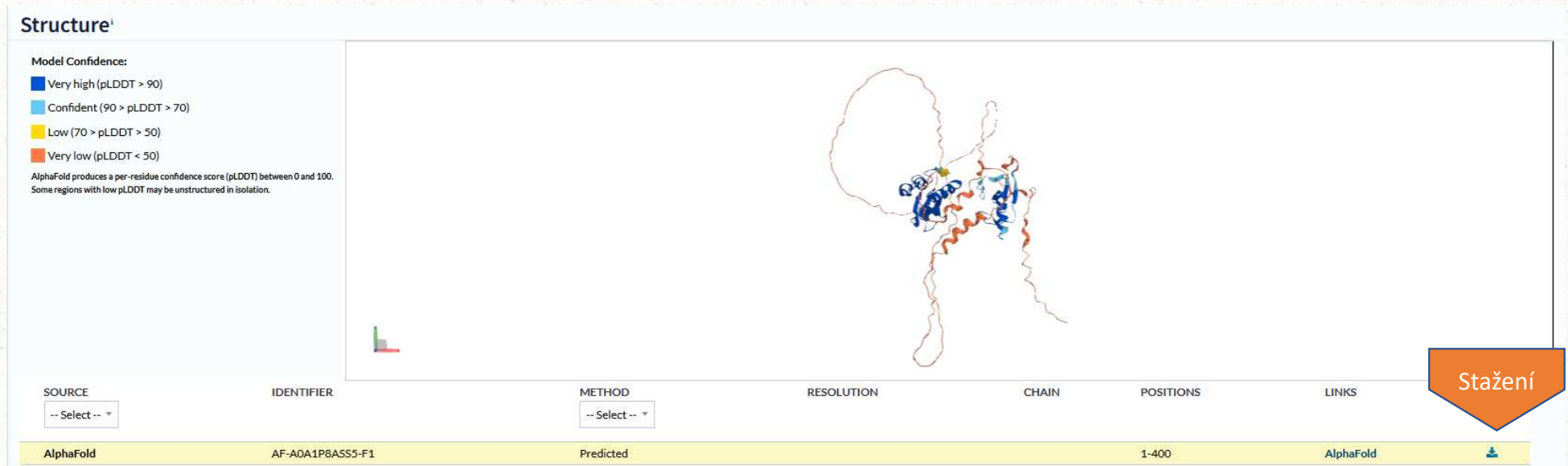
1) Využití databáze PDB (stažení PDB modelu) <https://www.rcsb.org/>



The screenshot shows the top navigation bar of the RCSB Protein Data Bank website. On the left is the PDB logo. In the center, it displays '203,607 Structures from the PDB' and '1,068,577 Computed Structure Models (CSM)'. On the right, there is a search bar with a dropdown menu set to '3D Structures', a search input field containing the placeholder text 'Enter search term(s), Ent', and a search button with a magnifying glass icon. Below the search bar are links for 'Advanced Search | Browse Annotations' and a 'Help' link.

2) UniProt (lze stáhnout)

<https://www.uniprot.org/uniprotkb/Q9ZW89/entry>



The screenshot shows the UniProt 'Structure' page for protein Q9ZW89. On the left, there is a 'Model Confidence' legend with four categories: 'Very high (pLDDT > 90)' in blue, 'Confident (90 > pLDDT > 70)' in light blue, 'Low (70 > pLDDT > 50)' in yellow, and 'Very low (pLDDT < 50)' in orange. Below the legend is a note: 'AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.' The main area contains a 3D ribbon diagram of the protein structure, colored according to the confidence scale. At the bottom, there is a table with columns: SOURCE, IDENTIFIER, METHOD, RESOLUTION, CHAIN, POSITIONS, LINKS, and a download icon. A yellow button labeled 'Stažení' (Download) is positioned over the download icon.

SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS	
-- Select --	AF-A0A1P8ASS5-F1	-- Select --			1-400	AlphaFold	📄

# AlphaFold – 3D struktury proteinu

3) EMBL – EBI

AlphaFold Protein Structure Database

<https://alphafold.ebi.ac.uk/>

## AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#) Help: [AlphaFold DB search help](#)

Feedback on structure: [Contact DeepMind](#)

# AlphaFold – predikce 3D struktury proteinu

- Využití online predikce AlphaFold
  - LatchBio: <https://console.latch.bio/workflows>

The screenshot displays the 'Workflows' page on the LatchBio console. At the top right, there is an 'Upload Workflow' button. Below the title, there are tabs for 'Workflows' (selected) and 'All Executions'. A search bar and a dropdown menu for sorting by 'Executions Ran' are also visible. Two workflow cards are shown:

- AlphaFold2** (Verified) with 10268 executions. Description: 'AlphaFold produces highly accurate protein structure predictions'. Created by Deepmind, version v2.3.1+1.
- ColabFold** (Verified) with 5572 executions. Description: 'The ColabFold version of AlphaFold2 is optimized for extremely fast predictions on small proteins. It uses the same basic architecture as AlphaFold2, but optimizes the sequence search procedure.' Created by Sergey D., version 0.0.48-fa4cfc.

- Umožňuje i predikci 3D struktury proteinového komplexu

# LatchBio AlphaFold

Inputs Outputs +1 v2.3.1+1

features.pkl

msas

**ranked\_0.pdb**

ranking\_debug.json

relax\_metrics.json

relaxed\_model\_1\_pred\_0.pdb

result\_model\_1\_pred\_0.pkl

timings.json

unrelaxed\_model\_1\_pred\_0.pdb

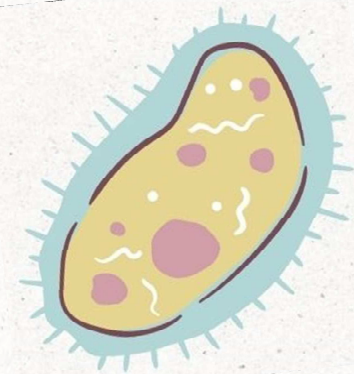
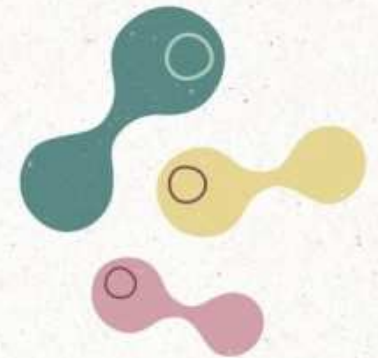
Executions

Workflow Version	Ran By	Date	Status
v2.3.1+1	Jitka Jemelková	4. 4. 2023 11:48:07	<span>Running</span>
v2.3.1+1	Jitka Jemelková	4. 4. 2023 10:34:28	<span>Failed</span>
v2.3.1+1	Jitka Jemelková	4. 4. 2023 11:48:07	<span>Succeeded</span>

- **unrelaxed\_model\_x.pdb**: output
- **relaxed\_model\_x.pdb**: after performing an Amber relaxation procedure on the unrelaxed structure prediction
- **ranked\_x.pdb**: A PDB format containing the relaxed predicted structures, after reordering by model confidence.
  - ranked\_0.pdb should contain the prediction with the highest confidence
  - ranked\_4.pdb the prediction with the lowest confidence

# Přednáška: Analýza proteinu

- Popis a funkce
- Konzervovanost (Alignment)
- Evoluce (fylogenetický strom)
- Zjištění přítomnosti domén daného proteinu
- Struktura proteinu – Alphafold/Colabfold
- Zjištění interakčních partnerů



# Databáze protein-proteinových interakcí

Informace o interakcích v literatuře  
Interakční databáze - odkazy v UniProt

## Interaction<sup>i</sup>

### Subunit<sup>i</sup>

Component of the Smc5-Smc6 complex which consists of KRE29, MMS21, NSE1, NSE3, NSE4, NSE5, SMC5 and SMC6.

