6 Amino Acids in Proteins

Proteins have many functions in the body. They serve as transporters of hydrophobic compounds in the blood, as cell adhesion molecules that attach cells to each other and to the extracellular matrix, as hormones that carry signals from one group of cells to another, as ion channels through lipid membranes, and as enzymes that increase the rate of biochemical reactions. The unique characteristics of a protein are dictated by its linear sequence of amino acids, termed its primary structure. The primary structure of a protein determines how it can fold and how it interacts with other molecules in the cell to perform its function. The primary structures of all of the diverse human proteins are synthesized from 20 amino acids arranged in a linear sequence determined by the genetic code.

General properties of amino acids. Each of the amino acids used for protein synthesis has the same general structure (Fig. 6.1). It contains a carboxylic acid group, an amino group attached to the α -carbon in an L configuration, a hydrogen atom, and a chemical group called a side chain that is different for each amino acid. In solution, the free amino acids exist as zwitterions, ions in which the amino group is positively charged and the carboxylate group is negatively charged. In proteins, these amino acids are joined into linear polymers called polypeptide chains through peptide bonds between the carboxylic acid group of one amino acid and the amino group of the next amino acid.

Classification of amino acids according to chemical properties of the side chains. The chemical properties of the side chain determine the types of bonds and interactions each amino acid in a polypeptide chain can make with other molecules. Thus, amino acids are often grouped by polarity of the side chain (charged, nonpolar hydrophobic, or uncharged polar) or by structural features (aliphatic, cyclic, or aromatic). The side chains of the nonpolar hydrophobic amino acids (alanine, valine, leucine, isoleucine, phenylalanine, and methionine) cluster together to exclude water in the hydrophobic effect. The uncharged polar amino acids (serine, threonine, tyrosine, asparagine, and glutamine) participate in hydrogen bonding. Cysteine, which contains a sulfhydryl group, forms disulfide bonds. The negatively charged acidic amino acids (aspartate and glutamate) form ionic (electrostatic) bonds with positively charged molecules, such as the basic amino acids (lysine, arginine, and histidine). The charge on the amino acid at a particular pH is determined by the pK_a of each group that has a dissociable proton.

Amino acid substitutions in the primary structure. Mutations in the genetic code result in proteins with an altered primary structure. Mutations resulting in single amino acid substitutions can affect the functioning of a protein or can confer an advantage specific to a tissue or a set of circumstances. Many proteins such as hemoglobin exist in the human population as polymorphisms (genetically determined variations in primary structure.)

Within the same individual, the primary structure of many proteins varies with the stage of development and is present in *fetal* and *adult isoforms*, such as *fetal* and *adult hemoglobin*. The primary structure of some proteins, such as *creatine kinase*, can also vary between tissues (*tissue-specific isozymes*) or between intracellular locations in the same tissue. *Electrophoretic separation* of tissue-specific isozymes has been useful in medicine as a means of identifying the tissue site of injury.



The genetic code is the sequence of three bases (nucleotides) in DNA containing the information for the

linear sequence of amino acids in a polypeptide chain (its primary structure). A gene is the portion of DNA that encodes a functional product, such as a polypeptide chain. Mutations, which are changes in the nucleotides in a gene, result in a change in the products of that gene that may be inherited. The genetically inherited disease sickle cell anemia is, for example, caused by a mutation in the gene encoding one of the subunits of hemoglobin. Hemoglobin is the protein present in red blood cells that reversibly binds O₂ and transports it to tissues. The adult hemoglobin protein comprises four polypeptide chains, 2α and 2 β . The α and β subunits differ in primary structure (i.e., they have different sequences of amino acids and are encoded by different genes). Sickle cell anemia is caused by a mutation in DNA that changes just one amino acid in the hemoglobin β chains from a glutamic acid to a valine.



Modified amino acids. In addition to the amino acids encoded by DNA that form the primary structure of proteins, many proteins contain specific amino acids that have been modified by phosphorylation, oxidation, carboxylation, or other reactions. When these reactions are enzyme-catalyzed, they are referred to as post-translational modifications.



тне WAITING ROOM

Will Sichel is a 17-year-old boy who came to the hospital emergency room with severe pain in his lower back, abdomen, and legs, which began after a 2-day history of nausea and vomiting caused by gastroenteritis. He was diagnosed as having sickle cell disease at age 3 years and has been admitted to the hospital on numerous occasions for similar vaso-occlusive sickle cell crises.

On admission, the patient's hemoglobin level in peripheral venous blood was 7.8 g/dL (reference range = 12–16 g/dL). The hematocrit or packed cell volume (the percentage of the total volume of blood made up by red blood cells) was 23.4% (reference range = 41-53%). His serum total bilirubin level (a pigment derived from hemoglobin degradation) was 2.3 mg/dL (reference range = 0.2-1.0 mg/dL). An x-ray of his abdomen showed radiopaque stones in his gallbladder. With chronic hemolysis (red blood cell destruction), the amount of heme degraded to bilirubin is increased. These stones are the result of the chronic excretion of excessive amounts of bilirubin from the liver into the bile, leading to bilirubinate crystal deposition in the gallbladder lumen.

Cal Kulis is an 18-year-old boy who was brought to the hospital by his mother because of the sudden onset of severe pain in the left flank radiating around his left side toward his pubic area. His urine was reddish-brown in color, and his urinalysis showed the presence of many red blood cells. When his urine was acidified with acetic acid, clusters of flat hexagonal transparent crystals of cystine were noted. An x-ray of his abdomen showed radiopaque calculi (stones) in both kidneys. There was no family history of kidney stone disease.



Di Abietes, who has type 1 diabetes mellitus, was giving herself subcutaneous injections of insulin regular NPH beef insulin twice daily after her disease was first diagnosed (see Chapters 4 and 5). Subsequently, her

physician switched her to synthetic human insulin. At this visit, her physician changed her insulin therapy and has written a prescription for Humalog mix 75/25 (Eli Lilly, Indianapolis, IN), a mixture of Humalog in protamine suspension and unbound Humalog insulin (25%).



Ann Jeina is a 54-year-old woman who is 68 inches tall and weighs 198 lb. She has a history of high blood pressure and elevated serum cholesterol levels. After a heated argument with a neighbor, Mrs. Jeina experienced a "tight pressure-like band of pain" across her chest, associated with shortness of breath, sweating, and a sense of light-headedness.

After 5 hours of intermittent chest pain, she went to the hospital emergency room, where her electrocardiogram showed changes consistent with an acute infarction of the anterior wall of her heart. She was admitted to the cardiac care unit. Blood was sent to the laboratory for various tests, including the total creatine kinase (CK) level and the MB ("muscle-brain") fraction of CK in the blood.





The term calculus is used to describe any abnormal concretion (concrete-like precipitate) of mineral salts. These almost always form within the cavity of a hollow organ, such as the kidney (kidney or renal stones) or the lumen of a duct (e.g., common bile duct stones).

The term angina describes a crushing or compressive pain. The term angina pectoris is used when this pain is located in the center of the chest, often radiating to the neck or arms. The most common mechanism for the latter symptom is a decreased supply of oxygen to the heart muscle caused by atherosclerotic coronary artery disease, which results in obstruction of the vessels that supply aterial blood to cardiac muscle.



