

Problem:	/	
Problem:	/	
Problem:	/	
Problem:	/	
Problem:	/	
Problem:	/	
Problem:	/	
Grade:	/	

Problem #1

Note: The page numbers below are for my studying later on and can be ignored during grading.

1.1) What is one of the key challenges facing biology today?

In the twentieth century, the link between DNA, RNA, and proteins was discovered. This led to the generation of large amounts of data on the sequences of these three molecules (i.e. DNA, RNA, and proteins). One of the key challenges facing biologists today is to organize, study, and draw conclusions from this data. This includes: identifying patterns within the sequences/data, documenting the structure and function of the molecules, and determining how all elements interact in the bodily functions. The field of studying this data and drawing the appropriate conclusions from it is known as bioinformatics. (See page 3)

1.2) What is the main role of DNA?

Information storage. It is the molecule that stores all genetic material and information that is needed to make and maintain all organisms. Moreover, DNA is what is passed on from one generation to the next as part of reproduction. (See page 4)

1.3) What is a nucleotide?

It is the building blocks of DNA and RNA. It is comprised of two primary parts. The first is the sugar-phosphate molecule and a nitrogenous base molecule. In DNA, there are four base molecules to differentiate nucleotides. They are thymine, guanine, cytosine, and adenosine. In RNA, thymine is replaced with uracil. When multiple nucleotides are linked via phosphodiester links, they form nucleotide sequences which are the foundations of DNA and RNA. What is more, in DNA, the sugar in the sugar-phosphate molecule is deoxyribose while in RNA it is ribose.

1.4) What DNA base is sometimes methylated? What effect does that base have on certain genes?

Cytosine bases can be methylated (i.e. have an added $-\text{CH}_3$ group). When cytosine is methylated, it can lead to a gene being rendered inactive. These methylated nucleotides are involved in the phenomenon known as *genomic imprinting*. In genetic imprinting, the change depends on whether the gene is maternally or paternally inherited. (See pages 6-7)

1.5) We know that errors occur during DNA replication. What is the error rate and why is it so low?

The error rate is approximately $1 \text{ in } 10^9$ (one in a billion). This low error rate is due to the way DNA polymerase builds the replicated strands, which includes mechanisms for checking and (if necessary) correcting erroneous base pairings. (See pages 8-9)

1.6) What is nucleic acid hybridization? Give three applications in which it is used.

DNA strands can pair with other nucleic acid sequences (e.g. DNA, RNA) of complementary sequences. These pairings result in double stranded DNA or DNA/RNA hybrids. This technique is used to identify identical or extremely similar segments of nucleic acids.

Nucleic acid hybridization is used in:

- a) DNA microarrays
- b) *In situ* hybridization to detect the activity of specific genes in cells.
- c) Fluorescence *in situ* hybridization (FISH) for visually locating genes within chromosomes.

(See page 9 and term definition on page 740)

1.7) What are the main working components of organisms that play the major role in almost all key processes of life?

Proteins (See page 10)

1.8) What part of DNA determines whether a gene is active (i.e. a protein is produced) or is inactive?

A gene is considered to have surrounding regions of noncoding DNA that act as **control regions**. These regions are involved in determining whether the gene is active (i.e. producing proteins or inactive). (See page 10)

1.9) What is the name of the RNA transcribed from a protein coding-gene?

messenger RNA – mRNA (See page 11)

1.10) What is meant by “a gene is expressed”?

Gene expression means that the gene in DNA has been transcribed into mRNA, and this RNA sequence is in turn directing protein synthesis in the cell. (See page 11)

1.11) Name one method of measuring the expression of many genes in a cell.

DNA or RNA expression microarrays (See page 11)

1.12) In which direction is the mRNA production during transcription? Give the direction with respect to the mRNA itself.

The mRNA strand is produced/built from the 5' end to the 3' end. However, the RNA polymerase transcribes (i.e. reads) the anticoding/template DNA strand in the 3' to 5' direction. (See page 12)

1.13) Where do overlapping genes most commonly occur? Do they also occur in humans?

Overlapping genes are where one or both strands in a part of DNA encode parts of different proteins.

Overlapping genes are most common in viruses due to the need to pack as much genetic information as possible into a small genome. Overlapping genes can occur in mammals with 774 pairs of overlapping genes identified in the human genome (see page 12).

