

1. Give the estimate of run time for aligning five sequences of 100 amino acids in length using summed pair scoring method. Briefly explain.

Sol: We know the formula is,

$$\text{DP run time for summed pair scoring method} = (2^{N-1}) (L^N)(N^2)$$

Here, we have 5 sequences to be aligned and all the letters in every sequence should be filled in the 100 positions

The No. of sequences = 5

Length of the sequence = 100

Substituting in the equation we have,

$$\text{DP run time for summed pair scoring method} = (2^{5-1}) (100^5)(5^2)$$

$$\text{Run time} = 2^4 * 25 * 10^{10}$$

$$= 4 * 10^{12}$$

2. One drawback of progressive alignment is described as “once a gap, always a gap”.

Explain why this is the case. What situation is the “gap” of “once a gap” inserted? What

situation is the “gap” in “always a gap” inserted?

Sol:

Let S1=ATTGA and S2=ATGCC & S3=ATCCAAT be two other sequences to be aligned with S1.

Let the pair wise alignment of S1 & S2 be:

S1= A T T G A

S2= A T G C C

Let the pair wise alignment of S1 & S3 be:

S1= A T \_ T G A \_

S3= A T C C A A T

Combining three sequences we have,

P      0   1   2 3   4   5   6

S1= A T \_ T G A \_

S2= A T \_ G C C \_

S3= A T C C A A T

We can observe that there are no gaps for the pair-wise alignment.

On the introduction of the alignment s3 the gap has been introduced at residue 2.

The gaps are inserted at residue 6 of both s1 and s2. The gap inserted in the main sequence cannot be changed. This is always a drawback.

3. Suppose the pair-wise alignment score is given as follows. The smaller the score,  
the more similarity between the two sequences. Use UPGMA method to build the similarity tree or the guide tree.

Sol:

	S1	S2	S3	S4
S1	0	1	4	3
S2	1	0	2.5	1.5
S3	4	2.5	0	2
S4	3	1.5	2	0

Distance from node 'u' of (s1, s2) =  $\frac{1}{2} = 0.5$

Distance between S3 and (S1, S2) =  $[(4*1) + (2.5*1)]/2 = 6.5/2 = 3.25$

Distance between S4 and (S1, S2) =  $[(3*1) + (1.5*1)]/2 = 4.5/2 = 2.25$

	(s1, s2)	S3	S4
(s1, s2)	0	3.25	2.25

S3	3.25	0	2
S4	2.25	2	0

Distance from node 'v' from (s3, s4) =  $2/2 = 1$  and group (s3,s4):

	(s1, s2)	(s3, s4)
(s1, s2)	0	2.75
(s3, s4)	2.75	0

Distance between (S3, S4) and (S1, S2) =  $[(3.25*1) + (2.25*1)]/2 = 5.5/2 = 2.75$

Distance from (s1, s2) & (s3, s4) to node r:  $2.75/2 = 1.375$

### Similarity Tree:

