

Name: \_\_\_\_\_

ID: \_\_\_\_\_

## Part II (2 problems; total 40 points)

### Problem 1 (4 questions, each 5 points: total 20)

Concanavalin A (ConA) is a lectin (carbohydrate-binding protein) originally extracted from the jack bean, *Canavalia ensiformis*. It binds specifically to certain sugars, glycoproteins, and glycolipids. The structure of concanavalin has been determined by X-ray crystallography, and is stored in the PDB. You are interested to know how similar this lectin is from the other lectins that are known, in particular to the lectin from peanut, whose structure is also known. First, you run BLAST, starting from the sequence of ConA. BLAST does find a match with the peanut lectin:

```
> pdb|2PEL|A S Chain A, Peanut Lectin
pdb|2PEL|B S Chain B, Peanut Lectin
pdb|2PEL|C S Chain C, Peanut Lectin
> 61 more sequence titles
Length=236
```

```
Score = 91.3 bits (225), Expect = 3e-22, Method: Compositional matrix adjust.
Identities = 53/116 (46%), Positives = 67/116 (58%), Gaps = 1/116 (1%)
```

```
Query 1 ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNMQDGKVGTAHIIYNSVDKR 60
A V VE DTY N++ DP H+GID+ SV S KT WN G V +IY+S K
Sbjct 114 AGHFVGVVEFDTYNSSEYNDPPTDHSVGDVNSVDSVKTVPWNSVSGAVVKVTVIYDSSSTKT 173

Query 61 LSAVVSYPNADATSVSYDVLNDVLPWVVRVGLSASTGL-YKETNTILSWSFTSKL 115
LS V+ N D T+++ VDL LPE V+ G SAS L ++ + I SWSFTS L
Sbjct 174 LSVAVTNDNGDITITIAQVVDLKAKLPERVKFGFSASGSLGGRQIHLIRSWSTSTL 229
```

```
Score = 76.3 bits (186), Expect = 1e-16, Method: Compositional matrix adjust.
Identities = 44/106 (42%), Positives = 64/106 (60%), Gaps = 7/106 (7%)
```

