

CS612 Homework Assignment 2

Solution

1. **The PDB, Multiple Sequence Alignment** The Protein Data Bank (PDB) has a Molecule of the Month section to feature examples of molecules archived in the PDB. To find an entry about ABO blood types, point your browser to <http://www.rcsb.org/pdb/101/motm.do?momID=156>

- (a) From the Molecule of the Month article, identify the name of the enzyme that determines blood type A. Identify the name of the related enzyme that determines that blood type B.

Answer: The ABO blood type is determined by the type of sugars that are used to build the carbohydrates that cover the surfaces of red blood cells. The carbohydrate is built around a core of 5-13 sugars, called the H-antigen, that ends in a fucose sugar. For people with blood type O that's it, but for people with blood types A and B, there is an enzyme (of type glycosyltransferase) that adds one more sugar to the chain. Blood types A and B differ in the type of sugar. For type A this sugar is N-acetylgalactosamine, and for type B, it is the slightly smaller sugar galactose. Notice that glycosyltransferase is a type of enzyme, not a specific name. The name suggests that: a. It is an enzyme (the suffix -ase indicates it). b. The enzyme transfer a glycosyl, which is a generic name for any functional group derived from a sugar by removing its hydroxy group. This is slightly beyond the scope of the class so I wasn't very picky about the answers.

- (b) Based on the two enzymes identified in part A, explain how the two enzymes function to determine blood type B.

Answer: See above plus an addition: A person with type B blood doesn't have any antibodies to bind to the type B sugars (or they would attack his/her own blood cells). They do, however, have the antibodies for type B sugars (and vice versa for type A). So, if a person with a blood type B receives a transfusion that contains the type A antigens, their immune system will attack them. Therefore, they can only get blood from either another person with a type B or type O (which has none of the antigens). As a matter of fact, type O is a universal donor.

- (c) Use the sequences shown below. Perform a ClustalO run with the three sequences. The server can be found at <https://www.ebi.ac.uk/Tools/msa/clustalo/>. Paste the three sequences to the input windows, leave the default parameters and hit "submit" Include the letter-by-letter sequence comparison results in your answer. This is found on the results page under the Alignment heading. You can use the download tab in the ClustalO alignment or cut and paste from the web browser into a text program, such as Notepad, that you can submit as a separate file. You may have to adjust the text formatting, such as using Courier New font, to keep the sequences aligned.

>blood_type1

```
MVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKY
VAFKLFLETAEKHFVGVHRVHYVFTDQPAAVPRVTLGTGRQLSVLEVRAYKRWQDV
SMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDHVGVVEILTPLF'GTLHPGFYGSSE
AFTYERRPQSQAYIPKDEGDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVW
HDESHLNKYLRLRHKPTKVLSPPEYLWDQQLLWPAVLRKLRFTAVPKNHQAVRNP
```

>blood_type2

```
MVYPQPKVLTPSRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKY
```

VAF LKLFLETA EKHF MVGHRVHYVFTDQPAAVPRVTLGTGRQLSVLEV GAYKRWQDV
 SMRRMEMISDFSE RRF LSEVDYLVCVDVDMEFRDHVGV EILTP LFGTLHP SFY GSSRE
 AFTYERRPQS QAYIPKDEGDFYYMGAFFGGSVQEVQRLTRACHQAMMVDQANGIEAVW
 HDESHLNKYLRLRHKPTKVLSP EYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP

>blood_type3

MVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVA
 VAF LKLFLETA EKHF MVGHRVHYVFTDQPAAVPRVTLGTGRQLSVLEV RAYKRWQDV
 SMRRMEMISDFCQRRFLSEVDYLVCVDVDMEFRDHVGV EILTP LFGTLHPGFY GSSRE
 AFTYERRPQS QAYIPKDEGDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVW
 HDESHLNKYLRLRHKPTKVLSP EYLWDQQLLGWPSVLRKLRFTAVPKNHQAVRNP

- (d) Based on the sequence alignment, which two sequences (blood type 1, 2, or 3) are the most similar? Explain how you arrived at your answer. Note: for ClustalO a * indicates an amino acid match. Any other symbol indicates an amino acid difference.