Fig. 6.2. Dissociation of the α -carboxyl and α -amino groups of amino acids. At physiologic pH (~7), a form in which both the α -carboxyl and α -amino groups are charged predominates. Some amino acids also have ionizable groups on their side chains.

The L-stereospecificity is often incorporated into the names of drugs derived from amino acids either as "L" or "levo" for the direction in which L amino acids rotate polarized light. For example, patients with Parkinson's disease are treated with L-DOPA (a derivative of the amino acid L-tyrosine), and hypothyroid patients are treated with levothyroxine (a different derivative of L-tyrosine)

I. GENERAL STRUCTURE OF THE AMINO ACIDS

Twenty different amino acids are commonly found in proteins. They are all α -amino acids, amino acids in which the amino group is attached to the α -carbon (the carbon atom next to the carboxylate group) (see Fig 6.1). The α -carbon has two additional substituents, a hydrogen atom and an additional chemical group called a side chain (-R). The side chain is different for each amino acid.

At a physiologic pH of 7.4, the amino group on these amino acids carries a positive charge, and the carboxylic acid group is negatively charged (Fig. 6.2). The pK_a of the primary carboxylic acid groups for all of the amino acids is approximately 2 (1.8–2.4). At pH values much lower than the pK_a (higher hydrogen ion concentrations), all of the carboxylic acid groups are protonated. At the pK_a, 50% of the molecules are dissociated into carboxylate anions and protons, and at a pH of 7.4, more than 99% of the molecules are dissociated (see Chapter 4). The pK_a for all of the α -amino groups is approximately 9.5 (8.8–11.0), so that at the lower pH of 7.4, most of the amino acid that has both a positive and a negative charge is called a zwitterion. Because these charged chemical groups can form hydrogen bonds with water molecules, all of these amino acids are watersoluble at physiologic pH.

In all of the amino acids but glycine, the α -carbon is an asymmetric carbon atom that has four different substituents and can exist in either the D or L configuration (Fig. 6.3). The amino acids in mammalian proteins are all L-amino acids represented with the amino group to the left if the carboxyl group is at the top of the structure. These same amino acids serve as precursors of nitrogen-containing compounds synthesized in the body, and thus human amino acid metabolism is also centered on L-amino acids. The amino acid glycine is neither D nor L because the α -carbon atom contains two hydrogen atoms.

The chemical properties of the amino acids give each protein its unique characteristics. Proteins are composed of one or more linear polypeptide chains containing



Fig. 6.3. L- and D-amino acids. The L forms are the only ones found in human proteins. Bonds coming out of the paper are shown by black arrows; those going in, by shaded arrows. The α -amino groups and H-atoms come toward the reader, and the α -carboxyl and side chains go away. The L and D forms are mirror images that cannot be superimposed by rotating the molecule. The reference for the L and D forms are the stereoisomers of glyceraldehyde.

hundreds of amino acids. The sequence of amino acids, termed the primary structure, is determined by the genetic code for the protein. In the polypeptide chains, amino acids are joined through peptide bonds between the carboxylic acid of one amino acid and the amino group of the adjacent amino acid (Fig. 6.4). Thus, the amino group, the α -carbon, and the carboxyl groups form the peptide backbone, and the side chains of the amino acids extend outward from this backbone. The side chains interact with the peptide backbone of other regions of the chain or with the side chains of other amino acids in the protein to form hydrophobic regions, electrostatic bonds, hydrogen bonds, or disulfide bonds. These interactions dictate the folding pattern of the molecule. The three-dimensional folding of the protein forms distinct regions called binding sites that are lined with amino acid side chains that interact specifically with another molecule termed a ligand (such as the heme in hemoglobin). Thus, the chemical properties of the side chains determine how the protein folds, how it binds specific ligands, and how it interacts with its environment (such as the aqueous medium of the cytoplasm).

The names of the different amino acids have been given three-letter and one-letter abbreviations (Table 6.1). The three-letter abbreviations use the first two letters in the name plus the third letter of the name or the letter of a characteristic sound, such as trp for tryptophan. The one-letter abbreviations use the first letter of the name of the most frequent amino acid in proteins (such as an "A" for alanine). If the first letter has already been assigned, the letter of a characteristic sound is used (such as an "R" for arginine). Single-letter abbreviations are usually used to denote the amino acids in a polypeptide sequence.

II. CLASSIFICATION OF AMINO ACID SIDE CHAINS

In Figure 6.5, the 20 amino acids used for protein synthesis are grouped into different classifications according to the polarity and structural features of the side chains. These groupings can be helpful in describing common functional roles or metabolic pathways of the amino acids. However, some amino acid side chains fit into a number of different classifications and are therefore grouped differently in different textbooks. Two of the characteristics of the side chain that are useful for classification are its pK_a and its hydropathic index, shown in Table 6.2 The hydropathic index is a scale used to denote the hydrophobicity of the side chain; the more positive the hydropathic index, the greater the tendency to cluster with other nonpolar molecules and exclude water in the hydrophobic effect. The more negative the hydropathic index of an amino acid, the more hydrophilic is its side chain.

A. Nonpolar, Aliphatic Amino Acids

Glycine is the simplest amino acid, and it really does not fit well into any classification because its side chain is only a hydrogen atom. Alanine and the branched chain amino acids (valine, leucine, and isoleucine) have bulky, nonpolar, aliphatic side chains. The high degree of hydrophobicitiy of the branched chain amino acid side chains is denoted by their high hydropathic index (see Table 6.2). Electrons are shared equally between the carbon and hydrogen atoms in these side chains, so that they cannot hydrogen bond with water. Within proteins, these amino acid side chains will cluster together to form hydrophobic cores. Their association is also promoted by van der Waals forces between the positively charged nucleus of one atom and the electron cloud of another. This force is effective over short distances when many atoms pack closely together.

The roles of proline and glycine in amino acid structure differ from those of the nonpolar amino acids. The amino acid proline contains a ring involving its α -carbon and its α -amino group, which are part of the peptide backbone. This rigid ring



Fig. 6.4. Peptide bonds. Amino acids in a polypeptide chain are joined through peptide bonds between the carboxyl group of one amino acid and the amino group of the next amino acid in the sequence.



According to some biochemists, "W" stands for the characteristic sound in "twyptophan" (see Table

 Table 6.1. Abbreviations for the Amino

 Acids

Acius						
Name	Abbreviations*					
	Three-Letter	One-Letter				
Alanine Arginine Asparagine Aspartate Cysteine Glutamate Glutamine Glycine Histidine Isoleucine Leucine Lysine	Three-Letter Ala Arg Asn Asp Cys Glu Gln Gly His Ile Leu Lys	A R N D C E Q G H H I L K				
Methionine Phenylalanine Proline Serine Threonine Tryptophan Tyrosine Valine	Met Phe Pro Ser Thr Trp Tyr Val	M F S T V V				

*Three-letter abbreviations are generally used. Oneletter abbreviations are used mainly to list the amino acid sequences of long protein chains.

The proteolytic digestive enzyme chymotrypsin cleaves the peptide bonds formed by the carboxyl groups of large, bulky uncharged amino acids. Which amino acids fall into this category?



