

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

Score: \_\_\_\_\_ / \_\_\_\_\_

## Exam1

Dear All,

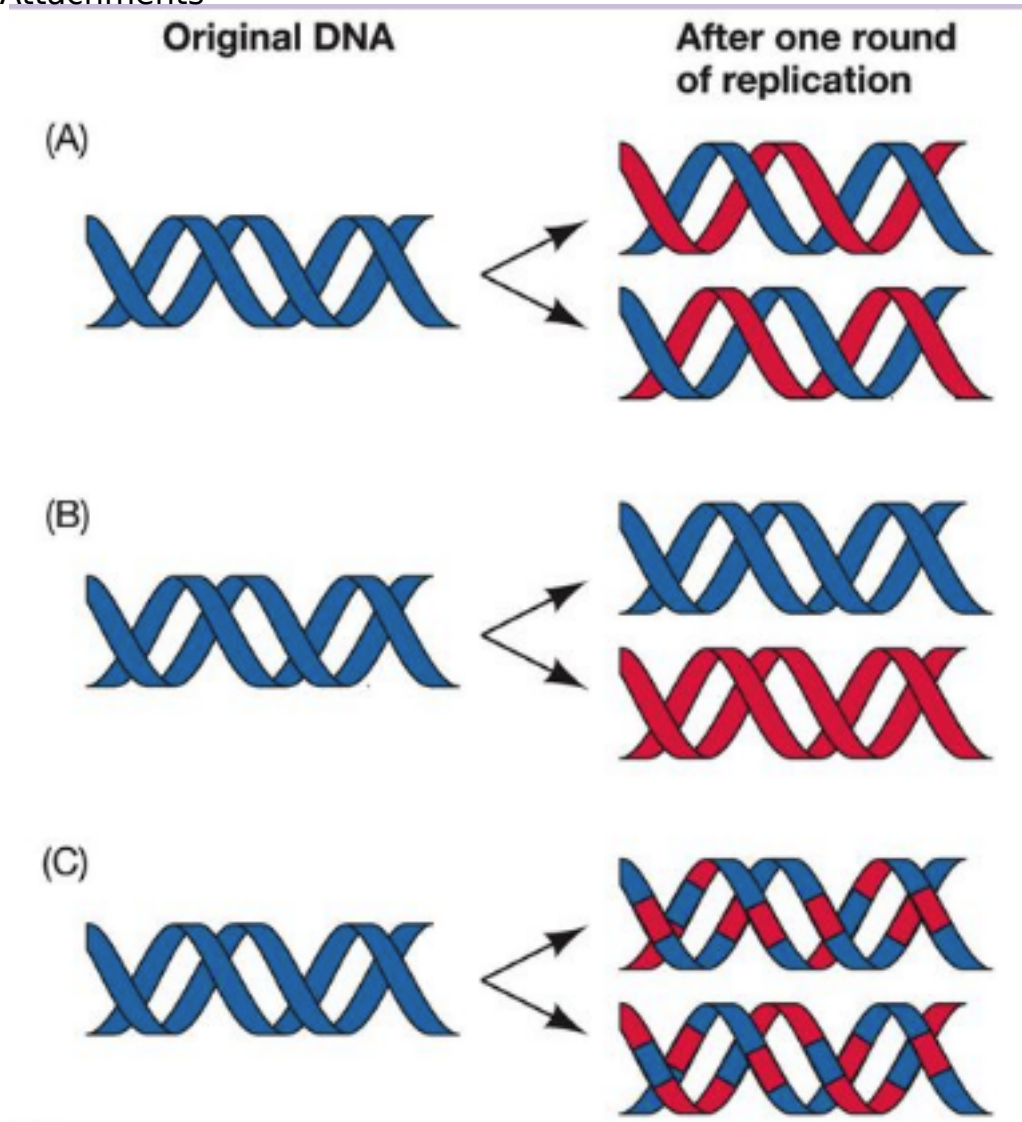
Here is your exam 1 for Bioinformatics (CS300 and Bio300). This exam is closed notes and closed book; you are not allowed to run any code on any computer to answer any questions. Unless you have made arrangements with the instructor, it is assumed that you will have 75 minutes to complete this exam. The exam is to be completed in Alden hall only. By the submission of your exam, you are agreeing to adhere to the honor code pledge.

