Biochemistry (I)

2nd Year Students – Science Group

Faculty of Education

First Term – 2022/2023

Dr./ Ibrahim Abdul-Motaleb Mousa

Chemistry of Carbohydrates

Introduction:

Carbohydrates received their name from the fact that the general empirical formula for many members of the class can be written $C_n(H_2O)_n$, hydrated carbon. (Note, however, that not all carbohydrates are represented by this formula.) Sugars, starches, and cellulose-compounds which have important structural and energy functions in the living materials-are all carbohydrates. (Sugars are water soluble carbohydrates).

It would be difficult to overestimate the importance of human beings of carbohydrates. We eat them directly in such foods as bread, potatoes, corn, and peas, and indirectly, in meat, eggs, and fats from animals that feed on carbohydrates in the form of grains and grasses. Cotton and linen, the traditional clothing fabrics, are both almost pure carbohydrates. Only in very recent times have synthetic polymers begun to replace these natural fibers. Wood consists largely of cellulose, and hence a good portion of the houses in which we live, as well as much of our furniture, is constructed of carbohydrates. Finally, paper is mostly carbohydrates. The importance of paper in modern civilization is enormous. Try to imagine life without paper: no paper money, checks, income-tax forms; no books, newspapers, birth certificates or milk cartons. College diplomas would have to be printed on sheepskin, and Playboy magazine would disappear from the corner drugstore. Truly, this will be a different world!

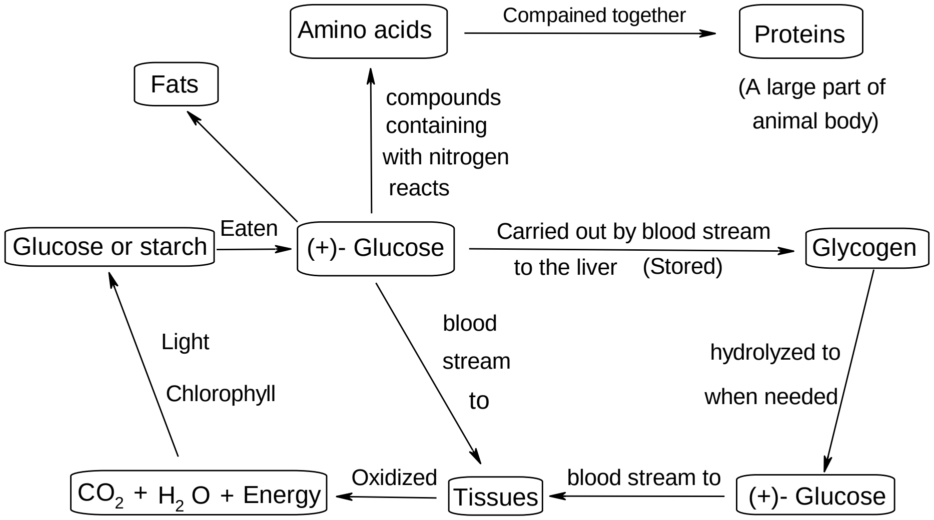
Photosynthesis:

The production of carbohydrates in nature occurs in green palnts by a process called photosynthesis. Plants contain the green pigments chlorophyll which catalyzes the conversion of carbon dioxide and water into sugar. The reaction is thermodynamically unfavorable, but proceed because the necessary energy is supplied by the sun in the form of sunlight.

$$6\,CO_2 + 6H_2O \underset{\text{- energy, animal metabolism}}{\overset{\text{+ Light, Chlorophyll}}{\rightleftharpoons}} C_6(H_2O)_6 + 6\,O_2$$

While plants build up carbohydrates from carbon dioxide and water, animals degrade carbohydrates to carbon dioxide and water. The animal obtain carbohydrates by eating plants and combine the carbohydrates with oxygen from the air to carry out the reverse of the photosynthesis reaction. The oxidation of carbohydrates supplies the animal with the energy (according to the above equation) necessary to sustain life, and it also regenerates carbon dioxide for use by the plants in photosynthesis.

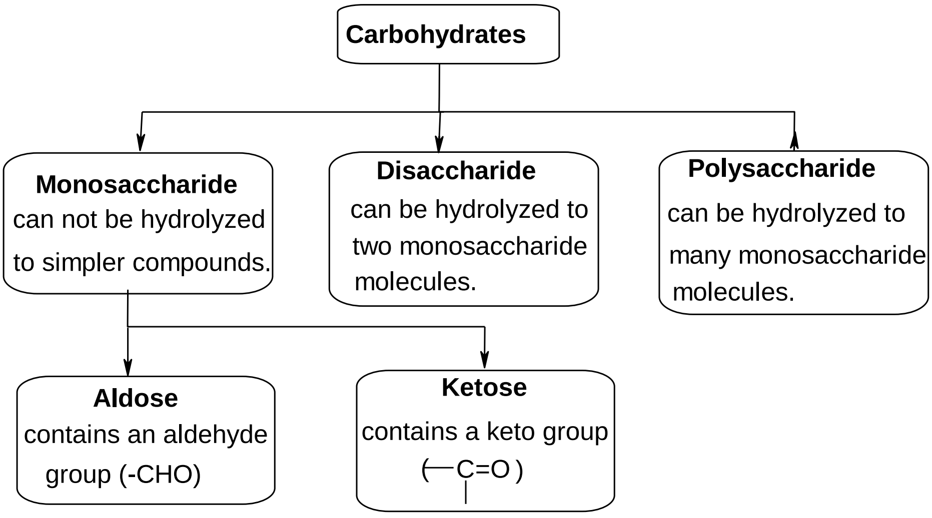
All carbohydrates are polyhydroxy aldehydes, polyhydroxy ketones, or molecules which yield polyhydroxy aldehydes or ketones on hydrolysis. Monosaccharides (Chapter 2), are the smallest carbohydrate molecules and include the four-, five-, and six-carbon sugars. Sucrose, table sugar, is one of the disaccharides (Chapter 3); disaccharides can hydrolyzed to two monosaccharides. Polysaccharides (Chapter 4), which include starch and cellulose, yield many monosaccharides molecules upon hydrolysis.



Definition and Classification:

Carbohydrates are polyhdroxy aldehydes, polyhdroxy ketones or compounds that can be hydrolyzed to them.

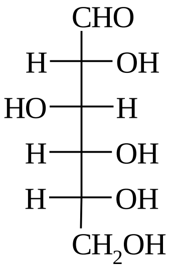
Depending upon the number of carbon atoms it contains, a monosaccharide known as a triose, tetrose, pentose. hexose.



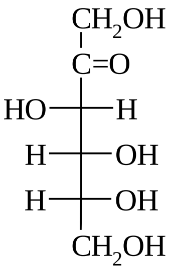
Example:

- A six carbon monosaccharide containing an aldehyde group is Aldohexose.
- A six carbon monosaccharide containing a keto group is kctohexosc.
- A five carbon monosaccharide containing an aldehyde group is Aldopentose.
- The -ose suffix is used to designate a carbohydrate.
- The aldo- and keto- prefixes designate the nature of the carbonyl group (aldehyde or ketone).

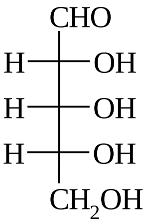
Examples:



Glucose
aldohexose



Fructose
ketohexose



Ribose
aldopentose

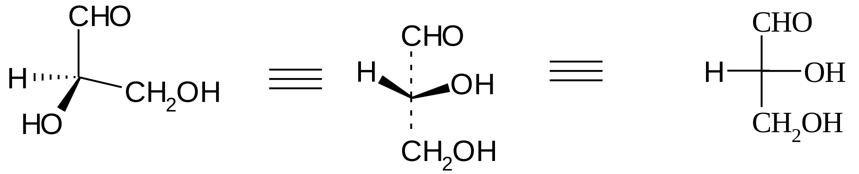
CHAPTER 2

Monosaccharides

$(C_6H_{12}O_6)$

Fischer Projection for Depicting Carbohydrates:

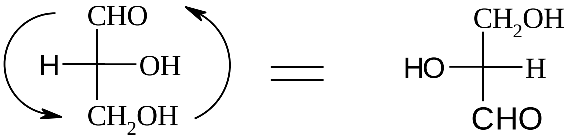
Teterahedral carbon atom is represented in a Fischer projection by two crossed lines as shown in the following.



(R)- glyceraldehyde

Fischer projection

Recall also that Fischer projection can be rotated on the page by 180 ºC without their meaning but not 90 ºC or 270 ºC.

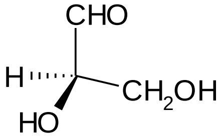


(R) - glyceraldehyde

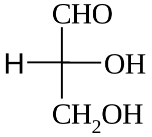
D, L-Sugars:

Three Dimensional

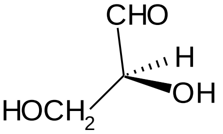
Fischer projection



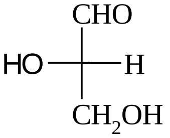
R(+)- glyceraldehyde



D- glyceraldehyde



S(-)- glyceraldehyde



L- glyceraldehyde

In Fischer Projection:

- All D-sugars have the hydroxyl group at the lowest chiral carbon atom on the right.
- All L-sugars have the hydroxyl group at the lowest chiral carbon atom on the left.
- None that the D-and L-natations have no relation to the direction in which a given sugar rotates plane polarized light, a D sugar may be either dextrorotatory or levorotatory.
- The D, L-system of carbohydrate nomenclature is of limited use, since it describes the configuration at only one stereogenic center and says nothing about other stereogenic centers that may by present. The advantage of the system, though, is that it allows a person to relate one sugar to another rapidly.
- Number of isomers depends on the number of stereogenic centers. number of isomers = 2^n.
- Where n is the number of stereogenic center.
- Glyceraldehyde have only one stereogenic center (chiral carbon atom) so it is of two isomers D and L.

Louis Fieser procedure for remembering the names and structures of the eight D-aldohexoses:

Louis Fieser suggested this procedure for remembering the names and structures of the eight D-aldohexoses:

1- Set up eight Fischer projections with the aldehyde group on the top and CH_2OH group at the bottom.

2- Indicate stereochemistry at C_5 by placing all eight hydroxyl groups to the right (D-series).

3- Indicate stereochemistry at C_4 by alternating four hydroxyl groups to the right and four to the left.

4- Indicate stereochemistry at C_3 by alternating two hydroxyl groups to the right, two to the left and so on.

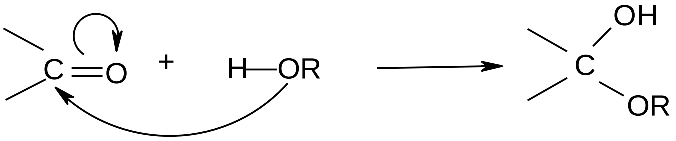
5- Indicate stereochemistry at C_2 by alternating hydroxyl groups to the right. left. right, left and so on.

6- Name the eight isomers according to the mnemonic "All altruists gladly make gum in gallon tanks".

Cyclic Structures of monosaccharides:

Hemiacetal Formation:

- Alcohols undergo a rapid and reversible nucleophilic addition reaction with ketones and aldehydes to form hemiacetals.

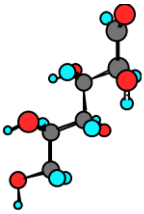


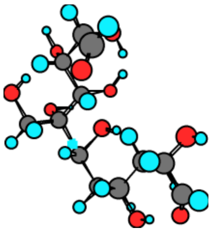
Aldehyde or Ketone
Hemiacetal

- If the hydroxyl group and the carbonyl group in the same molecule: an intramolecular nucleophilic addition can take place, leading to the formation of a cyclic hemiacetal.
- Five and six-membered cyclic hemiacetals are formed.
- Many carbohydrates therefore exist in an equilibrium between open-chain and cyclic forms.

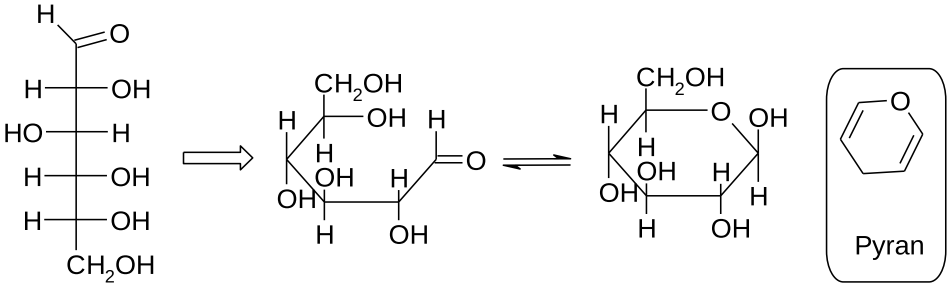
- Glucose exists in aqueous solution primarily as the Six-membered
- Pyranose ring formed by intramolecular nucleophilic addition of the hydroxyl group at C_5 to the C_1 aldehyde group.

Computer structure of D-glucose:





So that the pyranose form of glucose can be formed by the nucleophilic attack of the hydroxyl group on carbon number five on the aldehydic group as shown below.

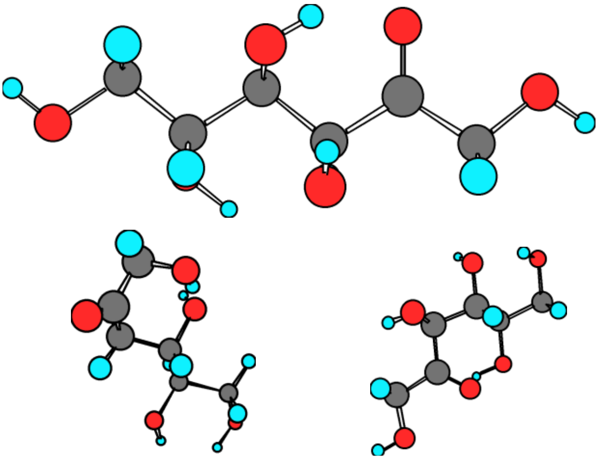


D-Glucose
(Fischer projection)

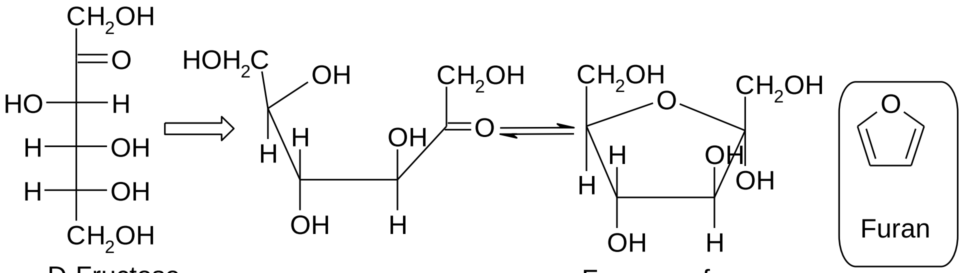
Pyranose form
(Haworth projection)

Fructose exists to the extent of about 20% as the five membered furanose ring formed by addition of the hydroxyl group at C_5 to the C_2 ketone group.

Computer structure of D-fructose:



So that the furanose form of fructose can be formed by the nucleophilic attack of the hydroxyl group of carbon number five on the ketonic group as shown below.



D-Fructose
(Fischer projection)

Furanose form
(Haworth projection)

The interconversion of Fischer and Haworth Projections in a Haworth Projection:

a- The hemiacetal ring is drawn as if it were flat and is viewed edge on with oxygen atom at the upper right.

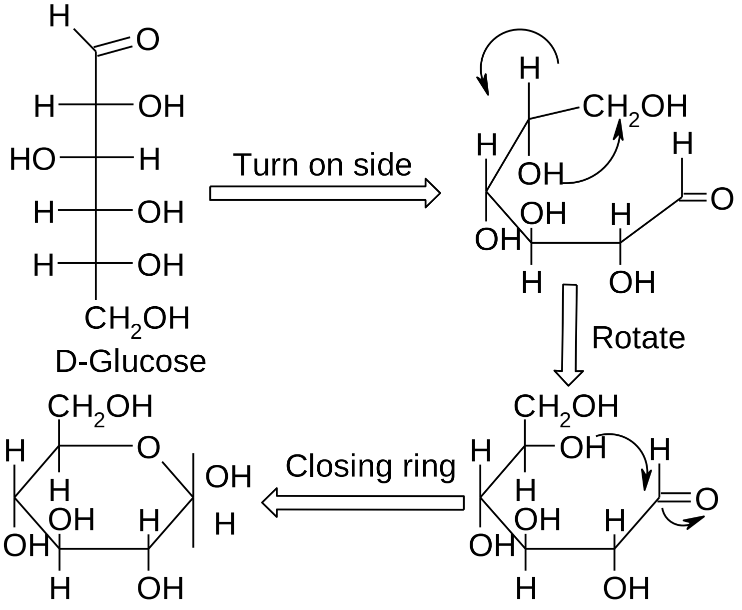
b- Hydroxyl group on the right in a Fischer projection is down in a Haworth projection.

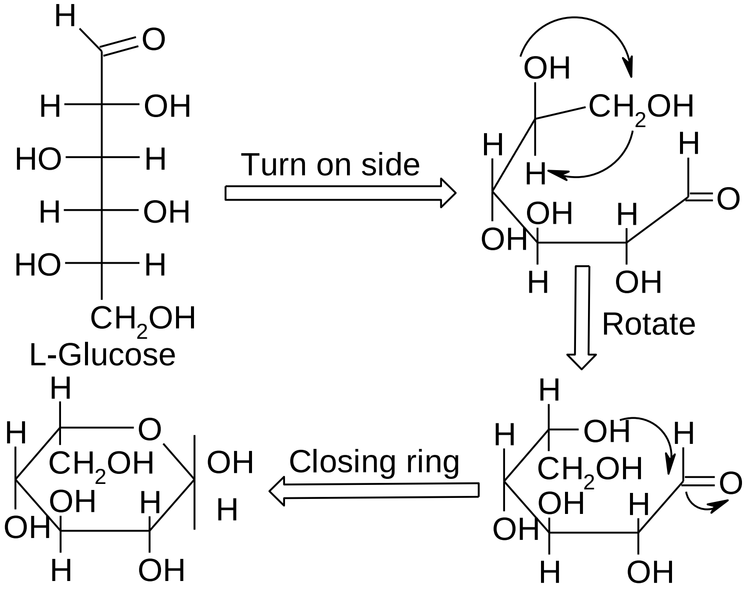
c- Hydroxyl group on the left in a Fischer projection is up in Haworth projection.

d- For D-sugars the terminal CH_2OH group is up in Haworth projection.

e- For L-sugars the terminal CH_2OH group is down in Haworth projection.

Examples:

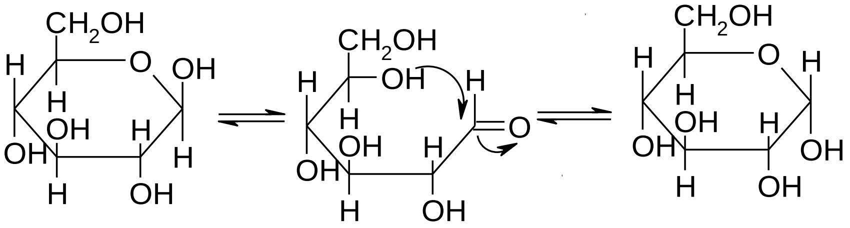




MUTAROTATION:

D-Glucose exists in two crystalline forms; one melts at 150ºC, the other at 146ºC. The fact that neither form shows a carbonyl frequency in the infrared suggests that these two crystalline forms are the α- and β - hemiacetals. X-ray diffraction studies confirm that this is indeed the case. The crystals melting at 150ºC are β-D-glucose, and they show the anomeric hydroxyl group to be in the equatorial position. The crystals of α-D-glucose melt at 146ºC, and crystallography shows the same molecular structure except for the anomeric hydroxyl, which here is in the axial position. A solution freshly prepared by dissolving β-D-glucose crystals in water gives a specific rotation of + 18.7º. This value slowly rises with time to +52.5º. The α-D-glucose shows a rotation, determined immediately upon dissolution, of +112º, but this value also changes slowly with time to a final value of + 52.5º.

The slow change of optical rotation in solution is called mutarotation, and it can be interpreted as involving the interconversion of hemiacetals through the aldehyde intermediate. The equilibrium mixture contains 64 percent of the β-isomer, 36 percent of the α-isomer, and only about 0.02 percent of free aldehyde. Although the equilibrium percentages will differ in different solutions, enough aldehyde is generally present in reaction mixtures to allow the occurrence of ordinary aldehyde reactions such, as oxidation. reduction, and the formation of carbonyl derivatives. The fact that the carbonyl carbon is not asymmetric in the aldehyde form while it is in the hemiacetal, made the original structural and stereochemical determinations more difficult.



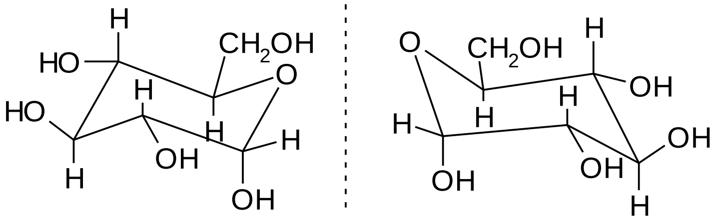
β-D-Glucopyranose $[\alpha]_D=+18.7^o$

Aldehyde form (Open-chain form)

α-D-Glucopyranose $[\alpha]_D=+112^o$

It should be understood that the mirror image of D is L, the mirror image of (+) is (-), but the mirror image of α is α and not β, as shown in the following example (i.e., α-implies an axial 1-hydroxyl group in both enantiomers).

Mirror images



α-D-(+) Glucopyranose α-L-(-) Glucopyranose

Conformations of Monosaccharides:

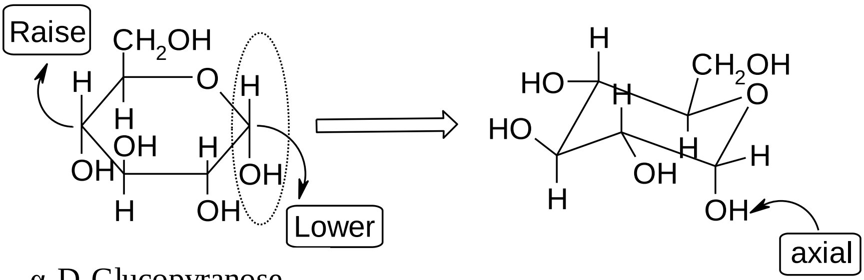
Pyranose rings like cyclohexane rings have a chair like geometry with axial and equatorial substituents.

Haworth projections can be converted into chair representations by the following three steps:

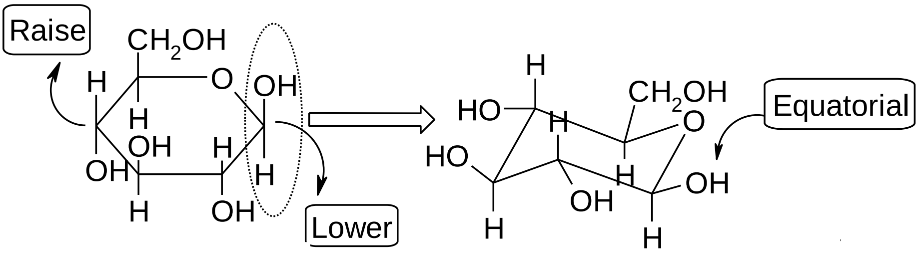
1- Draw the Haworth projection with the ring oxygen atom at the upper right.

2- Raise the left most carbon atom(C_4) above the ring plane.

3- Lower the anomeric carbon atom (C_1) below the ring plane.



α-D-Glucopyranose



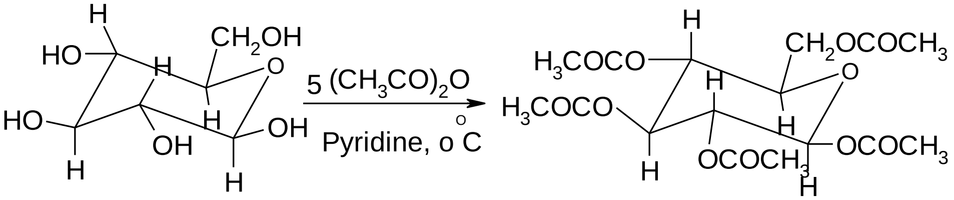
β-D-Glucopyranose

Note that in β-D-glucopyranose all the substituents on the ring are equatorial, thus β-D-glucopyranose is the least sterically and most stable of the eight D-aldohexoses.

CHAPTER 2
Reactions and Interconversion of Monosaccharides

1- Ester and Ether Formation:

Estrification is normally carried out by treating the carbohydrate with an acid chloride or acid anhydride in the presence of a base. All the hydroxyl groups react, including the anomeric one to produce penta-*O*-acetyl derivative.

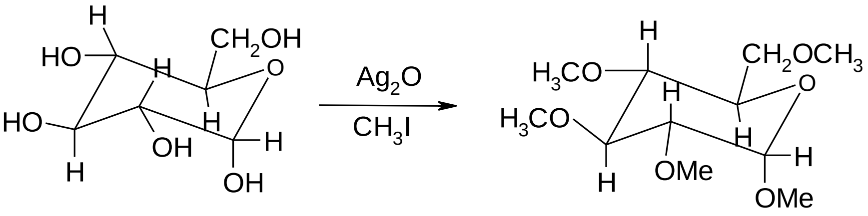


β-D-Glucopyranose

Penta-*O*-acetyl-β-*D*-Glucopyranose-91 %

Carbohydrates can be converted into ethers by treatment with an alkyl halide in the presence of base (the Williamson ether synthesis). Normal Williamson conditions using a strong base tend to degrade the Sensitive sugar molecules. In 1903 Purdie showed that silver oxide works particularly well and that high yields of ethers are obtained.

For example. α-D-glucopyranose is converted into its pentamethyl ether in 85% yield on reaction with iodomethane and silver oxide.

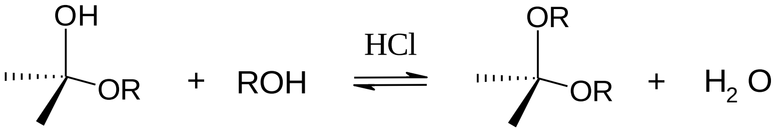


α-D-Glucopyranose

α-D-Glucopyranose Pentametyl ether-85 %

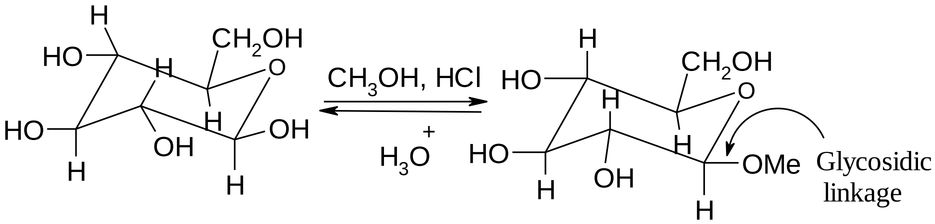
2- Glycoside Formation:

Treatment of hemiacetal with an alcohol and acid catalyst yields an acetal.



Hemiacetal Acetal

In the same way, treatment of a monosaccharide hemiacetal with an alcohol and an acid catalyst yields an acetal in which the anomeric hydroxyl has been replaced by an alkoxy group.



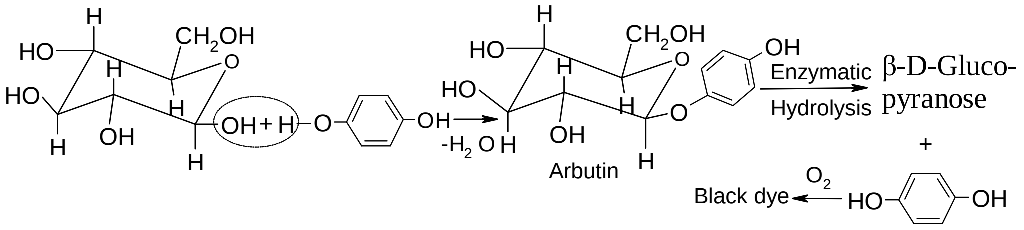
β-D-Glucopyranoside Methyl-β-D-Glucopyranoside (an acetal)

carbohydrate acetals are called glycosides. They are named by citing the alkyl group and adding the -oside suffix to the name of the specific sugar.

Glycosides are stable to water, and can be converted back to the free monosaccharide by hydrolysis with aqueous acid.

Sugar Derivatives in Nature:

Many natural products contain sugars attached to other types of chemical structures. In arbutin, the sugar is attached to a phenyl derivatives. Arbutin is hydroquinone β-D-glucoside, and it occurs in many plants. In the autumn, leaves from certain pear trees turn black instead of yellow and red. This black colour results from the fact that these leaves contain a high concentration of arbutin, which on enzymatic hydrolysis librates hydroquinones, which is oxidized by air to a black dye.

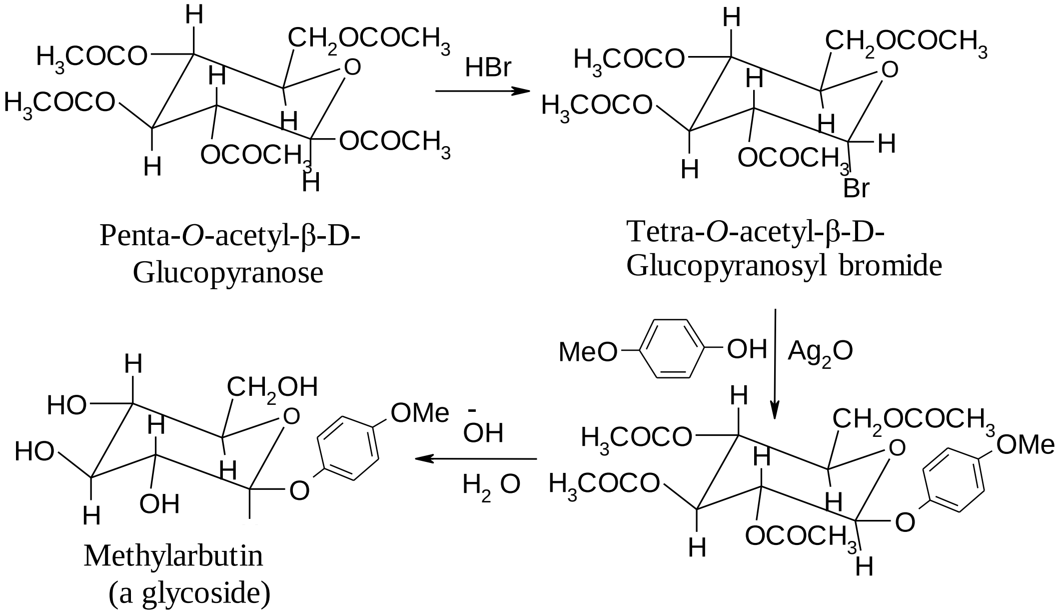


Arbutin

The moiety attached to the sugar is called the aglycone. When the aglycone is a dye or pigment such as an anthocyanin or anthoxanthin, the resulting compounds are the natural chemicals which provide the colours for flowers.

Koenigs Knorr reaction:

This reaction can be used for the preparation of only β-anomer from both α (alpha) and β (beta) anomers.



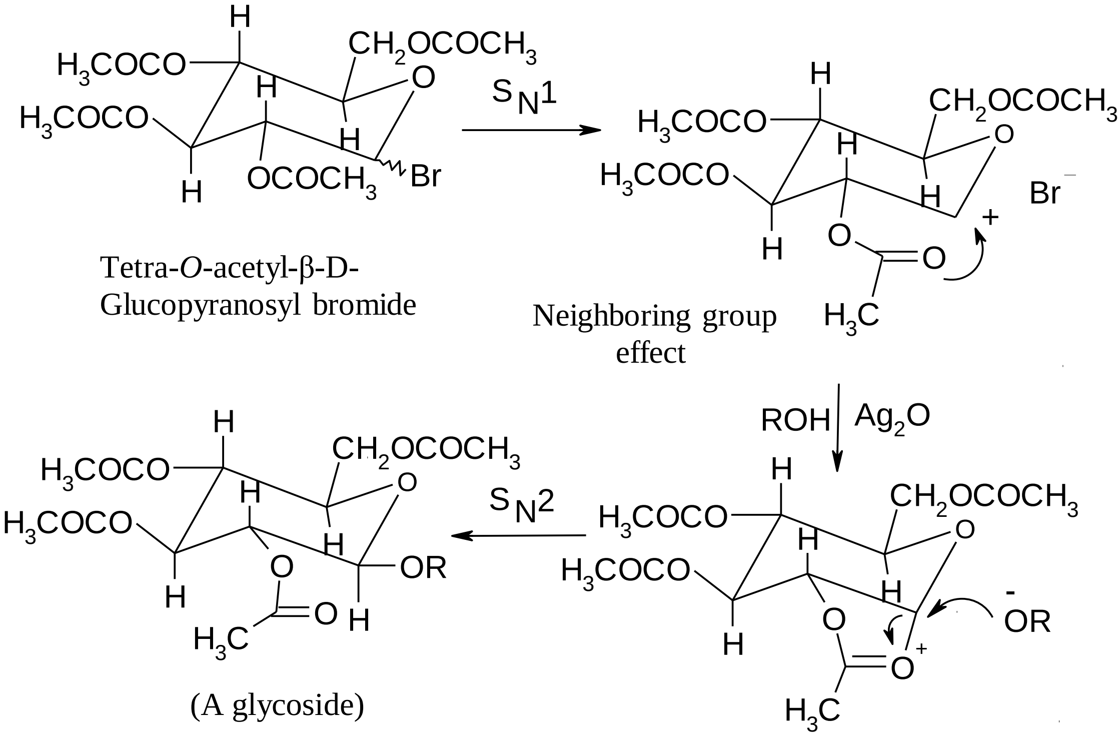
Penta-*O*-acetyl-β-D-Glucopyranose

Tetra-*O*-acetyl-β-D-Glucopyranosyl bromide

Methylarbutin
(a glycoside)

The mechanism of Koenigs Knorr reaction is as follows:

Both alpha and beta anomers of tetraacetyl-D-glycopyranosyl bromide give the same β-glycoside product. Suggesting that both anomers react by a common pathway.



