B1 Amino acid structure

Key Notes

Amino acids	All proteins are made up from the same set of 20 standard amino acids. A typical amino acid has a primary amino group, a carboxyl group, a hydrogen atom and a side-chain (R group) attached to a central α-carbon atom (C_α).
Enantiomers	Amino acids with four different groups arranged tetrahedrally around the C_α atom can exist in either the D or L configuration. These two enantiomers are nonsuperimposable mirror images that can be distinguished on the basis of their different rotation of plane-polarized light. Only the L isomer is found in proteins.
The 20 standard amino acids	The different side-chains or R groups display different physicochemical properties. They may be polar, acidic, basic, aromatic, bulky, conformationally inflexible, able to form hydrogen bonds, able to cross-link and be chemically reactive. Glycine (Gly, G) has a hydrogen atom as its R group. Alanine (Ala, A), valine (Val, V), leucine (Leu, L), isoleucine (Ile, I) and methionine (Met, M) have aliphatic side-chains of differing structures that are hydrophobic and chemically inert. The aromatic side-chains of phenylalanine (Phe, F), tyrosine (Tyr, Y) and tryptophan (Trp, W) are also hydrophobic in nature. The conformationally rigid proline (Pro, P) has its aliphatic side-chain bonded back on to the amino group and thus is really an imino acid. The hydrophobic, sulfur-containing side-chain of cysteine (Cys, C) is highly reactive and can form a disulfide bond with another cysteine residue. The basic amino acids arginine (Arg, R) and lysine (Lys, K) have positively charged side-chains, whilst the side-chain of histidine (His, H) can be either positively charged or uncharged at neutral pH. The side-chains of the acidic amino acids aspartic acid (Asp, D) and glutamic acid (Glu, E) are negatively charged at neutral pH. The amide side-chains of asparagine (Asn, N) and glutamine (Gln, Q), and the hydroxyl side-chains of serine (Ser, S) and threonine (Thr, T) are uncharged and polar, and can form hydrogen bonds.
Acids, bases and pH	pH is a measure of the concentration of H^+ in a solution. An acid is a proton donor; a base is a proton acceptor. Ionization of an acid yields its conjugate base, and the two are termed a conjugate acid–base pair, for example acetic acid (CH_3COOH) and acetate (CH_3COO^-). The pK of an acid is the pH at which it is half dissociated. The

Henderson–Hasselbalch equation expresses the relationship between pH, pK and the ratio of acid to base.

Buffers

An acid–base conjugate pair can act as a buffer, resisting changes in pH. From a titration curve of an acid, the inflection point indicates the pK value. The buffering capacity of the acid–base pair is the p$K \pm 1$ pH unit. In biological fluids, the phosphate and carbonate ions act as buffers. Amino acids, proteins, nucleic acids and lipids also have some buffering capacity.

Ionization of amino acids

The α-amino and α-carboxyl groups on amino acids act as acid–base groups, donating or accepting a proton as the pH is altered. At low pH, both groups are fully protonated, but as the pH is increased first the carboxyl group and then the amino group loses a hydrogen ion. For the standard 20 amino acids, the pK is in the range 1.8–2.9 for the α-carboxyl group and 8.8–10.8 for the α-amino group. Those amino acids with an ionizable side-chain have an additional acid–base group with a distinctive pK.

Related topics

(B2) Protein structure and function
(B3) Myoglobin and hemoglobin
(B4) Collagen

Amino acids

Amino acids are the building blocks of **proteins** (Section B2). Proteins of all species, from bacteria to humans, are made up from the same set of **20 standard amino acids**. Nineteen of these are α-amino acids with a **primary amino group** ($-NH_3^+$) and a **carboxylic acid** (carboxyl; $-COOH$) group attached to a central carbon atom, which is called the **α-carbon atom** (C_α) because it is adjacent to the carboxyl group (Figure 1a). Also attached to the C_α atom is a hydrogen atom and a **variable side-chain** or 'R' group. The one exception to this general structure is proline, which has a secondary amino group and is really an **α-imino acid**. The names of the amino acids are often abbreviated, either to three letters or to a single letter. Thus, for example, proline is abbreviated to Pro or P (Figure 2).