Fig. 6.5. The side chains of the amino acids. The side chains are highlighted. The amino acids are grouped by the polarity and structural features of their side chains. These groupings are not absolute, however. Tyrosine and tryptophan, often listed with the nonpolar amino acids, are more polar than other aromatic amino acids because of their phenolic and indole rings, respectively.

Table 6.2. Properties of the Common Amino Acids

Amino Acid	р <i>К</i> а1* (Carboxyl)	р <i>К</i> _{а2} (Amino)	p <i>K</i> ₄ _R (R Group)	Hydropathy Index**
Nonpolar aliphatic				
Glycine	2.4	9.8		-0.4
Proline	2.0	11.0		-1.6
Alanine	2.3	9.7		1.8
Leucine	2.4	9.6		3.8
Valine	2.3	9.6		4.2
Isoleucine	2.4	9.7		4.5
Aromatic				
Phenylalanine	1.8	9.1		2.8
Tyrosine	2.2	9.1	10.5	-1.3
Tryptophan	2.4	9.4		-0.9
Polar uncharged				
Threonine	2.1	9.6	13.6	-0.7
Serine	2.2	9.2	13.6	-0.8
Aspargine	2.0	8.8		-3.5
Glutamine	2.2	9.1		-3.5
Sulfur-containing				
Cysteine	2.0	10.3	8.4	2.5
Methionine	2.3	9.2		1.9
Charged negative				
Aspartate	1.9	9.6	3.9	-3.5
Glutamate	2.2	9.7	4.1	-3.5
Charged positive				
Histidine	1.8	9.3	6.0	-3.2
Lysine	2.2	9.0	10.5	-3.9
Arginine	2.2	9.0	12.5	-4.5
Average	2.2	9.5		

*When these amino acids reside in proteins, the pK, for the side chains may vary to some extent from the value for the free amino acid, depending on the local environment of the amino acid in the threedimensional structure of the protein.

**The hydropathy index is a measure of the hydrophobicity of the amino acid (the higher the number, the more hydrophobic). Values based on Kyte J, Doolittle RF. J Mol Biol 1982;157:105B132.

causes a kink in the peptide backbone that prevents it from forming its usual configuration. Because the side chain of glycine is so small compared with that of other amino acids, it causes the least amount of steric hindrance in a protein (i.e., it does not significantly impinge on the space occupied by other atoms or chemical groups). Therefore, glycine is often found in bends or in the tightly packed chains of fibrous proteins.

B. Aromatic Amino Acids

The aromatic amino acids have been grouped together because they all contain ring structures with similar properties, but their polarity differs a great deal. The aromatic ring is a six-membered carbon-hydrogen ring with three conjugated double bonds (the benzene ring or phenyl group). The substituents on this ring determine whether the amino acid side chain engages in polar or hydrophobic interactions. In the amino acid phenylalanine, the ring contains no substituents, and the electrons are shared equally between the carbons in the ring, resulting in a very nonpolar hydrophobic structure in which the rings can stack on each other (Fig. 6.6). In tyrosine, a hydroxyl group on the phenyl ring engages in hydrogen bonds, and the side chain is therefore more polar and more hydrophilic. The more complex ring structure in tryptophan is an indole ring with a nitrogen that can engage in hydrogen bonds. Tryptophan is therefore also more polar than phenylalanine.

C. Aliphatic, Polar, Uncharged Amino Acids

Amino acids with side chains that contain an amide group (asparagine and glutamine) or a hydroxyl group (serine and threonine) can be classified as aliphatic, polar, uncharged amino acids. Asparagine and glutamine are amides of the amino



Chymotrypsin's highest activity is toward peptide bonds formed by the carboxyl groups of aromatic amino acids (phenylalanine, tyrosine, tryptophan). The side chains of these amino acids are all large and uncharged. One of the chymotrypsin isozymes also exhibits activity toward leucine and methionine, which are similar in polarity.



Fig. 6.6. Hydrophobic and hydrogen bonds. A. Strong hydrophobic interactions occur with the stacking of aromatic groups in phenylalanine side chains. B. Examples of hydrogen bonds in which a hydrogen atom is shared by a nitrogen in the peptide backbone and an oxygen atom in an amino acid side chain or between an oxygen in the peptide backbone and an oxygen in an amino acid side chain.



Cal Kulis passed a renal stone shortly after admission, with immediate relief of flank pain. Stone analysis showed its major component to be cystine. Normally, amino acids are filtered by the renal glomerular capillaries into the tubular urine but are almost entirely reabsorbed from this fluid back into the blood via transport proteins in the proximal tubular cells of the kidney. Cal Kulis has cystinuria, a genetically inherited amino acid substitution in the transport protein that normally reabsorbs cystine, arginine, and lysine from the kidney lumen back into the renal tubular cells. Therefore, his urine contained high amounts of these amino acids. Cystine, which is less soluble than other amino acids, precipitates in the urine to form renal stones (calculi).



Will Sichel has sickle cell anemia caused by a point mutation in his DNA that changes the sixth amino acid in the β -globin chain of hemoglobin from glutamate to valine. What difference would you expect to find in the chemical bonds formed by these two amino acids?

acids aspartate and glutamate. The hydroxyl groups and the amide groups in the side chains allow these amino acids to form hydrogen bonds with water, with each other and the peptide backbone, or with other polar compounds in the binding sites of the proteins (see Fig. 6.6). As a consequence of their hydrophilicity, these amino acids are frequently found on the surface of water-soluble globular proteins. Cysteine, which is sometimes included in this class of amino acids, has been separated into the class of sulfur-containing amino acids.

D. Sulfur-Containing Amino Acids

Both cysteine and methionine contain sulfur. The side chain of cysteine contains a sulfhydryl group that has a pK_a of approximately 8.4 for dissociation of its hydrogen, so cysteine is predominantly undissociated and uncharged at the physiologic pH of 7.4. The free cysteine molecule in solution can form a covalent disulfide bond with another cysteine molecule through spontaneous (nonenzymatic) oxidation of their sulfhydryl groups. The resultant amino acid, cystine, is present in blood and tissues, and is not very water-soluble. In proteins, the formation of a cystine disulfide bond between two appropriately positioned cysteine sulfhydryl groups often plays an important role in holding two polypeptide chains or two different regions of a chain together (Fig. 6.7). Methionine, although it contains a sulfur group, is a nonpolar amino acid with a large bulky side chain that is hydrophobic. It does not contain a sulfhydryl group and cannot form disulfide bonds. Its important and central role in metabolism is related to its ability to transfer the methyl group attached to the sulfur atom to other compounds.

E. The Acidic and Basic Amino Acids

The amino acids aspartate and glutamate have carboxylic acid groups that carry a negative charge at physiologic pH (see Fig. 6.5). The basic amino acids histidine,



Fig. 6.7. A disulfide bond. Covalent disulfide bonds may be formed between two molecules of cysteine or between two cysteine residues in a protein. The disulfide compound is called cystine. The hydrogens of the cysteine sulfhydryl groups are removed during oxidation.

lysine, and arginine have side chains containing nitrogen that can be protonated and positively charged at physiologic and lower pH values. Histidine has a nitrogen-containing imidazole ring for a side chain, lysine has a primary amino group on the 6th or ε carbon (from the sequence α , β , γ , δ , ε), and arginine has a guanidinium group.

