MSE 212 – Exercise session 1.1 (2025)

1. Write your first name in amino acids (if the letter does not exist, leave it out)
    a. What 3-letter code corresponds with each of the letters?
    b. Give the structure of these amino acids corresponding to each of the letters.
    c. Are there any amino acids, in your name, that could enforce secondary and tertiary structure formation? Could you point them out and in case there are none, could you name two important examples?

2. You have the following DNA sequence on the template strand, that serves as template for transcription: TTCATTATCTAACAACTCCCA
    a. Please give the sequence of the complementary coding strand of this strand
    b. Please give the mRNA-sequence of this strand.
    c. Please give the corresponding amino acid sequence in single letter code

3. You would like to amplify the following DNA-strand with PCR: GGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAA
    a. Explain how the PCR technique works.
    b. Design 2 primers for the amplification of this sequence and give their sequence
    c. How many cycles do we need to obtain 1000 copies?
    d. Please name a potential application for PCR reactions.

4. DNA can be used as building material to create nanostructures for many biological applications:
    a. For what application could one use DNA-nanotechnology?
    b. Try to make a scaffold map of the Easter bunny
    c. You have made the green fluorescent protein (GFP) and want to connect it to your Easter bunny DNA scaffold, briefly explain how would you do?

Chart Key

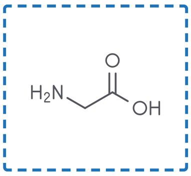
🟦 Alkyl ⬛ Aromatic 🟩 Neutral 🟨 Acidic 🟧 Basic ☐ Essential ⬚ Non-Essential

Note: The NH_2 and COOH values listed below are pK_a values.

