5.0 Introduction to Proteins

Objectives

1. Know the structures and be able to draw and name all AA's
2. Know the pKa ranges of all AA's
3. Understand ionization and be able to calculate isoelectric points (pI's) of oligopeptides
4. Read how the nonstandard AA's are made and used
5. Know the polypeptide formation and understand the polypeptide diversity
6. Understand protein purification techniques
7. Understand protein sequencing
8. Read about the protein evolution

5.1 Amino Acids

During the cellular translation process, mRNA is decoded into a sequence of amino acids, which may go on to polymerize and form full proteins. These 20 common (and 2 rare) amino acids are coded for in our genetic code. Amino acids come in three predominate forms, nonpolar, polar-charged, and polar-uncharged:

	Name	Abbreviations	pKa1 (-COOH)	pKa2 (-NH3)	pI
Nonpolar Aliphatic	Glycine	Gly, G	2.34	9.60	5.97
	Alanine	Ala, A	2.23	9.69	6.00
	Valine	Val, V	2.32	9.62	5.96
	Leucine	Leu, L	2.36	9.60	5.98
	Isoleucine	Ile, I	2.36	9.60	6.02

	Name	Abbreviations	pKa1 (-COOH)	pKa2 (-NH3)	pI
Nonpolar	Proline	Pro, P	1.99	10.60	6.30
	Methionine	Met, M	2.28	9.21	5.74

	Name	Abbreviations	pKa1 (-COOH)	pKa2 (-NH3)	pI
Nonpolar Aromatic	Phenylalanine	Phe, F	1.83	9.13	5.48
	Tyrosine	Tyr, Y	2.20	9.11	5.66
	Tryptophan	Trp, T	2.83	9.39	5.89

	Name	Abbreviations	pKa1 (-COOH)	pKa2 (-NH3)	pI
Polar	Serine	Ser, S	2.21	9.15	5.68
	Cysteine	Cys, C	1.96	10.28	5.07
	Threonine	Thr, T	2.09	9.10	5.60
	Asparagine	Asn, N	2.02	8.80	5.41
	Glutamine	Gln, Q	2.17	9.13	5.65

	Name	Abbreviations	pKa1 (-COOH)	pKa2 (-NH3)	pI
Polar (+)	Histidine	His, H	1.82	9.17	7.59
	Lysine	Lys, K	2.18	9.21	5.74
	Arginine	Arg, R	2.17	9.04	10.76

	Name	Abbreviations	pKa1 (-COOH)	pKa2 (-NH3)	pI
Polar (-)	Aspartic Acid	Asp, D	1.88	9.60	2.77
	Glutamic Acid	Glu, E	2.19	9.67	3.22

In the table above, pI represents the isoelectric point, the pH at which the molecule has a neutral overall charge and the zwitterion form is dominant.

The isoelectric point can be calculated for oligopeptides by assuming each peptide is fully protonated (pH=0) and incrementing charges on each group as the pH is raised and groups are deprotonated until the net charge is zero.

In addition to the amino acids listed above, there is a group of "nonstandard" amino acids that can be created in other organisms, as well as post-translational modification of the standard amino acids.

5.2 Peptides and the Peptide Bond

Amino acids may be convalently bonded to each other via amide, or peptide bonds, in structures known as polypeptides. These bonds form between the α-carboxylic acid group in one amino acid and the α-amino group in another.

5.3 Proteins: Polypeptides of Defined Sequence

When polypeptides reach a defined length and sequence they're referred to as proteins. This sequence of peptides is defined as the primary structure of the protein.

Determining proteins requires analysis which was tranditionally accomplished using protein purification methods such as anion exchange chromatography (IEC) and size exclusion chromatography (SEC). These techniques use features of the proteins, such as pI, polarity, solubility, and more, to purify proteins for further analysis.

5.4 From Gene to Protein

Proteins are ultimately determined by genetic sequences in a series of well-known steps. DNA is transcribed into sequences of mRNA, which is then translated into amino acids. These amino acids form peptides and polypeptides, eventually forming proteins in the body.

5.5 From Gene to Sequence to Protein Function

Though biochemistry has been historically more concerned with studying functions and structures of macromolecules, disciplines such as systems biology have recently started delving deeper into the function of proteins.

In the body, thousands of molecules may form complex interactions. In contemporary biochemistry, computers are being used to create simulations of these interactions to help determine the functions of proteins and gene sequences.

5.6 Protein Sequence Homology

Protein sequencing is a process that evaluates the total amino acid composition and other metrics to determine relationships between proteins and genetic sequences. Sequences are reconstructed from polypeptides that have been degraded using Edman Degradation followed by a polypeptide cleavage procedure.

Protein sequences are also used to determine evolutionary relationships between organisms. By comparing distributions of proteins, genetic similarities can be deduced that may provide information about evolutionary ancestry. Phylogenetics is the science concerned with mapping these relationships.