The positive charges on the basic amino acids enables them to form ionic bonds (electrostatic bonds) with negatively charged groups, such as the side chains of acidic amino acids or the phosphate groups of coenzymes (Fig. 6.8). In addition, lysine and arginine side chains often form ionic bonds with negatively charged compounds bound to the protein binding sites, such as the phosphate groups in ATP. The acidic and basic amino acid side chains also participate in hydrogen bonding and the formation of salt bridges (such as the binding of an inorganic ion such as Na⁺ between two partially or fully negatively charged groups).

The charge on these amino acids at physiologic pH is a function of their pK_as for dissociation of protons from the α -carboxylic acid groups, the α -amino groups, and the side chains. The titration curve of histidine illlustrates the changes in amino acid structure occurring as the pH of the solution is changed from less than 1 to 14 by the addition of hydroxide ions (Fig. 6.9). At low pH, all groups carry protons; amino groups have a positive charge, and carboxylic acid groups have zero charge. As the pH is increased by the addition of alkali (OH⁻), the proton dissociates from the carboxylic acid group, and its charge changes from zero to negative with a pK_a of approximately 2, the pH at which 50% of the protons have dissociated.

The histidine side chain is an imidazole ring with a pK_a of approximately 6 that changes from a predominantly protonated positively charged ring to an uncharged ring at this pH. The amino group on the α -carbon titrates at a much higher pH (between 9 and 10), and the charge changes from positive to zero as the pH rises. The pH at which the net charge on the molecules in solution is zero is called the isoelectric point (pI). At this pH, the molecules will not migrate in an electric field toward either a positive pole (cathode) or a negative pole (anode), because the number of negative charges on each molecule is equal to the number of positive charges.

Amino acid side chains change from uncharged to negatively charged, or positively charged to uncharged as they release protons (Fig. 6.10). The acidic amino acids lose a proton from their carboxylic acid side chains at a pH of roughly 4, and are thus negatively charged at pH 7.4. Cysteine and tyrosine lose protons at their pK_as (~ 8.4 and 10.5, respectively), so their side chains are uncharged at physiologic pH. Histidine, lysine, and arginine side chains change from positively charged to neutral at their pK_as . The side chains of the two basic amino acids, arginine and lysine, have pK_a values above 10, so that the positively charged form always predominates at physiologic pH. The side chain of histidine ($pK_a \sim 6.0$) dissociates near physiologic pH, so only a portion of the histidine side chains carry a positive charge (see Fig. 6.9).

In proteins, only the amino acid side chains and the amino group at the amino terminal and carboxyl group at the carboxyl terminal have dissociable protons. All of the other carboxylic acid and amino groups on the α -carbons are joined in peptide bonds that have no dissociable protons. The amino acid side chains might have very different pK_as than those of the free amino acids if they are involved in hydrogen or ionic bonds with other amino acid side chains. The pK_a of the imidazole group of histidine, for example, is often shifted to a higher value between 6 and 7 so that it adds and releases a proton in the physiologic pH range.

III. VARIATIONS IN PRIMARY STRUCTURE

Although almost every amino acid in the primary structure of a protein contributes to its conformation (three-dimensional structure), the primary structure of a protein Glutamate carries a negative charge on its side chain at physiologic pH and thus can engage in ionic bonds or hydrogen bonds with water or other side chains. Valine is a hydrophobic amino acid and therefore tends to interact with other hydrophobic side chains to exclude water. (The effect of this substitution on hemoglobin structure is described in more detail in Chapter 7.)





Electrophoresis is a technique used to separate proteins on the basis of charge that has been extremely useful in medicine to identify proteins with different amino acid composition. The net charge on a protein at a certain pH is a summation of all of the positive and negative charges on all of the ionizable amino acid side chains plus the N-terminal amino and Cterminal carboxyl groups. Theoretically, the net charge of a protein at any pH could be determined from its amino acid composition by calculating the concentration of positively and negatively charged groups from the Henderson-Hasselbalch equation (see Chapter 4). However, hydrogen bonds and ionic bonds between amino acid side chains in the protein make this calculation unrealistic.



Fig. 6.9. Titration curve of histidine. The ionic species that predominates in each region is shown below the graph. pI is the isoelectric point (at which there is no net charge on the molecule).



Fig. 6.10. Dissociation of the side chains of the amino acids. As the pH increases, the charge on the side chain goes from 0 to + or from + to 0. The pK_a is the pH at which half the molecules of an amino acid in solution have side chains that are charged. Half are uncharged.

can vary to some degree between species. Even within the human species, the amino acid sequence of a normal functional protein can vary somewhat among individuals, tissues of the same individual, and the stage of development. These variations in the primary structure of a functional protein are tolerated if they are confined to noncritical regions (called variant regions), if they are conservative substitutions (replace one amino acid with one of similar structure), or if they confer an advantage. If many different amino acid residues are tolerated at a position, the region is called hypervariable. In contrast, the regions that form binding sites or are critical for forming a functional three-dimensional structure are usually invariant regions that have exactly the same amino acid sequence from individual to individual, tissue to tissue, or species to species.

A. Polymorphism in Protein Structure

Within the human population, the primary structure of a protein may vary slightly among individuals. The variations generally arise from mutations in DNA that are passed to the next generation. The mutations can result from the substitution of one base for another in the DNA sequence of nucleotides (a point mutation), from deletion or insertions of bases into DNA, or from larger changes (see Chap.14). For many alleles, the variation has a distinct phenotypic consequence that contributes to our individual characteristics, produces an obvious dysfunction (a congenital or genetically inherited disease), or increases susceptibility to certain diseases. A defective protein may differ from the most common allele by as little as a single amino acid that is a nonconservative substitution (replacement of one amino acid with another of a different polarity or very different size) in an invariant region. Such mutations might affect the ability of the protein to carry out its function, catalyze a particular reaction, reach the appropriate site in a cell, or be degraded. For other proteins, the variations appear to have no significance.

Variants of an allele that occur with a significant frequency in the population are referred to as polymorphisms. Thus far in studies of the human genome, almost one third of the genetic loci appear to be polymorphic. When a particular variation of an allele, or polymorphism, increases in the general population to a frequency of over Is the substitution of a glutamate for a valine in sickle cell hemoglobin a conservative replacement? What about the substitution of an aspartate for a glutamate?

For the most part, human chromosomes occur as homologous pairs, with each member of a pair containing the same genetic information. One member of the pair is inherited from the mother and one from the father. Genes are arranged linearly along each chromosome. A genetic locus is a specific position or location on a chromosome. Alleles are alternate versions of a gene at a given locus. For each locus (site), we have two alleles of each gene, one from our mother and one from our father. If both alleles of a gene are identical, the individual is homozygous for this gene; if the alleles are different, he is heterozygous for this gene. Will Sichel has two identical alleles for the sickle variant of the β -globin gene that results in substitution of a valine for a glutamate residue at the sixth position of the β -globin chain. He is, therefore, homozygous and has sickle cell anemia. Individuals with one normal gene and one sickle cell allele are heterozygous. They are carriers of the disease and have sickle cell trait.