Enantiomers

All of the amino acids, except for glycine (Gly or G; Figure 2), have four different groups arranged **tetrahedrally** around the central C_α atom, which is thus known as an **asymmetric center** or **chiral center** and has the property of **chirality** (Greek; *cheir*, hand) (Figure 1b). The two **nonsuperimposable mirror images** are termed **enantiomers**. Enantiomers are physically and chemically indistinguishable by most techniques, but can be distinguished on the basis of their different optical rotation of plane-polarized light. Molecules are classified as dextrorotatory (D; Greek '*dextro*'=right) or levorotatory (L; Greek '*levo*'=left) depending on whether they rotate the plane of plane-polarized light clockwise or anticlockwise. D- and L-amino acids can also be distinguished by enzymes that usually only recognize one or other enantiomer. Only the **L-amino acids** are found in proteins. **D-Amino acids** rarely occur in nature, but are found in bacterial cell walls (Section A1) and certain antibiotics.

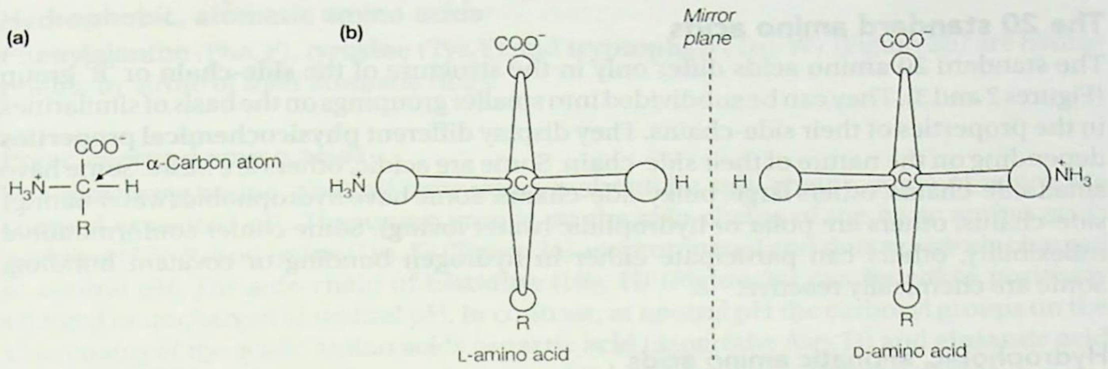

Figure 1. (a) Basic structure of an amino acid showing the four different groups around the central α-carbon atom; (b) the two enantiomers of an amino acid.

