**Chapter Review**

**Chapter 5: Genomics, Proteomics, and Systems Biology**

5.1

Bacterial and yeast genomes are compact, with protein-coding sequences accounting for most of the DNA. The *S. cerevisiae* genome contains about 6000 genes. Protein-coding sequences account for 10–25% of the genomes *C. elegans*, *Drosophila*, and *Arabidopsis*, which contain approximately 19,000, 14,000, and 26,000 genes, respectively. The human genome contains approximately 20,000 protein-coding genes—not much more than the number of genes found in simpler animals like *Drosophila* and *C. elegans*, and fewer than in *Arabidopsis* and other plants, emphasizing the lack of relationship between gene number and complexity of an organism. Enormous progress in the technology of DNA sequencing has now made it feasible to determine the complete sequence of individual genomes and of all the RNAs expressed in a cell.

5.2

Characterization of the complete protein complement of cells is a major goal of proteomics. Mass spectrometry provides a powerful tool for protein identification, which can be used to identify either isolated proteins or proteins present in mixtures. The protein compositions of subcellular organelles can be analyzed by mass spectrometry or large-scale immunofluorescence. The purification of protein complexes from cells and analysis of interactions of proteins introduced into yeast can identify interacting proteins and may lead to elucidation of the complex networks of protein interactions that regulate cell behavior.

5.3

Systems biology uses large-scale datasets for quantitative experimental analysis and modeling of biological systems, including genome-wide screens of gene function, regulation of gene expression, and quantitative modeling of regulatory networks. Synthetic biology is an engineering approach to designing new biological systems, including genetic circuits, metabolic pathways, and synthetic genomes.