Tetra-*O*-acetyl-β-D-Glucopyranosyl bromide

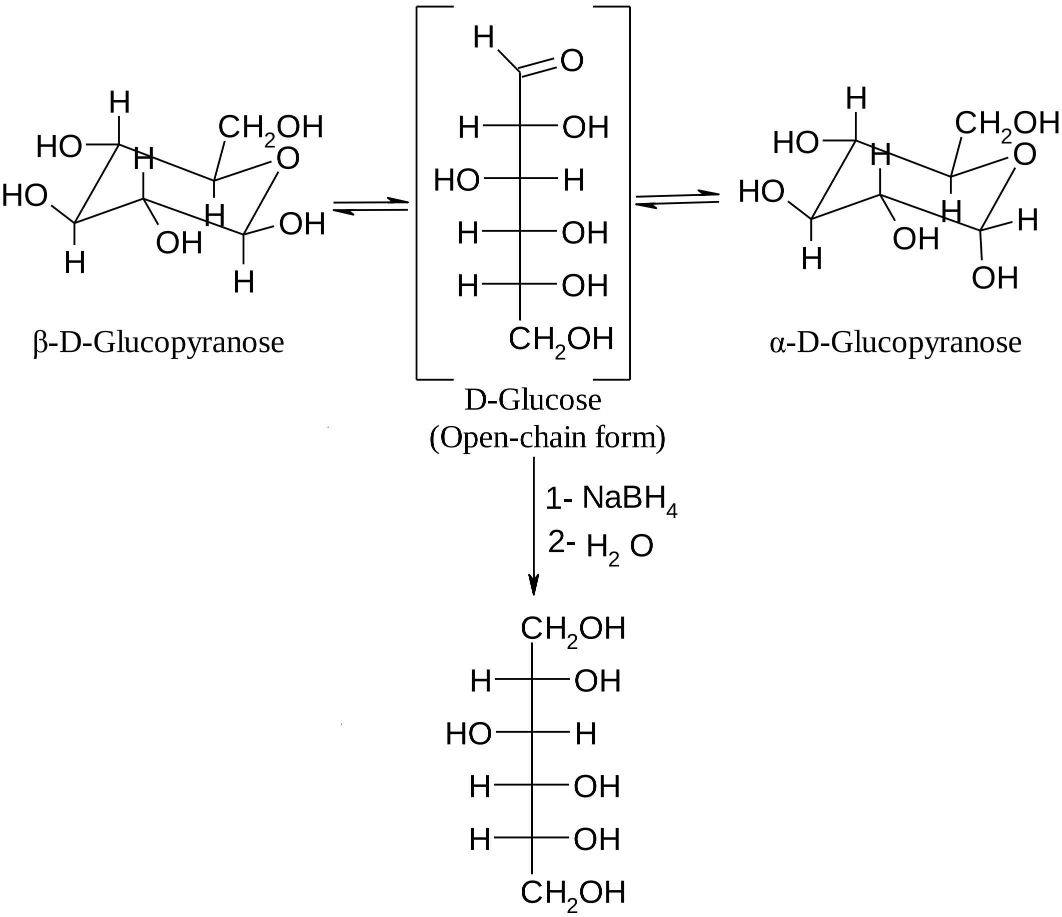
Neighboring group effect

(A glycoside)

Oxonium ion

3- Reduction of Monosaccharides:

Treatment of an aldose or a ketose with $NaBH_4$ reduces it to a polyalcohol called an alditol.



β-D-Glucopyranose

D-Glucose
(Open-chain form)

α-D-Glucopyranose

D-Glucitol (D-sorbital)
an alditol

4-Oxidation of Monosaccharides:

Like other aldehydes. aldoses are easily oxidized to yield carboxylic acids. Aldoses reacts with oxidizing reagents to yield the oxidized sugar and a reduced metallic.

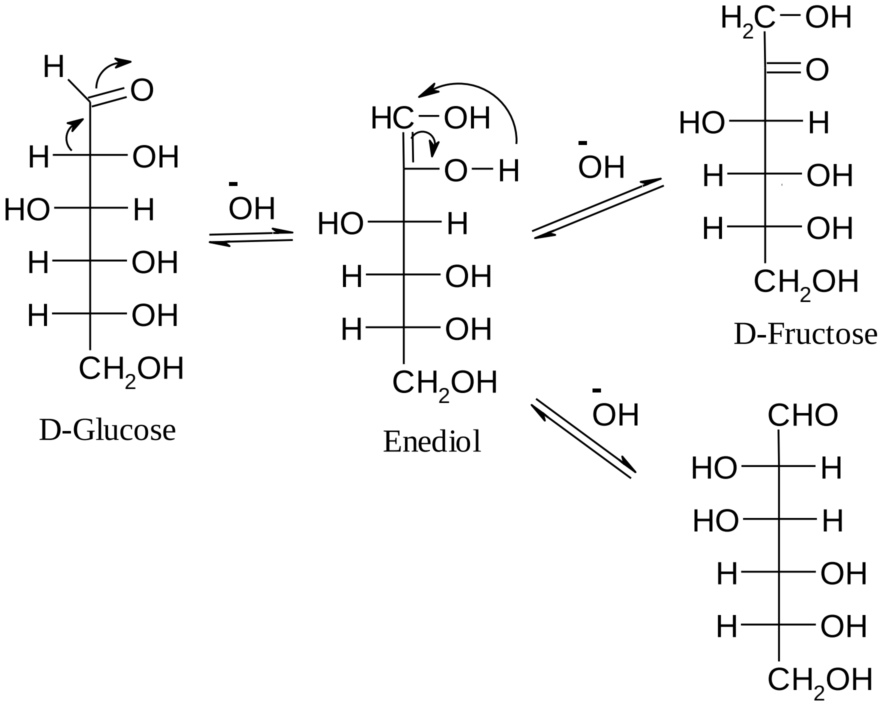
Oxidizing reagents like:

1- Tollen’s reagent (Ag^+ in aqueous ammonia) produce silver metal as a mirror.
2- Fehling’s reagent (Cu^{++} in aqueous sod. tartrate) produce reddish precipitate of cuprous oxide.
3- Benedict’s reagent (Cu^{++} in aqueous sod. citrate) produce reddish precipitate of cuprous oxide.

All aldoses are reducing sugars because they contain a free aldehyde group (or hemiacetal), but some ketoses (α-hydroxy ketones or hemiketal) are reducing sugars as well. Therefore, a free aldehyde (or hemiacetal), an α-hydroxy ketones or (hemiketal) is necessary for a positive test. For example, fructose reduces Tollen's, Fehling’s and Benedict’s reagents even though it contains no aldehyde group. This occurs because fructose is readily isomerized to an aldose in the basic solution by a series of Keto-enol tautomeric shifts. Once formed the aldose is oxidized normally. These tests can not be used to distinguish between aldoses and ketoses, since both react with the reagent.

ALKALINE ISOMERIZATION OF MONOSACCHARIDES:

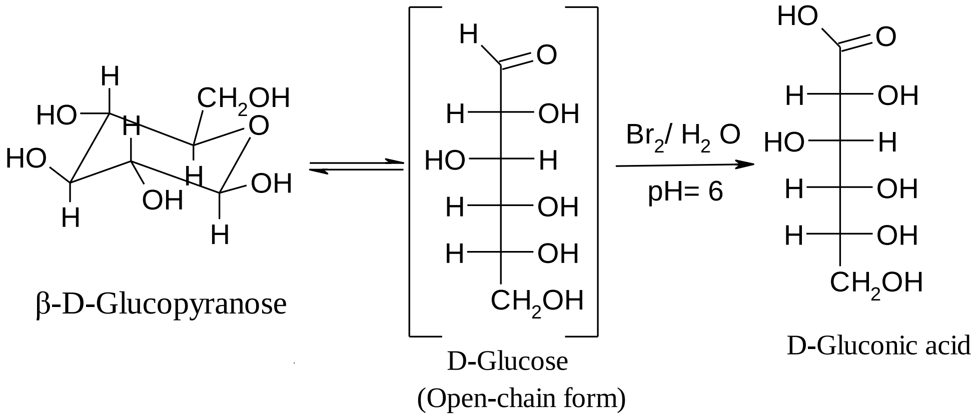
Chemical isomerization of glucose to fructose, an internal oxidation-reduction reaction, can be accomplished by treatment with alkali, but a variety of other isomeric and decomposition products are also produced. For example, when D-glucose is treated with base, the unstable enediol is formed. In this transient intermediate the stereochemistry of C-2 is lost. The enediol can then revert back to any of the three more stable hydroxy carbonyl compounds.



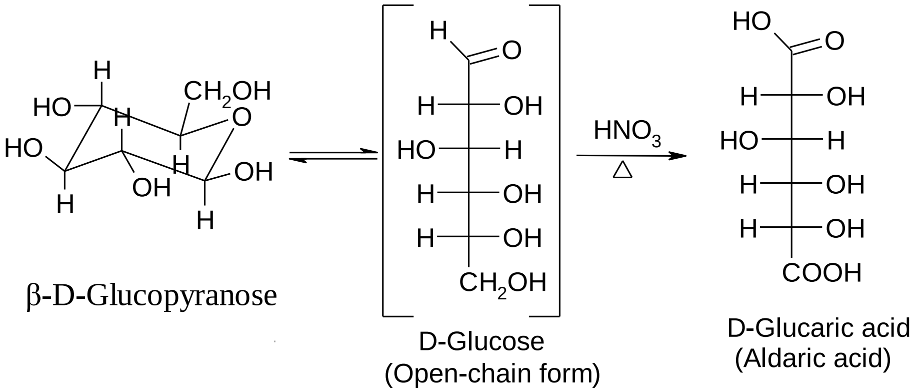
D-Mannose

The enzyme catalyzed interconversion of glucose and fructose plays a key role in carbohydrate metabolism in living organisms.

Glycosides however are nonreducing. They don’t react with Tollen's reagent because the acetal group can’t open to a free aldehyde (open-chain form) under basic conditions. Although the Tollen and Fehling reagents serve as useful tests for reducing sugars. They don’t give good yields of carboxylic acid products because the alkaline conditions used cause decomposition of the carbohydrate skeleton. It has been found however that a buffered solution of aqueous bromine oxidizes aldoses to monocarboxydic acids called aldonic acids. The reaction is specific for aldoses. ketoses are not oxidized by bromine water.



If more powerful oxidizing agent such as warm dilute nitric acid used, Aldoses oxidized to dicarboxylic acids called aldaric acids. Both CHO and CH_2OH are oxidized.



5-Synthesis and interconversion of the monosaccarides:

a- Conversion of an aldose into the next higher aldose (Chain Lengthening).

1- The Kiliani-Fischer synthesis:

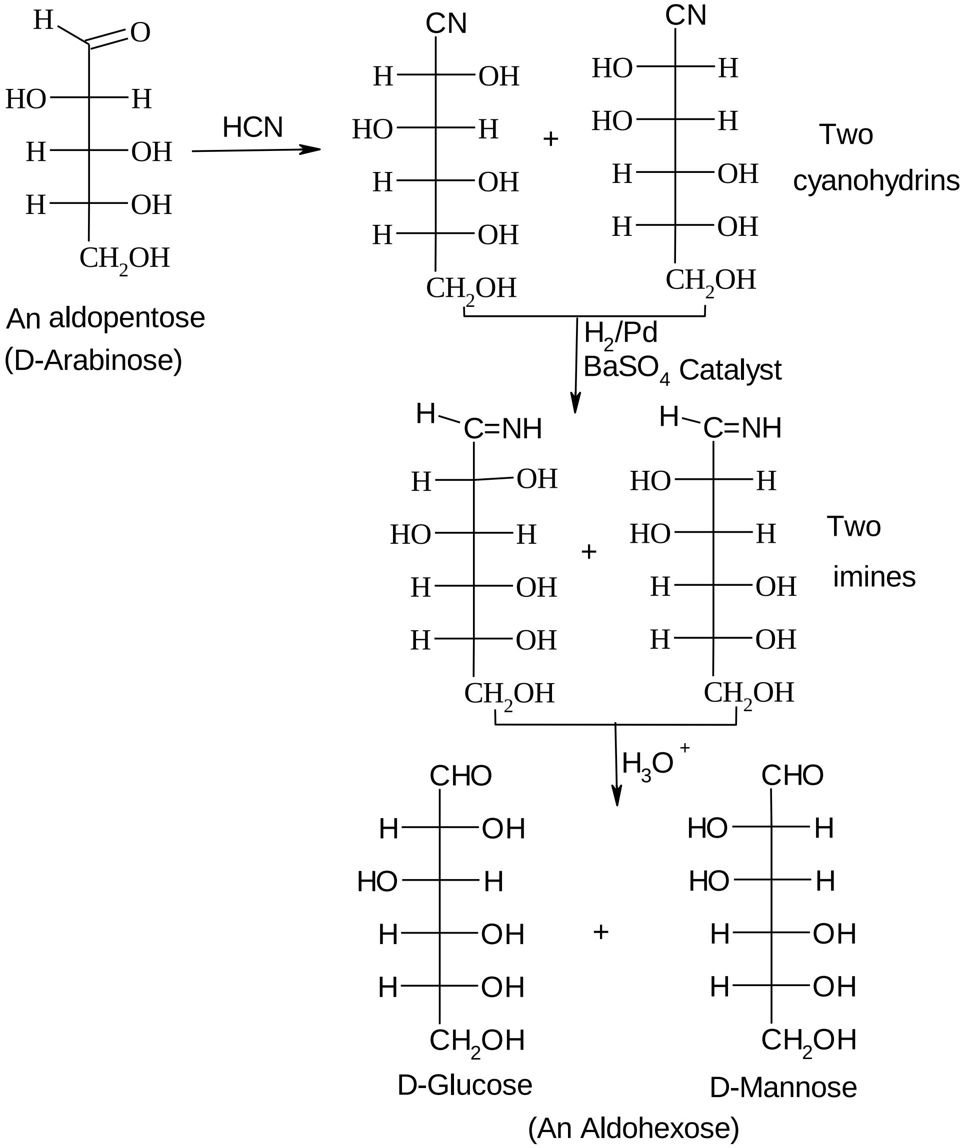
This method has been carried out as follows:

a- Aldose reacts with HCN to form cyanohydrins.

b- Conversion of the nitrile into an imine intermediate by catalytic hydrogenation over a palladium catalyst.

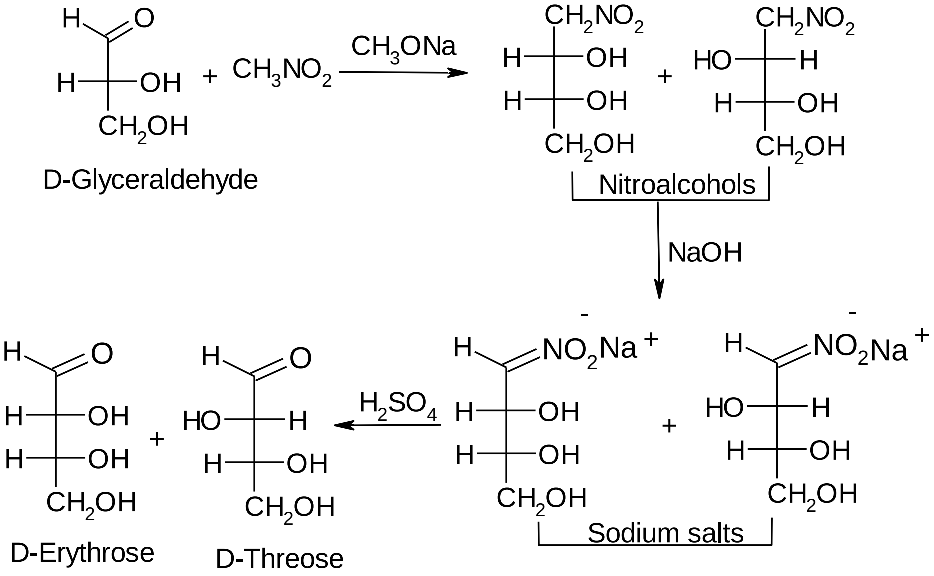
c- Hydrolysis of the intermediate yields an aldehydic group.

So that Kiliani-Fischer synthesis lengthens an aldose chain by one carbon.



2- Sowden-Fischer Nitromethane Synthesis:

This is a more recent method and involves the reaction of an aldose with nitromethane in the presence of a base to produce two different nitroalcohols which are separated by fractional crystallization. The individual nitroalcohols are next treated with sodium hydroxide solution to give the corresponding sodium salts which may then be decomposed to give the higher aldoses.



B) Conversion of an aldose into the next lower aldose (Chain Shortening):

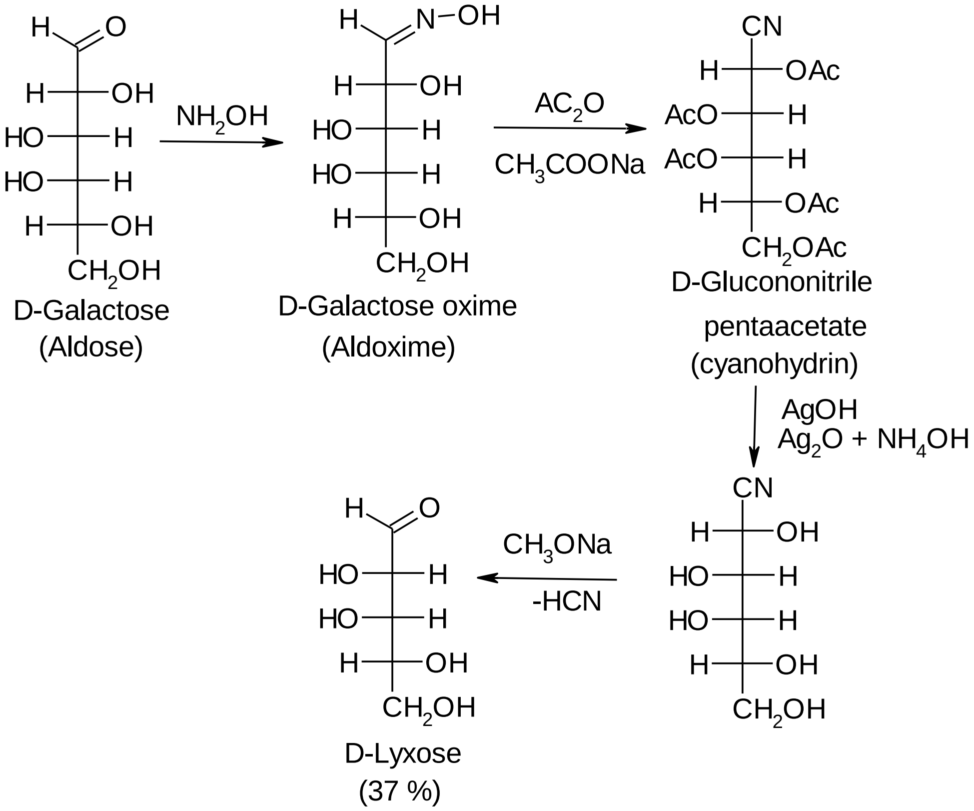
1- The Wohl Degradation:

This degradation can be carried out by the following:

a- The aldose aldehyde carbonyl group is first converted into oxime by treatment with hydroxylamine, and

b- The resulting cyanohydrin loss HCN under basic conditions (a retro nucleophilic addition reaction).

So that Wohl degradation shortens an aldose chain by one carbon.



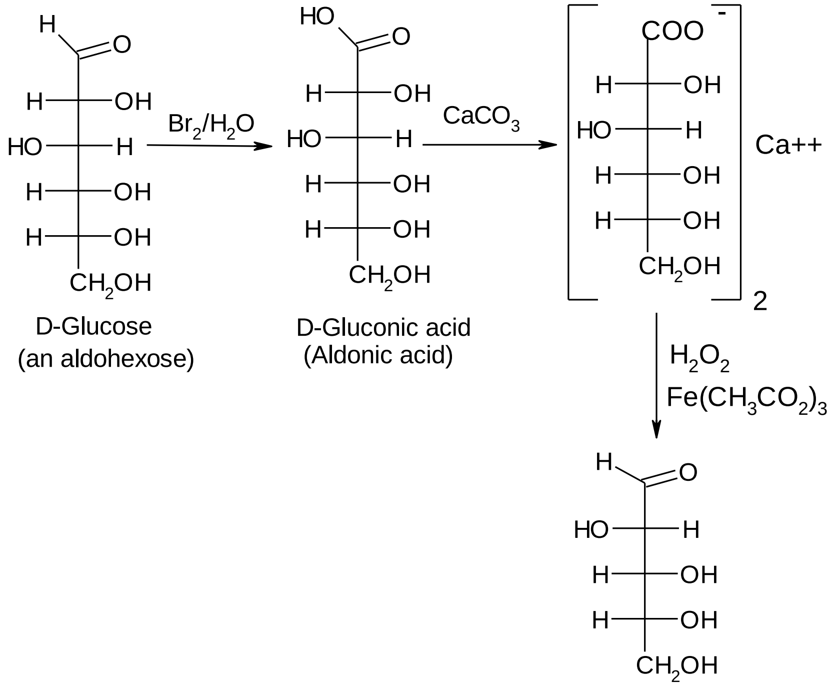
2- Ruff's Method:

In this method

1- The aldose is oxidized with bromine water to give the corresponding aldonic acid.

2- The aldonic acid is next treated with calcium carbonate to give the calcium salt of the acid.

3- This is then treated with hydrogen peroxide and ferric acetate (Fenton's reagent) so that CO_2 and H_2O are eliminated to give the lower aldose.



D-Glucose
(an aldohexose)

D-Gluconic acid
(Aldonic acid)

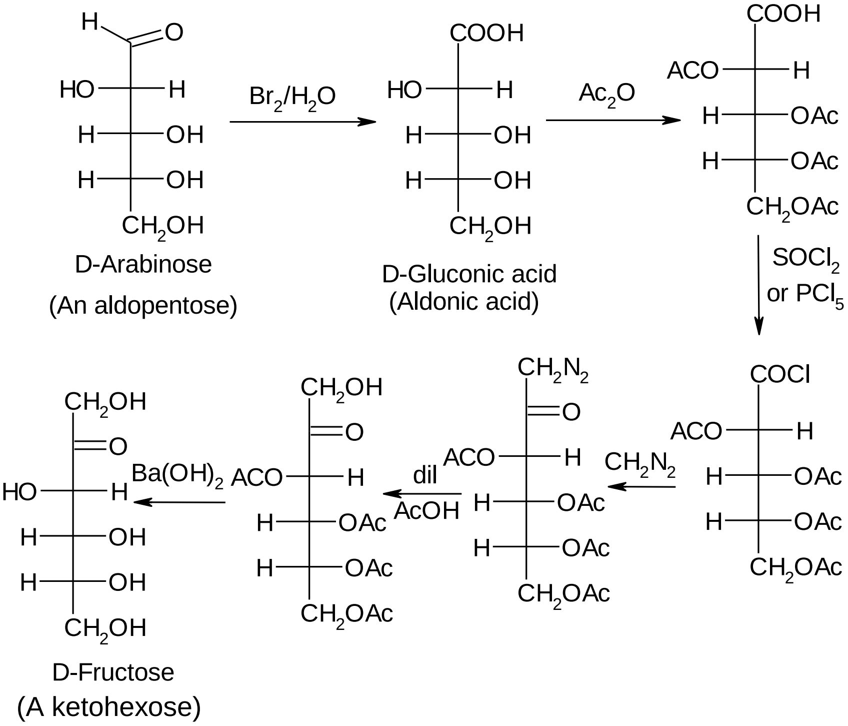
D-Arabinose
(An aldopentose)

C) Conversion of an aldose into the next higher ketose:

Wolfrom’s method:

In this method:

1- Oxidation of aldose to aldonic acid by Br_2/H_2O.
2- Acetylated with acetic anhydnide.
3- Acetylated aldonic acid is treated with $SOCl_2$ or PCl_5 to give the corresponding acid chloride.
4- Treatment of acid chloride with diazomethane followed by heating with aqueous acetic acid.
5- Deacetylation by alkaline hydrolysis gives next higher ketose.

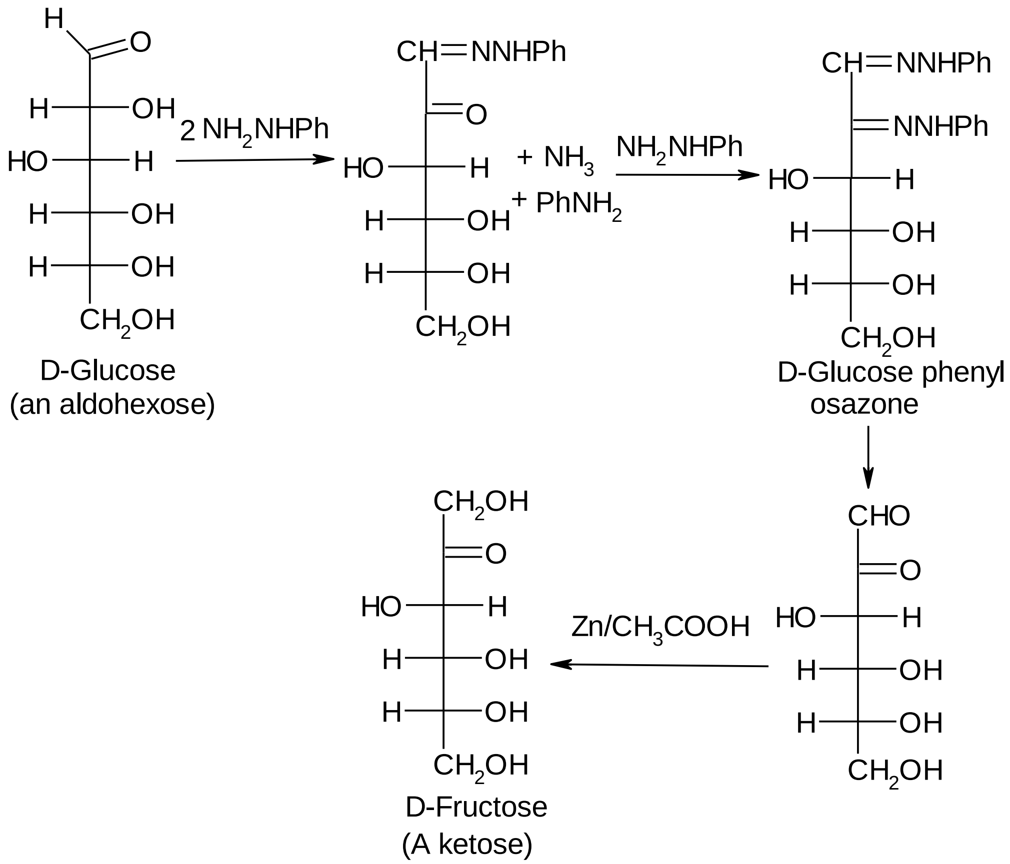


D) Conversion of an aldose into the corresponding ketose:

This conversion can be done by the following.

a- The aldose reacted with excess of phenyl hydrazine to give the corresponding osazone.

b-Hydrolysis with HCI to give the osone which is then reduced with zinc and acetic acid glacial to give ketose.



E) Conversion of the ketose into the corresponding aldose:

This conversion can be done by the following.

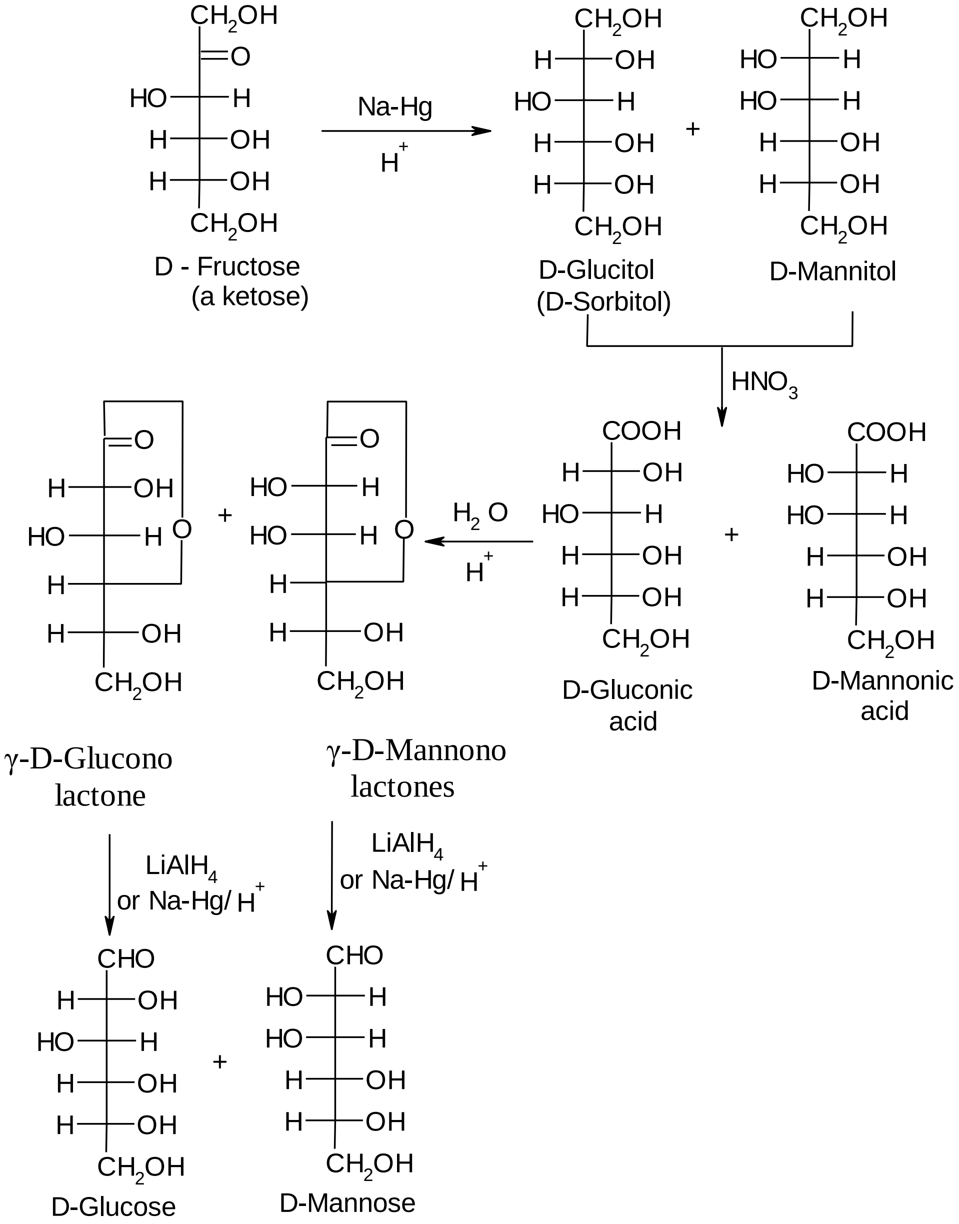
1- Ketose is reduced with sodium amalgam in the presence of a trace of acid to form polyhydric alcohol.

2- These alcohol are next oxidized with nitric acid to give monobasic aldonic acids.

3- The aldonic acids on treatment with dilute HCI give γ-lactones.

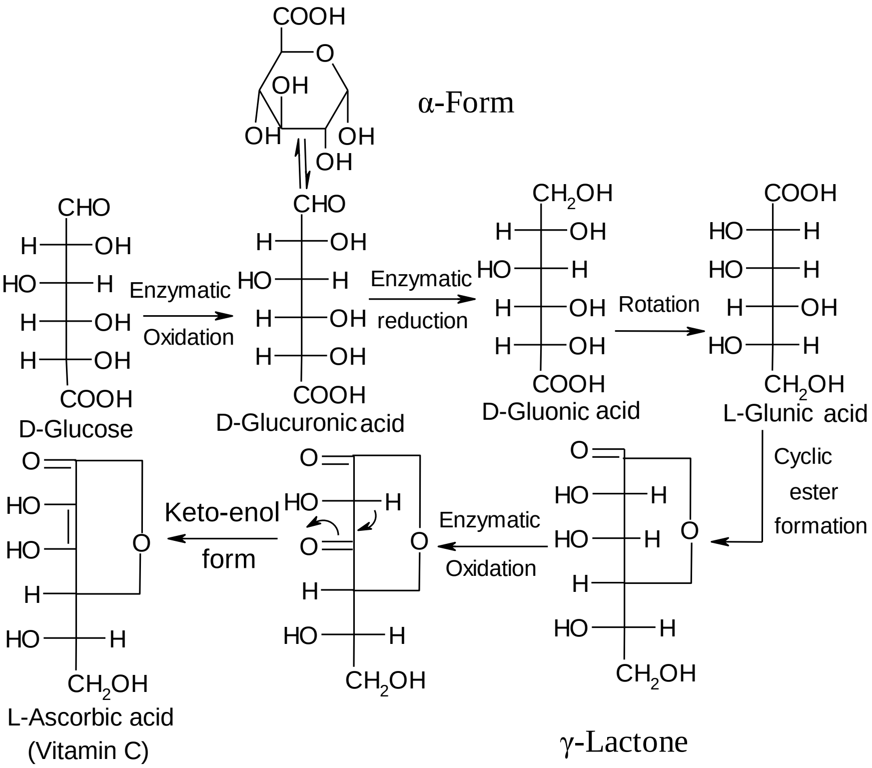
4- These lactones are solids and separated by fractional crystallization.