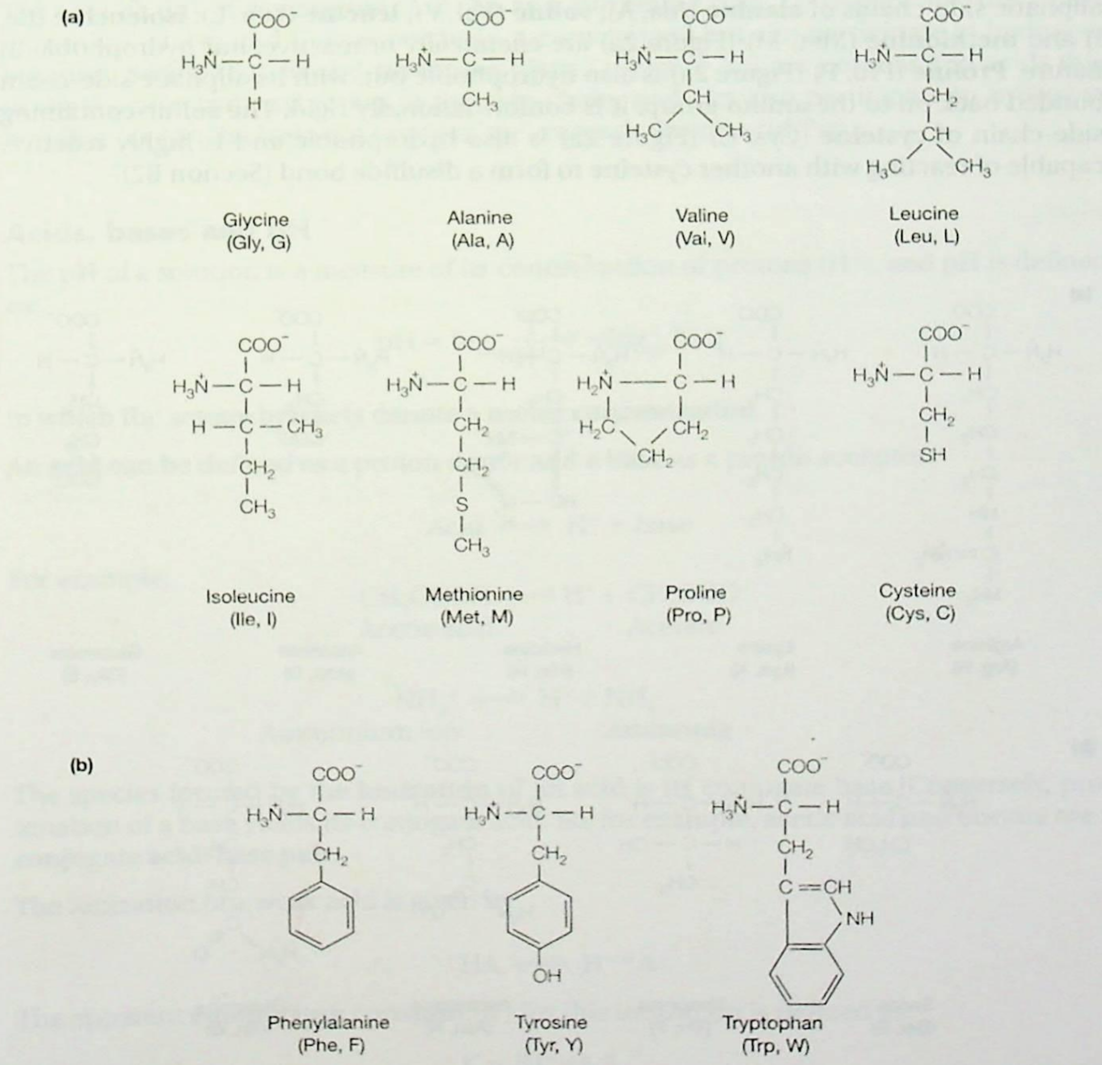

Figure 2. The standard amino acids: (a) hydrophobic, aliphatic R groups; (b) hydrophobic, aromatic R groups. The molecular weights of the amino acids are given in Table 1.

The 20 standard amino acids

The standard 20 amino acids differ only in the structure of the **side-chain** or 'R' group (Figures 2 and 3). They can be subdivided into smaller groupings on the basis of similarities in the properties of their side-chains. They display different **physicochemical properties** depending on the nature of their side-chain. Some are acidic, others are basic. Some have small side-chains, others large, bulky side-chains. Some have hydrophobic (water hating) side-chains, others are polar or hydrophilic (water loving). Some confer conformational inflexibility, others can participate either in hydrogen bonding or covalent bonding. Some are chemically reactive.

Hydrophobic, aliphatic amino acids

Glycine (Gly, G) (Figure 2a), the smallest amino acid with the simplest structure, has a hydrogen atom in the side-chain position, and thus does not exist as a pair of stereoisomers, since there are two identical groups (hydrogen atoms) attached to the C_α atom. The aliphatic side-chains of **alanine** (Ala, A), **valine** (Val, V), **leucine** (Leu, L), **isoleucine** (Ile, I) and **methionine** (Met, M) (Figure 2a) are chemically unreactive, but hydrophobic in nature. **Proline** (Pro, P) (Figure 2a) is also hydrophobic but, with its aliphatic side-chain bonded back on to the amino group, it is conformationally rigid. The sulfur-containing side-chain of **cysteine** (Cys, C) (Figure 2a) is also hydrophobic and is highly reactive, capable of reacting with another cysteine to form a disulfide bond (Section B2).

