

GENETICS

WITH EMPHASIS ON MOLECULAR STRUCTURE

BY

DEREK WAN

COMPUTER SCIENCE 61A TUTOR, FALL 2017

COMPUTER SCIENCE 61A INSTRUCTOR (TA), SPRING 2018 TO PRESENT

Preface

I compiled these notes while I was taking MCB 102 with Professor Ingolia mostly as a light supplement to the lectures. Prof. Ingolia himself writes practice questions for each lecture that ask students to think about what they've learned in a novel context. Although those questions don't require as much tedious memorization as the previous two modules of the course, memorization is still necessary. To fully capitalize on his materials, I suggest working through these very low-level questions first to test your basic understanding of each lecture (there is *a lot* of material in each lecture). It is very easy to complete Prof. Ingolia's worksheets by referencing lecture material—it's significantly harder (but more rewarding) to do the worksheets as mini-exams (closed book). These notes are intended to be a sort of stepping stone between lecture and Prof. Ingolia's problem sets.

Since I compiled these notes *while* I was taking the class, there may be certain parts of the text that include question marks/blanks/colloquial language that I never edited out. Feel free to reach out to me at [derek.wan11\[at\]berkeley.edu](mailto:derek.wan11[at]berkeley.edu) if you find any blatant mistakes or would like to offer any suggestions.

You can also check out my General Chemistry book and my Discussion slides for CS61A at my personal website: derekwan1.github.io (if that doesn't work I have probably switched to ddderek.com).

Post-semester update: I've been getting a lot of emails from people inquiring for either a) past exams or b) my exam stats (aka "how do I trust that you're qualified to write this?"). I'm not comfortable sharing exams that are not my property, and I don't want to run into any trouble with the professors for releasing exams without their consent—sorry about that. I really do understand the struggle to get my hands on past exams, but you'll have to go through the usual ways for that. As for my exam stats, I did 1.1 standard deviations above the mean on the first midterm, 0.99 standard deviation above the mean for the second midterm, and 1.8 or 1.9 standard deviations above the mean for the last exam. I believe I was the top score, or very close to the top ;)

Contents

Preface	2
1 Molecular Biology	5
1.1 Questions	5
1.2 Key	5
2 Nucleic Acid Chemistry	7
2.1 Questions	7
2.2 Key	7
3 DNA Topology and Chromatin	9
3.1 Questions	9
3.2 Key	9
4 DNA Replication	11
4.1 Questions	11
4.2 Key	12
5 DNA Repair and Recombination	14
5.1 Questions	14
5.2 Key	15
6 Episomes and Cloning	17
6.1 Questions	17
6.2 Key	17
7 Bacterial Transcription and Regulation	19
7.1 Questions	19
7.2 Key	20
8 Eukaryotic Transcription and Regulation	22
8.1 Questions	22
8.2 Key	23
9 RNA Processing	26
9.1 Questions	26
9.2 Key	27
10 Telomeres and Introduction to Translation	29
10.1 Questions	29
10.2 Key	30

11 More Translation	32
11.1 Questions	32
11.2 Key	32
12 Translation Regulation and micro-RNAs	34
12.1 Questions	34
12.2 Key	34
13 Protein Degradation and Genetic Engineering	36
13.1 Questions	36
13.2 Key	36

Chapter 1

Molecular Biology

1.1 Questions

1. What is the relationship between uracil, thymine, and cytosine?
2. Name the nucleotide/nucleoside of each nitrogenous base.
3. DNA/RNA most favorable positions?
4. Width of DNA helix and \AA per nucleotide/turn?
5. Dominant form of DNA/RNA?
6. Describe single-stranded RNA qualitatively.
7. A sequence reads TGCGAT 5' \rightarrow 3'. What is the sequence (also 5' \rightarrow 3') that would form a hairpin structure? Why is it more common in RNA than DNA?
8. Describe other secondary structures of single-stranded nucleic acids.
9. What is the wobble base?
10. What are the primary determinants for the 3D structure of RNA?
11. Which RNA structures melt first?
12. Rank base pairs and types of nucleic acid in decreasing order of melting temperature.
13. What are the main reasons for the higher melting temperature of GC?
14. Which type of gel do you use for different sizes of DNA?

1.2 Key

1. Thymine is 5-methyl-uracil. Uracil is deaminated cytosine.
2. (a) Adenine: Deoxyadenylate (nucleotide), deoxyadenosine (nucleoside)
(b) Guanine: Deoxyguanylate (nucleotide), deoxyguanosine (nucleoside)
(c) Thymine: Deoxythymidylate (nucleotide), deoxythymidine (nucleoside)
(d) Deoxycytidylate (nucleotide), deoxycytidine (nucleoside)
3. DNA: 2' endo. RNA: 3' endo. Both prefer anti conformations as opposed to syn (anti means the N-glycosidic bond is oriented in a way such that the base is not directly above the ribose sugar).

4. 3.4 Å per base pair, 36 Å per helical turn, 20 Å width
5. A-form for dsRNA, B form for DNA. The A-form is shorter and wider, due to the 3'-endo conformation.
6. (a) Tends to coil
(b) Flexible and dynamic
(c) Hydrophobic faces together, hydrophilic/charged backbone outward
7. ATCGCA, separated from the original strand by a few base pairs.
This is more common in RNA because ssRNA is more common than ssDNA.
8. (a) Bulge: extra bases on one side that doesn't base pair with anything on the opposite strand
(b) Loops: non-pairing regions
(c) Turns within hairpins, unpaired bases in a loop.
9. Guanine and uracil—angles and distances aren't normal.
10. H-bonding of bases, ribose, and phosphate
11. Tertiary structures melt first, then secondary
12. (a) GC, AU > GU > mismatches
(b) GC > AT
(c) RNA-RNA > RNA-DNA > DNA-DNA
13. Stacking geometry and structure.
14. Agarose gels for longer molecules (100s to 1000s), and acrylamide gels for shorter (10s-100s).

Chapter 2

Nucleic Acid Chemistry

2.1 Questions

1. What is the significance of base tautomerization? (3)
2. Why is DNA sequencing limited by roughly a 50,000 year time scale?
3. Usefulness of cytosine deamination?
4. Names of deaminated adenine?
5. What happens when cytosine is methylated (but no deamination)?
6. Why do cytosine methylation sites have higher rates of mutation?
7. Which bases are methylated?
8. Where are 5' and 3' positions on the simplified DNA backbone diagram?
9. What kind of gel do you use for Sanger sequencing?

2.2 Key

1. (a) Tautomers are **different chemicals**, not just resonance structures.
(b) They change the H-bond donor/acceptor patterns in the base.
(c) They pave the way toward spontaneous deamination, an irreversible change. Cytosine tautomer \rightarrow deamination leads to uracil. It occurs once in every 50,000 years ($5 * 10^{-13} s^{-1}$) for each individual cytosine, but there are 3 billion cytosines in each cell, so it happens quite often actually. This uracil then leads to an A on the opposite strand, and the A will lead to an AT base pair eventually. In this way, GC can become AT permanently.
2. To convert U to C requires energy. DNA repair dies when the cell dies, so there are a bunch of uracils in ancient DNA.
3. (a) Allows for diverse antibody genes in B lymphocytes
(b) mRNA editing can change protein sequence translated from mRNA (catalyzed by APOBEC)
(c) In viral RNAs, C \rightarrow U changes block viral replication (catalyzed by APOBEC)
4. Deaminated adenine: hypoxanthine (the nucleotide with this base is called inosine).
5. No change to the genetic information because H-bond donor/acceptor patterns have not been changed.