5- The individual lactones are then reduced with lithium aluminum hydride or sodium amalgam in a weakly acidic solution to yield aldoses which are isomeric with the original ketose.



OPTIONAL MATERIAL:

Ascorbic and glucuronic acids. Two other oxidation derivatives of monosaccharides are particularly important. Ascorbic acid is vitamin C, which is widely distributed in nature, especiall, in green plants. It functions as a biological oxidation-reduction reagent. acting as a hydrogen carrier.



Glucuronic acid, in which the primary hydroxyl group of glucose has been oxidized to a carboxylic acid while the aldehyde group remains intact, is a major building block in many naturally occurring polysaccharides.

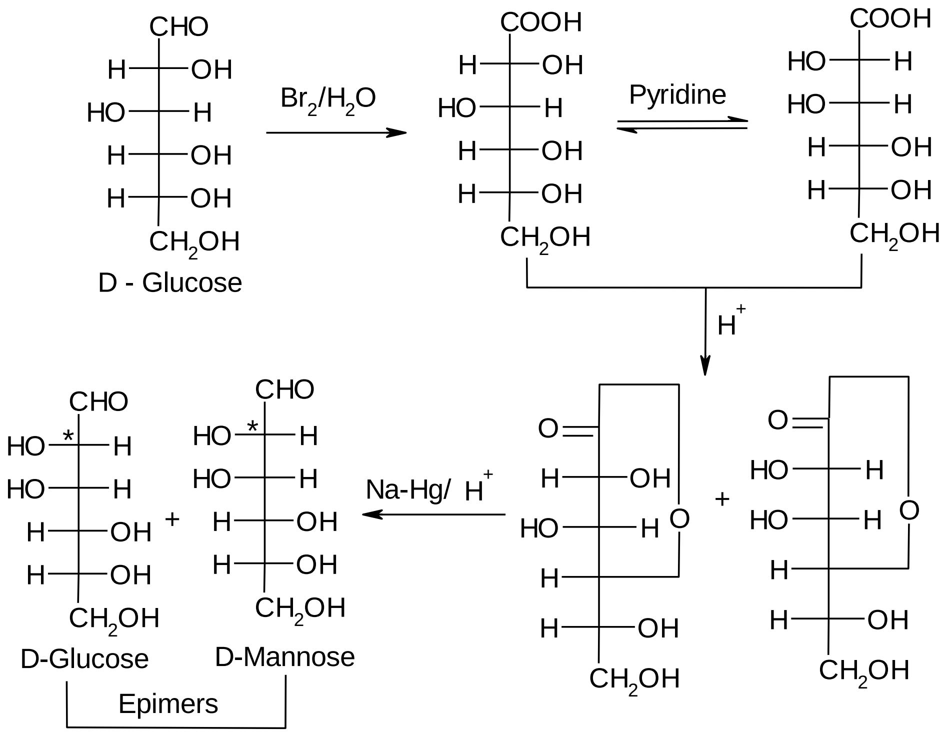
F) Conversion of an aldose into its epimeric aldose (Epimerisation):

This conversion can be done by the following.

a- The aldose is first oxidized to aldonic acid by Br_2/H_2O.

b- This aldonic acid is heated in pyridine or' quinoline to give an equilibrium mixture of the original acid and its isomer.

c-These aldonic acids are converted into lactones seprated and reduced to give aldose and its epimer at C_2.



This change of configuration of one asymmetric carbon in a compound containing two or more asymmetric carbon atom is know as epimerization.

Chapter 3

Disaccharides

When a hydroxyl group of one monosaccharide molecule acts as the alcohol to form a glycosidic linkage with the hemiacetal group of a second monosaccharide, the resulting glycoside is called a disaccharide. They are therefore acetals, formed from two molecules of monosaccharides by the elimination of one molecule of water. Conversely, hydrolysis of a disaccharide either by water in the presence of an acid or by enzymes yields two molecules of

$$\underset{\text{Monosaccharides}}{C_6H_{12}O_6 + C_6H_{12}O_6} \rightleftharpoons \underset{\text{Disaccharide}}{C_{12}H_{22}O_{11}} + H_2O$$

There for Disaccharides are carbohydrates that are made of two monosaccharide units. On hydrolysis a molecule of disaccharide yields two molecules of monosaccharide.

Among the most common disaccharides are:

1- (+) – Sucrose (Can or beet sugar)
2- (+) – Lactose (Milk sugar)
3- (+) – Maltose (Malt sugar)
4- (+) - Cellobiose.

As with the monosaccharrides, we shall focus our attention on the structure of these molecules on which monosaccharides make up the disaccharide, and how they are attached to each other. In doing this, we shall also learn something about the properties of these disaccharides.

Sucrose

(Cane sugar or Beet sugar $C_{12}H_{22}O_{11}$)

The most important one which is the table sugar and we eat every day.

Occurrence:

Sugar cant (16-to-20%) Sugar beets (10-to-I 5%)

Pineapples (10-to-12%)

Maple sap (2-to-4%). apricot, banana, mango. Almonds, Café, and honey.

Structure Determination of Sucrose:

The structure of sucrose has been derived from a consideration of facts and conclusions such as the following.

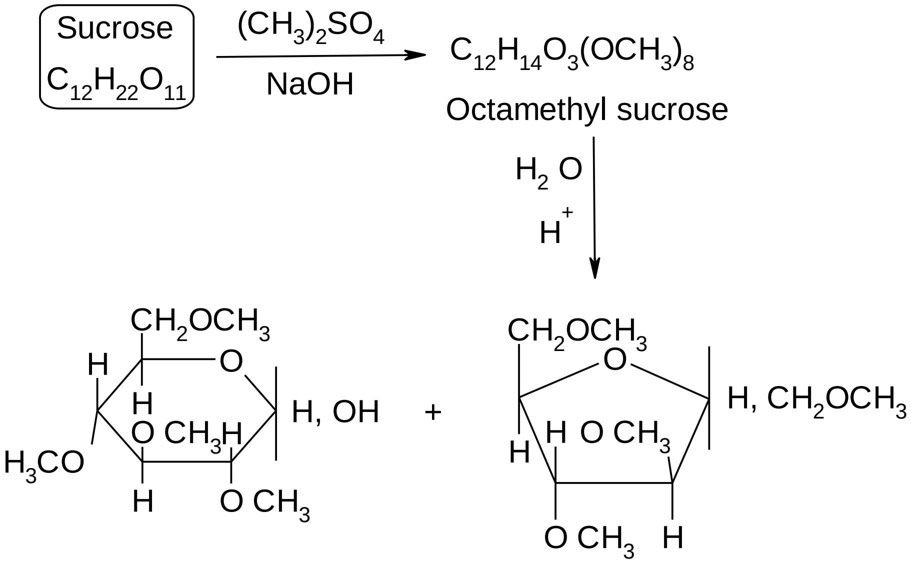
1- Elemental analysis and molecular weight determination show that the molecular formula of sucrose $C_{12}H_{22}O_{11}$.

2- Sucrose reacts with acetic anhydride in the presence of sodium acetate to form sucrose octaacetate. This reaction indicates the presence of eight hydroxyl groups in a sucrose molecule. Since sucrose is a stable compound. the eight hydroxyl groups roust be present on separate carbon atoms.

3- Hydrolysis of sucrose with dilute acids yields an equimolecular mixture of D-glucose and D-fructose. This indicates that the sucrose molecule is made up of one unit of each of these monosaccharides.

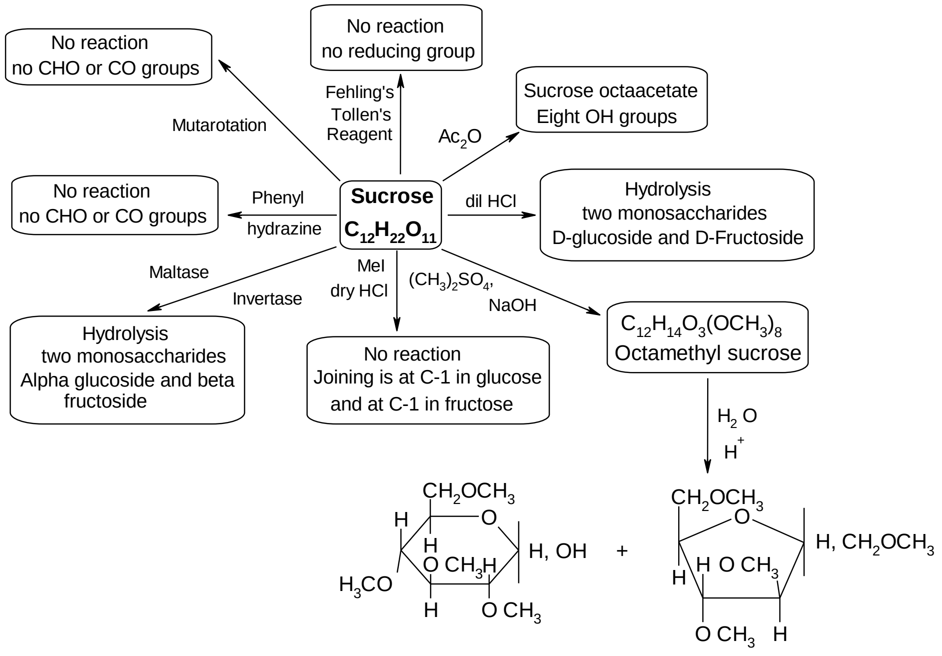
4- Sucrose does not reduce Tollen's reagent or Fehling’s solution. does not form an osazone (except on prolonged boiling, when glucosazone is formed due to hydrolysis of sucrose). Does not fond methyl glycosides. and dose not undergo mutarotation. All these observations indicate that the cyclic forms of glucose and fructose are joined together through glycosidic linkage at points where the carbonyl groups would otherwise become available, that, is C-1 in glucose and C-2 in fructose.

5- Sucrose reacts with dimethyl sulphate in an alkaline solution to form octamethyl sucrose which on hydrolysis yields a mixture of 2, 3, 4, 6-tetramethyl-D-glucopyranose and 1, 3, 4, 6-tetramethyl-D-fructofuranose. The formation of these compounds indicates that the glucose unit in sucrose has a pyranose form (6-membered ring), and the fructose unit the furanose form (5-membered).



2,3,4,6-Tetramethyl-D-Glucopyranose (α-and β-forms) 1,3,4,6-Tetramethyl-D-fructofuranose furanose (α-and β-forms)

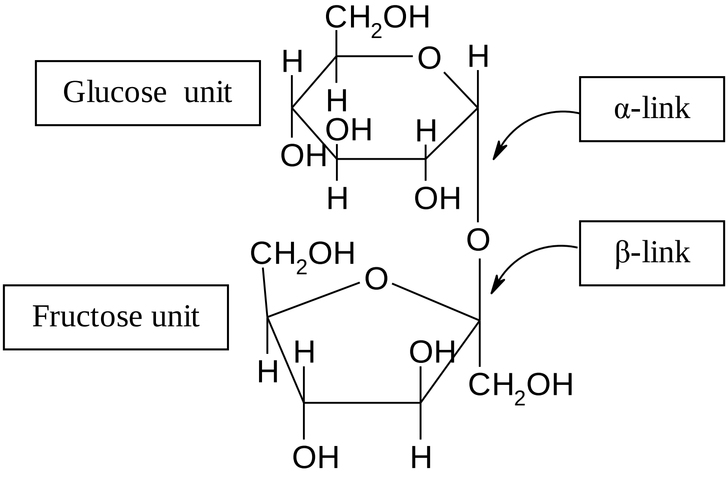
6- Sucrose is hydrolyzed by maltase. an enzyme that hydrolyses only α-glycosides. It is also hydrolyzed by invertase. an enzyme that hydrolyses β-but not α-fructofuranosides. These observations indicate that sucrose Is both an α-glucoside and a β-fructoside.



2,3,4,6-Tetra methyl-D-Glucopyranose (α and β- forms)

1,3,4,6-Tetra methyl-D-fructofuranose furanose (α and β- forms)

The above evidence clearly indicates that (+) - sucrose is made up of a D-glucose unit and a D-fructose unit. and has the following structure.



Sucrose

Physical Properties of Sucrose:

Sucrose is a colourless, odourless, crystalline solid, m.p. 184-5 °C. It very soluble in water, but only slightly soluble in alcohol. An aqueous solution of sucrose is dextrorotatory and does not exhibit mutarotation.

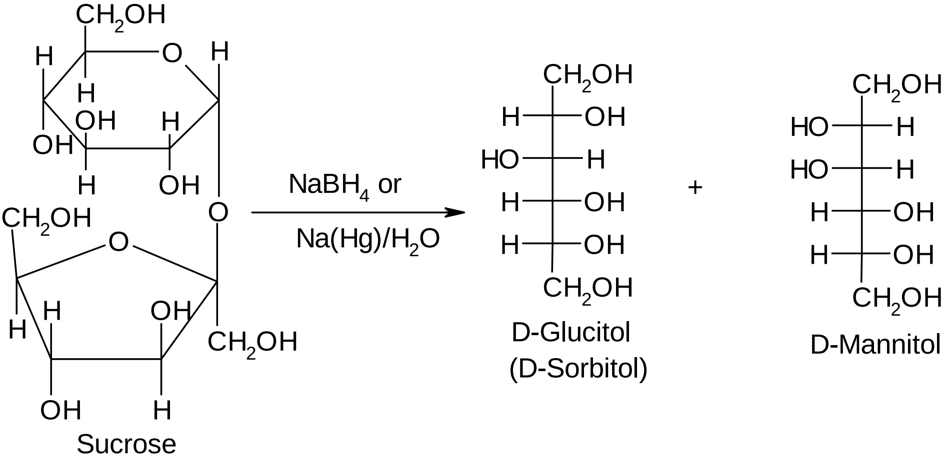
Chemical properties of Sucrose:

Sucrose is a on-reducing sugar. It does not react with Tollen’s reagent. Fehling's solution, hydrogen cyanide, hydroxylamine or phenyl hydrazine. However, it gives the following reactions.

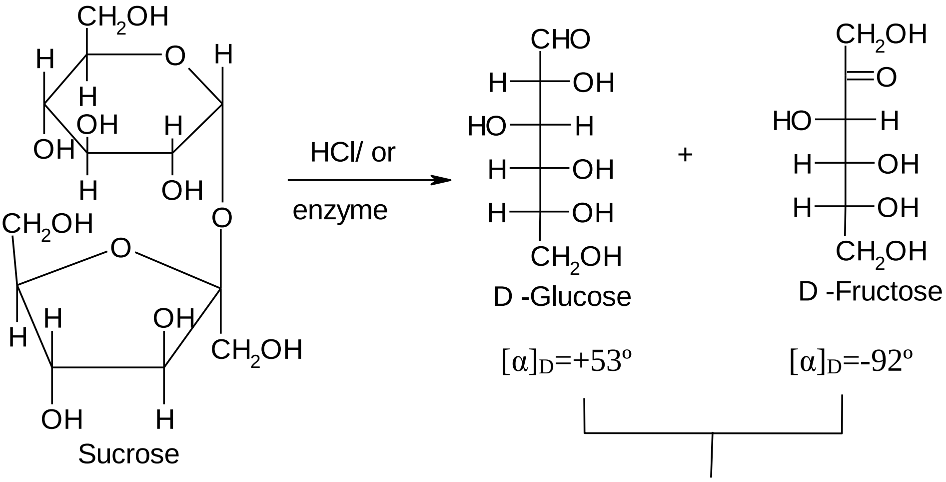
(1) **Effect of Heat:** When sucrose is heated to 210 °C. it forms a own mass known as Caramel which because of its colour and characteristic flavour, is used as colouring and flavouring material in foods and candies. At higher temperatures. sucrose chars to almost puro carbon and gives vapour of carbon dioxide, carbon monoxide. methane, ethylene, acetylene, acetone, formic acid, acetic acid, ethanal, and acrolein.

(2) **Oxidation:** Oxidation of sucrose with concentrated nitric acid yields a mixture of oxalic acid. tartaric acid. and D-glucanic acid.

(3) **Reduction (Hydrogenation)**: Reduction of sucrose with sodium borohydride or sodium amalgam in water under controlled conditions yields a mixture of D-sorbitol and D-mannitol



(4) **Hydrolysis (Invert Sugar or Invertose):** Hydrolysis of sucrose with hot dilute acid yields D-glucose and D-fructose.



$[\alpha]_D = +53^o$ $[\alpha]_D = -92^o$

$[\alpha]_D = +66.5^o$

Invert Sugar

$[\alpha]_D = (+53^o - 92^o) = -39^o$

Sucrose is dextrorotatory, its specific rotation being -66.5° D--glucose is also dextrorotatory, $[\alpha]_D$=+53°. but D-fructose has a large negative rotation, $[\alpha]_D$= -92°. Since D-fructose has a greater specific rotation than D-glucose. the resulting mixture is laevorotatory.

Because of this, the hydrolysis of sucrose is known as the Invert of sucrose, and the equimolecular mixture of glucose and fructose is known as invert sugar or invertose. The inversion (i.e., hydrolysis) of sucrose can also be brought about by the enzyme *invertase.* which is found in yeast.

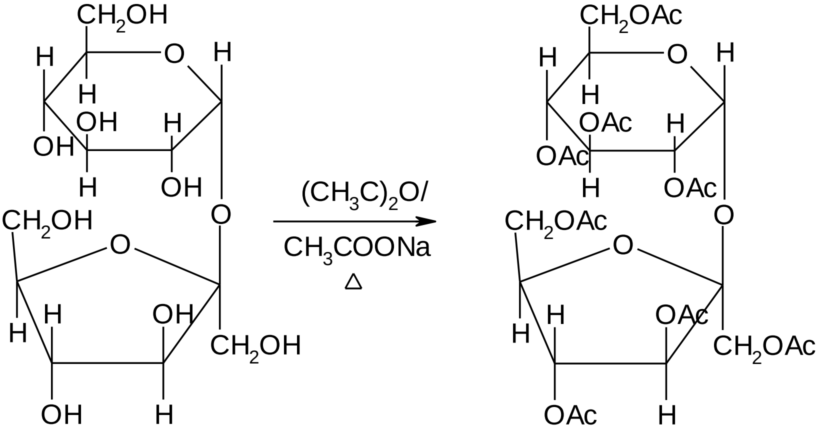
(5) **Reaction with Metallic Hydroxides** (Formation of Sucrosates). Sucrose in aqueous solution reacts with hydroddes of calcium strontium, and barium to produce insoluble compounds called sucrosates. These compounds are readily decomposed when carbon dioxide is pessed into their aqueous suspensions. The strontium compound is used for isolating pure sucrose from non-crystallisable molasses.

$$C_{12}H_{22}O_{11} + Sr(OH)_2 \xrightarrow{-H_2O} C_{12}H_{21}O_{11}SrOH \xrightarrow{CO_2} C_{12}H_{22}O_{11} + SrCO_3$$

Sucrose Sucrose

(6) **Reaction with Acetic anhydride** **(Acetylation).**

Sucrose reacts with acetic anhydride in the presence of sodium acetate to form sucrose octaacetate.



Sucrose

Ac= CH_3CO

Sucrose octaacetate
m.p 70 °C

(7) **Fermentation.** An aqueous solution of sucrose is readily fermented by yeast to give ethyl alcohol. The enzyme invertase present in yeast first converts sucrose into glucose and fructose. These sugars are then decomposed by the enzyme zymase (also present in yeast) to give ethyl alcohol and carbon dioxide.

$$\underset{\text{Sucrose}}{C_{12}H_{22}O_{11}} + H_2O \xrightarrow{\text{invertase}} \underset{\text{glucose}}{C_6H_{12}O_6} + \underset{\text{fructose}}{C_6H_{12}O_6}$$

$$\underset{\text{Glucose or Fructose}}{C_6H_{12}O_6} \xrightarrow{\text{zymase}} 2C_2H_5OH + 2\ CO_2$$

Uses of Sucrose:

Sucrose is used as a food. It is an ingredient of jellies, jams, canned fruits, preserves, confections, condensed milk and other foods. It is used in the manufacture of sucrose octaacetate which is employed to denature alcohol, to render paper transparent, to stiffen textiles. and as an ingredient of non aqueous adhesives.

Lactose

(milk sugar $C_{12}H_{22}O_{11}$)

Occurrence:

Lactose occurs hi the milk of all animals.

Caw's milk contains 4 to 6%

Human milk contains 5 to 8%.

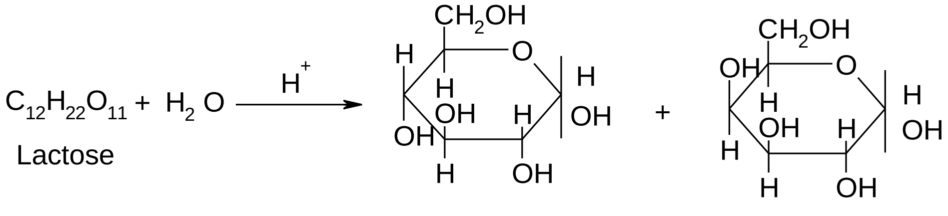
Structure Determination of Lactose:

The structure of lactose has been derived as follows:

1- Elemental analysis and molecular weight determination show that the molecular formula of lactose is $C_{12}H_{22}O_{11}$.

2- Lactose reacts with acetic anhydride in the presence of sodium acetate to form lactose octaacetate. This reaction indicates the presence of eight hydroxyl groups in a lactose molecule.

3- Hydrolysis of lactose with dilute acids yields an equimolecular mixture of D-glucose and D-galactose. This indicates that the lactose molecule is made up of one unit of each of these monosaccharides.

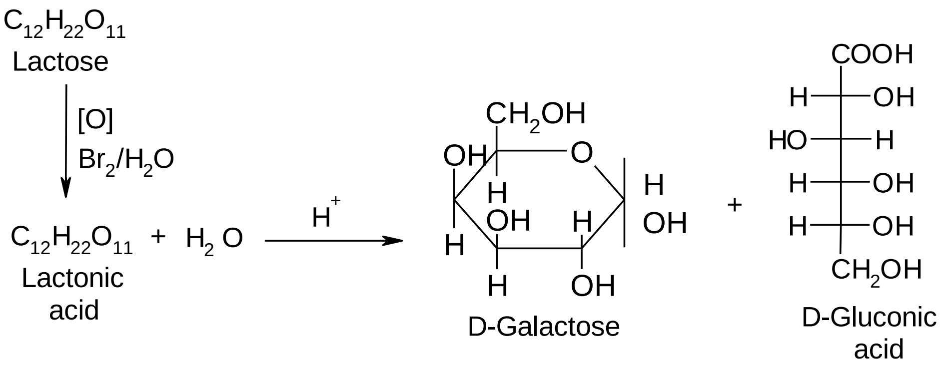


D-Glucose D-Galactose

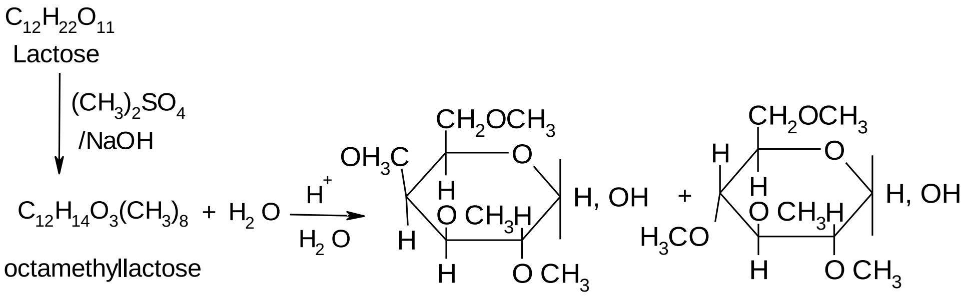
4- Lactose reduces Tollen's reagent and Fehling's solution, reacts with hydrogen cyanide, and forms an osazone. All these reactions indicate that one free hemiacetal group must be present and this is in equilibrium with some of the free aldehyde form.

5- Lactose can be isolated in two crystalline forms depending on how recrystallises ordinary lactose. If it is recrystallised from a concentrated aqueous solution at ordinary temperatures, the a-form of the sugar is obtained. Its melting point is 223°C and the specific rotation is +90°. However, if another portion of ordinary lactose is recrystalltsed from water at temperatures higher than 95°C, the 13-form isobtained: Its melting point is 252°C and the specific rotation is +35°, both α- and β-forms exhibit mutarotatlon until an equilibrium value of +55° is reached. This further confirms the presence of a free hemiacetal group in lactose.

6- Oxidation of lactose with bromine water gives lactonic acid, which on hydrolysis yields a mixture of D-galactose and D gluconic acid. This indicates that it is the glucose unit that contains the free hemiacetal-aldehyde group.



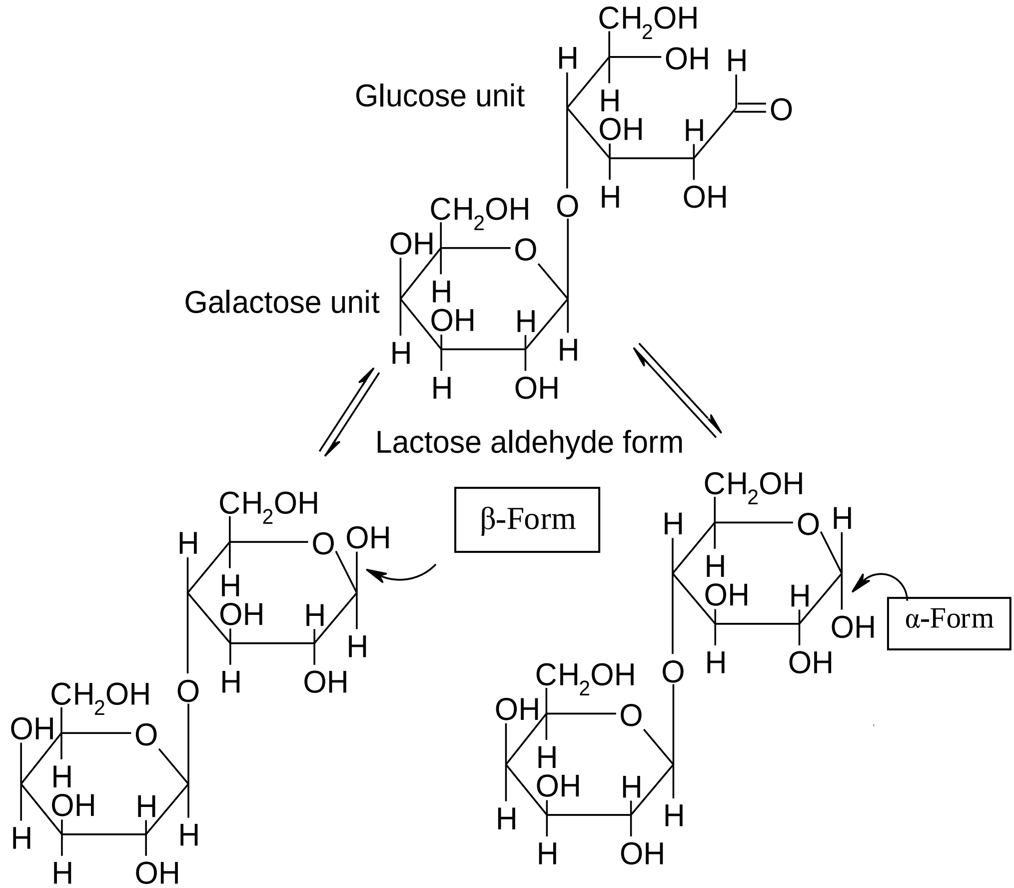
Lactose reacts with dimethyl sulphate in an alkaline solution to form octamethyl lactose which on hydrolysis yields a mixture of 2.3.4. 6-tetramethyl-D-galactose and 2,3,6-trimethyl-D-glacose. The formation of these compounds indicates that both units exist in 6-membered pyranose forms, and the glycosidic linkage involves the hydroxyl group at *C*-4 in glucose.



2,3,4,6-Tetramethyl-D-Galacose (α and β-forms)

1,2,3,6-Tetramethyl-D-Glucose (α-and β-forms)

Lactose is also hydrolyzed by emulsion, an enzyme that hydrolyses only β-glycosides rather than α-glycosides. This indicates that latose is a β-galactoside. The above evidence clearly shows that lactose has the following structure.



β-Lactose
m.p. 252 °C $[\alpha]_D$=+35°

α-Lactose
m.p. 223 °C $[\alpha]_D$=+90°

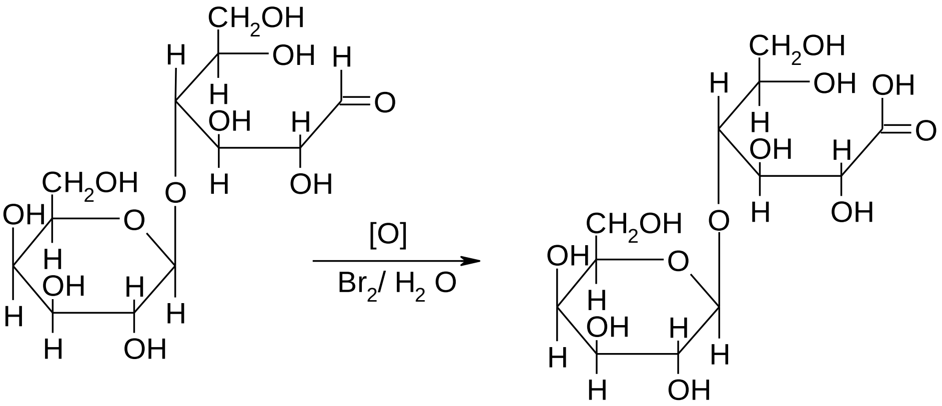
Properties of Lactose:

Physical properties Lactose (α-forms) is a colourless, odourless, crystalline solid. mp 223°C (with decomposition). It is soluble in water, but insoluble in alcohol and ether. An aqueous solution of lactose is dextrorotatory and exhibits mutarotation.

Chemical properties:

Lactose is a reducing sugar. that is, it reduces Fehling's solution and Tollen's reagent. Its reactivity is mainly due to presence of a free hemiacetal- aldedehyde group in the glucose unit of molecule. Some of the roost important reactions of lactose are given low.

(1) **Oxidation:** Oxidation of lactose with bromine water yields a onocarboxylic acid called lactonic acid or lactobionic acid.

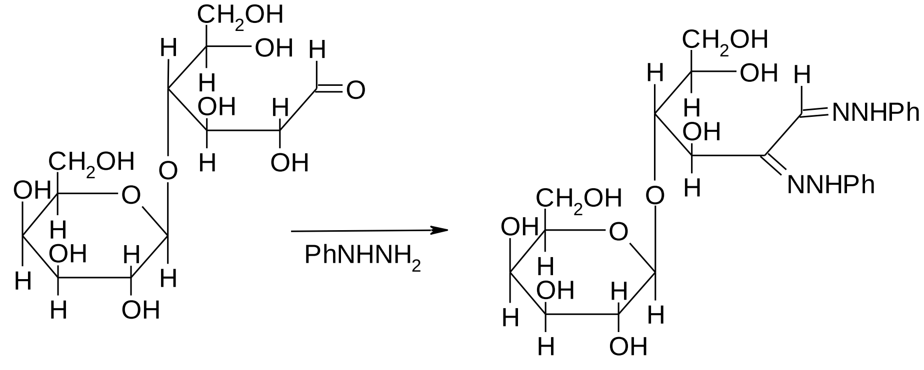


Lactose aldehyde form

Lactonic acid

(2) **Hydrolysis:** Hydrolysis of lactose with hot dilute acid or by the enzyme emulsin, yields a mixture of D-galactose and D-glucose.

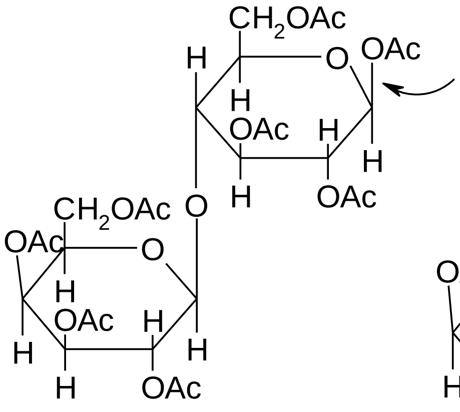
(3) **Reaction with phenyl hydrazine:** ***(Osazone formation)*** Lactose reacts with excess phenyl hydrazine in the presence of acetic acid to form lactosazone.

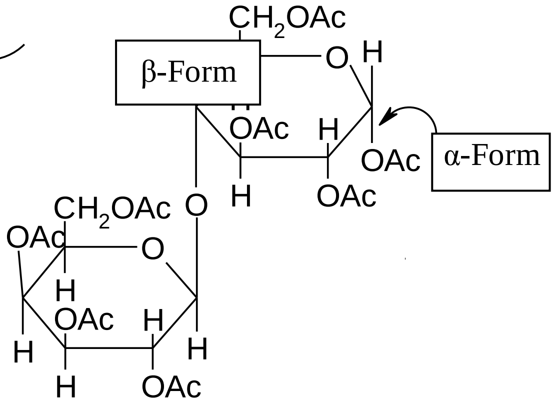


Lactose aldehyde form

Lactosazone
(Lactose phenylosazone)

(4) **Reaction with acetic anhydride: (Acetylation).** reacts with acetic anhydride in the presence of sodium acetate or zinc chloride to form two isomenic octaacetates. This is because in solution, both a- and b- forms of lactose are in equilibrium and each reacts separately to give a different compound.





