Due at the end of the class

Exercise 1: Nar issue databases

The Database Issue of Nucleic Acids Research is the in a series dedicated to factual biological databases. Such databases are an essential resource for working biologists and this compilation provides descriptions of the most important of these databases and serves to introduce newly compiled databases that provide specialist information in the biological area. NAR Online contains hotlinks to all of the databases in the compilation as well as brief summaries of their content.

Go to the NAR database (http://www.oxfordjournals.org/our_journals/nar/database/a/) Find information on e.g., Genbank (Nucleotide Sequence Databases), Genecards (General Human Genetic Databases), KEGG (Metabolic Pathways) and other databases that interest you.

1. Try to find out how many nucleotides and sequences are currently in GenBank

2. How can you download data from GenBank to your own computer?

Exercise 2: The GeneCards database.

In case you have the name of your gene, the GeneCards database is a very useful resource for retrieving information about it.

Go to the GeneCards database (http://www.genecards.org/) and retrieve the information for the *HBA2* gene. GeneCards provides links to many biological and biomedical databases. Follow some of the links to get an impression what kind of information you can retrieve about the HBA2 gene.

1. View three dimensional protein structure of a *HBA2* molecular with Protopedia and take a screenshot.

2. What is the molecular function of *HBA2* according to the Gene Ontology?

Exercise 4: Searching GenBank with Entrez.

1. Go the GenBank database (http://www.ncbi.nlm.nih.gov/nucleotide/) and retrieve the information about the mRNA sequence with accession code NM_000517. Which gene is this?

2. What kind of information is present in the retrieved database record? What is the sequence format?

3. Select the 'Graphic' to get an overview of the location of the features in this

mRNA sequence.

4. GenBank is also linked to many databases. To retrieve additional information you can use the 'Links' and then go to the Gene database. This provides you further information and links to other databases. Follow the links to PubMed, OMIM and the Protein database to retrieve additional information. Show additional information that you obtained from these three external databases.

Exercise 4: Searching GenBank with Entrez

1. Go to uniprot blast server (http://www.uniprot.org/blast/)

2. Copy the following **human** amino acid sequence

MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLF
EENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSR
DKKKDTVPWFPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRKQFADIA
YNYRHGQPIPRVEYMEEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCGF
HEDNIPQLEDVSQFLQTCTGFRLRPVAGLLSSRDFLGGLAFRVFHCTQYIRHGS
KPMYTPEPDICHELLGHVPLFSDRSFAQFSQEIGLASLGAPDEYIEKLATIYWFT
VEFGLCKQGDSIKAYGAGLLSSFGELQYCLSEKPKLLPLELEKTAIQNYTVTEFQ
PLYYVAESFNDAKEKVRNFAATIPRPFSVRYDPYTQRIEVLDNTQQLKILADSIN
SEIGILCSALQKIK

3. Run the search and identify the protein. Use the link provided to see the Swiss-Prot report.

• What is the primary accession number?

• What is the name of the protein?

• What is the gene called?

• What is the PAHdb?

• How many amino acid residues are there in the protein?

• What is the molecular mass of the protein?

• View and take a screen shot of a three-dimensional (3D) image of the protein that the gene codes for (Hint: PDB).

Exercise 5: using KEGG to identify known metabolic disorders. Hemolytic anemia can be caused by glucosephosphate isomerase (GPI) deficiency. The

KEGG database provides the tools to identify the pathway in which this enzyme is active.

1. Go to the KEGG database (http://www.genome.ad.jp/kegg/) and then go to the KEGG GENES database. Locate the gene *GPI* (in human, *hsa:GPI*). To which enzyme does this gene correspond?

2. Next click on the enzyme to proceed to the enzyme database. In which pathways is this enzyme active?

3. Proceed to the Glycolysis pathway map, which now indicates the enzymes (red outlined boxes) and the corresponding reactions steps. Show the pathway map.