Best of luck to all,  
Dr. Bonham-Carter

## Part 1

What model of replication is illustrated by part (a) of the included diagram?  
Explain your reasoning.

#### Attachments



- ☐ A. The semi-conservative model      Feedback: -----
- ☐ B. The min-max-conservative model      Feedback: -----
- ☐ C. The nieve cross-over dispersive model      Feedback: -----
- ☐ D. The conservative model      Feedback: -----
- ☐ E. The dispersive model      Feedback: -----

Answer Point Value: 10.0 points

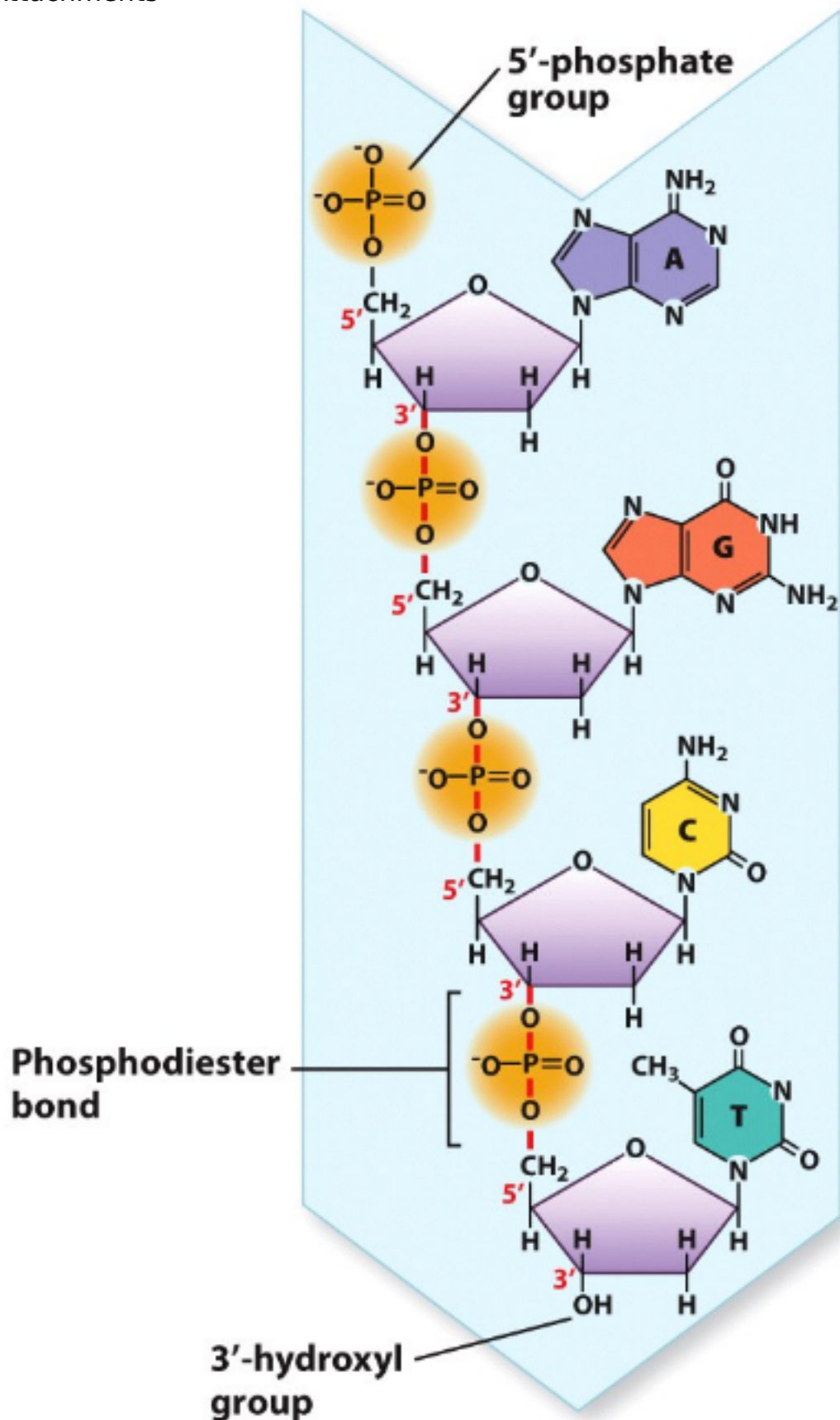
Answer Key: A

Correct Feedback: -----

Incorrect Feedback: -----

In the DNA sequence 5'-AGCT-3', (shown in the illustration) how is the phosphodiester linkage between the adenine and the guanine connected? Explain your reasoning.

Attachments



- ☐ A. The 4' end of the adenine to the 2' end of the guanine.

Feedback: -----

- ☐ B. The 5' end of the adenine to the 3' end of the guanine. Feedback: -----
- ☐ C. The 5' end of the adenine to the 5' end of the guanine. Feedback: -----
- ☐ D. The 3' end of the adenine to the 3' end of the guanine. Feedback: -----
- ☐ E. The 3' end of the adenine to the 5' end of the guanine. Feedback: -----

Answer Point Value: 10.0 points

Answer Key: E

Correct Feedback: -----

Incorrect Feedback: -----

3

In the DNA of certain bacterial cells, 16% of the nucleotides are adenine. What are the percentages of the other nucleotides in the bacterial DNA? Explain your reasoning.

- ☐ A. 34% thymidine, 34% guanine, 16% cytosine Feedback: -----
- ☐ B. 34% uracil, 16% guanine, 16% cytosine Feedback: -----
- ☐ C. 34% thymidine, 16% guanine, 34% cytosine Feedback: -----
- ☐ D. 16% thymidine, 34% guanine, 34% cytosine Feedback: -----
- ☐ E. 34% thymidine, 34% guanine, 16% cytosine Feedback: -----

