4 Basic Molecular Genetic Mechanisms

Review the Concepts

1. Watson-Crick base pairs are interactions between a larger purine and a smaller pyrimidine base in DNA. These interactions result in primarily G-C and A-T base pairing in DNA and A-U base pairs in double stranded regions of RNA. They are important because they allow one strand to function as the template for synthesis of a complementary, antiparallel strand of DNA or RNA.

2. The atoms on the edges of each base within the minor groove of DNA at the binding site for the TATA box–binding protein form a binding surface (see Figure 4-5). Multiple van der Waals interactions and hydrogen bonds form between atoms on the DNA-binding surface of the protein and atoms in the minor groove of DNA. In addition, the ability of DNA to bend along its long axis is important for binding by the TATA box–binding protein. DNA bending is possible because the hydrogen-bonds that give the double helix stability are perpendicular to the long axis of the helix. As a result, the bonds are not stretched greatly when the helix is bent. This is opposite to the situation in a protein α-helix where hydrogen bonds stabilizing the helix are close to parallel to the long axis of the helix. Bending of an α-helix would stretch the hydrogen-bonds on one side of the helix and compress them on the other side of the helix. As a result, a protein α-helix is very stiff compared to a DNA helix.

3. At 90^{O} C, the double-stranded DNA template will denature and the strands will separate. As the temperature slowly drops below the T_m of the plasmid DNA, the single-stranded oligonucleotide primer present at higher concentration than the plasmid DNA strands hybridizes to its complementary sequence on the plasmid template. The resulting molecules contain a short double-stranded stretch the length of the primer with a free 3′ OH that can be used by DNA polymerase enzyme in sequencing reactions.

4. RNA is less stable chemically than DNA because of the presence of a hydroxyl group on C-2 in the ribose moieties in the backbone. Additionally, cytosine (found in both RNA and DNA) may be deaminated to give uracil. If this occurs in DNA, which does not normally contain uracil, the incorrect base is recognized and repaired by cellular enzymes. In contrast, if this deamination occurs in RNA, which normally contains uracil, the base substitution is not corrected. Thus the presence of deoxyribose and thymine make DNA more stable and less subject to spontaneous changes in nucleotide sequence than RNA. These properties might explain the use of DNA as a long-term information-storage molecule.

5. In prokaryotes, many protein-coding genes are clustered in operons where transcription proceeds from a single promoter that gives rise to one mRNA encoding multiple proteins with related functions. In contrast, eukaryotes do not have operons but do transcribe intron sequences that must be spliced out of mature mRNAs. Eukaryotic mRNAs also differ from their prokaryote counterparts in that they contain a 5′ cap and 3′ poly(A) tail. Also, ribosomes have immediate access to nascent mRNAs in bacteria so that translation begins as the mRNA is being synthesized. In contrast, in eukaryotes, mRNA synthesis occurs in the nucleus, whereas translation by ribosomes occurs in the cytoplasm. Consequently, only fully synthesized and processed mRNAs are translated in eukaryotes.

6. A simple explanation is that the larger, membrane-spanning domain-containing protein and the small, secreted protein are encoded by the same gene which is differentially spliced. Specifically, the final exon of the gene could contain the information for the membrane-

spanning domain, and in the smaller, secreted protein, this exon could be omitted during splicing.

7. An operon is an arrangement of genes in a functional group that are devoted to a single metabolic purpose. In the case of tryptophan synthesis the DNA for five genes is arranged in a contiguous array that gets transcribed from a single promoter into a continuous strand of mRNA encoding five proteins. In this manner, the cell simply has to induce one promoter, which transcribes all the necessary genes encoding the proteins (enzymes) to make the amino acid tryptophan. Splicing out intronic sequences or transcribing multiple mRNAs from genes on different chromosomes, as seen in eukaryotic systems, is unnecessary; thus operons are a logical way to economize on the amount of DNA needed by genes to encode a number of proteins. In addition, this arrangement allows all the genes in an operon to be coordinately regulated by controlling transcription initiation from a single promoter.

