Solution key- 7.016 Problem Set 2- 2018

Question 1 (4pts)

Melatonin is a hormone that regulates sleep and is used to treat insomnia (inability to fall asleep). It acts by binding to the melatonin receptors (MT_R), which exists in different isoforms: MT_{R1}, encoded by *MT-1 gene* and MT_{R2}, encoded by *MT-2 gene.* The following is the sequence corresponding to base pairs (bp) 90-140 of *MT-1* and *MT-2* genes.

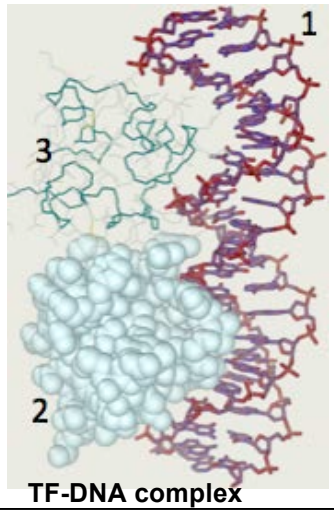
```
                90                                                  140
MT-1: 5'------GATATGCCCCCCCCGGCGCGCGATATGCCCCCCCCGGCGCGCGTGCGTGA-------3'
      3'------CTATACGGGGGGGGCCGCGCGCTATACGGGGGGGGCCGCGCGCACGCACT-------5'

                90                                                  140
MT-2: 5'------GATATGATATATATATATATAGATATGAAAAATTTTTATATAGTGCGTGA-------3'
      3'------CTATACTATATATATATATATCTATACTTTTTAAAAAATTATCACGCACT-------5'
```

a) Which of the above sequences will denature at a lower temperature and **why**?

Both MT-1 and MT- 2 sequences are of the same length i.e. they have same number of base pairs. However, Sequence MT-2 is A/T rich and therefore has fewer hydrogen bonds compared to MT-1. So MT-2 sequence needs less energy to disrupt the hydrogen bonds and denatures at a lower temperature compared to MT-1 sequence. (0.5 with 0.25 for explanation)

b) In the sequences above, which end receives the incoming nucleotide: the **5' or** the **3' end**? *3' end (0.25)*

c) Melatonin binds to MT_R and activates it. The activated MTR triggers a signaling pathway that activates two specific transcription factors (TF) that play a critical role in regulating sleep.



TF-DNA complex

i. On the schematic of the TF-DNA complex to the left, which macromolecule(s) are the likely transcription factors and **why**: **1/ 2/ 3**? *(0.5 with 0.25 for explanation)*

Transcription factors are proteins that regulate transcription by binding to the regulatory sequences (promoters or enhancers) of the genes. Molecule 1 is a DNA double helix unlike Molecules 2 and 3, which are proteins and therefore the likely TFs.

ii. You denature the TF-DNA complex. Which of these molecules may likely renature and **why**: **1/ 2/ 3**? *(0.5, 0.25 for explanation)*

Proteins have complex 3D- conformation that is stabilized by covalent and noncovalent interactions. They often fold with the help of chaperone proteins (You will learn about this in cell Biology lectures). So they are less likely to renature. In comparison, DNA is more likely to renature since it has only one possible double helical conformation, which is stabilized by hydrogen bonds between complementary base pairs and hydrophobic interactions between the aromatic rings of adjacent bases.

iii. Name a **region on a gene** to which the TFs can bind to regulate gene expression.

They can bind to the promoter sequence, which is part of a gene. They can also bind to the enhancer sequence, which is located far away from the gene but can influence its expression. Transcription factors may activate or inhibit gene expressions. (0.5)

iv. Which of the following amino acids in the TFs can potentially form an ionic bond with the DNA and **why**: **Serine/ Alanine/ Glutamic acid/ Lysine**? ***Note:*** *A chart of amino acids was given on the last page of Problem set 1.*

Lysine, since it has a positively charged side-chain that can form an ionic bond with the negatively charged phosphates of the sugar phosphate backbones of the DNA duplex. (0.25)

Question 1 continued

d) You identify three individuals with the following mutations in the *MT-1* gene. **Explain** how melatonin prescription would affect Individuals 1-3.