1 Publication

## Protein-protein interaction databases

BioGRID | 34366 [↗](#) 225 interactors

ComplexPortal | CPX-1364 [↗](#) SMC5-SMC6 SUMO ligase complex

DIP | DIP-1979N [↗](#)

IntAct | Q08204 [↗](#) 18 interactors

MINT | Q08204 [↗](#)

STRING | 4932.YOL034W [↗](#)

Protein je součástí komplexu

BioGRID  
IntAct  
MINT  
STRING

Obsahují podobné informace  
Nutné vyfiltrovat převážně fyzické interakce  
(protein-protein), binární

## Binary interactions<sup>i</sup>

Q08204 has binary interactions with 7 proteins

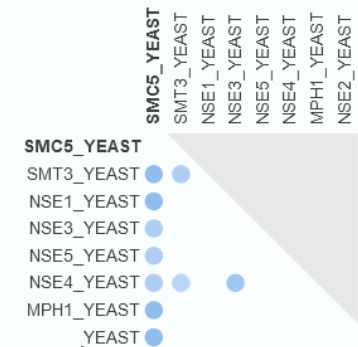
Filter

Subcellular location

Select...

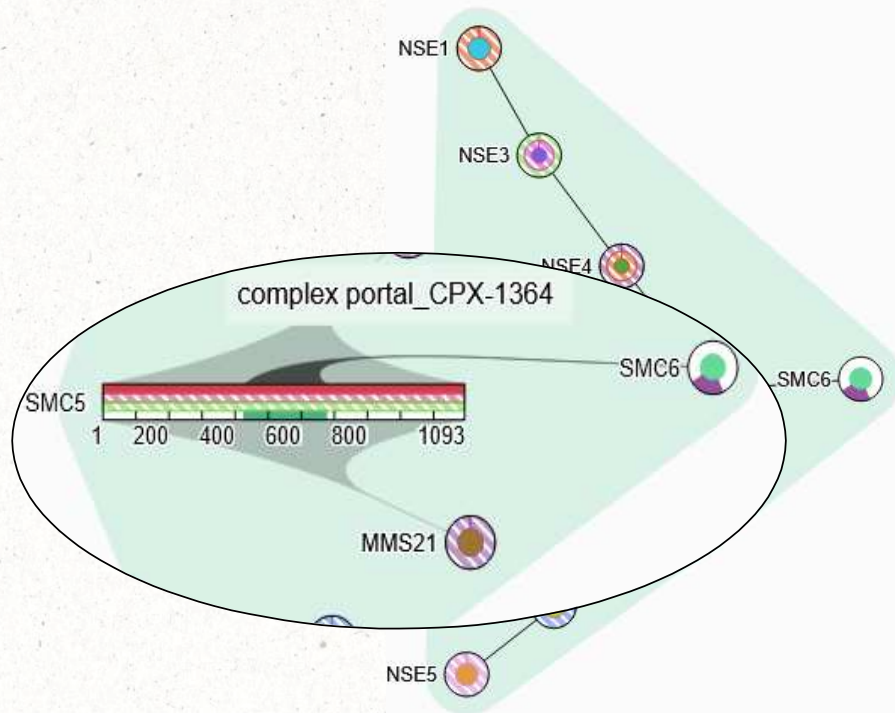
Diseases

Select...



# ComplexPortal

Protein je součástí komplexu



Legend	Description
	<b>protein - MMS21 (enzyme)</b> P38632 <a href="#">↗</a> E3 SUMO-protein ligase MMS21
	<b>protein - NSE4 (unspecified role)</b> P43124 <a href="#">↗</a> Non-structural maintenance of chromosome element 4
	<b>protein - SMC6 (unspecified role)</b> Q12749 <a href="#">↗</a> Structural maintenance of chromosomes protein 6
	<b>protein - KRE29 (unspecified role)</b> P40026 <a href="#">↗</a> DNA repair protein KRE29
	<b>protein - NSE5 (unspecified role)</b> Q03718 <a href="#">↗</a> Non-structural maintenance of chromosome element 5