1.14) What does it mean that the “genetic code is degenerate”?

It means that most amino acids can be specified by more than one codon (i.e. three base sequence). This means that while you can always determine the exact protein sequence from an mRNA molecule, the reverse process may not be possible. (See page 13)

1.15) Which amino acid is subsequently removed from newly synthesized proteins?

In most (but not all) cases, translation starts with an AUG (ATG) codon, which maps to the amino acid methionine. This methionine is often subsequently removed from the newly synthesized protein. (See page 13)

1.16) How many possible ways is there to translate a given DNA sequence?

Translation of bases to amino acids occurs in non-overlapping sets of three amino acids. As such, depending on which base is chosen as the start of the translation, there are three possible ways to translate any nucleotide sequence/strand. Since DNA is double stranded, the total number of ways to translate a given DNA sequence is six (three from each strand). (See page 13)

1.17) What is meant by an “open reading frame”?

An open reading frame (ORF) is a sequence segment that consists solely of a set of codons encoding for a protein. (See term definition on page 743) It consists of a reasonable length of uninterrupted protein code flanked by appropriate start and stop signals. (See page 13) It can *potentially* translate as a peptide chain although not all do. (See NCI Thesaurus definition)

1.18) What are RNAs mainly involved in? The three classes of RNA RNA: rRNA, mRNA, and tRNA. What is each one involved in?

RNA is primarily used in the transfer of information from DNA to use in the manufacture of proteins. mRNA is transcribed from DNA and contains the ordering/sequence information of the amino acids in the protein. rRNA and tRNA are involved in protein translation and protein synthesis. (See page 13)

1.19) What molecule is the physical link between the mRNA and the growing protein chain? Where does the molecule bind to?

Amino acids do not recognize the codons in mRNA directly. tRNA is the physical link between the mRNA and the growing protein chain. The tRNA binds to the ribosome. The tRNA contains the complementary anticodon for the codon in the mRNA and is what identifies which amino acid should be next in the polypeptide sequence. Enzymatic action in the ribosome joins amino acids together to form the peptide chain (with the assistance of rRNA). (See page 14)

1.20) What does the regulation of many processes interpret the information contained in a DNA sequence rely on?

The regulation of many processes that interpret the information contained in a DNA rely on the presence of short signal sequences in the DNA. These signal sequences are known as **regulatory elements**, and there are many different ones in a cell. (See page 15)

Problem #2

- A) Given the DNA nucleotide sequence below, write the complementary DNA sequence found on the other strand. Label the 5' and 3' ends of the sequence.

5'	C	A	C	T	G	T	C	A	T	G	G	C	T	T	T	T	G	A	T	C	A	A	A	A	A	A	A	3'
3'	G	T	G	A	C	A	G	T	A	C	C	G	A	A	A	A	C	T	A	G	T	T	T	T	T	T	T	5'

- B) Provide a link to an on-line tool that provides the complement of a DNA sequence.

http://www.bioinformatics.org/sms/rev_comp.html

Note: By default, this site provides the reverse complement and not the standard complement. To get the true complement, the user must select “complement” from the dropdown box.

- C) Suppose the DNA molecule from part A is transcribed and the lower (3' to 5') strand is used as the template strand. Provide the RNA sequence obtained from the transcription and label the 5' and 3' ends.

In DNA, the coding strand contains the information that will be in the mRNA after transcription. In this question, the template/anticoding/antisense strand is the lower strand. As such, the mRNA (top strand) is the complement of the complementary strand from part A (with minor modifications explained in part D).

5'	C	A	C	U	G	U	C	A	U	G	G	C	U	U	U	U	G	A	U	C	A	A	A	A	A	A	A	3'
3'	G	T	G	A	C	A	G	T	A	C	C	G	A	A	A	A	C	T	A	G	T	T	T	T	T	T	T	5'

- D) What is the difference between the RNA molecule you obtain and the given sequence in part A?

In RNA, the thymine (T) nucleotides/bases are replaced with uracil (U) (shown in green in part C). Moreover, the sugars in the sugar-phosphate backbones are ribose instead of deoxyribose. However, this latter difference is not captured when only providing a base sequence.