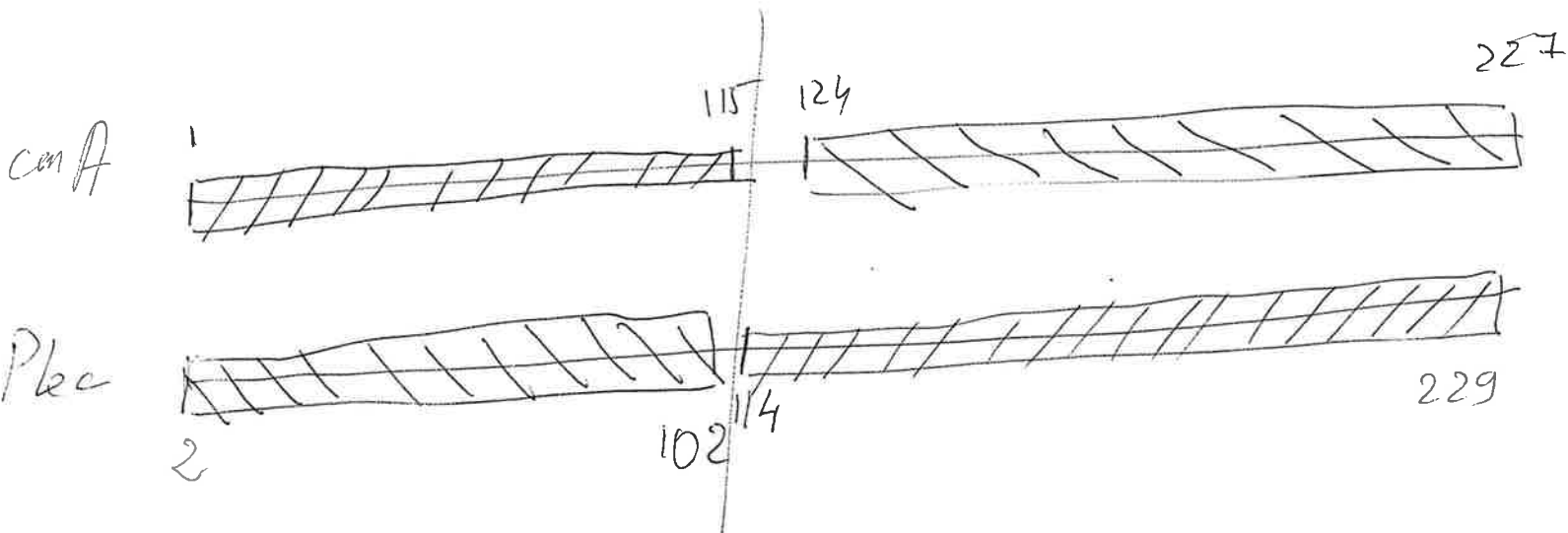
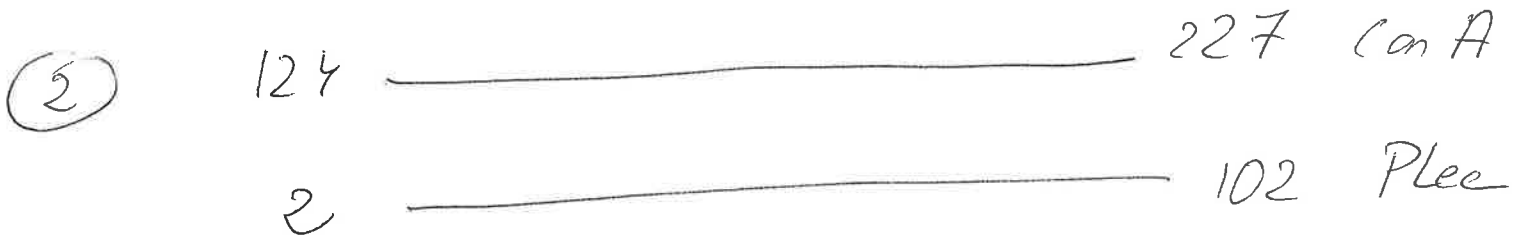
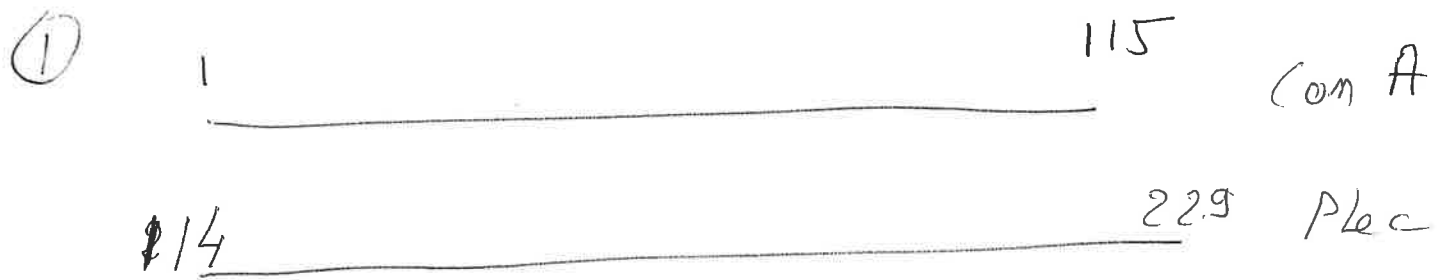
```
Query 124 DALHFMFNQFSKQKDLILQGDATTGTGDNLELTRVSSNGSPEGSSVGRALFYAPVHIWE 183
+ + F FN FS+ + QGD T ++GN++LT ++ + +SVGR L+ PV IW
Sbjct 2 ETVSFMFNFSFSEGNPAINFQGDVTVLSNGNIQLTNLN-----KVNSVGRVLYAMPVRIWS 56

Query 184 SSAATV-SFEATFAFLIKS-PDSDPADGIAFFISNIDSSIPSGSTG 227
S+ V SF +F+F +K D PADGI FFI+ D+ IP+GS G
Sbjct 57 SATGNVASFLTSFSFEMKDIKDYDPADGIIFFIAPEDTQIPAGSIG 102
```