I. **Individual 1:** The Histone proteins bound to the *MT-1* gene in Individual 1 are constitutively (always) acetylated: *(0.5 with 0.25 for explanation)*

Acetylation neutralizes the positive charges on Lys, Arg and His amino acids of histones thus decreasing their interactions with the DNA sequence. So the DNA can unwind and the TF can bind to the exposed regulatory sequences such as the promoters and enhancers of MT-1 thus promoting its expression. Since the MT-1 gene is constitutively expressed, there will be high levels of MT_{R1} protein, which can bind to melatonin to promote sleepiness.

II. **Individual 2:** Activating transcription factors (TF) for the *MT-1* gene expression are absent.

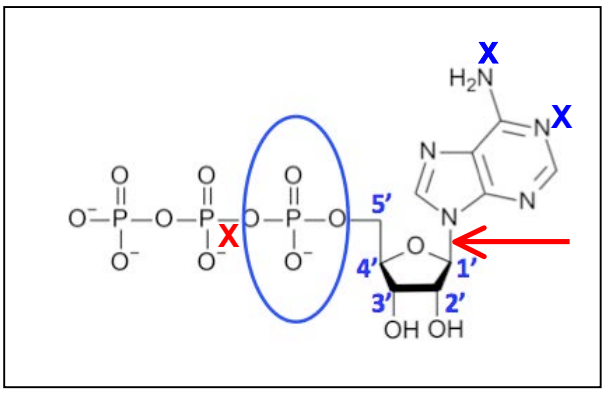
If the MT-1 gene specific transcription factor is absent, there is no MT-1 gene expression. So prescribed melatonin will have no effect on this person in the absence of the functional MT_{R1} receptor. This person will have insomnia. (0.5 with 0.25 for explanation)

III. **Individual 3:** Promoter region of *MT-1* gene does not have methylated bases:

DNA methylation generally represses transcription of a given gene. If the promoter sequence of MT-1 gene is unmethylated, the MT-1 gene will be hyper-transcribed. This person, like Individual 1, will have a heightened response to melatonin and will be very sleepy. (0.5 with 0.25 for explanation)

Question 2 (3pts)

Nucleic acids are made of five bases –A, T, G, C and U. The following is the chemical structure of one of the five bases. For parts **(a) – (d)**, on the schematic…



a) Number the carbon atoms of the sugar. *(0.25)*

b) Add three phosphates to the 5'C of the nucleoside (base + sugar) so that you now show it as a nucleotide (base + sugar + triphosphate). *Please note that the addition of an incoming base to the 3'OH end of a growing nucleic acid is an energy requiring reaction. This energy is derived by the hydrolysis of the bond (shown as X), which results in the formation of pyrophosphate (PPi) and AMP. It is the AMP, which gets added to the 3' OH end of the nucleic acid chain. (0.25)*

c) Put an **" X"** next to the atoms/ groups of the nucleotide that participate in hydrogen bonding with the complementary base. *(0.25)*

d) Circle the group that would participate in the formation of a phosphodiester bond if this nucleotide were **added to** the growing end of a nucleic acid chain. *(0.25)*

e) Identify the base on the schematic as **purine or pyrimidine**: **Purine** *(0.25)*

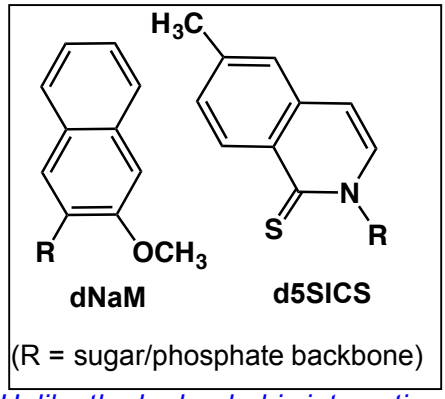
f) Name the **covalent bond** shown by an arrow in the schematic above: **Glycosidic bonds** *(0.25)*

g) Which nucleic acid is the nucleotide in the schematic above a building block for and **why**: DNA/ RNA?

