

Try to answer these questions about the comparison between human red opsin and the bovine rhodopsin in PDB 1F88:

1. How many corresponding residues, and what percent of the residues, do the two proteins have in common (exact matches)? **Identities: 136/317 (43%)**
2. How many and what percent of corresponding residues are similar in chemical properties? **Positives: 197/317 (62%)**
3. How many gaps did the alignment program introduce, and how many residues in each gap, to get best alignment between human red opsin and 1F88? **One gap, with three residues.**
4. Find the longest string of exact matches between the two proteins. How many matches does it contain, and what are the beginning and ending residue numbers? **12 residues**
TQKAEKEVTRMV 259-270

Questions

1. Start at the NCBI Map Viewer. How many genes in the human genome contain the term "homeo" in their name? To be sure you find them all, search for "*homeo*". Number found: 350 .
2. Which chromosome contains the largest number of these genes? How many? Chromosome # 7 ; Number of "homeo" genes on this chromosome: 60 .
3. Among the genes found in question 1, find one that has a role in insulin action. **Fastest way to search is to search for "*homeo* AND insulin" without the quotation marks. Capital AND, OR, NOT can be used.**
Name of the gene: insulin promoter factor 1 . Four-character ID: IPF! .
4. What chromosome contains this gene? Chromosome # 13 .
5. According to OMIM, what is the role of the protein encoded by this gene?
Role (limit to 25 words): it serves as a master control switch for expression of both the exocrine and endocrine pancreatic developmental programs .
6. Obtain the protein sequence of this gene, in FASTA format. File name: HmPrt.txt
Clicking on the link for IPF1, we get the Entrez Gene entry. Down under mRNA and Proteins, we click on the link for NP_000200.1. This shows the Entrez Protein entry, from which we obtain the FASTA format by clicking "FASTA" under the Display pull-down menu and saving the contents.
7. Go to UniProt. How many annotated human genes in UniProt contain the term "homeo"? Note that "*" is automatically used as prefix and suffix unless you specify otherwise. Number found: 69 .
8. Make a phylogenetic tree of the first 24 of these genes plus the insulin-related gene found in question 3, a total of 25 sequences. Use the insulin-related gene as the "outgroup".
Clicking on the "All" column name, all 25 displayed genes will be selected. Deselect the last one to obtain 24, and click the "Retrieve" button at the bottom. Select the option to save the file in FASTA format. Then add the insulin-related gene to this fasta file using a text editor.
9. According to your tree, what two entries in this group are the most similar? **Using NJ-Tree with**

Branch length from KEGG, we get the tree (attached below).

Entry numbers Q53FN3 and A8K3N8 .

10. What entry is most similar to the insulin-related gene? SwissProt entry number A4D190 .

11. What can you find out about the function of this similar gene? **As seen under “Protein existence” this gene was inferred by homology and has not been reviewed.**

★ Unreviewed, UniProtKB/TrEMBL **A4D190** (A4D190_HUMAN)

[Contribute](#)
[Send fee](#)

Last modified February 26, 2008. Version 9. [History...](#)

[Clusters with 100%, 90%, 50% identity](#) | [Third-party data](#) | [Customize display](#) [TEXT](#) [XML](#) [RDF/XML](#) [GFF](#) [F](#)

[Names and origin](#) · [General annotation \(Comments\)](#) · [Ontologies](#) · [Sequences](#) · [References](#) · [Cross-references](#) · [Entry information](#)

Names and origin		Hide
Protein names	Homeo box A11 (EMBL EAL24220.1) <i>Also known as:</i> Homeobox A11, isoform CRA_b	
Gene names	Name: HOXA11 (EMBL EAL24220.1) ORF Names: hCG_1745677 (EMBL EAW93888.1), tcag7.172 (EMBL EAL24220.1)	
Organism	Homo sapiens (Human) (EMBL EAL24220.1)	
Taxonomic identifier	9606 [NCBI]	
Taxonomic lineage	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo	
Protein existence	Inferred from homology.	

12. Go to the Protein Data Bank. Search for models of human homeodomain proteins.

Using the Search by sequence, enter the sequence in the HmPrt.txt file. Then click the plus button next to it to open a second query field (as in Figure below). Select Taxonomy, which will bring up a dialog box to select the organism from the pre-defined list. Simply enter “Homo sapiens” in the search field above, and click on the resulting HOMO SAPIENS link to select it. Finally click “Evaluate Query” at the bottom.

Match all of the following conditions: [Advanced Search Tutorial \(Requires Flash\)](#)

Structure Id

Chain Id

Sequence (Blast/Fasta)

Sequence

E Cut Off

Search Tool

Taxonomy

11811 Structures

Remove Similar Sequences at Identity

1. How many models do you find?

Number: 21 .

2. What method of structure determination produced the first of these models?

PDB ID code: 1b72 . Method: X Ray Diffraction .

3. View the first model on the list with QuickPDB or your favorite molecular viewer.

What are the main secondary structural elements (helix, sheet, coil) in this protein?

Secondary structural elements: four alpha helices .

4. Give beginning and ending residue numbers of three secondary structural elements.

1. start: 212 end: 223 .

2. start: 230 end: 240 .

3. start: 244 end: 262 .

13. Find a model of a human homeodomain/DNA complex.

Using the Advanced Search, select a third query for “Molecule / Chain Type” and select “Yes” for “Contains DNA”.

1. How many models do you find? Number: 10 .

2. What method of structure determination produced the first of these models?

PDB ID code: 1b72 . Method: X Ray Diffraction .

3. View the first model on the list with QuickPDB or your favorite molecular viewer.

What secondary structural element(s) (helix, sheet, coil) interact with DNA?

Secondary structural elements: four alpha helices .

4. Give beginning and ending residues of main secondary structural element in contact with DNA.

Using QuickPDB, we can click on the two inner alpha helices on chains A and B respectively that would be in contact with the DNA sequence. From here, we can find residue numbers corresponding to these chains. The residue numbers for these chains can be evaluated on the Sequence Details tab for the PDB entry. Thus for Chain A: Residue start: 244 end: 262 .and Chain B: Residue start: 277 end: 293 .