6. Methylated cytosine can be de-aminated to form thymine. Specialized U repair never gets a chance to fix the DNA, although mismatch repair is still available.
7. C and A
8. 5' is the pointy thing on top, 3' is the point lower than the phosphates.
9. Capillary gel electrophoresis.

Chapter 3

DNA Topology and Chromatin

3.1 Questions

1. Type I topoisomerase vs Type II topoisomerase
2. Describe histone structure.
3. What is the significance of histone tails?
4. Why is Type I nucleophilic attack reversible?
5. What is the effect of ethidium bromide on DNA topology?
6. Starting with $Lk = 200$ and $Tw = 200$, how can you achieve $Lk = 175$, $Tw = 200$, and $Wr = -25$ using EtBr?
7. What does plectonemic supercoiling refer to?
8. Describe the big-picture organization of DNA in bacteria
9. Since eukaryotes have no gyrase, how do they maintain negative supercoiling in their DNA? Describe this quantitatively.

3.2 Key

1. (a) Type I
 - i. Allows one twist in either direction. Lk moves up or down by 1
 - ii. Breaks one strand
 - iii. Uses no energy
 - iv. No bias for adding or removing twist
 - v. CANNOT handle two separate DNA molecules
 - vi. Change Tw and Lk , doesn't directly change Wr
- (b) Type II
 - i. Breaks two strands
 - ii. Lk goes up or down by 2
 - iii. Decatenates two double-helix circles
 - iv. CAN handle two separate DNA molecules as well as a supercoil of one double-helix
 - v. Uses ATP energy
 - vi. **Bacterial gyrase** is an example—it creates a negative supercoil, trading ATP energy for higher-energy DNA

2. Histone octamer, 2 each of H2A, H2B, H3, H4. They are rich in Arg and Lys residues, so they are positively charged at pH 7.
3. Post-translational modifications to the tails bookmark active or inactive genes in the genome.
4. The OH on tyrosine attacks a phosphodiester bond, creating another ester. Hence, since the original bond was also an ester, the nick is reversible.
5. Ethidium bromide binds DNA ("intercalating agent"), and reduces Tw.
6. (a) Add EtBr, bringing Tw down to 175, Wr to +25
(b) Add Type I topoisomerase, bringing Wr to 0. Overall: Tw = 175, Lk = 175
(c) Wash out topoisomerase
(d) Wash out EtBr – now, no more intercalating agent to prevent Tw from returning to default, so Tw goes back to 200, Wr goes to -25
7. It refers to supercoiling that occurs when DNA wraps around itself, as opposed to solenoidal supercoiling that occurs with histones.
8. DNA is organized into 10 kb loops whose boundaries are never well-defined. Topological linkage is limited to each loop (i.e., if Tw goes down by 1 in loop A, Wr only goes up by 1 in loop A). Gyrase adds plectonemic supercoiling to these loops.
9. They use histones, which create negative (LEFT-HANDED) solenoidal supercoiling. Each histone adds $-1\frac{2}{3}$ supercoil, introducing $+1\frac{2}{3}$ plectonemic supercoiling. A topoisomerase can come along and relax this plectonemic supercoiling to result in a net negative supercoil.

Each nucleosome consists of 146 bp DNA wrapped around, and nucleosomes are spaced apart roughly 200 bp.

Chapter 4

DNA Replication

4.1 Questions

1. What cofactor is necessary to stabilize the transesterification reaction in DNA synthesis? What amino acids are necessary?
2. What is processivity/processive? What is its opposite?
3. What are the terms used to describe the movement of polymerase as it adds nucleotides or falls off?
4. How does DNA polymerase ensure that there is correct base pairing? What is its error rate before proofreading?
5. When does proofreading occur? What actually happens during proofreading? What is the final error rate after proofreading?
6. Why is proofreading not 100% to the organism's benefit?
7. What is the name of the bacterium from which we derive our polymerase that survives many heating cycles?
8. DNA Polymerase I in *E. Coli* is very abundant. Why is it not suited to be the main DNA replicating enzyme?
9. In the Meselson-Stahl experiment, what would you expect to see after one round of heavy DNA replication in light dNTPs? What about 2 rounds?
10. How do the majority of eukaryotic genome replications differ from yeast replication?
11. What is the replicative speed of DNA polymerase I? What is its processivity?
12. What is the speed and processivity of DNA polymerase III? Why is DNA Polymerase III highly processive?
13. Describe the structure of DNA Polymerase III. Which subunits are located where?
14. What is the enzyme that unwinds parental DNA? Describe its structure and behavior.
15. Why do helicase and topoisomerase work in tandem?
16. While the lagging strand has not been replicated yet, it is exposed to the aqueous environment for a long time. Why isn't it affected?
17. Which three enzymes are involved in the synthesis of the lagging strand other than DNA Polymerase III?

18. What happens to beta clamps during DNA synthesis, leading/lagging strands? As a consequence, what happens to the clamp loader?
19. What makes DNA Pol I unique, and aptly suited for its role in degrading RNA primers?
20. How does DNA ligase seal nicks?
21. What is the structure of the protein clamp that aids DNA Polymerase III?
22. How do enzymes make replication initiation in bacteria relatively easy?
23. How do bacteria untangle the final completed DNA strands?
24. Compare/contrast the prokaryotic replisome with the eukaryotic one.
25. How do eukaryotic genomes replicate quickly?
26. How do most eukaryotic genomes ensure that replication only begins once per bubble?

4.2 Key

1. Mg^{2+} and aspartate are necessary.
2. Processivity is the number of cycles before polymerase dissociates—also describes the number of dNTPs added before falling off. A processive enzyme performs many reactions before dissociating. The opposite of a processive enzyme is distributive.
3. Translocate = moving one nucleotide down. Dissociation = falling off
4. If mispaired bases enter the active site of the polymerase, there will be steric clashes, disallowing energetically favorable mispairing (tautomers, weird conformations). Its error rate is 10^{-4} to 10^{-5} .
5. (a) Proofreading occurs after incorporation but before translocation.
(b) The mispaired nucleotide is moved into the $3' \rightarrow 5'$ exonuclease site, and the nucleotide is cleaved off. The terminus is then repositioned into the $5' \rightarrow 3'$ polymerase site.
(c) This gives a final error rate of 10^{-6} to 10^{-8} (improves fidelity 10^2 to 10^3).
6. (a) Proofreading slows down translocation. Translocation and proofreading are in kinetic competition with each other
(b) There is an energy cost. Sometimes correct bases are removed and resynthesized
7. *Thermus aquaticus*.
8. It moves too slowly and is not processive.
9. After one round, I would expect to see one layer that is not as heavy as the initial band. After two rounds, I would expect to see two bands, one intermediate (the light strand and the heavy strand), and one light.
10. Yeast have autonomously replicating sequences (ARS), which are defined replication origins. Most eukaryotes do not have defined replication origins.
11. It replicates about 20 bp per second. Its processivity is about 200 bp. (falls off every 200 bp)
12. 1000 bp per second, 500,000 bp processivity. It is processive because it has a protein, a β sliding clamp, that follows the polymerase and keeps the enzyme hooked to the DNA.
13. Two polymerase cores, composed of α , ϵ , and ω subunits (α polymerizes, ϵ proofreads—they are separate proteins). Behind each of those is a β sliding clamp. The polymerase cores are tethered to a clamp loader with τ subunits. The clamp loader itself is composed of γ , $(2)\tau$, δ , δ' subunits. It can lock the clamps around DNA.