Ac= CH_3CO

β-Lactose octaacetate
m.p. 90°C; $[\alpha]_D$=-4°

α-Lactose octaacetate
m.p. 152°C; $[\alpha]_D$=+54°

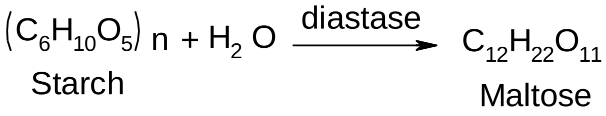
(5) **Fermentation.** Lactose is fermented by certain bacteria to gave actic acid which is responsible for the souring of milk. It is fermented by yeast.

(6) **Uses of Lactose.** lactose is used in baby foods and in pharmacy base for compressed tablets.

Maltose

(Malt Sugar $C_{12}H_{22}O_{11}$)

Maltose dose not occur in the free state in nature to an appreciable extent. It is obtained as a result of partial hydrolysis of starch by diastase (an enzyme present in malt).



$$\underset{\text{Starch}}{(C_6H_{10}O_5)n} + H_2O \xrightarrow{\text{diastase}} \underset{\text{Maltose}}{C_{12}H_{22}O_{11}}$$

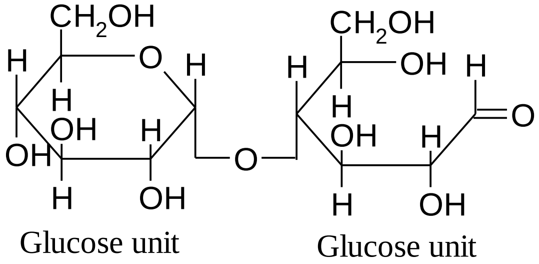
Structure Determination of Maltose:

(1) **Hydrolysis.** Hydrolysis of maltose with dilute acids yields only D-glucose. This indicates that the maltose molecule is made up of two glucose units.

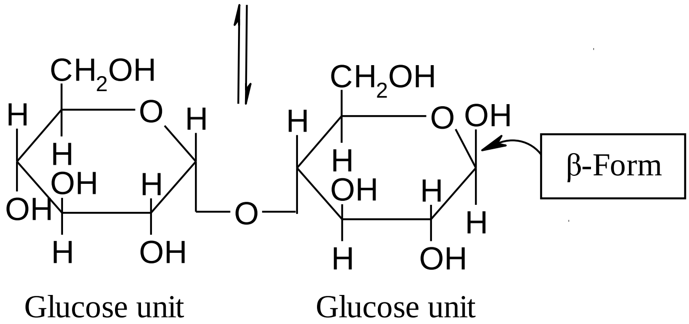
(2) **Other structural studies,** similar to those described under lactose. indicate that the two glucose units are joined by an α-glycosidic linkage between C-1 of one unit and C-4 of the other. Like, lactose, maltose exists in α-and β-forms, each of which exhibits, mutarotation. The values of specific rotation are:

α-Maltose = +168°, β-Maltose = +118°

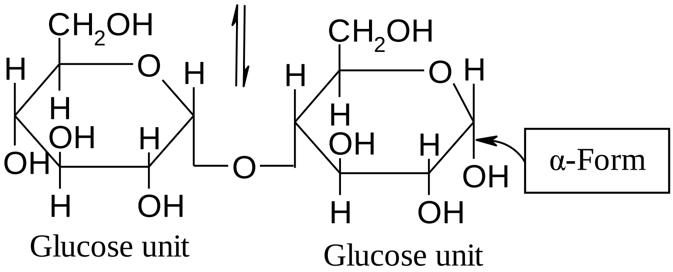
For equilibrium mixture = +136°



Maltose aldehyde form



β-Maltose, $[\alpha]_D = +118^o$



α-Maltose, $[\alpha]_D = +168^o$

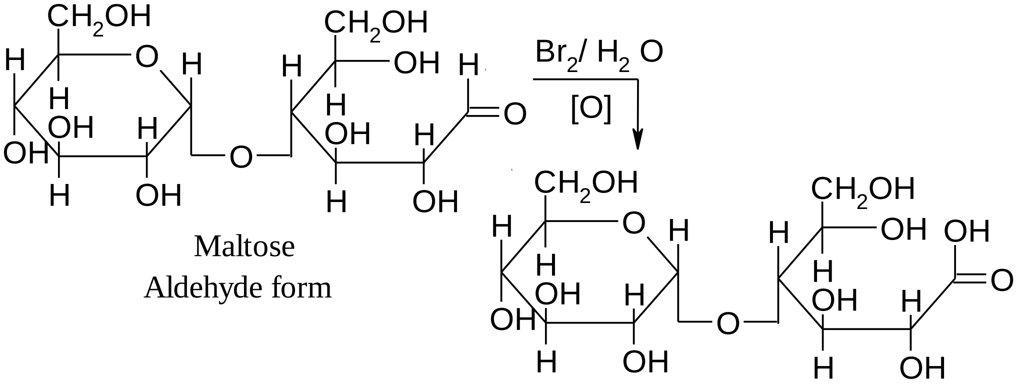
Properties of Maltose:

physical properties. Maltose (β-form) is a colourless. odourless. crystalline solid. mp l60-l65°C. It is soluble in water, but insoluble in alcohol or ether. An aqueous solution of maltose is dextrorotatory and exhibits mutarotation.

Chemical properties.

Maltose is a reducing sugar. Like lactose Its reactivity is also due mainly to the presence of a free hemiacetal group in one of the glucose units of its molecule.

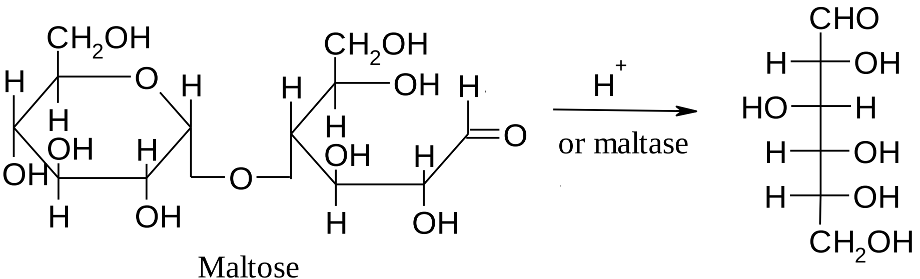
(1) **Oxidation.** Oxidation of maltose with bromine water yields a monocarboxylic acid called maltonic acid or maltobionic acid.



Maltose
Aldehyde form

Maltonic acid
or Maltobionic acid

(2) **Hydrolysis**. Hydrolysis of maltose with hot dilute acids or by enzyme maltase (specific for tri.-glycosides) yields only D-g]ucose.

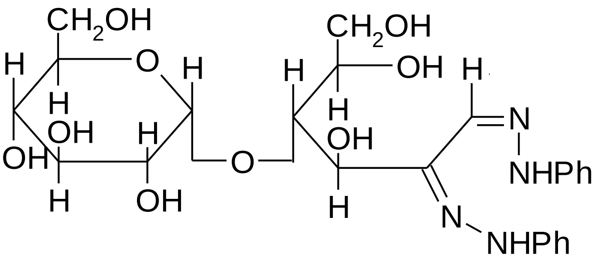


Maltose
aldehyde form

D-Glucose

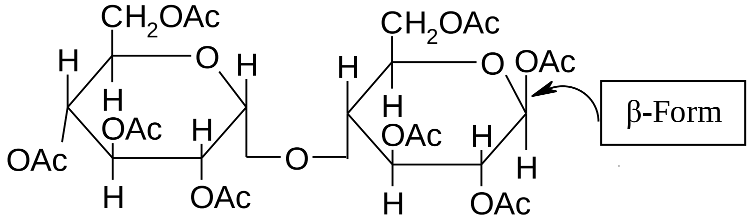
(3) **Reaction with phenyl hydrazine** (Osazone Formation):

Maltose reacts with excess phenyl hydrazine in the presence of acetic acid to form maltosazone. mp 206 °C.

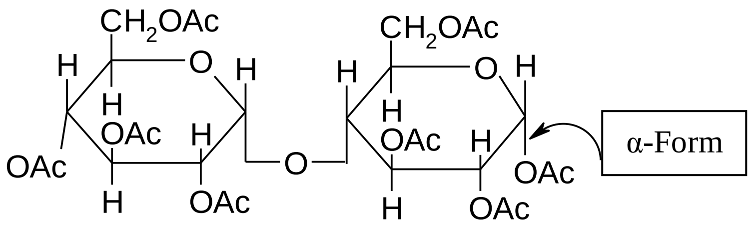


Maltosazone,. mp 206 °C

(4) **Reaction with Acetic anhydride (Acetylation):** Maltose reacts with acetic anhydride in the presence of zinc chloride or sodium acetate to give α-maltose octaacetate (mp 125 °C, [α]D=+123°) and β-maltose octaacetate (mp 160°C. [α]D=+63°).



β-Maltose octaacetate, mp 160°C. [α]D=+63°.



α-Maltose octaacetate, mp 125 °C, [α]D=+123°.

(5) **Fermentation.** An aqueous solution of maltose is fermented by yeast to give ethyl alcohol and carbon dioxide. The enzyme by enzyme *zymase* (also present in yeast to give ethyl alcohol and carbon dioxide.

Cellobiose

($C_{12}H_{22}O_{11}$)

Cellobiose is obtained by acetylating pure cellulose with acetic anhydride in the presence of sulphuric acid. The resulting cellobiose octaacetate is then hydrolyzed with potassium hydroxide or sodium methoxide to yield cellobiose.

Structure Determination of ceiloblose:

(1) Hydrolysis. Hydrolysis of cellobiose with dilute acids yields only D-glucose. This indicates that the cellobiose molecule is macic up of two glucose units.

(2) Other structural studies, similar to those described under lactose and maltose, indicate that thetwo glucose units are joined by an β-glycosidic linkage between C-1 of one unit and C-4 of the other. Like lactose and maltose ce]lobiose exists in α- and β-forms, each of which exhibits mutarotation. The values of specific rotation are:

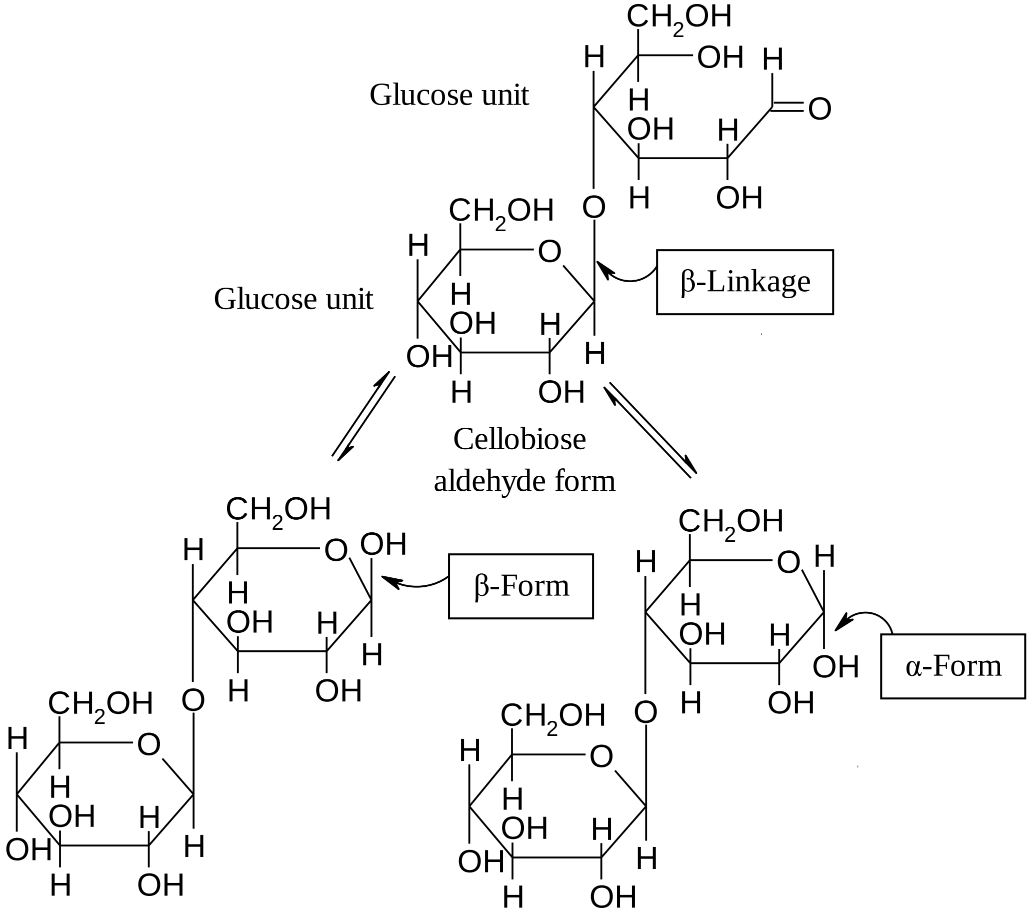
α-cellobiose = +72° β-cellobiose = +16°

For equilibrium mixture = +35°

Properties of Cellobiose:

Physical properties. Cellobiose (β-form) is a colourless, odourless crystalline solid, mp 225°C. It is soluble in water, but insoluble ether. An aqueous solution of cellobiose is dextrorotatory and exhibits mutarotation.

Chemical properties. Like maltose, cellobiose is a reducing sugar. Ake lactose and maltose its reactivity is also due mainly to the presence of a free hemiacetal group in one of the glucose units of its molecule. It undergoes all the reactions of maltose.



β-Cellobiose, $[\alpha]_D$= +16°

α-Cellobiose, $[\alpha]_D$= +72°

CHAPTER 4

Polysaccharides

Polysaccharides are compounds made up of many-hundreds or even thousands-monosaccharide units per molecule. As in disaccharides, these units are held together by glycoside. which can be broken by hydrolysis.

Polysaccharides are naturally occurring polymers, which can be considered as derived from aldoses or ketoses by polymerization with loss of water.

Polysaccharides are not reducing sugar and do not show mutarotation.

A polysaccharide derived from hexoses hzs 1.he general formula $\mathbf{(C_6H_{10}O_5)_n}$.

The most important polysaccharides are starch and cellulose. Both are produced in plants from carbon dioxide and water by thc p ncess of photosynthesis and both are made up of D-(+)-glucose units.

Starch makes up the reserve food supply of plants and occurs chiefly in seeds. It is more-soluble than cellulose, more essily hydrolyzed. and hence more readily digested. We use starch as a food: potatoes. corn, wheat, rice, cassava. etc.

Cellulose is the chief structural material of plants. giving the plants rigidity and form. It is probably the most widespread organic material known. We use cellulose for its structural properties: as wood for houses, as cotton or rayon for clothing, as paoer for communication and packaging.

So that both cellulose and starch are, of course, enormously important to us.

We need to know what the monosaccharide units are and how many there are in eaeh molecule; how they are joined to each other; and whether the huge molecules thus formed are straight-chained or branched, looped or coiled.

Starch:

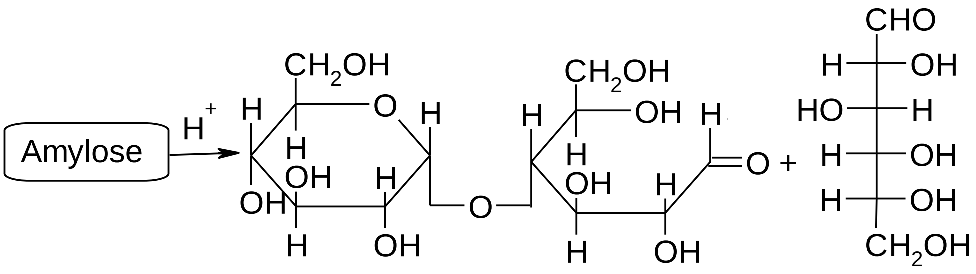
Starch occurs as granules whose size and shape are characteristic of the plant from which the starch is obtained. Starch contains about 20% of water-soluble fraction called amylose, and 80% of s waterinsoluhk Iractioa called amylopectin. Upon treatment with acid or under the influence of enzymes. the components of starch are hydrolyzed progressively to dextrin (a mixture of low molecular weight polysaccharides) (+)-maltose, and fina1ly D-(+)-glucose. Both amylose and amylopectin are made up of D-(+)-glucose units, but differ in molecular size and shape.

Structure of Amylose:

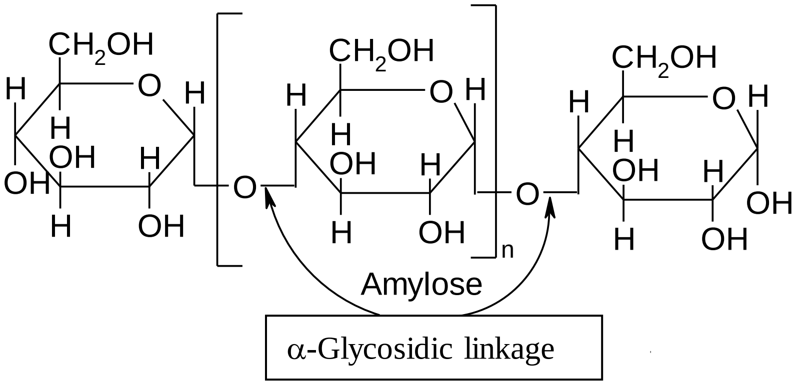
On hydrolysis of amylose yields:

1- (+)-Maltose (the only disaccharide), and

2- (+)-Glucose (the only monosaccharide).

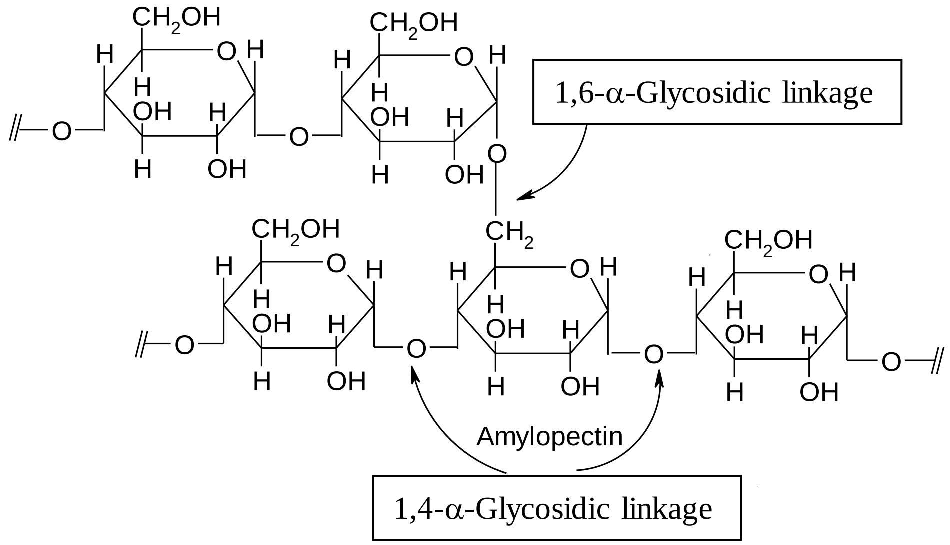


This indicates that arnylose is made up of chains of many D-(+)-glucose units, each unit joined by an alpha-glycoside linkage to C-4 of the next one.



Structure of Amylopectin:

On hydrolysis of amylopectin only (+)-maltose is obtained. Amylopectin is a branched chain polysaccharide. It is composed of chains of 24 to 30 D-glucose units joined by α-glycosidic linkages between C-1 of one glucose unit and C-4 of the next glucose unit. One end of each of these chains is joined through C-1 to a C-6 on the next chain.

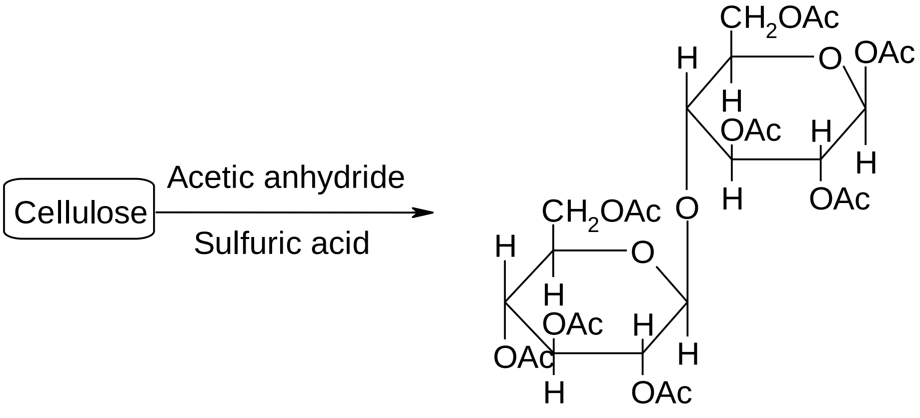


Cellulose

$(C_6H_{10}O_5)_n$

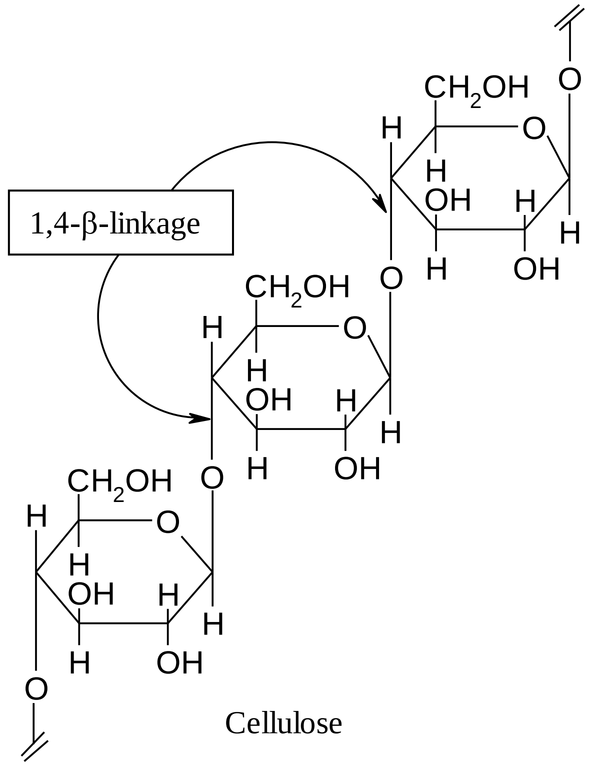
Structure of cellulose:

Cellulose differs from starch in the configuration of the glycoside linkage. Upon treatment with acetic anhydride and sulfuric acid, celiulcse yields octa-*O*-acetylcellobiose.



Octa-O-acetylcellobiose

This indicates that all glycosidic linkages in cellulose are beta linkage like the one in ceiiobiose. So that cellulose is made up of chains of D-glucose unit joined by a β-glycoside linkage to C-4 of the next.



Cellulose

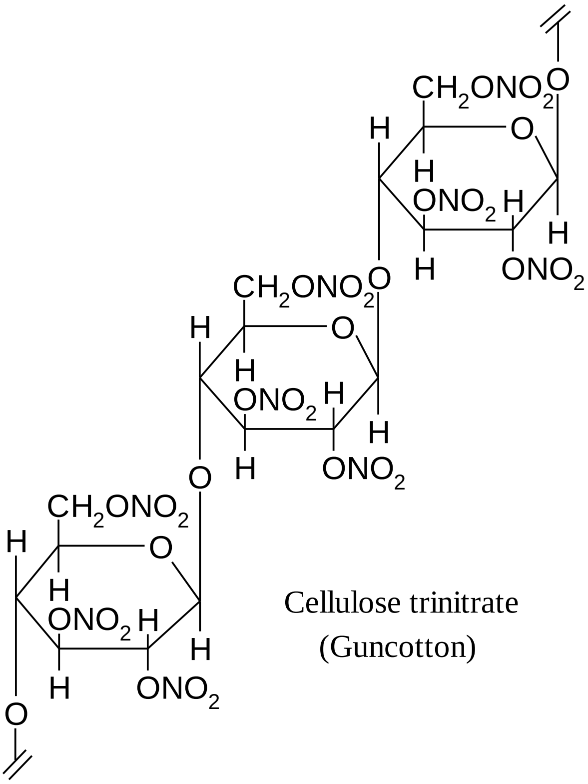
Reactions of Cellulose:

We have seen that the glycoside linkages of cellulose are broken by the action of acid, each cellulose molecule yielding many molecules of D-(+)-glucose. Now let us look briefly at reactions of cellulose in which the chain remains essentally intact. Each gluccse unit in cellulose contains three free -OH groups; these are the positions at which reaction occurs. These reactions of cellulose are of tremendous industrial importance.

1- Cellulose nitrate:

Like alcohol, cellulose forms esters. Treatment with a mixture of nitric acid and sulfuric acid converts cellulose into cellulose nitrate. The properties and uses of the product depend upon the extent of nitration.

Guncotton, in which of the -OH groups are replaced by -ONO_2 groups, that is. it contains three -ONO_2 groups per glucose unit, and is often tailed cdllulose trinitrate. Guncotton is looks something like ordinary cotton but is highly explosive. It is used in the manufacture of smokeless gunpowder.



Cellulose trinitrate
(Guncotton)

Pyroxylin. is less highly nitrated material containing between two and nitrate groups per glucose unit. It is used in the manufacture of placetics like celluloid and collodion. in photographic film, and in lacqures. It has the disadvantage of being flammable, and forms highly toxic nitrogen oxides upon burning.

2- Cellulose acetate.

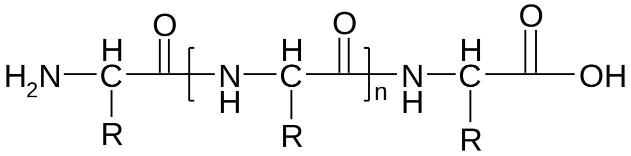
Cellulose is converted into cellulose triacetate by treatment with acetic acid, acetic anhydride. and a little sulfuric acid.

Amino Acids, Peptides, Proteins, and Nucleic Acids

5.1 INTRODUCTION:

Three types of organic polymers are essential to the life processes of every living cell. One of these, the polysaccharides, was discussed in Chapter 3. In this chapter we shall discuss the remaining two types. These are the nucleic acids and the proteins. The nucleic acids may be simply viewed as templates, from which the proteins are constructed. We shall return to them in Section 5.10. The proteins are polymers of amino acids. These ubiquitous macromolecules constitute nearly three-fourths of the dry weight of most animal tissues. and, indeed, they are involved in the structure and function of every living organism. Some proteins have a purely structural function (skin, hair, and muscle fiber, for example). Many others have a catalytic function (enzymes), which permits reactions to take place in living systems that would proceed so slowly in the absence of enzymes that the life could not be maintained. Other proteins have a regulatory function (hormones), and still others participate in the immunological defense mechanism of the organism (antibodies). In a human there are estimated to be about 5 million different proteins present, each of which is performing a function necessary for the well-being of the human. Other species of higher animals have similar numbers of proteins, and most of these proteins differ from one species to another. Some of them even differ from one individual to another.

Proteins are composed of α-amino acids joined together through amide linkages called peptide bonds:

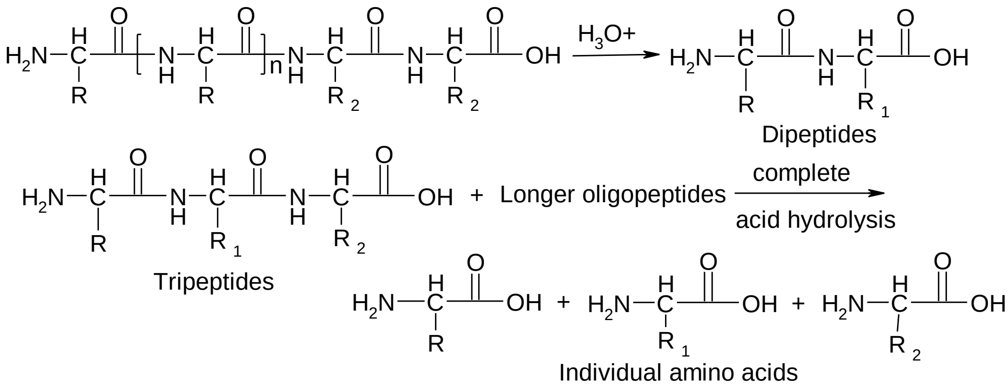


$$H_2N-\underset{R}{CH}-\overset{O}{\overset{\|}{C}}-\left[NH-\underset{R}{CH}-\overset{O}{\overset{\|}{C}}\right]_n-NH-\underset{R}{CH}-\overset{O}{\overset{\|}{C}}-OH$$

Partial hydrolysis of proteins by acids, bases, or enzymes yields smaller polyamides. Complete hydrolysis can be accomplished to give the individual amino acid components.

The molecular weights of proteins range from 6000 for insulin to 41.000,000 for the protein portion of tobacco mosaic virus. Polyamides of molecular weight less than 5000 are usually termed polypeptides. The large proteins are highly organized complexes of smaller subunits. In the case of tobacco mosaic virus protein, many identical subunits, each with a molecular weight of 17,500, are associated by noncovalent interactions. it seems likely that very few proteins of molecular weight greater than 100,000 will be found to consist of only one continuous polypeptide backbone.

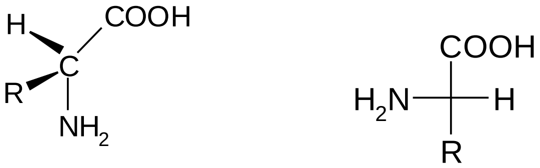
Natural polypeptides achieve their ability to carry out biological functions by virtue of their specific sequences of amino acids and their exact three-dimensional arrangement of these amino acids. The first step in the study of a protein is determination



of the amine acid science, called the primary structure. With the advent of more sophisticated techniques increasingly detailed aspects of protein structure are being investigated. These include the nature of the spatial relationship of near neighbors, sometimes called the secondary nurture; the gross folding of one chain, tertiary structure and the spinal relationship of one polypeptide chain to another quaternary structure.

5.2 NATURALLY OCCURRING AMINO ACIDS:

From all natural sources over 100 amino acids have been isolated and identified to date. The great majority of amino acids have the amino group attached to the cabon α to the carboxylic acid. With very few exceptions, the a carbon also bears a hydrogen atom The fourth bond of the a carbon is joined to a group which has over 100 variations. Thus, most of the naturally occurring amino acids differ only in the structure of the organic residue attached to the α-carbon. An interesting and important fact is that almost all amino acids isolated from proteins have the L-configuration at the α-carbon, although some amino acids isolated from microbiological sources are the mirror image isomers i.e., in the D-configuration (see p.5).



L-Configuration

Of the amino acids isolated from living material, only about 20 are naturally occurring components of proteins. The remainder are found as intermediates or end products of metabolism. All living species are able to synthesize amino acids. Many species, however, are deficient in their ability to synthesize within their own metabolic system all the amino acids necessary for the life of their species. The eight amino acids with this special significance for the human species are called essential amino acids (Table 5.1). They are essential not because they are the only amino acids required for human functioning but because they are essential in the diet of the human species

Table 5.1: Essential Amino Acids:

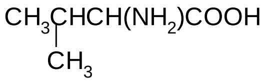
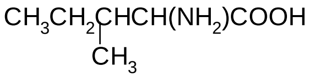
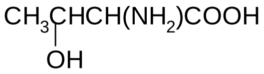
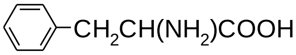
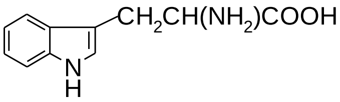
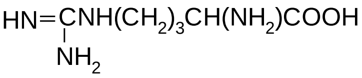
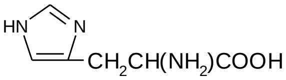
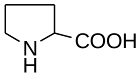
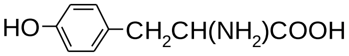
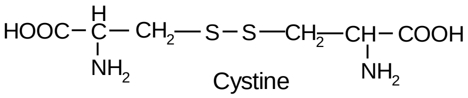
Structure	Name	Abbreviation
 $CH_3\underset{CH_3}{\underset{\vert}{C}}HCH(NH_2)COOH$	L-(+)- Valine	Val
$(CH_3)_2CHCH_2CH(NH_2)COOH$	L-(-)- Leucine	Leu
 $CH_3CH_2\underset{CH_3}{\underset{\vert}{C}}HCH(NH_2)COOH$	L-(+)- Isoleucine	Ile
 $CH_3\underset{OH}{\underset{\vert}{C}}HCH(NH_2)COOH$	L-(-)- Threonine	Thr
$CH_3S(CH_2)_2CH(NH_2)COOH$	L-(-)- Methionine	Met
	L-(-)- Phenlyalanine	Phe
	L-(-)- Tryptophan	Trp
$NH_2(CH_2)_4CH(NH_2)COOH$	L-(+)- Lysline	Lys

Table 5.2: Other Common Amino Acids[a]:

Structure	Name	Abbreviation
$CH_3CH(NH_2)COOH$	alanine	Ala
 $HN{=}\underset{NH_2}{\underset{\mid}{C}}NH(CH_2)_3CH(NH_2)COOH$	Arginine	Arg
$HOOCCH_2CH(NH_2)COOH$	Aspartic acid	Asp
$HSCH_2CH(NH_2)COOH$	Cysteine	CySH
$HOOC(CH_2)_2CH(NH_2)COOH$	Glutamic acid	Glu
H_2NCH_2COOH	Glycine	Gly
	Histidine	His
	Proline	Pro
$HOCH_2CH(NH_2)COOH$	Serine	Ser
	Tyrosine	Tyr
$NH_2COCH_2CH(NH_2)COOH$	Asparagine	Asn
$NH_2CO(CH_2)_2CH(NH_2)COOH$	Glutamine	Gln

[a] One of the amino acids commonly found in protein hydrolysates has the name cystine, and has the following structure:


$$HOOC-\underset{NH_2}{\underset{\mid}{\overset{H}{C}}}-CH_2-S-S-CH_2-\underset{NH_2}{\underset{\mid}{CH}}-COOH$$

Cystine (CyS-SCy)

It is clearly a dimer of cysteine. Where the thiol groups of the latter have been oxidized to form a disulfide linkage. The dimer actually results because of two monomers at widely spaced intervals in the polypeptide are joined together by a disulfide bridge. Thus the basic amino acid is cysteine: consequently, the dimer is not included here.

since our cells cannot synthesize them. The other 12 amino acids (Table 5.2) found in the biochemicals derived from human beings can be synthesized in individual cells from simpler starting materials that contain carbon, hydrogen. oxygen, and nitrogen.

