

Name: _____

ID : _____

ECS 129: Structural Bioinformatics
Midterm
February 18, 2026

Notes:

- 1) The midterm is open book, open notes.
- 2) You have 45 minutes, no more: I will strictly enforce this.
- 3) The midterm is divided into 3 parts and graded over 80 points (i.e. there is 5 extra credit points).
- 4) You can answer directly on these sheets (preferred), or on loose paper.
- 5) Please write your name at the top right of each page you turn in!
- 6) Please, check your work! **Show your work** when multiple steps are involved.

Part I (6 questions, each 10 points; total 60 points)

(These questions are multiple choices; in each case, find the **most plausible** answer)

- 1) Cytochrome P450 enzymes form a super-family of heme-containing oxygenases that are found in all kingdoms of life. Let us consider these three examples: (A) the human CYP46A1, (B) a human prostacyclin synthase, and (C) Xpla, a cytochrome P450 from rhodococcus (aerobic, gram-positive bacterium). (A), (B) and (C) are homologous proteins; what else can you say?
 - a) (A), (B) and (C) are orthologous
 - b) (A), (B) and (C) are paralogous
 - c) (A), (B) and (C) are analogous
 - d) (A) and (B) are paralogous, while (A) and (C) are orthologous
 - e) (A) and (B) are orthologous, while (A) and (C) are paralogous

- 2) The protein sequence alignment shown below has a total score of 20. Knowing that the score for an exact match is 5 and the score for a mismatch is -4, what is the score used for the gap penalty (constant, i.e. independent of length):
GCTGGAAG-GCA-T
GC-C-TAGAGCACT
 - a) -1
 - b) -2
 - c) -3
 - d) -4
 - e) Undefined (any value would give the same total score)

- 3) You are told that the DNA coding strand that encodes the peptide: Met-Trp-X-Trp-Met contains **60 % guanine (G)**. What is amino acid X?
 - a) Gly
 - b) Ala
 - c) Lys
 - d) Phe
 - e) Not enough information

Name: _____

ID : _____

- 4) We want to find the best alignment(s) between the DNA sequences TCATAGA and TCTAGC. The scoring scheme S is defined as follows: $S(i,j) = 3$ if $i = j$ (perfect match), $S(i,j) = 2$ if i and j are two different purines, $S(i,j)=1$ if i and j are two different pyrimidines, and $S(i,j)=0$ otherwise. There is a constant gap penalty of -1 (penalty for the first position counts; see table below). The score S_{best} and the number N of optimal alignments are (*show your final dynamic programming matrix and the best possible alignment (s) for full credit*):

	T	C	A	T	A	G	A
T	3	0	-1	2	-1	-1	-1
C	0						
T	2						
A	-1						
G	-1						
C	0						

- a) $S_{best} = 13, N = 2$
b) $S_{best} = 13, N = 1$
c) $S_{best} = 14, N = 1$
d) $S_{best} = 14, N = 2$
e) $S_{best} = 14; N = 3$
- 5) Dynamic programming, popular for sequence alignment, can also be used for spell checking. Assuming that a match is worth 10, a mismatch is worth 5, and a gap “costs” - 5, which of these five words is closest to the word “graffe” typed by a user? *Write the score of the optimal alignment next to each word (gaps at the start or at the end do not count).*
- a) gaff
b) graft
c) grail
d) giraffe
e) raft
- 6) A **single-stranded DNA** molecule contains: 30% adenine, and 50% pyrimidine. This strand is used as a template to synthesize its perfectly complementary strand. This complementary strand also contains 30% Adenine. What percentage of bases in the final double stranded DNA are guanines?
- a) 15%
b) 20%
c) 25%
d) 30%
e) Not enough information

Name: _____

ID : _____

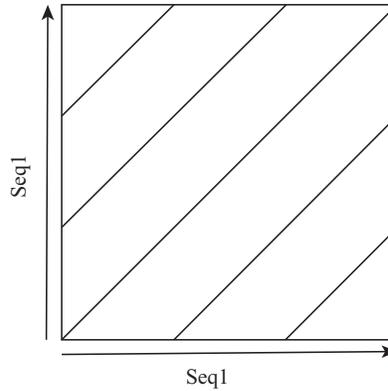
Part II (one question, 15 points)

Dotplots are useful, graphical tools for detecting information within a sequence. For example, consider the sequence:



ACDEFGHIACDEFGHIACDEFGHI

The dotplot below detects the presence of internal repeats as lines parallel to the main diagonal:



Draw schematically the dotplot for the following sequence:



ACDEFGHI IHGFEDCAACDEFGHI

Name: _____

ID : _____

Part III (total 10 points)

We want to find the best alignment(s) between the protein sequences WWYCTY and WCYTY. The scoring scheme S is defined as follows: $S(i,i) = P$ (perfect match), and $S(i,j) = M$ otherwise (mismatch). There is a constant gap penalty of G (gaps at the beginning count). The partial dynamic programming matrix is shown below:

	W	W	Y	C	T	Y
W						3
C						13
Y						26
T						28
Y	3	13				43

- a) Can you find P , M , and G ((note that we assume $P > 8$, $4 < M < 8$, and $-3 < G < 0$)). Explain your work
- b) Complete the dynamic programming matrix and find the best alignment(s)

Name: _____

ID : _____

Appendix:

Appendix A: Genetic Code

	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met/Start	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G