This ATP can be a part of RNA since it has a ribose sugar with a 2'-OH group. The nucleotides that make the DNA have deoxyribose sugar with an "H" at the 2'C position. (0.5 with 0.25 for explanation)

Question 2 continued

Organisms are defined by the information encoded in their genomes. The four bases in DNA (A, T, G, C) can be combined to form 64 different codons or "words". Recently ***Malyshev et al (Nature, 2014)*** have expanded the genetic code in bacteria to include two synthetic nucleotides, which are designed to form a new complementary base pair d5SICS–dNaM as shown below.



h) How many possible tri-nucleotide codons exist using the bases **A, T, G, C, d5SICS** and **dNaM**?
6 X 6 X 6 = 216 codons (0.25)

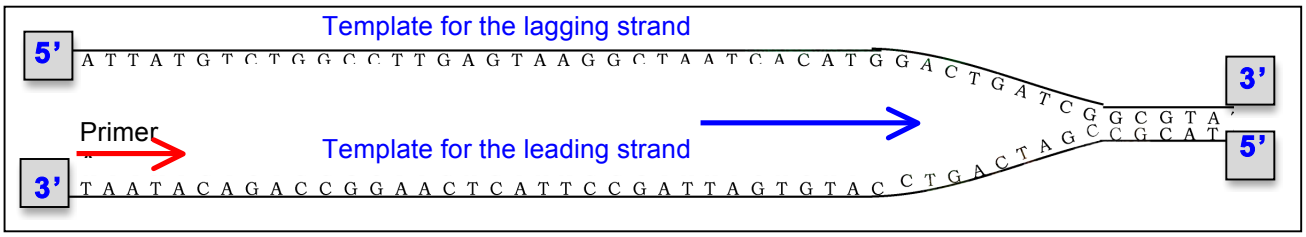
i) Circle the most likely interaction between the dNaM-d5SICS base pair: **covalent bond/ ionic bond/ hydrophobic interaction/ hydrogen bonding**. *(0.25)*

j) Explain how the interaction that you circled in part (i) above differs from the type of interactions between A/T and G/C base pairs

Unlike the hydrophobic interactions between the aromatic rings of the dNaM and d5SICS, the "A" forms two hydrogen bonds with the "T" and "C" forms three hydrogen bonds with the "G" of the complementary strands of a DNA duplex. (0.5)

Question 3 (3pts)

The *MT-1* gene is located on chromosome 11. The diagram below depicts a replication fork at one origin of replication site (ori) on Chromosome 11 in a cell undergoing DNA replication. **Note:** *The primer is shown by an arrow i.e. 5'→3'.*



a) On the schematic of the replicating DNA above, label the 5' and 3' ends of the template DNA strands by filling in the shaded boxes and show the direction of movement of the replication fork by an arrow. *(0.5pt with 0.25 for each)*

b) Label the parental strands as the templates for **leading (continuous)** or **lagging (discontinuous)** strands synthesis. *(0.25)*

c) Give the sequence of the five base long primer that is shown as an arrow in the diagram above: **5'AUUAU3'** ***(Note: primers with a cell are RNA primers)*** *(0.5)*

d) Give the sequence of the **next four bases** that would be added to the growing end of the primer: **5'GTCT3'** *(0.25)*

Question 3 continued

e) You grow the normal, healthy *MT-1 gene* expressing cells in the presence of nutrients and the following compounds. ***Note:*** *For this question you may assume that these compounds are able to get into the cells.*

- **Plate 1** contains TA-65, a compound that prevents the shortening of chromosomes.
- **Plate 2** contains doxorubicin, a compound that promotes DNA supercoiling.
- **Plate 3** contains an RNAse inhibitor.