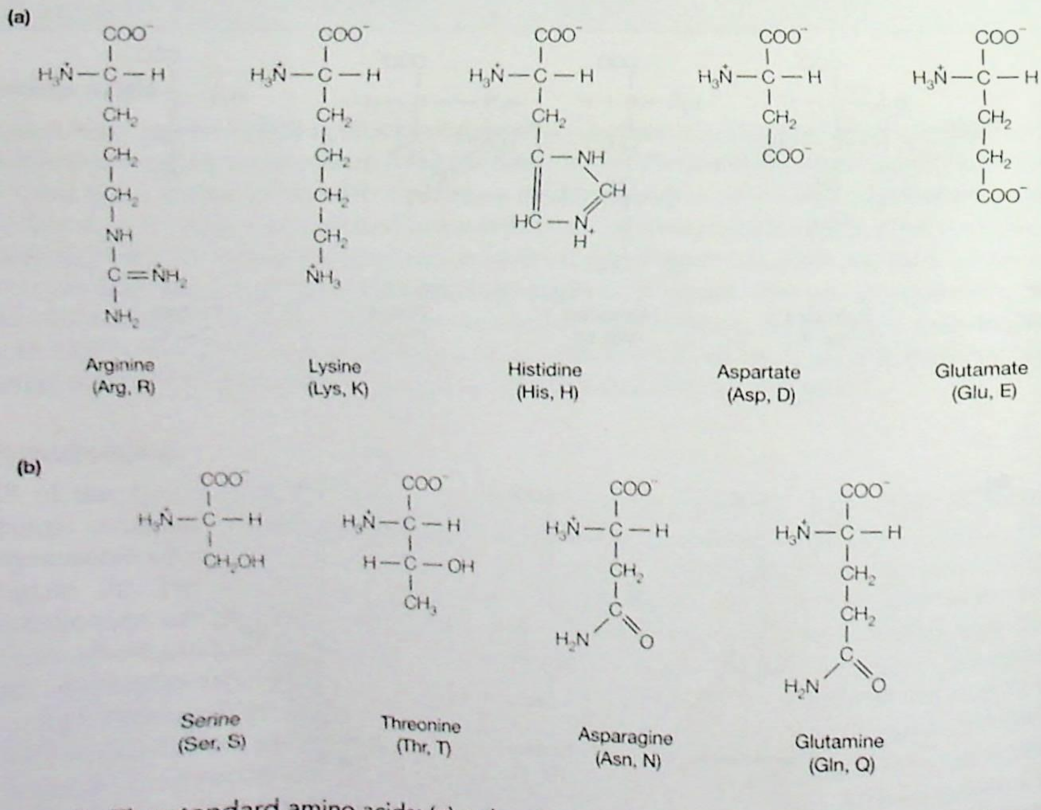

Figure 3. The standard amino acids: (a) polar, charged R groups; (b) polar, uncharged R groups. The molecular weights of the amino acids are given in Table 1.

Hydrophobic, aromatic amino acids

Phenylalanine (Phe, F), **tyrosine** (Tyr, Y) and **tryptophan** (Trp, W) (Figure 2b) are hydrophobic by virtue of their aromatic rings.

Polar, charged amino acids

The remaining amino acids all have polar, hydrophilic side-chains, some of which are charged at neutral pH. The amino groups on the side-chains of the basic amino acids **arginine** (Arg, R) and **lysine** (Lys, K) (Figure 3a) are protonated and thus positively charged at neutral pH. The side-chain of **histidine** (His, H) (Figure 3a) can be either positively charged or uncharged at neutral pH. In contrast, at neutral pH the carboxyl groups on the side-chains of the acidic amino acids **aspartic acid** (aspartate; Asp, D) and **glutamic acid** (glutamate; Glu, E) (Figure 3a) are de-protonated and possess a negative charge.

Polar, uncharged amino acids

The side-chains of **asparagine** (Asn, N) and **glutamine** (Gln, Q) (Figure 3b), the amide derivatives of Asp and Glu, respectively, are uncharged but can participate in hydrogen bonding. **Serine** (Ser, S) and **threonine** (Thr, T) (Figure 3b) are polar amino acids due to the reactive hydroxyl group in the side-chain, and can also participate in hydrogen bonding (as can the hydroxyl group of the aromatic amino acid Tyr).

Acids, bases and pH

The pH of a solution is a measure of its concentration of protons (H^+), and pH is defined as:

$$\text{pH} = \log_{10}\frac{1}{H^+} = -\log_{10}[H^+]$$

in which the square brackets denote a **molar concentration**.

An **acid** can be defined as a proton donor and a **base** as a proton acceptor:

$$\text{Acid} \rightleftharpoons H^+ + \text{base}$$

For example;

$$\underset{\text{Acetic acid}}{CH_3COOH} \rightleftharpoons H^+ + \underset{\text{Acetate}}{CH_3COO^-}$$

$$\underset{\text{Ammonium ion}}{NH_4^+} \rightleftharpoons H^+ + \underset{\text{Ammonia}}{NH_3}$$

The species formed by the **ionization** of an acid is its conjugate base. Conversely, protonation of a base yields its conjugate acid. So, for example, acetic acid and acetate are a **conjugate acid–base pair**.

