Finally, the term 'functional proteomics' should be introduced, as it refers to the 3-D structure determination of all proteins encoded by the genome of key organisms, a major focus of this study. The major goal of functional proteomics is the analysis of protein structures by an integrated approach combining computer-based technologies of bioinformatics and the in-depth analysis of 3-D protein structures through physical methods, such as nuclear magnetic resonance (NMR) spectroscopy or x-ray crystallography (see below).

B. Proteins and the biological organism

Proteins[23] support every aspect of biological activity.[24] Through their structural stability, diversity, and chemical reactivity, proteins influence and enable most of the key processes associated with life. They operate as catalysts[25], provide mechanical support and immune protection, transport and store other molecules such as oxygen, cause movement[26], transmit nerve impulses, and direct growth and differentiation.[27]

I. Amino acid sequences

In order to understand the functioning of proteins one must be aware that the term "protein structure" refers to three distinct levels of organization: primary, secondary, and tertiary. The primary structure refers to the amino acid sequence as such. The secondary structure describes the conformation or spatial relationship adopted by local regions of the polypeptide chain. Finally, "tertiary structure" expresses the entire folding of the polypeptide chain.[28]

23 The origin of the word "protein" is usually attributed to Jöns Jakob Berzelius (1779-1848) and has been ascribed to derivation from the Latin word primarius, or from the Greek word for "first thing" (in Greek πρωτεϊνη = first element), see Whitford, David, Proteins – Structure and Function, Chichester, West Sussex, England, 2005, 1.

24 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 9.

25 The term catalyst refers to substances that accelerate chemical reactions.

26 Schwaiger, Ingo/Sattler, Clara/Hostetter, Daniel R./Rief, Matthias, The Myosin coiled-coil is a truly elastic Protein Structure, 1 Nature Materials 2002, 232.

27 Berg, Jeremy M./Tymoczko, John L./Stryer Lubert, Biochemistry, New York, NY, 2005, 41.

28 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 81. The term "quaternary structure" further refers to a certain association of multiple 3-D folded proteins to form multi-subunit complexes.

All natural proteins are composed of the same set of 20 amino acids.[29] Each amino acid is constructed with a central tetrahedral carbon atom connected to an amino group, a carboxylic acid group, a distinctive side chain, and a hydrogen atom. The side chains of the 20 amino acid building blocks vary tremendously in size, shape, and the presence of functional groups. Amino acids can be grouped as follows: (1) aliphatic side chains: glycine, alanine, valine, leucine, isoleucine, methionine, and proline; (2) aromatic side chains: phenylalanine, tyrosine, and tryptophan; (3) hydroxyl-containing aliphatic side chains: serine and threonine; (4) sulfhydryl-containing cysteine; (5) basic side chains: lysine, arginine, and histidine; (6) acidic side chains: aspartic acid and glutamic acid; and (7) carboxamide-containing side chains: asparagine and glutamine. These groups are somewhat arbitrary and many other assemblies are possible.[30]

Overall, the bonds between amino acids are described as the primary structure of the protein. These bonds have several important characteristics. First, they are resistant to hydrolysis[31], so that proteins are kinetically remarkably stable. Second, the peptide group is planar because the C-N bond has a significant double-bond character. Third, each peptide bond has both a hydrogen-bond donor (the NH group) and a hydrogen-bond acceptor (the CO group). Hydrogen bonding between these backbone groups is a distinctive feature of protein structure. Ultimately, the peptide bond is uncharged, which allows proteins to form tightly packed globular structures having significant amounts of the backbone buried within the protein interior. Because they are linear polymers, proteins can be described as sequences of amino acids. Such sequences are written from the amino to the carboxyl terminus.[32] The complete amino acid sequences of more than 100,000 proteins are now known and documented. The primary structure can be illustrated as follows:[33]

29 Berg, Jeremy M./Tymoczko, John L./Stryer Lubert, Biochemistry, New York, NY, 2005, 53.

30 Berg, Jeremy M./Tymoczko, John L./Stryer Lubert, Biochemistry, New York, NY, 2005, 71-72.

31 Hydrolysis is a chemical process by which water reacts with a compound to produce other compounds. Specifically, a bond is split, and the hydrogen cation and the hydroxide anion of the water are added.

32 Berg, Jeremy M./Tymoczko, John L./Stryer Lubert, Biochemistry, New York, NY, 2005, 72.

33 Based on figure provided by National Human Genome Research Institute, Talking Glossary of Genetic Terms, available at: http://www.genome.gov/Pages/Hyperion//DIR/VIP/Glossary/Illustration/ amino_acid .shtml, last checked on October 16, 2005.