Alanine

Ala A

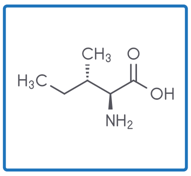
NH_2: 9.87 **COOH:** 2.35



Glycine

Gly G

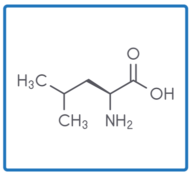
NH_2: 9.60 **COOH:** 2.34



Isoleucine

Ile I

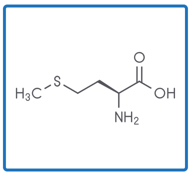
NH_2: 9.76 **COOH:** 2.32



Leucine

Leu L

NH_2: 9.60 **COOH:** 2.36



Methionine

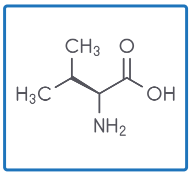
Met M

NH_2: 9.21 **COOH:** 2.28

Proline

Pro P

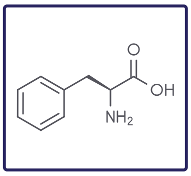
NH_2: 10.60 **COOH:** 1.99



Valine

Val V

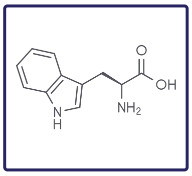
NH_2: 9.72 **COOH:** 2.29



Phenylalanine

Phe F

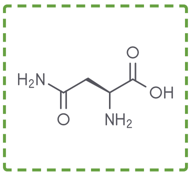
NH_2: 9.24 **COOH:** 2.58



Tryptophan

Trp W

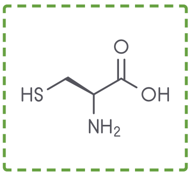
NH_2: 9.39 **COOH:** 2.38



Asparagine

Asn N

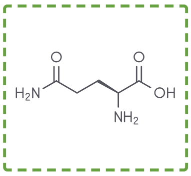
NH_2: 8.80 **COOH:** 2.02



Cysteine

Cys C

NH_2: 10.78 **COOH:** 1.71



Glutamine

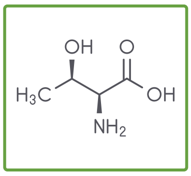
Gln Q

NH_2: 9.13 **COOH:** 2.17

Serine

Ser S

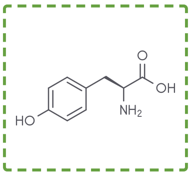
NH_2: 9.15 **COOH:** 2.21



Threonine

Thr T

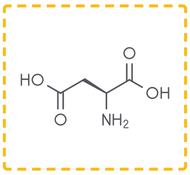
NH_2: 9.12 **COOH:** 2.15



Tyrosine

Tyr Y

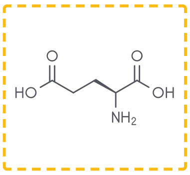
NH_2: 9.11 **COOH:** 2.20



Aspartic Acid

Asp D

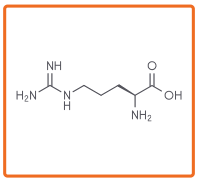
NH_2: 9.60 **COOH:** 1.88



Glutamic Acid

Glu E

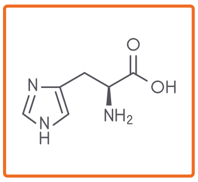
NH_2: 9.67 **COOH:** 2.19



Arginine

Arg R

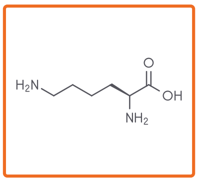
NH_2: 9.09 **COOH:** 2.18



Histidine

His H

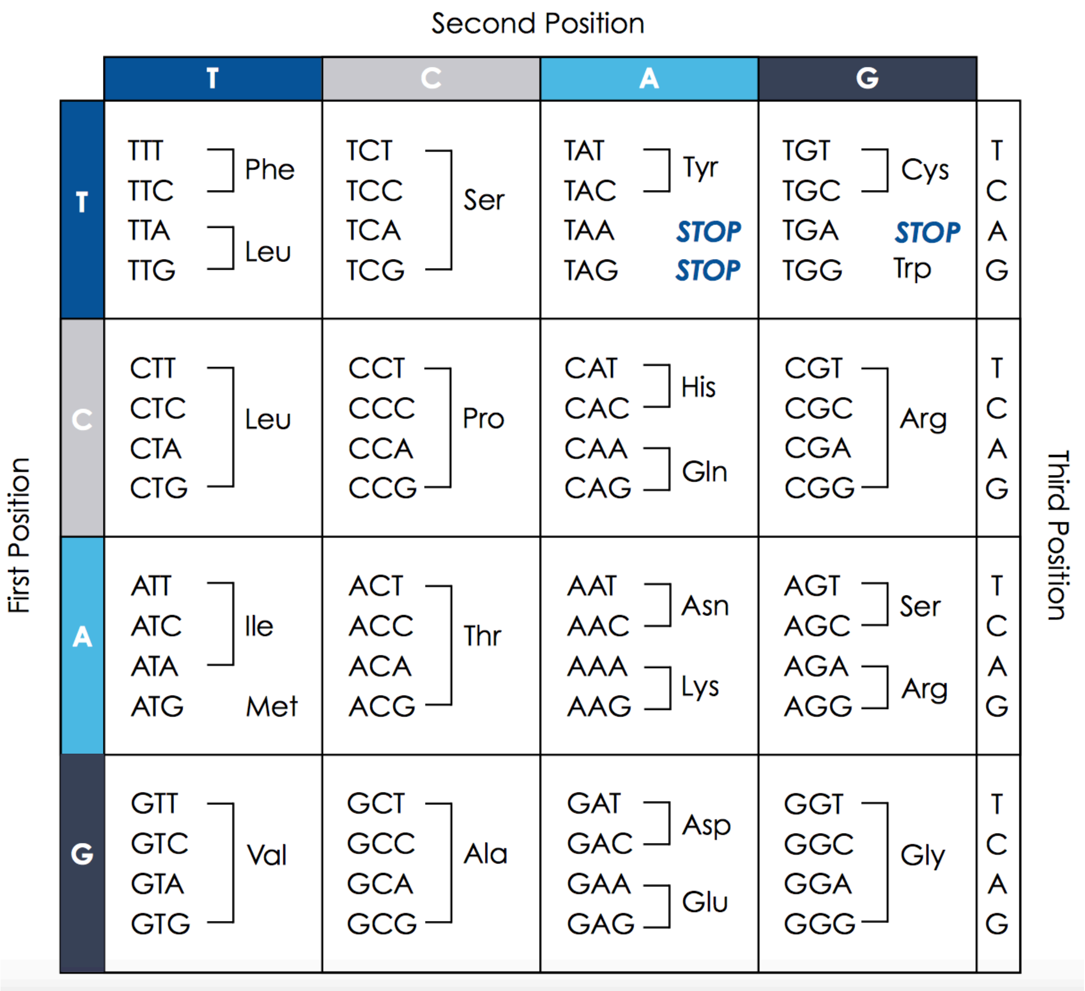
NH_2: 8.97 **COOH:** 1.78



Lysine

Lys K

NH_2: 10.28 **COOH:** 8.90



Second Position

First Position	T		C		A		G		Third Position