8. In prokaryotes, the eight nucleotide Shine-Dalgarno sequence located near the AUG start codon base pairs to a specific sequence near the 3′-end of the 16S rRNA in the small ribosome unit, positioning the start codon at the proper site in the small subunit. In eukaryotes, recognition of the start site involves other factors such as eIF4E that binds to the 5′ mRNA cap structure, other eIF4 subunits, eIF3 proteins, and Kozak sequences near the mRNA start site. eIF4E recognizes and binds to the 5′ cap structure on eukaryotic mRNAs, and eIF3 proteins and other eIF4E subunits are part of the preinitiation complex that is thought to scan along the mRNA, most often stopping at the first AUG. The Kozak sequences facilitate choice of the proper start site. Poliovirus initiates translation of its uncapped mRNA via an internal ribosome entry site (IRES), an RNA sequence of several hundred bases that folds into a three dimensional RNA structure that binds to the ribosome E and P sites, placing the AUG start codon at the proper position in the ribosome P site for translation initiation.

9. Even if all proteins are removed from the 23S rRNA, it is capable of catalyzing the formation of peptide bonds. X-ray crystallographic studies support this by revealing that no proteins lie near the site of peptide bond synthesis in the crystal structure of the large subunit.

10. Since poly(A)-binding protein I is involved in increasing the efficiency of translation, a mutant in poly(A)-binding protein I would have less efficient translation. Polyribosomes from such a mutant would not contain circular structures of mRNAs during translation because lack of the poly(A)-binding protein I would eliminate the 3′ binding site for eIF4G.

11. DNA synthesis is discontinuous because the double helix consists of two antiparallel strands and DNA polymerase can synthesize DNA only in the 5′ to 3′ direction. Thus one strand is synthesized continuously at the growing fork, but the other strand is synthesized utilizing Okazaki fragments that are joined by DNA ligase.

12. Base excision repair is responsible for repairing guanine-thymine mismatches caused by the chemical conversion of cytosine to uracil or by deamination of 5-methyl cytosine to thymine. Mismatch excision repair eliminates base pair mismatches and small insertions or deletions of nucleotides generated accidentally during DNA replication. Nucleotide excision-repair fixes DNA strands that contain chemically modified bases, which ensures that thymine-thymine dimers are repaired in the case of UV light damage.

13. UV irradiation causes thymine-thymine dimers. These are usually repaired by the nucleotide excision-repair system, which utilizes XP complexes and the transcriptional helicase TFIIH to unwind and excise the damaged DNA. The gap is then filled in by DNA polymerase. Ionizing radiation causes double-stranded breaks in DNA. Double-stranded breaks are repaired either by homologous recombination or nonhomologous DNA end-joining. Homologous recombination requires the BRCA1, BRCA2 and Rad51 proteins to use the sister chromatid as template for error-free repair. Nonhomologous DNA end-joining is error-prone because nonhomologous ends are joined together. Since formation of a malignant tumor requires multiple mutations, cells that have

lost DNA-repair function are more likely to sustain cancer-promoting mutations. Examples are xeroderma pigmentosum due to mutations in XP genes that prevent repair of thymine dimers and a genetic predisposition to breast cancer in individuals with germ-line mutations in the BRCA1 or BRCA2 genes.

14. Homologous recombination is the process that can repair DNA damage and also generate genetic diversity during meiosis. In both cases, repair is to double-strand breaks, RecA/Rad51-like proteins play key roles in the recombination process, and Holliday structures form, followed by cleavage and ligation to form two recombinant chromosomes. During DNA repair by homologous recombination, the damaged sequence is copied from an undamaged copy of the homologous DNA sequence on the homologous chromosome or sister chromatid. During meiosis, however, genetic diversity is generated by homologous recombination where large regions of chromosomes are exchanged between the maternal and paternal pair of homologous chromosomes. Also, in meiosis an exchange called crossing over is required for the proper segregation of the chromosomes during the first meiotic cell division.