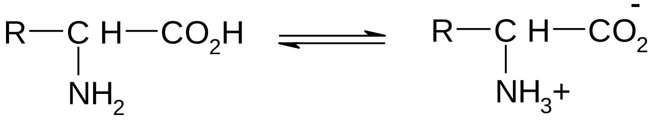
EXERCISE 5.1

Glycylglycine is a dipeptide composed of two molecules of glycine. Write its structure.

The various coexisting species have different sets of amino acids which they require-but are unable to synthesize. However, all creatures contain within themselves all amino acids. so that any animal may normally acquire needed amino acids from others. Man may acquire his daily minimum of essential amino acids by eating such things as filet mignon from the cow. Dover sole from the fish, and eggs Benedict from the chicken. Vegetarians can survive e because there are sources of plant protein that contain all the essential amino acids. Rice has a high protein content, as do legumes (peas, beans), Corn, wheat, and rye are other grains that have a significant quantity of plant protein which includes the essential amino acids. Proteins eaten by humans (and other animals) are completely hydrolyzed to amino acids, and these are then used as building blocks to construct the proteins of the individual.

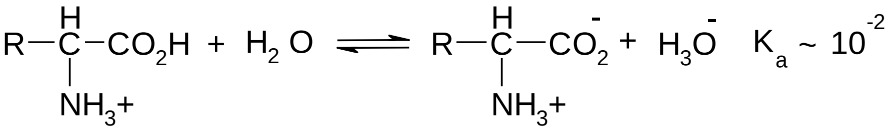
CHEMICAL AND PHYSICAL PROPERTIES OF AMINO ACIDS:

Amino acids are high-melting solids which, because of their two polar groups, would be expected to be insoluble in organic solvents but soluble in water. Since the carboxylic acid function is acidic and the amino group basic, the amino acids actually exist as dipolar ions (zwitterions), rather than in the un-ionized forms shown in the previous section.

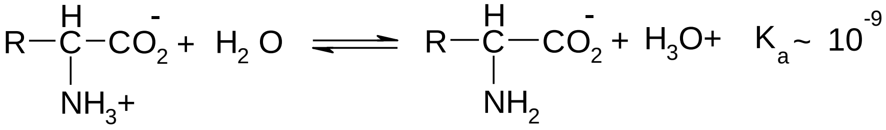


$$R-\underset{NH_2}{C}H-CO_2H \rightleftharpoons R-\underset{NH_3+}{C}H-CO_2^-$$

Amino acids with no ionizable side chains have two ionization constants with pK_a's of about 2 and 9.

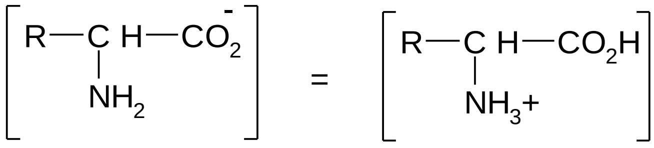


$$R-\overset{H}{\underset{NH_3+}{C}}-CO_2H + H_2O \rightleftharpoons R-\overset{H}{\underset{NH_3+}{C}}-CO_2^- + H_3O^- \quad K_a \sim 10^{-2}$$



$$R-\overset{H}{\underset{NH_3+}{C}}-CO_2^- + H_2O \rightleftharpoons R-\overset{H}{\underset{NH_2}{C}}-CO_2^- + H_3O+ \quad K_a \sim 10^{-9}$$

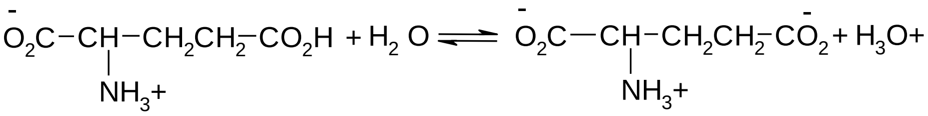
if an electrical potential is placed across two electrodes in a solution of an amino acid, the amino acid will migrate to the anode or the cathode, depending upon the pH. At one pH, called the isoelectric point, there is no *net* migration of he amino acid because the concentration' anion is the same as the concentration of the cation:



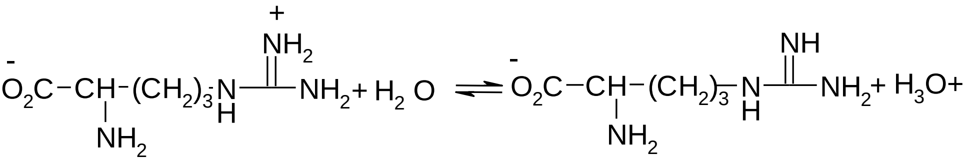
$$\left[R-\underset{NH_2}{C}H-CO_2^-\right] = \left[R-\underset{NH_3+}{C}H-CO_2H\right]$$

The isoelectric point is an individual characteristic of an amino acid; for example, it is pH 6.0 for glycine, pH 5.5 for phenylalanine, pH 11.2 for arginine, and pH 3.2 for

The amino acids with functional groups that are ionizable have ionization constants characteristic of those functional groups. For example. the side chain of glutamic acid has a pK_a of 4.3 and that of arginine has a pK_a of 13.2.



$${}^{-}O_2C-CH(NH_3^+)-CH_2CH_2-CO_2H + H_2O \rightleftharpoons {}^{-}O_2C-CH(NH_3^+)-CH_2CH_2-CO_2^{-} + H_3O^+$$



$${}^{-}O_2C-CH(NH_2)-(CH_2)_3-NH-C(=\overset{+}{N}H_2)-NH_2 + H_2O \rightleftharpoons {}^{-}O_2C-CH(NH_2)-(CH_2)_3-NH-C(=NH)-NH_2 + H_3O^+$$

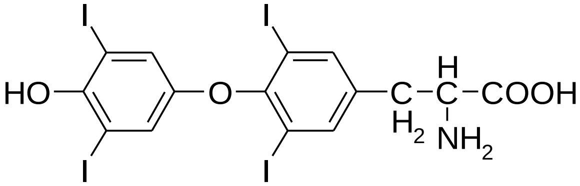
Saturated carboxylic acids absorb in the infrared at 1725-1700 cm^{-1}. Amino acids, however, absorb at 1400 and 1600 cm^{-1}, absorption frequencies characteristic of the carboxylate ion. When a neutral amino acid solution is made acidic, the 1720 cm^{-1} carbonyl stretching frequency of the carboxylic acid appears. This is consistent with the proposed dipolar structures.

Amino acids undergo most of the reactions characteristic of carboxylic acids and aliphatic amines. Amino acid esters are relatively unstable, and they are usually obtained as hydrochloride salts. The amino group reacts with nitrous acid, as do other aliphatic amino groups. The accompanying evolution of nitrogen is, in fact, often used to analyze for free amino groups in amino acids and their derivatives.

5.4 Primary structure and biological activity of polyamides

The number of possible random combinations of the 20 or so amino acids found in hormones, enzymes, and all other proteins is almost infinite. However, biological activity is not achieved by randomness but by a very precise ordering of the combined amino acids. Many scientists are now studying the primary structure-the amino acid sequence-of the polypeptides of biological importance for human beings. In this section we shall give some examples of the relationship between sequence and activity for polypeptides of varying chain length.

Even a single amino acid may exhibit potent biological activity: for example, thyroxine is a hormone that is an active principle for those animals which have a thyroid gland.



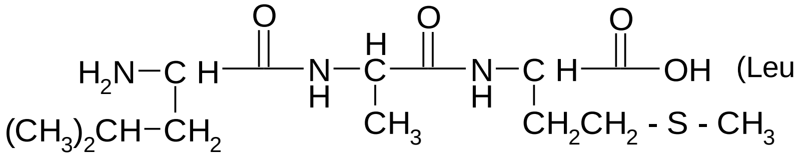
Thyroxine

This re1atively simple molecule exerts a profound effect upon the metabolism of almost every cell in the body.

Nomenclature:

In naming peptides, amino or N-terminal end refers to the end with the free amino group and carboxy or C-terminal end refers to the end with the free carboxyl group.

By convention, the N-terminal end is written to the left and the C-terminal end to the right. The amino acids are then named left to right, replacing ine with *yl,* except for the C-terminal amino acid, as in the example shown.



$$H_2N{-}\underset{\underset{CH_2{-}CH(CH_3)_2}{|}}{CH}{-}\overset{O}{\overset{\|}{C}}{-}\underset{H}{N}{-}\underset{\underset{CH_3}{|}}{\overset{H}{C}}{-}\overset{O}{\overset{\|}{C}}{-}\underset{H}{N}{-}\underset{\underset{CH_2CH_2{-}S{-}CH_3}{|}}{CH}{-}\overset{O}{\overset{\|}{C}}{-}OH$$

(Leu. Ala. Met.)

Leucylalanylmethionine

Sometimes a polypeptide will be found as a simple derivative, and this can also be indicated in the structure. For example, the formula Leu. Ala. Met(NH_2) would represent the above polypeptide, in which the free carboxyl group had been converted into an amide. Similarly, the formula (Ac) Leu. Ala. Met would indicate the acetyl derivative of the amino end of the molecule.

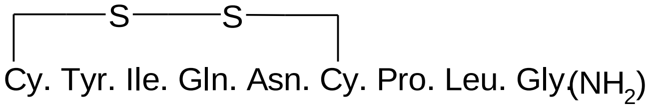
EXERCISE 5.2:

Write out the full structures that correspond to the following formulas:

(a) Pro.Val. Glu(NH_2)
(b) Arg. Gly. Phe. Ser
(c) Ser. Tyr. Arg. Asp

Oxytocin and bradvkinin are both polypeptides composed of nine amino acids, but different acids and in different sequences. Their resulting biological functions are strikingly different.

Oxytocin is one of the most physiologically active compounds known.



$$\overbrace{\text{Cy. Tyr. Ile. Gln. Asn. Cy.}}^{\text{S}-\text{S}}\text{ Pro. Leu. Gly.}(NH_2)$$

Oxytocin

Oxytocin is responsible for uterine contractions during childbirth and acts upon lactating mammary glands to stimulate the ejection of milk. It is interesting that only the female of the species produces this relatively simple polypeptide. It is even more interesting that this specific chemical is equally effective in causing a chicken to lay an egg or a cow to give down her milk to a farmer on a cold morning, or in causing a pregnant female human to give birth to a child. Chemically the oxytocin from chickens, cows, and hogs is identical Oxytocin obtained from chickens is used clinically for the induction of labor.

Bradykinin is also a very active substance. It is released by blood plasma globulins in response to a wasp sting, and is a potent pain-causing agent.

Arg. Pro. Pro. Gly. Phe. Ser. Pro. Phe. Arg

Bradykinin

Obviously any change in the amino acid sequence of bradykinin or oxytocin, whether it was substitution, deletion, or addition, would result in a profound modification of the biological activity.

Larger natural polypeptides may perform the same function in different species without being identical in primary structure. For example, insulin, a hormone that controls carbohydrate metabolism, differs in the arrangement of 4 of-its 51 amino acids -in each of many different species. However, bovine insulin may be used to compensate for the insulin deficiency of human beings suffering from diabetes.

Proteins of more than 100 amino acids are transferred between species with difficulty because of immunological problems (discussed in Section 5.5). For example, enzymes that perform identical functions in different species cannot simply be transferred between these species. There are differences in the amino acid sequences which result in a recognizably changed three-dimensional structure (Section 5.5). Very large proteins such as those in hair. muscle, and skin are present in such a complex arrangement that, except: for identical twins, no individual of a species will accept and use the hair, muscle, or skin of another.

A more dramatic example of the importance of the amino acid sequence is provided by the polypeptide globin. the protein moiety of hemoglobin (see else where for a discussion of the structure and function of hemoglobin). Globin has 146 amino acid residues jr. a very specific order, and for the human being. a substitution. a deletion, or an addition of even one amino acid to the number or the sequence may result in serious disease or possibly death. The disease called sickle-cell anemia is a molecular disease suffered by people whose globin differs from normal only in that the sixth amino acid in the series of 146 is valine rather than glutamic acid. Victims of this disease, which is hereditary, are unable to utilize oxygen at the normal rate and therefore must avoid high. oxygen-rare altitudes and any exercise that is physically taxing on their red blood cells. The formula below shows the substitution that distinguishes the globin of a normal human from one who has sickle-cell anemia.

Normal Globin:

VaL. His. Leu. Thr. Pro. Glu. Glu. Lys.

Sickle-Cell Globin:

Val. His. Leu. Thr. Pro. Val. Glu. Lys

Under conditions of oxygen deficiency or abnormal physical activity requiring rapid oxygen metabolism, the red blood cells of people with this disease take the shape of a "sickle" or quarter-moon. and they completely cease to function unless oxygen is administered effectively and immediately.

Enzymes are even more complex polypeptide material. Every living cell contains thousands of enzymes, each of which is responsible for catalyzing a single specific chemical reaction. The complete chemical structure has been elucidated for several enzymes. Recently, one enzyme, ribonuclease. which contains 124 amino acids in a specific sequence, has been synthesized (Section 5.9). Chymouypsin has been purified to the point where it has crystallized and the precise number of amino acid residues (246) and their precise sequence have been determined. X-ray structure determination has established its three-dimensional structure at a resolution of better than 3 A. At this resolution individual atoms cannot .be discerned, but the overall shape, and the twists and turns of various segments, can be seen (Section 5.7).

5.5 IMMUNOCHEMISTRY:

The human body reacts immediately whenever it is subjected to the introduction of any foreign substance, including larger polypeptides. It examines alien material very carefully for unfamiliar chemical structural characteristics and, should it recognize any, it causes the production of the specific polypeptide which is able to specifically bind to the foreign matter, so precipitating it from the surrounding medium.

In immunological terms. any such foreign substance is an antigen. The main encasing protein is gamma globulin of the blood of the host and is called an antibody. By a mechanism not completely understood, this initial antigen-antibody reaction elicits the production of greater amounts of the specific antibody required. Excess antibodies
remain in the bloodstream, where they afford the body a specific type of immunity for as long as they remain in excess. Should the offending antigen return while they are present, it will immediately be precipitated and the body will suffer no harmful effects. Unfortunately, immunity is not necessarily a permanent condition; its duration may range from several hours to a lifetime, depending on the nature of the antigen. Thus immunity from smallpox is normally long term, whereas that from the common cold lasts only a matter of days or weeks.

It is important to note that immunity is highly specific for a given antigen. Each new foreign cell that invades the body elicits a new supply of specific antibodies, which are stored in the blood as gamma globulins. Quite obviously, the gamma globulin blood fractions of individuals will vary to the extent to which they have been exposed to different foreign cells.

In this antigen-antibody phenomenon, we have a more precise explanation for the failure of skin and kidney and heart transplants among individual members of the same species. The body of the receptor looks upon the skin or kidney cells of the donor as foreign matter and immediately sets up an antibody-type rejection mechanism. Only between identical siblings, whose bodies are made up of protein materials that have identical structures and are therefore not "foreign," can such transplants be successful. One approach used today is to attempt to suppress production of gamma globulins. Drugs are known that accomplish this purpose. The patient is then able to receive a foreign substance, for example a heart, but has lost immunity and is therefore susceptible to many diseases from which he had previously been cured. Further. globulin synthesized in the body after the heart transplant still does not accept the heart as nonforeign material and finally rejects it by the antibody-antigen reaction.

5.6 DETERMINATION OF THE STRUCTURES OF PEPTIDES:

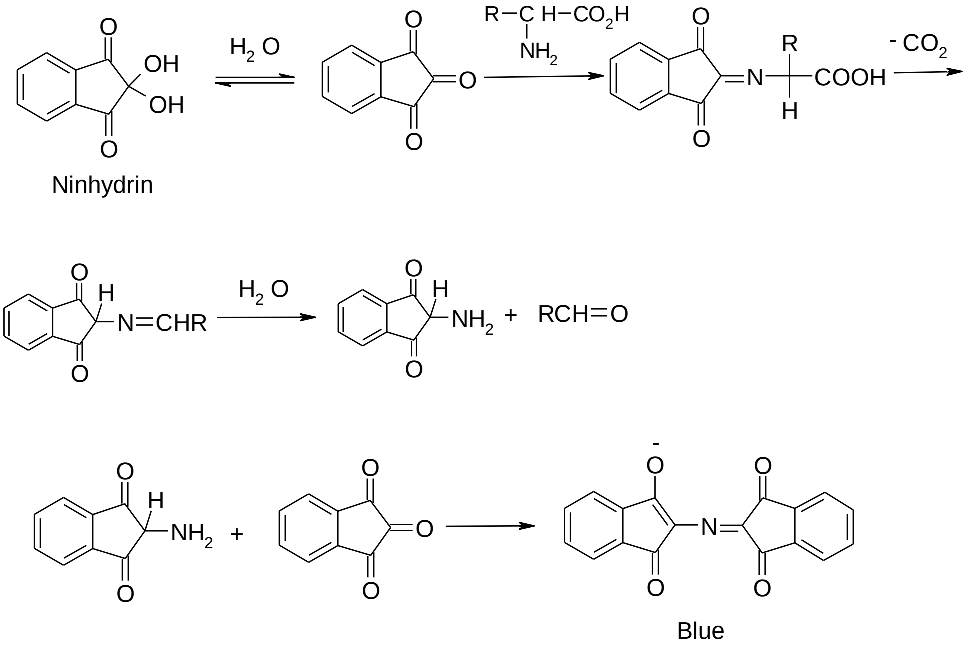
The first step in determining the primary structure of a peptide is to hydrolyze it to its individual amino acids and to assess which ones are present and how many of each. Although these analyses are somewhat complicated, the details for such procedures have been very thoroughly worked out, and the analyses themselves are now highly automated.

OPTIONAL MATERIAL:

Analysis of Amino Acids. Peptides are hydrolyzed in 6 N HC1 at 105 °C; base cannot be used because it racemizes the a carbons. Tryptophan. is sensitive to acid and is partially destroyed in the hydrolysis, which can be corrected for in quantitative studies. Instead of glutamine and asparagine, the corresponding acids and ammonia are isolated.

The mixture of amino acids obtained upon hydrolysis can be separated and analyzed by use of an “amino acid analyzer.” In this automatic equipment, aliquots of the mixture of amino acids are placed on separate columns of a sulfonic acid ion-exchange resin. One column is held at pH 5.3 and used for basic amino acids, ammonia, and tryptophan. A second is held at pH 3.25 for the other amino acids while 0 to 250 ml of elutant pass through the column: the pH is then increased to 4.25 while 250-500 ml of elutant passes through. Sodium citrate buffer solutions elute the acids, and the eluted solutions are mixed with ninhydrin and heated.

The reagent ninhydrin produces a blue color with primary α-amino acids by the following series of reactions:

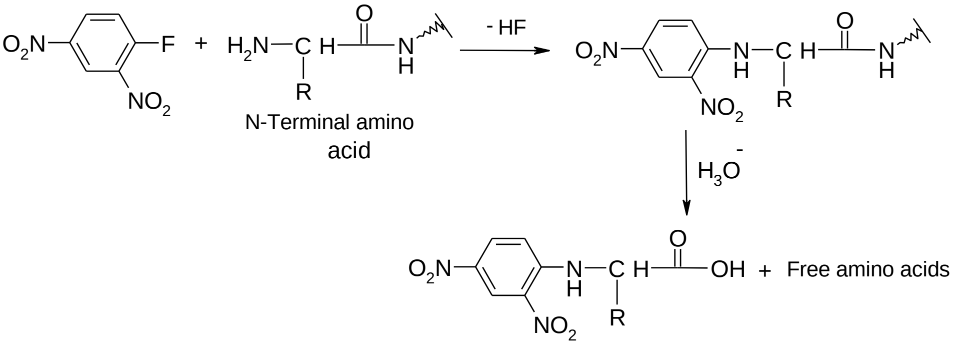


A spectrophotometer measures the optical absorption of the products of the ninhydrin reaction. and the recorder continuously plots the milliliters of eluate versus intensity of the ninhydrin color. The position of the absorption peak (in ml), which depends on the volume of buffer needed to elute a particular amino acid, is characteristic for each amino acid; the quantity of the acids is obtained from the areas under their peaks.

Gas chromatography is also useful for analyzing mixtures of small amounts of amino acids. Amino acids are too nonvolatile to be studied directly; their esters, for example trimethylsilyl esters, have sufficient volatility, however.

After determining the identities of the amino acids present, and their ratios, the next big problem is to determine the sequence of the amino acids in a peptide.

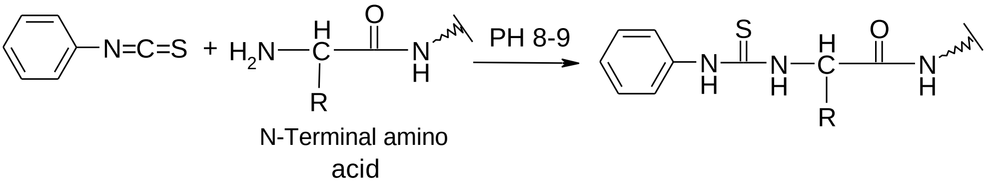
One common technique used to determine the N-terminal amino acid is to allow the amino group to react with 1-fiuoro-2,4-dinitrobenzene (nucleophilic aromatic substitution) and then to hydrolyze the peptide:



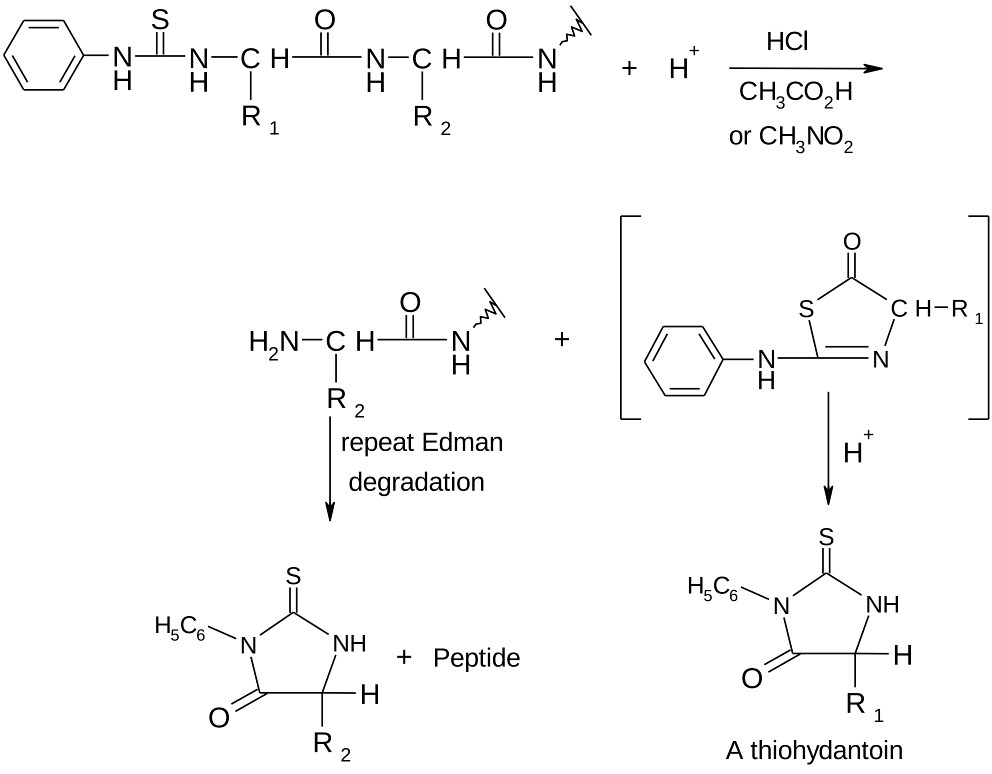
The N-(2,4-dinitrophenyl) derivative is isolated and identified.

An alternative procedure for determination of the N-terminal amino acid, which does not hydrolyze the peptide, is called Edman degradation.

An N-phenylthiocarbamyl derivative is prepared with phenyl isothiocyanate.

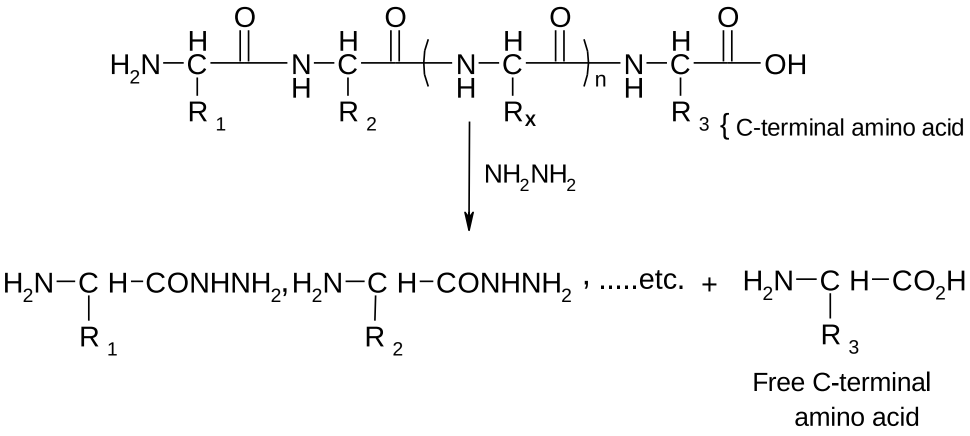


When this derivative is treated with hydrogen chloride in nitromethane or acetic acid, a thiohydantoin forms without destroying the remaining linkages.



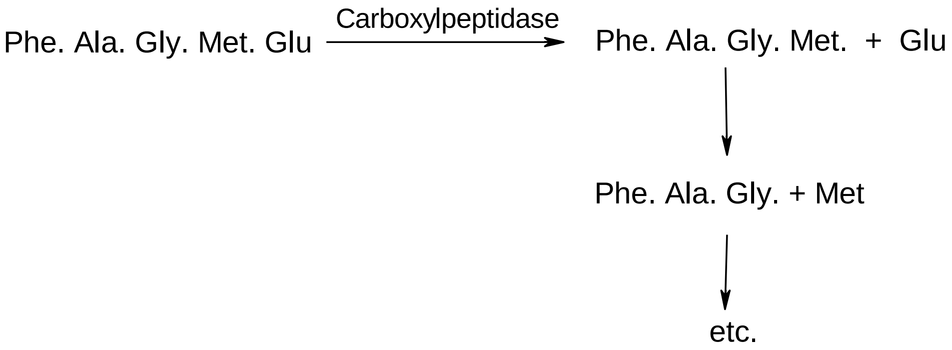
The thiohydantoin is identified and the nature of its R group characterizes the N-terminal amino acid. This procedure can be continued on the remaining chain; there is now available commercially an automated procedure which the manufacturer claims will perform 30 successive Edman degradations on a purified protein.

To determine the C-terminal end, the peptide can be heated with anhydrous hydrazine to convert the amide linkages iii the chain into hydrazides:



The C-terminal amino acid is identified as the free acid, wheieas the others in the chain are obtained as hydrazides.

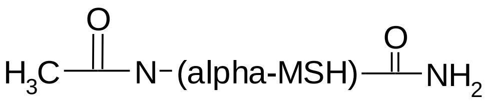
Certain carboxypeptidases, which are enzymes obtained from the pancreas, attack C-terminal peptide bonds and free the C-terminal amino acid. However, they continue acting on the remaining peptide and systematically liberate the newly formed C-terminal acids; for example,



The action of the enzyme cannot be stopped after it has released Gln. The enzyme proceeds to attack the tetrapeptide and release Met, and then to attack the tripeptide. and so on. Thus, the sequence of only a limited number of units can be obtained before the situation gets too mixed up to sort out The identities of the amino acids that are freed are determined as a function of incubation time, and as much sequencing is done as is possible from the data. If the C-terminal end is in the form of the amide, carboxy-peptidase does not act to free it.

To illustrate some of the techniques employed in amino acid sequence determination. we will follow an example of such a determination. Using an amino acid analyzer and a crude molecular-weight determination, α-MSH, a melanocyte-stimulating hormone from pituitary glands, was found to have the molecular formula: (Arg, Glu, Gly, His, Lys, Met, Phe, Pro, Ser_2, Trp, Tyr, Val, NH3) The commas between the abbreviations indicate that the sequence is unknown or unspecified. The acids were present in equimolar quantities, except for serine.

No N-terminal amino acid was found in the Edman degradation or dinitrophenylation reaction and no C-terminal amino acid was liberated by carboxypeptidase. As we will ace later, this is due to tie-up of the N-terminal end as the N-acetyl derivative and the C-terminal as the amide:



$$H_3C-\overset{O}{\overset{\|}{C}}-N-(\text{alpha-MSH})-\overset{O}{\overset{\|}{C}}-NH_2$$

It could also have meant that the peptide chain was cyclic. Thus the amino group and the carboxyl group must be free if these procedures are to work.

Chymotrjpsin, another pancreatic enzyme. preferentially attacks peptide bonds whose carbonyl function is furnished by one of the aromatic amino acids tyrosine, tryptophan, and phenylalanine. although it will also catalyze the hydrolysis of bonds with leucine, methionine, asparagine, and glutamine. After chymotryptic hydrolysis of α-MSH. three peptide fragments were isolated:

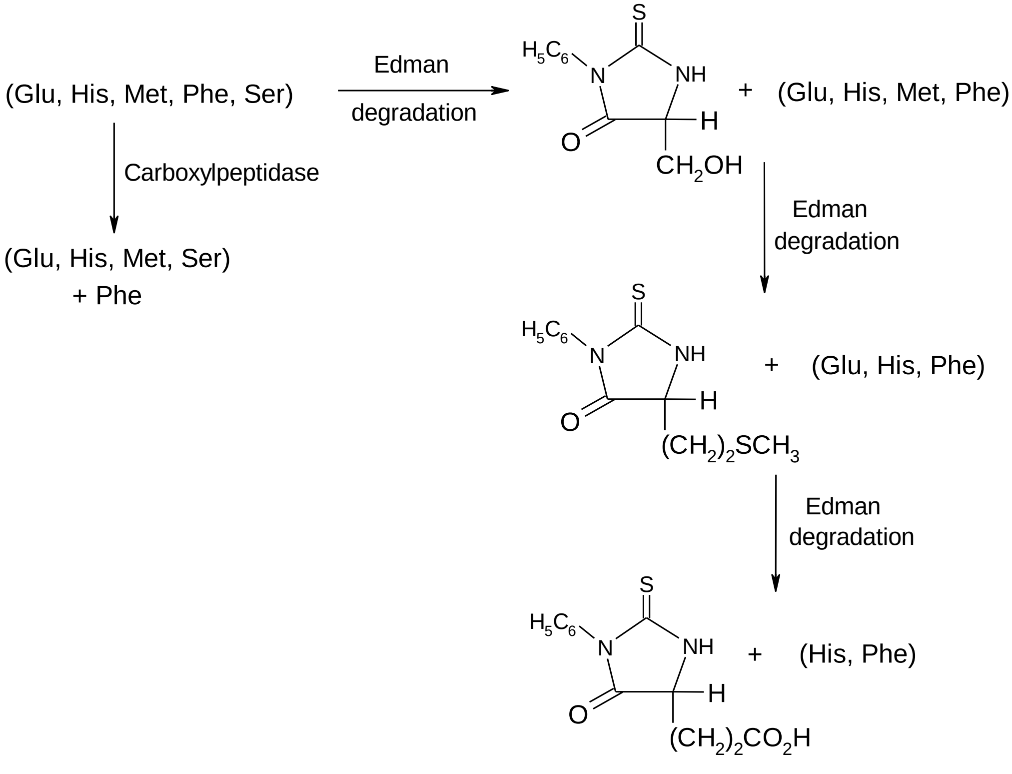
$$(\text{Arg, Gly, Gly, His, Lys, Met, Phe, Pro, Ser}_2\text{, Tyr, Val, NH}_3) \xrightarrow{\text{chymotrypsin}}$$

$$(\text{Ser, Tyr}) + (\text{Glu, His, Met, Phe, Ser}) + (\text{Arg, Gly, Lys, Pro, Trp, val})$$

Analysis of the serine in the dipeptide (Ser.Tyr) showed it to be in the form of the N-acetyl derivative. Thus Ser. Tyr is the N-terminal end of a α-MSH, and a partial structure of α-MSH is

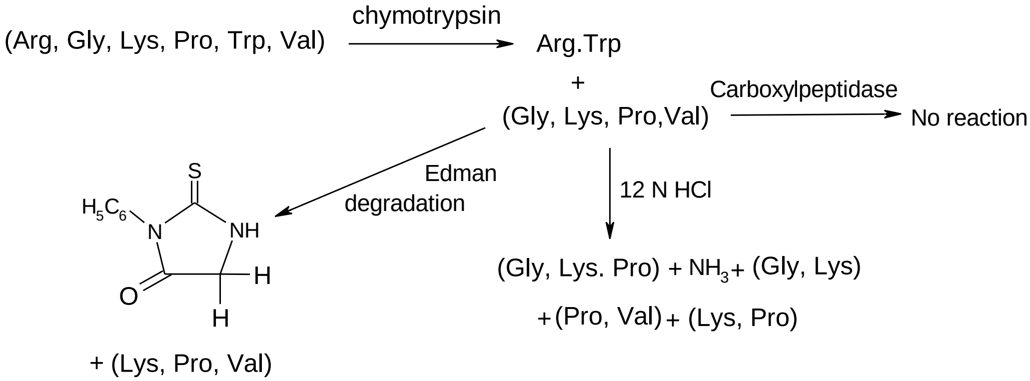
$$\text{(Ac) Ser. Tyr.} \left\{ \begin{array}{l} \text{(Glu, His, Met, Phe, Ser)} \\ \\ \text{(Arg, Gly, Lys, Pro, Trp, Val)} \end{array} \right\} -NH_2$$

The pentapeptide fragment gave phenylalanine as the C-terminal amino acid upon carboxypeptidase action, leading to the partial structure (Glu, His, Met, Ser). Phe. From an Edman degradation, it was found that serine is N-terminal: Ser. (Glu, His, Met). Phe. Two more Edman degradations gave, first, methionine, and second, glutamic acid:



Thus the pentapeptide is Ser. Met. Glu. His. Phe. Histidine is placed fourth in line because that is the only slot left.