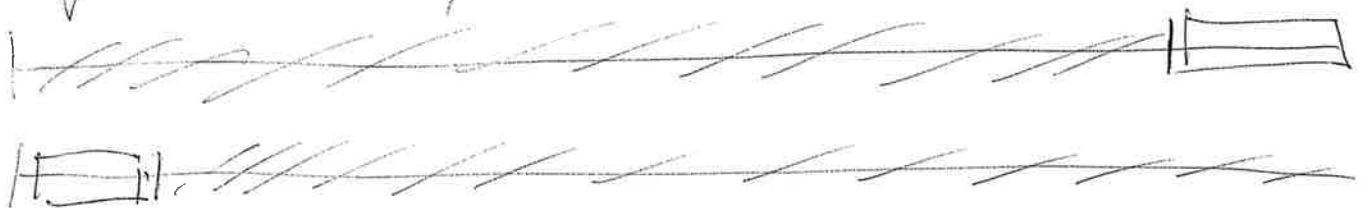
- a) BLAST found two alignments between subsets of the sequences of ConA and the peanut lectin. Are these two alignments significant? Justify your answer



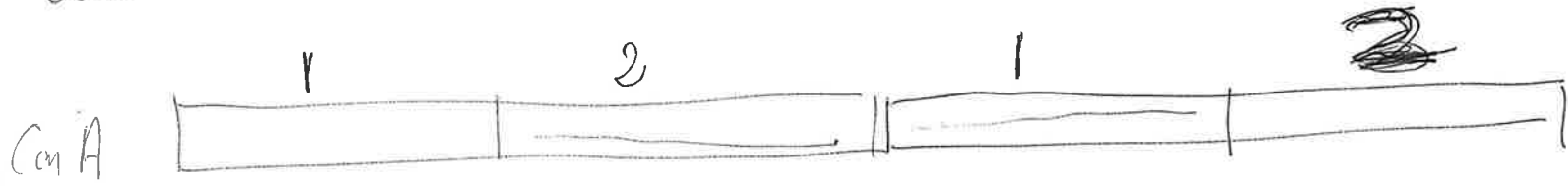
I found 2 subregions of can A  
 that are similar to 2 subregions of  
 Plec, with high significance. ①