<https://www.ebi.ac.uk/complexportal/home>

## Filtry – typ interakce Detekční metoda interakce – binární, koimunoprecipitace?

Filters: Interactor Species | Interactor Type | Interaction Type | Interaction Detection Method | Interaction Host Organism | Mutation | Expansion | Positive | MI Score

Export: Network | Table

### Network Tools

Redraw Network

### Interaction Network

Interactor Name

```
graph TD; SMC5 --- SPC24; SMC5 --- MPH1; SMC5 --- COG4; SMC5 --- URE2; SMC5 --- SMT3; SMC5 --- MUK1; SMC5 --- HUA2; SMC5 --- NSE4; SMC5 --- ATG17; SMC5 --- ATG19; SMC5 --- SAP1; SMC5 --- CHA4
```

### Legend

**Nodes**  
Color ~ Species  
● *Saccharomyces cerevisiae*  
Shape ~ Type  
● protein

**Edges**  
Color ~ MI Score  
Width ~ #Evidence


Layout:  
 Force directed  
 Circular  
 Bubbles




Edges:  
 Expand  
 Affected By Mutation

Group By:  
 Species



# IntAct



Select	Molecule A	Molecule B	Identifier A	Identifier B	Type A	Type B	Species A	Species B	Host Organism	Positive interaction	Detection Method	Publication IDs	Interaction Type
<input type="checkbox"/> 	SMC5	NSE4	<a href="#">UniProt Q08204</a>	<a href="#">UniProt P43124</a>	protein	protein	<a href="#">Saccharomyces cerevisiae</a>	<a href="#">Saccharomyces cerevisiae</a>	<a href="#">Saccharomyces cerevisiae (Baker's yeast)</a>	✓	2 hybrid	<a href="#">18719252</a>	<a href="#">physical association</a>
<input type="checkbox"/> 	SMC5	SMT3	<a href="#">UniProt Q08204</a>	<a href="#">UniProt Q12306</a>	protein	protein	<a href="#">Saccharomyces cerevisiae</a>	<a href="#">Saccharomyces cerevisiae</a>	<a href="#">Saccharomyces cerevisiae (Baker's yeast)</a>	✓	2 hybrid	<a href="#">18719252</a>	<a href="#">physical association</a>
<input type="checkbox"/> 	SMT3	SMC5	<a href="#">UniProt Q12306</a>	<a href="#">UniProt Q08204</a>	protein	protein	<a href="#">Saccharomyces cerevisiae</a>	<a href="#">Saccharomyces cerevisiae</a>	<a href="#">Saccharomyces cerevisiae (Baker's yeast)</a>	✓	pull down	<a href="#">25764370</a> <a href="#">10.1371/journal.pbio.1002089</a>	<a href="#">physical association</a>



Publikace

# AlphaFold – predikce 3D struktury proteinového komplexu

- Využití online predikce AlphaFold pro proteinový komplex
  - LatchBio: <https://console.latch.bio/workflows>

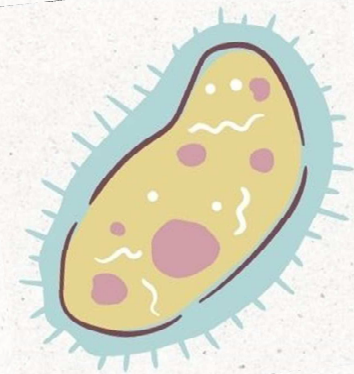
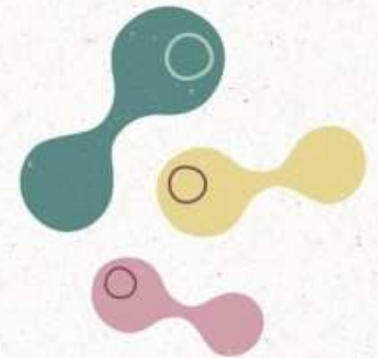
The screenshot displays the 'Workflows' section of the LatchBio console. At the top right is an 'Upload Workflow' button. Below it are tabs for 'Workflows' (selected) and 'All Executions'. A search bar and a 'Sort: Executions Ran' dropdown are also present. Two workflow cards are visible:

- AlphaFold2** (Verified): AlphaFold produces highly accurate protein structure predictions. Created by Deepmind, version v2.3.1+1. It has 10268 executions.
- ColabFold** (Verified): The ColabFold version of AlphaFold2 is optimized for extremely fast predictions on small proteins. It uses the same basic architecture as AlphaFold2, but optimizes the sequence search procedure. Created by Sergey D., version 0.0.48-fa4cfc. It has 5572 executions.