T	TTT TTC TTA TTG	Phe Leu	TCT TCC TCA TCG	Ser	TAT TAC TAA TAG	Tyr ***STOP*** ***STOP***	TGT TGC TGA TGG	Cys ***STOP*** Trp	T C A G
C	CTT CTC CTA CTG	Leu	CCT CCC CCA CCG	Pro	CAT CAC CAA CAG	His Gln	CGT CGC CGA CGG	Arg	T C A G
A	ATT ATC ATA ATG	Ile Met	ACT ACC ACA ACG	Thr	AAT AAC AAA AAG	Asn Lys	AGT AGC AGA AGG	Ser Arg	T C A G
G	GTT GTC GTA GTG	Val	GCT GCC GCA GCG	Ala	GAT GAC GAA GAG	Asp Glu	GGT GGC GGA GGG	Gly	T C A G

Exercise 1.2: Protein structure modeling with AlphaFold3.

Introduction

Sickle Cell Anemia (SCA) is a genetic blood disorder that primarily affects red blood cells, causing them to adopt an abnormal, sickle-like shape. This disease is caused by a mutation in the HBB gene, which encodes the β-globin subunit of hemoglobin, the protein responsible for carrying oxygen in red blood cells. Hemoglobin is a protein composed of two α-globin subunits together with two β-globin subunits and is vital for oxygen transport within the human body.

In people with sickle cell anemia, the mutation leads to the production of hemoglobin S (HbS) instead of the normal hemoglobin A (HbA). The change involves a single amino acid substitution in the β-globin chain, glutamic acid is replaced by valine at position 7. This seemingly small change has profound effects on the properties of hemoglobin and the overall function of red blood cells, as it encourages the clustering of multiple Hemoglobin proteins, and resultingly contributes to the deformation of red blood cells. Within this tutorial, we will use AlphaFold 3 (Google Deepmind) to predict the structure of hemoglobin manufactured from wild-type (healthy) and mutant-type genes to identify the cause of SCA.

