Glossary

1D gel (one-dimensional gel electrophoresis): See *electrophoresis*.

2D gel (two-dimensional electrophoresis or 2DE): See *electrophoresis*.

16S rRNA: RNA molecule (about 1500 nucleotides long) that combines with proteins to form the small subunit of the ribosome in prokaryotes. The gene for 16S rRNA is well studied, highly conserved, and present in all prokaryotic organisms, so variations in this gene's sequence can be used to determine relatedness (phylogenetic linkages) among prokaryotes. In eukaryotes, the small subunit of the ribosome contains an RNA molecule called 18S rRNA (about 1900 nucleotides long), which is analyzed to determine phylogenetic relationships.

ab initio: Type of method for predicting protein structure using first principles of physics and chemistry rather than comparisons with known homologous structures. Also used to describe gene-prediction methods based on analyzing the composition of raw genome sequence rather than using comparisons with other sequence data.

accurate mass and time tag (AMT): Peptide for which the liquid chromatography elution time and mass has been measured so accurately that it can be identified uniquely among all possible peptides predicted from a genome. In proteome mass spectrometry (MS), the proteome is digested enzymatically to produce protein fragments or peptides before MS analysis. A database of AMT mass data can be used to identify peptides in other samples analyzed by MS.

activator: Regulatory protein that binds the operator site and enhances transcription of genes in an operon. See also *repressor*.

adenine (A): Nitrogenous base, one member of the base pair AT (adenine-thymine) in DNA.

ADP (adenosine diphosphate): Molecule consisting of a nitrogenous base adenine linked to a ribose with two phosphate groups. Cells use solar energy from photosynthesis or the energy from oxidation of chemical compounds to synthesize ATP (adenosine triphosphate) by adding a phosphate group to ADP. Relative concentrations of ADP vs ATP control electron-transfer processes in cells.

affinity chromatography: Method for isolating or purifying a target molecule from a mixture by applying the mixture to a column containing immobilized ligands known specifically to bind the target. After nontargeted molecules in the mixture have passed through the column, it is washed with a solution that releases the target molecule from the immobilized ligand, allowing that molecule to be collected in a purified form.

affinity reagent: Antibody, peptide, nucleic acid, or other small molecule that specifically binds a target molecule of interest. Affinity reagents can be used to identify, track, capture, and influence the activity of larger proteins and molecular complexes in living systems.

agonist: Molecule that enhances the activity of another molecule.

algorithm: Formal set of instructions that tells a computer how to solve a problem or execute a task. A computer program typically consists of several algorithms.

alkylation: Modification of a molecule by adding a methyl ($-CH_3$), ethyl ($-CH_2CH_3$), or other alkyl group.

amensalism: Relationship in which one species inhibits the survival of another.

amino acid: Organic compound containing an amino group ($-NH_2$) on one end and a carboxyl group ($-COOH$) on the other. Any of 20 different amino acids are linked together in a linear fashion to form peptides or proteins. The sequence of amino acids in a protein, and hence protein function, are determined by the nucleotide sequence of genes.

analyte: Chemical substance to be experimentally measured.

ancillary pathway: Secondary biochemical pathway that supports a primary pathway of interest.

angstrom (C): One-tenth of a nanometer (10^{-10} meter).

annotation: Addition of biologically meaningful descriptions to data (e.g., by labeling regions of sequence data that encode a gene or regulatory region or by identifying the active site of a protein structure).

anoxic: Without oxygen.

antagonist: Molecule that interferes with the action of another.

anthropogenic: Resulting from human activity.

antibody: Protein molecule synthesized as part of the immune response in vertebrate animals. When an animal's cells recognize a substance as a foreign invader (the antigen), they produce an antibody capable of binding the invading molecule. Antibodies specific to a protein of interest can be synthesized and used as reagents to detect or isolate the protein.

aptamer: Short DNA segment that can fold into a structural shape that specifically binds to another target molecule not a nucleic acid (e.g., proteins or small molecules such as NADH or ATP).

Archaea: One of the three domains of life (along with Bacteria and Eukarya) distinguished through DNA sequence analysis. Archaea are structurally and metabolically similar to bacteria but share some features of their molecular biology with eukaryotes.

assimilation: Process of taking up essential elements (e.g., carbon, nitrogen, phosphorous) and converting them to biologically useful forms.

atomic force microscopy (AFM): Type of scanning probe microscopy that generates images with molecular and atomic detail by moving a probe over the surface of a biological structure. Any change in the probe's vertical position as it follows the structure's contour is detected by deflection of a laser beam pointed at the probe's tip.

atomic resolution: Level of resolution for a molecular structure that involves identifying the specific position of every atom in 3D space. Nuclear magnetic resonance spectroscopy and X-ray crystallography are used to determine molecular structures with atomic resolution.

ATP (adenosine triphosphate): Important energy carrier of all living cells. An ATP molecule consists of a nitrogenous base (adenine) linked to a ribose with three phosphate groups. Cleavage of ATP's terminal phosphate group yields ADP (adenosine diphosphate), inorganic phosphate, and the energy used to power cellular processes. Relative concentrations of ATP vs ADP control electron-transfer processes in cells.

attomole: Unit quantifying the amount of a chemical substance; equal to 10^{-18} mole where a mole represents 6.022 x 10^{23} items (e.g., molecules, atoms).

Bacteria: One of the three domains of life (along with Archaea and Eukarya) distinguished through DNA sequence analysis. Also a general term (sing., bacterium) referring to prokaryotic organisms that do not belong to the Archaea domain.

bacteriorhodopsin: Transmembrane protein that acts as a light-driven proton pump involved in ATP synthesis. Bacteriorhodopsins have been found in microorganism known to tolerate high levels of salinity and resemble light-sensitive rhodopsin proteins in the retinas of animals.

base pair (bp): Pair of weakly bonded nitrogenous bases (either adenine and thymine or guanine and cytosine) that hold together the two complementary DNA strands of a double helix.

biochemical characterization: Use of a variety of techniques to determine a protein's mechanism of action or biochemical function [e.g., its affinity for substrates and inhibitors, how it chemically modifies a substrate, how cofactors determine its mechanism of action, and how quickly it catalyzes a reaction (kinetics)].

biodiesel: Renewable alternative to petroleum diesel fuel; synthesized from the lipids of soybeans and plant materials, animal fats, and other biological sources.

biodiversity: Range of species living together in a particular environment.

bioethanol: Ethanol derived from biomass.

biofilm: Community of microorganisms living together on a surface and embedded in extracellular polymers they create.

biofuel: Liquid, solid, or gaseous fuel derived from renewable biomass. Biological materials can be used to produce such fuels as biodiesel, ethanol, methanol, methane, and hydrogen.

biogeochemistry: Study of how interactions among biological and geochemical processes influence the global cycling of such essential elements as carbon, nitrogen, phosphorous, and sulfur.

biohydrogen: Molecular hydrogen (H_2) gas generated from biological processes.

bioinformatics: Science of managing and analyzing biological data using advanced computing techniques.

biological pump: Collection of biological ocean processes that regulate the uptake, storage, and release of carbon.

biomarker: Chemical substance that can be used to detect the presence of a particular organism or biochemical activity in the environment.

biomass: Organic material from living organisms, typically plant matter including trees, grasses, and agricultural crops, that can be burned or converted to liquid or gaseous fuels for energy.

biomineralization: Process in which living organisms transform a substance into a mineral.

biomolecule: Molecule synthesized by living systems, including nucleic acids (DNA, RNA), proteins, lipids, carbohydrates, and metabolites.

biophotolysis: Biological process observed in green algae and cyanobacteria that can generate hydrogen from the photosynthetic splitting of water.

biophysical characterization: Use of a variety of analytical techniques to determine a molecular machine's composition and structure.

biophysics: Application of physical principles to the study of biological structures and processes.

bioreactor: Vessel in which biocatalysts or microorganisms involved in the production of some desired biological product are maintained. In industry, bioreactors typically house fermentation reactions and are called fermenters.

bioremediation: Use of biological organisms such as plants or microbes to degrade or chemically transform hazardous substances that have been released into the environment.

biosensor: Device that uses biological material (e.g., microorganisms, oligonucleotides, enzymes, antibodies) to detect other biological molecules or chemicals.

biosphere: Portion of earth and its atmosphere that supports life.

biotin: Water-soluble B vitamin that can be used to label macromolecules via a chemical reaction known as biotinylation. Molecules labeled with biotin are mixed with avidin proteins (from egg whites) labeled with a fluorescent compound or other reporter molecule. Avidin binds biotin very tightly, thus labeling the biotinylated molecule.

