

1 The three-dimensional structure of proteins

1.1 Structure of the native state

The human body contains the astonishing number of several 100,000 different proteins. Proteins are “smart” molecules each fulfilling largely specific functions such as highly efficient catalysis of biochemical reactions, muscle contraction, physical stabilization of the body, transport of materials in body fluids, and gene regulation. In order to optimally fulfill these functions, highly specific protein structures have evolved. The performance of humans, animals, and plants crucially depends on the integrity of these structures. Already small structural errors can cause diminishings of performance or even lethal diseases.

Proteins generally consist of thousands of atoms, such as hydrogen (H), carbon (C), nitrogen (N), oxygen (O), and sulfur (S). The van-der-Waals radii are about 1.0–1.4 Å for H, 1.6–2.1 Å for $-\text{CH}_3$, 1.4–1.8 Å for N, 1.4–1.7 Å for O, and 1.7–2.0 Å for S. Typical sizes of proteins range from a few nm to 200 nm. Since representations with atomic resolution of the whole molecule (Fig. 1.1a), or only its backbone (Fig. 1.1b), would be quite confusing for most proteins, it has become common to represent the protein structure as a ribbon of the backbone (Fig. 1.1c).

Multiple levels of structure are distinguished (see Nölting, 2005): The most basic is the primary structure which is the order of amino acid residues. The 20 common amino acids found in proteins can be classified into 3 groups: nonpolar, polar, and charged. Some physical properties of amino acids are given in Table 1.1. For the hydrophobicity of amino acids see Nölting, 2005. A typical protein contains 50–1000 amino acid residues. An interesting exception is titin, a protein found in skeletal muscle, containing about 27,000 residues in a single chain. The next level, the secondary structure, refers to certain common repeating structures of the backbone of the polypeptide chain. There are three main types of secondary structure: helix, sheet, and turns. That which cannot be classified as one of these three types is usually called “random coil” or “other”. Long connections between helices and strands of a sheet are often called “loops”. The third level, the tertiary structure, provides the information of the three-dimensional arrangement of elements of secondary structure in a single protein molecule or in a subunit of a protein molecule. The tertiary structure of a protein molecule, or of a subunit of a protein molecule, is the arrangement of all its atoms in space, without regard to its relationship with neighboring molecules or subunits. As this definition implies, a protein molecule can contain multiple subunits. Each subunit

consists of only one polypeptide chain and possibly co-factors. Finally, the quaternary structure is the arrangement of subunits in space and the ensemble of its intersubunit contacts, without regard to the internal geometry of the subunits. The subunits in a quaternary structure are usually in noncovalent association. Rare exceptions are disulfide bridges and chemical linkers between subunits.

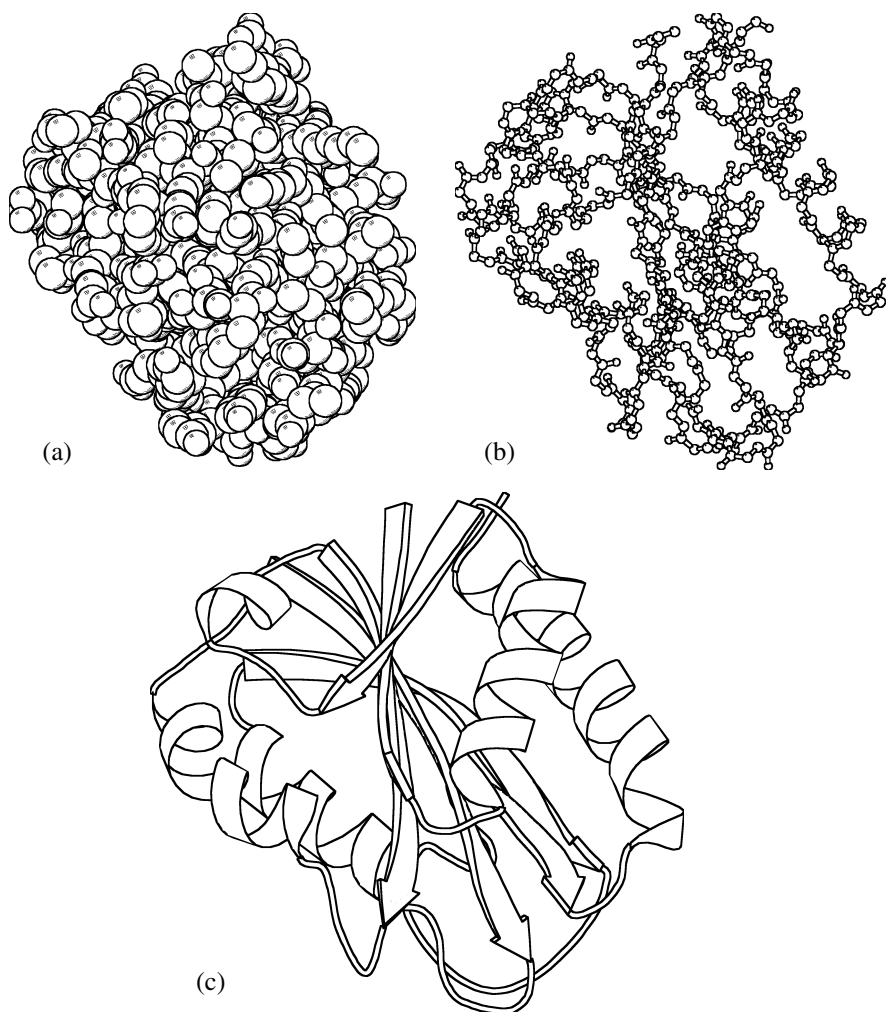


Fig. 1.1 The three-dimensional structure of the saddle-shaped electron transport protein flavodoxin from *Escherichia coli* (Hoover and Ludwig, 1997). (a) Space-filling representation of the complete molecule. (b) Ball-and-stick representation of the protein backbone. (c) Ribbon representation: ribbons, arrows, and lines symbolize helices, strands, and other, respectively. Coordinates are from the Brookhaven National Laboratory Protein Data Bank (Abola et al., 1997). The figure was generated using MOLSCRIPT (Kraulis, 1991)