14. Replicative helicase unwinds DNA. It is a hexamer (6 copies of the same protein) that forms a ring around the DNA, and it uses ATP to pull one strand through the ring, peeling apart the DNA
15. Topoisomerase needs to release the topological strain caused by the helicase.
16. It is coated with single strand binding (SSB) proteins until it is replicated.
17. DNA topoisomerase II which maintains a negative supercoil to promote untwisting. DNA-B helicase unwinds DNA. Primase adds the RNA primers
18. Leading strand: just one beta clamp is needed for each 500,000 bp. The beta clamp gets left behind at the 3' end of each Okazaki fragment, and the Pol-III enzyme jumps forward to the next beta clamp. Since we need new clamps constantly, the clamp loader must follow the replication fork, not just needed at the beginning.
19. It has a 5'→3' exonuclease (in addition to its 3'→5' exonuclease for proofreading, which most DNA polymerases have).
20.
 - (a) Its NH_3^+ group first conjugates (adenylylates) to AMP (from either ATP or NAD, releasing either PPi or NMN).
 - (b) The 5' phosphate attacks the phosphate on the AMP, removing AMP from the enzyme and adding an extra phosphate leaving group on the 5' end.
 - (c) The 3' OH attacks the 5' phosphate and kicks out the AMP as a leaving group.
21. It is a symmetric dimer with two copies of the same protein named beta.
22.
 - (a) DnaA protein binds to DNA at the origin, adds positive solenoidal supercoiling, and denatures the DNA unwinding element (DUE)
 - (b) To compensate for the positive linking number, DnaB replicative helicase is loaded with the help of DnaC + ATP and untwists/melts DNA.
23. They use topoisomerase IV (a type II topoisomerase).
24.
 - (a) Eukaryotic has two replicative polymerases (delta and epsilon)
 - (b) Eukaryotic sliding clamp is a homotrimer, a PCNA (proliferating cell nuclear antigen), which is a marker of a dividing cell
 - (c) Eukaryotic helicase (MCM2-7 complex) is still made up of 6 proteins, but they are slightly different (but related!)
25. Multiple replication origins.
26. The ORC (origin replication complex) assembles, then binds to DNA and ATP during G1. During S phase, ORC hydrolyzes the ATP and loads MCM2-7 helicase onto the strand. It cannot fire a second time during S phase.

Chapter 5

DNA Repair and Recombination

5.1 Questions

1. What are the main types of DNA damage/damage signatures/repair mechanisms?
2. What is an example of how the high priority cells give to maintaining genome integrity that involves guanine?
3. What is an AP site? How does it form? How does the cell recognize/fix it?
4. What is the difference between base excision and nucleotide excision?
5. How is mismatch repair dealt with?
6. How do mutations in mismatch repair and nucleotide excision pathways relate to different types of cancer?
7. What happens in bacteria if a damaged template strand is not repaired by the time a replicative polymerase gets there?
8. What happens in eukaryotes if a damaged template strand is not repaired by the time a replicative polymerase gets there?
9. What happens when there is a break in the sugar phosphate backbone on one strand in bacteria?
10. What happens in eukaryotes when there is non-homologous end joining?
11. MMEJ—explain it.

5.2 Key

Summary

DNA damage Chemical changes to bases can cause mis-pairing
 Rounds of replication turn mis-pairing into mutation
 Breaks and bulky lesions block replication entirely
 Mutagens and loss of DNA repair promote cancer

Repair pathway	Damage signature	Repair strategy
Direct repair	altered base, e.g. O ⁶ -Me-G	reverse chemical change
AP/abasic site	No base	cut (AP Endo), resynthesize (DNA Pol I), ligate (DNA Ligase)
Base excision	uracil or damaged base	remove base (glycosylase), AP repair
Nucleotide excision	distorted DNA backbone	cut on both sides (excinuclease), remove (helicase), resynth & ligate
Mismatch	normal but mis-paired bases	Identify “new” strand, degrade, resynthesizes & ligate

DNA breaks Homology directed repair — replication fork restart
 Process end, strand invasion, branch migration, resolution
 Non-homologous end (re-)joining: ligate them together ³¹

- (image courtesy of Professor Nicholas Ingolia)
- Methylated guanine is fixed back to its keto form using a cysteine-containing methyltransferase. The methyltransferase accepts the methyl group and is permanently inactivated—the cell synthesizes a protein for a single use (wow!).
- An AP site = apurinic/apyrimidinic, abasic site, which is
 - caused by spontaneous hydrolysis of the N-glycosidic bond between deoxyribose and the nitrogenous base. The nitrogenous base then just pops off the backbone.
 - AP endonuclease cuts the 5' end of the AP/abasic site, then DNA Pol I + DNA ligase repair the damaged strand, similar to Okazaki fragment concatenation.
- Base excision = there is a damaged base, and DNA glycosylase cuts out the base, leading to AP site. Then same process as 3 happens.
 - Nucleotide excision occurs when there is more significant DNA damage (UV light, for example). The helix has a funny shape, so exinuclease cuts the sugar-phosphate backbone (excises 13 nucleotides in bacteria, 29 nucleotides in eukaryotes). DNA helicase removes the damaged region. DNA Polymerase I fills in the gap (in E. Coli) and DNA polymerase epsilon fills in the gap in eukaryotes. DNA ligase then fixes the nick.
- MutS and MutL recognize the mismatch and hydrolyze ATP to bind to the site. Then, enzymes take advantage of the fact that “old” DNA has already been methylated (GATC is methylated in E. Coli, which is a motif the cell recognizes), and they remove the un-methylated “new” strand. MutH (+ATP to bind) + MutS + MutL make up the mismatch repair complex, which searches in both directions to find a methylated GATC. The mismatch is looped out of the sequence, and then the un-methylated strand of the hemi-methylated site is cut, either 5' or 3' relative to the mismatch. Then:
 - If the cut was 5' relative to the mismatch, exonuclease VII or RecJ nuclease comes in to degrade 5' → 3'

- (b) If the cut was 3' relative to the mismatch, exonuclease I or exonuclease X comes in.
- (c) In both cases, MutL-MutS and DNA helicase II are involved.

After the degradation occurs, SSB and DNA Polymerase III + DNA ligase closes the gap.

6. Mismatch repair takes care of replicative errors. Colon cells divide constantly and thus replicate DNA constantly. Therefore, mismatch repair defects (MSH, MLH, homologs to MutS and MutL) are particularly bad for colon cells, leading to Hereditary Non-polyposis Colon Cancer (HNPCC). Nucleotide excision takes care of UV damage, so xeroderma pigmentosum becomes prevalent in people with nucEx defects.
7. The SOS response occurs. A trans-lesion polymerase is required, something like polymerase IV and V. They are extremely error-prone because they can't proofread, but they allow replication to occur after the lesion, even if there are many mistakes.
8. Trans-lesion polymerases are also activated, but they only synthesize a few bases. There are many of these polymerases, some specialized for stuff like T-T dimers/UV damage (DNA Pol η for example).
9.
 - (a) The replication fork collapses, and the strand without the break will continue being replicated.
 - (b) The 5' end of the broken template is degraded (resection)
 - (c) RecA coats the loose end, promoting strand invasion
 - (d) Branch migration occurs, forming the Holliday intermediate e) Once DNA replication has passed the nick site, a special nuclease resolves the Holliday junction and restores the replication bubble
10. Ku end-binding protein joins the 2 broken ends directly. It cannot fix the replication fork collapse.
11. Microhomology-mediated end joining. It trims back DNA to find a micro homology of a few bases and anneals the homologies. Since it removes extra bases, some information is lost. But since both methods lose information, this is better.