BLAST: Computer program that identifies homologous (similar) genes in different organisms by comparing all sequence data available in public databases.

calorimetry: Measurement of heat released or taken up in a chemical reaction or physical process to derive thermodynamic data (e.g., dissociation constant, free-energy change, enthalpy change, entropy change) for molecular interactions. Two kinds of calorimetry important to biomolecular studies are isothermal titration calorimetry (ITC), which can be used to detect the number of binding sites on an enzyme; and differential scanning calorimetry (DSC), which monitors the energetics of conformational changes in proteins.

capillary electrophoresis (CE): Rapid, high-resolution electrophoresis technique that separates molecules by applying an electric current to a narrow tube (<1 mm in diameter) filled with a liquid or gel.

carbohydrate: Organic compound containing carbon, hydrogen, and oxygen; most simple carbohydrates or sugars contain three to seven carbons with a chemical composition represented by the general formula $(CH_2O)_n$. Many sugars can be linked together as linear or branched chains known as polysaccharides. Carbohydrates play key roles in a variety of cell functions including energy storage, structural support, and chemical modification of proteins and lipids.

carbon cycle: The global flow of carbon from one reservoir (carbon sink) to another. Each carbon exchange among reservoirs is mediated by a variety of physical, biogeochemical, and human activities.

carbon dioxide (CO_2): Gas that is an important part of the global carbon cycle. CO_2 is emitted from a variety of processes (e.g., cellular respiration, biomass decomposition, fossil-fuel use) and taken up primarily by the photosynthesis of plants and microorganisms. CO_2 is a greenhouse gas that absorbs infrared radiation and traps heat in the earth's atmosphere.

carbon fixation: Conversion of inorganic carbon dioxide to organic compounds by photosynthesis.

carbon flux: Rate of carbon movement as it flows from one carbon reservoir to another in the global carbon cycle, usually expressed in gigatons of carbon per year (GtC/yr).

carbon free: Describes an energy source that releases no carbon during its production and use.

carbon neutral: Describes an energy source that introduces no additional carbon to the global carbon cycle. For example, carbon dioxide released from the consumption of biofuels is recaptured by photosynthesis, which generates additional biomass.

carbon sequestration: Biological or physical process that captures carbon dioxide and converts it into inert, long-lived, carbon-containing materials.

carbon sink: Region of the earth that takes up carbon as it moves through the carbon cycle. Four main carbon sinks include the atmosphere, terrestrial environments, oceans, and sediments.

catalyst: Substance that speeds up a chemical reaction without being altered by the reaction.

CD: See *circular dichroism*.

cell-based expression system: Protein-production technique in which genes encoding proteins needed for analysis are introduced into living host cells (e.g., *E. coli*, yeast). The host's cellular machinery synthesizes the proteins, which then can be harvested from the cells and analyzed.

cell-free expression system: Protein-production technique that uses cell lysates (typically from *E. coli* or wheat germ) containing the molecular machinery needed to synthesize proteins. DNA segments encoding the proteins of interest are added to the cell lysate mixture, and the proteins are synthesized in vitro.

cell lysate: Inner contents of a cell released by rupturing the cell membrane.

cellulase: Enzyme involved in the conversion of cellulose to simple glucose molecules. Different types of cellulases work together as a cooperative system to carry out cellulose breakdown. The three main classes of cellulases are endoglucanases, exoglucanases, and cellobiases.

cellulolytic: Having the ability to hydrolyze or break down cellulose into carbohydrate subunits.

cellulose: Large, complex polysaccharide that provides structural support to plant cell walls and is synthesized by some bacteria. Each cellulose molecule is a linear chain of thousands of glucose subunits. Cellulose is the most abundant form of carbon in the biosphere.

centrifugation: Spinning of cells, proteins, or other particles in a centrifuge to separate them from the solution in which they are suspended. The centrifugal force from spinning causes the cells or proteins to form a pellet at the bottom of the sample tube. The pellet then can be separated from the solution.

chaperone: Type of protein that ensures proper folding of other proteins into functional, 3D structures in cells; also called chaperonins.

chemostat: Apparatus for the continuous cultivation of bacteria. Chemostats keep bacterial cultures in an optimal growth state by continually adding media and removing old cells.

chromatography: Method for separating mixtures of chemical compounds. In one form, liquid chromatography, a mixture is dissolved in a solvent and applied to or passed through an adsorbent solid material. Chemical compounds migrate through the solid material at different rates, thus separating the mixture's components. Other types include affinity, size-exclusion, gas, and high-performance liquid chromatography.

chromophore: Light-absorbing pigment that gives color to a molecule.

chromosome: Self-replicating molecular structure that contains an organism's genome. In most prokaryotes, the entire genome is packaged into a single chromosome consisting of a circular DNA molecule. Eukaryotic genomes are packaged into several different chromosomes, each consisting of a linear DNA molecule wrapped around proteins.

circular dichroism (CD): Spectroscopy technique that provides structural information about molecules such as proteins and peptides. Elements of asymmetry in proteins produce characteristic CD signals in the far UV region (190 to 250 nm) of the electromagnetic spectrum that can be used to determine how much of a protein is made up of alpha-helices, beta-sheets, or random coils. CD signals from the near-UV spectral region (250 to 350 nm) can be used to determine if a protein is folded into a well-defined structure or if protein-protein interactions or changes in environmental conditions cause conformational changes in a protein's tertiary structure.

climate model: Mathematical model used to understand, simulate, and predict climate trends by quantitatively analyzing interactions between the earth and its atmosphere.

clone: Exact copy of biological material such as a DNA segment (e.g., gene or other region), whole cell, or complete organism. Gene clones inserted into cloning vectors are used to produce proteins for laboratory analysis.

cloning: Technique used to produce multiple, exact copies of a single gene or other segment of DNA to obtain enough material for further study.

cloning vector: Self-replicating DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which a DNA fragment of interest is inserted. Vectors transfer DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, and yeast artificial chromosomes; vectors often are recombinant molecules containing DNA sequences from several sources.

cluster computing: Linking of many smaller, less expensive computers to obtain the throughput and computing power of a larger, more expensive machine; redundancy in the cluster provides greater protection from system failure. See also *grid computing.*

codon: Set of three consecutive nucleotides in mRNA that specify a particular amino acid in the protein synthesized during translation; a codon also may signal the beginning or end of the message to be translated (i.e., start codon, stop codon). See also *genetic code.*

codon bias: Preference for the use of certain codons by different organisms. Codon bias presents a problem for heterologous expression in which a gene rich in one type of codon is inserted into a host cell that rarely uses that codon, so there may not be enough of the corresponding tRNA to synthesize the protein.

cofactor: Small, nonprotein substance required for enzyme activity.

coimmunoprecipitation: Technique that uses antibodies to detect interacting proteins. An antibody that specifically binds a target protein is added to a cell lysate. The antibody forms a complex with its target and any protein or molecule bound to the target. Then an antibody-binding protein immobilized on a tiny bead is added and used to pull the antibody-protein complex out of solution.

colicin: Protein, secreted by certain strains of bacteria, that kills but does not lyse other strains.

colony: Cluster of cells originating from a single cell and growing together on a solid medium.

commensalism: Relationship in which only one party obtains some advantage.

community: All the different species of organisms living together and interacting in a particular environment.

comparative genomics: Field of study that compares DNA sequences of genes and genomes from different organisms to predict functions of newly discovered genes and gain insights into phylogenetic relationships among organisms.

competition: Relationship between two populations in which each is adversely affected by the other.

complementary sequence: Nucleic acid base sequence that can form a double-stranded structure with another DNA fragment by following base-pairing rules (A pairs with T and C with G). The complementary sequence to GTAC, for example, is CATG.

confocal microscopy: Type of microscopy that focuses a beam of light onto a fluorescently labeled specimen. As the laser scans the specimen within a narrow plane (<1 μm thick), light emitted from the excited fluorescent dye passes through a pinhole in a screen before reaching the light detector. The pinhole helps generate higher-resolution images by preventing out-of-focus light rays from reaching the detector and blurring the image. A computer digitizes these optical sections and develops a 3D representation of the specimen. Confocal microscopy is useful for viewing organisms that live at different depths within a biofilm. Also known as confocal scanning laser microscopy (CSLM) or laser scanning confocal microscopy (LSCM).

contaminant fate and transport model (transport model, fate model): Computer model that uses experimental data and known properties of subsurface constituents such as minerals to simulate groundwater conditions and predict how contaminants will move through and be chemically transformed by physical, chemical, and biological factors.

contaminant plume: Zone of contamination in soil, sediments, water, or air that originated from a point source.

cross-linker: Chemical group that forms a crosswise covalent connection between two parallel chains of a molecular complex.

cryoelectron microscopy (cryoEM): Type of electron microscopy (EM) that involves freezing samples to allow generation of high-resolution, 3D images of biological structures in their native, hydrated forms. Samples are

dipped in liquid ethane and chilled with liquid nitrogen in the electron microscope; samples are not stained or dried out, thus eliminating distortion associated with other EM techniques. CryoEM visualizes molecular complexes too large for nuclear magnetic resonance spectroscopy and X-ray crystallography (techniques that yield structural data with atomic resolution) and too small for conventional EM. CryoEM data are detailed enough to be used for molecular modeling.

crystallization: Formation of crystals (solid structures with highly ordered, three-dimensional, regularly repeated arrangements of atoms, ions, or molecules).

culturable: Cells capable of being grown on or in prepared media in the laboratory.

culture: Process of growing cells in the laboratory; the mass of cells produced during cultivation.

cyanobacteria: Division of bacteria capable of oxygen-producing photosynthesis and found in many environments including oceans, freshwater, and soils. Cyanobacteria contain chlorophyll *a* and other photosynthetic pigments in an intracellular system of membranes called thylakoids. Many cyanobacterial species also are capable of nitrogen fixation.

cysteine (cys): One of the amino acids linked together to form proteins. Cysteine is unique among all amino acids and important to protein structure because it contains a sulfhydryl group (-SH), which can form a disulfide bond with another cysteine.

cytochrome: Any of a family of iron-containing proteins that can serve as electron acceptors or donors in the electron-transfer reactions of cells.

cytoplasm: Liquid matrix, enclosed by the cell membrane, in which all inner contents of a cell are suspended.

cytosine (C): Nitrogenous base, one of the base pair GC (guanine and cytosine) in DNA.