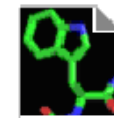
- Umožňuje i predikci 3D struktury proteinového komplexu
- Lze vložit i sekvence dvou proteinů – umožní predikci 3D modelu proteinového komplexu

# Přednáška: Analýza proteinu

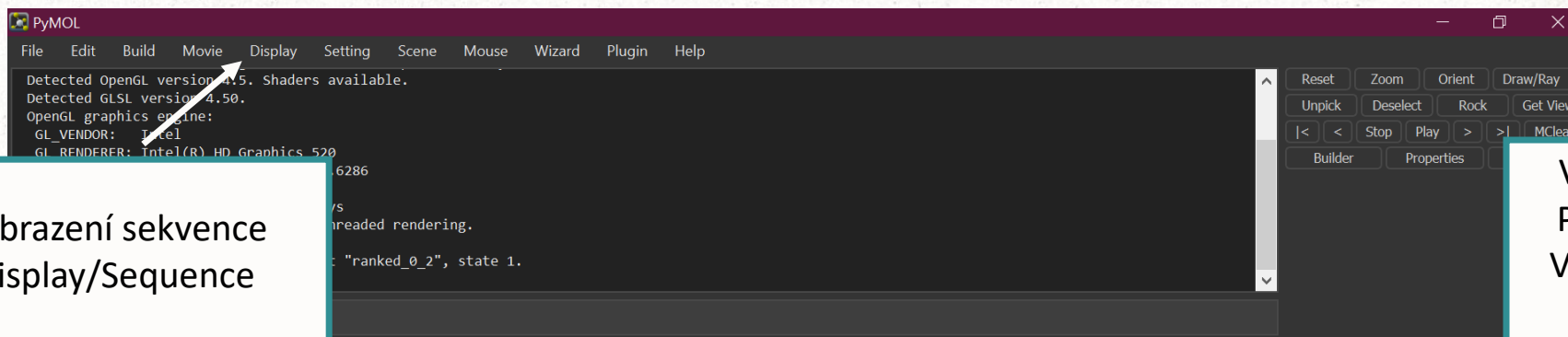
- Popis a funkce
- Konzervovanost (Alignment)
- Evoluce (fylogenetický strom)
- Zjištění přítomnosti domén daného proteinu
- Struktura proteinu – Alphafold/Colabfold
- Zjištění interakčních partnerů
- Úprava proteinového modelu v PyMol
  - Zvýraznění interakčních aminokyselin
  - Zvýraznění domén



# Zpracování 3D predikce (PyMOL)



ranked\_0(2)



