

Chapter 5: Sequence alignment.

Answers.

1. Aligning with:

AGG WCTTY and BWCTY
 $s(i,i) = 1$, $s(i,j) = 0$, no gap penalty.

	A	G	G	W	C	T	T	Y
G	0	1	1	0	0	0	0	0
W	0	0	1	2	1	1	1	
C	0	0	1	1	3	2	2	2
T	0	0	1	1	2	4	4	3
Y	0	0	1	1	2	3	4	5

of alignments with score 5: 14

2) Same as in 1, but now a gap penalty of 1 (i.e. -1)

	A	G	G	W	C	T	T	Y
G	0	1	1	-1	-1	-1	-1	-1
W	-1	0	0	1	-1	-1	-1	-1
C	-1	-1	0	0	2	0	0	0
T	-1	-1	-1	0	0	3	2	1
Y	-1	-1	-1	-1	0	1	3	3

Assume gaps at the beginning count.

3. The 3 sequence alignments are:

THESE SENTENCES ALIGN -- NICELY Score: 180
 THESE SENTENCES ALIGN -- NICELY
 THE SEQUENCE ---- ALIGNED NICELY
 (Red vertical lines indicate matches between the two sequences. Blue arrows labeled 'gap 1' and 'gap 2' point to the gaps in the second sequence.)

THESE SENTENCES ALIGN -- NICELY Score: 190
 THESE -- Q -- E N S E -- ALIGNED NICELY
 (Red vertical lines indicate matches. Blue arrows labeled 'gap 1', 'gap 2', and 'gap 3' point to gaps in the second sequence. A blue arrow labeled 'gap 4' points to a gap in the first sequence.)

THESE SENTENCES ALIGN -- NICELY Score: 190
 THE -- SEQ -- E N S E -- ALIGNED NICELY
 (Red vertical lines indicate matches. Blue arrows labeled 'gap 1', 'gap 2', and 'gap 3' point to gaps in the second sequence. A blue arrow labeled 'gap 4' points to a gap in the first sequence.)

All 3 alignments have the same score, 190

4. When we account for gaps:

Alignment 1: $Score = 190 - \underbrace{5 - 2 \times 4}_{\substack{\text{first gap} \\ \sqrt{\text{gap 1}}}} - \underbrace{5 - 2 \times 2}_{\substack{\text{second gap} \\ \sqrt{\text{gap 2}}}} = 168$

Alignment 2: $Score = 190 - \underbrace{5 - 2}_{\sqrt{\text{gap 1}}} - \underbrace{5 - 2 \times 2}_{\sqrt{\text{gap 2}}} - \underbrace{5 - 2}_{\sqrt{\text{gap 3}}} - \underbrace{5 - 2 \times 2}_{\sqrt{\text{gap 4}}} = 15$

Alignment 3: $Score = 190 - \underbrace{5 - 2 \times 2}_{\sqrt{\text{gap 1}}} - \underbrace{5 - 2}_{\sqrt{\text{gap 2}}} - \underbrace{5 - 2}_{\sqrt{\text{gap 3}}} - \underbrace{5 - 2 \times 2}_{\sqrt{\text{gap 4}}} = 15$

Alignment 1 is optimal.

5. As alignment B and C have similar gaps, we need to change the scores of individual matches = answer d

6. The size of the alphabet matters:
 you do better with proteins (20 letters)
 than with DNA (4 letters)

answer: C

7.8. ^{Scores:} ATGCA CTCCA GTAAA
 CTCCA 3 GTAAA 2.4 CTGCA 2.6
 ATGCA CTCCA
 GTAAA 2.9 CTGCA 3.6
 ATGCA 3.3
 CTGCA

Most dissimilar: CTCCA with GTAAA

Most similar: CTCCA with CTGCA

9. PAM250, just like BLOSUM, reflects
 the properties of amino acids: answer C

10. W and C are very different: W is
 aromatic, bulky, C is small, involved in
 disulfide bridges → answer B

11. All ~~5~~ amino acids (I, L, V, F, V) are hydrophobic
 (non polar) → answer A

12. $P_{\text{value}} \approx E_{\text{value}} = 10^{-4}$ answer C
 P value: probability of a random match with same score