

VYSOKÉ UČENÍ TECHNICKÉ FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

Bioinformatické databáze 1. cvičení

1 Srpkovitá anémie (Sickle Cell Disease)

Srpková anémie je genetická nemoc, která je nejčastější u Italů, Řeků, Afrických a Latinských Američanů. Projevuje se změnou tvaru buněk uvnitř červených krvinek. Buňky jsou pak méně flexibilní a mohou uvíznout v kapilárách, což způsobí blokaci toku krve do orgánů. Lidé se srpkovou anémií mohou mít symptomy, jako jsou například: zažloutlé oči a kůže, bledá pleť, bolesti kostí a kloubů, anémie nebo opožděný růst. Dožívají se věku 40 – 60 let. [1, 2]

1. Jaký přístupový kód má tento gen?

ACCESSION: NM_000518 (https://www.ncbi.nlm.nih.gov/nuccore/NM_000518.5)

2. Určete kódující sekvenci genu beta-globin a stáhněte ji ve FASTA formátu.

https://www.ncbi.nlm.nih.gov/nuccore/NM_000518.5?report=fasta

```
>NM_000518.5 Homo sapiens hemoglobin subunit beta (HBB), mRNA
ACATTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGA
GGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGC
AGGCTGTCTGGTGTCTACCCTTGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATG
CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCCTTTAGTGATGGCCTGGC
TCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT
CCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCA
CCCCACAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCA
CTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCTTTGTTCCCTAAGTCCAACCTACTAACT
GGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGCAA
```

3. Jaký je přístupový kód proteinové sekvence kódované genem beta-globin?

NP_000509 (<https://www.ncbi.nlm.nih.gov/protein/4504349>)

4. Stáhněte sekvenci proteinu kódovaného genem beta-globin ve FASTA formátu.

URL: <https://www.ncbi.nlm.nih.gov/protein/P68871.2?report=fasta>

```
>sp|P68871.2|HBB_HUMAN RecName: Full=Hemoglobin subunit beta; AltName:
    Full=Beta-globin; AltName: Full=Hemoglobin beta chain; Contains: RecName:
    Full=LVV-hemorphin-7; Contains: RecName: Full=Spinorphin
MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVVMGNPKVKAHGKKVLG
AFSDGLAHLNLTGTFATLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH
```

2 Hledání v databázích proteinových sekvencí

Vyhledejte záznam o proteinu beta-globin (Homo sapiens hemoglobin, beta, HBB) v proteinové databázi UniProtKB/Swiss-Prot.

1. Jaký přístupový kód má tento protein?

Kód: P68871 (<https://www.uniprot.org/uniprot/P68871>)

2. Z kolika aminokyselin je tento protein tvořen?

147

3. Jakou funkci má tento protein?

Zajišťuje přenos kyslíku z plic

4. Do jaké rodiny tento protein patří?

1.A.107.1.2 the pore-forming globin (globin) family

5. *Jaký efekt má mutace E7V?*

Produkce HbS (spojené s nemocí srpkové anémie) (<https://www.ncbi.nlm.nih.gov/clinvar/variation/15333/>)

6. *Zjistěte RS identifikátor pro mutaci E7V.*

rs334

7. *Stáhněte sekvenci tohoto proteinu ve FASTA formátu.*

<https://www.uniprot.org/uniprot/P68871.fasta>

```
>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAMGNPK
VKAHGKKVLGAFSDGLAHLDNLKGTFTLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTTPVQAAAYQKVAVGVALAHKYH
```

8. *Stáhněte z GenBank DNA sekvenci, která kóduje tento protein*

https://www.ncbi.nlm.nih.gov/nuccore/NC_000011.10?report=genbank&from=5225464&to=5227071&strand=true

