

Assignment #4: Sequence Alignment

1. Very briefly differentiate FASTA and BLAST (max. 300 words). [1 point]
2. Very briefly discuss the significance of employing a pseudofactor in PSSM building and the caveats that come with its introduction (max. 300 words). [1 point]
3. Given the peptides below, provide a possible alignment following the BLAST procedure (please report your full procedure and all your tables/matrices/etc.):

C C Y E K R R K H Y C Q H C N Q W W V E W
C C L V I L P G H C D D I E W

Parameters: $w=3$

acceptance = 25

extension = 20

matrix = BLOSUM50

gap open = -5

gap extend = -1

- Build the *wmer* matrix and find the HSPs [2 point]
 - Perform the extensions [2 point]
 - Connect final HSPs [3 points]
4. The sequences below are each a segment of the WNT3A genes involved in tail formation in several animals (already aligned):

RAT AGCAGAGAGTCAGTGAATACAGTGG
CAT ATCTCCAGCCCCAGGGGCCGGCGG
DOG CACATTTCTGAGGTGGGTCTGTG
DEER AAAATCCCATGGACAGAGGAGCTTG
CATTLE AAAATCCCATGGATAGAGGAGCCTG

- a. Build a PSSM matrix. Use $B=1$ and Random Freq. of Nucleotide = 25% each. Follow the three-step procedure as in the Exercise Session. [3 points]
- b. Below are the equivalent segments of the WNT3A gene in chimpanzees and humans that are believed to be involved in tail formation. Is there a greater likelihood that each sequence below is an instance of a “tail gene” or is it more likely that each is only a random sequence? How much is the likelihood? [1 point]

CHIMPANZEE CACGCCTCCCGGTCAGAGACAAGAG
HOMO SAPIENS CAAGCCTCCCGGTCAGAGACAAGAG