The ionization of a weak acid is given by:

$$HA \rightleftharpoons H^+ + A^-$$

The apparent **equilibrium constant** (K) for this ionization is defined as:

$$K = \frac{[H^+][A^-]}{[HA]} \qquad \text{(Equation 1)}$$

The **p*K*** of an acid is defined as:

$$pK = -\log K = \log \frac{1}{K}$$

The p*K* of an acid is the pH at which it is half dissociated, i.e. when $[A^-]=[HA]$.

The **Henderson–Hasselbalch equation** expresses the relationship between pH and the ratio of acid to base. It is derived as follows. Rearrangement of Equation 1 gives:

$$\frac{1}{[H^+]} = \frac{1}{K} \times \frac{[A^-]}{[HA]}$$

Taking the logarithm of both sides of this equation gives:

$$\log \frac{1}{[H^+]} = \log \frac{1}{K} + \log \frac{[A^-]}{[HA]}$$

Substituting pH for $\log 1/[H^+]$ and p*K* for $\log 1/K$ gives:

$$pH = pK + \log \frac{[A^-]}{[HA]}$$

which is the Henderson–Hasselbalch equation. This equation indicates that the p*K* of an acid is numerically equal to the pH of the solution when the molar concentration of the acid is equal to that of its conjugate base. The pH of a solution can be calculated from the Henderson–Hasselbalch equation if the molar concentrations of A^- and HA, and the p*K* of HA are known. Similarly, the p*K* of an acid can be calculated if the molar concentrations of A^- and HA, and the pH of the solution are known.

Buffers

An acid–base conjugate pair, such as acetic acid and acetate, is able to resist changes in the pH of a solution. That is, it can act as a **buffer**. On addition of hydroxide (OH^-) to a solution of acetic acid the following happens:

$$CH_3COOH + OH^- \rightleftharpoons CH_3COO^- + H_2O$$

A plot of the dependence of the pH of this solution on the amount of OH^- added is called a **titration curve** (Figure 4). There is an inflection point in the curve at pH 4.8, which is the p*K* of acetic acid. In the vicinity of this pH, a relatively large amount of OH^- (or H^+) produces

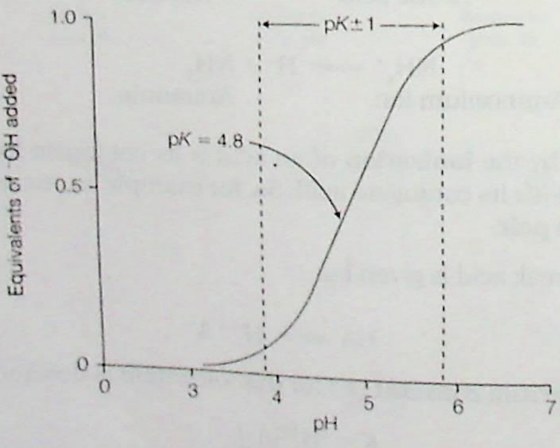

Figure 4. Titration curve of acetic acid.

little change in pH as the added OH^- (or H^+) reacts with CH_3COOH (or CH_3COO^-), respectively. Weak acids are most effective in buffering against changes in pH within 1 pH unit of the p*K* (Figure 4), often referred to as p*K* ± 1, the **buffering capacity**.

Biological fluids, including the cytosol and extracellular fluids such as blood, are buffered. For example, in healthy individuals the pH of the blood is carefully controlled at pH 7.4. The major buffering components in most biological fluids are the phosphate ion ($H_2PO_4^-$, p*K* 6.82) and the carbonate ion (HCO_3^-, p*K* 6.35) because they have p*K* values in this range. However, many biological molecules, including amino acids, proteins, nucleic acids and lipids, have multiple acid–base groups that are effective at buffering in the physiological pH range (pH 6–8).

When working with enzymes, proteins and other biological molecules it is often crucial to buffer the pH of the solution in order to avoid **denaturation** (loss of activity) of the component of interest (Section D3). Numerous buffers are used in laboratories for this purpose. One of the commonest is tris(hydroxy methyl)aminomethane or **Tris**, which has a p*K* of 8.08.