The hexapeptide (Arg, Gly, Lys, Pro, Trp, Val) was inert to carboxypeptidase, and, therefore, represents the C-terminal end of α-MSH. Further chymotryptic hydrolysis of this hexapeptide gave (Arg, Trp) and (Gly, Lvs, Pro, Val). The dipeptide must be Arg. Trp because chymotrypsin attacks the carbonyl function of tryptophan and not of arginine:



The tetrapeptide (Gly, Lys, Pro, Val) was subjected to the Edman degradation, and glycine was found to be the N-terminal amino acid. It was partially hydrolyzed in 12 NHC1 at 37 ºC for 120 hours. The following peptides were obtained, along with ammonia: (Gly, Lys, Pro) + (Gly, Lys) + (Pro, Val) + (Lys, Pro) + NH_3. The ammonia is formed because the C-terminal end is in the amide form. Since the Edman reaction established Gly as the amino terminal residue of this peptide, the isolation of the dipeptides (Gly, Lys), (Lys, Pro), and (Pro, Val) establishes the obligatory sequence Gly Lys. Pro’ Val(NH_2). The presence of tripeptide containing Gly, Lys, and Pro adds strength to the sequence assignment. The complete hexapeptide must, therefore, be Arg. Trp. Gly. Lys. Pro. Val(NH_2).

At this point, the information known about α-MSH is

(Ac)Ser. Tyr + Ser. Met. Glu. His. Phe + Arg. Trp. Gly. Lys. Pro. Val (NH_2)

If Tyr and Ser, and Phe and Arg, which were cleaved by chymotrypsin, are joined. there is only one way to put the sequence together:

This sequence is consistent with all data concerning the primary structure of α-MSH. A method for determination of the primary structures of proteins which is currently under investigation involves the use of a mass spectrometer to fragment the molecule and a computer to sort out and interpret the results.

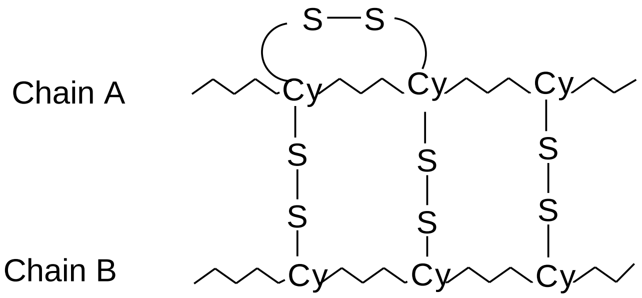
EXERCISE 5.3

A hexapeptide gave upon hydrolysis the following amino acids: Glu, Gly, Glv, Lvs, Ser, Tvr. When a partial hydrolysis was carried out, there were isolated two dipeptides and a tripeptide that had the following structures: Gly. Ser, Tyr. Lys and Gly. Glu. Gly. It was found that tyrosine occupied the N-terminal position. by an appropriate test. What is the structure of the hexapeptide?

5.7 PROTEINS:

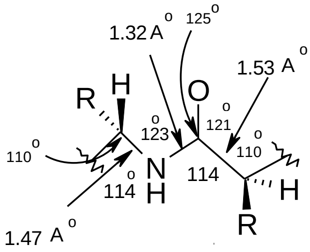
In spite of the fact that their molecular weights range up to millions, many proteins have been crystallized or at least purified until they behave as homogeneous substances. Care must be taken in the investigation of proteins, because they can be altered quite easily by changes in pH, by uv radiation, by heat. and by organic solvents. Such alteration is generally referred to as denaturation. A denatured protein, while very similar in its gross chemical structure to a native protein, no longer performs its important biological function. Simple proteins, for example the enzyme lysozyme, are hydrolyzed only to amino acids. Others contain non-amino acid portions, called prosthetic groups, and were originally referred to as conjugated proteins. In nucleoproteins (from cell nuclei), the prosthetic groups are nucleic acids: mucoproteins contain complex polysaccharides. Some prosthetic groups are much simpler. as exemplified by the oxidation-reduction enzymes known as fiavoproteins. which contain bound derivatives of the vitamin riboflavin.

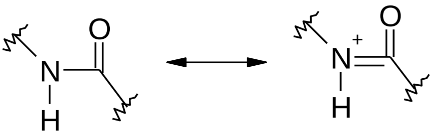
Proteins are amphoteric dipolar ions which migrate in an electric field and have characteristic isoelectric points. Even though the chain composing the backbone of the protein is comprised of relatively stable amide linkages, proteins are reactive and exhibit highly specific behavior. This reactivity is associated with the free active groups on the side chains, for example amino groups from lysine, guanido groups from arginine, or sulfhydryl groups from cysteine. Many proteins contain several peptide chains held together by cross linkages. Disulfide bonds between cysteines can link two chains, or even remote parts of the same chain: for example, beef insulin contains an A chain of 21 amino acids connected via two disulfide linkages to a B chain of 30 amino acids, forming a cyclic protein:



The discussion thus far has been involved with the characterization of polypeptides as a linear array of amino acids, that is, primary structure. One must not neglect, however, the manner in which these chains are arranged three-dimensionally. For instance, the finding that a particular amino acid side chain of an enzyme is involved in the catalysis of some reaction tells nothing about the details of its involvement; that is, we do not know whether it actually participates in the reaction, or is involved in the binding of the reactant (substrate) to the enzyme, or is merely necessary for maintaining the overall three-dimensional structural integrity of the enzyme molecule by interactions among the side chains of the constituent amino acids. All these roles, however, stipulate that the amino acid must be located very exactly. This spatial organization of proteins, as mentioned in the introduction to this chapter, is currently a topic of intense investigation in many laboratories.

An invaluable technique for studying three-dimensional protein structure is X-ray crystallography. An X-ray diffraction pattern is obtained from a crystal, and a structure is proposed, if possible, which would be expected to give such a pattern. From X-ray determinations of amino acid and peptide structures, the amide portion of the chain has been found to be planar and, anti. The following representation shows bond lengths and angles of a unit in the peptide chain:



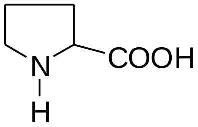


Anti coplanar

The carbon-nitrogen bond of the amide linkage has approximately 40 percent double-bond character, as a result of resonance. This resonance interaction strongly hinders rotation about that bond. Very importantly, however, rotations are free for bonds between the amide groups and the a carbons as well as for the a and the carbonyl carbons, thus permitting many conformations for the protein.

X-ray techniques were also instrumental in elucidating the two major ways in which the peptide backbone can interact with itself This level of organization is referred to as secondary structure. The first of these two types of interaction is shown in Figure 5.1 and is known as the a helix. Note that each amide group is hydrogen-bonded to the amide group, which is the third one from it in either direction along the chain. There are 3.6 amino acid units per turn of the helix. The side chains extend away from the axis of the helix. All natural amino acids are of the L-configuration and to date, all protein helices have been found to be right-handed. This is a very common structural component of proteins; an extreme example is the oxygen-carrying protein hemoglobin, which is about 75 percent a helix.

The cyclic nature of the amino acid proline forces the carbonyl group, which is attached to the proline nitrogen, to assume a conformational arrangement different from the one required for helix formation (which all the other amino acids can assume). Thus many proteins are found to consist of helical segments of different lengths, interrupted from time to time where the helix “goes a round corners.” The latter often occurs at the point where a proline residue is found in the amino acid sequence. In most proteins there are also regions where the amino acids are not ordered in any way that is simple to describe.



Proline

The interaction of various parts of the protein with each other via the amino acid side chains determines the tertiary structure of the protein. The bonds involved might be salt linkages, such as between an -NH_2 of Lys and a carboxyl of Asp; hydrogen bonds, such as between Ser and His; or van der Waals forces, such as between Tyr

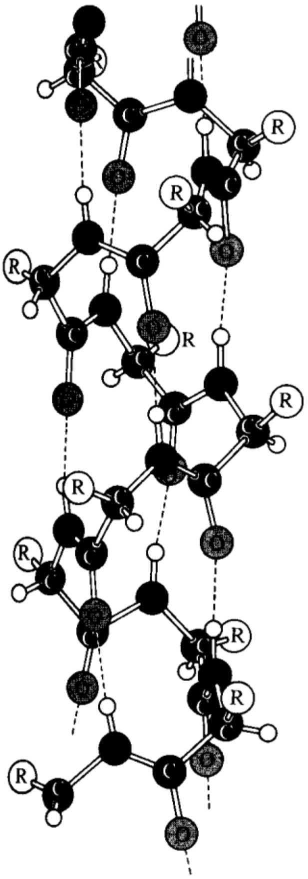


Figure 5.1: Alpha helix.

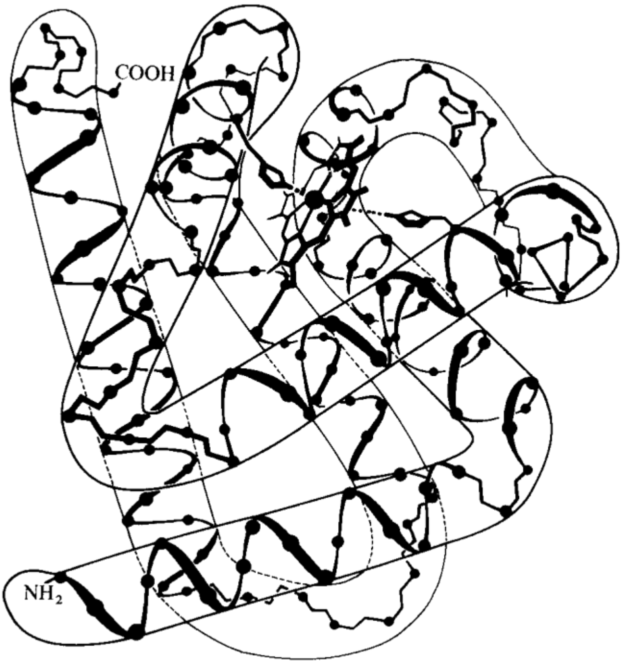


Figure 5.2: X-ray structure of myoglobin. (from R.E, Dickerson in H. Neurath(ed.) The proteins, Vol. 2, Academic Press, New York, 2nd ed., 1964, P 634).

and Phe. Disulfide bonds between cysteine residues on adjacent chains often help to stabilize the tertiary structure. X-ray crystallography has been used to determine the actual three-dimensional configurations of a number of proteins. Myoglobin, a protein similar to hemoglobin, has the shape shown in Figure 5.2. as determined by Perutz and Kendrew. The resolution of this X-ray study (and several similar studies) is not good enough to be able to see individual atoms: only gross shapes are discernible. The orientation of the peptide chain within this gross structure can be deduced since the primary structure of the protein is completely known. One of the consistent features of all protein structures studied so far is the presence of large numbers of polar amino acid residues on the surface of the molecule, with large clusters of nonpolar residues in the interior in contact with each other. Note the presence of several helical segments.

Quantitatively of less importance is the β, or pleated-sheet, structure, commonly seen in the fibrous proteins such as those found in silk, hair, and feathers. This is depicted in Figure 5.3. Notice that the chains are antiparallel. You can easily appreciate that steric crowding between R groups would make the straight-chain representation unfavorable. Thus, while silk, with a high percentage of Gly and Ala, could assume this configuration, bulky side chains would prohibit its formation. It is., of course, not obligatory that all parts of a protein molecule have either of these configurations.

There is one further level of organization in proteins, the quaternary structure, which describes the way multiple subunits (not always identical) can aggregate to form large complexes. As mentioned before, tobacco mosaic virus protein is actually a multiple of small subunits. As is usually the case with viruses, the protein complex forms a

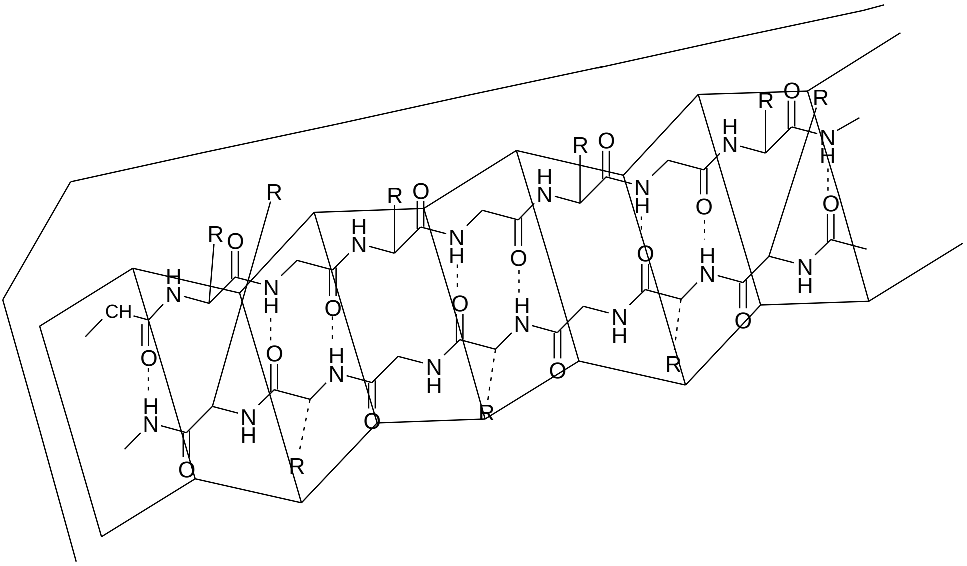


Figure 5.3: Beta, or Pleated-sheet, structure of proteins.

protective sheath around the nucleic acid core of the virus, which, of course, contains the genetic information required for the production of more virus particles.

Many sequential reactions in metabolism are efficiently catalyzed by well-organized complexes of enzymes which obviate the necessity of having the product of one enzymatic reaction float free in the cell waiting until it randomly collides with the enzyme required for its next transformation. Several of these complexes have been broken down into their individual enzyme components. The separated components can subsequently be reassociated in Vitro, and they will exhibit the original overall metabolic transformations. This experimental result shows that after synthesis of the individual enzymes, spontaneous assembly can produce the efficient complex observed in the cell.

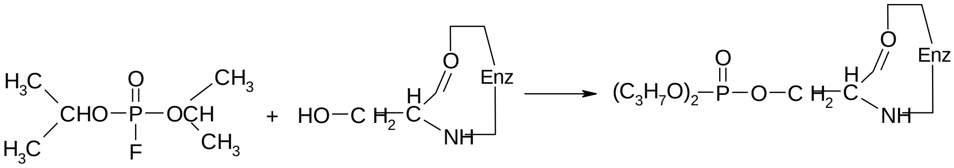
5.8 STRUCTURAL BASIS OF ENZYME CATALYSIS:

Having now looked at various factors that influence the overall structures of proteins, we are in a position to establish in a more meaningful way the correlation of structure with function which imparts to enzymes their extraordinary specificity and catalytic power. These characteristics (at least in enzymes that require no prosthetic groups for activity) must be determined solely by the specific spatial relationships among individual amino acid side chains of the polypeptide. No types of enzyme catalysis have yet been found that are mechanistically different from reactions carried out in test tubes, and similarly, no enzyme-catalyzed reactions have ever been documented that would not occur (eventually) without catalysis. It is noteworthy, however, that some enzymes have the capacity to speed up reactions by a factor of 1010 beyond their rates without catalysis.

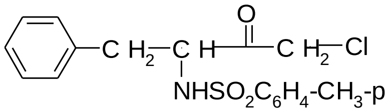
As an example to illustrate a structure-function relationship, we will use the enzyme chymotrypsin, which was mentioned in Section 5.6. As was described there, this proteolytic (peptide-hydrolyzing) enzyme preferentially attacks peptide bonds whose carbonyl function is furnished by an aromatic amino acid.

Some time before the three-dimensional structure of chymotrypsin had been established, a number of amino acids had been suspected by being components of the “active site” of the enzyme. A short description of several experiments that led to these suspicions will illustrate the type of approach used to investigate the mechanism of action of an enzyme.

1- Reaction of the enzyme with low concentrations of diisopropyl fluorophosphate led to rapid inactivation of the enzyme. Upon hydrolysis, the diisopropyl phosphate group was found to be covalently linked to a serine residue.



2- The pseudo-substrate 1-chloro-4-phenyl-3-(ρ-toluenesulfonamido)-2-butanone,



was found to react covalently with the enzyme in 1:1 stoichiometry, leading to complete loss of activity. Notice the designed similarity of the compound to a phenylalanine--containing peptide. suggestive that this compound is brought specifically into the region of the active site by the specificity-determining portions of the enzyme. Upon hydrolysis of the enzyme, the reagent was found to have reacted with a histidine residue.

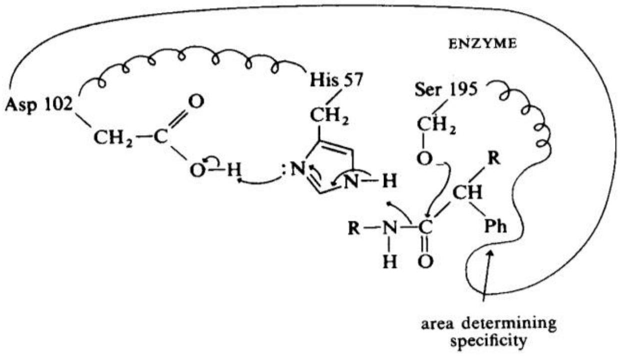
3- Further evidence implicating histidine as a component of the active site involved the drastic change in the enzyme's activity as the pH of the reaction was varied near the pK of an imidazole nitrogen.

4- When the enzyme reaction was carried out in D_2O rather than in water, the rate of hydrolysis decreased by over half, thus implicating a proton transfer (general acid-base catalysis) in at least the rate-determining step.

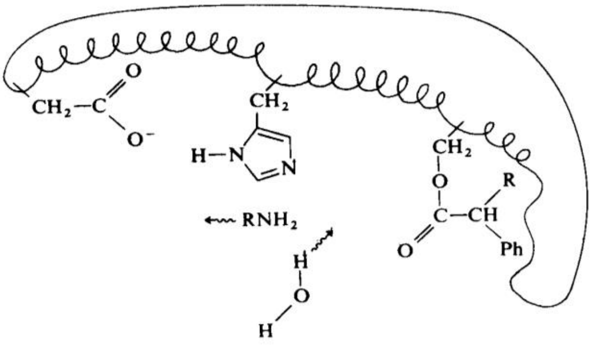
Subsequent X-ray analysis of the crystalline enzyme revealed that what would appear to be the active site is rather close to the surface of the molecule, as if designed to approach a rather large substrate. The high specificity of the enzyme would seem to be determined by two relatively short peptides within the enzyme which consist entirely of small nonpolar amino acids forming a kind of pocket into which the aromatic substrate would be held by van der Waals forces. Nearby were histidine and serine residues, and also in the same area was found an aspartic acid that could facilitate protonation of the histidine ring. Very importantly, these amino acids are not consecutive as one might at first think, but are actually separated from each other linearly by many other amino acids. Histidine is number 57, aspartic acid 102, and serine 195. The remainder of the protein is presumably involved in holding these catalytic and specific segments in the proper neighboring relationship.

The overall mechanism of action of chymotrypsin, shown in the following figures, was postulated before X-ray data were available (with the exception of the initial proton donation by aspartic acid). The confirmation by X-ray analysis of the feasibility of the proposed chemical mechanism is a striking example of the mutual benefit that can be derived from different but complementary lines of investigation of a single problem. The proposed mechanism of action of chymotrypsin is illustrated as follows:

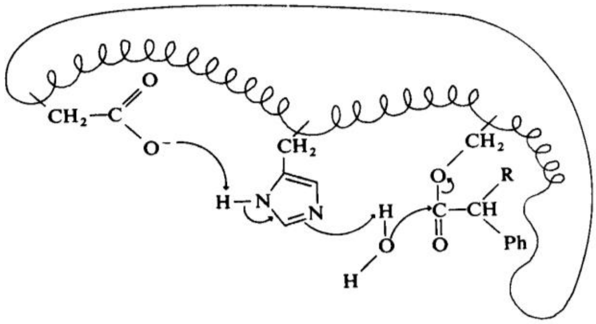
1. Initial Binding of the Substrate (after proton transfer from Ser to Asp via His)



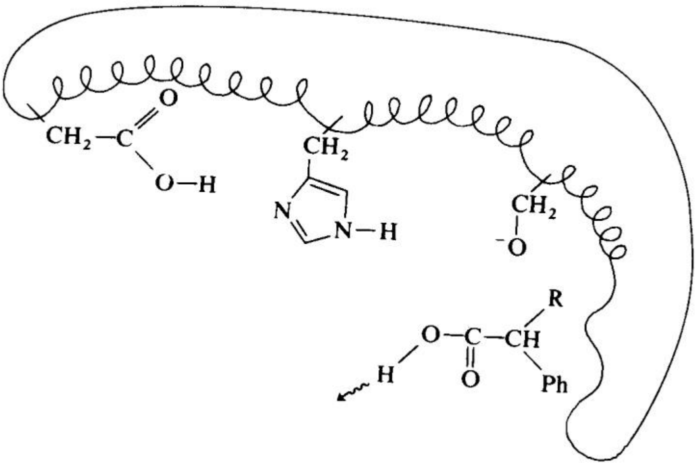
2. Formation of an Acyl Enzyme Intermediate and Release of Amine



3. Deacylation



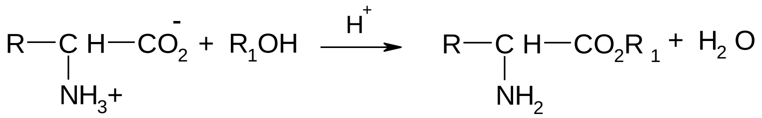
4. Release of Carboxylic Acid



5.9 PEPTIDE SYNTHESIS:

Peptide synthesis has long been an intriguing area of organic chemistry. The ability to duplicate in the laboratory one of the most complex processes of nature holds a fascination for the organic chemist. In any synthetic sequence directed toward the preparation of a particular polypeptide. there are difficulties to be encountered characteristic of the side-chain functional groups of the individual amino acids. These problems are too numerous and varied to go into here. However, some problems are common to all peptide syntheses. The two main difficulties seem to be (1) blocking the amino group of an amino acid while activating the acid group of the same molecule; and (2) blocking the acid group of an amino acid while leaving the primary amino group free to react.

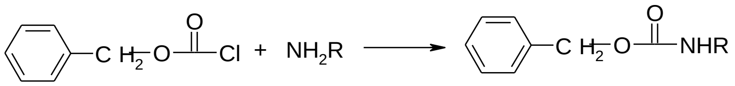
The second of the two problems is taken care of quite effectively by treating the amino acid with an alcohol to give an ester.



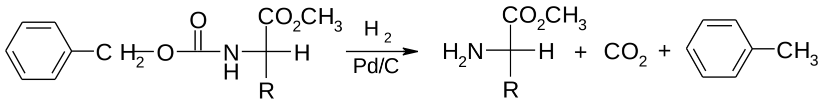
$$R-\underset{NH_3^+}{\underset{|}{C}}H-CO_2^- + R_1OH \xrightarrow{H^+} R-\underset{NH_2}{\underset{|}{C}}H-CO_2R_1 + H_2O$$

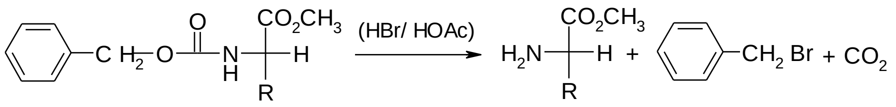
The first problem, that of blocking the amino group while activating the carboxyl group, is somewhat more involved and has been dealt with in many ways. The carbobenzoxy and t-butoxycarbonyl groups are widely used as amino blocking groups.

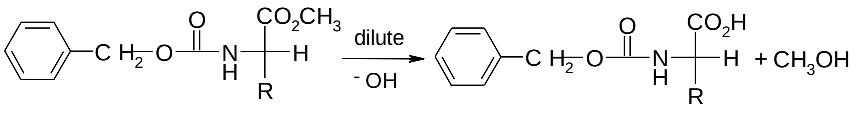
1- Carbobenzoxy. The N-carbobenzoxy (N-benzyloxycarbonyl) group is formed by treating an amino acid with benzyl chloroformate.



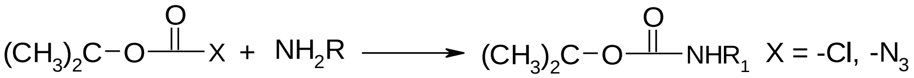
This protecting group has the advantage that it can be removed by hydrogenolysis or by acid hydrolysis, thereby generating the free amine, but it is relatively stable to dilute alkali.







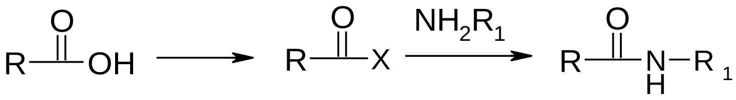
2- Butoxycarbonyl The N-t-butoxycarbonyl group is formed by treatment of the amino acid with t-butoxycarbonyl chloride (t-butyl chloroformate at 0 °C or by treatment of the amino acid with the more stable t-butoxycarbonyl azide at slightly higher temperatures.



$$(CH_3)_2C{-}O{-}\overset{O}{\overset{\|}{-}}{-}X + NH_2R \longrightarrow (CH_3)_2C{-}O{-}\overset{O}{\overset{\|}{-}}{-}NHR_1 \quad X = -Cl, -N_3$$

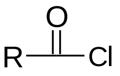
The N-t-butoxycarbonyl group can be removed by dilute acid, leaving the free amine. but it is unaffected by hydrogenolysis or dilute base.

As far as activating the acid function is concerned, it is necessary to convert the -OH of the acid to a better leaving group.



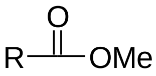
$$R{-}\overset{O}{\overset{\|}{-}}{-}OH \longrightarrow R{-}\overset{O}{\overset{\|}{-}}{-}X \xrightarrow{NH_2R_1} R{-}\overset{O}{\overset{\|}{-}}{-}\underset{H}{N}{-}R_1$$

Many such groups have been used and include the following:



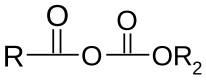
$$R{-}\overset{O}{\overset{\|}{-}}{-}Cl$$

Acid chlorides



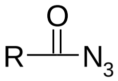
$$R{-}\overset{O}{\overset{\|}{-}}{-}OMe$$

Esters



$$R{-}\overset{O}{\overset{\|}{-}}{-}O{-}\overset{O}{\overset{\|}{-}}{-}OR_2$$

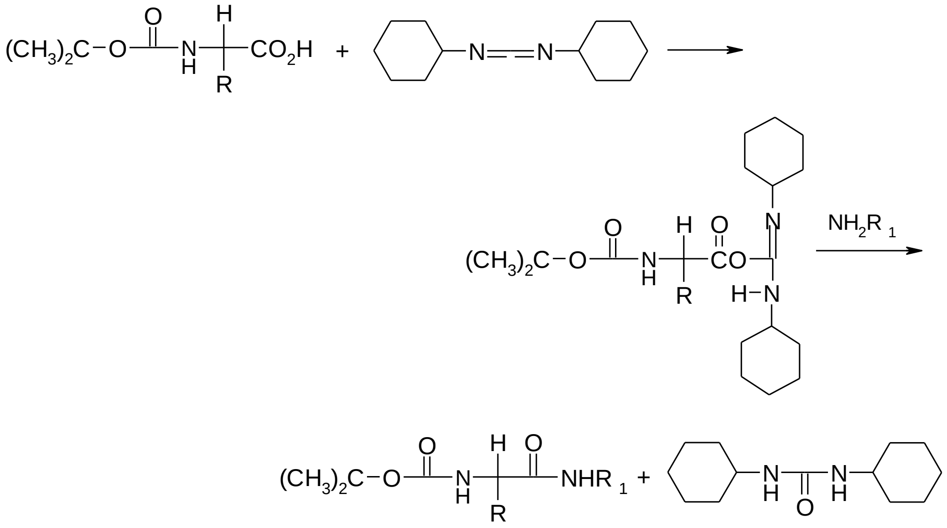
Mixed anhydrides with monoalkyl carbonates



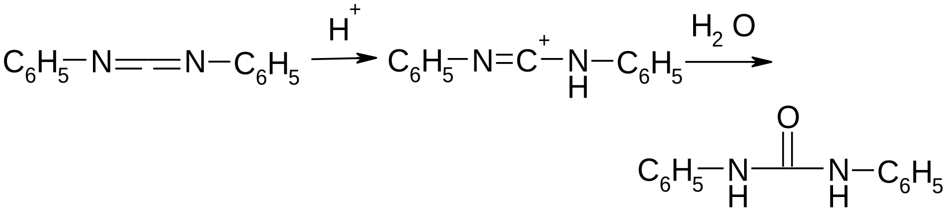
$$R{-}\overset{O}{\overset{\|}{-}}{-}N_3$$

Acyl azides

The use of these groups has been largely supplanted by that of dicyclohexylcarbodiimide-a reagent that in one step activates the carbonyl group and effects the coupling between an amino group and an acid with the removal of water.



Carbodiimides. These are a special class of imines of the general structure R-N=C=N-R: they are formally diimides of carbon dioxide. They add nucleophiles readily at the central carbon atom. For example, water adds to give substituted ureas; the reaction is acid-catalyzed.



$$C_6H_5-N=C=N-C_6H_5 \xrightarrow{H^+} C_6H_5-N=\overset{+}{C}-\underset{H}{N}-C_6H_5 \xrightarrow{H_2O} C_6H_5-\underset{H}{N}-\overset{O}{\overset{\|}{C}}-\underset{H}{N}-C_6H_5$$

N,N'- Diphenylurea

Carbodiimides can be prepared by dehydration of ureas, the reverse of the preceding reaction.

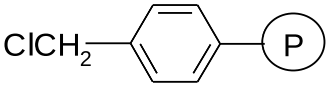
One can well imagine how it would be possible to build a chain of any length with amino acids in any desired sequence by adding on one at a time. In the laboratory, however, this is not the usual method. Normally small (two-, three-, or four-unit) polypeptides are constructed and then these units are coupled as if they were themselves amino acids. Du Vigneaud,' in fact, employed this method in his landmark synthesis of oxytocin in 1954. The main disadvantage of this synthetic approach is the racemization, which is nearly always a problem in each step; the purification of intermediates of two, three, or four polypeptide units to optical purity is consequently very difficult.

A recent and rather novel approach to polypeptide synthesis is that of Merrifield. The method, essentially, is to bind an amino acid through the carboxyl group to a highly porous polymeric resin. The t-butoxycarbonyl protection can then be removed from the amino group by merely washing the resin with acid. By then washing the resin with an activated acyl compound, the polypeptide chain is built up. In this method, the purification of smaller intermediates is eliminated, but, depending upon the accumulated amount of racemization, the problem of purification of the final polypeptide product still remains. By this method biologically active ribonuclease was synthesized by assembling the 124 amino acids using 369 chemical reactions and 11,931 steps of the "peptide synthesis machine," which carries out the operations with a minimum of human intervention.

The purity of the final protein presents a difficult synthetic problem, which can be understood using the ribonuclease synthesis as an example. Suppose that a 90 percent yield of pure product is obtained when each amino acid is added, the other 10 percent of the product being racemic or other impurity. As most of us know from ourlaboratory work, a 90 percent yield is usually pretty good After 124 steps, each with a 90 percent yield, the overall yield is (90 percent) = 0.0002 percent. A 95 percent yield on each step will give an overall yield of 0.2 percent, while if the yield can be raised to 99 percent on each step, the overall yield can be raised to 30 percent. Clearly, yield and purity are extremely critical problems in protein synthesis.

EXERCISE 5.4:

The polymeric resin used in the Merrifield "solid-phase" synthesis can be designated as:



Chloromethylated polystyrene

In the first step of a solid-phase synthesis, a N-t-butoxycarbonyl-protected amino acid (*t*-BOC-amino acid), as its sodium salt, is allowed to react with the polymer. The product thus obtained is then treated with dilute acid. Draw a diagram that would be representative of the solid phase at this stage, if phenylalanine were used as the amino acid.

EXERCISE 5.5:

Draw the structure of the material which would he obtained by a1lowing the product produced in Exercise 5.4 to react with t-BOC-glycine in the presence of dicyclohexyl-carbodiimide.

