

Bioinformatické databáze

Jakub Zárybnický (xzaryb00@stud.fit.vutbr.cz)

February 29, 2020

Contents

1	Hledání v databázích nukleotidových sekvencí	2
1.1	Jaký přístupový kód má tento gen?	3
1.2	Určete kódující sekvenci genu beta-globin a stáhněte ji ve FASTA formátu.	3
1.3	Jaký je přístupový kód proteinové sekvence kódované genem beta-globin?	3
1.4	Stáhněte sekvenci proteinu kódovaného genem beta-globin ve FASTA formátu.	4
1.5	Stáhněte článek věnující se genetickému onemocnění beta-talasemie (beta-thalassemia), jenž je důsledkem mutace v lidském hemoglobinu.	4
2	Hledání v databázích proteinových sekvencí	4
2.1	Jaký přístupový kód má tento protein?	4
2.2	Z kolika aminokyselin je tento protein tvořen?	4
2.3	Jakou funkci má tento protein?	4
2.4	Do jaké rodiny tento protein patří?	4
2.5	Jaký efekt má mutace E7V?	4
2.6	Stáhněte abstrakt článku popisujícího strukturu neokysličeného hemoglobinu s mutací způsobující srpkovitou anémii.	5
2.7	Zjistěte RS identifikátor pro mutaci E7V.	5
2.8	Stáhněte sekvenci tohoto proteinu ve FASTA formátu.	5
2.9	Stáhněte z GenBank DNA sekvenci, která kóduje tento protein.	6
3	Informace o SNP mutacích	7
3.1	Ověřte patogenicitu v databázi Clinvar.	7
3.2	Povšimněte si rozdílných pozic v různých verzích genomových map.	7

3.3	V jakém regionu lidského genomu se mutace nachází?	7
4	Porovnání textových vyhledávacích systémů	7
4.1	Prohlédněte si seznam získaných výsledků. Nalezněte záznam spojený s mutací beta-globinu pro nemoc alpha-thalassemia v databázi OMIM.	7
5	Informace o genomových projektech	7
5.1	Kolik bakteriálních a eukaryotických genomů bylo dosud os- ekvenováno a publikováno?	7
6	Vizualizace struktur molekul	8
6.1	V databázi PDB najděte článek popisující lidský beta-globin (tip: PDB-101, human hemoglobin). Podívejte se na animaci ukazující rozdíl mezi okysličenou a neokysličenou verzí. V článku nalezněte odkazy na PDB záznamy s okysličenou a neokysličenou verzí a také na verzi s genovou mutací způsobující srpkovitou anémií.	8
6.2	Srovnajte rozlišení a R-faktory výše nalezených struktur. Kterou ze struktur lze považovat za nejkvalitnější.	8
6.3	Stáhněte pdb soubor libovolné struktury.	8
6.4	Stáhněte sekvenci ve formátu FASTA. Použijte staženou sekvenci pro vyhledání struktury Beta-Globinu (tip: advanced search).	8
6.5	Vizualizujte jeden z těchto proteinů prostřednictvím nástroje / apletu JsMol.	8
6.6	Vyzkoušejte si různá zobrazení vybrané struktury.	8
6.7	Uložte si některá zobrazení ve formátu PNG.	8
7	Databáze PDBSum	8
7.1	Prohlédněte si Ramachandrov diagram. Jedná se o dobře definovanou strukturu? Srovnajte s jinými strukturami (např. 1CHR).	9
7.2	V záložce Proteins si prohlédněte informace o sekundární struktuře. Jaké z nich můžeme vyvodit závěry?	9
7.3	Jaké další informace lze dohledat v PDBSum?	9

1 Hledání v databázích nukleotidových sekvencí

Vyhledejte mRNA lidského genu beta-globinu (Homo sapiens hemoglobin, beta, HBB), který je součástí většího proteinu hemoglobinu. Použijte databázi

GenBank. Určete a stáhněte kódující oblasti tohoto genu a sekvenci proteinu, který je tímto genem kódován.