Chapter 6

Episomes and Cloning

6.1 Questions

1. Plasmids vs BAC?
2. What is the purpose of restriction endonucleases? How do they work?
3. Describe the concept of ligation compatibility.
4. What is a polylinker?
5. What is an example of seamless cloning, and how does it work?
6. What is the significance of the Tn3 transposon from *E. Coli*?
7. What does the presence of small duplications of "host" DNA indicate?
8. How are transposons involved in humans?
9. How does the "copy-and-paste" mechanism work in transposons?
10. What does recombinase use to perform site-specific recombination? How is it similar to topoisomerase?
11. How is the "copy-and-paste" mechanism related to viruses?
12. How do you determine whether a recombinase site will be inverted or cut out?
13. Where do Cre and Flp operate?
14. What are the less frequent uses for Cre and Flp? The most frequent?

6.2 Key

1. (a) Plasmids
 - i. Plasmids are "selfish" (replicate by forming additional copies of itself within the genome but don't necessarily contribute to fitness)
 - ii. Plasmids are not always passed to daughter cells
 - iii. May be low-copy (5-20) or high-copy (500)
- (b) BAC
 - i. BAC can hold more DNA (200 kb DNA inserts)
 - ii. BAC have replication origin that is different from plasmid ori
 - iii. *par* genes for partitioning BAC, one for each daughter cell after replication

- (c) Both
 - i. Often carry antibiotic resistance
 - ii. Both have origins of replication, but they are not the same
- 2. They cleave phage DNA but not the bacterial DNA because bacteria methylate their own genome. Usually they recognize 4 to 8 nucleotide palindromes and either cut down the middle to make blunt ends or leave sticky ends with free phosphates and OH's.
- 3. Compatibility in this context refers to the affinity for two cut ends to be re-base-paired. Two sticky ends cut by the same enzyme are compatible. All blunt ends are compatible. Sticky ends created by two different enzymes are *usually not* compatible.
- 4. A synthetic piece of DNA that researchers insert into plasmids. The insert contains many overlapping restriction sites for common enzymes.
- 5. Gibson Assembly. If two DNA strands have matching ends, a 5' → 3' exonuclease can chew up one end of both strands. Then, the two strands can anneal together, DNA polymerase can fill in the gaps, and DNA ligase seals the nicks. This is great because you can add sequences that match the plasmid DNA at the 5' end of your PCR primers, or you can concatenate multiple inserts
- 6. It is an example of a transposon that contains a transposase, a β -lactamase gene that codes for an enzyme that can destroy penicillin/ampicillin, and a resolvase (another enzyme that facilitates "jumping").
- 7. It indicates the presence of a transposon insertion. When transposase cuts into DNA, it often makes staggered cuts, leaving a gap on both ends after the transposon is inserted.
- 8. The generation of antibodies is mediated by transposon-like elements, RAG1 and RAG2. A B-cell chooses a V, D, and J segment and transposase cuts the unwanted DNA. Free OH groups on the desired portion can attack the opposite complementary strand and kick out the undesired DNA. Then double-strand break repair by end-joining.
- 9. The transposon is cut on each strand, and the dangling OH's attack another molecule of DNA, connecting the two DNA molecules. During replication, the cell will recognize this as a DNA break/Holliday junction that needs to be resolved. A special type of recombination separates the two molecules, called **site-specific recombination**.
- 10. Recombinase uses Tyrosine to cut the DNA. Both recombinase and topoisomerase are DNA-bound proteins.
- 11. Both utilize the same site-specific recombination. Site-specific recombinases can integrate viral DNA into the genome as well as excise it back out.
- 12. If the recombinase sites are facing the same direction, a circle will be cut out (direct repeats). If they are facing opposite directions (inverted repeats), they will be inverted.
- 13. Cre recombinates at loxP sites, Flp recombinates at FRT sites.
- 14. Most frequent: excision. Can also use for inversion, insertion (circle and target site must have aligned repeat sites), and translocation, whereby two different chromosomes are cut and pasted together (mimics cancer conditions).

Chapter 7

Bacterial Transcription and Regulation

7.1 Questions

1. What cofactors and amino acids might you need for RNA polymerase?
2. What is a consequence of transcription that requires another enzyme to be recruited?
3. Transcription moves 5' → 3' relative to the RNA strand or the template strand? Which way does a transcript arrow point?
4. Of the template strand that is transcribed and the reverse complement to the template strand, which is considered the coding strand?
5. Which is NOT random: the choice of coding strand, the choice of where to begin transcription.
6. Bacterial vs human genes in terms of transcript spacing in the genome.
7. Describe the structure of the *E. Coli* RNA polymerase.
8. RNA polymerase cannot recognize promoters, so how does it synthesize RNA?
9. Describe the promoter region. Which strand does the sigma factor recognize?
10. If there is a mutation in either the spacer or specific region of the promoter, will transcription still proceed?
11. Which type of RNA is the most common? Which has the "highest" affinity for promoters?
12. Since DNA-binding proteins don't break apart DNA, why are sigma-factors sensitive to 1 nucleotide additions, even if the mutation occurs in the spacer region? Why are they sensitive to mutations at all (i.e., if the mutation occurred in the specific part)?
13. Describe the steps in transcription.
14. Describe the two main ways that transcription can stop.
15. If a strand loses an activator protein or the binding site of the activator protein, will transcription still proceed? What about losing a repressor?
16. When do cells transcribe lacZYA? What do those genes represent?
17. How many lacI operator sites are there? What kind of repeat sequence are they? Describe the nature in which the lacI proteins bind to the operator.

18. Describe how the repressor protein interacts with the operator from a molecular standpoint.
19. Other than a repressor bound to allolactose, what else is required for high levels of gene expression?
20. If glucose is in high concentration and lactose is present, is there NO transcription or a little transcription?

7.2 Key

1. Aspartate conjugated to Mg^{2+} which is conjugated to phosphates of the incoming NTP.
2. Linking number cannot change, so negative supercoiling occurs behind the RNA polymerase and positive supercoiling occurs ahead of the polymerase. Topoisomerase is required to relieve this added tension.
3. $5' \rightarrow 3'$ relative to the RNA strand. The transcription arrow points $5' \rightarrow 3'$ relative to the RNA strand.
4. The **non-template** strand is considered the coding strand.
5. BOTH are NOT random. Transcription is specific in terms of both the strand selected and the place it begins.
6. Bacterial genes are more tightly packed but have little overlap. Large genomes like humans' have large non-transcribed regions. Small genomes (viral) are most likely to have overlapping transcripts.
7. Five proteins, 4 of which are unique. Two α , β , β' , ω . The active site is between the β and β' subunits. NO proofreading portion.
8. The sigma factor (default: sigma-70) binds the core RNA polymerase and promoter.
9. Two promoter-specific sequences separated by about 17 nucleotides (by convention, we use the sequence on the coding strand). The sigma factor can recognize either strand of DNA.
10. Yes. The more mutations, the lower the affinity of RNA polymerase for the promoter, but there will still be something produced.
11. Answer is rRNA for both. Up elements that bind the two core polymerase α subunits are most common for rRNA transcripts.
12. DNA-binding proteins interact with the major groove of DNA by inserting an alpha helix. All the protein's contacts lie on the same physical side of the double helix, so if there is an extra nucleotide in the spacer region, the important nucleotides might be shifted away from the protein side.

The proteins are sensitive to mutations in general because major grooves have more distinctive patterns of H-bond acceptors and donors on the surface than minor grooves. Minor grooves are also less accessible because they present less surface area.

Furthermore, protein side chains interact with specific bases. The wrong H-bond donor/acceptor, steric clashes, etc will all prevent favorable interactions from occurring between the protein and the bases.

13. (a) Sigma factor and core polymerase binds with DNA as a closed complex
- (b) DNA helix is unwound to form open complex 17 bp
- (c) 8-15 nucleotides are synthesized in the open complex, then released (abortive initiation). Happens multiple times
- (d) Eventually the favorable interactions that recruited the polymerase are broken (*promoter clearance*) and the polymerase synthesizes the rest of the sequence. Once it hits this stage, the polymerase is highly processive and cannot restart, although it can stall.