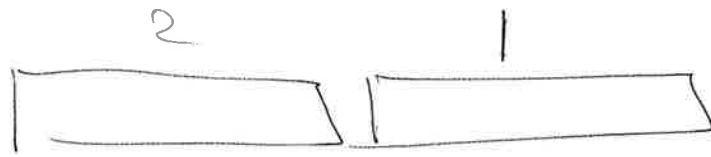
Possibly for other species



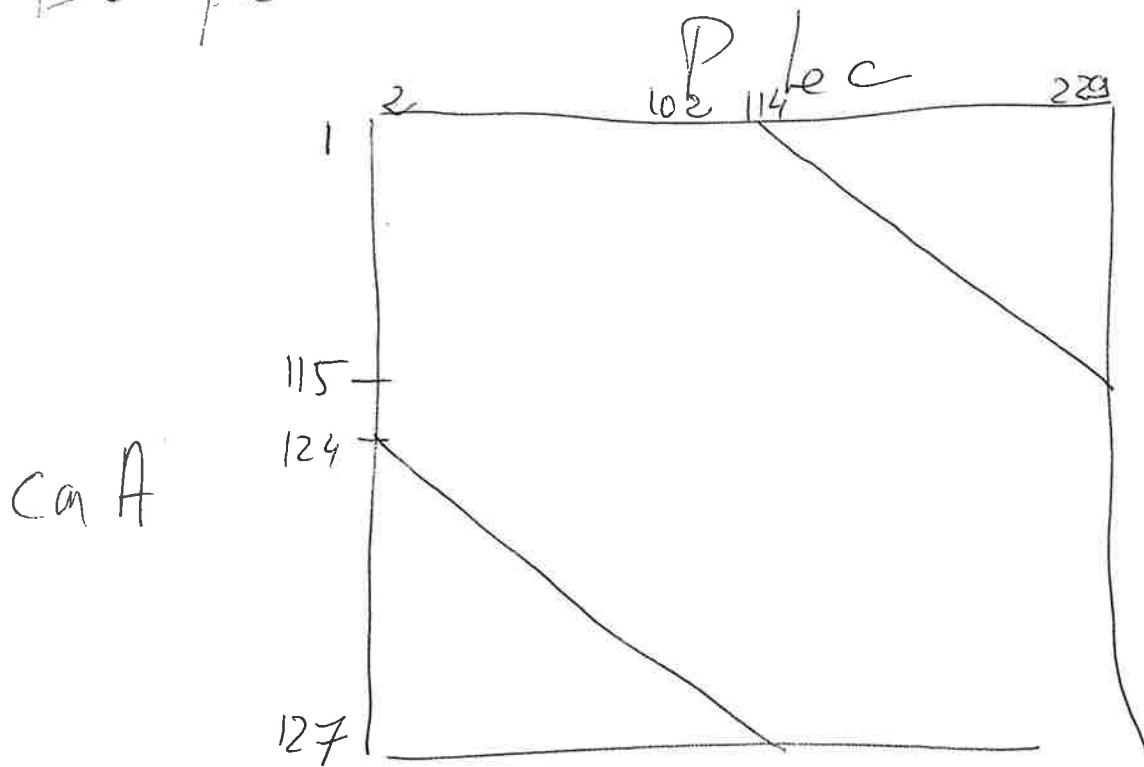
Strategy to detect permutations (circular) 2  
 with BLAST



Create an artificial sequence with 2 Can A repeated.



Dot plot



Name: \_\_\_\_\_

ID : \_\_\_\_\_

**Part II (2 problems; total 20 points)**

**Problem 1 (10 points)**

The **template** strand of a sample of double-helical DNA contains the sequence:  
(5') TTACGAGATCAT (3')

a) What is the sequence of the corresponding coding sequence (label its 5' and 3' end)?

b) What is the corresponding coding mRNA sequence (label its 5' and 3' end)?

c) What is the resulting amino acid sequence (label its Nter and Cter)?

**Problem 2 (10 points)**

A probe that was sent to the moon Titan of Saturn brings back a strange bacterium. Analysis of this bacterium shows that its DNA only contains 3 bases, D, N, and K, and that its proteins are made of up to 8 amino acids, which we will label as 1, 2, 3, 4, 5, 6, 7, and 8, for simplicity. You assume that the codon size associated with the unusual genetic code of this bacterium is 2 and it proves to be correct. You are assigned the task to find this genetic code. You test some DNA sequences and check the resulting proteins that you obtained. Your results can be summarized as follows:

Alphabets

DNA

Name: \_\_\_\_\_  
ID: \_\_\_\_\_

DNA contains	Protein contains
....KKKK....	Amino acid 1 only
....NNNN....	Amino acid 2 only
....DDDD....	Amino acid 3 only
...KNKNKN....	Amino acid 4 or amino acid 5
...KDKDKD....	Amino acid 6 or amino acid 7
....NDNDND....	Amino acid 8
....KNDKNDKND...	Amino acid 4, 6, or 8

The DNA sequences given are fragments. The reading frame could correspond to the first or the second nucleotide on each of these fragments. Can you fill up the genetic code table below?

		Second base		
		K	N	D
First base	K	1	4	7
	N	5	2	8
	D	6	8	3

KNKNKN → KN and NK 4, 5  
 KDKD → KD and DK 6, 7  
 KNDKND → (KN), (ND), (DK) (4), (6), (8)

**Part III extra credit (1 problem; total 5 points)**

You have isolated an important gene that regulates the size of a newly found frog from the island of Borneo. You have also been able to find the sequence of the protein encoded by this gene. You suspect that sequences similar to this sequence can be found in other organism, but with circular permutation:

