STRUCTURE AND FUNCTION OF PROTEINS

Julio F. Turrens
Pat Capps Covey College of Allied Health Professions, University of South Alabama, Mobile, AL 36688, USA

Keywords: Amino acids, Electrophoresis, Protein chemistry, Protein purification techniques, proteome.

Contents

Summary:

Most biological processes (movement, transport, metabolic processes, response to hormonal signals, etc.) are made possible by proteins. These fascinating molecules are polymers of alpha-amino acids joined by amide bonds (peptide bonds) in a predetermined sequence. Although proteins may contain hundreds of amino acids, there are only twenty different amino acids used in protein synthesis. Therefore, individual amino acids are repeated several times in the sequence. Newly synthesized proteins fold to form unique tridimensional structures which are responsible for their biological role.

In this chapter we describe the different levels of organization within a protein that contribute to their final tridimensional structure. We also describe various techniques used in purification, identification and sequencing of proteins.

1. Overview

The enormous complexity of biological systems is, to a large extent, a result of interactions among four different types of biomolecules: nucleic acids, carbohydrates, lipids and proteins. In addition, cells contain a variety of small molecules including ions, vitamin-derived co-enzymes, etc. Of these four types of biomolecules, proteins are by far the most diverse structures and, as result, they are able to carry out a variety of functions. Indeed, proteins are responsible for all of the following processes:

Transport: Some proteins are designed to bind and move specific compounds either inside the cell or throughout an entire organism. Examples of transport proteins include hemoglobin and myoglobin (binding oxygen), albumin (binding and transporting free fatty acids and other non-polar molecules), and transferrin (binding and carrying iron in the blood).

Catalysis: All biochemical processes involve a catalyst to allow reactions to occur under physiological temperatures and pH. With a few exceptions, these catalysts are always proteins (known as enzymes). There are thousands of enzymes with each typically catalyzing only a single reaction.

Structural roles: Some proteins have important structural roles, providing support or contributing to specific three-dimensional structures in biological systems. Examples of structural proteins include collagen (found in cartilages and bones), microtubules and microfilaments (cytoskeleton), etc.

Movement: Proteins such as actin and myosin are responsible for the contraction of muscle fibers as well as other structures such as flagella, responsible for the movement of microorganisms.

Decoding information: Specific proteins involved in gene expression are involved in recognition of promoter sequences in DNA and specific RNA sequences during translation.

Immune response: Antibodies may recognize and bind specific structures, starting a process that can lead to the elimination of foreign organisms and also trigger tissue rejection.

Proteins as toxins and other disease causing agents: There are many examples of proteins responsible for causing severe diseases. For example, cholera and diphtheria toxins are actual enzymes that catalyze the covalent modifications of critical proteins. In the case of cholera toxin, the target is the chloride channels in the intestine. Diphtheria toxin impacts protein synthesis in target cells. Prions, on the other hand, are proteins responsible for spongiform diseases (such as Kuru, Mad Cow disease and Creutzfeld-

Jacobs disease) using a mechanism that involves inducing conformational changes in existing proteins.

No other type of biomolecule has such diverse functions. In this review we will explore the structural features that force proteins to adopt only one three-dimensional configuration among many possible shapes.

2. Amino Acid Structure

Proteins are polymers of alpha-amino acids, which are organic molecules containing an amino group, a hydrogen atom and a side chain attached to the alpha carbon (the carbon positioned next to a carboxyl group, Figure 1).

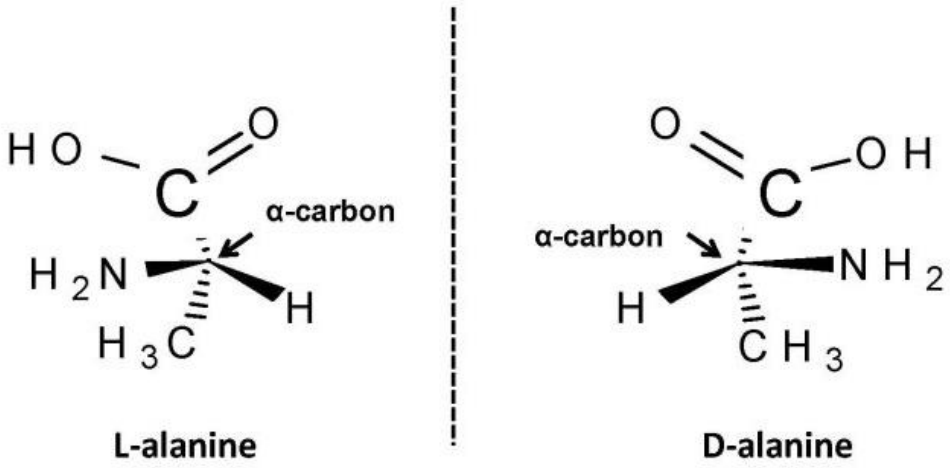
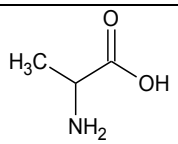
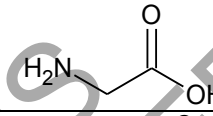
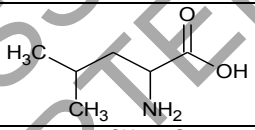
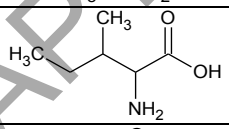
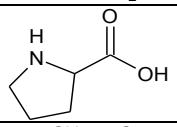
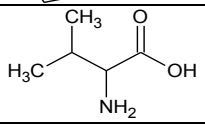
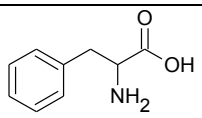
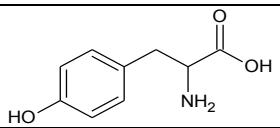
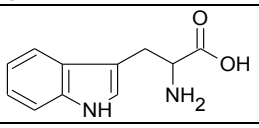
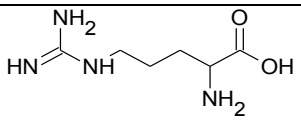


Figure 1. In organic acids, the carbon next to the acid group is also called the α-carbon. In all amino acids the alpha carbon contains an amino group, an organic side chain and a H atom. Since in 19 of the twenty amino acids alpha carbon has four different substituents, this carbon is a chiral carbon and may adopt two different stereo-configurations (L and D). L and D isomers of amino acids (in this case alanine) are mirror image of each other. Only L amino acids are used in the synthesis of proteins. (The chemical structures were drawn using ACD/ChemSketch Freeware, version 12.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2012).

Amino acids are bound to each other through amide bonds between the carboxyl group of one amino acid and the amino group of the next one. There are only twenty different amino acids used for protein synthesis, each one featuring a different side chain. Amino acid side chains include organic structures covering a vast array of chemical properties: saturated hydrocarbons (aliphatic), aromatic, basic, acidic, and, two side chains containing sulfur atoms (Table 1).

In addition to their common names, amino acids may be identified using two different types of abbreviations. When referring to individual amino acids, it is common to use a three-letter abbreviation which derives from the name of a particular amino acid (Table 1). In some cases we may need to list the entire sequence of amino acids in a given

protein which may include hundreds of them. In this case it is standard to use a one letter abbreviation. Some of the one letter abbreviations match the first letter of the particular amino acid, but because in several cases more than one amino acid has the same initial, alternative letters are required (Table 1). For example, the one letter abbreviation for the amino acid threonine is "T" while the abbreviation for tyrosine is "Y" and for tryptophan is "W".

Name	3 letter abb.	1 letter abb.	Structure
Aliphatic			
Alanine	Ala	A	
Glycine	Gly	G	
Leucine	Leu	L	
Isoleucine	Ile	I	
Proline	Pro	P	
Valine	Val	V	
Aromatic			
Phenylalanine	Phe	F	
Tyrosine	Tyr	Y	
Tryptophan	Trp	W	
Basic			
Arginine	Arg	R	

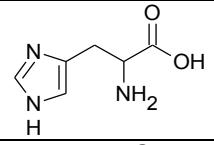
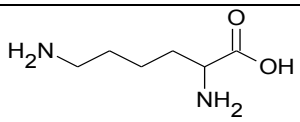
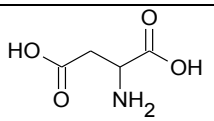
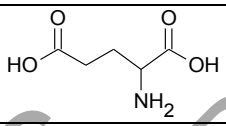
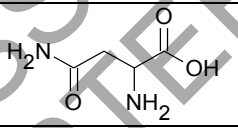
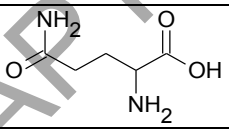
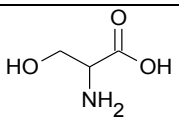
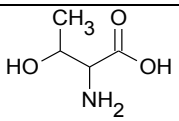
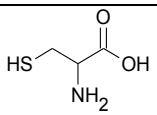
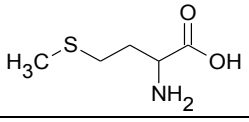
Histidine	His	H	
Lysine	Lys	K	
Acidic			
Aspartic acid	Asp	D	
Glutamic acid		E	
Amides			
Asparagine	Asn	N	
Glutamine	Gln	Q	
Non-aromatic alcohols			
Serine	Ser	S	
Threonine	Thr	T	
Sulfur-containing amino acids			
Cysteine	Cys	C	
Methionine	Met	M	

Table 1. Names and structures of the twenty amino acids found in proteins. The structures depicted in the table were downloaded from ACD/ChemSketch Freeware, version 12.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2012.

With the exception of the smallest of all amino acids (glycine, which has a second hydrogen in place of the side chain), the alpha carbon in the remaining nineteen structures is a chiral carbon with four different substituents. Chiral carbons may exist in two different optically active stereo configurations (D- and L-), which are mirror images

of each other (Figure 1). When plane-polarized light runs through a solution contacting molecules with chiral carbons, the axis of polarization rotates to the right or to the left, depending on the steric configuration of those carbons. The letters D and L come from the Latin words "Dextro" (right) and "Levo" (left) and in the case of glyceraldehyde the prefix indicates whether the stereoisomer will rotate a beam of polarized light to the right or the left. With more complex molecules it is impossible to predict the direction or magnitude of the rotation. Still, the two possible stereoisomers of amino acids are also identified as D- or L- based on their similarities to the structures of D- and L-glyceraldehyde (Figure 1), but in this case the prefixes are not indicative of the direction of rotation of polarized light. In other words, D and L only used to describe each of the two possible arrangements of the four substituents in the alpha carbon. Of the two possible steroisomers, only L-amino acids are used for the synthesis of proteins.

The physiological role of a given protein is a direct consequence of their three-dimensional structure and the chemical nature of each amino acid in the polymer. For example, binding of an antibody to its antigen is a result of the chemical nature of several amino acids located in the "binding site", which tightly associate with the antigen (sometimes described as a key-lock type of interaction). Likewise, the catalytic activity of an enzyme requires specific amino acids to be present in the "active site" of the protein, where the reaction will occur. In some enzymes these amino acids become directly involved in binding substrates providing an alternative path for the reaction. In other cases, the amino acids may be required to hold metals or additional organic molecules (prosthetic groups) which become involved in the reaction catalyzed by the enzyme. The amino acids needed for the reaction to occur do not need to be positioned near each other in the polymer, but become positioned in close proximity to each other as a result of a complex folding of the overall protein. This final three-dimensional configuration is a result of various levels of organization which lead to the final structure. These levels of structural organization are called primary, secondary and tertiary structures. In the case of oligomeric proteins (structures with more than one polypeptide chain), there is also a quaternary structure that describes the association among subunits.

3. Description of Each Level of Structural Organization

3.1. Primary Structure: The Peptide Bond

By definition, the primary structure of a protein is the sequence of amino acids in the order in which they are added to the polypeptide chain during protein synthesis. This sequence is unique to every protein and is coded as a sequence of nucleotides in DNA. As proteins are synthesized (translation) amino acids are attached to each other in a linear sequence through amide bonds between the carboxyl group of one amino acid and the amino group of another. These amide bonds between amino acids are called "peptide bonds" and have unique properties that will be described below. As protein synthesis starts, the carboxyl group of the first amino acid binds to the amino group of the second amino acid in the sequence, and so on. In the final polymer, the first amino acid will retain a free amino group (therefore becomes known as the N-terminus) while the other end of the polymer will have a free carboxyl group (called the C-terminus). By definition, the primary structure of a protein is the sequence of amino acids, starting

from the N-terminus (amino acid number one) and ending with the amino acid located in the C-terminus.

Peptide bonds are more stable than regular amide bonds due to the resonance of the double bond in the amide between carbon and oxygen and carbon and nitrogen (Figure 2). The bond between C and N is a double bond about 40% of the time, stabilizing the peptide bond and making it resistant to hydrolysis at physiological pH. In the absence of hydrolytic enzymes (peptidases), peptide bond hydrolysis can only occur under very acidic conditions (for example, 6 M hydrochloric acid). At pH 7.0, the half life of a peptide bond is about 1000 years.

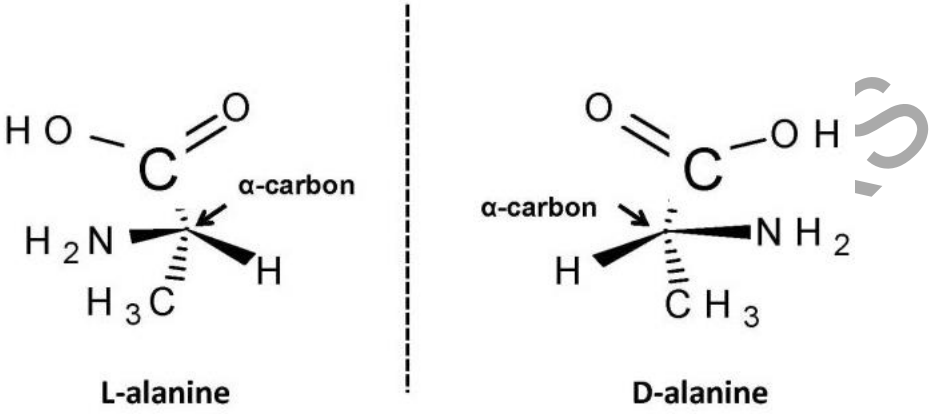


Figure 2. Resonance of the double bond in the carbonyl group with the C-N bond makes the peptide bond is rather planar limiting the free rotation of atoms around this bond.

This double bond prevents rotation around this link and also gives it a planar configuration. Of the two possible configurations (cis or trans), steric hindrance forces the alpha carbons to be positioned "trans" from each other (Figure 2). In summary, by limiting free rotation, the double bond character of the peptide bond makes the polypeptide chain behave like a series rather rigid planar structures connected through alpha carbons.

This rigidity is further exacerbated by the limitations imposed on rotation around the bonds between the alpha carbon and either the amino group and the carbonyl group. The angle of rotation between the alpha carbon and the amino group is called *phi* (φ) while the angle of rotation between the alpha carbon and the carbonyl group is called psi (ψ), and the rotation around these bonds is limited by steric hindrance and/or electrostatic repulsion among other groups in the chain (Figure 3). For φ, the angle is measured between two consecutive carbonyls along the polypeptide in the sequence C(O)-NH-C(α)-C(O). The ψ angle is measured between two consecutive amino groups in the sequence NH-C(α)-C(O)-NH. The value of these angles is positive if the groups rotate clockwise when looking from either the N (φ) or the carbonyl (ψ) towards the alpha carbon. Gopalasamundran Ramachandran studied the limitations of rotation around these bonds using chemical models and generated a graph that predicted the limit values for both φ and ψ (Ramachandran plots). The results show that about three quarter of all the possible combinations of φ and ψ are forbidden due to steric hindrance (Figure 3).