14. An RNA loop (hairpin) followed by a poly-U tract ends transcription. This is **intrinsic (ρ -independent) termination**

In ρ -*dependent* termination, an ATP-dependent helicase ρ is required.

- (a) ρ binds a *rut* sequence in the RNA
 - (b) ρ travels toward the 3' end using ATP hydrolysis
 - (c) Then somehow, once it catches up to the RNA polymerase, the helicase separates the mRNA from the DNA template.
15. No. Losing the activator means that the gene has become an uninducible (uninducible phenotype), and it is not transcribed.

Similar logic applies to losing a repressor. The loss of a repressor means that the gene is *always* transcribed, and the gene has become *constitutive*.

16. These genes are expressed when lactose is present AND glucose is absent. *lacZ* encodes β -galactosidase, which breaks down lactose into monomers, *lacY* encodes permease, which brings lactose into the cell, and *lacA* encodes acetylase. *lacZ* converts some lactose into allo-lactose as a side product, which is directly sensed within the cell as an indicator that lactose is present.
17. There are three *lacI* operator sites. The operator sites are symmetric, inverted repeat sequences of 7 bases. Two *lacI* proteins bind each operator as a dimer.
18. The *lac* repressor is a helix-turn-helix DNA binding protein. One α helix inserts into the major groove of DNA, and side chains that stick out from the helix interact with bases in the major groove, allowing the protein to recognize one AATTGT.
*However, keep in mind that there is still one *lacI* dimer per operator site.
19. cAMP-bound CRP (an activator protein). CRP interacts directly with α polymerase and increases binding energy. cAMP is only made in bacteria when glucose levels are low.
20. Low transcription.

Chapter 8

Eukaryotic Transcription and Regulation

8.1 Questions

1. Types of RNA polymerases and their functions?
2. Describe the structure of RNA polymerase II. Be sure to note which are analogous to the prokaryotic version.
3. What are some notable features about the promoter regions of Pol II?
4. How is the beginning of transcription in eukaryotes similar to that in prokaryotes?
5. Where is the C-terminus tail? What is the heptapeptide unit that repeats at the C-terminus of Pol II? What modifications can happen to that heptapeptide unit? How are those modifications made?
6. What is the basis for the different behavior of enzymes when they interact with the modified C-tail?
7. Which serines are phosphorylated during which stages of transcription?
8. Discuss briefly about the proteins involved in C-terminus tail modification during transcription.
9. What are the three main functions of TF_{II}H? Are its actions energy-dependent?
10. How does TF_{II}H switch to nucleotide excision repair?
11. What is an enhancer?
12. Describe the zinc finger.
13. Which is more specific: zinc finger or the protein that bends DNA? What protein bends DNA? Is it energy-dependent? On what type of energy?
14. Assume that the DNA has bent and that the activator is bound to the enhancer. How does RNA polymerase get recruited?
15. If eukaryotic DNA is wrapped up in chromatin and histones impede transcription, how do promoters/enhancers get recognized? What type of protein solves this problem?
16. Specifically, how does acetylation work?
17. Why is acetylation important in terms of transcription?

18. If an activator protein binds to an enhancer, but the DNA is wrapped around histone, how might the cell resolve this? In particular, what was the example given in class? What if we wanted to do the reverse (repressor binds to an operator to wrap DNA around histone)?
19. What are some important co-activators? Where do they bind?
20. Two types of chromatin?
21. Where does the methyl in methylation come from? What is the target of methylation? How does the opposite of methylation work?
22. How is heterochromatin like a disease?
23. Where does DNA methylation occur? What is its effect?
24. Discuss the inheritance of epigenetic factors.

8.2 Key

1. (a) Type I: rRNA
(b) Type II: mRNA
(c) Type III: tRNA
2. Made up of 12 different proteins, RPB1 through RPB12.
 β and β' in bacterial RNA Pol are orthologues of RPB1 and RPB2—these make up the active site. The α dimer orthologues are RPB3 and RPB11.
3. (a) No single element is universal
(b) TATA was discovered earliest but is by no means present in every promoter region
(c) Some Pol II elements lie downstream of the transcription start site.
4. They're actually very similar: both have the same abortive initiation cycle. General transcription factors bind RNA polymerase to the DNA (DNA still closed, preinitiation complex). Inr (initiator element) unwinds the DNA, begins the cycle.
5. It's the C-terminus of RPB1, one of the catalytic sites.
It is made up of repeating units of the heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser (Tyred, (the) serene programmer threw serpents probably seriously).
Many of these residues have OH groups: the three serines, tyrosine, and threonine can be phosphorylated at different times. Protein kinases use ATP to phosphorylate serine because they activate the OH to attack ATP, and phosphatases do the opposite. Mostly we care about **serine-2 and serine-5**.
6. The original OH on serine is very small and neutral. The resulting phosphoester is negatively charged and huge. This is a significant difference for active sites—hence, different enzymes will have different affinity for serine or phospho-serine.
7. (a) Upon initiation, serine-5 is phosphorylated
(b) As elongation continues, serine-2 is phosphorylated
(c) Serine-5 phosphorylation decreases as elongation continues
8. (a) The protein kinase that phosphorylates serine-5 is part of the $TF_{II}H$ (Transcription Factor II H) complex
(b) The protein kinase that phosphorylates serine-2 is part of the pTEFb (positive transcription elongation factor).
9. (a) Contains the protein kinase that phosphorylates Ser5 at initiation

- (b) It contains the XPB (DNA) helicase that unwinds for transcription ($3' \rightarrow 5'$)
 - (c) This same DNA helicase is the one in nucleotide excision repair.
 - (d) FYI XPD is $5' \rightarrow 3'$ helicase. XPB/XPD = Xeroderma Pigmentosum
Yes, this is ATP-dependent
10. XPB alone unwinds DNA for transcription initiation, but XPB and XPD unwind DNA for NER, one in each direction. Furthermore, if there is major UV damage on the **template strand** RNA polymerase will stall, triggering NER.
 11. An enhancer is an activator binding site that is located relatively far from a promoter.
 12. Small, sequence-specific DNA-binding protein structure. One zinc ion held in the middle by cysteine or histidine. The helix gets inserted into the major groove and contacts **3** bases. Proteins can have multiple zinc fingers (hence, recognizing bases in multiples of 3).
 13. (a) Zinc finger is more specific. Actually, the DNA-bending protein is NON-specific and merely binds to minor grooves.
(b) The HMG box protein is the architectural regulator that bends the DNA so that the enhancer is closer to the promoter
(c) It is energy-dependent: the HMG protein gets its energy from favorable binding interactions between DNA and protein.
 14. The activator protein binds to a (huge) mediator protein which then binds TATA-binding protein (TBP)/ TF_{II}B, then RNA polymerase II. *Also note that co-activators can bind the mediator
 15. Promoter-DNA is nucleosome-free. Nucleosome positioning can be driven by enzymes through chromatin remodeling.
SWI/SNF is an important ATP-driven nucleosome remodeling enzyme. It can:
 - (a) Eject (remove) a histone
 - (b) Insert a histone
 - (c) Slide a histone somewhere else
 - (d) Dimer exchange
 16. Histone acetyltransferase (HAT) adds an acetyl group to the NH_2 group of lysine. **The acetyl group comes from acetyl-CoA**. Histone de-acetylase (HDAC) hydrolyzes the *amide* bond.
 17. SWI/SNF binds H3K9-Ac and H3K14-Ac and relaxes the chromatin so that it is available for transcription.
 18. The Gal4p activator binds the enhancer, distant from the GAL gene. Gal4p is activated by galactose and other enzymes. Active Gal4p recruits SAGA, which contains a histone acetyltransferase, which lets SWI/SNF come in to relax the DNA and make it available for transcription.