Will Sichel's hemoglobin, HbS, comprises two normal α chains and 2 β -globin chains with the sickle cell variant ($\alpha_2\beta_2^s$). The change in amino acid composition from a glutamate to a valine in the β chain allows sickle hemoglobin to be separated from normal adult hemoglobin (HbA, or ($\alpha_2\beta_2^A$)) by electrophoresis. In electrophoresis, an aliquot of blood or other solution containing proteins is applied to a support, such as paper or a gel. When an electrical field is applied, proteins migrate a distance toward the anode (negative pole) or cathode (positive pole) that reflects small differences in their net charge. Electrophoresis of **Will Sichel's** blood shows that he is homozygous for the sickle cell variant, HbS, and has increased amounts of fetal hemoglobin, HbF. Individuals with sickle cell trait are heterozygous and have both HbA and HbS, plus small amounts of HbF ($\alpha_2\gamma_2$).

In heterozygous individuals with sickle cell trait, the sickle cell allele provides some protection against malaria. Malaria is caused by the parasite *Plasmodium fulciparum*, which spends part of its life cycle in red blood cells. The infected red blood cells of individuals with normal hemoglobin (HbA) develop protrusions that attach to the lining of capillaries. This attachment occludes the vessels and prevents oxygen from reaching cells in the affected region, resulting in cell death. In heterozygous individuals, HbS in infected cells aggregates into long fibers that cause the cell to become distorted. These distorted cells containing the parasite are preferentially recognized by the spleen and are rapidly destroyed, thus ending the life of the parasite.

In **Will Sichel** and other homozygous individuals with sickle cell anemia, the red blood cells sickle more frequently, especially under conditions of low oxygen tension (see Chapter 7). The result is a vaso-occlusive crisis in which the sickled cells clog capillaries and prevent oxygen from reaching cells (ischemia), thereby causing pain. The enhanced destruction of the sickled cells by the spleen results in anemia. Consequently, the sickle cell allele is of little advantage to homozygous individuals.

Because heterozygous individuals occur more frequently in a population than homozygous individuals, a selective advantage in a heterozygous state can outweigh a disadvantage in a homozygous state, causing the mutation to become a stable polymorphism in a population. As a consequence, the frequency of sickle cell anemia in parts of equatorial Africa in which malaria was endemic in the past is 1 in 25 births. Migration from Africa accounts for the high frequency of sickle cell anemia among blacks in the United States, which is approximately 1/400 at birth.



The substitution of a glutamate for a valine is a nonconservative replacement because a negatively

charged amino acid is substituted for a hydrophobic branched chain aliphatic amino acid. However, the substitution of an aspartate for a glutamate is a conservative replacement because the two amino acids have the same polarity and nearly the same size.



Homologous families of proteins are proteins that have the same ancestral protein and arose from the same gene. The term homologs includes both orthologs and paralogs. Orthologs are genes from different species that have evolved from a common ancestral gene as different species developed (for example, human and pork insulin). By contrast, paralogs are genes related by duplication within the genome of a single species (e.g., myoglobin and hemoglobin). Normally, othologs retain the same function in the course of evolution, whereas paralogs evolve new functions that may, or may not, be related to the original one.

1%, it is considered stable. The sickle cell allele is an example of a point mutation that is stable in the human population. Its persistence is probably attributable to selective pressure for the heterozygous mutant phenotype, which confers some protection against malaria.

B. Protein Families and Superfamilies.

A homologous family of proteins is composed of proteins related to the same ancestral protein. Groups of proteins with similar, but not identical, structure and function that have evolved from the same gene after the gene was duplicated are called paralogs and considered members of the same protein family. Once a gene has duplicated, one gene can continue to perform original function, and the second copy can mutate into a protein with another function or another type of regulation. This process is called divergent evolution. Very large families of homologous proteins are called a superfamily, which is subdivided by name into families of proteins with the most similarity in structure.

The paralogs of a protein family are considered different proteins and have different names because they have different functions. They are all present in the same individual. Myoglobin and the different chains of hemoglobin, for example, are paralogs and members of the same globin family that have similar, but not identical, structures and functions. Myoglobin, an intracellular heme protein present in most cells that stores and tranports O₂ to mitochondria, is a single polypeptide chain containing one heme oxygen-binding site. In contrast, hemoglobin is composed of four globin chains, each with a heme oxygenbinding site that is present in red blood cells and transports O₂ from the lungs to tissues. The gene for myoglobin is assumed to have evolved from gene duplication of the α -chain for hemoglobin, which evolved from duplication of the β chain. Figure 6.11 compares a region of the structure of myoglobin and the α and β chains of hemoglobin. Among these three proteins, only 15 invariant (identical) residues are present, but many of the other amino acid residues are conservative substitutions.

To compare the primary structure of two homologous polypeptide chains, the sequences are written left to right from the amino terminal to the carboxyl terminal. The sequences are aligned with computer programs that maximize the identity of amino acids and minimize the differences caused by segments that are present in one protein and not in the other.

The primary structure of human globin proteins

	1	5		10		15			
Myoglobin	gly	leu-ser-	-asp-gly-	glu-trp-	gln-leu-	val	-leu-asn-	val-trp-gly-lys-v	al-
β chain Hemoglobin	val-his-	leu-thr-	pro-glu-	glu-lys-	ser-ala-	va	-thr-ala-	leu-trp-gly-lys-v	al-
lpha chain hemoglobin	val	leu-ser-	-pro-ala-	asp-lys	-thr-asn-	val	-lys-ala-	ala-trp-gly-lys-v	al-
ζ chain Hemoglobin	met-ser-	leu-thr-	lys-thr-	glu-arg·	-thr-ile-	ile	-val-ser-	met-trp-ala-lys-il	e-
γ chain Hemoglobin	met-gly-his-	phe-thr	-glu-glu-	asp-lys	-ala-thr-	ile	-thr-ser-	leu-trp-gly-lys-v	al-

Fig 6.11. The primary structures of a region in human globin proteins. Gaps in the structure, indicated with dashes, are introduced to maximize the alignment between proteins in structure comparisons. They are assumed to coincide with mutations that caused a deletion. Regions of sequence similarity (identity and conservative substitution) are indicated in blue. Within these regions, there are smaller regions of invariant residues that are exactly the same from protein to protein. Myoglobin is a single polypeptide chain. The α and β chains are part of hemoglobin A ($\alpha_2\beta_2$). The ζ chain is part of embryonic hemoglobin ($\zeta_2\varepsilon_2$). The γ chain is part of fetal hemoglobin (HbF), $\alpha_2\gamma_2$.

C. Tissue and Developmental Variations in Protein Structure

Within the same individual, different isoforms or isozymes of a protein may be synthesized during different stages of fetal and embryonic development, may be present in different tissues, or may reside in different intracellular locations. Isoforms of a protein all have the same function. If they are isozymes (isoforms of enzymes), they catalyze the same reactions. However, isoforms have somewhat different properties and amino acid structure.

1. **DEVELOPMENTAL VARIATION**

Hemoglobin isoforms provide an example of variation during development. Hemoglobin is expressed as the fetal isozyme HbF during the last trimester of pregnancy until after birth, when it is replaced with HbA. HbF is composed of two hemoglobin α and 2 hemoglobin γ polypeptide chains, in contrast to the adult hemoglobin, hemoglobin A, which is 2 α and 2 β chains. During the embryonic stages of development, chains with a different amino acid composition, the embryonic ε and ζ chains, are produced (see Fig. 6.11). These differences are believed to arise evolutionarily from mutation of a duplicated α gene to produce ζ , and mutation of a duplicate α gene to produce ε . The fetal and embryonic forms of hemoglobin have a much higher affinity for O₂ than the adult forms, and thus confer an advantage at the low O₂ tensions to which the fetus is exposed. At different stages of development, the globin genes specific for that stage are expressed and translated.