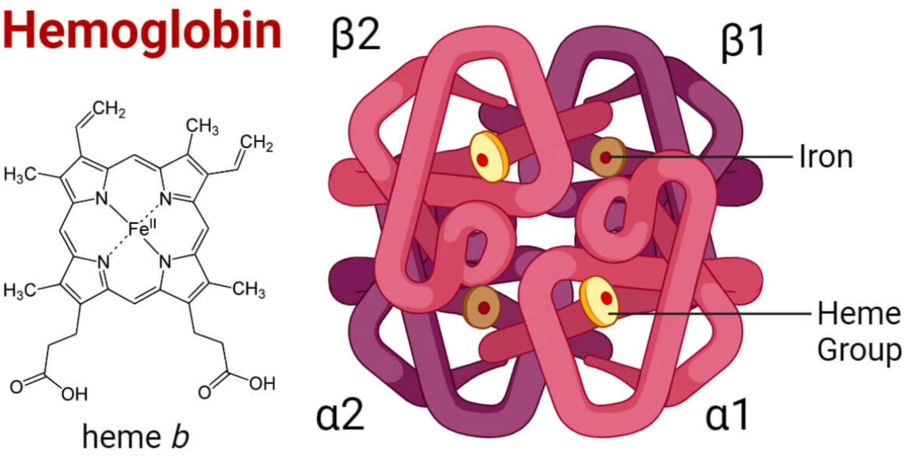
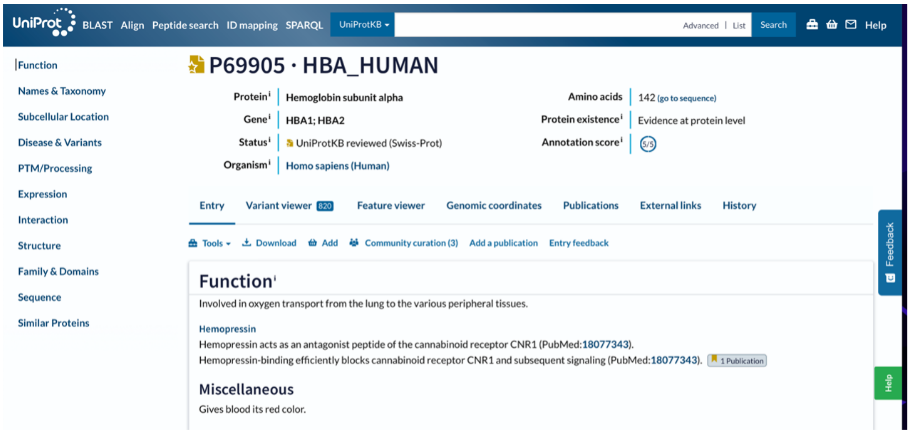


Figure 1: Structure of the hemoglobin composed of two α-globin subunits and two β-globin subunits. Figure from (https://microbenotes.com/hemoglobin).

1 Sequence and structural information about hemoglobin

To predict the structure of a protein with AlphaFold, we first need to know its amino acid sequence, and this information can be usually found in UniProt. UniProt (Universal Protein Resource) is a comprehensive and freely accessible database of protein sequences and functional information. It serves as a central hub for protein data, combining information from various sources including genome sequencing projects and scientific literature.

The protein we want to predict, hemoglobin, has α-globin subunits and β-globin subunits. The α-globin subunit sequence can be found under (https://www.uniprot.org/uniprotkb/P69905). The sequence of the β-globin subunit can be found under (https://www.uniprot.org/uniprotkb/P68871).



P69905 · HBA_HUMAN

Protein: Hemoglobin subunit alpha
Gene: HBA1; HBA2
Status: UniProtKB reviewed (Swiss-Prot)
Organism: Homo sapiens (Human)
Amino acids: 142 (go to sequence)
Protein existence: Evidence at protein level
Annotation score: 5/5

Function

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

Hemopressin

Hemopressin acts as an antagonist peptide of the cannabinoid receptor CNR1 (PubMed:18077343).

Hemopressin-binding efficiently blocks cannabinoid receptor CNR1 and subsequent signaling (PubMed:18077343). 1 Publication

Miscellaneous

Gives blood its red color.