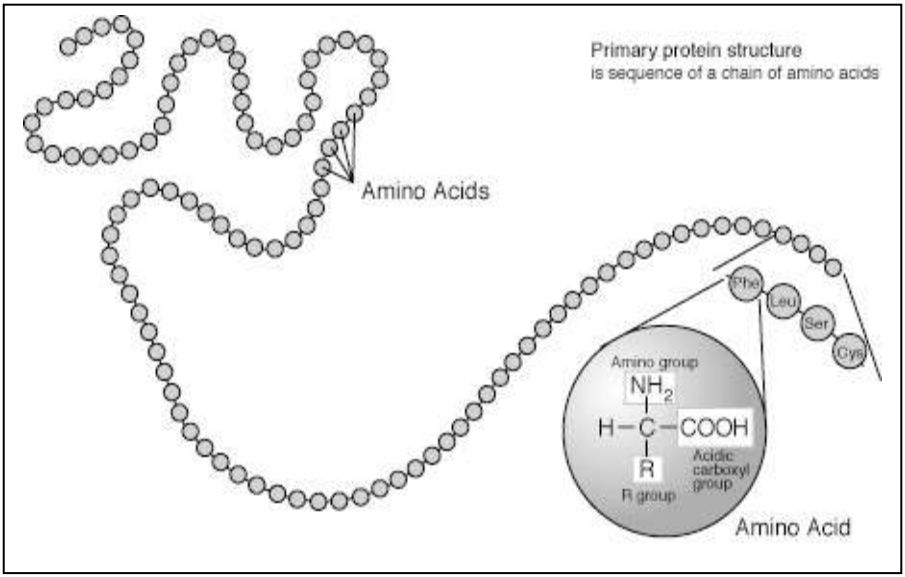

Figure 1: Primary protein structure

2. Secondary structure

As to the secondary structure, a spatial arrangement of amino acid residues exists in proximity to the sequence. Parts of the amino acid sequence are linked through hydrogen bonds. Polypeptide chains can fold into regular structures such as the α-helix, the beta sheet, and turns and loops. Two major forms of secondary structure are the α- and the β-strand. In the α-helix, the polypeptide chain is stabilized as a tightly packed rod. Within the helix, the CO group of each amino acid is hydrogen bonded to the NH group of the amino acid's four residues along the polypeptide chain. In the β-strand, the polypeptide chain is almost fully extended rather than being tightly wound as in the α-helix. Two or more β-strands linked by NH-to-CO hydrogen bonds unite to form β-sheets.

Most proteins have compact, globular shapes, requiring reversals in the direction of their polypeptide chains. Many of these reversals consist of a common structural element called the reverse turn. In many reverse turns, the CO group of residue *i* of a polypeptide is hydrogen bonded to the NH group of residue $i + 3$. This interaction adjusts to abrupt changes in the direction of the polypeptide chain. In other cases, structures that are more elaborate are responsible for chain reversals.

These structures are called loops, reflecting their overall shape. Unlike α-helices and β-strands, loops do not consist of regular, periodic structures. However, loop structures are often rigid and well defined. Turns and loops are consistently found on the surfaces of proteins and thus often play an important role in interactions between proteins and other molecules. The protein compound complex is typically the major focus of research related to drug development. Drug design only succeeds if the administered pharmaceuticals efficiently bind to the targeted protein. The commonly occurring secondary structure of a β-sheet can be illustrated as follows:[34]

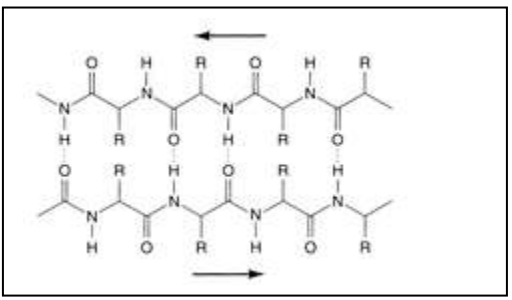

Figure 2: Secondary Structure of a β-sheet

3. Tertiary structure

The tertiary structure is the single most important determinant of the protein's biological function.[35] It refers to the folding of the basic protein, which is stabilized by interactions between the amino acid side chains on a single polypeptide forming the three dimensional molecule. Water-soluble proteins fold into compact structures with a nonpolar interior. Their asymmetric structures have two properties in common: (1) a core built of amino acids with hydrophobic side chains and (2) a surface formed largely of hydrophilic amino acids that interact with the aqueous[36] environment. The leading force behind the formation of the tertiary structure of water-soluble proteins is the hydrophobic interaction between the interior residues. Some proteins that exist in a hydrophobic environment in membranes display the inverse distribution of hydrophobic and hydrophilic amino acids. Within those proteins, the hydrophobic amino acids are on the surface to interact with the environment, whereas the hydrophilic groups are shielded from the environment in the interior of the protein.

34 Based on a figure provided by Wikipedia, available at: http://en.wikipedia.org/wiki/Beta_sheet, last checked on January 21, 2008.
35 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 39.
36 A solution in which the solvent is water.

