

Q1) Below is a multiple alignment of 35 human sequences. The sequences have been aligned around a donor splice. That site is indicated as the boundary between the 'Dark blue' and 'Dark red' colours.

Calculate the frequencies for positions 6-5 i.e. that is 2 columns per person.

| Position | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|---|---|---|---|---|---|
| Counts A |   |   |   |   |   |   |   |   |   |   |
| Counts T |   |   |   |   |   |   |   |   |   |   |
| Counts C |   |   |   |   |   |   |   |   |   |   |
| Counts G |   |   |   |   |   |   |   |   |   |   |
| % A      |   |   |   |   |   |   |   |   |   |   |
| % T      |   |   |   |   |   |   |   |   |   |   |
| % C      |   |   |   |   |   |   |   |   |   |   |
| % G      |   |   |   |   |   |   |   |   |   |   |

```

-----Exon|intron-----
01234567890123456789
tatcacaATGGTAGGTAACT
TCAACCAGGAGTAAGTCTTG
GTTGCACCCTGTAAGTCTCA
tatcacaATGGTAGGTAACT
TCAACCAGGAGTAAGTCTTG
CTTGCGAGAGGTGTGACATG
GCTCTACTCGGTAAGGTGAC
GCCTGGAGAGGTAATGACCC
CAAAACCATTGTGAGTAATC
GCCAGAGCAGGTAAAATATC
GAACAGTCAGGTCTGTTGCT
GAAGGCCAGGTGAGCATAA
TCCTCTACAGGTGGGTACAT
GGCGTCCCGGTAAGTATGG
CCTCGTGCAGGTAAGATTAA
TGCATGACAGGTGAGTGTTA
GAAATGTACAGTAAGTCTCT
GGTTCTCTGGGTAAGTAGAG
AAATGTACAGGTGAGTACTG
ACCTCGCTTGGTACGTGGGA
AATCAGACAGGTATAGAAAC
AGGACAGAAGGTAATTTTCT
AACTATTTGGGTAGGTAGCA
AAACTTGAAGGTATGTTGTT
CTGGGATAAGGTAAGGATAT

```

**TTGCACCCAGGTTAGTGGAT**  
**ACTTCAATCGGTATGTTTTTC**  
**ACAGAGAAAAGTAAATTCCT**  
**AATGGGAAAGGTAACAACAA**  
**CATGCTACAGGTAGGTGAAT**  
 ggctaggATGGTGAGGGCGC  
**CGACGCGGGCGTGAGAGGCG**  
**CATTGAGAATGTGAGTTATT**  
**AACAGAGCAGGTACTTGTAT**  
**TGAACCAAAGGTAAGACAT**

**Q2)** Calculate the Entropy and Information Content using the formula below

Eq.1  $H(p) = -\sum p_a \log_2(p_a)$

Eq.2  $I = -H(p) + 2.0$

| position            | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 |
|---------------------|---|---|---|---|---|---|---|---|---|---|
| Entropy             |   |   |   |   |   |   |   |   |   |   |
| Information content |   |   |   |   |   |   |   |   |   |   |

**Q3)** Where does the constant 2.0 come from in Eq.1 ?

**Q4)** Draw an approximate Logo Plot by hand on the White board

**Q5)** Submit the multiple alignment to the WebLogo server <http://weblogo.berkeley.edu/>

Make both the Logo plot and a frequency plot  
 Explain what you see on the two plots.

**Q6)** Lysozyme is an enzyme that is a part of the innate immune system cleaving glycoside bonds (1->4)-beta-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine. These linkages are found in the cell wall of bacteria and thus a bacteria coming into our system is destroyed by cleaving glycoside bonds in the cell wall.

```

>sp|P61626|LYSC_HUMAN Lysozyme C OS=Homo sapiens GN=LYZ PE=1 SV=1
MKALIVLGLVLLSVTVQGVKVFERCELARTLKRGLGMDGYRGISLANWMCLAKWESGYNTRA
TNYNAGDRSTDYGFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRD
PQGIRAWVAWRNRCQNRDVRQYVQGCYV
    
```

Use the Blast2logo <http://www.cbs.dtu.dk/biotools/Blast2logo-1.0/> tool to produce a logo plot.

What are the most conserved type of amino acids ?

What is the potential role of these amino acids ?