Ionization of amino acids

The 20 standard amino acids have two **acid–base groups**: the α-amino and α-carboxyl groups attached to the C_α atom. Those amino acids with an **ionizable side-chain** (Asp, Glu, Arg, Lys, His, Cys, Tyr) have an additional acid–base group. The **titration curve** of Gly is shown in Figure 5a. At low pH (i.e. high hydrogen ion concentration) both the amino group and the carboxyl group are fully protonated so that the amino acid is in the cationic form $H_3N^+CH_2COOH$ (Figure 5b). As the amino acid in solution is titrated with increasing amounts of a strong base (e.g. NaOH), it loses two protons, first from the carboxyl group, which has the lower **p*K*** value (p*K*=2.3), and then from the amino group, which has the higher p*K* value (p*K*=9.78). The pH at which Gly has no net charge is termed its **isoelectric point, pI**. The α-carboxyl groups of all the 20 standard amino acids have p*K* values in the range 1.8–2.5, whilst their α-amino groups have p*K* values in the range 8.7–10.7

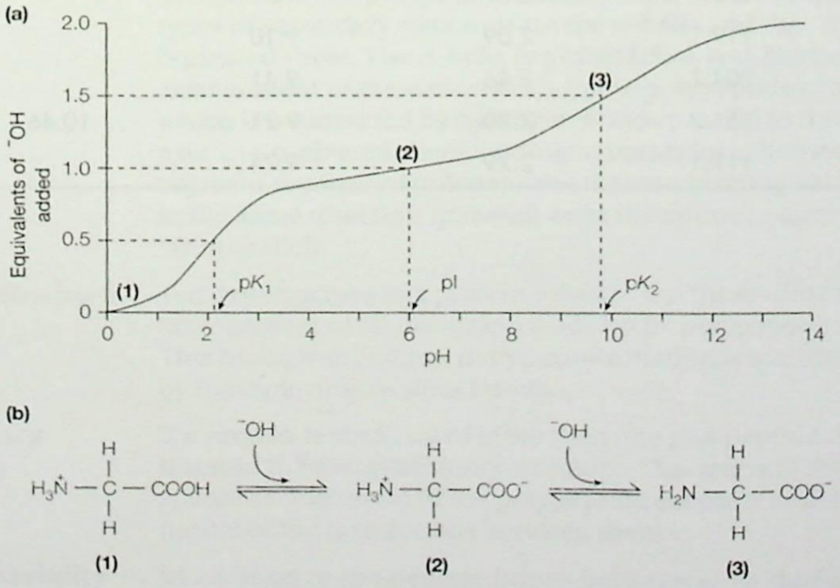

Figure 5. Ionization of glycine: (a) titration curve of glycine; (b) dissociation of glycine. Numbers in bold in parentheses in (a) correspond to the structures in (b).

(Table 1). The side-chains of the acidic amino acids Asp and Glu have p*K* values of 3.9 and 4.1, respectively, whereas those of the basic amino acids Arg and Lys have p*K* values of 12.5 and 10.5, respectively. Only the side-chain of His, with a p*K* value of 6.0, is ionized within the physiological pH range (pH 6–8). It should be borne in mind that when the amino acids are linked together in proteins, only the side-chain groups and the terminal α-amino and α-carboxyl groups of the polypeptide chain are free to ionize (Section B2).

Table 1. pK values and molecular weights of the 20 standard amino acids

Amino acid	Mol. Wt	p*K* α-COOH	p*K* $\alpha\text{-NH}_3^+$	p*K* side-chain
Alanine	89.1	2.35	9.87	
Arginine	174.2	1.82	8.99	12.48
Asparagine	132.1	2.14	8.72	
Aspartic acid	133.1	1.99	9.90	3.90
Cysteine	121.2	1.92	10.70	8.37
Glutamic acid	147.1	2.10	9.47	4.07
Glutamine	146.2	2.17	9.13	
Glycine	75.1	2.35	9.78	
Histidine	155.2	1.80	9.33	6.04
Isoleucine	131.2	2.32	9.76	
Leucine	131.2	2.33	9.74	
Lysine	146.2	2.16	9.06	10.54
Methionine	149.2	2.13	9.28	
Phenylalanine	165.2	2.20	9.31	
Proline	115.1	1.95	10.64	
Serine	105.1	2.19	9.21	
Threonine	119.1	2.09	9.10	
Tryptophan	204.2	2.46	9.41	
Tyrosine	181.2	2.20	9.21	10.46
Valine	117.1	2.29	9.74	