The folding of most proteins is complex and devoid of symmetry. A unifying principle becomes apparent from the distribution of side chains. The physical analysis of myoglobin[37] provided the first 3-D picture of a protein. Many of the basic rules governing tertiary structure rely on this discovery. In myoglobin, approximately 70% of the main chain is folded into eight α-helices and much of the remaining amino acids form turns and loops between helices. The interior consists almost entirely of nonpolar residues such as leucine, valine, methionine, and phenylalanine. Charged residues such as aspartate, glutamate, lysine, and arginine are absent from the inside of myoglobin. Only two polar residues reside inside the protein. Both are histidine residues and play critical roles in binding iron and oxygen. The outside of myoglobin consists of both polar and nonpolar residues.[38] There are approximately 200 different structures, including mutants of myoglobin.[39] Meanwhile, physical proteomics technologies, such as hX-ray crystallography and NMR approaches described below, have revealed the detailed three-dimensional structures of thousands of proteins.[40] To understand the protein's function, it is of fundamental importance to define the 3-D folding type.[41] An illustration of the 3-D folding structure of myoglobin is shown in Figure 3[42]:

37 The protein myoglobin is the oxygen carrier in muscles.

38 Polypeptides containing more than one polypeptide chain exhibit a fourth level of structural organization ("Quaternary structure"). Each polypeptide chain in such a protein is called a subunit. Quaternary structure describes the spatial arrangement of subunits and the nature of their interaction and can be as simple as two identical subunits or as complex as dozens of different subunits. In most cases, the subunits are held together by noncovalent bonds. One identical subunit organization is present in the DNA-binding protein CRo found in a bacterial virus called λ. A more complicated quaternary structure is human hemoglobin, the oxygen-carrying protein in blood, which consists of two subunits of one type and two subunits of another type. See Berg, Jeremy M./Tymoczko, John L./Stryer Lubert, Biochemistry, New York 2002, 63-64.

39 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 68.

40 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 346.

41 Matthias Mann, director of the Center for Experimental BioInformatics (CEBI) at the University of Southern Denmark, in his opening remarks at the HUPO 4th Annual World Congress, "From defining the proteome to understanding the function", held from August 29 to September 1, 2005 in Munich.

42 The figure is based on the represenatation provided by Wikipedia, available at http://en.wikipedia.org/wiki/Protein, last checked on January 21, 2008. Myoglobin was the the first protein structure revealed by X-ray crystallography. Max Perutz and Sir John Cowdery Kendrew discovered its structure in 1958; both men later received the Nobel Prize in Chemistry.

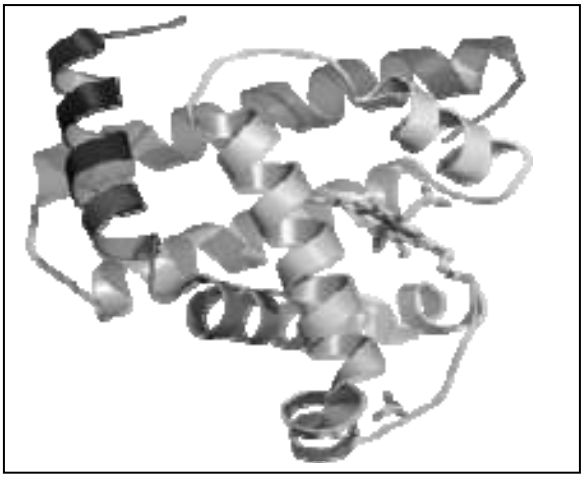

Figure 3: 3-D folding structure of myoglobin

III. Protein folding

1. Folding funnel theory of protein folding

How is a protein able to fold reliably into a predictable conformation? How can the mechanism be described in which the protein is carried from its unfolded random coil to a uniquely folded metastable state? Biochemical studies found that denatured proteins have all of their native three-dimensional structure disrupted. Yet, many of them refold efficiently and completely recover their biological activity when placed under conditions in which the folded form of the protein is stable.[43] Therefore, it is assumed that a native protein exists in some kind of thermodynamic configurational equilibrium. The biologically active state is the one with the lowest configurational energy.[44] The sequence of events guiding the protein folding is called the "protein folding pathway".[45] A random search among the entire conformation space for conformers would require an enormously long time.[46] Proteins, however, are able to

43 Berg, Jeremy M./Tymoczko, John L./Stryer Lubert, Biochemistry, New York, NY, 2005, 72-73.

44 Levinthal, Cyrus, Are there Pathways for Protein Folding? 65 Journal de Chimie Physique 1968, 44, 44.

45 Levinthal, Cyrus, Are there Pathways for Protein Folding?, 65 Journal de Chimie Physique 1968, 44, 44.

46 Although the protein is able to sample new configurations very fast, it will take at least 1027 years to try them all, see Zwanzig, R./Szabo, A./Bagchi, B., Levinthal's paradox, 89 Proceed-