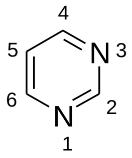
EXERCISE 5.6:

Beginning with the material obtained in Exercise 5.5, outline the steps necessary for the synthesis of the tetrapeptide Pro. Ala. Gly. Phe. Note that the completed peptide can be released from the polymer support by treatment with hydrogen bromide in trifluoroacetic acid.

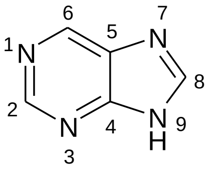
5.10 THE NUCLEIC ACIDS:

Of the three types of vitally important biopolymers. polysaccharides (Chapter 4) and proteins (Section 5.7) have already been discussed. The nucleic acids constitute the remaining type. Two varieties of nucleic acids are found in cells, ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). As discussed below. DNA constitutes the genetic material of cells. Both RNA and DNA are essential for the biosynthesis of proteins. Like the proteins themselves, RNA and DNA are very large molecules: molecular weights of up to 2,800,000.000 have been estimated for some DNAs.

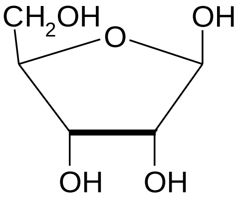
Following are shown the structures of some compounds we shah need to have in mind before going further with the discussion of nucleic acids. These compounds have all been mentioned in earlier chapters. but their structures will be repeated here for convenience. They are pyrimidine and purine and ribose.



Pyrimidine

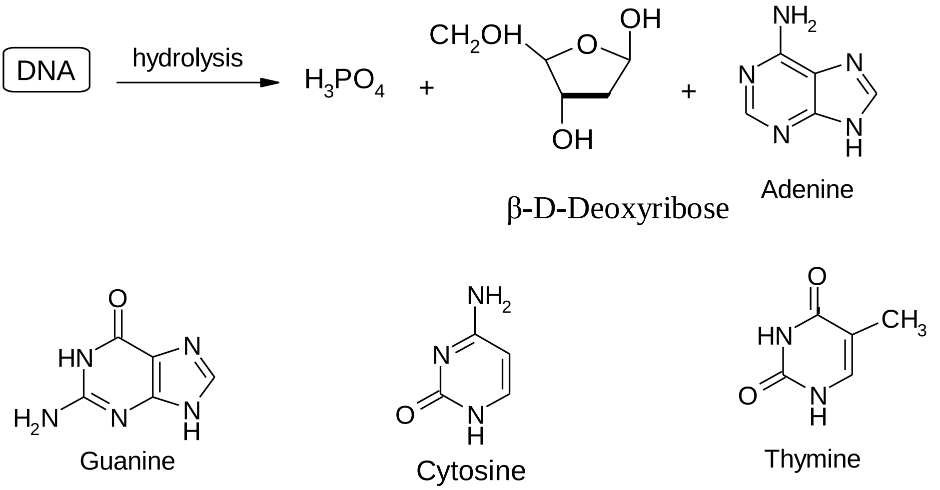


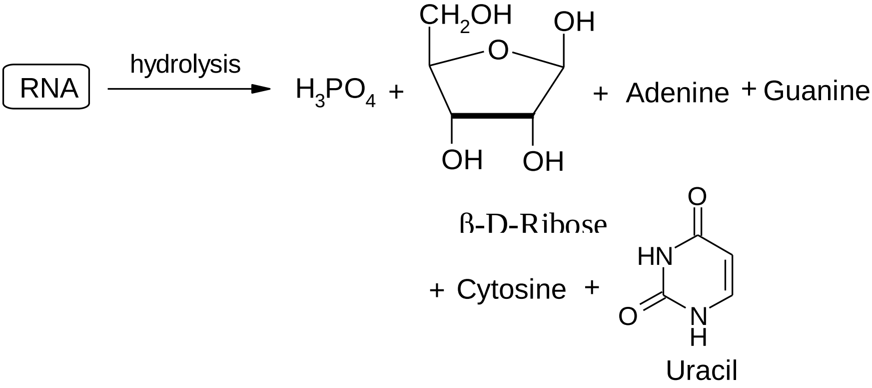
Purine



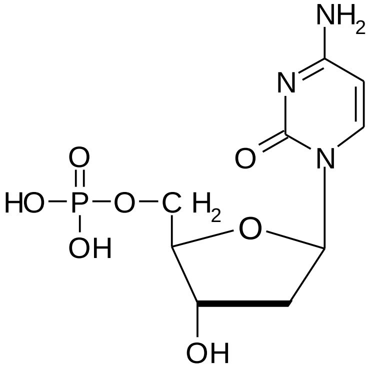
Ribose

On hydrolysis, both types of nucleic acids yield phosphoric acid. a sugar. and a mixture of purine and pyrimidine bases. The sugar from RNA is ribose. that from DNA is deoxyribose. The major bases from DNA are the purines adenine and guanine and the pyrimidines cytosine and thymine. RNA yields mainly adenine. guanine. Cytosine, and another pyrimidine base, uracil.

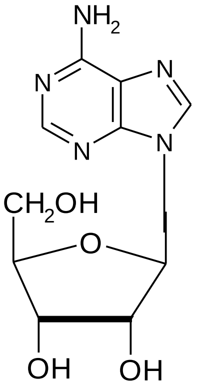




Mild degradation of a nucleic acid yields a mixture of acids known as nucleotides. Each nucleotide contains the elements of one purine or pyrimidine base. one phosphate unit, and one pentose unit The phosphate unit may be selectively removed by further careful hydrolysis to convert a nucleotide into a nucleoside. a molecule built up of a pentose joined to a purine or pyrimidine base. In a nucleotide, C-l of the sugar is joined to N-l of a pyrimidine or N-9 of a purine: the phosphoric acid unit is present as an ester at C-5 of the sugar.

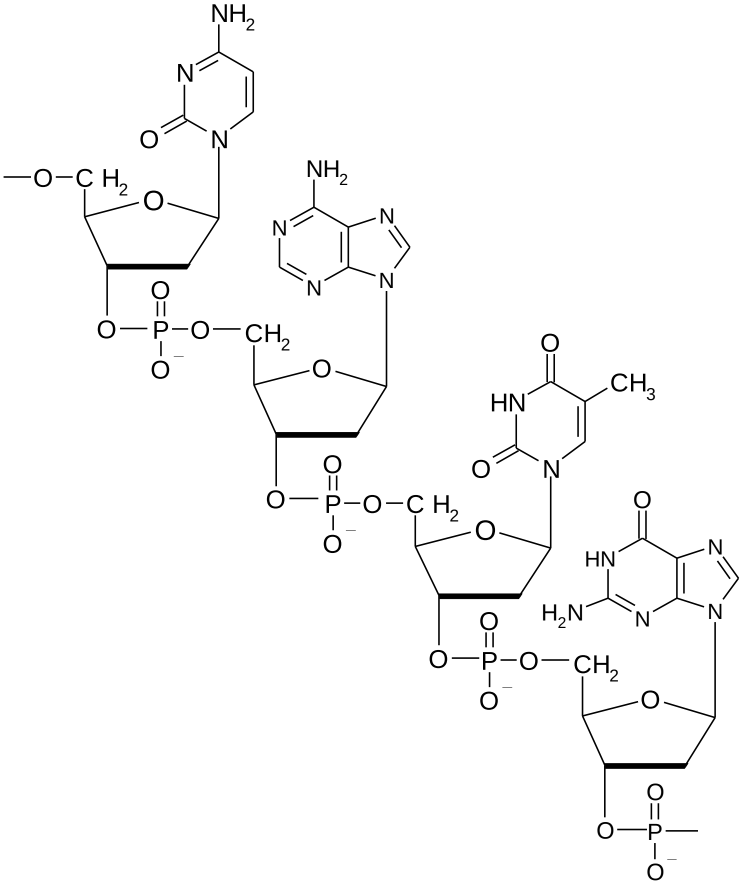


Deoxycytidylic acid
(a nucleotide)

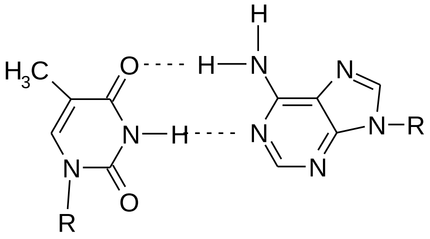


Adenosine
(a nucleoside)
(A)

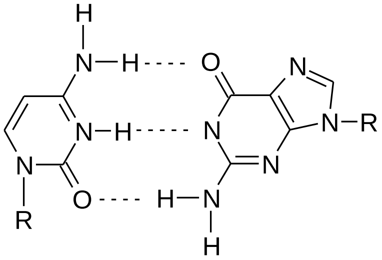
In a nucleic acid chain the phosphoric acid is esterified to form a bridge between C-5 of the sugar of one nucleoside and C-3 of the sugar of another nueleoside. In this way, the sugar-phosphate units can form a long backbone or framework, which bears purine and pyrimidine base substituents at regular intervals. A typical segment of a DNA chain is shown.



Watson and Crick in 1953 proposed the now-accepted double-helical structure of DNA. According to their analysis, the DNA molecule actually consists of two complementary strands that are twisted about a common axis as helices having the same chirality (handedness). Each adenine unit of one chain is specifically hydrogen-bonded to a thymine of the opposite chain, and each guanme of one chain is similarly bonded to a complementary cytosine unit.



Thymine-adenine
(T to A)
hydrogen bonding

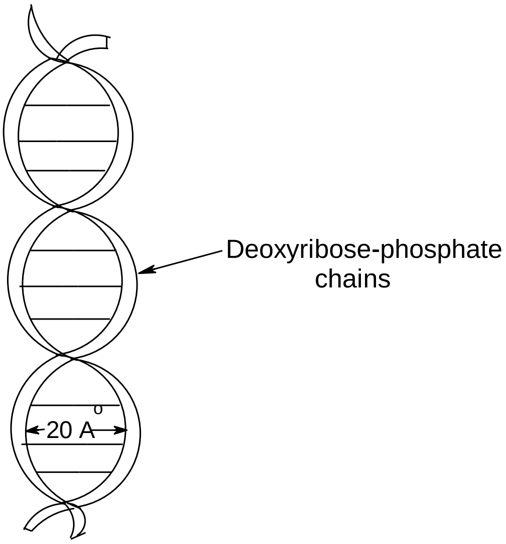


Cytosine-guanine
(C to G)
hydrogen bonding

The double-helical structure of DNA is shown schematically in Figure 5.4. The helical strands represent the sugar-phosphate backbones, which are held nicely in place by hydrogen bonding between the complementary base units. The order of the bases on the chain of the DNA molecule is extremely significant biologically; it is the fundamental source of the hereditary information of the genes.

A molecule of DNA reproduces itself by a remarkably simple mechanism. The two strands of the DNA molecule dissociate, and then free nucleotides hydrogen-bond with the nucleotides of the dissociated strands. An enzyme catalyzes the polymerization of these free nucleotides in an order complementary to that of the original strands, producing two new double-stranded DNA molecules identical to the original one. It

Figure 5.4: Double-helix structure of DNA.



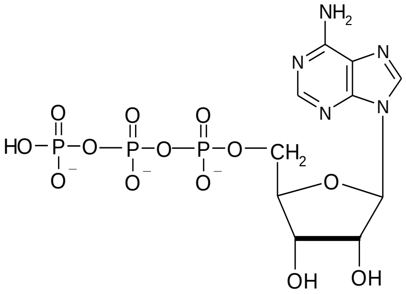
has been said that this simple process is the “secret of life.” If not the secret of lift, it is at least the secret of why children look much more like their relatives than like elephants or oak trees.

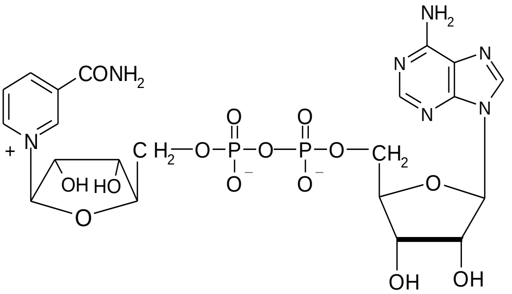
The DNA has two important functions. First it is the reference material or "book of instructions" as to how a given plant or animal is to be constructed. and it reproduces itself as explained above. Its second function is to act as a template in producing RNA. RNA is the material that actually carries out the synthesis of proteins.

Nucleosides and nucleotides serve other very important biochemical roles as portions of essential biological catalysts (coenzymes). Adenine units are most frequently encountered in these compounds.

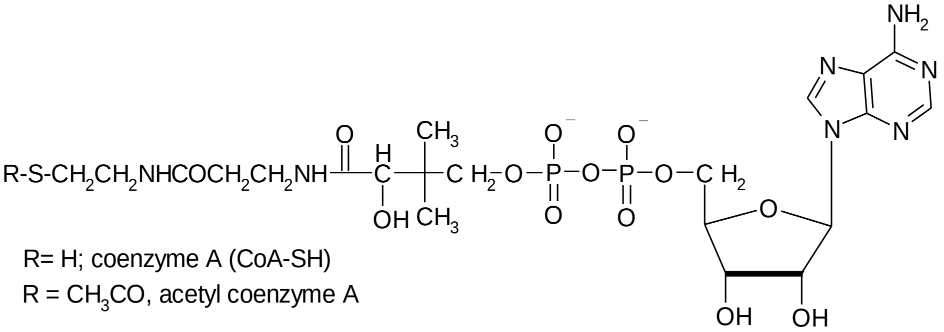
Numerous enzymes require the presence of a small nonprotein moiety more or less tightly bound to the protein for efficient performance of catalytic function. Since their nonprotein moieties are intimately involved with the overall reaction, they are termed coenzymes.

Examples of such nucleotides are nicotinamide adenine dinucleotide (NAD), and adenosine triphosphate (ATP), which functions as a pool for chemical energy in cells because of its energy-rich triphosphate unit.


Adenosine triphosphate
(ATP)

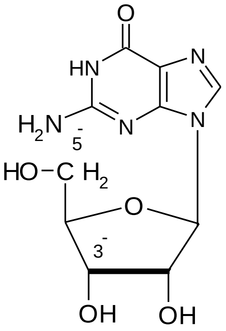

Nicotinamide adenine dinucleotide
(NAD)

A third very important adenine nucleotide derivative is coenzyme A (CoA-SH), which plays a very significant role in biosynthesis, especially in acyl group-transfer reactions. Acyl derivatives of CoA-SH are effective acylating agents because of diminished resonance interaction between the sulfur atom and the carbonyl group, as compared with oxygen-carbonyl delocalization in ordinary esters.

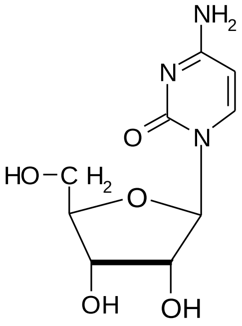

R= H; coenzyme A (CoA-SH)
R = CH_3CO, acetyl coenzyme A

5.11 PROTEIN BIOSYNTHESIS:

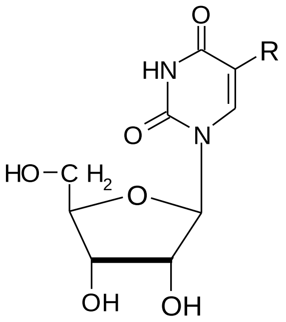
Ribonucleic acids have the general formula shown in the previous section. Note that the nitrogen bases cytosine, uracil, guanine, and adenine occur in a precise and characteristic sequence in any given ribonucleic acid chain. A more wieldy general formula results from the following abbreviations. The letters U, G, A, C, and T represent the nucleosides (ribose-nitrogen base conjugates) uridine, guanosine, adenosine (Section 5.10), cytidine, and thymidine, respectively.



Guanosine
(G)



Cytidine
(C)



Uridine (R= H) (U)
Thymidine (R= CH_3) (T)

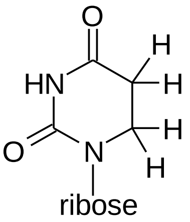
The letter *p* represents the phosphate unit. Notice (Section 5.10) that the phosphate esters link at 5′, 3′, which gives a directionality to the polymer. Thus, by convention the 5′ end of the chain is written to the left in our abbreviation and the 3′ end to the right. So the segment of DNA chain shown on page 712 can be abbreviated:

—CpApTpGp—

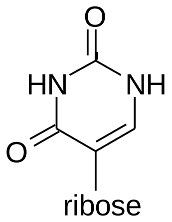
There are three major classifications of RNA; they are designated according to their functional properties. Messenger RNAs (mRNA) carry the genetic information from DNA for subsequent translation into specific protein sequences. Transfer RNAs (tRNA), previously called soluble RNAs, have molecular weights of 25,000-30,000, corresponding to 75-90 nucleotides. Ribosomal RNAs (rRNA) are much larger, having z molecular weights of 0.5-1 million. The two latter types of RNA will be discussed below.

The mechanisms by which the ribonucleic acids perform their varied and complicated functions are subject to intensive study among biochemists today. One key theory has emerged (mentioned in Section 5.10), which has defied attempts at disproof. It is the very specific hydrogen bonding that exists between AU base pairs and between GC base pairs, which is responsible for determining the eventual amino acid sequence in proteins. The diagram presented on page l2 actually depicts an AT pair. but inspection of the structures of T and U will show that both have the same hydrogen-bonding properties. No such analogous complementary hydrogen bonding can exist between AC, AG, UG, or LC pairs. Thus, a trinucleotide such as CpUpC will be strongly attracted to its complementary trinucleotide GpApG, less strongly to one such as CpApC, and virtually not at all to one such as ApCpU. This binding specificity controls the process of protein synthesis in the living cell.

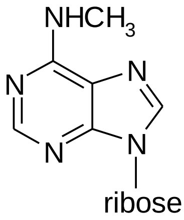
How does this synthesis take place? First, the free amino acids which have been synthesized by the body or derived from ingested nutrients become esterified to the 3′-OH end of a specific tRNA. This molecule is currently the subject of much attention and a short look at its structure as it is understood today will be profitable. The exact nucleotide sequence of a number of tRNAs has now been established, and preliminary structural characteristics have been described on the basis of X-ray data. As stated above, all tRNA molecules are approximately the same size and share other common features. Invariably, they terminate on the 3’ end with the sequence -pCpCpA-OH. Also, they all undergo a certain amount of chemical modification after their initial synthesis, leading to a number of modified bases. such as



Dihydrouridine
(diHU)



Pseudouridine

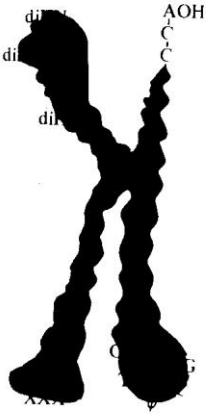


6-Methyladenosine
(mA)

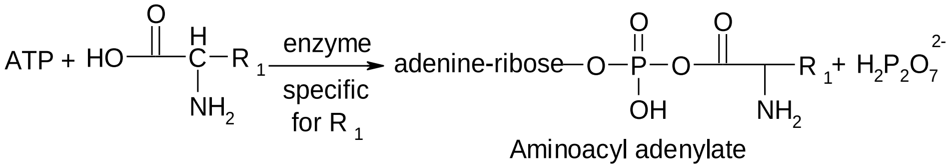
These unusual bases are present in every tRNA in the form of (1) a looped-out region with a sequence locally rich in diHU, and (2) another loop sequence, which is invariably —pCipTpΨpCpGp—. With all these similarities, where, then, are tRNAs different, and how does each amino acid invariably become attached to only one species of tRNA? The one answer with which all can agree involves a region very near the center of the chain and consists of a triplet of nucleotides whose sequence is different for each of the amino acid-specific tRNAs studied thus far. This trinucleotide sequence is called the anticodon region, for reasons soon to become apparent. There are also localized heterogeneities in the diHU loop from one species of tRNA to another.

A number of three-dimensional models of tRNAs have now been proposed. A tRNA is schematically represented in Figure 5.5. where XXX represents the anticodon, hydrogen bonding is indicated by dotted lines, and p's have been omitted for the sake of brevity.

Figure 5.5 Schematic diagram of transfer ribonucleic acid.



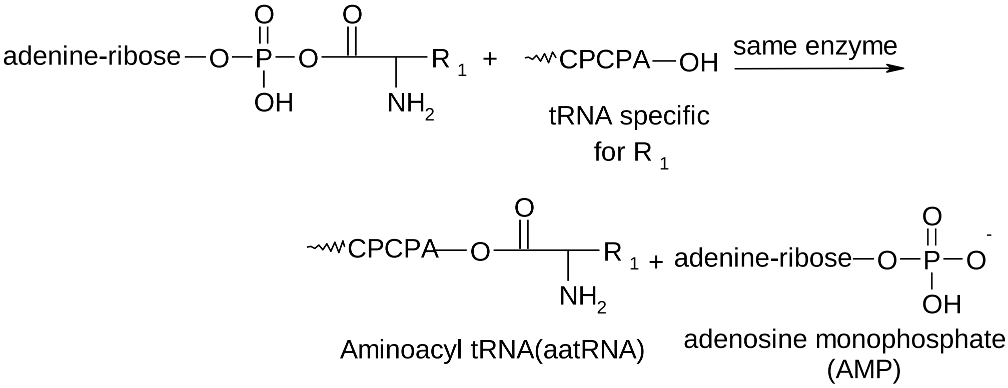
The enzyme that joins the amino acid to the tRNA catalyzes first the "activation" of the amino acid with the high-energy biochemical adenosine triphosphate (ATP) (Section 5.10).



$$\text{ATP} + \text{HO}-\overset{\overset{\text{O}}{\|}}{\text{C}}-\underset{\underset{\text{NH}_2}{|}}{\overset{\text{H}}{\text{C}}}-\text{R}_1 \xrightarrow[\text{specific for R}_1]{\text{enzyme}} \text{adenine-ribose}-\text{O}-\underset{\underset{\text{OH}}{|}}{\overset{\overset{\text{O}}{\|}}{\text{P}}}-\text{O}-\overset{\overset{\text{O}}{\|}}{\text{C}}-\underset{\underset{\text{NH}_2}{|}}{\text{C}}-\text{R}_1 + \text{H}_2\text{P}_2\text{O}_7^{2-}$$

Aminoacyl adenylate

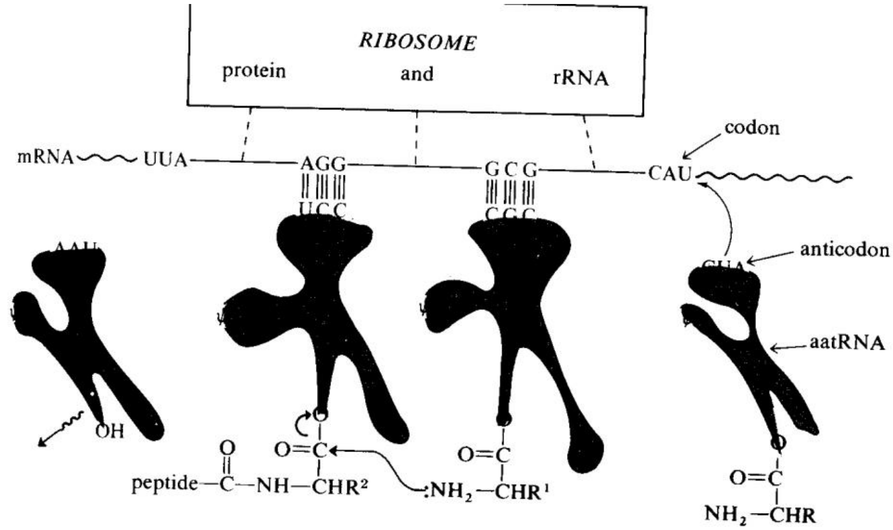
The second step involves the actual esterification:



This reaction is carried out with each amino acid. each time utilizing an amino acid specific enzyme and tRNA. until all amino acyl tRNA esters have been produced. each with a specific anticodon. It is in this form that the amino acids are ready to he polyrnerized into polypeptide linkages.

The actual formation of the peptide bond occurs on a nucleoprotein particle called the ribosome. Every living cell capable of synthesizing protein contains many ribosome, and even though there are minor differences depending on the source. they are always composed of one small and two large strands of RNA (rRNA), with which are associated some 60 or 80 smaller proteins. The functional nucleoprotein particle is roughly pear-shaped with a diameter of about 200 A. It is with the ribosome that the mRNA interacts prior to protein synthesis. After formation of the mRNA-ribosome complex, different amino acyl tRNA (aatRNA) molecules in the surrounding medium come into contact with it. If the exposed anticodon loop is exactly complementary to the specific triplet (the codon) on the ribosome-stabilized mRNA, it will form a hydroaen-bonded triplet of base pairs. Although the exact sequence of events that follows is not known with certainty, strong evidence indicates that the Ψ-containing loop of the tRNA becomes attached to the ribosome and further stabilizes the ternary aatRNA-mRNA-ribosome complex. If only two or three nucleotides are capable of pairing, the association will not be strong enough for the 41 loop to "lock" the complex, and the "wrong" aatRNA will diffuse away and be replaced by another until a correct match is made. After this happens, the following codon on the mRNA (the next three nucleotides) is in a position to react with its complementary anticodon on another tRNA. When two such alignments have been made, the carboxyl group of the first amino acid participates in amide bond formation with the free amino group of the recently incoming aminoacyl tRNA. This enzyme-catalyzed reaction, of course, frees the first amino acid from its tRNA, which by unknown means senses this change and diffuses away from the complex, leaving behind a dipeptidyl-tRNA-mRNA-ribosome complex. The ribosome then moves "down" the mRNA just enough to bring the next (third) codon into a position in which it can react with its own anticodon on vet another aminoacyl tRNA. This sequence of reactions is repeated until some chain-termination signal on the mRNA causes synthesis to stop and the finished protein to be released for use as a hormone, enzyme, or structural protein An intermediate stage in the process is depicted schematically in Figure 5.6.

Figure 5.6 Schematic diagram of polypeptide biosynthesis. The mRNA has become associated with the large ribosome. The codons of the mRNA are shown interacting with the complementary anticodons of the aatRNAs.



Using known values of hydrogen-bond energies, theorists have calculated that if the genetic code were a doublet or quadruplet one, life as we know it would not be possible. The attractive force between two hydrogen-bonded base pairs is inadequate to keen the large tRNA molecule in place on the ribosome sufficiently long for pentide bond formation to occur. Conversely, four base pairs would be so strongly bonded that proteins might require months or more to be synthesized. The actual measured rate of protein synthesis in the living cell is about two amino acids per second.

The next several years should see a tremendous increase in the knowledge not only of the details of this complex process but also of such intriguing topics as the transcriptional and translational control of protein synthesis. the design of specific drugs to control the growth of harmful bacteria and viruses, and control of the rejection problem in tissue and organ transplants.

PROBLEMS

1- Name the three types of natural organic polymers that are essential to life processes.

2- Give a definition and example of:

   (a) An essential amino acid
   (b) The peptide bond
   (c) Polypeptide
   (d) Oligopeptide
   (e) Tripeptide
   (f) Enzyme
   (g) Structural protein

3- What is meant by the primary, secondary, tertiary, and quaternary structure of polypeptides?

4- Define the isoelectric point of an amino acid and explain why arginine has an isoelectric point 5 pH units higher than does glycine, and glutamic acid has one about 3 pH units lower than that of glycine.

5- Write the structures that correspond to the following formulas:

   (a) Val. Trp. Lys. (NH_2).
   (b) (Ac) HiS. Gly. CySH. Gln.

6- Describe the molecular disease of sickle-cell anemia. Explain why a human being cannot use a blood transfusion from a monkey. but can use insulin from a cow.

8- Explain the difference in the chances of the success of a heart transplant between identical twins and a heart transplant between nonidentical twins.

9- Decribe the molecular basis of immunity.

10- Glutathione is a tripeptide that is an important regulator of the oxidation-reduction reactions of cells in animals. From the following experimental results, suggest a structure for glutathione:

(a) Enzymatic or acid hydrolysis gives glycine, cysteine, and glutamic acid in equimolar amounts.

(b) Mild hydrolysis gives two dipeptides: one on further hydrolysis gives cysteine and glutamic acid, and the other gives cysteine and glycine.

(c) Carboxvpeptidase liberates glycine.

(d) 2,4-Dinitrophenvlation Lives N-(24-dinitrophenvl) glutamine.

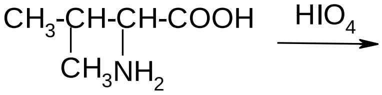
11- Describe by formula two N-terminal amino acid determinations.

12- Compare the merits of the chemical *C*-terminal amino acid determination with the enzymatic determination for the same purpose.

13- An octapeptide was found to contain the following amino acids: Ala, Ala. His. Leu. Lys. Pro. Thr. Tyr. Upon partial hydrolysis there were isolated from the resulting mixture four tri-peptides which had the following structures: Leu. Ala. Tyr. Thr. Pro. Leu, Lys. His. and His. Thr. Pro. A *C*-terminal amino acid determination showed that tyrosine occupied that position. What is the structure of the octapeptide?

14- Another octapeptide upon hydrolysis gave the following amino acids: Met. Asn, CySH. Lys. Pro. Thr, Thr, and Val. Partial hydrolysis gave a mixture from which it was possible to isolate four dipeptides and two tripeptides. which had the following structures: Met, Lys. Val. Thr. Asn. Val. Thr. Met. Pro. Thr. Asn. and Lys. CySH. Pro. What is the structure of the ecta-peptide?

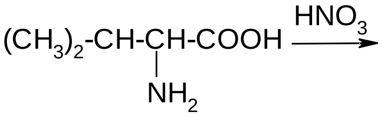
15 - Complete:

(a)


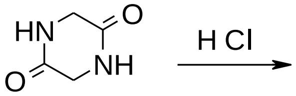
$$CH_3\text{-}\underset{CH_3}{CH}\text{-}\underset{NH_2}{CH}\text{-}COOH \xrightarrow{HIO_4}$$

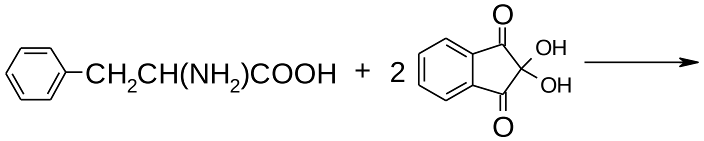
(b) $CH_3COOH \xrightarrow[(2)\ 2\ NH_3]{(1)\ Br_2,\ P}$

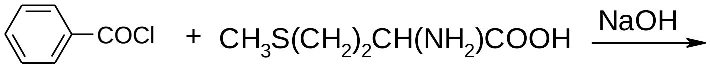
(c) $CH_3CHO \xrightarrow[(2)\ H_3O^+]{(1)\ NH_3\ ,HCN}$

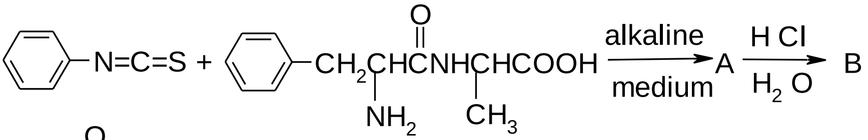
(d)


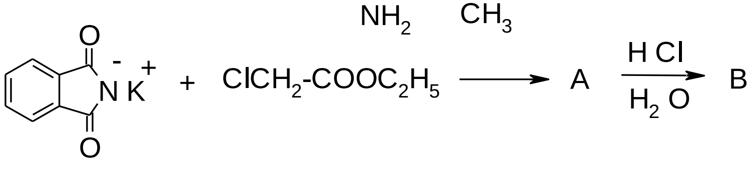
$$(CH_3)_2\text{-}CH\text{-}\underset{NH_2}{CH}\text{-}COOH \xrightarrow{HNO_3}$$

(e)


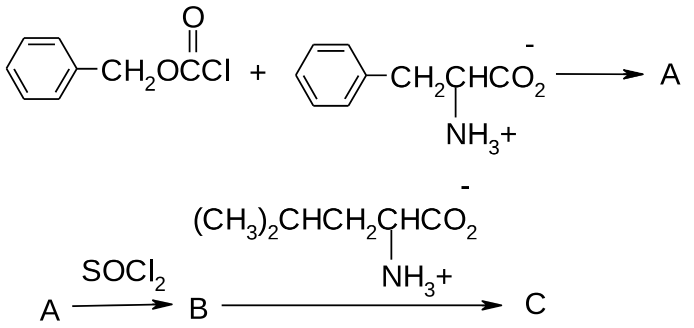
(f) (Optional)


(g)


(h)


(i)


(j)



16- Suggest a synthesis for

(a) Ala. Pro. Val.

(b) Asp. Ile. Gly.

(c) Thr. Arg. Lys.

17- Outline a synthesis for lysine from the oxime of cyclohexanone.

18- Discuss the stereochemical implications of obtaining L-Val. L-Ile. L-Lys. by resolution of DL-Val. DL-Ile. DL-LyS.

19- Vasopressin is a posterior pituitary hormone that acts on the kidney to reduce excretion of water and brings about a rise in blood pressure:

Cy. Tyr. Phe. Gln. Asn. Cy. Pro. Lys. Gly (NH_2)

Lysine vasopressin

(a) Devise a scheme for determining the peptide sequences.

(b) Devise a synthesis of the following part of the chain: CysH. Pro. Lys. Gly(NH_2).

20- Cells contain two types of nucleic acids. commonly referred to as RNA and DNA.

(a) What do the abbreviations RNA and DNA stand for?

(b) What are the similarities between RNA and DNA?

(c) What are the differences between RNA and DNA?

(d) In nucleic acid chemistry, the abbreviations A. C, G, T, and U are commonly employed. Give the structures and names for the compounds represented by these letters.