Dalton (Da): Unit of molecular mass equal to 1/12 the mass of a ^{12}C atom and typically used in the life sciences to describe the mass of large biomolecules.

data mining: Data-analysis techniques used to sift through large amounts of data and identify hidden patterns and relationships.

data model: Logical structure for representing data associated with a particular concept and relating it to other data in a database.

data standard: Set of specifications, established by community consensus or authorized by an official standards organization, for representing and organizing data in ways that promote the exchange, comparison, and integration of different data sets.

decrystallization: Breakdown of a solid, crystalline structure. To produce ethanol from cellulose, biomass must be pretreated with chemicals or steam to decrystallize or disrupt the highly ordered crystalline structure of cellulose and make the cellulose fibers more accessible to degradation by enzymes.

denaturation: Disruption of the native structures of proteins and nucleic acids that can be caused by increases in temperature, changes in pH, or exposure to certain chemicals. Proteins unfold and collapse into random coils, which results in loss of function; denaturation of DNA causes the two strands of the double helix to separate.

desorption: Removal of a substance that has permeated or attached to the surface of another substance; the opposite of absorption or adsorption.

detection limit: Lowest number or concentration of a particular kind of atom or molecule that can be detected by an analytical instrument or technique.

detergent: Chemical substance that contains both water-soluble (hydrophilic) and water-insoluble (hydrophobic) portions and can be used to solubilize proteins.

deuterium: Heavy isotope of hydrogen in which the nucleus contains one proton and one neutron. Also called heavy hydrogen, it is given the symbol ^{2}H or D. The most common form of hydrogen has one proton but no neutron in its nucleus.

dinoflagellate: Any of a group of eukaryotic microorganisms containing both plant-like and animal-like species that lives in marine and freshwater environments. These unicellular microorganisms use a pair of dissimilar cellular appendages called flagella for motility.

diatom: Type of microscopic, photosynthetic algae known for its intricately designed, silica-containing shell. Thousands of diatom species are known; most are unicellular, but some form colonies. Diatoms are responsible for a large portion of photosynthetic carbon assimilation in marine and freshwater environments.

direct CO_2 injection: Carbon-sequestration technique in which carbon dioxide is injected directly into the ocean depths.

directed evolution: Laboratory process used on isolated molecules or microbes to cause mutations and identify subsequent adaptations to novel environments.

directed mutagenesis: Alteration of DNA at a specific site and its subsequent reinsertion into an organism to study any effects of the change.

discovery-driven science: Research paradigm focused on creating resources and infrastructure to facilitate and advance hypothesis-based research. The Human Genome Project is an example of discovery-driven science undertaken to provide the scientific community with resources (sequence data, computational tools, technologies) to enable the pursuit of new hypothesis-based investigations. See also *hypothesis-driven science*.

disulfide bond: Structurally important covalent bond in protein complexes that can form between cysteine residues within the same or different polypeptide chains.

DNA (deoxyribonucleic acid): Molecule that encodes genetic information. DNA is a double-stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus, the base sequence of a single strand can be deduced from that of its partner.

DNA sequence: Relative order of base pairs in a DNA fragment, gene, chromosome, or entire genome.

docking: See *molecular docking*.

domain: Discrete portion of a protein with its own function; the combination of domains in a single protein determines its overall function. Alternately, "domain" may refer to one of the three main categories of living organisms (Archaea, Bacteria, and Eukarya), whose distinctions are based on DNA sequence analysis.

dynamic range: Range of concentrations that an instrument is capable of measuring.

ecogenomics: Approach for determining the genetic potential, structure, and functional capabilities of a natural microbial community by sequencing and analyzing DNA samples isolated from the environment.

ecophysiology: Study of the physiological functions of organisms as they pertain to their ecology or interactions with each other and their environment.

ecosystem: Set of living organisms (plants, animals, fungi, and microorganisms) and the physical and chemical factors that make up a particular environment.

electron acceptor: Substance that gains electrons from another substance in an oxidation-reduction reaction.

electron donor: Substance that loses electrons to another substance in an oxidation-reduction reaction.

electron microscopy (EM): Technique that uses electrons instead of light to obtain images of organelles or other structural components within cells. Imaging with electrons usually requires that a sample is analyzed in a vacuum, so living specimens cannot be visualized directly with EM. In an electron microscope, magnets and electrically charged surfaces are used to direct electrons toward a sample. As electrons pass through or are reflected by the sample, they are detected by a screen or camera that generates an image.

electron-transport chain: Series of membrane-bound proteins that receive electrons released from the oxidation of organic and inorganic compounds and mediate a sequence of electron-transfer reactions involved in the synthesis of ATP.

electrophoresis: Method of separating large molecules (such as DNA fragments or proteins) in a sample. An electric current is passed through a medium containing the sample; each molecule travels through the medium at a different rate, depending on its electrical charge, shape and size. Agarose and acrylamide gels are commonly used media for electrophoresis of proteins and nucleic acids. In one-dimensional (1D) gel electrophoresis, proteins and nucleic acids are separated in one direction on a gel, primarily by size. Two-dimensional (2D) gel electrophoresis is used in proteome analyses to separate complex protein mixtures using two separation planes (e.g., vertically down a gel by net charge and horizontally by molecular mass). Each unique protein mixture produces a characteristic pattern or fingerprint of protein separation on a 2D gel.

electrospray ionization (ESI): Method used to charge analytes as they transition from a liquid to a gaseous state. Analytes dissolved in a volatile liquid solvent are passed through a fine needle. A high voltage is applied to the analytes as they exit the needle, forming a fine mist of charged analytes (ions). Once the droplets of liquid solvent have evaporated from the ions, the ions are transported by a neutral carrier gas into the mass analyzer of a mass spectrometer.

ELSI: Ethical, legal, and social implications or issues relevant to new scientific-research initiatives.

endogenous: Originating from within a cell or organism.

environmental remediation: Removal from or immobilization of hazardous substances in a contaminated environment.

enzyme: Protein that acts as a catalyst, speeding the rate of a biochemical reaction but not altering its direction or nature.

epitope: Specific site on a protein to which an antibody will bind.

***Escherichia coli*:** Common bacterium that has been studied intensively by geneticists because of its small genome size, normal lack of pathogenicity, and ease of growth in the laboratory.

ethanol (CH_3CH_2OH): Simple alcohol containing only two carbon atoms. Ethanol is a product of the enzymatic breakdown of carbohydrates during microbial fermentation. Ethanol is combustible and can be used as a transportation fuel or fuel additive to improve gasoline combustion and reduce carbon monoxide emissions.

Eukarya: One of the three domains of life (along with Archaea and Bacteria) distinguished through DNA sequence analysis. Eukarya include animals, plants, fungi, and a variety of single-celled organisms that may have plant-, animal-, or fungi-like characteristics.

eukaryote: Cell or organism with membrane-bound, structurally discrete nucleus and other well-developed subcellular compartments. See also *prokaryote.*

exogenous DNA: DNA originating outside an organism that is introduced into the organism.

expression vector: Cloning vector engineered with regulatory signals in its DNA that enhance the transcription and translation (protein synthesis) of the gene clone inserted into it.

extracellular: Outside the cell.

extremophile: Type of microorganism that can survive extremes in temperature, salinity, pressure, and other environmental conditions detrimental to most forms of life.

exudate: See *root exudate*.

fatty acid: Long-chain carboxylic acid, typically containing 4 to 24 carbons, liberated from the hydrolysis of fats and oils.

fermentation: Metabolic pathway that breaks down organic compounds to generate cellular energy in the absence of oxygen. The production of ethanol using yeast is a fermentation pathway.

fermenter, fermentor: Large growth chamber (containing liters of liquid media) in which optimal conditions are maintained for the production of some desired product (e.g., ethanol) from microbial fermentation processes.