A repressor can do the opposite. A repressor can bind histone deacetylases to undo lysine acetylation.
 19. SAGA is an important co-activator. Co-activators do not recognize DNA—they bind activator proteins. SWI/SNF is another co-activator.
 20. Heterochromatin (tightly packed) and euchromatin (less dense, transcription active).
 21. Methyl comes from S-adenosylmethionine. Methylation targets lysine, and it can add 1, 2, or 3 methyls per lysine. In de-methylation, one can rip off an H from the methyl group, forming an imine. Spontaneous hydrolysis later leads to the release of formaldehyde, leaving behind just N and H.

Note that acetylation and methylation cannot happen at the same time—they are mutually exclusive

22. Heterochromatin can "spread." Heterochromatin has methylation on H3K9. Heterochromatin Protein 1 (HP1) binds histone tails with H3K9-Me. HP1 binds SUV39H1 histone methyltransferase, which methylates H3K9 on the next tail, creating a new HP1 binding site.
23. It occurs on the nitrogenous base of cytosine—specifically, it occurs on cytosines that are part of a CpG dinucleotide.

Proteins bind to methyl-cytosine and reduce transcription (i.e., chromatin remodeling might bind to methyl-C and make repressive chromatin).

24. A child inherits one methylated strand and one new unmethylated strand. Dnmt1 (DNA methyltransferase I) finds hemi-methyl DNA and methylates the other C on the child strand. This is *maintenance methylation*. New methylation patterns only happen during gamete formation and early embryogenesis by Dnmt3.

Chapter 9

RNA Processing

9.1 Questions

1. Why would an agent that wipes out snRNAs inhibit splicing?
2. Describe the structure of the 7-Me-guanosine cap.
3. What are the steps to adding the 5' cap to mRNA?
4. Describe what happens to the 5' cap during transcription.
5. When does splicing occur?
6. Which phosphate remains in eukaryotic mRNA splicing?
7. How do Group I self-splicing introns work?
8. How do Group II self-splicing introns work?
9. How is pre-mRNA splicing different from Group II self-splicing?
10. How does the spliceosome bind to the transcription complex?
11. What sequence is common to all introns that are not self-spliced?
12. Describe the specific binding of a spliceosome to an RNA sequence.
13. What is the purpose of U4, U5, and U6?
14. What is a lariat?
15. Whose phosphoester bond does the nucleophilic adenine attack during lariat formation?
16. What happens in between when an inactive spliceosome is converted to the active form?
17. What do you call a different mRNA version made from the same pre-mRNA?
18. How do cells determine which introns to cut out?
19. How would disabling splicing in *Drosophila* affect sex determination?
20. Describe the final steps of transcription.
21. Other than splicing, what are some ways to get alternate transcripts of the same gene?
22. What are the main ways in which RNA is degraded?
23. How are viral genes integrated into the host DNA?

24. How do cells repress retrotransposition?
25. Which kind of transposon is more effective at increasing copy number?
26. Which two proteins are essential for viruses to encode?
27. What is a retrovirus that cannot make a protein coat? What happens to its genetic information?
28. What are the two types of retrotransposons?

9.2 Key

1. Spliceosomes are made of five snRNPs (small nuclear ribonucleoproteins): U1, U2, U4, U5, U6. Each one is made of one snRNA and proteins that bind to it. Wiping out snRNAs would mean that this complex could not be assembled.
2. The 5'-O of guanosine is connected to the triphosphate. Hence, there is no free OH on the 5' end of mRNA if the guanosine cap has been added. The nitrogen of the pentagon section of the guanosine is methylated. The first 2 nucleotides other than guanosine are sometimes methylated on 2'. The ribose of the guanosine has an OH on both the 2' and 3' carbons.
3.
 - (a) One phosphate removed from the end by hydrolysis.
 - (b) The remaining phosphate attacks GTP, releasing inorganic pyrophosphate and restoring a triphosphate + guanine to the end of the RNA.
 - (c) Guanine-7-methyltransferase methylates the cap.
 - (d) 2'-OH of first and second ribose may also be methylated
4. The cap is added *during* transcription. The capping proteins bind to Pol II CTD, which has a serine-5 phosphorylated. Once it is capped, the nuclear cap-binding complex binds to the cap.

*Note: Recall that transcripts through Pol I and Pol III are not capped.
5. Splicing can occur during and after transcription.
6. The 5' phosphate of the first nucleotide of the 3' exon is left. The last base before the intron (5' end) loses the phosphate on its 3' oxygen.
7. The 3' OH of free guanosine monophosphate attacks the phosphate between exon #1 and the intron. This leaves an OH on the exon #1 side, and there is now a G on the 5' end of the intron. That OH then attacks the 5' phosphate on the first base on the second exon, kicking out the second phosphoester bond that still ties the intron to the exon.
8. A 2' OH of an adenosine from within the intron attacks the phosphate of an exon. This means that in addition to the 5' and 3' phosphodiester bonds, that adenosine also has a 2' phosphodiester bond. Then the resulting OH on the exon attacks the phosphate on the neighboring exon, kicking out the intron. Once again, the first phosphate on the second exon remains.
9. A spliceosome (RNA complex + protein) binds the intron to catalyze the initial OH attack, but mechanistically it is identical to Group II self-splicing.
10. It binds to the CTD.
11. All non self-splicing introns begin with 5'-GU and end with AG-3'. The other bases do not need to match up exactly.
12. U1 snRNA binds to the 5' end of the intron. U2 snRNA binds to the branch point, where the nucleophilic adenine/OH is located.

13. They bring U1/U2 physically closer together after they have binded to the RNA.
14. A lariat is the funny loopy thing that forms during RNA splicing, when adenine catalyzes an attack.
15. It attacks the 5'-phosphodiester bond of G, of the GU that begins the intron.
16. ATP is hydrolyzed, U1 and U4 are released, U6 replaces U1.
17. Splice isoform.
18. Splicing enhancers and splicing silencers favor/disfavor the use of splice sites at specific sequences. More specifically, recall that U1 and U2 snRNPs contain snRNAs that can base pair to specific sequences. Depending on the particular snRNA within the U1 and U2 subunits, difference sites will be recognizes and spliced.
19. Most likely, all flies would be male. In males, U2 binds to an earlier 3' splice site, so there is a shorter intron. The resulting exon includes a very early stop codon, and no **Tra protein** is made. In female flies, the Sx1 RNA-binding protein binds the early 3' splice site, so U2 is forced to bind a later 3' splice site. The resulting intron is longer and includes the stop codon. Thus, the Tra protein is translated, and the fly expresses a female phenotype.
20. An RNA sequence is recognized by a protein, which then cleaves extra base pairs off the RNA, and transcription stops. Poly-(A) polymerase adds a poly-(A) tail onto the mRNA without a template. poly-(A) binding protein (PAB) then binds to the A tail.
21. Alternative promoters upstream of the gene, or alternative cleavage (poly-A) sites downstream of the gene.
22. (a) Remove the 5' cap (DCP2), then Xrn1 exonuclease degrades 5' → 3'
(b) Deadenylate (with deadenylase), then exosome degrades 3' → 5'
(c) Endonuclease cleaves in the middle of the RNA, allowing Xrn1 and exosome to work both ways at the same time.
23. Integrase binds long terminal repeats which are at the end of each viral "genome" and inserts them into the host genome.
24. (a) APOBEC enzymes deaminate transposon RNA, mutating the DNA copy
(b) DNA methylation blocks transcription of retrotransposons.
25. Retrotransposons are better at increasing copy number.
26. Reverse transcriptase and integrase
27. Long-terminal-repeat (LTR) Retrotransposon is a retrovirus that cannot make a protein coat/virus.

