

11300A Bioinformatics

Homework 2

Nov. 14, 2023

1. Sequence Alignment Using Dynamic Programming (DP)

There are two amino acid sequences, seq1: COELACANTH and seq2: PELICAN. Obtain the global alignment by using DP (the Needleman-Wunsch algorithm).

$$\text{Scoring scheme: } \begin{cases} +1 & \text{for letter that match} \\ -1 & \text{for mismatches} \\ -1 & \text{for gaps} \end{cases}$$

	-	P	E	L	I	C	A	N
-	0	-1	-2	-3	-4	-5	-6	-7
C	-1	-1	-2	-3	-4	-3	-4	-5
O	-2	-2	-2	-3	-4	-4	-4	-5
E	-3	-3	-1	-2	-3	-4	-5	-5
L	-4	-4	-2	0	-1	-2	-3	-4
A	-5	-5	-3	-1	-1	-2	-1	-2
C	-6	-6	-4	-2	-2	0	-1	-2
A	-7	-7	-5	-3	-3	-1	1	0
N	-8	-8	-6	-4	-4	-2	0	2
T	-9	-9	-7	-5	-5	-3	-1	1
H	-10	-10	-8	-6	-6	-4	-2	0

Figure 1 The Matrix of the DP results

There are two kinds of alignments:

COELACANTH

Type1: **- P E L I C A N - -**

Type2: **P - E L I C A N - -**

2. Use two different dot matrix analysis servers to analyze the sequence of the human low density lipoprotein receptor (NP_000518). Please run a dot matrix analysis of this protein against itself.

A. First use Dottup

Set the word size to 2 (“word size” is basically the same as “window”).

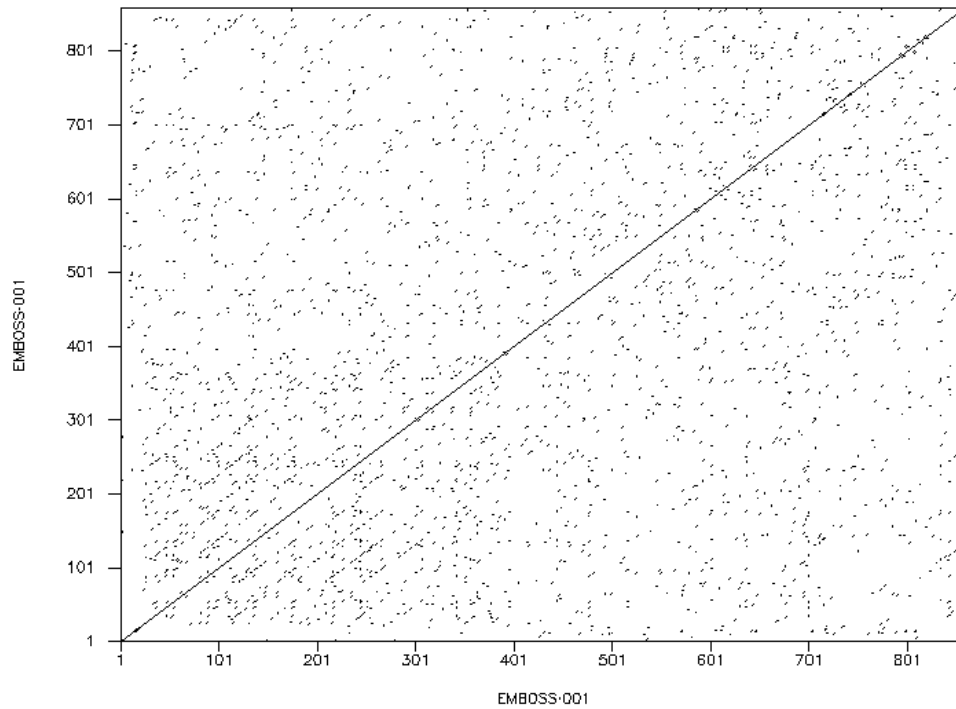
Using a word size of 2, the algorithm will scan a window of 2 amino acids and put one dot in the matrix when the two sequences have

identical amino acids. Dottup has no threshold, so it is simpler than Dotmatcher.

a) Print (or copy and paste) the output from Dottup and turn it in.

Answer:

Dottup: fasta::emboss·dottup—E20231107—063308—0461—95917...
Tue 7 Nov 2023 06:33:54



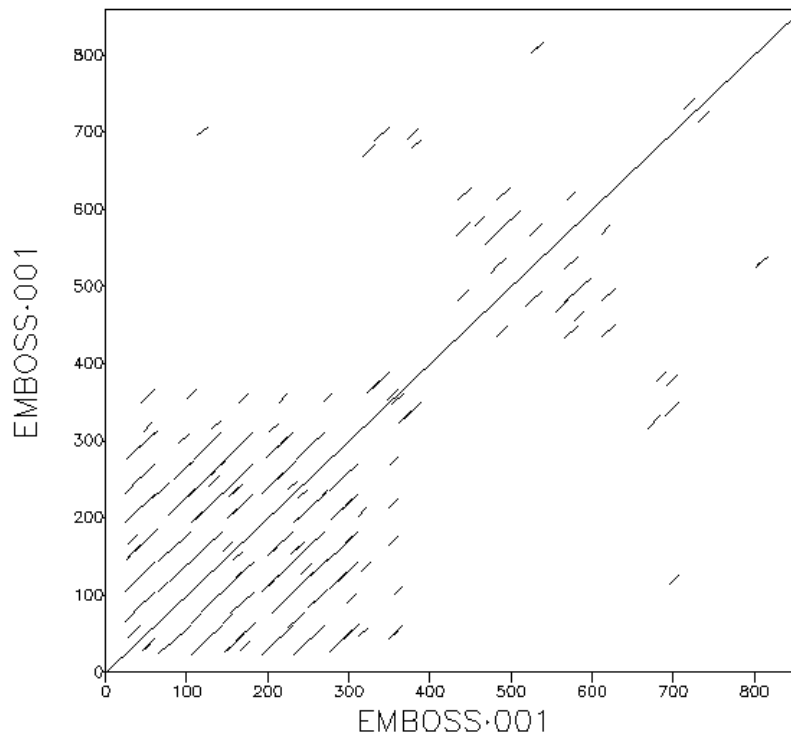
B. Now use Dotmatcher

Set the window to 10 and threshold to 23 in order to filter out the “noise.”

a) Print (or copy and paste) the output from Dotmatcher and turn it in.

Answer:

Dotmatcher: fasta::emboss-dotmatcher-E20231107-063550-08...
(windowsize = 10, threshold = 23.00 07/11/23)



C. Examine the two dot matrices and use them to answer the following questions.

a) What does the long diagonal from one corner to the other represent?

Answer:

The long diagonal from one corner to the other represents identical residues between the two sequences being compared. In this case, since the protein is being compared against itself, it represents the entirety of the sequence.

b) What do the shorter diagonals (mostly in the lower left corner) indicate about this protein?

Answer:

The shorter diagonals in the lower left corner are a representation of similar or repeated sequences within the protein. These could indicate domains or motifs that are repeated within the protein's structure.