FISH: See *fluorescence in situ hybridization*.

flow cytometry: Analysis of biological material by detection of light-absorbing or fluorescing properties of cells or subcellular components (e.g., chromosomes) passing in a narrow stream through a laser beam. An absorbance or fluorescence profile of the sample is produced. Automated sorting devices, used to fractionate samples, separate successive droplets of the analyzed stream into different fractions depending on the fluorescence emitted by each droplet.

fluorescence: Ability of a substance to emit light at one wavelength after it has been activated by absorption of light at a shorter (higher-energy) wavelength.

fluorescence in situ hybridization (FISH): Technique that uses fluorescent probes targeted to 16S rRNA to identify and locate different populations in a microbial community. In this technique, fluorescently labeled oligonucleotide probes (containing about 20 nucleotides) are designed that will hybridize with a unique sequence in the 16S rRNA of a microbial population of interest. Cells from a culture or environmental sample are treated so they will be permeable to the fluorescent probe and then immobilized on a microscope slide. After the probes have been applied and allowed to hybridize with rRNA in the cells, the sample can be imaged using confocal microscopy. Any cells that have hybridized with the florescent probes can be identified and localized within a microbial community.

fluorescence (Förster) resonance energy transfer (FRET): Fluorescence-labeling technique that uses two different fluorescent dye molecules (fluorophores) to identify interacting pairs of proteins in vivo. Two proteins are engineered genetically so one protein is tagged with a donor fluorophore and a second protein is tagged with an acceptor fluorophore. When the two flourophores are within a few nanometers of each other, light energy emitted from the donor excites the acceptor, which emits light at another wavelength. The acceptor cannot emit light without being close to the donor, therefore, any detection of acceptor fluorescence indicates where the two proteins are interacting inside a cell. FRET emissions can be detected using confocal microscopy.

fluorophore: Group of atoms capable of fluorescence. Fluorophores can be used to label and track proteins and other molecules in vivo.

Fourier transform ion cyclotron resonance (FTICR): Type of mass spectrometry with higher resolution and mass accuracy than other MS techniques. FTICR can be used to analyze the mass of large ions generated by electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI). FTICR uses electrical and magnetic fields to trap ions in a chamber. As the ions circulate inside the chamber, they generate an electrical signal that is received by a detector. A mathematical function (the Fourier transform) is used to convert the detected signal into a mass-to-charge ratio for each ion.

FRET: See *fluorescence (Förster) resonance energy transfer*.

fuel cell: Device that converts the chemical energy of a fuel (e.g., hydrogen) into electricity without combusting the fuel.

fusion protein: Protein formed by genetically fusing or combining a gene encoding a target protein of interest with a gene encoding a protein or portion of protein that adds a desired functionality to the target (e.g., the ability to fluoresce or bind a small molecule on an affinity column). The fused genes then can be used as a template for synthesizing the "fusion protein" (a target protein engineered to have some additional, desired functionality).

fusion tag: Short peptide, protein domain, or entire protein that can be fused to a target protein of interest to create a fusion protein. The fusion tag generally possesses a special functionality or biochemical property that can be added to the target protein. A fusion tag can be removed from the target protein by the enzymatic cleavage of a linker region that connects the tag to the target.

gas chromatography (GC): Analytical technique used to separate the chemical components of a mixture. A sample is vaporized and carried by a stream of inert gas through a separation column (millimeters in diameter and meters in length). The column contains a solid or liquid material through which the chemical components migrate at different rates. As each separated component exits the column, it generates a signal that can be used to determine the amount and identity of each chemical. A gas chromatograph can be interfaced with a mass spectrometer.

gene: Fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides, located in a particular position on a particular chromosome, that encodes a specific functional product (i.e., a protein or RNA molecule).

gene expression: Process by which a gene's coded information is converted into structures present and operating in the cell. Expressed genes include those transcribed into mRNA and then translated into proteins, as well as those transcribed into RNA but not translated into proteins [e.g., transfer (tRNA) and ribosomal RNA (rRNA)].

gene family: Group of closely related genes that make similar products.

gene prediction: Computer prediction identifying possible genes based on how well an unknown stretch of DNA sequence matches known gene sequences.

gene product: Biochemical material, either RNA or protein, resulting from expression of a gene. The amount of gene product is used to measure a gene's level of activity.

gene regulatory network (GRN): Intracellular network of regulatory proteins that control the expression of gene subsets involved in particular cellular functions. A simple GRN would consist of one or more input signaling pathways, regulatory proteins that integrate the input signals, several target genes (in bacteria a target operon), and the RNA and proteins produced from those target genes.

genetic code: Nucleotide sequence, coded in triplets along the mRNA, that determines the sequence of amino acids in a protein product. Each set of three nucleotides (codon) in a gene specifies a particular amino acid or signals the start or stop of protein synthesis.

genetic engineering: Alteration of the genetic material of cells or organisms to enable them to make new substances or perform new functions.

genome: All the genetic material in the chromosomes of a particular organism. Most prokaryotes package their entire genome into a single chromosome, while eukaryotes have different numbers of chromosomes. Genome size generally is given as total number of base pairs.

genome sequence: Order of nucleotides within DNA molecules that make up an organism's entire genome.

genomic plasticity: Alterable nature of prokaryotic genomes that enables the fluid exchange of DNA from one microorganism to another and allows prokaryotes to adapt their genomes rapidly so they can survive changes in environmental conditions. See also *lateral (horizontal) gene transfer.*

genomics: The study of genes and their function.

genotype: An organism's genetic constitution, as distinguished from its physical characteristics (phenotype).

geochemistry: The study of the chemical components that make up the earth's crust and the reactions and processes that influence the formation and cycling of those components.

gigabyte: Unit of computer storage equal to one billion (10^9) bytes.

gigaton (Gt): One billion metric tons; a metric ton is a unit of mass equal to 1000 kg (about 2200 lb).

global assay: Any of a variety of techniques that examine comprehensive sets of biomolecules (e.g., proteins, mRNA molecules, metabolites) present in a cell under certain conditions.

glucose: A six-carbon sugar with the chemical formula $C_6H_{12}O_6$. Glucose is a widely used carbon and energy source in biology and an important product of photosynthesis.

glycosyl hydrolase: Group of enzymes capable of hydrolyzing (breaking) the glycosidic bond that links a carbohydrate to another molecule. A cellulase is a type of glycosysl hydrolase that breaks the bond between glucose subunits in cellulose.

green fluorescent protein (GFP): Protein, originally isolated from jellyfish, containing a fluorophore molecule that emits green light when it absorbs UV light. GFP can be used to fluorescently label a target molecule and track it in vivo.

greenhouse gas (GHG): Heat-trapping gas such as carbon dioxide, methane, nitrous oxide, or dimethyl sulfide released into the atmosphere as a result of human activities (primarily fossil-fuel combustion) and natural processes (e.g., cellular respiration, biomass decomposition, volcanic activity).

grid computing: Large-scale computer processing tasks carried out by high-speed connections among many smaller computer systems housed at different locations and administered separately. The aggregation of unused computing resources can be applied to solving complex computational problems that otherwise would require more expensive supercomputers.

guanine (G): Nitrogenous base, one member of the base pair GC (guanine and cytosine) in DNA.

hemicellulose: Any of several polysaccharides (e.g., xylans, mannans, and galactans) that cross-link and surround cellulose fibers in plant cell walls. Hemicellulose molecules have less complicated structures than those of cellulose and are broken down more easily into their simple sugar subunits.

heterocyst: Specialized cell formed by some species of filamentous cyanobacteria when nitrogen sources in the environment have been depleted. A heterocyst has a thick, layered cell wall that minimizes the flow of oxygen into its interior; nitrogen-fixation reactions, which are sensitive to oxygen, take place within the heterocyst.

heterologous host: Host cell into which an expression vector is inserted and used to express large amounts of a protein naturally synthesized in a different species.

homology: Similarity in DNA or protein sequences among individuals of the same species or among different species.

homologous host: Host cell into which an expression vector is inserted and used to express large amounts of a protein derived from the host cell's genome. The expression vector may encode an altered form of the protein (e.g., a protein that has been mutated or labeled with a fluorophore).

horizontal gene transfer (or lateral gene transfer): Exchange of genetic material between two different organisms (typically different species of prokaryotes). This process gives prokaryotes the ability to obtain novel functionalities or cause dramatic changes in community structure over relatively short periods of time. See also *vertical gene transfer*.

humus: Long-lived mixture of organic compounds derived from the microbial decomposition of plant and animal matter in soils.

hybridization: Process of joining two complementary strands of DNA or one each of DNA and RNA to form a double-stranded molecule.

hybridoma: Cell line resulting from the fusion of a cancer cell with a lymphocyte (a cell that produces antibodies); hybridomas are used for the continuous production of antibodies.

hydrogenase: Enzyme capable of one or both of the following activities: (1) reduce (add electrons to) protons to generate molecular hydrogen (H_2) and (2) oxidize H_2 to generate protons and electrons (the electrons are used to reduce other molecules). Enzymes that generate H_2 are called evolving hydrogenases; those that oxidize hydrogen are called uptake hydrogenases; and those capable of catalyzing both types of reactions are called bidirectional hydrogenases.

hydrology: Study of the properties and movement of water in the atmosphere and the earth's lakes, streams, and groundwater. The study of marine waters is part of oceanography.