Explain how the compounds would affect DNA replication in the following plates.

i. **Plate 1:** *TA-65 will likely activate the telomerase enzyme so that it is able to maintain the telomeres at the ends of the chromosomes thereby preventing their shortening. So there will be no loss of genetic information following each round of replication. Hence the cells will continue replicating their DNA and hence continue to divide. (0.5 for explanation)*

ii. **Plate 2:** *Doxorubicin will likely inhibit the topoisomerase enzyme, which removes the supercoils thereby relieving stress on the DNA duplex. If the DNA remains supercoiled it will NOT be able to complete replication. (0.5 for explanation)*

iii. **Plate 3:** *In the presence of the RNAse inhibitor, the RNA primers will not be not be removed and replaced by DNA sequence. Also, the ligase will not join the Okazaki fragments to make a continuous DNA strand. So the replication will be abnormal and incomplete. (0.5 for explanation)*

Question 4 (2pts)

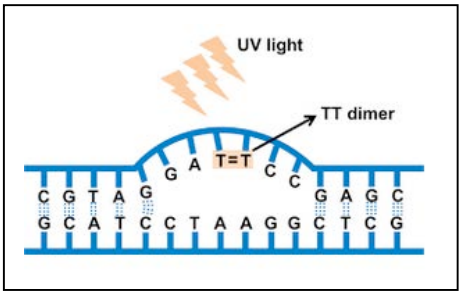
a) Prokaryotes such as bacteria only have one origin of replication (ori site) in the bacterial genome. If DNA polymerase enzyme adds nucleotides at the rate of 10,000 base pairs/minute ***in one direction*** and the bacterial cells replicate every 20 minutes, what is the size (in terms of base pairs) of the bacterial genome? **Show your work.** ***Note:*** *In your calculation be sure to account for the fact that replication is bidirectional.*

Replication occurs at the rate of 10,000bp/ minute in one direction. Number of bases replicating bidirectionally is 20,000bp/ min. In 20 minutes, number of bases replicating will be 20,000 x 20 = 4 x 10^5 bp. So the size of the genome is 4 x 10^5 bp. (0.25)

b) Unlike the prokaryotes, eukaryotic genomes have multiple ori sites. Using the following information, calculate the total number of ori sites on a single human chromosome that is 252 million base pairs long: *DNA polymerase adds nucleotides at the rate of 3,000 base pairs/minute in one direction. In your calculation be sure to account for the fact that replication is bidirectional. The S phase (DNA replication phase) of cell cycle is 5 hours long.*

At 3,000 bp/ min, each origin grows bi-directionally at 6,000 bp/ min. Replication is completed in 5 hours = 300 minutes. So the total number of replicated bp from one origin is 300 6000 =1.8 million. The total number of bp in chromosomes is 252 million base pairs. So the total number of origins on this chromosome is 252 million/ 1.8 million = 140. (0.25)*

c) UV radiation from the sun or tanning salons can result in the formation of thymidine dimers (T-T) in the DNA of skin and hair follicle cells. These T-T dimers, if left unrepaired can result in rapid aging of skin, freckles and even melanoma (a form of skin cancer). Which process will repair the T-T dimers: **DNA Proofreading/ mismatch repair/ Nucleotide excision repair**? **Explain** why you selected this process over the others. *The T-T dimerization can happen at any time following exposure to radiation. DNA proofreading happens only during replication. A mismatch can be repaired by mismatch repair mechanism, which happens immediately AFTER replication on the hypo-methylated strand. But there is no mismatched base pair here. SO the nucleotide excision repair will repair the T-T dimer. (0.5 with 0.25 for explanation)*



Question 4 continued

d) Does the formation of T-T dimers impact the…

i. Fidelity of Replication? If so, how? **Note:** *Your explanations may vary.*

The T-T dimer can generate kinks in the DNA strand changing its shape. So the DNA polymerase may get stalled or read the bases incorrectly thus impairing replication. Multiple explanations. (0.5 for explanation)