1.1 Jaký přístupový kód má tento gen?

GeneID 3043; ECYT6; CD113t-C

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

```
>NC_000011.10:5225464-5227071 Homo sapiens chromosome 11, GRCh38.p13 Primary Assembly
TTGCAATGAAAATAAATGTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAG
TTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTAGTG
ATACTTGTGGGCCAGGGCATTAGCCACACCAGCCACCCTTCTGATAGGCAGCCTGCACTGGTGGGGTG
AATTCTTTGCCAAAGTGATGGGCCAGCACACAGACCAGCACGTTGCCAGGAGCTGTGGGAGGAAGATAA
GAGGTATGAACATGATTAGCAAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAA
AATAAAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACATCAGTTACAAT
TTATATGCAGAAATATTTATATGCAGAGATATTGCTATTGCCTTAACCCAGAAATTATCACTGTTATTCT
TTAGAAATGGTGCAAAGAGGCATGATACATTGTATCATTATTGCCCTGAAAGAAAGAGATTAGGGAAAGTA
TTAGAAATAAGATAAAACAAAAAAGTATATTAAGAAGAAAGCATTTTTTTAAATTAACAAATGCAAAATT
ACCCTGATTTGGTCAATATGTGTACACATATTAACACATTACACTTTAACCATAAAATATGTATAATGAT
TATGTATCAATTAATAAATAAAGAAAAATAAAGTAGGGAGATTATGAATATGCAAATAAGCACACATATAT
TCCAAATAGTAATGTACTAGGCAGACTGTGTAAAGTTTTTTTTTAAGTTACTTAATGTATCTCAGAGATA
TTTCCTTTTGTTATACACAATGTTAAGGCATTAAGTATAATAGTAAAAATTGCGGAGAAGAAAAAAAAG
AAAGCAAGAATTAACAAAAAGAAAAACAATTGTTATGAACAGCAAATAAAGAAACTAAAACGATCCTGAG
ACTTCCCACTGATGCAATCATTCTGTCTGTTTCCATTCTAAACTGTACCCTGTTACTTATCCCCTTCCT
ATGACATGAACTTAACCATAGAAAAAGAGGGGAAAGAAACATCAAGCGTCCCATAGACTCACCTGAAG
TTCTCAGGATCCACGTGCAGCTTGTACAGTGCAGCTCACTCAGTGTGGCAAAGGTGCCCTTGAGGTTGT
CCAGGTGAGCCAGGCCATCACTAAAGGCACCGAGCACTTTCTTGCCATGAGCCTTACCTTAGGGTTGCC
CATAACAGCATCAGGAGTGGACAGATCCCCAAAGGACTCAAAGAACCTCTGGGTCCAAGGGTAGACCACC
AGCAGCCTAAGGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCAAGTGCCTATCAGAAACCAAGAGTCT
TCTCTGTCTCCACATGCCAGTTTCTATTGGTCTCCTTAAACCTGTCTTGTAACCTTGATACCAACCTGC
CCAGGGCTCACCACTTCCATCCAGTTTACCTTGCCCCACAGGGCAGTAACGGCAGACTTCTCCTC
AGGAGTCAGATGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTGAGAAGCAAATGT
```

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

cd08925, P68871

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

```
>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
VKAHGKKVLGAFSDGLAHLNLTGTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTTPVQAAAYQKVAVGVANALAHKYH
```

1.5 Stáhněte článek věnující se genetickému onemocnění beta-talasemie (beta-thalassemia), jenž je důsledkem mutace v lidském hemoglobinu.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2234194/pdf/zpq1620.pdf>

2 Hledání v databázích proteinových sekvencí

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

2.1 Jaký přístupový kód má tento protein?

P68871

2.2 Z kolika aminokyselin je tento protein tvořen?

146

2.3 Jakou funkci má tento protein?

"Involved in oxygen transport from the lung to the various peripheral tissues."

2.4 Do jaké rodiny tento protein patří?

" the pore-forming globin (globin) family"

2.5 Jaký efekt má mutace E7V?

VAR₀₀₂₈₆₃, "Hb S; at heterozygosity confers resistance to malaria", Sickle-cell anaemia

2.6 Stáhněte abstrakt článku popisujícího strukturu neokysličeného hemoglobinu s mutací způsobující srpkovitou anémii.

A variant Hb $\zeta 2\beta 2(s)$ that is formed from sickle hemoglobin (Hb S; $\alpha 2\beta 2(s)$) by exchanging adult α -globin with embryonic ζ -globin subunits shows promise as a therapeutic agent for sickle-cell disease (SCD). Hb $\zeta 2\beta 2(s)$ inhibits the polymerization of deoxygenated Hb S in vitro and reverses characteristic features of SCD in vivo in mouse models of the disorder. When compared with either Hb S or with normal human adult Hb A ($\alpha 2\beta 2$), Hb $\zeta 2\beta 2(s)$ exhibits atypical properties that include a high oxygen affinity, reduced cooperativity, a weak Bohr effect and blunted 2,3-diphosphoglycerate allostery. Here, the 1.95 Å resolution crystal structure of human Hb $\zeta 2\beta 2(s)$ that was expressed in complex transgenic knockout mice and purified from their erythrocytes is presented. When fully liganded with carbon monoxide, Hb $\zeta 2\beta 2(s)$ displays a central water cavity, a $\zeta 1-\beta(s)2$ (or $\zeta 2-\beta(s)1$) interface, intersubunit salt-bridge/hydrogen-bond interactions, C-terminal β His146 salt-bridge interactions, and a β -cleft, that are highly unusual for a relaxed hemoglobin structure and are more typical of a tense conformation. These quaternary tense-like features contrast with the tertiary relaxed-like conformations of the $\zeta 1\beta(s)1$ dimer and the CD and FG corners, as well as the overall structures of the heme cavities. This crystallographic study provides insights into the altered oxygen-transport properties of Hb $\zeta 2\beta 2(s)$ and, moreover, decouples tertiary- and quaternary-structural events that are critical to Hb ligand binding and allosteric function.