2. **TISSUE-SPECIFIC ISOFORMS**

Proteins that differ somewhat in primary structure and properties from tissue to tissue, but retain essentially the same function, are called tissue-specific isoforms or isozymes. The enzyme creatine kinase is an example of a protein that exists as tissue-specific isozymes, each composed of two subunits with 60 to 72% sequence homology. Of the two creatine kinases that bind to the muscle sarcomere, the M form is produced in skeletal muscle, and the B polypeptide chains are produced in the brain. The protein comprises two subunits, and skeletal muscle therefore produces an MM creatine kinase, and the brain produces a BB form. The heart produces both types of chains and therefore forms a heterodimer, MB, as well as an MM dimer. Two more creatine kinase isozymes are found in mitochondria, a heart mitochondrial creatine kinase, and the "universal" isoform found in other tissues. (In general, most proteins present in both the mitochondria and cytosol will be present as different isoforms.) The advantage conferred on different tissues by having their own isoform of creatine kinase is unknown. However, tissue-specific isozymes such as MB creatine kinase are useful in diagnosing sites of tissue injury and cell death.

The structure of proteins involved in the response to hormones has been studied in greater depth than many other types of proteins, and most of these proteins are present as several tissue-specific isoforms that help different tissues respond differently to the same hormone. One of these proteins present in cell membranes is adenylyl cyclase, an enzyme that catalyzes the synthesis of intracellular 3',5' cyclic adenosine monophosphate (cAMP) (Fig. 6.12). In human tissues, at least nine different isoforms of adenylyl cyclase are coded by different genes in different tissues. Although they have an overall sequence homology of 50%, the two intracellular regions involved in the synthesis of cAMP are an invariant consensus sequence with a 93% identity. The different isoforms help cells respond differently to the same hormone.



The term isozyme was originally defined as meaning enzymes with a different primary structure, catalyzing the same reaction, encoded by different genes. However, the term is now used more broadly. Some isozymes are now known to arise from alternate splicing of premRNA to form different mRNAs (mRNA is the final nucleic acid template used for protein synthesis.)

A myocardial infarction (heart attack) is caused by an atheromatous obstruction or a severe spasm in a coronary artery that prevents the flow of blood to an area of heart muscle. Thus, heart cells in this region suffer from a lack of oxygen and blood-borne fuel. Because the cells cannot generate ATP, the membranes become damaged, and enzymes leak from the cells into the blood.

Creatine kinase (CK or CPK) is one of these enzymes. The protein is composed of two subunits, which may be either of the muscle (M) or the brain (B) type. The MB form, containing one M and one B subunit, is found primarily in cardiac muscle. It can be separated electrophoretically from other CK isozymes and the amount in the blood used to determine if a myocardial infarction has occurred. On admission to the hospital, Ann Jeina's total CK was 182 units/L (reference range 5 38-174 U/L). Her MB fraction was 6.8% (reference range 5% or less of the total CK). Although these values are only slightly elevated, they are typical of the phase immediately following a myocardial infarction. Additional information was provided by myoglobin and troponin T (Tn-T) measurements.



Variations in protein structure among species have been used to develop a phylogenetic tree of ancestral relationships between species showing the progress of evolution. According to evolutionary theory, related species evolved from a common ancestor, and proteins all can be grouped into clusters of proteins (called orthologous proteins) that evolved from a common ancestral protein. The more similar the amino acid sequence of an orthologous protein from two different species, the closer is the relationship between those species.



Although bovine (beef) insulin is identical to human insulin in those amino acid residues essential for activity, the amino acid residues that are in the variable regions can act as antigens and stimulate the formation of antibodies against bovine insulin. Consequently,

recombinant DNA techniques have been used for the synthesis of insulin, identical in structure to native insulin, such as Humulin (intermediate-acting insulin) (see Chapter 17 for information on recombinant DNA technology.) Although Di Abietes had not yet experienced an allergic response to her beef insulin, the cost of Humulin is now sufficiently low that most patients are being changed from beef insulin to the synthetic forms of human insulin.



The 20 DNA-encoded amino acids form the linear sequence of amino acids known as the primary struc-

ture of the protein, and determine the folding pattern for the three-dimensional conformation of the protein. Posttranslational modifications usually occur once the protein has already folded into its characteristic pattern. They are like accessories (jewelry, ties, etc.) that provide additional variations to your basic wardrobe.

Fig 6.12. Invariant regions in the isoforms of adenylyl cyclase. The invariant regions are on the cytosolic side of the membrane in the C1 and C2 loops shown in blue. These amino acid residues participate in the catalytic function of the enzyme, synthesis of 3',5'-cyclic AMP. The protein also has several helical regions that span the membrane (M1 helices and M2 helices), represented as tubes. An oligosaccharide chain is attached to an extracellular domain. N is the amino terminus. (Copyrighted from Taussig R, Gilman AG, J Biol Chem Mammalian membrane-bound adenylyl cyclases 1995;270:1-4.)

D. Species Variations in the Primary Structure of Insulin

Species variations in primary structure are also important in medicine, as illustrated by the comparison of human, beef, and pork insulin. Insulin is one of the hormones that are highly conserved between species, with very few amino acid substitutions and none in the regions that affect activity. Insulin is a polypeptide hormone of 51 amino acids that is composed of two polypeptide chains (Fig. 6.13). It is synthesized as a single polypeptide chain, but is cleaved in three places before secretion to form the C peptide and the active insulin molecule containing the A and B chains. The folding of the A and B chains into the correct threedimensional structure is promoted by the presence of one intrachain and two interchain disulfide bonds formed by cysteine residues. The invariant residues consist of the cysteine residues engaged in disulfide bonds and the residues that form the surface of the insulin molecule that binds to the insulin receptor. The amino acid substitutions in bovine and porcine insulin (shown in blue in Fig. 6.13.) are not in amino acids that affect its activity. Consequently, bovine and pork insulin were used for many years for the treatment of diabetes mellitus. However, even with only a few different amino acids, some patients developed an immune response to these insulins.

IV. MODIFIED AMINO ACIDS

After synthesis of a protein has been completed, a few amino acid residues in the primary sequence may be further modified in enzyme-catalyzed reactions that add a chemical group, oxidize, or otherwise modify specific amino acids in the protein. Because protein synthesis occurs by a process known as translation, these changes are called posttranslational modification. More than 100 different posttranslationally modified amino acid residues have been found in human proteins. These modifications change the structure of one or more specific amino acids on a protein in a way that may serve a regulatory function, target or anchor the protein in membranes, enhance a protein's association with other proteins, or target it for degradation (Fig. 6.14).



