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.

<table>
<thead>
<tr>
<th>Position</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counts A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Counts T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Counts C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Counts G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

-----Exon|intron-----
01234567890123456789
tatcacaATGGTAGGTAACCT
TCAACCAGGAGTAAGTCTTG
GTTGCAACCCTGTAAGTCTCA
tatcacaATGGTAGGTAACCT
TCAACCAGGAGTAAGTCTTG
CTTGCGAGAGGTGACATG
GCTCTACTCGGTAAAGGTGAC
GCCCTGAGAGGTAAATGACCC
CAAAACCATTGTGAGTAATC
GCCAGAGCAGTTAAATATC
GAACGTCAGGTCTGTTGCT
GAAAGGCCAGGTGACATAA
TCCCTCTACAGGTGGTGACAT
GGCGTCCCGGTAAGTATGG
CCTCGTGCCAGGTTAAGTATTAA
TGCATGACAGGTGGTGTTA
GAAATGTACAGGTAAATGACCC
GAAATGTACAGGTGGAGAG
GTTCTCTGGGTAAAGTATAGAG
AAAATGTACAGGTGGTGACAT
ACCTCGCTGGGTACGGTGGA
AATCAGACAGGTATAGAATAC
AGGACAGAAGGTAAATTTTCT
AACTTTTGGGTGAGTACA
AAACTTGAAAGGTATGTTTT
CTGGGATAAGGTAAAAATAT
Q2) Calculate the Entropy and Information Content using the formula below

Eq.1 \[ H(p) = - \sum a \cdot p_a \log_2(p_a) \]

Eq.2 \[ I = -H(p) + 2.0 \]

<table>
<thead>
<tr>
<th>position</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 MKALIVALGLVGLLSVTVOQKVFERCELARTLKRLGMDYRGISLANWMCLAKWESGYNTRA TNYNAGDRSTDTDYGFOINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRD PQGIRAWVAWRNRCQNRDVRQYVQCGV
Use the Blast2logo [http://www.cbs.dtu.dk/biotools/Blast2logo-1.0/](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?

**Q7)** What residues are involved in the active site (type and position) - Lookup via [Uniprot](http://www.uniprot.org)

**Q8)** Are the active site residues conserved in the Logo plot?