15. The gene encoding the reverse transcriptase enzyme is unique in retroviruses and closely related retrotrasposons. These viruses contain RNA as their genetic material; a DNA copy of the viral RNA is made during infection and reverse transcriptase catalyzes this reaction. The human T-cell lymphotrophic virus, which causes T-cell leukemia, and human immunodeficiency virus, which causes AIDS, can infect only specific cell types because these cells possess receptors that interact specifically with viral envelope proteins of the progeny virus.

Analyze the Data

a. The data suggest that context matters, and that a change in the sequence surrounding the first AUG affects the efficiency of initiation from this start site. When comparing lanes 1 and 2, in which the mRNAs differ only at position (+4), one observes that a G, rather than a U, at this position reduces leaky scanning. More preCAT and less CAT is synthesized with the message used in lane 2 than with that used in lane 1. Although (–3)ACC**AUG**G(+4) is hypothesized to provide an optimal context in which the first AUG is presented, the data suggest that this sequence can be modified to ACC**AUG**A without significantly compromising efficiency of initiation from the first AUG. The only difference between the messages in lanes 3 and 4 is a change from G to A at position (+4), and in each case only preCAT is synthesized. With respect to the importance of ACC at positions (–3) to (–1), a comparison of lanes 4 and 5 reveal that a shift of ACC from position (–3) to (–1) (lane 4) to position (–4) to (–2) (lane 5) results in a loss of fidelity of initiation from the first start site. Replacement of (–3)ACC(–1) with (–3)UUU(–1) (compare lane 2 to lane 3) results in a loss of efficiency of translation from the first start site, so these data suggest that the ACC sequence and its position relative to the AUG matter.

b. In order to further test the importance of the nucleotide at the +4 position, it would be useful to undertake an analysis of CAT mRNA mutants with only single substitutions compared to the wild-type sequence shown in lane 3. The data shown provide evidence that G can be substituted with A (lane 4) at position (+4) and suggest that it cannot be substituted with U (lane 1). However, there are additional sequence changes in the mRNA in lane 1 other than the change at position (+4). It would be informative to change the mRNA used in lane 3 so that either U or C is substituted for G at position (+4). If each of these substitutions results in synthesis of some CAT, then one could deduce that an optimal context for the first start site can tolerate purines but not pyrimidines at the (+4) position. The data shown in the figure also do not examine the importance of the (–3)ACC(–1) sequence other than to change it completely to UUU. It would also be useful to change this sequence one nucleotide at a time to determine the relative importance of each of these nucleotides in helping the ribosomes pause at the first AUG site and begin translation. In each case, synthesis of CAT would be evidence that the particular substitution to the sequence results in a loss of fidelity of initiation at the first AUG. If the A at position (–3) is the most

important of the ACC sequence for generating efficiency of translation for the first AUG, then one would expect that changes to this nucleotide would result in synthesis of more CAT than would changes to the other two nucleotides.

c. The mutation in this family results in the introduction of a new AUG sequence in the hepcidin mRNA upstream from the original start site. Because the new AUG is not in frame with respect to the original start AUG, no hepcidin will be made if initiation begins exclusively at this new, upstream AUG. The fact that hepcidin is not synthesized in individuals who have inherited this mutation suggests that initiation of protein synthesis occurs at the new AUG with high efficiency and that the ribosomes do not scan through this site to begin synthesis downstream at the original start site. Thus, these findings support the hypothesis that initiation of protein synthesis in eukaryotes normally begins at the first 5′ AUG site. An examination of the context in which this new AUG start site is located reveals that it has the important G located at (+4), as does the original AUG start site. Whereas the original AUG has an A at the (−3) position (and, in fact, also has the consensus C at position (−2)), the new, upstream AUG does not have the consensus sequence in this position. However, given that no hepcidin is made, the new start site may be in a context that facilitates efficient recognition by the ribosomes and thereby does not result in any (or in any detectable) leaky scanning.