Answer Point Value: 10.0 points

Answer Key: D

Correct Feedback: -----

Incorrect Feedback: -----

The sequence (DNASeq) is a string of DNA and is shown below in the 5' to 3' direction. We note that this sequence is already the coding (i.e., non-template) strand of DNA.

DNASeq = 5'-ATGATCTCGTAA-3'

Find the corresponding RNA sequence. Explain your reasoning.

- A.  
☐ ATGATCTCGTAA Feedback: -----
- B.  
☐ TACTAGAGCATT Feedback: -----
- C.  
☐ TTACGAGATCAT Feedback: -----
- D.  
☐ AUGAUCUCGUAA Feedback: -----
- E.  
☐ AATGCTCTAGTA Feedback: -----

Answer Point Value: 10.0 points

Answer Key: D

Correct Feedback: -----

Incorrect Feedback: -----

The sequence (RNASeq) is a string of RNA and is shown below in the 5' to 3' direction and may be used to procure protein without any change to it. Use the displayed codon table to find the corresponding protein sequence. Explain your reasoning.

RNASeq = 5'-AAACGCGGCGCTGACGCGCGGAAA-3'

## Attachments

Standard genetic code									
1st base	2nd base								3rd base
	T		C		A		G		
T	TTT	(Phe/F)	TCT	(Ser/S) Serine	TAT	(Tyr/Y) Tyrosine	TGT	(Cys/C) Cysteine	T
	TTC	Phenylalanine	TCC		TAC		TGC		C
	TTA		TCA		TAA <sup>[B]</sup>	Stop (Ochre)	TGA <sup>[B]</sup>	Stop (Opal)	A
	TTG		TCG		TAG <sup>[B]</sup>	Stop (Amber)	TGG	(Trp/W) Tryptophan	G
C	CTT	(Leu/L) Leucine	CCT	(Pro/P) Proline	CAT	(His/H) Histidine	CGT	(Arg/R) Arginine	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine	ACT	(Thr/T) Threonine	AAT	(Asn/N) Asparagine	AGT	(Ser/S) Serine	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	ATG <sup>[A]</sup>	(Met/M) Methionine	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine	GCT	(Ala/A) Alanine	GAT	(Asp/D) Aspartic acid	GGT	(Gly/G) Glycine	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GTG		GCG		GAG		GGG		G