It is transcribed to RNA, reverse transcribed to DNA, then integrated into the host genome. This integration can happen because remember, it still has LTRs.
28. (a) LTR retrotransposons, like retroviruses
(b) Non-LTR retrotransposons: DNA is nicked, reverse transcription is primed using one strand, then no one knows what happens afterward.

Chapter 10

Telomeres and Introduction to Translation

10.1 Questions

1. Describe the end replication problem.
2. How does telomerase add to the ends of chromosomes?
3. What does it mean for a genetic code to be degenerate?
4. What are the two exceptions to the non-randomness of the genetic code wobbles, and what is the non-randomness?
5. What is some intuition behind the name "anti-codon"?
6. Given the above question, which is the wobble position? (assuming 1, 2, 3 numbering from left to right on the mRNA strand).
7. What are the allowed wobble pairs?
8. Which wobbles are synonymous?
9. True or false: Watson-Crick base pairing MUST be conserved no matter the wobble between the anticodon-codon.
10. What is the terminal "signature" of tRNA?
11. What are the arms of a typical tRNA, starting at the top, going clockwise?
12. Describe how the tRNA induces favorable base stacking energies.
13. Why do we say that a tRNA gets "charged" with an amino acid?
14. Three main points about aminoacyl-tRNA synthetases.
15. Describe the synthetase reaction (all possibilities).
16. Describe the proofreading mechanisms in aminoacyl-tRNA synthetase.
17. Describe how an amino acid is added to a growing peptide. Why can this reaction occur?

10.2 Key

1. The last Okazaki fragment leaves an RNA primer on the end of the template. Primer removal will leave a gap at the end of the sequence. In the next round of replication, information will be totally lost in one child strand. A telomere along with the TRF1 and TRF2 which are bound to it protect the important DNA.
2. It has reverse transcriptase activity—an RNA template base pairs with a few nucleotides on the terminus of a chromosome. It reverse transcribes RNA template into 6-unit DNA that adds to the existing strand. Then it translocates down the strand.
3. There is redundancy built-in to the code (64 possibilities for 20 amino acids).
4. Met and Trp are exceptions to the rule that the third base pair in a codon can wobble, but only wobble to another pyrimidine if it's a pyrimidine or to another purine if it's a purine. Met and Trp each only have one codon.
5. The sequence on the tRNA that base-pairs with the mRNA codon is the reverse complement of the mRNA sequence.
6. Position 3 (rightmost) of the mRNA strand is the wobble position, position 1 of the anticodon.
7. G:U. This is energetically favorable actually. It's just that DNA polymerase normally forbids this due to the shape of its active site.
8. Codons with either pyrimidine (C/T) at the third position are ALWAYS synonymous. With either purine, it's *usually* synonymous.
9. False. Inosine can be in the anticodon. Inosine naturally base pairs with C, wobbles with U and A (another purine!).
10. CCA
11. Amino acid arm, T Ψ C arm, anticodon arm, D arm
12. The T Ψ C arm undergoes co-axial stacking with the amino acid arm, and the D arm undergoes co-axial stacking with the anticodon arm.
13. The ester bond between terminal 3' OH of adenine and amino acid is higher in energy than a typical peptide bond. The ester bond is thermodynamically less stable and more reactive (kinetic).
14. (a) Energy of ATP \rightarrow AMP+PP_i is used to make a high-energy ester bond.
(b) One synthetase per amino acid, though a synthetase may charge a few different kinds of tRNA
(c) A synthetase "reads" the anticodon + other features in order to load the correct amino acid onto tRNA.
15. (a) COO- group on the amino acid attacks the α -phosphate on ATP. This resulting aminoacyl adenylate remains bound to the synthetase.
(b) Class I: 2' OH of terminal adenine of tRNA attacks the acyl group (O-C=O) and kicks out the adenosine with its one phosphate. However, eventually this will undergo a transesterification to switch the amino acid onto the 3' end.
(c) Class II: 3' OH attacks the acyl group.
16. (a) First, the amino acid binds to the synthetase and becomes activated to aminoacyl-AMP. If it's the wrong amino acid, the aminoacyl bond is hydrolyzed (pre-transfer).
(b) Second, there is a second active site on the synthetase. If the bound aminoacyl group fits into the second active site, the bond is also hydrolyzed (i.e., valine fits into the second active site on isoleucine's synthetase) (pre-transfer).

- (c) Third, most synthetases can hydrolyze the ester linkage between an amino acid and tRNAs if it's an incorrectly charged amino acid (post-transfer).

Final accuracy is about 3000 fold.

17. The amino group of the new amino acid attacks the terminal acyl group of the peptide. Some hydrogen swapping occurs, as the amino group also gives up one of its hydrogens to the tRNA, and the 2' OH and 3' OH are both restored. The new peptide is now on what used to be the aminoacyl-tRNA.

This reaction can occur because an amide bond is lower in energy than an ester. So ultimately, the energy of the ATP was used to make an ester, and that energy went into making the final peptide bond.

Chapter 11

More Translation

11.1 Questions

1. Proteins that recruit tRNA in bacteria? In eukaryotes? What do they bind? What do they do?
2. How do those proteins demonstrate energy cost for accuracy?
3. Where does the energy for peptide bond formation come from? What plays a role in catalyzing this reaction?
4. Enzymes involved in translocation? How do they work?
5. How does translocation preserve reading frames?
6. What is programmed frameshifting?
7. How is programmed frameshifting related to retroviruses?
8. How do release factors work?
9. How does initiation begin in bacteria?
10. What are the implications of the structure of methionyl tRNA?
11. How does initiation begin in eukaryotes?

11.2 Key

1. EF-Tu in bacteria
e-EF1 α in eukaryotes
They only bind CHARGED tRNAs, then bring them to the ribosome. It is a GTPase that binds GTP and hydrolyzes it in order to release tRNA into the ribosome. After hydrolysis, EF-Tu/eEF1 α leaves the ribosome, and the GDP is replaced with a GTP—now the enzyme can bind a new tRNA.
2. Incorrect tRNAs will leave the ribosome quickly. Correct tRNAs that stay long enough begin to hydrolyze GTP. After hydrolysis, there is another delay, giving bad tRNAs another chance to leave. So hydrolysis of GTP creates another chance for proofreading.
3. It comes from the high-energy ester bond between the peptide and the P-tRNA. Ribosomal RNA and tRNA contribute to catalysis.
4. EF-G in bacteria, eEF2 in eukaryotes. It is bound to GTP, binds to the ribosome after peptide bond formation, hydrolyzes GTP and the ribosome moves to the next location. A site tRNA is now in the P site, and P site tRNA is now in the E site.

5. The amino acid ends of tRNA move in the large subunit. The anticodon regions remain paired, then the large/small subunits rotate relative to each other. Then, the subunits rotate back, moving the mRNA one codon down.
6. When a stable secondary structure of mRNA blocks translocation **AND** there is a "slippery site" that allows tRNAs to re-bind with an alternate frame.
7. Many retroviruses have a frameshift before the *pol* gene encoding reverse transcriptase. About 10% of the time, the ribosome shifts frame and makes RT. The other 90% of the time, the ribosome stops at the end of *gag* (which is right before the *pol* gene).
8. They are proteins that resemble tRNAs. No tRNA anticodons will recognize stop codons. Hence, releasing factors interact with stop codons using favorable *amino-acid-to-codon* H-bonds. They then hydrolyze the ester linkage between tRNA and protein using an OH nucleophile instead of amino acid NH_2
9. The small (30S) subunit plus initiation factors bind at the start codon. The **Shine-Dalgarno** sequence base pairs with the 16S-rRNA (the better it pairs, the more initiation there is), IF-1 binds to A, IF-3 binds to E, and then IF2-GTP binds to IF-1, bringing methionyl-tRNA to the start codon (AUG is in the P site, so the methionyl-tRNA goes into the P site). Remember that the Shine-Dalgarno pairing occurs **BEFORE** initiator tRNA/mRNA start codon start pairing.

