I. Introduction

Proteins (or polypeptides) are biopolymers composed of hundreds to hundreds of thousands of repeating units. Each repeating unit includes an amide functional group (called a peptide bond) formed by the condensation reaction between two amino acids. While the main chain of the protein remain constant throughout the protein, side chain atoms (represented by R below) may consist of one of 20 different groups of atoms. These side chains determine the shape and properties of the protein, and are responsible for the nearly endless variety of protein sequences, shapes, and functions found in nature.

1. A generic amino acid is shown below. Circle and label the following items:
   - amine group  - carboxylic acid group  - side chain  - alpha carbon

   Why do you think this molecule is called an amino acid?

2. The identity of “R” determines which amino acid we have, and defines the properties of that amino acid (is it polar, nonpolar, electrostatically charged, big, small, etc.?). Naturally-occurring amino acids contain one of 20 different R groups. Some examples are shown below. For each, identify whether the R group is polar, nonpolar, or charged. If the R group could participate in a hydrogen bond, circle it. If the R group could participate in an ionic (or ion-dipole) interaction, put a box around it.
II. Chemistry of peptide formation.
Polymerization of amino acid building blocks is required to create the protein. This is conceptually similar to the

3. Two amino acids react in a/an ____________________________ reaction to form a/an ____________________________ and ____________________________.

4. Draw the two products formed in the reaction (side chains are shown as generic $R_1$ and $R_2$):

5. When many amino acids are combined, a ____________________________ is formed.

6. In the polypeptide (a.k.a. protein) shown below, circle the amide groups (a.k.a. peptide bonds) and the backbone. (The ends of the protein are shown charged because this is how they exist under physiological conditions.)

7. Analysis of a protein shows that it contains 153 amino acids. Assuming it was synthesized from individual amino acids, how many molecules of the side product were produced?
III. Protein Structure and Function.
The spatial arrangement of proteins can be described as a hierarchy of four structural levels. Each level is associated with particular types of bonding or intermolecular forces, as described below. (Figure taken from Branden and Tooze, Introduction to Protein Structure, 2nd ed.)

The primary structure of a protein is determined by its amino acid sequence. That is, the primary structure is the identity of which amino acids are in the protein, as well as the order in which that are connected. The protein begins with an amino group and ends with a carboxylate group, so the order of the amino acid sequence is unidirectional. (ex: The protein sequence Alanine-Histidine-Serine is not the same as the sequence Serine-Histidine-Alanine.) The primary sequence determines the identity and properties of the protein, as well as its final shape (tertiary structure, see below). The peptide bonds between amino acids are covalent bonds.

The secondary structure of a protein refers to the structural motifs that form when the protein folds up into a defined shape. Examples of these motifs are the alpha-helix (α-helix) and the beta-strand (β-strand). Multiple β-strands can form together in layers to make a β-sheet. These secondary structural elements are fairly short. They are held together primarily by hydrogen bonds through the protein main chain.

8. For the alpha helix shown at the right, describe what sort of interactions maintain the secondary structure.
The **tertiary structure** of a protein is its overall shape. **The folding of the protein into this three-dimensional form is defined by its primary structure** (although environmental factors, such as **type of solvent, ionic strength, and pH also play a role**). The tertiary structure of a protein usually is such that the hydrophobic (nonpolar) amino acids are gathered in the inside, or **core**, of the protein, and the polar and charged amino acids are found on the outside, or **surface** of the protein. This makes sense, since water is a polar molecule, and most proteins are found in aqueous solutions! The surface amino acids can make **dipole-dipole**, **ion-dipole**, or **hydrogen bonding** interactions with solvent (water) molecules, or **ion-dipole** or **ionic** interactions with ions dissolved in the solvent (of course London dispersion forces exist here too). Nonpolar amino acids buried in the core of the protein interact through London dispersion forces (and so-called “hydrophobic interactions”).

Some proteins consist of two or more distinct protein chains which associate specifically to carry out a physiological function. This is known as **quaternary structure**. The interactions between neighboring proteins may include any type of intermolecular force, depending on the nature of the protein side chains involved.

9. Consider the amino acids in question #2. Which ones would you expect to be in the protein interior? Which ones would you expect to be on the protein surface?

10. What are some biological functions of proteins?

11. When discussing proteins, the phrase “Function follows form” is often used, since the primary structure of the protein (along with environmental conditions, such as pH, ionic strength, and type of solvent) determines its tertiary structure. Give one example of a single amino acid substitution in a protein that results in a devastating loss of function.
NASA’s Astrobiology Institute seeks to identify potential sites for extraterrestrial life. This requires that scientists have a range of environmental limits within which to consider “habitable zones”. These environmental limits include temperature, pressure, pH, and salinity (salt content). Extremophilic organisms on Earth exhibit unusual growth behaviors and may serve as a model for extraterrestrial organisms. These organisms are predominantly single-celled organisms. Many of them are part of an ancient classification of organisms called archae; they are different from both bacteria and more-complex eukaryotic organisms. For example, hyperthermophilic organisms grow optimally in thermal vents in the ocean floor at temperatures ~100°C.

Images of hyperthermophilic life. (a) Deep-sea thermal vent where hyperthermophilic *Pyrococcus* species have been identified. (b) *Pyrococcus furiosus* cells (literally: “rushing fireball”). These organisms thrive at 100°C and 200 atm!

Halophilic organisms live in high-salt (ex: 5 M NaCl!) environments such as the Dead Sea.

Images of halophilic life. (a) Halobacteria at a salt works near San Quentin in Baja California Norte, Mexico. A particular light-absorbing protein (bacteriorhodopsin) is responsible for the red color. (b) *Halobacterium salinarum* cells.

12. Predict how a protein from a hyperthermophilic organism might differ from the same protein in an organism that grows optimally at physiological temperature of 37°C. Specifically, discuss how the amino acid arrangement on the protein surface might be different to maintain its tertiary structure at 100°C.

13. Predict how a protein from a halophilic organism might differ from the same protein in an organism that grows optimally at physiological salinity. Specifically, discuss how the amino acid arrangement on the protein surface might be different to maintain its tertiary structure in high-salt conditions.