C1. Structural genomics is the study of genome composition. Researchers attempt to map all of the genes in the genome and ultimately to determine the sequence of all the chromosomes. Functional genomics attempts to understand how genetic sequences function to produce the characteristics of cells and the traits of organisms. Much of functional genomics is aimed at an understanding of gene function. However, it also tries to understand the roles of other genetic sequences such as centromeres and repetitive sequences. Proteomics focuses on the functions of proteins. The ultimate goal is to understand how groups of proteins function as integrated units.

C2. There are two main reasons why the proteome is larger than the genome. The first reason involves the processing of pre-mRNA, a phenomenon that occurs primarily in eukaryotic species. RNA splicing and editing can alter the codon sequence of mRNA and thereby produce alternative forms of proteins that have different amino acid sequences. The second reason for protein diversity is posttranslational modifications. There are many ways that a given protein's structure can be covalently modified by cellular enzymes. These include proteolytic processing, disulfide bond formation, glycosylation, attachment of lipids, phosphorylation, methylation, and acetylation, to name a few.

C3. A database is a collection of many computer files in a single location. These data are usually DNA, RNA, or protein sequences. The data come from the contributions of many research labs. A major objective of a genome database is to organize the genetic information of a single species. A genome database will identify all of the known genes and indicate the map locations within the genome. In addition, a genome database may have other types of information such as a registry of workers and details that are specific to the particular species.

C4. Centromeric sequences, origins of replication, telomeric sequences, repetitive sequences, and enhancers. Other examples are possible.

C5. Sequence recognition involves the recognition of a particular sequence of an already-known function. For example, a program could locate start codons (ATG) within a DNA sequence. By comparison, pattern recognition relies on a pattern of arrangement of symbols but is not restricted to particular sequences.

C6. There are a few interesting trends. Sequences 1 and 2 are similar to each other, as are sequences 3 and 4. There are a few places where amino acid residues are conserved among all five sequences. These amino acids may be particularly important with regard to function.

C7. In genetics, the term *similarity* means that two genetic sequences are similar to each other. Homology means that two genetic sequences have evolved from the same ancestral sequence. Homologous sequences are similar to each other, but not all (short) similar sequences are due to homology.

C8. A. Correct.

B. Correct.

C. This is not correct. These are short genetic sequences that happen to be similar to each other. The *lac* operon and *trp* operon are not derived from the same primordial operon.

D. This is not correct. The two genes are homologous to each other. It is correct to say that their sequences are 60% identical.

C9. A gap is necessary when two homologous sequences are not the same length. Since homologous sequences are derived from the same ancestral gene, two homologous sequences were originally the same length. However, during evolution the sequences can incur deletions and/or additions that make the sequences longer or shorter than the original ancestral gene. If one gene incurs a deletion, a gap will be necessary in this gene's sequence in order to align it with a homologous gene. If an addition occurs in a gene's sequence, a gap will be necessary in a homologous gene sequence in order to align the two sequences.