2.7 Zjistěte RS identifikátor pro mutaci E7V.

rs334

2.8 Stáhněte sekvenci tohoto proteinu ve FASTA formátu.

```
>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2
MVHLTPVEKSAVTALWGKVNVDVGGGALGRLLVYPWTQRFFESFGDLSTPDAVMGNPK
VKAHGKKVLGAFSDGLAHLNLTGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTTPPVQAAYQKVVAGVANALAHKYH
```

2.9 Stáhněte z GenBank DNA sekvenci, která kóduje tento protein.

```
>NC_000011.10:5226570-5228834 Homo sapiens chromosome 11, GRCh38.p13 Primary Assembly
GACTCACCTGAAGTTCTCAGGATCCAGTGCAGCTTGTACAGTGCAGCTCACTCAGTGTGGCAAAGGT
GCCCTTGAGGTTGTCCAGGTGAGCCAGGCCATCACTAAAGGCACCGAGCACTTTCTTGCCATGAGCCTTC
ACCTTAGGGTTGCCATAACAGCATCAGGAGTGGACAGATCCCCAAAGGACTCAAAGAACCTCTGGGTCC
AAGGGTAGACCACCAGCAGCCTAAGGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCACTGCCTATCAG
AAACCAAGAGTCTTCTGTCTCCACATGCCAGTTTCTATTGGTCTCCTTAAACCTGTCTTGTAACCT
TGATACCAACCTGCCAGGGCCTCACCACTTCCACAGTTTACCTTGCCCCACAGGGCAGTAACG
GCAGACTTCTCCTCAGGAGTCAGATGCACCATGGTGTCTGTTGAGGTTGCTAGTGAACACAGTTGTGTC
AGAAGCAAATGTAAGCAATAGATGGCTCTGCCCTGACTTTTATGCCAGCCCTGGCTCCTGCCCTCCCTG
CTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGCTCCACAGGGTGAGGTCTAAGTGATGACAGCCGTA
CCTGTCTTGGCTCTTCTGGCACTGGCTTAGGAGTTGGACTTCAAACCCTCAGCCCTCCCTCTAAGATAT
ATCTCTTGGCCCCATACCATCAGTACAAATTGCTACTAAAAACATCCTCCTTGAAGTGTATTTACGTA
ATATTTGGAATCACAGCTTGGTAAGCATATTGAAGATCGTTTCCCAATTTTCTTATTACACAAATAAGA
AGTTGATGCACTAAAAGTGGAAGAGTTTGTCTACCATAATTCAGCTTTGGGATATGTAGATGGATCTCT
TCCTGCGTCTCCAGAATATGCAAAATACTTACAGGACAGAATGGATGAAAACCTTACCTCGGTTCTAAGC
ATATCTTCTCCTTATTTGGATTAAACCTTCTGGTAAGAAAAGAAAAAATATATATATATATGTGTGTAT
ATATACACACATACATATACATATATATGCATTCAATTTGTGTTGTTTTTCTTAATTTGCTCATGCATGC
TAATAAATTATGTCTAAAAATAGAATAAATACAAATCAATGTGCTCTGTGCATTAGTTACTTATTAGGTT
TTGGGAAACAAGAGATAAAAACTAGAGACCTCTTAATGCAGTCAAAAATACAAATAAATAAAAAAGTCAC
TTACAACCCAAAGTGTGACTATCAATGGGGTAATCAGTGGTGTCAAATAGGAGGTTAACTGGGGACATCT
AACTGTTTCTGCCTGGACTAATCTGCAAGAGTGTCTGGGGGAACAAAAGCCTCTGTGACTTAGAAAGTA
GGGGTAGGAGGGGAAAAGGCTTCTACTTGGCTCAGATTATTTTTTCTCTAGTCCACTAAGAATACTG
CGTTTTAAATCATTTCCTTGATTCAAGTTCCTATTTCTCTTTATATTTTGTGTTTAAACCTCCTTTA
CTAAAATTTACTCTTCTTCTCTATAGCTTCCCAACGTGATCGCCTTTCTCCCATCCCCCTGTACTTTTT
CCCCTTGTACTAAATTAACCTCCTCAGGTGAGGAAAAACCTTTGAAGTGCAGAGTTCTGCTTCCTGCTATT
AAAAGATGTAATTAACAGCAAAGGTAGCAAGCATTTATGAGGTGAGCGTAGGGTCTCAGTGTTCCCTA
AGGGCCCTGTGAGTCATCCTGAATCCTGCCCCTACCTGGAACCCATGTCGGTTTAGTAAGGAAAGTGTT
ATACTTTTACTTTGCATGTTTCTCCTACTTCTTCTTTTCTCAGCTCTAACACTCTGAACTACGATTACACA
AAATAAAATAAAATAAAATAAAATAAAACAATAAAATGAAATAAAATTTAGGTTAACCAAAAGAACTGG
ATCCTCTATTTCTAGTTATCAGAAGGAAATTTACAAATTTCTTATTTCCATTGCTTTATTCTCTTAAATG
CTTCTCTATTATTGCTAAATAAATAGAGATCTCTCACTTTTCTACCTGTCTCAACCTCATCAGGTAC
TTGTGAAAAAATCTCACTCTGATTATTCTCACACACGCAGAAAGTGTGTTGTTCTTCTATGGCTATCTGG
AGCCTAGGTATAAAATATGCCTATGTATGATTATAGAGGTAAGAGGGATAAAATTTAAGTATTTTCTT
TTTATATTCAATCCTCTGTAAAAAA
```