Structure of human insulin

Fig. 6.13. The primary structure of human insulin. The substituted amino acids in bovine (beef) and porcine (pork) insulin are shown in blue. Threonine 30 at the carboxy terminal of the B chain is replaced by alanine in both beef and pork insulin. In beef insulin, threonine 8 on the A chain is also replaced with alanine, and isoleucine 10 with valine. The cysteine residues, which form the disulfide bonds holding the chains together, are invariant. In the bioengineered insulin Humalog (lispro insulin), the position of proline at B28 and lysine at B29 is switched. Insulin is synthesized as a longer precursor molecule, proinsulin, which is one polpeptide chain. Proinsulin is converted to insulin by proteolytic cleavage of certain peptide bonds (squiggly lines in the figure). The cleavage removes a few amino acids and the 31–amino acid C-peptide that connects the A and B chains. The active insulin molecule, thus, has two nonidentical chains.

B. Glycosylation

Oligosaccharides (small carbohydrate chains) are bound to proteins by either *N*-linkages or *O*-linkages (see Fig. 6.14). *N*-linked oligosaccharides are found attached to cell surface proteins, where they protect the cell from proteolysis or an immune attack. In contrast, an *O*-glycosidic link is a common way of attaching oligosaccharides to the serine or threonine hydroxyl groups in secreted proteins. The intracellular polysaccharide glycogen is attached to a protein through an *O*-glycosidic linkage to a tyrosine.

C. Fatty Acylation or Prenylation

Many membrane proteins contain a covalently attached lipid group that interacts hydrophobically with lipids in the membrane. Palmitoyl groups (C16) are often attached to plasma membrane proteins, and the myristoyl group (C14) is often attached to proteins in the lipid membranes of intracellular vesicles (see Fig. 6.14). The farnesyl (C15) or geranylgeranyl group (C20) are synthesized from the five-carbon isoprene unit (isopentenyl pyrophosphate, see Fig. 5.1A) and are therefore called isoprenoids. These are attached in ether linkage to a specific cysteine residue of certain membrane proteins, particularly proteins involved in regulation.

D. Regulatory Modifications

Phosphorylation, acetylation, and adenosine diphosphate (ADP)-ribosylation of specific amino acid residues in a polypeptide can alter bonding by that residue and change the activity of the protein (see Fig. 6.14). Phosphorylation of an OH group on serine, threonine, or tyrosine by a protein kinase (an enzyme that transfers a phosphate group from ATP to a protein) introduces a large, bulky, negatively charged group that can alter the activity of a protein. Reversible acetylation occurring on lysine residues of histone proteins in the chromosome changes their

Adenylyl cyclase is posttranslationally modified (see Fig. 6.12). It has an oligosaccharide chain attached to the external portion of the protein. Some of the isozymes contain serine residues on the intracellular portion of the chain that may be phosphorylated by a protein kinase.

Carbohydrate addition

O-glycosylation: OH of ser, thr, tyr,



Lipid addition

Palmitoylation: Internal SH of cys



Myristoylation: NH of N-terminal gly







Regulation

Phosphorylation: OH of ser, thr, tyr



Acetylation: NH₂ of lys, terminus



ADP-ribosylation: N of arg, gln; S of cys



Modified amino acids



Fig. 6.14. Posttranslational modifications of amino acids in proteins. Some of the common amino acid modifications and the sites of attachment are illustrated. The added group is shown in blue. Because these modifications are enzyme-catalyzed, only a specific amino acid in the primary sequence is altered. R-O-hexagon refers to an oligosaccharide. In *N*-glycosylation, the attached sugar is usually N-acetylglucosamine (N-Ac).

interaction with the negatively charged phosphate groups of DNA. ADP-ribosylation is the transfer of an ADP-ribose from NAD⁺ to an arginine, glutamine, or a cysteine residue on a target protein in the membrane (primarily in leukocytes, skeletal muscles, brain, and testes). This modification may regulate the activity of these proteins

E. Other Amino Acid Posttranslational Modifications

A number of other posttranslational modifications of amino acid side chains alter the activity of the protein in the cell (see Fig 6.14). Carboxylation of the γ carbon of glutamate (carbon 4) in certain blood clotting proteins is important for attaching the clot to a surface. Calcium ions mediate this attachment by binding to the two negatively charged carboxyl groups of γ -glutamate and two additional negatively charged groups provided by phospholipids in the cell membrane. Collagen, an abundant fibrous extracellular protein, contains the oxidized amino acid hydroxyproline. The addition of the hydroxyl group to the proline side chain provides an extra polar group that can engage in hydrogen bonding between the polypeptide strands of the fibrous protein.

F. Selenocysteine

The unusual amino acid selenocysteine is found in a few enzymes and is required for their activity (Fig. 6.15). Its synthesis is not a posttranslational modification, however, but a modification to serine that occurs while serine is bound to a unique tRNA. The selenocysteine is then inserted into the protein as it is being synthesized.

CLINICAL COMMENTS

Will Sichel. Will Sichel was treated for 3 days with parenteral (intravascular) narcotics, hydration, and nasal inhalation of oxygen for his vasoocclusive crisis. The diffuse severe pains of sickle cell crises result from occlusion of small vessels in a variety of tissues, thereby causing damage to cells from ischemia (low blood flow) or hypoxia (low levels of oxygen). Vaso-occlusion occurs when HbS molecules in red blood cells polymerize in the capillaries, where the partial pressure of O_2 (pO_2) is low. This polymerization causes the red blood cells to change from a biconcave disc to a sickle shape that cannot deform to pass through the narrow capillary lumen. The cells aggregate in the capillaries and occlude blood flow.

In addition, **Will Sichel** was treated with hydroxyurea therapy, which increases the production of red blood cells containing fetal hemoglobin. HbF molecules cannot participate in sickling. **Will Sichel's** acute symptoms gradually subsided. Had his severe pain persisted, partial exchange blood transfusions would have been considered because no other effective therapy is currently available. Patients with sickle cell anemia periodically experience sickle cell crises, and Will's physician urged him to seek medical help whenever symptoms reappeared.

Cal Kulis. Mr. Kulis has cystinuria, a relatively rare disorder, with a prevalence that ranges between 1 in 2,500 to 1 in 15,000 births, depending on the population studied. It is a genetically determined disease with a complex recessive mode of inheritance resulting from allelic mutations. These mutations lead to a reduction in the activity of renal tubular cell transport proteins that normally carry cystine from the tubular lumen into the renal tubular cells. The transport of the basic amino acids (lysine, arginine, and ornithine, an amino acid

A number of pathogenic bacteria produce bacterial toxins that are ADP-ribosyl transferases (NAD⁺glycohydrolases). These enzymes hydrolyze the N-glycosidic bond of NAD⁺ and transfer the ADP-ribose portion to a specific amino acid residue on a protein in the affected human cell. Cholera A-B toxin, a pertussis toxin, and a diptheria toxin are all ADP-ribosyl transferases.

Selenocysteine

Fig. 6.15. Selenocysteine.

found in the urea cycle but not in proteins) is also often compromised, and they appear in the urine.

Because cystine is produced by oxidation of cysteine, conservative treatment of cystinuria includes decreasing the amount of cysteine within the body and, hence, the amount of cystine eventually filtered by the kidneys. Reduction of cysteine levels is accomplished by restricting dietary methionine, which contributes its sulfur to the pathway for cysteine formation. To increase the amount of cystine that remains in solution, the volume of fluid ingested daily is increased. Crystallization of cystine is further prevented by chronically alkalinizing the urine. Finally, drugs may be administered to enhance the conversion of urinary cystine to more soluble compounds. If these conservative measures fail to prevent continued cystine stone formation, existing stones may be removed by a surgical technique that involves sonic fracture. The fragmented stones may then pass spontaneously or may be more easily extracted surgically because of their smaller size.