Problem #3

A) Given the RNA base sequence of bases below, write the corresponding amino acid sequence:

5'--- **CACUGUCACGGCUUUAGAUC**AAAAAA --- 3'

Codon	Amino Acid
CAC	His (H) Histidine
UGU	Cys (C) Cysteine
CAC	His (H) Histidine
GGC	Gly (G) Glycine
UUU	Phe (F) Phenylalanine
AGA	Arg (R) Arginine
UCA	Ser (S) Serine
AAA	Lys (K) Lysine
AAA	Lys (K) Lysine

B) Provide a link to an online RNA translation tool.

<http://www.attotron.com/cybertory/analysis/trans.htm>

C) Write all six possible reading frames from 5' to 3' of the following sequence:

5' --- **GCACTAGTCATGGCTTTTGAC** --- 3'

There are three possible reading frames from this strand, and three additional possible reading frames from the complementary strand. For this strand, the three frames are:

- GCA (First Full Codon)
 - 5' --- **GCA CTA GTC ATG GCT TTT GAC** --- 3' (DNA)
 - 5' --- **GCA CUA GUC AUG GCU UUU GAC** --- 3' (RNA)
- CAC (First Full Codon)
 - 5' --- **G CAC TAG TCA TGG CTT TTG AC** --- 3' (DNA)
 - 5' --- **G CAC UAG UCA UGG CUU UUG AC** --- 3' (RNA)
- ACT (or ACU for RNA) (First Full Codon)
 - 5' --- **GC ACT AGT CAT GGC TTT TGA C** --- 3' (DNA)
 - 5' --- **GC ACU AGU CAU GGC UUU UGA C** --- 3' (RNA)

The complimentary strand for this sequence is (from 5' to 3') is:

5' --- **GTCAAAAGCCATGACTAGTGC** --- 3'

That makes the three complementary frames:

- GTC (or GUC for RNA) (First Full Codon)
 - 5' --- **GTC AAA AGC CAT GAC TAG TGC** --- 3' (DNA)
 - 5' --- **GUC AAA AGC CAU GAC UAG UGC** --- 3' (RNA)

- TCA (or UCA for RNA) (First Full Codon)
 - 5' --- **G** TCA **AAA** GCC **ATG** ACT **AGT** GC --- 3' (DNA)
 - 5' --- **G** UCA **AAA** GCC **AUG** ACU **AGU** GC --- 3' (RNA)
- CAA (First Full Codon)
 - 5' --- **GT** CAA **AAG** CCA **TGA** CTA **GTG** C --- 3' (DNA)
 - 5' --- **GU** CAA **AAG** CCA **UGA** CUA **GUG** C --- 3' (RNA)

Problem #4

5'	C	C	A	T	G	G	A	G	T	C	G	T	3'	DNA double helix
3'	G	G	T	A	C	C	T	C	A	G	C	A	5'	
5'	C	C	A	U	G	G	A	G	U	C	G	U	3'	mRNA transcribed
3'	G	G	U	A	C	C	U	C	A	G	C	A	5'	Appropriate tRNA Anticodon
Amino	Pro (P) Proline			Trp (W) Tryptophan			Ser (S) Serine		Arg (R) Arginine			Carboxyl	Amino acids incorporated into protein	

Proteins are synthesized in the amino to carboxyl direction.

Problem #5

Below is a segment of DNA. Assume the top strand is the template strand used by the RNA polymerase.

5' – GCTTCCCAA – 3'
3' – CGAAGGGTT – 5'

A-B) Draw the RNA transcript and label its 5' and 3' ends.

5' --- UUGGGAAGC --- 3'

In this question, the template/anticoding/antisense strand is the upper/top strand. As such, the mRNA is the essentially the complement of the top strand (or in other words the bottom strand) with the thymine (T) bases changed to uracil (U).

C-D) The amino acid sequence for this sequence is:

Amino (NH ₃)	
UUG	Leu (L) Leucine
GGA	Gly (G) Glycine
AGC	Ser (S) Serine
Carboxyl (COOH)	

Note: This answer assumes the first reading frame is at the beginning of the sequence.

E) Repeat this problem assuming the bottom strand is the template strand.