(e) What is the chemical basis of the Watson and Crick double-helix structure for DNA?

21- Show the mechanism of peptide coupling by dicyclohexylcarbodiimide.

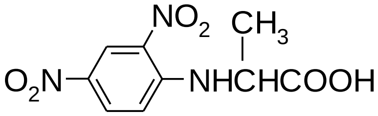
22- Describe the role of tRNA in protein biosynthesis.

23- Describe the function of the anticodon.

24- What is the maximum number of different amino acids possible in human protein?

25- A certain tetrapeptide is found to yield on hydrolysis 2 mol of alanine, 1 mol of glycine. and 1 mol of valine. Write the formula for each possible structurally isomeric tetrapeptide that could give this result.

26- The careful fractionation of the nonlipid portion of the spleen of a red snerd yielded a pure heptapeptide. Complete hydrolysis of this polypeptide yielded alanine (2 mol), cystine (1 mol), glutamic acid (2 mol). and glycine (2 mol). When the polypeptide was allowed to react with l-fluoro-2,4-dinitrobenzene and the product was hydrolyzed, the following compound was isolated:



The partial hydrolysis of the polypeptide yielded a mixture from which these three tripeptides could be isolated: CySH. Glu. Glu, Gly. Gly. CySH, and Glu. Ala. What is the structural formula of the heptapeptide?

27- Use the Merrifield solid-phase method to outline a synthesis of the tetrapeptide: Phe. Ala. Gly. Gly.

Enzymes

The word enzyme means “in yeast.” Even without any knowledge of their structures or functions, humans have used enzymes since prehistoric times in the production of wine, vinegar, and cheese. Pasteur thought that living yeast cells were necessary for fermentation processes. We now know that a living cell is not necessary; the proper enzymes, plus reaction conditions that do not cause denaturation, are all that are needed for enzymatic reactions.

An enzyme is a biological catalyst. A higher animal contains thousands of enzymes. Virtually every biochemical reaction is catalyzed by an enzyme. Even the equilibrium $CO_2 + H_2O \leftrightharpoons H_2CO_3$ is enzyme-catalyzed because the rate of the uncatalyzed equilibration does not produce carbonic acid fast enough for an animal’s needs.

Enzymes are more efficient catalysts than most laboratory or industrial catalysts (such as Pd in a hydrogenation reaction). Biological reactions in humans occur at 37^0C and in aqueous media. High temperature, high pressure, or very reactive reagents (such as NaOH or LiA1H4) are not available tt an organism. Enzymes also allow a selectivity of reactants and a control over reaction rate that can be obtained with no other class of catalyst.

All enzymes are proteins. Some are relatively simple in structure; however, most are complex. The structures of many enzymes are still unknown. For biological activity, some enzymes require prosthetic groups, or cofactors. These cofactors are nonprotein portions of the enzyme. A cofactor may be a simple metal ion; for example, copper ion is the cofactor for the enzyme ascorbic acid oxidase. Other enzymes contain nonprotein organic molecules as cofactors. An organic prosthetic group is frequently referred to as a coenzyme.

If an organism cannot synthesize a necessary cofactor, the cofactor must be present in small amounts in the diet. The active units of many cofactors are vitamins. Table 6.5 shows a few cofactors and the corresponding vitamins.

A. Naming enzymes:

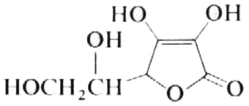
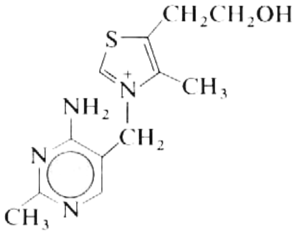
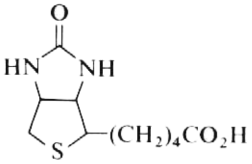
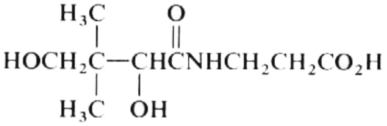
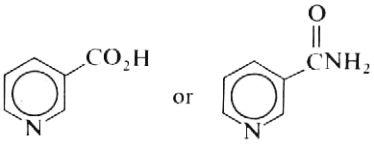
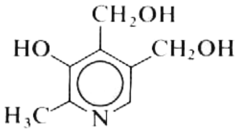
Most enzymes are named after the reactions that they catalyze. The ending for an enzyme name is usually -ase. The name may be general and refer to a class of enzymes that catalyze a general type of reaction. For example, a polyinerase is any enzyme that catalyzes a polymerization reaction, and a reductase is any enzyme that catalyzes a reduction reaction. An enzyme name may also refer to a specific enzyme: ascorbic acid oxidase is the enzyme that catalyzes the oxidation of ascorbic acid, while phosphoglucoisomerase catalyzes the isomerization of glucose 6-phosphate to fructose 6-phosphate.

STUDY PROBLEM:

6.9. Suggest the function of each of the following enzymes:

(a) an acetyltransferase;
(b) phenylalanine hydroxylase;
(c) pyruvate dehydrogenase.

Table 6.5. Some cofactors that contain vitamins

Name of cofactor	**Vitamin needed**	**Structure of vitamin**
Vitamin C (ascorbic acid)	Vitamin C	
Vitamin B_1 (Thiamine)	Vitamin B_1	
biotin	Biotin	
Coenzyme A	Pabtothenic acid	 $\underset{H_3C}{\overset{H_3C}{HOCH_2C}}-\underset{OH}{CH}\overset{O}{\overset{\Vert}{C}}NHCH_2CH_2CO_2H$
NAD^{+a}	Nicotinic acid (niacin) or nicotinamide (niacinamide)	
Pyridoxyl phosphate	Pyridoxyl	

Nicotinamide adenine dinucleotide, a biological oxidizing agent.

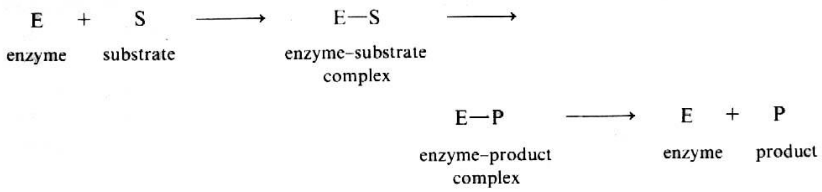
B. How enzymes work:

Some enzymes have been studied in detail, yet there is still much to learn about even the well-known enzymes. It is believed that an enzyme fits itself around the substrate (the molecule to be acted upon) to form an enzyme-substrate complex. The bonds of the substrate may be strained by attractions between itself and the enzyme. Strained bonds are of higher energy and are more easily broken; therefore, the desired reaction proceeds easily and yields an enzyme-product complex.

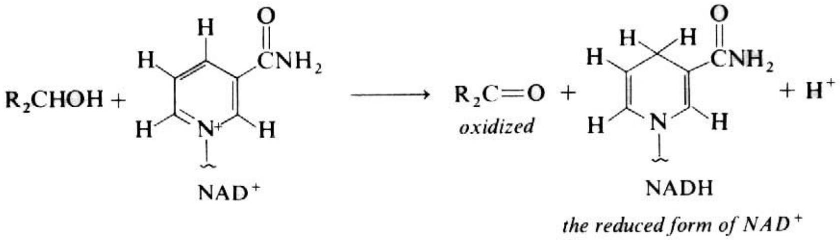
In many cases, the product is not the same shape as the reacting substrate; thus the fit between the product and the enzyme is no longer perfect. The altered shape of the product causes a dissociation of the complex, and the enzyme surface is ready to accept another molecule of substrate. This theory of enzyme activity is called the induced-fit theory.

$$\underset{\text{enzyme}}{F} + \underset{\text{substrate}}{S} \rightarrow \underset{\substack{\text{enzyme-substrate}\\\text{complex}}}{F - S} \rightarrow$$

$$\underset{\substack{\text{enzyme-product}\\\text{complex}}}{E - P} \rightarrow \underset{\text{enzyme}}{F} + \underset{\text{product}}{P}$$



Enzymes have molecular weights of 12,000-120,000 and higher. Most substrates (for example, an amino acid or a unit of glucose) are much smaller molecules. The specific location of the large enzyme structure where reaction occurs is called the active site. This site is where the prosthetic group (if any) is located. Metallic prosthetic groups are thought to serve as electrophilic agents and, in this way, catalyze the desired reactions. In NAD^+, the active site is the nicotinamide end of the cofactor. NAD^+ is readily reduced and therefore catalyzes oxidation reactions.



The rest of the enzyme molecule is not simply excess molecular weight! It is believed that this portion of the enzyme recognizes its substrate and holds it in place. It was suggested in the 1890's by Emil Fischer that enzymes are chiral molecules and that reactants must complement this chirality in order to undergo reaction. Fischer compared the fitting together of the substrate structure and the enzyme structure to a key fitting into a lock (see Figure 6.6).

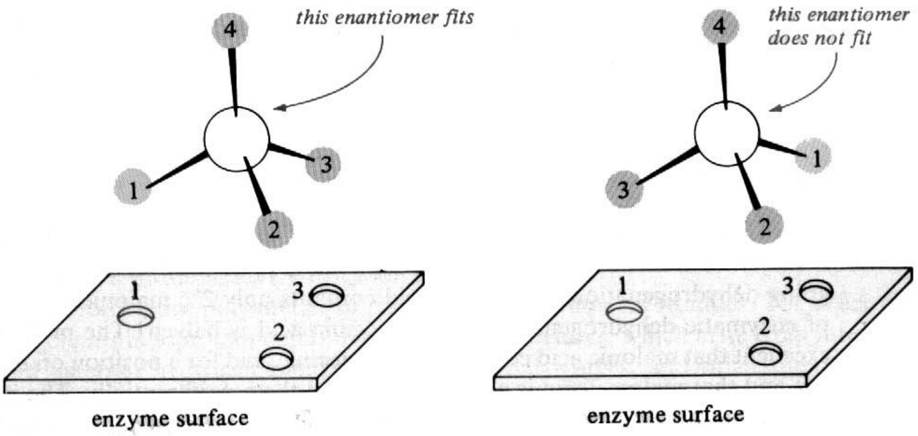
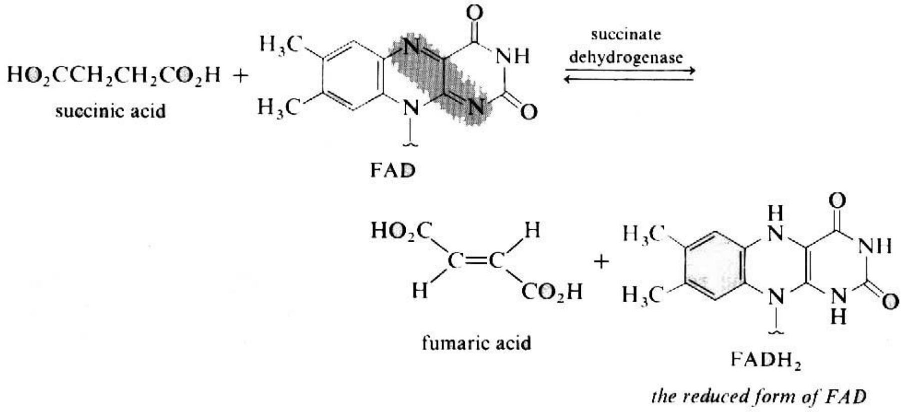


FIGURE 6.6. One enantiomer fits on the enzyme surface; its mirror image does not.

Recognition may occur by a series of dipole-dipole interactions, by hydrogen bonding, or by covalent bonding, in which the stereochemistry must be just right. In some cases, the rest of the enzyme molecule is folded to form a hydrophobic pocket that holds a nonpolar portion of the substrate in place. (We mentioned this type of structure for hemoglobin.) If the nonpolar end of a potential substrate does not fit in the pocket correctly, enzyme catalysis diminishes or is nonexistent. Therefore, the functional group to be acted upon must fit the active site on the enzyme, and the rest of the substrate molecule must fit together with other portions of the enzyme molecule for reaction to proceed. This dual type of recognition is the basis of the unique specificity of most enzymes.

Both the active site and the rest of an enzyme are important in enzyme activity. Let us look at one reaction in which the active site seems to be the more important factor in substrate recognition. The enzyme succinate dehydrogenase catalyzes the dehydrogenation of succinic acid to the trans-diacid fumaric acid. (The cisisomer, maleic acid, is not produced in this reaction.) The oxidizing agent in this reaction isflavin adenine dinucleotide (FAD), which is reduced by a 1,4-addition of two hydrogen atoms (plus two electrons). (We show only the functional portion of FAD here.)



Other diacids, such as oxalic acid, malonic acid, and glutaric acid, inhibit the dehydrogenation of succinic acid.

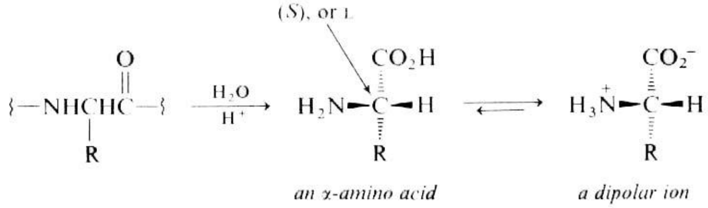
Inhibitors of succinate dehydrogenase:

$HO_2C—CO_2H$	$HO_2CCH_2CO_2H$	$HO_2CCH_2CH\ ,CH_2CO_2H$
oxalic acid	malonic acid	glutaric acid

Of these diacids, malonic acid has the greatest inhibiting effect. Malonic acid is very similar in structure to succinic acid, but is structurally incapable of undergoing dehydrogenation. If succinic acid contains only 200 malonic acid, the rate of enzymatic dehydrogenation of the succinic acid is halved! The probability is excellent that malonic acid competes with succinic acid for a position on the active site and that malonic acid is attracted and held there preferentially. The presence of malonic acid on the active site thus blocks the approach of succinic acid.

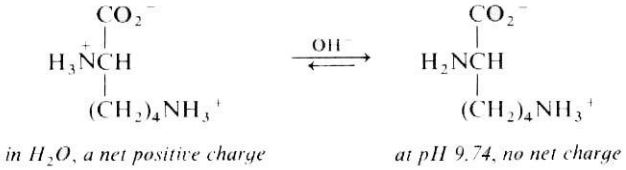
Summary:

A protein is a polyamide. Hydrolysis yields α-amino acids of (S)-configuration at the α carbon. Amino acids undergo an internal acid-base reaction to yield dipolar ions.



Essential amino acids are those that cannot be synthesized by an organism and must be present in the diet. Acidic amino acids are those with a carboxyl group in the side chain (R in the preceding equation). Basic amino acids contain an amino group in the side chain. Neutral amino acids contain neither -CO_2H nor -NH_2 in the side chain, but may contain OH, SH, or other polar group. Cross-linking in proteins may be provided by the SH group in cysteine, which can link with another SH in an oxidation reaction: 2 RSH → RSSR ± 2 H.

The isoelectric point of an amino acid is the pH at which the dipolar ion is electrically neutral and does not migrate toward an anode or cathode. The iso-electric point depends on the acidity or basicity of the side chain.



Racemic amino acids may be synthesized by a variety of routes.

A peptide is a polyamide of fewer than 50 amino acid residues. The N-terminal amino acid is the amino acid with a free cc-amino group, while the C-terminal amino acid has a free carboxyl group at carbon 1. End-group analysis to determine the C- and N-terminals and partial hydrolysis to smaller peptides are two techniques for peptide structure-determination.

In the synthesis of a peptide, reactive groups (except for the groups desired to undergo reaction) must be blocked. A carbamate group may be used to protect an amino group. A solid-phase peptide synthesis provides a blocking group for the C-terminal carboxyl group. The biosynthesis of proteins is accomplished by RNA. The order of incorporation of amino acids is determined by the order of attachment of the bases (N-heterocycles) in mRNA.

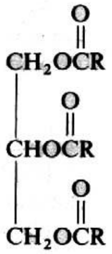
Proteins are polyamides of more than 50 amino acid residues. The order of side chains in a protein determines its higher structures, which arise from internal and external hydrogen bonding, van der Waals forces, and other interactions between side chains. The higher structures of proteins give them a variety of physical and chemical properties so that they may perform a variety of functions.

Denaturation is the disruption of hydrogen bonds and thus the disruption of the higher structure of the protein.

Enzymes are proteins that catalyze biochemical reactions. Enzymes are efficient and specific in their catalytic action. The specificity is provided for by the unique shape and by the polar (or nonpolar) groups contained within the enzyme structure. Some enzymes work in conjunction with a nonprotein cofactor, which may be organic or inorganic.

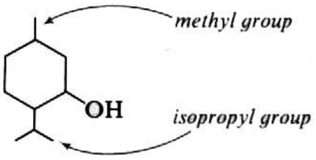
Lipids and Related Natural Products

A lipid: is defined as a naturally occurring organic compound that is insoluble in water, but soluble in nonpolar organic solvents such as a hydrocarbon or diethyl ether. This definition sounds as if it might include many types of compounds, and indeed it does. The various classes of lipids are related to one another by this shared physical property; but their chemical, functional, and structural relationships, as well as their biological functions, are diverse. We will discuss here the classes usually thought of as lipids: fats and oils, terpenes, steroids, and a few other compounds of interest. (Line formulas are generally used for terpenes and steroids, as the following examples show.



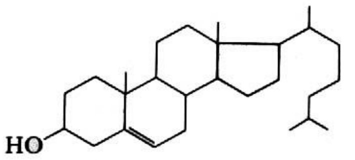
$$\begin{array}{l} CH_2O\overset{O}{\overset{\|}{C}}R \\ | \\ CHO\overset{O}{\overset{\|}{C}}R \\ | \\ CH_2O\overset{O}{\overset{\|}{C}}R \end{array}$$

a fat, or triglyceride



menthol

a terpene



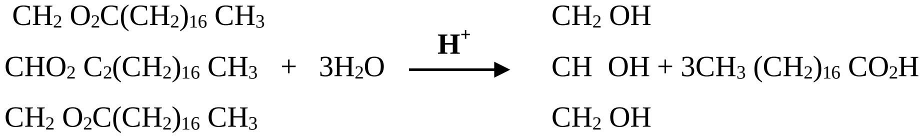
cholesterol

a steroid

Fats and Oils:

Fats and oils are triglycerides, or triacylglycerols, both terms meaning "triesters of glycerol." The distinction between a fat and an oil is arbitrary: at room temperature a fat is solid and an oil is liquid. Most glycerides in animals are fats, while those in plants tend to be oils; hence the terms *animal fats* (bacon fat, beef fat) and *vegetable oils* (corn oil, safflower oil).

The carboxylic acid obtained from the hydrolysis of a fat or oil, called a fatty acid, generally has a long, unbranched hydrocarbon chain. Fats and oils are often named as derivatives of these fatty acids. For example, the tristearate of glycerol is named tristearin, and the tripalmitate of glycerol is named tripalmitin.



$$\begin{array}{l} CH_2\,O_2C(CH_2)_{16}\,CH_3 \\ CHO_2\,C_2(CH_2)_{16}\,CH_3 \\ CH_2\,O_2C(CH_2)_{16}\,CH_3 \end{array} + 3H_2O \xrightarrow{H^+} \begin{array}{l} CH_2\,OH \\ CH\;\;OH \\ CH_2\,OH \end{array} + 3CH_3\,(CH_2)_{16}\,CO_2H$$

Tristearin
(glyceryl tristarate)
a typical fat

glycerol

stearic acid
a fatty acid

Fatty acids can also be obtained from waxes, such as beeswax. In these cases, the fatty acid is esterified with a simple long-chain alcohol.

$C_{25}H_{51}CO_2C_{28}H_{51}$	$C_{27}H_{55}CO_2C_{32}H_{65}$	$C_{15}H_{31}CO_2C_{16}H_{33}$
in beeswax	in carnauba wax	cetyl palmitate in spermaceti

Most naturally occurring fats and oils are *mixed* triglycerides-that is, the three fatty-acid portions of the glyceride are not the same. Table 7.1 lists some representative fatty acids, and Table 7.2 shows the fatty-acid composition of some plant and animal triglycerides.

Table 7.1. Selected fatty acids and their sources

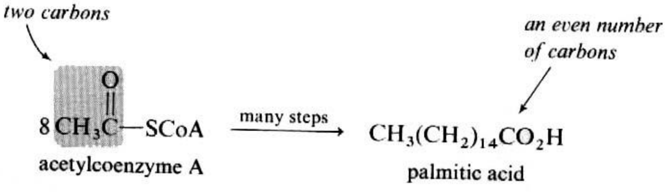
Name of acid	Structure	Source
Saturated:		
butyric	$CH_3(CH_2)_2CO_2H$	milk fat
Palmitic	$CH_3(CH_2)_{14}CO_2H$	animal and plant *fat*
Stearic	$CH_3(CH_2)_{16}CO_2H$	animal and plant fat
Unsaturated:		
Palmitoleic	$CH_3(CH_2)_5CHCH(CH_2)_7CO_2H$	animal and plant fat
Oleic	$CH_3(CH)_7CH=CH(CH_2)_7CO_2H$	animal and plant fat
Linoleic	$CH_3(CH_2)_4CHCHCH_2CHCH(CH)_7CO_2H$	plant oils
Linolenic	$CH_3CH_2CHCHCH_2CHCHCH_2CHCH(CH_2)_7CP_2H$	Linseed oil

Table7.2. Approximate fatty-acid composition of some common fats and oils

Composition (%)[a]

Source	*Palmitic*	*Stearic*	*Oleic*	Linoleic
corn oil	*10*	*5*	*45*	*38*
soybean oil	*10*	--	*25*	*55*
Lard	*30*	*15*	*45*	*5*
Butter	*25*	*10*	*35*	--
human fat	*25*	*8*	*46*	*10*
Other fatty acids are also found in lesser amounts.				

Almost all naturally occurring fatty acids have an *even* number of carbon atoms because they are biosynthesized from the two-carbon acetyl groups in acetylcoenzyme A.



$$8\,CH_3\overset{\overset{O}{\|}}{C}-SCoA \xrightarrow{\text{many steps}} CH_3(CH_2)_{14}CO_2H$$

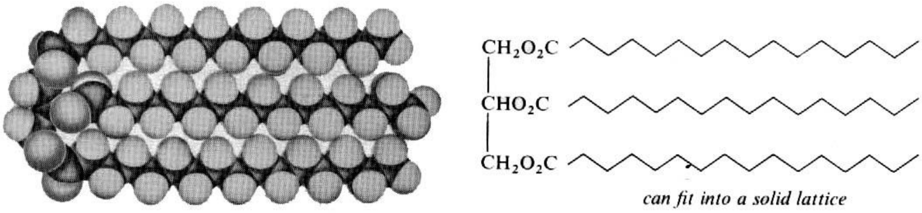
The hydrocarbon chain in a fatty acid may be saturated or it may contain double bonds. The most widely distributed fatty acid in nature, oleic acid, contains one double bond. Fatty acids with more than one double bond are not uncommon, particularly in vegetable oils; these oils are the so-called *polyunsaturates.*

The configuration around any double bond in a naturally occurring fatty acid is *cis*, a configuration that results in the low melting points of oils. A saturated fatty acid forms zigzag chains that can fit compactly together, resulting in high van der Waals attractions; therefore, saturated fats are solids. If a few *cis* double bonds are present in the chains, the molecules cannot form neat, compact lattices, but tend to coil; polyunsaturated triglycerides tend to be oils. Figure 7.1 shows models of the two types of chains.

Triglycerides are one of the three principal foodstuffs, carbohydrates and proteins being the other two. As an energy source, triglycerides are the most efficient: they provide 9.5 kcal/gram, while the proteins provide 4.4 kcal/gram and the carbohydrates provide 4.2 kcal/gram.

In an organism, ingested fats are hydrolyzed into monoglycerides, di-glycerides, fatty acids, and glycerol, all of which can be absorbed through the intestinal wall. The organism (1) uses these hydrolyzed or partially hydrolyzed fats as raw materials to synthesize its own fats; (2) converts the fatty acids to other compounds such as carbohydrates or cholesterol; or (3) converts the fatty acids to energy.

A saturated triglyceride:



An unsaturated triglyceride:

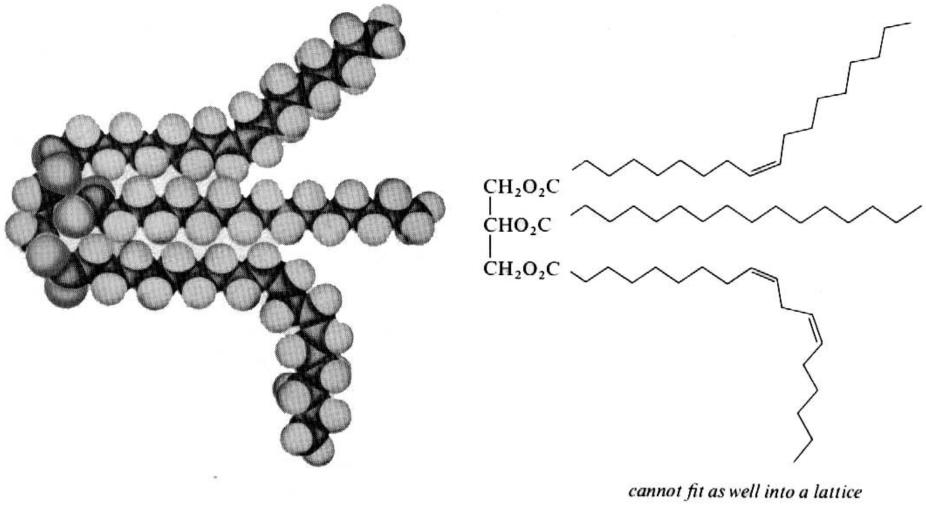
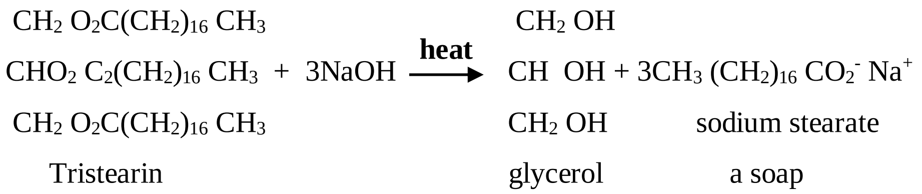


Figure 7.1. The shapes of saturated and unsaturated triglycerides. Adapted from William H. Brown and Judith A. McClarin, Introduction to Organic and Biochemistry, 3rd ed. (Willard Grant Press, Boston, 1981).

SECTION 7.2.

Soaps and Detergents:

The word *saponjfy* means “make soap.” Saponification of an ester with NaOH yields the sodium salt of a carboxylic acid. Saponification of a triglyceride yields a salt of a long-chain fatty acid, which is a soap. American pioneers used beef or pork fat and wood ashes (which contain alkaline salts, such as K_2C0_3) to make soap. (It was reported by Julius Caesar that Teutonic tribes of his era also made soap this way.)



$$\begin{array}{lcll} CH_2\,O_2C(CH_2)_{16}\,CH_3 & & CH_2\,OH & \\ CHO_2\,C_2(CH_2)_{16}\,CH_3 + 3NaOH & \xrightarrow{\textbf{heat}} & CH\ \ OH + 3CH_3\,(CH_2)_{16}\,CO_2^-\,Na^+ & \\ CH_2\,O_2C(CH_2)_{16}\,CH_3 & & CH_2\,OH & \text{sodium stearate} \\ \text{Tristearin} & & \text{glycerol} & \text{a soap} \end{array}$$

A molecule of a soap contains a long hydrocarbon chain plus an ionic end. The hydrocarbon portion of the molecule is hydrophobic and soluble in nonpolar substances, while the ionic end is hydrophilic and water-soluble. Because of the hydrocarbon chain, a soap molecule as a whole is not truly soluble in water. However, soap is readily suspended in water because it forms micelles, clusters of hydrocarbon chains with their ionic ends facing the water (see Figure 7.2).

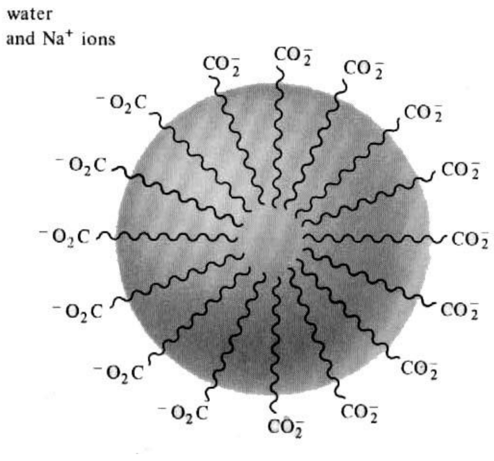
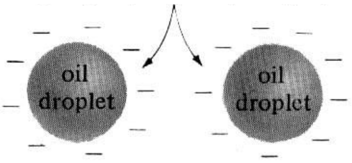


Figure 7.2. A micelle of the alkylcarboxylate ions of a soap.

The value of a soap is that it can emulsify oily dirt so that it can be rinsed away. This ability to act as an emulsifying agent arises from two properties of the soap. First, the hydrocarbon chain of a soap molecule dissolves in nonpolar substances, such as droplets of oil. Second, the anionic end of the soap molecule, which is attracted to water, is repelled by the anionic ends of soap molecules protruding from other drops of oil. Because of these repulsions between the soap-oil droplets, the oil cannot coalesce, but remains suspended.


in soapy water, oil droplets repel each other because of similar charges of soap's carboxylate groups

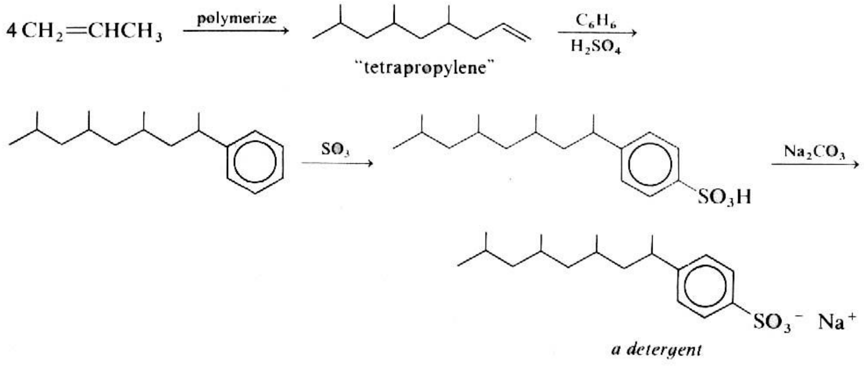
A disadvantage of soaps is that they form insoluble salts (bathtub ring) with Ca^{2+}, Mg^{2+} and other ions found in hard water. ("Softening" water involves exchanging these ions for Na^+)

$$2\ CH_3(CH_2)_{16}CO_2Na + Ca^{2+} \longrightarrow [CH_3(CH_2)_{16}CO_2]_2Ca + 2\ Na^+$$

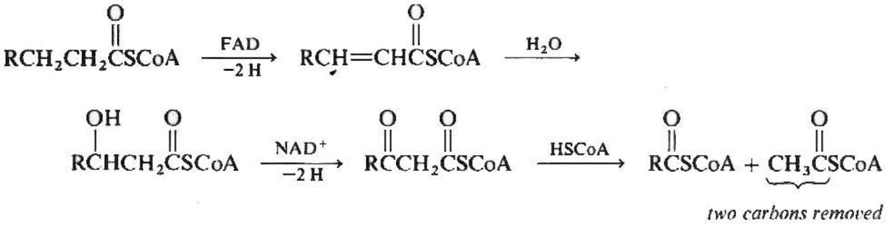
sodium stearate calcium stearate
insoluble

Most laundry products and many toilet “soaps” and shampoos are not soaps, but detergents. A detergent is a compound with a hydrophobic hydrocarbon end plus a sulfonate or sulfate ionic end. Because of this structure, a detergent has the same emulsifying properties as a soap. The advantage of a detergent is that most metal alkylsulfonates and sulfates are water-soluble; detergents do not precipitate with the metal ions found in hard water.

One of the first detergents in common use was a highly branched alkylbenzenesulfonate. The alkyl portion of this compound is synthesized by the polymerization of propylene and is attached to the benzene ring by a Friedel-Crafts alkylation reaction. Sulfonation, followed by treatment with base, yields the detergent.



Although the microorganisms in septic tanks or sewage-treatment plants can break down continuous-chain alkyl groups into smaller organic molecules, they cannot degrade branched chains. The reason for this difference in biodegradability is that long-chain hydrocarbons are degraded two carbons at a time by way of a keto ester. Branching interferes with the formation of the ketone group, and thus blocks the entire sequence. (FAD, NAD^+ and HSCoA, shown in the following equation, are discussed in Sections 13.8 and 19.14B.)



$$RCH_2CH_2\overset{O}{\overset{\|}{C}}SCoA \xrightarrow[-2\,H]{FAD} RCH{=}CH\overset{O}{\overset{\|}{C}}SCoA \xrightarrow{H_2O}$$

$$R\overset{OH}{\overset{|}{C}}HCH_2\overset{O}{\overset{\|}{C}}SCoA \xrightarrow[-2\,H]{NAD^+} R\overset{O}{\overset{\|}{C}}CH_2\overset{O}{\overset{\|}{C}}SCoA \xrightarrow{HSCoA} R\overset{O}{\overset{\|}{C}}SCoA + \underbrace{CH_3\overset{O}{\overset{\|}{C}}SCoA}_{\textit{two carbons removed}}$$

To prevent the build-up of detergents in rivers and lakes, present-day detergents are designed with biodegradability in mind. One type of biodegradable detergent is an alkylbenzenesulfonate with a continuous-chain, rather than a branched-chain, alkyl group. Another type of biodegradable detergent is a continuous-chain alkylsulfate.