Di Abietes. Di Abietes' treatment was first changed to daily injections of Humulin instead of beef insulin. Humulin is now mass-produced by recombinant DNA techniques that insert the human DNA sequences for the insulin A and B chains into the *Escherichia coli* or yeast genome (see Chapter 17). The insulin chains that are produced are then extracted from the media and treated to form the appropriate disulfide bonds between the chains. As costs have fallen for production of the synthetic human insulins, they have replaced pork insulin and the highly antigenic beef insulin.

Di's physician then recommended that she take Humalog, an insulin preparation containing lispro, an ultra-fast-acting bioengineered insulin analog in which lysine at position B29 has been moved to B28 and proline at B28 has been moved to B29 (hence, lispro) (see Fig. 6.13). With lispro, Di will be able to time her injections minutes before her consumption of carbohydrate-containing meals, rather than having to remember to give herself an insulin injection 1 hour before a meal.

Ann Jeina. Mrs. Jeina continued to be monitored in the cardiac care unit. At admission, her CK levels were elevated (182 units/L compared with a reference range of 38-174 U/L), and the MB fraction was high at 6.8% of the total (reference = <5% of the total). Her total CK continued to rise (228 units/L 12 hours after admission and 266 units/L at 24 hours), as did her MB fraction (8% at 12 hours and 10.8% at 24 hours). Within 2 hours of the onset of an acute myocardial infarction, the MB form of CK begins leaking from heart cells that were injured by the ischemic process. These rising serum levels of the MB fraction (and, therefore, of the total CK) reach their peak 12 to 36 hours later and usually return to normal within 3 to 5 days from the onset of the infarction (Fig. 6.16). In addition to the CK measurements, her blood levels were also analyzed for myoglobin and the heart isoform of troponin-T, a protein involved in muscle contraction (see Chapter 7).

BIOCHEMICAL COMMENTS

Enzyme and Protein Databases. Large databases of protein structure have been assembled to collate data from various laboratories around the world. The National Library of Medicine maintains a website, PubMed, which catalogues the medical literature, and a search engine that allows you to search for reference articles by topic or author

The switch in position of amino acids in lispro does not affect the action of this synthetic insulin on cells because it is not in a critical invariant region, but it does affect the ability of insulin to bind zinc. Normally, human insulin is secreted from the pancreas as a zinc hexamer in which six insulin molecules are bound to the zinc atom. When zinc insulin is injected, the binding to zinc slows the absorption from the subcutaneous (under the skin) injection site. Lispro cannot bind zinc to form a hexamer, and thus it is absorbed much more quickly than other insulins.

Creatine kinase isozymes in blood



Patient 24 hrs after myocardial infarction

Fig. 6.16. Electrophoretic separation of serum creatine kinase enzymes from a normal healthy adult and from a patient who had a myocardial infarction 24 hours previously. Creatine kinase catalyzes the reversible transfer of a phosphate from ATP to creatine to form phosphocreatine and ADP. The reaction is an important part of energy metabolism in heart muscle, skeletal muscle, and brain. Three different forms of the dimer exist: BB (or CK-1) found in brain, MB (or CK-2) found only in heart, and MM (or CK-3), found only in skeletal and heart muscle (cathode, -ve; anode, +ve).

(http://www.ncbi.nlm.nih.gov/pubmed). From the menu, you can directly enter a protein database or a structure database. These databases are linked, so that you can type in the name of a protein, such as human hemoglobin chain A, obtain a list of contributors to sequence data, and retrieve a complete amino acid sequence for many proteins. You can link to PubMed to find recent articles about the protein, or you can link to the structure database. A program called Cn3D can be downloaded from this site that allows you to view the three-dimensional versions of the protein structures.

These databases are only a few of more than 500 biologic databases that have been assembled to collate and exchange biologic information in the areas of DNA, RNA, genomics, gene mapping, and protein structure. The first issue of the *Journal of Nucleic Acid Research* each year provides a description of currently available

biologic databases. Their goal is to provide information that can relate a particular DNA sequence or mutation to the protein involved, to its function, and to the pathologic consequences of a particular amino acid substitution, by comparing proteins that have similar functional elements.

Will these databases be of any use to the practicing physician or to the practice-oriented medical student? Very few students will ever want to do protein modeling. However, students and physicians may wish to use the literature search in PubMed as part of their approach to evidence-based medicine. They also may wish to use it to track definitions or fundamental knowledge about particular topics. Thus, a move has begun to link biomedical and basic science textbooks to PubMed.

Suggested Reference

Tatusov RL, Koonin EV, Lipman, DJ. A genomic perspective on protein families. Science 1997;278:631-637.



REVIEW QUESTIONS—CHAPTER 6

- 1. In a polypeptide at physiologic pH, hydrogen bonding may occur between
 - (A) the side chains of a leucine residue and a lysine residue.
 - (B) the side chains of an aspartyl residue and a glutamyl residue.
 - (C) the terminal α -amino group and the terminal α -carboxyl group.
 - (D) the amide group in the peptide bond and an aspartyl side chain.
 - (E) the SH groups of two cysteine residues.
- 2. Which of the following shows the linear sequence of atoms joined by covalent bonds in a peptide backbone?
 - (A) –N–C–C–N–C–C–N–C–C
 - (B) -N-C-O-N-C-O-N-C-O-
 - (C) -N-C-C-O-N-C-C-O-N-C-C-O-
 - $(D) \ -N-H-C-C-N-H-C-C-N-H-C-C-$
 - $(E) \quad -N-H-C-O-H-N-H-C-O-H-N-H-C-C-\\$
- 3. **Di Abietes'** different preparations of insulin contain some insulin complexed with protamine that is absorbed slowly after injection. Protamine is a protein preparation from rainbow trout sperm containing arginine-rich peptides that bind insulin. Which of the following provides the best explanation for complex formation between protamine and insulin?
 - (A) Arginine is a basic amino acid that binds to negatively charged amino acid side chains in insulin.
 - (B) Arginine is a basic amino acid that binds to the α -carboxylic acid groups at the N-terminals of insulin chains.
 - (C) Arginine is a large bulky hydrophobic amino acid that complexes with leucine and phenylalanine in insulin.
 - (D) Arginine forms disulfide bonds with the cysteine residues that hold the A and B chains together.
 - (E) Arginine has a side chain that forms peptide bonds with the carboxyl terminals of the insulin chains.
- 4. Protein kinases phosphorylate proteins only at certain hydroxyl groups on amino acid side chains. Which of the following groups of amino acids all contain side chain hydroxyl groups?
 - (A) aspartate, glutamate, and serine
 - (B) serine, threonine, and tyrosine
 - (C) threonine, phenylalanine, and arginine
 - (D) lysine, arginine, and proline
 - (E) alanine, asparagine, and serine

- 5. In a single individual, the primary structures of enzymes catalyzing the same reaction
 - (A) are exactly the same from cell type to cell type, although the amount of enzyme may differ.
 - (B) stay the same throughout the lifetime of that individual.
 - (C) are identical if the enzymes are paralogs.
 - (D) are identical to all members of the homologous family.
 - (E) may differ between different cellular compartments of the same cell.