Zobrazení sekvence  
Display/Sequence

Vyhlazení obrazu  
Publikační kvalita  
Vytvoření obrázku  
pro publikaci



A: Action  
S: Show  
H: Hide  
L: Label  
C: Color

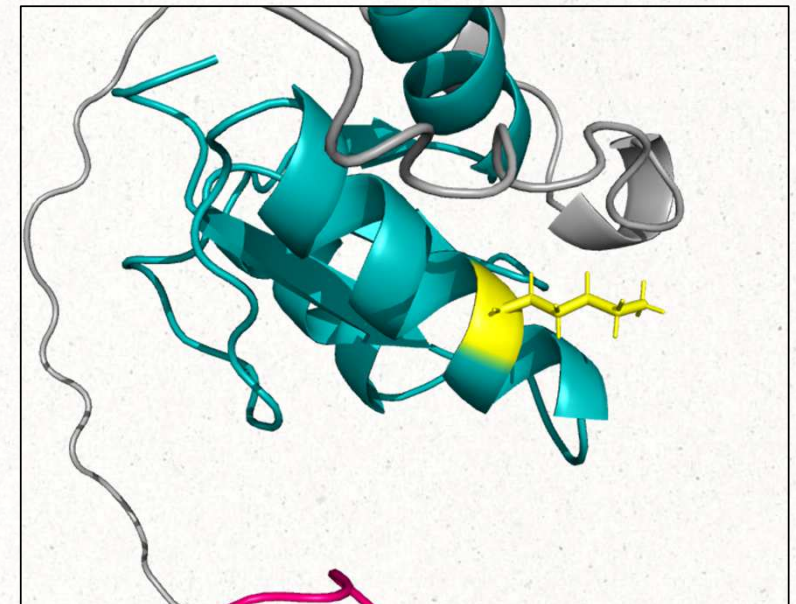
# Zvýraznění aminokyseliny

- Označení aminokyseliny
- Pravé tlačítko myši – Color
- Pravé tlačítko myši – Show → Side chain → Stick

```
16 21 26 31
sele SHSGASYVGMK
disable Color:
actions by element
color by chain
show by ss
hide by rep
preset spectrum
label auto