hydrolysis: Type of chemical reaction that uses water to cleave chemical bonds and break a large molecule into smaller components.

hypothesis-driven science: Approach in which experimental methods are used to test the validity of a hypothesis and answer specific scientific questions. See also *discovery-driven science*.

inducer: Substrate that enhances gene transcription by preventing a repressor from inhibiting the expression of a gene involved in the substrate's metabolism.

in silico: Using computers to simulate and investigate natural processes.

in situ: In a natural environment.

in vitro: "In a test tube" or outside a living organism.

in vivo: Within a living organism.

infrared (IR) spectroscopy: Technique used to characterize the structures of organic molecules. Infrared radiation has lower energy and longer wavelength (between 800 nm and 1 mm) than visible light. An infrared spectrometer measures a sample's transmission of infrared radiation. The covalent bonds within a molecule absorb infrared radiation at characteristic wavelengths. Molecules thus absorb infrared radiation in a unique pattern that can be used as a "fingerprint" for identifying that molecule.

interaction: Binding together of two or more molecules to carry out a specific cellular function.

interaction network: Diagram that shows numerous molecular interactions of a cell. Each point or node on the diagram represents a molecule (typically a protein), and each line connecting two nodes indicates that two molecules are capable of interacting.

interactome: Molecular interactions of a cell, typically used to describe all protein-protein interactions or those between proteins and other molecules.

ion: Atom or group of atoms that carry an electrical charge. Ions with a positive charge are called cations; ions with a negative charge are called anions.

ion suppression: In mass spectrometry, inhibition of ion formation of an analyte caused by the presence of less-volatile compounds. Analytes lost due to ion suppression never reach the detector, which results in artificially low readings for certain analytes.

iron fertilization: Delivery of iron-containing micronutrients to ocean regions to enhance the growth of phytoplankton that use carbon dioxide from the atmosphere to build biomass.

isoform: Any of a group of functionally similar proteins that vary slightly in amino acid sequence**.**

isomer: Molecule that has the same chemical formula as another but differs in how the atoms are bonded together or structurally arranged.

isotope: Atom that has the same number of protons as another atom but a different number of neutrons and hence atomic mass. For example, ^{13}C is an isotope of carbon that has one more neutron than the most common isotope of carbon, ^{12}C.

isotope-coded affinity tag (ICAT): Reagent used to label proteins analyzed by mass spectrometry. Each ICAT reagent has three parts: (1) a chemical group that reacts with a protein, (2) a linker chain that is synthesized in both light versions (e.g., containing hydrogen atoms) and heavy versions (e.g., containing isotopes such as deuterium atoms), and (3) an affinity tag (e.g., biotin). If two protein samples (e.g., from the same types of cells grown under different conditions) are each labeled with a different ICAT reagent and mixed together, the relative abundance of proteins in each sample can be determined by MS analysis.

kinase: Enzyme that catalyzes phosphorylation reactions (transfer of a phosphoryl group between ATP and another molecule). Phosphorylation reactions often have important roles in turning certain cellular processes on and off.

kinetics: Field of study that deals with determining the rates of biological, chemical, and physical processes (e.g., how quickly reactants are converted into products) under various conditions.

knockout: Deactivation of specific genes in an organism's genome; used in the laboratory to study gene function.

knowledgebase: Comprehensive collection of knowledge stored in databases and used to solve problems in a particular subject area.

lab on a chip: Device consisting of a silicon (sometimes glass) chip chemically etched and fitted with tiny tubes and compartments (microns in size) through which materials flow. Advantages of experimentation at such a small scale are faster analysis times and significant reduction in required sample size. Also known as a MEMS (microelectromechanical system) device.

labeling: Incorporation of traceable chemical group (e.g., containing an isotope or a fluorescent dye) into a protein or other biomolecule of interest so it can be tracked or quantified during experimental analysis. See *tag*.

laboratory information management system (LIMS): Computer system used by laboratories to track samples; automate data capture from laboratory instruments; and facilitate the storage, presentation, and sharing of data among collaborating researchers.

laser confocal microscopy: See *confocal microscopy*.

lateral gene transfer: See *horizontal gene transfer*.

ligand: Any small molecule that binds a larger molecule.

ligation: Process of joining molecules or molecular fragments via covalent-bond formation.

lignin: Complex, insoluble polymer whose structure, while not well understood, gives strength and rigidity to cellulose fibers in the cell walls of woody plants. Lignin makes up a significant portion of the mass of dry wood and, after cellulose, is the second most abundant form of organic carbon in the biosphere.

ligninase: Type of enzyme capable of breaking down the complex polymeric structure of lignin into aromatic acid subunits called phenylpropanoids. Ligninases are known to be secreted by certain species of white rot fungi.

LIMS: See *laboratory information management system*.

lipid: Diverse class of biomolecules that are insoluble or minimally soluble in water. Fatty acids are key components of many complex lipid molecules that can include sugars and amino acids. Lipids take on many important cellular roles; they are the primary components of biological membranes, provide long-term storage of cellular energy, and carry electrons between membrane-embedded molecular complexes in electron-transport chains.

lysate: See *cell lysate.*

lyse: To rupture a cell and cause it to release its inner contents.

macromolecule: Large molecule (typically with a mass greater than several thousand Daltons) such as a protein, carbohydrate, or nucleic acid.

mass analyzer: Component of a mass spectrometer that uses electrical and magnetic fields to separate ionized molecules by their mass-to-charge ratios. Different mass analyzers include quadropole, time-of-flight, sector, ion trap, and FTICR.

mass spectrum (pl., spectra): Data output from a mass spectrometer consisting of a viewgraph that appears as a series of sharp peaks with each peak representing a particular ion fragment. The placement of each peak on the X axis corresponds to an ion's mass-to-charge ratio, and the height of each peak represents the relative abundance of each ion.

mass spectrometer: Instrument that ionizes molecules and then separates the resulting ions by mass and charge. A mass spectrometer consists of three basic components that operate in a vacuum: Ion source, which imparts a charge on each sample molecule; mass analyzer, which uses electrical or magnetic fields to separate each ionized molecule by its mass-to-charge ratio; and detector, which detects each separated ion and amplifies its electronic signal. The electronic signal from the detector is sent to a computer, which generates a mass spectrum for each component in the sample.

mass spectrometry (MS): Analytical technique that uses a mass spectrometer to determine mass-to-charge ratios of ions formed from the molecules in a mixture. The resulting data are used to identify each chemical component in the mixture. In proteomics analyses, MS techniques can be used to determine the mass, amino acid sequence, and post-translational modification for each protein in a sample.

massively parallel processing (MPP): Type of high-performance computing that involves running multiple processors in parallel to execute a single program.

mass-to-charge ratio (m/z): Dimensionless value measured by a mass spectrometer for each ion in a sample and determined by dividing the ion's mass (m) by its charge number (z). For example, a molecule with an atomic mass of 180 mass units and a net charge of +1 would have a mass-to-charge ratio of 180/1 or 180. Another molecule with a mass of 360 and a net charge of +2 also would have a mass-to-charge ratio of 180 or 360/2.

matrix-assisted laser desorption ionization (MALDI): Ionization method used in the MS analysis of proteins and other large biomolecules. Sample biomolecules (e.g., proteins or DNA fragments) are embedded within a solid, crystalline matrix of organic molecules. A laser beam is directed at the sample, and, as the crystals absorb the laser energy, they protect the more fragile biomolecules from destruction. The excited matrix molecules are vaporized and converted to ions that can be carried into the mass analyzer of a mass spectrometer.

megabase (Mb): Unit of length for DNA fragments, equal to 1 million nucleotides.

membrane: Semipermeable biological barrier consisting of lipids, proteins, and small amounts of carbohydrate. Membranes control the flow of chemical substances (e.g., nutrients, protons, ions, and wastes) in and out of cells or cellular compartments. They also serve as structural supports for systems of membrane-embedded proteins that mediate important biological processes such as photosynthesis and cellular respiration.

