Making Sequence logos

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.

```
-----Exon|intron----
012345678901234567 89
tatcacaATGGTAGGTAACT
TCAACCAGGAGTAAGTCTTG
GTTGCACCCTGTAAGTCTCA
tatcacaATGGTAGGTAACT
TCAACCAGGAGTAAGTCTTG
CTTGCGAGAGGTGTGACATG
GCTCTACTCGGTAAGGTGAC
GCCTGGAGAGGTAATGACCC
CAAAACCATTGTGAGTAATC
GCCAGAGCAGGTAAAATATC
GAACAGTCAGGTCTGTTGCT
GAAGGCCCAGGTGAGCATAA
TCCTCTACAGGTGGGTACAT
GGCGTCCCGCGTAAGTATGG
CCTCGTGCAGGTAAGATTAA
TGCATGACAGGTGAGTGTTA
GAAATGTACAGTAAGTCTCT
GGTTCTCTGGGTAAGTAGAG
AAATGTACAGGTGAGTACTG
ACCTCGCTTGGTACGTGGGA
AATCAGACAGGTATAGAAAC
AGGACAGAAGGTAATTTTCT
AACTATTTGGGTAGGTAGCA
AAACTTGAAGGTATGTTGTT
CTGGGATAAGGTAAAAGTAT
TTGCACCCAGGTTAGTGGAT
ACTTCAATCGGTATGTTTTC
ACAGAGAAAAGTAAATTCCT
AATGGGAAAGGTAACAACAA
CATGCTACAGGTAGGTGAAT
ggctaggATGGTGAGGGCGC
CGACGCGGGCGTGAGAGGCG
CATTGAGAATGTGAGTTATT
AACAGAGCAGGTACTTGTAT
TGAACCAAAGGTGAAGACAT
```

Calculate the frequencies for positions 6-5. You have each been assigned one column on the upper right corner of the handout.

position	6	7	8	9	0	1	2	3	4	5
Counts A	15	12	20	4	0	0	23	23	5	6
Counts T	4	5	5	3	0	35	1	5	3	23
Counts C	8	13	4	2	0	0	1	2	1	1
Counts G	8	5	6	26	35	0	10	5	26	5
P(A)	0.43	0.34	0.57	0.11	0.0	0.0	0.66	0.66	0.14	0.17
P(T)	0.23	0.14	0.14	0.08	0.0	1.0	0.03	0.14	0.09	0.66
P(c)	0.11	0.37	0.11	0.05	0.0	0.0	0.03	0.06	0.03	0.03
P(G)	0.23	0.14	0.17	0.74	1.0	0.0	0.29	0.14	0.74	0.14

Q2) Calculate the Entropy (S) and Information Content (I) using the formula below

Eq.1 $$S(p) = -\sum_a p_a \log_2(p_a) = -\frac{1}{\log(2)} \sum_a p_a \log(p_a)$$

where $\log_2$ is the logarithm with base 2, and log is the logarithm with base 10 (or any base for that sake)

Eq.2 $I = 2.0 - S(p)$

position	6	7	8	9	**0**	**1**	**2**	**3**	**4**	**5**
Entropy	1.85	1.85	1.69	1.17	0	0	1.22	1.43	1.18	1.38
Information content	0.15	0.15	0.31	0.83	2	2	0.78	0.57	0.82	0.62

Q3) Where does the constant 2.0 come from in Eq.2?

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

If you have internet-access

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.