ORIGIN

```
1 acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcatc
61 tgactcctga ggagaagtct gccgttactg ccctgtgggg caaggtgaac gtggatgaag
121 ttggtggtga ggccctgggc aggttggtat caaggttaca agacaggttt aaggagacca
181 atagaaactg ggcatgtgga gacagagaag actcttgggt ttctgatagg cactgactct
241 ctctgcctat tggctctattt tcccaccctt aggctgctgg tggctctacc ttggacccag
301 aggttctttg agtcctttgg ggatctgtcc actcctgatg ctgttatggg caaccctaag
361 gtgaaggctc atggcaagaa agtgctcggg gccttttagt atggcctggc tcacctggac
421 aacctcaagg gcacctttgc cacactgagt gagctgcact gtgacaagct gcacgtggat
481 cctgagaact tcagggtgag tctatgggac gcttgatgtt ttctttcccc ttcttttcta
541 tggttaaagt catgtcatag gaaggggata agtaacaggg tacagtttag aatgggaaac
601 agacgaatga ttgcatcagt ttggaagtct caggatcggt ttagtttctt ttatttgctg
661 ttcataacaa ttgttttctt ttgtttaatt cttgctttct tttttttct tctccgcaat
721 ttttactatt atacttaatg ccttaacatt gtgtataaca aaaggaaata tctctgagat
781 acattaagta acttaaaaaa aaactttaca cagtctgcct agtacattac tatttggaat
841 atatgtgtgc ttatttgcatt attcataatc tccctacttt attttctttt atttttaatt
901 gatacataat cattatacat atttatgggt taaagtgtaa tgttttaata tgtgtacaca
961 tattgaccaa atcagggtaa ttttgcattt gtaattttta aaaatgcttt cttcttttaa
1021 tatacttttt tgtttatctt atttctaata ctttccctaa tctctttctt tcagggcaat
1081 aatgatacaa tgtatcatgc ctctttgcac cattctaaag aataacagtg ataatttctg
1141 ggttaaggca atagcaatat ctctgcatat aaatatttct gcatataaat tgtaactgat
1201 gtaagagggt tcatattgct aatagcagct acaatccagc taccattctg cttttatttt
1261 atggttggga taaggctgga ttattctgag tccaagctag gcccttttgc taatcatgtt
1321 catacctctt atcttctctc cacagctcct gggcaacgtg ctggtctgtg tgctggccca
1381 tcacttttggc aaagaattca cccaccaggt gcaggctgcc tatcagaaag tgggtggctgg
1441 tgtggctaata gccctggccc acaagtatca ctaagctcgc tttcttgctg tccaatttct
1501 attaaagggt cttttgttcc ctaagtccaa ctactaaact gggggatatt atgaaggggc
1561 ttgagcatct ggattctgcc taataaaaaa catttatttt cattgcaa
```

3 Informace o SNP mutacích

1. *Ověřte patogenicitu v databázi Clinvar.*

[https://www.ncbi.nlm.nih.gov/clinvar?term=\(\(30372\[AlleleID\]\)OR\(30214\[AlleleID\]\)\)](https://www.ncbi.nlm.nih.gov/clinvar?term=((30372[AlleleID])OR(30214[AlleleID])))

2. *Povšimněte si rozdílných pozic v různých verzích genomových map.*

Pozice 5227002 v GRCh38.p12

Pozice 5248232 v GRCh37.p13

3. V jakém regionu lidského genomu se mutace nachází?
Kódovací region

4 Porovnání textových vyhledávacích systémů

Prohlédněte si seznam získaných výsledků. Nalezněte záznam spojený s mutací beta-globinu pro nemoc alpha-thalassemia v databázi OMIM.

1. Prohlédněte si seznam získaných výsledků. Nalezněte záznam spojený s mutací beta-globinu pro nemoc alpha-thalassemia v databázi OMIM.
<https://omim.org/entry/604131>

5 Informace o genomových projektech

V databázi Genomes OnLine Database (GOLD) zjistíte informace o dokončených a probíhajících genomových projektech.

1. Kolik bakteriálních a eukaryotických genomů bylo dosud osekvenováno a publikováno?
32 233 eukaryotních organismů
163 524 bakteriálních organismů
<https://gold.jgi.doe.gov/>
2. Srovnajte rozlišení a R-faktory výše nalezených struktur. Kterou ze struktur lze považovat za nejkvalitnější.

	R-faktor	Rozlišení
Okysličená verze	22.3 %	2.1 Å
Neokysličená verze	16 %	1.74 Å
Verze s genovou mutací	16.5 %	2.05 Å

Tabulka 1: Porovnání hodnot R-faktoru a rozlišení u různých verzí hemoglobinu

Za nejkvalitnější lze považovat neokysličenou verzi, jelikož má nejlepší rozlišení a nejnižší hodnotu R-faktoru.