ss reds
zoom greens
orient blues
center yellows
origin magentas
drag cyans
clean oranges
```

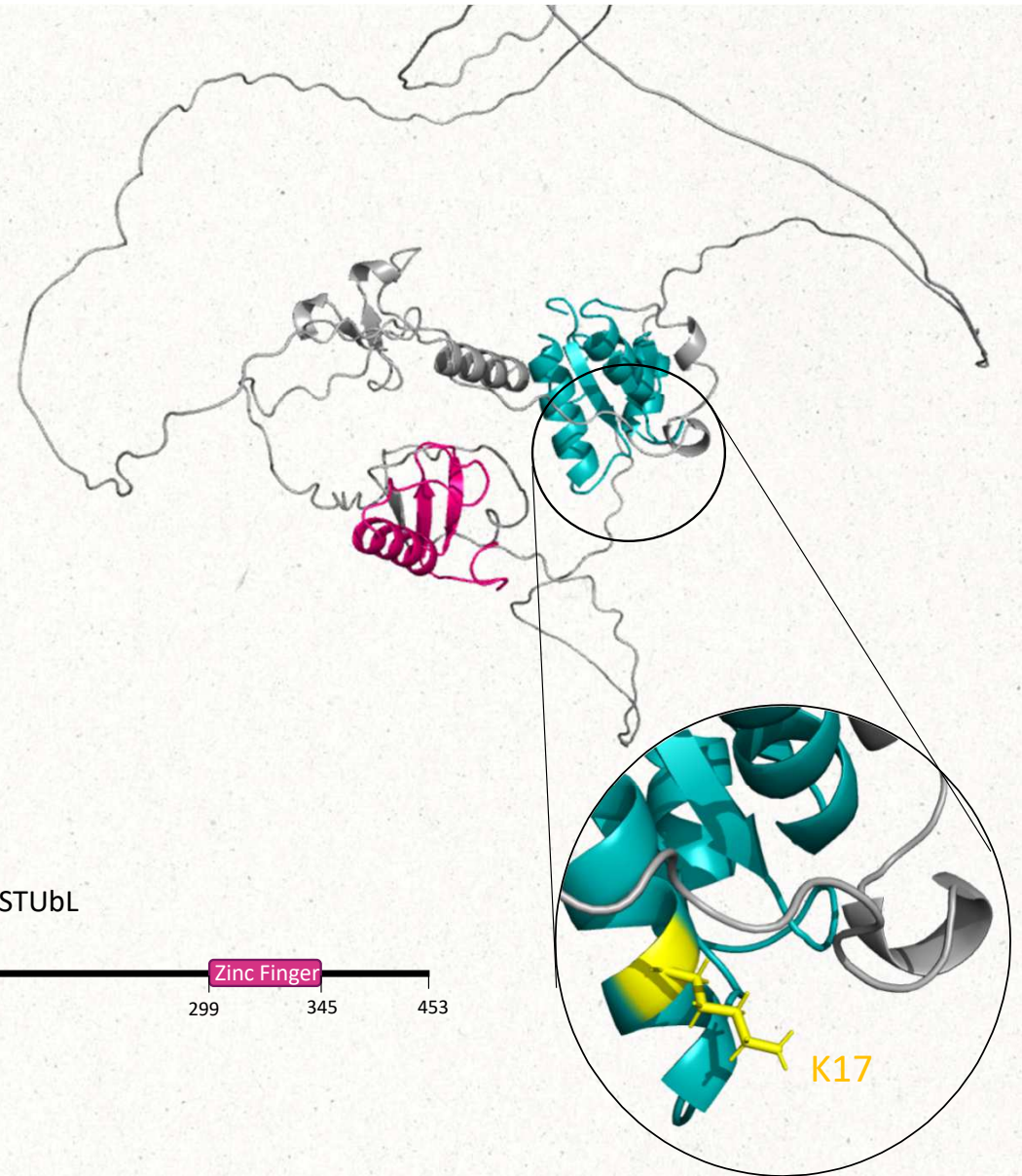
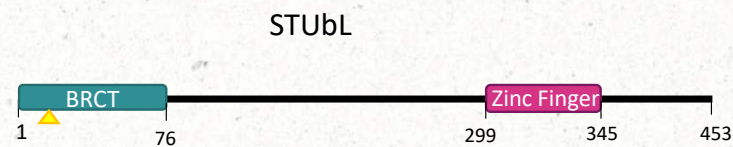
```
16 21 sticks 36 41
sele nb_spheres RSITHLVCI
disable ribbon
actions cartoon
color label
show cell
hide dots
preset spheres
label mesh
ss surface
zoom flag ignore
orient organic
center main chain
origin side chain
drag disulfides
clean valence
```

```
Show:
lines
sticks
spheres
```