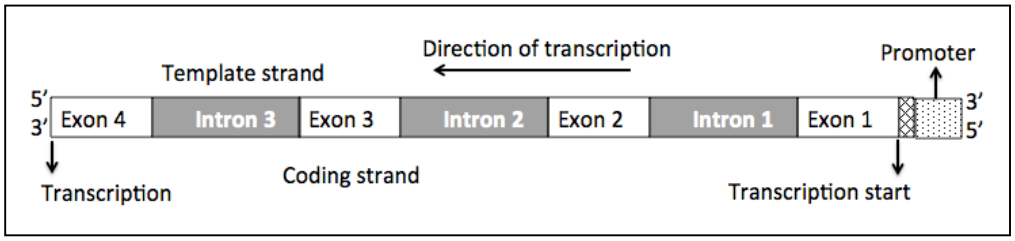
ii. Process of Transcription? If so, how? **Note:** *Your explanation may vary.*

Many explanations. It generates kinks in the DNA strand. This may change shape, impair normal base pairing, prevent RNA polymerase or transcription factors from binding to the promoter sequence to read the template strand to transcribe RNA. (0.5 for explanation)

Question 5 (3pts)

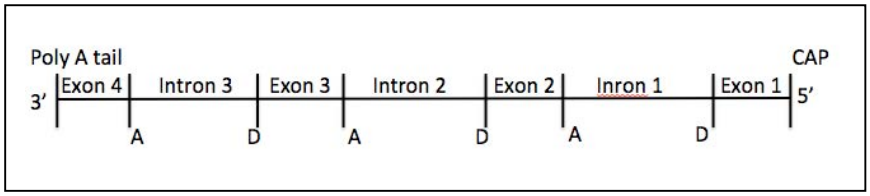
The *MT-1* gene is comprised of four exons (Exon 1, 2, 3, and 4) that are each 1.5 kilo base (kb) long and three introns (Introns 1, 2, and 3) that are each 3kb long.

a) Based on the information provided, draw the *MT-1* gene in the box below. Label its 5' and 3' ends, draw and label the promoter, the transcription start and stop sites, exons 1-4 (as blank boxes) and introns 1-3 (as shaded boxes). ***Note:*** *The arrow in the schematic shows the direction of transcription.*



(1pt)

b) Draw the **nascent (pre-spliced/ newly synthesized) mRNA** transcribed from the *MT-1* gene. Label its 5' and 3' ends, exons 1-4, introns 1-3 and any modifications to the mRNA. Label each splice donor site with a "D" and each splice acceptor site with an "A". *(0.5 with 0.25 for exons & 0.25 for ALL labels)*



c) Assuming all introns are spliced out, give **all** the possible **mature (spliced) mRNA transcripts** resulting from the *MT-1* gene.

3'AAAA....AA- Exon 4-Exon 3-Exon2-Exon 1-5'-CAP -> 6kb

3'AAAA....AA- Exon 4-Exon 3-Exon 1-5'-CAP -> 4.5kb

3'AAAA....AA- Exon 4-Exon 2-Exon 1-5'-CAP -> 4.5kb

3'AAAA....AA- Exon 4-Exon 1-5'-CAP -> 3kb

I. Label the 5' and 3' ends of each mature transcript. *(0.5)*

II. Specify the exons that make the transcripts and give the size (in terms of kb) for each mature transcript. *(0.5)*

III. List any modification(s) to the 5' and the 3' ends of the mature mRNA transcripts and **briefly describe** their significance. *(0.5 with 0.25 for each)*

The addition of 7methylguanosine CAP at the 5' end and Poly A tail at the 3' end provide stability to the mature mRNA and allow its export from the nucleus into the cytoplasm where it is translated.