Figure 2: The UniProt website of the α-globin subunit

Under the website of α-globin and β-globin subunit:

1. What organism do they belong to?
2. What are the lengths and amino acid sequences of them?

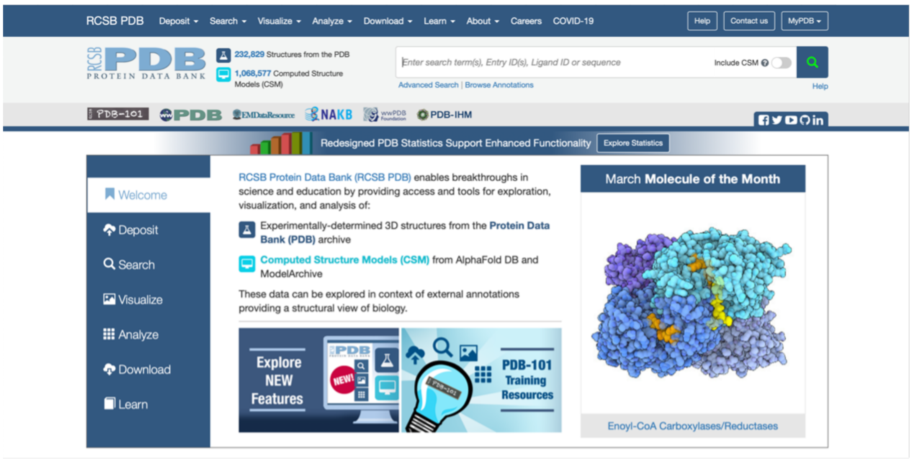
To further obtain the structural information of a protein, we can go to PDB. PDB (Protein Data Bank, https://www.rcsb.org) is a comprehensive database for three-dimensional structural data of large biological molecules, primarily proteins and nucleic acids. The structural data is obtained through experimental methods such as X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy.

3. Under the website of α-globin, in the “structure” session, can you find the PDB identifier of it?

4. Under the PDB website, with this PDB identifier, what protein do you find?

5. In the structure page of this protein, by rotating the protein with your mouse, can you find any α-helix and β-sheet in this protein? How many subunits are there in this protein?

6. How many α-globin and β-globin subunits are there in the hemoglobin? (Hint: you can also find the answer in Figure 1)



232,829 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

Redesigned PDB Statistics Support Enhanced Functionality

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

- Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive
- Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

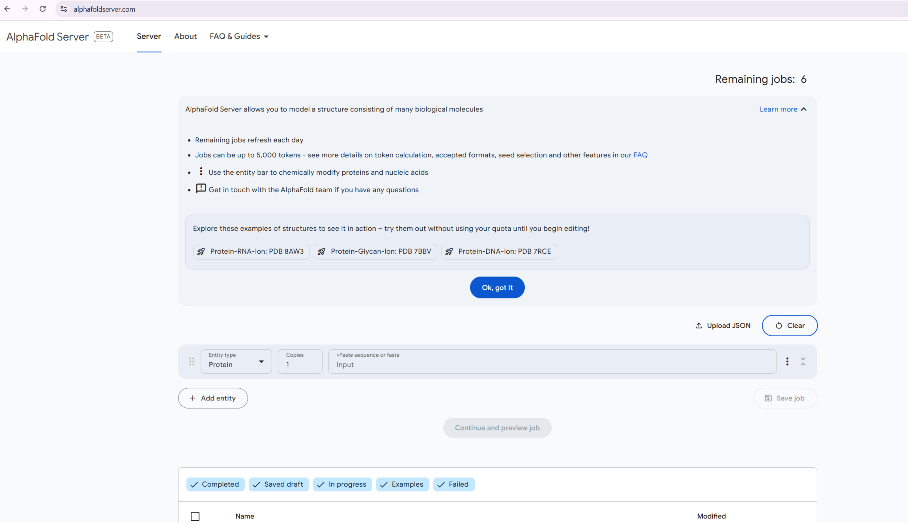
Figure 3: The PDB website

2 Introduction to AlphaFold 3

Read the abstract and the main section of (https://doi.org/10.1038/s41586-024-07487-w). Following that, describe in 1-2 sentences what AlphaFold 3 is and its intended purpose.

3 Predicting the structure of physiological (healthy) hemoglobin

Open AlphaFold 3 (https://alphafoldserver.com/) in any browser and click "Continue with Google". This does not work with the EPFL account, you will need a personal google account. Your screen should look something like the image below.