3 Informace o SNP mutacích

V databázi dbSNP vyhledejte informace o mutaci způsobující srpkovitou anémii (využijte RS identifikátor z předchozího úkolu).

3.1 Ověřte patogenicitu v databázi Clinvar.

Pathogenic, Hb SS disease: <https://www.ncbi.nlm.nih.gov/clinvar/variation/15333/>

3.2 Povšimněte si rozdílných pozic v různých verzích genomových map.

3.3 V jakém regionu lidského genomu se mutace nachází?

11p15.4

4 Porovnání textových vyhledávacích systémů

Vyhledejte záznam o proteinu beta-globin (Homo sapiens hemoglobin, beta, zkratka HBB) s použitím vyhledávacího systému GQuery.

4.1 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) zjistěte informace o dokončených a probíhajících genomových projektech.

5.1 Kolik bakteriálních a eukaryotických genomů bylo dosud osekvenováno a publikováno?

Bakterie: 14,863, eukarya: 19,163

6 Vizualizace struktur molekul

V největší strukturní databázi PDB hledejte informace o proteinu beta-globin.

- 6.1** V databázi PDB najděte článek popisující lidský beta-globin (tip: PDB-101, human hemoglobin). Podívejte se na animaci ukazující rozdíl mezi okysličenou a neokysličenou verzí. V článku naleznete odkazy na PDB záznamy s okysličenou a neokysličenou verzí a také na verzi s genovou mutací způsobující srpkovitou anémii.

<https://www.rcsb.org/structure/2hhb>
<https://www.rcsb.org/structure/1hho>
<https://www.rcsb.org/structure/2hbs>

- 6.2** Srovnajte rozlišení a R-faktory výše nalezených struktur. Kterou ze struktur lze považovat za nejkvalitnější.

2HHB - R-value 0.160, resolution 1.74 Å

- 6.3** Stáhněte pdb soubor libovolné struktury.

- 6.4** Stáhněte sekvenci ve formátu FASTA. Použijte staženou sekvenci pro vyhledání struktury Beta-Globinu (tip: advanced search).

e.g. <http://www.rcsb.org/structure/3W4U>

- 6.5** Vizualizujte jeden z těchto proteinů prostřednictvím nástroje / apletu JsMol.

<https://chemapps.stolaf.edu/jmol/jmol.php?pdbid=3W4U>

- 6.6** Vyzkoušejte si různá zobrazení vybrané struktury.

- 6.7** Uložte si některá zobrazení ve formátu PNG.

7 Databáze PDBSum

V databázi PDBSum vyhledejte strukturu 2HHB.

7.1 Prohlédněte si Ramachandrov diagram. Jedná se o dobře definovanou strukturu? Srovnajte s jinými strukturami (např. 1CHR).

2HHB	
Main-chain bond angles	-0.58*
OVERALL AVERAGE	-0.07
vs	
1CHR	
Main-chain bond angles	-3.27**
OVERALL AVERAGE	-1.23**

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

1 1.10.490.10 = Mainly Alpha Orthogonal Bundle

7.3 Jaké další informace lze dohledat v PDBSum?

Prvky a vlastnosti sekundární a terciární struktury proteinu