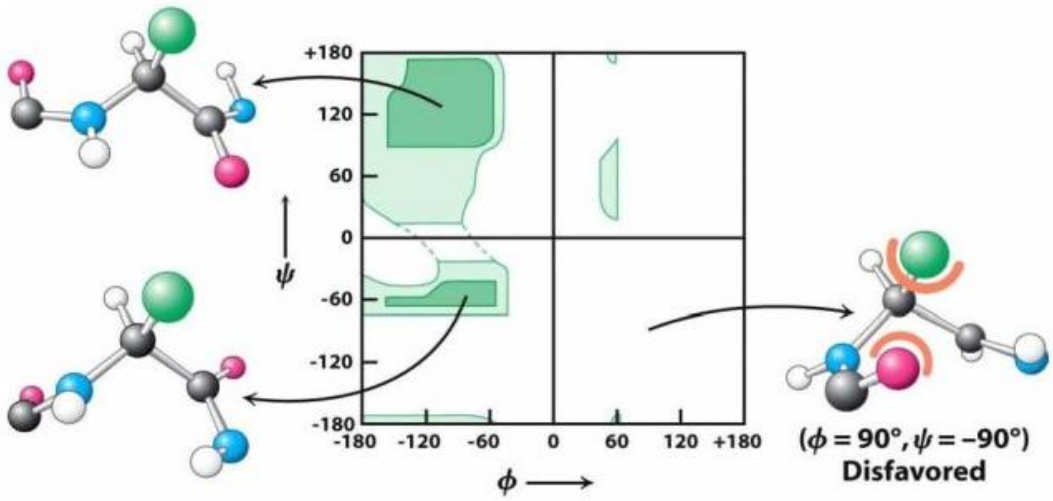


Figure 3. Ramachandran diagram showing different combinations of φ and ψ angles allowed in proteins. For φ, the angle is measured between two consecutive carbonyls (the O atoms is shown in red) along the polypeptide in the sequence C(O)-NH-C(α)-C(O). The ψ angle is measured between two consecutive amino groups (the N atom is shown in blue) in the sequence NH-C(α)-C(O)-NH. The darker areas represent the most frequent combinations. The configurations represented in the top left quadrant are most often observed in secondary structures known as beta-configurations while those represented the configuration in the bottom left quadrant are more common in alpha helices (details under "secondary structure" in the text). (The figure is reprinted from *Biochemistry: A Short Course*, 2nd edition, by Tymoczko et al., © 2013, 2010 by W.H. Freeman and Company. Used with the permission of W.H. Freeman and Company.)

In summary, although all the atoms in the protein backbone are linked through single bonds, in reality there is not much room for free rotation around any of those bonds. First, in the peptide bonds there is no free rotation between the carbonyl group and the nitrogen because of the double bond character of the C-N link, with the alpha carbons position "trans" from each other relative to the double bond. Second, there is limited free rotation around the alpha carbons because of electrostatic repulsion by carbonyl groups. Even at the level of the primary structure polypeptides are rather rigid, a factor that will contribute to the unique three-dimensional structure of a protein.

-

-

-

Bibliography

Buxmabum E. (2007) *Fundamentals of Protein Structure and Function.* Springer, ISBN-13: 978-0-387-26352-6 [This is a comprehensive text describing various aspects of protein chemistry and functions. The book is appropriate for graduate and undergraduate students].

Corley R. (2010) *A guide to Methods in Biomedical Sciences*, Springer, ISBN 1441935606, 9781441935601. [This book describes common methods used in biomedical research, including techniques used for protein purification].

Onuchic J.N. and Wolynes P. (2004) Theory of protein folding. *Current Opinion in Structural Biology*, **14,** 70-75. [This paper discusses the principles control the way a protein folds in order to reach its final tridimensional structure].

Protein Data Bank, www.rcsb.org/. [Protein database (PDB) is a free portal to a variety of information about protein structure. In addition to relevant articles and information, users may search for different proteins and download sequences, tridimensional structures, etc.].

Biographical Sketch

Dr. Julio F. Turrens received his PhD in Biological Chemistry from the University of Buenos Aires, Argentina. He continued his post-doctoral training at Duke University (Durham, NC) and Johns Hopkins University (Baltimore, MD). He is currently professor of Biomedical Sciences, Associate Dean and Director of Graduate Studies in the Pat Capps Covey College of Allied Health Professions at the University of South Alabama, in Mobile, AL. He teaches courses in biochemistry, topics in bioethics and responsible conduct of research. He has over 100 peer-reviewed publications, book chapters and abstracts in the areas of free radical metabolism in mammalian cells, and intermediate metabolism of parasitic protozoa.