MEMS: Microelectromechanical system. See *lab on a chip*.

messenger RNA (mRNA): RNA that serves as a template for protein synthesis. See also *transcription* and *translation*.

metabolic flux analysis (MFA): Method for measuring all the metabolic fluxes of an organism's central metabolism; ^{13}C-labeled substrate is taken up by an organism, and the distribution of ^{13}C throughout the metabolic network enables the quantification of labeled metabolite pools.

metabolism: Collection of all biochemical reactions that an organism uses to obtain the energy and materials it needs to sustain life. An organism uses energy and common biochemical intermediates released from the breakdown of nutrients to drive the synthesis of biological molecules.

metabolite: Small molecules (<500 Da) that are the substrates, intermediates, and products of enzyme-catalyzed metabolic reactions.

metabolome: All metabolites present in a cell at a given time.

metabolomics: Type of global molecular analysis that involves identifying and quantifying the metabolome.

metadata: Data that describe specific characteristics and usage aspects (e.g., what data is about, when and how data was created, who can access it, and available formats) of raw data generated from different analyses.

metagenome: Collective genomic DNA isolated from a community of organisms living in a particular environment.

metalloprotein: Protein that incorporates one or more metals into its molecular structure by binding individual metal ions [e.g., iron (Fe^{2+} or Fe^{3+}), zinc (Zn^{2+}), or magnesium (Mg^{2+})] or nonprotein organic compounds containing metals. Metalloproteins are important components of electron transport chains.

microarray: Analytical technique used to measure the mRNA abundance (gene expression) of thousands of genes in one experiment. The most common type of microarray is a glass slide onto which DNA fragments are chemically attached in an ordered pattern. As fluorescently labeled nucleic acids from a sample are applied to the microarray, they bind the immobilized DNA fragments and generate a fluorescent signal indicating the relative abundance of each nucleic acid in the sample. See also *protein chip*.

microbial genetics: The study of genes, gene function, and the transmission and regulation of genetic information in prokaryotic microorganisms.

microbial strain: See *strain*.

microgram (μg): Unit of mass equal to one-millionth (10^{-6}) of a gram or one-thousandth of a milligram.

micrometer (μm): Unit of length equal to one-millionth (10^{-6}) of a meter or one-thousandth of a millimeter.

micron: See *micrometer*.

microniche: 1. Specialized set of environmental conditions (e.g., pH, nutrient availability, electron-acceptor availability) that enables the survival of certain populations within a microbial community. 2. Function expressed by a microorganism or group of microorganisms living within a small portion of a community.

microorganism: Any unicellular prokaryotic or eukaryotic organism, sometimes called a microbe.

model organism: Organism studied widely by a community of researchers. Biological understanding obtained from model organism research is used to provide insights into the biological mechanisms of other organisms. Microbial model microorganisms include the bacteria *Escherichia coli* and *Bacillus subtilis*, the yeast *Saccharomyces cerevisiae*, and the green alga *Chlamydomonas reinhardtii*.

modeling: Use of statistical and computational techniques to create working computer-based models of biological phenomena that can help to formulate hypotheses for experimentation and predict outcomes of research.

moiety: Portion of a molecule that carries out a particular function or gives a molecule a particular chemical characteristic.

mole: Quantity equal to 6.022×10^{23} items (e.g., molecules, atoms).

molecular docking: Binding of a molecule (e.g., ligand or protein) to a specific site on another protein to form a three-dimensional complex.

molecular machine: Highly organized assembly of proteins and other molecules that work together as a functional unit to carry out operational, structural, and regulatory activities in cells.

molecular tag: See *fusion tag*.

monoculture: Batch of microbial cells belonging to a single microbial strain or species grown in a laboratory.

motif: A sequence motif is a characteristic sequence pattern observed in different proteins or nucleic acids and typically associated with a particular function such as molecular binding. A structural motif is a recurring three-dimensional arrangement of structural elements observed in different proteins.

mutagenesis: Any process that alters an organism's genetic material (DNA or RNA sequence).

mutation: Permanent change in DNA sequence. See also *polymorphism*.

mutualism: Relationship in which both parties benefit.

NADH: Reduced form of nicotinamide adenine dinucleotide (NAD^+), a molecular carrier of high-energy electrons in living cells. NADH is formed when NAD^+ accepts a pair of electrons released from oxidation reactions. NADH then transfers its electrons to other molecules in the cell. NADH is an important source of electrons for the electron-transport pathway that generates ATP during cellular respiration.

nanometer (nm): Unit of length equal to one-billionth (10^{-9}) of a meter or one-millionth (10^{-6}) of a millimeter.

nanowire: Conductive, extracellular appendages that some bacteria can grow under certain environmental conditions and use to transfer electrons to metals. Nanowires have been observed in *Shewanella* and *Geobacter* species.

niche: 1. Set of environmental conditions required for the survival of a particular organism or group of organisms. 2. The functional role taken on by an organism in a particular ecosystem.

nitrogenase: Enzyme that catalyzes the conversion of atmospheric nitrogen (N_2) to nitrate in nitrogen-fixing bacteria.

nitrogen fixation: Process carried out by certain species of bacteria in which atmospheric nitrogen (N_2) is converted to organic nitrogen-containing compounds that can be used by other living organisms.

NMR: See *nuclear magnetic resonance (NMR) spectroscopy*.

Northern blot: Gel-based laboratory procedure that locates mRNA sequences complementary to a piece of DNA used as a probe.

nuclear magnetic resonance (NMR) spectroscopy: Non-destructive technique that uses magnetic fields and radio-frequency (rf) pulses to analyze the structures of metabolites, proteins, or other molecules in solution. The magnetic nuclei of certain atoms (e.g., ^{1}H , ^{13}C, ^{31}P) can absorb rf energy of a particular frequency at different magnetic-field strengths. A detector within the NMR spectrometer monitors the absorbance of rf energy associated with different magnetic environments, and this information can be used to determine the position of each nuclei within a molecular structure.

nucleic acid: Large molecule composed of nucleotide subunits. See also *DNA* and *RNA*.

nucleotide: Chemical subunit of DNA or RNA consisting of a nitrogenous base (adenine, guanine, thymine, or cytosine in DNA; adenine, guanine, uracil, or cytosine in RNA), a phosphate molecule, and a sugar molecule (deoxyribose in DNA and ribose in RNA).

oligo: See *oligonucleotide*.

oligonucleotide: Short segment of 25 or fewer nucleotides that can hybridize with complementary sequence in a DNA sample.

ontology: Organized, hierarchical structure of concepts relevant to a particular knowledge domain. An ontology identifies which of several equivalent terms should be used to represent a concept and defines how different terms and concepts are related. Ontologies are developed to ensure the consistent use of language across multiple databases and information systems.

open reading frame (ORF): Sequence of DNA or RNA located between the start-code sequence (initiation codon) and the stop-code sequence (termination codon).

operon: In prokaryotic genomes, a linear group of genes transcribed together on the same mRNA molecule and controlled by the same regulatory element.

ORF: See *open reading frame*.

oxidation: Loss of one or more electrons from a chemical substance.

oxidative stress: In aerobic organisms, the cellular damage caused by reactive species of oxygen (e.g., free radicals and peroxides) that are by-products of the metabolism of oxygen.

oxygenic: Producing oxygen.

parasitism: Relationship in which one organism (parasite) obtains the resources it needs for survival from another organism (host) on which or within which it lives.

pathway: Series of molecular interactions that occur in a specific sequence to carry out a particular cellular process (e.g., sense a signal from the environment, convert sunlight to chemical energy, break down or harvest energy from a carbohydrate, synthesize ATP, or construct a molecular machine).

peptide: Two or more amino acids joined by a bond called a peptide bond.

petabyte: Unit of computer storage representing one quadrillion or 10^{15} bytes (equal to 1000 terabytes).

petaflop: Measure of computer speed representing one quadrillion or 10^{15} floating-point operations per second.

petascale: Level of high-performance computing capable of petaflop processing and the management of enormous petabyte data sets.

pH: Scale used to specify acidity or alkalinity. The hydrogen ion (H^+) concentration of a sample determines its pH ($pH = -\log_{10}[H^+]$); the higher the H^+ concentration, the lower the pH. A solution with a pH value of 7 is neutral; less than 7 is acidic; and greater than 7 is alkaline or basic.

phage: Virus for which the natural host is a bacterial cell.

phage display: Method used to detect interactions between peptides or proteins and other molecules. The gene for one of many random peptide or protein variants is fused to a gene encoding a coat protein expressed on the surface of bacteriophage. Libraries of phages, each displaying a different peptide on its surface, are created and applied to a target ligand immobilized on a solid support. Phage-displaying peptides that bind the ligand are purified and used to infect *E. coli*. DNA sequencing of the infected *E. coli* is used to identify the peptides that bound the target ligand.

phenotype: Physical characteristics of an organism.

phosphorylation: Type of chemical modification that adds a phosphate group (PO_4^{3-}) to a molecule. Phosphorylation is an important type of post-translational modification involved in the regulation of protein activity.

photolysis (photolytic): Use of light energy to break a chemical bond, such as cleavage of hydrogen-oxygen bonds in water to produce oxygen and hydrogen ions.

photon: Fundamental unit (quantum) of electromagnetic energy (e.g., light) that has no mass or electric charge.

photosynthesis: Process by which plants, algae, and certain types of prokaryotic organisms capture light energy and use it to drive the transfer of electrons from inorganic donors [e.g., water, thiosulfate (H_2S)] to carbon dioxide to produce energy-rich carbohydrates.

photosystem: Large, membrane-bound molecular complex consisting of multiple proteins containing pigment molecules (e.g., chlorophylls) that absorb light at a particular wavelength and transfer the energy from the absorbed photon to a reaction center that initiates a series of electron-transport reactions.

phylogenetic tree: Branching, hierarchical diagram that organizes species or other taxonomic units based on evolutionary relationships.

phylogeny: Evolutionary history that traces the development of a species or taxonomic group over time.

