Final (05.06.2012) (9:00-10:30)
_____ **0114820 – Introduction to Bioinformatics** _____

Q1(15)	Q2(20)	Q3(15)	Q4(15)	Q5(10)	Q6(10)	Q7(15)	Total(100)

Q1. An amino acid sequence is given as **DIK**. Determine the possible DNA sequences which results in the synthesis of the given amino acid sequence. (Use the genetic code table)   (15)

A1.

To determine possible DNA sequences, we need to apply phases of central dogma of MB in reverse order.
Looking at the genetic code table, Amino acids in DIK sequence can be translated from the following RNA triplets.

Abbreviation	Amino Acid	Possible triplets
D	Aspartic Acid	GAU, GAC
I	Isoleucine	AUU, AUC, AUA
K	Lysine	AAA, AAG

We have 12 possibilities to obtain the DIK sequence. We can visualize them in the following table.

*											
GAU						GAC					
AUU		AUC		AUA		AUU		AUC		AUA	
AAA	AAG	AAA	AAG	AAA	AAG	AAA	AAG	AAA	AAG	AAA	AAG

We, then apply reverse transcription to find possible DNA sequences by backsubstituting each RNA with its DNA counterpart.
12 Possible sequences are (RNA → DNA defines reverse transcription operation):

01-)   GAU AUU AAA→CTATAATTT
02-)   GAU AUU AAG→CTATAATTC
03-)   GAU AUC AAA→CTATAGTTT
04-)   GAU AUC AAG→CTATAGTTC
05-)   GAU AUA AAA→CTATATTTT
06-)   GAU AUA AAG→CTATATTTC
07-)   GAC AUU AAA→CTGTAATTT
08-)   GAC AUU AAG→CTGTAATTC
09-)   GAC AUC AAA→CTGTAGTTT
10-)   GAC AUC AAG→CTGTAGTTC
11-)   GAC AUA AAA→CTGTATTTT
12-)   GAC AUA AAG→CTGTATTTC

Q2. Given the following scoring rules: (match score = +1, mismatch score = 0, gap penalty = -1)

    a.   Fill in the dynamic programming matrix for local alignment of the sequences **ACTC** and **ACAGTA** (10)
    b.   Identify the best local alignment. (10)

----

a.

	-	A	C	A	G	T	A
-	0	0	0	0	0	0	0
A	0	1	0	1	0	0	1
C	0	0	2	1	1	0	0
T	0	0	1	2	1	2	1
C	0	0	1	1	2	1	2

$$S_{i,j} = \text{MAXIMUM} \; [\; S_{i-1,\,j-1} + s(a_i, b_j) ,$$
$$S_{i,j-1} + w ,$$
$$S_{i-1,j} + w ,$$
$$0 \qquad\qquad]$$

b. For the given scoring rules, there are 5 possible local alignments (all giving a maximum score of 2):

1.
```
A  C  A  G
A  C  T  C
1 + 1 + 0 + 0 = 2
```

2.
```
A  G  T  A
A  C  T  C
1 + 0 + 1 + 0 = 2
```

3.
```
A  G  T
A  C  T
1 + 0 + 1 = 2
```

4.
```
A  C  A
A  C  T
1 + 1 + 0 = 2
```

5.
```
A  C
A  C
1 + 1 = 2
```

==

Q3.
    a.   What are motivations for multiple sequence alignment? Explain (08)
    b.   List the approaches to multiple sequence alignment. (07)

----

a. Similar genes can be conserved across species that perform similar or identical functions. Many genes are represented in highly conserved forms across organisms. By performing a simultaneous alignment of multiple sequences having similar or identical functions we can gain information about which regions have been subject to mutations over evolutionary time and which are evolutionarily conserved. Such knowledge tells which regions or domains of a gene are critical to its functionality. Sometimes genes that are similar in sequence can be mutated or rearranged to perform an altered function. By looking at multiple alignments of such sequences, we can tell which changes in the sequence have caused a change in the functionality. Multiple sequence alignment yields information concerning the structure and function of proteins, and can help lead to the discovery of important sequence domains or motifs with biological significance while at the same time uncovering evolutionary relationships among genes.

b. There are four approaches to multiple sequence alignment:

       Dynamic Programming

       Progressive Alignment

       Iterative Alignment

       Statistical Modeling

===

Q5.
  a. What is a database system? Explain.   (07)
  b. How many database systems exist? Explain   (08)

----

  a. A database system is a computer program (or group of programs) that provides a mechanism to define and manipulate one or more databases.

  b.

    1. Personal database systems:
        Designed to run on PCs
            Access, Paradox, FileMaker, dBase
    2. Enterprise database systems:
        Designed to support efficient storage and retrieval of vast amount of data
            Interbase, Ingres, SQL Server, Informix, DB2, Oracle
    3. Open source database systems:
        Free (Usually for Linux OS)
            PostgreSQL, MySQL

===

Q6.   What does the following Perl one-liner do? Explain.   (10)

    $ perl -npe 'last if /\d{5}$/;' dna.dat

This is a one-liner that prints only those lines from the dna.dat disk-file that do not end in five digits.

equivalent perl program would given as:

```
while ( <> )
{
        last if /\d{5}$/;
}
continue
{

        print $_;
}
```

Q7.   Write a Perl code that tests if a specific username entered from the keyboard is "ogrenci".   (10)

```
print "Enter the username: ";
$username = <STDIN>;
chomp $username;
if       ($username =~ /ogrenci/) {
         print "Welcome ogrenci!\n\n";}
else     {print "Bad username, sorry!\n\n";}
```

===
Q4.   For the sequences given in the table;
       construct a dot plot using a sliding window of size 3 and a similarity
       cutoff of two nucleotides.   (15)

Compare triplets by shifting one residue at one time (starting at the first position):

	G	C	T	A	G	T	C	A
G	●							
A								
T			●					
G	●			●				
G					●			
T						●		
C							●	
A								

√   G A T - G C T   (This is a match)
X   GAT - CTA
X   GAT - TAG
X   GAT - GTC
X   GAT - TCA
X   GAT - CA-
X   ATG - GCT
              ●
              ●
              ●
√   G G T - G C T   (This is a match)
              ●
              ●
              ●
√   T C A - T C A   (This is a match)
              ●
              ●
              ●

Alternatively, the following solution is also possible (matching starts by considering a gap at the begining of the array and marking the box corresponding to middle residues for a match).

-	G	C	T	A	G	T	C	A
G								
A		●						
T								
G				●				
G		●			●			
T						●		
C							●	
A								●

X   -GA - -GC
X   -GA - GCT
              ●
              ●
√   G A T - G C T   (This is a match, dot is placed at the
              ●   intersection of A and C)
              ●
√   C A - - C A-   (This is a match, dot is placed at the
                               intersection of A and A)