Remaining jobs: 6

AlphaFold Server allows you to model a structure consisting of many biological molecules

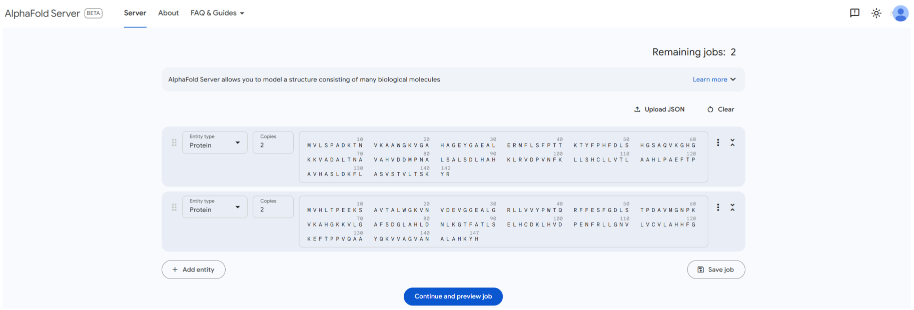
- Remaining jobs refresh each day
- Jobs can be up to 5,000 tokens - see more details on token calculation, accepted formats, seed selection and other features in our FAQ
- Use the entity bar to chemically modify proteins and nucleic acids
- Get in touch with the AlphaFold team if you have any questions

Explore these examples of structures to see it in action – try them out without using your quota until you begin editing!

Entity type: Protein
Copies: 1
>Paste sequence or fasta: Input

Firstly, copy the α sequence and paste it into the input field of AlphaFold, and put the right copy number. Subsequently, press ”Add entity”, choose Entity type ”Protein” and paste the β amino acid sequence. Again, put the right copy number.

Your screen should now look like below.



Remaining jobs: 2

AlphaFold Server allows you to model a structure consisting of many biological molecules

Entity type: Protein
Copies: 2

```
         10         20         30         40         50         60
MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS HGSAQVKGHG
         70         80         90        100        110        120
KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK LLSHCLLVTL AAHLPAEFTP
        130        140 142
AVHASLDKFL ASVSTVLTSK YR
```

Entity type: Protein
Copies: 2

```
         10         20         30         40         50         60
MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK
         70         80         90        100        110        120
VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG
        130        140      147
KEFTPPVQAA YQKVVAGVAN ALAHKYH
```

Click "Continue and preview Job". Name the Job "Hemoglobin" and, turn on manual seeding and select seed "1". Press "Confirm and submit Job". The job is now running. The calculation might take a couple of minutes. Please note, that we exclude the iron cations for simplicity. Now visualize the results by pressing the three dots next to the job and click "Open results'. What is the predominant secondary structure of the Hemoglobin protein?

4 Predicting the structure of Hemoglobin prone to SCA

In your list of jobs, find "Hemoglobin" press on the three dots and select "Clone and reuse", the two subunits are now loaded. Now change the glutamic acid (E) in the 7th position of the β-globin subunit to valine (V). You can then press "Continue and preview Job", name the job "Hemoglobin-SCA", select seed 1 and "Confirm and submit job". Once both jobs are finished, press the three dots and select "Open results" can you identify any major difference between the two results?

5 Understanding Sickle Cell Anemia

Click on the three dots next to the "Hemoglobin" job and click "Clone and reuse". Increase the number of copies for both the α and β-globin subunits to 8 each. This allows us to effectively visualize the interaction of 4 healthy Hemoglobin proteins (each composed of 2 α and 2 β subunits). Click continue and preview job, name the job "4Hemoglobin" select seed 1 and submit. Now press the three dots next to "Hemoglobin-SCA" and do the same (increase number of copies to 8 each, name the job "4Hemoglobin-SCA", select seed 1 and submit). Now visualize the results of both jobs by pressing the three dots and click "Open results". Can you now identify a significant difference between the two structures?