Question 6 *(3pts)*

The following is the DNA sequence for the transcription initiation region of the *MT-1* gene. ***Note:*** *Part of the* ***promoter region*** *is boxed and the dashes represent bases that are not shown. Transcription* ***begins at and includes*** *the bold and underlined* ***T/A*** *base pair.*

```
5′--TGGACTGCTATGATAGCAGTTCTGCTGAGACGATGGCCATACGGCCATGGTTCATAAAGT----3′  TOP
3′--ACCTGACGATACTATCGTCAAGACGACTCTGCTACCGGTATGCCGGTACCAAGTATTTCA----5′BOTTOM
```

a) Identify the template strand for transcription: **Top / Bottom**? Top strand *(0.25)*

b) Write the **first 9 nucleotides** of the newly transcribed *MT-1* mRNA: **5'*ACCAUGGCC*3'** *(0.25)*

c) Using the codon chart on the last page of this problem set, write the **first 2 amino acids** of the newly synthesized MT_{R1} receptor: **N-*Met-Ala*-C** *(0.25)*

d) Write the sequence of the **anticodon corresponding to the 2nd amino acid** of newly translated MT_{R1} protein: **5'-*GGC*-3'** *(0.25)*

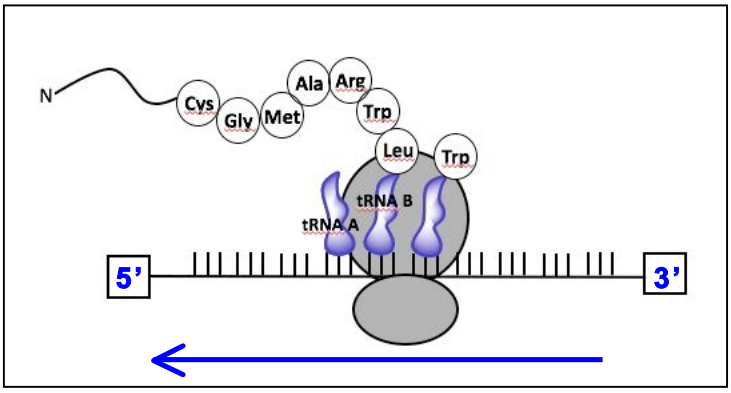
e) The last (C-terminal) 5 amino acids (296-300), which are **critical** for the normal function of MT_{R1} receptor are: **N –Val^{296}-Ser^{297}-Asn^{298}-Ser^{299}-Met^{300}- C**

The DNA sequence encoding the C- terminus of the wild- type and mutant forms of the MT_{R1} protein is included within the sequence below. The **point mutations** in mutants 1 and 2 are bold & shaded and the stop codon (5'TAG3') is underlined. ***Note:*** *The codon table is provided on the last page of the problem set.*

```
Wild- type:  5′-TCGTATCGAATTCCATGTAGC-3′
             3′-AGCATAGCTTAAGGTACATCG-5′

Mutant 1:    5′-TCGTATAGAATTCCATGTAGC-3′
             3′-AGCATATCTTAAGGTACATCG-5′

Mutant 2:    5′-TCGTATCGAACTCCATGTAGC-3′
             3′-AGCATAGCTTGAGGTACATCG-5′
```

Compared to the wild- type, which mutant allele will most likely not express the *MT-1* gene and **why**? *The mutation in Mutant 2 is a silent mutation, so the MT_{R1} receptor will still have the same primary sequence and will bind to melanin to regulate sleep. Mutant 1 has a nonsense mutation that produces a truncated, nonfunctional MT_{R1}, which cannot bind to melanin. A person with this mutation has insomnia and does not benefit from melanin prescription.* *(1 with 0.5 for explanation)*

f) Below is a drawing of a ribosome actively translating *MT-1* mRNA. The horizontal line represents the mRNA, each vertical line represents a base (A, U, G, C) and two tRNA molecules (tRNA A and tRNA B) are labeled and drawn in purple on the posted version of the problem set.