physiology: Study of the functions of living organisms and the factors that influence those functions.

phytoplankton: Microscopic photosynthetic organisms (e.g., algae, cyanobacteria, dinoflagellates) found in the surface layers of marine and freshwater environments.

plasmid: In prokaryotes, a circular DNA segment distinct from chromosomal DNA that autonomously replicates and can be transferred from one organism to another. Plasmids can be engineered and used to introduce modified or foreign genetic material into host organisms.

polar molecule: Molecule having an uneven distribution of electrons. A polar molecule has a region of partial positive charge and a region of partial negative charge and usually dissolves in other polar substances (e.g., water).

polymerase chain reaction (PCR): Rapid technique for generating millions or billions of copies of any piece of DNA. PCR also can be used to detect the existence of a particular sequence in a DNA sample.

polymerase (DNA or RNA): Enzyme that catalyzes the synthesis of nucleic acids on preexisting nucleic acid templates, assembling RNA from ribonucleotides or DNA from deoxyribonucleotides.

polymorphism: Common variation in a gene's DNA sequence observed among individuals of the same species.

polysemy: Diversity of meanings, as when the same term represents multiple concepts. Polysemy is one of many issues addressed when defining data standards. See also *synonymy*.

population: Collection of organisms of the same species living together in a given area. A microbial community comprises several different populations.

post-transcriptional regulation: Process that controls gene expression in cells by influencing the conversion of an mRNA transcript into protein.

post-translational modification: Any of several chemical modifications (e.g., phosphorylation, disulfide bond formation, cleavage of inactive sequence) involved in converting a newly translated amino acid sequence into a functional protein.

post-translational regulation: Process that controls the expression of gene products in cells by influencing the conversion of a newly translated amino acid sequence into a functional protein.

primary structure: Linear sequence of amino acids in a protein.

probe: Molecule used to isolate or detect the presence of certain biomolecules in a sample. A probe molecule may be a segment of DNA of known sequence that can hybridize complementary sequences in a genome or an antibody that specifically binds some protein of interest. A probe may be labeled with fluorescent groups or radioactive isotopes to facilitate isolation and detection.

prokaryote: Single-celled organism lacking a membrane-bound, structurally discrete nucleus and other subcellular compartments. Bacteria and archaea are prokaryotes. See also *eukaryote*.

promoter: DNA site to which RNA polymerase will bind and initiate transcription.

protein: Large molecule composed of one or more chains of amino acids in a specific order; the order is determined by the base sequence of nucleotides in the gene that codes for the protein. Proteins maintain distinct cell structure, function, and regulation.

protein chip: Glass slide onto which an ordered array of proteins has been chemically attached. Protein chips can be used to identify and measure the abundance of proteins in a sample, detect protein-protein interactions, or screen thousands of proteins simultaneously for a particular biochemical function.

protein complex: Aggregate structure consisting of multiple protein molecules.

protein digestion: See *proteolysis*.

protein family: Category of related proteins similar in structure and function. Members of a protein family have highly conserved sequence with greater than 50% sequence identity.

protein folding: Process that structurally arranges a linear polypeptide chain to form a three-dimensional, biologically active form of a protein in cells. Predicting a protein's functional 3D structure from its sequence is difficult due to the many possible different interactions that can occur between atoms in the same protein molecule.

protein interaction network: See *interaction network*.

proteolysis: Breakdown of a large protein into shorter polypeptide chains by the hydrolysis of peptide bonds.

proteome: Collection of proteins expressed by a cell at a particular time and under specific conditions.

proteomics: Large-scale analysis of the proteome to identify what proteins are expressed by an organism under certain conditions. Proteomics provides insights into protein function, modification, regulation, and interaction.

pulldown: Isolation of a protein or molecular complex from a mixture of molecules.

pull-down bait: Molecule (e.g., tagged protein or antibody) that binds a protein or molecular complex of interest and facilitates its isolation from a mixture of molecules.

QA/QC: See *quality assurance* and *quality control*.

quality assurance: Approach used to ensure that systems will perform to a required standard for quality.

quality control: Methods used to determine if the products of a process meet or exceed a defined standard for quality.

quantum dots: Inorganic nanocrystals that can be used as fluorescent tags in a variety of live-cell imaging techniques. When a quantum dot is excited by the absorption of light at one wavelength, it emits a narrow spectrum of light at other wavelengths depending on its size and shape. Multiple quantum dots of different shapes can be used simultaneously to label different components in a sample. In contrast, organic dyes (fluorescent molecules naturally synthesized in fireflies and jellyfish) are shorter lived and tend to emit light over broad, overlapping ranges of wavelengths, thus preventing the use of more than two or three organic dyes at once.

quaternary structure: Three-dimensional molecular complex consisting of two or more folded polypeptide chains.

quorum sensing: Mechanism by which bacteria communicate and coordinate activity by sensing the concentrations of signaling molecules they release into the environment.

recombinant: Type of molecule or organism created in the laboratory by combining DNA from two or more sources.

reductant: See *electron donor.*

reduction: Electron-transfer reaction in which a substance gains one or more electrons.

regulator: Protein (e.g., a repressor) that controls the expression or activity of other molecules in a cell.

regulatory elements: Segments of the genome (e.g., regulatory regions, genes that encode regulatory proteins or small RNAs) involved in controlling gene expression.

regulatory map: See *gene regulatory network*.

regulatory region or sequence: Segment of DNA sequence to which a regulatory protein binds to control the expression of a gene or operon.

regulon: Set of operons controlled by the same regulator. Operons belonging to the same regulon can be located in different regions of a genome.

repressor: Regulatory protein that binds the operator site and inhibits transcription of genes in an operon. See also *activator*.

resolution: 1. The smallest distance between two points that can be distinguished by a microscope as separate objects; the smaller the distance, the higher the resolution of the microscope. 2. The ability of a separation technique (e.g., electrophoresis, chromatography) to separate two similarly sized components in a sample. 3. In mass spectrometry, the ability of an instrument to separate ions that differ only slightly in their mass-to-charge ratios.

resolving power: See *resolution*.

respiration: Series of biochemical redox reactions in which the energy released from the oxidation of organic or inorganic compounds is used to generate cellular energy in the form of ATP.

rhizosphere: Zone immediately surrounding the root of a plant.

rhodopsin: Light-sensitive pigment protein found in the retinas of animals. It shares structural similarities to the bacteriorhodopsins found in prokaryotes.

ribosome: Molecular machine, composed of specialized RNA and proteins, which binds mRNA and uses mRNA sequence as a template for protein synthesis.

RNA (ribonucleic acid): Molecule that plays an important role in protein synthesis and other chemical activities of the cell. RNA's structure is similar to that of DNA. Classes of RNA molecules include messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and other small RNAs, each serving a different purpose.

root exudate: Chemical substance released from the root of a plant.

scanning electron microscopy (SEM): Type of electron microscopy in which a focused beam of electrons is scanned back and forth across the surface of a specimen, which is dehydrated and coated with a thin layer of a metal such as gold. The beam of primary electrons knocks off secondary electrons from the sample surface. The emitted secondary electrons generate signals that are amplified and used to build a 3D representation of the specimen.

scanning near-field optical microscopy (SNOM): Type of scanning probe microscopy in which a metal-coated optical fiber tip, positioned nanometers above a specimen, beams laser light onto the specimen's surface. An optical microscope detects the optical response of the laser light as it interacts with the sample. Passing light through a tiny aperture (25 to 100 nm in diameter) at the probe tip onto a specimen at such a close proximity produces an optical response that can be used to construct images with a resolution of about 50 to 80 nm, much higher than that of typical optical microscopes.

scanning probe microscopy (SPM): Any of several imaging techniques that involve sweeping a probe attached to a flexible cantilever across the surface of a specimen. Interactions between the probe and the specimen surface produce signals that can be used to generate an image of the specimen. Different types of scanning probe microscopy include atomic force microscopy, scanning tunneling microscopy, and scanning near-field optical microscopy.

scanning transmission electron microscopy (STEM): Type of electron microscopy that can be used to determine the mass and generate images of large biomolecular struc-

tures (e.g., proteins, DNA). A focused beam of electrons is scanned across the specimen, and a series of detectors within the instrument collect electrons that are transmitted through or scattered by the specimen. Signals from the detector may be used for compositional analysis of the molecular structure.

scanning tunneling microscopy (STM): Type of scanning probe microscopy that passes a sharp, conductive probe (consisting of a single atom at its tip) slightly above the surface of an electrically conductive speciman. A weak current of electrons, the "tunneling current," flows across the tiny gap between the tip of the needle and the specimen surface. The amount of current detected is related to the distance separating the tip and the specimen surface, and this information can be used to generate a 3D representation of a specimen's topography with atomic resolution.

secondary structure: Arrangement of a polypeptide chain into regions of recurring structural elements (e.g., alpha helices, beta sheets, turns) caused by hydrogen bonding among amino acids in the chain. Nucleotides of a single-stranded RNA molecule also interact to form secondary structures (e.g., the looped cloverleaf structure seen in tRNA).