Answer: Here is the alignment:

```

blood_type2 MVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVA
blood_type1 MVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVA
blood_type3 MVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVA
*****.*

blood_type2 FLKLFLETA EKHF MVGHRVHYVFTDQPAAVPRVTLGTGRQLSVLEV GAYKRWQDVSMRR
blood_type1 FLKLFLETA EKHF MVGHRVHYVFTDQPAAVPRVTLGTGRQLSVLEV RAYKRWQDVSMRR
blood_type3 FLKLFLETA EKHF MVGHRVHYVFTDQPAAVPRVTLGTGRQLSVLEV RAYKRWQDVSMRR
*****.*

blood_type2 MEMISDFSE RRF LSEVDYLVCVDVDMEFRDHVGV EILTP LFGTLHP SFY GSSRE AFTYER
blood_type1 MEMISDFCERRFLSEVDYLVCVDVDMEFRDHVGV EILTP LFGTLHPGFY GSSRE AFTYER
blood_type3 MEMISDFCQRRFLSEVDYLVCVDVDMEFRDHVGV EILTP LFGTLHPGFY GSSRE AFTYER
*****.:*****

blood_type2 RPQS QAYIPKDEGDFYYMGAFFGGSVQEVQRLTRACHQAMMVDQANGIEAVWHDESHLNK
blood_type1 RPQS QAYIPKDEGDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVWHDESHLNK
blood_type3 RPQS QAYIPKDEGDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVWHDESHLNK
*****:*.

blood_type2 YLLRHKPTKVLSP EYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP
blood_type1 YLLRHKPTKVLSP EYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP
blood_type3 YLLRHKPTKVLSP EYLWDQQLLGWPSVLRKLRFTAVPKNHQAVRNP
*****:*****

```

While obviously all three sequences are very similar, Blood type 1 and Blood type 3 are more similar to each other. They differ in only two positions, while Blood type 2 has at least 4 different positions.

2. **Protein structure search and classification:** Search the PDB with the entry 2BAA. Download the pdb file as text, and download FASTA sequence as text.

- a. How many atoms are there in the .pdb file?

Answer: A total of 1961 atoms, 1832 of which are protein atoms (I accepted both answers as correct).

- b. What atom type is atom 289?

Answer: CA for C – α (alpha carbon). Remember – this is NOT calcium!

- c. What is its amino acid 3 letter code?

Answer: ALA (for Alanine).

- d. What are the x,y,z coordinates of this atom?

Answer: -0.843 36.991 -3.128

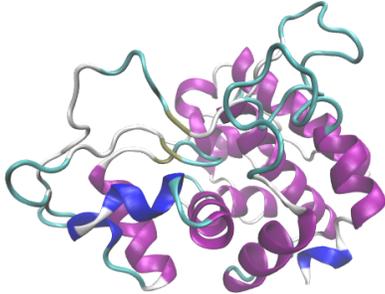
e. Search for 2BAA in SCOPe (at <http://scop.berkeley.edu>) – what is the class, fold and family of this protein? **Answer:**

- Class: Alpha and beta proteins (a+b)
- Fold: Lysozyme-like common alpha+beta motif for the active site region
- Superfamily: Lysozyme-like
- Family: Family 19 glycosidase

3. **Protein visualization:** Install VMD and Chimera, both available for free in the links given in class (handouts for this HW). Upload 2BAA.pdb (from the last question) in each viewer.

a. On vmd – upload 2BAA.pdb (from last question), display the protein in New Cartoon and color by secondary structure. How many helices do you see? (refer only to those colored in purple). How many beta strands? You can now render the display using File→Render. This will create a .tga file in your working directory. Print this file and submit it with the homework. You may want to change the color of the background from black to white to save ink. This can be done under the Graphics option of the menu, under Background. If the .tga doesn't work out you can print out a screenshot.

Answer: There were 9 or 10 helices (one short...) and three strands (all pretty short).



b. On chimera – upload the protein with the PDB code 1BWW (using File→fetch by id). Select all the beta strands (Select → structure → secondary structure → strand), and color them yellow by action → color and select yellow. Attach a printout by either saving a screenshot or File → save image. To save ink you can change the background color to white by selecting favorites → preferences and change "categories" to background. Click on color and change from black to white.