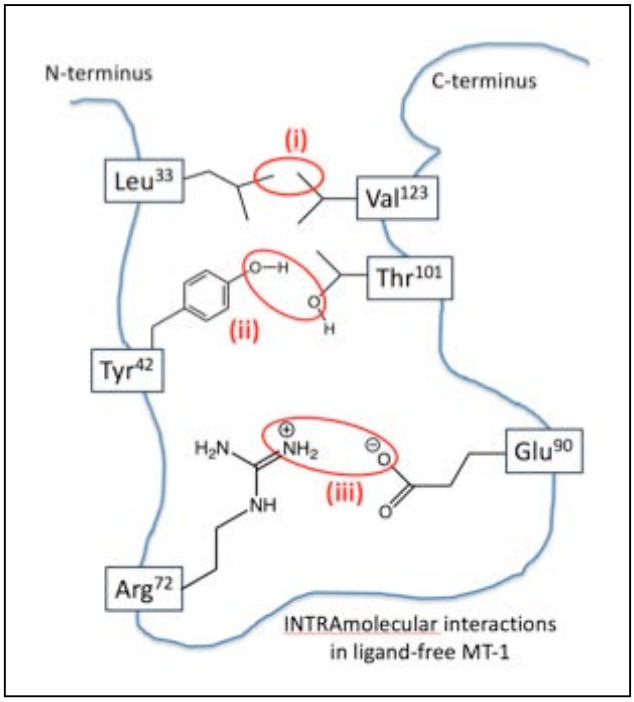


i. Draw an arrow to indicate the movement of mRNA relative to the ribosome during translation.*(0.5)*

ii. Label the 5' and the 3' ends of the mRNA. *(0.25)*

iii. Name the amino acid that was originally bound to **tRNA A**: ***Trp*** *(0.25)*

Question 7 (2pts)
The diagram below shows the intra-molecular interaction between the side-chains of amino acids at positions (i)-(iii) in the MT_{R1} protein. Each of these interactions is critical for the correct 3D folding of MT_{R1} protein, which allows it to bind to melatonin.



a) What is the strongest non-covalent interaction at positions (i)–(iii): **Ionic interaction/ hydrophobic interaction/ hydrogen bonding?**

i. ***Hydrophobic interaction*** *(0.25)*
ii. ***Hydrogen bonding*** *(0.5)*
iii. ***Ionic interaction*** *(0.25)*

b) You analyze both alleles of the *MT-1* gene in four individuals (1-4). Each of these individuals has a specific mutation in one or both alleles of the *MT-1* gene as specified in the table below. ***Note:*** *Please refer to the information at the beginning of Question 1 of this problem set.*

Individuals	Mutations
1	5'CUG3' (Leu) -> 5'CUA3' at (i)
2	5'UAU3' (Tyr) -> 5'UAA3' at (ii)
3	5'CGA3' (Arg) -> 5'CAA3' at (iii)
4	5'UAU3' (Tyr) -> 5'UAG3' at (ii)

Which of the above individuals are most likely to have insomnia and **why**?
Both 5'CUG3' and 5'CUA3' codons corresponding to the amino acids at position (i) of MT_{R1} in individual 1 code for amino acid leucine making this a ***silent mutation****. So this individual is NOT likely to have insomnia. In* ***Individuals 2 and 4****, mutation at position (ii) results in a premature stop codon. Hence a truncated, non-functional MT_{R1} is made which will not respond to melatonin. So this individual is likely to have insomnia.*

In ***Individual 3,*** *the mutation is a missense mutation that replaces Arg (a positively charged amino acid) with Gln (an uncharged hydrophilic amino acid) at position (iii). The changed amino acid (Gln) cannot undergo ionic interaction with Glu at position (iii). This may disrupt the 3D conformation of MT_{R1} and make it non-functional and unresponsive to melatonin. So it is very likely that this individual will have insomnia. (1 with 0.25 for each)*

7.016 Introductory Biology
Fall 2018