If the bottom strand is the template strand, then the top strand defines the mRNA with the exception that thymine bases are changed to uracil. Hence the new mRNA is:

5' – GCUUCCCAA – 3'

The resulting amino acid sequence is:

Amino (NH ₃)	
GCU	Ala (A) Alanine
UCC	Ser (S) Serine
CAA	Gln (Q) Glutamine
Carboxyl (COOH)	

Problem #6

Answer the following true or false questions:

- A) It is possible for an RNA polymerase molecule and a ribosome to be attached to a eukaryotic mRNA simultaneously.**

False. RNA synthesis (i.e. transcription) is catalyzed by the enzyme RNA polymerase. In a eukaryote, all DNA is stored in the nucleus. Hence, when RNA polymerase is attached, the mRNA is inside the nucleus in a eukaryote. Moreover, before mRNA can leave the nucleus, the mRNA must be spliced which is after transcription.

In contrast, ribosomes attach to the mRNA during protein synthesis (i.e. translation). This stage occurs outside the nucleus (i.e. in the cytoplasm) where ribosomes reside. Hence, given that the two molecules are involved in different mRNA stages and one is in the nucleus while the other is in the cytoplasm, they do not both attach simultaneously.

- B) At no time during protein synthesis does an amino acid make direct contact with the mRNA being translated.**

True. In the translation of mRNA, the tRNA contains the anticodon that couples with the codon on the mRNA inside the ribosome. The amino acid is on the **3'** prime end of the tRNA. Hence, the tRNA is the intermediary between the amino acid and the mRNA, which eliminates the need for them to ever make direct contact.

- C) Because the two strands of DNA are complementary, the mRNA of a given gene can be synthesized using either strand as a template.**

False. While the DNA strands are complementary, the orientation of the strands is in opposite directions with one strand being from 5' to 3' and the other being from 3' to 5'. What is more, the codons are not complimentary. For instance, for complementary strands to encode the same gene codon, UGU would need to encode the same amino acid as ACA. However, UGU encodes Cysteine (Cys) while ACA encodes Threonine (Thr).

Problem #7

A) In how many cases is the genetic code would you fail to know the amino acid specified by the codon if you knew only the first two nucleotides of the codon?

There are 8 (out of 16) pairs of nucleotides that do not lead to single amino acid. They are:

- 1) UUx – This can lead to Phe (UUU, UUC) or Leu (UUA, UUG).
- 2) UAx – The only amino acid possible from this codon is Tyr (UAU, UAC), but it could also be the stop codon (UAA, UAG).
- 3) UGx – Possible amino acids are Cys (UGU, UGC) and Trp (UGG). It can also be the stop codon (UGA).
- 4) CAx – Possible amino acids are His (CAU, CAC) and Gln (CAA, CAG)
- 5) AUx – Possible amino acids are Ile (AUU, AUC, AUA) and Met (AUG).
- 6) AAx – Possible Amino acids are Asn (AAU, AAC) and Lys (AAA, AAG)
- 7) AGx – Possible amino acids are Serine (AGU, AGC) and Arg (AGA, AGG).
- 8) GAx – Possible amino acids are Asp (GAU, GAC) and Glu (GAA, GAG)

Note: This answer included the case of UAx which has Tyr and the stop codon, which is not really an amino acid. If this case is excluded, then there would be only 7 pairs.

B) In how many cases would you fail to know the first two nucleotides of the code if you knew which acid is specified by it?

There are three cases. They are:

- 1) Leucine (Leu) – UUx and CUx
- 2) Arginine (Arg) – AGx and CGx
- 3) Serine (Ser) – AGx and UCx

Problem #8

A) What does CDS stand for?

CDS stands for “coding sequence.” It is a region of nucleotides (i.e. bases) that correspond with a sequence of amino acids in a protein. The CDS includes both the start and stop codons. The CDS is accompanied with an amino acid translation.

B) In the first of three CDS features, why does the “/translation” subfield start with the specific sequence of amino acids “SSIYN”?

The DNA sequence is shown in the origin section of the page. The first ten 20 characters in the sequence are:

GA **TCC** TCC **ATA** TAC **AAC** GGT

The region of the first protein is 1 to 206 with the first codon starting at the third base (i.e. T). Using the sequence shown above the first five amino acids are:

DNA Sequence	RNA Sequence	Amino Acid
TCC	UCC	Serine (S)
TCC	UCC	Serine (S)
ATA	AUA	Isoleucine (I)
TAC	UAC	Tyrosine (Y)
AAC	AAC	Asparagine (N)