# Úkony v PyMol

- Změna barvy proteinu
- Zvýraznění domén
- Zvýraznění aminokyselin
  - Změna barvy
  - Postranní řetězce
- Stahování obrázků
- Tvorba videa

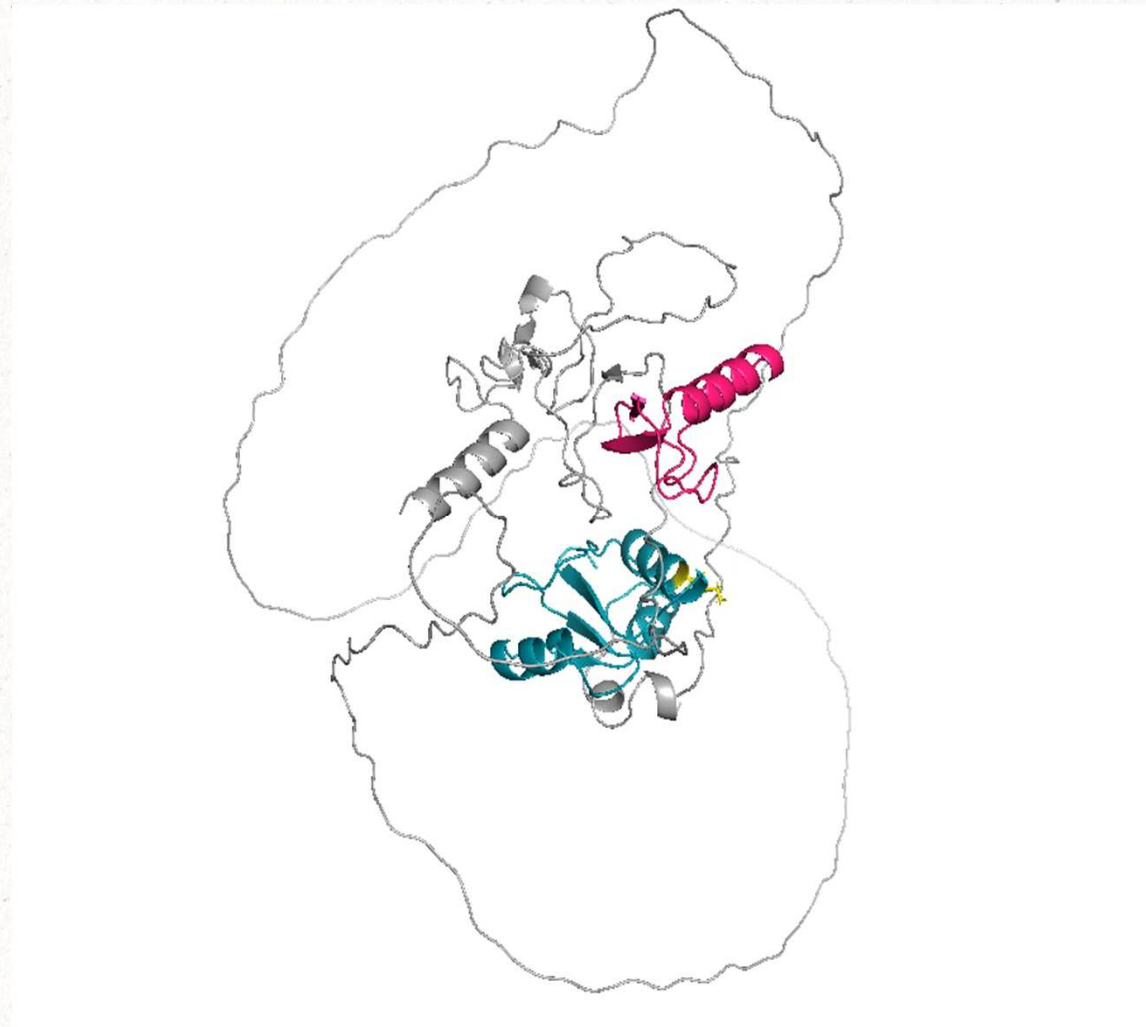


# Tvorba videa v PyMol

- Studentská verze PyMol nepodporuje stažení videí
- Využití aplikace BandiCam  
<https://www.bandicam.com/cz/>
  - Snímání obrazovky monitoru

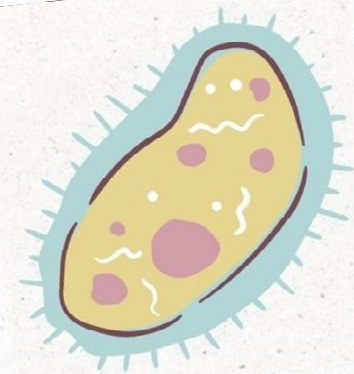
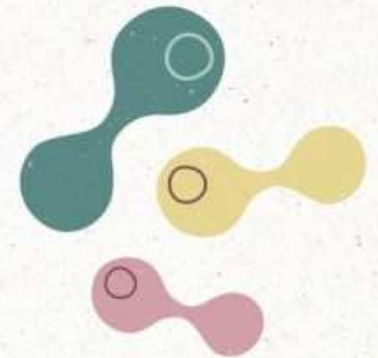
Příklad rotace v PyMol:

Movie → Program → Camera Loop  
→ Y roll → 16 second



# Přednáška: Analýza proteinu

- Popis a funkce
- Konzervovanost (Alignment)
- Evoluce (fylogenetický strom)
- Zjištění přítomnosti domén daného proteinu
- Struktura proteinu – Alphafold/Colabfold
- Zjištění interakčních partnerů
- Úprava proteinového modelu v PyMol
  - Zvýraznění interakčních aminokyselin
  - Zvýraznění domén





## Zkouška:

test 11.5. 2023 9:00 C2-211

+ přednáška/prezentace + nová data (výběr termínu)

Prezentace – Analýza proteinu

Konkrétní nová data – článek (< 5 let) o komplexu (nebo proteinu)

Ujasnit si souvislosti, rozšířit si znalosti, aplikovat poznatky z přednášek ...