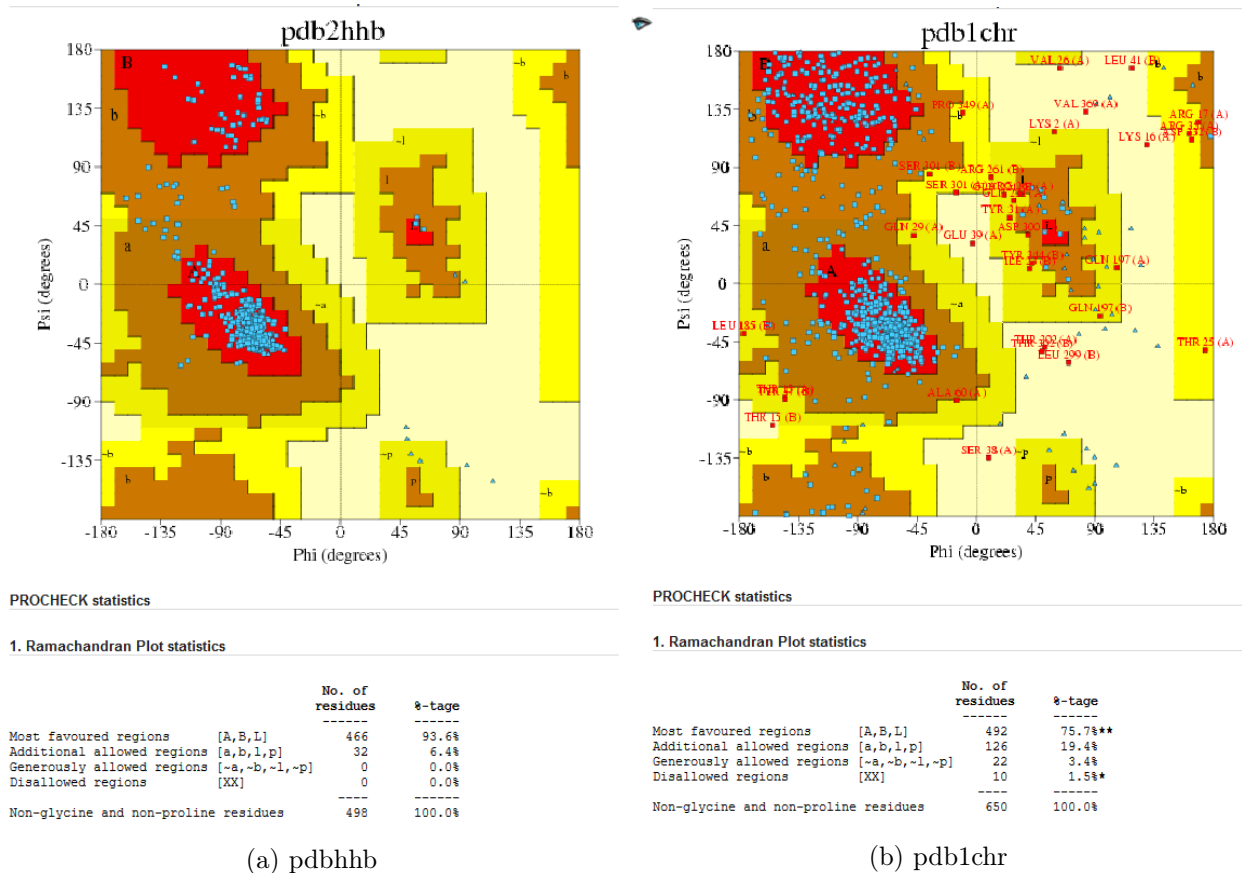
3. Stáhněte pdb soubor libovolné struktury. <https://files.rcsb.org/view/1BUW.pdb>
4. Stáhněte sekvenci ve formátu FASTA. Použijte staženou sekvenci pro vyhledání struktury Beta-Globinu (tip: advanced search).
Advanced options → Sequence → BLAST

6 Databáze PDBSum

V databázi PDBSum vyhledejte strukturu 2HHB.

<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl>

1. Prohlédněte si Ramachandrov diagram. Jedná se o dobře definovanou strukturu? Srovnajte s jinými strukturami (např. 1CHR).
Ano, jedná se o dobře definovanou strukturu, jelikož má model v nejvýhodnějších regionech více než 90 %. Oproti tomu struktura 1CHR má pouze 75,7 % a tedy se jedná o nepravděpodobnou strukturu (obrázek 1).



Obrázek 1: Ramachandran plots

2. V záložce *Proteins* si prohlédněte informace o sekundární struktuře. Jaké z nich můžeme vyvodit závěry?

Lze vidět, že se zde nachází šroubovice.

3. Jaké další informace lze dohledat v *PDBSum*?

Titulek, zdroj, R-faktor, rozlišení, autory, odkazy, ligandy, rozštěpy atd.

Reference

- [1] *Sickle Cell Disease - What Causes Sickle Cell Disease Video - About.com.m4v*. [Online, visited 19.3.2020]. URL <https://www.youtube.com/watch?v=9AHFHleYwdU&feature=related>
- [2] *Sickle Cell Disease*. [Online, updated 14.3.2020]. URL https://en.wikipedia.org/wiki/Sickle_cell_disease