- ☐ A. KRMLSGEP Feedback: -----
- ☐ B. KRGWGSi\* Feedback: -----
- ☐ C. SIKSGNM\* Feedback: -----
- ☐ D. KRGSCWAI Feedback: -----
- ☐ E. KRGADARK Feedback: -----

Answer Point Value: 10.0 points

Answer Key: E

Correct Feedback: -----

6

What output is produced by the following Python program? You may assume that the user enters the name as your first name and the number as \$300\$, corresponding to the number of this course. Show everything that is printed by the program to the terminal (you may include user input). Explain your reasoning.

```
number = input ("Enter a number: ")
# print(`Number is ', number)
name = raw_input ("Enter a name: ")
# print("Word is :", word)

print name, "You chose number ", number+10, "."
```

Answer Point Value: 10.0 points

Model Short Answer:

```
>>>number = input ("Enter a number: ")
```

Enter a number: 8

```
>>>name = raw_input ("Enter a name: ")
```

Enter a name: myName

```
>>>print name, "You chose number ", number+10, "."
```

myName You chose number 18 .

Feedback: -----

a) Complete the matrix aligning the two DNA sequences: CACTT and CAAT with the following rules: match score = +1, mismatch = -1, gap penalty = -1. Click on “Show/Hide Rich-Text Editor” link in your answer box, copy the table below to your answer box and then fill in the missing entries.

		C	A	C	T	T
	0	-1	-2	-3	-4	-5
C	-1	1	0	-1	-2	-3
A	-2	0	2	1	0	-1
A	-3	-1	1	2	1	
T	-4	-2	0	1		

b) What does the following Python code snapshot output for the matrix that you generated in part a)? Assume that you save the matrix from question 7 a) into the variable named table with N rows and M columns.

```
count = 0
for i in range(1, N+1):
    for j in range(1, M+1):
        if(table[i][j]==2):
            count +=1
print("Number of 2s is", count)
```

Answer Point Value: 10.0 points

Model Short Answer:

a)

		C	A	C	T	T
	0	-1	-2	-3	-4	-5
C	-1	1	0	-1	-2	-3
A	-2	0	2	1	0	-1
A	-3	-1	1	2	1	0
T	-4	-2	0	1	3	2

b) Number of 2s is 3

Feedback: -----



8

Explain when and why it might be more beneficial to use BLAST instead of Needleman-Wunsch type of algorithms?

Answer Point Value: 10.0 points

Model Short Answer: Needleman-Wunsch algorithm is a dynamic programming algorithm which gives an exact solution, but maybe impractical for real applications due to its computational complexity. BLAST is an approximation technique which performs faster than dynamic programming approaches, and produces results similar in accuracy.

Feedback: -----

9

After the program has run, what are the final values of the variables *a*, *b* and *c* which are used by the below python program? Justify your reasoning.

```
#!/usr/bin/env python
```

```
from Bio.Seq import Seq
```

```
dna_str = "atgcctcgattt"
```

```
sequence = Seq(dna_str)
```

```
a = Seq.transcribe(sequence)
```

```
b = Seq.back_transcribe(a)
```

```
c = Seq.translate(a)
```

```
print "a :", a
```

```
print "b :", b
```

```
print "c :", c
```

Answer Point Value: 10.0 points

Model Short Answer: a : augccucgauuu

b : atgcctcgattt

c : MPRF

Feedback: -----

One of the central tasks of bioinformatics is to compare genetic sequences using alignment algorithms such as the Needleman-Wunsch. In terms of tracking the spread of a virus across geographical distances, how is sequence-alignment a useful tool? Please use at least 150 words to justify your argument.

Answer Point Value: 10.0 points

Model Short Answer: Similar sequences indicate similar viruses and, often, similar remedies to handle outbreaks.

Feedback: -----