Then IF-2 hydrolyzes GTP, and the IF's all leave. Initiator tRNA then enters the P site of the large (50S) subunit. The ribosome has been assembled and is now ready for elongation

10. The methionine is modified on the amino group with an N-formyl group. This N-formyl group cannot attack the carboxyl group of another growing peptide, but its own carboxyl group is open to attack.
11. Cytoplasmic eIF4E binds to the cap, PABP binds to poly-A tail. Other eIFs bring the cap and tail together, resulting in a circularized mRNA.

In a separate process, eIF2-GTP binds tRNA-Met (NOT N-formylmethionine), forming a ternary complex (TC).

TC + small (40S) subunit binds near cap, where eIF4E + PABP + other eIFs are. The small subunit scans along the mRNA to find the right start codon.

When the start codon is found by the initiator tRNA, they base pair. eIF2 hydrolyzes GTP, and the IFs leave. Large subunit then joins with eIF5B-GTP, bringing the two subunits together. eIF5B-GTP then hydrolyzes GTP and leaves. Note that eIF5B is analogous to IF2.

Chapter 12

Translation Regulation and micro-RNAs

12.1 Questions

1. What is the translocon?
2. How is Trp regulation different from lac regulation?
3. What is the second layer of Trp regulation?
4. Anti-terminator vs terminator: which is kinetic/thermodynamic?
5. Can RNAs directly sense other small molecules?
6. How might RNA block translation instead of transcription?
7. How do ribosomes make sure that there is a balance of rRNA and ribosome proteins?
8. How do ribosomes make sure that there is a balance between tRNAs and amino acids?
9. Origin of siRNA/miRNA?
10. Describe how mRNAs might be silenced.

12.2 Key

1. Protein channel for secreted proteins. Proteins exit the ribosome directly into the translocon.
2. *trpEDCBA* is usually "on" but when Trp is present, Trp binds to the *trpR* repressor, and transcription is turned off. The lac operon is the opposite.
3. The *trpL* leader contains a good ρ -independent terminator. There is also an alternate RNA structure within *trpL* that is upstream of the terminator. One half of the terminator stem-loop can form a less-than-perfect secondary structure, and transcription can continue. On top of this, during transcription, *trpL* can be translated. When the ribosome translates *trpL*, it removes secondary structure, freeing up half the anti-terminator to form the terminator. This stops transcription. When Trp is low, translation stalls because the ribosome directly "senses" Trp availability. Thus it does not reach the anti-terminator in time to free half of it. Therefore, translation/transcription continues.
4. The anti-terminator is a kinetic effect (it forms because it is synthesized first). The terminator is a thermodynamic effect (more stable).

5. Yes. They are called riboswitches. One part of the RNA binds directly to a small molecule, and the other part of the RNA (allosterically) becomes a secondary structure that terminates transcription.
6. If a riboswitch forms a hairpin that includes the Shine-Dalgarno sequence, then initiation cannot occur.
7. If there is too much free ribosomal protein (meaning not enough rRNA to pair with it), it binds to its own mRNA, limiting translation.
8. A ribosome with an uncharged tRNA triggers the RelA protein to make ppGpp. ppGpp then inhibits RNA polymerase and reduces rRNA production.
9. (a) siRNA: long perfect duplex. Dicer cleaves the dsRNA, perfect match to the target.
(b) miRNA: Primary micro-RNA transcript, "normal" Pol II transcript (pri-miRNA) is capped/poly-A'd, sometimes even spliced. **Drosha** binds a hairpin within pri-miRNA and cleaves out the hairpin. The cleaved hairpin is a miRNA precursor hairpin (pre-miRNA) and is transferred to Exportin 5/Ran, which is a GTPase. Exportin 5/Ran complex exports the pre-miRNA out of the nucleus. When it cleaves GTP, pre-miRNA is transferred to a dicer, which cleaves the loop off the hairpin, leaving a 5' phosphate and 3' OH (most RNAses leave 5' OH and 3' phosphate). The dicer product is processed by RNA helicase, and one strand (the "true" strand) becomes the guide strand in RISC (RNA-induced silencing complex).
10. Small RNAs bind to a protein called Argonaute (Ago). The RNAs then base-pair with a target mRNA. If the pairing is perfect, then the target mRNA is cleaved and degraded (siRNA, silencing RNA). If the pairing is not perfect, Ago blocks translation initiation and removes the poly-A tail, slower decay. This is miRNA (micro RNAs).

Chapter 13

Protein Degradation and Genetic Engineering

13.1 Questions

1. Describe the structure of the proteasome.
2. Which organisms have similar ubiquitin structures?
3. Describe ubiquitin linkage.
4. Describe ubiquitin linkage in terms of the enzymes that catalyze the reaction.
5. Describe how ubiquitination is involved in hypoxia.
6. Name the important experimental techniques we talked about and how they work

13.2 Key

1. It is symmetric. Middle barrel is the large 20S subunit, composed of two α subunits and a β subunit between them. 19S regulatory lids that control access to the barrel. When the lid recognizes a K48 linked poly-ubiquitin, ATPases unfold the protein and feed it into the barrel.
2. All **eukaryotes** have almost exactly the same sequence.
3. K48 side chain of the previous ubiquitin is linked to the C-terminus of the second ubiquitin through a *iso-peptide bond*. When there are 4 or more ubiquitins linked in this fashion, the proteasome will recognize it is marked for degradation.
4. (a) **Ubiquitin-activating enzyme (E1)** hydrolyzes ATP to link itself to ubiquitin, replacing COO- of ubiquitin with thioester between E1 and ubiquitin.
(b) **Ubiquitin-conjugating enzyme (E2)** also has a cysteine on its end, so it is able to 'swap' out the thioester. E1-SH is released, and E1 is replaced with E2 on the ubiquitin. *E2 is responsible for different types of poly-Ub chains*. Some make K48 chains, some don't.
(c) **Ubiquitin Ligase E3** don't have an enzymatic purpose. They bind E2 and the substrate protein. **Target specificity occurs here**. E3's are specific for certain targets, and they are HIGHLY regulated.
5. In the presence of oxygen, Hypoxia Inducible Factor α (HIF α) is ubiquitylated and degraded. The E3 ligase is VHL (von Hippel-Lindau). In the absence of oxygen, HIF α is not degraded, and it is able to switch on genes for blood vessel growth, RBC production.

6. (a) Protein expression: ligate a gene near a good promoter and Shine-Dalgarno sequence.
- (b) Fusion Protein: Express GST with another protein in the same protein molecule. Then, stick glutathione into a column resin and the protein will stick to the resin when you run the column.
- (c) Use GFP to visualize how much of a protein is present and where it's located. Fuse the GFP gene with proteins in cells, and see where the proteins go. You could also fuse the GFP near a promoter, and then cells that turn on the promoter will glow.
- (d) Gene-targeting by homology. Put a gene with some "broken ends" and the cell will initiate homologous recombination, changing or removing endogenous genes.
- (e) CRISPR: Cas9 with guide sgRNA finds a sequence. It cuts the sequence. Then, repair pathways may introduce insertions/deletions to inactivate the gene, or you can add homologous DNA, leading to homology-directed repair.