sensitivity: Signal produced for a given amount of an analyte using an instrument or analytical technique.

sequence assembly: Arranging sequenced DNA fragments in their correct chromosomal positions.

sequestration: See *carbon sequestration*.

shotgun sequencing: Common approach to sequencing microbial genomes that involves breaking the genome into random fragments, which are cloned into vectors and sequenced. Computational analysis is used to compare all DNA sequence reads from random fragments and assemble the entire genome by aligning overlapping sequences.

siderophore: Chemical compound, secreted by certain species of microorganisms, that binds and solubilizes iron from the environment and facilitates the transport of iron into cells.

signal-transduction pathway: Series of biochemical reactions that receive extracellular chemical signals. These signals are transmitted and amplified within the cell and ultimately used to stimulate or repress a certain type of molecular activity (e.g., gene expression).

simulation: Combination of multiple models into a meaningful representation of a whole system that can be used to predict how the system will behave under various conditions. Simulations can be used to run in silico experiments to gain first insights, form hypotheses, and predict outcomes before conducting more expensive physical experiments.

small-angle neutron scattering (SANS): Type of molecular structural analysis carried out at facilities that have access to a neutron source. Neutrons are beamed at a sample in solution (no crystallization required). By measuring the angles at which neutrons are scattered, nanometer-scale information about the shape and structure of a molecule can be obtained.

small-angle X-ray scattering (SAXS): Type of molecular structural analysis in which X rays are beamed at a sample in solution (no crystallization required). By measuring the scattering pattern of the X rays after they have interacted with the sample, nanometer-scale information about the shape and structure of a molecule can be obtained.

small RNA molecule (sRNA): Functional RNA molecule, typically 350 nucleotides or fewer in length, that does not code for protein. sRNAs are known to regulate transcription, translation, and protein activity and can take on catalytic or structural functions as components of protein-RNA machines.

solubility pump: System of physical processes [e.g., changes in water temperature, ocean circulation, and gradient of carbon dioxide (CO_2) spanning the ocean depth] that influence the ocean's uptake of CO_2 from the atmosphere. In combination with ocean circulation, the solubility pump results in net CO_2 emissions at the equator and net CO_2 drawdown at high latitudes.

species: Taxonomic group of closely related organisms sharing structural and physiological features that distinguish them from individuals belonging to other species. In organisms capable of sexual reproduction, individuals of the same species can interbreed and generate fertile offspring. For microorganisms, a species is a collection of closely related strains.

spectromicroscopy: Combination of microscopy and spectroscopy techniques.

sporulation: Process by which certain species of bacteria produce differentiated cells called endospores. Under conditions unfavorable for growth (e.g., low nutrient availability, loss of hydration), endospores are formed and persist in a dormant state until favorable growth conditions return, causing the endospore to germinate and give rise to cells capable of normal growth and reproduction.

steady state: Growth state in which the concentration of bacterial cells is in equilibrium with the concentration of nutrients or substrates (i.e., the concentrations remain constant over time).

stochastic: Relating to a series of random events.

stoichiometry: Ratio of molecules in a structural complex.

strain: Representative of a species that differs genetically from others of the same species but not enough to be considered a new species. A strain of a microorganism often is created by genetically manipulating it to have some desired characteristic or phenotype.

structural genomics: The effort to determine the 3D structures of large numbers of proteins using both experimental techniques and computer simulation.

substrate: Substance transformed by enzymatic activity.

symbiosis: See *mutualism*.

synchrotron: Large machine that uses electric fields to accelerate charged particles to provide a continuous source of different types of electromagnetic radiation (e.g., infrared, ultraviolet, X ray) that can be used for a variety of applications, including the determination of molecular structure.

synonymy: Different terms having the same meaning. Synonymy is one of many issues addressed when defining data standards. See also *polysemy*.

synthetic biology: Field of study that aims to build novel biological systems designed to carry out particular functions by combining different biological "parts" or molecular assemblies.

syntrophy: Relationship in which two (or more) microbial populations metabolically interact to degrade a substance that one cannot metabolize alone.

systems biology: Use of global molecular analyses (e.g., measurements of all genes and proteins expressed in a cell at a particular time) and advanced computational methods to study how networks of interacting biological components determine the properties and activities of living systems.

systems microbiology: Systems biology approach that focuses on understanding and modeling microorganisms at molecular, cellular, and community levels.

tag: Molecule, chemical group, or amino acid sequence added to a protein of interest so it can be isolated or distinguished from other proteins in a mixture.

tandem mass spectrometry (MS/MS): Coupling of two mass spectrometers with a chamber known as a collision cell. The first mass spectrometer is used to separate and identify all ions in a sample. Selected ions are broken into smaller pieces in the collision cell before they enter the second instrument, which produces a mass spectrum for the pieces of selected ions. Analysis of the resulting mass spectrum provides structural information for each of the selected ions.

taxonomy: Hierarchical classification system for naming and grouping organisms based on evolutionary relationships.

terabyte: Unit of computer storage representing one trillion or 10^{12} bytes.

teraflop: Measure of a computer's speed representing one trillion floating-point operations per second.

terahertz (THz): Unit of frequency equivalent to 10^{12} hertz (10^{12} cycles per second).

tertiary structure: Arrangement of a polypeptide's folded secondary structural elements (e.g., alpha helices, beta sheets) into a three-dimensional structure.

terminal electron-accepting process (TEAP): Biochemical process in which electrons released from the oxidation of organic or inorganic compounds ultimately are transferred to a molecule or atom at the end of the electron-transport chain consisting of a series of intermediary electron-transfer reactions.

thylakoids: Membranes that contain the light-absorbing pigments, photosystems, and other proteins needed for photosynthesis. Thylakoids extend throughout the cytoplasm of photosynthetic prokaryotes. In eukaryotic plants and algae, thylakoids are housed in a special organelle called a chloroplast.

thymine (T): Nitrogenous base, one member of the base pair AT (adenine-thymine) in DNA.

transcript: RNA molecule (messenger RNA or mRNA) generated from a gene's DNA sequence during transcription.

transcription: Synthesis of an RNA copy of a gene's DNA sequence; the first step in gene expression. See also *translation.*

transcription factor: Protein that binds to regulatory regions in the genome and helps control gene expression.

transcriptome: All RNA transcripts present in a cell at a given time.

transcriptomics: Global analysis of expression levels of all RNA transcripts present in a cell at a given time.

transformation: Process by which genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome.

translation: Process in which the genetic code carried by mRNA directs the synthesis of proteins from amino acids. See also *transcription*.

transmembrane: Term used to describe a protein embedded within a membrane that spans the entire thickness of that membrane from its external surface to its internal surface.

transmission electron microscopy (TEM): Type of electron microscopy used to image the internal structure of specimens sliced into thin sections. A focused beam of electrons passes through the specimen and onto a fluorescent screen to generate a two-dimensional image. Less-dense portions of the sample transmit more electrons and are represented by brighter regions in the image; darker regions indicate the sample's denser portions. Stains can be used to enhance contrast between light and dark regions of an image.

transporter: Protein that transports a molecule from one location to another; in most cases, transporters are membrane proteins that control the movement of molecules in and out of cells.

ultrastructure: Cellular structure too small to be visualized with light microscopy; must be examined using higher-resolution imaging techniques such as electron microscopy.

ultraviolet (UV): Form of electromagnetic radiation having a wavelength roughly in the range from 100 to 400 nm. On the electromagnetic spectrum, it is found between the violet region of the visible light spectrum and X rays.

uracil: Nitrogenous base found in RNA but not DNA; uracil is capable of forming a base pair with adenine.

UV-CD (ultraviolet-circular dichroism): See *circular dichroism*.

vertical gene transfer: Inheritance or passing of genetic material from one generation to another. See also *horizontal gene transfer*.

virus: Noncellular biological entity that can replicate only by infecting a host cell and using its reproductive capabilities.

Western blot: Method for immunologically detecting the presence of a protein in a sample. Proteins are separated by electrophoresis and then transferred to a special paper. Labeled antibodies specific to the protein of interest are used to reveal the position of the immobilized target proteins.

wide-angle X-ray scattering (WAXS): Technique for studying how ligand binding causes changes in protein structure. A beam of X rays is directed at a solution containing the protein and its ligand. The X-ray scattering pattern is used to produce structural information for the ligand-protein complex. These data can be compared with structural information for the ligand-free protein to identify changes in protein structure. Although small-angle X-ray scattering is a similar technique, it is unable to detect small changes in protein structure. Since crystallization is not required, structural information is more quickly obtained with WAXS than with X-ray crystallography.

wild type: Form of an organism that occurs most frequently in nature.

wiring diagram: Visual representation of all components and connections that make up various cellular networks including signaling, regulatory, and metabolic networks.

X-ray crystallography: Technique used to obtain structural information for a substance (e.g., protein, molecular complex) that has been crystallized. A beam of X rays is focused on the crystals, and the scattering pattern of the X rays is used to create 3D representations of the crystal with atomic resolution.